Pro gradu -tutkielma
Fysiikka

Event Classification and Estimation of Low Mass Diffraction at the TOTEM Experiment at the LHC

Jan Welti
2011

Ohjaaja:	Kenneth Österberg
Tarkastajat:	Paula Eerola
	Kenneth Österberg

HELSINGIN YLIOPISTO
FYSIIKAN LAITOS

PL 64 (Gustaf Hällströmin katu 2)
00014 Helsingin yliopisto

HELSINGIN YLIOPISTO – HELSINGFORS UNIVERSITET – UNIVERSITY OF HELSINKI

Tiedekunta/Osasto – Fakultet/Sektion – Faculty/Section Faculty of Science	Laitos – Institution – Department Department of Physics

Tekijä – Författare – Author
Jan Welti

Työn nimi – Arbetets titel – Title
Event Classification and Estimation of Low Mass Diffraction at the TOTEM Experiment at the LHC

Oppiaine – Läroämne – Subject
Physics

Työn laji – Arbetets art – Level Master's Thesis	Aika – Datum – Month and year 03/2011	Sivumäärä – Sidoantal – Number of pages 64

Tiivistelmä – Referat – Abstract

The TOTEM experiment is a forward physics experiment at the Large Hadron Collider (LHC) at CERN. The goal of the experiment is to study the structure of the proton and proton interactions at high energies. This includes a precise measurement of the total proton-proton cross section and the measurement of the cross sections for individual processes like elastic, non-diffractive minimum bias and single and double diffractive. The experiment comprises three different types of detectors: Roman Pot silicon detectors for detecting protons scattered by very small angles (few micro radians) and T1 and T2 forward tracking telescopes for detecting charged particles from inelastic proton-proton scattering. Each detector has a digital readout.

The detectors for the T2 forward tracking telescope consist of three planes of gas electron multipliers. The radial coordinate is provided by the strip readout and the azimuthal coordinate and trigger are provided by the pad readout. Due to cross-talk and electronics noise, it is possible for the one particle to turn on many adjacent pads in a detector plane. Two different methods of clusterization of the digitized pad readout for the T2 detector are compared in order to determine the optimal way to recognize clusters corresponding to the ionization signal from one particle.

An event classifier is used to distinguish which process is leading to a specific inelastic event. This event classification is done with the charged particle tracks reconstructed from the T1 and T2 telescopes and the training of the classifier is done with simulated data, where the original process leading to the event is known. The classification efficiency and purity is studied using simulated data. Diffractive events can be distinguished with good efficiency and purity from non-diffractive in events that have reconstructed particles in both hemispheres. Events with reconstructed particles in only one hemisphere are mainly diffractive and do not classify well, which can be understood from simple physics arguments.

Not all diffractive events are detected because sometimes all final state particles are produced at pseudorapidities larger than the acceptance of T1 and T2. In order to determine the inelastic rate needed for the total cross section measurement, it is important to estimate the total number of inelastic events taking also into account the fraction of events that are not detected. These low diffractive mass events are mainly single diffractive, but a small fraction of double diffractive events also contribute. The number of missed inelastic events can be estimated by extrapolating the diffractive mass distribution in some suitable inverse power of the diffractive mass or by means of selecting events with only one "elastic-like" final state proton and no reconstructed particles in the forward tracking telescopes T1 and T2. The extrapolation is valid if the distribution is approximately linear, which is the case of events generated with PYTHIA for an extrapolation in the inverse mass squared. Adding the estimated number of events from such an extrapolation to the number of detected events, the ratio between the calculated and simulated cross section is measured to be 0.99±0.03.

Avainsanat – Nyckelord – Keywords

Säilytyspaikka – Förvaringställe – Where deposited

Muita tietoja – Övriga uppgifter – Additional information

Contents

1 Introduction

TOTEM is one of three small experiments at the Large Hadron Collider (LHC). TOTEM shares Interaction Point 5 (IP5) with the CMS experiment, which is one of four large experiments at the LHC. Interaction points are the positions on the LHC ring where the particle beams collide and where the experiments are situated. TOTEM studies the structure of the proton and intends to measure the total cross section for the collision of two protons (abbreviated as $p-p$) with the luminosity independent method as well as study elastic $p-p$ scattering and diffraction at the LHC at CERN.

TOTEM uses two types of forward telescopes to detect charged particles and Roman Pots to detect final state protons with small scattering angles with respect to the beam direction. The readout of the detectors is digital and the original event is reconstructed using software specifically designed for this purpose. One of the very first steps in reconstruction is the clusterization of hits. Two different methods for clusterization are presented and compared for TOTEM's T2 detectors.

Measurement of the total cross section requires knowledge about the inelastic rate and the diffractive mass distribution. To get information about either of these it is necessary to do event classification. Event classification aims at determining what kind of interaction happened given a set of detected variables from the final state. Event classification can be used to determine the relative rates of the different processes and used to filter out non-diffractive processes when calculating the diffractive mass distribution.

The studies presented here are done using a beam energy of 3.5 TeV, which is the current energy of the LHC. The total center of mass energy is thus 7 TeV. This is half the design energy and the final results with full energy beams can show different characteristics, since many variables vary with energy.

Most parts of this analysis is done using Monte Carlo data generated for the forward tracking telescopes. The Roman Pot system is used only when it provides significant improvement to the result of classification. The emphasis is on what is needed to be measured to determine the total $p-p$ cross section and diffractive mass spectrum. The event classification is in a key role in determining the inelastic rate, choosing events for diffractive mass measurements and for the measurement of individual process cross sections. The methods of event classification and the variables used are therefore discussed in detail. The classification is done separately for events with reconstructed charged particles in both hemispheres and events with reconstructed charged particles in only one hemisphere. The latter events are mainly diffractive and are used for measurement of the diffractive mass distribution. The largest uncertainty of the inelastic rate is the fraction of events with all particles in the pseudorapidity region $|\eta| > 6.5$ (outside the acceptance of T1 and T2). These are low diffractive mass M_X events and the number of these events is extrapolated using the diffractive mass distribution of the events detected with good efficiency.

The measured diffractive mass distribution is not equal to the true distribution because of limited acceptance and efficiency, but also largely due to limited resolution (the mass resolution of the TOTEM forward telescopes is of the order of $\Delta M_X/M_X \sim 1$). One aims at finding the true distribution given measured distribution via a process called unfolding. Diffractive masses $M_X < 4$ GeV are not observed at all and diffractive masses higher than 18 GeV are observed at lower masses (in the region 4 GeV $< M_X <$ 18 GeV) with the current detector systems. This is corrected for by unfolding the reconstructed diffractive mass spectrum using data from all generated events.

2 The TOTEM Experiment

The TOTEM (TOTal Elastic and diffractive cross-section Measurement) experiment is one of the experiments at the LHC (Large Hadron Collider) at CERN (European Organisation for Nuclear Research). TOTEM is located in the same interaction point IP5 as CMS (Compact Muon Solenoid). The TOTEM experiment studies the structure of the proton and has as its main focus to measure the total cross section for the two proton (a proton is abbreviated as p) $p-p$ interaction at LHC energies. This is done using the luminosity independent method applying the optical theorem. TOTEM differs from most other LHC experiments in having detectors only in the forward region. This allows for TOTEM to focus on forward physics.[1]

The most common types of processes in $p-p$ collisions are elastic scattering, single diffraction, double diffraction and inelastic non-diffractive processes. A graphical representation of the processes and estimates for cross sections at Tevatron and LHC are shown in figure 2.1. [1]

Elastic scattering is the process where both initial state protons remain intact (apart from changes in momenta). This is one of the most common processes and the most fundamental process, where the momentum transfer from one proton to the other is usually small. The Mandelstam variable t is defined as the square of the *four momentum* transfer and will be used extensively in this thesis.

$$t = (p_1 - p_3)^2 = (p_2 - p_4)^2, \tag{2.1}$$

where p_1 and p_2 are the four-momenta of the incoming protons and p_3 and p_4 their four-momenta in the final state (in the case of elastic scattering). This formula is valid when the *Minkowski Metric* $\eta^{\mu\nu}$ is defined as

$$\eta^{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}. \tag{2.2}$$

[2]

Single diffractive (SD) events are events with one of the incoming protons remaining intact, whereas the other proton dissociates into a system X of particles with a mass M_X. This mass is known as the diffractive mass of the system. Single diffractive dissociation is assumed to be the most common diffractive process.

In double diffractive (DD) events both of the incoming protons dissociate into systems of particles. Double diffractive events have two diffractive systems with two diffractive masses, often one in both hemispheres (the hemispheres are defined as the two sides of the experiment, in the direction of the LHC pipe and with respect to the interaction point). Double diffractive events have a clear rapidity gap in the central region.

Non-diffractive minimum bias (MB) events are more common than diffractive events. These events often have higher multiplicities of charged particles.

2.1 Measurement of the Total p-p Cross Section

TOTEM aims at a precise measurement of the total $p-p$ cross section σ_{tot}. This cross section has been measured previously for lower energies than the LHC energy. Extrapolations to a higher energy are, however, very unreliable. Earlier measurements at a collider are at the highest at a center of mass energy of 2 TeV at Tevatron (and this is for collisions between protons and anti-protons, $p-\bar{p}$, which can differ), while the LHC can use energies up to 7 times higher or more with future upgrades. The two Tevatron measurements at the *CDF* and *E170* experiments also differ by several standard deviations.[1]

TOTEM uses the Optical Theorem to measure the total cross section σ_{tot} and luminosity $\mathcal{L}$ at the same time. The Optical Theorem is derived from an important property of the S-matrix. This is the fact that the sum of the probabilities for all the final states must be exactly one. This can be expressed for final states f and initial states i as follows [3].

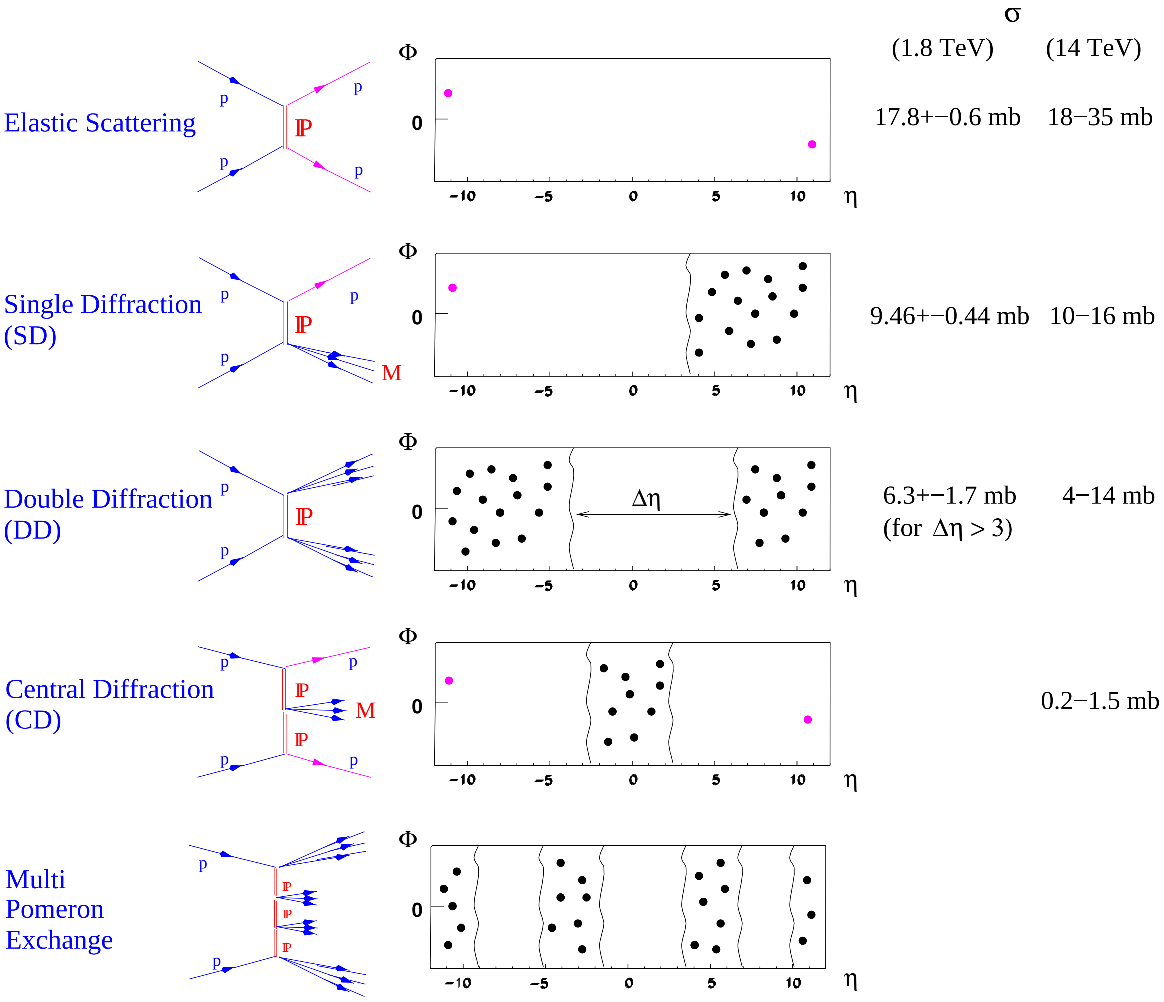

Fig. 2.1: Graphical representation of the most common event types in $p-p$ collisions. The leftmost pictures are graphical representations of the processes and in the middle typical angular and pseudorapidity distributions are shown. Rightmost there are measurements (Tevatron at 1.8 TeV) and estimates (LHC at 14 TeV) for the cross sections of each of the individual processes.

$$\sum_f |\langle f|S|i\rangle|^2 = 1 \tag{2.3}$$

Using orthonormality this can be generalized to give the following set of equations.

$$\sum_f \langle f|S|i\rangle \left(\langle f|S|i'\rangle\right)^* = \delta_{i,i'} \tag{2.4}$$

$$\sum_i \langle f|S|i\rangle \left(\langle f'|S|i\rangle\right)^* = \delta_{f,f'} \tag{2.5}$$

The previous formulae can be written more compactly using matrix notation.

$$S^*S = SS^* = 1 \tag{2.6}$$

This simply states that the S-matrix is a unitary matrix. It is often more convenient to express this using the transition matrix T.

$$\langle f|S|i\rangle = \delta_{f,i} + i(2\pi)^4\delta(Q^{(i)} - Q^{(f)})N_iN_f \langle f|T|i\rangle \,. \tag{2.7}$$

Here $Q^{(i)}$ and $Q^{(f)}$ denote the total energy momentum vectors of the initial and final states, respectively. Now equation 2.6 gives the following.

$$\delta_{f,i} = \delta_{f,i} + i(2\pi)^4\delta(Q^{(i)} - Q^{(f)})N_iN_f(\langle f|T|i\rangle - \langle i|T^*|f\rangle)+$$

$$+ (2\pi)^8 N_iN_f \sum_a N_a^2 \delta(Q^{(i)} - Q^{(a)})\delta(Q^{(a)} - Q^{(f)}) \langle f|T|a\rangle \langle a|T^*|i\rangle \tag{2.8}$$

$$\Im |\langle f|T|i\rangle| = \frac{1}{2}(2\pi)^4 \sum N_a^2 \delta(Q^{(i)} - Q^{(f)}) \langle f|T|a\rangle \langle a|T^*|i\rangle \,, \tag{2.9}$$

where $\Im$ stands for the imaginary part ($\Re$ is used for the real part).

An important case is when the final and initial states are the same. Using the normalization $N = \prod_{initial} \frac{1}{\sqrt{2VE_i}} \prod_{final} \frac{1}{\sqrt{2VE_f}}$ (where V is an arbitrary volume and E the energy of the state) and taking the limit when the normalization volume is very large leads the following expression:

$$\Im\left[\langle i|T|i\rangle\right] = \Im(f_{el}(0)) = \sqrt{\lambda(s, m_i^2, M_i^2)}\sigma_{tot}, \tag{2.10}$$

where $\lambda(x, y, z) = x^2 + y^2 + z^2 - 2xy - 2xz - 2yz$. This is known as the Optical Theorem.

The following relation applies for the total cross section:

$$\frac{d\sigma}{dt} = \frac{|T|^2}{16\pi\lambda}. \tag{2.11}$$

Combining the two previous formulae (2.10 and 2.11) on the limit of zero momentum transfer gives a relation between cross section, luminosity and the nuclear part of the elastic cross section dN_{el}/dt extrapolated to zero momentum transfer.

$$|T|^2 = 16\pi\lambda \left.\frac{d\sigma}{dt}\right|_{t=0}$$

$$|T|^2 = (\Im[T])^2 + (\mathcal{R}[T])^2$$

$$\sigma_{tot}^2 = \frac{(\Im[T])^2}{\lambda} = \frac{|T|^2 - (\mathcal{R}[T])^2}{\lambda} = \frac{|T|^2}{\lambda} - \frac{(\mathcal{R}[T])^2}{(\Im[T])^2}\sigma_{tot}^2$$

Defining $\rho = \mathcal{R}[T]/\Im[T]$, this can be written in the form shown below.

$$\sigma_{tot}^2 = 16\pi \left.\frac{d\sigma}{dt}\right|_{t=0} - \rho^2\sigma_{tot}^2$$

$$\sigma_{tot}^2 = \frac{16\pi}{1+\rho^2} \left.\frac{d\sigma}{dt}\right|_{t=0} \tag{2.12}$$

Using the relation $\mathcal{L}d\sigma/dt = dN_{el}/dt$ between cross section σ, luminosity $\mathcal{L}$ and number of elastic events N_{el} , the following useful formula is derived.

$$\mathcal{L}\sigma_{tot}^2 = \frac{16\pi}{1+\rho^2} \cdot \left.\frac{dN_{el}}{dt}\right|_{t=0} \tag{2.13}$$

The parameter ρ is taken from theoretical predictions and is of the order of $\rho \sim 0.14$ [4] and thus has a small impact due to the term $1+\rho^2$ being the only term with dependence on ρ.[1]

An additional relation for the cross section and the luminosity is given by the elastic and inelastic rates N_{el} and N_{inel} as follows.[1]

$$\mathcal{L}\sigma_{tot} = N_{el} + N_{inel} \tag{2.14}$$

Cross section and luminosity can be obtained by solving the system of equations given by 2.13 and 2.14. Solving the set of equations for σ_{tot} and $\mathcal{L}$ independently gives the following.

$$\sigma_{tot} = \frac{16\pi}{1+\rho^2} \cdot \frac{dN_{el}/dt|_{t=0}}{N_{el}+N_{inel}} \tag{2.15}$$

$$\mathcal{L} = \frac{1+\rho^2}{16\pi} \cdot \frac{(N_{el} + N_{inel})^2}{dN_{el}/dt|_{t=0}} \tag{2.16}$$

The cross section can thus be calculated by measuring the inelastic rate N_{inel} (measured by T1 and T2 tracking telescopes), the total nuclear elastic rate N_{el} (measured by Roman Pot detectors) and $dN_{el}/dt|_{t=0}$. Rate measurements are possible only because all TOTEM detectors systems are capable of triggering. The precision of the cross section and luminosity measurements is expected to be 5% with $\beta^* = 90\,\mathrm{m}$ beam optics as an intermediate step and as low as 1% with optimization and special beam optics with $\beta^* = 1540\,\mathrm{m}$. [1]

2.2 Diffraction

Diffractive scattering, comprising of Single Diffraction, Double Diffraction, Double Pomeron Exchange (also known as Central Diffraction) and higher order processes, together with elastic scattering, comprises a large fraction of the total $p-p$ cross section. The understanding of these processes, which are important in understanding the structure of the proton, is not yet good. The majority of events have intact final state protons (this always applies to both single diffraction and central diffraction), of which some can be detected with the Roman Pot detectors.[1]

At an early stage TOTEM can measure only large fractional momentum losses $\xi = \Delta p/p$, the four-momentum transfer squared t and the mass distributions for Single Diffractive and soft Double Pomeron exchange events. Diffractive events with one or more rapidity gaps cannot be measured precisely without the combination of TOTEM and CMS detectors, which is planned in the future. At the early stage the main focus is on single diffractive masses between 0.85 GeV and 2.5 GeV (for a center of mass energy of 7 TeV). [5, 1]

2.3 Detectors

The TOTEM experiment consists of three types of detectors: T1 and T2 tracking telescopes as well as Roman Pot proton detectors. The detector system is symmetric with respect to the interaction point. Thanks to the completely symmetric approach both hemispheres can be expected to give compatible results. Therefore it makes no difference for instance in a single diffractive event which hemisphere the diffractive system and which hemisphere the intact final state protons reside in.[1]

