Particle production yields induced by multi-GeV protons on nuclear targets

Dissertation
zur Erlangung des akademischen Grades
eines Doktors der Naturwissenschaften
des Fachbereichs Physik
der Universität Dortmund

Alexander Großheim

University of Dortmund, Germany
CERN, Geneva, Switzerland

December 2003

Contents

1 The Neutrino Factory

Experiments using conventional high-energy neutrino beams have played an important role in the development of particle physics. In the field of neutrino physics, they have provided increasingly sensitive searches for neutrino oscillations in short- and long-baseline experiments as well as demonstrated that muon neutrinos are different from electron neutrinos. With measurements on the structure of the nucleon via deep-inelastic scattering and precision tests on the standard model via measurements of charged- and neutral-current interactions they also contributed to particle physics not directly related to measurements of the properties of neutrinos.

Recent results from atmospheric and solar neutrino experiments indicate that neutrino oscillation may occur at rates within reach of the current generation of accelerator based experiments. Thus, they are expected to make important contributions in order to understand the current indications and to prove the presence of neutrino oscillation beyond any doubt.

However, in the long-term future of neutrino beams and related experiments, it will be required to drastically improve the intensity and composition of the beam. While conventional neutrino beams are made by allowing charged mesons to decay in-flight in a long decay channel, a new scheme has been proposed in which the beam is produced by the decay of *stored* muons: the so-called *Neutrino Factory* (NF).

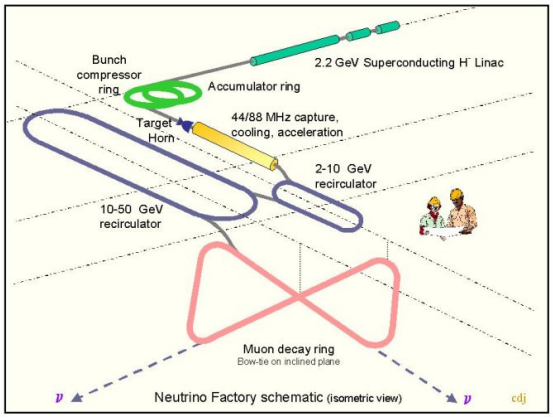


Figure 1: Possible layout of a CERN-based NF.

A NF is essentially a muon storage ring containing long straight sections to allow muons to decay to provide the high-intensity neutrino beam. The basic layout and the components of a NF are shown in Fig. 1. It consists of a proton driver delivering a multi-GeV proton beam on a pion production target surrounded by a capturing system designed to collect as many pions as possible. These pions enter a decay channel after which the phase space of the decay muons is reduced in order to fit the acceptance of an accelerator. The muons are then fed into the accelerator bringing them quickly up to a final energy of 10 – 50 GeV. Once they have the desired energy, the muons circulate in a storage ring with two or more straight sections pointing to neutrino detectors possibly at different distances.

Compared to a conventional neutrino beam the striking features of a NF are the much higher fluxes ($O(10^{21})$ per year) and the very well-known beam properties. Flux, spectrum and composition of a NF beam only depend on the parameters of the muon beam and the decay kinematics. In particular, the intrinsic background of conventional beams is absent. Moreover, by either storing negative or positive muons the beam can be charge-conjugated, i.e. it consists of ν_μ and $\overline{\nu}_e$ in case of μ^-, and ν_e and $\overline{\nu}_\mu$ if the stored particles are μ^+. This significant increase of statistics accompanied by a reduction of systematic errors makes the NF a unique tool to perform high precision measurements on the parameters of the neutrino oscillation phenomenology.

The components of a neutrino factory and related R&D activities are described below. Also, an outline of the rich physics programme which could be addressed with a NF is given.

1.1 Status of Neutrino Oscillation Searches

The current generation of neutrino experiments has provided strong indications for oscillation of solar and atmospheric neutrinos. Especially the results of SNO [1] and Super-Kamiokande [2] have established confidence in the presence of neutrino oscillation and therefore non-zero neutrino masses. The theory of neutrino oscillation and massive neutrinos shall not be described here; see for example reference [3] for an overview.

Up to now,[1] all indications for neutrino oscillation fall in the following classes:

- Six solar-neutrino experiments (Homestake [4], Kamiokande [5], SuperK [2], SAGE [6], GALLEX [7], SNO [1]) show a deficit in neutrino fluxes compared to the predictions by the standard solar model. This would be in agreement with $\nu_e \rightarrow \nu_x$ oscillations with a Δm^2_{sol} of about 10^{-5}eV^2 in the presence of matter effects (MSW[2]) with either large (≈ 1, LMA) or small ($\approx$ a few 10^{-3}, SMA) mixing angle $\sin^2\theta_{sol}$. For vacuum oscillations one obtains $\Delta m^2_{sol} \approx 10^{-10} eV^2$ at maximum mixing ($\sin^2\theta_{sol} \approx 1$).
- Anomalies in the arrival of atmospheric neutrinos observed by five experiments (Kamiokande [5], IMB [9], SuperK [2], Soudan [10], MACRO [11]) suggest $\nu_\mu \rightarrow \nu_x$ oscillations with $\Delta m^2_{atm} \approx 3 \times 10^{-3}\text{eV}^2$ and a maximal mixing $\sin^2\theta_{sol} \approx 1$. ν_x can only be a tau neutrino or a sterile one.
- Finally, a $\overline{\nu}_e$ appearance was measured by the LSND [12] experiment compatible with a $\overline{\nu}_\mu \rightarrow \overline{\nu}_e$ transition with parameters between $\sin^2\theta_{LSND} = 10^{-3}, \Delta m^2_{LSND} = 1$ eV^2 and $\sin^2\theta_{LSND} = 1, \Delta m^2_{LSND} = 7 \times 10^{-2}\text{eV}^2$. However, so far this result has not been confirmed by any other experiment.

With these data in hands and the constraints on mixing parameters given by theory one can set up various scenarios for the oscillation framework, including the postulation of a fourth, sterile neutrino.

The knowledge so far can be summarised as follows:

- There are at least three flavours participating in neutrino oscillations.
- $\sin^2\theta_{23} \approx 1$ (≥ 0.9 at 90% CL)
- $|\Delta m^2_{32}| \approx 3 \times 10^{-3} eV^2$
- $\Delta m^2_{21} \approx 5 \times 10^{-5} eV^2$ (LMA)
- $\sin^2\theta_{12} \approx 0.87$ (LMA)
- $\sin^2\theta_{13} < O(0.1)$

[1] Beginning of 2003.
[2] see [8]

A rather complete illustration of oscillation search results is shown in Fig. 2.

Currently, intense efforts are being undertaken to further study neutrino oscillations. New experiments in commissioning phase or with new data are K2K [14], KamLAND [15], MiniBooNE [16] and MINOS [17]. They will be able to prove that the existing evidence is indeed caused by neutrino oscillation and to make more precise measurements of at least a few oscillation parameters as mentioned above.

Recent results from SNO [18] prove that the solar neutrino flux reduction is caused by oscillation and the KamLAND experiment found evidence for the disappearance of reactor antineutrinos [19]. With these results, there is now a unique solution to the "solar neutrino puzzle" with $\Delta m^2 \approx 7 \times 10^{-5}$ *and* $\sin^2 2\theta_{12}$ *large, but smaller than maximal.*

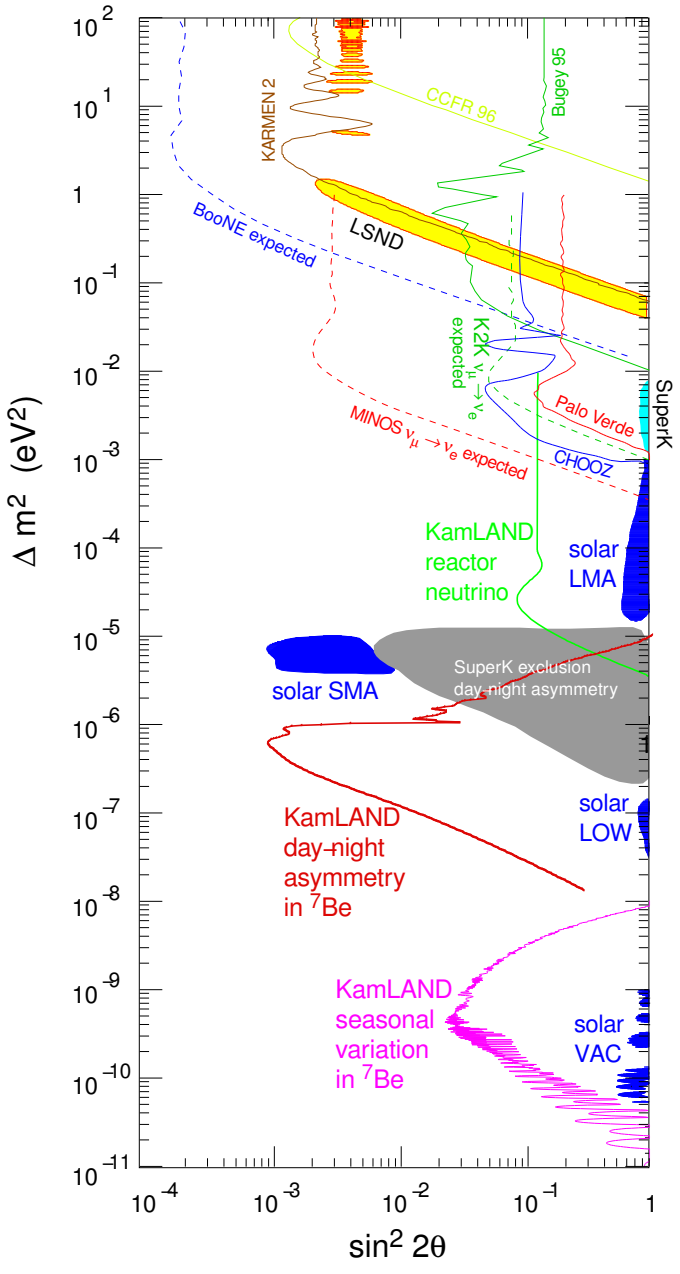


Figure 2: Compilation of oscillation search results (from [13]).

1.2 The Physics Programme at a Neutrino Factory

The physics potential of presently operating experiments is mainly limited by the properties of neutrino beams available. Thus, by using a different technique to generate the beam, a series of new measurements could be performed.

Note that all the measurements described below require suitable neutrino detectors at the right distances.

1.2.1 Conventional ν Beams versus NF

A conventional neutrino beam is made by allowing charged pions and kaons to decay in a long decay channel. The *pi* and K are produced as secondaries of a high-intensity proton beam impinging on a hadron production target. When positive hadrons are selected, the neutrino beam will contain mostly muon neutrinos from $\pi^+ \to \mu^+\nu_\mu$ and $K^+ \to \mu^+\nu_\mu$ decays. In addition, there is a $O(1\%)$ component of electron neutrinos caused by the three-body decay $K^+ \to e^+\pi^0\nu_e$ and a $O(3\%)$ contamination with $\overline{\nu}_\mu$ from μ^+ decay. If the proton energy is sufficiently high, the beam will also contain a small ν_τ component. It is evident that this beam composition is not ideal for most neutrino experiments: the small contamination with ν_e and ν_τ makes experiments difficult searching for appearance of tau or electron neutrinos; at the same time the ν_e content is not high enough to observe ν_e disappearance.

If the neutrino beam was produced from the decay of *stored* particles, the ν beam parameters (composition, flux, spectrum) only depend on the parameters of the primary beam and the decay kinematics of the stored particles. For the primary beam, only muons are suitable, since the lifetime of pions and taus is too short and electrons are stable. In fact, also radioactive nuclei can be used emitting ν_e via β-decay, see *Beta Beams*. Storing negative muons produces a beam with a precisely known mixture of 50% muon neutrinos and 50% electron antineutrinos, and charge-conjugated in case of μ^+. The neutrino and antineutrino energy and angular distribution only depend upon the parent muon energy, the decay angle and the muon polarisation. Thus, provided that the current, momentum and polarisation of the muon beam are accurately measured and controlled, the absolute neutrino fluxes can be precisely calculated. In addition, only *one* type of neutrino and *one* type of antineutrino is present in the beam (which can be charge-conjugated by selecting the polarity of the muon beam). This permits precise measurements with ν_e, $\overline{\nu}_e$, ν_μ and $\overline{\nu}_\mu$.

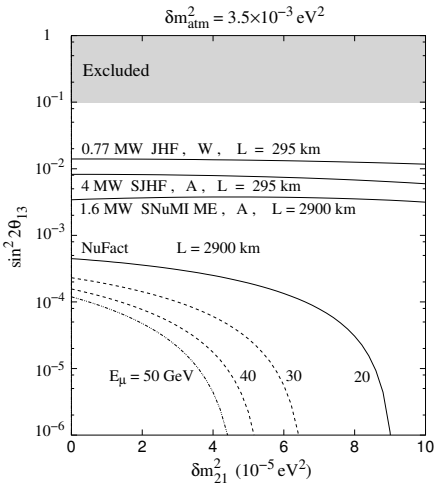


Figure 3: Summary of the 3σ sensitivity for the observation of $\nu_\mu \to \nu_e$ oscillations with various superbeams and $\nu_e \to \nu_\mu$ oscillations with NF beams [20]. The shaded area is excluded by present experiments. An experiment running period of 3 years is assumed.

Superbeams are a further development of conventional neutrino beams. Using a higher proton beam intensity (which is comparable to the one foreseen for the NF), the neutrino flux can be increased by a factor of ≈ 5. The much higher ν luminosity allows to exploit the decay kinematics of pions to produce a so-called off-axis beam. Accepting the lower overall flux at a few degrees from the beam axis, this leads to an important relative suppression of the electron neutrino contamination and a cut in the high momentum tail of the beam. A detailed comparison of the physics potential of superbeams and a NF can be found in [21, 22]. For illustration, a comparison of the physics possibilities for different NF beams parameters is given in Fig. 4.

Recently a neutrino beam from the β-decay of stored radioactive ions was proposed [23]. Thus, the generation of this so-called *Beta Beam* is similar to that of a NF. A unique feature of such a neutrino beam is its single-flavour content of either 100% ν_e or $\overline{\nu}_e$, which is advantageous for certain

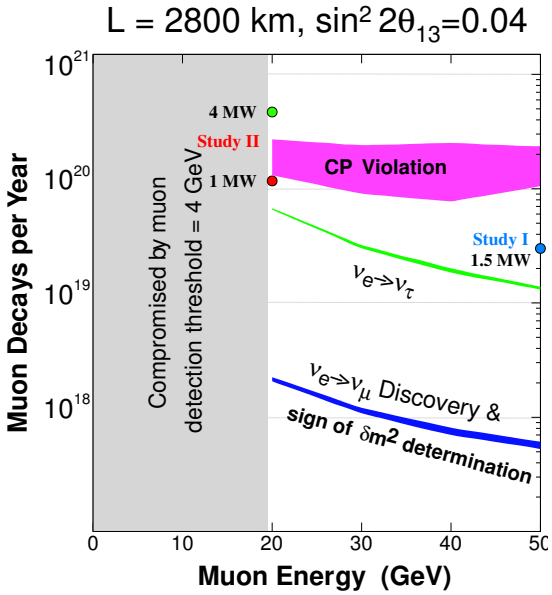


Figure 4: Physics cases for NF beams with a given flux and energy [20].

measurements.

1.2.2 Oscillation Physics

It is expected that a NF will not be build within the next ten years. By then, issues of interest are most likely the following:

- A precise determination of the parameters of neutrino oscillation. This includes all oscillation probabilities $P(\nu_e \to \nu_x)$ and $P(\nu_\mu \to \nu_x)$. With these data a forth, sterile neutrino could be confirmed or stringent limits on its existence are put. The measurements on $P(\nu_\mu \to \nu_\tau)$ and $P(\nu_\mu \to \nu_e)$ will be more accurate than with conventional beams, whereas a first observation of $P(\nu_e \to \nu_\tau)$ is unique to a NF. All mixing angles (especially in the difficult case of θ_{13}) should then be known or at least stringent limits will be placed.

- A measurement of the neutrino mass hierarchy, especially the sign of Δm^2_{32}. This will answer the question whether there are two 'heavy' neutrinos and one light one or vice versa. Since matter effects have to be exploited for this measurement, it calls for very long baselines of several thousand km. While those measurements can in principle be done with a conventional beam, low rates and systematic uncertainties will limit the precision significantly.

- The measurement of matter effects to confirm the significance of the MSW effect. The same scheme as for the determination of Δm^2_{32} can be used here; high statistics and low systematics which can only be provided by a NF permit this measurement.

- The observation or stringent limits on CP violation in the lepton sector. The CP test involves comparing $\nu_e \to \nu_\mu$ with $\overline{\nu}_e \to \overline{\nu}_\mu$ oscillation rates which can only be done with a NF with the required precision since backgrounds and systematic errors are very small.

1.2.3 Non-oscillation Physics

The high intensity beam of a NF can of course also be used to study non-oscillation topics:

- precise determination of neutrino-nucleon cross-sections in a near detector and precise measurement of nucleon structure functions

- tests of the electroweak model with the measurement of the mixing angle $\sin^2\theta_W$ to $\pm 10^{-4}$ and neutrino-electron cross-sections

- search for magnetic moment of neutrinos

The neutrino factory is also considered as a first step towards a multi-TeV muon collider [24]. A scenario leading to the construction of such a machine and its physics programme are presented in [25].

1.3 NF R&D

Recent studies have been performed for a NF on various sites in the US (BNL or FNAL), Europe (CERN) and Japan (JHF). The choice of these sites is mainly driven by the possibilities of reusing existing machines, equipment and infrastructure. Consequently also the R&D activities are oriented towards these scenarios.

Common to all present NF schemes is the use of an intense multi-MW, multi-GeV proton beam to produce low-energy pions and to collect them using a large-acceptance decay channel. The daughter muons from $\pi^{\pm}$ decay occupy a large phase-space volume which does not fit the acceptance of a conventional accelerator. To reduce the energy spread, the American and European studies suggest ionisation cooling and phase rotation. In Japanese studies an alternative strategy is pursued using fixed-field alternating-gradient (FFAG) accelerator modules which have a very large acceptance [26]. Finally, in all three schemes the muons are accelerated to their final energy of 10-50 GeV and fed into the storage ring.

It is widely agreed that a coherent and optimised design of all the modules needed for a neutrino factory can not be made before two especially critical issues have been investigated and tested in detail:

- the target station with the pion production target and the pion capturing system and
- the reduction of the phase space ("cooling") of the muons from pion decay.

Whereas cooling is a novel technology where no experience exists, the situation is different for the target. The available experimental data on hadron production is not sufficient to decide for the most efficient target setup. Open questions are shape and material of the production target, the optimal incident proton energy and the pion capturing system.

Thus, current R&D efforts concentrate on these two items. Among other physics goals, the HARP experiment was proposed and carried out to remove the lack of accurate hadron production cross-sections at multi-GeV energies which are of interest for the pion production front-end of a NF.

The single components of a NF are described below in some more detail. Emphasis will be given to the target and the pion collection system.

1.3.1 Proton Driver

To produce the required number of useful pions emerging from the target the proton driver needs to provide a beam power of 1–4 MW. While currently no proton accelerator is capable of delivering such intensities it is considered as challenging but feasible by accelerator experts. Of course it is desirable to upgrade an existing proton machine to NF specifications or at least to reuse existing equipment. Proposals are an upgraded JHF with a 4 MW 50 GeV proton synchrotron and a new 4 MW 2.2 GeV superconducting proton linac (SPL) with an accumulator ring to produce short pulses at CERN. At the suitable US sites BNL and FNAL, an upgraded Alternating Gradient Synchrotron (AGS) or a new fast cycling 16 GeV synchrotron are being studied. More scenarios can be envisaged, but the significantly different proton energies of the machines mentioned above shows that the results of pion production studies have a strong impact on the choice of the accelerator setup. Issues like bunch length and frequency also depend upon the target choice.

1.3.2 Target and Capture

Considering the high beam power of up to 4 MW it is clear that mechanical issues play an important role in the choice of the target station. Conventional solid (high-Z) targets will melt or break within a short time due to the high thermal power deposited in the material and the mechanical shocks caused by the impact of bunched protons. For liquid targets, this would also apply to eventual pipes or vessels exposed to the beam. In addition, a high pion yield has to be assured. Current knowledge of pion production cross-sections suggests that for energies above 10 GeV a high-Z target would provide the highest number of useful pions. Consequently, in most studies a liquid Hg jet is being considered as best option. However, solid low-Z targets like a long carbon-carbon composite cylinder and other high-Z options like metal beads cooled by a water flow and moving-band targets are also conceivable.

The pions produced in the target have to be collected so that they are useful for the final production of a high-quality muon beam. For the capturing of pions two options exist: a magnetic horn (for a later design study see reference [29]) and a

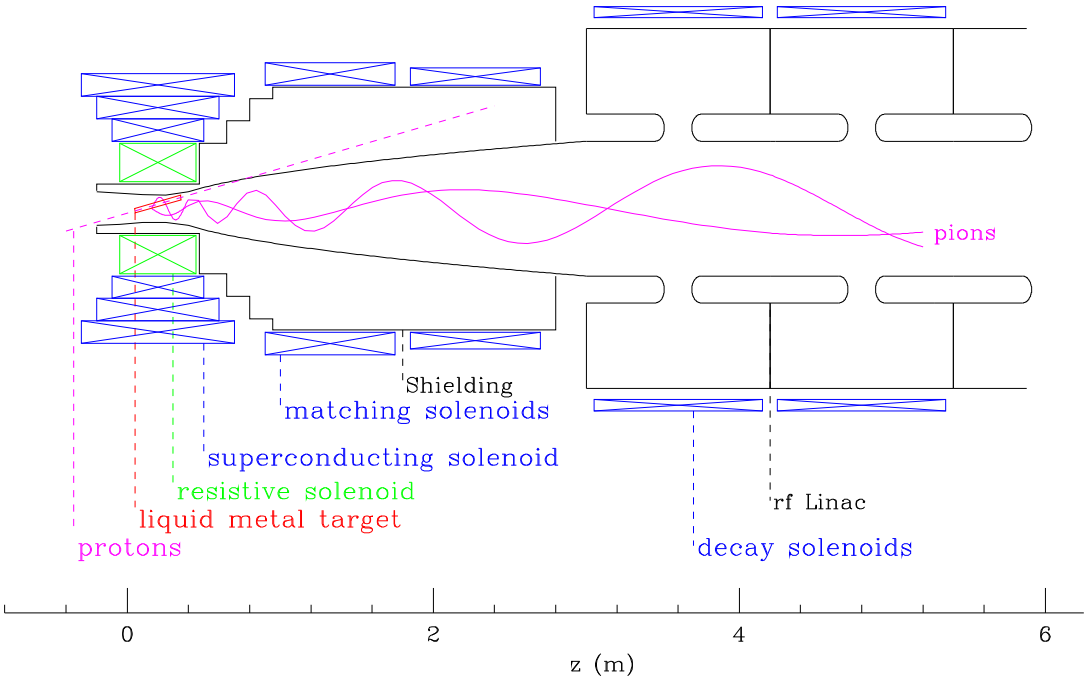


Figure 5: Target station and solenoidal capture system (from [27]).

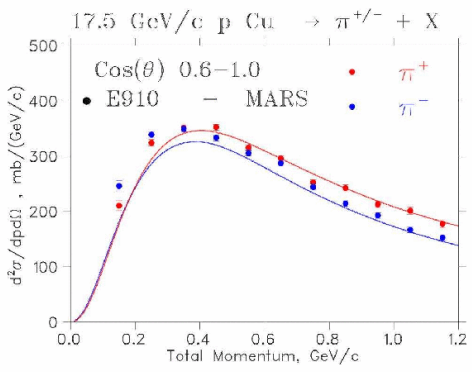


Figure 6: Pion production cross-sections as measured by the BNL experiment E910. Shown is also the comparison of the measurement with predictions from the MARS simulation code (from [28]).

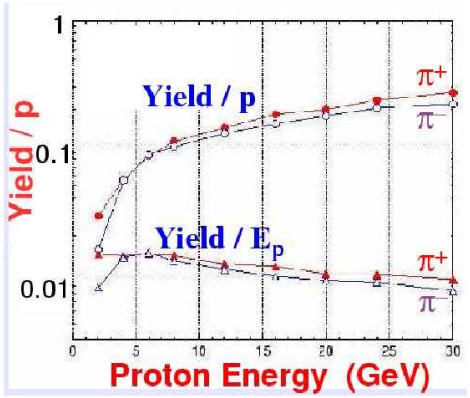


Figure 7: MARS simulation of the pion yield of a NF decay channel with a solenoidal capturing system as function of the incident proton energy. For details see reference [31]).

solenoidal capturing system (Figs. 5, 7). Simulations show, that both options would provide a similar pion yield [30].

It is evident that all components of the target and capturing system must be carefully chosen and optimised to provide a coherent design of the most efficient pion production front-end.

1.3.3 Phase Rotation, Bunching and Cooling

The pions and muons into which they decay are generated in the target over a very wide range of energies. To reduce this spread they undergo a so-called phase rotation. Using long induction linacs intersected with drifts, the early ones are decelerated and the late ones accelerated. These linacs also represent the decay channel.

In order to fit the acceptance of a conventional accelerator, the phase space of the decay muons has to be reduced. This so-called "cooling" follows the idea of reducing the total momentum by ionisation losses in matter and restoring the longitudinal momentum component with RF cavities. This can be done in several steps. Since muon cooling is a novel technology which has never been used before, a big R&D effort is underway to prove that it is feasible and performs as required. Notably the MICE [32] experiment planned to start up in 2004 aims at results which could be the basis of a technical proposal for the cooling system of a NF.

1.3.4 Acceleration

While depending on the actual parameters of the cooling and bunching system, the acceleration to the final energy of up to 50 GeV is planned to be done in three steps. Directly after the cooling, the muons are accelerated in a linac to a few GeV. Then, two recirculating linear accelerators (RLAs) are used to raise the energy in two steps first to $\approx$ 10 GeV and then to the final energy of 20, 30 or 50 GeV.

The most critical design requirement for the acceleration system is the time it takes until the accelerated muons enter the storage ring. Like this the muon-loss due to decay is kept at a minimum.

1.3.5 Storage Ring

The storage ring is a more or less straight-forward extrapolation of electron storage rings. However, various shapes and sizes are proposed. In any case, the ring should contain two or three straight sections where the decaying muons produce the desired neutrino beam. Thus, the straight parts should be as long as possible compared to the curves. Especially the orientation of the straight sections has to be carefully chosen in order to point to an optimal detector site. A triangular shape is possible as well as racetrack or bow-tie shapes, not necessarily in the horizontal plane (see Fig. 8). As pointed out in [22], details of the beam diagnostics also depend on the ring shape.

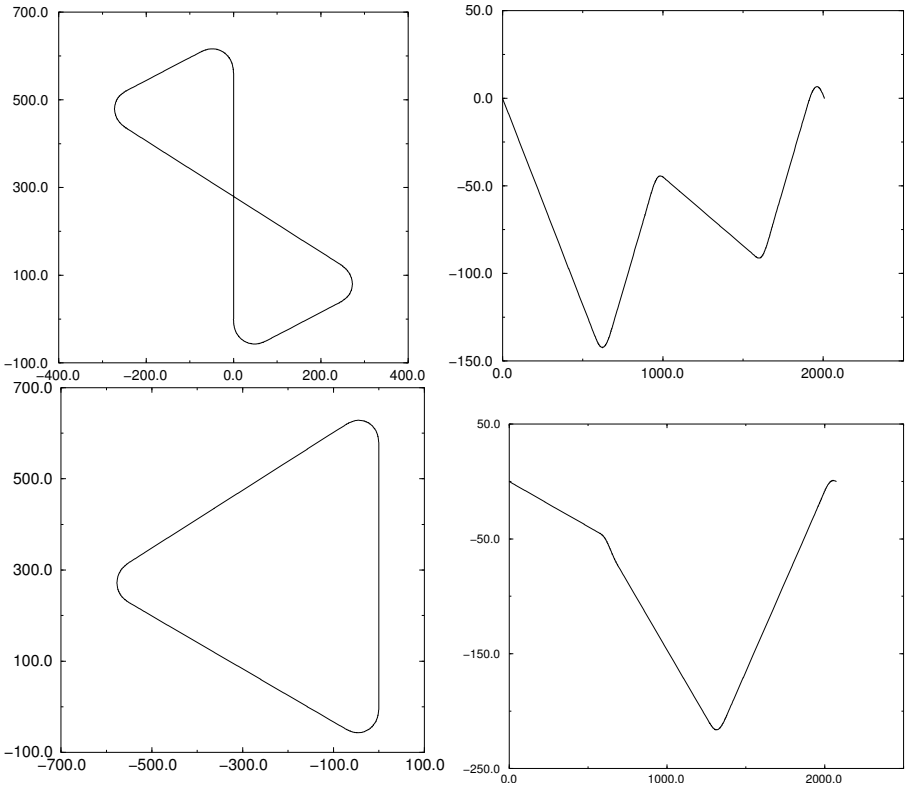


Figure 8: Possible shapes of a muon storage ring: bowtie (top) and triangle (bottom). Note that the left pictures show the ring in the horizontal plane (orientation arbitrarily) and the right ones the extension of the ring in the vertical plane (from [22]).

2 The HARP Experiment

HARP is a fixed-target experiment at the CERN proton synchrotron (PS) to measure hadron production cross-sections of protons on nuclear targets. Its main physics goals are high-accuracy pion production rates for the front-end of a neutrino factory and improvements of the prediction of atmospheric neutrino fluxes. HARP uses protons in the range 2 to 15 GeV/c on thin and thick targets of various Z and has an almost full solid angle acceptance for secondaries.

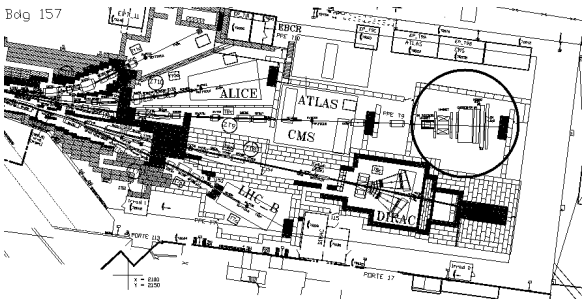


Figure 9: Location of the HARP spectrometer (circle) in the T9 experimental area at the CERN PS.

The experiment was proposed to the CERN SPSC on 01/12/1999 [33] and approved on 17/02/2000 by the CERN Research Board. A technical run with the partial detector was performed in 2000, the complete detector started physics data taking in August 2001. After taking 450 million physics triggers for about 300 different settings data taking ended in October 2002 (see Fig. 10). In total, more than 30 TB of data have been acquired. In February 2003, the dismantling of the detector was finished.

Physics motivation and the experimental setup will be described below.

2.1 Physics Goals

2.1.1 Neutrino Factory

The need for low-energy pion production measurements for an optimal design of the front-end of a neutrino factory or a muon collider is motivated in section 1.3.

The desired experimental data as stated in the HARP proposal [33] is given in Table 1.

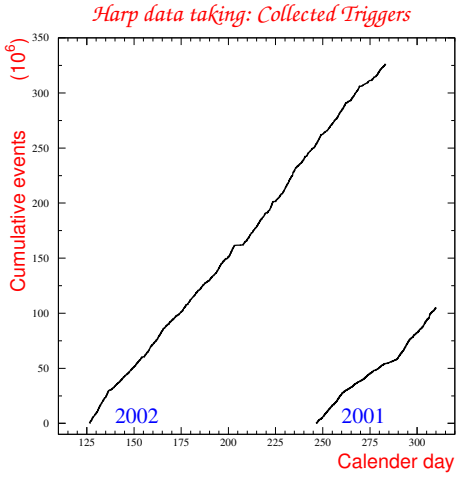


Figure 10: Amount of data taken in 2001 and 2002.

	Range	Precision
π p_L	100–700 MeV/c	$<$ 25 MeV/c
π p_T	0–250 MeV/c	$<$ 25 MeV/c
π/p yield		5%
π^+/π^- ratio		5%

Table 1: Desired experimental data for the Neutrino Factory.

2.1.2 Atmospheric Neutrinos

For the interpretation of results of neutrino experiments using atmospheric neutrinos it is crucial to correctly calculate the atmospheric neutrino flux.

Atmospheric neutrinos are generated via the following reaction chain: primary cosmic ray particles (p, He) react with the molecules of the atmosphere (O_2, N_2) and produce protons, kaons and pions. At the end of the decay chain primarily muons and electron- and muon-neutrinos arrive at earth's surface.

As discussed in [34], the dominant uncertainties in the neutrino flux are caused by the representation of hadron production in collisions of the primary cosmic rays with nitrogen and oxygen nuclei.

In this context two issues are of importance for the measurement:

- Full phase space coverage is needed.
- The normalisation precision of the data should be very good.

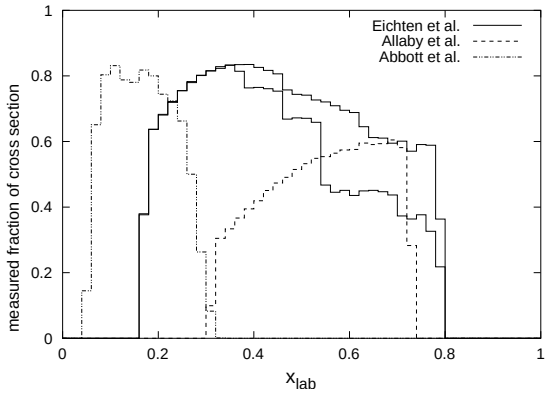


Figure 11: Fraction of pion production covered by data (from [34]). Note that only one experiment reaches in the x_{lab} region of ≤ 0.2 which forms an important contribution to the atmospheric lepton fluxes.

