Two Component General Multiplicity Distribution (GMD) At Current LHC Energies

MUHAMMAD JAMIL BIN AGUS RIZAL

A0074128H

National University of Singapore (NUS)

Supervisor

Assoc. Prof. CHAN Aik Hui, Phil

Co-Supervisor

Prof. OH Choo Hiap

A dissertation submitted to the National University of Singapore in partial fulfilment of PC4199: Honours Project in Physics for the degree of Bachelor of Science (Physics) Hons.

April, 2014

Abstract

As collider energy increases, the study of multiplicity distributions from high energy collisions has become even more interesting. On top of the very apparent "shoulder" structure first observed in the UA5 multiplicity data, the multiplicity data from LHC is now showing a sharper peak structure at low multiplicities. In this work, we first introduce the General Multiplicity Distribution (GMD) formalism before extending it to a two component GMD model. We adapt the interpretation that these two component refer to two different classes of events, one of which is known as the soft component, and the other, the semihard component. The soft component is understood as events without minijets while the semihard components as events with minijets. We then fit the multiplicity data from the Compact Muon Solenoid (CMS) at CERN using our two component GMD model. We show that this generally gives excellent results. By calculating the C_q moments, our two component GMD manages to show KNO Scaling approximately obeyed at low pseudorapidity cuts η_c but violated at higher η_c. Comparing with other available models that also perfomed fits on the same set of data, we found that our fitting gives χ^2/dof values lower than the other models such as the two component NBD, IPPI and QGSM indicating a better fit. However, we also found that our χ^2/dof values are higher when compared to the three component NBD model. We end this work by

attempting to predict the parameters that will give the best fit curve if the two component GMD model is employed to describe the awaited future multiplicity distribution data at 14 TeV from the LHC.

Acknowledgements

بِسْمِ ٱللَّهِ ٱلرَّحْمَٰنِ ٱلرَّحِيمِ

In the Name of Allah, the Most Beneficent, the Most Merciful.

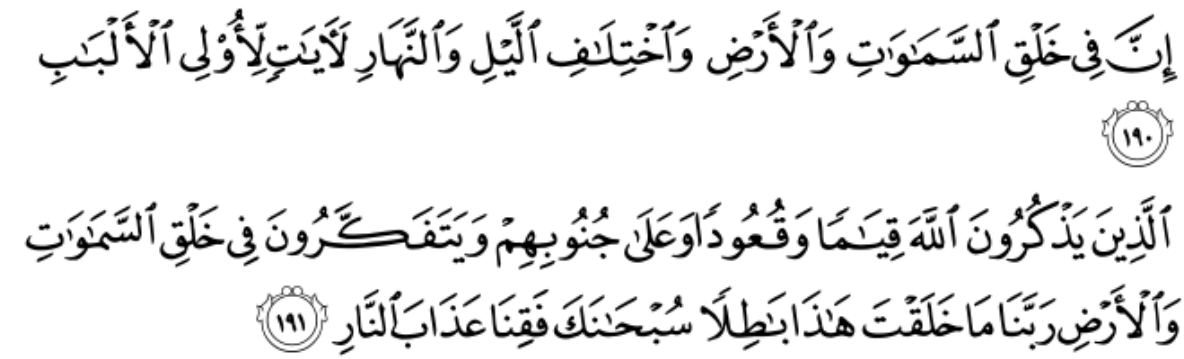
Praise be to Allah, Cherisher and Sustainer of the Universe. To Him belongs all the knowledge of the Physical Laws of the Universe. By his Grace and Guidance, I was able to complete my Final Year Project.

I would like to thank first and foremost, my supervisor, Assoc. Prof. Phil Chan for taking me under his wing for my final year project. He has been very supportive of my ideas and allowed me to explore the field in my own way. Under his guidance, I learned a lot about research in Physics. I would also like to thank my co-supervisor Prof Oh for his input and kind words towards my work.

My Physics journey in NUS would not have been enlightening without the guidance of my Professors and Tutors. Although unfortunately I am unable to list all of their names here, to them I am deeply thankful of their teachings. To my closest friends in Physics: Aren, John, Jacq, Ruth, Melvin, Koon Tong, Liesel and many more, I thank you very much for your company throughout my stay in NUS. Your presence in my journey in Physics has made the experience more enjoyable and less of a torture.

Special thanks to Liyi, who is closest to me in my final year journey in NUS, for always believing in me, encouraging me and keeping me in check.

“


إِنَّ فِي خَلْقِ السَّمَاوَاتِ وَالْأَرْضِ وَاخْتِلَافِ اللَّيْلِ وَالنَّهَارِ لَآيَاتٍ لِّأُولِي الْأَلْبَابِ ﴿١٩٠﴾
الَّذِينَ يَذْكُرُونَ اللَّهَ قِيَامًا وَقُعُودًا وَعَلَىٰ جُنُوبِهِمْ وَيَتَفَكَّرُونَ فِي خَلْقِ السَّمَاوَاتِ
وَالْأَرْضِ رَبَّنَا مَا خَلَقْتَ هَٰذَا بَاطِلًا سُبْحَانَكَ فَقِنَا عَذَابَ النَّارِ ﴿١٩١﴾

Indeed, in the creation of the heavens and the earth and the alternation of the night and the day are signs for those of understanding.

Who remember Allah while standing or sitting or [lying] on their sides and give thought to the creation of the heavens and the earth, [saying], “Our Lord, You did not create this aimlessly; exalted are You [above such a thing]; then protect us from the punishment of the Fire.”

— Al-Quran: Ali Imran, 190-191

Contents

Chapter 1

Introduction

The number of charged particles (the charged-particle multiplicity), n, is a global measure of the final state of high energy collisions. The multiplicity distribution, i.e. the probability distribution of obtaining a definite number of produced particles, is one of the simplest observables in hadron collisions, but yet imposes important constraints on the mechanisms of particle productions. This has made study of multiplicity distribution very important in understanding the hadronization mechanism in regions inexplicable by pertubative Quantum Chromodynamics (QCD). Pertubative QCD fails to explain multiplicity distributions satisfactorily from its first principles due to the complexity of the final states and the fact that the running coupling constant becomes larger at low energies. Therefore, phenomenological models are developed to attempt to explain the distribution. In studying multiplicity distributions, kinematic properties are neglected. This is a drastic reduction of complex information contained in the final state of a particle collisions as only the number of produced charged particles is considered. Nevertheless, multiplicity distribution still contains information about particle correlations. Reviews on the subject can be found in [**?**, **?**].

If final-state particles are produced independently, then multiplicity distributions will have a Poisson distribution. Deviations from a Poisson distribution indicate correlations. Early measurements of multiplicity distributions in e^+e^- collisions at center-of-mass energy $\sqrt{s} = 29$ GeV from the TASSO Collaboration [**?**] showed that the distribution is approximately Poissonian. $pp(\overline{p})$ collisions on the other hand showed broader multiplicity distributions and only approximately Poissonian up to the maximum ISR energy of $\sqrt{s} = 62$ GeV [**?**]. At higher energies, multiplicity distribution deviate from the Poisson distribution.

As multiplicity distribution deviates from the Poisson distribution, a distribution model that is able to describe the experimental data is needed. Based on an algorithm developed by Konishi *et al* [**?**], Giovannini [**?**] extended his work by considering the QCD jets as Markov (Stochastic) branching processes that describe the evolution of multi-particle production. He pointed out that the Negative Binomial Distribution (NBD) is a solution to his Stochastic branching equation.

NBD has successfully described multiplicity distributions well up to $\sqrt{s} = 540$ GeV at the UA5 experiment [**?**]. Deviations from the NBD were discovered by UA5 at $\sqrt{s} = 900$ GeV [**?**] and later confirmed at the Tevatron at $\sqrt{s} = 1800$ GeV [**?**]. A shoulder structure starts to appear at $n \approx 2\,\overline{n}$ in which one NBD function was inadequate to describe it. This led to the a two-component NBD model [**?**] in which Giovanni and Ugoccioni [**?**] interpreted as a combination of a soft and a semihard component.

Chew *et al* introduced another solution to the Stochastic branching equation, known as the Generalized Multiplicity Distribution (GMD) [**?**] which is the main focus of this work. GMD has given excellent fit for the e^+e^- data from the TASSO Collaboration by Chan *et al* [**?**] and reasonably good fits for $p\overline{p}$ data at the ISR energies [**?**] and at UA5 experiments [**?**]. To explain the shoulder structure at higher energies, Dewanto *et al* [**?**] has used a two weighted component GMD functions to describe the UA5 data. In that

same spirit, this work will attempt to fit the charged multiplicity distribution data from the Compact Muon Solenoid (CMS) [**?**] using a two component GMD model.

At this point in time the framework of the weighted sum of different classes of events has been extended from the two component NBD to a three component NBD model [**?**] for explaining possible new physics at the LHC at $\sqrt{s} = 14$ TeV. The third component is hoped to be able to describe the expected new class of high multiplicity events which would be manifested by the expected appearance of a new "elbow" structure in the tail of multiplicity distribution of pp collisions at the highest planned $\sqrt{s}$ at the LHC. However, hitherto no such elbow structure has been observed at LHC. Zborovský [**?**] has attempted to fit the LHC multiplicity data using a three component NBD with quite some success, but the third component turned out not to describe an "elbow" structure, but described the peaky region at low multiplicities of data better.

It is worth noting here that there are various other models developed to try to describe multiplicity distribution. A review by Wroblewski [**?**] mentioned other types of distribution such as modified negative binomial distribution (MNBD) [**?**], Krasznovszky-Wagner (KW) distribution [**?**] and lognormal distribution [**?**]. With the increase of collider energies, more recent models are proposed which include the Independent Pair Parton Interactions (IPPI) [**?**] model, the multi-ladder exchange or Quark-Gluon String (QGSM) model [**?**] and also Color Glass Condensate (CGC) Model [**?**]. Each of these models have its own advantages and disadvantages over one another.

The following sections in this Chapter will give a short description of the CMS detector and a description of the multiplicity data from the CMS. In Chapter **??**, we will formally introduce the GMD, its derivation and properties. Also in the same chapter, we will then extend the GMD formulation to the two component GMD model. Chapter **??** will describe our data analysis methodology while the results of our findings will be presented and discussed in Chapter **??**. We will also compare our results with fits done by other

works done on the CMS multiplicity data [**?**, **?**, **?**]. We will attempt to make predictions at higher energies at the LHC in Chapter **??** before presenting our conclusions and future works in the last chapter.

1.1 The Compact Muon Solenoid (CMS) Detector

The Compact Muon Solenoid (CMS) is one of the detectors at the Large Hadron Collider (LHC) [**?**] at CERN. It is a multi-purpose detector built for an array of experiments. A complete description of the CMS detector can be found in [**?**].

The CMS experiment uses a right-handed coordinate system, with the origin at the nominal interaction point (IP), the x-axis pointing to the centre of the LHC ring, the y-axis pointing up, and the z-axis along the counter clockwise beam direction.

The CMS detector boasts a superconducting solenoid of 6 m internal diameter, providing an axial magnetic field with nominal strength of 3.8 T. Surrounding the beam pipes, immersed in the magnetic field are the pixel tracker, the silicon-strip tracker (SST), the lead tungstate electromagnetic calorimeter, the brass/scintillator hadron calorimeter and the muon detection system. In addition to the barrel and endcap calorimeters, the steel/quartz-fibre forward calorimeter (HF) covers the pseudorapidity[1] region $2.9 < \eta < 5.2$.

The Beam Scintillation Counters (BSC) and Beam Pick-up Timing for the eXperiments (BPTX) devices are two of the CMS subdetectors acting as LHC beam monitors. These two were used to trigger the detector readout. The BSCs are located along the beam line on each side of the IP at a distance of 10.86 m and are sensitive in the range $3.23 < \eta < 4.65$. The two BPTX devices are located inside the beam pipe at distances

[1] See Appendix **??** for a review on Rapidity and Pseudorapidity

of 175 m from the IP . They are designed to provide precise information on the bunch structure and timing of the incoming beams, with a time resolution better than 0.2 ns.

The tracking detector consists of 1440 silicon-pixel and 15148 silicon-strip detector modules. The barrel part consists of 3 layers of pixel modules and 10 layers of SST modules around the IP at distances ranging from 4.4 cm to 1.1 m. Five out of the 10 strip layers are double sided and provide additional z coordinate measurements. The pseudorapidity acceptance is increased to $\eta < 2.5$ by the two endcaps, which consist of 2 disks of pixel modules and 12 disks of SST modules. The tracker provides an impact parameter resolution of about 100 μm and a transverse momentum p_T resolution of about 0.7% for 1 GeV/c charged particles at normal incidence.

1.2 Multiplicity Data from CMS

The multiplicity data from CMS refers to the charged hadron multiplicity distribution for inelastic non-single-diffractive (NSD) interactions. It is based on an event selection that retains a large fraction of the non-diffractive (ND) and double-diffractive (DD) events, while disfavouring single-diffractive (SD) events.[2]

The CMS multiplicity distributions are measured in five increasing pseudorapidity cuts $|\eta| < \eta_c$ from $\eta_c = 0.5$ to $\eta_c = 2.4$. The experimental analysis was based on about 132 000, 12 000 and 442 000 events at 3 different center-of-mass energy $\sqrt{s}$ =, 900, 2360 and 7000 GeV respectively. Figure **??** shows the plots of the multiplicity data at the three different $\sqrt{s}$ at the five different η_c.

From Fig **??**, it could be seen that in the largest pseudorapidity cut of $\eta_c = 2.4$, there is a change of slope in the distribution for $n > 20$. This "shouldering" feature becomes more pronounced with increasing $\sqrt{s}$, most notably at 7000 GeV.

[2]See Appendix **??** for the different types of events in pp collisions

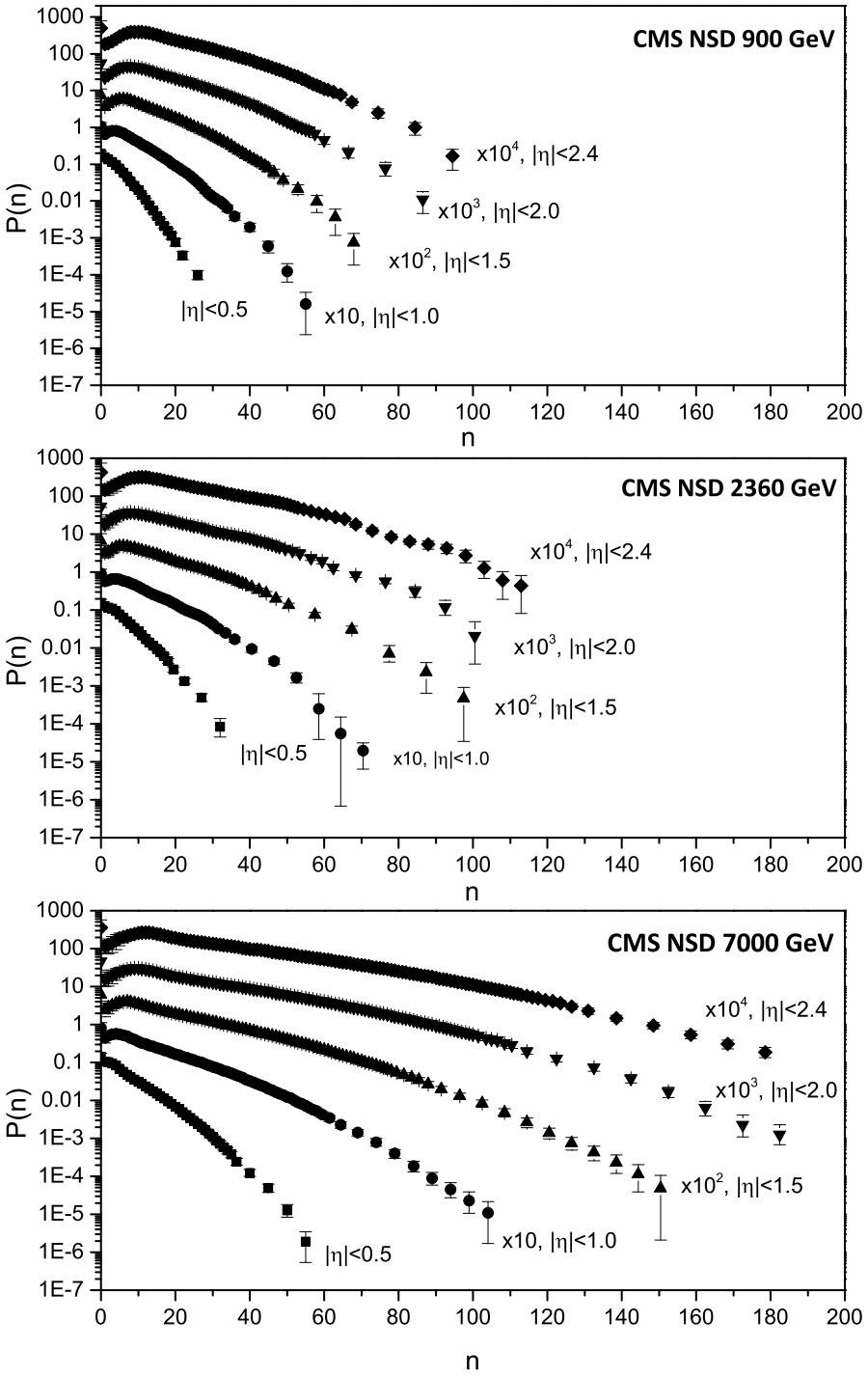


Figure 1.1: Multiplicity data from CMS at the 3 different $\sqrt{s}$, 900, 2360 and 7000 GeV and at the different pseudorapidity cuts η_c. The error bars include both the statistical and the systematic uncertainties. [**?**]

1.2.1 C_q Moments

Studying the moments of multiplicity distributions is convenient to describe their properties (e.g. as a function of $\sqrt{s}$). All moments together contain the information of the full distribution. However, in practice, only the first few moments can be calculated with reasonable uncertainties due to limited statistics. The reduced C_q moments are defined by:

$$C_q = \frac{\overline{n^q}}{\overline{n}^q} \tag{1.1}$$

where

$$\overline{n^q} = \sum_{n=0}^{\infty} n^q P_n \tag{1.2}$$

The total uncertainty in the C_q moments, E_q could be calculated as follows:

$$E_q^2 = \sum_{n=0}^{\infty} \left(\frac{\partial C_q}{\partial P(n)} \sigma_n \right)^2 \tag{1.3}$$

where

$$\frac{\partial C_q}{\partial P(n)} = \frac{(n^q - \overline{n^q})\overline{n} - \overline{n^q} q (n - \overline{n})}{\overline{n}^{q+1} \sum_m P(m)} \tag{1.4}$$

and σ_n is the uncertainty of the data point $P(n)$.

