

# DISSERTATION

## Application of Filter Methods to the Reconstruction of Tracks and Vertices in Events of Experimental High Energy Physics

ausgeführt zum Zwecke der Erlangung des akademischen Grades  
eines Doktors der technischen Wissenschaften

eingereicht an der Technischen Universität Wien  
Technisch-Naturwissenschaftliche Fakultät

von

Dipl.Ing. Rudolf Frühwirth  
geb. am 16.4.1952 in Wien

1040 Wien  
Schönburgstraße 50/20

Wien, im Oktober 1988

## Kurzfassung

Die rasante Entwicklung der experimentellen Hochenergiephysik in den letzten Jahren, die besonders auf dem Gebiet der Beschleuniger- und der Detektortechnik zu verzeichnen war, verlangt auch neue Methoden der Datenanalyse. Die vorliegende Arbeit befaßt sich mit zwei genau umgrenzten Teilaspekten der Analyse von experimentellen Daten: mit der Rekonstruktion von Spuren geladener Teilchen und mit der Rekonstruktion von Wechselwirkungsvertizes. Nach einer kurzen Einführung (Kapitel 1) werden in Kapitel 2 das experimentelle Umfeld, die wichtigsten Typen von Spurdetektoren und die Grundzüge der Analyse von Daten erläutert, die von diesen Detektoren produziert werden. In Kapitel 3 wird ein kurzer Überblick über herkömmliche Methoden der Spur- und Vertexrekonstruktion gegeben, und die Gründe für die Suche nach neuen Methoden werden erklärt. In Kapitel 4 wird gezeigt, daß Filteralgorithmen die Forderungen erfüllen, die von modernen Großexperimenten an die Spur- und Vertexrekonstruktion gestellt werden. Der Kalman-Filter, kombiniert mit dem Smoother nach Rauch, Tung und Striebel, ermöglicht die optimale Schätzung der Spurparameter in allen Meßpunkten auf zeitsparende Weise, auch für Spuren mit starker Vielfachstreuung. Dies eröffnet neue Möglichkeiten zur Behandlung von Ausreißern und zum Erkennen von Zerfällen, wie etwa der myonischen Zerfälle von  $\pi$ - und K-Mesonen. Ferner wird gezeigt, daß auch das herkömmliche Schätzverfahren zur Vertexrekonstruktion als Kalman-Filter interpretiert werden kann. Daraus ergibt sich die Möglichkeit, durch die Betrachtung von Sekundärspuren als Ausreißer Methoden zur Auffindung von Sekundärvertizes zu entwickeln. Außerdem werden robuste Varianten des Kalman-Filters zur Spur- und Vertexrekonstruktion vorgeschlagen, die den nach dem Prinzip der kleinsten Fehlerquadrate arbeitenden Kalman-Filter ergänzen können.

Die Anwendung der in Kapitel 4 entwickelten Filtermethoden in einem Großexperiment werden in Kapitel 5 präsentiert. Es handelt sich um das Experiment DELPHI am LEP-Speicherring im CERN (Genf), das Mitte nächsten Jahres in Betrieb gehen wird. Zunächst wird der Filteralgorithmus mit der klassischen Schätzmethode zur Spurrekonstruktion verglichen, und der Gewinn an Genauigkeit, der durch die Anwendung des Smoothers erzielt wird, demonstriert. Dann werden die vorgeschlagenen Methoden zur Behandlung von Ausreißern an Hand von simulierten Daten im zentralen Spurdetektor von DELPHI überprüft und verglichen. Die Verfahren zur Erkennung von Zerfällen werden im zentralen Spurdetektor untersucht, und zwar ohne und mit Zuhilfenahme der übrigen Spurdetektoren im Zentralbereich von DELPHI. Schließlich werden auch die zur Erkennung von sekundären Zerfallsvertizes entwickelten Algorithmen an Hand von simulierten Ereignissen überprüft und verglichen.

# Table of contents

1	INTRODUCTION	1
2	DATA ANALYSIS OF HIGH ENERGY PHYSICS EXPERIMENTS	3
2.1	The aims of high energy physics	3
2.2	The experimental environment	5
2.2.1	Detector systems	5
2.2.2	Tracking detectors	8
2.2.3	The DELPHI detector	11
2.3	The stepwise analysis of experimental data	13
2.3.1	Triggering, digitization and on-line selection	14
2.3.2	Pre-processing, pattern recognition and track search	14
2.3.3	Track reconstruction	15
2.3.4	Vertex reconstruction	16
2.3.5	Kinematic fit and particle identification	17
2.3.6	Event viewing	17
2.3.7	Data analysis software	18
2.4	Required properties of analysis methods	18
	Figures	20
3	TRADITIONAL METHODS OF TRACK AND VERTEX RECONSTRUCTION	28
3.1	Why least-squares?	28
3.2	Least-squares estimation of track parameters	29
3.2.1	The track model	31
3.2.2	Description of noise terms	31
3.2.3	The global method of estimation of track parameters	34
3.2.4	Test statistics	36
3.3	Estimation of vertex parameters	36
3.4	Some experimental results	39
3.5	Limits of traditional methods	39
	Figures	42

4	APPLICATION OF FILTER TECHNIQUES TO TRACK AND VERTEX FITTING	44
4.1	Properties of the Kalman filter	45
4.2	Estimation of track parameters with the Kalman filter	50
4.2.1	The track as a discrete dynamic system	50
4.2.2	The system and the measurement equation	50
4.2.3	The estimation procedure	51
4.2.4	Merging of track segments	51
4.2.5	Removal of measurements	53
4.3	Treatment of outliers in track reconstruction	54
4.3.1	Origin and modelling of outliers	54
4.3.2	Identification of outliers by chi-square tests	56
4.3.3	A robustified Kalman filter	59
4.4	Recognition of decays (kinks)	63
4.4.1	Physical motivation	63
4.4.2	A kink finding algorithm	65
4.5	Estimation of vertex parameters with the Kalman filter	67
4.5.1	The interaction vertex as a discrete dynamic system	67
4.5.2	The estimation procedure	68
4.6	Detection of secondary vertices	70
4.6.1	Detection of outlying tracks by chi-square tests	72
4.6.2	A robust vertex fit	73
	Figures	76
5	FILTER TECHNIQUES IN THE DATA ANALYSIS OF DELPHI	81
5.1	The DELPHI program chain for data analysis	81
5.1.1	Simulation	81
5.1.1	Data analysis	81
5.2	Performance studies of filter algorithms	83
5.2.1	Results of track reconstruction algorithms	83
5.2.2	Results of outlier studies	87
5.2.3	Results of kink finding studies	92

5.2.4 Results of secondary vertex detection studies	96
Tables	100
Figures	105
6 CONCLUSIONS	142
ACKNOWLEDGEMENTS	143
FIGURE CAPTIONS	144
TABLE CAPTIONS	153
REFERENCES	154

# CHAPTER 1

## INTRODUCTION

High energy physics is a fascinating subject. Researchers in this field set out to explain the fundamental structure of our universe. To this end it is necessary to study the smallest constituents of matter, the so-called elementary particles, the ultimate goal being a unified mathematical description of all particles and of the forces acting between them [1].

As in every branch of natural sciences, research is a continuous interaction between theory and experiment. In fact, experiments are frequently carried out in order to refute or to confirm certain theoretical predictions, but they also stimulate theoretical development by the discovery of unexpected and unexplained effects. In order to achieve the ambitious goal of a unified description of matter, highly sophisticated experimental techniques have to be applied. But a well-designed and well-working experimental apparatus is not enough. Equally important is the proper choice of methods of data analysis which extract the largest possible amount of information from the data recorded by the experiment. In view of the fact that a large high energy physics experiment costs tens of millions of dollars and involves several hundred researchers, engineers, technicians and programmers for several years, it cannot be justified if information is lost due to insufficient algorithms or sloppy analysis.

On the other hand, the huge amount of data collected by modern experiments imposes serious constraints on the data analysis programs, as far as computing time is concerned. With millions of events to be processed, and at least an equal number of events to be simulated, only a few seconds of computing time on a large mainframe computer can be spent per event. This inevitably restricts the choice of algorithms to fast and relatively simple procedures.

Other constraints are imposed by the social and psychological environment. The data analysis code has to be accepted, read, run and possibly modified by a large number of people, most of them non-specialists in the field of statistical data analysis. Therefore the methods employed have to be flexible, transparent and easily understandable to an experimental physicist. The last few years have brought a remarkable increase of capability, but also of complexity of the detector systems used in high energy physics experiments. In addition, due to progress in accelerator techniques, the energy scale of the processes under study has undergone a dramatic rise by at least a factor of ten during the last 8 years. As a consequence, events are far more complex and much more difficult to analyse than they used to be.

Considering these developments, one should not be surprised if some of the traditional methods of data analysis turn out to be insufficient for the new generation of experiments to be carried out at the new Large Electron-Positron accelerator (LEP). This huge circular accelerator, at present the largest among electron-positron machines, is currently under construction at the European Laboratory for Particle Physics (CERN) in Geneva. It is scheduled to be operational in summer of 1989. In order to keep pace with the development of the hardware, new algorithms have therefore to be developed. It is the aim of this work to present a new approach to two well-defined topics of data analysis: reconstruction of charged tracks and reconstruction of interaction vertices. We will show that filter techniques are particularly well adapted to fast, efficient and nevertheless robust estimation of track and vertex parameters, in the context of large modular detectors and complex multitrack events. In fact, one of the large experiments at LEP, DELPHI, will use this kind of methods in its data analysis programs.

The layout of the thesis is the following: In Chapter 2 first the experimental environment and the most important types of tracking detectors are described. Then the problems to be solved in data analysis and the basic requirements to the algorithms employed are formulated. Chapter 3 is a brief survey of commonly used methods of track and vertex fitting. The reasons for the search of new methods are explained. Chapter 4 shows how filter techniques can be applied to the estimation of track and vertex parameters. This approach also opens a new road towards the solution of problems which frequently arise in this context: the treatment of outliers, the recognition of kinks (charged 1-prong decays) and the detection of secondary vertices. In Chapter 5 we shall discuss the implementation of filter techniques in the data analysis program of DELPHI and investigate the performance of the proposed algorithms with simulated data. In Chapter 6 we will draw the conclusions and give some hints on possible future lines of research.

# CHAPTER 2

## DATA ANALYSIS OF HIGH ENERGY PHYSICS EXPERIMENTS

### 2.1 The aims of high energy physics

As stated in the introduction, the ultimate goal of high energy physics is the explanation of the fundamental structure of matter, and hence of our universe. Of course, at present one is still very far from a complete, unified description of the basic constituents of matter and of the forces acting between them; it is not even clear whether this goal can ever be reached at all. In order to understand the behaviour of the elementary particles one needs not only physical theories but also experiments which both serve as tests of the existing theories and at the same time try to find novel effects, thus provoking the development of new theories.

In the experiment the structure of matter is studied on the smallest possible scale. Due to the laws of quantum physics, in particular due to Heisenberg's uncertainty principle, probing matter at small distances requires small wavelengths or, equivalently, high energies. This can be understood by analogy to the optical microscope in which the resolution is limited by the wavelength of visible light, which is in the order of 500 nm. In an electron microscope the quanta of visible light (photons) are replaced by electrons accelerated to an energy of a few keV, about a factor 1000 larger than the energy of visible photons. According to the basic law that the energy is proportional to the frequency,  $E = h\nu$ , the electrons can be attributed wavelengths smaller by a factor 1000 than those of visible light. Therefore the electron microscope can be used to probe the structure of a crystal. If one compares the distance of two atoms in the lattice of a crystal ( $\approx 10^{-10}$  m) with the size of an atomic nucleus ( $\approx 10^{-15}$  m) one sees that particles with an energy of at least a few 100 MeV have to be used to study the nuclear structure. Since the aim of high energy physics is the understanding of the constituents of protons and neutrons (the quarks) and of the leptons (electrons, muons, neutrinos) the probing particles must have an energy of at least tens of GeV.

These highly energetic particles are obtained from particle accelerators. In such a machine, strong electric fields act on charged particles (in practice electrons or protons) to accelerate them to the required energy. Since in a linear accelerator the energy is limited by the length of the machine, the more widespread type is the circular accelerator in which the particles cross the accelerating field many times. Such machines are called **synchrotrons** [2].



A synchrotron can be operated in one of two different modes: In the first mode, the accelerated particles are extracted from the machine as soon as they have attained the required energy and are steered to the target, a block of matter, in which they undergo interaction with the material composing the target. This is called **fixed target** operation. Due to a kinematic effect not all of the total energy, but only a fraction proportional to its square root is available for the interaction itself; the remainder is consumed for boosting the center of mass of the beam-target system in the forward direction.

Therefore a second mode of operation has been invented, which uses the energy more efficiently: Two beams of particles, circulating inside a vacuum tube in opposite directions, are accelerated and kept in orbit during some hours. The beams are concentrated in a couple of bunches which collide at certain well-defined points along the machine. Now all of the total energy, which is the sum of the beam energies, is available for the interaction. Such a machine is called a **storage ring** or a **collider**.

As an example, the SPS (Super Proton Synchrotron) at CERN delivers protons with an energy of up to 450 GeV to fixed targets, but can also be operated as a proton-antiproton collider with an energy of 315 GeV per beam. The future large electron-positron collider LEP at CERN will initially have an energy of 50 GeV per beam, just enough to produce the recently discovered  $Z^0$  vector boson [3] in large numbers. In a later stage, after an upgrade, the energy will be 100 GeV per beam. In electron-positron colliders the energy is, however, limited by the losses due to synchrotron radiation. Therefore the latest trend is towards very long linear colliders.

In any experiment, the basic idea is to let an accelerated particle collide with a second particle, which is either at rest or itself accelerated, and to watch the outcome of the collision with a suitable apparatus, the **detector**. Such a collision is called an **event**; the event is the basic unit of data collection. Due to the statistical character of quantum mechanics the outcome of an individual event cannot be predicted; also, two events are independent in the stochastic sense. The possible outcome of an event varies over a large number of cases. In the simplest case the event may show elastic scattering of the two colliding particles (Fig. 2.1) with only two emerging particles. But, as we are in the domain of relativistic physics, the energy of the collision may also flow into the creation of new particles (Fig. 2.2). Many of these newly created particles are unstable and decay inside the detector, thereby creating another set of particles.

In order to describe completely the reaction of the two primary particles, it is necessary to determine **direction, momentum, charge and type** of all particles produced at the primary interaction vertex or at secondary decay vertices, as well as the location of all secondary decay vertices. If the reaction is sufficiently complicated it is nearly always impossible – and

frequently also unnecessary – to extract all of this information from the data recorded by the detector. The art of the experimenter consists to a large extent in designing the detector in such a way that the physically relevant information can be obtained with high probability and high precision.

## 2.2 The experimental environment

### 2.2.1 Detector systems

As indicated above, the main tasks of the experimental apparatus are:

- The measurement of the **momenta of charged particles**;
- The measurement of the **energy** (and possibly the direction) **of neutral particles**;
- The determination of the particle type or **particle identification**.

Due to the complexity of high energy collisions the apparatus consists of many detectors carrying out different types of measurements, so that it may be appropriately called a **detector system**.

The momentum of a charged particle is determined from a measurement of the curvature of the trajectory of the particle (its **track**) under the influence of a known stationary magnetic field. Therefore a detector system usually contains a **magnet**, sometimes superconducting, which is capable of supplying a field of sufficient strength. The trajectory of the charged particle can be determined from the trail it leaves in devices designed specially to this purpose, so-called **tracking detectors**. These are capable of detecting a charged track by means of secondary processes (mainly ionization) in the sensitive volume of the detector, which is filled with an appropriate gas. In order to compute the curvature, the position of the trajectory in space has to be measured in several points along the track. Thus a detector systems contains in most cases several tracking detectors. It has to be kept in mind, however, that a charged particle necessarily interacts with the material the detectors are made of. The resulting disturbance of the track has to be kept as small as possible by designing the innermost detectors, in particular the central tracking detectors, as light and "transparent" as their required mechanical strength permits. The basic types of tracking detectors will be discussed in Subsection 2.2.2. The combination of a deflecting magnet with one or several tracking detectors is sometimes called a **magnetic spectrometer** [4, 5].

As neutral particles, e.g. neutrons or photons, do not ionize matter significantly, they are detected in a destructive way by being absorbed in a block of dense matter, which is of suffi-

cient thickness, so that the incident particle deposits all its energy inside the detector volume in a cascade or shower of low-energy secondary particles. Such a device is called a **calorimeter**, although the heat produced by the absorption of a single particle is too small to be measured. Instead, one looks for another detectable signal, which should be a linear function of the energy of the incident particle. A frequently used calorimeter design consists of many layers of a dense material, such as iron, lead or uranium. Sandwiched between these are layers of sensing devices, usually scintillators or proportional chambers, which sample the energy deposited by the shower produced by the incident particle. Of course, the calorimeter is also sensitive to charged particles. Thus calorimeter information is frequently a useful complement to the information obtained from tracking detectors. A survey of calorimeter techniques is given in ref. [6, 7].

The layout of a calorimeter has to be optimized with respect to its purpose. So-called electromagnetic calorimeters are designed for the energy and position measurement of showers produced by electrons and photons, while the layout of hadron calorimeters takes into account the different characteristics of the showers produced by hadrons, e.g. protons, neutrons,  $\pi$ - or K-mesons.

Particles can be identified in several ways. The most important one is the measurement of the velocity of the particle. If the momentum is known in addition, the mass of the particle can be computed and the type can be inferred with a certain probability depending on the precision of the mass determination.

The velocity of a particle can be computed directly from a measurement of the time of flight between two counters with good time resolution; these are usually scintillation counters. It can also be determined from electromagnetic effects which depend on the velocity. Amongst these are [8]:

- Cherenkov radiation;
- Transition radiation;
- Energy loss due to the ionization of a medium.

Cherenkov and transition radiation are detected by devices specially designed to this purpose. Energy loss due to ionization can be measured by a tracking detector, provided it works in proportional mode and is equipped with the necessary electronics. As indicated above, information about the particle type can also be drawn from the properties of a shower in a calorimeter. If the particle is unstable and decays inside the detector, its type can also be inferred from an analysis of its decay products.

A special position, as far as particle identification is concerned, is occupied by muons. They distinguish themselves from all other kinds of charged particles by the fact that they are capa-

ble of penetrating several meters of a dense material, e.g. iron or lead. Hence the common technique of muon identification is the following one: A set of tracking detectors (**muon chambers**) is put behind an absorber of sufficient thickness (**muon filter**), so that only muons penetrate the absorber and are registered by the muon chambers. In a collider experiment the hadron calorimeter is normally used as the muon filter. Since in this case the muon chambers have to cover a large area of several hundred square meters, they are usually designed as drift chambers (see Subsection 2.2.2).

There are two basic types of detector systems, corresponding to the two possible modes of operation of the accelerator:

- Fixed target experiments, facing a host of energetic particles in the forward direction, require long magnetic spectrometers with high-precision tracking detectors and good two-particle resolution (Fig. 2.3). A fixed target spectrometer which employs many different techniques of particle detection is described in ref. [9].
- In collider experiments no particular direction is privileged by kinematics. Therefore a compact detector, covering the full solid angle is needed (Fig. 2.4). As the detector cannot be very large, due to limited space and money, a high magnetic field and precise tracking detectors are required. Again good two-particle resolution is a crucial feature, because of the high multiplicities produced in high energy collisions.

One of the most impressive aspects of modern detector systems is their sheer size. Let us consider for a moment the detector of the DELPHI collaboration [10] which will operate at LEP (Fig. 2.5). It has the shape of a cylinder about 10m wide and 10m long and weighs more than 3000 tons. It includes the largest superconducting coil built so far. The number of electronic channels required to read out all parts of the detector exceeds  $10^5$ . The total cost of the detector is in the order of 50 million dollars. A more detailed description of the detector system will be given in Subsection 2.2.3.

The design and construction of such a detector system is thus a major project. An important aspect of the design process is the optimization of the layout within the given limits of cost and space. The designing team has to make sure that the detector is capable of adequately detecting and recording the data arising from the events. Also, the layout and characteristics of magnets and tracking detectors will largely determine the momentum resolution which can be achieved and should be optimized accordingly.

It should be kept in mind that a detector not only provides information about a track, but that the track is also disturbed by the interactions with the material of the detector and of its infrastructure (supports, frames, vacuum vessels, etc.). Thus a massive detector in the wrong place can actually do more harm than good.

Another aspect of detector optimization is the interplay between hardware and software [11]. A detector designer should work in close contact with the people who choose the algorithms to be employed in the reconstruction of the events, and should even be ready to change his design accordingly. Otherwise the detector will be a continuous source of trouble during the analysis of the data.

These exceedingly complex optimization problems are solved by Monte Carlo methods. First events are generated randomly, resembling as closely as possible the processes likely to occur in the experiment. In a second step the response of the detector system is simulated in all details. In the third step, the event is reconstructed and the estimated quantities, like momenta and energies, are compared with the true values. This allows the designer to assess the performance of the hypothetical detector.

There exist specialized and standardized programs which perform the random generation of physics events [12] as well as the simulation of the detector response to a variable level of sophistication [13, 14]. With a large number of events the simulation may become quite lengthy, although in many cases the detailed simulation of a physical process can be replaced by a simple model giving statistically equivalent results [15] (see also Subsection 5.1.1). Therefore approximate formulae have been worked out for simplified layouts which give the designer a first idea of the resolution that may be expected from the detector [16].

### **2.2.2 Tracking detectors**

Only the most basic types of tracking detectors can be described here. For more details and references, see for instance refs. [17, 18].

#### **The multi-wire proportional chamber (MWPC)**

The invention of the MWPC was one of the milestones in the history of particle detectors [19]. The MWPC is based on the same principles as the well-known Geiger-Müller counter, invented already in 1928 [20], and is a clever development of this "single-wire chamber". In its simplest form a MWPC consists of a thin gas volume confined between two cathode planes. Midway between the cathode planes is an array of parallel signal wires which form the anode (Fig. 2.6). The chamber is put in such a position that most of the tracks cross it more or less perpendicularly.

A track passing the chamber ionizes the molecules of the gas and sets free electrons which are attracted by the nearest anode wire. In the vicinity of the thin wire the electric field is sufficiently high to accelerate the electrons to such energies that they in turn produce ionization

electrons. Finally an electron avalanche develops and produces a detectable electric signal in the wire. This signal is then amplified and registered by the electronics attached to the wire.

Under the assumption that only the nearest sense wire gives a signal, the uncertainty of the measured position of the track is distributed uniformly between  $-d/2$  and  $d/2$ , where  $d$  is the wire spacing, in practice between 0.5 mm and 3 mm. This uncertainty is in principle a deterministic function of the true track position, which, however, is itself random. Therefore the uncertainty of the measured coordinate may be regarded as a random variable. The standard deviation of the measurement error of a single coordinate measurement is then given by

$$\sigma = d/\sqrt{12}.$$

Thus the MWPC features reasonably good spatial resolution, and also good two-track separation and good multi-track efficiency. It allows high event rates and, although permanently at high voltage, is sufficiently fast to be used in a fast decision trigger [21] (see Subsection 2.3.1). Since a single MWPC measures only a single coordinate, nearly always several (at least three) MWPCs with different wire orientation are combined into a dense stack of chambers [22, 23, 24]. By a suitable linear combination of coordinates it is then possible to compute space points (and possibly directions). The error distribution of these approaches rapidly a normal distribution with increasing number of chambers (see also the Central Limit Theorem). The approximation is already quite good for only four chambers (Fig. 2.7).

The restriction that a single MWPC measures only a single coordinate can be overcome by picking up the signals which are induced in the cathode by the motion of the positive ions. By a suitable subdivision of the cathode into strips orthogonal to the wires or into small pads, two coordinates of the crossing point can be measured, the third one being given by the position of the chamber plane.

### **The drift chamber**

In MWPCs the attainable precision is limited by the wire spacing. Therefore a large MWPC which is to achieve good accuracy has a very large number of wires, each of which is equipped with its amplifier and other electronic devices. This leads to an unacceptably high power dissipation and also to tremendous costs. In order to overcome these constraints, the chamber is modified in the following way: The wire spacing is increased to a few centimeters, and the time lapse between the impact time of the particle and the arrival of the electron avalanche at the sense wire is measured to a precision of a few nanoseconds. Such a chamber is called a drift chamber [17, 25, 26] (Fig. 2.8).

Since in a suitable gas mixture the electron cloud drifts towards the anode wire at a constant speed of some 0.05 mm/nsec, the measured drift time is directly proportional to the distance

of the impact point from the wire. There is, however, an intrinsic ambiguity, since there is no information as to on which side of the wire the track has passed the chamber. The impact time is normally determined by independent scintillation counters or by the beam crossing in bunched colliders. Drift chambers do not stand as high particle rates as do MWPCs. Due to the relatively long drift times (in the microsecond range) they cannot be used easily in the fast decision trigger.

There are many possible arrangements of the signal wires and of the cathode, which may itself consist of wires. Typical for fixed target experiments is the flat drift chamber, whereas in collider experiments cylindrical drift chambers have been widely used as central tracking detectors (Fig. 2.9).

The measurement error of a drift chamber is generated by several effects. The most important of these are the following ones [16]:

- The random character of the primary ionization electrons, as far as their number and their spatial distribution is concerned.
- Diffusion along the path of the drifting electrons.
- Different path lengths of different primary electrons due to the inclination of the track and an inhomogeneity of the electric field.
- Variation of the drift velocity due to changes in temperature and pressure.
- Locally non-saturated drift velocity.
- The discretization error of the time-to-digital converter.
- Mechanical tolerances of the chamber.
- Gravitational sagging or electrostatic deflection of the sense wires.

The superposition of these independent effects leads to a measurement error which is normally distributed to a good approximation (Fig. 2.10). The actual variance of the measurement error may depend on the point of impact in the chamber. It has to be determined in a calibration experiment and constantly checked during data taking with real tracks. Standard deviations of 0.05 mm have been achieved for small drift distances; typical values are in the order of 0.1-0.2 mm.

### **The time projection chamber (TPC)**

A TPC in some sense is a special type of drift chamber, or rather of a dense stack of drift chambers [27] (Fig. 2.11). The main differences to a conventional drift chamber are the following:

- The drift space extends only to one side of the sense wire plane. Thus there is no left-right ambiguity.
- A magnetic field parallel to the drift direction leads to an additional confinement of the drifting electron cloud, preventing excessive diffusion. This allows drift lengths in the order of 1m and above.
- Opposite the sense wire plane there is a two-dimensional array of small pads. These pads receive pulses induced by the pulses in the sense wires. The properly weighted barycenter of the charges distributed over the pads gives the position of the electron avalanche along the wire.

Thus the TPC delivers genuine space points (wire position, barycenter of charges in the pads, drift distance) without ambiguities. The measurement errors are again normally distributed to a good approximation.

### 2.2.3 The DELPHI detector

There will be four experiments at the future electron-positron collider LEP. One of them is the detector of the DELPHI project. DELPHI is an international collaboration of some 40 groups from all over Europe, including one from the Institute of High Energy Physics of the Austrian Academy of Sciences in Vienna [10]. The huge detector system consists of two main parts: the central part or "barrel", an assembly of several detectors, more or less in the shape of concentric cylinders placed around the beam tube; and two symmetric endcaps which close the cylinder on both ends (Figs. 2.12, 2.13).

The heart of the detector system is a TPC (see Subsection 2.2.2). It is about 3m long, with a diameter of nearly 2.5m. The ionization electrons produced in the chamber volume by charged particles drift towards the end-plates. On both ends there is a plane of sense wires, and the cathode is interspersed with 16 rows of pads arranged in concentric circles around the origin. The TPC is capable of providing up to 16 space points ( $R$ ,  $R\Phi$ ,  $z$ ) per track. ( $R$  and  $\Phi$  are the cylindrical coordinates of the space point in the  $x$ - $y$ -plane, and  $z$  is the drift direction, which is parallel to the beams and orthogonal to  $x$  and  $y$ .) The standard deviation of the  $z$  measurement varies with the drift distance, the average being in the order of .5 mm. The standard deviation of the  $R\Phi$  measurement is about 0.2 mm [28]. For each space point,  $R$  is the radius of one of the sixteen pad rows and can be assumed to be known exactly, after a deterministic alignment correction computed from the data.

Towards the beam vacuum tube the TPC is complemented by the so-called Inner Detector (ID), a cylindrical high-resolution drift chamber. The inner part of the chamber ("jet part")



consists of 16 sectors with 24 sense wires each. A special feature of this chamber is the following: The electric field and the gas mixture are adjusted in such a way that the ionization electrons produced by a straight track coming from the origin arrive at all 24 wires of a sector at about the same time. This makes pattern recognition in this chamber a lot easier. The outer part ("trigger part") is less accurate and measures only  $z$  and the polar angle  $\theta$  of the track.

The beam tube will be surrounded by the Microvertex Detector, a device which uses state-of-the-art silicon technology with the extremely high resolution of about  $5\ \mu\text{m}$  in  $R\Phi$ .

An outstanding feature of the DELPHI detector is its capability of identifying charged particles and photons in a large energy range almost everywhere in solid angle. Apart from a measurement of energy loss due to ionization in the TPC and the analysis of shower profiles in the calorimeters, this is mainly achieved by Ring Imaging Cherenkov Counters ("RICH"es). A RICH measures the angle between the particle direction and the radiated Cherenkov light by determining the radius of the ring which is produced by the intersection of the Cherenkov light cone with a plane detector sensitive to photons [29]. The finding of the rings is a difficult problem of pattern recognition. But in view of the small number of Cherenkov photons which are produced at higher energies in gases this is the only solution, a ring being defined well enough by its center and a few photons along its circumference. The rings can obviously be found much easier if the intersection point of the particle with the photon sensitive detector in the RICH is known. In order to improve the matching of the tracks found in the TPC with the rings in the RICH covering the TPC, there is another tracking detector outside the RICH. It is a drift chamber, the so-called Outer Detector (OD). The OD is in turn enclosed by the central electromagnetic calorimeter, the HPC (High Density Projection Chamber).

All these central detectors are still inside the huge superconducting coil, which produces the field of 1.2 Tesla. The axis of symmetry of the field coincides with the beams. The field is constant in almost the whole of the central region. Outside the coil, still in the central region, is the return yoke of the magnet, which serves at the same time as a hadron calorimeter. It is supposed to absorb all particles with the exception of invisible neutrinos and not strongly interacting muons. The latter are detected by the muon chambers sitting on top of the yoke.

In the forward region we find a similar layout: Two wire chambers (FCA, an array of drift tubes operating in the streamer mode, and FCB, a set of drift chambers) behind the end-plate of the TPC, which improve the momentum resolution of forward tracks, and, sandwiched between them, a RICH. Immediately behind FCB are an electromagnetic calorimeter, a hadron calorimeter acting as the return yoke, and finally again the muon chambers.

## 2.3 The stepwise analysis of experimental data

In the context of high energy physics experiments, data analysis is a long and complex procedure. The input data consist of the response of the various detector components, i.e. of electric signals; these are either analog and carry information about physical quantities, or they are digital in which case they represent the outcome of a yes/no decision. The output is a collection of momenta and masses of the particles in the event, ready for physical interpretation.

Like any other complex task, data analysis is broken up into a number of steps. The input to each step is the output of the preceding one, and each step involves usually some **data reduction**. In a very complicated situation it may be required that some of the steps have to be repeated in the light of the results of later steps. Today all of these steps are performed either by hardwired processors or programmable computers. In fact, computers are omni-present in modern experiments [30]. They are used in all stages of the experiment: in experimental design, detector operation and monitoring, collection and analysis of data, bookkeeping, documentation, communication and publication of the results.

Breaking up data analysis into several independent sub-tasks has numerous advantages. The most important of them are the following ones:

- The software modules needed in each step can be developed independently, provided that the interfaces are well defined.
- In fact, one is forced to define clear interfaces between subsequent steps. This implies the design of a clear and efficient data structure, which helps in turn again in designing the software modules and in understanding the code.
- Every module can be debugged, modified and improved independently, without harmful side effects on other modules.
- Intermediate data can be stored and exchanged in a well defined format.
- In case a serious bug is found in one of the modules the data do not have to be reprocessed from scratch, but one can start from intermediate data. This may save a lot of computing time.

We shall now briefly describe the way of the data from the electric signal to the final data summary tape. An important distinction is the one between **on-line** and **off-line** processing of data, the borderline being the recording of data on mass storage (today still mostly magnetic tape). It has to be kept in mind that data rejected on-line are irretrievably lost. Therefore elaborate monitoring procedures are required to make sure that no valid data are inadvertently

discarded, because this would create an unrecognized and therefore dangerous bias in the recorded data sample.

### 2.3.1 Triggering, digitization and on-line selection

We have seen that a large LEP experiment like DELPHI has more than 100,000 electronic channels. In every event a large fraction of these channels may generate an electric signal which has to be processed. As the bunches of the beams in LEP cross at a rate of 50kHz, an amount of data in the order of 100 kbytes has to be dealt with in only 20 microseconds, so that the detector is ready for the next bunch crossing. Data are therefore generated at a rate in the order of 100 Mbyte/s up to 1 Gbyte/s, but have to be recorded on a mass storage medium which can be written to at a rate of less than 1Mbyte/s. Obviously some sort of selection and data reduction must take place before the event is recorded. This is done by passing the data at high speed through various levels of so-called **triggers**, each one more refined, more selective, and inevitably slower than the previous one [31]. First background events are filtered out, i.e. events arising not from the interaction of two beam particles, but from the interaction of a beam particle with residual gas in the beam tube or from reactions induced by synchrotron radiation. In the higher trigger levels some preliminary pattern recognition may be done, like track finding or shower finding in the calorimeters. In the highest level events may be tagged or selected according to physical relevance.

The lower level triggers are realized by dedicated hardware processors, whereas the high level trigger is a programmable computing device, e.g. a microprocessor, a micro- or a mini-computer or an emulator. The low level trigger works usually only on a subset of the data which can be read out fast, and while it is running the bulk of the signals is converted to numbers by analog-to-digital converters. If the event is accepted by the low level trigger these numbers are then fed into the high level trigger.

If the event is accepted by all trigger levels, it is written to mass storage by the central on-line computer.

### 2.3.2 Pre-processing, pattern recognition and track search

In the first step of off-line processing (or sometimes in the last step of on-line processing) the numbers recorded on tape are converted to physically meaningful quantities like pulse heights, drift time, spatial coordinates etc. This is known as **pre-processing** and requires as an additional input the **calibration constants** of the particular detector component. The data are now ready for the next stage, **pattern recognition**. The objective of pattern recog-

nition is the classification of all detector signal according to the criterion, that all signals in a class should be caused by the same particle. Of course there is one additional class containing all signals which cannot be matched to any particle.

There is a large variety of algorithms and heuristic methods which are used to this end; the choice of a particular method is also very much detector dependent. A review of widely used pattern recognition methods can be found in ref. [32].

In calorimeters the main goal of pattern recognition is to group the detector signals to showers, to analyze the properties of the showers, and to compute the energy and the position of the incident particles [32]. Difficulties can arise from overlapping showers or from showers not wholly contained in the detector and similar anomalies in the data.

Pattern recognition in the tracking detectors is called **track search** and tries to associate the coordinate values measured by all tracking detectors to track candidates. This may or may not be preceded by an intermediate step in which one tries to reconstruct space points independently in all detectors where it is possible. There is again a variety of methods in common use [32, 33]. The output of the track search is a number of track candidates which have to be confirmed or rejected by the following step. Of course track candidates may share an arbitrary number of coordinate values or space points and may contain ambiguous information, e.g. from drift chambers.

### 2.3.3 Track reconstruction

Track reconstruction serves a two-fold purpose: First, the **parameters describing the track geometrically**, i.e. position, direction and curvature, are estimated by a procedure, which should be optimal in the statistical sense; this is called **track fit**. Secondly, the results of the track search have to be confirmed or rejected, and ambiguities have to be solved. This requires **statistical test quantities** which describe the quality of the track candidate, i.e. whether it has high or low probability to correspond to a real physical track.

In addition to the estimated track parameters their **covariance matrix** should be available, so that the estimate can be fed into the next step of analysis with the proper weights. A short review of track fitting methods will be given in Section 3.2. An extensive survey of current techniques is available in ref. [16].

A track candidate may be rejected by the track fit for various reasons, among them the following ones:

- The track candidate apparently is a random combination of coordinate values and does not correspond at all to a real physical track. This is called a **ghost track** and should

occur fairly rarely if the pattern recognition is efficient. Ghost tracks may simply be rejected as junk.

- The track candidate exactly corresponds to a track, with the exception of one (or very few) coordinate value(s). In this case the track fit should not simply reject the track candidate, but try to **locate** the wrong coordinate value(s), which may be **ambiguities** or **outliers**, and to provide a **remedy** in order to save the track. The treatment of outliers will be discussed in Section 4.3.
- The track candidate consists of two segments which belong to two different physical tracks. This may arise if two tracks cross or if a particle decays with only one charged decay product ("**kink**"). The track fit should be able to recognize such cases and to split the track candidate properly. This problem is discussed in more detail in Section 4.4.

It is possible that after the track fit there are still incompatible tracks, i.e. tracks sharing a number of measurements. It is unlikely that all of these represent real tracks. On this level graph theoretical methods can be used to find the largest subset of compatible tracks [34].

### 2.3.4 Vertex reconstruction

After the track reconstruction one tries to associate each track with the interaction vertex at which it originates. Strictly speaking this is another step of pattern recognition; it is, however, closely linked to the **vertex fit** itself, i.e. the estimation of the vertex position and of the momentum vectors of the tracks attached to it [35].

Usually there is a fairly accurate prior knowledge of the position of the primary interaction vertex: the target size in fixed target experiments, the size of the interaction region in collider experiments. This information should be sufficient to find immediately decay products of longlived particles, e.g. of K-mesons, since they do not point to the primary vertex.

All other tracks have to be assumed to come from the primary vertex, for the time being. It is, however, of paramount physical importance to detect shortlived particles like heavy-flavoured hadrons and to determine their decay products. In order to be able to detect such shortlived particles with decay lengths in the order of a mm, one has to estimate the position of the primary vertex with utmost precision. In the vertex fit the decay products of the (invisible) short lived particle play the roles of outliers. Therefore the detection of decay vertices may be regarded as a multiple outlier problem. Due to the high track multiplicities (up to 50 tracks) a purely combinatorial procedure is ruled out; more feasible approaches will be discussed in Section 4.6. An additional benefit of the vertex fit is an improvement of the estimated mo-

menta of the particles participating in the fit, which is due to the additional constraint of a common vertex.

### 2.3.5 Kinematic fit and particle identification

Further information concerning a vertex (primary or secondary) can be obtained from the physical constraints of momentum and energy conservation. Normally only secondary or simple primary vertices will be subjected to a **kinematic fit**, in which the constraints are imposed on the 4-momentum vectors, usually via Lagrangian multipliers [36]. The purpose is in both cases the verification or the rejection of a certain kinematic hypothesis. Whether all outgoing particles have been found can be seen from an inspection of the momentum vectors. If a certain amount of momentum is missing one has to look for neutral particles in order to satisfy momentum conservation. If none can be found, then either the missing momentum has been carried away by an invisible neutrino, or it has disappeared in a hole in the detector, or one has to reconsider the association of tracks to the vertex.

If also energy conservation is to be taken into account, a mass has to be assigned to each particle participating in the kinematic fit. The mass of a particle can be determined by combining the information from the relevant detectors like the TPC, Cherenkov counters, RICHes, calorimeters etc. Of course these may give ambiguous or contradictory information, in which case the kinematic fit can be used to decide between the mass hypotheses, at least for low momentum tracks.

### 2.3.6 Event viewing

By now it should be obvious that in the many steps leading from the raw data to the final data summary tape many decisions have to be made by the data analysis programs. However well these programs are tuned, it is inevitable that sometimes wrong decisions are made. Therefore in doubtful cases human intervention is required. It has proved useful to represent important or difficult events graphically on a high-resolution screen and to let the operator, usually a physicist, take the final decisions. For instance, in the Nobel prize winning experiment UA1 [3] all events which were candidates for a vector boson decay were scanned on a high performance graphic device. For a systematic presentation of such techniques, see ref. [37].

### 2.3.7 Data analysis software

Today software is a major and indispensable part of every experiment. In the large experiments at LEP, 100 people or more are involved in writing the programs needed for the design of the experiment, the operation of the detector, the collection and the analysis of the data. The coordination of software development is thus a formidable task and special techniques have to be developed to ensure the coherence of data and programs [38, 39, 40, 41].

There are marked differences between on-line and off-line software as far as development and running is concerned. On-line software or, more specifically, data acquisition software runs in a single computer or in a number of strongly coupled processors, although the trend is clearly towards more distribution of intelligence and computing power around the experiment. Development of on-line software is usually confined to the laboratory where the experiment is carried out. An exception is the software monitoring the various pieces of the detector system, because nowadays it is supplied by the group building the particular detector component.

Off-line software, on the other hand, has to run on all large computers available to the groups participating in the experiment, i.e. on many different types of computers scattered all over Europe. One of the main requirements is therefore the **portability** of all programs, implying a strict adherence to the rules of standard FORTRAN 77. Also data have to be exchanged between different types of computing equipment, requiring a computer independent data format on mass storage (at present magnetic tape) [39].

Since most parts of the detector system are continuously improved or even replaced by better ones, the **flexibility** of the software is of paramount importance. This means that the programs must be designed in such a way that a piece of code may be changed easily without interfering with other pieces. Briefly, the design must be **modular**. This is also necessitated by the distributed development of the code, which is common practice in high energy physics. Finally, there is all too often a fast turnaround of people in the experiment. Therefore the software should be well documented, it should be easy to read and easy to understand, so that it can be upgraded by people other than the author, without unforeseen and dangerous side effects [42].

## 2.4 Required properties of analysis methods

If we assume that a big LEP experiment writes events onto tape at a rate of 1Hz, then in a data taking period of 2 months, with an availability of the accelerator of 50%, about 2.5 mil-

lion events are recorded. If we assume that the collaboration has 5 mainframe computers at its disposal, and that the data shall be analyzed within half a year, then the processing time of an event cannot be much larger than 20 seconds, particularly if we take into account that at the same time also extensive simulations have to be done. Considering the many steps of analysis to be performed and the complexity of the events to be analyzed, one realizes immediately that all algorithms employed have to be very **fast**. On the other hand, in view of the enormous costs of an experiment, it cannot be justified if the information contained in the data is not fully used. What is needed therefore are fast algorithms, which nevertheless are, in some statistical sense, **optimal**. This applies particularly to the estimation of track and vertex parameters, to a less degree to pattern recognition procedures, the performance of which can hardly be assessed in mathematical terms. In addition, one must not forget that the data can be affected by various sorts of defects. Therefore the estimation procedures should be **robust** to a certain degree. This is to say that they should be insensitive to data behaving in an unexpected way. This point has not found much attention up to now. We shall discuss some of the issues in Chapters 3 and 4. In any case a balance has to be found between precision, robustness and speed. This is a difficult subject, and one can hardly expect to get general results. However, one can learn quite a lot by analyzing simulated data and by comparing different algorithms. Results of such studies will be presented in Chapter 5.



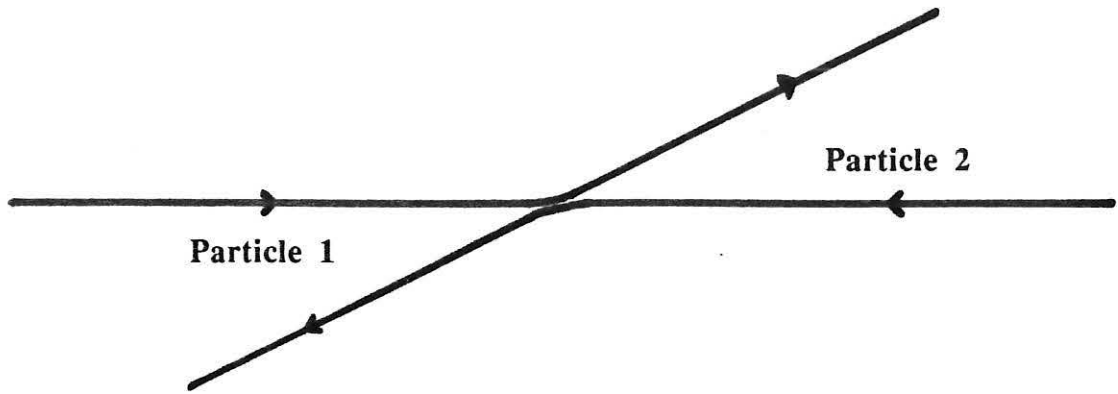


Fig. 2.1: Elastic scattering of two particles. Apart from a change in direction, the particles emerge essentially unchanged from the interaction.

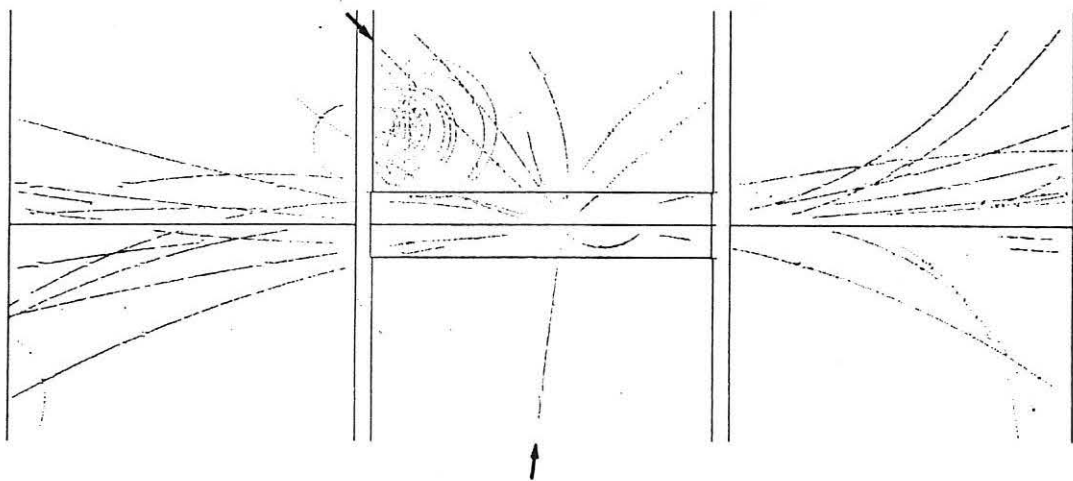


Fig. 2.2: Picture of a  $p\bar{p}$ -collision in the central drift chamber of the UA1 experiment. The energy in the center-of-mass is 540 GeV. A large part of the energy flows into the creation of new particles. The arrows point to the decay products of a  $Z^0$  vector boson.

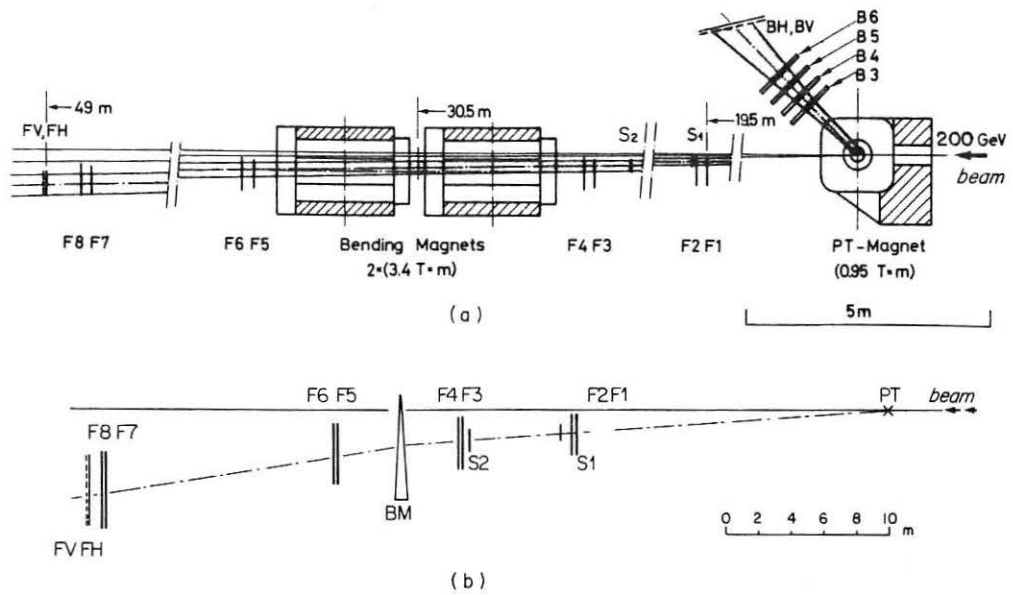


Fig. 2.3: A fixed target spectrometer (experiment WA6).  
 (From: G. Fidecaro et al., Phys. Lett. 105B (1981) 309)

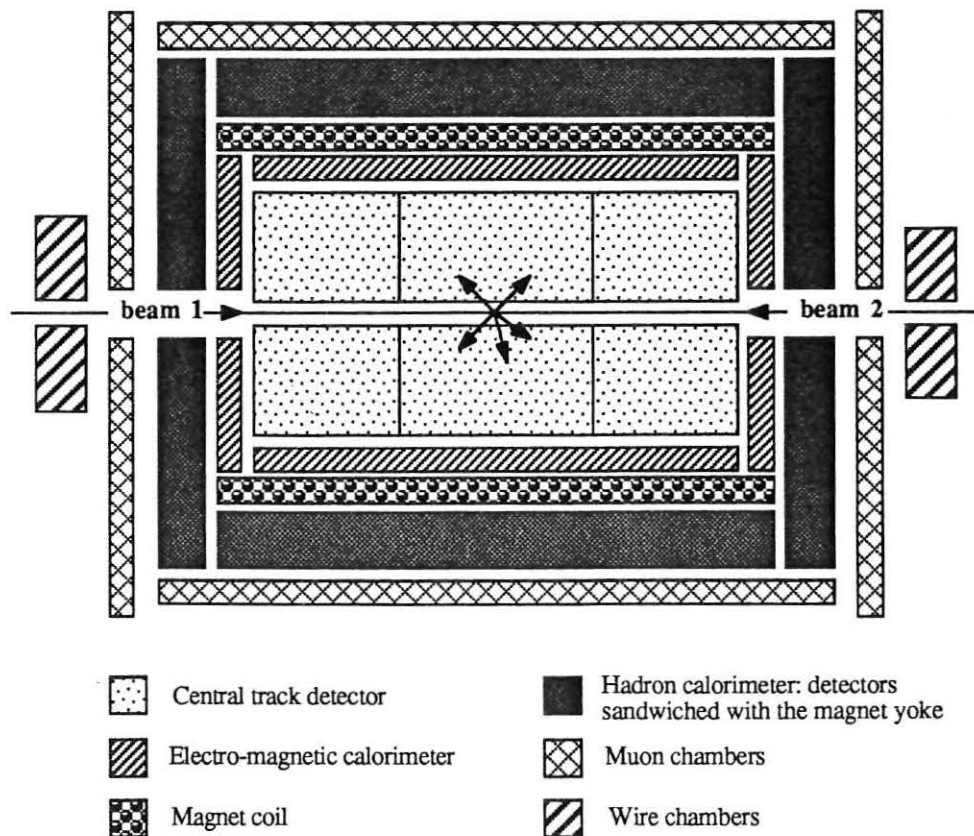
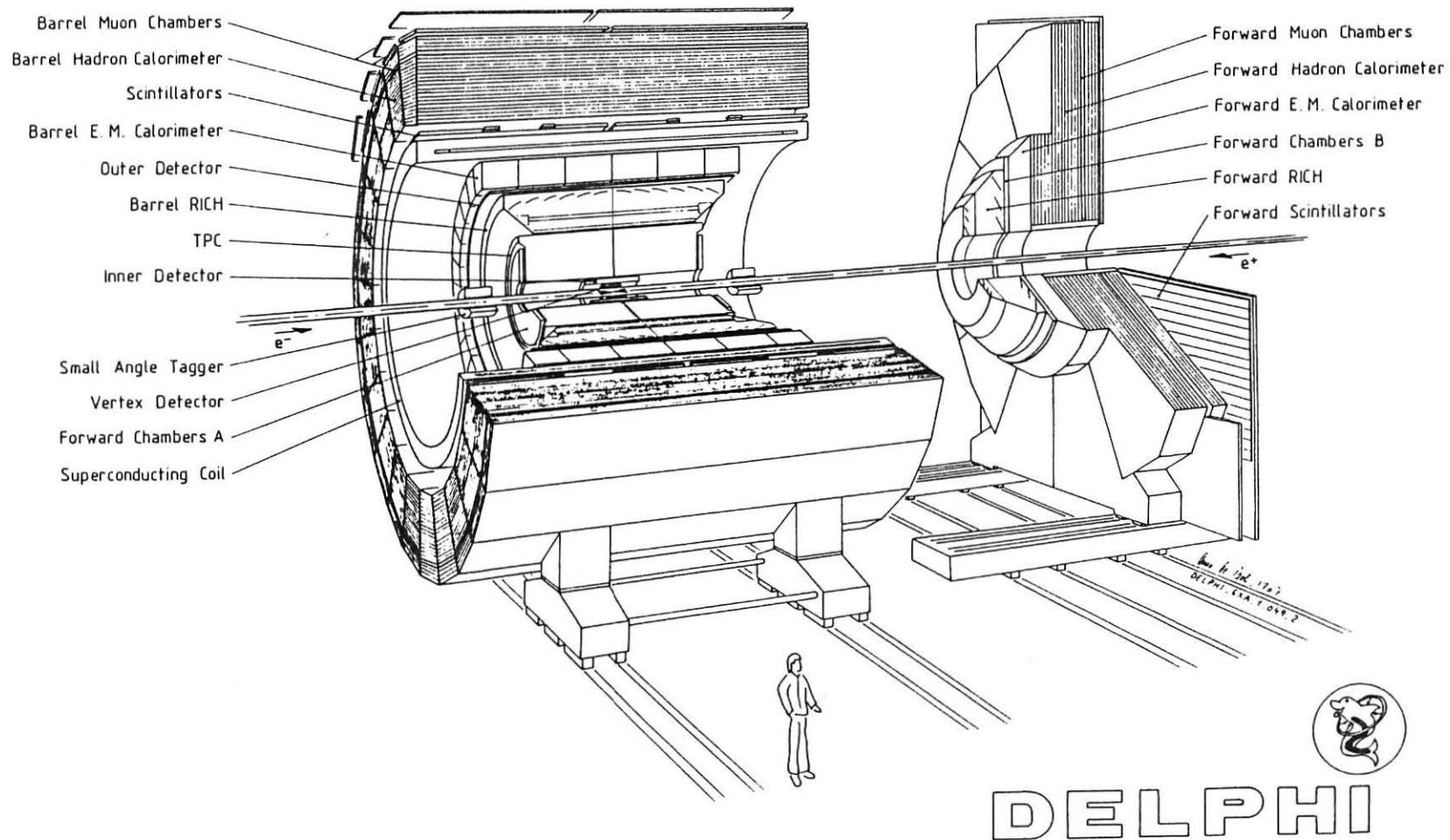


Fig. 2.4: A typical collider experiment.

Fig. 2.5: Artist's view of the DELPHI detector.



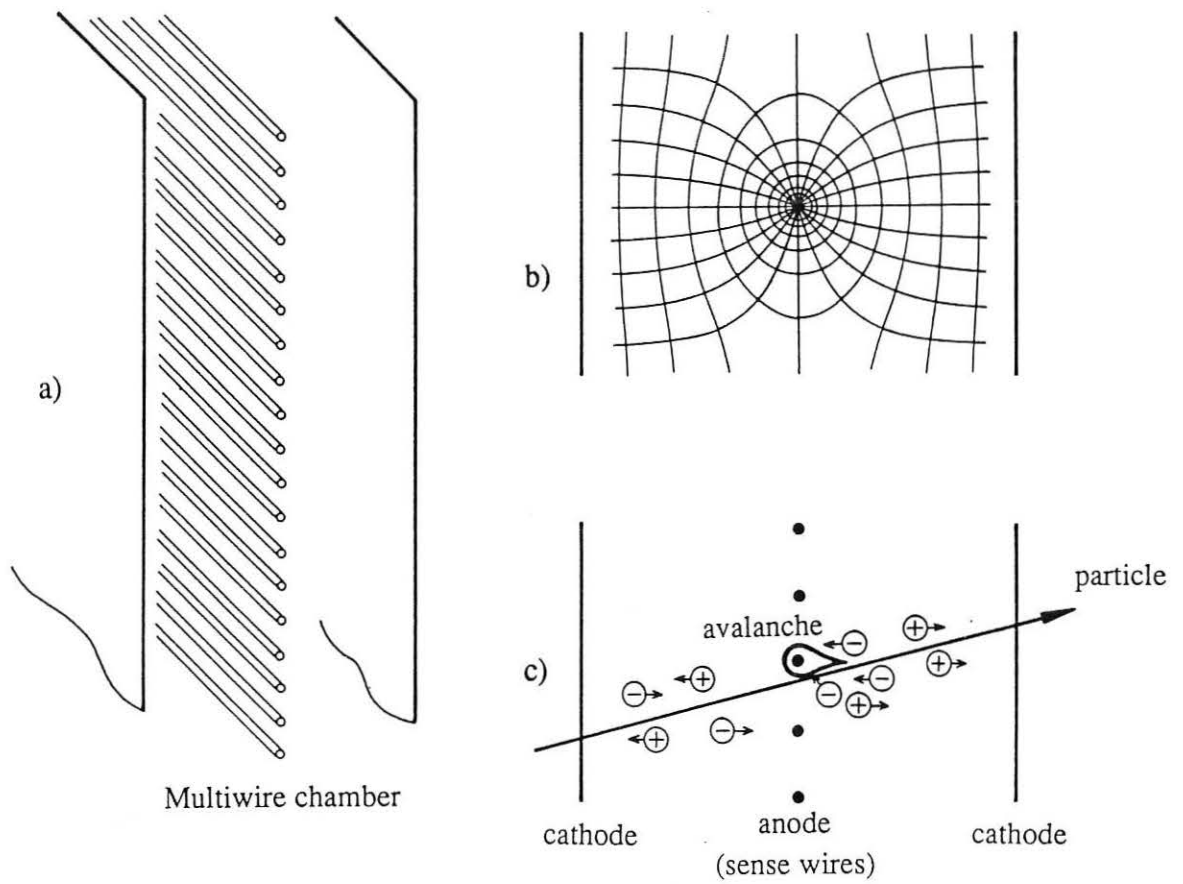


Fig. 2.6: Schematic view of a multi-wire proportional chamber (MWPC).