To illustrate the situation, Fig. 11 shows the coverage of proton-Beryllium cross-sections with experimental data. It should be recalled that this data has to be extrapolated from Be to nitrogen and oxygen nuclei. In addition, a fraction of the primary cosmic rays are bound in helium nuclei which requires another extrapolation since no experimental data of the pion yield of He-N_2 collisions is available.

2.1.3 Fluxes of Conventional Neutrino Beams

As discussed in section 1.2, one of the problems of conventional neutrino beams is that their composition, flux and spectrum is not known very well, which is a major limitation for experiments using those beams. These uncertainties can be traced back to inaccurate predictions of pion and kaon production yields of their primary targets.

In association with HARP, two running neutrino oscillation experiments will use HARP data to improve their beam composition knowledge: MiniBooNE [35] and K2K [36]. Particle production rates of replicae of the targets of both experiments are measured by HARP at their nominal proton beam energy: a thick Al target at 12.9 GeV for K2K and a thick Be target at 8.9 GeV for MiniBooNE.

2.1.4 Hadronic Generators

Most of the above mentioned needs for hadron production yields could in principle be satisfied using predictions from Monte-Carlo simulations. However, the results of current hadronic generators differ largely (see Fig. 12), which is due to two difficulties:

- There is practically no experimental data on hadronization at low (few GeV) energies available.
- It is difficult to describe the involved processes theoretically with a sufficient precision. In any case, predictions would have to be cross-checked with data.

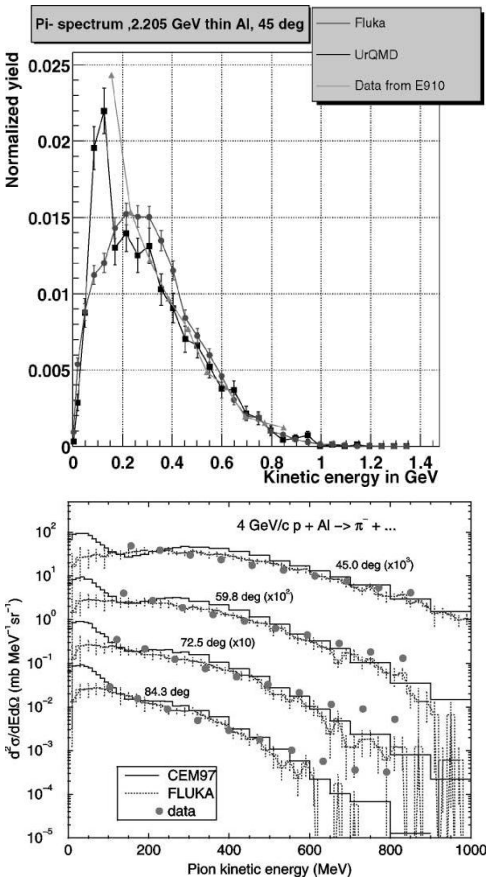


Figure 12: Comparison of pion spectra expected by different hadronic generators (taken from [37, 38]).

Therefore, HARP cross-section measurements will serve the GEANT4 [39] collaboration to refine

the implementations of hadronic generators.

2.2 The Detector

The HARP spectrometer can be divided into three major subsystems:

- The trigger and beam detectors are installed in order to reconstruct, identify and trigger on an incoming beam particle. In addition the trigger has to detect secondaries emerging from the target with almost full solid angle acceptance.
- Measuring secondaries at angles above $\approx 20°$ relies on the large-angle part of HARP. It includes the Time Projection Chamber (TPC) and the Resistive Plate Chambers (RPCs) installed around the TPC (Fig. 13).
- Secondaries at small production angles are measured by the forward spectrometer. It consists of drift chambers for tracking and a dipole magnet for momentum determination. A time-of-flight (TOF) system, a gas Cherenkov detector and an electron- and muon-identifier are employed for particle identification (PID) as can be seen in Fig. 13.

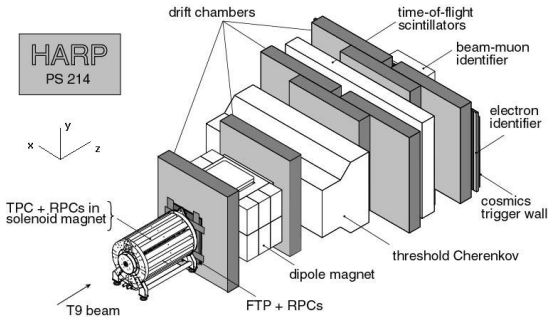


Figure 13: The HARP detector setup. The single components are described in the text.

2.2.1 Beam and Trigger Detectors

Beam detectors serve to identify a beam particle and to reconstruct its track parameters. Particles to be distinguished are p, π, K, μ and e in the momentum range 2 to 15 GeV/c; the track reconstruction should allow to find the particle impact point on the target with an accuracy better than 1 mm.

The detectors used in the trigger can be divided into two subgroups: the detection of a beam particle concerns all detectors in the beam-line upstream of the target, whereas the detection of secondaries emerging from the target relies on downstream detectors and detectors surrounding the target.

A schematic view of the relative position of all trigger and beam equipment is shown in Fig. 14. Below, all devices and their specific tasks are described.

Beam Cherenkov Counters

Two gas Cherenkov detectors (BCA, length 6000 mm and BCB, length 3000 mm) are inserted into the beam pipe upstream of focus A (see Fig. 14). They are filled with CO_2 at different pressures depending on the beam energy. At high energy, when the resolution of the beam TOF system is not sufficient, the Cherenkov counters are used to tag electrons and pions in the beam. Table 2 summarises the basic pressure settings of BCA and BCB for different beam energies. These settings have been adjusted during the data taking to meet different requirements for the trigger and to react on major variations of the beam composition.

Each detector has a sensitive diameter of about 180 mm and the light is collected with a mirror system to a single photomultiplier tube. In practice, the efficiency in tagging particles mainly depends on the correct pressure setting. Some redundancy can be achieved by choosing the pressures in such a way that both detectors are tagging the same type of particles.

beam momentum	BCA	BCB
≤ 3 GeV/c	0.0 (-)	1.05 (e)
5 GeV/c	0.60 (e)	2.50 (π, e)
≥ 8 GeV/c	1.25 (π)	1.50 (π)

Table 2: Default pressure settings for the beam Cherenkov counters and particle types which are tagged. Pressures in bar(absolute). At 3 GeV/c, BCA is evacuated in order to reduce multiple scattering of the beam.

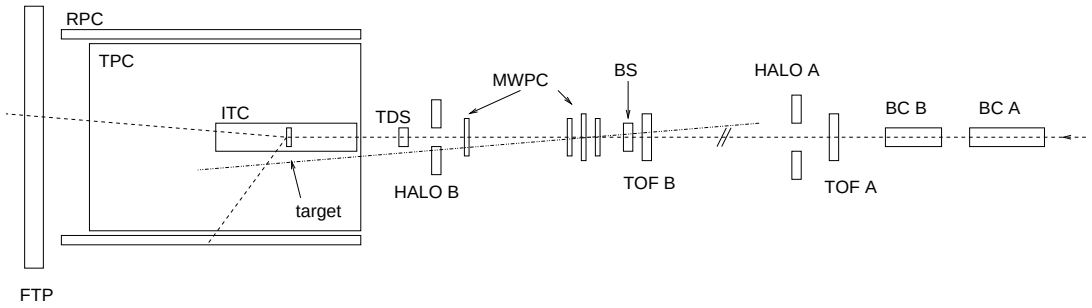


Figure 14: Schematic view of the arrangement of all trigger and beam equipment. Detailed descriptions in the text. The beam enters from the right.

Beam Halo Counters

Vetoing events where the beam particle is accompanied by a second particle in the halo of the beam is performed by two HALO counters.

HALO A is located near focus A and its primary purpose is to veto all particles which did not pass the beam Cherenkov counters. The geometry of HALO A is shown in Fig. 15. It has a double-sided readout with two Philips XP2020 photomultiplier tubes. The diameter of its hole is 9 cm.

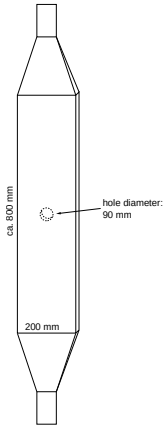


Figure 15: Schematic view of HALO A.

The final cleanup of the beam is done by HALO B (see Fig. 16) as near to the target as possible. With a minimum hole diameter of 3 cm it is complementary to the target defining scintillator (TDS, d = 2 cm) and the target (d = 3 cm). To convert gammas, an 11 mm $\approx 2 \times X_0$ thick lead plate is sandwiched between the two scintillator slabs. More details on HALO B can be found in [40].

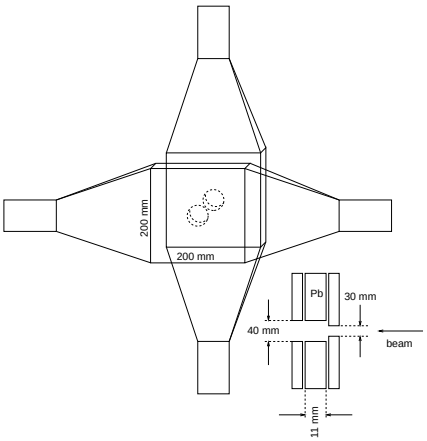


Figure 16: Schematic layout of HALO B. The counters are mounted with a rotation of 45 degrees around the beam axis. The size of a single scintillator slab is 200 × 200 × 5 mm^3. The beam direction which is indicated in the figure is valid for data taken in 2001. For the 2002 runs the scintillator planes were swapped.

To achieve a maximal veto-efficiency, each halo counter enters the trigger with a logic *or* of its channels.

Beam Multi-Wire Proportional Chambers

Four Multi-Wire Proportional Chambers (MWPCs) are installed in the beam-line to measure the track parameters of the beam particle. The tracks are extrapolated to the target to find the exact impact point with an accuracy of better than 1 mm. The angle of the particle with respect to the nominal beam axis is reconstructed as well.

As shown schematically in Fig. 14, all MWPCs

are located not too far away from the target in order to reduce multiple scattering effects between the track measurement and the target. However, the distance between the first and last chamber has to be large enough to allow an accurate reconstruction of the direction. The most important technical parameters of all chambers can be found in Table 3.

chamb.	z [mm]	area [mm^2]	# wires	spacing [mm]
1	-3961	96×96	96	1
2	-3126	96×96	96	1
3	-1263	96×96	96	1
4	-3543	192×192	48	4

Table 3: MWPC parameters: z with respect to focus B. The number of wires is given per plane. Chamber 4 is rotated by 45 degrees with respect to the chambers 1-3.

Each chamber has two independent (anode) sense wire planes rotated at 90 degrees with respect to each other. Their distance is ≈ 10 mm. In the middle, there is a common (cathode) field wire plane. For chambers 1,2 and 3 the sense wires are parallel to the $x-$ and $y-$axis of the HARP reference system (Fig. 13), in chamber 4 they are rotated by 45 degrees to facilitate the removal of ambiguities in the track reconstruction. The gas is a mixture of 50% Ar and 50% CO_2.

More information on technical details, alignment and track reconstruction of the MWPCs can be found in [41].

For the trigger decision, no information from the MWPCs is used. However, the offline selection of "good" beam particles also relies on the track reconstructed with the MWPC data.

Beam TOF system

A time-of-flight measurement system is used for the beam particle identification at low energies. It is complementary to the PID capabilities of the beam Cherenkov counters. In addition, the beam momentum can be calibrated by measuring various particles with different (known) masses at the same momentum as shown in Fig. 18.

Two identical TOF scintillation counters[3], TOFA (near focus A) and TOFB (surprise, surprise near focus B) with a distance of approximately 21.4 m provide a time-of-flight resolution of ≈ 150 ps. Each counter consists of eight single scintillator strips with double-sided readout (Fig. 17), i.e. 16 channels in total. For the trigger, a TOF counter is considered "hit" if at least one of the 16 photo-multipliers (PMs) gave a signal.

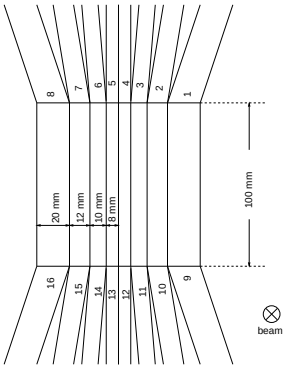


Figure 17: Layout of the TOFA/B counters. The total sensitive area is 100 × 100 mm^2, the thickness of the scintillators is 5 mm.

An example of the joint particle identification capabilities of beam-TOF and -Cherenkovs is given in Fig. 18.

Beam Scintillator

The beam scintillator (BS) is the detector which starts the trigger decision of the trigger system. It is located a few centimetres downstream of TOFB and in coincidence with a TOFB hit it represents the lowest-level trigger, the so-called *strobe*. As a consequence, the timing of all trigger signals is given by the original BS signal, i.e. BS is the timing reference for all ADC gates, TDC starts/stops etc.

BS is a single scintillator slab of 5 mm thickness and a sensitive area of 80 × 80 mm^2 centred on the nominal beam axis. It is read out by a single Philips XP2020 photomultiplier tube.

Target Defining Scintillator

The Target Defining Scintillator (TDS) is a scintillator disc of 20 mm diameter and 5 mm thickness which is viewed by four photomultiplier tubes

[3]The counters were recuperated from the NA 52 experiment.

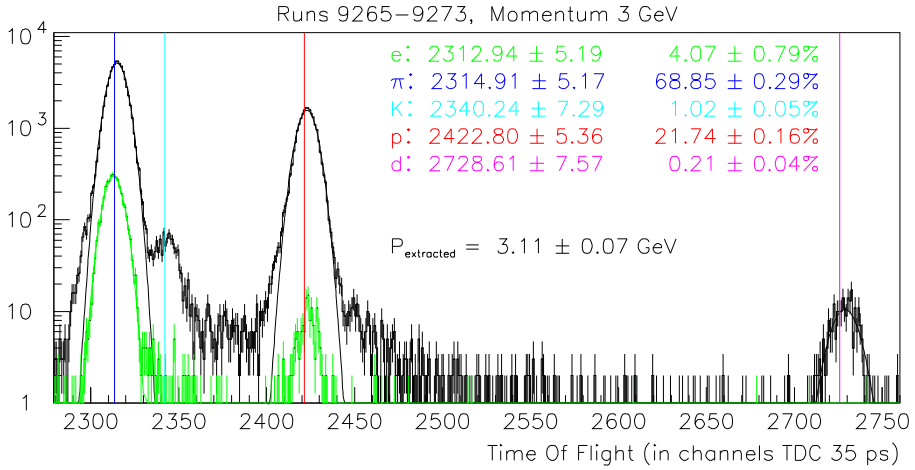


Figure 18: Example for beam particle identification with TOFs and Cherenkovs at a nominal beam momentum of 3 GeV/c. In these runs, electrons (overlaying plot) are tagged with both Cherenkovs. The purity of the beam particle identification is better than 99%.

(HAMAMATSU R1635P 3/8 inch), as shown in Fig. 19. The PMs are located at a distance of approximately 20 mm from the edge of the scintillator and optically connected via 'air light-guides', i.e. a tube formed with reflecting Mylar foil. They are housed in iron blocks in order to shield them from the magnetic stray field of the solenoid.

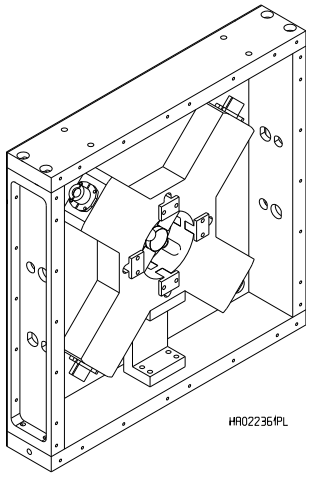


Figure 19: Design drawing of TDS. The iron block housing the four PMs can be seen. The scintillator disc in the centre is suspended by strings. Both sides of TDS are closed with black plastic sheets, except for the centre. Here, a thin aluminised plastic foil is used as light protection in order to reduce the amount of material in the beam.

The TDS is designed to have a very high efficiency ([42]), thus all particles hitting the target are also supposed to give a signal in TDS. Therefore it is located as near as possible to the entrance of the TPC and with its 20 mm diameter is somewhat smaller than that of the target being 30 mm. The TDS gives a signal if at least one PM was hit. An efficiency of well above 99.9% is assured.

During data taking it turned out that the TDS also has a sufficiently good time resolution and stability to be used as redundant part of the beam TOF system with the baseline TOFA-TDS.

Inner Trigger Cylinder

Triggering on large-angle secondaries emerging from the target is performed by the Inner Trigger Cylinder (ITC). It is mounted inside the inner field cage of the TPC and consists of an aluminium/carbon tube with a length of 1300 mm and an inner and outer diameter of 76 mm and 92 mm, respectively. Six layers of scintillating fibres (each of diameter 1 mm) are glued on the tube (Figs. 20 and 23).

The readout is performed by 24 PMs (HAMAMATSU R1635P 3/8 inch) which are contained in a soft iron magnetic shield on the upstream end cap of the solenoid magnet. They are connected to 24 fibre bunches according to the scheme shown in Fig. 21: the fibres of layers 1–4 are divided into

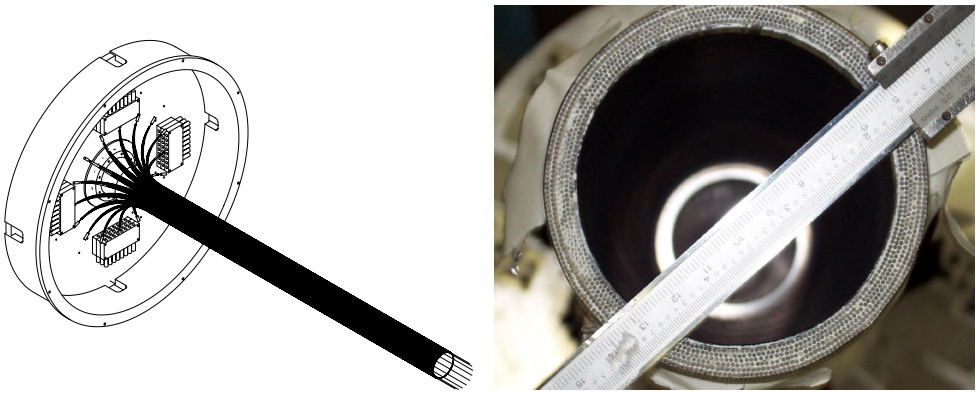


Figure 20: Left: Sketch of ITC. It is inserted into the inner field cage of the TPC. The compartment with the PM housings is fixed outside the magnetic field on the upstream solenoid end-cap. On the right, the downstream end of the ITC is shown after dismantling. The in total six layers of scintillating fibres can be seen. The fibre diameter is 1 mm.

2 times 8 segments (with constant ϕ for all z) and are connected to PMs 1–16. The fibre bunches of layers 5 and 6 are wound around the cylinder in opposite senses and connected to PMs 17-24.

Triggering on a logical *or* condition of all 24 channels provides a combined efficiency for a single track detection of well above 99%.

Forward Trigger Plane

Located directly at the downstream exit of the solenoid, the Forward Trigger Plane (FTP) covers the small angle region complementary to the ITC. It is made of two planes of scintillator slabs as shown in Fig. 22. One plane consists of 7 scintillators each of $1240 \times 200 \times 5$ mm^3, and therefore covering a total area of 124×140 cm^2. Each panel is readout on both sides by fine-mesh photomultiplier tubes. Both scintillators (Bicron BC-408) and PMs (Hamamatsu R2490) were previously used in the trigger system of the NOMAD experiment [43]. The two planes are rotated by 90° with respect to each other in order to minimise the insensitive areas due to the slits between the scintillators.

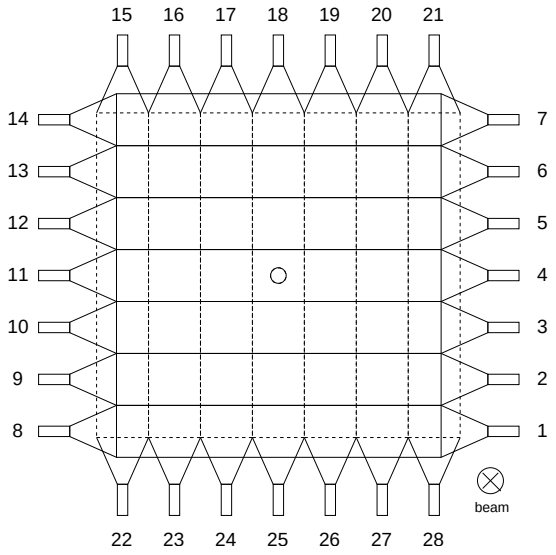


Figure 22: Layout of the FTP scintillator slabs and channel numbering scheme. The second (vertical) plane is downstream of the horizontal one.

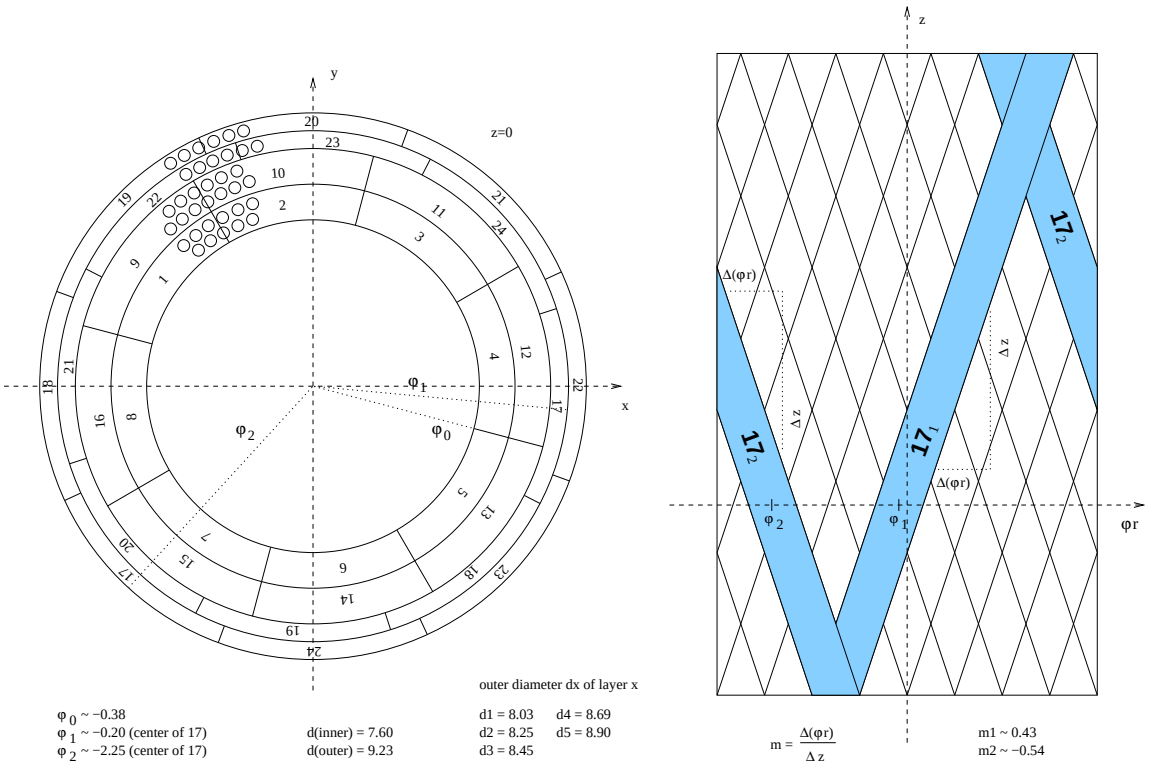


Figure 21: Alignment and channel numbering of the ITC. The alignment constants follow from an analysis using extrapolated TPC tracks to determine the impact point on the ITC.

2.2.2 Large-Angle Detectors

TPC

The HARP TPC is a time projection chamber of 150 cm length and 80 cm diameter surrounding the target. It provides tracking, momentum measurement and particle identification of secondary particles emerging from the target and covers the angular region of approximately 20° to 160°.

The TPC is inserted in a solenoid magnet providing a field of 0.7 Tesla parallel to the TPC axis as shown in Fig. 23. The magnet (45 cm gap radius, 224 cm gap length, 88 coils, 910 A current, 0.72 MW power consumption) provides a homogeneous field with $Br/Bz \leq 1\%$ over a length of 1.60 m, thus covering the whole TPC.

Drift Volume & Field Cages

The inner and outer field cage (IFC/OFC) are made of a 2 mm and 8 mm thick Stesalit cylinders, respectively (Tab. 4). The electric field gradient is provided by interleaved strips of conductive Mylar foil at precise potentials and positions.

Most important in the design of the field cages is to avoid high field gradients which could cause sparks or corona currents in the gas. Moreover, in the ideal TPC the field is homogeneous so that the drift velocity is stable and free of distortions over the whole gas volume. To achieve this, during the design of the field cages the electric field setup was extensively simulated.

	IFC	*OFC*
Length	799 mm	1541 mm
Diameter	104 mm	812 mm
Thickness	2 mm	8 mm

Table 4: Field cages construction details.

The choice of the gas mixture (91% Ar, 9% CH_4 at atmospheric pressure) was mainly driven by the fact that its basic characteristics are described in the literature. Mostly Ar atoms are ionised providing a ionisation yield of approximately 50 electrons per cm (Fig. 24). The gas volume in the TPC is about 700 l and the flux less than 100 l/h.

The validity of the overall design was proven by the stable operation during 1.5 years. The transversal drift distortions are less than 100 μm at more than 0.6 mm distance from the strips and no effect on the drift time at more than 1.2 cm was observed.

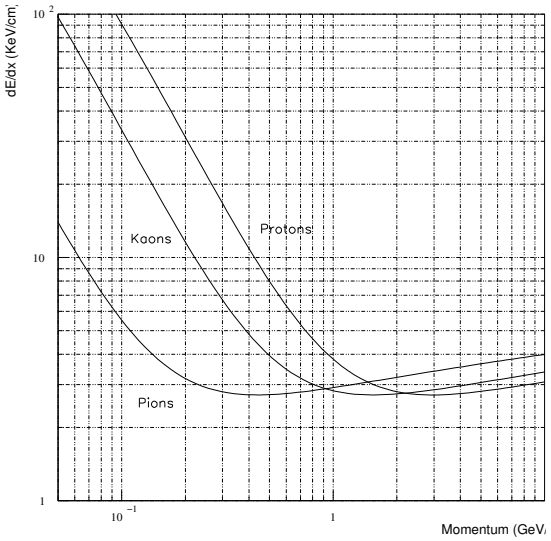


Figure 24: Energy loss of π, K, p in Ar at atmospheric pressure.

Details on design and field studies of the field cages can be found in [44].

Wire Chamber

The wire chamber at the upstream end of the TPC is made of the pad plane at 0 V, the anode wire plane at +1820 V and a cathode wire plane at 0 V, each at a distance of 5 mm. Each wire plane is made of a single 150 m long wire strung into a hexagonal frame in order to reduce dead areas due to support structures.

	Anode	*Cathode*	*Gate*
Material	W, Au	Cu, Be	Cu, Be
Thickness	20 μm	70 μm	70 μm
Pitch	4 mm	2 mm	4/2 mm
Length	150 m	150 m	-
Pad plane dist.	5 mm	10 mm	26 mm
Voltage	+1820 V	0 V	-67 V

Table 5: Technical details of the wire planes. The voltage given for the gating grid is for the open gate.

In addition, a gating grid is installed between the drift volume and the wire chamber. It is made of two interleaved single-wire planes and prevents ions to drift back into the TPC volume. The planes are

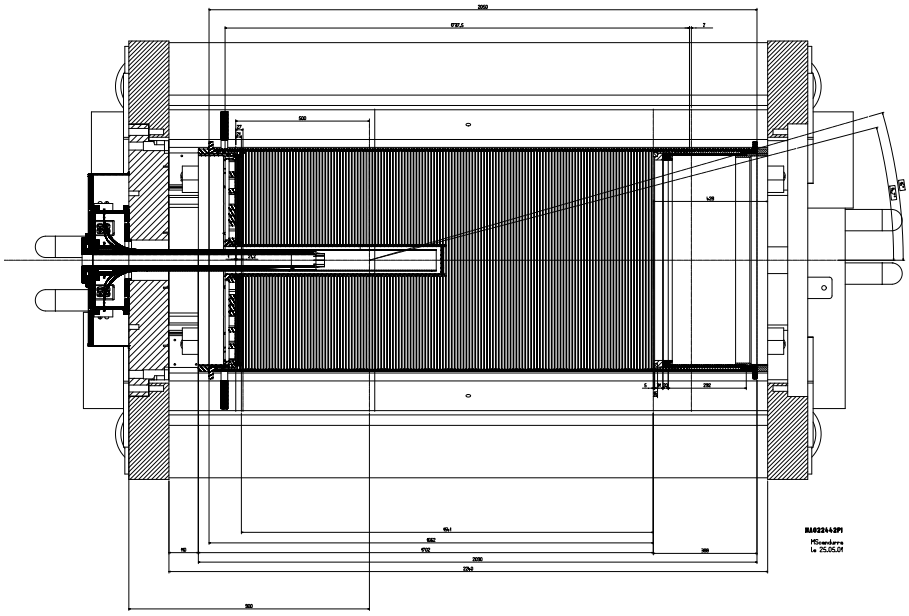


Figure 23: Cross-section of TPC and solenoid seen from the top. The beam enters from the left and travels through the ITC which is inserted into the inner field cage. The nominal position of the front face of the target is indicated by the starting point of two symbolic "tracks" (at angles 14.7° and 16°). It reaches 500 mm into the TPC volume. Target and target-holder are not shown.

at -67 V voltage in case of an open grid and at -35 V and +35 V when closed.

Technical details of the wire planes and the gating grid are given in Tab. 5.

Pad Plane

The pad plane is made of six identical sectors as shown in Fig. 25. A sector is a six-layer printed circuit board (PCB) with the pads on one side and electronics components on the other. The electronics serves for amplification, shaping and buffering of signals.

On one sector, the pads are arranged in 20 circular rows each containing between 11 and 55 pads. The pad size is not identical for all pads but is always approximately 6.5 × 15 mm^2 in ϕ and r, respectively. There are 662 pads per sector, hence 3972 pads for the whole TPC.

During data taking considerable cross-talk effects between pads were observed. These are mainly caused by the layout of the PCB: preamplifier output lines are in proximity of other pads. This represents a capacitive coupling between the output of one amplifier and the input of an other. Studies showed that about 50% of all pads cross-talk into other pads and the amplitude of the induced signals is generally less than 20% of the original signal. The effects of cross-talk might be corrected for in the reconstruction to a large extent.

Readout

The TPC readout is performed by ADC cards designed for the ALICE and CERES experiments [45]. One 9-unit VME module provides 10 MHz, 10 bit sampling for 48 channels. In addition, the cards perform real-time zero suppression and bunching of pulses with pre- and post-samples (see Fig. 26). The parameters determining the zero-suppression are loaded at the beginning of each run. Therefore, a few pedestal events[4] are taken, the values of the

[4] *Pedestal* signifies the noise level of the electronics as recorded by the ADC.

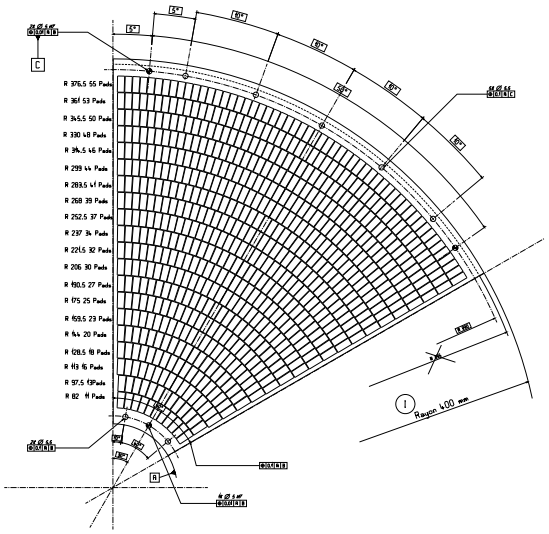


Figure 25: The pad layout on one sector. 662 pads are arranged in 20 rows containing between 11 pads (innermost row at 82 mm radius) and 55 pads (outermost row at 376.5 mm radius).

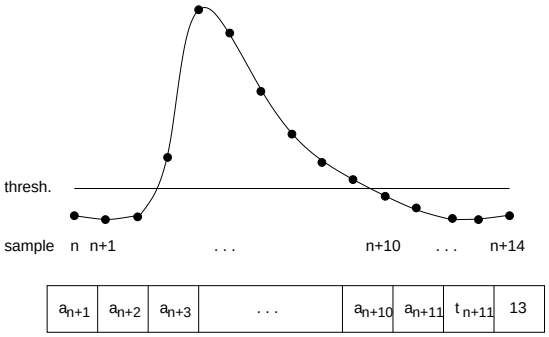


Figure 26: A typical pad signal: samples about threshold are merged with two pre- and two post-samples. For packing, the time (i.e. sample number) of the last sample and the length of the bunch are appended. Only bunches with two or more samples above threshold are accepted.

pedestal are calculated channel by channel and fed back into the ADC card. It is then used for the threshold. The data are stored bunch-wise in 10 bit words preceeded by a 32 bit header containing the pad identifier.

In total, 84 of these cards are distributed equally over six VME crates, each one containing 14 cards (thus 672 channels) for the readout of one sector (cf. Table 7). As can be seen in Fig. 35, the data volume and the readout time of the whole detector are dominated by these six TPC slave crates.

