LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN
FAKULTÄT FÜR PHYSIK

MASTERARBEIT

Entwicklung eines Triggers für versetzte Vertices zur Suche nach dunkler Materie am Belle II Experiment

Developing a Displaced Vertex Trigger for Dark Matter Searches at the Belle II Experiment

vorgelegt von
Elia Schmidt

unter Betreuung von
Prof. Dr. Christian Kiesling

März 2023

Abstract

For many well-motivated dark matter models, discovery potential is predicted at high-intensity particle colliders such as the Belle II experiment. A particularly interesting scenario is inelastic dark matter which features a long-lived scattering partner in addition to the dark matter particle. At Belle II it manifests itself through the signature of large missing energy with two associated tracks that originate at vertices far displaced from the primary collision point with almost no Standard Model background. However, in large parts of parameter space the Belle II trigger system is inefficient, and by exhausting the capabilities of the detector, new limits could be set on many dark matter models.

In this thesis, a trigger system is presented that significantly increases the sensitivity of the Belle II experiment towards displaced vertex signatures. Like the standard track triggers of the experiment, the displaced vertex trigger relies on the Hough transform of 2D hits into a generalised parameter space, where tracks can be found using clustering algorithms. The Hough transform is performed in parallel on multiple vertex hypotheses selecting a subgroup of probable vertex candidates. These candidates are then analyzed by a Neural Network in order to reliably discriminate background tracks. The algorithm is written in python and prepared to be implemented on programmable trigger boards installed at the Belle II facility, guaranteeing a low latency and a fast trigger response.

The trigger is tested in Monte Carlo (MC) simulations on example models, achieving an event efficiency between 40% and 80% at a fake rate of 1kHz. The efficiency does not significantly change in the entirety of the available parameter space and is found to be robust despite challenging background conditions.

Contents

Introduction

The Standard Model of particle physics (SM) is the pinnacle of modern science: No other theory has ever predicted measurements to the precision that the SM achieves routinely. The flagship measurement of quantum electrodynamics (QED), the anomalous magnetic moment of the electron, agrees to more than 10 decimal places with the predicted value - calculated within the framework of the SM - making it the most accurate scientific prediction in history. And yet, in spite of its enormous success, even the SM is an incomplete theory. Most notably, starting in the 1930s, evidence has piled up suggesting the existence of an unknown kind of matter not described in the SM. The motions of stars and galaxies seem to imply a gravitational pull far greater than what is expected from visible matter, pointing towards a new kind of invisible source of mass - dark matter (DM). Recent measurements are consistent with an overabundance of dark matter over visible matter of as much as 5 times - standing in stark contrast to the otherwise so precise SM.

Following the growing number of calculations of the abundance and distribution of dark matter based on independent measurements, scientists have started looking for simple theoretical frameworks fit to explain its peculiar nature. Many experiments were devised with the sole purpose of detecting some form of interaction - with no success. Therefore, a second generation of dark matter models has emerged - slightly more complicated, in order to also explain the reason for its non-detection in experiments.

Many of the newer dark matter models do promise discovery potential at particle colliders. Due to the extremely small interaction cross sections and the unusual signatures expected from dark matter, searches have to be extremely precise and require large statistics. Since dark matter is untraceable, the detectors must also be hermetic, i.e. they have to cover the whole solid angle around the interaction point. With such machines, dark matter can be indirectly measured as energy missing in the detector. Colliders at the high-intensity frontier, such as Belle II at the KEK facility located in Tsukuba, Japan, fulfill these requirements and are therefore best suited for the search for dark matter.

The Belle II detector is fed by SuperKEKB, an 11GeV electron-positron collider newly upgraded to reach a record breaking instantaneous luminosity of $6 \times 10^{35} \mathrm{cm}^{-2}\mathrm{s}^{-1}$. In order not to flood data with uninteresting background events, trigger systems are employed to only select events that are of interest for physics. Unfortunately, many dark matter models fall into a blind spot of triggers, as the interactions oftentimes involve long lived particles that decay into visible tracks at unusually large displacements from the primary collision point. As it turns out, many scenarios exist where SuperKEKB is actually capable of producing dark matter but Belle II is simply not triggering. This thesis presents an algorithm for a displaced vertex trigger (DVT) that greatly increases the sensitivity of the

Belle II detector towards tracks from displaced vertices. Since the experiment is currently shut down for upgrading and expected to resume data taking in early 2024, prospects for a timely implementation of the displaced vertex trigger are realistic. Therefore, the thesis serves as a proof-of-concept that a DVT is both efficient in theory and also within the hardware capabilities of Belle II.

The thesis is organized into 3 main parts: First, the evidence for the existence and distribution of dark matter, as well as its cosmology and particle nature are presented. The most basic theoretical models and experimental attempts at their detection are discussed, with a special emphasis on the framework of inelastic dark matter (iDM), which is a compelling minimal extension of the SM into the dark sector. The necessity of a displaced vertex trigger system that allows Belle II to catch events from iDM decays is motivated. Part two revolves around the experiment, its blueprint and the working principles of all of its subdetectors. It is explained how displaced signatures, as expected from iDM and other models, evade detection and how the ongoing upgrade of the Belle II hardware opens the door to new and more complex triggering strategies, including a displaced vertex trigger. Finally, the DVT algorithm is explained in detail in the third part: The new preprocessing system that it requires, the core concept of MacroCells and the algorithm that lies at its heart - the Hough transform. The DVT is tested in Monte Carlo simulations on an example scenario of inelastic dark matter, covering a large region of parameter space. Lastly, the model-dependency of the DVT and sources to potential biases are investigated.

Part I

Evidence and Theory

Chapter 1

Evidence for Dark Matter

A problem has plagued physicists for a long time: The contents of the universe seem to be very different depending on how they are measured: Mass can be measured photometrically by simply adding up all luminous matter in a given region of space. In order to catch very cold and very hot matter, all wavelengths of light are best taken into consideration. Another way is to measure the gravitational forces that matter exerts on stars and dust. The more trajectories are bent, the larger the gravitational pull and thus the attracting mass must be. Since these measurements are not particularly demanding technically, scientists have been able to perform them for a very long time, and have been finding large discrepancies in their results consistently. The most important measurements of this kind are explained in the following section, with increasing degree of complexity.

1.1   Fritz Zwicky and Virialized Galaxy Clusters

The term dark matter ("dunkle Materie") is usually attributed to swiss astronomer Fritz Zwicky, who suggested its existence in order to explain the great discrepancies between the measured luminous masses of "clusters of nebulae", which are now known to be galaxies, and their gravitative masses as derived from basic mechanical principles. In a paper submitted to the Astrophysical Journal in 1937 [1], Zwicky uses the virial theorem in order to calculate the total mass of the Coma cluster:

The virial theorem is a frequently used tool in classical and quantum mechanics, by which the average kinetic energy of the constituents of a dynamical system can be related to the potential that governs their motion. Applied to galaxy clusters, it allows to infer the mass of the entire system from the rather simple observation of the peculiar velocities of the galaxies within. Starting from the moment of inertia I, one can calculate the second derivative to find an expression for the kinetic energy and the so called virial:

$$\frac{1}{2}\frac{d^2 I}{dt^2} = \frac{1}{2}\frac{d^2}{dt^2}\left[\sum_k^N m_k \mathbf{r}_k^2\right] = \sum_k^N m_k \dot{\mathbf{r}}_k{}^2 + \sum_k^N m_k \mathbf{r}_k \ddot{\mathbf{r}}_k = 2\mathrm{E_{kin}} + \mathrm{Vir} \qquad (1.1.1)$$

The virial is simply the product of the force acted upon a galaxy by all the other galaxies in the cluster and its position vector, summed over all members k. These forces can be expressed as the gradient of the gravitational potential, such that

$$\text{Vir} \equiv \sum_k^N \mathbf{F}_k \mathbf{r}_k = -\sum_{k=2}^N \sum_{j=1}^{k-1} \frac{dV_{jk}}{dr_{jk}} r_{jk}. \tag{1.1.2}$$

Here, I have replaced the position vector $\mathbf{r}_k$ with the relative distance r_{jk} between galaxies j and k. In particular, if the potential is a power law of the distance vector,

$$\text{Vir} = -nV_{tot}, \tag{1.1.3}$$

which will later become important. If galaxies are approximated as spheres of uniform density and Newton's inverse square law of gravity is applied[1], equation 1.1.2 can be solved explicitly to find:

$$\text{Vir} = -G_N \frac{3M_{tot}^2}{5R_{tot}} \tag{1.1.4}$$

M_{tot} and R_{tot} denote the cluster's total mass and radius, respectively, while G_N is Newton's constant. The time derivative of equation 1.1.1 is now taken. If the cluster is sufficiently *virialized*, meaning that its moment of inertia averaged over a large time does not change, the left hand side vanishes. This is similar to thermal equilibrium in thermodynamics (in fact, the virial theorem is related to the equipartition theorem of statistical mechanics). Therefore, there is now a direct proportionality between the potential and the averaged kinetic energy, and the mass of the cluster can be calculated:

$$M_{tot} = \frac{5R_{tot}\bar{\bar{v}}^2}{3G_N}. \tag{1.1.5}$$

The twice barred v denotes the peculiar velocity of galaxies in the cluster averaged over time and mass, which can be measured: Zwicky observed the velocities along the line of sight of the constituents of the Coma cluster with an 18-inch telescope situated on Mount Palomar in California and concluded that the virial mass of the system was 500 times larger than the visible mass. Of course, not all galaxy clusters are virialized: Collisions between clusters often happen and the virial assumption only applies once the dust has settled. However, even today's measurements, done using modern astronomical tools and on larger statistical sets, give results in the same order of magnitude [2], in spite of the fact that Zwicky's calculation relied on a few assumptions now known to be incorrect (including the distance of the coma cluster and therefore its physical size).

Though the calculation points towards the existence of dark matter that contributes to the virial mass but escapes detection using telescopes, there is a second way out, as Zwicky himself acknowledges: He states that "the viral theorem as applied to clusters of nebulae [galaxies] provides a test of the validity of the inverse square law of gravitational forces". Looking at equation 1.1.3, it becomes clear that if a different power law of gravitation applies at very large distance scales, the virial changes proportionally to the exponent, possibly reconciling the discrepancy with the luminous mass measurements. This is the essence of MOdified Newtonian Dynamics (MONDs), as they represent an alternative explanation

---

[1]Also, one has to use the shell theorem first proven by Newton. It will be used extensively in the next section about the rotation curves of galaxies.

that makes dark matter (partially) obsolete. However, it will become clear when looking at more recent observations (specifically, mass distributions in colliding galaxy clusters) that MONDs have a hard time accommodating the ever increasing set of independent evidence in favor of dark matter.

1.2   Galactic Rotation Curves

The measurement of the rotation curves of galaxies is, in a way, a more quantitative version of the virial mass calculation from the previous section - done on a smaller scale. The velocity of stars in their orbit around the galactic center can be measured very precisely using the optical Doppler shift of prominent emission lines such as H-alpha and can be used to infer the mass of the galaxy.

Approximating the profile of a galaxy as a sphere of equal density and radius R, the orbit velocity of a probe mass can be calculated easily by invoking Newton's aforementioned shell theorem: The gravitational force on a body outside a closed spherical shell of matter is the same as it would be if all the shell's matter were concentrated into a point at its center. Conversely, if the body is a spherically symmetric shell, no net gravitational force is exerted by the shell on any object inside, regardless of the object's location within the shell.[2]

Thus, for any star that (circularly) orbits the center of the simplified galaxy with radius r_{star}, the force it feels is equal to the gravitation exerted by a point mass with $M \propto r_{\text{star}}{}^3$, as all the residual mass lies within a shell of radius $R > r_{\text{shell}} > r_{\text{star}}$ and, by virtue of the shell theorem, produces no net force. Orbital motion around a point mass is governed by Kepler's law,

$$v_{\text{Kepler}}(r) = \sqrt{\frac{G_N M}{r}}. \tag{1.2.1}$$

Therefore, if the orbital velocity of the stars with respect to their galacto-centric distance is measured (this is the so called rotation curve), the resulting curve is expected to split into two regimes:

$$v(r) \propto \begin{cases} \sqrt{\frac{M \propto r^3}{r}} = r & : \quad r < R \\ r^{-1/2} & : \quad r > R \end{cases} \tag{1.2.2}$$

A linear rise to the edge of the sphere, followed by a Keplerian fall-off for larger radii. Though the approximation is quite limiting - galaxies aren't balls of equal density - even more rigorous analytical calculations as well as numerical studies indicate a qualitatively similar picture, as outlined in [3].

However, real galaxies behave significantly differently. By the 1970s, measurements of rotational curves were getting precise enough to clearly indicate a departure from purely Keplerian motion at larger distances from the galactic core. Freeman [4] was the first one to suggest that "if they [the observational data] are correct, then there must be in these

---

[2]The proof of this theorem is tedious though straightforward. Mainly, it involves the calculation of the gravitational pull of 1) a ring, 2) a disc and finally 3) a sphere. The corresponding Wikipedia page features a detailed explanation.

galaxies additional matter which is undetected, either optically or at 21 cm. Its mass must be at least as large as the mass of the detected galaxy, and its distributions must be quite different from the exponential distributions which holds for the optical galaxy." More recent studies (for example [5]) support this hypothesis and give quantitative measures for the dark matter mass and distribution inside galaxies.

Figure 1.1 shows such a galactic rotation curve: In the innermost region it rises quickly due to the high matter density and gravitative friction that makes the galaxy behave like a rigid body. The steep nuclear rise is followed by a flat region that extends to the edge of the (visible) galaxy. The expected Keplerian fall-off (the curves labelled disk and gas) is not observed. In order to reconcile the laws of gravity with the visible matter distribution, a large massive halo of dark matter that encloses the galaxy must be postulated. The total mass of this dark matter halo is about 5 times larger than the visible components.
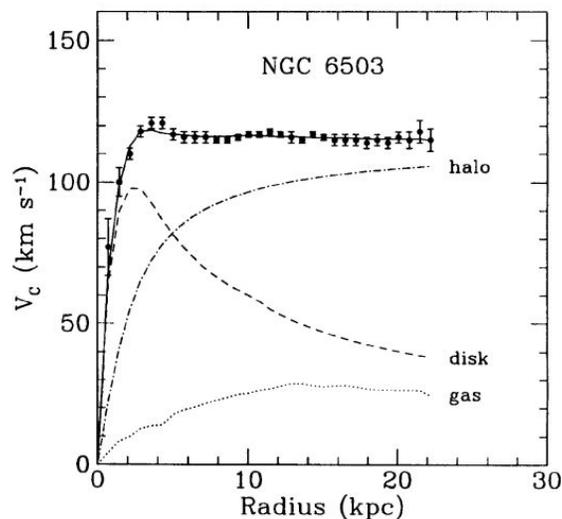
Figure 1.1: Rotation curve of spiral galaxy NGC 6503. The dotted and the dashed lines represent the Keplerian rotation curves expected from the measured disk and gas distributions. The dash-dotted line shows the dark matter distribution needed to recover a curve consistent with the observed data (solid line). Plot taken from the NASA/IPAC Extragalactic Database [6].

A constant rotation curve implies a mass distribution linear in r, as can be seen by inserting $M \propto r$ into equation 1.2.1. This corresponds to a density distribution $\rho(r) \propto r^{-2}$. In practice, numerical simulations of dark matter dynamics often use the so called isothermal sphere profile,

$$\rho(r) = \frac{\rho_0}{1 + (r/a)^\gamma}, \tag{1.2.3}$$

where ρ_0 is the DM density at $r = 0$ and a is the radius where the density has dropped off by half. Usually, $\gamma \approx 2$. This distribution shares the inverse square behavior for large radii but also accounts for the quasi-linear rise for the innermost region.

From this information two things about the nature of dark matter can be inferred: It

interacts with Standard Model particles very weakly at most, otherwise it would have been detected directly. But it also barely interacts with itself, or else the dark matter halo would have accreted to a disk due to scattering like ordinary matter. All mechanisms by which energy is dissipated are inefficient and therefore dark matter halos can exist for a very long time.

1.3 The Bullet Cluster

In the last two sections I arrived at two very different figures for the ratio of dark to luminous matter. In galaxy clusters, dark matter is a few hundred times more abundant than ordinary matter (in terms of mass), whereas the ratio is only about 5:1 in individual galaxies. How can these values be reconciled? The answer lies in the fact that galaxies are not the only contributors to the total mass of a galaxy cluster. In fact, most of baryonic matter is not gravitationally bound to galaxies and floats around in space as interstellar gas. Since it is trapped in the deep gravitational well of the cluster, it has a lot of energy - i.e. it is very hot and therefore constantly emits X-rays due to electromagnetic interactions.

Modern space telescopes, such as the *Chandra X-ray Observatory* can detect and accurately spatially resolve such X-rays to show the distribution of hot gas in galaxy clusters. Studies such as [7] report that interstellar gas accounts for approximately 90% of the baryonic mass. Only 10% is optically detectable matter inside galaxies. Of the *total* mass inside galaxy clusters, this means that $\sim 10\%$ is interstellar gas, $\sim 1\%$ is the galaxies and the remaining $\sim 89\%$ is *truly* dark matter.

This fact has a very important consequence for colliding galaxy clusters, such as the so called *Bullet Cluster*, which has allowed some of the most compelling evidence for the dark matter paradigm over MONDs: When two galaxy clusters collide, their constituents behave differently: Galaxies are very scarce and far away from each other, thus they most likely do not collide - they simply pass through each other. This can be quantified by calculating the mean free path of a galaxy travelling through a galaxy cluster: A typical galaxy cluster contains of the order 100-1000 galaxies and extends over a sphere with radius 0.5-2.5Mpc. An individual galaxy, on the other hand, can be approximated as a sphere of radius 1kpc. The mean free path of a galaxy then is

$$\lambda_{\text{free}} = \frac{1}{n\sigma}, \tag{1.3.1}$$

where n is the number density of galaxies and σ the cross section of two galaxies colliding. The former is easy to calculate and for argument's sake the most conservative value is taken (so the highest possible number density: 1000 galaxies in a 0.5Mpc sphere). As a first approximation, I use the classical cross section of two hard spheres colliding. For identical spheres, it is equal to $\pi(2r)^2$. With these numbers, the result is

$$\lambda_{\text{free}} = \frac{4/3\pi(0.5\text{Mpc})^3}{1000}\frac{1}{\pi(2\text{kpc})^2} = \frac{1}{24} \times 10^9\text{pc} \approx 100 \times R_{\text{cluster}}. \tag{1.3.2}$$

On average, a galaxy needs to pass 100 clusters in order to run into another galaxy. Therefore, it is safe to assume the galactic mass component to be mostly unaffected by the

"collision" of the clusters. Hot interstellar gas, on the other hand, is widespread and interacts not only gravitationally, but also electromagnetically. Therefore, it experiences ram pressure and essentially clumps up in the middle region due to friction as the two clusters pass through each other. As shown in the last section, dark matter does not interact much with itself. Hence, it should follow the distribution of the galaxies rather than the hot gas. If dark matter exists, it is by far the heaviest component of galaxy clusters, significantly affecting the overall mass distribution of the colliding galaxy clusters. It is expected to follow a dumbbell-shape, as the hot gas in the middle accounts only for $\sim 10\%$ of the total mass.
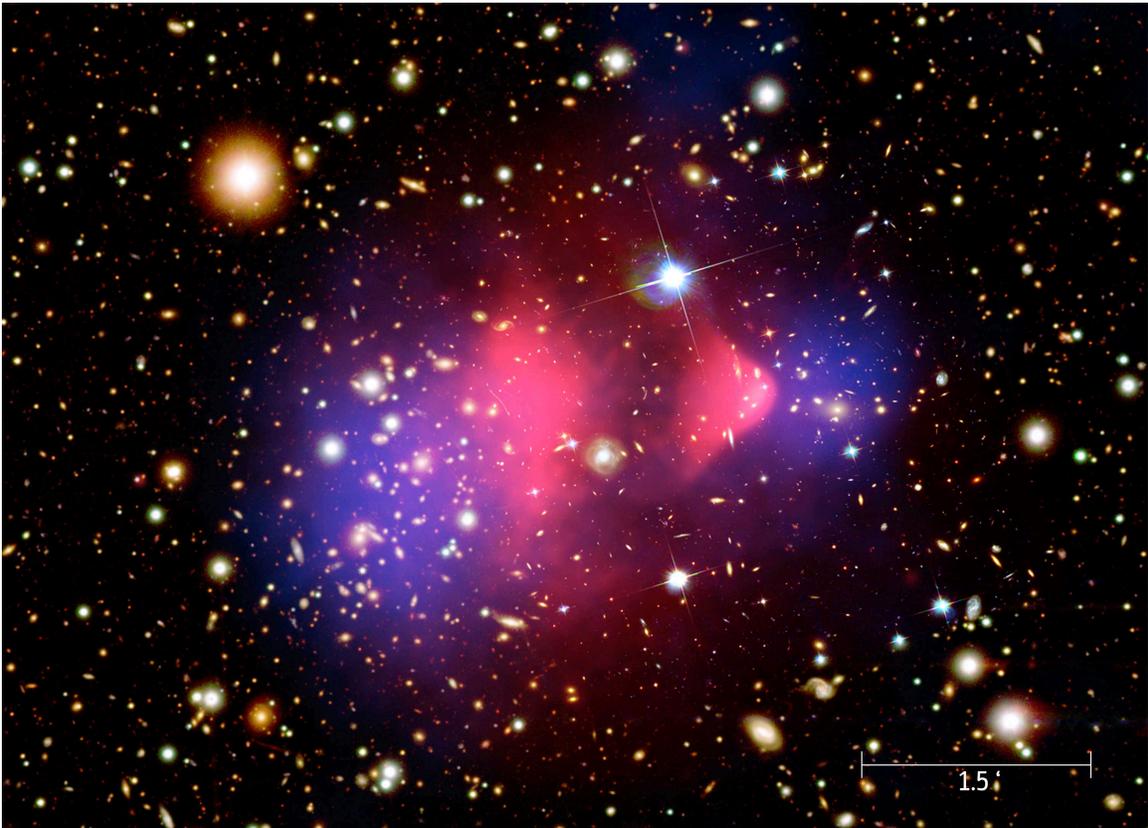
Figure 1.2: An image of the Bullet Cluster. X-Ray data (pink) is superimposed on an image taken in the visible spectrum. Image taken by the Chandra X-ray Observatory. Credit: NASA/CXC/SAO.

Figure 1.2 shows very well how the mass components were neatly separated during the collision: The galaxies have completed pass-through and are now on opposite sides of the structure. At the same time, X-Ray imaging reveals the distribution of the interstellar gas: A bow-shaped shock wave can be seen travelling through the gas, clearly indicating ram pressure. The gas components were stripped from the galaxies and remain in the middle region, closer to the center of mass.

Weak gravitational lensing is a very powerful tool used in astronomy to measure the absolute mass distribution of large-scale objects (like galaxy clusters) without depending

on their dynamics. In general relativity, light travels on geodesics of the spacetime manifold, which is bent in the presence of mass. Large masses can therefore deflect light off its straight (in the Euclidean sense) path and act as a lens. In extreme cases, this can result in the image of a distant galaxy being severely warped by a heavy object in line of sight. The galaxy may appear on either side of the heavy object or even in a complete ring around it. This is then called strong gravitational lensing. Weak lensing, on the other hand, is more subtle and involves the slight elongation and magnification of background objects to unnatural shapes without changing their overall brightness. Analysing how a field of background galaxies is bent and warped one can infer the mass distribution of the object that is in front. In [8] this weak lensing analysis was carried out in order to reconstruct the mass distribution of the Bullet Cluster.

The test is simple: If the dark matter paradigm is correct, then dark matter and the galaxies outweigh the interstellar gas, resulting in a dumbbell-shaped absolute mass distribution. If dark matter does not exist, interstellar gas makes up roughly 90% of the total cluster mass, significantly affecting the shape of the distribution. It must be much closer to an elongated ball, as the galaxies make up only a small amount of the mass.
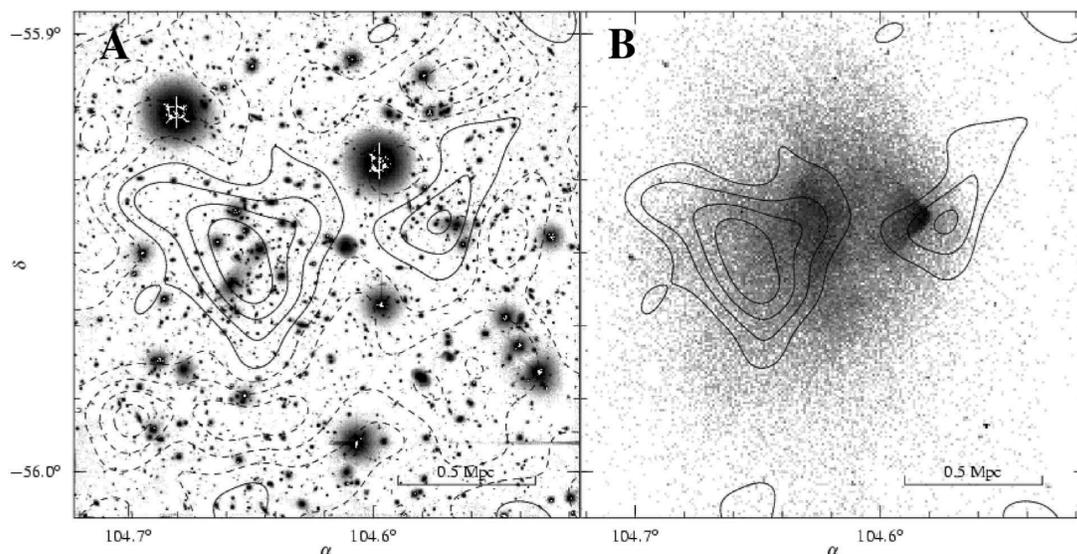
Figure 1.3: Left: Background galaxies used for lensing analysis. The inferred mass distribution is shown as contours of equal mass. Right: The mass distribution overlaid onto the X-Ray image of the Bullet cluster. This image was taken from the paper [8].

Figure 1.3 summarizes the results: The contour plots of the mass distribution clearly indicate a dumbbell shape with two sub-centers of mass neatly separated from each other. The interstellar gas does not seem to affect the mass distribution, as it is mostly located between the sub-centers. The 3 authors of the paper conclude that the mass distribution in the Bullet Cluster is "closer to, and consistent with, the optical component". Dark matter naturally explains this behavior, while the explanation in the MOND paradigm is extremely challenging.

14

Chapter 2

Cosmological Considerations

From gravitational evidence it is known that dark matter must be very abundant, interact very rarely with the Standard Model and also rarely with itself. Cosmology provides us with more information about dark matter and also some strong constraints on its nature. The evidence, however, is much less obvious than the gravitational observations mentioned before and some cosmological considerations have to be made before being able to extract it.

2.1 Thermal Relics

A central paradigm of modern cosmology is that of *thermal freeze-out*. It is used to explain the natural abundance of the charged leptons, of neutrinos and even of some light nuclei from some simple considerations about their mass and the type of reaction in which they are involved. The central idea is that every reaction, no matter the size of the cross section and no matter the density of the particles involved, at some point in the evolution of the universe happened frequently enough to keep the particles in chemical and thermal equilibrium with each other. So even particles that very rarely interact - like dark matter - were at some point in thermal equilibrium with each other and with the Standard Model. Every dark matter model can therefore be tested by means of thermal freeze-out to see if it produces the correct so called *relic densities* - the density of the particle species after it dropped out of thermal equilibrium. It needs to match the density calculated from the gravitational (and other) evidence. Thermal freeze-out can also be used to place some stringent constraints about the mass and the scattering cross section of dark matter.

A typical example is the thermal freeze-out of neutrinos. Shortly after the Big Bang, the universe is small, extremely hot, and, crucially, extremely homogeneous. The reaction

$$\nu + \bar{\nu} \rightleftharpoons l^+ + l^- \tag{2.1.1}$$

via the exchange of a neutral Z-boson is the only reaction that mediates between neutrinos and the charged leptons.[1] At energies much below the electroweak scale (80GeV) it behaves like a 4-point interaction with a cross section proportional to $G_F^2 E^2$, where G_F is Fermi's

---

[1]The scattering reaction $\nu + l^{\pm} \rightleftharpoons \nu + l^{\pm}$ via $W^{\pm}$- exchange keeps neutrinos in *thermal* but not in *chemical* contact with charged leptons, as it does not change the number of neutrinos in the universe.

constant and E the center-of-mass energy. The reaction rate scales with the number density of neutrinos to the second inverse power and is therefore minute today. However, in the early universe, when temperature and the number density of neutrinos were extremely high, the reaction was efficient in keeping the chemical equilibrium between the two particle species. In statistical physics, the conservation of entropy can be used to derive the following expression for the number density of any particle species:

$$n_{\mathrm{rel}} \propto T^3, \tag{2.1.2}$$

$$n_{\mathrm{non-rel}} \propto (mT)^{3/2} \exp\left(-\frac{m}{T}\right). \tag{2.1.3}$$

The above equation holds for relativistic, below for non-relativistic particle species. Neutrinos are very light and are well in the relativistic regime when they freeze out. Therefore, equation 2.1.2 is used. With the universe expanding and simultaneously cooling, the number density drops rapidly until the reaction rate falls below the rate of curvature. In other words, the typical time between reactions gets longer than the age of the universe. The reaction becomes inefficient and neutrinos fall out of chemical equilibrium. Their (comoving) number density *freezes out* and remains constant from this point on. Quantitatively, the typical time between reactions can be calculated using

$$t = \frac{1}{\Gamma} = \frac{1}{n_{\mathrm{rel}} \cdot \sigma \cdot v}, \tag{2.1.4}$$

where σ is the cross section and v the relative velocity at collision. Fully in the relativistic regime and using natural units, where $c \equiv 1$, it is simply equal to unity. On the other hand, n_{rel} scales with T^3 and $\sigma \propto G_F^2 E^2$, which can also be related to the temperature, since, in natural units, $E = T$. Putting everything together, the typical time between reactions must be

$$t \propto \frac{1}{T^3 \cdot G_F^2 T^2} = \frac{1}{G_F^2 T^5}. \tag{2.1.5}$$

With an expression that relates the temperature in the universe to its age (t_{cosm}), the temperature and the moment in time at which neutrinos chemically decoupled from all other particle species can be determined. This expression can be derived from the Friedmann equations in the case of a flat, radiation-dominated universe and is covered in all introductory cosmology textbooks, such as [9]. It reads:

$$t_{\mathrm{cosm}} = \left(\frac{3}{32\pi\kappa}\right)^{1/2} T^{-2}. \tag{2.1.6}$$

κ is the effective number of internal degrees of freedom of all particles present in the universe. If there were only photons, for example, $\kappa = 2$ for the two spin states etc. Equating 2.1.5 to 2.1.6 yields a freeze-out temperature of $T_\nu \sim 1\mathrm{MeV}$. [2], [3] This result confirms that, indeed,

---

[2] The exact calculation produces a value 0.5MeV larger.