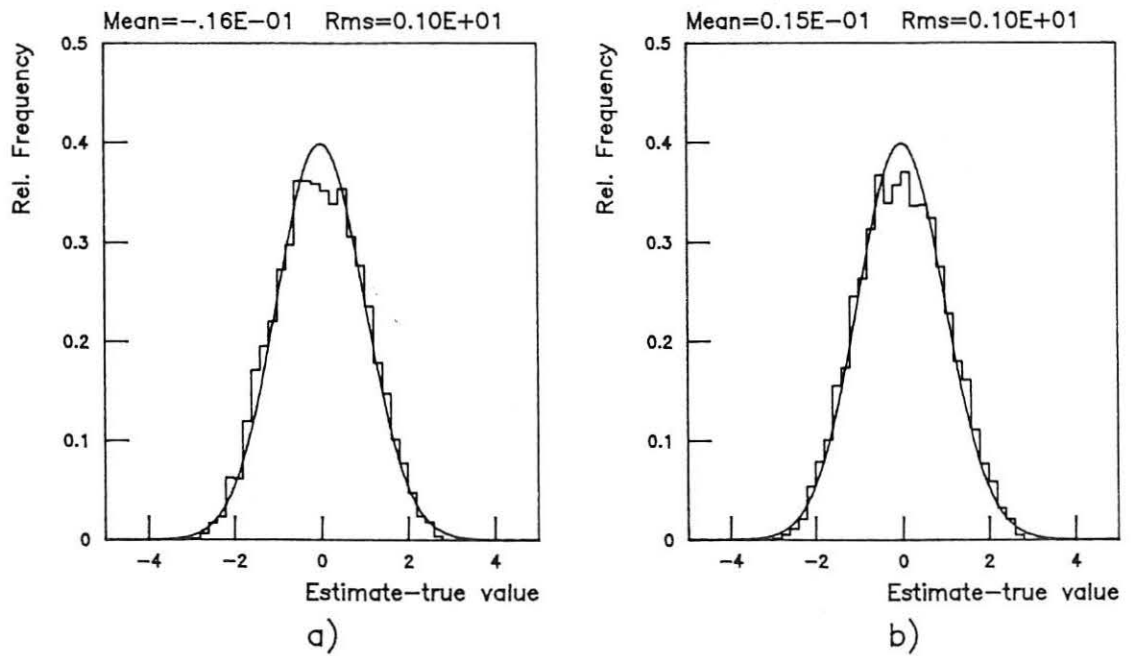


Fig. 2.7: Distribution of the measurement error of a stack of four MWPCs.

a) x-coordinate; b) y-coordinate (from simulated data).

The superimposed curve is a standard normal p.d.f.

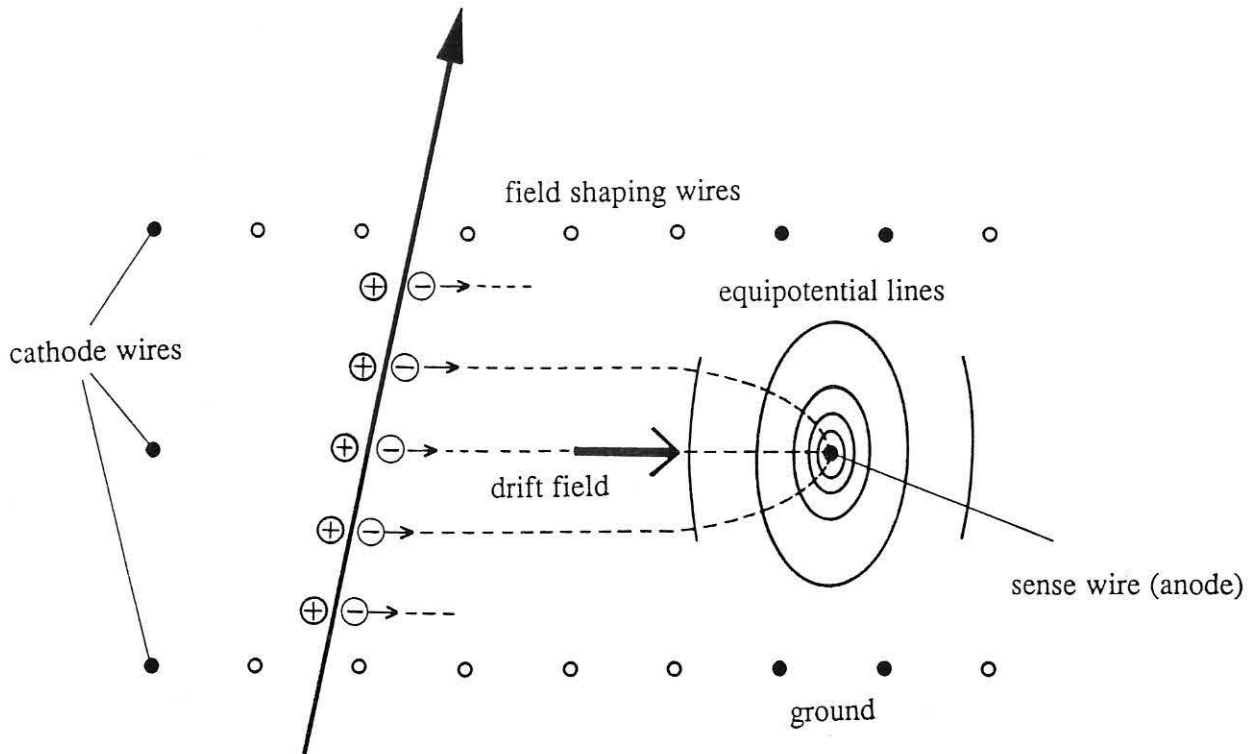


Fig. 2.8: Schematic view of a drift chamber.

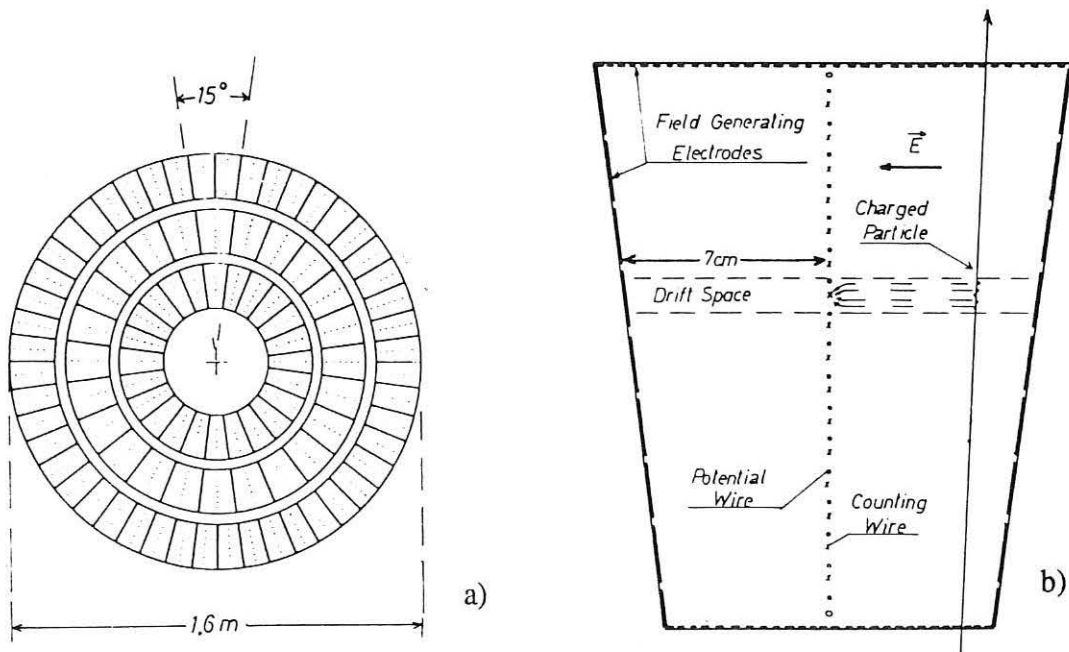


Fig. 2.9: The cylindrical drift chamber of the JADE experiment.  
(From: W. Farr et al., Nucl. Instr. and Meth. 156 (1978) 283)

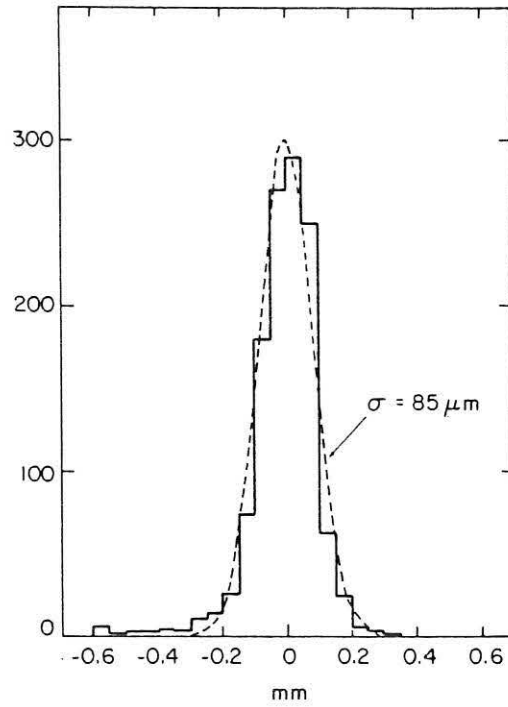


Fig. 2.10: The distribution of the measurement error of a drift chamber.  
 The tail on the left hand side is due to long-range delta-rays.  
 (From: F. Sauli, Nucl. Instr. and Meth. 156 (1978) 147)

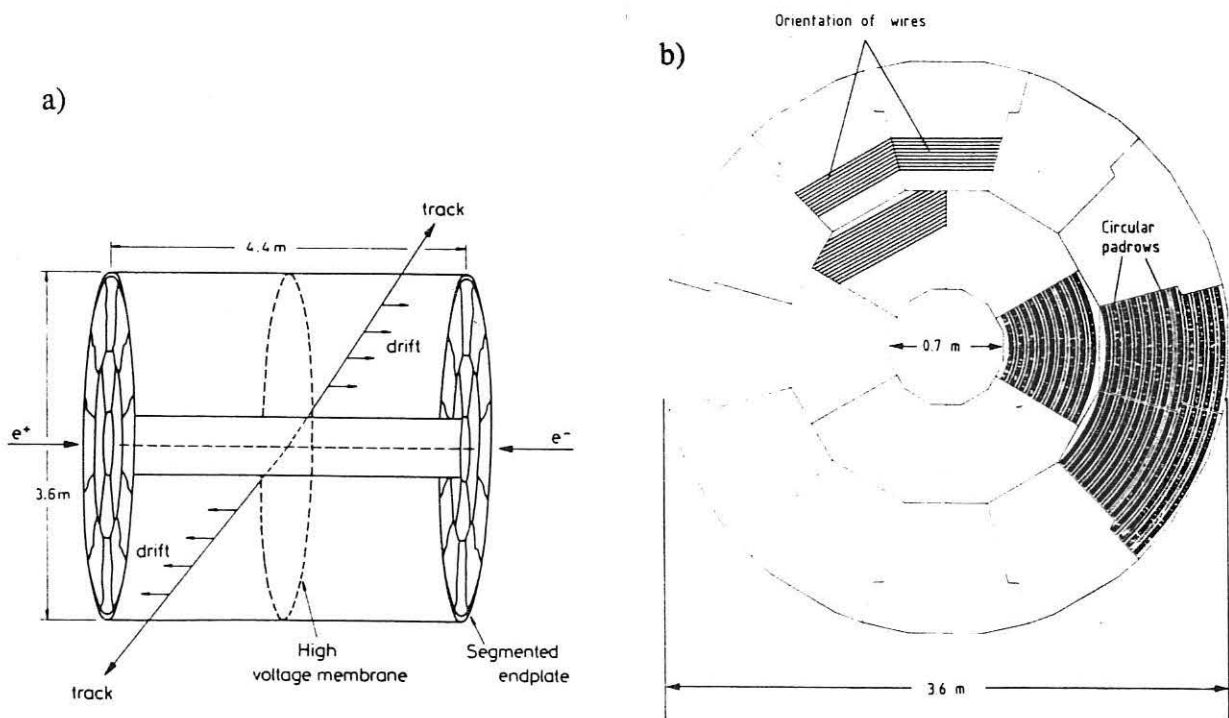


Fig. 2.11: Schematic view of the ALEPH time projection chamber (TPC).  
 (From: S.R. Amendolia et al., Nucl. Instr. and Meth. A252 (1986) 399)

The DELPHI detector. View in the z-y-plane.

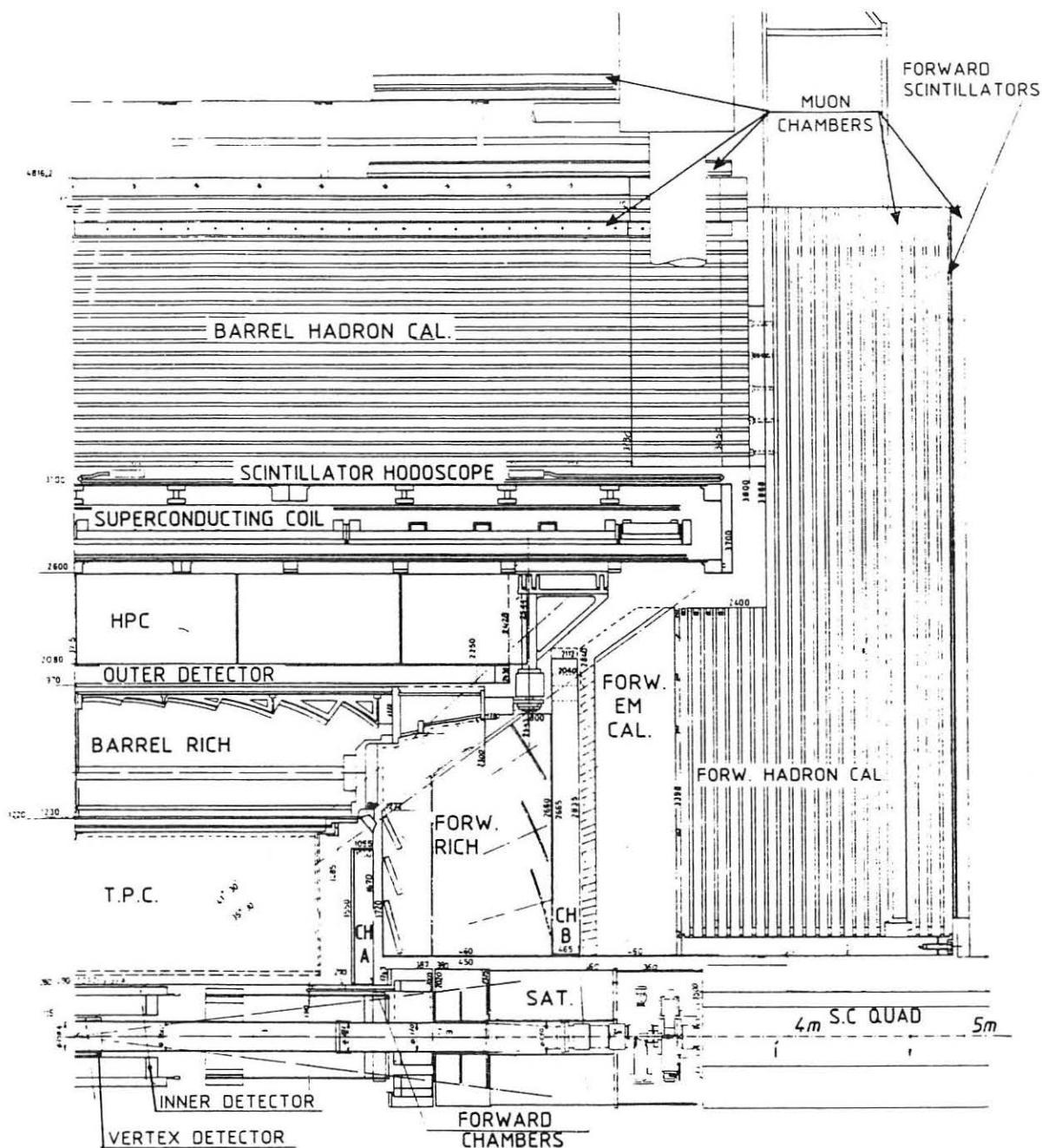


Fig. 2.12: The DELPHI detector. Cross section in the z-y-plane, parallel to the beam.

The DELPHI detector. View in the x-y-plane.

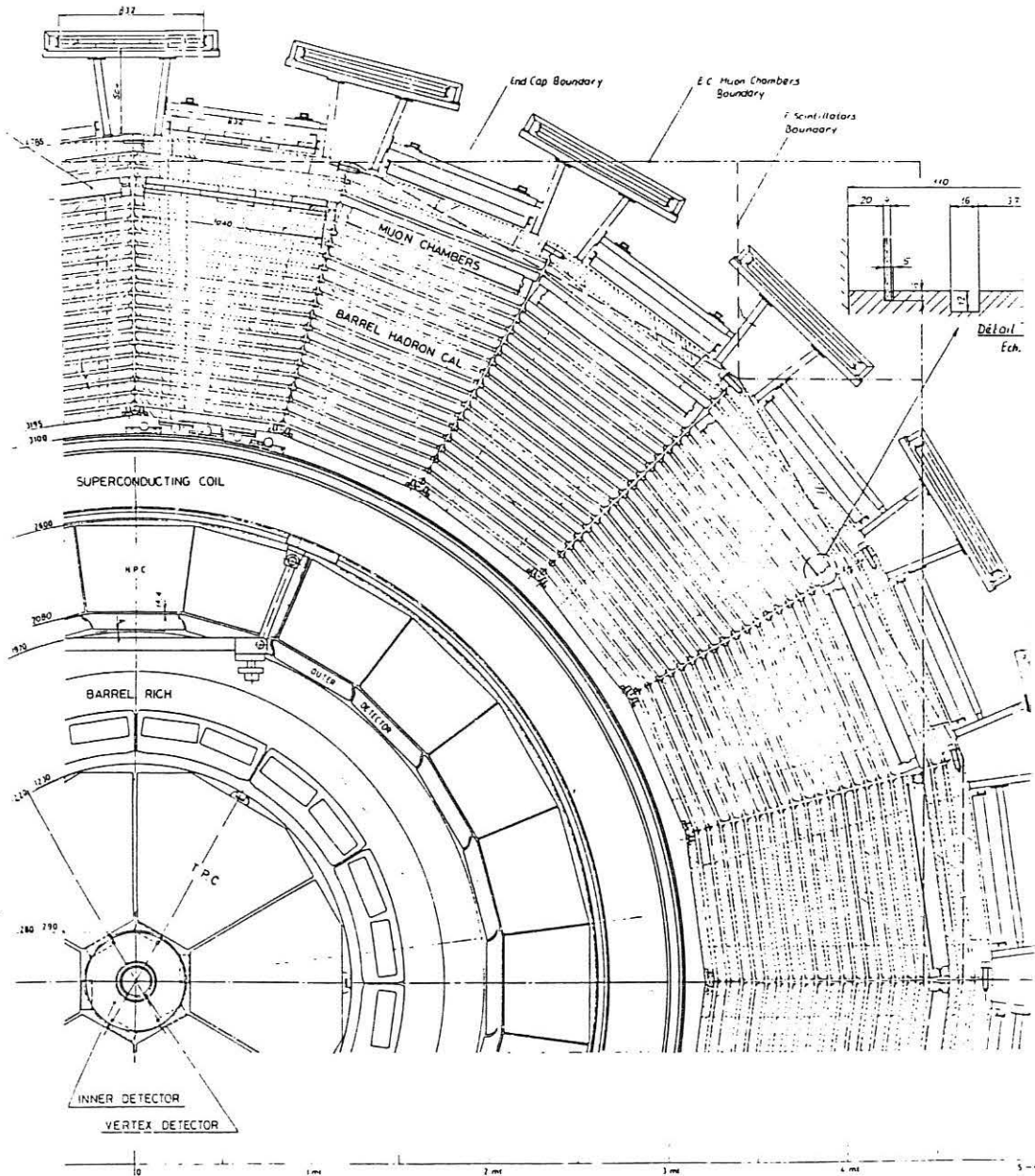


Fig. 2.13: The DELPHI detector. Cross section in the x-y-plane, perpendicular to the beam.



# CHAPTER 3

## TRADITIONAL METHODS OF TRACK AND VERTEX RECONSTRUCTION

In this chapter we shall describe the most commonly used methods of track and vertex fitting. An extensive review can be found in ref. [16].

A striking fact is the prominence of least-squares methods in the estimation of track and vertex parameters. The reasons for this are manifold; a few of them, which are the most important ones in our opinion, will be listed in Section 3.1.

Sections 3.2 and 3.3 deal with the application of least-squares estimators to track and vertex fitting. In Section 3.4 it will be demonstrated that least-squares methods have in fact yielded satisfactory results. With increasing complexity of detectors and events, however, the limits of conventional least-squares methods have become more and more obvious. This point will be discussed in Section 3.5.

### 3.1 Why least-squares?

With a few notable exceptions, e.g. ref. [43], the overwhelming majority of experiments have relied on least-squares estimators for the fitting of tracks and vertices. There are various reasons for this:

- In a linear model a least-squares estimator is a linear function of the vector of measurements; therefore an estimation algorithm is easy to code and its properties are easy to understand.
- If the measurement errors and other noise contributions (see Subsection 3.2.2) follow a normal distribution, the least-squares estimator in the linear model is the minimum variance bound estimator, if the correct weight matrix is used [44]. This is also true asymptotically in a non-linear model. Nevertheless the model should be made as closely linear as possible by the proper choice of the parameters. The assumption of normally distributed noise is frequently valid in very close approximation (see Subsections 2.2.2 and 3.2.2).

- Decision criteria, e.g. the  $\chi^2$ -statistic, are readily available and have well-known distributions. This permits an exact computation of the losses, e.g. in a  $\chi^2$ -test, without an excessive amount of simulation.
- Mean-squared quantities, like the expectation of the  $\chi^2$ -statistic or the variance of the residuals, remain the same also in the case of non-normal noise.
- The covariance matrix of the estimate emerges as a by-product in the calculation of the estimate and is therefore readily available for error propagation into the next stage of analysis.
- The computation of a least-squares estimate is very fast, provided the measurements are uncorrelated or there are only few correlated ones.
- Least-squares methods are familiar to most experimenters and have been applied with undubitable success in the analysis of many experiments. An example will be shown in Section 3.4.

On the other hand it cannot be denied that least-squares methods suffer from severe shortcomings. A discussion of these is deferred to Section 3.5. Also, the development of accelerators and detectors has made the usual estimation algorithms more and more unsuitable. This fact will also be explained in more detail in Section 3.5.

## 3.2 Least-squares estimation of track parameters

A track is uniquely described in each point by 6 parameters: 3 spatial coordinates, 2 directions (angles, cosines or tangents) and the curvature (or absolute value of the momentum with the proper sign). If the track is in vacuum, it is sufficient to know these parameters **in a single point**; the trajectory is then the solution of the **equation of motion** with these particular initial values.

The equation of motion of a charged particle in a stationary magnetic field can be derived from Maxwell's equations. It is a second order differential equation:

$$d^2\mathbf{x}/ds^2 = (kq/P) \cdot d\mathbf{x}/ds \otimes \mathbf{B}(\mathbf{x}(s)),$$

with:

- s.....path length,
- k.....constant of proportionality,
- q.....charge of the particle (signed),
- P.....absolute value of the momentum,
- $\mathbf{B}(\mathbf{x})$ .....stationary magnetic field.

In practice it is often sufficient to estimate the track parameters only on a single fixed surface, the so-called **reference surface**, which is usually close to the vertex, but can in principle be chosen arbitrarily. In this case one of the three spatial coordinates can be regarded as fixed, and in vacuum the trajectory is determined by **five initial values** or parameters **p**. **p** consists of two space coordinates, two directions (angles, cosines or tangents), and momentum or curvature (inverse momentum).

As soon as the particle interacts with matter, however, the trajectory is no longer an exact solution of the equation of motion and the track is **not** fully determined by the initial values **p**. This is due to the fact that the interaction of the particle with matter is a random process.

The main task of the track fit is thus the estimation of the track parameters **p** at a reference surface from the vector **m** of measurements associated to a track candidate by the track search. We assume for the time being that all elements of **m** do really belong to the same track.

If the detectors were perfect devices with no measurement errors at all, and if the particle did not interact with matter, **m** would be a deterministic function of the initial values **p**:

$$\mathbf{m} = \mathbf{f}(\mathbf{p}).$$

The function **f** is called the **track model**.

In reality charged particles do interact with matter along their path. The detector system itself consists of all sorts of materials, hence the trajectory is disturbed by the interactions of the particle with the detector itself. The resulting deviation from the exact track model is a random process and can therefore be described by a random vector  $\gamma$ . The measurement vector **m** can duly be regarded as a realization of a random vector **c**, defined by

$$\mathbf{c} = \mathbf{f}(\mathbf{p}) + \gamma.$$

**c** describes the actual physical trajectory.

If measurement errors are taken into account, a further random vector  $\epsilon$  has to be added to **c**:

$$\mathbf{c}' = \mathbf{f}(\mathbf{p}) + \gamma + \epsilon.$$

It should be noted that  $\gamma$  and  $\epsilon$  are stochastically independent.

Two questions arise immediately: How is the function **f** determined? What is the distribution of  $\gamma$  and  $\epsilon$ ? These questions will be discussed in Subsections 3.2.1 and 3.2.2, respectively.

### 3.2.1 The track model

We recall that the track model represents the set of exact solutions of the equation of motion. It is therefore uniquely determined by the magnetic field  $\mathbf{B}$  and by the arrangement of the tracking detectors.

If  $\mathbf{B}$  is equal to zero everywhere, the trajectory is a straight line and  $\mathbf{f}$  is a linear function, at least for plane detectors. If  $\mathbf{B}$  is constant (homogeneous) in a certain region, the trajectory is a helix and  $\mathbf{f}$  is obtained by the intersection of the helix with the detectors. Note that  $\mathbf{f}$  may still be a linear function, if the measurements and the track parameters are expressed in a suitable coordinate system.

If  $\mathbf{B}$  is inhomogeneous, i.e. if it depends on  $\mathbf{x}$ , it is normally not possible to find an analytic solution of the equation of motion, except in cases where the field is highly symmetric [45]. Therefore an algorithm has to be chosen which permits to compute the track model numerically. Such a procedure is called **track following**, since it follows the track from the initial values through the different detectors. The standard way of track following is a numerical integration of the equation of motion, usually by a Runge-Kutta method [46]. Depending on the step size and on the numerical representation of the magnetic field, track following may consume quite a lot of computing time. Therefore one may try to **approximate** the function  $\mathbf{f}$  by an **analytic function**, for example multidimensional polynomials or trigonometric polynomials. This is also known as a "parametrization" of  $\mathbf{f}$ . If  $\mathbf{f}$  is to be parametrized, one has to generate a **training sample** of tracks and to compute  $\mathbf{f}$  for all tracks in the training sample, e.g. by numerical integration. From the values of  $\mathbf{f}$  the coefficients of the approximation can be computed. Of course one has to take care that the training sample covers all values of  $\mathbf{p}$  which are likely to arise during subsequent data analysis. Finally it is advisable to check the quality of the approximation with an independent **test sample** of tracks. If no satisfactory approximation can be found it may help to divide the region of variation of  $\mathbf{p}$ , the **phase space**, into several cells.

Closely linked to the problem of track following is the question of a suitable numerical representation of the magnetic field. The parameters to be optimized are speed of computation and the size of the field map in computer memory. For a case study, see ref. [47].

### 3.2.2 Description of noise terms

We recall that two types of noise have to be considered:

- Measurement errors, i.e. the difference between the true value of the measured physical or geometrical quantity and the value registered by the detector.

- A random perturbation of the track due to interactions of the particle with material, i.e. the difference between the actual trajectory and the exact solution of the equation of motion.

As far as measurement errors are concerned, we note first that measurement errors arising in different detectors are stochastically independent, except for rare cases, in which two detectors are so close to each other that correlations can be observed. This occurs only in detectors with inherent discretization, e.g. in MWPCs (discretization of space) or drift chambers (discretization of time).

Generally speaking, the distribution of the measurement error can be described by a conditional probability density function (p.d.f.)  $g$ , the so-called resolution function:

$$g_k = g_k ( \varepsilon_k | \overset{t}{c}_k ) = g_k ( c_k - \overset{t}{c}_k | \overset{t}{c}_k ),$$

with:

- $c_k$  ..... vector of quantities measured by detector  $k$ ,
- $\overset{t}{c}_k$  ..... true values of  $c_k$ ,
- $\varepsilon_k$  ..... vector of measurement errors of detector  $k$ .

This includes the possibility that the distribution of the measurement error depends on the impact point of the track on the detector. An eventual dependence should be sufficiently weak, so that the measurement itself or an approximate impact point obtained from the track search can be substituted for the true impact point without ill effects.

We recall that for most tracking detectors in use today  $g_k$  can be approximated very well by a normal p.d.f. . As a normal p.d.f. is uniquely described by mean vector and covariance matrix, the problem of finding  $g_k$  is reduced in practice to the problem of determining the mean and the covariance matrix of the measurement error  $\varepsilon_k$ . For very simple detectors this can be done by theoretical considerations. In all other cases mean and covariance matrix have to be estimated from real data. Frequently a calibration experiment is carried out, in which the detector is exposed to a particle beam, a radioactive source, cosmic radiation or a laser beam. Once a sufficient number of tracks has been registered, standard techniques for the estimation of location and variance components can be applied (see e.g. ref. [48]).

The main source of random perturbations of a track is **multiple Coulomb scattering** (m.s.), i.e. elastic scattering of the particle off the electrons or nuclei of material. If a particle traverses a layer of material it suffers a random deflection  $\Theta$  and a random offset  $\gamma$ , with the exception of very thin layers for which the offset is negligible (Fig. 3.1).

Since these random effects can be ascribed to a superposition of a very large number of independent scattering processes, the derivative of the scattering angle can be represented mathematically as white noise [49]:

$$\vartheta(s) = d\Theta/ds,$$

$$\mathcal{E} \{ \vartheta(s_1) \cdot \vartheta(s_2) \} ds_1 ds_2 = c \cdot \delta(s_1 - s_2) ds_1 ds_2,$$

with:

$$\Theta(s) \dots\dots\dots \text{projected scattering angle at } s.$$

$c$  is the variance of the projected scattering angle per unit length. It depends on the mass and the momentum of the particle, and on the kind of material traversed:

$$c = \frac{k}{L_r} \frac{m^2 + p^2}{p^4},$$

with:

$m$ ..... mass,

$p$ ..... momentum,

$L_r$ ..... radiation length of the material (material constant),

$k$ ..... constant of proportionality (independent of the material).

We are now able to compute the covariance matrix of  $\gamma(L)$  and  $\Theta(L)$ :

$$\mathcal{E} \{ \Theta(L)^2 \} = c L,$$

$$\mathcal{E} \{ \Theta(L) \cdot \gamma(L) \} = c L^2/2,$$

$$\mathcal{E} \{ \gamma(L)^2 \} = c L^3/3.$$

Since  $\gamma$  and  $\Theta$  are due to the superposition of many independent scattering events, we may expect that they are normally distributed in good approximation.

Let us now consider again the original formula of the noise terms:

$$c = \mathbf{f}(\mathbf{p}) + \gamma + \varepsilon,$$

where  $\gamma$  describes the perturbation of the trajectory due to multiple scattering. Then the covariance matrix of  $\gamma$  is given by

$$cov \{ \gamma_i, \gamma_j \} = c \cdot \int_0^{\min(s_i, s_j)} \left( \frac{\partial f_i}{\partial \Theta_1} \frac{\partial f_j}{\partial \Theta_1} + \frac{\partial f_i}{\partial \Theta_2} \frac{\partial f_j}{\partial \Theta_2} \right) ds,$$

where  $\Theta_1$  and  $\Theta_2$  are two independent (orthogonal) projected scattering angles [49]. The integral can be computed analytically only in special cases, e.g. for a straight track. In all other

cases it is evaluated numerically. A simple trapezoidal rule is sufficiently accurate in most cases.

It should be noted that the covariance matrix of  $\gamma$  is **non-diagonal**.

### 3.2.3 The global method of estimation of track parameters

We are now able to specify fully the estimation problem:

$$\mathbf{c} = \mathbf{f}(\mathbf{p}) + \delta,$$

$$\delta = \gamma + \varepsilon,$$

$$\text{cov}\{\delta\} = \mathbf{V} = \text{cov}\{\gamma\} + \text{cov}\{\varepsilon\}.$$

$\text{cov}\{\gamma\}$  in general is **non-diagonal**, whereas  $\text{cov}\{\varepsilon\}$  is a **block-diagonal** matrix, with one block per detector. The vector of measurements,  $\mathbf{m}$ , is a realization of the random vector  $\mathbf{c}$ . The least-squares estimate of  $\mathbf{p}$  is obtained by minimizing the following objective function:

$$Q(\mathbf{p}) = \delta^T \mathbf{G} \delta = [\mathbf{m} - \mathbf{f}(\mathbf{p})]^T \mathbf{G} [\mathbf{m} - \mathbf{f}(\mathbf{p})], \mathbf{G} = \mathbf{V}^{-1}.$$

Differentiation of  $Q$  with respect to  $\mathbf{p}$  yields

$$\partial Q / \partial \mathbf{p} = -2 [\mathbf{m} - \mathbf{f}(\mathbf{p})]^T \mathbf{G} \partial \mathbf{f} / \partial \mathbf{p}.$$

In order to find a zero of  $\partial Q / \partial \mathbf{p}$ , we apply Newton's method and obtain the following recursion:

$$\tilde{\mathbf{p}}_i = \tilde{\mathbf{p}}_{i-1} + \{ \mathbf{A}^T \mathbf{G} \mathbf{A} - (\partial^2 \mathbf{f} / \partial \mathbf{p}_{i-1}^2)^T \mathbf{G} [\mathbf{m} - \mathbf{f}(\tilde{\mathbf{p}}_{i-1})] \}^{-1} \mathbf{A}^T \mathbf{G} [\mathbf{m} - \mathbf{f}(\tilde{\mathbf{p}}_{i-1})],$$

with:

$$\mathbf{A} = \partial \mathbf{f} / \partial \mathbf{p}(\tilde{\mathbf{p}}_{i-1}).$$

The derivatives are evaluated at  $\tilde{\mathbf{p}}_{i-1}$ . This procedure is equivalent to fitting a parabola to  $Q(\mathbf{p})$  at  $\tilde{\mathbf{p}}_{i-1}$  and computing the vertex of the parabola.

The starting point of the recursion,  $\mathbf{p}_0$ , must be sufficiently close to the true minimum of  $Q$ , in particular closer than the nearest inflection point; otherwise the iteration diverges. Since the track search yields usually a fairly good guess of the track parameters, this is rarely a problem.

The derivative matrix  $\mathbf{A}$  can be calculated in various ways. In common use are [16]:

- Numerical differentiation of  $\mathbf{f}$  by small variations of  $\mathbf{p}$ . This increases the time spent in track following sixfold, because the original track and five variations have to be followed.

- The partial derivatives of  $\mathbf{f}$  can be integrated in parallel to  $\mathbf{f}$  itself, by a modified Runge-Kutta algorithm [50]. For an exact treatment the derivatives of the magnetic field are required.
- The partial derivatives are approximated by analytical functions (see Subsection 3.2.1).

The covariance matrix of multiple scattering is computed in parallel to the track following (see Subsection 3.2.2). It depends normally only weakly on the initial conditions  $\mathbf{p}$  and thus doesn't need to be recomputed in every iteration. The covariance matrix of the measurement errors can be assumed to be known in this stage of the analysis.

If the track model is not known explicitly, the calculation of the tensor of second derivatives becomes prohibitively time consuming. In this case the term containing  $\partial^2 \mathbf{f} / \partial \mathbf{p}^2$  is neglected. The matrix in the Newton formula remains positive definite, but the curvature of  $Q$  is clearly no longer approximated correctly. The resulting formula is the same as the one which is obtained by approximating  $\mathbf{f}$  by a linear function:

$$\mathbf{f}(\mathbf{p}) \approx \mathbf{f}(\tilde{\mathbf{p}}_{i-1}) + \mathbf{A}_{i-1} (\mathbf{p} - \tilde{\mathbf{p}}_{i-1}), \quad \mathbf{A}_{i-1} = \partial \mathbf{f} / \partial \mathbf{p}(\tilde{\mathbf{p}}_{i-1}).$$

Then the standard linear least-squares estimate is given by

$$\tilde{\mathbf{p}}_i = \tilde{\mathbf{p}}_{i-1} + (\mathbf{A}_{i-1}^T \mathbf{G} \mathbf{A}_{i-1})^{-1} \mathbf{A}_{i-1}^T \mathbf{G} [\mathbf{m} - \mathbf{f}(\tilde{\mathbf{p}}_{i-1})].$$

Practice shows that rarely more than 3 iterations are required until the minimum is sufficiently approached, i.e. until

$$|Q(\tilde{\mathbf{p}}_i) - Q(\tilde{\mathbf{p}}_{i-1})| < \varepsilon.$$

A choice of  $\varepsilon = 0.01$  is adequate.

It should be noted that the convergence may be slowed down by neglecting the second derivatives. Therefore care should be taken that  $\mathbf{f}$  does not deviate too far from a linear function. This can be facilitated by the proper choice of the coordinate system in which the measurements and the track parameters are expressed. The right choice depends of course on the configuration of the magnetic field and of the tracking detectors. As an example, let us consider the frequently used setup of cylindrical detectors in a homogeneous magnetic field, which is parallel to  $z$  (Fig. 3.2). On each cylinder, two coordinates are measured:  $(R\Phi)_i$  and  $z_i$ . The track parameters are estimated on a cylindrical reference surface. Then the following choice of track parameters will give a track model which is close to linear [16]:

$$p_1 = (R\Phi)_{\text{ref}}, p_2 = z_{\text{ref}}, p_3 = (\tan \lambda)_{\text{ref}}, p_4 = \varphi_{\text{ref}} - \Phi_{\text{ref}}, p_5 = (1/r)_{\text{ref}},$$

where  $\lambda$  is the inclination angle of the helix with respect to the  $x$ - $y$ -plane, and  $r$  is the radius of the helix which is the exact solution of the equation of motion. The choice of  $1/r$  allows a change of sign of  $r$  without discontinuity.



### 3.2.4 Test statistics

The most important test statistic is the value of the objective function at the minimum. For a linear least-squares estimation with Gaussian noise this value is exactly  $\chi^2$ -distributed with  $n-m$  degrees of freedom, where  $n$  is the number of (functionally independent) measurements and  $m$  is the number of estimated parameters. We call it briefly the "chi-square of the fit". In track (and also in vertex) fitting it is in fact  $\chi^2$ -distributed to a very good approximation. Its importance as a test statistic lies in the fact that it tells us whether the hypothesis that all measurements belong to the same track (that all tracks belong to the same vertex) is justified. Since its distribution is known, the losses caused by a  $\chi^2$ -cut can readily be computed. Moreover it has the nice property that it is independent of the estimated parameters, so that a  $\chi^2$ -cut does not introduce a bias into the accepted portion of the sample, which is therefore representative. It has to be stressed, however, that a precise understanding of multiple scattering and detector resolution is a necessary prerequisite to a good performance of the  $\chi^2$ -cut.

In addition, the distribution of the chi-square of the fit can be used in first place to check whether the fit program is correct. A more demanding test in this respect is provided by the studentized residuals ("pulls") which ought to be distributed according to a standard normal p.d.f.:

$$r_i = (m_i - \tilde{m}_i) / \sqrt{s_{ii}}, \quad s_{ii} = (V - A(A^T G A)^{-1} A^T)_{ii}.$$

They can be computed for simulated as well as for real data and thus provide a continuous check of the quality of the data.

## 3.3 Estimation of vertex parameters

The estimation of vertex parameters follows much the same lines as the estimation of track parameters. In the vertex fit the track parameters  $\tilde{\mathbf{p}}_i$  estimated on a reference surface are considered as computed (virtual) measurements; their covariance matrix is also known from the track fit. The vertex parameters to be estimated are

- the vertex position  $\mathbf{v}$  (a point in 3-dimensional space),
- and the 3-momentum vectors  $\mathbf{q}_i$  of the  $n$  tracks belonging to the vertex.

If the reference surface of the track fit is chosen such that no multiple scattering occurs between the vertex and the reference surface, the estimation problem can be written down in the following way (with  $\mathbf{p}_i$  instead of  $\tilde{\mathbf{p}}_i$ ):

$$\mathbf{p}_i = \mathbf{h}_i(\mathbf{v}_k, \mathbf{q}_i) + \varepsilon_i, \quad cov\{\varepsilon_i\} = cov\{\mathbf{p}_i\} = \mathbf{V}_i = \mathbf{G}_i^{-1} \quad (1 \leq i \leq n).$$

$\mathbf{p}_i$  is the 5-dimensional vector of the estimated track parameters of track  $i$  with its covariance matrix  $\mathbf{V}_i$ . Note that the  $\varepsilon_i$  are stochastically independent random vectors. (The components of a single  $\varepsilon_i$  in general are correlated.)

Least-squares estimation of the parameters requires the objective function

$$Q(\mathbf{v}, \mathbf{q}_1, \dots, \mathbf{q}_n) = \sum_{i=1}^n [ \mathbf{p}_i - \mathbf{h}_i(\mathbf{v}, \mathbf{q}_i) ]^T \mathbf{G}_i [ \mathbf{p}_i - \mathbf{h}_i(\mathbf{v}, \mathbf{q}_i) ]$$

to be minimized.

If all tracks are exact helices, this non-linear least-squares problem can be solved by an iterative method proposed in [51]. We prefer to give a general algorithm which is based on a linear expansion of the functions  $\mathbf{h}_i$  [35]:

$$\mathbf{h}_i(\mathbf{v}, \mathbf{q}_i) \approx \mathbf{h}_i(\mathbf{v}_0, \mathbf{q}_{i,0}) + \mathbf{A}_i (\mathbf{v} - \mathbf{v}_0) + \mathbf{B}_i (\mathbf{q}_i - \mathbf{q}_{i,0}) = \mathbf{c}_i + \mathbf{A}_i \mathbf{v} + \mathbf{B}_i \mathbf{q}_i,$$

with:

$$\mathbf{A}_i = \frac{\partial \mathbf{h}_i}{\partial \mathbf{v}}(\mathbf{v}_0, \mathbf{q}_{i,0}), \quad \mathbf{B}_i = \frac{\partial \mathbf{h}_i}{\partial \mathbf{q}_i}(\mathbf{v}_0, \mathbf{q}_{i,0}),$$

$$\mathbf{c}_i = \mathbf{h}_i(\mathbf{v}_0, \mathbf{q}_{i,0}) - \mathbf{A}_i \mathbf{v}_0 - \mathbf{B}_i \mathbf{q}_{i,0}.$$

The linearized objective function is

$$Q^{\text{lin}} = \sum_{i=1}^n ( \mathbf{p}_i - \mathbf{c}_i - \mathbf{A}_i \mathbf{v} - \mathbf{B}_i \mathbf{q}_i )^T \mathbf{G}_i ( \mathbf{p}_i - \mathbf{c}_i - \mathbf{A}_i \mathbf{v} - \mathbf{B}_i \mathbf{q}_i ).$$

This linear least-squares problem can be solved by the standard procedure:

$$\begin{bmatrix} \tilde{\mathbf{v}} \\ \tilde{\mathbf{q}}_1 \\ \cdot \\ \cdot \\ \tilde{\mathbf{q}}_n \end{bmatrix} = \mathbf{M}^{-1} \mathbf{N} \begin{bmatrix} \mathbf{p}_1 - \mathbf{c}_1 \\ \cdot \\ \cdot \\ \mathbf{p}_n - \mathbf{c}_n \end{bmatrix},$$

with:

$$\mathbf{M} = \begin{bmatrix} \mathbf{S}_0 & \mathbf{S}_1 & \cdot & \cdot & \mathbf{S}_n \\ \mathbf{S}_1^T & \mathbf{T}_1 & \mathbf{0} & \cdot & \mathbf{0} \\ \cdot & \mathbf{0} & \mathbf{T}_2 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \mathbf{S}_n^T & \mathbf{0} & \cdot & \cdot & \mathbf{T}_n \end{bmatrix},$$

$$\mathbf{S}_0 = \sum_{i=1}^n \mathbf{A}_i^T \mathbf{G}_i \mathbf{A}_i, \quad \mathbf{S}_i = \mathbf{A}_i^T \mathbf{G}_i \mathbf{B}_i, \quad \mathbf{T}_i = \mathbf{B}_i^T \mathbf{G}_i \mathbf{B}_i \quad (1 \leq i \leq n),$$

and with:

$$N = \begin{bmatrix} A_1^T G_1 & \cdot & \cdot & \cdot & A_n^T G_n \\ B_1^T G_1 & \mathbf{0} & \cdot & \cdot & \mathbf{0} \\ \mathbf{0} & B_2^T G_2 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \mathbf{0} & \cdot & \cdot & \cdot & B_n^T G_n \end{bmatrix}.$$

The straight inversion of  $M$  would require about  $(3n+3)^3/2$  operations. It can be speeded up considerably by taking advantage of the internal block structure of  $M$  [35]. If we write  $M^{-1}$  in the form:

$$M^{-1} = \begin{bmatrix} C_{00} & \cdot & \cdot & C_{0n} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ C_{n0} & \cdot & \cdot & C_{nn} \end{bmatrix},$$

the submatrices  $C_{ij}$  can be computed explicitly:

$$\begin{aligned} C_{00} &= (S_0 - \sum S_i W_i S_i^T)^{-1}, \\ C_{0j} &= -C_{00} S_j W_j \quad (j>0), \\ C_{ij} &= \delta_{ij} W_j - W_i S_i^T C_{0j} \quad (i,j>0), \end{aligned}$$

with:

$$W_j = T_j^{-1} \quad (j>0).$$

The final estimate then reads:

$$\begin{aligned} \tilde{v} &= C_{00} \cdot \sum A_i^T G^{B_i} (p_i - c_i), \\ \tilde{q}_i &= W_i B_i^T G_i (p_i - c_i - A_i \tilde{v}), \end{aligned}$$

with:

$$G^{B_i} = G_i - G_i B_i W_i B_i^T G_i.$$

Obviously the covariance matrix of the estimate is  $M^{-1}$ .

The total  $\chi^2$ -statistic of the fit is given by

$$\chi^2 = \sum_{i=1}^n (p_i - c_i - A_i \tilde{v} - B_i \tilde{q}_i)^T G_i (p_i - c_i - A_i \tilde{v} - B_i \tilde{q}_i).$$

If necessary, the fit has to be iterated by re-expanding  $\mathbf{h}_i$  at the estimate  $\tilde{\mathbf{v}}, \tilde{\mathbf{q}}_i$ . Convergence is usually fast; more than two iterations are hardly ever needed. Note that the full covariance matrix does not have to be computed until the final iteration!

### 3.4 Some experimental results

In this section an example of the successful application of least-squares methods shall be presented. The experiment in question, called WA6 (Fig. 2.3), was carried out at CERN about 10 years ago [52]. Its aim was the measurement of the polarization parameter in elastic scattering of protons on free polarized protons. The success of the experiment depended on the clear separation of elastic events from quasi-elastic events (scattering on bound protons) and inelastic events. The final decision on the event type was taken on the basis of the kinematic fit of the elastic hypothesis. By a careful tuning of the measurements errors in the spectrometer, and by precise treatment of multiple scattering – particularly important for the slow recoil proton – it was possible to keep the losses of elastic events in the final  $\chi^2$ -cut of the kinematic fit at a level of only 6%. (In previous similar experiments losses of up to 30% had to be tolerated.) The  $\chi^2$ -probability distribution of the kinematic fit is shown in Fig. 3.3a. The inelastic events accumulate at very low probabilities. The same distribution after a 6%-cut is shown in Fig. 3.3.b. It is remarkably flat. It can be concluded that the measurement errors and the multiple scattering have been correctly understood, and that the information contained in the measurements has been optimally used on all levels of the analysis.

### 3.5 Limits of traditional methods

With the advent of more powerful accelerators and more complex detectors some shortcomings of the traditional methods of track and vertex fitting have made themselves felt. Even if one accepts least-squares estimators as an appropriate tool, there still remains the question whether the traditional implementation described in Sections 3.2 and 3.3 is really best suited to the latest requirements. In fact, closer inspection reveals some marked drawbacks:

- With the global method described in Subsection 3.2.3 the track parameters are estimated in a single point on the reference surface. As in the presence of multiple scattering the trajectory is not an exact solution of the equation of motion, the actual physical track may stray quite far from the "ideal" track, i.e. the extrapolation of the estimated parameters. But in the large modular detectors at LEP extrapolations from the far end of the track are needed in order to link together the pieces of information provided by different detectors,

e.g. central tracking detectors and muon chambers. In addition, the points of intersection of a track with certain detectors need to be computed, for instance with the Ring Imaging Cherenkov counters (see Subsection 2.2.3). Thus it is desirable that estimates of the parameters containing the whole information are available anywhere along the track.

- So far we have assumed that the track candidates produced by the track search algorithm are entirely correct. But this cannot be expected to be true in reality. The track candidate may contain wrong solutions of ambiguous measurements or measurements which are picked up from other tracks or are just noise signals. It is the duty of the track fit not to simply reject such a track candidate, but to find out which measurements might be anomalous. Attempts have been made to handle single outliers by rejecting the measurements with the largest residual whenever the latter exceeds a certain bound, and to recompute the estimate with an updated weight matrix. But in general the global method is not very well adapted to cope with multiple outliers, kinks or long tails in the distribution of the measurement errors.
- For tracks with many measurements and multiple scattering the inversion of the covariance matrix consumes too much time, and a faster track fit algorithm is needed.
- For the physics to be studied at LEP the detection of shortlived particles will be of crucial importance [53]. This means that secondary decay vertices very close to the primary interaction vertex will have to be found. Therefore all tracks emerging from the vertex region should be associated either with the primary vertex or with a secondary vertex. A correct association of all tracks is possible only if the primary vertex itself is computed with the utmost precision. The decision about which tracks are possibly secondary tracks is not easy. Due to the high primary multiplicities combinatorial procedures are ruled out: It is not feasible to recompute the vertex fit with different subsets of tracks removed. What is needed instead is an algorithm the computing time of which is essentially proportional to the total number of tracks.

The next chapter will show how these problems can be solved by the introduction of filter techniques, in particular the Kalman filter, without leaving the realm of least-squares estimation.

A more profound negative aspect of least-squares estimators in general is their lack of **robustness**. Robustness is a rather broad notion which embodies several mathematical concepts [54]. Loosely speaking, a robust estimator is one whose performance remains good if the actual distribution of the data deviates somehow from the assumed one. In particular, an estimator should not be excessively biased by the influence of one or a few outliers, i.e. measurements which do not follow the general pattern, either due to tails in the distribution of the measurement errors or because they are background (see Section 4.3). Robustness of

estimators has not found much, if any, attention up to now in track and vertex fitting. Therefore a first step towards the application of robust algorithms will be suggested in the next chapter.

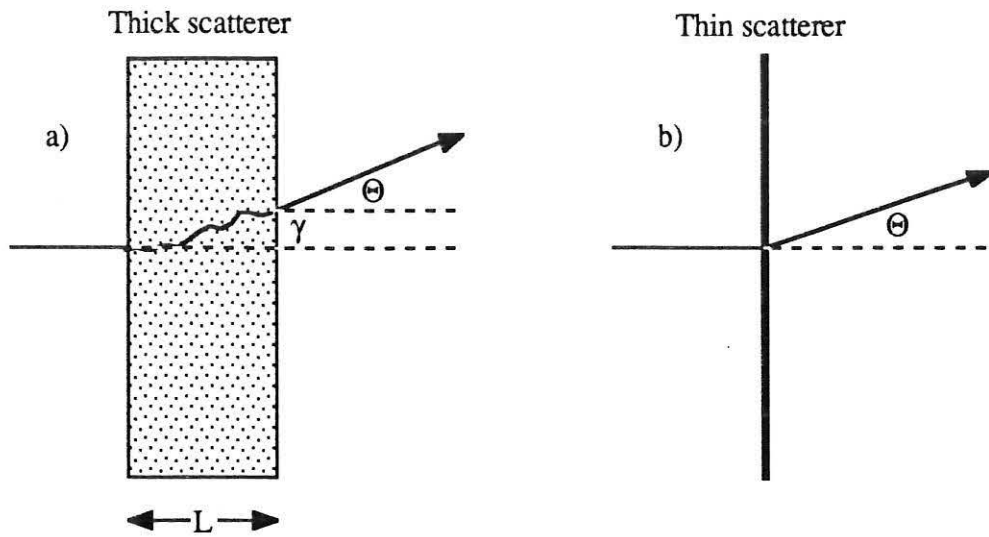


Fig. 3.1: Effect of multiple scattering on a track.

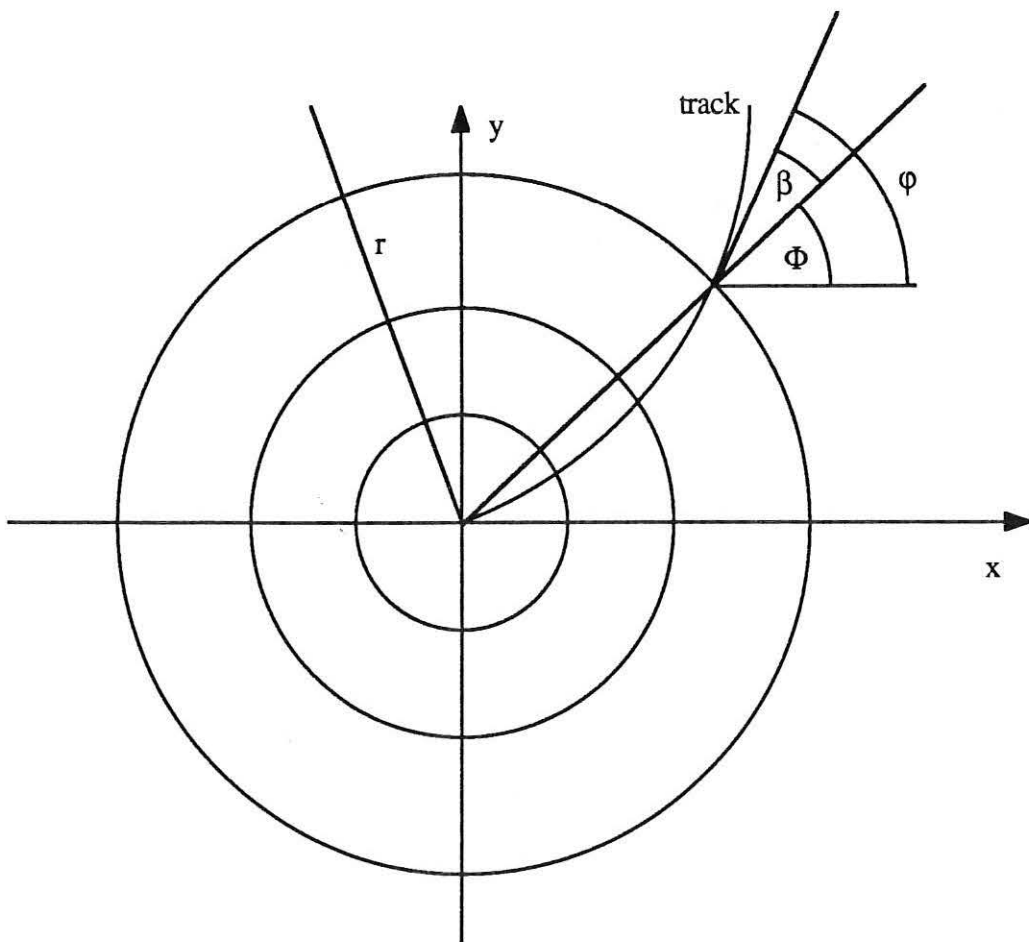


Fig. 3.2: Cylindrical detectors in a homogeneous magnetic field.

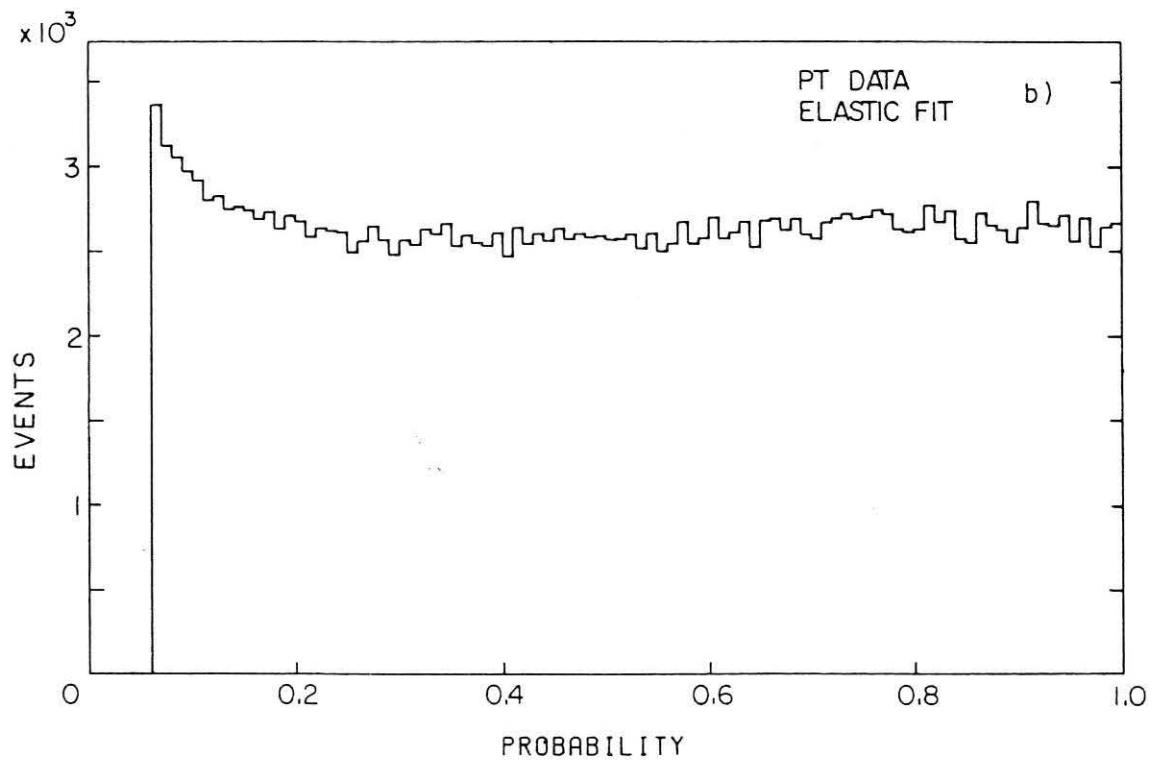
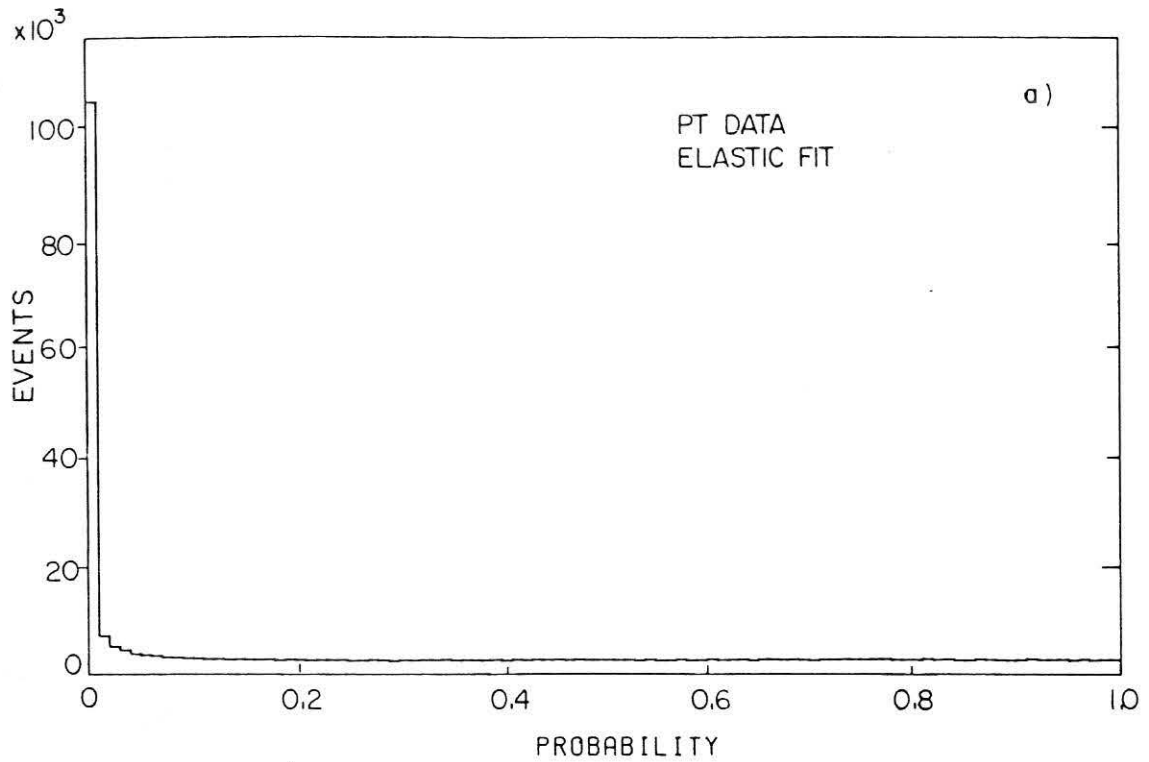


Fig. 3.3: Experiment WA6: Probability distribution of the  $\chi^2$  of the kinematic fit.

