Charged-Particle Multiplicity in Proton–Proton Collisions

Jan Fiete Grosse-Oetringhaus\textsuperscript{1}, Klaus Reygers\textsuperscript{2}

\textsuperscript{1}CERN, 1211 Geneva 23, Switzerland
\textsuperscript{2}Physikalisches Institut, Universität Heidelberg, Philosophenweg 12, 69120 Heidelberg, Germany

Abstract.
# Charged-Particle Multiplicity in Proton–Proton Collisions

## Contents

1. **Introduction**  
   1.1 Brief Overview of Multiplicity Measurements  
   1.2 Structure of the Review

2. **Multiplicity Distributions: Basic Theoretical Concepts**  
   2.1 Feynman Scaling  
   2.2 Moments  
   2.3 Koba–Nielsen–Olesen (KNO) Scaling  
   2.4 Negative Binomial Distributions  
   2.5 Two-Component Approaches  
   2.5.1 Combination of two NBDs  
   2.5.2 Interpretation in the Framework of Multiple-Parton Interactions  
   2.6 Similarities between $p + p$ and $e^+ e^-$ collisions and QCD predictions

3. **Multiplicity Distributions: Measurements**  
   3.1 Analysis Techniques  
   3.1.1 Event Classes  
   3.1.2 Unfolding of Multiplicity Distributions  
   3.2 Data Sample  
   3.3 Multiplicity Distributions from $\sqrt{s} = 20$ to 1800 GeV  
   3.4 $dN_{ch}/d\eta$ and $\langle N_{ch} \rangle$ vs. $\sqrt{s}$  
   3.5 Universality of Multiplicities in $p + p$ and $e^+ e^-$  
   3.6 Moments  
   3.7 NBD Parameters $\langle N_{ch} \rangle$ and $k^{-1}$  
   3.8 Two NBD Fit  
   3.9 Open Experimental Issues

4. **Predictions**

5. **Summary**

## Appendices

- **Appendix A**: Derivation of Feynman scaling
- **Appendix B**: Relation of NBD and BD
- **Appendix C**: Uncertainties on Moments
1. Introduction

The charged-particle multiplicity belongs to the simplest observables in collisions of hadrons, yet it provides important insights into the mechanisms of particle production. Experiments have been performed with cosmic rays, fixed target setups, and particle colliders. These data have been used to improve, or reject, models of particle production which are often available as Monte Carlo event generators. Just considering the number of produced charged particles is a drastic reduction of the complex information contained in a particle collision, especially the kinematic properties are neglected. Nevertheless, the multiplicity distribution, i.e., the probability distribution of obtaining a definite number of produced particle, still contains some information about particle correlations. By definition all information about correlations is removed when the information is reduced to the average charged particle multiplicity $\langle N_{ch} \rangle$. Distributions that still consider partly the kinematics are, e.g., the pseudorapidity density and the transverse momentum distribution which are one-dimensional projections of the kinematic properties. More sophisticated is the study the correlations of the different particles, a typically studied property which is still on a quite global level is the dependence of the average $p_T$ on the event multiplicity. Deeper correlation studies are in the realm of Hanbury-Brown and Twiss (HBT) interferometry where many-particle correlations as function of the particles’ momenta are studied.

This review focusses on the charged-particle multiplicity distribution and the pseudorapidity density. Correlations are only briefly discussed by studying the moments of the distributions. The main topics of this review are basic theoretical concepts and their applicability to data, the experimental challenges, experimental results, and predictions for the Large Hadron Collider (LHC) energies. The objective of this review is to give a general overview of the field, to discuss the relevant theoretical aspects without going into depth but giving references for the reader who wants to study certain topics deeper. By introducing collider experiments and their results the reader should obtain an understanding where the limitations of these theoretical descriptions are and what the experimental trends as function of center-of-mass energy are. Furthermore, an objective is to discuss the open experimental items to which data from the LHC can bring the needed clarifications. Generally, the study of the multiplicity of particles is an essential topic at the beginning of data-taking at the LHC. The characterization of the underlying event is a precondition for most of the flagship research topics of the LHC.

1.1. Brief Overview of Multiplicity Measurements

The charged-particle multiplicity is a key observable for the understanding of multi-particle production in collisions of hadrons at high-energy. The probability $P(n)$ for producing $n$ charged particles in the final state is related to the production mechanism of the particles. For a particle source without any correlations the multiplicity distribution follows a Poisson distribution and the dispersion $D = \sqrt{\langle n^2 \rangle - \langle n \rangle^2}$ would be related to the average multiplicity as $D = \sqrt{\langle n \rangle}$. Any kind of correlation manifests itself as a
Charged-Particle Multiplicity in Proton–Proton Collisions

deviation from a Poissonian. The measurement of multiplicity distributions provides significant constraints for particle-production models. However, the discrimination between models typically requires additional more differential measurements. Particle production models are typically based on Quantum Chromodynamics, however, they necessarily have a phenomenological component as the formation of particles involves a soft scale which limits the use of perturbative techniques.

Early measurements of multiplicity distributions in $e^+e^-$ collisions at $\sqrt{s} = 29\,\text{GeV}$ could approximately be described with a Poissonian distribution \cite{1, 2}. $p+p$ collisions, on the other hand, exhibited broader multiplicity distributions. The energy dependence of the dispersion in non-single diffractive $p+p$ collisions could approximately be described as $D \propto \langle n \rangle$ up to the maximum ISR energy of $\sqrt{s} = 62\,\text{GeV}$ \cite{3}. A simple interpretation was that the correlations in $p+p$ were related to superpositions of collisions with different impact parameters \cite{4}.