[3] Soon after chemical decoupling also the two W-mediated reactions cease to be efficient, as their cross sections are only slightly larger that the Z-mediated one due to the mass difference between the Z and W bosons.

neutrinos are relativistic when they freeze out, as T_ν vastly exceeds their rest mass. Figure 2.1 shows how the relic abundance of a particle species stays constant after it freezes out.
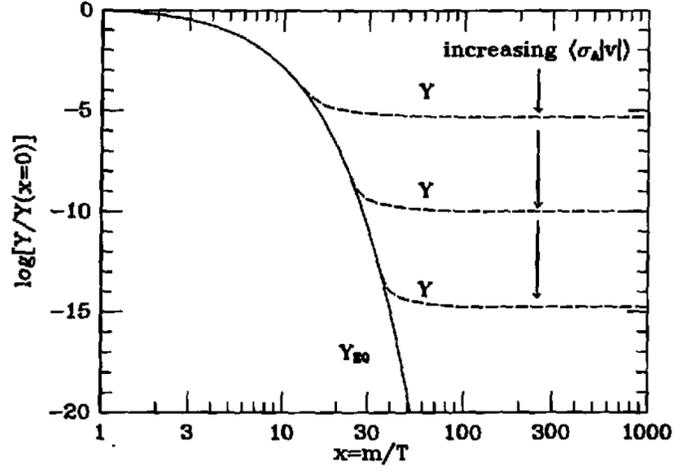
Figure 2.1: Relative abundance of a massive particle. Dashed lines: Abundances after freeze-out. Solid line: Abundance in thermal equilibrium. Image from [10].

Here lies the first distinction between two completely different types of dark matter models. Dark matter can be *hot* or *cold* depending on the temperature at which it freezes out and becomes independent. Light and weakly interacting dark matter behaves much like the neutrinos and is therefore relativistic when it decouples. Thus, it is called hot dark matter. Heavier and more interactive dark matter instead is already non-relativistic and thus cold by the time it freezes out of equilibrium.

Repeating the calculation of the freeze-out temperature for non-relativistic particles is more complicated than in the relativistic case, as 2.1.3 involves an exponential term. Assuming, for example, a dark matter particle of mass m_χ and a cross section σ for any process linking it to Standard Model particles, one needs to solve

$$\sqrt{\frac{m_\chi}{T}} \cdot \exp\left(-\frac{m_\chi}{T}\right) = \frac{\sqrt{\frac{32}{3}\pi\kappa}}{m_\chi\sigma}. \tag{2.1.7}$$

It is clear that from equation 2.1.3 the density of the particle species after freeze-out can be calculated. Due to the expansion of the universe, it will not remain constant, like the abundance, but change covariantly. Either way, for every pair of parameters m_χ and σ the corresponding relic density can be calculated. Since the actual value of the dark matter density is known, this in turn limits the viable parameter space of dark matter models to just one value for the mass for every cross section. Thus, a line can be traced in the $m_\chi - \sigma$ plane which produces the correct values. This line can be calculated to greater precision numerically and is typically shown in most exclusion plots, denoted as *thermal target*.

2.2 Anisotropies in the Cosmic Microwave Background

The Cosmic Microwave Background (CMB) contains some extremely valuable information about the universe as it is the earliest snapshot of the universe we possess. Even though it was first discovered in 1965, it is a source of new information still today. Recent high-precision measurements of its spectrum have revealed some intricate regularities that encode basic parameters of our universe, including the relative density of dark matter.
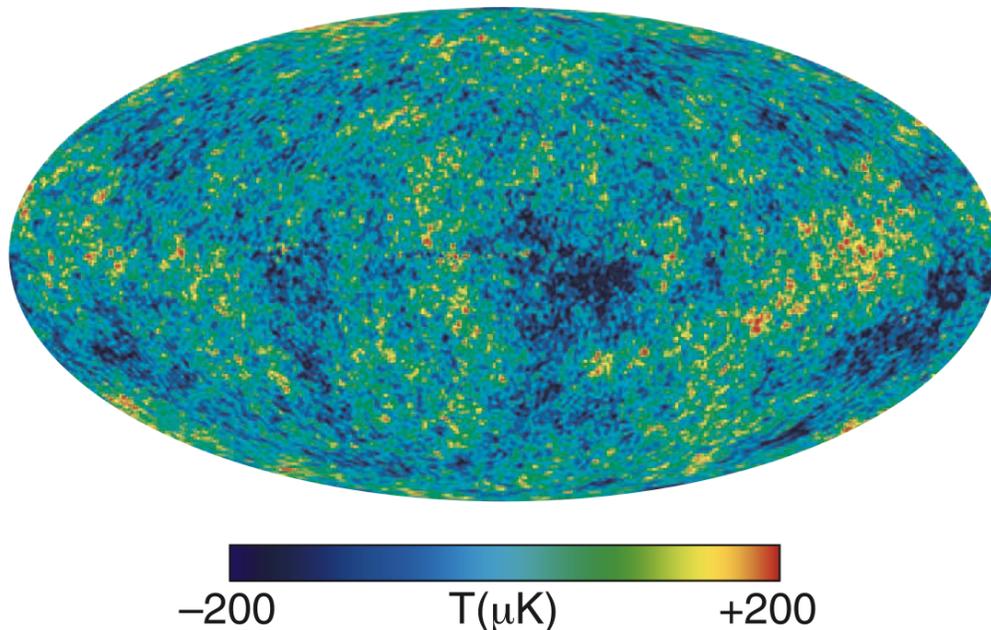
Figure 2.2: Temperature fluctuations from the mean in the CMB of the entire sky. Maps in different frequency bands were combined to subtract the milky way foreground. Image from [11].

Following the thermal freeze-out of dark matter, the universe was still hot enough to keep atoms ionized. In that state, electrons are separated from their nuclei and constantly interact with high-energy photons such that the mean free path of a photon between two (Thomson) scattering events off of electrons is very small. This in essence means that baryons and photons form a single fluid called the *Baryon-Radiation-Plasma* and that it is opaque. Anisotropies in the density of the baryons cause temperature fluctuations which in turn are directly reflected in the spectrum of the photons. As the temperature of the universe keeps decreasing, at some point it falls under the ionization-threshold. Charged nuclei combine with the free electrons to form neutral atoms that no longer interact with the photons[4] and the universe instantly becomes transparent. This effect is called recombination and it happened at $T \approx 0.34$eV, or, equivalently, 380.000 years after the Big Bang. Photons can now free stream into all directions and are still detectable today as the CMB. They

---

[4]There are of course scattering reactions between neutral atoms and photons, such as Rayleigh scattering. However, the cross section is very small and not efficient in keeping photons coupled to baryons.

form an extremely clean Planckian spectrum with a temperature of 2.75K since they were redshifted by the expansion of the universe.

Inhomogeneities in the baryon density right before recombination are therefore imprinted on the CMB and can be observed today as they were 380.000 after the Big Bang. The latest precision results were obtained by the WMAP spacecraft [11] and the resulting temperature map is shown in figure 2.2. The paradigm of *cosmic inflation* predicts inhomogeneities from quantum fluctuations that are subsequently stretched to all length scales. The amplitudes of these inhomogeneities are of order 10^{-4} with respect to the mean. At the end of inflation they are free to evolve and, according to their scale, either collapse gravitationally or behave like sound waves. The distinction is dictated by the Jeans wave length: Perturbations that are larger than the Jeans wave length collapse gravitationally, whereas small perturbations (in the baryon-radiation-plasma) are stabilized by pressure, forming sound waves. Before recombination, the Jeans wave length is very large, such that almost all perturbations propagate with sound speed $c_s \approx 0.5c$.

An overdense region after inflation has time to evolve until recombination freezes the phase of the wave and imprints it onto the CMB. Initially, dark matter and the baryon-radiation plasma share the same overdensities. After dark matter decoupling, the dark component no longer interacts and the perturbation can slowly grow due to gravitational collapse. At the same time, the baryon-radiation plasma starts to expand outwards radially at the speed of sound and essentially forms standing waves. A very important observation is the fact that since it is the overdensities to spark propagation, the distance between the two wavefronts (overdensities) is largest, while the troughs (underdensities) have half the distance from each other. Computing the spatial correlations shows that these odd numbered acoustic peaks correspond to overdensities while even numbered peaks correspond to underdensities. The larger the dark matter content is in the universe, the more gravity acts against pressure and overdensities grow faster and sharper, increasing the prominence of odd-numbered acoustic peaks and suppressing even-numbered ones. On the other hand, decreasing the dark matter content leads to pressure dominating over gravity, smearing out the overdensities and relatively boosting even-numbered peaks. A more in-depth analysis is carried out in [12], quantifying the exact amount of dark matter in order to create peaks of any given height and position. Importantly, the exact same parameters apply to all perturbations in the universe, as it is, at this point, homogeneous to one part in a thousand. Hence, the spatial correlations of all initial perturbations share the same scales and are frozen into the CMB after recombination.

It is crucial to study the spatial correlations in the CMB radiation as the location and relative height of the acoustic peaks depend sensitively on the amount of dark matter in the universe. In order to do so, a multipole expansion is carried out on the surface of the CMB-sphere. This corresponds to a two-dimensional, spherically symmetric Fourier transformation, or, in other words, a decomposition of the sky into the spherical harmonics. Some general remarks about the overall shape of the spectrum:

- The first harmonic (monopole moment, $l = 0$) is a constant and corresponds to the overall mean temperature of the CMB. It does not contain any information about the fluctuations.

- The dipole moment ($l = 1$) describes an overtemperature in one angular direction

and an undertemperature in the opposite direction due to Doppler-shift. It can be transformed away by Lorentz-boosting the system of the observer. It therefore also contains no information about the fluctuations but rather about the peculiar velocity of earth with respect to the cosmic background.

- Multipoles of order 3 and above do indeed carry information about fluctuations. The angular distance they represent can be roughly calculated by $\theta \approx \frac{60°}{l}$.

- For large angles (or small l) there is no correlation expected, as the angle corresponds to a distance at recombination that is larger than the causally connected regions.

- For very small angles (or large l) the acoustic peaks are expected to drop in intensity due to photons diffusing temperature differences in the baryon-radiation-plasma (Silk damping).
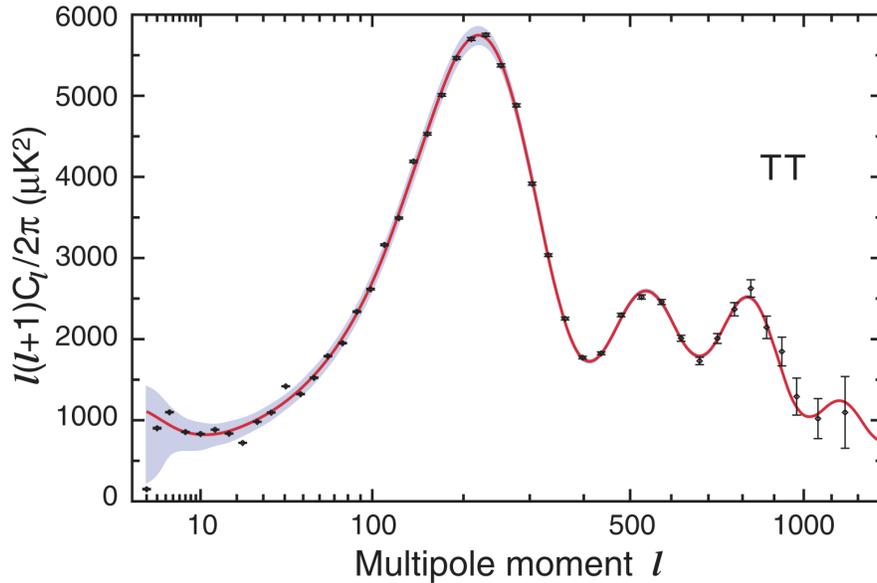
Figure 2.3: The power spectrum of the CMB radiation. The second peak is strongly suppressed with respect to the first and third. This indicates a large amount of dark matter in the universe. This image is also taken from [11].

The results obtained by the analysis of the WMAP data is shown in figure 2.3. Clearly, the first and third peak dominate over the second one, which is highly suppressed. This indicates a large amount of dark matter in the universe. Quantitatively, as reported in the table of cosmological parameters extracted from the data in [11], the calculated ratio of baryonic-to-dark matter is 0.2. This is very important as it represents a totally independent way of measuring the dark matter density that comes to the same conclusion as the analysis of gravitational evidence. From the analysis of WMAP data, the widely acclaimed Standard Model of cosmology (*concordance model*) is derived.

2.3 Baryon Acoustic Oscillations

One last and very compelling piece of evidence for the existence of dark matter comes from the distribution of visible matter in the current universe. Experiments such as the Sloan Digital Sky Survey (SDSS) measure the 3-dimensional correlations in the overall matter density (results in [13]). It is related to the power spectrum plot in the CMB. The point is the following: Given the initial conditions right after recombination with matter anisotropies in the order of one in a thousand, the evolution of these perturbations can be computed for different cosmological models. Today the inhomogeneities are evidently strongly non-linear, so from any cosmological model (MOND as well as the dark matter paradigm) we must demand that it predict first the correct scale of the inhomogeneities and, second, the correct power spectrum. This was analyzed by S. Dodelson in [14].
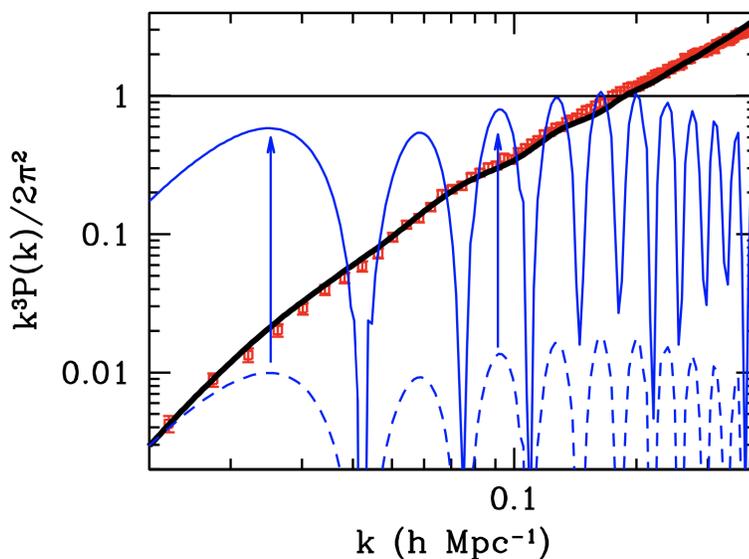
Figure 2.4: The power spectrum of the baryon acoustic oscillations. The data (red points) fit the concordance model very well (solid black line) while a universe without dark matter doesn't produce large enough inhomogeneities (dashed blue line) and TeVeS predicts peaks (solid blue line) in stark contrast to the data. This image was taken from [14]

Without dark matter, perturbations in the energy density are, until recombination, stabilized and diffused by pressure and can grow only very slowly. On the other hand, dark matter decouples from photons much earlier than baryons and has enough time to clump together. Therefore, dark matter seeds the inhomogeneities that can be seen in the CMB. Baryonic matter falls into these seeded potential wells and amplifies the gravitational instabilities. First and foremost, this enables timely structure formation and the non-linear perturbations observed today. Secondly, it strongly suppresses any correlations in the baryons similar to the CMB oscillations, as they are smeared out by the much larger underlying dark matter perturbations and essentially reduced to the single percent level.

Figure 2.4 summarizes Dodelson's results: The power spectra of the best-fit concordance model and data match very precisely. In a universe without dark matter but otherwise

identical to ours, there is simply not enough time between recombination and today for the initial 10^{-4} perturbations to grow to unity and beyond. Similarly, most MONDs cannot explain the perturbation amplitudes, either. One specific model, TeVeS (Tensor-Vector-Scalar gravity) does produce large inhomogeneities compatible with observation, but completely overpredicts oscillations in the power spectrum. Again, one can only conclude that the concordance model with a dark matter contribution to the overall density of the universe of approximately 20% is the most likely one.

Chapter 3

Dark Matter Models

If the cosmological and gravitational evidence provided in the previous chapters is accepted, and if the theoretical alternative of MONDs are discarded, still a vast array of different explanations for dark matter are left. A straight-forward classification of dark matter candidates is by the structure of their interaction with the Standard Model - also called the dark sector portal.

3.1 Dark Sector Portals

Before addressing different particle dark matter models, it is necessary to mention than not all of them even share a particle nature. An interesting alternative is the paradigm of MACHOs (massive compact halo objects). They are a class of macroscopic objects that float in space emitting little to no light with a typical mass scale somewhere between 10^{-8} and 10^2 solar masses. Some candidates are small black holes with no accretion disks around them, very light stars that are just not heavy enough to fuse nuclear hydrogen, or extremely old neutron stars and white dwarfs that have cooled down beyond detection through electromagnetic radiation. MACHOs fail to reproduce large parts of the dark matter phenomenology, namely the timely structure formation in the early universe or the shape of CMB-anisotropies mentioned previously. For completeness, however, I share a brief summary of the attempts that have been made to find MACHOs:

If present, MACHOs should be indirectly detectable by exploiting once again the microlensing effect when they pass through the line of sight to bright stars. In this context, microlensing manifests itself through variations in brightness of stars observed over periods of time between 2 hours and 2 years [15]. However, among others, measurements performed over large periods of time on the Magellanic clouds have ruled out a large portion of MACHOs as primary explanation for the missing mass in the universe [16].

Most common hypotheses for dark matter do assume new elementary particles. In order to escape detection, they have to interact very weakly with the Standard Model, but the exact form of the interaction can vary strongly between models. The simplest case is the weakly interacting massive particle, or WIMP. While originally WIMP meant a particle that could only interact via the Standard Model weak interaction, the term has come to describe any new particle with an interaction in roughly the scale of the weak force. WIMPs have

enjoyed much interest because of their theoretical simplicity and a promising observation often called the *WIMP miracle*: In the framework of thermal freeze-out, a hypothetical particle that interacts weakly and has a mass of the electroweak scale naturally suggests a relic abundance that matches the one inferred from gravitational measurements. The exact structure of the interaction with the Standard Model does not matter, so long as the cross section is of the correct magnitude. This made it a hot contender as dark matter candidates, pointing towards a new particle well within the energy scales reachable by experiments today. Recent results from direct detection, however, have ruled out large portions of the available parameter space [17] for this simple case, although it must be noted that by abandoning the assumption of the relic abundance being determined by freeze-out cosmology, search windows are still open for the WIMP hypothesis [18].

Still, for the most part, the search for dark matter has moved on to more complex models involving new mediators and symmetries. Considering the intricate complexity of the Standard Model, the existence of an entire dark sector rather than just one additional state should not be surprising. Arguably the simplest non-trivial dark sector model extends the Standard Model by an additional gauge boson called the dark photon. It couples to the Standard Model by kinetically mixing with the regular photon, creating a mechanism similar to the flavor oscillations observed in the neutrino sector. Couplings of dark matter particles to the Higgs-like particle are also possible. In fact, the model addressed in the present thesis features both the dark photon as well as the Higgs portal. It is described in detail in section 3.3. In certain regions of parameter space, dark photons can also explain the anomalous magnetic moment of the muon [19]. Further, well-motivated candidates are so called axions and their generalization, axion-like particles, because they primarily explain the well-known strong CP-problem and work as dark matter candidates in addition. Axions arise as pseudo-Goldstone bosons of a new symmetry and present quite a complex structure. Hence, the dark sector portal of axion-like particles is simply called the axion portal. A detailed description and phenomenology of axions and a broad range of other dark matter models can be found in [20].

3.2   Detection

In every model, dark matter somehow interacts with the Standard Model, albeit very weakly. However, this means that in principle dark matter should be detectable, if the sensitivity of the instruments used for the measurements is high enough.

3.2.1   Direct Detection and DAMA/LIBRA

Direct detection experiments seek to measure dark matter directly on earth by detecting the interaction of dark matter particles with scintillator material. Because of its revolution about the sun, the earth has a substantial velocity (about 30km per second) relative to the galactic disc with a periodicity of exactly one year. Around December 2nd, the earth moves exactly parallel to the rotation of the galaxy, while on June 2nd it moves against it. The dark matter halo must rotate, too, in order to stabilize against gravity. Due to the thermal equilibrium that the DM and the SM once shared, the sense of rotation must also be the same. Therefore, the flux of dark matter particles through earth reaches a maximum on

June 2nd and modulates yearly. If any interaction between DM and a detector happens regularly, its frequency must also modulate in the same manner and present a sinusoidal curve. This is the working principle of direct detection experiments.

The DAMA/LIBRA experiment and its predecessor, DAMA/NaI, both located at the Gran Sasso laboratory in Italy, have reported exactly the annual modulation of the signal expected from a WIMP for over 20 years, claiming a confidence limit of the combined exposures of 13.7σ [21]. DAMA uses an array of Thallium-doped sodium-iodide (NaI(Tl)) crystals that emit photons when subjected to ionising radiation. However, more sensitive experiments, such as CDMS and SuperCDMS, which rely not on NaI but rather cryogenic Germanium and Silicon crystals, have only reported results compatible with the null hypothesis [22]. While there have been some attempts to reconcile the results of DAMA with CDMS (see section 3.3), the general consensus among physicists is that the positive results is likely due to a wrongful background modelling, a suspicion enhanced by the non-publication of raw detector data by the DAMA collaboration. In particular, scientists at INFN in Pisa have pointed out that the subtraction of a yearly-averaged background rate from the measured event rates can produce sawtooth-shaped residuals that are compatible with the periodicities observed by DAMA [23].

ANAIS and COSINE intentionally employ NaI-crystals similar to the ones used at DAMA and have so far not found any evidence for dark matter-induced signal, though their sensitivity is not yet high enough to probe DAMA results [24].

3.2.2 Particle Colliders

Particle colliders also offer promising conditions for the detection of dark matter. Especially in the region where DM-nucleon scattering is suppressed, they promise greater discovery potential than direct detection experiments. While high-energy machines such as LHC have the potential to directly probe regions of parameter space otherwise unreachable, oftentimes the clean environment of low-energy but high-intensity electron-positron colliders is better suited for searches up to a few GeV. Even if the mediator masses are larger than the collision energy of a collider, off-shell production is still available, altering the cross sections of decays into Standard Model particles.

This section focuses on the search for dark photons, but similar strategies are employed at particle colliders for axion-like particles and other dark sector models, as well. In principle, the search strategy for any new intermediate state is different depending on its lifetime. A short-lived dark photon can be detected through the process $e^+e^- \rightarrow \gamma A' \rightarrow l^+l^-$ by measuring that the spectrum of measured lepton pairs is altered with respect to the SM. Searches like this have been performed at electron-positron colliders at BaBar and Belle, only yielding null results [25]. However, if the dark state is either so long-lived that it leaves the detector before decaying or if it only decays to untraceable final states, this search strategy is obviously unavailable. However, hermetic detectors, i.e. detectors that cover approximately the full solid angle around the interaction point, can still measure these decays indirectly by detecting missing energy, provided that there is a mechanism that tells the detector that an event has taken place - a trigger. In most cases, if an event involves a new particle X, high-energy photons emitted as initial state radiation (ISR) in the process $e^+e^- \rightarrow \gamma X$ (usually denoted in literature as $e^+e^-(\gamma) \rightarrow X$) can act as efficient triggers.

Since the process $e^+e^- \to \gamma$ without any X is kinematically forbidden, finding only one high-energy photon indicates that at least one further invisible particle was involved in the interaction. While not immediately proof for physics beyond the Standard Model (there are some rare SM processes with invisible neutrinos, for example), single (mono-) photon searches are a standard tool for many dark matter analyses. By employing mono-photon searches according to these principle, events of the form $e^+e^-(\gamma) \to A' \to$ invisible can be detected, too. The BaBar collaboration reported no signal consistent with this type of process [26].

The present thesis focuses on a third option sometimes referred to as the semi-visible signature. It happens when there is an intermediate dark matter state with a substantial lifetime that still deposits some energy in the detector. Charged tracks show up in conjunction with large amounts of missing energy and the center-of-mass energy is split up into the invisible and visible signatures. For very short lifetimes, i.e. prompt decays, the visible signature is recovered as the detector should be able to efficiently trigger and reconstruct the charged tracks. Conversely, lifetimes that are sufficiently long to make the intermediate state decay only once outside the detector, only ISR can be used for triggering, recovering the mono-photon signature. Interestingly, as pointed out in [19] and concretely for the case of Belle II in [27], there is a substantial range of displacements where the mono-photon search becomes inefficient due to tracks being found in the CDC alongside the ISR-photons. Thus, there is a blind spot of detectors towards these semi-visible signatures.

3.3 Inelastic Dark Matter

How exactly such a semi-visible signature may be realized in nature and why it makes for a successful dark-matter model is the subject of the next few sections. Namely, semi-visible decay signatures are a characteristic feature of so called inelastic dark matter models. They are a very interesting alternative to simple one-particle dark matter and appealing because of their theoretic simplicity, as they represent a minimal extension of the SM into a dark sector.

3.3.1 A Possible Solution to the DAMA Problem

Inelastic dark matter (iDM) was originally proposed as an attempt to reconcile the tensions between direct detection results from DAMA and CDMS. The WIMP hypothesis postulates equal nucleon-dark matter cross sections for all nucleons, but this picture changes dramatically when inelasticity is introduced: Rather than one state, two degenerate states (from now on denoted as χ_1 and χ_2) of slightly different mass are introduced. In this scenario, scattering off of nucleons can only happen when χ_1, the dark matter particle, transitions to its heavier inelastic partner, requiring some of the kinetic energy. The scattering process $\chi_1 N \to \chi_2 N$ is kinematically suppressed, since the end-products are heavier than the initial state. The exact constraint posed by this can be expressed in terms of the mass splitting δ:

$$\delta < \frac{\beta^2 m_\chi m_N}{2 \left(m_\chi + m_N \right)}, \tag{3.3.1}$$

where m_χ and m_N are the rest masses of dark matter and the nucleon acting as scattering center, respectively. Elastic scattering is even worse, as it is loop-suppressed due to the additional necessary transition to χ_2. Equation 3.3.1 shows that the constraint posed by δ is more severe the lighter the nucleon is. And since the scattering centers at DAMA with an atomic number of 127 are significantly heavier than the ones at CDMS (Germanium with atomic number 73) this opens the possibility of detection at only one site. It is shown in [28] that assuming $m_\chi = 100\text{GeV}$ and a typical DM velocity of 220km/s, a mass splitting of 13keV would be sufficiently small to make detection possible only at DAMA. Smaller DM masses or a larger splitting between the states would render any direct detection impossible.

3.3.2 Cosmology of iDM

Due to its relative simplicity, inelastic dark matter has become a compelling contender as likely dark matter model in its own right, not merely as explanation for the DAMA-CDMS-conundrum. iDM works by introducing a new interaction mediated by a new boson, the dark photon (A'), which can also decay into Standard Model particles, constituting the vector portal to the dark sector. Since the dark sector particles carry no SM charges, decays through the common vector bosons is impossible. The only allowed vertex is the 3-point vertex between a dark photon and the twin dark matter states, χ_1 and χ_2, shown in figure 3.1.
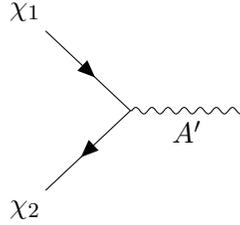
Figure 3.1: The only allowed 3-point vertex of inelastic dark matter. By absorbing a dark photon A', dark matter χ_1 transitions to the inelastic partner χ_2.

Vertices that feature two χ of the same type are, depending on the specific model, either strongly suppressed or completely forbidden. This means that a χ_2 can spontaneously decay to a χ_1 (provided that the mass splitting is large enough to have particles it can decay to), while the χ_1 is stable. The only way for a χ_1 to scatter is by transitioning to its heavier partner. The 2 lowest-order Feynman diagrams for DM-DM-scattering are shown in figure 3.2. Both process are kinematically suppressed, as either two dark photons or two χ_2 need to be produced from lighter initial states. Due to the mass splitting there is an energy boundary for interactions of χ_1 both with the Standard Model and with itself. Thus, if a cloud of χ_1 becomes too cold, the mean relative energy between particles is no longer sufficient to produce any inelastic partners and keep the cloud in thermal equilibrium with the SM and with itself. Once frozen out of equilibrium, the χ_1 becomes idle and interact only gravitationally - it is dark matter. The lack of DM self-interaction is also what keeps galactic dark matter halos from dissipating energy and collapsing onto a disk like baryonic matter.
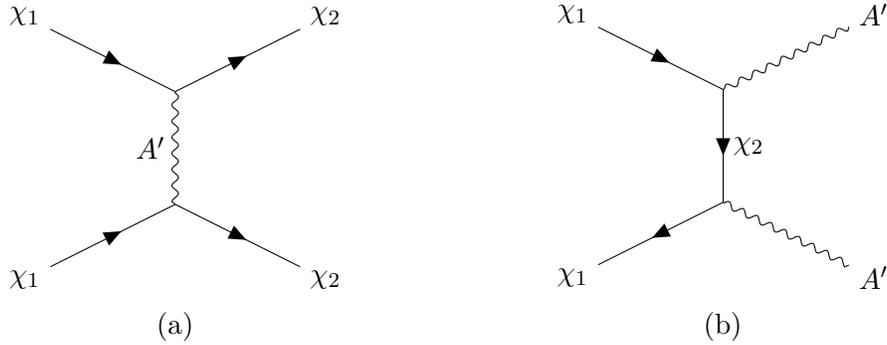
Figure 3.2: The two lowest order DM-DM scattering processes, via the exchange of a dark photon (a) or a χ_2 (b). Both cases are kinematically suppressed, as the final state is heavier than the initial state. Large dark photon masses or large mass splittings increase the suppression.

3.3.3 Theory

In principle, there is multiple ways of constructing a model that exhibits inelastic couplings, but the most natural one considers dark matter fermions[1] charged under a new gauge group most frequently denoted as X. While there is also the possibility of scalar inelastic dark matter, as described in [28], the more natural scenario is fermionic. The Weyl spinors of dark matter initially are symmetric under the new $U(1)_X$ gauge transformations but carry no other SM charge, making them inert to the common interactions. Inelastic dark matter also includes a Higgs-like scalar field ϕ which, like its SM counterpart, spontaneously breaks the $U(1)_X$ symmetry. As a consequence, the most general possible Lagrangian features both a Dirac mass term that couples the left-handed spinor to its right-handed counterpart and one Majorana mass term for each spinor generated by the coupling to the scalar (from now on referred to as the dark Higgs). Summarising,

$$\mathcal{L} \subset m_D \bar{\psi}\psi + \lambda_1 \phi \bar{\psi}_L^c \psi_L + \lambda_2 \phi \bar{\psi}_R^c \psi_R \tag{3.3.2}$$

are the arising mass terms for both spinors. The superscript c denotes the charge conjugated spinor, which is needed in order to construct Majorana mass terms. Here, m_D is the Dirac mass and $\lambda_{1/2}$ is the coupling of the dark Higgs field to the left-handed and the right-handed spinor, respectively. Since the gauge symmetry should be preserved in the limit of vanishing Majorana masses, it is natural for the symmetry-breaking terms to be small compared to the symmetry preserving one. Due to the off-diagonal mass terms, the mass eigenstates realized in nature are not aligned with the original Weyl spinors. Instead, the mass eigenstates must be calculated by constructing a mass matrix, defined by

$$\mathcal{L} \subset -\frac{1}{2} \begin{pmatrix} \bar{\psi}_L^c & \bar{\psi}_R \end{pmatrix} \begin{pmatrix} m_L & m_D \\ m_D & m_R \end{pmatrix} \begin{pmatrix} \psi_L \\ \psi_R^c \end{pmatrix}, \tag{3.3.3}$$

---

[1]Note that in all featured Feynman-diagrams χ shows up as arrows, indicating the fermionic nature of DM.

and finding its eigenvalues and vectors.[2] The mass eigenstates of inelastic dark matter are called χ_1 and χ_2 and are related to the Weyl spinors by:

$$\psi_L = \cos\theta\chi_{1,L} + i\sin\theta\chi_{2,L}, \tag{3.3.4}$$

$$\psi_R = \sin\theta\chi_{1,R} + i\cos\theta\chi_{2,R}, \tag{3.3.5}$$

yielding the mass eigenvalues

$$m_{1,2} = \sqrt{m_D^2 + (m_L - m_R)^2/4} \pm (m_L + m_R)/2. \tag{3.3.6}$$

Here, θ is the mass mixing angle. Crucially, the two stable eigenstates of the theory have a mass splitting between them, the magnitude of which is determined by the Higgs couplings. In the limit of vanishing Majorana masses, the mass eigenvalues are identical to the Dirac mass and the stable eigenstates are simply $\chi_1 = \chi_L$ and $\chi_2 = \chi_R$. A non-mixing theory is recovered and the gauge symmetry is preserved. When constructing the dark matter current, the result can be cast into 2 terms: The off-diagonal term connecting χ_1 to χ_2 and vice versa, labelled as the inelastic current $\mathcal{J}^\mu_{\text{iDM}}$, and diagonal couplings χ_1 to χ_1 and χ_2 to χ_2, labelled as the elastic current $\mathcal{J}^\mu_{\text{elastic}}$. In total, the dark matter current exhibits the following proportionalities:

$$\mathcal{J}^\mu =\propto m_D\mathcal{J}^\mu_{\text{iDM}}+ \propto (m_L - m_R)\mathcal{J}^\mu_{\text{elastic}} \tag{3.3.7}$$

[29]. In the natural scenario of $m_D > m_{L/R}$ and $m_L \sim m_R$, the diagonal coupling is suppressed and the dominant scattering process includes state transitions from χ_1 to χ_2. For simplicity, the limiting case of equal Majorana masses is assumed where the diagonal coupling vanishes completely.