1.2.2 KNO Scaling and its Violation

KNO scaling was suggested in 1972 by Koba, Nielsen, and Olesen [**?**]. In the KNO scaling, it is suggested that the probability distribution $P(n)$ scales with

$$\overline{n}P_n = \Psi(z) \tag{1.5}$$

where $z = n/\overline{n}$. $\Psi(z)$ is a universal, i.e., energy-independent function. This means that multiplicity distributions at all energies fall on one curve when plotted as a function of z. The C_q moments,

$$C_q = \int_0^\infty z^q \Psi(z)\, \mathrm{d}z \tag{1.6}$$

defines $\Psi(z)$ uniquely. Substituting $z = n/\overline{n}$ results in definition of C_q moments as in Eq. (**??**). If KNO scaling holds, C_q moments will be independent of energy.

At full space space (FPS), violation of KNO Scaling has been observed from the multiplicity data in the UA5 experiment [**?**] and confirmed by the CMS data [**?**] since it is demonstrated that the C_q moments are quite obviously increasing with energy. From the CMS data, a strong violation of KNO scaling is observed between $\sqrt{s} = 900$ to 7000 GeV at the largest pseudorapidity cut $\eta_c = 2.4$. KNO scaling holds approximately at small pseudorapidity cuts $\eta_c = 0.5$. Foor this work, we have decided to present our results without KNO scaling.

Chapter 2

General Multiplicity Distribution (GMD)

2.1 Formalism and Derivation of GMD

Giovannini [**?**] showed that the total multiplicity distribution of partons inside a jet calculus can be written in the following equation

$$\begin{aligned}\frac{dP_{n,m}}{dt} =& -(An + \tilde{A}m + Bn)P_{n,m} + A(n-1)P_{n-1,m} \\ &+ \tilde{A}mP_{n-1,m} + B(n+1)P_{n+1,m-2}\end{aligned} \tag{2.1}$$

also known as the stochastic branching equation, where

$$t = \frac{6}{11N_c - 2N_f} \ln\left[\frac{\ln(Q^2/\mu^2)}{\ln(Q_0^2/\mu^2)}\right] \tag{2.2}$$

is the QCD evolution parameter, with Q is the initial parton invariant mass, Q_0 is the hadronization mass, μ is a QCD mass scale (in GeV), $N_c = 3$ (number of colors), and

$N_f = 4$ (number of flavors). $P_{n,m}$ is the probability distribution of n gluons and m quarks at QCD evolution, with A, $\tilde{A}$ and B refer to the average probabilities of the branching processes $g \to gg$, $q \to qg$ and $g \to q\overline{q}$ respectively.

For a fixed m and number of n gluons at t, we can rewrite Eq. (**??**) to become

$$\begin{aligned}\frac{dP_n}{dt} &= -(An + \tilde{A}m + Bn)P_n + A(n-1)P_{n-1} \\ &\quad + \tilde{A}mP_{n-1} + B(n+1)P_{n+1}\end{aligned} \tag{2.3}$$

where GMD is a solution to it. To solve Eq. (**??**) analytically, denote a probability generating function to be

$$f(t,s) = \sum_{n=0}^{\infty} P_n s^n \tag{2.4}$$

Next we consider

$$\frac{\partial f}{\partial t} = \sum_{n=0}^{\infty} \frac{dP_n}{dt} s^n \tag{2.5}$$

It can be shown that[1]

$$\frac{\partial f}{\partial t} = (1-s)(B - As)\frac{\partial f}{\partial s} - \tilde{A}m(1-s)f \tag{2.6}$$

To solve this first order partial differential equation, we introduce the subsidiary equation

$$dt = \frac{ds}{(1-s)(As - B)} = \frac{df}{\tilde{A}m(1-s)f} \tag{2.7}$$

[1] See Appendix **??**

Thus solving the first and second term in Eq. (**??**) will result in

$$e^{t(A-B)}\left(\frac{1-s}{As-B}\right) = \text{constant} \tag{2.8}$$

while solving the second and third term in Eq. (**??**) will give

$$\frac{\tilde{A}}{A}m\ln(As-B)+\ln f = \text{constant} \tag{2.9}$$

Thus we can write the following relationship in term of a function Ψ

$$\frac{\tilde{A}}{A}m\ln(As-B)+\ln f = \Psi\left[e^{t(A-B)}\left(\frac{1-s}{As-B}\right)\right] \tag{2.10}$$

Setting the initial conditions to $f(t=0,s)=s^{k'}$ (hence k' is the initial number of gluons in average sense), one obtains[1]

$$f = \left[A\left(\frac{1+XB}{1+XA}\right)-B\right]^{\frac{m\tilde{A}}{A}}[As-B]^{-\frac{m\tilde{A}}{A}}\left[\frac{1+XB}{1+XA}\right]^{k'} \tag{2.11}$$

where $X = e^{t(A-B)}\left(\frac{1-s}{As-B}\right)$. At this stage, we neglect B as in [**?**] (i.e. $B=0$) so that (**??**) will reduce to the generating function of GMD

$$f = \left[s+(1-s)e^{At}\right]^{-k}\left[1+\frac{1-s}{s}e^{At}\right]^{-k'} \tag{2.12}$$

where $k = \dfrac{m\tilde{A}}{A}$ is related to the initial number of quarks in average sense.

Finally, using $P_n = \dfrac{1}{n!}\dfrac{\partial^n f}{\partial s^n}|_{s=0}$, we get the solution to Eq. (**??**), namely the generalized multiplicity distribution (GMD) for $B=0$[1] [**?**].

$$P_{GMD}(n) = \frac{(n+k-1)!}{(n-k')!(k'+k-1)!}\left(\frac{\overline{n}-k'}{\overline{n}+k}\right)^{n-k'}\left(\frac{k+k'}{\overline{n}+k}\right)^{k+k'} \tag{2.13}$$

Or in Gamma functions[2], Eq. (**??**) could be written as

$$P_{GMD}(n) = \frac{\Gamma(n+k)}{\Gamma(n-k'+1)\Gamma(k'+k)}\left(\frac{\overline{n}-k'}{\overline{n}+k}\right)^{n-k'}\left(\frac{k+k'}{\overline{n}+k}\right)^{k+k'} \tag{2.14}$$

2.2 Properties of GMD

2.2.1 Negative Binomial Distribution (NBD)

The GMD can be reduced to the popular NBD when k' is set to 0. Here we shall briefly illustrate the relationship between these two distributions and also describe some of the properties of the NBD.

The NBD is defined as

$$P_{p,k}^{NBD}(n) = \binom{n+k-1}{n}(1-p)^n p^k \tag{2.15}$$

It gives the probability of n failures and $k-1$ successes in any order for the kth success in a Bernoulli experiment with a success probability p. The NBD is a Poisson distribution for $k \to \infty$ and a geometrical distribution for $k = 1$. For the negative integer k and $\overline{n} \leq -k$ the distribution is a binomial distribution where $-k$ is the number of trials and $-\overline{n}/k$ the success probability.

[2]See Appendix **??**

The binomial could be written in terms of the Gamma function

$$\binom{n+k-1}{n} = \frac{(n+k-1)!}{n!(k-1)!} = \frac{\Gamma(n+k)}{\Gamma(n+1)\Gamma(k)} \tag{2.16}$$

while the mean of the distribution $\overline{n}$ is related to p by $p^{-1} = 1 + \overline{n}/k$. This leads to

$$P_{NBD}(n) = \frac{\Gamma(n+k)}{\Gamma(n+1)\Gamma(k)} \left(\frac{\overline{n}}{\overline{n}+k}\right)^n \left(\frac{k}{\overline{n}+k}\right)^k \tag{2.17}$$

Which is exactly GMD when $k' = 0$.

2.2.2 Furry-Yule Distribution (FYD)

GMD also reduces to the FYD proposed by Hwa and Lam [**?**] when $k = 0$

$$P_{GMD}(n) = \frac{\Gamma(n)}{\Gamma(n-k'+1)\Gamma(k')} \left(\frac{\overline{n}-k'}{\overline{n}}\right)^{n-k'} \left(\frac{k'}{\overline{n}}\right)^{k'} \tag{2.18}$$

Using data from UA5 [**?**], Hwa and Lam showed that the parameter k' has a strong indication to be independent of energy. Chan and Chew has attempted to fit FYD along with GMD in [**?**] using the data from UA5 as well and demonstrated that fits using FYD are not all acceptable as compared to fits using GMD or NBD. NBD and FYD are after all, special cases of the GMD.

It is probably also worth noting that the width of NBD is

$$\gamma \equiv \left(\frac{D}{\overline{n}}\right)^2 = \frac{1}{k} + \frac{1}{\overline{n}} \tag{2.19}$$

where as of FYD is

$$\gamma' \equiv \left(\frac{D}{\overline{n}}\right)^2 = \frac{1}{k} - \frac{1}{\overline{n}} \tag{2.20}$$

In both Eq.(**??**) and (**??**), D is the dispersion

$$D = \left(\overline{n^2} - \overline{n}^2\right)^{1/2} \tag{2.21}$$

2.3 Two Component GMD: The Soft and Semihard Components

It has been mentioned that as detector energy increases, one NBD function seem to fail to describe the multiplicity data well. Multiplicity distributions measured by UA5 have been successfully fitted using a combination two NBD functions [**?**]. Giovannini and Ugoccioni did a systematic investigation and interpreted this as a combination of a soft component and a semihard component [**?**]. The soft component can be understood as events without minijets while the semihard component as events with minijets. Here, the definition from the UA1 collaboration is used; a minijet is a group of particles having a total transverse energy larger than 5 GeV.

Dewanto *et al* [**?**] were first to use a two component GMD model to fit multiplicity data. In this work, following this spirit, we will use the two component GMD model to fit the multiplicity data from the CMS. In this approach, the multiplicity distribution depends on seven parameters that may all be dependent on $\sqrt{s}$

$$\begin{aligned} P_{total} = {} & \alpha_{soft} P_{GMD}(\overline{n}_{soft}, k_{soft}, k'_{soft}) \\ & + (1 - \alpha_{soft}) P_{GMD}(\overline{n}_{semihard}, k_{semihard}, k'_{semihard}) \end{aligned} \tag{2.22}$$

It is important to note that this approach combines two classes of events, not two different particle-production mechanisms in the same event. Therefore, no interference terms have to be considered and the final distribution is the sum of the two independent distributions.

Chapter 3

Data Analysis Methodology

In this work, we treat all the seven parameters in the two component GMD as free parameters and seek for the best fit curve using the method described below. The results of our fitting are based on the computation and minimization of the Chi-squared χ^2 value[1]. The value of χ^2 was computed as

$$\chi^2 = \sum_{m=m_1}^{m_2} \frac{(P_m^{ex} - P_m^{2GMD})^2}{\sigma_m^2} \tag{3.1}$$

over the finite range $< m_1, m_2 >$ of the multiplicity bins. Here, P_m^{ex} refers to the CMS experimental data and σ_m is the experimental uncertainty of P_m^{ex} (see section below for treatment of uncertainties in this fitting) while P_m^{2GMD} refers to our fitted two component GMD curve. The total number of bins considered in experiment is denoted by m_2. The value of m_1 in our fitting is taken to be 1 as we neglect the first data point at $n = 0$. This is because of its large uncertainty and being very off from the general trend of the data curve especially at larger pseudorapidity cuts η_c.

[1]See Appendix **??**

We have also neglected the last two data points in the multiplicity distribution for $\eta_c = 2.0$ and 2.4 at 7000 GeV. This is due to the computational limit in numerically calculating the Gamma function[2].

The values of P_m^{2GMD} are then renormalized to fulfil the condition

$$\sum_{m=m_1}^{m_2} P_m^{2GMD} = \sum_{m=m_1}^{m_2} P_m^{ex} \tag{3.2}$$

For this work, since we are taking into consideration of the $n = 1$ data point, we have set the constraint $k'_{soft} \leq 1$ to be consistent with GMD (since we must have $n \leq k$). However, we have more play over $k'_{semihard}$ as for the semihard component, since $\overline{n}$ has larger values causing the semihard component to be shifted to the right and start at larger n.

Treatment of Uncertainties

The statistical and systematic uncertainty are given for each of the data points in the CMS multiplicity data. In this work, the total uncertainty σ_m of a data point is calculated from a quadratic sum of the statistical and systematic uncertainties. The uncertainties of the data points from CMS are not symmetrical; generally, the uncertainties in the positive and negative directions are different. In this work, we adopt a simplistic method to handle this by taking the larger of the two uncertainties to be used in the minimization of the χ^2 value.

[2]See Appendix **??**

Mean Multiplicity $\overline{n}_{tot}$ and C_q Moments

The mean multiplicity is to be calculated using Eq. (**??**) with q set to 1. However, due to the limited statistics of the experimental data, we have to truncate the sum and calculate the mean multiplicity as follows

$$\overline{n}_{tot} = \sum_{m=m_1}^{m_2} m P_m^{2GMD/ex} \tag{3.3}$$

where m_1, m_2 and P_m are as in Eq. (**??**).

Similarly C_q moments are to be calculated using Eq. (**??**) while the uncertainty in the C_q moments of the experimental data is to be calculated using Eq. (**??**). However, as the case of the mean multiplicity, the summation has to be truncated and we calculate our C_q moments using the equations below

$$\overline{n^q} = \sum_{m=m_1}^{m_2} m^q P_m \tag{3.4}$$

The total uncertainty in the C_q moments, E_q could be calculated as follows:

$$E_q^2 = \sum_{m=m_1}^{m_2} \left(\frac{\partial C_q}{\partial P_m} \sigma_m \right)^2 \tag{3.5}$$

where $\partial C_q / \partial P_m$ is as in Eq. (**??**)

The mean multiplicities $\overline{n}_{tot}$ and C_q moments of the fitted two component GMD curve and the experimental data are compared in the Results and Discussion in Chapter **??**.

Chapter 4

Results and Discussion

As described in the previous chapter, we attempted to fit the multiplicity data from CMS by treating the seven parameters of the two component GMD function as free parameters with no constraints. The best fit curve for the data is determined by minimizing the χ^2 value.

Tables **??**, **??** and **??** shows the seven parameters as a result of the fitting. Also quoted is the value of χ^2 and the number of degrees of freedom (dof). Figs. **??**, **??** and **??** provides a graphical summary of the fitting on the data sets at 900, 2360 and 7000 GeV.

