

# Trigger and Reconstruction Farms in the HERA-*B* Experiment and Algorithms for a Third Level Trigger

## D I S S E R T A T I O N

zur Erlangung des akademischen Grades  
doctor rerum naturalium  
(dr. rer. nat.)  
im Fach Physik

eingereicht an der  
Mathematisch-Naturwissenschaftlichen Fakultät I  
Humboldt-Universität zu Berlin

von  
Herrn Dipl.-Phys. Ullrich Schwanke  
geboren am 23.07.1971 in Berlin

Präsident der Humboldt-Universität zu Berlin:  
Prof. Dr. J. Mlynek

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät I:  
Prof. Dr. B. Ronacher

Gutachter:

1. Prof. Dr. H. Kolanoski
2. Prof. Dr. T. Lohse
3. Dr. R. Waldi

eingereicht am: 31. Juli 2000  
Tag der mündlichen Prüfung: 30. Oktober 2000

## Abstract

The HERA- $B$  experiment at Deutsches Elektronen-Synchrotron (DESY) Hamburg aims at investigating the physics of particles containing  $b$  quarks. The experiment focusses on measuring CP violation in the system of neutral  $B$  mesons. It is expected that the precise determination of the CP asymmetry in the channel  $B^0(\bar{B}^0) \rightarrow J/\psi K_S^0$  will have an impact on the further development of the Standard Model of Elementary Particle Physics and cosmological theories.

The HERA- $B$  experiment uses the proton beam of the HERA storage ring in fixed-target mode.  $B$  hadrons are produced in pairs when protons from the beam halo interact with target nuclei. The interactions are recorded by a forward-spectrometer with roughly 600.000 readout channels. At the HERA- $B$  centre-of-mass energy of 42.6 GeV, the  $b\bar{b}$  cross section is only a tiny fraction of the total inelastic cross section. Only one in about  $10^6$  events contains  $b$  quarks, which turns the selection of signal events into a particular challenge. The selection is accomplished by a four-stage data acquisition and trigger system reducing the event rate from 10 MHz to about 20 Hz. Besides custom-made electronics, several hundreds of PCs are used in the trigger system. The computers are arranged in two so-called PC farms with more than 200 processors each. The PC farms provide the computing capacity for trigger decisions and the prompt analysis of event data. One farm executes fast trigger programs with a computing time of 1–100 ms per event. The other farm performs online reconstruction of the events before data are archived on tape. The computing time per event is in the range of several seconds.

This thesis covers two topics. In the beginning, the technical implementation of the trigger and the reconstruction farm are described. In doing so, emphasis is put on the software systems which make calibration data available to the farms and which provide a centralised view on the results of the executing processes.

The principal part of this thesis deals with algorithms for a Third Level Trigger. This trigger is to come into operation on the trigger farm together with existing programs. Processes of the type  $B^0(\bar{B}^0) \rightarrow J/\psi X$  have a very clean signature when the  $J/\psi$  decays to a  $e^+e^-$  or  $\mu^+\mu^-$  pair. The trigger system attempts to identify two unlike-sign leptons of the same flavour whose invariant mass matches the  $J/\psi$ . In later steps, the tracks are required to originate from a common vertex close to the target. It is assumed that these kinematic constraints are sufficient to pick out events of this type among the copious background processes. In contrast, the Third Level Trigger is to be applied to signal processes with fewer kinematic constraints. Such events occur for example when two  $B$  mesons, which were created in a proton-target collision, decay semileptonically. The trigger system selects merely the two leptons which do not originate from a common vertex in this case. The Third Level Trigger has 100 ms at its disposal to extract further criteria from the data which can serve to distinguish between signal and background events.

This thesis investigates with the aid of Monte-Carlo simulations how the data of the experiment's silicon vertex detector can contribute to the decisions of a Third Level Trigger. The trigger aims at reconstructing tracks from the decay cascade of  $B$  mesons in addition to the leptons selected by the preceding trigger levels. A fast pattern recognition for the vertex detector demonstrates that the reconstruction of all tracks and the application of trigger algorithms are possible within the given time slot of 100 ms. The determination of track parameters in the target region exploits the Kalman-filter method to account for the multiple scattering of particles in the detector material. The application of this method is, however, made difficult by two facts. First, the momentum of the reconstructed tracks is not known. And, second,

the material distribution in the detector cannot be taken into consideration in detail due to timing limitations. Adequate approximations for the momentum and the material traversed by a particle help to accomplish a sufficient accuracy of the track parameters. The reconstructed tracks constitute the starting point of several trigger algorithms, whose suitability to select signal events is investigated. Our studies indicate that the reconstruction of tracks with large impact parameters is a more promising approach than a search for secondary vertices.

**Keywords:**

HERA-*B* experiment, PC Cluster, pattern recognition, trigger algorithms

## Zusammenfassung

Das HERA- $B$ -Experiment am Deutschen Elektronen-Synchrotron (DESY) in Hamburg dient der Untersuchung der Physik von Teilchen, die  $b$ -Quarks enthalten. Der Schwerpunkt des Experimentes liegt auf der Messung der CP-Verletzung im System der neutralen  $B$ -Mesonen. Es wird erwartet, dass die präzise Bestimmung der CP-Asymmetrie im Zerfallskanal  $B^0(\bar{B}^0) \rightarrow J/\psi K_S^0$  großen Einfluss auf die Weiterentwicklung des Standardmodells der Elementarteilchenphysik und gängiger kosmologischer Theorien haben wird.

Das HERA- $B$ -Experiment nutzt den Protonenstrahl des HERA-Ringes, um in Kollisionen mit einem feststehenden Target paarweise  $B$ -Hadronen zu erzeugen. Die Wechselwirkungen werden in einem Vorwärtsspektrometer mit etwa 600.000 Auslesekanälen nachgewiesen. Aufgrund der relativ niedrigen Schwerpunktsenergie von 41.6 GeV sind Ereignisse mit  $b$ -Quarks im Vergleich zu Wechselwirkungen mit leichteren Quarks um etwa sechs Größenordnungen unterdrückt. Die Selektion von Signalereignissen stellt daher eine besondere Herausforderung dar. Sie wird von einem vierstufigen Datennahme- und Triggersystem übernommen, das die Ereignisrate von 10 MHz auf etwa 20 Hz reduziert. Neben speziell entwickelter Elektronik werden im Triggersystem mehrere hundert handelsübliche PCs eingesetzt. Die Computer sind in zwei so genannten PC-Farmen mit jeweils mehr als 200 Prozessoren angeordnet, die die Rechenkapazität für Triggerentscheidungen und die prompte Analyse der Ereignisdaten zur Verfügung stellen. Auf der einen Farm laufen schnelle Triggerprogramme mit einer Rechenzeit von etwa 1–100 ms pro Ereignis ab. Die andere Farm rekonstruiert die Ereignisse online, bevor die Daten auf Band dauerhaft archiviert werden. Die pro Ereignis aufgewandte Rechenzeit liegt dabei im Bereich einiger Sekunden.

Die vorliegende Arbeit behandelt zwei Themenkreise. Einerseits wird die technische Umsetzung der Trigger- und der Rekonstruktionsfarm beschrieben. Besonderes Augenmerk liegt dabei auf den Software-Systemen, die den Farmen erforderliche Kalibrationsdaten verfügbar machen und die zentrale Überwachung der Ergebnisse der ablaufenden Programme gestatten.

Der Hauptteil der Arbeit beschäftigt sich mit Algorithmen für eine dritte Triggerstufe, die zusätzlich zu existierenden Programmen auf der Triggerfarm zum Einsatz kommen sollen. Der Zerfall  $B^0(\bar{B}^0) \rightarrow J/\psi X$  hat eine sehr klare Signatur, wenn das  $J/\psi$  in ein  $e^+e^-$ - oder  $\mu^+\mu^-$ -Paar zerfällt. Im Triggersystem wird nach einem Paar entgegengesetzt geladener Leptonen des gleichen Typs gesucht, deren invariante Masse der des  $J/\psi$  entspricht und deren Spuren von einem gemeinsamen Vertex in der Nähe des Targets ausgehen. Es wird davon ausgegangen, dass die Ausnutzung aller kinematischen Zwangsbedingungen ausreicht, um diesen Zerfallskanal klar von Untergründereignissen zu trennen. Die dritte Triggerstufe soll dagegen auf Signalprozesse mit weniger kinematischen Beschränkungen angewandt werden. Solche Ereignisse entstehen zum Beispiel dann, wenn zwei in der Proton-Target-Kollision erzeugte  $B$ -Mesonen semileptonisch zerfallen. Das Triggersystem selektiert lediglich die beiden Leptonen, die aber hier nicht von einem gemeinsamen Vertex kommen. Die dritte Triggerstufe soll für derartige Zerfallstopologien innerhalb von 100 ms pro Ereignis weitere Kriterien zur Unterscheidung von Signal- und Untergrundprozessen aus den Daten extrahieren.

In der Arbeit wird anhand von Monte-Carlo-Studien untersucht, inwieweit die Daten des Silizium-Vertexdetektors des Experimentes zur Entscheidungsfindung einer dritten Triggerstufe beitragen können. Dabei wird die Rekonstruktion von Spuren aus der Zerfallskaskade der  $B$ -Hadronen zusätzlich zu den von der vorhergehenden Triggerstufe selektierten Leptonen angestrebt. Mithilfe einer schnellen Mustererkennung für den Vertexdetektor wird gezeigt, dass

das Auffinden aller Spuren und die Anwendung von Triggeralgorithmen innerhalb des vorgegebenen Zeitfensters von 100 ms möglich sind. Die Bestimmung der Spurparameter nahe der Targetregion macht von der Methode des Kalman-Filters Gebrauch, um der Vielfachstreuung im Detektormaterial Rechnung zu tragen. Dabei tritt das Problem auf, dass weder der Impuls der gefundenen Spuren bekannt ist, noch die Materialverteilung im Vertexdetektor aus Zeitgründen in aller Strenge berücksichtigt werden kann. Durch geeignete Näherungen gelingt es, eine ausreichende Genauigkeit für die Spurparameter zu erreichen. Die aufgefundenen Teilchen bilden den Ausgangspunkt für Triggeralgorithmen. Hierbei wird untersucht, welche Methoden am besten geeignet sind, um Signal- und Untergründereignisse voneinander zu trennen. Es erweist sich, dass das Auffinden von Spuren mit großem Impaktparameter aussichtsreichere Ansätze als eine Suche nach Sekundärvertices bietet.

**Schlagwörter:**

HERA-*B*-Experiment, PC Cluster, Mustererkennung, Triggeralgorithmen

# Acknowledgements

This thesis would not have been possible without the support of many people whom I would like to mention in this place. I thank my supervisor Prof. H. Kolanoski for the good collaboration and the numerous discussions that we had throughout the years. Many of his questions, hints, and suggestions proved to be fruitful for the work on the TLT studies. When looking back, I also remember vividly his lectures at the Humboldt-University, which were a great chance for me to prepare and conduct exercises for students.

I am indebted to Stefan Scharein who shared with me his experience on higher level triggers. His concise implementation of the TLT software as C++ classes made it easy to understand the code. Within a few weeks I could make own contributions to the TLT track finding. Our critical review of the existing class structure and the algorithms was a phase of intensive and enjoyable work. We were rewarded by a much faster and simpler to understand tracking program. Our discussions helped to formulate the TLT applications much clearer than it would have been possible for him or me alone.

I would like to express my gratitude to all other members of the HERA-*B* farm group, notably to Andreas Gellrich, Sun Fang, and Peter Wegner. It was really a pleasure to work, to discuss, and (last but not least) to celebrate with all of you. The basis idea of RHP and its name are due to Andreas who helped a lot to promote the package in the HERA-*B* collaboration and to make it a success. I received generous support from Andreas when discussing ideas for the TLT at various meetings at DESY Hamburg. He was the first to check what I had written for my thesis and gave many good comments. My special thanks to my room-mate Sun Fang with whom I worked on the slow-control software for the reconstruction farm and its WWW interface. There is no doubt that the concentrated atmosphere in our office was important for our daily work. Fortunately, this didn't prevent us from learning a lot from each other, be it about the situation in China or the rarely used features of C++. Peter Wegner provided me with valuable information that only computer centre people can give to physicists. It was instructive to learn about other available software technologies and packages for solving the problems that we encountered in connection with the online reconstruction farm.

I would like to thank also all the other friendly people at DESY Zeuthen and Humboldt-University who were just there when I had questions. Sigmund Nowak was a constant source of information about the details of the silicon vertex detector geometry and digitisation. His patience helped a lot when I was busy finding out that the amount of material in the SVD was strongly overestimated in the simulation. Some efficient discussions with Prof. T. Lohse about TLT problems proved to be very fruitful. I profited a lot from Jenny Ivarsson and Alexandre Somov who were always willing to share their ample experience concerning ARTE with me. Last but not least, I would like to thank Gerhard Bohm and Kristal Mauritz for proof-reading parts of this thesis.

# Contents

<b>Introduction</b>	<b>1</b>
<b>1 CP Violation</b>	<b>3</b>
1.1 CP Violation in the Standard Model . . . . .	4
1.2 Direct CP Violation . . . . .	6
1.3 CP Violation and Mixing . . . . .	7
1.4 The Golden Decay . . . . .	10
1.5 Other Angles of the Unitarity Triangle . . . . .	11
<b>2 The HERA-<i>B</i> Experiment</b>	<b>13</b>
2.1 Production of <i>b</i> quark Pairs . . . . .	13
2.2 Flavour Tagging . . . . .	14
2.3 Experimental Requirements . . . . .	16
2.4 Overview of the HERA- <i>B</i> Detector . . . . .	17
2.5 Wire Target . . . . .	18
2.6 Silicon Vertex Detektor . . . . .	19
2.6.1 Silicon Strip Detectors . . . . .	20
2.6.2 Detector Geometry . . . . .	21
2.6.3 Vertex Vessel . . . . .	23
2.6.4 Detector Modules . . . . .	25
2.7 Tracking System . . . . .	26
2.8 Particle Identification Devices . . . . .	27
<b>3 Data Acquisition and Trigger System</b>	<b>31</b>
3.1 Trigger Strategy . . . . .	32
3.2 Overview of the Trigger System . . . . .	33
3.3 Pretriggers . . . . .	35
3.4 First Level Trigger . . . . .	37
3.5 Second Level Trigger . . . . .	39
3.5.1 Tracking behind the Magnet . . . . .	40
3.5.2 Magnet Traversal . . . . .	41
3.5.3 Tracking and Vertexing in the SVD . . . . .	42
3.6 Structure of the Data Acquisition System . . . . .	43
3.7 Implementation of the SLT/TLT as a Farm . . . . .	47
<b>4 Online Reconstruction Farm</b>	<b>49</b>
4.1 Farm Concept . . . . .	49
4.2 Hardware Implementation . . . . .	50

4.3	Software Environment . . . . .	51
4.4	Control Software . . . . .	52
4.5	Application Software . . . . .	54
4.6	Data Logging . . . . .	55
4.7	Monitoring Software . . . . .	55
	4.7.1 Concept . . . . .	56
	4.7.2 Data Quality Assessment . . . . .	57
4.8	Online Calibration and Alignment System . . . . .	58
	4.8.1 Concept . . . . .	58
	4.8.2 Calibration Scheme . . . . .	58
4.9	Slow-Control Software . . . . .	60
4.10	Farm Summary . . . . .	61
<b>5</b>	<b>Third Level Trigger Studies</b>	<b>62</b>
5.1	Strategy at the Third Level Trigger . . . . .	62
5.2	Silicon Third Level Trigger . . . . .	63
5.3	Software Environment . . . . .	65
	5.3.1 Event Generation . . . . .	66
	5.3.2 Detector Simulation . . . . .	66
5.4	Fast Track and Primary Vertex Reconstruction in the SVD . . . . .	67
	5.4.1 Concept of the Track Reconstruction . . . . .	69
	5.4.2 Spacepoint Reconstruction . . . . .	70
	5.4.3 Track Reconstruction by Track Following . . . . .	71
	5.4.4 Track Refit . . . . .	73
	5.4.5 Compatibility Analysis of Tracks . . . . .	79
	5.4.6 Primary Vertex Reconstruction . . . . .	79
	5.4.7 Performance and Timing of Algorithms . . . . .	80
5.5	Data Samples . . . . .	83
5.6	Trigger Simulation . . . . .	85
5.7	Impact Parameter Studies . . . . .	86
	5.7.1 Impact Parameter with respect to the Target . . . . .	87
	5.7.2 Impact Parameter with respect to the Primary Vertex . . . . .	91
	5.7.3 Impact Parameter Cuts . . . . .	94
5.8	Vertexing with SLT Leptons . . . . .	97
	5.8.1 Event Sample . . . . .	97
	5.8.2 Vertex Resolution . . . . .	99
	5.8.3 Vertex Recognition . . . . .	101
	5.8.4 Efficiency . . . . .	103
5.9	Track Trigger . . . . .	105
	5.9.1 Strategy of the Algorithm . . . . .	105
	5.9.2 Track Trigger Algorithm . . . . .	106
	5.9.3 Efficiency . . . . .	108
5.10	Conclusion and Outlook . . . . .	110
	5.10.1 Algorithms . . . . .	110
	5.10.2 Further Studies . . . . .	111
	5.10.3 Technical Implementation of the TLT . . . . .	111



# List of Figures

1.1	The unitarity triangle . . . . .	5
1.2	Current knowledge of the unitarity triangle . . . . .	6
1.3	Box diagram for $B_d$ mixing . . . . .	8
1.4	Diagrams contributing to the decay $B \rightarrow J/\psi K_S^0$ . . . . .	10
1.5	Diagrams of the decays $B^0 \rightarrow \pi^+\pi^-$ and $B_s^0 \rightarrow \rho^0 K_S^0$ . . . . .	11
2.1	Decay scheme of a golden event . . . . .	15
2.2	Top view of the HERA- $B$ spectrometer . . . . .	18
2.3	Wire target . . . . .	19
2.4	Silicon strip detectors . . . . .	21
2.5	View of the silicon vertex detector . . . . .	22
2.6	Vertex vessel . . . . .	24
2.7	RF bands . . . . .	24
2.8	Various views of vertex detector modules . . . . .	25
2.9	View of a micro-strip gaseous chamber . . . . .	26
2.10	Layout of the electromagnetic calorimeter . . . . .	28
2.11	Calorimeter module . . . . .	29
2.12	Pad chamber module . . . . .	30
3.1	Pretrigger sources . . . . .	35
3.2	Functional view of the FLT . . . . .	37
3.3	Illustration of the Slicer algorithm . . . . .	40
3.4	Illustration of the magnet traversal . . . . .	41
3.5	Structure of the DAQ system . . . . .	45
3.6	The SHARC cluster board . . . . .	46
3.7	Software on a trigger PC . . . . .	47
4.1	Topology of the reconstruction farm . . . . .	51
4.2	Processes of the control software . . . . .	53
4.3	Application software . . . . .	54
4.4	Monitoring scheme . . . . .	57
4.5	Online calibration system . . . . .	59
4.6	Slow-control system . . . . .	60
5.1	TLT approaches . . . . .	64
5.2	Illustration of the spacepoint algorithm . . . . .	69
5.3	Pattern recognition of tracks . . . . .	71
5.4	Multiple scattering . . . . .	75
5.5	Multiple scattering between two measurements . . . . .	77

5.6	Momentum estimation . . . . .	78
5.7	Track finding efficiency and timing of algorithms . . . . .	82
5.8	Track finding efficiency . . . . .	83
5.9	Definition of impact parameters . . . . .	87
5.10	Impact parameter distributions . . . . .	88
5.11	Impact parameter resolution . . . . .	90
5.12	Pull of impact parameters . . . . .	92
5.13	Resolution of the primary vertex . . . . .	93
5.14	Efficiency of impact parameter cuts . . . . .	96
5.15	The decay chain $B^0 \rightarrow D^{*-} l^+ \nu_l$ . . . . .	98
5.16	Accuracy of the lepton-pion vertex . . . . .	100
5.17	Accuracy of the lepton-kaon vertex . . . . .	101
5.18	Reconstructed DCAs and vertex positions . . . . .	102
5.19	Results of the vertex search . . . . .	104
5.20	$\chi^2$ per degree of freedom . . . . .	107
5.21	Distributions of cut quantities . . . . .	109
5.22	Efficiency of the track trigger algorithm . . . . .	109

# List of Tables

2.1	SVD geometry . . . . .	21
2.2	Positions of SVD superlayers . . . . .	23
3.1	Present and future $B$ experiments . . . . .	31
3.2	Overview of the four trigger levels . . . . .	34
3.3	FLT performance . . . . .	39
3.4	SLT algorithms . . . . .	40
3.5	Event topologies in the SVD . . . . .	43
5.1	SVD configuration in 1999/2000 . . . . .	66
5.2	Parameters of the SVD simulation . . . . .	67
5.3	Materials in the sensitive area of SVD counters . . . . .	76
5.4	Impact parameter resolutions . . . . .	91
5.5	Parameters of the primary vertex finding . . . . .	93
5.6	Properties of $B$ and $D$ mesons in the decay $B^0 \rightarrow D^{*-} l^+ \nu_l$ . . . . .	98
5.7	Properties of the decay particles in $B^0 \rightarrow D^{*-} l^+ \nu_l$ . . . . .	100

# Introduction

Several decades after its discovery in weak decays of neutral  $K$  mesons, CP violation is still subject to intensive experimental and theoretical research. CP violation appears as a difference between the partial decay widths of particles and their charge-conjugated and mirrored counterparts. It is believed that such an asymmetry between particles and antiparticles might shed light on the question why the Universe visible to us is dominated by matter rather than antimatter. The leading theory, the Standard Model of Particle Physics, relates the origin of CP violation to the Cabibbo-Kobayashi-Maskawa (CKM) mixing matrix, which transforms the eigenstates of quark flavour to the eigenstates of the weak interaction. At the present state of knowledge, the elements of the mixing matrix cannot be derived from first principles but are fundamental constants taken from experiments. In the Standard Model, a non-vanishing complex phase in the CKM matrix is the lone source of CP violation in weak interactions. The concise determination of the CKM matrix elements and in particular of this CP-violating phase is the goal of many experimental efforts. It allows to check whether the current understanding of particle interactions suffices to describe all observable phenomena. Such a test is of paramount importance since cosmological arguments suggest that the amount of CP violation given by the Standard model mechanism is not sufficient to explain the extent of the observed matter-antimatter asymmetry. A precise measurement of CP-violating effects might point to theories beyond the Standard model providing new sources of CP violation.

Besides the system of neutral  $K$  mesons, decays of  $D$  and  $B$  mesons provide a way to explore CP violation. The predicted effects are tiny for the  $D$  system, but  $B$  mesons should exhibit sizable CP asymmetries which can be accessed experimentally. The HERA- $B$  experiment at DESY Hamburg was designed to study the physics of mesons and baryons containing  $b$  quarks. It started data-taking in 2000. Its main goal is the measurement of CP violation in the so-called *gold-plated channel*

$$B^0(\bar{B}^0) \rightarrow J/\psi K_S^0 \rightarrow l^+l^- \pi^+\pi^- \quad (1)$$

and in

$$B^0(\bar{B}^0) \rightarrow \pi^+\pi^-. \quad (2)$$

The  $B$  hadrons are produced in fixed-target mode using the 920 GeV proton beam of the HERA accelerator. Up to eight target wires are inserted into the beam halo to cause proton-nucleon collisions. The resulting cross section for  $b\bar{b}$  production at the centre-of-mass energy of 41.6 GeV is not yet known precisely. Extrapolations from other experiments and theoretical calculations indicate a value of  $\mathcal{O}(10\text{ nb})$  which is six orders of magnitude lower than the total inelastic cross section. Facing the tiny branching fractions of the aforementioned channels ( $< 10^{-4}$ ), event rates in excess of 40 MHz are needed to perform a significant measurement within one year. Since the bunch frequency of the proton beam is about 10 MHz, 4–5 interactions are superimposed and result in a harsh hadronic environment with about 150 tracks per event. The events are recorded by a forward-spectrometer with 600.000 readout channels employing

novel radiation-hard detector technologies.

The data acquisition and trigger system poses a particular challenge due to the high initial event rate and the tiny signal-to-background ratio. It has to pick out events with  $b$  quarks among the much more copious interactions with light quarks only. The reduction of background events by six orders of magnitude has to be accomplished at high efficiency for signal processes without introducing significant downtime. The technical implementation of the trigger system exploits fast dedicated electronics and several hundreds of microprocessors of the type used in customary personal computers. The latter provide the computing power needed for trigger decisions and the prompt analysis of the selected events. The processors are arranged in their normal hardware environment in two so-called PC farms. The processors of the first farm run fast trigger algorithms with a duration of 1–100 ms per event. The second farm is scheduled to perform online reconstruction of the events before data is written to tape. The reconstruction takes several seconds per event.

In the initial phase of his work at DESY, the author was concerned with the hardware and software setup of the online reconstruction farm, which will be described. The main part of this thesis is devoted to preparing studies for an additional trigger stage, which is aiming at decay modes of  $B$  hadrons different from (1) and (2). This Third Level Trigger (TLT<sup>1</sup>) is to come into operation together with existing trigger programs on the PC farm mentioned earlier. It is proposed to use the data of the experiment's Silicon Vertex Detector (SVD) for rapid track finding. The tracks are the input to algorithms allowing to distinguish signal and background events.

The first two chapters of this thesis give a short overview of the physical background of the HERA- $B$  experiment and the detector. Among the introduced detector components, the silicon vertex detector is treated in more detail. The third chapter describes the data acquisition system and the algorithms used for triggering. It sets the frame for the TLT studies and gives reasons why such an additional stage may be required. The setup of the reconstruction farm is reviewed in the following chapter. The comprehension of this part is, however, not necessary to follow the fifth chapter describing the Monte Carlo studies for a third level trigger. The trigger studies detail algorithms for fast track and vertex finding in the silicon vertex detector and investigate how to exploit the obtained information for trigger decisions.

---

<sup>1</sup>Specific abbreviations are introduced at the first occurrence in the text. An alphabetical index can be found in the appendix.

# Chapter 1

## CP Violation

The search for symmetries of physical phenomena has proven to be a fruitful concept throughout the development of particle physics. It is intriguing that many surmised symmetries turned out to hold only approximately or to be violated completely. This is quite evident in the description of weak interactions of elementary particles. The discovery that parity  $P$  is not conserved in weak interactions came as a surprise, even though this possibility had been suggested by theorists. In the historic experiment by Wu et al. [1], it was observed that electrons from the decay of a  $^{60}\text{Co}$

$$^{60}\text{Co}(J = 5) \rightarrow ^{60}\text{Ni}^*(J = 4) + e^- + \bar{\nu}_e$$

were emitted preferentially in opposite direction to the nuclear spin  $J$ . Why should an interaction distinguish between particles and their mirrored counterparts? Parity violation and the characteristic property of neutrinos and antineutrinos to occur left- and right-handed only were successfully included when formulating the  $V-A$ -theory of weak interactions.

Yet another symmetry is broken when neutrinos are involved. Charge conjugation  $C$ , reversing all charge-like quantum numbers, applied to the decay of a  $\pi^+$

$$\pi^+ \rightarrow \mu_R^+ + \nu_{\mu,L} \xrightarrow{C} \pi^- \rightarrow \mu_R^- + \bar{\nu}_{\mu,L}$$

yields a right handed  $\mu_R^-$  (in the  $\pi$ 's rest frame) and a left-handed  $\bar{\nu}_{\mu,L}$  which does not occur in Nature. The combined application of  $C$  and  $P$  results in an observable reaction

$$\pi^+ \rightarrow \mu_R^+ + \nu_{\mu,L} \xrightarrow{CP} \pi^- \rightarrow \mu_L^- + \bar{\nu}_{\mu,R},$$

which has suggested that  $CP$  is a good symmetry in weak interactions. In 1964,  $CP$  was found to be broken as well, but on a much smaller scale than  $P$  and  $C$ . It was demonstrated [2] that the long-lived neutral kaon  $K_L^0$  decays to final states with two ( $CP=+1$ ) and three (dominantly  $CP=-1$ ) pions. Obviously the mass eigenstate  $K_L^0$ , which can be viewed as a superposition of the states  $|K^0\rangle$  and  $|\bar{K}^0\rangle$  produced in strong interactions

$$|K_L^0\rangle = p|K^0\rangle + q|\bar{K}^0\rangle \quad p, q \in \mathbb{C},$$

is not a simultaneous eigenstate of  $CP$ . Since then many research activities were devoted to the  $K$  system. Only very recently it became possible to disentangle the effects of particle mixing and  $CP$ -violating decay amplitudes [3, 4].

Mesons containing  $b$  quarks are one of few systems in nature where  $CP$  violation is expected to be visible as well. In some channels, the extracted quantities can be related directly to parameters of the Standard Model. The concise measurement of these quantities is of interest since it constitutes a stringent test for our understanding of particle interactions and might reveal novel and unexpected phenomena.

## 1.1 CP Violation in the Standard Model

The Standard Model accommodates CP violation by virtue of a Lagrangian which is not invariant under CP transformations. CP non-invariance arises when there are complex coupling constants in the theory [5]. Since coupling constants appear in products with particle fields, which can be changed by unobservable phases, one must look for couplings, that cannot be made real by an appropriate redefinition of fields. Such a situation can occur in the  $W$ -mediated transitions between *up*-type and *down*-type quarks. The states taking part in the weak interaction are not identical to the mass eigenstates, but are conventionally constructed by applying a unitary transformation matrix  $V_{CKM}$  (Cabibbo-Kobayashi-Maskawa matrix) to the *down*-type quarks:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = V_{CKM} \begin{pmatrix} d \\ s \\ b \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}.$$

The  $V_{jk}$  are complex numbers to be determined from experiment. For three generations of quarks, unitarity constraints reduce the number of parameters in  $V_{CKM}$  from 18 to 9. A phase transformation of the quark fields replaces  $V_{jk}$  by  $V_{jk}e^{i(\phi_j - \phi_k)}$ . Four such relative phases (two between both *up*-type and *down*-type fields) and a global phase result in four remaining free parameters. Three of the parameters are angles, one is a phase. At least one non-vanishing phase rendering  $V_{CKM}$  complex occurs for more than two quark families only. This fact was a strong motivation to extend the Cabibbo matrix to three families prior to the discovery of *beauty* and *top* in order to adopt CP violation in a natural way.

It is customary to write down  $V_{CKM}$  in an empirical parametrization introduced by Wolfenstein [6] using four real parameters  $\lambda$ ,  $A$ ,  $\rho$  and  $\eta$ , which are all of order unity. This parametrization expresses the relative strength of transitions between quark families in powers of  $\lambda$ . In lowest order,  $\lambda$  is equal to the Cabibbo angle  $\sin \theta_C \approx 0.22$  and one obtains the two-generation quark mixing matrix. The CKM matrix is extended to three families by setting  $V_{tb} \simeq 1$  in a very good approximation. The weaker  $b \rightarrow c$  transition is described as  $V_{cb} = 0.04 = A\lambda^2$ . With this definition, the expansion for  $V_{ts}$  also starts in  $\lambda^2$ . The two smallest elements  $V_{ub}$  and  $V_{td}$  are proportional to  $\lambda^3$  and contain the complex phase. Up to  $\mathcal{O}(\lambda^3)$  the CKM matrix is then given by:

$$\begin{pmatrix} 1 - \frac{\lambda^2}{2} & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \frac{\lambda^2}{2} & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}(\lambda^4). \quad (1.1)$$

To date, not all elements of  $V_{CKM}$  have been measured directly. The magnitudes of  $V_{ud}$  and  $V_{us}$  are known with an accuracy better than 1%.  $V_{cd}$  and  $V_{cs}$  are known on a level of some per cent and  $V_{cb}$  has been determined with an error of about 5%. These measurements translate into errors for  $\lambda$  and  $A$  of better than 1% and 5%, respectively. The magnitudes of  $V_{ub}$  and  $V_{td}$  can be extracted from measurements of charmless  $B$  meson decays and of the  $B^0$ - $\bar{B}^0$  mixing. The achieved precision is roughly 30% and implies that  $\rho$  and  $\eta$  have a high uncertainty.

The unitarity of  $V_{CKM}$  allows to write down relations between matrix elements, which can be confronted with data. The most interesting relation is

$$V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0, \quad (1.2)$$

where all terms in the sum are proportional to  $\lambda^3$  and hence of similar magnitude. The equation above involves the poorly known matrix elements  $V_{td}$  and  $V_{ub}$  and can be used to approach these

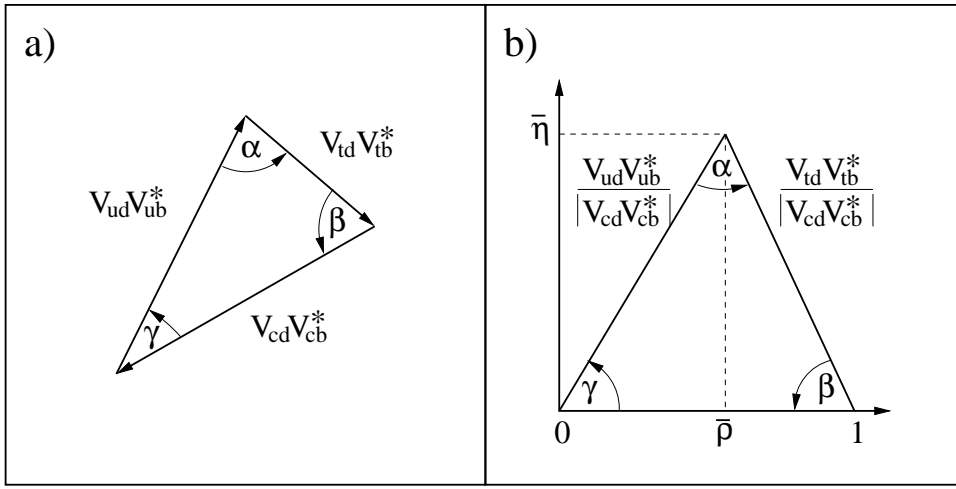


Figure 1.1: The unitarity triangle obtained by multiplying the first and third row of  $V_{CKM}$  (a). It assumes a simple form when divided by  $|V_{cd}V_{cb}^*|$  (b).

quantities. Equation (1.2) defines a triangle in the complex plane (Fig. 1.1) with angles  $\alpha$ ,  $\beta$  and  $\gamma$  given by

$$\alpha = \arg\left(-\frac{V_{td}V_{tb}^*}{V_{ud}V_{ub}^*}\right), \quad \beta = \arg\left(-\frac{V_{cd}V_{cb}^*}{V_{td}V_{tb}^*}\right), \quad \gamma = \arg\left(-\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*}\right). \quad (1.3)$$

When dividing Eq. (1.2) by  $|V_{cd}V_{cb}^*|$ , the unitarity triangle of Fig. 1.1 (b) emerges, which has a particularly simple shape. Using Eq. (1.1) up to  $\mathcal{O}(\lambda^3)$ , the tip of the triangle is given by  $(\bar{\rho}, \bar{\eta})$ , where  $\bar{\rho}$  and  $\bar{\eta}$  are connected to  $\rho$  and  $\eta$  via

$$\bar{\rho} = \rho\left(1 - \frac{\lambda^2}{2}\right) \quad \text{and} \quad \bar{\eta} = \eta\left(1 - \frac{\lambda^2}{2}\right).$$

The baseline of the triangle extends from  $(0, 0)$  to  $(1, 0)$  and the two other sides are given by  $\sqrt{\bar{\rho}^2 + \bar{\eta}^2}$  and  $\sqrt{(1 - \bar{\rho})^2 + \bar{\eta}^2}$ . The triangle contains all the information about CP violation in the Standard Model. If the triangle does not collapse into a line, i.e.  $\bar{\eta}$  is non-zero, CP is broken.

The present knowledge of the unitarity triangle is summarised in Fig. 1.2 [7]. The position of its tip is constrained by several measured quantities and their respective accuracy.

1. The measured ratio of

$$\frac{|V_{ub}|}{|V_{cb}|} = 0.093 \pm 0.016 = \lambda\sqrt{\rho^2 + \eta^2} = \lambda\frac{\sqrt{\bar{\rho}^2 + \bar{\eta}^2}}{\left(1 - \frac{\lambda^2}{2}\right)}$$

confines  $\bar{\rho}$  and  $\bar{\eta}$  to a circle with radius of about 0.4. The measurement error turns this circle into a ring.

2. The determination of the  $\epsilon_K$ -parameter in the system of neutral  $K$  mesons as

$$\epsilon_K = (2.280 \pm 0.019) \cdot 10^{-3}$$

defines hyperbolas in the range of which possible positions of the tip must fall.

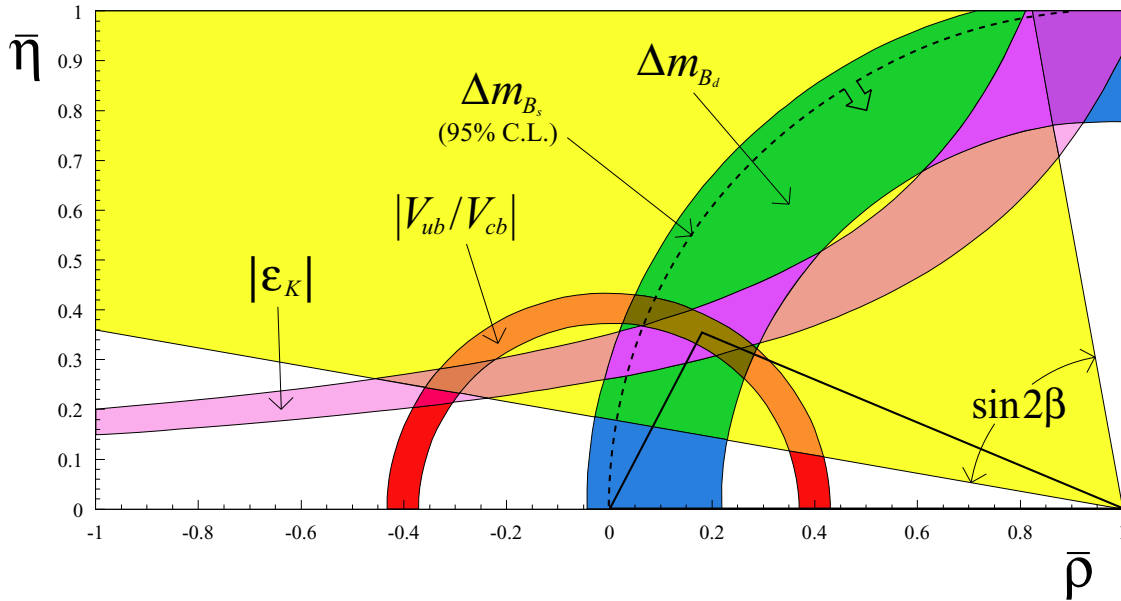


Figure 1.2: Current knowledge of the unitarity triangle [7]. The most probable shape of the triangle is indicated. The position of the tip is constrained by measured branching ratios, the  $B_d$ - $\bar{B}_d$  and  $B_s$ - $\bar{B}_s$  mixing, and CP violation measurements in the system of neutral  $K$  mesons. Except for  $\Delta m_{B_d}$ , the shaded areas denote 68 % C.L. regions.

3. The mass difference  $\Delta m_{B_d}$  of the mass eigenstates in the system of  $B_d$  mesons was determined in mixing experiments. The measured value of  $(0.471 \pm 0.010) \text{ ps}^{-1}$  cuts out a ring around  $(1, 0)$ .
4. The mixing of  $B_s$  mesons proceeds very rapidly and requires therefore an excellent resolution of vertex positions. To date, only a lower bound of

$$\Delta m_{B_s} \geq 14.3 \text{ ps}^{-1}$$

is known, which excludes a region in the  $\bar{\rho}$ - $\bar{\eta}$  plane.

There are many efforts to determine the angles and sides of the unitarity triangle to high precision. Once the triangle is overconstrained by measurements, one can check whether the Standard Model with three fermion generations can account for all observed effects. Any detected deviation is evidence for contributions from new physics beyond the Standard Model with possible additional sources of CP violation.

## 1.2 Direct CP Violation

An intuitive way to probe CP violation is to look for different decay widths of a process and its CP-mirrored counterpart, e.g in  $B^\pm \rightarrow \pi^\pm K^0$ . CP violation should show up as

$$\Gamma(B \rightarrow f) \neq \Gamma(\bar{B} \rightarrow \bar{f}).$$

(Here  $B$  stands for a generic  $B$  hadron of arbitrary charge.) The rate difference can be determined best as an asymmetry

$$a_{\text{asym}} = \frac{\Gamma(B \rightarrow f) - \Gamma(\bar{B} \rightarrow \bar{f})}{\Gamma(B \rightarrow f) + \Gamma(\bar{B} \rightarrow \bar{f})}, \quad (1.4)$$

where systematic uncertainties common to the measurement of both decay widths cancel. The amplitudes for both processes

$$A = \langle f | \mathcal{H} | B \rangle \quad \text{and} \quad \bar{A} = \langle \bar{f} | \mathcal{H} | \bar{B} \rangle$$

must have different magnitudes, i.e.

$$\left| \frac{\bar{A}}{A} \right| \neq 1,$$

to observe an asymmetry. This kind of CP violation is referred to as *direct* or as *CP violation in the decay*.

According to the Standard Model, the amplitudes  $A$  and  $\bar{A}$  differ only with respect to the sign of the involved weak phases. For the case of  $N$  contributing diagrams one can write

$$A = \sum_k^N A_k = \sum_k^N M_k e^{i\delta_k} e^{i\alpha_k} \quad \text{and} \quad \bar{A} = \sum_k^N \bar{A}_k = \sum_k^N M_k e^{-i\delta_k} e^{i\alpha_k}. \quad (1.5)$$

Here the  $\delta_k$  are weak phases changing sign when going to the CP-mirrored process while the strong phases  $\alpha_k$  remain unchanged. At least two amplitudes have to interfere in the above sums to generate an asymmetry in Eq. (1.4). With  $N=2$ ,  $\Gamma(B \rightarrow f)$  is proportional to  $|A_1 + A_2|^2$ :

$$\Gamma(B \rightarrow f) \propto A_1^2 + A_2^2 + 2\Re(A_1 A_2^*) = M_1^2 + M_2^2 + 2M_1 M_2 \cos(\phi_w + \phi_s),$$

with  $\phi_w = \delta_1 - \delta_2$  and  $\phi_s = \alpha_1 - \alpha_2$  being the difference of the weak and strong phases. For a  $\bar{B}$ ,  $\phi_w$  is changing sign

$$\Gamma(\bar{B} \rightarrow \bar{f}) \propto M_1^2 + M_2^2 + 2M_1 M_2 \cos(-\phi_w + \phi_s).$$

The resulting asymmetry is proportional to

$$M_1 M_2 \sin(\phi_w) \sin(\phi_s)$$

and can only be observed when the two amplitudes in the sum (1.5) contribute with different weak phases  $\delta_k$ . The interpretation of an observed asymmetry in terms of  $V_{CKM}$  is hampered by the presence of the strong phases  $\alpha_k$  which are in general not known. These hadronic uncertainties and the small magnitude of the expected asymmetries (<1%) make searches for direct CP violation in decays of charged or neutral  $B$  mesons difficult.

### 1.3 CP Violation and Mixing

In particular for charged  $B$  mesons, the mechanism described above is the only way to seek for CP violation. For pairs of neutral  $B$  mesons ( $B_d \bar{B}_d$ ,  $B_s \bar{B}_s$ ), related by a CP transformation, mixing is an important prerequisite when performing CP violation measurements. Time-evolved neutral  $B$  mesons are superpositions of flavour eigenstates due to flavour oscillations. These

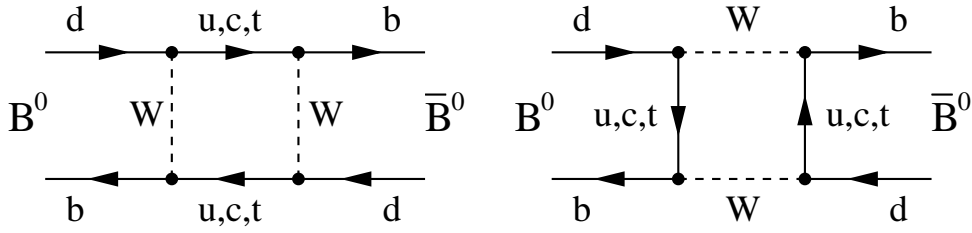


Figure 1.3: Box diagrams for  $B_d$ - $\bar{B}_d$  mixing.

oscillations proceed mainly via box diagrams (Fig. 1.3) with  $|\Delta B|=2$ . The mass eigenstates are defined by

$$\begin{aligned} |B_L\rangle &= p|B^0\rangle + q|\bar{B}^0\rangle \\ |B_H\rangle &= p|B^0\rangle - q|\bar{B}^0\rangle, \end{aligned} \quad (1.6)$$

where the subscripts  $L$  and  $H$  stand for *Light* and *Heavy*. When starting with a pure beam of  $B^0$  mesons, the probability that the time-evolved physical state  $B_{\text{phys}}(t)$  decays as  $B^0$  or  $\bar{B}^0$  varies in time as

$$P(B_{\text{phys}} \rightarrow B^0) = \frac{e^{-\Gamma t}}{2} \left( \cosh\left(\frac{\Delta\Gamma}{2}t\right) + \cos(\Delta m t) \right), \quad (1.7)$$

$$P(B_{\text{phys}} \rightarrow \bar{B}^0) = \frac{e^{-\Gamma t}}{2} \left( \cosh\left(\frac{\Delta\Gamma}{2}t\right) - \cos(\Delta m t) \right). \quad (1.8)$$

Here  $\Gamma$  is the average decay width of the two mass eigenstates,  $\Delta\Gamma$  their decay width difference and  $\Delta m = m_H - m_L$  their mass difference. For  $B_d$  mesons (to which we will restrict in the following) the decay width difference  $\Delta\Gamma$  is small. One can neglect  $\Delta\Gamma$  and set  $\cosh(\frac{\Delta\Gamma}{2}t) \simeq 1$ .

CP violation can occur simply because the mass eigenstates are no CP eigenstates. This is called *indirect* or *CP violation in mixing*. It is responsible for the majority of the observed effects in the system of neutral  $K$  mesons where the mixing probabilities  $|K^0\rangle \longleftrightarrow |\bar{K}^0\rangle$  differ strongly.

Formally, the mass eigenstates (1.6) can be made to be CP eigenstates when the difference between  $p$  and  $q$  is just a phase factor. Otherwise, the condition

$$|q/p| \neq 1 \quad (1.9)$$

holds and signals CP violation. CP violation could be observed for neutral  $B$  mesons through the asymmetry in semileptonic decays

$$a_{\text{sl}} = \frac{\Gamma(\bar{B}_{\text{phys}}(t) \rightarrow l^+ \nu_l X) - \Gamma(B_{\text{phys}}(t) \rightarrow l^- \bar{\nu}_l X)}{\Gamma(\bar{B}_{\text{phys}}(t) \rightarrow l^+ \nu_l X) + \Gamma(B_{\text{phys}}(t) \rightarrow l^- \bar{\nu}_l X)} = \frac{1 - |q/p|^4}{1 + |q/p|^4}. \quad (1.10)$$

The charge of the lepton determines whether the  $|B^0\rangle$  or the  $|\bar{B}^0\rangle$  admixture of the time-evolved state was responsible for the decay. However, it will be difficult to detect this kind of CP violation in decays of neutral  $B$  mesons. The deviation of  $|q/p|$  from unity is caused by decay modes common to  $B^0$  and  $\bar{B}^0$ . These have branching ratios smaller than  $10^{-3}$  and their contribution to the flavour oscillation is accordingly small. To a good approximation the ratio  $q/p$  is given by the diagrams in Fig. 1.3. Among these, the diagram with the heavy  $t$  quark in

the loop gives the dominating contribution. The vertices are proportional to  $V_{tb}^*$  and  $V_{td}$ . Using CKM parameters the ratio  $q/p$  can be expressed as

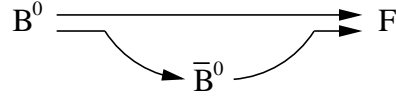
$$\frac{q}{p} = \frac{V_{tb}^* V_{td}}{V_{tb} V_{td}^*} = e^{-2i\beta},$$

where  $\beta$  is one of the angles in the unitarity triangle.