The sampling of the signals on all cards is started by the physics trigger. Then, 300 samples each of 100 ns are acquired to cover the total drift time of 30 μs.

Calibration

The gain of each pad has to be calibrated with good accuracy in order to allow the identification of a particle via its energy loss dE/dx and a precise track reconstruction. Since there is a priori no way to estimate the error on equalization constants, *two independent* methods are employed to determine the gain of each pad:

- Tracks of high energy cosmic muons are recorded in each inter-spill period as well as during periods without beam (machine development, downtime, repairs, etc.) They illuminate all pads in the same manner and allow the calculation of an equalization constant for each pad in order to normalize its gain.
- During longer shutdown periods, the TPC gas was enriched with radioactive ^{83m}Kr gas.[5] It mainly decays via γ-decay with a lifetime of 1.86 h. Most of the γ energy is transferred to electrons which then build a cluster with negligible size compared to the pad size. This cluster drifts to the pad plane and is recorded. The pulse-height distribution of each individual pad allows then to identify the typical peaks of the ^{83m}Kr decay.

In addition, a laser calibration system is employed to measure drift velocity[6] and eventual distortions of the magnetic or electric field. Therefore, 192 small Al tips connected via optical fibres to a KrF UV laser are installed in the HV membrane at the downstream end of the TPC gas volume. When these fibres are illuminated photo electrons will be released at a well known time and position. The fibre tips are equally distributed over the whole x/y plane. The drift velocity can then be calculated

[5] The ^{83m}Kr isotope is obtained by placing a ^{83}Rb source in the gas input line of the TPC.

[6] The drift velocity is very sensitive to composition, temperature and pressure of the TPC gas. Thus, it also serves as monitoring to detect problems during data taking and for long-term studies of the running conditions.

using the timing information. Deviations in the reconstructed position are used to determine field distortions.

After the end of datataking the TPC and its readout equipment were held operational to perform additional calibration measurements. In particular, the electronics gain of all pads and all crosstalk signals are measured. It is done by bringing a probe in direct contact with the pad and injecting a known amount of charge and reading out the response of all pads of the sector.

RPCs

At large production angles, the particle identification mainly relies on the dE/dx measurement of the TPC. However, in the momentum range 150 to 250 MeV/c electrons and pions can not be distinguished via their energy loss. Instead, a time-of-flight measurement is performed: the TPC is surrounded by 30 Resistive Plate Chambers (RPCs) providing a time resolution of better than 200 ps ("Barrel-RPCs"). With a minimal flight path of 45 cm this is sufficient to achieve the desired separation (see Tab. 6).

p [GeV/c]	$L = 0.45$ m	$L = 1.0$ m
0.10	1.08 ns	2.39 ns
0.15	0.55 ns	1.22 ns
0.20	0.33 ns	0.73 ns
0.25	0.22 ns	0.48 ns

Table 6: Time-of-flight difference between electrons and pions in the range 100 to 250 MeV/c for two typical flight path lengths L.

An RPC module is of 192 cm length, 10.6 cm width and 1.1 cm thickness. It consists of 4 gaps of 300 μm each, between 1 mm thick high-resistivity glass plates. In the longitudinal direction it is segmented into 8 × 8 pads each of size 104 × 29 mm^2. Each pad has a dedicated preamplifier, but after amplification the signals of 8 pads are added before they are recorded by the data acquisition (see Fig. 27).

The Barrel-RPCs are squeezed in 24 mm radial space between the outer field cage of the TPC and the heat shield of the solenoid and provide a double coverage of parts of the surface as shown in Fig. 28).

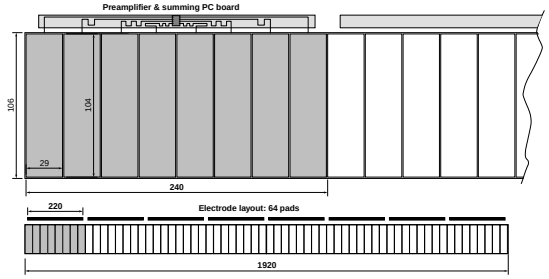


Figure 27: Pad layout of an RPC module with 64 pads. The signals of each 8 pads are summed so that a hit can only be assigned to an area of 240 × 106 mm^2.

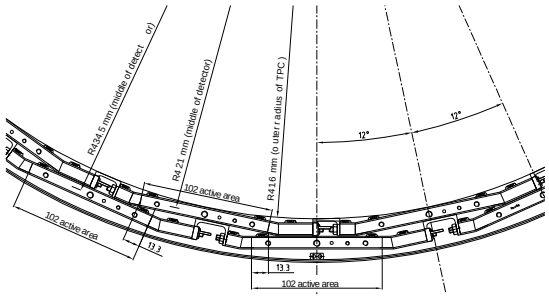


Figure 28: Barrel RPC detail. Note the partial double-coverage which is realized using the 24 mm radial space between TPC and heat shield of the solenoid.

16 additional chambers are installed at the downstream exit of the TPC and cover the forward direction outside the acceptance of the TOF wall ("Forward-RPCs").

Thus, with 368 RPC readout channels a sensitive area of about 8 m^2 is covered. The final time resolution was shown to be at the level of 150 ps with an efficiency of well above 95%.

The Barrel-RPCs were also used in the trigger as large angle trigger redundant to the ITC and as cosmics trigger (see section 3.2.5).

2.2.3 Small-Angle Detectors

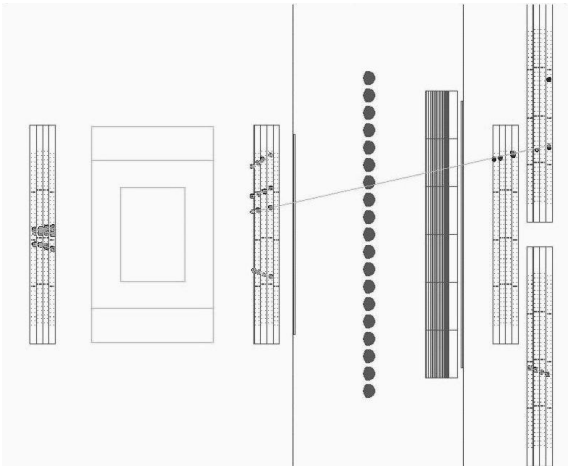


Figure 29: Event display with the main small-angle detectors. With (from left to right) a first drift chamber, followed by the dipole magnet and another drift chamber module for momentum determination. Further downstream, a threshold Cherenkov detector for particle identification, a set of drift chamber modules to facilitate track reconstruction. (cf. Fig. 13).

Drift Chambers

Tracking in the forward direction is performed by a set of drift chambers recuperated from the NOMAD experiment. The Nomad Drift Chambers (NDCs) cover an area of 3×3 m^2 and provided a spatial resolution of 150 μm at 97% efficiency in NOMAD [46].

For tracking, 20 chambers are used. They are arranged in five modules (the 5 upstream modules shown in Fig.13) of which each contains four chambers. Each chamber contains three wire planes: one with vertical oriented wires (parallel to the HARP y-axis), the other ones with the wires rotated by $+5°$ and $-5°$. There are 3×42 wires in one chamber. The gas gap is 8 mm and the maximum drift width 32 mm.

Three additional modules each with only one active chamber serve to define the position of clusters in the electromagnetic calorimeter. They are the most downstream modules between TOFW and electron-identifier as shown in Fig. 13.

For safety reasons, the flammable NOMAD gas (40% argon, 60% ethane) could not be used in HARP. Instead, the mixture was changed to 90% Ar, 9% CO2, 1% methane CH4. Unfortunately this caused (not unexpectedly) a reduced efficiency of the chambers.

Obviously, the momentum determination relies mostly on the NDC modules surrounding the dipole magnet[7]. Since its field is vertical (thus a charged particle is deviated in the horizontal plane), momentum determination requires a precise horizontal coordinate from the track reconstruction. As this is easier to achieve with vertical wires, the wire orientation was chosen correspondingly.

For an efficient and accurate track reconstruction it is mandatory that the wires of all chambers are very well aligned. This is done in a special procedure using data from non-interacting beam particles and nearly horizontal cosmic muons.

Cherenkov Counter

A Cherenkov counter is installed downstream of the second NDC module to provide discrimination between pions and protons at high momentum. It is filled with C_4F_{10} (perfluorobutane) gas[8] at atmospheric pressure and operated in threshold mode. The thresholds of C_4F_{10} are shown in Fig. 30.

The Cherenkov light is reflected by mirrors onto 38 8-inch photomultiplier tubes recuperated from the CHOOZ experiment. They are surrounded by magnetic shielding; additional focusing is provided by winston cones which are installed in front of their windows.

The layout and size of the detector is displayed in Fig. 31[9]

TOF Wall

An array of 3×13 scintillators is set up at a distance of ≈ 10 m from the target. It performs particle identification at low momenta complementary to the Cherenkov (mainly pion/proton separation).

[7] Technical data of the dipole magnet: gap height: 88 cm, gap width: 241 cm, gap depth: 171 cm, current: 2.9 kA, power: 0.36 kW, homogeneous vertical field: 0.5 T

[8] C_4F_{10} is not toxic and not flammable. It is 8 times heavier than air and very expensive. The costs of 30 kCHF for a complete fill of the detector volume of 31.4 m^2 made it mandatory to recuperate the gas.

[9] After construction it turned out that it is not trivial to build a gas-tight vessel of 31.4 m^3 volume. [47]

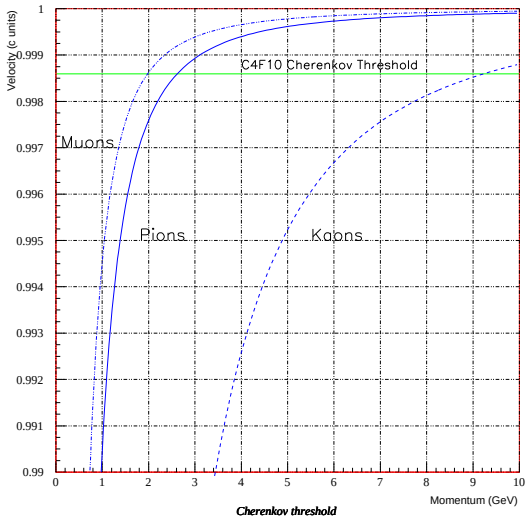


Figure 30: Cherenkov thresholds of C_4F_{10} at atmospheric pressure: μ: 2.0 GeV/c, π: 2.6 GeV/c, K: 9.3 GeV/c, p: 17.6 GeV/c. γ_{thresh} is 18.8.

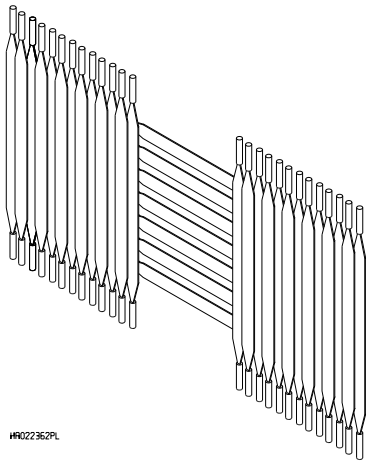


Figure 32: Schematic view of the TOF wall. The width of the scintillator slabs is 21 cm, their thickness 2.5 cm and their length 180 cm and 250 cm for the horizontal and vertical slabs, respectively. All scintillators are installed with a small geometrical overlap in order to avoid dead areas.

The setup is shown in Fig. 32. The scintillator slabs are 2.5 cm thick and equipped with double-sided readout in order to achieve a good time resolution of ≤ 200 ps. The sensitive area is approximately 7.4 m wide and 2.5 m high.

e- and μ-identifier

The electron identifier (Fig. 33) is intended to discriminate electrons from pions complementary to the Cherenkov counter. Moreover, below the pion threshold the Cherenkov might tag pions as electrons when they are accompanied by knock-on electrons.

The calorimeter modules (lead/scintillating fibres) have been recuperated from CHORUS: 62 "em" modules (5 X_0, total width 4.96 m, height 2.62 m) and 80 "had" modules (11 X_0, total width 6.4 m, height 3.35 m).

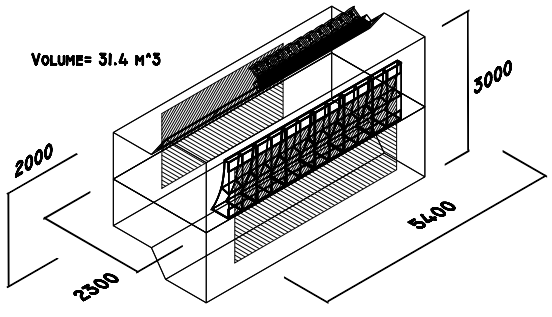


Figure 31: Schematic view of the Cherenkov detector. Only the instrumentation (mirrors and PMs) of the upper half is shown. See also its position in the complete HARP detector in Fig. 13.

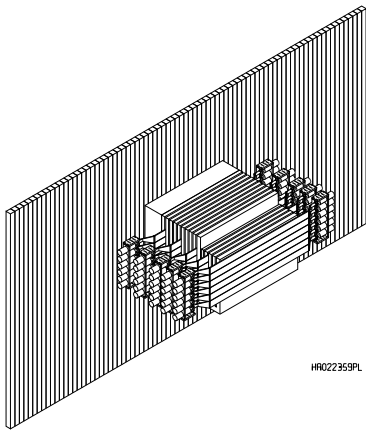


Figure 33: Schematic view of e- and μ-identifier.

The muon identifier (also shown in Fig. 33) is the most upstream part of the HARP detector. It serves to identify beam muons and consists of a set

of 140 cm wide iron plates interleaved with scintillator slabs. The thickness of the iron represents 6.44 λ_I. In front of the beam-muon identifier, a passive 32 cm layer of iron is placed.

Beam muons have to be identified since they would lead to a wrong cross-section if accounted as pions. At high momentum, the discrimination can be done by requiring non-showering in the muon identifier. At low momentum, muons can be tagged by their generally lower momentum (since they dominantly origin from pion decay in the beam line). Therefore, they are bent stronger in the magnetic dipole field.

3 The HARP Trigger System

The design of the trigger system is mainly based on the following requirements: for physics triggers the detection of an incoming beam particle and full solid angle acceptance for the detection of secondaries generated by that particle. In addition, calibration triggers have to be acquired between spills or at the beginning of each run. For all triggers the synchronous triggering of all DAQ readout equipment including the generation of ADC gates, etc. is performed. Moreover, the synchronisation with beam spills delivered to the T9 beam line is necessary. Features like high reliability, redundancy, efficiency, stability and speed complete the demands to the trigger.

To understand the full functionality of the trigger it is first necessary to review the HARP DAQ system. The various trigger detectors in the beamline and around the target which build the basis of trigger decisions are described in section 2.2.1.

3.1 DAQ

All readout of data in HARP is realized using PCs running Linux connected to VME crates housing standard or custom-made ADC and TDC modules. As DAQ software the ALICE DAQ prototype DATE [48] is used. In total, 16 crates with about 200 modules and $\approx$ 10000 channels are read out per event. The typical event size amounts to 30 kB and the readout time to about 500 μs per event.

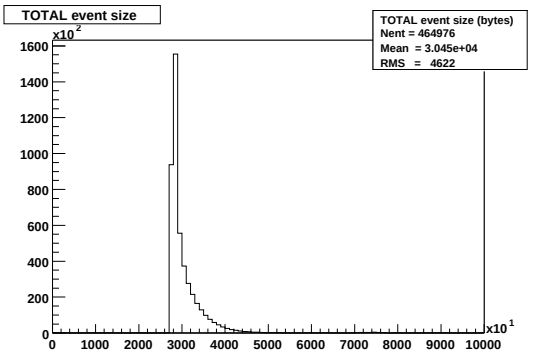


Figure 34: The event size distribution for complete events (run 18709, -3 GeV/c beam, 5% C target, accumulation of about 4000 spills). See also Fig. 35.

3.1.1 DATE

DATE performs the readout of all crates and the event building. In the nomenclature of DATE each VME crate equipped with (or connected to[10]) a PC running a DATE client is considered as LDC (Local Data Collector). Each LDC is required to have a triggering VME module at its disposal whose registers are permanently read out (polled) to detect the arrival of a trigger. Once this has happened, all modules in this crate are sequentially read out. The data is then transferred via the VME bus to the memory of the PC. In its function as LDC, the PC sends the data via TCP/IP network to the event builder(s), called GDC (Global Data Collector) in a DATE environment. The GDC waits until having received the data for a particular event from *all* LDCs and assembles the subevents to a consistent and complete event for the whole detector assuring this way the integrity of the data. On a run-by-run basis, the output of the GDC is written to a file which is the input for the next stage of data handling, the "objectification". Here, this data is merged with slow control data and other setting related data (like target, beam momentum, etc.) and fed into an object-oriented database system interfaced to the CASTOR mass storage facility provided by CERN-IT.

3.1.2 Triggering LDCs with CIRQs

A CIRQ interrupt request module [49] designed for and formerly used in the trigger of the CHORUS experiment [50] performs the triggering, gate and busy generation for a single crate. Each signal arriving in one of the six independent input channels generates an individual busy signal and, in addition, a common busy may be generated which can veto new input signals. The exact behaviour of the CIRQ is determined by the settings of a control register; especially important is the possibility to dis- and enable single channels.

For all HARP LDCs the configuration is the following:

- channel 3: physics trigger, always enabled
- channel 4: pedestal trigger, only enabled at the beginning of a run

[10] Except for one crate (which is interfaced to a standard PC box with a VME-PCI interface) all crates in HARP are equipped with a single-board VME PC.

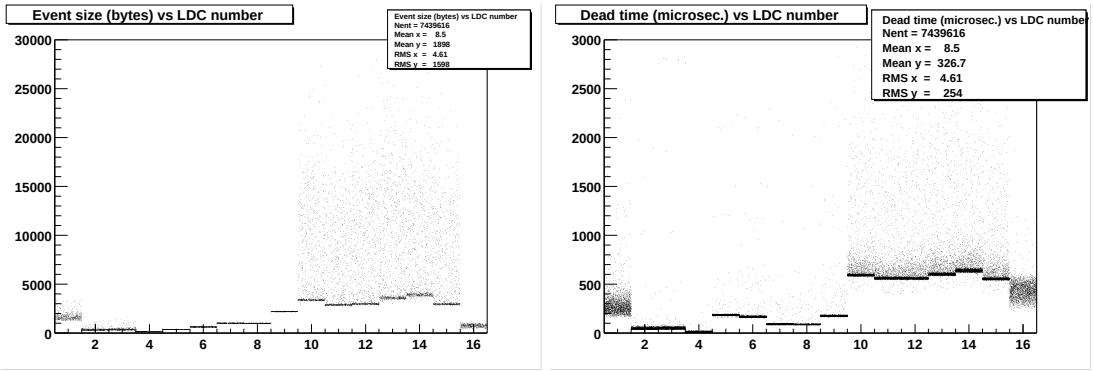


Figure 35: Event size and readout time per LDC. It can be seen that the six TPC slave crates produce the highest amount of data and show the largest fluctuations. Of course, the readout times are correlated to the event sizes (run 18709, -3 GeV/c, 5% C target, accumulation of about 4000 spills).

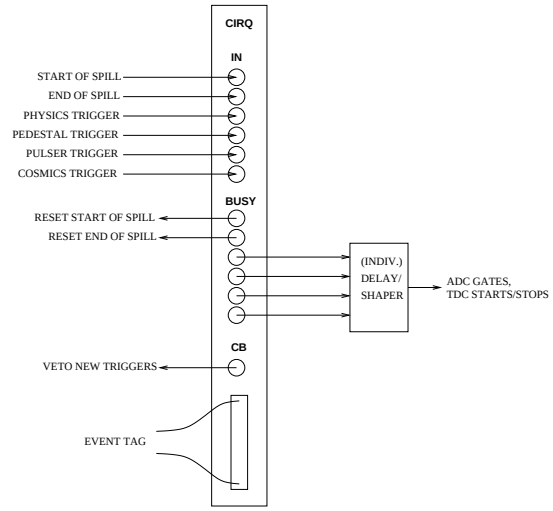


Figure 36: Schematic cabling/configuration of the CIRQ for all LDCs. The start/end of spill signals are produced by a Flip-Flop which is reset by the corresponding busy. At the same time as the trigger a centrally distributed 16 bit event tag arrives at all CIRQs and is latched into an internal register. It is read out and included in the event data. This allows additional event consistency checks.

- channel 5: pulser trigger, enabled only between spills and after cosmics triggers have been taken
- channel 6: cosmics trigger, enabled only between spills

Channels 1 and 2 are only used for the hardware-software communication to make the LDC aware of the beginning and the end of a spill, respectively. These signals are fed through a Flip-Flop (the arriving signal *sets* it whereas the busy *resets* it) making it impossible to miss them. This has to be done because they may arrive at any time and unlike trigger signals they cannot be vetoed by a busy logic.

All busys which are generated have to be cleared via software, which is done at the end of the readout in case of the triggers or after having acknowledged the beginning or end of the spill.

The advantages of this way of triggering are evident: the readout software is aware of *which type* of trigger arrived and can flag the data accordingly. This is in particularly important for the event builder, since only physics triggers are build for the whole detector, whereas cosmics triggers are build for a set of subdetectors and pulser/pedestal triggers are not build at all. Moreover, the individual busys allow to apply different delays and shapers for the gates of different triggers.

In the standard (global) configuration running the whole detector, all LDCs are receiving the inputs for the CIRQ (except for pulser trigger) from a centralised point and report back their common

busy (see Fig. 41). However, it is possible to run an LDC in standalone mode providing the trigger signals with a local trigger logics. This can be used for debugging and to take calibration data when no global run is active.

3.1.3 Readout of VME Modules

DATE has to be provided with a dedicated routine performing the readout for each module used. "Reading out" a module typically means to poll its status register for some kind of "data ready" or "conversion finished" bit which implies the earlier arrival of an ADC gate, TDC start/stop, etc. The data is then ready to be read via the VME bus. The modules used in HARP (see Table 7) are mostly standard 6-unit VME modules manufactured by LeCroy or CAEN. Exception are the MWPCs where BPUs (Buffered Pattern Units) [51] formerly used in the CHORUS experiment and the TPC where custom-made 9-unit ADC modules [45] designed for the ALICE experiment perform the readout.

LDC	detector		modules	channels
16	central trigger	3	CAEN V560 scaler	64
		1	LeCroy 1176 TDC	16
		2	PLU	64
1	local trigger	6	LeCroy 1176 TDC	96
		2	CAEN V775 TDC 35 *ps*	64
		6	CAEN V792 ADC	192
		1	CAEN V767 TDC	128
2	NDC(b)	12	CAEN V767 TDC	1536
3	NDC(t)	11	CAEN V767 TDC	1408
4	TPC master	1	CAEN V767 TDC	128
		1	CAEN V560 scaler	16
5	MWPC	11	BPU buffered pattern unit	704
6	TOFW	3	CAEN V775 TDC 35 *ps*	96
		3	CAEN V792 ADC	96
7	RPC(b)	6	CAEN V775 TDC 35 *ps*	192
		6	CAEN V792 ADC	192
8	RPC(t)	6	CAEN V775 TDC 35 *ps*	192
		6	CAEN V792 ADC	192
9	CALO	15	CAEN V792 ADC	480
		1	CAEN V767 TDC	128
10-15	TPC slaves	84	FEDC ADC	4032
16		187		10007

Table 7: VME readout modules used in HARP and their distribution to LDCs. The local trigger crate contains the readout equipment of the trigger detectors, all detectors in the beamline (except the MWPCs) and the Cherenkov.

3.2 Trigger Logic

3.2.1 Central Trigger

In the context of the trigger, the VME crates (LDCs) can be arranged in three logical levels, as shown in Fig. 37. In the global running mode with the complete detector, the distribution of physics trigger signals, the veto/busy logics and spill gates/signals are exclusively handled by the central trigger. The only exception to this rule are calibration triggers, which can be taken by each system between spills. The maximum number of those triggers is configured in the DAQ and the central trigger provides an interspill gate during which they are accepted.

Due to the large data volume and the high number of channels to be read out, the TPC readout is equally distributed to six crates (LDCs 10-15), each of them handling the data of one TPC sector. Those slave crates are driven by the master crate in the same way as all other crates are triggered by the central trigger.[11]

3.2.2 Spill Synchronisation

The central trigger generates several gates in order to assure that the spill and interspill period are clearly separated. The nominal duration of a T9 spill is 400 ms; the minimum delay between the beginning of two consecutive spills about 2 seconds.

The gate during which physics triggers are accepted is referred to as *ps-gate*, whereas the interspill gate for calibration triggers is called *cosmics-gate*. Most important in this context is that there is some margin between those two gates which will allow the DAQ to finish the readout of an accepted calibration/physics event before triggering the next physics/calibration event, respectively.

In addition, all LDCs are informed about the beginning of a spill (B-SPL) and its end (E-SPL). B-SPL and E-SPL are stored as special events.

The typical time structure of a spill and the various gates are displayed in Figs. 38 and 39, respectively.

[11] It should be mentioned that the slave crates do not contain a CIRQ but a rather similar module (an I/O register SIS3600 [52]) which basically performs the same tasks.

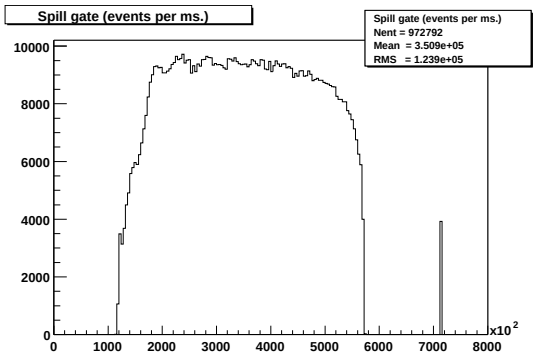


Figure 38: Time of the arrival of triggers as registered by the DAQ PCs (accumulated for about 4000 spills). Reference time (i.e. zero) is the begin-spill signal; the peak at $\approx 700\ ms$ is due to the end-spill events. See also the timing diagrams in Fig. 39. The spill length can be estimated to about 450 ms (run 18709, -3 GeV/c, 5% C target).

3.2.3 Physics Triggers

A beam particle having caused an interaction in the target is considered as *physics trigger*.

The PLU

Central part of the physics trigger decision is a programmable logic unit (PLU) [53] situated in the central trigger crate. Like the CIRQ, the PLU has been designed for the trigger of the CHORUS experiment [50]. When a strobe is received, 16 input bits are latched and according to an internal lookup table a pattern of 16 output bits is generated.

Input signals are summary logic signals of all detectors which are of importance for the trigger: TOFA, TOFB, BS, TDS, HALOA, HALOB, BCA, BCB as beam detectors and FTP, ITC, Cherenkov as interaction detectors.

Out of the 16 output signals 10 are used as physics trigger signals (as fan-out to the 10 LDCs) and one as *no trigger* signal to reset the trigger system (Fig. 41). Five output bits are set whenever special combinations of input signals occur and are fed into a scaler (Fig. 40).

Trigger Setting

Performing the trigger decision with a PLU allows a maximum flexibility for the trigger configuration: the lookup table is loaded at the beginning of each run and may therefore be adapted to different con-

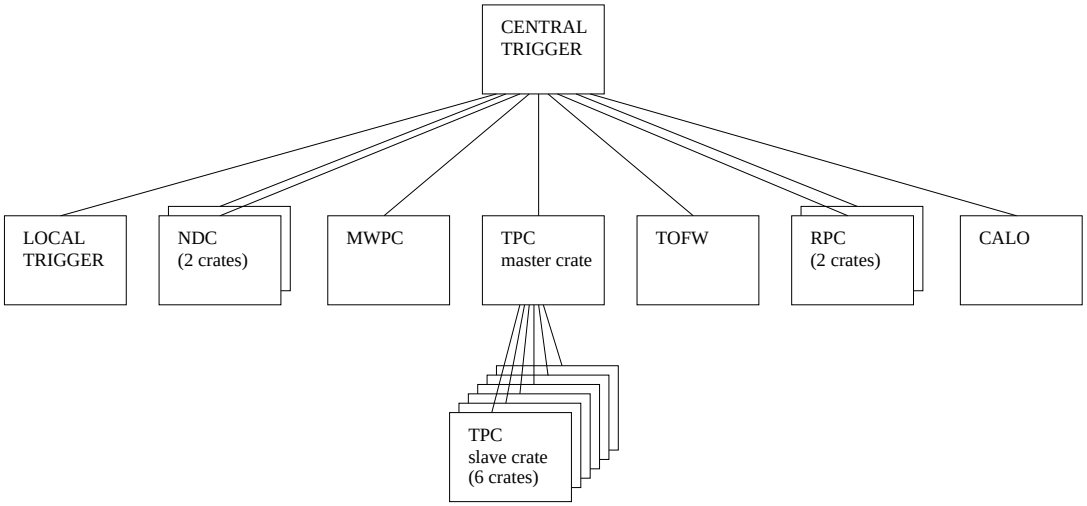


Figure 37: The central trigger (LDC 16) performs the global trigger decision and distributes the information to the CIRQs of the LDC of the second level (LDCs 1-9). The TPC master crate serves the six slave crates (LDCs 10-15) with trigger signals and communicates their busys back to the central trigger.

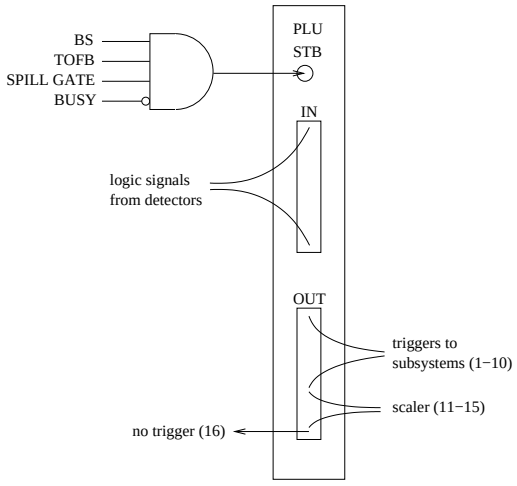


Figure 40: Simplified operation scheme of the PLU. Each time a coincidence of BS and TOFB occurs and the system is not busy a strobe is generated which causes a trigger decision to be done on the basis of the latched input signals from various detectors and the lookup table stored in the module. In case of a trigger condition, trigger signals are directly fanned out to the 10 subsystems.

ditions like beam momentum, beam composition and target. In addition, the timing for the trigger signals does not depend on the complexity of the decision like it would do for binary logics realized with standard NIM electronics only.

To generate the lookup tables[12] the special program *PluPatternGenerator* which is part of the HARP software is used. It produces a text file with all 2^{16} possible input combinations and their corresponding output words and assures reliability, flexibility and versioning of the generated patterns. Each pattern has a unique name and version number which is written into the data as part of the start-of-run event as well as the 2^{16} times 16 bits themselves.[13]

Since the trigger decision is crucial for the integrity of the experiment's data, its correctness has to be assured at any time. Thus, not only the input and output pattern are stored for each event but the decision is also checked with a copy of the complete lookup table in the memory of the DAQ PC. In case of a wrong decision the run will be

[12] In HARP nomenclature also referred to as "trigger pattern", "trigger setting" and "trigger file".

[13] The PluPatternGenerator also generates human-readable and html documentation for each pattern. The documentation has to be consistent with the actual patterns since it was used by the shifter to chose the correct setting during data taking.

immediately stopped. Another source of problems might be a false *loading* of the lookup table at the beginning of the run. Therfore, the DAQ initialisation software *reads back* the table from the PLU after loading and compares it with the original one.

A standard physics trigger condition for thin targets is $BS \times TOFA \times TOFB \times (NOT)HALOA \times (NOT)HALOB \times TDS \times (ITC + FTP)$. For thick targets a signal in FTP or ITC is not required since the probability for an interaction is almost 100%. Triggering on each beam particle also has the benefit of simplifying the normalisation. For beam configurations with extremely high electron content, a beam Cherenkov counter was used to veto these events on trigger level.

PLU Signal Timing
The strobe signal for the PLU is the lowest level indication of an incoming beam particle; it has been chosen as a coincidence of BS and TOFB.[14] Technically this coincidence is implemented in such a way that the timing of BS is preserved and all 16 signals entering the PLU to perform the decision are timed with respect to BS. To monitor and control this very important timing issue, a TDC is employed to measure the leading and trailing edges of the first 15 input signals and the strobe itself. Like this it can always be assured (especially in the offline analysis) that the signal levels of the inputs are well established before the strobe arrives.

Trigger Cycle
Of course, $BS \times TOFB$ coincidences can only strobe the PLU when the detector is not busy reading out an event and the PLU is not already occupied with a decision. This is guaranteed by a busy logic which vetoes these strobes. In addition it takes care that after the readout is finished *or* the event is rejected the busy is released. All strobes also have to pass the spill gate.

Implementation details can be learned from the electronics plans in Appendix A, measured timing diagrams are shown Fig. 42.

Scaler PLU
In addition to the PLU which performs the physics trigger decision ("trigger-PLU"), a second PLU ("scaler-PLU") is employed which receives the same input signals and strobe as the trigger-PLU. However, its strobe is *not* vetoed by a busy. The 16 outputs of this PLU represent 16 fixed logic combinations of input signals and are fed into a scaler. The output of this "deadtime-less" PLU can be used to determine how many triggers were missed during the busy period. Moreover, an eventual bias of events which are read out with respect to all theoretical triggers could be found.

[14] Due to the large upstream distance of TOFA (and therefore long cable lengths) this signal arrives too late to be incorporated into the strobe.