900 GeV								
η_c	α_{soft}	k_{soft}	k'_{soft}	$\overline{n}_{soft}$	$k_{semihard}$	$k'_{semihard}$	$\overline{n}_{semihard}$	χ^2/dof
0.5	1.00	1.73	0.00	3.88	0.00	9.00	11.23	0.51/14
1.0	0.77	2.48	0.00	5.74	6.50	0.00	14.67	4.23/31
1.5	0.78	2.55	0.00	8.69	6.15	0.57	21.84	2.43/43
2.0	0.81	2.55	0.00	12.11	0.00	5.55	30.37	3.94/53
2.4	0.82	2.60	0.00	14.80	0.00	6.31	36.42	6.79/59

Table 4.1: GMD parameters at $\sqrt{s}$ =900 GeV and the χ^2/dof values at the five different pseudorapidity cuts η_c as a result of our fitting (quoted to two decimal places).

2360 GeV								
η_c	α_{soft}	k_{soft}	k'_{soft}	$\overline{n}_{soft}$	$k_{semihard}$	$k'_{semihard}$	$\overline{n}_{semihard}$	χ^2/dof
0.5	0.99	1.37	0.13	4.89	15.28	11.89	15.45	1.58/14
1.0	0.87	1.94	0.00	8.06	8.09	0.00	21.81	5.54/31
1.5	0.94	0.53	1.00	13.82	0.00	7.64	32.95	5.31/41
2.0	0.71	2.52	0.00	13.14	0.00	4.92	36.48	8.54/51
2.4	0.75	2.51	0.00	16.62	0.00	5.77	44.78	8.10/61

Table 4.2: GMD parameters at $\sqrt{s}$ =2360 GeV and the χ^2/dof values at the five different pseudorapidity cuts η_c as a result of our fitting (quoted to two decimal places).

7000 GeV								
η_c	α_{soft}	k_{soft}	k'_{soft}	$\overline{n}_{soft}$	$k_{semihard}$	$k'_{semihard}$	$\overline{n}_{semihard}$	χ^2/dof
0.5	0.76	0.32	1.00	4.96	5.17	0.00	13.40	2.33/32
1.0	0.75	0.47	1.00	9.13	4.81	0.12	25.55	2.41/61
1.5	0.74	0.59	1.00	13.20	3.98	0.86	37.50	6.65/86
2.0	0.63	1.07	0.79	15.12	2.86	1.13	44.84	8.52/104
2.4	0.49	2.40	0.00	15.07	3.10	0.09	46.70	9.82/116

Table 4.3: GMD parameters at $\sqrt{s}$ =7000 GeV and the χ^2/dof value at the five different pseudorapidity cuts η_c as a result of our fitting (quoted to two decimal places).

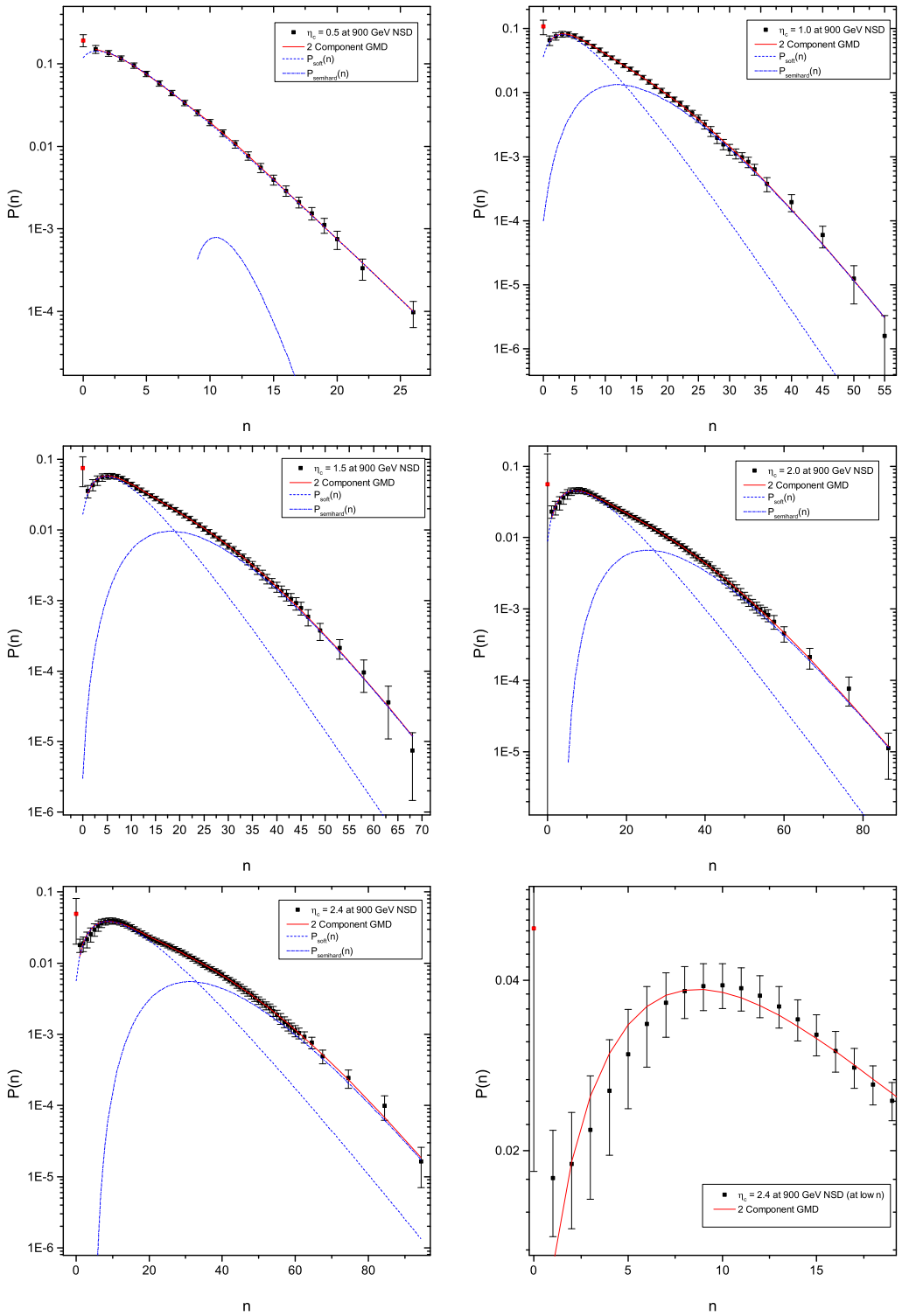


Figure 4.1: Multiplicity data at $\sqrt{s}$ =900 GeV fitted with two component GMD (red line) at all the different five pseudorapidity cuts η_c. The last graph shows the fitting at $n \leq 20$. The soft (blue dashed line) and semihard (blue dash-dotted lines) components are shown in blue lines. The error bars include both the statistical and the systematic uncertainties.

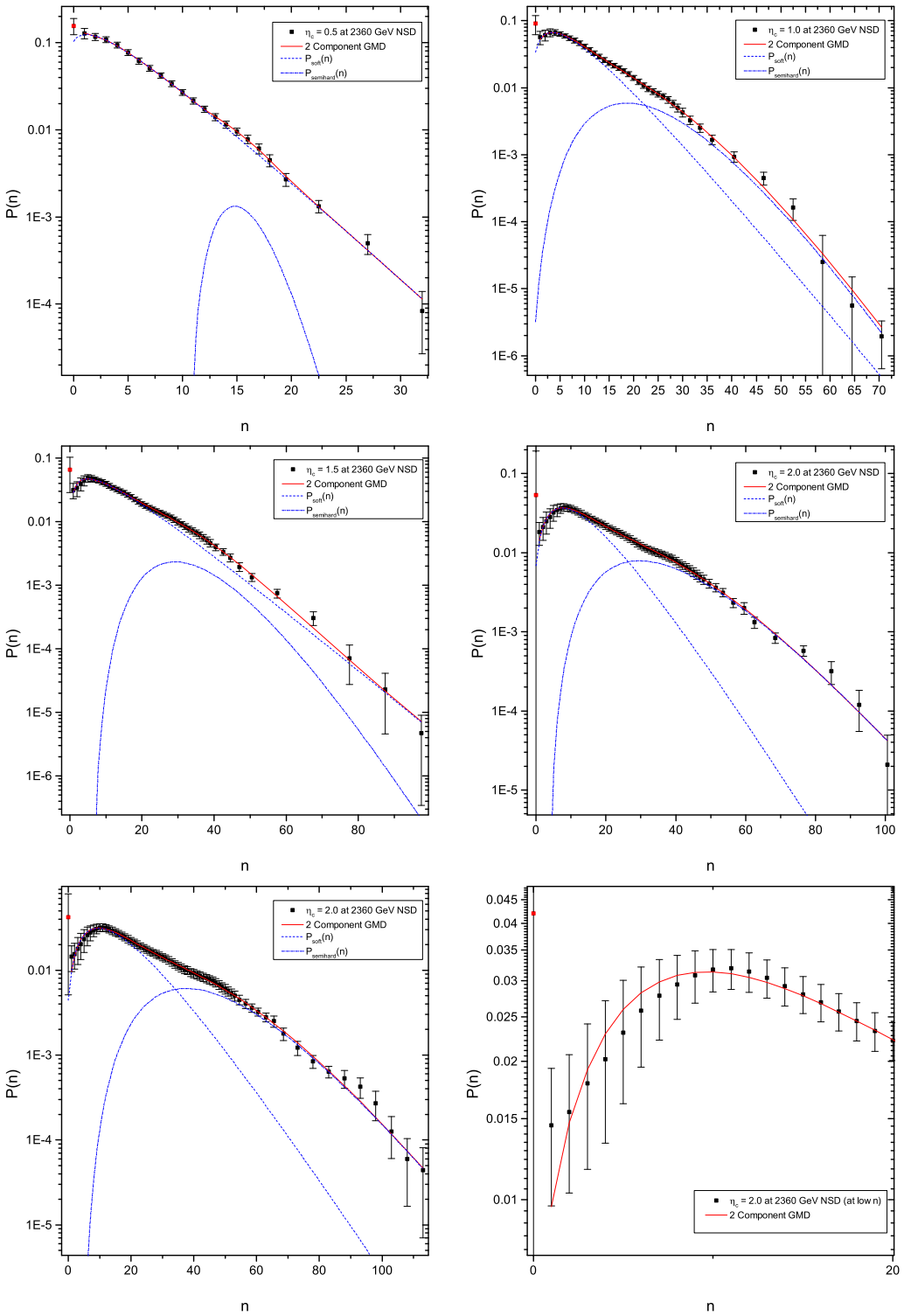


Figure 4.2: Multiplicity data at $\sqrt{s}$ =2360 GeV fitted with two component GMD (red line) at all the five different pseudorapidity cuts η_c. The last graph shows the fitting at $n \leq 20$. The soft (blue dashed line) and semihard (blue dash-dotted lines) components are shown in blue lines. The error bars include both the statistical and the systematic uncertainties.

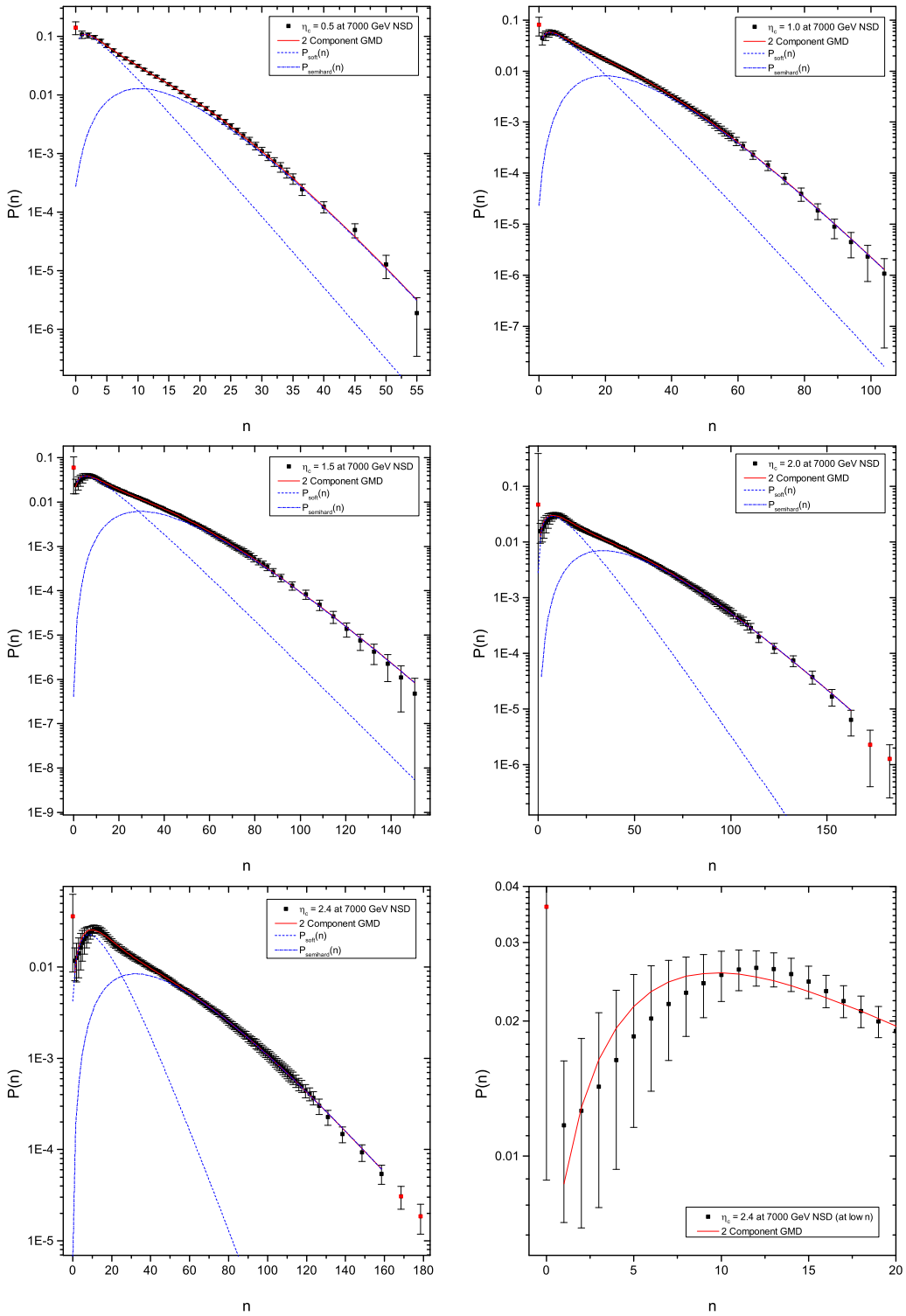


Figure 4.3: Multiplicity data at $\sqrt{s}$ =7000 GeV fitted with two component GMD (red line) at all the five different pseudorapidity cuts η_c. The last graph shows the fitting at $n \leq 20$. The soft (blue dashed line) and semihard (blue dash-dotted lines) components are shown in blue lines. The error bars include both the statistical and the systematic uncertainties.

Generally, the fits using the two component GMD model gives excellent results. The value of χ^2 for all the fitting are all satisfactorily low. However, at low n, the measured distribution is shifted towards higher multiplicities at all three $\sqrt{s}$ in comparison with the two component GMD fit. The fits are however, still within the uncertainties of the experimental values.

		Ghosh [**?**]		Zborovský [**?**]		Dremin and Nechitailo [**?**]	
$\sqrt{s}$ (GeV)	η_c	1NBD	2NBD	2NBD	3NBD	IPPI	QGSM
	0.5	6.12/22	-	2.0/20	0.8/16	-	-
	1.0	53.35/38	-	4.5/34	4.3/32	-	-
900	1.5	49.55/50	-	3.0/46	2.0/44	-	-
	2.0	36.69/60	-	5.2/56	2.4/54	-	-
	2.4	46.29/66	-	10.7/62	4.4/60	-	-
	0.5	6.38/21	-	-	-	-	-
	1.0	55.30/38	-	-	-	-	-
2360	1.5	24.79/48	-	-	-	-	-
	2.0	29.11/58	-	-	-	-	-
	2.4	29.76/68	-	-	-	-	-
	0.5	83.36/39	4.14/35	3.3/35	2.3/33	-	-
	1.0	152.65/68	6.97/64	3.8/64	0.9/62	-	-
7000	1.5	226.57/93	12.92/89	9.7/89	2.9/87	-	-
	2.0	208.56/113	14.93/109	14.2/109	3.8/106	-	-
	2.4	129.37/125	11.91/121	13.4/121	2.5/118	62/127	131/127

Table 4.4: Summary of χ^2/dof values reported by other papers that used different models.