Meanwhile, by introducing a new $U(1)_X$ charge, one new boson is introduced (called the dark photon) that must somehow be linked to the SM. This is done by use of the kinetic mixing portal (or dark photon portal) that arises naturally when writing down the most general kinetic term for the joint DM and SM Lagrangian. Only the relevant terms are shown in the following equation:

$$\mathcal{L} \subset -\frac{1}{4}F'_{\mu\nu}F'^{\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} - \frac{\epsilon_Y}{2}F'_{\mu\nu}B^{\mu\nu} - \frac{m_{A'}^2}{2}A'_\mu A'^\mu. \tag{3.3.8}$$

$B_{\mu\nu}$ and $F'_{\mu\nu}$ are the usual field strength tensors of the hypercharge boson and the dark photon, respectively. The third term is a theoretical necessity that introduces kinetic mixing proportional to the mixing parameter ϵ. Fine-tuning it to 0 would separate the fields completely but ultimately raise questions about the naturalness of such a scenario. After electroweak symmetry breaking:

$$\frac{\epsilon_Y}{2}F'_{\mu\nu}B_{\mu\nu} \to \frac{\epsilon_Y}{2}F'_{\mu\nu}\left(\cos\theta_w F_{\mu\nu} - \sin\theta_w Z_{\mu\nu}\right), \tag{3.3.9}$$

with the familiar photon and Z-boson fields, $F_{\mu\nu}$ and $Z_{\mu\nu}$. Interestingly, the electroweak mixing angle (Weinberg angle) θ_w shows up in this equation, as the dark photon must

---

[2]This is nothing new to the Standard Model: In the Weinberg-Salam theory of electroweak unification, the gauge bosons of the weak interaction (usually denoted as W^1 and W^3) are not simultaneously mass eigenstates. The familiar $W^\pm$-bosons are linear combinations of the two, related by a rotation matrix.

necessarily mix with both the photon and the Z-boson. Lastly, the coupling strength of the dark photon to SM currents is given by

$$g_{A'f\bar{f}} \approx -\epsilon_Y \frac{m_Z^2 \cos\theta_w e Q_f - m_{A'}^2 g Y_f}{m_Z^2 - m_{A'}^2}, \tag{3.3.10}$$

where Q_f is the electric and Y_f the weak hypercharge of a fermion f. It can be extracted from equation 3.3.10 that even for electrically neutral particles, a coupling to the dark photon exists, with a coupling strength proportional to the weak hypercharge. For example, a (left-handed) neutrino - while electrically neutral - could still interact with the dark photon as it carries weak hypercharge $Y = -1$. However, in the relevant region of iDM-parameter space, the Z-mass is significantly larger than the dark photon mass, and the hypercharge-coupling thus strongly suppressed. Through the dominating term proportional to $\epsilon \equiv \epsilon_Y \cos\theta_w$ the SM acquires a millicharge under $\mathrm{U(1)_X}$. It is this very weak coupling that gives the χ_2 its characteristically long lifetime.

Inelastic dark matter has 5 free parameters: The mass of the lighter DM particle, m_{χ_1}, the mass splitting between the states, Δ, the dark photon mass, m_A, the DM coupling strength to the dark photon, α_D and the SM coupling strength to the dark photon, determined by the dark mixing angle, ϵ_Y. Though some parameters can be constrained on cosmological grounds and by limits mostly set by collider experiments [27], the parameter space of iDM is vast. It is worth mentioning that due to the involvement of a new scalar field that provides DM fermions with Majorana mass, dark Higgs decays to SM particles are in principle possible, as well. The kinetic mixing of this dark Higgs with the SM Higgs is a naturally arising consequence as well, opening the Higgs portal to the dark sector, too. Due to the number of parameters involved, the phenomenology of this two-portal scenario is very complex.

3.3.4   Inelastic Dark Matter at Belle II
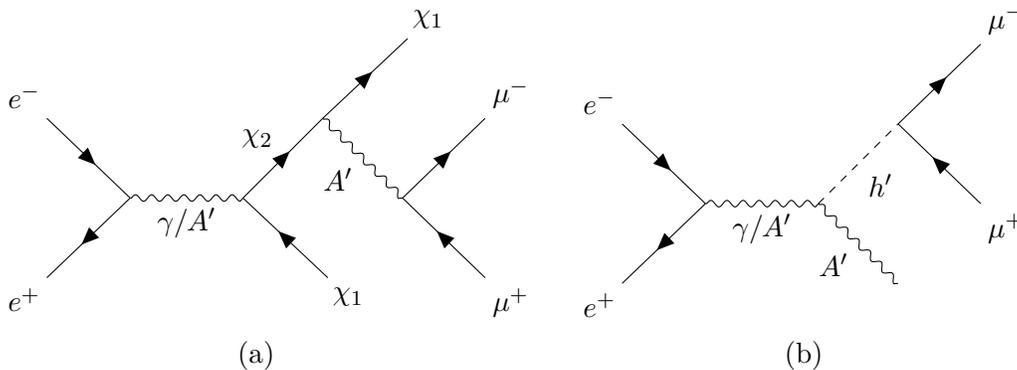
Figure 3.3: Feynman diagrams corresponding to the relevant inelastic dark matter decays (a) and dark Higgs decays (b) at Belle II. A final state involving a muon-antimuon pair is shown as an example.

The signature of inelastic dark matter events at Belle II can be very different depending

on the combination of model parameters. The most interesting case for the present thesis features small mass splittings between the dark sector states which lead to macroscopically displaced decay vertices. However, if the mass hierarchy $m(A') > m(h')$ is realized, also the dark Higgs boson is long lived, potentially creating displaced vertices on its own, too. The two scenarios are related and rely on the kinetic transition of the regular photon produced by the initial e^+e^--annihilation to the dark photon. The processes that can follow the transition are shown in figure 3.3.

Both events feature displaced decay vertices of an invisible state to a muon-antimuon pair. The reconstruction of the vertex, however, is more challenging in the inelastic case. The reason is that the dark Higgs decay is a two-body decay without loss of energy. The resulting topology has radial symmetry as the combined momentum of the muon pair points exactly in the same direction of the dark Higgs momentum. In inelastic decays, on the other hand, track momenta do not add up to the parent particle's momentum, as a fraction of it is carried away by the invisible χ_1. Vertices do not point back to the primary interaction point, spoiling the radial symmetry. The semi-visible signature of displaced vertices is also shared by other dark matter models, for example in [30]. In summary, a great deal of natural scenarios can be thought of that produce displaced vertices and missing energy in ranges fully available to current-generation colliders such as Belle II and almost undisturbed by Standard Model backgrounds. The next chapter offers detailed insights to the Belle II experiment, explaining concretely how these displaced signatures fall into a blind spot of the detector.

Part II

Belle II

Chapter 4

The Belle II Experiment

The Belle II experiment in Tsukuba, Japan offers promising conditions for the discovery of dark sector particles. It is the successor of the Belle experiment which was essential in discovering and quantifying CP-violations in B-meson-systems. Though both Belle and Belle II were built to produce as many B-mesons as possible, they are by no means the only particles that can be studied at the facility. Other fields of study include precision QCD measurements or searches for new physics in τ decays.
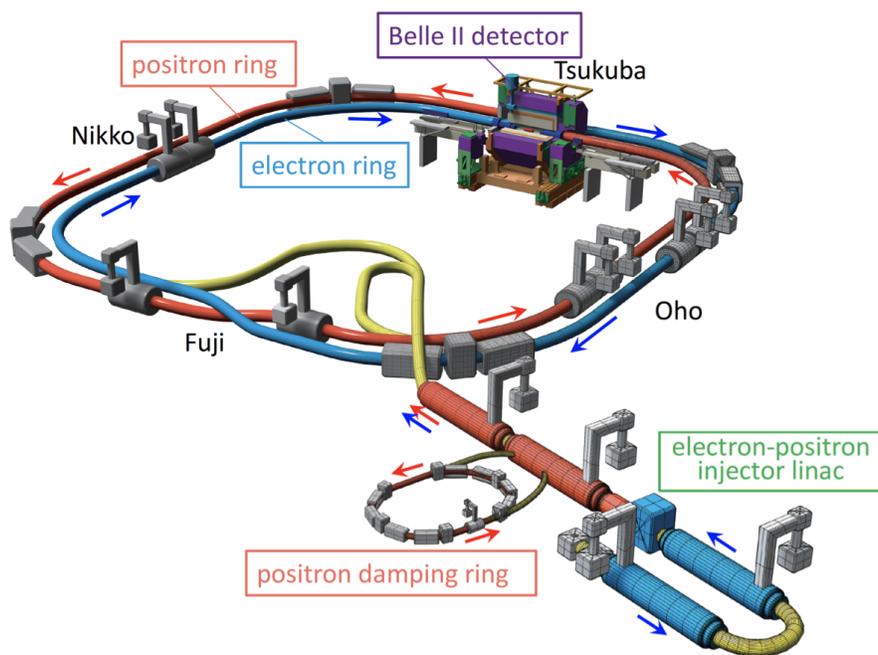
Figure 4.1: Schematic view of SuperKEKB.

4.1   SuperKEKB

SuperKEKB is the particle accelerator that feeds the Belle II experiment. During the upgrade of the detector from Belle to Belle II, it was also improved and partly replaced in order to achieve a higher luminosity. SuperKEKB is an asymmetric-energy electron-positron collider, meaning that the two beams have very different energies: Electrons are accelerated to 7GeV and positrons to 4GeV. Combined, the total collision energy is fine-tuned to the mass of the $\Upsilon(4S)$-resonance at 10.58GeV, which predominantly decays into B-mesons. Hence, accelerators like SuperKEKB are often called B-factories. The electrons and positrons are accelerated in bunches using the same linear accelerator (linac) and injected into two separate 3km-long storage rings. After half a turn, the beams cross paths at the so called *interaction point* (IP) at an angle, such that the particles can be separated and reinserted into the storage rings if no collision occurs.[1]. The Belle II detector is built around the IP and configured to efficiently capture the decay products of electron-positron-annihilations. A schematic view of the accelerator is shown in figure 4.1.

4.1.1   Luminosity

The (instantaneous) luminosity of a collider is a measure for the rate at which reactions happen. It has the dimension of a count rate per cross section (area). So, the typical luminosity of $10^{35}\text{cm}^{-2}\text{s}^{-1} \hateq 100(\text{nb})^{-1}\text{s}^{-1}$ means that on average, a hundred collisions with a cross section of 1nb (approximately the cross section of the important process $e^+e^- \to \Upsilon(4S) \to B\bar{B}$) are expected every second. In order to improve the precision of measurements of Standard Model parameters and to increase the sensitivity of the detector to very rare events after the upgrade, the accelerator was planned to reach a peak instantaneous luminosity of $6 \times 10^{35}\text{cm}^{-2}\text{s}^{-1}$ - 30 times higher than it's predecessor, KEKB. Integrated over the entire planned run time of Belle II, the accelerator is scheduled to deliver a total *integrated* luminosity of $50ab^{-1}$ ([31], [32]). For comparison, the complete Belle data set contains $1ab^{-1}$ worth of collisions. In general, the luminosity of an electron-positron-collider can be calculated using:

$$\mathcal{L} = \frac{N_+ N_- f}{4\pi \sigma_x \sigma_y} R_L. \tag{4.1.1}$$

Here, N_+ and N_- are the numbers of positrons and electrons in a bunch, respectively, f denotes the frequency of collisions and the factor in the denominator, $4\pi\sigma_x\sigma_y$ is the cross sectional area of the interaction point. R_L is called the luminosity reduction factor and accounts for a vast array of smaller effects that affect the luminosity, for example the fact that there is an angle between the two beams rather than a perfect head-on collision. From equation 4.1.1 it can be seen that increasing the luminosity can be achieved by increasing the collision frequency and the number of particles in the bunches and by decreasing the interaction cross section.

---

[1]Most bunches pass through each other without any pair-annihilation happening

4.1.2 Changes for the Upgrade

Significant changes with respect to the old KEKB system were necessary to reach the target instantaneous luminosity. Most notably, the so called *nanobeam scheme* with a significantly narrower beam cross section was adopted. In [33], the authors state that from a practical viewpoint including like hardware feasibility and operating costs, it was preferable to decrease the beam cross section rather than dramatically increasing the currents. Nevertheless, almost all parameters were improved, including but not limited to

- the maximal number of bunches in the storage ring (2503),

- the number of particles per bunch (order 10^{10}),

- the longitudinal extension of each bunch (down to $5 - 6$mm),

- the cross section of the beam (approximately 10μm $\times$ 50nm) and

- the crossing angle (83mrad).

The point of a large crossing angle is an effective decrease of the longitudinal size of the overlap region between the two beams. This is shown schematically in figure 4.2. With the
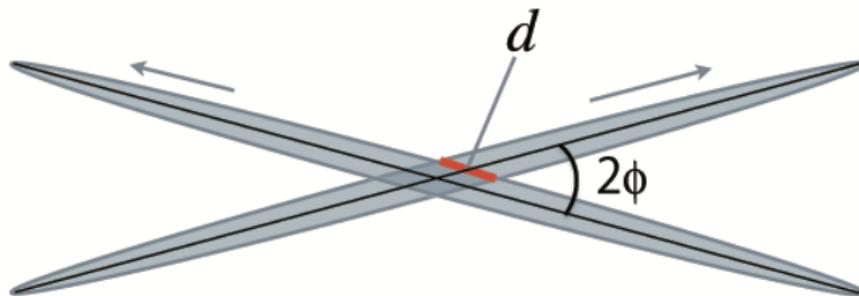
Figure 4.2: Schematic view of the beam crossing area in the nanobeam scheme. The effective intersection diameter d is smaller if the crossing angle ϕ is increased. Plot from [32].

increase in total particles in the storage rings at any given time, the beam currents were increased by a factor 2.2. For the positron beam, a damping ring (also shown in figure 4.1) was built in order to bundle the particles in a tighter volume of the transverse beam phase space (this volume is usually called emittance). Coupled with the significantly squeezed beam, this led to the desired increase in luminosity by a factor 30 [34].

However, because of the larger amount of charged particles squeezed inside a significantly smaller volume, unwanted beam-beam interactions and Touschek scattering rates at SuperKEKB are much higher than its predecessor. KEKB initially injected new particles into the storage rings every 70-80 minutes, but later on in its running pioneered the continuous injection mode, in which the beams are refilled by top-up injections from the linac many times per second. With a beam decay time of about 360sec at SuperKEKB, the only way to achieve the needed luminosity is by injecting new particles into the beams routinely

at 12.5-25Hz. Each pulse is accompanied by an *injection veto* which ignores all activities of the detector a few milliseconds after the injection. Due to the larger currents and frequent injections, SuperKEKB has to cope with a dramatic increase in beam background, which will be discussed in more detail in a later section.

In July 2022, the first data-taking period of Belle II was completed and SuperKEKB was shut down after setting a new world record for peak instant luminosity at $4.7 \times 10^{-34} \text{cm}^{-2} \text{s}^{-1}$. Currently, Belle II is in its long shutdown (LS1) phase, while general maintenance of all systems are be conducted and parts of the accelerator system as well as the detector are replaced and upgraded. The experiment is being prepared to reach its full design luminosity after data taking is resumed in early 2024. Taking advantage of LS1, the displaced vertex trigger presented in this thesis is planned to be implemented to hardware in order to be available during the next running period.

4.2   Detector

Belle II is a multi-layered hermetic detector built around the interaction point. A hermetic detector covers as large a solid angle as possible such that all particles that are created pass through the detector system. Of course, the entry points of the beams are not covered, leaving an acceptance region of 17° to 150° with respect to the beam axis. Furthermore, Belle II consists of many subdetectors that are specialized to capture a certain class of particles. Combining information from multiple subdetectors, Belle II can identify all final state particles, apart from the neutrinos. Final state particles are mostly photons, electrons, muons, pions and kaons.[2] As a virtue of the hermeticity of Belle II, even neutrinos can be indirectly detected if the calculated energies of all visible tracks do not sum up to the collision energy. A schematic view of Belle II is shown in figure 4.3.

The Belle II coordinate system which I adopt in this thesis is centered around the IP. The z-axis bisects the beams (pointing in the direction of travel of the electrons) and represents the symmetry axis of the detector, while x points towards the center of the accelerator ring (parallel to the ground) and y points upwards. Due to the symmetry, oftentimes a polar coordinate system is used. In that case, radial distances r are taken with respect to the z-axis and ϕ represents the azimuthal angle in the transverse plane, where $\phi = 0$ describes an axis parallel to the x-axis. Furthermore, tracks are typically parameterized in spherical coordinates. In that case, the definition of the azimuthal angle remains the same and an additional polar angle θ with respect to the z-axis is used rather than the z-coordinate itself.

Since the displaced vertex trigger proposed in the present thesis only uses data from one subdetector - the CDC - a special emphasis will be placed on that system while the other subdetectors will only be discussed briefly.

---

[2]Muons, pions and kaons are not intrinsically stable but at the boosts they typically receive during a collision their decay happens way outside of the detector.
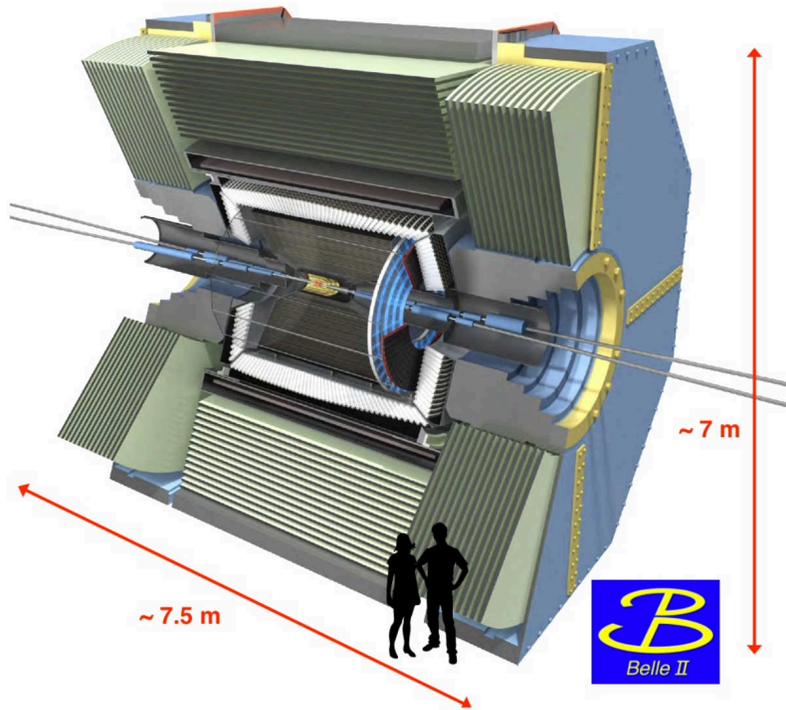
Figure 4.3: The Belle II detector.

4.3 Vertex Detector

During the upgrade, the width of the beam pipe was significantly reduced, allowing to move the innermost detector - the vertex detector (VXD) - even closer to the interaction point. It is a silicon detector sensitive to energetic charged particles and it is first and foremost used to determine the decay vertex of B-mesons to an accuracy of 50μm in order to perform one of the flagship measurements at Belle II: time-dependent CP-violation. The details of CP-violation measurements at Belle II can be found in the Physics Book [35]. Belle and Belle II alike use(d) a strip detector (SVD) with double sided silicon strips arranged in x and y direction. By combining the hit information from both sides, the two-dimensional hit position can be determined.

The small distance to the beam pipe is a double-edged sword as beam background rates scale roughly with the inverse square of the distance. Therefore, the VXD at Belle II is subjected to much higher background rates than its predecessor due to Touschek scattering and beam-beam interactions. These rates would be high enough to saturate the SVD completely. Therefore, the VXD was supplemented by an entirely new system based on single detector pixels (PXD) which directly contain the 2d-information of hits. With the fine granularity of pixels - each one measuring only 75μm $\times$ 50μm - the PXD can be placed 14mm away from the interaction point. The downside of this high-resolution system is that at the planned luminosities the data output of the entire map of pixels is too large to be saved let alone processed in real time for triggering. Instead, regions of interest are

37

extrapolated from track data provided by other detectors and used to improve resolution later on.

As discussed earlier, the signature of inelastic dark matter is a pair of leptons well displaced from the original interaction point. The tracks they create are in most models very far away from the interaction point. Therefore, the PXD does not serve a purpose in searching for iDM.

4.4 Central Drift Chamber

Surrounding the VXD lies the core tracking device of the Belle II detector - the central drift chamber (CDC). It ranges from a radial distance of 160mm to 1130mm to the beam axis and is filled with 14336 sense wires that fire when an energetic charged particle passes close to them. The entire CDC is submerged in a magnetic field of 1.5T directed along the z-axis which forces charged particles onto helical tracks. They appear as circles in the transversal plane, from which the curvature and sense of rotation can be measured, revealing information about the momentum and the charge of the particle: The radius of the circle is inversely proportional to the transverse component of the momentum and the sense of rotation indicates the charge (positive charges rotate clockwise).

4.4.1 Drift Cells

The CDC has a similar working principle as ordinary cloud chambers: It is filled with a gas mixture of 50% Helium and 50% Ethane (chapter 6.3 in [32]) that is ionized when fast charged particles pass through. On their paths through the CDC, they therefore leave a trail of free electrons, knocked out of their parent nuclei. In each so called *drift cell* 8 high-voltage field wires create an electric field that acts on these electrons which start to drift toward the center of the cell. There, a sense wire picks up on the charge the electrons deposit. The exact setup of the drift cells is shown in figure 4.4.

The gas mixture and the field strengths are chosen in such a way that the drift velocity of the electrons is, to good approximation, constant. Therefore, the time it takes an electron to reach the sense wire (i.e. the *drift time*) is proportional to its distance to it. This drift time is orders of magnitude larger than the typical time of flight of a particle through the CDC. Thus, the distance of closest approach of the track to the sense wire can be inferred from the drift time, narrowing down the exact position of the track from somewhere in the drift cell to just a circle with radius equal to the drift time multiplied by the drift velocity, which is precisely known.

Another crucial bit of information which will be discussed in greater detail in chapter 7.2 is the integrated charge deposited to a wire. This value is called the ADC (analogue-to-digital converter) count. Ionising particles roughly free up the same amount of electrons at a broad range of momenta, while various background sources can have very different ADC profiles. For example cross-talk between adjacent wires typically creates hits in a specific range of ADC values while synchrotron electrons, which stay close to one sense wire for a long time due to their lower energies (sometimes even circling around the same wire for the entire length of the CDC) leave a huge deposited charge. Therefore, the ADC count is a powerful tool in the separation of signal hits from random backgrounds.
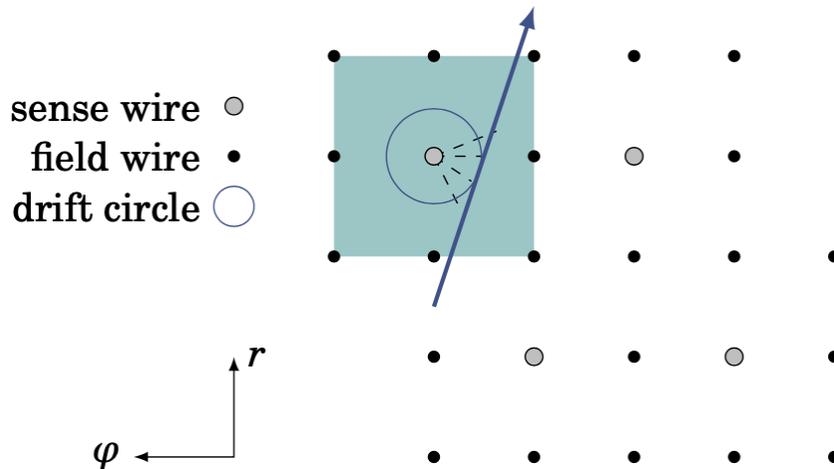
Figure 4.4: Structure of CDC wires. A particle creates a track of free electrons (blue arrow). In each drift cell (blue square) 8 field wires create a strong electric field which pushes the free electrons towards the central sense wire. This plot was taken from [36].

During its path through the CDC, a particle loses energy while ionizing gas molecules. The amount it loses depends on the mass and momentum and can be calculated with the Bethe-Bloch equation. Therefore, by analysing the track of a particle and, specifically, by measuring dE/dx, its mass and therefore particle type can be extrapolated. Even if a particle doesn't have enough momentum to reach the dedicated particle identification system (described in section 4.5), the CDC track is oftentimes enough to determine all the necessary parameters.

4.4.2 Wire Configuration

The 14336 sense wires are closely parallel to the z-axis and are arranged into 56 radial layers from 160mm to 1130mm. The 56 layers are further subdivided into 9 *superlayers* where every radial layer contains the same number of wires. The innermost superlayer is very close to the interaction point and therefore experiences much higher background rates than the other superlayers. In order to avoid an overly high occupancy in every event, the wire density must be increased. While all other superlayers contain 6 radial layers with an inter-wire distance of about 2cm, superlayer 1 has 8 layers and half the inter-wire distance. The wire configuration is summarized in table 4.1.

Another crucial point is that superlayers come in two types: *axial* and *stereo*, depending on the orientation of the wires within. If all wires of the CDC were parallel to each other, track parameters could only be determined in two dimensions, as the coordinate parallel to the wires would be degenerate. Therefore, a subset of wires, namely the wires inside stereo superlayers, are slightly skewed, allowing to complete the 3D picture. In the 5 axial superlayers all wires lie exactly parallel to the z-axis, allowing a precise determination of the transversal component of a track. Stereo wires are skewed between 2 and 5 degrees with respect to the z-axis, as figure 4.5 shows.

type and N.	N. of layers	wires/layer	radius [cm]	angle [mrad]
Axial 1	8	160	168.0-238.0	0
Stereo U 2	6	160	257.0-348.0	45.4-45.8
Axial 3	6	192	365.2-455.7	0
Stereo V 4	6	224	476.9-566.9	-(55.3-64.3)
Axial 5	6	256	584.1-674.1	0
Stereo U 6	6	228	695.3-785.3	63.1-70.0
Axial 7	6	320	802.5-892.5	0
Stereo V 8	6	352	913.7-1003.7	-(68.5-74.0)
Axial 9	6	384	1020.9-111.4	0

Table 4.1: Wire configurations for all superlayers. The skewing-direction of stereo super-layers alternates, hence the changing signs in the angle column.
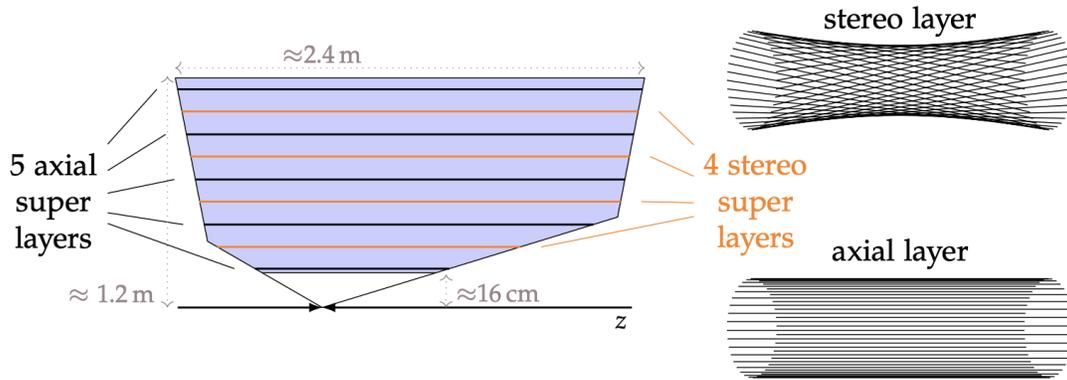
Figure 4.5: Wire configuration of the CDC. Schematically, one radial layer of an axial and a stereo layer are shown.

In the two-dimensional transversal plane, axial wires appear as points, as they have constant (r, ϕ) for all values of z. The transversal coordinates of stereo wires, on the other hand, are functions of z: $(r(z), \phi(z))$. A stereo wire projects to a straight line roughly 10cm long that starts where the wire attaches to the back cap and ends where it attaches to the front cap of the CDC. Each point in transversal plane is crossed by at least 5 stereo wires. *Which* stereo wires fired for any given 2D track parameters determines the polar angle of the 3D track. For clarity, event displays represent stereo wires not as lines, but rather as points. For this, the slice through the CDC at $z = 0$ is chosen as reference. Tracks that run exactly in the transversal plane appear as continuous circles and their stereo and axial hits align, while tracks at polar angles different from 90° appear to have offset stereo hits. The greater the offset the smaller is the angle between track and z-axis.

This point is illustrated in figure 4.6: Each dot represents a sense wire in the CDC that has fired during the event. In the left panel, axial hits are shown in red while stereo hits are blue. There are also background hits that do not belong to any (signal) particle, they are shown in grey. As the image shows, axial and stereo hits align very well on the left track (in the left panel) while the stereo hits of the other track zig-zag. This is simply due
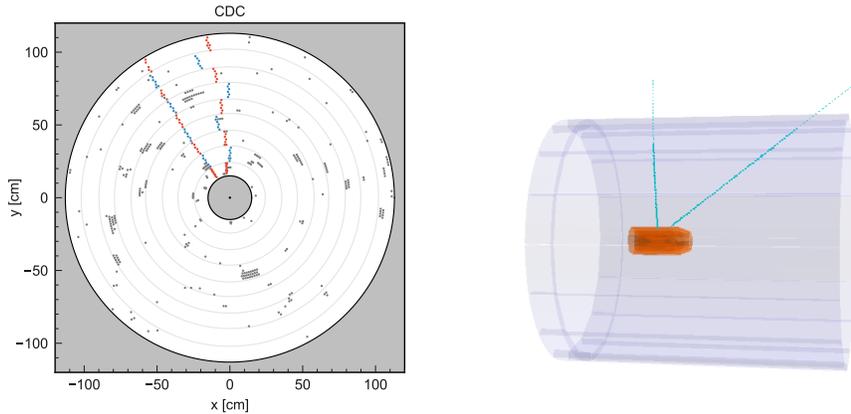
Figure 4.6: Display of the same event in the transversal plane (left) and in 3D (right). Stereo wires (blue, left plot) are in line for the track perpendicular to the z-axis but seem to zig-zag in the shallower track. For clarity, light background conditions has been chosen for this plot.

to the polar angle of the track, as the 3D rendering of the event shows: The left track lies almost perfectly in the transversal plane while the other one leaves the CDC at a shallow polar angle.