- a) The inelastic events accumulate in the high peak at low probabilities.
- b) The same distribution after a cut at 6% shows the expected flat shape.

(From: G. Fidecaro et al., Nucl. Phys. B173 (1980) 513)



# CHAPTER 4

## APPLICATION OF FILTER TECHNIQUES TO TRACK AND VERTEX FITTING

The idea of using filter methods in track fitting is not entirely new. P. Laurikainen mentioned already in 1971 the possibility of using stepwise estimation for the automatic measuring of bubble chamber tracks, but came to the conclusion that the method was too sophisticated for this purpose [55]. In 1984, P. Billoir proposed a recursive method of track fitting which is effectively equivalent to the Kalman filter, without recognizing this fact [56]. This method was called later "progressive method" [16, 35]. In 1987 the progressive method was put into the proper mathematical perspective by the author [57]. This opened up the possibility of complementing the progressive track fit by the elegant smoother algorithm. First we will show that the problems listed in Section 3.5 can be solved by the combination of the filter and the smoother. The full power of the combined filter/smoothing algorithm is exploited in the methods proposed below for handling of outliers and detection of kinks (charged 1-prong decays). We also will take advantage of the fact that the Kalman filter can be robustified with little additional computational effort.

In this chapter we shall first review the basic properties of the Kalman filter (Section 4.1). In Section 4.2 we show that a track can be interpreted as a non-linear discrete dynamic system and write down the system and the measurement equations. We also discuss briefly the merging of track segments. In Section 4.3 we discuss various possibilities of treating outliers and investigate a robustified filter which can readily be applied to track fitting. Section 4.4 deals with the important topic of the recognition of kinks in charged tracks. An algorithm based on the Kalman filter will be presented.

The rest of the chapter is dedicated to vertex fitting. In Section 4.5 we show that an interaction vertex can be regarded as a non-linear discrete dynamic system and write down the system and the measurement equations. Finally, in Section 4.6 we investigate the detection of secondary vertices. By interpreting it as an outlier problem we arrive at various procedures which are based on the Kalman filter and on a robust modification of the filter.

## 4.1 Properties of the Kalman filter

The Kalman filter is a method of analyzing a **linear discrete dynamic system**. The theory of the Kalman filter is described in many textbooks [58, 59, 60], and only a short review will be given here. A discrete dynamic system is in each point of (discrete) time, say  $t_k$ , characterized by a state vector  $\mathbf{x}_k$ . The evolution of the state vector in time is described by a time-dependent transformation, the **system equation**:

$$\mathbf{x}_k = \mathbf{f}_k(\mathbf{x}_{k-1}) + \mathbf{w}_k,$$

where  $\mathbf{f}_k$  is a deterministic function and  $\mathbf{w}_k$  is a random disturbance of the system, the so-called **process noise**.

The state vector does not need to be observed directly. In most cases a function of the state vector is observed, corrupted by **measurement noise**:

$$\mathbf{m}_k = \mathbf{h}_k(\mathbf{x}_k) + \epsilon_k,$$

where  $\mathbf{m}_k$  is the vector of observations at time  $t_k$ . We shall assume in the following that all  $\mathbf{w}_k$  and  $\epsilon_k$  are independent random vectors and have mean  $\mathbf{0}$  and a finite covariance matrix.

In the simplest case both  $\mathbf{f}_k$  and  $\mathbf{h}_k$  are linear functions for all  $k$ :

$$\mathbf{f}_k(\mathbf{x}_{k-1}) = \mathbf{F}_k \mathbf{x}_{k-1} + \mathbf{a}_k,$$

$$\mathbf{h}_k(\mathbf{x}_k) = \mathbf{H}_k \mathbf{x}_k + \mathbf{b}_k.$$

In the following all constants  $\mathbf{a}_k$  and  $\mathbf{b}_k$  are assumed to be  $\mathbf{0}$ , without loss of generality.

There are three types of operations to be performed in the analysis of a dynamic system (here described in terms of "time"):

- **Filtering** is the estimation of the "present" state vector, based upon all "past" measurements.
- **Prediction** is the estimation of the state vector at a "future" time.
- **Smoothing** is the improved estimation of the state vector at some time in the "past" based on all measurements taken up to the "present" time.

The Kalman filter is the optimum solution of these three problems in the sense that it minimizes the mean square estimation error. If the system is linear, and if  $\mathbf{w}_k$  and  $\epsilon_k$  are Gaussian random variables for all  $k$ , the Kalman filter is efficient, i.e. it is **the** optimal filter; no non-linear filter can do better. In other cases it is simply the optimal linear filter.

We give now the formulae for prediction, filtering and smoothing in a linear dynamic system, with the following notations and assumptions:

**System equation:**

$$\begin{aligned}\tilde{\mathbf{x}}_k^t &= \mathbf{F}_k \tilde{\mathbf{x}}_{k-1}^t + \mathbf{w}_k, \\ \mathcal{E} \{ \mathbf{w}_k \} &= \mathbf{0}, \quad \text{cov} \{ \mathbf{w}_k \} = \mathbf{Q}_k \quad (1 \leq k \leq n).\end{aligned}$$

**Measurement equation:**

$$\begin{aligned}\mathbf{m}_k &= \mathbf{H}_k \tilde{\mathbf{x}}_k^t + \varepsilon_k, \\ \mathcal{E} \{ \varepsilon_k \} &= \mathbf{0}, \quad \text{cov} \{ \varepsilon_k \} = \mathbf{V}_k = \mathbf{G}_k^{-1} \quad (1 \leq k \leq n),\end{aligned}$$

**Notations:**

$$\begin{aligned}\tilde{\mathbf{x}}_k^t &= \text{true value of the state vector at time } t_k, \\ \tilde{\mathbf{x}}_k^j &= \text{estimate of } \tilde{\mathbf{x}}_k^t, \text{ using measurements up to time } j, \\ (j < k: \text{ prediction, } j = k: \text{ filtered estimate, } j > k: \text{ smoothed estimate; } \tilde{\mathbf{x}}_k^k &\equiv \tilde{\mathbf{x}}_k), \\ \mathbf{C}_k^j &= \text{cov} \{ \tilde{\mathbf{x}}_k^j - \tilde{\mathbf{x}}_k^t \}, \\ \mathbf{r}_k^j &= \text{residual } \mathbf{m}_k - \mathbf{H}_k \tilde{\mathbf{x}}_k^j, \\ \mathbf{R}_k^j &= \text{cov} \{ \mathbf{r}_k^j \}.\end{aligned}$$

In the predicted state vector,  $j$  is usually equal to  $k-1$ ; in the smoothed state vector,  $j$  is usually equal to  $n$ .

**Prediction:**

Extrapolation of the state vector:

$$\tilde{\mathbf{x}}_k^{k-1} = \mathbf{F}_k \tilde{\mathbf{x}}_{k-1}$$

Extrapolation of the covariance matrix:

$$\mathbf{C}_k^{k-1} = \mathbf{F}_k \mathbf{C}_{k-1} \mathbf{F}_k^T + \mathbf{Q}_k$$

Predicted residuals:

$$\mathbf{r}_k^{k-1} = \mathbf{m}_k - \mathbf{H}_k \tilde{\mathbf{x}}_k^{k-1}$$

Covariance matrix of the predicted residuals:

$$\mathbf{R}_k^{k-1} = \mathbf{V}_k + \mathbf{H}_k \mathbf{C}_k^{k-1} \mathbf{H}_k^T$$

**Filter (Gain matrix formalism):**

Filtered state vector:

$$\tilde{\mathbf{x}}_k = \tilde{\mathbf{x}}_k^{k-1} + \mathbf{K}_k (\mathbf{m}_k - \mathbf{H}_k \tilde{\mathbf{x}}_k^{k-1})$$

Kalman gain matrix:

$$\mathbf{K}_k = \mathbf{C}^{k-1}_k \mathbf{H}_k^T ( \mathbf{V}_k + \mathbf{H}_k \mathbf{C}^{k-1}_k \mathbf{H}_k^T )^{-1} = \mathbf{C}_k \mathbf{H}_k^T \mathbf{G}_k$$

Covariance matrix of the filtered state vector:

$$\mathbf{C}_k = ( \mathbf{I} - \mathbf{K}_k \mathbf{H}_k ) \mathbf{C}^{k-1}_k$$

Filtered residuals:

$$\mathbf{r}_k = \mathbf{m}_k - \mathbf{H}_k \tilde{\mathbf{x}}_k = ( \mathbf{I} - \mathbf{H}_k \mathbf{K}_k ) \mathbf{r}^{k-1}_k$$

Covariance matrix of the filtered residuals:

$$\mathbf{R}_k = ( \mathbf{I} - \mathbf{H}_k \mathbf{K}_k ) \mathbf{V}_k = \mathbf{V}_k - \mathbf{H}_k \mathbf{C}_k \mathbf{H}_k^T$$

Chi-square increment:

$$\chi^2_{k,F} = \mathbf{r}_k^T \mathbf{R}_k^{-1} \mathbf{r}_k$$

Chi-square update:

$$\chi^2_k = \chi^2_{k-1} + \chi^2_{k,F}$$

**Filter (Weighted means formalism):**

Filtered state vector:

$$\tilde{\mathbf{x}}_k = \mathbf{C}_k [ ( \mathbf{C}^{k-1}_k )^{-1} \tilde{\mathbf{x}}^{k-1}_k + \mathbf{H}_k^T \mathbf{G}_k \mathbf{m}_k ]$$

Covariance matrix of the filtered state vector:

$$\mathbf{C}_k = [ ( \mathbf{C}^{k-1}_k )^{-1} + \mathbf{H}_k^T \mathbf{G}_k \mathbf{H}_k ]^{-1}$$

Chi-square increment:

$$\chi^2_{k,F} = \mathbf{r}_k^T \mathbf{G}_k \mathbf{r}_k + ( \tilde{\mathbf{x}}_k - \tilde{\mathbf{x}}^{k-1}_k )^T ( \mathbf{C}^{k-1}_k )^{-1} ( \tilde{\mathbf{x}}_k - \tilde{\mathbf{x}}^{k-1}_k )$$

**Smoother:**

Smoothed state vector:

$$\tilde{\mathbf{x}}^n_k = \tilde{\mathbf{x}}_k + \mathbf{A}_k ( \tilde{\mathbf{x}}^n_{k+1} - \tilde{\mathbf{x}}^k_{k+1} )$$

Smoother gain matrix:

$$\mathbf{A}_k = \mathbf{C}_k \mathbf{F}_{k+1}^T ( \mathbf{C}^k_{k+1} )^{-1}$$

Covariance matrix of the smoothed state vector:

$$\mathbf{C}^n_k = \mathbf{C}_k + \mathbf{A}_k ( \mathbf{C}^n_{k+1} - \mathbf{C}^k_{k+1} ) \mathbf{A}_k^T$$

Smoothed residuals:

$$\mathbf{r}_k^n = \mathbf{m}_k - \mathbf{H}_k \tilde{\mathbf{x}}_k^n = \mathbf{r}_k - \mathbf{H}_k (\tilde{\mathbf{x}}_k^n - \tilde{\mathbf{x}}_k)$$

Covariance matrix of the smoothed residuals:

$$\mathbf{R}_k^n = \mathbf{R}_k - \mathbf{H}_k \mathbf{A}_k (\mathbf{C}_{k+1}^n - \mathbf{C}_{k+1}^k) \mathbf{A}_k^T \mathbf{H}_k^T = \mathbf{V}_k - \mathbf{H}_k \mathbf{C}_k^n \mathbf{H}_k^T$$

An inspection of the formulae reveals the following facts:

- The gain matrix formalism and the weighted means formalism of the filter are equivalent. The choice between the two depends on the dimensions of the state vector and the measurement vector. If the dimension of the state vector is small, the computation by weighted means is usually faster.
- The filtered estimate of the state vector is unbiased and has minimum variance among all linear estimates using the same set of measurements. For Gaussian process noise and measurement errors it is efficient. The same is true for the smoothed estimates. Therefore the Kalman filter with a subsequent smoothing is equivalent to a global linear least squares fit which takes into account all correlations arising from the process noise.
- The computation time of the filter is basically proportional to the number of detectors and depends (in the weighted means formalism) very little on the number of measurements per detector.

If the intermediate results of the filter are retained the smoother consists only of a few matrix multiplications and is thus very fast, as long as the dimension of the state vector is small.

- $\mathbf{C}_k$  may also be expressed by the formula

$$\mathbf{C}_k = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{C}_{k-1} (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k)^T + \mathbf{K}_k \mathbf{V}_k \mathbf{K}_k^T.$$

One can show that this form is computationally superior to the form given above, although it consumes more computer time [58, p.305].

- If there is no process noise, i.e.  $\mathbf{Q}_k = \mathbf{O}$  for all  $k$ , smoothing is equivalent to back extrapolation, as can be verified directly from the smoother equations. In this situation a global fit is to be preferred.
- Inspection of the covariance matrix update equations gives the following results, which are intuitively obvious:

The variance of the filtered state vector is smaller than the variance of the predicted state vector (information from the measurement  $\mathbf{m}_k$ ); the mean squared filtered residual is smaller than the mean squared predicted residual (the state vector is pulled towards the

measurement); the variance of the smoothed state vector is smaller than the variance of the filtered state vector (information from all measurements); the mean squared smoothed residual is larger than the mean squared filtered residual (the state vector is pulled towards the true value by the additional constraints).

- The filtered residual vectors (also called innovations) are uncorrelated, in the Gaussian case even independent. This is a characteristic property of the Kalman filter. It also proves the chi-square update formula.

If  $\mathbf{h}_k$  is a non-linear function, the filter is computed by approximating  $\mathbf{h}_k$  by a linear function:

$$\mathbf{h}_k(\mathbf{x}_k) \approx \mathbf{h}_k(\tilde{\mathbf{x}}^{k-1}_k) + \mathbf{H}_k (\mathbf{x}_k - \tilde{\mathbf{x}}^{k-1}_k), \quad \mathbf{H}_k = (\partial \mathbf{h}_k / \partial \mathbf{x}_k)(\tilde{\mathbf{x}}^{k-1}_k).$$

The update of the state vector now reads

$$\tilde{\mathbf{x}}_k = \tilde{\mathbf{x}}^{k-1}_k + \mathbf{K}_k [ \mathbf{m}_k - \mathbf{h}_k(\tilde{\mathbf{x}}^{k-1}_k) ].$$

If necessary, the filter is iterated by reexpanding the function  $\mathbf{h}_k$  in the point  $\tilde{\mathbf{x}}_k$ , which is a better estimate of  $\mathbf{x}_k$  than  $\tilde{\mathbf{x}}^{k-1}_k$ . This is called an iterated Kalman filter. The decision whether an iteration is required is made on the basis of the second derivatives of  $\mathbf{h}_k$ : No further iteration is required if the influence of the second order term in the expansion of  $\mathbf{h}_k$  on the estimate is small as compared with the standard deviation of the estimate:

$$\{ \mathbf{K}_k [ \frac{1}{2} (\tilde{\mathbf{x}}_k - \tilde{\mathbf{x}}^{k-1}_k)^T \frac{\partial^2 \mathbf{h}_k}{\partial \mathbf{x}_k \partial \mathbf{x}_k} (\tilde{\mathbf{x}}_k - \tilde{\mathbf{x}}^{k-1}_k) ] \}_i \ll \sqrt{(\mathbf{C}_k)_{ii}}.$$

Note that  $\mathbf{h}_k$  is a vector of functions, and that the expression in square brackets is a vector of the dimension of  $\mathbf{h}_k$ . The decision does not need to be made for each track, but is made once for all on the basis of a representative sample of tracks.

If  $\mathbf{f}_k$  is non-linear, the predicted covariance matrix  $\mathbf{C}^{k-1}_k$  is computed from the approximation of  $\mathbf{f}_k$  by a linear function:

$$\begin{aligned} \mathbf{f}_k(\mathbf{x}_{k-1}) &\approx \mathbf{f}_k(\tilde{\mathbf{x}}_{k-1}) + \mathbf{F}_k (\mathbf{x}_{k-1} - \tilde{\mathbf{x}}_{k-1}), \quad \mathbf{F}_k = (\partial \mathbf{f}_k / \partial \mathbf{x}_{k-1})(\tilde{\mathbf{x}}_{k-1}), \\ \mathbf{C}^{k-1}_k &= \mathbf{F}_k \mathbf{C}_{k-1} \mathbf{F}_k^T + \mathbf{Q}_k. \end{aligned}$$

The natural choice of the expansion point is the filtered estimate  $\tilde{\mathbf{x}}_{k-1}$ .

Apart from the state vector extrapolation which now reads

$$\tilde{\mathbf{x}}^{k-1}_k = \mathbf{f}_k(\tilde{\mathbf{x}}_{k-1}),$$

the prediction, filter and smoother equations remain the same. This is called an extended Kalman filter. If both  $\mathbf{h}_k$  and  $\mathbf{f}_k$  are non-linear functions, the extended and the iterated filter have to be combined. If  $\tilde{\mathbf{x}}_{k-1}$  is far away from  $\mathbf{x}_{k-1}$ , the prediction  $\tilde{\mathbf{x}}^{k-1}_k$  may be heavily biased. In this case a special trick has to be applied (see Subsection 4.2.3).

## 4.2 Estimation of track parameters with the Kalman filter

### 4.2.1 The track as a discrete dynamic system

If we want to apply the Kalman filter to track fitting we have to interpret the trajectory in space as a dynamic system. This can be done quite naturally by identifying the **state vector** of the dynamic system with the **vector  $x$**  of the **five parameters** which describe the track uniquely at any given surface. We have seen earlier that the space points (and possibly directions) measured by a tracking detectors can be considered as lying on a fixed surface defined by the shape of the detector, which is usually either a plane or a cylinder. Consequently only two coordinates (and the directions) are affected by measurement errors, whereas the third one is known precisely. For the purpose of track fitting it is sufficient to consider the state vector in a discrete set of points, namely in the intersection points of the trajectory with those surfaces where measurements are available. Let us write  $x_k$  for the state vector of the track on the detector surface with the number  $k$ .

### 4.2.2 The system and the measurement equation

If the state vector, i.e. the vector of track parameters, is known on surface  $k-1$ , the trajectory can be extrapolated to surface  $k$  by means of the track model discussed in Subsection 3.2.1. But in the presence of multiple scattering the extrapolated track does not coincide with the actual physical track. If the perturbation of the track by **multiple scattering** is interpreted as a **process noise**, we can write down the following system equation:

$$x_k = f_k(x_{k-1}) + w_k \quad (1 \leq k \leq n).$$

$f_k$  is the **track propagator** from surface  $k-1$  to surface  $k$ ; we recall that it is computed either analytically or by numerical integration of the equation of motion, or else by an approximating analytical function. The random vector  $w_k$  describes the random deviation of the actual track from the extrapolated one due to multiple scattering between surface  $k-1$  and surface  $k$ . We recall that  $w_k$  is in good approximation normally distributed and that

$$E \{ w_k \} = 0, \quad cov \{ w_k \} = Q_k \quad (1 \leq k \leq n).$$

$Q_k$  is computed as indicated in Subsection 3.2.2, by integrating over an infinite number of infinitesimal scattering processes.

The quantities measured by detector  $k$  depend in some way on the state vector of the track in detector  $k$ . This dependence yields the measurement equation, which in general is non-linear:

$$\mathbf{m}_k = \mathbf{h}_k(\mathbf{x}_k) + \varepsilon_k, \quad \mathcal{E} \{ \varepsilon_k \} = \mathbf{0}, \quad \text{cov} \{ \varepsilon_k \} = \mathbf{V}_k = \mathbf{G}_k^{-1} \quad (1 \leq k \leq n).$$

The expectation of the measurement errors is in fact equal to 0 in virtually all tracking detectors, with the possible exception of boundary effects. If it is not equal to 0, the bias can be removed by a suitable modification of the function  $\mathbf{h}_k$ . The covariance matrix  $\mathbf{V}_k$  should only depend weakly on the track parameters  $\mathbf{x}_k$ .

The measurement equation can in principle always be made linear by the appropriate choice of the coordinate system in which the state vector is expressed. It is, however, more important that the track propagator  $\mathbf{f}_k$  is linear or nearly linear, so that  $\tilde{\mathbf{x}}_k^{k-1}$  is (nearly) unbiased for all  $k$ .

The covariance matrix of the measurement errors is known from theoretical considerations or from a calibration experiment (see Subsection 3.2.2). If the tracks are sufficiently well defined by other detectors, it can also be determined from the experimental data. We note that all  $\varepsilon_k$  and all  $\mathbf{w}_k$  are actually stochastically independent.

### 4.2.3 The estimation procedure

The track parameters are estimated by an extended Kalman filter which proceeds from one end of the track to the other one. At the end of the filter the information from all detectors is included in the estimate, with the exception of measurements which obviously do not belong to the track (see Section 4.3). **After a subsequent smoothing estimates of the track parameters using the full information are available on all detector surfaces.** This procedure has the following advantages as compared with the global fit:

- Extrapolations from both ends of the track as well as intersections of the trajectory with other detectors can be computed using the full information.
- The computation time of the filter is always proportional to the number of detectors in the track fit, independently of whether there is multiple scattering or not.
- The filter can be used for track finding and track fitting at the same time, provided that the track density is not too high. If the track density is high, track finding with the filter runs into combinatorial difficulties, since at any point several measurements may be compatible with the predicted state vectors. This is particularly serious in the first few steps of the track fit, when the state vector is only poorly defined.



- The linear approximation of the track model does not need to be valid along the entire trajectory, but only from one detector to the next one.

The time spent in the calculation of the track model, of its derivatives and of the covariance matrix of multiple scattering is the same with the Kalman filter as with a global fit; what is saved is the inversion of the global covariance matrix.

In order to be able to start the filter we need an initial value  $x_0$  with its covariance matrix  $C_0$ , except in the case that  $x_1$  can be uniquely determined from  $m_1$  alone. In all other cases we may take as  $x_0$  a rough estimate of the track parameters which is usually supplied by the track search. In order not to bias the exact filter,  $x_0$  should be given an "infinite" covariance matrix. In practice,  $C_0$  is a matrix with sufficiently large diagonal elements and zeroes otherwise.

In specific cases it may occur that the state vector is only poorly defined by the first measurement(s), and the extrapolation may go astray, although the track is in principle well defined by the subsequent measurements. Consider, for example, a track in the forward region of DELPHI (Fig. 5.9). The track fit starts in FCB and proceeds towards the vertex. The momentum measurement of FCB is rather poor, so that the prediction into FCA may be quite far from the measurement in FCA. In extreme cases the extrapolated track may even fail to intersect with the next detector. These problems are due to the fact that for a non-linear track model the prediction is not unbiased. If the estimated state vector is far away from the true value, the bias of the subsequent prediction may become arbitrarily large. In such a case the following trick can be applied [61]:

Instead of the state vector itself the **difference** between the state vector and some fixed **reference track** is filtered and smoothed, by a linear expansion of the track model around the reference track. If the reference track is close to the real track, the bias of the prediction is much reduced this way. The reference track is either supplied by the track search or is computed from a measurement which defines the track sufficiently well. In our example the reference track is taken from the results of the track search, which combines the information from all detectors in the forward region of DELPHI, although not in a statistically optimal way.

Another kind of problem may arise if the distance between two subsequent detectors is very large: The correlation between the track position and the track direction can get very close to 1 in the predicted covariance matrix. As a consequence, the latter becomes numerically singular. In this case the error propagation and the inversion have to be computed in double precision arithmetic.

#### 4.2.4 Merging of track segments

In a large modular detector like DELPHI it is frequently the case that a track is reconstructed not as a whole, but in several pieces which have then to be put together. We show now that this can be done efficiently with the Kalman filter/smoothen algorithm. For the sake of simplicity we will restrict ourselves to the case of two track segments; the generalization to several segments is obvious.

Let us assume that two track segments have been fitted individually in two separate detector modules (Fig. 4.1) with state vectors  $\tilde{\mathbf{x}}_j^i$  and  $\tilde{\mathbf{y}}_1^k$ , respectively. In order to merge the information from the two track segments we proceed as follows:

- Take the smoothed estimate  $\tilde{\mathbf{y}}_1^m$  at the start of track segment 2.
- Extrapolate  $\tilde{\mathbf{y}}_1^m$  to the end of track segment 1, i.e the surface of  $\tilde{\mathbf{x}}_n$ , including error propagation and multiple scattering. The extrapolation is called  $\tilde{\mathbf{y}}_1'$ .
- Compute the weighted mean of  $\tilde{\mathbf{y}}_1'$  and  $\tilde{\mathbf{x}}_n$  and store it in  $\tilde{\mathbf{x}}_n'$ .
- (Re)compute the smoother in track segment 1, starting with  $\tilde{\mathbf{x}}_n'$  in place of  $\tilde{\mathbf{x}}_n$ .

Note that only the smoother has to be (re)computed in all track segments, plus  $n-1$  extrapolation steps between adjacent segments. It is essential, however, that all segments are filtered along the same direction, and that the intermediate results of the filter are retained in all segments. This has to be taken into consideration during the design of the data structures of the track fitting program.

#### 4.2.5 Removal of measurements

Sometimes the necessity arises to **remove a measurement from a track**, for instance, if it is recognized as an outlier (see next section). In addition, it may be useful to compute an estimate  $\tilde{\mathbf{x}}_k^{n*}$  which contains the **full information with the exception of detector k**, e.g. for a check of the detector alignment and resolution with real tracks.  $\tilde{\mathbf{x}}_k^{n*}$  can be calculated by an "inverse Kalman filter" in the following way [57]. Since  $\tilde{\mathbf{x}}_k^{n*}$  is obtained from the smoothed state vector  $\tilde{\mathbf{x}}_k^n$  by removing  $\mathbf{m}_k$ ,  $\tilde{\mathbf{x}}_k^n$  is obtained from  $\tilde{\mathbf{x}}_k^{n*}$  by adding  $\mathbf{m}_k$ :

$$\begin{aligned}\tilde{\mathbf{x}}_k^n &= \mathbf{C}_k^n \cdot [ (\mathbf{C}_k^{n*})^{-1} \tilde{\mathbf{x}}_k^{n*} + \mathbf{H}_k^T \mathbf{G}_k \mathbf{m}_k ], \\ \mathbf{C}_k^n &= [ (\mathbf{C}_k^{n*})^{-1} + \mathbf{H}_k^T \mathbf{G}_k \mathbf{H}_k ]^{-1}.\end{aligned}$$

We solve for  $\tilde{\mathbf{x}}_k^{n*}$  and  $\mathbf{C}_k^{n*}$ :

$$\begin{aligned}\tilde{\mathbf{x}}_k^{n*} &= \mathbf{C}_k^{n*} \cdot [ (\mathbf{C}_k^n)^{-1} \tilde{\mathbf{x}}_k^n - \mathbf{H}_k^T \mathbf{G}_k \mathbf{m}_k ], \\ \mathbf{C}_k^{n*} &= [ (\mathbf{C}_k^n)^{-1} - \mathbf{H}_k^T \mathbf{G}_k \mathbf{H}_k ]^{-1},\end{aligned}$$

or, in gain matrix notation:

$$\begin{aligned}\tilde{\mathbf{x}}_k^{n*} &= \tilde{\mathbf{x}}_k^n + \mathbf{K}_k^{n*} (\mathbf{m}_k - \mathbf{H}_k \tilde{\mathbf{x}}_k^n), \\ \mathbf{K}_k^{n*} &= \mathbf{C}_k^n \mathbf{H}_k^T \cdot (-\mathbf{V}_k + \mathbf{H}_k \mathbf{C}_k^n \mathbf{H}_k^T)^{-1}, \\ \mathbf{C}_k^{n*} &= (\mathbf{I} - \mathbf{K}_k^{n*} \mathbf{H}_k) \mathbf{C}_k^n.\end{aligned}$$

Thus the inverse filter is a step of the Kalman filter with the covariance (or weight) matrix of  $\mathbf{m}_k$  taken negative.

The estimates  $\tilde{\mathbf{x}}_k^{n*}$  can be computed for all  $k$  in one single pass of the smoother. If  $\mathbf{m}_k$  has to be removed permanently from the track, the smoother has to be (re)computed with  $\tilde{\mathbf{x}}_k^{n*}$  and  $\mathbf{C}_k^{n*}$  in place of  $\tilde{\mathbf{x}}_k^n$  and  $\mathbf{C}_k^n$ . However, this does not update the smoothed estimates  $\tilde{\mathbf{x}}_j^n$  with  $j > k$ . If these are important the filter has to be recomputed, starting from  $\tilde{\mathbf{x}}_k^{k-1}$  and without using  $\mathbf{m}_k$ , followed by a pass of the smoother over the entire track.

## 4.3 Treatment of outliers in track reconstruction

### 4.3.1 Origin and modelling of outliers

In the context of track reconstruction an outlier is defined as a measurement which does not follow the expected behaviour. This may be put into statistical terms by saying that a measurement is considered as an outlier whenever its distance from the true track position is large under the assumption of normal measurement errors, the distance being expressed in terms of the covariance matrix attached to the measurement.

Outliers in track reconstruction may arise in different contexts. The simplest, but fairly frequent case is the presence of **ambiguities** in the track candidate. As we have seen in Subsection 2.2.2, some tracking detectors, in particular drift chambers, give rise to ambiguous information. The track search, being less restrictive than a rigorous track fit, is not always able to decide which of the two possible solutions is the correct one, and the decision must be deferred to the track fit. In the track fit the wrong solution is regarded as an outlier which has to be spotted.

This is an example of an outlier which is **track correlated**: It is a measurement generated by the particle the track of which is to be fitted, but it looks incompatible with the rest of the track. Another example of a track correlated outlier is the following one: In a gaseous tracking detector, e.g. a MWPC or a drift chamber, it happens occasionally that an electron knocked out from the gas molecule by the ionizing traversing particle has an energy which is so large

that it produces itself a long tail of secondary ionization along its path in the detector. These energetic electrons are called delta-rays. As a result, the signal produced by the incident particle is distorted by the additional ionization, the effect of which is a shift in the measured coordinate (see Fig. 2.10). Finally, it might be the case that the distribution of the measurement errors is far from being normal. Although the bulk of the data follows a normal distribution in most tracking detectors, there is nearly always a small fraction of the data which deviate from the normal law. These data show up as long tails in the error distribution and should be regarded as outliers. This does not necessarily imply that they should be discarded.

A second class comprises outliers which are **uncorrelated with the track**, i.e. signals which are not caused by the track, but are nevertheless picked up by the track search. They may be signals from adjacent tracks or genuine noise, either in the detector or in the electronics.

The distribution of track correlated outliers around the true track position depends first of all on the properties of the detector, and also, in the case of delta-rays, on the physics of the underlying secondary process. The distribution of track uncorrelated outliers is in principle uniform in the whole detector volume; in practice their distribution is imposed by the selection mechanism of the track search.

It is therefore very difficult, if not impossible, to specify the distribution of possible outliers a priori, only from theoretical considerations. The only practicable approach is the determination of the outlier distribution either from a detailed simulation or from the real data, i.e. from tracks which are well reconstructed in the adjacent tracking detectors.

There are two basic ways of modelling outliers in a regression problem [62]. The first one is the **mean-shift** model. It describes an outlier by the fact that its expectation is shifted by a certain amount with respect to the true track position:

$$\mathbf{m}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{d}_k + \varepsilon_k.$$

The mean-shift model describes well outliers picked up from adjacent tracks, outliers caused by delta-rays and wrong solutions of ambiguous measurements. The offset  $\mathbf{d}_k$ , however, is variable, so that the mean-shift model is not a very practical way to describe a large number of outliers.

A more useful description is obtained by considering the p.d.f. of outlying measurements. Since outlying measurements are farther away from the true track position than is indicated by their covariance matrix, their empirical distribution has **longer tails** and consequently **larger variance** than the distribution of "good" measurements. This is to say that the p.d.f. of a measurement has a larger variance than usual, if the measurement is an outlier:

$$g_k(\mathbf{m}_k - \mathbf{H}_k \hat{\mathbf{x}}_k \mid \mathbf{m}_k \text{ is outlier}) = g_k(\mathbf{0}, \mathbf{V}_k^{(2)}), \mathbf{V}_k^{(2)} > \mathbf{V}_k^{(1)},$$

where  $\mathbf{V}_k^{(1)}$  is the covariance matrix of good measurements in detector  $k$ . It is also assumed that the outliers are not biased, i.e. do not prefer a particular direction. This is called a **variance-inflation model** of outliers.

If the frequency of outliers is known, the unconditional p.d.f. of all measurements can be written down:

$$g_k = a_k^{(1)} \cdot \mathcal{N}(\mathbf{0}, \mathbf{V}_k^{(1)}) + a_k^{(2)} \cdot g_k^{(2)}(\mathbf{0}, \mathbf{V}_k^{(2)}), a_k^{(1)} + a_k^{(2)} = 1, \mathbf{V}_k^{(2)} > \mathbf{V}_k^{(1)},$$

where  $g_k^{(2)}$  is some p.d.f. with mean  $\mathbf{0}$  and covariance matrix  $\mathbf{V}_k^{(2)}$ , and  $a_k^{(2)}$  is the probability of an outlier occurring in detector  $k$ . Due to lack of information one is usually forced to assume that  $g_k^{(2)}$  is also a normal p.d.f. This is called a **mixture model** of outliers.

Being a least-squares method, the Kalman filter is rather sensitive to outliers, which may bias the final estimate unduly. If we consider a large sample of tracks, such a bias results in an inflated standard error of the estimate, which is not accounted for in the corresponding covariance matrix. One way of avoiding this is to try to **identify** outliers and to **remove** them from the track. An algorithm which accomplishes this for single and multiple outliers is presented in Subsection 4.3.2. Another solution consists in **accommodating** the outlier(s) in the estimate by giving them a **smaller weight**. This may be called a "robustification" of the Kalman filter and is discussed in Subsection 4.3.3, for the case of a mixture model.

Obviously outliers which are uncorrelated with the track should always be removed since they bear no information pertaining to the track. Of course, one can never be sure to which type an outlier belongs, so that downweighting of outliers is a somewhat dangerous procedure – except in the case of truly non-Gaussian measurement errors, where also outlying observations carry useful information (see Subsection 4.3.3).

### 4.3.2 Identification of outliers by chi-square tests

It is well known that least-squares estimators are sensitive to outliers. In the global method of track fitting, the total chi-square of the fit and the studentized residuals of the measurements (called "pulls" by particle physicists) can be used to check whether there are outliers present. Although the total chi-square is a powerful test against ghost tracks (see Subsection 2.3.3), it loses its power against single outliers with increasing number of measurements. The inspection of the largest studentized residual is a valid criterion for single outliers, but only if there is no or only little multiple scattering. If there is strong multiple scattering, the studentized residuals lose their power as test criteria, since the outliers are masked by the multiple scattering. An example showing this effect will be given in Chapter 5. Furthermore removal

of an outlier requires the weight matrix of all remaining measurements to be recomputed. We may conclude that it seems difficult to handle outliers with the global method in the presence of multiple scattering.

With the Kalman filter the situation is much better as ways can be found to recognize outliers and to remove them from the track with little additional computational effort.

Since an outlier in track fitting is defined as a measurement which is in some sense "too far away" from the true track position (and direction), it is natural to take as a decision criterion the **distance of the measurement from the estimated track position**. By using the predicted track position we obtain the following test statistic:

$$\chi^2_{k,P} = (\mathbf{r}^{k-1}_k)^T (\mathbf{R}^{k-1}_k)^{-1} \mathbf{r}^{k-1}_k.$$

It can easily be shown that the "predicted chi-square"  $\chi^2_{k,P}$  is equal to the "filtered chi-square"  $\chi^2_{k,F}$ :

$$\chi^2_{k,P} = \chi^2_{k,F} = \mathbf{r}_k^T \mathbf{R}_k^{-1} \mathbf{r}_k.$$

If there is no outlier at  $j < k$ , the filtered state vector is unbiased and normally distributed. If  $\mathbf{m}_k$  is not an outlier then  $\chi^2_{k,F}$  is  $\chi^2$ -distributed with  $m_k = \dim(\mathbf{m}_k)$  degrees of freedom. If  $\mathbf{m}_k$  is a mean-shift outlier with offset  $\mathbf{d}_k$  with respect to the true state vector:

$$\mathbf{m}_k = \mathbf{H}_k \tilde{\mathbf{x}}_k + \mathbf{d}_k + \varepsilon_k,$$

then  $\chi^2_{k,F}$  follows a non-central  $\chi^2$ -distribution with  $m_k$  degrees of freedom and non-central parameter  $q_{k,F}$ :

$$q_{k,F} = \mathbf{d}_k^T \mathbf{G}_k \mathbf{R}_k \mathbf{G}_k \mathbf{d}_k.$$

If  $\chi^2_{k,F}$  is larger than a given bound  $c$ ,  $\mathbf{m}_k$  is rejected as an outlier. If  $c$  is chosen as the  $(1-\alpha)$ -quantile of the  $\chi^2$ -distribution with  $m_k$  degrees of freedom, the size of the test is equal to  $\alpha$ . The probability of rejecting an outlier, i.e. the power of the test, depends on  $\alpha$ ,  $m_k$  and  $q_{k,F}$  [57]:

$$p(m_k, q_{k,F}, \alpha) = \alpha + \exp(-q_{k,F}/2) \exp(-c/2) \sum_{i=1}^{\infty} \frac{q_{k,F}^i}{2^i i!} \sum_{j=1}^i \frac{c^j}{2^j j!}.$$

If there is an (unrecognized) outlier at  $j < k$ , then  $\tilde{\mathbf{x}}_k$  is biased and  $\chi^2_{k,F}$  is no longer exactly  $\chi^2$ -distributed. Note, however, that a bias of the prediction  $\tilde{\mathbf{x}}^{k-1}_k$  is damped by the filter:

$$\text{bias}(\tilde{\mathbf{x}}_k) = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \text{bias}(\tilde{\mathbf{x}}^{k-1}_k).$$

A second test can be constructed by using the distance of the measurement from the smoothed track positions as a test statistic:

$$\chi^2_{k,S} = \mathbf{r}_k^{\text{T}} (\mathbf{R}_k^{\text{n}})^{-1} \mathbf{r}_k.$$

If  $\mathbf{m}_k$  is an outlier as above, and if it is the only outlier,  $\chi^2_{k,S}$  is again non-centrally  $\chi^2$ -distributed, but this time with a different non-central parameter  $q_{k,S}$  [57]:

$$q_{k,S} = \mathbf{d}_k^{\text{T}} \mathbf{G}_k \mathbf{R}_k^{\text{n}} \mathbf{G}_k \mathbf{d}_k.$$

Since the smoothed track position is better defined than the filtered one, this test should be more powerful than the previous one. Indeed we have:

$$q_{k,S} - q_{k,F} = \mathbf{d}_k^{\text{T}} \mathbf{G}_k (\mathbf{R}_k^{\text{n}} - \mathbf{R}_k) \mathbf{G}_k \mathbf{d}_k = \mathbf{d}_k^{\text{T}} \mathbf{G}_k \mathbf{H}_k (\mathbf{C}_k - \mathbf{C}_k^{\text{n}}) \mathbf{H}_k^{\text{T}} \mathbf{G}_k \mathbf{d}_k \geq 0,$$

since it can easily be shown by induction that  $\mathbf{C}_k - \mathbf{C}_k^{\text{n}}$  is positive semidefinite.

If there is an outlier at  $j \neq k$ ,  $\tilde{\mathbf{x}}_k^{\text{n}}$  is biased and  $\chi^2_{k,S}$  is not exactly  $\chi^2$ -distributed. This implies that the size of the test cannot be controlled precisely and has to be tuned with simulated data. The problem is aggravated if there are several outliers. The results of a related study will be presented in Chapter 5.

We conclude that a **rough selection of measurements should take place during filtering**, whereas the **final search for possible outliers should be carried out during smoothing**. If an outlier is found it can be removed from the track by means of the algorithm given in Subsection 4.2.5. If necessary, the filter has to be recomputed without the outliers.

Finally a few words about the choice of  $\alpha$  should be said. If  $\alpha$  is chosen too large, the variance of the final estimated track parameters will increase, because many "good" measurements are rejected as outliers. If  $\alpha$  is too small, outliers will remain unidentified, and the estimated track parameters will be biased. It is impossible to say in general which of the two effects is more harmful, because that depends on the number of measurements in the track, on the probability of an outlier occurring, and on the distribution of the outliers. Therefore the optimum value of  $\alpha$  can only be found by simulation studies, which try to mimic all processes leading to outliers as closely as possible. This requires feedback from detector experts and from real data.

The number of outliers that can be removed from a track is also restricted by the demand that the track is uniquely defined by the remaining measurements. Thus in some cases it may be better to keep an outlier instead of losing the whole track. As a rule, the track should not contain less than seven independent measurements.

### 4.3.3 A robustified Kalman filter

If the behaviour of outliers can be described by a mixture model, a robustified Kalman filter which takes into account the specific distribution of outliers can be constructed via a Bayesian approach [63]. It is almost as fast as the standard Kalman filter and is very easy to implement.

We start from a mixture model in which the distribution of measurements and outliers is described by a mixture of two normals:

$$\begin{aligned} \mathbf{m}_k &= \mathbf{H}_k \mathbf{x}_k + \boldsymbol{\varepsilon}_k, \\ \boldsymbol{\varepsilon}_k &\sim a_k^{(1)} \mathcal{N}(\mathbf{0}, \mathbf{V}_k^{(1)}) + a_k^{(2)} \mathcal{N}(\mathbf{0}, \mathbf{V}_k^{(2)}), \\ a_k^{(1)} + a_k^{(2)} &= 1, \mathbf{V}_k^{(2)} > \mathbf{V}_k^{(1)}, a_k^{(1)} \gg a_k^{(2)}, \end{aligned}$$

where  $a_k^{(2)}$  is the probability of an outlier occurring in detector  $k$ . This model is a reasonable assumption for measurements of positions and derived quantities, like directions or curvature. It is inadequate for measurements of energy loss which involve long-tailed distributions of a quite different type. It should be noted that the outliers lying within the range of the distribution of "good" measurements will appear as such and cannot be recognized.

The distribution of  $\mathbf{x}^{k-1}_k$  is assumed to be normal:

$$\begin{aligned} \mathbf{x}^{k-1}_k &= \tilde{\mathbf{x}}^{k-1}_k + \boldsymbol{\delta}_k, \\ \boldsymbol{\delta}_k &\sim \mathcal{N}(\mathbf{0}, \mathbf{C}^{k-1}_k), \end{aligned}$$

where  $\tilde{\mathbf{x}}^{k-1}_k$  is the predicted estimate obtained with the robust filter. For large values of  $k$  this assumption is not unreasonable because of the Central Limit Theorem.

By means of Bayes's theorem one obtains the following posterior distribution of  $\mathbf{x}_k$  [63]:

$$f(\mathbf{x}_k | \mathbf{m}^{(k)}) = \sum_{i=1}^2 b_k^{(i)} \varphi(\mathbf{x}_k; \tilde{\mathbf{x}}_k^{(i)}, \mathbf{C}_k^{(i)}),$$

with:

$$\begin{aligned} \mathbf{m}^{(k)} &= \{\mathbf{m}_1, \dots, \mathbf{m}_k\}, \\ \tilde{\mathbf{x}}_k^{(i)} &= \tilde{\mathbf{x}}^{k-1}_k + \mathbf{C}^{k-1}_k \mathbf{H}_k^T \mathbf{W}_k^{(i)} \mathbf{r}^{k-1}_k, \\ \mathbf{W}_k^{(i)} &= (\mathbf{V}_k^{(i)} + \mathbf{H}_k \mathbf{C}^{k-1}_k \mathbf{H}_k^T)^{-1}, \\ \mathbf{C}_k^{(i)} &= [(\mathbf{C}^{k-1}_k)^{-1} + \mathbf{H}_k^T \mathbf{G}_k^{(i)} \mathbf{H}_k]^{-1}, \\ \varphi(\mathbf{x}; \boldsymbol{\mu}, \mathbf{C}) &= \text{normal p.d.f. with mean } \boldsymbol{\mu} \text{ and covariance matrix } \mathbf{C}. \end{aligned}$$

The coefficients  $b_k^{(2)}$  and  $b_k^{(1)}$  can be interpreted as the posterior probabilities of  $\mathbf{m}_k$  being an outlier or not:



$$b_k^{(1)} = \left[ 1 + \frac{a_k^{(2)} |W_k^{(2)}|}{a_k^{(1)} |W_k^{(1)}|} \exp\left(\frac{1}{2} r^{k-1}_k{}^T D_k r^{k-1}_k\right) \right]^{-1},$$

$$b_k^{(2)} = 1 - b_k^{(1)},$$

with:

$$r^{k-1}_k = m_k - H_k \tilde{x}^{k-1}_k,$$

$$D_k = W_k^{(1)} - W_k^{(2)}.$$

The final estimate  $\tilde{x}_k$  and its covariance matrix  $C_k$  are obtained as the mean and the covariance matrix of the posterior distribution of  $x_k$ . The update of the state vector turns out to be a weighted sum of two Kalman filters, the weights being  $b_k^{(1)}$  and  $b_k^{(2)}$ :

$$\tilde{x}_k = \tilde{x}^{k-1}_k + C^{k-1}_k H_k^T (b_k^{(1)} W_k^{(1)} + b_k^{(2)} W_k^{(2)}) r^{k-1}_k,$$

$$C_k = C^{k-1}_k - C^{k-1}_k H_k^T (b_k^{(1)} W_k^{(1)} + b_k^{(2)} W_k^{(2)} - S_k) H_k C^{k-1}_k,$$

$$S_k = b_k^{(1)} b_k^{(2)} D_k r^{k-1}_k r^{k-1}_k{}^T D_k.$$

The posterior distribution of  $x_k$ , being a mixture of two normals with **different means**, is **asymmetric**; therefore  $\tilde{x}_k$  is **not** the maximum-likelihood estimate! For  $b_k^{(2)} = 0$  or, a fortiori,  $a_k^{(2)} = 0$ , the robust filter reduces to the standard Kalman filter.

The prior p.d.f. of the residual  $r^{k-1}_k$  is a mixture of two normals with mean zero:

$$f(r^{k-1}_k) = \sum_{i=1}^2 a_k^{(i)} \cdot \varphi(r^{k-1}_k; \mathbf{0}, V_k^{(i)} + H_k C^{k-1}_k H_k^T).$$

Therefore  $R^{k-1}_k$  is given by

$$R^{k-1}_k = a_k^{(1)} V_k^{(1)} + a_k^{(2)} V_k^{(2)} + H_k C^{k-1}_k H_k^T.$$

The filtered residual  $r_k$  is given by

$$r_k = m_k - H_k \tilde{x}_k = [I - H_k C^{k-1}_k H_k^T (b_k^{(1)} W_k^{(1)} + b_k^{(2)} W_k^{(2)})] r^{k-1}_k.$$

We may also compute the derivative of  $r_k$  w.r.t.  $r^{k-1}_k$ :

$$\partial r_k / \partial r^{k-1}_k = I - H_k J_k,$$

$$J_k = C^{k-1}_k H_k^T (b_k^{(1)} W_k^{(1)} + b_k^{(2)} W_k^{(2)} - S_k).$$

$J_k$  may be regarded as a modified gain matrix, which reduces to the Kalman gain matrix  $K_k$  if  $b_k^{(2)} = 0$ .

Unfortunately  $r^{k-1}_k$  cannot be expressed explicitly as a function of  $r_k$ ; otherwise the p.d.f. of  $r_k$  could be computed by using the transformation theorem for probability density functions.

We can, however, compute  $R_k$  by linear error propagation:

$$\mathbf{R}_k = (\mathbf{I} - \mathbf{H}_k \mathbf{J}_k) \mathbf{R}^{k-1}_k (\mathbf{I} - \mathbf{J}_k^T \mathbf{H}_k^T).$$

We look now for a generalized  $\chi^2$ -statistic of the filter step. We observe that the  $\chi^2$ -increment can be computed as in the standard case:

$$\chi^2_{k,F} = \mathbf{r}_k^T \mathbf{G}_k \mathbf{r}_k + (\tilde{\mathbf{x}}_k - \tilde{\mathbf{x}}^{k-1}_k)^T (\mathbf{C}^{k-1}_k)^{-1} (\tilde{\mathbf{x}}_k - \tilde{\mathbf{x}}^{k-1}_k).$$

The weight matrix  $\mathbf{G}_k = \mathbf{V}_k^{-1}$  can be computed either with the prior or with the posterior probabilities, yielding two different statistics. Their statistical properties will be investigated in more detail in Chapter 5.

We have seen that in case the distribution of  $\mathbf{x}^{k-1}_k$  is normal, the posterior distribution of  $\mathbf{x}_k$  is a mixture of two normals. An exact prediction to the next step would result in a posterior of  $\mathbf{x}_{k+1}$  which would be a mixture of four normals, and continuing this way would yield an exponentially increasing number of terms. In order to keep the filter simple, the posterior distribution of  $\mathbf{x}_k$  is approximated in each step by a **single normal** with covariance matrix  $\mathbf{C}_k$ . It has been shown that this approximation is optimal in the sense that it minimizes the Kullback-Liebler distance between the two density functions [64]. However, it should be kept in mind that the filtered estimate, the residuals and the  $\chi^2$ -statistics do not follow an exact normal and  $\chi^2$ -distribution, respectively. The statistical properties of these quantities will be investigated in Chapter 5 by large samples of simulated tracks.

The smoother is not affected by the robustification of the filter and remains the same.

Since the robustified filter takes into account the specific distribution of the outliers, it may be expected to yield better results than a standard filter, in which the mixture of measurement and outlier distributions is approximated by a single normal with the same covariance matrix:

$$\begin{aligned} \mathbf{m}_k &= \mathbf{H}_k \hat{\mathbf{x}}_k + \varepsilon_k, \\ \varepsilon_k &\sim \mathcal{N}(\mathbf{0}, \mathbf{a}_k^{(1)} \mathbf{V}_k^{(1)} + \mathbf{a}_k^{(2)} \mathbf{V}_k^{(2)}). \end{aligned}$$

A detailed comparison of the performance of the standard filter and of the robust filter will be presented in Chapter 5. Note that the standard filter is the optimal linear filter and yields exact mean-squared quantities, like the average chi-square or the variance of the state vector and of the residuals. The distribution of these quantities, however, is distorted with respect to a chi-square or normal distribution.

Finally, it should be repeated that the robust filter is equally useful if the distribution of the measurement errors is not normal, but can be approximated reasonably well by a mixture of two normals.

The robust filter accommodates outliers by giving them a smaller weight than to "good" measurements; this property should render a  $\chi^2$ -cut on the measurements unnecessary. If the se-

lection criteria of the track search, however, are too weak, the generalized chi-square must be used to reject measurements which are incompatible with the state vector even with the long-tailed error distribution. This is a nuisance, since its exact distribution is not known. Note that the  $\chi^2$ -statistic using the prior probabilities must be employed to reject measurements or to resolve ambiguous measurements.

Another situation in which the robust filter may come in useful, is an eventual non-normal behaviour of the process noise. We have seen that multiple Coulomb scattering leads to a nearly normal process noise. But there are also other processes which contribute to the interaction of a particle with matter, like nuclear scattering and hard electromagnetic scattering. These occur rarely, but with large scattering angles. Hard electromagnetic scattering is particularly serious for electrons, because of their low mass. Hadrons may undergo nuclear scattering in the beam tube or in other massive walls in the detector.

One can try to take these rare processes into account by a mixture model of the process noise:

$$\begin{aligned} \mathbf{w}_k &\sim a_k^{(1)} \mathcal{N}(\mathbf{0}, \mathbf{Q}_k^{(1)}) + a_k^{(2)} \mathcal{N}(\mathbf{0}, \mathbf{Q}_k^{(2)}), \\ a_k^{(1)} + a_k^{(2)} &= 1, \mathbf{Q}_k^{(2)} > \mathbf{Q}_k^{(1)}, a_k^{(1)} \gg a_k^{(2)}. \end{aligned}$$

Then the distribution of  $\mathbf{x}^{k-1}_k$  is a mixture of two normals:

$$f(\mathbf{x}^{k-1}_k | \mathbf{m}^{(k-1)}) = \sum_{i=1}^2 a_k^{(i)} \varphi(\mathbf{x}_k; \tilde{\mathbf{x}}^{k-1}_k, \mathbf{C}^{k-1}_k{}^{(i)}),$$

with:

$$\mathbf{C}^{k-1}_k{}^{(i)} = \mathbf{F}_k \mathbf{C}_{k-1} \mathbf{F}_k^T + \mathbf{Q}_k^{(i)}.$$

If the distribution of  $\mathbf{m}_k$  is assumed to be normal with covariance matrix  $\mathbf{V}_k$ , one obtains the following posterior p.d.f. of  $\mathbf{x}_k$ :

$$f(\mathbf{x}_k | \mathbf{m}^{(k)}) = \sum_{i=1}^2 b_k^{(i)} \varphi(\mathbf{x}_k; \tilde{\mathbf{x}}_k^{(i)}, \mathbf{C}_k^{(i)}),$$

with:

$$\begin{aligned} \tilde{\mathbf{x}}_k^{(i)} &= \tilde{\mathbf{x}}^{k-1}_k + \mathbf{K}_k^{(i)} \mathbf{r}^{k-1}_k, \\ \mathbf{K}_k^{(i)} &= \mathbf{C}^{k-1}_k{}^{(i)} \mathbf{H}_k^T (\mathbf{V}_k + \mathbf{H}_k \mathbf{C}^{k-1}_k{}^{(i)} \mathbf{H}_k^T)^{-1}, \\ \mathbf{C}_k^{(i)} &= (\mathbf{I} - \mathbf{K}_k^{(i)} \mathbf{H}_k) \mathbf{C}^{k-1}_k{}^{(i)}, \\ b_k^{(1)} &= \left[ 1 + \frac{a_k^{(2)}}{a_k^{(1)}} \frac{|\mathbf{W}_k^{(2)}|}{|\mathbf{W}_k^{(1)}|} \exp\left(\frac{1}{2} \mathbf{r}^{k-1}_k{}^T \mathbf{D}_k \mathbf{r}^{k-1}_k\right) \right]^{-1}, \\ b_k^{(2)} &= 1 - b_k^{(1)}, \end{aligned}$$

$$\mathbf{W}_k^{(i)} = (\mathbf{V}_k + \mathbf{H}_k \mathbf{C}^{k-1}_{k^{(i)}} \mathbf{H}_k^T)^{-1},$$

$$\mathbf{D}_k = \mathbf{W}_k^{(1)} - \mathbf{W}_k^{(2)}.$$

The filter is again a weighted sum of two Kalman filters:

$$\tilde{\mathbf{x}}_k = b_k^{(1)} \tilde{\mathbf{x}}_k^{(1)} + b_k^{(2)} \tilde{\mathbf{x}}_k^{(2)} = \tilde{\mathbf{x}}_k^{k-1} + (b_k^{(1)} \mathbf{K}_k^{(1)} + b_k^{(2)} \mathbf{K}_k^{(2)}) \mathbf{r}^{k-1}_k,$$

$$\mathbf{C}_k = b_k^{(1)} \mathbf{C}_k^{(1)} + b_k^{(2)} \mathbf{C}_k^{(2)} +$$

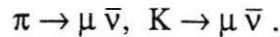
$$+ b_k^{(1)} b_k^{(2)} (\mathbf{K}_k^{(1)} - \mathbf{K}_k^{(2)}) \mathbf{r}^{k-1}_k \mathbf{r}^{k-1}_k{}^T (\mathbf{K}_k^{(1)} - \mathbf{K}_k^{(2)})^T.$$

It seems that the question whether a mixture of two normals is an adequate model for the superposition of multiple Coulomb scattering with nuclear and hard electromagnetic scattering has not yet been investigated until now. Nor are values of  $a_k^{(2)}$  or  $\mathbf{Q}_k^{(2)}$  available in the literature. A study of these points would be an interesting research topic and could yield considerable improvements in the reconstruction of electron tracks.

## 4.4 Recognition of decays (kinks)

### 4.4.1 Physical motivation

We have seen that many of the particles produced in an event are unstable and decay with an exponentially distributed lifetime. Due to the basic law of conservation of electric charge a charged particle decays into an odd number of charged particles, plus a certain number of neutral particles (Fig. 4.2). Of course the total number of decay products is limited by the law of energy conservation. If one of the charged decay products is emitted at a small angle, its track and the track of the decaying particle may be misinterpreted by the track search as a single track, particularly if there is only one charged decay product (Fig. 4.3). For obvious reasons this configuration is called a **kink**. The most frequent sources of kinks are the muonic decays of charged  $\pi$ - and  $K$ -mesons:



If such kinks are not recognized, the muons produced by the decay may be misinterpreted as prompt muons, i.e. as coming from the primary vertex. In addition, the kink may fake a high momentum of the presumed prompt muon, particularly if the muon is emitted in the forward direction, so that the muon momentum is high and the kink angle is small (see below). High-momentum prompt muons are often the signature of an interesting event, e.g. the decay of a

W or a Z vector boson or the semileptonic decay of a heavy quark. Therefore unrecognized kinks are a potential source of background.

The recognition of kinks relies on two effects: first, on the angle formed between the tracks of the decaying meson and of the emerging muon; secondly, on the momentum difference between the meson and the muon. (The latter has nearly always a smaller momentum.) Both the kink angle and the momentum difference are random variables; their distributions depend on the meson mass, on the muon mass and on the momentum of the meson. For a two-body decay like the  $\pi$ - and K-decay under consideration, there is a rather simple and illustrative graphical method of determining the kink angle and the momentum of a decay product, in this case a muon. We first consider the decay in the **rest system** of the meson. In this reference frame the decay is **isotropic**, which means that there is no privileged direction for either of the decay products. (We assume that the meson is in ground state and has no angular momentum.) We now write down the momentum vector  $\mathbf{p}$  of the muon in the following way:

$$\mathbf{p} = \begin{bmatrix} p_L \\ p_T \end{bmatrix},$$

where  $p_L$ , the **longitudinal momentum**, is the component of  $\mathbf{p}$  parallel to the momentum vector  $\mathbf{P}$  of the meson, and  $p_T$ , the **transverse momentum**, is the component orthogonal to  $\mathbf{P}$ . It follows that

$$p_T = p \sin\theta,$$

$$p_L = p \cos\theta,$$

with:

$$p = |\mathbf{p}| = \sqrt{p_T^2 + p_L^2}.$$

Hence  $\mathbf{p}$  lies on a circle of radius  $p$  (Fig. 4.4). The momentum vector of the neutrino is equal to  $-\mathbf{p}$ , because of momentum conservation.  $p$  can be determined from the law of energy conservation:

$$p = \frac{M^2 - m^2}{2M},$$

where  $M$  is the meson mass and  $m$  is the muon mass. The neutrino can be assumed to be massless.