Table **??** shows a table summary of χ^2/dof values reported by other papers which used various other models. Comparing χ^2/dof values of a single NBD model fit by Ghosh [**?**] on the CMS multiplicity data, it could be seen clearly that the values of χ^2/dof of our two component GMD is lower; indicating a better fit. In the same paper, Ghosh also made a two component NBD fit on the $\sqrt{s} = 7000$ GeV data. Although the values of the χ^2/dof is lower than that for the single NBD fit, our two component GMD fit

at $\sqrt{s} = 7000$ GeV still have lower values of χ^2/dof therefore more superior than a two component NBD fit. Zborovský in [**?**] reported χ^2/dof values of a two component NBD fit on the the CMS $\sqrt{s} = 900$ and 7000 GeV data. His values of χ^2/dof for the 7000 GeV data are lower than that quoted by Ghosh, but still not as low as our two component GMD fit. Our values of χ^2/dof for the 900 GeV data are also lower than the two component NBD fit quoted in Zborovský's article. However, Zborovský's main work in that paper of his is on a three component NBD fit on the multiplicity data from LHC. In comparison, Zborovský's three component NBD fit on the CMS data generally all has lower values compared to our two component GMD model (with a marginal exception of the χ^2/dof values at $\eta_c = 0.5$ and 1.0 at 900 GeV). However, the success of three component NBD relies on a total of 8 parameters; one parameter more than our two component GMD.

It is probably worth mentioning also that fits using other models such as Independent Pair Parton Interactions (IPPI) [**?**] model and the multiladder exchange or Quark-Gluon String (QGSM) model [**?**] on the CMS data for $\eta_c = 2.4$ at 7000 GeV reported by Dremin and Nechitailo [**?**] has χ^2/dof values higher than the other models.

4.0.1 The Two Component GMD Parameters

Figures ?? to ?? shows the Two Component GMD parameters plotted out against η_c and also against $\sqrt{s}$ as a result of the χ^2 fitting.

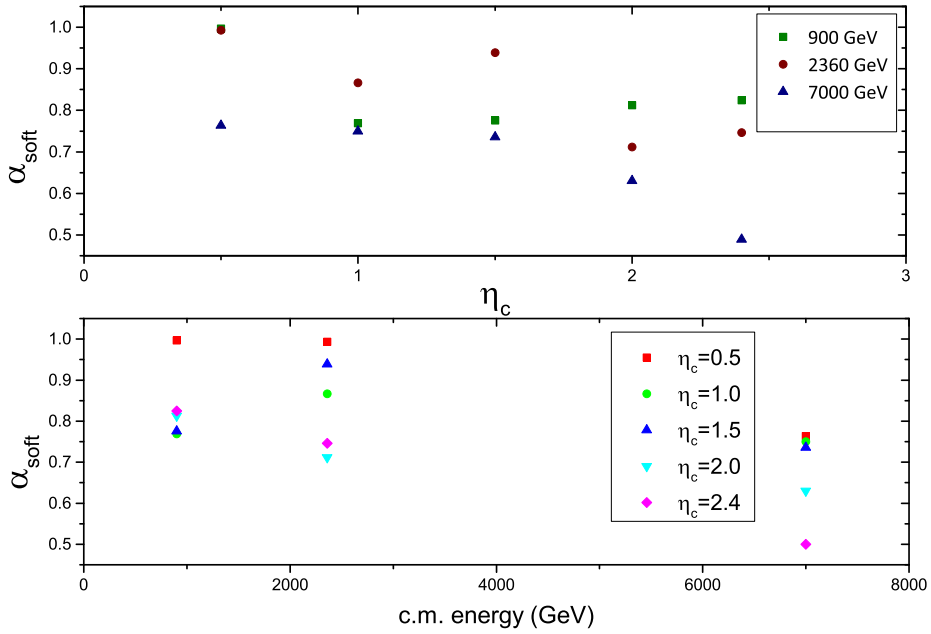


Figure 4.4: *Top:* α_{soft} plotted against pseudorapidity cuts η_c at all the three different center-of-mass energy $\sqrt{s}$. *Bottom:* α_{soft} plotted against center-of-mass energy $\sqrt{s}$ at all the five different pseudorapidity cuts η_c.

From Fig. ??, there seem to be an inverse relationship between α_{soft} and $\sqrt{s}$. A possible explanation would be that the semihard component becomes more important at higher $\sqrt{s}$. The value of α_{soft} seem to decrease with η_c for $\sqrt{s} = 900$ and 7000 GeV. However, the value of α_{soft} seem to increase over a small range for $\sqrt{s} = 2360$ GeV.

From Figs ?? and ??, the values of $\overline{n}_{soft}$ and $\overline{n}_{semihard}$ increases approximately linearly with η_c. This is natural as more particles will be observed as we increase η_c. Analysis on the values of $\overline{n}_{soft}$ and $\overline{n}_{semihard}$ shows that $\overline{n}_{semihard} \not\approx 2\,\overline{n}_{soft}$ as suggested by UA1 analysis on minijets mentioned in [?]. However the alternative postulate mentioned in

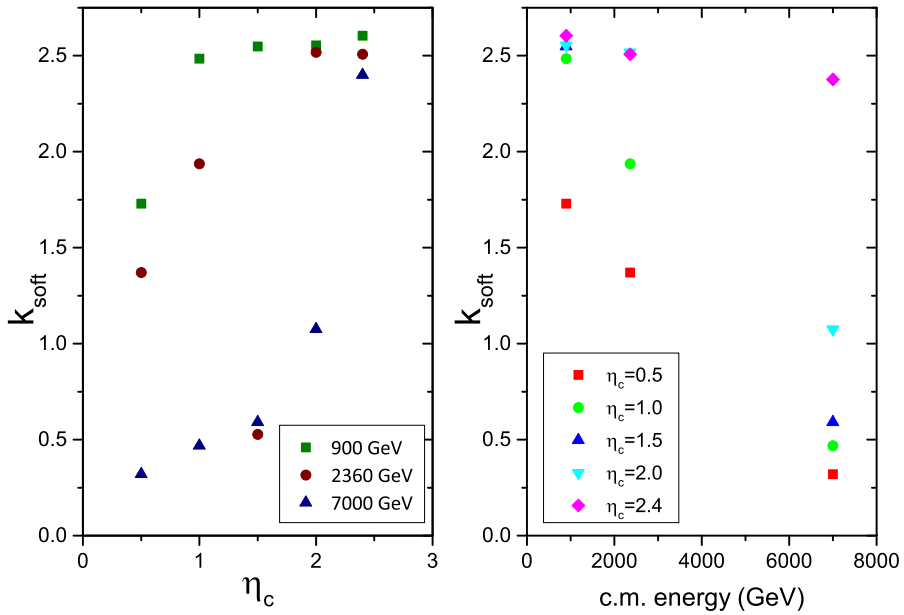


Figure 4.5: *Left:* k_{soft} plotted against pseudorapidity cuts η_c at all the three different center-of-mass energy $\sqrt{s}$. *Right:* k_{soft} plotted against center-of-mass energy $\sqrt{s}$ at all the five different pseudorapidity cuts η_c.

the same paper that $\overline{n}_{semihard}$ increases more rapidly with $\sqrt{s}$ according to

$$\overline{n}_{semihard} \approx 2\overline{n}_{soft} + c' \ln^2(\sqrt{s}) \tag{4.1}$$

seems to be more appropriate. Our analysis on the CMS data showed that $c' \approx 0.1$ as suggested in [**?**] except for $\eta_c = 2.0$ and 2.4 at 2360 GeV and 7000 GeV where the value of c' is closer to 0.2.

From Eq. (**??**), it is suggestive that the total multiplicity $\overline{n}_{tot}$ is

$$\overline{n}_{tot} = \alpha_{soft}\overline{n}_{soft} + (1 - \alpha_{soft})\overline{n}_{semihard} \tag{4.2}$$

We have plot in Fig. **??** the $\overline{n}_{tot}$ resulted from our two component GMD fit (calculated from α_{soft}, $\overline{n}_{soft}$ and $\overline{n}_{semihard}$ using Eq. (**??**)) against η_c and $\sqrt{s}$ and compared it with the experimental values of $\overline{n}_{tot}$ from CMS. It can be seen that the $\overline{n}_{tot}$ from our two component

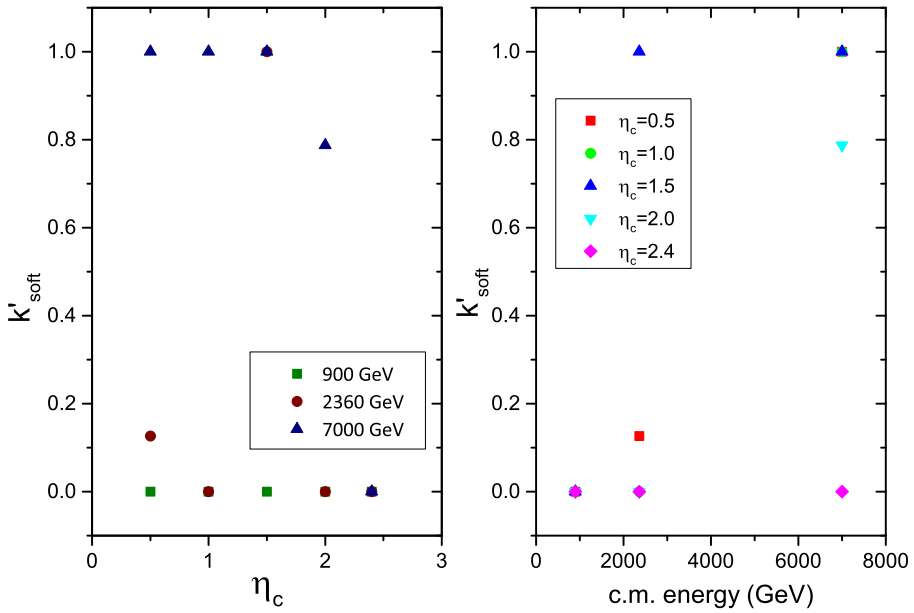


Figure 4.6: *Left:* k'_{soft} plotted against pseudorapidity cuts η_c at all the three different center-of-mass energy $\sqrt{s}$. *Right:* k'_{soft} plotted against center-of-mass energy $\sqrt{s}$ at all the five different pseudorapidity cuts η_c.

GMD model tend to overestimate the experimental values of $\overline{n}_{tot}$. However, with the exception for at $\sqrt{s} = 2360$ GeV, the values of $\overline{n}_{tot}$ still lies within the uncertainties of the experimental values.

From Figs. **??**, **??**, **??** and **??**, the behavior of k and k' for both soft and semihard components are rather erratic. k'_{soft} is virtually 0 for $\sqrt{s} = 900$ and 2360 GeV indicating an NBD behavior for the soft component of the multiplicity distribution. The behavior k'_{soft} seem to suggest that it will saturate at 1 at higher energies. k_{soft} tend to increase with η_c at a given energy but the behavior suggests a saturation at some value around 2.5 at $\eta_c = 2.4$. However, k_{soft} seems to have an inverse relationship with energy. At a given $\sqrt{s}$, the value of $k_{semihard}$ seem to start high before decreasing with η_c. The slope seem to become less negative with energy. $k'_{semihard}$ seem to have similar behavior at a given $\sqrt{s}$ but seems to tend to 0 at high energies. The behavior of $k'_{semihard}$ is too erratic to have a trend at low pseudorapidity cuts η_c but seem to decrease with energy at wider

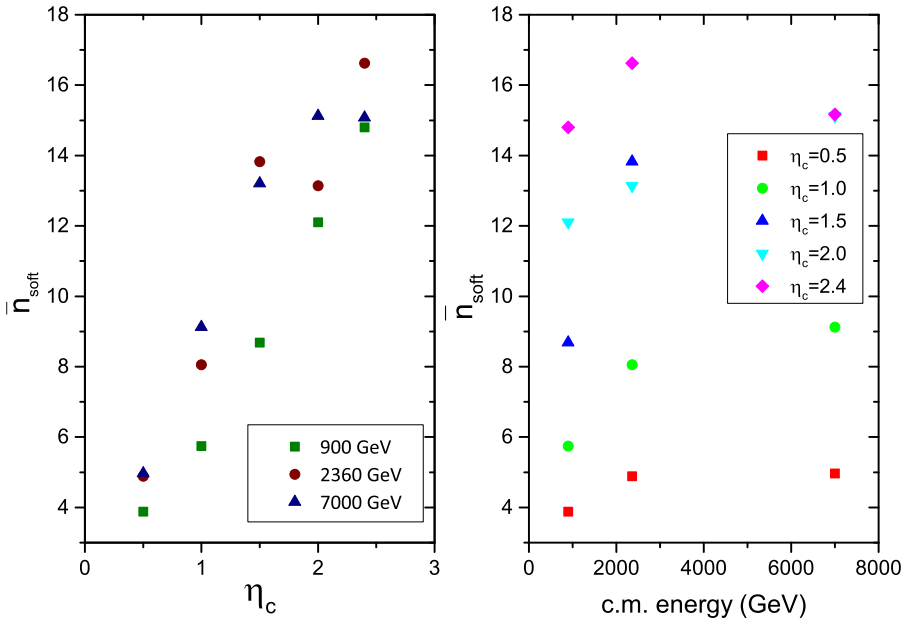


Figure 4.7: *Left:* $\overline{n}_{soft}$ plotted against pseudorapidity cuts η_c at all the three different center-of-mass energy $\sqrt{s}$. *Right:* $\overline{n}_{soft}$ plotted against center-of-mass energy $\sqrt{s}$ at all the five different pseudorapidity cuts η_c.

η_c. It should be noted that k and k' for the semihard component are less likely to be reliable at narrow η_c as the values of α_{soft} is close to 1.

In our derivation of GMD, k refers to the initial number of quarks in the average sense while k' refers to the initial number of gluons in the average sense. From our results, since the soft component seem to exhibit NBD behavior, it seems to suggest that gluons do not tend to take part in the soft component of the parton branching before hadronization. However, there is a problem in this interpretation because our results show that the values of k' at narrower pseudorapidity cuts η_c is larger than at wider cuts. Does this mean more gluons is involved in parton branching if we observe a narrower η_c as opposed to observing a wider η_c? If the interpretation that k' refers to the initial number of gluons in the average sense holds, we would expect that k' would either be approximately constant or increase with η_c

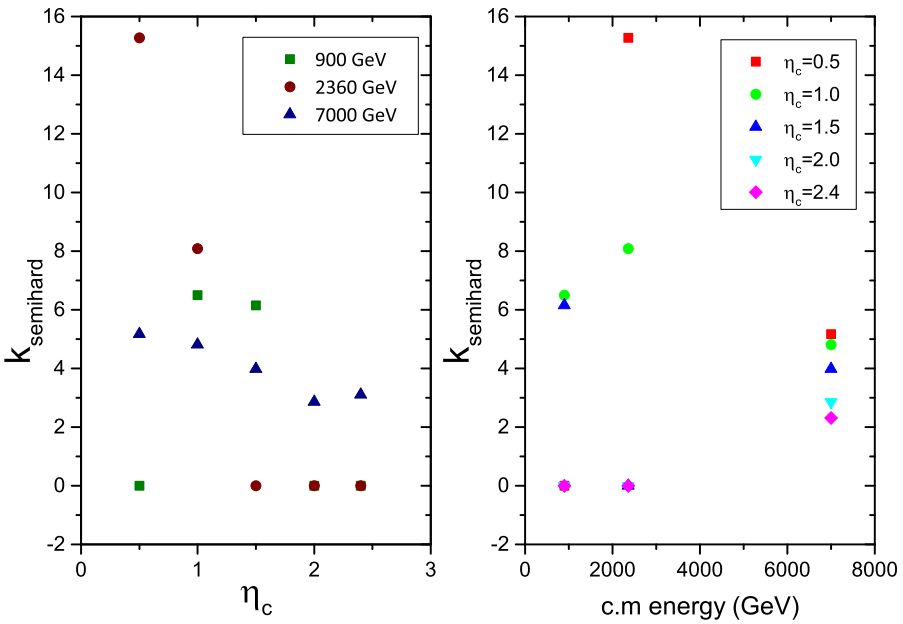


Figure 4.8: *Left:* $k_{semihard}$ plotted against pseudorapidity cuts η_c at all the three different center-of-mass energy $\sqrt{s}$. *Right:* $k_{semihard}$ plotted against center-of-mass energy $\sqrt{s}$ at all the five different pseudorapidity cuts η_c.