4.5 Particle Identification System

The next subdetector radially outwards is the particle identification system (PID). It is comprised of two different kinds of detectors that are situated just outside the forward caps and the barrel of the CDC. Both the two systems make use of the Cherenkov effect to identify the velocity of particles, but with different designs due to the different locations.

When travelling through a refractive medium (a medium with a refractive index greater than unity), particles can in principle move faster than light, provided they have enough momentum. The Cherenkov effect occurs if that medium also happens to be dielectric - meaning that it can be polarized - and the particle is charged. If this is the case, as the particle passes, polarization waves ripple through the medium with a phase velocity equal to the speed of light in the medium. In analogy to the sonic boom that happens if objects move faster than the speed of sound in the atmosphere, wavefronts trail behind the particle and compress to a Mach cone. Since polarization waves are nothing but periodic fluctuations of electric fields, these wavefronts are just electromagnetic radiation - light. The inside angle of the cone is

$$\cos\theta = \frac{1}{n\beta_p} \tag{4.5.1}$$

where n is the refractive index of the medium and $\beta_p = \frac{v_p}{c}$ is the the velocity of the particle. More details about the Cherenkov effect and detector applications thereof can be found in [37].

In the forward cap lies the aerogel ring-imaging Cherenkov detector (ARICH). It consists of a 2cm thick layer of an aerogel which serves as the dielectric medium. Charged particles that pass through the aerogel emit a cone of light at a fixed Cherenkov angle determined by equation 4.5.1. During a 20cm propagation phase, the cone is allowed to grow out radially and is eventually captured by photomultiplier tubes with a spatial resolution of order 1mm (chapter 8 in [32]). From the size of the ring and the exact position of incidence of the particle into the aerogel (extrapolated from CDC tracks), the Cherenkov angle is measured and thus the velocity of the particle is calculated. Comparing the velocity to the momentum, which is known from the track curvature in the CDC, the mass of the particle can be uniquely determined. Using the ARICH, Belle II can reliably distinguish charged kaons and pions, which otherwise behave very similarly.
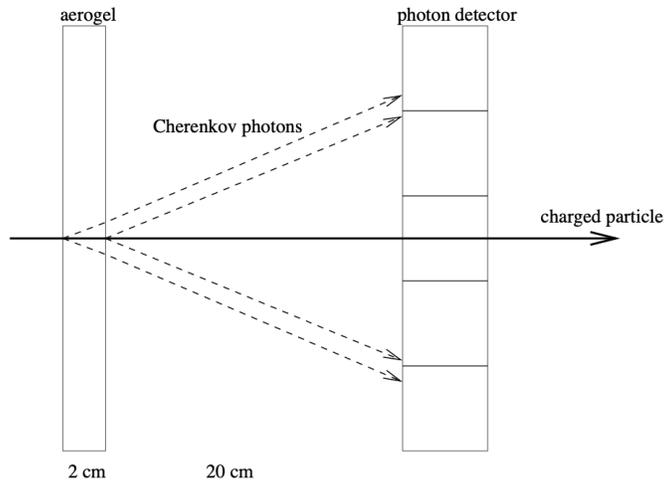
Figure 4.7: Working principle of the ARICH. Depending on their velocity, charged particles create Cherenkov cones of light with a specific inner angle while passing through the aerogel. The cone of light is captured by the photon detector 20cm later, where the size of the resulting ring is measured. Image from the technical design report [32].

The other constituent of the PID is the time of propagation detector (TOP), located outside the barrel of the CDC. While the exact design is different from ARICH, the principle of operation is the same.

4.6 Electromagnetic Calorimeter

Up to this point, all presented subdetectors were only capable of sensing charged particles - neutral particles escape detection by the PXD, the CDC and the PID and are also unaffected by the magnetic field. The electromagnetic calorimeter (ECL), on the other hand, can also detect neutrals, which is why it is of great importance - especially in the context of dark matter detection.

It consists of 8736 cuboid CsI(Tl) (thallium-doped caesium-iodide) scintillator crystals with an average size of $6 \times 6 \text{cm}^2$ and a length of 30cm, arranged in such a way that all

crystals point to the interaction region. It wraps around the complete CDC and thus has a larger polar acceptance region of $12.4° < \theta < 144.1°$ (apart from two $\sim 1°$-holes where the caps meet the barrel). Photomultiplier tubes are attached to the back side of each crystal, detecting scintillation photons. More specifics about the ECL design can be found in chapter 9 of the design report [32].

Gamma rays can in principle interact with matter through 3 different processes: The photoelectric effect, Compton scattering and pair production. What all of these processes have in common is that the energy of the initial photon is converted to other particles, which in turn interact electromagnetically. In the case of pair production, the incident photon has enough energy to create an electron-positron pair. The two particles then interact with the detector material, radiating off more photons in the process. Eventually, the positron annihilates with an electron from the material and creates a new, lower energy photon. In total, the initial photon creates a shower of photons that are captured by the photomultipliers in the back of the crystal. Since the cross sections of the 3 interaction processes increase for heavier nuclei (the photoelectric effect even scales with Z^5) high-Z and dense scintillator materials are especially suited (like CsI) for gamma ray detection.

In order to cope with the high luminosity-induced background rates and to better discriminate particle types from one another, a novel pulse shape discrimination algorithm was implemented in the Belle II ECL [38]. By analysing their shape and size, hadronic showers caused by neutral pions and kaons can be distinguished from gamma ray showers, and the energies of the incident particles can be reconstructed, too. Furthermore, the ECL also detects clusters from electrons, which are fully absorbed by the crystals, while heavier charged particles such as muons pass through the detector depositing only a small fraction of their energy and thus do not generate enough activity over the standard electronic noise to be detected. The ECL reaches an energy resolution of $\sigma_E/E = 1.6\% - 4\%$ and an angular resolution of $3\text{mrad} - 13\text{mrad}$ depending on the photon energy.

4.7   Kaon and Muon Detector

The last subdetector down the line is the long-lived kaon (K_L^0) and muon detector (KLM). It is built out of 4.7cm-thick iron plates sandwiched between active detector elements. 14 of these layers are arranged vertically (in the xy-plane) at the end plates and 15 in radial layers outside the barrel. The iron plates provide 3.9 interaction lengths worth of material for hadronic showers to form and simultaneously serve as a flux return yoke for the magnetic field. As the name suggests, its task is to detect neutral kaons and muons. The former produce hadronic showers that can be matched to a shower in the PID system, while muons produce tracks in the detector elements consistent with tracks in the CDC. Both neutral and charged pions usually do not reach the KLM[3], so they can be distinguished from muons and long-lived kaons, respectively.

---

[3]In some rare cases, charged pions do reach the KLM before decaying and do not interact strongly. In this case the pion is wrongly identified as a muon. However, the muon fake-rate is, depending on the energy, between 1 and 4% (Section 10.3.1 in [32])

Chapter 5

Triggering

SuperKEKB constantly supplies the detector with electron-positron-reactions. However, not all reactions are of physical interest. In fact, Belle II is dominated by QED-events that are very well understood and need not be analyzed for physics. By far the most frequent event that follows the e^-e^+-collision is Bhabha scattering, where after annihilation, the virtual photon promptly decays back into an electron and a positron. Often, it is accompanied by initial or final-state radiated photons, in which case it is called radiative Bhabha scattering. Table 5.1 summarizes the most frequent events at Belle II.

process	cross section [nb]	rate [Hz]
$e^+e^- \to \tau^+\tau^-$	0.9	540
$e^+e^- \to \Upsilon(4S) \to B\bar{B}$	1.1	660
$e^+e^- \to \mu^+\mu^-$	1.1	660
$e^+e^- \to q\bar{q}$	3.4	2040
$e^+e^- \to e^+e^-$ (Bhabha)	44.0	26400

Table 5.1: A selection of the most frequent events at Belle II with the rates expected at full luminosity. The cross section and reaction rate of Bhabha scattering includes only events where both final state particles are within the CDC acceptance region.

Evidently, with no selection-system in place, the Belle II data set would be flooded with unnecessary Bhabha scatterings the frequency of which scales with luminosity. Bhabha scattering is thus regarded as a luminosity-induced background. In addition, more backgrounds of other origin exist: Rogue high-momentum particle tracks are often caused by beam-beam interactions without annihilation or by the beam colliding with left-over gas molecules inside the beam pipe. These are called beam backgrounds. Lastly, random hits caused by wire cross-talk or by electronic readout noise can create fake tracks that occur in absence of any particle. In order to only save interesting events to tape, Belle II employs a host of *triggering* algorithms that perform a rough analysis of the event and react to physically interesting signatures. Later, during *reconstruction* the tracks and clusters are evaluated in depth and the best-fit trajectories and particle species computed. A good trigger system must have a high efficiency, i.e. detect as many signal events as possible, while maintaining a high rejection of background, i.e. not pollute the data sample with

uninteresting events.

The *online* processing (while the detector is running) of subdetector data is done on Field Programmable Gate Arrays (FPGAs). FPGAs are the middle-ground between general purpose processors and application specific integrated circuits (ASICs), which are extremely fast but can only perform a predefined set of tasks. FPGAs can be reprogrammed and customized after production and therefore offer a great platform for online processing and triggering. If, for example, trigger definitions change or new bits are introduced, they can be quickly implemented without the need of manufacturing a new ASIC board. Until long shutdown 1, Belle II trigger components were implemented onto the so called universal trigger board 3 (UT3), featuring a virtex-6 FPGA. During LS1, they will be upgraded to UT4 (either a Xilinx XCVU160 or a Xilinx XCVU80 FPGA) with significantly increased bandwidth. This will allow for more complex preprocessing algorithms to be implemented.

The main issue for displaced tracks from inelastic dark matter is that - of course depending on the length scale of the displacements - they are triggered with a very low efficiency [39]. It could therefore be that dark matter decays do happen at Belle II but the detector is simply blind to them. In the following subsections the current main trigger systems are discussed, with a special emphasis on their sensitivity to dark matter models.

5.1   The Track Segment Finder

In order to allow quick online processing, data from the CDC must be cleaned and compressed very quickly and efficiently, in light of the fact that background rates are expected to be very high. Fortunately, most background hits look very different from signal: Particles leave continuous hits along their path whereas random noise background hits have no spatial correlation and other unwanted processes such as cross-talking wires create clumps of hits with no preferred direction. Discarding such uncorrelated hits is the job of the track segment (TS) finder. One superlayer at a time, the geometrical arrangement of hits is compared to what is expected from a track. Track segments are defined as hourglass-shaped wire configurations (see figure 5.1) and a simple rule is imposed: If 4 different layers within the track segment are hit, the track segment is valid. Subsequently, the position of the wire with the smallest drift time (i.e. the wire closest to the physical track) is written to tape as a representation of the TS. If a problem is encountered during the determination of drift times, the so called priority wire is used. Due to hardware constraints, the outer layer in each except for the first superlayer is not analysed. Similarly, in the first superlayer, the inner 3 layers have very high occupancy due to their proximity to the IP and are therefore not used for triggering. The missing layers are included in offline reconstruction in order to increase the fit precision. If the hits within a superlayer do not fulfill the track segment requirement, they are discarded and not used in the subsequent tracking.

While the TS-finder is very efficient for tracks that originate at $(x, y) = (0, 0)$, the hourglass shape has significant drawbacks when it comes to detecting shallow tracks that cross few superlayers. In the context of dark matter, many signal tracks are expected to be inconsistent with the vertex hypothesis of the IP. Tracks often cross TS at an angle close to 90 degrees and may sometimes even curl back into the CDC. The hourglass-TSs are not designed for such tracks. Therefore, new TSs are needed. One proposal for such TSs is used for the DVT algorithm and discussed in great detail in section 7.1.
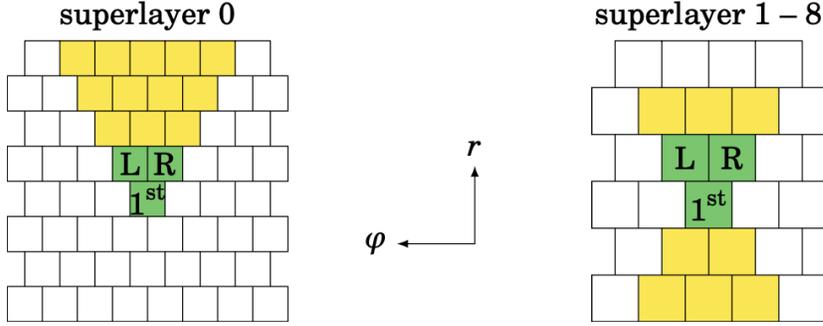
Figure 5.1: Shape of a track segment in the innermost superlayer (left) and in all other superlayers (right). Each cell represents a drift cell in the CDC. The green cells are the so called *priority* wires, which serve as a reference wire to the track segment. Plot from [36].

5.2 The CDC Track-Finder and the Trigger Menu

SuperKEKB is built primarily to produce as many B-mesons as possible. As with all reactions that feature strongly interacting particles, such events are characterized by a large amount of particles. In fact, the average $e^+e^- \rightarrow B\bar{B}$ process contains 12 tracks. Complementary to that, other equally important fields of study at Belle II, like dark matter searches or τ-physics, focus on *low multiplicity* events, where only few tracks are present in the CDC. They are much more challenging to efficiently extract from the background.

After the picture of the CDC is cleaned up with the TS-finder, the axial hits are combined to 2D tracks using a Hough transform (discussed in depth in chapter 6). The Hough transform is also the centerpiece of the displaced vertex trigger. In short, it uses the IP as a vertex hypothesis and defines tracks that are compatible with the hits. At least 4 TS hits are needed to constitute a track. In this case, it is called a short track as opposed to full tracks that contain TS hits in all 5 axial superlayers, traversing the entirety of the CDC. Every track is parametrized by $\left(\phi, \frac{1}{r}\right)$, where ϕ is the azimuthal angle and $\frac{1}{r}$ is the track curvature (the reciprocal of the circle radius).

These simple trigger objects from the CDC as well as the ECL, KLM and TOP are combined using logic operators to define a so called *trigger bit*. For example, one trigger bit (denoted cdcecl2) requires that there be two 2D tracks in association with ECL clusters. Trigger bits can be switched on and off for any accelerator run and constitute the first-level (L1) *trigger menu*. For each event, the global decision logic (GDL) compares the trigger objects it received with the trigger menu and takes a yes-or-no decision to write out the event information. A special veto for Bhabha scattering can overrule trigger decisions if the topological properties of a typical Bhabha event are satisfied. The distinct signature of Bhabha scattering - that is, high energy tracks approximately back-to-back to each other - is exploited, while more complex topologies, like radiative Bhabha scattering are more challenging to veto. Because of hardware limitations, the GDL is designed to ultimately achieve a first-level trigger rate of 30kHz [40]. If a trigger bit is found to fire too often, it is prescaled by a factor X, meaning that only every Xth event is accepted by the GDL - at the expense of efficiency.

Therefore, first-level triggers are not unlimited in their trigger rate and generally aim for a high rejection rate in order not to waste resources on background. An example trigger bit suited for the search of displaced vertices is the two-track trigger - it requires two full tracks that do not fulfill the Bhabha veto conditions. However, full tracks are extremely unlikely for displacements that exceed the outer radius of the first superlayer (23.8cm). Furthermore, with increasing luminosity many fake and background tracks are found, leading to a high trigger rate for the two-track trigger. For these reasons, it was no longer active in recent runs.

5.3   Neuro Tracks and the Single Track Trigger

At Belle II, many physics analyses related to tau physics or the search for dark matter (among others) feature events with very low track multiplicities. Therefore, triggering must be efficient even if very little topological information is available. However, with little to no constraints on tracks, the system simultaneously becomes very susceptible to background tracks, too. Signal events that actually come from an e^+e^--annihilation must therefore be reliably distinguished from rogue background tracks that are caused by the beam running into obstacles along the beam pipe such as the quadrupole magnets that enclose the interaction region. A track that is emitted from somewhere along the beam axis but, crucially, *not* from the interaction point can be safely assumed to be background and needs to be discarded. Figure 5.2 shows the z-distribution of all tracks. Evidently, a substantial fraction of them is caused by machine background. Therefore, if the z-impact of a track is measured with enough precision, it can be used to discriminate background tracks.
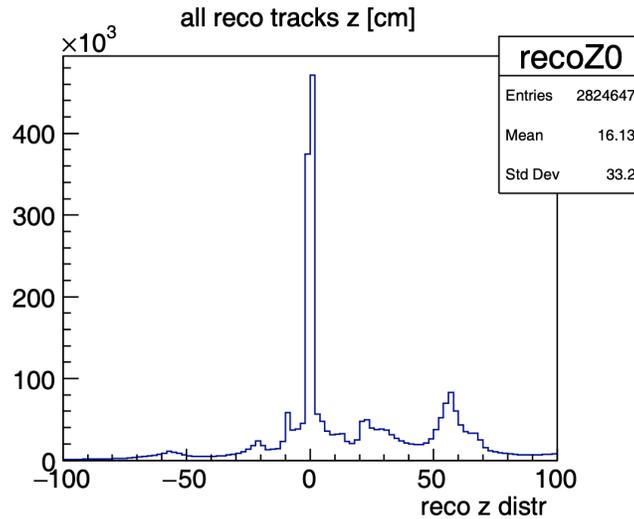
Figure 5.2: Distribution of the z-impact of reconstructed tracks from the last experiment before the long shutdown. No selection is applied. Only events with $z \approx 0$ are products of e^+e^--annihilations, the rest is beam background.

In [36], a neural network is introduced specifically for the task of estimating the z-impact and the polar angle of single tracks. Instead of relying on analytic calculations based on axial and stereo track segments, tracks found by the 2D track-finder are analyzed by a neural network and converted to 3-dimensional so called neuro tracks. Specifically, the input array contains the crossing angle, drift time and azimuthal wire position of all track segment hits - 27 numbers in total. For stereo wires, the azimuthal wire position is the distance between the position of the wire at $z = 0$ and the extrapolated intersection of the 2D track and the TS layer. It contains the information about the polar angle of the track: The larger the absolute value of this azimuthal distance, the larger the polar angle. The azimuthal distance is what causes the zig-zagging of stereo hits in figure 4.6. Axial wires, by construction, have an azimuthal distance of 0. It is only due to tracking errors if this number is not zero. Still, it is passed to the neural network because it can improve the fit quality of the track parameters. The inputs are passed to a single hidden layer of 81 nodes and subsequently to the output layer, which consists of two nodes, one for the z-impact, the other for θ. Combined with the initial 2D parameters, the complete 3D trajectory can be computed. If the resulting track is compatible with the IP (i.e. if $z < 15cm$), it is kept as a neuro track object. 2D triggers such as the two-track trigger described in the previous section are upgraded by requiring that one of its tracks actually be a neuro track (thus fulfilling the IP-requirement). This way, trigger bits that would otherwise exceed acceptable fake rates are kept from being prescaled or deactivated altogether.

A few years following the implementation of the neural track-finder, the single track trigger (STT) was introduced in [41]. The STT is a minimum-bias first-level track trigger, meaning that no prior assumption about the event topology or track type is needed for triggering - only a single track in the CDC must be present. With such little constraints, the STT is the best bet at triggering displaced tracks among the trigger bits currently active. Single neuro tracks are the basis of the STT, imposing a more stringent z-cut of 10cm and a further momentum cut of 0.7GeV in order to keep the trigger rate low. Lower-momentum tracks can only be triggered in combination with other tracks in the right kinematic conditions, such as opening angle. The STT has a very high efficiency and is constantly being maintained and retrained on new data, as background conditions frequently change between experiments and runs.

Despite its great success, the STT performs poorly on tracks from displaced vertices, see figure 5.3. This has two major reasons: For one, as already mentioned in section 5.1, the TS-finder which feeds the STT is not designed for shallow tracks and therefore records no hits that could be passed to the neural networks. Furthermore, even if the tracks are steep enough to be caught by the TS-finder, the inner layers are often missing. If the radial displacement surpasses the 40cm-mark - approximately the outer edge of the second axial superlayer - the tracks only have 3 axial superlayers left before exiting the CDC.[1] But since the 2D track-finder requires axial hits in at least 4 layers, the STT does not fire. The efficiency therefore drops off very rapidly after 40cm displacement.

The other reason lies within the main purpose of the STT: To discard tracks from outside the interaction region. Even if a track produces 4 or more TS hits along the entirety of its path through the CDC, the STT likely would estimate its z-impact to be large. Since the

---

[1]Unless one of the particles is emitted backwards. This does happen, however, depending on the model parameters, this is rather unlikely and never the dominant topology.
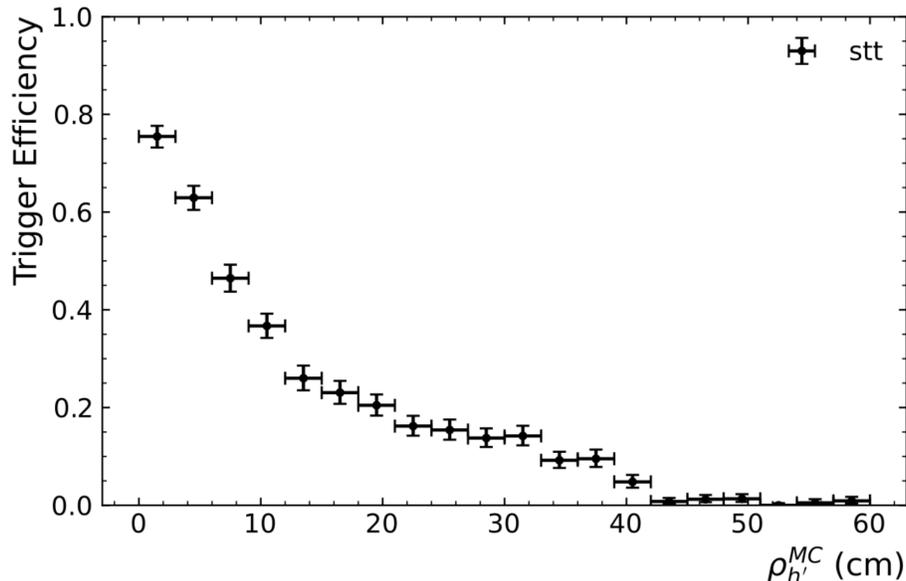
Figure 5.3: Efficiency of the STT against vertex displacement, calculated on simulated data with light background conditions matching the early phase 3 running period. Plot used with permission from Patrick Ecker, KIT.

2D track-finder and therefore also the STT assume tracks to come from a point along the beam axis (that is, with coordinates $(x, y) = (0, 0)$), for displaced tracks, the z-impact will be calculated as the projection of their θ-component back to the origin. With the currently very stringent z-cut set to 5cm, unless the tracks are exactly pointing back to the IP, they are classified as background and rejected. With increasing beam background, the cut is likely to be further decreased, lowering the iDM-detection rates even more. Lastly, the momentum cut of the STT precludes interesting regions of iDM parameter space, as two undetectable particles take away energy from the CDC tracks.

5.4 ECL-based Triggers

If there are no sufficiently clear CDC tracks, the ECL can provide crucial sensitivity to trigger dark matter events. If neutral long lived particles have very large decay times, they exit the CDC without decaying and therefore leave no wire hits. In this case, the only way of detecting such events is by detecting initial state radiation and a large amount of missing energy (i.e. mono-photon searches, described in section 3.2.2). The trigger required for such a search is the so called hie - the high energy single cluster trigger. Like the STT, it is a minimum bias trigger, since it requires only one ECL cluster. Due to high background rates, a strict threshold on the total deposited energy in the cluster must be imposed. In the present configuration, the hie requires clusters with an energy exceeding 1GeV and is still at risk of being prescaled. The main background for the hie is radiative Bhabha scattering, where an electron pair is produced in association with a single photon. If the electrons are not reconstructed due to inefficiencies in the CDC or

because they escape the fiducial detector region, only the ECL cluster is found. Although such partial reconstructions are rare, radiative Bhabha rates are large enough to lead to a sizeable background which proportionally increases with luminosity. In order to cope with these backgrounds, the energy threshold must therefore be gradually increased, too. However, this necessarily decreases the efficiency of the detector with respect to signal events. Furthermore, the detected ISR-photon takes away energy from the main reaction. If it is too high, the remaining energy might not be enough to create a dark matter pair anymore. Thus, increasing the energy threshold of the hie-trigger also cuts away parts of the iDM parameter space that can be probed due to kinematic constraints.

The other problem with mono-photon searches is their susceptibility to Bhabha vetoes. If a single photon cluster is found while unrelated tracks are present in the CDC, the event is likely to be vetoed by anti-Bhabha systems. Since inelastic dark matter models generally present CDC tracks even in events that feature initial state radiated photons, this significantly affects the sensitivity of Belle II towards such models, as outlined in [39].

Finally, for final states featuring electrons and an ISR-photon, there is a further possibility of using the electron clusters in the ECL in order to trigger. Due to the higher number of clusters the energy threshold can be significantly reduced. Exemplary trigger lines include the lml12 which requires 3 clusters, one of which needs to exceed an energy deposition of 500MeV, or the newly implemented dpee which is used to keep signatures with one electron and one photon. Requiring final-state electrons, however, also reduces parameter space, especially in the context of dark Higgsstrahlung, since the dark Higgs coupling strength is proportional to mass, favoring muons, tauons or light mesons in the final state. Furthermore, background levels of pair conversions from $ee \rightarrow \gamma\gamma$ followed by $\gamma \rightarrow ee$ are expected to be challenging. This issue is discussed in more detail in [27], where the Belle II sensitivity to generic dark photon models is explored.

5.5 Displaced Vertex Trigger

In conclusion, it is clear that the current trigger menu of Belle II is inefficient for many dark matter models because the trigger systems and in fact the entire detector were not built to catch tracks from displaced vertices. A dedicated displaced vertex trigger, on the other hand, would specifically search for such signatures, without relying on information from any other subdetector. A standalone displaced vertex trigger would significantly increase the regions of DM-parameter space in which Belle II is sensitive, most prominently in the context of inelastic dark matter, as shown in figure 5.4.

At the very least, a displaced vertex trigger would provide the necessary redundancy even in regions of parameter space that would already be covered otherwise, as background levels are expected to rise.

5.6 Simulation

The Belle II analysis software (BASF2) [42] features a very powerful event simulation based on the toolkit GEANT4 [43], which was originally inspired by a joint investigation of CERN and KEK. Coupled with independent event generators, GEANT4 features a Monte Carlo
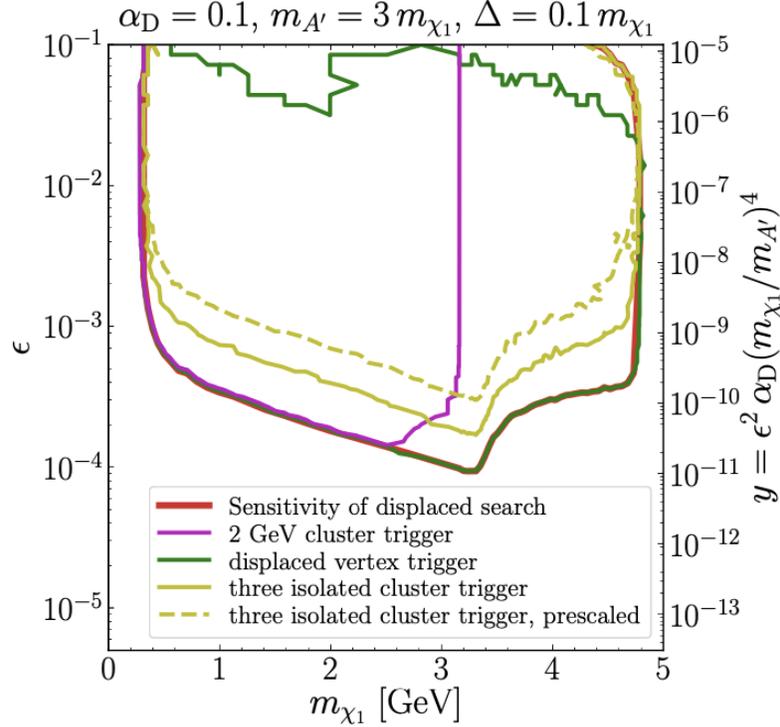
Figure 5.4: Sensitivity of the Belle II experiment to inelastic dark matter. Since inelastic dark matter features 5 free parameters but exclusion plots are typically shown in 2D, 3 parameters need to be fixed. They are defined in the plot title. The displaced search is the sum of all future search strategies, including ECL-based triggers as well as a displaced vertex trigger. Plot taken from [27].

simulation for particle production according to a wide range of physics models. Such models include the well-understood perturbative parts of the Standard Model as well as lattice QCD and, most importantly, new physics models such as inelastic dark matter.

Event generation packages create particles with specified kinematic ranges as well as background samples. Particles are then propagated through material, according to the exact geometry of the detector. The exact response of all subdetectors to particles, including triggering and track reconstruction is simulated as well. In short, BASF2 reliably simulates Belle II from particle production to physics analysis.

The obvious advantage of using simulated data is that the entire underlying information about all particles involved in an event is always known. For physics analyses, a standard procedure is the selection of suitable kinematic regions where the relative background rates for the studied decay are minimized. This is done exclusively in Monte Carlo before looking at real data in order to avoid biasing the end result. In trigger studies the efficiencies can be easily calculated because the true number of signal events is precisely known. For example, all hits in the CDC can be matched to the MC particle that produced them. Of course, studies done exclusively on simulated data need to be taken with a grain of salt. Although the fidelity of GEANT4 to real physics is very high, the risk of artefacts, especially in

previously unexplored regimes, is never zero.

The development of the displaced vertex trigger was, in fact, done with simulated data - it is the obvious choice when studying new physics and there are hardly any Standard Model processes that resemble inelastic dark matter signature. The models used feature both an on-shell and an off-shell dark photon and dark Higgsstrahlung with different Higgs masses ranging from 0.5 to 4GeV, where kinematically allowed. Background conditions were chosen to represent the early physics runs of Belle II with lower luminosity (early phase 3, abbreviated as EP3) as well as the expected background rates in full luminosity running (nominal phase 3, NP3). Furthermore, an interesting possibility is the use of the decay $K^0 \rightarrow \pi^+\pi^-$. Neutral kaons predominantly decay to a charged pion pair with a significantly long lifetime. Typically, the decay length of a neutral kaon is of the order 1-10cm, providing a macroscopically displaced vertex. Offline reconstructed events that feature such a kaon decay somewhere in the CDC may be used to monitor the response of the displaced vertex trigger on real data, not placing too much load on simulation.

Part III

The Displaced Vertex Trigger

Chapter 6

The Hough Transform

At the core of the displaced vertex trigger lies the Hough transform - the same tool used by the 2D track-finder of the CDC to calculate track parameters from TS hits. Now a standard tool in image analysis, it was first proposed by Paul V. C. Hough in order to automate the analysis of charged tracks in a hydrogen bubble chamber in Berkeley. Following its success the transform was patented in 1962 [44].

