

Measurement of Antideuteron Production



Bachelor Thesis at the faculty of physics
at the
Ludwig-Maximilians-University Munich

submitted by
Jakob Roth
born in Hamburg

Munich, the 29th June, 2018

Supervisor:
Prof. Dr. Thomas Kuhr

Messung von Antideuteron-Produktion



Bachelorarbeit an der Fakultät für Physik
der
Ludwig-Maximilians-Universität München

vorgelegt von
Jakob Roth
geboren in Hamburg

München, den 29. Juni 2018

Betreuer:
Prof. Dr. Thomas Kuhr

Abstract

A measurement of the antideuteron production rate is an important baseline for indirect search of dark matter. The analysis presents an antideuteron identification method for the Belle II detector, which allows a production rate measurement. Thereby the necessity of an additional deuteron track hypothesis is considered. The expected statistical uncertainty of a Belle II measurement is estimated and discussed.

Contents

1	Introduction	1
1.1	Standard Model	1
1.2	Purpose of this study	1
1.3	Coalescence model	2
1.4	Previous measurements	2
1.5	Belle II experiment	4
2	Methods	7
2.1	Classification operators	7
2.2	Typical branching ratio measurements	8
2.3	Fisher’s linear discriminant	9
3	Antideuteron Identification	11
3.1	General idea of antideuteron identification	11
3.2	Simulated signal events	11
3.3	Simulated background events	12
3.4	Particle identification methods	12
3.5	Matter and antimatter detection	13
3.6	Cuts for antideuteron identification	15
3.7	Distribution of deuteronID	17
3.8	Likelihood distributions	18
3.9	Separating background an signal with Fisher’s linear discriminant	19
3.10	Deuteron track hypothesis	20

<i>CONTENTS</i>	III
4 Sensitivity Estimation	23
4.1 General assumptions	23
4.2 Statistical error of continuum measurement	23
4.3 Limits for antideuteron production in $\Upsilon(4S)$ events	24
4.4 Comparison to BaBar measurement	25
5 Conclusion and Outlook	27
5.1 Conclusion	27
5.2 Outlook	27
Bibliography	29

Chapter 1

Introduction

1.1 Standard Model

The most precise description of energy, matter and the elementary forces is given by the Standard Model. In this model the constituents of matter are the three generations of quarks and leptons. Each generation consists of two spin $1/2$ particles, quarks with charge $2/3$ and $-1/3$ and leptons, one with charge -1 and the other a neutral neutrino. The quarks form mesons and baryons, which – together with the leptons – build up the matter we observe in our everyday life. For each of these particles there exists an antiparticle, together they can annihilate. The three fundamental forces electromagnetic, strong and weak interaction are described by so-called mediator particles with spin 1.

With the Standard Model, most observations of matter and antimatter can be explained with astonishing precision. But still some aspects are missing. The majority of cosmological models postulate the existence of additional forms of matter and energy, so-called dark matter and dark energy. Although dark matter is a regular feature in cosmological models, its behavior is not understood. From a theoretical point of view, it is not clear how this additional form of matter can be included in our particle physics models. For this reason further experimental results are needed.

1.2 Purpose of this study

There are multiple approaches to dark matter observation, one of them is the indirect detection via secondary particles in dark matter self annihilation. These indirect searches are based on the kinematic differences between particles created in dark matter annihilation compared to astrophysical sources. For such searches it is important that the background of astrophysical sources is very low. It is unlikely that antimatter particles originate from

astrophysical sources because the visible universe is out of matter. Especially antideuteron has a very low background, since the threshold energy for antideuteron is higher than for antiprotons or positrons. The precise measurement of possible dark matter self-annihilation requires not only a low background but also a thorough understanding of the production of the observed particles from the fragments of the annihilation. Due to the low astrophysical background, antideuteron is a promising candidate for the search of dark matter. However, the measurements suffer from relatively large uncertainties in the antideuteron production models. The purpose of this study is to investigate how the antideuteron production rate could be measured by the Belle II experiment. A Belle II measurement could help to better understand antideuteron production and thereby increase the precision of indirect dark matter search.

1.3 Coalescence model

Mechanisms of deuteron and antideuteron formation are probably identical up to a high precision since they are connected via charge conjugation. Therefore the following paragraph refers to both deuteron and antideuteron formation. A model which is often used, although it is not proven by experiment, is the coalescence model [1]. In the coalescence model, deuteron originates from the binding of a proton and a neutron with a small relative momentum. Thereby the simplifying assumption is made that all antiprotons and antineutrons with a relative momentum less than p_0 fusion to antideuteron. This coalescence condition is applied by Monte Carlo generators on an event-by-event basis.¹ The yield of deuterons in coalescence model is fitted to measured production rates; thereby the empirical constant p_0 is determined. Thus measurements of deuteron production are needed in order to validate the coalescence model and estimate the parameter p_0 .

1.4 Previous measurements

The antideuteron production rate has already been measured by other experiments. Although the production rate for deuteron and antideuteron is assumed to be identical, experiments focus on antideuteron since antideuteron has a lower background of particles from detector material and beam pipe gas. Previous measurements for antideuteron production are listed in Table 1.1.

On the basis of these measurements, the coalescence model was tested and the coalescence parameter p_0 determined. Despite these multiple measurements, the uncertainty on p_0 is

¹Before 2009 the further assumption of isotropic and uncorrelated proton and neutron momenta was made, which allows to derive an analytic expression for the deuteron spectrum, in terms of the neutron and proton spectra. However this assumption dose not hold, since baryons are mostly produced in QCD jets [2].

Table 1.1: Previous measurements of antideuteron formation.

Experiment	Antideuteron production measurement in
ALICE [3]	p^+p^+ continuum at $\sqrt{s} = 7$ TeV
ALEPH [4]	Z decay
ZEUS [5]	e^+p^+ and e^-p^+ continuum at $\sqrt{s} \approx 318$ GeV
ISR [6]	p^+p^+ continuum at $\sqrt{s} = 53$ GeV
BaBar [7]	$\Upsilon(2S)$, $\Upsilon(3S)$ decays and e^+e^- continuum at $\sqrt{s} = 10.58$ GeV
CLEO [8]	$\Upsilon(2S)$, $\Upsilon(3S)$ decays
ARGUS [9]	$\Upsilon(2S)$, $\Upsilon(3S)$ decays

still substantial because the different experiments yield inconsistent values. A summary of the resulting values for p_0 for different Monte Carlo generators is given in Figure 1.1. Currently, it is an open question what causes this inconsistency.

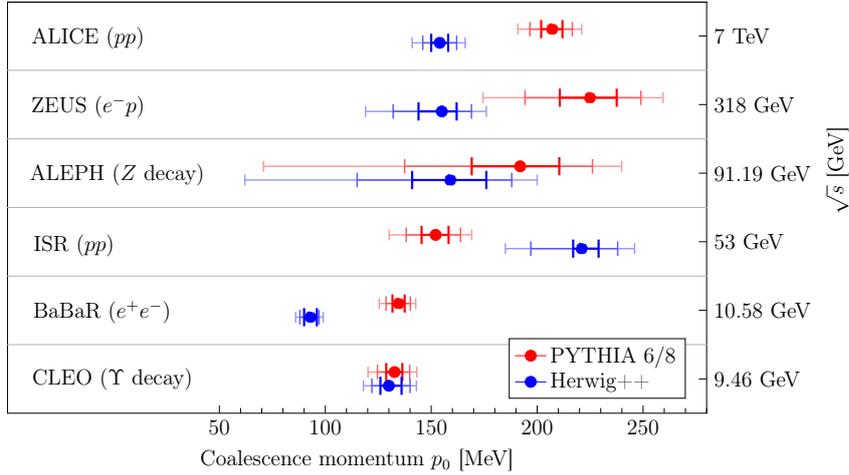


Figure 1.1: Results for p_0 from different experiments. The values of p_0 are inconsistent for different Monte Carlo generators and experiments, from [10].

But even if one assumes the coalescence model to be correct, the uncertainty about p_0 is significant. Besides the antideuteron space propagation model, p_0 contributes to the leading uncertainty on the cosmic antideuteron flux [11], which is the most important quantity for indirect dark matter search.

Beyond this uncertainty, associated with the estimation of the parameter, the coalescence model might be an oversimplification anyway. This could be a reason for its failure to describe all experimental results with one value of p_0 . The authors of "Alternative formation model for antideuterons from dark matter" [1] come to the result that processes like $pp \rightarrow d\pi^+$ or $nn \rightarrow d\pi^-$ make a similar contribution to the deuteron formation as the direct proton neutron binding.