Interest in multiplicity distributions was stimulated by a paper of Koba, Nielsen, and Olesen in 1972 in which they derived theoretically that multiplicity distributions should follow a universal scaling at high energies (KNO scaling). With $z = n/\langle n \rangle$ the function $\Psi(z) = \langle n \rangle P(n)$ was expected to asymptotically reach a universal energy-independent form. KNO scaling was derived based on Feynman scaling, i.e., based on the assumption that the rapidity density $dN_{ch}/dy$ reaches a limiting value at $y = 0$ above a certain energy which corresponds to an asymptotic scaling of the total multiplicity as $\langle n \rangle \propto \ln \sqrt{s}$. Bubble chamber data between $\sqrt{s} \approx 6\,\text{GeV}$ and $24\,\text{GeV}$ indicated an onset of KNO scaling already at $\sqrt{s} \approx 10\,\text{GeV}$ \cite{5}. At the ISR among other observations the relation $D \propto \langle n \rangle$ implied that KNO scaling was satisfied \cite{3} (although deviations were noted in \cite{6}). It was found at the ISR that the average multiplicity increased faster with energy than $\ln \sqrt{s}$. Thus, the theoretical basis for KNO scaling appeared to be gone. In 1985, breaking of KNO scaling was observed by the UA5 collaboration in $p+\bar{p}$ collisions at $\sqrt{s} = 540\,\text{GeV}$ \cite{7}. In a later publication UA5 concluded that KNO scaling was already violated at $\sqrt{s} = 200\,\text{GeV}$ \cite{8}.

UA5 found that the multiplicity distributions up to $\sqrt{s} = 540\,\text{GeV}$ can be well described by a negative binomial distribution (NBD) \cite{7} which is determined by two parameters $\langle n \rangle$ and $k$. The parameter $k$ determines the width. It was found that $1/k$ increases approximately linearly with $\ln \sqrt{s}$ whereas KNO scaling corresponds to a constant, energy-independent $1/k$. However, deviations from the NBD were discovered by UA5 at $\sqrt{s} = 900\,\text{GeV}$ and later confirmed at the Tevatron at $\sqrt{s} = 1800\,\text{GeV}$ \cite{9}. A shoulder structure appeared at $n \approx 2\langle n \rangle$ which could not be described with a single NBD. This led to a two-component model by Giovannini and Ugoccioni in 1999 who described the measured data by the combination of two NBDs, interpreting one as a soft and one as a semi-hard component. An alternative description interprets the results in favor of multiple-parton interactions which become more important at higher energies.

The superposition of several interactions has influence on the multiplicity distribution and therefore explains the deviation from the scaling found at lower energies.

Multiplicity measurements in $e^+e^-$ between $10 \lesssim \sqrt{s} \lesssim 91.2\,\text{GeV}$ showed that also
in this system the dispersion $D$ scaled approximately linearly with $\langle n \rangle$ and that KNO scaling was approximately satisfied \cite{10}. Thus, the Poissonian shape at $\sqrt{s} \approx 30$ GeV was only accidental. Also in $e^+e^-$ the NBD provided a useful description of the data. The Delphi experiment found that at $\sqrt{s} = 91.2$ GeV multiplicity distributions in restricted rapidity intervals exhibited a shoulder structure similar to the observations in $p + p(\bar{p})$ \cite{11}. This shoulder was attributed to three- and four-jet events which have larger average multiplicities than two-jet events.

1.2. Structure of the Review

The review is divided into two parts. The first part introduces basic theoretical concepts, the second part concentrates on experimental data.

The first part discusses scaling properties of the multiplicity, i.e., Feynman and KNO scaling. We recall the definitions of various moments that are used in this review. Furthermore, negative binomial distributions (NBDs) and two component models are discussed. Among the latter is one combining two NBDs identifying one component with soft and the other with semi-hard interactions; a different one identifies single and multi-parton interactions in the distribution. Similarities between $p + p$ and $e^+e^-$ collisions are investigated in the context of QCD.

The second part starts with the introduction of important aspects of the multiplicity analysis. It is motivated why measured multiplicity distributions need to be unfolded to obtain the originating (true) distribution. Subsequently, references of a representative data sample in the center-of-mass energy range from 23.6 GeV to 1.8 TeV are given. Their analysis methods and error treatments are discussed. Out of these selected results are shown of the pseudorapidity density and the multiplicity distribution and their consistencies with the theoretical descriptions introduced in the first part are assessed. The dependence of the multiplicity on the collision energy is analyzed. Results from hadron and lepton colliders are compared and their universalities and differences investigated. The behavior of the moments of the multiplicity distribution as a function of $\sqrt{s}$ are studied. Then we investigate how NBDs and the combination of two NBDs describe the distributions. The part concludes with a discussion of open experimental issues.

An overview of available predictions for the LHC energy range is given in the final section of the review.

2. Multiplicity Distributions: Basic Theoretical Concepts

Before introducing various analytical descriptions of the multiplicity distribution, it is important to remark that in case the underlying production process can be described by uncorrelated emission, i.e., the production of an additional particle is independent from the already produced particles, the multiplicity distribution is expected to be of Poissonian form. Any difference to this, indicates correlations between the produced
Charged-Particle Multiplicity in Proton–Proton Collisions

6

particles. Forward–backward correlations have in fact been measured, e.g., by UA5 in p+\bar{p} collisions \cite{12} but are not further discussed here.

Many authors have tried to identify simple analytical forms that reproduce the multiplicity distributions at different \sqrt{s} requiring only a simple rescaling or change of a few parameters as function of \sqrt{s}. Considering the wealth of processes that are expected at large \sqrt{s} one may ask the question if such an approach can be successful at all. At LHC energies a regime is reached where the average collision contains multiple parton interactions whose products might undergo final-state interactions. Not to mention parton branchings and fragmentation to the final particles whose existence is well-established. As unsatisfying as it sounds we may face an energy regime where MC programs that consider all known effects are more likely to reproduce collisions than a much more appealing because simple approach using analytical forms.