6.1 Principle

In the bubble chambers that Hough originally analysed, no magnetic field was present, so the tracks of charged particles were straight lines. Photographs of the bubble chamber were taken against a dark background such that bubbles appeared as white dots - quite similar to the wire hits in the CDC of Belle II. In the Hough transform, each point in geometrical space (i.e. a hit) is transformed to a straight line in parameter space called the hit curve (see figure 6.1). Each coordinate in parameter space, in turn, corresponds to a track in geometrical space - one coordinate describes the slope of the track, the other one the x-axis intercept. The interpretation of the hit curve is that each point on in represents a track that passes through the hit in geometrical space - a possible trajectory the particle could have taken. Therefore it is clear that the actual track that produced the hit must lie somewhere on that hit curve.

The Hough transform is then repeated for every hit found in geometrical space, filling parameter space with hit curves. But each one of them must pass through the same point that represents the slope and x-intercept of the true track, as illustrated in figure 6.1. Thus, by finding the intersection of all hit curves, the parameters of the track can be read off.

However, there is a problem in the chosen parametrization of tracks that Hough himself noticed: If tracks lie close to parallel to the x-axis, their x-intercept approaches infinity. Parameter space is thus unbounded (in both dimensions). Hough solved this by demanding that the transform be conducted once more with inverted axes. This way, the track with x-intercept at infinity would in the perpendicular transform have a finite intercept. However, this problem can be avoided all together by choosing another parametrization for the tracks, as explained in depth in [45].

The key is the normal parametrization of straight lines given by the closest distance to
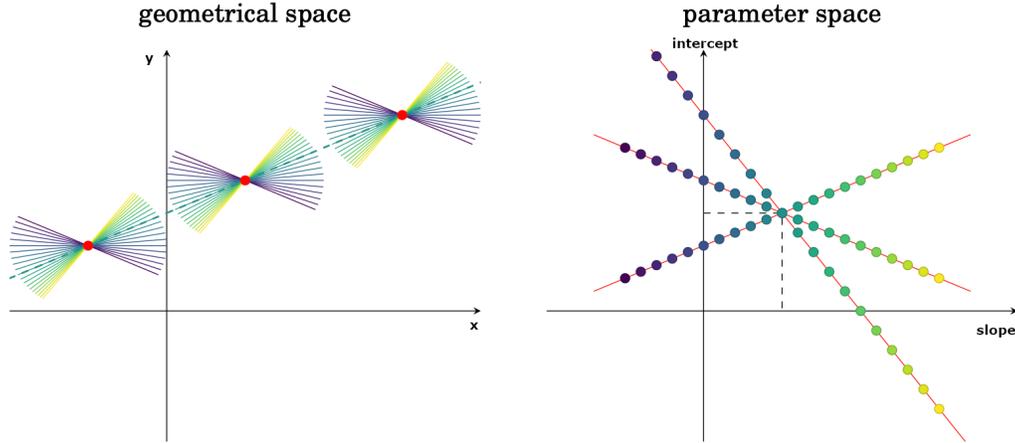
Figure 6.1: Working principle of the Hough transform: Points in geometrical space (red dots in the left panel) are transformed to hit curves in parameter space (red lines in the right panel). The intersection of all curves marks the track parameters of the true track. Image taken from [36].

the origin, ρ, and the angle of its normal (i.e. the line that connects the origin and the closest point) θ rather than the slope. Using the normal parameterization, a track parallel to the x-axis is no problem, as both the angle and the closest distance are finite numbers. Parameter space is still unbounded in ρ, as the line may pass the origin at arbitrary distances, but for a compact region of geometrical space, like a photograph or a CDC, it is compact, too. The points (x, y) that lie on a line parametrized by (ρ, θ) satisfy the equation:

$$\rho = x \cdot \cos\theta + y \cdot \sin\theta \tag{6.1.1}$$

The Hough equation 6.1.1 is also the main equation of the Hough transform, only that x and y are fixed and ρ is calculated for different θ. The resulting hit curves are thus sinusoidal and no longer straight lines. The operation can be easily adapted as a computer calculation by binning parameter space (then called the Hough matrix) and solving the equation by running through θ bins. For every tuple (ρ, θ) that satisfies the equation, one value is added to the corresponding pixel in the Hough matrix, thus filling it each hit at a time. The intersection of hit curves in parameter space is simply the pixel in the Hough matrix with the highest value - precisely equal to the number of hits on the track (at least, if the granularity of the Hough matrix is fine enough). In principle, the track parameters can be found by locating the peak in the Hough matrix.

6.2 Conformal Mapping

Unlike Hough's bubble chambers, the CDC has a magnetic field and tracks are not straight lines but circles when projected to the transversal plane (i.e. the plane that runs perpendicular to the magnetic field). Thus, before applying the Hough transform, the circular track paths must be mapped to lines. This is possible with a conformal map function that can

be derived from the equation of a circle which by definition passes through the origin, i.e. the trajectory of a charged particle with production vertex at $(0, 0)$:

$$\begin{aligned} r^2 &= (x - r\cos\phi)^2 + (y - r\sin\phi)^2 \\ &= x^2 + y^2 - 2r\,(x\cos\phi + y\sin\phi) \end{aligned} \tag{6.2.1}$$

Here, $(r\cos\phi, r\sin\phi)$ is the position of the circle center, expressed using standard polar coordinates. By dividing by $x^2 + y^2$ and rearranging, the equation can be recast to fit the form of the Hough equation 6.1.1:

$$\begin{aligned} \frac{1}{r} &= \frac{2x}{x^2 + y^2}\cos\phi + \frac{2y}{x^2 + y^2}\sin\phi \\ &\equiv x'\cos\phi + y'\sin\phi \end{aligned} \tag{6.2.2}$$

Therefore, by identifying $x' = 2x/(x^2+y^2)$ and $y' = 2y/(x^2+y^2)$ the circular path of a track can be mapped to a straight line, as illustrated in figure 6.2. The distance to the closest point of the line to the origin, ρ, becomes the inverse of the circle radius and the angle of the circle center maps to the angle of the normal of the line. There is an implicit assumption
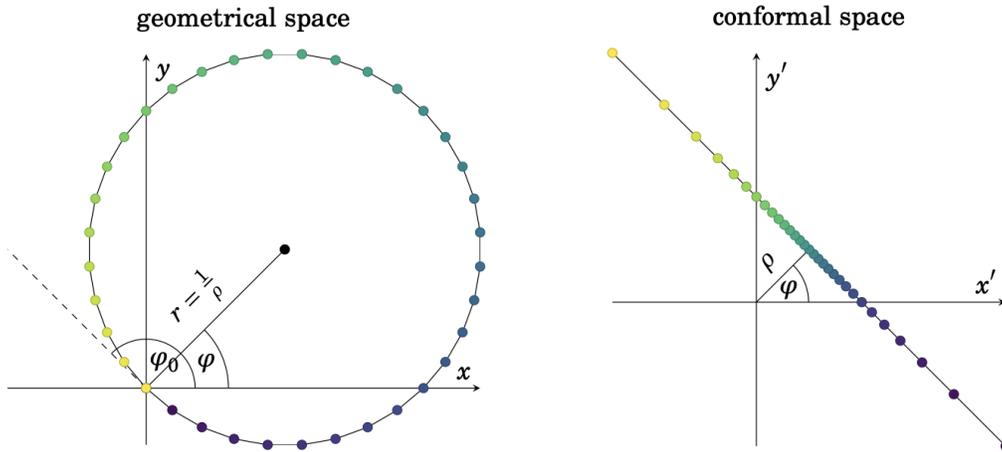
Figure 6.2: Conformal mapping of a circle to a line, also taken from [36].

about the vertex position in the conformal map that becomes apparent in equation 6.2.1: The point $(0, 0)$ trivially solves the equation for all ρ and ϕ, therefore the conformal map assumes the origin to be the vertex of the track. Before performing the Hough transform, all hits must be transformed according to the conformal map $x \mapsto x'$ and $y \mapsto y'$. For any vertex assumption, the full Hough algorithm has 4 major steps:

- Load the hit map of an event

- Perform the conformal transformation (assuming a vertex)

- Solve Hough equation in θ bins

- Find the intersection of the hit curves in parameter space

6.3 Track-finding

By construction, the Hough matrix has a periodicity of 2π. However, there is an ambiguity in the Hough equation (6.1.1) inherited from the axial symmetry of a circle: If ρ is allowed to be negative, shifting θ by π has the same effect as flipping the sign of ρ. In other words, a circle with negative curvature is identical to the same circle rotated by half a turn.

This can be exploited to distinguish the charge of particles in the CDC. As pointed out in [36], the ambiguity can be solved by either limiting the Hough matrix to positive values of ρ (or, equivalently, to $\theta \in [0, \pi)$) or by limiting the hit curves to represent only outgoing tracks - i.e. half circles. By choosing the second option, positive values of ρ exclusively create clockwise-turning, outgoing tracks, which, due to the polarization of the Belle II solenoid, correspond to positively charged particles. The respective opposite holds for negative values of ρ. The Hough matrix is kept to its full extent, with $\rho \in \mathbb{R}$ and $\theta \in [0, 2\pi]$.
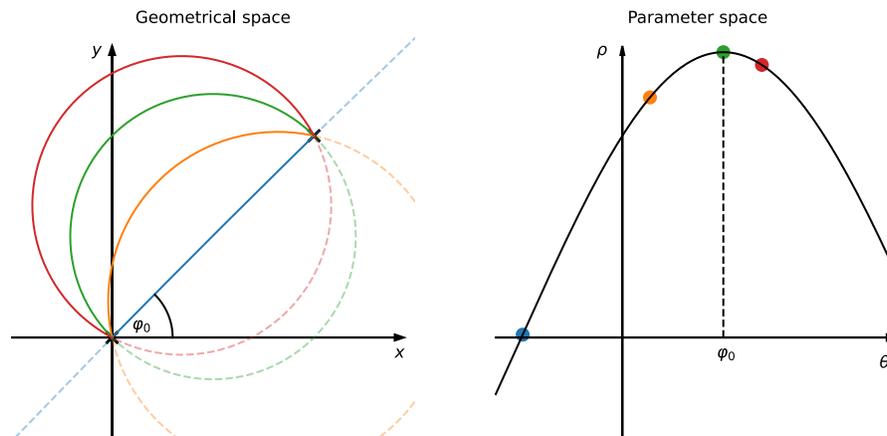
Figure 6.3: A hit in geometrical space (black cross) produces a sinusoidal curve in parameter space (black curve in right panel). A few selected tracks with positive curvature and their respective points in parameter space are shown.

But how to limit the hit curves? Looking at the parameter space curve of a single hit (figure 6.3), it is clear that the track with maximal curvature (green track in the figure) must satisfy $\theta = \varphi_0$ where φ_0 is the polar angle of the hit and θ is the angle of the center of the track, as illustrated in figure 6.2. Every track with a higher curvature does not reach the hit.[1] If θ is increased (that is, the track is rotated counter-clockwise), the only way to reach the hit is by curling back. The resulting track is larger than a half-circle. On the other hand, decreasing θ leads to tracks reaching the hit earlier in their arc and the track

---

[1]Mathematically, this manifests itself in the boundedness of the Hough equation. For a given point (x, y) the sum $\rho = x \cdot \cos\theta + y \cdot \sin\theta$ has an upper and a lower limit. Any value of ρ outside of these bounds is never reached by the function.

curvature decreases until a quarter turn out of phase (so at $\theta = \varphi_0 - \pi/2$) the track reaches minimal curvature (blue track). The track radius approaches infinity as the arc becomes a straight line. Further decreasing θ flips the sign of the curvature and the track now turns counter-clockwise until at $\theta = \varphi_0 - \pi$ tracks start curling again.

Therefore, limiting the hit curves to the range $\theta \in [\varphi_0 - \pi, \varphi_0]$, that is, the range in which the hit curve is monotonically increasing, makes sure only outgoing tracks are fitted, excluding curling tracks.

6.4   Advantages of the Hough Transform

The Hough transform is a great tool for fast image analysis because it allows to utilize prior knowledge about the shape of features in the image. In this case, particle tracks in the CDC are known to be (to reasonable approximation) circular in the plane transversal to the magnetic field. The algorithm does not have to learn the spatial correlations between the hits on its own, it is already implicit in the conformal mapping.

A different approach to track-finding is the use of Graph Neural Networks (GNNs). GNNs sequentially build up tracks by analysing correlations of hits, essentially learning their shapes from scratch. This allows them to find tracks where the Hough transform fails, for example when tracks start spiralling. However, Hough transforms can be fully parallelized and contain, at their core, only elementary calculations. This makes them very suited for the task of triggering and fast track-finding.

6.5   The Hough Matrix of the DVT

Particularly low-momentum tracks are sometimes not able to reach the end of the CDC and thus curl back into the CDC volume. Many dark matter models predict a large amount of these curling tracks due to two main reasons: First, inelastic dark matter features two invisible particles that leave the CDC undetected, taking away energy from the two tracks in the CDC. Low energy tracks mean high curvatures. Furthermore, the tracks do not always point radially outwards, because they may come from 3-body decays. If momentum is carried away by an invisible particle, the symmetry of a two-body decay is spoiled. Therefore, for large displacements, curling tracks with relatively high energies are possible, since they have more CDC volume to cross.

Curling tracks also pose problems in standard offline track reconstruction, therefore I adopt the convention of limiting the hit curves to fit outgoing track. The decrease in efficiency due to misreconstruction of curling tracks is far outweighed by the overall cleaner picture of parameter space if clipped hit curves are used. If full circles are fitted, background hits on the opposite side of the vertex create hit curve intersections where no spatial correlation is present, polluting otherwise clean intersections with background curves. However, in order to increase the efficiency of low-energy tracks, the maximal curvature track the Hough matrix represents is increased with respect to the 2D track-finder.

The parameters that define the size and structure of the DVT-Hough-matrix are provided in table 6.1. They are compared to their counterparts in the 2D track-finder used in the first-level trigger, taken from [36]. It should be noted that the standard track-finder

utilizes the drift times of the priority wires for more precise tracking. This number is not used by the DVT. Since the track curvature is allowed to be negative, the value $\rho = 0$ describes a straight track with no curvature. The track radius r is therefore not bounded from above. However, the Hough matrix has to be limited, i.e. there must be a maximal value ρ can reach. In the case of the DVT this value corresponds to a heavily curling track of 25cm radius, or about 100MeV of momentum.

variable	2D finder	DVT
θ bins	160	160
$1/r$ bins	34	40
$r_{\min}$	≈ 67cm	25cm
$\min(p_T)$	≈ 290MeV	≈ 100MeV

Table 6.1: Comparison of the Hough matrix of the 2D track-finder and the displaced vertex trigger

Chapter 7

Data reduction

7.1 The New Track Segment Finder

The average number of hits in an event with early phase 3 background conditions is about 400 and 10 times more for the severe backgrounds of high-luminosity running. It is clear that the vast majority of hits are background. Performing an individual Hough transform on each hit would have an enormous consumption of computational resources. Further, the Hough matrix would be completely occupied and peakfinding would likely not work anymore. Evidently, some kind of preprocessing needs to be performed before the Hough transform. As discussed in section 5.1, for the normal 2D track-finder this is done with the TS-finder, greatly reducing the amount of independent Hough transforms that need to be performed. However, the current hourglass-shaped TSs perform poorly for shallow tracks. Their efficiency with respect to the crossing angle of the track through the track segments is shown in figure 7.1.
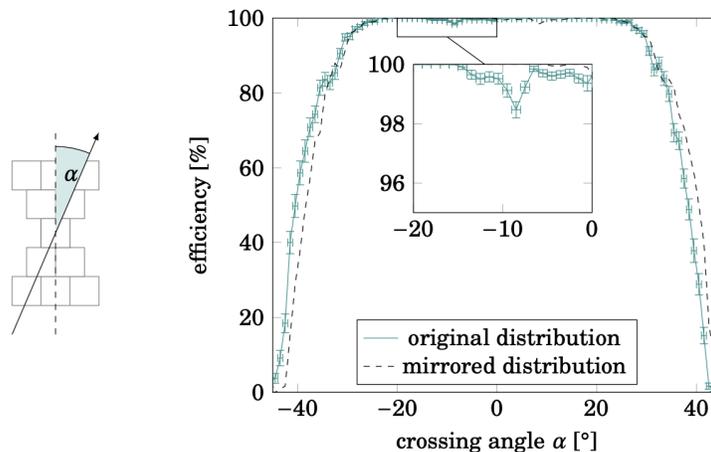
Figure 7.1: Efficiency of the hourglass TS-finder against crossing angles of tracks through track segments. The mirrored distribution is superimposed onto the original to highlight a slight asymmetry towards positive angles. Plot from [36].

The larger the displacement of a vertex gets, the more shallow the tracks that originate

from that vertex become on average. If the parent particle is massive or has low momentum, the opening angle is large, making tracks even shallower. It is not uncommon for one track to have constant r - that means a crossing angle of $90°$ or even to curl back into the CDC. The hit efficiency for many inelastic dark matter models is thus very poor. For the purpose of displaced vertex decays, it is necessary to put a new track segment finder in place. Such a system was in fact studied in [46]. Rather than relying on the hourglass TSs with a hit in every layer, the new TS-finder works with a set of look-up tables (LUTs) of hit patterns and a track-background-classifier. LUTs of different sizes are shown in figure 7.2.
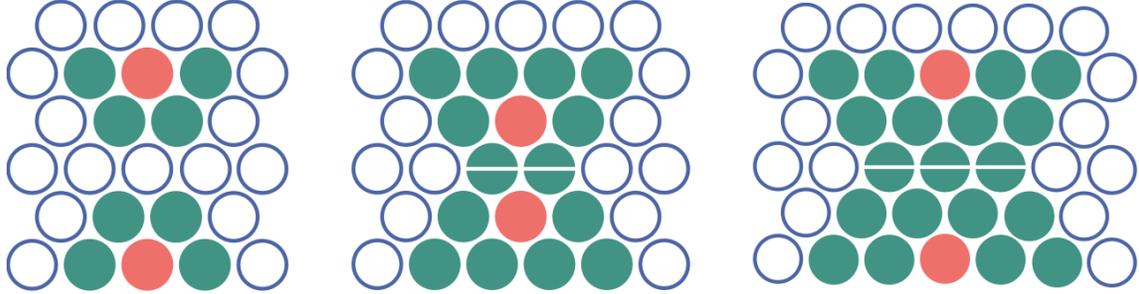
Figure 7.2: Track segments for different cell sizes: 5 wires (=bits) in the left, 9 in the middle and 12 in the right panel. Priority wires (red circles) were tentatively placed within the segments but not used in the final version of the DVT. Image adapted from [46].

Like with the hourglass TSs, only 5 layers within a superlayer are used for the trigger decision. The last layer is added in offline track reconstruction, where more time and computational resources are available. All possible hit patterns within a given cell are analysed using MC samples of inelastic dark matter models (the same samples used for the development of the DVT). A score is given to each single hit pattern in the TSs based on how often the pattern was set off by signal tracks and how often by background. A subset of patterns is then chosen based on hit accuracy and stored in a LUT. The amount of patterns that are accepted is a free parameter of the algorithm and can be optimized. During online processing, the actual hit patterns from the CDC are then compared with the ones in the LUT and only valid ones are kept. This way, noise is suppressed and the number of hits per event reduced significantly.

The current TS-finder uses priority wires to represent the track segments. The hits in all layers within a track segment are collapsed onto one point (the priority wire) and only one Hough transform is performed. A similar system was tentatively put in place for the LUT-based TS-finder, but eventually discarded, as the upgrade from UT3 to UT4-boards during LS1 will make it possible to pass the full hit information to the 2D-finder (and the DVT), increasing tracking precision. Therefore, the full picture of the CDC is used for the DVT and no priority wire algorithm is implemented.

The data reduction performance of the new TS-finder is shown in figure 7.3. The ROC-curves for different cell sizes as well as for the hourglass TS-finder are calculated on an iDM-model with different background conditions. Efficiency is evaluated at hit level and defined as the fraction of signal hits that survive TS-preprocessing. Similarly, rejection is the fraction of background hits from all sources that is left after preprocessing. From the figure

it can be read off that the conventional hourglass TS-finder indeed rejects background hits very well. However, it also captures very few signal hits. The trained 12-bit LUT TS-finder scores significantly better, while no large deterioration is observed for a similar 9-bit LUT TS-finder. Since the hit information that is passed to the TS-finder only contains 5 layers per superlayer, even a perfectly trained TS-finder cannot achieve 100% efficiency. Furthermore, a significant discrepancy in efficiency is observed between EP3 and NP3 samples that cannot be explained by the incomplete hit information, because it should affect both backgrounds equally. The discrepancy is rather due to issues in data simulation that are well understood and explained in appendix B.1. Due to these issues, a TS-finder trained on only EP3 data is used for both background conditions, significantly affecting performance in the severe case. After preprocessing by the TS-finder, the signal hit multiplicity of EP3 events is reduced to approximately 57 hits on average. For NP3, only about 42 signal hits remain.
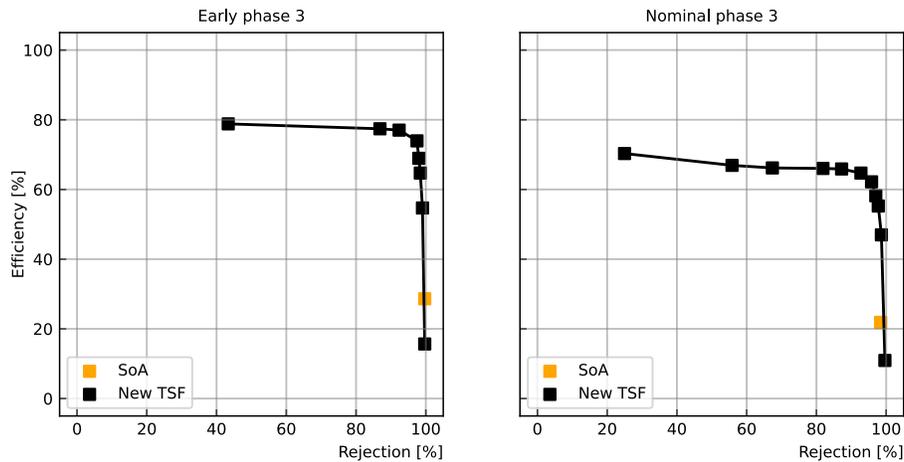
Figure 7.3: Efficiency against background rejection on hit level for different numbers of accepted 12-bit hit patterns for early phase 3 (left panel) and nominal phase 3 (right panel) running. SoA (state of the art) represents the presently used hourglass-shaped TS-finder.

7.2 ADC Values

Another possibility for the data reduction step before track-finding is a cut based on the deposited charge on CDC wires. The upgrade of the trigger boards will make it possible for ADC information to be utilized in some form (probably 2 bits per wire) as a means of reducing the amount of hits in the CDC. The distribution of ADC values for signal and background hits in a MC-generated data sample of inelastic dark matter with a dark Higgs boson is shown in figure 7.4.

Background hits dominate the regions of very low ADC values (under approximately 10 counts) as well as very high values (above approx 250). Furthermore, there is a peak in the background distribution around the value 140. This has multiple reasons: Cross-talk
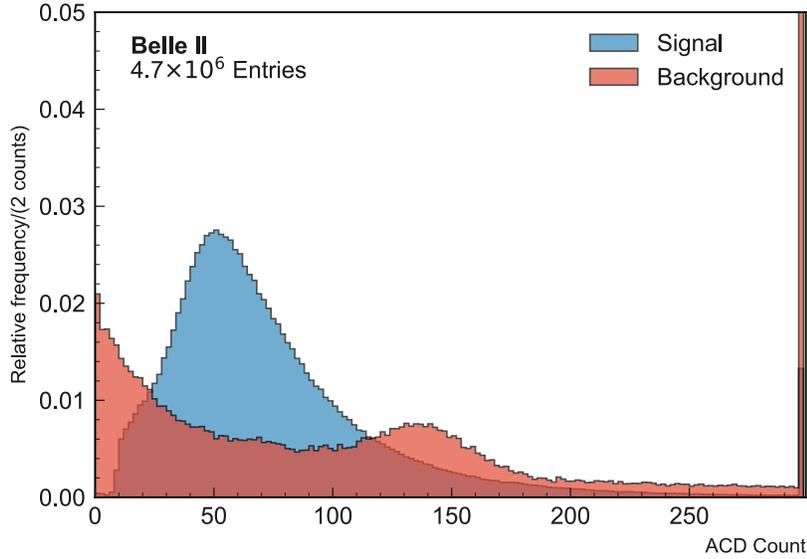
Figure 7.4: Distribution of ADC counts for signal hits and background, for an iDM + DH model and with nominal phase 3 background conditions. Both histograms have been normalized to the number of hits in their respective category.

between wires can cause fake hits to be registered, these typically lie at the 140-peak. Fake hits also show up at very low values of deposited charge. On the other side, synchrotron electrons that curl around few sense wires with a very high curvature constantly deposit charge along a wire until they either leave the detector or lose enough energy to fall under the ionization threshold. This leads to very large charge counts on these wires and a long tail in the distribution for large ADC values. Cutting hits that present ADC values in the extreme ranges increases the signal-to-noise ration significantly.

Early phase 3 background conditions are not very severe and the TS-finder is able to weed out enough background hits for the remaining picture to be very clean. Applying an ADC cut is not necessary at this stage and could even lower efficiency due to signal hits being discarded. Nominal phase 3 presents backgrounds that are orders of magnitude more abundant. Here, the ADC cut is necessary, since the improvement in background rejection vastly outweighs the decrease in signal efficiency. The order in which the pre-processing techniques are used is also clear: Applying the ADC cut first disrupts the spatial correlations of signal hits and the TS-finder efficiency is worse. For example, hit patterns that contain only one or two hits always present low precision and they are never accepted. One could imagine a scenario where in one superlayer there are 3 spatially correlated signal hits one of which has a low ADC count. Applying the cut first would remove the outlier and pass only two hits to the TS-finder. Since two-hit patterns are discarded, the entire superlayer disappears. Instead, the TS-finder is applied first. It sees 3 correlated hits and identifies them as signal, the subsequent ADC cut still removes one hit, but the rest are passed. The example shows that the order of operations is not irrelevant and that it is favorable to apply

first the TS-finder and then the ADC cut. For nominal phase 3, a cut of values below 10 and above 150 is found to be best suited.

7.3 Backgrounds

While very effective in reducing readout noise, cross-talk between wires and synchrotron tracks, the combined data reduction processes cannot suppress all backgrounds. There are background processes that create clean tracks (like protons from nuclear spallation or Bhabha-electrons) that are virtually indistinguishable from signal tracks, including realistic ADC distributions. Background tracks can cause a positive trigger response in two cases: A clean background track combined with a fake track created by bunched-up hits, or the coincidence of two clean background tracks in the same event, mimicking a displaced vertex. Studies performed at KIT have shown that with nominal phase 3 conditions, background tracks with more than 50 hits are expected frequently. Rejection of these backgrounds is only possible by analysing the relation between tracks.

Chapter 8

MacroCells

The displaced vertex trigger proposed in this thesis is a standalone system - it only needs unprocessed CDC data and does not require additional information from other subdetectors. If two tracks with opposite charges originate from a common vertex, the DVT should trigger the event and output rough estimates of the vertex position and the track parameters.
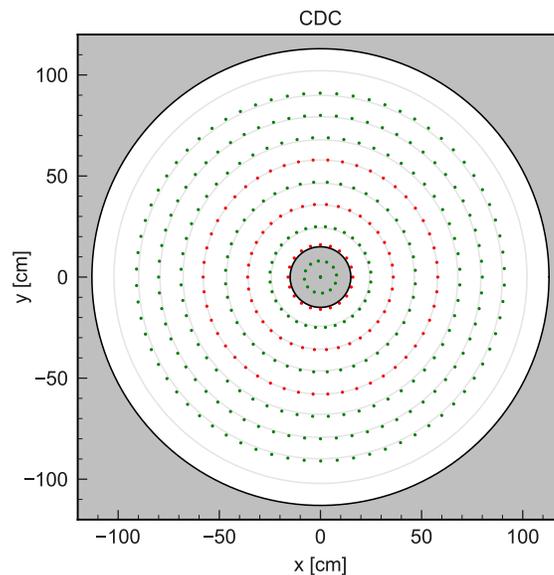
Figure 8.1: Configuration of the MacroCells. 387 reference vertices are placed inside the CDC and the inner region in accordance to the wire structure (light gray circles). In parallel, they probe if two tracks could have originated in that region. MacroCells are subdivided into two categories (green and red dots) based on the superlayer types.

8.1 Basic Principle

The tool that is used to fit tracks is the Hough transform. However, as explained in chapter 6, the Hough transform (specifically the conformal mapping) implicitly contains an assumption about the origin of the tracks. The 2D-finder uses the IP as reference vertex which is an excellent assumption for its purposes but ultimately fails when analyzing tracks with substantial displacements. For the DVT, the CDC is thus subdivided into 387 regions called MacroCells, arranged with polar symmetry around the IP and based on the CDC wire configuration. Each MacroCell represents a reference vertex that it tested in every event. The number of MacroCells is a free parameter of the algorithm - it affects the trigger efficiency and the spatial resolution of the vertex fit but is limited by hardware constraints. The number of MacroCells in the preliminary configuration is, after a rough analysis carried out by members of the ITIV at KIT in Karlsruhe, within the capabilities of the new FPGAs used for the new UT4 boards.

layer	type	radius	wires per layer	MacroCells per layer	offset
0	A	0cm	nan	1	0
1	A	8cm	nan	12	$\pi/6$
2	B	16cm	160	20	0
3	A	25cm	160	32	$\pi/8$
4	B	36cm	192	38	0
5	A	47cm	224	44	$\pi/11$
6	B	58cm	256	50	0
7	A	69cm	288	56	$\pi/14$
8	A	80cm	320	64	0
9	A	91cm	352	70	$\pi/17$

Table 8.1: MacroCell Configuration. The first two layers have no corresponding CDC superlayer. Layer 0 corresponds to the IP. Since the CDC begins at a radial distance of 16cm, there is no CDC wire equivalent for the inner two MacroCell layers.

For every event, all 387 MacroCells are analysed in parallel as potential sources for the hits using modified Hough transforms and a neural clustering algorithm. In order to trigger an event, two oppositely charged tracks must originate at a MacroCell. With this simple but clear signature, background rates from cosmic muons and beam interactions are suppressed. Such events present only one track and are not triggered. Even if two tracks happen to coincide, they have no physical reason to be in any way correlated, as opposed to a particle decay like dark matter transitions.

The configuration of MacroCells is shown in figure 8.1 and is based on the CDC wire configuration. For each superlayer there is one layer of MacroCells located at the inner radius of the superlayer and two additional layers in the inner region before the CDC. In order to save hardware resources, the layer farthest from the IP is not used as for such large displacements the efficiency is expected to be very small low and the number of MacroCells would be very large. MacroCells are distributed into 2 categories (type A and type B), roughly depending on the superlayer they are in. Type A MacroCells are located in the 2 innermost layers as well as MacroCells in stereo superlayers and have a minimum distance

to hits, since only axial hit information is passed to CDC triggers. Type B MacroCells, on the other hand, encounter hits at arbitrarily small distances. The Hough matrices of type A and type B MacroCells therefore look different and there is a difference in performance between the two. Because of these performance asymmetries in the two classes, they are treated separately. Details about this are provided in chapter 8.3. A summary of the MacroCell structure is provided in table 8.1.