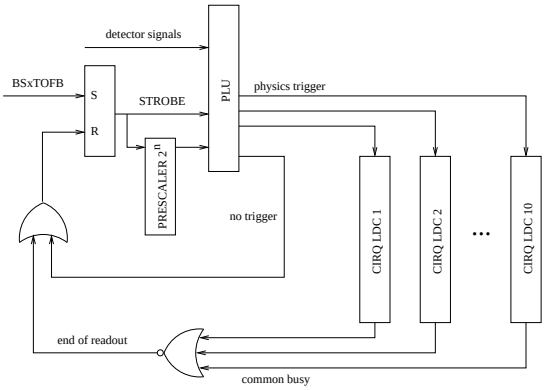


Figure 41: A physics trigger decision cycle commences with a $BS \times TOFB$ coincidence which sets the strobe flip-flop. Then, on the basis of its lookup table, the PLU performs the physics trigger decision taking into account the signals from the detectors. In case of a trigger it is sent to the CIRQs in the LDCs and the STROBE is reset after the readout of the event, when all CIRQs have released their busy. A dedicated "no trigger" signal resets the STROBE for an event not satisfying the physics trigger criteria.

3.2.4 Normalisation Triggers

For the precision requirements on the normalisation in HARP it is mandatory that a downscaled sample of minimum-biased data is recorded. Obviously, the least biased trigger condition is the $BS \times TOFB$ coincidence. Thus, a certain fraction of all strobes reaching the PLU is readout, no matter whether they caused an interaction or not. Technically this is realized by using a prescaler to generate a *downstrobe* signal for each i'th strobe. This signal is fed into one of the PLU inputs and forces a trigger. Since all input bits of the PLU are recorded it is straightforward to select those downscaled events in the analysis.

The downscale value i is adjusted in such a way that the downstrobe events do not substantially reduce the number of physics triggers. At the same time it has to be assured that enough of them are acquired to meet the requirements on the statistical error for the normalisation.[15]

3.2.5 Calibration Triggers

Apart from physics and normalisation triggers provided by the PLU the following calibration events have to be taken: pedestal events, pulsed events and events induced by cosmic muons. Common to these events is that they are taken separately by a single subsystem or a set of subsystems. The number of such triggers is determined by the configuration of the DAQ and the central trigger opens an interspill gate to allow their acquisition.

Pedestal Events

Pedestal events are taken once at the beginning of a run. They are triggered by a clock (about 1 kHz frequency) and are forced to be out of spill. These data are used to monitor the pedestal of all ADC channels as well as calculate its value and feed it back into the module for an online zero-suppression. Pedestal events are not built and each LDC can take an individual number of them (or none at all).

Pulser Events

Systems taking pulser triggers are the TOF wall (monitoring of timing stability with a laser system) and the local trigger crate (monitoring and calibration of the gain of the PMTs of the Cherenkov with a LED and laser flashing system). At the same time as the respective CIRQs are triggered, a signal is sent to the laser/LED flashing system.

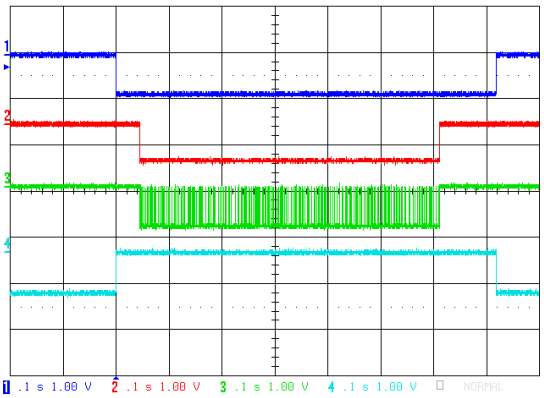
Cosmics Events

Cosmic muon events are taken in the interspill periods and are used for calibration and alignment. Those events are built by the eventbuilders to allow crosschecks, efficiency determination and cross-alignment of the concerned detectors. Here, the HARP detector is divided into two subsystems which take independently data via two separate trigger- and busy logics (see Appendix A):

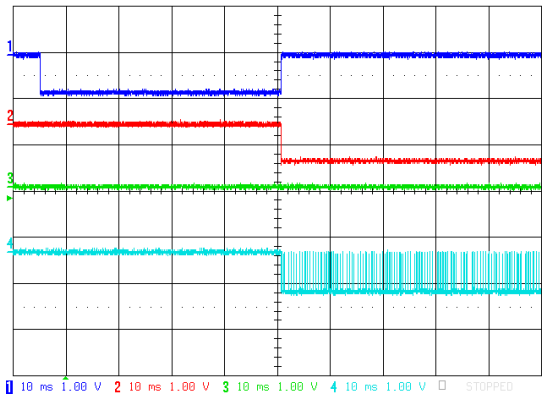
- TPC cosmics trigger (also referred to as "upstream" cosmics trigger): TPC and RPCs are read out. The triggering signal is provided by a coincidence of signals in two opposite RPC modules. Those straight tracks are used for calibration of TPC pads and testing the TPC clustering and track finding/reconstruction algorithms.
- NDC cosmics trigger ("downstream" cosmics trigger): NDCs and calorimeter are read out. Triggering signal is a coincidence of FTP, TOF wall and a dedicated large scintillator wall installed downstream of the last NDC modules[16] (Fig. 13. It is aimed to trigger on nearly horizontal high energy muons which provide means to align the NDC wires as well as monitoring and calibration of the e- and μ-identifier.

[15] Typical values are $i = 32$ and $i = 64$.

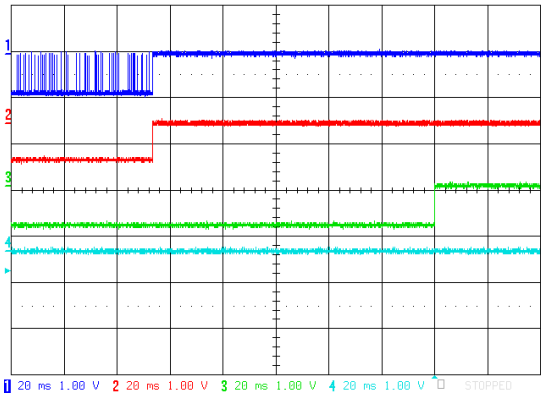
[16] The *cosmics wall* is an array of 32 scintillators each of 320×20 cm^2.



(a) A whole spill cycle: *ps-long(1)* defines the spill period. Shorter, with some margin to finish the acquisition of events, *ps-gate(2)* is the gate in which *strobes(3)* are accepted. Thus, the *cosmics-gate(4)* during which calibration triggers are accepted is the inverse of *ps-long(1)*. Scale: x: 100 ms, y: 1.0 V.

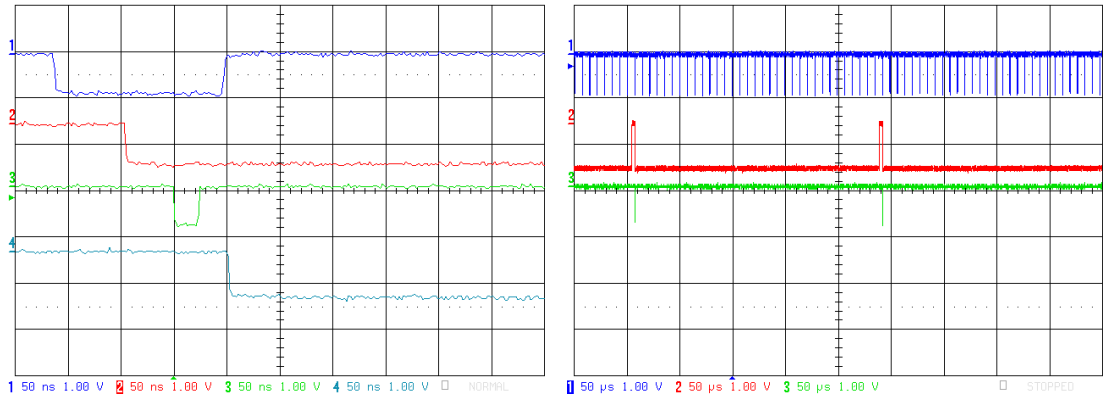


(b) Beginning of a spill: *ps-delay(1)* covers the margin between the beginning of ps-long (not shown) and *ps-gate(2)*. During this period no triggers at all can be taken. The pulse signalling the beginning of the spill (*b-spl*(3)) to all subsystems can not be seen in the screenshot since it is too short. Its position is indicated by the little triangle at the bottom of the grid (at 10 ms). Finally, *strobes(4)* are accepted during *ps-gate(2)*. Scale: x: 10 ms, y: 1.0 V.



(c) End of a spill: *strobes(1)* are accepted until *ps-gate(2)* ends. At the end of *ps-long(3)* (coinciding with *e-spl(4)* as indicated by the triangle at the bottom at 160 ms) calibration triggers can be taken again. Scale: x: 20 ms, y: 1.0 V.

Figure 39: Timing of spill synchronisation signals. The screenshots were taken at the end of data taking with a digital 2.5 Gs/s LeCroy oscilloscope and a faked $BS \times TOFB$ signal. The trigger setting was chosen in such a way that all strobes arriving at the PLU generate a physics trigger. Description of the *signals* from top to bottom. The naming of the signals is the same as in the electronics plans in Appendix A.



(a) A $BS \times TOFB$ *signal(1)* generates a *strobe(2)* for the PLU. In case of a positive trigger decision a *physics trigger(3)* and consequently a *busy signal level(4)* is established. Scale: x: 50 ns, y: 1.0 V.

(b) $BS \times TOFB$ *signals(1)* arrive permanently at the beginning of the trigger decision chain. Only when the system is not *busy(2)*, a *physics trigger(3)* can be generated, causing a new *busy(2)*. From this plot, the readout time can be estimated to ≈ 0.23 ms (empty event). Scale: x: 0.05 ms, y: 1.0 V.

Figure 42: Timing of strobe, trigger and busy signals. The screenshots were taken at the end of data taking with a digital 2.5 Gs/s LeCroy oscilloscope and a faked $BS \times TOFB$ signal. The trigger setting was chosen in such a way that all strobes arriving at the PLU generate a physics trigger. Description of the *signals* from top to bottom.

4 The TPC Simulation

The TPC is the most complicated detector in HARP. To understand its data and to validate the reconstruction a detailed Monte Carlo (MC) simulation has been developed.

Before describing the underlying technical and physical models of the TPC Monte Carlo (MC) and their implementation, the purpose of creating a fully featured simulation will be explained.

4.1 Purpose of the MC and Design Philosophy

Usually there are three main usages of MC simulations, corresponding to the typical three phases of an experiment:

In the design phase of a detector it is often not trivial to determine the optimal values for certain technical quantities. Typically, these are parameters such as geometry, magnetic and electric fields, materials, granularity, etc. Is is then useful to simulate the behaviour of the detector and to study its physics performance as a function of these parameters and to find a suitable set of parameters for the desired physics performance. Here the physics performance compromises efficiency, resolution, background, etc.

After comparing first experimental data with the MC prediction one might find (bigger or smaller) differences and the MC for the perfect (*'up-to-specs'*) detector does not reproduce the data satisfactorily. At this point all problems must be understood and their sources need to be traced back. Typically, effects like crosstalk in electronics, misalignments, field distortions, etc. were not foreseen and can be the cause for such differences. For each of these a model has to be found to describe how the data is affected. It can then be implemented in the MC and its impact on the physics performance can be studied. Moreover, any correction algorithm in the reconstruction can then be benchmarked against the MC.

Once the MC reproduces the data an obviously important application is to study the resolution and acceptance of the detector and the expected background in the measurements.

4.2 Implementation of the TPC MC According to Specifications

The basic simulation of the HARP TPC is done in two steps: The *physics processes* reproduce the generation of drift electrons and their signals in the wire chamber of the TPC. The *electronics part* performs the sampling, digitisation and packing of the data.

At the end of the simulation chain the data is available in the same format as real physics data and is unpacked and reconstructed with the same software.

Technically the simulation is realized by using the GEANT4 (G4) toolkit within the GAUDI framework. A set of GAUDI algorithms create and act on various objects in the event model stored in the GAUDI transient store.

The Physics Processes

For all charged particles in the TPC gas volume which have a non-zero energy deposit, drift electrons are generated and tracked to the cathode wire plane. Then, the avalanche to the nearest wire is reproduced and its charge distributed to the pads.

Result is a charge time series for each pad which received at least one charge deposit.

4.2.1 Generation of Drift Electrons

The generation of drift electrons is done on a step-by-step basis, where a *step* is a G4 step. The total energy deposit E_{dep} for this step is retrieved and the number of drift electrons n_e calculated with

$$n_e = \frac{E_{dep}}{E_{pair}} \tag{1}$$

with E_{pair} the energy required to produce an electron-ion pair (see Fig.43). It is an average value taking into account primary and secondary ionisation.

Single drift electrons can be created separately. In this case they are distributed one by one randomly over the length of the step.

However, since the large number of electrons can cause technical problems (in particular running time), they can also be grouped to one charge cluster *per step*. The step size can be controlled by G4 and a step size of 2 mm was found to be a

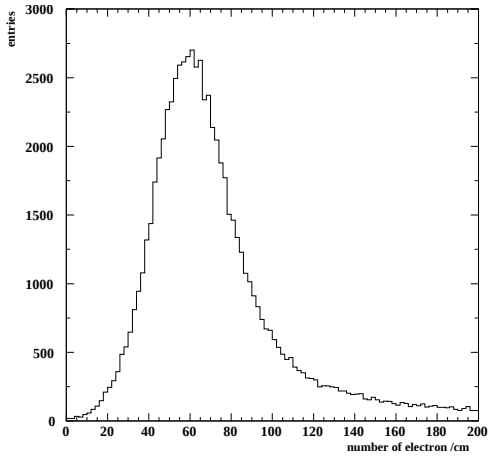


Figure 43: Distribution of the number of drift electrons produced for a 1 cm step [58].

reasonable and appropriate value. The position of these charge clusters on the step length is Gaussian distributed around the centre of the step with a $\sigma = stepLength/6$. A position which is not contained in the step is rejected.

Drift electrons and charge clusters will be treated in the same way in the following chapters.

4.2.2 Tracking of Drift Electrons

First, electrons are drifted to the cathode wire plane over a driftlength of $L = \Delta z$. The new position of the electron is calculated by adding a Gaussian smearing in both transversal and longitudinal direction. The amount of this diffusion depends on the gas choice and the strength of the magnetic field which especially suppresses the transversal component. In particular, a low transversal diffusion is essential for a good space resolution of the TPC.

The transverse diffusions Δx and Δy are random values following a Gaussian distribution, i.e. with a probability density function of

$$g(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp -(x-\mu)^2/2\sigma^2 \qquad (2)$$

with $\mu = 0$, $\sigma = \sigma_t \cdot \sqrt{L}$ and σ_t the transverse diffusion coefficient for 1 cm.

The longitudinal diffusion only affects the time since it is considered as variation from the average drift velocity v_D. z will be set to $z_{cathode}$. Using the probability density function (2), Δt is calculated with

$$\Delta t = \frac{g(x;0,\sigma_l \cdot \sqrt{L})}{v_D} \qquad (3)$$

Distributions of the diffusion Δx, Δy and Δt are shown in Fig. 44.

4.2.3 Avalanche to the Wires

After checking that the position of an electron is not in one of the blind areas of the spokes it is moved to the nearest wire in the anode wire plane. To determine the nearest wire, the actual position vector of the electron is projected on the centre radius of the sector. This also determines the Δr for the new electron position (see Fig. 45). Normally, the electron would travel directly to the nearest wire, i.e. perpendicular to the wire. However, in the wire chamber the E-field is not parallel to the magnetic field anymore which causes the so-called *wire E×B effect*. Due to an E×B component of the fields the electron direction contains a component $\Delta\phi$ parallel to the wire as illustrated in Fig. 45. The deviation angle (we will refer to it as *Lorentz angle* ψ) depends on the fields, the gas and the geometry of the wire chamber. Typically this angle is about 30 degrees. For the ALEPH TPC it was measured to be 32 degrees [56]; this value is adopted as default for the MC.

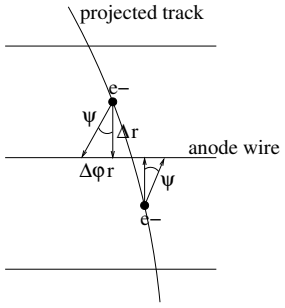


Figure 45: Illustration of the wire E×B effect. Electrons do not move perpendicular to the anode wire but instead under the angle ψ. The signal left by a track on this wire will therefore be smeared.

At this point also a gas gain η_{gas} is calculated and assigned to the electron. Its distribution should be exponential and is obtained from

$$\eta_{gas} = -gasGain \cdot log(u(x;0,1)) \qquad (4)$$

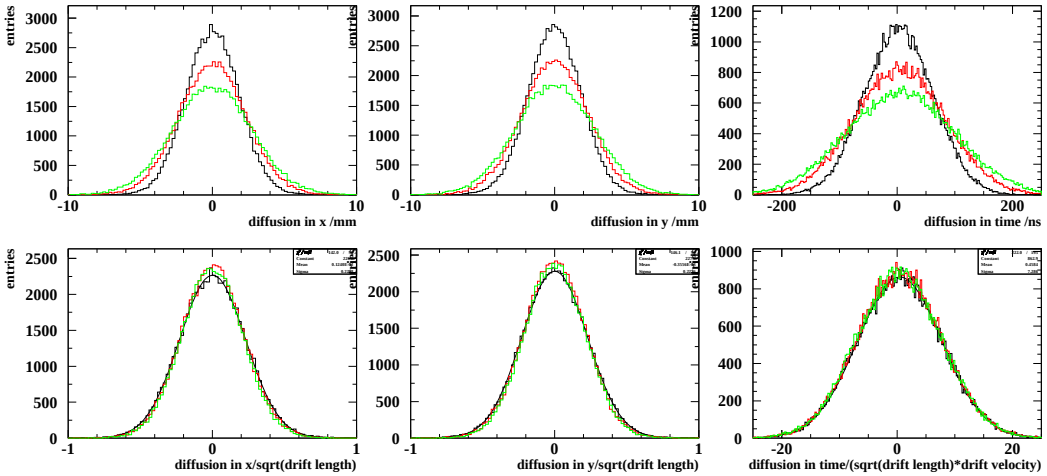


Figure 44: Diffusion in x (left), y (centre) and time (right) for a drift distance of 50 cm (black), 100 cm (red) and 150 cm (green)[58].

with *gasGain* being the scaling parameter and u a random number with the probability density function

$$u(x;a,b) = \begin{cases} 1/(b-a) & a \le x \le b, \\ 0 & \text{otherwise.} \end{cases} \quad (5)$$

A typical distribution of the gas gain is shown in Fig. 46.

For tests, an option is provided to force the gas gain to be 1. Moreover, to model the saturation of the chamber, a value for η_{gas} can be rejected if it is above a certain threshold.

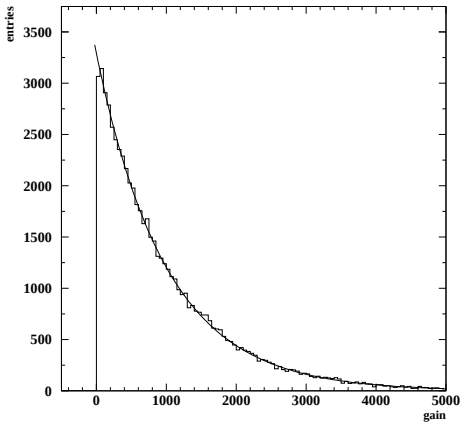


Figure 46: Distribution of the gas gain η_{gas} with a scaling of $gasGain = 1000$. The value for the gas gain is limited to < 5000 [58].

4.2.4 Calculation of the Charge in a Pad

The amount of charge deposited by an avalanche at a certain position into a particular pad is determined by the pad response function.

The charge for a pad is calculated taking into account the distance from the final avalanche position and the geometry of the pad. For wire chambers one can empirically describe the charge distribution as a function of the geometry: for a chamber with distance D between anode and cathode plane and width W of the pad (parallel to the wires) and the total charge density λ the charge P induced is ([56])

$$P = \int_{x-W/2}^{x+W/2} -\frac{\lambda}{2D} \cdot \frac{1}{cosh(\pi x/D)} dx \quad (6)$$

$$= \left| C \cdot \arctan\left(\exp\left(\pi \cdot \frac{x}{2D}\right)\right)\right|_{x-W/2}^{x+W/2} \quad (7)$$

For the HARP TPC readout chamber this inte-

gral is calculated for both dimensions P_r and P_ϕ in r and ϕ, respectively. Then, the absolute value of both dimensions is multiplied and scaled with the gas gain η_{gas} for this avalanche:

$$P_{tot} = \eta_{gas} \cdot |P_r| \cdot |P_\phi| \tag{8}$$

P_{tot} is shown in Fig. 47 for $\eta_{gas} = 1.0$.

To speed up the program, the pad response function is only calculated for pads which have their centre of less than $r_{padsearch}$ from the position of the avalanche.

All charges bigger than a threshold value are gathered in one charge time-series per pad. Note that these digits are just created one-by-one for the charge deposited by each avalanche. There is no sampling/amplification yet.

The Electronics Behaviour

The behaviour of the following two electronic components has to be reproduced: the amplifier on the motherboard and the FEDC which is used to read out the TPC.

4.2.5 Amplification and ADC Sampling

The sampling is done in 'bunches'. Each bunch is expected to be the charge deposited from a single original track and well separated in time (by 2 samples=200 ns) from other bunches. [17] After each bunch, the charge time series is checked for the presence of a later charge which belongs to a new bunch.

For each bunch, the pulse shape is applied which is the transfer function of the amplifier. It is assumed to be a gamma function of the form

$$\left(\frac{t}{\tau}\right)^2 \cdot \exp\left(-\frac{t}{\tau}\right) \tag{9}$$

τ is calculated from the length of the original bunch. Dedicated simulations of the electronics of the HARP TPC have shown that values between 70 ns and 300 ns are expected for different track inclinations (and therefore bunch lengths) [57]. Thus, τ for a given bunch length is found by interpolation to meet this requirement. t is the time between the first charge deposit of this bunch and the end of the current sample (c.f. Fig. 48).

The samples of the bunch are calculated by weighting the transfer function with the sum of charges deposited so far which damps the rising edge of the pulse.

Each ADC which is calculated this way is multiplied with scaling factors to match the range of ADC values of the real TPC which is 0 to 1023 (10 bit). At this point an eventual difference of the pad gains can be applied. A relative error on the equalisation constants can be applied, too.

Usually the sampling of one bunch is stopped because the calculated ADC values have fallen below threshold. For technical reasons also the maximal length of a bunch is limited.

4.2.6 Bunching the Raw ADC Time Series

To simulate the operation mode of the FEDC cards which are used for the TPC readout, the raw ADC time series have to be 'bunched'. Mainly the application of a threshold (i.e. real-time zero-suppression) has to be performed here. In addition, a minimum number of samples above threshold is required as well as post- and presamples to be added. A bunch is considered as the necessary number of samples above threshold with their pre- and postsamples and the corresponding sample numbers (i.e. the time). Samples which overlap due to post- or presamples are merged.

4.2.7 Packing

The simulation chain is completed by the packing of the MC result into raw (binary) data format as delivered by the DAQ readout of the FEDC.

The FEDC packs data in 32 bit words. For each pad there is a 32 bit header containing

- the software address (bits 31 to 16)
  - 31 to 29 contain the sector number
  - 28 is masked
  - 27 to 23 contain the row number
  - 22 is masked
  - 21 to 16 contain the pad number

[17] This condition is usually valid. Of course the limits of this approximation are cosmics/halo particles parallel to the z axis and events with a topology such as two particles which are 'hitting' each other in the TPC. The exact method would be to perform a Fourier transformation of the charge time series.

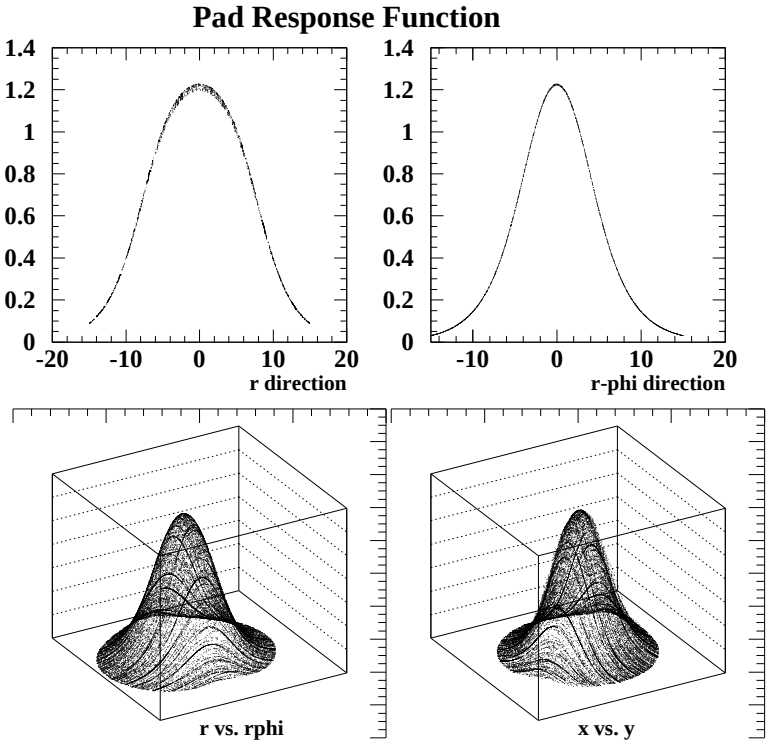


Figure 47: Simulated pad response function. The pad reaches from -7.5 mm to +7.5 mm in r and from -3.0 mm to +3.0 mm in $r\phi$ [58].

- the hardware address (bits 15 to 9, not used)
- the total number of 10 bit words for that pad (bits 9 - 0)

with the numbering of sectors, rows and pads as described in [55]. The event data is packed into 10 bit words (ADC values are between 0 and 1023) of which three are contained in one 32 bit word. Bunch by bunch it is followed by the (10 bit) time of the last sample in the bunch and the (10 bit) total number of 10 bit words for this bunch. Unused bits are set to zero. For example, the raw data block for a pad with one bunch containing seven samples is shown in Table 8.

These raw data from the front-end electronics is then gathered sector by sector and tagged with event and data headers. Their exact format is determined by the DATE DAQ software and will not be described here.

word↓/bits→	29-20	19-10	9-0
0	adc2	adc1	adc0
1	adc5	adc4	adc3
2	9	time6	adc6
3	soft	hard	12

Table 8: Data packing for a pad which contains one bunch with seven samples. The seven ADC values *adc0-adc6* are followed by the time *time6* of the last sample and the total number of 10-bit words for this bunch, i.e. 9. The pad data block is finished with the total number of 10-bit words for the pad (12) plus hardware and software address. In all 32-bit words bits 30 and 31 are not used and set to zero.

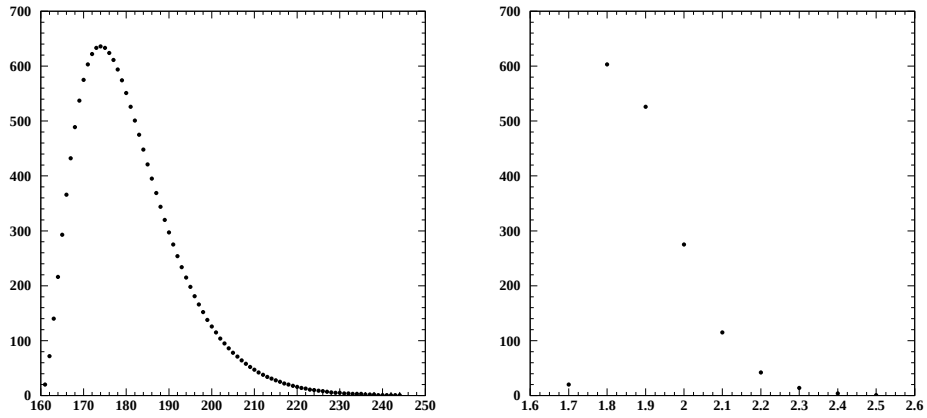


Figure 48: Typical pad signals. Sampled with 10 ns (left) and 100 ns (right) [58].

4.3 Cross-talk

As already mentioned in section 2.2.2, an undesired cross-talk of the front-end electronics on the pad plane was discovered during data taking. This effect has to be implemented into the MC for two reasons:

- Understanding how cross-talk diminishes the spatial and energy-resolution by "switching" cross-talk on or off.
- Development and testing of an eventual cross-talk correction or recovery algorithm.

4.3.1 The Technical Causes for Cross-talk

Cross-talk is caused by capacitive coupling of the output line of one amplifier with the input line of another one (or the same one in case of self-talk). It is due to the proximity of the vias in the pad plane carrying the signals before and after amplification (see Fig. 49).

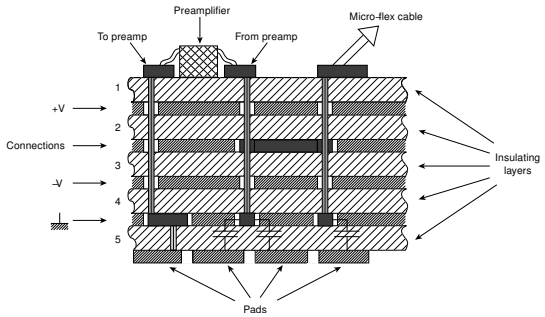


Figure 49: Cross-section of the pad plane. The points of possible couplings are marked with capacitors. The vertical vias carrying the signals traverse almost all the pad plane.

Thus, in the simplest case the pad plane layout can be represented by the circuit shown in Fig. 50. Consequently, cross-talk signals can be predicted by calculation if the couplings are known.

4.3.2 Cross-talk Measurement

A dedicated measurement campaign has been performed in order to identify all cross-talk relations and coupling between pads. During data taking it was done with a spare sector of the pad plane, after the end of data taking the pad plane has been dismantled and the measurement was repeated on the sectors which have actually been used.

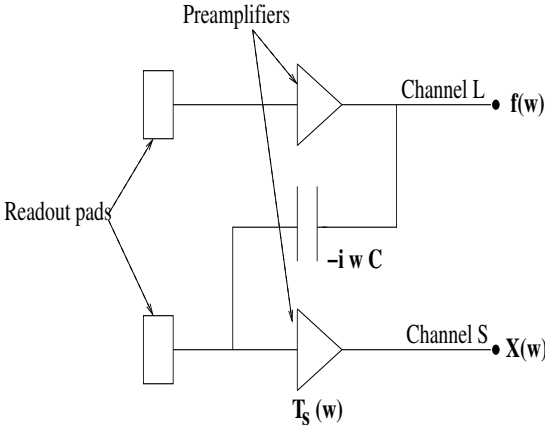


Figure 50: Schematic layout of the preamplifier output-input coupling. In this case, the signal induced by the output line of channel L into the input line of channel S can be calculated if the coupling C and both transfer functions are known.

Each of the 6×622 pads has been excited multiple times and the response of the whole sector recorded. It is done by bringing a probe in direct contact with the pad and feeding it with a step function of 900 mV which reaches the pad through a capacitor of 15 pF.

The signal recorded in the pad which was exited can be considered as its "measured" (or "effective") transfer function. It includes the electronics gain of this channel and eventual self-talk. The signals in other pads are due to cross-talk and are used to find relations and couplings.

Several thousands of events are acquired for each pad. The recorded signals are overlayed using the phase-lock technique (described in [59]). This compensates for the unsynchronised trigger time and the information loss due to the 100 ns binning of the TPC readout.

4.3.3 The Fourier Cross-talk Model

With the pad plane layout and the measurement in hands a model has been developed to predict cross-talk signals. It basically describes the signals which are expected to be induced via the capacitive coupling in the pad plane. To calculate a cross-talk signal the transfer function (TF) of the mother pad

and the daughter pad are multiplied in the Fourier-space.[18] Thus, by fitting the measured signals it is possible to find all couplings if the TFs are known.

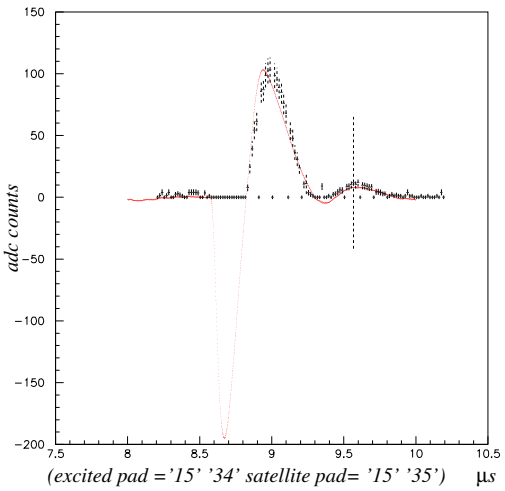


Figure 51: Comparison of the measured and fitted cross-talk signal for one relation in row 15. Mother and daughter pads are 34 and 35, respectively. The negative part of the signal cannot be recorded by the DAQ but is confirmed by using an oscilloscope for visualisation.

As an example, the predicted and measured cross-talk signals for one relation are shown in Fig. 51.

An analysis performed for sector 6 revealed 725 capacitive couplings. From all 622 pads 159 do not have any daughters and 325 no parents. Hence, the signals of roughly half of the pads are not affected by cross-talk.

4.3.4 Cross-talk in the MC

The information used by the MC to reproduce cross-talk are

- the effective TFs for all pads, no matter whether they are affected by cross-talk or not,
- a map of cross-talk relations with mother and daughter pads ID and the coupling.

[18] A "mother" pad is a pad inducing a signal into another pad (the "daughter"). A pad can be mother and daughter at the same time.

In a first step, all primary pad signals are generated using the measured TFs. Thus, the concept of a fixed (calculated) pulse-shape as it is used for the description of the "ideal" TPC is replaced by a convolution of the charges on a pad with its TF.