2.1. Feynman Scaling

Feynman derived that the mean total number of any kind of particle rises logarithmically with \sqrt{s} \cite{13}:

\[ \langle N \rangle \propto \ln W \propto \ln \sqrt{s} \quad \text{with} \quad W = \sqrt{s}/2. \] (1)

Based on phenomenological arguments about the exchange of quantum numbers between the colliding particles, he concludes that the number of particles with a given mass and transverse momentum per \( dp_z \) interval depends on the energy \( E = E(p_z) \) as:

\[ \frac{dN}{dp_z} \sim \frac{1}{E} \] (2)

This is extended to the probability of finding a particle of kind \( i \) with mass \( m \) and transverse and longitudinal momentum \( p_T \) and \( p_z \):

\[ f_i(p_T, p_z/W) \frac{dp_z}{E} d^2p_T \] (3)

with the energy of the particle:

\[ E = \sqrt{m^2 + p_T^2 + p_z^2}. \] (4)

\( f_i(p_T, x = p_z/W) \) is a function describing the distribution of the particles and Feynman’s hypothesis is that \( f_i \) becomes independent of \( W \) at high energies. This assumption is the Feynman scaling and \( f_i \) is called scaling function or Feynman function. \( x = p_z/W \) is the fraction of the longitudinal momentum of the particle \( p_z \) and the total energy of the incident particle \( W \). \( x \) is now called Feynman-\( x \). Integrating Eq. (3) results in \( \langle N \rangle \propto \ln W \). A derivation is given in Appendix A.

Considering that the maximum reachable rapidity in a collisions increases also with \ln \sqrt{s}, it follows that:

\[ \frac{dN}{dy} = \text{const.} \] (5)

The height of the rapidity distribution around mid-rapidity, the so-called plateau, is independent of \sqrt{s}. Equivalently, the pseudorapidity at mid-rapidity \( dN/d\eta|_{\eta=0} \) is...
approximately constant when Feynman scaling applies. Here the transformation from $y$ to $\eta$ has to be taken into account. It depends on the average $m_T$ which, however, is only weakly energy-dependent (a rough estimate of the change in the transformation factor is $1 - 2\%$ from $\sqrt{s} = 100$ GeV to 1 TeV). Furthermore, this transformation causes a dip in the distribution around $\eta \approx 0$ which is not present in the rapidity distribution itself (see Section 3.4 where measured $dN_{ch}/d\eta$ distributions are shown).

2.2. Moments

Multiplicity distributions are functions with a non-trivial shape. To assess the behavior, e.g., as a function of $\sqrt{s}$, it is convenient to study the moments of the multiplicity distribution. All moments together contain the information of the full distribution. However, in practice only the first few moments can be calculated due to the limited statistics. Moments that will be presented in this review are $C$, $F$, and $D$-moments.

The reduced $C$-moments are defined by:

$$C_q = \frac{\langle n^q \rangle}{\langle n \rangle^q}.$$  

The normalized factorial $F$-moments are defined by:

$$F_q = \frac{\langle n(n - 1) \ldots (n - q + 1) \rangle}{\langle n \rangle^q}.$$  

It can be shown that these moments contain in integrated form the correlations of the system of particles (see e.g. [14]). For a Poisson distribution one obtains $F_q = 1$ for all $q$. The $D$-moments are defined by:

$$D_q = \langle (n - \langle n \rangle)^q \rangle^{1/q}. $$  

$D_2 \equiv D$ is referred to as dispersion.

The equations that are used to calculate the uncertainties on these moments are given in Appendix C. Further moments of interest are the $K$-cumulants and the $H$-moments:

$$K_q = \frac{1}{\langle n \rangle^q} \frac{d^q \ln G(z)}{dz^q} \bigg|_{z=0}$$

$$H_q = \frac{K_q}{F_q}$$

where $G$ is the so-called generating function of the multiplicity distribution, that replaces the set of probabilities at a given energy by an analytical function:

$$G(z) = \sum_{n=0}^{\infty} P(n)(1 + z)^n.$$  

The $H$-moments are interesting to study because higher-order QCD calculations predict that these oscillate as a function of the rank $q$ [15]. $H$-moments are only briefly discussed in the following. For more details about the definitions of the moments see, e.g., [14].
2.3. Koba-Nielsen-Olesen (KNO) Scaling

KNO scaling was suggested in 1972 by Koba, Nielsen, and Olesen \[16\]. Their main assumption is Feynman scaling.

KNO scaling is derived by calculating
\[
\langle n(n-1)\ldots(n-q+1)\rangle = \int f^{(q)}(x_1,p_{T,1};\ldots;x_q,p_{T,q}) \frac{dp_{z,1}}{E_1} \frac{dp_{z,2}}{E_2} \cdots \frac{dp_{z,q}}{E_q} \frac{dp_{p_{T,q}}}{E_T} \tag{12}
\]
which is an extension of the expression used in the derivation of Feynman scaling (see Eq. (A.4) on page 43) but using a function \( f^{(q)}(q) \) that describes \( q \)-particle correlations \( (q \) particles with energy \( E_q \), longitudinal momentum \( p_{z,q} \), transverse momentum \( p_{T,q} \), and Feynman-x \( x_q \)). Integration by parts is performed for all \( x_i \) and it is proven that the resulting function is uniquely defined by moments. This yields a polynomial in \( \ln s \).

