Springer Theses
Recognizing Outstanding Ph.D. Research

Thomas Keck

# Machine Learning at the Belle II Experiment

## The Full Event Interpretation and Its Validation on Belle Data

Springer

**Springer Theses**

Recognizing Outstanding Ph.D. Research

## Aims and Scope

The series "Springer Theses" brings together a selection of the very best Ph.D. theses from around the world and across the physical sciences. Nominated and endorsed by two recognized specialists, each published volume has been selected for its scientific excellence and the high impact of its contents for the pertinent field of research. For greater accessibility to non-specialists, the published versions include an extended introduction, as well as a foreword by the student's supervisor explaining the special relevance of the work for the field. As a whole, the series will provide a valuable resource both for newcomers to the research fields described, and for other scientists seeking detailed background information on special questions. Finally, it provides an accredited documentation of the valuable contributions made by today's younger generation of scientists.

## Theses are accepted into the series by invited nomination only and must fulfill all of the following criteria

- They must be written in good English.
- The topic should fall within the confines of Chemistry, Physics, Earth Sciences, Engineering and related interdisciplinary fields such as Materials, Nanoscience, Chemical Engineering, Complex Systems and Biophysics.
- The work reported in the thesis must represent a significant scientific advance.
- If the thesis includes previously published material, permission to reproduce this must be gained from the respective copyright holder.
- They must have been examined and passed during the 12 months prior to nomination.
- Each thesis should include a foreword by the supervisor outlining the significance of its content.
- The theses should have a clearly defined structure including an introduction accessible to scientists not expert in that particular field.

More information about this series at http://www.springer.com/series/8790

Thomas Keck

# Machine Learning at the Belle II Experiment

## The Full Event Interpretation and Its Validation on Belle Data

Doctoral thesis accepted by the Karlsruhe Institute of Technology, Karlsruhe, Germany

Springer

*Author*
Dr. Thomas Keck
Institute of Experimental Particle Physics
Karlsruhe Institute of Technology
Karlsruhe, Germany

*Supervisor*
Prof. Michael Feindt
Karlsruhe Institute of Technology
Karlsruhe, Germany

ISSN 2190-5053 ISSN 2190-5061 (electronic)
Springer Theses
ISBN 978-3-319-98248-9 ISBN 978-3-319-98249-6 (eBook)
https://doi.org/10.1007/978-3-319-98249-6

Library of Congress Control Number: 2018950801



This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

# Supervisor's Foreword

In this thesis, Thomas Keck has automated and optimized a large part of the reconstruction work of highly skilled elementary particle physicists at the Belle and Belle II experiments in Japan performed by large international collaborations (450 and >700 researchers from the whole world) by means of a hybrid (expert knowledge and machine learning) hierarchical artificial intelligence system. Based on a first version for the Belle experiment developed in 2011, he improved the performance by another factor of 2 to an overall factor of 4 compared to human performance in reconstruction efficiency at constant purity. He invented several new machine learning methods making possible trainings independent of systematic uncertainties in Monte Carlo simulations. He made it work both for the existing Belle data and the future Belle II data. In addition, he performed an analysis of an important rare decay of B mesons with the potential to find new physics beyond the current standard model. Here, he pointed out previously unknown systematic uncertainties.

The thesis poses an important step in improving the human expert efficiency in exploiting expensive fundamental science experiments by means of artificial intelligence/machine learning (AI/ML). It is a highly enjoyable, readable, successful and valuable application of ML methods and shows real value and progress in elementary particle physics beyond the current hype around AI, a superb example for the new development of data science for science. It also shows that deep human understanding and intelligence is the basis for successful application of artificial intelligence. Notwithstanding my high-level view in this foreword: the thesis is a scientific paper containing excellent reviews and new developments and results about both experimental B flavour physics and machine learning.

Karlsruhe, Germany
July 2018

Prof. Feindt Michael

# Acknowledgements

I thank all the people who contributed to the success of this work.