This clearly shows that a better understanding of the antideuteron formation is an essential prerequisite for an astrophysical indirect dark matter search via antideuteron. Hence the theoretical community needs more data to further analyse deuteron formation models. Besides the measurements of deuteron or antideuteron production in low energy proton proton collisions, high precision measurements in electron positron annihilation are considered most valuable [10]. A measurement with Belle II would be crucial, since Belle II is the only ongoing experiment in this category.

1.5 Belle II experiment

The Belle II is a particle collider experiment in Tsukuba, Japan operated by an international collaboration. The main goal of the Belle II experiment is to search for answers to open questions in flavour physics. For example, new flavour symmetries and new flavour symmetry violating processes which explain the structure of the CKM matrix¹ are an important issue. In this context, rare decays of B mesons play a pivotal role since in these processes the mechanisms of the Standard Model are strongly suppressed and new physics can come to light. Anomalies in the field of B meson physics were already observed with the predecessor Belle and also at the BaBar experiment. The mission of Belle II is to provide higher statistics for the measurement of these processes. Belle II and the accelerator SuperKEKB are upgrades of the Belle detector and KEKB accelerator. The purpose of the upgrade is to reach higher luminosities with the accelerator and to adapt the detector to the requirements for those. As KEKB was, SuperKEKB is an electron positron collider with a center of mass energy of 10.58 GeV. This is the resonance energy of the $\Upsilon(4S)$ meson, which decays mainly into B mesons. It is planned to collect data at the $\Upsilon(4S)$ resonance, representing an integrated luminosity of 50 ab^{-1} . In comparison to Belle, the statistical error will scale down substantially because Belle II will produce 50 times more data. The high luminosity of Belle II will also be the main reason for the small statistical error of this analysis.

The Belle II detector (Figure 1.2) is a cylindrical construct with multiple layers of different detector components. The inner detectors are the Pixel Detector (PXD), Strip Vertex Detector (SVD) and the Central Drift Chamber (CDC). These detectors measure the tracks, dE/dX ,² and momenta of charged particles. The next detector is the Time-Of-Propagation (TOP) detector which not only measures the propagation time but is also sensitive to Cherenkov photons³ emitted in quartz bars. Surrounding the TOP Detector is the Electromagnetic Calorimeter (ECL). All these components are enclosed by a 1.5 T

¹The CKM matrix describes the flavour mixture in weak interaction. Flavour mixture results from the fact that the eigenstates of strong and weak interaction are not the same.

² dE/dX denotes the energy deposited per unit of track length.

³Charged particles traveling through a medium exceeding the speed of light in this medium cause the emission of electromagnetic waves by polarization of medium.

magnetic field. The last component is a K_L and Muon (KLM) detector. The end-caps contain the Aerogel Ring-Imaging Cherenkov (ARICH) detector; a variation of the TOP detector and again the KLM detector.

The analysis presented here focusses on the SVD and CDC detectors because up to date the other detectors are not functioning properly for antimatter identification. More details about this issue can be found in section 3.5. The functionality of these two detectors will be briefly explained now. The SVD detector consists of four layers of p and n-doped silicon strips. The p and n strips are perpendicular to each other. Particles create electron electron-hole pairs when crossing the p-n junction. These pairs are separated by an external electric field and detected at the strips. The position of the hit in the detector is reconstructed by the junction point of the activated p and n-strips. The CDC detector is built of 14,336 sense wires and 42,240 field wires in a gas mixture. A charged particle crossing the detector ionises the gas atoms. The electrons resulting from the ionisations drift to the sense wires. Depending on the distance to the next wire, a gas amplification process takes place. The drift length of the electrons can be reconstructed in combination with the known time of the initial electron positron collision. For this reason, the spatial resolution of the CDC is much better than the distance of the wires.

The tracks of charged particles are bent since these detectors are enclosed in a magnetic field. Via the magnitude of the curvature, the momentum of the particles can be determined. By the amount of collected charges from the ionizations along the track, the dE/dX of the particle is ascertained.

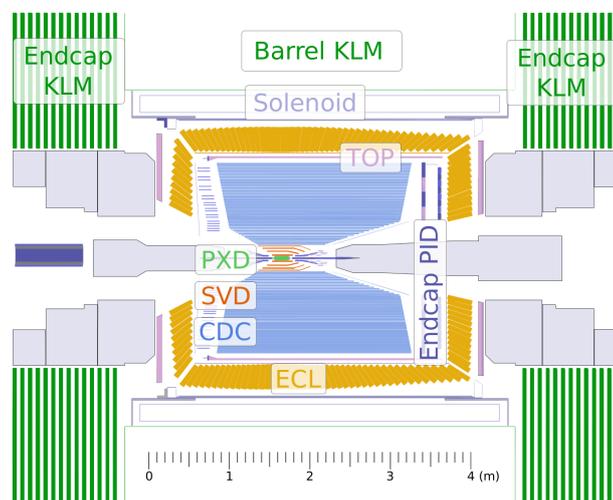


Figure 1.2: Schematic construction of the Belle II detector, from [12].

Chapter 2

Methods

2.1 Classification operators

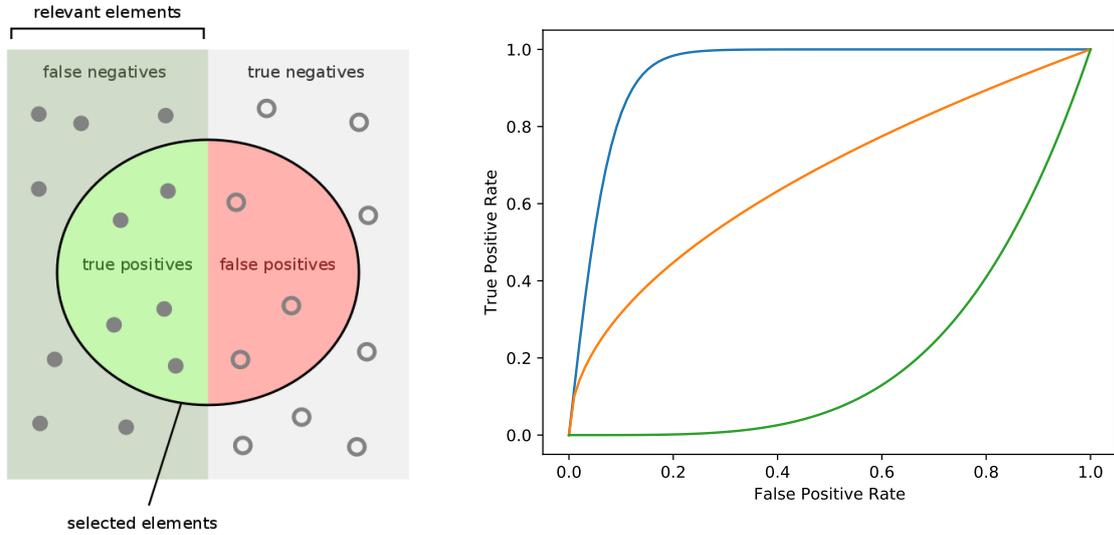
Classification operators are functions which are supposed to identify elements with a specific property in a given set. An example of a classification operator is a medical test determining whether a patient has a disease or not. In particle physics, classification operators are used to identify particle species. For classification operators, multiple characteristic quantities are defined. Most important for the analysis below are the following:

- **True Positive:** These elements have the desired property and are selected by the operator. They are correctly classified.
- **True Negative:** True Negative elements do not have the desired property and are not selected by the operator. They are correctly classified.
- **False Positive:** These elements do not have the desired property but are selected by the operator. They are not correctly classified.
- **False Negative:** These elements have the desired property but are not selected by the operator. They are not correctly classified.

A graphical overview is shown in Figure 2.1a. Important corresponding rates are:

- **True Positive Rate:** $\frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$
- **False Positive Rate:** $\frac{\text{False Positive}}{\text{False Positive} + \text{True Negative}}$

The performance of an operator is commonly illustrated by a receiver operating characteristic curve (ROC curve). In an ROC curve the True Positive rate is plotted as a function of the False Positive rate. A desirable operator has a high True Positive rate even at low False Positive values. An example is in Figure 2.1b.



(a) The filled circles should be selected. The categories of selected and not selected elements are colored [13].