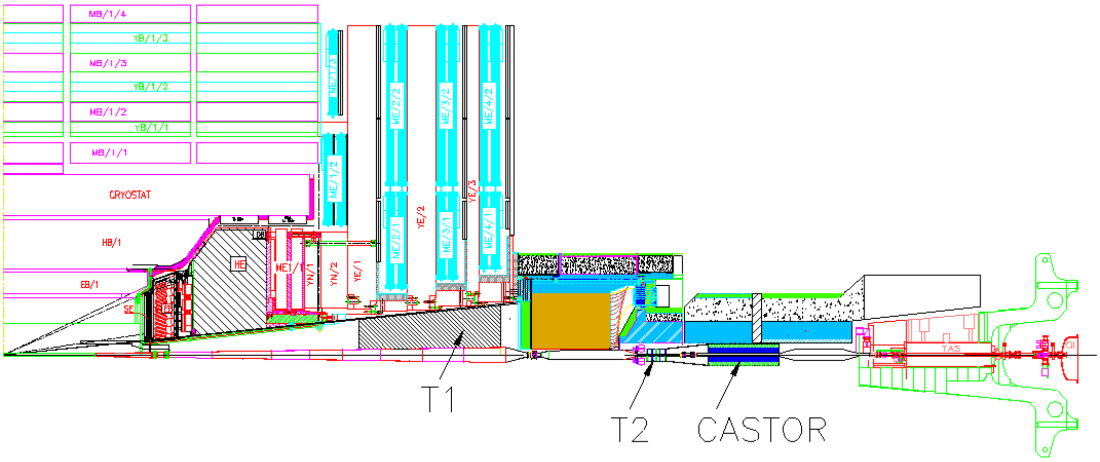

Fig. 2.2: The positioning of the T1 and T2 forward tracking telescopes in the CMS forward region. The layout on the other side of the interaction point is similar.[6]

T1 Detectors

The T1 detectors are based on cathode strip chambers (CSC), which are multi-wire proportional chambers with segmented cathode readout. The T1 detectors have a geometric acceptance of $3.0 < |\eta| < 4.6$ for charged particles and are installed in the endcaps of CMS. The inelastic trigger plays an important role in measuring the total cross section.[1]

The telescope consists of two arms on each side of the interaction point at a distance of between 7.5 m and 10.5 m. Each arm of the telescope is built up of two half arms, consisting of 5 CSC planes equally spaced in Z (the coordinate parallel to the beam). Each of the CSC planes consists of 3 CSCs, each covering a region of approximately 60° in the azimuthal plane. There is some overlap in order to achieve continuous efficiency in region between two adjacent detectors.[1]

T2 Detectors

The T2 detectors are based on gas electron multipliers (GEM). The T2 detectors have a geometric acceptance of $5.3 < |\eta| < 6.5$ for charged particles. The T2 detector system consists of two detector telescopes (one on each side of the interaction point at a distance of 13.5 m). Each of the telescopes consists of two half telescopes, which in turn consist of 10 GEM-planes each. The GEM planes are almost semi-circular and alternate on each side of the beam giving complete angular coverage around the beam in the θ-plane. The design of the telescopes allows good coverage of forward physics processes at low and moderate luminosities. Key benefits of the gas electron multiplier detectors are high rate capability, good spatial resolution, robust mechanical structure and excellent ageing properties. GEM technology was previously used at the COMPASS experiment at the SPS (*Super Proton Synchrotron*) at CERN.[1, 7, 8]

The GEM-detector is of semi-circular shape having an area covering an azimuthal angle of 192° and a radius from 43 mm up to 144 mm. Each GEM plane consists of three layers of GEM foils, used to amplify the signal by multiplying the electrons formed in the initial ionisation of the gas. The multiplication is done in stages to reduce probability of discharge below 10^{-12}. Each stage gives on average a multiplication of 20 and thus the final multiplication after all three stages is roughly 8000. The voltage divider is also designed such that the voltage gradually decreases going down from the uppermost foil to the lowermost. The GEM foils consist of 50 μm polyimide foils with 5 μm copper cladding on each side. The holes are wet etched into the foils using a bidirectional approach, producing conical holes with diameters of 65 μm in the middle and 80 μm at the surface. The support structures for the GEM foils are manufactured from fiberglass reinforced epoxy plates. [1, 9]

Readout is provided by a two dimensional printed circuit readout board with a pattern of pads and strips. The radial coordinate is determined using 2×256 concentric strips, while the azimuthal coordinate and trigger is provided by 1560 pads. The strips are isolated from the pads by a layer of polyimide. The strips are split into two sections of 256 strips, which are read out at both ends of the chamber. Pads are radially divided into 65 sectors, which are further divided into 24 pads, differing in size depending on their radial position. [1]

A mixture of 70/30 Ar/CO_2 is used as fill gas in the GEM chambers. Gas is supplied through channels engraved in the fiberglass frame beneath the polyimide foil.[1]

Roman Pot Detectors

The Roman Pot detectors are silicon strip detectors in special beam insertion devices called Roman Pots. They have their own vacuum in order to prevent outgassing of detector materials to interfere with the LHC vacuum. They are specialised detectors suited for detecting the protons that scatter in small angles. The Roman Pot detectors are located far away (147 and 220 m) from the interaction point and symmetrically on both sides of the interaction point in order to be able to detect protons scattered in small angles with respect to the beam. The Roman Pot stations at 147 m and 220 m consist of two units with two vertical pots and one horizontal pot each. The horizontal pots are placed on the outer side of the LHC ring and completes acceptance for diffractively scattered protons, whose momentum loss deviates them towards the

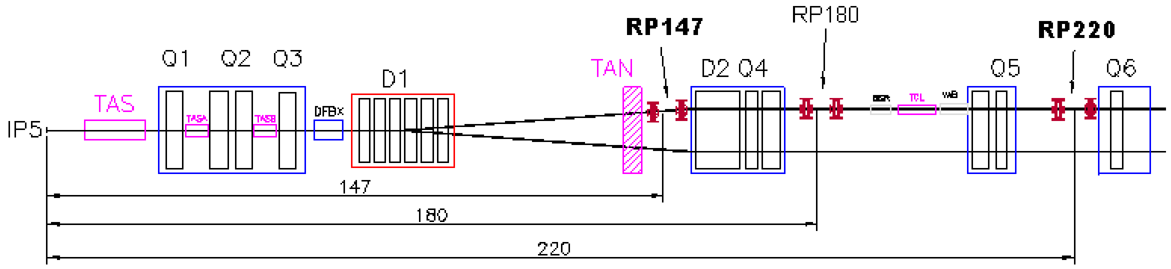

Fig. 2.3: The layout of the Roman Pot system.[6]

pot. The horizontal pots also provide overlap with the other pots and can be used for relative alignment of the pots. No protons produced by interaction are expected on the inner side of the LHC ring and thus there is no need for a detector there.[1]

2.4 Detector Readout and Reconstruction Chain

Readout of all TOTEM detectors is done using second generation VFAT chips (*Very Fast ATLAS and TOTEM* chip) that provide a digital readout. This requires that the parameters, like the threshold used for determining if a sensitive element of a detector is on or off, are optimized before measurement. The main advantage of digital readout is smaller amounts of data that need to be transferred and the possibility to detect errors in transmission using control bits. [1]

For T2 detectors this data consists of matrices of pads and strips. Each matrix element is either true or false and corresponds to a pad or strip in the detector being on or off, respectively. Due to diffusion of electrons and delta rays there is a possibility that a particle produces an active region of two or more adjacent pads or strips. Such large active regions need to be recognized in order to reduce noise. This is done with clusterization algorithms, that detect and store the data of such clusters. Also T1 and RP detectors have similar reconstruction done.

Overlapping pad and strip clusters are detected and stored as what is called hits. The ideal would be that each of these reconstructed hits corresponds to a particle hitting the region of the detector, but due to noise and inefficiency this is not always true. These reconstructed hits are used for track reconstruction. The tracks are fitted to the hits and track parameters like angles are calculated.

2.5 T2 Pad Clusterization

Whenever the signal level in one of the pads in the T2 detectors (there are 1560 pads in each plane) is higher than a given threshold (this threshold is adjusted prior to data taking), the pad is on. This happens mostly due to particle hits, but sometimes this can also happen due to electronics noise. This noise, however, often affects large areas, whereas true hits only put on a few adjacent pads.

Because the interest lies in finding the multiplicities and tracks of real particles, such clusters of many pads caused by one particle need to be identified. The identification is done at the TOTEM experiment using a clusterization algorithm, implemented in a clusterization module. The clusters produced by the clusterization module are considered the individual particle hits, instead of the active pads. The clusterization is done for each pad plane separately.

So far, the clusterization has been done using a fixed window algorithm (the module implementing the fixed window algorithm is called *T2MakeCluster*), which looks at a window of three rows and five columns. This works well when the clusters are small. This would be the ideal case in absence of electronic noise and detector malfunction. In reality the case is often not so simple. The output of a clusterizer of this type (with this amount of rows and columns in the window) is always a cluster of size of 1-15 pads (in the current code a pad on in the lower left corner of such a window is not included in the cluster, thus limiting the size to a maximum of 14).

A new clusterizer (*T2MakeCluster2*) was implemented in order to better satisfy the needs of events with high electronics noise. This clusterizer uses a nearest neighbour approach. Each pad is examined one by one starting from one corner and then going through each of the 65 rows and 24 columns. The construction of a new cluster is started when an active pad is found. This pad is added to the cluster and marked as inactive to prevent it from being added to another cluster. All the neighbours (to the left, to the right, above, below and diagonal) are checked for activity. If active, they are added to the cluster and to a list of pads, whose neighbours need checking. This process is repeated for all the entries in the list of clusters to be checked until the list is empty. This way the constructed cluster comprises of all the adjacent active pads.

There are two reasons why it is important to have the big clusters constructed properly.

Firstly, if a big cluster is divided into many smaller clusters, all of these are interpreted as passing particles in the subsequent track reconstruction. This can act as a source of spurious tracks, that do not reflect the true event. This also distorts the position information of the cluster, which can also affect the reconstruction of the parameters of the resulting tracks.

Secondly, the proper reconstruction of big clusters allows one to identify electronics noise. Based on results from Monte Carlo simulations, a true particle is not expected to produce very large clusters. Looking at the clusters one can discriminate between noisy and clean events. Noise can be reduced by several means based on cluster size.

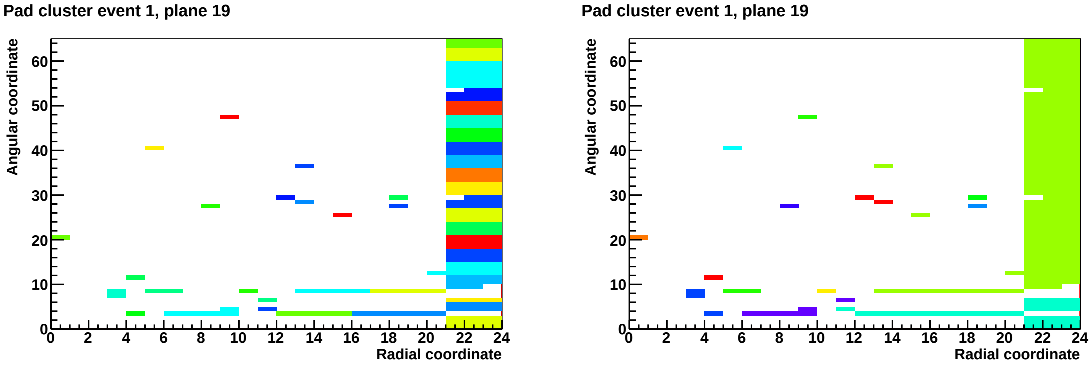

Fig. 2.4: Comparison between clusterization of a pad plane with the fixed window (left) and the nearest neighbour (right) approaches. Each of the figures is a representation of pads on in the same single plane in the same single event. Colours are arbitrary; each block of one colour represents one cluster.

Large clusters can be discarded from reconstruction. It is also possible to discard the whole detector or plane, or even the whole event, when such clusters are found.

Further improvement of the clusterization can be done by taking into account dead (dead refers to a pad that is always or a considerable fraction of the time off) or noisy (noisy refers to a pad that is a considerable fraction of the time on) pads. Dead pads could for instance split a cluster into two, whereas noisy pads could form spurious clusters or enlargen otherwise good clusters. This can relatively easily be taken into account with the nearest neighbour approach. A pad considered dead could be considered as being on, if several of its neighbours are on. A pad considered noisy could be discarded if all nearby pads are off.

Histograms of the pad and strip cluster size distributions are made separately for pad and strip clusters for each plane. This helps in distinguishing planes with unusually large clusters (noisy planes) or planes with a large fraction of small clusters (clean planes). For pads, the distributions are studied using both clusterization algorithms. The impact of the chosen algorithm is clearly visible in the resulting distributions (see figure 2.5). As expected, the mean number of pads in one cluster is larger using the nearest neighbour algorithm, whereas the number of pad clusters is smaller. The integral of the two is the same (thus this complies with the number of pads on and the number of events being the same).

2.6 T2 Timing and Threshold Study

Due to the digital readout of the T2 detector, optimal data acquisition can only be achieved after adjusting parameters affecting readout. Each detector plane and readout chip is different and parameters that are optimal for one detector plane might

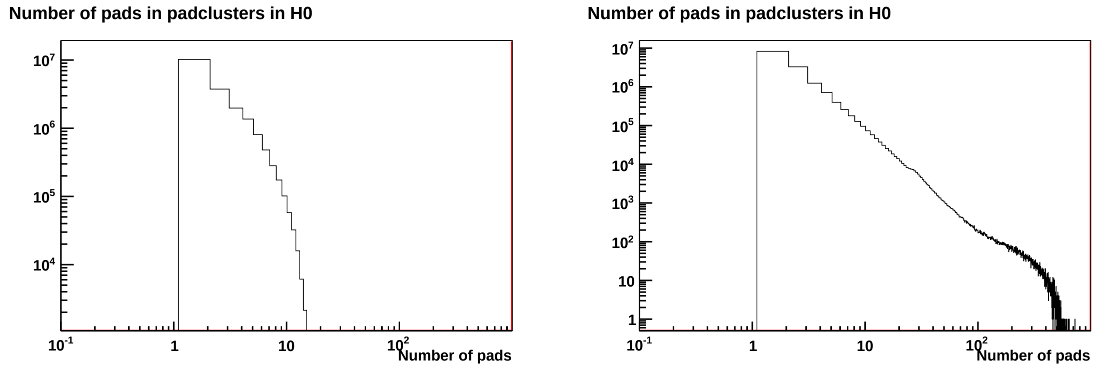

Fig. 2.5: Comparison between pad cluster sizes for the same run and plane using the fixed window (left) and nearest neighbour (right) clusterization algorithms. Note that the scales are logarithmic.

be inoptimal for another. Some of the most important parameters are the threshold voltage (the measured voltage above which the *analog to digital converter* (ADC) considers an area of the detector “ON”), the latency (the amount of clock ticks after triggering when the electrons produced in the GEMs have drifted and been digitized) and the voltage difference applied over the detector plane.

Extensive testing has been done to determine the optimal latencies and threshold voltages. This is done by measuring collisions with different combinations of parameters; this is best achieved by keeping the threshold constant and varying the latency for a set of measurements and vice versa. The effect of varying parameters is studied by comparing parameters describing the occupancy (the fraction of pads or strips “ON” with respect to their total numbers) of planes or half-telescopes of the T2 system. The occupancy of a plane or as an average occupancy for all the planes in a half-telescope can be shown as a distribution, but a large number of distributions is difficult to compare. Therefore the comparison is done between mean occupancies and the fraction of events with an entire plane “OFF” (this is done individually for pads and strips). Also the cluster sizes can be studied in order to see if the level of noise changes for instance when lowering the threshold voltage.

Effects of varying latency and threshold on pad occupancies and cluster numbers and sizes are presented here (effects on the corresponding variables for strips are similar). The results from a latency scan shown in figures 2.6 and 2.7 show that there is an optimal latency, which is not same for all half-telescopes. The maximum occupancy is found with a latency of 142 clocks in one half-telescope, but 140 for another (also the fraction of events with planes off is smallest with these latency values). The number of pad clusters and the size of pad clusters are highest at these latencies.

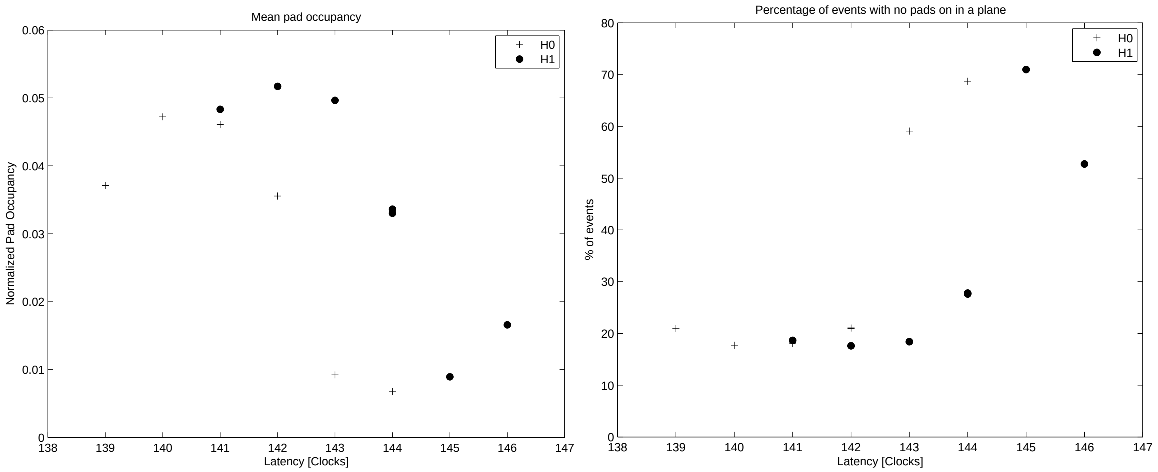

Fig. 2.6: Results of a latency scan with constant threshold for mean occupancies in two half-telescopes (labelled H0 and H1). The mean pad occupancy is on the left and the percentage of events with all pads "OFF" on the right.

The results from a threshold scan shown in figures 2.8 and 2.9 show, as expected, that the lower the threshold is, the higher the occupancy. The larger cluster size with a low threshold is a sign of possible additional noise. Due to this additional noise, the lowest threshold is not the best.

2.7 Software

TOTEM uses the same software framework CMSSW (CMS SoftWare)[10] as CMS for most of the simulation and reconstruction. Using the same software framework as CMS makes sense because it has all the basic functionality implemented and software compatibility makes a possible joint physics program with CMS easier to accomplish. CMSSW has a high degree of integration with the analysis framework ROOT[11], which provides functionality for filling and plotting histograms, fitting of functions to distributions, event classification. Unfolding can be added to the ROOT framework using an additional library called RooUnfold.

Additional tools used in simulation are the Monte Carlo event generators PYTHIA[12, 13] and PHOJET[14] as well as the GEANT4[15, 16] program, which simulates the passage of particles and their interaction with matter. Geant4 information is also used to get information about simulated detector response using a module implemented within CMSSW.

Each step of simulation, reconstruction and analysis is done using small modular pieces of software implemented within the CMSSW framework. Each step of simulation and reconstruction can be saved into files in the format used by ROOT. The ROOT format is a memory dump of the C++ objects, which is compressed and

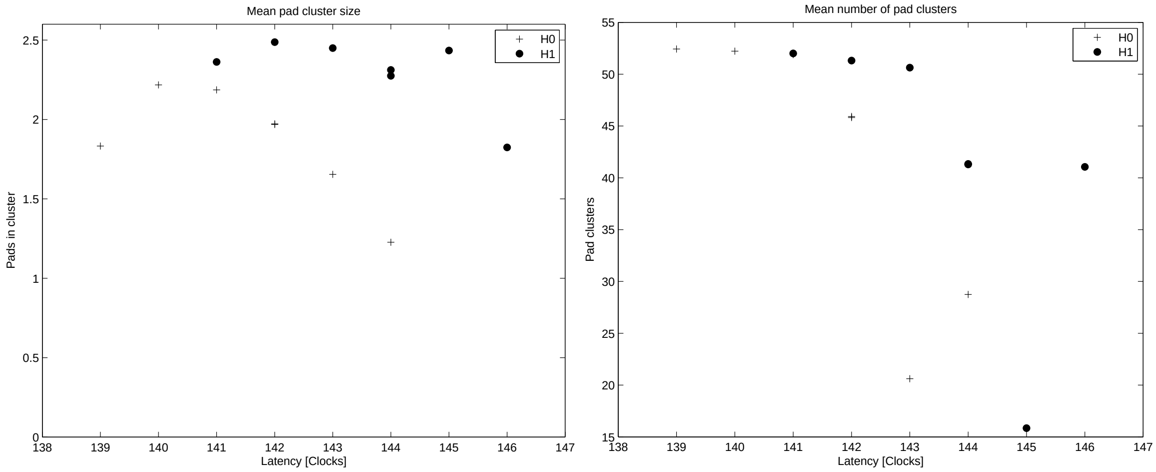

Fig. 2.7: The effect of latency on the mean pad cluster size (left) the mean number of pad clusters (right) in two half-telescopes.