It is in general not straightforward to relate the results of measurements of direct and indirect CP violation to the fundamental parameters of the Standard Model. But there exists yet another method to determine the degree of CP violation. It is called *CP violation in the interference between decays with and without mixing* and provides a direct handle to the phases of the quantities  $q/p$  and  $\bar{A}/A$ . One considers the decay of neutral  $B$  mesons to a CP eigenstate  $F$  which can be reached by *both*  $B^0$  and  $\bar{B}^0$ . The relevant amplitudes are

$$A = \langle F | \mathcal{H} | B^0 \rangle \quad \text{and} \quad \bar{A} = \langle F | \mathcal{H} | \bar{B}^0 \rangle .$$

The extraction of a CP-violating phase is possible due to the interference of two different ways along which the decay to the CP eigenstate  $F$  can proceed. A  $B^0$ , for example, can decay directly to the state  $F$  or after oscillating into a  $\bar{B}^0$ :



In contrast to the interference of at least two diagrams with different strong phases as required in (1.5), no hadronic uncertainties obscure the interpretation from the outset.

The quantity relevant for the classification of this kind of CP violation is

$$\lambda = \frac{q}{p} \frac{\bar{A}}{A}.$$

CP is violated when  $\lambda \neq 1$ . The property  $q/p = e^{-2i\beta}$ , which made the measurement of the asymmetry (1.10) more difficult, results now in a considerable simplification. For processes, where in addition  $A$  and  $\bar{A}$  differ by just a phase, i.e.

$$\frac{\bar{A}}{A} = e^{2i\delta}, \tag{1.11}$$

the magnitude of  $\lambda$  is 1, but its imaginary part is

$$\Im(\lambda) = -\sin(2(\beta - \delta)).$$

The experimental technique involves the measurement of the time-dependent asymmetry of the number of  $B^0$  and  $\bar{B}^0$  mesons decaying to the same eigenstate  $F$ . The asymmetry

$$a_F = \frac{\Gamma(\bar{B}_{\text{phys}}(t) \rightarrow F) - \Gamma(B_{\text{phys}}(t) \rightarrow F)}{\Gamma(\bar{B}_{\text{phys}}(t) \rightarrow F) + \Gamma(B_{\text{phys}}(t) \rightarrow F)} \tag{1.12}$$

is only visible after allowing for some time where mixing can occur. At a time  $t$ , the decay rate for a state starting as a  $B^0$  at  $t = 0$  is

$$\Gamma(B_{\text{phys}}(t) \rightarrow F) = |A|^2 e^{-\Gamma t} \left( \frac{1 + |\lambda|^2}{2} + \frac{1 - |\lambda|^2}{2} \cos(\Delta m t) - \Im(\lambda) \sin(\Delta m t) \right), \tag{1.13}$$

and similar for the case when a  $\bar{B}^0$  was present in the beginning

$$\Gamma(\bar{B}_{\text{phys}}(t) \rightarrow F) = |A|^2 e^{-\Gamma t} \left( \frac{1 + |\lambda|^2}{2} - \frac{1 - |\lambda|^2}{2} \cos(\Delta m t) + \Im(\lambda) \sin(\Delta m t) \right). \quad (1.14)$$

The asymmetry (1.12) assumes the form

$$a_F = \frac{(1 - |\lambda|^2) \cos(\Delta m t) - 2\Im(\lambda) \sin(\Delta m t)}{1 + |\lambda|^2}$$

and simplifies to

$$a_F = -\Im(\lambda) \sin(xt), \quad x = \frac{\Delta m}{\Gamma}, \quad (1.15)$$

in the approximation  $|\lambda|=1$ . The quantity  $x$  is the ratio of the mass difference between the mass eigenstates and their width. The time  $t$  is measured in units of the  $B$  lifetime. The observable asymmetry is proportional to  $\Im(\lambda)$  and varies in time according to the frequency of the  $B^0$ - $\bar{B}^0$  oscillation.

## 1.4 The Golden Decay

The prerequisite for (1.11), which guarantees  $|\lambda|=1$ , is readily found. All contributing amplitudes must have the same weak phase  $\delta$ , i.e.

$$A = \sum_k^N M_k e^{i\delta_k} e^{i\alpha_k} \stackrel{!}{=} e^{i\delta} \sum_k^N M_k e^{i\alpha_k}.$$

The decay  $B^0 \rightarrow J/\psi K_S^0$  fulfills this requirement to a good approximation and is regarded as the most promising candidate to quantify CP violation in the  $B$  system. The final state is

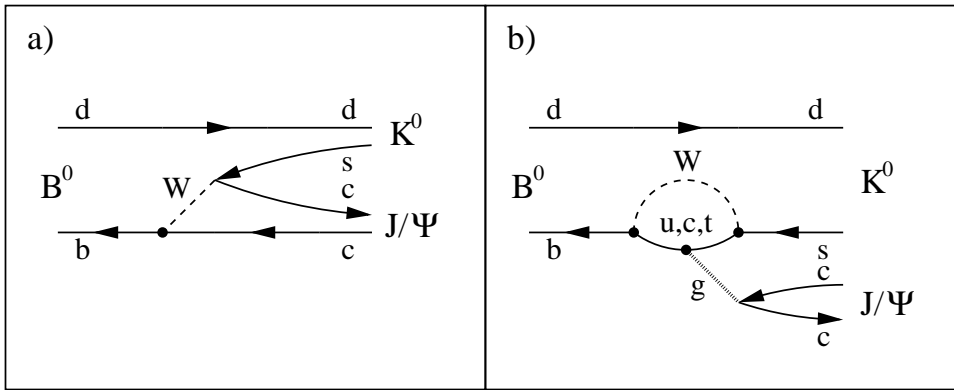


Figure 1.4: Diagrams contributing to the golden decay. Both tree (a) and penguin graph (b) enter with the same weak phase into the computation of the asymmetry.

a CP eigenstate with  $\text{CP}=-1$ . The decay proceeds primarily along the color-suppressed tree diagram depicted in Fig. 1.4 (a). The tree amplitude  $A_T$  is proportional to  $V_{cb}^* V_{cs} = A\lambda^2$  plus corrections in order  $\lambda^4$ . The same final state can be reached as well via penguin diagrams (Fig. 1.4 (b)) with  $up$ -type quarks in the loop. All penguin diagrams are suppressed by the additional vertices. The smallest contribution comes from the diagram with the  $u$  quark in

the loop, the amplitude of which is proportional to  $V_{ub}^*V_{us}$  and hence only of order  $\lambda^4$ . With a  $c$  quark in the loop the same combination of CKM matrix elements occurs as for  $A_T$ . The  $t$  quark diagram is proportional to  $V_{tb}^*V_{ts} = -A\lambda^2 = -A_T$ . When neglecting contributions in  $\mathcal{O}(\lambda^4)$  all diagrams enter into the calculation of the CP asymmetry with the same weak phase.

The final state with a  $K_S^0$  involves necessarily the  $K^0$ - $\bar{K}^0$  mixing, which contributes to the calculation of  $\lambda_{J/\psi K_S^0}$ . The relevant amplitudes are proportional to  $V_{cs}$  and  $V_{cd}$  which are free of the CP-violating phase to  $\mathcal{O}(\lambda^4)$ . The imaginary part of  $\lambda_{J/\psi K_S^0}$  is therefore not altered. The ratio of the decay amplitudes (1.11) is real ( $\delta=0$ ) and the magnitude of the asymmetry (1.15) is then

$$\Im(\lambda_{J/\psi K_S^0}) = -\sin(2(\beta - \delta)) = -\sin 2\beta.$$

The decay  $B^0(\bar{B}^0) \rightarrow J/\psi K^0$  has a small branching fraction of only  $8.9 \cdot 10^{-4}$ . Its observation in experiments is, however, relatively uncomplicated provided that a strong enough source of neutral  $B$  mesons is available. When the  $J/\psi$  decays to a lepton pair and the  $K_S^0$  to two charged pions, a clear experimental signature arises. This decay chain provides many kinematic constraints which help to cleanly isolate events of this type.

## 1.5 Other Angles of the Unitarity Triangle

A measurement of the CP asymmetry in the golden decay channel alone will allow for a precise determination of the CP-violating phase in the CKM matrix. To overconstrain the unitarity triangle using high-precision data, experimental values for the other angles in the triangle,  $\alpha$  and  $\gamma$ , would be most welcome. One can then check with high sensitivity whether the triangle closes and whether there is evidence for conceivable contributions from processes not included in the Standard Model.

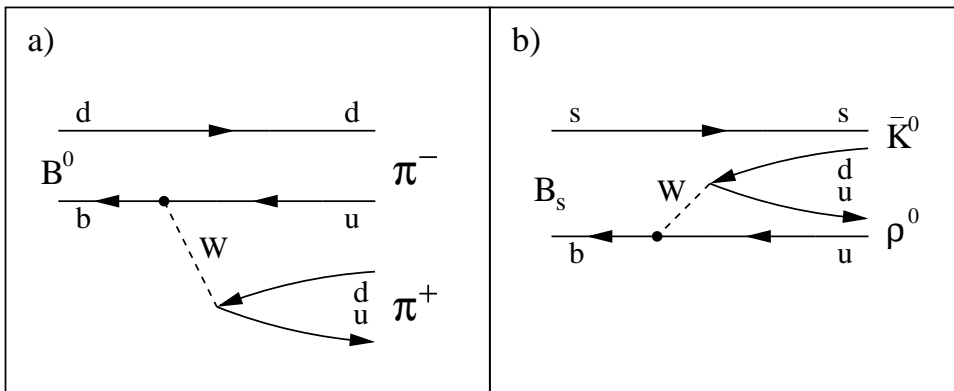


Figure 1.5: Diagrams for the decays  $B^0 \rightarrow \pi^+\pi^-$  (a) and  $B_s^0 \rightarrow \rho^0 K_S^0$  (b).

The most promising candidate for the determination of the angle  $\alpha$  is the decay  $B^0(\bar{B}^0) \rightarrow \pi^+\pi^-$ . The decay, shown in Fig. 1.5 (a), proceeds through the quark transition  $b \rightarrow u \bar{u}d$ . The amplitude is the sum of the tree diagram and a sizable contribution from penguin diagrams. All diagrams are of order  $\lambda^3$ . The amplitude of the dominating penguin diagram with a  $t$  quark in the loop is proportional to  $V_{td}^*V_{tb}$  and has a phase different from the tree diagram. The admixture of the penguin diagrams provokes that the magnitude of  $\lambda_{\pi^+\pi^-}$  is not unity. The uncertainties due to the penguin contributions can be eliminated by means of an isospin

analysis. In a first approximation without penguin pollution, the amplitude ratio is

$$\frac{\bar{A}}{A} \simeq \frac{V_{ub}V_{ud}^*}{V_{ub}^*V_{ud}} = e^{-2i\gamma},$$

and  $\Im(\lambda_{\pi^+\pi^-}) = -\sin(2(\beta + \gamma)) = -\sin(2(\pi - \alpha))$  measures  $\sin 2\alpha$ .

The largest difficulties are associated with a measurement of the angle  $\gamma$ . There are no decay channels with sufficiently high branching ratios where the penguin contributions are small. A conceivable candidate is the decay  $B_s(\bar{B}_s) \rightarrow \rho^0 K_S^0$  (Fig. 1.5 (b)), which proceeds through the same quark transition as  $B^0(\bar{B}^0) \rightarrow \pi^+\pi^-$  but with a different contribution from the  $B_s$ - $\bar{B}_s$  mixing. Extremely high fluxes of  $B_s$  mesons are required to approach  $\gamma$  in this channel and the rapid mixing deteriorates the chances to establish a time-dependent asymmetry. It is conceivable that one will resort to decays of charged  $B$  mesons to extract  $\gamma$  [8].

# Chapter 2

## The HERA- $B$ Experiment

There are currently several experiments striving for a measurement of CP violation in the channel  $B^0(\bar{B}^0) \rightarrow J/\psi K_s^0$ . Their experimental techniques differ widely. BaBar at SLAC<sup>1</sup> and BELLE at KEK<sup>2</sup> [9, 10] are both operated at asymmetric  $e^+e^-$  colliders producing the  $\Upsilon(4S)$  resonance, which decays to pairs of charged and neutral  $B$  mesons with roughly equal probabilities. The  $b\bar{b}$  cross section at the centre-of-mass energy of 10.6 GeV is about 1 nb, implying that unprecedented luminosities are needed to obtain statistically relevant samples of golden decays within a run period. Higher cross sections can be reached when heavy quarks are created in strong rather than in electroweak interactions. The CDF experiment at Fermilab<sup>3</sup> records symmetric  $p\bar{p}$  collisions at  $\sqrt{s} = 2$  TeV, where the  $b\bar{b}$  cross section is about 50  $\mu\text{b}$ . The CDF detector is not fully optimised to  $B$  physics and has limited access to the kinematic distributions since it emphasises the central over the forward region. The CDF experiment has extracted a value [11] of

$$\sin 2\beta = 0.79_{-0.44}^{+0.41}.$$

The analysis is based only on 400 selected events. The given error combines the statistical and systematic uncertainties. The result means that  $\sin 2\beta$  is in the interval  $[0,1]$  with a probability of 93%. The large error of the measurement imposes only a weak constraint on the shape of the unitarity triangle. The arising constraint is shown in Fig. 1.1 along with the exclusion areas resulting from other measurements.

HERA- $B$  is a large-acceptance experiment at DESY Hamburg [12, 13] dedicated to  $B$  physics. It will study  $B$  hadrons generated in hadronic collisions. Its principal goal is the measurement of the CP-asymmetry in the golden decay channel and in  $B \rightarrow \pi^+\pi^-$ . It might as well contribute to the detection of asymmetries in other channels, such as  $B_s^0 \rightarrow J/\psi\phi$ . Further fields of study are the measurement of the CKM matrix elements  $V_{ub}$  and  $V_{cb}$ , the mixing of  $B_s$  mesons, and the properties of  $b$  baryons.

### 2.1 Production of $b\bar{b}$ Pairs

The HERA- $B$  experiment exploits the 920 GeV proton beam of the HERA storage ring as a source of  $b$  quarks. Particles from the beam halo are directed on thin wires in fixed target mode. Heavy flavours are dominantly produced by gluon-gluon fusion and quark-antiquark

---

<sup>1</sup>Stanford Linear Accelerator Center (USA)

<sup>2</sup>High Energy Accelerator Research Organization, Japan

<sup>3</sup>Fermi National Laboratory, Batavia (USA)

annihilation in inelastic collisions of protons with target nucleons

$$p + N \rightarrow b\bar{b} + X. \quad (2.1)$$

The centre-of-mass energy of the proton-nucleon system is 41.6 GeV. The actual cross section of the process (2.1) is not known reliably. Both theoretical predictions and experimental data have large errors. The calculation of the cross section in perturbation theory is difficult due to the slow convergence of the series expansion in the strong coupling  $\alpha_s$ . It is known that the cross section is dominated by gluon fusion in lowest order while quark annihilation gives only a contribution of 15 %. Calculations up to  $\alpha_s^3$  resulted in predicted values between 6 and 150 nb. The measured values for the  $b\bar{b}$  cross section in the kinematic region relevant for HERA-*B* come from two Fermilab experiments. Both E789 and E771 used 800 GeV-protons and a fixed target to determine the  $b\bar{b}$  cross section for restricted phase space regions. The total cross section was then obtained as

$$\begin{aligned} \sigma_{b\bar{b}} &= (5.7 \pm 2.0) \text{ nb} && \text{(E789 [14]) and} \\ \sigma_{b\bar{b}} &= 43_{-18}^{+28} \text{ nb} && \text{(E771 [15])} \end{aligned}$$

by integration over the kinematic variables. The large errors are due to the limited number of events and model-dependent assumptions, which were made to perform the extrapolation to the total cross section.

Throughout the design phase of the HERA-*B* experiment, a value of  $\sigma_{b\bar{b}} = 12$  nb was assumed in all calculations. At that time, the proton-beam energy was only 820 GeV. It was increased to 920 GeV in 1998. HERA-*B* should profit from this change by a 10–15 % larger  $b\bar{b}$  cross section. Nevertheless, the actual value of  $\sigma_{b\bar{b}}$  at 920 GeV remains unknown until HERA-*B* has performed a precise measurement itself. This measurement, though not required explicitly for the determination of CP asymmetries, is one of the first tasks for the experiment.

After the creation of  $b$  quarks in pairs, they fragment independently into  $B$  hadrons. The fragmentation proceeds not completely symmetrically with respect to the production of particles and antiparticles. Due to the given initial state (2.1) the valence quarks of the nucleons will be enhanced in the created particles. This renders the creation of anti- $b$  baryons ( $\bar{b}\bar{q}\bar{q}$ ) very unlikely and might result in a production asymmetry for  $B^0$  ( $\bar{b}d$ ) and  $\bar{B}^0$  ( $b\bar{d}$ ). Simulations predict that the probability to create a  $\bar{B}^0$  in the fragmentation is 42 % and therewith slightly smaller than that of a  $B^0$  (43 %). Such production asymmetries must be measured thoroughly before CP asymmetries can be extracted. They can be determined by comparing the decay rates to final states which are no CP eigenstates, e.g.  $B^0 \rightarrow J/\psi K^{*0} (\rightarrow K^+\pi^-)$ .

## 2.2 Flavour Tagging

The neutral  $B$  mesons created in the proton-nucleon collisions exist as time-evolved states before decaying into a CP eigenstate like  $J/\psi K_S^0$ . The final state can be reached by both  $B^0$  and  $\bar{B}^0$  and does not contain a reference to the initial state anymore. Therefore, the second  $B$  hadron in the event is used to detect whether the neutral  $B$  was initially a  $B^0$  or a  $\bar{B}^0$ . This determination of the quark content of the decaying neutral  $B$  at the production time is called *flavour tagging*.

The methods employed for flavour tagging are partly illustrated in Fig. 2.1. One distinguishes between *exclusive* and *inclusive* tags. The former involve the complete reconstruction of the second  $B$  in the event. In that way, a  $B^-$  ( $b\bar{u}$ ) indicates that the time-evolved neutral  $B$  was as a  $B^0$  at  $t=0$ . Exclusive tags are very reliable, but can only be applied for those

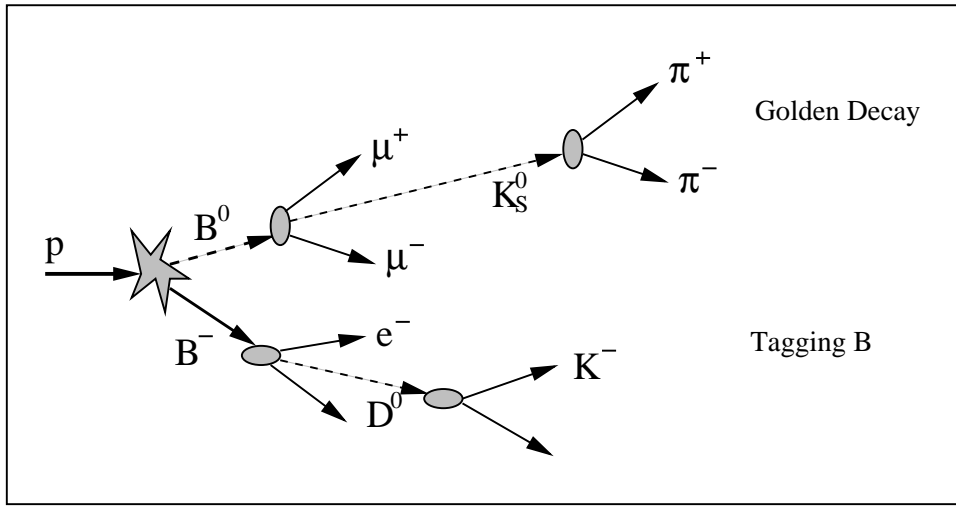


Figure 2.1: Decay scheme of a golden event. The flavour at production time of the  $B$  decaying to  $J/\psi K_S^0$  (upper half of plot) is tagged by the second  $B$  in the event (lower half). Inclusive tags involve the charge of the lepton from semileptonic decays of the second  $B$  (lepton tag) or the charge of the kaon created in the  $b \rightarrow c \rightarrow s$  transition (kaon tag). If the second  $B$  is not neutral, the sum of all charges of the particles emerging from the decay vertex of the second  $B$  determines the  $b$  flavour as well (charge tag). (The sketch is not to scale. The  $K_S^0$  travels on average more than 1 m while the  $B$  mesons decay after about 10 mm.)

events where the complete reconstruction of the decay chain was possible. This limits their efficiency severely and one is forced to resort to inclusive tags. Inclusive tags are based on less reconstructed information from the decay chain of the tagging  $B$ :

1. **Lepton tag:** When the tagging  $B$  decays semileptonically via  $b \rightarrow c + W^-$  or  $\bar{b} \rightarrow \bar{c} + W^+$  the charge of the lepton from the  $W$  boson defines the quark content unambiguously. The high-energy lepton occurring in proximity to the  $B$  hadrons must be reconstructed. The lepton tag will give a wrong answer if the lepton did not come from the  $b$  quark but from the following semileptonic  $c \rightarrow s$  transition.
2. **Kaon tag:** The charge of a kaon from the dominant decay chain  $b \rightarrow c \rightarrow s$  determines the initial quark flavour. A  $K^-$  ( $s\bar{u}$ ) results from the decay chain of a  $b$  quark, a  $K^+$  denotes a  $\bar{b}$ .
3. **Charge tag:** If the tagging  $B$  is charged, its sign identifies the quark flavour. For the charge tag one sums up the charges of all tracks emerging from the tagging  $B$  hadron. A negative (positive) charge indicates a  $b$  ( $\bar{b}$ ). The charge tag cannot be applied when the tagging  $B$  is neutral.

Besides the tags enumerated above, the  $B$  meson decaying to a CP eigenstate can reveal its flavour content itself. This is called *same-side tagging* since it works independently of the second  $B$  hadron in the event. Local charge conservation in the fragmentation makes it likely that a  $B^0$  is accompanied by a  $\pi^+$  rather than a  $\pi^-$ . If the pion can be reconstructed, it signals the flavour of the decaying  $B$ . Likewise, a flavour assignment can be made when the neutral  $B$  stems from an excited  $B$  state. This is the case for roughly 30% of all events and the  $b$  flavour can be derived from the charge of pions in transitions like  $B^{**} \rightarrow B^0 \pi^+$  ( $B^{**}$  tag).

None of the listed methods is the ideal solution. The thorough understanding of their assets and drawbacks is crucial for the determination of CP asymmetries. The quality and efficiency of the different tagging methods have been studied in great detail using simulations [12, 16]. Tagging errors reduce the magnitude of the observable CP asymmetry which is then given by

$$a_{J/\psi K_S^0} = -D_T D_M \sin 2\beta \sin(\Delta m t). \quad (2.2)$$

The "dilution" factors  $D_T$  and  $D_M$  are both smaller than unity.  $D_T$  arises from mistagging the  $B^0$  or  $\bar{B}^0$  flavour using one of the methods described above.  $D_M$  is a irreducible dilution inherent to events where the tagging  $B$  is neutral and can mix. If the second  $B$  oscillates into its antiparticle before the decay, the tag will inevitably give a wrong result.

In the beginning, HERA- $B$  will measure the time-integrated instead of the time-resolved CP asymmetry since the latter requires a thorough understanding of the decay length measurement of the  $B$  mesons. Integrating Eqs. (1.13) and (1.14) over all times  $t \geq t_0$  yields

$$N = \int_{t_0}^{\infty} \Gamma(B_{\text{phys}}^0 \rightarrow J/\psi K_S^0) dt \quad \text{and} \quad \bar{N} = \int_{t_0}^{\infty} \Gamma(\bar{B}_{\text{phys}}^0 \rightarrow J/\psi K_S^0) dt$$

as the total number of events where a time-evolved  $B^0$  or  $\bar{B}^0$  decayed to the golden channel. The integration starts at  $t=t_0$  rather than at  $t=0$  because a minimal lifetime<sup>4</sup> of the  $B$  mesons must be required to sufficiently reduce backgrounds. The losses introduced by this cut are small, since the mixing-induced asymmetry reaches its first maximum after roughly two  $B$  meson lifetimes. The time-integrated asymmetry is given by

$$a_{J/\psi K_S^0}(t_0) = \frac{N - \bar{N}}{\bar{N} + N} = -D_T D_M \sin 2\beta \xi(t_0),$$

with  $\xi(t_0) = (x \sin x t_0 + \cos x t_0)/(1 + x^2)$ . For a background-free sample of  $N_B$  golden events the uncertainty of  $\sin 2\beta$  is then

$$\Delta \sin 2\beta = \frac{1}{\xi(t_0)} \frac{1}{D_M D_T \sqrt{\varepsilon_{\text{tag}}}} \frac{1}{\sqrt{N_B}}, \quad (2.3)$$

where  $\varepsilon_{\text{tag}}$  denotes the combined efficiency of all tagging methods. The goodness of the tagging is measured by the *tagging power*  $P = D_M D_T \sqrt{\varepsilon_{\text{tag}}}$ . Simulations indicate that a value of  $P = (0.31 \pm 0.02)$  [17] can be reached. The lifetime cut will be placed around  $t_0 = 0.7 \tau_B$ , which is more than ten times the expected resolution in the  $B$  decay vertex position. Inserting these values into (2.3) indicates that a sample of roughly 2000 events is required to measure  $\sin 2\beta$  with an absolute error of 0.1.

## 2.3 Experimental Requirements

At HERA- $B$  energies, the centre-of-mass energy on the parton level is still close to the threshold for pair production of  $b$  quarks. Therefore, the reaction  $p + N \rightarrow b\bar{b} + X$  is still strongly suppressed. The total inelastic cross section is about 13mb. Interactions with light quarks occur  $10^6$  times more often than  $b\bar{b}$  events.

---

<sup>4</sup>In practice, the cut is imposed on the flight distance of the  $B$  mesons. The translation to a lifetime is not well-defined due to the momentum distribution of the  $B$  mesons. This point is ignored in the discussion.

The number of interactions  $N_{\text{inel}}$  that HERA- $B$  has to examine in order to obtain a sample of roughly 2000 golden decays can be estimated as follows. The number of extracted golden events  $N_B$  is related to  $N_{\text{inel}}$  by

$$N_B = N_{\text{inel}} \cdot \frac{\sigma_{b\bar{b}}}{\sigma_{\text{inel}}} \cdot P_B \cdot \text{BR}_{\text{gold}} \cdot (\varepsilon_e + \varepsilon_\mu). \quad (2.4)$$

Here  $P_B$  denotes the probability that at least one neutral  $B$  is created in the fragmentation (about 0.8) and  $\text{BR}_{\text{gold}}$  is the branching ratio for the decay chain  $B \rightarrow J/\psi K_S^0 \rightarrow l^+l^- \pi^+\pi^-$  (about  $2 \cdot 10^{-5}$ ). The branching fractions of the  $J/\psi$  to the electron and muon channel of 6% agree within the experimental errors. The efficiency to record golden events is thus the sum of the efficiencies for the electron ( $\varepsilon_e$ ) and the muon channel ( $\varepsilon_\mu$ ). According to Monte Carlo simulations, the joint efficiency to find all decay particles in the geometrical acceptance of the detector, to trigger and to reconstruct them completely is 8.5% for electrons and 16.6% for muons [13]. Taking these numbers at face value and  $\sigma_{b\bar{b}}=14$  nb, Eq. (2.4) results in a value of about  $5 \cdot 10^{14}$  for  $N_{\text{inel}}$ . The generation of this huge number of proton-nucleon collisions within 1 year ( $10^7$  s) of data-taking requires interactions to occur on the wire target with a rate of 50 MHz.

The HERA proton ring is operated at a frequency of 10.4 MHz. Proton bunches can circulate in 220 RF buckets along the ring circumference of 6.3 km. One bucket is 96 ns long. Only 180 of the 220 buckets are filled with bunches to allow for beam injection and extraction. The effective frequency at which proton bunches cross the HERA- $B$  target is therefore 8.5 MHz. When interaction rates as high as several 10 MHz should be reached, more than one inelastic proton-nucleon collision must occur per bunch crossing. The highest interaction rate, which is considered as manageable for HERA- $B$ , is 40 MHz. At this rate, 4–5 interactions are superimposed in one event. The multiple interactions provoke a harsh hadronic environment with particle fluxes of about  $10^7 \text{ cm}^{-2} \text{ s}^{-1}$  in the region close to the proton beam. Including secondary particles, roughly 150 tracks are created per event.

The HERA- $B$  detector must be able to resolve the multiple interactions per event, implying that particle detection systems with an appropriate granularity are required. The occupancies of detection elements should be limited to at most 20% to allow for pattern recognition. For reconstruction of the decay chains of  $B$  hadrons good vertex and mass resolutions are mandatory. A further requirement is an efficient identification of leptons, kaons, pions and protons. The subdetectors have to provide signals fast enough to cope with the close bunch spacing of 96 ns. They should take advantage of radiation-hard detector technologies to resist the high particle flux, especially in the region close to the proton beam.

## 2.4 Overview of the HERA- $B$ Detector

The implemented forward-spectrometer depicted in Fig. 2.2 is a compromise between performance and cost. It extends to 250 and 160 mrad in the bending and non-bending plane, respectively. This corresponds to 90% coverage in the centre-of-mass system. The wire target and a silicon vertex detector reside in a vessel in front of the spectrometer magnet. Tracking is performed by different kinds of position-sensitive detectors, which are arranged around the proton beam pipe. Particle identification is accomplished by means of a Ring Imaging Cherenkov Counter (RICH) and a Transition Radiation Detector (TRD). An electromagnetic calorimeter (ECAL) measures particle energies. Particles penetrating both the ECAL and several thick

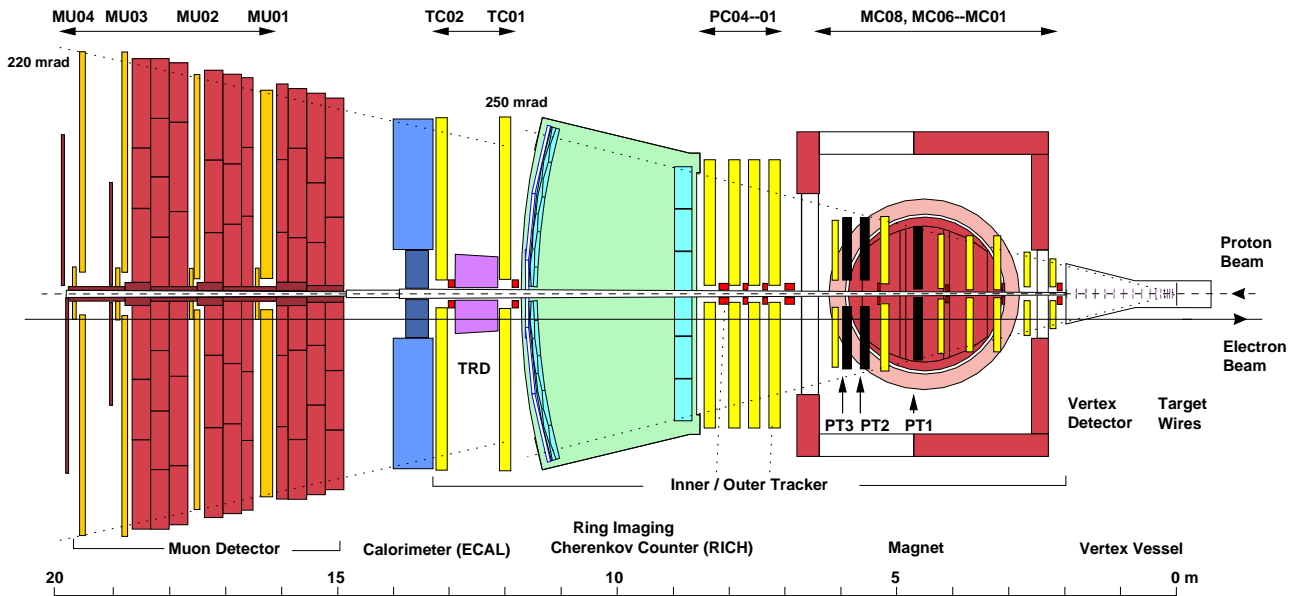


Figure 2.2: Top view of the HERA-*B* spectrometer. The proton beam enters from the right and passes through a vessel housing both wire target and silicon vertex detector. The main tracking system consists of a dipole magnet and different kinds of position-sensitive detectors arranged along the proton beam. A Ring Imaging Cherenkov Counter and a Transition Radiation Detector provide particle identification. The energies of electrons and photons are measured by an electromagnetic calorimeter, which is situated in front of the huge muon system at the end of the spectrometer. The electron beam pipe passes everywhere untouched through the detector.

absorber layers are detected in the muon system. The total number of readout channels is about 600.000.

## 2.5 Wire Target

The basic idea of the HERA-*B* target is to make the halo protons, which would get lost from the beam anyway, interact in the target wires. The natural loss rate of the proton beam can be calculated from the number of protons (about  $2 \cdot 10^{13}$ ) and its average lifetime. The natural lifetime of the proton beam is roughly 1000 hours and is reduced by a factor of 10 when *ep* collisions occur for the collider experiments H1 and ZEUS. The natural proton loss rate amounts then to several 10 MHz already. Ideally, the target wires would absorb all protons drifting away without affecting the beam lifetime further. To obtain the interaction rate for the experiment, the wires are operated at  $4-6 \sigma$ , where  $\sigma$  describes the beam profile in horizontal or vertical direction. In practice, a target efficiency (defined by the ratio of the number of protons interacting in a wire and the total number of particles getting lost) of 40–50% [18] is reached routinely. The lifetime of the proton beam is reduced to about 50 hours, which is still acceptable for the other experiments at HERA, since the overall schedule is mainly determined by the shorter lifetime of the electron beam.

The target comprises eight wires to spread the interactions spatially and to ease the assignment of tracks to interactions. There are two target stations separated by about 4 cm along the beam direction (Fig. 2.3). Each station features two horizontal and two vertical wires mounted on forks, which can be moved with respect to the beam. During target operation, the wires are

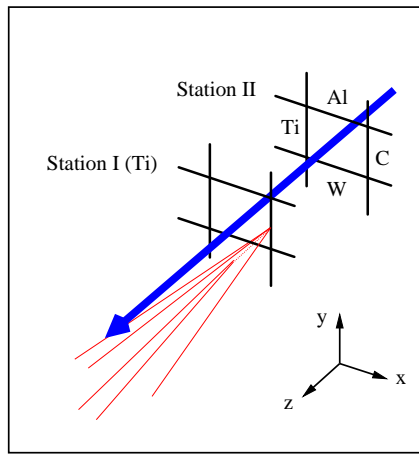


Figure 2.3: Sketch of the target. The two target stations with four wires each are separated by about 4 cm. The wires are operated at a distance of 2–3 mm from the proton beam. The plot reflects the target configuration in the running period 1999/2000. The four wires of station I (further downstream) consisted of titanium. Four different wires of aluminum, carbon, tungsten and titanium were deployed at station II.

steered to provide a constant interaction rate. The wire dimensions are adjusted to match the resolution of the nearby vertex detector. The wires have a thickness of  $50\ \mu\text{m}$  in the transverse plane. Their length along the beam is  $500\ \mu\text{m}$ . Several elements (Al, C, Cu, W, Ti) have been tested as material for the target wires. Since the  $b\bar{b}$  cross section grows approximately linearly with the mass number  $A$  while the total cross section scales only with  $A^{2/3}$ , some benefit can be expected from heavy elements. On the other hand, the multiplicity of charged tracks is proportional to  $A^{0.2}$  and takes away from the gain of heavy target materials. In the 1999/2000 running period, the target station I (further downstream) was equipped with titanium wires throughout. At station II, four wires made of carbon, tungsten, aluminum and again titanium were operated. (The coordinate system used in HERA-B and in this thesis is indicated in the lower right corner of Fig. 2.3. The  $z$ -axis is parallel to the beam direction, the  $y$ -axis is vertical, and the  $x$ -axis points to the centre of the HERA ring. The wires of station I and II are at  $z=-1\ \text{cm}$  and  $-5\ \text{cm}$ , respectively.)

## 2.6 Silicon Vertex Detektor

In proximity to the target, the silicon vertex detector records the trajectories of particles emerging from the interactions on the wires. This information allows to reconstruct the primary and secondary vertices in the recorded events. The vertex positions serve to reject background processes without extended lifetime and to determine the decay length of the  $B$  mesons needed for the measurement of the time-resolved CP asymmetry. The accuracy of the decay length measurement is crucial for the performance of the experiment and its reach in terms of  $B$  physics. The requirements to the vertex detector are the following:

- \* **Large geometrical acceptance for particles from events with heavy quarks.** The SVD should allow to record tracks with angles to the beam between 10 and 250 mrad.

- \* **Precise determination of the decay length  $l$  of  $B$  mesons.** For CP violation studies, the error  $\sigma_l$  has to be much smaller than the average decay length of about 10 mm itself. One demands

$$\frac{\sigma_l}{l} < \frac{1}{10},$$

implying that  $\sigma_l$  should fall below 1 mm. To resolve the rapid  $B_s$  oscillations, a better resolution is clearly welcome. A value of  $\sigma_l = 500 \mu\text{m}$  and a transverse resolution of about  $25 \mu\text{m}$  should be reached to separate primary and secondary vertices efficiently.

- \* **Feasibility of a self-contained pattern recognition of tracks.** The high track density in the HERA- $B$  detector can be coped with best, when track finding is started independently before and after the magnet. The benefits of the resulting redundancy are twofold. If a real track is found due to a insufficient number of hits in one system only, the prolongation to the other system will recover the associated hits. Likewise, track segments which cannot be matched to a segment in the other system will be marked as ghosts, since they comprise most likely hits caused by different particles. (The impact of a possible local pattern recognition in the vertex detector on triggering will become apparent later.)
- \* **Radiation-hard implementation.** The need for radiation hardness is a consequence of the first two items of this list and the extremely high particle flux close to the beam. The active detector elements cannot be replaced more often than once a year.

Two quantities must be minimised to achieve the best precision for track parameters and vertices. One is the distance between the particle origin and the first measurement of its trajectory, the other the amount of material seen by the particle, in particular before the first measurement. On the other hand, the performance degradation due to radiation damage sets a lower limit to the distance from the beam, where silicon detectors can be operated over extended time periods. At HERA- $B$ , the total particle flux at  $R=1 \text{ cm}$  is about  $3 \cdot 10^7 \text{ cm}^{-2} \text{ s}^{-1}$ . This corresponds to an annual radiation dose of typically 100 kGy [13], which silicon detectors can resist when their temperature is kept around  $0^\circ\text{C}$  by active cooling.

### 2.6.1 Silicon Strip Detectors

The requirements listed above were met by mounting double-sided silicon strip detectors perpendicularly to the proton beam (Fig. 2.4 (a)). The detectors are deployed in Roman Pots and can be moved as close as 10 mm to the proton beam. The used silicon strip detectors have a dimension of  $53 \times 73 \text{ mm}^2$  and a thickness of  $280 \mu\text{m}$ . The active area covered with strips on a chip is  $50 \times 70 \text{ mm}^2$ . The strip pitch is about  $25 \mu\text{m}$  and every second strip is connected to the readout. Amplifiers sample the amount of charge created by a penetrating particle close to the strip. Combining the information of neighbouring strips allows to determine the impact point of a track with an accuracy of about  $12 \mu\text{m}$ , which is slightly better than the resolution expected from the readout pitch of  $50 \mu\text{m}$  alone. The strips run under an angle of  $2.5^\circ$  with respect to the detector edges. When the detectors are aligned along the  $x$ - and  $y$ -axis, one strip layer measures a particle trajectory under a stereo angle  $\alpha_{\text{st}}$ . On a double-sided chip, the strips on the front and the back side measure different views ( $x$  and  $y$ ). One layer with 1280 readout strips is (approximately) parallel to the shorter edge, the other with 1024 strips runs (approximately) parallel to the longer edge. The resulting strip densities keep the occupancies below the level of 5 %.

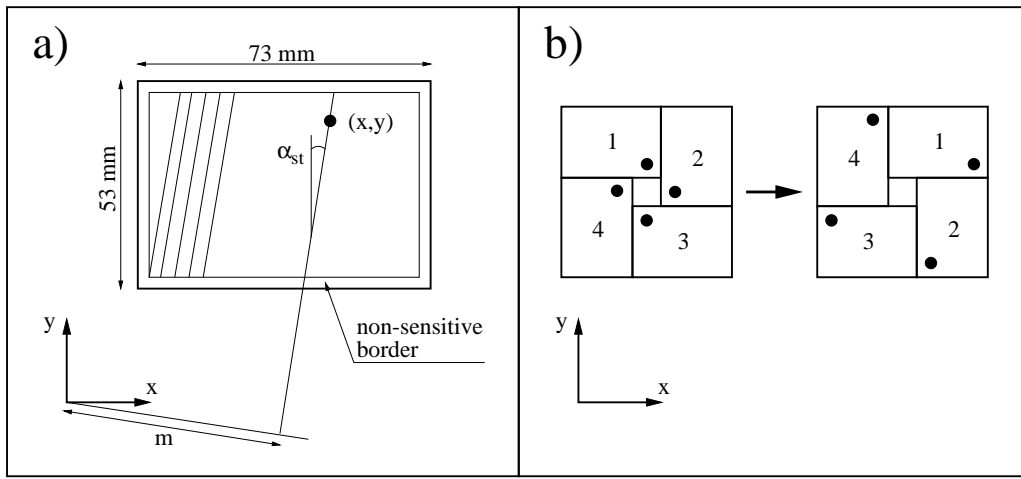


Figure 2.4: (a) Sketch of a silicon strip detector mounted perpendicularly to the proton beam. The strips run under a stereo angle  $\alpha_{st}$  of  $2.5^\circ$  with respect to the long or short detector edges. One strip layer measures a projection  $m = x \cdot \cos \alpha_{st} - y \cdot \sin \alpha_{st}$  of a track impact point  $(x, y)$ . (b) The silicon strip detectors are arranged around the beam to cover the entire acceptance. From time to time, their position is changed between the two configurations shown in order to expose all detector parts to the particle flux uniformly.

Table 2.1: Each SVD superlayer has four layers of silicon strips whose arrangement depends on the sector in question. The table lists the four layers in the direction of increasing  $z$ . The letter  $x$  or  $y$  indicates to which axis the strips are almost parallel, the sign specifies whether the stereo angle is  $+2.5^\circ$  or  $-2.5^\circ$  with respect to this axis.

	SL 1-7
Sector 1	$-y + x - x + y$
Sector 2	$+x - y + y - x$
Sector 3	$-y + x - x + y$
Sector 4	$+x - y + y - x$

## 2.6.2 Detector Geometry

The described silicon strip detectors are arranged in eight superlayers along the beam (upper half of Fig. 2.5). One superlayer consists of four quadrants, which are referred to as sectors and labelled from 1 to 4 (sector 1-4). One sector comprises two double-sided silicon chips, which are mounted such that all four possible stereo angles ( $\pm 2.5^\circ$  with respect to the  $x$ - and  $y$ -axis) are represented. The actual sequence of strip orientations depends on the sector in question (cf. Table 2.1), but in either case one superlayer can give rise to up to four measurements of a particle trajectory. The  $z$ -positions of the superlayers were adjusted such that every particle with a slope between 10 and 250 mrad hits three superlayers and can be found as a self-contained track with up to 12 hits (lower half of Fig. 2.5). The measurement of tracks under very small angles is assisted by an additional superlayer (SL 8), which also takes advantage of silicon strip detectors but rather belongs to the main tracking system. Within a superlayer, the silicon strip detectors in the different sectors have to overlap in order to cover the entire geometrical acceptance. This implies that they cannot reside at the same  $z$ -position (cf. Table 2.2), but

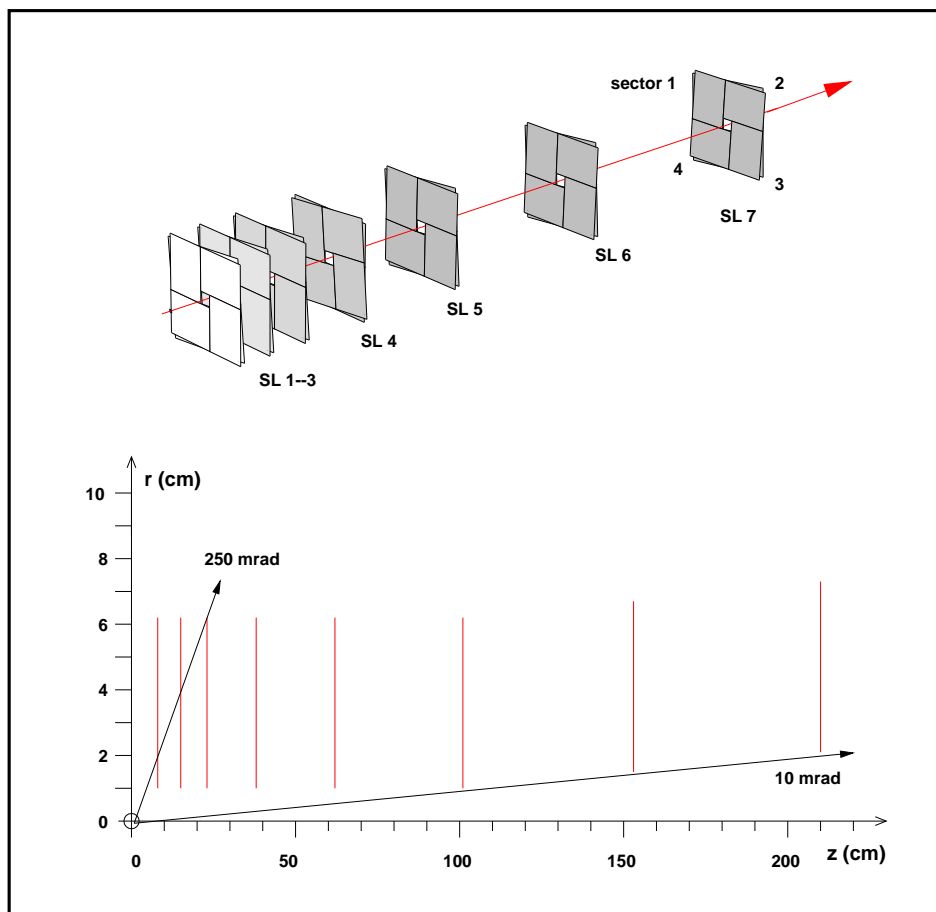


Figure 2.5: The silicon vertex detector consists of four telescopes mounted around the beam. In the upper plot, the different stereo angles of the strip layers are indicated by a rotation of the chips, which are aligned along  $x$  and  $y$  in reality. Eight silicon strip detectors make up one superlayer consisting of four quadrants, which are referred to as sectors (sectors 1–4). There are eight superlayers (lower plot) whose arrangement guarantees that particles hit three superlayers independently of their polar angle. Only one sector is shown in the lower plot. Superlayer 8 does not reside in the vertex vessel. This last silicon superlayer is regarded as a part of the main tracking system. It ensures that also tracks under very low angles hit three superlayers.

Table 2.2: Positions of the SVD superlayers 1–7 in  $z$ . The minimal (maximal) position is given by the first (last) strip layer in a superlayer.

Superlayer	1	2	3	4	5	6	7
$z_{\min}$ (cm)	8.1	13.4	20.3	37.7	61.4	99.1	151.2
$z_{\max}$ (cm)	10.7	16.0	22.9	40.4	64.3	102.1	154.2

must be slightly displaced along the beam. The superlayers acquire therefore a thickness of roughly 2.5 cm.

### 2.6.3 Vertex Vessel

Both target and the SVD superlayers 1–7 reside in a vacuum vessel (Fig. 2.6) connected to the proton beam pipe on both sides. The vessel has a length of 2.6 m and consists almost entirely of steel. Only the vaulted exit window, which has to be crossed by all particles in the acceptance, is made of 3 mm thick aluminum. The target system is located close to the point where the proton beam enters the vessel. At five positions further downstream, at each case four pots surround the beam and can accommodate silicon strip detectors to form the different sectors of the superlayers. Pot 1 contains the superlayers 1–3, while superlayer 4–7 are mounted in individual pots (2–5). The two silicon strip detectors within one sector of a superlayer are separated from the primary ring vacuum in the vessel ( $p = 10^{-8}$  mbar) by aluminum caps. These caps have a thickness of only  $150 \mu\text{m}$  and are hit by the tracks almost perpendicularly. They pose a much smaller amount of material than a normal beam pipe with similar dimensions parallel to the beam. The caps shield the detectors from the proton beam and protect the primary ring vacuum from the outgassing of the variety of materials used in assembling the silicon detector modules. They are, however, by far not rigid enough to constitute the separation between the ring vacuum and atmospheric pressure. The volume of the pots is therefore evacuated down to a pressure of  $10^{-6}$  mbar.