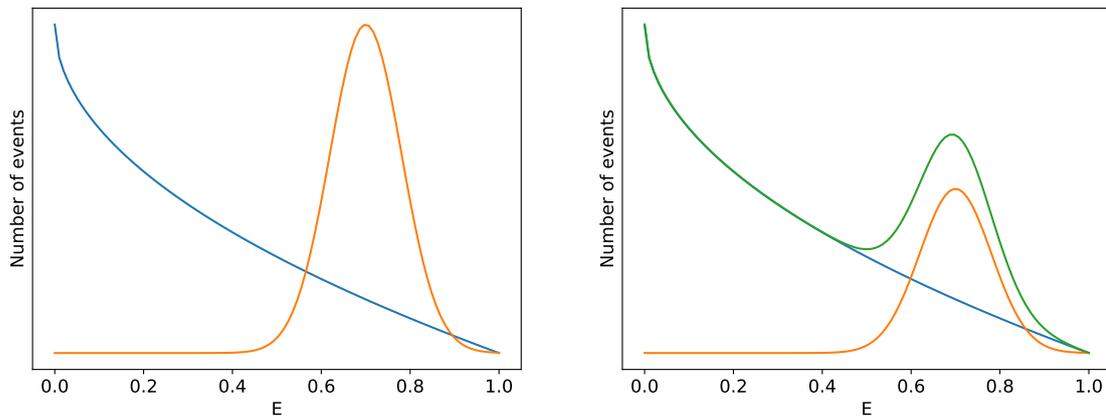
(b) Plot of ROC curves. The blue line represents a highly discriminating operator, whereas the orange and especially the green curves correspond to less effective operators.

Figure 2.1: Operator classification

2.2 Typical branching ratio measurements

In a particle physics analysis, it is necessary to find a method for signal event identification. The validation of the method is done on Monte Carlo simulated data where it is known which events belong to the signal and which to the background. Because the methods of signal identification are never perfect, one has to estimate the proportion of real signal from the events that get selected by the identification method. This is a further step of the analysis. For this step there exist different techniques. A simple approach is to determine the number of expected background events with the simulation. Based on that, the number of signal events in the real data is the difference between the selected events and the expected background events. This method is called counting experiment. There is a more accurate approach, which does not assume that the simulation yields the correct number of background events. This method works roughly as follows: With the help of the Monte Carlo information, one determines the probability distribution of a variable such as the energy E separately for signal and background. Ideally, the distributions for signal and background are different as for instance in Figure 2.2a. If this is the case, the fraction of signal events in the real data can be determined by varying the fraction of signal in the sum of signal and background, thereby fitting the measured data (Figure 2.2b). In case that the distributions of signal and background are very similar, however, this procedure will not

produce stable results. Then one has to improve the identification method or look for other variables with distinguishable background and signal distributions.



(a) The blue line represents the background and the orange line the signal distribution of E . In this case, the shape of background and signal distribution can be distinguished.

(b) The green line shows the measured data. By varying the proportion of signal (orange) and background (blue), the green distribution can be fitted.

Figure 2.2: Finding the proportion of signal in real data.

2.3 Fisher's linear discriminant

Fisher's linear discriminant is a technique to predict the categories of multidimensional two categorical data. This method determines a hyperplane which best separates the categories. In the three dimensional case, the categories are separated by a plane, in the two dimensional case by a line. The hyperplane, or more precisely the orthogonal of the hyperplane is found by maximizing the ratio $\sigma_{between}/\sigma_{within}$, with $\sigma_{between}$ and σ_{within} being the variance between the mean of the classes and the variance within one class, respectively. The projection of the data to the orthogonal of the hyperplane is an $n-1$ dimensional subspace of the n dimensional data in which the categories are maximally separated. Thus this method is ideal to reduce the dimension of a data set while keeping as much information about the separation of the categories as possible.

Chapter 3

Antideuteron Identification

3.1 General idea of antideuteron identification

The particle identification at the Belle II experiments works via likelihood variables. From quantities sensitive to particle discrimination, likelihoods are calculated that describe how probable a given measurement is under the hypothesis of a particle species. These likelihoods are stored for the different tracks rather than the discriminating measurement values themselves. In this chapter, a procedure for antideuteron identification with these likelihoods is explained. A reliable identification is the baseline for a branching ratio or cross section measurement as described in section 2.2.

3.2 Simulated signal events

In this analysis, the antideuteron production rate is analysed for $e^+e^- \rightarrow \bar{d}X$ events around $\sqrt{s} \approx 10.6$ GeV, where X stands for an arbitrary number of further particles. As a specific signal event, the decay $\Upsilon(4S) \rightarrow \bar{d} + p + n + \pi^+ + \pi^+ + \pi^- + \pi^- + \pi^0$ was simulated to study identification efficiencies.¹ The proton and neutron ensure baryon number conservation. By restricting the phase space, the pions cause a realistic momentum distribution for antideuteron. In an event of real deuteron production, there are not necessarily five pions, but they are a representative average for additional particles. Due to baryon number conservation, it is necessary that two baryons are produced. Production of a proton and a neutron is not the only but the most probable case. The distribution of the momentum of the special decay might differ from the momentum distribution of a specific $e^+e^- \rightarrow \bar{d}X$

¹The momentum distributions for antideuteron in $\Upsilon(4S)$ decays and e^+e^- continuum events at $\sqrt{s} = 10.58$ GeV are very similar since the center of mass energy is identical. Thus it does not make a difference which of these is simulated.

event. But since the five pions are a representative average, the momentum distribution of $\Upsilon(4S) \rightarrow \bar{d} + p + n + \pi^+ + \pi^+ + \pi^- + \pi^- + \pi^0$ is assumed to be identical with average momentum distribution of $e^+e^- \rightarrow \bar{d}X$.

For the analysis 200,000 events were simulated using EvtGen [14] and GEANT4 [15] within the Belle II analyses framework. Based on the BaBar measurement,² this equals an integrated luminosity of 20 ab^{-1} .

3.3 Simulated background events

As background both $\Upsilon(4S)$ decays and continuum events are used. The $\Upsilon(4S)$ background are generic decays of $\Upsilon(4S) \rightarrow B^0\bar{B}^0$ and $\Upsilon(4S) \rightarrow B^+B^-$. The hadronisation of $u\bar{u}$, $d\bar{d}$, $s\bar{s}$ and $c\bar{c}$ quarks constitutes the continuum background. Besides this, decays of tau leptons are also background. For each background type, one data block containing 1000 files was processed. Unfortunately, these data blocks do not represent equal luminosities. Therefore, the observed number of background events is scaled for the different background types such that they equal the same luminosity. The processed luminosities for the different background types are presented in Table 3.1

Table 3.1: Background luminosities used to determine the number of expected background events.

background	$B^0\bar{B}^0$	B^+B^-	$u\bar{u}$	$d\bar{d}$	$s\bar{s}$	$c\bar{c}$	$\tau^+\tau^-$
luminosity in fb^{-1}	280	164	164	560	587	141	286

3.4 Particle identification methods

All particle species have known distributions of dE/dX . Based on these distributions, the probability of specific dE/dX values is calculated under the hypothesis of a particle species. These probabilities are determined for multiple other characteristic quantities and are stored as likelihood variables \mathcal{L}_x . At the current status of the Belle II analysis software, the likelihoods for particles and antiparticles are the same. The likelihood of combined detector components is the product of the likelihoods of the sub-detectors. The normal value of interest is not the likelihood, but the probability for a particle species given a measurement. Related to this quantity are the particle ID and pidProbability variables. Hitherto, the particle ID is the standard method and is calculated as the particle likelihood

²From the BaBar measurement the cross section $\sigma = 9.36 \text{ fb}$ for antideuteron production in continuum events was assumed. The production in $\Upsilon(4S)$ decays is probably negligible since up it could not be measured to-date.

divided by the sum of the particle and the pion likelihoods. For example, Equation 3.1 specifies the ID variable for the deuteron.

$$\text{deuteronID} = \frac{\mathcal{L}_d}{\mathcal{L}_d + \mathcal{L}_\pi} \quad (3.1)$$

Besides this standard ID, there is the global ID or pidProbability method. It is defined as the likelihood for a specific particle divided by the sum of the likelihoods of all other particle species. Equation 3.2 is the pidProbability variable for the deuteron. This approach is not only sensitive to confusion with pions but also with other particle species.

$$\text{pidProbability}(\text{deuteron}) = \frac{\mathcal{L}_d}{\mathcal{L}_d + \mathcal{L}_\pi + \mathcal{L}_e + \mathcal{L}_\mu + \mathcal{L}_p + \mathcal{L}_K} \quad (3.2)$$

The performance of the two methods for antideuteron identification is compared through the corresponding ROC curves. As visible in Figure 3.1, the pidProbability method has higher true positive rates at the same false positive value than the particle ID. The first method is used in this analysis, because up to an unacceptable false positive rate of 0.1 the pidProbability method is more precise than the particleID. It turned out that the pidProbability has generally higher True Positive rates at low false positive values, which is why it will be the standard method with the release 02-00-00 of the Belle II analysis software.