We now transform the decay into the reference frame of the observer. The Lorentz transformation is particularly simple in the coordinate system of  $p_L$  and  $p_T$  [65]:

$$p_T' = p_T = p \sin\theta,$$

$$p_L' = \gamma p_L + \beta \gamma E = \gamma p \cos\theta + \beta \gamma E,$$

with:

$$\gamma = \frac{\sqrt{P^2 + M^2}}{M}, \quad \beta = \frac{P}{\sqrt{P^2 + M^2}}, \quad E = \frac{M^2 + m^2}{2M}.$$

We recall that  $P$  is the momentum of the decaying meson. It is obvious that the circle is transformed into an ellipse, its center being at  $\beta\gamma E$ , and its half-axes being equal to  $p$  and  $\gamma p$ , respectively. The momentum vector  $\mathbf{p}'$  and the kink angle  $\theta'$  of the muon in the observer frame can be read directly from the ellipse (Fig. 4.4). Since the decay is isotropic,  $\cos\theta$  is uniformly distributed in the interval  $[-1,1]$ , and the distribution of  $\theta'$  can be computed in a straightforward manner from the Lorentz transformation.

In order to give an idea of the kink angle and momentum difference to be expected from the decay of a K-meson, Fig. 4.5 shows the kink angle  $\theta'$  and the momentum  $p'$  of the muon as a function of  $\cos\theta$ , for two values of the momentum of the K: 2 GeV/c (Fig. 4.5 a,b) and 40 GeV/c (Fig. 4.5 c,d). In the region around  $\cos\theta=1$  ( $\theta=0^\circ$ ) the chances of detecting the decay are rather small, as both the kink angle and the momentum difference are close to zero. The kink angle then rises with  $\theta$  and attains the maximum at  $\cos\theta \approx -0.95$  ( $\theta \approx 155^\circ$ ); then it drops sharply to 0 at  $\cos\theta=-1$  ( $\theta=180^\circ$ ). But at  $\theta=180^\circ$  the momentum of the muon attains its minimum value close to zero, so that the decay is very likely to be recognized there.

For  $\pi$ -decays (Fig. 4.6) the situation is less favourable: first, the kink angle is much smaller due to the small mass difference between the  $\pi$  and the muon; secondly, also the momentum difference between the  $\pi$  and the muon is less pronounced. The simulation results in Chapter 5 will confirm this.

#### 4.4.2 A kink finding algorithm

We now show that the information provided by the filter and smoother is sufficient to allow the construction of a fast and relatively efficient kink finding algorithm. Normally a track will be subjected to the kink finding procedure only if it is identified as a muon track and if it is in some way suspicious, e.g. if it has an unusually large chi-square. Only in some rare event types all tracks may be passed through the kink finder.

If the track is regarded as a discrete dynamic system a kink is nothing else but a **sudden change of the state vector** (Fig. 4.7), a change both in direction and in momentum (or curvature).

If the kink occurs somewhere between detectors  $k$  and  $k+1$ , we may expect that the state vector  $\tilde{\mathbf{x}}_k$  obtained from the first track segment  $\{\mathbf{m}_1, \dots, \mathbf{m}_k\}$  is significantly different from the back extrapolation  $\tilde{\mathbf{x}}_k^{k+1(b)}$  from the second track segment  $\{\mathbf{m}_{k+1}, \dots, \mathbf{m}_n\}$ . The back

extrapolation is either computed from an actual **backward filter**, i.e. a filter starting at  $\mathbf{m}_n$  and proceeding towards  $\mathbf{m}_1$ , or from the smoother by "removing"  $\tilde{\mathbf{x}}_k$  from  $\tilde{\mathbf{x}}_k^{n_k}$  (the superscript <sup>(b)</sup> denotes the backward filter):

$$\begin{aligned}\tilde{\mathbf{x}}_k^{k+1(b)} &= \mathbf{C}^{k+1}_k(b) ( \mathbf{C}^{n_k-1}_k \tilde{\mathbf{x}}_k^{n_k} - \mathbf{C}_k^{-1} \tilde{\mathbf{x}}_k ), \\ \mathbf{C}^{k+1}_k(b) &= ( \mathbf{C}^{n_k-1}_k - \mathbf{C}_k^{-1} )^{-1}.\end{aligned}$$

In order to decide whether a kink has occurred between detectors  $k$  and  $k+1$ , we can test whether the difference

$$\Delta_k = \tilde{\mathbf{x}}_k^{k+1(b)} - \tilde{\mathbf{x}}_k$$

is significantly different from  $\mathbf{0}$ , by means of the  $\chi^2$ -statistic  $\chi^2_{k,\Delta}$ :

$$\begin{aligned}\chi^2_{k,\Delta} &= \Delta_k^T \cdot (\text{cov} \{ \Delta_k \})^{-1} \Delta_k = \\ &= \Delta_k^T ( \mathbf{C}_k + \mathbf{C}^{k+1}_k(b) )^{-1} \Delta_k = \\ &= \Delta_k^T \mathbf{C}_k^{-1} ( \mathbf{C}_k - \mathbf{C}^{n_k}_k ) \mathbf{C}_k^{-1} \Delta_k.\end{aligned}$$

If there is no kink,  $\chi^2_{k,\Delta}$  is  $\chi^2$ -distributed with 5 degrees of freedom for all  $k$ . Therefore a cut at the  $(1-\alpha)$ -quantile will yield a test of size  $\alpha$ .

This test, however, is not robust against an outlier somewhere along the track. In fact, an outlier leads to a distortion of  $\chi^2_{k,\Delta}$  which as a consequence exceeds the cut more often than expected, thus faking kinks where there are actually none. In addition, the test on  $\chi^2_{k,\Delta}$  does not tell us where the kink is situated, as it is by no means obvious that the largest  $\chi^2_{k,\Delta}$  corresponds to the position of the kink.

These considerations lead us to propose the following modified test. We demand that

- both track segments  $\{\mathbf{m}_1, \dots, \mathbf{m}_k\}$  and  $\{\mathbf{m}_{k+1}, \dots, \mathbf{m}_n\}$  have small total chi-squares
- and that  $\chi^2_{k,\Delta}$  is large.

These conditions could be checked independently, but it is more convenient to combine them into a single test-statistic:

$$F_k = \frac{ ( \chi^2_{k,\Delta} / 5 ) }{ [ ( \chi^2_k + \chi^2_{k+1(b)} ) / n_k ] },$$

where  $\chi^2_k$  is the total chi-square of the first track segment,  $\chi^2_{k+1(b)}$  is the total chi-square of the second track segment, and  $n_k$  is the sum of the respective numbers of degrees of freedom.  $\chi^2_{k+1(b)}$  is either computed during the backward filter or from the following relation:

$$\chi^2_n = \chi^2_k + \chi^2_{k+1(b)} + \chi^2_{k,\Delta}.$$

Let us consider first the case that the track contains no outliers. If there is no kink,  $F_k$  is F-distributed with 5 and  $n_k$  degrees of freedom for all  $k$ . Note that  $F_k$  is also robust against an overall scale error of the measurement variances.

If there is a kink between  $k$  and  $k+1$ , the numerator of  $F_k$  will be large and the denominator will be small, resulting in a large value of  $F_k$ .  $F_k$  has slightly less power than  $\chi^2_{k,\Delta}$  though, as will be shown in Chapter 5. If there is a kink at  $j \neq k$ , one of the chi-squares in the denominator will be large, thereby preventing  $F_k$  from getting too large. Therefore we propose to locate the kink by looking for the largest  $F_k$ .

If there is an outlier somewhere along the track, the situation is more complicated. If there is no kink, but an outlier at  $j$ ,  $\chi^2_{k,\Delta}$  will be distorted, particularly if  $k$  is close to  $j$ . But also one of the chi-squares in the denominator will be large, so that a kink will be faked with much less probability than is the case with the test on  $\chi^2_{k,\Delta}$ . If there is a kink between  $k$  and  $k+1$  and an outlier somewhere, both the numerator and one of the chi-squares in the denominator of  $F_k$  will be large, so that the kink may be partially masked by the outlier. Therefore the detection of kinks has to be preceded by or combined with the removal or accommodation of outliers. The results of a study of these issues with simulated tracks will be presented in Chapter 5.

It may of course happen in practice that two track segments which do not form a kink have been combined into a track candidate, e.g. if two tracks cross in a region of high track density. If there is reason to doubt that an actual kink has been found, for instance from a visual inspection of the track on a graphic screen, an additional test can be imposed on the closest distance in space of the two track segments.

## 4.5 Estimation of vertex parameters with the Kalman filter

### 4.5.1 The interaction vertex as a discrete dynamic system

Throughout this section we assume that  $n$  tracks are to be fitted to a common vertex. For each track participating in the vertex fit, the five estimated track parameters  $\mathbf{p}_k$  are given on a reference surface, together with their covariance matrix  $\mathbf{V}_k$  (Fig. 4.8). The vector  $\mathbf{p}_k$  is considered as a (virtual) measurement in the vertex fit. If required, contributions of multiple scattering between the vertex and the reference surface are added to the matrix  $\mathbf{V}_k$ . The parameters to be estimated are the vertex position  $\mathbf{v}$  and the 3-momentum vectors  $\mathbf{q}_k$  of all tracks.



We now show that the fast vertex fit described in Section 3.3 can also be derived as a Kalman filter [57].

Initially, the vertex is described by the prior knowledge of the vertex position,  $\mathbf{v}_0$  and  $\mathbf{C}_0 = \text{cov} \{ \mathbf{v}_0 \}$ .  $\mathbf{v}_0$  may be a very crude estimate, and  $\mathbf{C}_0$  accordingly very large. Now one track after the other is added to the vertex. In each step of the filter the vertex position is updated, and the state vector is augmented by the 3-momentum vector  $\mathbf{q}_k$  of the track at the vertex. Thus after  $k$  steps the state vector consists of  $\mathbf{v}_k, \mathbf{q}_1, \dots, \mathbf{q}_k$ . The corresponding covariance matrices are

$$\begin{aligned} \mathbf{C}_k &= \text{cov} \{ \tilde{\mathbf{v}}_k \}, \\ \mathbf{D}_i &= \text{cov} \{ \tilde{\mathbf{q}}_i \} \quad (1 \leq i \leq k), \\ \mathbf{E}_i &= \text{cov} \{ \tilde{\mathbf{q}}_i, \tilde{\mathbf{v}}_i \} \quad (1 \leq i \leq k). \end{aligned}$$

The system equation is particularly simple:

$$\mathbf{v}_k^t = \mathbf{v}_{k-1}^t.$$

Note the absence of any process noise.

The measurement equation is non-linear:

$$\mathbf{p}_k = \mathbf{h}_k(\mathbf{v}_k^t, \mathbf{q}_k^t) + \boldsymbol{\varepsilon}_k, \quad \text{cov} \{ \boldsymbol{\varepsilon}_k \} = \text{cov} \{ \mathbf{p}_k \} = \mathbf{V}_k = \mathbf{G}_k^{-1} \quad (1 \leq k \leq n).$$

Note that the estimated track parameters are stochastically independent from each other.

#### 4.5.2 The estimation procedure

Since the measurement equation is non-linear, the state vector is estimated by an iterated Kalman filter. In each step the function  $\mathbf{h}_k$  is expanded into a linear function at some point  $(\mathbf{v}_k^{(0)}, \mathbf{q}_k^{(0)})$ :

$$\begin{aligned} \mathbf{h}_k(\mathbf{v}, \mathbf{q}_k) &= \mathbf{h}_k(\mathbf{v}_k^{(0)}, \mathbf{q}_k^{(0)}) + \mathbf{A}_k (\mathbf{v} - \mathbf{v}_k^{(0)}) + \mathbf{B}_k (\mathbf{q}_k - \mathbf{q}_k^{(0)}) = \\ &= \mathbf{c}_k^{(0)} + \mathbf{A}_k \mathbf{v} + \mathbf{B}_k \mathbf{q}_k, \end{aligned}$$

with:

$$\begin{aligned} \mathbf{A}_k &= (\partial \mathbf{h}_k / \partial \mathbf{v})(\mathbf{v}_k^{(0)}, \mathbf{q}_k^{(0)}), \quad \mathbf{B}_k = (\partial \mathbf{h}_k / \partial \mathbf{q}_k)(\mathbf{v}_k^{(0)}, \mathbf{q}_k^{(0)}), \\ \mathbf{c}_k^{(0)} &= \mathbf{h}_k(\mathbf{v}_k^{(0)}, \mathbf{q}_k^{(0)}) - \mathbf{A}_k \mathbf{v}_k^{(0)} - \mathbf{B}_k \mathbf{q}_k^{(0)}. \end{aligned}$$

$\mathbf{v}_k^{(0)}$  is conveniently taken equal to  $\tilde{\mathbf{v}}_{k-1}$ , while  $\mathbf{q}_k^{(0)}$  is obtained by tracking back from the estimated track parameters  $\mathbf{p}_k$  towards the approximate vertex position  $\mathbf{v}_k^{(0)}$ . Since there is no prior information about  $\mathbf{q}_k$  a zero weight matrix is assigned to the "predicted" vector  $\mathbf{q}^{k-1}_k$ ,

which as a consequence drops out of the filter equations. Then the prediction equations look as follows:

$$\tilde{\mathbf{v}}_k^{k-1} = \tilde{\mathbf{v}}_{k-1},$$

$$\mathbf{C}_k^{k-1} = \mathbf{C}_{k-1},$$

and we have the following filter equations in the weighted means formulation:

$$\begin{bmatrix} \tilde{\mathbf{v}}_k \\ \tilde{\mathbf{q}}_k \end{bmatrix} = \begin{bmatrix} \mathbf{C}_k & \mathbf{E}_k \\ \mathbf{E}_k^T & \mathbf{D}_k \end{bmatrix} \cdot \begin{bmatrix} \mathbf{C}_{k-1}^{-1} \tilde{\mathbf{v}}_{k-1} + \mathbf{A}_k^T \mathbf{G}_k (\mathbf{p}_k - \mathbf{c}_k^{(0)}) \\ \mathbf{B}_k^T \mathbf{G}_k (\mathbf{p}_k - \mathbf{c}_k^{(0)}) \end{bmatrix},$$

with

$$\begin{bmatrix} \mathbf{C}_k & \mathbf{E}_k \\ \mathbf{E}_k^T & \mathbf{D}_k \end{bmatrix} = \begin{bmatrix} \mathbf{C}_{k-1}^{-1} + \mathbf{A}_k^T \mathbf{G}_k \mathbf{A}_k & \mathbf{A}_k^T \mathbf{G}_k \mathbf{B}_k \\ \mathbf{B}_k^T \mathbf{G}_k \mathbf{A}_k & \mathbf{B}_k^T \mathbf{G}_k \mathbf{B}_k \end{bmatrix}^{-1}.$$

After some matrix algebra (see e.g. ref. [66]) we obtain the following results which are identical to the ones in Section 3.3:

$$\tilde{\mathbf{v}}_k = \mathbf{C}_k [ \mathbf{C}_{k-1}^{-1} \tilde{\mathbf{v}}_{k-1} + \mathbf{A}_k^T \mathbf{G}_k^B (\mathbf{p}_k - \mathbf{c}_k^{(0)}) ],$$

$$\tilde{\mathbf{q}}_k = \mathbf{W}_k \mathbf{B}_k^T \mathbf{G}_k (\mathbf{p}_k - \mathbf{c}_k^{(0)} - \mathbf{A}_k \tilde{\mathbf{v}}_k),$$

$$\mathbf{C}_k = (\mathbf{C}_{k-1}^{-1} + \mathbf{A}_k^T \mathbf{G}_k^B \mathbf{A}_k)^{-1},$$

$$\mathbf{D}_k = \mathbf{W}_k + \mathbf{W}_k \mathbf{B}_k^T \mathbf{G}_k \mathbf{A}_k \mathbf{C}_k \mathbf{A}_k^T \mathbf{G}_k \mathbf{B}_k \mathbf{W}_k,$$

$$\mathbf{E}_k = -\mathbf{W}_k \mathbf{B}_k^T \mathbf{G}_k \mathbf{A}_k \mathbf{C}_k,$$

with:

$$\mathbf{W}_k = (\mathbf{B}_k^T \mathbf{G}_k \mathbf{B}_k)^{-1},$$

$$\mathbf{G}_k^B = \mathbf{G}_k - \mathbf{G}_k \mathbf{B}_k \mathbf{W}_k \mathbf{B}_k^T \mathbf{G}_k,$$

$$\text{cov} \{ \tilde{\mathbf{v}}_k \} = \mathbf{C}_k, \text{cov} \{ \tilde{\mathbf{q}}_k \} = \mathbf{D}_k, \text{cov} \{ \tilde{\mathbf{v}}_k, \tilde{\mathbf{q}}_k \} = \mathbf{E}_k.$$

The chi-square statistic of the filter step is given by

$$\chi_{k,F}^2 = (\mathbf{p}_k - \mathbf{c}_k^{(0)} - \mathbf{A}_k \tilde{\mathbf{v}}_k - \mathbf{B}_k \tilde{\mathbf{q}}_k)^T \mathbf{G}_k (\mathbf{p}_k - \mathbf{c}_k^{(0)} - \mathbf{A}_k \tilde{\mathbf{v}}_k - \mathbf{B}_k \tilde{\mathbf{q}}_k) +$$

$$+ (\tilde{\mathbf{v}}_k - \tilde{\mathbf{v}}_{k-1})^T \mathbf{C}_{k-1}^{-1} (\tilde{\mathbf{v}}_k - \tilde{\mathbf{v}}_{k-1}),$$

$$\chi_k^2 = \chi_{k-1}^2 + \chi_{k,F}^2.$$

If necessary, the linear expansion can now be repeated in the new point

$$\mathbf{v}_k^{(0)} = \tilde{\mathbf{v}}_k, \mathbf{q}_k^{(0)} = \tilde{\mathbf{q}}_k,$$

and the filter can be recomputed, until there is no significant change either in the chi-square or in the estimate.

If the association of tracks with the vertex is known to be correct, the intermediate results of the filter do not need to be computed, and all  $n$  steps can be combined into a single one (see also Section 3.3 and ref. [35]):

$$\tilde{\mathbf{v}}_n = \mathbf{C}_n \cdot \sum_{k=1}^n \mathbf{A}_k^T \mathbf{G}^{\mathbf{B}_k} (\mathbf{p}_k - \mathbf{c}_{k,0}),$$

$$\mathbf{C}_n = \left[ \sum_{k=1}^n \mathbf{A}_k^T \mathbf{G}^{\mathbf{B}_k} \mathbf{A}_k \right]^{-1}.$$

Since there is no process noise, the smoother is extremely simple:

$$\begin{aligned} \tilde{\mathbf{v}}_k^n &= \tilde{\mathbf{v}}_n, \\ \tilde{\mathbf{q}}_k^n &= \mathbf{W}_k \mathbf{B}_k^T \mathbf{G}_k (\mathbf{p}_k - \mathbf{c}_k^{(0)} - \mathbf{A}_k \tilde{\mathbf{v}}_n), \\ \mathbf{C}_k^n &= \mathbf{C}_n, \\ \mathbf{D}_k^n &= \mathbf{W}_k + \mathbf{W}_k \mathbf{B}_k^T \mathbf{G}_k \mathbf{A}_k \mathbf{C}_n \mathbf{A}_k^T \mathbf{G}_k \mathbf{B}_k \mathbf{W}_k, \\ \mathbf{E}_k^n &= -\mathbf{W}_k \mathbf{B}_k^T \mathbf{G}_k \mathbf{A}_k \mathbf{C}_n. \end{aligned}$$

If there is a significant change in the smoothed vertex position, it may be worthwhile to recompute the derivative matrices  $\mathbf{A}_k$  and  $\mathbf{B}_k$ .

The residuals and their covariance matrices have the following form:

$$\begin{aligned} \mathbf{r}_k &= \mathbf{p}_k - \mathbf{c}_k^{(0)} - \mathbf{A}_k \tilde{\mathbf{v}}_k - \mathbf{B}_k \tilde{\mathbf{q}}_k, \\ \mathbf{R}_k &= \mathbf{V}_k (\mathbf{G}^{\mathbf{B}_k} - \mathbf{G}^{\mathbf{B}_k} \mathbf{A}_k \mathbf{C}_k \mathbf{A}_k^T \mathbf{G}^{\mathbf{B}_k}) \mathbf{V}_k, \\ \mathbf{r}_k^n &= \mathbf{p}_k - \mathbf{c}_k^{(0)} - \mathbf{A}_k \tilde{\mathbf{v}}_n - \mathbf{B}_k \tilde{\mathbf{q}}_k^n, \\ \mathbf{R}_k^n &= \mathbf{V}_k (\mathbf{G}^{\mathbf{B}_k} - \mathbf{G}^{\mathbf{B}_k} \mathbf{A}_k \mathbf{C}_n \mathbf{A}_k^T \mathbf{G}^{\mathbf{B}_k}) \mathbf{V}_k. \end{aligned}$$

## 4.6 Detection of secondary vertices

A secondary vertex is created by the decay of an unstable particle. Two types of decays have to be distinguished, because they bear completely different features:

- The particle is longlived (typically a  $K^0$  or a  $\Lambda$ ) and decays in the active volume of the detector. Since the tracks of the decay products do not point to the primary vertex, they are easily identified as secondary tracks. The combination of such tracks to secondary vertices is a problem of global pattern recognition and will not be treated in this section.

- The particle is shortlived (typically a meson or baryon carrying charm or beauty) and decays before it reaches a detector, e.g. inside the beam tube in a collider experiment. The tracks of the decay products point roughly to the primary vertex. The association of such tracks to secondary vertices requires a fit of the primary vertex with the utmost precision.

Since certain shortlived particles are currently the focus of considerable attention, it is of great physical interest to separate their secondary decay vertices from the primary interaction vertex. The shortlived particles in question have average decay lengths between a few hundred micrometer and a few millimeter. Therefore the task is not a trivial one.

The smallest decay length which can be resolved depends mainly on two factors, one of them connected with the detector, the other one with the accelerator. The first one is the quality of the track reconstruction, including effects from detector alignment; the second one is the extension of the interaction region, i.e. the size of the beam spot (in colliders) or of the target (in fixed target experiments).

In the reconstruction of secondary vertices two main problems have to be solved:

- First, the tracks which do not originate at the primary vertex have to be identified. This can be regarded as an outlier problem, in fact a **multiple outlier problem**, since there are normally several tracks emerging from each decay vertex, and there may be several decay vertices. Given the high multiplicities (up to 50) and the fact that the number of "outlying" tracks is not known a priori, it is not possible – for reasons of computing time – to work through all subsets of tracks for several values of  $k$ , say  $2 \leq k \leq 8$ . What is needed is a **sequential procedure**, the computing time of which is basically proportional to  $n$ , the total number of tracks.
- Secondly, the secondary tracks found in the previous step have to be associated to physically possible decay vertices. Thereby it is essential that the secondary vertex is not contaminated by wrongly associated primary or secondary tracks which happen to pass close to the secondary vertex. Such a track would inevitably spoil the kinematic fit of the decay or lead to a wrong charge and mass being assigned to the decaying particle. Therefore outliers have to be detected and removed also in the geometrical fit of the secondary vertex.

It should be noted that these two outlier problems exhibit fairly different features. In the case of the primary vertex, there is usually a high multiplicity and possibly a very precise prior knowledge of the beam position; in the case of a secondary vertex, the multiplicity is small and there is no prior information on the vertex position. Therefore the latter case is certainly the more difficult one. The final strategy of vertex reconstruction depends on the properties of the beam, on the availability of a micro-vertex detector and on the quality of track reconstruc-

tion and particle identification in general. It can only be arrived at by detailed simulation studies for a particular detector at a particular accelerator.

#### 4.6.1 Detection of outlying tracks by chi-square tests

We propose three different approaches. Their respective merits and performance will be investigated in Chapter 5, where the results of a Monte-Carlo study will be presented. The first two have already been used in a study on vertex evaluation in the DELPHI experiment [67].

##### Sequential test against a reliable subset

This test is only useful for the primary vertex. It is based on the fact that some tracks are unlikely to come from a secondary vertex, i.e. that it may be possible to select tracks very likely produced at the primary vertex according to physical criteria, e.g. high momentum or a certain angular region. If this can be done, a preliminary vertex can be fitted from those tracks. The remaining tracks are then subjected to a  $\chi^2$ -test. As a test statistic the chi-square increment of the filter,  $\chi^2_{k,F}$ , is used. If the track is accepted, the vertex may or may not be updated.

The prior information on the beam position plays a crucial role in the construction of the preliminary vertex. For instance, in the linear collider SLC at SLAC the transverse size of the beam is in the order of only 0.01 mm. If the beam position is monitored to a similar precision, it provides a preliminary vertex position which is much more accurate than the position obtained from the track information of a single event. (The position of the beam is monitored by taking the average vertex position of the events sampled in an interval which is long enough to provide a sufficient number of events, but not so long that beam conditions change by an appreciable amount.) Unfortunately, the situation is less favourable at LEP: The size of the beam is in the order of a few hundred micrometer, and also the estimated vertex position is less accurate, because of the larger beam tube ( $R=8$  cm at LEP versus  $R=2.5$  cm at SLC).

##### Sequential test of the entire sample

In this procedure first all tracks are fitted to a common vertex. Then for each track  $k$  ( $1 \leq k \leq n$ ) an updated vertex is computed, which results from the omission of track  $k$  from the entire sample. This is done by the inverse Kalman filter:

$$\begin{aligned} C_k^{n*} &= (C_n^{-1} - A_k^T G_k^B A_k)^{-1}, \\ \tilde{v}_k^{n*} &= C_k^{n*} [C_n^{-1} \tilde{v}_n - A_k^T G_k^B (p_k - c_{k,0})]. \end{aligned}$$

The distance of the track from the new vertex is expressed by the "smoothed chi-square":

$$\chi^2_{k,S} = \mathbf{r}_k^n T \mathbf{G}_k \mathbf{r}_k^n + (\tilde{\mathbf{v}}_n - \tilde{\mathbf{v}}_k^{n*})^T (\mathbf{C}_k^{n*})^{-1} (\tilde{\mathbf{v}}_n - \tilde{\mathbf{v}}_k^{n*})$$

$\chi^2_{k,S}$  is used as the test statistic. If it exceeds the  $(1-\alpha)$ -quantile of the  $\chi^2$ -distribution with two degrees of freedom, the track is considered as an outlier and rejected from the vertex. It is also useful to label doubtful tracks with a chi-square between the  $(1-\alpha)$ -quantile and the  $(1-\beta)$ -quantile ( $\alpha < \beta < 1$ ).

This procedure works properly only if the vertex estimated from the entire sample is not noticeably biased by the outlying tracks. This is very likely to be true in the case of the primary vertex, since there are usually only few outliers relative to the total multiplicity. It is less likely to be true in the case of a secondary vertex. (It is assumed that obvious outliers have been rejected already during filtering.)

If there is a serious possibility that the estimated vertex is biased by outliers, we propose two solutions: Either the filter is robustified (see next subsection) or the sequential test is refined in the following way:

- Set  $i=1$ .
- Step 1: Look for the track with the largest  $\chi^2_{k,S}$ ; suppose it is track  $k_i$ .
- Step 2: If  $\chi^2_{k_i,S} \leq c$ , stop. If  $\chi^2_{k_i,S} > c$ , go to Step 3.
- Step 3: Remove track  $k_i$  from the vertex and update the momenta of all remaining tracks. Set  $i \leftarrow i+1$  and go to Step 1.

Note that the number of operations is no longer proportional to  $n$ , but rather to  $m \cdot n$ , where  $m$  is the number of outliers found.

#### 4.6.2 A robust vertex fit

It is obvious that outlying tracks can be detected more frequently if the estimated position of the vertex is not biased by the outliers, particularly if the multiplicity is not very large, as in the case of a secondary vertex. Therefore we propose to study the application of a robust estimation procedure to the vertex fit. Such a procedure should suppress or at least reduce the influence of outlying tracks. The robust estimator which is most easily adapted to the problem of vertex fitting is in our opinion the M-estimator [68]. An M-estimator may be regarded as a generalization of the least-squares estimator.

We consider a linear model of the following form:

$$\mathbf{y} = \mathbf{A} \mathbf{x} + \boldsymbol{\varepsilon}, \quad \text{cov} \{ \boldsymbol{\varepsilon} \} = (\delta_{ij} \sigma_i^2).$$

The residuals are denoted by  $r_i$ :

$$r_i = y_i - \sum_{j=1}^m a_{ij} x_j \quad (1 \leq i \leq n).$$

Then the least-squares estimator of  $\mathbf{x}$  is defined by

$$\sum_{i=1}^n (r_i / \sigma_i)^2 \stackrel{!}{=} \min,$$

or

$$\sum_{i=1}^n (r_i / \sigma_i) a_{ij} / \sigma_i = 0 \quad (1 \leq j \leq m).$$

An M-estimator of  $\mathbf{x}$  is defined by

$$\sum_{i=1}^n \rho(r_i / \sigma_i) \stackrel{!}{=} \min,$$

or

$$\sum_{i=1}^n \psi(r_i / \sigma_i) a_{ij} / \sigma_i = 0 \quad (1 \leq j \leq m),$$

where  $\rho$  is a function devised in such a way that the influence of large residuals  $r_i$  on the estimate is reduced, and  $\psi(t) = d\rho(t)/dt$ . (We note that by setting  $\psi(t) = t$  we recover the least-squares estimator.) We will restrict our investigations to Huber's monotone psi-function:

$$\psi_H(t) = \begin{cases} t, & |t| \leq c, \\ c \cdot \text{sgn}(t), & |t| > c. \end{cases}$$

$c$  is called the robustness constant. The choice of  $c$  determines the "degree of robustness" of the M-estimator. For  $c \rightarrow \infty$  the M-estimator approaches the ordinary LS-estimate, for  $c \rightarrow 0$  we obtain the  $L_1$ -estimator.  $c$  is commonly set to a value between 1 and 2.

The definition of the M-estimator can be rewritten as follows:

$$\sum_{i=1}^n \frac{r_i}{\sigma_i} \cdot \frac{\psi(r_i / \sigma_i)}{(r_i / \sigma_i)} \cdot a_{ij} / \sigma_i = 0 \quad (1 \leq j \leq m).$$

Thus an M-estimator may be expressed as a **weighted least-squares estimator** with weights

$$w_i = \frac{\psi(r_i / \sigma_i)}{(r_i / \sigma_i)}.$$

The introduction of the weights  $w_i$  is equivalent to the replacement of all residuals which are larger than  $c \cdot \sigma_i$  by  $c \cdot \sigma_i$ , i.e. a "winsorization" of the residuals. Unfortunately the weights

depend on the residuals, so that the M-estimate of  $\mathbf{x}$  cannot be defined explicitly. Algorithms based on iterative procedures are given in [54, 69]. The iterated weighted least squares (IWLS) algorithm described in ref. [54] can easily be adapted to the robust estimation of the vertex. First we note that correlations between measurements cannot be accommodated in the definition of the M-estimator. In the case of the vertex fit, this implies that the 5-dimensional vector  $\mathbf{p}_k$  of each track has to be transformed to a vector of uncorrelated random variables before the iteration starts. The first step of the iteration is an ordinary least-squares estimation of the vertex position and of the momentum vectors, as described in Section 4.5, including the computation of the smoothed residuals  $\mathbf{r}_{k,i}^n$ . These are used to calculate the weights for each track:

$$w_{i,k} = \psi [ (\mathbf{r}_{k,i}^n) / \sigma_{i,k} ] / [ (\mathbf{r}_{k,i}^n) / \sigma_{i,k} ],$$

with:

$$\sigma_{i,k} = \sqrt{(\mathbf{V}_k)_{ii}}, \quad \mathbf{V}_k = \text{cov} \{ \mathbf{p}_k \},$$

where  $\mathbf{V}_k$  is the (diagonalized) covariance matrix of the estimated parameters  $\mathbf{p}_k$  of track  $k$ . The least-squares fit is repeated with the modified weight matrices:

$$\mathbf{G}_k' = \mathbf{G}_k \mathbf{W}_k, \quad \mathbf{G}_k = \mathbf{V}_k^{-1}, \quad \mathbf{W}_k = \text{diag} \{ w_{1,k}, \dots, w_{5,k} \}.$$

The new estimate is used in turn to determine new weights, and so on, until the objective function is close to its minimum.

Once an estimate has been obtained, the statistic  $\chi^2_{k,S}$  (see Subsection 4.6.1) is used to test on outlying tracks. The final fit of the primary and of eventual secondary vertices is done with the standard Kalman filter and smoother.

The performance of the robust vertex estimator will be compared to the ordinary least-squares estimator in Chapter 5.



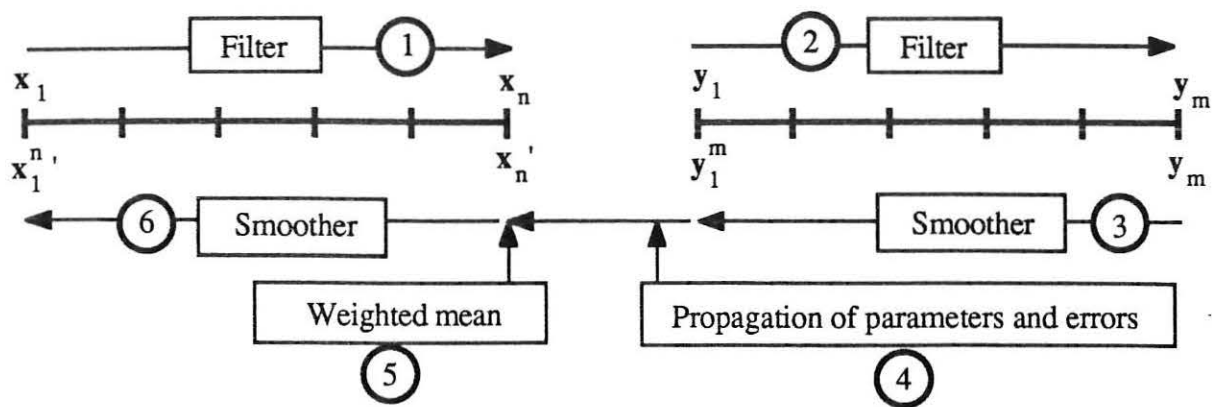


Fig. 4.1: Merging of two track segments by successive filtering and smoothing.

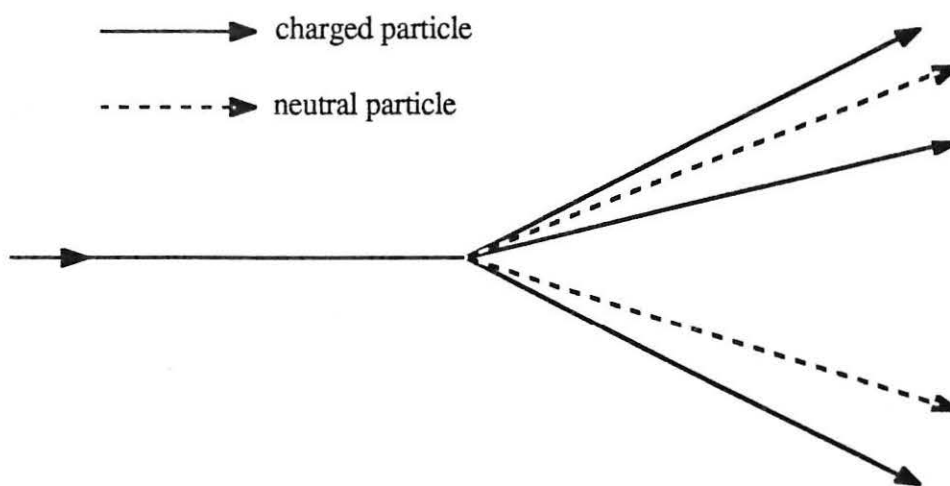


Fig. 4.2: The decay of a charged particle.

Because of charge conservation, the number of charged decay products is always odd.

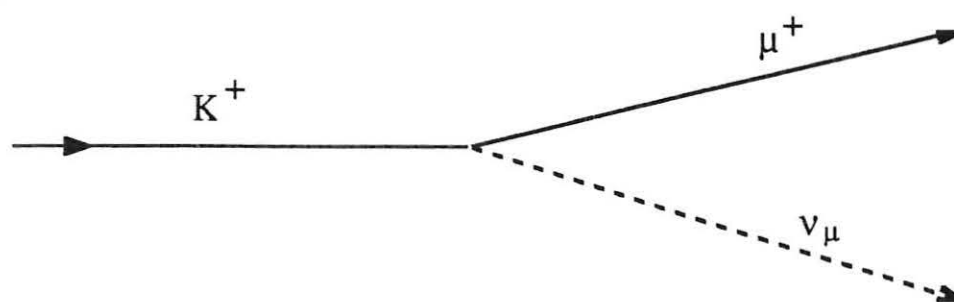


Fig. 4.3: A 1-prong decay (kink).

Lorentz transformation of the muon momentum in K- and  $\pi$ -decays

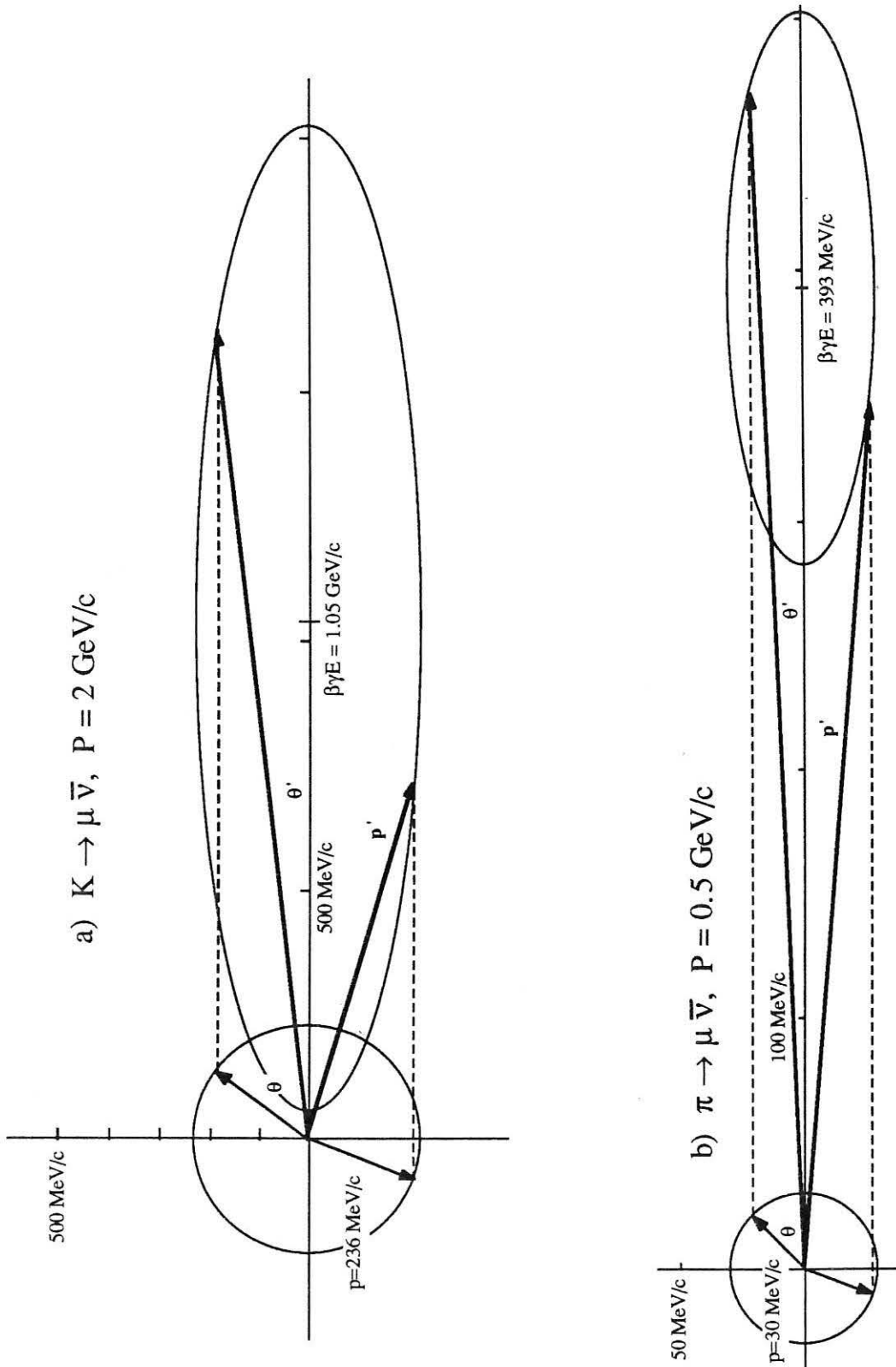
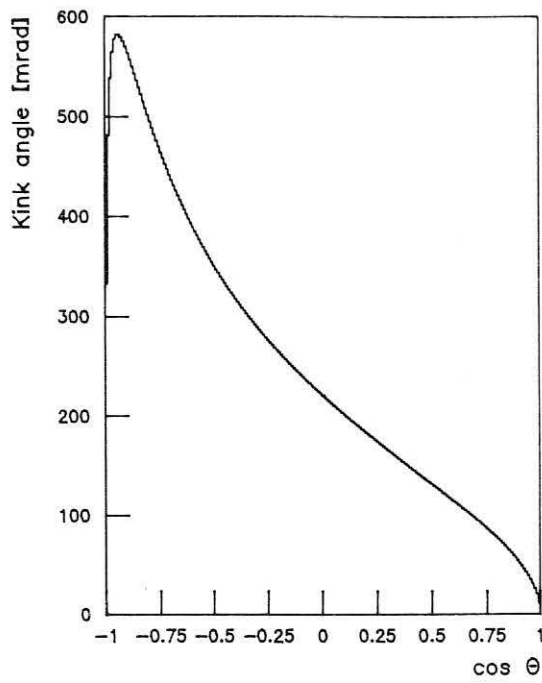


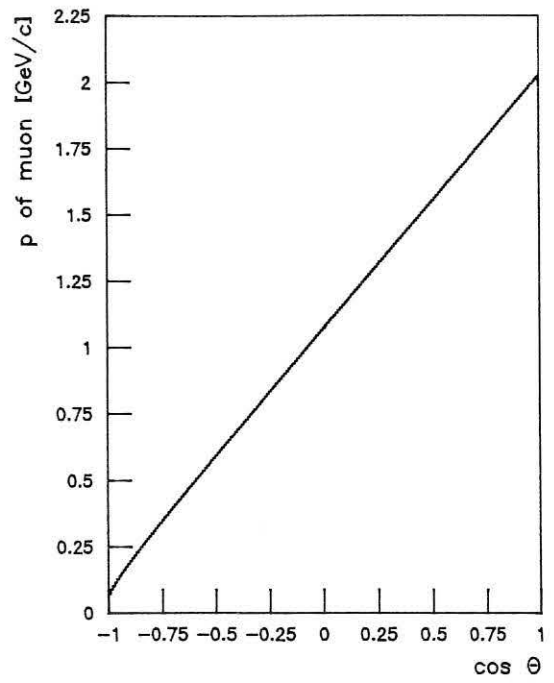
Fig. 4.4: Lorentz transformation of the muon momentum in K- and  $\pi$ -decays.

a)  $K \rightarrow \mu \bar{\nu}$ ,  $P = 2 \text{ GeV}/c$ ; b)  $\pi \rightarrow \mu \bar{\nu}$ ,  $P = 0.5 \text{ GeV}/c$ .

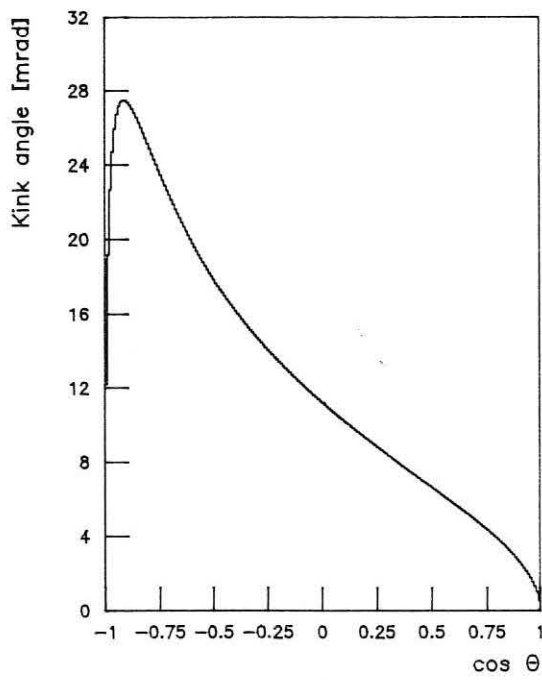
## Kinematics of K-decay



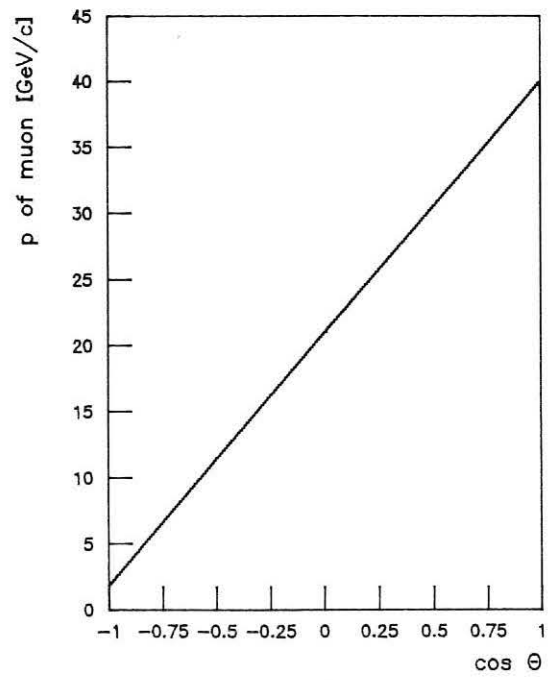
a)



b)



c)



d)

Fig. 4.5: Kink angle and momentum of muon as a function of  $\cos \theta$  for the decay  $K \rightarrow \mu \bar{\nu}$ .  
a,b)  $P = 2 \text{ GeV}/c$ ; c,d)  $P = 40 \text{ GeV}/c$ .

## Kinematics of $\pi$ -decay

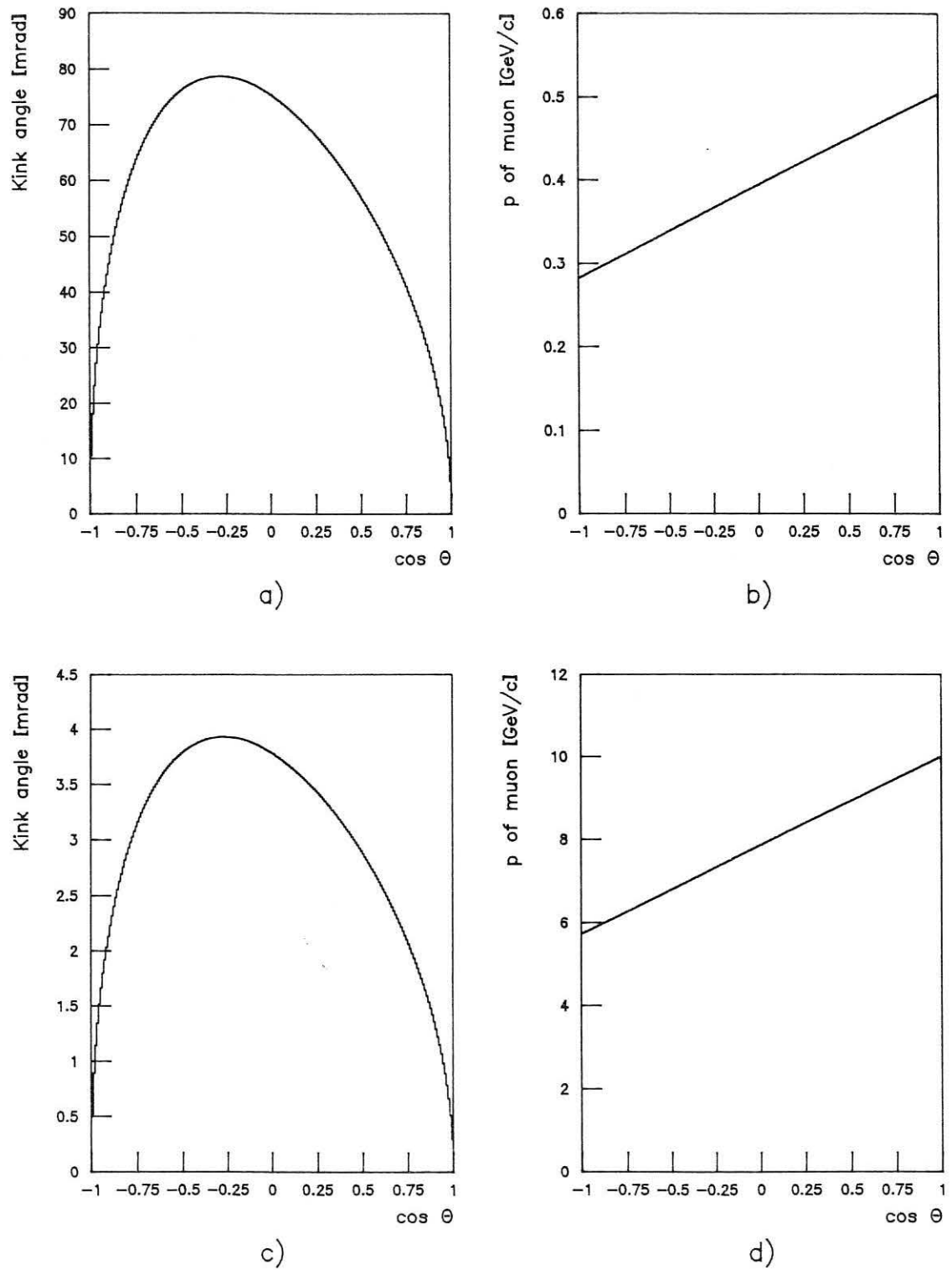


Fig. 4.6: Kink angle and momentum of muon as a function of  $\cos \theta$  for the decay  $\pi \rightarrow \mu \bar{\nu}$ .  
a,b)  $P = 0.5 \text{ GeV}/c$ ; c,d)  $P = 10 \text{ GeV}/c$ .

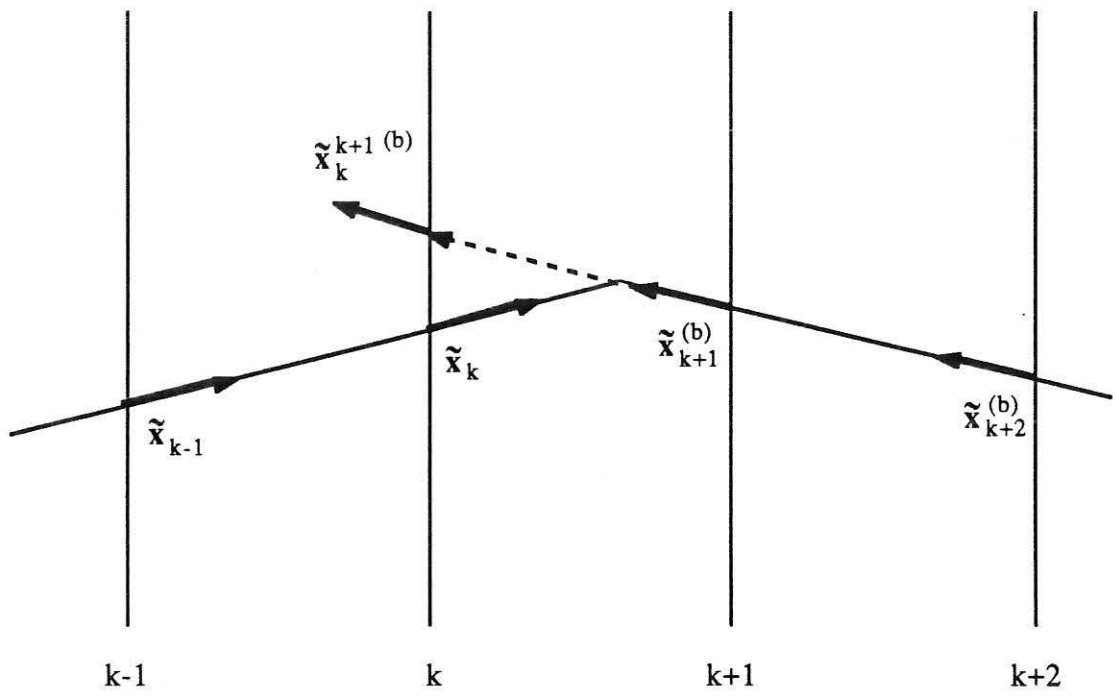


Fig. 4.7: A kink as a sudden change of the state vector.

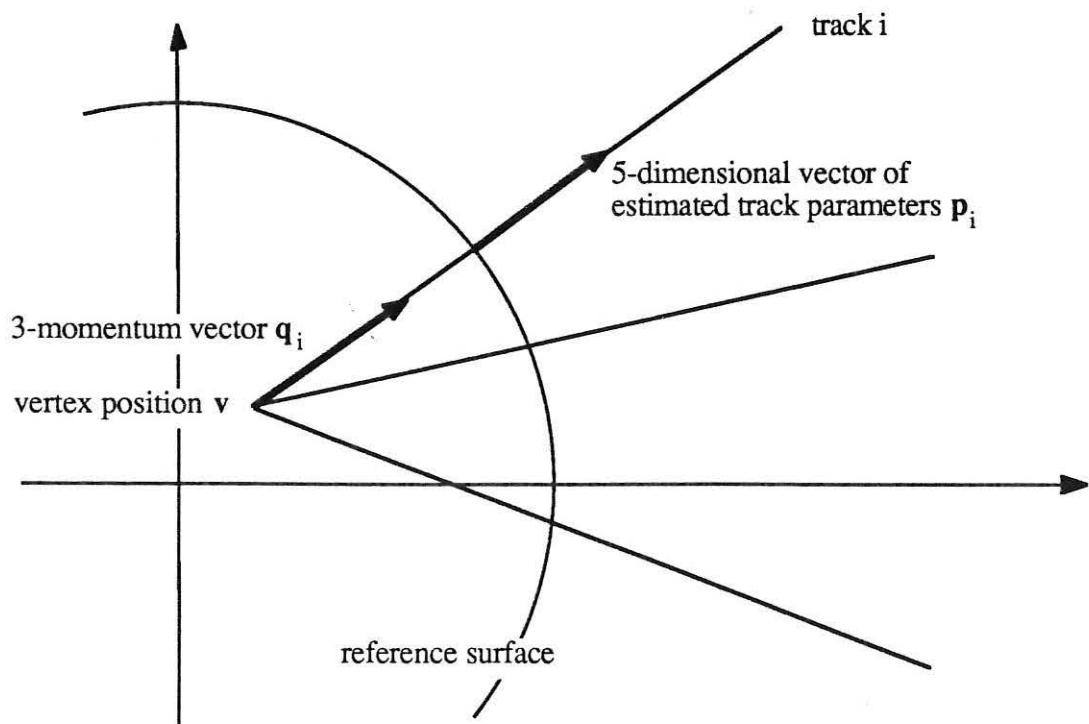


Fig. 4.8: Measurements and estimated parameters of the vertex fit.

# CHAPTER 5

## RESULTS OF FILTER TECHNIQUES IN THE DATA ANALYSIS OF DELPHI

### 5.1 The DELPHI program chain for data analysis

In this section we will give a brief survey of the software which is available now in DELPHI for simulation and track reconstruction. Of course the software will not cease to evolve during the next few years, therefore the status is the one of autumn 1988.

#### 5.1.1 Simulation

There are two mainstreams of event simulation programs in DELPHI: Full simulation (DELSIM, [14]) and fast simulation (FASTSIM, [70]). Both programs consist of two main parts: generation of events and simulation of detector response.

The generation of events is not of our concern here; suffice it to say that there are various types of programs available which simulate interactions of elementary particles according to different physical theories and models. The output of an event generator consists of a list of the particles produced in the interaction. For each particle the type and the 3-momentum vector at the production vertex is given, for secondary particles also the decay vertex of its mother particle.

In the second part of the simulation program, each particle is followed through the detector system, until it either decays, leaves the detector or interacts in one of the detector components. In each detector component crossed by the particle the response of the particular detector is simulated and stored.

As it is, the response of a detector is by no means uniquely defined. The detector can be simulated to various levels of detail, and the simulation may in fact anticipate the first steps of the data analysis. As an example we consider the simulation of the response of a drift chamber (see Subsection 2.2.2). A fully detailed simulation of the chamber starts with the generation of the primary ionizations electrons in the gas, tracks the electrons to the anode (sense) wire, simulates the formation of the avalanche, computes the shape of the electric signal in the wire, passes the signal through the amplifier and the digitizer, and finally generates a drift time.

Although such a detailed simulation – or parts of it – may be required during the design and optimization phase of the detector and of the electronics, it is obviously much too slow to be used for the production of a large number of events. Therefore it is common practice to replace the detailed simulation of a physical process or a sequence of such processes by a **statistically equivalent model**. In our example, the drift time is simply computed from the shortest distance of the track from the sense wire, and all effects occurring between the primary ionization and the digitizer are absorbed by two random numbers. One of them is added to the computed drift time as the "measurement error"; the other one is the "efficiency", i.e. the probability of the detector response. Of course one has to make sure that the distribution of these random numbers is sufficiently close to the distribution arising from a fully detailed simulation.

The concept of the statistically equivalent model can be carried further up in the simulation of the detector response. Thus not only drift times, but also drift distances, space points, track elements and even whole tracks can be simulated by a model which is roughly equivalent to the respective stage of data analysis. For some problems, e.g. the checking of the data analysis code, a simplified model of the detector response is absolutely necessary.

After this digression into the technique of event simulation we can state briefly the main differences between DELSIM and FASTSIM:

- DELSIM simulates events down to the level of raw data, i.e. of electronic signals (mostly drift times and electric charges). It uses a very detailed geometrical and functional description of the detector. It takes a few minutes on a large mainframe computer to simulate a typical event with DELSIM.
- FASTSIM simulates events down to the level of track elements. For example, in a tracking detector a space point and the track direction is simulated, possibly also a momentum measurement. FASTSIM uses a very much simplified geometrical description of the detector. It is at least 50 times faster than DELSIM.

### 5.1.2 Data analysis

We will now give a brief review of the present status of DELANA, DELPHI's data analysis program.

The first stage of data analysis is **local pattern recognition** in all detector components. For a tracking detector, the aim of local pattern recognition is the production of **track elements** from the raw data. A track element is either a space point, or a space point plus directions, or a complete 5-dimensional vector of track parameters. (A single coordinate value is

accepted only in some special cases.) The methods involved do not concern us here. The interested reader can find them in ref. [71]. The track elements are stored in so-called TE banks, which are manipulated by DELPHI's data management system TANAGRA [40].