> If I have seen further it is by standing on the shoulders of giants.
>
> —Isaac Newton

I can look back on 4 wonderful years at the Institute of Experimental Particle Physics. Therefore, I would like to thank my colleagues and friends for the wonderful time and the creative environment. Because of you, I looked forward to every single day at the institute and learned something new on each of these days.

I would like to mention the technical contributions of my colleagues separately here and would like to express my special thanks to them in doing so.

I got the idea for a Belle to Belle II data conversion during a discussion with Dr. Bastian Kronenbitter. The starting point for the `belle_legacy` library was a cleaned up version of the `Belle Analysis Software Framework`, which I received from Dr. Ryosuke Itoh.

Dr. Martin Ritter supported me in integrating this library into BASF2 and connecting it to the `Belle II Condition Database`. The `b2bii` package was developed in cooperation with the Belle II `b2bii` task force under the direction of Dr. Anze Zupanc. Among the first users who were particularly important for feedback were Felix Metzner, Moritz Gelb and Markus Prim.

My knowledge of multivariate methods was strongly influenced by Prof. Dr. Michael Feindt. The original ideas for the data-driven methods and for the further development of sPlot technique originate from him. The first implementation of sPlot technology in the Belle II software framework was developed during Benjamin Lipp's bachelor's thesis. I first heard the term deep learning from Dr. Martin Heck. Essential contributions to the integration of deep learning technologies into the MVA package were made during Jochen Gemmler's master's thesis and doctoral thesis, and Dennis Weyland's master's thesis. Nils Braun and Oliver Frost were a driving force in integrating the MVA package into the reconstruction software.

The first version of the FEI was strongly influenced by the full reconstruction, which was developed by many doctoral students at the EKP between 2008 and 2012. This first version was developed in close cooperation with Dr. Christian Pulvermacher. I received many suggestions and the idea for the specific FEI from Dr. Martin Heck. The calibration of the FEI on Belle data was carried out by Judith Schwab in her master's thesis. The first application of the FEI on Belle data was carried out by Felix Metzner in his master's thesis.

The analysis $B^+ \to \tau^+ \nu_\tau$ is based on the doctoral theses of Dr. Bastian Kronenbitter, Dr. Johannes Grygier and Dr. Michael Ziegler.

The content and language of this work were corrected: Nils Braun, Moritz Gelb, Jochen Gemmler, Dr. Pablo Goldenzweig, Dr. Thomas Hauth, Dr. Martin Heck, Felix Metzner, Dr. William Sutcliffe and Judith Schwab.

Equivalent to these professional contributions, the success of this work is based on the support of my supervisors.

I thank Dr. Martin Heck, who has taken over as deputy head of the Feindt working group and has continued the group with great commitment. I thank Prof. Dr. Güter Quast for taking over the correferat and especially for the motivating discussions. I thank my supervisor and lecturer Prof. Dr. Michael Feindt. During my time as a doctoral student, he has enabled me to gain many valuable experiences that I could not have gained in any other working group.

# Contents

# Chapter 1
# Introduction

Physics deals with the development of quantitative models that describe nature. The most fundamental and precise models at present are: general relativity, which describes gravity as an interaction between the curvature of space-time and matter; and the Standard Model of particle physics (SM), which describes matter and the remaining electromagnetic, weak and strong interaction in the form of relativistic quantum fields.

The current high-energy physics (HEP) research focuses on the precise determination of the 19 free parameters of the SM and the search for new physics phenomena beyond the SM.

The Belle II experiment is part of this effort. It is located at the SuperKEKB electron-positron collider in Tsukuba, Japan. It is designed to perform a wide range of high-precision measurements in all fields of heavy flavour physics, including: B meson decays; $B_s^0$ meson decays; charm physics; $\tau$ lepton physics; hadron spectroscopy; and pure electroweak measurements. These measurements will constrain the parameter space of the SM as well as some of its extensions.