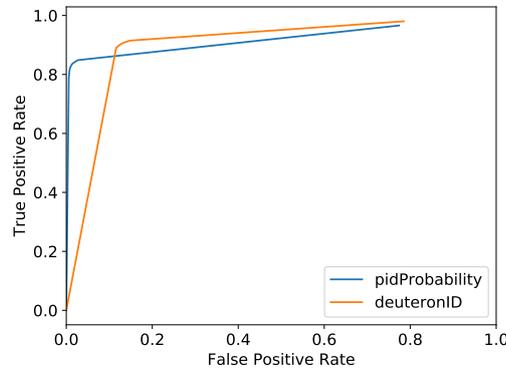


Figure 3.1: ROC curves of the deuteronID and pidProbability methods. For low false positive rates the pidProbability is better than the deuteronID.

3.5 Matter and antimatter detection

The pidProbability method allows to define subsets of detectors for the particle identification. Figure 3.2a shows the ROC curves of the antideuteron identification for all

detector components. It revealed that the identification using all detector components is less efficient than identification employing only the SVD or CDC detector. In Figure 3.2a, the ROC curve of the SVD detector exceeds the curve of the combined detectors and thus provides higher True Positive rates at the same False Positive Rate. In theory, this should not be the case because with all detector components there is at least the same measurement information available as with only one or two components. For this reason, the identification using all detector components should be more precise. As described in section 3.4, the likelihood of combined detectors is the product of the likelihoods from the single detector components. This implies that some detector components determine false likelihoods for antideuterons. To check whether this is a specific problem of deuterons or an issue with antimatter particles in general, the charge conjugate decay was also studied. For deuterons, identification with all detector components provides the best results. In Figure 3.2b, the curve of the combined components exceeds the others. Comparing deuteron and antideuteron identification, it is remarkable that the ROC curves of the single components are very similar except for the TOP detector. The TOP detector is much better for deuterons than for antideuterons. But the incorrect behaviour of the combined detectors can not be explained by the TOP detector alone. Also, the combination of all except the TOP detector is not better than the combination of SVD and CDC detectors.

In addition to deuteron and antideuteron identification, proton and antiproton identification is investigated, too. With protons, the situation is more complicated. Again, identification of the antiparticle with all components does not improve the performance. But some detectors such as the SVD show superior discrimination analysing antimatter than normal matter. The SVD detector achieves much higher True Positive Rates for antiproton identification than for proton identification as plotted in Figure 3.2c and Figure 3.2d.

At the current status, likelihoods for matter and antimatter particles are calculated identically. However, the misbehaviour of the combined detector components observed here suggests that it might be necessary to employ different methods to matter and antimatter likelihoods.

As shown in Figure 3.2b, mainly the SVD and CDC detectors account for the deuteron identification; the other detectors contribute only very little since the curve of the combined components exceeds these two only slightly. For this reason, it is to be expected that detection rates using the SVD and CDC detectors will be merely worse than with all detector components used. Hence only the SVD and CDC detectors are employed for likelihood and pidProbability variables in this analysis.

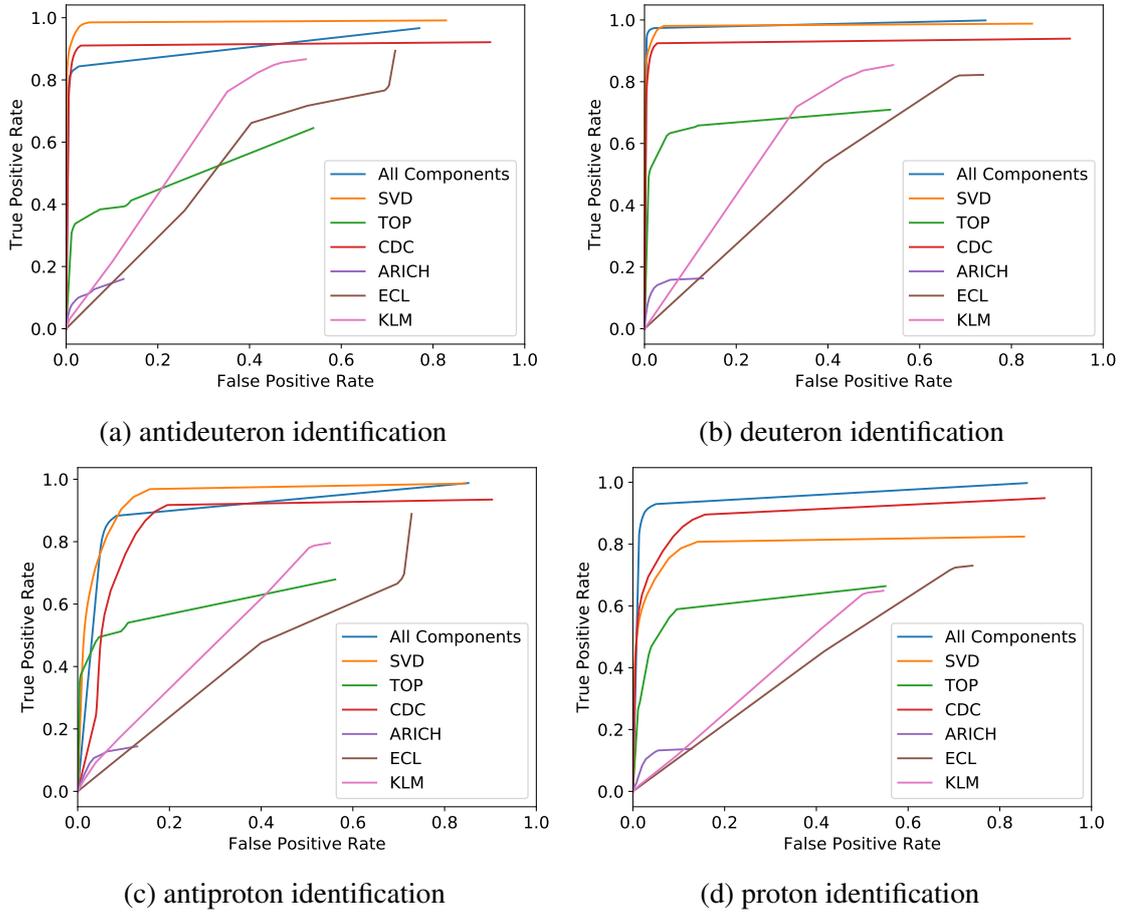


Figure 3.2: ROC curves for particle and antiparticle identification with the `pidProbability` variable and detector subsets. In both antimatter cases, identification with all detector components employed is malfunctioning.

3.6 Cuts for antideuteron identification

To improve the purity of the selected events, cuts are made on certain variables. For antideuteron identification, the most important set of variables for cuts are the tracking variables. Tracking variables describe the quality of the track measurement and reconstruction. Cuts on the number of hits in the PXD, SVD and CDC detectors (`nPXDHits`, `nSVDHits`, `nCDCHits`) are done. The cuts on the hits in the tracking detectors ensure a precisely measured track and dE/dX , which is necessary for a reliable particle identification because the identification in the SVD and CDC detectors is mainly based on this quantity. Some signal events get lost with a cut on the number of hits, but the benefit is a reduction of background. Different cuts are applied in the momentum bins (0.5 – 1.5) GeV and (1.5 – 2.5) GeV because the dE/dX and therefore the number of hits of a particle is momentum dependent.

Table 3.2: Final cuts for antideuteron identification and efficiencies. The efficiencies refer to their respective bin.

momentum [GeV/c]	0.5 - 1.5	1.5 - 2.5
nPXDHits	≥ 2	≥ 2
nSVDHits	≥ 8	≥ 8
nSVDHits	≥ 30	≥ 40
d0 [cm]	-0.02 - 0.02	-0.01 - 0.01
z0 [cm]	-0.05 - 0.05	-0.05 - 0.05
omega	≤ 0	≤ 0
pidProbability(SVD, CDC)	≥ 0.95	≥ 0.95
efficiency	72%	64%

Furthermore, cuts on the distance of the closest approach to the beam collision point in z direction and $r - \phi$ plain ($z0$, $d0$) are applied. Tracks not originating from the collision point do not belong to any particle created by the e^+e^- collision and are therefore definitely background.

