Physics for particle detectors and particle detectors for physics

Timing performance of semiconductor detectors with internal gain and constraints on high-scale interactions of the Higgs boson

Philipp Jonas Windischhofer

Merton College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity Term 2022

Abstract

Experimental particle physics is a science of many scales. A large number of physical processes spanning energies from meV to TeV must be understood for modern collider experiments to be designed, built, and conducted successfully. This thesis contributes to the understanding of phenomena across this entire dynamic range.

Low-energy physics governs the operation of particle detectors, limits their performance, and guides the development of novel instrumentation. To formalise these aspects, classical electrodynamics is used to derive a general description of the formation of electrical signals in detectors, and ideas from quantum mechanics are applied to the study of charge avalanche amplification in semiconductors. These results lead to a comprehensive analytical characterisation of the time resolution and the efficiency of single-photon avalanche diodes, and isolate the most important design variables. They also reveal the applicability of these devices in precision timing detectors for charged particles, which is experimentally verified in a high-energy hadron beam.

Large detector systems at hadron colliders probe fundamental physics at the energy frontier. Data collected with the ATLAS detector during Run 2 of the Large Hadron Collider are used to measure the cross-section for the production of a Higgs boson together with an electroweak boson as a function of the kinematic scale of the process. This measurement provides the finest granularity available to date for this process. It is highly informative of the structure of interactions beyond the direct kinematic reach of the experiment, and new limits are set on the couplings of such interactions within an effective field theory.

Preface

The following is a story about two different kinds of physical laws. It is a story about the symbiotic relationship between low-energy processes and the fundamental dynamics of nature at extremely high energy scales, between the known physical laws and those that are yet to be discovered, and also between the physics of the past and that of the future.

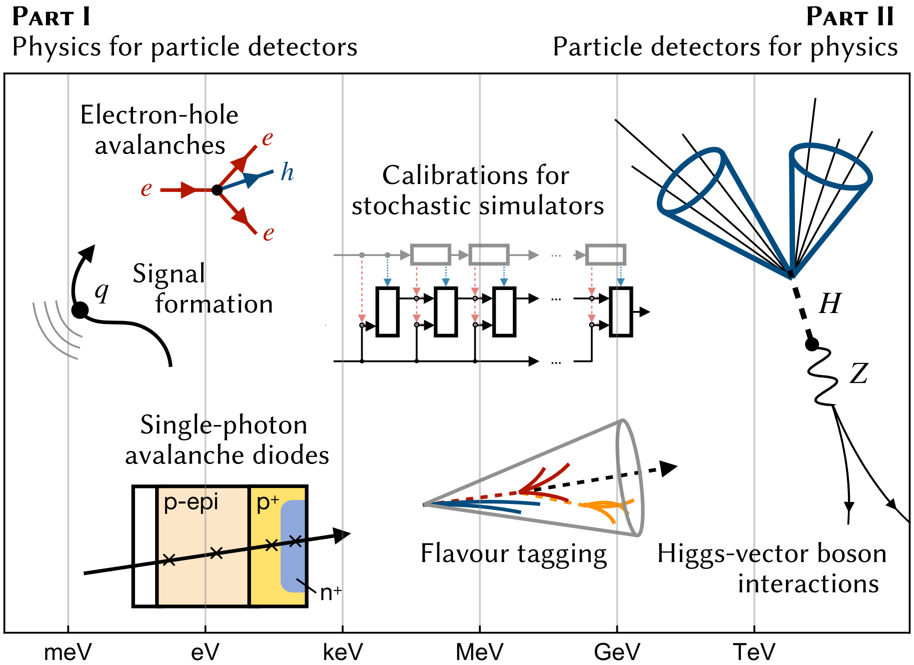
The subject of this thesis is the experimental science of fundamental physics, the study of physical processes that occur at the highest energies and the smallest distances presently accessible in the laboratory. Human experimenters, however, are macroscopic entities. And so, bridging the gap that separates us from the energy frontier requires the mastery of physics at many scales. It involves the construction of complex experimental apparatuses, which *themselves* can be described by the laws of physics—laws that once were at the forefront of our knowledge are now exploited to catalyse the discoveries that still lie ahead.

This document is divided into two parts, illustrated by the figure on the next page.

Part I: Physics for particle detectors

This first half begins at very low energies. It treats modern particle detectors as well-defined physical systems and studies the processes that underpin their functioning. The investigation uses analytical calculations to expose the essence of these phenomena, simulations to study them in more detail, and experimental measurements to validate the conclusions.

Chapter 1 treats the electrodynamics of the formation of signals in detectors. Published in Refs. [1, 2], my contributions led to the formulation of a very general signal theorem covering all devices that detect fields or radiation from charged particles and opening up new possibilities for the simulation of antenna arrays used in cosmic-ray physics. The mathematical high-water mark is reached in Chapter 2, which studies the physics of a powerful signal-amplification mechanism, charge avalanche multiplication in semiconductors. Using methods from stochastic mechanics, I developed a comprehensive analytical theory of electron-hole avalanches, capable of describing their entire evolution. A part of this work is published in Ref. [3] and another publication is in preparation.



Following these principles, I have contributed to studying the properties of single-photon avalanche diodes [4]. As Chapter 3 shows theoretically, these devices are not only capable of detecting individual photons with superb time resolution but can also directly detect charged particles without requiring external scintillators. Chapter 4 investigates this claim experimentally by exposing prototype devices to a high-energy hadron beam. For this campaign, I designed the high-bandwidth electronics necessary for the readout of the detector. I also devised and assembled the experimental setup, operated it during the data-taking period, and analysed the data. This is more recent work and a dedicated publication is forthcoming.

Part II: Particle detectors for physics

This second part describes the exploration of the interactions of the Higgs boson at very high energy scales, using the data set recorded by the ATLAS detector at the Large Hadron Collider. Chapters 5 and 6 briefly review the theoretical and experimental foundations on which this work is based.

Chapter 7 presents a measurement of the associated production of a Higgs boson H with an electroweak boson V, followed by the decay $H \to b\bar{b}$. This work is published in Ref. [5]. I performed the statistical analysis to measure the VH cross-section as a function of the kinematic scale of the process and refined and improved the signal model. I led the effort to interpret these measurements in terms of constraints on the couplings of effective interactions of mass-dimension six in the

Warsaw basis, leading to the first such constraints from a VH measurement. I developed software tools for the parameterisation of the effects of these modifications on the experimental observables and for the statistical inference. These have also been used in a $VH(\to b\bar{b})$ measurement specifically targeting high vector-boson transverse momenta [6] as well as in a search for the $VH(\to c\bar{c})$ process [7]. I was also responsible for the projection of the sensitivity to $VH(\to b\bar{b})$ expected at the end of the high-luminosity phase of the Large Hadron Collider, which is published in Ref. [8].

Finally, Chapter 8 presents a combination of two $VH(\to b\bar{b})$ measurements which together allow the VH cross-section to be measured across a wide kinematic scale with the highest granularity available to date. For this work, published in Ref. [9], I was one of three principal analysers and editor of the internal documentation. I contributed significantly to the design of the combination strategy and the development of the statistical model.

Everything in between: connecting the scales

The path that connects low-energy detector information to the details of physics at the energy frontier is often obstructed by processes occurring at intermediate scales.

These hurdles are particularly prominent in jet flavour tagging, the art of identifying the nature of the high-energy particle at the origin of a collimated cascade of hadrons. Within the ATLAS flavour-tagging group, I developed and maintained a simulation-driven method which allows existing algorithms for the identification of b-jets to be applied across the entire kinematic spectrum contained in the data. This work directly benefited the di-b-jet resonance search published in Ref. [10], for which I was granted exceptional authorship. The method itself, published in Ref. [11], is also part of the recommended flavour-tagging procedures used within the Collaboration, and I directly contributed to several such releases. Sophisticated simulations, together with modern machine-learning techniques, allow powerful summary statistics to be constructed, often combining many observables across different detector systems. I am the main author of Ref. [12], which introduced a new method to render such discriminants insensitive to certain aspects of the underlying input data. These discriminants, however, cannot be better than the—imperfect—simulations used to define them. Building on the theory of optimal transportation, I co-developed a very versatile scheme for the correction of such deficiencies in Ref. [13], taking the next step towards exploiting the full potential of our instrumentation. None of this work is documented here.

Contents

PART I

Physics for particle detectors

Chapter 1

The formation of electrical signals in particle detectors

The detection of moving charged particles is an important part of the experimental methodology in many branches of physics, oftentimes enabling direct access to detailed information about the physical system under study. The technologies invented for this purpose are as numerous and diverse as the investigated physical processes themselves.

In astronomy and astrophysics, the detailed study of active galactic nuclei (AGNs), accreting supermassive black holes, provides important information about the evolution of the early universe. Electrons and positrons gyrating in the intense magnetic fields of the accretion disk and the ejected jets are powerful sources of synchrotron radiation with a wide range of frequencies, part of which can be imaged with radio telescopes (Figure 1.1a) [14]. AGNs are also believed to be sources of ultra-high energy neutrinos, the detection of which requires very large instrumented volumes. A compelling technique monitors the Askaryan radiation [15] produced by neutrino-induced showers in the radio-transparent arctic ice by means of an antenna array [16] positioned close to the surface (Figure 1.1b).

Terrestrial high-energy scattering experiments allow fundamental physical processes to be studied in the laboratory. Particle accelerators and storage rings must be able to safely handle the required high-intensity particle beams, requiring fast diagnostic devices [17] such as beam-current transformers (Figure 1.1c). The energetic final-state particles produced in these scattering reactions must be measured with superb precision in space (in some cases also in time), and at very high rates. Modern semiconductor detectors, often based on silicon [18], provide the required perfor-

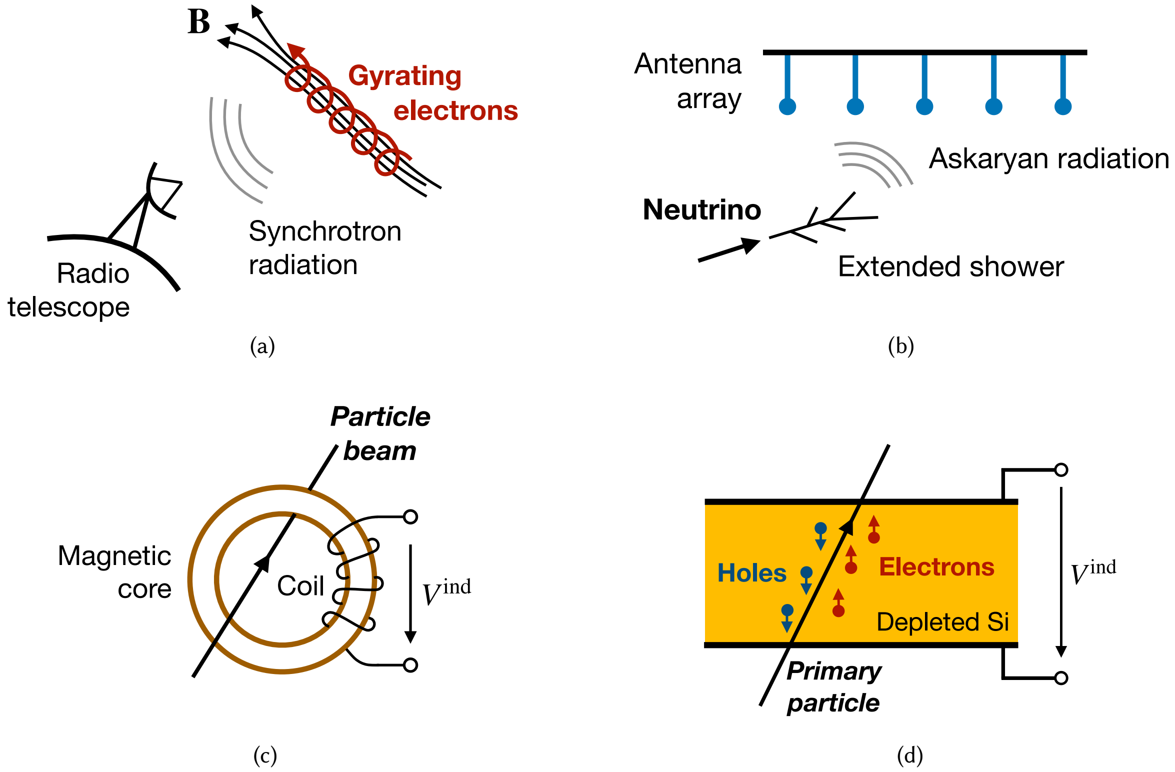


Figure 1.1: Detector technologies used in different areas of experimental physics to detect moving charged particles: (a) a radio telescope imaging the synchrotron emission of an AGN, (b) an array of antennas embedded in the arctic ice to detect the radio signature of neutrino-induced showers, (c) a beam-current transformer as a diagnostic device for storage rings, (d) a modern silicon detector for a collider experiment.

mance (Figure 1.1d).

There are countless other examples. In the situations mentioned above, the detector system delivers an electrical signal upon exposure to a moving charged particle. This signal, ultimately, is generated by the electromagnetic interaction between the primary particle and the detector. For radio telescopes and radio neutrino observatories, this coupling happens via electromagnetic radiation; beam current transformers detect the field of the passing particle beam, and silicon detectors record the drift of secondary charge carriers produced by ionisation of the detector material. As dissimilar as the practical implementations of these principles might appear, the physical origin of the detector signals is always the same: the classical dynamics of charges and fields.

The systematic, rigorous, and complete treatment of the signal formation is the subject of this chapter, starting from a description of the detector and the trajectory of the moving charged particle. The developed strategy is very physically intuitive and straightforward to implement numerically, yet no approximations are necessary and all electrodynamic effects are naturally captured. Maxwell's equations are written down in Section 1.1 as the starting point for the ensuing discussions. Reciprocity relations, almost as old as electrodynamics itself and at the heart of the method, are also reviewed there. Section 1.2 formalises the problem of charged-particle detection and provides its fully general solution. The nonrelativistic limit of signal formation is studied in detail in Section 1.3. Methods for the description of signals in detectors based on secondary ionisation (in gases or in solids) emerge effortlessly as special cases of the general result. Finally, Section 1.4 studies electro*dynamic* effects in the formation of detector signals in transmission lines and antenna arrays.

1.1 Reciprocity relations in classical electrodynamics

The discussion in this section is greatly simplified by keeping the treatment as general as possible, introducing approximations only at the very end. The electrodynamic equations for a general, anisotropic, linear material are a suitable starting point.

Such a material is characterised by a position- and frequency-dependent permittivity matrix $\hat{\boldsymbol{\epsilon}}(\mathbf{x},\omega)$, permeability matrix $\hat{\boldsymbol{\mu}}(\mathbf{x},\omega)$, and conductivity matrix $\hat{\boldsymbol{\sigma}}(\mathbf{x},\omega)$. These 3×3 matrices define the constitutive relations of the material, in which they relate the electric field[1] $\mathbf{E}(\mathbf{x},\omega)$ to the displacement field $\mathbf{D}(\mathbf{x},\omega)$ and the local current density $\mathbf{J}(\mathbf{x},\omega)$, and the magnetic field strength $\mathbf{H}(\mathbf{x},\omega)$ to the flux density $\mathbf{B}(\mathbf{x},\omega)$,

$$\begin{aligned}\mathbf{D}(\mathbf{x},\omega) &= \hat{\boldsymbol{\epsilon}}\mathbf{E}(\mathbf{x},\omega),\\ \mathbf{J}(\mathbf{x},\omega) &= \hat{\boldsymbol{\sigma}}\mathbf{E}(\mathbf{x},\omega),\\ \mathbf{B}(\mathbf{x},\omega) &= \hat{\boldsymbol{\mu}}\mathbf{H}(\mathbf{x},\omega).\end{aligned}$$

Maxwell's equations [19] connect these fields to their corresponding sources, the local charge density $\rho(\mathbf{x},\omega)$ and the externally imposed current density $\mathbf{J}^e(\mathbf{x},\omega)$. (The current density $\mathbf{J}^e$ is conceptually different from the current density $\hat{\boldsymbol{\sigma}}\mathbf{E}$. The latter is an automatic reaction of the material to a locally present electric field, whereas the former is entirely under the control of the experimenter and is imposed on the system. Both currents act, of course, as sources for the magnetic field.) Using SI units and taking a harmonic time dependence $e^{i\omega t}$ for all fields, Maxwell's equations read

$$\boldsymbol{\nabla}\cdot(\hat{\boldsymbol{\epsilon}}\mathbf{E}) = \rho, \tag{1.1a}$$

[1] Throughout this document, $SO(3)$ vectors are printed in boldface. The Euclidean norm of a vector is denoted by the unbolded symbol, e.g. $E = \|\mathbf{E}\|$. Individual components of a vectorial quantity are denoted by superscripts, e.g. E^x denotes the component of the vector $\mathbf{E}$ along the x-direction of a Cartesian coordinate system.

$$\nabla \cdot (\hat{\mu}\mathbf{H}) = 0, \tag{1.1b}$$

$$\nabla \times \mathbf{E} = -i\omega\hat{\mu}\mathbf{H}, \tag{1.1c}$$

$$\nabla \times \mathbf{H} = \mathbf{J}^e + \hat{\sigma}\mathbf{E} + i\omega\hat{\epsilon}\mathbf{E}. \tag{1.1d}$$

For a fixed material distribution, the resulting electric and magnetic fields are determined as a function of the sources ρ and $\mathbf{J}^e$; different sources, or different materials, generally lead to different field configurations.

1.1.1 Lorentz reciprocity

Two field configurations that emerge in the same material distribution (or two closely related material distributions) as a result of differing source terms are, however, not entirely independent. Maxwell's equations imply relations between the two situations, referred to as *reciprocity relations*. These relationships generally take the form of integral equations that involve the field distributions in the two cases. (The concept of reciprocity as a way to relate nonequivalent configurations of the same physical system has a long and colourful history. Refs. [20–23] show classical results in the context of electrodynamics. Classical mechanics also has reciprocal properties, as elucidated in Refs. [24–30].)

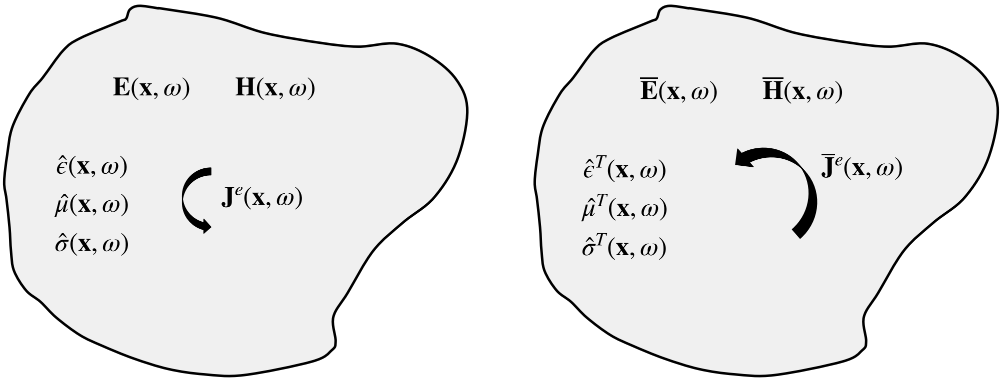
A very powerful such relation, the Lorentz reciprocity theorem [31], relates the two situations shown in Figure 1.2. An arbitrary external current density $\mathbf{J}^e(\mathbf{x}, \omega)$ is applied to a material distribution $(\hat{\epsilon}, \hat{\mu}, \hat{\sigma})$ in Figure 1.2a, and another arbitrary current density $\bar{\mathbf{J}}^e(\mathbf{x}, \omega)$ is applied to a material distribution with transposed response matrices $(\hat{\epsilon}^\mathsf{T}, \hat{\mu}^\mathsf{T}, \hat{\sigma}^\mathsf{T})$ in Figure 1.2b. The current $\mathbf{J}^e$ generates electric and magnetic fields $\mathbf{E}(\mathbf{x}, \omega)$ and $\mathbf{H}(\mathbf{x}, \omega)$, and $\bar{\mathbf{J}}^e$ generates fields $\bar{\mathbf{E}}(\mathbf{x}, \omega)$ and $\bar{\mathbf{H}}(\mathbf{x}, \omega)$. (The response matrices of most practical materials are symmetric or approximately symmetric. In these cases the material distributions of the two situations compared by the theorem are identical.)

The expression $\nabla \cdot (\mathbf{E} \times \bar{\mathbf{H}})$ may be used to derive the relationship between these two situations. Using Eqs. 1.1,

$$\begin{aligned}\nabla \cdot (\mathbf{E} \times \bar{\mathbf{H}}) &= \bar{\mathbf{H}} \cdot (\nabla \times \mathbf{E}) - \mathbf{E} \cdot (\nabla \times \bar{\mathbf{H}}) \\ &= -\mathbf{E} \cdot \bar{\mathbf{J}}^e - i\omega\bar{\mathbf{H}} \cdot \hat{\mu}\mathbf{H} - \mathbf{E} \cdot (\hat{\sigma}^\mathsf{T} + i\omega\hat{\epsilon}^\mathsf{T})\bar{\mathbf{E}}.\end{aligned}$$

Starting from $\nabla \cdot (\bar{\mathbf{E}} \times \mathbf{H})$ instead leads to

$$\nabla \cdot (\bar{\mathbf{E}} \times \mathbf{H}) = \mathbf{H} \cdot (\nabla \times \bar{\mathbf{E}}) - \bar{\mathbf{E}} \cdot (\nabla \times \mathbf{H})$$



(a) (b)

Figure 1.2: (a) An arbitrary current density $\mathbf{J}^e$ produces fields $\mathbf{E}$ and $\mathbf{H}$ in a situation with material distribution $(\hat{\boldsymbol{\epsilon}}, \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\sigma}})$. (b) Another current density $\overline{\mathbf{J}}^e$ leads to fields $\overline{\mathbf{E}}$ and $\overline{\mathbf{H}}$ in a related material $(\hat{\boldsymbol{\epsilon}}^\mathsf{T}, \hat{\boldsymbol{\mu}}^\mathsf{T}, \hat{\boldsymbol{\sigma}}^\mathsf{T})$.

$$= -\overline{\mathbf{E}} \cdot \mathbf{J}^e - i\omega \mathbf{H} \cdot \hat{\boldsymbol{\mu}}^\mathsf{T} \overline{\mathbf{H}} - \overline{\mathbf{E}} \cdot (\hat{\boldsymbol{\sigma}} + i\omega\hat{\boldsymbol{\epsilon}})\mathbf{E},$$

and subtracting these two expressions results in

$$\boldsymbol{\nabla} \cdot (\mathbf{E} \times \overline{\mathbf{H}} - \overline{\mathbf{E}} \times \mathbf{H}) = \overline{\mathbf{E}} \cdot \mathbf{J}^e - \mathbf{E} \cdot \overline{\mathbf{J}}^e.$$

Upon integration over a (finite) volume V with surface ∂V and an application of Gauss' theorem this becomes

$$\oint_{\partial V} d\mathbf{A} \cdot (\mathbf{E} \times \overline{\mathbf{H}} - \overline{\mathbf{E}} \times \mathbf{H}) = \int_V dV (\overline{\mathbf{E}} \cdot \mathbf{J}^e - \mathbf{E} \cdot \overline{\mathbf{J}}^e). \tag{1.2}$$

The domain of integration is now extended over all of space. If all sources are contained in a compact region (centred, say, on the origin of the coordinate system), the boundary integral on the left-hand side of Eq. 1.2 vanishes [32]. This can happen in a variety of ways. It is simplest to suppose that (as for any realistic situation) the material distributions reduce to vacuum as the boundary at infinity is approached. In this case, the asymptotic radiation fields are transverse electromagnetic waves that decay inversely with the distance to the source and whose field strengths are related as $\mathbf{H} \sim \hat{\mathbf{r}} \times \mathbf{E}$. The radial unit vector $\hat{\mathbf{r}}$ points from the position of the source to the observer. The integrand becomes proportional to $\mathbf{E} \times (\hat{\mathbf{r}} \times \overline{\mathbf{E}}) - \overline{\mathbf{E}} \times (\hat{\mathbf{r}} \times \mathbf{E})$, which vanishes identically.

This leads to the Lorentz reciprocity theorem [31]

$$\int_{\mathbb{R}^3} dV\, \overline{\mathbf{E}}(\mathbf{x}, \omega) \cdot \mathbf{J}^e(\mathbf{x}, \omega) = \int_{\mathbb{R}^3} dV\, \mathbf{E}(\mathbf{x}, \omega) \cdot \overline{\mathbf{J}}^e(\mathbf{x}, \omega), \tag{1.3}$$

which relates the external current distributions and the electric fields between the two situations.

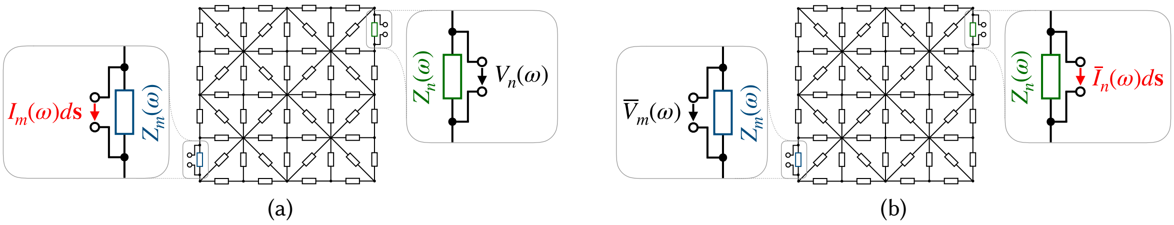


Figure 1.3: A network of N impedance elements $Z_1(\omega), \ldots, Z_N(\omega)$. (a) The current source $I_m(\omega)$ connected in parallel with $Z_m(\omega)$ produces a voltage $V_n(\omega)$ across $Z_n(\omega)$. (b) The current source $\bar{I}_n(\omega)$ connected in parallel with $Z_n(\omega)$ produces a voltage $\overline{V}_m(\omega)$ across $Z_m(\omega)$.

A detailed examination shows that $\mathbf{J}^e$ and $\bar{\mathbf{J}}^e$ must both be nonzero for the theorem to be useful. If one of the current densities vanishes identically, say, $\mathbf{J}^e \equiv 0$, the relation reduces to a tautology.

1.1.2 Network reciprocity

The Lorentz reciprocity relation in Eq. 1.3 leads to a host of interesting and useful results. Since the theorem holds for arbitrary linear material distributions $(\hat{\boldsymbol{\epsilon}}, \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\sigma}})$, it applies in particular to electrical circuits where a number of discrete but otherwise fully general impedance elements $Z_i(\omega)$ are connected with conducting wires. For practical circuits, the response matrices may also be taken to be symmetric, i.e. the reciprocity relation compares two different configurations of the same circuit.

Figure 1.3 shows a complicated circuit consisting of N arbitrary impedance elements Z_i, and connection pads are available at the locations of Z_m and Z_n. In order to simplify the following argument, the two terminals of each pair are taken to be in close proximity, i.e. separated only by an infinitesimal distance $d\mathbf{s}$. In Figure 1.3a, a current source is connected to the impedance element $Z_m(\omega)$ of the circuit, imposing a line current $I_m(\omega)d\mathbf{s}$. A voltage $V_n(\omega)$ appears across the terminals of Z_n in reaction to this current. It is given by $V_n(\omega) = d\mathbf{s} \cdot \mathbf{E}(\mathbf{x}_n, \omega)$, where $\mathbf{x}_n$ is the location of the terminal. In the situation of Figure 1.3b, a current source $\bar{I}_n(\omega)d\mathbf{s}$ is instead connected in parallel to the other element $Z_n(\omega)$, resulting in the voltage $\overline{V}_m(\omega) = d\mathbf{s} \cdot \overline{\mathbf{E}}(\mathbf{x}_m, \omega)$.

The two line currents I_m and $\bar{I}_n$ define the external current densities $\mathbf{J}^e$ and $\bar{\mathbf{J}}^e$ that appeared in Section 1.1.1. Of course, currents also flow throughout the network as dictated by Kirchhoff's rules. They are a consequence of the finite conductivity encoded in the material distribution that defines the circuit, and so not explicitly visible in the Lorentz reciprocity theorem. Applying Eq. 1.3 to this situation, the volume integrals reproduce the definitions of the voltages V_n and $\overline{V}_m$, and the

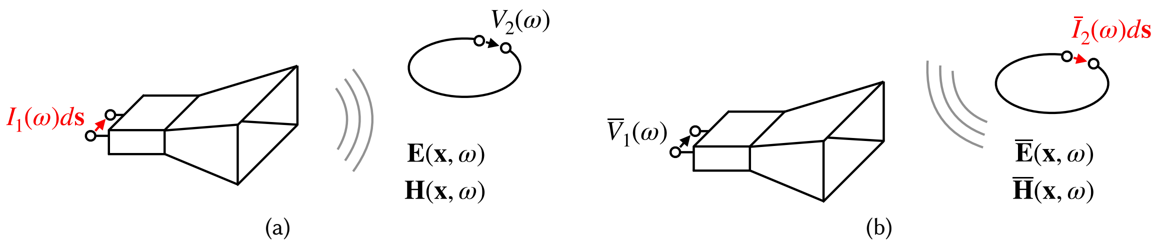


Figure 1.4: (a) A horn antenna is fed by a current source $I_1(\omega)$ connected to its terminals, inducing a voltage $V_2(\omega)$ across the terminals of a loop antenna. (b) The loop antenna is instead connected to a current source $\overline{I}_2(\omega)$, producing a voltage $\overline{V}_1(\omega)$ in the horn antenna.

relation simplifies to

$$I_m(\omega)\overline{V}_m(\omega) = \bar{I}_n(\omega)V_n(\omega). \tag{1.4}$$

This is called the network reciprocity theorem [19].

The object $T_{mn}(\omega) = V_m(\omega)/I_n(\omega)$ may be identified as the “transfer function” between nodes m and n of the circuit. It describes the voltage that can be observed at node m in reaction to a current injected at node n. In these terms, the network reciprocity theorem states that $T_{mn} = T_{nm}$, i.e. that the transfer function remains invariant if the positions of sender and receiver are exchanged.

1.1.3 Antenna reciprocity

The limit of infinitely large, densely connected, circuits leads to a continuous medium that allows the propagation of waves. It does not come as a surprise, therefore, that important reciprocity relations exist that relate the transmission and reception characteristics of antennas.

In Figure 1.4a, an external line current $I_1(\omega)d\mathbf{s}$ is impressed between the two terminals of a horn antenna. The resulting electromagnetic radiation induces a voltage $V_2(\omega)$ across the terminals of a loop antenna in its vicinity. Figure 1.4b reverses the roles of transmitter and receiver: the loop antenna is now connected to a current source $\overline{I}_2(\omega)d\mathbf{s}$, resulting in a signal $\overline{V}_1(\omega)$ being delivered by the horn antenna.

Applying the Lorentz reciprocity theorem in Eq. 1.3 to this situation and following the same steps as in Section 1.1.2 gives the antenna reciprocity relation

$$\frac{V_2(\omega)}{I_1(\omega)} = \frac{\overline{V}_1(\omega)}{\overline{I}_2(\omega)}. \tag{1.5}$$

This shows that the communication channel consisting of the two antennas and the surrounding medium is again symmetric. Since Eq. 1.5 holds for arbitrary antenna geometries and material dis-

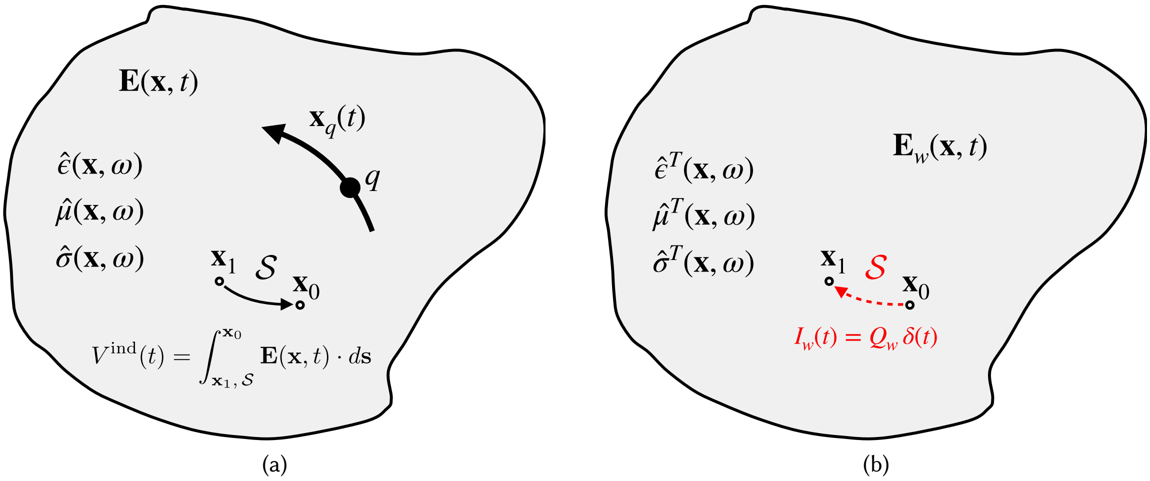


Figure 1.5: (a) A moving point charge produces an electric field $\mathbf{E}(\mathbf{x}, t)$ and therefore a voltage signal V^{ind} across the detector terminals at $\mathbf{x}_1$ and $\mathbf{x}_0$. (b) A line current $I_w(t) = Q_w\delta(t)$ flowing between the detector terminals produces the weighting field $\mathbf{E}_w(\mathbf{x}, t)$.

tributions, it also implies the remarkable result that the reception and transmission characteristics must be identical for each antenna *individually*.

1.2 Signals induced by a moving point charge: the general case

Using these reciprocity relations, a fully general result for the computation of the signal in a particle detector is almost immediate. The elements of the detector (electrodes, wires, cooling circuits) can again be absorbed into the material parameters $\hat{\boldsymbol{\epsilon}}(\mathbf{x}, \omega)$, $\hat{\boldsymbol{\mu}}(\mathbf{x}, \omega)$, $\hat{\boldsymbol{\sigma}}(\mathbf{x}, \omega)$, as shown in Figure 1.5a. A point charge q moves along an arbitrary trajectory $\mathbf{x}_q(t)$ in the vicinity of the detector, which generates the external current density

$$\mathbf{J}^e(\mathbf{x}, t) = q \cdot \dot{\mathbf{x}}_q(t)\delta(\mathbf{x} - \mathbf{x}_q(t)). \tag{1.6}$$

In response to the field $\mathbf{E}(\mathbf{x}, t)$ generated by this current, the detector delivers a signal through two terminals, located at positions $\mathbf{x}_1$ and $\mathbf{x}_0$. In a traditional particle detector $\mathbf{x}_1$ is a point on the signal electrode and $\mathbf{x}_0$ corresponds to the ground reference. The integral of the electric field along a particular path $\mathcal{S}$ connecting $\mathbf{x}_1$ and $\mathbf{x}_0$ is a quantity related to a voltage $V^{\mathrm{ind}}(\omega)$ between these two points. Parameterising $\mathcal{S}$ as $\mathbf{x}_s(s)$ with $\mathbf{x}_s(s_1) = \mathbf{x}_1$ and $\mathbf{x}_s(s_0) = \mathbf{x}_0$, this voltage is

$$V^{\mathrm{ind}}(\omega) = \int_{\mathbf{x}_1, \mathcal{S}}^{\mathbf{x}_0} d\mathbf{s} \cdot \mathbf{E}(\mathbf{x}, \omega). \tag{1.7}$$

The specification of the full path of integration in this definition is important: in general, $\boldsymbol{\nabla} \times \mathbf{E} \neq 0$, and there is no unique potential difference between the two points. Only once a particular path $\mathcal{S}$ is chosen does the voltage $V^{\text{ind}}(\omega)$ become unambiguous and well-defined. Eq. 1.7 may be regarded as the *definition* of the detector signal in which $\mathcal{S}$ plays the role of a prescription for the practical measurement of $V^{\text{ind}}(\omega)$.

Figure 1.5b shows a second electrodynamic configuration which is related to the situation described above through the Lorentz reciprocity theorem. The point charge is removed from this picture and a line current $I_w(\omega)$ is instead set to flow from $\mathbf{x}_0$ to the detector terminal at $\mathbf{x}_1$ along the same path $\mathcal{S}$. The current is embedded in the material distribution with transposed response matrices, as discussed in Section 1.1.1. (The current is defined to flow in the sense opposite to that in which the voltage V^{ind} is defined. This choice is made to respect commonly used sign conventions in Section 1.3.) The current $I_w(\omega)$ creates an electric-field distribution $\mathbf{E}_w(\mathbf{x}, \omega)$, referred to as the "weighting field".

Applying Eq. 1.3 to this pair of configurations leads to the relation

$$V^{\text{ind}}(\omega) = -\frac{1}{I_w(\omega)} \int_{\mathbb{R}^3} dV\, \mathbf{E}_w(\mathbf{x}, \omega) \cdot \mathbf{J}^e(\mathbf{x}, \omega), \tag{1.8}$$

where the current $\mathbf{J}^e$ is defined in Eq. 1.6. The line current $I_w(\omega)$ is arbitrary. The discussion simplifies if the current is chosen to be independent of the frequency ω, i.e. $I_w(\omega) = Q_w$, where Q_w has the dimension of a charge. This corresponds to a delta-like current pulse $I_w(t) = Q_w\delta(t)$ in the time domain.

Expressing Eq. 1.8 in the time domain and using Eq. 1.6 gives

$$V^{\text{ind}}(t) = -\frac{1}{Q_w} \int_{-\infty}^{\infty} dt' \int_{\mathbb{R}^3} dV\, \mathbf{E}_w(\mathbf{x}, t - t') \cdot \mathbf{J}^e(\mathbf{x}, t') \tag{1.9a}$$

$$= -\frac{q}{Q_w} \int_{-\infty}^{\infty} dt'\, \mathbf{E}_w(\mathbf{x}_q(t'), t - t') \cdot \dot{\mathbf{x}}_q(t'). \tag{1.9b}$$

This is a very general theorem for the computation of the detector signal V^{ind} that holds for arbitrary electrodynamic situations. It explicitly combines information about the particle trajectory with the properties of the detector, as encoded in the weighting field. (The weighting field may also be regarded as a Green's function for the detector signal.) As the derivation of Eq. 1.9 shows, $\mathbf{E}_w$ may be computed once and for all for a given detector geometry as the field response to a delta-like current pulse flowing along the signal-defining path. Analytic expressions for the weighting field exist for simple geometries, illustrated by the examples in Section 1.4. For realistic detectors, the weighting field must be obtained from numerical solutions of Maxwell's equations.

Eq. 1.9 allows the detector signal to be computed assuming that the particle trajectory is known. The trajectory depends in turn on the environment or the field configuration inside the detector itself, and must be determined separately.

The need for a physical regulator The Lorentz reciprocity theorem in Eq. 1.3 has been derived from Maxwell's equations assuming that the Fourier transforms of all participating objects exist and are smooth functions in ω. A detailed inspection shows that this assumption is violated in situations where charges are present in the interior of the domain in the far past or in the far future, i.e. at $t \to \pm\infty$. These technical complications can be avoided rather straightforwardly through the following regularisation procedure.

All participating charged particles are slowly brought in from spatial infinity during $-\infty < t < -T$, where the time $T > 0$ is chosen to be much larger than all physical timescales of the problem under consideration. The interactions between the charged particles and the detector occur in $-T < t < T$, during which $V^{\text{ind}}(t)$ may be computed using Eq. 1.9. All charges are then moved back out to spatial infinity during $T < t < \infty$.

It is needless to say that this regularisation prescription does not lead to any limitations in the scope of the situations that can be analysed. The procedure merely imitates what is anyways true in the real world, namely that detectors do not exist for an infinite period of time, and neither do charged particles in their vicinity. Section 1.3 below illustrates its application in several important example situations.

1.2.1 Induced signals in the presence of background fields

Particle detectors often contain electric or magnetic background fields $\mathbf{E}_{\text{bkg}}$ and $\mathbf{B}_{\text{bkg}}$ that are static on the timescales of the induced signal, such as bending fields in spectrometers or space charge fields in depleted silicon sensors. At a fundamental level, these background fields are generated by current and charge distributions $\mathbf{J}_{\text{bkg}}$ and ρ_{bkg}.

The total electric and magnetic fields are decomposed into static background fields and the contributions arising from the moving charged particle as $\mathbf{E} = \mathbf{E}_{\text{bkg}} + \mathbf{E}_{\text{signal}}$ and $\mathbf{B} = \mathbf{B}_{\text{bkg}} + \mathbf{B}_{\text{signal}}$. The constitutive equations of the material may depend nonlinearly on the background fields, e.g. in a saturated magnet return yoke. For most practically relevant cases, however, the back-reaction of the induced signal on the detector material can be neglected, and the material properties linearised around the background fields.

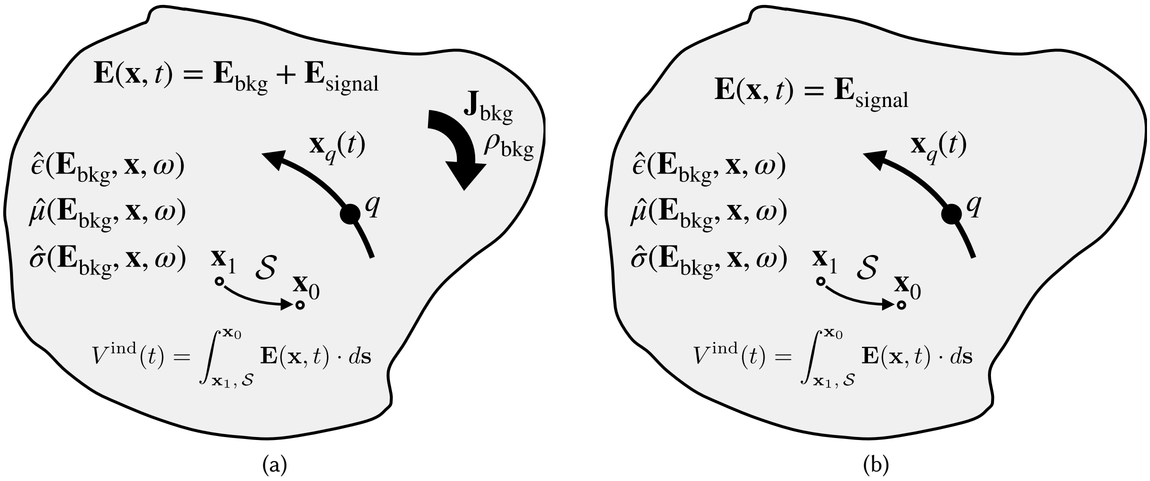


Figure 1.6: (a) Signal induced by a moving charged particle in the presence of background fields $\mathbf{E}_{\text{bkg}}$ and $\mathbf{B}_{\text{bkg}}$, created by current and charge distributions $\mathbf{J}_{\text{bkg}}$ and ρ_{bkg}. (b) The signal contribution that is due to the moving charge is computed in the absence of the background sources.

In this approximation, the signal-inducing charge propagates through a fixed material distribution with parameters $\hat{\boldsymbol{\epsilon}}(\mathbf{E}_{\text{bkg}}, \mathbf{B}_{\text{bkg}})$, $\hat{\boldsymbol{\mu}}(\mathbf{E}_{\text{bkg}}, \mathbf{B}_{\text{bkg}})$, and $\hat{\boldsymbol{\sigma}}(\mathbf{E}_{\text{bkg}}, \mathbf{B}_{\text{bkg}})$, shown in Figure 1.6a. The component of the signal caused by the moving point charge is computed as in Figure 1.6b. The explicit sources of the background fields are removed, but the material properties are kept at their original, background field-dependent, values. The signal theorem in Eq. 1.9 then carries over without modifications. Any static signal components arising solely from the background fields are added at the end of the calculation.

This demonstrates that the presence of background field is relevant for the computation of the detector signal only through a modification of the material properties used to obtain the weighting field. The sources $\mathbf{J}_{\text{bkg}}$ and ρ_{bkg} do not enter explicitly [33–35].

1.2.2 Processing and filtering of the detector signal

The induced signal V^{ind} is typically processed by a linear signal-processing chain (consisting of amplifiers, shapers, etc.) with an overall transfer function $G(\omega)$. The delivered output signal V^{out} is $V^{\text{out}}(\omega) = G(\omega)V^{\text{ind}}(\omega)$. It may be obtained rather efficiently without first having to compute, and then process, V^{ind}.

Choosing $I_w(\omega) = Q_w G(\omega)$ in Eq. 1.8 and transforming back to the time domain yields

$$V^{\text{out}}(t) = -\frac{q}{Q_w} \int_{-\infty}^{\infty} dt' \, \mathbf{K}_w(\mathbf{x}_q(t'), t - t') \cdot \mathbf{x}_q(t'), \tag{1.10}$$

where $\mathbf{K}_w$ is the weighting field computed as the response to the line current $I_w(t) = Q_w g(t)$, where $g(t)$ is the impulse response of the signal-processing chain. The transfer function $G(\omega)$ has a low-pass character for many relevant cases. Short time- and distance scales are thus effectively removed from the filtered weighting field $\mathbf{K}_w(\mathbf{x}, t)$, simplifying its computation with numerical methods.

1.3 Nonrelativistic limit

The signal theorem in Eq. 1.9 holds for arbitrary particle trajectories and includes radiation effects from accelerated charges. Many detectors, e.g. those based on ionisation in gases or solids, however operate in the nonrelativistic limit of signal formation. In case all charges move sufficiently slowly, the situation may be approximated by a sequence of purely electrostatic configurations and Eq. 1.9 specialises to several well-known results. These are important for the work presented in Chapters 2–4 and are thus briefly reviewed and summarised below.

1.3.1 Voltage induced on insulated electrodes

The simplest relevant case is that of a detector consisting of N perfectly conducting electrodes embedded in a material with (symmetric) permittivity matrix $\hat{\boldsymbol{\epsilon}}(\mathbf{x})$ and vanishing conductivity $\hat{\boldsymbol{\sigma}}(\mathbf{x})$. In the quasi-electrostatic limit, there are no magnetic effects and the permeability matrix $\hat{\boldsymbol{\mu}}(\mathbf{x})$ of the material becomes irrelevant. All electrodes are taken to be insulated, as shown in Figure 1.7a. The voltage V_n^{ind} measured between electrode n and ground is the detector signal. It does not depend on the chosen path $\mathcal{S}$.

To compute the weighting field for V_n^{ind}, a current $I_w(t) = Q_w \delta(t)$ is set to flow from the ground reference onto electrode n, as shown in Figure 1.7b. This current applies a charge Q_w to electrode n at $t = 0$, while all other electrodes remain uncharged. The resulting weighting field is the gradient of a scalar "weighting potential" $\phi_{w,n}$,

$$\mathbf{E}_w(\mathbf{x}, t) = -\boldsymbol{\nabla}\phi_{w,n}(\mathbf{x}; Q_w)\,\Theta(t), \tag{1.11}$$

where $\Theta(t)$ is the Heaviside step function. The notation emphasises that this potential corresponds to the situation where only electrode n carries a charge of Q_w.

A point charge q moving along a trajectory $\mathbf{x}_q(t)$ represents a charge density $\rho^e(\mathbf{x}, t) = q \cdot \delta(\mathbf{x} - \mathbf{x}_q(t))$ and a current density $\mathbf{J}^e(\mathbf{x}, t) = q \cdot \dot{\mathbf{x}}_q(t)\delta(\mathbf{x} - \mathbf{x}_q(t))$. Using the signal theorem in Eq. 1.9a

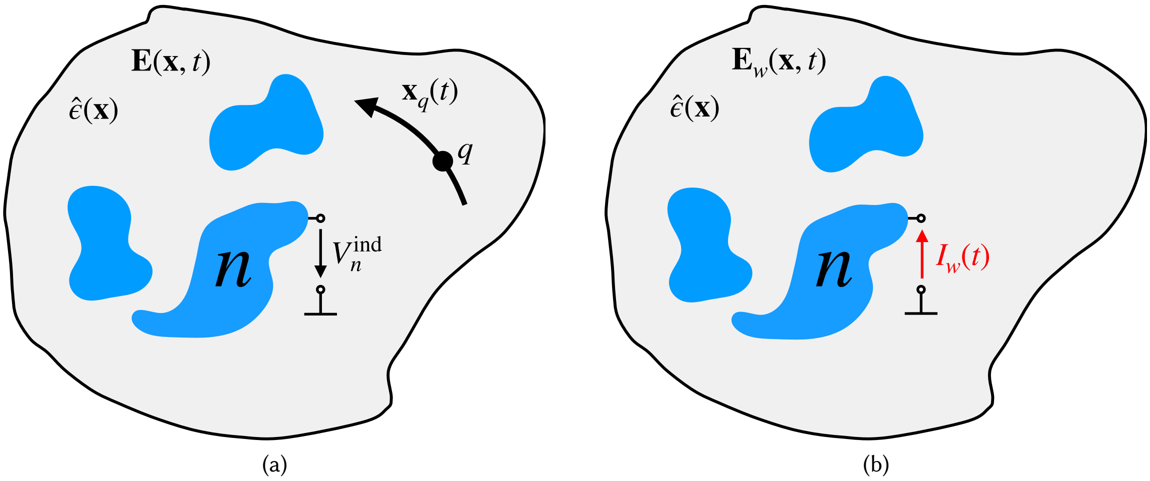


Figure 1.7: (a) Detector setup consisting of N insulated, perfectly conducting electrodes embedded in a material with permittivity $\hat{\epsilon}(\mathbf{x})$. A moving point charge induces a voltage V_n^{ind} on electrode n. (b) Situation to compute the weighting field $\mathbf{E}_w(\mathbf{x}, t)$ for the n-th electrode, which is produced by a current $I_w(t) = Q_w \delta(t)$.

and the weighting field in Eq. 1.11, V_n^{ind} becomes

$$\begin{aligned} V_n^{\text{ind}}(t) &= \frac{1}{Q_w} \int_{-\infty}^{\infty} dt' \Theta(t - t') \int_{\mathbb{R}^3} dV \; \boldsymbol{\nabla} \phi_{w,n}(\mathbf{x}; Q_w) \cdot \mathbf{J}^e(\mathbf{x}, t') \\ &= -\frac{1}{Q_w} \int_{-\infty}^{\infty} dt' \Theta(t - t') \int_{\mathbb{R}^3} dV \; \phi_{w,n}(\mathbf{x}; Q_w) \; \boldsymbol{\nabla} \cdot \mathbf{J}^e(\mathbf{x}, t'). \end{aligned}$$

With the continuity equation for the current $\mathbf{J}^e$,

$$\frac{\partial}{\partial t} \rho^e(\mathbf{x}, t) + \boldsymbol{\nabla} \cdot \mathbf{J}^e(\mathbf{x}, t) = 0,$$

and integrating by parts, the signal becomes

$$V_n^{\text{ind}}(t) = \frac{1}{Q_w} \int_{-\infty}^{\infty} dt' \delta(t - t') \int_{\mathbb{R}^3} dV \; \phi_n(\mathbf{x}; Q_w) \rho^e(\mathbf{x}, t') - \lim_{t' \to -\infty} \frac{1}{Q_w} \int_{\mathbb{R}^3} dV \; \phi_n(\mathbf{x}; Q_w) \rho^e(\mathbf{x}, t').$$

According to the regularisation procedure outlined in Section 1.2, the charge must be brought in from spatial infinity at $t \ll -T$. Far away from the detector assembly, only the monopole term in the weighting potential survives, and $\phi_{w,n}(\mathbf{x}) \sim 1/|\mathbf{x}|$. The boundary term in the above equation thus vanishes and the remaining contribution evaluates to

$$V_n^{\text{ind}}(t) = \frac{q}{Q_w} \phi_n(\mathbf{x}_q(t); Q_w). \tag{1.12}$$

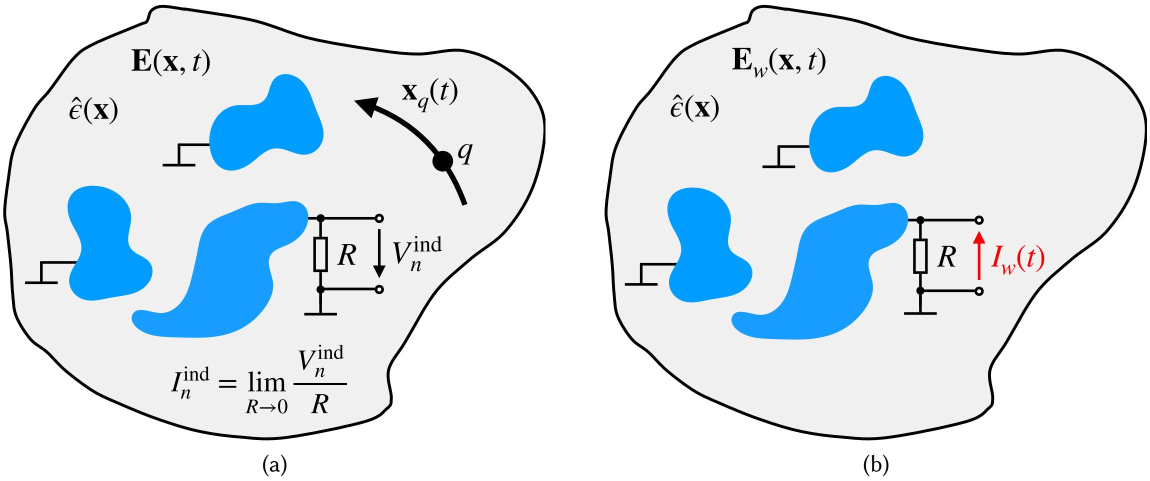


Figure 1.8: (a) Detector setup consisting of N perfectly conducting electrodes connected to ground and embedded in a material with permittivity $\hat{\epsilon}(\mathbf{x})$. A moving point charge induces a current I_n^{ind} flowing from electrode n to ground. (b) The weighting field $\mathbf{E}_w(\mathbf{x}, t)$ for the n-th electrode is produced by the current $I_w(t) = Q_w\delta(t)$.

1.3.2 Current induced on grounded electrodes

Another common case is that of a detector whose electrodes are grounded. The detector signal takes the form of a current I_n^{ind} defined as flowing from electrode n to ground potential. For simplicity, the permittivity matrix continues to be taken to be symmetric. To measure I_n^{ind}, a small ohmic resistance R is introduced between electrode n and ground, as shown in Figure 1.8a. The induced current I_n^{ind} in the unperturbed circuit is given by

$$I_n^{\text{ind}}(t) = \lim_{R\to 0} \frac{V^{\text{ind}}(t)}{R}, \tag{1.13}$$

where V_n^{ind} is the induced voltage across the terminals of the resistor.

The weighting field $\mathbf{E}_w(\mathbf{x}, t)$ for this situation is computed as shown in Figure 1.8b. As in Section 1.3.1, the effect of the current $I_w(t) = Q_w\delta(t)$ is to apply a charge Q_w to electrode n at $t = 0$. For $t > 0$, a current $I_R(t)$ flows through the resistor R towards ground and discharges the electrode with a time constant that depends on R and the mutual capacitances between the electrodes. The weighting field $\mathbf{E}_w$ is thus no longer static. Denoting the potential on the surface of electrode n as $V_n(t) = R \cdot I_R(t)$, it may be parameterised as

$$\mathbf{E}_w(\mathbf{x}, t) = \frac{V_n(t)}{V_w}\mathbf{E}_{w,n}(\mathbf{x}; V_w).$$

The field $\mathbf{E}_{w,n}(\mathbf{x}; V_w)$ corresponds to the situation where electrode n is kept at a constant potential of V_w and all other electrodes remain grounded. The above relation exploits the fact that the quasi-static electric field created by the charged electrode n is homogeneous in its potential V_n.

Using the signal theorem in Eq. 1.9b to compute I_n^{ind} in Eq. 1.13 gives

$$I_n^{\text{ind}}(t) = \lim_{R\to 0} -\frac{q}{Q_w R}\int_{-\infty}^{\infty} dt' \, \frac{V_n(t-t')}{V_w}\, \mathbf{E}_n(\mathbf{x}_q(t'); V_w)\cdot \dot{\mathbf{x}}_q(t'). \tag{1.14}$$

In the limit $R \to 0$, the discharge of electrode n becomes arbitrarily fast, i.e.

$$\lim_{R\to 0} V_n(t) = R\, Q_w\, \delta(t).$$

The normalisation is determined by the requirement that the entire charge Q_w must flow through the resistor R, i.e. $\int_0^\infty dt\, I_R(t) = Q_w$. The induced current then becomes

$$I_n^{\text{ind}}(t) = -\frac{q}{V_w}\mathbf{E}_n(\mathbf{x}_q(t); V_w)\cdot \dot{\mathbf{x}}_q(t). \tag{1.15}$$

This formula is known in the literature as the "Ramo-Shockley theorem" [36, 37]. It is conventionally derived from Green's reciprocity relation [38, 39] which holds only for quasi-electrostatic situations but emerges here as a special case of a much more general theorem.

1.3.3 Signals induced on electrodes embedded in a circuit

Detector electrodes are typically neither grounded nor perfectly insulated but rather connected to readout electronics with a nontrivial input impedance. Figure 1.9a shows the general situation for the case of a detector with two readout electrodes. A general impedance element $Z_{12}(\omega)$ connects the electrodes with each other. The electrodes are furthermore connected to ground potential through the impedance elements $Z_{10}(\omega)$ and $Z_{20}(\omega)$.

The detector signal is the current I_{20}^{ind} flowing through Z_{20} (which could, for instance, represent the input impedance of the front-end amplifier). As before, a small ohmic resistance R is connected in series with Z_{20} to measure the induced current.

The weighting field for this situation is computed by applying a current $I_w(t) = Q_w\delta(t)$ across the terminals of R, as shown in Figure 1.9b. Following the same arguments as in Section 1.3.2, the weighting field takes the form

$$\mathbf{E}_w(\mathbf{x}, t) = \frac{V_2(t)}{V_w}\mathbf{E}_{w,2}(\mathbf{x}; V_w). \tag{1.16}$$

The field $\mathbf{E}_{w,2}(\mathbf{x}; V_w)$ corresponds to the situation where electrode 2 is kept at a constant potential

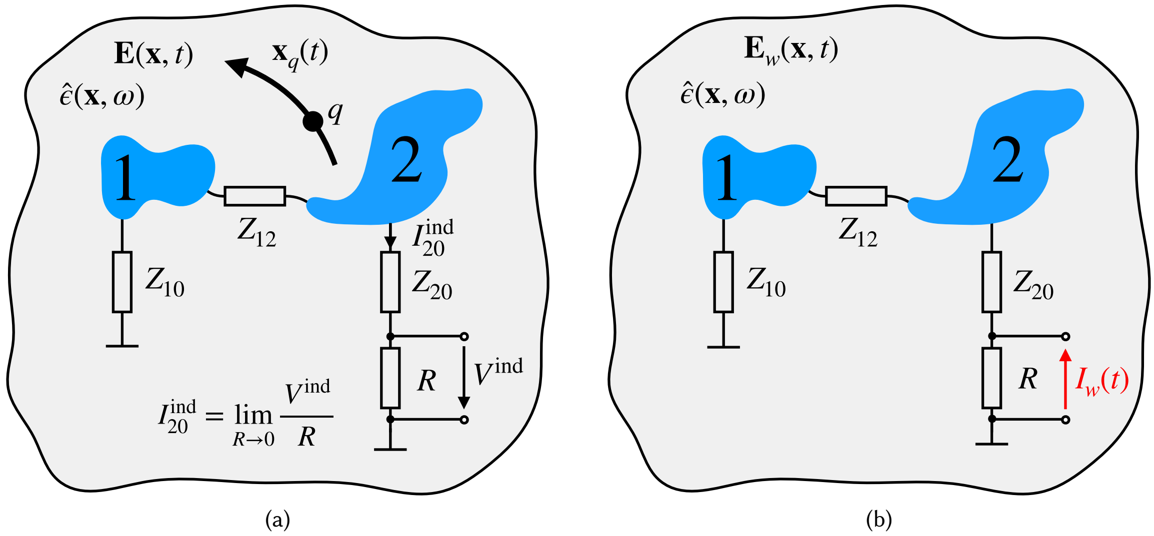


Figure 1.9: (a) Detector setup consisting of two conducting electrodes that are interconnected with general impedance elements. A moving point charge induces a current signal I_{20}^{ind} through Z_{20}. (b) Situation to compute the weighting field $\mathbf{E}_w(\mathbf{x}, t)$ for the signal.

V_w. It is scaled by the time-dependent potential $V_2(t)$ of this electrode that arises as a consequence of the current $I_w(t)$. With Eq. 1.9, and in analogy to Eq. 1.14, the induced current I_{20}^{ind} is

$$I_{20}^{\text{ind}}(t) = \lim_{R \to 0} -\frac{q}{Q_w V_w} \int_{-\infty}^{\infty} dt' \, \frac{V_2(t-t')}{R} \, \mathbf{E}_{w,2}(\mathbf{x}_q(t'); V_w) \cdot \dot{\mathbf{x}}_q(t'). \tag{1.17}$$

Network reciprocity may be used to bring $\lim_{R \to 0} V_2/R$ into a more physically illuminating form.

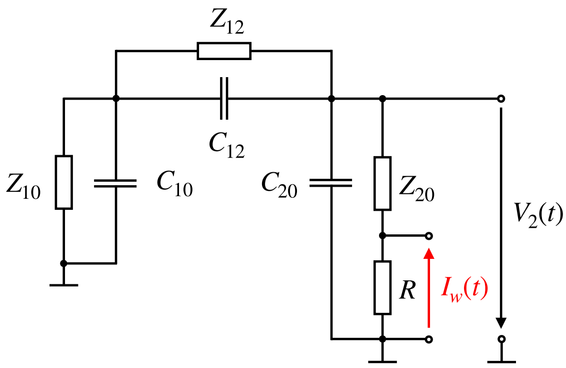
The voltage $V_2(t)$ also appears in the equivalent circuit of the considered situation, shown in Figure 1.10a. A different configuration of this circuit is shown in Figure 1.10b, where a delta-like current $I_2(t) = Q_w \delta(t)$ is instead applied to electrode 2 directly, generating a voltage $V_R(t)$. The network reciprocity relation from Section 1.1.2 allows these two situations to be compared, producing the relation $V_2(t) = V_R(t)$. In the limit of small R, it thus follows that

$$\lim_{R \to 0} \frac{V_2(t)}{R} = \lim_{R \to 0} \frac{V_R(t)}{R} = I_{20}\left(t; I_2(t) = Q_w \delta(t)\right), \tag{1.18}$$

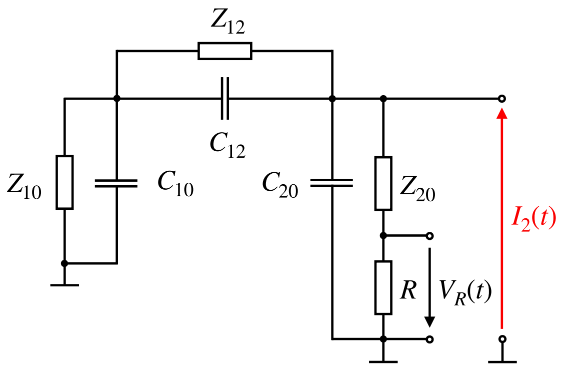
where $I_{20}\left(t; I_2(t) = Q_w \delta(t)\right)$ is the current flowing through Z_{20} in reaction to the delta-like stimulus $I_2(t)$ applied to a circuit where $R = 0\,\Omega$.

With the relation in Eq. 1.18, the induced current in Eq. 1.17 becomes

$$\begin{aligned} I_{20}^{\text{ind}}(t) &= \int_{-\infty}^{\infty} dt' \, \frac{1}{Q_w} I_{20}\left(t-t'; I_2(t) = Q_w \delta(t-t')\right) \cdot \left(-\frac{q}{V_w} \mathbf{E}_2(\mathbf{x}_q(t'); V_w) \cdot \dot{\mathbf{x}}_q(t')\right) \\ &= \int_{-\infty}^{\infty} dt' \, \frac{1}{Q_w} I_{20}\left(t-t'; I_2(t) = Q_w \delta(t-t')\right) \cdot I_2^{\text{ind}}(t'). \end{aligned} \tag{1.19}$$



(a)



(b)

Figure 1.10: Equivalent circuit of the situation in Figure 1.9, including the discrete impedances Z_{ij} as well as the mutual capacitances C_{ij}. (a) The current $I_w(t) = Q_w\delta(t)$ produces the voltage $V_2(t)$. (b) Applying a current $I_2(t) = Q_w\delta(t)$ directly on electrode 2 leads to a voltage $V_R(t)$ across R.

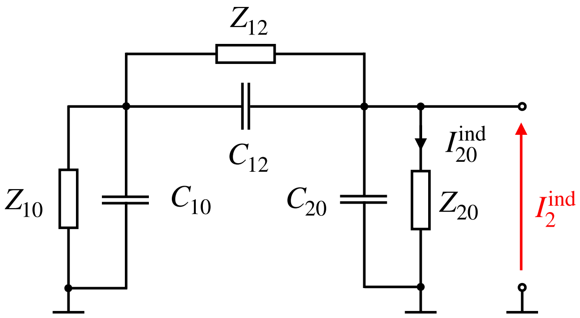


Figure 1.11: The current $I_2^{\text{ind}}(t)$ induced on the grounded detector electrode is inserted into the equivalent circuit of the detector to compute the physical signal $I_{20}^{\text{ind}}(t)$.

The expression in parentheses is the current $I_2^{\text{ind}}(t)$ from Eq. 1.15 that would be induced on electrode 2 if all electrodes were connected to ground. The above formula interprets $I_2^{\text{ind}}(t)$ as originating from a current source connected as shown in Figure 1.11. It then convolves this current with the Green's function of the circuit to find $I_{20}^{\text{ind}}(t)$.

This makes it possible to factorise the calculation: one first computes the current induced on grounded electrodes with the Ramo-Shockley theorem in Eq. 1.15, and then includes this current as a source in the simulation of the equivalent circuit of the detector [40, 41]. In practice, the latter step is best performed with simulation tools such as SPICE [42].

1.3.4 Detectors with resistive media

The signal induced in a detector setup involving materials with finite conductivity $\hat{\boldsymbol{\sigma}}$ may be computed directly with Eq. 1.9b. To compute the (time-dependent) weighting field, a delta-like line current is applied as detailed in Section 1.2. If radiation effects can be neglected, the weighting field can be computed in the electrostatic approximation. The procedure then immediately reduces to the one developed in Refs. [41, 43–45].

1.4 Applications and examples

The results discussed in Section 1.3 cover many traditional particle detectors, including most devices based on ionisation in gas or silicon. This section explores some of the consequences of the signal theorem in Eq. 1.9 that go beyond the quasi-electrostatic approximation.

1.4.1 Signals induced in long drift tubes

Drift tubes are cylindrical gas-based detectors. As illustrated in Figure 1.12a, the gas volume is delimited by a metallic inner conductor with radius a and a grounded cathode with radius b. The inner electrode is typically implemented as a thin wire to which a potential $V > 0$ is applied. In the following, the drift tube is taken to be oriented along the z-direction.

Gas molecules are ionised by a primary particle traversing the chamber at longitudinal coordinate $z = z_0$. The ionisation electrons drift towards the inner conductor in the purely radial electric field $E^r \sim 1/r$. Close to the wire, the diverging field strength enables impact ionisation which leads to a proportional amplification of the primary electrons. The ions produced in this process drift radially outwards and are responsible for the dominant part of the induced signal [40], as shown in Figure 1.12b. The signal propagates along the transmission line formed by the inner conductor and the outer mantle. It is registered by an amplifier located at the end of the chamber at $z = 0$, often several metres away from the position of the charge avalanche at $z = z_0$. The opposite end of the drift tube is terminated with an impedance Z_T. To prevent reflections, Z_T, and the input impedance of the amplifier, Z_{in}, are matched to the characteristic impedance of the transmission line, Z_0, i.e. $Z_{\text{in}} = Z_T = Z_0$.

A rigorous and coherent description of the signal formation process must include the propagation of electromagnetic waves in a transmission line, a phenomenon that cannot be described in the quasi-electrostatic approximation.

As usual, the weighting field $\mathbf{E}_w(\mathbf{x}, t)$ for this situation is the result of a line current $I_w(t) = Q_w \delta(t)$ applied at the terminals of the amplifier, as shown in Figure 1.12c. It has the character of a narrow pulse propagating along the transmission line, the details of which depend on the field configurations that are allowed to propagate. If the material distributions of the detector and the

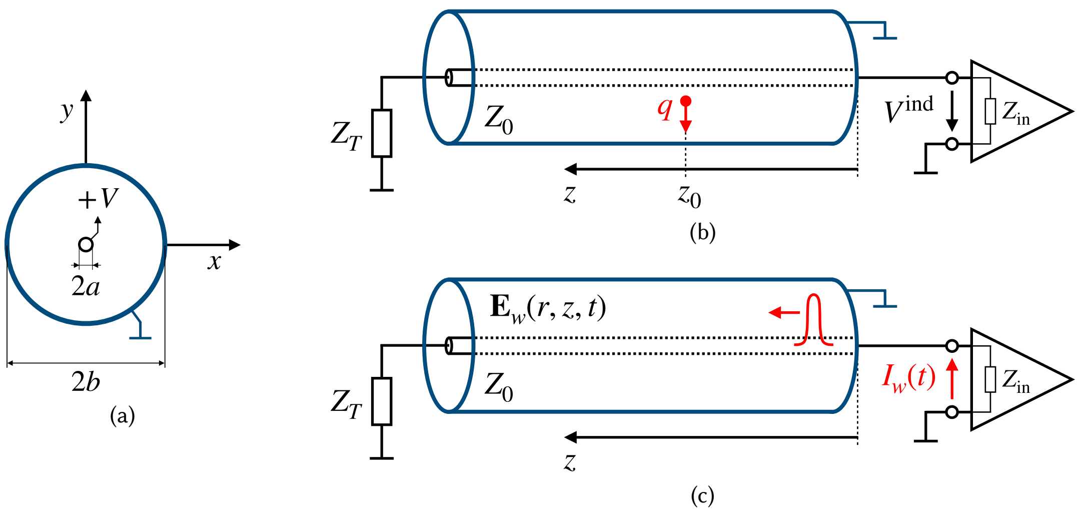


Figure 1.12: Signal formation in a long drift tube. (a) The drift tube consists of a central conductor of radius a, kept at a positive voltage V. It is surrounded by a grounded cathode with radius b. (b) A positive charge q drifts radially outwards at $z = z_0$, producing a signal V^{ind} at the terminals of the front-end amplifier with input impedance Z_{in}. The other side of the drift tube is terminated with an impedance Z_T. (c) To compute the weighting field, a current $I_w(t) = Q_w \delta(t)$ is applied as shown.

environment are sufficiently homogeneous along the direction of propagation, transverse electromagnetic (TEM) modes are most important.

For TEM modes, all fields are purely transverse, i.e. the vector $\mathbf{E}_w$ lies in the xy-plane. Maxwell's equations in the transverse plane reduce to the laws of electrostatics and magnetostatics [46], and propagation of waves is only possible along the longitudinal z-direction. This implies that the weighting field has the structure

$$\mathbf{E}_w(r, z, t) = V(z, t) \frac{-\boldsymbol{\nabla} \phi_w(r; V_w)}{V_w}, \tag{1.20}$$

where $\phi_w(r; V_w)$ is the electrostatic potential in the transverse plane and $r = \sqrt{x^2 + y^2}$. It corresponds to the situation where a potential V_w is applied to the anode wire while the outer cathode is kept at ground. The voltage $V(z, t) = \int_{r=a}^{r=b} d\mathbf{s} \cdot \mathbf{E}_w(r, z, t)$ is the instantaneous potential difference between the inner conductor and the outer mantle. In a TEM field configuration, it is unambiguous and not dependent on the path of integration. Its full functional form is determined by the telegrapher's equations [46], but this will not be relevant for what follows.

Using Eqs. 1.20 and 1.9, the signal induced by the drifting charge is

$$V^{\text{ind}}(t) = \frac{1}{Q_w} \int dt' \, V(z_0, t - t') I^{\text{ind}}(z_0, t'), \tag{1.21}$$

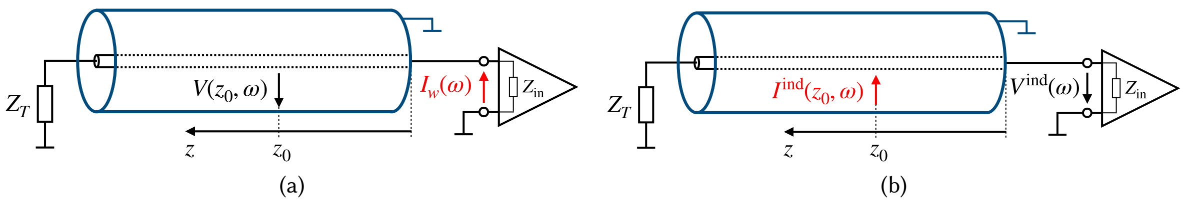


Figure 1.13: The situations in (a) and (b) relate the quantities $V^{\text{ind}}(\omega)$, $I^{\text{ind}}(z_0, \omega)$, $V(z_0, \omega)$ and $I_w(\omega)$ by network reciprocity.

where the current $I^{\text{ind}}(z_0, t)$ is given by

$$I^{\text{ind}}(z_0, t) = \frac{q}{V_w} \boldsymbol{\nabla} \phi_w(r_q(t); V_w) \cdot \dot{\mathbf{x}}_q(t).$$

This result is reminiscent of Eq. 1.19. The Ramo-Shockley theorem in Eq. 1.15 identifies $I^{\text{ind}}(z_0, t)$ as the current induced on the inner conductor at the position of the primary ionisation event, computed in the electrostatic approximation. Eq. 1.21 propagates this current through the transmission line to obtain the signal at the position of the amplifier. In the frequency domain, it reads $V^{\text{ind}}(\omega) = V(z_0, \omega) \cdot I^{\text{ind}}(z_0, \omega)/I_w(\omega)$, which Figure 1.13 reveals as an application of the network reciprocity relation in Eq. 1.4.

This calculation demonstrates a physically very intuitive fact: the induced signal can be obtained by first computing the induced current at the location of the drifting charge, and then transmitting this signal along the axis of the drift tube. The physics of signal induction factorises from the propagation of this signal through the circuit.

1.4.2 Dipole antennas as particle detectors

Antennas are detectors which directly measure electric and magnetic fields emitted by moving charged particles, often over very long distances. Figure 1.14 shows infinitesimal electric and magnetic dipole antennas as two very simple realisations. (The following discussion also applies to antennas of finite extent, provided that their size remains small compared to the wavelength of the incoming radiation field.)

The two terminals $\mathbf{x}_1$ and $\mathbf{x}_0$ of the electric dipole in Figure 1.14a are separated by an infinitesimal distance $d\mathbf{s}$ along the z-axis. The antenna is positioned at the origin of the coordinate system. The detector signal is formed by the voltage $V^{\text{ind}}(t)$ measured between these two points along the path $\mathcal{S}_|$, as shown in the figure. The magnetic dipole in Figure 1.14b consists of an infinitesimal loop $\mathcal{S}_\circ$ in the xy-plane situated at the origin and enclosing an area dA. The space surrounding the antennas

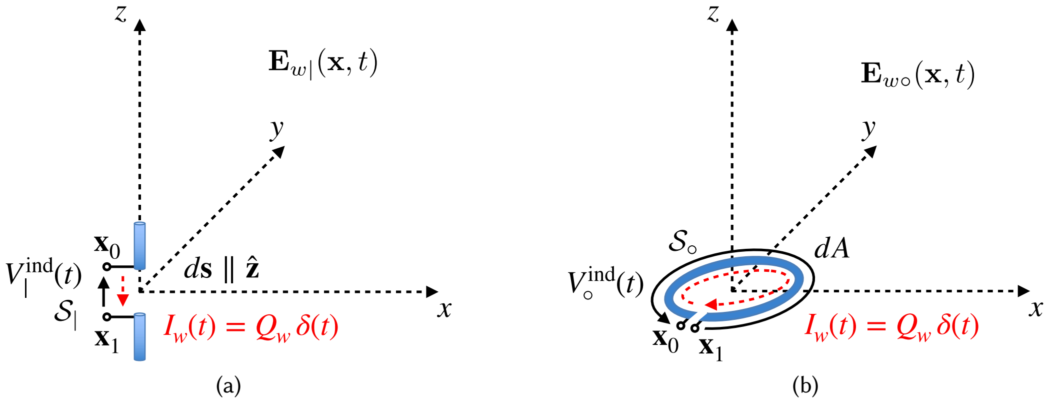


Figure 1.14: Infinitesimal antennas as particle detectors. (a) An infinitesimal electric dipole antenna oriented along the z-axis and its weighting field $\mathbf{E}_{w|}(\mathbf{x}, t)$. (b) An infinitesimal magnetic dipole antenna in the xy-plane and its weighting field $\mathbf{E}_{w\circ}(\mathbf{x}, t)$.

is taken to be filled with a homogeneous, isotropic material with permittivity ϵ and permeability μ.

The weighting field $\mathbf{E}_{w|}(\mathbf{x}, t)$ for the electric dipole is computed as the response to a line current $I_w(t) = Q_w\delta(t)$ flowing along the path $\mathcal{S}_{|}$, as illustrated in Figure 1.14a. Maxwell's equations are most conveniently solved in spherical coordinates with radial coordinate r, polar angle θ (measured from the z-axis) and azimuthal angle φ (measured from the x-axis). Only the radial and polar components $E^r_{w|}$ and $E^\theta_{w|}$ are nonvanishing. They read

$$E^r_{w|}(r,\theta) = -2\frac{Q_w ds}{4\pi\epsilon}\frac{\cos\theta}{r^3}\left[\Theta\left(t-\frac{r\,n}{c}\right)+\frac{r\,n}{c}\delta\left(t-\frac{r\,n}{c}\right)\right], \tag{1.22a}$$

$$E^\theta_{w|}(r,\theta) = -\frac{Q_w ds}{4\pi\epsilon}\frac{\sin\theta}{r^3}\left[\Theta\left(t-\frac{r\,n}{c}\right)+\frac{r\,n}{c}\delta\left(t-\frac{r\,n}{c}\right)+\left(\frac{r\,n}{c}\right)^2\delta'\left(t-\frac{r\,n}{c}\right)\right], \tag{1.22b}$$

$$E^\varphi_{w|}(r,\theta) = 0, \tag{1.22c}$$

where $\delta'(x)$ is the derivative of the Dirac delta distribution. The weighting field contains a shock front propagating outwards with a velocity $v = c/n$ where $n = c\sqrt{\mu\epsilon}$ is the refractive index of the material and c is the speed of light in vacuum. For $t > r\,n/c$ the weighting field corresponds to the field of a static electric dipole with dipole moment $\mathbf{p} = -Q_w\,ds\,\hat{\mathbf{z}}$ situated at the origin, where $\hat{\mathbf{z}}$ is the unit vector in z-direction.

A similar calculation returns the weighting field $\mathbf{E}_{w\circ}$ of the magnetic dipole antenna. The current $I_w(t) = Q_w\delta(t)$ is now set to flow along the path $\mathcal{S}_\circ$ in the direction shown in Figure 1.14b. The result

$$E^\theta_{w\circ}(r,\theta) = 0, \tag{1.23a}$$

$$E^r_{w\circ}(r,\theta) = 0, \tag{1.23b}$$

$$E^{\varphi}_{w\circ}(r,\theta) = \frac{Q_w\, dA\, \mu}{4\pi} \frac{\sin\theta}{r^2} \left[\delta'\left(t - \frac{r\, n}{c}\right) + \frac{r\, n}{c}\delta''\left(t - \frac{r\, n}{c}\right)\right] \tag{1.23c}$$

displays a similar shock front extending outwards from the position of the antenna.

The weighting fields in Eqs. 1.22 and 1.23 contain "Coulomb terms" with a $1/r^2$-dependence and "radiation terms" which decay like $1/r$. They may be convolved with a general trajectory $\mathbf{x}_q(t)$ of a moving charged particle according to Eq. 1.9. The calculation is slightly lengthy and reproduced in full in Ref. [1]. It shows that the signal $V_{|}^{\text{ind}}$ measured by the infinitesimal electric dipole corresponds to the electric field $\mathbf{E}$ produced by the moving particle evaluated at the position of the antenna and projected onto its orientation,

$$V_{|}^{\text{ind}}(t) = \mathbf{E}(\mathbf{x} = \mathbf{0}, t) \cdot \hat{\mathbf{z}}. \tag{1.24}$$

Similarly, the infinitesimal magnetic dipole measures the rate of change of the magnetic flux $dA\, \mathbf{B} \cdot \hat{\mathbf{z}}$ produced by the moving particle,

$$V_{\circ}^{\text{ind}}(t) = -dA\, \frac{\partial \mathbf{B}(\mathbf{x} = \mathbf{0}, t)}{\partial t} \cdot \hat{\mathbf{z}}, \tag{1.25}$$

consistent with the Maxwell-Faraday equation $\boldsymbol{\nabla} \times \mathbf{E}(\mathbf{x}, t) = -\partial_t \mathbf{B}(\mathbf{x}, t)$.

1.4.3 Radio emissions of showers induced by cosmic rays

An important application of the results obtained in Section 1.4.2, and in particular Eqs. 1.24 and 1.25, is the experimental study of the radio signature of showers induced by ultra-high energy cosmic rays.

Charged cosmic rays that impinge on the atmosphere produce particle showers which develop over length scales of several kilometres [47]. Similar showers may be triggered in dense media, such as water ice, by astrophysical neutrinos with energies of the order of petaelectronvolts. The number of particles contained in the shower grows approximately exponentially during the early stages of its formation. These showers are not globally charge neutral and so represent an electric current that leads to the production of electromagnetic radiation which provides an excellent experimental signal [48]. In particular, the characteristics of the radiation pattern from the excess of negative charge produced at the shower front were first described in detail by Askaryan [15].

Large-scale arrays of radio antennas are either planned [16] or already in operation [49] to study the radio signature of showers developing in air or in water ice. Both media are remarkably transparent to electromagnetic radiation with frequencies of a few GHz. A radio array positioned on the

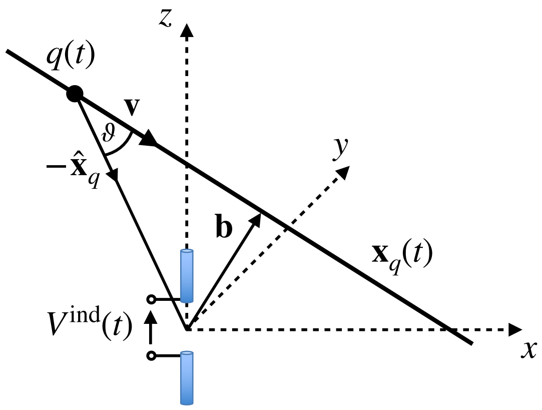


Figure 1.15: Simple model for an electromagnetic shower. The shower front is represented by a point charge $q(t)$ moving at a constant velocity $\mathbf{v}$. The radiation produced by the shower is received by an infinitesimal electric dipole antenna positioned at the origin.

surface can thus effectively generate the vast instrumented volumes that are required to achieve good sensitivity to very rare events.

To expose the salient physical properties of these radio emissions, it is not necessary to include all phenomenological details of the shower evolution, nor is it important to have a precise description of the environment in which the shower develops. At distances that are much larger than the lateral extent of the shower, a one-dimensional model is sufficient. In this model, the shower front and its growing negative charge excess are represented by a time-dependent point charge $q(t) < 0$. This charge moves along a straight trajectory $\mathbf{x}_q(t) = (x_q(t), y_q(t), z_q(t))^{\mathsf{T}} = \mathbf{b} + \mathbf{v}t$ with $\mathbf{v} = \boldsymbol{\beta}c$ through a medium with refractive index n. The fields and radiation produced by the resulting current $\mathbf{J}^e(\mathbf{x}, t) = q(t)\,\mathbf{v}\,\delta(\mathbf{x} - \mathbf{x}_q(t))$ are observed by an infinitesimal dipole antenna situated at the origin and oriented along the z-direction. The situation is illustrated in Figure 1.15.

The signal $V_|^{\text{ind}}$ produced by the antenna is computed using the signal theorem in Eq. 1.9 and the weighting field from Eq. 1.22. It reads

$$\begin{aligned} V_|^{\text{ind}}(t) = &-\frac{ds}{4\pi\epsilon}\int_{-\infty}^{\infty} dt' q(t')\,\Theta\left(t - t' - \frac{r_q\,n}{c}\right)\frac{d}{dt'}\frac{z_q}{r_q^3} \\ &-\frac{ds}{4\pi\epsilon}\int_{-\infty}^{\infty} dt' q(t')\,\delta\left(t - t' - \frac{r_q\,n}{c}\right)\frac{r_q\,n}{c}\frac{d}{dt'}\frac{z_q}{r_q^3} \\ &+\frac{ds}{4\pi\epsilon}\int_{-\infty}^{\infty} dt' q(t')\,\delta'\left(t - t' - \frac{r_q\,n}{c}\right)\frac{n^2}{c^2 r_q^3}\left(z_q\,\mathbf{x}_q\cdot\dot{\mathbf{x}}_q - r_q^2\dot{z}_q\right), \end{aligned}$$

where $r_q = \|\mathbf{x}_q\|$ and the dependence of $\mathbf{x}_q$ on t' is suppressed for clarity. The remaining manipulations simplify if the shower front moves at subluminal velocities $n\beta < 1$. Evaluating the above integrals for this case shows that the signal contains three contributions with different physical origins, $V_{|}^{\text{ind}} = V_{\text{Front}}^{\text{ind}} + V_{\text{Ions}}^{\text{ind}} + V_{\text{Askaryan}}^{\text{ind}}$. The first contribution, $V_{\text{Front}}^{\text{ind}}$, is the (Lorentz-boosted) Coulomb field of the moving charge $q(t)$ representing the shower front. The moving shower front leaves behind a trail of positive charges at rest, whose Coulomb field is given by the second term, $V_{\text{Ions}}^{\text{ind}}$. The third signal component, $V_{\text{Askaryan}}^{\text{ind}}$, is the Askaryan radiation term associated with the shower front [15]. It depends explicitly on $\dot{q}(t)$. Explicitly, these signal components read

$$V_{\text{Front}}^{\text{ind}}(t) = -\frac{ds}{4\pi\epsilon}\left[\frac{q(t)}{|1-n\beta\cos\vartheta|^3}\frac{1}{r_q^2}\left(1-n^2\beta^2\right)\left(\frac{z_q}{r_q}+n\beta^z\right)\right]_{t_{\text{ret}}}, \tag{1.26a}$$

$$V_{\text{Ions}}^{\text{ind}}(t) = \frac{ds}{4\pi\epsilon}\int_{-\infty}^{t_{\text{ret}}} dt'\dot{q}(t')\frac{z_q}{r_q^3}, \tag{1.26b}$$

$$V_{\text{Askaryan}}^{\text{ind}}(t) = -\frac{ds}{4\pi\epsilon}\left[\frac{\dot{q}(t)}{|1-n\beta\cos\vartheta|(1-n\beta\cos\vartheta)}\frac{n}{r_q c}\left(n\beta^z+n\beta\cos\vartheta\frac{z_q}{r_q}\right)\right]_{t_{\text{ret}}}, \tag{1.26c}$$

with $\hat{\mathbf{x}}_q = \mathbf{x}_q/r_q$ and ϑ is the angle between $\mathbf{v}$ and $-\hat{\mathbf{x}}_q$. The retarded time t_{ret} is related to the coordinate time t as $t_{\text{ret}} + r_q(t_{\text{ret}})\,n/c = t$ [50]. For the shower trajectory studied here, it is

$$t_{\text{ret}} = \frac{1}{1-n^2\beta^2}\left[t-\frac{n}{c}\sqrt{b^2+\beta^2(c^2t^2-n^2b^2)}\right].$$

For some applications it is convenient to express Eq. 1.26 as an explicit function of the coordinate time. To eliminate the retarded time t_{ret}, the following relations between retarded and non-retarded quantities are useful [50],

$$\begin{aligned}
\left[r_q^2(1-n\beta\cos\vartheta)^2\right]_{t_{\text{ret}}} &= b^2(1-n^2\beta^2)+\beta^2c^2t^2, \\
[\cos\vartheta]_{t_{\text{ret}}} &= n\beta\left(1-\frac{c\,t}{[n\,r_q]_{t_{\text{ret}}}}\right), \\
\left[\frac{z_q}{r_q}\right]_{t_{\text{ret}}} &= \frac{b^z+\beta^z ct}{[r_q]_{t_{\text{ret}}}}-n\beta^z, \\
\left[r_q^2\right]_{t_{\text{ret}}} &= b^2+\frac{\beta^2c^2}{(1-n^2\beta^2)^2}\left[t-\frac{n}{c}\sqrt{b^2+\beta^2(c^2t^2-n^2b^2)}\right]^2.
\end{aligned}$$

With the result in Eq. 1.24, the three terms in Eq. 1.26 may be completed into an expression for the electric field at the position of the dipole antenna,

$$\begin{aligned}
\mathbf{E}(t) = &-\frac{ds}{4\pi\epsilon}\left[\frac{q(t)}{|1-n\beta\cos\vartheta|^3}\frac{1}{r_q^2}\left(1-n^2\beta^2\right)\left(\hat{\mathbf{x}}_q+n\boldsymbol{\beta}\right)\right]_{t_{\text{ret}}} \\
&+\frac{ds}{4\pi\epsilon}\int_{-\infty}^{t_{\text{ret}}} dt'\dot{q}(t')\frac{\mathbf{x}_q}{r_q^3}- \\
&-\frac{ds}{4\pi\epsilon}\left[\frac{\dot{q}(t)}{|1-n\beta\cos\vartheta|(1-n\beta\cos\vartheta)}\frac{n}{r_q c}\left(n\boldsymbol{\beta}-\left(n\boldsymbol{\beta}\cdot\hat{\mathbf{x}}_q\right)\hat{\mathbf{x}}_q\right)\right]_{t_{\text{ret}}},
\end{aligned}$$

which reveals that the electric field due to the Askaryan effect is oriented along the velocity com-

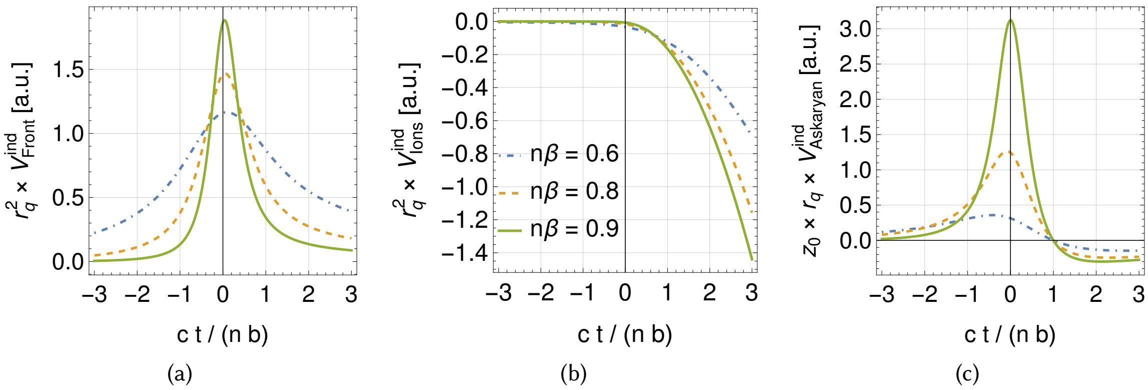


Figure 1.16: Signal components induced in the dipole antenna by the developing shower. All components are scaled by suitable powers of $r(t)$ and z_0 to show their intrinsic strength, and $z_0/b = 10$ is chosen. (a) The signal $V^{\text{ind}}_{\text{Front}}$ produced by the Coulomb field of the shower front. (b) The signal $V^{\text{ind}}_{\text{Ions}}$ produced by the Coulomb field of the ion tail. (c) The signal $V^{\text{ind}}_{\text{Askaryan}}$ attributed to the Askaryan effect.

ponent $\boldsymbol{\beta}_\perp = \boldsymbol{\beta} - (\boldsymbol{\beta} \cdot \hat{\mathbf{x}}_q)\hat{\mathbf{x}}_q$ perpendicular to the line of sight [51].