During beam injection and luminosity tuning, the silicon strip detectors cannot be operated at distances of only 10 mm to the beam. They must be retracted to 22–35 mm to allow for an enlarged free aperture. The SVD modules are connected to so-called manipulators providing the required mechanical degrees of freedom along with a high-precision steering. Each pot can be moved in the transverse plane by means of stepping motors. The movements serve primarily to position the silicon strip detectors with respect to the beam for data-taking. On a timescale of months, the detectors are shifted such as to move those regions exposed to the highest radiation doses further away from the beam (cf. Fig. 2.4 (b)). This proceeding ensures a more uniform distribution of the accumulated particle flux over individual chips and should help to prolong their lifetime.

The vacuum vessel as such acts as a cavity and would thus interfere with the proton beam without further measures. Induced wake-fields might provoke resonant beam losses. The screening of the proton beam is accomplished by four thin steel ribbons as a substitute for a complete beam pipe. These so-called radio frequency bands (RF bands) have a width of 13 mm, a length of 195 cm, and a thickness of  $5 \mu\text{m}$ . They surround the beam almost completely (cf. Fig. 2.7) and run in parallel from the first superlayer up to the exit window. The bands are pushed towards the beam by the SVD superlayers when these are moved in. During data-taking, their distance to the beam is about 0.8 cm.

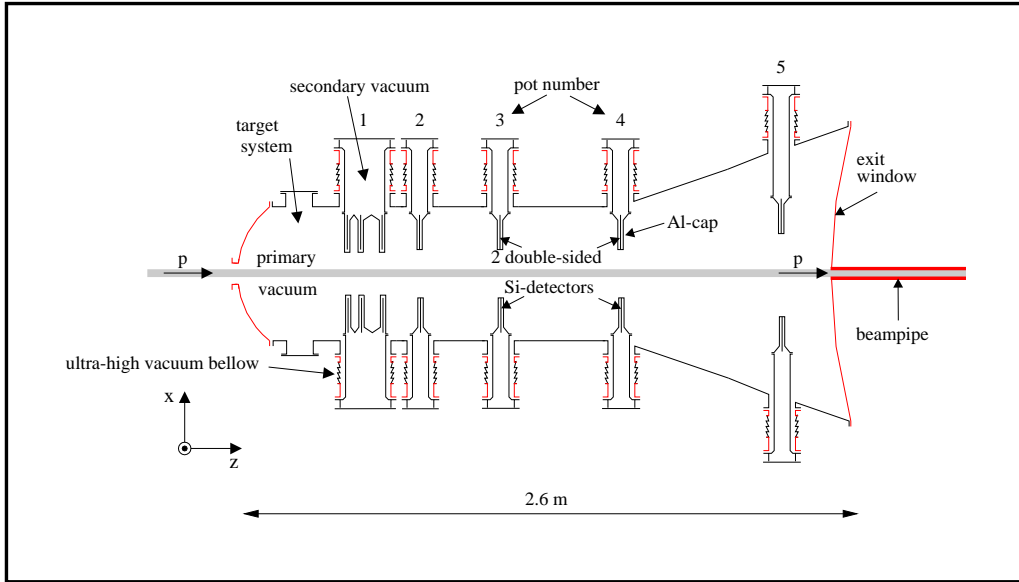


Figure 2.6: Schematical drawing of the vessel housing both target and the vertex detector superlayers 1–7. The target system is deployed in the first part on the very left. The SVD superlayers are installed in  $4 \times 5$  pots around the proton beam. The superlayers 1–3 share one pot (1), the superlayers 4–7 have individual pots (2–5). Two double-sided detectors are deployed per superlayer and sector in the pots. The pots are separated from the ring vacuum by means of thin aluminum caps. The particles generated in the interactions on the target wires leave the vessel through an exit window. (The plot is not to scale.)

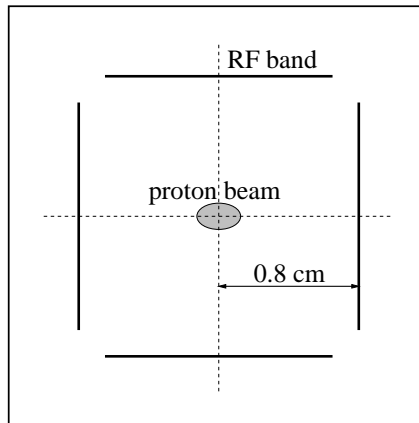


Figure 2.7: Four steel ribbons serve to prevent electromagnetic interference between the proton beam and the vertex vessel. They are 1.3 cm wide and have a thickness of  $5 \mu\text{m}$ . During SVD operation, the RF bands run at a distance of 0.8 cm parallel to the beam.

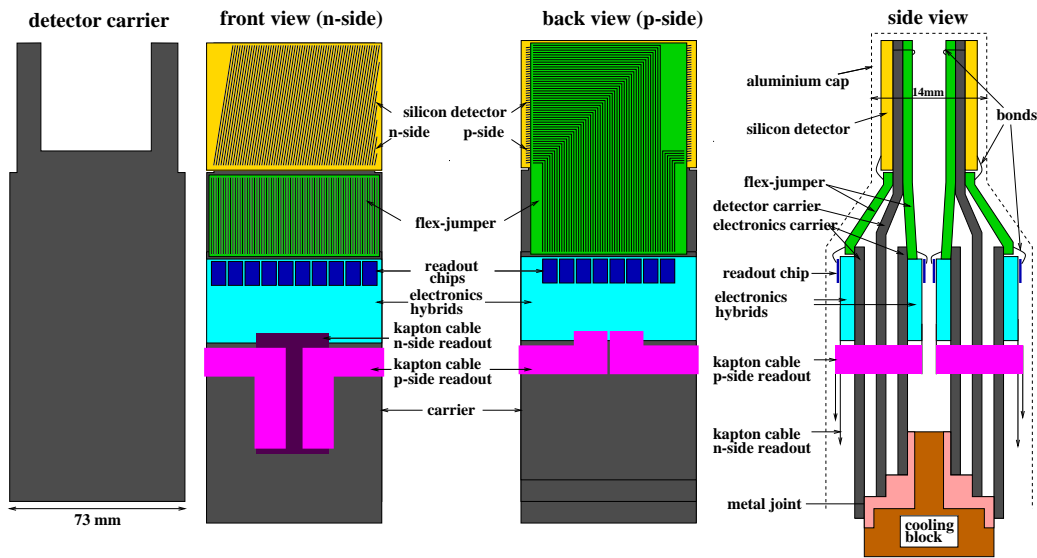


Figure 2.8: Various views of detector elements. The pictured module is designated for superlayer 5 or 6. The detector carrier is shown on the very left. It consists of carbon fibre and has a fork-like shape to minimise the amount of material in the acceptance. The silicon detectors are fixed on the carrier by gluing. The two plots in the middle show the front and back side of a complete module. The connection of the silicon strips to the readout chips (HELIX) differs on the two sides due to the orientation of the strips with respect to the detector edges. One HELIX chips covers 128 channels. The further signal transmission from the readout chips to the data acquisition system employs kapton cables. Since the entire structure resides in the secondary vacuum, additional cooling is required to dissipate the heat generated by the electronics. The actual detectors and the readout chips are mounted on different carriers (cross section on the right) to avoid a direct transfer of heat between them. The three carriers attached to one silicon detector are connected to a cooling block. The entire module has a thin cover of aluminum separating it from the ring vacuum.

## 2.6.4 Detector Modules

The actual implementation of the detector modules, which are scheduled for mounting in one pot, is illustrated in Fig. 2.8. The base element is a detector carrier. It consists of carbon fibre and has a fork-like shape to limit the amount of material in the geometrical acceptance. One silicon strip detector is mounted on the fork by gluing. The readout chip employed for the SVD (HELIX) is radiation-hard ( $>1\text{ kGy}$ ) but cannot be operated at distances smaller than 10 cm from the beam. One chip provides 128 channels, implying that 10 and 8 chips are required to cover the 1280 and 1024 strips on the two sides of one silicon strip detector. The connection between readout strips and the HELIX chips is provided by a so-called flex-jumper. The flex-jumper consists of a flexible kapton foil with conductor paths made of copper. The electrical connections between silicon strips and flex-jumper and between flex-jumper and HELIX chips are ensured by short bonding wires. The readout chips are attached to electronics hybrids, which provide the required operating voltages. The further signal transmission from the HELIX chips to the data acquisition system employs kapton cables. The electronics hybrids have their own carbon fibre carriers to avoid a direct transfer of heat produced by the electronics to the silicon detectors. One complete module with two silicon detectors (shown on the very right of Fig. 2.8) comprises six carriers, which are connected to a common cooling block. The cooling block consists of metal and dissipates the heat generated by the module. Each block is

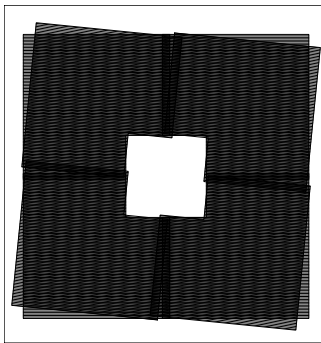


Figure 2.9: View of the sensitive areas of four MSGCs surrounding the beam. The area covered is about  $40 \times 40 \text{ cm}^2$ . Two layers ( $0$  and  $5^\circ$ ) are shown.

connected to a system of pipes through which a cooling liquid is pumped. The cooling liquid used is a mixture of water, alcohol, and ice which keeps the cooling blocks at a temperature of  $-15^\circ$ . The resulting temperature gradient within a module is such that the active detector elements are operated at a temperature of roughly  $0^\circ\text{C}$ .

## 2.7 Tracking System

The tracking system for momentum determination extends from the exit window of the vertex vessel up to the muon system. It consists of a dipole magnet ( $2.2 \text{ Tm}$ ) and 14 superlayers each of which provides several measurements in different views for a passing particle. The track density scales like

$$\frac{dN}{dA} \approx \frac{30 \text{ MHz}}{R^2}$$

with the perpendicular distance  $R$  from the proton beam pipe. Detector technology and granularity are adjusted to this particle flux. Occupancies are limited to at most 20% to allow for independent pattern recognition in the tracking system. The detector technology used is

- silicon strip detectors for  $R$  in the range of 1–6 cm,
- micro-strip gaseous chambers (MSGC) for  $R$  up to 19 cm (Inner Tracker (ITR)), and
- drift chambers at all larger  $R$  (Outer Tracker (OTR)).

The silicon strip detectors in the main tracking system are restricted to one superlayer (SL 8) between the vertex tank and the magnetic field. The remaining 13 superlayers are divided into three distinct groups: seven superlayers in the magnet (*magnet chambers* MC01–MC06, MC08), four pattern recognition stations in the following field-free region (*pattern recognition chambers* PC01–PC04), and two superlayers for triggering (*trigger chambers* TC01, TC02) in front of the ECAL. The momentum measurement is performed by detecting the curvature of tracks in the magnet using the MC and PC chambers. In the offline analysis of events, pattern recognition starts in the PC chambers where the track model is linear. Found candidates are extrapolated towards the MC chambers (for momentum measurement) and towards the TC stations (for linking of tracks with the calorimeter and the muon system).

About 50% of all tracks are detected in the ITR. In all superlayers except MC02, MC04 and MC08, four L-shaped MSGC modules form the detection layers. The MSGCs cover an area of

$40 \times 40 \text{ cm}^2$  (Fig. 2.9) and measure particles in views of  $0^\circ$  and  $\pm 5^\circ$  with respect to the vertical direction. They consist of a glass substrate with anode and cathode strips in a box filled with drift gas. The length of the drift region is 6 mm along  $z$ . The anode pitch is  $300 \mu\text{m}$  and the measured coordinate is derived from the centre-of-gravity of charge depositions on neighbouring strips. The resulting position resolution of a single layer is  $80 \mu\text{m}$ . To prevent strong aging effects, a Gas Electron Multiplier (GEM) foil is used. This thin layer divides the drift region in two sections. Gas amplification occurs both at holes drilled into the copper-coated foil and at the anodes. The pre-amplification at the GEM foil allows for smaller gains and subsequently lower voltages at the anode strips. The GEM technique is an effective measure to prevent sparking between cathode and anode strips.

The outer tracking regions are covered entirely by drift tubes with hexagonal cross section. The cathodes consist of thin ( $75 \mu\text{m}$ ) polycarbonate foils, which are metal coated to enhance conductivity and to avoid aging effects. The anode wires have a diameter of  $25 \mu\text{m}$ . There are two types of drift cells. Close to the beam their size is 5 mm. At larger distances modules with 10 mm cells take over. The wires run in the plane perpendicular to the  $z$ -axis. They are either oriented vertically ( $0^\circ$  view) or have a stereo angle of  $\pm 5^\circ$ . The length of drift tubes varies widely and can be as large as 4.5 m in the large TC superlayers. Inside the long tubes, the length of sensitive areas is limited by a subdivision. The signals from wires in sensitive areas are guided to amplifiers by means of thick wires where no gas amplification can occur.

Inside the magnet, there are three additional stations with chambers, which are referred to as *high- $p_T$  chambers*. They are used in connection with the trigger (Section 3.3). Each station covers roughly  $2.5 \text{ m}^2$ . PT1 is situated between MC05 and MC06, PT2 and PT3 between MC06 and MC08. They have the common feature that already a single layer of detection elements can provide all three spatial coordinates of a passing particle. Four independent gas pixel chambers form the inner part of the PT superlayers up to roughly 25 cm from the beam. The cells of the chambers have a quadratic structure. Four potential wires surround one sense wire and create one cell with a size of  $8 \times 8 \text{ mm}^2$ . The wires are parallel to the beam axis and have a length of 3.5 cm, which provides more than 1 cm of fully efficient gas amplification. The outer region above 25 cm is equipped with straw tube chambers having their sense wires aligned vertically. The straw tubes have a diameter of roughly 6 mm and lengths between 1.5–2.0 m. The base material of the straws is highly resistive carbon-loaded Kapton. A three-dimensional position measurement is accomplished by pads, which are connected to the readout in addition to the anode wires. The pad readout is realised by positioning a printed circuit board against the tubes in order to pick up the signals provoked by the gas avalanches. The signals are induced on rectangular pads whose sizes range from  $1.5 \times 3.2 \text{ cm}^2$  to  $3.0 \times 6.0 \text{ cm}^2$ .

The combined momentum resolution of the tracking system is very good. For muons from a  $J/\psi$  decay one obtains for example [19]

$$\frac{\Delta p}{p} = (8.1 \pm 0.3) \cdot 10^{-3},$$

which means that the mass of a  $B$  undergoing the golden decay can be measured with an accuracy of 8–10 MeV.

## 2.8 Particle Identification Devices

Identification of leptons and hadrons is accomplished by four different devices:

- the Ring Imaging Cerenkov Counter for separation of pions and kaons up to 50 GeV,

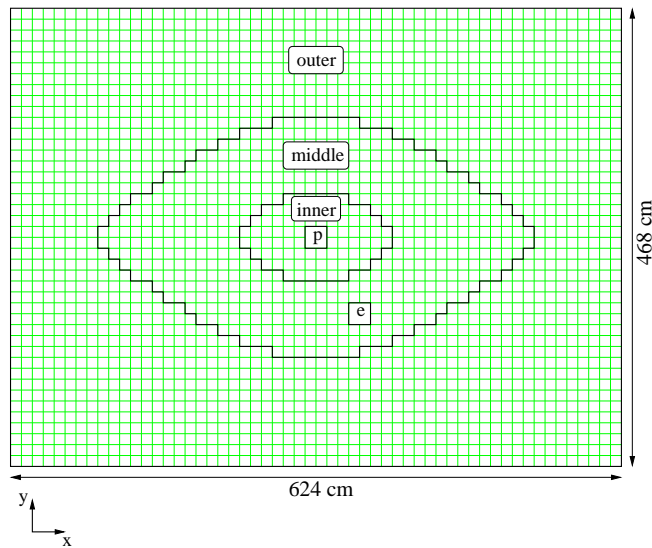


Figure 2.10: Layout of the electromagnetic calorimeter for electron and photon detection. In the outer region, one shown module (square) is equivalent to one cell, while one module in the middle and inner part houses 4 and 25 cells, respectively. The cell size is  $2.2 \times 2.2$ ,  $5.6 \times 5.6$ , and  $11.2 \times 11.2$  cm<sup>2</sup> in the inner, middle, and outer section.

- the Electromagnetic Calorimeter for the detection of showers induced by electrons and photons,
- the Transition Radiation Detector for electron-hadron separation, and
- a muon system.

The RICH detector is a large tank filled with  $C_4F_{10}$  (Freon) located between the PC and TC chambers. Particles travelling through the 2.7 m long gas volume emit light in a cone, if their velocity is above the Cherenkov threshold. The Cherenkov angle for particles with  $\beta=1$  is 55.6 mrad. A system of spherical and planar mirrors is employed to direct the photons onto photo multiplier tubes (PMT) outside of the particle flux. The radius of rings detected in the focal plane translates into the opening angle of the light cone, from which the velocity of the particle follows. The ring position determines the track slope. In combination with the momentum obtained by the tracking system, the masses of particles can be inferred.

The ECAL detects electrons and photons as showers. It is situated between the tracking and the muon system and has a geometrical acceptance of 68% for electrons and positrons from a golden decay. The ECAL is subdivided into three sections (outer, middle, inner) with increasing granularity towards the beam pipe in order to ensure an overall occupancy below 10% in spite of the widely varying particle flux (Fig. 2.10). About 6000 cells sample shower particles in a scintillator material sandwiched with absorber layers (Fig. 2.11). Wavelength shifter fibres are deployed perpendicularly to the alternating scintillator and absorber layers ("shashlik" technique) and guide the light to PMTs. The absorber in the middle and outer region is lead and the overall depth of a cell amounts to  $33 X_0$ . In the inner section with a depth of  $20 X_0$ , absorber layers of a tungsten alloy keep the shower radius small.

In the region of highest particle density, the electron identification system is completed by a transition radiation detector assisting the ECAL in the electron-hadron separation. The TRD is positioned in front of the calorimeter between TC01 and TC02. Its outer dimensions are

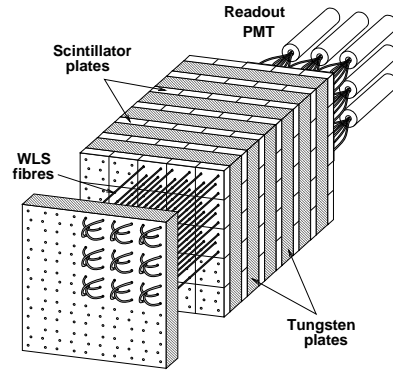


Figure 2.11: Sketch of an ECAL module with 25 cells. The wavelength shifter fibres run perpendicularly to the scintillator and absorber layers and are connected to the photo cathodes of the PMTs behind the module.

given by the geometrical extension of the inner ECAL. It consists of four large rectangular modules, which are rotated by  $\pm 30^\circ$  to match the shape of the inner ECAL. The TRD is based on straw proportional counters embedded in a polypropylene fibre radiator. Radiator sheets and sensitive layers form a periodic structure with 32 layers in total. The Kapton straws with a diameter of 6 mm serve as ionization chambers (for transition radiation photons) and drift chambers (for passing tracks) at the same time. The presence of transition radiation is detected by counting ionization clusters above a threshold, while the drift time information is used for tracking. The TRD is to accomplish a pion rejection of up to  $10^{-2}$  at 95 % efficiency for electron identification [20].

The muon system is the last part of the detector. It consists of four superlayers (MU01–MU04 in direction of increasing  $z$ ) interleaved with concrete and steel absorbers imposing an effective cutoff for muons with momenta below 4.5 GeV. The muon system is not expected to contribute to the momentum measurement but was rather designed to link track segments and to deliver muon track seeds for the event reconstruction and the trigger (Section 3.3). Its geometrical acceptance is slightly lower (220 mrad in the bending plane) than in the main tracking system. No absorber is deployed between MU03 and MU04 to obtain a clean measurement of track directions after the filter, which is not affected by multiple scattering.

The demands with respect to detector granularity are much weaker than in the main tracking system due to the lower occupancy. The maximal size of detector elements is limited by the requirement that all drift times fit into one bunch crossing. In the region close to the beam ( $< 33$  cm), gas pixel chambers are used in all superlayers. These chambers have their 30 mm long sense wires oriented along the  $z$ -axis. One drift cell of  $9 \times 9$  mm<sup>2</sup> (and slightly larger in MU04) is formed by four potential wires surrounding the sense wire. The detector technology applied at larger radii varies from superlayer to superlayer. In MU01–MU02, tube chambers consisting of sense wires in rectangular ( $15 \times 12$  mm<sup>2</sup>) aluminium profiles are used. They measure particles in projections of  $-20^\circ$ ,  $0^\circ$  and  $+20^\circ$  with respect to the  $y$ -axis. The chambers in the outer region of MU03 and MU04 provide a measurement in both  $x$  and  $y$  with one hit for trigger purposes (Section 3.3). These pad chambers have a cell size of  $14 \times 12$  mm<sup>2</sup> (Fig. 2.12) and vertical sense wires only. Three sides of a single cell are formed by a rigid aluminum structure. The cell is closed by a Cu covered G10 board with a pad structure.

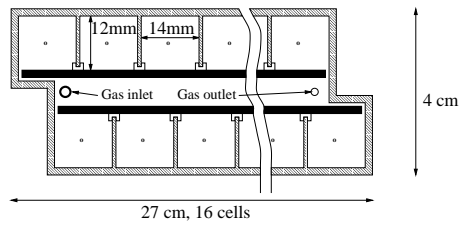


Figure 2.12: Schematical drawing of a muon pad chamber module. The relative displacement of cells ensures uniform efficiency throughout the module.

# Chapter 3

## Data Acquisition and Trigger System

Hadronic  $B$  factories like HERA- $B$  offer higher initial rates of  $b\bar{b}$  events than experiments at  $e^+e^-$  storage rings. Provided the  $b\bar{b}$  cross section is indeed of order 10 nb, events with  $b$  quarks are generated by the HERA- $B$  target at a rate of roughly 40 Hz when running at an interaction rate of 40 MHz. For experiments at the  $\Upsilon(4S)$  the  $b\bar{b}$  rate is limited by the smaller cross section and the attainable luminosity. At peak luminosities of  $3 \cdot 10^{33} \text{ cm}^{-2} \text{ s}^{-1}$  events with  $b$  quarks occur only at a rate of 3 Hz (cf. Table 3.1).

The higher  $b\bar{b}$  rate is bought at the expense of a less clean operating environment in hadronic collisions. Events with  $b$  quarks constitute only a small fraction of the total cross section, implying an unfavourable signal-to-background ratio. At the  $\Upsilon(4S)$ , this ratio is of order unity and the total event rate is in principle low enough to direct the resulting data stream directly to a permanent storage media. This is clearly impossible for hadronic  $B$  experiments. Even at the LHC centre-of-mass energy of 14 TeV, events with  $b$  quarks will be suppressed by more than two orders of magnitude and must be selected by means of sophisticated trigger systems before they can be stored at a rate of about 100 Hz. The selection itself is made difficult by the complicated topology of the hadronic events. The tracks from the decay of heavy quarks are accompanied by lots of particles from the underlying event. Resolving these events requires detectors with high granularity and consequently a large number of readout channels.

Table 3.1: Comparison between selected present and future  $B$  experiments.

experiment	BaBar	HERA- $B$	LHCb
accelerator	PEP II ( $e^+e^-$ )	HERA ( $pN$ )	LHC ( $pp$ )
centre-of-mass energy	10.6 GeV	41.6 GeV	14 TeV
$\sigma_{b\bar{b}}$	1 nb	$\approx 14$ nb	500 $\mu\text{b}$
$b\bar{b}$ rate	3 Hz	$\approx 40$ Hz	250 kHz
$\sigma_{\text{inel}}$	3.4 nb	13 mb	80 mb
$\sigma_{b\bar{b}}/\sigma_{\text{inel}}$	0.28	$\approx 10^{-6}$	$6.2 \cdot 10^{-3}$
read-out channels	80.000	600.000	1.100.000
time between bunches	4.2 ns	96 ns	25 ns

With respect to triggering, the HERA- $B$  experiment poses a particular challenge. The ratio of  $b\bar{b}$  cross section and total cross sections is only  $10^{-6}$ . This means that even after reducing events without heavy quarks by a factor  $10^3$ , the fraction of signal events is still lower than in a sample obtained by a random trigger at LHC energies.

The rate at which events can be recorded for permanent storage is restricted to some 10 Hz. The HERA- $B$  trigger system must therefore be capable of reducing the primary event rate by six orders of magnitude. On behalf of an efficient data-taking, this should be done without introducing dead-time. The efficiency of selecting  $B$  events should be of order unity, notably in the golden channel and in  $B \rightarrow \pi^+\pi^-$ . Due to the large number of readout channels and the tight bunch spacing, the trigger system has to cope with huge data streams. They can only be treated by employing up-to-date electronics, fast data links, and a seamless interplay of data acquisition and trigger system.

At the time of writing, the data acquisition and trigger system is in a commissioning phase and experiences improvements on short time scales. The following description refers therefore to the plans detailed in the HERA- $B$  Technical Design Report [13].

### 3.1 Trigger Strategy

The trigger strategy is shaped by the goal to record as many golden events  $B \rightarrow J/\psi K_s^0$  as possible. It is based on the following physical properties of  $B$  hadrons and their decay products:

- \* **The  $b$  quark is distinctly heavier than the quarks of the first and second family:** The decay particles of  $B$  hadrons feature a larger average transverse momentum ( $p_T \approx 0.8$  GeV) than tracks from interactions with light quarks due to the large mass of  $B$  hadrons and their transverse momentum distribution with respect to the proton beam. In decays, where the phase space is shared between few comparatively light daughter particles (e.g.  $B^0 \rightarrow \pi^+\pi^-$ ,  $B_s^0 \rightarrow K^+K^-$ ), tracks with high momentum are created. The bending of such tracks in a magnetic field is relatively small.
- \* **The lifetime  $\tau_B$  of  $B$  mesons amounts to about 1.5 ps:** They decay via  $b \rightarrow c$  transitions dominantly in  $D$  mesons, the lifetime of which is lower, but still sizable. Consequently, the daughter particles emerge from secondary (and possibly tertiary) vertices which are spatially separated from the point where the interaction occurred. The average distance is given by  $\beta\gamma c\tau_B$  and amounts to about 10 mm. The daughter particles have measurable impact parameters with respect to the primary vertex.
- \* **Charged leptons with high energy can be readily identified among the various hadrons generated in proton-nucleon collisions:** Electrons and positrons cause electromagnetic showers in a calorimeter. Muons can be detected thanks to their characteristic aptitude to penetrate thick absorbers by signals in adequate chambers.
- \* **Invariant masses of decaying particles provide valuable kinematical constraints:** Pairs of opposite-sign leptons (hadrons) whose invariant mass is compatible with 3.1 GeV (5.3 GeV) signal the presence of a  $J/\psi$  ( $B$ ) in the event.

The decay of the  $J/\psi$  to a lepton pair provides a clear signature which should be exploited for both the electron and the muon channel. Electron and muon candidates are identified as ECAL showers and track segments in the muon system. These localised signals define *regions of interest* (RoIs), which are then developed into tracks in the region behind the magnet. A target constraint allows to estimate the momentum of the tracks and the invariant mass of track pairs with unlike charges. Events are accepted if two leptons of the same flavour with unlike charges occur whose invariant mass matches the  $J/\psi$  mass.

This  $J/\psi$  trigger grants access to  $b\bar{b}$  events where at least one  $B$  meson decays via

$$B \rightarrow J/\psi(\rightarrow l^+l^-)X.$$

The branching fraction of these decays is about 1%. Since the trigger is derived completely from one  $B$  meson, the other  $B$  particle can be studied without any constraint. In addition, the trigger on events with a  $J/\psi$  provides a handle to the physics of directly produced  $J/\psi$  and  $\chi$  states. The cross section for directly produced  $J/\psi$  is about 400 nb per nucleon, implying that these events occur at a substantial rate.

The decay channel  $B \rightarrow \pi^+\pi^-$  for the measurement of the angle  $\alpha$  of the unitarity triangle can be accommodated in a trigger on a large invariant mass as well. For this purpose, the unlike-sign leptons are replaced by two stiff pions, which have to fulfil the requirement of an even higher invariant mass. Likewise, decays like  $\Upsilon \rightarrow l^+l^-$  can be selected.

To some extent, the trigger on two lepton tracks with a large invariant mass includes events in which both  $B$  mesons decay semileptonically. If the leptons are of the same type and their charges are unlike, they can be selected when the calculated invariant mass comes close to 3.1 GeV. Of course, this applies only to a small fraction of the double-semileptonic events and excludes for instance cases where one neutral  $B$  meson oscillated into its antiparticle. The HERA- $B$  trigger strategy opens a broader window on  $B$  physics by a general *dilepton trigger*. An appropriate trigger setting is to look for a combination of two leptons of (almost) arbitrary flavour and to place cuts on the transverse momentum and separation from the primary vertex.

The implementation of the described triggers requires the extraction of lepton and hadron tracks from the event data. With these tracks at hand, the  $J/\psi$  and the dilepton trigger can be complemented by generalised trigger settings. The trigger criterion is then a combination of lepton and hadron tracks whose large transverse momentum makes it likely that they originate from  $B$  hadrons.

## 3.2 Overview of the Trigger System

The features listed at the beginning of the last section are the basis of trigger decisions. Different levels of complexity are required when extracting these features from the raw data of an event. The amount of involved data and the needed processing time per event vary widely. In general, the event rate cannot be reduced sufficiently in one step, which enforces to breakdown the trigger system into several levels. Each trigger level accomplishes a distinct rejection of ordinary inelastic events (suppression) at a certain efficiency for signal events. In general, suppression and efficiency cannot be maximised simultaneously. The actual values for each trigger level are a compromise with respect to physical goals and technical (and financial) boundary conditions. The output rate of level  $N$  needs to match the input rate that level  $N+1$  is able to master. The entire system has to reduce the event rate to a value acceptable for writing to mass storage.

In most cases, the largest suppression is achieved by the lowest level facing the highest input rate to cut down the overall data stream as early as possible. Dedicated electronics is used at this stage to analyse and transmit events to the next level sufficiently fast. The trigger decision is based on parts of the event data only. Going up in the trigger system, the available time and the amount of data included in the decision for the event are increasing. Standardised hardware and – in contrast to customised electronics – freely programmable triggers can be used. For HERA- $B$ , the general scheme is the following (cf. also Table 3.2):

- \* **Level-1 Trigger (First Level Trigger, FLT):** The FLT consists of a network of roughly 100 custom-built massively parallel and pipelined processors. Each processor is assigned

to a particular detector region. It receives the corresponding detector data at the full initial event rate of 10 MHz. The processing of an event is started by pretrigger signals (see below). The pretriggers point to a region of interest in the tracking system, which is developed into tracks. The trigger decision is made by applying cuts on the transverse momentum of tracks and the invariant mass of track pairs. The time available is  $12\ \mu\text{s}$  and a suppression of 200 must be achieved.

- \* **Level-2 Trigger (Second Level Trigger, SLT):** The SLT can handle an input rate of 50 kHz including rate fluctuations up to 100 kHz. It is implemented with programs running on a farm of 240 Linux PCs. In principle, each computer can access the entire data of an event over a fast network. For reasons of processing time, the SLT takes over the validated RoIs from the FLT and requests additional event data. It refines the trigger decision by utilising the full chamber resolution (i.e. drift times) and adding of tracking planes not connected to the FLT. The average processing time per event is about 4 ms. The attainable suppression depends on the kind of event in question. In particular for the  $J/\psi$  trigger the rate can be reduced to a level where no inspection by the next trigger level is needed anymore. For other classes an anticipated reduction factor of 100 results in an input rate of 500 Hz for the subsequent trigger level.
- \* **Level-3 Trigger (Third Level Trigger, TLT):** The Level-3 trigger is the first system having the entire data of an event for rapid access in its local memory. Therefore the algorithms used are not confined to the RoIs but can include any part of the data in their processing. The TLT shall be executed on the same PC farm as the SLT. With a processing time below 100 ms per event it should reduce the rate by another factor of 10. Conceivable algorithms for such a system are subject to Chapter 5 of this thesis.
- \* **Level-4 Trigger (Fourth Level Trigger, 4LT):** The input rate to the Level-4 trigger is about 50 Hz. A second large farm of Linux PCs, totalling 200 processors, performs online reconstruction, basic classification and selection of events. The usage of the complete information about an event should allow for a further reduction by a factor 2.5 before data are stored on magnetic tape. The implementation of the reconstruction farm is detailed in Chapter 4.

In the following, the functionality of FLT and SLT is described as a prerequisite for the studies presented in Chapter 5. When reviewing the overall structure of the data acquisition system, some emphasis is put on the implementation of the SLT as a farm of commodity PCs.

Table 3.2: Overview of the four trigger levels.

level	input rate	reduction	time per event	hardware
1	10 MHz	200	$12\ \mu\text{s}$	custom-built processors
2	50 kHz	100	4 ms	PC farm
3	500 Hz	10	100 ms	PC farm
4	50 Hz	2.5	2–4 s	PC farm

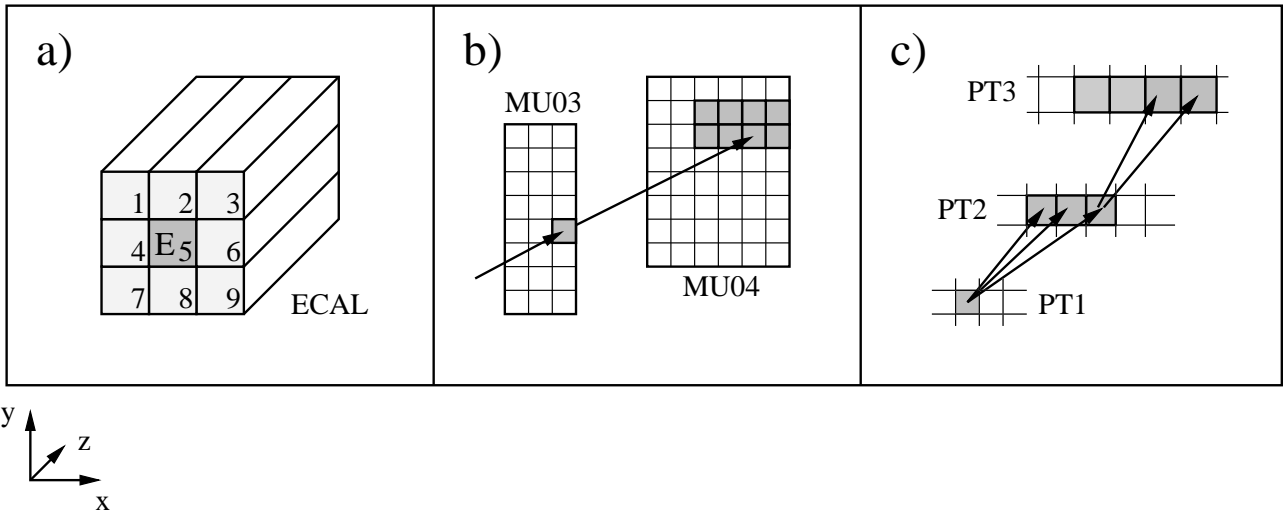


Figure 3.1: Illustration of the conditions that must be fulfilled in order to initiate a pretrigger in ECAL (a), in the two rearmost muon superlayers (b), or in the high- $p_T$  chambers (c). For (a) the smallest unit shown is a calorimeter cell, for (b) and (c) a pad.

### 3.3 Pretriggers

The processing at the first two trigger levels uses the data from regions of interest. Pretrigger electronics supplies several RoIs per event to the FLT. It does not cause a reduction itself since in general at least one RoI is found in every event. The RoIs form the starting point of the FLT processing. There are three sources of pretriggers:

- cluster in ECAL (electron pretrigger),
- track segments in the muon system (muon pretrigger), and
- track segments in the high- $p_T$  chambers in the magnetic field (high- $p_T$  pretrigger).

The pretriggers have the common feature that trigger signals are derived from fixed coincidences between neighbouring detector regions. Their output information is a bit pattern, which describes the established RoI and contains a unique event identification (bunch crossing number). It is fed into the FLT electronics.

#### Electron Pretrigger

The electron pretrigger identifies electromagnetic showers with a minimum transverse energy in the ECAL. For this purpose each calorimeter cell is put in coincidence with its eight direct neighbours (Fig. 3.1 (a)). An electromagnetic shower profile is detected by the requirement that 50% of the total energy  $E_{\text{total}}$  is released in the central cell:

$$E_5 > \frac{1}{2}E_{\text{total}} = \frac{1}{2} \sum_{i=1}^9 E_i.$$

The total energy must exceed a threshold, which depends on the distance between the cluster candidate's position at  $(x, y)$  and the beam axis ( $R = \sqrt{x^2 + y^2}$ ):

$$E_{\text{threshold}} = K_{\text{trig}} \cdot \left( \frac{1}{R} + \frac{1}{\sqrt{x^2 + |y|^3}} \right).$$

The asymmetric treatment of  $x$ - and  $y$ -coordinates accounts for the bending of tracks in the  $x$ - $z$  plane. The adjustable parameter  $K_{\text{trig}}$  ( $\approx 700$  GeV cm) is the product of the minimal transverse momentum required for the cluster (typically 0.5 GeV) and the spatial separation of target and ECAL ( $\approx 1400$  cm). It is an effective handle to control the event rate to and from the FLT. According to simulations, a  $K_{\text{trig}}$  of about 750 GeV cm gives rise to 5 RoIs per event from the ECAL when running at 40 MHz [21].

Beyond the formation of coincidences, the pretrigger electronics calculates the centre-of-gravity of clusters to obtain a better position estimate, which allows for RoIs smaller than the ECAL cell size. Furthermore, it attempts to recover energy losses due to Bremsstrahlung. The cluster position of photons emitted before the magnet can be calculated from the electron (or positron) candidate's cluster position, its energy and the known  $p_T$ -kick in the magnet ( $\approx 650$  MeV) and using the assumption that the particle comes from the target region. There are two such positions, corresponding to the electron and positron hypothesis. The energy of a cluster found in either position is added to the primary cluster. Including the Bremsstrahlung correction, the processing time can be as high as  $2 \mu\text{s}$ . The message passed to the FLT contains cluster energy and position estimate.

## Muon Pretrigger

The muon pretrigger detects three-dimensional muon track seeds whose origin is compatible with the target region. It exploits coincidences between two pads in the muon chambers of the superlayers MU03 and MU04. About 7800 pads are arranged in a projective geometry. In proximity of the proton beam, they are formed by an OR of neighbouring channels in the gas pixel chambers. At larger radii, a pad readout is accomplished by an appropriate structuring on the cathodes of the proportional chambers. A pretrigger is generated under the condition of simultaneous signals on one pad in MU03 and several associated pads in MU04 (Fig. 3.1 (b)). The absence of absorber layers between MU03 and MU04 helps to avoid a deterioration of the muon candidate's track parameters by multiple scattering. The derived position and direction of the detected coincidence are packed into a message and sent to the FLT. The maximum extension of found RoIs is 12 cm. Due to the smaller occupancies behind the hadron filter the muon pretrigger finds on average only two RoIs for events with 5 superimposed interactions. Therefore, the decay  $J/\psi \rightarrow \mu^+ \mu^-$  can be isolated cleanly and has a higher trigger efficiency than  $J/\psi \rightarrow e^+ e^-$ .

## High- $p_T$ Pretrigger

The electronics of the high- $p_T$  pretriggers receives its input from the three PT superlayers in the magnetic field. Signals are derived from the gas pixel chambers and the straw tubes with pad readout. Owing to the higher track densities in the magnet area in comparison with the muon system, a triple coincidence between pads is required (Fig. 3.1 (c)). A mutual shift of the pads ensures that only particles with a transverse momentum greater than 1.5 GeV (or random coincidences of several tracks) can initiate a pretrigger. The messages to the FLT comprise the

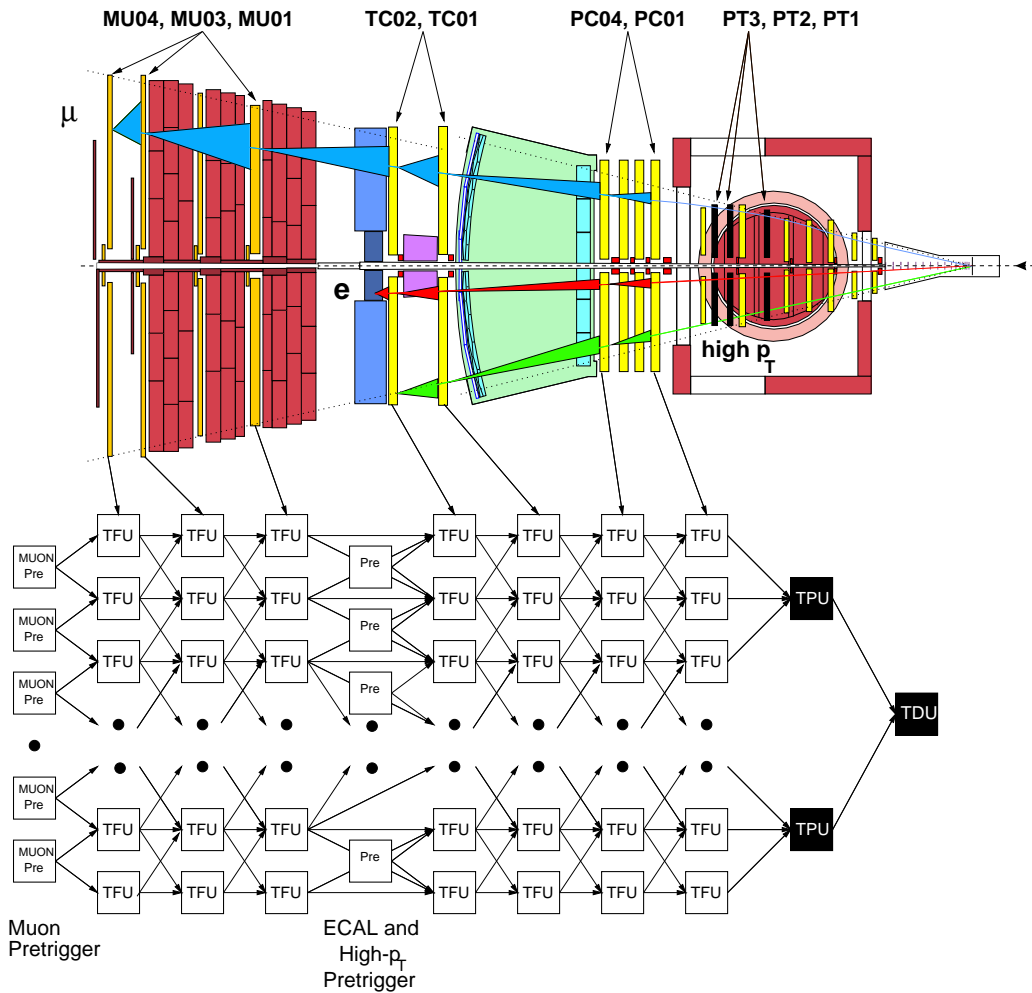


Figure 3.2: Functional view of the First Level Trigger. Track Finding Units (TFU) are assigned to regions in selected superlayers. The tracks search is initiated by pretrigger signals derived from the muon system, the ECAL, or the PT superlayers in the magnet. Tracks are propagated in RoIs using the hit information from the detector by message exchange between TFUs.

parameters of found coincidence roads. The high- $p_T$  pretrigger gives rise to about 3 RoIs per event with 5 interactions.

### 3.4 First Level Trigger

At the pretrigger stage, electromagnetic clusters of electrons cannot be distinguished from photon clusters without associated track. Likewise, the track seeds of muon and high- $p_T$  pretrigger need to be verified. For the high- $p_T$  pretrigger for example, random coincidences caused by different particles are ten times more abundant than real tracks. The FLT evolves the track seeds and cluster candidates into tracks using hit information from selected superlayers behind the magnet. The programmable trigger decision involves cuts on the number and flavour of found tracks ( $e$ ,  $\mu$ , hadron  $h$ , photon  $\gamma$ ), their transverse momenta calculated using a target constraint, and invariant masses of track pairs.

The FLT operates on data from all views of the four ITR/OTR superlayers (PC01, PC04,

TC01, TC02) and from three muon stations (MU01, MU03, MU04). In some detector regions, neighbouring channels are combined in a logical OR to reduce the granularity and to ensure almost 100% hit efficiency. The overall structure of the FLT is illustrated in Fig. 3.2. The FLT hardware is a network of so-called Track Finding Units (TFU) with data links for message exchange. Each TFU carries several dedicated processors and is assigned to a particular region in a superlayer. It receives patterns of fired detector channels synchronously for each bunch crossing and buffers them in a local wire memory for the last 128 events. A second, but asynchronous input to a TFU are messages with estimated RoIs and track parameters stemming either from a pretrigger or a TFU board further away from the target. The FLT tracking is performed on the basis of the hit data and the messages exchanged between different TFUs. It is completely message driven and uses techniques from Kalman filtering to propagate track candidates towards the target. The parameters of a track candidate are refined in each step by adding more hits.

At a muon pretrigger for example (indicated in the upper half of the detector in Fig. 3.2), a message with the extension of the RoI defined by the coincidence road between MU03 and MU04 is sent to the TFU responsible for the region in MU04 covered by the RoI. The TFU checks in its wire memory for hits in the RoI defined by the track seed. If no hit is found the track candidate dies. Otherwise the track parameters are updated by adding the information of matching hits from the different views. The RoI is extrapolated to the next tracking station upstream using the new track parameters. A message with the RoI and refined track parameters is sent to the responsible TFU which in turn checks for matching hits and contacts the next TFU in line. Of course it can happen that there is more than one way to prolong a track using the local data of one TFU. In that case, one new RoI is created for each group of matching hits and passed on to the responsible TFU. Likewise, if a geometrical RoI falls in a channel range covered by two neighbouring TFUs, the extrapolated RoI is sent to both modules and the track search takes place concurrently.

The described procedure is similar for electron and high- $p_T$  pretrigger, except that the pretrigger track candidates are extrapolated to TC02 and initiate the track finding in one of the TFUs there. At the most upstream superlayer connected to the FLT (PC01), the TFUs contact – after adding the last hits successfully – another type of electronics module. Two Track Parameter Units (TPU) translate the arriving messages into track parameters at the target and sort out multiply reconstructed tracks. The tracks are assumed to come from the target. At this stage the wire which caused the interaction is not known yet. A mean target position serves therefore to estimate the track momentum from the slope behind the magnet. The momentum estimation makes use of lookup tables and reaches a resolution better than 5%. The TPUs calculate the transverse momenta of track candidates and reject tracks below a programmable threshold (typically 0.5 GeV). Electron candidates<sup>1</sup> are rejected if the found momentum does not match the ECAL energy. Photons are established as very high- $E_T$  ECAL clusters ( $E_T > 3.5 \text{ GeV}$ ) without associated track. The final decision whether to keep the event is taken by a Trigger Decision Unit (TDU) on the basis of the derived track parameters.

The propagation of track candidates in the FLT network is an asynchronous procedure. Each TFU processes tracks from different events in parallel. The first task of the TDU is therefore to identify tracks belonging to the same event. The number of tracks above a programmable  $p_T$ -threshold is counted separately for different kinds of particles ( $e$ ,  $\mu$ ,  $h$ ,  $\gamma$ ). Tracks with unlike charges are combined into pairs and the corresponding invariant masses are calculated. The obtained resolutions of the track parameters allow to estimate the invariant mass of a  $J/\psi$

---

<sup>1</sup>The ECAL cluster energy is part of the messages exchanged by the TFUs for electrons and positrons.

decaying to  $\mu^+\mu^-$  with an accuracy of 3%. There are two distinct classes of triggers issued by the TDU:

- **Lookup Table Triggers:** The trigger decision is based on a track pair. The input to the lookup table (LUT) is bit pattern which is coding the flavours ( $e, \mu, h, \gamma$ ), the charges, the momenta, and the transverse momenta of the two involved tracks. The pattern is completed by the invariant mass of the track pair and serves to address a single bit in the LUT. For the  $J/\psi$  trigger one requires a  $e^+e^-$  or  $\mu^+\mu^-$  pair with an invariant mass exceeding typically 2.0 or 2.5 GeV, respectively. The lower value for electrons accounts for Bremsstrahlung losses. Also decays with larger invariant mass ( $B \rightarrow h^+h^-, \Upsilon \rightarrow l^+l^-$ ) are selected by this trigger.
- **Count Triggers:** These triggers complement the LUT trigger when an invariant mass is not a powerful constraint. Instead, one demands a number (1–8) of tracks of a certain flavour, optionally in combination with a  $p_T$ -cut. Also the presence of certain combinations of particle flavours, like ( $e, h$ ), can serve as trigger criterion. Count triggers are aiming for instance at the dilepton trigger where both  $B$  mesons decay semileptonically. One requires two leptons of arbitrary charge and flavour with a  $p_T > 1$  GeV. Simulations show, however, that the copious case of two electrons must be excluded to avoid a too high trigger rate at this stage.