In the next stage, the **track search**, track elements are put together to **track candidates**, which are stored in TS (track string) banks. This is done separately in the central ("barrel") and in the forward region of DELPHI.

In the third stage, each track candidate is submitted to the **track fit**, which computes an estimate of the track parameters, its covariance matrix and a chi-square. The results of the track fit are stored in TK banks. There are two track fit modules, one for the barrel and one for the forward region. Both use filter techniques for track fitting. In the forward fit, which is being developed in the Institute of High Energy Physics in Vienna, the full filter/smoothing is operational [72]. The performance of the forward track fit will be presented in more detail in the next section. In the barrel fit which is being developed at Collège de France in Paris, smoothing is implemented by the combination of a forward and a backward filter [73].

After the track fit, a second pass of local pattern recognition is done. It is supposed to clear up any remaining ambiguities with the help of the information gained by the track fit. This is still under development.

Finally, tracks are associated with a vertex, either the primary interaction vertex or a secondary decay vertex. All vertices that have been found are fitted, the results being stored in TV banks. The programs for track association and vertex fitting are developed in the Vienna institute, along the lines proposed in sections 4.5 and 4.6 [74].

## 5.2 Performance studies of filter algorithms

In this section we present the results obtained from the application of the filter algorithms proposed in Chapter 4 to problems of track and vertex reconstruction in the DELPHI detector. Since real data are not yet available, we are forced to resort to simulated data. Clearly, simulated data never show precisely the same behaviour as real data, but we trust that the simulation is sufficiently close to reality, so that the conclusions drawn from simulated data will essentially remain valid for real data.

### 5.2.1 Results of track reconstruction algorithms

This subsection is mainly devoted to a comparison between the global method of track fitting (see Subsection 3.2.3) and the Kalman filter/smoothing algorithm.



An important characteristic of an algorithm is its speed. In order to compare the execution times of global fit and filter methods, we have simulated tracks in an idealized detector, which consists of 16 measurement planes. Each plane supplies a measurement vector which consists either of two spatial coordinates or of two coordinates plus two directions. The track model is an exact helix. Fig. 5.1 shows the execution time per track as a function of the number of planes, with and without multiple scattering in the detector volume. The execution time includes the computation of the  $\chi^2$ -statistic and of the "pulls". Without multiple scattering, the computation times of both the filter and the global fit rise linearly. The filter is slower, because the estimated state vector is computed in every measurement plane. The smoother amounts to a simple back extrapolation and is redundant without multiple scattering. With multiple scattering, the picture changes considerably. The filter still rises linearly, and so does the smoother. The global fit, however, exhibits a third order rise, which is due to the inversion of the full covariance matrix of the measurements, and ceases quickly to be competitive, particularly if there are 4 measurements per plane.

Our next investigation is concerned with the behaviour of global and filter methods in a non-linear model. In a linear model, the global estimate, the filtered estimate and the smoothed estimate are strictly the same. In a non-linear model, differences can arise between the three estimates. Also, residuals and chi-square statistics do no longer obey their exact (normal or chi-square) distributions.

For the study of non-linear track models, we pass to a more realistic detector, which we will call a "simplified TPC (STPC)" (see Subsection 2.2.3 and Fig. 5.2). STPC consists of sixteen cylinders concentric with the z-axis. Each cylinder provides a measured space point  $(R_i, R_i\Phi_i, z_i)$  in cylindrical coordinates. The measurement error is normally distributed with a s.d. of .2 mm in  $R\Phi$  and .6 mm in z. The radii  $R_i$  are assumed to be known exactly. The sixteen cylinders are equally spaced between 36.5cm and 106.625cm, with a pitch of 4.675 cm. The reference surface is the innermost measurement cylinder. The magnetic field of 1.2 Tesla is homogeneous and parallel to z. Hence the track model is an exact helix.

The non-linearity of the model depends on the choice of the state vector. We choose the state vector such that the model is not too far from linear (see Subsection 3.2.3):

$$\mathbf{x} = ( \Phi, z, \theta, \beta, 1/r ),$$

with:

- $\Phi$ ..... polar angle of space point,
- $z$ ..... z of space point,
- $\theta$  ..... polar angle of track direction,

$\beta \dots\dots\dots = \varphi - \Phi$ , where  $\varphi$  is the azimuthal angle of the track direction,  
 $r \dots\dots\dots$  radius of the helix.

The curvature is chosen rather than the radius because it enters directly into the equation of motion.

In this detector 5000 tracks were simulated with the following initial conditions:

$$x=y=z=0, \pi/4 \leq \theta \leq 3\pi/4, 0 \leq \varphi \leq 2\pi, 0.5 \leq p \leq 1.$$

The initial values of both the global fit and the filter/smoothen were chosen at a distance of five standard deviations from the true values.

Figs. 5.3 a)-c) show the distributions of three estimates of  $1/r$ , centered around the true value of  $1/r$ , at the reference surface: the global estimate (a), the filtered estimate (filtered from 16 to 1, b), and the filtered/smoothed estimate (filtered from 1 to 16, c). The difference of the estimate and the true value is divided by the respective standard deviation. Also shown are standard normal probability density functions. The agreement between the experimental and the theoretical distribution is remarkably good, which means that the track model is indeed close to a linear one. Figs. 5.3 d)-f) show the pairwise differences of the three estimates, again divided by the standard deviation of the estimate. The r.m.s. of the normalized difference is in all cases below 3%. Figs. 5.4 a)-c) show the probability transform of the  $\chi^2$ -statistics of the three estimates, and Figs. 5.4 d)-f) show the pairwise differences of the  $\chi^2$ -statistics. The distributions of the probability transforms are perfectly flat. It can be concluded that no iterations are required in order to approach the minimum of the objective function sufficiently closely. The  $\chi^2$ -statistics agree very well.

In order to check the influence of a strong process noise (multiple scattering) on the behaviour of the estimates, a second sample of 5000 tracks was generated. In this sample, STPC was assumed to be filled with a material with a radiation length of 200cm, so that each track crosses about 50% of a radiation length. The resulting distributions are shown in Figs. 5.5 and 5.6. We observe that the filtered and the smoothed estimates of the curvature agree perfectly, whereas the global estimate is slightly biased (by about 4% of the standard deviation). The  $\chi^2$ -statistics agree very well on average; the differences exhibit a larger spread than without multiple scattering. We conclude that the non-linear model is handled very well by all three fitting methods; it has, however, to be stressed that this is due to the proper choice of the track parameters and to the simple homogeneous field.

As stated in Chapter 4, one of the assets of the combined filter/smoothen algorithm is the fact that optimal estimates are available anywhere along the track. We have checked the gain in precision due to smoothing both in the barrel and in the forward region of DELPHI.

In the barrel region a stripped down version of FASTSIM was used to simulate tracks in ID, TPC and OD. The arrangement of the detectors and the scattering surfaces is summarized in Fig. 5.7 and Table 5.1. In this quite realistic setup tracks were simulated with the following initial conditions:

$$R=0.1, \pi/2 \leq \Phi \leq 3\pi/2, z=0, \pi/4 \leq \theta \leq 3\pi/4, \varphi=\Phi, 1 \leq p \leq 25.$$

We compare two sets of estimates of the track parameters in OD: the estimate obtained by extrapolation of the filtered estimate in ID to OD, and the estimate obtained by smoothing back from ID to OD, which contains the full information. Fig. 5.8 shows the standard error of the difference of both estimates and the true value, as a function of the momentum  $p$ . The smoothed estimates of  $R\Phi$  (Fig. 5.8 a) and  $\varphi$  (Fig. 5.8 d) are clearly much better than the extrapolation from ID. Closer inspection reveals that the s.d. of the smoothed estimates of  $R\Phi$  and  $\varphi$  are virtually equal to the s.d. of the corresponding measurements in OD. This can be explained by the fact that there is a long lever arm and a lot of multiple scattering between TPC and OD, so that the smoothed estimate is dominated by the precise measurement of  $R\Phi$  and  $\varphi$  in OD. On the other hand,  $z$  (Fig. 5.8 b) is measured very poorly, and  $\theta$  (Fig. 5.8 c) is not measured at all in OD, so that the smoothed estimate is not better than the extrapolated one. For larger momenta, the effect of multiple scattering disappears, and the two estimates coincide.

A similar behaviour is found in the forward region. Standard FASTSIM was used to simulate tracks in the forward tracking detectors ID, TPC, FCA and FCB (Fig. 5.9). The measurement errors in ID and TPC depend strongly on the polar angle  $\theta$  of the track below a value of  $\theta \approx 30^\circ$ . In order to get a transparent behaviour, tracks were simulated at constant  $\theta = 25^\circ$ , again with  $1 \leq p \leq 25$  GeV/c and isotropic in  $\varphi$ . All tracks come from the origin.

The estimates to be compared are the smoothed estimate in FCB and the extrapolation from the filtered estimate in ID. The results are shown in Fig. 5.10. The precision of the smoothed estimate of position in FCB is dominated by the one of the measurement in FCB; the track direction in FCB is very much improved by the smoother. Therefore **the smoother is particularly important** for extrapolation into the forward RICH, next to FCB, which needs a very precise prediction of the track position. This result was obtained with the standard forward track fit in DELANA.

## 5.2.2 Results of outlier studies

In Section 4.3 various algorithms for the detection, removal or accommodation of outliers were proposed. In this subsection we will study the performance of these algorithms and compare their respective merits.

We start with a relatively simple case, namely the simplified TPC (STPC) introduced in the preceding subsection. We recall that STPC has 16 cylindrical measurement surfaces, each of which provides a measurement of  $R\Phi$  and  $z$ . Again we assume normally distributed measurement errors with a standard deviation of .2mm and .6mm, respectively. The basis of our investigation is formed by three samples of 5000 tracks each, which were simulated in STPC by a dedicated simulation program. In order to investigate the effect of multiple scattering on the performance of our algorithms, the amount of material crossed by a track is different in the three samples: In sample I, there is no material; in sample II (III) about 10% (50%) of a radiation length of material is spread evenly across the detector volume. The precise characteristics of the three samples are summarized in Table 5.2.

### A) Finding outliers with the filter

In Subsection 4.3.2 we saw that the "chi-square of the filter",  $\chi^2_{k,F}$ , can be used as an outlier test. This test will be called briefly "CF-test". We start the investigation of the CF-test with the simple case of a single variance-inflation outlier (see Subsection 4.3.1). To this end a new sample Ia of tracks was created, by contaminating each track of sample I with a single variance-inflation outlier with threefold standard deviation, i.e.:

$$V_k^{(2)} = 9 \cdot V_k^{(1)} \quad (1 \leq k \leq 16).$$

The probability of the outlier occurring at position  $k$  is the same for all  $k$ . The size of the test was chosen to be 1%. We cannot expect to find outliers in the initial track segment of the first three measurements, so testing for outliers starts at  $k=13$ .

There are now two possible ways of proceeding if an outlier is found. Either it can be removed immediately, or it is only marked as an outlier and removed in a second pass of the filter, or by the smoother (see Subsection 4.2.5). It turns out that slightly more outliers are found in sample Ia, if they are removed immediately (30.4% vs. 29.9%), but that the losses of good measurements are nearly three times as high (7.6% vs. 2.7%). The reason for this is the fact that an outlier in the initial track segment of the first three points, in a certain number of cases, leads to the loss of the whole track, since all subsequent points are rejected as outliers, and the track has no chance to recover from the bad initial segment.

This situation can be improved in the following way: Before the outlier test is performed, the quality of the track segment filtered so far is checked by another  $\chi^2$ -test. If the track segment fails, it is considered as unreliable, and the CF-test is skipped. This brings down the losses to 6.4% and 1.8%, respectively, the number of found outliers being virtually the same. The improvement is smaller than one would expect. The reason is that a track segment of three points is not constrained at all in the plane of curvature; therefore the curvature, being just defined by three points, may be totally wrong due to an outlier in  $R\Phi$ .

A further improvement can be achieved by increasing the initial track segment by one point, i.e. to start testing for outliers only at  $k=12$ . This brings again down the losses, to 4.5% and 1.6%, respectively, but reduces also the power of the test a bit, namely to 29% and 28.6%. We conclude that the optimal strategy for finding outliers with the filter requires a check on the quality of the current track segment and a second pass for removing the outliers. This makes the procedure rather unattractive as compared to finding outliers with a combined filter and smoother, which is much more powerful.

## B) Finding outliers with the smoother

We have seen in Subsection 4.3.2 that there is a test which has larger power than the CF-test: the test on the "chi-square of the smoother",  $\chi^2_{k,S}$ , henceforth to be called "CS-test". This is borne out by the results of the CS-test applied to sample Ia. Fig. 5.11a shows the probability of detecting an outlier with the CS-test (blank) and with the CF-test (shaded), as a function of  $d$ , where  $d$  is, roughly speaking, the distance of the outlier from the true track parameters:

$$d = \sqrt{(\mathbf{m}_k - \mathbf{H}_k \hat{\mathbf{x}}_k)^T \mathbf{G}_k (\mathbf{m}_k - \mathbf{H}_k \hat{\mathbf{x}}_k)} .$$

The power of the CF-test is well below the power of the CS-test. Fig. 5.11b shows the same probabilities, as a function of the position  $k$  of the outlier along the track. The power of the CS-test is largest in the center of the track, dropping slightly towards both ends. The CF-test doesn't of course find any outliers in the initial segment; the rise of the power towards the other end of the track reflects the increasing amount of information contained in the filtered estimate.

Fig. 5.12 shows the distribution of the  $\chi^2$ -probability transform of the total  $\chi^2$ , in (a) without removal of outliers, in (b) with outliers removed by the CF-test, in (c) with outliers removed by the CS-test. A subsequent  $\chi^2$ -cut would reject a large number of otherwise good tracks in cases (a) and (b), whereas the p.d.f. in (c) is nearly flat, showing even a slight depletion at large values of  $\chi^2$ . The loss of good measurements with the CS-test is 1.3%, only slightly more than the nominal value of 1%, the excess being due to a bias of the smoothed estimate caused by unrecognized outliers. No tracks were lost at all.

### C) Finding outliers with the global fit

Also the global estimate can be used to look for outliers, in the same way as the filtered and the smoothed estimate. It suffices to define an analogous  $\chi^2$ -statistic  $\chi^2_{k,G}$  by

$$\chi^2_{k,G} = (\mathbf{r}^G_k)^T (\mathbf{R}^G_k)^{-1} \mathbf{r}^G_k,$$

with:

$$\mathbf{r}^G_k = \mathbf{m}_k - \mathbf{f}^{(k)}(\tilde{\mathbf{p}}),$$

$$\mathbf{R}^G_k = \text{cov} \{ \mathbf{r}^G_k \},$$

where  $\mathbf{f}^{(k)}$  is the projection of the track model vector  $\mathbf{f}$  on the subspace corresponding to measurement  $k$ , and  $\mathbf{R}^G_k$  is the projection of the global covariance matrix of residuals

$$\mathbf{R}^G = \mathbf{V} - \mathbf{A} (\mathbf{A}^T \mathbf{G} \mathbf{A})^{-1} \mathbf{A}^T,$$

on the same subspace. The test based on the statistic  $\chi^2_{k,G}$  will henceforth be called "CG-test".

If there is no multiple scattering, the extrapolation from the global estimate carries the full information, and  $\chi^2_{k,G}$  is equal to  $\chi^2_{k,S}$ . With multiple scattering, however,  $\chi^2_{k,G}$  is a less powerful test statistic than  $\chi^2_{k,S}$ . This was checked by contaminating also samples II and III by a single variance-inflation outlier with threefold standard deviation. The contaminated samples are called IIa and IIIa.

Fig. 5.13a shows the power of the CS-test (blank), of the CG-test (shaded) and of the CF-test (doubly shaded) as a function of the distance  $d$  (see above), for sample IIa. Fig. 5.13b shows the respective powers as a function of the outlier position  $k$ . The effect of multiple scattering is still rather small, although the power function of the CG-test is clearly below the power function of the CS-test for values of  $k$  above 8. Fig. 5.14 shows the same power functions for sample IIIa. Here the effect of multiple scattering is obvious – the outliers in the rear end of the track are to a large extent masked by multiple scattering for the CG-test, whereas the CS-test (and also the CF-test) is hardly affected.

### D) Finding multiple outliers

Until now only single outliers have been considered. Now we want to see how well multiple outliers are handled by the filter/smoothen on the one side and by the global fit on the other side. Therefore sample I and III were contaminated with 2 and also with 3 variance-inflation outliers. Again the outliers were spread uniformly along the track. The results are summarized in Table 5.3 and in Figs. 5.15–5.18.

Table 5.3 shows the power of the various tests and the losses of good measurements, for samples I and III, with 0,1,2 and 3 outliers. The power drops slightly with increasing number of outliers, and the losses of good measurements become larger. This is due to distortions of the estimate by the outliers. The power of the CG-test is well below the power of the CS-test in sample III, and also the losses are higher. Fig. 5.15 (5.16) shows the probability transforms of the different  $\chi^2$ -statistics after removal of outliers, for sample I (III) with 0 (a), 1 (b), 2 (c) and 3 (d) outliers. The peaks at low probabilities which can be observed with the CG-test are due to unrecognized outliers masked by multiple scattering.

Finally, Fig. 5.17 (5.18) gives an idea of how much the removal of outliers affects the quality of the estimated track parameters for sample I (III). For each of the parameters the estimated standard error w.r.t. to the true value is plotted as a function of the number of outliers in the track. In sample I (no multiple scattering) all parameters are about equally affected by outliers: 3 outliers result in an increase of about 60% of the standard error. As expected, the CS-test and the CG-test perform about equally well, while the CF-test is obviously worse. In sample III (strong multiple scattering) the track position is more affected by outliers than the track direction, and the curvature is hardly affected at all. The CS-test does somewhat better than the CG-test, though not much. This can be explained by the fact that the outliers missed by the CG-test are mainly in the rear end of the track (see Fig. 5.14b) and have little influence on the global estimate, as they get smaller weights due to the multiple scattering. Again the CF-test is worst.

### E) Accommodation of outliers with the robust filter

In order to study the statistical properties of the robust filter described in Subsection 4.3.3 with tracks simulated in the simplified TPC, sample I was contaminated with outliers generated according to a mixture model (see Subsection 4.3.1):

$$\begin{aligned} \varepsilon_k &\sim a_k^{(1)} \mathcal{N}(\mathbf{0}, \mathbf{V}_k^{(1)}) + a_k^{(2)} \mathcal{N}(\mathbf{0}, \mathbf{V}_k^{(2)}), \\ a_k^{(1)} + a_k^{(2)} &= 1, \mathbf{V}_k^{(2)} > \mathbf{V}_k^{(1)}, a_k^{(1)} \gg a_k^{(2)}. \end{aligned}$$

We chose  $\mathbf{V}_k^{(2)} = 9 \cdot \mathbf{V}_k^{(1)}$  for all  $k$ , and  $a_k^{(2)}$  varied between  $1/16$  and  $1/4$ . For each value of  $a_k^{(2)}$  the number of outliers follows a binomial distribution. The average number of outliers is equal to  $na_k^{(2)}$  ( $n=16$ ).

First, we check the quality of the estimated track parameters by plotting the normalized differences of estimated and true track parameters. This is shown in Fig. 5.19 for the maximum contamination ( $a_k^{(2)} = 1/4$ ). Even here the agreement with a normal p.d.f. is fairly close, although the tails are somewhat heavier. We can also check how close the distribution of the

estimated parameters is to a normal distribution with mean  $\mathbf{0}$  and covariance matrix  $\mathbf{C}_1$ , by plotting the probability transform of the following  $\chi^2$ -like statistic:

$$\chi^2_E = (\tilde{\mathbf{x}}_1 - \hat{\mathbf{x}}_1)^T \mathbf{C}_1^{-1} (\tilde{\mathbf{x}}_1 - \hat{\mathbf{x}}_1).$$

The distributions are shown in Fig. 5.20 for different values of  $a_k^{(2)}$ : 1/16 (a), 1/8 (b), 3/16 (c) and 1/4 (d). One sees nicely an increasing deviation from a flat distribution with increasing outlier content.

Finally, we look at the distribution of the generalized  $\chi^2_{k,F}$  of the filter step and of the total generalized chi-square. As mentioned in Subsection 4.3.3,  $\chi^2_{k,F}$  can be generalized in two ways:

$$\chi^2_{k,F^{(i)}} = \mathbf{r}_k^T \mathbf{G}_k^{(i)} \mathbf{r}_k + (\tilde{\mathbf{x}}_k - \tilde{\mathbf{x}}^{k-1}_k)^T (\mathbf{C}^{k-1}_k)^{-1} (\tilde{\mathbf{x}}_k - \tilde{\mathbf{x}}^{k-1}_k),$$

with:

$$\mathbf{G}_k^{(1)} = (a_k^{(1)} \mathbf{V}_k^{(1)} + a_k^{(2)} \mathbf{V}_k^{(2)})^{-1},$$

$$\mathbf{G}_k^{(2)} = (b_k^{(1)} \mathbf{V}_k^{(1)} + b_k^{(2)} \mathbf{V}_k^{(2)})^{-1}.$$

Fig. 5.21 shows the  $\chi^2$ -probability transforms of  $\chi^2_{k,F^{(1)}}$  and  $\chi^2_{k,F^{(2)}}$  for all  $k$ , as well as of the two corresponding total  $\chi^2$ -statistics, for an average number of outliers of 2. Neither of them is really satisfactory.

We have compared the performance of the robust filter with two other filter methods: first, with the standard filter/smoothing with removal of outliers by the CS-test; secondly, with the linear filter which is optimal in the mixture model, i.e. a Kalman filter with the following covariance matrix of the measurements:

$$\mathbf{V}_k = a_k^{(1)} \mathbf{V}_k^{(1)} + a_k^{(2)} \mathbf{V}_k^{(2)}.$$

Fig. 5.22 shows the standard errors of the estimated track parameters, as a function of the average number of outliers, for the three filters. The optimal linear filter performs worst, and the robust filter does better than the smoother with removal of outliers. If, however, the average number of outliers per track is only 1 or 2, the difference between the two is very small. In this case the smoother has the advantage that its total  $\chi^2$ -statistic is more closely  $\chi^2$ -distributed than the one of the robust filter.

## F) Detection of outliers in FCA

The last example in this subsection is concerned with the detection of outliers in the chamber FCA, which is part of the forward spectrometer of the DELPHI detector. The detection of outliers is particularly important in FCA, because pattern recognition in FCA is made difficult



by the large number of secondary tracks which are expected to be produced in the massive endplate of the TPC. The track fit in the forward region of DELPHI proceeds from FCB towards the vertex (see Fig. 5.9). Since the momentum is measured very poorly in FCB, the predicted track parameters in FCA have rather large errors, so that there is very little chance to detect outliers during filtering. Consequently, the smoother is crucial for the detection of outliers in FCA. In our study, 20000 tracks were simulated with FASTSIM, at  $\theta=25^\circ$  and  $1 \leq p \leq 25$  GeV/c. Then outliers were generated in FCA, according to a variance-inflation model, with

$$V_{FCA}^{(2)} = 9 \cdot V_{FCA}^{(1)}.$$

The tracks were then passed through the forward track search and the forward track fit (see Subsection 5.1.2). More than 99.9% of the tracks were found by the track search, in spite of the outlier. In the track fit, the outlier was looked for with the CS-test at the 1% level. The power of this test is shown in Fig. 5.23, as a function of the distance  $d$  of the outlier from the true track position. (There are no outliers beyond  $d=16$ .) The CS-test is fully efficient at  $d=13$ , which is roughly the 1%-cut of a  $\chi^2$ -distribution with four degrees of freedom. Since FCA measures four quantities, we conclude that the chi-square of the smoother is dominated by the measurement errors of FCA.

### 5.2.3 Results of kink finding studies

We now turn to the detections of kinks in the DELPHI detector. We investigate first  $\pi$ - and K-decays in the TPC, and then show how the efficiency of kink finding can be improved by including the other two tracking detectors in the barrel region, ID and OD.

As mentioned in Subsection 4.4.1, the most frequent sources of kinks are the decay of a charged  $\pi$ -meson or a charged K-meson into a muon and a neutrino. In the case of the  $\pi$  this type of decay is virtually the only one; for the K its branching ratio is about 63.5%. The second most frequent decay type of the charged K-meson is also a two-body decay, namely the decay into a charged  $\pi$  and a  $\pi^0$ . This type, however, is less interesting in this context, because it is not a potentially dangerous source of background muons, as is the other one. Also, the  $\pi^0$  is detected much more easily than the neutrino which is virtually invisible in a typical storage ring detector.

In order to test the kink finders proposed in Subsection 4.4.2, three samples of 5000 tracks each were simulated in STPC: a sample of  $\pi$ -decays into muon plus neutrino ( $\Pi$ ), a sample of K-decays into muon plus neutrino (K), and a sample of tracks without decays for cross-checking (N). The precise characteristics of the three samples are listed in Table 5.4. It

should be noted that the decay radius, i.e. the radial distance of the decay vertex from the origin, was simulated unphysically, namely uniformly in a certain region of STPC, not according to the exponential life time distribution of the decaying particle. This was done in order to achieve a uniform population of the volume of STPC with decays. The decays were simulated with a standard decay routine [75]. The dependence of the simulated kink angle on the momentum of the  $\pi$  (K) is shown in Fig. 5.24a (b) (see also Fig. 4.4).

We recall that two tests for the detection of kinks were proposed in Subsection 4.4.2: a test based on the statistic  $\chi^2_{k,\Delta}$ , and a test based on the statistic  $F_k$ . The actual test statistics used were the following:

$$CK = \max \chi^2_{k,\Delta} \quad (3 \leq k \leq 14),$$

$$FK = \max F_k \quad (3 \leq k \leq 14).$$

Obviously a kink can be detected only if the track segments in front of and behind the decay are sufficiently well determined, i.e. if both of them contain at least three measurements in STPC. Because of this  $k$  ranges from 3 to 14. Both the  $\chi^2_{k,\Delta}$  and the  $F_k$  are **not** independent random variables; hence the exact distributions of CK and FK are difficult to compute. Therefore the 99%-quantile used in the test was determined empirically from sample N (no kinks).

If we want to specify the performance of the tests, the following questions arise:

- How many kinks are detected?
- How well can the location of the kink be determined?
- How well can the kink angle be determined?
- How does the efficiency of the kink finder depend on the momentum of the  $\pi$  (K), on the decay radius, and on the kink angle?
- How does the efficiency compare to that of a simple 1%-cut on the total chi-square?
- How many kinks are faked by outliers?

The answers to these questions can be gathered from Table 5.5 and Figs. 5.25-5.28. Table 5.5 shows in column 3 the percentage of kinks found, i.e. the efficiency, for both tests and both decay samples. Also shown is the fraction of tracks rejected by a 1%-cut on the total  $\chi^2$ -statistic. These numbers are meant only for comparison; their absolute size depends of course very much on the range of momentum of the decaying meson. We see immediately, that the FK-test is indeed less powerful than the CK-test; also the total  $\chi^2$  is a nearly as good indicator of a kink. Of course it doesn't tell us where the kink might be.

Next we consider the difference between the found position of the kink, i.e. the value of  $k$  where the maximum  $\chi^2_{k,\Delta}$  or  $F_k$  occurs, and the actual position (if the kink is between  $k$  and

$k+1$ , the position is  $k$ ). This is shown in Fig. 5.25 for the CK-test (a,c) and the FK-test (b,d), for both the  $\pi$ -sample (a,b) and the K-sample (c,d). If the found position differs from the true position by not more than 1, we define that the kink has been found in the "correct position". The fraction of tracks where the kink was found in the correct (wrong) position is given in the 4th (5th) column of Table 5.5. In brackets is the number of kinks found in the correct (wrong) position as a fraction of all found kinks.

The next two figures show the dependence of the kink finding efficiency on the momentum (a), on the position of the decay (b), and on the kink angle (c), for the  $\pi$ -sample (Fig.5.26) and for the K-sample (Fig. 5.27). The blank area corresponds to the CK-test, the shaded area to the FK-test. For  $\pi$ -decays the efficiency drops sharply with increasing momentum, and approaches zero at 10 GeV/c. For K-decays, it is still in the order of 50% at 40 GeV/c. The dependence on the position of the decay is rather weak in all cases; a slight loss of efficiency, however, can be observed towards both ends of the detector. As far as the dependence on the kink angle is concerned, the picture is more or less the same in all 4 cases: The efficiency rises steeply from zero and attains 100% at a value of about 30 mrad.

If a kink is found at  $k$ , the kink angle can be reconstructed from the state vectors of the two track segments in front of and behind the kink (see Subsection 4.4.2). Fig. 5.28 shows the difference between the reconstructed and the true kink angle, for the CK-test (a,c) and the FK-test (b,d), for both the  $\pi$ -sample (a,b) and the K-sample (c,d). In all cases there is a clear central peak which reflects the uncertainty of the track parameters of both track segments, and long tails which are due to kinks found in the wrong position. The tails are less pronounced in the case of the FK-test.

So far it has been assumed that all tracks were measured perfectly, without disturbance by outliers. We investigate now the influence of outliers on the kink finding algorithms. To this end the three samples  $\Pi$ , K and N (see Table 5.4) were contaminated with a single variance-inflation outlier (with threefold standard deviation both in  $R\Phi$  and  $z$ ). The position of the outlier was distributed uniformly over all measurements. The contaminated samples are called  $\Pi 1$ , K1 and N1.

The kink finding algorithms were then applied to the contaminated samples. The results are summarized in Table 5.6. We see immediately that with the CK-test a single outlier fakes a kink in no less than a quarter of the tracks in sample N1! As expected, the FK-test does much better: It fakes a kink in only about 8% of the tracks. For the samples with  $\pi$ - and K-decays there is an apparent increase in efficiency. Closer inspection, however, reveals that the increase is due to a much larger number of kinks found in the wrong position; the number of kinks found in the correct position actually drops in all but one cases, so that the additional kinks can be said to be faked by the outlier. As expected, the FK-test is much more robust

against the outlier. The price to be paid is a greater loss of efficiency than with the CK-test. The close correlation between the position of a faked kink and the position of the outlier can be seen in Fig. 5.29, where the difference of the position of the outlier and the found position of the kink is histogrammed for all tracks where a kink was found, but in the wrong position (a: Sample N1, b: Sample  $\Pi$ 1, c: Sample K1).

The obvious idea would be to remove the outlier before looking for a kink. There is, however, a fundamental problem here: If the procedure of removing outliers as described in subsection 5.2.2 is applied to a track with a kink, usually a large number of outliers is found. This is so because the estimated track does not fit very well the two tracks forming the kink – particularly, if the kink angle is large. The crucial point is therefore that outliers may only be removed if very few of them are found, preferably only one! Actually the problem is less serious than it may seem by now, because large kinks are already recognized by the pattern recognition and are not passed through the kink finder.

Consequently, the contaminated tracks were subject to the following procedure:

- Outliers were looked for with the  $\chi^2_{k,S}$ -test (see Subsection 5.2.2).
- If only one outlier was found it was removed.
- The track was passed through the kink finder.

The 1%-cuts on CK and FK were readjusted with sample N. The results of the foregoing procedure are summarized in Table 5.7, both for the original and the contaminated samples. A comparison with Table 5.5 shows that on the original samples the loss of efficiency is fairly small. For the contaminated samples an important reduction of faked kinks can be seen, but at the price of a somewhat lower efficiency.

We conclude that single outliers can be handled reasonably well with this procedure; in the presence of more than one outlier, however, the kink finder generates a lot of faked kinks which are difficult to get rid of without a serious loss of efficiency.

Our final investigation is concerned with finding kinks in the barrel region of DELPHI, by using not only information from the TPC, but from ID and OD as well (Fig. 5.7). We consider first decays inside the volume of TPC.

A sample  $\Pi'$  of 5000  $\pi$ -decays and a sample K' of 5000 K-decays were simulated in ID, TPC and OD, with the same characteristics as samples  $\Pi$  and K, with the exception of the decay radius being uniformly distributed in the entire volume of the TPC:

$$36.5 \text{ cm} \leq r_{\text{Decay}} \leq 106.625 \text{ cm.}$$

Due to the additional information provided by ID and OD, we expect a higher efficiency of kink finding than with the TPC alone. This is confirmed by the results obtained from the simulated tracks.

Table 5.8 shows the efficiency of the kink finder for samples  $\Pi'$  and  $K'$ , as well as the percentage of kinks found in the correct position. A direct comparison with the kink finder in TPC alone can be found in Fig. 5.30. Fig. 5.30 a (c) shows the efficiency of the CK-test for  $\pi$ -decays (K-decays), with all detectors (blank) and with only TPC (shaded), plotted as a function of the momentum of the  $\pi$  (K). The same function is shown in Figs. 5.30 b,d for the FK-test. Finally, Fig. 5.31 shows the efficiency with all detectors (blank) and with only TPC (shaded) as a function of the decay position, with the CK-test (a,c) and with the FK-test (b,d), for both the  $\pi$ -sample (a,b) and the K-sample (c,d). In the  $\pi$ -sample, there is a clear "plateau" of efficiency between the 3rd and the 10th measurement surface; towards OD the efficiency drops more than towards ID. The effect is less visible in the K-sample, due to the high overall efficiency. The detection of  $\pi$ -decays is clearly very much improved by the additional information provided by ID and OD.

As far as kinks outside the volume of the TPC are concerned, the situation is more difficult. First we note that the statistic  $\chi^2_{k,\Delta}$  at OD is basically the same as the chi-square of the filter,  $\chi^2_{k,F}$ , and thus only has three degrees of freedoms. Therefore it is in principle not possible to distinguish between an outlier in OD and a kink somewhere between TPC and OD. If  $\chi^2_{k,F}$  exceeds the cut, it can only be stated that the prediction from TPC and the measurement in OD are not compatible. The track segment in OD is not defined well enough to allow kinks to be recognized unambiguously.

#### 5.2.4 Results of secondary vertex detection studies

The results of a comparative study of the algorithms proposed in Section 4.6 will be presented here. Before this is done, however, a word of warning is required. It was not attempted to simulate and reconstruct real physical events in a real detector; therefore one should not attribute significance to the absolute performance of the algorithms with respect to a particular reaction. The aim of the study was solely to **compare** the performance under conditions which are hopefully not entirely different from the ones in real life. It should be kept in mind, however, that for the sake of convenience the simulation of events is very much simplified.

We start with the description of the 6 test samples used in the study, each of them containing 5000 events. The reference surface of the vertex fit is a cylinder of radius  $R = 10$  cm. The track parameters simulated on this reference surface are  $\Phi$ ,  $z$ ,  $\tan \lambda$  and  $\varphi$  (Fig. 5.32). It is

assumed that there is no magnetic field, hence the momentum measurement is not simulated. The measurement errors of the simulated track parameters correspond roughly to what can be expected in the DELPHI barrel region:

$$\sigma(\Phi) = 0.2 \text{ mrad}, \quad \sigma(z) = 0.2 \text{ mm}, \quad \sigma(\tan \lambda) = 0.001, \quad \sigma(\varphi) = 0.0003 \text{ rad}.$$

No correlations were taken into account. The vertex parameters to be estimated are the position  $(x,y,z)$  and the directions  $(\tan \lambda, \varphi)$  for each track.

At the primary vertex tracks were simulated with a uniform distribution of  $\lambda$  and  $\varphi$ :

$$\lambda \sim U[-\pi/4, \pi/4],$$

$$\varphi \sim U[0, 2\pi].$$

The secondary vertex was simulated according to the decay



which accounts for about 11% of all  $D^+$ -decays. The momentum of the D-meson was chosen to be 1 and 5 GeV/c. The decay length was not simulated according to the true exponential life-time distribution, but uniformly between 0 and 0.5 cm. The average decay length of a particle grows with the momentum due to the relativistic time dilatation, and is equal to about 0.6 mm (1.5 mm) for a D with 1 (5) GeV/c.

The first two samples were simulated with a primary multiplicity of 12 and a single secondary vertex, the momentum of the D being 1 GeV/c (sample A1) and 5 GeV/c (sample A5), respectively. The next two samples (B1 and B5) differ from the first two only by the primary multiplicity, which is equal to 6. In the last two samples (C1 and C5) there is no primary vertex, but two secondary vertices generated by two back-to-back D-mesons with a momentum of 1 (5) GeV/c each. The distance between the two decay vertices is again uniformly distributed between 0 and 0.5 cm.

The search for outlying tracks was performed with three different algorithms:

- Algorithm CS. This is a straightforward 1%-cut on the statistics  $\chi^2_{k,S}$  ("chi-square of the smoother") as defined in Subsection 4.6.1. All tracks exceeding the cut are considered as outliers (secondary tracks).
- Algorithm CR. This is the refined sequential test of Subsection 4.6.1. It first looks for the track with the largest  $\chi^2_{k,S}$ . If this exceeds a 1%-cut, the track is flagged as a secondary track and **removed immediately** from the fitted vertex, i.e. the smoother is recomputed without the track. This procedure is repeated until no  $\chi^2_{k,S}$  exceeds the 1%-cut.

- Algorithm CM. The vertex position is estimated by the M-estimator described in Subsection 4.6.2. The robustness constant was chosen to be equal to 1.5. The average number of iterations required until convergence is between 3 and 5, depending on the sample. Then the standard smoother is computed and the statistics  $\chi^2_{k,S}$  are subjected to a 1%-cut. All tracks exceeding the cut are considered as secondary tracks.

In all cases the vertex was fitted without any prior information on the vertex position. All three algorithms were checked with a sample of events with a primary multiplicity of six and without secondary vertices. The losses of primary tracks were at the expected 1%-level. The probability transform of the total chi-squares are shown in Fig. 5.33, for the CS-algorithm after removal of outliers (a) and for the CM-algorithm (b). Both show a slight depletion at small probabilities (large chi-squares).

We first consider the results of the secondary vertex search on the high multiplicity samples A1 and A5, which are summarized in Table 5.9. It shows the average number of secondary tracks found per event, the average and the maximum number of primary tracks lost per event, the fraction of events where **all** secondary tracks were found, and the fraction of events where all secondary tracks were found **without loss of primary tracks**. (In brackets is the last fraction relative to the last but one.) The following facts can be observed:

- The momentum of the D has a strong influence on the results. Because of kinematic effects, the decay products point more closely to the primary vertex if the D has a higher momentum. Therefore in sample A1 the "outlyingness" of the secondary tracks is much more pronounced than in sample A5. This is fortunate, as it compensates for the shorter decay lengths of the low momentum D.
- The CS-test is completely unsuitable to this type of outlier problem. The primary vertex is distorted to such an extent that in sample A1 **no or only one track survives the 1%-cut in more than 40% of the events**. The situation is somewhat better in sample A5, but there are still heavy losses of primary tracks.
- The CR-test does remarkably well in both sample A1 and A5. The losses of primary tracks never exceed two tracks per event. In sample A5, the CM-test does even better, although the losses are slightly higher. A complete picture of the losses of primary tracks is given in Fig. 5.34. In sample A1, the CM-test is slightly worse than the CR-test, and the losses are still somewhat higher. In one event out of 5000, all tracks are above the cut. This indicates that this particular M-estimator is still not robust enough, i.e. that still too much weight is given to far outlying tracks.

A similar behaviour, even more pronounced, is shown by samples B1 and B5, with a much lower primary multiplicity (Table 5.10). The failure of the CS-test is even more spectacular in

sample B1: In more than 50% of the events all tracks are above the 1%-cut on  $\chi^2_{k,S}$ . In sample B5 the CM-test is again slightly better than the CR-test; in sample B1, however, the losses of the CM-test are quite serious. The efficiency of secondary vertex detection shows the expected drop when passing from 12 primary tracks to 6 primary tracks; the drop is however surprisingly small.

An important point is the dependence of the efficiency of secondary vertex detection on the decay length. Fig. 5.35 shows the fraction of secondary vertices found with the CR-test (a) and with the CM-test (b) as a function of the decay length, for samples A1 and A5 (shaded). Fig. 5.36 a(b) shows the efficiency of the CR-test for samples A1 (A5) and B1 (B5) (samples B1 and B5 shaded).

If we turn to samples C1 and C5 which contain events with two back-to-back decays, the question to be asked is now: How well can two decay vertices be separated, provided that all tracks belonging to them have been identified? The answer is given in Table 5.11, which shows the fraction of perfectly separated decay vertices for samples C1 and C5. (The separation is called perfect if all tracks are assigned to the correct vertex.) Both the CS- and the CM-algorithm fail completely in this symmetric configuration; the reason is that in both cases the estimated vertex is somewhere between the two actual decay vertices, and all tracks appear more or less to be outliers. The CR-algorithm, however, is different: Once a track has been removed from vertex 1, the new estimate moves towards vertex 2, so that there is a fair probability that the next track will also be removed from vertex 1, and so on, until the estimate settles on one of the two vertices. The efficiency of the CR-test as a function of the distance between the vertices is shown in Fig. 5.37, for samples C1 and C5 (shaded).

We conclude that the CR-algorithm is very well suited to the fast and efficient detection of outlying tracks. If there are no outlying tracks, there is no additional computation to be done. The M-estimator with Huber's  $\psi$ -function is slightly better in some cases, but not robust enough in other ones. It fails completely in the symmetric case of two decay vertices with similar characteristics. Clearly other types of robust estimators are needed to handle this case, e.g. those with high break-down points. This is, however, beyond the scope of the present work.



Surface	Radius [cm]	Rad. length [%]	$\sigma(R\Phi)$ [cm]	$\sigma(z)$ [cm]	$\sigma(\theta)$ [rad]	$\sigma(\varphi)$ [rad]	$\sigma(1/p_T)$ [c/GeV]
ID1	17.	0.	0.002	0.022	0.015	0.0007	0.1375
ID2	22.3	2.8	—	—	—	—	—
TP4	30.75	2.5	—	—	—	—	—
TPC1	36.5	0.	0.02	0.06	—	—	—
...							
TPC16	106.625	0.	0.02	0.06	—	—	—
TP5	127.1	22.3	—	—	—	—	—
RI1	147.3	6.6	—	—	—	—	—
RI2	188.	9.3	—	—	—	—	—
OD	200.	0.	0.012	2.	—	0.006	—

Table 5.1

Detectors and scatterers in the barrel region of DELPHI.

Sample	Nr. of tracks	Rad. length	Initial conditions
I	5000	—	$R=36.5, 0 \leq \Phi \leq 2\pi,$
II	5000	1000.	$z=0., \pi/4 \leq \theta \leq 3\pi/4,$
III	5000	200.	$\varphi = \Phi, 0.5 \leq p \leq 5.$

Table 5.2

Characteristics of the three samples of tracks used in the outlier studies in STPC.

Sample	Outliers	CS-test		CG-test		CF-test	
		Power[%]	Losses[%]	Power[%]	Losses[%]	Power[%]	Losses[%]
I	0	—	1.0	—	1.0	—	1.0
I	1	53.7	1.3	54.2	1.6	29.5	1.8
I	2	54.4	1.7	55.5	2.2	25.6	2.4
I	3	52.6	2.0	53.9	2.7	20.2	2.8
III	0	—	1.0	—	1.0	—	1.0
III	1	50.6	1.3	36.1	1.9	27.3	1.9
III	2	51.5	1.8	39.2	3.2	22.6	2.5
III	3	49.6	2.2	39.0	4.3	19.3	2.9

Table 5.3

Power of CS-, CG- and CF-test for samples I and III.

Sample	Nr. of tracks	Decay type	Decay radius [cm]	Momentum range [GeV/c]
N	5000	————	————	$0.5 \leq p \leq 10.0$
$\Pi$	5000	$\pi \rightarrow \mu\bar{\nu}$	$45.85 \leq r \leq 97.275$	$0.5 \leq p \leq 10.0$
K	5000	$K \rightarrow \mu\bar{\nu}$	$45.84 \leq r \leq 97.275$	$2.0 \leq p \leq 40.0$

Table 5.4

Characteristics of the three samples used in the kink finding studies in STPC.

Sample	Test	Found [%]	Found correctly [%]	Found wrongly [%]	Not found [%]
$\Pi$	CK	38.8	25.2 (65.8)	13.2 (34.2)	61.2
$\Pi$	FK	32.1	22.3 (69.4)	9.8 (30.6)	67.9
$\Pi$	total $\chi^2$	29.9	————	————	70.1
K	CK	80.9	68.5 (84.7)	12.4 (15.3)	19.1
K	FK	75.4	65.9 (87.4)	9.5 (12.6)	24.6
K	total $\chi^2$	74.6	————	————	25.4

Table 5.5

Efficiency of CK-test, FK-test and total  $\chi^2$ .

Sample	Test	Found [%]	Found correctly [%]	Found wrongly [%]	Not found [%]
N1	CK	24.9*	————	24.9 (100.)	————
N1	FK	8.2*	————	8.2 (100.)	————
$\Pi 1$	CK	57.3	27.2 (47.5)	30.1 (52.5)	42.7
$\Pi 1$	FK	32.6	19.2 (59.1)	13.3 (40.9)	67.4
K1	CK	87.1	65.1 (74.8)	21.9 (25.2)	12.9
K1	FK	73.0	59.7 (81.7)	13.3 (18.3)	27.0

Table 5.6

Efficiency of CK-test and FK-test in the presence of a single outlier.

(\* Faked by the outlier)

Sample	Test	Found [%]	Found correctly [%]	Found wrongly [%]	Not found [%]
N	CK	1.0	—————	1.0 (100.)	—————
N	FK	1.0	—————	1.0 (100.)	—————
Π	CK	35.0	23.2 (66.4)	11.8 (33.6)	65.0
Π	FK	29.8	21.1 (70.6)	8.8 (29.4)	70.2
K	CK	77.8	66.6 (85.0)	11.7 (15.0)	22.1
K	FK	73.5	64.2 (87.3)	9.3 (12.7)	26.5
N1	CK	11.3	—————	11.3 (100.)	—————
N1	FK	3.6	—————	3.6 (100.)	—————
Π1	CK	45.9	23.6 (51.4)	22.3 (48.6)	54.1
Π1	FK	30.1	18.2 (60.4)	11.9 (39.6)	69.9
K1	CK	83.0	63.2 (76.1)	19.8 (23.9)	17.0
K1	FK	71.6	58.7 (82.0)	12.9 (18.0)	28.4

Table 5.7

Efficiency of CK-test and FK-test with removal of outliers.

Sample	Test	Found [%]	Found correctly [%]	Found wrongly [%]	Not found [%]
Π'	CK	81.4	47.8 (58.8)	33.5 (41.2)	18.6
Π'	FK	78.1	46.4 (59.4)	31.7 (40.6)	21.9
Π'	total $\chi^2$	74.6	—————	—————	25.4
K'	CK	93.8	81.5 (86.8)	12.3 (13.2)	6.2
K'	FK	92.5	81.2 (87.7)	11.3 (12.3)	7.5
K'	total $\chi^2$	90.8	—————	—————	9.2

Table 5.8

Efficiency of kink finding in TPC with information from ID, TPC and OD.

Sample	A1	A1	A1	A5	A5	A5
Test	CS	CR	CM	CS	CR	CM
Average # of secondary tracks found	2.76	2.78	2.78	2.48	2.41	2.44
Average # of primary tracks lost	8.64	0.13	0.43	4.90	0.13	0.18
Maximum # of primary tracks lost	12	2	12	12	2	4
Fraction of secondary vertices found [%]	85.1	86.9	86.3	66.2	61.8	63.5
Fraction of sec. vertices found without losses [%]	0.94 (1.1)	76.1 (87.6)	56.2 (65.2)	8.54 (12.9)	54.0 (87.4)	51.9 (81.7)
Average # of iterations			4.6			3.2

Table 5.9

Efficiency of secondary vertex detection with CS-, CR- and CM-test for samples A1 and A5.

Sample	B1	B1	B1	B5	B5	B5
Test	CS	CR	CM	CS	CR	CM
Average # of secondary tracks found	2.71	2.67	2.72	2.49	2.36	2.42
Average # of primary tracks lost	4.90	0.28	2.49	3.12	0.10	0.41
Maximum # of primary tracks lost	6	6	6	6	4	6
Fraction of secondary vertices found [%]	81.5	82.0	81.4	67.2	58.1	61.6
Fraction of sec. vertices found without losses [%]	0.16 (0.2)	75.6 (92.1)	8.5 (10.4)	5.2 (7.7)	54.0 (93.0)	36.1 (58.6)
Average # of iterations			4.7			3.4

Table 5.10

Efficiency of secondary vertex detection with CS-, CR- and CM-test for samples B1 and B5.

Sample	C1	C1	C1	C5	C5	C5
Test	CS	CR	CM	CS	CR	CM
Fraction of perfectly separated decay vertices [%]	0.12	66.1	1.04	0.18	39.5	2.6

Table 5.11  
Efficiency of secondary vertex separation with CS-, CR- and CM-test  
for samples C1 and C5.

### Execution times of global fit, filter and smoother

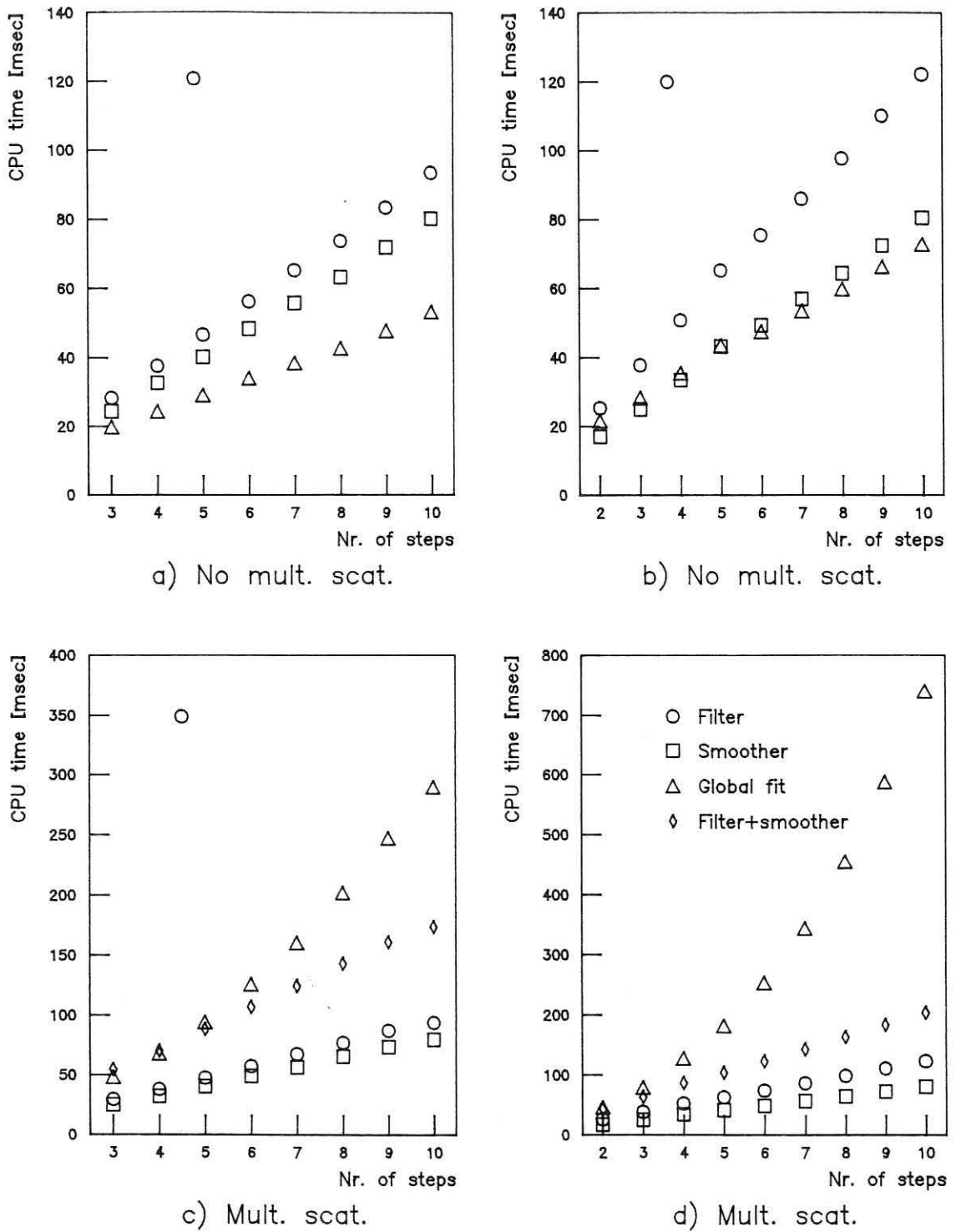


Fig. 5.1: Execution times of global fit, filter and smoother ( $\mu$ VAX II).

- a) Helix tracks, 2 measurements/step, no multiple scattering;
- b) Helix tracks, 4 measurements/step, no multiple scattering;
- c) Helix tracks, 2 measurements/step, with multiple scattering;
- d) Helix tracks, 4 measurements/step, with multiple scattering.

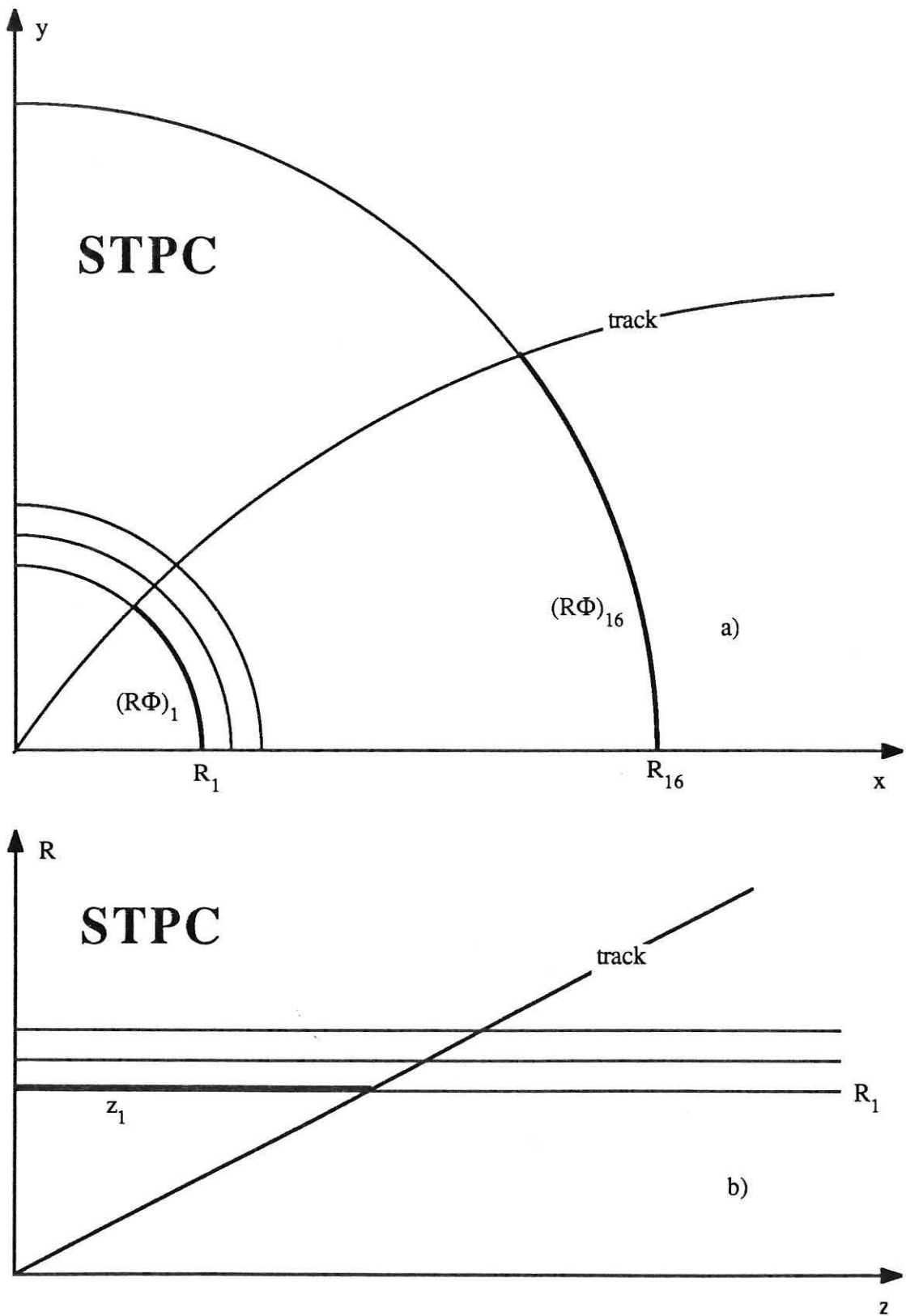


Fig. 5.2: View of the simplified TPC (STPC).  
 a) Cross section perpendicular to the beam; b) Cross section parallel to the beam.

### Comparison of filter, smoother and global fit (no multiple scattering)

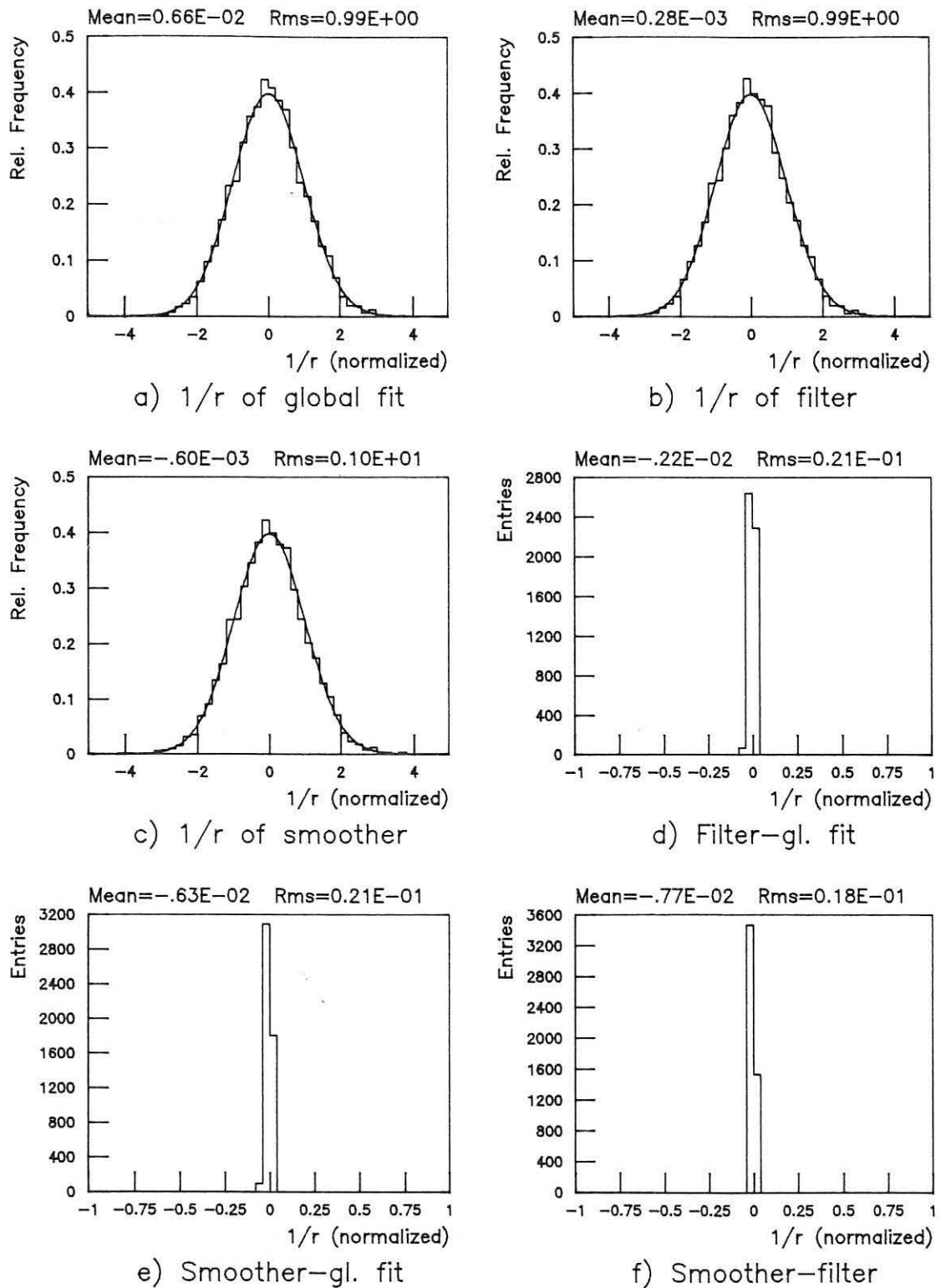


Fig. 5.3: Comparison of filter, smoother and global fit (no m.s.).

a) - c) Normalized difference of estimated  $1/r$  and true  $1/r$  at  $R = 36.5$ ; standard normal p.d.f.

d) - f) Pairwise differences of estimates of  $1/r$ , normalized by the standard deviation.