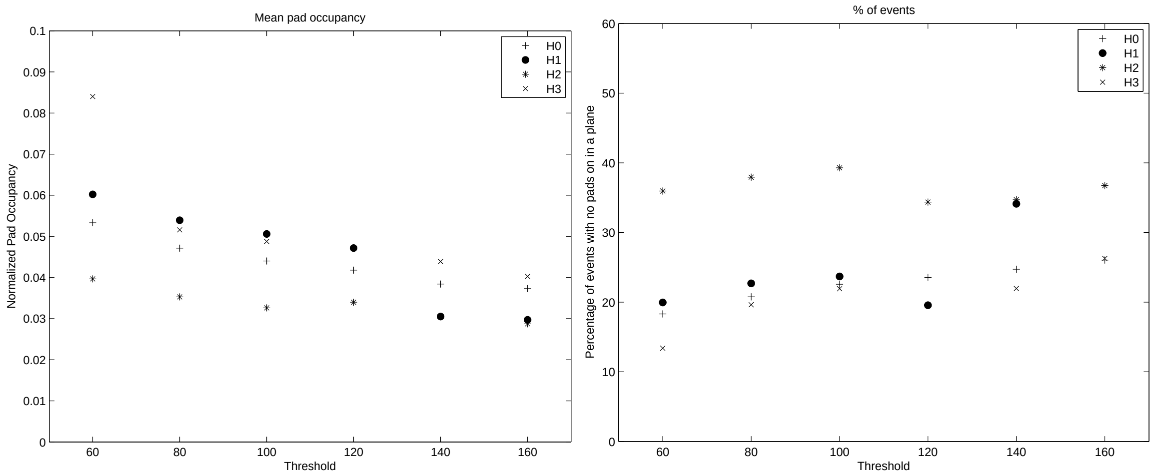

Fig. 2.8: Results of a threshold scan with constant latency for mean occupancies in all four half-telescopes (labelled H0, H1, H2 and H3). The mean pad occupancy is on the left and the percentage of events with all pads "OFF" on the right.

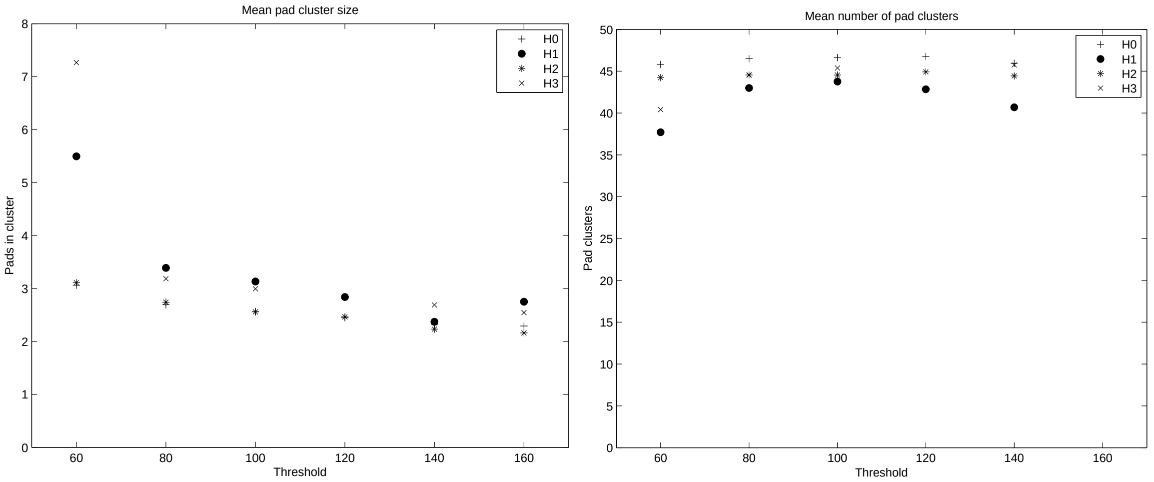

Fig. 2.9: The effect of threshold on the mean pad cluster size (left) the mean number of pad clusters (right) in all four half-telescopes.

information about the names of the variables and their types is retained.

Data acquired from detectors is in an intermediary VMEA format as given by the VFAT-chips, but is converted to the ROOT format used by CMSSW before reconstruction. This allows for more easy access to data and allows an event by event and step by step access to different stages of reconstruction. This also allows later validation and analysis of raw data to be done.

Classification of events is done using two separate modules for CMSSW. One reads the data and allows for a classifier to be trained while the other loops through the events classifying them one by one. These modules are called *T1T2Reader* and *T1T2Analyzer*, respectively. This implementation is because of the way the classification toolbox in ROOT works. It requires the variables to be given in a ROOT file as simple TTree objects. TTree is a data format, which allows simple storage of data given to ROOT as standard C programming language structures. Events in CMSSW are stored using the Event Data Model (EDM), which stores all the data for one event in one block and the events in a sequential manner.[17, 18]

3 Event Generators

Simulation of collisions is done in order to be able to develop and test the reconstruction software without the immediate need for real data. Monte Carlo methods are used to simulate collisions and detector response. The first step collision process is simulated by the event generator (for instance PYTHIA, PHOJET or ELEGENT). This event generator uses a mathematical model to simulate the collision process and the types and directions of final state particles. Monte Carlo simulations are often tuned to be compatible with results from earlier experiments and thus they give reasonable results at least in some region of collision energy. The following step comprises the transfer of particles in vacuum or a medium, often in presence of an electric or magnetic field. This step is often done with separate simulation software (GEANT4), which uses Monte Carlo methods and models for electromagnetic interactions etc., but also simulates simple particle decays. Simulations are usually used to determine how accurately measurements can be made given a certain set of detectors. Simulated data can also be used to correct for the regions where the detector is inefficient. PYTHIA is used as the primary generator for the studies presented in the later sections. PHOJET is used for comparison to estimate the systematic errors for parameters extracted using the PYTHIA sample.

3.1 Simulation Chain

Simulation of events starts of with the simulation of the collision between protons. This is done with the Monte Carlo event generators PYTHIA or PHOJET. The event generator uses a mathematical model to estimate final state particle types (masses and charges) and momenta (including absolute value and direction).

Information about particles produced in the collision are given to the GEANT 4 simulation software, which in turn simulates their tracks in magnetic and electric fields and their interactions with materia as well as their lifetimes and decay modes. Interactions with detector materia lead to an observable signal, which is first simulated as an analog response and then digitized. This digitized information corresponds to the information received from the real detectors.[19]

3.2 Monte Carlo Event Generators

Three different Monte Carlo event generators are used to simulate the proton-proton collisions at the interaction point. The generators used are PHOJET, PYTHIA 6.4 (old version written in Fortran) and PYTHIA 8.1 (new version written in C++). The generators use different mathematical models to simulate the interactions and therefore show different behaviour. At this stage mainly the older version of PYTHIA and PHOJET are used, but the new version of PYTHIA can possibly come into use in the near future at TOTEM.[19]

PYTHIA uses perturbative QCD for both low and high regions in transverse momentum. Regions in low transverse momentum play a significant role in inelastic scattering. Perturbative QCD is divergent at low transverse momentum. This is corrected by one of two methods, either a simple cut-off or a more complex correction factor.[14]

PHOJET uses the Dual Parton Model and another mechanism than perturbative QCD in the region of low transverse momentum. PHOJET is considered to use a two component model with a smooth transition between low and high transverse momentum. PHOJET has been tuned with data from CDF at 1800 GeV.[14]

The total cross section σ_{tot} for proton-proton collisions is calculated in PYTHIA using Regge theory using the following sum of powers.[14]

$$\sigma_{tot}[\text{mb}] = 21.70 s[\text{GeV}]^{0.0808} + 56.08 s[\text{GeV}]^{-0.4525}$$

These terms correspond to Pomeron exchange and Reggeon exchanges, respectively. The second term becomes negligible at high energy. In PHOJET, on the other hand, the cross section is calculated using the dual component Dual Parton Model and the optical theorem. In the region of $700-800\,\text{GeV}$ both generators agree well with earlier experimental data. With higher energies the predictions are different; at full LHC energy of 14 TeV the predictions are $\sigma_{tot} = 101.5$ mb for PYTHIA and $\sigma_{tot} = 119$ mb for PHOJET. A list of cross sections for the most common processes simulated with PYTHIA and PHOJET is given in table 1. An overview of total cross sections for the Monte Carlo generators PYTHIA and PHOJET and experimental data is in figure 3.1. From this figure the total cross section for the current LHC energy $\sqrt{s} = 7$ TeV is found to be $\sigma_{tot} \sim 90$ mb for PYTHIA and $\sigma_{tot} \sim 105$ mb for PHOJET.[14, 20, 21]

The total cross section σ_{tot} is the sum of the elastic and inelastic cross sections:

$$\sigma_{tot} = \sigma_{elastic} + \sigma_{inelastic}, \tag{3.1}$$

where the inelastic cross section $\sigma_{inelastic}$ can be further divided between different process types like single diffractive (SD), double diffractive (DD) and non-diffractive (ND) [14].

$$\sigma_{inelastic} = \sigma_{SD} + \sigma_{DD} + \sigma_{ND} + ..., \tag{3.2}$$

where σ_{SD}, σ_{DD} and σ_{ND} represent single diffractive, double diffractive and non-diffractive cross sections. The additional terms come from less common processes like central diffraction. The fractions of the total cross section that the generator designates to each of these depends on the generator.[14]

Process	PYTHIA σ [mb]	PHOJET σ [mb]	Difference [%]	Theoretical predictions [mb]	
Elastic	22.2	54.9	34.4	20.1[22]	20.9[23]
Inelastic	79.3	84.5	6.6	67.9[22]	71.2[23]
Minimum bias	55.2	68.0	23.1	-	-
Single diffraction	14.3	11.0	22.0	13.3[22]	11.8[23]
Double diffraction	9.8	4.06	58.5	13.4[22]	6.08[23]
Central diffraction	-	1.42	-	-	-
Total cross-section	101.5	119	17.2	88.0[22]	92.1[23]

Tab. 1: Cross sections for the most common processes at full LHC energy of $\sqrt{s} = 14\,\text{TeV}$.[14]

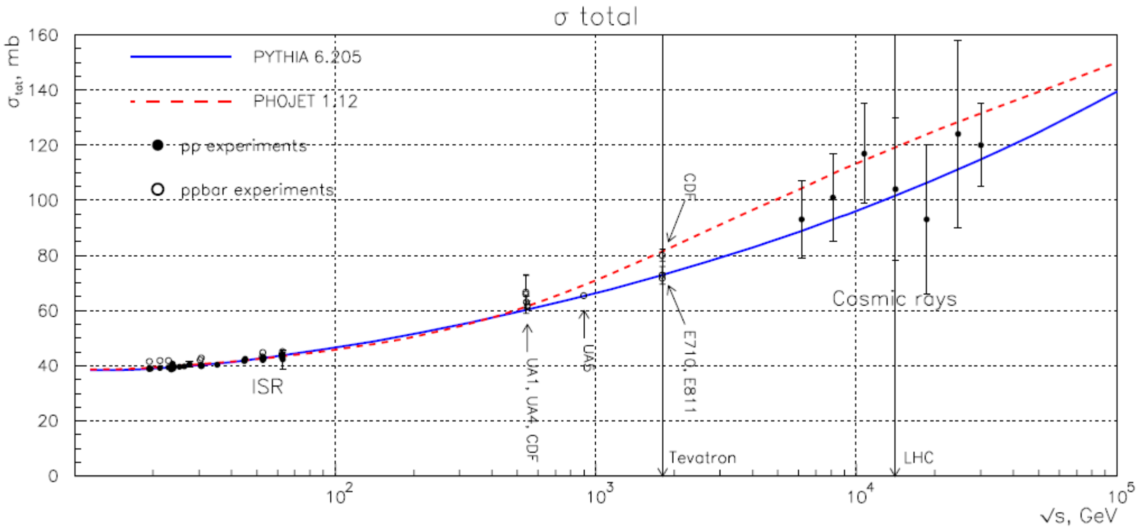

Fig. 3.1: Cross sections for different energies using PYTHIA (solid line), PHOJET (dotted line) and from experimental data from pp-colliders (black circles) and $p\bar{p}$-colliders (white circles).[14]

4 Event Classification

Event classification aims at recognizing the type of interaction that has occured for each event by using only detector responses. Event classification is needed to distinguish between diffractive and non-diffractive events. Diffractive masses only make sense in diffractive events and thus non-diffractive background has to be eliminated.

4.1 Definitions

Detector response is characterized by several variables. Definitions of multiplicities and rapidity gaps are introduced in order to understand the variables that are used in the following event classification.

4.1.1 Multiplicities

The multiplicity is defined as the number of reconstructed particles in an event. The detectors do not detect particles in all angles and therefore the multiplicities covered here are experimental multiplicities. Multiplicities are studied in both hemispheres as well as the T1 and T2 detectors separately. It should be noted that the T1 and T2 detectors detect only charged particles and thus the reconstructed multiplicities are multiplicities for charged particles (or *charged multiplicities*).

Multiplicity is one of the main signatures of the process. Elastic collisions have low multiplicities, because the protons remain intact and thus one can expect to have only one proton in each hemisphere. Single diffractive events can have a multiplicity of one in one hemisphere, but a higher multiplicity in the other, whereas double diffractive events have a higher multiplicity in both hemispheres. Non-diffractive minimum bias events have on average even higher multiplicity than diffractive events.

4.1.2 Rapidity Gaps

Two kinds of rapidity gaps (low η and high η) are defined per hemisphere in order to aid with the classification work. With the acceptances of the T1- and T2-detectors being $3.0 \leq |\eta| \leq 4.6$ and $5.3 \leq |\eta| \leq 6.5$, the low eta rapidity gap is defined as

$$\Delta\eta_{low} = |\eta_{min}| - 3.0 \tag{4.1}$$

and the high eta rapidity gap is defined as

$$\Delta\eta_{high} = 6.5 - |\eta_{max}|, \tag{4.2}$$

where η_{min} and η_{max} are the smallest and largest values of the reconstructed track pseudorapidities in each hemisphere of the event. The rapidity gaps are limited to the interval from 0 to 3.5, so that negative values given by the formula above are replaced by 0 and values higher than 3.5 by 3.5. The defined gaps in pseudorapidity give information about the angular distribution of detected tracks with respect to the beam line.

The rapidity gaps defined here differ from the conventional definition of rapidity gap. Usually the rapidity gap is defined as the size of the gap in the central region of the experiment. Thus the defined low η rapidity gaps and the conventional definition are correlated as follows.

$$\Delta\eta = \Delta\eta_{low+} + \Delta\eta_{low-} + 6.0, \tag{4.3}$$

where $\Delta\eta$ is the standard rapidity gap, $\Delta\eta_{low+}$ the low η rapidity gap in the positive hemisphere and $\Delta\eta_{low-}$ in the negative hemisphere. This is under the assumption that there are no undetected particles in the region $|\eta| < 3.0$. If there are particles in this region, it is impossible to accurately define the conventional rapidity gap with only data from the T1 and T2 detectors.

Rapidity gaps can often be used to understand what kind of interaction each event represents. In many cases this requires very precise measurement of the rapidity gap, which is often not possible due to limited acceptance. Rapidity gaps also provide a means of determining the diffractive mass of a diffractive system. Rapidity gaps are closely related to the kinematics of the process.

4.2 Event Selection

Events for classification and for diffractive mass measurement are selected in order to have enough information to make classification or calculation of diffractive mass feasible. Event selection requires that an event must have a total of at least two reconstructed particles in T1 and T2 combined. This limits the analysis to events with some non-zero multiplicities and non-trivial rapidity gaps.

4.3 Method

The type of event the detected particles originate from is of high interest. For Monte Carlo simulated events this information is available, but for real events statistical methods are needed to deduce the event type. Only data received from the detectors and the beam parameters is available. The behaviour of the particles and detectors is simulated and this gives an insight to how classification can be done. Multivariate analysis techniques can be used to transform a set of known variables into one of many predefined classes. One of the classical methods is the Fisher discriminant analysis method, but newer methods allow for more complex classification and can give better results. There can, however, be a remarkable difference between real and simulated data; this can be for instance due to errors in the Monte Carlo model or errors due to noise in detectors or other devices. Simulation of noise is often done with a simplified approximative model.

Classification of events at TOTEM is done using multivariate analysis methods. Classification is done in one or two steps: the first step is to determine whether the event is diffractive or non-diffractive and for diffractive events the second step is to determine whether the event is single or double diffractive. A better result might be obtained by using a multi class variant of the classifier. This would be most likely to aid in the classification of diffractive events.

Nine variables are used for events with detected particles in both hemispheres. The variables used are:

- Multiplicity of tracks in the T1 detector in the positive hemisphere (fig. 4.2 and 4.9)
- Multiplicity of tracks in the T1 detector in the negative hemisphere (fig. 4.1)
- Multiplicity of tracks in the T2 detector in the positive hemisphere (fig. 4.4 and 4.10)
- Multiplicity of tracks in the T2 detector in the negative hemisphere (fig. 4.3)
- High eta rapidity gap in the positive hemisphere $\Delta\eta_{high+}$ (fig. 4.6 and 4.11)
- High eta rapidity gap in the negative hemisphere $\Delta\eta_{high-}$ (fig. 4.5)
- Low eta rapidity gap in the positive hemisphere $\Delta\eta_{low+}$ (fig. 4.8 and 4.12)
- Low eta rapidity gap in the negative hemisphere $\Delta\eta_{low-}$ (fig. 4.7)
- Number of protons detected in the Roman Pot detectors

Events with tracks in only one hemisphere are flipped in the analysis software so that they always have the reconstructed tracks in the positive hemisphere. This is done in order to make classification easier and to eliminate the need for variables in the negative hemisphere, while eliminating only null data. Flipping is possible thanks to the symmetric detector systems. In case the hemispheres would differ with respect to acceptance this would not be feasible. Also events with data from both hemispheres are flipped. In this case the condition is that if the high eta rapidity gap $\Delta\eta_{high}$ in the positive hemisphere is larger than in the negative hemisphere. In single diffractive events this most often flips the event such that the intact final state proton is in the positive hemisphere. It is important to do both the training of the classifier and the final classification with the same criteria of flipping or classification performance weakens due to incorrect training of the classifier.

Events with tracks in only one hemisphere can be automatically chosen as diffractive events, due to non-diffractive minimum bias events on average having large multiplicities in both hemispheres. Classification can in theory get some of the non-diffractive minimum bias background removed, which makes the diffractive mass distributions more accurate, but cutting the non-diffractive minimum bias events with a classifier cuts a large fraction of diffractive events as well.

Roman Pot data is currently only used after the classification is done to move all inelastic events with final state protons into the single diffractive category. Also using the number of protons as a variable for the classifier itself has been tested, but this was found to make the classification of single diffractive events with no detected final state protons much less accurate.

The classification algorithms depend heavily on the multiplicities (figures 4.1-4.4 for two hemisphere events and 4.9 and 4.10 for one hemisphere events); especially the non-diffractive minimum bias events have on average higher multiplicities than diffractive events, making these the most important variables for distinguishin diffractive events. The rapidity gap distributions (figures 4.5-4.8 for two hemisphere events and 4.11 and 4.12 for one hemisphere events) give some additional information. This information can be used to distinguish some single and double diffractive events and also aids in identifying non-diffractive minimum bias events. Non-diffractive minimum bias events are usually two hemisphere events and therefore the distributions for one hemisphere non-diffractive minimum bias events have poor statistics.

4.4 Fisher Linear Discriminant Analysis

In multivariate analysis the goal is to classify sets of variables with certain values using only the information about their values. The idea behind the Fisher discriminant is to project a vector, with these variables as components, into a 1-dimensional plane in a way that maximizes the separation between the classes. If the value of the projection is above a certain threshold, the event belongs into one class and if below it belongs to the other class [24].

In mathematical terms, we have a set of N measured variables per event. A linear combination X of the measured variables, which maximizes the ratio of the difference

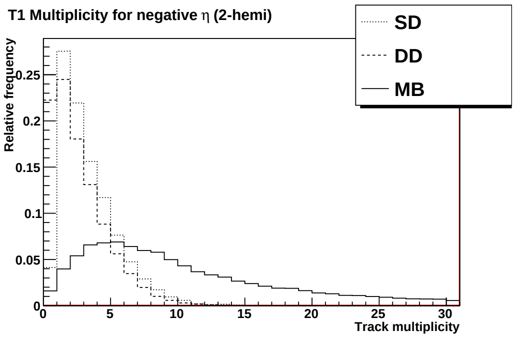

Fig. 4.1: Multiplicities in T1 in negative hemisphere for events with tracks in both hemispheres.

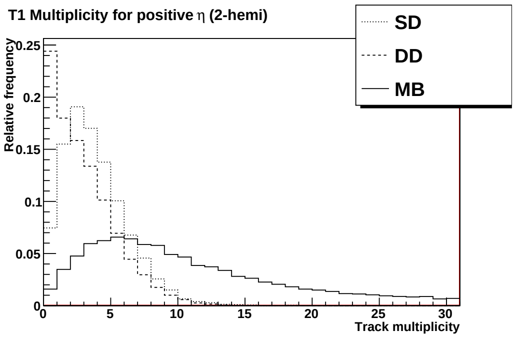

Fig. 4.2: Multiplicities in T1 in positive hemisphere for events with tracks in both hemispheres.

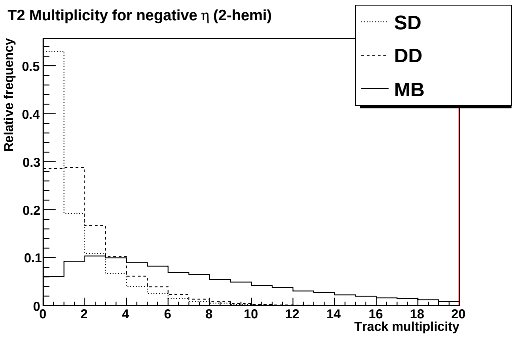

Fig. 4.3: Multiplicities in T2 in negative hemisphere for events with tracks in both hemispheres.