The FLT can operate several triggers in parallel. Each trigger class can be downscaled not to overrun the SLT. For accepted events, the FLT provides the SLT with a list of the satisfied triggers and a list of all tracks.

Table 3.3: FLT performance.

Process	Efficiency
$B \rightarrow J/\psi K_S^0 \rightarrow e^+e^- K_S^0$	55 %
$B \rightarrow J/\psi K_S^0 \rightarrow \mu^+\mu^- K_S^0$	35 %
$B \rightarrow \pi^+\pi^-$	25 %

The FLT is a sophisticated system allowing to trigger on events with a  $J/\psi$  already at the earliest level. Its performance for the most important channels was derived from simulations [22] and is summarised in Table 3.3. The quoted efficiencies include the geometrical acceptance of the spectrometer and the pretrigger requirements.

### 3.5 Second Level Trigger

The comparatively coarse FLT tracking is susceptible to wrong tracks combining hits from different particles. 75% of the track candidates delivered by the FLT are such ghosts. Most of the electron-positron pairs found by the FLT do not come from the target, but are due to photon conversions before and in the magnet. Other potential background sources are electromagnetic clusters caused by hadrons, photon clusters accidentally overlaid with hadron tracks, thus faking an electron, or muons from hadron decays in flight. The SLT repeats the tracking in the regions of interest using the full chamber resolution to eliminate RoIs that do not contain real tracks. For good-quality tracks established behind the magnet, kinematic cuts can be applied with improved momentum resolution. Eventually, the SLT can explore a region not accessible

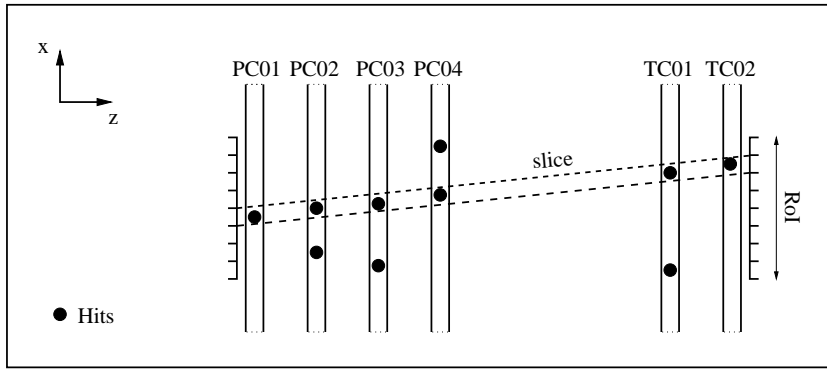


Figure 3.3: Illustration of the Slicer algorithm used for fast track verification in the six superlayers behind the magnet (PC01–PC04, TC01, TC02). It is checked that at least one slice contains an appropriate number of hits. (The sketch is not to scale. One bullet denotes several possible hits in the  $x$ -view of a superlayer. See text for further explanations.)

for the FLT. The tracks are propagated through the magnetic field defining a RoI at the exit window of the vertex tank. In the last step, the tracking is extended to the vertex detector to resolve the target region. Table 3.4 provides an overview of the SLT processing.

Table 3.4: SLT algorithms.

Algorithm	Used Detectors	Integrated Suppression
Slicer	PC01–04, TC01–02 ( $x$ -view)	4
$x$ - $z$ -Refit	PC01–04, TC01–02 ( $x$ -view)	7
$y$ - $z$ -Refit	PC01–04, TC01–02 ( $x$ - and $y$ -view)	13
L2Magnet	MC01	13
L2Sili and L2Vertex	SVD	100

### 3.5.1 Tracking behind the Magnet

The verification of tracks found by the FLT is accomplished by three distinct algorithms. They operate on data from all six large superlayers in the region behind the magnet and in front of the calorimeter, i.e. PC02 and PC03 are added to the stations already used in the FLT. The tracking algorithms are applied to all track candidates. The event is discarded if less than two tracks pass the current algorithm’s selection criterion.

The first step, called **Slicer**, combines hit counting and a variant of a Hough transform. The SLT requests the data from the  $0^\circ$ -layers ( $x$ -view) in the superlayers for all channels in a narrow band around the track candidate. Slicer verifies that there is at least one hit in each superlayer. In the bands remaining after this cut, a fast pattern recognition is attempted, which is illustrated in Fig. 3.3. Both ends of a band in PC01 and TC02, respectively, are divided into  $N$  sections. Connecting one section in PC01 with one in TC02 defines a slice,  $N \times N$  slices in total. It is checked whether there is at least one slice containing at least one hit in each superlayer as well as at least 12 hits in total. A value of  $N=8$  was found to be a good compromise between rejection power, efficiency and execution time [23]. Simulations indicate that a suppression of 4 is attainable.

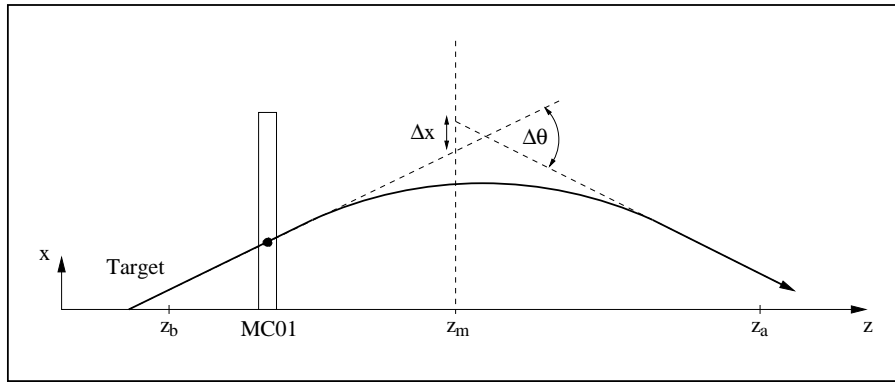


Figure 3.4: Sketch of a bending track in the magnetic field. In the magnet part of the SLT algorithms, the effect of the magnetic field is described as change of the slope  $\Delta\theta$  and a shift  $\Delta x$  at some plane  $z_m$  in the magnet. (See text for further explanation.)

The second step in the tracking, **x-z-Refit** [24], operates on the same data as Slicer but also takes advantage of the drift times for the outer tracker. It is a Kalman fit in the  $x$ - $z$  plane resolving all ambiguities and calculating a  $\chi^2$  for the candidate. The  $\chi^2$ -test is a good handle to reject ghosts efficiently. The overall suppression after the refit is about 7. More suppression can be gained when confronting the track candidates with data from the rotated views. These data are requested by the SLT for the **y-z-Refit**. The  $y$ - $z$ -Refit establishes three-dimensional track candidates and applies ultimate track quality cuts.

The combined suppression of the tracking algorithms is about 13. At this stage also the RoIs in background events contain real tracks thus enforcing to exploit the information from detectors closer to the target in order to achieve further reduction.

### 3.5.2 Magnet Traversal

The propagation of tracks found as straight lines in the pattern tracker through the magnetic field is a difficult and time-consuming task due to the non-linear track model and local field inhomogenities. The momentum obtained from a target constraint allows to estimate the initial bending. In the final reconstruction of events [25], optimised numerical methods [26] and an accurately measured field map are employed to propagate track candidates from chamber to chamber and to look for confirming hits. For trigger applications this track-following is much too slow. Instead, a faster method [27] utilising lookup tables, which does not need data from the chambers in the magnet, is used.

The **L2Magnet** algorithm translates the track parameters after the magnet (slopes  $t_{x,a}$  and  $t_{y,a}$ , intercepts  $x_a$  and  $y_a$  at  $z_a$ ) to their counterparts before the magnet ( $t_{x,b}$ ,  $t_{y,b}$ ,  $x_b$ ,  $y_b$  at  $z_b$ ) and estimates charge over momentum  $Q/p$ . (Here the indices  $a$  and  $b$  stand for *after* and *before*.) The relation between the track parameters before and after the magnet is described at a plane  $z_m$  in the centre of the magnet (Fig. 3.4). In this plane, the bending results in a shift  $\Delta x$  of the track and a change of its slope  $\Delta\theta$ . In the  $x$ - $z$  plane the following equations hold (and similarly in  $y$ ):

$$\begin{aligned}
 t_{x,a} &= t_{x,b} + \Delta\theta \\
 x_a &= x_b + (z_m - z_b) \cdot t_{x,b} + \Delta x + (z_a - z_m) \cdot t_{x,a}
 \end{aligned}
 \tag{3.1}$$

The algorithm is based on the observation that  $\Delta\theta$  and  $\Delta x$  can be parameterised as polynomials in  $Q/p$ :

$$\begin{aligned}\Delta\theta &= \frac{Q}{p_{xz}} \cdot A1(t_{x,b}, t_{y,b}) \\ -\Delta\theta + \Delta x &= \frac{Q}{p_{xz}} \cdot A5(t_{x,b}, t_{y,b})\end{aligned}\tag{3.2}$$

$A1$  and  $A5$  are first and second order integrals of the magnetic field. Their values for all slopes in the detector acceptance were extracted on a grid of  $23 \times 15$  ( $t_{x,b} \times t_{y,b}$ ) using a detector simulation and the measured field map of the HERA- $B$  magnet. Using Eqs. (3.1) and (3.2) one can create lookup tables for the track parameters before the magnet as a function of their values behind the magnet.

L2Magnet extracts the track parameters before the magnet and of  $Q/p$  in two iterations. In the first iteration,  $x_b$  and  $y_b$  are assumed to coincide with the mean target position. The found track parameters define the centre of a new region of interest in the first tracking superlayer before the magnet (MC01, see Fig. 3.4). The boundaries of the new RoI are the result of a proper error propagation for all involved quantities including effects like Bremsstrahlung and multiple scattering. Hits found in the RoI provide better estimates for  $x_b$  and  $y_b$  than the previously used mean target position. After the second iteration, the resolution for  $Q/p$  is about  $5 \cdot 10^{-4}$ , which is a factor of 10 worse than in the reconstruction program, but the result was obtained more than one order of magnitude faster [27].

The efficiency of L2Magnet exceeds 90% for electron and muon pairs from  $J/\psi$  decays. The track parameters of all processed particles are transported to the exit window of the vertex vessel. The RoIs in this position are the starting points for the subsequent silicon tracking.

### 3.5.3 Tracking and Vertexing in the SVD

The last section of the SLT algorithms has two parts: track propagation to the target and vertexing. The vertex detector (described in Chapter 2) measures a particle trajectory with strips under four different angles ( $\pm 2.5^\circ$  with respect to the  $x$ - and  $y$ -axis). The track propagation in the SVD, **L2Sili** [28], is a Kalman filter algorithm which transports the track parameters from one SVD superlayer to the next and checks for hits in the projected RoI. For timing reasons the  $x$ - and  $y$ -view are treated separately. A track candidate is regarded as confirmed if at least three matching hits in both  $x$  and in  $y$  are found and the  $\chi^2$  is acceptable. The condition of at least six hits can be fulfilled for tracks hitting two or more superlayers. It rejects particles that originate from conversions in one of the rear silicon superlayers or the exit window. The track parameters of surviving candidates at the  $z$ -position of the target are estimated by giving larger errors to the hits at larger  $z$  to account for the fact that the resolution in the SVD is governed by multiple scattering (Section 5.4.4). The track parameters allow to verify that at least two trigger tracks come from the same target wire. This requirement removes events where tracks from interactions on different wires initiated a trigger. The suppression attainable in this step scales with the number of operated target wires.

Further cuts applied to the trigger tracks must result in the overall suppression of 100 expected from the SLT. Their actual implementation depends on the type of event in question and the performance of the trigger levels passed up to this point. For event topologies with two trigger particles having the same origin, the vertexing part of the silicon algorithms (**L2Vertex**) determines the vertex position. This applies to  $B \rightarrow J/\psi X$ ,  $B \rightarrow h^+ h^-$  (Table 3.5 (a)) and to

Table 3.5: Event topologies for signal and background processes encountered by the SLT in the SVD. The black circle denotes the position of a target wire. The solid lines are particles found in the RoIs. The symbol  $l$  stands for  $e$  or  $\mu$ ,  $X$  and  $Y$  denote any other particles. (Further explanation in the text.)

	Topology in SVD	Processes
a)		$B \rightarrow J/\psi X, B \rightarrow h^+ h^-$
b)		direct $J/\psi$
c)		$B\bar{B} \rightarrow lXlY, D\bar{D} \rightarrow lXlY$
d)		$B \rightarrow D l X \rightarrow l Y l X$
e)		Particle Misidentification

the direct  $J/\psi$  production (b). The  $\chi^2$  of such a vertex fit alone should allow for an effective control of the SLT output rate regardless of the vertex position with respect to the wire. In addition, one can exploit momentum conservation and check whether the two known momenta sum up to point back to the target wire. Such events will be accepted without further stringent examination by subsequent trigger levels.

The situation is more complicated when there are no vertices formed by two trigger particles. This is the case for the dilepton trigger. It is intended to trigger events in which a pair of  $B$  or  $D$  mesons decays semileptonically (Table 3.5 (c)) or the decay cascade of one  $B$  generates two leptons with sufficient  $p_T$  (d). The leptons are found by FLT and SLT in the respective RoIs. For such events the SLT can determine the distance of closest approach (DCA) of both tracks and their impact parameters with respect to the target wire. The DCA will not be very small, otherwise the particles would form a good common vertex. On the other hand a large DCA ( $> 3\text{ mm}$ ) is very unlikely when both particles really come from the same interaction, i.e. the same primary vertex. Further cuts can be applied on the measured impact parameters to keep the rate under control. The efficiency of a cut at  $100\ \mu\text{m}$  is about 60% for leptons from  $B$  mesons and about 30% for leptons from  $D$  mesons. The dilepton trigger is, however, much more susceptible to background processes which are sketched in Table 3.5 (e). Tracks coming from the primary vertex, for instance misidentified hadrons, can fake large impact parameters due to resolution effects if their rate is not kept low enough in the previous trigger steps. For the dilepton trigger additional suppression might be necessary on the next trigger level. We will come back to this question and to a more detailed discussion of impact parameters and their resolutions in Chapter 5. For the moment, we turn to the structure of the data acquisition system which is connecting the four trigger levels.

### 3.6 Structure of the Data Acquisition System

The data acquisition system moves the event data from the detector to archive. It has to ensure the synchronisation of subdetectors and their readout with respect to the bunch structure of the HERA proton beam. Already this synchronisation is a challenging task since the time of flight of particles through the 20 m long HERA- $B$  detector is only 30 ns smaller than the bunch spacing. We do not pursue the techniques used for synchronisation [29] but rather describe how the data acquisition system makes data available to the trigger levels.

The implementation of the HERA-*B* data acquisition system resembles the systems designed for the future LHC experiments due to the large number of readout channels and similar requirements with respect to the data rates. Wherever possible, the number of different hardware modules used was restricted to the necessary minimum. This measure should limit the expertise needed in building and operating the various components and allows for a homogeneous system. A functional view of the data acquisition system is shown in Fig. 3.5. Amplifiers directly mounted on the **Detector** hardware in the experimental area guide the signals from 600.000 electronic channels to about 1400 so-called front end driver boards (**FED**). The custom-made FEDs are subdetector-specific and serve as a uniform interface to the data acquisition system. Besides signal processing (ADC or TDC conversions), they fulfil three different tasks. Parts of the event data are sent via intermediate electronics (**1**) to the **FLT**. This intermediate hardware performs a preprocessing of the data, like generating of bit patterns of fired channels for tracking devices. The FEDs keep the event data in a pipeline during the FLT processing. The pipeline depth of (at least) 128 bunch cycles results in  $96 \text{ ns} \cdot 128 = 12.3 \mu\text{s}$  available for execution of FLT algorithms and communication of a possible FLT accept to all FEDs. Events accepted by the FLT are communicated (**2**) to the **Fast Control System** (FCS). The FCS electronics initiates the readout of the data belonging to this event from the pipeline by sending synchronised signals to all FEDs (**3**). Each FED sends out its data along with a header denoting subdetector type, channel range, and packet length. The length of the packets varies from 64 to 300 bytes depending on the subdetector.

The raw data of a single event encompass about 450 kB. The readout of FLT accepted events from the pipeline at 50 kHz involves thus the transmission of a 22 GB/s data stream. The destination of the data packets sent by the FEDs is another event buffer, called Second Level Buffer (**SLB**). The SLB holds the data during the SLT step and answers requests for parts of the event data. It consists of digital signal processors (DSP) of the type ADSP21060 (SHARC<sup>2</sup>) connected to the FEDs by means of fast data links, which allow to meet the requirement of a 22 GB/s data input rate. The entire SLB consists of 134 so-called SHARC cluster boards in VME standard, each carrying six DSPs (Fig. 3.6). On a single SHARC cluster board, a fast global bus (240 MB/s) and DMA capabilities allow for communication of the six CPUs and rapid access to 512 kB memory per processor. Each DSP has six data links (SHARC-links) at its disposal. On the links half-bytes are transmitted at a 80 MHz cycle, resulting in a bandwidth of 40 MB/s. All 36 SHARC-links of a cluster board are available for connections on the front panel. The DSPs are interconnected to the FEDs, to other DSP boards, and finally to the SLT by appropriate cabling. In this way, the relation between a particular FED and a SHARC cluster board is fixed, i.e. one board stores the data from a defined detector region, but for several hundreds of events. In contrast to the pipelined event storage in the FEDs, a managed buffer system is used in the SLB. An event buffer is released and overwritten only after the SLT rejects an event or all data were read out from the SLB. This managed buffer allows the SLT to operate under the weaker constraint of maximum average times instead of maximum absolute times for trigger decisions.

For an estimated average SLT processing time of 3.8 ms, 190 processors are needed to cope with the 50 kHz output rate of the FLT. The PC farm used for the **SLT/TLT** comprises 240 machines in total to allow in addition for TLT processing. In the SLT/TLT, one event stored in the SLB is assigned to one processor under supervision of an **Event Controller** (**4**). The Event Controller software [30] is implemented on a dedicated SHARC cluster board. It sends and receives control messages via SHARC-links and schedules the assignment of events to SLT

---

<sup>2</sup>SHARC stands for Super Harvard Architecture. It is a trade mark of Analog Devices.

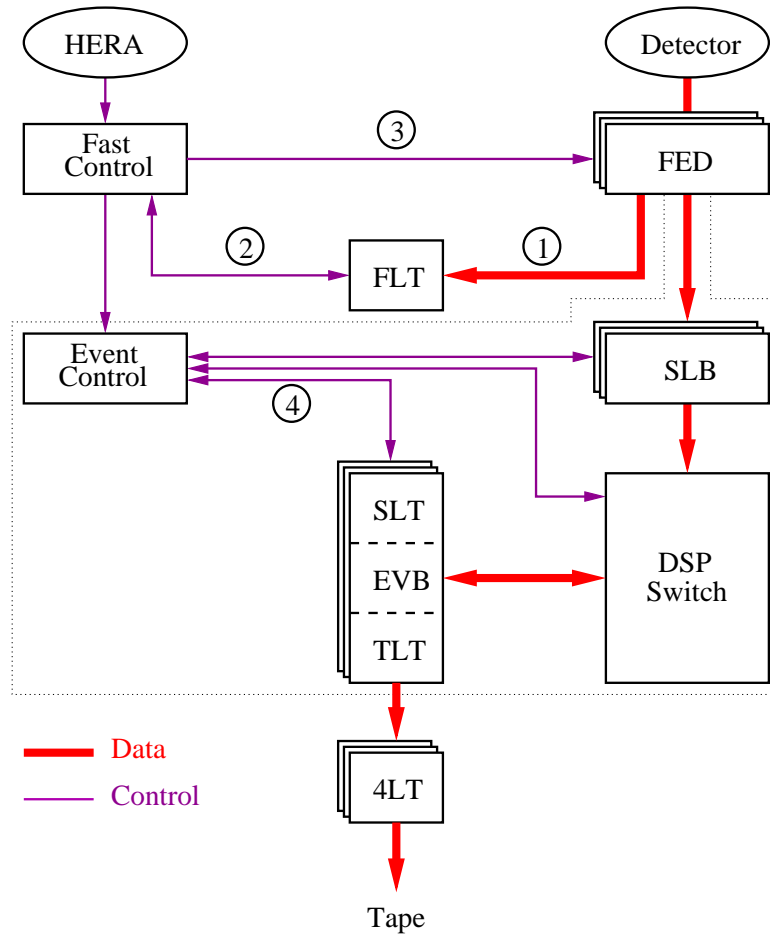


Figure 3.5: Overall structure of the data acquisition system. The data flows from the pipeline on the FEDs to the SLB where it is kept during SLT processing. The event building takes place after the SLT step. All data belonging to one event are read from the SLB over a fast network (DSP Switch). In the dotted region, event data and steering messages are exchanged using SHARC-links. Event data are sent from the SLT/TLT farm to the online reconstruction farm via ethernet. (See text for a more detailed description.)

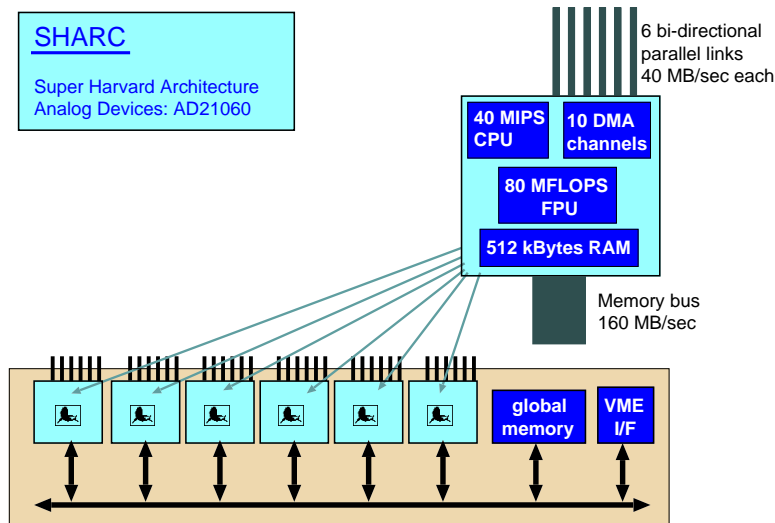


Figure 3.6: The SHARC cluster board is a 6U VME-module carrying six DSP ADSP21060 with 512 kB memory per processor. Internal communication uses a global bus (240 MB/s) while message transfer to the outside world uses six SHARC-links per processor (40 MB/s per link).

processors. The designated SLT processor applies the SLT algorithms described above to the event. Due to the restriction to RoIs, the fraction of data requested per event is smaller than 0.5 %. The resulting data stream of about 50 MB/s is rather moderate in absolute numbers, but consists of many small ( $< 100$  bytes) request messages at a very high sustained rate ( $> 100$  kHz). These messages can be directed from any PC in the SLT/TLT farm to any DSP in the SLB requiring about  $10^5$  point-to-point communication channels. A fast network with a high degree of connectivity is needed. It must support to establish a connection between any two points with a minimal startup time. This startup latency should be negligible in comparison with the SLT processing time.

Such a network (**DSP Switch**) was realised for HERA-*B* using the same hardware as in the SLB. The entire system is made up of 44 SHARC cluster boards connected via SHARC-links in an appropriate way. Data portions and messages are passed from one DSP to another on the same SHARC cluster board and sent over SHARC-links to boards closer to the final destination. Each processor's decision where to move a message is taken from routing tables. They are loaded when initialising the data acquisition system. The average time spent by a SLT PC for requesting and receiving a data portion from the SLB is about  $70 \mu\text{s}$  [30].

The data of an event accepted by a SLT processor are gathered from the SLB over the DSP Switch. This procedure is called event building. The event-builder software (**EVB**) contacts each SHARC cluster board in the SLB to extract the respective data portion for the event. Event building at 500 Hz causes an overall data stream of 250 MB/s between SLB and SLT/TLT farm. After this step, the event data reside in the local memory of the SLT processor. In the event assembly, the application of zero-suppression algorithms reduces the event size to 150–250 kB. The **TLT** is executed on the same processor to avoid a needless transfer of event data to another CPU. Accepted events are sent to the online reconstruction farm (**4LT**) via fast-ethernet.

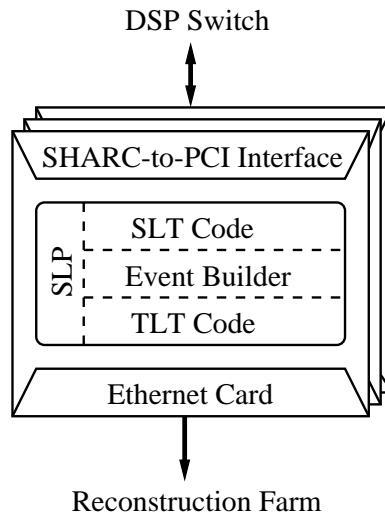


Figure 3.7: Each PC of the SLT/TLT farm communicates with the SLB and the Event Controller over SHARC-links. A SHARC-to-PCI interface allows for data transfer in both directions. All functionality is put into a single process (SLP) housing SLT algorithms, event builder software and TLT code. Accepted events are sent to the reconstruction farm using ethernet.

### 3.7 Implementation of the SLT/TLT as a Farm

Except for Slicer and x-z-Refit, the SLT algorithms use different and independent parts of the event data. This suggested at the time of the HERA-*B* Technical Design Report [13] to build the SLT as a parallel and pipelined system. Like in the FLT, specialised processors assigned to detector regions were to refine RoIs. Event building was in the realm of a processor farm running both TLT and online reconstruction. With the availability of increasingly powerful processors it was realised that a system made up of commodity processors in their standard hardware environment could do the same work, exploiting the favourable price/performance ratio of these chips. Another important step was the transition from a RoI-based to an event-based processing. In the old scheme, one processor would take over all calculations needed for one RoI, implying that several CPUs are devoted to the same event. Another central instance and additional communication were needed to combine the results of all RoIs into a trigger decision for the event. In the new design, there is a one-to-one correspondence between events and processors.

The SLT/TLT farm consists of 240 standard PCs with Pentium CPUs running the Linux operating system. The configuration of both hardware and software reflects that it is an online-like system, aiming to minimise the latency of context switching and data exchange with the SLB. The machines are equipped with 32 MB of memory, a fast-ethernet card, a simple graphics card, a floppy drive, but no hard disk. Apart from booting, a disk provides no benefit since it is too slow to be accessed during triggering. The booting of machines proceeds with the help of a floppy disk. The only non-standard hardware component is a card in each PC allowing for a connection between SHARC-links and the PCI-bus (SHARC-to-PCI interface) [31]. It transfers data bi-directional at a rate of 40 MB/s. Each SLT/TLT PC equipped with such a card has two SHARC-links which are used separately for reading and writing. The choice of Linux as operating system was motivated by its widespread use in the experiment and the benefits of the open source policy. The installed system is a lean version of a standard-

distribution. Since Linux as such is lacking realtime capabilities, all attempts are made to limit the number of processes needed for operation. To avoid frequent context switches the entire functionality is put into one single executable. This so-called Second Level Process (SLP) comprises SLT algorithms, event builder software and TLT code (Fig. 3.7). The SLP is the exclusive owner of the SHARC-to-PCI interface and runs with root privileges for this reason. The SLP communicates with the Event Controller and the SLB over SHARC-links. Messages are received by polling from the SHARC-to-PCI and the ethernet interface. Accepted events are sent out to the online reconstruction farm over ethernet. In the SLT and EVB part of the SLP, the frequent reallocation of memory blocks is usually avoided for performance reasons. The SLT uses, for example, one large static structure to hold the RoIs, hits and tracks of an event. This implies a maximal number of these quantities which can be handled per event. The structure is dimensioned such that practically no event is discarded due to overflow.

The SLT/TLT farm is an efficient and flexible solution for triggering. Its major benefits are scalability and the possibility to implement arbitrary trigger algorithms as long as timing constraints are obeyed. The latter feature was exploited widely in the development and commissioning phase of HERA-*B*. In partial absence of the pretrigger and the FLT, the functionality of these components was emulated on the farm. This provided the muon group with an event sample enriched with tracks in the muon system which can hardly be obtained when taking data with a random trigger. Another successful example is connected with the calorimeter. At a time, when the ECAL pretrigger could give rise to a single cluster only, a version of the offline ECAL reconstruction software was incorporated into the SLP. The cluster finding was applied to the ECAL data pulled from the SLB to the respective SLT processor. Events without a second electromagnetic cluster above an energy threshold were rejected.

The operation of a farm with several hundreds of running processes necessitates careful monitoring and surveillance. To be successful in terms of triggering, each SLP must be provided with proper calibration constants at all times. The SLT performance should be checked centrally by filling monitoring histograms. The mechanisms used for these purposes are very similar to solutions developed for the online reconstruction farm. The online reconstruction farm is the subject of the next chapter.

# Chapter 4

## Online Reconstruction Farm

The handling of the recorded data and in particular the reconstruction of the events pose a challenge to HERA-*B* for two reasons. The first reason is the large amount of data that the experiment will record per running period. Archiving of event data is foreseen at a rate of 20 Hz, which results in a data volume of 20 TB per year assuming 100 kB/event and  $10^7$  s/year. The principal storage media for data sets of this size is still the magnetic tape, even though the application of hard disks for larger parts of the event data is anticipated for the future LHC experiments. The speed at which data can be transferred to and read from the central tape robot in the DESY computer centre is limited to roughly 5 MB/s.

The second reason, influencing the data handling to a high degree, is the time required for reconstruction of events. The high multiplicity of the hadronic events recorded by the spectrometer makes an impact on the algorithms used for pattern recognition, track fitting, and particle identification. The sophisticated algorithms developed for event reconstruction have a computing time of 2–4 s per event on today's microprocessors.

Such a relatively long computing time per event requires a huge number of CPUs to preserve a reasonable ratio between the rates at which events are recorded and reconstructed. On the other hand, it is impractical to read back the entire data volume and to distribute it to the CPUs possibly several times. For this reason, HERA-*B* is following the approach of taking the reconstruction program to the data rather than vice versa. The event data leaving the data acquisition and trigger system are subjected to an online reconstruction and classification by means of a dedicated reconstruction farm.

### 4.1 Farm Concept

One of the major benefits of the online reconstruction farm project is a shortened cycle between data analysis and a possible feedback to the data-taking. Immediately after data-recording, information about the data quality is available, which allows for rapid tuning of detector and trigger settings. Another advantage is the possible usage of the farm as an event filter. Trigger algorithms used at this stage can exploit the complete information about an event. In addition, the online reconstruction allows for an early event classification and selection, providing the physicist who is carrying out the data analysis with dedicated data sets. This last item is especially pronounced for HERA-*B*, since the rate of recorded *B* events is well below 1 Hz and therewith just a small fraction of the entire event data stream.

The promising features of an online reconstruction are bought at the expense of new problems to be solved, though. An online version of the reconstruction program is needed, which

does not access an input file, but receives data directly from the data acquisition system. There must be a mean to distribute the event data stream to the reconstructing CPUs and to bring it together again for centralised mass storage. The key problem of online reconstruction is, however, the availability of the correct calibration and alignment constants at the time of event processing on the farm. Additional software systems are mandatory in order to meet this requirement.

The number of reconstructing CPUs can be estimated from the event rate passing the TLT and the average processing time per event:

$$50 \text{ Hz} \cdot \frac{4 \text{ s}}{\text{event/CPU}} = 200 \text{ CPUs.}$$

As in the case of the SLT/TLT farm, none of the microprocessors has to communicate with another CPU in the farm, which poses a simplification of the overall system. Each CPU applies a copy of the reconstruction program to events assigned to it by an external steering mechanism. The input bandwidth into the farm is about 5 MB/s, which is well within the realm of standard network technology.

Facing the comparatively large number of CPUs required, pricing considerations influence the implementation of the reconstruction farm. Besides a more favourable price/performance ratio than offered by multiprocessor mainframes, an easy scalability of the farm's computing capacity is desired. This demand can be met when it is comparatively simple to increase the number of CPUs or to replace them by the next generation of faster, up-to-date processors. Under the viewpoint of hardware maintenance and availability of components through the lifetime of the experiment, the usage of standardised off-the-shelf products is advisable. The software environment, notably the operating system, should be compliant with the offline reconstruction program, which is developed on UNIX platforms exclusively. The ultimate goal is to be in a position where an overall reconstruction program can be started on the farm without any further modifications.

## 4.2 Hardware Implementation

A farm of standard PCs running the freely available Linux operating system fulfils the demands described above to a high degree. It allows to stick to off-the-shelf hardware everywhere in the system. Every PC is equipped with a 13 GB hard disk, 128 MB of memory per processor and a simple graphics card. In contrast to the situation in the SLT/TLT farm, the latency time of context switching and data transfer does not play an important role, as it is completely negligible in comparison with the long computing time per event. Therefore, the most efficient solution is the usage of motherboards with two PentiumIII processors (500 MHz). Both CPUs can share the same network interface card for communication with the outside world. The performance decrease due to the non-exclusive access of either CPU to the common main memory was shown to be very small. This holds only, if the memory deployed is sufficient to prevent a paging of memory segments to disk (swapping).

The used network technology is fast-ethernet (100 Mb/s). A group of 16 PCs, called **mini-farm** (Fig. 4.1), is plugged into a common switch. All minifarms are connected via a **farm switch** to the SLT/TLT farm, the network at DESY, and three **farm host computers**. One of these host computers (**Srv**) provides the application software for the entire farm centrally by functioning as an NFS server (Network File System). The other two machines take over responsibilities in conjunction with data logging (Section 4.6) and farm monitoring (Section 4.9).

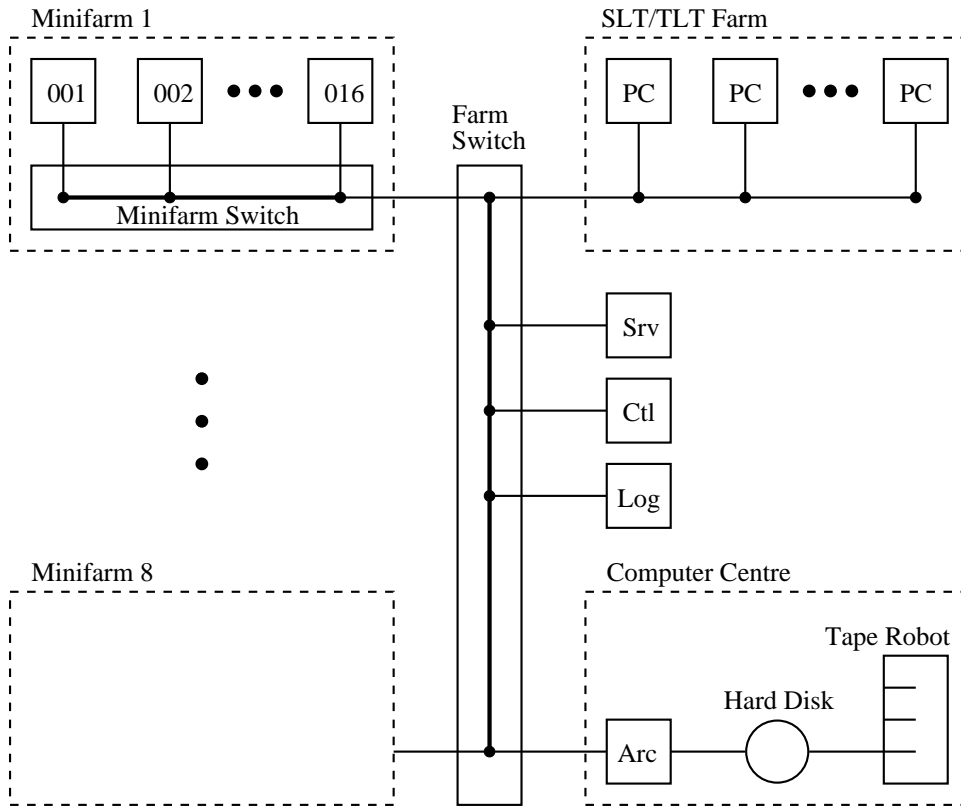


Figure 4.1: Scheme of the reconstruction farm and its network topology. (See text for explanation.)

The arrangement of 100 mini-tower PCs with two CPUs each in four shelves results in quite compact setup, which requires only limited infrastructure. The farm occupies one room with good air-cooling above the experimental area in the western hall of HERA.

### 4.3 Software Environment

The following software systems are needed to distribute events to the processor farm, to perform the reconstruction in a proper and checkable way, and to pass the event output to central data storage:

- The **control software** solves the assignment of an event coming from any of the  $N$  SLT/TLT PCs to one of the  $M$  reconstruction CPUs. It is a farm-wide system.
- Each CPU runs a copy of the **application software**, receiving events under the supervision of the control software and applying the reconstruction program to the data. Finally, the reconstructed events are sent out to the
- **logging software** for permanent storage.
- During farm operation, event data quality and proper functioning of the reconstruction program must be checked. This is the task of the **monitoring software**, allowing to inspect histograms with, say, correlations between subdetectors or residuals of track fits.

- The output of the monitor software may indicate a change in the calibration or alignment of a subsystem. The **calibration and alignment software** takes over the calculation of a new set of usable constants, either on basis of the observed deviations or after acquiring additional data. After this step, it has to manage the distribution of its results to all interested instances in the data acquisition and trigger system.
- Furthermore, monitoring of the basic functionality of PC hardware and software (temperatures, fan speeds, processing load) is desirable. It should be decoupled from the event processing on the farm. The **slow-control software** provides a centralised view on the mentioned quantities and issues alarms when appropriate.

As mentioned earlier, the reconstruction farm is completely dominated by the computing time, while all latencies of context switching and data transfer can be neglected. In the software design, one can take full advantage of all UNIX features, like multiprocessing. The given tasks can be solved by cooperating processes on the same CPU, exchanging information and accessing data in memory concurrently. So the software makes use of Inter Process Communication (IPC), with facilities like message queues, semaphores and shared memory. The network communication with other components in the data acquisition and trigger system employs software layers on top of the protocols UDP (User Datagram Protocol) and TCP/IP (Transmission Control Protocol/Internet Protocol). UDP is a connectionless service, transmitting small (<1000 bytes) data packets with minimal overhead, but does not guarantee a message to reach its destination. This safety in delivering messages is inherent to TCP/IP. TCP/IP is connection-oriented protocol. Every incoming data portion is acknowledged by the receiver with an appropriate message. Losses in the data transmission can be detected and trigger the source to resend the data.

TCP/IP is effective and comes close to the maximal bandwidth possible on the network hardware, when a large amount of data (> 50 kB) is to be transferred between two persistent partners. The time to establish and close the connection is then much smaller than the overall duration of the data transmission. In the farm environment, TCP/IP is used for passing the events, whereas short control messages are exchanged with UDP.

In the following, the software systems are described in more detail. The emphasis is on the last three items of the list, due to the participation of the author.

## 4.4 Control Software

One PC of the SLT/TLT farm is scheduled to execute the control software. The PC is equipped with both an ethernet card and a SHARC-to-PCI interface. The control software communicates with the SLT/TLT farm via the DSP Switch and with the reconstruction farm via ethernet lines.

In a (naive) implementation, all reconstructing processes might register at startup with the control software, which inserts them one after the other in a waiting queue. Whenever a SLT/TLT PC accepts an event, it requests the network address of a reconstruction CPU from the control software to send out the event data. The control software returns the address of the currently first PC in the queue and moves it to the end of the list. With such a round-robin approach, any of the  $M$  reconstructing processors is given every  $M$ -th event. This simple scheme, however, does not account for computing times fluctuating from event to event, be it due to the intrinsic properties of the events or different processing speeds of the farm CPUs. These fluctuations introduce a potential inefficiency into the system. Either a reconstructing

CPU is idle when it has finished the last event and no new event is supplied, or the SLT/TLT processor must wait if the destination CPU is late. The latter case results in an larger average duration of the event data transfer, reducing the time available for triggering. Moreover, in the system sketched, the treatment of possibly died reconstruction processes is poor. A crash is not noticed by the control software, which keeps on giving the affected PC's address to SLT/TLT computers, obscuring them – and the entire network – by unreachable destinations.

The addressed problems can be solved by a buffer for several events on the receiving CPUs and an improved steering mechanism [32]. A sufficiently deep buffer guarantees that there is one event available for immediate reconstruction at any time and diminishes in that way the computing time fluctuations. The SLT/TLT computers can send out data at the lowest possible latency when the receiving side is able to put yet another event into the buffer.

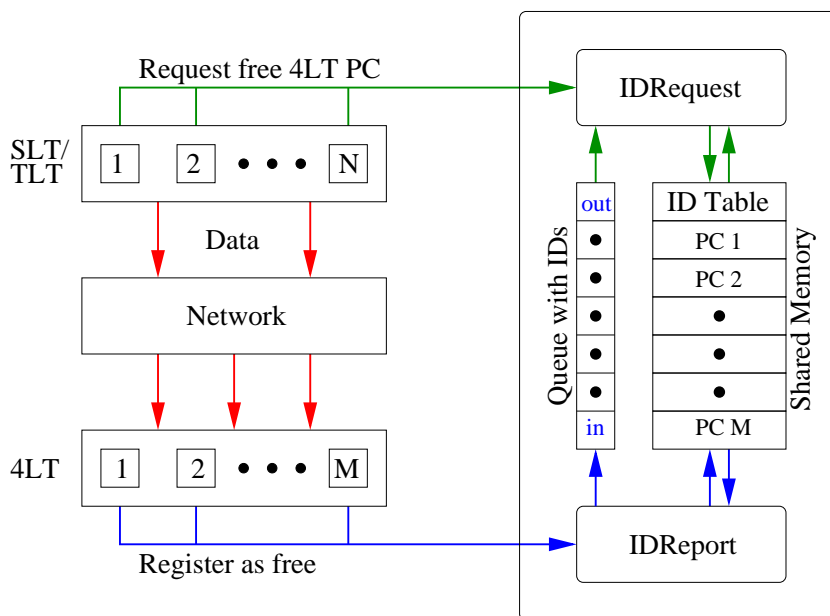


Figure 4.2: Scheme of the control software, steering the event flow from the  $N$  SLT/TLT PCs to the  $M$  reconstruction CPUs. (See text for explanation.)

In the implemented control software scheme, a reconstruction farm CPU registers anew as being ready for receiving data as soon as its buffer can hold at least one more event. More powerful CPUs will report the presence of free buffer space more often, thus receiving a greater share of the event stream. Possibly crashed reconstruction computers will simply not register anymore and are ignored by the system. The control software (Fig. 4.2) consists of two processes accessing a shared-memory segment and both ends of a message queue. One process (**IDRequest**) communicates exclusively with the SLT/TLT farm, the other (**IDReport**) with the 4LT farm. The latter deals with incoming registrations and inserts a CPU-specific identifier (ID) at the end of the queue. If a SLT/TLT PC has finished event processing with an accept, it turns to IDRequest in order to obtain a destination for the data. IDRequest removes the first ID from the queue and translates it using an **ID Table** to the corresponding PC's network address, which is then returned. The ID Table was filled by IDReport at the very first registration of each reconstruction farm CPU.

The described procedure assumes that the control software has at least one farm PC ready for data receiving in the queue whenever a request arrives. This is the case when the reconstruc-

tion farm and the logging system behind can handle the input rate. Otherwise the SLT/TLT PC must wait, hence introducing deadtime into the system. The resulting back-pressure affects all trigger levels and throttles the FLT in the end.

The implementation of the control software as two separate processes instead of one large program reduces the number of transactions, which one process has to handle per time interval. The highest demands to the system occur in situations with many involved computers and small event sizes. The robustness of the control software was shown in tests using events of a few kB in autumn 1999. Event rates exceeding 120 Hz were reached in a configuration with 45 SLT/TLT and 40 4LT PCs.

## 4.5 Application Software

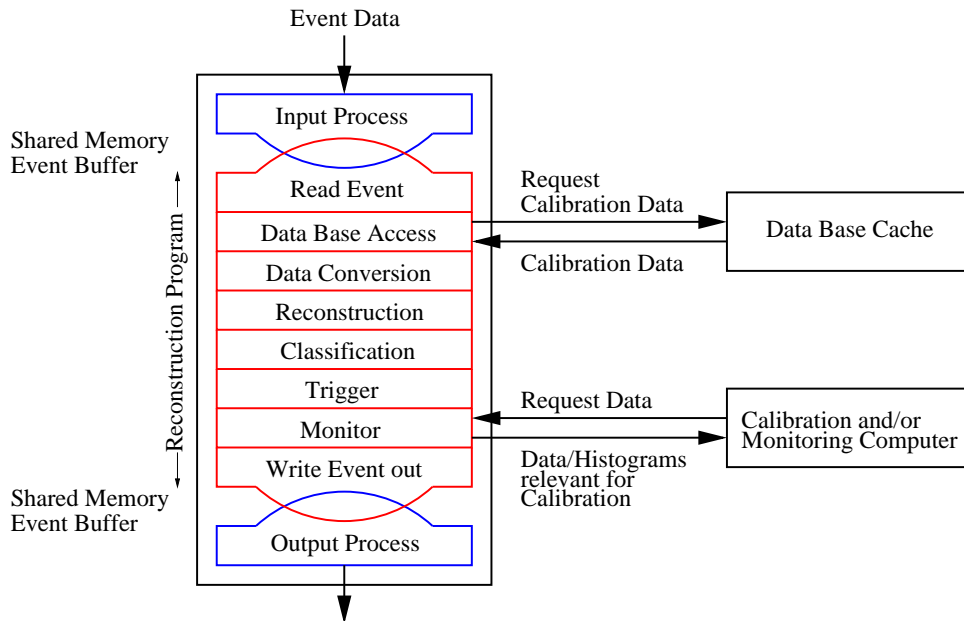


Figure 4.3: Overview of the software executing on each CPU of the online reconstruction farm. (See text for details.)

The application software consists of three processes, the first of which was already implicitly mentioned in the preceding section. It is an **Input Process** (Fig. 4.3) which communicates with the control software and copies arriving data to the event buffer. The buffer comprises several shared-memory segments ( $\approx 10$ ). The segments are dimensioned such that they can hold even the largest events. They are filled one after the other. The **Reconstruction Program** reads the next event data block from the first segment filled, thereby releasing it. The input-output routines of the HERA-*B* reconstruction program were modified to allow for reading from a shared-memory segment. The program checks at startup whether a name for an event data file was given. If not, it automatically attaches itself to a shared-memory buffer. This mechanism implies that there is no dedicated online-version of the reconstruction program. Every statically linked executable prepared for Linux and the farm's processor architecture can be run there. The separation between offline and online was overcome by making the libraries for shared-memory access part of every release of the HERA-*B* analysis software.

The event processing has several distinguished phases. After reading the event from the buffer, a calibration-key in its header is inspected. A change of this mark since the last event triggers the loading of new calibration and alignment data (Section 4.8). The next step is the transformation of raw detector data into the format expected by the various reconstruction modules (e.g. tables with hits). The succeeding call of the actual reconstruction routines generates new tables with found tracks and clusters, which become part of the event data. In the next phase, all pending requests for control histograms and data relevant for calibration and alignment are answered by returning appropriate data (see Section 4.7). Afterwards, classification and selection algorithms are applied to the event. They assign the event to a certain class and mark it accordingly. In the extreme case, the event is discarded in a trigger step, using the complete knowledge about the data.

Accepted events are scheduled for archiving, which includes another network transfer. The communication with the data logger is taken over by an **Output Process** in order to decouple the reconstruction program from this task. A buffer of shared-memory segments ensures that the reconstruction program can continue with the next event immediately after copying the present event to an unoccupied segment. The output process establishes a connection to the data logger and sends out events in the buffer.