### Comparison of filter, smoother and global fit (no multiple scattering)

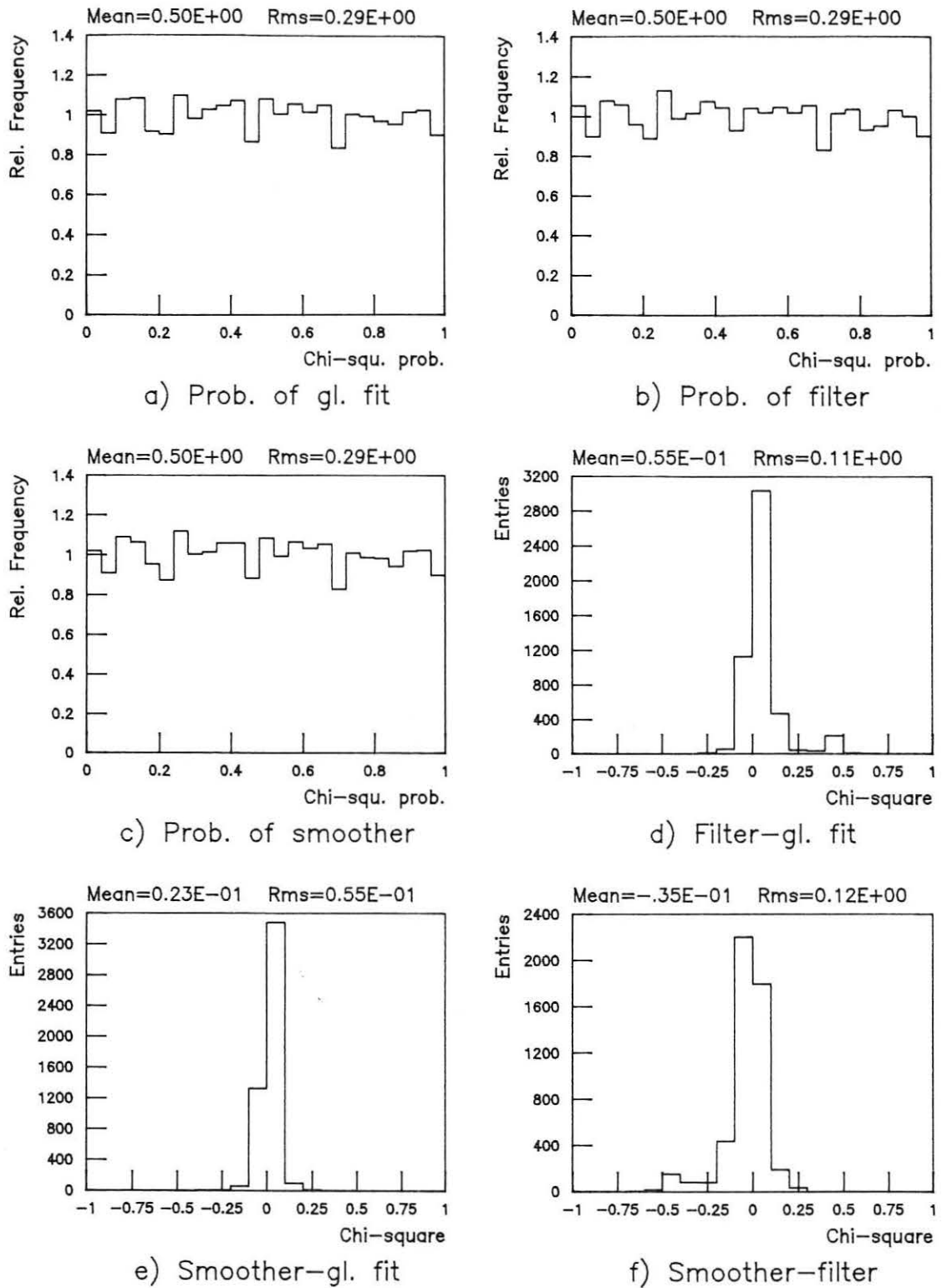


Fig. 5.4: Comparison of filter, smoother and global fit (no m.s.).

a) - c) Probability transforms of total chi-squares;

d) - e) Pairwise differences of total chi-squares.

### Comparison of filter, smoother and global fit (strong multiple scattering)

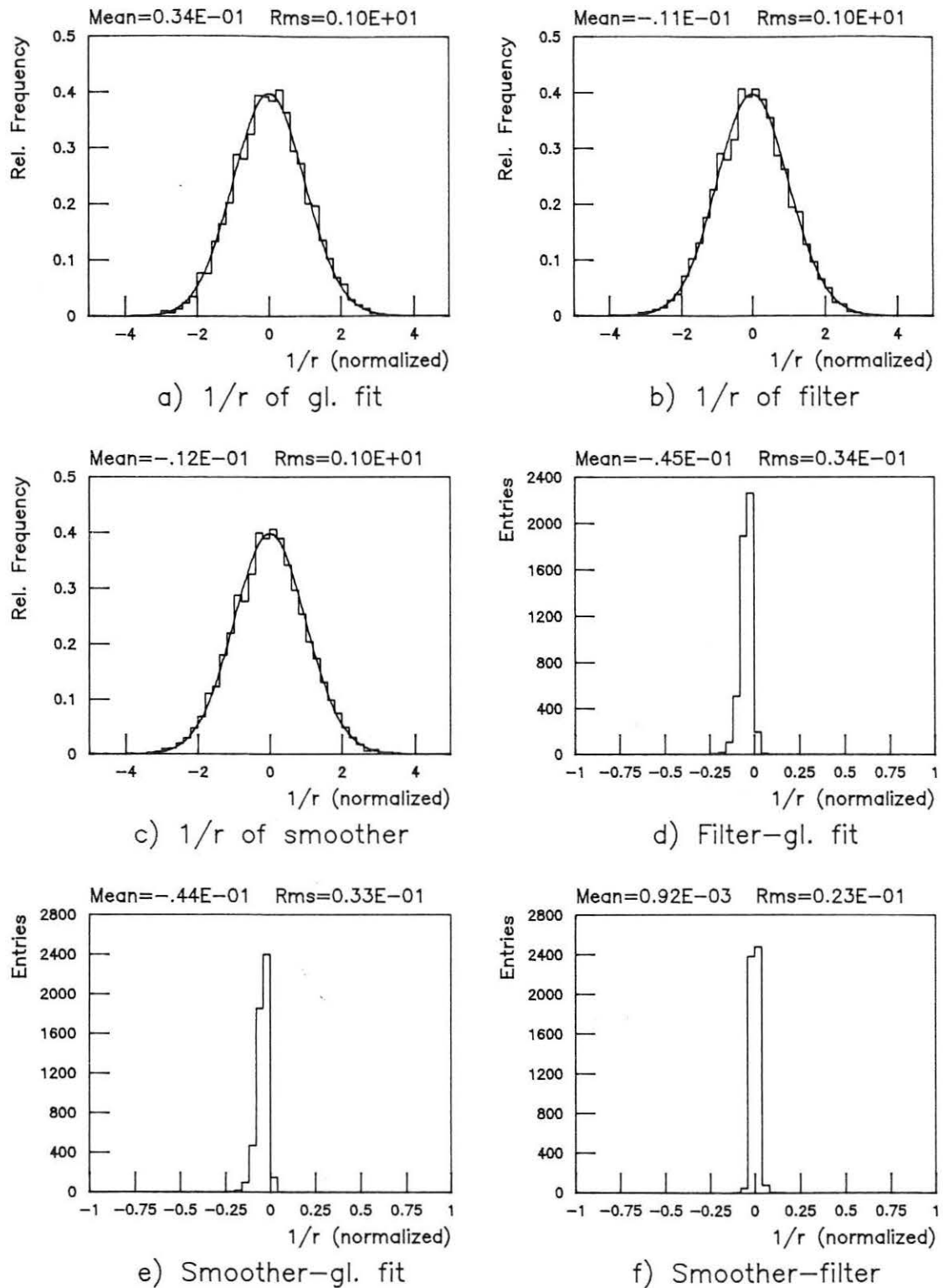


Fig. 5.5: Comparison of filter, smoother and global fit (strong m.s.).

a) - c) Normalized difference of estimated  $1/r$  and true  $1/r$  at  $R = 36.5$ ; standard normal p.d.f.

d) - f) Pairwise differences of estimates of  $1/r$ , normalized by the standard deviation.

Comparison of filter, smoother and global fit (strong multiple scattering)

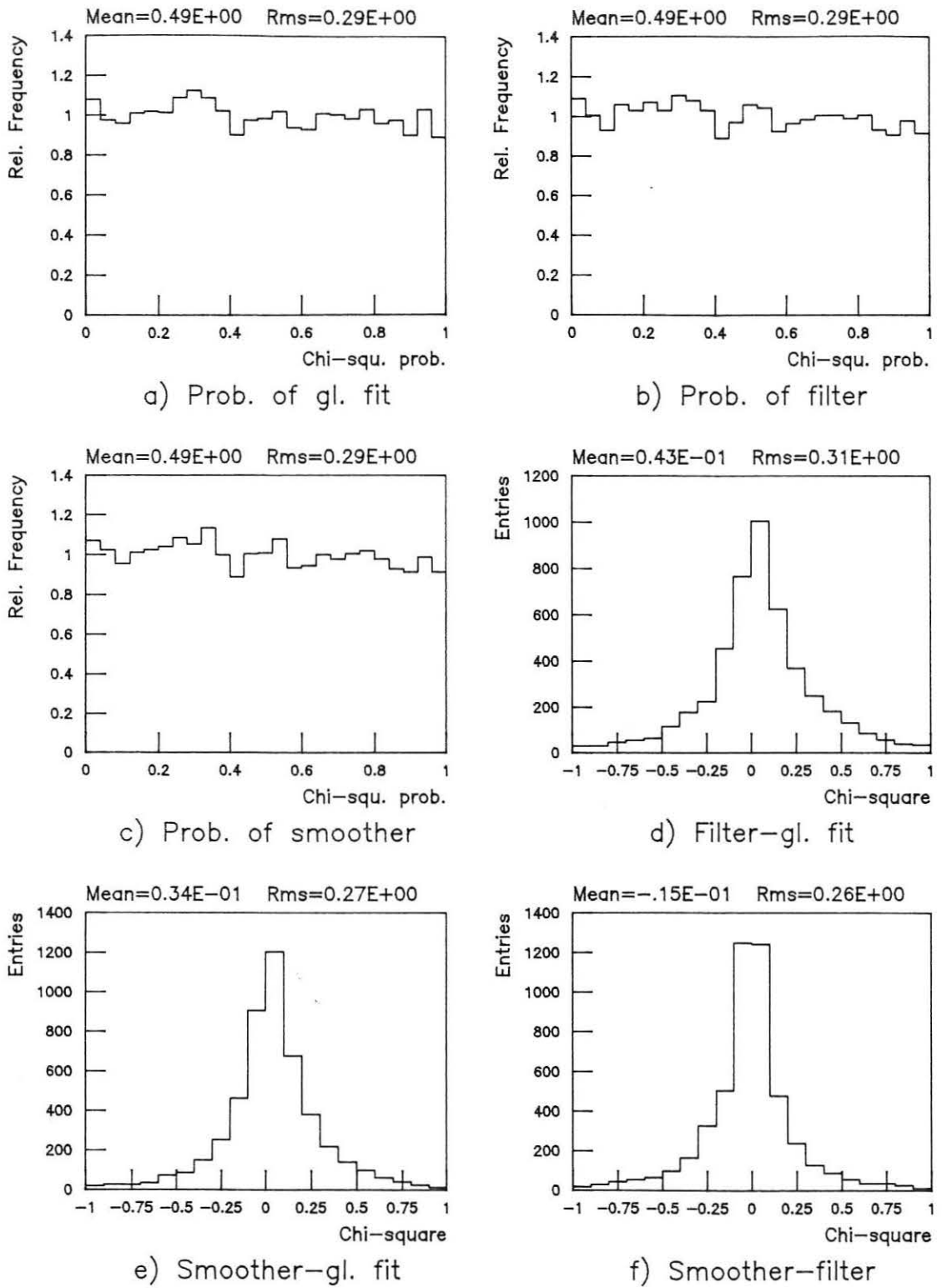


Fig. 5.6: Comparison of filter, smoother and global fit (strong m.s.).

a) - c) Probability transforms of total chi-squares;

d) - e) Pairwise differences of total chi-squares.

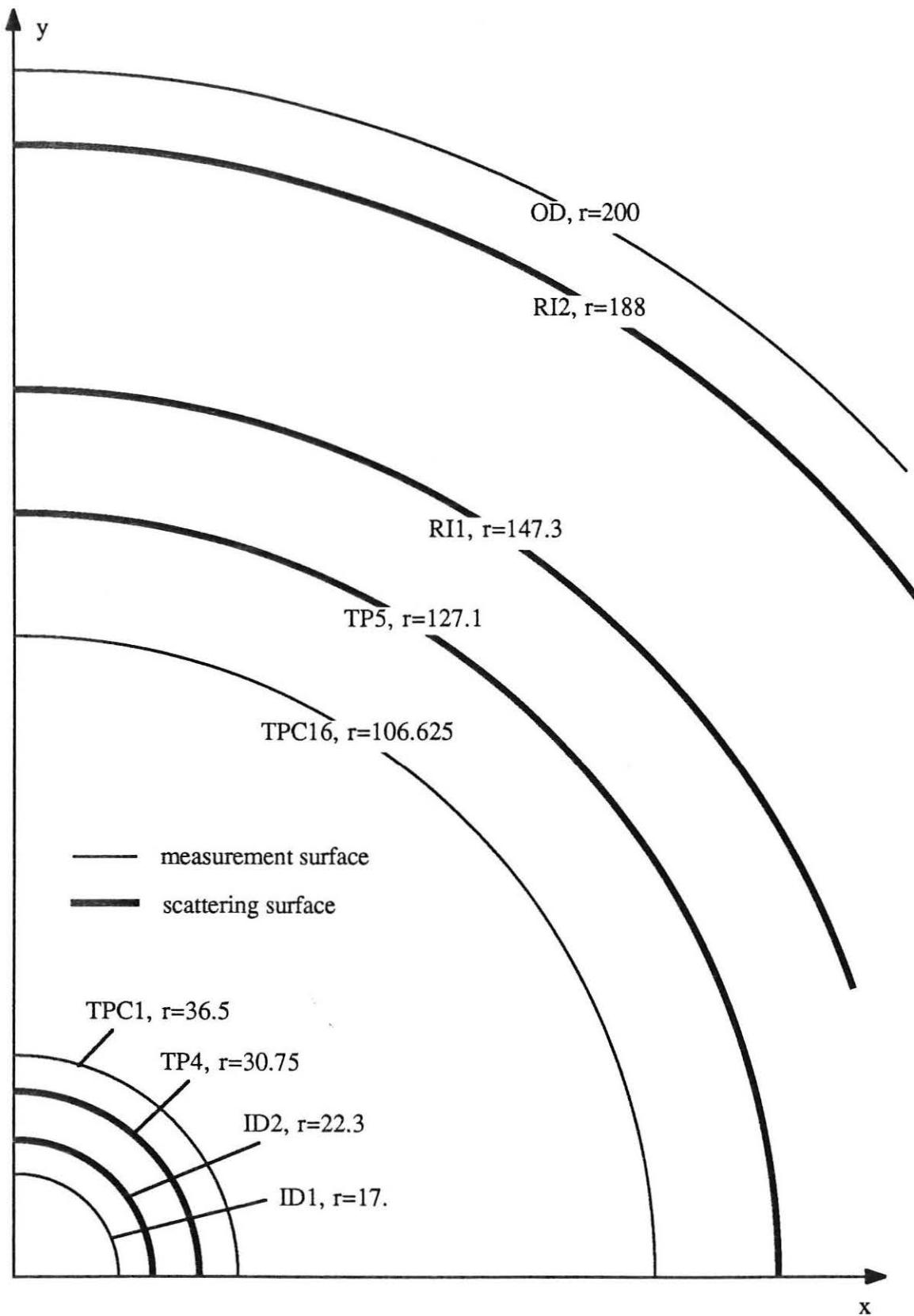


Fig. 5.7: The detector configuration in the barrel region of DELPHI (see also Table 5.1).

## Comparison of smoothed and extrapolated estimate in OD

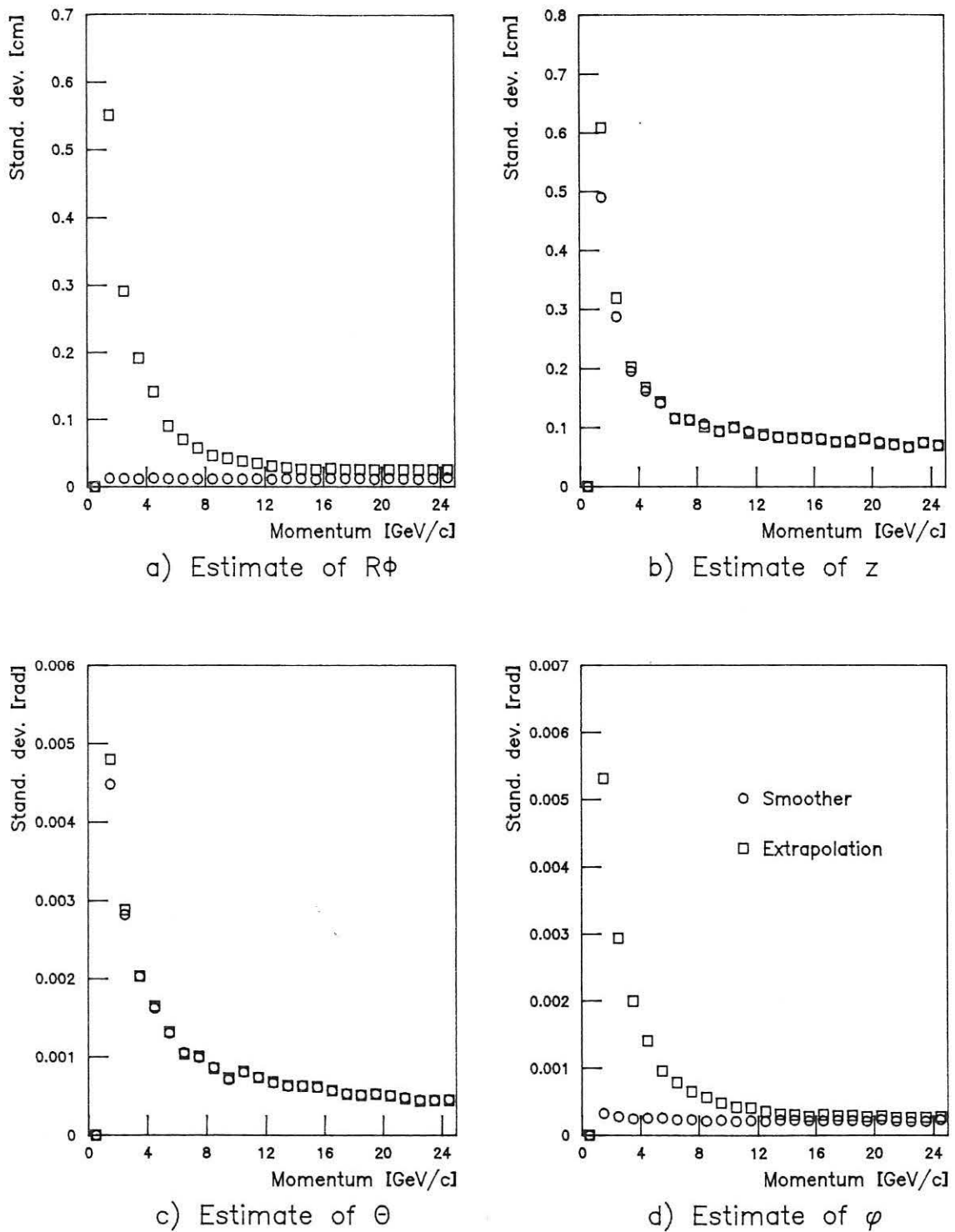
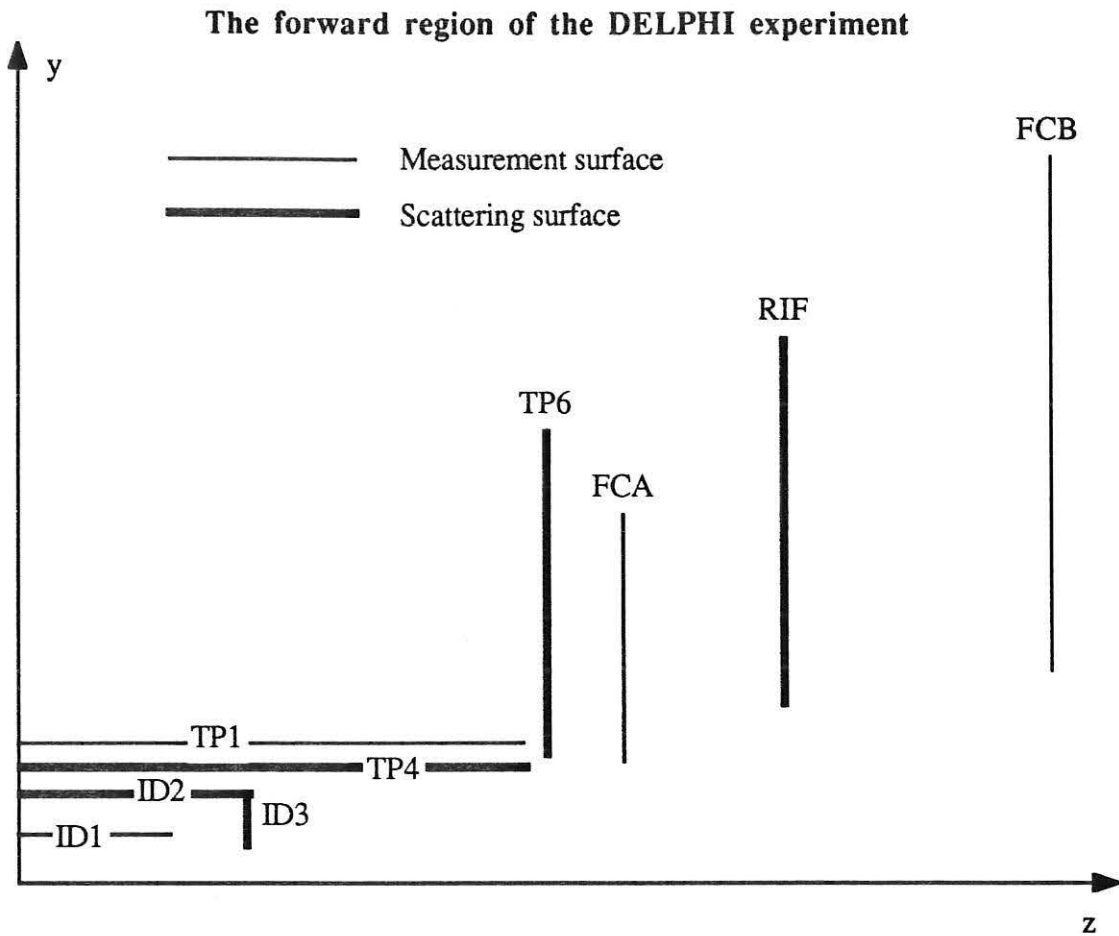


Fig. 5.8: Standard errors of the smoothed track parameters in OD and of the filtered track parameters in ID, extrapolated to OD.



Surface	$R_{\min}$ [cm]	$R_{\max}$ [cm]	$z_{\max}$ [cm]	Rad. length [%]
ID1	12.	12.	40.	0.
ID2	22.3	22.3	60.	2.8
ID3	12.2	22.3	60.	26.3
TP4	30.75	30.75	134.	2.5
TP1	36.5	36.5	134.	0.
TP6	34.	117.	140.	30.
FCA	31.7	98.	160.	0.
RIF	46.8	147.3	203.	28.1
FCB	55.3	192.	274.	0.

Fig. 5.9: The detector configuration in the forward region of DELPHI as used by FASTSIM.  
View in the y-z-plane.

### Comparison of smoothed and extrapolated estimate in FCB

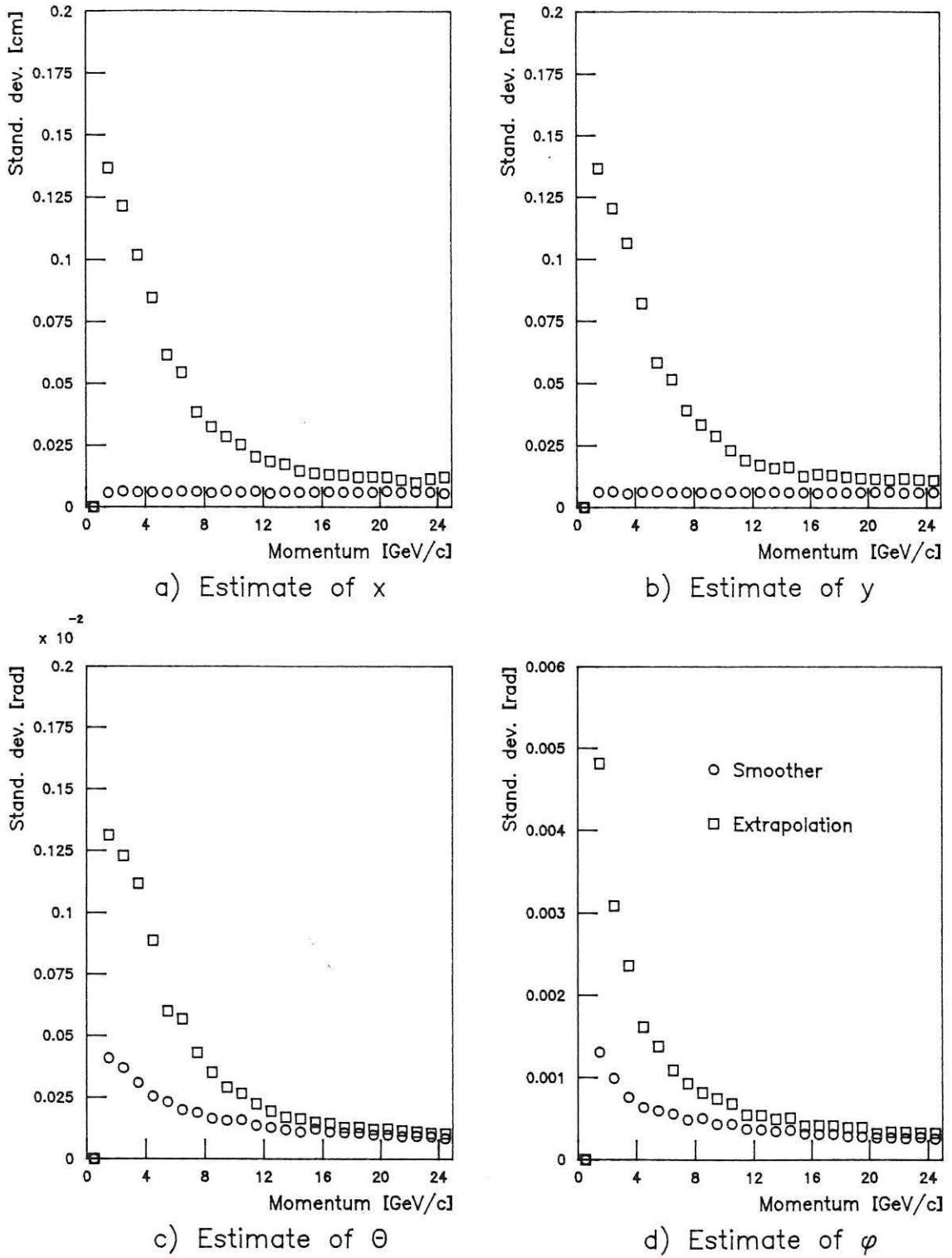
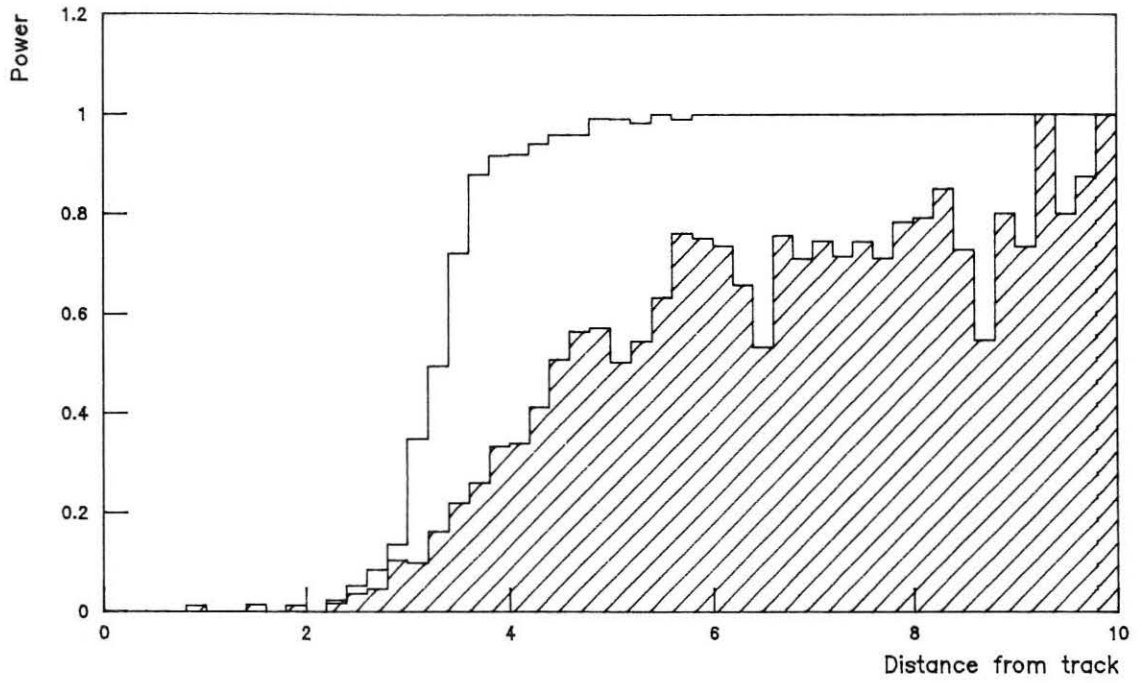
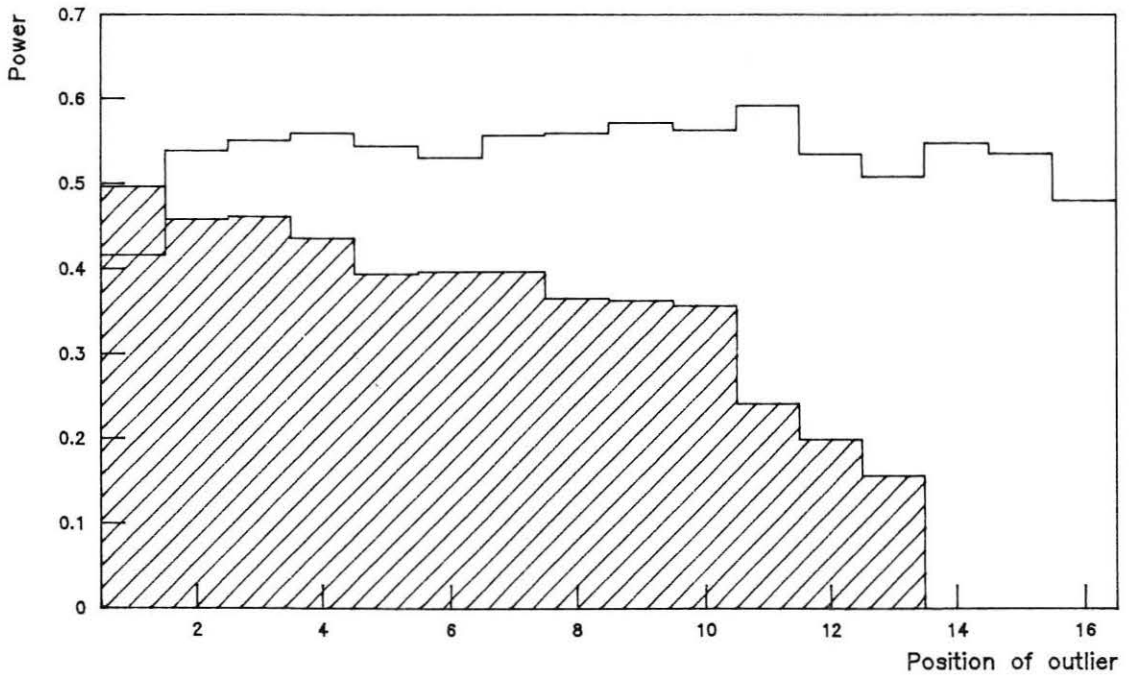


Fig. 5.10: Standard errors of the smoothed track parameters in FCB and of the filtered track parameters in ID, extrapolated to FCB.

Single variance-inflation outlier, no multiple scattering



a) Comparison of power



b) Comparison of power

Fig. 5.11: Power of CS-test (blank) and CF-test (shaded) for sample Ia.

a) As a function of the distance  $d$  of the outlier from the true track position;

b) As a function of the position  $k$  of the outlier.



### Single variance-inflation outlier, no multiple scattering

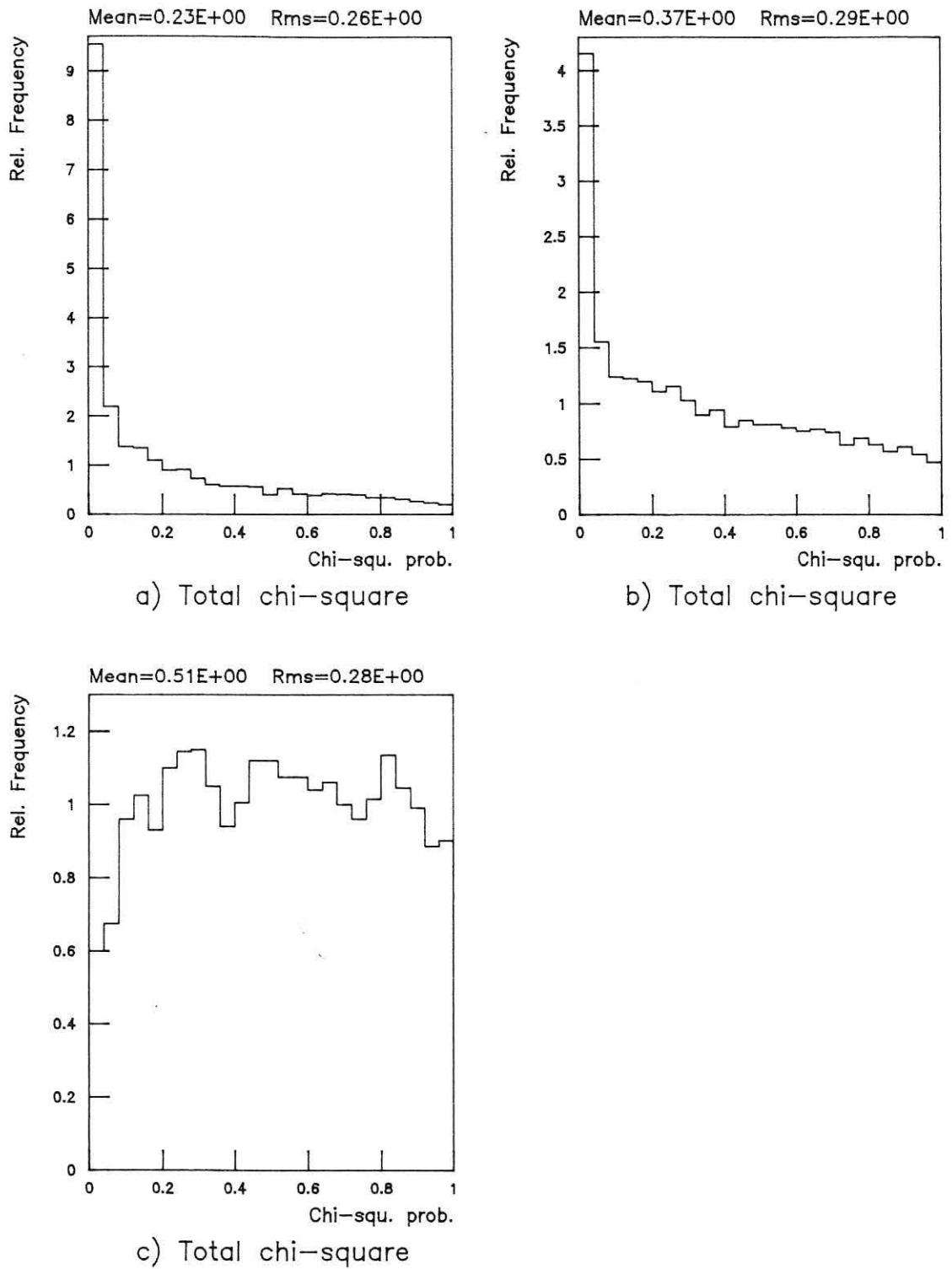
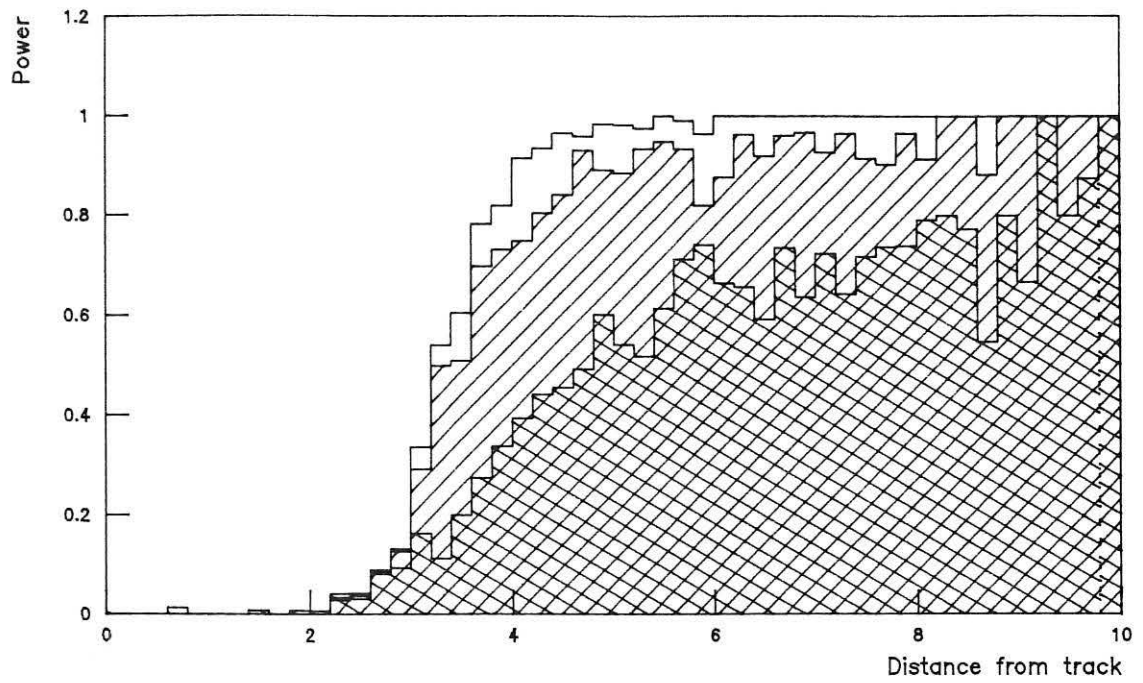


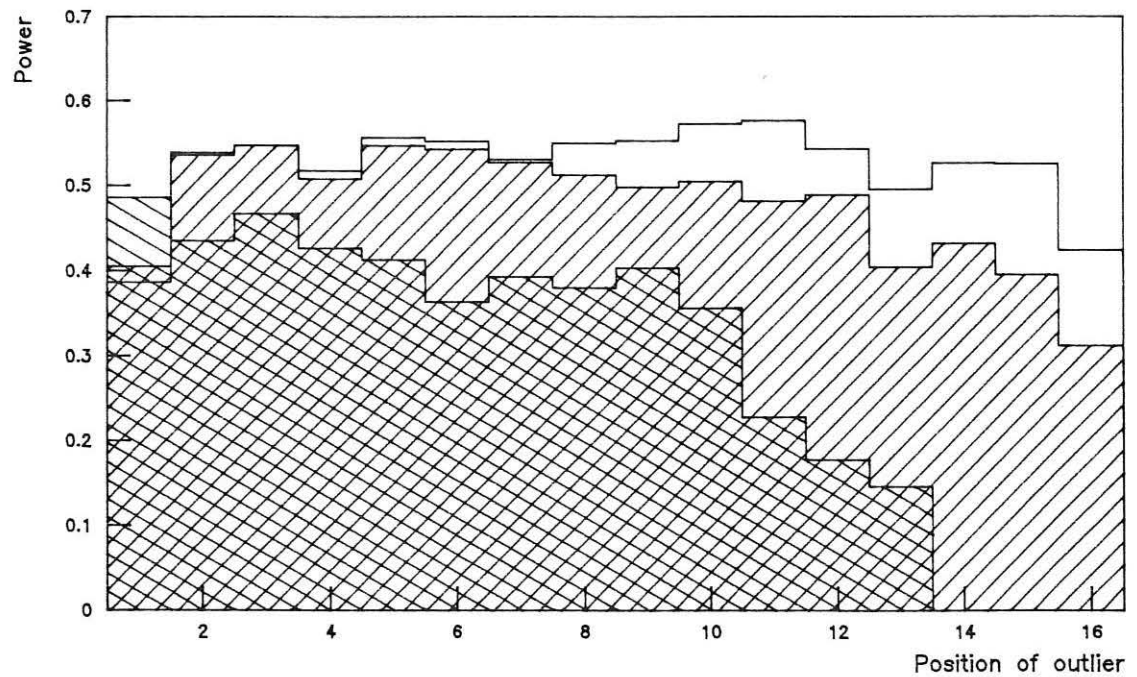
Fig. 5.12: Probability transform of total chi-square (Sample Ia).

a) Outliers not removed; b) Outliers removed with CF-test; c) Outliers removed with CS-test.

Single variance-inflation outlier, little multiple scattering



a) Comparison of power



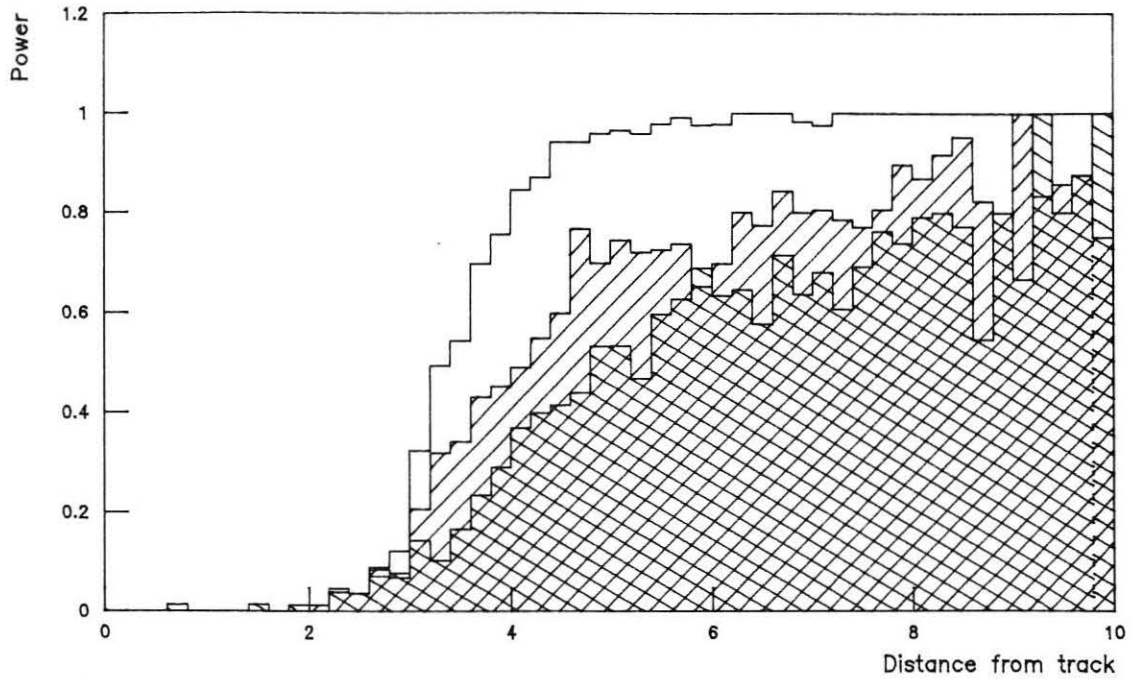
b) Comparison of power

Fig. 5.13: Power of CS-test (blank), of CG-test (shaded) and of CF-test (doubly shaded) for sample IIa.

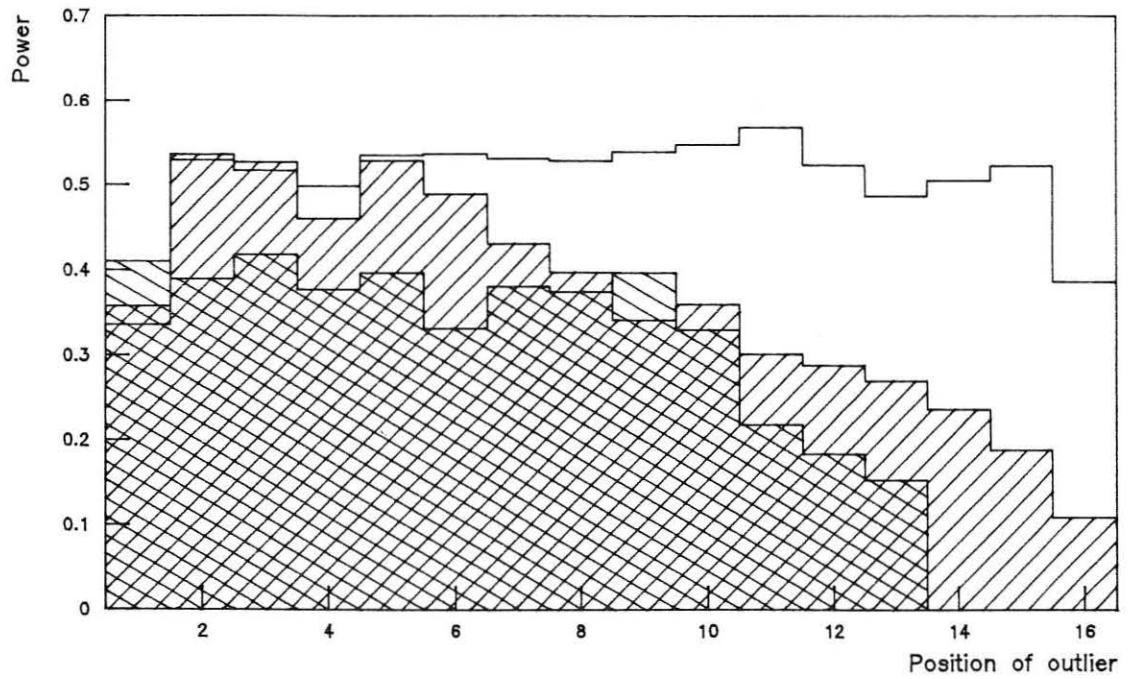
a) As a function of the distance  $d$  of the outlier from the true track position;

b) As a function of the position  $k$  of the outlier.

**Single variance-inflation outlier, strong multiple scattering**



a) Comparison of power

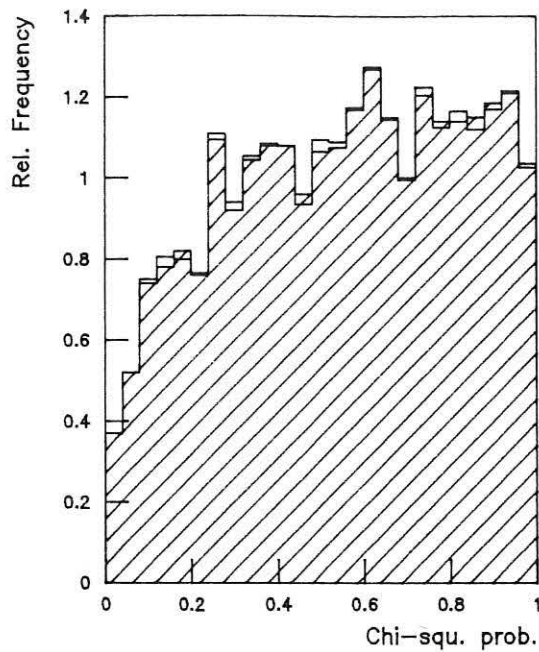


b) Comparison of power

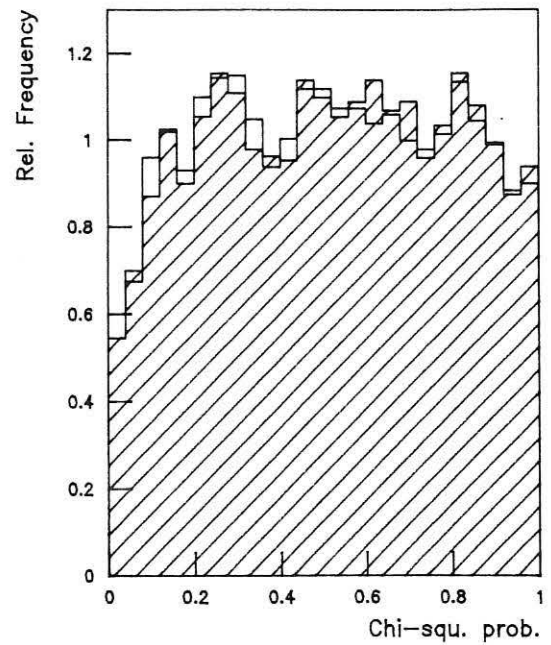
**Fig. 5.14: Power of CS-test (blank), of CG-test (shaded) and of CF-test (doubly shaded) for sample IIIa.**

- a) As a function of the distance  $d$  of the outlier from the true track position;
- b) As a function of the position  $k$  of the outlier.

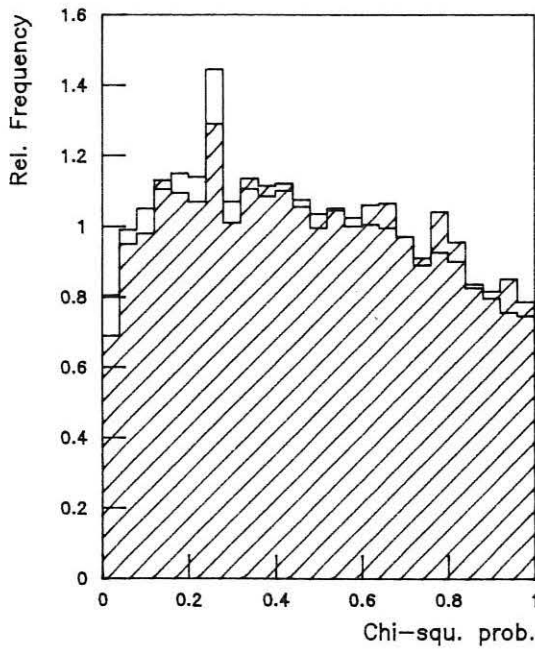
### 0-3 variance-inflation outliers, no multiple scattering



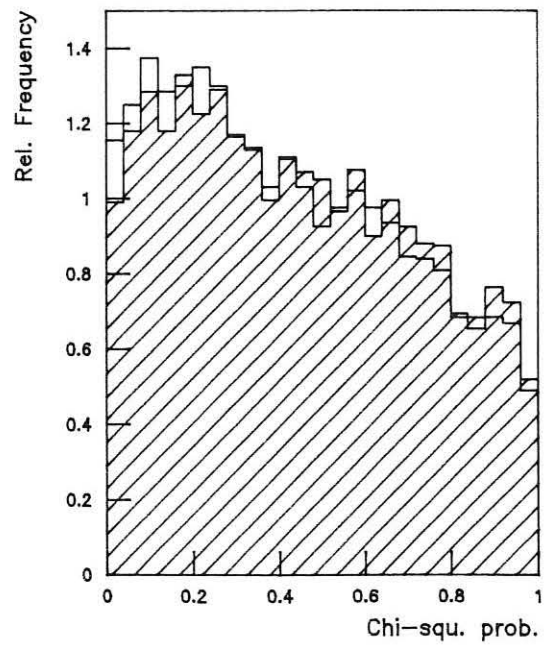
a) No outliers



b) 1 outlier



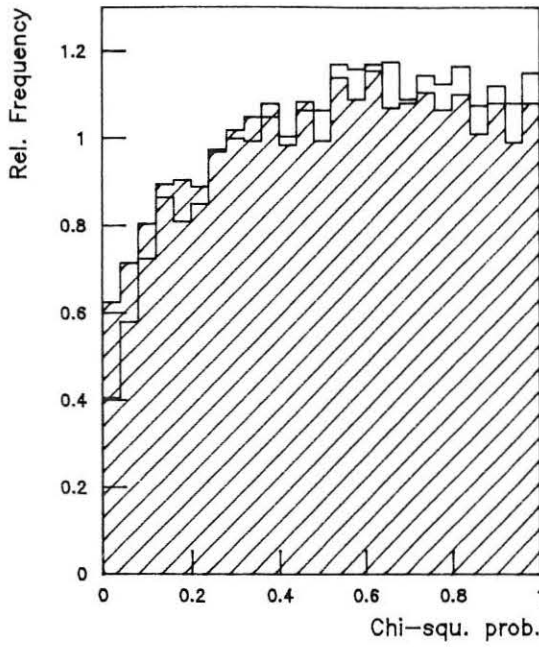
c) 2 outliers



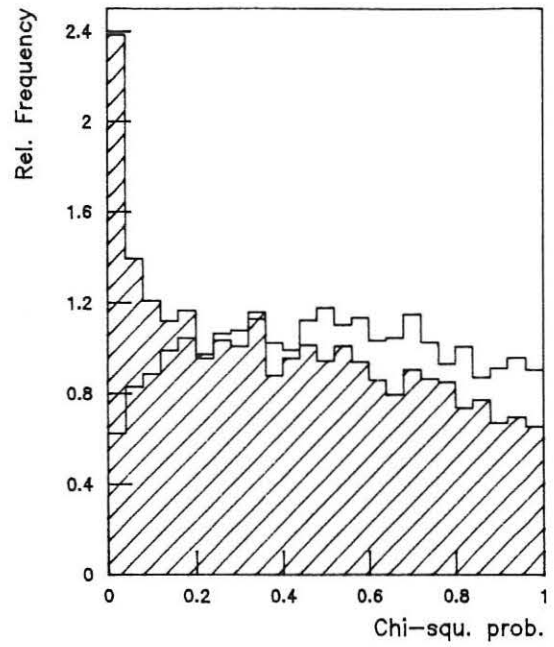
d) 3 outliers

Fig. 5.15: Probability transforms of the total chi-square of the smoother (blank) and of the global fit (shaded) after removal of outliers (Sample I).

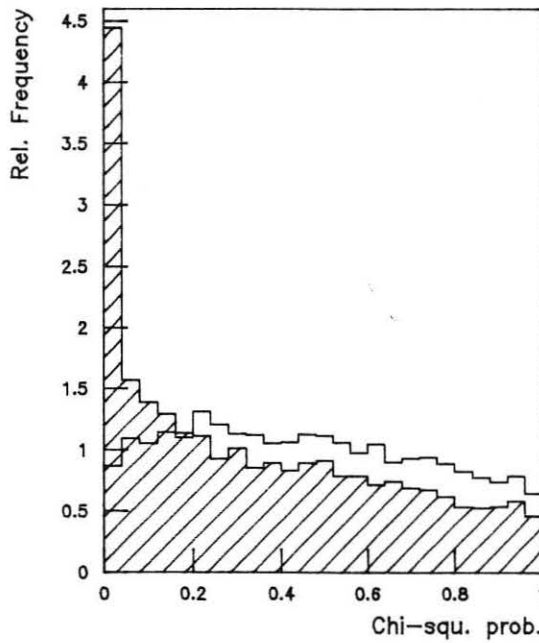
0-3 variance-inflation outliers, strong multiple scattering



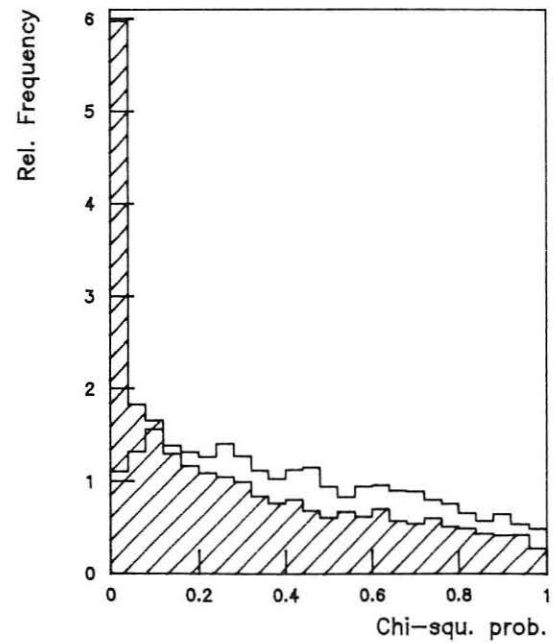
a) No outliers



b) 1 outlier



c) 2 outliers



d) 3 outliers

Fig. 5.16: Probability transforms of the total chi-square of the smoother (blank) and of the global fit (shaded) after removal of outliers (Sample III).