In addition, instead of the standard 100 ns binning, a binning 10 times finer is used. This has to be done in order to obtain smooth cross-talk signals. Each order of cross-talk is then generated by passing all pad signals with 10 ns binning to a software module which returns a new set of pads with cross-talk added. In detail, the cross-talk calculation in this module is done in the following steps:

- All pads which have daughters are searched using the cross-talk map.
- For each relation, the mother signal $S_M(x)$ is transformed in Fourier space $S_M(y)$ with a Fast Fourier Transformation (FFT). The same is done with the transfer function $T_D(x)$ of the daughter pad. The product of both is then multiplied with $i\omega$ and the coupling C and gives the resulting daughter signal $S_D(y)$ in the Fourier space.

  $$S_D(y) = -i\omega \cdot C \cdot S_M(y) \cdot T_D(y) \qquad (10)$$

  The signal $S_D(x)$ recorded in the daughter pad is then obtained by applying an inverse FFT to $S_D(y)$.
- If the daughter pad received already a (primary) signal it is added to the cross-talk signal. Otherwise a new pad is created which might then cause second-generation cross-talk itself.

4.4 Distortions

Inhomogeneities of the electric and magnetic fields in the TPC may lead to displacements of drift-electrons and therefore wrong reconstruction of clusters and tracks.

The magnetic field of the solenoid has been has been experimentally mapped with a hall probe and can be taken into account in the simulation. The electric field inhomogeneities arise from voltage misalignments between the inner and outer field cages and are more difficult to describe. Preliminary measurements indicate a misalignment of or-

der of 1% at the inner field cage voltage. The resulting $E \times B$ field component rotates the electrons during their drift.

A prototype plug-in for the MC has been developed which rotates the position of the drift electrons before they arrive at the first wire plane of the chamber. The rotation angle depends on the r-position of the electron and the length of its drift. This is illustrated in Fig. 52.

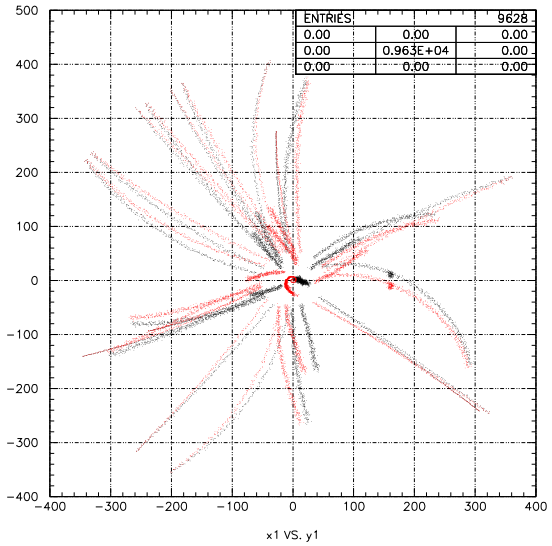


Figure 52: Illustration of distortions for the tracks of a few events. During their drift, the electrons are rotated clock-wise, depending on r and the drift-length.

5 Unfolding

5.1 The Unfolding Problem

A common task in the analysis of a High Energy Physics experiment is to construct the probability density function (pdf) $f(x)$ of a variable x by a series of measurements of x. The measurement can usually be represented by a histogram with N bins and a vector $\boldsymbol{\nu}$ of bin entries. However, if the measurement of x is subject to systematic and statistical errors it is not valid to consider $\boldsymbol{\nu}$ as estimates for the expectation values of the *true* bin content. Instead, an additional quantity y (the measurement) has to be introduced with its own pdf $g(y)$. The histograms $\nu_{1..N}$ and $\mu_{1..M}$ are then the discrete representation $g(y)$ and $f(x)$, respectively.

The pdfs $f(x)$ and $g(y)$ of the true and measured variable, respectively are then correlated by the following integral:

$$g(y) = \int S(x,y)\ f(x)\ dx \tag{11}$$

where $S(x,y)$ is the *response function*.

If both pdfs are represented as histograms we can formulate

$$\nu_i = \sum_j^M S_{ij}\ \mu_j \tag{12}$$

and $\mathbf{S}$ becomes the response or smearing matrix. S_{ij} describes the conditional probability that an event has the true value in bin j and the observed one in bin i. If background $\boldsymbol{\beta}$ is included Eq. 12 becomes

$$\boldsymbol{\nu} = \mathbf{S}\ \boldsymbol{\mu} + \boldsymbol{\beta}. \tag{13}$$

The process of using relation (12) to determine a transformation for the measurement to obtain the expectation values $\boldsymbol{\mu}$ is called *unfolding* or *deconvolution*.

Depending on the kind of problem different unfolding methods can be applied. The most simple and obvious solutions are matrix inversion and bin-by-bin correction. They can be applied if acceptance and efficiency are the dominating errors of the measurement.

Bin-by-bin correction is a relatively simple method to find estimators for $\boldsymbol{\mu}$. Here, a MC simulation is used to determine a correction factor $C_i = \mu_i/\nu_i$ for each bin ν_i to calculate μ_i. Even if migration effects and correlation are small, the result is biased towards the MC. Moreover, the errors are underestimated since they are dominated by the Poisson statistics of the data.

Matrix inversion is the attempt to invert the response matrix S_{ij} and to apply it to the measurement. While appearing very natural this method often yields disappointing results. The reason is that in the case of too large off-diagonal elements in the response matrix (i.e. correlation between bins) the result is dominated by catastrophically large variances and strong negative correlation between neighbouring bins. One way out of this problematic aspect is to damp the parts of the result which are dominated by statistical fluctuations. This process is called *regularisation* and the basic idea of the unfolding method proposed by Blobel [60], [61]. Here, the solution is decomposed into a set of orthogonal functions and high frequency parts are identified by imposing a certain *smoothness* on the solution. Another regularisation scheme is proposed by Höcker and Kartvelishvili [62] which is based on the Singular Vector Decomposition of the response matrix.

An iterative method not trying to cure the deficiencies of the inverted response matrix has been proposed by D'Agostini [63]. It will be described in more detail in the next section.

It can be concluded that unfolding beyond simple matrix inversion or the application of correction factors should be envisaged in the following cases:

- The resolution of the measurement is low. In this case there is only a statistical relation between the true value x and the measurement y and the correlations between adjacent bins are not negligible.
- Efficiency and acceptance of the measurement are depending on x, i.e. the loss of events is biased with x.
- Effects like migration/bias/shift are not negligible. This is the case if the true and measured quantity are related via an unknown transformation.
- Limited statistics in MC and data.

5.2 Bayesian Unfolding

5.2.1 Method

The "multidimensional unfolding method based on Bayes' theorem" [63] (from now on referred to as *Bayesian unfolding*) performs the calculation of the unfolding matrix $\mathbf{M}$ in an iterative way. $\mathbf{M}$ is a two-dimensional matrix connecting the *effects space* (i.e. the measurement) to the *causes space* (the true values). Both spaces are binned and the binning can be represented by a multidimensional histogram with a total number of n_E and n_C cells, respectively. We can then solve Eq. 12 for $\boldsymbol{\mu}$ and rewrite as:

$$\hat{n}(C_i) = \sum_j M_{ij}\ n(E_j) \tag{14}$$

with $\hat{n}(C_i) \equiv \boldsymbol{\mu}$ and $n(E_j) \equiv \boldsymbol{\nu}$.

At this point it is important to notice the difference between bins and cells: for example, a three-dimensional histogram with 50 bins in each dimension will have $50^3 = 125000$ cells.[19]

Starting point of the iteration is the following formulation of Bayes' theorem for the conditional probability $P(C_i|E_j)$ for effect E_j having been caused by cause C_i:

$$P(C_i|E_j) \propto P(E_j|C_i)\ P_0(C_i) \tag{15}$$

$P(E_j|C_i)$ is the smearing matrix $\mathbf{S}$ and therefore the probability for cause C_i to produce effect E_j. $P_0(C_i)$ the initial probability for the ith cause to happen. $\mathbf{S}$ is either provided as matrix or calculated from MC events. If MC events are used it is best to generate data in such a way that all (kinematic) areas of both causes and effects histograms are well populated. $P_0(C_i)$ can either be a uniform distribution or a prediction extracted from a theory or a model.

Next, the numbers $\hat{n}(C_i)$ of expected events in each cause cell are calculated with

$$\hat{n}(C_i) = \sum_j n(E_j)\ P(C_i|E_j) \tag{16}$$

Here, $n(E_j)$ is the measurement, i.e. the number of events observed in each effect cell. Notice that $\hat{\mathbf{n}}(\mathbf{C})$ are *expected* values, not the *true* ones $\mathbf{n}(\mathbf{C})$.

$\hat{n}(C_i)$ is then used to calculate a new $P(C_i)$ which is inserted back into Eq. 15 for the next iteration. It is possible to smooth $P(C_i)$ before an iteration to damp oscillations due to statistical fluctuations. Between the results of two sequential iterations a χ^2 test is performed. When the difference between the results becomes very small, the iteration is stopped by default. However, it can be reasonable to perform fewer or more steps. The optimal number should be found using tests with the MC: too many steps reproduce the fluctuations one would have obtained by plain matrix inversion; too few steps will bias the result towards the measurement.

The final result is the unfolded distribution $\hat{\mathbf{n}}(\mathbf{C})$ and its covariance matrix.

For simplicity, additions to cover normalisation, inefficiencies and background are not included in the above formulas.

5.2.2 Usage

As already pointed out above, one of the most crucial steps in the unfolding process is the choice of variables and binning in the causes and effects space. Once this is decided, the further application of the method is rather straight-forward thanks to the implementation provided by D'Agostini. The following treatment of MC and measured data has to be performed in order to employ this implementation:

- All events in the MC sample are evaluated for the cause and effect cell into which they belong. The result of this process is a matrix S'_{ji} of dimension (n_E, n_C) describing how many events are generated for the i^{th} cause cell and reconstructed in the j^{th} effect cell. $\mathbf{S}$ and $\mathbf{S'}$ only differ by the normalisation, but since $\mathbf{S'}$ contains absolute event numbers it is possible to take the statistical error of the MC sample into account.

- The number of MC events generated for each cause is calculated. This is not necessarily identical to a projection of $\mathbf{S'}$ because *causes* might be generated which are reconstructed outside the acceptance of the effects binning. Those events are not represented in $\mathbf{S'}$. Thus,

[19] If both effects and causes would be represented by such a histogram, $\mathbf{M}$ would have $125000^2 = 1.5625 \cdot 10^{10}$ entries which will clearly cause technical problems (memory consumption and CPU time). It is therefore essential to keep n_E and n_C as small as possible. This shows that the choice of variables and their binning in both spaces is delicate.

at this point inefficiencies are taken into account.

- An initial distribution for $P_0(C_i)$ must be found. If no prediction is available, a flat distribution can be used. In either case P_0 must be normalised, i.e.

$$\sum_i P_0(C_i) = 1 \tag{17}$$

- The measurement result is binned using the *effects* binning. The result is the vector $\mathbf{n}(\mathbf{E})$ of the number of events reconstructed in each effects cell.

The unfolding can now be performed. For the start one can leave the choice of the ideal number of iterations to the program and in order to accelerate the execution, error calculation can be limited or switched off. In particular, it is then possible to study the result and – if needed – to try a different selection of variables or binning. It should also be studied whether smoothing can be applied between two steps.

5.2.3 Summary

Bayesian unfolding has been chosen for the analysis for the following reasons:

- The migration matrix $\mathbf{M}$ is not calculated by matrix inversion of $\mathbf{S}$, i.e. $\mathbf{M} \neq \mathbf{S^{-1}}$. This avoids the difficulties described above.
- The covariance matrix of the results is provided.
- The starting distribution for the first iteration does not necessarily have to be realistic to obtain convergence.
- It can unfold multidimensional distributions.
- The distributions in the causes and effects space can have different dimensions and binning. In particular this also means that the variables in both spaces do not have to be the same.
- A clean implementation by D'Agostini in the form of a small, fast and easy to use FORTRAN subroutine is available.
- The program does not deal with single events but instead with distributions. Thus, rerunning the program with different parameters does not require to re-process all events.
- The number of adjustable parameters is small. Moreover, their meaning and interpretation is close to the application.

5.3 Unfolding of Mock Data

Before applying unfolding to real data it should be demonstrated that it performs well under controlled conditions. Therefore, the MC is used to generate mock data which is then unfolded. Of course the MC program which generates the mock data and the unfolding input is the same, however if the MC reproduces the real data and all the migration effects with good accuracy this comparison proves the validity of the unfolding.

5.3.1 TPC Track Reconstruction

At this point it is useful to review the steps of the reconstruction of TPC data. It will help to understand the differences between MC truth and reconstructed tracks and certain features of the unfolding. It must be pointed out that at the time of this study the TPC reconstruction is not yet available in its final version and therefore its performance is diminished.

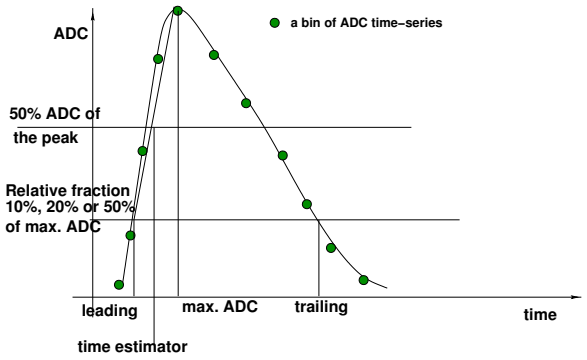


Figure 53: Calculation of the ADC value and time estimator for a pad signal (from [67]).

Starting from the raw FEDC time series for all pads the following steps are performed to reconstruct the spatial track parameters of all tracks in an event:

- Equalisation of signals heights. All recorded signals are corrected using the gain of each pad. This gain is the product of the gain of the chamber at the pad's position and the gain of its amplifier and will be referred to as *pad gain*. It is found using cosmic muons traversing the TPC and dedicated calibration runs as described earlier.

- Clustering of pads. For each pad with a signal, a time and charge estimator is calculated (see Fig. 53). Due to the effects of crosstalk, the charge is determined by the maximum ADC value and not by the sum of all samples. In addition, the centre of the rising edge is considered as best time estimator in presence of crosstalk. All pads of one row belonging to a single track are then merged to a *cluster* (or *point*) with time and charge being the weighted mean of the estimators of its pads. The present clustering algorithm is described in detail in [67].

- Pattern recognition. The pattern recognition identifies points belonging to one track. It is adapted from the algorithm used in the CHORUS experiment [68].

- Track fit. All points belonging to a track are fitted assuming a helix model for the track. The helix is described by five track parameters as shown in Fig. 54:

  ρ is the radius of the track (it should be mentioned that in some conventions it is the inverse of the radius). ρ is signed: if the track bends clock-wise (seen from the centre) it is positive, otherwise negative.

  d_0 is the impact parameter in the x/y-plane. If the z-axis is inside the helix it is positive, otherwise negative.

  ϕ_0 is the track angle in the x/y projection at the impact point.

  $\tan\lambda$ is the dip angle in the z/s-plane where s is the length of the track starting from the impact point.

  z_0 is the z-coordinate at the impact point (i.e. at $x^2 + y^2 = d_0^2$).

- Kalman fit. The track fitted to a helix is passed to a Kalman filter. Currently, it only re-fits the track and estimates a χ^2 for the fit quality.

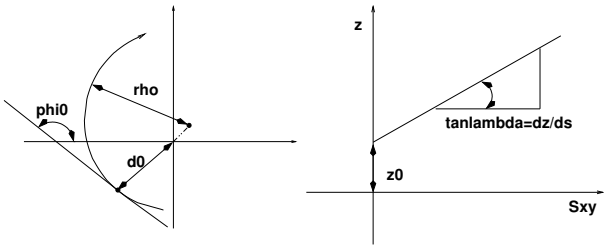


Figure 54: Helix parametrisation of a TPC track (from [69]). The parameters are described in the text.

5.3.2 MC Truth vs. Reconstruction

For both the mock data sample which will be unfolded and the MC sample which is needed as input for the unfolding it is essential to relate the information from the MC (the *MC truth*) with the output/result of the reconstruction. This is not trivial since due to correlations and the vast amount of information it is not practical to carry the complete MC truth through the whole simulation and reconstruction starting from each single drift electron and ending with the final track fit. Thus, a different method has been used to match reconstructed tracks with MC particles.

Since at the time of writing the HARP reconstruction is not yet available in its final functionality and quality the particle identification (via dE/dx and TOF) is not considered and therefore replaced by the MC truth information.

In general, MC particle tracks will be considered as *visible* in the TPC if their inclination angle θ is between 10 and 140 degrees and the (MC-) particle performed not less than 100 simulation steps in the TPC gas with a standard step length of 2 mm.

For each of the visible particles a corresponding reconstructed track is searched. Only particles are accepted which can be – within a certain accuracy – uniquely assigned to one reconstructed track. In detail, the matching between each MC particle ('particle') and the reconstructed tracks ('tracks') is performed as follows.

By comparing reconstructed track parameters with the true ones it was found that in general the reconstruction of the initial track angles θ and ϕ

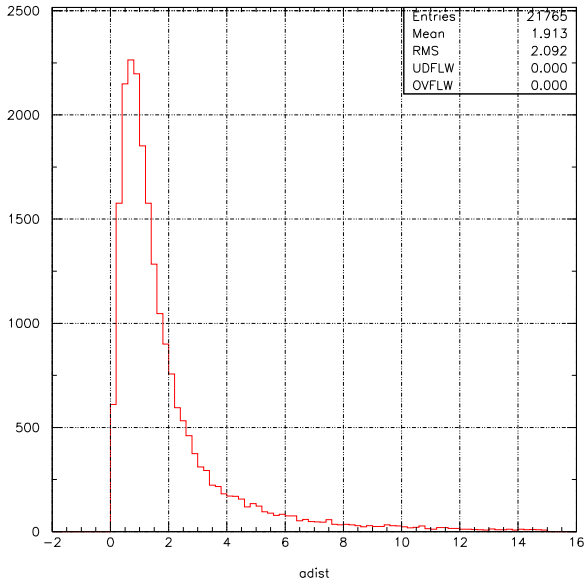


Figure 55: Distribution of χ^2 for the angles of all reconstructed (i.e. matched) charged pions in the mock data sample. The distribution is not an exact χ^2 because the errors of the reconstructed angles θ_{rec} and ϕ_{rec} are neither exactly Gaussian nor constant over the whole range. Values above 15 are cut.

(i.e. the direction) is the most accurate one. Even if the reconstructed absolute of the momentum is far away from the true one the direction is usually correct. In particular, the reconstructed θ_{rec} is very robust: only a nearby track (similar θ *and* ϕ) causes problems. On the other hand, the error on ϕ_{rec} increases not only in case of nearby tracks but also at the boundaries of sectors, the spokes. Here, the pattern recognition for the TPC is not yet tuned to handle these blind areas with high efficiency and accuracy. Based on these observations, for the angles of each track a χ^2 comparison with the true values is made and they are ordered by this χ^2. If the best χ^2 is small enough and no other track has a similar value it is considered as matched.[20] Compared to other matching schemes this method yields the most reliable and stable results. Another benefit is that additional cuts do not improve things and are therefore not needed.

For illustration, the distribution of these χ^2s for matched tracks in the mock data sample which will be unfolded is shown in Fig. 55.

5.3.3 Mock Data

Mock data is generated using the standard HARP MC with a 3 GeV/c proton beam on a Ta target of 5% λ_I thickness. As described earlier, only matched pions will be considered and their identification relies on the information from the MC. The reconstruction of tracks is performed using the default reconstruction software and parameters.[21]

The goal will be to unfold the two-dimensional p_t vs. p_l distribution of charged pions generated in the target.

The true and measured distributions are shown in Fig. 57. The parameters for this unfolding (variables, binnings) are summarised in Tab. 9.

			range [MeV/c]		
	dim.	variable	from	to	bins
causes	0	p_t	-100	1000	16
	1	p_l	0	600	18
effects	0	p_t	-100	1000	20
	1	p_l	0	700	20

Table 9: Parameters for the unfolding of the mock data. The naming convention using *causes* and *effects* and their two *dimensions* will be used throughout the rest of this chapter.

In our case, variables (and therefore also the number of dimensions) are the same for the causes and effects. Of course, in general this is not mandatory.

The choice of variables should be motivated by the errors (i.e. smearing and migration effects) of the measurement and their impact on the actual measured quantities and their relation to the *wanted* physics quantities. For the HARP TPC, these physics quantities are indeed p_t and p_l. p_t directly follows from the measured radius r of the track in the solenoid field ($r \propto p_t$) and p_l from the inclination angle θ of the track. The binning of the unfolding histograms is adapted to the errors

[20] In order to make the calculation of the χ^2 more reliable, the expected errors on both angles are adapted to other track parameters: for tracks with a very low or very high θ, larger σ_ϕ and σ_θ are expected. In addition, the error on ϕ is larger if the particle is close to a spoke (at $\pm 30°, \pm 90°, \pm 150°$) or its absolute charge is higher than one.

[21] The software used for the generation of both mock data and MC samples is basically the one of the HARP software release *v5r15* with minor adaptions to the TPC simulation part.

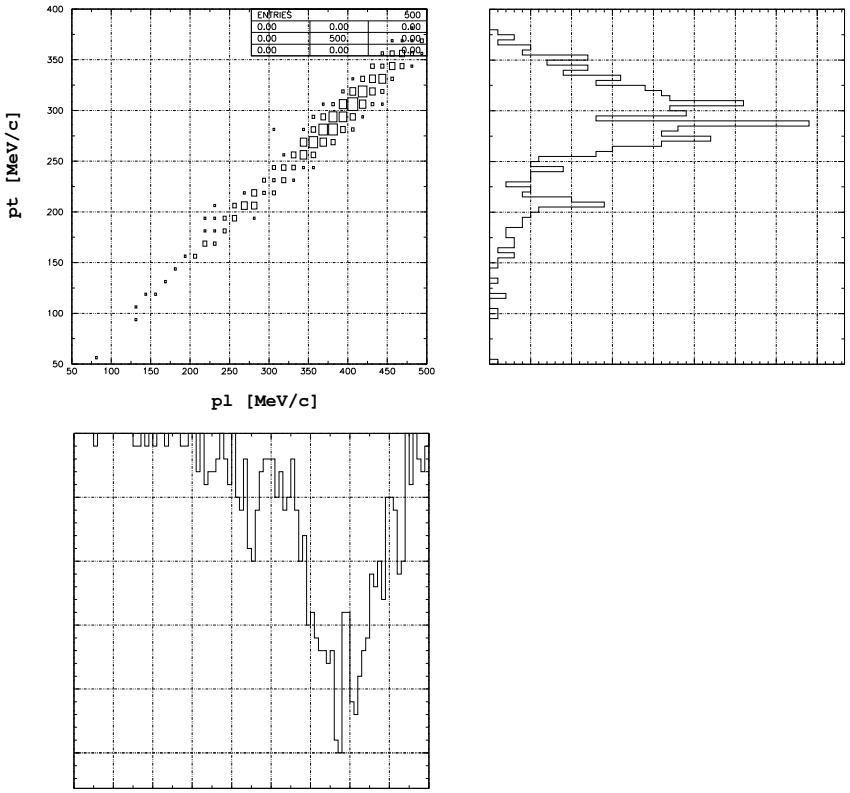


Figure 56: Illustration of migration effects. Shown is the reconstructed p_t and p_l for all tracks in the mock data sample having a true p_t and p_l between 300 and 310 MeV/c and 400 and 410 MeV/c, respectively. As it can be expected from the reconstruction method (section 5.3.1), p_l and p_t are linearly correlated since the only momentum measurement is p_t. Moreover, for certain tracks the momentum seems to get systematically underestimated. This happens indeed if tracks do contain points with large errors on the inner or outer rows. Consequently the track fit will assign a smaller curvature to the track and therefore a smaller momentum.

of the measurement and the expected errors of the unfolding result. For example, a look at the covariance matrix of the unfolding output can help to identify non-optimal binning. When determining the ranges, it is important not to cut into distributions. In this case, the results near the boundaries become inaccurate. For the mock data unfolding the ranges are chosen to contain almost the whole distributions.

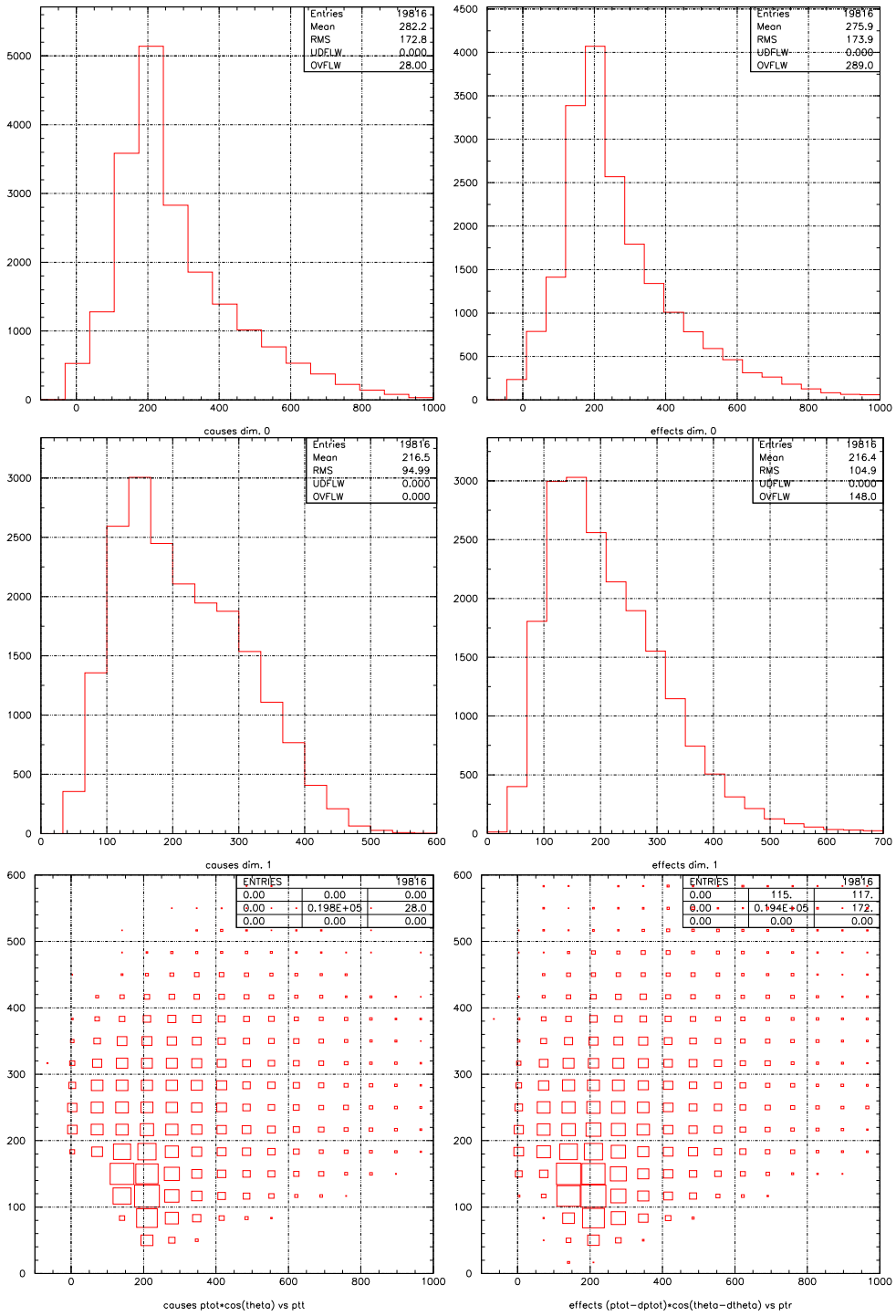


Figure 57: Momentum distributions of the sample to be unfolded. On the left the true (causes) distributions: p_l (top), p_t (middle) and p_t versus p_l (bottom), c.f. Table 9. For comparison, the corresponding measured distributions (effects) are shown on the right. The 2D effects plot is re-binned to the causes binning in order to facilitate the comparison. Obviously, unfolding should reproduce the left distributions from the right ones. All scales are in MeV/c.

5.3.4 MC Input

Two different samples of simulated data are used as MC input for the unfolding:

- The *single-track sample* consists of an accumulation of muon tracks generated at the nominal vertex point in the target. Their directions and momenta do not follow any particular distribution.[22] However, it is essential that all kinematic areas are well covered.
- The *interaction sample* is basically the same kind of data as the mock data itself. The particle spectra are therefore according to the expected (realistic) distributions. The statistics of this sample is about ten times the one of the mock data. The mock data itself is *not* part of this sample.

The true and measured p_l and p_t distributions for the single-track sample and the interaction sample are shown in Figs. 58 and 59, respectively.

As already pointed out earlier, one important prerequisite for an unfolding algorithm is that it does not bias the unfolding results towards the MC input. This is *particularly* important for HARP since its results are supposed to be used to tune the hadronic MC generators.

Thus, the unfolding results obtained from the two MC samples are compared in order to show that the unfolding result is indeed independent of the particle generation model in the MC.

[22] This sample is composed of several tens of smaller samples each with well defined angular and momentum distributions.

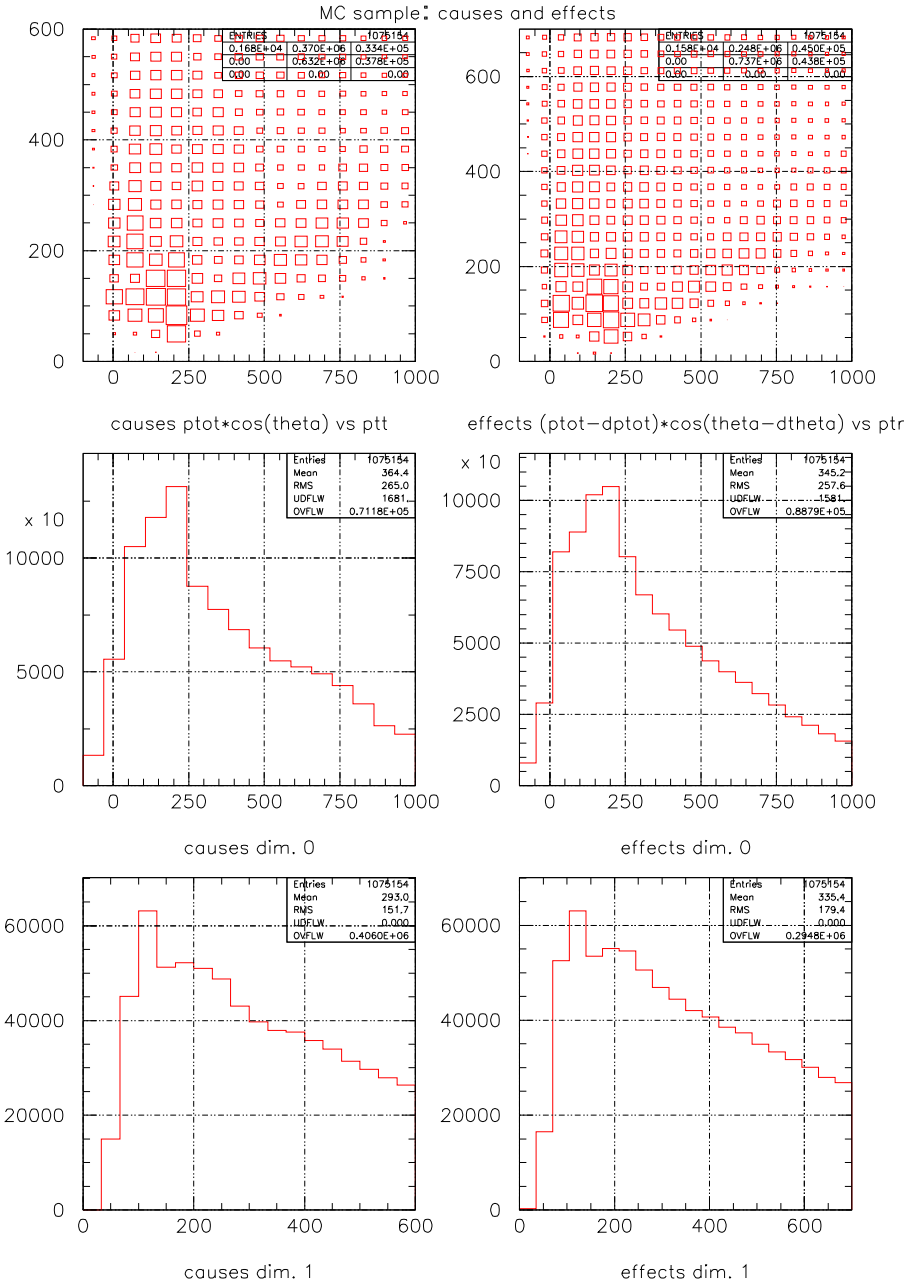


Figure 58: MC single-track sample. It is composed of several tens of overlapping samples each covering a certain kinematic region. Causes histograms are shown on the left, effects on the right. Dimension 0 is p_l and dimension 1 p_t (c.f. Table 9).