This work focuses on the development of software, in particular machine learning algorithms, to advance scientific progress, and to enable and improve a wide range of physics measurements at Belle II. This thesis summarizes my contributions to the Belle experiment and its successor the Belle II experiment.

Four major topics are covered. The conversion of the data recorded by the Belle experiment into the new data-format used by Belle II in Chap. 2. The integration of state-of-the-art machine learning algorithms and novel data analysis techniques into the Belle II Software Framework (BASF2) in Chap. 3. The development of the `Full Event Interpretation` exclusive tagging algorithm, which is unique to the Belle II experiment in Chap. 4. And the validation of the entire analysis software stack using the benchmark measurement of the branching fraction of the rare decay $B \to \tau\nu_\tau$ in Chap. 5.



T. Keck, *Machine Learning at the Belle II Experiment*, Springer Theses,
https://doi.org/10.1007/978-3-319-98249-6_1

# Chapter 2
# From Belle to Belle II

During this thesis, the full $\Upsilon(4S)$ dataset of the Belle experiment and large amounts of the available Monte Carlo data were converted into the new data-format used by Belle II. Therefore, it is possible to evaluate the reliability and performance of the newly developed analysis methods, in particular data-driven techniques, before a comparable dataset from the Belle II experiment is available.

In the following I give a brief overview of the Belle experiment (Sect. 2.1), its successor the Belle II experiment (Sect. 2.2) and describe the technical aspects of converting the Belle dataset into the Belle II data-format (Sect. 2.3).

## 2.1 The Belle Experiment

From June 1st 1999 until June 30th 2010, the Belle experiment recorded 988 $fb^{-1}$ of data at the KEKB asymmetric $e^+e^-$ collider [1].

This summary of the KEKB accelerator and the Belle detector is based on [1, 2]. Sections 2.1.1 and 2.1.2 are adapted from my master's thesis [3].

### *2.1.1 KEKB Accelerator*

The KEKB collider was dedicated to B physics and operated in the energy range of the $\Upsilon$ resonances. Its asymmetric beam energies induced a Lorentz boost $\beta\gamma = 0.42$ of the center of mass frame relative to the laboratory system, enabling the precise observation of the time evolution of B meson decays. During its runtime between 1998 and 2010 KEKB achieved the highest instantaneous luminosity of $2.1 \times 10^{34}\,cm^{-2}\,s^{-1}$ ever achieved by a collider [1, Sect. 1.3]. The machine parameters of KEKB are summarized in Table 2.2.



T. Keck, *Machine Learning at the Belle II Experiment*, Springer Theses,
https://doi.org/10.1007/978-3-319-98249-6_2

### 2.1.2 Belle Detector

The sole interaction point (IP) was surrounded by the Belle detector to detect and identify particles produced by the collisions. Like the accelerator, the detector was specifically designed for the precise observation of B meson decays. This includes precise measurement of secondary vertices and good particle identification capabilities.

Figure 2.1 shows a schematic side view of the Belle detector.

Going outwards from the interaction point the Belle detector consisted of:

- a double-walled Beryllium beam pipe with a radius of 20 mm cooled by He gas;
- radiation-hard Bismuth Germanate crystals used as an extreme forward calorimeter (EFC) as well as a beam and luminosity monitor;
- four layers of double-sided Silicon strip detectors (SVD) for precise vertex detection;
- a central drift chamber (CDC), which measured momentum and energy loss of charged particles;
- an Aerogel Cherenkov counter (ACC) system for particle identification (PID);
- a time-of-flight (TOF) detector system with plastic scintillation counters;
- a segmented array of CsI (Tl) crystals with silicon photodiode readout for electromagnetic calorimetry (ECL);
- a superconducting solenoid which provided a homogeneous magnetic field of 1.5 T;
- and an iron support structure, which served as the return path of the magnetic flux and was instrumented with glass-electrode resistive plate counters for $K_L$ and muon detection (KLM).