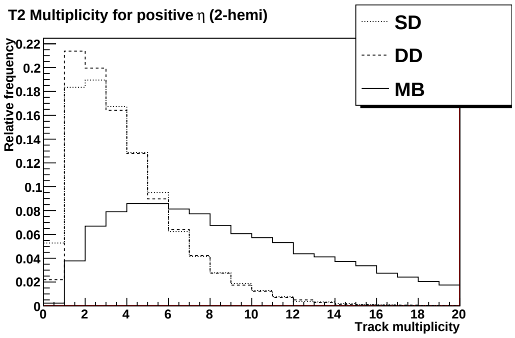

Fig. 4.4: Multiplicities in T2 in positive hemisphere for events with tracks in both hemispheres.

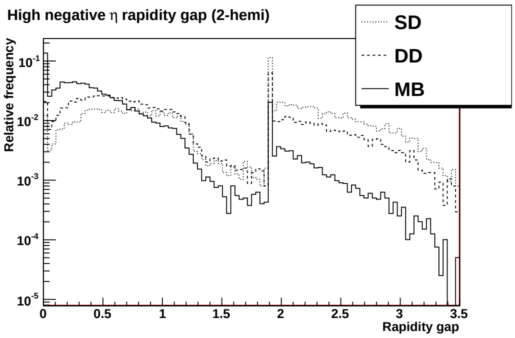

Fig. 4.5: High η rapidity gaps in negative hemisphere for events with tracks in both hemispheres. Due to event flipping all events with non-trivial high η rapidity gaps in the negative hemisphere are two hemisphere events.

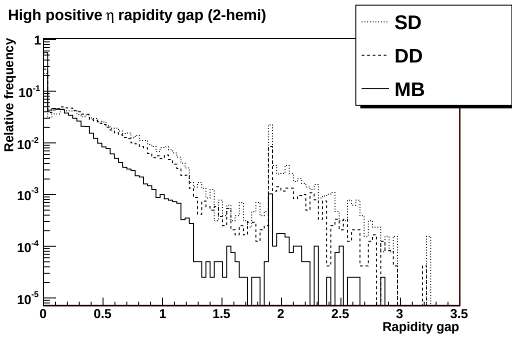

Fig. 4.6: High η rapidity gaps in positive hemisphere for events with tracks in both hemispheres. This variable shows very little power for separation.

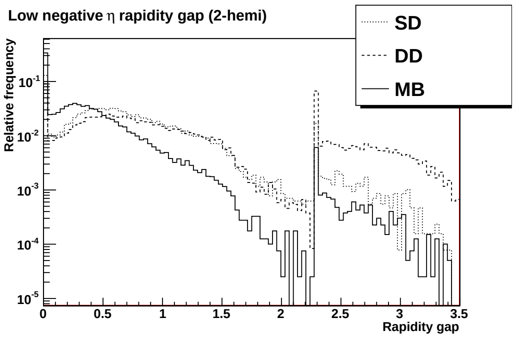

Fig. 4.7: Low η rapidity gaps in negative hemisphere for events with tracks in both hemispheres.

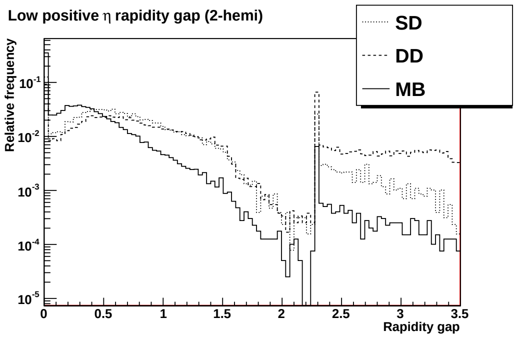

Fig. 4.8: Low η rapidity gaps in positive hemisphere for events with tracks in both hemispheres.

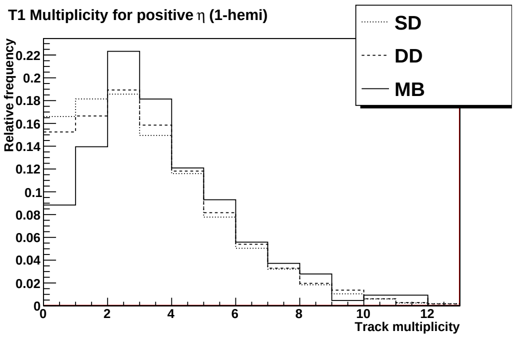

Fig. 4.9: Multiplicities in T1 in negative hemisphere for events with tracks in only one hemisphere.

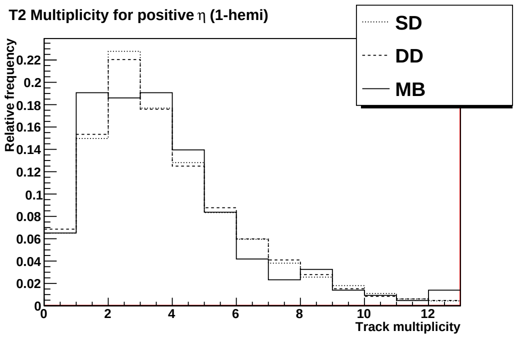

Fig. 4.10: Multiplicities in T2 in positive hemisphere for events with tracks in only one hemisphere.

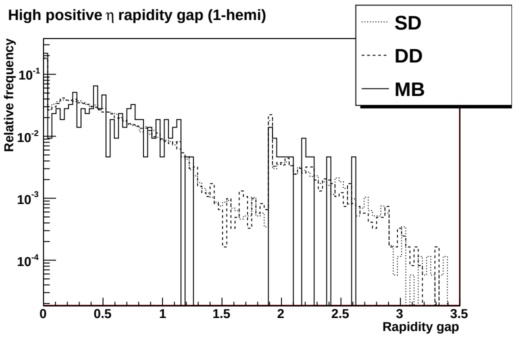

Fig. 4.11: High η rapidity gaps for events with tracks in only one hemisphere.

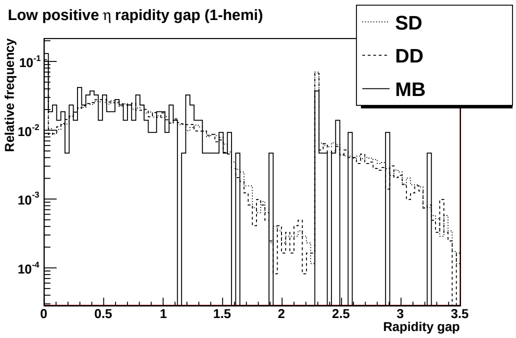

Fig. 4.12: Low η rapidity gaps for events with tracks in only one hemisphere.

between the means and the sum of standard deviations of the groups, is sought. X is formulated as

$$X = \sum_{i=1}^{N} \lambda_i x_i, \tag{4.4}$$

where x_i are the values of the measurements and λ_i are parameters, which are calculated to weight each variable so that optimal separation between the groups is obtained . The differences are represented here by $d_i = \overline{x_i^a} - \overline{x_i^b}$, $i = 1...N$, where x^a and x^b are the sets of measured variables in groups a and b and $\vec{\mu_a}$ and $\vec{\mu_b}$ their means. Their covariances are Σ_a and Σ_b. This gives a difference D between means of X in the two groups as:

$$D = \sum_{i=1}^{N} \lambda_i d_i \tag{4.5}$$

and the variance S of X within groups is proportional to

$$S = \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j S_{ij}, \tag{4.6}$$

where S_{ij} is defined as

$$S_{ij} = (x_i^a - \vec{\mu_a})(x_j^b - \vec{\mu_b})$$

[24].

The linear function which best discriminates the groups is the one with the greatest D^2/S ratio, which is given by individual variation of the coefficients λ_i as shown below [24].

$$\frac{D}{S^2}(2S\frac{\partial D}{\partial \lambda} - D\frac{\partial S}{\partial \lambda}) = 0 \tag{4.7}$$

In practice the Fisher discriminant is calculated using multivariate analysis software or using the formula below.

$$\tilde{\lambda} = (\Sigma_a + \Sigma_b)^{-1} (\vec{\mu_a} - \vec{\mu_b}) \tag{4.8}$$

Using the coefficients λ_i the projection to the 1-dimensional plane can be calculated. The threshold T is given by:

$$T = \vec{\lambda} \cdot \left(\frac{\vec{\mu}_a + \vec{\mu}_b}{2} \right). \tag{4.9}$$

4.5 Artificial Neural Networks

Artificial neural networks are a branch of artificial intelligence. They are simulated collections of interconnected neurons; a neural network can be considered a non-linear discriminant. Each neuron produces a certain response when given certain input signal. Given an external signal to input neurons the network is put in a definite state, which can be measured from one or more output neurons. The artificial neural network therefore maps a set of input variables to one or more output variables, which can be used to classify an event. An artifical neural network can use nonlinear mapping in one or more neurons. [25, 26]

The neural network, which turned out to be the most succesful in the classification process was the newest addition to TMVA, called MLP. The MLP is a feed-forward multilayer perceptron.

How the artificial neural network behaves depends on the layout of the neurons, the weights of the connections between neurons and the response of the neurons to input. The neurons in a multilayer perceptron are organized in layers. Only connections between two adjacent layers are allowed. The first layer of the neural network is called the input layer and the last layer is called the output layer. All layers in between are hidden layers. The input layer for a classification problem with n variables contains n neurons, while the output layer holds one variable. Every connection between neurons is given a weight.[25]

Each neuron has a response function. It maps the neuron input to the neuron output. The weights between neurons are calculated using a method called backpropagation.[25]

4.6 Classifier Response

The output from the classifiers used is a decimal number, which can be either positive or negative. The better the separation between the two classes, the better the classifier works. This section describes the response from the classifiers for PYTHIA simulated processes with a center of mass energy of 7 TeV (thus the beam energy is 3.5 TeV). Events with tracks in both hemispheres (two hemisphere events) have twice the amount of classification variables that events with tracks in only one hemisphere have (one hemisphere events).

The output distributions for the various steps of classification are shown in figures 4.13-4.16. These distributions give a good picture about the usability of the classifier. The distribution for the classifier for two hemisphere events that determines whether an event is diffractive or not (fig. 4.13), shows that the non-diffractive minimum bias

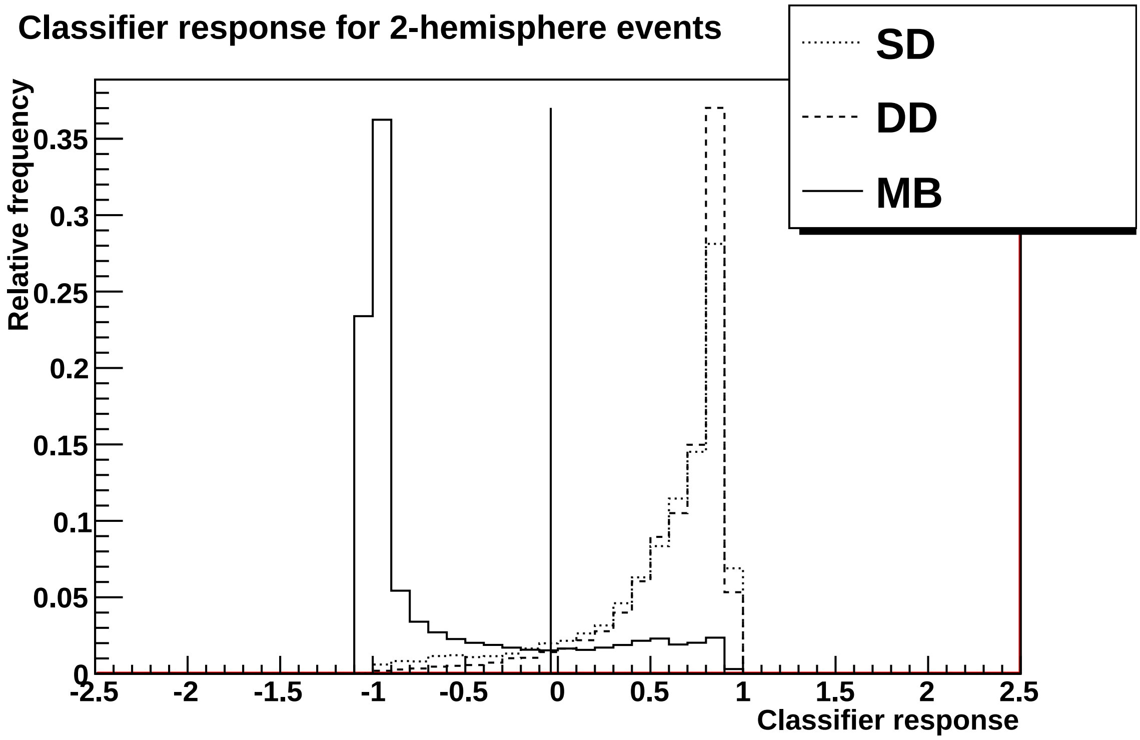

Fig. 4.13: Classifier output for two hemisphere events being classified into diffractive and non-diffractive events. The response is shown for three different data sets; MB stands for non-diffractive minimum bias, DD for double diffractive and SD for single diffractive. The black vertical line shows the threshold for classification; values under this threshold are considered non-diffractive and values above it diffractive.

events are most of the time clearly separated from the diffractive events. The further classification into single and double diffractive events is not as good, but the classifier can still clearly distinguish some single and double diffractive events from each other (fig. 4.14).

Classification of one hemisphere events is poor due to limited information and similar distributions for the classification variables for all event types (fig. 4.15 and 4.16). Therefore classification is not applied for one hemisphere events. However, due to the small number of non-diffractive minimum bias events falling into the one hemisphere category this classifier is not important. Classification between single and double diffractive events is not possible for one hemisphere events. This can be understood from the underlying physics. As seen in figure 2.1, the single and double diffractive processes look the same if particles are seen in only one hemisphere.

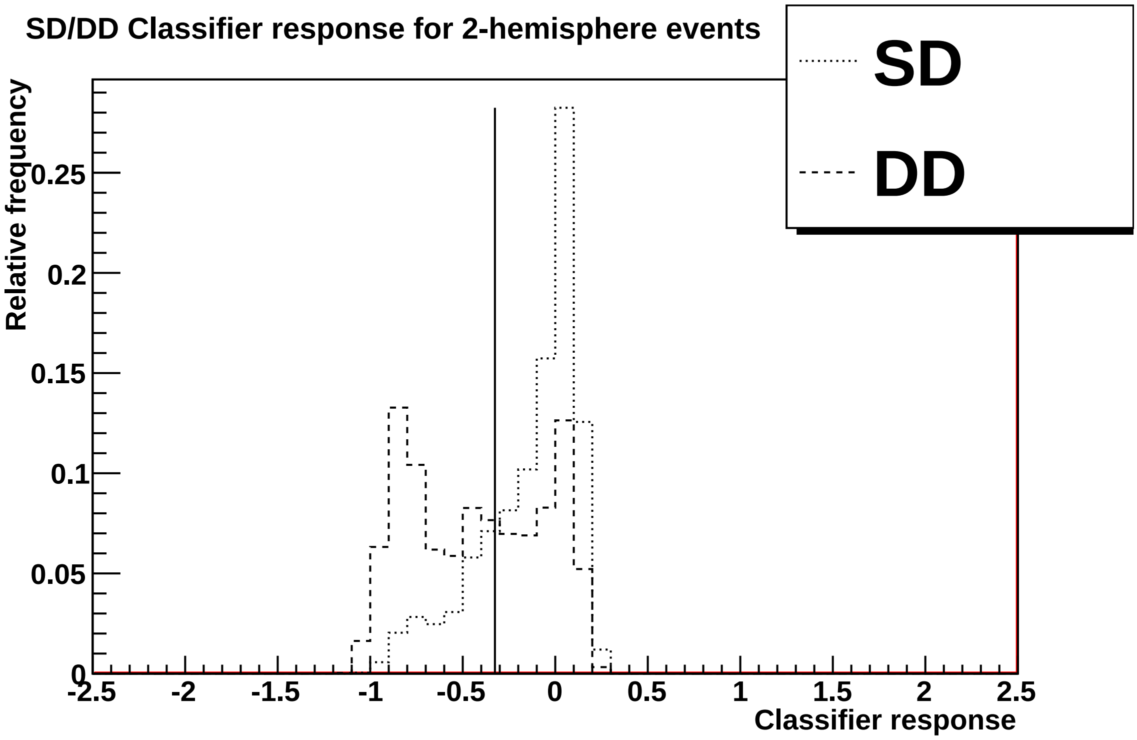

Fig. 4.14: Classifier output for two hemisphere events being classified into single and double diffractive. These events have already in the previous step been classified as diffractive. The black vertical line shows the threshold for classification; values under this threshold are considered double diffractive and values above it single diffractive.

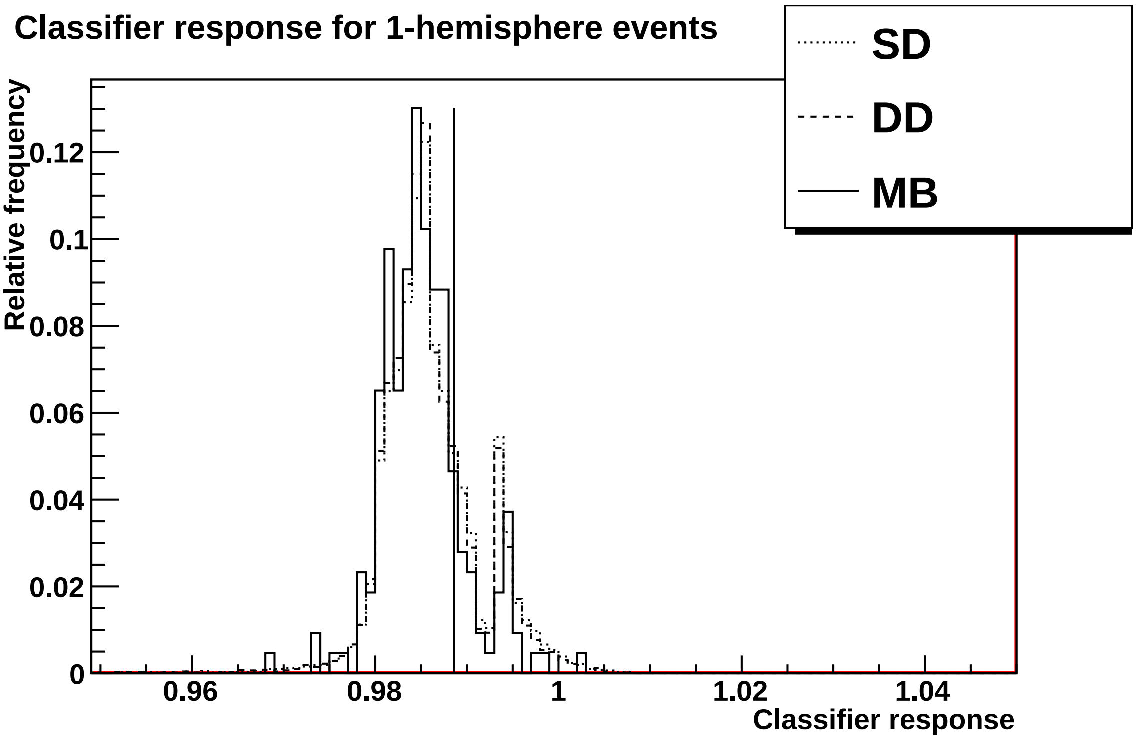

Fig. 4.15: Classifier output for one hemisphere events being classified into diffractive and non-diffractive events. The black vertical line shows the threshold for classification; values under this threshold are considered non-diffractive and values above it diffractive.

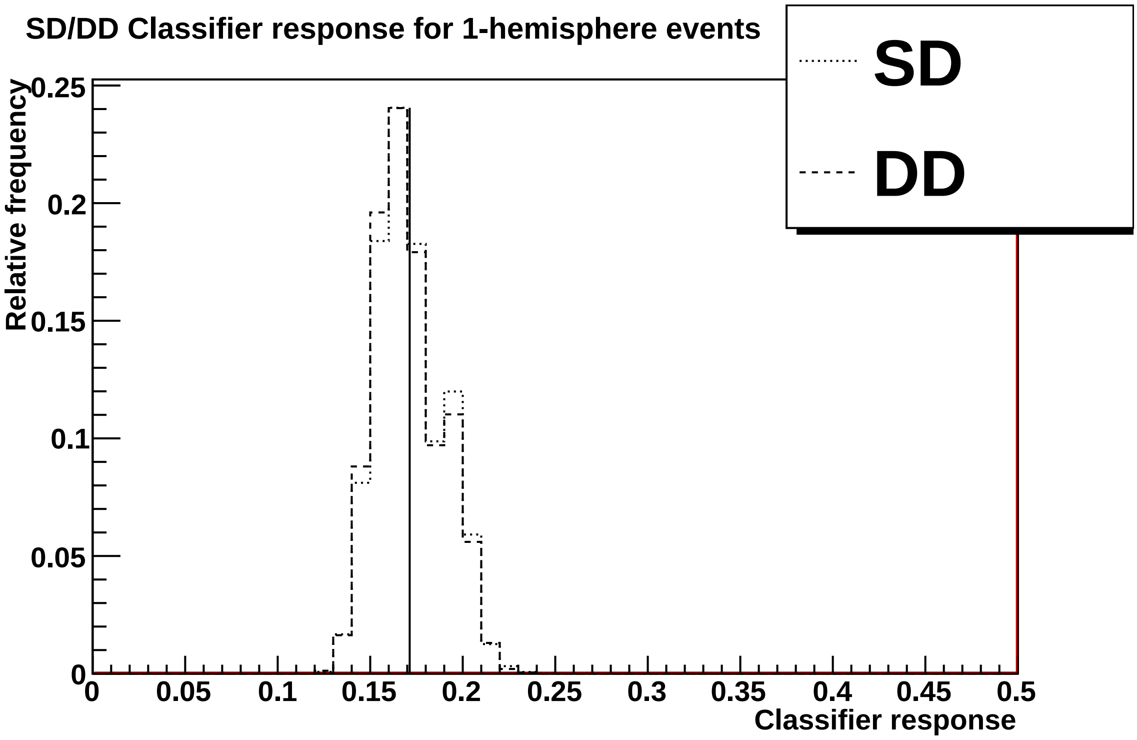

Fig. 4.16: Classifier output for one hemisphere events being classified into single and double diffractive. These events have already been classified as diffractive in the previous step. The black vertical line shows the threshold for classification; values under this threshold are considered double diffractive and values above it single diffractive.