The three signal components in Eq. 1.26 are illustrated in Figure 1.16 for a shower developing in the xz-plane parallel to the x-axis. The net charge of the shower front is taken to grow exponentially with time, $q(t) = q \exp\left(\frac{\beta c t}{z_0}\right)$, where the quantity z_0 parameterises the longitudinal length scale of the shower development. This simple model is, of course, not accurate in the late stages of the shower but provides an excellent approximation in cases where the shower maximum is reached only well beyond its point of closest approach to the antenna. The contributions $V^{\text{ind}}_{\text{Front}}$ and $V^{\text{ind}}_{\text{Ions}}$ both lead to unipolar detector signals. The polarity of the Askaryan signal $V^{\text{ind}}_{\text{Askaryan}}$ changes as the charge moves past the observer, arising from a sign flip in $\boldsymbol{\beta}_\perp$. It also becomes relatively more important at high shower energies.

Signal simulations for radio neutrino observatories Accurate simulations of the radio signature of the shower are important for the present and next generation of cosmic-ray radio detection experiments. This is particularly pertinent for in-ice neutrino observatories [16] such as the one in Figure 1.1b. A number of instrumental and physics backgrounds must be understood to be able to reconstruct the neutrino energy, e.g. contaminations from the thermal noise emitted by the ice or air-shower cores impacting the ice from above [52].

However, the direct, exact computation of the radio emissions through the numerical solution of Maxwell's equations presents significant challenges for realistic scenarios and detector geometries. The instrumented volume of $\mathcal{O}(\text{km}^3)$ must be discretised with a resolution of $\mathcal{O}(\text{cm})$ to capture propagation and diffraction of waves in the GHz domain. The incoming shower can have an arbitrary orientation with respect to the array, leaving no useful symmetries to be exploited. The presently available simulation infrastructure therefore typically uses simplified parameterised models or geometric optics approximations [53].

The signal theorem in Eq. 1.9 has the potential to avoid these shortcomings and significantly extend the scope of current simulation tools. Even for practically relevant detector geometries and ice models, the weighting field continues to have an approximate cylindrical symmetry. This greatly simplifies its numerical calculation and realistic weighting fields are readily obtained with available numerical methods [54] and then discretised and stored. The electric field at the position of the antenna, or the filtered detector signal V^{out}, is then determined through a numerical convolution of the stored weighting field with the particle trajectories from a Monte Carlo evolution of the shower.

1.5 Conclusions

For charged-particle detectors consisting of linear, inhomogeneous, anisotropic, non-symmetric materials, the induced signal may be computed with the signal theorem in Eq. 1.9, rigorously explaining the characteristics of this signal as consequences of classical electrodynamics. The theorem applies to all devices that detect fields or radiation from moving charges. It explicitly shows how the signal arises as the convolution of the trajectory of the particle with the weighting field, which encodes the geometry of the detector and its environment.

This result allows electro*dynamic* situations to be described that are not easily accessible with previous methods, and reduces to these prescriptions in various special cases. It has important applications in situations where the signal formation involves a combination of electrostatic induction and wave propagation, such as in long drift tubes; or where the detector signal is predominantly or entirely caused by electromagnetic radiation, as it is the case for radio neutrino observatories.

Chapter 2

The statistics of electron-hole avalanches in semiconductors

Charge avalanche amplification is an important experimental technique indispensable in many applications where small amounts of initial charge need to be detected or amplified. An avalanche can form only when conditions are such that the initial charge carriers may participate in processes that produce additional charge carriers of the same type. Provided these multiplication reactions can be sustained for a sufficiently long time, exponentially large gains are possible, allowing macroscopically large signals to be induced by even a single initial charge carrier.

Charge avalanches in gases were studied in detail by Townsend [55], who identified electrons as the charge carrier species propelling the growth of the avalanche and electron impact ionisation as the physical mechanism behind charge multiplication. In recent decades, advances in the manufacturing of semiconductors have made it possible to construct devices that exhibit avalanche multiplication in the solid state. Charge carriers of both polarities, electrons and holes, generally participate in these avalanches.

The understanding of the physics behind Townsend discharges has sparked the development of a new class of sensitive instruments, with the Geiger counter as one of the most impactful examples. Similarly, exploiting avalanche multiplication in semiconductors has allowed particle detectors with superb timing capabilities to be constructed [56].

Therefore, it is pertinent to develop a rigorous understanding of the phenomenology of the electron-hole avalanche and the ultimate performance limits of any detection device employing avalanche amplification. Such a treatment need not necessarily deal with all the details and in-

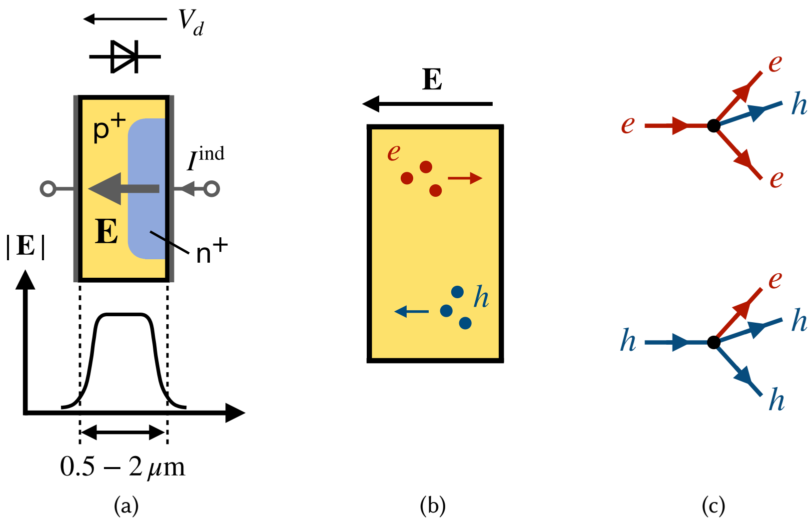


Figure 2.1: Formation of electron-hole avalanches in silicon devices. (a) A highly-doped p–n junction to which a reverse-bias voltage V_d is applied creates very strong electric fields in a region with a thickness of a few microns. The junction is bounded by two conducting electrodes. (b) Electrons (e) and holes (h) drift in opposite directions in the, generally position dependent, electric field. (c) The charge carriers receive energy from the field and undergo impact ionisation, leading to the formation of additional electron-hole pairs.

tricacies of practical devices. Rather, deliberately reducing the complexity of the situation, while retaining the essential character of the avalanche, helps to clearly expose its central features and often allows them to be condensed into simple, analytic formulae.

Figure 2.1a shows the situation that forms the starting point of this discussion. Upon reverse-biasing, a strongly doped p–n junction exhibits strong electric fields of the order of 20–60 V/µm in a thin layer with a thickness of about 0.5–2 µm. The position dependence of the field is set by the doping profile. Free charge carriers, the quasiparticles electrons and holes, drift in opposite directions in this electric field, as shown in Figure 2.1b. As in gases, they may undergo impact ionisation reactions such as those shown in Figure 2.1c, leading to the creation of additional pairs of electrons and holes and therefore the amplification of the initial charge. According to the Ramo-Shockley theorem in Eq. 1.15, the instantaneous number of drifting charges determines the current I^{ind} that is induced on the metal contacts of the junction and which forms the signal produced by the device.

Of particular interest to the treatment here is the limit of high avalanche gain. For sufficiently strong electric fields above the "breakdown limit" of the diode, the electron-hole avalanche becomes self sustaining, i.e. the number of charge carriers in the avalanche continues to grow exponentially as long as the electric field is maintained. Below the breakdown limit, impact ionisation is insufficient

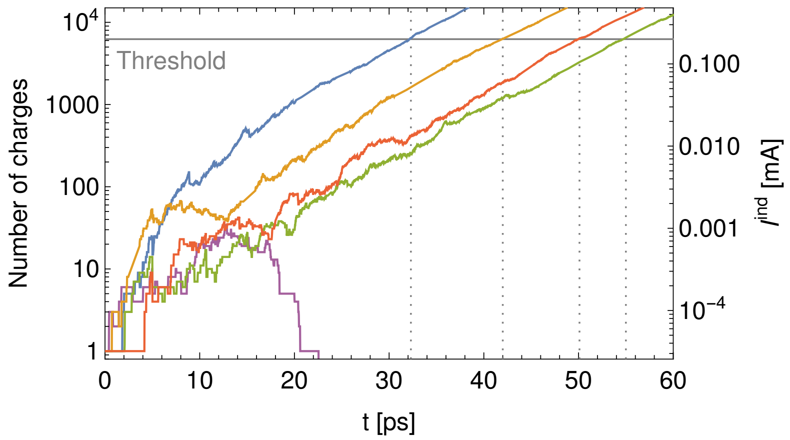
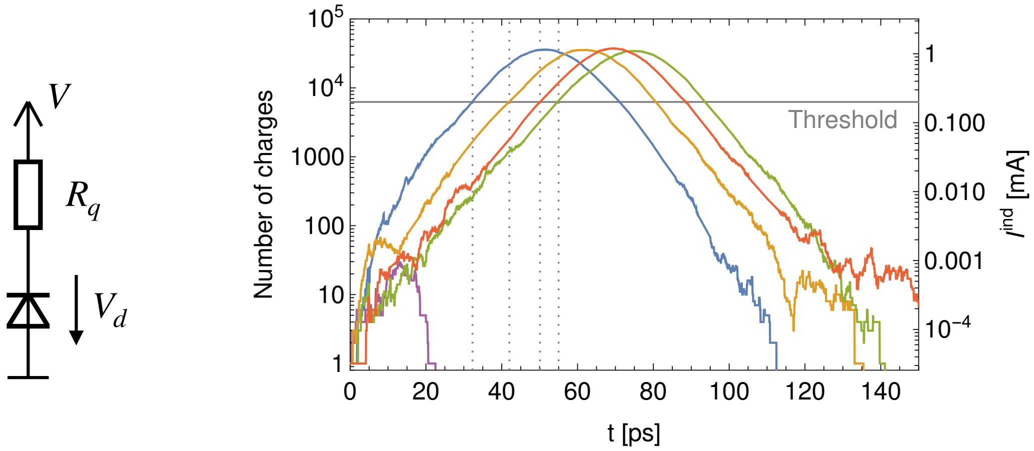


Figure 2.2: The evolution of the total number of charges (electrons and holes) participating in a series of avalanches obtained from a Monte Carlo simulation. Each avalanche is initiated by a single electron and develops in a layer of silicon with a thickness of 0.5 µm to which a reverse bias voltage of 21.8 V is applied. A threshold is applied on the induced current I^{ind}. A vertical dashed line indicates the time at which each avalanche crosses the threshold. One avalanche fails to remain self-sustaining and does not cross the threshold.

to maintain continued growth. The avalanche collapses and stops after a finite time. This case is extensively described in Ref. [57] and not further discussed here. Figure 2.2 shows several simulated avalanches evolving in the regime above breakdown, illustrating several important aspects of their phenomenology that are quantitatively studied in the following.

Impact ionisation reactions occur stochastically, leading to fluctuations in the development of the avalanche. These avalanche fluctuations are particularly pronounced at early times when the avalanche consists of only a small number of charge carriers. In this regime, there is a nontrivial probability for all charge carriers to drift out of the high-field region without undergoing any further impact ionisation reactions, which ends the avalanche discharge. Even above breakdown, the "efficiency" for a certain initial charge deposit to trigger a self-sustaining avalanche remains finite.

For sufficiently large avalanches, the avalanche fluctuations average out and only a deterministic, exponential growth remains. Nonetheless, the initial fluctuations have important consequences for the avalanche evolution at late times and for the operation of a practical device. Such a device may be operated as a counting detector which registers an event when the avalanche current I^{ind} exceeds a certain predetermined threshold. The threshold is usually chosen to correspond to a current large enough to simplify the design of the detector electronics. As Figure 2.2 shows, the avalanche fluctuations generate variations in the threshold crossing time and therefore limit the time resolution of the detector.



(a) (b)

Figure 2.3: (a) Simple passive circuit to ensure quenching of large avalanches. (b) Continued evolution of the avalanches from Figure 2.2.

In any practical detector, the exponential growth of the avalanche must be stopped after the threshold crossing and all charge carriers removed from the junction so that the device is prepared for the detection of the next event. Figure 2.3a shows a simple example of such a “quenching” circuit, where the p–n junction is connected to the bias voltage supply through a resistance R_q. As the avalanche grows, the induced current leads to a reduction of the bias voltage V_d to values below the breakdown limit and eventually to the collapse of the avalanche, shown in Figure 2.3b. The dynamics of the quenching is determined by the interplay of the avalanche and the external quenching circuitry. Provided that the avalanche becomes self-sustaining, it ensures that the signals produced by the device for subsequent events have a constant amplitude and identical, deterministic shape, independent of the amount or distribution of the initial charge.

The remainder of this chapter covers the full evolution of the electron-hole avalanche as set out above. Section 2.1 formulates a simplified model of the avalanche which forms the basis for all subsequent investigations. Extracting predictions from any kind of model requires a well-adapted set of mathematical tools. Stochastic avalanches are most effectively described using the language of many-body quantum mechanics, suitably modified to apply to classical systems. These modifications are detailed in Section 2.2. The next two sections study in detail the emergence of avalanche fluctuations at early times. Section 2.3 lays the groundwork by considering the idealised scenario of an infinitely extended semiconductor and a constant electric field. Realistic geometries encountered in a practical device, i.e. thin avalanche regions and inhomogeneous electric fields, are considered in Section 2.4. In some cases, the analytic description of the avalanche development requires calculations of a certain technical complexity. The main steps of these calculations are outlined in the main text and summaries of the most important conclusions are available at the end of each section. All arguments are sufficiently general to apply to a variety of semiconductor materials. Silicon is of great practical relevance for the construction of detectors, so that a more detailed discussion of this case is in order, given in Section 2.5. Finally, Section 2.6 describes the quenching process.

2.1 Avalanche model and assumptions

In practical realisations of the device in Figure 2.1a, the p–n junction is often part of a pixel structure. The pixel size in the direction transverse to the electric field, denoted as D, is of the order of 5–15 µm [58]. This is much larger than the longitudinal thickness d of the high-field region. In the limit $D/d \gg 1$, the configuration of the electric field is effectively one-dimensional and the development of the avalanche can be modelled as such with good approximation. (This does not apply to cases where the avalanche develops close to the transverse boundary of the diode, where the electric field has a more complicated structure.) In a physical device, the avalanche discharge forms a cylindrical "microplasma tube" with a nontrivial phenomenology [59]. In a parallel field geometry, its transverse extent is determined by diffusion and drift of the participating charge carriers. This does not affect the statistics of the avalanche process and is thus irrelevant for many of the phenomena discussed here.

Figure 2.4a lays out the coordinate system used to describe the high-field region, also referred to as the "avalanche region" in the following. The electric field generated by the reverse bias voltage V_d is taken to point in the negative x-direction. The drift velocities of electrons and holes depend on the local electric-field strength $\mathbf{E}(x)$. With the chosen field direction, electrons drift towards positive x with velocity $v_e(x)$ and holes drift towards negative x with velocity $v_h(x)$. Impact ionisation processes are modelled as shown in Figure 2.4b. The probability for an electron to multiply in the small spatial interval dx is taken to be $\alpha(x)dx$. Analogously, the probability for a hole to multiply is given by $\beta(x)dx$. The probabilities for a multiplication to occur in a small time interval dt are then $\alpha(x)v_e(x)dt$ and $\beta(x)v_h(x)dt$. The impact ionisation coefficients $\alpha(x)$ and $\beta(x)$ encode the details of the underlying scattering processes in the material. They are strongly dependent on the local electric field strength. In analogy to avalanches in gases, α and β are sometimes referred to as "Townsend coefficients". They are treated here as external input parameters for which parameteri-

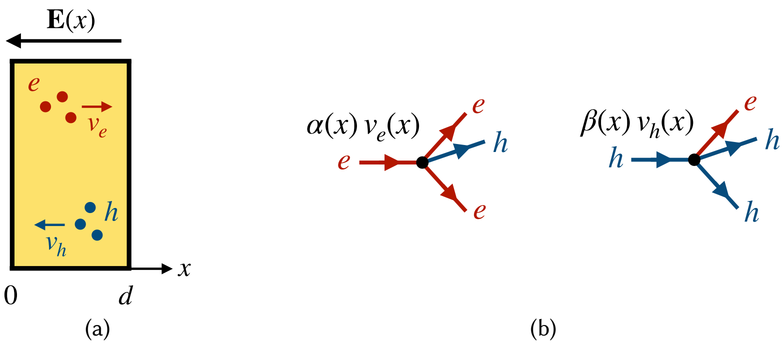


Figure 2.4: (a) Electrons and holes drift in opposite directions with velocities $v_e(x)$ and $v_h(x)$. (b) Impact ionisation processes proceed with a rate $\alpha(x)v_e(x)$ when initiated by an electron, and with a rate $\beta(x)v_h(x)$ when initiated by a hole.

sations are available in the literature for specific materials (cf. Section 2.5 for the case of silicon).

The impact ionisation coefficients are assumed to be local quantities, parameterising the probability for impact ionisation to occur at a certain position x. Any dependence on the history of the charge carriers, such as on the distance travelled by the charge carrier since the last ionisation event, are neglected here. These "dark space" effects [60] modify the details of the avalanche evolution but do not drastically alter its phenomenology.

Avalanche fluctuations arise in the early stages of the evolution where the total number of charge carriers is small and the backreaction on the externally impressed field caused by space-charge effects is negligible. In the following, all charge carriers in the avalanche are therefore assumed to evolve independently, irrespective of the total number of charges already present in the avalanche.

2.2 Avalanches as stochastic many-body systems

With the assumptions laid out in Section 2.1, the avalanche satisfies the Markov property [61]: a full description of the avalanche at a certain time t_0, the current "state" of the avalanche, contains enough information to allow the state at any future time $t > t_0$ to be computed.

As a stochastically evolving system, a complete description of the state of the avalanche is equivalent to a specification of the probability $p(\mathcal{A})$ with which a certain avalanche configuration $\mathcal{A}$ occurs. An "avalanche configuration" is a sufficiently detailed description of a concrete *realisation* of an avalanche. Exactly which amount of information constitutes a "sufficient" set depends on the situation. In the most general (and most complex) case, $\mathcal{A}$ contains the spatial positions of all participating electrons and holes, or equivalently, their spatial densities $n_e(x)$ and $n_h(x)$. In situations where both the material and the electric field are homogeneous, it is sufficient to describe $\mathcal{A}$ through the total numbers of electrons and holes, N_e and N_h. Both cases are treated in the following.

The probability $p(\mathcal{A})$ changes with time as the avalanche evolves, denoted as $p(\mathcal{A}, t)$. For a Markov process, the time evolution is described by an equation of the form

$$\frac{d}{dt}p(\mathcal{A}, t) = \sum_{\mathcal{A}'} \left[T(\mathcal{A}' \to \mathcal{A})\, p(\mathcal{A}', t) - T(\mathcal{A} \to \mathcal{A}')\, p(\mathcal{A}, t)\right] . \tag{2.1}$$

The two terms on the right-hand side correspond to the two ways in which the probability $p(\mathcal{A}, t)$ may change in the small time interval $[t, t + dt]$. First, an avalanche in configuration $\mathcal{A}'$ may evolve into $\mathcal{A}$, thereby increasing $p(\mathcal{A}, t)$ by an amount proportional to $p(\mathcal{A}', t)$. This is captured by the first (positive) term, where $T(\mathcal{A}' \to \mathcal{A})$ labels the rate at which the transition occurs. The second (negative) term represents avalanches that are evolving away from the configuration $\mathcal{A}$ into another configuration $\mathcal{A}'$, thereby reducing $p(\mathcal{A}, t)$ by a proportional amount.

The set of all transition rates $T(\mathcal{A} \to \mathcal{A}')$ is collectively referred to as the "transition matrix". It encodes the complete dynamics of the avalanche process. Expressing the matrix elements explicitly can involve significant (technical) complexity, particularly in situations where the description of the avalanche configuration $\mathcal{A}$ is very detailed.

The remainder of this section describes an efficient formalism for the solution of Eq. 2.1. It is based on the occupation-number representation familiar from quantum mechanics; a vector in a Fock space is used to represent the state of the avalanche. Operators are defined that may act on this state and modify it in well-defined ways to describe the time evolution of the system. When expressed in terms of these objects, Eq. 2.1 closely resembles the Schrödinger equation. This in no way implies that the avalanche is interpreted as a quantum system. Rather the opposite: the way in which many-body quantum mechanics is conventionally formulated provides a very powerful framework for the description of Markov processes.

The treatment given here is rather loose and certain mathematical liberties are taken to emphasise the physical connections between classical avalanches and quantum mechanics. More complete and more rigorous descriptions of the formalism are available in the literature [62–65] and in the original publication of this work [3].

2.2.1 Avalanche configurations and state vectors

An avalanche configuration $\mathcal{A}$ is associated with a vector $|\mathcal{A}\rangle$, which is an element of a Fock space. A more explicit notation includes the degrees of freedom of the avalanche (its "quantum numbers") as $|\mathcal{A}\rangle = |N_e, N_h\rangle$ or $|\mathcal{A}\rangle = |n_e(x), n_h(x)\rangle$. For now, it is more convenient to suppress these details and to refer to the abstract vector $|\mathcal{A}\rangle$ whenever the nature of the underlying attributes is not relevant. The configuration without any charges (the "vacuum") receives a dedicated symbol, $|0\rangle$.

The vectors $|\mathcal{A}\rangle$ serve as a basis of the vector space from which more complicated states may be assembled. In the most general case, the avalanche is found in configuration $\mathcal{A}$ with probability $p(\mathcal{A}, t)$. The vector $|\psi(t)\rangle$ representing this generic state is given by the "superposition" of the basis states in which the probabilities $p(\mathcal{A}, t)$ take the role of the coefficients,

$$|\psi(t)\rangle = \sum_{\mathcal{A}} p(\mathcal{A}, t) |\mathcal{A}\rangle . \tag{2.2}$$

In this expression, the $\sum$ symbol is used to represent both discrete or continuous sums, depending on the nature of the degrees of freedom in $\mathcal{A}$. This decomposition is compatible with the interpretation of $|\mathcal{A}\rangle$ as states having well-defined avalanche configurations. Eq. 2.2 encodes the full information about the set of probabilities $p(\mathcal{A}, t)$. They may be recovered by projecting $|\psi\rangle$ onto $|\mathcal{A}\rangle$, i.e. $p(\mathcal{A}, t) = \langle\mathcal{A}|\psi\rangle \,/\, \langle\mathcal{A}|\mathcal{A}\rangle$. (The basis states are orthogonal but generally not ortho*normal* with respect to the inner product.)

Eq. 2.2 also exposes the "normalisation condition" imposed on any well-formed state vector: as probabilities, the linear sum over all coefficients must evaluate to unity, i.e. $\sum_{\mathcal{A}} p(\mathcal{A}, t) = 1$.

2.2.2 Observables and expectation values

Given a state $|\psi(t)\rangle$, an important task is to find the expectation value of an observable $\mathcal{O}(\mathcal{A})$ that depends on the avalanche configuration, i.e. to compute the sum

$$\langle\mathcal{O}(\mathcal{A})\rangle = \sum_{\mathcal{A}} \mathcal{O}(\mathcal{A}) p(\mathcal{A}, t). \tag{2.3}$$

Important observables include the total number of electrons or holes in the avalanche, $\mathcal{O}(\mathcal{A}) = N_e$ or $\mathcal{O}(\mathcal{A}) = N_h$, as well as the total number of charge carriers, $\mathcal{O}(\mathcal{A}) = N_e + N_h$. To evaluate Eq. 2.3, it is convenient to introduce a special dual vector $\langle\text{avg}|$. Its defining property is that it has unit overlap with all basis vectors $|\mathcal{A}\rangle$, i.e. $\langle\text{avg}|\mathcal{A}\rangle = 1$ for all $|\mathcal{A}\rangle$. (It is not obvious at this point that such a vector even exists. The explicit form of $\langle\text{avg}|$ is given below in Sections 2.3.1 and 2.4.1.) Each observable $\mathcal{O}(\mathcal{A})$ is associated to an operator $\hat{\mathcal{O}}$. In the basis spanned by the vectors $|\mathcal{A}\rangle$, the action of $\hat{\mathcal{O}}$ is diagonal: the *operator* $\hat{\mathcal{O}}$, acting on $|\mathcal{A}\rangle$, reduces to the *function* $\mathcal{O}(\mathcal{A})$ multiplying the state, i.e. $\hat{\mathcal{O}}\,|\mathcal{A}\rangle = \mathcal{O}(\mathcal{A})\,|\mathcal{A}\rangle$.

With these definitions, the expectation value in Eq. 2.3 may be computed for a general state $|\psi(t)\rangle$ by acting on the state with the operator $\hat{\mathcal{O}}$ and taking the inner product with $\langle\text{avg}|$. This operation gives

$$\langle \text{avg}|\hat{\mathcal{O}}|\psi(t)\rangle = \sum_{\mathcal{A}} p(\mathcal{A},t)\,\langle \text{avg}|\hat{\mathcal{O}}|\mathcal{A}\rangle = \sum_{\mathcal{A}} p(\mathcal{A},t)\mathcal{O}(\mathcal{A})\,\langle \text{avg}|\mathcal{A}\rangle = \langle \mathcal{O}(\mathcal{A})\rangle\,. \tag{2.4}$$

This makes manifest that expectation values are linear functions of the state vector, and thus of the probabilities encoded therein. This is to be contrasted with quantum mechanics where the expectation value of an operator $\hat{\mathcal{O}}$ is $\langle\psi|\hat{\mathcal{O}}|\psi\rangle$ and thus depends quadratically on the amplitudes.

2.2.3 Time evolution

The development of the avalanche is described through the time evolution of the underlying state vector and of important expectation values. The time evolution of $|\psi(t)\rangle$ is implemented by a time translation operator $\hat{U}(dt)$, which encodes the dynamics of the system over an infinitesimal time step. Its defining property is its action on a generic state $|\psi(t)\rangle$,

$$\hat{U}(dt)\,|\psi(t)\rangle = |\psi(t+dt)\rangle\,. \tag{2.5}$$

In the basis $\{|\mathcal{A}\rangle\}$, the time translation operator may be expanded as

$$\hat{U}(dt) = \sum_{\mathcal{A},\mathcal{A}'} p(\mathcal{A}\to\mathcal{A}';dt)\frac{|\mathcal{A}'\rangle\,\langle\mathcal{A}|}{\langle\mathcal{A}|\mathcal{A}\rangle}, \tag{2.6}$$

where $p(\mathcal{A}\to\mathcal{A}';dt)$ labels the probability to transition from configuration $|\mathcal{A}\rangle$ to $|\mathcal{A}'\rangle$ within the time interval dt. (To see that Eq. 2.6 implements the definition in Eq. 2.5, it is sufficient to act with Eq. 2.6 on the expansion of a general state $|\psi(t)\rangle$ in Eq. 2.2.) The time translation operator is closely related to the transition matrix elements $T(\mathcal{A}\to\mathcal{A}')$ from Eq. 2.1,

$$p(\mathcal{A}\to\mathcal{A}';dt) = dt\,T(\mathcal{A}\to\mathcal{A}'),\quad \mathcal{A}\neq\mathcal{A}' \tag{2.7a}$$

$$p(\mathcal{A}\to\mathcal{A};dt) = 1 - dt\sum_{\mathcal{A}'\neq\mathcal{A}} T(\mathcal{A}\to\mathcal{A}'). \tag{2.7b}$$

The operator $\hat{U}(dt)$ may be brought into the form $\hat{U}(dt) = \mathbb{1} + dt\hat{H}$, where $\hat{H}$ is referred to as the "Hamiltonian" of the system. Using Eqs. 2.6 and 2.7, the Hamiltonian may be expressed as

$$\hat{H} = \sum_{\mathcal{A}}\sum_{\mathcal{A}'\neq\mathcal{A}} T(\mathcal{A}\to\mathcal{A}')\frac{|\mathcal{A}'\rangle\,\langle\mathcal{A}|}{\langle\mathcal{A}|\mathcal{A}\rangle} - \sum_{\mathcal{A}}\sum_{\mathcal{A}'\neq\mathcal{A}} T(\mathcal{A}\to\mathcal{A}')\frac{|\mathcal{A}\rangle\,\langle\mathcal{A}|}{\langle\mathcal{A}|\mathcal{A}\rangle}. \tag{2.8}$$

It appears in the time-evolution equation for $|\psi(t)\rangle$,

$$\frac{d}{dt}|\psi(t)\rangle = \hat{H}|\psi(t)\rangle . \tag{2.9}$$

This equation is equivalent to Eq. 2.1: an explicit evolution equation for $p(\mathcal{A},t)$ may be recovered by taking the inner product with $\langle\mathcal{A}|$,

$$\frac{d}{dt}p(\mathcal{A},t) = \frac{d}{dt}\frac{\langle\mathcal{A}|\psi(t)\rangle}{\langle\mathcal{A}|\mathcal{A}\rangle} = \frac{\langle\mathcal{A}|\hat{H}|\psi(t)\rangle}{\langle\mathcal{A}|\mathcal{A}\rangle}. \tag{2.10}$$

Eq. 2.9 also gives rise to time-evolution equations for expectation values. From Eq. 2.4,

$$\frac{d}{dt}\langle\mathcal{O}(\mathcal{A})\rangle = \frac{d}{dt}\langle\text{avg}|\hat{\mathcal{O}}|\psi(t)\rangle = \langle\text{avg}|\hat{\mathcal{O}}\hat{H}|\psi(t)\rangle . \tag{2.11}$$

Time evolution must preserve the normalisation of the state $|\psi\rangle$, i.e. $\langle 1\rangle = \sum_{\mathcal{A}} p(\mathcal{A},t) = 1$ for all times t, and thus $\frac{d}{dt}\langle 1\rangle = 0$. This is compatible with Eq. 2.11 only if the Hamiltonian obeys $\langle\text{avg}|\,\hat{H} = 0$. This constraint is the analogue of the hermiticity requirement $\hat{H}^\dagger = \hat{H}$ in quantum mechanics. Eq. 2.11 may thus also be written in the form

$$\frac{d}{dt}\langle\mathcal{O}(\mathcal{A})\rangle = \langle\text{avg}|\left[\hat{\mathcal{O}},\hat{H}\right]|\psi(t)\rangle , \tag{2.12}$$

with the commutator $\left[\hat{\mathcal{O}},\hat{H}\right] = \hat{\mathcal{O}}\hat{H} - \hat{H}\hat{\mathcal{O}}$. The evolution equation in this form is extremely important. It may be used to extract information about the time evolution of the system even in complicated cases where Eq. 2.9 itself can no longer be solved explicitly.

Since $\left[\hat{\mathcal{O}},\mathbb{1}\right] = 0$, the time-evolution operator may be used interchangeably with the Hamiltonian in the evaluation of commutation relations,

$$\left[\hat{\mathcal{O}},\hat{H}\right] = \frac{1}{dt}\left[\hat{\mathcal{O}},\hat{U}(dt)\right].$$

This relation is used extensively in Section 2.4.

2.3 Avalanches in an infinite semiconductor with constant electric field

The simplest interesting scenario is that in which an electron-hole avalanche forms in an infinitely extended semiconductor embedded in a constant electric field, shown in Figure 2.5. The absence of any spatial boundaries across which charges may be removed allows the intrinsic dynamics of the

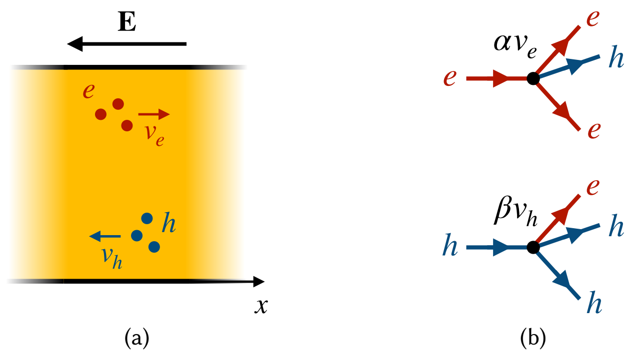


Figure 2.5: (a) A one-dimensional semiconductor of infinite spatial extent, exposed to a constant electric field **E**. Electrons and holes drift in opposite directions with constant velocities v_e and v_h. (b) Multiplication processes occur in the material at a rate independent of the position x.

avalanche to be studied in isolation and the most salient features of its development to be exposed.

In these conditions, the drift velocities v_e and v_h as well as the impact ionisation coefficients α and β become independent of the position x. All spatial locations are equivalent and the evolution of the avalanche is fully determined in terms of the total numbers of electrons and holes, N_e and N_h. It is not necessary, therefore, to follow the evolution of the spatial carrier densities $n_e(x)$ and $n_h(x)$, which greatly simplifies the discussion.

2.3.1 Fock space and Hamiltonian

The formalism introduced in Section 2.2 is straightforwardly specialised to the situation at hand. The particle content of the basis vectors is now made explicit in the notation: $|\mathcal{A}\rangle = |N_e, N_h\rangle$ represents an avalanche configuration consisting of N_e electrons and N_h holes.

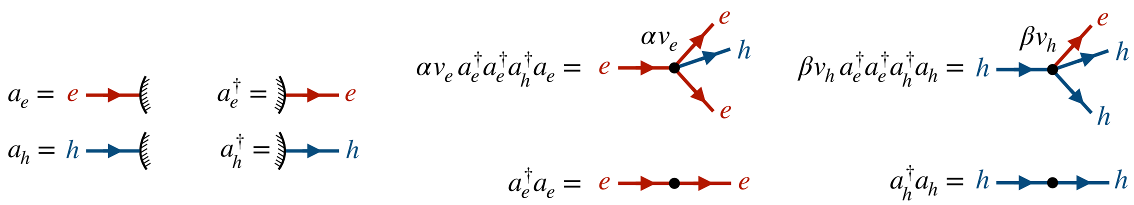
It is convenient to introduce several operators that allow these states to be manipulated.

Ladder algebra The basis vectors may be constructed from the empty avalanche $|0\rangle$ by adding the required numbers of charges,

$$|N_e, N_h\rangle = \left(\hat{a}_e^\dagger\right)^{N_e} \left(\hat{a}_h^\dagger\right)^{N_h} |0\rangle ,$$

where the operators $\hat{a}_e^\dagger$ and $\hat{a}_h^\dagger$ are the creation operators for electrons and holes. They increase the occupation numbers of the corresponding particle species by one,

$$\hat{a}_e^\dagger |N_e, N_h\rangle = |N_e + 1, N_h\rangle , \qquad \hat{a}_h^\dagger |N_e, N_h\rangle = |N_e, N_h + 1\rangle . \tag{2.13}$$



(a) (b)

Figure 2.6: (a) Diagrammatic representation of annihilation operators "absorbing" a charge and of creation operators "emitting" a charge. The hatched blob may be replaced with an arbitrary diagram. (b) Diagrams corresponding to the four terms in the Hamiltonian in Eq. 2.16.

The corresponding annihilation operators $\hat{a}_e$ and $\hat{a}_h$ reduce the corresponding occupation numbers,

$$\hat{a}_e \left| N_e, N_h \right\rangle = N_e \left| N_e - 1, N_h \right\rangle, \qquad \hat{a}_h \left| N_e, N_h \right\rangle = N_h \left| N_e, N_h - 1 \right\rangle. \tag{2.14}$$

Eqs. 2.13 and 2.14 imply that $\hat{a}_e^\dagger$ and $\hat{a}_e$ (as well as $\hat{a}_h^\dagger$ and $\hat{a}_h$) are noncommuting operators. They satisfy the commutation relations familiar from quantum mechanics, which are collected in Appendix A.1. The special vector $\langle \text{avg}|$ arising in Eq. 2.4 may be expressed in terms of the annihilation operators, $\langle \text{avg}| = \langle 0| \, e^{\hat{a}_e + \hat{a}_h} = \langle 0| \sum_{n=0}^{\infty} \frac{(\hat{a}_e + \hat{a}_h)^n}{n!}$.

Operators to determine the occupation numbers are constructed in the usual way. The number operators for electrons and holes are

$$\hat{N}_e = \hat{a}_e^\dagger \hat{a}_e, \qquad \hat{N}_h = \hat{a}_h^\dagger \hat{a}_h,$$

and the states $|N_e, N_h\rangle$ are eigenstates for these operators,

$$\hat{N}_e \left| N_e, N_h \right\rangle = N_e \left| N_e, N_h \right\rangle, \qquad \hat{N}_h \left| N_e, N_h \right\rangle = N_h \left| N_e, N_h \right\rangle. \tag{2.15}$$

Avalanche Hamiltonian The avalanche Hamiltonian is readily obtained from Eq. 2.8. Transitions occur between the states $|N_e, N_h\rangle$ and $|N_e + 1, N_h + 1\rangle$ when one of the N_e electrons or one of the N_h holes undergoes a multiplication reaction and produces an additional electron-hole pair. By definition of the impact ionisation coefficients in Section 2.1, the transition rate is

$$T(N_e, N_h \to N_e + 1, N_h + 1) = \alpha v_e N_e + \beta v_h N_h.$$

Inserting this relation into Eq. 2.8 and using Eqs. 2.13 and 2.15 yields the Hamiltonian

$$\hat{H} = \alpha v_e (\hat{a}_e^\dagger \hat{a}_h^\dagger \hat{N}_e - \hat{N}_e) + \beta v_h (\hat{a}_e^\dagger \hat{a}_h^\dagger \hat{N}_h - \hat{N}_h). \tag{2.16}$$

It admits a very intuitive diagrammatic interpretation. Creation and annihilation operators are conveniently represented as lines entering or leaving any given diagram as shown in Figure 2.6a. With this convention, the four terms in Eq. 2.16 lead to the diagrams depicted in Figure 2.6b. They correspond to the possible interactions that are available to each charge carrier. The probability per unit time for an interaction to happen appears as a numerical coefficient in the Hamiltonian and plays the role of the coupling constant of the corresponding interaction vertices.

2.3.2 Evolution of the full state

The evolution of the probabilities $p(\mathcal{A}, t) \equiv p(N_e, N_h, t)$ is determined by Eq. 2.10. With the Hamiltonian in Eq. 2.16, it becomes

$$\frac{d}{dt}p(N_e, N_h, t) = p(N_e - 1, N_h - 1, t)\left[\alpha v_e(N_e - 1) + \beta v_h(N_h - 1)\right] \\ - p(N_e, N_h, t)\left[\alpha v_e N_e + \beta v_h N_h\right], \tag{2.17}$$

which is the explicit form of the general Eq. 2.1. Its right-hand side expresses the rates of the allowed transitions between the configurations $|N_e, N_h\rangle$ and $|N_e + 1, N_h + 1\rangle$. Eq. 2.17 may be solved in the general case of an avalanche initiated by N_e^0 electrons and N_h^0 holes at $t = 0$, corresponding to the initial condition $p(N_e, N_h, t = 0) = \delta_{N_e, N_e^0}\,\delta_{N_h, N_h^0}$.

Evolution of the number of additional electron-hole pairs For $t > 0$, the avalanche can only reach configurations which differ from its initial state by an integer number of electron-hole pairs. The number of additional electron-hole pairs produced is denoted as K_{eh} and directly related to N_e and N_h as $N_e(K_{eh}) = N_e^0 + K_{eh}$ and $N_h(K_{eh}) = N_h^0 + K_{eh}$. It is thus natural to first study the evolution of this variable. Eq. 2.17 may be expressed as an equation for the probabilities $p(K_{eh}, t)$,

$$\frac{d}{dt}p(K_{eh}, t) = (K_{eh} - 1)\left(\alpha v_e + \beta v_h\right) p(K_{eh} - 1, t) - K_{eh}\left(\alpha v_e + \beta v_h\right) p(K_{eh}, t) + \\ + \left(\alpha v_e N_e^0 + \beta v_h N_h^0\right)\left[p(K_{eh} - 1, t) - p(K_{eh}, t)\right], \tag{2.18}$$

together with the initial condition $p(K_{eh}, t = 0) = \delta_{K_{eh}, 0}$.

This is an infinite family of coupled ordinary differential equations, which is most easily analysed in the z-domain. Upon inserting the z-transform [66] of $p(K_{eh}, t)$, defined as $P(z, t) := \sum_{K_{eh}=0}^{\infty} p(K_{eh}, t) z^{-K_{eh}}$, Eq. 2.18 turns into a single partial differential equation in t and z. Ap-

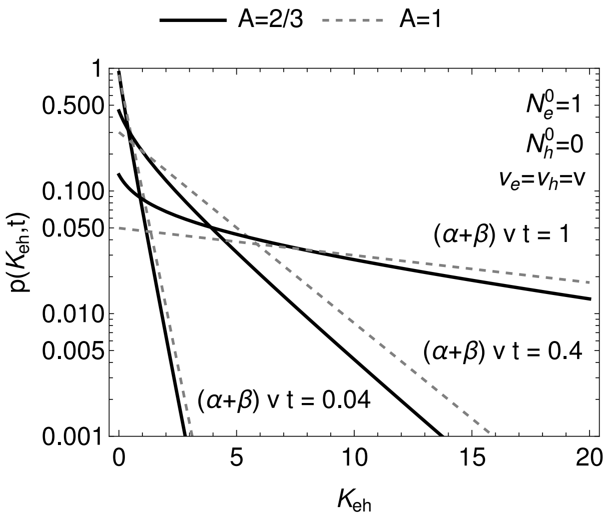


Figure 2.7: Probability $p(K_{eh}, t)$ to find K_{eh} additional electron-hole pairs at time t in an avalanche triggered by a single electron at $t = 0$. The probability $p(K_{eh}, t)$ is defined only for integer values of K_{eh} but is drawn here as a continuous function for visual clarity. The drift velocities of electrons and holes are taken to be equal, $v_e = v_h = v$. The dashed grey line corresponds to an avalanche in which only electrons multiply, $\beta = 0$. The solid black line shows an electron-hole avalanche with identical $\alpha + \beta$ but $\beta \neq 0$, leading to $A = 2/3$ in this example.

plying the inverse z-transform to its solution yields the following expression for $p(K_{eh}, t)$,

$$p(K_{eh}, t) = \frac{\Gamma(A + K_{eh})}{\Gamma(A)\Gamma(1 + K_{eh})} \left(\frac{1}{\nu(t)}\right)^A \left(1 - \frac{1}{\nu(t)}\right)^{K_{eh}}. \tag{2.19}$$

with

$$\nu(t) = e^{(\alpha v_e + \beta v_h)t}, \qquad A = \frac{\alpha v_e N_e^0 + \beta v_h N_h^0}{\alpha v_e + \beta v_h}. \tag{2.20}$$

The "avalanche parameter" A is very important. It characterises the avalanche fluctuations and appears in expressions for many experimentally relevant expectation values derived in the following. The quantity $\nu(t)$ sets the absolute scale of the exponential growth of the avalanche, to which both electron-induced and hole-induced multiplication reactions contribute symmetrically. The function $\Gamma(z)$ is the gamma function $\Gamma(z) = \int_0^\infty dx\, x^{z-1} e^{-x}$.

Figure 2.7 illustrates $p(K_{eh}, t)$ for different values of the avalanche parameter A. Here and in the following, the drift velocities of electrons and holes are assumed to be identical for simplicity. This is correct to good approximation for saturated drift velocities in silicon, as shown in Section 2.5. For $A = 1$, a case studied more closely in Section 2.3.4, the distribution reduces to an exponential distribution with scale parameter $\nu(t)$. Small avalanches thus remain the most probable ones at all times, but both mean and standard deviation of the distribution increase with t. For general noninteger A, an additional power-law tail appears at small K_{eh}, but the qualitative behaviour of the solution remains unchanged.

For late times and large avalanches, Eq. 2.19 simplifies to

$$p(K_{eh},t) \approx \frac{1}{\Gamma(A)} \left(\frac{1}{\nu(t)}\right)^{A} K_{eh}^{A-1} e^{-\frac{K_{eh}}{\nu(t)}}.$$

In this limit, K_{eh} is distributed according to a gamma distribution with shape parameter A and scale parameter $\nu(t)$. The mean of K_{eh} can be read off as their product, $\langle K_{eh} \rangle = A\,\nu(t)$. The asymptotic distribution may thus also be written in a more physically expressive form as

$$p(K_{eh},t) \approx \frac{1}{\Gamma(A)} \left(\frac{A}{\langle K_{eh} \rangle}\right)^{A} K_{eh}^{A-1} e^{-A\frac{K_{eh}}{\langle K_{eh} \rangle}}. \tag{2.21}$$

Evolution of the total number of charges For the computation of the current I^{ind} induced on the readout electrodes, only the total number N of charge carriers is relevant, $N = N_e + N_h$. It is related to K_{eh} as $N = N^0 + 2K_{eh}$, where $N^0 = N_e^0 + N_h^0$ is the total number of initial charge carriers. The solution in Eq. 2.19 may be used to derive an expression for the distribution of N,

$$p(N,t) = \begin{cases} \frac{\Gamma(A+\frac{N-N^0}{2})}{\Gamma(A)\Gamma(1+\frac{N-N^0}{2})} \left(\frac{1}{\nu(t)}\right)^{A} \left(1-\frac{1}{\nu(t)}\right)^{\frac{N-N^0}{2}} & \text{if } \frac{N-N^0}{2} \in \mathbb{N}, \\ 0 & \text{otherwise.} \end{cases} \tag{2.22}$$

In the limit of large avalanches where $K_{eh} \gg N^0$, the random variables N and K_{eh} become directly proportional, and $p(N,t)$ limits to a gamma distribution of the form of Eq. 2.21,

$$p(N,t) \approx \frac{1}{\Gamma(A)} \left(\frac{A}{\langle N \rangle}\right)^{A} N^{A-1} e^{-A\frac{N}{\langle N \rangle}}. \tag{2.23}$$

Its mean $\langle N(t) \rangle$ is computed below in Section 2.3.3. This result may alternatively be written as

$$p(N,t) \approx \int_0^{\infty} dk_\infty \, \frac{k_\infty^{A-1} e^{-k_\infty A}}{\Gamma(A) A^{-A}} \, \delta\left(N - k_\infty \langle N \rangle\right). \tag{2.24}$$

This formalises what is intuitively apparent from Figure 2.2: at late times, all avalanche realisations become proportional to the average avalanche and grow deterministically at the same rate, $N(t) = k_\infty \langle N(t) \rangle$. The stochasticity generated by the avalanche fluctuations merely affects the overall normalisation controlled by the multiplicative factor k_∞. Eq. 2.24 shows that k_∞ is a random variable that obeys a gamma distribution with a shape parameter of A and a scale parameter of A^{-1}.

2.3.3 Moments

The moments $\langle K_{eh}^m \rangle = \sum_{K_{eh}=0}^{\infty} K_{eh}^m \, p(K_{eh}, t)$ encode information about the average evolution of the avalanche and its fluctuations. They may be obtained directly from the exact solution in Eq. 2.19. The mean and the variance of K_{eh} are the most important moments,

$$\langle K_{eh} \rangle = A \left(\nu(t) - 1 \right), \tag{2.25}$$

$$\sigma(K_{eh})^2 = \left\langle K_{eh}^2 \right\rangle - \langle K_{eh} \rangle^2 = A \, \nu(t) \left(\nu(t) - 1 \right). \tag{2.26}$$

More generally, the expectation values $\langle N_e^m N_h^n \rangle$ may be obtained as solutions to evolution equations of the form of Eq. 2.11. Using $\hat{\mathcal{O}} = \hat{N}_e^m \hat{N}_h^n$ as the observable of interest and evaluating its commutator with the Hamiltonian from Eq. 2.16 returns

$$\frac{d}{dt} \langle N_e^m N_h^n \rangle = \sum_{(l,k) \in S} \binom{m}{l} \binom{n}{k} \left(\alpha v_e \left\langle N_e^{l+1} N_h^k \right\rangle + \beta v_h \left\langle N_e^l N_h^{k+1} \right\rangle \right). \tag{2.27}$$

The sum runs over the set $S = \{(l,k) | 0 \leq l \leq m, 0 \leq k \leq n\} \setminus (m,n)$ which contains all combinations of l and k except that in which $l = m$ and $k = n$ simultaneously. An inspection of the right-hand side of Equation 2.27 shows that it only contains terms $\left\langle N_e^l N_h^k \right\rangle$ with $l + k \leq m + n$. Iterating this equation thus gives rise to a closed system of ordinary differential equations that may be solved systematically for arbitrary m and n. The initial conditions at $t = 0$ are $\langle N_e^m N_h^n \rangle = \left(N_e^0 \right)^m \left(N_h^0 \right)^n$.

For the first moments $\langle N_e \rangle$ and $\langle N_h \rangle$, the evolution equations read

$$\frac{d}{dt} \langle N_e \rangle = \alpha v_e \langle N_e \rangle + \beta v_h \langle N_h \rangle, \qquad \frac{d}{dt} \langle N_h \rangle = \alpha v_e \langle N_e \rangle + \beta v_h \langle N_h \rangle, \tag{2.28}$$

with the general solution

$$\langle N_e \rangle = A\nu(t) + \beta v_h \frac{N_e^0 - N_h^0}{\alpha v_e + \beta v_h}, \qquad \langle N_h \rangle = A\nu(t) - \alpha v_e \frac{N_e^0 - N_h^0}{\alpha v_e + \beta v_h},$$

and

$$\langle N \rangle = 2A\nu(t) - (\alpha v_e - \beta v_h) \frac{N_e^0 - N_h^0}{\alpha v_e + \beta v_h}. \tag{2.29}$$

The three second moments $\left\langle N_e^2 \right\rangle$, $\left\langle N_h^2 \right\rangle$, and $\langle N_e N_h \rangle$ follow the evolution equations

$$\frac{d}{dt} \left\langle N_e^2 \right\rangle = \alpha v_e \langle N_e \rangle + \beta v_h \langle N_h \rangle + 2\alpha v_e \left\langle N_e^2 \right\rangle + 2\beta v_h \langle N_e N_h \rangle, \tag{2.30a}$$

$$\frac{d}{dt}\left\langle N_h^2\right\rangle = \alpha v_e \left\langle N_e\right\rangle + \beta v_h \left\langle N_h\right\rangle + 2\beta v_h \left\langle N_h^2\right\rangle + 2\alpha v_e \left\langle N_e N_h\right\rangle, \tag{2.30b}$$

$$\frac{d}{dt}\left\langle N_e N_h\right\rangle = \alpha v_e \left\langle N_e\right\rangle + \beta v_h \left\langle N_h\right\rangle + \alpha v_e \left\langle N_e^2\right\rangle + \beta v_h \left\langle N_h^2\right\rangle + (\alpha v_e + \beta v_h)\left\langle N_e N_h\right\rangle. \tag{2.30c}$$

The quantity $\sigma(N)^2/\left\langle N\right\rangle^2$ measures the magnitude of fluctuations of the avalanche around its average evolution. Similar measures of "relative fluctuations" may be defined for N_e and N_h individually. The solutions of Eqs. 2.30 reveal the following important relation between the avalanche parameter A and the relative fluctuations at late times

$$\lim_{t\to\infty} \frac{\sigma(N)^2}{\left\langle N\right\rangle^2} = \lim_{t\to\infty} \frac{\sigma(N_e)^2}{\left\langle N_e\right\rangle^2} = \lim_{t\to\infty} \frac{\sigma(N_h)^2}{\left\langle N_h\right\rangle^2} = \frac{1}{A}. \tag{2.31}$$

Eqs. 2.25 and 2.26 show that the timescale on which the relative fluctuations approach this limiting value is given by the "saturation time" $T_{\text{sat}} = (\alpha v_e + \beta v_h)^{-1}$. Notably, this timescale is identical to that governing the average growth of the avalanche.

Eqs. 2.30 also imply that N_e and N_h are maximally Pearson correlated at all times,

$$\rho(N_e, N_h) = \frac{\text{cov}[N_e, N_h]}{\sigma(N_e)\sigma(N_h)} = 1, \tag{2.32}$$

as expected from the fact that impact ionisation symmetrically produces both charge carrier species.

2.3.4 Avalanches driven by a single species

Avalanches driven by a single charge-carrier species form an important special case. For $\beta = 0$, only electrons multiply and the situation reduces to that of a Townsend avalanche in a gas discharge. (The converse case where $\alpha = 0$ and only holes multiply may not be of practical significance.) If one impact ionisation coefficient vanishes but the other one does not, the avalanche parameter A attains an integer value. The expression for $p(K_{eh}, t)$ in Eq. 2.19 then simplifies to

$$p(K_{eh}, t) = \binom{K_{eh} + A - 1}{K_{eh}} \left(\frac{1}{\nu(t)}\right)^{A} \left(1 - \frac{1}{\nu(t)}\right)^{K_{eh}}.$$

For an electron avalanche initiated by N_e^0 electrons, $A = N_e^0$ and

$$p(N_e, t) = \binom{N_e - 1}{N_e - N_e^0} \left(\frac{N_e^0}{\left\langle N_e\right\rangle}\right)^{N_e^0} \left(1 - \frac{N_e^0}{\left\langle N_e\right\rangle}\right)^{N_e - N_e^0},$$

where $\left\langle N_e\right\rangle = N_e^0 e^{\alpha v_e t}$. This is the well-known Yule-Furry law [67, 68] for electron avalanches.

Another important special case of significant practical importance is that of an avalanche initiated by a number of electron hole *pairs*, i.e. $N_e^0 = N_h^0 = N_{eh}^0$. The avalanche parameter is $A = N_{eh}^0$

and the evolution is conveniently described in terms of the total number of electron-hole pairs, $N_{eh} = N_{eh}^0 + K_{eh}$. The statistics of such avalanches is identical to those driven by electrons alone,

$$p(N_{eh}, t) = \binom{N_{eh} - 1}{N_{eh} - N_{eh}^0} \left(\frac{N_{eh}^0}{\langle N_{eh} \rangle}\right)^{N_{eh}^0} \left(1 - \frac{N_{eh}^0}{\langle N_{eh} \rangle}\right)^{N_{eh} - N_{eh}^0},$$

where $\langle N_{eh} \rangle = N_{eh}^0 e^{(\alpha v_e + \beta v_h)t}$.

2.3.5 Time-response function

The current $I^{\text{ind}}(t)$ induced on the readout electrodes of the device may be computed with the Ramo-Shockley theorem, Eq. 1.15. The weighting field for the parallel-plate geometry is constant throughout the semiconductor, $\mathbf{E}_w = E_w \hat{\mathbf{x}}$, and the current produced by the avalanche is

$$I^{\text{ind}}(t) = e_0 \frac{E_w}{V_w} \left(v_e N_e + v_h N_h\right), \tag{2.33}$$

where e_0 is the elementary charge. Eq. 2.33 may also be written as

$$I^{\text{ind}}(t) = e_0 \frac{E_w}{V_w} \frac{v_e + v_h}{2} \left(\frac{v_e - v_h}{v_e + v_h}(N_e^0 - N_h^0) + N\right), \tag{2.34}$$

which shows that the current is fully determined by the total number of charges in the avalanche.

The time at which the current crosses an applied threshold of I_{th} is denoted as t_{th}. The threshold-crossing time is a random variable whose distribution determines the time resolution of the device. Eq. 2.34 shows that the threshold current I_{th} may be converted into an equivalent threshold N_{th} on the total number of charges, which is not generally experimentally accessible. It is thus sufficient to study the statistics of the threshold-crossing time for this case.

The probability distribution of t_{th} is referred to as the "time-response function" $\rho(N_{\text{th}}, t)$. It is closely related to $p(N, t)$,

$$\rho(N_{\text{th}}, t) dt = p(N_{\text{th}}, t) \left[\alpha v_e N_e(N_{\text{th}}) + \beta v_h N_h(N_{\text{th}})\right] dt. \tag{2.35}$$

The factor $[\alpha v_e N_e(N_{\text{th}}) + \beta v_h N_h(N_{\text{th}})]\, dt$ multiplying $p(N_{\text{th}}, t)$ expresses the probability that an avalanche containing N_{th} charges crosses this threshold in a short time interval $[t, t+dt]$ through one further multiplication of an electron or a hole. The time-response function is properly normalised such that $\int_0^\infty dt\, \rho(N_{\text{th}}, t) = 1$. With Eq. 2.22, it may be written as

$$\rho(N_{\text{th}}, t) = (\alpha v_e + \beta v_h) \frac{\Gamma\left(1 + \frac{N_{\text{th}} - N^0}{2} + A\right)}{\Gamma(A)\Gamma\left(1 + \frac{N_{\text{th}} - N^0}{2}\right)} \left(\frac{1}{\nu(t)}\right)^A \left(1 - \frac{1}{\nu(t)}\right)^{\frac{N_{\text{th}} - N^0}{2}}, \tag{2.36}$$

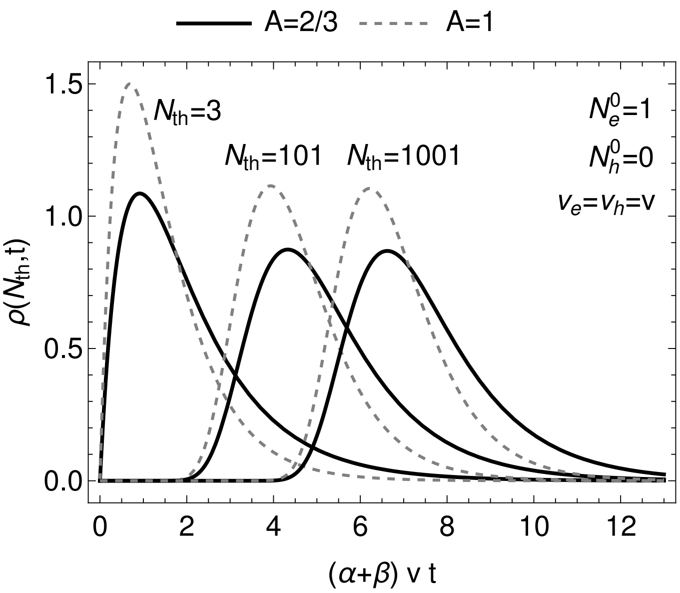


Figure 2.8: Time response function $\rho(N_{\text{th}}, t)$ for an avalanche initiated by a single electron for thresholds $N_{\text{th}} = 3, 101, 1001$. The dashed grey line corresponds to $\beta = 0$, i.e. a pure electron avalanche. The solid black line shows an electron-hole avalanche with identical $\alpha + \beta$ but $\beta \neq 0$, i.e. a non-integer value for the avalanche parameter A.

where it is again understood that $(N_{\text{th}} - N^0)/2 \in \mathbb{N}$. For large thresholds (and late times), the time-response function is approximately given by

$$\rho(N_{\text{th}}, t) \approx \frac{\alpha v_e + \beta v_h}{\Gamma(A)} \exp\left(A \log \frac{N_{\text{th}}}{2\nu(t)} - \frac{N_{\text{th}}}{2\nu(t)}\right). \tag{2.37}$$

Figure 2.8 shows that the shape of the time-response function changes appreciably only for very small N_{th}, caused by the build-up of avalanche fluctuations in this regime. For sufficiently large thresholds, these fluctuations saturate and the shape of the distribution remains stable. In this regime, only a shift in the time argument results as the expected time of the threshold-crossing, $\langle t_{\text{th}} \rangle$, is moved to larger values. This is discussed further in the following section.

2.3.6 Time resolution from time-response function

The time-response function in Eq. 2.36 determines all moments $\langle t_{\text{th}}^n \rangle = \int_0^\infty dt\, t^n \rho(N_{\text{th}}, t)$ of the threshold-crossing time. The first two moments are sufficient for the study of the time resolution. They can be derived as

$$\langle t_{\text{th}} \rangle = \frac{1}{\alpha v_e + \beta v_h} \left[-\psi_0\,(A) + \psi_0 \left(1 + \frac{N_{\text{th}} - N^0}{2} + A\right)\right], \tag{2.38a}$$

$$\langle t_{\text{th}}^2 \rangle = \langle t_{\text{th}} \rangle^2 + \frac{1}{(\alpha v_e + \beta v_h)^2} \left[\psi_1(A) - \psi_1 \left(1 + \frac{N_{\text{th}} - N^0}{2} + A\right)\right]. \tag{2.38b}$$

The functions ψ_0 and ψ_1 belong to the family of polygamma functions ψ_k, defined in terms of derivatives of the gamma function, $\psi_k(z) = d^{k+1} \ln \Gamma(z)/dz^{k+1}$.

The standard deviation of the threshold crossing time, $\sigma_{\text{th}} = \sqrt{\langle t_{\text{th}}^2 \rangle - \langle t_{\text{th}} \rangle^2}$, is used here to quantify the time resolution. (This is sometimes referred to as the "RMS time resolution". The full width at half maximum (FWHM) of the time-response function serves as an alternative measure but is more difficult to control theoretically.)

With Eqs. 2.38, σ_{th} becomes

$$\sigma_{\text{th}}(N_{\text{th}}) = \frac{1}{\alpha v_e + \beta v_h} \sqrt{\psi_1(A) - \psi_1 \left(1 + \frac{N_{\text{th}} - N^0}{2} + A\right)}. \tag{2.39}$$

The time resolution σ_{th} deteriorates as the threshold N_{th} is increased, as anticipated by the discussion following Eq. 2.37. For large thresholds, it saturates at

$$\sigma_{\text{th},\infty} = \lim_{N_{\text{th}} \to \infty} \sigma_{\text{th}}(N_{\text{th}}) = \frac{\sqrt{\psi_1(A)}}{\alpha v_e + \beta v_h}. \tag{2.40}$$

The overall scale of the time resolution is set by the inverse asymptotic growth rate, $(\alpha v_e + \beta v_h)^{-1}$. The dimensionless expression in the numerator depends on the avalanche parameter and describes the magnitude of the avalanche fluctuations.

The convergence of Eq. 2.39 to the limiting value is extremely rapid. As Figure 2.9 shows, the time resolution is virtually identical to its limiting value already for thresholds of $N_{\text{th}} \approx 30$–40. All asymptotic results derived above for "large" thresholds thus hold in excellent approximation for most situations. This result also suggests that quenching the avalanche at early times in order to improve the time resolution is not practical.

For the case of an electron avalanche ($\beta = 0$) initiated by N_e^0 electrons, Eq. 2.40 reduces to

$$\sigma_{\text{th},\infty} = \frac{\sqrt{\psi_1(N_e^0)}}{\alpha v_e} \approx \frac{1}{\alpha v_e} \frac{1}{\sqrt{N_e^0}},$$

where the approximation is valid for large N_e^0. The time resolution improves like $\sigma_{\text{th}} \sim 1/\sqrt{N_e^0}$ as the primary charge increases, as expected from the central limit theorem. For a single initial electron, $N_e^0 = 1$, the time resolution becomes

$$\sigma_{\text{th},\infty} = \frac{1}{\alpha v_e} \frac{\pi}{\sqrt{6}},$$

as originally derived in Ref. [69] for the case of resistive plate chambers.

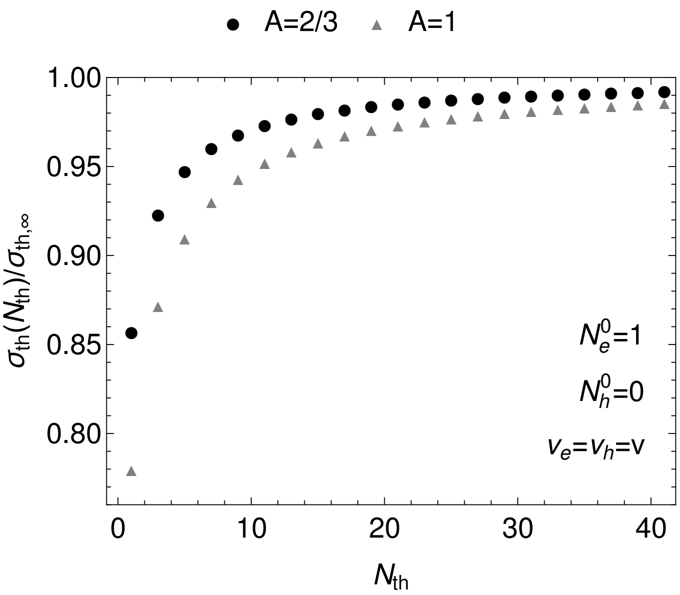


Figure 2.9: Convergence of the time resolution $\sigma_{\text{th}}(N_{\text{th}})$ to its limiting value $\sigma_{\text{th},\infty}$ as N_{th} increases. The grey triangles correspond to $\beta = 0$, i.e. a pure electron avalanche. The black circles are valid for an electron-hole avalanche with identical $\alpha + \beta$ but $\beta \neq 0$, i.e. for a non-integer value for the avalanche parameter A.

2.3.7 Time resolution for large thresholds

The asymptotic time resolution $\sigma_{\text{th},\infty}$ may be computed in a physically more intuitive way that does not require the full time response function $\rho(N_{\text{th}}, t)$ and generalises better to more complicated situations. Eq. 2.24 shows that the avalanche evolves as $N(t) = k_\infty \langle N(t) \rangle$ for late times. With Eq. 2.29, fluctuations of the threshold-crossing time t_{th} are related to fluctuations of the overall normalisation of the avalanche as controlled by the random variable k_∞,

$$t_{\text{th}} = \frac{1}{\alpha v_e + \beta v_h} \log\left(\frac{N_{\text{th}}}{k_\infty}\right) + \text{const}.$$

Its standard deviation $\sigma_{\text{th},\infty}$ is

$$\sigma_{\text{th},\infty} = \frac{\sigma(\log k_\infty)}{\alpha v_e + \beta v_h} = \lim_{t\to\infty} \frac{\sigma(\log N)}{\alpha v_e + \beta v_h}. \tag{2.41}$$

This argument is very general and holds for any multiplication process that leads to (asymptotically) exponential growth. It shows that the asymptotic time resolution is determined by the fluctuations of the logarithm of the number of charges in the avalanche, $\sigma(\log N)$.

The moments $\langle \log N \rangle$ and $\langle \log^2 N \rangle$, and thus $\sigma(\log N)$, may be computed using the general Eq. 2.11 and the Hamiltonian in Eq. 2.16. Evaluating the commutators $[\log \hat{N}, \hat{H}]$ and $[\log^2 \hat{N}, \hat{H}]$

shows that these observables do not, however, admit closed evolution equations. This technical issue can be overcome by constructing operators that are not generally identical to $\log \hat{N}$ and $\log^2 \hat{N}$ but limit to these expressions as the avalanche grows, and for which the evolution equations may be integrated. This calculation is summarised in Appendix A.2. It gives

$$\lim_{t\to\infty} \sigma(\log N) = \sqrt{\psi_1(A)}, \tag{2.42}$$

and so recovers the expression in the numerator of Eq. 2.40.

2.3.8 Summary

The most important properties of electron-hole avalanches developing in an unbounded domain are:

(i) The average charge content of the avalanche grows exponentially at a rate of $\alpha v_e + \beta v_h$ (cf. Eqs. 2.20 and 2.29).

(ii) The magnitude of avalanche fluctuations around the average evolution is determined by the avalanche parameter A, which depends on the initial conditions (cf. Eq. 2.20 and 2.31).

(iii) Avalanche fluctuations are predominantly created at early times. They saturate on a timescale $(\alpha v_e + \beta v_h)^{-1}$, which is identical to the inverse asymptotic growth rate (cf. Eq. 2.31).

(iv) Asymptotically, i.e. at late times, each individual avalanche grows at the same rate as the average avalanche. In this regime, avalanche fluctuations appear as fluctuations of the total charge content and follow a gamma distribution (cf. Eq. 2.24).

(v) The time resolution for large thresholds, i.e. saturated fluctuations, is given by Eq. 2.40. The overall scale of the time resolution is determined by the inverse asymptotic growth rate.

2.4 Avalanches in a thin semiconductor with arbitrary electric field

The treatment is now generalised to the more realistic geometry of a finite avalanche region and a general position-dependent electric field, as summarised in Figure 2.10. This scenario shares many properties with the simplified situation studied in Section 2.3 but also introduces significant additional technical challenges. However, the basic character of the electron-hole avalanche persists, and the intuition gained from the previous discussion helps to guide the following calculations.

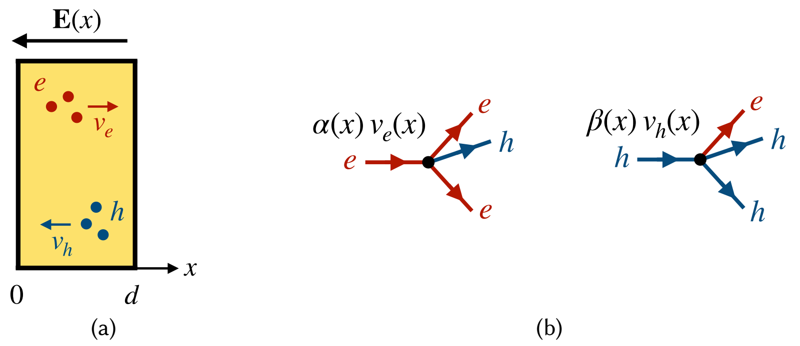


Figure 2.10: (a) An electron-hole avalanche develops in a semiconductor with thickness d, which is exposed to a position-dependent electric field $\mathbf{E}(x)$. Electrons and holes drift in opposite directions with velocities $v_e(x)$ and $v_h(x)$. (b) Impact ionisation processes proceed with a rate $\alpha(x)v_e(x)$ when initiated by an electron, and with a rate $\beta(x)v_h(x)$ when initiated by a hole.

As before, fluctuations of the induced signal I^{ind} continue to be determined by the fluctuations of the instantaneous number of charges present in the avalanche region. The evolution is now complicated by the removal of charges from the avalanche through the boundaries at $x = 0$ and $x = d$, with instantaneous outflow rates proportional to $v_e\, n_e|_{x=d}$ and $v_h\, n_h|_{x=0}$, respectively. The presence of the boundaries modifies the average evolution of the avalanche and also changes the magnitude of its fluctuations around this average. To rigorously describe both effects, it is necessary to determine the evolution of the full spatial carrier densities $n_e(x,t)$ and $n_h(x,t)$ as random variables defined on $x \in [0,d]$.

Applying the methods developed in Section 2.2 to this problems leads to systems of partial differential equations which can be solved efficiently with numerical methods. Analytic solutions of these equations exist in the special case of position-independent electric fields. They are very physically illuminating and descriptive of the general character of the avalanche, complementing the numerical prescriptions.

2.4.1 Fock space and time-evolution operator

An avalanche configuration represented by the basis vector $|\mathcal{A}\rangle$ is now labelled by the spatial charge carrier densities, i.e. $|\mathcal{A}\rangle = |n_e(x), n_h(x)\rangle$. The (integer) number of electrons in a small interval $[x, x+dx]$ is $n_e(x)dx$ and the number of holes in the same region is $n_h(x)dx$. The Fock space construction shown in Section 2.3.1 is straightforwardly generalised to the present higher-dimensional setting.

Operator algebra The basis states $|n_e(x), n_h(x)\rangle$ may again be constructed starting from the vacuum. Schematically,

$$|n_e(x), n_h(x)\rangle = \prod_{\{x\}} \hat{a}_e^\dagger(x)^{n_e(x)dx}\ \hat{a}_h^\dagger(x)^{n_h(x)dx}\,|0\rangle\,,$$

where the creation operators $\hat{a}_e^\dagger(x)$ and $\hat{a}_h^\dagger(x)$ produce an additional electron or hole at position x. Corresponding annihilation operators $\hat{a}_e(x)$ and $\hat{a}_h(x)$ are also introduced. Their commutation relations directly generalise the algebra introduced in Section 2.3.1, again summarised in Appendix A.1. Ladder operators at identical positions continue to obey the relations introduced earlier, and operators at different locations always commute. In terms of these operators, the definition of the averaging state $\langle\text{avg}|$ becomes $\langle\text{avg}| = \langle 0|\, e^{\int dx\, \hat{a}_e(x) + \int dx\, \hat{a}_h(x)}$.

The number operators $\hat{N}_e$ and $\hat{N}_h$ generalise to the number *density* operators

$$\hat{n}_e(x) = \hat{a}_e^\dagger(x)\hat{a}_e(x), \qquad\qquad \hat{n}_h(x) = \hat{a}_h^\dagger(x)\hat{a}_h(x),$$

which determine the electron- or hole-content at position x,

$$\hat{n}_e(y)\,|n_e(x), n_h(x)\rangle = n_e(y)\,|n_e(x), n_h(x)\rangle\,, \qquad \hat{n}_h(y)\,|n_e(x), n_h(x)\rangle = n_h(y)\,|n_e(x), n_h(x)\rangle\,.$$

An additional class of operators is required to implement spatial translations and thus describe the drift of charge carriers in the electric field. A general translation operation T_Δ modifies the position of a charge carrier initially located at x by a position-dependent amount $\Delta(x)$, i.e. it maps $x \mapsto y = x + \Delta(x)$. Rearranging the charge carriers in the way prescribed by T_Δ induces a nontrivial change in the charge densities. An initial density $n(x)$ turns into the transformed density $T_\Delta n(y)$, which incurs an additional Jacobian factor

$$T_\Delta n(y)dy = n(x)dx \Rightarrow T_\Delta n(y) = \frac{n(x)}{1 + \frac{d\Delta(x)}{dx}}. \tag{2.43}$$

Translation operators $\hat{T}_e$ and $\hat{T}_h$ are defined to apply the transformation in Eq. 2.43 to the electron- and hole density, respectively. Their action on the basis states is

$$\hat{T}_e[\Delta(x)]\,|n_e(x), n_h(x)\rangle = |T_\Delta n_e(x), n_h(x)\rangle\,,$$
$$\hat{T}_h[\Delta(x)]\,|n_e(x), n_h(x)\rangle = |n_e(x), T_\Delta n_h(x)\rangle\,,$$

and the resulting commutation relations are collected in Appendix A.1.

Time-evolution operator It is more convenient in the present case to specify the time-evolution operator $\hat{U}(dt)$ instead of the Hamiltonian $\hat{H}$ itself, which may be obtained by generalising the result in Eq. 2.16 for the unbounded and position-independent geometry

Impact ionisation reactions at different positions proceed at their respective local rates. To account for the position-dependence of the impact ionisation coefficients and the drift velocities, the replacements $\hat{N}_e \to \hat{n}_e(x)$, $\hat{N}_h \to \hat{n}_h(x)$, $\alpha \to \alpha(x)$, $\beta \to \beta(x)$, $v_e \to v_e(x)$, and $v_h \to v_h(x)$ must be made in Eq. 2.16 and the expression integrated over all positions x. In the diagrammatic language of Figure 2.6, this introduces a separate interaction vertex for each position x with its appropriate local coupling constant. The translation operators $\hat{T}_e$ and $\hat{T}_h$ are used to implement the drift of all charges during the infinitesimal time interval dt. The full expression for $\hat{U}(dt)$ becomes

$$\begin{aligned}\hat{U}(dt) = \hat{T}_e[v_e(x)dt]\hat{T}_h[-v_h(x)dt]\Big[&\mathbb{1}\\ &+ dt \int dx\, \alpha(x)v_e(x) \left(\hat{a}_e^\dagger(x)\hat{a}_h^\dagger(x)\hat{n}_e(x) - \hat{n}_e(x)\right)\\ &+ dt \int dx\, \beta(x)v_h(x) \left(\hat{a}_e^\dagger(x)\hat{a}_h^\dagger(x)\hat{n}_h(x) - \hat{n}_h(x)\right)\Big].\end{aligned} \tag{2.44}$$

2.4.2 Average development of the avalanche

The first moments $\langle n_e(x,t)\rangle$ and $\langle n_h(x,t)\rangle$ contain information about the average charge content at a particular position x and time t. Their time evolution is determined by the general Eq. 2.12, evaluated for the observables $\hat{n}_e(x,t)$ and $\hat{n}_h(x,t)$. Using Eq. 2.44 and the relations in Appendix A.1 to compute the commutators $[\hat{n}_e(x,t), \hat{U}(dt)]$ and $[\hat{n}_h(x,t), \hat{U}(dt)]$, the evolution equations are

$$\frac{\partial}{\partial t}\langle n_e(x)\rangle + \frac{\partial}{\partial x}v_e(x)\langle n_e(x)\rangle = \alpha(x)v_e(x)\langle n_e(x)\rangle + \beta(x)v_h(x)\langle n_h(x)\rangle\,, \tag{2.45a}$$

$$\frac{\partial}{\partial t}\langle n_h(x)\rangle - \frac{\partial}{\partial x}v_h(x)\langle n_h(x)\rangle = \alpha(x)v_e(x)\langle n_e(x)\rangle + \beta(x)v_h(x)\langle n_h(x)\rangle\,. \tag{2.45b}$$

For general fields $\mathbf{E}(x)$ these equations must be solved numerically with the boundary and initial conditions

$$\begin{aligned}\langle n_e(0,t)\rangle &= 0, & \langle n_h(d,t)\rangle &= 0,\\ \langle n_e(x,0)\rangle &= n_e^0(x), & \langle n_h(x,0)\rangle &= n_h^0(x).\end{aligned}$$

The analytic solutions for constant electric fields are discussed in detail in Section 2.4.2.2.

Eqs. 2.45 are the position-dependent analogues of Eq. 2.28 and immediately reduce to the latter for position-independent impact ionisation coefficients and drift velocities upon integration over x.

They may also be interpreted as continuity equations for the average electron- and hole densities in the avalanche region, for which the terms on the right-hand side represent the rates at which charge carriers are being generated.

2.4.2.1 Condition for breakdown

The above evolution equations contain information about the conditions under which breakdown is achieved and self-sustaining avalanches are possible [70, 71].

Solutions of Eqs. 2.45 take the form $\langle n_e \rangle = n_e(x)e^{St}$ and $\langle n_h \rangle = n_h(x)e^{St}$, where the functions $n_e(x)$ and $n_h(x)$ are yet to be determined. The parameter S represents the average growth rate of the avalanche and $S > 0$ above breakdown. Inserting this ansatz into the evolution equations and setting $S = 0$ shows that the onset of breakdown is located at

$$\int_0^d dx'\, \alpha(x') \exp\left[- \int_0^{x'} dx''\, \left(\alpha(x'') - \beta(x'')\right)\right] = 1. \tag{2.46}$$

For position-independent impact ionisation coefficients, this condition simplifies. Self-sustaining avalanches are possible for

$$d > \frac{\log \frac{\alpha}{\beta}}{\alpha - \beta}. \tag{2.47}$$

It is visualised in Figure 2.11. Both expressions depend only on the impact ionisation coefficients. For a specific material, a parameterisation of the impact ionisation coefficients in terms of the electric field strength may be inserted into Eq. 2.47 to compute the breakdown field E_{br} or the breakdown voltage $V_{\text{br}} = E_{\text{br}}\, d$.

2.4.2.2 Analytic solution for constant electric field

For position-independent impact-ionisation coefficients and drift velocities, Eqs. 2.45 may be solved analytically. The most important properties of these solutions are briefly summarised in the following. A more detailed analysis including a summary of all derivations is available in Ref. [3].

The average densities $\langle n_e(x,t) \rangle$ and $\langle n_h(x,t) \rangle$ may be expressed as a linear combination of "eigenfunctions" $f^e_\lambda(x,t)$ and $f^h_\lambda(x,t)$, indexed by the "eigenvalues" λ,

$$\langle n_e(x,t) \rangle = \sum_\lambda C(\lambda) f^e_\lambda(x,t), \tag{2.48a}$$

$$\langle n_h(x,t) \rangle = \sum_\lambda C(\lambda) f^h_\lambda(x,t). \tag{2.48b}$$

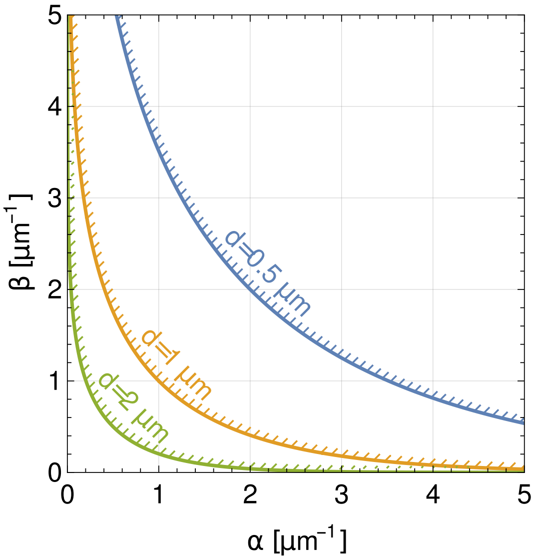


Figure 2.11: Impact ionisation coefficients for which breakdown is possible, as a function of the thickness d of the avalanche region. Breakdown is achieved in the shaded region. The impact ionisation coefficients are assumed to be constant throughout the material.

The constants $C(\lambda)$ are complex-valued coefficients. All components of this solution are discussed in the following. Partial solutions to Eqs. 2.45 have previously been derived in Refs. [72, 73].

Properties of the eigenfunctions The eigenfunctions are given by

$$f^e_\lambda(x,t) = e^{\gamma v^* t} e^{ax} \sinh \kappa x, \tag{2.49a}$$

$$f^h_\lambda(x,t) = \frac{1}{\beta d} \frac{v_e}{v_h} e^{\gamma v^* t} e^{ax} \left[\lambda \sinh \kappa x + \bar{\kappa} \cosh \kappa x\right], \tag{2.49b}$$

with the definitions

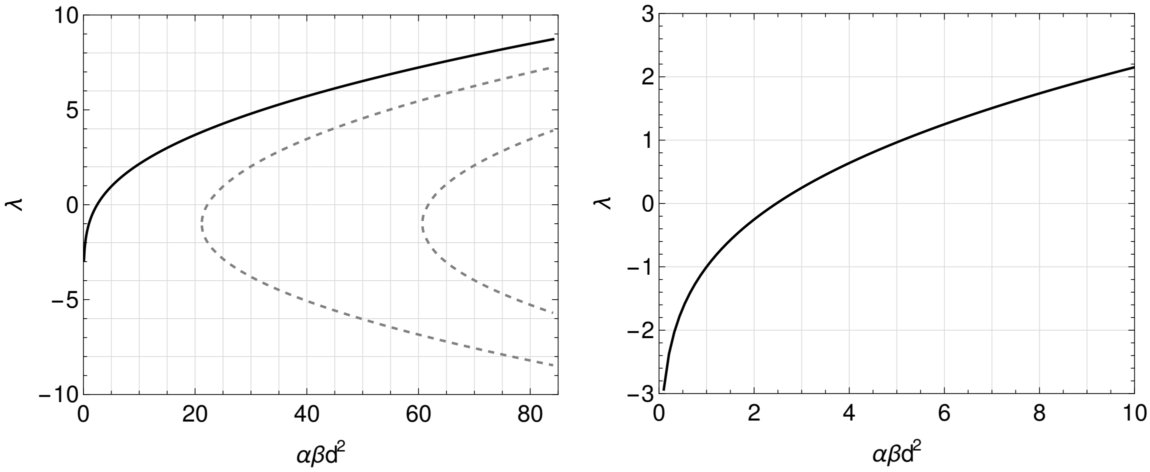
$$\gamma = \frac{\alpha+\beta}{2} + \frac{\lambda}{d}, \tag{2.50a}$$

$$v^* = \frac{2 v_e v_h}{v_e + v_h}, \tag{2.50b}$$

$$a = \frac{v^*}{2}\left(\frac{\alpha}{v_h} - \frac{\beta}{v_e}\right) + \frac{v^*}{2d}\left(\frac{1}{v_h} - \frac{1}{v_e}\right)\lambda, \tag{2.50c}$$

$$\kappa = \frac{\bar{\kappa}}{d} = \frac{1}{d}\sqrt{\lambda^2 - \alpha\beta d^2}. \tag{2.50d}$$

Both f^e_λ and f^h_λ grow exponentially in time with the time constant γv^*. The parameter γ plays the role of an "effective" impact-ionisation coefficient. It depends on the material parameters α and β and on the eigenvalue λ. Eigenfunctions corresponding to different eigenvalues thus have different growth rates. The velocity v^* is the "effective" drift velocity that is relevant for the growth of the avalanche. The spatial dependence of the eigenfunctions is controlled by the parameters a



(a) (b)

Figure 2.12: (a) Real-valued solutions of Eq. 2.51 as a function of the characteristic quantity $\alpha\beta d^2$. The solid black line indicates the largest eigenvalue. The dashed grey lines represent the remaining eigenvalues. (b) The largest eigenvalue for small values of $\alpha\beta d^2$.

and κ, which describe the length scales on which the charge-carrier densities may vary.

Properties of the eigenvalues The sums in Eqs. 2.48 extend over all allowed eigenvalues λ. The eigenvalues form a discrete spectrum and are solutions of the equation

$$\lambda + \bar{\kappa} \coth \bar{\kappa} = 0, \tag{2.51}$$

which depends only on the characteristic dimensionless quantity $\alpha\beta d^2$. The eigenvalue equation has a finite number of real-valued solutions. Above breakdown, the largest real-valued eigenvalue λ_1 leads to a positive time constant $\gamma_1 v^*$, which describes the exponential growth of the avalanche and is analogous to the parameter S introduced in Section 2.4.2.1. Further positive and negative eigenvalues may exist depending on the value of the combination $\alpha\beta d^2$. They introduce additional, subleading, time scales as shown in Figure 2.12.

Eq. 2.51 also admits an infinite number of complex eigenvalues which appear in complex conjugate pairs. These lead to an oscillating behaviour of the eigenfunctions, necessary to describe the wave-like propagation of perturbations in the average densities carried by the drifting charges.

Properties of the coefficients The coefficients $C(\lambda)$ in Eq. 2.48 control the overall normalisation and the absolute average charge content of the avalanche. They are determined from the initial conditions. For general charge-carrier densities $n_e^0(x)$ and $n_h^0(x)$ deposited at $t = 0$, the coefficients

$C(\lambda)$ may be computed as

$$C(\lambda) = \frac{1}{\mathcal{N}(\lambda, t=0)} \int_0^d dx \left[\alpha v_e f_\lambda^e(d-x, t=0) n_e^0(x) + \beta v_h f_\lambda^h(d-x, t=0) n_h^0(x)\right], \quad (2.52)$$

where the normalisation $\mathcal{N}(\lambda, t)$ is

$$\mathcal{N}(\lambda, t) = -e^{2\gamma v^* t} e^{ad} \frac{\alpha d}{2\bar{\kappa}} \frac{v_e}{v_h} (v_e + v_h)(1+\lambda) \sinh \bar{\kappa}. \quad (2.53)$$

For N_e^0 electrons and N_h^0 holes at the position $x = x_0$, i.e. with $n_e^0(x) = N_e^0\, \delta(x - x_0)$ and $n_h^0(x) = N_h^0\, \delta(x - x_0)$, Eq. 2.52 simplifies to

$$C(\lambda) = \frac{1}{\mathcal{N}(\lambda, t=0)} \left[\alpha v_e N_e^0 f_\lambda^e(d - x_0, t=0) + \beta v_h N_h^0 f_\lambda^h(d - x_0, t=0)\right]. \quad (2.54)$$

A similar formula for $C(\lambda_1)$ is mentioned in Ref. [74] for the special case $\alpha = \beta$ and $v_e = v_h$.

Induced current and late-time behaviour For a position-independent weighting field E_w/V_w throughout the avalanche region, the average current induced on the readout electrodes is given by (cf. Eqs. 1.15 and 2.33)

$$\left\langle I^{\text{ind}}(t) \right\rangle = e_0 \frac{E_w}{V_w} \left[v_e \langle N_e(t) \rangle + v_h \langle N_h(t) \rangle\right]. \quad (2.55)$$

The expected total numbers of electrons and holes in the avalanche, $\langle N_e \rangle$ and $\langle N_h \rangle$, are obtained from Eqs. 2.48

$$\langle N_e(t) \rangle = \int_0^d dx\, \langle n_e(x,t) \rangle = \sum_\lambda C(\lambda) \int_0^d dx\, f_\lambda^e(x,t) = \sum_\lambda N_e(\lambda) e^{\gamma v^* t}, \quad (2.56a)$$

$$\langle N_h(t) \rangle = \int_0^d dx\, \langle n_h(x,t) \rangle = \sum_\lambda C(\lambda) \int_0^d dx\, f_\lambda^h(x,t) = \sum_\lambda N_h(\lambda) e^{\gamma v^* t}, \quad (2.56b)$$

where the coefficients $N_e(\lambda)$ and $N_h(\lambda)$ are

$$N_e(\lambda) = C(\lambda) \frac{\kappa + e^{ad}\left(a \sinh \bar{\kappa} - \kappa \cosh \bar{\kappa}\right)}{a^2 - \kappa^2},$$

$$N_h(\lambda) = C(\lambda) \frac{1}{\beta d} \frac{v_e}{v_h} \frac{\kappa\lambda - a\bar{\kappa} - \kappa e^{ad}\left(\lambda \cosh \bar{\kappa} + \bar{\kappa} \sinh \bar{\kappa}\right)}{a^2 - \kappa^2}.$$

Eq. 2.56 is to be compared with the equivalent expression for avalanches in unbounded domains, Eq. 2.29. It shows that, in general, infinitely many time constants $\gamma(\lambda) v^*$ now contribute to the average growth of the avalanche. At late times, however, the average charge content continues to grow at a well-defined rate which is determined by the largest time constant $\gamma_1 v^* = \gamma(\lambda_1) v^*$. In

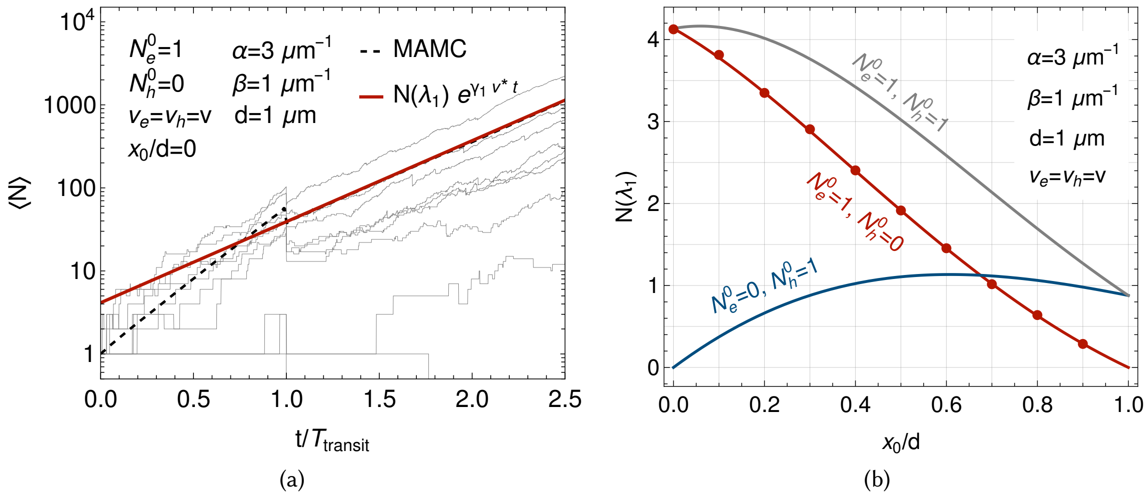


Figure 2.13: (a) Evolution of the expected total number of charges $\langle N \rangle$ as a function of time for an avalanche initiated by a single electron placed at $x_0 = 0$. The dashed black line shows the average obtained from the MAMC simulation model and the solid red line represents the fastest-growing component normalised according to the coefficient $N(\lambda_1)$. The thin grey lines correspond to individual avalanches in the ensemble. (b) The coefficient $N(\lambda_1)$ as a function of the position x_0 of the primary charge, for the case of a single electron (red), a single hole (blue) and a single electron-hole pair (grey). The red markers correspond to the MAMC prediction.

this regime, the average number of charges in the avalanche has the form

$$\langle N \rangle \approx N(\lambda_1) e^{\gamma_1 v^* t}, \tag{2.57}$$

where $N(\lambda_1) = N_e(\lambda_1) + N_h(\lambda_1)$.