### Influence of outliers on estimates of track parameters (no m.s.)

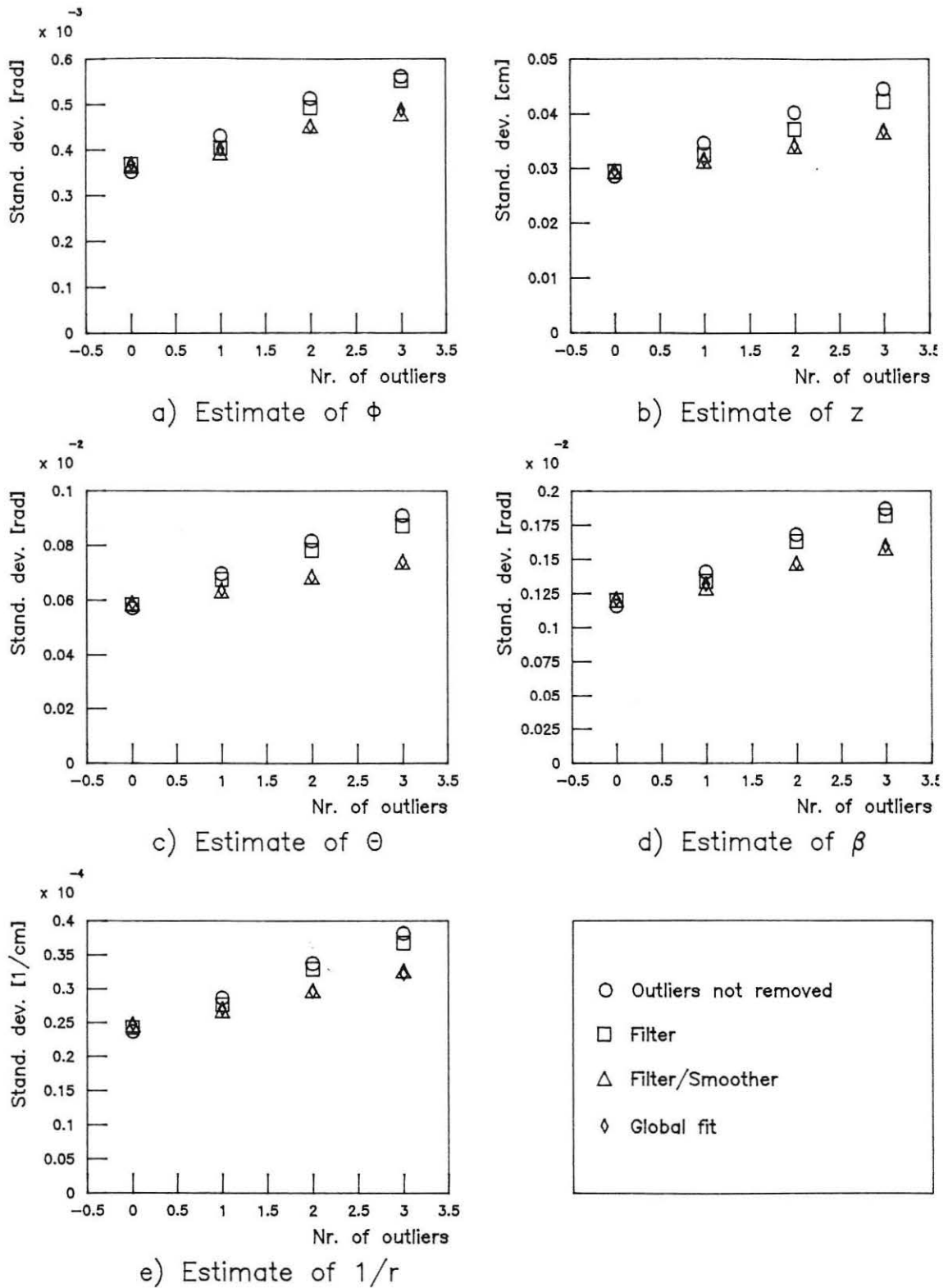


Fig. 5.17: Standard errors of estimated track parameters as a function of the number of outliers (Sample I).

### Influence of outliers on estimates of track parameters (strong m.s.)

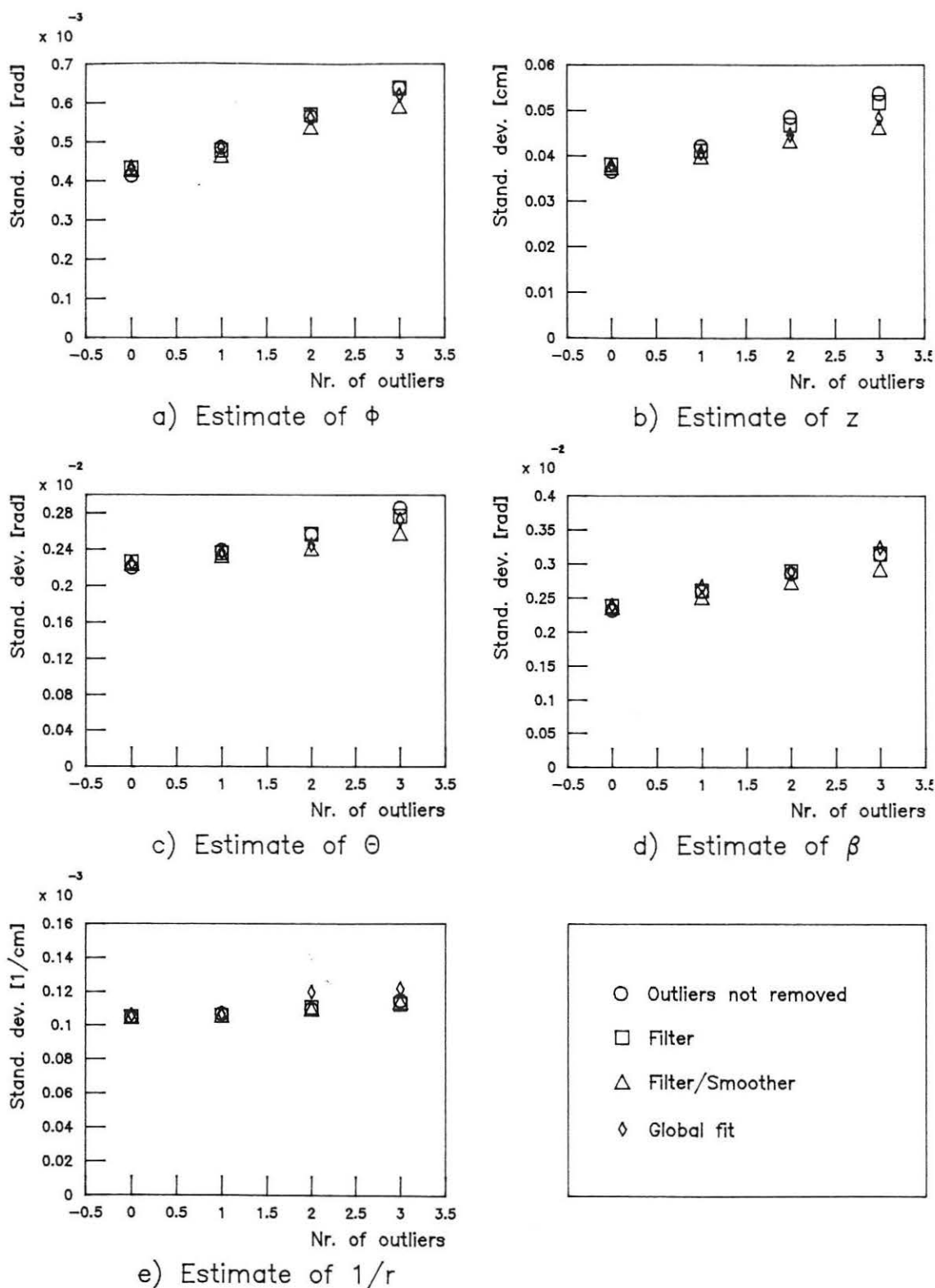


Fig. 5.18: Standard errors of estimated track parameters as a function of the number of outliers (Sample III).

### Mixture-model outliers, robust filter

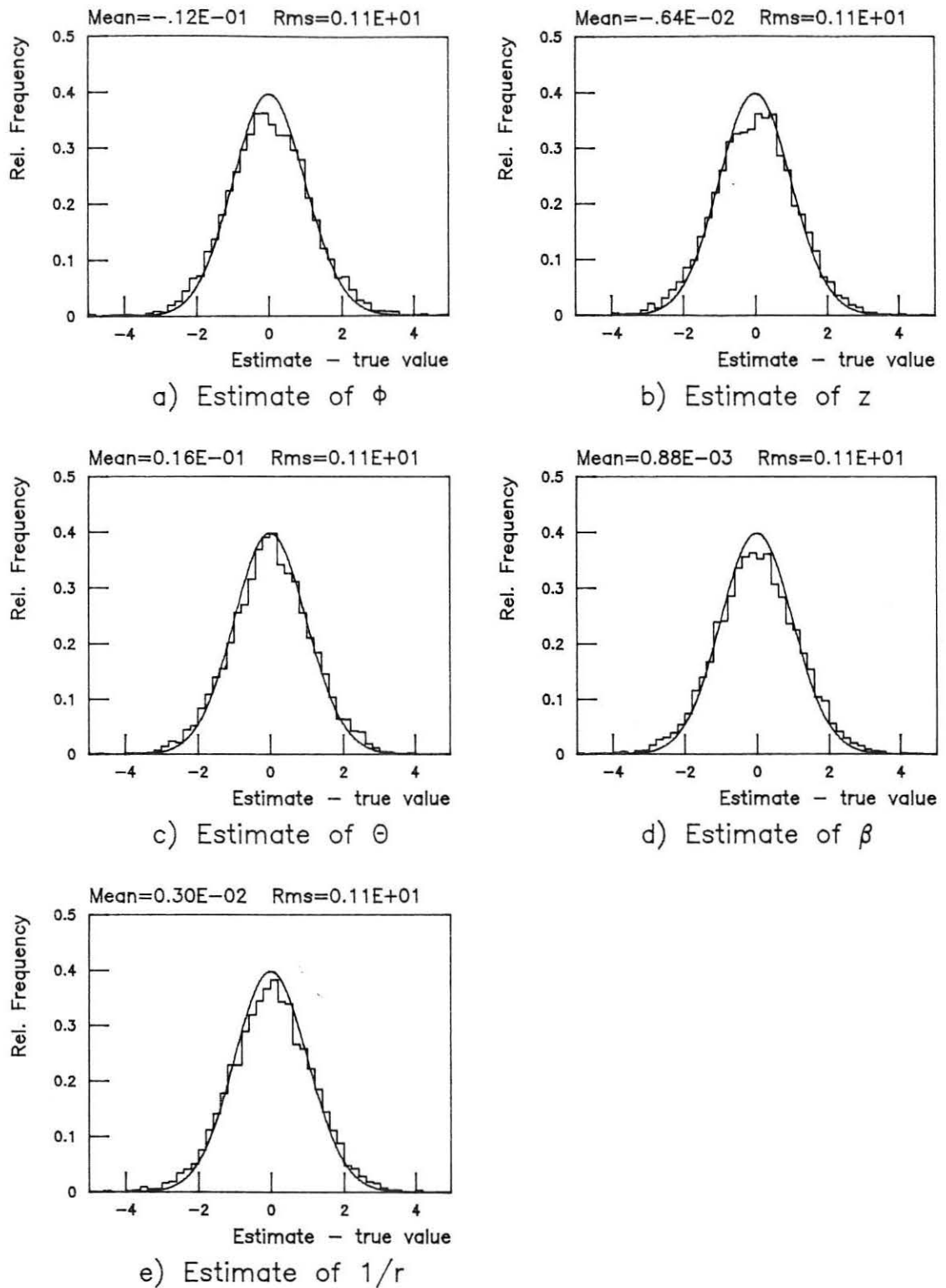


Fig. 5.19: Normalized residuals of track parameters estimated with the robust filter. Sample I with mixture-model outliers. The average number of outliers is 4. The superimposed curve is a standard normal p.d.f.



## Mixture-model outliers, robust filter

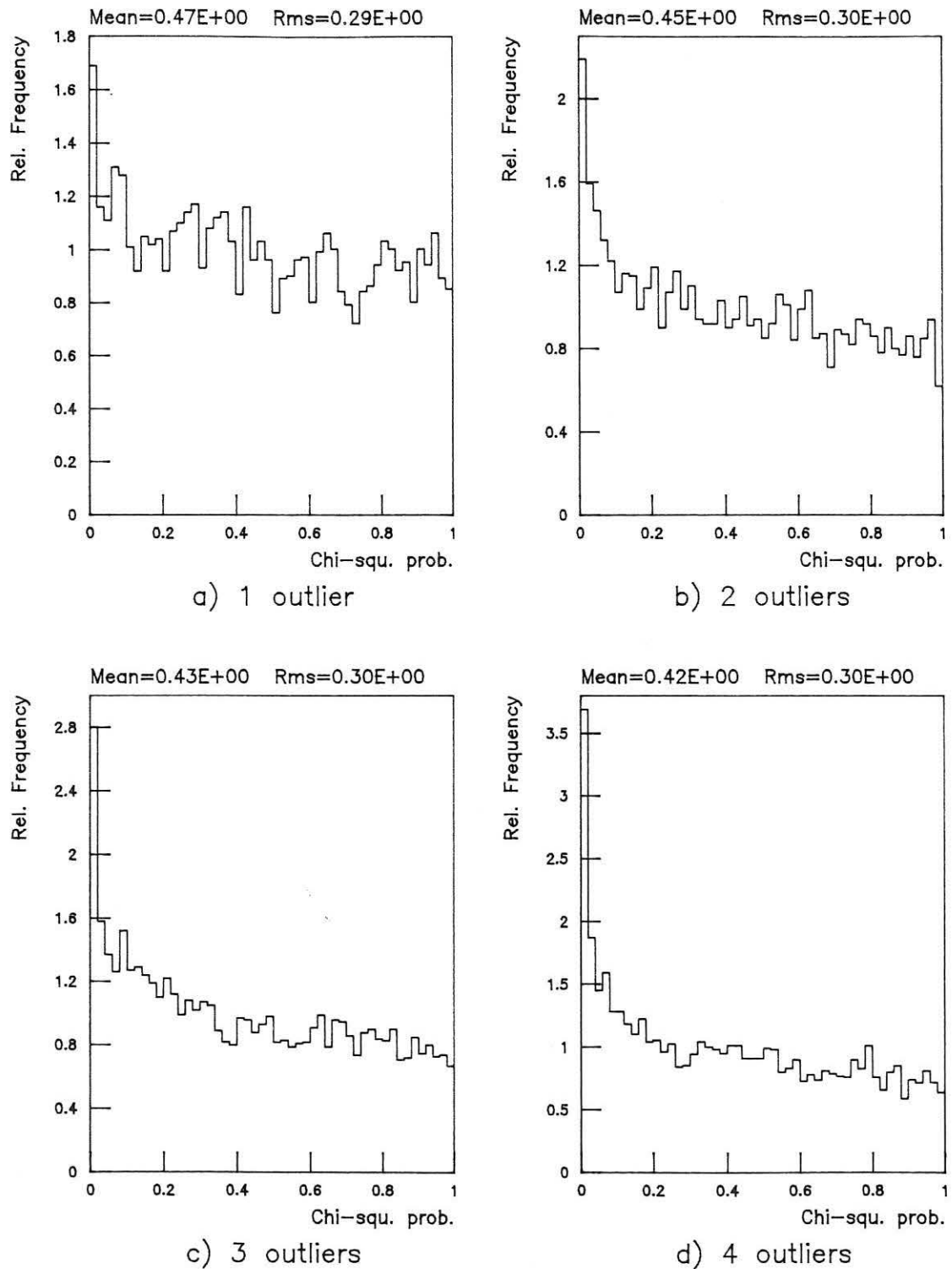


Fig. 5.20: Probability transform of  $\chi^2_E$  of robust filter (for the definition of  $\chi^2_E$  see text).

### Mixture-model outliers, robust filter

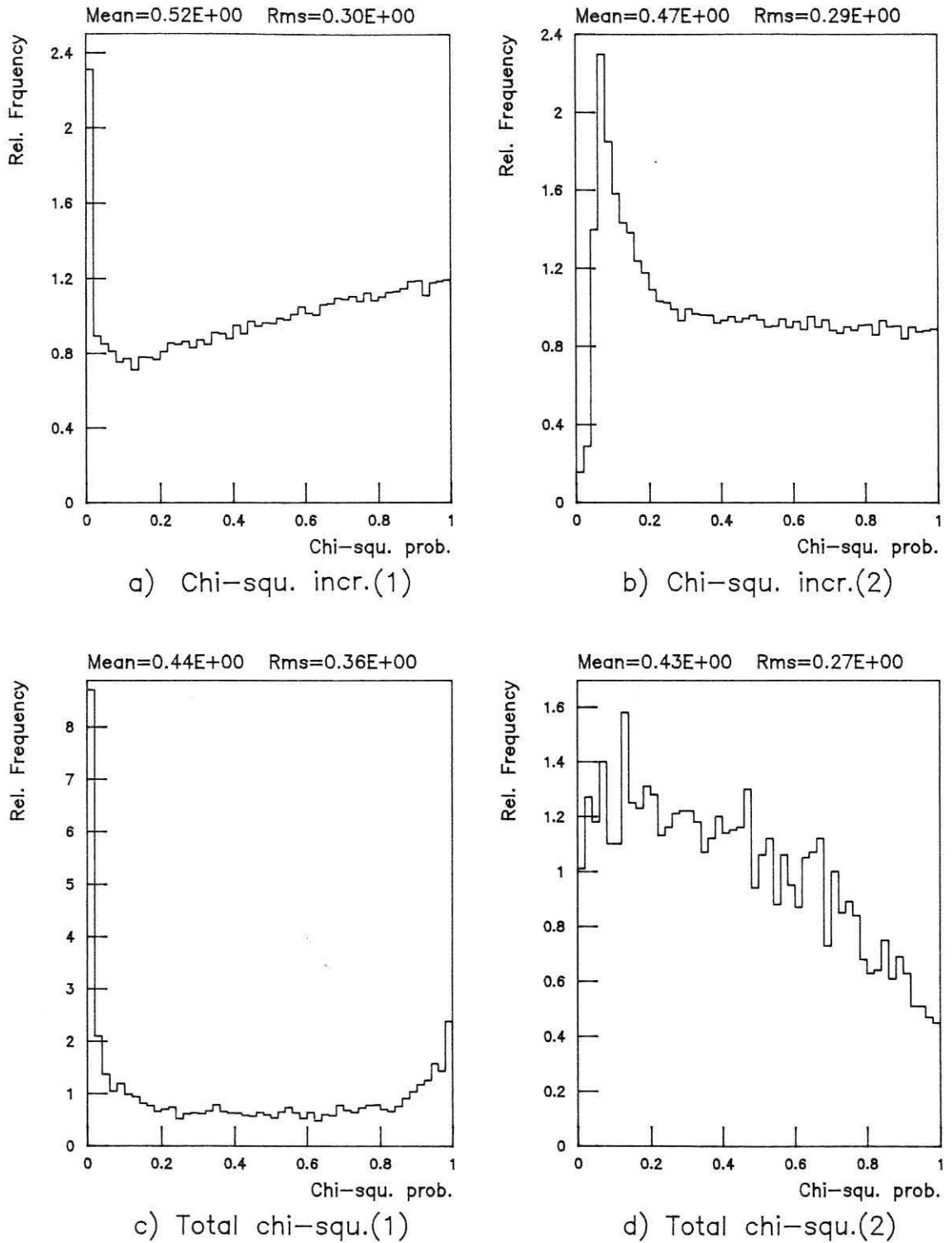


Fig. 5.21: Probability transforms of various generalized  $\chi^2$ -statistics of the robust filter (for a definition see text).

a)  $\chi^2_{k,F(1)}$  for all  $k$ ; b)  $\chi^2_{k,F(2)}$  for all  $k$ ; c) total  $\chi^2(1)$ ; d) total  $\chi^2(2)$ .

### Influence of mixture-model outliers on estimates of track parameters

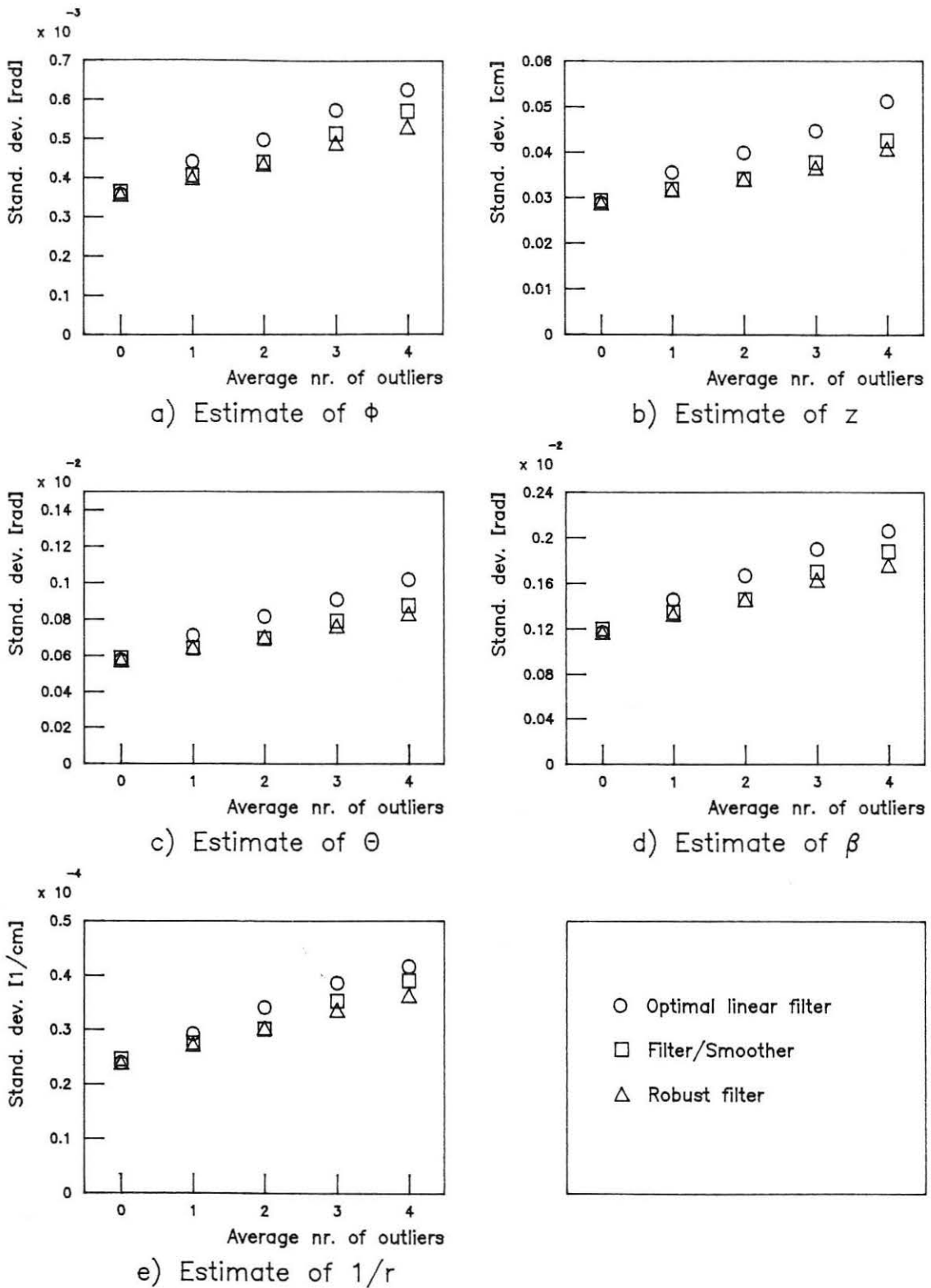


Fig. 5.22: Standard errors of estimated track parameters as a function of the average number of outliers. Sample I with mixture model outliers.

### Single variance-inflation outlier in FCA

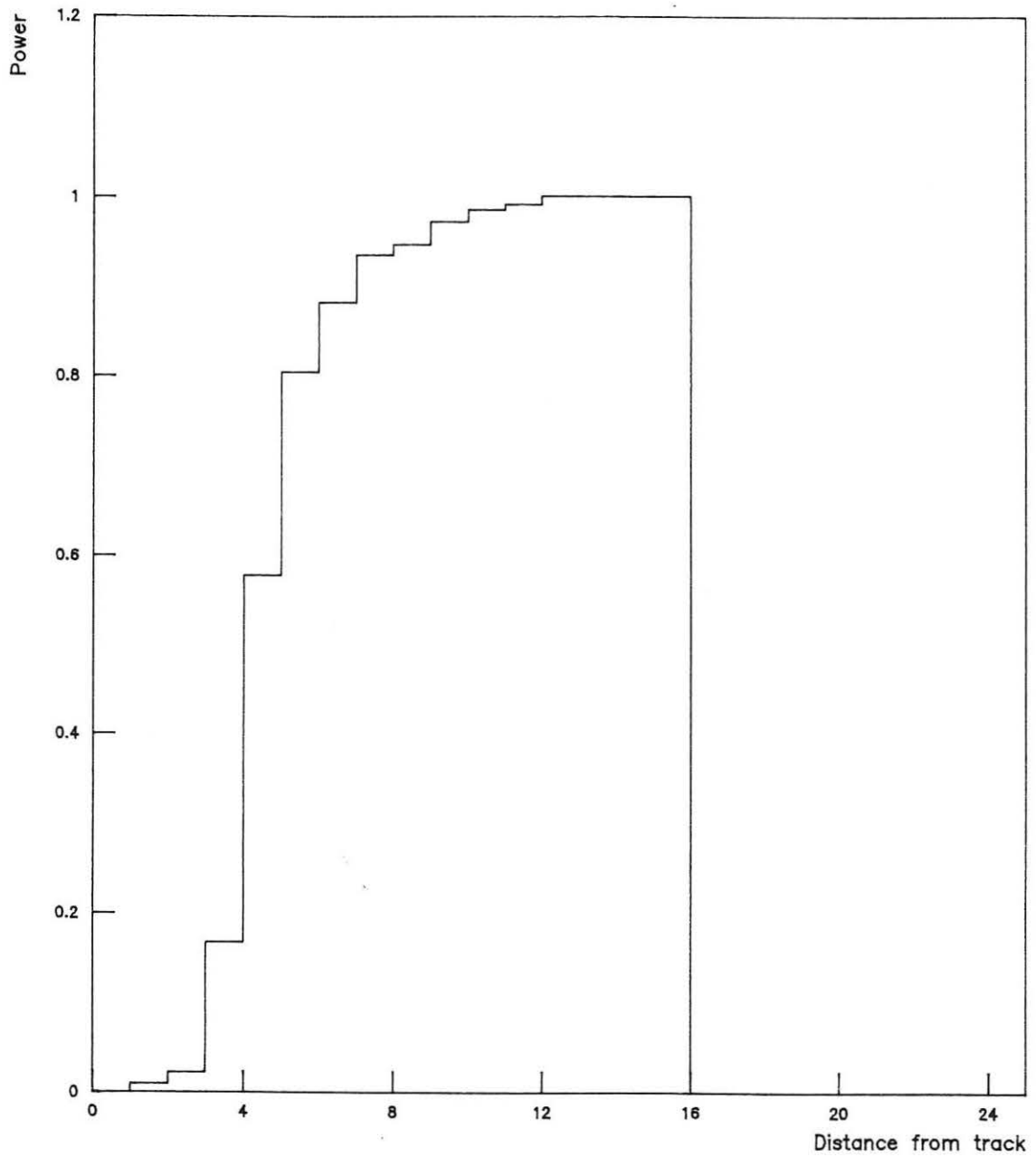
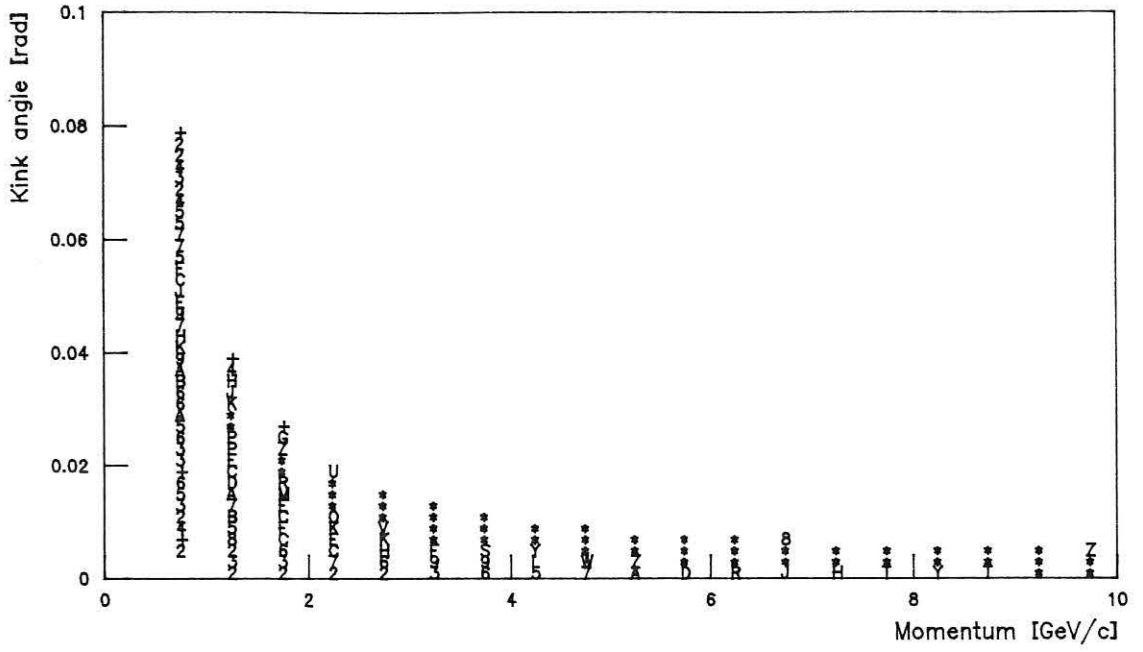
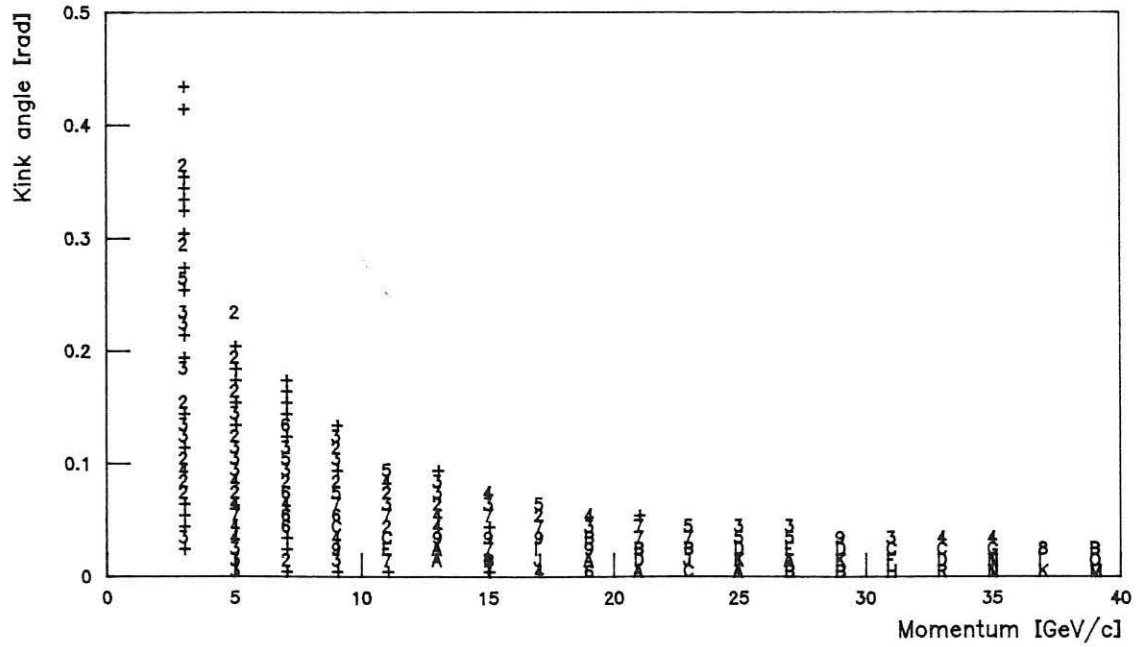


Fig. 5.23: Power of the CS-test in FCA as a function of the distance  $d$ .  
Tracks were simulated with FASTSIM, with a variance-inflation outlier in FCA.  
There are no outliers beyond  $d=16$ .

### Kink angle versus momentum of $\pi$ (K)



a)  $\pi$ -decays



b)  $K$ -decays

Fig. 5.24: Momentum of decaying meson versus simulated kink angle.

## Performance of kink finder in TPC

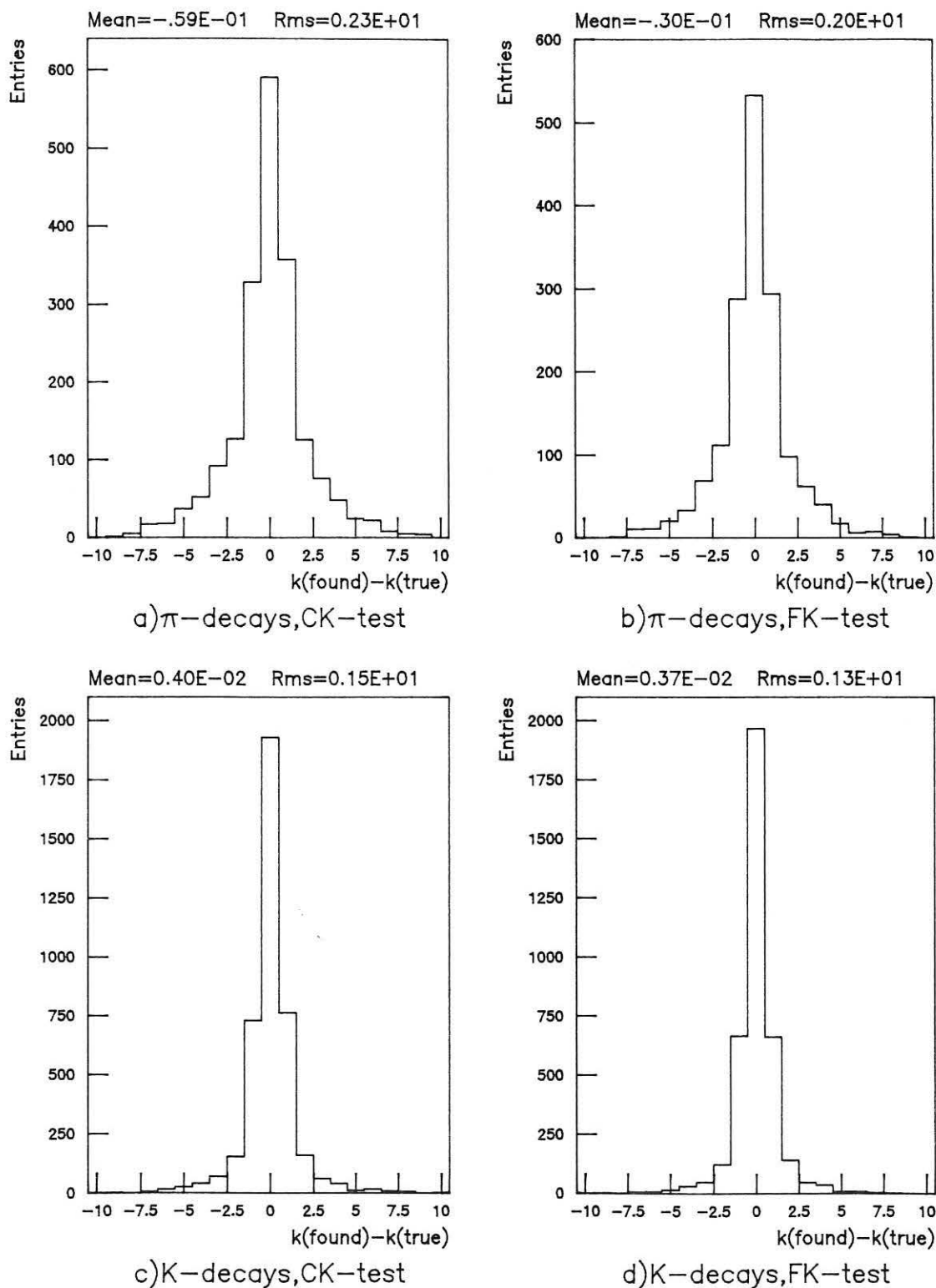


Fig. 5.25: Difference of the found position and the true position of the kink in STPC.

### Performance of kink finder in TPC ( $\pi$ -decays)

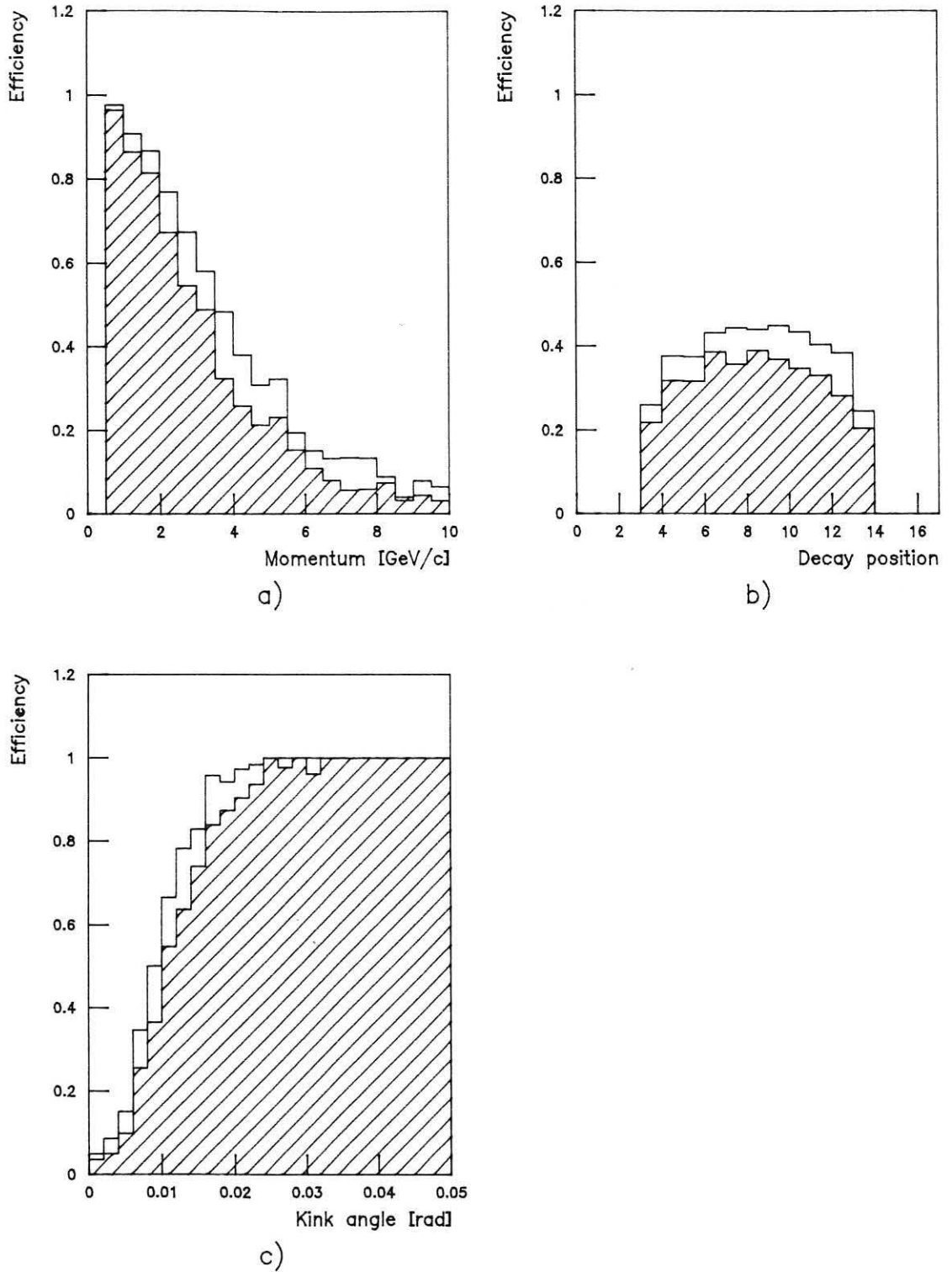


Fig. 5.26: Efficiency of CK-test (blank) and FK-test (shaded) in sample  $\Pi$  as a function of a) momentum; b) position of the decay; c) kink angle.

### Performance of kink finder in TPC (K-decays)

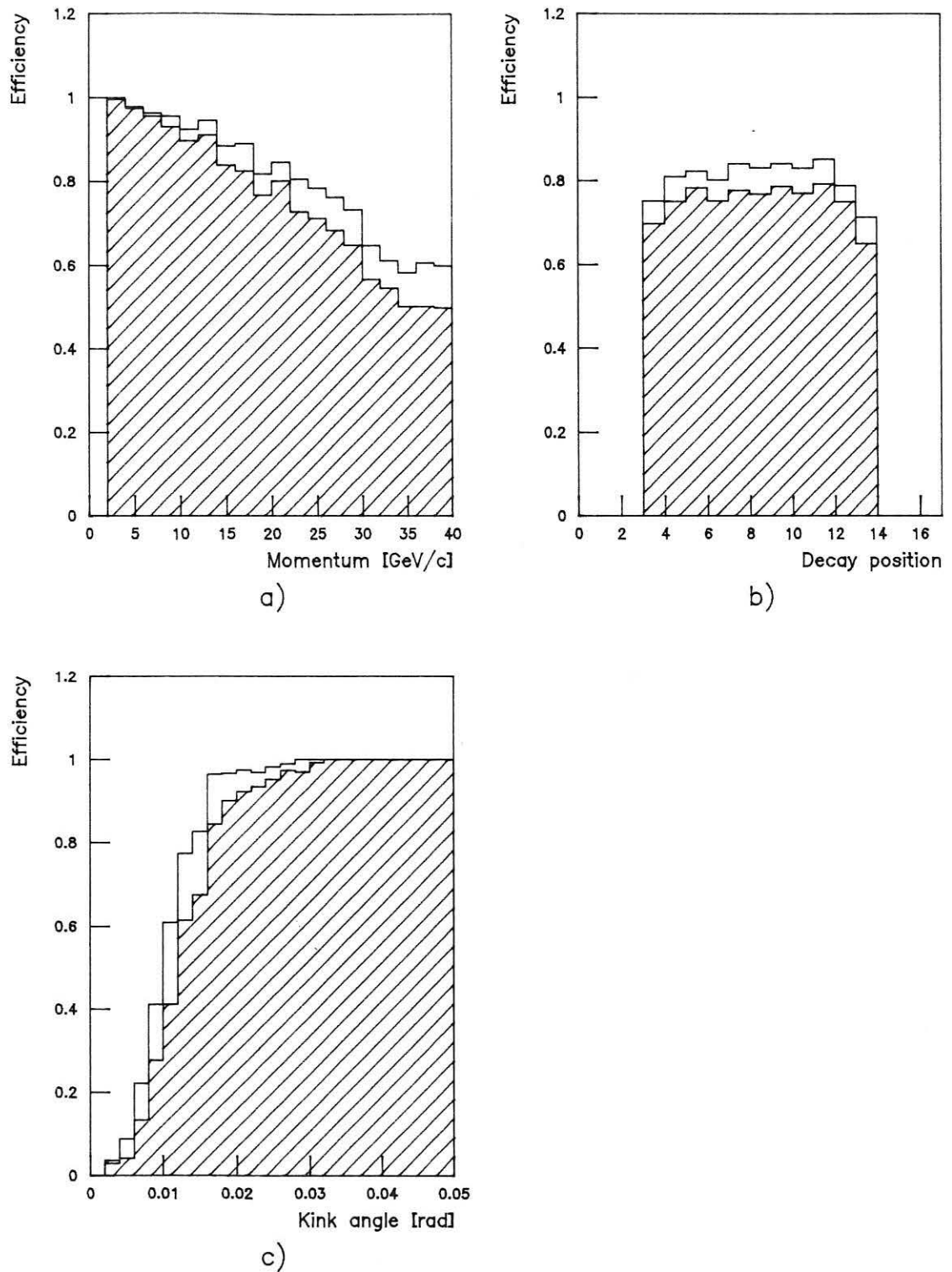


Fig. 5.27: Efficiency of CK-test (blank) and FK-test (shaded) in sample K as a function of a) momentum; b) position of the decay; c) kink angle.



## Performance of kink finder in TPC

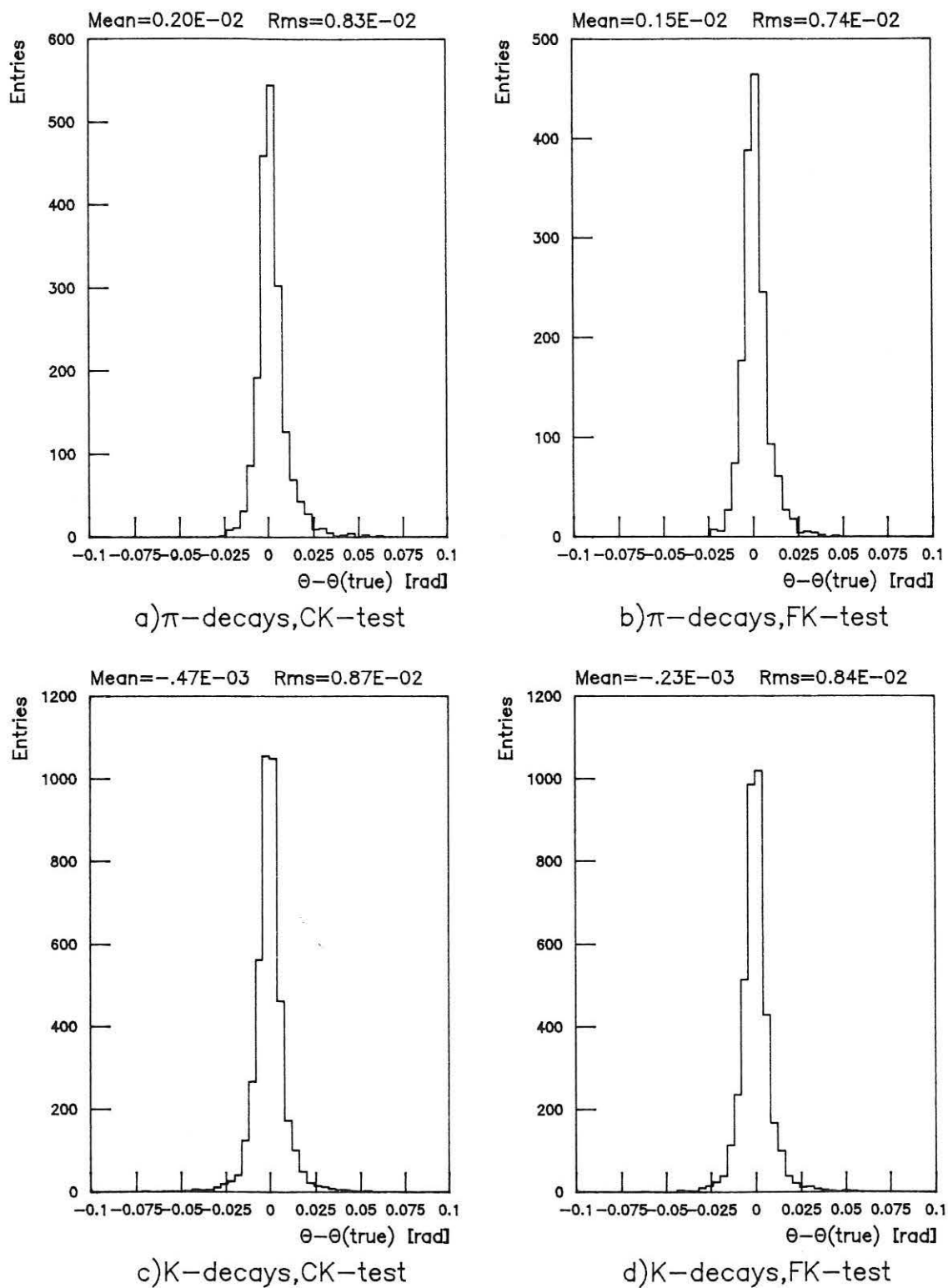


Fig. 5.28: Difference of the reconstructed and the true kink angle.

## Performance of kink finder in TPC

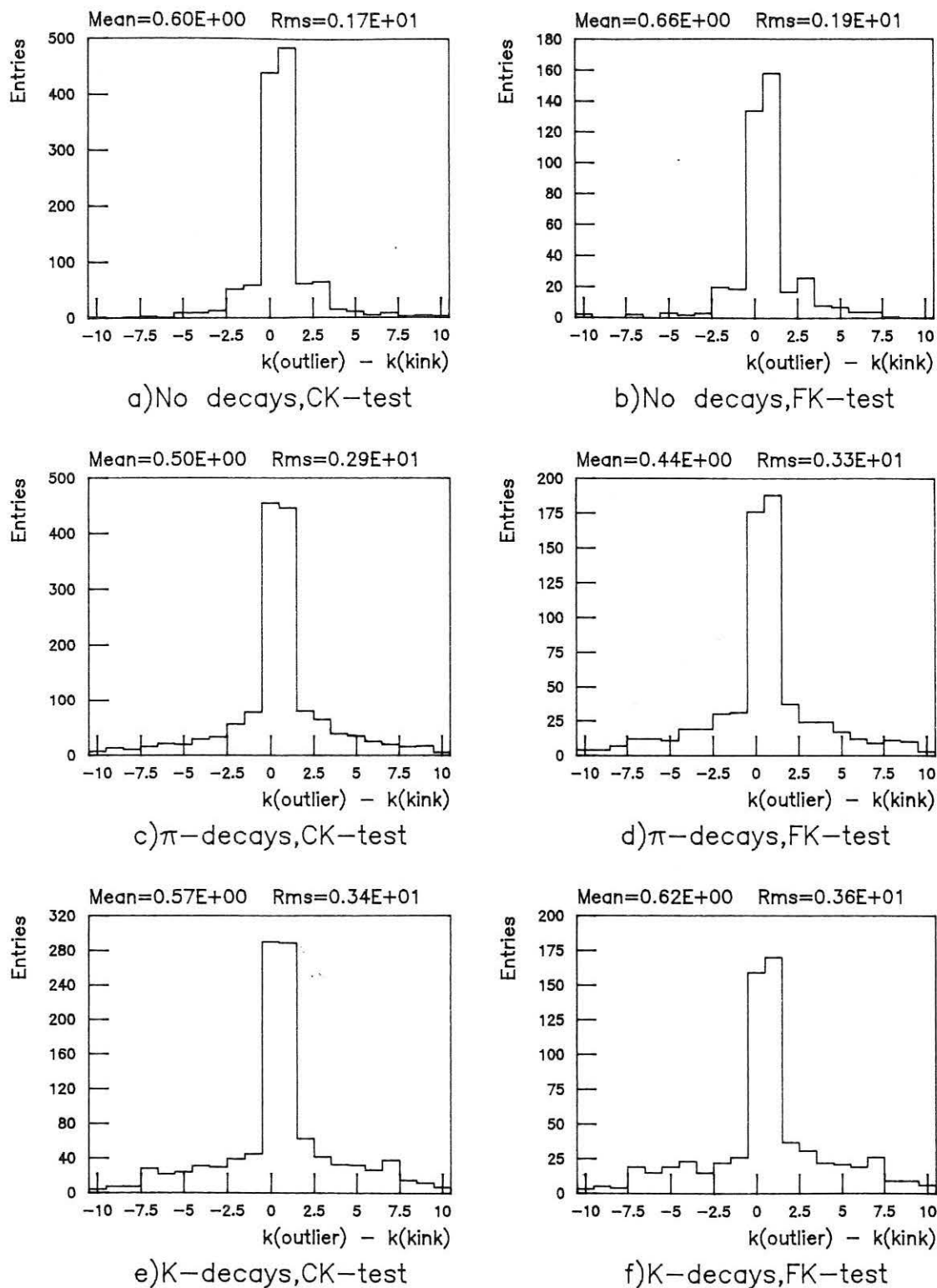


Fig. 5.29: Difference of the position of the outlier and the found position of the kink, for kinks found in the wrong position.

### Performance of kink finder with ID, TPC and OD

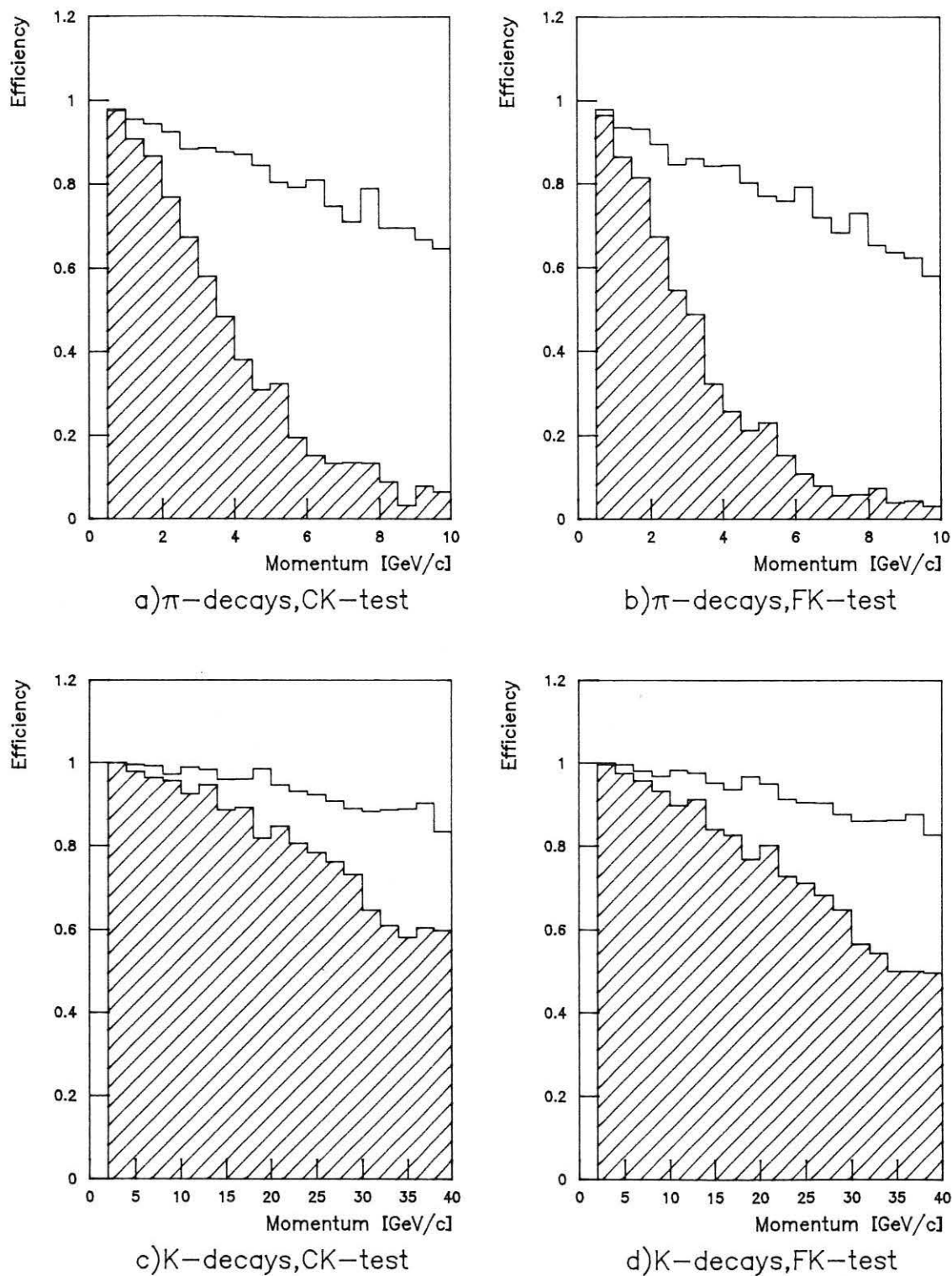


Fig. 5.30: Efficiency of kink finder with ID, TPC and OD (blank) and with TPC only (shaded), as a function of the momentum. The kink is inside the TPC.

### Performance of kink finder with ID, TPC and OD

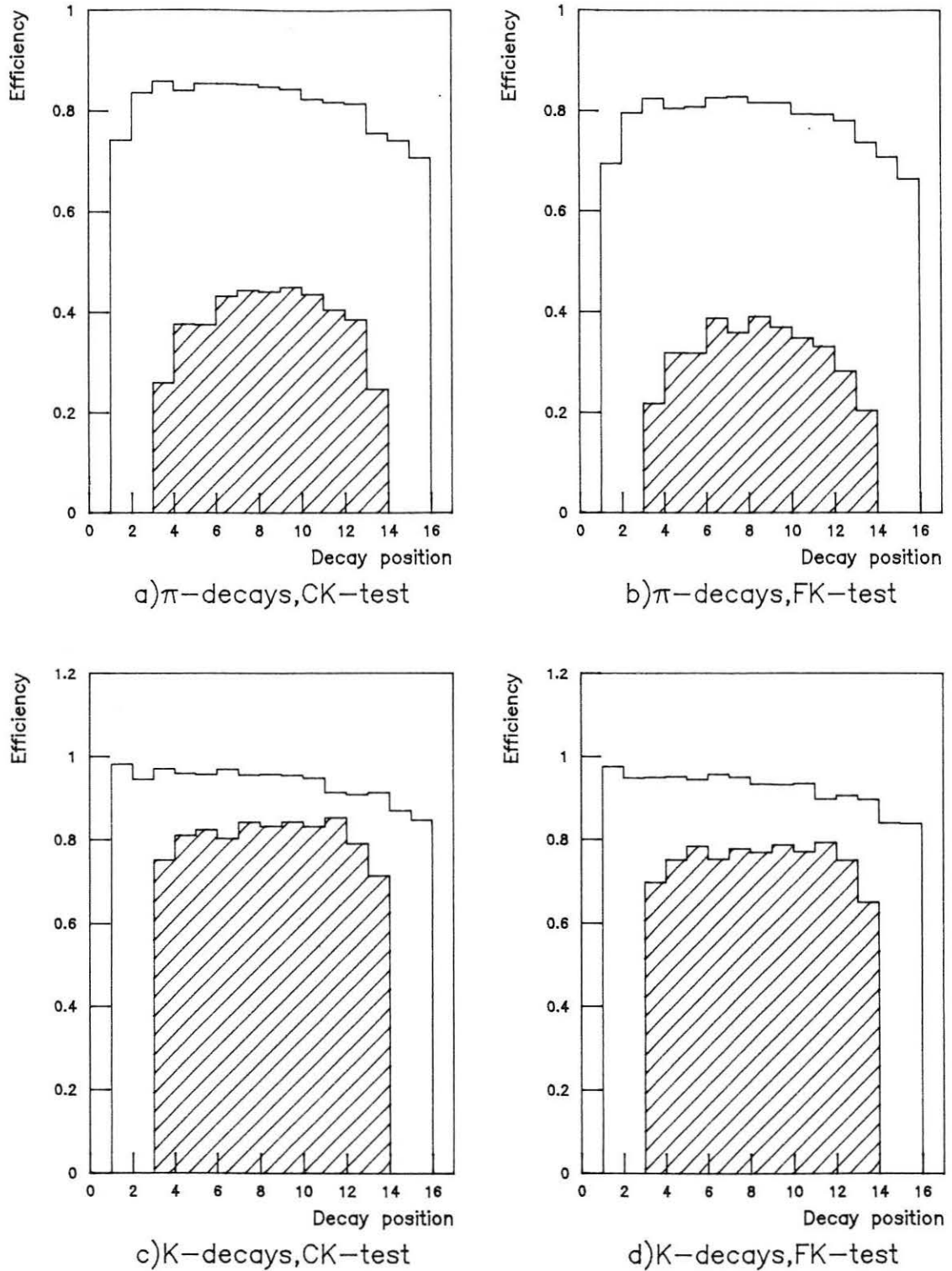


Fig. 5.31: Efficiency of kink finder with ID, TPC and OD (blank) and with TPC only (shaded), as a function of the decay position. The kink is inside the TPC.