At last, a cut is done on the curvature of the track (omega). The curvature corresponding to the charge of the particle has to be negative because antideuteron has charge -1 . The plots of the cut variables are shown in Figure 3.3. The final cuts and efficiencies are reported in Table 3.2

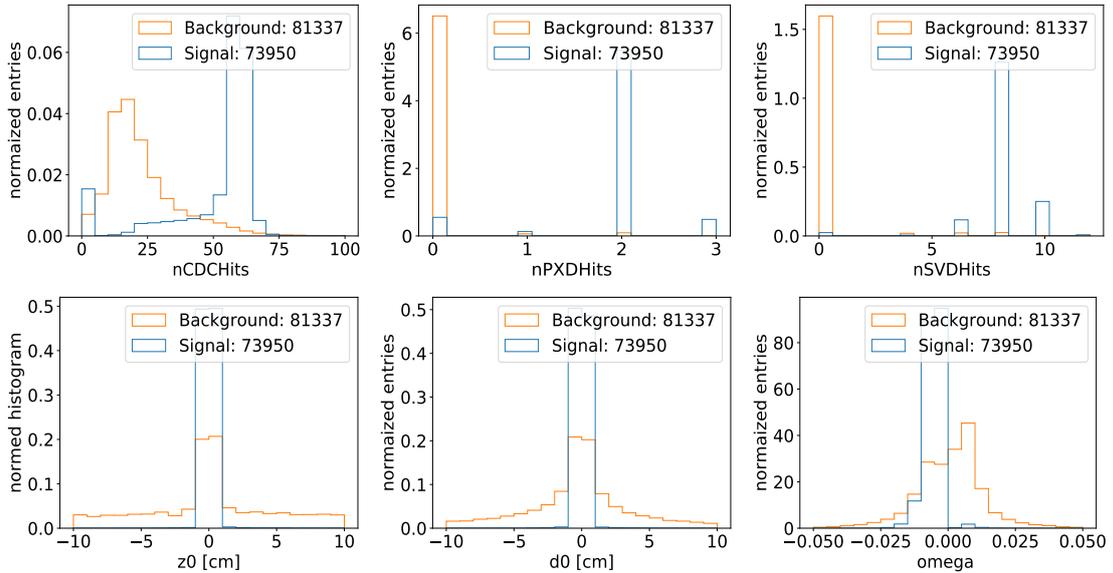
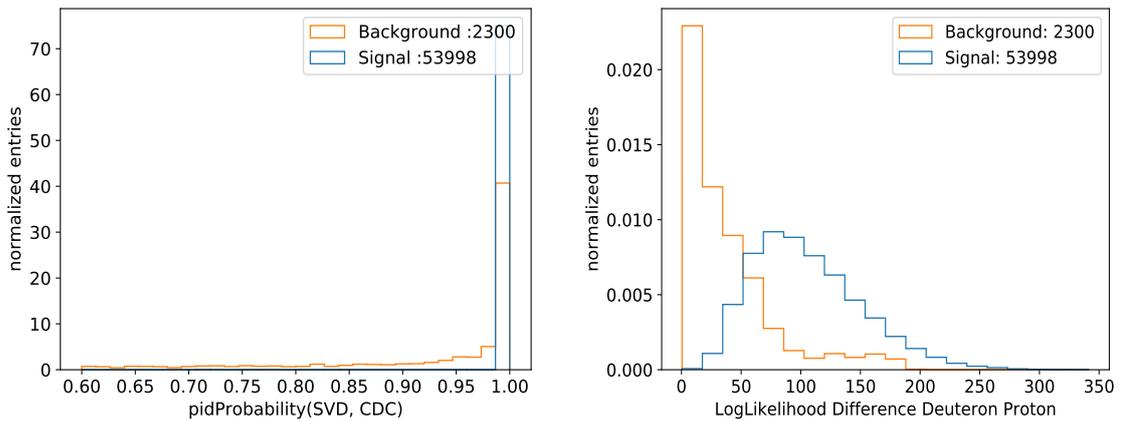


Figure 3.3: Normed histograms of tracking variables.

3.7 Distribution of deuteronID

In the BaBar and CLEO experiment, the dE/dX distribution was used to discriminate between background and signal. In the Belle II experiment, the dE/dX information measured by the tracking detectors is not stored following the completion of reconstruction. Nevertheless, the dE/dX information is the most relevant basis for particle identification via the likelihoods of the SVD and CDC detectors. For this reason, the pidProbability for deuteron was initially used in this analysis. With this variable the problem arose that the signal and background distribution both peaked at 1 and, consequently a stable discrimination between signal and background proved impossible. First, it was assumed that the peak in the pidProbability distribution for background would vanish with tighter cuts. Therefore, the maximal allowed distance to the interaction point was reduced and more hits were required in the tracking detectors. These cuts reduced the number of background events observed in the local available background samples so far that the shape of the background distribution was no longer recognisable. To process more background data, the Grid system was used. Even with the higher statistics and new cuts, the pidProbability of the background in the momentum bin (0.5 – 1.5) GeV still peaked at 1, as visible in Figure 3.4a. Hence further discrimination methods had to be employed. To understand the origin of this behaviour, the ratio of the deuteron pidProbability and the proton pidProbability was investigated. The idea is that the proton is most similar to the deuteron and therefore this ratio might be sensitive to discrimination between antideuteron and fake antideuteron. Because this value seemed promising, as shown in Figure 3.4b, the deuteron and proton likelihoods were investigated directly.



(a) pidProbability distribution for tracks surviving the cuts with a momentum between 0.5 GeV and 1.5 GeV.

(b) Logarithm of the proton and deuteron likelihood ratio for tracks surviving the cuts with a momentum between 0.5 GeV and 1.5 GeV.

Figure 3.4: Normed histograms of pidProbability and likelihood ratio.

3.8 Likelihood distributions

Figure 3.5 visualizes a two dimensional plot of the deuteron and proton likelihoods. Since the distributions for signal and background are very different in the 2D plane, it is possible to discriminate between them. For signal events, the deuteron log-likelihood is near zero, whereas the log-likelihood for background events is mostly very small.

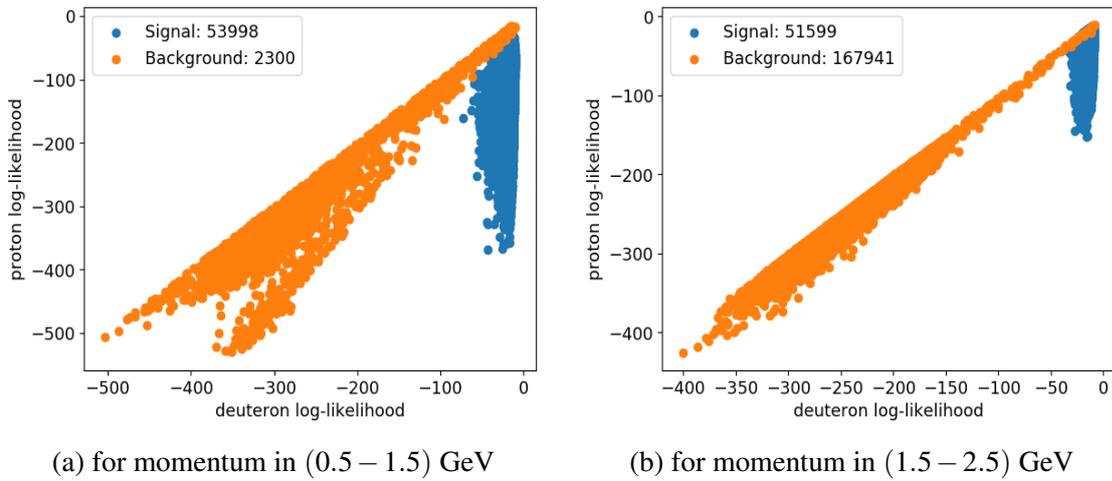


Figure 3.5: 2D Plot of the deuteron and proton log-likelihoods. Background and signal are clearly separated.

A small log-likelihood does not imply a small `pidProbability` since the `pidProbability` is calculated as a likelihood ratio (Equation 3.2). For many background events, not only the nominator but also the denominator of the likelihood ratio is very small. This explains the peak at 1 in the `pidProbability` distribution. In consequence, separation of background and signal through the log-likelihood variables is much more precise than through the `pidProbability` variable. This is visualised in Figure 3.6: Signal and Background are separated along the log-likelihood but not along the `pidProbability` axis.

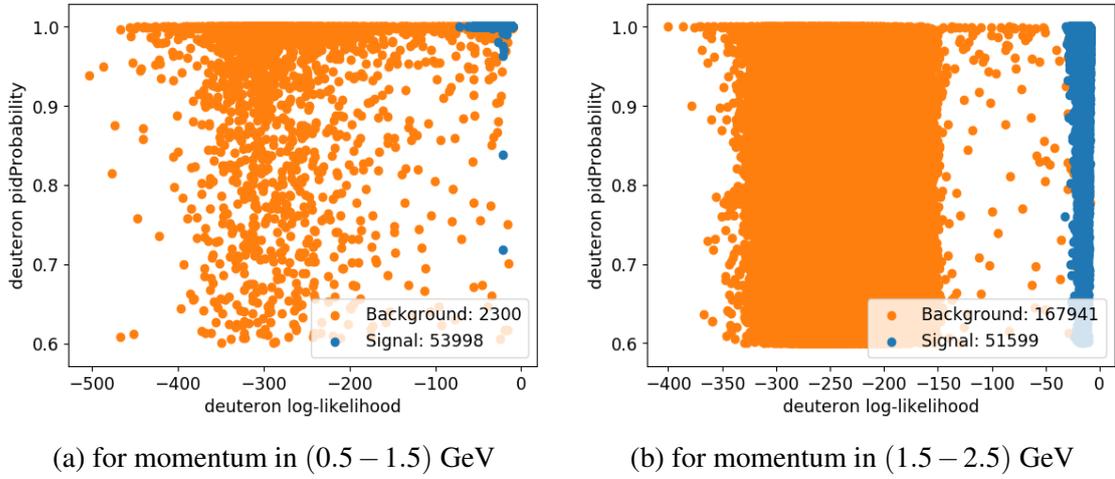


Figure 3.6: 2D Plot of the deuteron log-likelihood and pidProbability. The log-likelihood separates background and signal whereas the pidProbability does not.