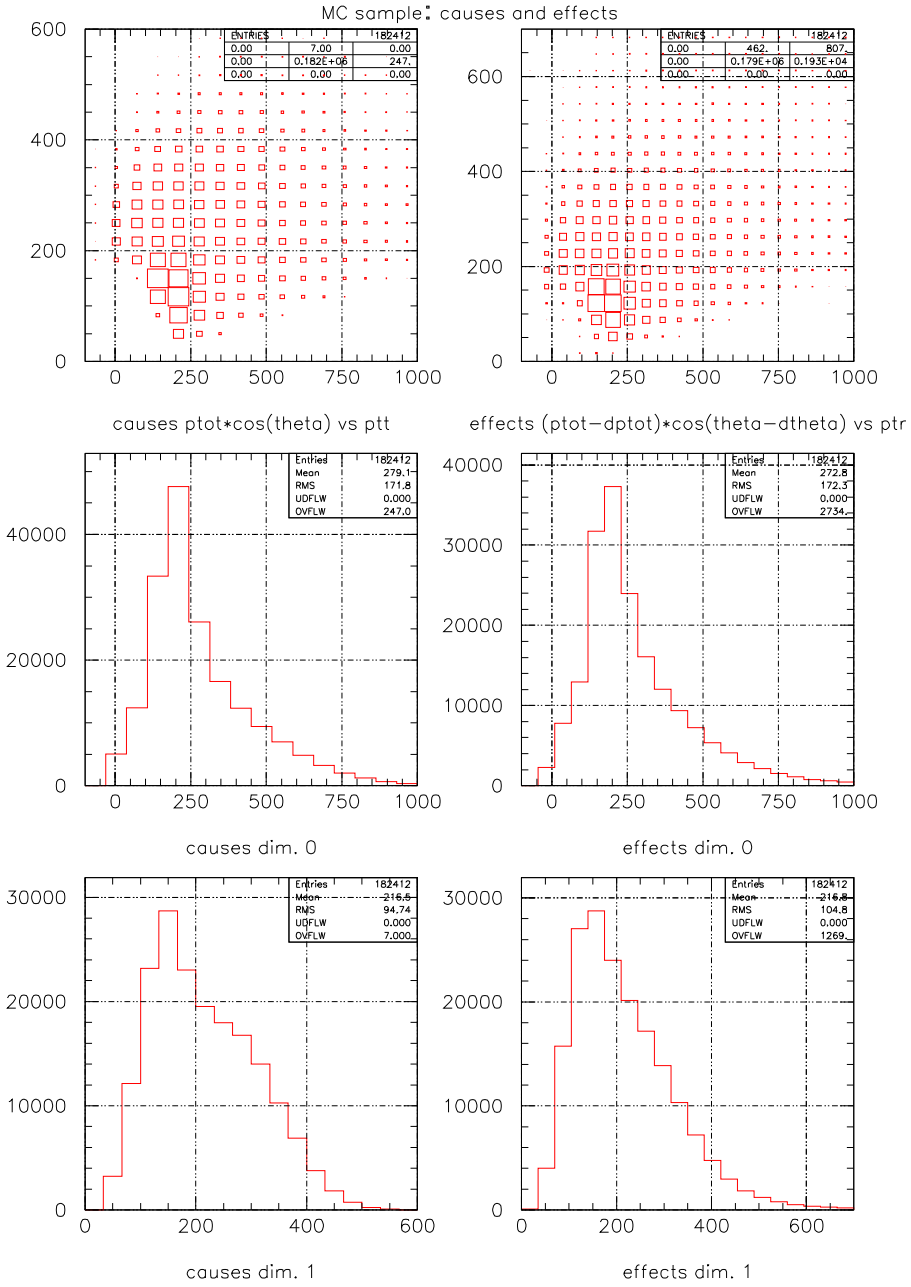


Figure 59: Same as Fig. 58 but for the interaction MC sample. Of course, the shape of the distributions and the visible migration effects are the same as in the mock data sample (Fig. 57). The statistics of this MC sample is about ten times the one of the mock data.

5.3.5 Unfolding Results

The basic requirement to the unfolding is of course that the unfolded distributions are within errors compatible with the true distributions. Even if this is not the case for every single bin, the unfolding result should be closer to the truth than the measurement. In order to verify this, the unfolding result is compared with both the measurement and the truth.

On the following pages the results are shown in detail and discussed.

Projections
In general, the unfolding performs as expected. For the projections p_t and p_l we are interested in, the unfolding manages for most bins to reproduce the truth within the error, as can be seen in Figs. 61 and 62. The top plots show the distributions for truth, measurement and unfolding. In order to see more clearly in which areas the agreement is not satisfactory the middle plots show the ratios between measurement (circles)/unfolding (crosses) and the truth. Here it becomes obvious that the unfolding has problems at the upper tail of the distributions. However, the deviation from unity is only so large because of the small numbers in the true histogram. In the same plots for the *differences* (bottom) these areas look already less alarming. The difference plots clearly show that even if the unfolded bin content is not consistent with the truth it is usually much closer than the measurement. Basically, the results obtained with the single-track and interaction MC sample are the same, however around the peaks of both distributions the interaction sample seems to performs slightly better.

A graphical representation of the covariance matrices for the four unfolded distributions is shown in Fig. 63. They do not show too high off-diagonal elements, but it can be seen that one could expect some improvements by using variable bin widths for the histograms.

2D
It should be recalled that the actual unfolding is performed on the two-dimensional p_l-p_t histogram. Evaluating the projections is already rather demanding on the quality of unfolding since eventual errors of cells which are added up for the projection can compensate or amplify. Therefore a comparison of the measured, true and unfolded 2D histograms is instructive.

For this reason the bin content of each of the 288 bins before and after unfolding is compared with the true one. In order to quantify the quality of the unfolding the relative errors between the unfolded, measured and true bin content are histogrammed. This is shown in Fig. 64 using the single track and interaction MC sample. It is here where the earlier impression of the interaction MC sample producing better results gets confirmed. Its distributions of errors with respect to the truth are narrower and more symmetric while the distribution obtained with the single-track sample looks similar to the measured ones.

Stability
To evaluate the stability of the results discussed so far the following unfolding conditions/parameters were modified observing the impact on the results:

- *Number of iterations*
  The number of iterations proposed by the program to obtain both mock data unfolding results is between 10 and 15. In this case a uniform distribution has been used as starting point. If under these conditions the number of iterations is forced to be up to five more or five less, the results do not change significantly. This is a sign of good convergence.

- *Starting distribution*
  In all cases a uniform starting distribution for the iteration performs well. When random bin contents are used as start the results are in general not usable.

- *MC sample*
  As already pointed out above, the single-track MC sample consists of tens of smaller samples each covering a certain kinematic area (with very large overlaps though). If a few of these samples are removed it can be observed that the accuracy of the unfolding in this area degrades. Thus, care must be taken that all kinematic areas are well covered by the MC. Naturally, reducing the statistics of the MC interaction samples does not have this immediate impact.

- *Smoothing*
  In general, in an unfolding procedure some kind of smoothing can be applied in order to remove unwanted and statistically dominated oscillations. For the mock data unfolding it was found that smoothing the 2D histogram between iteration steps[23] worsens the results considerably. Comparing the shape of the histogram before and after smoothing shows that this problem is caused by the smoothing itself not performing as expected. This is independent of whether the smoothing is performed at the beginning, middle or end of the iteration. In the best case the unfolding re-establishes the original results from before the smoothing. It can be concluded that smoothing should only be applied if there are really oscillations to damp and a reliable smoothing method is used which does not change too much the shape of the distribution. The smoothing of the 1D projections is of course not possible in the current scheme.

Errors

The errors provided by the unfolding program are checked by unfolding ten only stochastically different mock data samples with the same MC sample. For each bin in the projected p_l and p_t histograms the difference between true and unfolded bin content is calculated and divided by the unfolding error. The histogram of these values is shown in Fig. 60.

It can be seen that the distribution is well centred around zero and its shape is Gaussian with a width of three. However, the width should be equal to one since the errors which are histogrammed are in units of σ. It must be concluded that the errors provided by the unfolding program are systematically too small by a factor of three. This is consistent with the mock data unfolding results as shown in Figs. 61 and 62: if the MC describes perfectly the detector – certainly true for the mock data – the unfolding should reproduce the truth within errors for a large majority of the bins, which is not the case.[24]

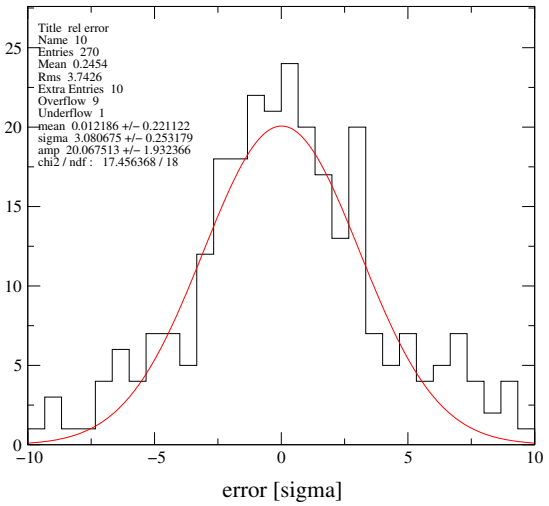


Figure 60: Errors of the unfolding result. The normalised unfolding errors of ten mock data samples are histogrammed. Each sample contributes with 16 bin errors in p_t and 18 bin errors in p_l. Only bins containing more than 50 (true) events are considered.

[23] The standard smoothing function for two-dimensional histograms provided by PAW [65] has been used.

[24] The error calculation inside the program has not been further validated. According to the author (G. D'Agostini) certain approximations can indeed lead to errors being underestimated. This issue should be addressed before using this unfolding scheme for the production of the final results.

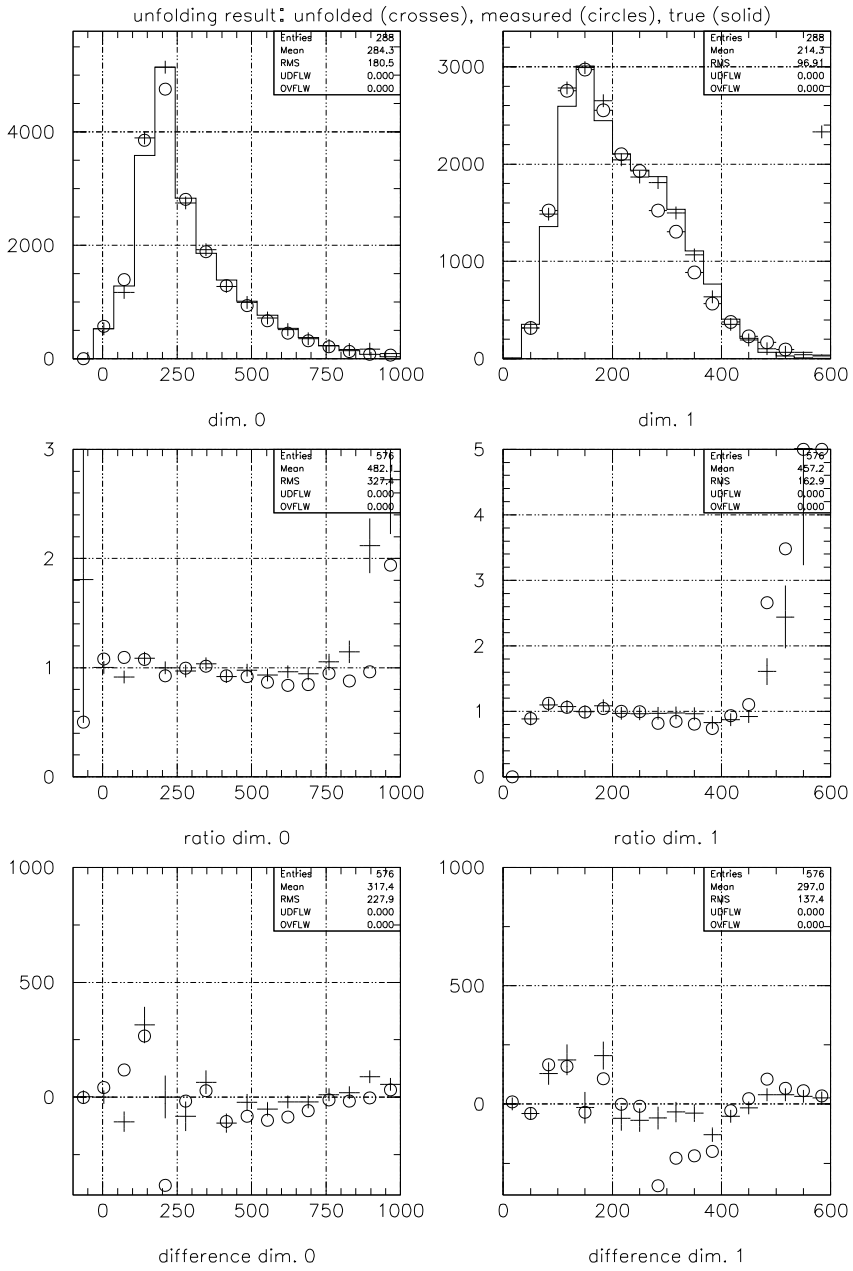


Figure 61: Unfolding result using the single-track MC sample. On the top, the measured, unfolded and true projections of the unfolded 2D histogram are overlayed. The middle plots show the relative bin content of unfolding and measurement with respect to the true one. The absolute difference of the bin content is displayed in the bottom plots.

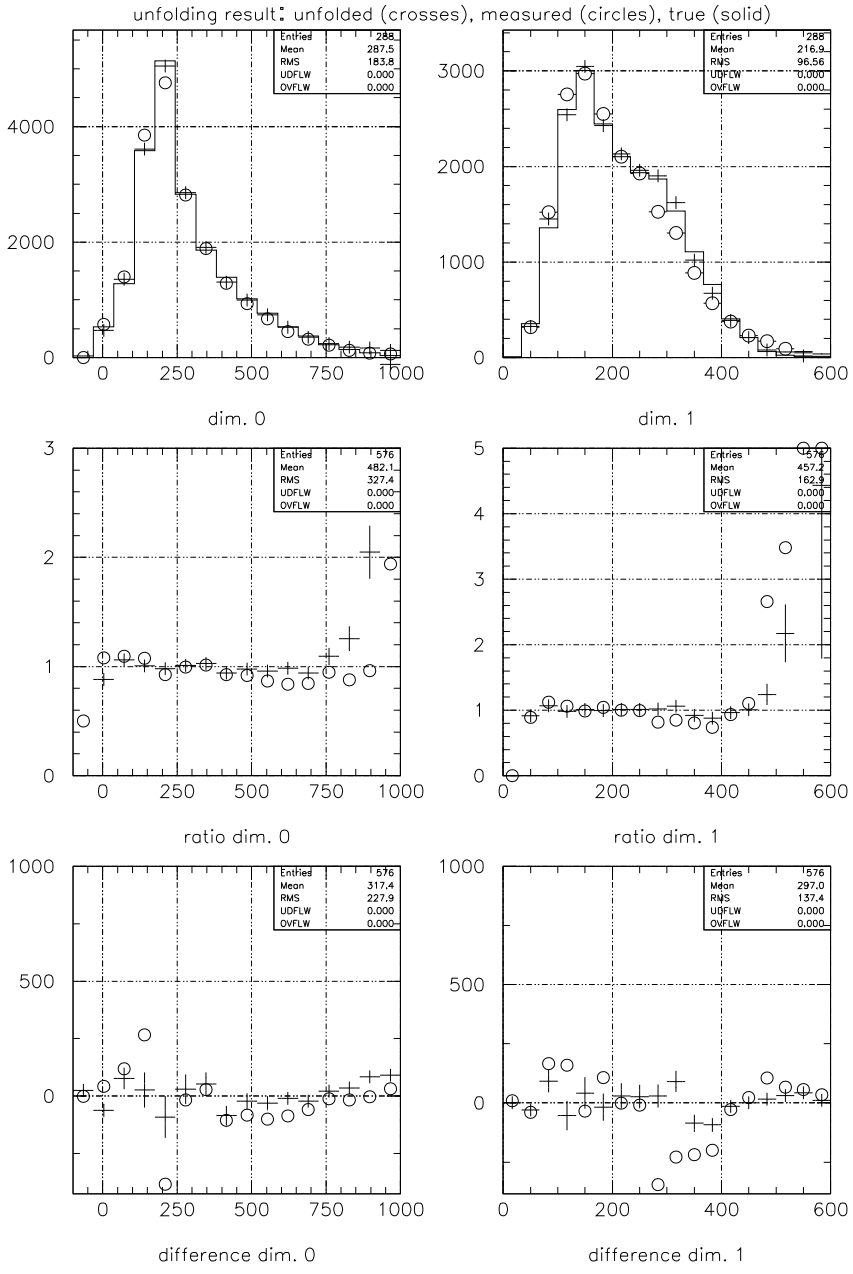


Figure 62: Same as Fig. 61 but using the interaction MC sample.

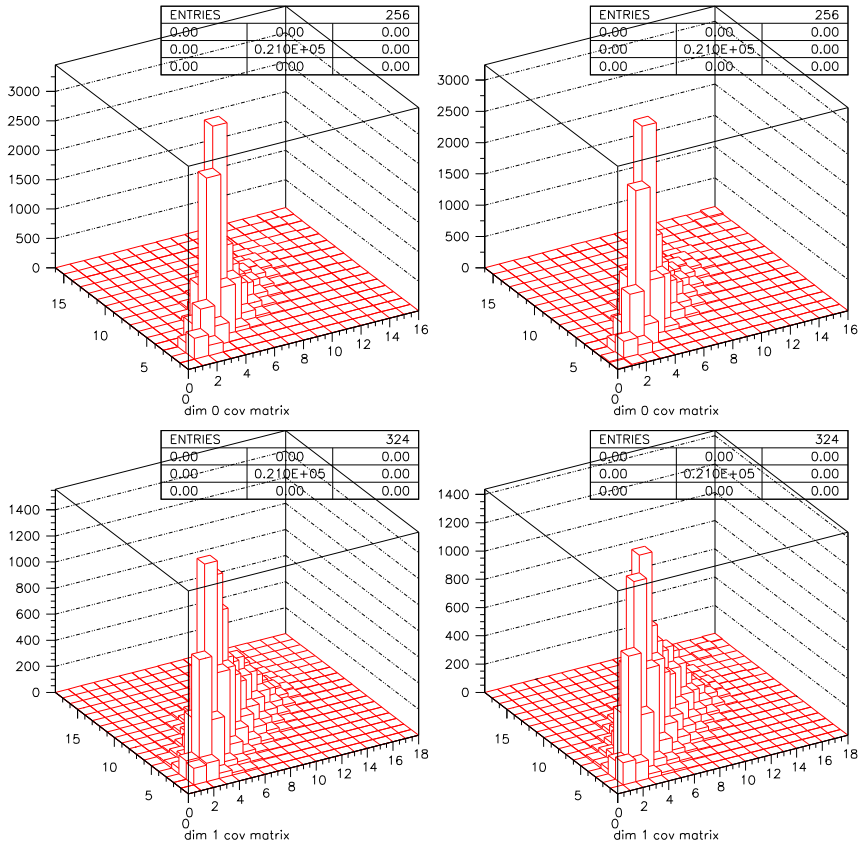


Figure 63: Graphical representation of the covariance matrices for the distributions of p_l (top) and p_t (bottom) unfolded with the single-track and interaction MC samples (left and right, respectively).

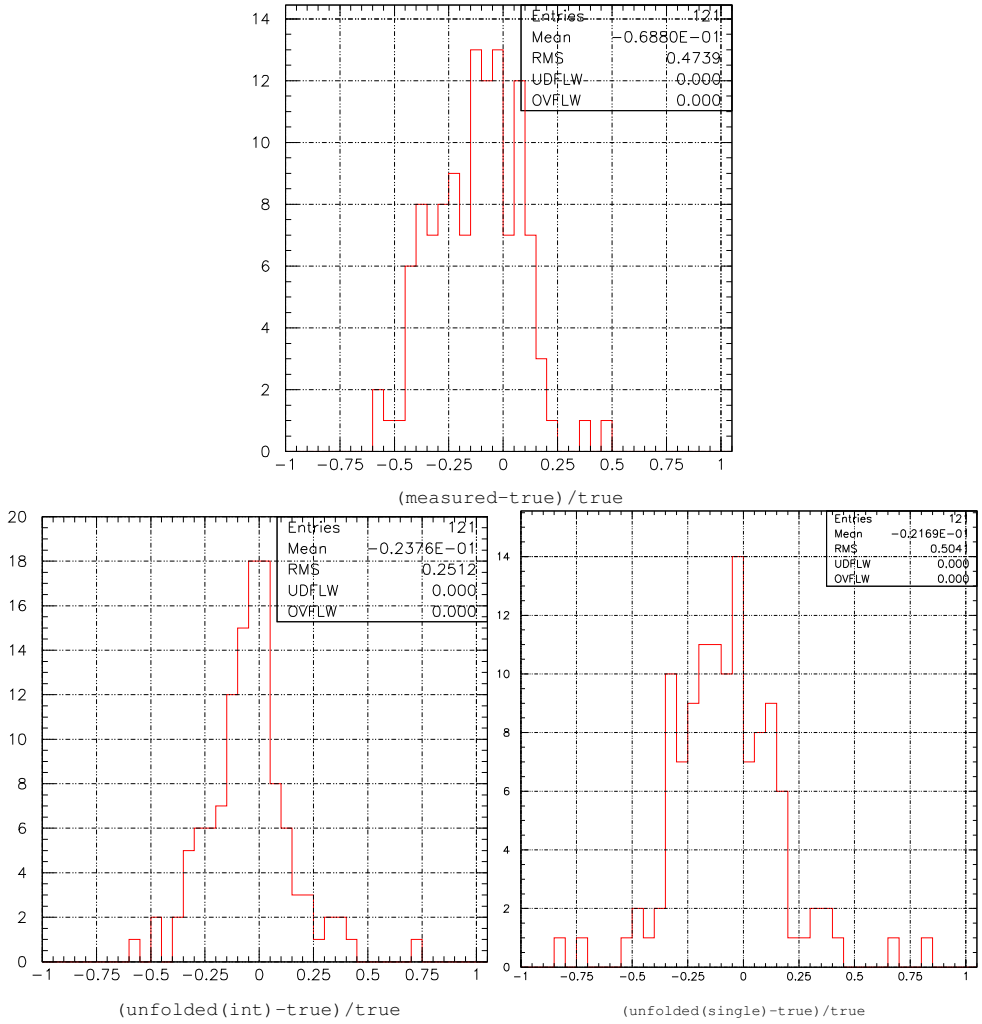


Figure 64: Comparison of the quality of the unfolding result. The histograms show the relative difference between the true and measured/unfolded bin content for the 2D histogram on which the unfolding is actually performed. Only cells with more than 20 events in the true histogram are considered. It can be seen that the measurement (top) indeed has larger errors and is slightly biased to lower values due to inefficiencies. The unfolding with single-track events (bottom right) yields a result with similar width but better centred around zero. The unfolding result obtained using the interaction MC sample (bottom left) is again centred around zero and has a much smaller width than the others.

5.3.6 Conclusion

The unfolding with each MC sample yields good results. Nevertheless, as can be seen in the comparison in Fig. 65, the interaction sample is slightly more accurate. Most likely the reason for this is that not all kinematic areas of the interaction spectrum are covered well enough.[25] Of course, in the interaction sample (which is of ten times the statistics of the mock data sample) all areas are covered naturally. Using *both* MC samples at the same time yields results similar to the ones obtained with the interaction sample only.

For the unfolding of real data the following can be concluded.

- Using MC samples either based on single tracks or on interaction events provides similar results. It might be possible to use a mixture of both or to check the consistency between results obtained with both samples.
- In most cases, a uniform distribution can be used as starting point for the iteration.
- The determination of the number of steps for the iteration can be left to the program. A few iterations less or more than suggested do not change the results significantly. A very low (< 5) or high (> 30) number of iterations is usually caused by problems in the data or MC sample (binning, statistics, etc.) or a non-suitable starting distribution.
- Care must be taken that both causes and effects histograms contain the full distributions without having to cut a significant amount of events. If this is not possible, the unfolded content of the bins near the boundaries are not correct. Most likely this problem can be overcome by using variable bin widths for the histograms so that long tails of distributions can be included.
- The errors provided by the unfolding are underestimated. A global correction of the errors by a factor of three should be applied when unfolding real data.

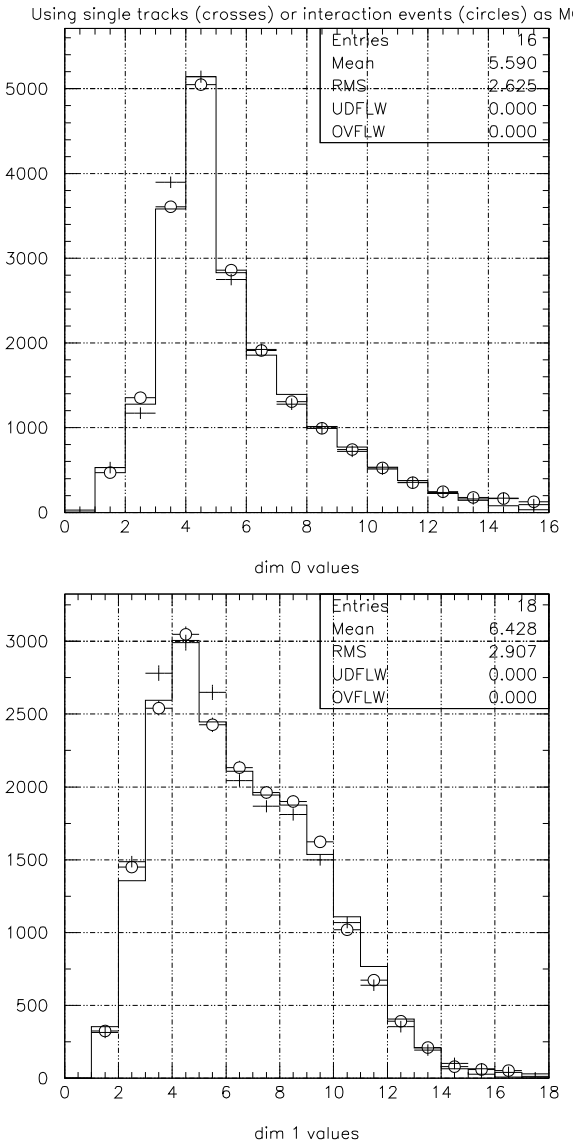


Figure 65: Comparison of the unfolding results of the same measurement using different MC samples. In most bins, the interaction sample reconstructs the truth with better accuracy. The scale on the horizontal axes is in bin numbers of the causes histograms. As before, dimensions 0 and 1 in the causes space are p_l and p_t, respectively.

[25] It is not trivial to generate a suitable MC sample with single-tracks. To cover well all areas in the smearing matrix, four quantities (two dimensions in both effects and causes space) have to be controlled since they decide about the cell of this track in the matrix.

6 Particle Production Yields

After having motivated and validated the unfolding procedure in the previous chapter, it should now be applied to real data.

The particle production yields will be determined for various target materials and momenta, and a comparison will be made.

6.1 Data Sets

In total, nine settings which comprise three different beam momenta (3, 5 and 8 GeV/c) each with three different targets (Be, Al and Ta) are selected. All runs are done with positive beam and thick targets, i.e. the target length is one nuclear interaction length λ_I. A summary of the data sets which have been chosen for this study is given in Table 10. As described earlier, for thick target runs triggering is performed on each incoming beam particle so that interaction trigger efficiencies are not an issue. On all data, basic data quality (DQ) issues like incomplete or crashed runs, detector or slow-control failure, etc. are taken into account. However, this does not include any channel-based DQ selection like a map of dead channels.

It should be pointed out that the following limitations of data reconstruction will put stringent boundaries to the scope of this study:

- The crosstalk of the TPC electronics (Section 4.3) is not corrected in the reconstruction. Crosstalk deteriorates the quality of the (geometrical) track reconstruction and reduces dramatically the accuracy of dE/dx measurements in the TPC.
- Still not clearly identified distortions are shifting reconstructed tracks. Currently being under study, the most likely source is distortions in the electric field. This would bias the reconstruction depending on the sign of the particle charge.
- TOF measurements between target (vertex) and barrel RPCs are not possible due to a lack of calibration for the RPCs.
- There is no special treatment of tracks crossing sector boundaries. Depending on the overlap of the track with these blind areas the reconstructed track parameters have large errors.
- No general vertex reconstruction/fit is available. This concerns not only primary but also secondary vertices.

In particular, particle identification (PID) in the large-angle region which is relying on TOF and dE/dx is concerned by these deficiencies.

To simplify the analysis, the following assumptions can therefore safely be made:

- Efficiency of the proton identification and purity of the proton sample for the beam PID are assumed to be 100%. Preliminary analyses of the performance of the beam detectors indicate that both is higher than 95%.
- Any background will be considered as zero. The estimation of background relies on the analysis of the *empty target* data samples which has not been done for this study.
- The re-interaction of secondaries in the target is neglected.

All data has been processed with version 5 release 12 of the HARP software. The reconstruction software and all calibration data are identical to the ones used for the *phase 1 calibration production* within the HARP analysis project.

6.2 Selection of Events and Tracks

The selection criteria for both events and tracks in selected events will be kept as simple and broad as possible. In detail, the following cuts are applied.

6.2.1 Event Selection

As already discussed, the incoming beam contains not only protons but in addition a considerable number of other positive particles emerging from the primary target. The beam composition (in addition to protons mainly muons, pions and electrons) strongly varies with beam momenta and the material of the primary target. Therefore, the detectors in the beam-line can separate protons from any other particle at any beam momentum which is used (see Section 2.2.1).

Thus, for the beam momenta of the settings which are considered the following cuts are applied: At 3 GeV/c, the beam TOF system (Section 2.2.1) provides a protons separation with several sigma. At 5 GeV/c, the TOF measurement is

p_{beam} [GeV/c]	target (Z,l)	setting no	run range	date	number events	p.o.t.
3	Be (4, 407 mm)	777	18992-19017	09/2002	839187	166323
5	Be (4, 407 mm)	796	19019-19034	09/2002	735889	233825
8	Be (4, 407 mm)	794	19034-19052	09/2002	831708	386165
3	Al (13, 394 mm)	420	13357-13372	05/2002	1065641	257065
5	Al (13, 394 mm)	407	13042-13060	05/2002	1044905	372310
8	Al (13, 394 mm)	398	12944-12959	05/2002	1174983	657487
3	Ta (73, 111 mm)	279	8944-8989	10/2001	1029152	254948
5	Ta (73, 111 mm)	409	13079-13090	05/2002	861387	308538
8	Ta (73, 111 mm)	404	12982-13003	05/2002	1156963	671017

Table 10: Summary of the nine data sets which are considered.

complemented by the pulse-height measurement in beam Cherenkov B (BCB) and at 8 GeV/c both beam Cherenkovs, BCA and BCB (Section 2.2.1) perform the particle identification.

The track parameters of the incoming proton are measured using the MWPCs (Section 2.2.1). To assure that the each beam particle hits the target, its track is extrapolated to the z-position of the target. Events in which this extrapolation does not reach the target (diameter 30 mm) are discarded.

Moreover it has to be assured, that there was *only one* incoming particle. Therefore, also events in which more than one TOF measurement was done or more than one MWPC track was reconstructed are discarded.

After applying these cuts it will be possible to normalise all results to the number of protons on target (p.o.t.). The number of events with one proton on target in each data set is given in Table 10.[26]

6.2.2 Track Selection

The selection of tracks in the TPC is uniquely aimed to reduce the amount of "bad" tracks in the sample. As such the following tracks should be discarded:

- Tracks not emerging from the target. This is mainly done by requiring the $z0$ of the track (for the conventions on track parameters see Section 5.3.1) to be within the length of the target. Since the reconstruction of track parameters in the TPC is not yet done with the final accuracy, a margin of 100 mm will be allowed at the upstream and downstream ends of the target.
- "Ghost tracks". If a track is reconstructed and fitted with only five points or less it will be discarded. Note that the maximum number of points in a track is equal to the number of pad rows of the TPC, i.e. 20. A track is also removed from the sample if an error occurred during the fit (track radius too large or too small) or the fit quality is extremely bad.
- Low-p_t tracks. Tracks with a transversal momentum with less than 20 MeV/c are discarded.

This leaves a sample of mostly pions, protons and electrons without further PID. From preliminary MC studies it can be assumed that the *direction* of the reconstructed momentum is quite accurate while its *amount* (and therefore also p_l and p_t) is not. If this is the case, the unfolding should provide already reasonable results.

6.3 The MC Samples

The MC samples used for the final unfolding are basically similar to the ones for the mock data unfolding (see Section 5.3.4). Again, two different samples are produced: one containing full interaction events, the other one single pion tracks distributed over the full phase space. In all new samples the TPC electronics crosstalk is modelled, including the correct electronics gain for each pad and a map of dead pads. However, at the time of this study, the exact crosstalk parameterisation is only available for one of the six sectors of the TPC

[26] The data analysed here does not necessarily contain the complete HARP data set for a given setting, but a large fraction of it.

pad-plane. It is assumed that – by construction – the crosstalk in the electronics of other sectors is similar.

Interaction Sample
The interaction sample is contains 434400 events of 3 GeV/c protons on a thick Tantalum target. The beam is slightly shifted (x = -0.3 mm, y = 1.6 mm) and smeared (σ_x = 4.7 mm, σ_y = 5.1 mm) according to MWPC measurement of the real 3 GeV/c beam. The beam direction is parallel to the z-axis. Regardless of the thickness of the target, the MC forces an interaction to happen in each event. In total, 510226 tracks are taken from this sample.

Single-track Sample
This sample contains 615545 single pion tracks distributed uniformly over total momenta from 150 to 2150 MeV/c. The distribution in the azimuthal angle ϕ is uniform, too, while the cosine of the polar angle θ is Gaussian distributed with a mean of cos 70 and a sigma of cos 55. The pions are injected at the centre of the coordinate system, i.e. the middle of the TPC where the coordinates x and y are Gaussian smeared with a width of 5 mm.

If not mentioned otherwise, a merger of both samples will be used for the unfolding.