Apart from layer number 2, the number of MacroCells in each radial layer is roughly equal to 1/20 of the number of wires. Layer 2 corresponds to the innermost superlayer where the number of wires is greatly reduced due to high occupancy. Furthermore, in order to minimize the mean inter-cell distance and to avoid patterns with polar symmetry, MacroCell layers have azimuthal offsets in alternation.

8.2   Limitations of the Hough Transform

The Hough transform takes the spatial coordinates of hits in the transversal plane and outputs 2D tracks characterized by two independent parameters. It is therefore a 2D fit. The DVT, however, requires to not only obtain the tracks but also the x and y position of the vertex - two more numbers. Since the Hough transform is only suited to fit two parameters, demanding that it, too, determine the vertex position yields unreliable results.

The natural idea for the DVT-algorithm would be to simply perform a Hough transform on each MacroCell and extract the two highest peaks. The signature of interest has two tracks, thus two peaks in the Hough matrix should be enough. Then, the MacroCell with the two highest peaks could be selected and the event triggered, if the value surpasses some threshold. Unfortunately, this simple algorithm fails even when analysing events without background. The main problem is that reference vertices located far away from all hits manifest peaks much higher even than the peaks calculated at the correct vertex. Compare the two exemplary Hough transforms in figure 8.3 performed for the Monte Carlo vertex and for a wrong vertex hypothesis. There are 3 major reasons for this bias:

- The active area of one Hough matrix pixel is the set of points in geometrical space that under the Hough transform map to that pixel. There, the hit curves of all these points intersect exactly, yielding a fan of divergent intersecting Hough curves (top right panel of figure 8.3). The left and right borders of this active area diverge when moving away from the reference vertex, reaching a maximal distance at the apex of the track. Therefore, there is a bigger tolerance for uncertainty if hits are far away from the reference vertex. If the active area is large, more hits fall into the same bin, even if they are background hits, resulting in a higher average peak. Conversely, at close distances to the vertex, even slight deviations from the theoretical track cause hits to fall out of the active area of the correct pixel. The hit curve of such a hit therefore miss the theoretical intersection in the Hough matrix. Hit deviations are inherent to the CDC structure, since the inter-wire distances of up to 2cm are sufficient to cause such effects. Secondly, the positions of the reference vertices is bound by the structure of the MacroCell grid, with distances of 7cm on average. For the current 2D track-finder this is not a problem, because the IP is the only reference vertex and it
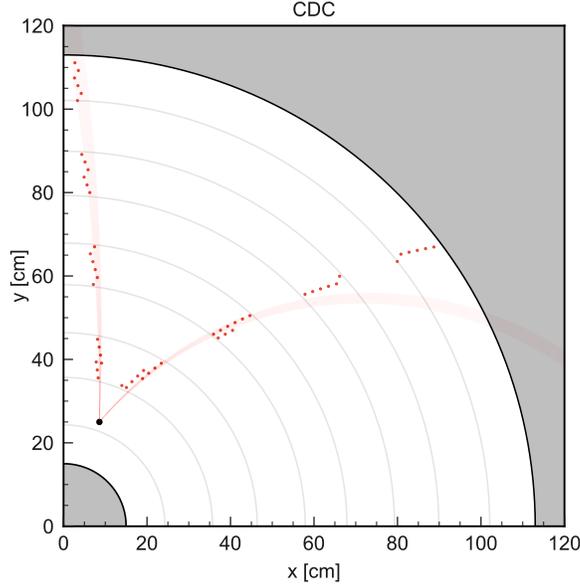
Figure 8.2: Active area (red area) of the Hough matrix peaks in an example event. The MacroCell is not perfectly aligned with the tracks, causing a shift in track parameters. Not all hits are inside the active area, their hit curves do not intersect at the theoretical intersection. This can cause imprecise track parameters.

has a distance of 16cm to the closest hits determined by the inner radius of the CDC. The effect this has on MacroCells is illustrated in figure 8.2.

- In principle, for every value of θ there is a track that reaches any given hit, since the Hough equation is continuous. However, there is not a track for every value of r, since there is a minimal track radius equal to half the distance from the vertex to the hit. The amplitude of the sinusoidal hit curves is inversely proportional to the hit distance: Hits that are far away from their parent vertex produce curves with small amplitudes - the occupied region of the Hough matrix is limited. On the flip side, very close hits can have track curvatures that even exceed the maximal value of the Hough matrix (see table 6.1) and the hit curves are essentially cut off.

  Thus, Hough curves of far-away hits are bounded while Hough curves of close hits are stretched to occupy the entire $\frac{1}{r}$-range of the Hough matrix. The result: The average density of hit curves must be higher for far-away hits, adding to the bias.

- The third effect is a consequence of the finite granularity of the Hough matrix. Digitising a curve, i.e. binning its parameters, has the drawback that there is a maximal graph slope that can be displayed continuously. In a matrix with equal bin density in both dimensions, a continuous curve has the maximal slope of ± 1. It is a diagonal line of pixels that share one corner. Any steeper than this and bins in y start to get skipped as the step from $y(x)$ to $y(x+1)$ becomes larger than the bin size in y. The Hough matrix suffers from this issue, too. It is especially problematic in this

case, as a very steep hit curve might miss the intersection and make the peak-finding challenging.

The complete calculation of the maximal continuous slope of the Hough matrix is carried out in appendix A - the result is that above an absolute slope value of ≈ 19.6 (in the units of Hough space) curves are no longer continuous.

The slope of hit curves can be calculated from equation 6.2.2. By applying trigonometric identities it can be seen that hit curves are just phase-shifted sine curves

$$
\begin{aligned}
\rho &= x' \cos \theta + y' \sin \theta \\
&= \sqrt{x'^2 + y'^2} \cdot \cos(\theta + \phi)
\end{aligned}
\tag{8.2.1}
$$

with the phase factor $\phi = \arctan(-y'/x')$. The point of maximal slope of an unmodified sine curve lies at its intersection with the x-axis and is equal to its amplitude. Transforming from conformal back to cartesian coordinates yields

$$
\rho'_{\max} = \sqrt{x'^2 + y'^2} = \frac{2}{r}.
\tag{8.2.2}
$$

Evidently, the closer a hit lies to its parent vertex, the higher the maximal slope of the hit curve becomes. Comparing this quantity to the maximal resolvable slope of the Hough matrix, it follows that under a distance of ≈ 40cm, hit curves become discontinuous. Curvature bins start being skipped and intersections in the Hough matrix get more rare, hindering the peak-finding algorithm.

This effect can be suppressed by increasing the resolution of the Hough matrix in θ, or even by altering the entire algorithm to allowing multiple $\frac{1}{r}$-bins for every angle. All of these possibilities, however, are limited by hardware capabilities as the algorithm should be minimally complex.

Figure 8.3 exemplifies the above effects: Although the "tracks" found from the random vertex are clearly inconsistent with the hit positions, the peaks in the Hough matrix are almost twice as high than the correct peaks found from the MC-vertex - 30 and 28 versus 17 and 14, respectively. Evidently, it is not enough to perform unmodified Hough transforms and simply add peak values together. However, the shapes of the peaks are completely different: While hit curves from true tracks form nice fans coalescing to clean and symmetric intersections, the peak artefacts from the random vertex do not show any of these characteristics, showing bar-like patterns of the intersection area. A cluster-pattern algorithm that analyzes the crossing regions is expected to perform much better than a simple threshold, though at the expense of increased complexity. Applying clustering algorithms to 400 MacroCells in parallel may become problematic. Thus, a simpler solution for the geometric limitations of the Hough transform that does not add much complexity to the algorithm is needed.

8.3 Weighted Hough Transform

This simpler solution is given by a special weighted Hough transform. It is introduced to counteract the bias towards far-away vertex hypotheses. In the original Hough transform
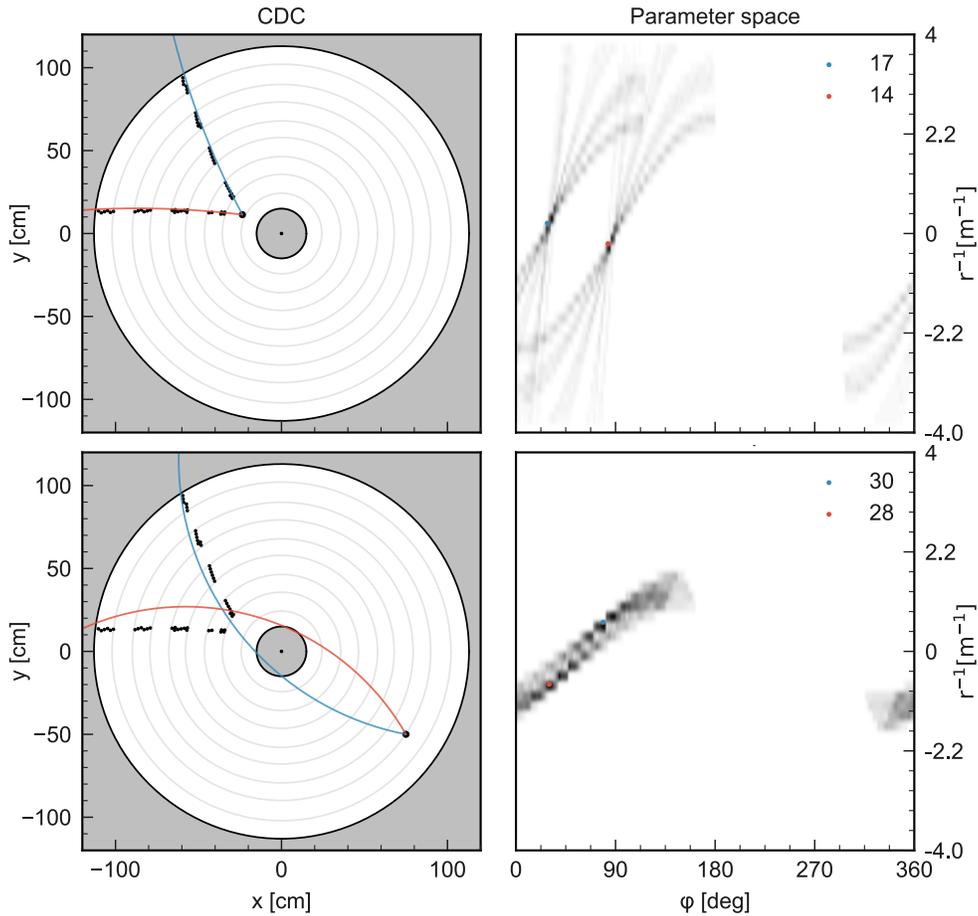
Figure 8.3: For the same event, Hough transforms are shown from the correct vertex position (top right panel) and a random, far-away vertex (bottom right panel). The tracks are obtained from simple peak-finding of the Hough matrix and are color coded to match the peaks.

algorithm, the value of each pixel through which the hit curve passes is increased by one. In other words, the hit curve has a weight equal to unity. In the modified Hough transform, the hit curves are first weighted according to their distance to the assumed vertex. Conformal mapping already necessitates the calculation of the vertex distance so using it to determine weights does not add any new expensive computations. A weight function in the form of a 3-bit look-up table is then applied to retrieve the corresponding hit-weight. Generally, the closer a hit is to the vertex, the higher its weight should be, see figure 8.4. Lastly, the pixel values in the Hough matrix are incremented by the weight of the hit curve rather than simply by one. By setting all values in the LUT to unity, the original unweighted Hough transform can be recovered. The Hough weights used in the final version of the DVT are listed in table 6.1.

As a second measure, the hit curves corresponding to very close hits (i.e. hits closer
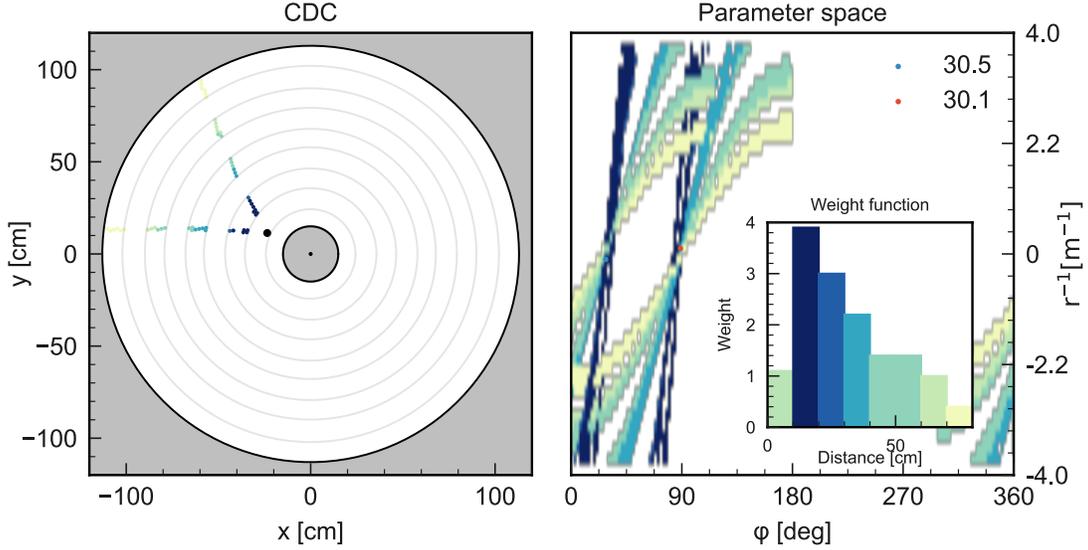
Figure 8.4: Illustration of the working principle of the weighted Hough transform. The same event as in figure 8.3 is shown, using the weighted Hough transform. Hits are weighted according to the function shown in the nested plot, the resulting color-mapped distribution is shown in the left panel. The Hough matrix is shown in the right panel using the same color scheme.

than 20cm to the selected reference vertex) are artificially broadened by one pixel to each side in $\frac{1}{r}$-direction. This way, discontinuous hit curves are suppressed and the probability of a hit curve to pass through the theoretical intersection is increased. In the example event depicted in figure 8.3, the weighted Hough transform assigns a total score (that is, the sum of the two peak values) of 60.6 to the true vertex and only 16.0 to the wrong vertex hypothesis in the lower panel, correcting for the bias of the unmodified transform.

I want to note that both hit curves with weights and broader than a single pixel are not unprecedented and were proposed - albeit for a different purpose - in [47], where a 3-dimensional Hough transform is proposed for the neural track trigger.

8.3.1 Peak-finding and Vertex Preselection

In the Hough matrix, peaks are found starting from a threshold. However, the region around the intersections of hit curves in the Hough matrix usually contains multiple over-threshold pixels. Accepting each would lead to up to 5 or 6 fake tracks being found per cluster. Instead, after every accepted peak, a mask is laid over the surrounding area of the Hough matrix, blocking any more peaks from the near vicinity. The mask is rectangular and spans 7 bins in θ-direction and 12 bins in $\frac{1}{r}$-direction, corresponding to an angle between the tracks of approximately 15 degrees. Tracks with smaller opening angles cannot be resolved into a pair and are often misidentified as one. The drop in efficiency due to tracks that are genuinely this close in the Hough matrix is negligible (approximately 2%) and fake tracks

from the same cluster are reliably suppressed. The DVT has the specific task of finding exactly two tracks from one vertex. Thus, the peak search can be limited to two peaks, truncating the iterative process to deterministic computation times needed for the FPGAs.

Even with a weighted Hough transform and broadened hit curves, a more involved clustering algorithm is necessary to reliably distinguish two-particle vertices from fake tracks and artefacts in the CDC. Since there is a large number of MacroCells each already demanding an involved computation, it is not possible to apply any kind of complex clustering on every MacroCell in parallel. However, simple peak-finding after a weighted Hough transform is precise enough to substantially narrow down the number of vertex candidates to take into consideration. Most vertices are discarded right away because their peaks do not reach the peak-finding threshold. Therefore, the displaced vertex trigger features a two-step process: First, 6 type A MacroCells and 4 type B MacroCells are selected as vertex candidates according to the heights of their 2 peaks, while all remaining vertices are discarded. This step is called preselection. Treating the two classes of MacroCells separately rather than simply selecting any 10 candidates yields better results due to a significant difference in performance between the two classes. As pointed out in section 8.2, the Hough transform encounters issues when fitting hits that lie close to the reference vertex. CDC triggers only see 2D data and never get stereo hit information, therefore tracks have periodic holes whenever they pass through a stereo layer. A MacroCell that lies within a stereo layer never faces the problem of very close hits and on average performs better. Using only one class of MacroCells would lead to almost no axial MacroCells being selected, even if the correct vertex lies within an axial superlayer. In total, 10 vertex candidates are selected and further analysed using a Neural Clustering Algorithm. Since the clustering algorithm is only applied to a small subset of MacroCells, the hardware limitations can be met. The exact number of vertex candidates is not fixed and can be optimized later.

It must be noted that while 10 is the current *maximal* number of vertex candidates passed on the neural clustering, in order to even be considered a candidate, the Hough transform computed with respect to a reference vertex must present two distinct peak that surpass the threshold. If this condition is not met, the vertex cannot become a vertex candidate. Therefore, there can in principle be events that have fewer than 10 or even no vertex candidates at all after vertex preselection. This is indeed observed to happen in pure background events with particularly clean background conditions in early phase 3 samples.

Chapter 9

Neural Clustering

The remaining vertex candidates are analysed with a neural clustering algorithm that eventually makes the trigger decision by exploiting the different cluster shapes observed from true and wrong vertex hypotheses. Since the vertex candidates are already fixed and the natural favoring of close hits described above is no longer an issue, the Hough matrix for clustering can be built using a different set of weights. In order to favor a precise track-parameter estimation, exclusively axial hits are used. In the Hough matrix, clusters are built iteratively from initial seeds and simple cluster parameters are calculated. The parameters corresponding to two tracks are combined to an array and passed to a neural network with a single output node. The network classifies clusters into two categories: clusters that likely correspond to tracks with the desired signature and fake tracks. The trigger decision is made based on a threshold on the output value.

9.1 Cluster Algorithm

The clustering algorithm starts from an initial seed found with the same peak-finding algorithm used in the preselection step. Clusters are built iteratively but with a fixed number of steps in order to keep computation times deterministic. Starting from the seed, neighboring Hough pixels are checked against the threshold and added to the cluster, if they pass. In the second iteration, the process is repeated for each neighbor of the new cluster members. After 5 such iterations the process is truncated and the full cluster established. The clustering algorithm is illustrated in figure 9.1.

The Hough matrix is periodic in θ, this periodicity is built into the clustering algorithm, such that clusters are allowed to wrap around the boundaries of the matrix. This periodicity does not hold for the omega-dimension and thus clusters are allowed to extend only to the upper and lower boundaries of omega. Cluster expansion is stopped if the cluster extends to the edge of the Hough matrix, but the cluster is still kept. This leads to imprecisions in the track parameter resolution for curved low-momentum tracks (starting from track radii smaller than about 31cm corresponding to roughly 125MeV). Tracks with such high curvature are rare and the loss in efficiency in test samples lies between 0.1% and 1% and therefore negligible. In the case of models featuring particularly soft tracks, increasing the maximal allowed track curvature is no burden.
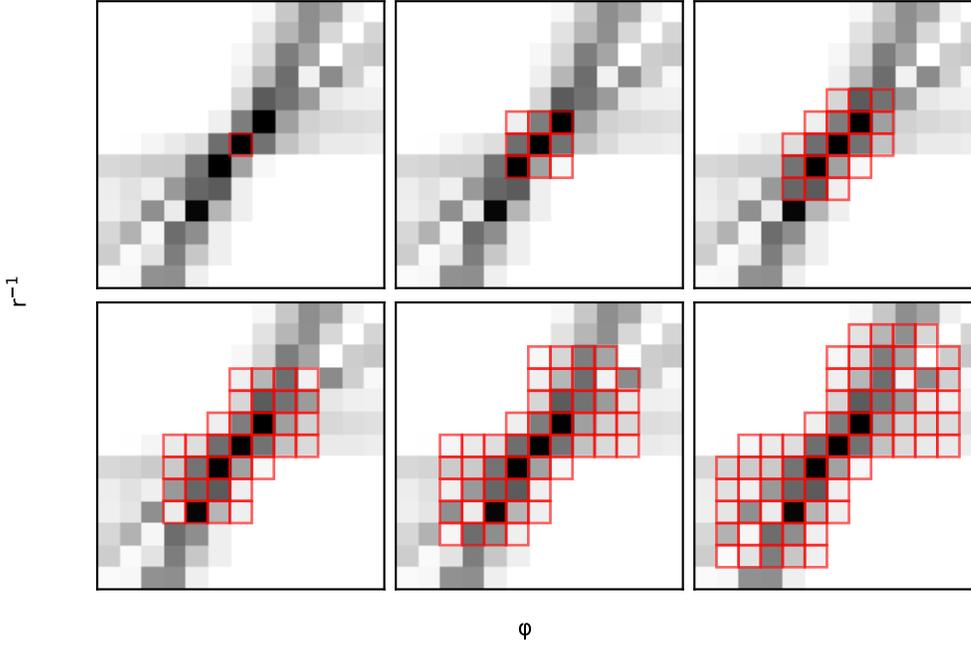
Figure 9.1: 5-step iterative clustering algorithm in the Hough matrix. The red cells represent Hough pixels that are part of the cluster. The initial seed (top left panel) is calculated with a simple peak-finding algorithm and iteratively grown into a cluster. After 5 iterations (bottom right panel) the process is truncated.

9.2 Cluster Parameters

Using a clustering algorithm instead of a threshold has the benefit that the true position of the hit-curve intersection can be estimated to much higher accuracy. The position of the highest peak is discrete due to the binned parameter space and its resolution limited to the granularity of the Hough matrix. Calculating the center of gravity of the entire cluster, on the other hand, allows the position to take values independent from the bin width. This leads to a (Gaussian) track parameter resolution of $\sigma(\varphi) \approx 0.9°$ and $\sigma\left(\mathrm{r}^{-1}\right) \approx 0.1\mathrm{m}^{-1}$.[1] Note that the track fits used to calculate the resolutions were performed at the known MC vertices, since the parameters are very sensitive to shifts of the reference vertex. This is done to illustrate the precision of track finding with the clustering algorithm in principle. Because of the rotational symmetry in the transversal plane, the φ track parameter is of no particular interest for the neural network. The curvature, on the other hand, contains information about the track momentum and is thus a useful variable. The complete set of variables that are computed for each cluster is the following:

- **Track curvature**. Simply the $\frac{1}{r}$-coordinate of the center of mass.

---

[1]the value is given in m^{-1} because the Hough matrix is linear in curvature, the inverse of the radius

- **Extent**. Vertical extent of the cluster ($\frac{1}{r}$-direction).

- **Orientation**, defined as the sum of all cluster values in the upper right and bottom left quadrant minus the sum of the values in the upper left and bottom right quadrants.

- **Peak Height**. The largest value in the cluster.

- **Sigma**. The mean distance of cluster elements weighted by their value in the Hough matrix. Definition:

$$\sigma \equiv \frac{\sum_i w_i \times \Delta r_i}{\sum_i w_i},$$ (9.2.1)

where w_i is the value of cluster element i in the Hough matrix and Δr_i its distance from the center of mass measured in bins.

- **Heavy elements**: The fraction of elements of the cluster to have a value larger than 70% of the peak.

- **Prominence**. The difference between the largest value and the average value in the cluster.

- **Off-diagonals**. Hit curves always have a positive slope (see section 6), forming a rising diagonal shape. The off-diagonal values to the top left and bottom right of the seed are summed and divided by the peak height. If the resulting value is high, the cluster is likely a fake.

- **Charge**. It is calculated as the distance of the center of mass from the bisector of the Hough matrix with a tanh activation function.

Two further parameters are calculated that are not intrinsic properties of a single cluster, but of both clusters at the same time:

- **2D opening angle**. In the Hough plane, this parameter is particularly easy to calculate. It is simply the difference in the θ-coordinate of the two clusters. In physical space, it corresponds to

$$\alpha \equiv \arccos\left(\frac{\vec{p_1} \cdot \vec{p_2}}{\|\vec{p_1}\| \|\vec{p_2}\|}\right),$$ (9.2.2)

where $\vec{p_1}$ and $\vec{p_2}$ are the transversal (2D) momenta of the two leptons produced at the displaced vertex. Since equation 9.2.2 is symmetric under the exchange of the two momenta, it doesn't matter which particle is assigned to which momentum.

- **Occupancy**. The fraction of Hough matrix elements that contain a value larger than zero after all hit curves have been traced. For computational simplicity, the negated value is used.

The distributions for all used cluster variables are shown in the appendix in figures C.1 and C.2. Many variables discriminate tracks from fakes by exploiting the symmetry of clusters: Typical track-clusters look like slanted fans, see figure 9.1: They have a narrow waist and

wide tails, displaying point-symmetry around the intersection. Clusters that arise from random hits in the CDC generally have no symmetry and no narrow waist. The purpose of the DVT is to find two tracks from a common vertex, not single tracks. Therefore, the neural network takes as input not the cluster parameters of one track candidate, but of two, allowing it to learn relations between the two tracks, such as charge conservation and the opening angle. Analysed individually, two tracks with the same sign of curvature could be perfectly fine, but they can't be a vertex candidate as charge conservation would be violated. Thus, if the neural network is fed the charges of both tracks, it can reject vertices based on track curvature. Furthermore, fake tracks are often created by the hit curves of true tracks spilling onto smaller peaks in the Hough matrix, artificially increasing the peak height. These fake track clusters have a typical position relative to the original cluster, leading to characteristic opening angles. The distribution of opening angles of fake tracks is peaked at this characteristic value and offers a good parameter for discrimination: Including the opening angle as cluster parameter allows the network to learn what artefacts look like. Of course, some iDM models produce lepton vertices with opening angles that fall into this typical range of opening angles from fakes, resulting in an increased risk of wrongly classifying true tracks. Therefore, the efficiency of the DVT with respect to the true opening angle is carefully studied.

Although it is not used in the present configuration of the DVT, another very powerful tool for fake-rejection can be employed by exploiting the pointing nature of signal vertices. In some dark matter models, most notably in the scenario of dark Higgsstrahlung used for testing the DVT, the displaced vertex is the product of a two-body decay with no loss of visible energy. In such decay topologies, the vector formed by the sum of the track momenta (i.e. the thrust vector) of the two daughter particles always points exactly away from the IP. The same is obviously not true for the vast majority of fake tracks. In principle, the scalar product of the thrust vector and the vector connecting the IP to the reference vertex represents a strongly discriminating parameter. However, if there are dark matter particles in the final state, the pointing quality of the displaced vertex is lost due to missing momentum. It is precisely because of this reason that this parameter is not utilized. Nevertheless, should the fake rate of the DVT turn out to be unacceptably high, there is always the possibility of switching on thrust-based discrimination, greatly decreasing fake rates but reducing the sensitivity to models that do not exhibit pointing vertices.

9.3 Neural Network

With ever increasing computing capabilities over the last couple of decades, neural networks (NN) have become a very powerful tool of modern data analysis. If there is a deterministic relation between the input set and the target value of a data sample, neural networks can be trained to approximate this relation even when it does not follow any obvious behavior. As the name suggests, (artificial) neural networks are based on the biological neural networks of the human brain. Individual neurons, artificial and biological alike, behave according to very simple rules, but the vast amount of neurons makes it possible to approximate complex relations. A data set that is known to contain a pattern that is too complex to extract analytically is suited for analysis using neural networks.

In 1957, Andrey Kolmogorov formulated a theorem now known as the Kolmogorov-

Arnold Superposition Principle [48] which states that any multivariate continuous function can be written as a superposition of functions of only one variable each. This theorem has strong implications for neural networks, as it shows that they can be used to approximate any continuous function, though with unknown activation functions. Based on the superposition theorem, is was later showed explicitly that even with an arbitrary squashing function and with as few as one hidden layer, any measurable function can be approximated to arbitrary precision, depending only on the number of hidden layer nodes used [49]. Therefore, by using a neural network with one hidden layer, the latency of the calculation can be minimized while maintaining the universal approximation property.

9.3.1 Architecture

The neural network of the DVT is a multi-layer perceptron (MLP) with fully connected nodes and one hidden layer. The number of input layer nodes is equal to the number of cluster parameters (20) while the hidden layer is chosen to be twice this size. There is only one output node which represents the vertex score given by the MLP. Adding one bias node for each non-input node leads to a total number of network parameters equal to

$$N_{\text{network}} = N_{\text{input}} * N_{\text{hidden}} + N_{\text{hidden}} + N_{\text{biases}} = 881. \qquad (9.3.1)$$

The ReLU-function is chosen as the activation function for hidden nodes and a Sigmoid function for the output node in order to compactify the domain of the output, allowing a probabilistic interpretation of the output value.

9.3.2 Training and Data

For large neural networks there is always the risk of overfitting its free parameters to a specific data sample. When this happens, the network learns the data by heart including random variations therein, rather than learning the underlying structure. Trying to classify new data with an overfitted network yields bad results even if the sample is similar to the original training data, since the random variations in the new sample cannot be expected to be the same. Overfitting happens if the number of free parameters in the neural network approaches the number of training examples. Two measures are taken in order to avoid overfitting:

- The example data are split into one set specifically designated for training and one set only used for validation. The validation sample is used to monitor the network performance and never to update the weights. 20% of examples are used as validation sample, the remaining 80% for training. The entire DVT-algorithm is tested exclusively on samples used neither as training nor as validation.

- The free parameters of the neural network are vastly outnumbered by the amount of training examples. As a rule of thumb, at least 10 times more examples than free parameters should be the minimum. Depending on the background conditions, the data samples used for DVT-training contained between 80,000 and 150,000 examples for 881 network parameters. The condition is therefore met by a comfortable margin.

Training samples are taken from simulation. The true position of the decay vertex is therefore always known. Signal examples of two-lepton vertices are acquired by running a Hough transform with respect to the specific MacroCell that lies closest to the true vertex. If two tracks are found, their cluster parameters are written to a data file. However, large displacements and extreme polar angles often cause tracks to have very few to no hits, making a correct reconstruction for all events in a sample impossible. In order to retain a clean example set for training, the track parameters obtained in the Hough transform are first checked against the values in the Monte Carlo simulation. Only if the centers of mass of both clusters lie within a 4x3 box (θ bins $\times \frac{1}{r}$ bins) with the theoretical intersections are the parameters kept.

Background examples are obtained from the same samples, but stripped of their signal hits. The result are samples of pure beam background, including occasional background tracks. The weighted Hough transform is performed for each beam background event and the cluster parameters calculated from the 10 vertex candidates are written out as background examples. Since the occurrence of two background tracks in the same event is extremely low, the second cluster is almost guaranteed to be a fake track. Therefore, the neural networks are trained specifically to reject fake tracks.

The PyTorch library [50] is used for training, using the ADAM optimizer for weight updating. ADAM calculates the direction of steepest descent on the error manifold and takes into account also momentum and higher-order parameters. Further, input values are normalized to a [0,1]-interval and classes kept equal in size for fast convergence. PyTorch also provides a system of optimized weight-initialization that is based on the variance of the input parameters. The data are subdivided into batches of 32 examples each, and a standard learning rate of 10^{-5} is used. As is usual with binary classifiers, the binary cross-entropy is used as a loss function to be optimized. The NP3 data set contained a total of 1.1×10^5 examples divided into a set of 8.8×10^4 for training and 2.2×10^4 for validation. After 10.000 learning epochs, the error converges to approximately 0.08 (the average discrepancy of the target and the predicted value) with a validation accuracy of roughly 97%. The networks are tested directly as part of the complete DVT algorithm with a completely independent data set containing 1.2×10^5 example events.