3.9 Separating background an signal with Fisher's linear discriminant

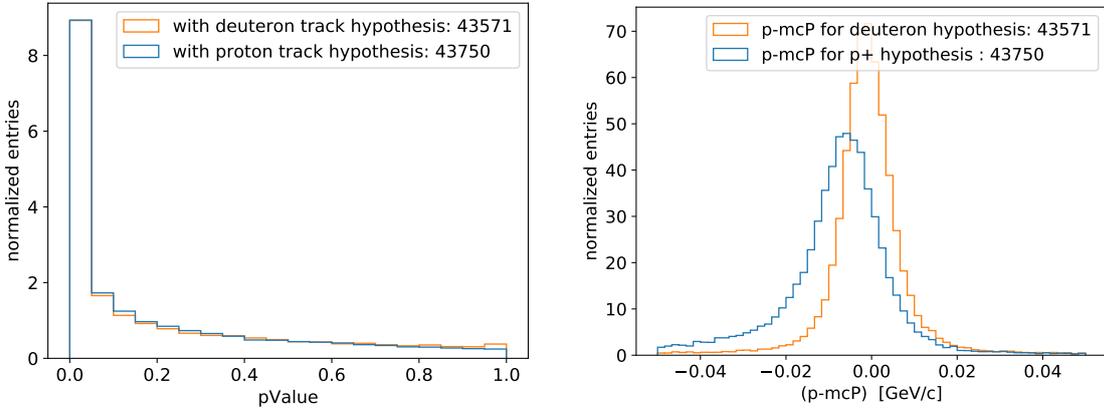
With a one dimensional variable it is easier to count events in the signal region or to fit the probability distributions of signal and background. For this reason, a projection to a one dimensional variable, optimally separating signal and background, is helpful. This projection was found via the Fisher's linear discriminant technique. The principles of Fisher's linear discriminant are explained in section 2.3. Compared with the deuteron likelihood, this approach separates signal and background much more precisely and reduces the background roughly by a factor of two. Table 3.3 specifies the number of expected background events in the signal region for a luminosity of 50 ab^{-1} . The cuts for the signal region for Fisher's and likelihood separation are chosen such that they both have an efficiency of 99%. By this approach, the number of background events can be directly compared.

The projection operator of Fisher's linear discriminant is $(0.057, -0.004)$ in the lower momentum bin and $(0.062, -0.012)$ in the higher momentum bin. The first coordinate denotes the deuteron log-likelihood $\ln(\mathcal{L}_d)$ and the second coordinate the proton log-likelihood $\ln(\mathcal{L}_p)$. Thus the projection from the 2D plain to the one dimensional variable is carried out as following: $0.057 \cdot \ln(\mathcal{L}_d) - 0.004 \cdot \ln(\mathcal{L}_p)$, respectively $0.062 \cdot \ln(\mathcal{L}_d) - 0.012 \cdot \ln(\mathcal{L}_p)$. Hence the deuteron log-likelihood is the more important value for the discrimination than the proton log-likelihood since its weight in the projection is one order of magnitude larger.

Table 3.3: Expected number of background events for a luminosity of 50 ab^{-1} . The results of both separation methods via deuteron log-likelihood and Fisher’s discriminant are quoted.

background	$B^0\bar{B}^0$	B^+B^-	$u\bar{u}$	$d\bar{d}$	$s\bar{s}$	$c\bar{c}$	$\tau^+\tau^-$	sum
likelihood	1216	1520	1824	4560	3400	1824	608	14952
Fisher	608	304	912	3648	1824	912	304	8512

3.10 Deuteron track hypothesis



(a) χ^2 probability of track fit for proton and deuteron hypothesis.

(b) Shift of measured momentum for proton hypothesis.

Figure 3.7: Investigation fo deuteron track fit hypothesis.

The track fitting in the SVD and CDC requires a mass hypothesis for the particle which might have caused the track. As a default, the tracks are fitted with the hypothesis of pion, proton and muon mass. If particle species are reconstructed whose mass hypothesis is unavailable, the next closest hypothesis is used. It is possible to use additional hypotheses for the track fit. In the following, the impact of adding the deuteron hypothesis is studied. In this case, the deuteron hypothesis track fit is used for particles in the deuteron list¹ instead of the proton hypothesis as in the default case. The plausibility of a fit to a track is given by the χ^2 probability. A comparison of the χ^2 probability distributions for deuteron and proton hypotheses is shown in Figure 3.7a. The distributions are very similar and the quality of the track fit does not seem significantly improved. It is astonishing that with the use of the deuteron hypothesis slightly less track fits are successful than with the use of the proton hypothesis. The track hypothesis also has an influence on the measured momentum. As shown in Figure 3.7b, the reconstructed momentum tends to be too small

¹In the Belle II analysis software particle candidates are listed.

for tracks fitted with the proton hypothesis. In Figure 3.8, the shift is plotted for different momentum bins. The figure shows that the momentum is only shifted for low values. For higher momenta, the false track hypothesis has no influence. The absolute value of the momentum shift is only in the low MeV/c range and seems insignificant. In conclusion, the use of the deuteron hypothesis does not seem compelling for the purpose of this analysis, because only deuterons with a momentum larger than 0.5 GeV/c are considered. The cost of an additional hypothesis would be increased computational effort for measurement reconstruction. However, one should be aware of the fact that for momenta lower than 0.5 GeV/c the proton hypothesis has a bias of several percent towards lower momenta.

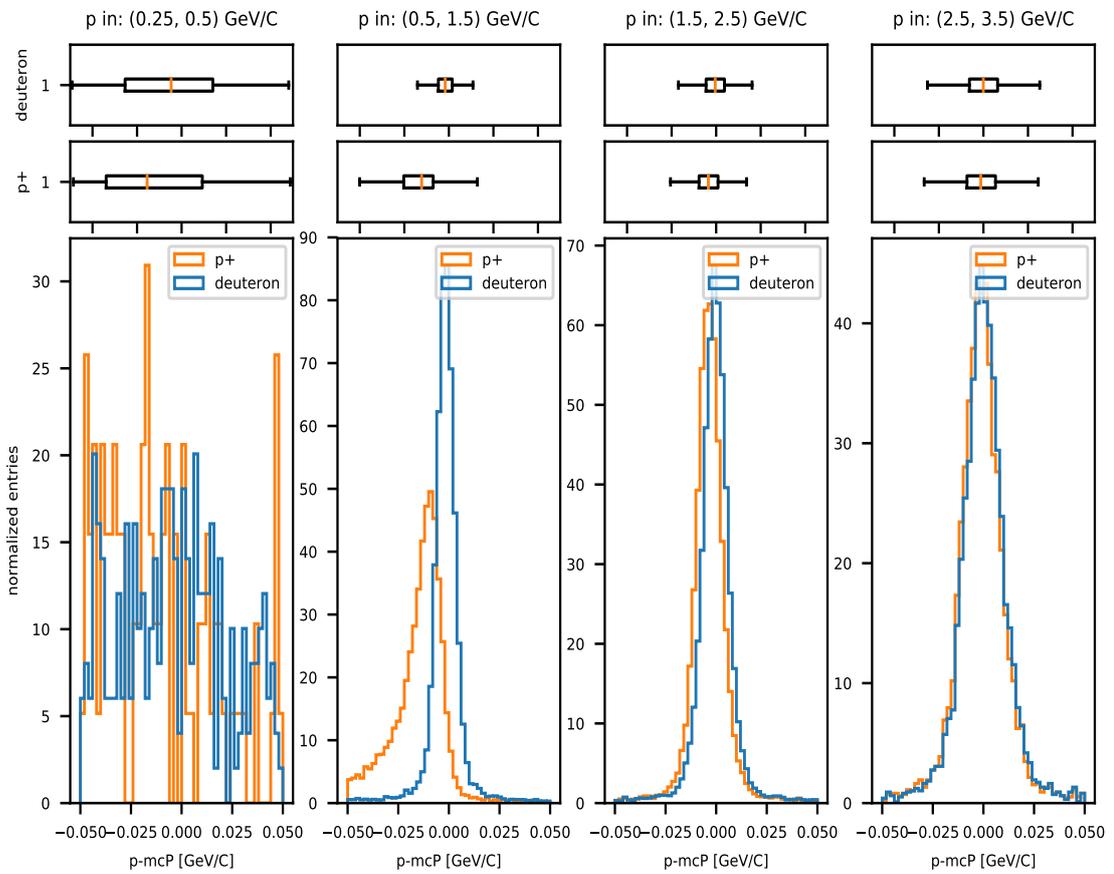


Figure 3.8: Shift between the measured momentum with proton and deuteron track hypotheses. The results with the proton hypothesis are systematically too small at low momenta. Furthermore the momentum distribution for the incorrect track hypothesis is broader than for the deuteron hypothesis as visible in the box plots. The box covers the first to the third quantile. The width of the whiskers is 1.5 times the interquartile range.