4.7 Results of Classification

Given the classifier output distributions shown in figures 4.13-4.16, thresholds for each of the classes can be determined. If the output from the classifier is below the threshold the event belongs to one class and if above it belongs to the other. Thresholds $T_{Diff,2}$, $T_{SD/DD,2}$, $T_{Diff,1}$ and $T_{SD/DD,1}$ (for classification into non-diffractive or diffractive and single or double diffractive for two hemisphere events and one hemisphere events, respectively) can be chosen by using weighed means of the distributions for different classes, like shown in the formulae below.

$$T_{Diff,2} = \frac{N_{DD,2} \cdot \overline{X_{DD,2}} + N_{SD,2} \cdot \overline{X_{SD,2}} + N_{MB,2} \cdot \overline{X_{MB,2}}}{N_{DD,2} + N_{SD,2} + N_{MB,2}}, \qquad (4.10)$$

where $N_{DD,2}$,$N_{SD,2}$ and $N_{MB,2}$ are the numbers of events in the distribution for two hemisphere classifier response (figure 4.13) for each of the event classes (double diffractive, single diffractive and non-diffractive minimum bias, respectively) and the $\overline{X}$ with identical subscripts are the means of the distributions.

$$T_{SD/DD,2} = \frac{N_{D,2} \cdot \overline{X_{D,2}} + N_{S,2} \cdot \overline{X_{S,2}}}{N_{D,2} + N_{S,2}}, \qquad (4.11)$$

where the N and $\overline{X}$ with subscripts are named in an identical fashion (but now correspond to the distribution in figure 4.14) and S stands for single diffractive and D stands for double diffractive.

The thresholds for events with particles in only one hemisphere can be calculated using the same formulae but using the corresponding distributions for one hemisphere events and renaming the subscripts in the formulae. Using these thresholds each of the events is classified into one and only one of the given classes.

Classification using the thresholds obtained this way (also thresholds are shown in figures 4.13-4.16), result in figures 4.17 (for events with reconstructed particles in both hemispheres) and 4.18 (for events with reconstructed particles in only one hemisphere). The no tracks class in these plots refers to events with no reconstructed particles, whereas the bad tracks class is for events which do not fulfill the criteria for event selection (as explained in section 4.2). Results for two hemisphere events show that 81 % of non-diffractive minimum bias events are recognized and the misclassification into the non-diffractive class is small. Diffractive events are not equally well recognized.

One hemisphere events are more difficult to classify (fig. 4.18). All the negligible non-diffractive minimum bias background is classified into single and double diffractive classes in order not to eliminate a large fraction of the diffractive sample.

An estimate for systematical errors on classification is calculated by adding errors of ±1 on non-zero multiplicities. Table 2 shows the results. Especially classification into non-diffractive minimum bias and single diffractive classes is sensitive to changes in multiplicity. Another estimate for the systematical errors is obtained by comparing to results for data generated with PHOJET.

Classification was also analyzed with data from one half of the T1 telescope removed. This scenario is equal to one half of the T1 system missing in the real detector system. The removed half is in the negative hemisphere and the T1 detector in the positive hemisphere are left intact. This study shows the importance of the T1 detector as part of the detector system. Classification of events with reconstructed particles in both hemispheres resulted in figure 4.19. Some contamination between event types exists as can be seen in the figures 4.17 and 4.19 . It should also be noted that the number of events with tracks in both hemispheres is reduced with one of the T1 arms removed. Classification of the remaining two hemisphere events is somewhat weakened.

Classification results with only half of the T1 system for events with particles reconstructed only in one hemisphere are analyzed separately for the hemisphere with no T1 detector present and the hemisphere with the whole detector present. Especially the classification of non-diffractive minimum bias events is much better in the hemisphere with the T1 detector present; in the hemisphere with the T1 detector present 78 % of the events are correctly classified and in the other hemisphere only 49 % (fig. 4.20). This is explained by non-diffractive events having many particles close to the central region, which the T1 detector are much closer to than the T2 detector. Without the full T1 detector it is useful to use event classification to distinguish between non-diffractive minimum bias and diffractive events also in the case of one hemisphere events. This is because a significant number of non-diffractive minimum bias events are considered one hemisphere events in this scenario (approximately 4 %). Also classification of diffractive events is better with the T1 detector present.

5 Estimation of the Number of Low Mass Diffractive Events

Diffractive mass of a single diffractive event can be obtained from the momentum of the final state proton, when the energies of the initial state protons are known. The square of the diffractive mass M_X is given by:

$$M_X^2 = (p_1 + p_2 - p_3)^2, \tag{5.1}$$

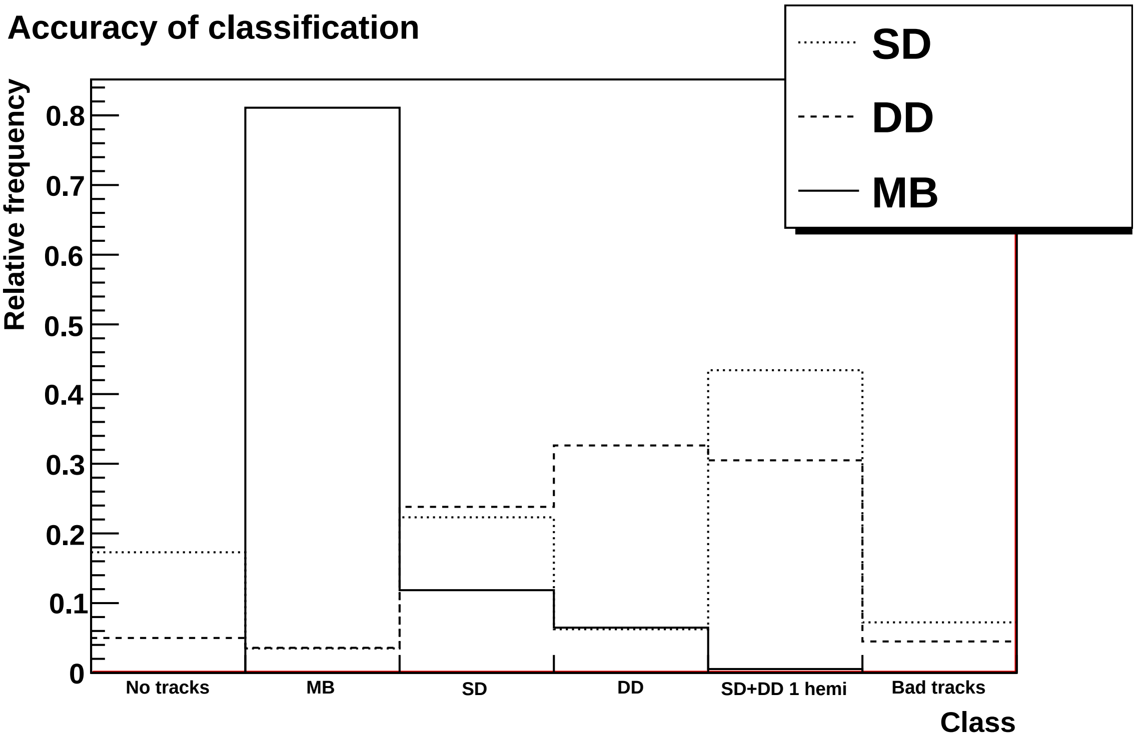

Fig. 4.17: Classification results for events with tracks in two hemispheres. The bins represent the classes which events have been classified into and the different types of lines are for the data sets of events from different classes (continuous line for non-diffractive minimum bias, coarsely dotted line for double diffractive and finely dotted line for single diffractive). Also the fraction of events with reconstructed particles in only one hemisphere (labelled SD+DD 1-hemi) is shown.

Data class	No tracks	MB	SD	DD	1-h	Bad tracks
SD	17^{+3}_{-0}(sy.2)	3 ± 2(sy.1)$^{+5}_{-0}$(sy.2)	22^{+6}_{-0}(sy.1)$^{+0}_{-10}$(sy.2)	$6.2^{+0}_{-1.5}$(sy.1)$^{+0}_{-5}$(sy.2)	43^{+8}_{-0}(sy.1)$^{+10}_{-0}$(sy.2)	$7.2^{+0.4}_{-0}$(sy.2)
MB	$0.01^{+0.03}_{-0}$(sy.2)	81^{+4}_{-8}(sy.1)$^{+0}_{-3}$(sy.2)	12^{+3}_{-2}(sy.1)$^{+2}_{-0}$(sy.2)	7 ± 2(sy.1)$^{+0}_{-2}$(sy.2)	1^{+2}_{-0}(sy.1)$^{+0}_{-0}$(sy.2)	$0.03^{+0.02}_{-0}$(sy.2)
DD	$5.0^{+0}_{-0.6}$(sy.2)	4^{+3}_{-2}(sy.1)$^{+30}_{-0}$(sy.2)	24^{+1}_{-7}(sy.1)$^{+0}_{-7}$(syt.2)	33^{+0}_{-7}(sy.1)$^{+0}_{-13}$(sy.2)	30^{+17}_{-0}(sy.1)$^{+0}_{-4}$(sy.2)	$4.5^{+0}_{-0.9}$(sy.2)

Tab. 2: The classification results in percentages of events from each data set being classified into each category (with systematic errors sy.1 obtained by adding an error of ±1 on non-zero multiplicities and sy.2 errors by comparing to samples generated using PHOJET). Each column represents one of the classes obtained using the classification procedure and each row represents the class the events were simulated for.

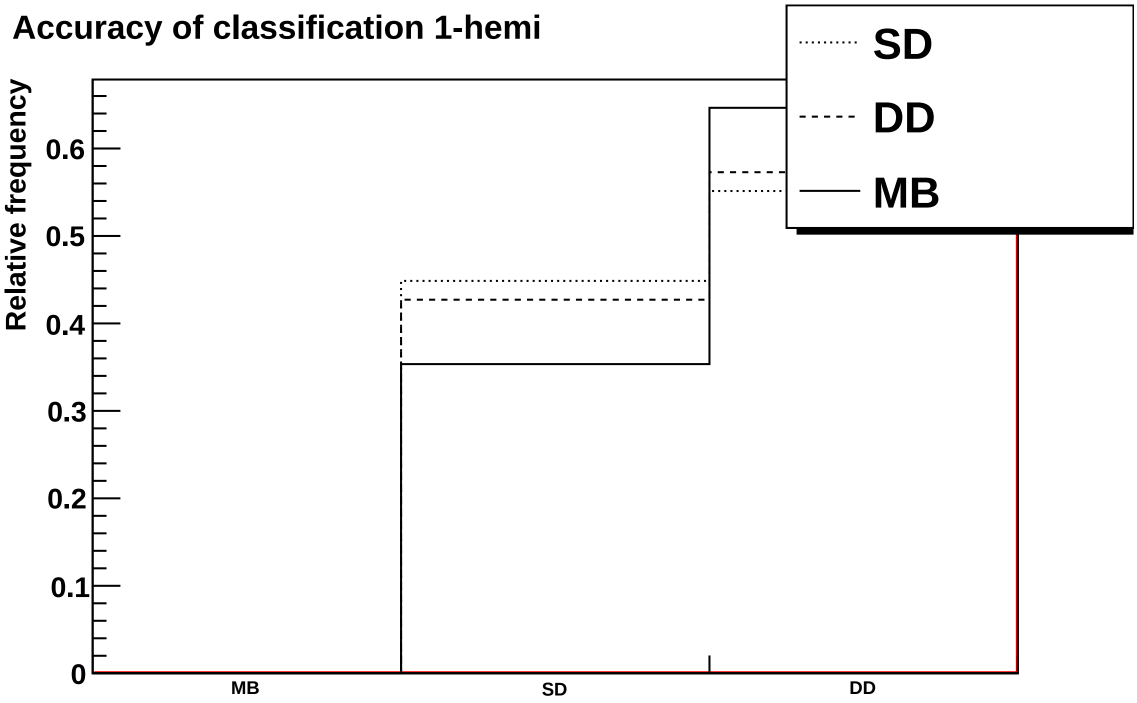

Fig. 4.18: Classification results for events with tracks in only one hemisphere. The bins represent the classes which events have been classified into and the different types of lines are for the data sets of events from different classes (continuous line for non-diffractive minimum bias, coarsely dotted line for double diffractive and finely dotted line for single diffractive).

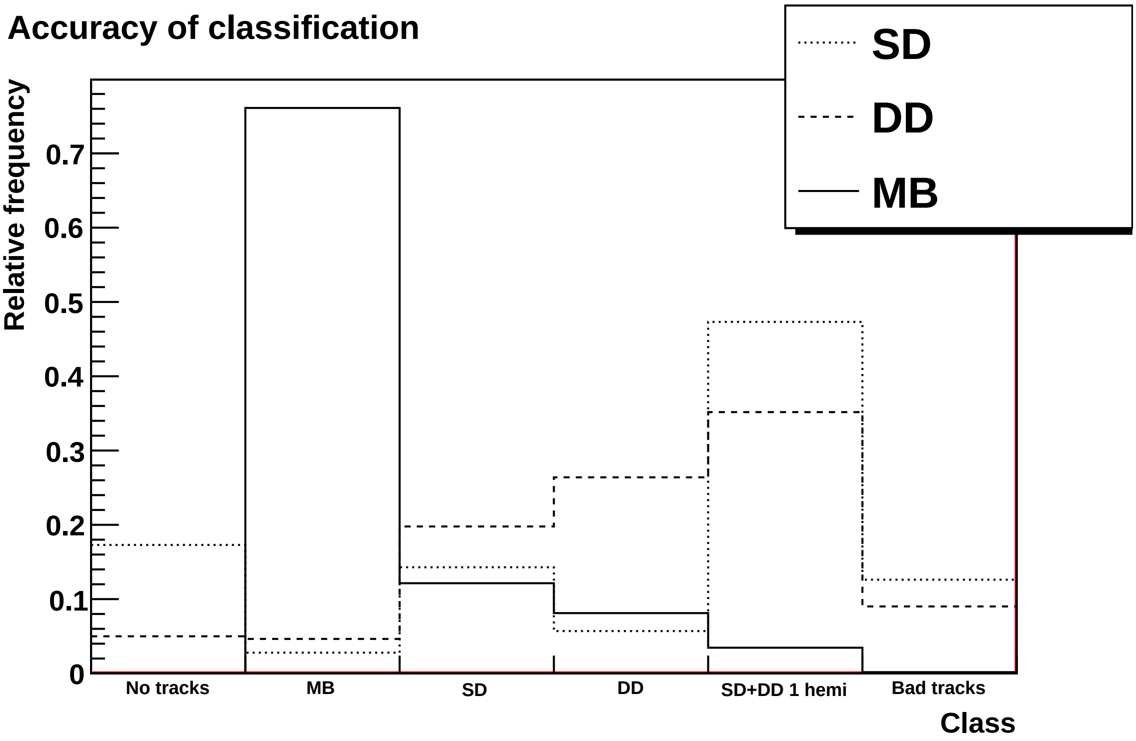

Fig. 4.19: Classification of events with tracks in both hemispheres with T1 detectors only in the positive hemisphere.

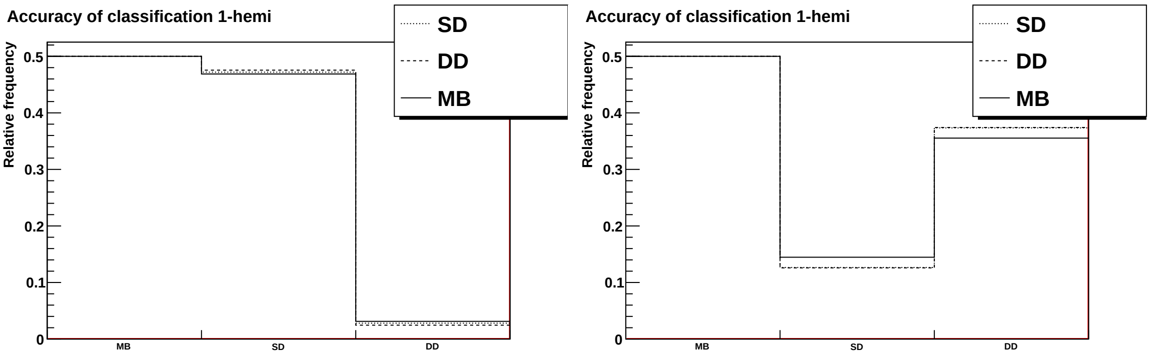

Fig. 4.20: Classification accuracy for events with tracks in only one hemisphere and only one half of the T1 system installed. The figure on the left is for the hemisphere with no T1 detectors and the one on the right for the hemisphere with all T1 detectors present.

where p_1, p_2 and p_3 are the four-momenta of the protons corresponding to each of the two initial state protons and the final state proton, respectively.[27]

This result can be greatly simplified. The metrics tensor $\eta^{\mu\nu}$used in derivation of the final result is:

$$\eta^{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} . \tag{5.2}$$

When factorised, equation 5.1 becomes:

$$M_X^2 = E_1^2+E_2^2+E_3^2+2E_1\cdot E_2-2E_1\cdot E_3-2E_2\cdot E_3-\vec{p}_1^2-\vec{p}_2^2-\vec{p}_3^2-2\vec{p}_1\cdot\vec{p}_2+2\vec{p}_1\cdot\vec{p}_3+2\vec{p}_2\cdot\vec{p}_3, \tag{5.3}$$

where $\vec{p}_i$ are the three-momenta of the protons and E_i their energies. Assuming high energy the energies and momenta in a head-on collision of two beams with energy E_i and a final state proton with energy E_f can be approximately described by the following set of equations using natural units (a system of units where the velocity of light $c = 1$ and the Plank constant $\hbar = 1$ and energy and momentum have the same unit eV). It is also assumed that the final state proton has a large longitudinal momentum and negligible transverse momentum (this is true for one hemisphere events).

$$E_1 = E_2 = E_i \tag{5.4}$$

$$E_3 = E_f \tag{5.5}$$

$$p_{1x} = p_{2x} = p_{3y} = 0 \tag{5.6}$$

$$p_{1y} = p_{2y} = p_{3y} = 0 \tag{5.7}$$

$$p_{1z} = -p_{2z} = E_i \tag{5.8}$$

$$p_{3z} = E_f \tag{5.9}$$

Given these equations equation 5.3 becomes in the high energy limit:

$$M_X^2 = 4E_i^2 - 4E_iE_f = 4E_i^2(1 - \frac{E_f}{E_i}) \tag{5.10}$$

and the diffractive mass is therefore:

$$M_X = 2E_i\sqrt{1 - \frac{E_f}{E_i}}. \tag{5.11}$$

Further expressing this in terms of the Mandelstam variable s, which is the square of the center-of-mass energy (that is $s = (2E_i)^2$ and the final state momentum p_f (which is the same as the energy in the high energy limit), the equation can be expressed as follows.

$$M_X = \sqrt{s(1 - \frac{p_f}{0.5\sqrt{s}})} \tag{5.12}$$

This equation for the diffractive mass works well if the beam energy is precisely $0.5\sqrt{s}$. In case the energy of the initial state protons differ from this energy, it can affect the value of the calculated diffractive mass by several GeV.

The diffractive mass distribution is calculated for the simulated sample of diffractive events. The sample needs to be as pure as possible to allow for a reasonably accurate distribution (not much contribution from non-diffractive minimum bias events). Even with no reconstructed final state leading protons, the diffractive masses can be calculated using the rapidity gaps for events with tracks in only one hemisphere. Here the rapidity gap $\Delta\eta$ is defined as

$$\Delta\eta = \Delta\eta_{low} - (y_{cm})_{max} = \Delta\eta_{low} - \ln(\frac{\sqrt{s}}{m_p}), \tag{5.13}$$

where $\Delta\eta_{low}$ is the rapidity gap as defined in equation 4.1 for the hemisphere with reconstructed particle tracks, $(y_{cm})_{max}$ is the maximal rapidity allowed by kinematics. Here it is assumed that the energy of the particles is high enough to neglect rest masses, in which case rapidity and pseudorapidity are approximately equal. Furthermore, $\sqrt{s}$ is the center of mass energy for the colliding protons and m_p is the proton mass.