## 4.6 Data Logging

At the data logger, all events assigned to the various farm CPUs by the control software come together again in one central place. It has to be capable of receiving and storing the event data stream of about 2 MB/s coming from the farm. The logging software runs on a workstation in the DESY Computer Centre (labelled as **Arc** in Fig. 4.1). It creates one thread of execution for each farm CPU taking part in the run. The transmission of event data from the output-processes to the logger employs TCP/IP. The events are stored in files with a size of currently 400 MB. The name of the file is derived from the run number and a counter, which is increased when opening a new file. The files are created initially on hard disks with a total storage capacity of 107 GB. A daemon-process copies them automatically to the tape robot.

On the event of a breakdown of data storage in the computer centre, the farm should allow for local recording. For that purpose, one of the farm host computers (**Log** in Fig. 4.1) was equipped with a hard disk capacity of 36 GB. At a data rate of 2 MB/s, HERA-*B* can continue data-taking for about 5 hours. This period can be extended by deploying a tape drive with a sufficient speed.

## 4.7 Monitoring Software

The primary task of the monitoring software is to provide information about the quality of the raw data and the reconstruction during an ongoing run. If this information points out changes in the calibration or alignment of a subdetector, the data required for a new calibration (like track segments or histograms with residuals) should also be transmitted.

The reconstruction farm as a whole can deliver monitoring data with high statistics due to the large number of involved CPUs. On the other hand the required data are generated locally on each processor during event reconstruction. They must therefore be collected in a central place for further analysis.

### 4.7.1 Concept

The design of such a system started from the consideration that the transmission of monitoring data has lower priority than the event data stream. The amount of network bandwidth needed was difficult to estimate at an early stage. The share of the monitoring data on the overall bandwidth should not be fixed by the loaded reconstruction program, but be subject to external adjustment by farm steering processes. In contrast to the event data, a fractional loss of monitoring data is acceptable, lowering the demands for a safe transmission over the network.

As a general rule, farm PCs do not send out monitoring data as histograms and other data objects actively (PUSH architecture), but rather answer requests by special monitoring computers for such information merely when need arises (PULL architecture). In this scheme, each reconstruction program running on a farm CPU is a server replying to requests of monitor programs (clients). Owing to the various subdetectors and different modules in the reconstruction program, there will be many clients which are interested in different data.

A software package to ease the writing of appropriate clients and servers was developed and documented [33] under the name of RHP (Remote Histogramming Package). It provides in a similar way to HBOOK<sup>1</sup> n-tuples and histograms, which can be sent over networks using a UDP-based protocol. In addition, the transmission of blocks with arbitrary data relevant for calibration is supported. RHP is hiding the network access and the distributed nature of data objects. From the user's point of view, requesting a histogram from all farm PCs proceeds in the same manner as for a single computer. One of the major goals of RHP is to allow the creation of general client programs which have to make as few as possible assumptions about the objects booked in a server. In practice, there is no need to recompile a client when adding new histograms or changing the parameters of existing histograms in a server. Another RHP feature is tailored to the needs of the SLT/TLT farm. Each object has a flag which can be toggled by clients. A deactivation indicates that the client is not interested in requesting the data at present. Time-critical servers can query this flag and skip possibly lengthy calculations and filling in of data.

There are other client-server relations besides farm-like configurations with  $M$  data sources and one client. They occur for instance when several users want to display a histogram which is being filled permanently by a monitoring task. In this case, concurrent requests by more than one client (display program) make no sense since the monitoring programs resets the histogram after each query to accumulate new data. In situations like that, a subscription mechanism is used [34]. An arbitrary number of clients can register as a subscriber for a specific histogram in a server. Once the histogram is full it is published and afterwards reset by the server. Each registered client receives a copy of the histogram upon publishing.

The functionality of RHP is general enough to build up different monitoring applications with varying number of involved processes for the data acquisition and trigger system. Interfaces to other histogram packages (HBOOK, Root<sup>2</sup>) and to the data base allow for graphics display and permanent storage of data objects. RHP is distributed with the versions of the analysis software ARTE to ensure that any reconstruction programs can run on the farm without modifications. Its application for the 4LT farm is now detailed with a concrete example.

---

<sup>1</sup>HBOOK is a histogramming package widely used in high energy physics.

<sup>2</sup>Root is a data analysis framework developed at CERN.

## 4.7.2 Data Quality Assessment

The quality of taken data is checked by filling appropriate RHP histograms over many events on the reconstruction farm. At the lowest level, the functioning of the subdetectors and their readout electronics is judged using ADC and TDC spectra. The graphical display of correlations between hits in different subdetectors as two-dimensional histograms shows whether all data in the event belong to the same bunch crossing, which is only the case for a proper synchronisation of all subsystems. At the highest level, selected results of the reconstruction are inspected, like the Cerenkov angle in the RICH for particles with  $\beta=1$  and the number of tracks found in the SVD.

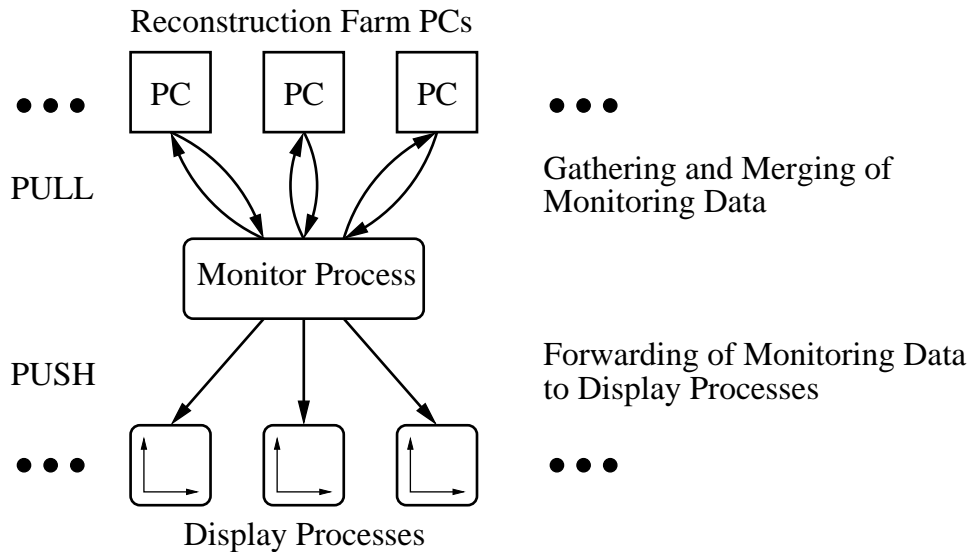


Figure 4.4: Acquisition and display of monitoring data using the reconstruction farm. (See text for explanation.)

The monitoring system is responsible for gathering histograms from the reconstruction farm PCs and to make them available to processes concerned with display and archiving. The used scheme is sketched in Fig. 4.4. The central part is a **Gatherer Process** acting as client (receiver of histograms) for the farm computers and as a server (source of histograms) for the various display programs. At system initialisation, the gatherer is started after all reconstruction processes. The gatherer finds out about the number of reconstruction processes and contacts one of them. (It does not matter which one, as all book and fill the same histograms.) The gatherer requests the parameters of all histograms and books an empty version in its context. Now display processes can be run as well. They locate the gatherer process and request a list with the key features (histogram ID and title) of all objects. The list is presented to the user who can select the desired histograms and receive them using the subscription service. The display process registers as a subscriber for the selected histograms at the gatherer and books an empty version of the histogram to prepare for arriving data. At regular intervals, the gatherer requests for each histogram, which has at least one subscriber, the data from all farm computers and merges them into one high-statistics histogram. The resulting object is published to every subscriber for display.

The described scheme illustrates that the gatherer determines the network bandwidth used up by monitoring data within the farm. The gatherer decides at what time intervals how many

histograms are requested. It could be steered to gather histograms less often when more clients register.

## 4.8 Online Calibration and Alignment System

Online reconstruction of events is sensible only when proper calibration and alignment<sup>3</sup> constants are available. A similar statement holds for the SLT and TLT depending on the presence of these information as well. Both SLT and TLT use the silicon vertex detector which can serve as an example. The SVD is a high-resolution device whose superlayers are moved individually with respect to the proton beam at the beginning and end of a run. An alignment on the level of  $10\ \mu\text{m}$  is vital for the correct functioning of the triggers. Any deviations from the assumed positions result in reduced trigger efficiency. They should be detected and corrected as fast as possible to resume effective data-taking.

Of course the calibration and alignment system is not confined to the reconstruction farm, but affects the entire data acquisition and trigger system. It is described in this chapter because many of the basic ideas emerged in connection with the reconstruction farm.

### 4.8.1 Concept

The main goal of the online calibration system is to reduce the time between the acquisition of data relevant for calibration and the distribution of updated constants to all levels as much as possible. The PCs of SLT/TLT and the reconstruction farm are the principal sources of data needed for calibration. They calculate for many events quantities like the energy of ECAL clusters and residuals of track fits. The basic idea is to gather this information by dedicated computers and to apply calibration algorithms to them in case of deviations from the current set of constants. The newly calculated constants should be distributed to all systems depending on them, ideally without starting a new run. The system must ensure that an event is processed on all levels with the same set of calibration constants and that the entire set can be identified unambiguously later in the offline analysis.

For some applications not triggered but randomly selected event samples are needed. Depending on the required rate either a dedicated calibration run is started or a subset of events is treated differently. In the latter case, random-trigger events are sent to predefined farm PCs. A monitoring process is used to gather calibration-relevant data from these computers.

### 4.8.2 Calibration Scheme

The demands to an online calibration system are met in the following way (Fig. 4.5): At the SLT, each event gets a marker (calibration key) denoting the entire set of calibration constants. Technically it is the version number of a data base table. This table in turn contains a list with the versions of all possible calibration data records. Usually there are several data records per subdetector. Whenever there is a new version of a record, a new data base table is created. The version of the table is the new calibration key which is put into the event headers from then on. The PCs of the reconstruction farm infer from a changed key that new calibration data must be read from an appropriate data base.

---

<sup>3</sup>In the following, detector alignment is included implicitly when the term "calibration" is used.

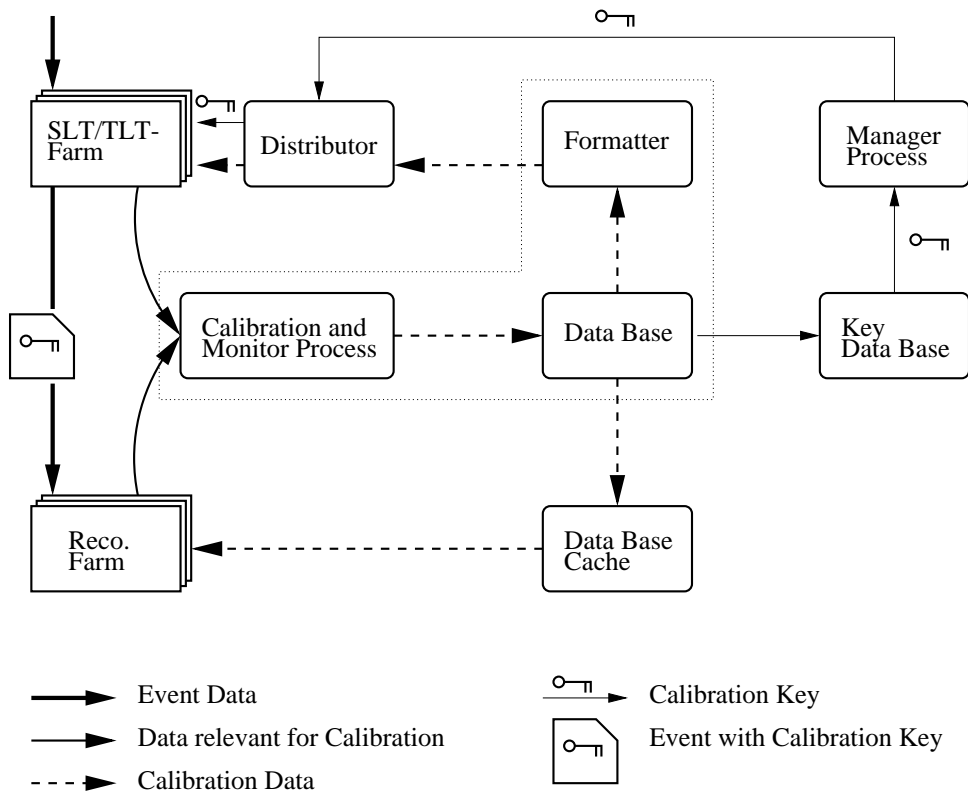


Figure 4.5: Overview of the systems involved in the online calibration system. The flow of control messages is not shown. The systems surrounded by the dotted line exist once per calibration record, the others are global. (Further explanation in the text.)

The **Gatherer and Calibration Programs** are the sources of new calibration constants. They are subdetector-specific processes and use RHP to request data relevant for calibration from the PCs of the trigger or reconstruction farm. New versions of calibration constants are stored in a **subdetector data base** containing all versions ever written. Using a subscription mechanism, a global **key data base** is notified. It creates a new table including the latest version. The central part of the calibration system is a global **Manager Process** which is informed by the key data base that new calibration records are available. The manager supervises the distribution of constants to the data acquisition and trigger system. The new table is read to extract which records have changed. In general, the format of calibration data in the data base differs from what the SLPs expect. The required processing is done by a **Formatter**. It receives the new version number of the affected record from the manager, reads it from the subdetector data base, and performs the formatting. Afterwards, control is given back to the manager. The manager instructs the data acquisition system not to accept new events for a short period. The SLT/TLT PCs empty all event buffers. A **Distributor** process takes over the transmission of the new calibration data record to all machines in the trigger farm. It decouples the formatter processes from the actual farm structure (number and location of PCs used in the current run). Under the command of the manager the data records are pushed from the formatter over the distributor to all SLPs. The new calibration key is sent from the manager to all SLT/TLT PCs using the distributor. Finally, the manager signals to resume accepting events. The new calibration key is put into the events entering the trigger farm. In that way it reaches the reconstruction farm and is used to extract the updated records

from the data base. In contrast to the pushing of records to the SLT/TLT computers, the data base access has to resemble the offline-case for the reconstruction farm. There is a potential bottleneck when  $\mathcal{O}(100)$  reconstruction programs try to request records from a data base at almost the same time. That many requests cannot be handled over the network by a data base stored on hard disks. The implemented solution is a **Data Base Cache**. It resides in the main memory of a dedicated data base machine situated close to the reconstruction farm. The data base cache holds only the last versions of calibration constants. It is updated like the subdetector data base prior to the distribution of new calibration records (lower branch in Fig. 4.5).

The steering software for online calibration provides not only updates of records but uses the described mechanism to distribute the latest version at run startup. Both the detector geometry and the parameterization of the magnetic field are read from the data base in that way.

## 4.9 Slow-Control Software

Operating more than 100 computers requires a sufficient reliability of the employed hardware to exclude frequent technical failures. An acquisition and monitoring of fan speeds, board temperatures and chip operating voltages is a mean to detect evolving problems early. Such a system was implemented for the reconstruction farm. It displays the hardware status of the farm together with the current load of all machines as an Internet page, which is updated once per minute.

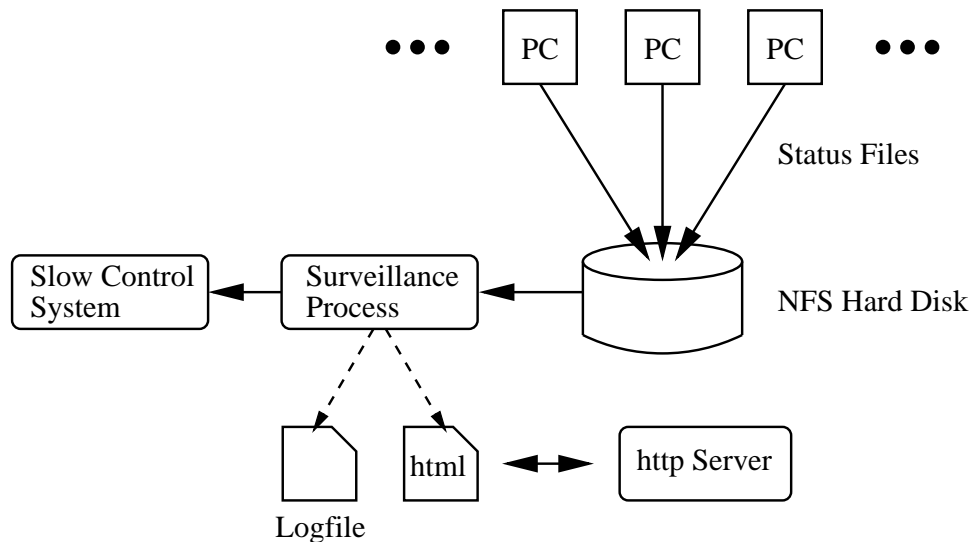


Figure 4.6: Sketch of the farm slow-control system. (Explanation in the text.)

Recent PC-motherboards are equipped with sensors for a measurement of the aforementioned quantities. A special software module was inserted into the Linux-kernel on each computer. The module creates a plain ASCII file with slow-control information on the local hard disk which is updated frequently. At booting time, a daemon-process is started on each machine. It reads the file once per minute and copies it together with information about the present load to a status file on a central **NFS-disk** which can be accessed by all farm PCs.

The internet-page is generated dynamically by a **Surveillance Process**. This process inspects all available status files once per minute and infers from the access-time which PCs are functioning. Deviations of measured quantities from allowed ranges are recorded in logfiles which can be visualised as history-plots using an Internet browser.

An additional logfile summaries at what points in time machines started or ceased to write their status files. This feature allows to detect otherwise unnoticed crashes of computers and a subsequent reboot. Warnings and serious failures are passed on to the overall **Slow-Control System** transmitting them to the shift crew in the control room.

The described system was found to operate stably for extended periods. It is also used for the monitoring of PCs having responsibilities in the data acquisition system and for the data base.

## 4.10 Farm Summary

The application of standard PCs and network technology proved to be an adequate solution for online reconstruction, data logging, and event reprocessing in the experiment. The usage of Linux as operating system and changes to the input-output routines of the HERA-*B* data analysis framework made it possible to blur the border between offline and online. Every reconstruction program that was statically linked on Linux can run on the farm without modifications. The design of the farm software employs a push architecture for the event data stream, while monitoring and calibration data are pulled from the farm machines to dedicated computers. This proceeding gives highest priority to the transmission of the event data. The concept of the control software allows to assign events to reconstruction farm CPUs with sufficient speed and reliability.

The farm has been included in the data path since inception. In the development phase of the data acquisition and trigger system, a farm of reduced size was sufficient to handle the recorded data. Up to 45 machines performed partial event reconstruction and data quality checks before event data was archived to tape. The designed event and data rates could be exceeded clearly. At event sizes of  $\mathcal{O}(10\text{ kB})$  logging rates of up to 7.5 MB/s were obtained.

The reconstruction farm has been upgraded to the planned number of about 200 processors when the installation of the detector hardware was completed. Since the beginning of the data-taking, online reconstruction is performed routinely. The data quality assessment on the farm provides valuable input for the selection of data sets which are scheduled for analysis. Data are archived in data summary tapes (DSTs) containing both raw data and the reconstructed information. For faster access, selected results of the reconstruction are stored in so-called MINIs, which are typically a factor of 10 smaller than DSTs and can be kept on disks.

The online calibration system ensures the distribution of calibration and alignment information to all processes in the data acquisition and trigger system. In connection with the data base, it keeps a record of the versions of calibration constants which were used for processing on the different trigger levels. The inclusion of all subdetectors into the online calibration system is still an ongoing process.

# Chapter 5

## Third Level Trigger Studies

At the time of the HERA-*B* Technical Design Report, the third level trigger was thought of as a complementary system to the SLT [13]. The TLT was to be executed after the event building, but its processing would still be restricted to the particles in the RoIs. Its anticipated main task was the full refit of the lepton and hadron tracks found and refined by the FLT and SLT. The trigger decisions would be based on the properties of these particles, like positions of common vertices or impact parameters with respect to the target wires. Already at this time, the usage of general-purpose processors for the TLT was envisaged. The TLT should execute a variant of the offline analysis algorithms with looser cuts to account for imperfect calibration. Its planned working environment was a joint third level and reconstruction processor farm. Besides complete event reconstruction at a rate of 50 Hz, the computing capacity of this farm should allow to spend about 10ms per event for the TLT, which was expected to handle an input rate of about 500 Hz.

Besides other reasons, the sketched design was a consequence of the proposed implementation of the SLT as a network of DSPs connected to specific detector regions. Tracks were to be established locally in a way similar to the FLT by passing RoIs from processor to processor in the direction towards the target. The combination of track properties into a trigger decision for an event was therefore shared between a global instance in the SLT and the TLT, where the latter was to examine the entire data belonging to a trigger track with improved accuracy.

The implementation of the second level trigger as a farm of commodity processors and the transition from a RoI-based to an event-based processing (cf. Section 3.7) changed the role of the TLT significantly. The SLT could take over much of the TLT's functionality and it became clear that the SLT processor dedicated to an event would perform the event building as well. It is then only reasonable to execute TLT algorithms whenever needed on the same processor before sending the accepted events to the reconstruction farm. Also in the final design of the trigger and data acquisition system, the TLT retains rapid local access to the entire data of an event. Its main characteristics in contrast to the SLT is now, however, the ability to include data outside of the RoIs in the decision. A suppression of about 10 should be reached within 100 ms per event. This means that at an estimated input rate of 500 Hz, about 50 of the 240 PC of the SLT/TLT farm would execute the TLT step at a time.

### 5.1 Strategy at the Third Level Trigger

The TLT is now being viewed as a system which is applied to event classes where the information available to the SLT might be insufficient to bring down the rate to a level acceptable for

input into the online reconstruction farm without impairing the efficiency for signal events. As discussed briefly in Section 3.5.3, the SLT alone provides enough suppression for events in which two trigger particles found in the RoIs form a secondary vertex. This is the case for the two most important physics channels in HERA-*B*, the  $J/\psi$  trigger for  $B \rightarrow J/\psi X$  and decays of  $B$  mesons into charged hadron pairs ( $B \rightarrow h^+ h^-$ ). The trigger is derived entirely from one  $B$  meson in the event, and the second  $B$  can be studied without any restriction on its decay channel. A good-quality vertex detached from one of the target wires will be required. In addition, the SLT can check whether the measured momenta of the trigger particles sum up to a vector pointing from the vertex to the nearby target wire. Since the position of the primary vertex on the target wire is not known yet, this cut can only be applied in the plane perpendicular to the wire. The main focus of the TLT will be on event classes with fewer kinematical constraints, i.e. without a vertex formed by two trigger particles. This is the case when the trigger particles come from different  $B$  hadrons in the event or from different stages of the decay chain of one  $B$ . Copious events of interest are those where both  $B$ s decay semileptonically

$$b\bar{b} \rightarrow B_1 B_2 \rightarrow l_1 X_1 l_2 X_2 \quad (5.1)$$

or two leptons from the decay chain of one  $B$  initiate a trigger

$$b\bar{b} \rightarrow B_1 B_2 \rightarrow l_1 X_1 B_2 \rightarrow l_1 l'_1 X'_1 B_2. \quad (5.2)$$

Such events are selected by the FLT if the invariant mass of two opposite-sign leptons of the same flavour happens to fall in the window around the  $J/\psi$  mass or the FLT allows for count triggers (cf. Section 3.4) in addition. In the latter case the cut on the invariant mass is replaced by the requirement of a minimal transverse momentum of about 1 GeV for the leptons. If count triggers are enabled, one can also select a combination of high- $p_T$  hadrons and leptons. This adds events in which one  $B$  decays semileptonically and the second  $B$  gives rise to a high- $p_T$  hadron

$$b\bar{b} \rightarrow B_1 B_2 \rightarrow l_1 X_1 h_2 X_2 \quad (5.3)$$

to (5.1) and (5.2). Presumably, the required  $p_T$  has then to be well above 1 GeV to achieve a sufficient suppression of hadrons. In either case, the SLT verifies that the two trigger particles come from the same target wire. The only handle available to reject pairs of background tracks is a cut on an appropriate combination of the impact parameter of the particles.

## 5.2 Silicon Third Level Trigger

The third level trigger has up to 100 ms at its disposal to find additional features discriminating between signal and background events. This time slot is big enough to allow for local pattern recognition in tracking and particle identification devices. The particle identification devices like RICH and ECAL have the potential disadvantage that their information alone is not sufficient in order to assign a particle to the interaction, which caused the trigger at the first and second level. One possible direct approach for triggering, which has already been pondered in the HERA-*B* Proposal [12], is the reconstruction of tracks in the silicon vertex detector. There are several ways how a partial or complete SVD reconstruction can contribute to accomplish the suppression of background events and the selection of signal events at a higher efficiency:

- \* **Reconstruction of secondary vertices in signal events:** A well-fitted vertex downstream of the target wire identified by the SLT points to an event with particles having

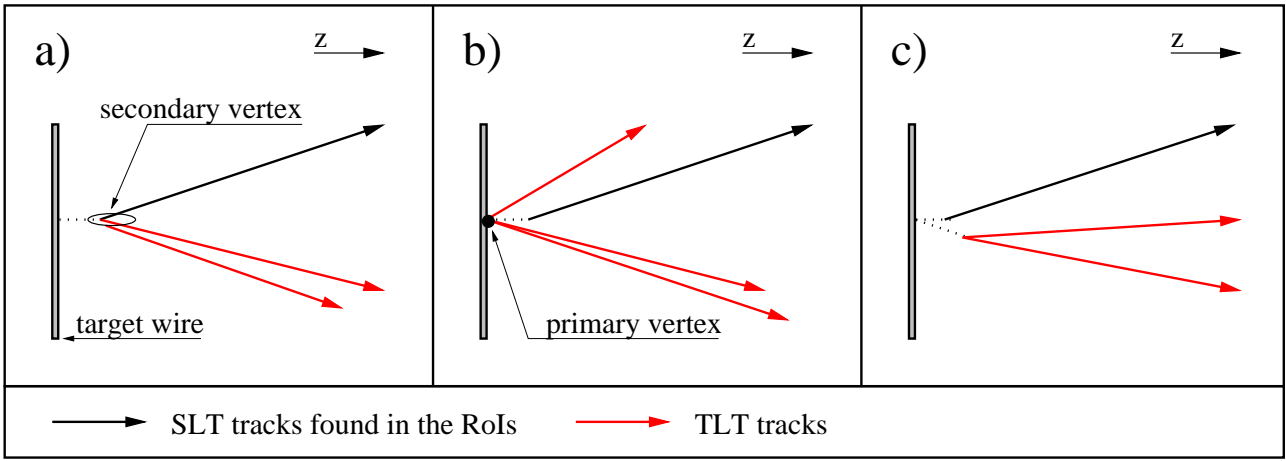


Figure 5.1: The TLT using the silicon vertex detector aims at the reconstruction of additional tracks outside of the RoIs. Track from the decay chain of  $B$  hadrons which are not selected by the preceding trigger levels form displaced secondary vertices with the SLT trigger particles (a) or feature non-vanishing impact parameters (c). The determination of the primary vertex (b) allows to apply impact parameter cuts not only on the distance between track and wire, but on the separation between the track and the position of the interaction.

an extended lifetime. Such a displaced vertex between any two particles is a powerful constraint. A vertex formed by a particle found by the SLT and a TLT track (Fig. 5.1 (a)) might recover some of the leptons, which would be rejected by the SLT, because their impact parameters to the target wire are small.

- \* **Reconstruction of the primary vertex:** After finding almost all tracks in the SVD, the position of the primary vertex on the target wire can be determined readily (Fig. 5.1 (b)). With this knowledge, the impact parameter cuts imposed by the SLT on the trigger particles can be refined. Cuts can be placed on the impact parameter with respect to the primary vertex, rather than on the distance to the target wire. Here, the strategy would be to use less stringent cuts at the SLT and to postpone the final decision to the TLT. This proceeding might help to keep a larger fraction of the signal events.
- \* **Reconstruction of tracks with large impact parameters:** Most of the tracks from the decay cascade of  $B$  hadrons are not found in the RoIs, but feature non-vanishing impact parameters as well. The detection of such tracks in the SVD is a strong indication for an event with heavy quarks (Fig. 5.1 (c)).

The feasibility of a TLT using only the vertex detector depends on the availability of a sufficiently fast pattern recognition program and the quality of the found tracks. Facing the computing power of the SLT/TLT processors and the scalability of the trigger farm, the first item should not pose a real problem. The crucial question is in fact whether the precision obtained with the SVD alone allows for strong conclusions, when separating signal from background events. The intrinsic difficulty one must cope with is the missing momentum information for the tracks found in the SVD outside of the RoIs. The missing momentum prohibits to estimate the multiple scattering in the SVD properly and to look for decays kinematically by forming the invariant masses of particles.

To study the problems listed above, a software environment for a third level trigger at HERA-*B* has been developed<sup>1</sup>. The software comprises a fast track reconstruction including primary vertex finding in the silicon vertex detector and code for conceivable TLT algorithms and their evaluation using recorded data and Monte Carlo events. In the following, we briefly describe the software, the principle of the track reconstruction, the primary vertex finding, and their respective performance. When reviewing the track reconstruction, we emphasise the treatment of multiple scattering in absence of momentum information, rather than the details of the algorithms, which can be found in [35]. We then present studies conducted along the items of the list at the beginning of this section.

### 5.3 Software Environment

The final operating environment of the TLT software is the Second Level Process (SLP) (cf. Section 3.7), a copy of which is executed on each PC of the SLT/TLT farm. For ease of algorithm development using simulated event data, the TLT software can be linked with both SLP and the HERA-*B* framework for event processing ARTE [36]. ARTE and SLP are distinct environments adapted to their specific tasks. ARTE provides event generation and houses the various packages used for offline event reconstruction and data analysis. Quantities like tracks, hits, and geometrical descriptions of detectors are stored in so-called ARTE-tables with ZEBRA<sup>2</sup> performing memory management. An ARTE environment built for Linux can run on the on-line reconstruction farm (Section 4.5), but is nonetheless rather offline-like. In contrast, the SLP emphasises features indispensable for running online, like low-latency communication with DAQ components for SLT software and event builder. Memory management and structured data storage are shaped by the widespread use of the C programming language. ARTE and the SLT software agree, however, on the way of storing important data. The format of the data structures describing for instance the detector geometry in the SLT framework is derived directly from the corresponding ARTE-tables.

A TLT dealing with the vertex detector exclusively needs comparatively few inputs, namely the geometrical arrangement of silicon counters and target wires, and the hits associated with an event. This eases the implementation of clean interfaces to SLP and ARTE. In the TLT software, compile options are used to switch to either ARTE or SLP. In ARTE, the appropriate tables are read while dedicated functions provided by the SLT group are employed to retrieve detector geometry and the SVD part of the event data within the SLP. This was exercised successfully in April 1999 when the TLT software reconstructed tracks online within an SLP environment [35].

The development of the TLT track reconstruction and the studies presented in the following were conducted entirely within ARTE. Files with Monte Carlo events were generated to implement, test, and benchmark TLT algorithms. Different event classes (signal and background) served to determine performance figures. Before turning to the track reconstruction, the event generation and the detector simulation are described.

---

<sup>1</sup>This work was carried out in cooperation with S. Scharein [35] who contributed the central ideas of the pattern recognition algorithm.

<sup>2</sup>ZEBRA is a memory management package provided by CERN.

### 5.3.1 Event Generation

ARTE provides a simple interface to event generation of various processes relevant for HERA-*B*'s physics program [37]. The generation of events with heavy flavours is accomplished by two well-known packages, PYTHIA [38] and FRITIOF [39]. FRITIOF models proton-nucleon collisions with hard parton scattering and gluon emission, but creates light quarks only. PYTHIA supports heavy flavour production, but does not allow for nuclei as targets. The two packages are therefore combined in a consistent way to create proton-nucleus collisions.

In the event generation, PYTHIA is called with a 920 GeV proton and a nucleon, the type of which is selected randomly according to the ratio of protons and neutrons in the target material. Among the particles generated in the hadronization, all those not associated with heavy quarks are removed. Conservation laws allow to infer the amount of energy the proton has lost due to the creation of heavy quarks. A proton with reduced energy and a target nucleus are then handed over to FRITIOF. FRITIOF simulates the underlying event emerging from the collision, taking into account effects like energy loss in the nucleus and multiple scattering of partons.

### 5.3.2 Detector Simulation

The simulated HERA-*B* detector corresponds to the running configuration in 1999/2000. Four titanium wires were deployed at the first target station closer to the SVD (cf. Fig. 2.3 on Page 19), while the second station was equipped with various target materials to study the dependence of physical quantities on the mass number  $A$ . In the simulation, all wires of station I were assumed to contribute to the interaction rate. Such a configuration is well-suited for running at high interaction rates, since it simplifies the assignment of reconstructed tracks to wires.

The operated vertex detector differed from the final design shown in Table 2.1 on Page 21. In superlayer 1–3, the first (further upstream) module was single-sided in all sectors. All other modules were double-sided. With this arrangement (cf. Table 5.1), a particle can cause up to 3 (4) hits per superlayer in each of the superlayers 1–3 (4–7). The interactions generated in

Table 5.1: SVD Configuration in 1999/2000. The table lists the four layers in the direction of increasing  $z$ . The letter  $x$  or  $y$  indicates to which axis the strips are almost parallel, the sign specifies whether the stereo angle is  $+2.5^\circ$  or  $-2.5^\circ$  with respect to this axis. The missing strip layers in superlayer 1–3 appear as blanks.

Superlayer	1–3		4–7	
Sector 1	$-y$	$-x + y$	$-y + x$	$-x + y$
Sector 2	$+x$	$+y - x$	$+x - y$	$+y - x$
Sector 3	$-y$	$-x + y$	$-y + x$	$-x + y$
Sector 4	$+x$	$+y - x$	$+x - y$	$+y - x$

ARTE are spread over the active target wires. The distribution of primary vertices is derived from the horizontal and vertical profile of the proton beam. The propagation of tracks through the detector is taken over by the GEANT simulation program. GEANT traces the particles and their decay products through the various detector volumes, and calculates energy loss and multiple scattering. The impact points of tracks  $(x, y)$  impinging on the vertex detector modules are the starting points for the digitization. The relation between  $(x, y)$  and the projected

coordinate  $m$  measured by a layer of silicon strips with a stereo angle  $\alpha_{\text{st}}$  is given by (cf. Fig. 2.4 (a) on Page 21):

$$m = x \cos \alpha_{\text{st}} - y \sin \alpha_{\text{st}}. \quad (5.4)$$

The digitization transforms these impact points into measured coordinates  $m_i$  for all crossed layers  $i$ . In doing so, it uses macroscopic properties of the silicon counters derived from measurements and simulations (Table 5.2). The intrinsic hit resolution  $\sigma_0$  specifies the error of the measured projected coordinate  $m$ . It is smaller than the resolution given by the  $50 \mu\text{m}$  strip pitch due to the analogue readout. A value of  $12 \mu\text{m}$  [40] is used to smear out the coordinates. A hit efficiency  $\varepsilon_{\text{hit}}$  of 98 % is used to determine whether an impact point gives rise to a measurement or not. Particles passing through a detector plane at almost the same position

Table 5.2: Parameters of the SVD Simulation.

parameter	value
hit resolution $\sigma_0$	$12 \mu\text{m}$
double-hit resolution $\Delta h$	$200 \mu\text{m}$
hit efficiency $\varepsilon_{\text{hit}}$	98 %

create a charge deposition, which cannot be resolved into several hits. In the simulation, hits are therefore merged if their spatial separation is smaller than the double-hit resolution  $\Delta h$  of  $200 \mu\text{m}$ .

## 5.4 Fast Track and Primary Vertex Reconstruction in the SVD

In a region without matter and deflecting fields particle trajectories are straight lines. Therefore, in the vertex detector, deviations from a linear propagation of tracks originate only from multiple scattering in the detector material. Particles experience multiple scattering exclusively in the SVD superlayers, while the vacuum between them is a good approximation to empty space. At least locally, tracks can be described as straight lines. The angles with respect to the  $z$ -axis of tracks in the geometrical acceptance of the SVD are smaller than  $250 \text{ mrad}$ . Tracks can be parameterised as a function of  $z$ :

$$\begin{pmatrix} x(z) \\ y(z) \end{pmatrix} = \begin{pmatrix} x(z_{\text{ref}}) \\ y(z_{\text{ref}}) \end{pmatrix} + (z - z_{\text{ref}}) \begin{pmatrix} t_x(z_{\text{ref}}) \\ t_y(z_{\text{ref}}) \end{pmatrix}. \quad (5.5)$$

Here  $x(z)$  and  $y(z)$  are the local transverse coordinates at a position  $z$ , and  $x(z_{\text{ref}})$  and  $y(z_{\text{ref}})$  their respective values at an arbitrary reference position  $z_{\text{ref}}$ . The slopes in  $x$  and  $y$ ,  $t_x(z_{\text{ref}})$  and  $t_y(z_{\text{ref}})$ , specify the direction of tangents along the particle trajectory at  $z_{\text{ref}}$ . They are given in terms of momentum components  $p_{x,y}$  and angles  $\theta_{x,y}$  with respect to the  $z$ -axis by:

$$t_x(z) = \frac{\partial x(z)}{\partial z} = \frac{p_x(z)}{p_z(z)} = \tan \theta_x(z) \quad (5.6)$$

$$t_y(z) = \frac{\partial y(z)}{\partial z} = \frac{p_y(z)}{p_z(z)} = \tan \theta_y(z). \quad (5.7)$$

Lacking momentum information, the vector of track parameters  $\vec{t}$  is four-dimensional:

$$\vec{t}(z_{\text{ref}}) = \begin{pmatrix} x(z_{\text{ref}}) \\ y(z_{\text{ref}}) \\ t_x(z_{\text{ref}}) \\ t_y(z_{\text{ref}}) \end{pmatrix}.$$

For later reference we write down how track parameters are transported from a plane  $z$  to a position  $z'$  and the relation between a parameter vector and a measurement. Both operations are linear and can be expressed conveniently as matrix multiplication. The transport of track parameters is accomplished by:

$$\vec{t}(z') = \mathbf{F}(z, z') \vec{t}(z) = \begin{pmatrix} 1 & 0 & z' - z & 0 \\ 0 & 1 & 0 & z' - z \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \vec{t}(z). \quad (5.8)$$

In the following, we assume that the vertex detector modules are parallel to the  $x$ - $y$  plane. A layer of silicon strips at a position  $z_i$  with a stereo angle  $\alpha_i$  measures according to Eq. (5.4) a projected coordinate  $m_i$ . The relation between  $m_i$  and the vector of track parameters at  $z_{\text{ref}}$  is given by ( $\Delta z = z_i - z_{\text{ref}}$ )

$$m_i = x(z_i) \cos \alpha_i - y(z_i) \sin \alpha_i = (x(z_{\text{ref}}) + \Delta z t_x) \cos \alpha_i - (y(z_{\text{ref}}) + \Delta z t_y) \sin \alpha_i. \quad (5.9)$$

It can be re-expressed using a matrix  $\mathbf{H}_i(z_i, z)$  as

$$m_i = \mathbf{H}_i(z_i, z) \vec{t}(z) = \begin{pmatrix} + \cos \alpha_i \\ - \sin \alpha_i \\ + \Delta z \cos \alpha_i \\ - \Delta z \sin \alpha_i \end{pmatrix} \vec{t}(z). \quad (5.10)$$

At least four measured hits  $m_i$  in different projections are required to define a track candidate. The task of a track finding algorithm is to form sets of hits which were caused by the same particle. For trigger applications it should be sufficiently fast to meet the timing constraints imposed by the number and computing capacity of the SLT/TLT processors. A second requirement is the need for a self-contained method exploiting only information from the SVD. In the offline reconstruction, pattern recognition of tracks is tried separately in the SVD and in the pattern tracker behind the magnet. When connecting the established candidates, one can discard tracks that are found in one system exclusively. This proceeding allows to reject tracks candidates that comprise hits from different particles, so-called ghost tracks. Without this measure, one must aim for algorithms delivering a low number of ghost tracks.

When TLT algorithms were investigated for the first time, there was one program for SVD reconstruction available (SUSi<sup>3</sup>, [40]). It formed track candidates from almost all possible sets of four SVD hits in different superlayers. These initial candidates were propagated through the detector looking for confirming hits. The SUSi algorithm reached a good track finding efficiency of 91.6 % with 8.7 % ghost tracks [40], but resulted in unacceptable execution times. When rescaled to the processors of the SLT/TLT farm, 11 s were needed per event [35], which is clearly not acceptable for trigger purposes.

---

<sup>3</sup>Searching Utility for the Silicon.

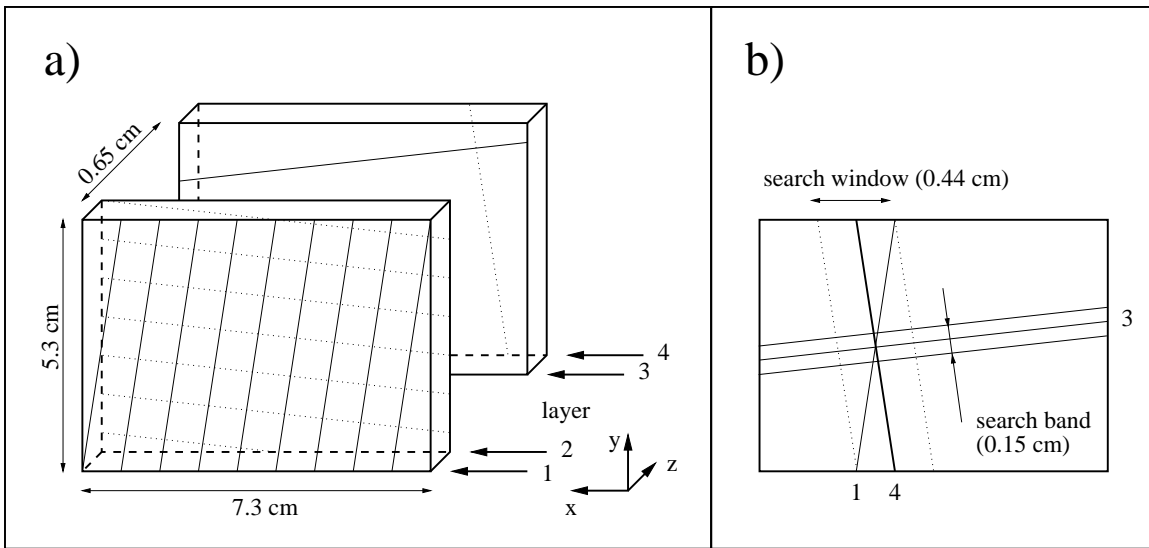


Figure 5.2: **(a)** A sector in a SVD superlayer is equipped with two double-sided silicon detectors separated by about 0.65 cm. A passing particle is measured in four layers (1–4) with different stereo angles. There are two pairs of layers with almost vertical (1+4) and horizontal (2+3) strips. Within such a pair, the stereo angles differ by only  $5^\circ$ . Layer 4 is called the *associated stereo layer* of layer 1, while 2 and 3 are referred to as *rotated layers*. **(b)** A fired strip in a layer (1) defines a comparatively narrow search window in the associated stereo layer (4). Hits caused by the same particle are expected in this window. In the spacepoint algorithm, initial spacepoints are created from two intersecting strips (1+4). The rotated views (layer 2+3, only layer 3 is indicated) are then searched for confirming hits in a narrow band.

The speed of a tracking algorithm strongly depends on the number of processed track candidates. In the SUSi algorithm, the number of track candidates grows with the fourth power of the number of SVD hits

$$N_{\text{cand}} \propto N_{\text{hits}}^4.$$

The about  $10^5$  hypotheses to be considered per event outnumber the real tracks by far since the latter scale only linearly with  $N_{\text{hits}}$ . An attempt to devise a faster algorithm was made.

### 5.4.1 Concept of the Track Reconstruction

The close arrangement of two double-sided silicon modules in one single superlayer is the prerequisite for the construction of faster algorithms. One can exploit efficiently that a passing particle hits up to four strips at almost the same  $z$ -position (Fig. 5.2 (a)). The two double-sided detectors are separated by about 0.65 cm. The four strip layers feature all different stereo angles. There are two pairs of layers with almost vertical and horizontal strips. Within a pair, the two stereo angles differ by only  $5^\circ$ . Such strips are called *stereo layers* and the strips in the other pair are referred to as *rotated layers*. Two arbitrary hits in two stereo layers can only be caused by the same particle, when the fired strips cross in a projection on the  $x$ - $y$  plane. This suggests an algorithm that tries to identify the hits in a superlayer, which could be due to the same particle. The found hits can be combined into a three-dimensional spacepoint. Tracks are then defined in terms of spacepoints lying approximately on a straight line. Track candidates can be derived from any two spacepoints. The number of track candidates grows thus only

quadratically with the number of spacepoints

$$N_{\text{cand}} \propto N_{\text{spacepoints}}^2$$

and should allow for a considerable gain in execution speed.

The spacepoint approach is the basic new property of the TLT track finding. The implemented track finding has four distinct phases:

1. **Spacepoint reconstruction:** The SVD reconstruction starts with the creation of three-dimensional points from the SVD hits.
2. **Track reconstruction:** Track candidates are then identified by finding those spacepoints which might be related to the same particle.
3. **Kalman refit:** For each track candidate, the best estimate of track parameters is obtained by applying a Kalman refit to the hits contained in the spacepoints.
4. **Compatibility analysis:** In a final step, quality criteria are imposed to select a set of compatible tracks, which can then be used for making trigger decisions.

### 5.4.2 Spacepoint Reconstruction

The spacepoint reconstruction is applied sequentially to all superlayers and sectors. A fired strip in a module (Fig. 5.2 (b)) defines a search window in the associated stereo layer. The extension of the search window is dependent on the strip orientation. In the example in Fig. 5.2 (b), the overlap region is 0.44 cm wide, i.e. much smaller than the module size. Only about 90 of the more than 1000 readout strips in the stereo layer must be checked for hits.

The algorithm starts with two stereo layers. It calculates the position of the search window for each fired strip in one layer and extracts hits in the associated stereo layer. A spacepoint candidate is formed with each stereo hit. Its  $z$ -coordinate is chosen as the mean  $z$ -position of the involved hits. The exact spacepoint position in  $x$  and  $y$  depends on the unknown angle between silicon wafer and particle trajectory. An estimate for this angle is obtained under the assumption that the track originates from the target region. An average target position is derived from all active wires. The same approximation is used to look for confirming hits in the two rotated layers. The straight line made up of the two-hit spacepoint and the mean target position is extrapolated to the  $z$ -position of the two rotated layers. Each fired strip in a search band around the predicted position is associated to the spacepoint. The width of this search band accounts for the uncertainty in the track angle estimate. It is an important parameter of the algorithm. If chosen too small, not all real spacepoints are found, while a too large value will give rise to many false combinations. An optimal value for each superlayer and sector was derived from a Monte Carlo study. The found values drop from 0.44 cm at the first superlayer to 0.11 cm at superlayer 7 where the approximation of a point-like target is best.

If more than one matching hit is found in each of the two rotated layers, all possible four-hit spacepoints are constructed. If the  $\chi^2$  of a fit calculating the spacepoint position exceeds a certain value ( $\chi^2 > 75$ ), it is assumed that one of the two hits in the rotated layers does not belong to the spacepoint. In this case, two spacepoints with three hits are formed instead. These two spacepoints comprise the two hits from the initial stereo layers and one hit from the first and second rotated layer, respectively. Only the spacepoint with the better  $\chi^2$  is kept.

The algorithm described above is now repeated with the other two stereo layers as starting point for combination of two fired strips. This is mandatory when one of the stereo hits is

missing due to detector inefficiencies. In the normal case, this proceeding has the side effect of reconstructing some spacepoints more than once. The last step is therefore to remove identical copies of spacepoints containing exactly the same hits as another candidate. Points, whose hits are just a subset of another spacepoint, are sorted out as well.

The implementation of the algorithm was strongly optimised. Fitting matrices prepared beforehand are used whenever spacepoint positions are calculated from hits. The efficiency for finding spacepoints is greater than 99 % in all superlayers when the SVD simulation uses the parameters from Table 5.2. Clearly, not all ambiguities in the assignment of hits to passing particles can be resolved. The properly found spacepoints are accompanied by a considerable number of "ghost points". This is most pronounced in the first superlayers because of the comparatively poor approximation of a point-like target. In the extreme case – the most upstream sectors of superlayer 1 – only 30 % of all reconstructed spacepoints combine hits caused by the same particle. This percentage is 70 % in superlayer 7, its value averaged over all superlayers is about 50 %.