Figure 2.13a shows a comparison with results obtained from a Monte Carlo (MC) implementation of the Markov avalanche model of Section 2.1, referred to as "MAMC". Its technical implementation is described in Appendix A.3. The avalanche is initiated by a single electron placed at the upstream end of the avalanche region at $x_0 = 0$. The chosen material parameters are typical for a silicon device. Already after one transit time, $T_{\text{transit}} = d/v^*$, the asymptotic solution in Eq. 2.57 describes the average charge content with excellent precision. The average growth rate of the avalanche changes at $t = T_{\text{transit}}$, caused by electrons leaving the high-field region across the boundary at $x = d$.

Figure 2.13b shows the dependence of the overall normalisation $N(\lambda_1)$ on the position of the primary charge. Avalanches that are initiated by a carrier placed at the corresponding upstream end of the avalanche region, i.e. at $x_0 = 0$ for an initial electron and at $x_0 = d$ for an initial hole, show large overall normalisations. For the chosen impact-ionisation coefficients, the maximum for hole-initiated avalanches is attained in the interior of the avalanche region at around $x_0/d \approx 0.6$.

It occurs at the location for which the losses of electrons and holes across both boundaries are simultaneously minimised.

2.4.3 Development of spatial correlations and fluctuations

The three second moments $\langle n_e(x)n_e(y)\rangle$, $\langle n_h(x)n_h(y)\rangle$, and $\langle n_e(x)n_h(y)\rangle$ contain information about local fluctuations of the charge-carrier densities and their correlations across different spatial locations. Evolution equations for these quantities may be obtained from the general Eq. 2.12 for the observables $\hat{n}_e(x)\hat{n}_e(y)$, $\hat{n}_h(x)\hat{n}_h(y)$, and $\hat{n}_e(x)\hat{n}_h(y)$. Evaluating their commutators with $\hat{U}(dt)$ from Eq. 2.44 yields the following system of equations,

$$\begin{aligned}\frac{\partial}{\partial t}\langle n_e(x)n_e(y)\rangle + \frac{\partial}{\partial x}v_e(x)\langle n_e(x)n_e(y)\rangle + \frac{\partial}{\partial y}v_e(y)\langle n_e(x)n_e(y)\rangle \\ = \alpha(x)v_e(x)\delta(x-y)\langle n_e(x)\rangle + \beta(x)v_h(x)\delta(x-y)\langle n_h(y)\rangle \\ + \beta(y)v_h(y)\langle n_e(x)n_h(y)\rangle + \beta(x)v_h(x)\langle n_e(y)n_h(x)\rangle \\ + [\alpha(x)v_e(x)+\alpha(y)v_e(y)]\langle n_e(x)n_e(y)\rangle,\end{aligned} \tag{2.58a}$$

$$\begin{aligned}\frac{\partial}{\partial t}\langle n_h(x)n_h(y)\rangle - \frac{\partial}{\partial x}v_h(x)\langle n_h(x)n_h(y)\rangle - \frac{\partial}{\partial y}v_h(y)\langle n_h(x)n_h(y)\rangle \\ = \alpha(x)v_e(x)\delta(x-y)\langle n_e(y)\rangle + \beta(x)v_h(x)\delta(x-y)\langle n_h(y)\rangle \\ + \alpha(y)v_e(y)\langle n_e(y)n_h(x)\rangle + \alpha(x)v_e(x)\langle n_e(x)n_h(y)\rangle \\ + [\beta(x)v_h(x)+\beta(y)v_h(y)]\langle n_h(x)n_h(y)\rangle,\end{aligned} \tag{2.58b}$$

$$\begin{aligned}\frac{\partial}{\partial t}\langle n_e(x)n_h(y)\rangle + \frac{\partial}{\partial x}v_e(x)\langle n_e(x)n_h(y)\rangle - \frac{\partial}{\partial y}v_h(y)\langle n_e(x)n_h(y)\rangle \\ = \alpha(x)v_e(x)\delta(x-y)\langle n_e(y)\rangle + \beta(x)v_h(x)\delta(x-y)\langle n_h(y)\rangle \\ + \alpha(y)v_e(y)\langle n_e(x)n_e(y)\rangle + \beta(x)v_h(x)\langle n_h(x)n_h(y)\rangle \\ + [\alpha(x)v_e(x)+\beta(y)v_h(y)]\langle n_e(x)n_h(y)\rangle,\end{aligned} \tag{2.58c}$$

with the boundary conditions

$$\begin{aligned}\langle n_e(0,t)n_e(y,t)\rangle &= 0, & \langle n_h(d,t)n_h(y,t)\rangle &= 0,\\ \langle n_e(0,t)n_h(y,t)\rangle &= 0, & \langle n_e(x,t)n_h(d,t)\rangle &= 0,\end{aligned}$$

and the initial conditions

$$\langle n_e(x,0)n_e(y,0)\rangle = n_e^0(x)n_e^0(y),$$

$$\langle n_h(x,0)n_h(y,0)\rangle = n_h^0(x)n_h^0(y),$$

$$\langle n_e(x,0)n_h(y,0)\rangle = n_e^0(x)n_h^0(y).$$

These equations are the direct position-dependent analogues of the evolution equations for the moments $\langle N_e^m N_h^n \rangle$ encountered in Section 2.3. For constant impact-ionisation coefficients and drift velocities they reduce to Eqs. 2.30 upon integration over x and y.

The second moments determined by Eqs. 2.58 still carry information about the absolute scale of the avalanche which is already contained in the solutions for $\langle n_e \rangle$ and $\langle n_h \rangle$. It is convenient to subtract this contribution and to consider instead the covariances

$$\text{cov}[n_e(x), n_e(y)] = \langle n_e(x)n_e(y)\rangle - \langle n_e(x)\rangle \langle n_e(y)\rangle\,, \tag{2.59a}$$

$$\text{cov}[n_h(x), n_h(y)] = \langle n_h(x)n_h(y)\rangle - \langle n_h(x)\rangle \langle n_h(y)\rangle\,, \tag{2.59b}$$

$$\text{cov}[n_e(x), n_h(y)] = \langle n_e(x)n_h(y)\rangle - \langle n_e(x)\rangle \langle n_h(y)\rangle\,. \tag{2.59c}$$

The diagonal components ($x = y$) of $\text{cov}[n_e(x), n_e(y)]$ and $\text{cov}[n_h(x), n_h(y)]$ correspond to the variances of the charge- carrier densities and measure the magnitude of fluctuations at a certain spatial location. The off-diagonal entries ($x \neq y$) contain information about the correlations of the charge densities at different positions.

Starting from Eqs. 2.58, it is possible to derive evolution equations for the covariances defined in Eqs. 2.59 [3]. These equations may be solved numerically for general field profiles.

2.4.3.1 Analytic solution for constant electric field

For position-independent impact-ionisation coefficients and drift velocities, analytic expressions for the covariances in Eqs. 2.59 may be derived. Of particular significance is the limit $t \to \infty$ of these solutions, which describes the fluctuations and the correlation structure of large avalanches, and thus contains information about the asymptotic time resolution. The detailed analysis is summarised in Ref. [3], with the most important results outlined below.

In analogy to Eq. 2.48, the covariances may be expanded in terms of eigenfunctions. In the late-time limit, $t \to \infty$, this decomposition is given in terms of the functions f_λ^e and f_λ^h (cf. Eqs. 2.49),

$$\text{cov}[n_e(x), n_e(y)] \approx \sum_{\lambda_a, \lambda_b} C(\lambda_a, \lambda_b, t) f_{\lambda_a}^e(x,t) f_{\lambda_b}^e(y,t), \tag{2.60a}$$

$$\text{cov}[n_e(x), n_h(y)] \approx \sum_{\lambda_a, \lambda_b} C(\lambda_a, \lambda_b, t) f_{\lambda_a}^e(x,t) f_{\lambda_b}^h(y,t), \tag{2.60b}$$

$$\text{cov}[n_h(x), n_h(y)] \approx \sum_{\lambda_a, \lambda_b} C(\lambda_a, \lambda_b, t) f^h_{\lambda_a}(x, t) f^h_{\lambda_b}(y, t), \tag{2.60c}$$

where $C(\lambda_a, \lambda_b, t)$ are complex-valued coefficients. The sums over λ_a and λ_b run over the set of solutions of Eq. 2.51.

Late-time behaviour The fastest-growing component in Eqs. 2.60 is that for which $\lambda_a = \lambda_b = \lambda_1$. The corresponding coefficient $C(\lambda_1, \lambda_1, t)$ approaches a constant $C_\infty(\lambda_1, \lambda_1)$ as $t \to \infty$,

$$C(\lambda_1, \lambda_1, t) \approx C_\infty(\lambda_1, \lambda_1) \left(1 - e^{-\gamma_1 v^* t}\right), \tag{2.61}$$

where the saturation timescale T_{sat} is given by $T_{\text{sat}} = (\gamma_1 v^*)^{-1}$, in complete analogy to the result for unbounded geometries discussed below Eq. 2.31. The constant C_∞ may be computed as

$$\begin{aligned} C_\infty(\lambda_1, \lambda_1) = 2\alpha\beta v_e v_h \int_0^\infty dt' \frac{1}{\mathcal{N}(\lambda_1, t')^2} \int_0^d dx \int_0^x dy \, \Big[\alpha v_e f^e_{\lambda_1}(d - x, t') f^e_{\lambda_1}(d - y, t') \\ + \beta v_h f^h_{\lambda_1}(d - x, t') f^h_{\lambda_1}(d - y, t')\Big] \, f^{eh}_1(x, y, t'), \end{aligned} \tag{2.62}$$

where $\mathcal{N}(\lambda, t)$ is taken from Eq. 2.53, and f^{eh}_1 is defined as

$$\begin{aligned} f^{eh}_1(x, y, t) = \frac{1}{v_e + v_h} \exp\left(\frac{\alpha v_e + \beta v_h}{v_e + v_h}(x - y)\right) \big[\alpha v_e \left\langle n_e\left(\bar{x}, t_{\text{ret}}\right)\right\rangle + \beta v_h \left\langle n_h\left(\bar{x}, t_{\text{ret}}\right)\right\rangle \\ + \alpha v_e f^{ee}_1\left(\bar{x}, t_{\text{ret}}\right) + \beta v_h f^{hh}_1\left(\bar{x}, t_{\text{ret}}\right)\Big], \end{aligned} \tag{2.63}$$

with

$$t_{\text{ret}} = t - \frac{x - y}{v_e + v_h}, \qquad \bar{x} = \frac{v_h x + v_e y}{v_e + v_h}.$$

The integral in Eq. 2.62 also exists analytically, but the explicit expression is too lengthy to be of much practical use. Retaining only the most important contribution, as $t \to \infty$, the covariances in Eqs. 2.60 become

$$\text{cov}[n_e(x), n_e(y)] \approx C_\infty(\lambda_1, \lambda_1) f^e_{\lambda_1}(x, t) f^e_{\lambda_1}(y, t), \tag{2.64a}$$

$$\text{cov}[n_h(x), n_h(y)] \approx C_\infty(\lambda_1, \lambda_1) f^h_{\lambda_1}(x, t) f^h_{\lambda_1}(y, t), \tag{2.64b}$$

$$\text{cov}[n_e(x), n_h(y)] \approx C_\infty(\lambda_1, \lambda_1) f^e_{\lambda_1}(x, t) f^h_{\lambda_1}(y, t). \tag{2.64c}$$

The correlation structure in Eqs. 2.64 has important consequences for the behaviour of the avalanche. In particular, it implies that the Pearson correlation coefficients between the charge-carrier densities at positions x and y are

$$\lim_{t\to\infty} \rho\left(n_e(x), n_e(y)\right) = \lim_{t\to\infty} \frac{\mathrm{cov}[n_e(x), n_e(y)]}{\sigma(n_e(x))\sigma(n_e(y))} = 1,$$
$$\lim_{t\to\infty} \rho\left(n_h(x), n_h(y)\right) = \lim_{t\to\infty} \frac{\mathrm{cov}[n_h(x), n_h(y)]}{\sigma(n_h(x))\sigma(n_h(y))} = 1,$$
$$\lim_{t\to\infty} \rho\left(n_e(x), n_h(y)\right) = \lim_{t\to\infty} \frac{\mathrm{cov}[n_e(x), n_h(y)]}{\sigma(n_e(x))\sigma(n_h(y))} = 1,$$

i.e. that the charge distributions inside the avalanche region become maximally correlated across space at late times. Eqs. 2.64 also show that the absolute numbers of electrons and holes become maximally Pearson-correlated at late times,

$$\lim_{t\to\infty} \rho(N_e, N_h) = \lim_{t\to\infty} \frac{\mathrm{cov}[N_e, N_h]}{\sigma(N_e)\sigma(N_h)} = 1.$$

The only fluctuating degree of freedom of the densities of electrons and holes at late times is thus their joint overall normalisation; the spatial distribution of the charges is completely determined.

Eqs. 2.64 moreover determine the asymptotic magnitude of avalanche fluctuations. The relative fluctuations defined in Section 2.3.3 become

$$\lim_{t\to\infty} \frac{\sigma(N_e)^2}{\langle N_e\rangle^2} = \lim_{t\to\infty} \frac{\sigma(N_h)^2}{\langle N_h\rangle^2} = \lim_{t\to\infty} \frac{\sigma(N)^2}{\langle N\rangle^2} = \lim_{t\to\infty} \frac{\sigma(I^{\mathrm{ind}}(t))^2}{\langle I^{\mathrm{ind}}(t)\rangle^2} = \frac{C_\infty(\lambda_1, \lambda_1)}{C(\lambda_1)^2}, \qquad (2.65)$$

where the induced current I^{ind} is taken from Eq. 2.55. The first three equalities demonstrate again the maximal correlation between the electron and hole content.

The calculations leading to these results assumed position-independent electric fields. However, the strategy for the solution of Eqs. 2.58 and in particular the structure of the decomposition in Eq. 2.60 continues to hold in the general case of non-uniform impact-ionisation coefficients and drift velocities. For late times, the position dependence of the covariances in Eqs. 2.59 continues to be given by products of the eigenfunctions of Eqs. 2.45 (which must now be found numerically). Maximal correlation occurs regardless of the shape of the field profile $\mathbf{E}(x)$. If only numerical solutions are available for the covariances, the relative fluctuations in Eq. 2.65 should be computed as

$$\lim_{t\to\infty} \frac{\sigma(N)^2}{\langle N\rangle^2} = \lim_{t\to\infty} \frac{\mathrm{cov}[n_e(x), n_e(y)]}{\langle n_e(x,t)\rangle^2} = \lim_{t\to\infty} \frac{\mathrm{cov}[n_h(x), n_h(y)]}{\langle n_h(x,t)\rangle^2}.$$

2.4.4 Asymptotic behaviour and time resolution

The above results demonstrate that the asymptotic evolution of avalanches in thin semiconductors is very similar to that in the unbounded geometry investigated in Section 2.3. At late times, all spatial degrees of freedom freeze out and only the overall charge content of the avalanche remains dynamic. One important conceptual difference between the two situations remains, however. An avalanche developing in a bounded semiconductor is not guaranteed to propagate and become self-sustaining, in contrast to the situation where there are no boundaries.

The probability distribution $p(N,t)$ of the total charge content therefore has the structure

$$p(N,t) = [1-\epsilon(t)]\;\delta_{N,0} + \epsilon(t)\,p_S(N,t), \tag{2.66}$$

where $1-\epsilon(t)$ is the fraction of all avalanches that have already ended at time t, i.e. for which no free charge carriers remain in the avalanche region. The distribution $p_S(N,t)$ describes those avalanches that have not (yet) terminated, referred to as the "starters". Once an avalanche has ended, it can never grow again, i.e. $\epsilon(t)$ is monotonically falling with t. Its limiting "efficiency" $\epsilon = \lim_{t\to\infty}\epsilon(t)$ depends on the applied electric field and on the initial charge deposits $n_e^0(x)$ and $n_h^0(x)$. It is easily computed with the method developed in Refs. [70, 71]. These calculations are reviewed and applied in the context of Chapter 3. For now, the efficiency ϵ is simply taken as an external input parameter.

The form of $p_S(N,t)$ is generally unknown. For most practical purposes, it can be approximated based on the results obtained in Sections 2.4.2 and 2.4.3, as explained in the following.

Conditional and unconditional averages All expectation values discussed so far are defined as in Eq. 2.3. They are *unconditional* averages in the sense that all possible avalanche configurations are summed over, using the full distribution $p(N,t)$ in Eq. 2.66. Many expectation values of this kind can easily be converted into *conditional* averages which only involve diverging avalanches, and therefore sum over the distribution of starters, $p_S(N,t)$. To clearly distinguish between conditional and unconditional expectation values, the former are denoted as $\langle\cdot\rangle_S$.

For any observable $\mathcal{O}$ that is homogeneous in N, i.e. for which $N=0$ implies $\mathcal{O}(N)=0$, unconditional and conditional averages are related through

$$\langle\mathcal{O}\rangle_S = \sum_N \mathcal{O}(N)\,p_S(N,t) = \frac{1}{\epsilon(t)}\sum_N \mathcal{O}(N)\,p(N,t) = \frac{1}{\epsilon(t)}\langle\mathcal{O}\rangle\,.$$

With Eq. 2.57, the average charge content of the starters is simply

$$\langle N\rangle_S \approx \frac{N(\lambda_1)}{\epsilon}e^{\gamma_1 v^* t} \tag{2.67}$$

at late times. Similarly, the asymptotic relative fluctuations, restricted to the starters, become

$$\lim_{t\to\infty}\frac{\sigma(N)_S^2}{\langle N\rangle_S^2}=\lim_{t\to\infty}\epsilon\left(1+\frac{\sigma(N)^2}{\langle N\rangle^2}\right)-1=\epsilon\left(1+\frac{C_\infty(\lambda_1,\lambda_1)}{C(\lambda_1)^2}\right)-1, \tag{2.68}$$

where the last equality makes use of the explicit solution for constant electric fields.

Approximating the distribution of starters For avalanches in unbounded semiconductors, the asymptotic distribution $p(N,t)$ in Eq. 2.23 is fully determined by the two parameters $\langle N\rangle$ and A, which control its first and second moment, respectively. The interactions of the avalanche with the boundaries of a finite domain generally cause modifications of all moments of this distribution. This is explicitly reflected in Eqs. 2.67 and 2.68 for the first two moments.

These results allow $p_S(N,t)$ to be determined under the approximation that there are no additional, independent boundary effects on the higher moments, i.e. the relationships between these moments remain identical to those implied in Eq. 2.23. This assumption fixes $p_S(N,t)$ to correspond to a gamma distribution whose first two moments correctly reproduce Eqs. 2.67 and 2.68. An even coarser approximation, shown in the following to be useful for certain applications, is obtained by considering modifications to the first moment only.

In analogy to Eq. 2.31, it is convenient to take the result in Eq. 2.68 as the *definition* of a modified, "effective" avalanche parameter A_{eff},

$$\frac{1}{A_{\text{eff}}}:=\lim_{t\to\infty}\frac{\sigma(N)_S^2}{\langle N\rangle_S^2}. \tag{2.69}$$

As $t\to\infty$, $p_S(N,t)$ may then be expressed in terms of A_{eff} and $\langle N\rangle_S$ as

$$\begin{aligned} p_S(N,t) &\approx \frac{1}{\Gamma(A_{\text{eff}})}\left(\frac{A_{\text{eff}}}{\langle N\rangle_S}\right)^{A_{\text{eff}}} N^{A_{\text{eff}}-1}\exp\left(-A_{\text{eff}}\frac{N}{\langle N\rangle_S}\right)\\ &= \frac{1}{\Gamma(A)}\left(\frac{h}{e^{\gamma_1 v^* t}}\right)^{A_{\text{eff}}} N^{A_{\text{eff}}-1}\exp\left(-\frac{Nh}{e^{\gamma_1 v^* t}}\right), \end{aligned} \tag{2.70}$$

where $h=\epsilon A_{\text{eff}}/N(\lambda_1)$ and the second equality holds for the case of constant electric fields.

Time-response function and time resolution Only the starters cross an applied threshold and are thus relevant for the computation of the time resolution. In exact analogy to the discussion in Section 2.3.5, Eq. 2.70 implies the following asymptotic time-response function for large thresholds N_{th},

$$\rho(N_{\text{th}},t)\approx\frac{\gamma_1 v^*}{\Gamma(A_{\text{eff}})}\exp\left(A_{\text{eff}}\log\frac{hN_{\text{th}}}{e^{\gamma_1 v^* t}}-\frac{hN_{\text{th}}}{e^{\gamma_1 v^* t}}\right). \tag{2.71}$$

Correspondingly, the time resolution for large thresholds is

$$\sigma_{\text{th},\infty} = \frac{\sqrt{\psi_1(A_{\text{eff}})}}{\gamma_1 v^*}. \tag{2.72}$$

This shows that the overall scale of the time resolution continues to be determined by the asymptotic growth rate of the avalanche, $\gamma_1 v^*$, also in the presence of boundaries. It generally accounts for the predominant part of the material- and field-dependence of the time resolution.

The behaviour of the dimensionless coefficient $\sqrt{\psi_1(A_{\text{eff}})}$ in the numerator is studied in Figure 2.14. It summarises results obtained with the MAMC simulation model that show the dependence of this quantity on the impact-ionisation coefficient β and the thickness d of the avalanche region. Its dependence on the material properties is typically rather weak. The approximate result in Eq. 2.72 is accurate to within about 10% across a wide range of the ratio α/β. For avalanches that are initiated by an electron, it becomes exact in the limit $\alpha \gg \beta$. This case is relevant for detectors based on silicon, as discussed further in Chapter 3. The effective avalanche parameter A_{eff} depends only weakly on the position x_0 of the initial charge. Variations in $\sqrt{\psi_1(A_{\text{eff}})}$ are limited to about 30% as x_0 traverses the medium.

Neglecting boundary effects on the avalanche fluctuations, i.e. approximating $A_{\text{eff}} = A$, generally reproduces the correct scaling of $\sigma_{\text{th},\infty}\gamma_1 v^*$ with β. By construction, this approximation does not capture the residual dependence on the thickness d and the initial position x_0. Depending on the values of the impact ionisation coefficients and the thickness d, the presence of the boundary may increase or decrease the magnitude of avalanche fluctuations compared to an unbounded geometry. This effect has its origins in the correlation structure of the avalanche at early times.

2.4.5 Summary

The most important properties of electron-hole avalanches developing in a bounded domain are:

(i) The average charge content of the avalanche grows exponentially at a rate of $\gamma_1 v^*$ (cf. Eq. 2.57), determined by the impact ionisation coefficients and the thickness of the avalanche region.

(ii) Avalanche fluctuations saturate on a timescale $(\gamma_1 v^*)^{-1}$ (cf. Eq. 2.61).

(iii) At late times, fluctuations of the charge-carrier densities at different spatial locations become maximally correlated. Only the absolute normalisation continues to fluctuate. Each individual avalanche becomes proportional to the average.

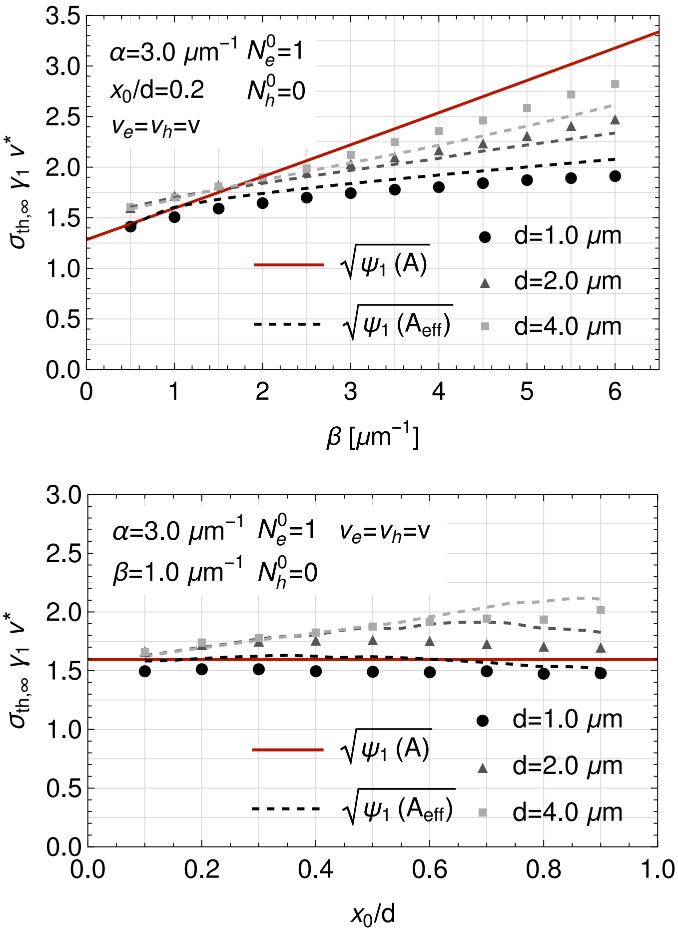
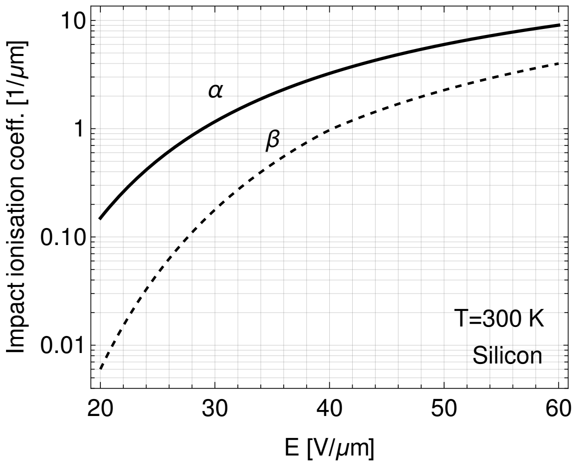


Figure 2.14: The black and grey markers show the quantity $\sigma_{\text{th},\infty}\gamma_1 v^*$ obtained from the MAMC simulation model, the dashed lines correspond to the approximation from Eq. 2.72, and the solid red line sets $A_{\text{eff}} = A$ in addition. The top panel shows the dependence on the impact ionisation coefficient β and the thickness d of the semiconductor for an avalanche initiated by a single electron. The bottom panel shows variations of the position x_0/d of the initial charge carrier for fixed impact ionisation coefficients.

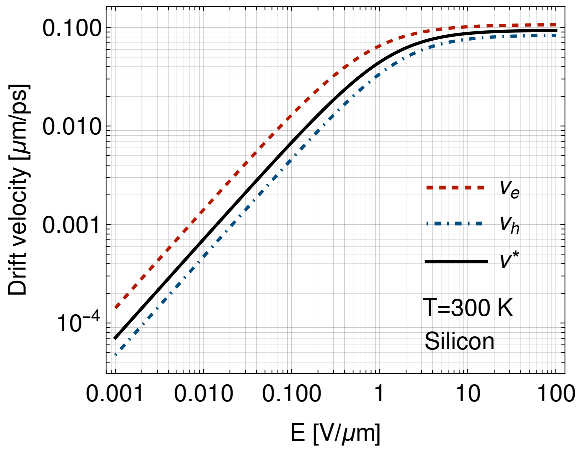
(iv) The time resolution for large thresholds is approximately given by Eq. 2.72. Its overall scale is determined by the inverse of the asymptotic growth rate. The time resolution depends weakly on the position of the initial charge carrier. There is a nontrivial dependence on the impact-ionisation coefficients and the thickness of the avalanche region.

2.5 Avalanches in silicon: transport parameters, breakdown, and time scales

The results obtained above are easily applied to the important case of silicon-based devices by inserting the appropriate parameterisations for the impact ionisation coefficients and the drift velocities.



(a)



(b)

Figure 2.15: Transport parameters for silicon. (a) The local impact ionisation coefficients α and β as a function of the electric field in the material. Data from Ref. [75]. (b) The drift velocities v_e and v_h as well as the effective drift velocity v^*. Data from Ref. [76].

These material parameters are briefly reviewed in the following.

Impact ionisation coefficients Figure 2.15a shows the local impact ionisation coefficients α and β. They depend exponentially on the applied electric field, following the parameterisation

$$\alpha(E) = \alpha_\infty e^{-\frac{a}{E}}, \qquad \beta(E) = \beta_\infty e^{-\frac{b}{E}},$$

where the constants $\alpha_\infty, \beta_\infty, a$, and b are listed in Ref. [75]. For a field of $E_0 = 40\,\mathrm{V/\mu m}$, the previous expressions become $\alpha(E_0) \approx 3.24$ and $\beta(E_0) \approx 0.97$. The impact-ionisation coefficients have negative temperature coefficients, i.e. impact ionisation becomes less likely at high temperatures. This effect is caused by increased energy losses of the drifting charge carriers into the sea of phonons, caused by elevated scattering rates at high temperatures [77].

Breakdown Figure 2.16a evaluates the breakdown condition from Eq. 2.47 for the impact ionisation coefficients in Figure 2.15a. Breakdown is achieved by applying a voltage of around 35 V across an avalanche region with a thickness of $d = 1\,\mathrm{\mu m}$. Correspondingly lower field strengths are required for thicker avalanche regions.

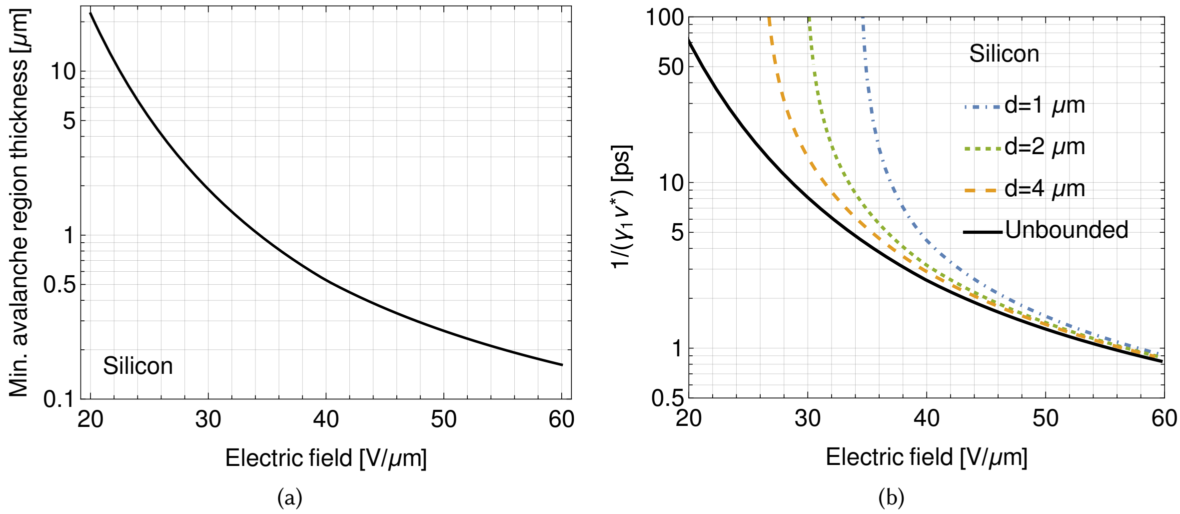


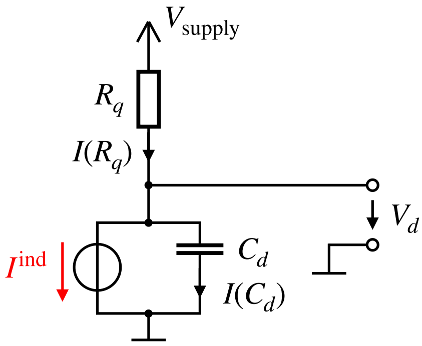
Figure 2.16: (a) Minimum required thickness d of the avalanche region to achieve breakdown, shown as a function of the applied electric field. (b) The characteristic time scale $(\gamma_1 v^*)^{-1}$ as a function of the electric field in the avalanche region and its thickness d. Shown in black is the corresponding result for an unbounded geometry.

Drift velocities Figure 2.15b shows the drift velocities v_e and v_h obtained in Ref. [76] as a function of the electric field. The experimental data are described by the functional form

$$v_e(E) = \frac{\mu_e E}{\left[1 + \left(\frac{\mu_e E}{v_e^{\text{sat}}}\right)^{\beta_e}\right]^{\frac{1}{\beta_e}}}, \qquad v_h(E) = \frac{\mu_h E}{\left[1 + \left(\frac{\mu_h E}{v_h^{\text{sat}}}\right)^{\beta_h}\right]^{\frac{1}{\beta_h}}},$$

with the low-field carrier mobilities μ_e and μ_h, the saturation velocities v_e^{sat} and v_h^{sat} and the empirical fit parameters β_e and β_h. For velocities much less than the saturation velocities, the drift velocity is directly related to the applied field through the carrier mobilities. At high electric fields, increased interactions with the crystal lattice and enhanced phonon scattering rates lead to a saturation of the drift velocities. Also shown in Figure 2.15b is the averaged drift velocity v^* computed according to Eq. 2.50b. In the saturation regime, $v_e \approx v_h \approx v^* \approx 0.1\,\mu\text{m/ps}$ to within about 20%.

Time resolution The scale $(\gamma v^*)^{-1}$ of the time resolution as obtained from Eqs. 2.50a and 2.51 is shown in Figure 2.16b. At extremely high field strengths, the dependence on the thickness d becomes weak and $(\gamma v^*)^{-1}$ approaches the value attained in an unbounded geometry. In this limit, the contribution to the time resolution originating from avalanche fluctuations is significantly less than 10 ps. Close to the breakdown limit, the achievable time resolution diverges and differences between different detector geometries become noticeable.



(a)

Circuit:

$$\frac{dV_d}{dt} = \frac{I(C_d)}{C_d}$$
$$I(R_q) = I^{\text{ind}} + I(C_d)$$
$$V_{\text{supply}} = V_d + I(R_q)R_q$$

(b)

Avalanche:

$$\mathbf{E}(x,t) = \mathbf{E}_{\text{ext}}(x, V_d(t)) + \mathbf{E}_{\text{sc}}(x,t)$$
$$\alpha(x,t) = \alpha(E(x,t))$$
$$\beta(x,t) = \beta(E(x,t))$$

(c)

Figure 2.17: (a) Equivalent circuit for the computation of the signal, consisting of the diode capacitance C_d and the quench resistor R_q. The current source implements the current I^{ind} induced by the avalanche on the grounded detector electrodes. (b) System of algebraic-differential equations describing the circuit. (c) General relations between the external electric field $\mathbf{E}_{\text{ext}}$, the field $\mathbf{E}_{\text{sc}}$ due to space charge, and the impact ionisation coefficients.

2.6 Quenching dynamics and large avalanches

The passive circuit shown in Figure 2.3a represents the simplest practical quenching method. Figure 2.17 defines the model used to describe the quenching process. Following Section 1.3.3, the p–n junction is represented by its capacitance C_d. The voltage V_d measured between the capacitor plates is the source of the external electric field $\mathbf{E}_{\text{ext}}(x, V_d)$. The current I^{ind} induced by the avalanche is given by Eq. 2.33.

The device is prepared by biasing the junction above its breakdown voltage, $V_d = V_{\text{supply}} > V_{\text{br}}$, and $I^{\text{ind}} = 0$ initially. A primary charge carrier injected into the high-field region triggers the formation of the avalanche which discharges the diode capacitance and lowers V_d and $\mathbf{E}_{\text{ext}}$. For sufficiently large avalanches, the space-charge field $\mathbf{E}_{\text{sc}}(x)$ created by the carriers themselves is no longer negligible and may contribute significantly to the total electric field $\mathbf{E} = \mathbf{E}_{\text{ext}} + \mathbf{E}_{\text{sc}}$ in the multiplication region. This field distribution in turn determines the impact-ionisation coefficients according to Figure 2.17c, tightly coupling the evolution of the avalanche with the dynamics of the equivalent circuit. If this process reduces the field sufficiently far below the breakdown limit, the avalanche collapses and $I^{\text{ind}} = 0$ as soon as all charge carriers have drifted out of the avalanche region. The quench resistor then recharges the diode capacitance on timescales of $\tau_{\text{rec}} = R_q C_d$ and prepares the detector for the next event. The dead time for passively quenched devices is typically a few tens of nanoseconds.

Section 2.6.1 demonstrates that the salient features of this quenching process may be described analytically as long as space-charge effects are neglected. Space charge is considered explicitly in Section 2.6.2, where a simplified simulation shows that it does not fundamentally alter the dynamics of the quenching but only leads to modifications of the signal charge and its temporal evolution. (In recent years, more sophisticated quenching strategies have been developed [78]. These techniques use active circuitry to monitor and stop the development of the avalanche at a suitable moment [79], and to reduce τ_{rec} and the dead time as much as possible. They are not considered here.)

2.6.1 Adiabatic model for passive quenching without space charge

To ensure quenching, R_q must be chosen such that τ_{rec} is much longer than the timescale on which the avalanche develops [80]. The initial phase of the quenching process, whose characteristics determine the fast component of the detector signal, thus remains unchanged as $R_q \to \infty$, and this limit is taken in what follows. For simplicity, the discussion here is limited to position-independent external electric fields $\mathbf{E}_{\text{ext}}(t) = -V_d(t)/d\,\hat{\mathbf{x}}$, for which Section 2.4 contains a complete analytical description of the avalanche. The treatment extends naturally to the case of general field profiles $\mathbf{E}_{\text{ext}}(x, V_d(t))$ by replacing all quantities with their equivalents extracted from the numerical solution of the corresponding evolution equations.

Section 2.4.4 establishes that the total numbers of electrons N_e and holes N_h in the multiplication region become maximally correlated for sufficiently large avalanches developing in a time-independent electric field. The evolution of each avalanche is identical to its average evolution, $N(t) \equiv \langle N(t) \rangle_S$, up to fluctuations in the time argument with standard deviation $\sigma_{\text{th},\infty}$. The time resolution is factored out of the present discussion, which is concerned only with the deterministic component of the avalanche growth and quenching. The spatial distributions of the charge carriers are described by the eigenfunctions $f^e_{\lambda_1}$ and $f^h_{\lambda_1}$ defined in Eqs. 2.49, which grow exponentially with a rate of $\gamma_1 v^*$.

These results do not immediately generalise to the dynamical case considered here, where the discharge of C_d leads to a time-dependent electric field $\mathbf{E} = \mathbf{E}_{\text{ext}}$. Under the assumptions that $\gamma_1 v^*$ adiabatically follows the evolution of $\mathbf{E}$ (or, equivalently, V_d) and that the subleading time constants $\gamma_n v^*$, $n > 1$ remain negligible at all times, the evolution of the system follows the equations

$$\frac{dI^{\text{ind}}}{dt} = \gamma_1(V_d) v^*(V_d) I^{\text{ind}}, \tag{2.73a}$$

$$\frac{dV_d}{dt} = -\frac{I^{\text{ind}}}{C_d}. \tag{2.73b}$$

This is similar to the rate model constructed in Ref. [81] but in contrast to the calculations performed there automatically includes boundary effects. The quantities $\gamma_1(V_d)$ and $v^*(V_d)$ are determined by the material parameters.

The drift velocities v_e and v_h typically saturate in the avalanche region, i.e. the velocity v^* becomes independent of V_d. At the breakdown voltage, $\gamma_1(V_{\text{br}}) = 0$. Provided that V_d remains close to the breakdown voltage during the entire evolution of the system, the function $\gamma_1(V_d)$ may be replaced by its linear expansion around V_{br},

$$\gamma_1(V_d) \approx \left.\frac{d\gamma_1}{dV_d}\right|_{V_{\text{br}}} (V_d - V_{\text{br}}) =: \gamma'_{1,\text{br}}(V_d - V_{\text{br}}). \tag{2.74}$$

Using Eqs. 2.50a, 2.50d, and 2.51, the parameter $\gamma'_{1,\text{br}}$ may be expressed as

$$\gamma'_{1,\text{br}} = \frac{1}{d}\left.\frac{d\gamma_1}{dE}\right|_{E_{\text{br}}} = \frac{\alpha'_{\text{br}} + \beta'_{\text{br}}}{2d} + \frac{(\sinh 2\bar{\kappa}_{\text{br}} - 2\bar{\kappa}_{\text{br}})\left(\alpha'_{\text{br}}\beta_{\text{br}} + \alpha_{\text{br}}\beta'_{\text{br}}\right)}{2\bar{\kappa}_{\text{br}}\left(\cosh 2\bar{\kappa}_{\text{br}} - 1\right) - d\left(\sinh 2\bar{\kappa}_{\text{br}} - 2\bar{\kappa}_{\text{br}}\right)\left(\alpha_{\text{br}} + \beta_{\text{br}}\right)},$$

where $\alpha_{\text{br}} = \alpha(E_{\text{br}})$, $\beta_{\text{br}} = \beta(E_{\text{br}})$, $\alpha'_{\text{br}} = d\alpha/dE|_{E_{\text{br}}}$, $\beta'_{\text{br}} = d\beta/dE|_{E_{\text{br}}}$, and $\bar{\kappa}_{\text{br}} = \bar{\kappa}|_{E_{\text{br}}} = d|\alpha_{\text{br}} - \beta_{\text{br}}|/2$. With these additional simplifications, Eqs. 2.73 admit the following solutions,

$$I^{\text{ind}}(t) = \frac{1}{2}\gamma'_{1,\text{br}}v^* C_d V_{\text{ex}}^2 \left[1 - \tanh^2\left(\frac{1}{2}\gamma'_{1,\text{br}}v^* V_{\text{ex}} t\right)\right], \tag{2.75a}$$

$$V_d(t) = V_{\text{supply}} - V_{\text{ex}}\left[1 + \tanh\left(\frac{1}{2}\gamma'_{1,\text{br}}v^* V_{\text{ex}} t\right)\right], \tag{2.75b}$$

with the excess voltage $V_{\text{ex}} = V_{\text{supply}} - V_{\text{br}}$. The time coordinate is conveniently chosen so that $V_d(t=0) = V_{\text{br}}$. In this model, the evolution of the current I^{ind} is symmetric around $t = 0$, where the avalanche attains its maximum size. The FWHM of the signal is

$$\text{FWHM}\left(I^{\text{ind}}\right) = \frac{4\,\text{artanh}\,\frac{1}{\sqrt{2}}}{\gamma'_{1,\text{br}}v^* V_{\text{ex}}} \approx \frac{3.53}{\gamma'_{1,\text{br}}v^* V_{\text{ex}}}.$$

The observable voltage step $\Delta V_d = \lim_{t\to-\infty} V_d(t) - \lim_{t\to\infty} V_d(t)$ is $\Delta V_d = 2V_{\text{ex}}$, and the total released charge is $Q = \int_{-\infty}^{\infty} dt I^{\text{ind}}(t) = 2C_d V_{\text{ex}}$. The signal charge is thus directly proportional to the excess voltage at which the junction is operated. It does not depend on the primary charge deposit which initiated the avalanche. In the adiabatic approximation, predictions for the charge carrier densities $n_e(x,t) \equiv \langle n_e(x,t)\rangle_S$ and $n_h(x,t) \equiv \langle n_h(x,t)\rangle_S$ at late times are given by the eigenfunctions corresponding to $\gamma_1(V_d)$, i.e by the functions $f^e_{\lambda_1(V_d)}$ and $f^h_{\lambda_1(V_d)}$.

Figure 2.18 compares the analytic solution in Eq. 2.75 to the trajectory obtained from a combined simulation of the avalanche and the quenching circuit, using the material parameters for silicon from Section 2.5. The avalanche is simulated by the MAMC model (cf. Appendix A.3) and the evolution of

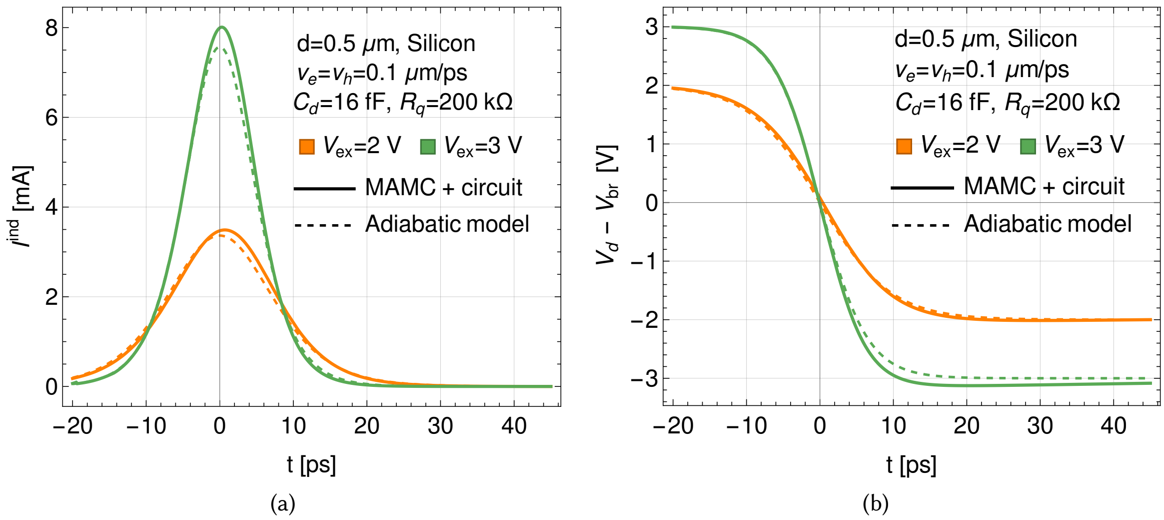


Figure 2.18: Comparison between the MAMC + circuit simulation model (thick lines) and the adiabatic approximation in Eqs. 2.75 (dashed lines), for excess voltages of $V_{\text{ex}} = 2\,\text{V}$ (orange) and $V_{\text{ex}} = 3\,\text{V}$ (green). The induced current I^{ind} is shown in (a) and the voltage $V_d - V_{\text{br}}$ in (b).

the electric field in the avalanche region is determined from the relations in Figures 2.17b and 2.17c. The circuit equations are integrated with the forward Euler method [82]. The resulting simulation model is referred to as "MAMC + circuit". It is similar to the simulation described in Refs. [83, 84]. The chosen thickness of the avalanche region is $d = 0.5\,\mu\text{m}$ and its diameter is $D = 10\,\mu\text{m}$. Assuming a parallel field geometry and using a relative permittivity of $\epsilon_r = 11.7$ [85], the junction capacitance is $C_d \approx 16\,\text{fF}$. A quench resistor with $R_q = 200\,\text{k}\Omega$ and saturated drift velocities $v_e = v_h = 0.1\,\mu\text{m/ps}$ are used. Eq. 2.47 determines the breakdown voltage for this geometry to be $V_{\text{br}} = 20.34\,\text{V}$ and the model parameter $\gamma'_{1,\text{br}}$ becomes $\gamma'_{1,\text{br}} = 1.05\ (\text{V}\,\mu\text{m})^{-1}$.

The adiabatic model correctly reproduces the simulated shape of the signal. Small deviations arise as V_d crosses the breakdown voltage, resulting in a slight underestimation of ΔV for large excess voltages. This is caused by additional dynamic effects originating from changes in the shapes of the spatial carrier densities of electrons $n_e(x,t)$ and holes $n_h(x,t)$. The evolution of these densities is compared in Figure 2.19. (For the simulation, the average densities $\langle n_e \rangle$ and $\langle n_h \rangle$ are shown to ensure a meaningful comparison with the adiabatic model even at early times where the number of charge carriers is small.) Charge carriers of both polarities are present throughout the entire avalanche region. The highest densities are attained at the corresponding downstream ends, i.e. at $x = d$ for electrons and at $x = 0$ for holes. The shape of the distributions, and in particular the curvature of $n_h(x)$, depend on V_d.

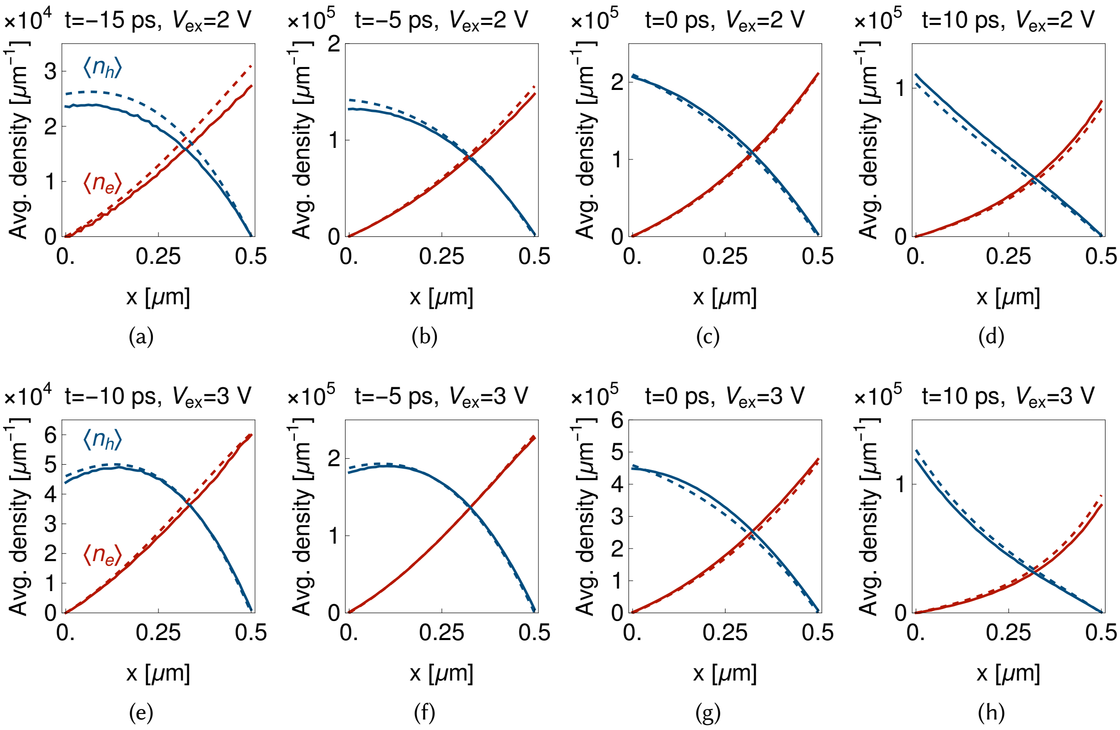


Figure 2.19: Evolution of the charge carrier densities $n_e(x,t)$ and $n_h(x,t)$ with time, for $V_{\text{ex}} = 2\,\text{V}$ in (a)–(d) and $V_{\text{ex}} = 3\,\text{V}$ in (e)–(h). The thick lines show the average densities obtained from the MAMC + circuit simulation model, and the dashed lines correspond to the predictions from the adiabatic model. All parameters are identical to those used in Figure 2.18.

2.6.2 The role of space charge

A full description of the effects of space charge (SC) requires a three-dimensional simulation of the avalanche together with the computation of a self-consistent solution for the space-charge field $\mathbf{E}_{\text{sc}}(\mathbf{x},t)$ at every time step. In particular, $\mathbf{E}_{\text{sc}}$ depends on the spread of the avalanche in the direction transverse to the external field $\mathbf{E}_{\text{ext}}$.

The situation is treated here in a simplified way as illustrated in Figure 2.20. The evolution of the avalanche continues to be determined in a one-dimensional manner by the MAMC + circuit simulation model, i.e. all charge carriers are assumed to be located on the x-axis. A parameterised model is used to describe the development of the avalanche in the transverse yz-plane, defining a radially symmetric charge profile $\rho(r,x,t)$. Different physical effects may be included in this parameterisation, e.g. charge-carrier diffusion as well as drift due to nonvanishing radial electric-field components. The prescribed charge distribution ρ is then used to compute the field $\mathbf{E}_{\text{sc}}(r=0,x,t)$ along the x-axis, where the charge carriers are located. By symmetry, only the x-component of the space-charge field is nonvanishing, i.e. $\mathbf{E}_{\text{sc}}(r=0,x,t) = E^x_{\text{sc}}(x,t)\,\hat{\mathbf{x}}$. Together with the exter-

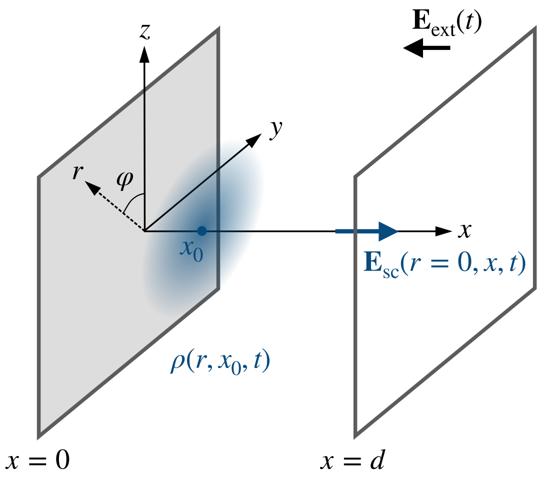


Figure 2.20: For the computation of the space charge field, the avalanche region is delimited by two conducting plates of infinite transverse extent, located at $x = 0$ and $x = d$. Charge carriers are constrained to move along the x-axis. This charge distribution is extended into a three-dimensional charge density $\rho(r, x, t)$, illustrated here in the plane $x = x_0$. The field $\mathbf{E}_{\text{sc}}(r, x, t)$ is evaluated on the x-axis as the electric field generated by $\rho(r, x, t)$. The external field $\mathbf{E}_{\text{ext}}(t)$ also contributes to the total electric field $\mathbf{E}(x, t)$ relevant for the avalanche development.

nal field $\mathbf{E}_{\text{ext}} = -V_d/d\,\hat{\mathbf{x}}$, this defines the electric-field distribution used for the simulation of the avalanche. The resulting simulation model is referred to as "MAMC + circuit + SC". It is analogous to the "1.5-dimensional" model of space-charge effects in resistive plate chambers discussed in Ref. [86].

For the results presented here, only transverse diffusion is considered for the construction of $\rho(r, x, t)$. This simplified model is likely to overestimate the true effect of space charge on the avalanche development. First, the space-charge field $\mathbf{E}_{\text{sc}}$ itself is overestimated: the field originating from a radially symmetric charge density $\rho(r, x, t)$ attains a maximum at $r = 0$. All charge carriers are thus subject to the maximal field, while regions of lower field strengths at $r > 0$ are not probed by the avalanche. Second, the transverse size of the avalanche is underestimated by only considering diffusion but ignoring carrier drift. This leads to a narrower distribution of charges and also enhances the computed space-charge field.

Transverse spread of the avalanche Diffusion is modelled by prescribing a normal distribution for the transverse distribution of charge carriers, while the longitudinal distribution continues to be given by the densities $n_e(x)$ and $n_h(x)$. The densities ρ_e and ρ_h used to compute $\mathbf{E}_{\text{sc}}$ are thus

$$\rho_e(r, x, t) = -\frac{e_0 n_e(x, t)}{2\pi\sigma_e^2(t)} \exp\left[-\frac{r^2}{2\sigma_e^2(t)}\right], \tag{2.76a}$$

$$\rho_h(r, x, t) = \frac{e_0 n_h(x,t)}{2\pi\sigma_h^2(t)} \exp\left[-\frac{r^2}{2\sigma_h^2(t)}\right]. \tag{2.76b}$$

The transverse extent is governed by the transverse-diffusion coefficients D_T for electrons and holes, i.e. $\sigma_e(t) = \sqrt{2D_{T,e}t}$ and $\sigma_h(t) = \sqrt{2D_{T,h}t}$. Values of $D_{T,e} = 15\,\mathrm{cm}^2/\mathrm{s}$ [87] and $D_{T,h} = 11\,\mathrm{cm}^2/\mathrm{s}$ [88] are used in the following. (To the best of our knowledge, no measurements of the transverse-diffusion coefficients are available in the literature for the very strong longitudinal fields $|\mathbf{E}_{\mathrm{ext}}| > 20\,\mathrm{V}/\mu\mathrm{m}$ that occur above breakdown. The selected values correspond to a field of $5\,\mathrm{V}/\mu\mathrm{m}$ in silicon. The transverse diffusion coefficients depend only relatively weakly on the longitudinal electric field [89], and it is expected that these values give at least an approximate description of diffusion in the relevant regime. This is sufficient for the qualitative assessment of space-charge effects attempted here.)

Computation of the electric field The Green's function for the electrostatic potential ϕ in the parallel-plate geometry of Figure 2.20 is [90, 91]

$$G(r, r', \varphi, \varphi', x, x') = \frac{1}{2\pi\epsilon} \sum_{m=-\infty}^{\infty} \int_0^\infty dk\, e^{im(\varphi-\varphi')} J_m(kr) J_m(kr') f(k, x, x'),$$

where the J_m are the Bessel functions of the first kind, and

$$f(k, x, x') = \frac{\sinh(kx)\sinh(k(d-x'))}{\sinh(kd)}\Theta(x'-x) + (x \leftrightarrow x').$$

A given charge density $\rho(r, x, t)$ generates an electric field $E^x = -\partial\phi/\partial x|_{r=0}$ along the x-axis,

$$E^x(r=0, x, t) = -\frac{\partial}{\partial x}\int_0^\infty dr' r' \int_0^d dx' \int_0^{2\pi} d\varphi' \rho(r', x', t) G(r=0, r', \varphi, \varphi', x, x').$$

Using the densities in Eqs. 2.76, the contributions to the space charge resulting from electrons and holes are

$$E^x_{\mathrm{sc},e}(x,t) = \frac{e_0}{2\pi\epsilon}\int_0^\infty dk \exp\left[-\frac{k^2\sigma_e^2(t)}{2}\right]\int_0^d dx' n_e(x', t)\frac{\partial}{\partial x} f(k, x, x'),$$
$$E^x_{\mathrm{sc},h}(x,t) = -\frac{e_0}{2\pi\epsilon}\int_0^\infty dk \exp\left[-\frac{k^2\sigma_h^2(t)}{2}\right]\int_0^d dx' n_h(x', t)\frac{\partial}{\partial x} f(k, x, x'),$$

and the complete space-charge field is given by their sum, $E^x_{\mathrm{sc}} = E^x_{\mathrm{sc},e} + E^x_{\mathrm{sc},h}$

Quenching with space charge Figure 2.21 shows the evolution of the carrier densities, the space-charge field, and the induced signal. At early times, space-charge effects are of subleading importance and the charge-carrier densities are given by $f^e_{\lambda_1}$ and $f^h_{\lambda_1}$ in good approximation. The sepa-

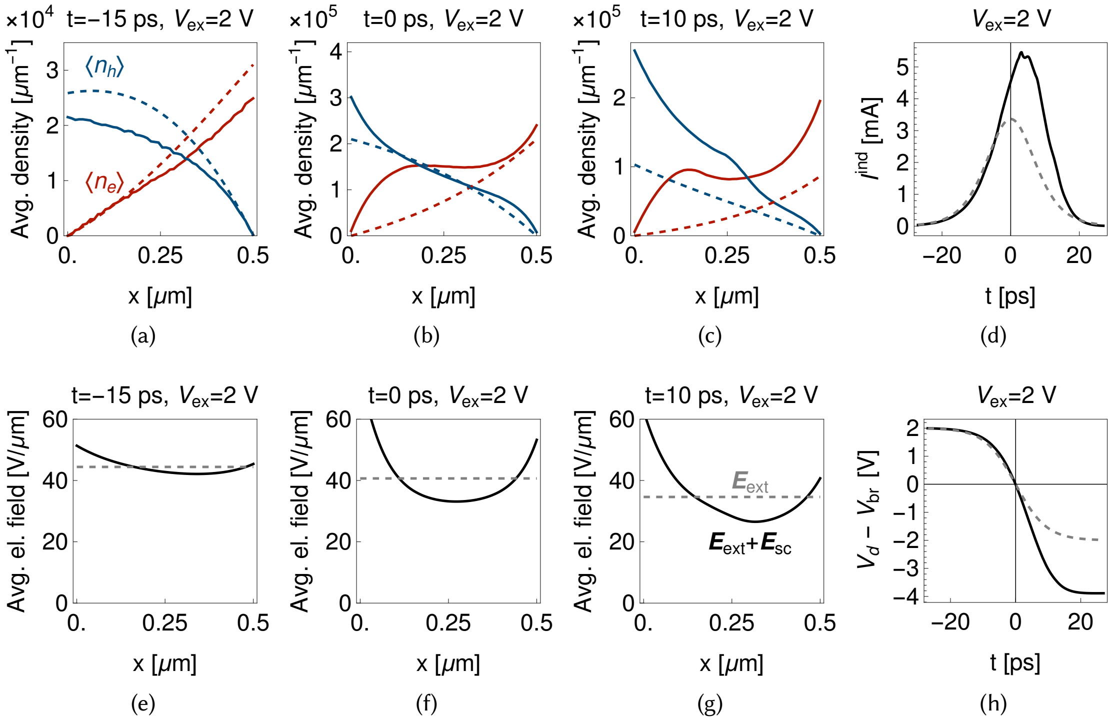


Figure 2.21: (a)–(c) Evolution of the charge carrier densities $n_e(x,t)$ and $n_h(x,t)$ with time, including space charge effects, for $V_{\text{ex}} = 2\,\text{V}$. The corresponding electric field configurations are shown in (e)–(g). The current I^{ind} is shown in (d) and the voltage $V_d - V_{\text{br}}$ in (h). In all plots, the thick lines correspond to averages obtained from the MAMC + circuit + SC simulation model, and the dashed lines show the predictions from the adiabatic model which does not include space charge effects.

ration of charge caused by $\mathbf{E}_{\text{ext}}$ leads to a space-charge field which reduces the electric field in the centre of the avalanche region and enhances it close to its boundaries. (Space charge cannot alter the voltage V_d measured across the avalanche region, i.e. $\int_0^d d\mathbf{s} \cdot \mathbf{E}_{\text{sc}} = 0$.) As a result of the modified field configuration, the carrier densities deviate from the adiabatic prediction at later times.

Impact ionisation continues to occur in regions where $\mathbf{E}_{\text{sc}}$ enhances the local electric field, even when V_d is already below V_{br}. For the chosen parameters, this has the net effect of delaying the collapse of the avalanche and moderately enhancing the total charge content of the signal. Space charge does not, however, affect the quenching mechanism itself, which continues to be given by the discharge of C_d by the avalanche current.

This is in stark contrast to the situation in resistive plate chambers, where space-charge effects lead to a similar distortion of the electric-field profile along the gas gap. In this case, only electrons are relevant for the development of the avalanche. Their density is concentrated in regions where the electric field, and thus the gas gain, is suppressed. This reduces the overall signal charge by many orders of magnitude [86] and identifies space-charge effects as the central quenching mechanism.

Chapter 3

Time resolution and efficiency of single-photon avalanche diodes

Single-photon avalanche diodes (SPADs), also referred to as "Geiger-mode avalanche diodes", are semiconductor detectors that combine single-photon sensitivity with precision timing capabilities. Initially developed in the 1960s [92, 93], SPADs are now commercially available and widely used in many applications, including LIDAR [94], biophotonics [95], and quantum photonics [96]. Modern devices achieve single-photon time resolutions of around 10 ps FWHM and dead times of only a few nanoseconds [97].

Figure 3.1a illustrates the main functional components of a typical SPAD. A strongly doped reverse-biased p–n junction generates a thin gain layer with a thickness d, in which the electric field exceeds the breakdown limit and diverging electron-hole avalanches may form, as discussed in Chapter 2. The avalanche is initiated by the charges deposited by the incoming photon and provides the gain mechanism that creates the observable signal. Short-wavelength photons are preferentially absorbed close to the surface of the semiconductor and may thus deposit an electron-hole pair directly in the gain layer. Infrared photons have significantly longer absorption lengths. To achieve good absorption efficiencies also for longer wavelengths, a separate conversion layer with a thickness w in the range 10–100 µm is often used. The conversion layer exhibits a weak electric field used to drift one of the deposited charge carriers, in most cases the electron, to the gain region.

The device is augmented by external circuitry to quench the avalanche and define the signal. A simple example is shown in Figure 3.1b, consisting of a quench resistor R_q and a capacitor C_C to couple the signal $V_d \sim V^{\text{ind}}$ to the on-detector electronics. In some applications, the current

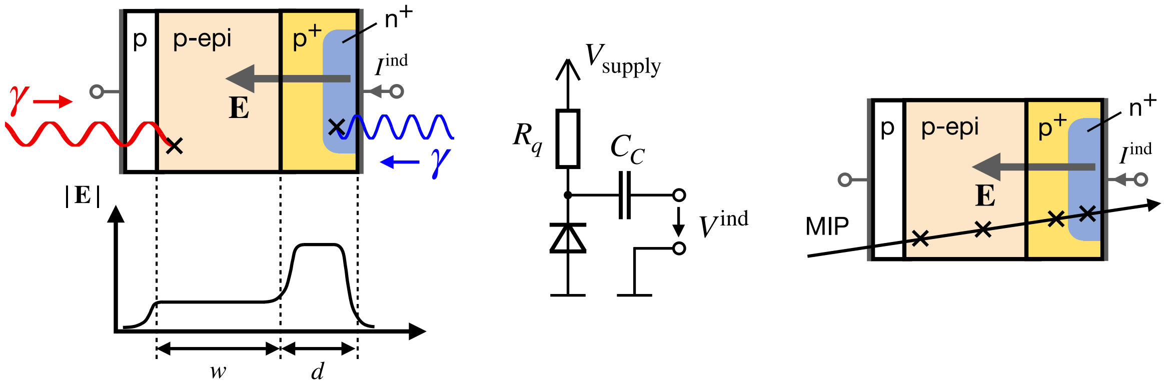


Figure 3.1: (a) Functional parts of an exemplary SPAD, consisting of a thin gain layer with thickness d and a conversion layer with thickness w. (b) Simple operating circuit for a SPAD with a quench resistor R_q and a capacitor C_C which couples the SPAD to the front-end electronics. (c) A minimum-ionising particle crosses a SPAD. A series of ionisation clusters are produced along its trajectory and initiate the avalanche.

through R_q is monitored instead.

According to the discussion in Section 2.6, a single SPAD is a binary detector for which the signal contains no information about the deposited primary charge or the number of incident photons. The capability to manufacture SPADs in CMOS technology [56] has sparked the development of large-scale SPAD arrays containing as many as 10^6 individual pixels, or "microcells" [98]. By monitoring the state of all microcells, these "silicon photomultipliers" (SiPMs) have photon-number resolving, and sometimes imaging, capabilities [99]. SiPMs optically coupled to scintillators can replace conventional photomultiplier tubes in many applications. Several experimental collaborations in particle physics [100, 101] are considering this option for the readout of crystal calorimeters.

Because of their capability to sense individual charge carriers, SPADs and SiPMs may serve as fast detectors for charged particles even without coupling to an external scintillator. As Figure 3.1c shows, a minimum ionising particle that crosses a SPAD produces a number of charge clusters along its track. As discussed below, the average distance between two ionisation clusters is significantly less than 1 µm for silicon. Very thin devices are thus expected to achieve good detection efficiencies for minimum ionising particles, and a dedicated conversion layer might not be necessary in all cases.

This chapter provides a theoretical study of the properties of silicon-based SPADs as detectors for photons and charged particles, focusing on the achievable efficiency and time resolution. The emphasis is put on an analytic description of the underlying physical phenomena rather than a detailed, simulation-based characterisation that would necessarily be device-specific. The efficiency which which a diverging avalanche is triggered in a SPAD is computed in Section 3.1. The avalanche theory developed in Chapter 2 forms the foundation for the computation of the achievable time resolution. It is extended in Section 3.2 to cover the more complicated initial charge deposits encountered in a SPAD. Section 3.3 covers SPADs as photon detectors and outlines the most important contributions to the single-photon time resolution for different wavelengths. Finally, Section 3.4 studies the properties of SPADs as charged-particle detectors.

There are many aspects that are relevant for the practical operation of SPADs and SiPMs which are not discussed here and go beyond the reach of the calculational methods developed [102, 103]. These include the (optical) coupling between neighbouring SPADs in the pixel matrix of an SiPM ("cross-talk"), the phenomenon of repeated discharge of the same pixel ("afterpulsing") due to crystal defects and the trapping of charge carriers in the semiconductor lattice, as well as the spurious discharge of a SPAD cell in the absence of any external stimulus ("dark count"). Some of these topics are experimentally studied in Chapter 4.

3.1 Efficiency

The efficiency ϵ introduced in Eq. 2.66 represents the probability that a given primary charge distribution leads to a diverging avalanche. It may be computed by an explicit enumeration of all trajectories that result in breakdown, as discussed in Refs. [70, 71] and reviewed below.

In the simplest relevant scenario, the avalanche discharge is triggered by a single electron placed at position $x_0 - dx_0$, for which the corresponding efficiency is denoted as $\epsilon_e(x_0 - dx_0)$. As the electron traverses a small interval dx_0, a diverging avalanche may form in two mutually exclusive ways. If the charge does not multiply during dx_0 (which happens with probability $1 - \alpha dx_0$), the probability for breakdown continues to be given (by definition) by the function ϵ_e, but evaluated at the new position x_0. Conversely, if the electron does undergo impact ionisation during dx_0 (which happens with probability αdx_0), the probability for breakdown is given by the efficiency that two electrons and one hole at position x_0 trigger a diverging avalanche. This yields the recursion relation

$$\epsilon_e(x_0 - dx_0) = (1 - \alpha(x_0)dx_0)\,\epsilon_e(x_0) + \alpha(x_0)dx_0\left[1 - (1 - \epsilon_e(x_0))^2\,(1 - \epsilon_h(x_0))\right].$$

An analogous argument starting with a single hole placed at $x_0 + dx_0$ gives for the corresponding breakdown probability $\epsilon_h(x_0)$,

$$\epsilon_h(x_0 + dx_0) = (1 - \beta(x_0)dx_0)\,\epsilon_h(x_0) + \beta(x_0)dx_0\left[1 - (1 - \epsilon_e(x_0))\,(1 - \epsilon_h(x_0))^2\right].$$

These relations may also be written in the form of (nonlinear) differential equations,

$$\frac{d\epsilon_e(x_0)}{dx_0} = -\alpha(x_0)\left(1-\epsilon_e(x_0)\right)\left[\epsilon_e(x_0)+\epsilon_h(x_0)-\epsilon_e(x_0)\epsilon_h(x_0)\right], \tag{3.1a}$$

$$\frac{d\epsilon_h(x_0)}{dx_0} = \beta(x_0)\left(1-\epsilon_h(x_0)\right)\left[\epsilon_e(x_0)+\epsilon_h(x_0)-\epsilon_e(x_0)\epsilon_h(x_0)\right], \tag{3.1b}$$

which are readily solved numerically for given impact-ionisation coefficients $\alpha(x_0)$ and $\beta(x_0)$ and the boundary conditions $\epsilon_e(d) = \epsilon_h(0) = 0$.

A careful analytic investigation of Eqs. 3.1 provides important insights into the structure of the solution. The term in square brackets on the right-hand side represents the efficiency for an electron-hole pair at x_0 to trigger a diverging avalanche, $\epsilon_{eh}(x_0) = 1-\left(1-\epsilon_e(x_0)\right)\left(1-\epsilon_h(x_0)\right)$. Using Eqs. 3.1, the efficiency ϵ_{eh} is seen to satisfy the equation

$$\frac{d\epsilon_{eh}(x_0)}{dx_0} = -(\alpha(x_0)-\beta(x_0))(1-\epsilon_{eh}(x_0))\epsilon_{eh}(x_0),$$

subject to the boundary condition $\epsilon_{eh}(0) = \epsilon_e(0) = \epsilon_0$. In terms of the unknown constant ϵ_0, its solution is

$$\epsilon_{eh}(x_0) = \frac{\epsilon_0}{\epsilon_0 + (1-\epsilon_0)\exp\left[\int_0^{x_0} dx_0'\left(\alpha(x_0')-\beta(x_0')\right)\right]}. \tag{3.2}$$

Using Eq. 3.2, Eqs. 3.1 may be integrated and the solutions take the form

$$\epsilon_e(x_0) = 1-\exp\left[-\int_{x_0}^{d} dx_0'\alpha(x_0')\epsilon_{eh}(x_0')\right], \tag{3.3a}$$

$$\epsilon_h(x_0) = 1-\exp\left[\int_0^{x_0} dx_0'\beta(x_0')\epsilon_{eh}(x_0')\right]. \tag{3.3b}$$

With the boundary condition $\epsilon_e(0) = \epsilon_0$, Eq. 3.3a gives rise to the following condition from which the constant ϵ_0 may be determined,

$$\epsilon_0 = 1-\exp\left[-\int_0^{d} dx_0'\frac{\epsilon_0\alpha(x_0')}{\epsilon_0+(1-\epsilon_0)\exp\left[\int_0^{x_0'} dx_0''\left(\alpha(x_0'')-\beta(x_0'')\right)\right]}\right]. \tag{3.4}$$

For position-independent impact-ionisation coefficients, the expressions for the efficiencies simplify,

$$\epsilon_{eh}(x_0) = \frac{\epsilon_0}{\epsilon_0+(1-\epsilon_0)e^{(\alpha-\beta)x_0}}, \tag{3.5a}$$

$$\epsilon_e(x_0) = 1-e^{-\alpha(d-x_0)}\left[\frac{e^{(\alpha-\beta)d}(1-\epsilon_0)+\epsilon_0}{e^{(\alpha-\beta)x_0}(1-\epsilon_0)+\epsilon_0}\right]^{\frac{\alpha}{\alpha-\beta}}, \tag{3.5b}$$

$$\epsilon_h(x_0) = 1-e^{-\beta x_0}\left[(1-\epsilon_0)e^{(\alpha-\beta)x_0}+\epsilon_0\right]^{\frac{\beta}{\alpha-\beta}}, \tag{3.5c}$$

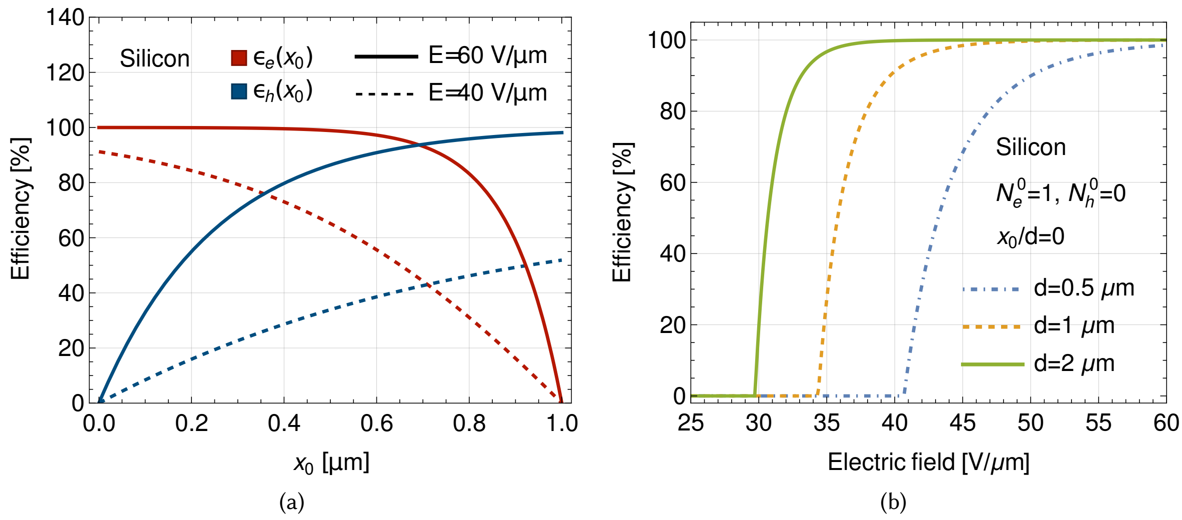


Figure 3.2: (a) Efficiency to obtain a diverging avalanche when a single electron (ϵ_e) or a single hole (ϵ_h) is placed at position x_0 in a gain layer with a thickness of $d = 1\,\mu$m, for two different (position-independent) electric fields. (b) Efficiency ϵ_e for the case of a single electron placed at the upstream end of the gain layer at $x_0/d = 0$, as a function of the electric field and the thickness d.

and Eq. 3.4 results in the following expression from which ϵ_0 may be determined,

$$e^{-(\alpha-\beta)d} = \frac{1}{\epsilon_0}\left[(1-\epsilon_0)^{1-\frac{\beta}{\alpha}} - (1-\epsilon_0)\right].$$

Figure 3.2a illustrates the dependence of the efficiencies $\epsilon_e(x_0)$ and $\epsilon_h(x_0)$ on the position of the initial charge carrier. Both functions are monotonic and the highest efficiencies are achieved at the corresponding upstream ends of the gain layer. In silicon, $\beta < \alpha$ (cf. Figure 2.15a), which makes holes less efficient than electrons for initiating avalanche breakdown. To maximise the quantum efficiency of the detector, the field configuration should be chosen such that the deposited electrons trigger the avalanche. This is relevant for the arrangement of the conversion layer relative to the gain layer.

Figure 3.2b shows the efficiency as a function of the electric field. A single electron is placed at $x_0 = 0$, e.g. coming from the conversion layer of the device in Figure 3.1a. Above the breakdown field, the efficiency rises steeply and quickly saturates at its maximum value.

The efficiency for more general initial condition may be obtained from the functions $\epsilon_e(x_0)$ and $\epsilon_h(x_0)$. For N_e^0 initial electrons at positions $x_{e,i}$ and N_h^0 initial holes at positions $x_{h,j}$, the efficiency becomes

$$\epsilon = 1 - \prod_{i=1}^{N_e^0}\left[1-\epsilon_e(x_{e,i})\right]\prod_{j=1}^{N_h^0}\left[1-\epsilon_h(x_{h,j})\right],$$

and for initial charge carrier densities $n_e^0(x)$ and $n_h^0(x)$, it is

$$\epsilon = 1 - \exp\left[\int_0^d dx_0\, \left(n_e^0(x_0)\log\left[1-\epsilon_e(x_0)\right] + n_h^0(x_0)\log\left[1-\epsilon_h(x_0)\right]\right)\right]. \tag{3.6}$$

In case the avalanche is initiated by a density of electron-hole pairs, i.e. $n_e^0(x) = n_h^0(x) = n_{eh}^0(x)$, the above expression simplifies to

$$\epsilon = 1 - \exp\left[\int_0^d dx_0\, n_{eh}^0(x_0)\log\left[1-\epsilon_{eh}(x_0)\right]\right]. \tag{3.7}$$

3.2 Stochastic initial conditions

Chapter 2 studied the emergence of avalanche fluctuations starting from fixed initial conditions. The deposit of charge carriers in silicon through absorption or ionisation is, however, itself stochastic. This leads to additional fluctuations and creates additional contributions to the time resolution.

The time-response function in Eq. 2.71 may be used to quantify these effects. Here and in the following, the approximation $A_{\text{eff}} \approx A$ is made. This neglects the (small) effects of boundaries on the avalanche fluctuations. Section 2.4.4 shows that these are of subleading importance; and using this approximation exposes the underlying physical effects without introducing undue complexity. The time-response function thus reads (cf. Eq. 2.71)

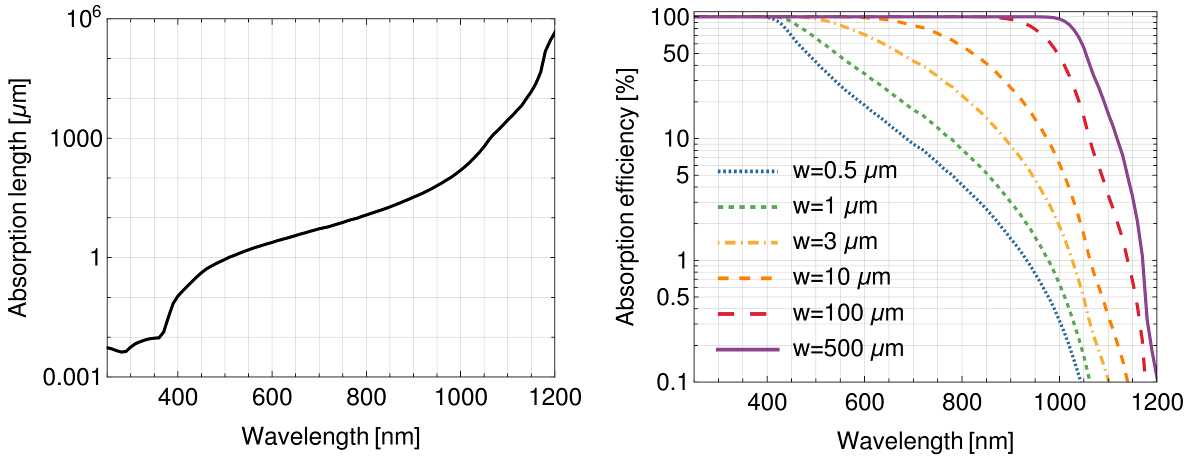
$$\rho(N_{\text{th}}, t) \approx \frac{\gamma_1 v^*}{\Gamma(A)} \exp\left(A \log\frac{hN_{\text{th}}}{e^{\gamma_1 v^* t}} - \frac{hN_{\text{th}}}{e^{\gamma_1 v^* t}}\right), \tag{3.8}$$

where

$$h = \frac{\epsilon A}{N(\lambda_1)}, \qquad A = \frac{\alpha v_e N_e^0 + \beta v_h N_h^0}{\alpha v_e + \beta v_h}.$$

This expression depends on the initial conditions in several ways. The avalanche parameter A depends only on the absolute number of primary electrons and holes, while the efficiency ϵ and the normalisation $N(\lambda)$ depend additionally on their spatial distribution. (In a more precise calculation that uses A_{eff} directly, the dependence of this parameter on the number of primary charge carriers and their positions must be taken into account according to Eq. 2.69.) The asymptotic time resolution $\sigma_{\text{th},\infty}$ is computed by including these parameters into the computation of expectation values. The n-th moment of the threshold crossing time t_{th} becomes

$$\langle t_{\text{th}}^n \rangle = \int_0^\infty dt\, t^n \int d(A,h)\, p(A,h)\, \rho(t, N_{\text{th}}).$$



(a) (b)

Figure 3.3: Optical properties for silicon. Data from Refs. [104, 105]. (a) Absorption length for photons in silicon as a function of the wavelength. (b) Efficiency $\epsilon_{\text{abs}}(w)$ to absorb a photon with a certain wavelength, depending on the thickness w of the conversion layer or gain layer.

The joint probability distribution $p(A, h)$ of the parameters A and h is determined by the (distribution of the) initial conditions. With this, the time resolution $\sigma_{\text{th},\infty}$ is

$$\sigma^2_{\text{th},\infty} = \frac{1}{(\gamma_1 v^*)^2} \left[\langle\psi_1(A)\rangle + \text{var}\left[\log h - \psi_0(A)\right]\right], \tag{3.9}$$

where the variance and the angle brackets symbolise the remaining expectation values over the fluctuating initial conditions.

As discussed in Section 2.4.4, the scale of the time resolution is set by the asymptotic growth rate of the avalanche. The two terms in parentheses on the right-hand side of Eq. 3.9 are dimensionless numbers and represent different physical effects. The first contribution represents the average contribution of avalanche fluctuations. The second term implements contributions due to fluctuations in the overall normalisation of the *average* avalanche. It depends on the positions and the number of the initial charge carriers as laid out in Section 2.4.2.2.

The relative importance of both contributions is determined by the physical mechanism by which primary charges are deposited in the sensor. They are studied in detail for photoabsorption (for photon detection) and ionisation (for charged particle detection) in Sections 3.3 and 3.4, respectively.

3.3 Photon detection

The properties of SPADs as photon detectors depend to a significant degree on the physics of photon absorption in silicon. The most important physical scale is the photon-absorption length l_a, shown

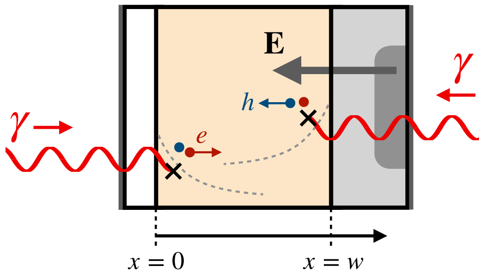


Figure 3.4: Long-wavelength photons may reach the device from the side of the substrate at $x = 0$ ("back-side illumination") or from the side of the gain layer at $x = w$ ("front-side illumination"). In both cases, the photons are primarily absorbed in the conversion layer and deposit an electron-hole pair.

in Figure 3.3a. It determines the efficiency ϵ_{abs} that a photon is absorbed after traversing a distance w in the medium, $\epsilon_{\text{abs}}(w) = 1 - e^{-\frac{w}{l_a}}$.

The absorption efficiency ϵ_{abs} is plotted in Figure 3.3b. Photons with a wavelength below 400 nm are efficiently absorbed even by very thin silicon layers with a thickness of only a few hundred nanometres. The gain layer itself is often sufficient to achieve good efficiencies [106] for this case. Conversely, infrared photons with wavelengths around 1 µm require conversion layers with a thickness of several tens of microns to achieve good efficiencies; photon absorption in the thin gain layer is negligible. Both wavelength regimes are considered in what follows.

3.3.1 Absorption in conversion layer

As Figure 3.4 shows, photons may impinge on the device from either direction and get absorbed in the conversion layer with thickness w. For illumination from the side of the substrate ("back-side illumination"), the probability for the photon to convert in a small interval dx_0 centred around a position x_0 is

$$p_{\text{abs}\rightarrow}(x_0) = \frac{1}{1 - e^{-\frac{w}{l_a}}} \frac{1}{l_a} e^{-\frac{x_0}{l_a}}. \tag{3.10}$$

It is normalised to the absorption efficiency, i.e. $\int_0^w dx_0\, p_{\text{abs}\rightarrow}(x_0) = 1$.

The photon absorption leads to a contribution σ_{abs} to the time resolution. The electron deposited at $x = x_0$ drifts for a time $t_d = (w - x_0)/v_e$ until it reaches the gain layer at $x = w$ and triggers the avalanche. The electron arrival time is a random variable, and Eq. 3.10 determines its distribution to be

$$\rho^{(1)}_{\text{arr}\rightarrow}(t) = \int_0^w dx_0\, p_{\text{abs}\rightarrow}(x_0)\, \delta\left[t - \frac{w - x_0}{v_e}\right] = \frac{1}{T_d l_a} \frac{w}{e^{w/l_a} - 1} \exp\left[\frac{tw}{T_d l_a}\right], \tag{3.11}$$

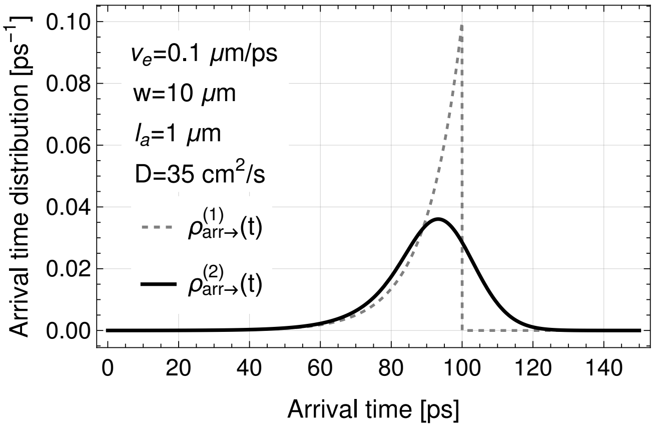


Figure 3.5: The distribution of the arrival time of the drifting electron at the gain layer, for a typical situation. The electron drift velocity v_e is taken to be close to the saturation velocity v_e^{sat}. The selected diffusion constant corresponds to the low-field limit [107]. Both $\rho^{(1)}_{\text{arr}\to}$ from Eq. 3.11 and $\rho^{(2)}_{\text{arr}\to}$ from Eq. 3.13 are shown.

where $T_d = w/v_e$ sets the drift timescale. It is visualised in Figure 3.5. The variance $\sigma^2_{\text{abs}\to}$ of this distribution is

$$\sigma^2_{\text{abs}\to} = T_d^2 \left[\frac{l_a^2}{w^2} - \frac{1}{4\sinh\left(\frac{w}{2l_a}\right)^2} \right]. \tag{3.12}$$

It is determined solely by the thickness w of the conversion layer and the exponential character of photon absorption. For illumination from the side of the gain layer ("front-side illumination"), the variance $\sigma^2_{\text{abs}\leftarrow}$ is found to be identical, i.e. $\sigma^2_{\text{abs}} = \sigma^2_{\text{abs}\to} = \sigma^2_{\text{abs}\leftarrow}$.