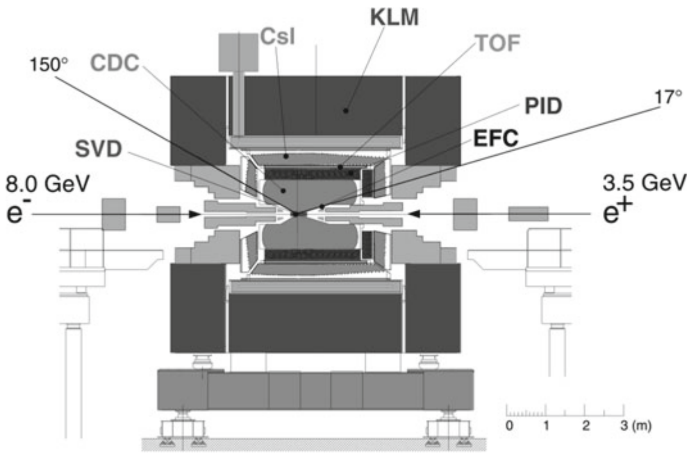


**Fig. 2.1** Side view of the Belle detector. Adapted from [2]

**Table 2.1** Summary of the integrated luminosity collected by Belle, broken down by center-of-mass energy. Adapted from [1]

| Resonance | On-resonance luminosity ($fb^{-1}$) | Off-resonance luminosity ($fb^{-1}$) |
|---|---|---|
| $\Upsilon(5S)$ | 121.4 | 1.7 |
| $\Upsilon(4S)$ – SVD1 | 140.0 | 15.6 |
| $\Upsilon(4S)$ – SVD2 | 571.0 | 73.8 |
| $\Upsilon(3S)$ | 2.9 | 0.2 |
| $\Upsilon(2S)$ | 24.9 | 1.7 |
| $\Upsilon(1S)$ | 5.7 | 1.8 |
| Scan $> \Upsilon(4S)$ | n/a | 27.6 |

### *2.1.3 Recorded Dataset*

The Belle dataset is grouped into 31 experiments[1]; each experiment marks the time period between two major shutdowns. Experiments 7 to 27 were recorded with the first Silicon Vertex Detector (SVD1). The remainder was recorded with the second (SVD2) detector and, in addition, was reprocessed in 2009 with an improved version of the reconstruction software. Each experiment is further subdivided into a number of runs.

Most of the data was recorded at the center-of-mass energy of the $\Upsilon(4S)$ resonance. In addition, data was also recorded at the $\Upsilon(1S)$, $\Upsilon(2S)$, $\Upsilon(3S)$ and $\Upsilon(5S)$ resonances. Apart from that, off-resonance data 60 MeV below the resonances was collected to estimate the continuum background from data instead of Monte Carlo simulation. Table 2.1 shows a summary of the integrated luminosity collected by Belle, broken down by the center-of-mass energy.

The raw data coming from the detector was calibrated, reconstructed and stored on tape using PANTHER-based[2] data summary tape (DST) files. After each experiment the calibration constants were recomputed by detector experts or computed directly from data, and stored in the Belle Condition Database (based on PostgreSQL). Finally, the data of the completed experiment was reprocessed and stored in a compact form called mDST files.[3]

For each experiment, ten times the real integrated luminosity in $b\bar{b}$ events and six times that in continuum events were simulated using EvtGen and GEANT3, and reconstructed with the same software as was used for the detector data.

---

[1]Enumerated from 7 to 73 using only odd numbers and skipping the numbers 29, 57 and 59 for reasons unknown to the author.

[2]The PANTHER format consists of tables, which are compressed by the zlib library. The table formats are defined by ASCII header files.

[3]A reduced and compressed form of the data summary tape files.