### 5.4.3 Track Reconstruction by Track Following

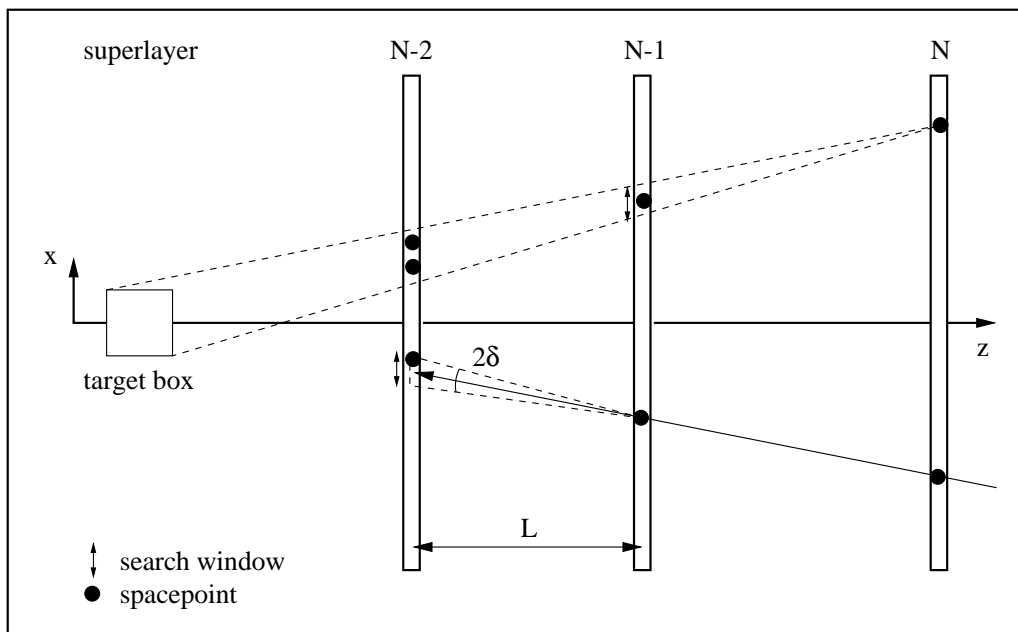


Figure 5.3: For pattern recognition of tracks, each spacepoint in superlayer  $N$  ( $N = 7 \dots 3$ ) is regarded as a seed (upper half of plot). A projective search window is calculated at the  $z$ -position of superlayer  $N-1$  by assuming that all particle trajectories originate from a box around the target region. In case more than one spacepoint is found in the window, the track candidates are propagated concurrently. Once a track has at least two spacepoints (lower half of plot), a straight-line fit to their coordinates predicts a spacepoint position in the next superlayer upstream ( $N-2$ ). A search window is put up in both  $x$  and  $y$  around the prediction. It is determined by an angle  $\delta$  around the predicted direction to account for deviations from linearity due to multiple scattering and the approximate nature of spacepoint positions. The boundaries of the search window are calculated from  $\delta$  and the extrapolation distance  $L$ .

An exhaustive method for deriving tracks from spacepoints was already sketched in Sec-

tion 5.4.1. A full initial track parameter vector can be obtained by combining any two spacepoints from different superlayers into a line. Matching spacepoints can be picked up by extrapolation to the remaining superlayers. This approach has the potential benefit that tracks with an arbitrary angle to the beam (including tracks from decays behind the first superlayers) can be found. The number of track candidates grows, however, quadratically with the number of spacepoints.

The track finding algorithm implemented in the TLT software is a faster local method. It considers every spacepoint as a track seed and assumes that the particle originates from the target region. Such a method is not very well suited for finding the products of late-decaying particles. This shortcoming is well acceptable for a trigger algorithm aiming for daughter particles of hadrons containing heavy quarks. The effort of this method is only linear in the number of spacepoints and it provides effective handles to limit the number of followed track candidates.

The pattern recognition method is sketched in Fig. 5.3. A valid track is required to have at least three spacepoints. Starting in superlayer 7 and ending in superlayer 3, the tracking algorithm turns every spacepoint into a track seed. The target region is modelled as a box around the wires. When beginning in superlayer  $N$ , the spatial extension of the target box and projective geometry are used to calculate a search window for spacepoints at superlayer  $N-1$ . For each spacepoint in the window a new track is spawned. Each branch is propagated separately towards the target. When a track candidate has acquired at least two spacepoints, their coordinates are used to predict the likely position of a spacepoint in the next superlayer upstream. A search window around this position is scanned for points. The window dimensions must account for the extrapolation error. This error combines the uncertainty in the determination of spacepoint coordinates and the influence of multiple scattering occurring between the measurements. In the algorithm, the extrapolation error is introduced as an adjustable parameter  $\delta$ . It specifies the angular extension of the search window around the predicted position when viewed from the last spacepoint added. Width and height of the search window are calculated from  $\delta$  and the extrapolation distance.

The propagation of a track stops when either superlayer 1 is reached or no spacepoint is found in the last search window. The latter condition occurs automatically when the search window falls outside the sensitive areas in a superlayer. Tracks with three or more spacepoints are stored for further evaluation.

The implemented tracking algorithm has several built-in handles to find a good compromise between the efficiency for finding tracks and execution speed. Given that the latter depends on the number of track candidates followed, one can adapt the following parameters accordingly:

1. **The size of the target box.** In the present TLT software, a single box surrounding all active wires is used to find all tracks in the SVD. For real triggering applications the wire, which caused a possibly interesting interaction, is already identified at the SLT. It is suggestive to use as many target boxes as active wires and to switch between them on an event-to-event basis. In this way, tracks coming from one of the remaining wires can be excluded from the reconstruction. This measure should speed up the tracking part by a factor equal to the number of target wires operated.
2. **The search angle  $\delta$ .** A comparatively low value will exclude low-momentum tracks suffering from considerable multiple scattering at an early stage. It should reduce the number of track branches initiated by spacepoints which combine hits caused by different particles.

3. **The number of spacepoints considered in a search window.** In the tracking, the creation of track branches can be restricted to those spacepoints in a search window which match the prediction best. This might be especially useful when a noisy detector region gives rise to many spacepoints at the same position. This cut can depend on the stage of the track propagation. One might consider all combinations when looking for the second spacepoint using the target box, but decide to select only the best match in all later steps.
4. **The treatment of search windows without any found spacepoint.** Stopping the propagation for candidates with "missing" spacepoints limits the number of track branches at an early stage. On the other hand, to allow for one superlayer without spacepoint might recover some tracks in case of considerable detector inefficiencies. The benefit should be small, however, when a track is required to have at least three spacepoints since only few particles hit more than three superlayers.

In this study the parameters<sup>4</sup> were adapted for track reconstruction in the entire vertex detector. It was avoided to reject tracks early to gain a thorough understanding of relations between track quality and momentum. The angle  $\delta$  was adjusted to 2 mrad based on Monte Carlo studies. This value should be compared with the maximal possible deflection due to multiple scattering, which a particle can experience when traversing a superlayer. The Gaussian width  $\sigma$  of this distribution for a 1 GeV-particle in 280  $\mu\text{m}$  of silicon is about 0.7 mrad. The selected value for  $\delta$  constitutes therefore a  $3\sigma$ -interval. This particular setting for  $\delta$  translates into maximal and minimal search window dimensions of about  $0.2 \times 0.2 \text{ cm}^2$  and  $0.02 \times 0.02 \text{ cm}^2$ . They occur when extrapolating from superlayer 6 to 5 and from superlayer 2 to 1, respectively. Up to six spacepoints matching the centre of the search window best can give rise to new track branches. This limitation of the spacepoints per search window rejects only a small fraction of points. The propagation of track candidates was stopped when no spacepoint was found in the next superlayer. This criteria excludes tracks without hits in consecutive superlayers from the outset. About 10% of all tracks are affected by this cut.

#### 5.4.4 Track Refit

In the track propagation from superlayer to superlayer, the spacepoint coordinates provide a first estimate of the track parameters which are then used to calculate the centre of the search window. The track parameter vector  $\vec{t}$  of candidates with at least three spacepoints is determined with improved accuracy by a track refit. The refit is not applied to the spacepoint coordinates but directly to the contained hits. It is a Kalman filter with a twofold aim. Besides the  $\chi^2$  as quality criteria for track selection in the compatibility analysis, it supplies the best estimate of track parameters in the target region and their covariance matrix which are indispensable for the determination of impact parameters.

#### Kalman Filter

The Kalman filter method allows to obtain best estimates for the track parameter vector at any position along the trajectory in presence of multiple scattering. It is used widely for both pattern recognition and fitting of tracks in HERA-B [41]. In contrast to a global track fit it proceeds from hit to hit and refines the vector of track parameters whenever a new measurement

---

<sup>4</sup>A detailed listing of the used parameter settings including the dimensions of the target box can be found in the appendix.

is added. In the Kalman filter, the track parameters known so far are transported to the position of the next measurement. If they are known at  $z_{i-1}$  and the next measurement is available at  $z_i$  the track parameter vector and its covariance matrix  $\mathbf{C}(z_i)$  are given by (cf. Eq. (5.8) on Page 68):

$$\hat{\vec{t}}(z_i) = \mathbf{F}(z_i, z_{i-1}) \vec{t}(z_{i-1}) \quad (5.11)$$

$$\mathbf{C}(z_i) = \mathbf{F}(z_i, z_{i-1}) \mathbf{C}(z_{i-1}) \mathbf{F}^T(z_i, z_{i-1}). \quad (5.12)$$

The basic idea of the Kalman method is to regard this prediction  $\hat{\vec{t}}$  as an additional measurement. At  $z_i$ , that particular parameter vector  $\vec{t}(z_i)$  is calculated, which minimises the least-squares distance to both  $\hat{\vec{t}}$  and the available measurement  $m_i$  simultaneously ( $\mathbf{H}_i$  has been defined in Eq. (5.10)):

$$(\vec{t} - \hat{\vec{t}})^T \mathbf{C}'^{-1} (\vec{t} - \hat{\vec{t}}) + (\mathbf{H}_i \vec{t} - m_i)^T \mathbf{V}^{-1} (\mathbf{H}_i \vec{t} - m_i) \longrightarrow \min. \quad (5.13)$$

The appropriate weighting is ensured by the covariance matrices of the measured coordinate  $m_i$  ( $\mathbf{V}$ ) and of  $\hat{\vec{t}}$ . If for instance the track parameter vector is only known with low precision, i.e. all elements of its covariance matrix are large, the result will be dominated by the measured coordinate. The influence of multiple scattering is taken into account by the covariance matrix of the predicted track parameters. The elements of the matrix are increased according to the scattering the fitted track might have experienced. This is the reason for replacing  $\mathbf{C}$  by  $\mathbf{C}'$  in Eq. (5.13). The latter is given by

$$\mathbf{C}' = \mathbf{C} + \mathbf{Q}_{\text{MS}}, \quad (5.14)$$

where the matrix  $\mathbf{Q}_{\text{MS}}$  describes the multiple scattering.

It is obvious from the description above that the track parameters found at a certain point include only the knowledge from all preceding hits. As a consequence, the track parameters are optimal only at the last measurement in line. The ability to identify hits, which were associated wrongly with a track, is low in the first steps when the track parameter vector is not known yet with good precision. The best estimate at any point can be obtained by reversing the Kalman filter and transporting back the information from the last hits. This so-called Kalman smoothing [42] can serve well to repair pattern recognition errors by removing outliers.

For most applications one is interested in the track parameters in the target region. In the TLT software, the Kalman filter proceeds therefore from hit to hit towards the target to reach its final accuracy there. In the usual setting, the smoothing step is not executed. The outlier removal using the Kalman smoothing was found not to improve the track parameters in the target region significantly.

The Kalman filter is in general rather time consuming. Therefore, the implementation in the TLT software had to be optimised. The optimisation exploits the structure of the involved matrices with many elements which are zero<sup>5</sup>.

## Treatment of Multiple Scattering

The matrix  $\mathbf{Q}_{\text{MS}}$  specifies in how far the track parameters are washed out by multiple scattering between two measurements. A particle traversing a medium with a thickness  $x$  and a radiation

---

<sup>5</sup>A very good and detailed description of the formulas is available in Ref. [43].

length  $x_0$  is deflected stochastically by angles  $\theta$  due to multiple scattering. The distribution of deflection angles in a plane is approximately Gaussian for small angles with a width

$$\theta_{\text{MS}}(x/x_0) = \frac{13.6 \text{ MeV}}{\beta c p} \hat{z} \sqrt{x/x_0} [1 + 0.038 \ln(x/x_0)]. \quad (5.15)$$

Here  $p$ ,  $\beta$  and  $\hat{z}$  stand for the particle's momentum, velocity and charge, respectively. The above formula can be used to calculate how much the error of the track slopes  $t_x$  and  $t_y$  is increased after passing a scatterer. Most detector planes and layers of non-sensitive material are perpendicular to the  $z$ -axis. If their thickness along  $z$  is  $x$ , one must account for the increased path length due to the slopes by setting

$$x' = x \sqrt{1 + t_x^2 + t_y^2}.$$

The complete matrix  $\mathbf{Q}_{\text{MS}}$  assumes the following form [41]:

$$\mathbf{Q}_{\text{MS}} = (1 + t_x^2 + t_y^2) \theta_{\text{MS}}^2 (x'/x_0) \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 + t_x^2 & -t_x t_y \\ 0 & 0 & -t_x t_y & 1 + t_y^2 \end{pmatrix}. \quad (5.16)$$

The proper treatment of multiple scattering in the Kalman filter requires the knowledge of  $\mathbf{Q}_{\text{MS}}$  for any volume of silicon detectors and additional supporting material traversed by particles. In the TLT this is impractical for two reasons. First, the momentum of particles is not known. A second complication is of more technical nature. The detector has many non-sensitive parts and the time effort for checking whether a track hits a particular volume is rather high. Though the required information is included in ARTE, it is not easily available within the SLP. The

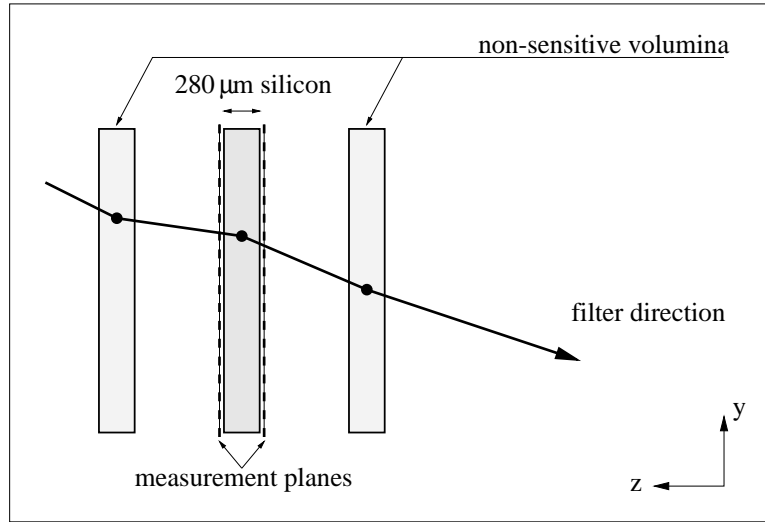


Figure 5.4: The treatment of multiple scattering in the Kalman refit is based on the assumption that the deflection occurs suddenly at a point in the middle of each scatterer. The silicon strip detectors are approximated as a scatterer between two measurement planes. The various non-sensitive layers between two measurements are combined into one effective scatterer with appropriate thickness and radiation length.

TLT uses therefore approximations for both the multiple scattering occurring between two measurements and the momentum. The amount of multiple scattering was estimated under the assumption that the deflection occurs in the middle of scatterer (Fig. 5.4). A silicon module is considered as a  $280\ \mu\text{m}$ -thick wafer sandwiched between two "virtual" strip layers. All non-sensitive volumes between two hits are merged into one single scatterer with an effective  $x/x_0$ . If  $M$  such planes with a thickness  $x^{(i)}$  and radiation length  $x_0^{(i)}$  are traversed, the effective value is given by

$$\frac{x}{x_0}\Big|_{\text{effective}} = \sum_{i=1}^M \frac{x^{(i)}}{x_0^{(i)}}.$$

The effective values were extracted from the simulation by following Monte Carlo tracks through all detector volumes. The amount of multiple scattering between two measurements depends of course on the exact positions where a particles hits a silicon counter. All particles have to cross the aluminum caps separating primary and secondary vacuum and the silicon wafers (cf. Table 5.3 [44] and Fig. 2.8 on Page 25). Additional contributions result from the readout cables and the detector carriers. Especially the design of the latter depends strongly on the superlayer in question. The material distribution of the fork-shaped carriers varies with the position on a detector module. This could be accounted for by storing the average effective  $x/x_0$  as a binned function of a track's impact point on the detector surface. It turned out, however, that a sufficient precision can be reached by just introducing dependencies on the superlayer and the layer number. (The sector number is included to be in accord with the geometrical description.) The effective  $x/x_0$  is averaged over the surface of a layer of silicon strips according to the radial distribution of reference tracks. This simplified procedure results in  $7(\text{superlayers})\cdot 4(\text{sectors})\cdot 4(\text{layers})=112$  numbers to store for fast evaluation of multiple scattering.

Table 5.3: Compilation of materials in the sensitive area of SVD counters.

detector part	material	$x_0$	thickness $x$	$x/x_0$
cap	aluminum	8.9 cm	$150\ \mu\text{m}$	0.17 %
wafer	silicon	9.36 cm	$280\ \mu\text{m}$	0.30 %
detector carrier	carbon fibre	19–26 cm	$600\text{--}1100\ \mu\text{m}$	0.23–0.58 %
flex-jumper	capton	28.6 cm	$420\ \mu\text{m}$	0.15 %
	copper	1.46 cm	$8\ \mu\text{m}$	0.06 %

The resulting average radiation lengths for superlayer 1 are shown in Fig. 5.5 (left plot). In each of the four sectors there are two silicon modules. The module at lower  $z$  is single-sided, the other double-sided. To follow a common scheme for all superlayers the single-sided modules are treated as double-sided modules, which have one detector layer with zero efficiency. From this point of view there are four layers of silicon strips in each of the four sectors of a superlayer. The plot shows the extracted average effective  $x/x_0$  between the last possible measurement at larger  $z$  and a hit in a detector plane for all planes in superlayer 1. For layer 1 and 3 the traversed material is always just the silicon wafer with 0.3 % of a radiation length. The values of  $x/x_0$  for the layers 2 and 4 describe the distribution of material between the two modules of a superlayer and between superlayers. They show some spread due to the inhomogeneous distribution of material over a the sensitive detector area of which the average is taken. As pointed out above, only the mean value is used.

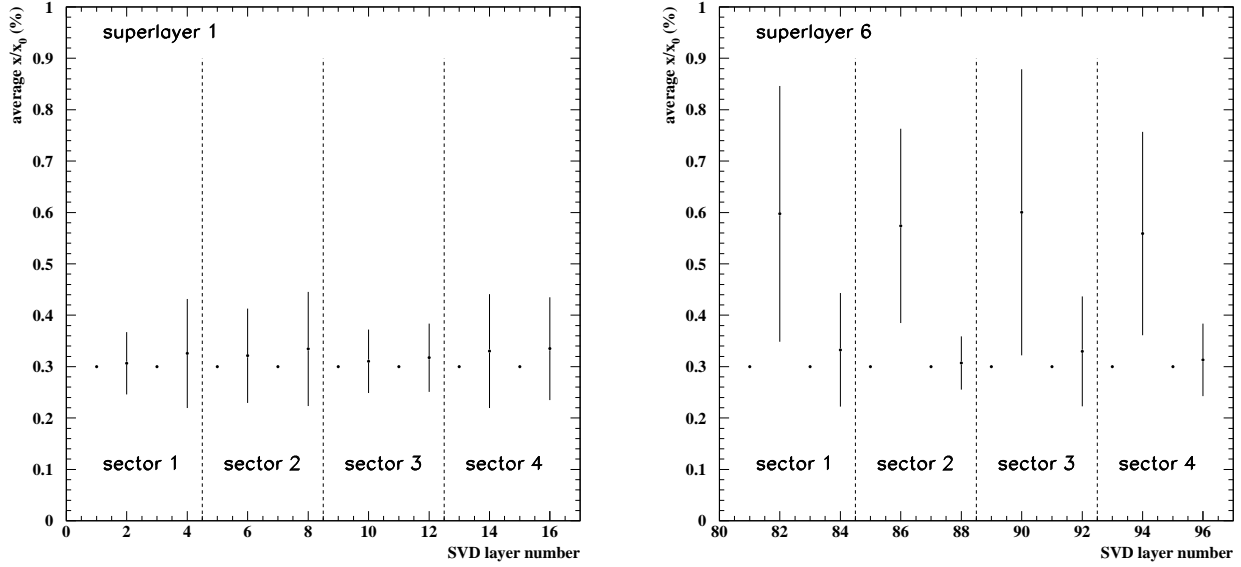


Figure 5.5: Before processing the next measurement at  $z_i$  in the Kalman filter, the error of the track parameters is increased according to the amount of material the particle traversed between  $z_{i-1}$  and  $z_i$ . The average thickness of material in units of the radiation length  $x/x_0$  is shown for all 16 detector layers in superlayer 1 (left). The mean value of  $x/x_0$  for a layer  $n$  serves to update the covariance matrix of the track parameters *before* a possible hit in this layer is processed. The mean values were extracted from Monte Carlo events by following tracks through all detector volumes. The error bars indicate the RMS of the distribution rather than the uncertainty of the mean. The amount of material depends on the construction of the SVD modules. This is demonstrated on the right for superlayer 6 which is "thicker" in terms of  $x/x_0$ .

The second ingredient needed for the estimation of  $\mathbf{Q}_{MS}$  besides  $x/x_0$  is the momentum of particles. The track slopes in  $x$  and  $y$  are the only available information connected with the momentum of a particles. Tracks under small angles with respect to the  $z$ -axis have on average higher momenta than particles further away from the beam. The momentum of particles is estimated using an average transverse momentum  $\langle p_T \rangle$  and the known slopes as:

$$p_{\text{estimated}} = \langle p_T \rangle \sqrt{1 + \frac{1}{t_x^2 + t_y^2}}. \quad (5.17)$$

The proper inclusion of multiple scattering is especially important for low-energy tracks due to the  $1/p$ -dependence of the scattering angle in Eq. (5.15). For the amount of material in the SVD, particle trajectories show severe deviations from a straight line for momenta below about 10 GeV. This can be seen by inspecting the  $\chi^2$  per degree of freedom of a straight-line fit to the hits of reference tracks. Tracks with momenta below 10 GeV have a  $\langle p_T \rangle$  of about 0.4 GeV. When adopting this value the real momentum of particles can be estimated to an accuracy of about 4 GeV (Fig. 5.6) for low energies. The particular choice of 0.4 GeV for  $\langle p_T \rangle$  gives too small values for the momentum of high-energy particles. This can be accepted since the resulting overestimation of multiple scattering does not deteriorate the calculated track parameter vector  $\vec{t}$  significantly.

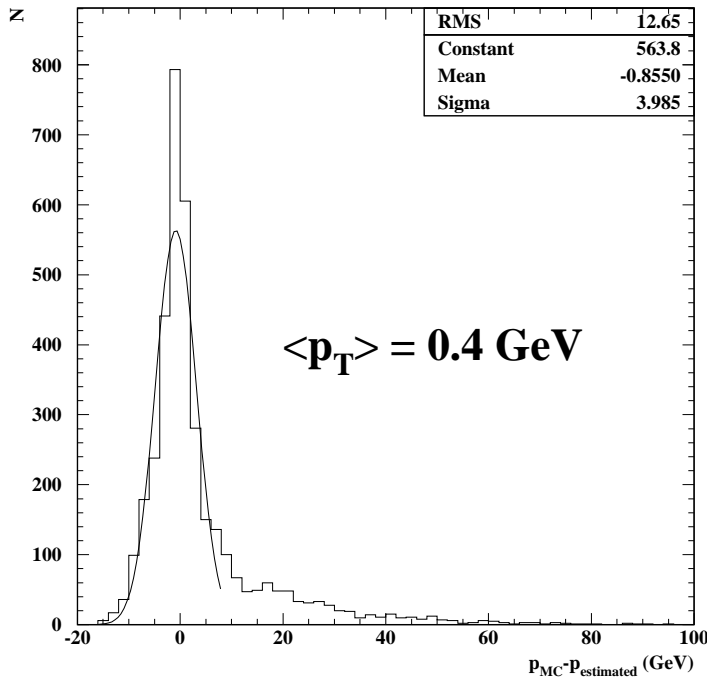


Figure 5.6: Difference between the Monte Carlo momentum  $p_{MC}$  and the estimated momentum  $p_{estimated}$  calculated with Eq. (5.17) and  $\langle p_T \rangle = 0.4 \text{ GeV}$ . The plot was prepared for  $B$  events in which the mean momentum for all reference tracks from the decay chain of either  $B$  is about 17 GeV. The reference tracks from the underlying event have an average momentum of about 6.5 GeV. The value of 0.4 GeV underestimates the real momentum for high-energy particles as can be seen from the tail on the right hand side. Beyond the tail, the momentum is reproduced with an accuracy of about 4 GeV.

The Kalman filter using the above approximations for momentum and multiple scattering starts with the detector plane at highest  $z$  which has a hit. The vector of track parameters  $\vec{t}$  contains the rough values which were calculated in the track reconstruction when extrapolating to the next superlayer. The covariance matrix of the track parameters  $\mathbf{C}$  is initialised with large errors for slopes and intercepts in  $x$  and  $y$  [35]. Both  $\vec{t}$  and  $\mathbf{C}$  are now moved to the position of the next detector layer closer to the target according to Eq. (5.12). The matrix  $\mathbf{Q}_{MS}$  is calculated by inserting the present values of  $\vec{t}$  and the effective average  $x/x_0$  for this layer into Eq. (5.16). In case a hit is present, it is used to improve the estimation of  $\vec{t}$  and to calculate its  $\chi^2$ -contribution. Then the filter proceeds to the next measurement layer. It stops at layer 1 of the last superlayer with a hit. This end condition ensures that also the multiple scattering after the last hit (which might occur for instance already in layer 2) in the same superlayer is included. After leaving the last superlayer with a hit, there are two major places where additional multiple scattering can occur. A track can hit non-sensitive material in the next superlayer upstream. This applies to about 12% of all tracks. The most likely place where this happens is the aluminum cap which is closest to the beam. Even a full silicon wafer can be traversed without giving rise to a hit when a particle passes the non-sensitive border surrounding the area with strips (cf. Fig. 2.4 (a)). In either case, the exact calculation of the passed material is rather difficult. For this reason the effect is neglected in the TLT software.

Scatterers which are taken into account are the four steel bands (cf. Fig. 2.7 on Page 24) of the RF shielding. About 87% of all reference tracks hit one of the  $5\ \mu\text{m}$  thick steel ribbons. The path length in the material varies widely with the particle's angle of incidence. It is calculated in the track refit. The average pathlength is about  $115\ \mu\text{m}$  which corresponds to 0.65% of a radiation length. For particles hitting one of the ribbons, the covariance matrix is moved to the  $z$ -position of the impact point and updated according to the material traversed. In the last step, the covariance matrix is transported to a reference position which is taken as  $z=0$ .

### 5.4.5 Compatibility Analysis of Tracks

In the track following, no attempt is made to mark spacepoints, which were already picked up by a track, as "used" in order to exclude them from the seeding step. When this option was tried, it impaired the efficiency for finding tracks too much to be applicable. This means, however, that a track with spacepoints in four consecutive superlayers is very likely to be found twice, first with all spacepoints and then again as a candidate with three spacepoints. Furthermore, the creation of track branches for all spacepoints found in a search window results frequently in several almost parallel tracks differing by just one spacepoint. This is a consequence of the fact that the same hits can be contained in more than one spacepoint.

The selection of the most likely track among similar candidates is accomplished by an additional algorithm after the Kalman fit. The compatibility analysis bases its decision on the number of hits shared by two tracks and the  $\chi^2$ -probability of the fit as track quality criteria. An unbiased method to find a set of compatible tracks is to determine for each track how many of its hits are contained in other track candidates as well. In each iteration the track with the maximum number of such hits is removed until a limit is reached. The result of such an algorithm does not depend on the sequence for processing tracks, but the method requires frequent comparisons between any two tracks in the sample. In the TLT software, an algorithm needing less comparisons was implemented.

The algorithm works with two track lists and is steered by a parameter  $N_{\text{max}}$ . The first list contains all candidates delivered by the track reconstruction. The second list serves as destination for the selected tracks and is empty in the beginning. A track candidate is either discarded entirely or moved to the second list as a new track or to replace an already inserted track. The decision for each track candidate is derived by finding that track in the second list, which has the maximum number of common hits with the candidate. If the candidate has less than  $N_{\text{max}}$  hits in common with any accepted track it is added to the second list. Otherwise it replaces the found track in any case if it has more hits. For pairs of tracks with the same number of hits, the one with the higher  $\chi^2$ -probability is kept.

The sketched algorithm is not unbiased because the final set of tracks can depend on the sequence of tracks in the initial list. This bias seems to be small in practice and the algorithm rejects badly reconstructed or ghost tracks effectively. With a value of 6 for the parameter  $N_{\text{max}}$  it sorts out multiply reconstructed tracks without impairing the efficiency for finding of real tracks.

### 5.4.6 Primary Vertex Reconstruction

With the reconstructed tracks of an event at hand, one can assign the origin of particles to wires and determine the position of the primary vertices. This task is performed within the TLT framework by an existing and proven package for vertex recognition and vertex fitting.

The VT-package [45] is a general-purpose tool for vertex finding in a region without magnetic field where tracks are straight lines. It expects the parameters of reconstructed tracks and their covariance matrices as input. Vertex fits of two or more tracks are performed using the Kalman filter technique [45, 42]. Each track assigned to a vertex is considered as a measurement of the vertex position. The fit determines the best estimate of the vertex position in space and of the new parameters of all tracks, which are now constrained to originate from the vertex. In contrast to the track fitting described in Section 5.4.4, the relation between a measurement (i.e. a new track parameter vector added to the vertex) and the resulting parameters of the vertex fit is non-linear. In the Kalman filter technique, the relation is linearised and the result is calculated by iteration. A second important difference is that the Kalman smoothing has to be executed in any case, since otherwise the calculated vertex position depends on the sequence used for adding tracks.

The VT-package implements vertex fits using the Kalman filter technique efficiently and allows to impose additional constraints like masses of decaying particles. The TLT software uses only a small fraction of VT’s functionality, namely the search for primary vertices and two-particle vertex fits without constraints. In the standard setting, only the primary vertex search is executed. To this end, the target configuration, and the parameters and covariance matrices of the reconstructed tracks are given to VT. The routine for pattern recognition of primary vertices assigns each track to the target wire with the smallest distance of closest approach. A histogramming technique is used to identify track clusters on each target wire separately. The impact point with the  $x$ - $y$  plane at the target position  $z_T$  is determined for all tracks. The quantity relevant for histogramming is the position of the impact point along the direction of a target wire (for instance its  $x$ -coordinate for horizontal wires). A cluster algorithm extracts groups with at least 3 tracks as vertex candidates and determines the vertex position using the dimensions of the target wire as constraints. When necessary tracks with large  $\chi^2$ -contributions are removed from the vertex. Only primary vertices with at least 3 tracks are stored as final and given back to the TLT software.

### 5.4.7 Performance and Timing of Algorithms

The performance and the timing of the tracking and vertexing algorithm were evaluated using the Monte Carlo simulation of the SVD with the parameters of Table 5.2. It is common practice [46] to quote the efficiency of reconstruction modules for those particles inside the geometrical acceptance, which are of relevance in terms of physics. To remove low-energy tracks only particles with a minimal momentum are considered. All those particles with a momentum greater than 1 GeV, which have at least one hit in three or more silicon superlayers, constitute the set of  $N_{\text{ref}}$  *reference tracks*. The comparison of reconstructed and reference tracks uses the assigned hits rather than the found track parameters as criterion. A reference track is regarded as found when there is a reconstructed track sharing at least 70% of its hits. Tracks which cannot be assigned to a reference particle are counted as ghosts. It can also happen that a reference particle is reconstructed more than once. Such redundant tracks are called *clones*. When  $N_{\text{rec}}$  tracks were reconstructed and  $N_{\text{found}}$  of them could be assigned to a reference particle, track efficiency and percentage of ghost are defined by

$$\varepsilon_{\text{track}} = \frac{N_{\text{found}}}{N_{\text{ref}}} \quad \text{and} \quad \varepsilon_{\text{ghost}} = \frac{N_{\text{rec}} - N_{\text{found}}}{N_{\text{ref}}}. \quad (5.18)$$

The percentage of clones  $\varepsilon_{\text{clone}}$  is determined as the number of clones per reference track. If  $N_{\text{rec}}^{(i)}$  reconstructed tracks were assigned to a reference particle  $i$ , the number of clones  $N_{\text{clone}}^{(i)}$  is

$$N_{\text{clone}}^{(i)} = \begin{cases} N_{\text{rec}}^{(i)} - 1 & \text{if } N_{\text{rec}}^{(i)} \geq 2 \\ 0 & \text{otherwise.} \end{cases}$$

The percentage of clones is then given by

$$\varepsilon_{\text{clone}} = \frac{1}{N_{\text{ref}}} \sum_{i=1}^{N_{\text{ref}}} N_{\text{clone}}^{(i)}.$$

The dependence of the track efficiency and the percentage of ghosts and clones on the number of interactions in the detector was determined by superimposing inelastic events. They are shown in Fig. 5.7 versus the interaction rate. The tracking efficiency for reference tracks is about 85 % for single interactions (up to 5 MHz) and falls up smoothly to 82 % at higher rates. This comparatively low efficiency is mostly due to the explicit requirement of spacepoints in three consecutive superlayers. Only about 90 % of all reference tracks meet this requirement. These tracks are then found with an efficiency of 95 %. A remarkable feature of the algorithm is the small fraction of ghosts among the reconstructed tracks at low interaction rates. Only for very high interaction rates it reaches a level of 19 %. The percentage of clones is in the 3–7 % range.

The efficiency for primary vertex finding is strongly correlated with the track multiplicity of an event. For minimum-bias events the average number of reference tracks is about 6, implying that there is a considerable fraction of interactions where the requirement of at least 3 tracks per primary vertex cannot be met. The overall efficiency for such events was found to be 85 %. It is higher in signal events due to the increased multiplicity (about 10 reference tracks per interaction) and reaches about 96 %.

The TLT software is fast enough to allow for track reconstruction, primary vertex finding, and execution of trigger algorithms within 100 ms. The timing behaviour can be seen in Fig. 5.7 where the time spent in the different parts of the TLT software is shown. At track densities equivalent to running at 40 Mhz, the total time needed is about 80 ms on a PentiumII (333 MHz). This leaves enough time for the execution of the actual trigger algorithms.

The TLT software encompasses a full-fledged track reconstruction which is distributed with every version of the ARTE analysis framework. Besides its demonstrated ability to operate within the SLP, it also proved useful when analysing recorded data in an ARTE environment. The TLT track reconstruction including primary vertex finding is part of the data quality assessment for the SVD running on the online reconstruction farm (cf. Section 4.7.2). It has been operated routinely for extended periods within this framework and provided early estimates of the SVD performance. Among the quantities archived as data quality histograms were the number of reconstructed tracks and the positions of the primary vertices on the target wires.

The TLT track reconstruction was the first algorithm, which used spacepoints for pattern recognition in the HERA-*B* vertex detector. This method made it possible to push the time for a complete SVD reconstruction below 100 ms. Somewhat later, the nowadays leading algorithm for offline SVD reconstruction in HERA-*B*, CATS<sup>6</sup> [47], was developed within the SVD group. CATS takes advantage of spacepoints as well, but employs another method for association of

---

<sup>6</sup>Cellular Automaton for Tracking in Silicon.

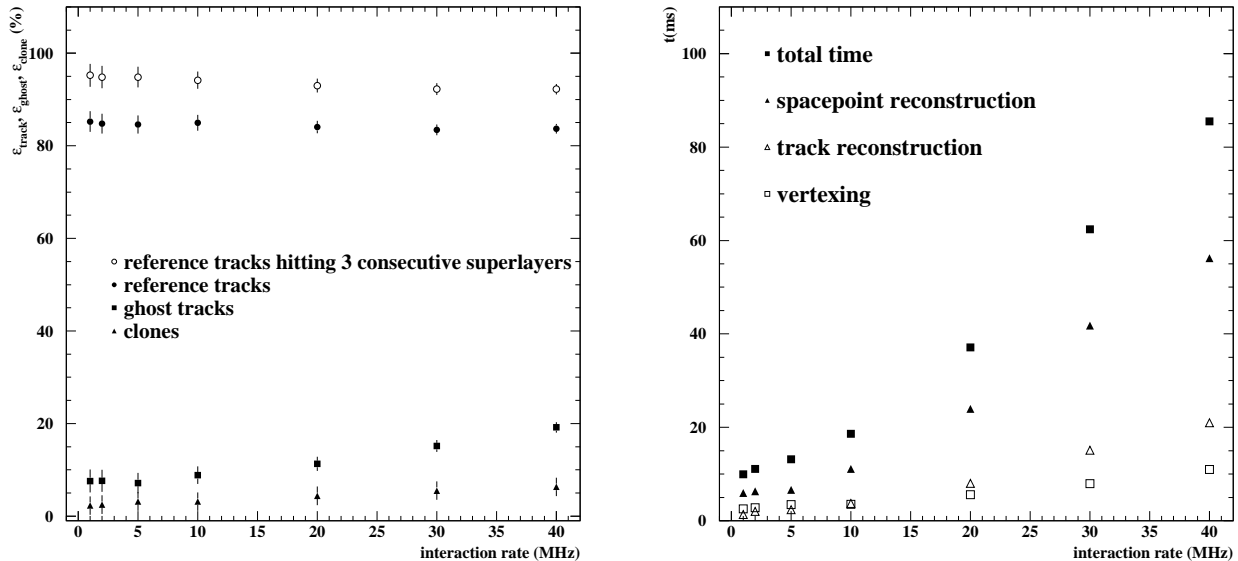


Figure 5.7: **(Left:)** Efficiency and ghost rate of the TLT tracking algorithm measured for inelastic events. The black dots denote the efficiency for all reference tracks which is about 85 % with a slow degradation towards higher interaction rates. The white dots show the efficiency for tracks with hits in three consecutive superlayers. Such tracks constitute about 90 % of all reference tracks. Only they can be found by the tracking algorithm. The percentage of ghosts (squares) is fairly small for low multiplicities but rises considerably with the interaction rate. The fraction of clones (triangles) is in the range of 3 to 7%. **(Right:)** Timing of TLT algorithms using superimposed inelastic events. The program was executed on a PentiumII 333 MHz. The total time is the time spent in the spacepoint reconstruction, in the track reconstruction and in the vertexing part.

spacepoints to tracks. The algorithm has become fast enough to be a very competitive contender for trigger applications [47]. This was motivation enough to investigate it for TLT purposes as well. CATS can be run in the framework of the TLT software within ARTE<sup>7</sup>. It then is used, however, for pattern recognition of tracks only. The TLT software adopts the assignment of hits to tracks and applies its Kalman refit to all segments. Only segments with hits in at least three superlayers are kept. The performance figures for CATS were determined with the same event samples as for the TLT tracking (Fig. 5.8). The efficiency was found to vary between 97 % and 93 % depending in the interaction rate. This superior efficiency is valid for all tracks from the reference set and not only for particles hitting three consecutive superlayers. It is partially bought at the expense of a larger fraction of ghost tracks which exceeds 20 % even for low interaction rates. On the other hand, multiply reconstructed tracks are practically no problem for the CATS tracking. The fraction of clones is smaller than 0.4 % for all interaction rates.

In the following, we turn to the studies of conceivable TLT algorithms which use the reconstructed tracks and vertices as input. These studies were conducted with both CATS and the TLT tracking program.

<sup>7</sup>Note that this is not yet possible in the SLP.

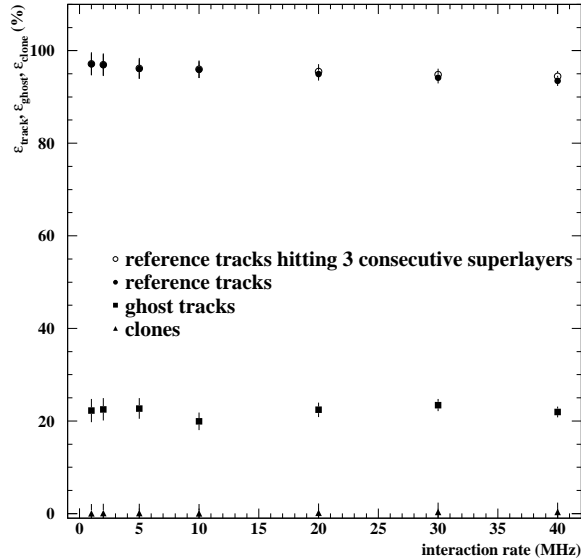


Figure 5.8: Efficiency and ghost rate for the CATS tracking algorithm. The performance figures were determined with the same event samples as for Fig. 5.7. CATS has excellent efficiency for all reference tracks regardless whether they hit three consecutive superlayers or not. This higher efficiency is accompanied by a comparatively large fraction of ghost tracks.

## 5.5 Data Samples

The functionality of a TLT sketched in Section 5.2 encompasses the reconstruction of tracks outside of the RoIs and of the primary vertices. The effort for the determination of this additional information is only justified when it indeed serves to record data at a higher efficiency for signal events. The calculation of a signal efficiency for a particular TLT algorithm makes no sense without a statement about the attainable suppression of background events. The strategy followed was therefore to compare the combined signal efficiency for SLT and TLT with the efficiency obtained by the SLT alone under the requirement that the overall background suppression should assume the same value in either case.

This proceeding needs data samples of both signal and background events, which the TLT software will process. The **signal sample** used contained double-semileptonic decays of  $B$  mesons. Such events can be generated readily using the tools mentioned in Section 5.3.1. Already at generator level, one can impose cuts on the transverse momentum of the leptons. A  $p_T$  greater than 1 GeV was required keeping  $(37 \pm 2)\%$  of the events according to the PYTHIA generator. The trigger particles that the FLT and SLT are expected to find are the leptons from the decay cascade of both  $B$  hadrons in the event.

To obtain a sample with background events, which a third level trigger is likely to face, is much more difficult. Up to now, the exact composition of such a sample is not known yet because the computational effort for its generation is prohibitively large and some input quantities (like the  $p_T$ -spectra of particles in inelastic interactions) are not known reliably. Even with a concise description of the physical processes at hand, all conceivable background sources had to be subjected to a complete simulation of the detector response and the influence of the trigger levels. Given the large suppression of the FLT and SLT of about  $2 \cdot 10^4$ , millions of events

had to be generated, stored and processed in order to find out what fraction will survive all cuts and leave the SLT. Since the generation of a single event takes about 1 minute on today's microprocessors this is not a manageable task. When studying the SLT algorithms in detail, the complete treatment of background events had to be stopped after Slicer and Refit (cf. Section 3.5.1) due to lack of Monte Carlo statistics [48]. The needed background samples will become available automatically from the experiment once the trigger chain is set up completely. Until then, one has to resort to plausible assumptions about the most important background sources, which can be studied using limited Monte Carlo samples. In doing so, one can exploit the ability of the ARTE interface to the Monte Carlo generators [37] to populate those regions of the phase space with events that the triggers are likely to select.

The SLT will require at least two tracks in the SVD, which were identified as electron, muon, or high- $p_T$  hadron by the pretriggers. When concentrating on events with trigger particles, which do not necessarily form a common vertex, the following processes can give rise to this signature:

- \* **Dileptons from decays of  $c$  quark pairs.** Events of the type  $gg \rightarrow c\bar{c} \rightarrow D\bar{D} \rightarrow llXY$  result in two genuine leptons. The charm cross section at the HERA- $B$  centre-of-mass energy is well below  $35 \mu\text{b}$  [49]. Less than 6% of all  $D\bar{D}$  pairs undergo a double-semileptonic decay. When running at 40 MHz interaction rate, events with two leptons from charm decays occur at a rate of about 6 kHz. The rate can be limited by applying a cut on the transverse momentum of the leptons. A study at generator level has shown that already a cut at 1 GeV suppresses such events by a factor of  $(510 \pm 50)$ . When folding in the geometrical acceptance of the spectrometer and the trigger efficiencies, double-semileptonic charm events will be of interest in the offline analysis rather than in the trigger system.
- \* **Hadrons misidentified as leptons.** Hadrons coming from the target region can be misidentified as leptons. Fake leptons can arise from punch-through in the muon system and from showers with a strong electromagnetic component, which were induced by a hadron. The impact of such events depends on the misidentification probability. The calorimeter is more susceptible to fake leptons than the muon system due to the higher occupancy, resulting in overlapping showers. In the HERA- $B$  Proposal, a misidentification probability of 3% (1%) was assumed for tracks impinging on the calorimeter with an energy of 5–20 GeV (>20 GeV). The impact of the hadron misidentification probabilities was estimated using a Monte Carlo sample with high- $p_T$  inelastic events. It was required that there are at least two reference tracks in the SVD with a  $p_T$  greater than 1 GeV. Given the momentum of these reference tracks one can estimate the misidentification probability for such events which was found to be about  $10^{-3}$ . This means that the  $p_T$  cut alone has to provide another factor of  $10^3$  suppression to reduce the rate of such events sufficiently.
- \* **Tracking errors in the SLT.** There are chances that the SVD track segment found by the SLT does not belong to the lepton or hadron candidate identified by the pretrigger and the FLT. This can happen when the SVD segment is a pure ghost, or many track candidates behind the magnet point to the seed and a wrong track is followed to the target region. Such cases are most likely to occur in high-multiplicity events with large background of ghost tracks. Most of these events will be rejected presumably by one of the SLT algorithms when the number of track candidates exceeds some maximal value.

Another option might be to restrict the size of the RoI in the track propagation, which should cut down the number of combinations significantly. The problem is again that such phenomena are difficult to study in the Monte Carlo because many parameters (settings of the SLT algorithms, noise levels in the detectors) are involved.

- \* **Conversions in one of the front SVD superlayers.** Conversions of photons to electron pairs in the material of the vertex detector give rise to tracks which might survive the selection criteria of the FLT and SLT. The tracks are emitted into the flight direction of the photon and point therefore back to the interaction region.

Of course, the described background processes will occur as a mixture in the events. A trigger could be generated for example by a real lepton from a  $D$  meson and a misidentified hadron. In how far the various processes can be excluded depends on the performance of the detector and the trigger system.

The most serious background for HERA- $B$  are inelastic events surviving the trigger criteria. To be accepted by FLT and SLT they must contain more than one track with a sufficient transverse momentum which can be followed back to the target region. The **background sample** was assumed to comprise high- $p_T$  inelastic interactions with light quarks only where at least two tracks have a  $p_T$  greater than 1 GeV and are measured in the SVD. Such events have roughly twice as much measurable tracks in the vertex detector than ordinary inelastic events. The trigger particles are taken as the SVD tracks above the  $p_T$  threshold of 1 GeV.

To cover both the case of misidentification and of tracking errors in the SLT, the background sample was used in two different ways. First, the SLT was assumed to follow the high- $p_T$  tracks correctly into the vertex detector. We will refer to the background events in this mode as **misidentification sample**. In the second scenario, the case was studied that the SLT finds a real track segment in the SVD which belongs, however, not to the high- $p_T$  track. It was assumed that the momentum of the high- $p_T$  track is measured correctly by L2Magnet but the propagation into the SVD results in a wrong track segment for exactly one trigger particle. For this purpose one high- $p_T$  track was exchanged with another particle from the event which was well measurable in the SVD. The wrong track should be reasonably close to the trigger track in order to have a chance to be found in the RoIs. The particle was selected from the set of all tracks with a transverse momentum below 1 GeV by the requirement that its track parameters are as similar as possible to the high- $p_T$  track. Similarity was defined simply as the Euclidian distance between points in the four-dimensional space of track parameters in the SVD. The momentum of the wrong track was taken over from the original track. It is clear that this procedure will result in a SLT track of poor quality with lower average momentum which will suffer more from multiple scattering. The extraction of track parameters is obscured by the fact that not the appropriate momentum is used. The background events in the second scenario will be referred to as **tracking error sample**.