Superimposed on the (deterministic) drift motion of the electron are (stochastic) interactions with other charge carriers and the lattice. Their effect on the arrival time distribution are approximately taken into account by convolving the distribution in Eq. 3.10 with the kernel

$$k(x, x_0, t) = \frac{1}{\sqrt{2\pi}\sqrt{2Dt}} \exp\left[-\frac{[x - (x_0 + v_e t)]^2}{2(2Dt)} \right].$$

which is a solution of the diffusion equation. It determines the probability density to find the electron at coordinates (x, t) if it originated at x_0 at $t = 0$. This results in the following improved approximation of the arrival time distribution,

$$\begin{aligned} \rho^{(2)}_{\text{arr}\to}(t) &= \int_0^w dx_0\, p_{\text{abs}\to}(x_0) k(x = w, x_0, t) \\ &= \frac{v_e}{2l_a} \frac{e^{(D + l_a v_e)t/l_a^2}}{e^{w/l_a} - 1} \left(\text{Erf}\left[\frac{\sqrt{Dt}}{l_a} + \frac{v_e}{2}\sqrt{\frac{t}{D}} \right] - \text{Erf}\left[\frac{\sqrt{Dt}}{l_a} + \frac{v_e}{2}\sqrt{\frac{t}{D}} - \frac{w}{2\sqrt{Dt}} \right] \right), \end{aligned} \tag{3.13}$$

where D is the longitudinal diffusion constant for silicon and $\text{Erf}(x) = 2/\sqrt{\pi}\int_0^x dz\, e^{-z^2}$. Diffusion contributes a new term $\sigma^2_{\text{diff}\rightarrow}$ to the variance of the arrival time,

$$\sigma^2_{\text{diff}\rightarrow} = T_d \frac{2D}{v_e^2} \left(\frac{1}{1 - e^{-w/l_a}} - \frac{l_a}{w} \right) . \tag{3.14}$$

For front-side illumination, the contribution from diffusion is instead

$$\sigma^2_{\text{diff}\leftarrow} = T_d \frac{2D}{v_e^2} \left(\frac{l_a}{w} + \frac{1}{1 - e^{w/l_a}} \right) . \tag{3.15}$$

The total variance of the electron-arrival time, and hence the full contribution of the conversion layer to the time resolution of the device, is $\sigma^2_{\text{abs}} + \sigma^2_{\text{diff}}$. Its exact numerical value depends critically on several design choices (drift field, conversion layer thickness) and must be determined for a specific device. It is worthwhile to evaluate the expressions derived above for some limiting cases to reveal the overall scale of the expected time resolution. Expressions for the cases $l_a \gg w$ (where absorption occurs throughout the entire conversion layer) and $d \ll l_a \ll w$ (where absorption preferentially occurs close to the surface of the conversion layer, but absorption in the gain layer is negligible) are highlighted in Table 3.1.

The scale of σ_{abs} is set by the drift timescale, T_d. If the absorption length is long compared to the thickness of the conversion layer, the position x_0 at which the conversion happens is uniformly distributed, leading to the appearance of a factor $1/\sqrt{12}$. If the photon converts close to the surface of the conversion layer, the fluctuations in x_0 and thus also σ_{abs} are reduced. For a drift velocity close to saturation, i.e. $v_e \approx 0.1\,\mu\text{m/ps}$, the drift timescale is $T_d = 10/100/1000\,\text{ps}$ for a conversion layer with thickness $w = 1/10/100\,\mu\text{m}$.

The contribution σ_{diff} is characterised by the timescale $T_{\text{diff}} = \sqrt{DT_d}/v_e$. For $l_a \gg w$, diffusion is not sensitive to the direction of the arriving photons. Naturally, diffusion is less important if the photon absorption occurs close to the gain layer. The longitudinal diffusion coefficient of electrons in silicon is bounded from above by $D \approx 35\,\text{cm}^2/\text{s}$ [107], attained at low electric fields. The diffusion timescale is thus approximately $T_{\text{diff}} = 1.6/8.4/26.5\,\text{ps}$ for $w = 1/10/100\,\mu\text{m}$ and a saturated electron drift velocity.

3.3.2 Absorption in gain layer

For photon absorption in the gain layer, the efficiency and the overall normalisation of the avalanche depend on the conversion point, as discussed in Sections 2.4.2.2 and 3.1. The avalanche is initiated by an electron-hole pair at x_0. For this initial condition, $A = 1$; and the contribution to the time

		$l_a \gg w$	$d \ll l_a \ll w$
Diffusion	Front-side illumination	$\sigma_{\text{diff}\leftarrow} = \frac{\sqrt{DT_d}}{v_e}$	$\sigma_{\text{diff}\leftarrow} = \sqrt{\frac{2l_a}{w}}\frac{\sqrt{DT_d}}{v_e}$
	Back-side illumination	$\sigma_{\text{diff}\rightarrow} = \frac{\sqrt{DT_d}}{v_e}$	$\sigma_{\text{diff}\rightarrow} = \sqrt{2}\frac{\sqrt{DT_d}}{v_e}$
Absorption		$\sigma_{\text{abs}} = \frac{T_d}{\sqrt{12}}$	$\sigma_{\text{abs}} = T_d \frac{l_a}{w}$

Table 3.1: Contributions to the time resolution from photon absorption and charge carrier diffusion in the limits $l_a \gg w$ and $d \ll l_a \ll w$, based on the expressions in Eqs. 3.12, 3.14, and 3.15.

resolution originating from the avalanche is quantified by Eq. 3.9,

$$\sigma^2_{\text{th},\infty} = \sigma^2_{\text{av}} + \sigma^2_{\text{pos}} = \frac{1}{(\gamma_1 v^*)^2}\left[\frac{\pi^2}{6} + \text{var}\left[\log h\right]\right],$$

with $h = \epsilon/N(\lambda_1)$. The first term represents the contribution from avalanche fluctuations, where $\pi^2/6 = \psi_1(1)$. Written explicitly, the variance in the second term is

$$\text{var}\left[\log h\right] = \int_0^d dx_0\, p_{\text{abs}}(x_0) \log^2 h(x_0) - \left[\int_0^d dx_0\, p_{\text{abs}}(x_0) \log h(x_0)\right]^2, \tag{3.16}$$

where $p_{\text{abs}}(x_0)dx_0$ is the probability that the photon is absorbed in a small interval dx_0 centred on x_0 in the gain layer. This probability distribution is given by an exponential distribution analogous to Eq. 3.10.

An estimate for the overall scale of the contribution in Eq. 3.16 may be obtained by taking p_{abs} to be a uniform distribution, i.e. $p_{\text{abs}}(x_0) = 1/d$. The variance may then be computed using the analytic expressions for the efficiency in Eqs. 3.5 and the normalisation of the avalanche in Eq. 2.57. Figure 3.6 compares the result to the universal contribution from the avalanche fluctuations.

The term σ_{pos} depends significantly on the electric field and on the thickness d of the gain layer. It is generally of the same order as the contribution σ_{av} from avalanche fluctuations but may become dominant for high electric fields and thick gain layers. This dependence originates from the behaviour of the coefficient $N(\lambda_1)$ defined in Eq. 2.54, which controls the overall normalisation of the avalanche.

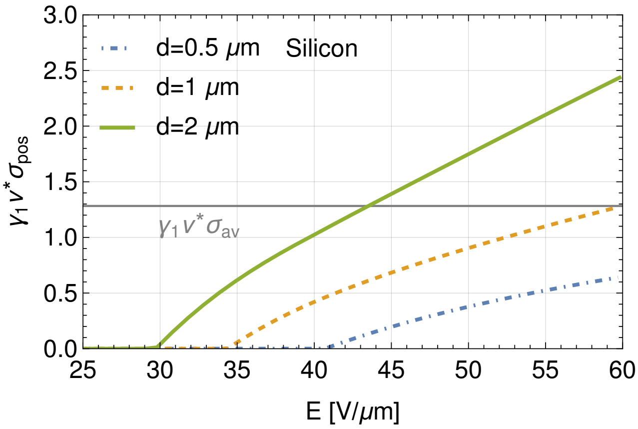


Figure 3.6: The contribution σ_{pos} from the variations in the position x_0 for different values of the electric field and the thickness d of the gain layer, assuming x_0 to be uniformly distributed throughout the gain layer. For comparison, the thin horizontal line shows σ_{av} originating from the avalanche fluctuations.

3.4 Charged particle detection

The properties of SPADs as fast detectors for charged particles are grounded on the interactions of these particles with matter. Section 3.4.1 briefly reviews some of these aspects insofar as they are relevant for the ensuing discussions. Based on these results, Section 3.4.2 investigates the detection efficiencies that SPADs can achieve for charged particles, and the time resolution is studied in Section 3.4.3.

3.4.1 Interaction of charged particles with matter

The situation of interest is the detection of fast ($\beta\gamma \gtrsim 3$), heavy (heavier than electrons) charged particles. Such minimum-ionising particles (MIPs) interact electromagnetically with the detector material and deposit energy primarily through ionisation. The details of the interaction are described by the differential cross-section $d\sigma/dE$ to deposit a certain amount of energy E in the material in a collision.

Description of the interaction The maximal energy E_{max} that can be transferred in a collision is determined by the kinematics of the scattering process, and for a MIP is much less than the total kinetic energy of the particle. The particle may thus independently undergo a series of interactions as it traverses the material. The mean free path λ between two ionisation events is the inverse of the expected number of collisions per unit length. It is related to the number density of scattering centres in the detector material, N, and the total cross-section σ as

$$\lambda^{-1} = N\sigma = N \int_{E_{\text{min}}}^{E_{\text{max}}} dE\, \frac{d\sigma}{dE},$$

where the minimum energy transfer E_{min} is given by the accessible excitations of the medium and E_{max} is determined by the kinematics of the scattering. For silicon, the ionisation mean free path is approximately $\lambda \approx 0.23\,\mu\text{m}$ at $\beta\gamma = 3$ [108]. The first moment of the differential cross section is the expected energy loss per unit length, referred to as the "stopping power" $\langle dE/dx \rangle$,

$$\left\langle \frac{dE}{dx} \right\rangle = N \int_{E_{\text{min}}}^{E_{\text{max}}} dE\, E \frac{d\sigma}{dE}. \tag{3.17}$$

As defined here, x has dimensions of length.

The number of interactions of the particle as it traverses a medium with thickness d follows a Poisson distribution with a mean of d/λ. The energy E deposited in a collision event leads to the formation of a cluster of n electron-hole pairs located at the position x at which the interaction took place. The expected number of electron-hole pairs in the cluster $\langle n \rangle$ and its variance $\sigma^2(n)$ are

$$\langle n \rangle = \frac{E}{E_0}, \qquad \sigma^2(n) = F \langle n \rangle\,,$$

where E_0 is the average energy required to excite an electron-hole pair, and F is a Fano factor [109]. For silicon, $E_0 \approx 3.7\,\text{eV}$ [110] and $F < 0.1$ [108].

The distribution of n, the "cluster-size distribution" $p_{\text{clu}}(n)$, is related to the cross-section as

$$p_{\text{clu}}(n) = \frac{1}{\sigma} \int dE\, \frac{d\sigma}{dE}\, p(n; E),$$

where $p(n; E)$ is the probability to create a cluster of n electron-hole pairs from an energy deposit E. In the limit of small Fano factors, $p(n; E) \approx \delta(n - E/E_0)$ and the cluster-size distribution becomes approximately proportional to the cross-section,

$$p_{\text{clu}}(n) \approx \frac{E_0}{\sigma} \frac{d\sigma}{dE}\bigg|_{nE_0}. \tag{3.18}$$

For M produced clusters, the distribution of the total number of electron-hole pairs is given by the M-fold convolution of p_{clu} with itself, denoted $(p_{\text{clu}})^{*M}$. Correspondingly, the probability to deposit a total of N_{eh}^0 electron-hole pairs in a material of thickness d becomes

$$p(N_{eh}^0) = \sum_{M=0}^{\infty} \text{Po}\left(M; \frac{d}{\lambda}\right) (p_{\text{clu}})^{*M}\,(N_{eh}^0), \tag{3.19}$$

where $\mathrm{Po}(M; d/\lambda) = e^{-d/\lambda}(d/\lambda)^M/M!$ denotes the Poisson distribution.

Classical dielectric theory and the photoabsorption ionisation model A classical treatment is adequate in the limit of a slow and quasi-continuous energy loss. In electrodynamics, a moving point-charge q polarises the medium that it traverses; the polarisation, in turn, influences the electric field **E** at the position of the particle. The component of **E** projected along the direction of motion $\hat{\boldsymbol{\beta}}$ of the particle determines the backreaction of the material, and, therefore, the energy loss,

$$\left\langle \frac{dE}{dx} \right\rangle = q\mathbf{E} \cdot \hat{\boldsymbol{\beta}}.$$

An explicit calculation [111] shows that this quantity may be expressed purely in terms of the dielectric function $\epsilon(k, \omega)$ that appears in Gauss' law in Eq. 1.1a,

$$\begin{aligned}\left\langle \frac{dE}{dx} \right\rangle = -\frac{2q^2}{\beta^2 \pi} \int_0^\infty d\omega \int_{k_{\mathrm{min}}(\omega)}^\infty dk \\ \times \left[\omega k \left(\beta^2 - \frac{\omega^2}{k^2 c^2} \right) \mathrm{Im}\left(\frac{1}{-k^2 c^2 + \epsilon(k,\omega)\omega^2} \right) + \frac{\omega}{kc^2} \mathrm{Im}\left(-\frac{1}{\epsilon(k,\omega)} \right) \right]\end{aligned} \tag{3.20}$$

The integration involves an integration over the entire accessible kinematic plane. The dielectric function $\epsilon(k, \omega)$ describes the coupling of the material to the electromagnetic field generated by the particle. It can be computed from first principles only for a few simple materials. For practical purposes, a simple model of $\epsilon(k, \omega)$ based on optical data is often sufficient.

The integral over the first term in Eq. 3.20 is dominated by on-shell photons for which $k = \sqrt{\epsilon}\,\omega/c$. In this optical limit, the dielectric function depends only on the frequency, i.e. $\epsilon = \epsilon(\omega)$. The photoabsorption ionisation (PAI) model [111, 112] extends this approximation also into the off-shell region of the kinematic plane. In this model, also the second term of Eq. 3.20 is expressed in terms of $\epsilon(\omega)$. The optical-loss function $\mathrm{Im}(1/\epsilon(\omega))$ summarises the properties of the material and that determines the energy loss. By comparing Eqs. 3.20 and 3.17, the differential cross section $d\sigma/dE$ may be approximated in terms of the loss function.

Figure 3.7 shows experimental data for the loss function $\mathrm{Im}(-1/\epsilon(\omega))$ of silicon. Photoionisation becomes possible above the band gap of intrinsic silicon. The plasmon peak at around 16 eV is due to collective excitations of the valence electron gas. The K and L edges of atomic silicon at 99 eV and 1.84 keV are also visible.

At high energy transfers E, the electrons in the material appear to be unbound and the cross-section scales as $d\sigma/dE \sim 1/E^2$, familiar from Rutherford scattering. For large n, the cluster-size

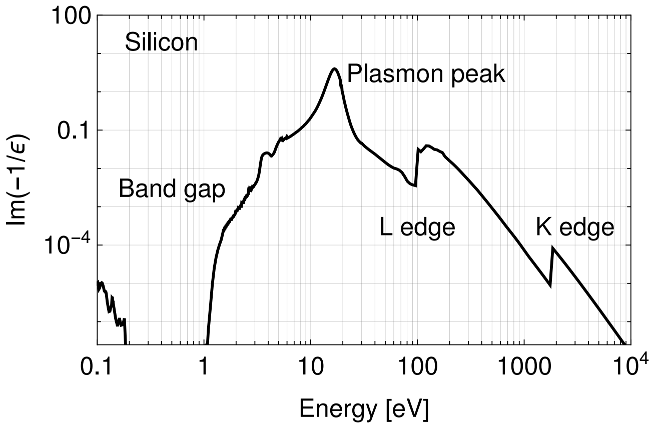


Figure 3.7: Loss function Im$(-1/\epsilon(\omega))$ for silicon, plotted in units of energy $E = \hbar\omega$. Data from Refs. [113–115].

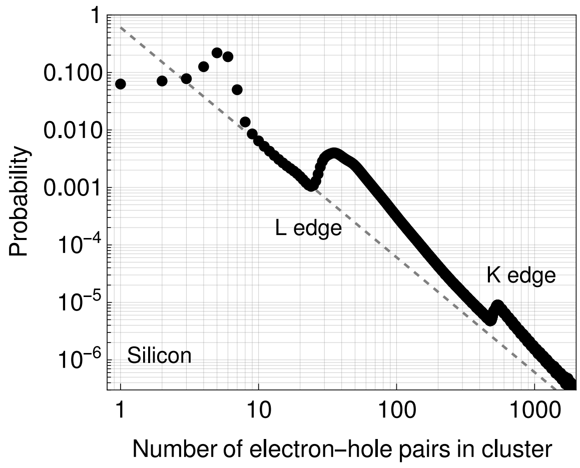
distribution is thus expected to show the same scaling behaviour, $p_{\text{clu}}(n) \sim 1/n^2$. Figure 3.8a compares this naive expectation with the full cluster-size distribution for silicon, as computed with the implementation of the PAI model available in HEED [116]. The distribution mirrors the structures visible in the optical data of Figure 3.7, as anticipated by Eq. 3.18, and deviates significantly from the simple Rutherford scaling law.

Figure 3.8b shows the distribution of the total number of generated electron-hole pairs in the material according to Eq. 3.19. The expected number of charges is substantial compared to photoabsorption. Even for a very thin layer with $d = 0.5\,\mu\text{m}$, the most probable number of electron-hole pairs lies significantly above one.

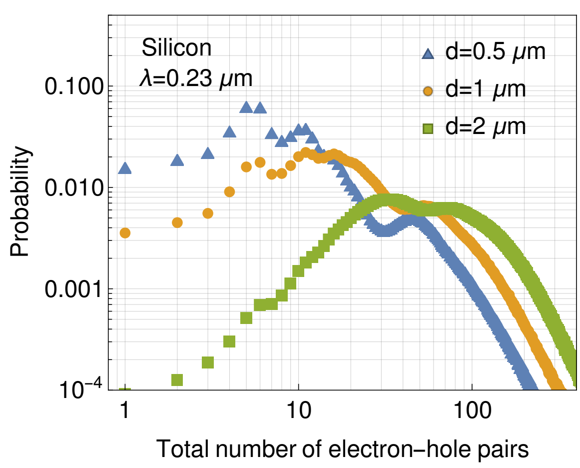
HEED relies on the photoabsorption cross-section of atomic silicon to extract the loss function Im$(-1/\epsilon(\omega))$. Collective effects are therefore not included and $p_{\text{clu}}(n)$ is not expected to be accurately modelled for small clusters. Ref. [117] compares this result with a more careful calculation that uses the full loss function from Figure 3.7. This improvement leads to differences in the predicted energy-loss distributions for extremely thin silicon layers. However, neither model has been validated against experimental data, and the (simpler) HEED calculations continue to be used in the following.

3.4.2 Efficiency

Eq. 3.7 determines the efficiency for a fixed initial density $n^0_{eh}(x)$ of electron-hole pairs. To compute the efficiency ϵ_{MIP} for MIP detection, this expression must be averaged over the distribution of n^0_{eh}



(a)



(b)

Figure 3.8: (a) The black markers show the cluster-size distribution $p_{\text{clu}}(n)$ for silicon, computed with the PAI model as implemented in HEED [116]. The grey dashed line corresponds to a $1/n^2$ distribution as expected from Rutherford scattering. (b) Distribution of the total number of electron-hole pairs deposited by a MIP for silicon layers with different thickness d.

produced by the interaction of the MIP with the sensor,

$$\epsilon_{\text{MIP}} = \int Dn^0_{eh}\, p[n^0_{eh}]\, \epsilon[n^0_{eh}], \tag{3.21}$$

where $\epsilon[n^0_{eh}]$ is the expression in Eq. 3.7. The square brackets are used to indicate its functional dependence on the density of initial electron-hole pairs. For the following calculation, only interactions in the gain layer are considered.

The distribution $p[n^0_{eh}]$ is determined by the interactions with the medium as discussed in Section 3.4.1. The passing MIP leads to initial densities of the form

$$n^0_{eh}(M, \{x_i\}, \{n_i\}) = \sum_{j=1}^{M} n_j \delta(x - x_j),$$

where M is the number of interactions. The resulting charge clusters are located at positions $x_1, \ldots, x_M$ and contain $n_1, \ldots, n_M$ electron-hole pairs, respectively. The number of clusters follows a Poisson distribution with a rate parameter of d/λ. The cluster sizes n_i are independently distributed according to $p_{\text{clu}}(n)$, and the cluster positions x_i are uniformly distributed throughout the gain layer.

This determines the distribution $p[n^0_{eh}]$ to be

$$p[n^0_{eh}] = \sum_{M=0}^{\infty} \text{Po}\left(M; \frac{d}{\lambda}\right) \int_0^d \frac{dx^M}{d^M} \sum_{n_1,\ldots,n_M} \prod_{i=1}^{M} p_{\text{clu}}(n_i)\, \delta\left[n^0_{eh} - n^0_{eh}(M, \{x_i\}, \{n_i\})\right]. \tag{3.22}$$

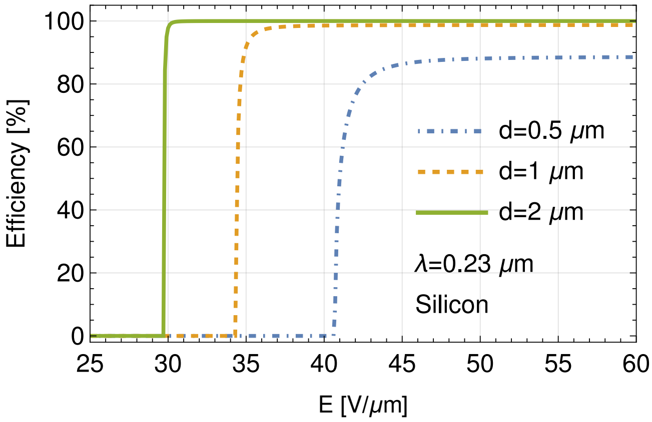


Figure 3.9: Efficiency ϵ_{MIP} for a MIP to trigger a diverging avalanche in silicon, shown as a function of the electric field in the gain layer and its thickness d.

From Eq. 3.21, the efficiency ϵ_{MIP} becomes

$$\epsilon_{\text{MIP}} = 1 - e^{-\frac{d}{\lambda}} \exp\left[\frac{1}{\lambda}\sum_{n=1}^{\infty} p_{\text{clu}}(n) \int_0^d dx\,(1-\epsilon_{eh}(x))^n\right]. \tag{3.23}$$

Figure 3.9 evaluates this expression for the cluster-size distribution from Figure 3.8a. Above breakdown, the efficiency rises steeply and saturates at $\epsilon_{\text{MIP}} = 1 - e^{-d/\lambda}$. At high fields, $\epsilon_{eh} \approx 1$ and the efficiency becomes limited by the probability for the MIP to undergo a single interaction in the gain layer. This confirms that a thin gain layer is sufficient to provide high efficiencies and that a dedicated conversion layer is not needed.

3.4.3 Time resolution

Similarly, also the time resolution of the device is dominated by the interactions of the MIP in the gain layer. It may be computed analogously to the case of photon detection discussed in Section 3.3.2, starting from Eq. 3.9.

The expectation values and variances must now be computed with respect to the distribution in Eq. 3.22. The avalanche parameter A is given by the total number of electron-hole pairs deposited in the gain layer, $A = \sum_i n_i = N^0_{eh}$, and the quantity $\log h$ is $\log h = \log A + \log \epsilon - \log N(\lambda_1)$.

As demonstrated above in Section 3.4.2, the averaged efficiency ϵ_{MIP} is close to maximal above breakdown. This implies that the efficiency for each initial density n^0_{eh} included in $p[n^0_{eh}]$ must itself be close to maximal. The quantity $\log \epsilon$ is thus constant to an excellent approximation and does not contribute to the overall variance. Removing this term from the computation of the variance, Eq. 3.9

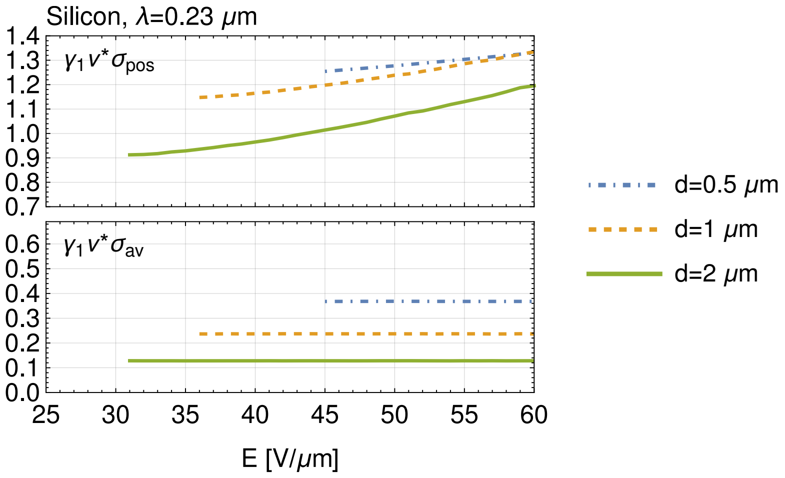


Figure 3.10: The top panel shows the contribution σ_{pos} arising from the fluctuations in the cluster positions and the cluster sizes, as a function of the electric field and for different gain layer thickness. For comparison, the terms σ_{av} from the fluctuations of the avalanche are shown at the bottom.

becomes

$$\begin{aligned}\sigma^2_{\text{th},\infty} = \sigma^2_{\text{av}} + \sigma^2_{\text{pos}} &\approx \frac{1}{(\gamma_1 v^*)^2}\left[\langle \psi_1(N^0_{eh})\rangle + \text{var}\left[\log N^0_{eh} - \log N(\lambda_1) - \psi_0(N^0_{eh})\right]\right] \\ &\approx \frac{1}{(\gamma_1 v^*)^2}\left[\langle \psi_1(N^0_{eh})\rangle + \text{var}\left[\log N(\lambda_1)\right]\right]. \end{aligned} \tag{3.24}$$

In the second line, the approximation $\psi_1(N^0_{eh}) \approx \log N^0_{eh}$ is made, which is accurate to about 10% for $N^0_{eh} > 4$, where most of the probability mass lies (cf. Figure 3.8b).

The two contributions in Eq. 3.24 are compared and visualised in Figure 3.10. Avalanche fluctuations are reduced by the large amount of primary charge, and significantly suppressed compared to the detection of individual photons. Their contribution σ_{av} to the total time resolution is not important relative to the effects of varying cluster positions and cluster sizes contained in σ_{pos}. In contrast to photon detection, this term *decreases* with increasing thickness d: on average, thicker gain layers contain more charge clusters, which reduces the effect of position fluctuations.

3.5 Summary

The theory of electron-hole avalanches developed in Chapter 2, together with the physics of charge deposition in silicon, explains the main functional properties of SPADs, predicts the ultimate time resolution that can be achieved with these devices, and clearly identifies the most important design parameters. Such analytic results are important for the validation of detailed, device-specific, simulations; they also guide the development of sensors with improved characteristics.

Self-sustaining electron-hole avalanches in silicon are shown to provide a powerful amplification mechanism that can achieve intrinsic time resolutions significantly below 10 ps for sufficiently strong electric fields (cf. Figure 2.16b). This fundamental limit arises from the fluctuations in the avalanche development. Other sources also contribute to the achievable time resolution of a practical SPAD.

For single-photon detection, the timing performance depends significantly on the targeted wavelength regime. For infrared photons, it is limited by drift and diffusion in the thick conversion layer, which can easily contribute 50–100 ps of additional jitter (cf. Table 3.1). Short-wavelength photons are efficiently absorbed in the gain layer itself. In this case, the overall scale of the time resolution improves with the electric field in the gain layer. It receives a contribution from fluctuations in the position of the photon conversion, which is of a similar order as the intrinsic avalanche fluctuations but can become dominating for relatively thick gain layers (cf. Figure 3.6).

For the detection of charged particles, the statistics of the energy loss in the gain layer through ionisation is the most important aspect, and avalanche fluctuations only play a subleading role. Even for very thin devices without a dedicated conversion layer, the detection efficiency is close to maximal (Figure 3.9).

Remarkably, the design criteria to optimise the time resolution in these two applications are not conflicting, and the timing performance of SPADs is shown to be remarkably universal even across different particle species. SPADs designed for precision photon timing applications are thus expected to achieve an almost identical time resolution also for the detection of charged particles. The capability to detect photons and charged particles simultaneously is interesting for the next generation of collider experiments. It could enable the design of sensors that combine the functionality of a time-of-flight detector for charged particles with particle identification through the imaging of Cherenkov emission [118].

Chapter 4

In-beam performance of single-photon avalanche diodes

The discussion in Section 3.4 provides the theoretical justification and motivation for applying SPADs as precision timing detectors for charged particles. Before this technology can be employed, its performance in an experimental setup must be characterised and understood.

Several studies investigating the properties of individual SPADs (or SPAD arrays) as charged particle detectors have been performed in the past. They find good sensitivity to charged particles [119, 120] and confirm the expectation that the time resolution is similar to that for single photons [121, 122]. Nonetheless, many aspects remain to be understood, including the interplay between timing properties, position resolution, and efficiency, and the selection of the optimal operating conditions.

This chapter contributes to this effort and characterises the in-beam performance of a prototype CMOS SPAD array. All results have been obtained with a versatile experimental setup constructed for this measurement campaign. It features a high-frequency readout pipeline for the amplification and digitisation of the signal delivered by the sensor. This is combined with a silicon pixel detector, allowing the crossing charged-particle tracks to be reconstructed and the hit position to be determined relative to the pixel structure of the SPAD array. The combination of tracking and high-fidelity analogue information represents a significant advance compared to previous studies and makes it possible to investigate the functional properties of the tested sensor in much greater detail.

The most relevant properties of the studied device are listed in Section 4.1. Sections 4.2–4.5

Parameter	Symbol	Value
Pixel pitch	p	25 µm
Array size	–	$6 \times 16 = 96$
Effective area	$w \times h$	150 µm × 400 µm
Diode diameter	d	≈ 20 µm
Fill factor	–	≈ 50 %
Thickness of epitaxial layer	–	≈ 10 µm

Table 4.1: Relevant characteristics of the tested SPAD arrays, referred to as "SPAD/P96" in the main text. The fill factor is defined as the fraction of the total pixel area covered by the diode.

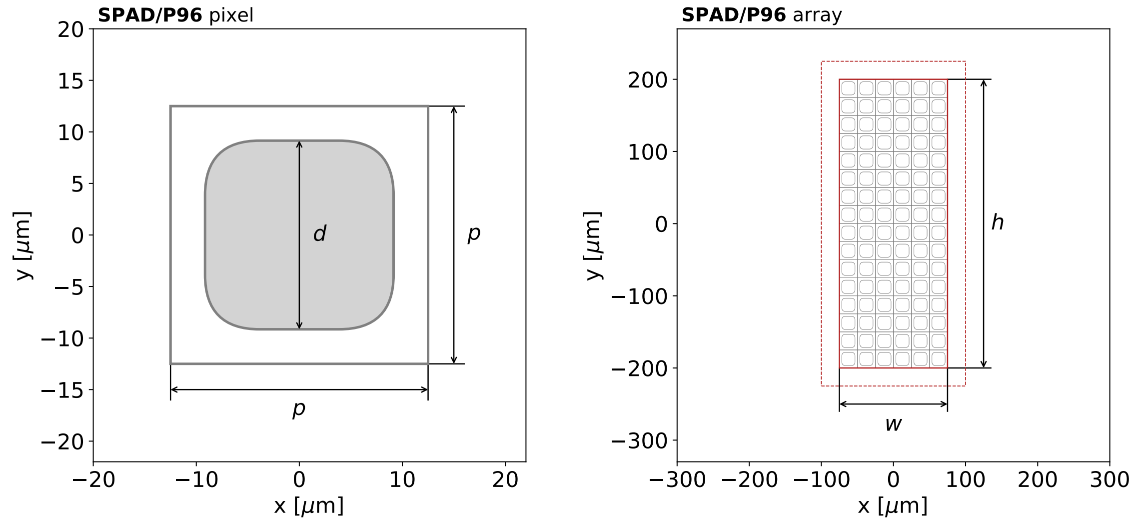
describe the developed readout electronics and the remainder of the experimental setup in more detail, and a summary of the results obtained from a first beam test campaign is available in Section 4.6.

4.1 Device information

The device under test (DUT) is a small, rectangular array of $6 \times 16 = 96$ SPAD pixels manufactured in a CMOS process. Table 4.1 summarises important device parameters, and the array and pixel geometries are illustrated in Figure 4.1. Information on doping profiles is proprietary and not available. Two samples with nominally identical characteristics are used in the following, referred to as SPAD/P96/1 and SPAD/P96/2.

The breakdown voltage V_{br} at room temperature is extracted from the relationship between bias voltage and reverse current as explained in Ref. [123] and found to be around 21 V, compatible with information from the foundry. (A more detailed in situ measurement of the breakdown voltage is shown in Section 4.6.1.) The temperature coefficient of V_{br} is positive and determined by the temperature dependence of the impact ionisation coefficients (cf. Section 2.5). It is measured in a separate device sample which consists of only a single SPAD cell and found to be around 25 mV / K.

Figure 4.2a shows the in-pixel circuit components arranged around the p–n junction, consisting of the quench resistor R_q and the elements C_c and R_c to couple out the signal. According to the discussion in Section 2.6 the avalanche discharge induces a voltage signal V^{ind} with negative polarity. The slew rate of the signal is determined by the dynamics of the avalanche and the presence of additional parasitic elements (not shown) that reflect the effects of bonding wires and the device



(a) (b)

Figure 4.1: Geometry of the tested SPAD array. Figure (a) shows the structure of a single SPAD pixel with pitch p. The diode has a diameter $d < p$, defined as shown. Figure (b) illustrates the entire array. The thick red rectangle marks the nominally active array and the dashed red line indicates the presence of a dummy row and dummy column on each side.

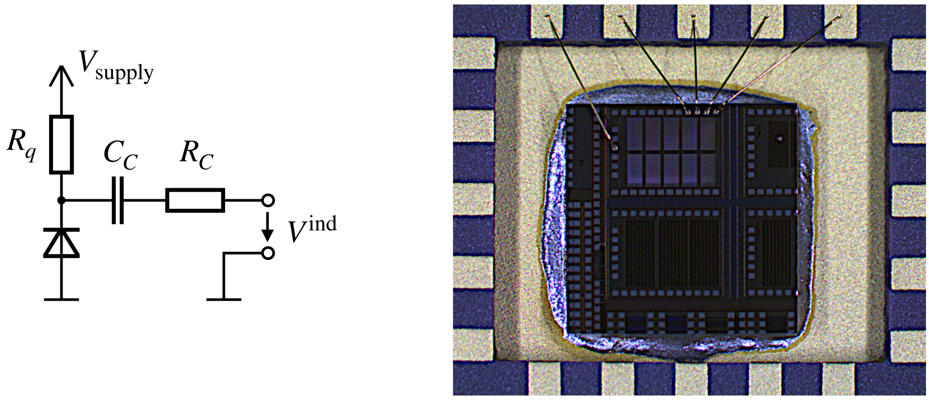
package on the electrical signal path.

The signal lines from all pixels in the array are connected in parallel and only the summed signal is routed outside the package. An additional dummy row and an additional dummy column of pixels surround the active part array on either side. The SPADs in these regions are not connected to the supply and the common signal line.

4.2 Readout electronics

A custom front-end amplifier is used to increase the signal amplitude and provide a standardised impedance of $Z_0 = 50\,\Omega$ while preserving its analogue bandwidth as much as possible.

Figure 4.3 shows a simplified schematic diagram of the developed broadband amplifier, targeting a frequency band from 100 MHz to 5 GHz with $50\,\Omega$ ports. It consists of two separate amplification stages. The first stage is designed around the low-noise SiGe hybrid bipolar transistor Q3 (BFP840FESD, see Ref. [124]), operated in a common emitter configuration. It acts as an inverting amplifier with a voltage gain of around 19 dB. The operating point of Q3 is actively controlled by the low-frequency transistors Q1 and Q2, which also automatically compensate temperature variations. The second (non-inverting) stage consists of the integrated amplifier U1 (HMC313, see Ref. [125]) which increases the overall gain to about 35 dB. The full circuit achieves a (simulated) 3 dB-cutoff



(a) (b)

Figure 4.2: (a) In-pixel circuit, consisting of the quench resistor $R_q \approx 300\,\mathrm{k\Omega}$ and the elements $C_c \approx 10\,\mathrm{fF}$ and $R_c \approx 100\,\mathrm{k\Omega}$ to extract the signal. (b) Magnified view of the silicon die containing the device SPAD/P96/2.

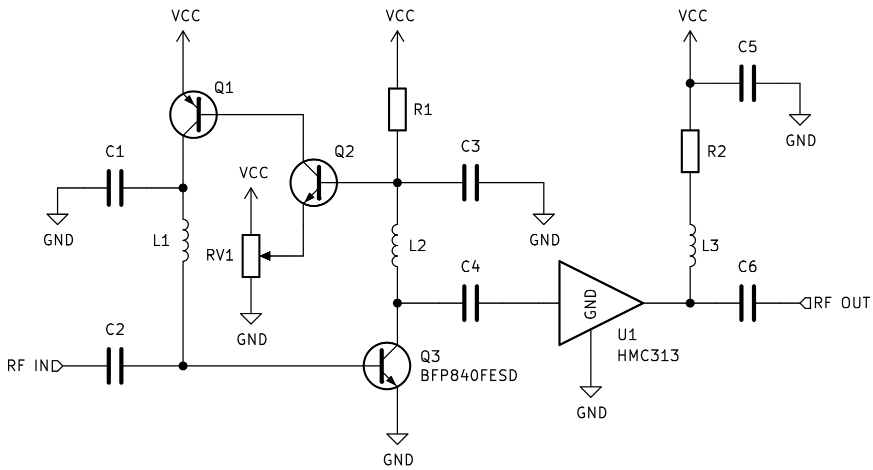


Figure 4.3: Simplified schematic diagram of the front-end amplifier, consisting of a discrete common-emitter transistor amplifier (Q3) as first stage, followed by an integrated amplifier (U1) as second stage. The signal input is labelled as "RF IN" and the amplified output signal is available at the terminal marked "RF OUT".

frequency of 5.3 GHz. It was demonstrated to be stable independent of the impedances of the signal source and the load using the methods summarised in Ref. [126].

The first amplification stage was routed on a 4-layer printed circuit board (PCB), pictured in Figure 4.4. All traces carrying high-frequency signals are implemented as grounded coplanar waveguides [127, 128] with a characteristic impedance of $50\,\Omega$, and kept strictly separate from supply lines or low-frequency signals. The second stage is housed on a small separate PCB. Both stages are connected and combined in a common aluminium enclosure for shielding purposes. A dedicated amplifier is used for each SPAD array.

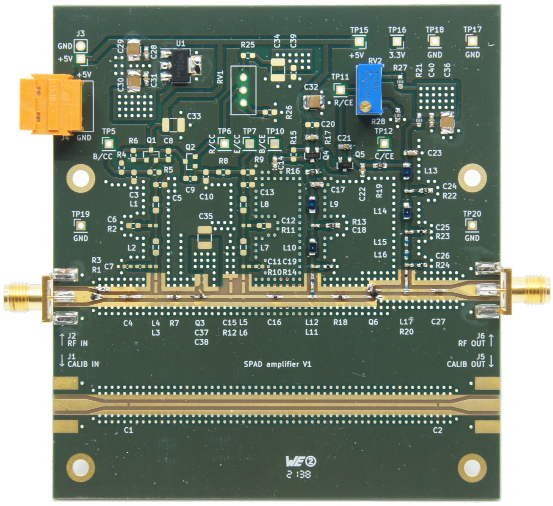


Figure 4.4: Printed circuit board housing the first amplification stage, shown to scale 1:1.5.

4.3 Experimental setup

The two DUTs together with the front-end amplifiers described in Section 4.2 are embedded in an experimental setup consisting of a beam telescope, a trigger system, and ancillary equipment for data acquisition, monitoring, and slow control. Figure 4.5 gives an overview of the full system. It serves as a guide to the following description, which covers the most important functional components of the apparatus.

4.3.1 Beam telescope

The beam telescope is the only component exposed to the particle beam during the measurement campaign. It combines the two DUTs, four tracking planes, and three trigger scintillators in a linear arrangement whose layout is shown schematically in Figure 4.6. The longitudinal z-axis is chosen to be parallel to the beam direction.

4.3.1.1 Tracking planes

The tracking layers consist of ALPIDE monolithic active pixel sensors [129, 130]. Each ALPIDE sensor contains 1024×512 pixels distributed over an active area of $3\,\mathrm{cm} \times 1.5\,\mathrm{cm}$. Each pixel has a size of around $29\,\mu\mathrm{m} \times 27\,\mu\mathrm{m}$. The in-pixel circuitry includes a discriminating amplifier with a peaking time of about $2\,\mu\mathrm{s}$ and a series of digital buffers to store the hit information upon receipt of an externally provided trigger signal. The sensor is read out through a serial interface. It achieves a sub-pixel position resolution of below $5\,\mu\mathrm{m}$. The fake hit rate is below $10^{-6}\,\mathrm{pixel}^{-1}\,\mathrm{event}^{-1}$ for detection

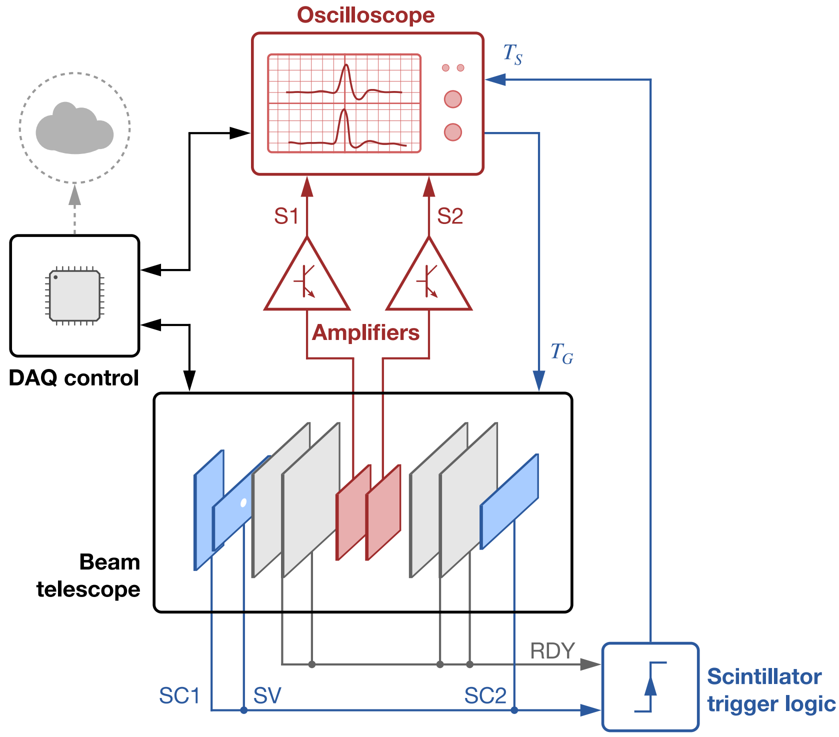


Figure 4.5: Overview of the beam-test setup, consisting of the beam telescope, the analogue pipelines for the SPAD signals (red), a trigger system involving scintillators (blue), and a central controller for data acquisition and control (DAQ). The signal labels are explained in the main text.

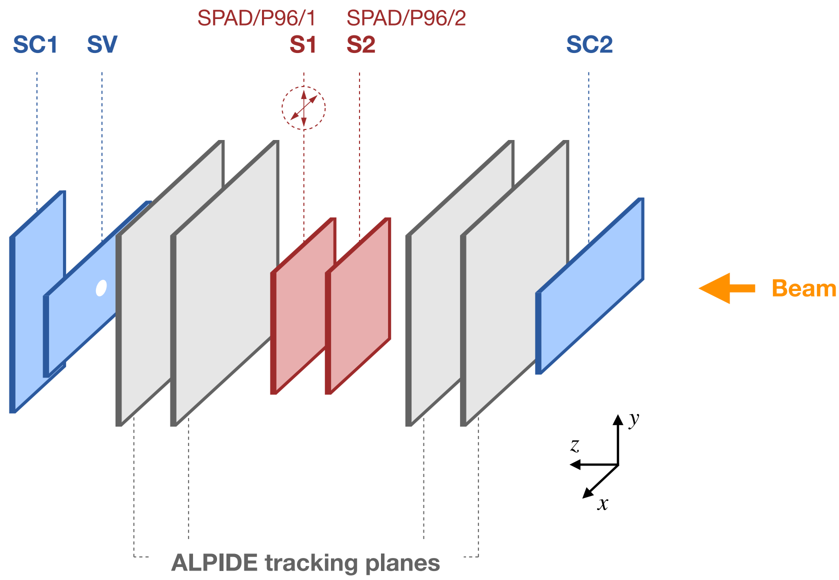


Figure 4.6: Layout of the beam telescope, consisting of four ALPIDE tracking layers (grey), the two DUTs, SPAD/P96/1 and SPAD/P96/2, and three trigger scintillators, SC1, SC2, and SCV. SPAD/P96/1 can be moved in the transverse xy-plane for it to be aligned with SPAD/P96/2.

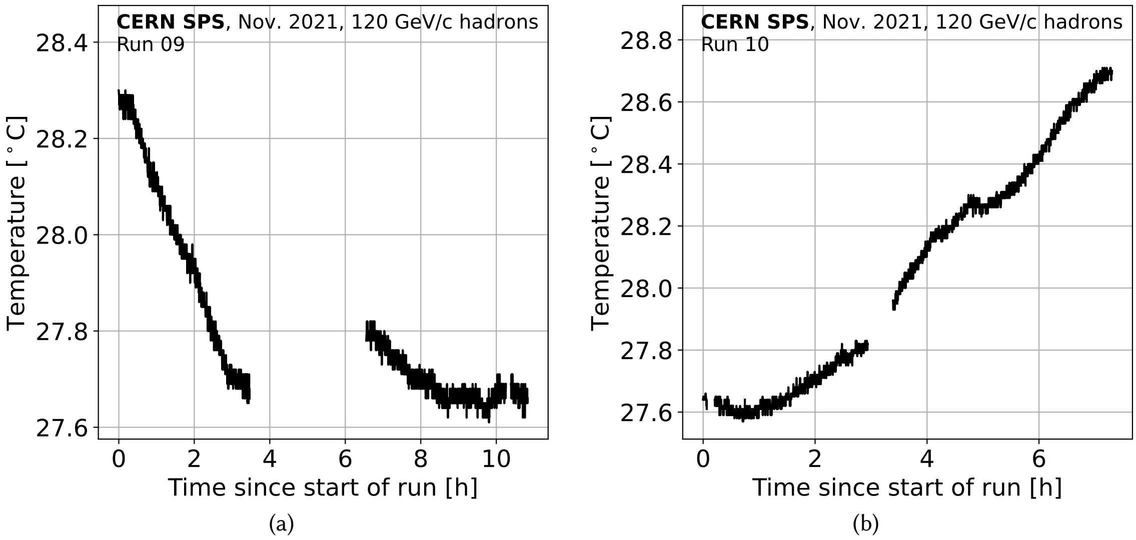


Figure 4.7: Evolution of the temperature measured inside the beam telescope enclosure during two selected data-collection periods. The run numbers refer to Table 4.2.

efficiencies above 99.5% [129].

4.3.1.2 SPAD planes

The DUTs are mounted on carrier PCBs, allowing them to be precisely positioned in the space-constrained telescope environment. One plane is attached to a motorised linear movement stage to adjust the relative position of the two devices with sub-micron precision [131]. The SPAD reverse-bias voltage V_{supply} is provided by a remote-controlled power supply [132] augmented by an additional ripple-rejection circuit. The front-end amplifiers are located outside the telescope enclosure and are connected to the DUTs through coaxial cables with a characteristic impedance of $Z_0 = 50\,\Omega$.

A temperature sensor is placed close to the position of the DUTs inside the telescope to monitor environmental conditions. Figure 4.7 shows the temperature evolution during two typical data-taking runs in nominal operating conditions, indicating that the temperature is stable to about $1°$C. For each run, the bias voltage is set to a fixed, predetermined value. No temperature feedback is attempted given the small temperature coefficient of V_{br} measured in Section 4.1.

4.3.1.3 Trigger scintillators

Three plastic scintillators, read out by photomultiplier tubes [133], are used to preselect events in which beam particles are present in a region of interest centred on the position of the DUTs. Two scintillators, SC1 and SC2, are located at the extreme ends of the telescope and operated in coincidence [134]. The third scintillator, SCV, contains a hole with a diameter of 5 mm and implements a veto. It is positioned such that both DUTs are aligned with the hole, thus defining the region that is enhanced by the scintillator trigger.

The scintillator trigger signal T_S is computed as

$$T_S = (\mathrm{SC1}\uparrow) \wedge (\mathrm{SC2}\uparrow) \wedge \left(\overline{\mathrm{SCV}\uparrow}\right) \wedge \mathrm{RDY}, \tag{4.1}$$

where the $\uparrow$ symbol denotes the presence of a rising edge on the respective signal. The level-sensitive signal RDY is asserted only if all tracking planes are ready to acquire data. T_S is generated by a field-programmable gate array (FPGA) [135], clocked at $f_c = 100$ MHz. The coincidence time window T_c is implicit in Eq. 4.1 and fixed by the clock period of the FPGA to $T_c = 1/f_c = 10$ ns. The time resolution of the trigger signal T_S is limited by the coincidence time T_c to around $T_c/\sqrt{12} \approx 3$ ns (RMS).

4.3.2 Global trigger logic

The active area of each DUT is smaller than the area enhanced by the scintillator trigger by a factor of more than 300 (cf. Table 4.1). To improve the efficiency of selecting events in which a beam particle traverses one, or both, of the DUTs, the readout of the tracking detectors is requested only upon assertion of the global trigger signal T_G. It combines T_S with hit information extracted from the analogue SPAD signals S1 and S2, defined by discriminating the analogue signals delivered by the front-end amplifiers against fixed voltage thresholds. Two different global trigger configurations, $T_{G,1}$ and $T_{G,2}$, are used,

$$T_{G,1} = (T_S\uparrow) \wedge (\mathrm{S1}\uparrow)\,, \qquad\qquad T_{G,2} = (T_S\uparrow) \wedge (\mathrm{S2}\uparrow)\,. \tag{4.2}$$

For T_G to be asserted, the scintillator trigger signal T_S must arrive within a window of 100 ns after the signals S1 or S2. This accounts for the relative delay between T_S and the SPAD signals, caused by the longer scintillator rise time and the latency introduced by the scintillator trigger FPGA.

4.3.3 SPAD signal acquisition

The amplified SPAD signals are digitised at a sampling rate of 80 GS/s by an oscilloscope with an analogue bandwidth of 36 GHz [136]. Upon assertion of the global trigger signal, a waveform buffer covering a period of 400 ns immediately preceding the trigger signal is recorded.

Run number	Particles / extraction	SPAD bias voltage	Global trigger
09	$\approx 5 \cdot 10^4$	23 V	$T_{G,1}, T_{G,2}$
10	$\approx 5 \cdot 10^4$	23 V, 23.5 V, 24 V, 24.5 V, 25 V	$T_{G,1}, T_{G,2}$
14	$\approx 5 \cdot 10^6$	23 V	$T_{G,1}, T_{G,2}$

Table 4.2: Overview of data taking conditions for the runs discussed here. The runs are ordered chronologically.

4.3.4 Online data processing and slow control

Data from the tracking detectors are read out and decoded with the EUDAQ2 framework [137] running on a Raspberry Pi single-board computer [138]. For each event, the waveform recorded by the oscilloscope is associated with the data provided by the tracking detectors.

4.4 Data collection and beam conditions

The experimental setup is placed in the H6 beam line in the EHN1 area at the CERN Super Proton Synchrotron (SPS). Protons with an energy of 450 GeV are extracted from the SPS ring over the course of 4.8 s [139] and directed onto a beryllium target. A secondary mixed hadron beam with a momentum of 120 GeV/c is selected and directed to the experimental area. It mainly consists of protons and positive pions. The particle flux is controlled by means of collimators located at different positions along the beam line.

Data were taken over a period of three days in November 2021. Different beam conditions and SPAD bias voltages were explored, as summarised in Table 4.2. The global trigger logic was switched between the two conditions in Eq. 4.2 in blocks of 1000 events. Unless noted otherwise, data collected with both triggers are used in the analysis.

SPAD/P96/1 was moved in the transverse plane until it was aligned with SPAD/P96/2. The absolute positions of both DUTs were determined from hits recorded in the closest tracking planes as well as the intercepts of tracks reconstructed in the coarsely aligned telescope, following the methodology outlined in Section 4.5.

4.5 Offline event reconstruction and signal processing

Offline event reconstruction was performed using the Corryvreckan framework [140, 141]. The tracking planes were aligned using the Millepede algorithm [142] as implemented in Corryvreckan.

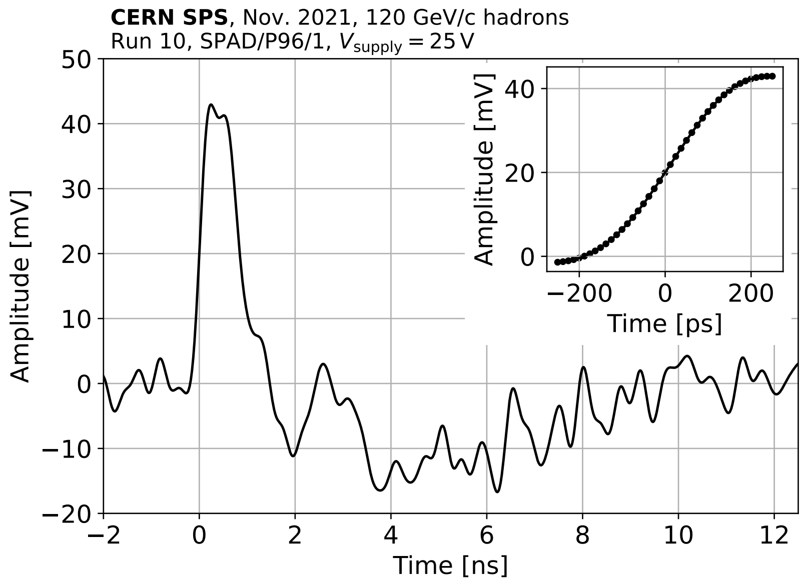


Figure 4.8: Typical SPAD signal, filtered as described in the main text. The inset shows a magnified version of the leading edge with a rise time of approximately 300 ps. The points indicate the individual samples taken during digitisation.

Following alignment, tracks were reconstructed using a straight-track model. Exactly four hits, one per tracking plane, are required for each track.

The bandwidth of the recorded analogue waveforms was limited to 3 GHz by means of a finite-impulse-response low-pass filter with 1000 coefficients. The filter coefficients are constructed with the least-squares design method [143] and multiplied by a Blackman-Harris window function [144] to improve stopband attenuation. This reduces the contribution from sampling noise in frequency domains in which the spectral density of the SPAD signal is low. Figure 4.8 shows a typical signal obtained after these processing steps. The inverting nature of the front-end amplifier leads to a positive signal polarity.

Figure 4.9 shows the transverse intercepts of a sample of reconstructed tracks. The beam profile, the region enhanced by the scintillator veto, and the position of the DUTs are clearly visible.

4.6 Results

The findings summarised in the following give an overview of the behaviour of both DUTs under a variety of different operating conditions, with a focus on those aspects that are relevant to assess their suitability as practical charged-particle detectors.

Section 4.6.1 completes the basic characterisation of the DUTs and measures the breakdown

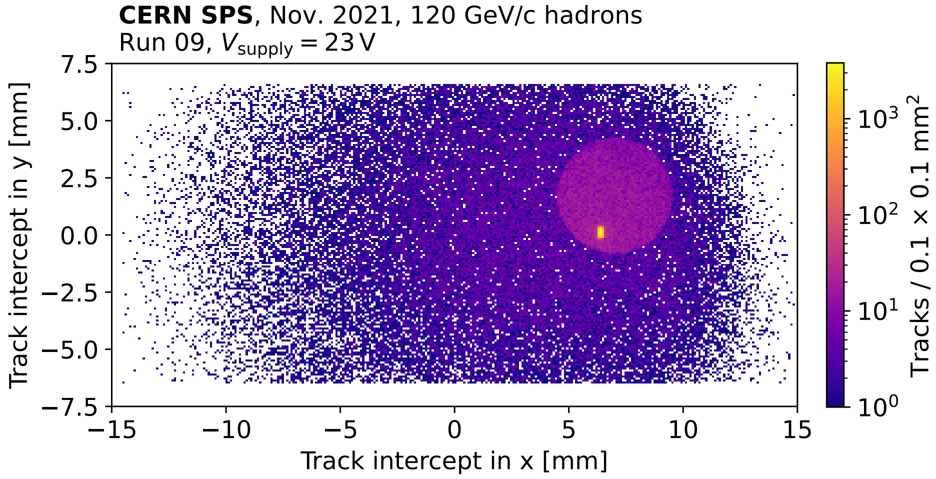


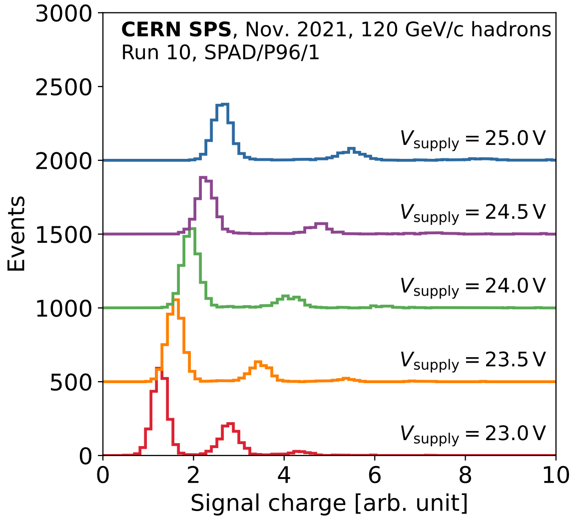
Figure 4.9: Intercepts of reconstructed tracks in the transverse plane for a z-coordinate that is equidistant between the two DUTs, collected with the global trigger. The recorded event rate is enhanced by the scintillator trigger in the circular region visible in the top-right corner. A large fraction of the recorded tracks cross at least one of the DUTs, visible at $(x, y) \approx (6.5\,\mathrm{mm}, 0\,\mathrm{mm})$.

voltage for the environmental conditions in the beam area. An *in-situ* measurement of the rate of dark counts is shown in Section 4.6.2, and consequences for the radiation tolerance of the studied devices are discussed. As Chapter 3 shows, transport processes such as drift and diffusion of charge carriers can have significant implications for the timing characteristics of the device. Section 4.6.3 confirms the importance of these effects and points out striking differences in the behaviour of the two DUTs. Finally, Section 4.6.4 builds on these results and measures the time resolution.

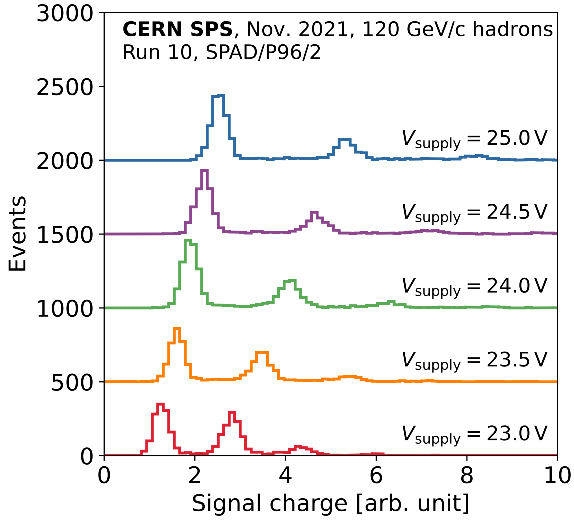
4.6.1 Breakdown voltage

The SPAD breakdown voltage is extracted from the linear dependence of the signal charge on the operating voltage derived in Section 2.6. (Ref. [123] refers to this quantity as the "gain breakdown voltage" and shows that it may differ slightly from the "current breakdown voltage" obtained from the dependence of the reverse bias current on the bias voltage. This distinction is not important for what follows.)

For this analysis, only signals which occur within the coincidence time window of the global trigger are considered (cf. Section 4.3.2). To maximise the size of the available data set, events collected with both triggers, $T_{G,1}$ and $T_{G,2}$, are used. The time integral over the positive pulse of the acquired voltage waveform is used to determine the number of fired pixels. This quantity is loosely referred to as the "signal charge" in the following. It is found to be a more robust estimator for the number



(a)



(b)

Figure 4.10: Distribution of the signal charge at different operating voltages for beam-induced events, for SPAD/P96/1 in (a) and for SPAD/P96/2 in (b). The histograms for different voltages are vertically offset for visual clarity.

of fired microcells than the maximum amplitude of the signal, which is subject to electronic noise. (As the signal is capacitively coupled, the signal charge in the strict sense, i.e. the time integral over the full waveform, vanishes.) The chosen integration window has a width of 4 ns for SPAD/P96/1 and is centred on the rising edge of the signal. For SPAD/P96/2, a slightly longer window of 4.5 ns is used to ensure that the entire positive pulse is captured. Differences in the signal shapes between the devices are discussed in Section 4.6.4.3.

Figure 4.10 shows the distribution of this observable for various operating voltages. The discrete nature of the spectrum is apparent. The predominant part of the population is concentrated in events with ≤ 3 active SPAD cells, with only a single active pixel in most events. The mean signal charge for each operating point is determined from the location parameter of a normal distribution fitted to the primary peak. Figure 4.11 highlights the extracted linear relationship between the mean signal charge and the operating voltage.

The breakdown voltage is obtained by extrapolating this linear dependence towards zero signal charge, i.e. zero gain. The values obtained are listed in Figure 4.11 and found to be compatible with the result reported in Section 4.1. The statistical uncertainty in this measurement is around 10 mV. Contributions from systematic uncertainties are estimated by removing individual data points from the linear fit. This leads to variations in the extracted breakdown voltage of the order of 50 mV, which is similar to the impact expected from variations of the ambient temperature during data

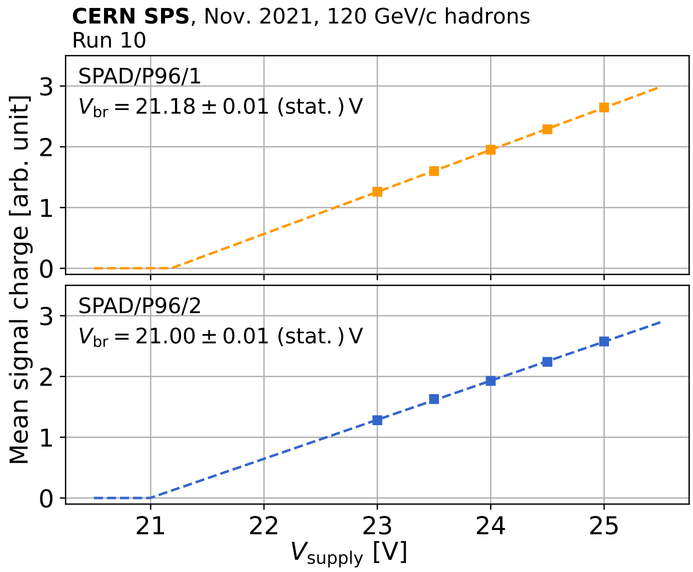


Figure 4.11: Mean signal charge produced by a single active SPAD cell at different operating voltages, shown in the top pane for SPAD/P96/1 and in the bottom pane for SPAD/P96/2. The dashed line results from a linear fit to all data points. Its intercept with the horizontal axis corresponds to the breakdown voltage V_{br}.

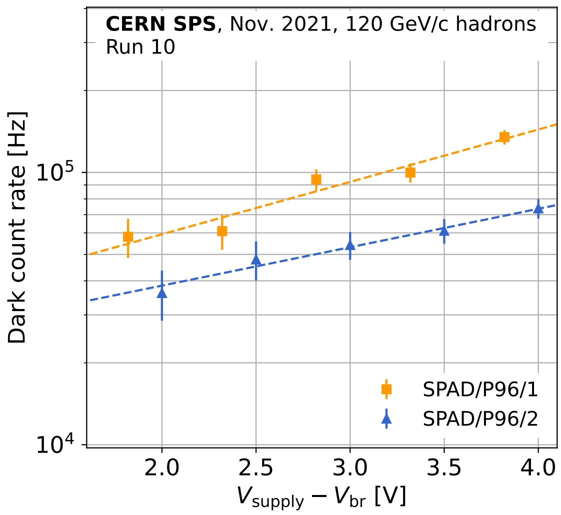
taking (cf. Figure 4.7).

4.6.2 Dark-count rate

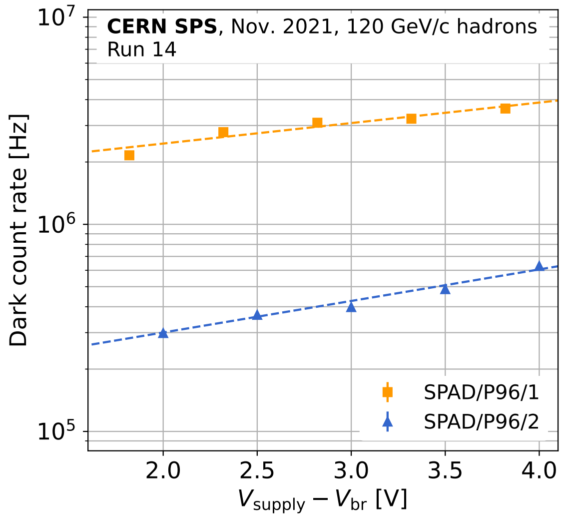
The dark-count rate (DCR) is estimated from a sample of events for which the trigger is not due to a passing beam particle. Beam-induced events are efficiently removed by requiring no reconstructed tracks in a region of $600\,\mu\text{m} \times 600\,\mu\text{m}$ centred on the position of the DUTs, and considering only the portion of the recorded waveforms that lies outside the coincidence window of the global trigger, i.e. during a period of 300 ns at the beginning of the acquisition buffer.

Dark counts may be triggered by the thermal excitation of charge carriers or by their tunnelling into the conduction band [102, 145, 146]. These processes occur independently at different positions in the semiconductor. The time Δt between two subsequent dark counts thus follows an exponential distribution, $p(\Delta t; \text{DCR}) = \text{DCR} \cdot \exp\left(-\Delta t \cdot \text{DCR}\right)$. At the lowest observed rates, the expected number of dark-count events contained in each acquired waveform buffer is significantly less than unity. A direct measurement of the DCR from an exponential fit to the observed distribution of Δt is thus not possible [102]. The DCR was instead measured as follows.

An estimate DCR_{est} was obtained by dividing the number of observed dark counts by the total time period spanning multiple events. To remove contamination from afterpulsing or (delayed) cross



(a)



(b)

Figure 4.12: Dark-count rates as function of the excess voltage $V_{\text{ex}} = V_{\text{supply}} - V_{\text{br}}$, corrected according to Eq. 4.3, for run 10 in (a) and run 14 in (b).

talk, two subsequent signals were counted as distinct if they were separated by a period of more than $\Delta t_{\text{min}} = 20$ ns. (For operating conditions that produce sufficiently high dark-count rates, it has been explicitly checked that ΔT is distributed exponentially for $\Delta T > \Delta t_{\text{min}}$, i.e. that no significant contamination remains. This measurement cannot discriminate between genuine dark counts and delayed afterpulsing or delayed crosstalk that occurs on time scales of the buffer length of 300 ns.)

The estimate DCR_{est} captures only a fraction $\int_{\Delta t_{\text{min}}}^{\infty} d\Delta t\, p(\Delta t; \text{DCR}) = \exp\left(-\Delta t_{\text{min}} \cdot \text{DCR}\right) < 1$ of all genuine dark counts. It is thus related to the true DCR through the relation

$$\text{DCR}_{\text{est}} = \text{DCR} \cdot \exp\left(-\Delta t_{\text{min}} \cdot \text{DCR}\right), \tag{4.3}$$

which may be solved numerically and thus allows DCR_{est} to be corrected. (This equation has two solutions for the corrected value DCR, one much larger than the other. It is easy in practice to eliminate the unphysical solution through a manual inspection of the waveform buffer.)

Figure 4.12a shows the measured DCR measured for run 10, which involved comparatively low beam intensities. It exhibits a characteristic exponential enhancement with the electric field [146]. SPAD/P96/2 shows a DCR of 36–74 kHz, depending on the operating voltage. For identical voltage, the DCR is higher for SPAD/P96/1 and ranges between 58–135 kHz. As this measurement exploits the trigger inefficiency (which is low if the product of dark-count rate and coincidence time window is small), statistical uncertainties are sizeable and of the order of 10–15%.

Figure 4.12b shows the DCR integrated over run 14, which was characterised by significantly higher beam intensities (cf. Table 4.2). The observed DCR is significantly increased compared to Figure 4.12a, reaching an average value of 2–4 MHz for SPAD/P96/1. A further, smaller, increase of about 30% was observed over the course of the run. SPAD/P96/2 continues to show lower DCRs, now between 300–600 kHz, stable to within about 10% during the run. (Even at these extreme rates, the correction introduced by Eq. 4.3 is subleading and attains a value of around 7% at most.)

Exposure to energetic hadrons can introduce bulk lattice defects at intermediate energy levels between the valence and conduction bands. This enhances the generation of free charge carriers through thermal excitation and tunnelling [147], and an increase in the DCR is thus a typical signature of radiation damage. Figure 4.12 therefore implies a substantial radiation sensitivity of both DUTs. Since stable data taking resumed only approximately three hours after beam conditions were changed, it was impossible to quantify the exact value of the integrated flux the devices have accumulated.

It is also unclear why the two devices respond differently to an identical radiation environment. Inconsistent behaviour is also seen in other observables, outlined below, which is suggestive of differences in the electric field configuration between the DUTs.

4.6.3 Charge collection and detector response

It is important to understand the dependence of the detector response on the position of the signal-inducing charged particle. This concerns in particular the time of the SPAD signal relative to the time of the track, as well as the number of triggered SPAD cells. The former can help to identify the processes that transport the deposited charge to the avalanche region and the latter contains information about the spatial configuration of the electric field that collects and guides this charge. Both aspects are considered in the following.

In describing the hit position of the beam particle relative to the SPAD array, three functionally distinct regions are identified. First, the "diode" area, which defines the approximate lateral extent of the avalanche region (cf. Figure 4.1a). Second, the area of the pixel which is *not* covered by the diode and where the resistors R_c and R_q are located, referred to as the pixel "periphery" in the following. Third, the area outside the nominally sensitive part of the array, labeled "inactive". It includes the dummy pixels but also the surrounding silicon area within a certain region of interest as defined below.

To allow events to be assigned unambiguously to one of these categories, events are required to

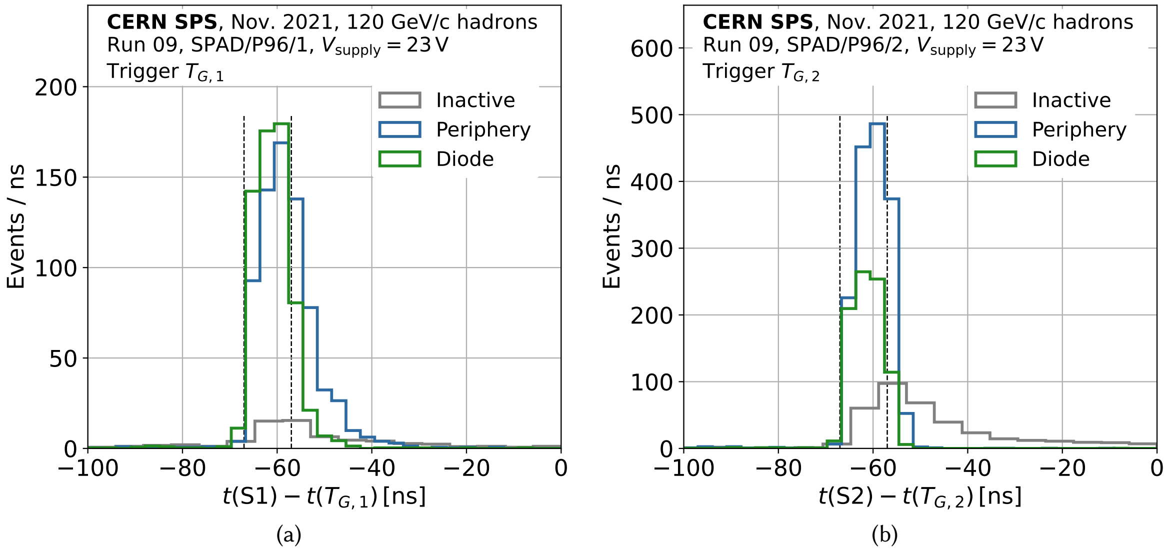


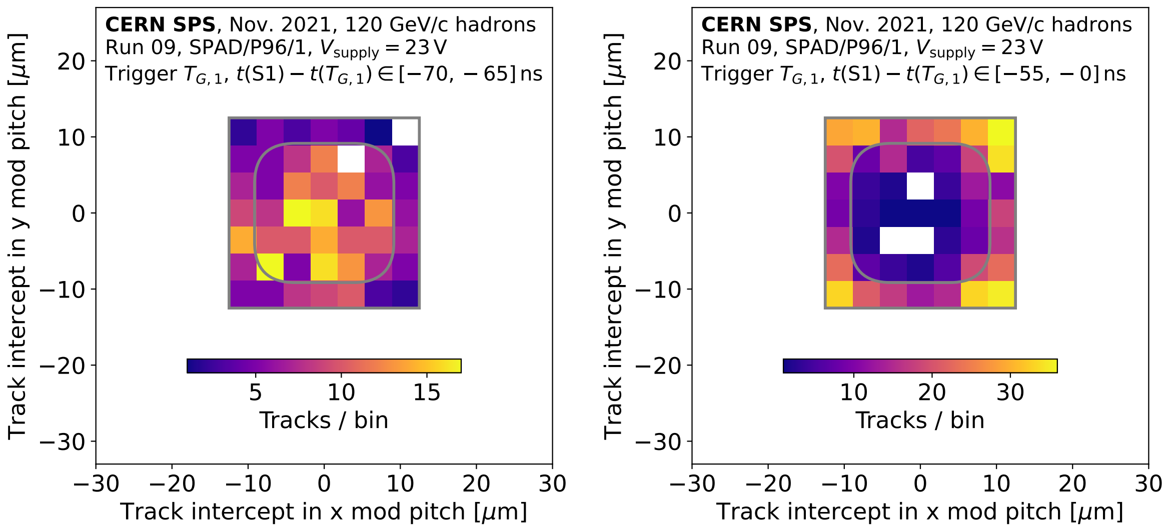
Figure 4.13: Distributions of the signal arrival times $t(\mathrm{S1}) - t(T_{G,1})$ in (a) and $t(\mathrm{S2}) - t(T_{G,2})$ in (b), shown as a function of the position of the primary charged deposited by the charged particle. The dashed vertical lines correspond to one clock period T_c of the scintillator trigger FPGA.

contain exactly one track in a region of $600\,\mu\mathrm{m} \times 600\,\mu\mathrm{m}$ centred on the position of the DUTs. The fraction of events with two or more tracks in this region is very small and only a few permille of all events are rejected by this cut. Only self-triggered events are considered in the study of each device, i.e. using the trigger selection $T_{G,1}$ for SPAD/P96/1 and $T_{G,2}$ for SPAD/P96/2. This avoids any bias introduced by the (potentially spatially inhomogeneous) trigger efficiency of the other DUT.

4.6.3.1 Signal arrival time

The time $t(T_G)$ at which the global trigger signal is asserted defines an absolute reference with respect to which the arrival time of the SPAD signals, $t(\mathrm{S1})$ and $t(\mathrm{S2})$, are measured. Figure 4.13 shows the distributions of the thus defined arrival times. (As mentioned in Section 4.3.1.3, the scintillator trigger signal T_S, and therefore also T_G, has a time resolution of about 3 ns, which is longer than the duration of a typical SPAD signal. No particular care is required in the exact definition of the times $t(\mathrm{S1})$ and $t(\mathrm{S2})$. For the remainder of this section, they correspond to the times at which the respective signals attain 50% of their maximal amplitudes.)

As shown in Figure 4.13, a prompt detector response is caused by a primary charge deposit within the area of the diode, close to the avalanche region, for both devices. Hits in the pixel periphery require charges to be transported to the high-field regions and thus delay the onset of the avalanche. This is illustrated by Figures 4.14a and 4.14b, which show the track intercepts within a single pixel



(a) (b)

Figure 4.14: Track intercepts in bins of the signal arrival time $t(\text{S1}) - t(T_{G,1})$ for SPAD/P96/1. Tracks across the entire active area of the device are considered and shown in local coordinates where the centre of each pixel is mapped to the origin. The resulting pixel outline and the diode are shown in grey (cf. Figure 4.1a). Prompt signals are shown in (a) and events in the tail of the arrival distribution in (b).

for prompt and delayed signals, respectively. A substantial fraction of events is triggered by charges deposited outside the active area of the array, particularly for SPAD/P96/2. This class is mainly composed of events triggered by hits in the dummy pixels but also includes a small number of cases where charge transport over distances of up to 50 μm is involved. The delay in the formation of the avalanche is even more pronounced in this regime and can amount to a few tens of nanoseconds.

Figure 4.13 also shows the clock period of the scintillator trigger FPGA, which determines the variance of the observed arrival-time distribution in the limit where the jitter in the SPAD signals is negligible compared to that in $t(T_G)$. For SPAD/P96/1, the arrival-time distribution shows pronounced tails beyond T_c, demonstrating that charge transport occurs on time scales of the order of $\tau_t \approx 10\,\text{ns}$ for an appreciable fraction of events. The time τ_t does not depend significantly on the operating voltage of the device, suggesting that charge-carrier diffusion contributes significantly. Using the value of the longitudinal diffusion constant D from Section 3.3.1, the characteristic diffusion length scale evaluates to $\Delta x \sim \sqrt{2D\tau_t} \approx 8\,\mu\text{m}$, which is comparable to the thickness of the epitaxial layer and the radius of the diode. The fraction of events in the diffusive tail reduces at high voltages, compatible with an enhanced drift contribution.

Charge collection appears to proceed on much shorter time scales for SPAD/P96/2, suggestive of a higher electric field throughout the bulk of the device and a diffusion contribution that is sup-

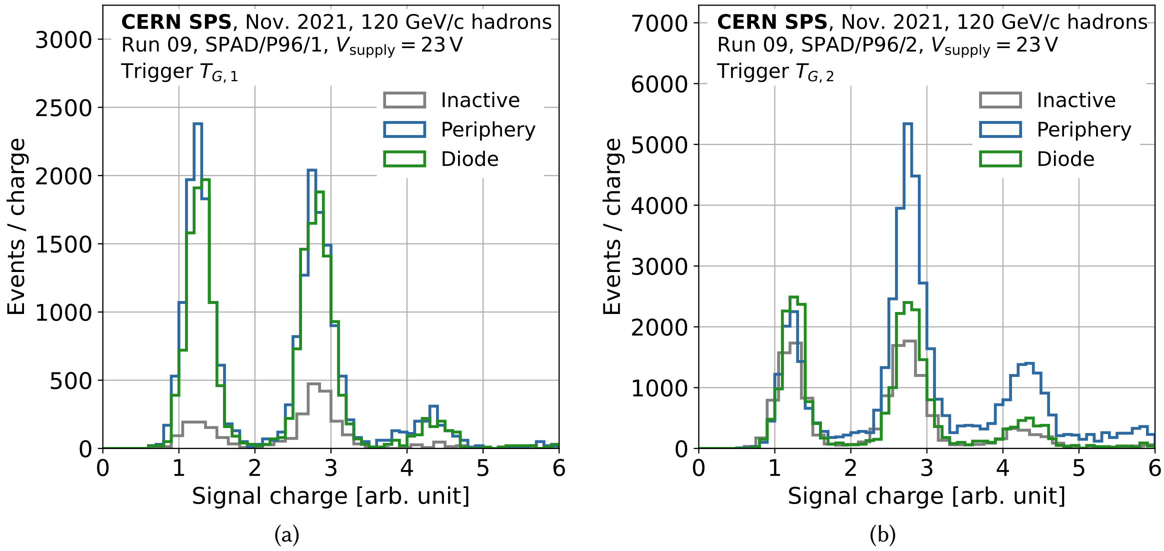


Figure 4.15: Distributions of the signal charge as a function of the position of the primary charged deposited by the charged particle, for SPAD/P96/1 in (a) and for SPAD/P96/2 in (b).

pressed relative to charge-carrier drift. The fraction of events triggered by charges deposited outside the active part of this array is significantly higher compared to SPAD/P96/1. It increases from about 23% for a bias voltage of 23 V to around 42% at 25 V, consistent with a drift field that extends beyond the array itself.

4.6.3.2 Pixel multiplicity

Figure 4.15 shows the dependence of the signal charge distribution on the hit position for an operating voltage of 23 V. The fraction of events for which more than one SPAD cell is triggered, referred to as the "crosstalk", is around 60% at this operating point. This is significantly higher than the number expected from the intrinsic coupling of the individual pixels; the cross talk for dark-count events is only 7% at this voltage. This is consistent with the observations made in Ref. [122]. The crosstalk reduces to around 25–35% at $V_{\text{supply}} = 25\,\text{V}$ but remains above the level measured for dark counts at all voltages. This implies that charge is predominantly collected by a single pixel at high voltages.

For SPAD/P96/1 in Figure 4.15a, the signal charge spectrum does not significantly depend on the hit position. If an underlying dependence exists, it might be partially washed out by carrier diffusion during charge collection. The yield of events triggered by hits in the diode and the pixel periphery is comparable, indicating that both areas are sensitive to charged particles and that the overall efficiency of the device is not limited by its relatively low geometric fill factor (cf. Table 4.1). A similar effect has also been observed for commercial devices with higher fill factors in Ref. [122].

(This work also quantified the detection efficiency through a standard "tag and probe" measurement. A robust and systematics-free application of this technique is not possible with the experimental setup used here, where dark counts are present in both the "tag" and the "probe" device, and is therefore not attempted.)

In the case of SPAD/P96/2 in Figure 4.15b, the event-yield ratio is skewed considerably at $V_{\text{supply}} = 23$ V, where the detection efficiency for a hit in the periphery is about twice that for a particle crossing the diode area. This suggests that the drift-field configuration is such that it impedes the collection of charges deposited in the detector volume underneath the avalanche region. Most such events also provoke avalanche breakdown in multiple pixels, demonstrating significant charge sharing at low voltages. For higher operating voltages, the charge-collection efficiency in the diode area recovers and the detector response becomes uniform again already for $V_{\text{supply}} = 24$ V.

The inclusive trigger efficiency remains lower for SPAD/P96/1 for all operating conditions due to its higher dark-count rate relative to SPAD/P96/2, leading to a lower number of recorded beam-induced events.

4.6.4 Timing characteristics and time resolution

The timing properties of the DUTs are investigated based on a sample of events for which both devices produce a beam-induced signal. Each signal was assigned a timestamp. In the following, the signal timestamps $t(\text{S1})$ and $t(\text{S2})$ were defined to correspond to the times at which the analogue signals S1 and S2 cross fixed voltage thresholds $V_{\text{th},1}$ and $V_{\text{th},2}$. The individual threshold-crossing times cannot themselves be accurately measured, but only the time difference $\Delta t_{12} = t(\text{S1}) - t(\text{S2})$ is experimentally accessible with good precision.

The observable standard deviation $\sigma_{\text{th},12}$ is related to the fluctuations in the threshold-crossing times of the two DUTs, $\sigma_{\text{th},1}$ and $\sigma_{\text{th},2}$, as follows

$$\sigma_{\text{th},12} = \sqrt{\text{var}\left[\Delta t_{12}\right]} = \sigma_{\text{th},1} \oplus \sigma_{\text{th},2},$$

where the $\oplus$ symbol denotes addition in quadrature. From a measurement of $\sigma_{\text{th},12}$, various contributions to $\sigma_{\text{th},1}$ and $\sigma_{\text{th},2}$ may be partially reconstructed, as demonstrated in the following.

4.6.4.1 Timing contributions

A multitude of different physical processes contribute to σ_{th}, which may be decomposed as

$$\sigma_{\text{th}}(V_{\text{supply}}, V_{\text{th}}, \mathbf{x}) \approx \sigma_{\text{SPAD}}(V_{\text{supply}}) \oplus \sigma_{\text{transp}}(\mathbf{x}, V_{\text{supply}}) \oplus \sigma_{\text{noise}}(V_{\text{th}}, V_{\text{supply}}). \tag{4.4}$$

In this expression, σ_{SPAD} represents the contribution from the discharge of an individual SPAD pixel. It includes avalanche fluctuations (σ_{av}) and the stochasticity in the primary charge deposit (σ_{pos}), $\sigma_{\text{SPAD}}(V_{\text{supply}}) = \sigma_{\text{av}}(V_{\text{supply}}) \oplus \sigma_{\text{pos}}(V_{\text{supply}})$, as discussed in Section 3.4.3. Chapters 2 and 3 show that both components depend primarily on the electric field in the multiplication region, or equivalently, on the bias voltage V_{supply}.

Transport processes such as drift and diffusion determine the arrival time of the primary charge carriers in the avalanche region and generate the term $\sigma_{\text{transp}}(\mathbf{x}, V_{\text{supply}})$. As demonstrated in Section 4.6.3.1, it depends on the position $\mathbf{x}$ of the charge deposit relative to the pixel structure, as well as on the global field configuration of the device, and thus on V_{supply}. In the following, the position-dependence is studied by distinguishing between $\sigma^{\text{diode}}_{\text{transp}}$ and $\sigma^{\text{periph}}_{\text{transp}}$, corresponding to a particle hit in the central diode region and the pixel periphery, respectively. Differences in the response between different SPADs in the array are neglected.

Finally, $\sigma_{\text{noise}}(V_{\text{th}}, V_{\text{supply}})$ represents the jitter introduced by the electronic noise present in the recorded signal. This includes the intrinsic thermal noise [148] produced by the DUT and any additional noise injected during the amplification and digitisation process. A noise signal $n(t)$ has a standard deviation of $\sigma_n = \sqrt{\langle n^2(t)\rangle_t - \langle n(t)\rangle_t^2}$, where the expectation values are taken with respect to time. The quantity σ_n is related to the contribution σ_{noise} through the slew rate $d\text{S}/dt$ of the detector signal at the moment of threshold-crossing, i.e.

$$\sigma_{\text{noise}} = \frac{\sigma_n}{\left.\frac{d\text{S}}{dt}\right|_{V_{\text{th}}}}, \tag{4.5}$$

where S is the underlying detector signal in the absence of noise.

Eq. 4.4 implicitly assumes that only a single SPAD cell is triggered. The signal pile-up caused by multiple pixels firing simultaneously leads to an additional term in the decomposition of σ_{th}. It can, in principle, be eliminated by slewing corrections [40], which correct the signal timestamps in each event so that the *average* signal time becomes independent of the signal amplitude. This correction is applied separately for each DUT. To simplify the following discussion, the pile-up contribution

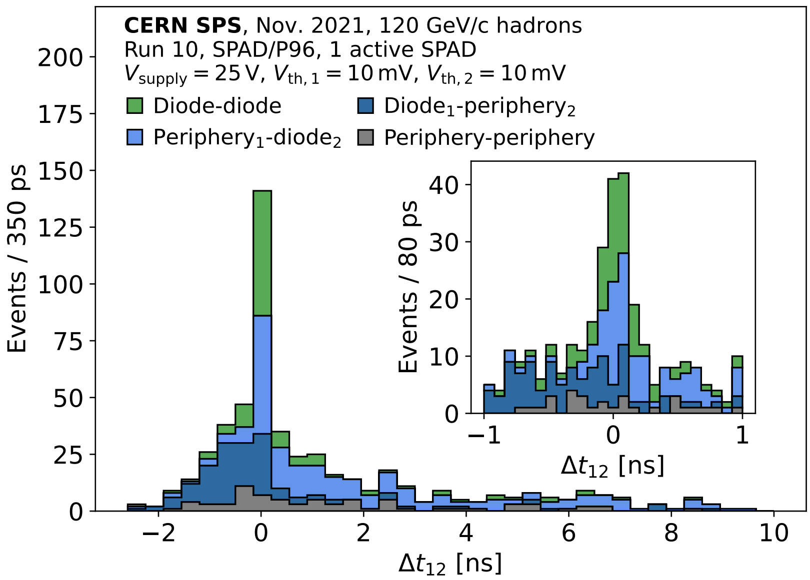


Figure 4.16: Distribution of the time difference Δt_{12}, depending on the hit position. The green component collects events where the charge deposit occurred in the diode region in both devices ("Diode-diode"). The dark blue histogram ("Diode$_1$-periphery$_2$") corresponds to a hit in the diode region of SPAD/P96/1 and the pixel periphery of SPAD/P96/2; and the roles are reversed for the light blue contribution ("Periphery$_1$-diode$_2$). In a small number of cases, the primary charge is deposited in the periphery of both SPAD arrays ("Periphery-periphery"), shown in grey.

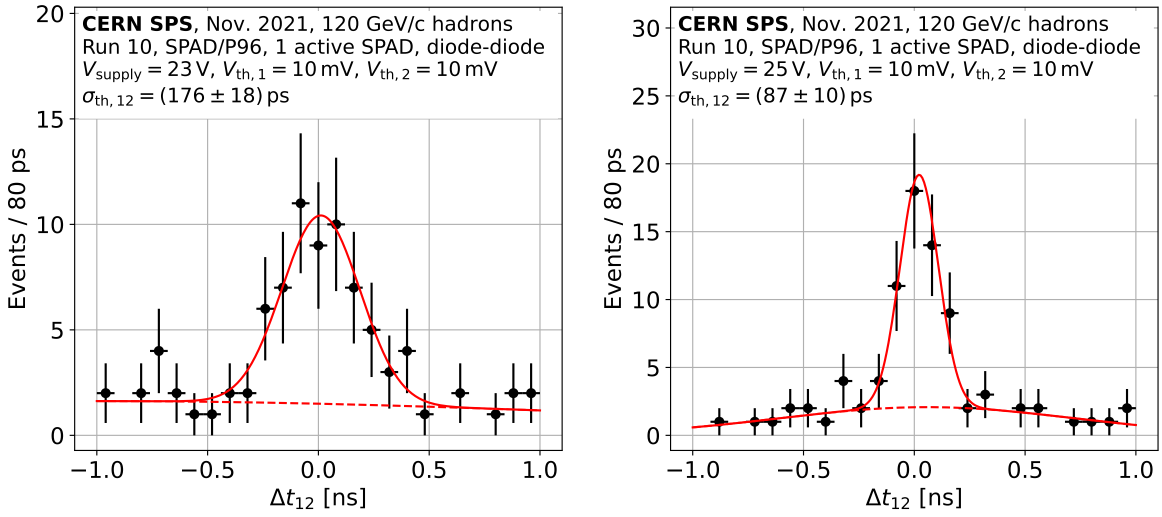
to the time resolution (and the need for slewing corrections) is removed by considering only events with a single active SPAD.

4.6.4.2 Timing hierarchy and position dependence

Figure 4.16 shows the observed distribution of Δt_{12} as a function of the hit position in the two DUTs. A clear hierarchy between the different cases is visible, reflecting the relative importance of σ_{SPAD} and σ_{transp}.

The distribution is very narrow in the case in which charges are deposited exclusively in the diode regions, allowing the sum $\sigma_{\text{SPAD}} \oplus \sigma_{\text{transp}}^{\text{diode}} \oplus \sigma_{\text{noise}}$ to be constrained to not more than a few tens of picoseconds. This case is studied in more detail below in Section 4.6.4.3.