Chapter 4

Sensitivity Estimation

4.1 General assumptions

The statistical uncertainties are estimated by counting signal and background events in the signal region after Fisher's discriminant analysis. It is assumed that signal and background are independently Poisson distributed. Confidence intervals with a confidence level of α are computed via the connection to the χ^2 distribution:

$$\frac{1}{2}\chi^2(\alpha/2, 2n) \leq \mu \leq \frac{1}{2}\chi^2(1 - \alpha/2, 2n + 2) \quad (4.1)$$

where $\chi^2(\cdot, \cdot)$ is the quantile function of the χ^2 distribution and μ the mean and n the number of events.

A systematic bias resulting from the fit of the signal and background distribution is not included because the distributions are not fitted. Furthermore, the systematic uncertainties of the Belle II detector are not yet known. Hence only the statistical error is given.

4.2 Statistical error of continuum measurement

The entire number of antideuterons N_{AntiD} is given by $N_{AntiD} = N_S/\text{eff}$ where N_S is the number of observed signal events and eff the efficiency of antideuteron identification. The number of signal events N_S is calculated as the number of events N minus the expected number of background events N_B . The error for the number of antideuterons is therefore given by Equation 4.2:

$$\Delta N_{AntiD} = \frac{1}{\text{eff}}\sqrt{N} = \frac{1}{\text{eff}}\sqrt{N_S + N_B} \quad (4.2)$$

Table 4.1: Expected statistical errors for $\sigma(e^+e^- \rightarrow \bar{d}X)$ at $\sqrt{s} = 10.58$ GeV for a luminosity of $L = 50 \text{ ab}^{-1}$. The number of signal events is derived from the outcome of the BaBar measurement. The error corresponds to a confidence level of 1σ .

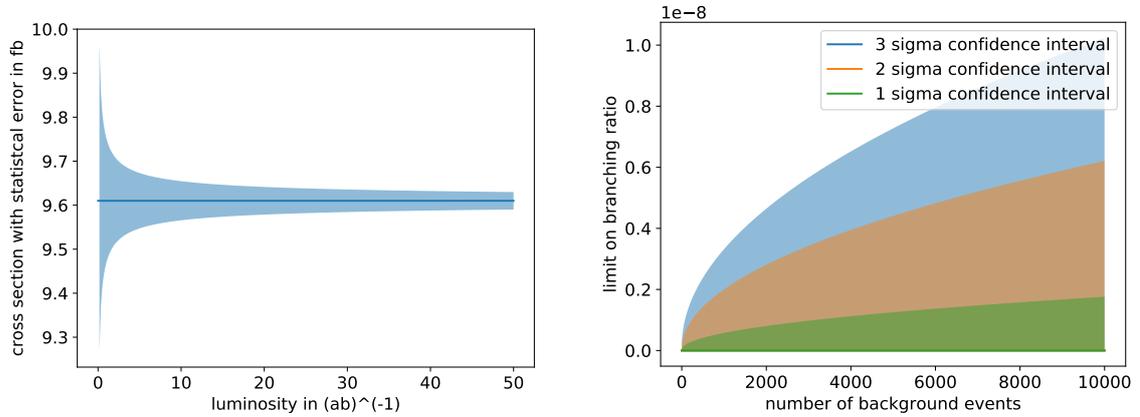
momentum [GeV/c]	0.5 - 1.5	1.5 - 2.5	0.5 - 2.5
eff	$0.37 \cdot 0.71 = 0.26$	$0.37 \cdot 0.63 = 0.23$	0.49
N_S	$1.3 \cdot 10^5$	$1.1 \cdot 10^5$	$2.4 \cdot 10^5$
N_B	6688	1824	8512
ΔN_{AntiD}	1417	1418	1003
$\Delta\sigma(e^+e^- \rightarrow \bar{d}X)$ [fb]	0.029	0.029	0.021

Because of the cut on the momentum in the lab frame, the efficiency eff is determined by two factors: Firstly, the fraction of antideuterons within the momentum interval, secondly, the efficiency of the antideuteron selection in this momentum interval. The two factors of the efficiency are determined via the simulated $\Upsilon(4S) \rightarrow \bar{d} + p + n + \pi^+ + \pi^+ + \pi^- + \pi^- + \pi^0$ events. Thereby it is assumed that the momentum distribution of these events equals the momentum distribution of $e^+e^- \rightarrow \bar{d}X$ at $\sqrt{s} = 10.58$ GeV.

For evaluation of the statistical error, the number of signal events is needed. The uncertainties given here are based on an expected number of signal events resulting from the cross section $\sigma(e^+e^- \rightarrow \bar{d}X) = 9.63$ fb measured at the BaBar experiment [7]. Under this assumption, the overall statistical 1σ uncertainty of the Belle II experiment using the method presented here and data corresponding to an integrated luminosity of 50 ab^{-1} would result in ± 0.021 fb or $\pm 0.22\%$. The statistical errors of the two bins are listed in Table 4.1 for a luminosity of 50 ab^{-1} . Figure 4.1a shows the 1σ confidence interval as a function of the luminosity.

4.3 Limits for antideuteron production in $\Upsilon(4S)$ events

Assuming one could discriminate between antideuterons coming from continuum events and antideuterons from $\Upsilon(4S)$ decays, one could also look at the branching ratio $\Upsilon(4S) \rightarrow \bar{d}X$. This branching ratio is presumably tiny because the available phase space for antideuterons in B meson decays is very small and B mesons are the almost exclusive decay products of the $\Upsilon(4S)$. This is the reason why the antideuteron production in $\Upsilon(4S)$ decays could not be measured in previous experiments [7]. Since an assumption about the number of signal events is needed, one can not calculate the expected uncertainty of a potential Belle II measurement. Under the hypothesis that the production rate for antideuteron is also under the sensitivity of Belle II, Figure 4.1b shows the upper confidence limits of antideuteron production over the number of expected background events. As quoted in Table 4.1, the method presented here yields 8512 background events. This would result in an upper limit of a branching ratio $\mathcal{B} = 1.6 \cdot 10^{-9}$ on a 1σ confidence level.



(a) Statistical 1σ error of deuteron production cross section $\sigma(e^+e^- \rightarrow \bar{d}X)$ at $\sqrt{s} = 10.58$ GeV over the luminosity.

(b) Limits on branching ratio of antideuteron production from $\Upsilon(4S)$ over the number of background events under the hypothesis that the real branching ratio is below the sensitivity of Belle II and $L = 50 \text{ ab}^{-1}$. The method of this analysis would yield 8512 background events.

Figure 4.1: Statistical uncertainties for potential Belle II measurements.

4.4 Comparison to BaBar measurement

In the BaBar experiment, the production of antideuteron was measured in $\Upsilon(2S)$ and $\Upsilon(3S)$ decays and in e^+e^- continuum events at $\sqrt{s} = 10.58$ GeV.

The continuum measurement of the BaBar experiment had a statistical uncertainty of $\Delta\sigma = 0.41 \text{ fb}$ or 4.3% and was based on data corresponding to an integrated luminosity of 429 ab^{-1} . The reduction of the statistical error at Belle II is mainly due to the fact that Belle II will have 116 times more data. Furthermore, the BaBar experiment only used the momentum range of $(0.5 - 1.5) \text{ GeV}/c$, because for other momenta they could not identify the antideuterons via dE/dX . As indicated in Table 4.1, this reduces the overall efficiency approximately by a factor of two. Consequently, the statistical uncertainty expands by a factor of $\sqrt{2}$ (see Equation 4.2). Using only the momentum range $(0.5 - 1.5) \text{ GeV}$ and data with $L = 429 \text{ ab}^{-1}$, the method of this analysis would result in a statistical error of $\Delta\sigma = 0.31 \text{ fb}$ (3.2%). The BaBar experiment also took statistical uncertainties for the antideuteron detector material interaction into account.¹ This causes the remaining difference in the statistical uncertainties, because the efficiency of this method and the BaBar method are very similar for $p \in (0.5 - 1.5) \text{ GeV}/c$. In conclusion, the statistical error is reduced by three factors: a factor 11 with higher statistics, factor $\sqrt{2}$ due to a

¹The BaBar experiment simulated antiproton interactions with the detector and scaled the cross section, since they could not simulate antideuteron.

higher efficiency and factor 1.3 because of a better simulation.