For 7.0 TeV center of mass energy $(y_{cm})_{max}$ becomes

$$(y_{cm})_{max} = \ln(\frac{7 * 10^6\,\text{MeV}}{938.3\,\text{MeV}}) = 8.917,$$

where the proton mass is taken as 938.3 MeV.[28]

The measured diffractive mass distribution differs significantly from the simulated diffractive mass distribution. The main reasons for this are the limited acceptance of the detectors, mainly in the central region in pseudorapidity (that is $|\eta| < 3.0$) and

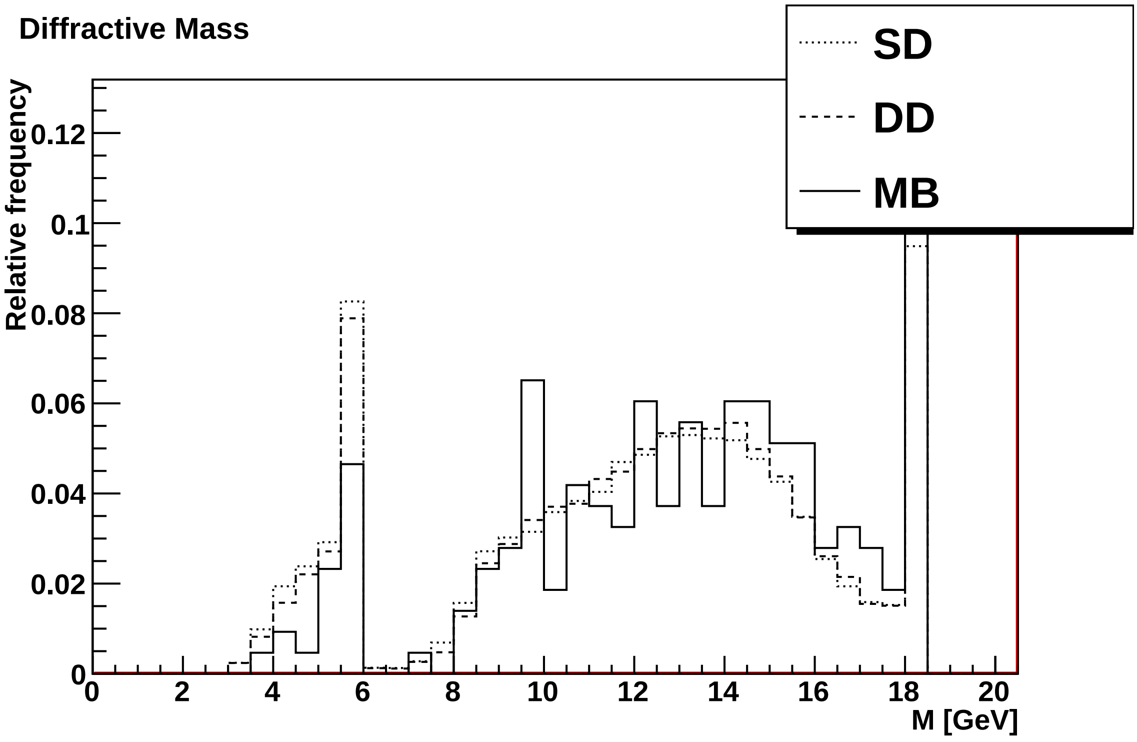

Fig. 5.1: Reconstructed diffractive mass for events with reconstructed particles in one hemisphere and classified as diffractive.

the gap between the T1 and T2 acceptance for $4.6 < |\eta| < 5.3$. The gap is visible in the mass range $M_X = 6 - 8\,\text{GeV}$ in the reconstructed diffractive mass distribution in figure 5.1. Masses over 18 GeV are mainly seen as a peak at 18 GeV, due to the lower limit of acceptance $|\eta| \approx 3.0$ and masses below 2.5 GeV are not seen due to the upper limit at $|\eta| \approx 6.5$.

The effect on the reconstructed diffractive mass distribution of having the T1 detector in only one hemisphere was also examined. The effect of missing half of the T1 detector can be seen in figure 5.2. Not having the T1 detector clearly cuts of the higher masses, which are seen with T1 detectors present. The simulated mass distribution extends beyond the acceptance of T1, but T1 often sees some of the particles of such events and they become reconstructed at smaller diffractive masses. Without the T1 detector these events are reconstructed with a mass of approximately 6 GeV.

5.1 Unfolding

A method of obtaining an efficiency and acceptance corrected diffractive mass distribution is needed to be able to estimate the total number of diffractive events. The measured distribution is distorted by limited acceptance and efficiency. By using both

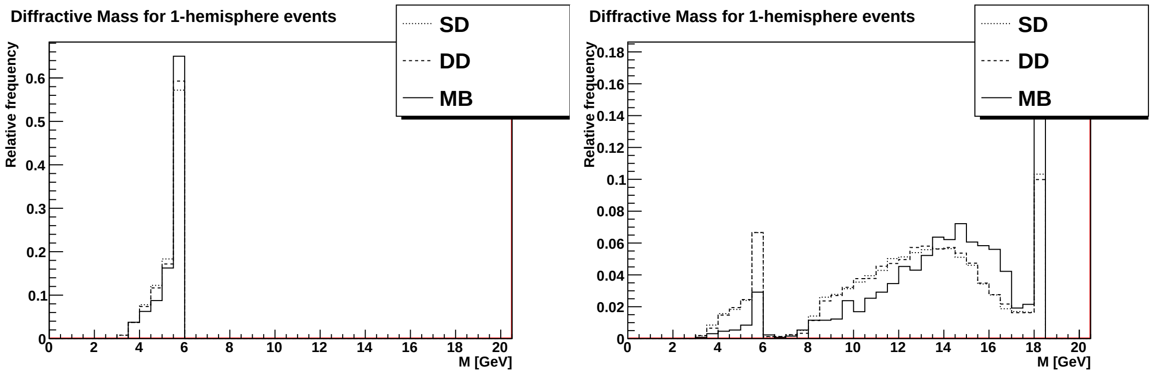

Fig. 5.2: Diffractive mass distributions for one hemisphere events with one half of the T1 detector system missing. The figure on the left is for the hemisphere with no T1 detector and the one on the right for the hemisphere with the T1 detector present.

the simulated and reconstructed diffractive mass distributions an unfolding algorithm can be built. Unfolding transforms the reconstructed distribution to reproduce the simulated distribution. As long as the Monte Carlo event generator is realistic, this can produce a usable unfolding algorithm. To be useful, the unfolding algorithm should be able to reproduce the true distribution also with real data. This is extremely difficult to validate, because the true distribution is not known. In essence the unfolding algorithm should work like an inverse matrix for the matrix given by the correlation between simulated and reconstructed diffractive masses (see figure 5.3). This figure clearly shows that there is an upper limit on reconstructed diffractive mass given by the acceptance of T1 and that many events are reconstructed with lower diffractive mass than their real mass.

Unfolding is not only useful to get a result comparable with theory, but it is also required to produce results which are comparable with other experiments and configurations. The unfolded result is directly comparable with both theoretical predictions and results from other experiments. Unfolding can be used to compensate for physics and detector effects in an experiment.[29, 30]

The unfolding problem is formulated as follows. A random variable x is measured with limited resolution, presence of contaminating background as well as less than full efficiency. Limited resolution means here that due to random fluctuations the measured value of x can differ from its true value. Each observed value of x is accompanied by the true value, represented by y. The true value is only known in case of simulation. In many cases a true value y is outside the acceptance of the detectors, in which case there might not even be a measured value x. This causes limited detection efficiency ε, which often depends on the value of y.

In order to be able to construct a usable estimator for the probability density

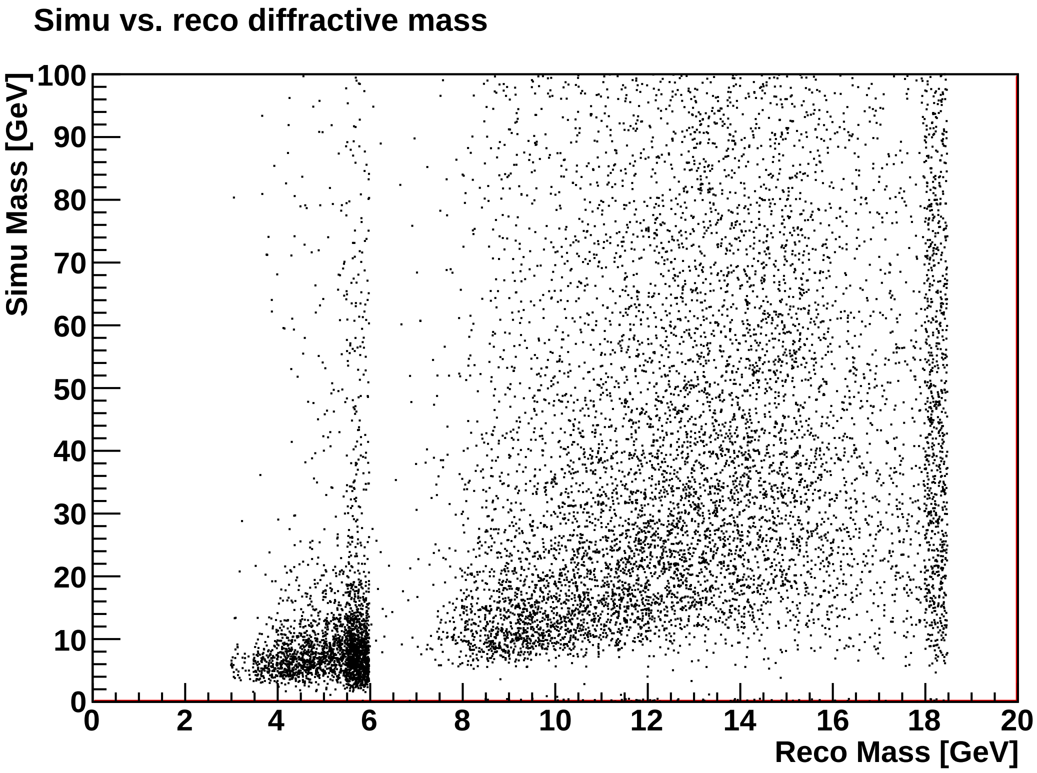

Fig. 5.3: The correlation between simulated diffractive mass and reconstructed diffractive mass.

function ($p.d.f.$) of the true values $f_{true}(y)$, it must be represented by means of a finite set of parameters. Even without knowledge of the functional form of $f_{true}(y)$, it can be represented by a normalized histogram with M bins. The probability p_j of finding y in bin j is the integral over the bin.

$$p_j = \int_j f_{true}(y)\,dy \tag{5.14}$$

The total number of events that occur when performing an experiment n_{tot} is often not the same as the number observed. An expectation value for the total number of events can be given as $\mu_{tot} = E[n_{tot}]$, which gives the expected number of events in bin j as $\mu_j = \mu_{tot} p_j$. These numbers μ_j are the contents of the bins of the true histogram. In this case the diffractive mass distribution given by simulation is considered the true histogram. These numbers μ_j are not directly observable.

Reconstructed diffractive masses are filled into the histogram of observed values, with n_i entries in the i:th bin and a total number of entries $n_{tot} = \sum_i n_i$. These variables n_i are considered independent Poisson variables with expectation values ν_i. This gives the probability of observing n_i entries in bin i as

$$P(n_i, \nu_i) = \frac{\nu_i^{n_i} \times e^{-\nu_i}}{n_i!}. \tag{5.15}$$

A sum of Poisson variables is in itself a Poisson variable and therefore n_{tot} will follow a Poisson distribution with expectation value $\nu_{tot} = \sum_i \nu_i$. The law of probability gives the expectation values $\nu_i = E[n_i]$ as:

$$\nu_i = \mu_{tot} P(event\ observed\ in\ bin\ i)$$

$$= \mu_{tot} \int dy\ P(observed\ in\ bin\ i \mid true\ value\ y\ and\ event\ detected)\ \varepsilon(y)\ f_{true}(y)$$

$$= \mu_{tot} \int_{bin\ i} dx \int dy\ s(x \mid y)\varepsilon(y)\ f_{true}(y), \tag{5.16}$$

where $s(x \mid y)$ is the conditional probability distribution function for the measured value x given the true value y under the assumption that the event was observed somewhere, giving a normalization $\int s(x \mid y)\, dx = 1$. The true distribution is *folded* with the response function and the estimation of f_{true} is called *unfolding*.

Turning the integral over y into a sum over bins the expected number of observed entries in bin i can be written as:

$$\nu_i = \sum_{j=1}^{M} \frac{\int_{bin\ i} dx \int_{bin\ j} dy\ s(x \mid y)\varepsilon(y)\ f_{true}(y)}{(\mu_j/\mu_{tot})} \mu_j = \sum_{j=1}^{M} R_{ij}\mu_j, \tag{5.17}$$

with the response matrix R defined as:

$$R_{ij} = \frac{\int_{bin\ i} dx \int_{bin\ j} dy\ s(x \mid y)\varepsilon(y)\ f_{true}(y)}{\int_{bin\ j} dy\ f_{true}(y)}$$

$$= \frac{P(observed\ in\ bin\ i\ and\ true\ value\ in\ bin\ j)}{P(true\ value\ in\ bin\ j)}$$

$$= P(observed\ in\ bin\ i \mid true\ value\ in\ bin\ j). \tag{5.18}$$

The response matrix elements R_{ij} represent the conditional probability of finding an event with measured value x in bin i given that the true value y was in bin j. This matrix depends on the probability distribution function $f_{true}(y)$. Since the goal is to determine $f_{true}(y)$, this is unknown. However, with small enough bins, the direct dependence on $f_{true}(y)$ cancels out. This is assumed to be true in this analysis. The response matrix R_{ij} is usually not symmetric. Summing over the first index i over the bins in the observed histogram and using the normalization $\int s(x \mid y)\,dx = 1$ gives:

$$\begin{aligned}\sum_{i=1}^{N} R_{ij} &= \sum_{i=1}^{N} \frac{\int_{bin\,i} dx \int_{bin\,j} dy\; s(x \mid y)\varepsilon(y)\; f_{true}(y)}{(\mu_j/\mu_{tot})} \\ &= \frac{\int_{bin\,j} dy\; \varepsilon(y)\; f_{true}(y)}{\int_{bin\,j} f_{true}(y)\; dy} \\ &\equiv \varepsilon_j, \end{aligned} \tag{5.19}$$

which is the average value of the efficiency over bin j.

In addition to limited resolution and efficiency also possible background needs to be taken into account. The detector can detect a value x which does not correspond to a true value y. The contribution of background processes and noise β_i in bin i must be added to equation 5.17 to get the true expectation value. This equation is thus modified as follows.

$$\nu_i = \sum_{j=1}^{M} R_{ij}\mu_j + \beta_i \tag{5.20}$$

The complete set of vector quantities that constitute the unfolding problem are thus:

1. The true histogram $\mu = (\mu_1, ..., \mu_M)$,
2. The normalized true histogram $p = (p_1, ..., p_M) = \mu/\mu_{tot}$,
3. Expectation values of the observed numbers of entries $\nu = (\nu_1, ..., \nu_N)$,
4. Actual number of entries observed $n = (n_1, ..., n_N)$,
5. Efficiencies $\varepsilon = (\varepsilon_1, ..., \varepsilon_M)$ and
6. Expected background values $\beta = (\beta_1, ..., \beta_N)$.

Either the probability distribution for the data n or the covariance matrix $V_{ij} = \mathrm{cov}[n_i, n_j]$ is required to construct the likelihood function or a χ^2 function, respectively. In addition the response matrix R_{ij} and the contribution from background β must be known. The relation betveen these can be expressed as vectors (as in equation 5.20 for the elements) as

$$\nu = R\mu + \beta, \tag{5.21}$$

where μ, ν and β are understood as column vectors. The goal is to construct estimators $\hat{\mu}$ for the true histogram or esimators $\hat{p}$ for the probabilities. This can be done by inverting the response matrix, but this often leads to a poor solution. Inversion of the response matrix is also not possible if the matrix is singular and there is no reason for the inverse matrix to exist. This can be seen by taking two bins of true quantities, both of which have the same probability to be observed in each of the bins of the measured quantity. This suggests that treating the probability density functions as vectors in space is not the correct approach. Even if the response matrix can has an inverse, the method of inversing the matrix does not work with large statistical fluctuations. Negative terms in the inverse matrix could lead to negative numbers of entries in bins of the unfolded distribution. [29, 30]

An alternative is the method of correction factors imposed bin-by-bin. This leads to a highly model dependent result, which is not acceptable because the training is done based on a Monte Carlo model, which probably gives rise to different distributions than what the real processes in reality would give. One major problem with the bin-by-bin approach is also that it cannot take into account migration from one bin to another. This model can thus be used only if the migration is negligible and the standard deviation of the smearing is smaller than the bin size. Therefore the best choice is to use regularized unfolding, which is not as model dependent. This is done by imposing a measure of smoothness on the estimators for the true histogram μ. Regularized unfolding usually provides satisfactory results. [29, 30]

Acceptable solutions are chosen by choosing a certain region of μ-space around the maximum likelihood solution, which has a good level of agreement between the predicted expectation values ν and the data n. In addition to determining the acceptability of the solution, a measure of smoothness is defined by introducing a function $S(\mu)$, which is known as the *regularization function*. The solution with the highest degree of smoothness out of all acceptable solutions is chosen.

The regularization function can be based on the entropy H of a probability distribution $p = (p_1, ..., p_M)$. The entropy H is defined as

$$H = -\sum_{i=1}^{M} p_i \log p_i. \tag{5.22}$$

[29, 31]

The entropy is interpreted as a measure of smoothness of the histogram $\mu = (\mu_1, ..., \mu_M)$. The following is then used as regularization function.

$$S(\mu) = H(\mu) = -\sum_{i=1}^{M} \frac{\mu_i}{\mu_{tot}} \log \frac{\mu_i}{\mu_{tot}} \tag{5.23}$$

This kind of estimator is said to be constructed according to the *principle of maximum entropy*. The relation between entropy and smoothness is found by considering in how many ways a particular histogram $\mu = (\mu_1, ..., \mu_M)$ can be constructed out of μ_{tot} entries (assuming the values of μ_j to be integer), which is given by

$$\Omega(\mu) = \frac{\mu_{tot}!}{\mu_1!\mu_2!...\mu_M!}. \tag{5.24}$$

By taking the logarithm of equation 5.24 and using the Stirling approximation $\log n! \approx n(\log n - 1)$ for large n the equation becomes:

$$\begin{aligned}\log\Omega &\approx \mu_{tot}(\log\mu_{tot} - 1) - \sum_{i=1}^{M} \mu_i(\log\mu_i - 1)\\ &= -\sum_{i=1}^{M} \mu_i \log\frac{\mu_i}{\mu_{tot}}\\ &= \mu_{tot}S(\mu), \end{aligned} \tag{5.25}$$

which can also be generalized to non-integer μ_i.

Bayes' theorem then gives for $f(\mu \mid n)$:

$$f(\mu \mid n) \propto L(n \mid \mu)\pi(\mu), \tag{5.26}$$

where $L(n \mid \mu)$ is the likelihood function multiplied by the prior density $\pi(\mu)$, which represents the knowledge of μ before using the data n.

In this approach, it is assumed that the total number of events μ_{tot} is an integer, whereas in the classical approach it is an expectation value of a random integer variable (which can be non-integer). The prior probability $\pi(\mu)$ is interpreted to be given by the total number of ways of distributing the entries $\Omega(\mu)$ given by equation 5.24. Thus the following can be written for $\pi(\mu)$:

$$\begin{aligned}\pi(\mu) = \Omega(\mu) &= \frac{\mu_{tot}!}{\mu_1!\mu_2!...\mu_M!}\\ &= \exp(\mu_{tot}H), \end{aligned} \tag{5.27}$$

where the entropy H is given by equation 5.22.

A strict Bayesian approach would be to determine $f(\mu \mid n)$, but this is not practical to report completely, because it is a function of as many variables as there are bins in the unfolded distribution. A way of summarizing the function is to use a single vector $\hat{\mu}$ as Bayesian estimator. This vector is chosen as the μ, which maximizes the logarithm of $f(\mu \mid n)$. This is found by maximizing equation 5.26.[29]

$$\log f(\mu \mid n) \propto \log L(\mu \mid n) + \log \pi(\mu)$$

$$= \log L(\mu \mid n) + \mu_{tot} H(\mu) \tag{5.28}$$

This corresponds to using a regularization function $S(\mu)$.

$$S(\mu) = \mu_{tot} H(\mu) = -\sum_{i=1}^{M} \mu_i \log \frac{\mu_i}{\mu_{tot}} \tag{5.29}$$

The regularization paramter α is now set equal to 1. If all the efficiencies ε_i are equal, the requirement $\nu_{tot} = n_{tot}$ requires that μ_{tot} be constant. This is equivalent to using the regularization function $S(\mu) = H$ with $\alpha = \frac{1}{\mu_{tot}}$. If the efficiencies are not equal, constant ν_{tot} does not imply constant n_{tot} and the distribution of maximum $S(\mu) = \mu_{tot} H(\mu)$ is not uniform. The distribution of $S = H$ is easy to justify, because the distribution of maximum H is always uniform.

The principle of maximum entropy can be developed in the framework of Bayesian statistics using an iterative approach. Some advantages of using a iterative approach based on Bayes' theorem are:

- it is well grounded in theory,
- it can handle cells of different sizes for the true and measured distributions,
- the domains of definition for measured and true values can differ,
- any smearing and migration can be taken into account,
- background can be taken into account,
- no matrix inversion required and
- it usually gives the best results.