A charge deposit in the pixel periphery significantly broadens the distribution and degrades the achievable time resolution. In the case of the diffusion-dominated device SPAD/P96/1, $\sigma_{\text{transp}}^{\text{periph}}$ involves time scales of the order of 10 ns, as already anticipated in Section 4.6.3.1. For SPAD/P96/2, the contribution $\sigma_{\text{transp}}^{\text{periph}}$ is significantly faster and of the order of 1 ns.



(a) (b)

Figure 4.17: Modelling of the distribution of Δt_{12} by a mixture of two normal distributions for the extraction of $\sigma_{\text{th},12}$, for a bias voltage of 23 V (a) and 25 V (b). The dashed line represents the fitted background component and the solid line is the full model prediction.

4.6.4.3 SPAD time resolution and electronic noise

The sum $\sigma_{\text{SPAD}} \oplus \sigma^{\text{diode}}_{\text{transp}} \oplus \sigma_{\text{noise}}$ limits the ultimate time resolution of the device. It is measured through an unbinned maximum-likelihood fit to Δt_{12}, which only includes events with charge deposits in the diode regions (the "diode-diode" contribution in Figure 4.16). A mixture model of two normal distributions is used to describe the data. One component models the narrow peak around $\Delta t_{12} = 0$, and the other captures a small but wider background. This background is partly due to a residual contamination of events which are misclassified due to the finite tracking resolution of the telescope, as well as boundary effects caused by charges deposited close to the edge of the diode.

Figure 4.17 shows the resulting modelling of the data for two different operating voltages and fixed thresholds. The extracted value of $\sigma_{\text{th},12}$ depends on the voltage, ranging from about 180 ps at 23 V to about 90 ps at 25 V. Corrections to account for the time slewing caused by residual variations in the signal amplitude of a single active SPAD were investigated but do not lead to a measurable improvement. No such corrections are therefore included.

The dependence of $\sigma_{\text{th},12}$ on the supply voltage and the thresholds $V_{\text{th},1}$ and $V_{\text{th},2}$ is investigated in more detail in Figure 4.18. One threshold is varied across the leading edge of the corresponding signal while the other threshold is kept fixed at a value of 9 mV. This corresponds to about $3\sigma_n$ and represents the lower end of the usable threshold regime. The extracted value of $\sigma_{\text{th},12}$ is largely independent of $V_{\text{th},1}$, as expected from avalanche statistics (cf. Figure 2.9). The jitter $\sigma_{\text{th},12}$ increases

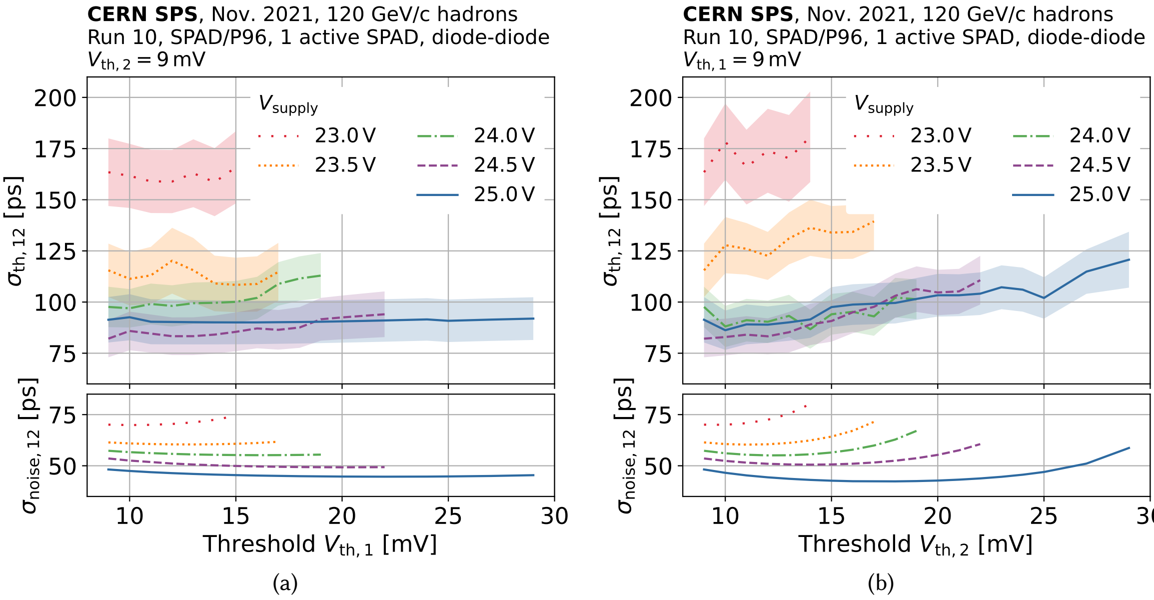
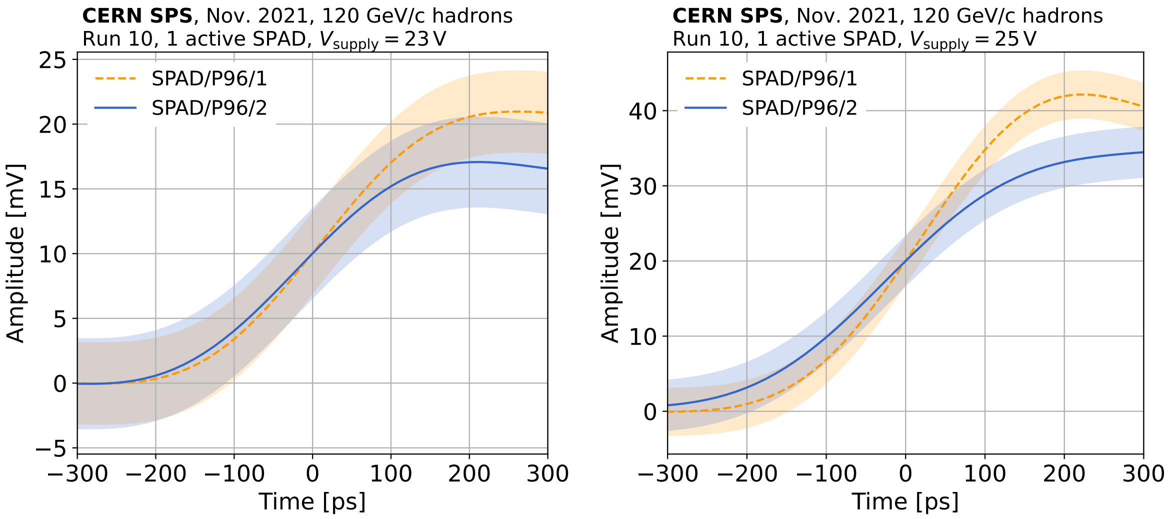


Figure 4.18: Dependence of $\sigma_{\text{th},12}$ on the operating voltage and the thresholds $V_{\text{th},1}$ and $V_{\text{th},2}$ placed on the respective signals. The threshold $V_{\text{th},1}$ is varied in (a) and $V_{\text{th},2}$ is kept fixed, and vice versa in (b). The bottom panels show the total contribution from electronic noise, $\sigma_{\text{noise},12}$, as defined in the main text.

mildly with $V_{\text{th},2}$, although the predominant part of the variance continues to be generated at low thresholds and small avalanches. Increasing the operating voltage from 23 V to 25 V significantly reduces the measured fluctuations.

The dependence of the total noise contribution $\sigma_{\text{noise},12} = \sigma_{\text{noise},1} \oplus \sigma_{\text{noise},2}$ on the thresholds $V_{\text{th},1}$ and $V_{\text{th},2}$ is shown in the bottom panels of Figure 4.18. The quantities $\sigma_{\text{noise},1}$ and $\sigma_{\text{noise},2}$ are estimated through Eq. 4.5, where the average signal is used to approximate the noise-free detector signal S, shown in Figure 4.19 for different operating voltages. SPAD/P96/1 produces signals with a higher peak amplitude and faster slew rate for the same bias voltage. This hints at a higher avalanche growth rate $\gamma_1 v^*$ and a stronger electric field in the gain layer, as already suggested in Section 4.6.2. The standard deviation of the electronic noise, σ_n, is around 3 mV for both devices. The noise contribution σ_{noise} to the time resolution is subleading relative to $\sigma_{\text{SPAD}} \oplus \sigma_{\text{transp}}$.

The single-device time resolution σ_{th} in Eq. 4.4 may be estimated from the measured value of $\sigma_{\text{th},12}$ under the additional assumption that both DUTs behave identically, $\sigma_{\text{th}} \approx \sigma_{\text{th},12}/\sqrt{2}$. As the above discussion shows, this is not fully justified, but $\sigma_{\text{th},12}/\sqrt{2}$ nevertheless continues to be a useful figure of merit for comparison with different devices. Figure 4.20 shows the evolution of this quantity with the operating voltage, and Table 4.3 collects all numerical values. The measured time resolution σ_{th} saturates at around 60 ps for an operating voltage of 24 V, corresponding to an intrinsic



(a) (b)

Figure 4.19: Average detector signals at an operating voltage of 23 V in (a) and 25 V in (b). The shaded band corresponds to the 1σ envelope of the total noise contribution.

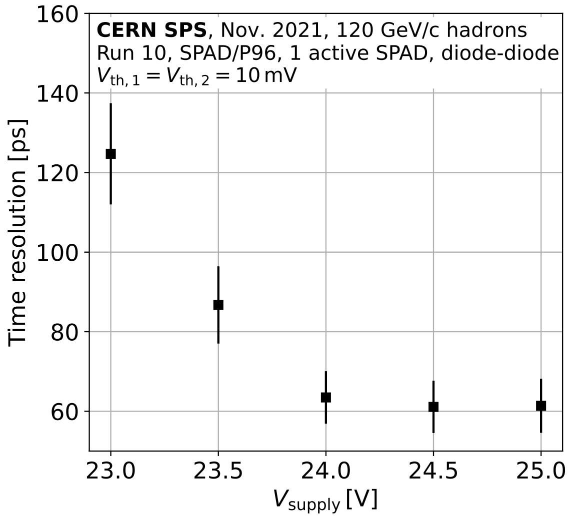


Figure 4.20: Equivalent time resolution σ_{th} of a single device as a function of the operating voltage.

V_{supply}	Measured time resolution $(\sigma_{\text{th},12}/\sqrt{2})$	Estimated noise contribution (σ_{noise})	Estimated detector time resolution $\left(\sigma_{\text{SPAD}} \oplus \sigma_{\text{transp}}^{\text{diode}}\right)$
23 V	125 ± 13 ps	49 ps	115 ps
23.5 V	87 ± 10 ps	43 ps	76 ps
24 V	63 ± 7 ps	39 ps	50 ps
24.5 V	61 ± 7 ps	36 ps	49 ps
25 V	61 ± 7 ps	32 ps	52 ps

Table 4.3: Numerical values for the equivalent time resolution (second column), the estimated contribution from noise (third column), and the detector time resolution with the noise contribution subtracted (fourth column).

detector time resolution $\sigma_{\text{SPAD}} \oplus \sigma_{\text{transp}}^{\text{diode}}$ of about 50 ps. Further increasing the supply voltage to $V_{\text{supply}} = 25\,\text{V}$ does not lead to any additional improvements. The measured values are very similar to the time resolution of commercial SPAD devices and prototypes determined in Ref. [122].

4.7 Summary

A systematic experimental investigation confirms that SPADs can function as fast timing detectors for charged particles. The tested pixel geometry is relevant for the design of future position-resolving detector systems and provides information about the expected performance of such an instrument.

The area sensitive to charged particles is not restricted to the high-field regions but extends to the immediately surrounding pixel periphery and, depending on the operating conditions, even outside the nominal boundaries of the pixel array. The achievable detection efficiency is not limited by the fill factor of the device, but values close to unity can, in principle, be achieved.

The time resolution depends strongly on the position of the charged particle hit relative to the pixel structure. If charges are deposited directly in the avalanche regions, a time resolution of around 60 ps is possible with the tested device. Primary charges deposited in other locations must first be collected through diffusion or drift. These transport processes introduce additional fluctuations and degrade the time resolution to about 1 ns or more, depending on the conditions. Multiple detector layers operating in coincidence, offset by a fraction of the pixel pitch, might be required to provide uniform timing capabilities over an extended area. MC transport simulations taking into account the three-dimensional electric field configuration throughout the device are necessary for a more detailed study and for the design of pixel geometries that minimise these effects.

The two tested devices, nominally identical to one another, show markedly different behaviour in several metrics that indicate differences in the electric field configuration. Reproducibility of performance across multiple devices and manufacturing batches must be considered when scaling up the usage of SPADs for larger-scale experiments. Radiation tolerance also remains a concern. Both devices show a substantial increase in the rate of dark counts after only a moderate but not very well controlled exposure accumulated during the measurement campaign. Further experimental studies are required to investigate the intrinsic limitations of this detector technology and to determine if it can be used in high-rate environments at future accelerator facilities.

PART II

Particle detectors for physics

Chapter 5

An effective theory of fundamental physics

Remarkably, one can understand *something* about the universe without having to comprehend *everything* at once. In a very precise sense, physical phenomena occurring at a certain energy scale can be described almost entirely in terms of concepts that reside at the same scale; knowledge about the fine details of physics at much higher energy scales is not needed. In this way, scattering processes at current collider experiments can be described without any recourse to Planck-scale physics. This chapter shows that this is an inevitable consequence of the fact that the universe is *both* relativistic and quantum mechanical.

Section 5.1 explores these aspects by considering scattering processes between elementary particles. Reasoning directly from first-principles physics, it shows that the primary contribution to the scattering dynamics is generated at the scale of the process itself, and that the allowed form of these dominant interactions is extremely constrained. The same argument reveals that phenomena originating at much higher energy scales may be observable with low-energy experiments. The form of these effects is not constrained by symmetry; they are thus highly non-unique, and parametrically suppressed. Section 5.2 then begins to organise these general conclusions within the formalism of quantum field theory. It recognises the structure of the leading interactions as being generated by the Lagrangian formulation of the Standard Model of particle physics. This model is extended in Section 5.3 to allow subleading contributions originating at disparate energy scales to be effectively parameterised. Together, this establishes the methodology and the notation used in the interpretation of the work discussed in Chapters 7 and 8.

5.1 A reductionist's guide to the universe

The starting point of the following discussions is the set of fundamental principles underlying quantum mechanics and special relativity. The causal structure of relativity posits that interactions occur locally in spacetime, and quantum mechanics demands that these interactions form part of the unitary evolution of the state vector of the world. Taken together, locality and unitarity impose extremely strong constraints which significantly limit the form that physical interactions can take.

What is a particle? A "particle" is a quantum state whose possible attributes are directly determined by the symmetries of spacetime. Translation invariance imparts a continuous identity to a propagating such state, and rotation invariance renders this identity independent of the direction of propagation. More formally, the representation theory of the Poincaré group [149, 150] identifies mass and spin (or helicity, for massless particles) as good quantum numbers of one-particle states. A massive particle with spin s has $2s+1$ spin degrees of freedom, while a massless particle fundamentally exhibits only a single helicity state in the absence of a global parity symmetry.

Scattering amplitudes Information about the interactions between particles is encoded in scattering amplitudes of the form $\mathcal{M} = {}_{\text{out}}\langle\{p_f, \sigma_f\}|\{p_i, \sigma_i\}\rangle_{\text{in}}$, where $\{p_i, \sigma_i\}$ and $\{p_f, \sigma_f\}$ label the sets of momenta and spin quantum numbers of the incoming ("initial state") and outgoing ("final state") particles, respectively. (In the following, units are chosen such that $c = \hbar = 1$.)

Locality has important consequences for the analytic structure of the amplitude. This is apparent already in the simple case of the four-particle amplitude illustrated in Figure 5.1 for the $2 \to 2$ scattering process $1+2 \to 3+4$. If the external momenta p_1 and p_2 are such that the production of an intermediate one-particle state I becomes possible, $s = (p_1+p_2)^2 = m_I^2$, the four-particle amplitude must develop a pole [151]. The residue of this pole must be the product of the two corresponding three-particle amplitudes $1+2 \to I$ and $I \to 3+4$. The internal quantum numbers of I are summed over. Analogous poles exist in the t- and u-channels, where poles are located at $(p_1+p_3)^2 = m_I^2$ and $(p_1+p_4)^2 = m_I^2$, respectively. This factorisation property is a reflection of the physicality of the propagating intermediate state I, created and destroyed in local three-point interactions.

The structure of the three-particle amplitudes, in turn, is completely determined by spacetime symmetries up to an overall factor. They may be catalogued for arbitrary particle spin [151–153]. In

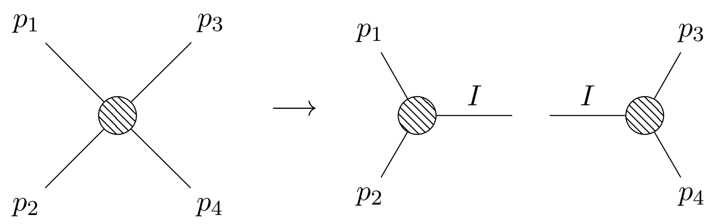


Figure 5.1: The four-particle amplitude for the scattering process $1 + 2 \to 3 + 4$ factorises into two three-particle amplitudes in the s-channel, involving the intermediate state I, whose quantum numbers are summed over.

the limit of weak coupling (where simple poles are the only singularities of the amplitudes), there is thus a network of relations between three-particle processes and higher-point amplitudes [154].

Interacting particle spectra Consistent factorisation of the four-particle amplitude is not automatic but possible only for a very restricted set of particle spectra and types of interactions [151, 155–157].

For massless particles with spins $s > 2$, factorisation cannot be achieved for *any* kind of interaction. Such particles must thus be noninteracting, and hence unobservable.

Interactions of massless spin-2 particles are highly constrained. Their coupling strength to all other states must be universal, and if multiple species of spin-2 particles exist, they must be mutually noninteracting. There can thus be only one kind of gravitation, and it must abide by the equivalence principle [156].

Interactions between different species a, b, and c of massless spin-1 particles are possible if the coupling strength f^{abc} is fully antisymmetric in the particle label and satisfies a Jacobi identity. In particular, this implies that a massless vector boson cannot interact with itself. If the coupling of a vector boson of type a to fermions belonging to species i and j is proportional to a factor T^a_{ij}, consistent factorisation is ensured if and only if these couplings form the generators of a Lie algebra whose structure constants are the f^{abc}, i.e. $[T^a, T^b]_{ij} = f^{abc}T^c_{ij}$.

Finally, massless fermions and massless scalars may undergo Yukawa-type [158] interactions with one another.

Leading and subleading interactions The above discussion only applies to the leading interactions, i.e. to those that dominate the scattering amplitude of a process at a specific characteristic energy scale E. They are responsible for the pole structure of the amplitude and are thus uniquely determined by locality. These interactions, together with the known spectrum of elementary particles, define the Standard Model (SM) of particle physics.

Scattering amplitudes may also contain additional pieces that are polynomials in the kinematic invariants and thus grow polynomially in E. Their form is left unconstrained by locality and a large number of such terms may, in principle, be present in a particular amplitude. As they do not contribute to its factorisation properties, such terms cannot be interpreted as originating from a sequence of resolved interactions between one-particle states. They parameterise the effect of local physical effects operating at much higher energy scales $\Lambda \gg E$ in the form of short-distance "contact" interactions. Their coupling constants are dimensionful quantities that contain negative powers of the scale Λ.

This is a remarkable conclusion: whatever form physics at a scale Λ might take, its effects on processes with characteristic energies around E are polynomially suppressed in the ratio E/Λ. At a fixed order in this ratio, these effects can manifest themselves only in a finite number of possible ways.

5.2 The Standard Model as a field theory

The preceding discussion considered scattering amplitudes as the primary objects and identified the properties of their analytic structure that allow them to be interpreted as resulting from a local evolution in spacetime. It is also possible to describe scattering processes in a way that makes locality manifest.

In quantum field theory (QFT), particles are represented by a set of fields $\Phi(x)$ which transform as irreducible representations of the Lorentz group. The dynamics of these fields is determined by the Lagrangian density $\mathcal{L}[\Phi]$ of the theory (hereafter simply referred to as the "Lagrangian"). It generates the action $S[\Phi] = \int d^4x\, \mathcal{L}[\Phi]$ and allows scattering amplitudes to be computed through the path integral, $\mathcal{M} \sim \int D\Phi\, e^{iS[\Phi]}$ [158]. For weak coupling, amplitudes may be constructed in terms of a diagrammatic expansion which directly corresponds to the narrative of fundamentally local interactions.

Embedding physical particles into field representations may introduce additional unphysical degrees of freedom. This is the case for massless vector bosons, whose fields A^μ contain four independent degrees of freedom. The field configurations themselves cannot, therefore, carry physical meaning, but only equivalence classes under a corresponding gauge redundancy do. To ensure independence of physical observables on the choice of gauge, the field-theoretic formulation of in-

Field	Mass dimension	Spin	Lorentz group representation	Gauge group representation
Gluons A_μ^A				$(\mathbf{8}, \mathbf{1}, 0)$
Electroweak bosons W_μ^I	1	1	$(\frac{1}{2}, \frac{1}{2})$	$(\mathbf{1}, \mathbf{3}, 0)$
Hypercharge boson B_μ				$(\mathbf{1}, \mathbf{1}, 0)$
Left-chiral quarks $Q_L = (u_L, d_L)$			$(\frac{1}{2}, 0)$	$(\mathbf{3}, \mathbf{2}, \frac{1}{6})$
Right-chiral up-type quarks u_R	$\frac{3}{2}$	$\frac{1}{2}$	$(0, \frac{1}{2})$	$(\mathbf{3}, \mathbf{1}, \frac{2}{3})$
Right-chiral down-type quarks d_R			$(0, \frac{1}{2})$	$(\mathbf{3}, \mathbf{1}, -\frac{1}{3})$
Left-chiral leptons $L_L = (\nu_L, e_L)$	$\frac{3}{2}$	$\frac{1}{2}$	$(\frac{1}{2}, 0)$	$(\mathbf{1}, \mathbf{2}, -\frac{1}{2})$
Right-chiral charged leptons e_R			$(0, \frac{1}{2})$	$(\mathbf{1}, \mathbf{1}, -1)$
Higgs $H = (H^+, H^0)$	1	0	$(0, 0)$	$(\mathbf{1}, \mathbf{2}, \frac{1}{2})$

Table 5.1: Fields in the SM, along with their mass dimensions and the spin of the particle they represent, as well as the representations in which they transform under the Lorentz group $\mathrm{SO}^+(3,1)$ and the gauge group $SU(3) \times SU(2)_L \times U(1)_Y$.

teracting massless vector bosons requires a Lagrangian which is invariant under the set of gauge transformations, the gauge group, of the theory.

Table 5.1 lists the representations of the known elementary particles in terms of fields, along with their mass dimensions [150] and gauge quantum numbers. The Lagrangian of the SM specifies the leading non-gravitational interactions between these fields as[1]

$$\mathcal{L}_{\mathrm{SM}} = -\frac{1}{4} G^A_{\mu\nu} G^{\mu\nu A} - \frac{1}{4} W^I_{\mu\nu} W^{\mu\nu I} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} + (D_\mu H)^\dagger (D^\mu H) + \sum_{\substack{\psi = Q_L, u_R, d_R, \\ L_L, e_R}} \bar{\psi} i \not{D} \psi$$
$$- \left[H^\dagger \bar{d}_R Y_d Q_L + \tilde{H}^\dagger \bar{u}_R Y_u Q_L + H^\dagger \bar{e}_R Y_e L_L + \text{h.c.} \right] - \lambda \left(H^\dagger H - \frac{v^2}{2} \right)^2 , \tag{5.1}$$

where the various terms and parameters are defined below. The Lagrangian is invariant under the gauge group $SU(3) \times SU(2)_L \times U(1)_Y$, where the first factor is associated to the gauge theory of the strong force (QCD) and the remaining two factors describe the electroweak (EW) interaction. The full gauge group is spontaneously broken to $SU(3) \times U(1)_Q$ by the vacuum expectation value of the Higgs doublet, $\langle H^\dagger H \rangle = v^2/2$. All Lorentz indices in Eq. 5.1 are suppressed for clarity. Generation (flavour) indices are also suppressed unless important, when they are denoted as $\{p, r, s, t\}$. Gauge indices are $\{A, B, C\}$ for the adjoint representation of $SU(3)$; and $\{I, J, K\}$ and $\{i, j, k\}$ for the adjoint and fundamental representations of $SU(2)_L$, respectively. The gauge-covariant derivative D_μ is $D_\mu = \partial_\mu + i g_3 T^A A^A_\mu + i g_2 t^I W^I_\mu + i g_1 Y B_\mu$. The matrices T^A and t^I are the $SU(3)$ and

[1] Additional gauge-fixing and ghost terms [158] are necessary for its consistent quantisation but not shown explicitly.

$SU(2)_L$ generators, respectively. The latter are related to the Pauli matrices σ^I as $t^I = 1/2\sigma^I$. The g_i are the dimensionless gauge couplings and the strong coupling constant α_s is defined as $\alpha_s = g_3^2/(4\pi)$. The Yang-Mills field strength tensors $G^A_{\mu\nu}$, $W^I_{\mu\nu}$, and $B_{\mu\nu}$ may be expressed in terms of the commutator $[D_\mu, D_\nu]$. The slash notation is used, where $\not{D} = \gamma^\mu D_\mu$ and the gamma matrices satisfy the Clifford algebra relation $\{\gamma^\mu, \gamma^\nu\} = 2\eta^{\mu\nu}$ with the Minkowski metric $\eta^{\mu\nu}$. The field $\tilde{H}$ is defined as $\tilde{H}_i = \epsilon_{ij} H^{\dagger j}$, where the fully antisymmetric $SU(2)_L$ tensor has $\epsilon_{12} = 1$. With these conventions, the relation between the hypercharge quantum number Y, the weak isospin eigenvalue t^3 and the electric charge Q is $Q = t^3 + Y$. The Yukawa-coupling matrices for down-type and up-type quarks and charged leptons are Y_d, Y_u, and Y_e, respectively. The Higgs quartic coupling is λ.

It will be useful later to introduce the following shortcuts to label the hermitian action of D_μ onto a generic pair of fields Φ and $\Phi^\dagger$,

$$\Phi^\dagger i\overleftrightarrow{D}_\mu \Phi = i\Phi^\dagger \left(D_\mu - \overleftarrow{D}_\mu \right) \Phi, \qquad \Phi^\dagger i\overleftrightarrow{D}^I_\mu \Phi = i\Phi^\dagger \left(\sigma^I D_\mu - \overleftarrow{D}_\mu \sigma^I \right) \Phi.$$

Mass dimensions, dimensional analysis, and energy scales A generic Lagrangian is composed of different operators $\mathcal{Q}_i$, each of which is multiplied by its characteristic coupling constant C_i,

$$\mathcal{L} = \sum_i C_i \, \mathcal{Q}_i. \tag{5.2}$$

The action $S = \int d^4x \, \mathcal{L}$ is dimensionless (it appears in the exponent in the path integral); the Lagrangian is thus required to have mass dimension four. The mass dimension of a particular operator $\mathcal{Q}_i$ is the sum of the mass dimensions of the participating fields, cf. Table 5.1. The spacetime derivative ∂_μ has a mass dimension of unity. This implies that the coupling constant C_i of an operator with mass dimension d has a mass dimension of $4-d$. It may thus be written as $C_i = c_i/M^{d-4}$, where c_i is a dimensionless coefficient and M is a characteristic mass scale. By dimensional analysis, the contribution of $\mathcal{Q}_i$ to the action scales as E^d, where E is the characteristic energy scale of the scattering process and d is the mass dimension of the operator. With $d^4x \sim E^{-4}$, the total contribution to the action generated by the operator $\mathcal{Q}_i$ scales as

$$S \ni \int d^4x \, C_i \, \mathcal{Q}_i \sim \left(\frac{E}{M} \right)^{d-4}. \tag{5.3}$$

The SM Lagrangian in Eq. 5.1 is built exclusively from operators with mass dimension $d = 4$ (with the notable exception of the Higgs mass-squared operator $H^\dagger H$ with $d = 2$). According to

Eq. 5.3, the contribution of such operators to the action is independent of the energy scale E of the process under consideration. Their contributions to scattering amplitudes $\mathcal{M}$ thus survive down to very low energies E. It is in this sense that the Lagrangian formulation of the SM encodes the leading interactions, which mirror precisely those identified by the general considerations in Section 5.1.

5.3 Separation of scales and effective descriptions

What remains to be catalogued are the "subleading" interactions originating at very high characteristic energy scales Λ that were encountered at the end of Section 5.1. Eq. 5.3 shows that operators with mass dimension $d > 4$ (and coupling constants with negative mass dimension) generate contributions to scattering amplitudes that are polynomially suppressed as $(E/\Lambda)^{d-4}$. For experiments performed at $E \ll \Lambda$, they introduce small modifications to scattering processes between SM states. Constructing all possible such operators and appending them to the Lagrangian of Eq. 5.1 represents the most general way in which the scale Λ may affect the field theory of physics at energies E. This extends the QFT of the SM into the effective field theory (EFT) referred to as the SMEFT [159],

$$\mathcal{L}_{\mathsf{SMEFT}} = \mathcal{L}_{\mathsf{SM}} + \sum_{d=5}^{\infty} \mathcal{L}^{(d)}, \qquad \mathcal{L}^{(d)} = \sum_i C_i^{(d)} \mathcal{Q}_i^{(d)} = \sum_i \frac{c_i^{(d)}}{\Lambda^{d-4}} \mathcal{Q}_i^{(d)}, \tag{5.4}$$

where $\mathcal{L}^{(d)}$ contains operators of mass dimension d. It is given by a sum of the form of Eq. 5.2, where each operator $\mathcal{Q}_i^{(d)}$ is assigned an independent coupling constant C_i. Its mass dimension is made explicit and factored out into inverse powers of the energy scale Λ; the remaining dimensionless coefficients $c_i^{(d)}$ are referred to as Wilson coefficients [159].

Eq. 5.4 is an explicit expansion of the SM in powers of E/Λ. For well separated scales, $\Lambda \gg E$, effects originating at higher mass dimension are increasingly suppressed. Conversely, the expansion becomes invalid as the ratio E/Λ approaches unity and non-SM states with mass scales of order Λ can appear in the final state as dynamical degrees of freedom.

All operators $\mathcal{Q}_i^{(d)}$ are constructed from the SM field content in Table 5.1 and must be symmetric under the full (unbroken) SM gauge group. This reflects the fact that spontaneous symmetry breaking is a phenomenon limited to the electroweak scale and independent of the scale Λ [160]. The number of operators in $\mathcal{L}^{(d)}$ that satisfy the above requirements grows exponentially in d, reaching $7.5 \cdot 10^9$ for $d = 15$ [161]. At each mass dimension, a minimal nonredundant set of operators,

Operator class	# CP-even operators, general	# CP-even operators, $U(3)^5$
F^3	2	2
H^6	1	1
D^2H^4	2	2
F^2H^2	4	4
ψ^2H^3	27	3
ψ^2FH	72	8
$D\psi^2H^2$	51	8
$\bar{\psi}\psi\bar{\psi}\psi$	1191	30
Total	1350	58

Table 5.2: Classes of $d = 6$ operators in the Warsaw basis. The symbol F refers to a generic gauge-field-strength tensor, ψ labels a generic fermion field, H is the Higgs doublet, and D symbolises a derivative. The second and third columns show the number of CP-even operators in the respective categories for the most general flavour structure of the theory, and when imposing the $U(3)^5$ symmetry discussed in Section 5.3.2.1.

an operator basis, may be systematically constructed with methods from algebraic geometry [162–165].

Some important properties of the interactions induced by these operators depend strongly on their dimensionality. If an operator $\mathcal{Q}^{(d)}$ violates baryon number by an amount ΔB and lepton number by an amount ΔL, then $\mathcal{Q}^{(d)}$ must have a mass dimension of $d \geq 9/2|\Delta B| + 3/2|\Delta L|$ [166]. Operators of odd mass-dimension *must* violate $B - L$ conservation.

5.3.1 Phenomenology of dimension-five operators

Contributions from operators with $d = 5$ are formally the least suppressed (cf. Eq. 5.3). Only a single new operator in this class is compatible with the gauge structure of the SM [167]. It provides the simplest way to generate (Majorana) neutrino-mass terms [168] with $|\Delta L| = 2$.

5.3.2 Phenomenology of dimension-six operators

The leading interactions compatible with all accidental global symmetries of the SM arise at $d = 6$. As discussed below, many new operators may contribute, modifying numerous aspects of SM phenomenology. The modifications relevant for Higgs physics are briefly reviewed below in preparation for the discussion in Section 7.9.

Using the Warsaw basis [169], the contributing operators fall into eight distinct categories, shown in Table 5.2. The SMEFT in the Warsaw basis has been completely renormalised [170–174]. Operators of the H^6 and D^2H^4 types modify the Higgs potential and thus affect the dynamics of spontaneous symmetry breaking, or renormalise the Higgs field. Similar renormalisations of the gauge fields as well as mixing between the B_μ and W^I_μ fields are effected by operators in the F^2H^2 class. They also modify the Higgs-gauge interactions. The operators of the types ψ^2H^3, ψ^2FH, and $D\psi^2H^2$ generate three- and four-point interactions between the Higgs field, the gauge fields, and fermions. The four-fermion operators $\bar{\psi}\psi\bar{\psi}\psi$ are by far the most numerous due to the many different flavour combinations and contractions possible for the SM matter content.

5.3.2.1 Flavour symmetry

The kinetic terms in the SM Lagrangian in Eq. 5.1 show a $U(3)^5$ flavour symmetry; separate (global) unitary transformations performed on the flavour indices of the five fields u_R, d_R, Q_L, L_L, and e_R leave Eq. 5.1 invariant, i.e. the flavour symmetry group corresponds to

$$U(3)^5 = \mathcal{U}(u_R) \times \mathcal{U}(d_R) \times \mathcal{U}(Q_L) \times \mathcal{U}(L_L) \times \mathcal{U}(e_R).$$

This symmetry is broken by the fermion mass terms; it is a genuine global symmetry of the SM only for vanishing Yukawa coupling matrices Y_u, Y_d, and Y_e.

This characteristic flavour structure of the SM can be extended to the entire SMEFT by retaining only $d = 6$ operators that are either fully symmetric under $U(3)^5$, or break this symmetry group in a minimal way, i.e. contain the smallest allowed number of Yukawa couplings [175]. (This is most easily done by promoting the Yukawa couplings to spurion fields as described in Ref. [176].) These additional assumptions on the flavour structure significantly reduce the parameter count of the theory (cf. Table 5.2). Certain flavour contractions are forbidden entirely, while the couplings of the leading $U(3)^5$-breaking operators become proportional to the Yukawa couplings and thus primarily affect the third generation of fermions. This is the case e.g. for the operator $\mathcal{Q}_{dH} = (H^\dagger H)(\bar{Q}_{Lp} d_{Rs} H)$, for which only the contraction of $\bar{Q}_L$ and d_R with $Y_d^\dagger$ remains,

$$\mathcal{Q}_{dH} \xrightarrow{U(3)^5} \left[Y_d^\dagger\right]_{ps} (H^\dagger H)(\bar{Q}_{Lp} d_{Rs} H),$$

as also shown in Table 5.3.

The significantly reduced number of independent Wilson coefficients in the $U(3)^5$-symmetric limit of the SMEFT greatly simplifies cross-section calculations and allows the parameters of the

Operator class	Operator	Wilson coefficient	Operator definition
D^2H^4	$\mathcal{Q}_{H\Box}$	$c_{H\Box}$	$(H^\dagger H)\partial_\mu\partial^\mu(H^\dagger H)$
	$\mathcal{Q}_{HD}$	c_{HD}	$(H^\dagger D^\mu H)^*(H^\dagger D_\mu H)$
F^2H^2	$\mathcal{Q}_{HW}$	c_{HW}	$H^\dagger H W^I_{\mu\nu}W^{\mu\nu I}$
	$\mathcal{Q}_{HB}$	c_{HB}	$H^\dagger H B_{\mu\nu}B^{\mu\nu}$
	$\mathcal{Q}_{HWB}$	c_{HWB}	$H^\dagger \sigma^I H W^I_{\mu\nu}B^{\mu\nu}$
ψ^2H^3	$\mathcal{Q}_{dH}$	c_{dH}	$\left[Y_d^\dagger\right]_{ps}(H^\dagger H)(\bar{Q}_{Lp}d_{Rs}H)$
$D\psi^2H^2$	$\mathcal{Q}^{(3)}_{H\ell}$	$c^{(3)}_{H\ell}$	$(H^\dagger i\overleftrightarrow{D}^I_\mu H)(\bar{L}_{Lp}\sigma^I\gamma^\mu L_{Lp})$
$\bar{\psi}\psi\bar{\psi}\psi$	$\mathcal{Q}^{(1)}_{\ell\ell}$	$c^{(1)}_{\ell\ell}$	$(\bar{L}_{Lp}\gamma_\mu L_{Ls})(\bar{L}_{Ls}\gamma^\mu L_{Lp})$

Table 5.3: Selected $d = 6$ operators in the $U(3)^5$-symmetric SMEFT that are important for electroweak phenomenology and discussed in the main text. Sums over repeated flavour indices are implied.

theory to be constrained by contemporary measurements, as shown in Chapter 7. For this reason, the $U(3)^5$ flavour structure will be implied in what follows, unless stated otherwise.

5.3.2.2 Electroweak symmetry breaking

The Higgs potential $V(H)$ in the SMEFT receives a contribution from the operator $\mathcal{Q}_H = (H^\dagger H)^3$,

$$V(H) = \lambda\left(H^\dagger H - \frac{v^2}{2}\right)^2 - \frac{c_H}{\Lambda^2}\left(H^\dagger H\right)^3,$$

which modifies the vacuum expectation value v_T of the Higgs field to

$$v_T = v\left(1 + \frac{3c_H v^2}{8\lambda\Lambda^2}\right),$$

so that $\langle H^\dagger H\rangle = v_T^2/2$. In the unitary gauge, the Higgs doublet is parameterised as

$$H = \frac{1}{\sqrt{2}}\begin{bmatrix} 0 \\ v_T + (1+\Delta_H)h \end{bmatrix}, \qquad \text{where} \qquad \Delta_H = \frac{v^2}{\Lambda^2}\left(c_{H\Box} - \frac{1}{4}c_{HD}\right), \tag{5.5}$$

where $c_{H\Box}$ and c_{HD} are the Wilson coefficients of the operators $\mathcal{Q}_{H\Box}$ and $\mathcal{Q}_{HD}$, defined in Table 5.3. The kinetic term for the scalar field h also receives contributions from these operators. To canonically normalise the kinetic term, h must be rescaled by a factor $1 + \Delta_H$, which depends on their Wilson coefficients and appears as a universal multiplicative factor in all interaction terms involving the scalar field h.

Upon breaking of the electroweak symmetry, similar contributions to the kinetic terms of W_μ^I and B_μ are caused by the operators $\mathcal{Q}_{HW}$ and $\mathcal{Q}_{HB}$. Kinetic mixing between W_μ^3 and B_μ is effected by $\mathcal{Q}_{HWB}$. The relevant part of the SMEFT Lagrangian in Eq. 5.4 is

$$\mathcal{L}_{\text{SMEFT}} \ni -\frac{1}{4} W_{\mu\nu}^I W^{\mu\nu I} \left(1 - 2c_{HW}\frac{v^2}{\Lambda^2}\right) - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} \left(1 - 2c_{HB}\frac{v^2}{\Lambda^2}\right) - \frac{c_{HWB}}{2}\frac{v^2}{\Lambda^2} W_{\mu\nu}^3 B^{\mu\nu}.$$

The Wilson coefficients c_{HW} and c_{HB} appear in the rescaling of the fields W_μ^I and B_μ to canonically renormalise their kinetic terms. To preserve the gauge-covariant derivative, also the couplings g_1 and g_2 are simultaneously redefined,

$$\begin{aligned} W_\mu^I &\to \left(1 + c_{HW}\frac{v^2}{\Lambda^2}\right) W_\mu^I, & B_\mu &\to \left(1 + c_{HB}\frac{v^2}{\Lambda^2}\right) B_\mu, \\ g_2 &\to \left(1 - c_{HW}\frac{v^2}{\Lambda^2}\right) g_2, & g_1 &\to \left(1 - c_{HB}\frac{v^2}{\Lambda^2}\right) g_1. \end{aligned}$$

As a result, $\mathcal{Q}_{HW}$ and $\mathcal{Q}_{HB}$ do not appear in pure-gauge interactions but only affect the Higgs-gauge coupling. Conversely, c_{HWB} appears in the transformation that relates the gauge and mass eigenstates (W_μ^3, B_μ) and (Z_μ, A_μ). It modifies the weak mixing angle θ_w and thus leads to a relative modification of all photon and Z-boson couplings,

$$\tan\theta_w = \frac{g_1}{g_2} + \frac{c_{HWB}}{2}\frac{v^2}{\Lambda^2}\left(1 - \frac{g_1^2}{g_2^2}\right). \tag{5.6}$$

5.3.2.3 Input parameters

The 19 free parameters (summarised in a vector $\mathbf{g}$) of the SM Lagrangian are determined by inverting predictions for a set of experimentally accessible quantities $\mathcal{O}_i$, i.e. from relations of the form $\mathcal{O}_i = F_i(\mathbf{g})$. In the SMEFT, the relations F_i also depend on the Wilson coefficients, taken as free parameters. At $d = 6$ the parameters $\mathbf{g}$ can thus be determined only up to terms of order v^2/Λ^2.

By construction, SMEFT predictions for the input quantities $\mathcal{O}_i$ must be independent of the Wilson coefficients. Corrections to these observables instead get shifted into the predictions made with the fully renormalised theory [176, 177].

Electroweak parameters To determine the four parameters $\mathbf{g} = (g_1, g_2, v, \lambda)$ of the electroweak Lagrangian, four independent input observables $\mathcal{O}_i$ are necessary. Different combinations are possible; here the set $\{m_W, m_Z, m_H, G_F\}$ is used, consisting of the measured masses of the W, Z, and Higgs bosons, and the Fermi constant. Tree-level predictions for these observables may be read off from the Lagrangian $\mathcal{L}_{\text{SMEFT}}$,

$$G_F = \frac{1}{\sqrt{2}v_T^2}\left(1 + 2c_{H\ell}^{(3)}\frac{v^2}{\Lambda^2} - c_{\ell\ell}^{(1)}\frac{v^2}{\Lambda^2}\right), \tag{5.7a}$$

$$m_W^2 = \frac{v_T^2}{4}g_2^2, \tag{5.7b}$$

$$m_Z^2 = \frac{v_T^2}{4}\left(g_1^2 + g_2^2\right)\left(1 + \frac{c_{HD}}{2}\frac{v^2}{\Lambda^2} + c_{HWB}\frac{2g_1g_2}{g_1^2+g_2^2}\frac{v^2}{\Lambda^2}\right), \tag{5.7c}$$

$$m_H^2 = 2v_T^2\lambda\left(1 + 2\Delta_H - \frac{3c_H}{2\lambda}\frac{v^2}{\Lambda^2}\right). \tag{5.7d}$$

The operators that contribute to these relations are again summarised in Table 5.3. As for the SM, the vacuum expectation value v_T may be determined via Eq. 5.7a from a measurement of G_F. Together with the experimentally determined boson masses, Eqs. 5.7b–5.7d fix the electroweak gauge couplings and the Higgs quartic-coupling parameter in terms of the relevant Wilson coefficients.

Fermion masses The ψ^2H^3-operators generate Yukawa-like interactions between the Higgs boson and the fermions. Focusing on the down-type quarks, the relevant terms in the Lagrangian are

$$\mathcal{L}_{\text{SMEFT}} \ni H^\dagger \bar{d}_R Y_d Q_L + \frac{c_{dH}}{\Lambda^2}\left[Y_d^\dagger\right]_{ps}(H^\dagger H)(\bar{Q}_{Lp} d_{Rs} H) + \text{h.c.}$$

The eigenvalues of the fermion mass matrix now depend explicitly on c_{dH} (cf. Table 5.3). Fixing them to the experimental values of the quark masses shifts the modifications due to $\mathcal{Q}_{dH}$ into the Yukawa couplings.

Chapter 6

High-energy scattering experiments

The matrix element $\mathcal{M} = {}_{\text{out}}\langle\{p_f\}|\{p_i\}\rangle_{\text{in}}$ determines the cross-section σ for the scattering reaction $\{p_i\} \to \{p_f\}$. Performing such scattering experiments in a controlled laboratory environment is a powerful method for the study of the spectrum of interacting fundamental particles. The initial state $|\{p_i\}\rangle$ relevant for the work presented here consists of two counterpropagating beams of ultrarelativistic protons (pp) prepared by the Large Hadron Collider (LHC), while the final state particles $\{p_f\}$ are measured with the ATLAS detector. Section 6.1 sets the notational conventions used to describe scattering processes and points out a few of their most important properties. The consequences for the design of the LHC and ATLAS are described in Sections 6.2 and 6.3, respectively.

6.1 Kinematics of proton-proton collisions

To describe the scattering reaction, it is convenient to employ a right-handed coordinate system in which the initial-state protons move along the z-axis, i.e. they have momenta $p_1 = (p, 0, 0, p)$ and $p_2 = (p, 0, 0, -p)$. The y-axis is chosen to point vertically upwards, the polar angle θ is measured from the z-axis, and the azimuthal angle φ is defined in the transverse xy-plane. Quantities defined in the transverse plane are denoted by a subscripted 'T', such as the transverse momentum p_T. The longitudinal z-component of a vectorial quantity receives a subscripted 'L', e.g. the longitudinal momentum $p_L = p^z$. Instead of θ, the pseudorapidity $\eta = -\log\tan(\theta/2)$ is often used to describe the trajectory of a final state particle, $\eta \approx 0/1/2/4$ for $\theta = 90°/40°/15°/2°$. The distance ΔR between two points $[\eta_1, \varphi_1]^\mathsf{T}$ and $[\eta_2, \varphi_2]^\mathsf{T}$ in the pseudorapidity-azimuth plane is measured as $\Delta R^2 = (\eta_1 - \eta_2)^2 + (\varphi_1 - \varphi_2)^2$. The pseudorapidity coincides with the rapidity $y = 1/2 \log\left[(E + p_L)/(E - p_L)\right]$ up to terms of order $m^2/\mathbf{p}^2$, where m is the invariant mass and

E and $\mathbf{p}$ denote the energy and the three-momentum of the particle, respectively.

Protons are composite particles; the initial state relevant for short-distance interactions is a pair of partons, quarks or gluons, carrying fractions x_1 and x_2 of the proton momenta, respectively. The parton distribution function (PDF) $f_i(x, \mu_F^2)$ parameterises the momentum distribution of the parton species i inside the proton at a characteristic energy set by the factorisation scale μ_F. The dependence of the PDF on μ_F is determined by its renormalisation group equation, which may be solved perturbatively. The PDFs relate the cross-section for a proton-induced scattering process, $\sigma(\{p_1, p_2\} \to \{p_f\})$, to the cross-sections $\hat{\sigma}_{ij}$ for the equivalent reaction initiated by the partons i and j,

$$\sigma(\{p_1, p_2\} \to \{p_f\}) = \sum_{ij} \int dx_1 dx_2\, f_i(x_1; \mu_F^2) f_j(x_2; \mu_F^2) \hat{\sigma}_{ij}(\{x_1 p_1, x_2 p_2\} \to \{p_f\}; \mu_F^2, \mu_R^2).$$

The parameter μ_R sets the scale at which the perturbatively computed cross-section $\hat{\sigma}_{ij}$ is renormalised. The squared proton-proton centre-of-mass (COM) energy is $s = (p_1 + p_2)^2$ and that of the parton system is $\hat{s} \approx x_1 x_2 s$. In the absence of other intrinsic energy scales, the partonic cross-section scales as $\hat{\sigma} \sim 1/\hat{s}$. The rate with which a particular scattering reaction occurs, the interaction rate R, is the product of its cross-section σ and the instantaneous luminosity L, which is related to the incoming flux of the initial-state protons [178].

For $\mu_F^2 \gg (1\,\text{GeV})^2$ and $x \ll 1$, the gluon PDF $f_g(x)$ of the proton dominates over the PDFs of other parton flavours and gluon-induced processes contribute significantly to the overall cross-section. As $x \to 0$, $f_g(x) \approx x^{-2}$ and the cross-section for gluon-induced processes scales as $\sigma \sim s^2/\hat{s}^3$. The COM energy enters at the fourth power. The quark PDFs are more important for $x \lesssim 1$; quark-initiated processes are thus subject to higher scales $\hat{s}$ [179, 180]. Based on the above, the characteristic cross-section for processes with $\hat{s} \approx (1\,\text{TeV})^2$ is $\sigma \approx \alpha_s^2/(1\,\text{TeV})^2 \approx 1\,\text{pb}$, requiring an instantaneous luminosity of around $10^{33}\,\text{cm}^{-2}\text{s}^{-1}$ to achieve an interaction rate R of around 1 mHz.

For $f(x) \sim x^{-\lambda}$, the rapidity of the partonic COM system is uniformly distributed in the region $|y_{\text{COM}}| < 1/2 \log s/\hat{s}$. For processes at the electroweak scale, $\sqrt{\hat{s}} \approx v$ studied at $\sqrt{s} = 13\,\text{TeV}$, this implies $|\eta_{\text{COM}}| \lesssim 4$. The final-state products of such processes are concentrated in the "forward" direction, at very shallow angles with the beam. High-mass states close to the kinematic limit, i.e. for $\hat{s} \to s$, appear in the "central" region at small rapidities.

6.2 The Large Hadron Collider

The LHC [181, 182] is a two-ring synchrotron and collider designed to provide an instantaneous proton-proton luminosity of around 10^{34} cm^{-2}s^{-1} in two interaction regions (IRs) at an energy of up to $\sqrt{s} = 14$ TeV. It is installed in the 26.7 km tunnel at CERN, originally built for the Large Electron-Positron Collider, and connected to the remainder of the accelerator complex which acts as an injector. Achieving this performance requires storing and accelerating two very high-intensity counter-rotating particle beams while maximising their density and geometrical overlap in the IRs.

In a strong-focusing synchrotron [183], the transverse beam envelope is determined by the main quadrupoles and main dipoles, which also limit the maximum beam energy. The LHC main dipoles and quadrupoles use superconducting NbTi coils operated at $T \approx 1.9$ K (significantly below the critical temperature) to maximise the achievable field and field gradient [181].

The longitudinal beam structure is generated by the radiofrequency (RF) system, which operates at a nominal frequency of 400.8 MHz and creates the short bunches (with a 4σ envelope of around 1 ns) required for high-luminosity operation. The bunch intensity is limited to about $1.2 \cdot 10^{11}$ particles by the mechanical aperture of the main magnets and the beam-beam interactions in the IRs [182]. The bunch spacing is set to 10 RF periods, i.e. around 25 ns, selected to maximise the instantaneous luminosity given the constraints set by the above machine limits and the injector complex [184].

The LHC operated at $\sqrt{s} = 13$ TeV during Run 2 in the period from 2015–2018. The peak luminosity reached $L \approx 2.1 \cdot 10^{34}$ cm^{-2}s^{-1} in 2018 and the total delivered integrated luminosity during Run 2 was around $\int dt\, L \approx 156\,\text{fb}^{-1}$ [185]. This excellent performance, together with the bunch structure described above, led to an average number of inelastic pp interactions per bunch crossing of about 13 in 2015, increasing to about 37 in 2017 [185].

6.3 The ATLAS Experiment

ATLAS [186] is a large multi-purpose detector system centred on one of the two high-luminosity IRs of the LHC. The detector volume is approximately cylindrical and immersed in a magnetic field generated by a central superconducting solenoid and three superconducting toroids. Figure 6.1 gives an overview of the arrangement of the detector subsystems necessary to operate in the challenging luminosity environment provided by the LHC.

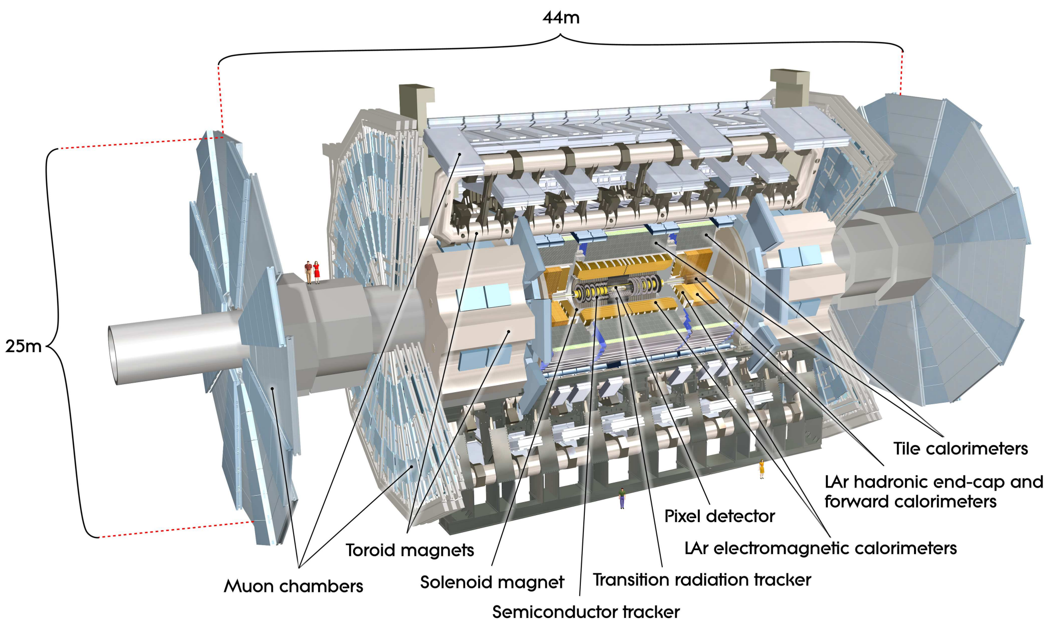


Figure 6.1: Cutaway illustration of the ATLAS detector, reproduced from Ref. [186].

Tracking and vertexing elements are located close to the IR. They permit the primary collision vertices within the luminous region to be identified, tracks produced by charged final-state particles emerging from these vertices to be reconstructed, and displaced secondary vertices signalling the decay of τ-leptons and long-lived heavy-flavour hadrons to be found. Several detector elements are used for this purpose, located within the central solenoidal field and collectively referred to as the "inner detector". It comprises high-resolution semiconductor detectors with pixel and strip geometries as well as a straw tube tracker. Electromagnetic and hadronic calorimetry systems surround the inner detector and provide energy measurements across the wide pseudorapidity acceptance of ATLAS. Both are sampling calorimeters and use a combination of liquid argon and plastic scintillators as active media. The muon spectrometer is the outermost detector system. Based on hit information delivered by gaseous detector systems, it tracks muons through the magnetic field provided by the air-core toroid. The total inelastic proton-proton cross-section of about $1/m_p^2 \sim 80\,\text{mb}$ implies an overall interaction rate of around 800 MHz at $L = 10^{34}\,\text{cm}^{-2}\text{s}^{-1}$. Limitations in the detector readout and data processing require the total inelastic rate to be suppressed by about six orders of magnitude and hard-scattering events to be selected with maximum efficiency. For this purpose, ATLAS is equipped with a two-level trigger system. The first-level trigger logic is implemented in hardware and receives information from the calorimeter and the muon spectrometer. Table 6.1 summarises the detector layout.

Detector system	Technology	$\lvert\eta\rvert$ coverage		Resolution
		Measurement	Trigger	
Inner detector: tracking & vertexing				
Pixel	Silicon pixel	< 2.5	–	$\sigma(p_T)/p_T = a\,p_T \oplus b\,\cosh^{\frac{1}{2}}\eta$
SCT	Silicon strip	< 2.5	–	$\sigma(d_0) = a \oplus b\,p_T^{-1}\cosh^{\frac{1}{2}}\eta$
TRT	Drift chamber	< 2.0	–	$\sigma(z_0) = a \oplus b\,p_T^{-1}\cosh^{\frac{3}{2}}\eta$
Calorimeters				
EM	LAr (Pb)	< 3.2	< 2.5	$\sigma(E)/E = a\,E^{-\frac{1}{2}} \oplus b\,E^{-1} \oplus c$
Tile	Scintillator (Fe)	< 1.7	< 1.7	
HEC	LAr (Cu)	$[1.5, 3.2]$	$[1.5, 3.2]$	
FCal	LAr (Cu + W)	$[3.1, 4.9]$	$[3.1, 4.9]$	
Muon system				
RPC	RPC	< 1.05	< 1.05	$\sigma(p_T)/p_T = a\,p_T^{-1} \oplus b\,p_T \oplus c\,\cosh^{\frac{1}{2}}\eta$
TGC	MWPC	$[1.05, 2.4]$	$[1.05, 2.4]$	
MDT	Drift chamber	< 2.7	–	
CSC	MWPC	$[2.0, 2.7]$	–	

Table 6.1: Overview of the detector technologies used in ATLAS. The names of the subdetector systems in the first column are explained in the main text. The acronyms RPC and MPWC stand for resistive plate chamber and multi-wire proportional chamber, respectively. All calorimeter systems are sampling calorimeters, for which the second column lists the material used for the active medium (absorber). The last column shows the characteristic composition of the resolution of the measured quantities. The constants a, b, and c typically depend on the geometry of the detector and the magnetic field.

6.3.1 Inner detector

The inner detector (ID) [187] system consists of a silicon pixel detector, a silicon strip tracker (semiconductor tracker, SCT), and a straw-tube transition-radiation tracker (TRT). The silicon tracking detectors are arranged in concentric cylinders in the central detector and in the form of end-cap disks in the forward regions. This layout provides eight precision measurements per track throughout the entire tracking volume for $\lvert\eta\rvert < 2.5$. The TRT contributes 36 additional (lower-precision) hits per track on average and adds particle-identification capabilities. Figure 6.2 gives an overview.

Tracking and vertexing performance The most important performance characteristics of a tracking detector can be understood separately from the details of its practical realisation. An idealised tracker geometry consists of N cylindrical detector layers, equidistantly spaced between an

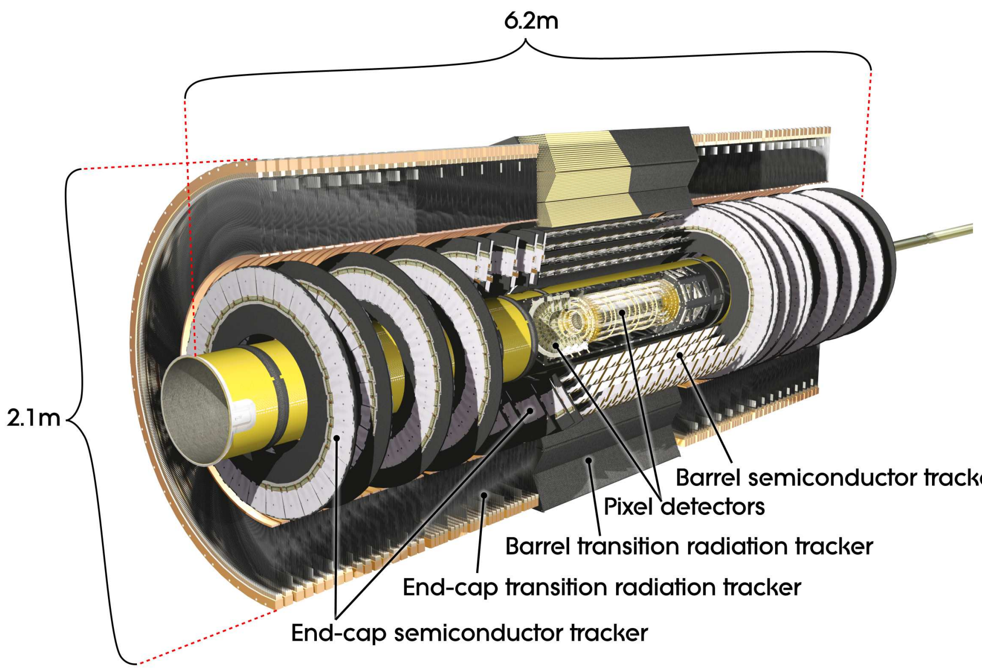


Figure 6.2: Overview of the systems comprising the inner detector. Reproduced from Ref. [186].

inner radius r_0 and an outer radius $r_0 + L$. It is immersed in a homogeneous longitudinal magnetic field B. In the practical limit $r_0 \ll L$, the relative transverse momentum resolution is [188, 189]

$$\frac{\sigma(p_T)}{p_T} \sim \left(\frac{\sigma_{r\varphi}}{BL^2} \sqrt{\frac{5}{N+5}} \right) p_T \oplus \left(\frac{\sqrt{d_{\text{tot}}/X_0}}{BL} \right) \cosh^{\frac{1}{2}} \eta, \tag{6.1}$$

where d_{tot}/X_0 is the total material budget in units of the radiation length X_0 [190] and $\sigma_{r\varphi}$ is the position resolution in the $r\varphi$-plane. Additional dimensionful constants independent of the detector geometry are suppressed for simplicity. The dependence of the resolution on N shown in Eq. 6.1 is approximate and accurate to around 6% for $N \geq 2$.

The first term in Eq. 6.1 describes the intrinsic detector resolution, limited by the precision of the measurement in the bending plane. Multiple scattering is captured by the second term and limits the resolution at low p_T and high $|\eta|$. Both contributions improve with the magnetic bending power BL and only have a weak (or no) dependence on the number of detector layers, N. Additional layers are, however, important to improve the robustness of the track reconstruction in dense environments.

The impact parameters of reconstructed tracks are highly predictive of the presence of separated secondary vertices and play an important role in jet flavour tagging, as further detailed in Section 7.3.3. The transverse impact parameter d_0 (the track perigee) is defined as the minimum transverse distance between the track and the z-axis, and the longitudinal impact parameter z_0 labels the

z-coordinate of the track perigee. In the relevant limit where $r_0 \ll L$, their resolutions are [188]

$$\sigma(d_0) \sim \frac{\sigma_{r\phi}}{\sqrt{N+5}} \oplus \left(r_0\sqrt{\frac{d}{X_0}}\right) p_T^{-1} \cosh^{\frac{1}{2}} \eta, \qquad \sigma(z_0) \sim \frac{\sigma_z}{\sqrt{N+3}} \oplus \left(r_0\sqrt{\frac{d}{X_0}}\right) p_T^{-1} \cosh^{\frac{3}{2}} \eta,$$

where d/X_0 is the material budget of the innermost detector layer. For high transverse momenta, the position resolution along the respective coordinates, $\sigma_{r\varphi}$ and σ_z, limit the accuracy of the extrapolation. Multiple scattering is important at low p_T; the resolution scales linearly with the radius r_0 of the first detector layer.

Silicon pixel detector Pixels with a minimum size of $50\,\mu\text{m} \times 400\,\mu\text{m}$ in $r\varphi \times z$ are used for the outer three layers in the barrel region as well as the end-cap disks, where they are oriented along $r\varphi \times r$. The intrinsic position resolutions are $\sigma_{r\varphi} = 10\,\mu\text{m}$ and $\sigma_z = 400\,\mu\text{m}/\sqrt{12}$. The innermost barrel layer, referred to as the IBL [191], is installed at $r_0 \approx 33\,\text{mm}$ and uses pixels with a reduced size of $50\,\mu\text{m} \times 250\,\mu\text{m}$ in $r\varphi \times z$. The track measurement in the IBL dominates the impact-parameter resolution [192].

Semiconductor tracker In the barrel, eight concentric silicon-strip layers surround the pixel detector, arranged in pairs at four equidistant radial locations. The strips have a pitch of 80 µm and are oriented along the z-axis. Neighbouring strips subtend a stereo angle of 40 mrad, resulting in position resolutions of $\sigma_{r\varphi} = 17\,\mu\text{m}$ and $\sigma_z = 580\,\mu\text{m}$. In the end-cap disks, the strips are oriented radially.

Transition radiation tracker In the barrel, the TRT consists of 73 layers of straw drift tubes oriented along the z-axis for $|\eta| < 2.0$, and radially mounted straws in the end-caps. The drift tubes have a diameter of $4\,\text{mm}$ and provide a position resolution of $130\,\mu\text{m}$ in the plane transverse to the tube. No position information is available for the coordinate parallel to the straw. The straws are interspersed with hydrocarbon radiators, in which relativistic charged particles excite transition radiation which is used in conjunction with other detector systems to identify electrons and converted photons [193]. Despite the high occupancies in the TRT of $\mathcal{O}(10\%)$, the large number of additional hits at larger radii significantly improve the robustness of the pattern recognition during track finding. The outer radius of the TRT at $r_0 + L = 1082\,\text{mm}$ defines the total extent of the tracker and the outermost straw layers contribute significantly to the momentum measurement. The bending power of the solenoid deteriorates for $|\eta| > 1.85$, correspondingly reducing the momentum resolution.

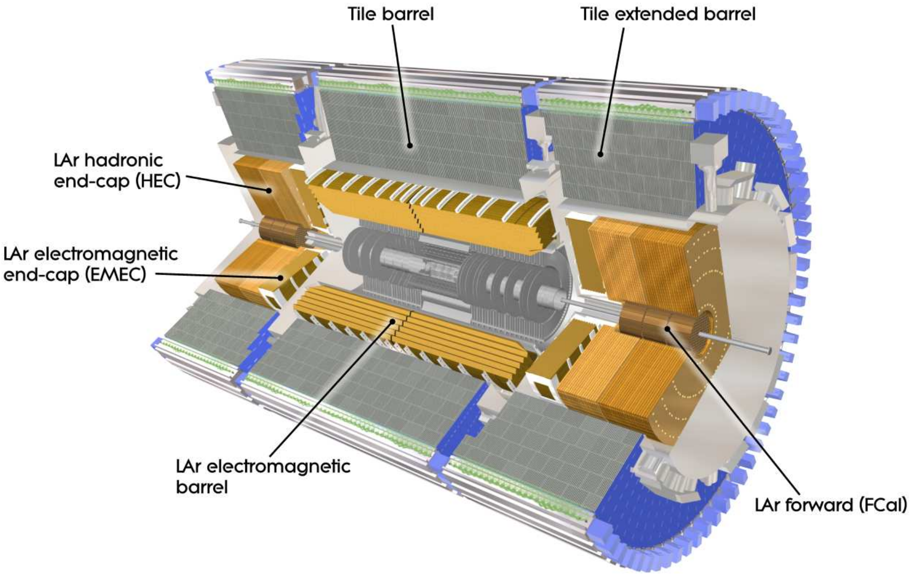


Figure 6.3: Overview of the electromagnetic and hadronic calorimeter systems. Reproduced from Ref. [186].

6.3.2 Calorimetry

The calorimetry system is composed of several different sampling calorimeters which provide hermetic coverage in the azimuthal direction. The inner electromagnetic calorimeter uses liquid argon (LAr) as active medium and covers the region $|\eta| < 3.2$ across barrel and end-caps. The outer hadronic calorimeter is based on steel absorbers and plastic scintillator tiles in the barrel for $|\eta| < 1.7$, while LAr is employed by the hadronic end-cap calorimeters (HECs). A dedicated forward calorimeter (FCal), also using LAr, provides both electromagnetic and hadron calorimetry and extends the acceptance up to $|\eta| < 4.9$. Calorimeter trigger signals representing energy deposits at a reduced granularity are derived from all systems [194]. Figure 6.3 gives an overview.

Sampling calorimetry A sampling calorimeter [195] intersperses a passive absorber medium with instrumented volumes in which the developing particle shower is sampled and the deposited energy measured. This measurement is typically done through ionisation processes, i.e. is sensitive only to the charged component of the cascade. For a simple calorimeter with constant spacing $\Delta d/X_0$ between samples and a thickness of each active element of $\Delta x/X_0$ (both in units of the radiation length), the energy resolution $\sigma(E)/E$ has the structure [196]

$$\frac{\sigma(E)}{E} \sim \left(\sqrt{\frac{\Delta d/X_0}{Z}} \oplus \log^{-1} \frac{\Delta x}{X_0} \oplus S_{\mathrm{EM}} \right) E^{-\frac{1}{2}} \oplus b\, E^{-1} \oplus c. \tag{6.2}$$

The terms proportional to $E^{-\frac{1}{2}}$ arise from the different physical contributions to the intrinsic fluctuations of the measurable energy deposition. The finite interval $\Delta d/X_0$ results in only a (fluctuating) fraction of the full cascade being sampled, described by the first "sampling" term. Using absorber materials with higher atomic-number Z results in a final state with greater particle multiplicity, thereby improving the counting statistics. The amount of energy deposited along a single track over $\Delta x/X_0$ follows a Landau distribution. Its fluctuations are particularly pronounced for thin layers of active media, as shown by the second term. This contribution is usually negligible in calorimeters which use liquid or solid active media, such as those discussed here. In a hadronic cascade, a large fraction of the total energy is usually invisible, and the total energy reconstructed from the detectable electromagnetic component is thus subject to strong fluctuations. This is parameterised by the term S_{EM}, which is dominant in a hadronic calorimeter. Lateral or longitudinal leakage of the shower beyond the nominal volume of the calorimeter leads to similar effects.

The contribution from electronic (thermal) noise and the intrinsic radioactivity of the employed materials is energy-independent; these effects are subsumed in the parameter b. Imperfections in the relative calibration between different calorimeter elements in a larger assembly are captured by the constant c, limiting the resolution for very high energies.

Liquid argon calorimeters The precision electromagnetic calorimeter [197] uses lead absorbers and LAr as active material, arranged in an accordion geometry in both barrel and end-caps. Three (for $|\eta| < 2.5$) or two (for $|\eta| \in [2.5, 3.2]$) longitudinal layers are available to extract pointing information for photons. The first layer is finely segmented to aid discrimination between photons and neutral pions, and to identify converted photons. The energy resolution is $\sigma(E)/E = 10\%\, E[\text{GeV}]^{-\frac{1}{2}} \oplus 0.7\%$. It is dominated by sampling fluctuations.

The hadronic end-cap calorimeters (HECs) [197] use a cylindrical geometry in which the copper absorber is organised into several wheels oriented perpendicular to the z-axis. The LAr drift gaps are segmented as $\Delta\eta \times \Delta\varphi = 0.1 \times 0.1$.

For the FCal [197] station providing electromagnetic calorimetry, cylindrical LAr gaps are embedded in a copper absorber matrix. The same geometry is used for the outer two hadronic calorimetry stations, which employ tungsten rather than copper to create a high-density design, respect tight space constraints, and ensure good hadronic-shower containment. Its energy resolution is $\sigma(E)/E = 100\%\, E[\text{GeV}]^{-\frac{1}{2}} \oplus 10\%$.

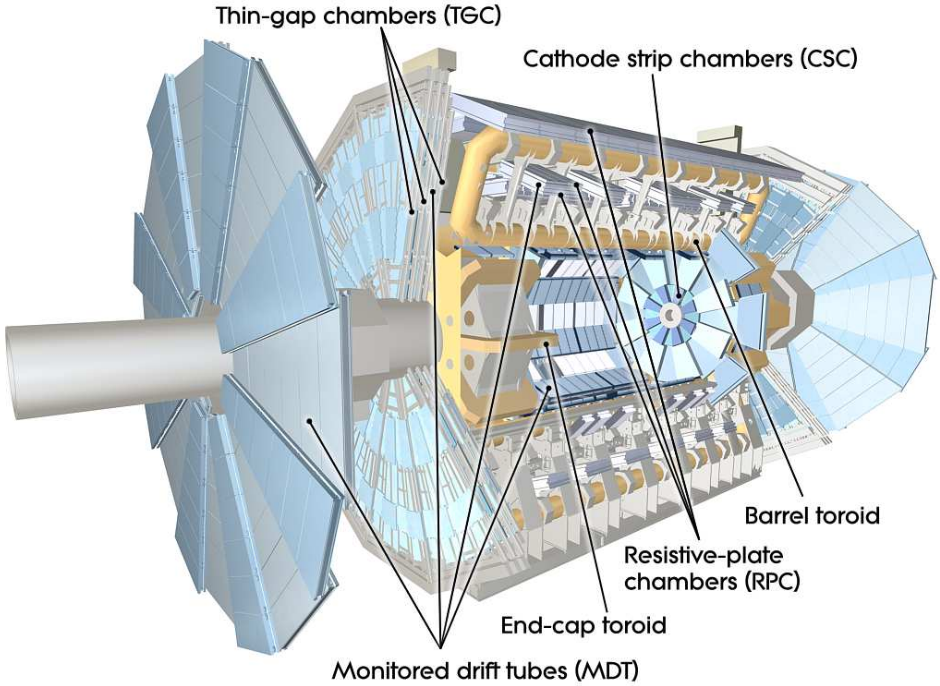


Figure 6.4: Overview of the muon spectrometer. Reproduced from Ref. [186].

Scintillator tile calorimeter For $|\eta| < 1.7$, a steel / plastic-scintillator tile calorimeter [198] is used, read out by fibre-coupled photomultiplier tubes with a segmentation of $\Delta\eta \times \Delta\varphi = 0.1 \times 0.1$ and three longitudinal layers. Its intrinsic resolution is $\sigma(E)/E = 50\%\,E[\text{GeV}]^{-\frac{1}{2}} \oplus 3\%$, where the constant term is limited by the energy calibration.

6.3.3 Muon spectrometer

The muon spectrometer (MS) [199] tracks muons through the magnetic field created by a large air-core toroid in the barrel and two smaller toroids in the end-caps. It consists of a system of monitored drift tubes (MDTs) and cathode strip chambers (CSC), which instrument the region $|\eta| < 2.7$. They provide three precision measurements of muon tracks in the principal bending direction throughout the full acceptance of the spectrometer. Within $|\eta| < 2.4$, the MS is also equipped with fast resistive-plate chambers (RPCs) in the barrel and thin-gap chambers (TGCs) in the end-caps to sample tracks along the orthogonal coordinate as well as for hardware triggering.

Precision tracking The MDT system achieves a single-wire position resolution of 80 µm, to be compared with the tube diameter of 30 mm. Each muon station consists of several layers of drift tubes to improve the position resolution and provide direction information. The maximum drift time is significantly longer than the interval between two subsequent bunch crossings. The trigger chambers (discussed below) are used to resolve ambiguities. The CSCs are multi-wire proportional chambers with cathodes segmented in orthogonal directions. They are used in the innermost end-cap plane in $\eta \in [2, 2.7]$, where particle rates are highest and shorter drift times are required. An optical alignment system is used to monitor the relative positions between the different (systems of) precision tracking chambers.

The p_T resolution of the MS is limited by the hit position resolution for high momenta ($p_T > 100\,\text{GeV}$) and by multiple scattering in the spectrometer for $p_T \in [30, 100]$ GeV (cf. Eq. 6.1). Fluctuations of the muon energy loss in the calorimeter become noticeable for $p_T < 30\,\text{GeV}$. They are approximately independent of the muon momentum and thus contribute to the resolution as $\sigma(p_T)/p_T \sim a\, p_T^{-1}$, where the constant a depends mainly on the material budget of the calorimeter. This limitation is removed if information from both the ID and the MS is used to reconstruct the muon track, improving the resolution to $\sigma(p_T)/p_T \lesssim 2\%$ for $p_T < 30\,\text{GeV}$. The toroidal field geometry renders the p_T resolution approximately independent of η, although the resolution deteriorates in the transition region $\eta \in [1.4, 1.6]$ between barrel and end-cap toroids, where the field geometry is more complicated and lower bending power is achieved.

Trigger chambers Three layers of RPCs are located within the volume of the barrel toroid. They provide a coarse position measurement in both coordinates with a resolution of around $1\,\text{cm}$ and a time resolution of approximately $1\,\text{ns}$. Coincident hits in multiple RPC planes compatible with a muon trajectory are used to select events with a well-defined p_T-trigger threshold [200].

In the end-caps, TGCs are used as trigger chambers, located entirely outside the magnet. They are thin multi-wire proportional chambers with high rate capabilities and finer segmentation but worse time resolution, allowing a similar trigger strategy to be used.

6.3.4 Trigger system

The hardware trigger receives information from the calorimeters and the muon system at reduced granularity and reduces the event rate to at most 100 kHz [201, 202]. The second trigger level is implemented in software. It performs reconstruction of the event using information from all subdetectors and further reduces the event rate to around 1 kHz. The work presented in Chapters 7 and 8 uses triggers that select events containing electrons, muons, or missing transverse energy (defined as the negative vectorial sum of calorimeter energy deposits).

Chapter 7

Measurement of WH and ZH production in the $H \to b\bar{b}$ channel

The SM in the form of Eq. 5.1 is sufficient to consistently describe most experimental results at energy scales E up to $E \lesssim 1\,\text{TeV}$. In particular, the prescribed dynamics of the Higgs field h is compatible with the experimentally observed interactions of a neutral scalar boson with a mass of $m_H \approx 125\,\text{GeV}$ discovered by the ATLAS and CMS Collaborations using the data from LHC Run 1 [203, 204]. The model parameter m_H itself is not computable from known $d = 4$ physics, suggesting that the origins of the electroweak scale, and thus the mass scales of all other known particles, lie at energies $\Lambda \gg E$ that are potentially beyond the direct kinematic reach of the LHC [205]. Similarly, other phenomena, such as experimental indications for the violation of lepton-flavour universality in semileptonic b-meson decays [206], are conjectured to be associated to new interactions occurring at high scales.

The data set collected during Run 2 of the LHC is sufficient to place meaningful constraints on $d = 6$ modifications of the SM. Such a program is an important part of the legacy of the LHC experiments, and must ultimately be based on measurements of a comprehensive set of SM observables to be successful. This chapter summarises one contribution to this effort, consisting in the experimental study of the interactions of the Higgs boson with the electroweak bosons and their impact on the structure of $d = 6$ physics.

The experimental vehicle used for this analysis is the production of an on-shell Higgs boson together with an electroweak boson $V = W, Z$ [207] from the pp initial state, $pp \to VH$. A concise review of the salient physical properties of this process and its most important backgrounds is given

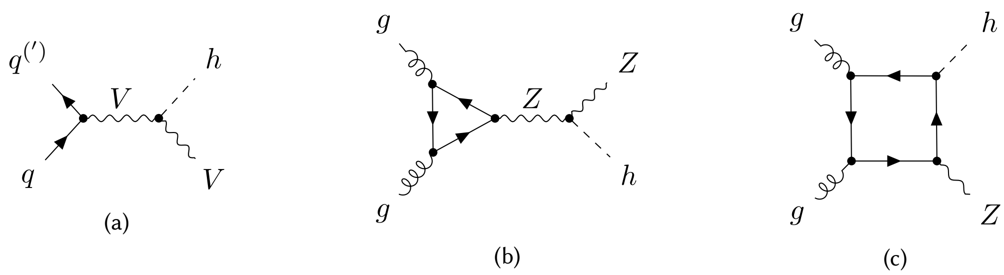


Figure 7.1: Leading-order Feynman diagrams contributing to $pp \to VH$, starting from a quark initial state in (a), or a gluon initial state in (b) and (c).

in the following. These considerations guide the design of the data analysis, to which the remainder of this chapter is dedicated.

Signal characteristics The characteristic energy scale E of $pp \to VH$ is set by the partonic COM energy, $E^2 \sim \hat{s} = \left(p^V + p^H\right)^2 = m^2_{VH}$. It scales with the transverse momentum of the vector boson, p_T^V, which can be measured with good precision. The latter thus indicates the overall kinematic scale of the VH process, and is used as a convenient surrogate for $\sqrt{\hat{s}}$ in what follows, $\sqrt{\hat{s}} \sim p_T^V$.

Both $qq^{(\prime)}$ and gg initial states contribute to the amplitude for VH production [208]. The quark-induced Drell-Yan like component in Figure 7.1a is available for both WH and ZH production, $q\bar{q} \to ZH$ and $q\bar{q}' \to WH$. The gluon-induced processes in Figure 7.1b and 7.1c proceed primarily through a top-quark loop and exist solely for ZH. Their cross-sections are suppressed by about an order of magnitude relative to the Drell-Yan contribution, but their importance is enhanced around the energy scale introduced by the top-quark loop, $\sqrt{\hat{s}} \approx 2m_t$ [209].

As discussed in Section 6.1, the contribution of the Drell-Yan signal component falls off less steeply with $\hat{s}$ compared to that of gluon-induced processes, among which are some of the most important SM backgrounds (discussed below) but also other Higgs-boson production modes [210]. Performing an experimental measurement in regions of different $\sqrt{\hat{s}} \sim p_T^V$ thus helps to maximise the sensitivity to the VH signal itself; and differential cross-sections binned in terms of this quantity are important observables to isolate and quantify $d = 6$ modifications suppressed by powers of $\hat{s}/\Lambda^2$.

At hadron colliders, the fully hadronic final state (where $V \to q\bar{q}^{(\prime)}$ and $H \to q\bar{q}$) suffers from major irreducible multi-jet continuum backgrounds which make this channel intractable with present technology. Selecting the leptonic decay of the vector boson provides a cleaner signature which naturally suppresses these background contributions, enables efficient triggering, and allows

WH and ZH production to be separated based on the number of observed charged leptons in the final state [211]. This channel provides good sensitivity to the dominant $H \to b\bar{b}$ decay mode, thereby helping to constrain the overall Higgs-boson width [212, 213].

The experimental signature of the Higgs boson in this decay channel consists of two b-jets whose axes have an angular separation of

$$\Delta R \approx \frac{m_H}{p_T^H} \frac{1}{\sqrt{z(1-z)}} \approx \frac{2m_H}{p_T^H}, \tag{7.1}$$

where $p_T^H \sim p_T^V$ is the transverse momentum of the Higgs boson and z labels the momentum fraction carried by one of the two b-quarks. In the kinematic regime $p_T^H \lesssim 300\,\text{GeV}$, where the bulk of the signal cross-section is located, $\Delta R \gtrsim 1$. The two b-quarks can be resolved into two well-separated hadronic jets that may be reconstructed individually, provided that the jet radius is chosen to be sufficiently small. (Jets of this kind are, somewhat imprecisely, referred to as "small-radius" jets in the following.) At significantly higher scales where $p_T^H > 400\,\text{GeV}$, $\Delta R < 1$ and the two jets begin to overlap and merge into one "large-radius" jet cone. Dedicated boosted-event reconstruction techniques are required to efficiently access this regime.

Backgrounds Despite the significant reduction of the multi-jet continuum in final states containing leptons, a diverse set of backgrounds with intrinsic mass scales similar to that of the signal process remains. The production of a vector boson in association with a pair of b-jets, $V + bb$, forms an important irreducible component. Jets of other flavours also contribute, but their importance can be greatly reduced by applying flavour-tagging techniques. The yield of the gluon-induced component (shown in Figure 7.2a) typically dominates over the quark-initiated process in Figure 7.2b due to the relatively enhanced PDF.

Vector-boson pair production, VZ with $Z \to b\bar{b}$, shares many kinematic features with the VH signal process. The mass scale of this process, m_Z, is sufficiently similar to the scale m_H of the signal to act as a well-characterised "standard candle" for validating the analysis procedure but still different enough from the latter to be experimentally separable.

An important reducible background, particularly for WH, is the production of top-quark pairs followed by the decay $t \to Wb$, shown in Figure 7.2c. Reducing this component requires excellent detector coverage for soft leptons and jets at high pseudorapidities to prevent misreconstructing a leptonically or hadronically decaying W boson. Top-quark pairs are preferentially produced close to threshold and are thus relatively more important for the signal at low p_T^V.

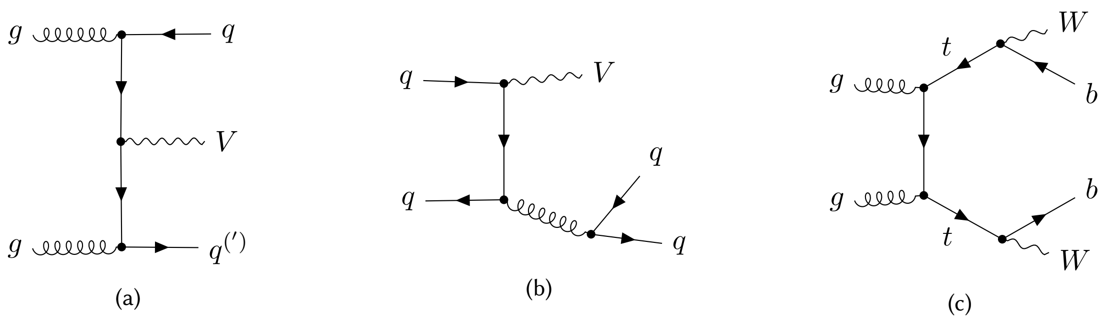


Figure 7.2: Example processes contributing to gluon-induced and quark-induced V+jets production in (a) and (b), and top-quark pair production in (c).

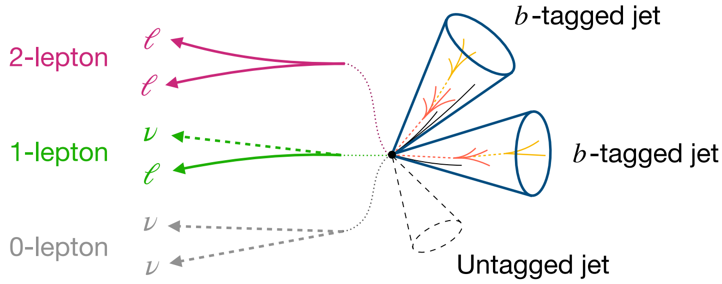


Figure 7.3: Detector signature of the process $pp \to V(\to \text{lep.})H(\to b\bar{b})$ targeted by this measurement, containing two b-tagged jets and, depending on the decay of the vector boson, zero, one, or two charged leptons.

Additional smaller background components arise from the production of single top quarks in association with a jet or a W boson, as well as from the remaining multi-jet production.

Analysis overview The measurement described in this chapter investigates the $pp \to V(\to \text{lep.})H(\to b\bar{b})$ process in its resolved regime, i.e. with a focus on $p_T^H \lesssim 300\,\text{GeV}$ [5]. It uses the full Run 2 data set collected with the ATLAS detector and includes numerous improvements over analyses previously conducted in this channel [214–216]. The measurement separately targets the processes $Z(\to \nu\nu)H(\to b\bar{b})$, $W(\to \ell\nu)H(\to b\bar{b})$, and $Z(\to \ell\ell)H(\to b\bar{b})$. It selects events with zero, one, or two charged leptons ℓ (electrons or muons), and exactly two b-tagged small-radius jets to represent the Higgs-boson candidate, as illustrated in Figure 7.3. Events containing τ-leptons are not explicitly targeted. Additional non-b-tagged jets, e.g. originating from gluon radiation off the initial-state partons, are also accepted [217]. The selected events are partitioned according to p_T^V into several mutually disjoint regions that are analysed separately, starting at $75\,\text{GeV}$ and extending to beyond $250\,\text{GeV}$.

The three analysis channels, hereafter referred to as "0-lepton", "1-lepton", and "2-lepton", have very different characteristics. The experimentally clean 2-lepton channel receives background contributions from $Z(\to \ell\ell) + bb$ and dileptonic top-quark pair production. Good signal purity is achieved despite the relatively low event yield. The dominant backgrounds in the 1-lepton channel are $W(\to \ell\nu) + bb$ and semileptonic $t\bar{t}$. The 0-lepton channel has a high signal yield due to the large $Z \to \nu\nu$ branching ratio. Its background composition is very diverse; $W + bb$, $Z + bb$ and $t\bar{t}$ contribute at similar levels.

The signal characteristics are extracted through a fit of template histograms describing the signal and all backgrounds to the binned distribution of a multivariate discriminant evaluated on data. The templates are constructed using a combination of simulation and data-driven methods. Among the central results of this analysis is a measurement of fiducial VH cross-sections binned in p_T^V, and its interpretation in the context of the SMEFT.

The following pages describe various aspects of the analysis procedure in more detail. Section 7.1 defines the cross-section bins targeted by this measurement; MC simulations employing the generators listed in Section 7.2 play a central role in the analysis design. The event reconstruction, selection, and categorisation are summarised in Sections 7.3–7.5, while Sections 7.6 and 7.7 define the model used in the statistical analysis of the data. A summary of the results is given in Sections 7.8 and 7.9.

7.1 Simplified template cross-sections

Simplified template cross-sections (STXS) are cross-sections for the production of an on-shell Higgs boson in certain, well-defined fiducial volumes [218–220]. The binning distinguishes between different Higgs-boson production modes and, for a specific production mode, is defined in terms of kinematic variables. No attempt is made to separate decay modes of the Higgs boson, which allows measurements to be combined across different final states.

The kinematic binning is operationally defined in terms of information from the truth record of the MC event generator as explained below. Detector effects are unfolded as part of the measurement procedure. The binning closely corresponds to the acceptance of the detector and so minimises extrapolations beyond the observed kinematic regions. Variations of the experimental acceptance within each bin are required wherever possible to be small relative to the expected statistical precision. The required levels of acceptance uniformity are not as stringent as for truly differential cross-section measurements, and the definition of the fiducial volumes are correspondingly simplified. This allows multivariate analysis techniques (which encode information about the kinematics

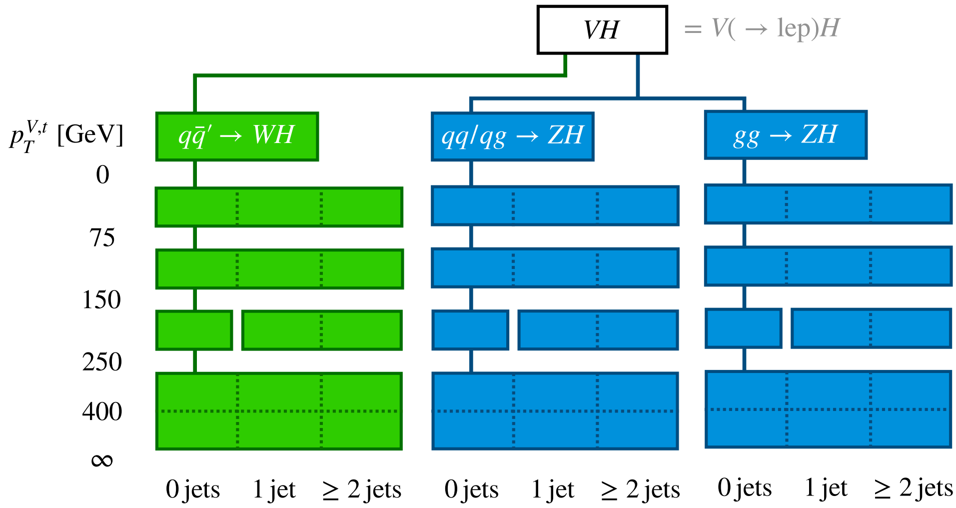


Figure 7.4: STXS binning as defined in Stage 1.2 [222]. Coloured dashed lines within a bin indicate sub-bin boundaries.

of the targeted signal process in an implicit way) to be used in an STXS measurement, maximising the precision of the result while limiting the theory uncertainties that are folded into it.

The STXS follow a staged definition in which the granularity of the binning is increased as measurements become more precise. STXS bins may also be merged in case the experimental sensitivity is insufficient for a simultaneous unfolding of all regions.