6.4 Unfolding Parameters

To find the best parameters for the unfolding of all nine data sets, the Tantalum 5 GeV/c sample has been studied in more detail. Taking into account the conclusions drawn from the mock data unfolding, the optimal variables, binning, number of iterations and error treatment are determined.

6.4.1 Considerations on Variables and Binnings

The choice of variables and binnings in the causes space is determined by the desired quantities and distributions for the result. On the obtained particle production yields various cuts on the polar angle θ and on the transversal momentum p_t should be applied. Thus, those two quantities will be the ones used in the causes space.

Causes (Results)
The bin-width for p_t will be fixed at 50 MeV/c, which roughly corresponds to the expected resolution. The p_t range of interest basically extends from 0 to about 600 MeV/c. However, for the unfolding it is recommended to exclude problematic areas or at least to chose the ranges wide enough that they do not deteriorate the regions of interest. Towards lower p_t, the resolution, purity and efficiency of the track reconstruction starts to decrease below approximately 100 MeV/c. Those tracks have a small bending radius and might spiral in the TPC. At the same time the probability of crossing a sector boundary increases. In addition, the average number of points used to build a track is lower because the projection of the track onto the pad-plane does not cover all pad rows. At higher p_t (> 800 MeV/c), tracks become straight and small errors of the points in the innermost or outermost rows can bias the measurement of the bending radius (and therefore p_t) systematically towards smaller values. The upper boundary should therefore not be smaller than 800 MeV/c. Following the conclusions from the unfolding of mock data, it should also be taken into account that any upper p_t limit will cut into the distribution of the true (i.e. actually produced) transversal momenta. Is has been demonstrated that the unfolding produces unreliable results close to these boundaries which will therefore be fixed at 950 MeV/c, sufficiently far away from the region of interest.

In the other causes dimension θ, the bin width is not determined by the expected measurement errors. Instead, a binning is chosen which will allow to conveniently create p_t spectra for certain slices of θ. Slices finer than 100 mrad will not be needed and the accuracy of the θ measurement is certainly better than that. For the ranges, similar considerations as for p_t must be made: at low θ

			range		
	dim.	variable	from	to	bins
causes	0	θ [rad]	0.0	1.6	16
	1	p_t [MeV/c]	50	950	18
effects	0	θ [rad]	0.0	2.8	22
	1	p_t [MeV/c]	0	1200	22

Table 11: Parameters for the unfolding of the data. The naming convention using *causes* and *effects* and their two *dimensions* will be used throughout the rest of this chapter.

(i.e. in forward-direction), the TPC starts to become insensitive at less than 300 mrad. Low-angle tracks suffer from large reconstruction errors: like for small p_t, the number of points (i.e. covered rows) is rather low and due to the large drift length of the drift electrons, the affect of eventual field distortions plays an important role. Nevertheless, in order not to cut too much of the true distribution, the lower θ limit will be 0 mrad. Since the main region of interest for this study is between 20° and 60°, large θ of more than 90° (i.e. about 1.6 rad) are cut.

Effects (Measurement)
In the effects space, the choice of variables and binnings basically follows the causes. In this case both θ and p_t are also the two measured quantities which determine the track parameters of interest for this study. The ranges are chosen such that almost all tracks which can be measured in the TPC are included. The choice of the number of bins is mainly limited for practical reasons. Too many bins need too much computer memory and a too long time to calculate the complete unfolding results. This is not critical since the results are rather insensitive to this binning.

The unfolding parameters following these considerations are summarised in Table 11.

6.4.2 Number of Iterations and Errors

Iterations
The problem of finding the number of iterations which has to be performed before the unfolding result is stable is considerably different from the one for the mock data unfolding. In the mock data, there were no problematic areas, the number of events which were cut at the boundaries of distributions was small and the MC perfectly described the "data". Due to these features, the unfolding converged for all cells of the unfolded histogram. However, none of this is the case for real data. This results in unstable and non-converging bin contents in the unfolded histogram. This effect is illustrated in Fig. 66. In the first three and the last bin in θ the unfolded bin contents run away with an increasing number of iterations. Here it is crucial to notice that this unwanted behaviour is limited to the bins where it was anticipated, namely at the boundaries of the histogram. As explained above, the binning is chosen in a way that these areas are not in the main region of interest. However, also here a (considerably smaller) drift of bin contents can be observed. This is best seen in the θ slices of p_t shown in Fig. 67. Thus, there is no real convergence of the bin contents, which makes it difficult to find the point where no further iterations should be performed. In this case, it seems to be most natural to stop the iterations if the difference of the results of two successive steps remains stable. This change can then be attributed to the constant drift. In order to do this, the unfolding program computes a χ^2-comparison between the results of the steps. While not being interested in the absolute value, it can be seen in Fig. 68 that indeed this χ^2 remains rather constant after about five steps. With higher step numbers it starts to oscillate and in the results histogram tracks are reshuffled into a few preferred cells confirming the trend shown in Fig. 67. It should be mentioned that by construction, this unfolding method is not supposed to *converge*. As pointed out in [63], the best way to decide when to stop the iterations is to consider the differences between the results of two successive steps. Carrying out hundreds of iterations leads unavoidably to results dominated by statistical fluctuations, even in the case of the mock data unfolding.

This problem does not require further investigations because it is expected to disappear once the "problematic" areas in the unfolding histograms are eliminated. This will be the case with the upcoming versions of the TPC track reconstruction (including PID and crosstalk corrections) and the TPC MC (including the parameterisation of field distortions and the crosstalk in all sectors). Thus, the number of iterations to be performed will be set to five if not otherwise stated.

Errors
As discussed earlier (Section 5.3.6) the errors provided by the unfolding itself are underestimated, which would require further investigations to understand. While being aware that this treatment can only be preliminary, all unfolding errors are scaled by a factor of 3.0.

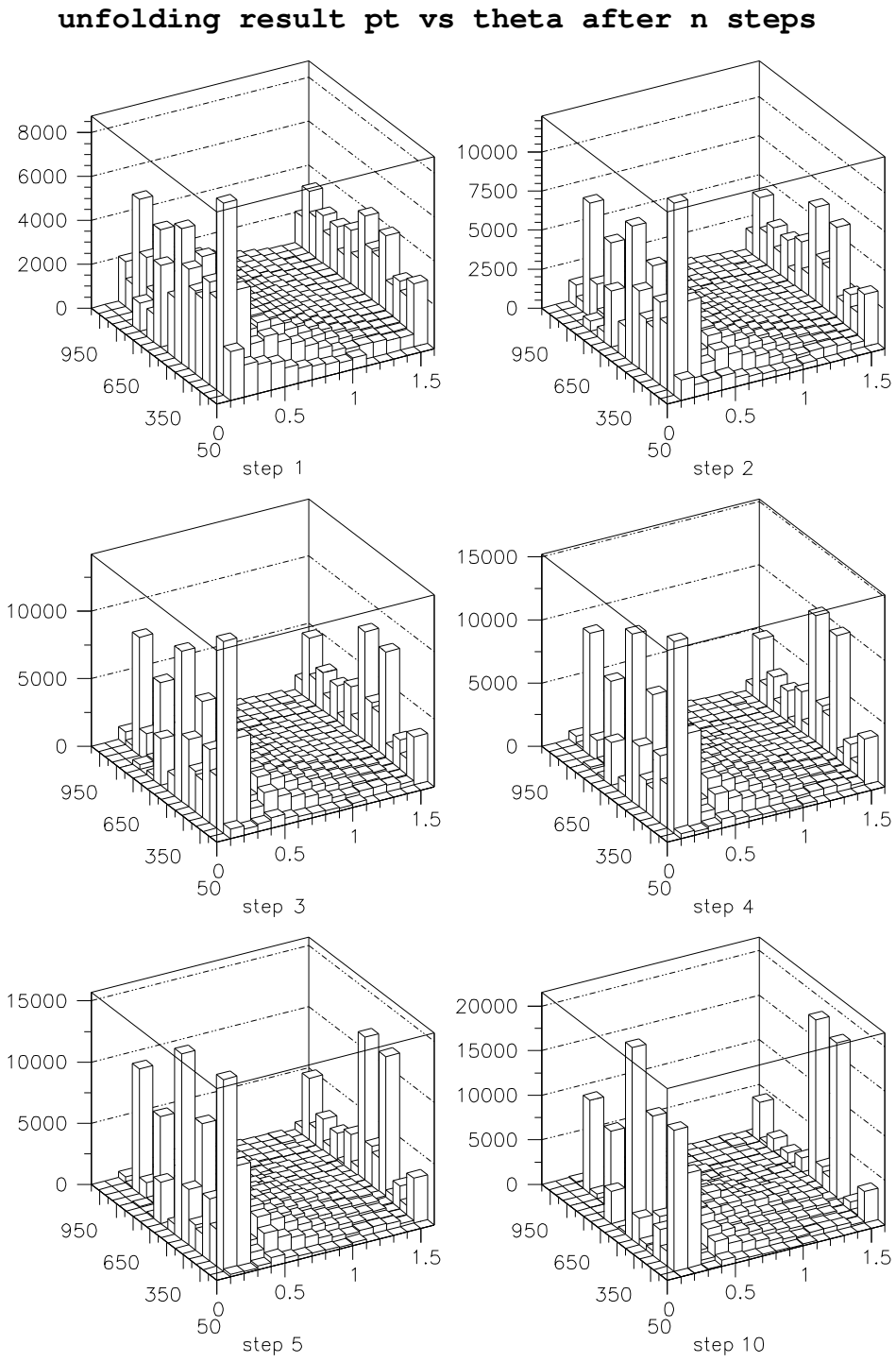


Figure 66: Visualisation of the unfolding result after 1, 2, 3, 4, 5 and 10 iterations. The pathologic areas where the bin contents remain zero or grow each step are restricted to the first three and the last slice in θ. Smoothing between steps has therefore not been done. The θ slices are shown in Fig. 67.

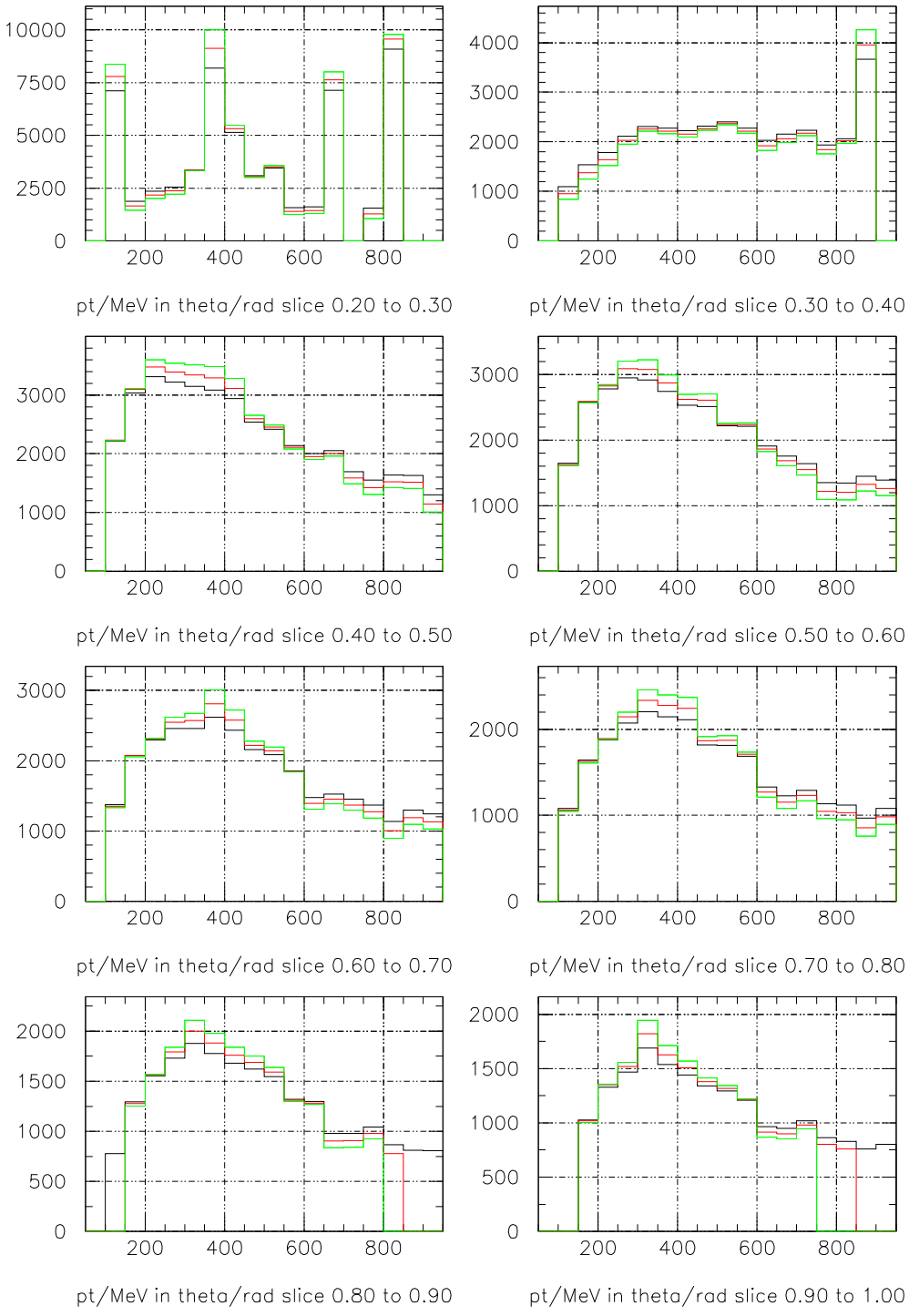


Figure 67: Different θ slices of p_t after 4 (black), 5 (red/grey) and 6 steps (green/light-grey). The drift which can be seen continues for more than 100 steps resulting in a completely unusable histogram. All tracks are then accumulated in a few preferred cells.

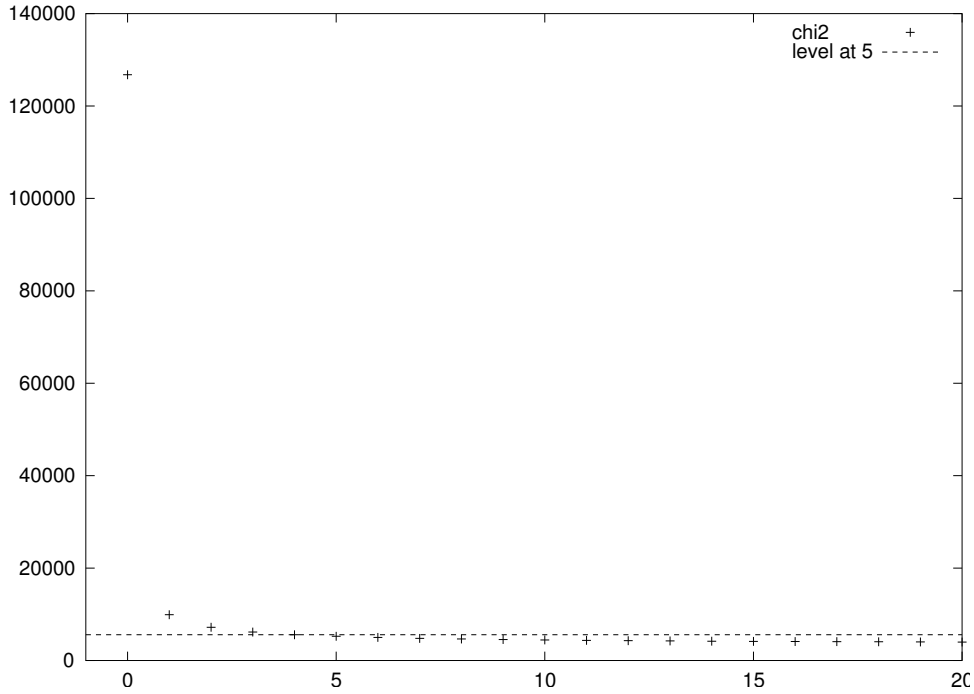


Figure 68: χ^2-comparison of unfolding results after step n and $n-1$ for the first 15 steps. It can be concluded that most parts of the result remain rather stable after the first iteration. The horizontal line marks the level of step 5 which is the number of iterations to be performed.

6.5 p_t Spectra and Particle Production Yields

As final result, the p_t spectra and the integrated particle yields are shown.

The p_t spectra are divided into ten θ slices each of 100 mrad width from 300 to 1300 mrad. They are shown in Figs. 69 and 70. Each row contains the spectra for one θ slice; each column the data for one beam momentum (3, 5 and 8 GeV/c from left to right). The graphs for the three target materials are overlayed (Ta: black, Al: red/grey, Be: green/light-grey). All scales are identical and extend from 50 to 800 MeV/c in p_t on the horizontal axis and 0 to 0.03 in particles per proton on target (p.o.t.) on the vertical axis.

In Fig. 71, the final particle production rates for all nine settings are shown. They correspond to the sum of all particles found in a certain kinematic area defined by the polar angle θ and the transversal momentum p_t. In detail, three different areas are defined:

- The full region of interest for this study. This comprises $0.3 < \theta < 1.4$ rad and $50 < p_t < 600$ MeV/c (top of Fig. 71).
- A region limited in p_t, comprising $0.3 < \theta < 1.4$ rad and $50 < p_t < 300$ MeV/c (middle of Fig. 71).
- A region limited in θ, containing $0.3 < \theta < 0.6$ rad and $50 < p_t < 600$ MeV/c (bottom of Fig. 71).

All error bars follow from the full covariance matrix of the unfolding result. In case of the plots in Fig. 71 they are too small to bee seen.

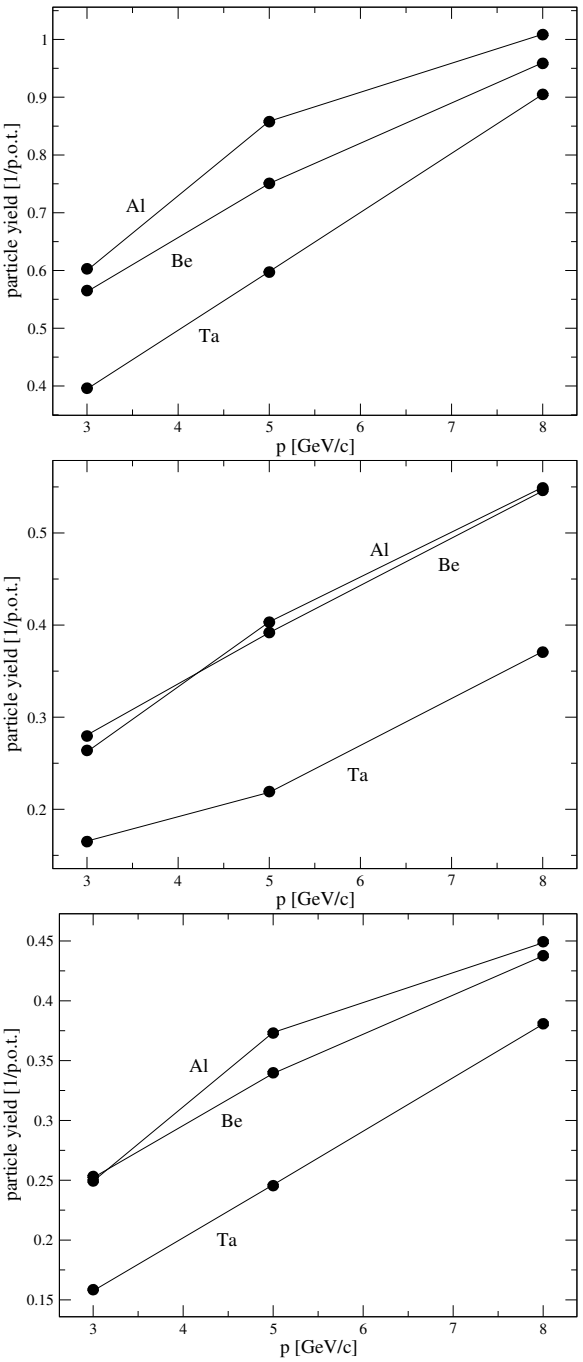


Figure 71: Final particle production yields for all nine settings. Top: $0.3 < \theta < 1.4$ rad and $50 < p_t < 600$ MeV/c, middle: $0.3 < \theta < 1.4$ rad and $50 < p_t < 300$ MeV/c, bottom: $0.3 < \theta < 0.6$ rad and $50 < p_t < 600$ MeV/c.

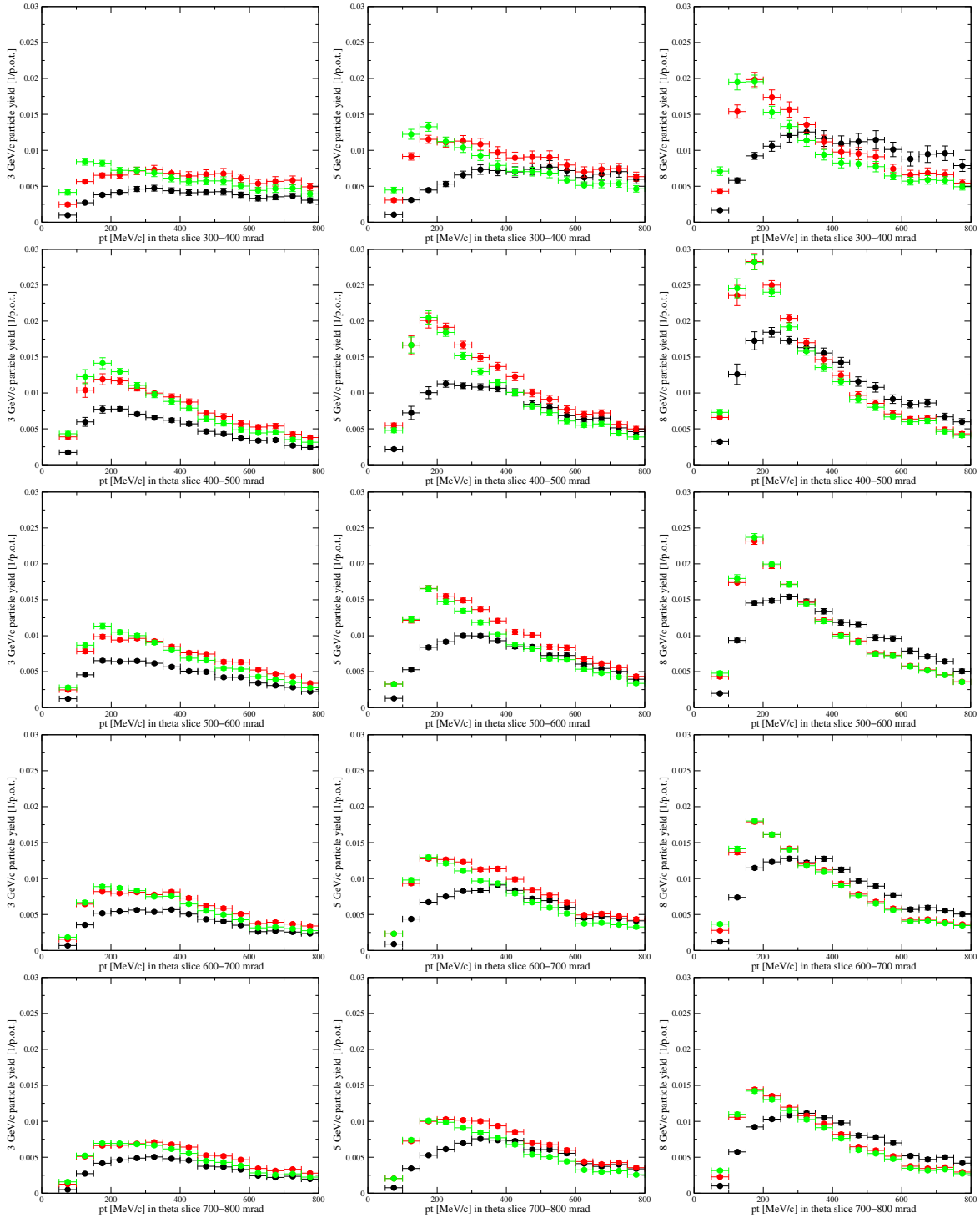


Figure 69: p_t spectra for θ slices from 300 to 800 mrad (see text for further explanations).

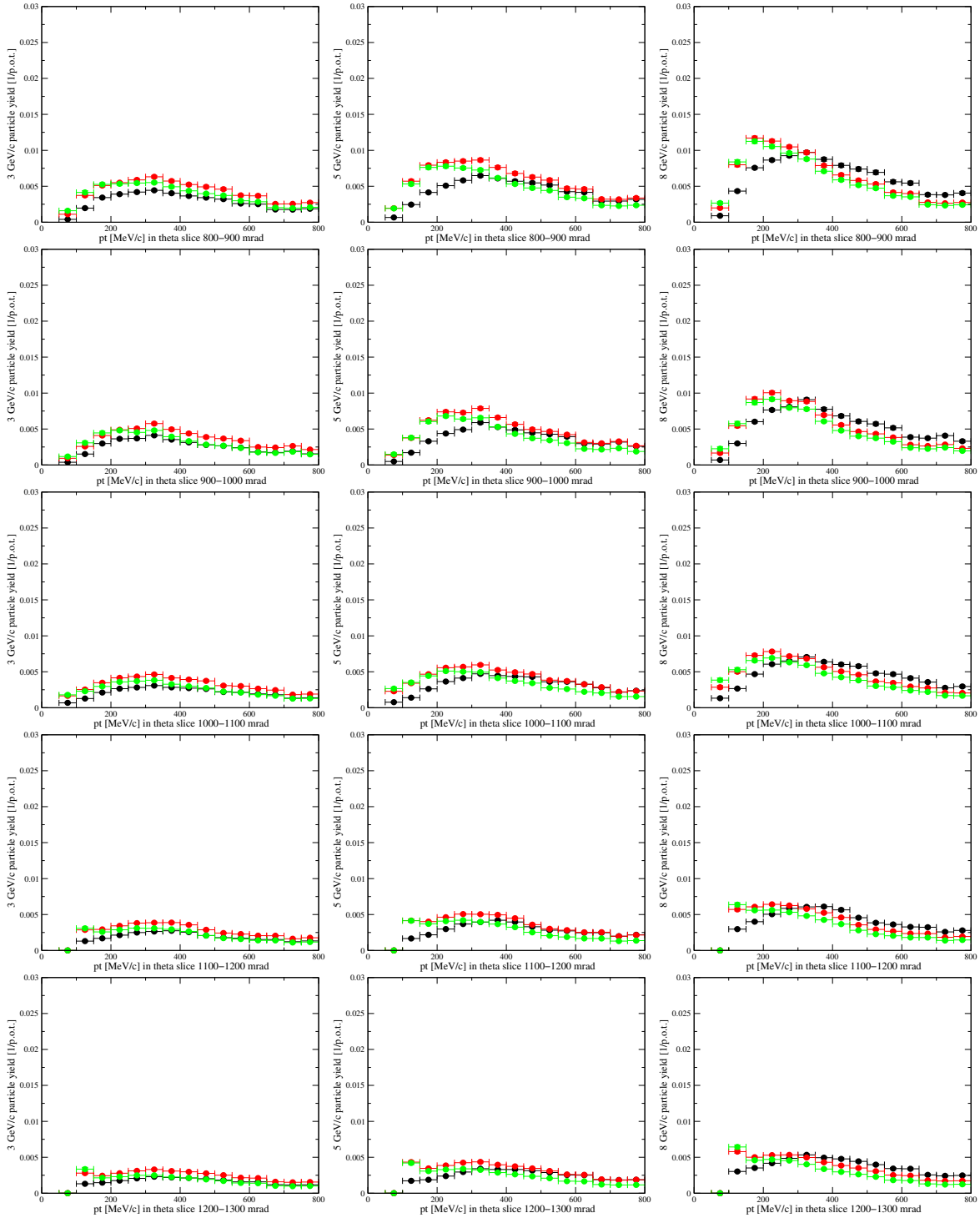


Figure 70: p_t spectra for θ slices from 800 to 1300 mrad (see text for further explanations).

6.6 Conclusions

From the results shown in Figs. 69, 70 and 71 the following can be concluded:

Beam Momentum
For all studied settings, the particle production rates scale almost linearly with the beam momentum. Only in case of the Al target, the rates at 3 GeV/c seem smaller than expected.

Target Material
The integrated yields are smallest for Ta while the highest are the ones for Al. Surprisingly, the values for Be are similar to those of Al but slightly smaller.

Between the "light" targets Al and Be on one side and the high-Z Ta on the other, there are considerable differences in the dependence of the yields on p_t:

Above 400 MeV/c, the p_t spectra have a similar shape for all beam momenta and targets. It is interesting to note that in this region, the rates for Ta are lower than those of Be and Al with a 3 GeV/c beam. At 5 GeV/c, they are about equal and at 8 GeV/c, the Ta yields are higher than for the light targets at all θ.

Towards lower p_t, the spectra of light and heavy targets separate between 200 and 300 MeV/c. In particular in the forward direction, the distributions for Be and Al have much higher peaks than for Ta. Most likely this is an effect of the reinteraction probability in the target which is considerably higher for Ta. This would also explain that this effect is strongest in the forward direction where all secondaries have to travel a longer path inside the target. However, in order to fully explain this surprising behaviour, particle identification would be needed. For obvious reasons, attempts to estimate the particle composition of the spectra using the predictions from the MC have not been made.

Angular Dependance
With respect to θ, the yields for all targets and momenta peak between 400 and 500 mrad. Towards larger angle, they decrease steadily down to 10–20% of the maximum value at 1200 mrad.

It should be pointed out that already the results obtained for the θ slice of 300–400 mrad might be affected by the lower resolution and reconstruction efficiency of TPC tracks at small angles.

p_t Spectra
Independent of θ, the p_t distributions have peaks at about 200 MeV/c for the light targets and between 200 and 400 MeV/c for Ta. The position of the peaks shifts slightly towards lower p_t for smaller beam momenta.

Once again it should be pointed out that the reliability and the scope of this study is limited by the missing particle identification and the spatial track reconstruction not yet being at its final performance. As a consequence of this, the MC is not yet perfectly describing the detector, thereby creating problems for the unfolding. For these reasons, no final conclusions concerning the best target and best proton beam momentum for a Neutrino Factory will be possible now.

7 Summary

In this study a first analysis is presented which extracts large-angle particle production rates from data taken by the HARP experiment. The analysis is carried out for nine different settings in order to compare the results in view of their dependence on the beam momentum and the target material.

To achieve this goal, a fully-featured simulation for the relevant subdetector, the TPC, has been developed. This simulation was then used to unfold the measured momentum and angular distributions of secondaries. The unfolding scheme has been validated with simulated data.

Due to the lack of particle identification at the time of this study, the results are not analysed in view of the best parameters for the target station of a Neutrino Factory. Nevertheless, the ratios and trends of the results presented here should already be reliable and have to be confirmed by future analyses.

Appendix

A Trigger Electronics Layout

The trigger system is implemented in standard NIM, CAMAC and VME electronics. In total, 10 crates are used as partially shown in the photograph 72.

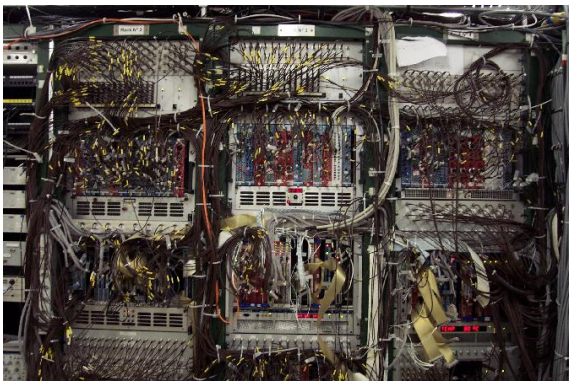


Figure 72: Photograph of the trigger racks and crates in the counting room.

Figures 74 to 80 describe the details of the implementation of the trigger electronics. The symbols which are used are the following and shown in Fig. 73: (a) a signal entering or leaving the trigger system or generated/used in a different subsystem of the trigger (i.e. on a different plan), (b) any kind of signal type converter, (c) a dual timer, usually of type NIM CAEN N93B, (d) a simple cable delay, (e) a logic fan in/fan out, usually of type NIM LeCroy 429, (f) a (strobed) coincidence unit, usually of type NIM LeCroy 365, 465 or 622 and (g) any type of discriminator or shaper. Signal lengths, thresholds and signal names are given as they are available.

Since a few parts of the system are extremely sensitive to timing issues, the length of cable connections (in ns) are given as well. The numbering scheme of modules/signals is *rack-crate-module-channel*. During maintenance and eventual repairs these plans served as reference. They are not 100% complete but nevertheless cover the most sensitive parts of the trigger and DAQ system.

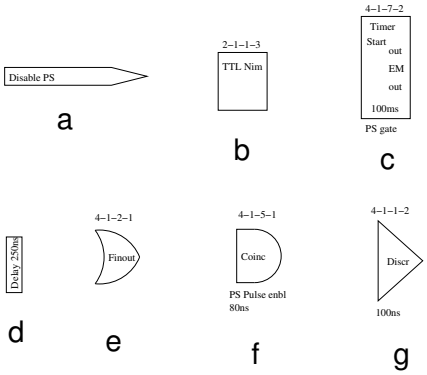


Figure 73: Explanation of symbols used in the electronics layouts.