The upper and lower bounds of the normalizations are summarized in table 9.1. The importance of all parameters is also given for the nominal-phase network, calculated as follows: In a sample of example clusters, all instances of a given parameter are replaced by a random number between 0 and 1, while the rest of the parameters are left untouched. The loss function of the resulting network estimates is compared to the loss obtained without replacing any value. The drop (in percent) of the loss is defined as the importance.

parameter	N.	lower bound	upper bound	importance [%]	example
$\frac{1}{r}$-COM	1	0	40	390	25.6
$\frac{1}{r}$-COM	2	0	40	230	-
$\frac{1}{r}$-Extent	1	4	11	45	11
$\frac{1}{r}$-Extent	2	4	11	15	-
Orientation	1	0	50	45	46
Orientation	2	0	50	20	-
Peak value	1	7	40	930	94.9
Peak value	2	12	35	220	-
Sigma	1	20	120	10	116.5
Sigma	2	20	120	10	-
Heavy elements	1	0	0.6	75	0.04
Heavy elements	2	0	0.6	45	-
Prominence	1	5	30	440	76.1
Prominence	2	4	30	380	-
Off-diagonals	1	0	0.3	25	0.01
Off-diagonals	2	0	0.4	35	-
Charge	1	-1	1	135	-1
Charge	2	-1	1	95	-
Occupancy	both	0.1	0.5	230	-
Opening angle	both	0	$\pi(3.14)$	75	-

Table 9.1: Normalizations and importances of the relevant cluster parameters used for nominal phase 3. Most parameters come in pairs, one for each cluster (hence the second column). The last column shows the values of the example shown in figure 9.1.

Chapter 10

Studies and Results

In order to evaluate the performance of the DVT-algorithm, it is tested on data samples of dark Higgs decays with different model parameters. Furthermore, the DVT is applied to pure beam background events in order to determine its background rejection capabilities. This procedure is carried out for background levels corresponding to early phase 3 and nominal phase 3 data taking.
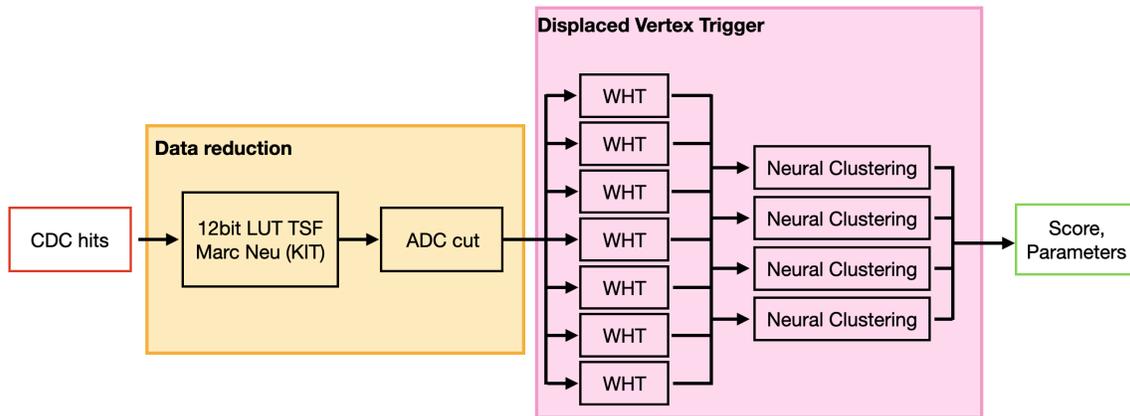
Figure 10.1: Dataflow diagram of the DVT.

Figure 10.1 illustrates the data flow of the complete DVT algorithm including the data reduction that precedes it. In summary: The LUT-based TS-finder separates signal hits from the background according to spatial correlations. Subsequently, an ADC cut between 10 and 150 counts is applied, further cleaning the picture. In parallel, weighted Hough transforms are applied to the hits from all MacroCells, narrowing down the number of reference vertices to 10 candidates. Finally, the candidates are analysed with a neural clustering algorithm that assigns a score between 0 and 1 to each vertex, based on the likelihood of the reference vertex being the production vertex of two charged tracks. If a MacroCell surpasses a certain neural threshold, the trigger fires and vertex position as well as rough estimates of the track parameters are outputted. A summary of all DVT-parameters is given in table 10.1.

For both levels of background conditions an individual neural network is trained. The

name	value	note
ADC cut	[10,150]	-
TS-finder patterns	50	Accuracy metric
Hough matrix size	160x40 bins	-
Minimal track radius	25cm	-
Peak-finding threshold	12	-
Clustering threshold	1	-
Number of vertex candidates	[6, 4]	[type A, type B]
Size of distance mask	[7, 12] bins	$[\theta, \frac{1}{r}]$
Clustering iterations	5	-
Hough weights (pre.)	1.1,3.9,3.0,2.2,1.4,1.4,1.0,0.4	in 10cm bins
Hough weights (nn.)	2.0,3.3,3.2,2.5,1.9,1.5,1.3,0.5	-
Number of peaks	2	-

Table 10.1: Summary of all MacroCell parameters. The abbreviations *pre* and *nn* stand for preselection and neural clustering, respectively. The weight configurations are different for those two cases.

network can be retrained post-implementation an thus adapt to changing background conditions. The cluster parameters remain the same, though their normalizations may differ. For a list of cluster parameters and their normalization values, see table 9.1.

10.1 Data Samples

The events were generated by Patrick Ecker at KIT using MadGraph5 version 3.3.1 [51] and simulated and reconstructed with BASF2 release 06-00-07. The background hits are overlaid onto the signal from dedicated Belle II background campaigns. For signal, the underlying model is that of inelastic dark matter featuring a dark Higgs boson h' in bins of mass. The decay path is the following:

$$e^+e^- \to A'h' \quad A' \to \chi_2\chi_1 \quad h' \to \mu^+\mu^- \tag{10.1.1}$$

The dark photon A' has a mass of 4GeV and is thus produced on-shell, decaying promptly to a dark matter doublet. The χ_2 then transitions to the lighter DM-particle (χ_1) and a pair of leptons. The reconstruction of this secondary reaction, however, is not addressed and only events in which it happens outside of the detector are analysed. The displaced tracks of interest are produced by the pointing vertex that corresponds to the reaction $h' \to \mu^+\mu^-$. Muons are preferred as final state as the dark Higgs couples to mass and therefore favors decays to heavier particles. Electrons are suppressed with respect to muons by a factor

$$\frac{\mathcal{B}(h' \to \mu^+\mu^-)}{\mathcal{B}(h' \to e^+e^-)} \propto \left(\frac{m_\mu}{m_e}\right)^2 \approx 4.28 \times 10^4. \tag{10.1.2}$$

Muons are also interesting because they produce no ECL clusters and seldom reach the KLM owing to their low momenta. The only system that reliably detects them is therefore the CDC. Since the decay times are chosen to be flat and the DM doublet is not considered,

only the dark Higgs and dark photon masses are relevant parameters and the kinematics are fixed.

The momentum-spectrum of the dark Higgs is delta-peaked because initial state radiation is excluded and the dark photon is produced on-shell, fixing all kinematic quantities. The dark Higgs itself decays into two muons isotropically in its rest frame. This represents the simplest displaced decay topology. However, since the DVT only works in two dimensions, the distributions of the effective two-dimensional quantities of opening angle and Higgs momentum are smeared. The decay lengths in the data samples are chosen to be flat, i.e. every (3D) displacement is equally likely, rather than the exponential decay lengths expected from a physical model. The trigger must work on a vast spectrum of models with different parameters, so training and testing is done in equal bins of displacement. The kinematics change for different Higgs masses. Lighter Higgs particles lead to larger boosts, resulting in on average low final state momenta and opening angles that are small and narrow-peaked. Conversely, heavy Higgses release much energy in their decay to muons, such the momenta and the opening angles are large. These kinematic differences are shown in figure 10.2 and affect the DVT performance.
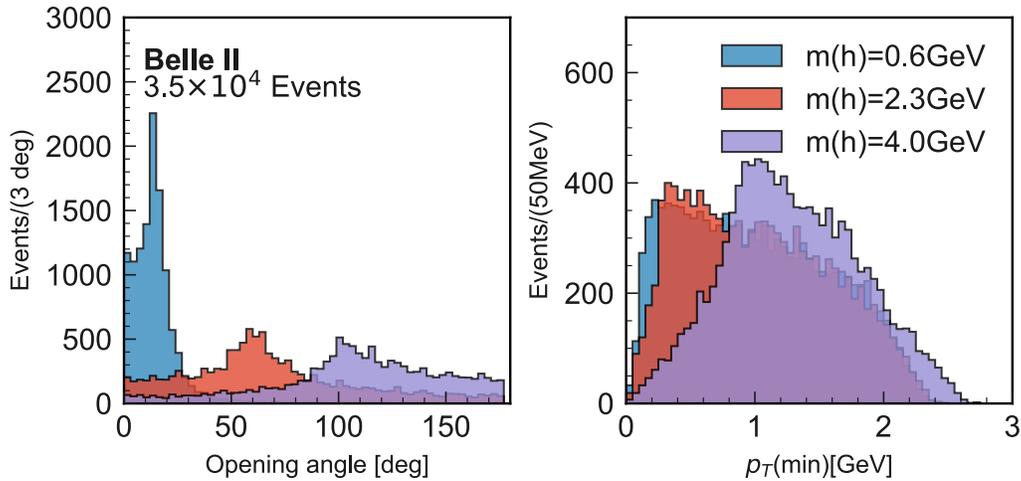
Figure 10.2: Opening angle (left) and minimum transverse momentum (right) for 3 selected dark Higgs masses. The DVT performance strongly depends on these kinematic variables.

Half of the data samples was reserved for training (odd numbered Higgs masses), the other half is used to test the algorithm. For early phase 3, 18 samples of approximately 12k events each are analysed, in bins of dark Higgs mass, ranging from 0.6GeV to 4.0GeV. Due to the drastically larger data files, for nominal phase 3 the range was reduced to $1.2\mathrm{GeV} - 4.0\mathrm{GeV}$. Furthermore, equally many beam background events are produced from the same data samples by removing all hits with a relation to a MC-particle. Since background hits are obtained separately and overlaid to signal, this procedure is unbiased and produces pure beam background samples.

10.2 Rejection Rate and Efficiency

The output distribution of the neural network is shown in figure 10.3. The neural threshold is essentially a free parameter to be chosen according to the required background rejection rate. The higher it is the more strict the condition for the trigger, leading to lower fake rates at the expense of efficiency. As a side effect, increasing the threshold also increases track parameter and vertex resolution as the neural network assigns higher scores to clean events. For the purposes of this thesis, a fake rate of 1kHz is targeted, although this number is somewhat arbitrary and open for change, depending on the hardware limitations and the rates of concurring active trigger bits. The neural threshold is an elegant solution to controlling trigger rates, avoiding to prescale the trigger. It has been an ongoing challenge to simulate beam background reliably, including background tracks, ADC distributions and event timing. The frequency of pure beam background events is hard to determine but at the same time sensitively affects how open the DVT is allowed to be.
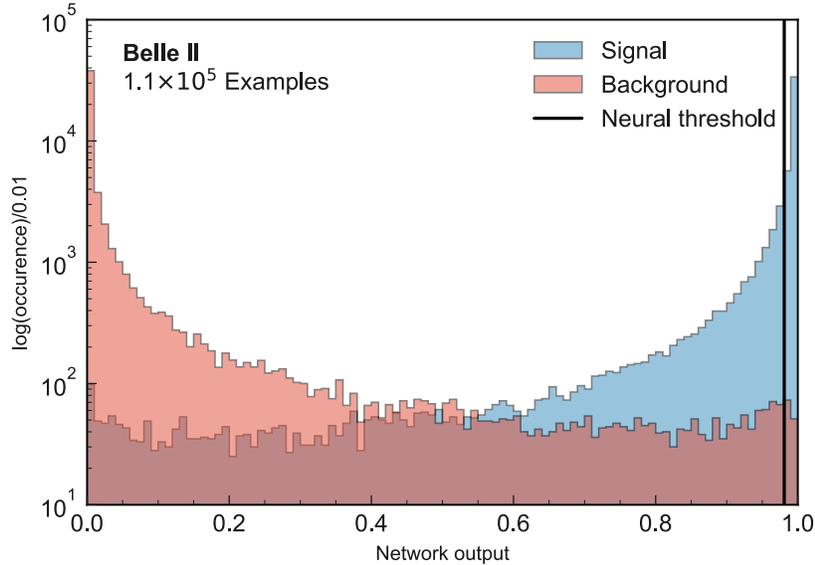
Figure 10.3: Output distribution of the neural network specialized on NP3 data. The neural threshold (black line) at 0.98 determines the trigger decision and is set by the background rejection condition.

The detector is not always live during a running period. Due to very high observed backgrounds shortly after beam injections, an injection veto of the order 10^{-3}s is applied, shortening the live time of the detector. Furthermore, after each trigger decision, the data registers must be cleared before being able to record new events, adding a recovery time of 500ns after each trigger decision. The time during which a trigger is stopped from capturing an event is called dead time, and it is planned to not exceed around 5% at nominal phase 3 running conditions at a trigger rate of 30kHz. Triggers are independent from each other and their rates are orders of magnitude slower than the rates at which events take place at

Belle II. They follow Poisson statistics, governed by the distribution

$$P(n, \lambda) = \frac{1}{n!} \lambda^n \exp(-\lambda), \tag{10.2.1}$$

which describes the probability of n events occurring in a given time interval, in which, on average, λ such events would be expected. Applied to Belle II, n is the number of trigger decisions and $\lambda = \Delta t \times R$ the expected number calculated from the mean trigger rate (R). The dead time can be calculated from the Poisson distribution as the probability of 2 or more triggers firing during a recovery time. As an example, the dead time resulting from two or more trigger decisions occurring during one recovery time can be calculated as:

$$\begin{aligned} P(n \geq 2) &= 1 - P(0) - P(1) \\ &= 1 - \exp(-\lambda) - \lambda \exp(-\lambda)|_{\lambda=0.015} \\ &\approx 0.011\%, \end{aligned} \tag{10.2.2}$$

with the expected average number of triggers during the recovery time of $\lambda = 500\text{ns} \times 30\text{kHz} = 0.015$. The figure is orders of magnitude smaller than the anticipated dead time of 5%. Clearly, the true dead time is dominated by the injection veto. Calculating the mean interval between events numerically using the same approach but fixing the dead time to 5% yields an approximate time of $12\mu s$ which leads to a maximal trigger rate of 80kHz. Background studies from simulation and early running suggest a slightly more pessimistic value of $\mathcal{O}(100\text{kHz})$ [47]. For the present thesis, a conservative background rate of 100kHz is assumed to evaluate the trigger performance, requiring a background rejection of 99% in order to reach the desired fake rate of 1kHz. The neural threshold is fine-tuned to reach this rejection rate, setting the efficiency.

Before evaluating it, the definition of efficiency used for the purposes of the DVT needs to be discussed. In the physical sense, efficiency is simply the fraction of displaced decays that the DVT triggers. However, this kind of efficiency is not very useful for a technical evaluation, since the data sample itself - with its flat dark-matter lifetimes - is not physical. In approximately 10% of events at least one of the tracks originating from the displaced vertex exits the CDC without leaving a single hit, as illustrated in figure C.4. Large displacements or extreme polar angles can produce tracks without any hits - triggering on such an event is impossible. Measuring the efficiency on these data samples in the physical sense does not reflect the efficiency obtained on physical events. Therefore, the chosen convention is to measure efficiency in a technical sense - as the fraction of events that are triggered out of the subset of events that *can* be triggered in principle. These are events in which both lepton tracks present 5 or more axial hits. This way the performance of the trigger system can be precisely evaluated, decoupling it from the hit distributions and from the efficiency of the TS-finder.

10.3   Results

Under early phase 3 and nominal phase 3 background conditions, the DVT achieves a mean efficiency of 81% and 35%, respectively, averaged over all available samples and all

displacements. A total of 1.7×10^5 signal and 2.0×10^5 pure background events with EP3 background conditions as well as 1.2×10^5 signal and 1.7×10^5 background events with full-luminosity backgrounds were used for the full calculation. The effect of the neural threshold on efficiency and background rejection is shown in figure 10.4. The performance on early
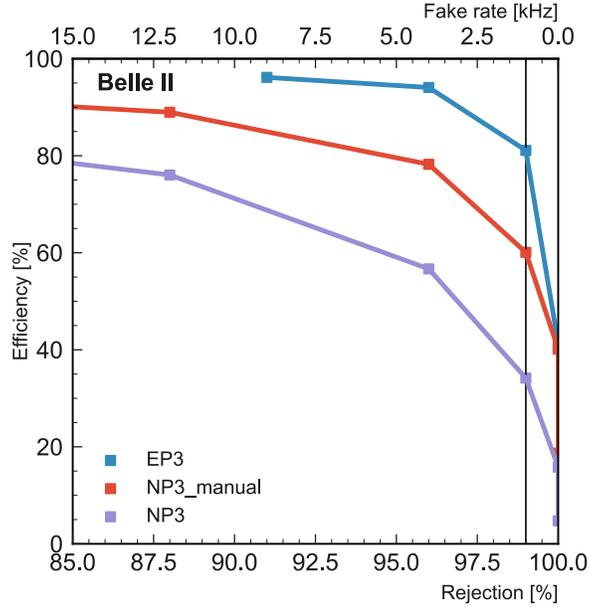
Figure 10.4: Efficiency against background rejection of the DVT under variation of the neural threshold. Curves are shown for different background conditions as well as the value corresponding to a fake rate of 1kHz (vertical line).

phase 3 is substantially better than on nominal phase 3 data, owing to the much larger amount of background tracks in the high luminosity sample. For the low-background sample (blue curve), the TS-finder works so well that in most pure background events, not enough hits are present to even trigger any vertex candidates in the preselection step. Therefore, the background rejection never dips below 90%, not even if the neural threshold is set to 0, explaining the end of the EP3-curve. For the case of NP3 (purple), this stopping point is much lower and with the neural threshold set to 0, the rejection nears 20%. Evidently, with a similar discriminatory power of the two neural nets, two very different end results are obtained.

An almost linear correlation exists between the efficiency and the number of signal hits on the weaker of the two tracks, i.e. the track that exhibits fewer hits. Even for NP3 samples, the efficiency improves to 60% if both tracks have more than 15 signal hits (equivalent to roughly 3 axial superlayers) after preprocessing by the TS-finder. Evidently, the DVT performance is very sensitively dependent on the signal efficiency of the TS-finder and suffers from the problem described in appendix B.1. As a rough approximation of the degree to which this issue affects the overall performance, a cross-check is performed on modified data with a manual background overlay, described in detail in appendix B.2. An increase in efficiency of 25% is observed (red curve), confirming the hypothesis that issues

in simulation of NP3 samples indeed impact efficiency significantly.
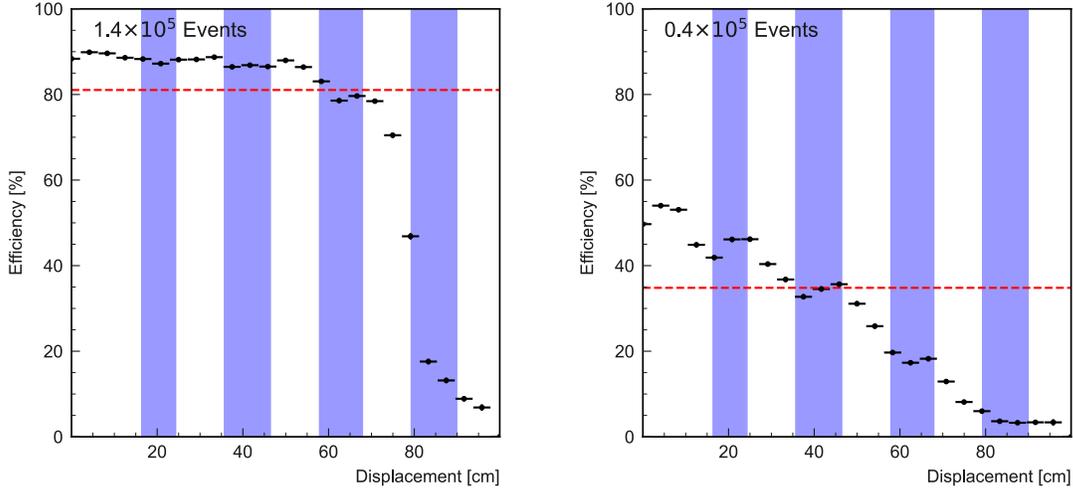
Figure 10.5: Efficiency with respect to the transversal vertex displacement for early phase 3 (left) and nominal phase 3 (right). The radial positions of the axial superlayers are highlighted in light blue and the overall efficiencies marked with the dashed red line.

The reported technical efficiencies do not reflect the physical efficiencies expected for realistic models with exponential lifetimes. They depend on the performance of the trigger with respect to transversal displacement, shown in figure 10.5, averaged over all data samples. Transversal displacement is defined as

$$\rho \equiv \sqrt{x_{\text{vertex}}^2 + y_{\text{vertex}}^2}, \tag{10.3.1}$$

ignoring the z-dimension. Especially for large displacements, the efficiency is significantly improved with respect to the STT, shown in figure 5.3. For nominal phase 3, the neural threshold must be very strict in order to obtain the assumed allowed rejection rate of 1kHz. Therefore, the triggered signal events are very precise and patterns in the efficiency crystallize that are washed out in early phase 3 samples. For example, with high backgrounds, a significant oscillatory behavior of the efficiency consistent with the position of the axial layers can be observed. Efficiency is worse at the inner edge of axial superlayers and best and the outer edge. Since the DVT sees only axial wires, hits lie very close to the best MacroCell only if the decay vertex happens to be located inside an axial superlayer. As outlined in section 8.2, close hits are problematic and generally cause a bad performance. Therefore, it is to be expected that the efficiency is higher in stereo superlayers and lower in axial layers. This is supported by the data. Further, regardless of the background conditions, a drop in efficiency is observed for displacements beyond approximately 70cm consistent with the outer edge of the third axial superlayer (67.9cm). At these distances from the IP, only two more axial layers are left until the edge of the CDC, corresponding to 10 active wires. In most cases, these are too few to significantly elevate a track from the beam background.
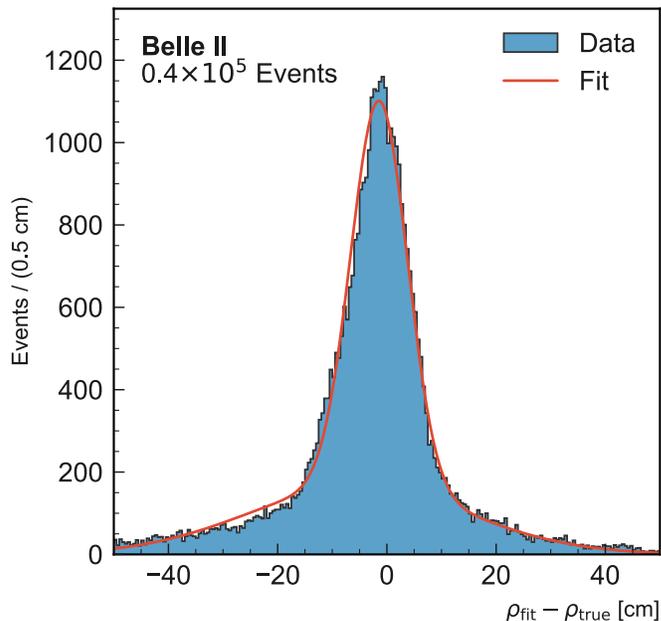
Figure 10.6: Vertex resolution of the DVT under high-luminosity background conditions: The true displacement is subtracted from the estimated value and a double Gaussian is fitted to the resulting distribution. The inner Gaussian has a width of 5.2cm.

In order to obtain the vertex resolution, the discrepancy between the estimated and the true displacement is computed according to $\rho_{\mathrm{fit}} - \rho_{\mathrm{true}}$ and subsequently fitted using a double Gaussian, as illustrated in figure 10.6. The resolution is defined as the width of the inner of the two Gaussians. Due to the stringent neural threshold, the resolution on NP3 data is actually improved with respect to EP3: The resulting values are 5.6cm for EP3 and 5.2cm for NP3 and therefore roughly equal to half the distance between two MacroCells.

Since the model parameters of possible iDM-states are not known, potential sources of bias towards certain kinds of kinematics are investigated. As explained in chapter 9.2, the 2D opening angle of the two tracks is used as an input parameter for the neural networks in order to reject certain kinds of fake tracks. It must be excluded that signal events are rejected due to their opening angle. Therefore, the efficiency is plotted against the (2D) lepton opening angle of signal tracks in figure 10.7 (following the definition given in equation 9.2.2) for high-luminosity backgrounds data samples. The efficiency is observed to remain stable across the entire range of opening angles, apart from a drop for values below approximately 30° accompanied by a strongly worsened vertex resolution. No evidence for any bias towards specific opening angles is observed. The opening angle distribution (given in figure C.1 in the supplementary materials) of the data used for network training does not exhibit patterns consistent with the variations observed in figure 9.2.2. The drop at small opening angles instead is well understood, as the Hough matrix clusters of two close tracks are frequently misidentified as one (see chapter 8.3.1), decreasing the efficiency at the
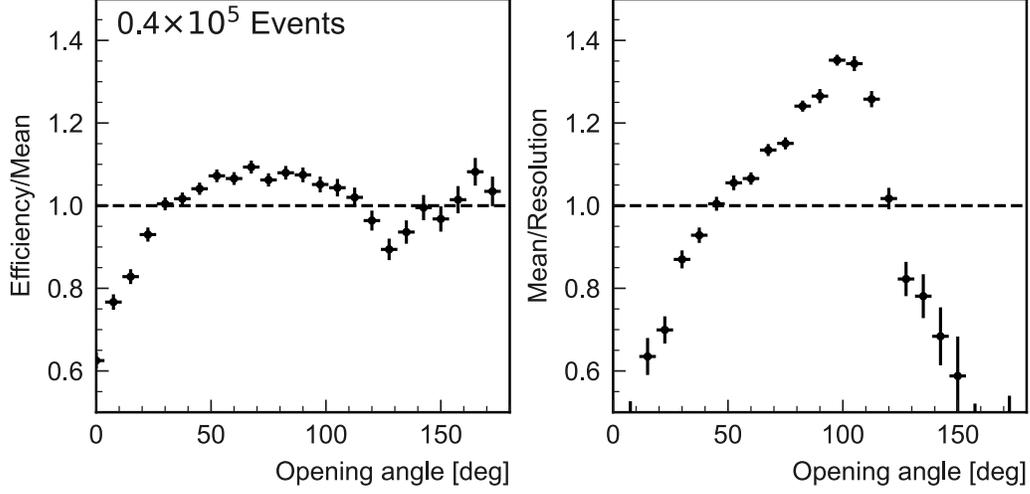
Figure 10.7: Relative change in efficiency (left) and resolution (right) of the DVT with respect to the opening angle (alpha) obtained on NP3 data. Note: Since small values of resolution are desirable, the convention is flipped in the resolution plot. Therefore, a value of 1.4 means an over-performance of 40%.

best MacroCell. However, fitted from a reference vertex somewhat shifted along the track momentum, the same tracks may appear to diverge more, evading the two-cluster resolution limit at the expense of vertex fit precision. In such events the vertex displacement tends to be overestimated, explaining the decrease in resolution. On the other extreme of opening angles approaching 180°, the drop in efficiency is not as severe, though it becomes more challenging to find the true origin of the tracks, also degrading the resolution.

Only a small dependency of efficiency on minimum transverse momentum is observed with a slight drop in the region below 0.5GeV (plot shown in the supplementary material C.3). The stability of the trigger performance for tracks across all kinematically available regions justifies the use of kinematic variables such as opening angle and track curvature as input for the neural network. No overfitting to any specific value is present. Recalling the distributions of kinematic variables of different dark Higgs masses given in figure 10.2, it is expected for the efficiency to not show any significant variation among different dark Higgs masses. Due to small transversal opening angles being strongly favored in models with light Higgses, the resolution is expected to decrease with the mass. This behavior is indeed confirmed by data and shown in figure 10.8.

Lastly, although no parameter in the network input makes use of the pointing nature of vertices, a further cross-check is performed in order to make sure no overfitting to pointing vertices occurs, given the fact that the training sample consists of exclusively two-body decays. After examining the kinematic variables of triggered fake tracks, no bias towards pointing vertices is observed. On the contrary, fake tracks are actually more likely to point back into the CDC, likely due to two reasons: Firstly, as seen from a displaced reference vertex, only a small portion of the CDC remains in the region pointing away from the IP while the entire opposing hemisphere of the CDC lies in the same direction as the
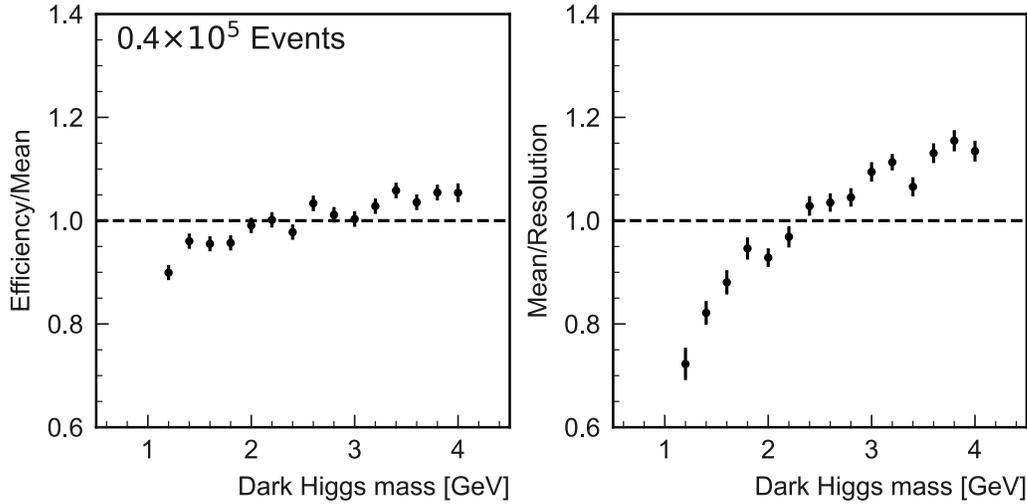
Figure 10.8: DVT efficiency and resolution for different dark Higgs masses. Full luminosity backgrounds are used.

IP. Therefore, the probability to find fake tracks is simply higher looking towards the IP rather than away from it. Secondly, notwithstanding all measures taken to counteract this problem, distant hits are still sometimes favored over close hits (see section 8.2), creating a bias towards the MacroCell at the furthest possible distance from an arrangement of hits. The fake tracks produced by this phenomenon are often exactly symmetric with respect to the axis pointing back to the IP, mimicking back-scattering with a thrust vector exactly pointing towards the IP. In short, the neural network did not learn to discriminate track candidates according to the thrust vector and the performance is expected to remain unchanged in the case of non-pointing topologies.

While all plots are taken from fits under full-luminosity background conditions, the same general behavior is observed in the case of EP3 although attenuated due to the neural threshold being more open. Since the DVT is intended as a first-level trigger and the exact determination of the vertex position is only of secondary importance, the decrease in vertex resolution is completely acceptable.

10.4 Conclusion and Outlook

For early phase 3 running conditions, the displaced vertex trigger obtains approximately an efficiency of 80% on all available data samples and for radial displacement well outside the reach of other CDC trigger systems while keeping a maximum fake rate of 1kHz. The performance in full-luminosity running is more challenging to determine, but likely decreased by as much as 20% with respect to EP3.