STXS binning for VH Only truth Higgs bosons with a rapidity of $|y_H| < 2.5$, approximately identical with the tracker coverage of the experiments, are considered. The truth vector boson is defined as the particle attached to the production vertex of the Higgs boson in the truth record of the generator. Only leptonically decaying vector bosons are considered, and all three charged lepton and neutrino flavours are included. Truth jets are clustered with the anti-k_T algorithm [221] with a radius parameter of $R = 0.4$, starting from the set of all stable truth particles (defined as having a lifetime greater than 10 ps) but excluding the decay products of the truth Higgs boson as well as the truth vector boson. Jets are required to have a minimum transverse momentum of 30 GeV.

The VH process is split into WH and ZH based on the identity of the truth vector boson, and ZH is further split into its quark- and gluon-initiated components. This distinction is not without ambiguities. These are resolved in an ad-hoc fashion based on the content of the available samples of simulated events, described in Section 7.2. The kinematic binning is then defined in terms of the transverse momentum of the truth vector boson, $p_T^{V,t}$, and the number of truth jets, n_{jet}^t. The $p_T^{V,t}$ bin boundaries are chosen relative to the mass scale set by m_H.

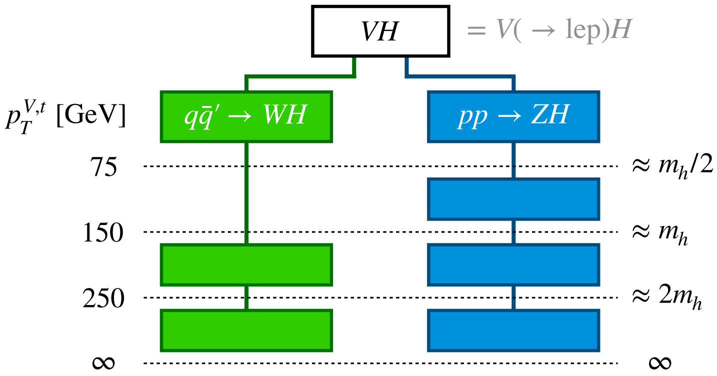


Figure 7.5: Reduced STXS binning, obtained by merging bins defined in STXS Stage 1.2.

The full binning corresponding to STXS Stage 1.2 is shown in Figure 7.4 [222]. Beside the explicit bin boundaries, it also defines a number of sub-bins. Additional explicit splits may be introduced at these locations if permitted by the experimental sensitivity. The sub-bins also enter the parameterisation of theoretical uncertainties, as laid out in Section 7.7.2.3.

A coarser binning is employed for the analysis defined here, shown in Figure 7.5. It only distinguishes between WH and ZH, does not resolve the exclusive n^t_{jet} binning, and only covers a subset of all defined $p_T^{V,t}$ bins.

7.2 Event samples

7.2.1 Data sample

The data analysed for this measurement were recorded during Run 2 of the LHC, cf. Section 6.2. A number of unprescaled single-lepton and missing-transverse-momentum triggers were used, described in more detail in Section 7.4. Events were retained only if all relevant detector systems are known to have been operating nominally. The selected data set corresponds to a total integrated luminosity of $139.0 \pm 2.4\,\text{fb}^{-1}$ [223, 224].

7.2.2 Samples of simulated events

Samples of simulated events are used to describe the VH signal as well as most background processes. With few exceptions, discussed below, the samples were generated to at least next-to-leading-order (NLO) accuracy in α_s. They were normalised to calculations of the inclusive cross-section where available, typically at a higher order. The cross-section values computed by the generator were taken otherwise. Unless noted otherwise, EvtGen (v1.6.0) [225] was used to consistently model the decays of charm or bottom hadrons across different samples. Following the simulation of the inelastic collision, minimum-bias events were superimposed to describe the effects of additional "pile-up" interactions in the same, or neighbouring, bunch crossings. They were obtained by simulating soft QCD processes with Pythia 8.186 [226] using the A3 set of tuneable parameters [227] and the NNPDF2.3LO set of PDFs [228]. Events are reweighted so as to reproduce the pile-up distribution observed during data taking. All samples are processed by a high-fidelity simulation of the ATLAS detector [229] based on GEANT [230]. Table 7.1 summarises the samples used for the nominal modelling of events, with some further details given below.

Additional samples of simulated events were used to assess the impact of theoretical uncertainties. They use alternative matrix-element generators or choose different models for parton showering and hadronisation; more information is available in Section 7.7.2.

Signal Samples were generated separately to NLO(QCD) accuracy for the qq/qg-initiated VH process, and to LO(QCD) accuracy for the gg initial state. (The gluon-induced process is separately finite and gauge invariant [217], motivating this ad-hoc distinction between the initial states. The same definition is adopted for the STXS bins in Figure 7.4.) In both cases, events were first generated with the first PDF in the NNPDF3.0NLO set and then reweighted to PDF4LHC15NLO [180] in order to be able to employ the parameterisation of uncertainties in the latter PDF set [269].

The $qq/qg \to VH$ process is normalised to NNLO(QCD)+NLO(EW) accuracy. The simulated events were further reweighted differentially in p_T^V to account for electroweak effects to NLO accuracy. This correction factor was computed with Hawk [270, 271]. Contributions from $q\gamma$-induced processes were included in the overall normalisation of the WH sample [272].

Top-quark backgrounds Top-quark pair production ($t\bar{t}$) is simulated to NLO(QCD) accuracy. The resummation damping parameter h_{damp} of the employed Powheg generator is set to $1.5m_t$, shown to result in good modelling [273].

The production of single top quarks was included in the s- and t-channel processes; Wt production was also simulated. The diagram removal scheme [274] was used to eliminate doubly-resonant diagrams overlapping with top-quark pair production.

Vector boson + jets The production of a vector boson in association with jets (V+jets) was simulated to NLO(QCD) accuracy for up to two partons in the final state, and to LO(QCD) for three-

Process	Matrix element generator	PDF for Matrix element	Parton shower and Hadronisation	Underlying event model tune	Cross-section order	Cross-section [pb]
Signal, $m_H = 125$ GeV and $B(H \to b\bar{b}) = 58.2\%$ [231–236]						
$qq/qg \to WH$ $\to \ell\nu b\bar{b}$	Powheg-Box v2 [237] + GoSam [241] + MiNLO [242, 243]	NNPDF3.0NLO [238]	Pythia 8.212 [239]	AZNLO [240]	NNLO(QCD) + NLO(EW) [244–250]	0.26904
$qq/qg \to ZH$ $\to \nu\nu b\bar{b}/\ell\ell b\bar{b}$	Powheg-Box v2 + GoSam + MiNLO	NNPDF3.0NLO	Pythia 8.212	AZNLO	NNLO(QCD) + NLO(EW)	0.13391
$gg \to ZH$ $\to \nu\nu b\bar{b}/\ell\ell b\bar{b}$	Powheg-Box v2	NNPDF3.0NLO	Pythia 8.212	AZNLO	NLO + NLL [251–255]	0.02153
Top quark, $m_t = 172.5$ GeV						
$t\bar{t}$	Powheg-Box v2 [256]	NNPDF3.0NLO	Pythia 8.230	A14 [257]	NNLO+NNLL [258]	831.76
s-channel single top	Powheg-Box v2 [259]	NNPDF3.0NLO	Pythia 8.230	A14	NLO [260]	10.32
t-channel single top	Powheg-Box v2 [259]	NNPDF3.0NLO	Pythia 8.230	A14	NLO [261]	216.99
Wt	Powheg-Box v2 [262]	NNPDF3.0NLO	Pythia 8.230	A14	Approx. NNLO [260]	71.7
Vector boson + jets						
$W \to \ell\nu$	Sherpa 2.2.1 [263–265]	NNPDF3.0NNLO	Sherpa 2.2.1 [266, 267]	Default	NNLO [268]	59702
$Z/\gamma^* \to \ell\ell$	Sherpa 2.2.1	NNPDF3.0NNLO	Sherpa 2.2.1	Default	NNLO	6268.4
$Z \to \nu\nu$	Sherpa 2.2.1	NNPDF3.0NNLO	Sherpa 2.2.1	Default	NNLO	11251
Diboson						
$qq/qg \to WW$	Sherpa 2.2.1	NNPDF3.0NNLO	Sherpa 2.2.1	Default	NLO (generator)	112.7
$qq/qg \to WZ$	Sherpa 2.2.1	NNPDF3.0NNLO	Sherpa 2.2.1	Default	NLO (generator)	50.3
$qq/qg \to ZZ$	Sherpa 2.2.1	NNPDF3.0NNLO	Sherpa 2.2.1	Default	NLO (generator)	15.6
$gg \to VV$	Sherpa 2.2.2	NNPDF3.0NNLO	Sherpa 2.2.2	Default	NLO (generator)	6.4

Table 7.1: The MC generators used for the nominal simulation of the considered signal and background processes. Samples are generated including decays into all three charged lepton (ℓ) and neutrino (ν) flavours. Unless noted otherwise, the order of the cross-section calculation refers to the expansion in powers of α_s. The last column lists the cross-sections used to normalise the samples. Adapted from Ref. [5].

and four-parton final states. For Z+jets, contributions induced by off-shell photons, $\gamma^* \to \ell\ell$, were also included. The MEPS@NLO algorithm [275–277] was used to merge the generated samples with exclusive jet multiplicities. Decays of heavy-flavour hadrons were also handled by the Sherpa generator.

Diboson The production of a pair of on-shell vector bosons (ZZ, WW, and WZ) is simulated separately for the qq/qg- and gg-induced processes. For the former, the simulation is accurate to NLO(QCD) for up to one additional parton in the final state, and to LO(QCD) for two- and three-parton final states. The samples were normalised according to the cross-sections calculated by the generator at NLO(QCD). The gluon-induced process is included at LO(QCD) for up to one additional jet. The merging procedure is identical to that employed for the V+jets sample. The interference with the VH signal is irrelevant at present levels of experimental precision and is therefore ignored. Decays of heavy-flavour hadrons were also handled by the generator.

7.3 Event reconstruction and physics object definition

Information from different subdetectors is combined and reconstructed into analysis objects representing final-state particles. The leptonic decay of the vector boson produces charged leptons (for $Z(\to \ell\ell)H$ and $W(\to \ell\nu)H$) as well as missing transverse energy (for $W(\to \ell\nu)H$ and $Z(\to \nu\nu)H$). The $H \to b\bar{b}$ decay of the Higgs boson leads to jets containing heavy-flavour hadrons. The definition of the corresponding physics objects is briefly summarised in the following, with more details available in the references provided.

7.3.1 Charged leptons

Depending on the flavour, charged leptons were reconstructed based on information from the inner detector, the calorimeters, and the muon spectrometer.

Electrons Electrons were reconstructed from topological clusters in the electromagnetic calorimeter [278–281], geometrically matched to a track reconstructed in the inner detector [282]. Topological clusters are groups of calorimeter energy deposits that are estimated to have been caused by the same electromagnetic shower. Spatially separate energy deposits caused by bremsstrahlung photons are also included in the definition of the electron energy [279]. Reconstructed tracks were also used to determine the position of the collision vertices and to associate electron candidates to these vertices. The primary vertex was defined to be the reconstructed vertex with the highest sum of squared transverse momenta of associated tracks [283, 284]. Any remaining vertices are referred to as pile-up vertices.

Only electrons with $p_T > 7\,\text{GeV}$ and $|\eta| < 2.47$ were used in this measurement. "Loose" electrons were defined to originate from the primary vertex, satisfy a track-based isolation criterion, and meet a loose quality requirement based on the electromagnetic-shower profiles, the quality of the track fit and the track-cluster matching. In addition, "tight" electrons had to satisfy more stringent quality requirements and have well-isolated calorimeter energy deposits.

Muons The dominant fraction of all muons considered in this measurement were reconstructed as a continuous track constrained by hits in both the inner detector as well as the muon spectrometer [285, 286]. To limit acceptance losses in regions of poor detector coverage, muon candidates consisting of only a track in the inner detector, combined with energy losses in the calorimeters consistent with a minimum ionising particle, and individual hits in the muon spectrometer were also considered.

All muons were required to originate from the primary vertex, be in the acceptance of the muon system with $|\eta| < 2.7$, and have a transverse momentum of $p_T > 7\,\text{GeV}$. "Loose" muons were required to have limited neighbouring-momentum activity and pass loose track-quality requirements to maximise the reconstruction efficiency. "Tight" muons were defined to be loose muons that additionally meet more stringent track quality standards.

Taus Hadronically decaying τ-leptons were reconstructed from their individual charged and neutral decay products [287, 288]. Identification of τ-candidates against hadronic activity initiated by quarks or gluons was performed with a multivariate discriminant based on the calorimeter shower shape and tracking information.

τ-leptons were not explicitly used in this measurement but were only employed to prevent double-counting and to avoid overlap with other physics objects.

7.3.2 Jets

Jets were reconstructed by combining topological clusters in the electromagnetic and hadronic calorimeters [289, 290] using the anti-k_T algorithm with a radius parameter of $R = 0.4$ [221]. Cleaning criteria were applied to identify and veto events containing spurious jets from beam-induced backgrounds, cosmic-ray showers overlapping with collision events, or noisy calorimeter cells [291]. For central jets ($|\eta| < 2.5$) with $p_T < 120$ GeV, tracking information was used to tag and remove jets that originate from a pile-up vertex [292].

Central jets with $p_T > 20$ GeV were used in this measurement. Forward jets ($2.5 \leq |\eta| < 4.5$), for which jet vertex tagging is not available, were instead required to have $p_T > 30$ GeV to limit contributions from pile-up. All jets were corrected with the standard jet-energy-scale calibration [293, 294], which subtracts energy due to pile-up and accounts for inhomogeneities in the calorimeter response and for differences between quark- and gluon- initiated jets.

Jets that were b-tagged (defined below) received additional energy corrections. A muon reconstructed in the vicinity of the jet axis was interpreted as coming from the semileptonic decay of a b-hadron, and the muon momentum was added to the momentum of the jet to improve the energy measurement. A correction, derived from simulation and dependent on the jet p_T, accounted for energy leakage outside the jet cone, which is more prominent in heavy-flavour jets. It improves the di-b-jet mass resolution by about 20%. In the 2-lepton channel, the final state of the targeted $Z(\to \ell\ell)H$ signal was fully reconstructed. A per-event likelihood uses the reconstructed vector boson decay to adjust the jet-energy measurement and thus improve the di-jet mass resolution by about 40%.

7.3.3 Jet-flavour tagging

The decay chain of the heavy-flavour hadron present in a b-jet produces secondary and sometimes also tertiary vertices [295, 296]. These are separated from one another and from the primary vertex of the event by the decay lengths of the respective b- and c-hadrons. This striking flavour substructure allows such heavy-flavour jets to be identified experimentally.

The MV2 [297] discriminant was used to identify b-jets in the central region of the detector. MV2 is a boosted decision tree which processes information about the impact parameters of the tracks associated to the jet and the properties of reconstructed secondary and tertiary decay vertices in the jet. For flavour-tagging applications (and in contrast to Section 6.3.1), the impact parameter was defined relative to the primary vertex. The selected working point achieves a b-tagging efficiency of 70% on a reference sample of simulated $t\bar{t}$ events. The c-jet and light-flavour jet misidentification efficiencies were 12.5% and 0.3%, respectively. (Light-flavour jets were defined as those *not* initiated by b- or c-quarks.)

The tagging efficiencies for b- and c-jets were measured from samples of $t\bar{t}$ events up to a jet p_T of about 600 GeV [297, 298]. Tagging efficiencies for light jets were extracted from a sample of multi-jet events [299]. These results defined a set of data-to-simulation scale factors that adjust the simulation to ensure that the expected tagging performance in data is faithfully modelled. A simulation-driven method was used to extrapolate this calibration to the high-p_T regime for which no efficiency measurements are available [11].

7.3.4 Missing transverse energy

The missing transverse energy $\mathbf{E}_T^{\text{miss}}$, as a vectorial quantity, is defined as the negative sum of the transverse momenta of charged leptons, photons, jets, hadronically decaying τ-leptons, and the "soft term" $\mathbf{p}_T^{\text{miss,st}}$ [300]. The latter is the sum of all tracks associated with the primary vertex but not used in the definition of any other object. A related observable is the track-based missing transverse momentum $\mathbf{p}_T^{\text{miss}}$, computed as the negative sum of the transverse momenta of all tracks associated to the primary vertex. It is a measure of the total momentum imbalance carried by the charged particle content of the event.

7.4 Event selection

Selected events must have exactly two b-tagged jets, referred to as $\mathbf{b}_1$ and $\mathbf{b}_2$, which were used to reconstruct the Higgs boson candidate. At most one additional untagged jet was permitted in the 0- and 1-lepton channels, where the $t\bar{t}$ background increases considerably for higher jet multiplicities. No requirement on the number of additional jets was used in the 2-lepton channel, which almost doubles the signal acceptance.

The transverse momentum of the vector boson, $\mathbf{p}_T^V$, was reconstructed as the missing transverse energy $\mathbf{E}_T^{\text{miss}}$ in the 0-lepton channel, the vectorial sum of $\mathbf{E}_T^{\text{miss}}$ and the transverse momentum of the charged lepton in the 1-lepton channel, and the transverse momentum of the dilepton system in the 2-lepton channel.

The analysis focuses on the regime of high scales, $p_T^V \gtrsim m_H$, where the signal-to-background ratio is more favourable. Events with $p_T^V \geq 150$ GeV were analysed in the 0-lepton channel (limited by the trigger acceptance) and the 1-lepton channel, where experimental sensitivity at lower scales was limited due to significant top-quark backgrounds. For the 2-lepton channel, this requirement was relaxed to $p_T^V \geq 75\,\text{GeV}$.

Table 7.2 summarises the nominal event selection. Additional details are given below.

0-lepton channel The online selection used unprescaled E_T^{miss} triggers. The trigger thresholds depend on the data-taking period and range from 70 GeV in 2015 up to 110 GeV in 2018. Depending on the trigger, the efficiency plateau was reached for an offline E_T^{miss} of between 180–200 GeV; the trigger was thus not fully efficient for a subset of events passing the offline selection. The trigger efficiency was measured for the $t\bar{t}$ and V+jets processes; differences between the measured and simulated efficiencies amount to 5% at most and were corrected. A cut on the scalar sum of the transverse momenta of the jets, H_T, removed a small number of events where the simulation fails to account for the dependence of the trigger efficiency on the distribution of jet activity in the event.

Events containing any "loose" leptons were rejected. The multi-jet background was efficiently rejected by the high E_T^{miss} threshold. Events where high levels of E_T^{miss} are due to mismeasured energy deposits in the calorimeters were removed by cuts on the angular separation between $\mathbf{E}_T^{\text{miss}}$, $\mathbf{p}_T^{\text{miss}}$, the axes of the reconstructed jets, and the di-b-jet system, **bb**.

1-lepton channel The event selection depends on the flavour of the reconstructed lepton. The electron sub-channel is comprised of events passing single-electron triggers with p_T thresholds of 24 GeV in 2015 and 26 GeV in 2016–2018. In the muon sub-channel, the same E_T^{miss} triggers as for the 0-lepton channel were used. (As explained in Section 6.3.4, muons do not form part of the online E_T^{miss} calculation.)

To minimise contributions from the multi-jet background, events were required to contain exactly one "tight" lepton and no additional "loose" leptons. Leptons were required to have a p_T of at least 27 GeV (25 GeV) in the electron (muon) sub-channels. In addition, the requirement $E_T^{\text{miss}} > 30\,\text{GeV}$ was imposed in the electron sub-channel to remove multi-jet events where a soft jet was misreconstructed as an electron.

2-lepton channel In analogy to the 1-lepton channel, electron and muon sub-channels were defined. The trigger requirement in the electron sub-channel was identical to that used in the 1-lepton channel. Events in the muon sub-channel were required to pass at least one of several single-muon triggers with p_T thresholds of 20–26 GeV.

Exactly two "loose" leptons of identical flavour were required, at least one of which must have a

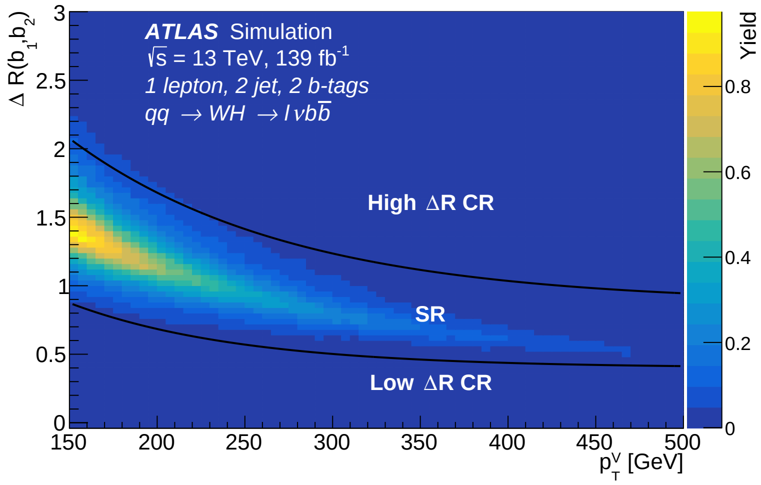


Figure 7.6: Predicted signal yield in the 2-jet category of the 1-lepton channel, shown in terms of p_T^V and $\Delta R(\mathbf{b}_1, \mathbf{b}_2)$. The definition of the SR and the two CRs is overlaid. Published in Ref. [5].

transverse momentum $p_T > 27$ GeV. Events with two muons were retained only if the muons are of opposite sign. The charge misidentification probability was large for electrons with high momenta; no charge requirement was imposed in the electron sub-channel to maintain the signal acceptance. Only the resonant contribution is of interest and the dilepton invariant mass $m_{\ell\ell}$ was required to be within 10 GeV of the nominal Z-boson mass of 91 GeV [190].

7.5 Event categorisation

Selected events were categorised into several non-overlapping regions depending on the reconstructed value of p_T^V, the number of reconstructed jets and charged leptons, and the kinematics of the di-b-jet system. The definitions of the event categories are summarised in Table 7.3.

The categorisation in terms of p_T^V was chosen following the STXS binning introduced in Figure 7.5. This results in two p_T^V regions in the 0- and 1-lepton channels, $p_T^V \in [150, 250)$ GeV and $p_T^V \geq 250$ GeV. The 2-lepton channel employs $p_T^V \in [75, 150)$ GeV as an additional third region. Events with different jet multiplicities represent different levels of signal purity and are kept separate for each p_T^V region. The 0- and 1-lepton channels define exclusive 2-jet and 3-jet categories. Following the event selection, there is no upper limit on the jet multiplicity in the 2-lepton channel, and events were sorted into a 2-jet and a $\geq$ 3-jet category.

Each of these regions was further divided into one signal region (SR) and two control regions (CRs). The SR was defined in terms of a requirement on the angular separation $\Delta R(\mathbf{b}_1, \mathbf{b}_2)$ which

Selection	0-lepton	1-lepton: e sub-channel	1-lepton: μ sub-channel	2-lepton
Trigger	E_T^{miss}	Single lepton	E_T^{miss}	Single lepton
Leptons	0 *loose* leptons	Exactly 1 *tight* electron 0 additional *loose* leptons $p_T > 27$ GeV	Exactly 1 *tight* muon 0 additional *loose* leptons $p_T > 25$ GeV	Exactly 2 *loose* leptons Leading $p_T > 27$ GeV Same-flavour Opposite-sign charges ($\mu\mu$)
E_T^{miss}	> 150 GeV	> 30 GeV	–	–
$m_{\ell\ell}$	–	–	–	$81 \text{ GeV} < m_{\ell\ell} < 101$ GeV
Jet p_T	> 20 GeV for $\lvert\eta\rvert < 2.5$ > 30 GeV for $2.5 < \lvert\eta\rvert < 4.5$			
b-jets	Exactly 2 b-tagged jets			
Leading b-tagged jet p_T	> 45 GeV			
Number of untagged jets	≤ 1	≤ 1		≥ 0
H_{T}	> 120 GeV (2 jets), > 150 GeV (3 jets)	–		–
$\min[\Delta\varphi(\mathbf{E}_T^{\text{miss}}, \mathbf{jets})]$	$> 20°$ (2 jets), $> 30°$ (3 jets)	–		–
$\Delta\varphi(\mathbf{E}_T^{\text{miss}}, \mathbf{bb})$	$> 120°$	–		–
$\Delta\varphi(\mathbf{b_1}, \mathbf{b_2})$	$< 140°$	–		–
$\Delta\varphi(\mathbf{E}_T^{\text{miss}}, \mathbf{p}_{\text{T}}^{\text{miss}})$	$< 90°$	–		–

Table 7.2: Summary of the selection of events used for the nominal analysis. The symbols are defined in the main text. Adapted from Ref. [5].

Category	0-lepton	1-lepton	2-lepton
p_T^V categories	– $150\,\text{GeV} < p_T^V < 250\,\text{GeV}$ $p_T^V > 250\,\text{GeV}$	– $150\,\text{GeV} < p_T^V < 250\,\text{GeV}$ $p_T^V > 250\,\text{GeV}$	$75\,\text{GeV} < p_T^V < 150\,\text{GeV}$ $150\,\text{GeV} < p_T^V < 250\,\text{GeV}$ $p_T^V > 250\,\text{GeV}$
Jet categories	Exactly 2 / Exactly 3 jets	Exactly 2 / Exactly 3 jets	Exactly 2 / ≥ 3 jets
Signal regions	$\Delta R(\mathbf{b}_1, \mathbf{b}_2)$ signal selection		
Control regions	High and low $\Delta R(\mathbf{b}_1, \mathbf{b}_2)$ side bands		

Table 7.3: Summary of the event categorisation based on the kinematics of the vector boson, the Higgs boson candidate, and the jet multiplicity.

isolates a domain preferentially populated by the signal, as illustrated in Figure 7.6. This selection depends continuously on p_T^V and is designed to achieve a signal efficiency approximately independent of the kinematic scale. The SR contains approximately 93% of all 2-jet signal events and about 81% of all 3-jet signal events that pass the event selection. (The signal efficiency in the ≥ 3-regions of the 2-lepton channel is lower and about 68%.)

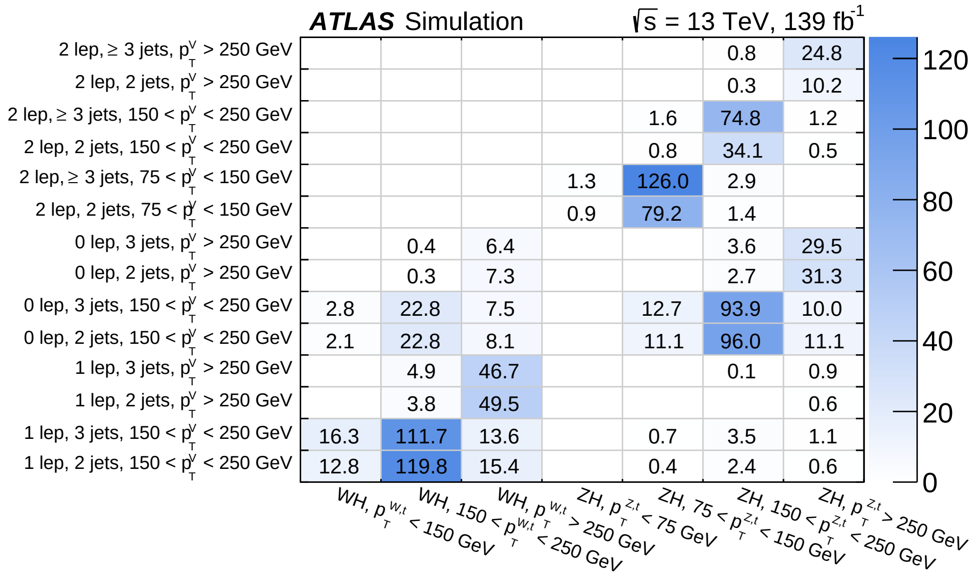
The $\Delta R(\mathbf{b}_1, \mathbf{b}_2)$ side bands define the CRs, referred to as the "high ΔR" and "low ΔR" CRs. They are used to measure the normalisations of the most critical backgrounds. The high ΔR CRs are enriched in $t\bar{t}$ events in the 1-lepton channel and in Z+jets events in the 2-lepton channel. The low ΔR CRs in the 1-lepton channel have a reasonably pure W+jets sample. The composition of the CRs in the 0-lepton channel is more complicated and both V+jets and $t\bar{t}$ contribute.

In total, this categorisation results in 14 SRs and 28 CRs.

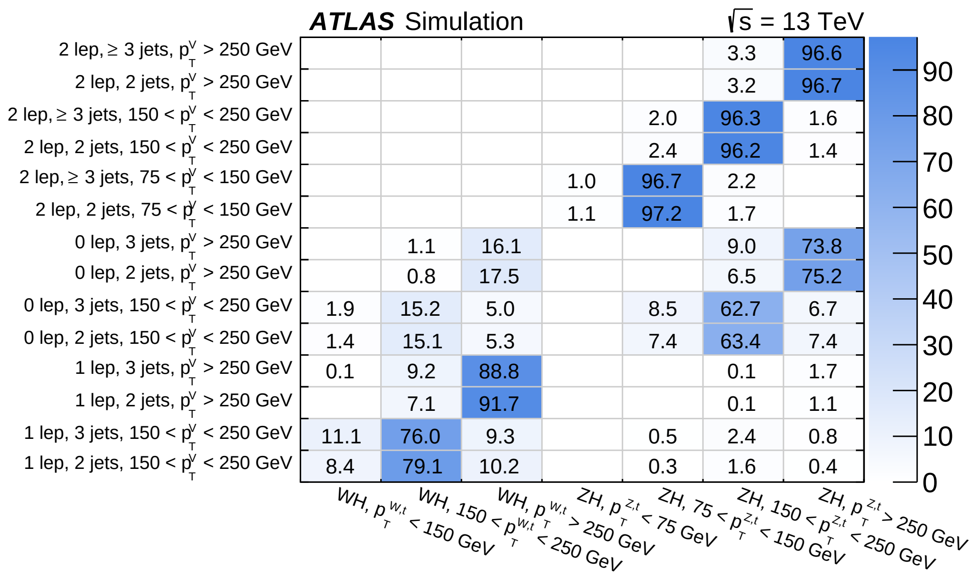
The analysis categories were constructed to have maximal signal acceptance for the corresponding targeted STXS fiducial volumes. Figure 7.7 shows the distribution of the signal yield associated with a particular STXS bin across the reconstructed regions. The accepted signal yields across both jet multiplicities are comparable, and the fraction of events from a particular $p_T^{V,t}$ bin migrating to an incorrect p_T^V region is small. A sizeable fraction of WH events in the 0-lepton channel involves the $W \to \tau(\to \text{had.})\nu_\tau$ decay chain. The analysis acceptance of WH events for $p_T^{V,t} < 150\,\text{GeV}$ and ZH for $p_T^{V,t} < 75\,\text{GeV}$ was negligible, resulting in very few reconstructed events of this type.

7.6 Multivariate discriminant

A set of multivariate discriminants, collectively denoted as BDT_{VH}, was used as a summary statistic for the statistical analysis described in Section 7.7. It was computed by a number of boosted decision trees (BDTs), trained to separate the VH signal ("sig") from the sum of all background components



(a)



(b)

Figure 7.7: Correspondence between the reconstructed analysis categories, shown on the vertical axis, and the STXS fiducial volumes, shown on the horizontal axis. (a) Predicted event yields for the VH signal process. Entries with < 0.1 events are suppressed for clarity. (b) The predicted fraction of signal events in a particular analysis category originating from a certain STXS bin. Entries with a signal fraction $< 0.1\%$ are suppressed. Each row sums to unity within rounding errors.

("bkg"). Based on a set of event variables $\{e\}$, this discriminant implements a function from which the likelihood ratio $p(\{e\}|\text{sig})/p(\{e\}|\text{bkg})$ may be reconstructed [301]. The likelihood ratio is the uniformly most powerful test statistic for testing simple hypotheses [302].

The training was performed on the samples of simulated events described in Section 7.2 using the gradient boosting algorithm available in the TMVA package [303]. Separate BDTs were derived in eight different kinematic regions, obtained by merging subsets of analysis categories to ensure sufficient training statistics. Kinematic information describing the vector-boson and Higgs-boson candidates forms the most important set of input variables. Among these are the invariant mass of the Higgs-boson candidate (m_{bb}), information on the relative orientation of the b-tagged jets and their transverse momenta ($\Delta R(\mathbf{b}_1, \mathbf{b}_2)$, $\Delta\eta(\mathbf{b}_1, \mathbf{b}_2)$, $p_{T,b1}$, and $p_{T,b2}$), the reconstructed value of p_T^V, and the pseudorapidity gap between the vector-boson and the Higgs-boson candidate ($\Delta y(\mathbf{V}, \mathbf{bb})$ and $\Delta\eta(\mathbf{V}, \mathbf{bb})$ were used, depending on the lepton channel). For events with three (or more) reconstructed jets, the p_T of the (leading) untagged jet ($p_{T,j3}$) and the invariant mass of the three-jet system (m_{bbj}) were used in addition.

Some additional event variables were employed depending on the lepton channel. In the 0-lepton channel, the effective mass $m_{\text{eff}} = p_{T,b1} + p_{T,b2} + p_{T,j3} + E_T^{\text{miss}}$ measures the energy scale of the event and helps to separate the signal from the (generally softer) backgrounds. Events containing top-quark pairs can enter this channel due to misreconstructed leptons or jets. The soft term $p_T^{\text{miss,st}}$ captures (part of) this event activity and improves discrimination against this background.

In the 1-lepton channel, the transverse mass m_T^W of the W candidate (defined as the invariant mass computed from the p_T of the reconstructed lepton and $\mathbf{E}_T^{\text{miss}}$, assuming massless particles) and the reconstructed mass of the top quark (m_{top}) were used as additional inputs. The latter assumes that the event originates from the $t\bar{t}$ process and is defined as the invariant mass of the charged lepton, the neutrino, and the b-tagged jet leading to the lower mass value. The four-momentum of the neutrino was reconstructed from $\mathbf{E}_T^{\text{miss}}$ and the assumption that the lepton-neutrino system originates from the decay of an on-shell W boson. The azimuthal angle between the lepton and the closest b-tagged jet, min $[\Delta\varphi(\boldsymbol{\ell}, \mathbf{b})]$, provided additional discriminating power in this channel.

In both the 0-lepton and 1-lepton channels, the MV2 discriminant values of the two b-tagged jets (MV2($\mathbf{b}_1$) and MV2($\mathbf{b}_2$)) help to identify $t\bar{t}$ and Wt events containing $W \to cq$ decays. (In the 2-lepton channel, these backgrounds are already very pure in their irreducible bb components and these variables provided only insignificant improvements.) The MV2 discriminant enters in the

Variable	0-lepton	1-lepton	2-lepton
m_{bb}	$\times$	$\times$	$\times$
$\Delta R(\mathbf{b}_1, \mathbf{b}_2)$	$\times$	$\times$	$\times$
$p_{T,b1}$	$\times$	$\times$	$\times$
$p_{T,b2}$	$\times$	$\times$	$\times$
p_T^V	$\equiv E_T^{\text{miss}}$	$\times$	$\times$
$\Delta\varphi(\mathbf{V}, \mathbf{bb})$	$\times$	$\times$	$\times$
MV2($\mathbf{b}_1$)	$\times$	$\times$	
MV2($\mathbf{b}_2$)	$\times$	$\times$	
$\lvert\Delta\eta(\mathbf{b}_1, \mathbf{b}_2)\rvert$	$\times$		
m_{eff}	$\times$		
$p_T^{\text{miss,st}}$	$\times$		
E_T^{miss}	$\times$	$\times$	
$\min\left[\Delta\varphi(\boldsymbol{\ell}, \mathbf{b})\right]$		$\times$	
m_T^W		$\times$	
$\lvert\Delta y(\mathbf{V}, \mathbf{bb})\rvert$		$\times$	
m_{top}		$\times$	
$\lvert\Delta\eta(\mathbf{V}, \mathbf{bb})\rvert$			$\times$
$E_T^{\text{miss}}/\sqrt{S_\text{T}}$			$\times$
$m_{\ell\ell}$			$\times$
$\cos\theta(\boldsymbol{\ell}^-, \boldsymbol{Z})$			$\times$
		(Only in 3-jet events)	
$p_T^{\text{jet}_3}$	$\times$	$\times$	$\times$
m_{bbj}	$\times$	$\times$	$\times$

Table 7.4: Input variables used for the multivariate discriminant BDT$_{VH}$. The "$\times$" symbol indicates the inclusion of a variable. Adapted from Ref. [5].

form of a categorical variable with two bins, corresponding to b-tagging efficiencies of 0–60% and 60%–70%, respectively. The continuous MV2 distribution was not calibrated against data and could not, therefore, be used.

In the 2-lepton channel, the polarisation of the Z boson carries important information. The polar angle between the three-momentum of the negative lepton (in the rest frame of the Z boson) and the three-momentum of the Z boson (in the detector frame), $\cos\theta(\boldsymbol{\ell}^-, \boldsymbol{Z})$, was used, following Ref. [304]. The significance of E_T^{miss}, $E_T^{\text{miss}}/\sqrt{S_T}$, where S_T labels the scalar sum of the momenta of reconstructed jets and leptons, helps to discriminate against dileptonic $t\bar{t}$ events, as does the dilepton invariant mass $m_{\ell\ell}$.

A complete list of all employed input variables is available in Table 7.4. A second set of BDTs, BDT$_{VZ}$, was defined using the same methodology and the same set of input variables but instead trained to identify the $VZ, Z \to b\bar{b}$ diboson process against the VH signal and the remaining backgrounds. It is used in Section 7.8.1 to validate the nominal measurement.

7.7 Statistical model

Samples of simulated events were used to model the binned distributions of BDT$_{VH}$ for the signal and most of the backgrounds, with two exceptions, as discussed in Section 7.7.1. The resulting histogram templates are subject to parameterised systematic uncertainties, summarised in Section 7.7.2. The measurement results were obtained through a joint fit of this statistical model to the observed data, as explained in Section 7.7.3.

7.7.1 Background modelling

Most background processes were modelled from simulation using the methodology described in Section 7.7.1.1. Data-driven techniques were used to extract the top backgrounds in the 2-lepton channel (where a clean data-driven determination is possible) and to estimate the multi-jet background in the 1-lepton channel (for which reliable simulations are not available). These aspects are briefly summarised in Sections 7.7.1.2 and 7.7.1.3, respectively.

7.7.1.1 Simulation-driven background modelling

The MV2 discriminant achieves substantial c-jet and light-flavour-jet rejections. Directly applying the b-tagging requirement in the event selection to simulated events would significantly deplete the samples corresponding to these flavours and generate large statistical uncertainties in the resulting histogram templates. Simulated events containing jets initiated by a c-quark or a light-flavour parton (determined from the truth record of the generator) were instead weighted by the per-jet b-tagging efficiency, parameterised in terms of the transverse momentum and the pseudorapidity of the jet.

7.7.1.2 Data-driven top-background estimation

Events from the $t\bar{t}$ and Wt backgrounds passing the event selection of the 2-lepton channel typically originate from processes where both W bosons decay leptonically; the event yields are independent of the lepton flavour. This allows both the yield and the shape of the BDT$_{VH}$ distribution for these backgrounds to be extracted from a set of dedicated control regions, referred to as the $e\mu$-CRs. They follow the nominal event selection and categorisation in the 2-lepton channel but replace the same-flavour condition on the two reconstructed leptons (ee or $\mu\mu$) with the requirement of reconstructing an $e\mu$ pair. Contributions from non-top backgrounds were negligible ($< 1\%$). Possible differences in the detector acceptance or the trigger efficiencies between the $e\mu$-CRs and the nominal analysis categories were evaluated in simulation and found to be negligible compared to the statistical uncertainty resulting from the event yield in the $e\mu$-CRs.

7.7.1.3 Multi-jet background estimation

The angular cuts in the 0-lepton channel and the selection of resonant dilepton events in the 2-lepton channel render the multi-jet background negligible in the respective analysis categories.

In the 1-lepton channel, the shape of the BDT$_{VH}$ distribution for the remaining multi-jet events and its normalisation were separately extracted from the data. Two control regions enriched in multi-jet events, MJ-CR-$1b$ and MJ-CR-$2b$, were defined for each signal region by inverting the tight lepton isolation cuts in the nominal event selection and requiring exactly one, or exactly two, b-tagged jets. The CR with relaxed b-tagging requirement, MJ-CR-$1b$, contained sufficient statistics to allow a BDT$_{VH}$ template to be extracted upon subtraction of the non-multi-jet components (taken from simulation). To ensure that the shape of the resulting template is representative of the multi-jet distribution in the SRs, the variables MV2($\mathbf{b}_1$) and MV2($\mathbf{b}_2$) in these events were replaced with samples drawn from their joint distribution (determined from MJ-CR-$2b$), and the discriminant BDT$_{VH}$ was subsequently recomputed.

The normalisation of the multi-jet background in a specific SR was determined from a fit of the multi-jet template to the m_T^W distribution of the data following the *nominal* event selection, where the shapes of non-multi-jet backgrounds were again modelled by simulation, but their normalisations were allowed to vary. The variable m_T^W offers good discrimination between multi-jet events and the remaining backgrounds, thus leading to a stable fit.

7.7.2 Systematic uncertainties

Systematic uncertainties originate from two principal sources. The understanding of the detector, the performance of the event reconstruction, and the calibration of discriminating variables built from reconstructed objects all contribute to the experimental uncertainty. Modelling uncertainties parameterise the degree of uncertainty in the simulation of signal and background processes. The treatment of both types of uncertainties is summarised in the following.

7.7.2.1 Experimental uncertainties

The most important sources of experimental systematic uncertainties enter through the calibration of the jet-energy scale and the flavour-tagging performance. Uncertainties in the data-to-simulation scale factors for the b-tagging efficiencies originate from a series of (experimental and modelling) uncertainty sources considered in the data-driven tagging-efficiency measurements for each jet flavour. These uncertainties were decomposed into several independent components, of which there are 44 for b-jets and 19 for both c-jets and light-flavour-jets [297, 298]. Uncertainties in the jet-energy scale measurement were likewise parameterised in terms of 29 independent components [293].

Uncertainty sources affecting the reconstruction of charged leptons were considered [281, 285] but did not significantly impact the results of this measurement. Uncertainties on all relevant physics objects were accounted for in the calculation of E_T^{miss}, including uncertainties on the reconstruction of tracks in the inner detector, which are important for the soft term $p_T^{\text{miss,st}}$ [305].

Additional uncertainties arise from the measurement of the integrated luminosity [223, 224] and the correction of differences in the average number of interactions per bunch crossing between simulation and data.

7.7.2.2 Background modelling uncertainties

Uncertainties in the simulation-based modelling of backgrounds arise from free parameters in the MC simulation models or from assumptions made in their design. The impact of these uncertainties was assessed by comparing predictions for different settings of these internal parameters, across different generator setups, or between the nominal simulated sample and data in a dedicated modelling control region.

These differences may be parameterised in terms of three main classes of uncertainties. "Absolute normalisation uncertainties" affect the overall normalisation of a background component, e.g. due to uncertainties in the inclusive cross-section of the corresponding process. "Relative normalisation uncertainties" capture possible relative differences in the background event yields between a pair of analysis regions with different acceptances. They are also referred to as "relative acceptance uncertainties". Acceptance uncertainties may also exist between background components that contribute to the same analysis category but involve different final-state signatures, e.g. where the same background process may produce jets of different flavour. Finally, "shape uncertainties" model variations in the differential distribution of a particular kinematic variable (or collection of such variables) within an analysis category, or a set of analysis regions.

Care was taken in the parameterisation of these uncertainties to avoid double-counting. Relative acceptance uncertainties were normalised to not impact the overall normalisation across the two analysis regions (or sample components) they concern. Similarly, shape uncertainties did not modify the total event yield in the region in which they act. A shape uncertainty defined on a collection of analysis regions may, however, change the event yield between these regions, and thus additionally act as a relative acceptance uncertainty.

The absolute normalisations of the main backgrounds (V+jets and $t\bar{t}$) are not constrained by simulation but extracted from the data in the CRs ("floating normalisation"). They are thus not subject to any absolute normalisation uncertainties. Depending on the background, independent normalisation factors were defined for the different kinematic regions shown in Figure 7.8, further discussed below. In cases where the same floating normalisation applies to multiple analysis categories, shape uncertainties and relative acceptance uncertainties were defined to allow the relative background normalisations to vary across these regions. Absolute normalisation uncertainties remain for the subleading backgrounds for which no dedicated CRs are available.

For the diboson, Z+jets, and single-top backgrounds, shape uncertainties were assessed for the two most important inputs to the multivariate discriminant, m_{bb} and p_T^V. These were found to be sufficient to cover the observed variations in the distribution of BDT$_{VH}$. For W+jets and $t\bar{t}$, shape uncertainties were defined to capture variations in the joint distribution of all input variables. For each modelling uncertainty source considered, a dedicated BDT (referred to as R_{BDT}) was trained to estimate the ratio [301] between the nominal BDT$_{VH}$ distributions and those extracted from an alternative simulated sample. It was subsequently used to reweight the nominal sample, thereby smoothly interpolating between the two generator setups. R_{BDT} also smooths out statistical fluctuations in the alternative samples, typically containing substantially fewer simulated events. Variations in the p_T^V distribution were factored out of R_{BDT} and implemented as a separate and independent shape uncertainty.

Tables 7.5 and 7.6 summarise the most important systematic uncertainties, with more details provided below.

V+jets production Simulated V+jets events were partitioned into a number of components, depending on the flavours of the b-tagged jets. The dominant component, responsible for 80%–90% of the overall V+jets yield, consists of the flavour combinations bb, bc, bl, and cc, collectively denoted as

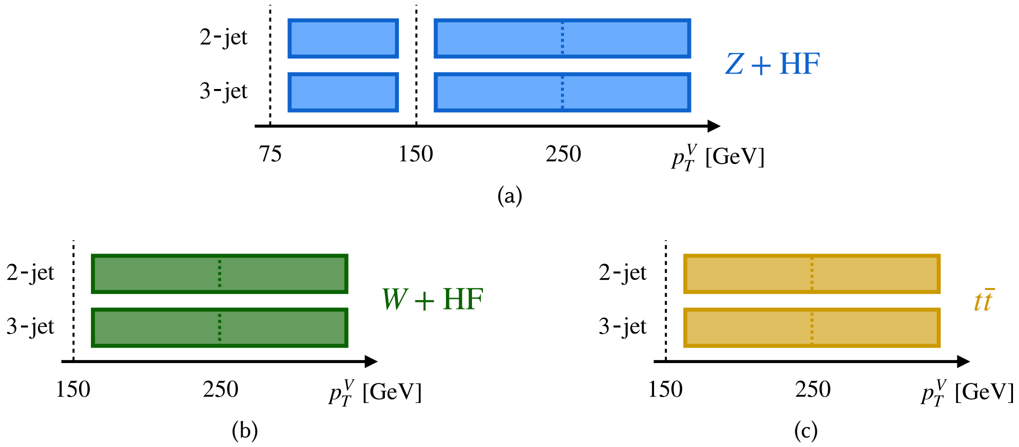


Figure 7.8: Strategy for the extraction of the normalisation of the main backgrounds from data, for Z+jets in (a), W+jets in (b), and $t\bar{t}$ in (c). The coloured boxes indicate the (combinations of) analysis regions for which a common floating normalisation is used. For V+jets, the normalisation factors apply only to the heavy flavour (HF) components, defined in the main text. Coloured dashed lines indicate the presence of uncertainties that allow the relative background yields to be adjusted between regions controlled by the same floating normalisation.

V+HF ("heavy flavour"). Its normalisation was extracted from data. The Z+HF yield is determined primarily from the 2-lepton channel, and the W+HF normalisation is constrained by the 1-lepton channel.

Relative acceptance uncertainties were assigned to control the extrapolation of these normalisation factors into the 0-lepton channel. Similar extrapolation uncertainties exist between the high ΔR and low ΔR CRs and the SR. Uncertainties on the flavour composition of V+HF were parameterised as relative acceptance uncertainties between the dominant bb component and the remaining flavour combinations. The above uncertainties subsume several components, of which the most important one is the comparison of the nominal simulated sample with an alternative generator (MadGraph5_aMC@NLO + Pythia 8 [306, 307]).

For Z+HF, shape uncertainties for m_{bb} and p_T^V were derived by comparing the nominal Sherpa simulation to data in the m_{bb} side bands to either side of the Higgs-boson peak. For W+HF, shape uncertainties originate from a comparison with the MadGraph5_aMC@NLO sample.

The remaining cl and ll flavour components are subleading, and only absolute normalisation uncertainties are defined.

$t\bar{t}$ production Only the $t\bar{t}$ contributions in the 0- and 1-lepton channels, where simulation is used to model this background, suffer from modelling uncertainty. A common floating normalisation was used for both channels, together with an uncertainty to account for acceptance differences. As for

V+jets, the dominant jet flavour combination is bb. There also exists a nonvanishing bc component, whose m_{bb} distribution contains a broad peak at a similar position as the VH signal. Flavour composition uncertainties between bb, bc, and the remaining (subleading) flavour components were derived, and shape uncertainties were also defined.

All uncertainties are based on a comparison of the nominal (Powheg + Pythia 8) simulation with alternative samples, generated using different matrix element (MadGraph5_aMC@NLO + Pythia 8) or parton shower (Powheg + Herwig 7 [308, 309]) generators.

Single-top production Normalisation uncertainties were assigned separately for s- and t-channel single-top production, as well as for Wt. They contain contributions resulting from variations of the renormalisation and factorisation scales, and from uncertainties on the PDFs and the strong coupling constant.

Uncertainties on the Wt acceptance were evaluated depending on the jet flavours. For $Wt(bb)$, the comparison of two different schemes to resolve the overlap with $t\bar{t}$ production, diagram removal and diagram subtraction [274], was the largest source. For non-bb flavours, denoted as Wt(other), the acceptance uncertainty is dominated by the parton-shower generator, assessed from a comparison of the nominal sample (Powheg + Pythia 8) with an alternative (Powheg + Herwig ++ [308, 309]) configuration.

Multi-jet production Several shape and normalisation uncertainties on the multi-jet templates were considered. They were evaluated from variations of the definition of the MJ-CRs, the simulation used for the subtraction of non-multi-jet components, and the m_T^W range used for the template fit that determined the normalisation.

Diboson production Independent absolute normalisation uncertainties exist for WZ, ZZ, and the (much less important) WW process. They are composed of differences in the cross-section computed for different scale choices by the nominal generator (Sherpa) and by an alternative generator setup (Powheg + Pythia 8). Uncertainties on the diboson acceptance, the relative acceptance between the 0- and 1-lepton channels (for WZ) and the 0- and 2-lepton channels (for ZZ), as well as shape uncertainties for p_T^V and m_{bb} were derived from similar comparisons.

Z+jets		
Normalisation	Z+HF	Floating, cf. Figure 7.8a
	$Z+ll$	18%
	$Z+cl$	23%
Flavour composition	$bc \leftrightarrow bb$	30%–40%
	$cc \leftrightarrow bb$	13%–16%
	$bl \leftrightarrow bb$	20%–28%
Relative acceptance	SR $\leftrightarrow$ low ΔR CR	3.8%–9.9% $^{(\star)}$
	SR $\leftrightarrow$ high ΔR CR	2.7%–4.1% $^{(\star)}$
	0-lepton $\leftrightarrow$ 2-lepton	7%
Shape $^{(\star)}$	p_T^V, m_{bb}	

W+jets		
Normalisation	W+HF	Floating, cf. Figure 7.8b
	$W+ll$	32%
	$W+cl$	37%
Flavour composition	$bc \leftrightarrow bb$	15%–30%
	$cc \leftrightarrow bb$	10%–30%
	$bl \leftrightarrow bb$	23%–26%
Relative acceptance	SR $\leftrightarrow$ CRs	3.6%–15%
	0-lepton $\leftrightarrow$ 1-lepton	5%
Shape	p_T^V $^{(\dagger)}$, R_{BDT} (generator)	

$t\bar{t}$ *(0-lepton and 1-lepton)*		
Normalisation		Floating, cf. Figure 7.8c
Flavour composition	$bc \leftrightarrow bb$	1.3%–8.2%
	other $\leftrightarrow bb$	0.3%–13%
Relative acceptance	0-lepton $\leftrightarrow$ 1-lepton	8%
Shape	p_T^V $^{(\dagger)}$, R_{BDT} (matrix element) $^{(\dagger)}$, R_{BDT} (parton shower) $^{(\ddagger)}$	

Single top *(0-lepton and 1-lepton)*		
Normalisation	s-channel	4.6%
	t-channel	4.4%
	Wt	6.2%
Acceptance $^{(\dagger)}$	t-channel	17%–20%
	$Wt(bb)$	51%–55%
	Wt(other)	21%–24%
Shape	p_T^V, m_{bb}	

Multi-jet *(1-lepton)*		
Normalisation $^{(\dagger)}$		30%–200%
Shape	BDT$_{VH}$	

Table 7.5: Summary of the most important modelling uncertainties for the V+jets, top quark, and multi-jet backgrounds. For relative normalisation uncertainties, the $\leftrightarrow$ symbol is used to denote the two sample components or analysis regions affected by the uncertainty. $(\star)$ Independent uncertainty components were applied in the p_T^V regions with separate normalisation factors. $(\dagger)$ Independent uncertainty components were applied in the 2-jet and 3-jet regions. $(\ddagger)$ Independent uncertainty components were applied in the 0-lepton and 1-lepton channels.

	ZZ	
Normalisation		20%
Relative acceptance	0-lepton $\leftrightarrow$ 2-lepton	6%
Acceptance		3%–18%
Shape	p_T^V, m_{bb}	
	WZ	
Normalisation		26%
Relative acceptance	0-lepton $\leftrightarrow$ 1-lepton	11%
Acceptance		4%–21%
Shape	p_T^V, m_{bb}	
	WW	
Normalisation		25%

Table 7.6: Summary of the most important modelling uncertainties for the diboson backgrounds. For relative normalisation uncertainties, the $\leftrightarrow$ symbol is used to denote the two sample components or analysis regions affected by the uncertainty.

7.7.2.3 Signal-modelling uncertainties

Signal-modelling uncertainties broadly belong to two categories: those primarily affecting the VH production cross-section and those mainly impacting the modelling of the $H \to b\bar{b}$ decay. Table 7.7 summarises the most important contributions.

Uncertainties on the inclusive VH production cross-sections and the $H \to b\bar{b}$ branching ratio were assessed following the guidelines published in Refs. [213, 310, 311]. The inclusive cross-section of the $gg \to ZH$ process is known only to NLO(QCD) accuracy, and higher-order corrections are expected to be significant. This leads to a scale-uncertainty estimate of about 25%, which is the dominant signal-modelling uncertainty in this measurement. (The cross-section of $qq/qg \to VH$ comes from an NNLO(QCD)+NLO(EW) calculation which achieves sub-percent scale uncertainties.) The dominant contributions to the uncertainty on the branching ratio $B(H \to b\bar{b})$ arise from the limited precision of measurements of the b-quark mass and α_s, and from missing higher-order corrections.

Scale uncertainties on the differential VH production cross-sections were parameterised separately for $gg \to ZH$ and $qq/qg \to VH$, following the procedure documented in Refs. [269, 312, 313]. First, scale uncertainties on the cross-sections for a number of *inclusive* $p_T^{V,t}$ and n_{jet}^t bins were assessed. The bin boundaries correspond to the sub-bins defined in STXS Stage 1.2 (cf. Section 7.1), which are sufficiently granular for present levels of experimental precision. This results in a

$\boldsymbol{VH, H \to b\bar{b}}$

Scale variations		
Inclusive cross-section	$qq/qg \to VH$	0.7%
	$gg \to ZH$	25%
Differential cross-section *(STXS Stage 1.2)*	$qq \to WH$	3.0%–3.0%
	$qq/qg \to ZH$	6.7%–12%
	$gg \to ZH$	37%–100%
Shape	p_T^V, m_{bb}	
PDF+α_s variations		
Differential cross-section *(STXS Stage 1.2)*	$qq \to WH$	1.8%–2.2%
	$qq/qg \to ZH$	1.4%–1.7%
	$gg \to ZH$	2.9%–3.3%
Shape	m_{bb}	
Parton shower variations		
Acceptance	$qq/qg \to VH$	1%–5%
	$gg \to ZH$	5%–20%
Shape	m_{bb}	
Branching fraction		
$B(H \to b\bar{b})$		1.7%

Table 7.7: Summary of the most important systematic uncertainties on the modelling of the VH signal.

total of four inclusive $p_T^{V,t}$ bins, $p_T^{V,t} \geq \{75, 150, 250, 400\}$ GeV with corresponding uncertainties Δ_{75}, Δ_{150}, Δ_{250}, and Δ_{400}; and two inclusive n^t_{jet} bins, $n^t_{\text{jet}} \geq \{1, 2\}$, with uncertainty components Δ_1 and Δ_2. These were then combined with the uncertainty on the inclusive cross-section, Δ_y, as described above, to yield the complete parameterisation. This procedure ensures a realistic assessment of scale uncertainties and avoids spurious cancellations between physically unrelated terms in the perturbative expansion [313]. Residual effects originating from modifications of kinematic distributions within each Stage 1.2 sub-bin were assessed and found to be negligible. Cross-section modifications caused by uncertainties on the PDF set and α_s were also derived in the Stage 1.2 sub-bins.

The modelling of the $H \to b\bar{b}$ decay, dominated by the parton shower, affects the signal acceptance. Acceptance uncertainties were assigned based on a comparison between the nominal signal simulation (Powheg + Pythia 8) and an alternative sample of events (Powheg + Herwig 7).

Shape uncertainties for the distributions of p_T^V and m_{bb} were defined to parameterise effects from variations in the renormalisation and factorisation scales, the PDFs, and parton shower models.

Uncertainty scheme for STXS measurement Theoretical uncertainties on the inclusive cross-section of a targeted signal component are relevant for determining the signal yield relative to the SM prediction (discussed below in Section 7.8.2), but they do not contribute to the measurement of its cross-section.

The scale-uncertainty model defined above was modified for the unfolding of the STXS in the binning used by this measurement (cf. Figure 7.5). The uncertainties $\Delta_{p_T^{V,t}}$ and $\Delta_{n^t_\text{jet}}$ were redefined by subtracting their impact on the total cross-section for each targeted STXS bin [269]. The remaining contribution thus takes the form of a shape uncertainty on the differential cross-section within these bins. As the unfolded STXS bins are fully inclusive in n^t_jet, this subtraction is trivial for the $\Delta_{n^t_\text{jet}}$ components. An analogous redefinition was performed for all PDF and α_s uncertainty components.

7.7.3 Statistical analysis

All results were obtained through a binned maximum likelihood fit [314] of the statistical model of the analysis to the observed data distributions in the SRs and CRs. The likelihood function L follows the template defined in Ref. [315],

$$L(\boldsymbol{\mu}, \boldsymbol{\theta}) = \prod_{\beta \in \text{bins}} \text{Po}(N_\beta; \nu_\beta(\boldsymbol{\mu}, \boldsymbol{\theta})) \prod_{\substack{k \in \text{constrained} \\ \text{NPs}}} f_k(\theta_k), \tag{7.2}$$

where N_β is the observed number of events in bin β and $\nu_\beta(\boldsymbol{\mu}, \boldsymbol{\theta})$ is the expected event yield predicted for this bin by the statistical model, defined below. The vector of signal strength parameters, also referred to as "parameters of interest" (POIs), $\boldsymbol{\mu}$, describes the inferred yield of the targeted signal processes. The vector of nuisance parameters (NPs), $\boldsymbol{\theta}$, implements the effects of the parameterised systematic uncertainties. Certain classes of NPs are subject to constraints, expressed by the second term in Eq. 7.2.

The maximum likelihood estimators (MLE) $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\theta}}$ were used to obtain point estimates for the model parameters $\boldsymbol{\mu}$ and $\boldsymbol{\theta}$. Confidence intervals for $\boldsymbol{\mu}$ were constructed using the profile likelihood test statistic [316],

$$t(\boldsymbol{\mu}) = -2 \log \frac{L(\boldsymbol{\mu}, \hat{\hat{\boldsymbol{\theta}}})}{L(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\theta}})}, \tag{7.3}$$

where $\hat{\hat{\boldsymbol{\theta}}}$ is the MLE of the vector of NPs conditional on $\boldsymbol{\mu}$, i.e. $\hat{\hat{\boldsymbol{\theta}}} = \hat{\hat{\boldsymbol{\theta}}}(\boldsymbol{\mu})$. Confidence intervals of this form achieve exact coverage only for the points $(\boldsymbol{\mu}, \hat{\hat{\boldsymbol{\theta}}})$. For a d-dimensional confidence interval, this test statistic is asymptotically distributed according to a (noncentral) χ^2_d-distribution [317, 318].

The asymptotic approximation is used throughout the remainder of this chapter. The MLEs and confidence intervals determined from an Asimov data set were used to describe the results expected for particular true values of the model parameters. The Asimov data set is constructed following Ref. [316]. The technical implementation uses Minuit [319], RooFit [320], and RooStats [321].

The discovery test statistic t_0 [316] was used to quantify the significance of an observed positive signal, or individual signal component, $\mu \geq 0$,

$$t_0 = \begin{cases} t(0) & \hat{\mu} \geq 0, \\ 0 & \hat{\mu} < 0. \end{cases}$$

The p-value of t_0 under the $\mu = 0$ hypothesis, p_0, was used to define the significance $Z = \Phi^{-1}(1 - p_0)$, where Φ is the cumulative distribution function of the standard normal distribution.

7.7.3.1 Statistical model

The likelihood in Eq. 7.2 includes all SRs and CRs defined in Section 7.5. In the SRs, the binned distributions of the final discriminant (BDT$_{VH}$ or BDT$_{VZ}$) were fitted, while the event yields were used in the CRs. The binning in the SRs was chosen dynamically to generate narrow bins in signal-enriched domains, while ensuring that the background estimate did not suffer from excessive statistical uncertainties. The number of generated bins depends on the total event yield expected in each SR.

The model prediction ν_β is a linear combination of histogram templates for the participating processes, obtained from the nominal simulation or, for the cases described in Section 7.7.1, from data. The signal strength parameters $\boldsymbol{\mu}$ act as scale factors for the signal templates, i.e. they measure the signal abundance relative to the SM prediction. The normalisations of the floating background components (cf. Section 7.7.2.2) are controlled by similar, unconstrained, scale factors.

A separate constrained NP θ_k was introduced for each independent modelling or experimental systematic-uncertainty component k. The NPs act linearly, modifying the relative normalisations of the nominal histogram templates in different analysis regions (for relative acceptance uncertainties), or interpolating the binned distribution between its nominal and alternative configurations (for shape uncertainties). Each NP was standardised such that its modification vanishes for $\theta_k = 0$ and attains the parameterised magnitude for $|\theta_k| = 1$ (cf. Section 7.7.2). The constraint term $f_k(\theta_k)$ in Eq. 7.2 is $f_k(\theta_k) \sim \exp(-\theta_k^2/2)$. Uncertainty components which have a negligible impact on the binned distribution in a given analysis category are pruned from the model.

The effect of statistical uncertainties on the template histograms was included using the Barlow-Beeston "light" method [322], which introduces a single additional constrained NP for each bin.

7.7.3.2 Fit configurations

Several different fit configurations were investigated. To validate the analysis, a measurement of the overall VZ signal strength was performed through a fit of the BDT$_{VZ}$ distributions. This defines a likelihood with a single POI, μ^{bb}_{VZ}. A second validation fit simultaneously measured the WZ and ZZ signal strengths, parameterised by the POIs μ^{bb}_{WZ} and μ^{bb}_{ZZ}. In both cases, the Higgs boson signal was set to its SM expectation within a conservative normalisation uncertainty of 50%.

The measurements of the VH process used BDT$_{VH}$ as final discriminant. Signal strengths were measured both inclusively for VH (μ^{bb}_{VH}), as well as separately for WH and ZH (μ^{bb}_{WH} and μ^{bb}_{ZH}).

The measurement of the five STXS bins shown in Figure 7.5 is based on a likelihood L^{STXS} which introduces a separate signal-strength parameter μ^b for each STXS bin b. With the signal systematics model defined in Section 7.7.2.3 for cross-section measurements, these parameters measure the fiducial production cross-sections times the branching ratio into $b\bar{b}$ relative to the SM expectation. Signal contributions from STXS bins not targeted by the analysis were set to their SM expectations within uncertainties.

7.8 Results

The results summarised in the following focus on the aspects relevant for the developments in Section 7.9. Section 7.8.1 validates design choices made in the construction of the analysis through a measurement of the diboson signal strengths. The measured WH and ZH signal strengths are listed in Section 7.8.2 and the results of the STXS measurement for these processes are summarised in Section 7.8.3. Additional information and results from an alternative analysis strategy which does not rely on the multivariate discriminant BDT$_{VH}$ is available in the original publication [5] and in a HEPData repository [323].

7.8.1 Diboson validation

The signal strength of the VZ process is measured to be

$$\mu^{bb}_{VZ} = 0.93^{+0.16}_{-0.13} = 0.93^{+0.07}_{-0.06}\,(\text{stat.})\,^{+0.14}_{-0.12}\,(\text{syst.}).$$

The signal strengths for WZ and ZZ obtained from a simultaneous fit are

$$\mu^{bb}_{WZ} = 0.68^{+0.26}_{-0.25} = 0.68^{+0.15}_{-0.15}\,\text{(stat.)}\,^{+0.21}_{-0.19}\,\text{(syst.)},$$
$$\mu^{bb}_{ZZ} = 1.00^{+0.18}_{-0.15} = 1.00^{+0.08}_{-0.08}\,\text{(stat.)}\,^{+0.16}_{-0.13}\,\text{(syst.)},$$

with a Pearson correlation of 3% between these values. All values are compatible with the SM prediction within one standard deviation.

7.8.2 Signal-strength measurements

Figure 7.9 shows the modelling of the observed BDT$_{VH}$ distribution in the three lepton channels for different SRs. Good modelling is observed in all regions. The extracted floating normalisation of the $t\bar{t}$ background in the 0- and 1-lepton channels is compatible with the SM expectation. The floating normalisation of the W+HF background component is corrected upwards by about 6% (15%) relative to the nominal simulation for the 2-jet (3-jet) categories. The Z+HF background is increased by 9%–28% relative to the simulation, with stronger corrections for lower values of p_T^V. Similar effects have been observed in previous results [214–216], indicating deficiencies in the modelling of the p_T^V spectra of these processes.

The inclusive VH signal strength is measured to be

$$\mu^{bb}_{VH} = 1.02^{+0.18}_{-0.17} = 1.02^{+0.12}_{-0.11}\,\text{(stat.)}\,^{+0.14}_{-0.13}\,\text{(syst.)},$$

and the WH and ZH signal strengths are

$$\mu^{bb}_{WH} = 0.95^{+0.27}_{-0.25} = 0.95^{+0.18}_{-0.18}\,\text{(stat.)}\,^{+0.19}_{-0.18}\,\text{(syst.)},$$
$$\mu^{bb}_{ZH} = 1.08^{+0.25}_{-0.23} = 1.08^{+0.17}_{-0.17}\,\text{(stat.)}\,^{+0.18}_{-0.15}\,\text{(syst.)},$$

with a Pearson correlation of 2.7%, illustrated in Figure 7.10. The ZH process is observed for the first time with a significance of 5.3 standard deviations (5.1 s.d. expected), and the above result shows strong evidence for the WH process with an observed significance of 4.0 standard deviations (4.1 s.d. expected). The inclusive VH process is observed at 6.7 standard deviations (6.7 s.d. expected).

The total statistical uncertainty is defined as the uncertainty in the signal strength when all NPs are fixed at their MLEs, except for those implementing the floating background normalisations and the effect of limited data statistics in the $e\mu$-CRs. The total systematic uncertainty is the difference in quadrature between the total uncertainty on the signal strength and the total statistical

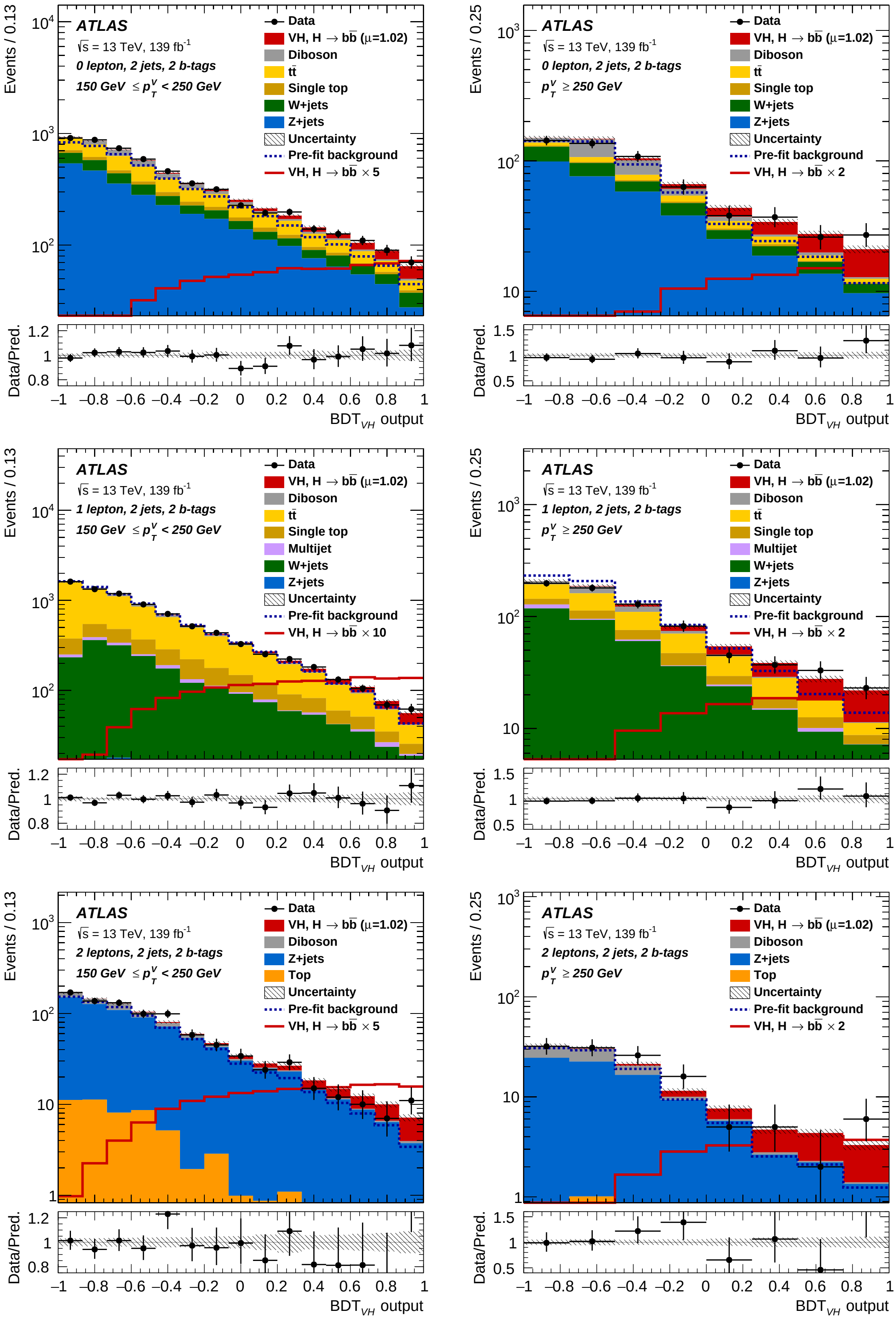
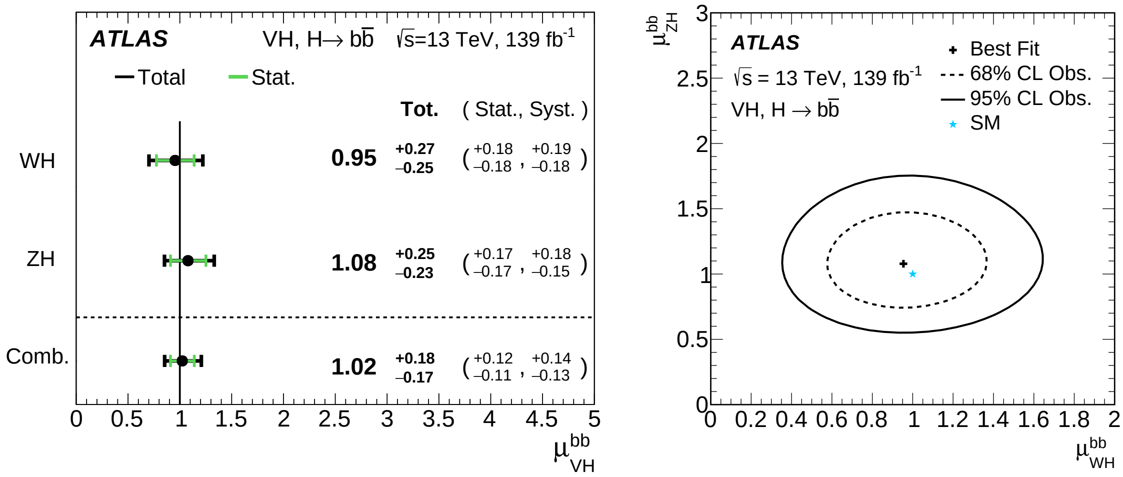


Figure 7.9: Binned distributions of BDT_{VH} in the 0-lepton (top), 1-lepton (middle), and 2-lepton (bottom) channels, for $p_T^V \in [150, 250)$ GeV (left) and $p_T^V > 250$ GeV (right). The signal and background components are shown as stacked histograms, with the dashed line indicating the expected total background contribution. The signal contribution, scaled by the extracted signal strength, is additionally indicated by the red line. The total background uncertainty induced by the uncertainty in the fit parameters $(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\theta}})$ is shown by the hatched band.



(a)

(b)

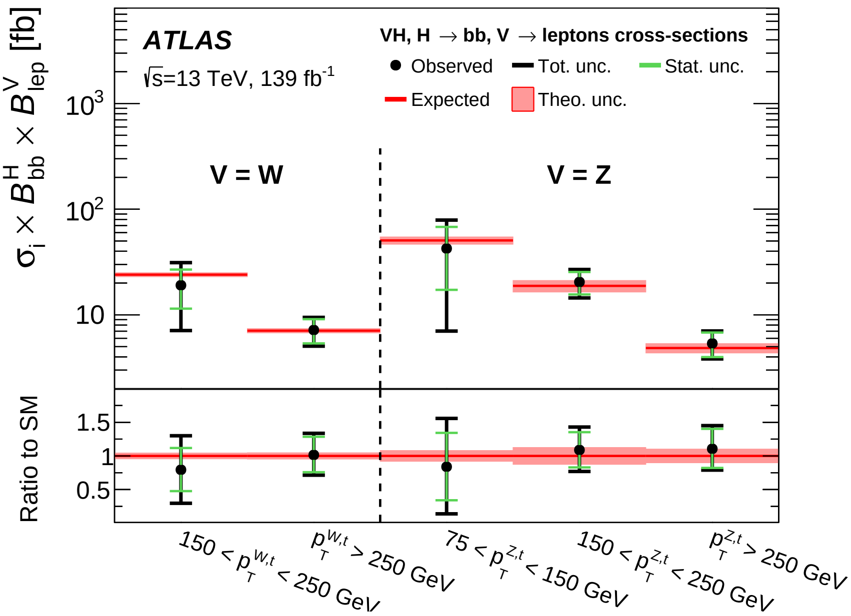
Figure 7.10: (a) Results for the signal strengths μ^{bb}_{WH} and μ^{bb}_{ZH} from a simultaneous fit, as well as the inclusive signal strength μ^{bb}_{VH}. (b) Two-dimensional confidence intervals for the signal-strength parameters μ^{bb}_{ZH} and μ^{bb}_{WH}, at 68% and 95% confidence level (CL). The best-fit value is marked by a cross, and the SM expectation is indicated by a blue star.

uncertainty. Both components contribute about equally to the results for the WH and ZH signal strengths. The most important experimental systematic uncertainties originate from the data-driven measurement of the b-tagging efficiencies for b- and c-jets, the calibration of the jet-energy scale, and the reconstruction of E_T^{miss}. For ZH, the dominant modelling uncertainties originate from scale uncertainties of the gluon-induced signal process. Modelling uncertainties on the W+HF background are the leading contribution for WH.

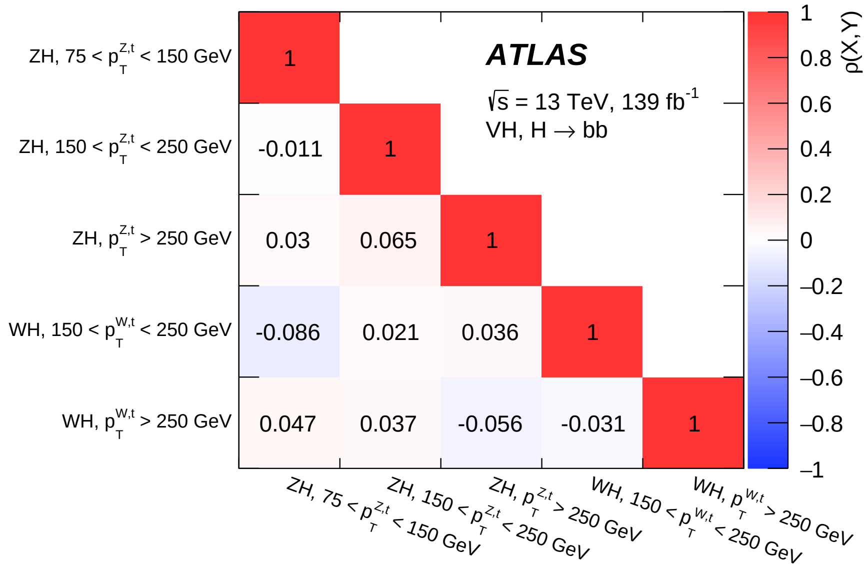
The three lepton channels contribute approximately equally to the result for VH, with the 2-lepton channel showing slightly lower systematic uncertainties and higher statistical uncertainties compared to the 0- and 1-lepton channels. The analysis categories selecting $p_T^V > 250$ GeV are the most sensitive.

7.8.3 STXS measurements

Figure 7.11a shows the measured cross-sections in the five considered STXS bins (cf. Figure 7.5). They are obtained as the products of the signal-strength parameters μ^b with the SM cross-section for leptonic VH production in STXS bin b and the $H \to b\bar{b}$ branching fraction, $\left[\sigma^b_{V(\text{lep.})H} \times B(H \to b\bar{b})\right]^{\text{SM}} = \left[\sigma^b_{VH} \times B(H \to b\bar{b}) \times B(V \to \text{lep.})\right]^{\text{SM}}$. The STXS definition includes decays into all three generations of charged leptons; these are extrapolated from the measurements targeting electrons and muons. All results are compatible with the SM expectation. The Pearson correlations between



(a)



(b)

Figure 7.11: (a) Measured fiducial cross-sections $\sigma^{b}_{V(\text{lep.})H} \times B(H \to b\bar{b})$ in the five considered STXS bins. The SM cross-sections and uncertainties in the theoretical predictions are indicated by the red bars. (b) Pearson correlation coefficients between the extracted STXS values.

Process	STXS bin $p_T^{V,t}$ region	SM prediction [fb]			Exp. result [fb]			Stat. unc. [fb]		Syst. unc. [fb] Th. sig.		Th. bkg.		Exp.	
WH	[150,250) GeV	24.0	±	1.1	19.0	±	12.1	±	7.7	±	0.9	±	5.5	±	6.0
WH	>250 GeV	7.1	±	0.3	7.2	±	2.2	±	1.9	±	0.4	±	0.8	±	0.7
ZH	[75,150) GeV	50.6	±	4.1	42.5	±	35.9	±	25.3	±	5.6	±	17.2	±	19.7
ZH	[150,250) GeV	18.8	±	2.4	20.5	±	6.2	±	5.0	±	2.3	±	2.4	±	2.3
ZH	>250 GeV	4.9	±	0.5	5.4	±	1.7	±	1.5	±	0.5	±	0.5	±	0.3

Table 7.8: Numerical values of the measured cross-sections $\sigma^b_{V(\text{lep.})H} \times B(H \to b\bar{b})$ in the five targeted STXS bins, compared to the SM prediction and its theoretical uncertainties. The total uncertainty on the experimental result is split into a statistical uncertainty (Stat. unc.), systematic components resulting from the modelling of the signal (Th. sig.) and the backgrounds (Th. bkg.), as well as experimental sources (Exp.). Adapted from Ref. [5].

the extracted cross-sections are shown in Figure 7.11b. The correlations are below 10% in all cases, most substantially so, reflecting the close correspondence between the analysis categories and their targeted STXS signals (cf. Figure 7.7). In some cases correlations are reduced by the presence of NPs and thus reach levels that are *below* those expected from the residual STXS signal contamination in adjacent analysis regions.

Table 7.8 lists the numerical values of the measured cross-sections, with the total uncertainties split into a statistical component (defined as above), as well as systematic uncertainties originating from experimental sources and the modelling of the signal and the backgrounds. The total uncertainty on the measurement ranges from around 85% in the lowest $p_T^{Z,t}$ bin to about 30% in the highest $p_T^{W,t}$ bin. The precision is limited by statistical uncertainties in all bins. Signal modelling uncertainties are substantially larger for ZH compared to WH due to the limited theoretical knowledge of the $gg \to ZH$ process.

A more detailed breakdown of the uncertainties is available in Table 7.9. The systematic uncertainty attributed to a particular subset of NPs is defined as the difference in quadrature between the total uncertainty from the nominal measurement and the uncertainty observed when all NPs in the subset are fixed to their MLEs. The total statistical uncertainty is further decomposed into the intrinsic uncertainty due to the limited size of the recorded data set ("Data statistical"), as well as the components from floating background normalisations and the limited statistics in the $t\bar{t}$ $e\mu$-CR used in the 2-lepton channel. The assigned total uncertainty differs from the sum in quadrature of the individual components due to the presence of nontrivial correlations between the NPs.

The contribution from E_T^{miss} is among the most important sources of experimental uncertainties

Source of uncertainty	Uncertainty on $\mu^b = \sigma^b_{V(\text{lep.})H} \times B(H \to b\bar{b}) / \left[\sigma^b_{V(\text{lep.})H} \times B(H \to b\bar{b})\right]^{\text{SM}}$				
	WH $p_T^{W,t} \in [150, 250)$ GeV	WH $p_T^{W,t} \in [250, \infty)$ GeV	ZH $p_T^{Z,t} \in [75, 150)$ GeV	ZH $p_T^{Z,t} \in [150, 250)$ GeV	ZH $p_T^{Z,t} \in [250, \infty)$ GeV
Total	0.502	0.311	0.710	0.330	0.330
Statistical	0.320	0.263	0.501	0.262	0.291
Systematic	0.386	0.166	0.503	0.200	0.156
Statistical uncertainties					
Data statistical	0.298	0.252	0.421	0.243	0.284
$t\bar{t}\ e\mu$ control regions	0.032	0.007	0.221	0.039	0.023
Floating normalisations	0.157	0.050	0.181	0.095	0.047
Experimental uncertainties					
Jets	0.145	0.054	0.266	0.082	0.040
E_T^{miss}	0.171	0.009	0.235	0.027	0.016
Leptons	0.019	0.018	0.027	0.007	0.007
b-tagging: b-jets	0.049	0.023	0.176	0.082	0.041
b-tagging: c-jets	0.109	0.060	0.028	0.020	0.006
b-tagging: light-flavour jets	0.004	0.005	0.006	0.013	0.015
Pile-up	0.017	0.006	0.046	0.009	0.003
Luminosity	0.017	0.015	0.012	0.016	0.017
Theoretical and modelling uncertainties					
Signal	0.035	0.050	0.110	0.096	0.091
Z + jets	0.038	0.011	0.271	0.089	0.071
W + jets	0.159	0.072	0.020	0.019	0.008
$t\bar{t}$	0.152	0.037	0.108	0.036	0.025
Single top quark	0.135	0.032	0.044	0.015	0.015
Diboson	0.040	0.034	0.073	0.044	0.029
Multi-jet	0.015	0.019	0.009	0.008	0.005
MC statistical	0.112	0.068	0.168	0.057	0.055

Table 7.9: Breakdown of the contributions to the uncertainty on μ^b for the five considered STXS bins. The total uncertainty is decomposed into a statistical uncertainty as well as systematic components (defined in the main text) originating from the modelling of signal and background components, and from experimental sources. All uncertainties are taken to be symmetric.

in the $p_T^{W,t} \in [150, 250)$ GeV bin, originating from the $p_T^{\text{miss,st}}$ contribution and the important role of this variable in the multivariate discriminant. A similar impact exists for the $p_T^{Z,t} \in [75, 150)$ GeV bin, which emerges from the uncertainty on $E_T^{\text{miss}}/\sqrt{S_T}$, also used as input variable to BDT$_{VH}$. Uncertainties from the tagging efficiency for c-jets are more important for WH compared to ZH due to the presence of a larger sample of events with $W \to cq$ decays relative to $Z \to c\bar{c}$.

7.9 Constraints on effective interactions

To explore the consequences of this result for the structure of interactions at energy scales $\Lambda \gg \sqrt{\hat{s}}$, the STXS measurement is reinterpreted in terms of constraints on Wilson coefficients of $d = 6$ operators in the SMEFT.

Based on the formalism established in Section 5.3, Section 7.9.1 defines a set of operators that modify the VH production dynamics or the $H \to b\bar{b}$ decay. Sections 7.9.2 and 7.9.3 derive SMEFT predictions for the cross-sections in the STXS fiducial volumes in terms of the Wilson coefficients of these operators. This parameterisation is incorporated into the statistical model in Section 7.9.4, and the resulting operator constraints are summarised in Section 7.9.5.

7.9.1 Selection of operators

In agreement with the definition of the STXS regions, the Higgs boson is assumed to be on-shell in what follows, i.e. modifications to the $V(\to \text{lep.})H$ production cross-section and the $H \to b\bar{b}$ decay arise separately. Only CP-even operators in the $U(3)^5$-symmetric SMEFT are considered. CP-odd operators do not modify the distribution of the CP-even kinematic variable p_T^V that defines the binning of the analysis, and cannot therefore be constrained by the present measurement.

Of the remaining 58 operators, 17 were found to modify the signal process at leading order [324]. The $V(\to \text{lep.})H$ production cross-section receives corrections from the 13 operators listed in Table 7.10. A subset of these operators also alters the branching fraction $B(H \to b\bar{b})$.

A further four operators, $\mathcal{Q}_{dH}$, $\mathcal{Q}_{eH}$, $\mathcal{Q}_{uH}$, and $\mathcal{Q}_{HG}$, exclusively affect $B(H \to b\bar{b})$, either by modifying the partial or the total decay width. They are listed in Table 7.11. Their effect on the measured observables is completely degenerate, and only $\mathcal{Q}_{dH}$ was retained for further analysis. This operator directly modifies the partial width of the $H \to b\bar{b}$ decay.

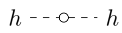
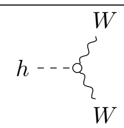
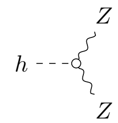
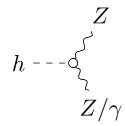
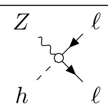
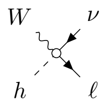
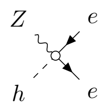
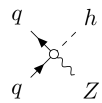
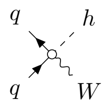
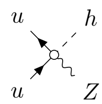
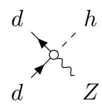
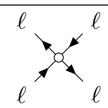
	Operator	Wilson coefficient	Operator definition	Example vertex
D^2H^4	$\mathcal{Q}_{H\Box}$	$c_{H\Box}$	$(H^\dagger H)\partial_\mu\partial^\mu(H^\dagger H)$	
	$\mathcal{Q}_{HD}$	c_{HD}	$(H^\dagger D^\mu H)^*(H^\dagger D_\mu H)$	
F^2H^2	$\mathcal{Q}_{HW}$	c_{HW}	$H^\dagger H W^I_{\mu\nu}W^{\mu\nu I}$	
	$\mathcal{Q}_{HB}$	c_{HB}	$H^\dagger H B_{\mu\nu}B^{\mu\nu}$	
	$\mathcal{Q}_{HWB}$	c_{HWB}	$H^\dagger\sigma^I H W^I_{\mu\nu}B^{\mu\nu}$	
$D\psi^2H^2$	$\mathcal{Q}^{(1)}_{H\ell}$	$c^{(1)}_{H\ell}$	$(H^\dagger i\overleftrightarrow{D}_\mu H)(\bar{L}_{Ls}\gamma^\mu L_{Ls})$	
	$\mathcal{Q}^{(3)}_{H\ell}$	$c^{(3)}_{H\ell}$	$(H^\dagger i\overleftrightarrow{D}^I_\mu H)(\bar{L}_{Ls}\sigma^I\gamma^\mu L_{Ls})$	
	$\mathcal{Q}_{He}$	c_{He}	$(H^\dagger i\overleftrightarrow{D}_\mu H)(\bar{e}_{Rs}\gamma^\mu e_{Rs})$	
	$\mathcal{Q}^{(1)}_{Hq}$	$c^{(1)}_{Hq}$	$(H^\dagger i\overleftrightarrow{D}_\mu H)(\bar{Q}_{Ls}\gamma^\mu Q_{Ls})$	
	$\mathcal{Q}^{(3)}_{Hq}$	$c^{(3)}_{Hq}$	$(H^\dagger i\overleftrightarrow{D}^I_\mu H)(\bar{Q}_{Ls}\sigma^I\gamma^\mu Q_{Ls})$	
	$\mathcal{Q}_{Hu}$	c_{Hu}	$(H^\dagger i\overleftrightarrow{D}_\mu H)(\bar{u}_{Rs}\gamma^\mu u_{Rs})$	
	$\mathcal{Q}_{Hd}$	c_{Hd}	$(H^\dagger i\overleftrightarrow{D}_\mu H)(\bar{d}_{Rs}\gamma^\mu d_{Rs})$	
$\bar\psi\psi\bar\psi\psi$	$\mathcal{Q}^{(1)}_{\ell\ell}$	$c^{(1)}_{\ell\ell}$	$(\bar{L}_{Lp}\gamma_\mu L_{Ls})(\bar{L}_{Ls}\gamma^\mu L_{Lp})$	

Table 7.10: Operators with $d = 6$ that affect the VH production cross-section or the $H \to b\bar{b}$ decay, along with their respective Wilson coefficients. Sums over repeated flavour indices in the operator definitions are implied. The last column shows example interaction vertices contained in the $U(3)^5$-symmetric SMEFT in terms of mass eigenstates.