With a substitution of the form \( \Psi(n) = \frac{1}{n} \Psi\left(\frac{n}{\langle n \rangle}\right) + O\left(\frac{1}{(\langle n \rangle)^2}\right) \) \( \tag{13} \)
where the first term results from the leading term in \( \ln s \), that is \( (\ln s)^q \). The second term contains all other terms in \( \ln s \), i.e., \( (\ln s)^q \) for \( q' < q \). \( \Psi(z := n/\langle n \rangle) \) is a universal, i.e., energy-independent function. This means that multiplicity distributions at all energies fall onto one curve when plotted as a function of \( z \). However, \( \Psi(z) \) can be different depending on the type of reaction and the type of measured particles.

The moments \( C_q \) define \( \Psi(z) \) uniquely \[16\]:
\[
C_q = \int_0^\infty z^q \Psi(z) dz. \tag{14}
\]
Substituting \( z = n/\langle n \rangle \) results in Eq. (14).

Studying the moments of the distribution shows if the scaling hypothesis holds; in this case the moments are independent of energy. For example an experimentally accessible possibility is to calculate \( D_2^2 = \langle n^2 \rangle - \langle n \rangle^2 \); the relation \( D_2/\langle n \rangle = \text{const.} \) follows from Eq. (13) (if \( \Psi(z) \) is not a \( \delta \) function, see \[16\]).

It has been pointed out \[17\] that the conclusion that the multiplicity distribution follows a universal function is only an approximation (neglecting the second term in Eq. (13)). Therefore the exact result is that the factorial moments (see Eq. (7)) are required to be constant, not the reduced moments (which follow from Eq. (14)). This is addressed further in Section 3.6.

The description of discrete data points with a continuous function in Eq. (13) is an approximation valid for \( \langle n \rangle \gg 1 \). A generalized KNO scaling which avoids this problem is described in \[18, 19\]. Moreover, different scaling laws for multiplicity distributions were proposed (see e.g. \[14, 20\]), among those the so-called log-KNO scaling \[21\] which predicts a scaling of the form
\[
P(n) = \frac{1}{\lambda(s)} e^{\frac{\ln n + c(s)}{\lambda(s)}} \tag{15}
\]
with energy-dependent functions $\lambda(s)$ and $c(s)$ which correspond to $\langle n \rangle$ and the multiplicity related to the leading particles, respectively.

### 2.4. Negative Binomial Distributions

The **Negative Binomial Distribution (NBD)** is defined as

$$P_{n,k}^{NBD}(n) = \binom{n + k - 1}{n} (1 - p)^n p^k.$$  \hfill (16)

It gives the probability for $n$ failures before $k$ successes in a Bernoulli experiment with a success probability $p$ (the failures and the first $k - 1$ successes can be in any order). The NBD is a Poisson distribution for $k^{-1} \rightarrow 0$ and a geometrical distribution for $k = 1$. For negative integer $k$ the distribution is a binomial distribution (for a proof see Appendix B). The continuation to negative integer $k$ is performed by writing the binomial in terms of the $\Gamma$ function and using the equation $\Gamma(x + 1) = x\Gamma(x)$:

$$\binom{n + k - 1}{n} = \frac{(n + k - 1)!}{n!(k - 1)!} = \frac{\Gamma(n + k)}{\Gamma(n + 1)\Gamma(k)} = \frac{(n + k - 1) \cdot (n + k - 2) \cdot \ldots \cdot k}{\Gamma(n + 1)}. \hfill (17)$$

The mean of the distribution $\langle n \rangle$ is related to $p$ by $p^{-1} = 1 + \langle n \rangle/k$. This leads to the form of the NBD that is used to describe multiplicity distributions [22, 23]:

$$P_{\langle n \rangle,k}^{NBD}(n) = \binom{n + k - 1}{n} \left(\frac{\langle n \rangle/k}{1 + \langle n \rangle/k}\right)^n \frac{1}{(1 + \langle n \rangle/k)^k}. \hfill (18)$$

Figure 1 shows normalized NBDs with different sets of parameters. Also the case of a large $k$ where the distribution approaches a Poissonian is shown and the case with
a negative $k$ where the function gets binomial. $P_{(n),k}^{\text{NBD}}(n)$ follows KNO scaling if $k$ is constant (energy-independent). This can be seen from the KNO form

$$
\Psi_{\text{NBD}}(z) = \frac{k^k}{\Gamma(k)} z^{k-1} e^{-kz}
$$

which holds in the limit $\langle n \rangle / k \gg 1$, $z = n / \langle n \rangle$ fixed [24]. Therefore, studying $k$ as a function of $\sqrt{s}$, of multiplicity distributions that are described by NBDs, directly shows whether KNO scaling is fulfilled.

The physical origin of a multiplicity distribution following a negative binomial shape has not been ultimately understood. One approach is to use the recurrence relation of collisions of multiplicities $n$ and $n + 1$ [25]. This relation is defined such that for uncorrelated emission it is constant; any departure shows the presence of correlations.

Evaluating

$$
g(n) = \frac{(n + 1)P(n + 1)}{P(n)}
$$

for a Poisson distribution $P(n) = \lambda^n e^{-\lambda} / n!$ (representing uncorrelated emission), yields that $g(n) = \lambda = \text{const.}$ is indeed constant. The term $n + 1$ in Eq. (20) can be understood by considering that the particles are in principle distinguishable, e.g., by their momenta, therefore it has to be taken into account that a collision of multiplicity $n + 1$ can be related to $n + 1$ collisions of multiplicity $n$ (by removing any single one of the $n + 1$ particles).

For NBDs, Eq. (20) can be written as

$$
g(n) = a + bn
$$

with $k = a/b$ and $\langle n \rangle = a / (1 - b)$

or $a = \langle n \rangle k / (\langle n \rangle + k)$ and $b = \langle n \rangle / (\langle n \rangle + k)$.