The BaBar experiment could not detect antideuteron production in $\Upsilon(4S)$ decays. They did not determine explicit limits on the branching ratio but checked that the branching ratio is below their sensitivity by comparing on and off peak data sets around the $\Upsilon(4S)$ resonance.

Chapter 5

Conclusion and Outlook

5.1 Conclusion

A better understanding of the production rate of antideuteron in Standard Model processes is the baseline for the indirect dark matter search via antideuterons. This thesis describes a method for antideuteron identification at the Belle II experiment, which constitutes the foundation of a measurement of the production rate. The described method identifies antideuterons with high efficiency in a wide momentum range from $(0.5 - 2.5)$ GeV/c. In combination with the high luminosity of the Belle II experiment, this results in a significant reduction of the statistical errors of the $e^+e^- \rightarrow \bar{d}X$ cross section. Only the SVD and CDC detectors are used here since the antideuteron detection is currently not functioning properly with the other detectors. The results of an additional deuteron track hypothesis is compared to the default proton hypothesis. No urgent need for the deuteron hypothesis is found, since it yields only a minor improvement.

5.2 Outlook

Because of an already high purity, the signal region of the $e^+e^- \rightarrow \bar{d}X$ channel will probably be dominated by real signal events. This means it is reasonable to look for a high signal efficiency to reduce statistic errors since the background events have only a small impact on the statistical error. This approach was used in the analysis presented here and is the main reason why the statistical error was reduced by more than a factor of 11 compared to BaBar. In the case of $\Upsilon(4S) \rightarrow \bar{d}X$, the situation is different. Here, most events in the signal region are expected to be background events. Hence the sensitivity could be improved via a reduction of background, even though the signal efficiency would thereby decrease substantially. Figure 4.1b shows the potential for improvement on the limits on

the antideuteron branching ratio when there is less background. For this reason it might prove helpful to use two different methods for continuum and $\Upsilon(4S)$ measurement. The signal efficiency optimised method of this analysis is good for the continuum case, whereas a background reduction optimised method would be better suited for a measurement of the $\Upsilon(4S)$ branching ratio. The method presented in this thesis would have to be expanded for the $\Upsilon(4S)$, because no continuum separation is included.

In the sensitivity estimation, the systematical uncertainties are neglected although they might dominate the combined error. This was the case in the BaBar measurement. Here, the systematic uncertainty was ${}^{+1.17}_{-1.01}$ fb where the statistical error was only 0.41 fb. This means that the reduction of the statistical error via the large Belle II data is only worthwhile if the systematic uncertainties also decrease. The systematic error at the BaBar measurement mainly arose from uncertainties in the reconstruction efficiencies, background model and kinematic acceptance. Furthermore, the selection on the distance of the closest approach of the track to the collision point had a huge systematic uncertainty. Hopefully, the uncertainty about the corresponding cuts on d_0 and z_0 will improve with Belle II due to the better focus of the beams before the collision. The improvement on the tracking system is another goal of the detector update. Thereby also the systematic uncertainty about the tracking efficiency might decrease. The systematic bias of the fit of background and signal model was compared to the other systematic uncertainties negligible in the BaBar analysis. Therefore, no further systematic uncertainties should arise as compared to the counting experiment assumed in this analysis.

Bibliography

- [1] L. A. Dal and A. R. Raklev. “Alternative formation model for antideuterons from dark matter”. In: *Phys. Rev. D* 91.12 (2015). [Erratum: *Phys. Rev. D* 92,no.8,089901(2015)], p. 123536. DOI: 10.1103/PhysRevD.91.123536. arXiv: 1504.07242 [hep-ph].
- [2] Mario Kadastik, Martti Raidal, and Alessandro Strumia. “Enhanced anti-deuteron Dark Matter signal and the implications of PAMELA”. In: *Phys. Lett. B* 683 (2010), pp. 248–254. DOI: 10.1016/j.physletb.2009.12.005. arXiv: 0908.1578 [hep-ph].
- [3] Francesco Barile. “Light (Hyper-)Nuclei production at the LHC measured with ALICE”. In: *EPJ Web Conf.* 95 (2015), p. 04003. DOI: 10.1051/epjconf/20159504003. arXiv: 1411.1941 [hep-ex].
- [4] S. Schael et al. “Deuteron and anti-deuteron production in e^+e^- collisions at the Z resonance”. In: *Phys. Lett. B* 639 (2006), pp. 192–201. DOI: 10.1016/j.physletb.2006.06.043. arXiv: hep-ex/0604023 [hep-ex].
- [5] S. Chekanov et al. “Measurement of (anti)deuteron and (anti)proton production in DIS at HERA”. In: *Nucl. Phys. B* 786 (2007), pp. 181–205. DOI: 10.1016/j.nuclphysb.2007.06.022. arXiv: 0705.3770 [hep-ex].
- [6] W. et al. Gibson. “Production of deuterons and antideuterons in proton- proton collisions at the CERN ISR”. In: *Lettere al Nuovo Cimento* 21 (1978), pp. 189–194.
- [7] J. P. Lees et al. “Antideuteron production in $\Upsilon(nS)$ decays and in $e^+e^- \rightarrow q\bar{q}$ at $\sqrt{s} \approx 10.58$ GeV”. In: *Phys. Rev. D* 89.11 (2014), p. 111102. DOI: 10.1103/PhysRevD.89.111102. arXiv: 1403.4409 [hep-ex].
- [8] D. M. Asner et al. “Anti-deuteron production in Upsilon(nS) decays and the nearby continuum”. In: *Phys. Rev. D* 75 (2007), p. 012009. DOI: 10.1103/PhysRevD.75.012009. arXiv: hep-ex/0612019 [hep-ex].
- [9] H. Albrecht et al. “Study of Anti-deuteron Production in e^+e^- Annihilation at 10-GeV Center-of-mass Energy”. In: *Phys. Lett. B* 236 (1990), pp. 102–108. DOI: 10.1016/0370-2693(90)90602-3.

- [10] T. Aramaki et al. “Review of the theoretical and experimental status of dark matter identification with cosmic-ray antideuterons”. In: *Phys. Rept.* 618 (2016), pp. 1–37. DOI: 10.1016/j.physrep.2016.01.002. arXiv: 1505.07785 [hep-ph].
- [11] Umberto Tamponi. *Constraining the astrophysical models for \bar{d} production at Belle II*. <https://kds.kek.jp/indico/event/25459/session/1/contribution/160/material/slides/0.pdf>. Accessed: 2018-13-10.
- [12] Christian Pulvermacher. *dE/dx particle identification and pixel detector data reduction for the Belle II experiment*. KIT, Diplomarbeit. 2012. URL: <https://ekp-invenio.physik.uni-karlsruhe.de/record/48263>.
- [13] FeanDoe Wikipedia. *Sensitivity and specificity*. https://en.wikipedia.org/wiki/Sensitivity_and_specificity. Accessed: 2018-13-10.
- [14] David J. Lange. “The EvtGen particle decay simulation package”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 462.1 (2001), pp. 152–155. DOI: 10.1016/S0168-9002(01)00089-4.
- [15] S. Agostinelli et al. “Geant4—a simulation toolkit”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (2003), pp. 250–303. DOI: 10.1016/S0168-9002(03)01368-8.

Acknowledgements

I am thankful to Prof. Dr. Kuhr for the opportunity to write my bachelor thesis in his working group about such an interesting topic. Besides Prof. Dr. Kuhr, I also want to thank Dr. Martin Ritter and James Kahn for many helpful suggestions concerning my work in the group meetings and their tips about the basf2 software. Without their help, my work would not have been possible. Futhermore, I owe Tobias Jenegger a big thank you for running my Grid jobs. I am also grateful to Dr. Michael Bender and Gordian Edenhofer for proofreading my thesis and to my farther and brother for correcting spelling mistakes. Finally, I want to thank the other bachelor students Anna Bertolini, Christoph Ames, Daniel Moritz, Gordian Edenhofer, Nan-Hee Kang and Yasin Silyanoglu for their help and the pleasant atmosphere in our office.

Erklärung zur Bachelorarbeit

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.

Bachelor's thesis statement of originality

I hereby confirm that I have written the accompanying thesis by myself, without contributions from any sources other than those cited in the text and acknowledgments.

.....

Date and location

Signature