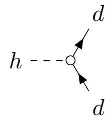
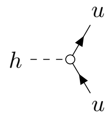
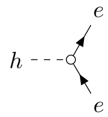
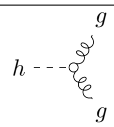
	Operator	Wilson coefficient	Operator definition	Example vertex
$\psi^2 H^3$	$\mathcal{Q}_{dH}$	c_{dH}	$\left[Y_d^\dagger\right]_{ps}(H^\dagger H)(\bar{Q}_{Lp} d_{Rs} H)$	
	$\mathcal{Q}_{uH}$	c_{uH}	$\left[Y_u^\dagger\right]_{ps}(H^\dagger H)(\bar{Q}_{Lp} u_{Rs} \tilde{H})$	
	$\mathcal{Q}_{eH}$	c_{eH}	$\left[Y_e^\dagger\right]_{ps}(H^\dagger H)(\bar{L}_{Lp} e_{Rs} H)$	
$F^2 H^2$	$\mathcal{Q}_{HG}$	c_{HG}	$H^\dagger H G^A_{\mu\nu} G^{\mu\nu A}$	

Table 7.11: Operators with $d = 6$ that exclusively affect the branching ratio of the $H \to \bar{b}b$ decay, along with their respective Wilson coefficients. Sums over repeated flavour indices in the operator definitions are implied. The last column shows example interaction vertices contained in the $U(3)^5$-symmetric SMEFT in terms of mass eigenstates.

7.9.2 Parameterisation of modifications to observables

The Lagrangian $\mathcal{L}_{\text{SMEFT}}$ in Eq. 5.4 allows predictions for arbitrary observables to be made. To $\mathcal{O}(\Lambda^{-4})$, a generic scattering amplitude $\mathcal{M}$ has the structure

$$\mathcal{M} = \mathcal{M}_{\text{SM}} + \sum_i \frac{c_i^{(6)}}{\Lambda^2}\mathcal{M}_i^{(6)} + \sum_{ij} \frac{c_i^{(6)} c_j^{(6)}}{\Lambda^4}\mathcal{M}_{ij}^{(6)} + \sum_i \frac{c_i^{(8)}}{\Lambda^4}\mathcal{M}_i^{(8)}. \tag{7.4}$$

The amplitude $\mathcal{M}_{\text{SM}}$ is computed from $\mathcal{L}_{\text{SM}}$, $\mathcal{M}_i^{(6)}$ contains exactly one insertion of the operator $\mathcal{Q}_i^{(6)}$, and $\mathcal{M}_{ij}^{(6)}$ contains a double insertion of $\mathcal{Q}_i^{(6)}$ and $\mathcal{Q}_j^{(6)}$. The last term arises at $d = 8$. The expansion in Eq. 7.4 is terminated at second order, implying that contributions from the operators $\mathcal{Q}_i^{(6)}$ are not resummed, e.g. into corrections to the total widths of SM states. This would lead to a more general dependence of $\mathcal{M}$ on the Wilson coefficients.

A cross-section σ depends on the Wilson coefficients through the corresponding matrix element-squared. To second order in Λ, $|\mathcal{M}|^2$ reads

$$\begin{aligned}|\mathcal{M}|^2 = |\mathcal{M}_{\text{SM}}|^2 &+ \frac{2}{\Lambda^2}\sum_i \text{Re}\left(c_i^{(6)}\mathcal{M}_i^{(6)}\mathcal{M}_{\text{SM}}^*\right) + \frac{1}{\Lambda^4}\Big|\sum_i c_i^{(6)}\mathcal{M}_i^{(6)}\Big|^2 \\ &+ \frac{2}{\Lambda^4}\sum_{ij}\text{Re}\left(c_i^{(6)}c_j^{(6)}\mathcal{M}_{\text{SM}}^*\mathcal{M}_{ij}^{(6)}\right) + \frac{2}{\Lambda^4}\sum_i \text{Re}\left(c_i^{(8)}\mathcal{M}_i^{(8)}\mathcal{M}_{\text{SM}}^*\right).\end{aligned} \tag{7.5}$$

To $\mathcal{O}(\Lambda^{-2})$, only the interference contributions in the first line of Eq. 7.5 are relevant. They depend linearly on the $c_i^{(6)}$ and are fully determined at $d = 6$. These terms are suppressed as $(E/\Lambda)^2$ and thus typically contain the dominant modifications to the SM cross-section.

In certain scenarios, however, the terms of $\mathcal{O}(\Lambda^{-4})$ might represent the leading contribution, despite formally being of higher order in the scale ratio and quadratic in the Wilson coefficients. This can be the case when the SM amplitude itself is suppressed, e.g. in a certain kinematic limit or when the presence of selection rules forbids the interference contributions altogether [325]. Terms of this order are not fully determined in terms of the $c_i^{(6)}$ alone. Diagrams with double insertions of $d = 6$ operators are renormalised by $d = 8$ interactions, although they are entirely defined in terms of the $c_i^{(6)}$ at tree-level. The $d = 8$ matrix elements also contribute directly through their interference with the SM amplitude.

A complete calculation of these terms is beyond the reach of presently available SMEFT MC codes. It is nevertheless crucial to understand the relative importance of the first- and second order corrections for the process at hand. The studies in Refs. [326, 327] suggest using the contribution from $\left|\mathcal{M}_i^{(6)}\right|^2$ to estimate the generic size of the remaining $\mathcal{O}(\Lambda^{-4})$ terms and evaluate their importance for the resulting operator constraints. In the following discussion, only the terms in the first line of Eq. 7.5 are therefore retained. These are individually gauge invariant and may be translated between different operator bases without ambiguity [328]. Comparisons are made with a fully linearised calculation in which only the interference contributions are considered.

7.9.2.1 Parameterisation of the production cross-section

With these assumptions, the leptonic VH production cross-section in STXS bin b, $\sigma^b_{V(\text{lep.})H}$, depends polynomially on the Wilson coefficients $c_i^{(6)}$,

$$\sigma^b_{V(\text{lep.})H} = \left[\sigma^b_{V(\text{lep.})H}\right]^{\text{SM}} \left(1 + \sum_i \alpha_i^b c_i^{(6)} + \sum_{ij} \beta_{ij}^b c_i^{(6)} c_j^{(6)}\right). \tag{7.6}$$

The cross-section $\left[\sigma^b_{V(\text{lep.})H}\right]^{\text{SM}}$ is computed with the SM Lagrangian. The coefficients α_i^b parameterise the interference of the SM amplitude with those containing a single $d = 6$ operator insertion. The parameterisation constants $\beta_{ij}^b = \beta_{ji}^b$ represent the partial $\mathcal{O}(\Lambda^{-4})$ contributions as explained above. Parameterisations of the type of Eq. 7.6 are referred to as "linear" if they include only interference terms and "linear + quadratic" if $\mathcal{O}(\Lambda^{-4})$ contributions are also present.

The parameterisation coefficients are obtained from MC calculations of the cross-section σ^b_{VH}

based on the Lagrangian $\mathcal{L}_{\text{SM}} + \mathcal{L}^{(6)}$. They depend implicitly on the scale Λ, set to $\Lambda = 1\,\text{TeV}$ by convention. This implies that also experimental constraints placed on the dimensionless Wilson coefficients $c_i^{(6)}$ are subject to this artificial scale dependence. They may easily be converted to limits at a different scale Λ' through the relation $c_i^{(6)}/\Lambda^2 = \left(c_i^{(6)}/\Lambda^2\right)'$. This is only admissible, of course, provided that the assumption $E/\Lambda' \ll 1$ remains valid.

Practical details of the calculation MadGraph 5 (v2.6.5) [306] was employed as matrix-element generator, together with the implementation of the Warsaw basis available in the SMEFTsim package (v2.1) [176]. The parameters in the electroweak SMEFT Lagrangian were determined based on the input parameters m_W, m_Z, m_H, and G_F, as explained in Section 5.3.2.3. Pythia 8 (v8.240) [239] acts as parton shower (PS) generator, and Rivet [329] is used to classify the generated events according to the definition of the STXS fiducial volumes. Only information extracted from the truth record of the generator is necessary and the simulation of the ATLAS detector is not included. This is further justified in Section 7.9.3.

The matrix elements for the processes $qq \to H\ell^+\ell^-$, $qq \to H\nu\bar{\nu}$, and $qq \to H\ell\nu$ are computed at LO in QCD. All three charged lepton and neutrino flavours are included. Off-shell W and Z bosons are explicitly allowed. A consistent computation of the loop-induced $gg \to ZH$ production mode in the SMEFT is not possible with these tools, and this process is not considered. The widths of all particles are fixed to their SM values computed at LO to ensure a manifestly polynomial dependence of the cross-section on the Wilson coefficients. To improve the efficiency of the calculation, the masses of the first two generations of quarks and leptons are set to zero. The CKM matrix is assumed to be diagonal. All generator-level cuts are aligned with the ATLAS recommendations published in Ref. [324].

Eq. 7.6 defines a linear system of equations for α_i^b and β_{ij}^b, provided that computations of $\sigma_{V(\text{lep.})H}^b$ for sufficiently many distinct sets of Wilson coefficients are available. The described generator setup allows the SM cross-section as well as the linear and quadratic modifications to be computed separately, significantly reducing the number of MC samples necessary. The size of each simulated sample is chosen such that the relative statistical uncertainty on the parameterisation coefficients is not larger than 10^{-3}. This ensures that the parameterisation in Eq. 7.6 is accurate for Wilson coefficients $c_i^{(6)} \lesssim \sqrt{10^3} \approx 30$, far exceeding the regime $c_i^{(6)} \approx 1$ expected for a natural theory.

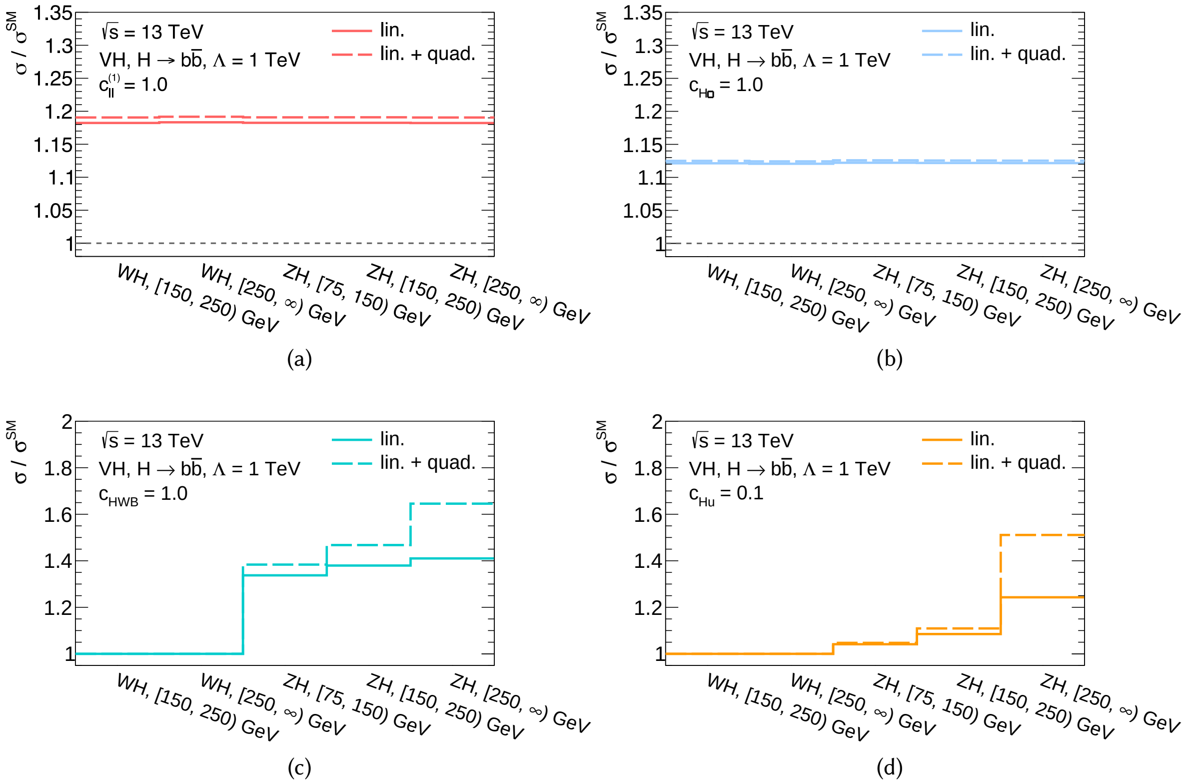


Figure 7.12: Relative modification of the VH production cross-section $\sigma^{b}_{V(\text{lep.})H}$ for STXS bin b in the SMEFT for canonical values of the Wilson coefficients of individual operators, $c^{(1)}_{\ell\ell} = 1$ in (a), $c_{H\Box} = 1$ in (b), $c_{HWB} = 1$ in (c), and $c_{Hu} = 0.1$ in (d). Effects are shown when only the linear interference terms are parameterised (thick lines) and for the case where also the quadratic contributions are included (dashed lines).

Characteristics of the modifications As discussed following Eq. 7.5, the contribution of an operator $\mathcal{Q}^{(6)}_i$ to the cross-section generically scales as E^2/Λ^2. When embedded in a particular process, however, this leading component may not contribute and only a more modest growth with the kinematic scale may remain. This is illustrated in Figure 7.12 for operators modifying $\sigma^{b}_{V(\text{lep.})H}$.

The four-fermion operator $\mathcal{Q}^{(1)}_{\ell\ell}$ modifies the relationship between G_F and the electroweak coupling g_2, cf. Eqs. 5.7a and 5.7b. Its Wilson coefficient $c^{(1)}_{\ell\ell}$ thus leads to an overall scaling of the WH and ZH cross-sections independent of the event kinematics (cf. Figure 7.12a). The operator $\mathcal{Q}_{H\Box}$ enters in a similar manner through the factor Δ_H defined in Eq. 5.5, as shown in Figure 7.12b.

The term $\mathcal{Q}_{HWB}$ modifies the HZZ coupling both directly and through its appearance in the modified weak mixing angle in Eq. 5.6. Depending on the polarisation state of the Z boson, this leads to an energy-independent contribution (longitudinal polarisation), or a linearly growing cross-section (transverse polarisation). The polarisation-averaged impact is shown in Figure 7.12c. The operator $\mathcal{Q}_{HW}$ shows similar characteristics but affects both WH and ZH [330].

Operators of the type $D\psi^2 H^2$ such as $\mathcal{Q}_{Hq}^{(3)}$, $\mathcal{Q}_{Hq}^{(1)}$, $\mathcal{Q}_{Hu}$, and $\mathcal{Q}_{Hd}$ create genuine $qqZH$ and $qqWH$ contact interactions. In this case, the leading energy dependence remains and the modifications of the cross-sections increase quadratically with the kinematic scale p_T^V. The inclusion of the $\mathcal{O}(\Lambda^{-4})$ term in the parameterisation further enhances the magnitude of these changes. The operator $\mathcal{Q}_{Hu}$ is chosen as a representative, and its impact is shown in Figure 7.12d. The remaining operators in this class produce very similar effects.

7.9.2.2 Parameterisation of the $H \to b\bar{b}$ decay

In analogy to Eq. 7.6, the Higgs-boson decay width into the final-state f has the structure

$$\Gamma_{H\to f} = [\Gamma_{H\to f}]^{\mathsf{SM}} \left(1 + \sum_i A_i^f c_i^{(6)} + \sum_i B_{ij}^f c_i^{(6)} c_j^{(6)}\right), \tag{7.7}$$

with numerical coefficients A_i^f and $B_{ij}^f = B_{ji}^f$. For a set $\{f\}$ of final states, the Higgs-boson total width Γ_H may be approximated as the sum of the respective partial widths, $\Gamma_H = \sum_{\{f\}} \Gamma_{H\to f}$. With Eq. 7.7, this generates the parameterisation

$$\Gamma_H = [\Gamma_H]^{\mathsf{SM}} \left(1 + \sum_i A_i c_i^{(6)} + \sum_{ij} B_{ij} c_i^{(6)} c_j^{(6)}\right), \tag{7.8}$$

where the coefficients A_i and B_{ij} are given by

$$A_i = \sum_{\{f\}} \left[\frac{\Gamma_{H\to f}}{\Gamma_H}\right]^{\mathsf{SM}} A_i^f, \qquad B_{ij} = \sum_{\{f\}} \left[\frac{\Gamma_{H\to f}}{\Gamma_H}\right]^{\mathsf{SM}} B_{ij}^f.$$

All numerical coefficients were again obtained from MC calculations of the partial and total decay widths as explained in Section 7.9.2.1. Using the relations in Eqs. 7.7 and 7.8, the branching fraction $B(H\to f) = \Gamma_{H\to f}/\Gamma_H$ may be expressed as a rational function of the Wilson coefficients,

$$B(H\to f) = [B(H\to f)]^{\mathsf{SM}} \frac{1 + \sum_i A_i^f c_i^{(6)} + \sum_{ij} B_{ij}^f c_i^{(6)} c_j^{(6)}}{1 + \sum_i A_i c_i^{(6)} + \sum_{ij} B_{ij} c_i^{(6)} c_j^{(6)}}. \tag{7.9}$$

Again, Eq. 7.9 implies a "linear" parameterisation in which the terms depending on B_{ij} and B_{ij}^f are dropped. The additional approximation of linearising the remaining expression is not made, i.e. a rational dependence in terms of the Wilson coefficients is retained also in this case.

Practical details of the calculation The generator setup is very similar to the one described in Section 7.9.2.1. MadGraph is used to integrate the decay matrix elements and compute the inclusive decay widths $\Gamma_{H\to f}$ and Γ_H. Parton showering and STXS classification are not necessary.

The total width Γ_H is computed as the sum of the partial decay widths that are most important for the SM Higgs boson. This includes two-body decays into fermions, $H \to f\bar{f}$ where $f \in \{b, c, s, \tau, \mu, e\}$, as well as the bosonic decays into two photons, two gluons, and $H \to Z\gamma$. The following four-body decays were considered: $4q$, 4ℓ, 4ν, $2q2\ell$, $2q2\nu$, $2q\ell\nu$, and $2\nu2\ell$, where $q \in \{b, c, s, d, u\}$ and all lepton and neutrino flavours are included. The decays into electroweak gauge bosons, $H \to WW$ and $H \to ZZ$, are implicitly included.

7.9.2.3 Parameterisation of production cross-section times branching fraction

Relevant for the comparison with the measured observables is the product $\sigma^b_{V(\text{lep.})H} B(H \to b\bar{b})$. In terms of Eqs. 7.6 and 7.9, its parameterisation is given by

$$\sigma^b_{V(\text{lep.})H} B(H \to b\bar{b}) = \left[\sigma^b_{V(\text{lep.})H} B(H \to b\bar{b})\right]^{\text{SM}} \left(1 + \sum_i \alpha^b_i c^{(6)}_i + \sum_{ij} \beta^b_{ij} c^{(6)}_i c^{(6)}_j\right) \times \frac{1 + \sum_i A^f_i c^{(6)}_i + \sum_{ij} B^f_{ij} c^{(6)}_i c^{(6)}_j}{1 + \sum_i A_i c^{(6)}_i + \sum_{ij} B_{ij} c^{(6)}_i c^{(6)}_j}, \quad (7.10)$$

and the corresponding "linear" parameterisation is constructed as described above.

It is convenient for the discussion in Section 7.9.5.3 to also introduce a parameterisation which is manifestly linear in the Wilson coefficients and absorbs all branching-fraction modifications into a separate parameter $\Delta B(H \to b\bar{b}) / \left[B(H \to b\bar{b})\right]^{\text{SM}}$ which also acts linearly. This quantity does not play the role of a Wilson coefficient (as it does not multiply a gauge-invariant operator) but is more akin to a phenomenological parameter that can be experimentally constrained. For this case, the parameterisation in Eq. 7.10 reduces to

$$\sigma^b_{V(\text{lep.})H} B(H \to b\bar{b}) = \left[\sigma^b_{V(\text{lep.})H} B(H \to b\bar{b})\right]^{\text{SM}} \left(1 + \sum_i \alpha^b_i c^{(6)}_i + \frac{\Delta B(H \to b\bar{b})}{\left[B(H \to b\bar{b})\right]^{\text{SM}}}\right). \quad (7.11)$$

7.9.3 Modifications of acceptance and multivariate discriminant

The above parameterisations capture changes in the overall cross-sections in the STXS fiducial volumes. These are not directly observable but experimentally accessible only through the STXS signal strengths μ^b in terms of which the statistical model L^{STXS}, designed and optimised for the SM $VH(\to b\bar{b})$ process, is formulated. SMEFT modifications of the STXS are expected to be good representations of effects on the signal strengths only if the behaviour of the analysis procedure itself is not altered appreciably.

This is the case if residual modifications to the p_T^V distribution within each STXS bin are small, and effects on kinematic variables other than p_T^V are such that the distribution of the multivariate discriminant used in the fit is not impacted in any significant way. Also, modifications to the signal acceptance of the analysis must be subleading. These could either enhance or suppress changes in the observable signal yield and thus deviate from the expectation encoded in the parameterisations of the STXS. These aspects are investigated below for the operator to which the analysis shows the highest expected sensitivity, $\mathcal{Q}_{Hq}^{(3)}$.

Modifications to kinematic distributions and the multivariate discriminant Modifications to kinematic distributions are studied in a simplified setting, where the analysis selection is implemented based on information from the truth record of the generator and reconstruction effects are absent. Samples of simulated events are generated following the procedure outlined in Section 7.9.2.1 and the distributions of important event variables studied in individual analysis categories.

Figures 7.13a–7.13c show modifications to the distributions of selected kinematic quantities when the Wilson coefficient $c_{Hq}^{(3)}$ is placed at its expected $\pm 1\sigma$ limits, obtained with the statistical model described in Section 7.9.4. The distributions of the dimensionful observables $p_{T,b1}$ and m_{top} are modified significantly, following the characteristic modifications of the p_T^V spectrum caused by the induced contact interaction. As $\mathcal{Q}_{Hq}^{(3)}$ does not couple to leptons, the angular observable $\cos\theta(\boldsymbol{\ell}^-, \boldsymbol{Z})$ in the 2-lepton channel is unaffected.

Figure 7.13d shows the effects on the distribution of the multivariate discriminant BDT_{VH}, using the same binning that is employed in the statistical analysis described in Section 7.7.3.1. Variables that are not well-defined based on truth quantities alone but required as inputs to the BDT are sampled from their respective marginal distributions determined from the full detector simulation, similar to the procedure used in the data-driven estimate of the multi-jet background (cf. Section 7.7.1.3). This concerns $p_T^{\text{miss,st}}$, MV2($\mathbf{b}_1$), and MV2($\mathbf{b}_2$). These quantities are only very weakly correlated with each other and with the remainder of the inputs. The impact of $\mathcal{Q}_{Hq}^{(3)}$ on the most signal-enriched bins is of the order of 10% in all analysis regions. Identical studies performed for $\mathcal{Q}_{Hu}$, $\mathcal{Q}_{HW}$, and $\mathcal{Q}_{HWB}$ show that these important operators introduce similar modifications of the BDT_{VH} distribution.

Modifications to analysis acceptance SMEFT modifications of the signal acceptance are evaluated in a similar manner. The signal event yield in each analysis category, computed based on

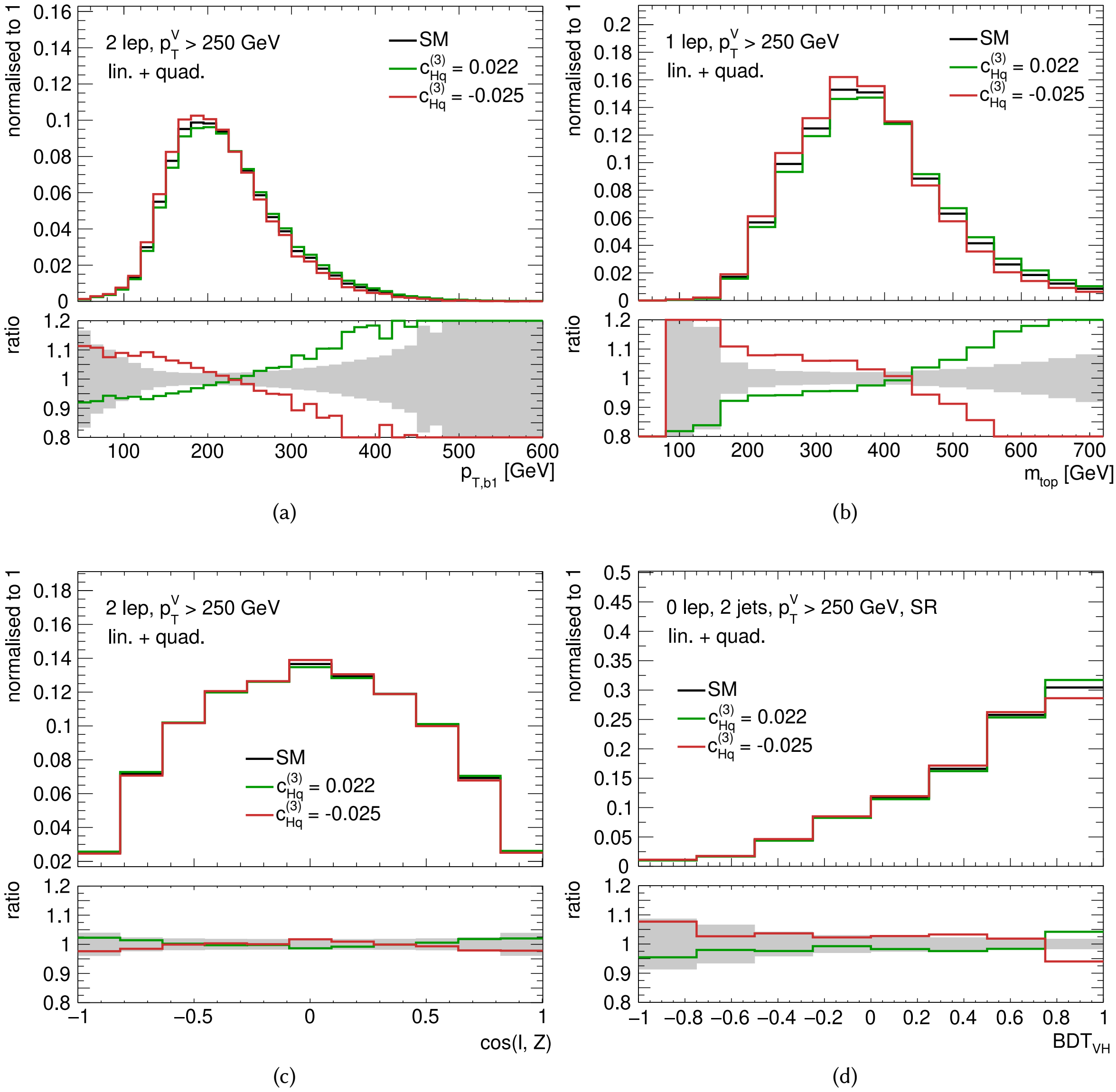


Figure 7.13: SMEFT modifications to the distributions of kinematic variables in different analysis categories, shown for the expected $\pm 1\sigma$ limits of the Wilson coefficient $c_{Hq}^{(3)}$, for $p_{T,b1}$ in (a), m_{top} in (b), and $\cos\theta(\ell^-, Z)$ in (c). The discriminant BDT$_{VH}$, computed as described in the main text, is shown in (d). The grey band corresponds to the statistical uncertainty of the sample of simulated events.

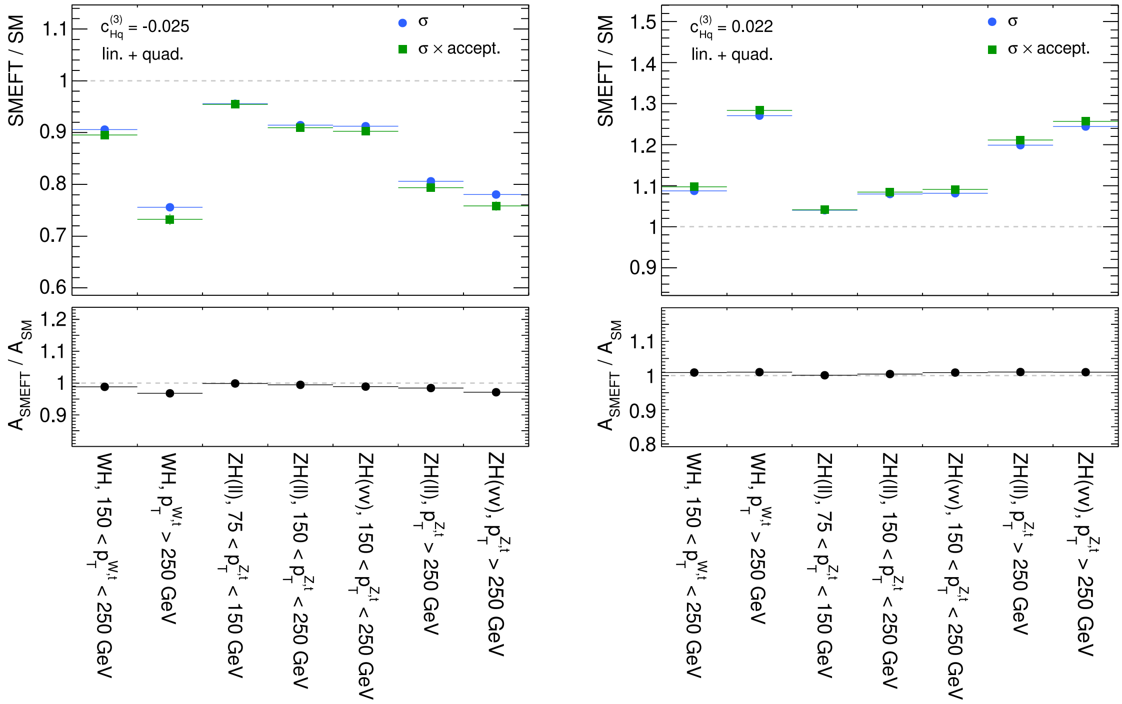


Figure 7.14: Modification of the fiducial cross-section in a specific truth fiducial volume (blue) and the observed event yield in the corresponding analysis category (green), when $c_{Hq}^{(3)}$ is set to its expected -1σ limit in (a) and $+1\sigma$ limit in (b). The horizontal axis refers to both a fiducial volume according to the STXS scheme, as well as the related analysis category. Both interference and quadratic terms are taken into account.

generator truth information as described above, is studied as a function of each Wilson coefficient. Comparing the evolution of this quantity to the parameterisation of the STXS allows the variation of the analysis acceptance to be studied.

Figure 7.14 quantifies these changes when $c_{Hq}^{(3)}$ is set to its expected $\pm 1\sigma$ limits. The impact on the scale of the analysis binning is minor and of the order of a few percent at most. A more detailed inspection shows that significant acceptance changes are visible in the extreme tails of the distribution for $p_T^V \gtrsim 600\,\text{GeV}$. In this regime, the analysis acceptance is dominated by the efficiency to reconstruct a highly boosted Higgs boson decay in the form of two isolated small-R jets. Operators such as $\mathcal{Q}_{Hq}^{(3)}$ or $\mathcal{Q}_{Hu}$ can significantly enhance these tails and lead to acceptance losses of about 10%–30% in the strongly boosted regime. Even in these extreme conditions, acceptance losses due to changes in kinematic variables other than p_T^V remain subleading. This shows an intrinsic limitation of the resolved reconstruction strategy, making a robust exploration of these high scales impossible. This issue is mitigated in Chapter 8.

7.9.4 Statistical model

A new likelihood function $L^{\text{EFT}}(\boldsymbol{c}^{(6)}, \boldsymbol{\theta})$ formulated in terms of Wilson coefficients was constructed based on $L^{\text{STXS}}(\boldsymbol{\mu}, \boldsymbol{\theta})$ described in Section 7.7.3. For notational convenience, the investigated Wilson coefficients are grouped into the vector $\boldsymbol{c}^{(6)}$.

As shown in Section 7.9.3, modifications of the shape of the template histograms and the signal acceptance are of the order of 10%. In addition, SMEFT modifications and higher-order QCD corrections are known to factorise to within about 10–15% for WH [331]. A similar factorisation is expected to hold for ZH. Working to this precision, the parameterisations determined in Section 7.9.2 at an accuracy of LO(QCD) + PS may be interpreted as scale factors for the nominal NLO(QCD) + PS VH signal templates encoded in L^{STXS}. SMEFT modifications generally also affect the normalisations and shapes of the background templates. These changes are expected to be largely absorbed by the floating normalisations and extrapolation uncertainties defined in the background model (cf. Section 7.7.2.2) and are thus not explicitly parameterised.

To construct L^{EFT}, the STXS signal-strength parameters μ^b in L^{STXS} are replaced by the parameterisations of $\sigma^b_{V(\text{lep.})H} B(H \to b\bar{b}) / \left[\sigma^b_{V(\text{lep.})H} B(H \to b\bar{b})\right]^{\text{SM}}$ in Eqs. 7.10 or 7.11. All nuisance parameters present in the original model were kept and theoretical uncertainties on the overall cross-sections in the STXS fiducial volumes are reintroduced. Confidence intervals for the Wilson coefficients are obtained in analogy to Eq. 7.3 with the test statistic

$$t(\boldsymbol{c}^{(6)}) = -2 \log \frac{L^{\text{EFT}}(\boldsymbol{c}^{(6)}, \hat{\hat{\boldsymbol{\theta}}})}{L^{\text{EFT}}(\hat{\boldsymbol{c}}^{(6)}, \hat{\boldsymbol{\theta}})},$$

where hatted quantities label unconditional MLEs and $\hat{\hat{\boldsymbol{\theta}}}$ is the MLE of the NPs conditional on $\boldsymbol{c}^{(6)}$.

7.9.5 Results

Beside the structural assumptions on the action of $d = 6$ operators discussed above, a separate question concerns the number of those operators that are allowed to contribute simultaneously. Different types of operator constraints are derived in the following to explore various choices.

7.9.5.1 One-dimensional limits

In the most restrictive case, only a single Wilson coefficient is assumed to be nonzero and the associated operator coherently modifies both the VH production and the $H \to b\bar{b}$ decay. Constraints

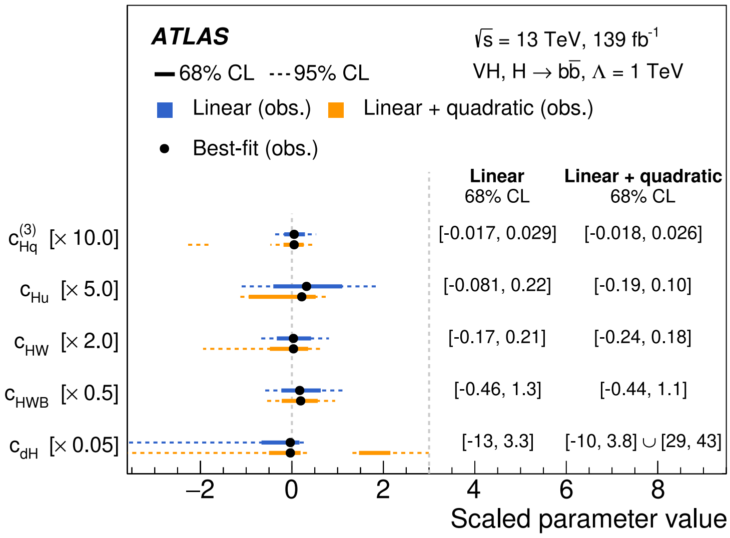


Figure 7.15: Summary of observed (obs.) one-dimensional confidence intervals at 68% and 95% CL for the most important Wilson coefficients. Results are shown for linear (blue) and linear + quadratic (orange) parameterisations.

of this type typically do not directly discriminate between particular models of local physics at the scale Λ but rather serve as useful benchmarks of the sensitivity of the analysis to various signatures.

Figure 7.15 summarises the one-dimensional confidence intervals for the operators to which the experimental sensitivity is greatest. The result for the parameter c_{dH}, to which this analysis is uniquely sensitive, is also included. The strength of the constraint on a particular Wilson coefficient depends significantly on the operator under consideration. Parameters such as $c_{Hq}^{(3)}$, c_{Hu}, or $c_{Hq}^{(1)}$ are constrained to levels of $\mathcal{O}(0.01\text{–}0.1)$ by this measurement, driven by the strong modifications of the cross-section in the $p_T^{V,t} > 250$ GeV STXS bins produced by the generated contact interactions. The likelihood profiles for $c_{Hq}^{(3)}$ and c_{Hu} are shown explicitly in Figure 7.16. The constraints on operators such as $\mathcal{Q}_{\ell\ell}^{(1)}$, which enter only indirectly, are significantly weaker. The one-dimensional confidence intervals for all investigated operators are listed in Table 7.12.

In many cases, the obtained constraints show considerable differences between linear and linear + quadratic parameterisations. This indicates sizeable $\mathcal{O}(\Lambda^{-4})$ contributions, compatible with the observations made below Figure 7.12.

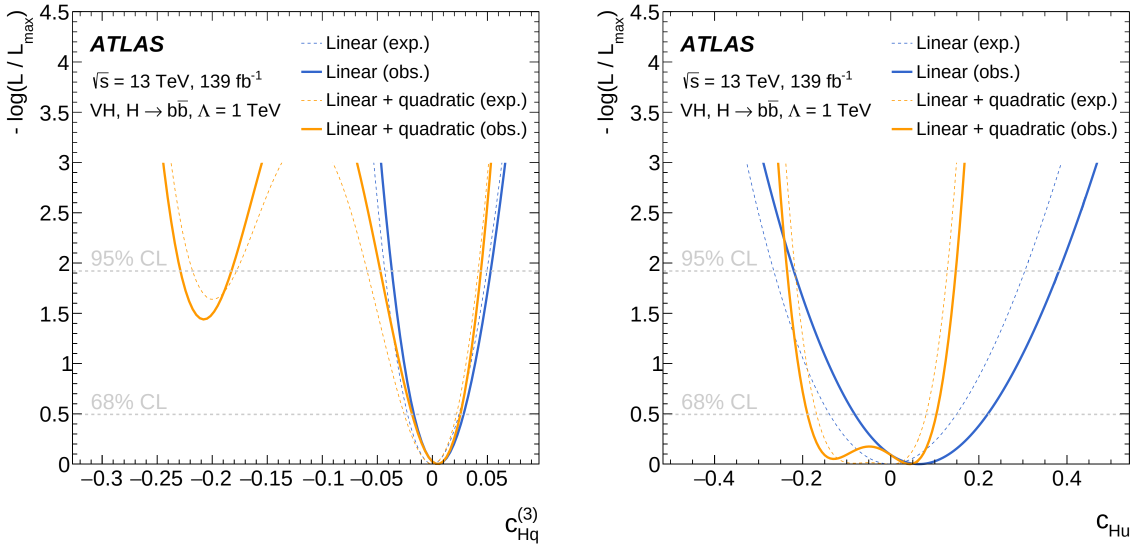
7.9.5.2 Two-dimensional limits

A natural relaxation of the above situation is to consider two-dimensional confidence intervals, i.e. to allow pairs of Wilson coefficients to be present at the same time. The form of the parameterisation itself is not modified, and modifications of both VH production and $H \to b\bar{b}$ decay are included.

Wilson coefficient	Linear		Linear + quadratic	
	Exp. (68% CL)	Obs. (68% CL)	Exp. (68% CL)	Obs. (68% CL)
c_{HWB}	[-0.75, 0.86]	[-0.46, 1.3]	[-0.66, 0.89]	[-0.44, 1.1]
c_{HW}	[-0.19, 0.2]	[-0.17, 0.21]	[-0.27, 0.17]	[-0.24, 0.18]
c_{Hu}	[-0.14, 0.15]	[-0.081, 0.22]	[-0.17, 0.08]	[-0.19, 0.1]
$c_{Hq}^{(3)}$	[-0.022, 0.024]	[-0.017, 0.029]	[-0.025, 0.022]	[-0.018, 0.026]
$c_{Hq}^{(1)}$	[-0.32, 0.29]	[-0.47, 0.18]	[-0.08, 0.11]	[-0.1, 0.13]
$c_{ll}^{(1)}$	[-1.2, 1.3]	[-1, 1.4]	[-1.2, 1.2]	[-1, 1.3]
$c_{Hl}^{(3)}$	[-0.99, 0.91]	[-1.1, 0.78]	[-0.93, 0.97]	[-1, 0.83]
$c_{Hl}^{(1)}$	[-6.2, 5.9]	[-8.6, 3.9]	[-4, 11]	[-5.1, 13]
c_{HD}	[-8.4, 8.2]	[-5.5, 9.8]	[-7.4, 5.7]	[-7, 7.3]
c_{Hd}	[-0.47, 0.44]	[-0.69, 0.26]	[-0.13, 0.18]	[-0.16, 0.21]
$c_{H\Box}$	[-1.6, 1.7]	[-1.3, 1.8]	[-1.6, 1.6]	[-1.4, 1.7]
c_{HB}	[-1.5, 1.4]	[-1.1, 1.8]	[-0.44, 0.58]	[-0.38, 0.56]
c_{He}	[-8.6, 8.3]	[-12, 5.2]	[-5.8, 17]	[-7.3, 19]
c_{dH}	[-11, 3.7]	[-13, 3.3]	[-8.8, 4.2] $\cup$ [29, 42]	[-10, 3.8] $\cup$ [29, 43]

Wilson coefficient	Linear		Linear + quadratic	
	Exp. (95% CL)	Obs. (95% CL)	Exp. (95% CL)	Obs. (95% CL)
c_{HWB}	[-1.4, 1.8]	[-1.2, 2.3]	[-1.2, 1.8]	[-1.1, 1.9]
c_{HW}	[-0.37, 0.4]	[-0.34, 0.41]	[-0.9, 0.3]	[-0.97, 0.31]
c_{Hu}	[-0.27, 0.3]	[-0.22, 0.38]	[-0.22, 0.13]	[-0.24, 0.15]
$c_{Hq}^{(3)}$	[-0.043, 0.05]	[-0.037, 0.053]	[-0.22, -0.18] $\cup$ [-0.059, 0.042]	[-0.23, -0.18] $\cup$ [-0.047, 0.044]
$c_{Hq}^{(1)}$	[-0.65, 0.56]	[-0.81, 0.47]	[-0.12, 0.15]	[-0.14, 0.16]
$c_{ll}^{(1)}$	[-2.2, 2.6]	[-2, 2.7]	[-2.5, 2.4]	[-2.2, 2.5]
$c_{Hl}^{(3)}$	[-2, 1.7]	[-2.1, 1.6]	[-1.8, 2]	[-1.9, 1.8]
$c_{Hl}^{(1)}$	[-12, 12]	[-15, 9.7]	[-6.7, 14]	[-7.6, 15]
c_{HD}	[-17, 16]	[-13, 17]	[-11, 9.2]	[-12, 10]
c_{Hd}	[-0.95, 0.84]	[-1.2, 0.69]	[-0.2, 0.25]	[-0.22, 0.27]
$c_{H\Box}$	[-3, 3.4]	[-2.7, 3.5]	[-3.3, 3.1]	[-3, 3.2]
c_{HB}	[-3.3, 2.6]	[-2.7, 2.9]	[-0.72, 0.9]	[-0.67, 0.85]
c_{He}	[-17, 16]	[-21, 13]	[-9.6, 21]	[-11, 23]
c_{dH}	[-2700, 5.4]	[$-\infty$, 5.2]	[-241, 7] $\cup$ [26, 266]	[$-\infty$, 6.6] $\cup$ [26, ∞]

Table 7.12: Expected (Exp.) and observed (Obs.) one-dimensional confidence intervals for the considered Wilson coefficients, at 68% CL (top) and at 95% CL (bottom). In each case, numbers are shown separately for linear and linear + quadratic parameterisations.



(a) (b)

Figure 7.16: The likelihood profile $t(c_i^{(6)})/2$ as a function of a single Wilson coefficient, for $c_{Hq}^{(3)}$ in (a), and for c_{Hu} in (b). All other Wilson coefficients are assumed to vanish. Curves for linear (blue) and linear + quadratic (orange) parameterisations are shown separately. The thin dashed lines indicate expected results (exp.); observed results (obs.) are shown in thick lines. Horizontal lines defining the confidence intervals at 68% and 95% CL are also visible.

Figure 7.17 shows two-dimensional confidence intervals for different operator pairings. These constraints show significant correlations, particularly when both linear and quadratic terms are included in the parameterisation. The sensitivity to a particular linear combination of Wilson coefficients can thus vary significantly. This has its origins in the interplay between the modifications caused by independent operators, leading to (partial) compensations and cancellations.

7.9.5.3 Eigenoperators

In principle, higher-dimensional (and increasingly model-independent) confidence intervals may be obtained by generalising the procedure outlined above. In practice, this leads to a further enhancement of the correlations between constraints. The limited amount of information contained in the five available STXS bins does not allow all linear combinations of Wilson coefficients to be constrained simultaneously, and additional observables must be included in the fit to lift the underlying degeneracies.

Here, a different approach is investigated, which determines those operator combinations to which this analysis shows good sensitivity. They are referred to as "eigenoperators", for reasons explained below. The eigenoperators parameterise the set of corrections to the SM that *can* be con-

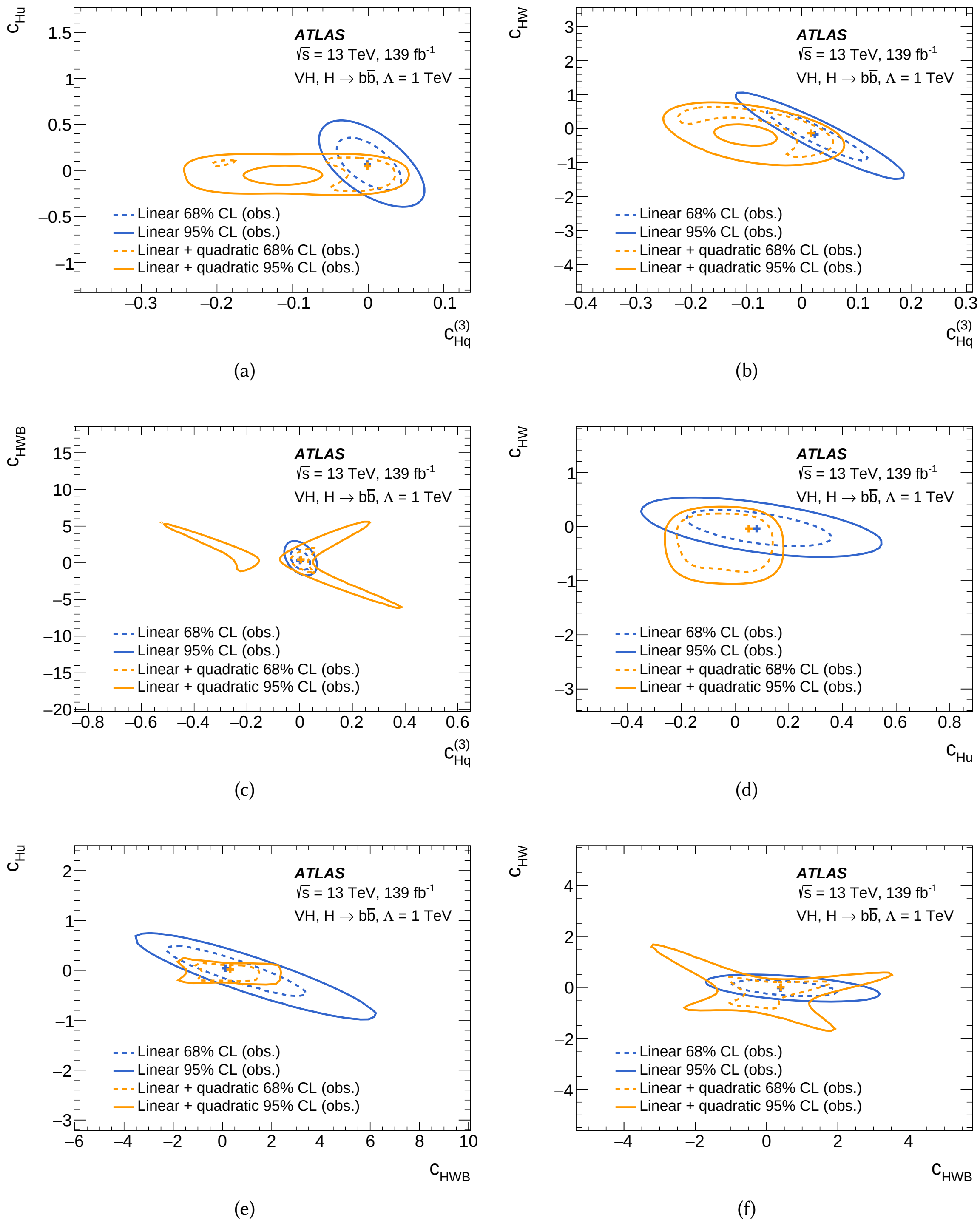


Figure 7.17: Observed (obs.) confidence intervals for pairs of Wilson coefficients at 68% and 95% CL. All Wilson coefficients except those explicitly considered in each case are assumed to vanish. Results are shown for linear (blue) and linear and quadratic (orange) parameterisations, and the best-fit value is indicated by a cross.

strained simultaneously by this measurement. This avoids introducing the unphysical assumption that only a limited number of operators contribute, and allows the available data to be interpreted in the most natural sense.

Determination of the eigenoperators The eigenoperators are defined as those linear combinations of Warsaw-basis operators that diagonalise the covariance matrix of a simultaneous fit [324].

The covariance matrix returned by a fit of the original likelihood model $L^{\mathsf{EFT}}(\boldsymbol{c}^{(6)}, \boldsymbol{\theta})$ (or, equivalently, its Hessian H^{EFT} evaluated at the best-fit values) is in general not diagonal,

$$H^{\mathsf{EFT}} = -\left.\frac{\partial^2 \log L^{\mathsf{EFT}}}{\partial c_i^{(6)} \partial c_j^{(6)}}\right|_{\hat{\boldsymbol{c}}^{(6)}} \neq \mathrm{diag}(\boldsymbol{\lambda}),$$

where $\boldsymbol{\lambda}$ is the vector of eigenvalues. For N_{op} parameterised Warsaw-basis operators and N_{STXS} measured STXS bins, H^{EFT} contains $N_{\mathrm{op}} - N_{\mathsf{STXS}}$ zero eigenvalues. Their eigenvectors label directions along which the likelihood does not vary, i.e. they identify linear combinations of operators that are degenerate given the available observables.

The eigenvectors associated with the nonzero eigenvalues represent a first-order estimate of the eigenoperators $\mathcal{Q}_{Ei}$. Due to interactions with nuisance parameters in the fit model, reformulating the likelihood model in terms of the (eigen-) Wilson coefficients c_{Ei} does not typically lead to a complete elimination of all mutual correlations, although these are usually reduced substantially. Iterating the diagonalisation procedure gives rise to the final eigenoperators so that $H^{\mathsf{EFT}} \to \mathrm{diag}(\boldsymbol{\lambda})$.

The operators $\mathcal{Q}_{Ei}$ are naturally ordered according to the corresponding eigenvalues λ_i. As diagonal elements of the Hessian, they represent a measure of the sensitivity of the analysis to the parameters c_{Ei},

$$\sigma^2(c_{Ei}) \approx \frac{1}{\lambda_i}. \tag{7.12}$$

The definition of the eigenoperators as linear combinations of Warsaw-basis operators is only strictly correct in the case of fully linear parameterisations. For more general dependencies, the eigendirections are given by curved trajectories in the space of operators (or Wilson coefficients).

Results The eigenoperators are determined for the manifestly linear parameterisation defined in Eq. 7.11. Table 7.13 shows the five eigenoperators $\mathcal{Q}_{E0}$–$\mathcal{Q}_{E4}$ that diagonalise the covariance matrix in a fit to data.

The leading eigenoperator $\mathcal{Q}_{E0}$ consists almost exclusively of $\mathcal{Q}_{Hq}^{(3)}$, which is also the operator

Eigenoperator	Eigenvalue	Composition
$\mathcal{Q}_{E0}$	2000	$0.98 \cdot \mathcal{Q}_{Hq}^{(3)}$
$\mathcal{Q}_{E1}$	38	$0.85 \cdot \mathcal{Q}_{Hu} - 0.39 \cdot \mathcal{Q}_{Hq}^{(1)} - 0.27 \cdot \mathcal{Q}_{Hd}$
$\mathcal{Q}_{E2}$	8.3	$0.70 \cdot \Delta B(H \to b\bar{b}) / \left[B(H \to b\bar{b})\right]^{\text{SM}} + 0.62 \cdot \mathcal{Q}_{HW}$
$\mathcal{Q}_{E3}$	0.2	$0.74 \cdot \mathcal{Q}_{HWB} + 0.53 \cdot \mathcal{Q}_{Hq}^{(1)} - 0.32 \cdot \mathcal{Q}_{HW}$
$\mathcal{Q}_{E4}$	$6.4 \cdot 10^{-3}$	$0.65 \cdot \mathcal{Q}_{HW} - 0.60 \cdot \Delta B(H \to b\bar{b}) / \left[B(H \to b\bar{b})\right]^{\text{SM}} + 0.35 \cdot \mathcal{Q}_{Hq}^{(1)}$

Table 7.13: Eigenoperators expressed as linear combinations of Warsaw basis operators, ordered according to their eigenvalues. Each eigenoperator is scaled to unit norm. For clarity, operators contributing with a coefficient less than 0.2 are omitted.

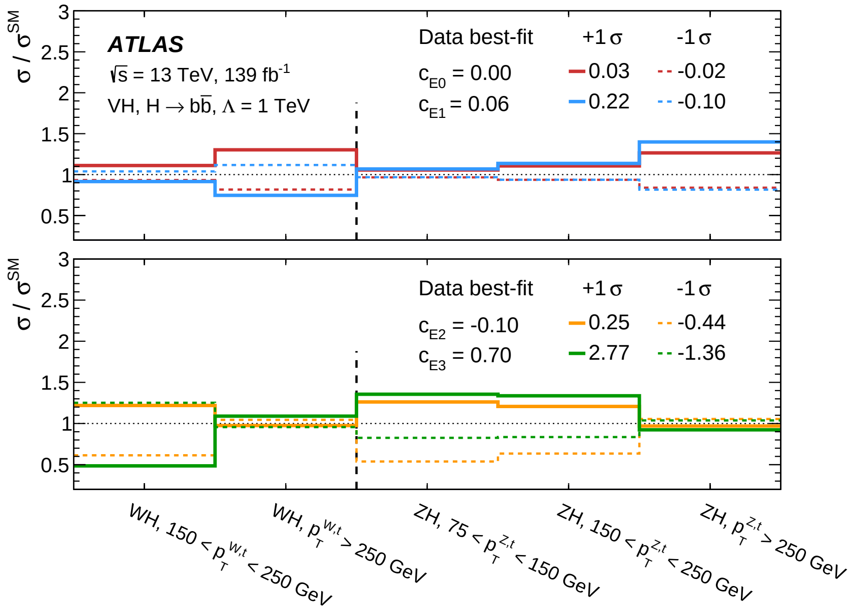


Figure 7.18: Modifications of the STXS bins caused by the four leading eigenoperators $\mathcal{Q}_{E0}$–$\mathcal{Q}_{E3}$, when the corresponding Wilson coefficients c_{E0}–c_{E3} are set to their $\pm 1\,\sigma$ values obtained from a simultaneous fit to data.

with the most significant one-dimensional constraint in Section 7.9.5.1. The second eigenoperator $\mathcal{Q}_{E1}$ consists of a linear combination of $\mathcal{Q}_{Hu}$, $\mathcal{Q}_{Hd}$, and $\mathcal{Q}_{Hq}^{(1)}$, exposing the degeneracy between them that was already pointed out in Section 7.9.2.1. The operators $\mathcal{Q}_{E2}$ and $\mathcal{Q}_{E3}$ show limited sensitivity to linear combinations of $\Delta B(H \to b\bar{b}) / \left[B(H \to b\bar{b})\right]^{\text{SM}}$, $\mathcal{Q}_{HW}$, and $\mathcal{Q}_{HWB}$. Although the measured STXS bins represent five degrees of freedom, only four eigenoperators show reasonable sensitivity, leaving the last operator, $\mathcal{Q}_{E4}$, virtually unconstrained.

The constraints on the four leading Wilson coefficients c_{E0}–c_{E3} resulting from a simultaneous fit are summarised in Figure 7.18, which also shows the modifications to the five measured STXS bins. The sensitivity closely matches the expectation from Eq. 7.12.

Chapter 8

Combination of measurements of VH production in the $H \to b\bar{b}$ channel

The data set recorded by ATLAS to date harbours great potential for the study of the VH process in extreme configurations. Although the cross-section for $pp \to V(\to \text{lep.})H(\to b\bar{b})$ in the region $p_T^{V,t} > 400\,\text{GeV}$ is only about 2 fb [269], several hundred events of this type are expected to have been produced during Run 2 of the LHC. Modifications effected by $d = 6$ interactions are enhanced in this domain, rendering those events highly informative about the structure of fundamental dynamics at large energy scales Λ. Such high-$p_T^{V,t}$, high-m_{VH} events are spectacular, with production restricted to well-instrumented central rapidities, heavily suppressed backgrounds, and striking trigger signatures. Nevertheless, it remains challenging to design analysis procedures that allow this potential to be realised and the available data set to be optimally utilised.

The p_T^V acceptance of the resolved measurement presented in Chapter 7 is intrinsically limited by the chosen strategy of reconstructing the Higgs-boson decay into two separate small-radius jets. Following Eq. 7.1, the two jet cones begin to touch for $p_T^H \approx 300\,\text{GeV}$. Their geometrical overlap becomes significant for $p_T^H > 600\,\text{GeV} \gg m_H$, considerably reducing the analysis acceptance and preventing a detailed study of the highly boosted regime. It becomes more efficient, then, to capture the collimated Higgs-boson decay products in a single large-radius jet with a distinct, two-pronged substructure [332].

A boosted VH measurement of this kind was presented in Ref. [6] for the first time. It is briefly summarised in Section 8.1. Section 8.2 recognises the complementarity between the resolved and boosted analysis techniques and devises a strategy for a rigorous combination of both approaches.

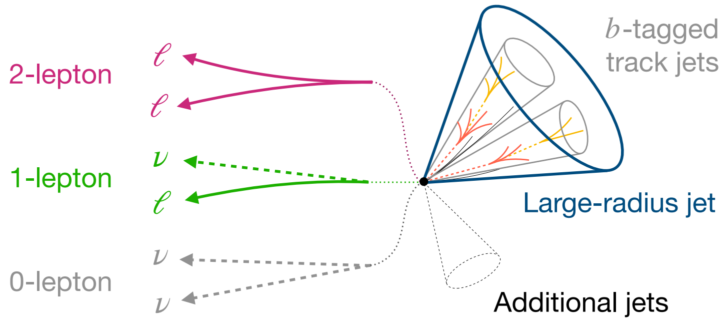


Figure 8.1: Detector signature of a highly boosted $V(\to \text{lep.})H(\to b\bar{b})$ event, where the Higgs boson decay products are contained in a single large-radius jet whose substructure is captured by two b-tagged track jets. Additional small-radius jets can also be present. The leptonic decay of the vector boson is reconstructed in the same manner as for resolved VH events.

The results of this procedure are summarised in Section 8.3. They extend the physics reach of either measurement and characterise the kinematic structure of the VH process with the highest granularity available to date.

8.1 Measurement of boosted VH production

The signature of the events selected by the boosted $V(\to \text{lep.})H(\to b\bar{b})$ analysis of Ref. [6] is illustrated in Figure 8.1. Like the resolved measurement presented in Chapter 7, the analysis is organised in three lepton channels to target WH and ZH production separately. The trigger strategies, selection criteria for charged leptons, and the definition of the kinematic scale p_T^V are identical to those listed in Section 7.3. Only events with $p_T^V > 250$ GeV are considered.

Large-radius jets are clustered with the anti-k_T algorithm using a radius parameter of $R = 1$, starting from topological calorimeter clusters. At least one large-radius jet must be present. It is required to be within the tracker acceptance, $|\eta| < 2.0$, and satisfy $p_T > 250$ GeV. The invariant mass of the large-radius jet, m_J, is computed from calorimeter energy deposits and tracks reconstructed in the inner detector [333].

The energy of a large-radius jet is susceptible to contamination from pile-up. To improve the jet mass resolution, the contents of the large-radius jet are re-clustered using the k_T algorithm [221] and soft sub-jets are subsequently removed. This grooming technique is referred to as “jet trimming” [334, 335].

The large-radius jet with the highest p_T forms the Higgs-boson candidate and must contain at least two sub-jets, which are clustered with a p_T-dependent radius parameter $(0.02 < R < 0.4)$ using tracks in the inner detector [336–338]. These jets are referred to as “variable-radius track jets”.

The two leading sub-jets are required to pass a b-tagging requirement using the MV2 algorithm and a working point similar to that defined in Section 7.3.3.

SRs are defined for two p_T^V regions, $p_T^V \in [250, 400)$ GeV and $p_T^V > 400$ GeV. In the 0- and 1-lepton channels, these are further split depending on the number of additional small-radius ($R = 0.4$) jets in the event. This defines a "high-purity" (HP) and a "low-purity" (LP) SR, containing events with no, or at least one, additional jet, respectively. Dedicated CRs for the $t\bar{t}$ background contain events with at least one additional b-tagged variable-radius track jet not associated with the Higgs-boson candidate. For such events, the Higgs-boson candidate typically captures a fully hadronic $t \to W(\to cq)b$ decay. In the low-yield 2-lepton channel, only one SR (for all jet multiplicities) was defined, and a dedicated $t\bar{t}$ CR was not required.

The signal contribution was extracted through a maximum-likelihood fit of histogram templates to the binned m_J distributions in all SRs and CRs. The small multi-jet background was extracted using a data-driven technique similar to that described in Section 7.7.1.3. All remaining backgrounds were modelled from simulation. This analysis provides unique sensitivity to the STXS Stage 1.2 sub-bin boundaries at $p_T^{V,t} = 400$ GeV (cf. Figure 7.4). Four STXS bins were measured, $p_T^{V,t} \in [250, 400)$ GeV and $p_T^{V,t} > 400$ GeV for both WH and ZH.

More details are available in the original publication.

8.2 Combination strategy

In the regime $p_T^H \in [300, 600)$ GeV, a significant fraction of all $H \to b\bar{b}$ decays may be reconstructed either in two resolved small-radius jets or as one large-radius jet. This leads to a substantial overlap in the events selected by the corresponding resolved and boosted analyses, illustrated in Figure 8.2.

8.2.1 Overlap removal

For the combined measurement presented here, the event selections are made orthogonal in the following way. Higgs-boson decays in events with $p_T^V < 400$ GeV were always reconstructed using the resolved analysis strategy (cf. Chapter 7), and the boosted strategy was used for events with $p_T^V \geq 400$ GeV. This allows the sensitivity provided by the multivariate discriminant BDT$_{VH}$ to be fully exploited in the region that dominates the inclusive VH cross-section, while simultaneously ensuring good signal acceptance at high scales. (The boosted-event selection achieves a signal acceptance that is about 30% higher than the acceptance of the resolved-event selection for events

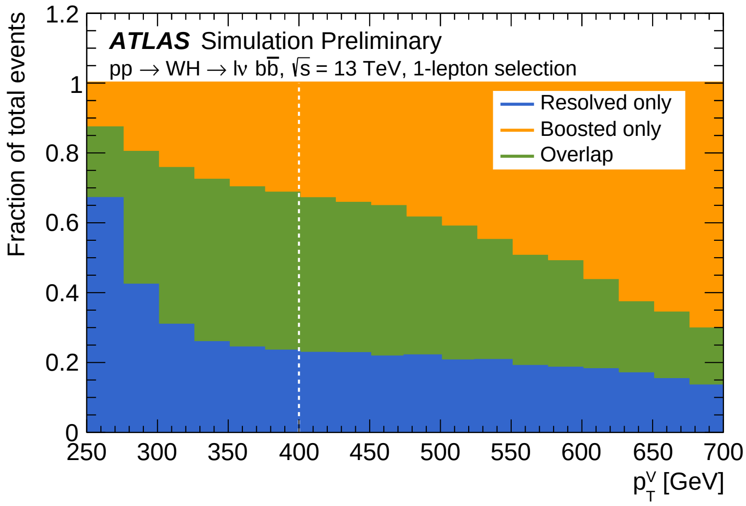


Figure 8.2: Fraction of $W(\to \ell\nu)H(\to b\bar{b})$ events selected in the 1-lepton channel of only the resolved analysis (blue), only the boosted analysis (orange), or both (green), shown as a function of p_T^V. Published in Ref. [9].

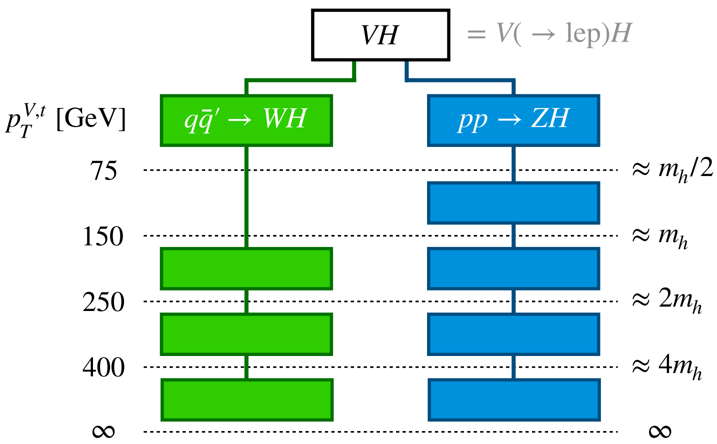
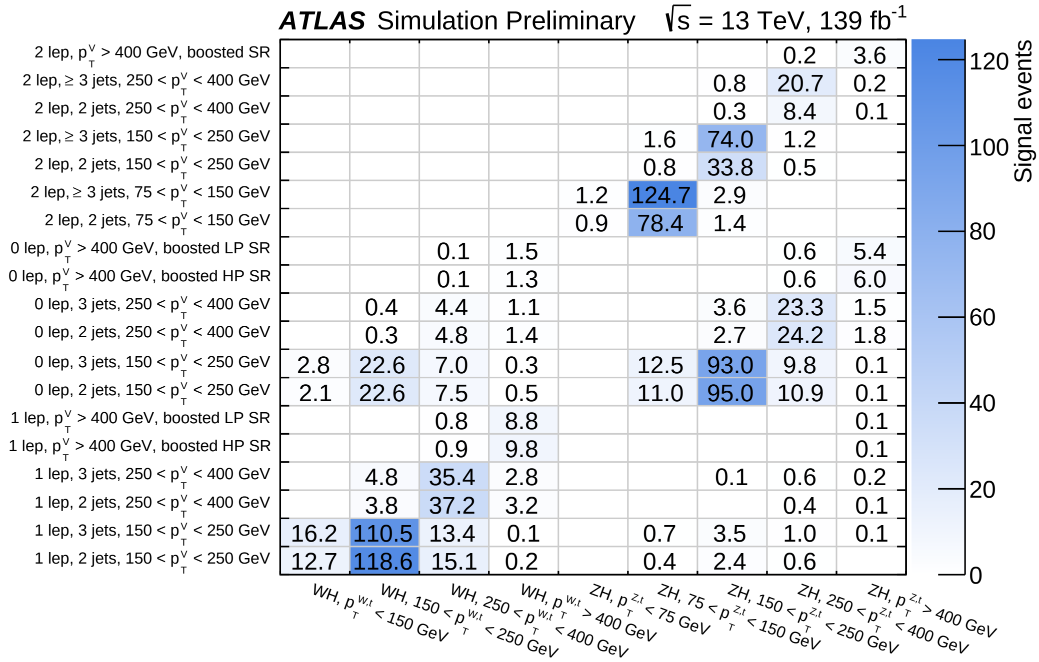


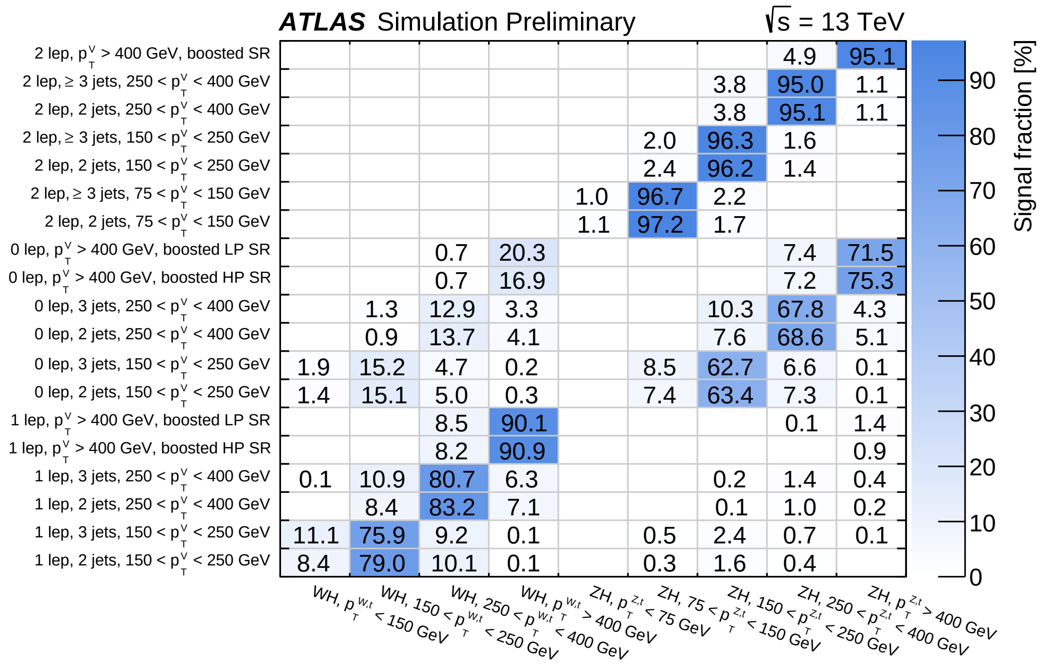
Figure 8.3: Reduced STXS binning targeted by the combined VH measurement, obtained by merging bins defined in STXS Stage 1.2.

with $p_T^V \geq 400$ GeV.) The transition at $p_T^V = 400$ GeV is well-aligned with the $p_T^{V,t} = 400$ GeV sub-bin boundary in the STXS Stage 1.2 scheme [222].

To implement this overlap-removal strategy, the event categorisation of the resolved measurement (cf. Table 7.3) is adjusted and the regions for $p_T^V > 250$ GeV were redefined to correspond to $p_T^V \in [250, 400)$ GeV instead. The analysis categories for this p_T^V bin used in the original boosted VH measurement were removed. This combination leads to a set of analysis categories well-suited to extracting the seven STXS bins illustrated in Figure 8.3. The correspondence between the analysis categories thus defined and these STXS bins is shown in Figure 8.4.



(a)



(b)

Figure 8.4: Correspondence between the analysis categories, shown on the vertical axis, and the STXS fiducial volumes, shown on the horizontal axis. (a) Predicted event yields for the VH signal process. Entries with < 0.1 events are suppressed for clarity. (b) The predicted fraction of signal events in a particular analysis category originating from a certain STXS bin. Entries with a signal fraction $< 0.1\%$ are suppressed.

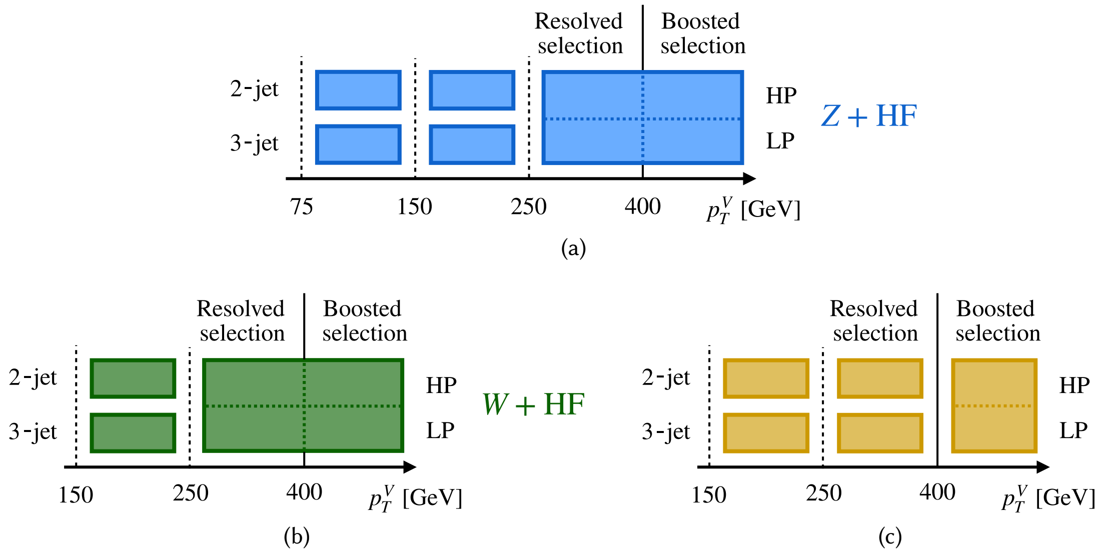


Figure 8.5: Strategy for the extraction of the normalisations of the main backgrounds from data, for Z+HF in (a), W+HF in (b), and $t\bar{t}$ in (c). The coloured boxes indicate the (combinations of) analysis regions for which a common floating normalisation is used. Coloured dashed lines indicate the presence of uncertainties that allow the relative background yields to be adjusted between regions controlled by the same floating normalisation.

8.2.2 Statistical model

The structure of the likelihood model and the inference strategy is identical to that described in Section 7.7.3. The statistical model now extends over the enlarged set of analysis categories defined above. Modifications are made to the treatment of the floating normalisations of the main backgrounds. These aspects are briefly summarised below, along with the form of the combined systematics model.

Background normalisation strategy The normalisations of the W+HF, Z+HF, and $t\bar{t}$ background components continue to be determined from the available CRs. The definitions of the corresponding normalisation factors were changed compared to Section 7.7.2 to account for the inclusion of the boosted analysis regions in the fit model and to ensure a reliable background modelling in all categories. The normalisation-factor scheme used in the combined analysis is summarised in Figure 8.5.

For W+HF and Z+HF, the background normalisation factor is shared between the highest resolved p_T^V bin and the boosted HP and LP SRs. Common normalisations were used for the 2-jet and ($\geq$)3-jet regions in this regime. Extrapolation uncertainties ranging from 10% to 30% were used to account for differences in the relative acceptance between the resolved and boosted regimes, and between the different jet multiplicities. They were derived from comparisons of the nominal simulation with alternative generators (cf. Section 7.7.2.2 and Ref. [6]). In the resolved domain for $p_T^V < 250$ GeV, separate normalisation factors were used for different analysis categories to reduce the reliance on simulation for the description of the p_T^V distribution.

The $t\bar{t}$ CRs in the boosted 0- and 1-lepton channels allow the normalisation of this background to be separately determined from data. Dedicated normalisation factors were used for the four p_T^V and jet-multiplicity combinations in the resolved domain. In the 2-lepton channel, the resolved top background contribution was determined from the $e\mu$-CR as described in Section 7.7.1.2. The boosted $t\bar{t}$ contribution was negligible in this channel.

Systematic uncertainties Experimental systematic uncertainties affecting trigger efficiencies, the luminosity determination, and the reconstruction and identification of leptons are equally present in both boosted and resolved categories. They were described by common ("correlated") NPs. The algorithms for the reconstruction of calorimeter jets and flavour tagging differ between the two regimes. The NPs parameterising these uncertainties were hence kept separate ("uncorrelated").

The parameterisation of background modelling uncertainties was primarily determined by the design and structure of the analysis and differed between the boosted and resolved regimes. The corresponding modelling NPs were thus kept largely uncorrelated, with few exceptions. Correlated NPs were used only for the description of flavour-composition uncertainties in the prediction of the V+HF backgrounds (which use equivalent parameterisations) and for uncertainties on the overall normalisation of the remaining subleading background components.

Theoretical uncertainties on the description of the VH signal were parameterised as described in Section 7.7.2.3 across all analysis categories and hence use correlated NPs.

8.3 Results

8.3.1 Signal-strength measurements

Figure 8.6 shows the modelling of the observed m_J distributions in the HP SRs and the $t\bar{t}$ CR in the boosted 1-lepton channel, obtained from the combined fit. The modelling of the data distributions in the resolved-analysis categories is very similar to that shown in Figure 7.9.

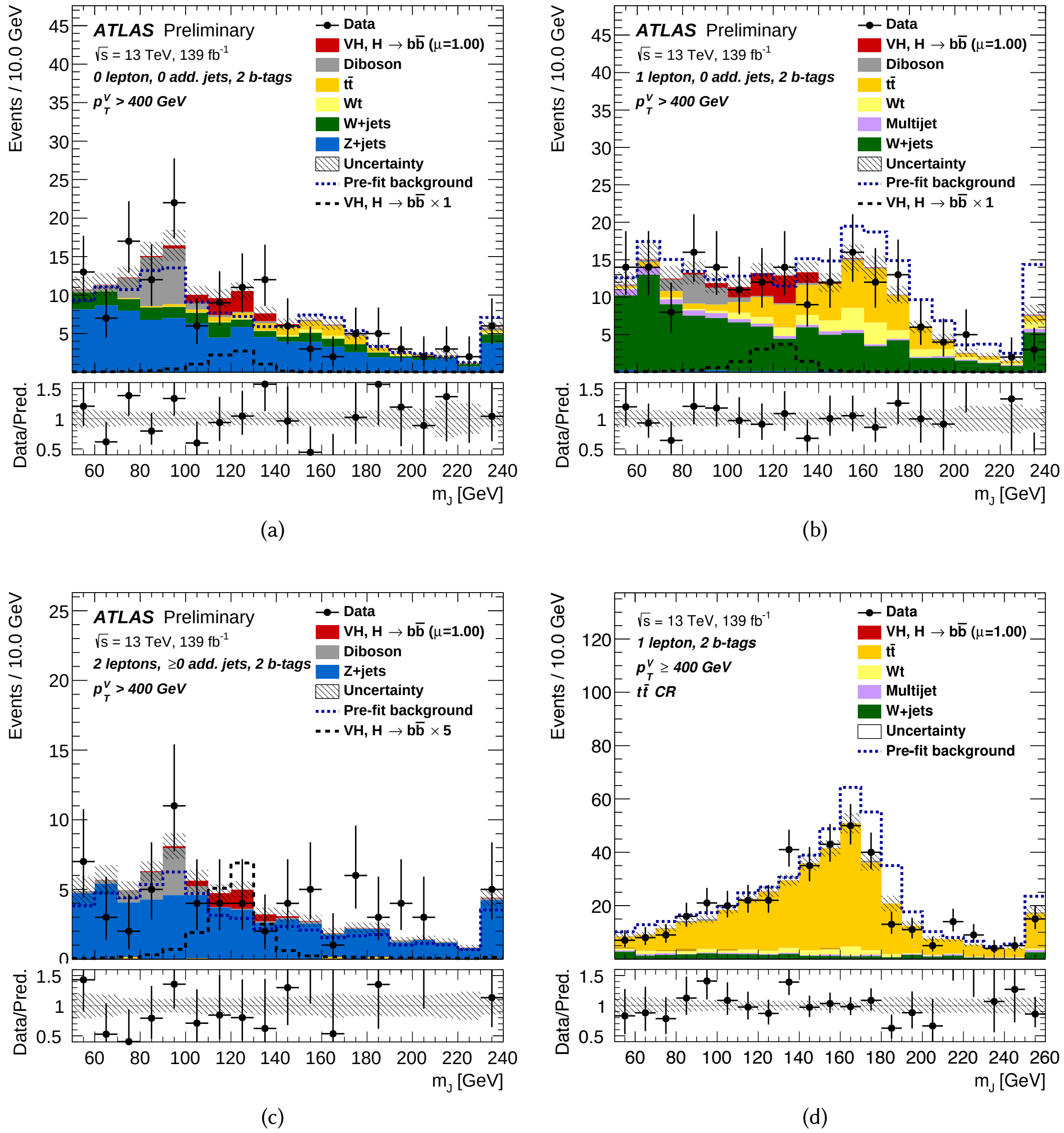


Figure 8.6: Binned distributions of m_J in the high-purity SRs for $p_T^V > 400$ GeV, in (a) for the 0-lepton channel, in (b) for 1-lepton, and in (c) for 2-lepton. The distribution in the $t\bar{t}$ CR in the boosted 1-lepton channel is shown in (d). The signal and background components are shown as stacked histograms, with the blue dashed line indicating the expected total background contribution. The signal contribution, scaled by the extracted signal strength, is additionally indicated by the black dashed line. The total background uncertainty is shown by the hatched band.

STXS bin		SM prediction			Exp. result			Stat. unc.		Syst. unc. [fb]				
Process	$p_T^{V,t}$ region	[fb]			[fb]			[fb]		Th. sig.		Th. bkg.		Exp.
WH	[150,250) GeV	24.0	$\pm$	1.1	19.6	$\pm$	12.4	$\pm$	8.1	$\pm$ 0.8	$\pm$	7.3	$\pm$	5.7
WH	[250,400) GeV	5.8	$\pm$	0.3	6.4	$\pm$	2.3	$\pm$	2.0	$\pm$ 0.3	$\pm$	0.9	$\pm$	0.6
WH	>400 GeV	1.3	$\pm$	0.1	1.9	$\pm$	1.1	$\pm$	0.9	$\pm$ 0.2	$\pm$	0.4	$\pm$	0.3
ZH	[75,150) GeV	50.6	$\pm$	4.1	49.5	$\pm$	36.9	$\pm$	25.9	$\pm$ 6.3	$\pm$	18.5	$\pm$	20.6
ZH	[150,250) GeV	18.8	$\pm$	2.4	20.0	$\pm$	6.4	$\pm$	5.1	$\pm$ 1.8	$\pm$	2.5	$\pm$	2.2
ZH	[250,400) GeV	4.1	$\pm$	0.5	4.0	$\pm$	1.6	$\pm$	1.5	$\pm$ 0.4	$\pm$	0.4	$\pm$	0.3
ZH	>400 GeV	0.7	$\pm$	0.1	0.2	$\pm$	0.6	$\pm$	0.5	$\pm$ 0.1	$\pm$	0.3	$\pm$	0.2

Table 8.1: Numerical values of the measured cross-sections $\sigma^{b}_{V(\text{lep.})H} \times B(H \to b\bar{b})$ in the seven targeted STXS bins, compared to the SM prediction and its theoretical uncertainties. The total uncertainty on the experimental result is split into a statistical uncertainty (Stat. unc.), systematic components resulting from the modelling of the signal (Th. sig.) and the backgrounds (Th. bkg.), as well as experimental sources (Exp.). Adapted from Ref. [9].

This results in a measured inclusive VH signal strength of

$$\mu^{bb}_{VH} = 1.00^{+0.18}_{-0.17} = 1.00^{+0.12}_{-0.11}\,(\text{stat.})\,^{+0.14}_{-0.13}\,(\text{syst.}),$$

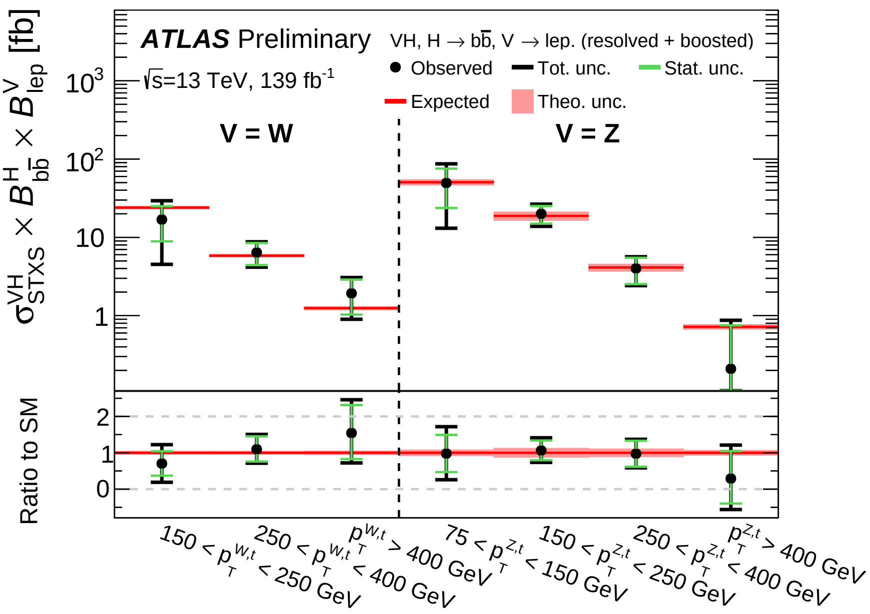
and in WH and ZH signal strengths of

$$\mu^{bb}_{WH} = 1.03^{+0.28}_{-0.27} = 1.03^{+0.19}_{-0.19}\,(\text{stat.})\,^{+0.21}_{-0.19}\,(\text{syst.}),$$
$$\mu^{bb}_{ZH} = 0.97^{+0.25}_{-0.23} = 0.97^{+0.17}_{-0.17}\,(\text{stat.})\,^{+0.18}_{-0.15}\,(\text{syst.}),$$

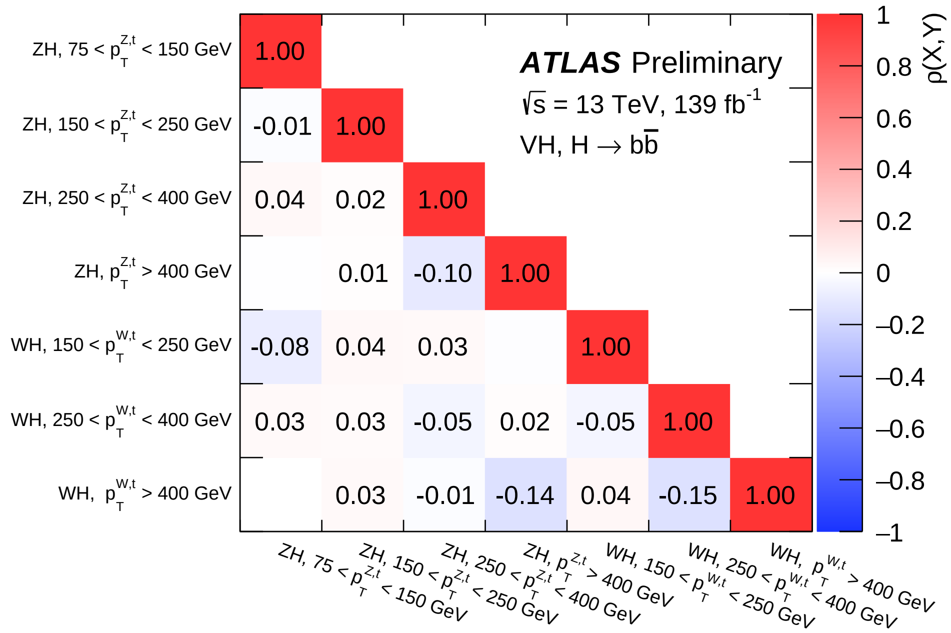
with a Pearson correlation of 2%. These results correspond to a rejection of the background-only hypotheses with observed (expected) significances of 6.4 (6.3) standard deviations for VH, and 4.1 (3.9) and 4.6 (5.0) standard deviations for WH and ZH, respectively. The sensitivity to the inclusive VH signal is slightly reduced compared to the results shown in Section 7.8.2.

8.3.2 STXS measurements

The cross-sections measured in the seven targeted STXS bins are shown in Figure 8.7a, and Table 8.1 summarises the numerical values. All values are in good agreement with the SM predictions, with relative uncertainties ranging between 35%–300%, depending on the p_T^V region. The Pearson correlations between these observables are shown in Figure 8.7b. The highest correlations of around 10%–15% are attained between the WH and ZH cross-sections for $p_T^{V,t} > 400$ GeV and between this bin and the $p_T^{V,t} \in [250, 400)$ GeV regions.

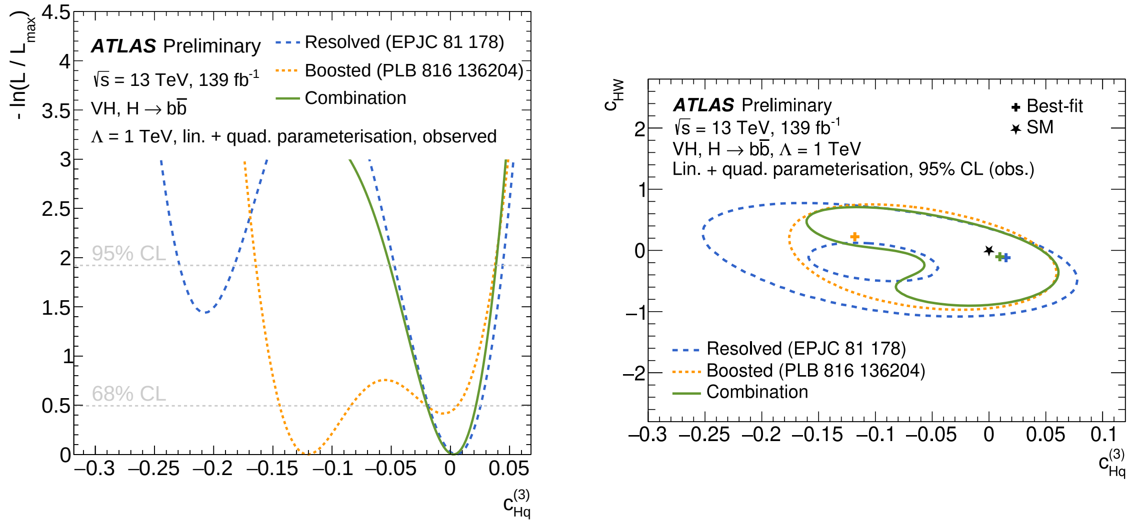


(a)



(b)

Figure 8.7: (a) Measured fiducial cross-sections $\sigma^{b}_{V(\text{lep.})H} \times B(H \to b\bar{b})$ in the seven considered STXS bins. The SM cross-sections and uncertainties in the theoretical predictions are indicated by the red bars. (b) Pearson correlation coefficients between the extracted STXS values.



(a) (b)

Figure 8.8: Constraints on the Wilson coefficients of $d = 6$ operators in the Warsaw basis, obtained from the STXS measurement of Figure 8.7, shown in green. The dashed lines in blue and orange show the comparison with the constraints set by the individual resolved and boosted measurements documented in Section 7.9 and Ref. [6], respectively. (a) Likelihood profile $t(c_i^{(6)})/2$ and one-dimensional confidence interval for $c_{Hq}^{(3)}$. (b) Two-dimensional confidence interval for $c_{Hq}^{(3)}$ and c_{HW}. Published in Ref. [9].

8.3.3 Constraints on effective interactions

Using the methodology developed in Section 7.9, the extended STXS measurement was used to place constraints on the Wilson coefficients of selected operators with $d = 6$ in the Warsaw basis.

Figure 8.8a illustrates the likelihood profile as a function of $c_{Hq}^{(3)}$ and compares it to the curves obtained from the individual boosted and resolved measurements. As discussed below Figure 7.12, the contact operator $\mathcal{Q}_{Hq}^{(3)}$ induces modifications that depend strongly on the kinematic scale p_T^V. The improved granularity and kinematic reach of the combined analysis allow this Wilson coefficient to be constrained more strongly. The approximate degeneracy for $c_{Hq}^{(3)} < 0$ is now resolved.