A model of partially stimulated emission identifies $a$ in Eq. (21) with the production of particles which is independent of the already present particles and $bn$ with emission that is enhanced by already present particles (Bose–Einstein interference). From $g(n) = a(1 + n/k)$ (which follows from Eq. (21)) one sees that $k^{-1}$ is the fraction of the already present particles $n$ stimulating emission of additional particles. Following these rather simple assumptions results in two facts that are found experimentally: 1) $k$ increases when the considered $\eta$-interval is enlarged (because the range of Bose–Einstein interference is finite, the fraction of present particles stimulating further emission reduces); 2) $k$ decreases with increasing $\sqrt{s}$ for a fixed $\eta$-interval (the density of particles in the same interval increases because $\langle n \rangle$ increases) [25].

The multiplicity distribution can be deduced as being of negative binomial shape within the so-called clan model [25, 26, 27]. It describes the underlying production by cascades of particles. In the clan model a particle can emit additional particles, e.g., by decay and fragmentation. A clan (or cluster) contains all particles that stem from the same ancestor. The ancestors are produced independently.

The production of ancestors, and thus clans, is governed by a Poisson distribution $P(N, \langle N \rangle)$ where $\langle N \rangle$ is the average number of produced clans.
The probability to produce \( n_c \) particles in one clan \( F_c(n_c) \) can be concluded from the following considerations: One defines that without particles there is no clan:
\[
F_c(0) = 0
\]  
(22)
and assumes that the production of an additional particle in a clan is proportional to the number of already existing particles with some probability \( p \) (see also Eq. (20)):
\[
\frac{(n_c + 1)F_c(n_c + 1)}{F_c(n_c)} = pn_c.
\]  
(23)
By iteration, the following expression is obtained:
\[
F_c(n_c) = F_c(1) \frac{p^{n_c-1}}{n_c}.
\]  
(24)
The multiplicity distribution that takes into account the distribution of clans and the distribution of particles among the different clans is:
\[
P(n) = \sum_{N=1}^{n} P(N, \langle N \rangle) \sum_{i=1}^{N} F_c(n_1)F_c(n_2)...F_c(n_N),
\]  
(25)
where \( \sum_i \) runs over all combinations \( n_i \) for which \( n = \sum_{i=1}^{N} n_i \) is valid. It can be shown that Eq. (25) is a NBD, identifying \( \langle n \rangle = \langle N \rangle F_c(1)/(1 - p) \) and \( k = \langle N \rangle F_c(1)/p \) [25].

For the case of \( n = 2 \) it can be shown using Eqs. (23) and (25) that \( k \) is the relative probability of obtaining one clan with two particles with respect to obtaining two clans with one particle each.

2.5. Two-Component Approaches

2.5.1. Combination of two NBDs Multiplicity distributions measured by UA5 have been successfully fitted with a combination of two NBD-shaped components [28]. A systematic investigation has been performed by Giovannini and Ugoccioni who interpret the two components as a soft and a semi-hard one [29]. These can be understood as events with and without minijets, respectively (the authors of [29] use a definition from the UA1 collaboration: a minijet is a group of particles having a total transverse energy larger than 5 GeV): the fraction of semi-hard events found corresponds to the fraction of events with minijets seen by UA1. It is important 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.

In this approach, the multiplicity distribution depends on five parameters, that may all be \( \sqrt{s} \) dependent:
\[
P(n) = \alpha_{\text{soft}} \times P_{(n)\text{soft},k_{\text{soft}}}^{\text{NBD}}(n) \\
+ (1 - \alpha_{\text{soft}}) \times P_{(n)\text{semi-hard},k_{\text{semi-hard}}}^{\text{NBD}}(n).
\]  
(26)
The parameters and their dependence on \( \sqrt{s} \) are found by fitting data from experimental measurements. Note that \( \langle n \rangle \) is about two times larger in the semi-hard component than
in the soft component. Furthermore, the fits show that the soft component follows KNO scaling, while the semi-hard component violates KNO scaling. This is discussed in more detail in Section 3.8.

2.5.2. Interpretation in the Framework of Multiple-Parton Interactions  Above CERN ISR energies parton-parton interactions with high momentum transfer (i.e. hard scatterings) are expected to contribute significantly to the total charged-particle multiplicity in \( p + p(\bar{p}) \) collisions [30, 31, 32]. Hard parton-parton scatterings resulting in QCD jets above a lower transverse momentum threshold can be described with perturbative QCD. Softer interactions either require a recipe for regularizing the diverging QCD jet cross section for \( p_T \to 0 \) [31] or models for soft-particle production, see e.g. [32]. The transverse momentum scale \( p_{T,0} \) that controls the transition from soft to hard interactions is typically on the order of 2 GeV/\( c \). In these QCD-inspired models two or more independent hard parton-parton scatterings frequently occur in the same \( p+p(\bar{p}) \) collision [31, 33]. These models explain many observed features of these collisions including the rise of the total inelastic \( p + p(\bar{p}) \) cross section with \( \sqrt{s} \), the increase of \( \langle p_T \rangle \) with the charged particle multiplicity \( N_{ch} \), the increase of \( \langle p_T \rangle \) with \( \sqrt{s} \), and the increase of \( dN_{ch}/d\eta \) with \( \sqrt{s} \). High-multiplicity collisions in these models are collisions with a large number of minijets. In the minijet model of ref. [32] up to 8 independent parton-parton scattering are expected to contribute to the high-multiplicity tail of the multiplicity distribution at \( \sqrt{s} = 1.8 \text{TeV} \). In this model also the violation of KNO scaling is attributed to the onset of minijet production. Strong correlations between multiple parton interactions and the shape of the multiplicity distribution have also been demonstrated for the Pythia event generator [31]. In Pythia the multiplicity distribution turns out to be strongly related to the density profile of the proton [31, 33]. A purely analytic model based on multiple parton interactions is the IPPI model [35]. In this model the multiplicity distribution is a superposition of negative binomial distributions where each NBD represents the contribution of collisions with a given number of parton-parton scatterings.