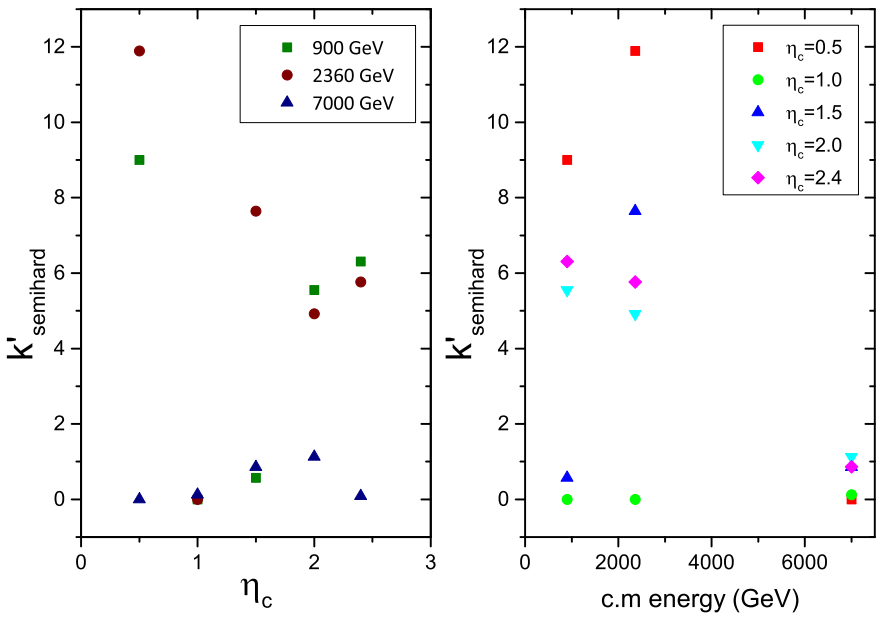


Figure 4.9: *Left:* $k'_{semihard}$ plotted against pseudorapidity cuts η_c at all the three different center-of-mass energy $\sqrt{s}$. *Right:* $k'_{semihard}$ plotted against center-of-mass energy $\sqrt{s}$ at all the five different pseudorapidity cuts η_c.

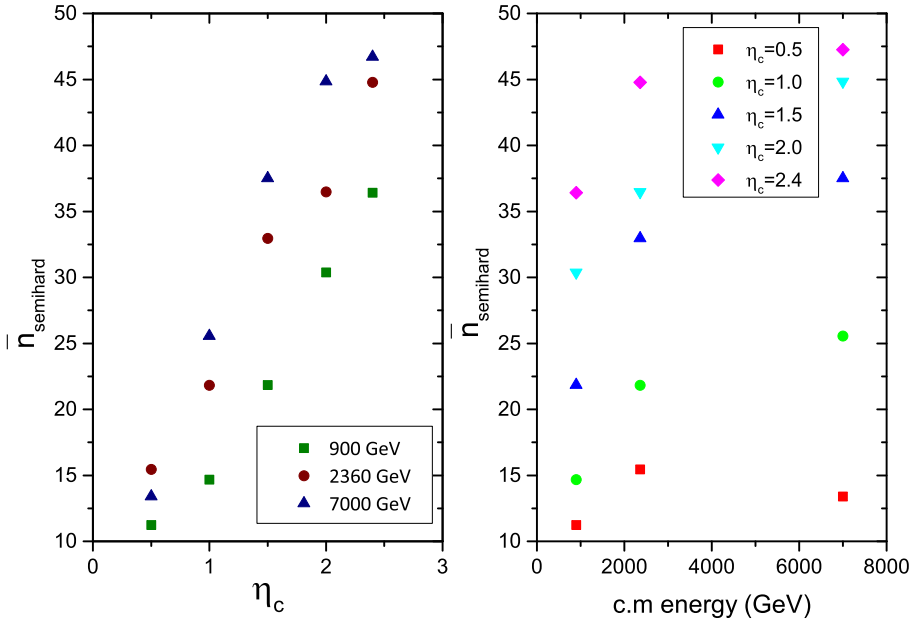


Figure 4.10: *Left:* $\overline{n}_{semihard}$ plotted against pseudorapidity cuts η_c at all the three different center-of-mass energy $\sqrt{s}$. *Right:* $\overline{n}_{semihard}$ plotted against center-of-mass energy $\sqrt{s}$ at all the five different pseudorapidity cuts η_c.

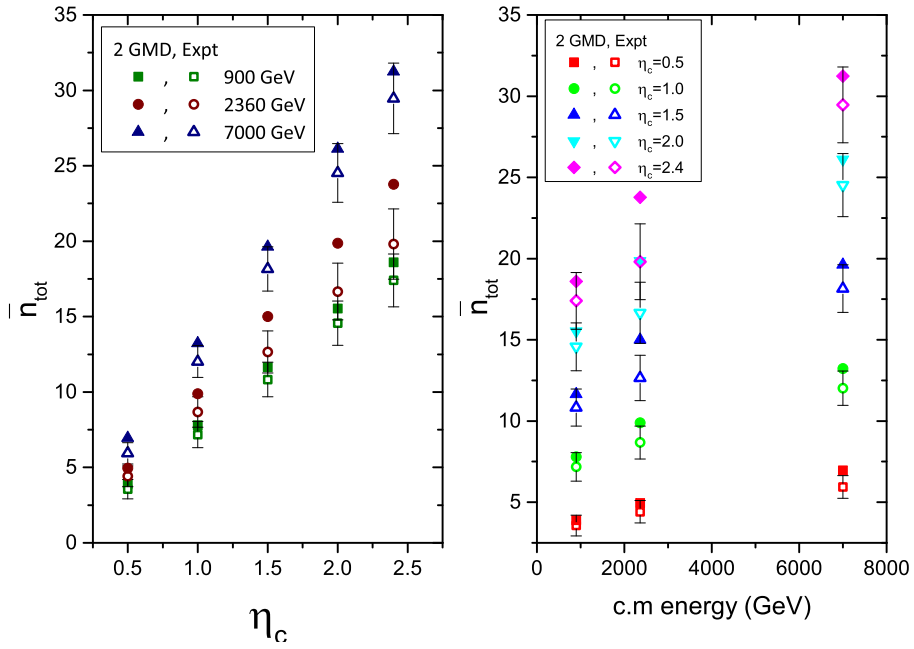


Figure 4.11: *Left:* $\overline{n}_{tot}$ from our two component GMD plotted against pseudorapidity cuts η_c at all the three different center-of-mass energy $\sqrt{s}$ and compared with $\overline{n}_{tot}$ from experiment. *Right:* $\overline{n}_{tot}$ from our two component GMD plotted against center-of-mass energy $\sqrt{s}$ at all the five different pseudorapidity cuts η_c and compared with $\overline{n}_{tot}$ from experiment.

4.0.2 C_q Moments (Results)

The C_q moments of the fits using the two component GMD model for the three different center-of-mass energies $\sqrt{s}$ are calculated and given in Tables **??** to **??** below. We compare them with the experimental C_q moments and plot Graphs in Figs. **??** to **??**.

900 GeV								
	C_2		C_3		C_4		C_5	
η_c	2 GMD	Expt	2 GMD	Expt	2 GMD	Expt	2 GMD	Expt
0.5	1.96	1.96	5.21	5.21	17.0	17.0	64.0	64.0
1.0	1.76	1.76	4.12	4.13	11.6	11.6	37.3	37.6
1.5	1.67	1.67	3.63	3.64	9.42	9.46	27.6	27.8
2.0	1.60	1.60	3.29	3.30	7.98	8.03	21.7	21.9
2.4	1.56	1.56	3.08	3.10	7.14	7.18	18.4	18.6

Table 4.5: C_q moments from our two component GMD model compared with the experimental values at $\sqrt{s} = 900$ GeV.

2360 GeV								
	C_2		C_3		C_4		C_5	
η_c	2 GMD	Expt	2 GMD	Expt	2 GMD	Expt	2 GMD	Expt
0.5	1.90	1.96	4.78	4.79	14.5	14.5	50.4	50.5
1.0	1.75	1.76	3.99	4.00	10.8	10.8	33.1	33.4
1.5	1.68	1.67	3.59	3.60	9.01	9.08	25.6	25.9
2.0	1.64	1.60	3.38	3.40	8.19	8.24	22.4	22.6
2.4	1.60	1.56	3.26	3.28	7.79	7.90	21.2	21.7

Table 4.6: C_q moments from our two component GMD model compared with the experimental values at $\sqrt{s} = 2360$ GeV.

We can see that the C_q moments of the two component GMD model lie very closely to the C_q moments of the experimental data indicating an excellent fit. We can also see

7000 GeV								
	C_2		C_3		C_4		C_5	
η_c	2 GMD	Expt	2 GMD	Expt	2 GMD	Expt	2 GMD	Expt
0.5	2.02	2.02	5.64	5.65	19.3	19.4	76.0	76.5
1.0	1.86	1.86	4.68	4.70	14.3	14.3	49.4	49.8
1.5	1.78	1.78	4.23	4.26	12.0	12.2	38.8	39.4
2.0	1.72	1.73	3.93	3.97	10.7	10.8	32.5	33.1
2.4	1.68	1.69	3.71	3.75	9.69	9.83	28.2	28.8

Table 4.7: C_q moments from our two component GMD model compared with the experimental values at $\sqrt{s} = 7000$ GeV.

in Fig. **??** that C_q moments are more approximately constant at narrower pseudorapidity cuts η_c but has an obvious linear increase at wider η_c. This suggests that KNO scaling is approximately obeyed at narrow pseudorapidity cuts η_c but violated at wider η_c.

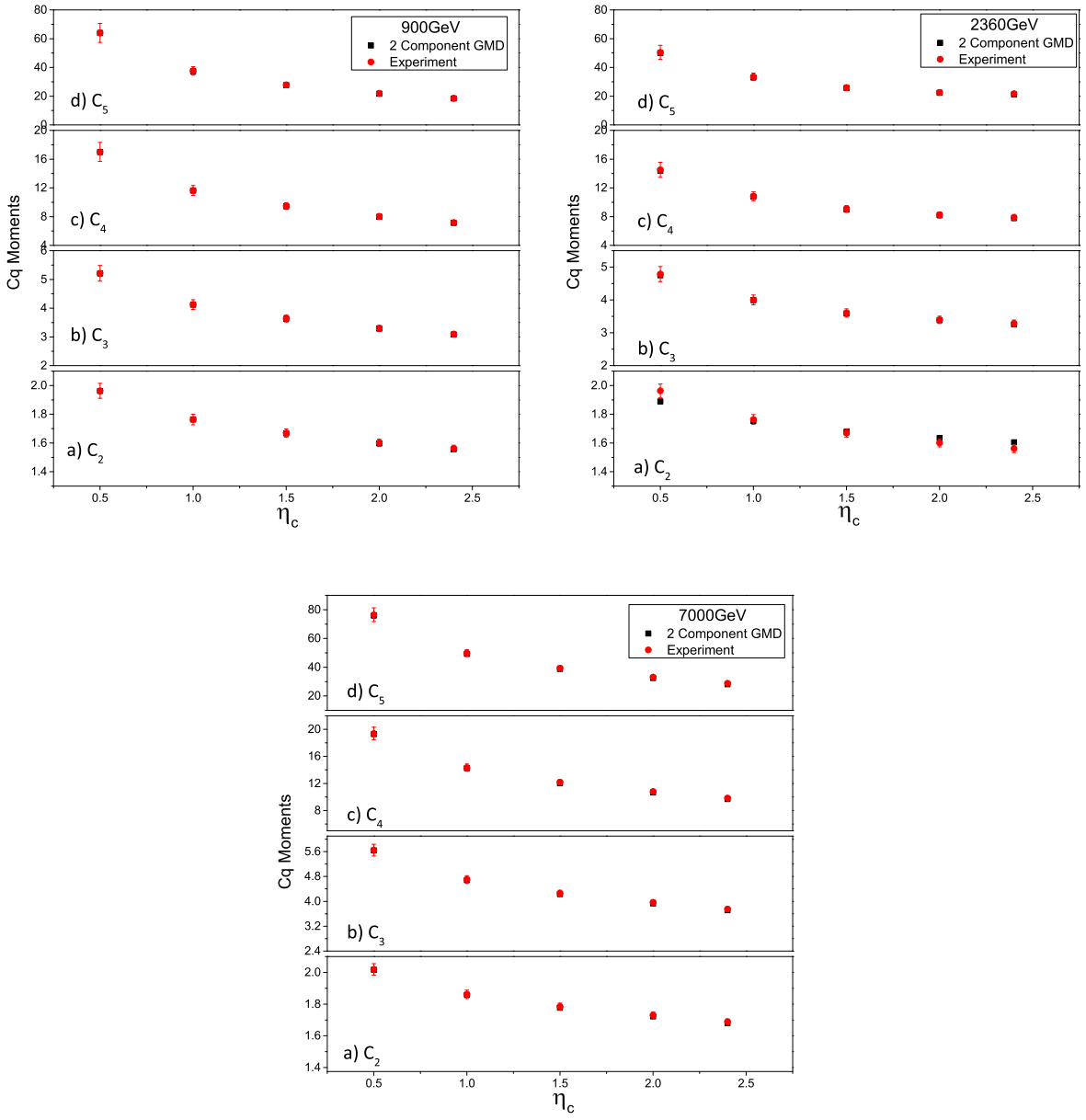


Figure 4.12: C_q moments plotted against η_c for all the three different center-of-mass energies $\sqrt{s}$.

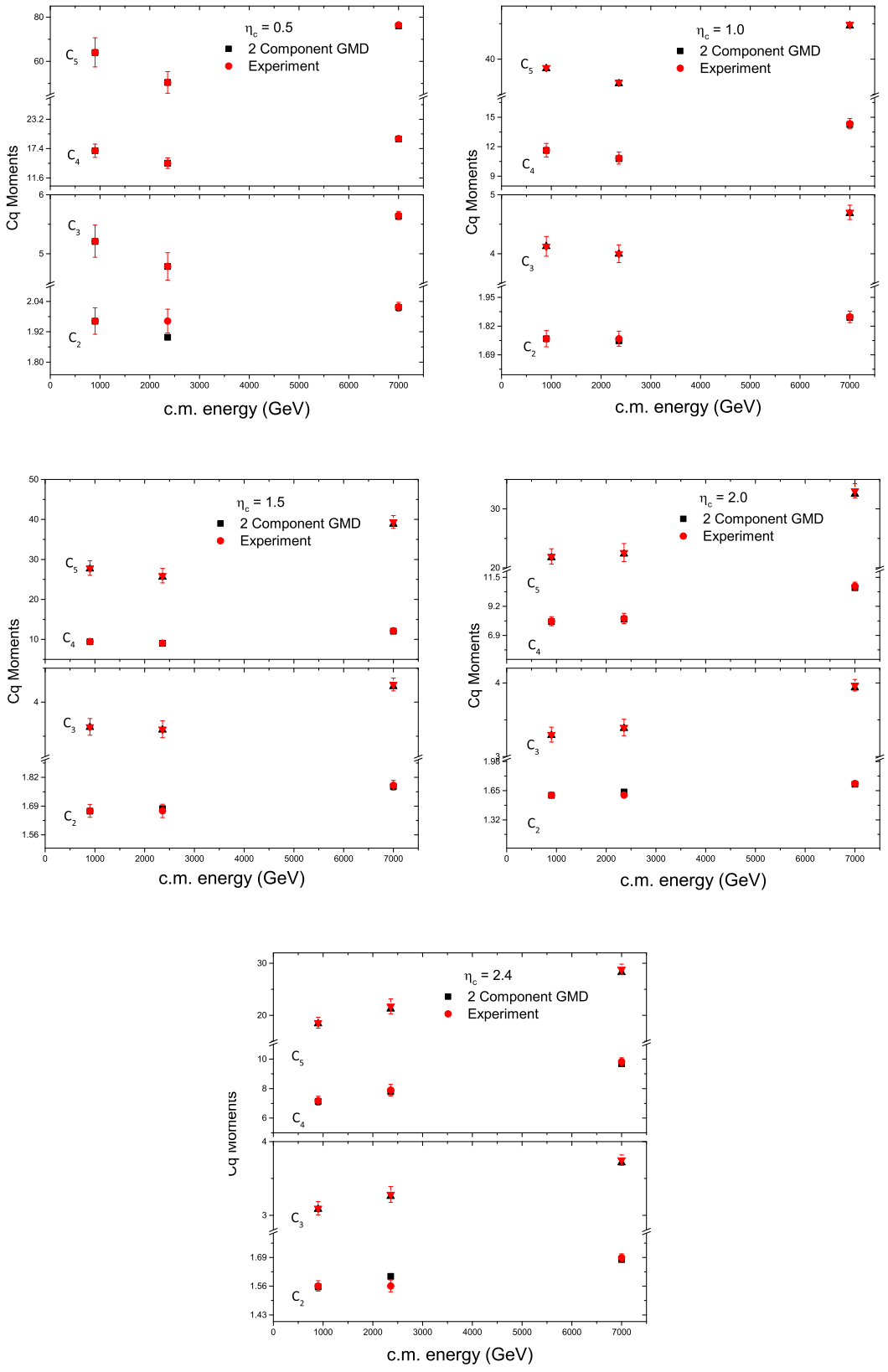


Figure 4.13: C_q moments plotted against center-of-mass energies $\sqrt{s}$ for all the five different pseudorapidity cuts η_c.