## 5.6 Trigger Simulation

Unfortunately, the two event samples could not be subjected to the full FLT and SLT simulation since no working version of either package was available for the used ARTE version. The impact of FLT and SLT was therefore emulated by the TLT software on the basis of the following considerations.

The First Level Trigger does not include the vertex detector in its decision. Its functionality was assumed to be equivalent to the  $p_T$ -cut applied to the trigger particles in combination with

the requirement that the tracks fall into the geometrical acceptance of the detector. Any combination of lepton flavour and charge was accepted for the signal sample.

The Second Level Trigger was assumed to improve the resolution of the momentum determination to such an extent that effects resulting from the finite momentum resolution at the FLT can be neglected safely. Only tracks with a  $p_T$  greater than 1 GeV and at least three hits in both the  $x$ - and  $y$ -view of the SVD (i.e. at least six hits in total) were considered as primary SLT tracks. For the signal sample, only leptons were passed to the SLT while for the background sample the tracks above the  $p_T$ -cut (or the wrong counterpart for one track) were taken into account. The SLT emulator assigned each trigger track to a target wire by calculating the distance of closest approach. Only events with exactly two trigger particles close to the same wire were passed on to the TLT. Events with less than two tracks will be rejected by the SLT, events with more than two tracks are very rare and occur only in the signal sample when the decay cascade of the  $B$  mesons creates more than two leptons with sufficient  $p_T$ . Presumably, three high- $p_T$  leptons assigned to the same target wire are a very strong argument for an interesting event, which should be kept by the SLT without further inspection by the TLT. The SLT will presumably impose cuts on the distance of closest approach (DCA) between tracks to remove events in which the trigger particles come from different interactions. The DCA was required to fall below 3 mm. This keeps more than 99.5 % of all events in each sample but removes some events from the background sample where the trigger tracks are separated by distances as large as some centimetres. Such events occur when a track from a hadron decay in flight points to a position far outside of the rectangle formed by the four target wires. The information given from the SLT to the TLT comprised the two trigger tracks and the target wire they were assigned to. The first step of the TLT processing after track and primary vertex finding was in any case to discard those TLT tracks sharing more than 50 % of the hits with one of the two trigger particles.

The assignment of SVD hits to SLT tracks was extracted from the simulation. There is good reason to believe that pattern recognition errors can be brought down to a negligible level for the SLT tracks by applying the Kalman smoothing for outlier removal. Nothing like that is implemented in the SLT software at the moment, but it might be done there or constitute the first step of the TLT processing. The smoothing should be possible without spending too much time since the number of SLT tracks is small. The track parameters of the trigger particles were estimated by applying the Kalman refit using the Monte Carlo momentum for the signal and the misidentification sample. To neglect the error of the momentum determination in the SLT is justified because it can be surmised that the result of most studies will be dominated by the poorer quality of TLT tracks. In the case of the tracking error sample the wrong tracks were fitted using the momentum of the original high- $p_T$  track.

## 5.7 Impact Parameter Studies

The SLT will apply cuts on the impact parameters that the trigger particles have with respect to the target wire. After the reconstruction of all tracks in the TLT step, also the position of the primary vertex is known and the distance between track and vertex can be calculated. In the following, the accuracy of the different kinds of impact parameters which is decisive for their application as a cut criterion is discussed. We then proceed with a study on various ways of combining impact parameters into a trigger decision and try to estimate the gain which can be expected from the firm knowledge of the primary vertex.

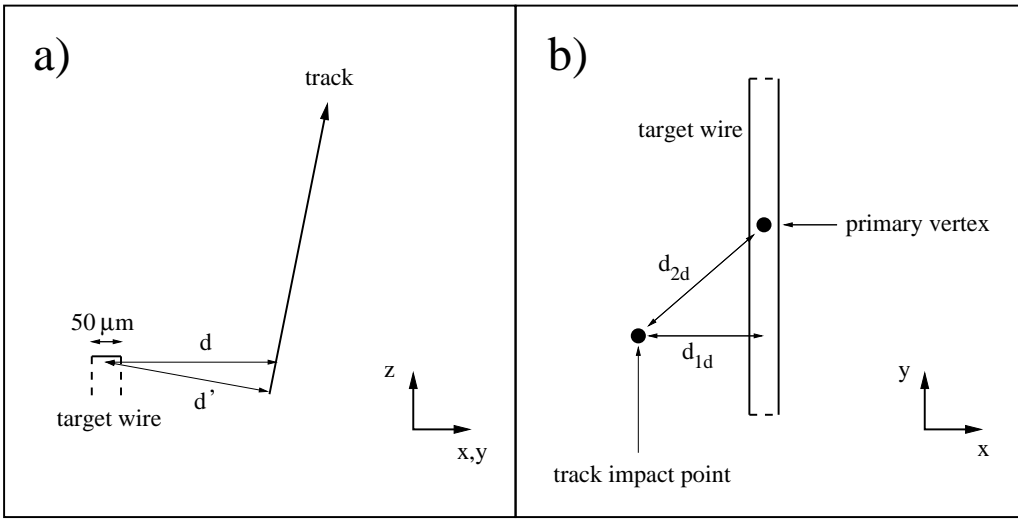


Figure 5.9: (a) In the TLT software, the impact parameter with respect to the target is defined as the distance between track and wire in the  $x$ - $y$  plane at the  $z$ -position of the wire ( $d$ ). The difference to the quantity  $d'$ , the distance of closest approach between wire and track, is only 3%. (b) In the SLT, cuts are placed on the distance between the track and the target wire ( $d_{1d}$ , *one-dimensional impact parameter*). Once the position of the primary vertex is known, one can determine  $d_{2d}$  as *two-dimensional impact parameter*.

### 5.7.1 Impact Parameter with respect to the Target

The impact parameter is given by the distance of closest approach between a track and a wire at the  $z$ -position  $z_w$  ( $d'$  in Fig. 5.9 (a)). One can equally well use the shortest distance between the measured impact point of the track with the  $x$ - $y$  plane at  $z_w$  and the wire ( $d$  in Fig. 5.9 (a)) since both quantities differ only due to the track slope in the plane perpendicular to the wire in question. Even for the largest slopes the difference between  $d$  and  $d'$  is only 3%. Throughout this thesis,  $d$  is the discussed quantity and is referred to as *one-dimensional impact parameter*. The distributions of the calculated impact parameters of the trigger particles from the data samples introduced in Section 5.5 are shown in Fig. 5.10. The leptons from the decay cascade of  $B$  hadrons have an average impact parameter of  $263 \mu\text{m}$  and the spectrum falls off exponentially. Most of the trigger particles in the misidentification sample come directly from the primary vertex and acquire an impact parameter due to the limited accuracy of the measured track parameters. Very large impact parameters (greater than  $300 \mu\text{m}$ ) stem from decays of hadrons in flight. The average impact parameter is  $55 \mu\text{m}$ . The tracking error sample combines one high- $p_T$  track and one particle with lower momentum. The average impact parameter of  $100 \mu\text{m}$  is therefore larger than in the misidentification sample.

The accuracy of the one-dimensional impact parameter of a track depends on the properties of the vertex detector and on the momentum of the particle which influences the amount of multiple scattering. The first measurement of a particle trajectory behind the target determines its coordinate with an accuracy  $\sigma_0$  which is about  $12 \mu\text{m}$ . The error of the impact parameter is the sum of  $\sigma_0$  and the extrapolation error

$$\sigma_{\text{impact}}^2 = \sigma_0^2 + (\sigma_{\text{slope}} \cdot D)^2. \quad (5.19)$$

Here  $D$  denotes the distance between the first measurement and the point where the impact parameter is calculated. In absence of multiple scattering, the error of the track slope  $\sigma_{\text{slope}}$

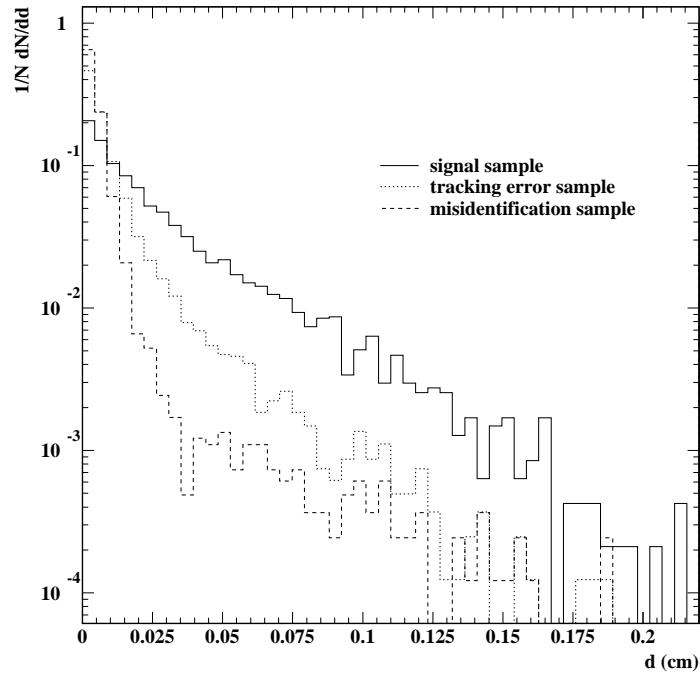


Figure 5.10: Distributions of the impact parameters with respect to the target wire for the trigger tracks in signal events (solid histogram) and the background sample. The dotted and dashed histogram represent the tracking error and the misidentification sample. All histograms were scaled to represent the probability distributions rather than absolute numbers. The mean impact parameter is  $263 \mu\text{m}$ ,  $100 \mu\text{m}$  and  $55 \mu\text{m}$ , respectively (top to bottom).

is a function of the number of hits on the track, their resolution  $\sigma_0$  and the  $z$ -coordinates of the measurements. In principle, it could be made smaller by sampling the particle trajectory more often. The attainable precision of the track slope is limited, however, by the multiple scattering in the detector material. The error of the track slope  $\sigma_{\text{slope}}$  is dominated by the multiple scattering in the first measurement layer and in the material traversed between the particle origin and the first measurement layer. It can be written using Eq. (5.15) as

$$\sigma_{\text{slope}} \cdot D \simeq \theta(x/x_0) \cdot D = \theta(x/x_0) \cdot \frac{p}{p_T} R_T. \quad (5.20)$$

In the rightmost term,  $R_T$  denotes the radial distance of the first measurement from the beam. Clearly, the error of the impact parameter can be minimised by keeping  $D$  as low as possible and by aiming for the lowest amount of material  $x/x_0$  between the first measurement and the target. The HERA- $B$  vertex detector was designed to provide the first measurement of a particle trajectory at a constant distance of about 1 cm from the beam independent of the track angle. With  $R_T \simeq \text{const.}$ , the error of the impact parameter is to good approximation a function of the transverse momentum

$$\sigma_{\text{impact}} = \sqrt{A^2 + (B/p_T)^2} = A \oplus (B/p_T) \quad (5.21)$$

with constants  $A$  and  $B$ .

Besides the mechanisms elucidated above, pattern recognition errors and the approximate treatment of multiple scattering in the Kalman refit influence the impact parameter resolution. Hits which were assigned wrongly to a track pull the fit into a false direction and spoil the knowledge of the track parameter vector. The consequences are most dramatic when such a wrong assignment occurs in the first superlayer with a hit. The unprecise knowledge of the particle momentum and of the material passed in the Kalman refit prohibits to describe the influence of multiple scattering between two measurements correctly. This prevents the TLT software from obtaining really the best estimate of the final track parameters at the target. To quantify the various effects the impact parameter resolution was determined for three different scenarios:

- The Kalman refit was applied to the hits contained in the reference tracks (ideal pattern recognition) using also the correct momentum.
- As before, but the momentum was estimated from Eq. (5.17) on Page 77.
- The TLT track finding is executed in the normal setting. The pattern recognition is done by the TLT algorithm or by CATS and the Kalman refit employs the momentum approximation (5.17).

The accuracy of the one-dimensional impact parameter was determined by comparing the reconstructed values  $d$  with the Monte Carlo truth  $d_{\text{MC}}$ . For the case of TLT pattern recognition only tracks classified as found were taken into account. (These are tracks sharing at least 70 % of the hits with a reference track.) It follows that very badly reconstructed tracks with a lower fraction of correctly assigned hits are ignored. The Gaussian width  $\sigma_{\text{impact}}$  of the quantity  $d - d_{\text{MC}}$  was extracted for all tracks in bins of  $p_T$ . For this purpose, the impact parameter was treated as a signed quantity defined as the difference of track impact point and wire position. The parameters  $A$  and  $B$  can be obtained by fitting the found values as a function of  $p_T$  using (5.21). The results are summarised in Fig. 5.11 and Table 5.4. Using ideal pattern recognition

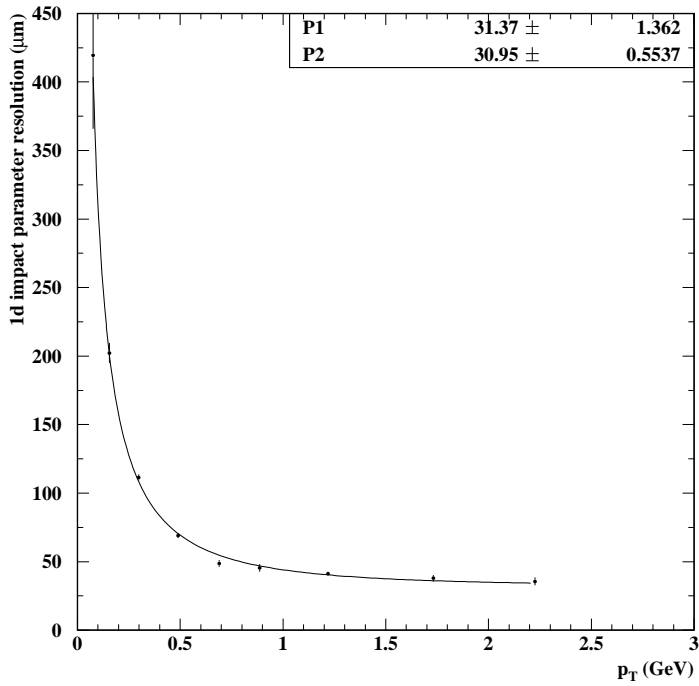


Figure 5.11: Resolution of the impact parameter with respect to the target wire which is reached by the TLT track finding. Only reconstructed tracks sharing at least 70% of the hits with a reference track were taken into account.

and the known momentum the obtained impact parameter resolution is  $(23 \oplus 31/p_T) \mu\text{m}$  ( $p_T$  in GeV). This value is very close to the impact parameter resolution quoted in the HERA-*B* Proposal of  $(25 \oplus 30/p_T)$ . Note that the Kalman fit was not perfect even for known momentum, because the amount of material describing multiple scattering is averaged over the surface of a silicon counter according to the radial distribution of reference tracks. The impact parameter resolution suffers slightly when the momentum of particles can only be estimated. For the TLT, it was found to be  $(31 \oplus 31/p_T) \mu\text{m}$ . This value is almost independent on whether the TLT software or the CATS algorithm was executed for track finding and agrees well with the figure obtained for the case of ideal pattern recognition and unknown momentum. One can conclude that pattern recognition effects are small and have no significant influence on the resolution. This is confirmed by the observation that for both TLT track finding and CATS the fraction of incorrectly assigned hits is smaller than 0.6%.

Table 5.4: Impact Parameter Resolutions.

	A ( $\mu\text{m}$ )	B ( $\mu\text{m}\cdot\text{GeV}$ )
ideal pattern recognition, momentum known	$22.9 \pm 1.2$	$31.4 \pm 0.6$
ideal pattern recognition, momentum estimated	$31.2 \pm 1.2$	$30.6 \pm 0.6$
TLT tracking	$31.4 \pm 1.4$	$31.0 \pm 0.6$
CATS tracking	$30.7 \pm 1.3$	$32.5 \pm 0.7$
HERA- <i>B</i> Proposal	25	30

According to Table 5.4 the deterioration of the impact parameter resolution due the lacking momentum information at the TLT is relatively small. A more serious problem is the fact that the error of a measured impact parameter  $\sigma_{\text{impact}}$  cannot be estimated correctly. This in demonstrated in Fig. 5.12 where the normalised residual

$$\frac{d - d_{\text{MC}}}{\sigma_{\text{impact}}}$$

of the one-dimensional impact parameter is shown. For ideal pattern recognition and known momentum, the covariance matrices of the track parameters reflect the uncertainty properly and are a valuable input to the error propagation. The normalised residual of the impact parameter (Fig. 5.12 (left)) takes the shape of a Gaussian distribution centred at zero. The standard deviation is compatible with unity and there are no pronounced tails. When the Kalman refit is repeated for the same track sample but the momentum estimation (5.17) is used, the resulting distribution (Fig. 5.12 (middle)) is clearly not Gaussian anymore and has long tails. The additional influence of pattern recognition errors is completely negligible as can be seen from Fig. 5.12 (right) where the TLT software instead of the simulation provided the assignment of hits to tracks before the Kalman fit determined the track parameter vector at the target. The observed behaviour of the normalised residuals in Fig. 5.12 indicates that one has to be cautious when using the quantity  $d/\sigma_{\text{impact}}$  to reject or accept TLT tracks as candidates for particles from decays of *B* hadrons.

## 5.7.2 Impact Parameter with respect to the Primary Vertex

The impact parameter with respect to the primary vertex is defined as the distance of closest approach between the vertex and a track. When projected on the *x-y* plane is has components

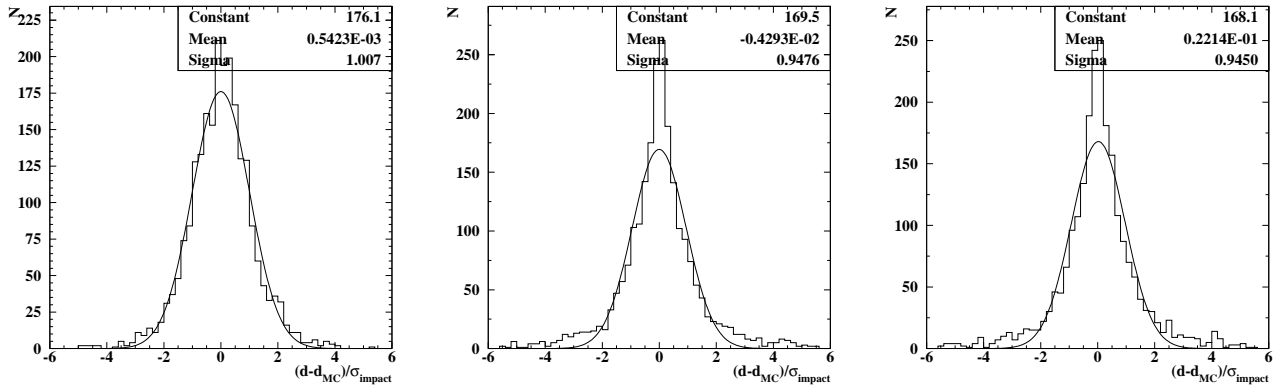


Figure 5.12: Normalised residual  $(d - d_{MC})/\sigma_{\text{impact}}$  of the one-dimensional impact parameter. In the left plot, ideal pattern recognition and the momentum from the simulation were used for fitting of tracks. The error of the impact parameter cannot be calculated properly when the momentum is not known. This can be seen in the middle and right plot where the momentum was only estimated. The assignment of hits to tracks was taken from the simulation (middle) and from the TLT software (right), respectively.

in both  $x$  and  $y$  (cf. Fig. 5.9 (b)) and is therefore referred to as the *two-dimensional impact parameter*. Its resolution cannot be expressed conveniently by a single formula like Eq. (5.21) since it has implicit dependencies on various parameters varying from event to event. The resolution is governed by both the accuracy of the track and of the primary vertex found. The precision of the reconstructed primary vertex position was studied for the signal and the background sample. To this end, the output of the VT-package was compared with the vertex coordinates from the simulation.

As mentioned above, the VT-package exploits the wire dimensions as an additional constraint for the position of the primary vertex. The resolutions along the wire and perpendicular to it differ therefore strongly. The expected resolution perpendicular to the wire  $\sigma_T$  is given by the width of the ribbon  $W$  as  $W/\sqrt{12}$ , i.e. about  $14 \mu\text{m}$  for  $W = 50 \mu\text{m}$ . This value is reached indeed (Fig. 5.13 (left)) and it follows that the uncertainty of the coordinate perpendicular to the wire is dominated by the track (as it should be for the benefit of the measurement of one-dimensional impact parameters at the SLT). The error of the one-dimensional impact parameter perpendicular to the wire is at least twice as large as  $\sigma_T$ , even for high-energy tracks where multiple scattering does not play an important role. The resolution of the primary vertex along the wire  $\sigma_L$  and in  $z$  improves with increasing track multiplicity per interaction. In the signal sample, the average number of reference tracks per interaction is about 10 and the mean number of tracks per primary vertex is 8. This determines the coordinate along the wire and in  $z$  with an accuracy of  $66 \mu\text{m}$  and  $182 \mu\text{m}$ , respectively. In the inelastic events, about 12 tracks per interaction are measured in the geometrical acceptance of the SVD and 11 of them help to pinpoint the primary vertex position with an accuracy of  $28 \mu\text{m}$  along the wire and  $143 \mu\text{m}$  in  $z$  (cf. Fig. 5.13 (middle and right) and Table 5.5).

The efficiency for finding the primary vertex is strongly correlated with the track multiplicity. The larger the overall number of tracks, the higher is the probability to find a primary vertex with at least three tracks. The efficiency of the VT primary vertex search was evaluated for the signal and the background sample. Agreement between a reconstructed vertex and the Monte

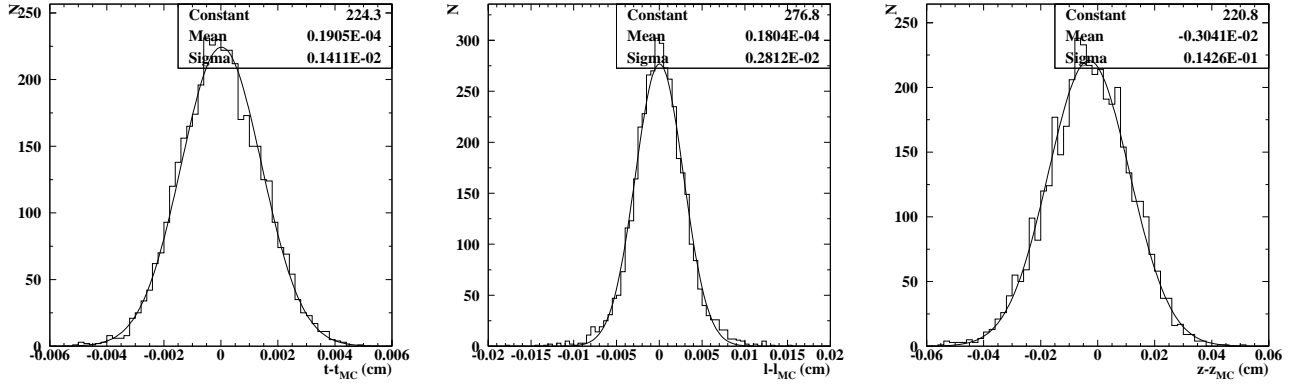


Figure 5.13: Resolution of the primary vertex position in the inelastic sample. The plots show the difference between the reconstructed and the simulated coordinate perpendicular to the wire (left), along the wire (middle) and in  $z$ . The standard deviations are  $14 \mu\text{m}$ ,  $28 \mu\text{m}$  and  $143 \mu\text{m}$  (left to right).

Table 5.5: Parameters of the primary vertex finding in the signal and background sample.  $N_1$  is the average number of reference tracks per interaction,  $N_2$  the mean number of tracks on a reconstructed primary vertex. The values for  $\sigma_T$  and  $\sigma_L$  specify the resolution in the  $x$ - $y$  plane perpendicular to and along the wire,  $\sigma_z$  is the resolution in  $z$ . The efficiency for primary vertex finding is shown in the column  $\varepsilon_{3\sigma}$  ( $\varepsilon_{4\sigma}$ ) under the requirement that the reconstructed coordinate is in a  $3\text{-}\sigma$  ( $4\text{-}\sigma$ ) environment in  $x$ ,  $y$  and  $z$  around the primary vertex position taken from the simulation.  $R$  is the average number of reconstructed vertices per Monte Carlo vertex if at least one vertex was reconstructed.

sample	$N_1$	$N_2$	$\sigma_T$ ( $\mu\text{m}$ )	$\sigma_L$ ( $\mu\text{m}$ )	$\sigma_z$ ( $\mu\text{m}$ )	$\varepsilon_{3\sigma}$	$\varepsilon_{4\sigma}$	$R$
signal	10.3	7.9	18	66	182	90 %	95 %	1.18
background	12.0	11.0	14	28	143	96 %	97 %	1.11

Carlo was checked merely by parameter matching, i.e. the reconstructed and the simulated coordinate were compared. It was not judged whether the tracks assigned to a vertex originate indeed from the primary interaction or are due to secondary particles. A simulated primary vertex was counted as found when the coordinate of a reconstructed vertex lay in a  $3\text{-}\sigma$  interval around the Monte Carlo position in  $x$ ,  $y$  and  $z$ . Using this criterion, the efficiency is 90% in the signal sample, and 96% in the background events. When going to a  $4\text{-}\sigma$  interval, even 95% and 97% of all primary vertices are classified as found. This is a good efficiency for finding of the primary vertex. It was found, however, that the VT package frequently reconstructs more than one primary vertex when exactly one was predicted by the simulation. The amount of additional vertices in close vicinity to the one matching the Monte Carlo best, is 18% in the signal sample and 11% in the inelastic events. The higher fraction of ghost vertices in the  $B$  sample is most likely a consequence of the different event topology with the frequent occurrence of tracks with sizable impact parameters. When confronted with more than one primary vertex on a target wire, the one with the lowest distance to a track was used whenever two-dimensional impact parameters were calculated.

### 5.7.3 Impact Parameter Cuts

In the following, we want to investigate whether the known primary vertex position can serve to refine the impact parameter cuts applied to the trigger particles. A possible scenario could be that the SLT cuts on the one-dimensional impact parameter to bring down the rate to a level acceptable for the TLT. The TLT determines then the primary vertex and refines the trigger decision using the two-dimensional impact parameter.

The prospects of such an approach are determined by the resolution of the coordinate perpendicular and along the target wire. The resolution of the primary vertex along the wire was found to be  $66\text{ }\mu\text{m}$  in the signal sample. This means that this coordinate is known with lower precision than the one-dimensional impact parameter as far as the trigger particles are concerned. According to Table 5.4 the resolution is better than  $40\text{ }\mu\text{m}$  for all particles with a transverse momentum above 1 GeV. To evaluate the benefit of the primary vertex finding quantitatively, we have compared the signal efficiency at a given suppression for background events, which can be reached when cuts are imposed on the one-dimensional or the two-dimensional impact parameter of the trigger particles.

The SLT will use a combination of the impact parameters of the trigger particles as cut criterion. For simplicity, only the presumably most abundant case of exactly two trigger particles assigned to the same target wire is treated here. There are various ways to combine two measured impact parameters  $d_1$  and  $d_2$  into a decision. One can either use a direct function  $f(d_1, d_2)$  or rely on the error  $\sigma_{d_i}$  of the impact parameter calculated in the silicon part of the SLT and use a function  $g(d_1/\sigma_{d_1}, d_2/\sigma_{d_2})$ . The second form might be more appropriate when the two trigger tracks differ significantly with respect to their number of hits and the measured momentum. On the other hand, the SLT will accept only tracks passing a quality criterion anyway and the impact parameter resolution for particles with a transverse momentum greater than 1 GeV remains almost constant (cf. for example Fig. 5.11). This suggests that a function  $f(d_1, d_2)$  which is symmetric in its arguments is a plausible choice. We have used four different cut functions to compare their performance:

- The sum of the two squared impact parameters  $f = \sqrt{d_1^2 + d_2^2}$  prefers the larger of the two.

- The product of impact parameters  $f = d_1 \cdot d_2$  was advocated already in Ref. [50] to disentangle double-semileptonic decays of  $B$  and  $D$  meson pairs.
- $f = \text{MAX}(d_1, d_2)$  bases the trigger decision entirely on the larger of the two impact parameters. It is equivalent to the Gaussian addition when the magnitudes of  $d_1$  and  $d_2$  differ strongly.
- The sum  $f = d_1 + d_2$  of the two impact parameters.

In this study, the SLT was assumed to evaluate  $r = f(d_1, d_2)$  to reject events with  $r$  lower than some  $r_{\text{cut}}$ . It is difficult to make plausible assumptions about the actual value of  $r_{\text{cut}}$  for the different cut functions since the required suppression not known. At the present stage of the trigger chain, the projection of tracks into the silicon vertex detector alone provides enough suppression. The event rate is then low enough to pass on the data stream to the online reconstruction farm. The cut value to use will be extracted from the experiment by adjusting it in a way that the output rate of the SLT is manageable. On the other hand, any cut imposed on the one-dimensional impact parameter would bring down the Monte Carlo statistics available for these studies strongly because the corresponding distributions are very steep. Therefore, no cut was imposed on the one-dimensional impact parameters before performing the primary vertex search at the TLT. This is equivalent to a situation where the requirement of two trigger tracks assigned to the same target wire results in a SLT output rate of about 500 Hz, which is the designed input rate of the TLT.

The TLT track finding was applied to all events of the signal and the background sample to reconstruct primary vertices on the wire denoted by the SLT emulation. The two-dimensional impact parameter can be calculated from the found vertex position and the track parameters of the trigger particles. In events with more than one reconstructed primary vertex, the smallest resulting two-dimensional impact parameter was kept conservatively. The cut functions described above were evaluated for both pairs of one-dimensional and two-dimensional impact parameters. The particular values for  $r_{\text{cut}}^{(1d)}$  and  $r_{\text{cut}}^{(2d)}$  were adjusted independently to obtain the desired suppression using the background sample. The efficiency for the signal sample is then the fraction of events kept when the selected cuts are imposed. This procedure was repeated for different suppressions and the four proposed cut functions. The suppression factors were varied around the value of 10 expected from the TLT, with  $10^2$  as the maximal value.

The result is depicted in Fig. 5.14 where the efficiency for signal events is shown versus the efficiency for events from the misidentification sample. The four plots correspond to the different cut functions tested. The black dots denote the efficiency obtained using cuts on the one-dimensional impact parameter, the white circles correspond to the case that the trigger decision took advantage of the two-dimensional impact parameters. It is seen that neither cut function shows superior performance in comparison with the others. At the same suppression for background events, very similar efficiencies for signal events are obtained. The usage of the two-dimensional impact parameter results in general in a very small efficiency gain. When the product of the two impact parameters is used as trigger criterion, a cut on the two-dimensional impact parameter shows even a slightly worse performance than the one-dimensional impact parameter. The result is reproduced when the tracking error sample constitutes the background.

The benefits of the primary vertex finding seem to be rather limited in practice. The knowledge of the primary vertex can serve to identify those trigger tracks in background events which have a large spatial separation from the interaction. Such tracks are most likely to arise from hadrons decaying in flight or pattern recognition errors in the SLT. On the other hand, the

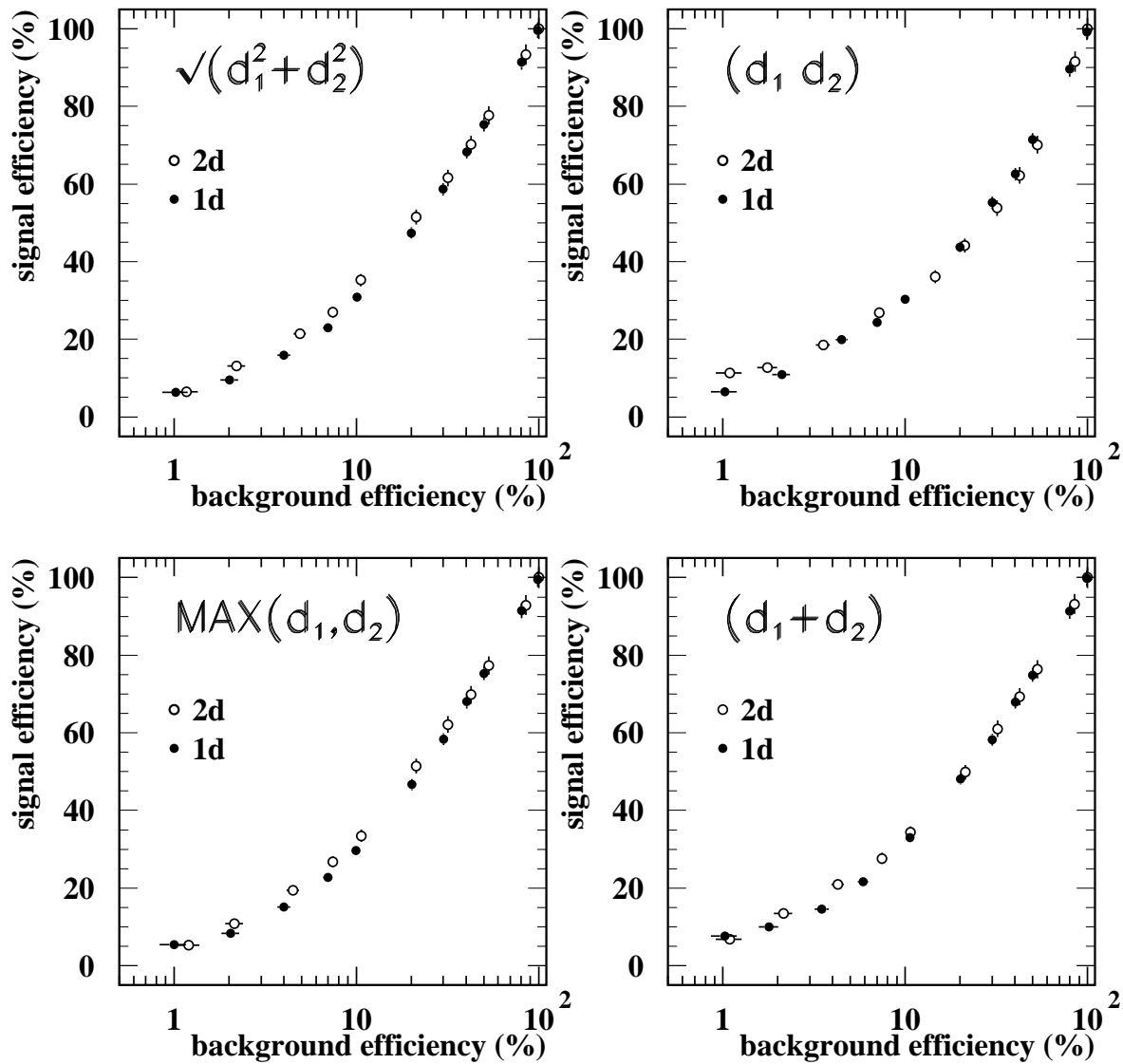


Figure 5.14: Signal efficiency versus efficiency for events from the tracking error sample. The black dots denote the efficiency obtained by the SLT when imposing cuts on a suitable combination of the two one-dimensional impact parameters  $d_1$  and  $d_2$  with respect to the target wire. The SLT was assumed to cut on the sum of the squared impact parameters (upper left), the product of the two (upper right), the larger of the impact parameters (lower left), and their sum (lower right). Once the primary vertex is found at the TLT, it can impose cuts on the full impact parameter. The resulting signal efficiency for cuts on the two-dimensional impact parameter is shown with white dots.

SLT itself can reject such events already by imposing a moderate cut on the maximal allowed distance of closest approach between any two trigger tracks. For this proceeding the knowledge of the primary vertices is not even needed. Placing cuts on the two-dimensional rather than on the one-dimensional impact parameter is probably not worth the effort.

## 5.8 Vertexing with SLT Leptons

Secondary or even tertiary vertices are strong indications for an event with heavy quarks. In principle, one can combine any two tracks found by the joint effort of SLT and TLT in the SVD into a vertex candidate. In doing so one will restrict the selection to tracks in the vicinity of the target wire identified by the SLT. The most likely candidates are tracks, which were not assigned to the primary vertex in the first place, and pairs with a reasonably small distance of closest approach. An attempt to devise such a general search for secondary vertices was made in the framework of the VT-package and proved to be a very challenging task [51] due to the difficult vertex recognition step and the limited resolution arising from the presence of multiple scattering in the detector.

The above result was confirmed by inspecting the output of the VT vertex recognition routines. Own attempts to develop a general-purpose secondary vertex recognition failed and suggested to follow a less ambitious strategy for the TLT. Instead of combining arbitrary tracks, only vertices formed by one SLT particle and one TLT track are searched for. This proceeding guarantees that the parameters of one particle on a vertex candidate are well-defined since the momentum of SLT tracks is known. It exploits the SLT information as starting point and limits the number of track pairs to be considered in a natural way.

### 5.8.1 Event Sample

The vertex search is targeting at decays where only one of the involved particles is found in the RoIs. An example is the semileptonic transition of a  $B^0$  to a  $D^{*-}$  (In the following, a reference to a particular charge state also implies its charge conjugate.):

$$B^0 \rightarrow D^{*-} (\rightarrow \bar{D}^0 \pi_s^-) l^+ \nu_l. \quad (5.22)$$

The decay (5.22) is of relevance for the determination of the CKM matrix element  $V_{cb}$  and of the  $B^0$ - $\bar{B}^0$  oscillation frequency. The branching fraction for the first decay with  $l = e, \mu$  is substantial and amounts to 9.2%. The subsequent decay of the  $D^{*-}$  to a charged pion occurs with a probability of 68.3% [52]. A principle sketch of the decay chain (5.22) is shown in Fig. 5.15. The neutral  $B$  meson has an average decay length of roughly 11 mm. The lepton created in the semileptonic  $B$  decay is found by the FLT and SLT if it falls in the geometrical acceptance of the detector and has sufficient  $p_T$ . The excited  $D^{*-}$  decays by virtue of the strong interaction and creates a charged pion in vicinity to the lepton. This pion track is not selected by the triggers but can be reconstructed by the TLT. Further charged tracks emerge close to the lepton-pion vertex when the  $\bar{D}^0$  decays almost immediately, e.g to  $\bar{D}^0 \rightarrow K^+ \pi^-$ . The average distance covered by the  $\bar{D}^0$  is about 4 mm.

The decay chain (5.22) is well suited for vertex studies. The spatial separation between the lepton from the  $B^0$  and a charged track, which is detectable in the SVD, is minimised since the  $D$  meson from the  $b \rightarrow c$  transition is in an excited state. A vertex can either be found with the pion from the  $D^{*-}$  or with one of the daughter particles of the  $\bar{D}^0$ . Only the latter option would be available when replacing the  $D^{*-}$  by a  $D^-$ .

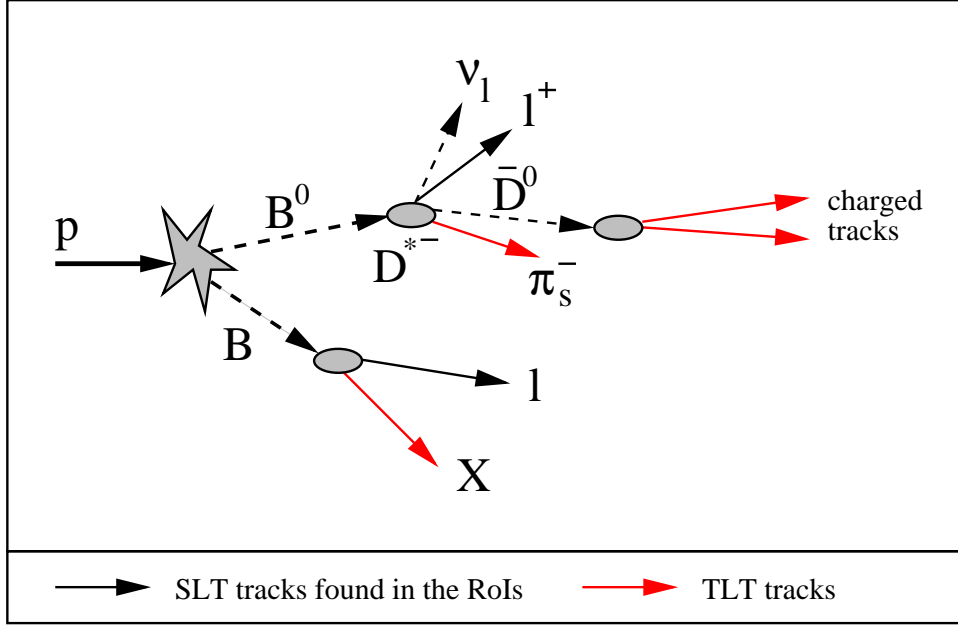


Figure 5.15: Sketch of the decay chain  $B^0 \rightarrow D^{*-}(\rightarrow \bar{D}^0 \pi_s^-) l^+ \nu_l$ . The  $B^0$  travels on average about 11 mm. The  $\bar{D}^0$  covers an average distance of about 4 mm. In the simulation, it decayed freely without any prescribed final state. The other  $B$  in the event (lower half of plot) decayed semileptonically to initiate a dilepton trigger. The charged lepton tracks from both  $B$  mesons are found by the FLT and SLT if they fall in the geometrical acceptance of the detector and have sufficient  $p_T$ . The TLT aims at reconstructing additional tracks from the decay chain of the two  $B$  mesons.

Table 5.6: Properties of the  $B$  and  $D$  mesons in the decay  $B^0 \rightarrow D^{*-}(\rightarrow \bar{D}^0 \pi_s^-) l^+ \nu_l$ .

Particle	$c\tau$ ( $\mu\text{m}$ )	$\langle \beta\gamma c\tau \rangle$ (mm)	$\langle p \rangle$ (GeV)	$\langle p_T \rangle$ (GeV)
$B^0$	468	11.2	131	2.4
$D^{*-}$	$\sim 0$	$\sim 0.0$	67	1.6
$\bar{D}^0$	124	4.2	63	1.5

The double-semileptonic  $B$  sample described in Section 5.5 contains always one  $B^0$  decaying according to (5.22) and was prepared with this study in mind. No restriction was imposed on the decay mode of the  $\bar{D}^0$ . The sample was used to study the physical properties of the particles involved in the decay (cf. Table 5.6) and to attempt the development of a suitable pattern recognition algorithm for vertices.

## 5.8.2 Vertex Resolution

The possible impact of a found vertex on the trigger decision for an event is given by its quality and position with respect to the target wire. The relevant quality criterion is the  $\chi^2$  of the vertex fit. The position of good-quality vertices will be required to be distinctly downstream of a target wire. This measure should reject vertices with a trigger lepton and a track coming directly from the primary vertex. What spatial separation one must demand, depends on the accuracy, which can be reached for the vertex position.

The attainable accuracy determines as well whether the fitted vertex position is a better handle than the impact parameter of the lepton to infer the finite lifetime of the  $B$  meson. A simple calculation can demonstrate what resolution in  $z$  is needed to compete with impact parameter cuts. According to Table 5.4, the impact parameter resolution for particles with a transverse momentum larger than 1 GeV is better than  $39 \mu\text{m}$ . The average impact parameter for highly relativistic  $B$  mesons is  $c\tau_B$  since the increase in the average decay length by a factor  $\beta\gamma$  is cancelled by the decrease of the angle under which the lepton is emitted with respect to the flight direction of the  $B$ . This angle is proportional to  $\gamma^{-1}$  [53]. The mean lifetime of neutral  $B$  mesons translates into a  $c\tau_B$  of  $468 \mu\text{m}$ . The corresponding impact parameter  $d$  is established as being different from zero with a significance of

$$\frac{d}{\sigma_d} = \frac{468 \mu\text{m}}{\sqrt{2} \cdot 39 \mu\text{m}} = 8.5. \quad (5.23)$$

The factor  $\sqrt{2}$  accounts roughly for the fact that the measured one-dimensional impact parameter is only a projection of the full impact parameter on the direction perpendicular to the target wire. A  $B$  meson moving along the  $z$ -axis and decaying exactly after one average lifetime gives rise to a vertex at a distance  $\beta\gamma \cdot c\tau_B = 11.2 \text{ mm}$  from the wire with the primary vertex. To reach the same level of significance as in (5.23) the resolution along  $z$  must be better than  $1.3 \text{ mm}$ .

The attainable resolution was studied for events where the  $\pi_s^-$  from the decay (5.22), or a charged particle from the  $\bar{D}^0$  (kaon, lepton, pion) was found by the TLT tracking. The assignment of tracks to vertices was extracted from the simulation and is referred to as *ideal vertex recognition*. The vertex positions were reconstructed by means of the VT-package and compared with the Monte Carlo. The result for the lepton-pion vertex is shown in Fig. 5.16 where the difference between the reconstructed coordinate and the Monte Carlo truth is plotted. The distributions have a non-Gaussian shape and long tails. Fitting a Gaussian to the central peak yields a resolution of about  $65 \mu\text{m}$  in the transverse plane ( $x$  and  $y$ ) and of about  $2.2 \text{ mm}$  in  $z$ . The longitudinal resolution is a factor of 5 worse than those of the  $J/\psi$  vertex in the decay  $B \rightarrow J/\psi(\rightarrow l^+l^-)$ . This effect has two reasons. The first is the low momentum of the slow  $\pi_s^-$  due to the limited phase space in the transition  $D^{*-} \rightarrow \bar{D}^0$ . The average pion momentum is only  $3.5 \text{ GeV}$ . The strong multiple scattering of such low-momentum particles prohibits a more precise determination of their track parameters. The second reason is again the missing momentum information for the  $\pi_s^-$ . If the vertex fit is repeated after the Monte

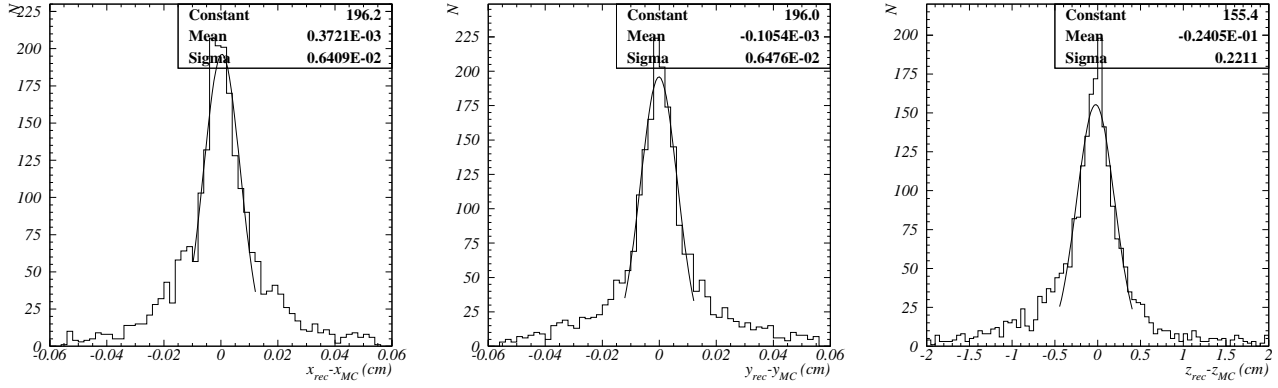


Figure 5.16: Accuracy for the reconstructed position of the lepton-pion vertex in  $x$ ,  $y$  and  $z$  (left to right). The difference between the reconstructed coordinate and the Monte Carlo truth is shown. The assignment of tracks to vertices was taken from the simulation. Fitting a Gaussian to the central part of the distributions yields a standard deviation of roughly  $65 \mu\text{m}$  in  $x$  and  $y$ . The resolution in  $z$  is about  $2.2 \text{ mm}$ .