**Table 2.2** Achieved parameters of KEKB and design parameters of SuperKEKB [4]

| Machine Parameter | KEKB | | SuperKEKB | |
|---|---|---|---|---|
| | $e^-$ | $e^+$ | $e^-$ | $e^+$ |
| Beam current (A) | 1.64 | 1.19 | 3.60 | 2.61 |
| Energy (GeV) ($E_{HER}/E_{LER}$) | 8.0 | 3.5 | 7.0 | 4.0 |
| $\beta_y^*$ (mm) | 5.9 | 5.9 | 0.27 | 0.41 |
| Crossing angle (mrad) | 22 | | 83 | |
| Beam lifetime (min) | 200 | 150 | 10 | 10 |
| Luminosity ($10^{34}$ cm$^{-2}$ s$^{-1}$) | 2.11 | | 80 | |

## 2.2 The Belle II Experiment

This summary of the SuperKEKB accelerator and the Belle II detector is based on the detailed description in the technical design report [4] and is an updated and condensed version of a summary presented in my master's thesis [3].

### *2.2.1 SuperKEKB Accelerator*

The accelerator was shut down in June 2010 and upgraded to SuperKEKB to increase the instantaneous luminosity to $8 \times 10^{35}$ cm$^{-2}$ s$^{-1}$, which is 40 times the peak instantaneous luminosity of KEKB [5, Sect. 2].

The higher instantaneous luminosity is obtained by adopting the Nano-Beam scheme [5, Sect. 2], which requires a larger crossing angle to fit the final focusing magnets [4, Sect. 3.1]. Moreover, the beam current in both rings is doubled. The beam energy asymmetry was reduced to mitigate the shortened beam lifetime due to the Touschek effect.[4] In consequence the Lorentz boost is reduced to $\beta\gamma = 0.28$. The relevant machine parameters of KEKB and SuperKEKB are summarized in Table 2.2.

### *2.2.2 Belle II Detector*

The original detector is currently upgraded to match the higher instantaneous luminosity of SuperKEKB. The most important objectives for the upgraded detector are higher physics and background rate tolerance, better physics performance despite smaller Lorentz boost, and improved radiation hardness [5].

[4]The Touschek effect is a loss mechanism due to large angle coulomb scattering inside a bunch.

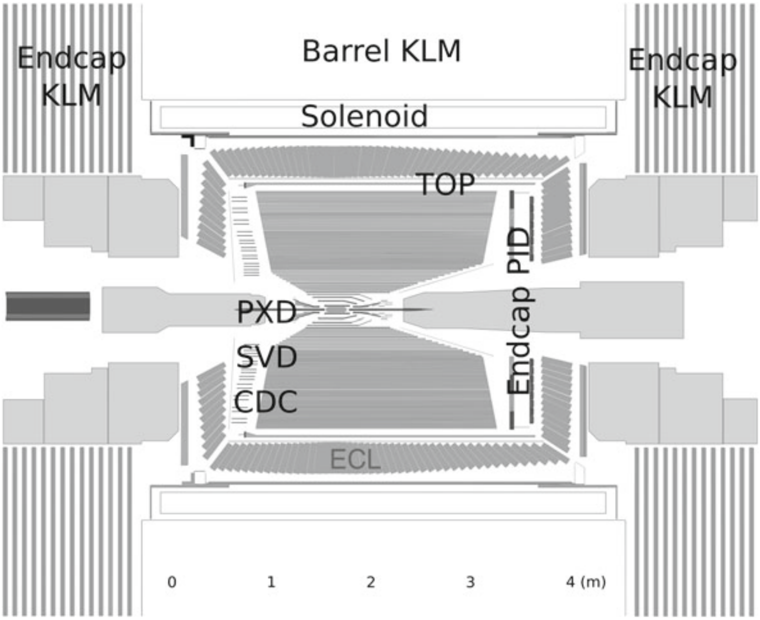


**Fig. 2.2** Side view of the upgraded Belle II detector [6]

Figure 2.2 shows the Belle II detector.