Chapter 5

Predictions at Higher Energies

As of the writing of this work, the LHC is currently offline for maintenance and upgrades. When it reopens in 2015, it is expected to run about double the previous energy (expected to be between 13-14 TeV). We will attempt to make predictions based on the results of our work. We shall assume that the CMS will operate at 14 TeV and continue to collect multiplicity data at the same pseudorapidity cuts η_c as it has done so in the previous three center-of-mass energies $\sqrt{s}$. Using this assumption, we will attempt to predict the behaviour of the multiplicity distribution at the five different pseudorapidity cuts η_c.

Based on previous multiplicity data, the total average multiplicity $\overline{n}_{tot}$ is best described by a quadratic fit of the form

$$\overline{n}_{tot} = a + b\ln(\sqrt{s}) + c\ln^2(\sqrt{s}) \tag{5.1}$$

where $a = 3.01$, $b = -0.474$ and $c = 0.754$ for the FPS data [**?**]. We assume a similar behaviour at different pseudorapidity cuts η_c. By using the corresponding average multiplicity values from UA5 [**?**, **?**] at the same η_c, we fit the data using Eq. (**??**) and find the parameters a, b and c. Since UA5 did not collect data at $\eta_c = 2.4$, we therefore

are forced to use the only 3 data points we have from CMS to fit it with Eq. (**??**). In this case, we have fixed a to be 3.01. Using the parameters found above, we then extrapolate the curve to predict the values of $\overline{n}_{tot}$ at 14 TeV.

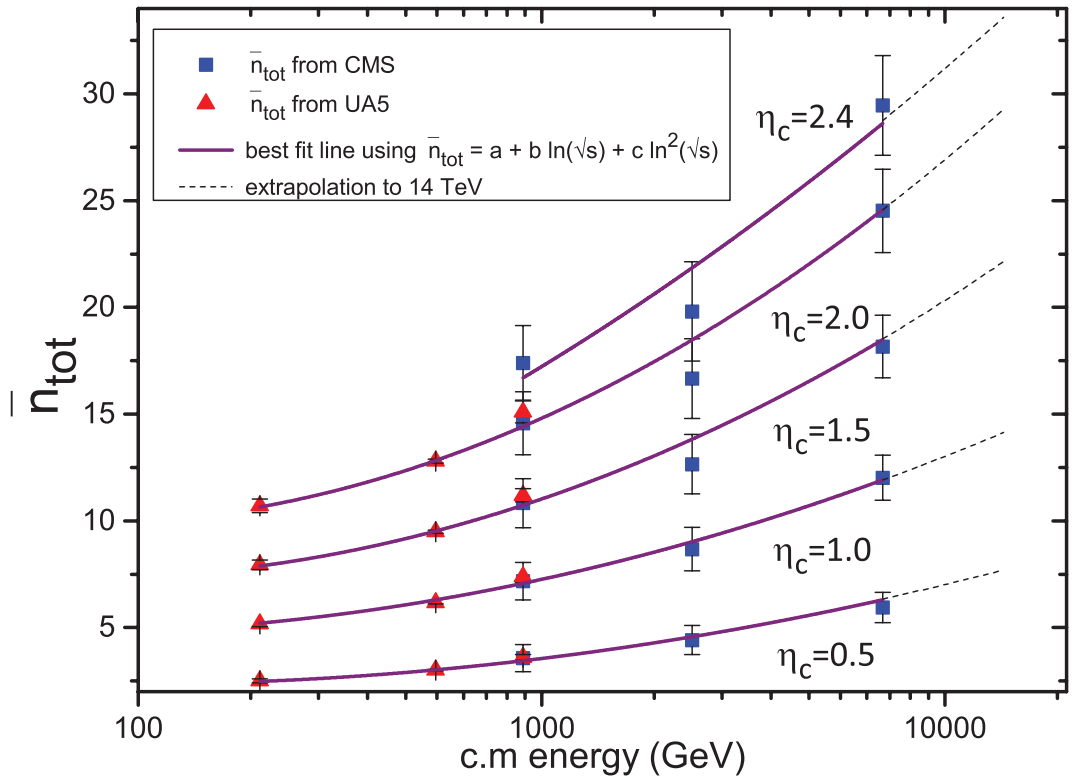


Figure 5.1: $\overline{n}_{tot}$ from CMS and UA5 plotted against center-of-mass energy $\sqrt{s}$ and fitted using Eq. (**??**) (solid purple lines) at all the five different pseudorapidity cuts η_c. The parameters a, b and c obtained from the fittings are used to extrapolate the curves (black dotted lines) to predict the value of $\overline{n}_{tot}$ at 14 TeV.

To predict the behaviour of a two component GMD at 14 TeV, we therefore require a means to guess the value of α_{soft} at that energy. Looking at Fig. **??**, We take the most simple approach by assuming that α_{soft} is linearly decreasing with energy. We then extrapolate the linear fit to find the probable values of α_{soft} at 14 TeV. We then calculate the predicted values of $\overline{n}_{soft}$ and $\overline{n}_{semihard}$ using Eqs. (**??**) and (**??**) by setting $c' = 0.1$.

The values of k and k' for both soft and semihard components are harder to predict. It is probably good to assume that k'_{soft} to be saturated at 1 as hinted by our results. We leave the other three parameters to vary when the multiplicity data at 14 TeV becomes

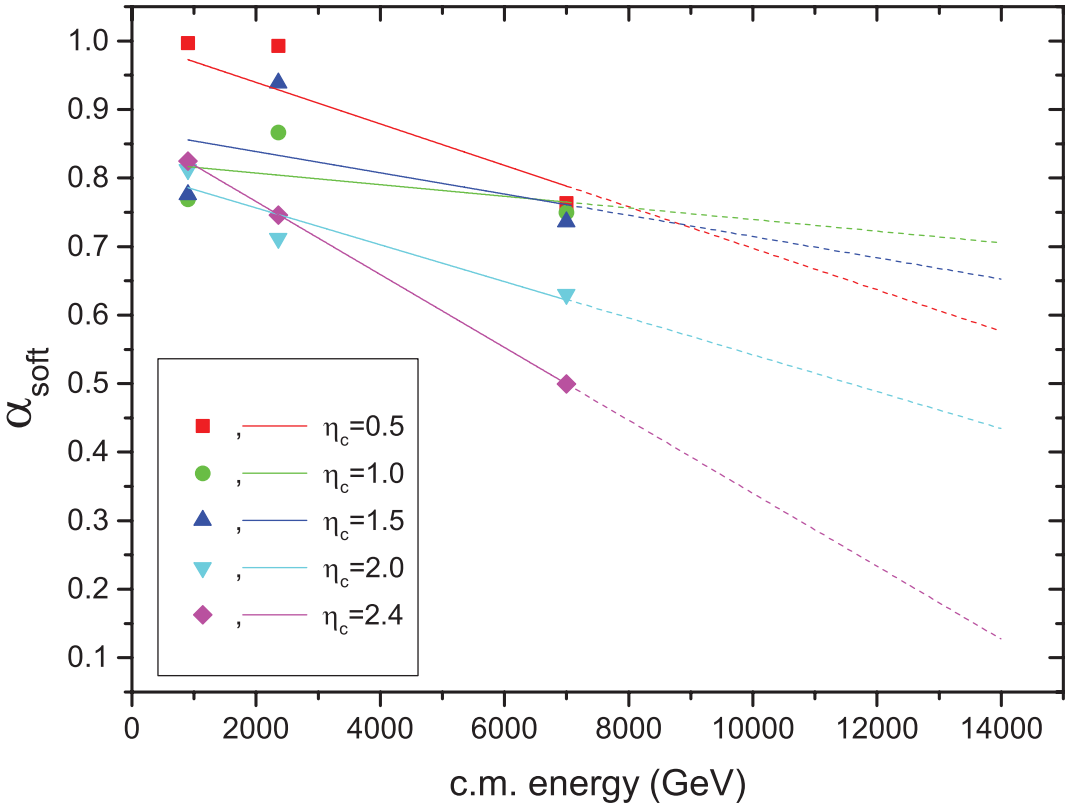


Figure 5.2: α_{soft} plotted against center-of-mass energy $\sqrt{s}$ fitted using a linear fit (solid lines) at all the five different pseudorapidity cuts η_c. The fittings are then extrapolated (dotted lines) to predict the value of α_{soft} at 14 TeV. c.f. Fig. **??**.

available. Table **??** gives a summary of our predicted values of the two component GMD parameters at 14 TeV. Fig. **??** shows possible shape of the distribution of the data at 14 TeV based on the values in Table **??**.

η_c	α_{soft}	k_{soft}	k'_{soft}	$\overline{n}_{soft}$	$k_{semihard}$	$k'_{semihard}$	$\overline{n}_{semihard}$	$\overline{n}_{tot}$
0.5	0.58	*	1.00	5.10	*	*	11.15	7.66
1.0	0.71	*	1.00	10.71	*	*	22.38	14.15
1.5	0.65	*	1.00	16.17	*	*	33.28	22.11
2.0	0.43	*	1.00	18.36	*	*	37.68	29.31
2.4	0.13	*	1.00	17.42	*	*	35.79	33.45

Table 5.1: Summary of our prediction of the values of the two component GMD model at 14 TeV. Cells marked with * indicates parameters to be varied when the experimental data is available for fitting.

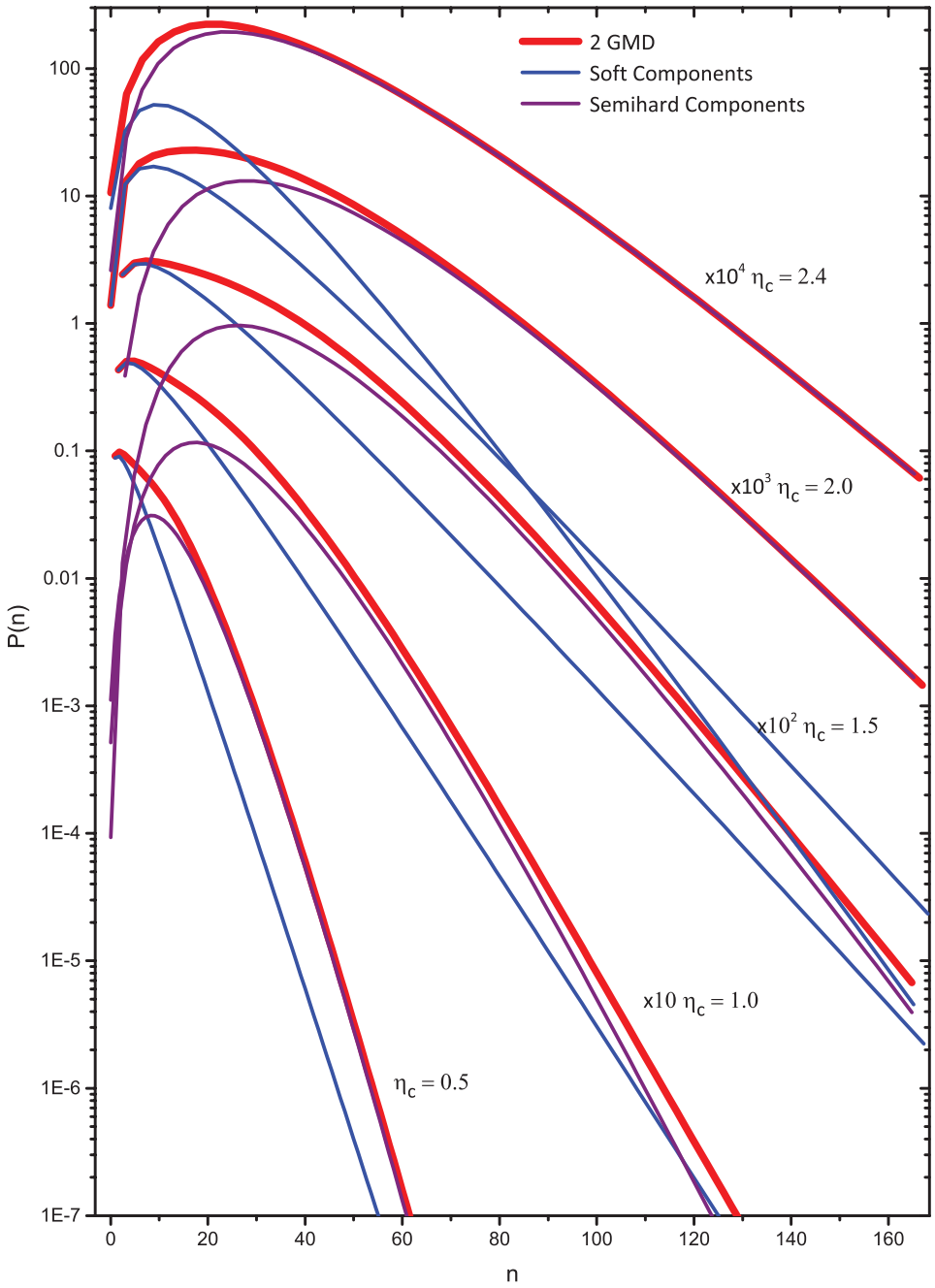


Figure 5.3: Possible distribution shapes of multiplicity distributions at 14 TeV at all the five different η_c. η_c labels on the graph refer to the red lines, which are the predicted two component GMD curve at 14 TeV. The blue lines (lines that peak at low multiplicities) and the purple lines (lines that peak at slightly higher multiplicities) are the soft and semihard components of the two components at the different η_c respectively. In this figure, the curve is generated using k_{soft}, $k_{semihard}$ and $k'_{semihard}$ from our results in Table **??**.

From Fig. **??**, it can be seen that our predictions dictate that the semihard component becomes more and more prominent with increase in η_c to the extent that the shouldering almost disappears as the peak of the multiplicity distributions shift to higher values. There is approximately a linear relationship between the length of the tail of the distribution and $\sqrt{s}$. For example, we would expect that at 14 TeV at $\eta_c = 2.4$, the tail of the distribution would reach $n \approx 320$. However, in Fig. **??**, we unfortunately have to cut the predicted curve at $n \approx 160$ due to the limitation is our computation of the value of the Gamma function[1].

From the results in Zborovský's paper [**?**] there is an indication that a two component model may be inadequate to describe the anticipated multiplicity data at 14 TeV. The indication comes from the elongation of the tail of the multiplicity distribution and also the peak structure at low multiplicities. The two component model seems to be inadequate to describe the peak and the tail of the multiplicity distribution well enough simultaneously. The erratic behavior of k and k' is also another indication of how a two component model may not be very suitable at higher energies. The three component NBD model however, is able to describe the peak structure of the multiplicity distribution better than the two component NBD. It is also demonstrated that the values of the parameter $1/k$ for all 3 components of NBD are approximately invariant. It is perhaps possible to employ a three component GMD at 14 TeV (or variants of it eg. GMD+FYD+Poisson). This will however, come at the expense of increasing the number of fitting parameters. We could consider this to be an angle of attack for future work.

[1] See Appendix **??**

Chapter 6

Conclusion and Future Work

We have first introduced the General Multiplicity Distribution (GMD) formalism for multiplicity distributions before extending it to a Two Component GMD model. We then fit the multiplicity data from CMS with the two component GMD model and demonstrated the goodness of fit through the χ^2/dof and C_q moments. We also compared our fit results with other models on the same data sets. We proceeded to analyze the GMD parameters from the results and attempted to make some predictions at 14 TeV.

Much work still has to be done to understand the hadronization process that leads to the observed multiplicity distribution. While we await for the data at 14 TeV, one may try to fit the multiplicity distribution from CMS at large transverse momentum cut of $p_T > 500$ MeV. These corresponds to harder processes in the high energy collisions. Fits using two component GMD could also be attempted on data from ATLAS and ALICE which have comparable conditions as the multiplicity data of that of CMS. It would also probably be interesting to see how GMD fairs using the multiplicity data from the LHCb which deals with multiplicities at pseudorapidity range close to the beam line.