In this approach, the μ are treated as random variables and the joint probability density $f(\mu \mid n)$ is the degree of belief that the true histogram is given by μ. Bayes' theorem in terms of N different causes C_i, which can produce one effect E, knowing the initial probability of the causes $P(C_i)$ and the conditional probability of cause C_i to produce the effect $P(E \mid C_i)$ can be stated in a formula as follows.[29, 30]

$$P(C_i \mid E) = \frac{P(E \mid C_i)P(C_i)}{\sum_{i=1}^{N} P(E \mid C_i)P(C_i)} \tag{5.30}$$

This can be understood as if a single event is observed, the probability that it has been due to the i:th cause is proportional to the probability of the cause times the probability of the cause to produce this effect. $P(C_i|E)$ clearly depends on the initial probability of the causes. [30]

If $n(E)$ events with effect E are observed, then the expected number of events attributed to each of the causes is:

$$\hat{n}(C_i) = n(E)P(C_i \mid E). \tag{5.31}$$

Each measurement can give several possible effects E_j for each cause C_i. The Bayes' formula is true for each of them. For several possible effects Bayes' formula (equation 5.30) then becomes as follows.[30]

$$P(C_i \mid E_j) = \frac{P(E_j \mid C_i)P_0(C_i)}{\sum_{i=1}^{N} P(E_j \mid C_i)P_0(C_i)} \tag{5.32}$$

A set of N observations gives rise to a distribution of frequencies $n(E) = \{n(E_1), n(E_2), ..., n(E_N)\}$. The expected number of events associated to each of the causes due to the observed events is calculated using equation 5.31.

$$\hat{n}(C_i)|_{obs} = \sum_{j=1}^{N} n(E_j)P(C_i \mid E_j) \tag{5.33}$$

Adding the effect of efficiencies $\varepsilon_i \neq 0$ the formula becomes as follows.

$$\hat{n}(C_i)|_{obs} = \frac{1}{\varepsilon_i}\sum_{j=1}^{N} n(E_j)P(C_i \mid E_j) \tag{5.34}$$

Using these unfolded events an estimate of the true total number of events can be made as well as the final probabilities of the causes and the overall efficiency.[30]

$$\hat{N}_{true} = \sum_{i=1}^{N} \hat{n}(C_i) \tag{5.35}$$

$$\hat{P}(C_i) = P(C_i \mid n(E)) = \frac{\hat{n}(C_i)}{\hat{N}_{true}} \tag{5.36}$$

$$\hat{\varepsilon} = \frac{N_{obs}}{\hat{N}_{true}} \tag{5.37}$$

The need of knowledge of the initial distribution can be overcome by using an iterative procedure for the Bayes approach. Such an approach can also be extended to multidimensional problems. If the initial distribution $P_0(C)$ is not consistent with the data, it will not agree with the final distribution $\hat{P}(C)$. The agreement gets better as the initial distribution gets closer to the true distribution. For simulated data, $\hat{P}(C)$ lies between $P_0(C)$ and the true one. Therefore an iterative approach is usable. This can be done by starting of with an initial distribution $P_0(C)$, which is chosen according to the best of knowledge. This implies also choosing the expected number of events $n_0(C_i) = P_0(C_i)N_{obs}$. Without prior knowlede, the initial distribution can be chosen as a uniform distribution $P_0(C_i) = \frac{1}{N}$. After the distribution is chosen $\hat{n}(C)$ and $\hat{P}(C)$ are calculated. The χ^2 of $\hat{n}(C)$ and $n_0(C)$ are compared. After this $P_0(C)$ is replaced by $\hat{P}(C)$ and $n_0(C)$ by $\hat{n}(C)$. This process is repeated until the χ^2 value is small enough.[30]

Limited acceptance and efficiency clearly causes the reconstructed diffractive mass distribution to differ especially on very low mass and very high mass and in the region between the acceptance of T1 and T2 telescopes. Unfolding this distribution with the iterative Bayes' method described above gives satisfactory result. Because the difference between the two distributions is mainly an effect of acceptance it can be corrected by unfolding. The main areas that need correction are the gap between T1 and T2 and the high values diffractive masses.

Unfolding is done using the additional unfolding library for ROOT, called RooUnfold: ROOT Unfolding Framework, which supports uniform or non-uniform binning. Also ROOT supports variable bin sizes, but in the case of 2-dimensional histograms, variable bin sizes are limited to one of the axes. Therefore one or both of the distributions need to be given with fixed bin sizes and the correlation histogram must have identical binning.

The fraction of events reconstructed as one hemisphere events depends on the diffractive mass (see figure 5.4). Most of these events originate from single diffractive processes and some from double diffractive processes. The approximate cross sections used in these studies are $\sigma_{SD} = 12.5$ mb for single diffractive, $\sigma_{DD} = 7.5$ mb for double diffractive and $\sigma_{MB} = 50$ mb for non-diffractive minimum bias (these values are compatible with theoretical predictions shown in table 1).

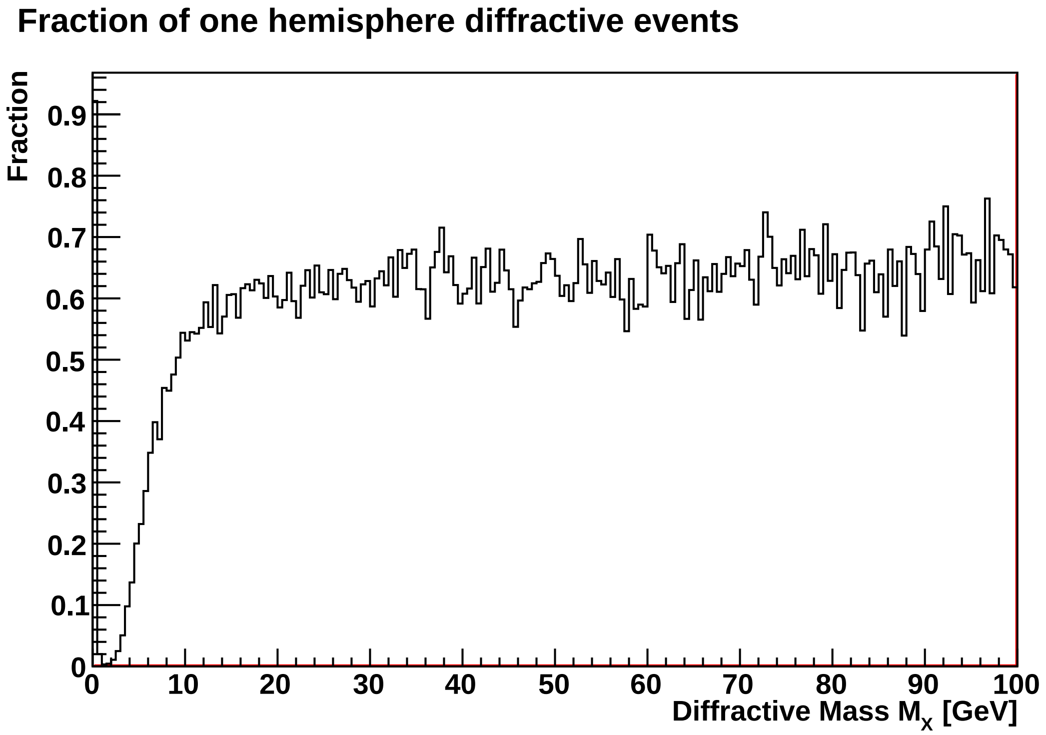

Fig. 5.4: The percentage of events reconstructed as 1-hemisphere events of all simulated events as a function of simulated diffractive mass. This is a weighted mean of single diffractive and double diffractive distributions with weights as their approximative cross sections.

5.2 Estimation of Number of Events

Because of limited or no acceptance in the region of low diffractive mass ($M_X <$ 10 GeV), a majority of diffractive events with diffractive masses in this region are missed. An estimate of the number of missed events is needed for the calculation of the inelastic rate. There are two possible approaches to calculate this estimate. One option is to choose events with only one proton reconstructed in the final state and no particles detected in the forward trackign telescopes. These events are likely to be single diffractive events with a low mass diffractive system in the hemisphere opposite the detected proton. This accounts for most of the low mass single diffractive events, but low mass double diffractive events are missed. The majority of low mass diffractive events are, however, single diffractive.

The alternative approach is to do the estimation through extrapolation of the efficiency and acceptance corrected inverse diffractive mass distribution multiplied by the square of the bin centers; this corresponds to the distribution ($1/M_X^p$, dN/dM_X^2), where the power p is of the order of $2 - 3$ (see figure 5.5 for the distribution with PYTHIA generated events showing the squared inverse diffractive mass distribution). For this distribution, the same event selection criteria are imposed as for the classification (explained in section 4.2). The distribution shown is the result of unfolding of the reconstructed one hemisphere diffractive mass distribution (this distribution contains both single and double diffractive events weighted by their approximate relative cross sections) using the iterative Bayesian approach presented earlier. The experimental error on the inverse squared diffractive masses is obtained from the error on the diffractive mass distribution, which is given by equation 5.38 (ΔM_X being the error of the diffractive mass).

$$\frac{\Delta M_X}{M_X} = 1 \tag{5.38}$$

$$M_X = \Delta M_X \tag{5.39}$$

$$\frac{d}{dM_X}(\frac{1}{M_X^2}) = -\frac{2}{M_X^3} \tag{5.40}$$

$$\Delta(\frac{1}{M_X^2}) = \frac{2\Delta M_X}{M_X^3} = \frac{2}{M_X^2} \tag{5.41}$$

A linear fit is made on the unfolded distribution in the region of acceptance (the region used is M_X^{-2} = [0.003, 0.04] GeV^{-2}, which corresponds to the region $M_X =$ $[5, 18.3]$GeV). This linear fit is extrapolated to higher values of $1/M_X^p$ and a histogram with the same binning as for the simulated diffractive mass distribution is filled with values given by this extrapolated function. This histogram is divided by the bin centers squared to obtain the extrapolated $1/M_X^p$ distribution. This distribution is integrated from $M_X < 10$ GeV (this is roughly the mass below which efficiency is degrading) up to the maximum simulated inverse squared diffractive mass value. This gives the extrapolated number of events outside acceptance. The unfolded distribution is used to determine the number of events with masses above 10 GeV and the sum of these is the estimated total number of one hemisphere events. The extrapolation is valid if the differential cross section $d\sigma/dM_X^2$ is proportional to the chosen power p of the inverse diffractive mass $1/M_X$. A reasonable choice for the parameter for events generated with PYTHIA is $p = 2$ and for PHOJET $p = 3$.

$$\frac{d\sigma}{dM_X^2} \propto \frac{1}{M_X^p} \tag{5.42}$$

When this assumption is true or approximately true, the estimated number of one hemisphere events can be used. In this context the set of one hemisphere events is extended to contain all events that do not have reconstructed particle tracks in both hemispheres. Most of the missed events are expected to originate from single diffractive events (see table 2). The estimated fraction of one hemisphere events (R is the ratio of the estimated number of one hemisphere events and the total number of simulated events) using events generated with PYTHIA is $R = 0.54 \pm 0.01$ (the corresponding value for PHOJET is 0.67 ± 0.01, but the distribution is not as linear (see figure A.20), which makes the extrapolation more uncertain).

The total simulated inelastic cross section used for this study is 70 mb. The estimate of the total reconstructed inelastic cross section is calculated using the following formula:

$$\sigma_{inelastic} = \sum_i Q_i \sigma_i + R(\sigma_{SD} + \sigma_{DD}),$$

where the individual cross sections are summed with weights Q_i (the number of two hemisphere events divided by the total number of events) to obtain the two hemisphere event cross section and to which the one hemisphere cross section, obtained using the calculated ratio R, is added. Using the sample generated with PYTHIA the following is obtained.

$$\sigma_{inelastic} = 0.32 \pm 0.12 \cdot 12.5\,\text{mb} + 0.994 \pm 0.009 \cdot 50\,\text{mb} + 0.60 \pm 0.06 \cdot 7.5\,\text{mb}$$

Process	2-h Q_i [%]	1-h W_i [%]	0-h [%]	Cross section σ_i [mb]
SD	32_{-12}	43^{+9}	24.6^{+3}	12.5
MB	$99.4_{-0.9}$	$0.5^{+0.9}$	0.032 ± 0	50
DD	60^{+6}	$30.5_{-4.9}$	$9.5_{-1.1}$	7.5

Tab. 3: The fractions of one and two hemisphere events as well as events with no reconstructed particles in T1 and T2 for each type of process studied plus the assumed cross sections. Systematic errors on fractions obtained from comparison with events generated for these processes with PHOJET.

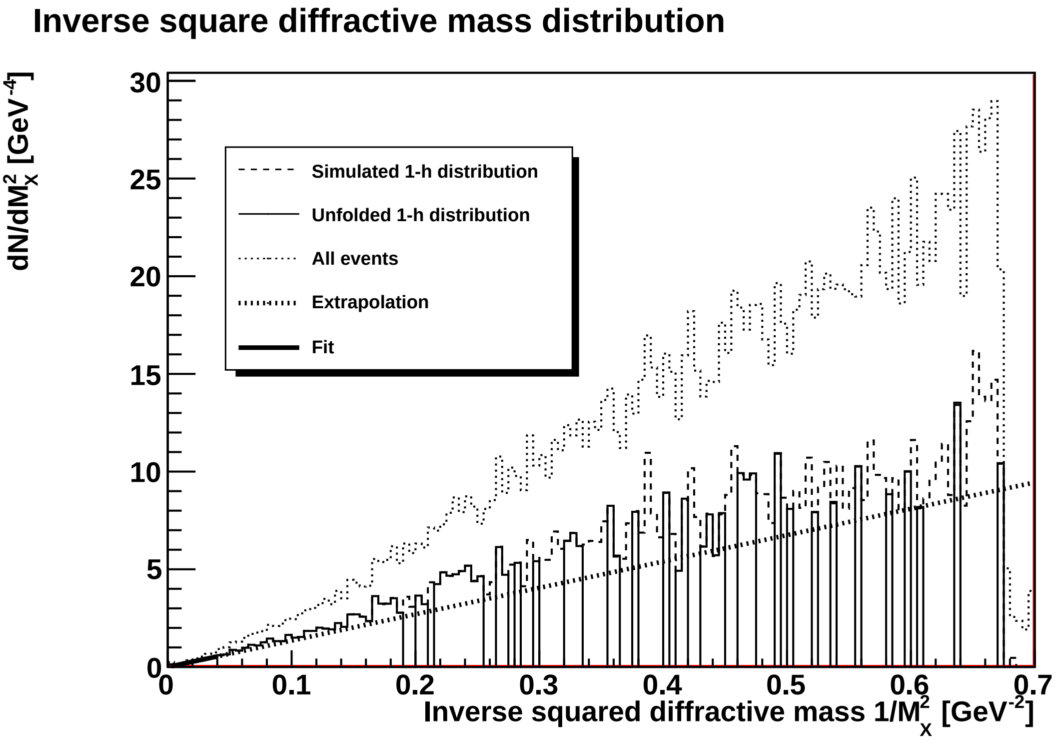

Fig. 5.5: The unfolded inverse square diffractive mass distribution for one hemisphere events. The value of the parameter a of the $y = ax$ fit is 13.5 ± 0.7.

$$+0.54 \pm 0.01 \cdot (12.5\,\mathrm{mb} + 7.5\,\mathrm{mb}) = 69 \pm 2\,\mathrm{mb}$$

This result shows that the estimated cross section is $99 \pm 3\%$ of the simulated cross section. In this case the extrapolation is usable, because the distribution generated by PYTHIA is nearly linear. Applying the same method (but using the inverse cubed diffractive mass distribution) on events generated with PHOJET, a cross section $\sigma_{inelastic} = 70 \pm 2$ mb is observed. The largest uncertainty comes from the single diffractive events with no or bad reconstructed particle tracks. A small fraction of double diffractive events have no charged particles within the acceptance of T1 and T2 and no final state leading protons (the fraction obtained from simulations done with PYTHIA is 1.8% and with PHOJET 0.9%).

6 Conclusions

Measurement of the total proton-proton cross section depends on many variables. Clusterization of hits is quite straight-forward and solid in case the data from the detector is clean. The track reconstruction algorithm usually has several data points for use in the reconstruction, which makes it reasonably reliable, even under the assumption that the data from previous stages of reconstruction is not ideal.

Event classification depends heavily on the previous stages of reconstruction, but also on the classifier itself. Training data sets for the classifier can differ from the data set, which is being classified. Any differences in the simulation chain versus the real interaction can cause differences in classification. In many events reliable classification is not possible due to similar signatures for certain events of different event types in the variables used.

Diffractive mass measurement depends on the classification to filter out most of the non-diffractive events, but in the case of one hemisphere events the non-diffractive background is negligible. Limited acceptance and efficiency require an unfolding method to be used to extract the true M_X distribution. To determine the number of undetected events a fit is done for the distribution within the region of good acceptance and the fitted function is extrapolated outside the region of acceptance. Both the unfolding and the extrapolation are sources of error.

The importance of T1 detectors for both event classification and diffractive mass measurement is clear. Much of the charged multiplicity of non-diffractive minimum bias events is in the acceptance of T1 and thus the T1 detectors allow the identification of non-diffractive events more precisely. Also the fraction of non-diffractive minimum bias events identified as diffractive increases substantially without complete T1 and thus a background in the diffractive mass distribution is introduced. Finally the reconstructed diffractive mass distribution is heavily distorted in the hemisphere without T1.

In the following the results and errors are examined more closely for each of the steps studied.

6.1 Pad Hit Clusterization

Both clusterization algorithms studied work well with data that is free of noise and cross-talk. If there are significant levels of noise present, fixed window algorithms give much higher pad cluster multiplicities than nearest neighbour algorithms. On the other hand, the sizes of clusters are smaller with fixed window algorithms. This is because fixed window algorithms artificially create additional clusters, while nearest neighbour algorithms allow easier identification of noise and cross-talk.

Numbers of clusters and cluster sizes affect the reconstruction of tracks. Differently reconstructed tracks can in turn lead to different multiplicities and thus lead to different classification.

6.2 Event Classification

Event classification works well if the signatures of different event types are really different. It turns out non-diffractive minimum bias, single diffractive and double diffractive events can, however, show many very different signatures and there is some overlap. A majority of non-diffractive events have large multiplicities and the classification variables for these events differs from those for diffractive events with on average smaller multiplicities.

Fisher Linear Discriminant Analysis works almost as well as the Artificial Neural Networks. This suggests that the problem can be well approximated with a linear approach. The ROOT TMVA toolkit allows for the use of other methods as well. MLP_ANN is chosen to be used because it is more flexible and once training is done the classification step is nearly as fast as with less complex methods.

The event classification gives good results in distinguishing between diffractive and non-diffractive two hemisphere events, but does not work as well for distinguishing between single and double diffractive events. Classification of one hemisphere events is not possible, which can be understood from simple physics arguments.

6.3 Estimation of Number of Inelastic Events

The diffractive mass extrapolation is prone to many sources of error. The major sources of error are: limited acceptance and efficiency of detectors, unfolding and extrapolation. Detector efficiency can be corrected for if the detector efficiencies for different pseudorapidites are known. Limited acceptance is more difficult, because there is no data from regions outside of acceptance. Linearity of the distribution can be checked in the region of acceptance and the power of inverse diffractive mass used can be chosen accordingly. Extrapolation is always prone to problems if the distribution is not as expected in the region where the extrapolation is done. This can lead to results which are far off from the true value.

Major issues for diffractive mass extrapolation are also the assumptions applied when performing the extrapolation. In this case equation 5.42 is assumed to hold in the region estimated using extrapolation, but this is not necessarily the case and the inverse power of M_X needs to be chosen correctly. In the case of PYTHIA generated events, the $1/M_X^2$ approximation is usable, but for PHOJET a better fit can be done in the $1/M_X^3$ distribution. The extrapolation of the diffractive mass distribution gives an estimate that by using this technique for correcting the diffractive rates, a cross section of 69±2mb out of the simulated 70mb is measured assuming the PYTHIA description of diffraction. In a real measurement the estimate obtained using extrapolation to low diffractive mass needs to be compared to the estimate based on proton only events.

References

[1] The TOTEM Collaboration, The TOTEM experiment at the CERN Large Hadron Collider, Journal of Instrumentation 3 (2008) S08007.

[2] F. Halzen, A. Martin, Quarks & Leptons: An Introductory Course in Modern Particle Physics, John Wiley & Sons, 1984.

[3] G. Källén, Elementary Particle Physics, Addison-Wesley Publishing Company, Inc., 1964.

[4] J. R. Cudell, et al., Benchmarks for the forward observables at RHIC, the Tevatron-Run II, and the LHC, Phys. Rev. Letters 89 (2002) 201801.

[5] G. Antchev, et al., Diffraction at TOTEM, arXiv:0812.3338v1 [hep-ex] 17 Dec 2008 (12 2008).

[6] The TOTEM Collaboration, TOTEM Technical Design Report, CERN, 2004, CERN-LHCC-2004-002.

[7] C. Altunbas, et al., Construction, test and commissioning of the triple-GEM tracking detector for COMPASS, Nucl. Instrum. Meth. A 490 (2002) 177.

[8] B. Ketzer, et al., Performance of triple GEM tracking detectors in the COMPASS experiment, Nucl. Instrum. Meth. A 535 (2004) 314.

[9] M. Alfonsi, et al., High-rate particle triggering with triple-GEM detector, Nucl. Instrum. Meth. A518 (2004) 106.

[10] `https://twiki.cern.ch/twiki/bin/view/CMS/WorkBook`.

[11] `http://root.cern.ch/drupal/`.

[12] T. Sjöstrand, S. Mrenna, P. Skands, Pythia 6.4 physics and manual, Journal of High Energy Physics 2006 (05) (2006) 026.
URL `http://stacks.iop.org/1126-6708/2006/i=05/a=026`

[13] `http://home.thep.lu.se/~torbjorn/Pythia.html`.