Table 5.7: Kinematic properties of the charged particles in the transition  $B^0 \rightarrow D^{*-} (\rightarrow \bar{D}^0 \pi_s^-) l^+ \nu_l$  and in the subsequent decay of the  $\bar{D}^0$ . The transverse and longitudinal resolution  $\sigma_{x,y}$  and  $\sigma_z$  specify the accuracy for the position of the lepton-pion vertex which is reached when combining the trigger lepton  $l^+$  and the respective track in a vertex fit. (For illustration the vertexing was also applied to the  $e^+e^-$  pair from the decay of a  $J/\psi$ . Using ideal pattern recognition and the Kalman refit the results do *not* differ significantly when the momentum of the leptons is taken from the simulation or only estimated like for TLT tracks.)

particle	$\langle p \rangle$ (GeV)	$\langle p_T \rangle$ (GeV)	$\sigma_{x,y}$ ( $\mu\text{m}$ )	$\sigma_z$ ( $\mu\text{m}$ )
$B^0 \rightarrow D^{*-} (\rightarrow \bar{D}^0 \pi_s^-) l^+ \nu_l$				
$l^+$	31.9	1.73		
$\pi_s^-$	3.5	0.15	65	2200
$\bar{D}^0 \rightarrow$ charged particles				
kaons	17.5	0.84	50	1100
pions	10.3	0.54	60	1400
leptons	11.2	0.62	55	1100
$B \rightarrow J/\psi X \rightarrow e^+e^-X$				
electrons	32.8	1.64	30	450

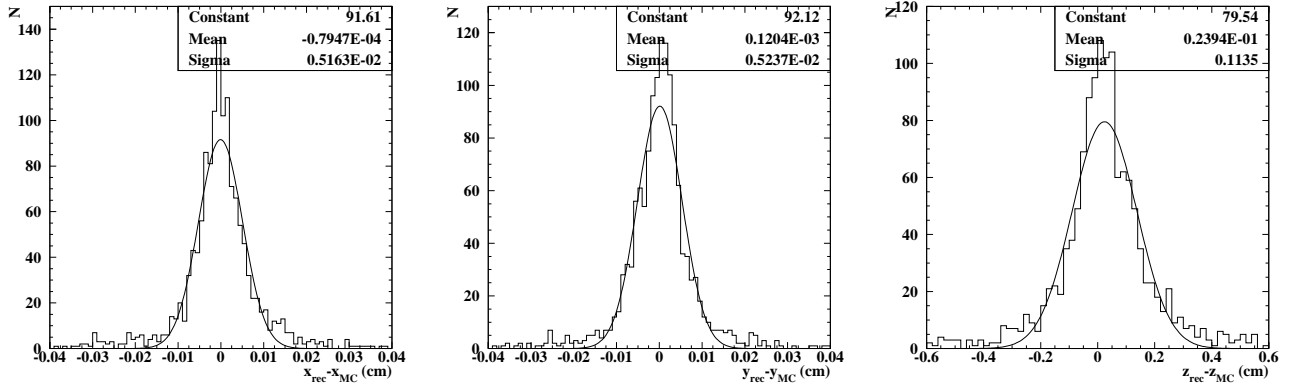


Figure 5.17: The position of the lepton-pion vertex is well reproduced by a vertex fit applied to the lepton  $l^+$  and a kaon from the decay of the  $\bar{D}^0$ . The plots show the difference between the reconstructed vertex position and the Monte Carlo coordinate of the lepton-pion vertex for  $x$ ,  $y$  and  $z$  (left to right). The distributions are narrower than in Fig. 5.16 due to the higher average momentum of the kaons of 17.5 GeV. The distribution in  $z$  appears to be slightly shifted to the right due to the distance travelled by the  $\bar{D}^0$  prior to its decay. The shift is much smaller than the longitudinal resolution of about 1 mm.

Carlo momentum of the  $\pi_s^-$  was used to determine its track parameters in the refit, the resulting longitudinal resolution is improved substantially and amounts to about 1.8 mm. Due to the limited resolution in  $z$  the  $\pi_s^-$  is not a very suitable handle when looking for secondary vertices. The situation is more relaxed for the charged daughter particles of the  $\bar{D}^0$  (cf. Table 5.7), which have higher average momenta. A vertex fit applied to the trigger lepton and one of these daughter particles reproduces the position of the lepton-pion vertex with better accuracy in spite of the distance covered by the  $D$  meson. For the abundant kaons from the decay of the  $\bar{D}^0$  (Fig. 5.17) the transverse resolution is about  $50 \mu\text{m}$ . The longitudinal resolution was found to be about 1 mm which is decisively better than that of the lepton-pion vertex itself. The finite lifetime of the  $\bar{D}^0$  provokes that the reconstructed  $z$ -coordinate is on average slightly larger than the value from the simulation. The magnitude of the shift is only  $(240 \pm 50) \mu\text{m}$  which is still small in comparison with the longitudinal resolution of roughly 1 mm.

### 5.8.3 Vertex Recognition

In spite of the encouraging resolutions for the lepton-kaon vertex, the recognition of secondary vertices was found to be a very difficult task. Vertex recognition was studied using ideal pattern recognition of tracks for the signal and background events. The two trigger particles served as starting points for the vertex recognition algorithm.

The following quantities can be exploited to select reconstructed tracks, which are subjected to a vertex fit with one of the trigger particles:

- The DCA of the trigger particle and the TLT track.
- Kinematic correlations of the two tracks. The kinematics of the  $B$  decay might restrict for instance the possible range of angles between the two candidate particles.
- The assignment of a track to a primary vertex.

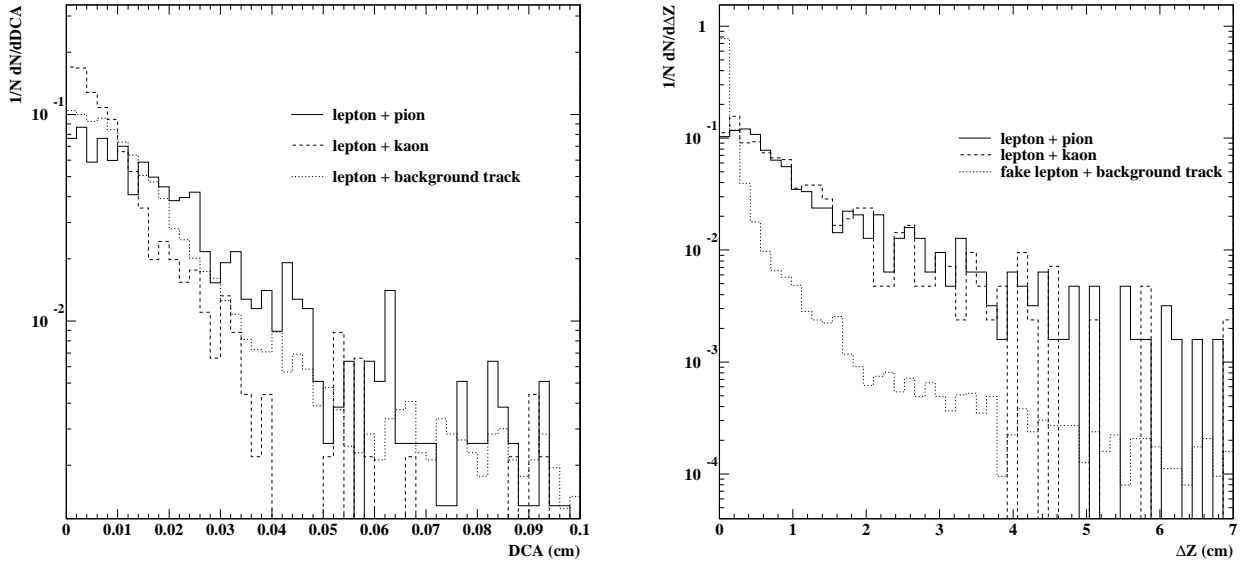


Figure 5.18: **(Left:)** Distributions of the reconstructed distances of closest approach of the lepton from the  $B^0$  and other tracks in the signal sample. The solid (dashed) histogram represents the case where the lepton is paired with the pion from the  $D^{*-}$  (with a kaon from the  $\bar{D}^0$ ). The dotted histogram shows the corresponding distribution when a track from the underlying event is used. **(Right:)** Positions of reconstructed vertices in the signal and the background sample. The shown quantity  $\Delta Z = z_{\text{vertex}} - z_{\text{wire}}$  is the difference of the vertex and the wire coordinate in  $z$ . Vertices were formed with all tracks having a DCA smaller than  $500 \mu\text{m}$ . The solid (dashed) histogram shows the vertices of lepton and slow pion (lepton and a kaon from the  $\bar{D}^0$ ) in the signal sample. The dotted histogram represents the distribution of reconstructed vertex positions in the background sample, where a fake lepton was paired with another track from the event.

- The quality of the reconstructed track.
- The impact parameter of the track.

A potential source of problems are tracks not related to the decay of the  $B^0$  but appearing accidentally close to the trigger lepton. Figure 5.18 shows on the left-hand side the reconstructed distance of closest approach between the lepton from the decay (5.22) and tracks from the decay cascade of the  $D^{*-}$ . The solid histogram is obtained for the DCA of the lepton and the slow pion. Taking into account all DCAs smaller than 1 mm the mean value of the distribution is  $206 \mu\text{m}$ . The dashed histogram represents the case that the lepton is paired with a kaon from the decay of the  $\bar{D}^0$ . Though separated from the lepton-pion vertex by the distance travelled by the  $\bar{D}^0$ , the average reconstructed DCA of  $105 \mu\text{m}$  is smaller than that of lepton and pion. This confirms that the trajectory of the kaon can be measured in the SVD to much better accuracy than that of the low-momentum pion. The dotted histogram shows the background of false combinations. The average DCA of the lepton and any other track not related to the decay of either  $B$  particle in the event is  $165 \mu\text{m}$ . This observation renders the extraction of candidates for a lepton-pion vertex rather hopeless when only the DCA is used.

A further study investigated in how far kinematic correlations between the lepton and the other tracks from the decay (5.22) can help to reject wrong combinations. For this purpose the

properties of track pairs given by the trigger lepton and either the  $\pi_s^-$  or a daughter track of the  $\bar{D}^0$  were scrutinised. Among these quantities were the angle between the tracks and the angle between either track and the flight direction of the  $B^0$ . The flight direction was estimated by the vector pointing from the closest primary vertex to the vertex formed by the two tracks. The same quantities were inspected for the case that the lepton was combined with a track coming directly from the primary vertex. The resulting distributions do not exhibit apparent discriminating features, implying that correlations between the trigger lepton and the other tracks are weak. This result is not unexpected since correlations with the trigger lepton are most likely washed out by the intermediate  $D^{*-}$  and  $\bar{D}^0$  decay.

It was also attempted to restrict the formation of vertex candidates to tracks which were not assigned to a primary vertex by the VT-package. This measure cuts down the number of candidates significantly but is not very productive since it excludes many of the tracks the search is targeting at.

The vertex recognition in its present shape does not try to exploit any correlations between tracks, nor does it obey the assignment of TLT tracks to primary vertices. No attempt is made to use only tracks having a minimal impact parameter. This is justified because both impact parameter and vertex position carry the same information about the lifetime of the decaying particle whose daughter tracks are detected. Vertices with tracks coming directly from the primary vertex will be rejected based on their position with respect to the target wire. The vertex recognition is applied to all SLT tracks sequentially. For each TLT track whose DCA to the lepton falls below  $500\ \mu\text{m}$  a secondary vertex is calculated. It is accepted if the  $\chi^2$  of the fit is smaller than a tuned cut value (5) and the vertex position is inside a window extending along the beam direction from  $z_{\min}$  to  $z_{\max}$ . The value of  $z_{\min}$  is given by the  $z$ -coordinate of the target wire, selected by the two trigger leptons, plus an offset  $\bar{z}$ . The offset can be varied to reject pseudo-vertices with a track from the underlying event in the signal sample or vertices found in the background events. The maximal accepted vertex position  $z_{\max}$  is 7 cm downstream of the target wire. Almost 99.9% of the  $B$  mesons decay before reaching this point.

### 5.8.4 Efficiency

Given the difficulties to disentangle true and false combinations of tracks into vertices, a variation of  $\bar{z}$  remains the only way to extract vertices with particles from a  $B$ . The right-hand plot in Fig. 5.18 shows the separation  $\Delta Z$  of the reconstructed vertex coordinate and the target wire for both signal and background events. The solid (dashed) histogram represents the case where a vertex of lepton and pion (lepton and kaon) is fitted. The dotted histogram results when the trigger tracks in the background sample are used for vertexing. There is a clear distinction between the vertex positions in signal and background events. A value of roughly 8 mm for  $\bar{z}$  suppresses pseudo-vertices in background events by almost two orders of magnitude. It is also seen that even in the background sample good-quality vertices are found several centimetres downstream of a target wire. The signal efficiency and the suppression of background events were studied as a function of the value of  $\bar{z}$ . For this purpose, all events were given to the TLT without imposing impact parameter cuts on the trigger leptons. For a given value of  $\bar{z}$  the efficiencies for both samples can be calculated from the percentage of events with one or more secondary vertices downstream of  $z_{\min}$ . Figure 5.19 (left) shows the result for the misidentification sample. The black triangles represent the efficiency for signal events as a function of the efficiency for background events when at least one vertex was required. To obtain a background suppression of roughly 10, all vertices closer than 1.3 cm to the target must be rejected. The

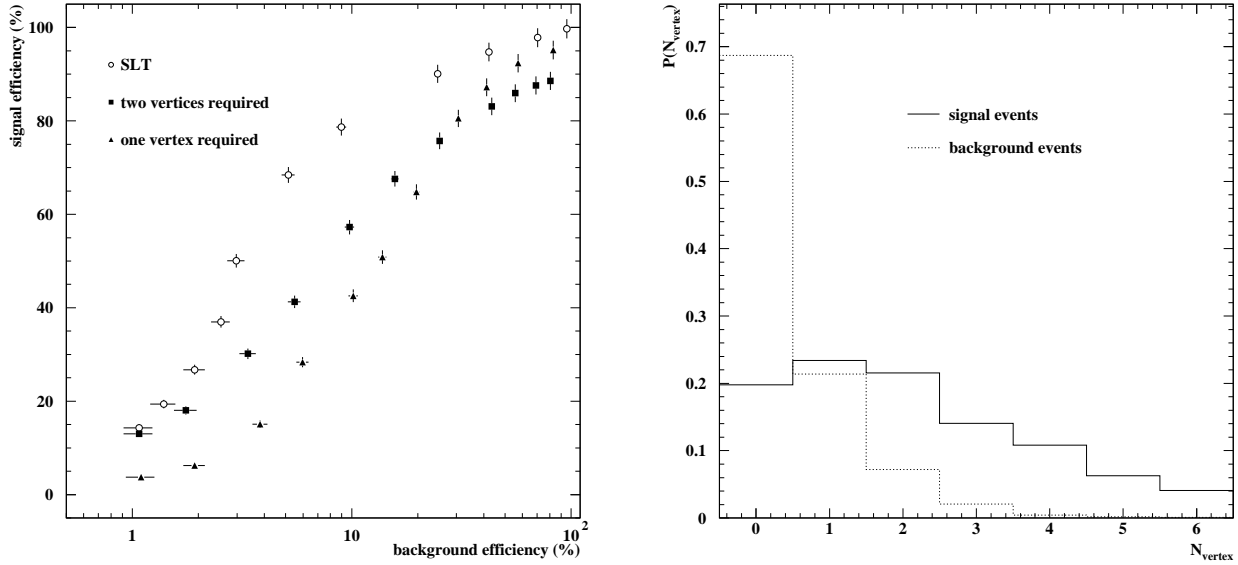


Figure 5.19: **(Left:)** Efficiency versus background suppression for the vertexing algorithm (black symbols). The parameter varied is the distance  $\bar{z}$  that a found vertex is required to be downstream of a target wire. The squares (triangles) denote the case where events with a least two found vertices (at least one found vertex) were accepted. The open dots denote the efficiency that the SLT would obtain by a cut on the one-dimensional impact parameters of the trigger particles. **(Right:)** Number of reconstructed vertices in signal (solid) and background (dotted) events. A value of  $\bar{z} = 4$  mm was used.

signal efficiency is then about 42 % for the vertexing algorithm. This efficiency is higher than expected from the pure exponential decay law for neutral  $B$  mesons. Using the average  $B^0$  decay length from Table 5.6, only 31 % of all particles are left after flight distances greater than 1.3 cm. The efficiency increase is due to reconstructed vertices combining a trigger lepton and tracks either from the underlying interaction or from the decay cascade of the other  $B$  meson in the event. The fraction of vertices of a trigger lepton and a charged track from the decay cascade of the same  $B$  rises from 40 % to 60 % when  $\bar{z}$  is varied from 0.5 to 5 cm. This implies that roughly every second signal event is selected based on a pseudo-vertex. Still, the number of vertices found in signal events is significantly higher than in background events as is demonstrated on the right-hand side of Fig. 5.19. When the longitudinal resolution of the vertex position is about 1.5 mm a value of  $\bar{z} = 4$  mm should reject a good fraction of the background events. With value of 4 mm for  $\bar{z}$  no vertex is found in 69 % of the background events but only in 20 % of the signal events. The presence of fake vertices suggests to accept only events with more than one vertex. The black squares in Fig. 5.19 (left) denote the resulting signal efficiency. A factor of 10 background suppression is accomplished by placing a cut on  $\bar{z}=4$  mm. The signal efficiency is then about 57 % and therewith higher than in the case where only one vertex was required in order to reach the same background suppression. Except for very moderate cuts on  $\bar{z}$ , the requirement of at least two vertices results in a better signal efficiency than the demand for only one vertex.

Unfortunately, the vertexing algorithm in its present state cannot compete with cuts that the SLT would place on the trigger particles. The white dots in Fig. 5.19 (left plot) specify the efficiency of a cut on the quantity  $r = \sqrt{d_1^2 + d_2^2}$ , the  $d_i$  being the one-dimensional impact parameters of the two trigger particles. Such a cut is superior to the search for secondary vertices for all values of the background suppression.

## 5.9 Track Trigger

The preceding section has shown that the search for secondary vertices at the TLT does not result in a better efficiency for signal events than cuts on the impact parameters of the trigger particles. Quantities, which can be extracted easier than vertices, are the impact parameters with respect to the target wire or the primary vertex. The TLT can try to refine the SLT decision by including the information from the additional tracks. The trigger decision will be based on tracks with large impact parameters, which are more likely to come from the decay cascade of  $B$  mesons. The algorithms used are referred to as track triggers.

### 5.9.1 Strategy of the Algorithm

There are various ways to detect the presence of additional tracks from the decay chain of the  $B$  mesons. A possible requirement might be, that every trigger particle is accompanied by at least one TLT track with a larger impact parameter than the trigger particle itself. Such a rule has the disadvantage that it does not provide a flexible way to control the output rate of the trigger. The strategy is therefore to find a discriminator function

$$S = \mathcal{F}(d_1, \dots, d_N; \sigma_1, \dots, \sigma_N; \dots),$$

which depends on the impact parameters of all  $N$  reconstructed tracks and their respective errors. Different other quantities, like track quality criteria, should be included as well to

suppress the influence of ghost tracks and particles, which suffered severely from multiple scattering. The discriminator function returns a single number  $S$  which specifies to what extent the event in question resembles the signal or the background. The cut quantity  $S$  should change smoothly on an interval when going from signal to background events. The triggering is done by accepting all events with  $S > S_{\text{cut}}$  in case a large value for  $S$  denotes a likely signal interaction. The value of  $S_{\text{cut}}$  is then adjusted such that the output rate of the trigger is manageable. The requirement of two additional tracks with large impact parameters, for example, can then be rephrased in such a way, that there are at least two tracks besides the trigger particles with an impact parameter above a tunable threshold.

## 5.9.2 Track Trigger Algorithm

This study aimed at finding a discriminator function  $\mathcal{F}$ , which is particularly suitable to distinguish between signal and background events. Facing the results of Section 5.7.3, the one-dimensional impact parameters of the trigger particles and the TLT tracks are used as arguments for  $\mathcal{F}$ . Since the impact parameter resolutions vary widely, it seems appropriate to base decisions on the the impact parameter divided by its error rather than on its absolute value. Badly reconstructed tracks with a high  $\chi^2$  from the track fit and particles with very large impact parameters, which cannot be assigned unambiguously to a target wire, should be excluded from the decision. A possible *ansatz* for  $\mathcal{F}$  is

$$S = \frac{1}{N} \sum_{i=1}^N w_i \left( \frac{d_i}{\sigma_{d_i}} \right)^q. \quad (5.24)$$

The sum runs over all  $N$  tracks, which are left after basic quality cuts, and includes the trigger particles. Dividing by the number of involved tracks ensures that the multiplicity is taken into account in the trigger decision. The  $w_i$  are individual weights, which reflect the quality of the track  $i$ . The exponent  $q$  is a free parameter.

The most obvious choice is to set  $q=2$  and  $w_i=1$  for all  $i$ . The quantity  $S$  is then the sum of the normalised squared one-dimensional impact parameters divided by  $N$ . It should be compatible with unity for inelastic events where all tracks come directly from the primary vertex. In signal events, larger values for  $S$  are expected due the presence of particles with extended lifetime. This applies only when the error of the impact parameters  $\sigma_{d_i}$  can be estimated properly which is not the case for the TLT tracks (cf. Fig. 5.12). In any case the efficiency for the signal and the background sample can be extracted by calculating the fraction of events with a value for  $S$  above some cut.

Our studies indicate that a value of  $q=1$  results in a better efficiency for signal events at the same suppression of background events. Further improvements are possible when tracks with relatively small impact parameters are excluded from (5.24). These tracks would only smear out the classification of signal and background events. The track trigger algorithm skips tracks with impact parameters smaller than some  $d_{\text{min}}$ . A good choice for  $d_{\text{min}}$  is the smaller of the impact parameters of the two trigger particles. Likewise, it is advisable to reject tracks with very large impact parameters. To exclude the track with the largest contribution to (5.24) deteriorated the efficiency for signal events. An appropriate measure is, however, to ignore tracks with impact parameters greater than 1 mm. Most of the tracks from the decay cascade of  $B$  mesons have smaller impact parameters, but this cut removes ghost tracks and trigger particles in the background events, which stem from hadrons decaying in flight.

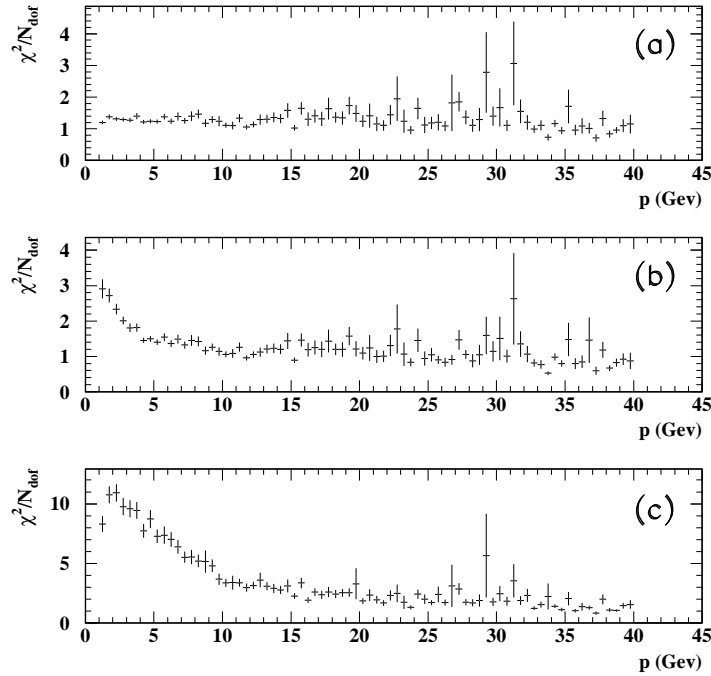


Figure 5.20: Average  $\chi^2$  per number of degrees of freedom versus the momentum of the fitted particle. Plot (a) results from a Kalman fit when the momentum of the particle is known. Plot (b) represents the case where the momentum approximation (5.17) was used in the Kalman fit. The last plot (c) results from a global straight-line fit. The error bars specify the uncertainty of the mean value, not the width of the distribution.

The quality weights in (5.24) were set to

$$w_i = \frac{1}{\sqrt{\chi^2/N_{\text{dof}}}},$$

where the  $\chi^2$  is calculated from a straight-line fit to the hits of the  $i$ -th track. The number of degrees of freedom  $N_{\text{dof}}$  is the number of hits of the track minus the dimension of the track parameter vector, which is four. The reasoning for this proceeding is the following. The proper evaluation of the error of the impact parameters suffers from the unknown momentum of TLT the tracks. On the other hand, it is advisable to give less weight to low-momentum tracks, which are more susceptible to multiple scattering. In principle, tracks which experience strong multiple scattering can be identified by detecting the deviation of the particle trajectory from a straight line. The  $\chi^2/N_{\text{dof}}$  of the Kalman fit is not suitable for this purpose, since the hits do not contribute uniformly to the  $\chi^2$ . Moreover, the Kalman fit allows to accomodate particle trajectories which are not straight lines. When the proper momentum is used in the Kalman fit, the  $\chi^2/N_{\text{dof}}$  has no dependence on the momentum (Fig. 5.20 (a)). The momentum approximation (5.17) used for TLT tracks introduces a dependence of the  $\chi^2/N_{\text{dof}}$  on the momentum as can be seen from Fig. 5.20 (b). The average  $\chi^2/N_{\text{dof}}$  of the Kalman fit is then greater than unity for momenta below roughly 10 GeV. This behaviour is even more pronounced for a straight-line fit (Fig. 5.20 (c)), which does not take into account multiple scattering. The  $\chi^2/N_{\text{dof}}$  of a straight-line fit is therefore the best way to detect strong multiple scattering.

### 5.9.3 Efficiency

The application of the cuts developed in the last section results in the distributions of the quantity  $S$ , which are shown on the left-hand side of Fig. 5.21 for the signal and the two background samples. The solid histogram belongs to the signal sample, the dashed (dotted) histogram results from the misidentification (tracking error) sample. As expected, the signal events have larger values for  $S$ . The mean value of the distribution is 3.8. The misidentification sample features a very narrow distribution at small values for  $S$  with a mean value of 0.9. For the misidentification sample, already a cut at  $S > 3$  reduces this background by roughly two orders of magnitude. The tracking error sample takes an intermediate position between the other two. It contains always at least one badly reconstructed trigger track which is included in the sum (5.24). The distribution for  $S$  reaches therefore out to larger values than for the misidentification sample and has a mean of 1.3.

The efficiency for signal and background events was calculated by varying  $S_{\text{cut}}$ . The result is displayed separately in Fig. 5.22 for the misidentification (left plot) and the tracking error sample (right). The black symbols specify the signal efficiency as a function of the suppression of the respective background sample. The black dots denote the case where ideal pattern recognition was used. The curve with the black triangles results when CATS provides the assignment of hits to tracks. The efficiency obtained using CATS comes close to the efficiency which is possible for ideal pattern recognition. When the TLT tracking is used instead, the resulting efficiencies are roughly 10% smaller. (They are not shown in Fig. 5.22 for clarity.) The plots confirm that the tracking error sample is a more severe background than the misidentification sample. At a suppression of about 10 for events from the misidentification sample, more than 75% of the signal events are kept. At the same suppression the signal efficiency is only 55% when the tracking error sample is used instead. For comparison with a scenario without the TLT, it has been assumed again that the SLT imposes a cut on the quantity  $D = \sqrt{d_1^2 + d_2^2}$  where the  $d_i$

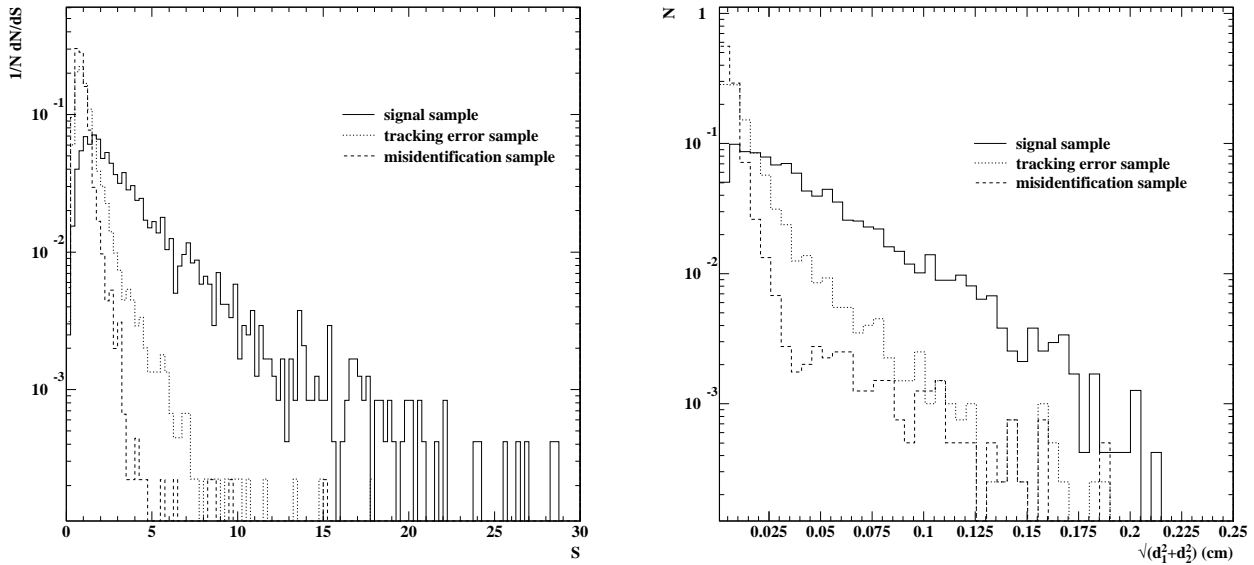


Figure 5.21: Distributions of the cut parameter  $S$  (left-hand plot) and the quantity  $\sqrt{d_1^2 + d_2^2}$  where the  $d_i$  are the one-dimensional impact parameters of the two trigger particles (right-hand plot). All histograms were scaled to represent the probability distribution for the shown quantities. The solid histogram results for the signal sample, the dashed (dotted) histogram shows the distribution for the misidentification sample (the tracking error sample).

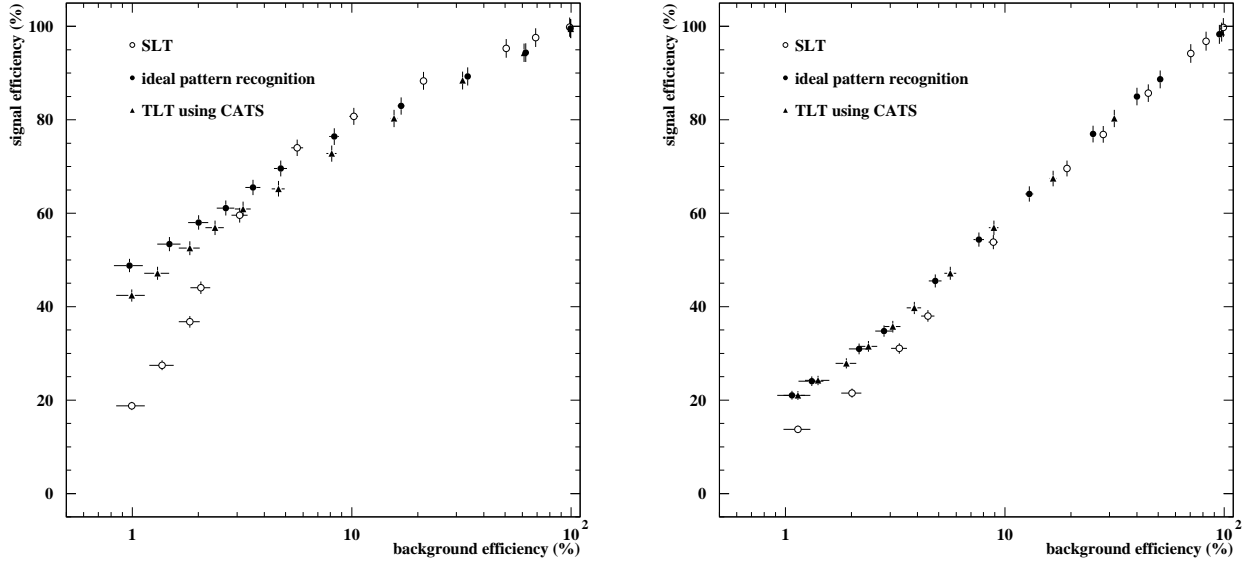


Figure 5.22: Efficiency for signal events versus suppression of background events when using the track trigger algorithm (black symbols). The curve with the black dots (triangles) is obtained with ideal pattern recognition (with CATS for tracking). For comparison, the efficiency obtained by the SLT using cuts on the impact parameters of the trigger particles is shown as well (white dots). For the figure on the left (right) the misidentification (tracking error) sample was used as background.

are the one-dimensional impact parameters of the two trigger particles. The resulting signal efficiency is displayed in Fig. 5.22 as well (white dots). The shape of the resulting curves differs strongly for the two background samples. For the misidentification sample (left), it is seen that the signal efficiency decreases moderately as long as a large fraction of the background events is kept. At suppressions larger than about 20, the drop of the signal efficiency is steeper than before. For the tracking error sample in turn (right), the signal efficiency decreases at a constant rate when the background efficiency goes down exponentially. This behaviour is a consequence of the different impact parameter distributions for the trigger particles in the two background samples. When they are combined into  $D$  the resulting distributions differ strongly (right plot in Fig. 5.21). For the tracking error sample, large values of  $D$  are suppressed exponentially. In contrast the corresponding distribution for the misidentification sample has a pronounced tail towards large values for  $D$ . This tail results from events where at least one trigger track comes from a hadron decaying in flight. The tail develops at values for  $D$  of about  $300\ \mu\text{m}$  and provokes the change of slope in the left plot of Fig. 5.22.

The most important feature in Fig. 5.22 is, however, that the tracks found by the TLT improve the efficiency for signal events. For moderate cuts imposed on the impact parameters of the trigger particles (high background efficiency) the signal efficiency obtained by the SLT alone and the TLT is almost the same. In this region, the trigger particles have small impact parameters and this should also apply to the additional tracks from the decay cascade of either  $B$  meson. The impact parameters of the additional tracks are of the same size as for particles from the underlying event. The quantity  $S$  is therefore smeared out by the relatively large number of contributing tracks. In fact, the efficiency obtained using  $S$  is slightly smaller than for cuts on the impact parameters of the trigger particles. Only for events with increased decay lengths of the  $B$  mesons and subsequently larger impact parameters of the daughter particles the TLT tracks have a significant influence. The possible increase in efficiency depends on the background sample. For the tracking error sample the gain is 5–10%, for the misidentification sample it can be as high as 30%.

## 5.10 Conclusion and Outlook

We have presented studies for a conceivable approach to a Third Level Trigger. The main characteristics of the trigger is its restriction to the silicon vertex detector data. The study of trigger applications was focussed on signal events where two  $B$  mesons decay semileptonically.

It has been shown that it is possible to find all tracks and primary vertices in the SVD, and to process trigger algorithms within a time slot of 100 ms. An efficient implementation of the Kalman filter method serves to determine the track parameters at the target. In spite of the missing momentum information, reliable estimates of the track parameters can be obtained when using a momentum approximation and a dedicated method for taking into account the material distribution in the detector. The found tracks can serve as input for trigger algorithms allowing to distinguish between signal and background events.

### 5.10.1 Algorithms

The track fit as implemented in the TLT software allows to determine the impact parameters of particles with respect to the target wire with good accuracy. The missing momentum information results in a deterioration of the impact parameter resolution and prevents the exact

calculation of the errors of measured impact parameters. The influence of both effects is still acceptable for trigger applications.

The determination of the primary vertices allows to extract the full impact parameter of particles as the distance of closest approach of track and vertex. It was found, however, that the usage of the primary vertex does not result in a significantly better efficiency for signal events than cuts imposed directly on the separation between track and target wire.

A trigger based on the presence of secondary vertices in double-semileptonic events of the type (5.22) proved not to be a promising option. Difficulties in the recognition of real vertices in combination with a resolution of 1–2 mm in the beam direction make such an approach less efficient than impact parameter cuts applied directly by the SLT to the leptons from both  $B$  hadrons.

According to our studies, a TLT considering only SVD information can contribute most to data-taking at an increased efficiency for signal events by finding more tracks with non-vanishing impact parameters. These tracks can serve together with the SLT tracks to refine the trigger decision. The gain is largest when the SLT is forced to reject events with a comparatively small combined impact parameter.

### 5.10.2 Further Studies

The description of the tracking algorithm in Ref. [35] and this thesis are only a first approach to conceivable algorithms for higher level triggers. The ongoing completion of the detector and the trigger chain up to the SLT will make the high-statistics data samples needed for trigger studies available. TLT studies will become more realistic when the output of the SLT can be inspected for different trigger settings ( $J/\psi$  trigger, dilepton trigger). It will become possible to investigate the quality of the SLT tracks, which were responsible for the trigger decision, and to include this knowledge into the TLT algorithms.

The track trigger algorithm (Section 5.9) is a first attempt to separate signal and background events. It should be worthwhile to consider more elaborated methods making trigger decisions based on the properties of reconstructed tracks. One might think of feeding the  $\chi^2$  of each track, its impact parameter, the error of the impact parameter, and other associated quantities into a high-dimensional decision function such as a neural network. The application of such smoother cuts can help to keep a larger fraction of the signal events at the same suppression for background processes.

### 5.10.3 Technical Implementation of the TLT

The existing TLT software is a suitable framework for applications on the SLT/TLT farm and for trigger studies. In its present implementation, it can either use the TLT tracking or CATS for pattern recognition. In the studies presented in this thesis, CATS has been used as is without striving for a deeper understanding of its properties (e.g. the increased ghost rate in comparison with the TLT track finding). On the long run it seems advisable to use CATS for pattern recognition. This is motivated by its superior track finding efficiency and its widespread application for offline analysis. Once CATS was made to function within the SLP, all experience from the data analysis can be carried over directly to the trigger system. For the track trigger algorithm, CATS has shown better performance than the original TLT tracking. We are confident that the TLT Kalman refit of tracks is the appropriate method to obtain the best possible estimate of track parameters lacking momentum information and the

full knowledge of the material distribution in the SVD.

# Bibliography

- [1] C.S. Wu et al., Phys. Rev. 105 (1957) 1413.
- [2] J.H. Christenson et al., Phys. Rev. Lett. 13 (1964) 138.
- [3] KTeV, A. Alavi-Harati et al., Phys. Rev. Lett. 83 (1999) 22.
- [4] NA48, V. Fanti et al., Phys. Rev. B 465 (1998).
- [5] P.F. Harrison and H.R. Quinn, editors, The BaBar Physics Book (SLAC-R-504, 1998).
- [6] L. Wolfenstein, Phys. Rev. Lett. 51 (1983) 1945.
- [7] M. Bargiotti et al., HERA-*B* Note 00-012 (2000).
- [8] D. Atwood et al., FERMILAB-PUB-96/448-T (1996).
- [9] BaBar, D. Boutigny et al., SLAC-R-95-457 (1995).
- [10] BELLE, KEK Report 95-1 (1995).
- [11] CDF, T. Affolder et al., Phys. Rev. D 61 (2000) 072005.
- [12] T. Lohse et al., DESY-PRC 94/02 (1994).
- [13] HERA-*B*, E. Hartouni et al., DESY-PRC 95/01 (1995).
- [14] E789, D.M. Jansen et al., Phys. Rev. Lett. 74 (1995) 3118.
- [15] E771, T. Alexopoulos et al., Phys. Rev. Lett. 82 (1999) 41.
- [16] T. Lohse, HERA-*B* Note 97-085 (1997).
- [17] M. Langer, Bestimmung der Beauty-Quantenzahl bei HERA-*B* mit Hilfe Neuronaler Netze, Master's thesis, Humboldt-University, Berlin, 1998, in German.
- [18] S. Spratte, private communication.
- [19] R. Mankel, Nucl. Instr. and Meth. A 395 (1997).
- [20] V. Saveliev, Nucl. Instr. and Meth. A 408 (1998) 289.
- [21] D. Rensing, Nucl. Instr. and Meth. A 384 (1996) 131.
- [22] T. Fuljahn et al., IEEE Trans. Nucl. Sci. 45 (1998) 1782.

- [23] M. Medinnis, Status Report: Slicer Performance, 1997, <http://www-hera-b.desy.de/subgroup/slt/notes/list.html>.
- [24] S. Erhan et al., HERA-*B* Note 96-102 (1996).
- [25] R. Mankel and A. Spiridonov, HERA-*B* Note 98-154 (1998).
- [26] A. Spiridonov, HERA-*B* Note 98-133 (1998).
- [27] S. Xella, The HERA-*B* Experiment: Preparation and first Results, PhD thesis, University of Copenhagen, 1999.
- [28] S. Schmidt, Second Level Trigger Studies at the HERA-*B* Experiment, PhD thesis, University of Copenhagen, 2000.
- [29] T. Fuljahn, Aufbau und Charakterisierung des schnellen Kontrollsystems für das Experiment HERA-*B*, PhD thesis, University of Hamburg, 1999, in German.
- [30] G. Wagner, private communication.
- [31] K.H. Sulanke, PCI-Bus to SHARC-Link Interface, <http://www-hera-b.desy.de/subgroup/daq/>.
- [32] F. Sun and P. Wegner, Higher Level Farm Event Control and Level 4 Farm Node ID Server, 1997.
- [33] U. Schwanke, The Remote Histogramming Package, <ftp://ftp.desy.de/pub/herab/daq/welcome.html>.
- [34] Subscription Service, <ftp://ftp.desy.de/pub/herab/daq/welcome.html>.
- [35] S. Scharein, Schnelle Spurerkennung für einen Level-3 Trigger im HERA-*B*-Experiment, PhD thesis, in preparation, in German.
- [36] H. Albrecht et al., ARTE. The event reconstruction and analysis tool for HERA-*B*, 1995, [http://www-hera-b.desy.de/subgroup/software/arte/arte\\_manual.ps](http://www-hera-b.desy.de/subgroup/software/arte/arte_manual.ps).
- [37] T. Lohse, HERA-*B* Note 99-098 (1999).
- [38] T. Sjostrand, Comp. Phys. Comm. 82 (1994) 74.
- [39] H. Pi, Comp. Phys. Comm. 71 (1992) 173.
- [40] J. Rieling, Entwicklung eines Silizium-Vertex-Detektors für das HERA-*B*-Experiment, PhD thesis, Ruprecht-Karls-University, Heidelberg, 1997, in German.
- [41] R. Mankel, HERA-*B* Note 95-239 (1995).
- [42] R.K. Bock et al., Data analysis techniques for high-energy physics experiments (Cambridge University Press, 1990).
- [43] D. Emelianov and I. Kisel, HERA-*B* Note 00-032 (2000).

- [44] W. Wagner, Auswertung der Daten des HERA-*B* Vertexdetektors im Hinblick auf die physikalischen Eigenschaften der verwendeten Siliziumstreifenzähler, PhD thesis, Ludwigs-Maximilians-University, Munich, 1999, in German.
- [45] T. Lohse, HERA-*B* Note 95-013 (1995).
- [46] R. Mankel, HERA-*B* Note 99-062 (1999).
- [47] I. Kisel and S. Masciocchi, HERA-*B* Note 99-242 (1999).
- [48] M. Medinnis, private communication.
- [49] A. Kharchilava, private communication.
- [50] J. Ivarsson, Potential for B-physics measurements with a fixed-target proton-collision experiment, PhD thesis, Lund University, 1999, LUNDFD6/(NFFL-7173).
- [51] T. Lohse, private communication.
- [52] C. Caso et al., The European Physics Journal C3 (1998).
- [53] W.B. Atwood and J.A. Jaros, Lifetimes, B decays, edited by S. Stone, pp. 364–394, World Scientific, Singapore, 1994.

# Appendix A

## Abbreviations

ARTE	HERA- <i>B</i> software environment for simulation and data analysis
CATS	A reconstruction software for the vertex detector
DCA	Distance of Closest Approach
DAQ	Data Acquisition
DSP	Digital Signal Processor
ECAL	Electromagnetic Calorimeter
FCS	Fast Control System
FED	Front End Driver
FLT	First Level Trigger
GEM	Gas Electron Multiplier
ITR	Inner Tracker
LHC	Large Hadron Collider
LUT	Lookup Table
MSGC	Micro-Strip Gaseous Chamber
OTR	Outer Tracker
PMT	Photo Multiplier Tube
RHP	Remote Histogramming Package
RICH	Ring Imaging Cerenkov Counter
RoI	Region of Interest
SHARC	Super Harvard Architecture
SL	Superlayer
SLB	Second Level Buffer
SLP	Second Level Process
SLT	Second Level Trigger
SVD	Silicon Vertex Detector
TCP/IP	Transmission Control Protocol/Internet Protocol
TLT	Third Level Trigger
4LT	Fourth Level Trigger
TRD	Transition Radiation Detector
UDP	User Datagram Protocol
VME	Versa Module Eurocard

# Appendix B

## Parameters of the TLT Tracking Algorithms

The TLT tracking is steered by several parameters which can be set using a configuration file. They are listed in the following in the order of their appearance in the algorithm.

### Spacepoint Reconstruction

Initial spacepoints candidates are created by combining two fired strips from two silicon strip layers in the same superlayer and sector whose stereo angles differ by just  $5^\circ$ . In the algorithm, the geometrical width of the search window (cf. Fig. 5.2 (b)) for crossing strips is enlarged by a small safety margin  $\Delta w$  in order to account for the unknown track angle. The size of  $\Delta w$  varies from superlayer to superlayer between 70 and  $300\ \mu\text{m}$ , corresponding to roughly 1 to 6 readout strips. The value of  $\Delta w$  decreases from superlayer 1 to 7 since the angle between the silicon chips and impinging tracks comes closer to  $90^\circ$ :

Superlayer	1	2	3	4	5	6	7
$\Delta w$ (cm)	0.0300	0.0180	0.0120	0.0120	0.0120	0.0100	0.0070

When looking for confirming hits in the rotated layers, the size of the search band  $\Delta b$  is a function of the distance between target and superlayer. Even within one superlayer this distance varies considerably. The value of  $\Delta b$  is therefore a function of both the superlayer and the sector:

Superlayer	1	2	3	4	5	6	7
$\Delta b$ (cm) in sector 1	0.44	0.28	0.22	0.15	0.14	0.13	0.11
$\Delta b$ (cm) in sector 2	0.22	0.16	0.13	0.15	0.14	0.13	0.11
$\Delta b$ (cm) in sector 3	0.44	0.28	0.22	0.15	0.14	0.13	0.11
$\Delta b$ (cm) in sector 4	0.22	0.16	0.13	0.15	0.14	0.13	0.11

### Track Reconstruction

In the track reconstruction, the target region is modelled as a box around the active wires. The box dimensions are calculated by inspecting the wire positions. The smallest and the largest wire coordinate is identified for  $x$ ,  $y$  and  $z$ , respectively. When these are denoted by  $x_{\min}$  and  $x_{\max}$  (and similarly in  $y$  and  $z$ ) the target box extends

- in  $x$  from  $x_{\min} - \Delta W/2$  to  $x_{\max} + \Delta W/2$ ,

- in  $y$  from  $y_{\min} - \Delta W/2$  to  $y_{\max} + \Delta W/2$  and
- in  $z$  from  $z_{\min} - \Delta Z_{\min}$  to  $z_{\max} + \Delta Z_{\max}$ .

The quantities  $\Delta W$ ,  $\Delta Z_{\min}$  and  $\Delta Z_{\max}$  are parameters of the algorithm with the following values:

$\Delta W$ (cm)	0.8
$\Delta Z_{\min}$ (cm)	2.0
$\Delta Z_{\max}$ (cm)	3.0

When extrapolating from superlayer to superlayer, the size of the search window for spacepoints in both  $x$  and  $y$  is determined by the angle  $\delta$  (cf. Fig. 5.3). In the Kalman refit applied to found tracks, the average transverse momentum is estimated to be  $\langle p_T \rangle$ . In the compatibility analysis of tracks, candidates are considered as new tracks when they have less than  $N_{\max}$  hits in common with any other track from the list of selected track candidates.

$\delta$ (rad)	0.002
$\langle p_T \rangle$ (GeV)	0.4
$N_{\max}$	6

# Lebenslauf

Name: Ullrich Schwanke  
geboren am: 23.07.1971  
in: Berlin  
Nationalität: deutsch

1988 – 1990 Abitur an der Erweiterten Oberschule "Albert Einstein"

10/1990 – 08/1991 Zivildienst

10/1991 – 04/1997 Studium an der Humboldt-Universität zu Berlin  
in der Fachrichtung Physik

10/1995 – 12/1996 Diplomarbeit "Aufbau und Durchführung von Testexperimenten  
mit Wabendrifthkammern für das HERA-*B*-Experiment"

07/1997 – 07/2000 Doktorand am Deutschen Elektronen-Synchrotron (DESY) Zeuthen,  
Arbeit in der Prozessorfarm-Gruppe des HERA-*B*-Experimentes

seit 08/2000 Wissenschaftlicher Mitarbeiter am DESY Zeuthen

# Selbständigkeitserklärung

Hiermit erkläre ich, die vorliegende Arbeit selbständig ohne fremde Hilfe verfasst zu haben und nur die angegebene Literatur und Hilfsmittel verwendet zu haben.

Ullrich Schwanke  
31. Juli 2000