The erratic behavior of k and k' from our results poses questions to the interpretation of these two GMD parameters. While the two component GMD model gives excellent fits to the multiplicity data, interpreting k and k' as initial number of gluons and quarks on the average sense respectively raises further questions. One may like to try if it is possible to arrive at the GMD formulation not necessarily from the stochastic branching equation (**??**). From Eqs. (**??**) and (**??**), we remember that k and k' describes the width of NBD and FYD respectively. We would therefore imagine that the width of GMD is a convolution of k and k'. However, no one has yet to write down this relationship down explicitly. Understanding the behavior of the width of GMD may give us more insight on the behavior of GMD. In conclusion, a better understanding of k and k' is needed for GMD for us to take advantage of its ability to describe multiplicity data well.

It is exciting to see what the multiplicity data will be like at higher energies. Will the tail of the distribution continue to become longer and the peak at low multiplicities become sharper? Or will there be unexpected structures appearing at higher energies? The path to understand the hadronization mechanism seems to be still far from sight. We can only hope that as we push the energy limits of our colliders to even greater heights, the emerging patterns from the data would throw more light into further understanding the ever elusive hadronization mechanism.

Appendix A

Rapidity and Pseudorapidity

A.1 Rapidity

Rapidity and Pseudorapidity are two variables that are in common use in accelerator physics. In accelerators the incident velocities of the particles taking part in a collision are along the beam axis. This leads to the definition of various quantities that are either with respect to boosts to the rest frames of observers moving at different velocities parallel to the beam axis, or others that although they are not invariant have transformation properties that are easy to handle and useful for analysis.

Taking the beam axis to be the z direction, the Rapidity y of a particle is defined as

$$y = \frac{1}{2} \ln \left(\frac{E + p_z c}{E - p_z c} \right) \tag{A.1}$$

where E is the energy of the particle, c is the speed of light in vacuum and p_z is the momentum of the particle in the z direction.

Rapidity is defined in this way to deal with the very high energy product of collisions in high energy experiments. Suppose a particle is directed essentially in the xy-plane, perpendicular to the beam direction. Then the p_z will be small, and the rapidity will be close to 0. For a highly relativistic particle directed predominantly down the beam axis (eg. the $+z$ direction), $E \approx p_z c$ and $y \to \infty$. Similarly, a particle travelling down the negative beam axis, $E \approx -p_z c$, then $y \to -\infty$.

Using $\tanh = (e^{\theta} - e^{-\theta})/(e^{\theta} + e^{-\theta})$, it could be shown that **??** could be written as

$$y = \tanh^{-1}\left(\frac{p_z c}{E}\right) \tag{A.2}$$

Under a Lorentz Boost parallel to the z-axis, it could be shown that the transformation of $y \to y'$ is

$$y' = y + \ln\sqrt{\frac{1-\beta}{1+\beta}} = y - \tanh^{-1}\beta \tag{A.3}$$

where $\beta = \dfrac{v}{c}$

This particularly simple transformation law for y has an important consequence. Suppose we have two particles ejected after a collision, and they have rapidities y_1 and y_2 when measured by some observer. Now, let some other observer measure these same rapidities, and obtain y_1' and y_2'. The difference between the rapidities of the two frames becomes

$$y_1' - y_2' = (y_1 - \tanh^{-1}\beta) - (y_2 - \tanh^{-1}\beta) = y_1 - y_2 \tag{A.4}$$

This shows that the rapidities of two particles is invariant with respect to Lorentz boosts along the z-axis. This is the main reason why rapidities are so crucial in accelerator physics.

A.2 Pseudorapidity

For highly relativistic particles, however, it can be very hard to measure their rapidities because both the energy and the total momentum is needed. In reality it is often difficult to get the total momentum vector of a particle, especially at high values of the rapidity where the z component of the momentum is large, and the beam pipe can be in the way of measuring it precisely. The way around this is by defining another quantity: pseudorapidity that is almost the same thing for high highly energetic particles.

$$\begin{aligned} y &= \frac{1}{2}\ln\left(\frac{E+p_z c}{E-p_z c}\right) \\ &= \frac{1}{2}\ln\left(\frac{\sqrt{p^2c^2+m^2c^4}+p_z c}{\sqrt{p^2c^2+m^2c^4}-p_z c}\right) \end{aligned} \tag{A.5}$$

For a highly relativistic particle, $pc \gg mc^2$. We factor out pc from square root terms in the numerator and denominator and use a binomial expansion to approximate the inside

$$\begin{aligned} y &= \frac{1}{2}\ln\left(\frac{pc\left(\sqrt{p^2c^2+m^2c^4}\right)+p_z c}{pc\left(\sqrt{p^2c^2+m^2c^4}\right)-p_z c}\right) \\ &= \frac{1}{2}\ln\left(\frac{pc+p_z c+\frac{m^2c^4}{2pc}+\dots}{pc-p_z c+\frac{m^2c^4}{2pc}+\cdots}\right) \\ &= \frac{1}{2}\ln\left(\frac{1+\frac{p_z}{p_c}+\frac{m^2c^4}{2p^2c^2}+\dots}{1+\frac{p_z}{p_c}+\frac{m^2c^4}{2p^2c^2}+\cdots}\right) \end{aligned} \tag{A.6}$$

Now $p_z/p = \cos\theta$, where θ is the angle made by the particle trajectory with the beam pipe, and hence we have

$$1+\frac{p_z}{p} = 1+\cos\theta = 1+\left(\cos^2\frac{\theta}{2}-\sin^2\frac{\theta}{2}\right) = 2\cos^2\frac{\theta}{2} \tag{A.7}$$

and

$$1 - \frac{p_z}{p} = 1 - \cos\theta = 1 - \left(\cos^2\frac{\theta}{2} - \sin^2\frac{\theta}{2}\right) = 2\sin^2\frac{\theta}{2} \tag{A.8}$$

Substituting back to **??**, we obtain

$$\begin{aligned} y &\approx \frac{1}{2}\ln\frac{\cos^2\frac{\theta}{2}}{\sin^2\frac{\theta}{2}} \\ &\approx -\ln\tan\frac{\theta}{2} \end{aligned} \tag{A.9}$$

We therefore define the pseudorapidity η as

$$\eta = -\ln\tan\frac{\theta}{2} \tag{A.10}$$

so that for highly relativistic particles, $y \approx \eta$. Pseudorapidity is particularly useful in hadron colliders such as the LHC, where the composite nature of the colliding protons means that interactions rarely have their centre of mass frame coincident with the detector rest frame, and where the complexity of the physics means that η is far quicker and easier to estimate than y. Furthermore, the high energy nature of the collisions mean that the two quantities may in fact be almost identical.

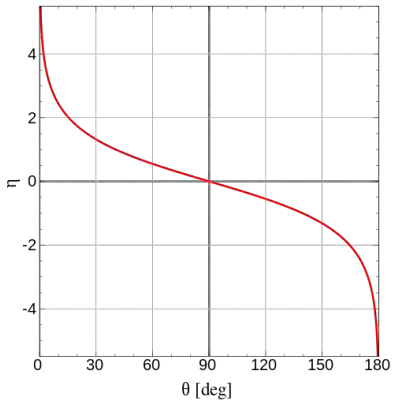


Figure A.1: A plot of polar angle θ vs. pseudorapidity η.

Appendix B

Event classes

pp collisions can be divided into elastic and inelastic ones. Elastic collisions are collisions in which both protons remain intact after the collision and are detected at (usually) the high rapidity regions of the detector with no production of any other particles. Inelastic collisions are commonly divided into non-diffractive (ND) events, single-diffractive (SD) and double-diffractive (DD). Figure **??** shows rapidity distributions of these three classes obtained with the event generator Pythia to illustrate their differences. A non-diffractive collision have many particles detected in the central rapidity region of the detector with their yield steeply falling towards higher rapidities. In a single-diffractive collision, only one of the protons breaks up to produce particles. This leads to particles detected at high rapidities on one side. The other proton, still intact and with slightly altered momentum is found near the rapidity of the beam on the other side of the detector. In a double diffractive collision, both protons break up and produce particles. Most of the particles will be found at the higher rapidities compared to the central rapidity region of the detector. Integrating the three graphs shows that the average total multiplicity is about a factor of 4 higher in non-diffractive collisions than in diffractive collisions.

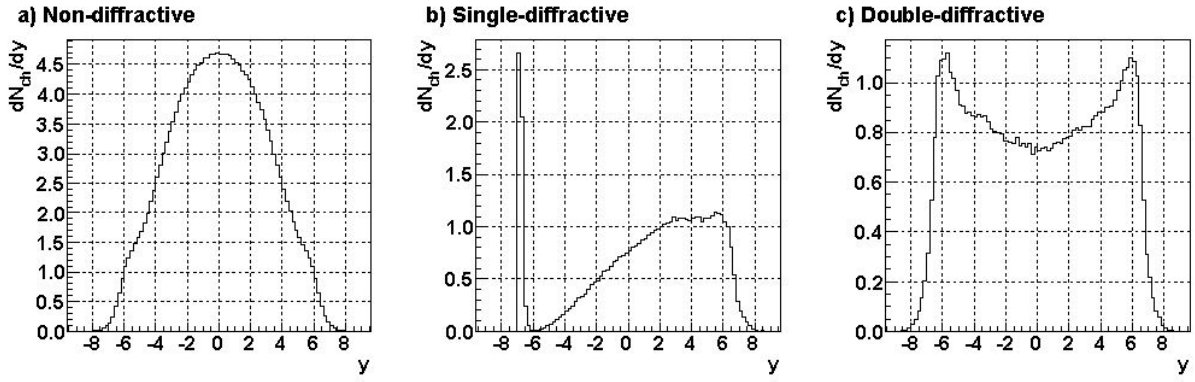


Figure B.1: Rapidity distributions of chraged particles per event for different processes (a) Non-diffractive (ND) (b) Single-diffractive (SD) (c) Double-diffractive (DD) generated using the event generator Pythia at $\sqrt{s} = 900$ GeV. Note the different scales of the three distributions. Image from [**?**].

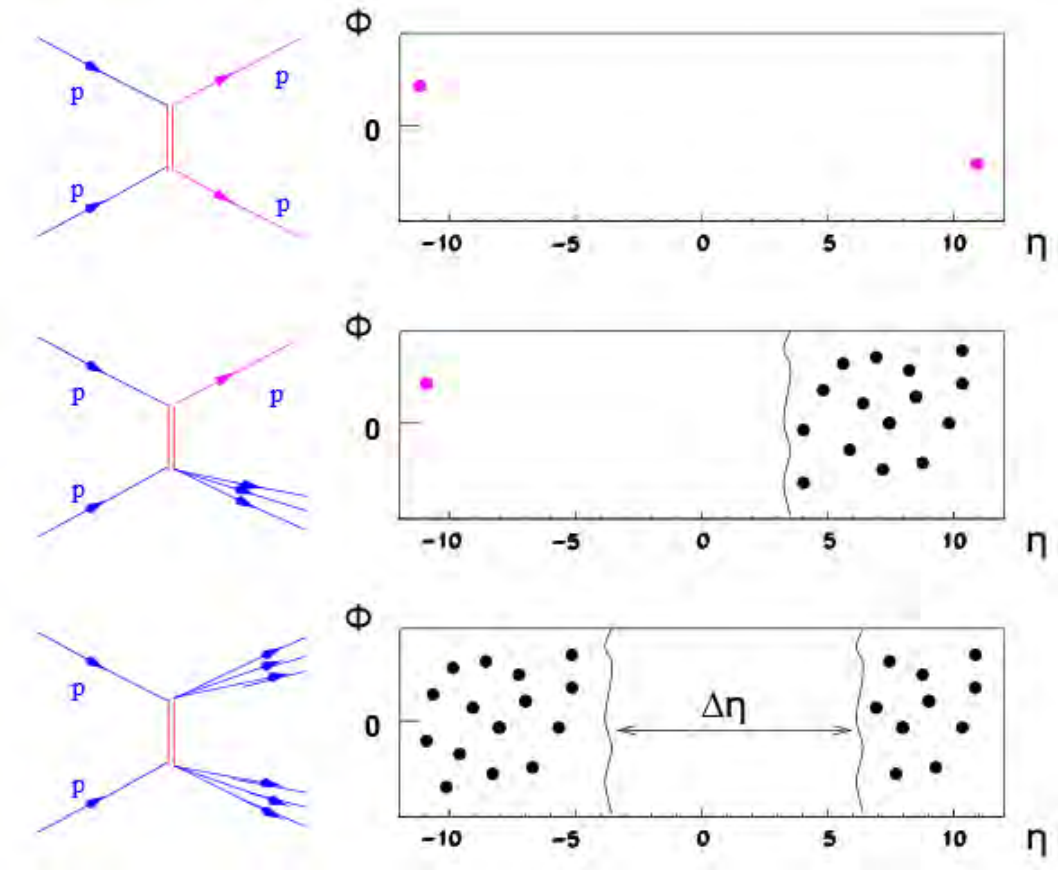


Figure B.2: Graphical representation of the most common event classes in *pp* collisions. The pictures on the left column are graphical representations of the processes and in the right columnar are typical angular and pseudorapidity distributions. *Top row:* Elastic Scattering *Middle row:* Single-diffraction *Bottom row:* Double-diffraction. Adapted from [**?**].

Appendix C

Proofs in the Derivation of Generalized Multiplicity Distribution

To show Eq. (**??**),

$$
\begin{aligned}
\frac{\partial f}{\partial t} &= \frac{\partial}{\partial t}\left[\sum_{n=0}^{\infty} P_n s^s\right] = \sum_{n=0}^{\infty} \frac{dP_n}{dt} s^n \\
&= \sum_{n=0}^{\infty} \left[-AnP_n s^n - \tilde{A}mP_n s^n - BnP_n s^n + A(n-1)P_n s^n - \tilde{A}mP_{n-1}s^n + B(n+1)P_{n+1}s^n\right] \\
&= \sum_{n=0}^{\infty} \left[-AnP_n s^n - \tilde{A}mP_n s^n - BnP_n s^n + AnP_n s^{n+1} + \tilde{A}mP_n s^n - BnP_n s^{n-1}\right] \\
&= (B - Bs - As + As^2)\sum_{n=0}^{\infty} P_n n s^{n-1} - \tilde{A}m(1-s)\sum_{n=0}^{\infty} P_n s^n \\
&= (1-s)(B-As)\sum_{n=0}^{\infty} P_n \frac{ds^n}{ds} - \tilde{A}m(1-s)\sum_{n=0}^{\infty} P_n s^n \\
&= (1-s)(B-As)\frac{\partial f}{\partial s} - \tilde{A}m(1-s)f
\end{aligned}
$$

□

To show Eq. (**??**), we first substitute $f(t=0,s)=s^{k'}$ to Eq. (**??**)

$$\frac{\tilde{A}}{A}m\ln(As-B)+\ln s^{k'}=\Psi\left(\frac{1-s}{As-B}\right)$$

Denote $u=\dfrac{1-s}{As-B}\Rightarrow s=\dfrac{uB+1}{uA+1}$, and substitute s to the equation above

$$\frac{\tilde{A}}{A}m\ln\left[A\left(\frac{uB+1}{uA+1}\right)-B\right]+k'\ln\left[\frac{uB+1}{uA+1}\right]=\Psi\left(u\right)$$

Denote $X=e^{t(A-B)}u$, and we need $\Psi\left(e^{t(A-B)}\left(\dfrac{1-s}{As-B}\right)\right)=\Psi(X)$

$$\begin{aligned}
\Psi(X)&=\frac{\tilde{A}}{A}m\ln\left[A\left(\frac{XB+1}{XA+1}\right)-B\right]+k'\ln\left[\frac{XB+1}{XB-1}\right]\\
&=\frac{\tilde{A}}{A}m\ln(As-B)+\ln f\qquad\text{from Eq. (??)}\\
\Leftrightarrow\ln f&=\frac{\tilde{A}}{A}m\ln\left[A\left(\frac{XB+1}{XA+1}\right)-B\right]-\frac{\tilde{A}}{A}m\ln(As-B)+k'\ln\left[\frac{XB+1}{XB-1}\right]\\
&=\ln\left[A\left(\frac{XB+1}{XA+1}\right)-B\right]^{\frac{\tilde{A}}{A}m}+\ln(As-B)^{-\frac{\tilde{A}}{A}m}+\ln\left[\frac{XB+1}{XB-1}\right]^{k'}\\
&=\ln\left[\left(A\left(\frac{XB+1}{XA+1}\right)-B\right)^{\frac{\tilde{A}}{A}m}(As-B)^{-\frac{\tilde{A}}{A}m}\left[\frac{XB+1}{XB-1}\right]^{k'}\right]\\
\Leftrightarrow f&=\left(A\left(\frac{XB+1}{XA+1}\right)-B\right)^{\frac{\tilde{A}}{A}m}(As-B)^{-\frac{\tilde{A}}{A}m}\left[\frac{XB+1}{XB-1}\right]^{k'}
\end{aligned}$$