The displaced vertex trigger does not come without its issues: Clusters in Hough space are very reductive as they compress information from many hits onto few cells inside a cluster. The discrimination of genuine and fake tracks on the basis of cluster parameters is a challenging task - especially considering the coarse granularity of the MacroCell grid

which can oftentimes distort track parameters. Furthermore, having almost 400 Macro-Cells also means that a background event must necessarily be rejected independently from almost 400 points. Therefore, the performance of the algorithm depends very sensitively on the TS-finder that precedes it in the trigger pipeline. In light of the issues with hit relations observed in high-backgrounds Monte Carlo samples, retraining the TS-finder on corrected data is expected to yield improved results which will likely benefit the DVT as well. Lastly, while no model dependence is observed, testing the algorithm on a broader range of kinematic parameters as well as qualitatively different topologies such as explicitly non-pointing vertices or events with multiple displaced vertices is needed. In order to give a better estimate on the performance of the DVT on real data unbiased by Monte Carlo issues, testing on $K_S \to \pi^+\pi^-$-samples should be performed.

While the end results look very promising, the demand on hardware resources of the DVT is high, mostly due to the large number of MacroCells operating in parallel, each one calculating hundreds of Hough transformations. However, because of the stability of the efficiency even for MacroCells at larger distances from the true decay vertex, decreasing the granularity is a viable option. Furthermore, the implementation of a system of priority wires to the novel TS-finder would drastically decrease the number of hits having to be Hough-transformed. Among others, these measures can be taken in order to reduce the resource demands of the complete algorithm.

The present thesis showcases a proof-of-concept that a displaced vertex trigger based on parallel Hough transforms is a promising option for the future of data taking at Belle II. With 9 months until the end of the long shutdown, there is enough time for optimization and implementation of the system, adding an important new trigger to the menu and increasing the sensitivity of the experiment to a large range of dark matter scenarios.

Appendix A

Aspect ratio of the Hough Matrix

As already established, hit curves are phase-shifted sine curves with amplitudes equal to the inverse distance between the respective hit and the reference vertex. These sine curves have a periodicity of 2π and are therefore not stretched in the x-direction. Computing the first derivative it is easily seen that the maximal slope is equal to the amplitude and happens at every zero-crossing. When a curve is displayed, the aspect ratio of the axes affects the apparent slope of the curve. Trivially, on a coordinate grid with a 1:1 aspect ratio, the displayed slope matches the mathematical one. If the grid is stretched by a factor a in y-direction, slope is amplified by a, too. Conversely, stretching the grid in x-direction by b scales down the slope by b. Slope maps in proportion to the aspect ratio of the displaying surface.

The Hough matrix has a $\theta - \frac{1}{r}$ coordinate system, thus slope has the unit $\frac{1}{\text{cm rad}}$. As table 6.1 shows, there are much more θ-bins than $\frac{1}{r}$-bins in the Hough matrix, essentially stretching hit curves by a factor of ≈ 19.6. The aspect ratio of the Hough matrix is equal to the vertical bin density divided by the horizontal bin density:

$$f_{\text{aspect}} = \frac{40\text{bins}}{0.08\text{cm}^{-1}} \div \frac{160\text{bins}}{2\pi} \approx 19.6\text{cm rad}, \tag{A.0.1}$$

where 0.08cm^{-1} is the range of track curvatures covered by the Hough matrix. A curve calculated to have a slope equal to unity appears with a displayed slope of 19.6.

Appendix B

Issues in Background Simulation

B.1 TS-finder Training Problems

Simulating background is notoriously hard due to the amount of different sources (cross-talk, synchrotron electrons, nuclear spallation etc.) that can create backgrounds with very different signatures. For the sake of highest possible fidelity, backgrounds are recorded directly at the detector in background campaigns under different running conditions. Background files are created from these campaigns to be overlaid onto signal data generated by outside programs. Because of this, it often happens that wires are hit both by signal and by background, since the two classes are created separately. However, as of now, the core framework of BASF2 does not allow for one hit to have multiple relations. BASF2 must therefore choose whether to assign a hit to background or to signal. This decision is taken under consideration of timing information about the hit and often leads to signal hits being discarded in favor of background. The result is that the number of signal hits associable to signal depends on the severity of the background overlay used in data production. The already quite low signal hit multiplicity of EP3 events (see figure C.4) is observed to be another 10% smaller in NP3 data. We believe the wire overlap to be one (though likely not the only) effect to cause this drastic decrease. From a physical point of view, there should be no difference in the number of hits caused by signal particles.

If wire overlaps are the only effect at play, there is no problem for track reconstruction, as all hits are present and located where they should be. It is only the *relation* to a signal particle that is missing. However, this effect causes a problem in the training of the TS-finder: In order to be trained, the TS-finder needs to learn what pure signal data and pure background look like. The separation of data samples into signal and background is done using the hit relation to signal particles. Therefore, in the case of NP3, the sample used for the TS-finder training is biased and contains too few signal hits. The spatial correlations of the signal hits are affected and the TS-finder can not reach the efficiency it needs to obtain.

Another problem encountered in the training of a TS-finder especially when it is dedicated to preprocessing NP3 data is the presence of frequent background tracks. Apart from different event time distributions, no physical difference between a background and a signal track exist. Training is sub-optimal if tracks appear in background files, because they are indistinguishable from signal. The ideal solution is the exclusion of background tracks from beam background files, ensuring that the TS-finder truly learns to distinguish background

from signal. This is an improvement planned in future development of the TS-finder.

B.2   Manual Background Overlay
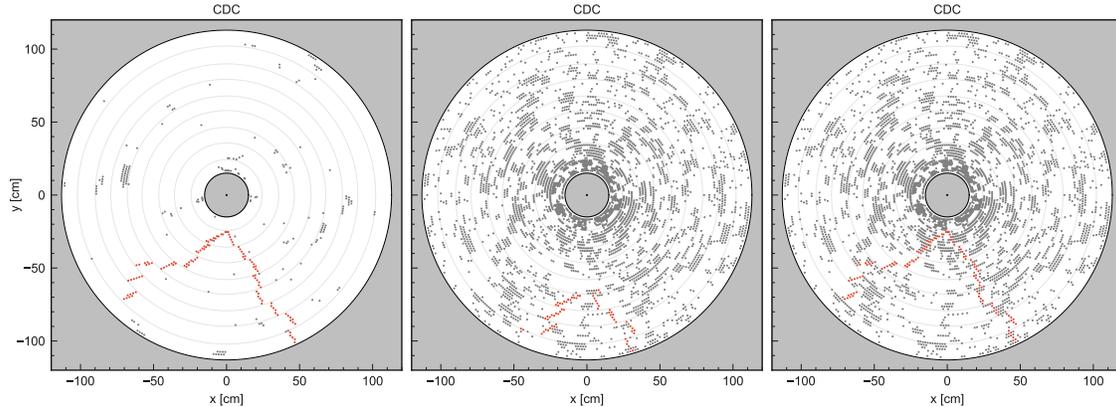
Figure B.1: Manual background overlay procedure. Signal hits from EP3 (left) are super-imposed on background hits from NP3 (center). Overlapping hits are removed, yielding a new sample (right).

As outlined in appendix B.1, the signal hit multiplicity of NP3 is significantly decreased with respect to EP3 samples. This leads to overall worse performance. Since there is no physical explanation for why the number of signal hits should be in any way affected by the background conditions, there is reason to believe that there might be a slight mismodelling of hit distributions in high-luminosity files. In order to measure the impact of this issue, a manual background overlay procedure is employed: Signal hits are extracted from an EP3 data sample and superimposed onto pure beam background, taken from a NP3 file, as illustrated in figure B.1. For each overlapping hit, the background hits are carefully removed, keeping the relation to the signal particle intact. The resulting events have therefore the signal features of EP3 data and the background conditions of NP3. No double hits are present. The signal hit multiplicity of EP3 is obtained, while the change of the background hit distribution is negligible (a reduction of the order 0.5% is observed).

The TS-finder efficiency is drastically improved with respect to the original NP3 sample: The mean signal hit multiplicity (after the ADC cut) increases from 36 to 53, reaching the same level of TS-finder efficiency as the EP3 version. Meanwhile, the number of surviving background hits is unchanged. For this reason the DVT algorithm performs significantly better, obtaining an efficiency of over 60%.

Appendix C

Supplementary Plots and Figures
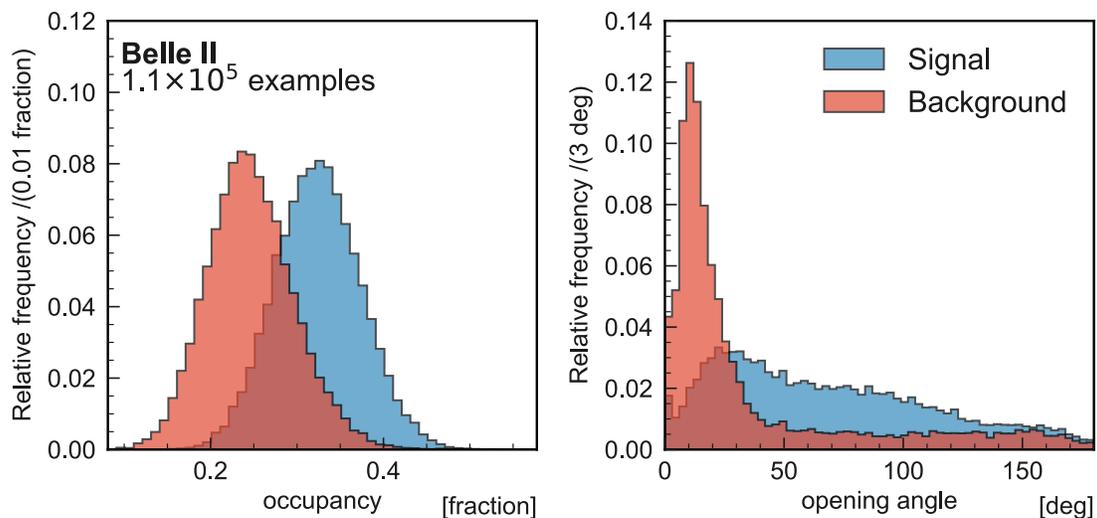
Figure C.1: Signal and background parameter distributions obtained with NP3 background conditions. The parameters shown in this plot are not calculated from individual clusters. Rather, they are shared from both clusters in an event and therefore only appear once in the network input.
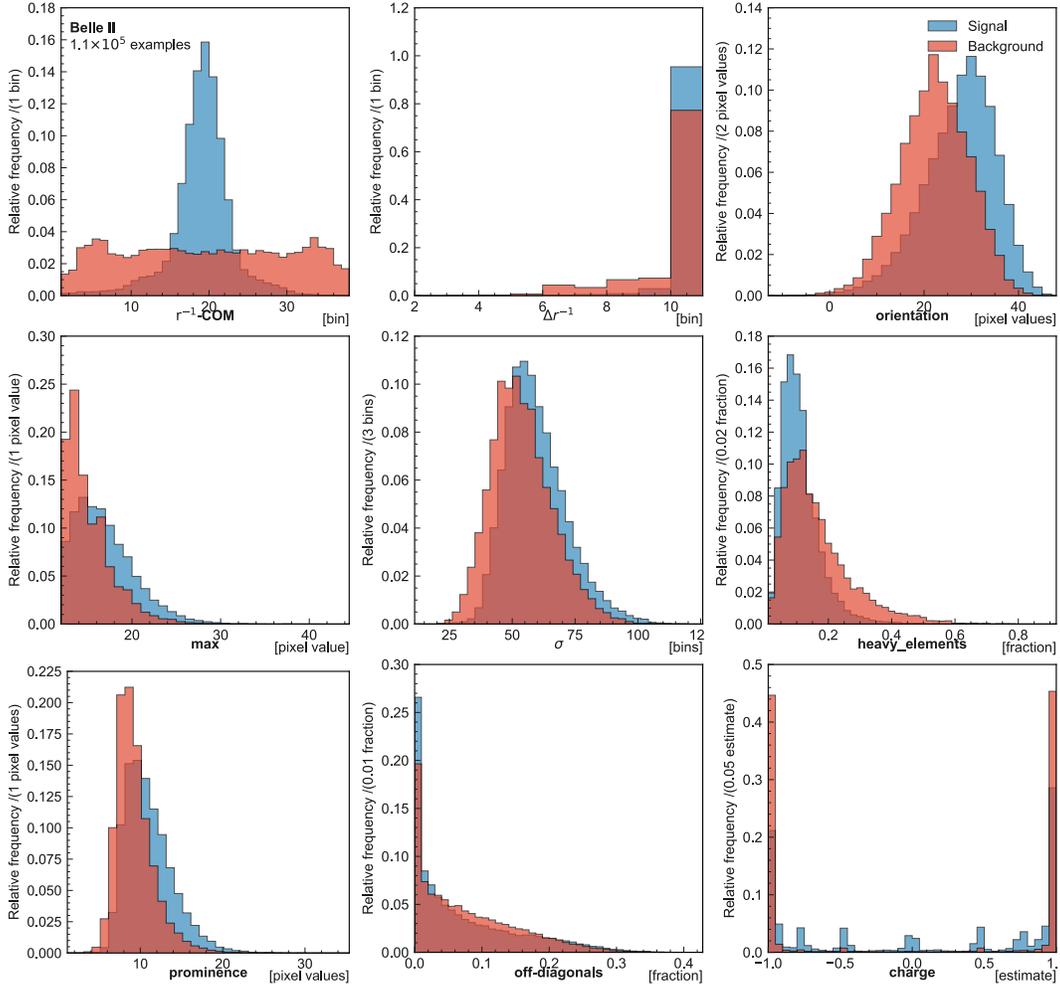
Figure C.2: Signal and background distributions of individual cluster variables obtained with NP3 background conditions. These variables enter the input array of the neural network twice, once for each cluster.
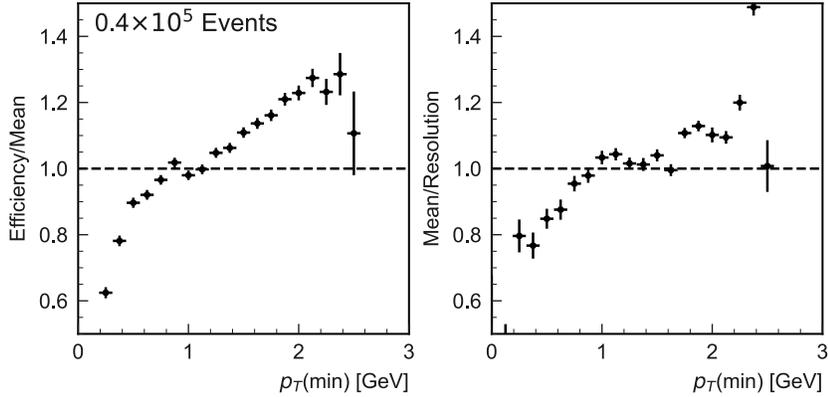
Figure C.3: Relative efficiency and resolution change in bins of minimum momentum, i.e. the momentum of the softer lepton. Approaching the limit of maximal track curvature allowed in the Hough matrix (25cm curvature, corresponding to track momenta of 100MeV), the performance deteriorates. No unexpected trend is observed.
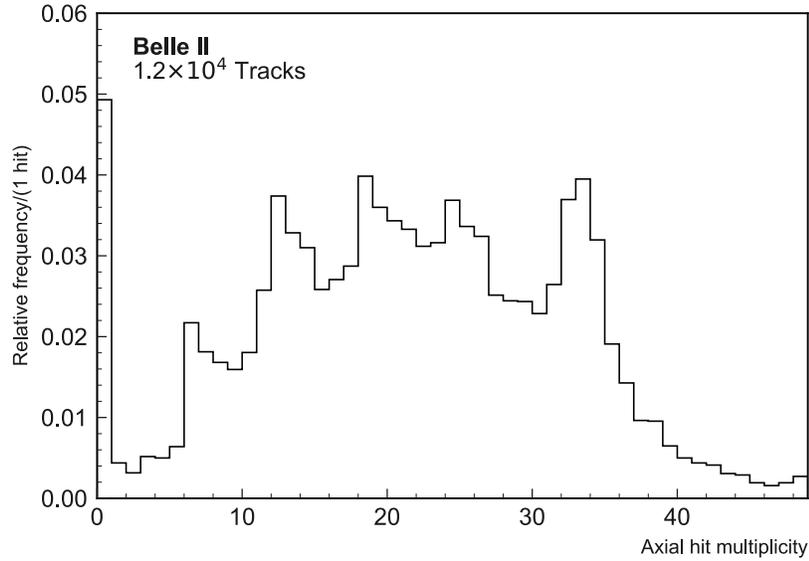
Figure C.4: Distribution of axial hit numbers per track in 12.000 dark Higgs decays in the favorable case of EP3 backgrounds. 5 peaks corresponding to 5 fully traversed superlayers are clearly visible in the distribution. However, the most frequent hit multiplicity is 0, owing to the limited CDC acceptance region combined with large displacements. Extrapolated to the two-track topology, this means that in approximately 10 percent of all events at least one track is missing.

Acknowledgements

Bibliography

[1] Fritz Zwicky. On the masses of nebulae and of clusters of nebulae. *The Astrophysical Journal*, 86(3):217–246, 1937.

[2] N. A. Bahcall, L. M. Lubin, and V. Dorman. Where is the dark matter? *The Astrophysical Journal*, 447:L81–L86, 1995.

[3] Stefano Profumo. *An Introduction to Particle Dark Matter*. World Scientific, 2017.

[4] C. K. Freeman. On the disks of spiral and s0 galaxies. *The Astrophysical Journal*, 160:811–830, 1970.

[5] Yoshiake Sofue and Vera Rubin. Rotation curves of spiral galaxies. *Annual Review of Astronomy & Astophysics*, 39:137–177, 2000.

[6] G. Helou, B. F. Madore, M. Schmitz, M. D. Bicay, X. Wu, and J. Bennett. The NASA/IPAC extragalactic database. In M. A. Albrecht and D. Egret, editors, *Databases and On-line Data in Astronomy*, volume 171 of *Astrophysics and Space Science Library*, pages 89–106, January 1991.

[7] S. W. Allen, P R. W. Schmidt, and A. C. Fabian. Cosmological constraints from the x-ray gas mass fraction in relaxed lensing clusters observed with chandra. *Mon. Not. R. Astron. Soc.*, 334:L11–L15, 2002.

[8] D. Clowe, A. Gonzales, and M. Markevitch. Weak lensing mass reconstruction of the interacting cluster 1e0657-558: Direct evidence for the existence of dark matter. *The Astrophysical Journal*, 604:596–603, 2008.

[9] V. Mukhanov. *Physical Foundations of Cosmology*. Cambridge University Press, 2005.

[10] Edward W. Kolb and Michael S. Turner. *The Early Universe*. Addison-Wesley Publishing Company, 1994.

[11] C. L. Bennett, D. Larson, J. L. Weiland, N. Jarosik, G. Hinshaw, N. Odegard, K. M. Smith, R. S. Hill, B. Gold, M. Halpern, E. Komatsu, M. R. Nolta, L. Page, D. N. Spergel, E. Wollack, J. Dunkley, A. Kogut, M. Limon, S. S. Meyer, G. S. Tucker, and E. L. Wright. Nine-year wilkinson microwave anisotropy probe (wmap) observations: Final maps and results. *The Astrophysical Journal Supplement Series*, 208(2):20, 2013.

[12] Wayne Hu, Naoshi Sugiyama, and Joseph Silk. The physics of microwave background anisotropies. *Nature*, 386(6620):37–43, 1997.

[13] Will J. Percival, Robert C. Nichol, Daniel J. Eisenstein, Joshua A. Frieman, Masataka Fukugita, Jon Loveday, Adrian C. Pope, Donald P. Schneider, Alex S. Szalay, Max Tegmark, Michael S. Vogeley, David H. Weinberg, Idit Zehavi, Neta A. Bahcall, Jon Brinkmann, Andrew J. Connolly, and Avery Meiksin. The shape of the sloan digital sky survey data release 5 galaxy power spectrum. *The Astrophysical Journal*, 657(2):645–663, 2007.

[14] Scott Dodelson. The real problem with mond. *International Journal of Modern Physics D*, 20(14):2749–2753, 2011.

[15] B. Paczynski. Gravitational Microlensing by the Galactic Halo. , 304:1, 1986.

[16] P. Tisserand, L. Le Guillou, C. Afonso, J. N. Albert, J. Andersen, R. Ansari, É. Aubourg, P. Bareyre, J. P. Beaulieu, X. Charlot, C. Coutures, R. Ferlet, P. Fouqué, J. F. Glicenstein, B. Goldman, A. Gould, D. Graff, M. Gros, J. Haissinski, C. Hamadache, J. de Kat, T. Lasserre, É. Lesquoy, C. Loup, C. Magneville, J. B. Marquette, É. Maurice, A. Maury, A. Milsztajn, M. Moniez, N. Palanque-Delabrouille, O. Perdereau, Y. R. Rahal, J. Rich, M. Spiro, A. Vidal-Madjar, and L. Vigroux and. Limits on the macho content of the galactic halo from the EROS-2 survey of the magellanic clouds. *Astronomy &amp Astrophysics*, 469(2):387–404, 2007.

[17] Leszek Roszkowski, Enrico Maria Sessolo, and Sebastian Trojanowski. WIMP dark matter candidates and searches—current status and future prospects. *Reports on Progress in Physics*, 81(6):066201, may 2018.

[18] Jason Arakawa and Tim Tait. Is a miracle-less WIMP ruled out? *SciPost Physics*, 11(2), aug 2021.

[19] Gopolang Mohlabeng. Revisiting the dark photon explanation of the muon anomalous magnetic moment. *Physical Review D*, 99(11), jun 2019.

[20] R. Essig, J. A. Jaros, W. Wester, P. Hansson Adrian, S. Andreas, T. Averett, O. Baker, B. Batell, M. Battaglieri, J. Beacham, T. Beranek, J. D. Bjorken, F. Bossi, J. R. Boyce, G. D. Cates, A. Celentano, A. S. Chou, R. Cowan, F. Curciarello, H. Davoudiasl, P. deNiverville, R. De Vita, A. Denig, R. Dharmapalan, B. Dongwi, B. Döbrich, B. Echenard, D. Espriu, S. Fegan, P. Fisher, G. B. Franklin, A. Gasparian, Y. Gershtein, M. Graham, P. W. Graham, A. Haas, A. Hatzikoutelis, M. Holtrop, I. Irastorza, E. Izaguirre, J. Jaeckel, Y. Kahn, N. Kalantarians, M. Kohl, G. Krnjaic, V. Kubarovsky, H-S. Lee, A. Lindner, A. Lobanov, W. J. Marciano, D. J. E. Marsh, T. Maruyama, D. McKeen, H. Merkel, K. Moffeit, P. Monaghan, G. Mueller, T. K. Nelson, G. R. Neil, M. Oriunno, Z. Pavlovic, S. K. Phillips, M. J. Pivovaroff, R. Poltis, M. Pospelov, S. Rajendran, J. Redondo, A. Ringwald, A. Ritz, J. Ruz, K. Saenboonruang, P. Schuster, M. Shinn, T. R. Slatyer, J. H. Steffen, S. Stepanyan, D. B. Tanner, J. Thaler, M. E. Tobar, N. Toro, A. Upadye, R. Van de Water, B. Vlahovic, J. K. Vogel, D. Walker, A. Weltman, B. Wojtsekhowski, S. Zhang, and K. Zioutas. Dark sectors and new, light, weakly-coupled particles, 2013.

[21] R. Bernabei et al. Recent Results from DAMA/LIBRA and Comparisons. *Moscow Univ. Phys. Bull.*, 77(2):291–300, 2022.

[22] R. Agnese, T. Aramaki, I. J. Arnquist, W. Baker, D. Balakishiyeva, S. Banik, D. Barker, R. Basu Thakur, D. A. Bauer, T. Binder, M. A. Bowles, P. L. Brink, R. Bunker, B. Cabrera, D. O. Caldwell, R. Calkins, C. Cartaro, D. G. Cerdeño, Y. Chang, Y. Chen, J. Cooley, B. Cornell, P. Cushman, M. Daal, P. C. F. Di Stefano, T. Doughty, E. Fascione, E. Figueroa-Feliciano, M. Fritts, G. Gerbier, R. Germond, M. Ghaith, G. L. Godfrey, S. R. Golwala, J. Hall, H. R. Harris, Z. Hong, E. W. Hoppe, L. Hsu, M. E. Huber, V. Iyer, D. Jardin, A. Jastram, C. Jena, M. H. Kelsey, A. Kennedy, A. Kubik, N. A. Kurinsky, B. Loer, E. Lopez Asamar, P. Lukens, D. MacDonell, R. Mahapatra, V. Mandic, N. Mast, E. H. Miller, N. Mirabolfathi, B. Mohanty, J. D. Morales Mendoza, J. Nelson, J. L. Orrell, S. M. Oser, K. Page, W. A. Page, R. Partridge, M. Penalver Martinez, M. Pepin, A. Phipps, S. Poudel, M. Pyle, H. Qiu, W. Rau, P. Redl, A. Reisetter, T. Reynolds, A. Roberts, A. E. Robinson, H. E. Rogers, T. Saab, B. Sadoulet, J. Sander, K. Schneck, R. W. Schnee, S. Scorza, K. Senapati, B. Serfass, D. Speller, M. Stein, J. Street, H. A. Tanaka, D. Toback, R. Underwood, A. N. Villano, B. von Krosigk, B. Welliver, J. S. Wilson, M. J. Wilson, D. H. Wright, S. Yellin, J. J. Yen, B. A. Young, X. Zhang, and X. Zhao and. Results from the super cryogenic dark matter search experiment at soudan. *Physical Review Letters*, 120(6), 2018.

[23] Dario Buttazzo, Paolo Panci, Nicola Rossi, and Alessandro Strumia. Annual modulations from secular variations: relaxing DAMA? *Journal of High Energy Physics*, 2020(4), apr 2020.

[24] J. Amaré , S. Cebrián, D. Cintas, I. Coarasa, E. García, M. Martínez, M. A. Oliván, Y. Ortigoza, A. Ortiz de Solórzano, J. Puimedón, A. Salinas, M. L. Sarsa, and P. Villar. Annual modulation results from three-year exposure of ANAIS-112. *Physical Review D*, 103(10), may 2021.

[25] the BaBar collaboration. Search for a dark photon in e^+e^- collisions at babar. *Physical Review Letters*, 113(20), nov 2014.

[26] The BaBar collaboration. Search for invisible decays of a dark photon produced in e^+e^- collisions at babar. *Physical Review Letters*, 119(13), sep 2017.

[27] Torben Ferber, Camilo Garcia-Cely, and Kai Schmidt-Hoberg. Belle II sensitivity to long–lived dark photons. *Physics Letters B*, 833:137373, 2022.

[28] David Smith and Neal Weiner. Inelastic dark matter. *Physical Review D*, 64(4), jul 2001.

[29] Eder Izaguirre, Gordan Krnjaic, and Brian Shuve. Discovering inelastic thermal relic dark matter at colliders. *Physical Review D*, 93(6), mar 2016.

[30] Asher Berlin, Nikita Blinov, Stefania Gori, Philip Schuster, and Natalia Toro. Cosmology and accelerator tests of strongly interacting dark matter. *Physical Review D*, 97(5), mar 2018.

[31] Yukiyoshi Ohnishi, Tetsuo Abe, Toshikazu Adachi, Kazunori Akai, Yasushi Arimoto, Kiyokazu Ebihara, Kazumi Egawa, John Flanagan, Hitoshi Fukuma, Yoshihiro Funakoshi, Kazuro Furukawa, Takaaki Furuya, Naoko Iida, Hiromi Iinuma, Hoitomi Ikeda, Takuya Ishibashi, Masako Iwasaki, Tatsuya Kageyama, Susumu Kamada, and Zhanguo Zong. Accelerator design at superkekb. *Progress of Theoretical and Experimental Physics*, 2013, 03 2013.

[32] T. Abe et al. Belleii technical design report, 2010.

[33] Kazunori Akai, Kazuro Furukawa, and Haruyo Koiso on behalf of the SuperKEKB accelerator team. Superkekb collider. 2018.

[34] J.W. Flanagan and Y. Ohnishi. Letter of intent for kek super b factory, part iii: Accelerator design. *Tech. rep. KEK Report 04-4*, 2004.

[35] E Kou et al. The belle II physics book. *Progress of Theoretical and Experimental Physics*, 2019(12), 2019.

[36] Sara Pohl. *Track Reconstruction at the First Level Trigger of the Belle II Experiment.* PhD thesis, Ludwig-Maximilians-Universität, Munich, 2018.

[37] Konrad Kleinknecht. *Detektoren für Teilchenstrahlung.* Teubner Studienbücher, 1992.

[38] S. Longo, J.M. Roney, C. Cecchi, S. Cunliffe, T. Ferber, H. Hayashii, C. Hearty, A. Hershenhorn, A. Kuzmin, E. Manoni, F. Meier, K. Miyabayashi, I. Nakamura, M. Remnev, A. Sibidanov, Y. Unno, Y. Usov, and V. Zhulanov. CsI(tl) pulse shape discrimination with the belle II electromagnetic calorimeter as a novel method to improve particle identification at electron–positron colliders. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 982:164562, dec 2020.

[39] Michael Duerr, Torben Ferber, Christopher Hearty, Felix Kahlhoefer, Kai Schmidt-Hoberg, and Patrick Tunney. Invisible and displaced dark matter signatures at belle II. *Journal of High Energy Physics*, 2020(2), 2020.

[40] Yoshihito Iwasaki, ByungGu Cheon, Eunil Won, Xin Gao, Luca Macchiarulo, Kurtis Nishimura, and Gary Varner. Level 1 trigger system for the Belle II experiment. *IEEE Trans. Nucl. Sci.*, 58:1807–1815, 2011.

[41] Christian Kiesling et al. Belle2-note-tr-2023-006. 2023. preliminary.

[42] T. Kuhr, C. Pulvermacher, M. Ritter, T. Hauth, and N. Braun. The belle II core software. *Computing and Software for Big Science*, 3(1), 2018.

[43] S. Agostinelli et al. Geant4—a simulation toolkit. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 506(3), 2003.

[44] Paul V.C. Hough. Method and means for recognizing complex patterns, 1962. US patent n. 3.069.654.

[45] Dr. Peter Hart. How the hough transform was invented. *IEEE Signal Processing Magazine*, 2009.

[46] Marc Neu. Data-driven design of the belle ii track segment finder for dark matter decays. Master's thesis, Karlsruhe Institute of Technology (KIT), Institute for Information Processing Technologies (ITIV), 2022.

[47] Sebastian Skambraks. *Efficient Physics Signal Selectors for the First Trigger Level of the Belle II Experiment based on Machine Learning*. PhD thesis, Ludwig-Maximilians-Universität, Munich, 2020.

[48] A. Kolmogorov. On the representation of continuous functions of many variables by superposition of continuous functions of one variable and addition. *Doklady Akademii Nauk SSSR*, 114(5):953–956, 1957.

[49] Kurt Hornik. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366, 1989.

[50] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Köpf, Edward Yang, Zach DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library, 2019.

[51] Johan Alwall, Michel Herquet, Fabio Maltoni, Olivier Mattelaer, and Tim Stelzer. MadGraph 5: going beyond. *Journal of High Energy Physics*, 2011(6), jun 2011.

Erklärung

Hiermit erkläre ich, die vorliegende Arbeit selbständig verfasst zu haben und keine anderen als die in der Arbeit angegebenen Quellen und Hilfsmittel benutzt zu haben.

München, den 11. März 2023