A data-driven approach to define and identify double parton interactions and thus a second component in the multiplicity distribution is by plotting the multiplicity distribution in KNO variables and subtracting the part of the distribution for which KNO scaling holds [36]. This is done by comparing the distribution to a KNO fit that is valid at ISR energies. Due to the large errors in the low-multiplicity bins, \( \langle n \rangle \) cannot be satisfactorily determined. Therefore, it is found by using the empirical relation \( \langle n \rangle \approx 1.25 n_{\text{max}} \), where \( n_{\text{max}} \) is the most probable multiplicity which is inferred from the KNO fit at ISR energies. The authors find an interesting feature when the part that follows the KNO fit is subtracted and the remaining part plotted (not shown here). The remaining part does not follow KNO, its most probable value is 2, and its width is about \( \sqrt{2} \) times the width of the KNO distribution.

This procedure to identify the second component is similar to the one described in the previous section. However, the authors of [36] conclude that the second part of
the distribution is the result of two independent parton–parton interactions within the
same collision. The cross-sections of the two contributions \( \sigma_1, \sigma_2 \) can be calculated as
a function of \( \sqrt{s} \). It is found that \( \sigma_1 \) is almost independent of \( \sqrt{s} \), while \( \sigma_2 \) increases
with \( \sqrt{s} \). However, it is unclear if two parton–parton interactions in the same collision
evolve independently to their final multiplicity due to final-state interactions.

The same reasoning and data is used in \[37\] to identify a third component, three
independent parton–parton interactions. This is extended in \[38\] to predict a multiplicity
distribution for 14TeV which is included in Section 4.

2.6. Similarities between \( p + p(\bar{p}) \) and \( e^+ e^- \) collisions and QCD predictions

The theoretical description of the formation of hadrons necessarily involves a soft
scale so that perturbative QCD cannot be directly applied. Therefore, models for soft
interactions like those from the large class of string models are often used to describe
multiplicity distributions in collisions of hadrons (see for example the Dual Parton Model
\[39\] or the Quark–Gluon String Model \[40\]). It is instructive to compare multi-particle
production in \( e^+ e^- \) and \( p + p(\bar{p}) \) collisions. In \( p + p(\bar{p}) \) collisions without a hard parton-
parton interaction as well as in \( e^+ e^- \) collisions particle productions can be viewed as
resulting from the fragmentation of color-connected partons. In \( e^+ e^- \) collisions the color
field stretches along the jet axis whereas in \( p + p(\bar{p}) \) it stretches along the beam axis.
Based on this analogy it is not unreasonable to expect some similarities between particle
production in \( e^+ e^- \) and \( p + p(\bar{p}) \) collisions. However, the configurations of the strings in
the two cases are different. Moreover, processes with different energy dependencies will
contribute significantly to the overall particle multiplicity at high energies \[20\]: minijet
production in hard parton-parton scatterings in case of \( p + p(\bar{p}) \) collisions and hard
gluon radiation in \( e^+ e^- \) collisions. Thus, theory does not provide convincing arguments
for this simple analogy, yet it will be worthwhile to compare particle production in these
two systems.

In \( e^+ e^- \) collisions moments of the multiplicity distributions are rather well
described in an analytical form with perturbative QCD in the ‘modified leading
logarithmic approximation’ (MLLA) \[11\]. The evolution of the parton shower is
described perturbatively to rather low virtuality scales close to the hadron mass. Using
the hypothesis of ‘local parton-hadron duality’ (LPHD) one then assumes a direct
relation between parton and hadron multiplicities. In \( e^+ e^- \) collisions the next-to-leading-
order (NLO) prediction for the average multiplicities is given by

\[
\langle N_{ch}(\sqrt{s}) \rangle = A_{\text{LPHD}} \cdot \alpha_s^b(\sqrt{s}) \cdot \exp \left( \frac{a}{\sqrt{\alpha_s(\sqrt{s})}} \right) + A_0
\]

(27)

with \( a = \sqrt{6\pi}12/23 \) and \( b = 407/838 \) for 5 quark flavors \[12\] \[13\] \[14\]. Fixing the
strong coupling constant \( \alpha_s \) at the \( Z \) mass to \( \alpha_s(M_Z^2) = 0.118 \) leaves \( A_0 \) and \( A_{\text{LPHD}} \nfree parameters. An excellent parametrization of the experimental multiplicities can
be obtained in this way (see Section \[5\]). An analytical form at next-to-next-to-next-to-leading order (3NLO) is available in \[14\] \[45\].
The second factorial moment for multiplicity distributions in $e^+e^-$ collisions

$$F_2 = \frac{\langle n(n - 1) \rangle}{\langle n \rangle^2} = 1 + \frac{D^2}{\langle n \rangle^2} - \frac{1}{\langle n \rangle}, \quad (28)$$

is given at NLO by

$$F_2(\sqrt{s}) = \frac{11}{8}(1 - 0.55\sqrt{\alpha_s(\sqrt{s})}). \quad (29)$$

This QCD prediction for $F_2$ is $\sim 10\%$ above the experimental values for $\sqrt{s} = 10 - 91.2$ GeV [47]. The calculation of higher moments shows that the theoretical multiplicity distribution in $e^+e^-$ collisions is well approximated by a negative binomial distribution [46] with

$$1/k \approx 0.4 - 0.88\sqrt{\alpha_s}. \quad (30)$$

This implies that asymptotically ($\alpha_s \to 0$ as $s \to \infty$) the multiplicity distributions in $e^+e^-$ satisfy KNO scaling. However, the KNO form of the multiplicity distributions up to the maximum LEP energy (corresponding to $\alpha_s \gtrsim 0.1$) differs significantly from the asymptotic form.