□

To show Eq. (**??**), we start with

$$\begin{aligned} f &= \left[s+(1-s)e^{At}\right]^{-k}\left[1+\frac{1-s}{s}e^{At}\right]^{-k'} \\ &= \left[s+(1-s)e^{At}\right]^{-k}\left[s+(1-s)e^{At}\right]^{-k'}s^{k'} \\ &= \left[s+(1-s)e^{At}\right]^{-(k+k')}s^{k'} \\ &= \left[se^{-At}+(1-s)\right]^{-(k+k')}e^{-At(k+k')}s^{k'} \\ &= \frac{1}{\left[se^{-At}+(1-s)\right]^{(k+k')}}e^{-At(k+k')}s^{k'} \\ &= \frac{1}{\left[1-s(1-e^{-At})\right]^{(k+k')}}e^{-At(k+k')}s^{k'} \end{aligned} \tag{C.1}$$

and use the expansion $\frac{1}{(1-z)^{n+1}} = \sum_{m\geq 0}\binom{n+m}{n}$, where $\binom{n+m}{n} = \frac{(n+m)!}{n!m!}$, to rewrite

$$\begin{aligned} f &= \left[\sum_{m\geq 0}\binom{k+k'-1+m}{k+k'-1}s^m(1-e^{-At})^m\right]e^{-At(k+k')}s^{k'} \\ &= \sum_{m\geq 0}\binom{k+k'-1+m}{k+k'-1}e^{-At(k+k')}(1-e^{-At})^m s^{m+k'} \end{aligned}$$

and denote $n = m + k'$ to rewrite

$$\begin{aligned} f &= \sum_{n-k'\geq 0}\binom{k+n-1}{k+k'-1}e^{-At(k+k')}(1-e^{-At})^{n-k'}s^n \\ &= \sum_{n\geq k'}\frac{(k+n-1)!}{(k+k'-1)!(n-k')!}\left(\frac{k+k'}{\overline{n}+k}\right)^{k+k'}\left(\frac{\overline{n}-k'}{\overline{n}+k}\right)s^n \\ &= \sum_{n\geq k'}\frac{(k+n-1)!}{(k+k'-1)!(n-k')!}\left(\frac{k+k'}{\overline{n}+k}\right)^{k+k'}\left(\frac{\overline{n}-k'}{\overline{n}+k}\right)s^n \end{aligned}$$

with $\overline{n} = (k' + k)e^{At} - k$. And comparing this with Eq. (**??**), we have

$$P_n = \frac{(n+k-1)!}{(n-k')!(k'+k-1)!} \left(\frac{\overline{n}-k'}{\overline{n}+k}\right)^{n-k'} \left(\frac{k+k'}{\overline{n}+k}\right)^{k+k'} \tag{C.2}$$

□

Appendix D

The Gamma Function

The gamma function is an extension of the factorial function, with its argument shifted down by 1, to real and complex numbers.

$$\Gamma(n) = (n-1)! \tag{D.1}$$

The gamma function is defined for all complex numbers except the negative integers and zero. For complex numbers with a positive real part, it is defined via a convergent improper integral:

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt \tag{D.2}$$

The numerical computation of the value of $\Gamma(x)$ in our work has a computational limit such that it will give a missing value if the value of x is too large (about 163.264). The algorithm used to compute $\Gamma(x)$ in this work is the s14aac (nag_gamma) function [**?**] that is based on a Chebyshev expansion for $\Gamma(1+u)$ and uses the property $\Gamma(1+x) = x\Gamma(x)$.

Appendix E

OriginPro 9.0 and Theory of Non-Linear Curve Fitting

Our fitting of the multiplicity distribution data sets were done using the OriginPro 9.0 software. OriginPro uses Origin C, an ANSI C compatible programming language native to Origin, which also utilizes elements of C++ and C. Origin C is used in our work to create the two component GMD fitting function and used with Origin's curve fitting module, NLFit. Also, Origin C is used to call computational routines from the NAG Library, which in our case where our computation of the Gamma function comes from. Below is the code used taken from the Code Builder in Origin Pro 9.0.

```
Calling computational routines from the NAG Library#pragma
    numlittype(push, TRUE)
void _nlsfWeightedTwoGMD(
// Fit Parameter(s):
double weight, double asoft, double ksoft, double kprimesoft, double
     nbarsoft, double ksemihard,
double kprimesemihard, double nbarsemihard,
// Independent Variable(s):
```

```
double n,
// Dependent Variable(s):
double& Pgmd)
{
  // Beginning of editable part
  Pgmd = weight
  *(
  asoft
  *((gamma(n+ksoft))/((gamma(n - kprimesoft + 1))*(gamma(kprimesoft
   + ksoft))))
  *((nbarsoft - kprimesoft)/(nbarsoft + ksoft))^(n - kprimesoft)
  *((ksoft + kprimesoft)/(nbarsoft + ksoft))^(ksoft + kprimesoft)
  + (1 - asoft)
  *((gamma(n+ksemihard))/((gamma(n - kprimesemihard + 1))*(gamma(
   kprimesemihard + ksemihard))))
  *((nbarsemihard - kprimesemihard)/(nbarsemihard + ksemihard))^(n -
    kprimesemihard)
  *((ksemihard + kprimesemihard)/(nbarsemihard + ksemihard))^(
   ksemihard + kprimesemihard)
  )
  // End of editable part
}
```

Listing E.1: Origin C Code used for our two component GMD model

We will illustrate here briefly the theory of non-linear curve fitting. A general non-linear model can be expressed with the following equation

$$\mathbf{Y} = f(\mathbf{X}, \theta) + \boldsymbol{\epsilon} \tag{E.1}$$

where $\mathbf{Y} = (y_1, y_2, \ldots, y_i)$ represents the data to be modelled, $\mathbf{X} = (x_1, x_2, \ldots, x_k)$ are independent variables, $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_p)$ are model parameters and $\boldsymbol{\epsilon}$ are the residuals or errors.

The aim of non-linear fitting is to estimate the parameter values which best describe the data. The standard way of finding the best fit is to choose the parameter values that minimize the residuals, i.e., the deviations of the theoretical curve(s) from the experimental data points. This method, as mentioned in Chapter **??** is also called Chi-square χ^2 minimization, with χ^2 more generally defined as follows

$$\chi^2(\boldsymbol{\theta}) = \sum_{i=1}^{n} \left[\frac{y_i - f(x_i, \boldsymbol{\theta})}{\sigma_i} \right]^2 \tag{E.2}$$

where x_i is the row vector for the ith $(i = 1, 2, \ldots, n)$ observation and σ_i is the uncertainty in y_i.

To estimate $\boldsymbol{\theta}$ with the least squares method, we need to solve the normal equations, i.e., the equations that minimize the residuals. We do this by setting the partial derivatives of χ^2 with respect to each θ_p to zero

$$\frac{\partial \chi^2}{\partial \boldsymbol{\theta}} = -2 \sum_{i=1}^{n} \frac{1}{\sigma_i^2} \left[y_i - f(x_i, \boldsymbol{\theta}) \right] \left[\frac{\partial f(x_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right] = 0 \tag{E.3}$$

Since there are no explicit solutions to the normal equations, we employ an iterative strategy to estimate the parameter values. This process starts with some initial values, $\boldsymbol{\theta}_0$. With each iteration, a χ^2 value is computed and then the parameter values are adjusted so as to reduce the χ^2. When the χ^2 values computed in two successive iterations are small enough (compared with the tolerance), it is said that the fitting procedure has converged.

Origin uses the Levenberg-Marquardt algorithm to adjust the parameter values in the iterative procedure. This algorithm, which combines the Gauss-Newton method and the

steepest descent method, works for most cases. To start a minimization, we have to first provide an initial guess for the parameter vector, $\boldsymbol{\theta}_0$. In cases with only one minimum, an uninformed standard guess like $\boldsymbol{\theta}_0^T = (1, 1, ..., 1)$ will work fine. However, in cases with multiple minima, the algorithm converges only if the initial guess is already somewhat close to the final solution.

In each iteration step, the parameter vector $\boldsymbol{\theta}$ is replaced by a new estimate, $\boldsymbol{\theta} + \boldsymbol{\delta}$. To determine $\boldsymbol{\delta}$, the functions $f(x_i, \boldsymbol{\theta} + \boldsymbol{\delta})$ are approximated by their linearizations

$$f(x_i, \boldsymbol{\theta} + \boldsymbol{\delta}) \approx f(x_i, \boldsymbol{\theta}) + J_i \boldsymbol{\delta} \tag{E.4}$$

where

$$J_i = \frac{\partial f(x_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \tag{E.5}$$

is the gradient (in this case, row-vector) of f with respect to $\boldsymbol{\theta}$.

At the minimum of $\chi^2(\boldsymbol{\theta})$, the gradient of χ^2 with respect to $\boldsymbol{\delta}$ will be zero. The above first order approximation of $f(x_i, \boldsymbol{\theta} + \boldsymbol{\delta})$ gives

$$\chi^2(\boldsymbol{\theta} + \boldsymbol{\delta}) \approx \sum_{i=1}^{n} \frac{1}{\sigma_i^2} \left(y_i - f(x_i, \boldsymbol{\theta}) - J_i \boldsymbol{\delta}\right)^2 \tag{E.6}$$

Or in vector notation,

$$\chi^2(\boldsymbol{\theta} + \boldsymbol{\delta}) \approx \left| \frac{\mathbf{Y} - \mathbf{f}(\boldsymbol{\theta}) - \mathbf{J}\boldsymbol{\delta}}{\boldsymbol{\sigma}^2} \right|^2 \tag{E.7}$$

Taking the derivative with respect to $\boldsymbol{\delta}$ and setting the result to zero gives

$$(\mathbf{J^T J})\boldsymbol{\delta} = \mathbf{J^T} \left[\mathbf{Y} - \mathbf{f}(\boldsymbol{\theta})\right] \tag{E.8}$$

where $\mathbf{J}$ is the Jacobian matrix whose ith row equals J_i, and where $\mathbf{f}$ and $\mathbf{Y}$ are vectors with ith component $f(x_i, \boldsymbol{\theta})$ and y_i respectively. This is a set of linear equations which can be solved for $\boldsymbol{\delta}$.

Levenberg replaced this equation by a "damped version",

$$(\mathbf{J^T J} + \lambda \mathbf{I})\boldsymbol{\delta} = \mathbf{J^T}\left[\mathbf{Y} - \mathbf{f}(\boldsymbol{\theta})\right] \tag{E.9}$$

where $\mathbf{I}$ is the identity matrix, giving as the increment, $\boldsymbol{\delta}$, to the estimated parameter vector $\boldsymbol{\theta}$.

The (non-negative) damping factor, λ, is adjusted at each iteration. If reduction of χ^2 is rapid, a smaller value can be used, bringing the algorithm closer to the Gauss-Newton algorithm, whereas if an iteration gives insufficient reduction in the residual, λ can be increased, giving a step closer to the gradient descent direction. Note that the gradient of χ^2 with respect to $\boldsymbol{\theta}$ equals $-2(\mathbf{J^T}\left[\mathbf{Y} - \mathbf{f}\left(\boldsymbol{\beta}\right)\right])^{\mathbf{T}}$. Therefore, for large values of λ, the step will be taken approximately in the direction of the gradient. If either the length of the calculated step, $\boldsymbol{\delta}$, or the reduction of sum of squares from the latest parameter vector, $\boldsymbol{\theta} + \boldsymbol{\delta}$, fall below predefined limits, iteration stops and the last parameter vector, $\boldsymbol{\theta}$, is considered to be the solution.

Levenberg's algorithm has the disadvantage that if the value of damping factor, λ, is large, inverting $\mathbf{J^T J} + \lambda \mathbf{I}$ is not used at all. Marquardt provided the insight that we can scale each component of the gradient according to the curvature so that there is larger movement along the directions where the gradient is smaller. This avoids slow convergence in the direction of small gradient. Therefore, Marquardt replaced the identity matrix, $\mathbf{I}$, with the diagonal matrix consisting of the diagonal elements of $\mathbf{J^T J}$, resulting

in the Levenberg-Marquardt algorithm

$$(\mathbf{J^T J} + \lambda\, \mathbf{diag}(\mathbf{J^T J}))\boldsymbol{\delta} = \mathbf{J^T}\,[\mathbf{Y} - \mathbf{f}(\boldsymbol{\theta})] \tag{E.10}$$

The Levenberg-Marquardt algorithm is a very popular curve-fitting algorithm used in many software applications for solving generic curve-fitting problems. However, as for many fitting algorithms, the Levenberg-Marquardt algorithm finds only a local minimum, which is not necessarily the global minimum.

Bibliography

[1] W. Kittel and E. A. De Wolf, *Soft Multihadron Dynamics* (World Scientific, 2005).

[2] J. F. Grosse-Oetringhaus and K. Reygers, Journal of Physics G: Nuclear and Particle Physics **37**, 083001 (2010).

[3] M. Althoff *et al.*, Zeitschrift fur Physik C **22**, 307 (1984).

[4] A. Breakstone *et al.*, Physical Review D **30**, 528 (1984).

[5] K. Konishi, A. Ukawa, and G. Veneziano, Nuclear Physics B **157**, 45 (1979).

[6] A. Giovannini, Nuclear Physics B **161**, 429 (1979).

[7] G. Alner *et al.*, Physics Letters B **160**, 193 (1985).

[8] R. E. Ansorge *et al.*, Zeitschrift für Physik C Particles and Fields **43**, 357 (1989).

[9] C. S. Lindsey *et al.*, Nuclear Physics A **544**, 343 (1992).

[10] C. Fuglesang, UA5 multiplicity distributions and fits of various functions, in *to appear in Proceedings of Multiparticle Dynamics*, p. 19, 1989.

[11] A. Giovannini and R. Ugoccioni, Nuclear Physics B-Proceedings Supplements **71**, 201 (1999).

[12] C. Chew, D. Kiang, and H. Zhou, Physics Letters B **186**, 411 (1987).

[13] A. Chan and C. Chew, Zeitschrift für Physik C Particles and Fields **55**, 503 (1992).

[14] A. Chan and C. Chew, Il Nuovo Cimento A **101**, 409 (1989).

[15] A. Chan and C. Chew, Physical Review D **41**, 851 (1990).

[16] A. Dewanto, A. Chan, and C. Oh, International Journal of Modern Physics E **16**, 3295 (2007).

[17] S. Chatrchyan *et al.*, Jinst **3**, S08004 (2008).

[18] A. Giovannini and R. Ugoccioni, Physical Review D **68**, 034009 (2003).

[19] I. Zborovský, Journal of Physics G: Nuclear and Particle Physics **40**, 055005 (2013).

[20] A. Wroblewski, Soft hadron physics, in *Proceedings of the 25th International Conference on High Energy Physics, Singapore, 1990*, pp. 125–143, 1990.

[21] P. Chliapnikov and O. Tchikilev, Physics Letters B **282**, 471 (1992).

[22] S. Krasznovszky and I. Wagner, Physics Letters B **228**, 159 (1989).

[23] R. Szwed, G. Wrochna, and A. K. Wróblewski, Modern Physics Letters A **5**, 1851 (1990).

[24] I. Dremin and V. Nechitailo, Physical Review D **70**, 034005 (2004).

[25] A. Kaidalov and M. Poghosyan, The European Physical Journal C **67**, 397 (2010).

[26] F. Gelis, T. Lappi, and L. McLerran, Nuclear Physics A **828**, 149 (2009).

[27] P. Ghosh, Physical Review D **85**, 054017 (2012).

[28] I. Dremin and V. Nechitailo, Physical Review D **84**, 034026 (2011).

[29] L. Evans and P. Bryant, Journal of Instrumentation **3**, S08001 (2008).

[30] Z. Koba, H. Nielsen, and P. Olesen, Nuclear Physics B **43**, 125 (1972).

[31] R. C. Hwa and C. Lam, Physics Letters B **173**, 346 (1986).

[32] P. Ghosh, Physics Letters B **705**, 356 (2011).

[33] A. Giovannini and R. Ugoccioni, Physical Review D **59**, 094020 (1999).

[34] J. Welti, (2011).

[35] O. Lab, NAG library function document: nag_gamma (s14aac), 2009.

List of Figures

List of Tables