Going outwards from the interaction point the Belle II detector consists of:

- a double-walled Beryllium beam pipe with a radius of 12 mm cooled by paraffin;
- a pixel detector based on the DEPFET[5] technology (PXD) for precise vertex detection;
- four layers of double-sided silicon strip detectors covering the full $17° - 150°$ acceptance of the Belle II detector (SVD) to extrapolate the tracks reconstructed in the CDC to the PXD and to reconstruct low-momentum tracks;
- a central drift chamber (CDC), which measures momentum and energy loss of charged particles, and provides a fast trigger signal for the Level-1 (L1) trigger system;
- a proximity-focusing[6] Aerogel Ring-Imaging Cherenkov detector (ARICH) for particle identification (PID) in the forward end-cap;
- a time-of-propagation counter (TOP) in the barrel region using an array of 16 quartz bars between the outer CDC cover and the calorimeter's inner surface providing PID information and a timing signal with a resolution of a few nanoseconds to the trigger system;

---

[5]DEPleted Field Effect Transistor.

[6]An increasing refractive index is used to reduce the spread of the ring image due to emission point uncertainty.

- a segmented array of CsI(Tl) crystals in the barrel region and pure CsI crystals in the end-caps with photodiode readout for electromagnetic calorimetry (ECL);
- a superconducting solenoid which provides a homogeneous magnetic field of approximately 1.5 T;
- and an iron support structure, which serves as the return path of the magnetic flux, and is instrumented with glass-electrode resistive plate counters (RPC) in the barrel region and scintillator strip in the end-caps for $K_L$ and muon detection (KLM).

### *2.2.3 Anticipated Dataset*

By 2025, Belle II will record $50\,\mathrm{ab}^{-1}$ of data, which corresponds to 50 times the integrated luminosity of Belle. The current (March 17th 2017) luminosity projection is shown in Fig. 2.3.

#### 2.2.3.1 Data Acquisition

Belle II uses a two-level trigger system, with an FPGA-based Level-1 (L1) trigger decision and a high level trigger (HLT) farm.

At the design luminosity the nominal average L1 trigger rate is expected to be up to 30 kHz. For each trigger signal the data is read out by the data acquisition system

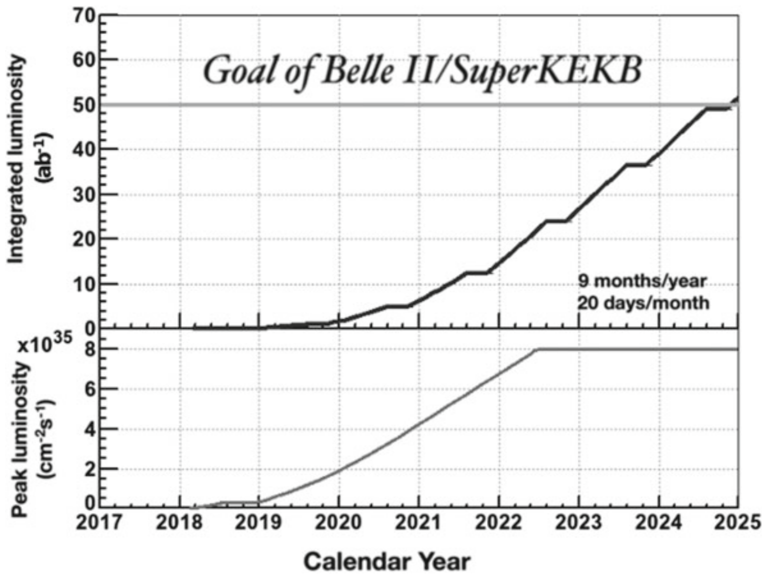


**Fig. 2.3** SuperKEKB integrated and instantaneous luminosity projection [7] (March 17th 2017)