Even before QCD was known Polyakov found that KNO scaling occurs naturally in a picture of hadron production in a self-similar scale-invariant branching process [48, 49]. For $e^+e^-$ collisions the KNO form was given as

$$\psi(z) \propto a(z) \exp(-z^\mu) \quad \text{with} \quad \mu > 1, \quad (31)$$

where $a(z)$ is a monomial. Thus, $\psi(z)$ is a gamma distribution in $z^\mu$. In the ‘double logarithmic approximation’ (DLA) of QCD, valid at asymptotic energies, the KNO form of the multiplicity distribution in jets can be calculated [50, 51]. Higher-order correction to this form were found to be large [52] so that the preasymptotic distributions, e.g., at LEP energies, are quite different from the asymptotic DLA form [53].

3. Multiplicity Distributions: Measurements

This part of the review presents $p + p(\bar{p})$ measurements that have been performed by experiments at hadron colliders, i.e., the ISR, SppS, and Tevatron. The Intersecting Storage Rings (ISR), the very first hadron collider, was operating at CERN between 1971 and 1984. It collided $p$ on $p$, $\bar{p}$, and $\alpha$ at a maximum center-of-mass energy of 63 GeV. The Super Proton Synchrotron (SPS) which has operated at CERN since 1976 has accelerated in its lifetime electrons, positrons, protons, anti-protons, and ions. After modification to a collider, it provided $p$ on $\bar{p}$ collisions with a maximum $\sqrt{s}$ of 900 GeV, at that time it was called SppS. The Tevatron at the Fermi National Accelerator Laboratory (FNAL) came into operation in 1983. It provides $p + \bar{p}$ collisions at energies up to $\sqrt{s} = 1.96$ TeV. In addition, results from bubble chamber experiments are included where appropriate.

References of measurements of experiments at these colliders are given, discussing their analysis methods and error treatments. A representative selection of measurements...
of experiments at these colliders are shown to assess the validity of the models that have been described in the previous part of the review. Additionally, the experimental challenges are recalled and unresolved experimental inconsistencies discussed.

3.1. Analysis Techniques

3.1.1. Event Classes

Inelastic $p+p$ collisions are commonly divided into non-diffractive (ND), single-diffractive (SD), and double-diffractive (DD) events. Figure 2 shows rapidity distributions of those classes obtained with Pythia to illustrate their differences. Non-diffractive collisions (left panel) have many particles in the central region, steeply falling to higher rapidities. In a single-diffractive collision only one of the beam particles breaks up and produces particles at high rapidities on one side. In the center panel only those single-diffractive collisions are shown where the particle going to positive $y$ breaks up. The other incoming particle, nearly uninfluenced, is found at the rapidity of the beam. In a double-diffractive collision (right panel) both beam particles break up and produce particles at positive and negative high rapidities. A dip can be seen in the central region. The different scales of the three distributions should be noted. Integrating the histograms demonstrates that the average total multiplicity is about a factor four higher in non-diffractive collisions than in diffractive collisions.

Measurements are usually presented for the sample of all inelastic collisions or non-single-diffractive (NSD) collisions, i.e., not considering the SD component. The reason for the latter choice is that triggering detectors are usually less sensitive to SD events due to their topology: few particles are found in the central region and only the nearly uninfluenced incident proton is found on one side. To select a pure NSD sample for the analysis, depending on the detector geometry, SD events that pass the trigger can be rejected by their reconstructed topology. E.g. to reject events where in one hemisphere no track or only one track is found that has 80% of the incident proton momentum. Today’s detectors may have too limited phase space acceptance to allow a sufficient event-by-
Figure 3. The need for unfolding. The left panel shows a measured spectrum in a limited region of phase space superimposed with the true distribution that caused the entries in one single measured bin (exemplarily at multiplicity 30 indicated by the line). Clearly the shape of this true distribution depends on the shape of the multiplicity distribution given by the model used (a suggestive example is if the true spectrum stopped at a multiplicity of 40: the true distribution that contributed to the measured multiplicity of 30 would clearly be different, still events at a multiplicity of 30 would be measured). Inversely, in the right panel, the true distribution is shown superimposed with the measured distribution caused by events with the true multiplicity 30 (exemplarily). The shape of this measured distribution depends only on the detector simulation, i.e., the transport code and reconstruction, and not on the multiplicity distribution given by the model (only events with multiplicity 30 contribute to the shown measured distribution).

3.1.2. Unfolding of Multiplicity Distributions

Given a vector $T$ representing the true spectrum, the measured spectrum $M$ can be calculated using the detector response matrix $R$:

$$M = RT.$$  \hfill (32)

The aim of the analysis is to infer $T$ from $M$. Simple weighting, i.e., assuming that a measured multiplicity $m$ is caused ‘mostly’ by a true multiplicity $t$, would not be correct. This is illustrated in Figure 3. Analogous, adding for each measured multiplicity the corresponding row of the detector response matrix to the true distribution is incorrect. This is model-dependent and thus in principle not possible. On the other hand the measured spectrum which is the result of a given true multiplicity is only determined by the detector simulation and is model-independent.
Given a measured spectrum, the true spectrum is formally calculated as follows:

\[ T = R^{-1} M. \]  

(33)