Figure 8.8b shows a two-dimensional confidence interval for $c_{Hq}^{(3)}$ and c_{HW}. A significant increase in the strength of the constraint is again visible for $c_{Hq}^{(3)}$. The operator $\mathcal{Q}_{HW}$ leads to an approximately linear enhancement of the cross-section with p_T^V (cf. Figure 7.12). In this case, the increased granularity of the STXS measurement is not as important and the limit on c_{HW} improves less. Constraints on additional (combinations of) operators are available in the original publication in Ref. [9].

These results are also utilised in a combined analysis of ATLAS Higgs-boson measurements which includes in addition the $H \to \gamma\gamma$, $H \to ZZ^*$, $H \to WW^*$, and $H \to \tau\tau$ decay channels [339]. The sensitivity to $c_{Hq}^{(3)}$ is almost completely determined by the combined $VH, H \to b\bar{b}$ measurement and the improvements described above directly enhance the global constraint.

Chapter 9

Conclusions and outlook

Experimental particle physics truly is a science of many scales, a science where the transverse momentum of a Higgs boson measured in units of TeV can give rise to detector signals with an energy of only a few meV. Physical processes across this extreme range, and the interplay between them, must be understood and controlled to ensure the continued development of the global experimental programme.

The analytical methods developed in Part I help tackle some of this complexity, condensing the essence of intricate dynamical processes into a few explicit formulae. They clearly expose fundamental limitations and commonalities, show scaling laws, identify the most important physical parameters, and thus help to guide and direct further work based on more detailed simulations.

A fully general but yet efficient method for the computation of detector signals is a powerful tool. It unifies many previous results and provides a coherent and principled modelling strategy for all devices that detect fields or radiation from moving charged particles. Applied to antenna arrays used in the study of cosmic rays, the theorem opens up exciting new possibilities for building high-fidelity signal simulations that go beyond contemporary techniques and have the potential to enhance the physics reach of future experiments.

The analytic description of electron-hole avalanche statistics sheds light on the magnitude of fluctuations and the origins of subtle correlation patterns. This theory provides the groundwork for the rigorous description of single-photon avalanche diodes and explains the superior timing capabilities of these devices from first principles. It also identifies their potential for precision timing measurements of charged particles in collider experiments. Although first experimental studies in a hadron beam have demonstrated good performance, additional studies are needed before large-area, radiation-hard detectors can become a reality.

In a reductionistic universe, most problems (and their answers) are destined to lie at the bottom, at the smallest of length scales. Part II builds on the fantastic capabilities of the ATLAS detector and the Large Hadron Collider to obtain the most granular cross-section measurements of the $VH(\to b\bar{b})$ process available to date. This analysis combines resolved and boosted event reconstruction techniques to cover the entire accessible kinematic domain, measuring fiducial production cross-sections in the STXS scheme with uncertainties from 30%–300%, depending on the category. These measurements lead to important constraints on contact interactions generated by processes beyond the direct reach of present colliders. Effects from operators to which the analysis is most sensitive are shown to be absent up to energy scales of 1–4 TeV, depending on the assumptions made.

To further tighten our grip on the TeV scale and beyond, we must continue to improve and expand our experimental capabilities. The future of the experimental programme at the Large Hadron Collider is a bright and luminous one, promising to deliver profound surprises and enable stringent tests of the global structure and internal coherence of our model of the world. But yet, if the physics we understand is to help us master phenomena we don't yet comprehend, we must also not shy away from exploring new, complementary ideas. In the long term, we must learn to harvest the full power contained in the laws of physics, all the way up to the Standard Model in its present form, exploit this potential for the construction of novel instruments, and thus begin to shrink the gap between our experiments and their target.

Acknowledgements

First and foremost, I would like to thank my advisor, Daniela Bortoletto, for giving me as much intellectual freedom as I could wish for, and for letting me go wherever my curiosity would lead me. My time in Oxford has been a truly amazing period, and one that I will not easily forget!

Special thanks also go to Werner Riegler, who taught me everything I know about particle detectors. Our forays into electrodynamics kept me sane during the hectic “unblinding” days, and discussing exponentially growing charge avalanches with you was a most welcome distraction during the early days of exponentially growing case numbers.

Many thanks also to Magnus Mager for adopting me into the ITS3 lab team, for rekindling my fascination for electronics, and for periodically reminding me that even an RF circuit board will eventually need an enclosure. And to Chris Pollard for long conversations about category theory, geodesics, optimal transport, and other crazy ideas; and for putting some of these ideas into practice together. It’s been a lot of fun!

Over the course of the last three years, I have had the privilege to meet, work with, and learn from a huge number of talented people, both at Oxford and elsewhere. Thanks to Thomas Calvet, Tatsuya Masubuchi, and Elisabeth Schopf for your leadership in bringing the VH analysis to its conclusion; and in particular to Thomas for his companionship during 3 a.m. fit adventures and for introducing me to the Higgs Cross-Section Working Group community. And to the entire $H \to b\bar{b}$ community in ATLAS, you have taught me what it means to never stop asking questions. Thanks to Bingxuan Liu, Carlo Schiavi, and Valerio Dao for willingly sharing your expertise on so many things, flavour tagging being only one of them. Thanks to Francesca Carnesecchi and Alperen Yüncü for the fun beam test times, and to the entire paradise club for making me feel at home immediately. Special thanks go to Sue Geddes, who makes the impossible possible on a daily basis.

I can not even begin to express my gratitude to a few very special people who have shaped me immensely, and who made sure that I am a more complete person now than I have been three years ago. It would not suffice to just thank you here, in these very last lines the unrelenting page limit permits me to write. You know who you are, and I owe you everything.

Appendices

Appendix A

The statistics of electron-hole avalanches in semiconductors: auxiliary material

A.1 Commutation relations

A.1.1 Ladder operators

The ladder operators $\hat{a}^\dagger$ and $\hat{a}$ exhibit the usual commutation relations

$$\left[\hat{a}_e, \hat{a}_e^\dagger\right] = \left[\hat{a}_h, \hat{a}_h^\dagger\right] = 1,$$

while all commutators combining operators from different species, e.g. $[\hat{a}_e, \hat{a}_h^\dagger]$, vanish. The commutation relations of the position-dependent ladder operators $\hat{a}^\dagger(x)$ and $\hat{a}(x)$ are direct generalisations of this expression. Operators at different locations always commute,

$$\left[\hat{a}_e(x), \hat{a}_e^\dagger(y)\right] = \left[\hat{a}_h(x), \hat{a}_h^\dagger(y)\right] = \delta(x-y),$$

and commutators mixing operators from different species continue to vanish.

A.1.2 Number operators

The nonvanishing commutation relations between the number operators and the ladder operators are

$$\begin{aligned} \left[\hat{N}_e, \hat{a}_e^\dagger\right] &= \hat{a}_e^\dagger, & \left[\hat{N}_e, \hat{a}_e\right] &= -\hat{a}_e, \\ \left[\hat{N}_h, \hat{a}_h^\dagger\right] &= \hat{a}_h^\dagger, & \left[\hat{N}_h, \hat{a}_h\right] &= -\hat{a}_h, \end{aligned}$$

and

$$\left[\hat{n}_e(x), \hat{a}_e^\dagger(y)\right] = \hat{a}_e^\dagger(x)\delta(x-y), \qquad [\hat{n}_e(x), \hat{a}_e(y)] = -\hat{a}_e(x)\delta(x-y),$$
$$\left[\hat{n}_h(x), \hat{a}_h^\dagger(y)\right] = \hat{a}_h^\dagger(x)\delta(x-y), \qquad [\hat{n}_h(x), \hat{a}_h(y)] = -\hat{a}_h(x)\delta(x-y).$$

A.1.3 Translation operators

With the definition in Eq. 2.43, the commutators of the translation operators with the number density operators are

$$\left[\hat{n}_e(y), \hat{T}_e[\Delta]\right] = \hat{T}_e[\Delta]\left(\frac{\hat{n}_e(x)}{1+\frac{d\Delta}{dx}} - \hat{n}_e(y)\right),$$
$$\left[\hat{n}_h(y), \hat{T}_h[\Delta]\right] = \hat{T}_h[\Delta]\left(\frac{\hat{n}_h(x)}{1+\frac{d\Delta}{dx}} - \hat{n}_h(y)\right).$$

A position-dependent drift velocity $v(x)$ acting for a short time interval dt induces the infinitesimal translation $\Delta(x) = v(x)dt$. For electrons and holes, the above commutation relations specialise to

$$\left[\hat{n}_e(x), \hat{T}_e[v_e(x)dt]\right] = -dt\,\hat{T}_e[v_e(x)dt]\,\frac{d}{dx}v_e(x)\hat{n}_e(x),$$
$$\left[\hat{n}_h(x), \hat{T}_h[-v_h(x)dt]\right] = dt\,\hat{T}_h[-v_h(x)dt]\,\frac{d}{dx}v_h(x)\hat{n}_h(x).$$

In these expressions, the action of a derivative on a generic operator $\hat{\mathcal{O}}(x)$ is to be understood as

$$dx\frac{d}{dx}\hat{\mathcal{O}}(x) := \hat{\mathcal{O}}(x+dx) - \hat{\mathcal{O}}(x),$$

which is inspired by the fact that expectation values and spatial derivatives commute for sufficiently smooth distributions, $\left\langle \frac{d}{dx}\hat{\mathcal{O}}(x)\right\rangle = \frac{d}{dx}\left\langle \hat{\mathcal{O}}(x)\right\rangle$.

A.2 Logarithmic fluctuations at late times

With Eq. 2.12, the evolution equation for the observable $\log \hat{N}_e$ reads

$$\frac{d}{dt}\langle \log N_e\rangle = \langle \text{avg}|\,[\log \hat{N}_e, \hat{H}]\,|\psi\rangle \ni \alpha v_e \left\langle \log\left(1+\frac{1}{N_e}\right)N_e\right\rangle. \tag{A.1}$$

The right-hand side contains terms that are not determined by the equation itself, and an ad-hoc closure scheme would be required to determine its solution. Expanding Eq. A.1 for large N_e shows that this issue disappears for large avalanches, where the right-hand side becomes constant, $\lim_{N_e\to\infty}\left\langle \log\left(1+\frac{1}{N_e}\right)N_e\right\rangle = 1$. A similar phenomenon occurs in the evolution equation for the second moment $\langle \log^2 N \rangle$.

This suggests that there exist observables which are not generally identical to $\log \hat{N}$ and $\log^2 \hat{N}$ but which limit to these quantities for large avalanches and whose evolution equations can be integrated for all t. Two such observables, $\hat{f}_1$ and $\hat{f}_2$, are defined as

$$\hat{f}_1 := \frac{1}{\alpha v_e + \beta v_h} \psi_0 \left(\frac{\alpha v_e \hat{N}_e + \beta v_h \hat{N}_h}{\alpha v_e + \beta v_h} \right),$$

and

$$\hat{f}_2 := \frac{1}{2(\alpha v_e + \beta v_h)^2} \left[\psi_0 \left(\frac{\alpha v_e \hat{N}_e + \beta v_h \hat{N}_h}{\alpha v_e + \beta v_h} \right)^2 + \psi_1 \left(\frac{\alpha v_e \hat{N}_e + \beta v_h \hat{N}_h}{\alpha v_e + \beta v_h} \right) \right],$$

with the polygamma functions ψ_k. Their limiting behaviour is correct,

$$\lim_{N_e, N_h \to \infty} \hat{f}_1 = \frac{1}{\alpha v_e + \beta v_h} \left[\log \hat{N} + C_0 \right],$$
$$\lim_{N_e, N_h \to \infty} \hat{f}_2 = \frac{1}{2(\alpha v_e + \beta v_h)^2} \left[\left(\log \hat{N} + C_0 \right)^2 \right],$$

where C_0 is an irrelevant constant. With these definitions, the asymptotic value of $\sigma(\log N)$ may be expressed in terms of the expectation values $\langle f_1 \rangle$ and $\langle f_2 \rangle$ as

$$\lim_{t\to\infty} 2 \langle f_2 \rangle - \langle f_1 \rangle^2 = \lim_{t\to\infty} \frac{1}{(\alpha v_e + \beta v_h)^2} \left[\langle \log^2 N \rangle - \langle \log N \rangle^2 \right] = \lim_{t\to\infty} \frac{\sigma^2(\log N)}{(\alpha v_e + \beta v_h)^2}. \quad \text{(A.2)}$$

The time evolution equations of these expectation values involve the commutators $[\hat{f}_1, \hat{H}]$ and $[\hat{f}_2, \hat{H}]$, where $\hat{H}$ is the Hamiltonian from Eq. 2.16. These commutators are trivial,

$$\left[\hat{f}_1, \hat{H} \right] = \hat{a}_e^\dagger \hat{a}_h^\dagger, \qquad \left[\hat{f}_2, \hat{H} \right] = \hat{a}_e^\dagger \hat{a}_h^\dagger \hat{f}_1,$$

as are their time evolution equations,

$$\frac{d}{dt} \langle f_1 \rangle = 1, \qquad \frac{d}{dt} \langle f_2 \rangle = \langle f_1 \rangle .$$

This shows that the combination $2 \langle f_2 \rangle - \langle f_1 \rangle^2$ that appears in Eq. A.2 is time-independent, and thus

$$\lim_{t\to\infty} \sigma^2(\log N) = (\alpha v_e + \beta v_h)^2 \left(2 \langle f_2 \rangle - \langle f_1 \rangle^2 \right) \Big|_{t=0} = \psi_1(A),$$

as claimed in Eq. 2.42.

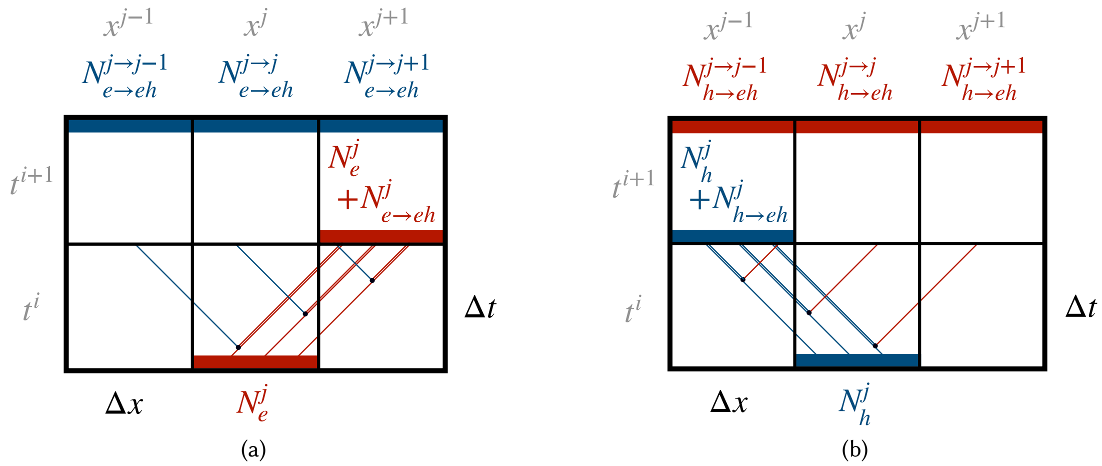


Figure A.1: Illustration of the update step $t^i \to t^{i+1}$ for the evolution of electrons in (a) and for the evolution of holes in (b). The numbers of electrons N_e^j and holes N_h^j contained in a bin j are represented by coloured bars within the respective bin. A number of exemplary drift lines are shown to illustrate the movement of charge carriers. Impact ionisation events are indicated by small black dots.

A.3 Markov avalanche Monte Carlo simulation model

To simulate the development of the avalanche for equal carrier drift velocities, $v_e = v_h = v^*$, the avalanche region $x \in [0, d]$ is partitioned into N equidistant bins of width $\Delta x = d/N$. The units are chosen such that charges drift a distance of Δx during the simulation time step Δt, $\Delta t = \Delta x / v^*$. The bin centre of bin j is located at coordinate x^j and the simulation time steps are indexed as $t^i = i\Delta t$.

The electric field distribution $\mathbf{E}(x,t) = -E^x(x,t)\hat{\mathbf{x}}$ and the resulting impact ionisation coefficients $\alpha(x,t)$ and $\beta(x,t)$ are discretised into $E^j(t^i) = E^x(x^j, t^i)$, $\alpha^j(t^i) = \alpha(x^j, t^i)$, and $\beta^j(t^i) = \beta(x^j, t^i)$. The charge carrier densities $n_e(x,t)$ and $n_h(x,t)$ are represented by the numbers of electrons $N_e^j(t^i)$ and holes $N_h^j(t^i)$ in each bin. These carriers are assumed to be uniformly distributed within the bin.

The discretised electric field profile $E^j(t^i)$ and the initial carrier densities $N_e^j(t^0)$ and $N_h^j(t^0)$ form the inputs to the simulation.

In the absence of impact ionisation, the N_e^j holes present in bin j at time t^i drift into bin $j+1$ at t^{i+1}, and the N_h^j holes move into bin $j-1$. Impact ionisation is simulated as illustrated in Figure A.1. During the time interval Δt, the N_e^j electrons (N_h^j holes) in bin j produce $N_{e\to eh}^j$ ($N_{h\to eh}^j$) additional electron-hole pairs. These numbers are drawn from a Poisson distribution with rate parameters

$\alpha^j \Delta x$ and $\beta^j \Delta x$, respectively,

$$N^j_{e\to eh}(t^i) \sim \text{Po}\left(N^j_e(t^i); \alpha^j(t^i)\Delta x\right), \qquad N^j_{h\to eh}(t^i) \sim \text{Po}\left(N^j_h(t^i); \beta^j(t^i)\Delta x\right).$$

Impact ionisation occurs uniformly along Δx. The $N^j_{e\to eh}$ electrons created in bin j at t^i drift into bin $j+1$ at t^{i+1}, while the $N^j_{e\to eh}$ holes spread across multiple bins: $N^{j\to j-1}_{e\to eh}$ holes reach bin $j-1$, $N^{j\to j}_{e\to eh}$ remain in bin j, and $N^{j\to j+1}_{e\to eh}$ enter bin $j+1$. The same consideration holds for the $N^j_{h\to eh}$ electrons created from hole-initiated impact ionisation events, leading to $N^{j\to j-1}_{h\to eh}$, $N^{j\to j}_{h\to eh}$, and $N^{j\to j+1}_{h\to eh}$ electrons in bins $j-1$, j, and $j+1$, respectively. These numbers are drawn from a multinomial distribution with event probabilities of $p_{j\to j-1} = 0.25$, $p_{j\to j} = 0.5$, and $p_{j\to j+1} = 0.25$, respectively,

$$\begin{aligned}
\{N^{j\to j-1}_{e\to eh}(t^i), N^{j\to j}_{e\to eh}(t^i), N^{j\to j+1}_{e\to eh}(t^i)\} &\sim \text{Mult}\left(N^j_{e\to eh}(t^i), \{p_{j\to j-1}, p_{j\to j}, p_{j\to j+1}\}\right)\\
\{N^{j\to j-1}_{h\to eh}(t^i), N^{j\to j}_{h\to eh}(t^i), N^{j\to j+1}_{h\to eh}(t^i)\} &\sim \text{Mult}\left(N^j_{h\to eh}(t^i), \{p_{j\to j-1}, p_{j\to j}, p_{j\to j+1}\}\right).
\end{aligned}$$

The updated charge carrier distributions are then computed as

$$\begin{aligned}
N^j_e(t^{i+1}) &= N^{j-1}_e(t^i) + N^{j-1}_{e\to eh}(t^i) + N^{j+1\to j}_{h\to eh}(t^i) + N^{j\to j}_{h\to eh}(t^i) + N^{j-1\to j}_{h\to eh}(t^i),\\
N^j_h(t^{i+1}) &= N^{j+1}_h(t^i) + N^{j+1}_{h\to eh}(t^i) + N^{j+1\to j}_{e\to eh}(t^i) + N^{j\to j}_{e\to eh}(t^i) + N^{j-1\to j}_{e\to eh}(t^i).
\end{aligned}$$

The current $I^{\text{ind}}(t^i)$ induced by the drifting charges is computed according to Eq. 2.33 and assuming a uniform weighting field $E_w = V_w/d$,

$$I^{\text{ind}}(t^i) = \frac{e_0 v^*}{d} \sum_j \left(N^j_e(t^i) + N^j_h(t^i)\right).$$

In the continuum limit where $\alpha\Delta x \to 0$ and $\beta\Delta x \to 0$, this simulation model converges to the avalanche model defined in Section 2.1. For the results presented in the main body, the avalanche region is divided into $N = 10^3$ bins.

Bibliography

[1] W. Riegler and P. Windischhofer, "Signals induced on electrodes by moving charges, a general theorem for Maxwell's equations based on Lorentz reciprocity", *Nucl. Instrum. Meth. A* 980, 164471 (2020), DOI: 10.1016/j.nima.2020.164471.

[2] P. Windischhofer and W. Riegler, "Electrical signals induced in detectors by cosmic rays: a reciprocal look at electrodynamics", *PoS* ICRC2021, 184 (2021), DOI: 10.22323/1.395.0184.

[3] P. Windischhofer and W. Riegler, "The statistics of electron-hole avalanches", *Nucl. Instrum. Meth. A* 1003, 165327 (2021), DOI: 10.1016/j.nima.2021.165327.

[4] W. Riegler and P. Windischhofer, "Time resolution and efficiency of SPADs and SiPMs for photons and charged particles", *Nucl. Instrum. Meth. A* 1003, 165265 (2021), DOI: 10.1016/j.nima.2021.165265.

[5] ATLAS Collaboration, "Measurements of WH and ZH production in the $H \to b\bar{b}$ decay channel in pp collisions at 13 TeV with the ATLAS detector", *Eur. Phys. J. C* 81, 178 (2021), DOI: 10.1140/epjc/s10052-020-08677-2.

[6] ATLAS Collaboration, "Measurement of the associated production of a Higgs boson decaying into b-quarks with a vector boson at high transverse momentum in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector", *Phys. Lett. B* 816, 136204 (2021), DOI: 10.1016/j.physletb.2021.136204.

[7] ATLAS Collaboration, "Direct constraint on the Higgs-charm coupling from a search for Higgs boson decays into charm quarks with the ATLAS detector", arXiv: 2201.11428 [hep-ex] (2022).

[8] ATLAS Collaboration, "Extrapolation of ATLAS sensitivity to $H \to b\bar{b}$ and $H \to c\bar{c}$ decays in VH production at the HL-LHC", ATL-PHYS-PUB-2021-039 (2021), URL: http://cdsweb.cern.ch/record/2788490.

[9] ATLAS Collaboration, "Combination of measurements of Higgs boson production in association with a W or Z boson in the $b\bar{b}$ decay channel with the ATLAS experiment at $\sqrt{s} = 13$ TeV", ATLAS-CONF-2021-051 (2021), URL: http://cdsweb.cern.ch/record/2782535.

[10] ATLAS Collaboration, "Search for new resonances in mass distributions of jet pairs using 139 fb^{-1} of pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector", *JHEP* 2020, 145 (2020), DOI: 10.1007/JHEP03(2020)145.

[11] ATLAS Collaboration, "Simulation-based extrapolation of b-tagging calibrations towards high transverse momenta in the ATLAS experiment", ATL-PHYS-PUB-2021-003 (2021), URL: http://cdsweb.cern.ch/record/2753444.

[12] P. Windischhofer, M. Zgubič, and D. Bortoletto, "Preserving physically important variables in optimal event selections: a case study in Higgs physics", *JHEP* 2020, 1 (2020), DOI: 10.1007/JHEP07(2020)001.

[13] C. Pollard and P. Windischhofer, "Transport away your problems: Calibrating stochastic simulations with optimal transport", *Nucl. Instrum. Meth. A* 1027, 166119 (2022), DOI: 10.1016/j.nima.2021.166119.

[14] P. Padovani et al., "Active galactic nuclei: what's in a name?", *A&AR* 25, 2 (2017), DOI: 10.1007/s00159-017-0102-9.

[15] G. A. Askaryan, "Excess negative charge of an electron-photon shower and its coherent radio emission", *Sov. Phys. JETP* 41, 616–618 (1961).

[16] J. A. Aguilar et al., "The Next-Generation Radio Neutrino Observatory–Multi-Messenger Neutrino Astrophysics at Extreme Energies", arXiv: 1907.12526 [astro-ph.HE] (2019).

[17] J.-C. Denard, "Beam current monitors", CERN-2009-005.141 (2009), URL: https://cds.cern.ch/record/1213275.

[18] G. Lutz, "Semiconductor radiation detectors", 1st ed., Springer (1999).

[19] J. C. Maxwell, "A Treatise on Electricity and Magnetism", 2nd ed., vol. 1, Clarendon Press, Oxford (1881), pp. 371–373.

[20] J. R. Carson, "A generalization of the Reciprocal Theorem", *Bell Syst. Tech. J.* 3, 393–399 (1924), DOI: 10.1002/j.1538-7305.1924.tb00009.x.

[21] J. R. Carson, "Reciprocal Theorems in Radio Communication", *Proc. IRE* 17, 952–956 (1929), DOI: 10.1109/JRPROC.1929.221772.

[22] J. R. Carson, "The Reciprocal Energy Theorem", *Bell Syst. Tech. J.* 9, 325–331 (1930), DOI: 10.1002/j.1538-7305.1930.tb00369.x.

[23] S. Ballantine, "Reciprocity in Electromagnetic, Mechanical, Acoustical, and Interconnected Systems", *Proc. IRE* 17, 927–951 (1929), DOI: 10.1109/JRPROC.1929.221771.

[24] J. C. Maxwell, "On the Calculation of the Equilibrium and Stiffness of Frames", *Phil. Mag.* 27, 294–299 (1864), DOI: 10.1080/14786446408643668.

[25] H. Lamb, "On Reciprocal Theorems in Dynamics", *Proc. London Math. Soc.* 19, 144–151 (1887), DOI: 10.1112/plms/s1-19.1.144.

[26] H. Helmholtz, "Theorie der Luftschwingungen in Röhren mit offenen Enden", *J. Reine Angew. Math.* 1860, 57, 1–72 (1860), DOI: 10.1515/crll.1860.57.1.

[27] H. Helmholtz, "Ueber die physikalische Bedeutung des Prinicips der kleinsten Wirkung", *J. Reine Angew. Math.* 1887, 100, 137 (1887), DOI: 10.1515/crll.1887.100.137.

[28] J. W. Strutt, "The Theory of Sound", Macmillan and Co., London (1877), pp. 97–117.

[29] J. W. Strutt, "Some General Theorems Relating to Vibrations", *Proc. London Math. Soc.* 4, 357–368 (1871), DOI: 10.1112/plms/s1-4.1.357.

[30] B. J. Hoenders, "On the invalidity of Helmholtz's reciprocity theorem for Green's functions describing the propagation of a scalar wave field in a non empty- and empty space", *Optik* 54, 373–380 (1979).

[31] H. A. Lorentz, "Het theorema van Poynting over de energie in het electromagnetisch veld en een paar algemeene stellingen over de voortplanting van het licht", *Vers. Konig. Akad. Wetensch.* 4, 176–187 (1895).

[32] M. Stumpf, "Electromagnetic Reciprocity in Antenna Theory", Wiley-IEEE Press (2017), pp. 11–13.

[33] G. Cavalleri et al., "Extension of Ramo's theorem as applied to induced charge in semiconductor detectors", *Nucl. Instrum. Meth. A* 92, 137–140 (1971), DOI: 10.1016/0029-554X(71)90235-7.

[34] L.-A. Hamel and M. Julien, “Comments on Ramo’s theorem”, *Proc. SPIE* 4507, 255–263 (2001), DOI: 10.1117/12.450766.

[35] L.-A. Hamel and M. Julien, “Generalized demonstration of Ramo’s theorem with space charge and polarization effects”, *Nucl. Instrum. Meth. A* 597, 207–211 (2008), DOI: 10.1016/j.nima.2008.09.008.

[36] W. Shockley, “Currents to Conductors Induced by a Moving Point Charge”, *J. Appl. Phys.* 9, 635–636 (1938), DOI: 10.1063/1.1710367.

[37] S. Ramo, “Currents induced by electron motion”, *Proc. IRE* 27, 9, 584–585 (1939), DOI: 10.1109/JRPROC.1939.228757.

[38] N. M. Ferrers, ed., “Mathematical Papers of George Green”, AMS Chelsea Publishing (1970).

[39] G. Green, “An Essay on the Application of mathematical Analysis to the theories of Electricity and Magnetism”, *J. Reine Angew. Math* 1852, 44, 356 (1852), DOI: 10.1515/crll.1852.44.356.

[40] W. Blum, W. Riegler, and L. Rolandi, “Particle detection with drift chambers”, 2nd ed., Springer, Berlin, Heidelberg (2008).

[41] E. Gatti, G. Padovini, and V. Radeka, “Signal evaluation in multielectrode radiation detectors by means of a time dependent weighting vector”, *Nucl. Instrum. Meth. A* 193, 651–653 (1982), DOI: 10.1016/0029-554X(82)90265-8.

[42] T. Quarles et al., “The Spice Page”, URL: http://bwrcs.eecs.berkeley.edu/Classes/IcBook/SPICE (visited on 04/11/2022).

[43] W. Riegler, “Induced signals in resistive plate chambers”, *Nucl. Instrum. Meth. A* 491, 258–271 (2002), DOI: 10.1016/S0168-9002(02)01169-5.

[44] W. Riegler, “Extended theorems for signal induction in particle detectors”, *Nucl. Instrum. Meth. A* 535, 287–293 (2004), DOI: 10.1016/j.nima.2004.07.129.

[45] W. Riegler, “An application of extensions of the Ramo–Shockley theorem to signals in silicon sensors”, *Nucl. Instrum. Meth. A* 940, 453–461 (2019), DOI: 10.1016/j.nima.2019.06.056.

[46] C. R. Paul, “Analysis of Multiconductor Transmission Lines”, 2nd ed., Wiley-IEEEPress (2007).

[47] R. Engel, D. Heck, and T. Pierog, “Extensive Air Showers and Hadronic Interactions at High Energy”, *Annu. Rev. Nucl. Part. Sci.* 61, 467–489 (2011), DOI: 10.1146/annurev.nucl.012809.104544.

[48] F. G. Schröder, “Radio detection of cosmic-ray air showers and high-energy neutrinos”, *Prog. Part. Nucl. Phys.* 93, 1–68 (2017), DOI: 10.1016/j.ppnp.2016.12.002.

[49] M. P. van Haarlem et al., “LOFAR: The LOw-Frequency ARray”, *A & A* 556, A2 (2013), DOI: 10.1051/0004-6361/201220873.

[50] J. D. Jackson, “Classical Electrodynamics”, 1st ed., John Wiley & Sons (1962).

[51] J. Alvarez-Muñiz, A. Romero-Wolf, and E. Zas, “Čerenkov radio pulses from electromagnetic showers in the time domain”, *Phys. Rev. D* 81, 123009 (2010), DOI: 10.1103/PhysRevD.81.123009.

[52] J. A. Aguilar et al., “Reconstructing the neutrino energy for in-ice radio detectors”, *Eur. Phys. J. C* 82, 147 (2022), DOI: 10.1140/epjc/s10052-022-10034-4.

[53] C. Glaser et al., “NuRadioMC: simulating the radio emission of neutrinos from interaction to detector”, *Eur. Phys. J. C* 80, 77 (2020), DOI: 10.1140/epjc/s10052-020-7612-8.

[54] C. Deaconu et al., "Measurements and modeling of near-surface radio propagation in glacial ice and implications for neutrino experiments", *Phys. Rev. D* 98, 043010 (2018), DOI: 10.1103/PhysRevD.98.043010.

[55] J. S. Townsend, "The theory of ionization of gases by collision", Constable & Company, London (1910).

[56] A. Rochas et al., "Single photon detector fabricated in a complementary metal–oxide–semiconductor high-voltage technology", *Rev. Sci. Instrum.* 74, 7, 3263–3270 (2003), DOI: 10.1063/1.1584083.

[57] R. J. McIntyre, "The distribution of gains in uniformly multiplying avalanche photodiodes: Theory", *IEEE Trans. on Electron Devices* 19, 6, 703–713 (1972), DOI: 10.1109/T-ED.1972.17485.

[58] M. Perenzoni, L. Pancheri, and D. Stoppa, "Compact SPAD-Based Pixel Architectures for Time-Resolved Image Sensors", *Sensors* 16, 5, 745 (2016), DOI: 10.3390/s16050745.

[59] O. Marinov, M. J. Deen, and Tejada J. A., "Theory of microplasma fluctuations and noise in silicon diode in avalanche breakdown", *J. Appl. Phys.* 101, 064515 (2007), DOI: 10.1063/1.2654973.

[60] Y. Okuto and Crowell C. R., "Ionization coefficients in semiconductors: A nonlocalized property", *Phys. Rev. B* 10, 4284–4296 (1974), DOI: 10.1103/PhysRevB.10.4284.

[61] R. Durret, "Probability: Theory and Examples", 5th ed., Cambridge University Press (2019).

[62] L. P. Kadanoff and J. Swift, "Transport Coefficients near the Critical Point: A Master-Equation Approach", *Phys. Rev.* 165, 310–322 (1978), DOI: 10.1103/PhysRev.165.310.

[63] P. Grassberger and M. Scheunert, "Fock-Space Methods for Identical Classical Objects", *Fortschr. Phys.* 28, 547–578 (1980), DOI: 10.1002/prop.19800281004.

[64] M. Doi, "Second quantization representation for classical many-particle system", *J. Phys. A: Math. Gen.* 9, 1465–1477 (1976), DOI: 10.1088/0305-4470/9/9/008.

[65] J. Baez and J. D. Biamonte, "Quantum Techniques in Stochastic Mechanics", World Scientific (2018), DOI: 10.1142/10623.

[66] A. V. Oppenheim and R. W. Schafer, "Discrete-Time Signal Processing", 3rd ed., Pearson (2010).

[67] W. H. Furry, "On Fluctuation Phenomena in the Passage of High Energy Electrons through Lead", *Phys. Rev.* 52, 569–581 (1937), DOI: 10.1103/PhysRev.52.569.

[68] G. U. Yule, "A Mathematical Theory of Evolution: Based on the Conclusions of Dr. J. C. Willis, F.R.S.", *Philos. Trans. Royal Soc. B* 213, 21–87 (1925), URL: http://www.jstor.org/stable/92117.

[69] W. Riegler, "Time response functions and avalanche fluctuations in resistive plate chambers", *Nucl. Instrum. Meth. A* 602, 377–390 (2009), DOI: 10.1016/j.nima.2008.12.207.

[70] R. J. McIntyre, "On the Avalanche Initiation Probability of Avalanche Diodes above the Breakdown Voltage", *IEEE Trans. Electron Devices* 20, 7, 637–641 (1973), DOI: 10.1109/T-ED.1973.17715.

[71] W. G. Oldham, R. R. Samuelson, and P. Antognetti, "Triggering phenomena in avalanche diodes", *IEEE Trans. on Electron Devices* 19, 9, 1056–1060 (1972), DOI: 10.1109/T-ED.1972.17544.

[72] L. H. Holway, "Electron-hole avalanches with constant ionization coefficients", *IEEE Trans. on Electron Devices* 26, 6, 991–993 (1979), DOI: 10.1109/T-ED.1979.19532.

[73] F. Sellberg, "Large-signal theory for rectangular-voltage operation of a uniform avalanche zone in IMPATT diodes", *Electron. Lett.* 7, 154–156 (1971), DOI: 10.1049/el:19710100.

[74] F. Sellberg, "Comments on 'Electron-hole avalanches with constant ionization coefficients'", *IEEE Trans. Electron Devices* 26, 1978 (1979), DOI: 10.1109/T-ED.1979.19805.

[75] R. Van Overstraeten and H. De Man, "Measurement of the ionization rates in diffused silicon p–n junctions", *Solid-State Electron.* 13, 583–608 (1970), DOI: 10.1016/0038-1101(70)90139-5.

[76] C. Canali et al., "Electron and hole drift velocity measurements in silicon and their empirical relation to electric field and temperature", *IEEE Trans. Electron Devices* 22, 1045–1047 (1975), DOI: 10.1109/T-ED.1975.18267.

[77] D. J. Massey, J. P. R. David, and G. J. Rees, "Temperature Dependence of Impact Ionization in Submicrometer Silicon Devices", *IEEE Trans. Electron Devices* 53, 2328–2334 (2006), DOI: 10.1109/TED.2006.881010.

[78] S. Cova et al., "Avalanche photodiodes and quenching circuits for single-photon detection", *Appl. Opt.* 35, 1956–1976 (1996), DOI: 10.1364/AO.35.001956.

[79] Y. Yu et al., "A Review of Quenching Circuit Design Based on Geiger-Mode APD", *2018 IEEE International Conference on Mechatronics and Automation (ICMA)*, 28–33 (2018), DOI: 10.1109/ICMA.2018.8484615.

[80] V. Savuskan et al., "Selecting Single Photon Avalanche Diode (SPAD) Passive-Quenching Resistance: An Approach", *IEEE Sens. J.* 13, 2322–2328 (2013), DOI: 10.1109/JSEN.2013.2253603.

[81] A. Inoue et al., "Modeling and Analysis of Capacitive Relaxation Quenching in a Single Photon Avalanche Diode (SPAD) Applied to a CMOS Image Sensor", *Sensors* 20, 3007 (2020), DOI: 10.3390/s20103007.

[82] K. E. Atkinson, "An Introduction to Numerical Analysis", 2nd ed., John Wiley & Sons (1989).

[83] D. A. Ramirez et al., "New perspective on passively quenched single photon avalanche diodes: effect of feedback on impact ionization", *Opt. Express* 20, 1512–1529 (2012), DOI: 10.1364/OE.20.001512.

[84] M. M. Hayat et al., "Model for passive quenching of SPADs", *Proc. SPIE* 7608, 76082B (2010), DOI: 10.1117/12.847005.

[85] W. C. Dunlap Jr. and R. L. Watters, "Direct Measurement of the Dielectric Constants of Silicon and Germanium", *Phys. Rev.* 92, 1396–1397 (1953), DOI: 10.1103/PhysRev.92.1396.

[86] C. Lippmann and W. Riegler, "Space charge effects in Resistive Plate Chambers", *Nucl. Instrum. Meth. A* 517, 54–76 (2004), DOI: 10.1016/j.nima.2003.08.174.

[87] D. J. Bartelink and G. Persky, "Diffusion of Electrons in Silicon Transverse to a High Electric Field", *Appl. Phys. Lett.* 16, 191 (1970), DOI: 10.1063/1.1653157.

[88] J. M. Hinckley and J. Singh, "Anisotropic high-field diffusion of holes in silicon", *Appl. Phys. Lett.* 66, 2727 (1995), DOI: 10.1063/1.113502.

[89] G. Persky and D. J. Bartelink, "High-Field Diffusivity of Electrons in Silicon", *J. Appl. Phys.* 42, 4414 (1971), DOI: 10.1063/1.1659788.

[90] T. Heubrandtner et al., "Static electric fields in an infinite plane condensor with one or three homogeneous layers", CERN-OPEN-2001-074 (2001), URL: https://cds.cern.ch/record/525925.

[91] T. Heubrandtner et al., "Static electric fields in an infinite plane condenser with one or three homogeneous layers", *Nucl. Instrum. Meth. A* 489, 439–443 (2002), DOI: 10.1016/S0168-9002(02)00805-7.

[92] R. J. McIntyre, "Theory of Microplasma Instability in Silicon", *J. Appl. Phys.* 32, 6, 983–995 (1961), DOI: 10.1063/1.1736199.

[93] R. H. Haitz, "Model for the Electrical Behavior of a Microplasma", *J. Appl. Phys.* 35, 5, 1370–1376 (1964), DOI: 10.1063/1.1713636.

[94] K. T. Son and C. C. Lee, "Multiple-Target Laser Range finding Receiver Using a Silicon Photomultiplier Array", *IEEE Trans. Instrum. Meas.* 59, 3005–3011 (2010), DOI: 10.1109/TIM.2010.2047127.

[95] C. Bruschini et al., "Single-photon avalanche diode imagers in biophotonics: review and outlook", *Light Sci. Appl.* 8, 87 (2019), DOI: 10.1038/s41377-019-0191-5.

[96] F. Ceccarelli et al., "Recent Advances and Future Perspectives of Single-Photon Avalanche Diodes for Quantum Photonics Applications", *Adv. Quantum Technol.* 4, 2000102 (2021), DOI: 10.1002/qute.202000102.

[97] F. Gramuglia et al., "A Low-noise CMOS SPAD Pixel with 12.1 ps SPTR and 3 ns Dead Time", *IEEE J. Sel. Top. Quantum Electron.* 28, 2, 1–9 (2021), DOI: 10.1109/JSTQE.2021.3088216.

[98] K. Morimoto et al., "Megapixel time-gated SPAD image sensor for 2D and 3D imaging applications", *Optica* 7, 4, 346–354 (2020), DOI: 10.1364/OPTICA.386574.

[99] C. Piemonte and A. Gola, "Overview on the main parameters and technology of modern Silicon Photomultipliers", *Nucl. Instrum. Meth. A* 926, 2–15 (2019), DOI: 10.1016/j.nima.2018.11.119.

[100] N. Atanov et al., "Design and test of the Mu2e undoped CsI + SiPM crystal calorimeter", *Nucl. Instrum. Meth. A* 936, 94–97 (2019), DOI: 10.1016/j.nima.2018.09.043.

[101] F. De Guio on behalf of the CMS Collaboration, "First results from the CMS SiPM-based hadronic endcap calorimeter", *J. Phys.: Conf. Ser.* 1162, 012009 (2019), DOI: 10.1088/1742-6596/1162/1/012009.

[102] F. Acerbi and S. Gundacker, "Understanding and simulating SiPMs", *Nucl. Instrum. Meth. A* 926, 16–35 (2019), DOI: 10.1016/j.nima.2018.11.118.

[103] R. Klanner, "Characterisation of SiPMs", *Nucl. Instrum. Meth. A* 926, 36–56 (2019), DOI: 10.1016/j.nima.2018.11.083.

[104] M. A. Green and M. J. Keevers, "Optical properties of intrinsic silicon at 300 K", *Prog. Photovolt.: Res. Appl.* 3, 3, 189–192 (1995), DOI: 10.1002/pip.4670030303.

[105] M. A. Green, "Self-consistent optical parameters of intrinsic silicon at 300 K including temperature coefficients", *Sol. Energy Mater. Sol. Cells* 92, 1305–1310 (2008), DOI: 10.1016/j.solmat.2008.06.009.

[106] F. Acerbi et al., "NUV Silicon Photomultipliers With High Detection Efficiency and Reduced Delayed Correlated-Noise", *IEEE Trans. Nucl. Sci.* 62, 1318–1325 (2015), DOI: 10.1109/TNS.2015.2424676.

[107] R. Brunetti et al., "Diffusion coefficient of electrons in silicon", *J. Appl. Phys.* 52, 11, 6713–6722 (1981), DOI: 10.1063/1.328622.

[108] H. Bichsel and H. Schindler, "The Interaction of Radiation with Matter", in: *Particle Physics Reference Library*, ed. by C. W. Fabjan and H. Schopper, vol. 2, Springer, Cham (2020), pp. 5–44, DOI: 10.1007/978-3-030-35318-6.

[109] U. Fano, "Ionization Yield of Radiations. II. The Fluctuations of the Number of Ions", *Phys. Rev.* 72, 26 (1947), DOI: 10.1103/PhysRev.72.26.

[110] R. H. Pehl et al., "Accurate determination of the ionization energy in semiconductor detectors", *Nucl. Instrum. Meth.* 59, 45–55 (1968), DOI: 10.1016/0029-554X(68)90342-X.

[111] W. W. M. Allison and J. H. Cobb, "Relativistic Charged Particle Identification by Energy Loss", *Ann. Rev. Nucl. Part. Sci.* 30, 253–298 (1980), DOI: 10.1146/annurev.ns.30.120180.001345.

[112] G. Hall, "Ionisation energy losses of highly relativistic charged particles in thin silicon layers", *Nucl. Instrum. Meth.* 220, 356–362 (1984), DOI: 10.1016/0167-5087(84)90296-5.

[113] J. Geist, "Silicon Revisited (1.1–3.1 eV)", in: *Handbook of Optical Constants of Solids*, ed. by E. D. Palik, vol. 3, Academic Press (1997), pp. 519–529, DOI: 10.1016/B978-012544415-6.50116-3.

[114] D. F. Edwards, "Silicon", in: *Handbook of Optical Constants of Solids*, ed. by E. D. Palik, vol. 1, Academic Press (1997), pp. 547–569, DOI: 10.1016/B978-012544415-6.50027-3.

[115] B. L. Henke, E. M. Gullikson, and J. C. Davis, "X-ray interactions: photoabsorption, scattering, transmission, and reflection at E=50–30000 eV, Z=1–92", *At. Data Nucl. Data Tables* 54, 181–342 (1993), DOI: 10.1006/adnd.1993.1013.

[116] I. B. Smirnov, "Modeling of ionization produced by fast charged particles in gases", *Nucl. Instrum. Meth. A* 554, 474–493 (2005), DOI: 10.1016/j.nima.2005.08.064.

[117] H. Schindler, "Microscopic Simulation of Particle Detectors", CERN-THESIS-2012-208 (2012), URL: https://cds.cern.ch/record/1500583.

[118] ALICE Collaboration, "Letter of intent for ALICE 3: A next generation heavy-ion experiment at the LHC", CERN-LHCC-2022-009 (2022), URL: https://cds.cern.ch/record/2803563.

[119] M. M. Vignetti et al., "Development of a 3D silicon coincidence avalanche detector for charged particle tracking in medical applications", *2016 IEEE Nuclear Science Symposium, Medical Imaging Conference and Room-Temperature Semiconductor Detector Workshop*, 1–7 (2016), DOI: 10.1109/NSSMIC.2016.8069553.

[120] N. D'Ascenzo et al., "Silicon avalanche pixel sensor for high precision tracking", *JINST* 9, C03027 (2014), DOI: 10.1088/1748-0221/9/03/c03027.

[121] F. Gramuglia et al., "Sub-10 ps Minimum Ionizing Particle Detection with Geiger-Mode APDs", arXiv: 2111.09998 [physics.ins-det] (2021).

[122] F. Carnesecchi et al., "Direct detection of charged particles with SiPMs", arXiv: 2202.04160 [physics.ins-det] (2022).

[123] V. Chmill et al., "Study of the breakdown voltage of SiPMs", *Nucl. Instrum. Meth. A* 845, 56–59 (2017), DOI: 10.1016/j.nima.2016.04.047.

[124] Infineon Technologies AG, "SiGe:C NPN RF bipolar transistor", BFP840FESD, v2.0, URL: https://www.infineon.com/dgdl/Infineon-BFP840FESD-DS-v02_00-EN.pdf?fileId=5546d46265f064ff01663896c6294eb9 (visited on 02/28/2022).

[125] Analog Devices Inc., "GaAs InGaP HBT MMIC Broadband Amplifier Gain Block, DC–6 GHz", HMC313, v07.0418, URL: https://www.analog.com/media/en/technical-documentation/data-sheets/hmc313.pdf (visited on 02/28/2022).

[126] D. Woods, "Reappraisal of the unconditional stability criteria for active 2-port networks in terms of S parameters", *IEEE Trans. Circuits Syst.* 23, 73–81 (1976), DOI: 10.1109/TCS.1976.1084179.

[127] M. Steer, "Microwave and RF Design: A Systems Approach", SciTech Publishing, Raleigh, NC (2010).

[128] B. C. Wadell, "Transmission Line Design Handbook", 6th ed., Artech House (1991).

[129] M. Mager on behalf of the ALICE Collaboration, "ALPIDE, the Monolithic Active Pixel Sensor for the ALICE ITS upgrade", *Nucl. Instrum. Meth. A* 824, 434–438 (2016), DOI: 10.1016/j.nima.2015.09.057.

[130] ALICE Collaboration, "Technical Design Report for the Upgrade of the ALICE Inner Tracking System", CERN-LHCC-2013-024 (2013), URL: http://cds.cern.ch/record/1625842.

[131] Zaber, "X-LSM-E Series Datasheet", X-LSM025A, URL: https://www.zaber.com/ufs/PdfsStore/6AwPTABX6yP5KSuYg/X-LSM-E-legacy-datasheet.pdf (visited on 02/28/2022).

[132] Rohde & Schwarz GmbH & Co. KG, "HMP Power Supply Family", HMP4040, v2.0, URL: https://scdn.rohde-schwarz.com/ur/pws/dl_downloads/dl_common_library/dl_brochures_and_datasheets/pdf_1/HMP_dat_en_5215-4981-32_v0200.pdf (visited on 02/28/2022).

[133] Hamamatsu, "Head-on PMT", H10492-003, URL: https://www.hamamatsu.com/content/dam/hamamatsu-photonics/sites/documents/99_SALES_LIBRARY/etd/m-h10492e.pdf (visited on 02/28/2022).

[134] W. Bothe and H. Geiger, "Über das Wesen des Comptoneffekts; ein experimenteller Beitrag zur Theorie der Strahlung", *Z. Physik* 32, 639–663 (1925), DOI: 10.1007/BF01331702.

[135] Lattice Semiconductor, "ice40 LP/HX Family Data Sheet", iCE40HX-1k, URL: https://www.latticesemi.com/~/media/LatticeSemi/Documents/DataSheets/iCE/iCE40LPHXFamilyDataSheet.pdf (visited on 02/28/2022).

[136] Teledyne LeCroy, "LabMaster 10 Zi-A High Bandwidth Modular Oscilloscopes", LabMaster 10 Zi-A 36 GHz, URL: http://cdn.teledynelecroy.com/files/pdf/labmaster-10zi-a-datasheet.pdf (visited on 02/28/2022).

[137] Y. Liu et al., "EUDAQ2—A flexible data acquisition software framework for common test beams", *JINST* 14, P10033 (2019), DOI: 10.1088/1748-0221/14/10/P10033.

[138] Raspberry Pi Ltd., "Raspberry Pi 4 Datasheet", Model B, URL: https://datasheets.raspberrypi.com/rpi4/raspberry-pi-4-datasheet.pdf (visited on 02/28/2022).

[139] M Pari et al., "Characterization of the slow extraction frequency response", *Phys. Rev. Accel. Beams* 24, 083501 (2021), DOI: 10.1103/PhysRevAccelBeams.24.083501.

[140] J. Kröger, S. Spannagel, and M. Williams, "User Manual for the Corryvreckan Test Beam Data Reconstruction Framework, Version 1.0", arXiv: 1912.00856 [physics.ins-det] (2019).

[141] J. Kröger and L. Huth, "Efficient Analysis of Test-beam Data with the Corryvreckan Framework", arXiv: 2011.09205 [physics.ins-det] (2020).

[142] V. Blobel, "Software alignment for tracking detectors", *Nucl. Instrum. Meth. A* 566, 5–13 (2006), DOI: 10.1016/j.nima.2006.05.157.

[143] T. W. Parks and C. S. Burrus, "Digital Filter Design", John Wiley & Sons, Inc. (1987).

[144] F. J. Harris, "On the use of windows for harmonic analysis with the discrete Fourier transform", *Proc. IEEE* 66, 51–83 (1978), DOI: 10.1109/PROC.1978.10837.

[145] M. Ghioni et al., "Progress in Silicon Single-Photon Avalanche Diodes", *IEEE J. Sel. Top. Quantum Electron.* 13, 852–862 (2007), DOI: 10.1109/JSTQE.2007.902088.

[146] Y. Xu, P. Xiang, and X. Xie, "Comprehensive understanding of dark count mechanisms of single-photon avalanche diodes fabricated in deep sub-micron CMOS technologies", *Solid State Electron.* 129, 168–174 (2016), DOI: 10.1016/j.sse.2016.11.009.

[147] E. Garutti and Y. Musienko, "Radiation damage of SiPMs", *Nucl. Instrum. Meth. A* 926, 69–84 (2019), DOI: 10.1016/j.nima.2018.10.191.

[148] S. O. Rice, "Mathematical analysis of random noise", *Bell Syst. Tech. J.* 23, 282–332 (1944), DOI: 10.1002/j.1538-7305.1944.tb00874.x.

[149] E. Wigner, "On Unitary Representations of the Inhomogeneous Lorentz Group", *Ann. Math.* 40, 149–204 (1939), DOI: 10.2307/1968551.

[150] S. Weinberg, "The Quantum Theory of Fields", vol. 1, Cambridge University Press (1995), DOI: 10.1017/CBO9781139644167.

[151] N. Arkani-Hamed, T.-C. Huang, and Y.-T. Huang, "Scattering amplitudes for all masses and spins", *JHEP* 2021, 70 (2021), DOI: 10.1007/JHEP11(2021)070.

[152] P. Benincasa and F. Cachazo, "Consistency Conditions on the S-Matrix of Massless Particles", arXiv: 0705.4305 [hep-th] (2007).

[153] P. Benincasa and E. Conde, "On the tree-level structure of scattering amplitudes of massless particles", *JHEP* 2011, 74 (2011), DOI: 10.1007/JHEP11(2011)074.

[154] R. Britto et al., "Direct Proof of the Tree-Level Scattering Amplitude Recursion Relation in Yang-Mills Theory", *Phys. Rev. Lett* 94, 181602 (2005), DOI: 10.1103/PhysRevLett.94.181602.

[155] S. Weinberg, "Feynman Rules for Any Spin II Massless Particles", *Phys. Rev.* 134, B882–B896 (1964), DOI: 10.1103/PhysRev.134.B882.

[156] S. Weinberg, "Photons and Gravitons in S-Matrix Theory: Derivation of Charge Conservation and Equality of Gravitational and Inertial Mass", *Phys. Rev.* 135, B1049–B1056 (1964), DOI: 10.1103/PhysRev.135.B1049.

[157] S. Weinberg, "Photons and Gravitons in Perturbation Theory: Derivation of Maxwell's and Einstein's Equations", *Phys. Rev.* 138, B988–B1002 (1965), DOI: 10.1103/PhysRev.138.B988.

[158] S. Weinberg, "The Quantum Theory of Fields", vol. 2, Cambridge University Press (1996), DOI: 10.1017/CBO9781139644174.

[159] I. Brivio and M. Trott, "The standard model as an effective field theory", *Phys. Rep.* 793, 1–98 (2019), DOI: 10.1016/j.physrep.2018.11.002.

[160] W. Buchmüller and D. Wyler, "Effective Lagrangian Analysis of New Interactions and Flavour Conservation", *Nucl. Phys. B.* 268, 621–653 (1986), DOI: 10.1016/0550-3213(86)90262-2.

[161] R. M. Fonseca, "Enumerating the operators of an effective field theory", *Phys. Rev. D* 101, 035040 (2020), DOI: 10.1103/PhysRevD.101.035040.

[162] B. Henning et al., "Hilbert series and operator bases with derivatives in effective field theories", *Commun. Math. Phys.* 347, 363–388 (2016), DOI: 10.1007/s00220-015-2518-2.

[163] B. Henning et al., "2, 84, 30, 993, 560, 15456, 11962, 261485, ...: higher dimension operators in the SM EFT", *JHEP* 2017, 16 (2017), DOI: 10.1007/JHEP08(2017)016.

[164] L. Lehman and A. Martin, "Hilbert Series for Constructing Lagrangians: expanding the phenomenologist's toolbox", *Phys. Rev. D* 91, 105014 (2015), DOI: 10.1103/PhysRevD.91.105014.

[165] L. Lehman and A. Martin, "Low-derivative operators of the Standard Model effective field theory via Hilbert series methods", *JHEP* 2016, 81 (2016), DOI: 10.1007/JHEP02(2016)081.

[166] A. Helset and A. Kobach, "Baryon number, lepton number, and operator dimension in the Standard Model", *Phys. Lett. B* 800, 135132 (2020), DOI: 10.1016/j.physletb.2019.135132.

[167] S. Weinberg, "Baryon- and Lepton-Nonconserving Processes", *Phys. Rev. Lett.* 43, 1566–1570 (1979), DOI: 10.1103/PhysRevLett.43.1566.

[168] S. F. King, "Neutrino mass models", *Rep. Prog. Phys.* 67, 107 (2003), DOI: 10.1088/0034-4885/67/2/R01.

[169] B. Grzadkowski et al., "Dimension-Six Terms in the Standard Model Lagrangian", *JHEP* 2010, 85 (2010), DOI: 10.1007/JHEP10(2010)085.

[170] E. E. Jenkins, A. V. Manohar, and M. Trott, "Renormalization group evolution of the standard model dimension six operators. I: formalism and λ dependence", *JHEP* 2013, 87 (2013), DOI: 10.1007/JHEP10(2013)087.

[171] E. E. Jenkins, A. V. Manohar, and M. Trott, "Renormalization group evolution of the Standard Model dimension six operators II: Yukawa dependence", *JHEP* 2014, 35 (2014), DOI: 10.1007/JHEP01(2014)035.

[172] R. Alonso et al., "Renormalization group evolution of the Standard Model dimension six operators III: gauge coupling dependence and phenomenology", *JHEP* 2014, 159 (2014), DOI: 10.1007/JHEP04(2014)159.

[173] E. E. Jenkins, A. V. Manohar, and M. Trott, "Naive dimensional analysis counting of gauge theory amplitudes and anomalous dimensions", *Phys. Lett. B* 2013, 697–702 (2013), DOI: 10.1016/j.physletb.2013.09.020.

[174] C. Grojean et al., "Renormalization Group Scaling of Higgs Operators and $h \to \gamma\gamma$ decay", *JHEP* 2013, 16 (2013), DOI: 10.1007/JHEP04(2013)016.

[175] D. A. Faroughy et al., "Flavour symmetries in the SMEFT", *JHEP* 2020, 166 (2020), DOI: 10.1007/JHEP08(2020)166.

[176] I. Brivio, Y. Jiang, and M. Trott, "The SMEFTsim package, theory and tools", *JHEP* 2017, 70 (2017), DOI: 10.1007/JHEP12(2017)070.

[177] I. Brivio, "SMEFTsim 3.0—a practical guide", *JHEP* 2021, 73 (2021), DOI: 10.1007/JHEP04(2021)073.

[178] W. Herr and B. Muratori, "Concept of luminosity", CERN-2006-002.361 (2006), URL: https://cds.cern.ch/record/941318.

[179] E. Eichten et al., "Supercollider physics", *Rev. Mod. Phys.* 56, 579–707 (1984), DOI: 10.1103/RevModPhys.56.579.

[180] J. Butterworth et al., "PDF4LHC recommendations for LHC Run II", *J. Phys. G: Nucl. Part. Phys.* 43, 023001 (2016), DOI: 10.1088/0954-3899/43/2/023001.

[181] L. Evans and P. Bryant, "LHC Machine", *JINST* 3, S08001 (2008), DOI: 10.1088/1748-0221/3/08/S08001.

[182] O. Brüning et al., eds., "LHC Design Report", CERN-2004-003-V-1 (2004), URL: https://cds.cern.ch/record/782076.

[183] E. D. Courant, M. S. Livingston, and H. S. Snyder, "The Strong-Focusing Synchrotron—A New High Energy Accelerator", *Phys. Rev.* 88, 1190–1196 (1952), DOI: 10.1103/PhysRev.88.1190.

[184] K. Eggert, K. Honkavaara, and A. Morsch, "Luminosity considerations for the LHC", CERN-AT-94-04-DI (1994), URL: https://cds.cern.ch/record/260711/.

[185] ATLAS Collaboration, "Luminosity Public Results: Run 2", URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2 (visited on 03/16/2022).

[186] ATLAS Collaboration, "The ATLAS Experiment at the CERN Large Hadron Collider", *JINST* 3, S08003 (2008), DOI: 10.1088/1748-0221/3/08/S08003.

[187] ATLAS Collaboration, "ATLAS inner detector: Technical Design Report, Volume 1", ATLAS-TDR-4 (1997), URL: https://cds.cern.ch/record/331063.

[188] Z. Drasal and W. Riegler, "An extension of the Gluckstern formulae for multiple scattering: Analytic expressions for track parameter resolution using optimum weights", *Nucl. Instrum. Meth. A* 910, 127–132 (2018), DOI: 10.1016/j.nima.2018.08.078.

[189] F. Ragusa and L. Rolandi, "Tracking at LHC", *New J. Phys.* 9, 336 (2007), DOI: 10.1088/1367-2630/9/9/336.

[190] Particle Data Group, "Review of Particle Physics", *PTEP* 2020, 083C01 (2020), DOI: 10.1093/ptep/ptaa104.

[191] ATLAS Collaboration, "ATLAS Insertable B-Layer Technical Design Report", ATLAS-TDR-19 (2010), URL: https://cds.cern.ch/record/1291633.

[192] M. Backhaus on behalf of the ATLAS Collaboration, "The upgraded Pixel Detector of the ATLAS Experiment for Run 2 at the Large Hadron Collider", *Nucl. Instrum. Meth. A* 831, 65–70 (2016), DOI: 10.1016/j.nima.2016.05.018.

[193] ATLAS Collaboration, "Particle Identification Performance of the ATLAS Transition Radiation Tracker", ATLAS-CONF-2011-128 (2011), URL: https://cds.cern.ch/record/1383793.

[194] ATLAS Collaboration, "The ATLAS Level-1 Calorimeter Trigger", *JINST* 3, P03001 (2008), DOI: 10.1088/1748-0221/3/03/P03001.

[195] C. W. Fabjan and F. Gianotti, "Calorimetry for particle physics", *Rev. Mod. Phys* 75, 1243 (2003), DOI: 10.1103/RevModPhys.75.1243.

[196] U. Amaldi, "Fluctuations in Calorimetry Measurements", *Phys. Scr.* 23, 409–424 (1981), DOI: 10.1088/0031-8949/23/4A/012.

[197] ATLAS Collaboration, "ATLAS liquid-argon calorimeter: Technical Design Report", ATLAS-TDR-2 (1996), URL: https://cds.cern.ch/record/331061.

[198] ATLAS Collaboration, "ATLAS tile calorimeter: Technical Design Report", ATLAS-TDR-3 (1996), URL: https://cds.cern.ch/record/331062.

[199] ATLAS Collaboration, "ATLAS muon spectrometer: Technical Design Report", ATLAS-TDR-10 (1997), URL: https://cds.cern.ch/record/331068.

[200] ATLAS Collaboration, "Performance of the ATLAS RPC detector and Level-1 muon barrel trigger at $\sqrt{s} = 13\,\text{TeV}$", *JINST* 16, P07029 (2021), DOI: 10.1088/1748-0221/16/07/P07029.

[201] ATLAS Collaboration, "Performance of the ATLAS Trigger System in 2015", *Eur. Phys. J. C* 77, 317 (2017), DOI: 10.1140/epjc/s10052-017-4852-3.

[202] ATLAS Collaboration, "Operation of the ATLAS trigger system in Run 2", *JINST* 15, P10004 (2020), DOI: 10.1088/1748-0221/15/10/P10004.

[203] 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, 1–29 (2012), DOI: 10.1016/j.physletb.2012.08.020.

[204] CMS Collaboration, “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”, *Phys. Lett. B* 716, 30–61 (2012), DOI: 10.1016/j.physletb.2012.08.021.

[205] M. Dine, “Naturalness Under Stress”, *Annu. Rev. Nucl. Part. Sci.* 65, 43–62 (2015), DOI: 10.1146/annurev-nucl-102014-022053.

[206] D. Buttazzo et al., “B-physics anomalies: a guide to combined explanations”, *JHEP* 2017, 44 (2017), DOI: 10.1007/JHEP11(2017)044.

[207] S. L. Glashow, D. V. Nanopoulos, and A. Yildiz, “Associated production of Higgs bosons and Z particles”, *Phys. Rev. D* 18, 1724–1727 (1978), DOI: 10.1103/PhysRevD.18.1724.

[208] A. Stange, W. Marciano, and S. Willenbrock, “Associated production of Higgs and weak bosons, with $H \to b\bar{b}$, at hadron colliders”, *Phys. Rev. D* 50, 4491–4498 (1994), DOI: 10.1103/PhysRevD.50.4491.

[209] C. Englert, M. McCullough, and M. Spannowsky, “Gluon-initiated associated production boosts Higgs physics”, *Phys. Rev. D* 89, 013013 (2014), DOI: 10.1103/PhysRevD.89.013013.

[210] K. Becker et al., “Precise predictions for boosted Higgs production”, arXiv: 2005.07762 [hep-ph] (2020).

[211] R. V. Harlander et al., “Exploiting the WH / ZH symmetry in the search for new physics”, *Eur. Phys. J. C* 78, 760 (2018), DOI: 10.1140/epjc/s10052-018-6234-x.

[212] R. Lafaye et al., “Measuring the Higgs sector”, *JHEP* 2009, 9 (2009), DOI: 10.1088/1126-6708/2009/08/009.

[213] LHC Higgs Cross Section Working Group, “Handbook of LHC Higgs Cross Sections: 3. Higgs Properties”, CERN-2013-004 (2013), URL: http://cds.cern.ch/record/1559921.

[214] ATLAS Collaboration, “Evidence for the $H \to b\bar{b}$ decay with the ATLAS detector”, *JHEP* 2017, 24 (2017), DOI: 10.1007/JHEP12(2017)024.

[215] ATLAS Collaboration, “Observation of $H \to b\bar{b}$ decays and VH production with the ATLAS detector”, *Phys. Lett. B* 786, 59–86 (2018), DOI: 10.1016/j.physletb.2018.09.013.

[216] ATLAS Collaboration, “Measurement of $VH, H \to b\bar{b}$ production as a function of the vector-boson transverse momentum in 13 TeV pp collisions with the ATLAS detector”, *JHEP* 2019, 141 (2019), DOI: 10.1007/JHEP05(2019)141.

[217] Gonçalves D. et al., “Higgs-Strahlung: Merging the NLO Drell-Yan and Loop-Induced 0+1 jet Multiplicities”, *Phys. Rev. D* 92, 073006 (2015), DOI: 10.1103/PhysRevD.92.073006.

[218] LHC Higgs Cross Section Working Group, “Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector”, CERN-2017-002-M (2016), URL: https://cds.cern.ch/record/2227475.

[219] S. Badger et al., “Les Houches 2015: Physics at TeV Colliders Standard Model Working Group Report”, arXiv: 1605.04692 [hep-ph] (2016).

[220] M. Boggia et al., “The HiggsTools Handbook: Concepts and observables for deciphering the Nature of the Higgs Sector”, arXiv: 1711.09875 [hep-ph] (2017).

[221] M. Cacciari, G. P. Salam, and G. Soyez, “The anti-k_t jet clustering algorithm”, *JHEP* 2008, 63 (2008), DOI: 10.1088/1126-6708/2008/04/063.

[222] N. Berger et al., “Simplified Template Cross Sections—Stage 1.1”, arXiv: 1906.02754 [hep-ph] (2019).

[223] ATLAS Collaboration, “Luminosity determination in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector at the LHC”, *Eur. Phys. J. C* 76, 653 (2016), DOI: 10.1140/epjc/s10052-016-4466-1.

[224] G. Avoni et al., “The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS”, *JINST* 13, P07017 (2018), DOI: 10.1088/1748-0221/13/07/P07017.

[225] D. J. Lange, “The EvtGen particle decay simulation package”, *Nucl. Instrum. Meth. A* 462, 152 (2001), DOI: 10.1016/S0168-9002(01)00089-4.

[226] T. Sjöstrand, S. Mrenna, and P. Skands, “A brief introduction to PYTHIA 8.1”, *Comput. Phys. Commun.* 178, 852–867 (2008), DOI: 10.1016/j.cpc.2008.01.036.

[227] ATLAS Collaboration, “The Pythia 8 A3 tune description of ATLAS minimum bias and inelastic measurements incorporating the Donnachie–Landshoff diffractive model”, ATL-PHYS-PUB-2016-017 (2016), URL: https://cds.cern.ch/record/2206965.

[228] R. Ball et al., “Parton distributions with LHC data”, *Nucl. Phys. B* 867, 244–289 (2013), DOI: 10.1016/j.nuclphysb.2012.10.003.

[229] ATLAS Collaboration, “The ATLAS Simulation Infrastructure”, *Eur. Phys. J. C* 70, 823 (2010), DOI: 10.1140/epjc/s10052-010-1429-9.

[230] S. Agostinelli et al., “GEANT4: a simulation toolkit”, *Nucl. Instrum. Meth. A* 506, 250–303 (2003), DOI: 10.1016/S0168-9002(03)01368-8.

[231] A. Bredenstein et al., “Precise predictions for the Higgs-boson decay $H \to WW/ZZ \to 4$ leptons”, *Phys. Rev. D* 74, 013004 (2006), DOI: 10.1103/PhysRevD.74.013004.

[232] A. Bredenstein et al., “Precise calculations for the Higgs decays $H \to WW/ZZ \to 4$ leptons”, *Nucl. Phys. Proc. Suppl.* 160, 131 (2006), DOI: 10.1016/j.nuclphysbps.2006.09.104.

[233] A. Bredenstein et al., “Radiative corrections to the semileptonic and hadronic Higgs-boson decays $H \to WW/ZZ \to 4$ fermions”, *JHEP* 2007, 80 (2007), DOI: 10.1088/1126-6708/2007/02/080.

[234] M. Spira, “QCD Effects in Higgs physics”, *Fortschr. Phys.* 46, 203–284 (1998), DOI: 10.1002/(SICI)1521-3978(199804)46:3\%3C203::AID-PROP203\%3E3.0.CO;2-4.

[235] A. Djouadi, J. Kalinowski, and M. Spira, “HDECAY: A program for Higgs boson decays in the Standard Model and its supersymmetric extension”, *Comput. Phys. Commun.* 108, 56 (1998), DOI: 10.1016/S0010-4655(97)00123-9.

[236] A. Djouadi, M. M. Mühlleitner, and M. Spira, “Decays of supersymmetric particles: The Program SUSY-HIT (SUspect-SdecaY-Hdecay-InTerface)”, *Acta Phys. Polon. B* 38, 635 (2007).

[237] S. Alioli et al., “A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX”, *JHEP* 2010, 43 (2010), DOI: 10.1007/JHEP06(2010)043.

[238] R. D. Ball et al., “Parton distributions for the LHC run II”, *JHEP* 2015, 40 (2014), DOI: 10.1007/JHEP04(2015)040.

[239] T. Sjöstrand et al., “An Introduction to PYTHIA 8.2”, *Comput. Phys. Commun.* 191, 159–177 (2015), DOI: 10.1016/j.cpc.2015.01.024.

[240] ATLAS Collaboration, “Measurement of the Z/γ^* boson transverse momentum distribution in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector”, *JHEP* 2014, 145 (2014), DOI: 10.1007/JHEP09(2014)145.

[241] G. Cullen et al., "Automated one-loop calculations with GOSAM", *Eur. Phys. J. C* 72, 1889 (2012), DOI: 10.1140/epjc/s10052-012-1889-1.

[242] K. Hamilton, P. Nason, and G. Zanderighi, "MINLO: multi-scale improved NLO", *JHEP* 2012, 155 (2012), DOI: 10.1007/JHEP10(2012)155.

[243] G. Luisoni et al., "$HW^{\pm}/HZ$ + 0 and 1 jet at NLO with the POWHEG BOX interfaced to GoSam and their merging within MiNLO", *JHEP* 2013, 83 (2013), DOI: 10.1007/JHEP10(2013)083.

[244] O. Brein et al., "Top-quark mediated effects in hadronic Higgs-Strahlung", *Eur. Phys. J. C* 72, 1868 (2012), DOI: 10.1140/epjc/s10052-012-1868-6.

[245] G. Ferrera, M. Grazzini, and F. Tramontano, "Higher-order QCD effects for associated WH production and decay at the LHC", *JHEP* 2014, 39 (2014), DOI: 10.1007/JHEP04(2014)039.

[246] G. Ferrera, M. Grazzini, and F. Tramontano, "Associated ZH production at hadron colliders: The fully differential NNLO QCD calculation", *Phys. Lett. B* 740, 51–55 (2015), DOI: 10.1016/j.physletb.2014.11.040.

[247] J. M. Campbell, R. K. Ellis, and C. Williams, "Associated production of a Higgs boson at NNLO", *JHEP* 2016, 179 (2016), DOI: 10.1007/JHEP06(2016)179.

[248] G. Ferrera, M. Grazzini, and F. Tramontano, "Associated Higgs-W-Boson Production at Hadron Colliders: A Fully Exclusive QCD Calculation at NNLO", *Phys. Rev. Lett.* 107, 152003 (2011), DOI: 10.1103/PhysRevLett.107.152003.

[249] M. L. Ciccolini, S. Dittmaier, and M. Krämer, "Electroweak radiative corrections to associated WH and ZH production at hadron colliders", *Phys. Rev. D* 68, 073003 (2003), DOI: 10.1103/PhysRevD.68.073003.

[250] O. Brein, A. Djouadi, and R Harlander, "NNLO QCD corrections to the Higgs-strahlung processes at hadron colliders", *Phys. Lett. B* 579, 149–156 (2004), DOI: 10.1016/j.physletb.2003.10.112.

[251] B. Hespel, F. Maltoni, and E. Vryonidou, "Higgs and Z boson associated production via gluon fusion in the SM and the 2HDM", *JHEP* 2015, 65 (2015), DOI: 10.1007/JHEP06(2015)065.

[252] R. V. Harlander et al., "Soft gluon resummation for gluon-induced Higgs Strahlung", *JHEP* 2014, 82 (2014), DOI: 10.1007/JHEP11(2014)082.

[253] R. V. Harlander, S. Liebler, and T. Zirke, "Higgs Strahlung at the Large Hadron Collider in the 2-Higgs-doublet model", *JHEP* 2014, 23 (2014), DOI: 10.1007/JHEP02(2014)023.

[254] L. Altenkamp et al., "Gluon-induced Higgs-strahlung at next-to-leading order QCD", *JHEP* 2013, 78 (2013), DOI: 10.1007/JHEP02(2013)078.

[255] O. Brein, R. V. Harlander, and T. Zirke, "vh@nnlo - Higgs Strahlung at hadron colliders", *Comput. Phys. Commun.* 184, 998–1003 (2013), DOI: 10.1016/j.cpc.2012.11.002.

[256] S. Frixione, P. Nason, and G. Ridolfi, "A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadron production", *JHEP* 2007, 126 (2007), DOI: 10.1088/1126-6708/2007/09/126.

[257] ATLAS Collaboration, "ATLAS Pythia 8 tunes to 7 TeV data", ATL-PHYS-PUB-2014-021 (2014), URL: https://cds.cern.ch/record/1966419.

[258] M. Czakon and A. Mitov, "Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders", *Comput. Phys. Commun.* 185, 2930 (2014), DOI: 10.1016/j.cpc.2014.06.021.

[259] S. Alioli et al., “NLO single-top production matched with shower in POWHEG: s- and t-channel contributions”, *JHEP* 2009, 111 (2009), DOI: 10.1088/1126-6708/2009/09/111.

[260] N. Kidonakis, “Two-loop soft anomalous dimensions for single top quark associated production with a W^- or H^-”, *Phys. Rev. D* 82, 054018 (2010), DOI: 10.1103/PhysRevD.82.054018.

[261] N. Kidonakis, “Next-to-next-to-leading-order collinear and soft gluon corrections for t-channel single top quark production”, *Phys. Rev. D* 83, 091503 (2011), DOI: 10.1103/PhysRevD.83.091503.

[262] E. Re, “Single-top Wt-channel production matched with parton showers using the POWHEG method”, *Eur. Phys. J. C* 71, 1547 (2011), DOI: 10.1140/epjc/s10052-011-1547-z.

[263] E. Bothmann et al., “Event Generation with Sherpa 2.2”, *SciPost Phys.* 7, 34 (2019), DOI: 10.21468/SciPostPhys.7.3.034.

[264] F. Cascioli, P. Maierhöfer, and S. Pozzorini, “Scattering Amplitudes with Open Loops”, *Phys. Rev. Lett.* 108, 111601 (2012), DOI: 10.1103/PhysRevLett.108.111601.

[265] T. Gleisberg and S. Höche, “Comix, a new matrix element generator”, *JHEP* 2008, 39 (2008), DOI: 10.1088/1126-6708/2008/12/039.

[266] S. Schumann and F. Krauss, “A parton shower algorithm based on Catani-Seymour dipole factorisation”, *JHEP* 2008, 38 (2008), DOI: 10.1088/1126-6708/2008/03/038.

[267] S. Höche et al., “QCD matrix elements + parton showers. The NLO case”, *JHEP* 2013, 27 (2013), DOI: 10.1007/JHEP04(2013)027.

[268] S. Catani et al., “Vector Boson Production at Hadron Colliders: A Fully Exclusive QCD Calculation at Next-to-Next-to-Leading Order”, *Phys. Rev. Lett.* 103, 082001 (2009), DOI: 10.1103/PhysRevLett.103.082001.

[269] ATLAS Collaboration, “Evaluation of theoretical uncertainties for simplified template cross section measurements of V-associated production of the Higgs boson”, ATL-PHYS-PUB-2018-035 (2018), URL: https://cdsweb.cern.ch/record/2649241/.

[270] A. Denner et al., “Electroweak corrections to Higgs-strahlung off W/Z bosons at the Tevatron and the LHC with Hawk”, *JHEP* 2012, 75 (2012), DOI: 10.1007/JHEP03(2012)075.

[271] A. Denner et al., “HAWK 2.0: A Monte Carlo program for Higgs production in vector-boson fusion and Higgs strahlung at hadron colliders”, *Comput. Phys. Commun.* 195, 161–171 (2015), DOI: 10.1016/j.cpc.2015.04.021.

[272] J. de Favereau de Jeneret et al., “High energy photon interactions at the LHC”, arXiv: 0908.2020 [hep-ph] (2009).

[273] ATLAS Collaboration, “Studies on top-quark Monte Carlo modelling for Top2016”, ATL-PHYS-PUB-2016-020 (2016), URL: https://cds.cern.ch/record/2216168.

[274] S. Frixione et al., “Single-top hadroproduction in association with a W boson”, *JHEP* 2008, 29 (2008), DOI: 10.1088/1126-6708/2008/07/029.

[275] S. Höche et al., “QCD matrix elements and truncated showers”, *JHEP* 2009, 53 (2009), DOI: 10.1088/1126-6708/2009/05/053.

[276] S. Höche et al., “A critical appraisal of NLO+PS matching methods”, *JHEP* 2012, 49 (2012), DOI: 10.1007/JHEP09(2012)049.

[277] S. Catani et al., “QCD Matrix Elements + Parton Showers”, *JHEP* 2001, 63 (2001), DOI: 10.1088/1126-6708/2001/11/063.

[278] ATLAS Collaboration, "Improved electron reconstruction in ATLAS using the Gaussian Sum Filter-based model for bremsstrahlung", ATLAS-CONF-2012-047 (2012), URL: https://cds.cern.ch/record/1449796.

[279] ATLAS Collaboration, "Electron and photon reconstruction and performance in ATLAS using a dynamical, topological cell clustering-based approach", ATL-PHYS-PUB-2017-022 (2017), URL: https://cds.cern.ch/record/2298955.

[280] ATLAS Collaboration, "Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton-proton collision data", *JINST* 14, P12006 (2019), DOI: 10.1088/1748-0221/14/12/P12006.

[281] ATLAS Collaboration, "Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton-proton collision data at $\sqrt{s} = 13$ TeV", *Eur. Phys. J. C* 79, 639 (2019), DOI: 10.1140/epjc/s10052-019-7140-6.

[282] ATLAS Collaboration, "Performance of the ATLAS Track Reconstruction Algorithms in Dense Environments in LHC run 2", *Eur. Phys. J. C* 77, 673 (2017), DOI: 10.1140/epjc/s10052-017-5225-7.

[283] ATLAS Collaboration, "Reconstruction of primary vertices at the ATLAS experiment in Run 1 proton–proton collisions at the LHC", *Eur. Phys. J. C* 77, 332 (2017), DOI: 10.1140/epjc/s10052-017-4887-5.

[284] ATLAS Collaboration, "Vertex Reconstruction Performance of the ATLAS Detector at $\sqrt{s} = 13$ TeV", ATL-PHYS-PUB-2015-026 (2015), URL: https://cds.cern.ch/record/2037717.

[285] ATLAS Collaboration, "Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s} = 13$ TeV", *Eur. Phys. J. C* 76, 292 (2016), DOI: 10.1140/epjc/s10052-016-4120-y.

[286] ATLAS Collaboration, "Muon reconstruction and identification efficiency in ATLAS using the full Run 2 pp collision data set at $\sqrt{s} = 13$ TeV", *Eur. Phys. J. C* 81, 578 (2021), DOI: 10.1140/epjc/s10052-021-09233-2.

[287] ATLAS Collaboration, "Reconstruction of hadronic decay products of tau leptons with the ATLAS experiment", *Eur. Phys. J. C* 76, 295 (2016), DOI: 10.1140/epjc/s10052-016-4110-0.

[288] ATLAS Collaboration, "Measurement of the tau lepton reconstruction and identification performance in the ATLAS experiment using pp collisions at $\sqrt{s} = 13$ TeV", ATLAS-CONF-2017-029 (2017), URL: https://cds.cern.ch/record/2261772.

[289] W. Lampl et al., "Calorimeter Clustering Algorithms: Description and Performance", ATL-LARG-PUB-2008-002 (2008), URL: https://cds.cern.ch/record/1099735.

[290] ATLAS Collaboration, "Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1", *Eur. Phys. J. C* 77, 490 (2017), DOI: 10.1140/epjc/s10052-017-5004-5.

[291] ATLAS Collaboration, "Selection of jets produced in 13 TeV proton-proton collisions with the ATLAS detector", ATLAS-CONF-2015-029 (2015), URL: https://cds.cern.ch/record/2037702.

[292] ATLAS Collaboration, "Performance of pile-up mitigation techniques for jets in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector", *Eur. Phys. J. C* 76, 581 (2016), DOI: 10.1140/epjc/s10052-016-4395-z.

[293] ATLAS Collaboration, "Jet energy scale measurements and their systematic uncertainties in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector", *Phys. Rev. D* 96, 072002 (2017), DOI: 10.1103/PhysRevD.96.072002.

[294] ATLAS Collaboration, "Jet energy scale and resolution measured in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector", *Eur. Phys. J. C* 81, 689 (2021), DOI: 10.1140/epjc/s10052-021-09402-3.

[295] J. D. Bjorken, "Properties of hadron distributions in reactions containing very heavy quarks", *Phys. Rev. D* 17, 171 (1978), DOI: 10.1103/PhysRevD.17.171.

[296] ATLAS Collaboration, "Comparison of Monte Carlo generator predictions for bottom and charm hadrons in the decays of top quarks and the fragmentation of high-p_T jets", ATL-PHYS-PUB-2014-008 (2014), URL: http://cds.cern.ch/record/1709132.

[297] 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, 970 (2019), DOI: 10.1140/epjc/s10052-019-7450-8.

[298] ATLAS Collaboration, "Measurement of b-tagging Efficiency of c-jets in $t\bar{t}$ Events Using a Likelihood Approach with the ATLAS Detector", ATLAS-CONF-2018-001 (2018), URL: https://cds.cern.ch/record/2306649.

[299] ATLAS Collaboration, "Calibration of light-flavour b-jet mistagging rates using ATLAS proton-proton collision data at $\sqrt{s} = 13$ TeV", ATLAS-CONF-2018-006 (2018), URL: https://cds.cern.ch/record/2314418.

[300] ATLAS Collaboration, "Performance of missing transverse momentum reconstruction with the ATLAS detector using proton-proton collisions at $\sqrt{s} = 13$ TeV", *Eur. Phys. J. C* 78, 903 (2018), DOI: 10.1140/epjc/s10052-018-6288-9.

[301] K. Cranmer, J. Pavez, and G. Louppe, "Approximating Likelihood Ratios with Calibrated Discriminative Classifiers", arXiv: 1506.02169 [stat.AP] (2015).

[302] J. Neyman and E. S. Pearson, "On the problem of the most efficient tests of statistical hypotheses", *Philos. trans. R. Soc. Lond. A* 231, 289–337 (1933), DOI: 10.1098/rsta.1933.0009.

[303] A. Hoecker et al., "TMVA - Toolkit for Multivariate Data Analysis", arXiv: physics/0703039 [physics.data-an] (2007).

[304] Gonçalves D. and J. Nakamura, "Role of the Z polarization in the $H \to b\bar{b}$ measurement", *Phys. Rev. D* 98, 093005 (2018), DOI: 10.1103/PhysRevD.98.093005.

[305] ATLAS Collaboration, "E_T^{miss} performance in the ATLAS detector using 2015-2016 LHC pp collisions", ATLAS-CONF-2018-023 (2018), URL: https://cds.cern.ch/record/2625233.

[306] J. Alwall et al., "MadGraph 5: Going Beyond", *JHEP* 2011, 128 (2011), DOI: 10.1007/JHEP06(2011)128.

[307] J. Alwall, R. Frederix, S. Frixione, et al., "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations", *JHEP* 2014, 79 (2014), DOI: 10.1007/JHEP07(2014)079.

[308] J. Bellm et al., "Herwig 7.0 / Herwig++ 3.0 release note", *Eur. Phys. J. C* 76, 196 (2016), DOI: 10.1140/epjc/s10052-016-4018-8.

[309] M. Bähr et al., "Herwig++ physics and manual", *Eur. Phys. J. C* 58, 639–707 (2008), DOI: 10.1140/epjc/s10052-008-0798-9.

[310] LHC Higgs Cross Section Working Group, "Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables", CERN-2011-002 (2011), URL: http://cds.cern.ch/record/1318996.

[311] LHC Higgs Cross Section Working Group, "Handbook of LHC Higgs Cross Sections: 2. Differential Distributions", CERN-2012-002 (2012), URL: http://cds.cern.ch/record/1416519.

[312] J. Bendavid et al., "Les Houches 2017: Physics at TeV Colliders Standard Model Working Group Report", arXiv: 1803.07977 [hep-ph] (2018).

[313] I. W. Stewart and F. J. Tackmann, "Theory uncertainties for Higgs mass and other searches using jet bins", *Phys. Rev. D* 85, 034011 (2012), DOI: 10.1103/PhysRevD.85.034011.

[314] M. Kendall and A. Stuart, "The Advanced Theory of Statistics", vol. 2, Hafner Publishing Company, New York (1961).

[315] K. Cranmer et al., "HistFactory: A tool for creating statistical models for use with RooFit and RooStats", CERN-OPEN-2012-016 (2012), URL: https://cds.cern.ch/record/1456844.

[316] G. Cowan et al., "Asymptotic formulae for likelihood-based tests of new physics", *Eur. Phys. J. C* 71, 1554 (2011), DOI: 10.1140/epjc/s10052-011-1554-0.

[317] S. S. Wilks, "The large-sample distribution of the likelihood ratio for testing composite hypotheses", *Ann. Math. Statist.* 9, 60–62 (1938).

[318] A. Wald, "Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large", *Trans. Amer. Math. Soc.* 54, 426–482 (1943).

[319] F. James and M. Roos, "Minuit - a system for function minimization and analysis of the parameter errors and correlations", *Comput. Phys. Commun.* 10, 343–367 (1975), DOI: 10.1016/0010-4655(75)90039-9.

[320] W. Verkerke and D. Kirkby, "The RooFit toolkit for data modeling", arXiv: physics/0306116 [physics.data-an] (2003).

[321] L. Moneta et al., "The RooStats project", *PoS* ACAT2010, 57 (2011), DOI: 10.22323/1.093.0057.

[322] R. Barlow and C. Beeston, "Fitting using finite Monte Carlo samples", *Comput. Phys. Commun.* 77, 219–228 (1993), DOI: 10.1016/0010-4655(93)90005-W.

[323] ATLAS Collaboration, "Measurements of WH and ZH production in the $H \to b\bar{b}$ decay channel in pp collisions at 13 TeV with the ATLAS detector", HEPData (2020), DOI: 10.17182/hepdata.94800.

[324] ATLAS Collaboration, "Methodology for EFT interpretation of Higgs boson Simplified Template Cross-section results in ATLAS", ATL-PHYS-PUB-2019-042 (2019), URL: https://cds.cern.ch/record/2694284.

[325] A. Azatov et al., "Helicity Selection Rules and Non-Interference for BSM Amplitudes", *Phys. Rev. D* 2017, 065014 (2017), DOI: 10.1103/PhysRevD.95.065014.

[326] C. Hays et al., "On the impact of dimension-eight SMEFT operators on Higgs measurements", *JHEP* 2019, 123 (2019), DOI: 10.1007/JHEP02(2019)123.

[327] S. Alte, M. König, and W. Shepherd, "Consistent searches for SMEFT effects in non-resonant dilepton events", *JHEP* 2019, 144 (2019), DOI: 10.1007/JHEP07(2019)144.

[328] I. Brivio et al., "Truncation, validity, uncertainties", CERN-LHCEFTWG-2021-002 (2022), URL: https://cds.cern.ch/record/2798082.

[329] A. Buckley et al., "Rivet user manual", *Comput. Phys. Commun.* 184, 2803–2819 (2013), DOI: 10.1016/j.cpc.2013.05.021.

[330] J. Brehmer et al., "Benchmarking simplified template cross sections in WH production", *JHEP* 2019, 34 (2019), DOI: 10.1007/JHEP11(2019)034.

[331] C. Degrande et al., “Electroweak Higgs boson production in the standard model effective field theory beyond leading order in QCD”, *Eur. Phys. J. C* 77, 262 (2017), DOI: 10.1140/epjc/s10052-017-4793-x.

[332] J. M. Butterworth et al., “Jet substructure as a new Higgs search channel at the LHC”, *AIP Conf. Proc.* 1078, 189–191 (2008), DOI: 10.1063/1.3051904.

[333] ATLAS Collaboration, “In situ calibration of large-radius jet energy and mass in 13 TeV proton-proton collisions with the ATLAS detector”, *Eur. Phys. J. C* 79, 135 (2019), DOI: 10.1140/epjc/s10052-019-6632-8.

[334] D. Krohn, J. Thaler, and L.-T. Wang, “Jet trimming”, *JHEP* 2010, 84 (2010), DOI: 10.1007/JHEP02(2010)084.

[335] ATLAS Collaboration, “Performance of jet substructure techniques for large-R jets in proton-proton collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector”, *JHEP* 2013, 76 (2013), DOI: 10.1007/JHEP09(2013)076.

[336] D. Krohn, J. Thaler, and L.-T. Wang, “Jets with variable R”, *JHEP* 2009, 59 (2009), DOI: 10.1088/1126-6708/2009/06/059.

[337] ATLAS Collaboration, “Variable Radius, Exclusive-k_T, and Center-of-Mass Subjet Reconstruction for Higgs$\rightarrow b\bar{b}$ Tagging in ATLAS”, ATL-PHYS-PUB-2017-010 (2017), URL: https://cds.cern.ch/record/2268678.

[338] ATLAS Collaboration, “Identification of boosted Higgs bosons decaying into b-quark pairs with the ATLAS detector at 13 TeV”, *Eur. Phys. J. C* 79, 836 (2019), DOI: 10.1140/epjc/s10052-019-7335-x.

[339] ATLAS Collaboration, “Combined measurements of Higgs boson production and decay using up to $139\,\text{fb}^{-1}$ of proton-proton collision data at $\sqrt{s} = 13$ TeV collected with the ATLAS experiment”, ATLAS-CONF-2021-053 (2021), URL: http://cds.cern.ch/record/2789544.

[340] E. D. Palik, ed., “Handbook of Optical Constants of Solids”, Academic Press (1997).