[14] J. Guillaud, A.Sobol, Simulation of diffractive and non-diffractive processes at the LHC energy with the PYTHIA and PHOJET MC event generators., lAPP-EXP 2004-06 (7 2004).

[15] S. Agostinelli, et al., Geant4 - a simulation toolkit, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 506 (2003) 250.

[16] `http://geant4.cern.ch/`.

[17] The ROOT team, The ROOT users guide.
URL `http://root.cern.ch/drupal/content/users-guide`

[18] Elmer, et al., Experience with the CMS event data model, Journal of Physics: Conference Series 219 (3) (2010) 032022.
URL `http://stacks.iop.org/1742-6596/219/i=3/a=032022`

[19] V. Avati, et al., Offline software for the TOTEM experiment at the LHC (2010).
URL `http://villaolmo.mib.infn.it/ICATPP11th_2009/accepted/SW%20Applications/Avati.pdf`

[20] R. Engel, Photoproduction within the two-component Dual Parton Model: Amplitudes and cross sections, Zeitschrift für Physik C Particles and Fields 66 (1995) 203–214, 10.1007/BF01496594.
URL `http://dx.doi.org/10.1007/BF01496594`

[21] K. Hagiwara, et al., Review of particle physics, Physical Review D 66:1.

[22] M. Ryskin, A. Martin, V. Khoze, Soft diffraction at the LHC: a partonic interpretation, The European Physical Journal C - Particles and Fields 54 (2008) 199–217, 10.1140/epjc/s10052-007-0514-1.
URL `http://dx.doi.org/10.1140/epjc/s10052-007-0514-1`

[23] Gotsman, et al., A QCD motivated model for soft interactions at high energies, Eur.Phys.J. C57 (2008) 689–709. `doi:10.1140/epjc/s10052-008-0704-5`.

[24] R. A. Fisher, The use of multiple measurements in taxonomic problems, Annuals of Eugenics 7 (1936) 179–188.

[25] A. Hoecker, et al., TMVA 4 Toolkit for Multivariate Data Analysis with ROOT users guide, CERN-OPEN 007.

[26] M. W. Gardner, S. R. Dorling, Artificial neural networks (the multilayer perceptron)–a review of applications in the atmospheric sciences, Atmospheric Environment 32 (14-15) (1998) 2627 – 2636. `doi:10.1016/S1352-2310(97)00447-0`.

[27] J. Armitage, et al., Diffraction dissociation in proton-proton collisions at ISR energies, Nuclear Physics B194 (1982) 365–372.

[28] K. Nakamura, et al., Review of particle physics, J. Phys. G G 37 (2010) 075021.

[29] G. Cowan, Statistical Data Analysis, Oxford University Press, 1998.

[30] G. D'Agostini, A multidimensional unfolding method based on Bayes' theorem, Nucl. Instrum. Meth. A 362 (1995) 487–498.

[31] C. Shannon, A mathematical theory of communication, Bell Sys. Tech. J. 27 (1948) 379, 623.

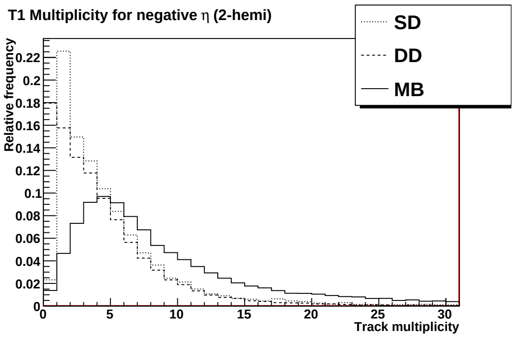

Fig. A.1: Multiplicities in T1 in negative hemisphere for events with tracks in both hemispheres.

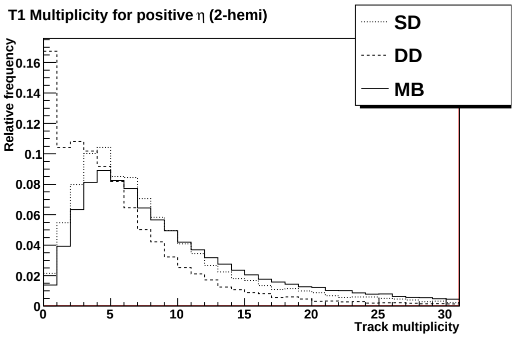

Fig. A.2: Multiplicities in T1 in positive hemisphere for events with tracks in both hemispheres.

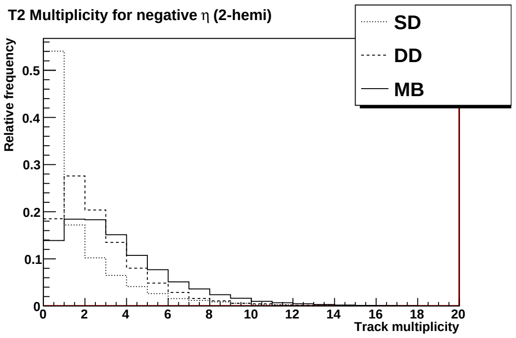

Fig. A.3: Multiplicities in T2 in negative hemisphere for events with tracks in both hemispheres.

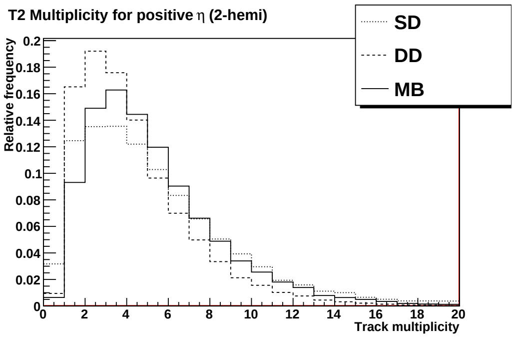

Fig. A.4: Multiplicities in T2 in positive hemisphere for events with tracks in both hemispheres.

A Plots from Events Generated with PHOJET

Most of the results from events generated with PHOJET are compatible with the results from PYTHIA. Based on the classifier response A.13 it, however, looks like PHOJET has a diffractive component in the minimum bias sample, which is not the case with PYTHIA. Also the diffractive mass distributions in figures A.19 and A.20 are different from the ones generated by PYTHIA. The extrapolation technique for estimating the number of non-detected diffractive events works with PHOJET when using the $1/M_X^3$ distribution.

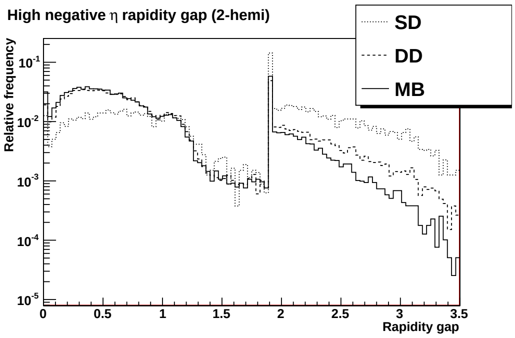

Fig. A.5: High η rapidity gaps in negative hemisphere for events with tracks in both hemispheres. Due to event flipping all events with non-trivial high η rapidity gaps in the negative hemisphere are 2-hemisphere events.

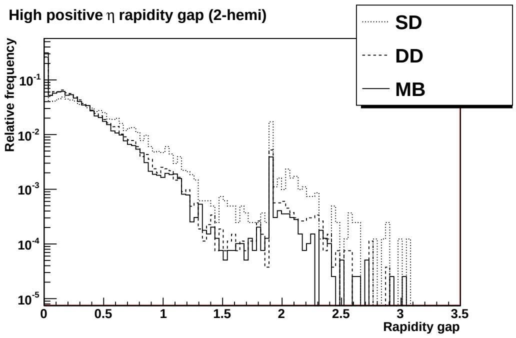

Fig. A.6: High η rapidity gaps in positive hemisphere for events with tracks in both hemispheres. This variable shows very little power for separation.

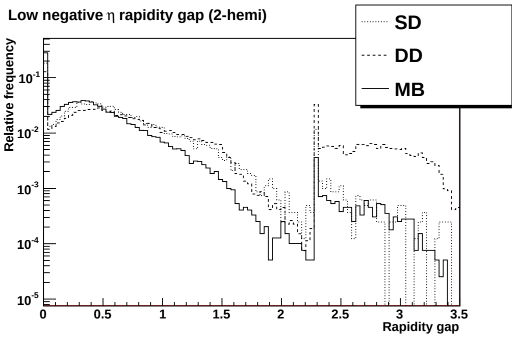

Fig. A.7: Low η rapidity gaps in negative hemisphere for events with tracks in both hemispheres.

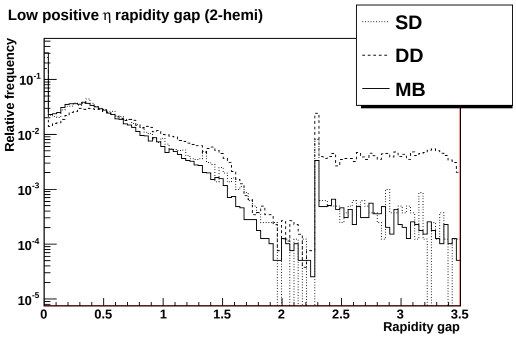

Fig. A.8: Low η rapidity gaps in positive hemisphere for events with tracks in both hemispheres.

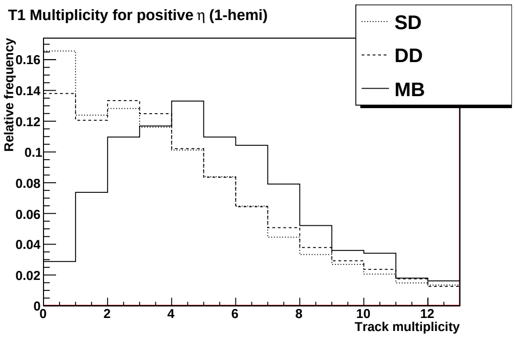

Fig. A.9: Multiplicities in T1 in negative hemisphere for events with tracks in only one hemisphere.

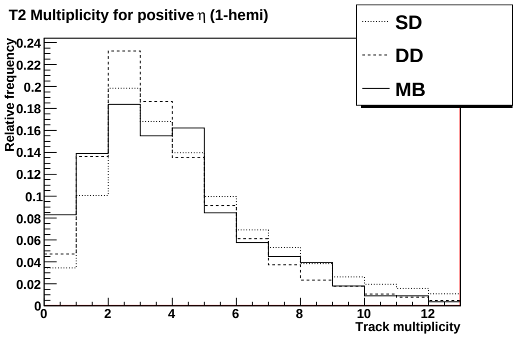

Fig. A.10: Multiplicities in T2 in positive hemisphere for events with tracks in only one hemisphere.

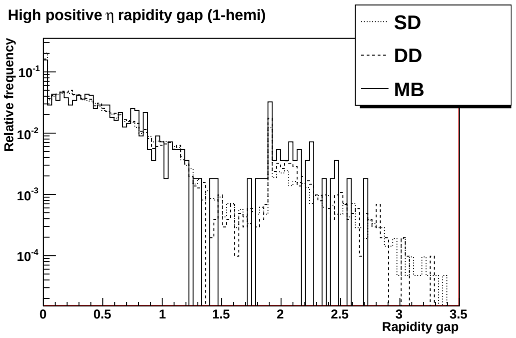

Fig. A.11: High η rapidity gaps for events with tracks in only one hemisphere.

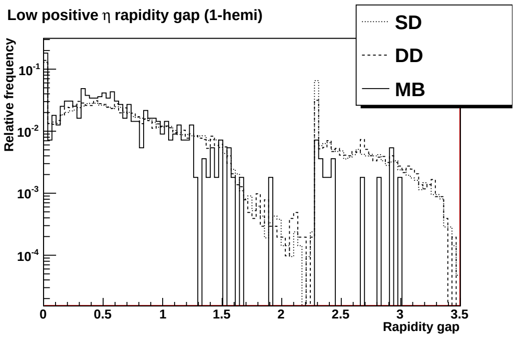

Fig. A.12: Low η rapidity gaps for events with tracks in only one hemisphere.

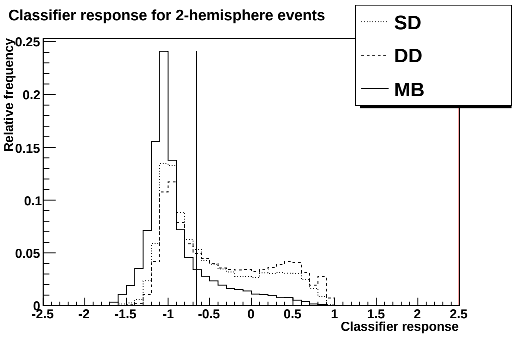

Fig. A.13: Classifier output for two hemisphere events being classified into diffractive and non-diffractive events. The black vertical line shows the threshold for classification; values under this threshold are considered non-diffractive and values above it diffractive.

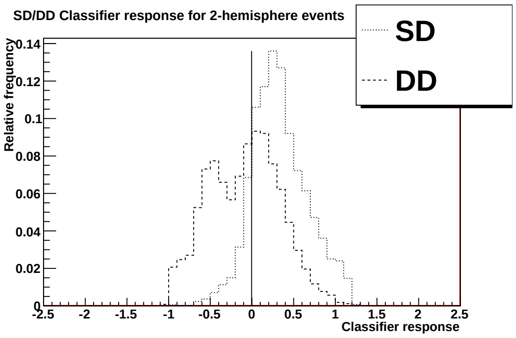

Fig. A.14: Classifier output for two hemisphere events being classified into single and double diffractive. These events have already in the previous step been classified as diffractive. Based on the classifier response it is not possible to make a good distinction between single and double diffractive events. Also non-diffractive minimum bias events that have passed through the earlier classification phase as diffractive give a similar response. The black vertical line shows the threshold for classification; values under this threshold are considered double diffractive and values above it single diffractive.

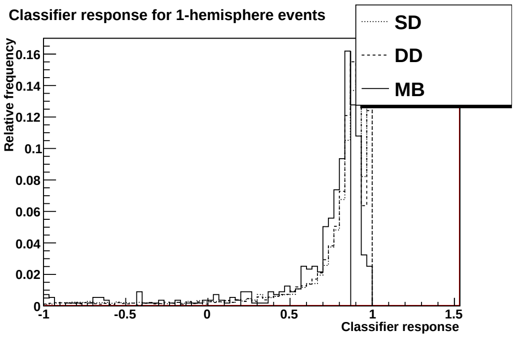

Fig. A.15: Classifier output for one hemisphere events being classified into diffractive and non-diffractive events. As can be seen the classifier cannot do much with the limited data available for one hemisphere events. The black vertical line shows the threshold for classification; values under this threshold are considered non-diffractive and values above it diffractive.

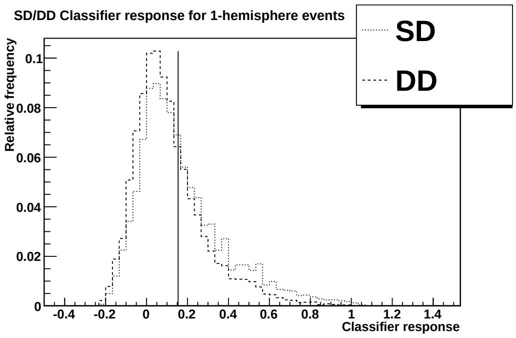

Fig. A.16: Classifier output for one hemisphere events being classified into single and double diffractive. These events have already been classified as diffractive in the previous step. The black vertical line shows the threshold for classification; values under this threshold are considered double diffractive and values above it single diffractive.

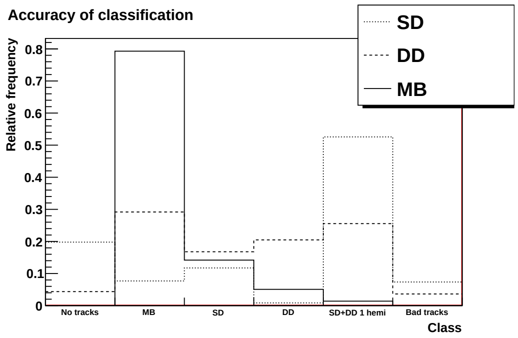

Fig. A.17: Classification results for events with tracks in two hemispheres. The bins represent the classes which events have been classified into and the different types of lines are for the data sets of events from different classes (continuous line for non-diffractive minimum bias, coarsely dotted line for double diffractive and finely dotted line for single diffractive). Also the fraction of events with reconstructed particles in only one hemisphere (labelled SD+DD 1-hemi) is shown.

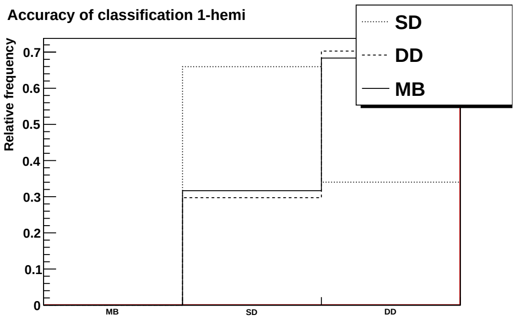

Fig. A.18: Classification results for events with tracks in only one hemisphere. The bins represent the classes which events have been classified into and the different types of lines are for the data sets of events from different classes (continuous line for non-diffractive minimum bias, coarsely dotted line for double diffractive and finely dotted line for single diffractive).

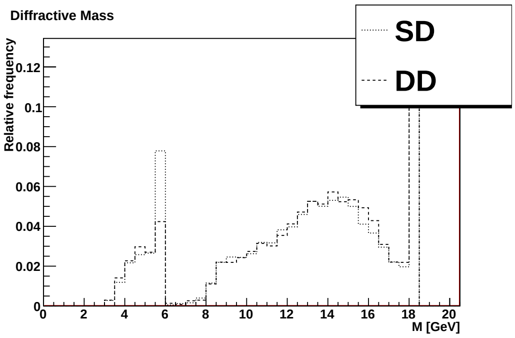

Fig. A.19: Reconstructed diffractive mass for events with reconstructed particles in one hemisphere and classified as diffractive.

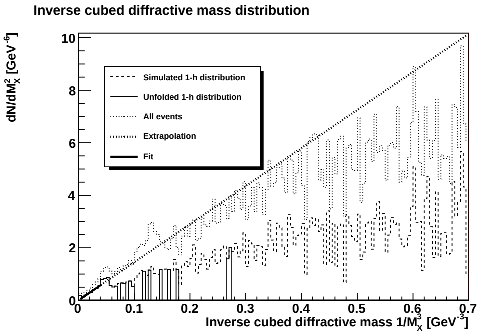

Fig. A.20: The unfolded inverse cubed diffractive mass distribution for one hemisphere events (the fit is better in the inverse cubed distribution with PHOJET). The value of the parameter a of the $y = ax$ fit is 14.5 ± 0.8.

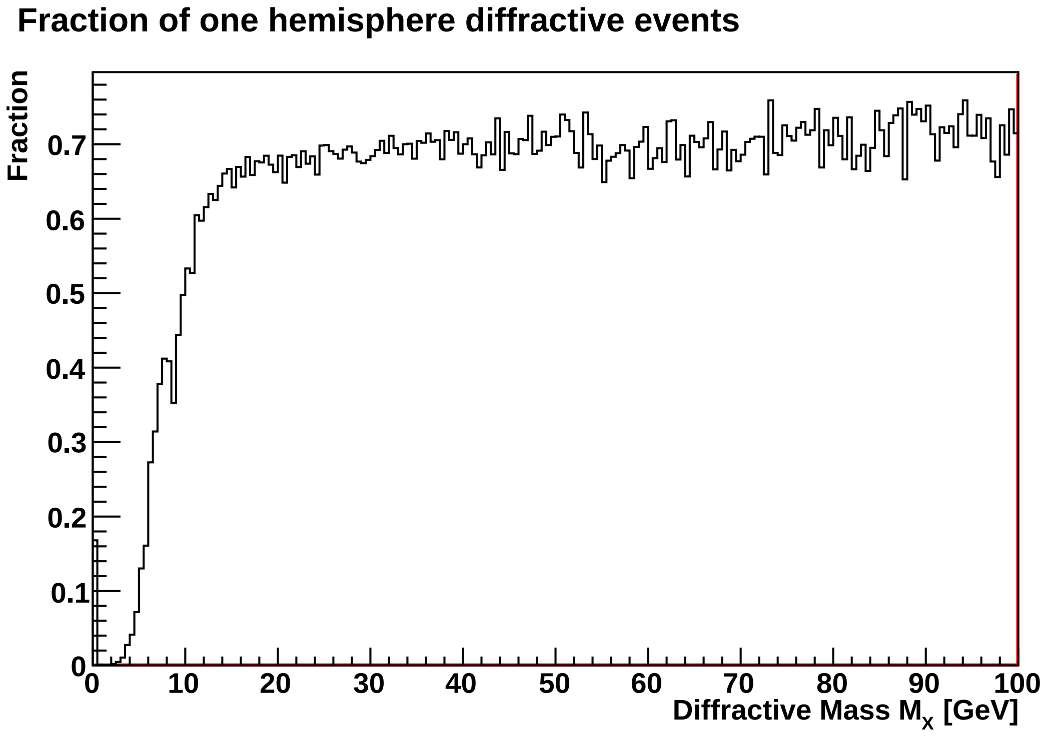

Fig. A.21: Fraction of one hemisphere diffractive events as function of diffractive mass (average weighted distribution for single and double diffractive events).