Coordinate system used in secondary vertex study

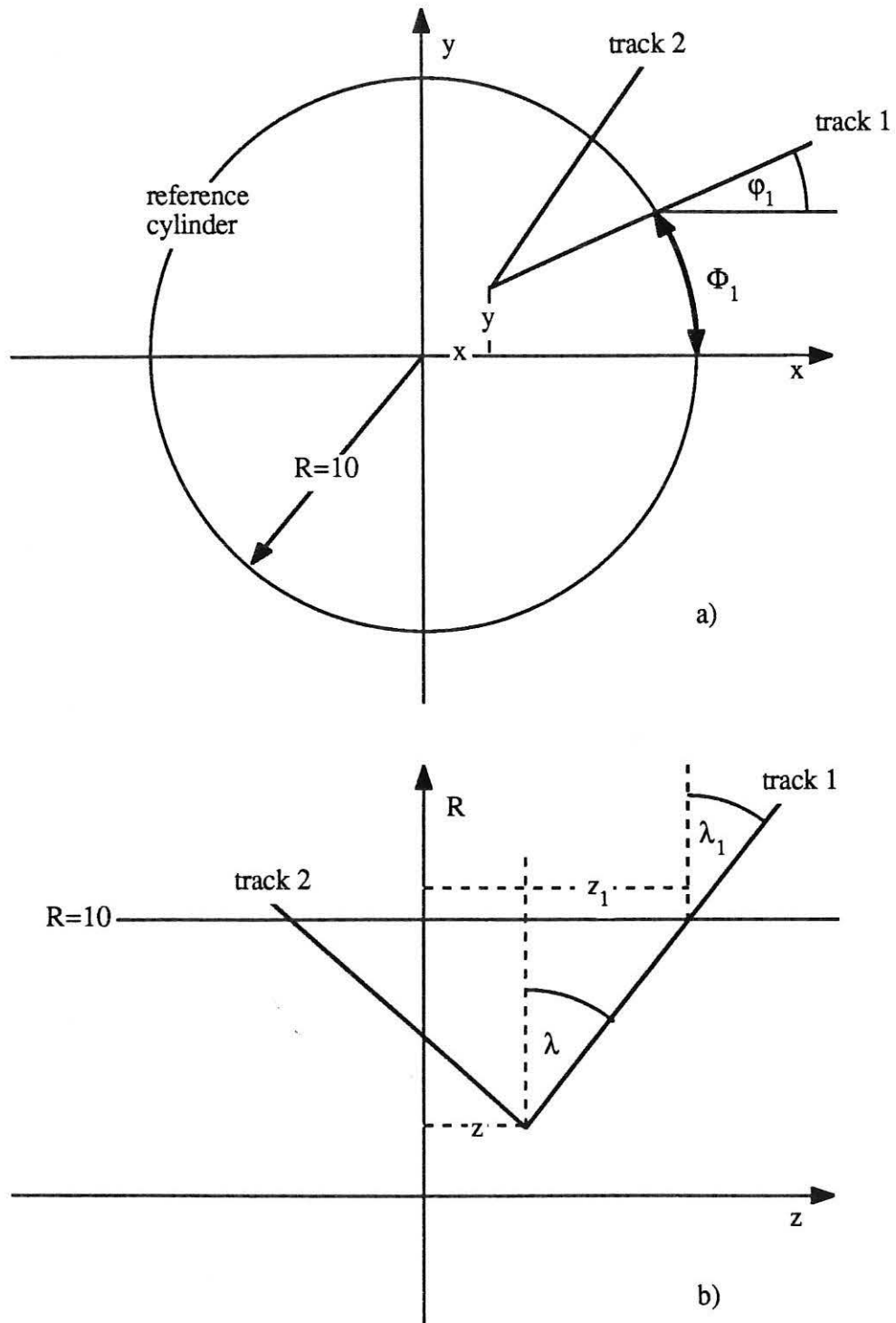
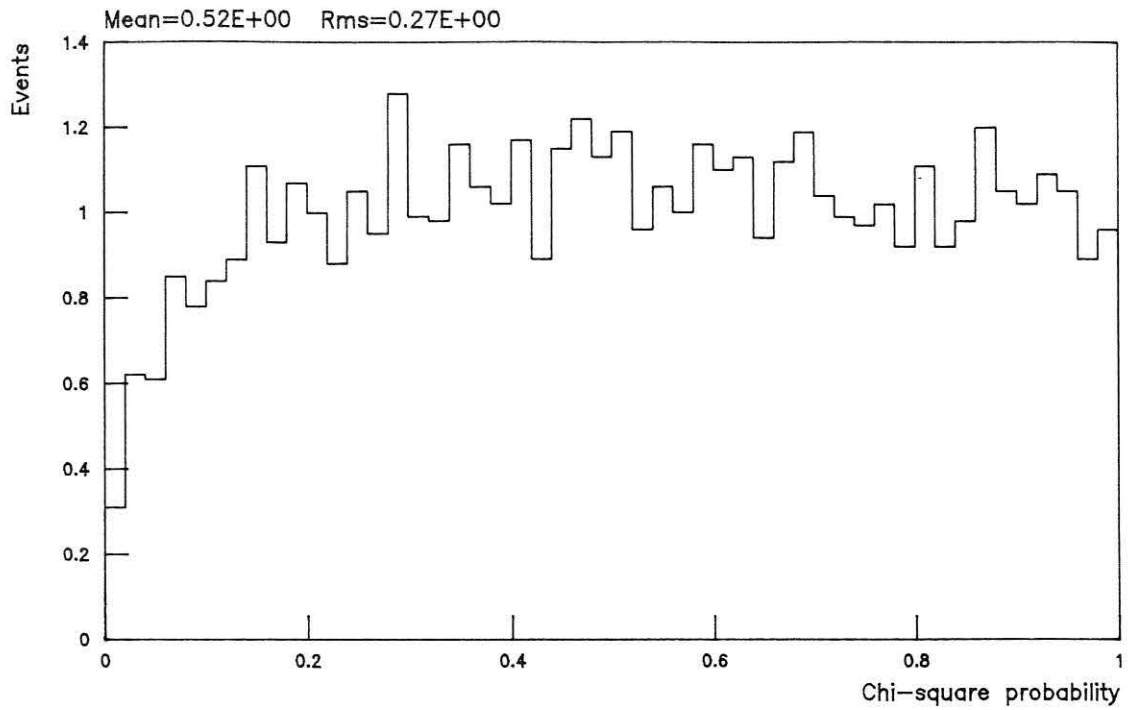
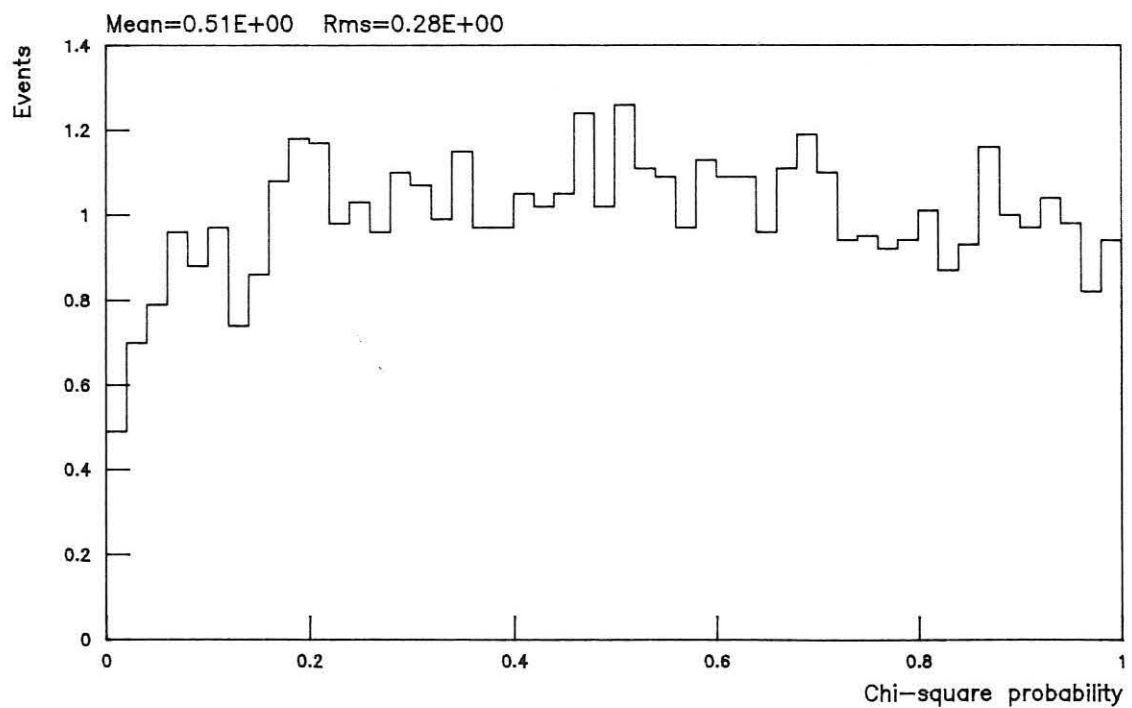


Fig. 5.32: Coordinate system used in the study of secondary vertex detection.  
 a) View in the x-y-plane; b) View in the z-R-plane.

### Total chi-square probabilities without secondary vertex



a) CS-test, outliers removed



b) CM-test

Fig. 5.33: Probability transforms of total chi-squares. No secondary vertex.

a) CS-algorithm, after removal of outlying tracks; b) CM-algorithm.

### Losses of primary tracks in secondary vertex detection

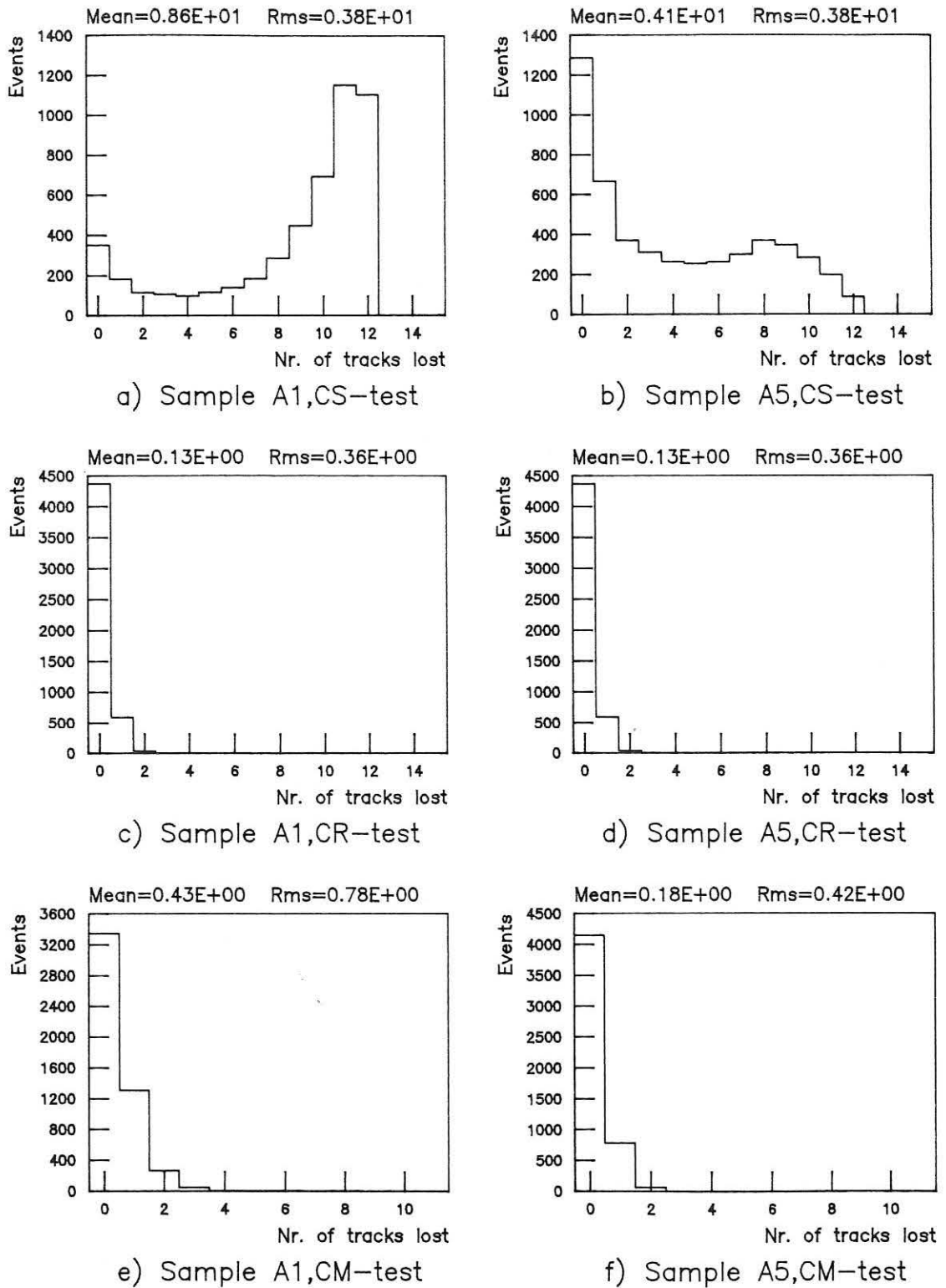
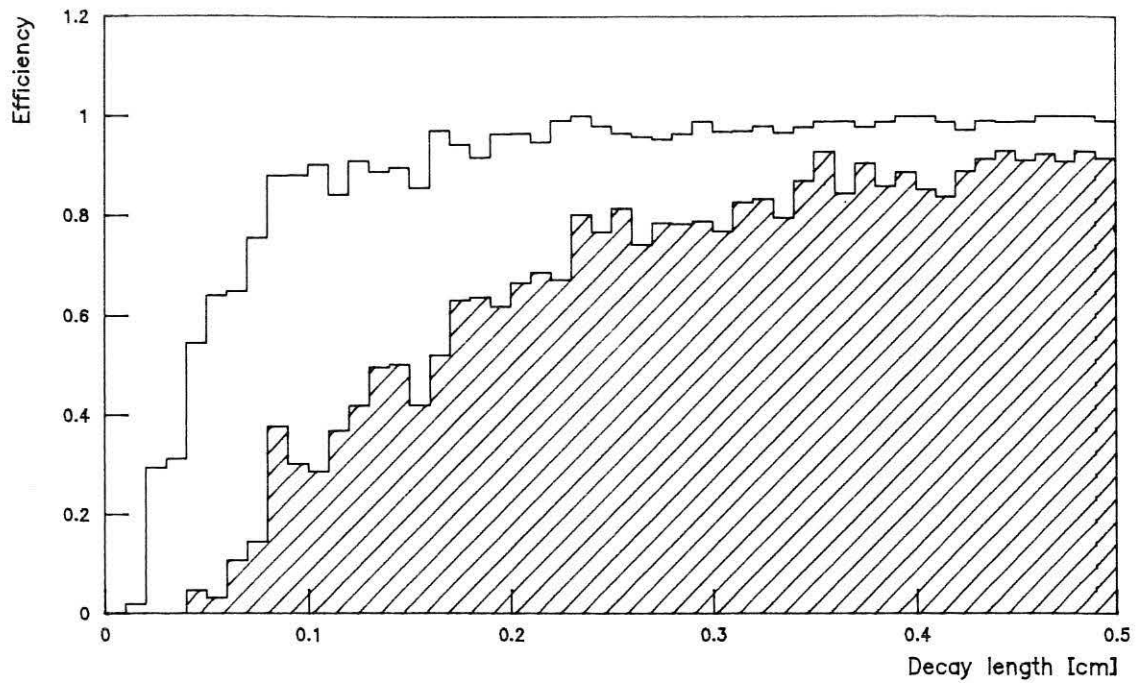


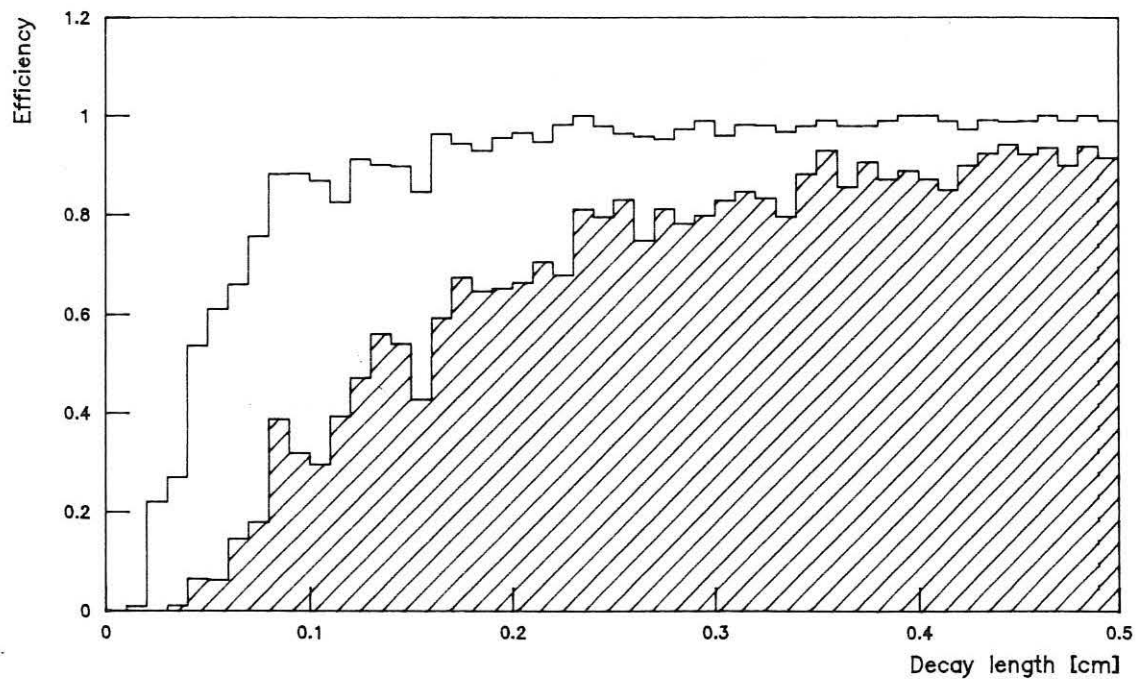
Fig. 5.34: Number of primary tracks lost in the detection of secondary vertices.

- a) Sample A1, CS-test; b) Sample A5, CS-test; c) Sample A1, CR-test;  
d) Sample A5, CR-test; e) Sample A1, CM-test; f) Sample A5, CM-test.

## Efficiency of secondary vertex detection



a) CR-test, samples A1 and A5



b) CM-test, samples A1 and A5

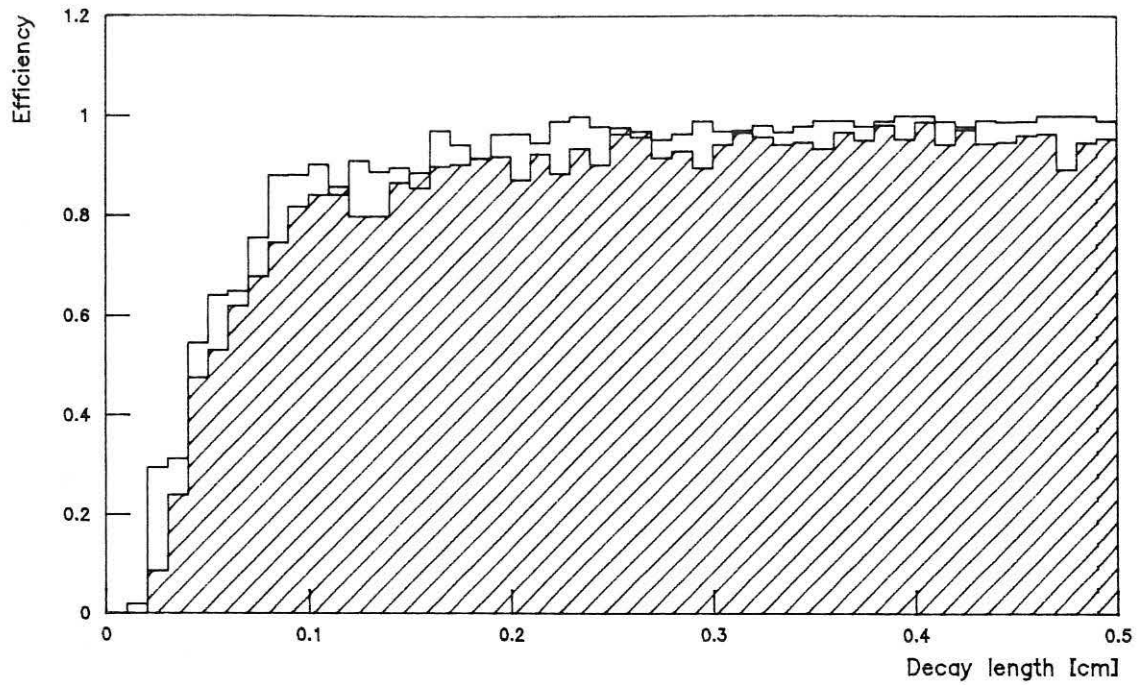
Fig. 5.35: Fraction of secondary vertices found as a function of the decay length.

a) CR-test, sample A1 (blank) and sample A5 (shaded);

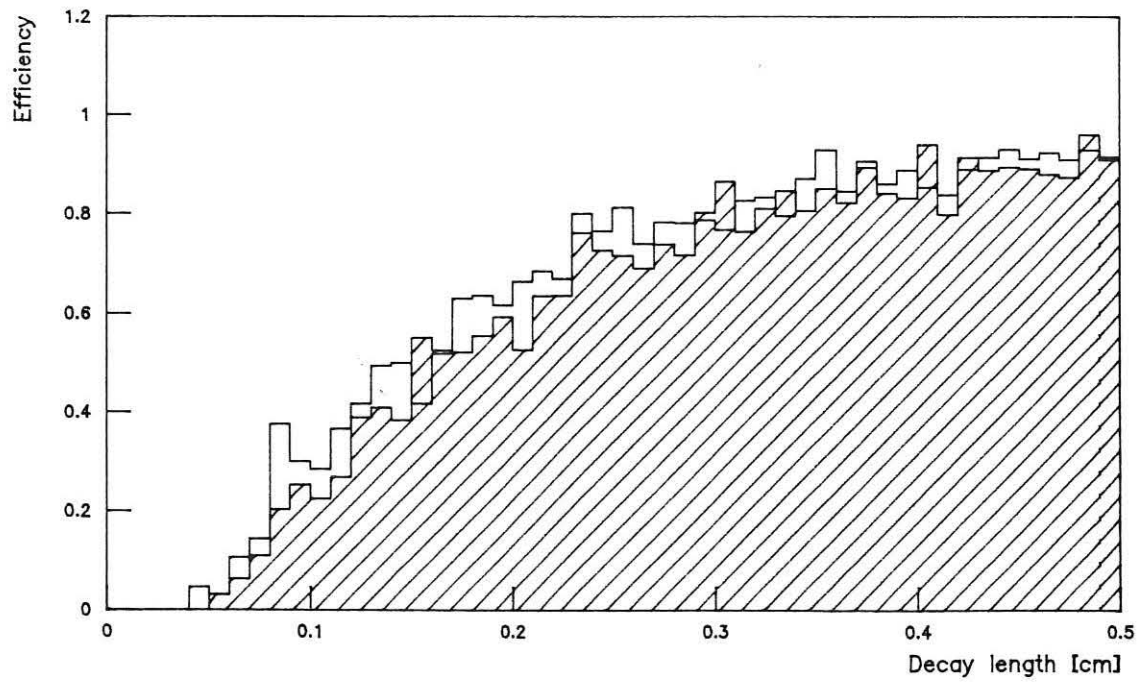
b) CM-test, sample A1 (blank) and sample A5 (shaded).



### Efficiency of secondary vertex detection



a) CR-test, samples A1 and B1



b) CR-test, samples A5 and B5

Fig. 5.36: Fraction of secondary vertices found as a function of the decay length.

a) CR-test, sample A1 (blank) and sample B1 (shaded);

b) CR-test, sample A5 (blank) and sample B5 (shaded).

## Efficiency of secondary vertex separation

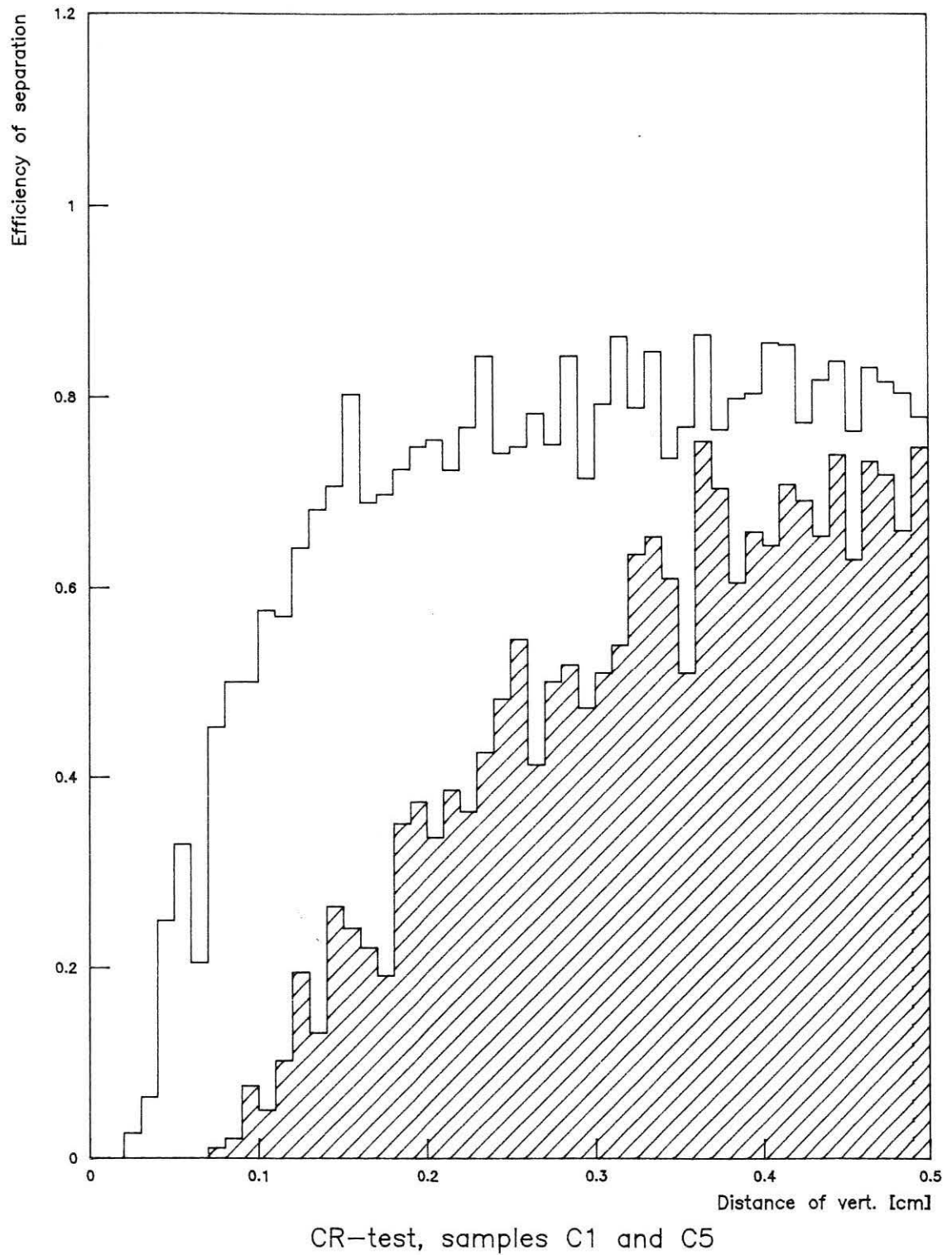


Fig. 5.37: Fraction of perfectly separated decay vertices as a function of their distance.  
CR-test, sample C1 (blank) and sample C5 (shaded).

# CHAPTER 6

## CONCLUSIONS

We have shown that filter methods meet virtually all requirements to track and vertex fitting which have arisen during the last years due to the rapid developments of accelerators and detectors.

- With the combined Kalman filter/smoothing algorithm it is possible to compute optimal estimates of the track parameters anywhere along a track. Hence optimal extrapolations from both ends of a track as well as optimal interpolations can be computed efficiently. This is important for the task of linking track segments in complex modular detectors.
- Outlying measurements can be found and removed with little additional computation.
- Kinks (1-prong decays) can be detected and located with good efficiency.
- The filter can be robustified in order to handle non-normal measurement errors or non-normal models of multiple scattering.
- The computing time of the track fit with the filter is a linear function of the number of measurements, both with and without multiple scattering.
- The application of the Kalman filter/smoothing to vertex fitting yields a procedure the computing time of which is proportional to the number of tracks involved. This allows precise estimation of vertex parameters also for high multiplicities.
- This procedure of vertex fitting is a very flexible tool which allows sequential testing for outlying (secondary) tracks and fast removal of found secondaries. It can also be robustified by the introduction of M-estimates.

We see two main research topics of further interest:

- First, the application of the robust filter to the reconstruction of electron tracks. This would require a mixture model of the process noise, i.e. a model of multiple Coulomb scattering, hard electromagnetic scattering, bremsstrahlung etc. which can be approximated sufficiently well by a mixture of two normals.
- Secondly, the application of robust estimators to vertex fitting. In particular, robust estimators with high break-down point could be a powerful tool for the separation of two (or more) vertices with low multiplicities, e.g. in  $b\bar{b}$ -decays.

## Acknowledgements

This thesis is based on work carried out in the Institute for High Energy Physics of the Austrian Academy of Sciences in Vienna. It is intended for application in the DELPHI experiment, which will start taking data in the middle of next year.

First I want to thank my colleague and friend Doz. Dr. M. Regler. It was he who introduced me to the field of data analysis in the context of experimental particle physics, and I owe a great deal to our long and fruitful collaboration. Without his constant support and warm encouragement this thesis would never have been written.

It is with great pleasure that I take the opportunity to express my thanks to Prof. R. Dutter of the Institute of Statistics and Probability Theory at the Technical University in Vienna. His interest in this work has yielded valuable suggestions concerning the application of robust algorithms to track and vertex fitting. I also appreciate many helpful discussions with Dr. S. Schnatter about Kalman filtering and Bayesian estimation.

My thanks are also due to many members of the DELPHI collaboration, in particular to Dr. P. Billoir (Paris) and Dr. P. Laurikainen (Helsinki). They have contributed many useful ideas. I also want to thank my colleagues in the Institute for High Energy Physics in Vienna. I am greatly indebted to Dr. T. Kreuzberger and P.K. Lichtenwagner for the implementation of the Kalman filter/smoother algorithm in the DELPHI data analysis program. N. Neumeister has written the program with which most of the figures were produced. Finally, I want to express my thanks to the Director of the Institute for High Energy Physics, Prof. W. Majerotto, and to my group leader, Dr. W. Mitaroff, for granting me the opportunity of completing this work, and for providing the pleasant working conditions in the Institute.

## Figure captions

- Fig. 2.1 .....20  
Elastic scattering of two particles. Apart from a change in direction, the particles emerge essentially unchanged from the interaction.
- Fig. 2.2 .....20  
Picture of a  $p\bar{p}$  collision in the central drift chamber of the UA1 experiment. The energy in the center-of-mass is 540 GeV. A large part of the energy flows into the creation of new particles. The arrows point to the decay products of a  $Z^0$  vector boson.
- Fig. 2.3 .....21  
A fixed target spectrometer: the experiment WA6.  
a) B3-6 and F1-8 are MWPCs; FV, FH, BV and BH are scintillator hodoscopes; S1 and S2 are scintillator counters.  
b) Sketch of the forward arm with scale.  
From: G. Fidecaro et al., Phys. Lett. 105B (1981) 309
- Fig. 2.4 .....21  
A typical collider experiment. No particular direction is privileged by kinematics. The detector is as hermetic as possible.
- Fig. 2.5 .....22  
Artist's view of the DELPHI detector.
- Fig. 2.6 .....23  
Schematic view of a multi-wire proportional chamber (MWPC).  
a) Cross section of a chamber with the two cathode planes and the array of anode wires (sense wires).  
b) Field lines and equipotential lines. The sense wire is in the centre.  
c) The electrons which have been set free by the charged particle crossing the chamber are attracted to the sense wire. In the region of high field gradient around the wire an avalanche develops, which leads eventually to a detectable signal.
- Fig. 2.7 .....23  
Distribution of the measurement error of a stack of four MWPCs, from simulated data. The four wire directions are  $0^\circ$ ,  $90^\circ$ ,  $45^\circ$  and  $135^\circ$ . The x- and the y-coordinate are estimated from the four wire positions under the assumption that the track is orthogonal to the chamber. The superimposed curve is a standard normal p.d.f.

Fig. 2.8 .....	24
Schematic view of a drift chamber. The time between the passage of the particle and the arrival of the electrons at the sense wire is (roughly) proportional to the distance between the track and the wire.	
Fig. 2.9 .....	24
The cylindrical drift chamber of the JADE experiment.	
a) Total view, cross section perpendicular to the beam;	
b) One drift cell.	
From: W. Farr et al., Nucl. Instr. and Meth. 156 (1978) 283	
Fig. 2.10.....	25
The distribution of the measurement error of a drift chamber. The central part of the distribution has a nicely Gaussian shape. The tail on the left hand side is due to long-range delta-rays.	
From: F. Sauli, Nucl. Instr. and Meth. 156 (1978) 147	
Fig. 2.11.....	25
Schematic view of the ALEPH time projection chamber (TPC).	
a) The whole chamber;	
b) One of the end plates.	
From: S.R. Amendolia et al., Nucl. Instr. and Meth. A252 (1986) 399	
Fig. 2.12.....	26
The DELPHI detector. Cross section in the z-y-plane, parallel to the beam.	
Fig. 2.13.....	27
The DELPHI detector. Cross section in the x-y-plane, perpendicular to the beam.	
Fig. 3.1 .....	42
Effect of multiple scattering on a track.	
a) In a thick scatterer the track is deflected and emerges with an offset.	
b) In a thin scatterer the offset is negligible, but the deflection is not.	
Fig. 3.2 .....	42
Cylindrical detectors in a homogeneous magnetic field. The track is an exact helix.	
Fig. 3.3 .....	43
Experiment WA6: Probability distribution of the $\chi^2$ of the kinematic fit.	
a) The inelastic events accumulate in the high peak at low probabilities. They can be removed by a $\chi^2$ -cut at 6%.	
b) The same distribution after a cut at 6% shows the expected flat shape.	
From: G. Fidecaro et al., Nucl. Phys. B173 (1980) 513	

Fig. 4.1 .....	76
Merging of two track segments by successive filtering and smoothing.	
Fig. 4.2 .....	76
The decay of a charged particle. Because of charge conservation, the number of charged decay products is always odd.	
Fig. 4.3 .....	76
A 1-prong decay (kink). The two charged tracks can be misinterpreted as a single one, if the angle between them is small.	
Fig. 4.4 .....	77
Lorentz transformation of the muon momentum in K- and $\pi$ -decays. The circle in the rest system of the decaying particle is transformed into an ellipse in the observer system. The kink angle $\theta'$ and the momentum $\mathbf{p}'$ in the observer system can be read directly from the ellipse.	
a) $K \rightarrow \mu \bar{\nu}$ , $P = 2 \text{ GeV}/c$ ;	
b) $\pi \rightarrow \mu \bar{\nu}$ , $P = 0.5 \text{ GeV}/c$ .	
Fig. 4.5 .....	78
Kink angle (a,c) and momentum of muon (b,d) from K-decay as a function of $\cos\theta$ of the muon in the rest system of the K.	
a,b) $P = 2 \text{ GeV}/c$ ;	
c,d) $P = 40 \text{ GeV}/c$ .	
Fig. 4.6 .....	79
Kink angle (a,c) and momentum of muon (b,d) from $\pi$ -decay as a function of $\cos\theta$ of the muon in the rest system of the $\pi$ .	
a,b) $P = 0.5 \text{ GeV}/c$ ;	
c,d) $P = 10 \text{ GeV}/c$ .	
Fig. 4.7 .....	80
A kink as a sudden change of the state vector. The kink can be found by comparing the state vector $\tilde{\mathbf{x}}_k$ of the forward filter with the back extrapolation $\tilde{\mathbf{x}}^{k+1}_k(b)$ of the backward filter. If they are significantly different, there is reason to assume a kink.	
Fig. 4.8 .....	80
Measurements and estimated parameters of the vertex fit. The 5-dimensional vectors of track parameters estimated in the track fit serve as (virtual) measurements of the vertex fit. The parameters to be estimated are the vertex position $\mathbf{v}$ and the 3-momentum vectors $\mathbf{q}_i$ of all tracks.	

Fig. 5.1 .....	105
Execution times of global fit, filter and smoother ( $\mu$ VAX II).	
a) Helix tracks, 2 measurements/step, no multiple scattering;	
b) Helix tracks, 4 measurements/step, no multiple scattering;	
c) Helix tracks, 2 measurements/step, multiple scattering;	
d) Helix tracks, 4 measurements/step, multiple scattering.	
Fig. 5.2 .....	106
Cross section of the simplified TPC (STPC). In each of the 16 circular pad rows $R\Phi$ is measured with a standard deviation of 0.2 mm. The z-coordinate is determined from the drift time with a standard deviation of 0.6 mm.	
Fig. 5.3 .....	107
Comparison of three estimates in STPC (5000 tracks, no multiple scattering).	
a–c) Difference of estimated $1/r$ and true $1/r$ at $R=36.5\text{cm}$ , divided by the standard deviation of $1/r$ :	
a) global fit; b) filter; c) smoother.	
The superimposed curve is a standard normal p.d.f.	
d–f) Pairwise differences of the three estimates of $1/r$ , divided by the standard deviation of $1/r$ :	
d) filter – global fit; e) smoother – global fit; f) smoother – filter.	
Fig. 5.4 .....	108
Comparison of three estimates in STPC (5000 tracks, no multiple scattering).	
a–c) Probability transforms of the total chi-squares:	
a) global fit; b) filter; c) smoother.	
d–f) Pairwise differences of total chi-squares:	
d) filter – global fit; e) smoother – global fit; f) smoother – filter.	
Fig. 5.5 .....	109
Same as Fig. 5.3, but with multiple scattering in STPC.	
Fig. 5.6 .....	110
Same as Fig. 5.4, but with multiple scattering in STPC.	
Fig. 5.7 .....	111
The detector configuration in the barrel region of DELPHI, as used by FASTSIM. View in the x-y-plane. The characteristics of detectors and scatterers are summarized in Table 5.1.	



Fig. 5.8 .....	112
Standard errors of the smoothed track parameters in OD and of the extrapolation of the filtered track parameters in ID to OD, as a function of the momentum $p$ : a) $R\Phi$ ; b) $z$ ; c) $\theta$ ; d) $\phi$ .	
Fig. 5.9 .....	113
The detector configuration in the forward region of DELPHI as used by FASTSIM. View in the $y$ - $z$ -plane.	
Fig. 5.10 .....	114
Standard errors of the smoothed track parameters in FCB and of the extrapolation of the filtered track parameters in ID to FCB, as a function of the momentum $p$ : a) $x$ ; b) $y$ ; c) $\theta$ ; d) $\phi$ .	
Fig. 5.11 .....	115
Power of outlier tests (Sample Ia). a) Power of the CS-test (blank) and of the CF-test (shaded) as a function of the distance $d$ of the outlier from the true track position. b) Power of the CS-test (blank) and of the CF-test (shaded) as a function of the position $k$ of the outlier in the track.	
Fig. 5.12 .....	116
Probability transform of total chi-square (Sample Ia). a) Outliers not removed; b) Outliers removed with CF-test; c) Outliers removed with CS-test.	
Fig. 5.13 .....	117
Power of outlier tests (Sample IIa). a) Power of the CS-test (blank), of the CG-test (shaded) and of the CF-test (doubly shaded) as a function of the distance $d$ of the outlier from the true track position. b) Power of the CS-test (blank), of the CG-test (shaded) and of the CF-test (doubly shaded) as a function of the position $k$ of the outlier in the track.	
Fig. 5.14 .....	118
Power of outlier tests (Sample IIIa). a) Power of the CS-test (blank), of the CG-test (shaded) and of the CF-test (doubly shaded) as a function of the distance $d$ of the outlier from the true track position. b) Power of the CS-test (blank), of the CG-test (shaded) and of the CF-test (doubly shaded) as a function of the position $k$ of the outlier in the track.	

Fig. 5.15.....	119
Probability transform of the total chi-square of the smoother (blank) and of the global fit (shaded), after removal of outliers (Sample I).	
a) no outliers; b) 1 outlier; c) 2 outliers; d) 3 outliers	
Fig. 5.16.....	120
Probability transform of the total chi-square of the smoother (blank) and of the global fit (shaded), after removal of outliers (Sample III).	
a) no outliers; b) 1 outlier; c) 2 outliers; d) 3 outliers	
Fig. 5.17.....	121
Standard errors of estimated track parameters as a function of the number of outliers (Sample I).	
a) $\Phi$ ; b) $z$ ; c) $\theta$ ; d) $\beta = \varphi - \Phi$ ; e) $1/r$	
○ Outliers not removed	
□ Outliers removed by CF-test	
△ Outliers removed by CS-test	
◇ Outliers removed by CG-test	
Fig. 5.18.....	122
Standard errors of estimated track parameters as a function of the number of outliers (Sample III).	
a) $\Phi$ ; b) $z$ ; c) $\theta$ ; d) $\beta = \varphi - \Phi$ ; e) $1/r$ .	
○ Outliers not removed	
□ Outliers removed by CF-test	
△ Outliers removed by CS-test	
◇ Outliers removed by CG-test	
Fig. 5.19.....	123
Normalized residuals of track parameters estimated with the robust filter. Sample I with mixture model outliers. The average number of outliers is 4. The superimposed curve is a standard normal p.d.f.	
a) $\Phi$ ; b) $z$ ; c) $\theta$ ; d) $\beta = \varphi - \Phi$ ; e) $1/r$ .	
Fig. 5.20.....	124
Probability transform of $\chi^2_E$ of the robust filter (for the definition of $\chi^2_E$ see text).	
a) average number of outliers = 0;	
b) average number of outliers = 1;	
c) average number of outliers = 2;	
d) average number of outliers = 3.	

Fig. 5.21.....	125
Probability transforms of various generalized $\chi^2$ -statistics of the robust filter (for the definition see text).	
a) $\chi^2_{k,F(1)}$ for all k;	
b) $\chi^2_{k,F(2)}$ for all k;	
c) total $\chi^2_{(1)}$ ;	
d) total $\chi^2_{(2)}$ .	
Fig. 5.22.....	126
Standard errors of estimated track parameters as a function of the average number of outliers. Sample I with mixture-model outliers.	
a) $\Phi$ ; b) z; c) $\theta$ ; d) $\beta = \varphi - \Phi$ ; e) $1/r$ .	
○ Optimal linear filter	
□ Filter/smoother with removal of outliers by CS-test	
△ Robust filter	
Fig. 5.23.....	127
Power of the CS-test in FCA as a function of the distance d of the outlier from the true track position. Tracks were simulated with FASTSIM in the forward region of DELPHI. A variance-inflation outlier with threefold standard deviations was simulated in FCA. There are no outliers beyond d=16.	
Fig. 5.24.....	128
Momentum of decaying meson versus simulated kink angle.	
a) Decay of $\pi$ -meson (Sample $\Pi$ );	
b) Decay of K-meson (Sample K).	
Fig. 5.25.....	129
Difference of the found position and the true position of the kink in STPC.	
a) Sample $\Pi$ , CK-test;	
b) Sample $\Pi$ , FK-test;	
c) Sample K, CK-test;	
d) Sample K, FK-test.	
Fig. 5.26.....	130
Efficiency of CK-test (blank) and FK-test (shaded) in sample $\Pi$ :	
a) as a function of the momentum of the $\pi$ ;	
b) as a function of the position of the decay;	
c) as a function of the kink angle.	

Fig. 5.27.....	131
Efficiency of CK-test (blank) and FK-test (shaded) in sample K:	
a) as a function of the momentum of the K;	
b) as a function of the position of the decay;	
c) as a function of the kink angle.	
Fig. 5.28.....	132
Difference of the reconstructed and the true kink angle.	
a) Sample $\Pi$ , CK-test;	
b) Sample $\Pi$ , FK-test;	
c) Sample K, CK-test;	
d) Sample K, FK-test.	
Fig. 5.29.....	133
Difference of the position of the outlier and the found position of the kink, for kinks found in the wrong position.	
a) Sample N1, CK-test;	
b) Sample N1, FK-test;	
c) Sample $\Pi$ 1, CK-test;	
d) Sample $\Pi$ 1, FK-test;	
e) Sample K1, CK-test;	
f) Sample K1, FK-test.	
Fig. 5.30.....	134
Efficiency of kink finder with ID, TPC and OD (blank) and with TPC only (shaded), as a function of the momentum. The kink is inside the TPC.	
a) Samples $\Pi, \Pi'$ , CK-test;	
b) Samples $\Pi, \Pi'$ , FK-test;	
c) Samples K, K', CK-test;	
d) Samples K, K', FK-test.	
Fig. 5.31.....	135
Efficiency of kink finder with ID, TPC and OD (blank) and with TPC only (shaded), as a function of the decay position. The kink is inside the TPC.	
a) Samples $\Pi, \Pi'$ , CK-test;	
b) Samples $\Pi, \Pi'$ , FK-test;	
c) Samples K, K', CK-test;	
d) Samples K, K', FK-test.	

Fig. 5.32.....	136
Coordinate system used in the study of secondary vertex detection.	
a) View in the x-y-plane;	
b) View in the z-R-plane.	
Fig. 5.33.....	137
Probability transforms of total chi-squares. No secondary vertex.	
a) CS-algorithm, after removal of outlying tracks;	
b) CM-algorithm.	
Fig. 5.34.....	138
Number of primary tracks lost in the detection of secondary vertices.	
a) Sample A1, CS-test;	
b) Sample A5, CS-test;	
c) Sample A1, CR-test;	
d) Sample A5, CR-test;	
e) Sample A1, CM-test;	
f) Sample A5, CM-test.	
Fig. 5.35.....	139
Fraction of secondary vertices found as a function of the decay length.	
a) CR-test, sample A1 (blank) and sample A5 (shaded);	
b) CM-test, sample A1 (blank) and sample A5 (shaded).	
Fig. 5.36.....	140
Fraction of secondary vertices found as a function of the decay length.	
a) CR-test, sample A1 (blank) and sample B1 (shaded);	
b) CR-test, sample A5 (blank) and sample B5 (shaded).	
Fig. 5.37.....	141
Fraction of perfectly separated decay vertices as a function of their distance.	
CR-test, sample C1 (blank) and sample C5 (shaded).	

## Table captions

Table 5.1.....	100
Description of the detectors and scatterers in the barrel region of DELPHI as used in the simulation of tracks.	
Table 5.2.....	100
Characteristics of the three samples of tracks used in the outliers studies in STPC.	
Table 5.3.....	100
Power of CS-test, CG-test and CF-test for samples I and III.	
Table 5.4.....	101
Characteristics of the three samples of tracks used in the kink finding studies in STPC.	
Table 5.5.....	101
Efficiency of CK-test, FK-test and total $\chi^2$ .	
Table 5.6.....	101
Efficiency of CK-test and FK-test in the presence of a single outlier.	
Table 5.7.....	102
Efficiency of CK-test and FK-test with removal of outliers.	
Table 5.8.....	102
Efficiency of kink finding with information from ID, TPC and OD.	
Table 5.9.....	103
Efficiency of secondary vertex detection with CS-, CR- and CM-test, for samples A1 and A5 (primary multiplicity 12).	
Table 5.10.....	103
Efficiency of secondary vertex detection with CS-, CR- and CM-test, for samples B1 and B5 (primary multiplicity 6).	
Table 5.11.....	104
Efficiency of secondary vertex separation with CS-, CR- and CM-test, for samples C1 and C5 (two back-to-back decays).	

## References

- [1] Y. Nambu: *Quarks*. (World Scientific, Philadelphia, Singapore, 1985)
- [2] K. Johnsen: *Introduction to Accelerator Physics*.  
Elementary Particles, XCII Corso (Soc. Italiana di Fisica, Bologna, 1987)
- [3] G. Arnison et al., Phys. Lett. **126B** (1983) 398
- [4] H. Leutz und A. Minten: *Nachweisgeräte der Hochenergiephysik*.  
Physik in unserer Zeit, 11. Jahrgang, Nr. 3 (Verlag Chemie, Weinheim, 1980)
- [5] W.J. Willis, Physics Today, vol. **31**, No. 10 (1978)
- [6] C.W. Fabjan and T. Ludlam, Ann. Rev. Nucl. Part. Sci. Vol. **32** (1982) 335
- [7] C.W. Fabjan: *Calorimetry in High Energy Physics*. In: T. Ferbel (ed.),  
*Concepts and Techniques in High Energy Physics III*.  
(Plenum Press, New York, 1985)
- [8] W.W.M. Allison and P.R.S. Wright: *The Physics of Charged Particle Identification:  
dE/dx, Cherenkov and Transition Radiation*.  
In: T. Ferbel (ed.), *Experimental Techniques in High Energy Physics*.  
(Addison-Wesley, Reading (Mass.), 1987)
- [9] M. Aguilar-Benitez et al., Nucl. Instr. and Meth. **205** (1983) 79-97
- [10] DELPHI Technical Proposal, CERN/LEPC/83-3 (CERN, Geneva, 1983)
- [11] M. Regler, Comp. Phys. Comm. **22** (1981) 167
- [12] T. Sjöstrand, Comp. Phys. Comm. **39** (1986) 347
- [13] R. Brun, F. Bruyant and A.C. McPherson: *GEANT3 User's Guide*.  
CERN/DD/EE/84-1 (CERN, Geneva, 1984)
- [14] DELPHI collaboration: *DELSIM User's Guide*. DELPHI 87-96 PROG-99  
DELPHI collaboration: *DELSIM Reference Manual*. DELPHI 87-97 PROG-100  
(CERN, Geneva, 1987)
- [15] R. Frühwirth und M. Regler: *Monte-Carlo-Methoden*.  
(Bibliographisches Institut, Mannheim, Wien, Zürich, 1983)
- [16] M. Regler: *Track and Vertex Fitting*. In: M. Regler (ed.),  
*Data Analysis Techniques in High Energy Physics Experiments*.  
(Cambridge University Press, Cambridge, 1989)
- [17] G. Charpak, Physics Today, vol. **31**, No. 10 (1978)
- [18] G. Charpak and F. Sauli: *High-Resolution Electronic Particle Detectors*.  
In: T. Ferbel (ed.), *Experimental Techniques in High Energy Physics*.  
(Addison-Wesley, Reading (Mass.), 1987)
- [19] G. Charpak et al., Nucl. Instr. and Meth. **62** (1968) 262

- [20] W. Bartl, G. Neuhofer und M. Regler, *Elektrotechnik und Maschinenbau* **7** (1983) 303
- [21] M. Turala, *Nucl. Instr. and Meth.* **176** (1980) 51
- [22] P. Dechelette et al., *Nucl. Instr. and Meth.* **156** (1978) 133
- [23] R. Bouclier et al., *Nucl. Instr. and Meth.* **115** (1974) 235
- [24] R. Bouclier et al., *Nucl. Instr. and Meth.* **125** (1975) 19
- [25] A.H. Walenta, J. Heinze and B. Schürlein, *Nucl. Instr. and Meth.* **92** (1971) 373
- [26] D. Schmidt, *Nucl. Instr. and Meth.* **176** (1980) 39
- [27] J.N. Marx and D.R. Nygren, *Physics Today*, vol. **31**, No. 10 (1978)
- [28] C. Brand et al., *Nucl. Instr. and Meth.* **A252** (1986) 413
- [29] R. Arnold et al., *Nucl. Instr. and Meth.* **A252** (1986) 188
- [30] M. Metcalf: *Computers in High Energy Physics*.  
Advances in Computers, vol. **25** (1986)
- [31] D. Notz: *Real Time Data Triggering and Filtering*. In: M. Regler (ed.),  
*Data Analysis Techniques in High Energy Physics Experiments*.  
(Cambridge University Press, Cambridge, 1989)
- [32] H. Grote and R.K. Bock: *Pattern Recognition*. In: M. Regler (ed.),  
*Data Analysis Techniques in High Energy Physics Experiments*.  
(Cambridge University Press, Cambridge, 1989)
- [33] H. Grote and P. Zanella, *Nucl. Instr. and Meth.* **176** (1980) 29
- [34] A. Fröhlich et al., CERN/DD/76-5 (CERN, Geneva, 1976)
- [35] P. Billoir, R. Frühwirth and M. Regler, *Nucl. Instr. and Meth.* **A243** (1986) 173
- [36] W.I. Smirnov: *Lehrgang der Höheren Mathematik*, Teil 1.  
(VEB Deutscher Verlag der Wissenschaften, Berlin, 1971)
- [37] R.K. Bock: *Tools for Statistical Analysis*. In: M. Regler (ed.),  
*Data Analysis Techniques in High Energy Physics Experiments*.  
(Cambridge University Press, Cambridge, 1989)
- [38] *Final Report of the DELPHI Software Planning Group*.  
DELPHI 84-3 PROG-2 (CERN, Geneva, 1984)
- [39] CERN Program Library, *ZEBRA Reference Manual* (CERN, Geneva)
- [40] D. Bertrand and L. Pape: *TANAGRA User's Guide*.  
DELPHI 87-95 PROG 98 (CERN, Geneva, 1987)
- [41] CERN Program Library, *PATCHY Reference Manual* (CERN, Geneva)
- [42] R.K. Bock: *Management of the Program Life Cycle*. In: M. Regler (ed.),  
*Data Analysis Techniques in High Energy Physics Experiments*.  
(Cambridge University Press, Cambridge, 1989)
- [43] F. James, *Nucl. Instr. and Meth.* **211** (1983) 145



- [44] M. G. Kendall and A. Stuart: *The Advanced Theory of Statistics*, vol II. (Charles Griffin, London, 1967)
- [45] R. Birsa et al., Nucl. Instr. and Meth. **146** (1977) 357
- [46] J. Stoer und R. Bulirsch: *Einführung in die Numerische Mathematik II*. (Springer, Berlin, Heidelberg, New York, 1973)
- [47] R. Frühwirth, Comp. Phys. Comm. **22** (1981) 223
- [48] R. Frühwirth, Nucl. Instr. and Meth. **A243** (1985) 173
- [49] M. Regler, Acta Phys. Austr. **49** (1977) 37
- [50] J. Myrheim and L. Bugge, Nucl. Instr. and Meth. **160** (1979) 43
- [51] G. Patrick and B. Schorr, Nucl. Instr. and Meth. **A241** (1985) 132
- [52] G. Fidecaro et al., Nucl. Phys. **B173** (1980) 513
- [53] R. Settles: *Vertex detection at LEP*. In: J. Ellis and R. Peccei (eds.), *Physics at LEP*, CERN 86-02 (CERN, Geneva, 1986)
- [54] N. Stockinger and R. Dutter, Kybernetika vol. **23** (1987)
- [55] P. Laurikainen, Report Series in Physics **35** (University of Helsinki, 1971)
- [56] P. Billoir, Nucl. Instr. and Meth. **225** (1984) 352
- [57] R. Frühwirth, Nucl. Instr. and Meth. **A262** (1987) 444
- [58] A. Gelb (ed.): *Applied Optimal Estimation*. (MIT Press, Cambridge and London, 1977)
- [59] K. Brammer und G. Siffling: *Kalman-Bucy-Filter*. (R. Oldenbourg, München, Wien, 1975)
- [60] A. Jazwinski: *Stochastic Processes and Filtering Theory*. (Academic Press, New York, San Francisco, London, 1970)
- [61] P. Billoir, Private Communication
- [62] R.J. Beckman and R.D. Cook, Technometrics **25** (1983) 119
- [63] I. Guttman and D. Pena: *A Bayesian Approach to Robustifying Kalman Filtering*. In: J.C. Spall (ed.), *Bayesian Analysis of Time Series and Dynamic Models*. (Marcel Dekker, Inc., New York, 1988)
- [64] I. Guttman and D. Pena: *Optimal Collapsing of Mixture Distributions in Robust Recursive Estimation*. (unpublished)
- [65] R. Hagedorn: *Relativistic Kinematics*. (W.A. Benjamin, New York, Amsterdam, 1963)
- [66] R. Zurmühl und S. Falk: *Matrizen und ihre Anwendungen*. (Springer, Berlin, Heidelberg, New York, Tokyo, 1984)
- [67] W. Mitaroff: *Vertex Evaluation in the DELPHI Experiment*. DELPHI 88-46 PROG-112 (CERN, Geneva, 1988)
- [68] P.J. Huber: *Robust Statistics*. (Wiley, New York, 1981)

- [69] R. Dutter and P.J. Huber, *J. Statist. Comput. Simul.* **13** (1981) 79
- [70] J. Cuevas et al.: *Fast Simulation for DELPHI. Reference Manual.*  
DELPHI 87-27 PROG-72 (CERN, Geneva, 1987)
- [71] DELPHI Data Analysis Group: *Report on Local Pattern Recognition Methods.*  
DELPHI 86-56 PROG-51 (CERN, Geneva, 1986)
- [72] R. Frühwirth et al.: *The Forward Track Fit.* In print (CERN, Geneva, 1988)
- [73] P. Billoir: *Evaluation of Precision on Momentum and Extrapolation with Barrel Detectors.* DELPHI 88-44 PROG-100 (CERN, Geneva, 1988)
- [74] W. Mitaroff: *Vertex Evaluation in the DELPHI Experiment.*  
DELPHI 88-46 PROG-112 (CERN, Geneva, 1988)
- [75] W. Mitaroff: *Utility Library for Helix Tracking and Error Propagation (UHLIB).*  
DELPHI 87-51 PROG-85 (CERN, Geneva, 1987)

## Lebenslauf

Ich, Rudolf Frühwirth, wurde am 16.4.1952 in Wien als Sohn des Rudolf Frühwirth und der Ernestine Frühwirth, geb. Burgmann, geboren. Von 1958 bis 1962 besuchte ich die Volksschule in Wien 13, Steinlechnergasse; von 1962 bis 1970 besuchte ich das humanistische Gymnasium in Wien 13, Fichtnergasse, wo ich 1970 die Reifeprüfung ablegte. Nach der Ableistung des ordentlichen Präsenzdienstes begann ich im Herbst 1971 mit dem Studium der Technischen Mathematik an der Technischen Universität Wien, das ich im Juni 1976 mit der 2. Diplomprüfung abschloß. Im Februar 1977 wurde ich am Institut für Hochenergiephysik der Österreichischen Akademie der Wissenschaften angestellt; seither arbeite ich an Problemen der Datenanalyse in der experimentellen Hochenergiephysik. In der Zeit vom Oktober 1979 bis zum Mai 1984 war ich im Europäischen Laboratorium für Teilchenphysik (CERN) in Genf tätig, zuerst als "Scientific Associate" des CERN, dann im Auftrag des Instituts. Im Herbst 1987 begann ich, die Resultate meiner Arbeit an verschiedenen Experimenten in Form der vorliegenden Dissertation niederzulegen; ich schloß sie im Oktober 1988 ab.