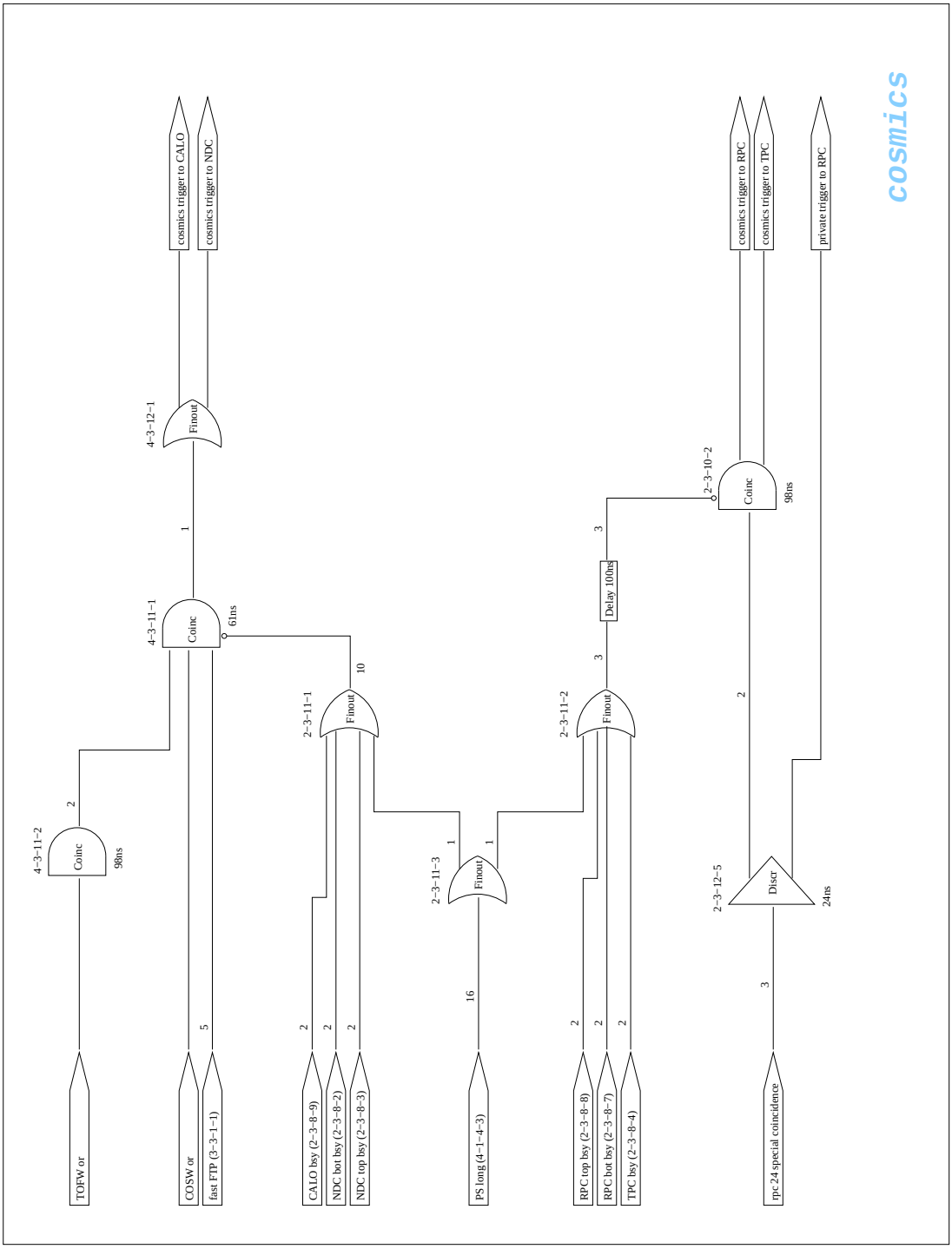


Figure 74: Electronics layout of both cosmics triggers. A coincidence of adjacent barrel RPC modules triggers TPC and RPCs. The forward cosmics trigger (calorimeter, TOF wall and NDCs) for horizontal cosmics is formed by a coincidence of the cosmics wall, TOF wall and FTP.

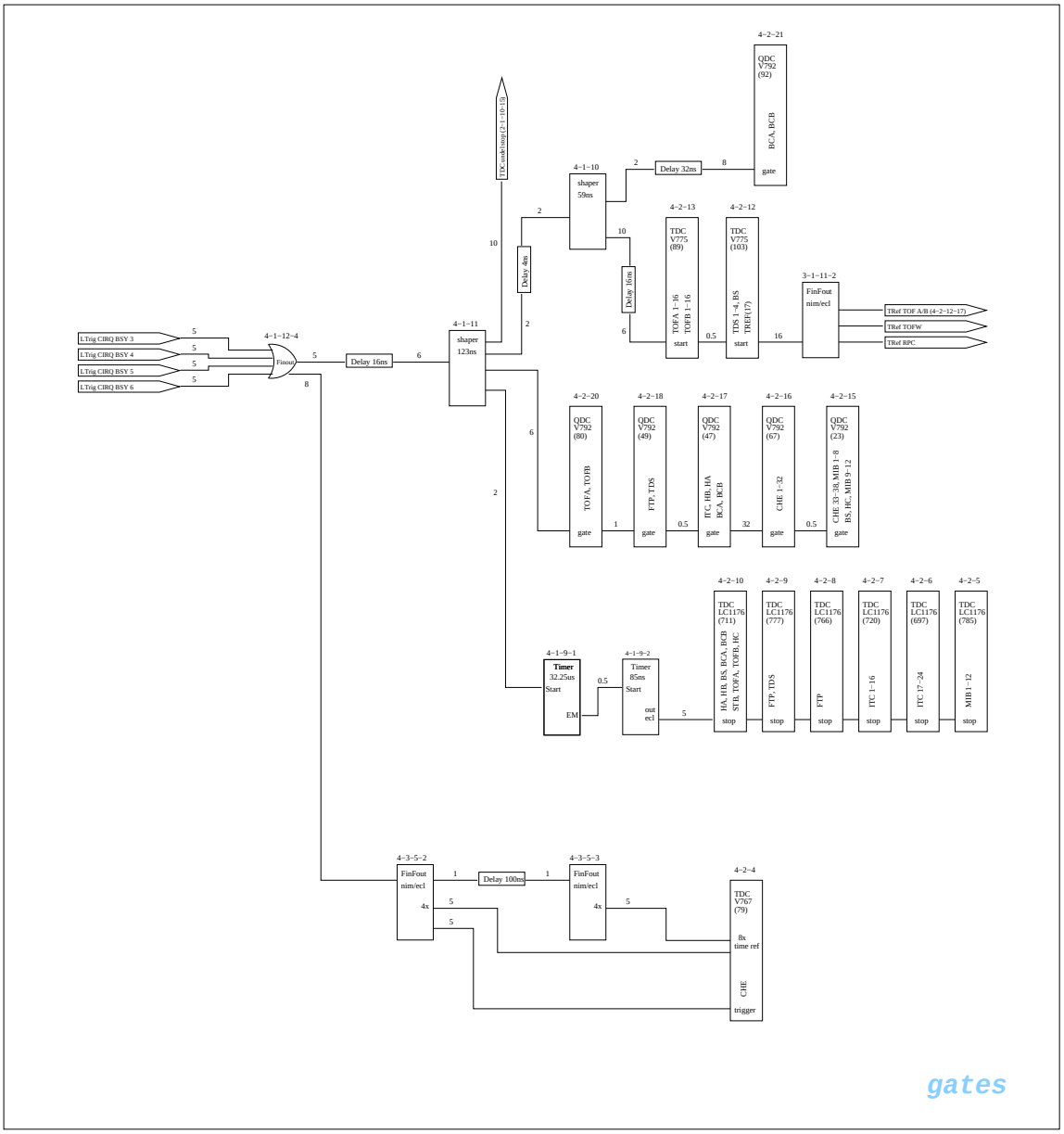


Figure 75: Generation of ADC gates and TDC start/stops for the readout equipment in the trigger/beam VME crate. The outputs of the CIRQ are shaped and suitably delayed to serve as trigger signals with stable duration and timing.

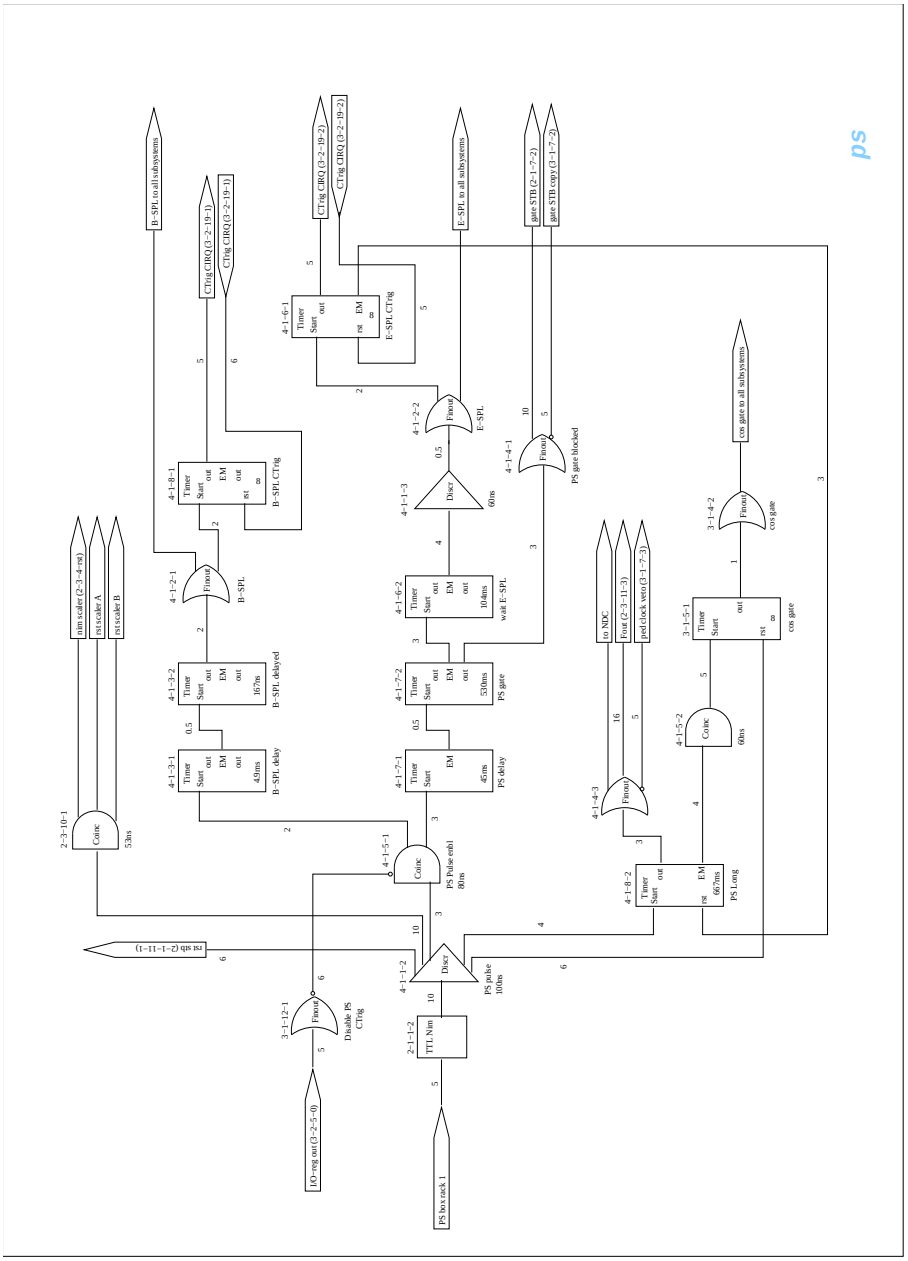


Figure 76: The synchronisation with the PS accelerator. Before each spill PS sends a synchronisation signal which is used to generate gates for physics. The begin-spill and end-spill signals are sent to the subdetectors to create cosmics and calibration gates between spills.

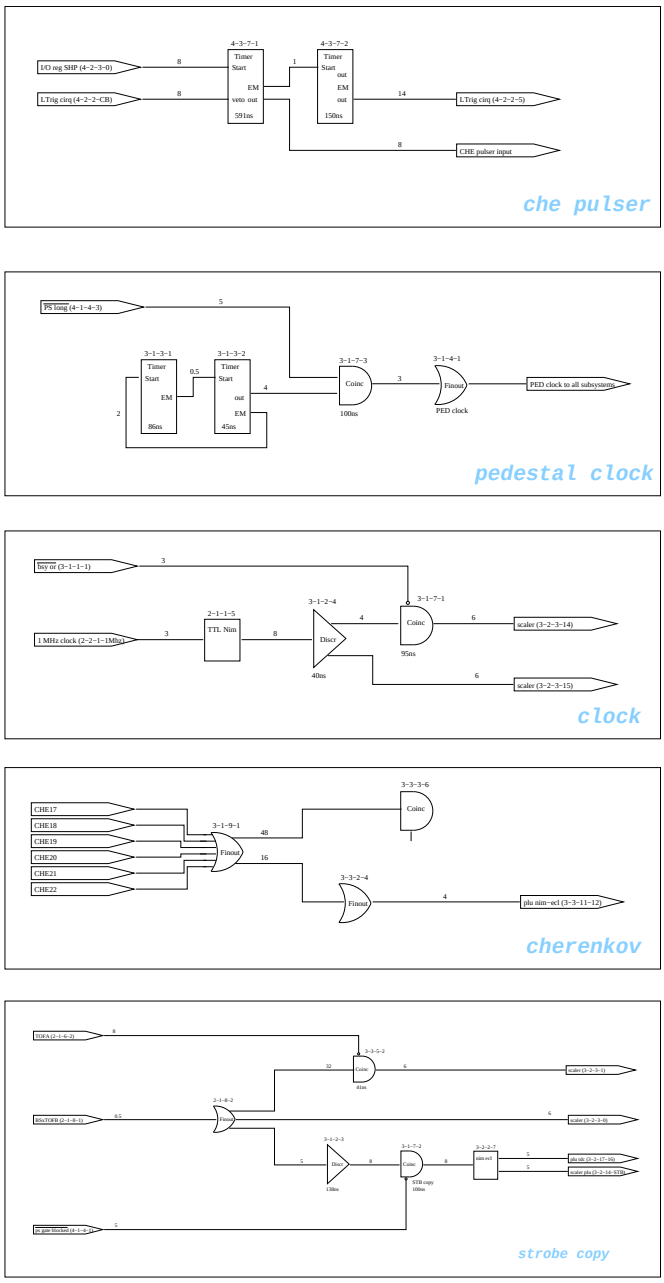


Figure 77: Electronics layout of minor separated systems. From top to bottom: Cherenkov pulser system (correctly timed to match the ADC gate), the pedestal clock, the MHz timing clock, the generation of the Cherenkov signal for the trigger and the strobe copy (generated to control the strobe timing).

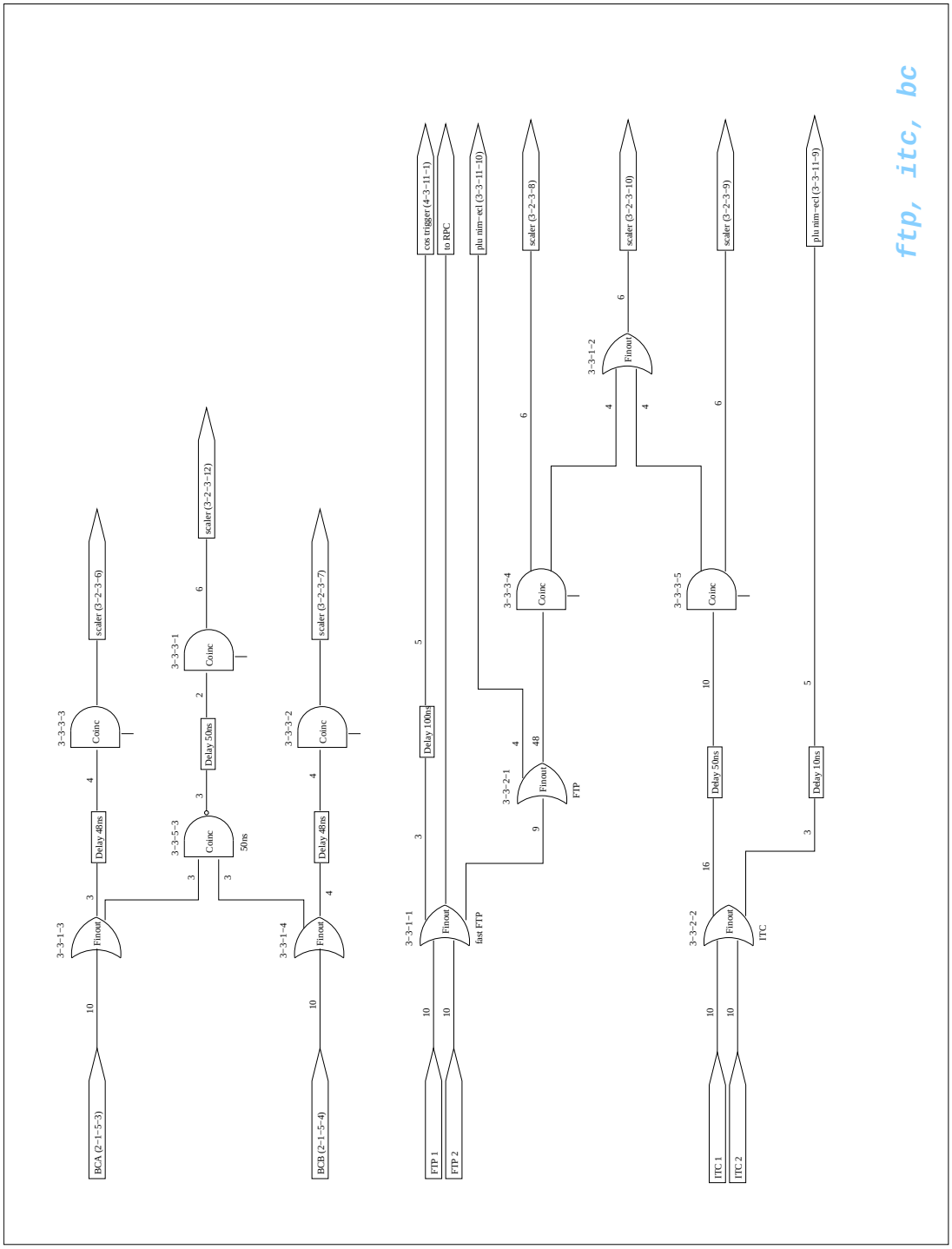


Figure 78: Before FTP, ITC and beam Cherenkov signals enter the trigger system (i.e. scalers and PLU) they are treated like shown. The summary signal of FTP is also used to generate the small-angle cosmics trigger and as calibration trigger for standalone RPC runs.

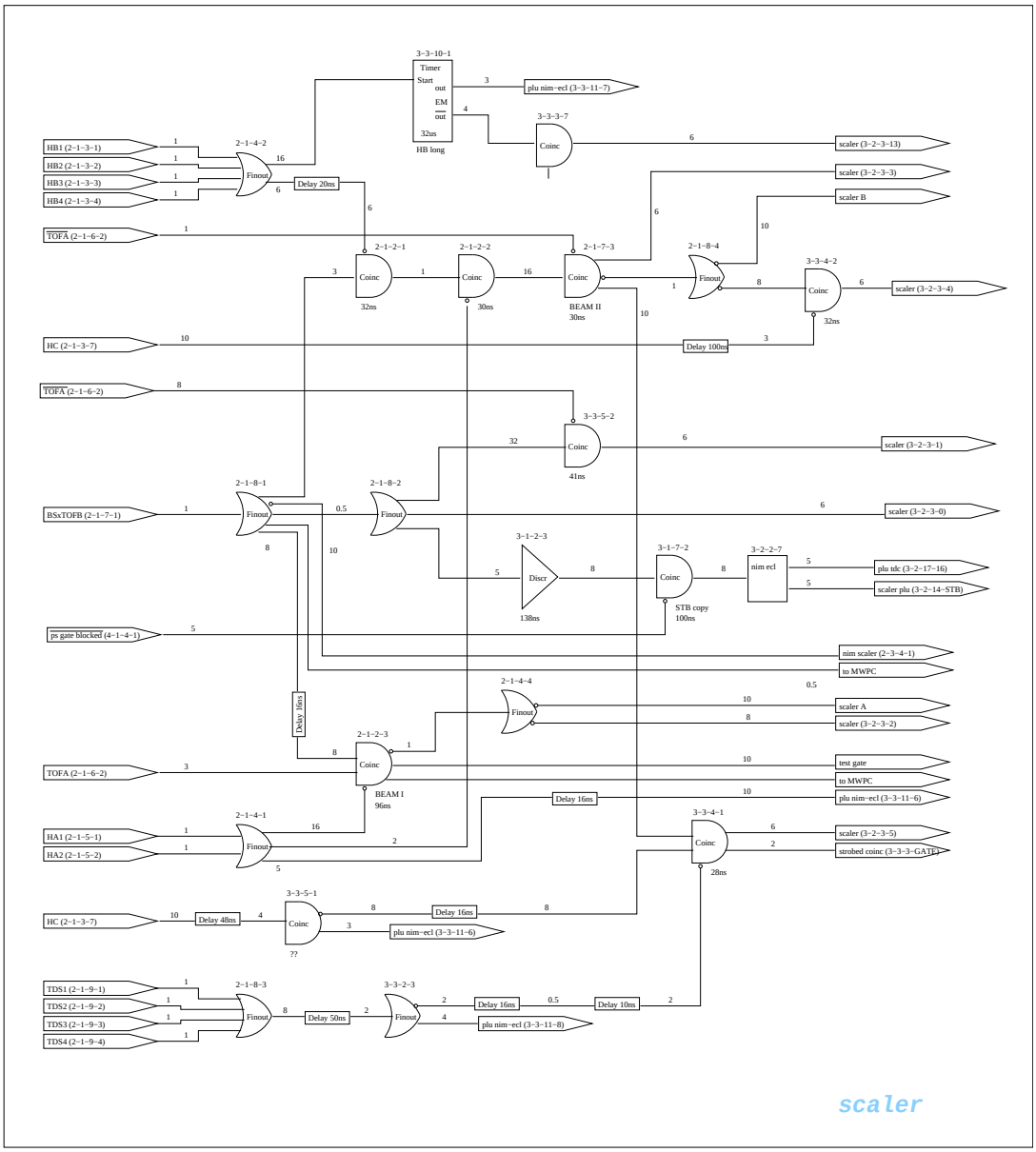


Figure 79: Generation of coincidences for the scalers. Respecting their different timings, the discriminated signals from the detectors are merged in various NIM logic units to provide coincidences read out by the VME scalers. In addition, coincidence signals are created and displayed in the control room of the experiment to inform the shifter about beam conditions.

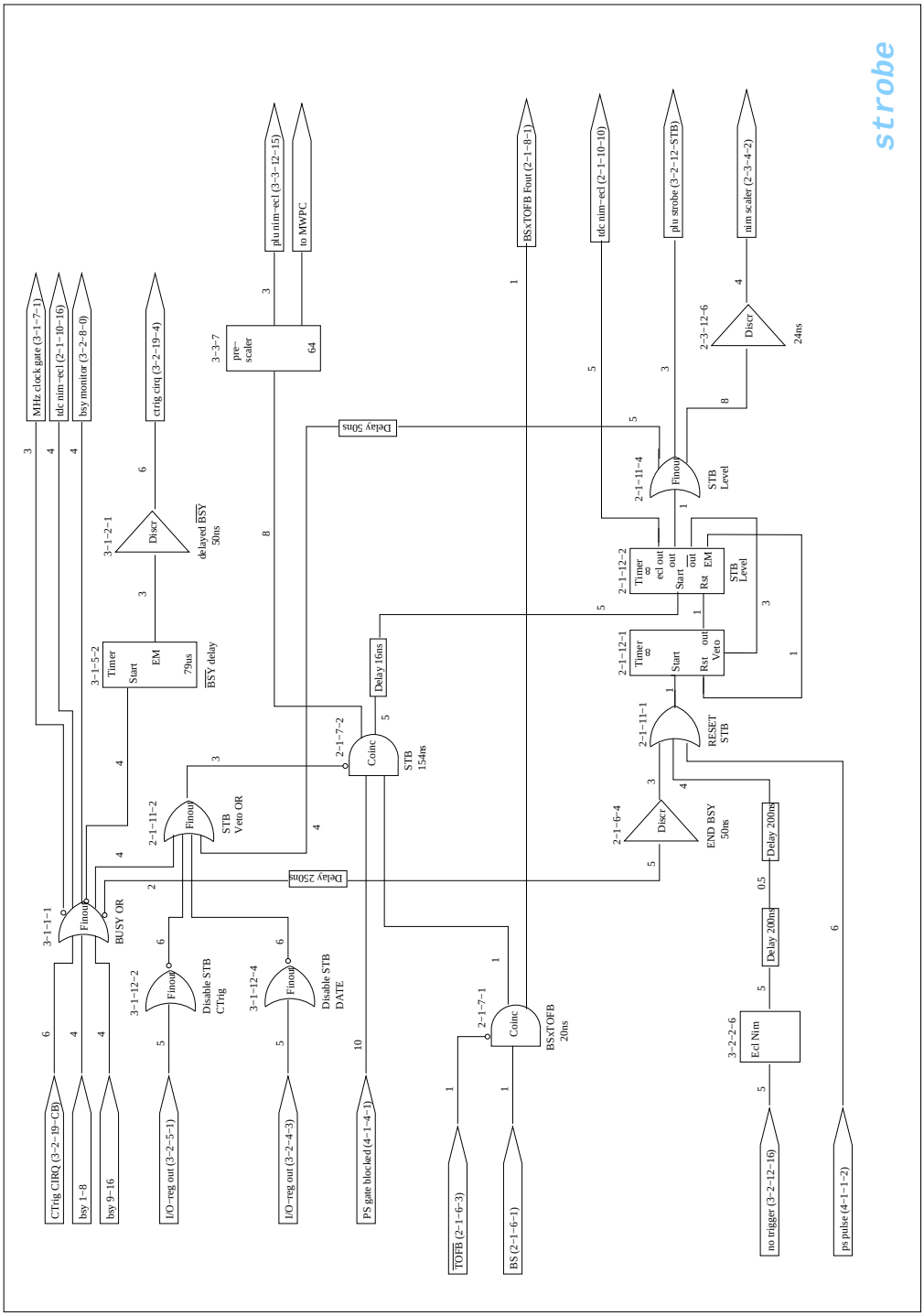


Figure 80: Strobe electronics layout. Basic input signals are the discriminated hits from TOFB and BS. A coincidence of those detectors is accepted as PLU strobe if the system is not busy and the DAQ control allows strobes. Moreover, strobes are only accepted if they happen during a spill. Strobe signals are counted by scalers at various levels of vetoing ('plain' strobes, 'missed' strobes during busy time, accepted strobes).

List of Figures

References

[1] Q.R. Ahmad et al., Phys. Rev. Lett. 87 (2001) 071301

[2] Y. Fukuda et al., Phys. Rev. Lett. 81 (1998) 1158

[3] K. Zuber, Phys. Rep. 305 (1998) 295

[4] R. Davis, D.S. Harmer and K.C. Hoffman, Phys. Rev. Lett. 20 (1968) 1205

[5] Y. Fukuda et al., Phys. Rev. Lett. 77 (1996) 1638

[6] J.N. Abdurashitov et al., Phys. Rev. Lett. 83 (1999) 4686

[7] P. Anselmann et al., Phys. Lett. B 314 (1993) 284

[8] N. Hata, P. Langacker, Phys. Rev., D 48 (1993) 2937

[9] R.A. Becker-Szendy et al., Nucl. Instrum. Methods Phys. Res., A 324 (1993) 363

[10] W.W. Allison et al., Phys. Rev. Lett. B 449 (1999) 137 and references therein

[11] M. Ambrosio et al., Phys. Lett. B 517 (2001) 59

[12] C. Athanassopoulos et al., Nim A 388 (1997) 149

[13] S.F. King, hep-ph/0105261

[14] M.H. Ahn et al., Phys. Rev. Lett. 90 (2003) 041801

[15] A. Bandyopadhyay, S. Choubey, S. Goswami, S.T. Petcov, hep-ph/0309236

[16] A. Bazarko et al., hep-ex/0009056

[17] M.V. Diwan et al., hep-ex/0211026

[18] Q.R. Ahmad et al., Phys. Rev. Lett. 89 (2002) 011302 and 011301

[19] K. Eguchi et al., Phys. Rev. Lett. 90 (2003) 021802

[20] M.M. Alsharo'a et al., hep-ex/0207031

[21] P. Huber et al., hep-ph/0204352

[22] M. Campanelli et al., hep-ph/0210192

[23] P. Zucchelli, Phys. Lett. B 532 166 (2002)

[24] D. Neuffer, μ^+-μ^- colliders, CERN yellow report 99-12

[25] B. Autin, A. Blondel, J. Ellis (eds.), Prospective study of muon storage rings at CERN, CERN yellow report 99-02

[26] K.R. Symon et al., Phys. Rev. 103 (1956) 6

[27] D.M. Kaplan, hep-ph/0001037

[28] S. Geer, hep-ex/0111016

[29] S. Gilardoni et al., CERN-NUFACT-Note 129

[30] S. Gilardoni et al., CERN-NUFACT-Note-86

[31] T. Anderson et al., A Feasibility Study of a Neutrino Source based on a Muon Storage Ring, FERMILAB-Pub-00/108-E (2000)

[32] R. Edgecock, J. Phys. G: Nucl. Part. Phys. 29 (2003) 1601

[33] M.G. Catanesi et al., Proposal to study hadron production for the neutrino factory and for the atmospheric neutrino flux, CERN-SPSC/99-35, SPSC/P315

[34] R. Engel et al., hep-ph/9911394

[35] I. Stancu et al., MiniBooNE collaboration, FERMILAB-TM-2207

[36] M.H. Ahn et al., K2K collaboration, Phys. Rev. Lett. 90 (2003) 041801

[37] J. Collot et al., Nucl. Instrum. and Meth. A 451 (2000) 327

[38] J. Pasternak et al., Nucl. Instrum. and Meth. A 472 (2001) 557

[39] S.Agostinelli et al., Nucl. Instr. and Meth. A 506 (2003) 250

[40] Martin Mass, Diplomarbeit, Universität Dortmund, 2001

[41] P. Gorbunov and L. Scotchmer, HARP Note MWPC-2, May 2001

[42] F.J. Braid, CERN Summer Student Report, September 2000

[43] B. Schmidt et al., Nucl. Instrum. and Meth. A 428 (1999) 299

[44] A. Lundborg, Electrostatic simulation of parts of the HARP TPC, HARP internal note, 25 August 2000

[45] J. Baechler et al., Front-End Electronics for the ALICE TPC-detector, CERN-ALI-98-022, 16 October 1998

[46] J. Dumarchez et al., Nucl. Instrum. and Meth. A 481 (2002) 339

[47] J.S. Graulich, private communication

[48] ALICE DATE User's Guide v3.7, ALICE Internal Note/DAQ, ALICE-INT-2000-31 v.2

[49] V451-CIRQ VME/VSB Interrupt Request Module, Francois Bal, ECP-EDA/DQ, February 8, 1993

[50] J. Panman et al., Nucl. Instrum. and Meth. A 427 (1999) 587

[51] L. Bonnet, VME-Buffered Pattern Unit, Chorus internal note, 1994

[52] SIS3600 VME Multi-Event Latch, SIS GmbH, Hamburg, Germany

[53] J.P. Dewulf, Programmable Logic Unit, Chorus internal note, 1994

[54] F. Sauli (edit.), Instrumentation in High Energy Physics, World Scientific Publishing, 1993

[55] G. Prior and G. Vidal-Sitjes, http://harp.web.cern.ch/harp/ Classified/Sub_detectors/TPC/documents/ mapping.ps

[56] W. Rolandi and L. Blum, Particle Detection with Drift Chambers, Springer Verlag, 1993

[57] R. Veenhof, http://r.home.cern.ch/r/rjd/www/Harp/

[58] L.C. Howlett, private communication

[59] G. Vidal-Sitjes, PhD thesis, Universidad de Valencia, Spain, 2003

[60] V. Blobel, Unfolding methods in High Energy Physics, CERN 85-02

[61] V. Blobel, The RUN manual, OPAL Technical Note TN361, 1996

[62] A. Höcker and V. Kartvelishvili NIM A 372 (1996) 469

[63] G. D'Agostini, NIM A 362 (1995) 487

[64] HARP status report, March 2002, CERN-SPSC/2002-013

[65] PAW Physics Analyis Workstation, CERN Program Library Q121, CERN, Geneva 1995

[66] G. Vidal-Sitjes et al., Status of the TPC analog electronics cross-talk measurement, HARP internal note, HARP-MEMO-03-006

[67] S. Borghi et al., TPC clusterization, HARP analysis note, 2003

[68] J.W.E. Uiterwijk et al., A C++ framework for track finding with 3-dimensional hits, in prep.

[69] M.C. Morone, PhD thesis 3418, Université de Genève, Switzerland, 2003

Acknowledgements

Of course this work would never have been neither started nor finished without the support of my thesis supervisors, C. Gößling, J. Panman and K. Zuber. They made it possible that I could spend three very instructive years at CERN and constantly supported me in a complementary way.

I also wish to express my gratitude to all collaborators of HARP and the scientific, administrative and technical staff of the Lehrstuhl für Experimentelle Physik IV of my university.

I am grateful of having had the pleasure to work with (in order of office distance) I. Papadopoulos, F.J. Braid, G. Prior, S. Piperov, C.H. Wiebusch, Z. Kroumchtein, I.M. Rusinov, P. Petev, E. Radicioni, P. Gorbounov, S. Giani, P. Zucchelli, G. Vidal-Sitjes, S. Borghi, J.S. Graulich, M.C. Morone, C. Detraz and J. Mulon, S. Robbins, L.C. Howlett.

Special thanks to J. Großheim and K. Zimnick.

This work was supported by the CERN Doctoral Student Programme.