\( R^{-1} \) cannot be calculated in all cases, because \( R \) may be singular; e.g. when a poor detector resolution causes two rows of the matrix to be identical. This can in most cases be solved by choosing a more appropriate binning (combining the entries in question). But even if \( R \) can be inverted, the result obtained by Eq. (33) contains usually severe oscillations (due to statistical fluctuation caused by the limited statistics of events used to create the response matrix). This can be illustrated with the following example [54]; a square response matrix is assumed to describe the detector (rows: measured multiplicities; columns: true multiplicities):

\[ R = \begin{pmatrix} 0.75 & 0.25 & 0 & \cdots \\ 0.25 & 0.50 & 0.25 & 0 \\ 0 & 0.25 & 0.50 & 0.25 \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \]  

(34)

A true distribution \( T \) is assumed, and the expected measured distribution \( M \) is calculated with Eq. (32). The distribution \( M \) is used to generate a sample of 10000 measurements: \( \tilde{M} \). Using Eq. (33) the corresponding true distribution \( \tilde{T} \) is calculated. Figure 4 shows these four distributions. Although the resolution effect on the shape of the measured distribution (left histogram) is very small, the unfolded solution (right histogram) suffers from large non-physical fluctuations. Clearly, this is not the spectrum that corresponds to the true one.

The information that is lost due to the resolution cannot be recovered in principle. To work around this problem the result is usually constrained with a priori knowledge about the smoothness of the function. Methods that allow the recovery of the true distribution are \( \chi^2 \) minimization (e.g. used in [55]) that leads to the true spectrum by minimizing a \( \chi^2 \) function and Bayesian unfolding [56, 57], an iterative method based on Bayes’ theorem. Both methods are described and evaluated in detail in [58].

Clearly, for the true distribution in full phase space only even numbers of particles occur due to charge conservation. Due to efficiency and acceptance effects in the measured spectrum also odd numbers occur. This has to be taken into account in the detector response matrix, i.e., every column corresponding to an odd number of generated particles is empty; for the number of measured particles even and odd values occur. For the correction to limited phase space this constraint does not arise.

3.2. Data Sample

The data sample considered in this review is summarized in Table 1. In the following details about the different analyses are given ordered by detector and collider. Unless otherwise stated, the correction procedures described in the publications consider the effect of decays of strange and neutral particles as well as secondary particles.
The **Split Field Magnet detector (SFM)** \[69\] at the ISR measured the multiplicity distribution for NSD and inelastic $p+p$ events at $\sqrt{s} = 30.4, 44.5, 52.6, \text{and} 62.2 \text{GeV} \[3\]. Between 26,000 and 60,000 events were collected for each of the energies.

The SD component was removed from the sample by means of its topology: events are considered SD when in one of the hemispheres no track or only one track carrying 80% of the incident proton’s energy is found. Systematic errors have been evaluated and include the error that arises from the corrections and in the low-multiplicity region from the subtraction of elastic events.

A detector based on streamer chambers \[6\] at the ISR measured pseudorapidity and multiplicity distributions for inelastic events at center-of-mass energies of 23.6, 30.8, 45.2, 53.2, and 62.8 GeV. Between 2300 and 5900 events were measured for each energy. In the analysis corrections for the acceptance, the low-momentum cut-off (about 45 MeV/c), and secondary particles due to interactions with the material are taken into account.

The **UA1 (Underground Area 1) experiment** measured the multiplicity distribution for NSD events in the interval $|\eta| < 2.5$ at $\sqrt{s} = 200, 500, \text{and} 900 \text{GeV} \[59\]$. 188,000 events were used, out of which 34% were recorded at the highest energy. The SppS was operated in a pulsed mode where data was taken during the energy ramp from 200 GeV to 900 GeV and vice versa. Therefore the data at 500 GeV is in fact taken in an energy range from 440 GeV to 560 GeV. Only tracks with $p_T$ larger than 150 MeV/c are considered for the analysis to reduce the contamination by secondaries.
Table 1. Listed are the references of data used in this chapter. For each reference it is indicated at which energy the sample was taken, to which event class it is corrected, and what kind of data ($dN_{ch}/d\eta$ and/or multiplicity distribution) are presented.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Ref.</th>
<th>Energy</th>
<th>$dN_{ch}/d\eta$</th>
<th>Mult.</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>SFM</td>
<td>[3]</td>
<td>30.4, 44.5, 52.6, 62.2 GeV (INEL, NSD)</td>
<td>X</td>
<td></td>
<td>a</td>
</tr>
<tr>
<td>Detector</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UA1</td>
<td>[59]</td>
<td>200, 500, 900 GeV (NSD)</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[60]</td>
<td>540 GeV (NSD)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UA5</td>
<td>[61]</td>
<td>53 GeV (INEL)</td>
<td>X</td>
<td>X</td>
<td>b</td>
</tr>
<tr>
<td></td>
<td>[63]</td>
<td>546 GeV (INEL, NSD)</td>
<td>X</td>
<td>X</td>
<td>c</td>
</tr>
<tr>
<td></td>
<td>[64]</td>
<td>540 GeV (NSD)</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[22]</td>
<td>540 GeV (NSD)</td>
<td>X</td>
<td></td>
<td>d</td>
</tr>
<tr>
<td></td>
<td>[23]</td>
<td>200, 900 GeV (NSD)</td>
<td>X</td>
<td></td>
<td>e</td>
</tr>
<tr>
<td>P238</td>
<td>[65]</td>
<td>630 GeV (NSD)</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CDF</td>
<td>[66]</td>
<td>0.63, 1.8 TeV (NSD)</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[9]</td>
<td>1.8 TeV (NSD)</td>
<td>X</td>
<td></td>
<td>f</td>
</tr>
<tr>
<td>E735</td>
<td>[36]</td>
<td>0.3, 0.5, 1.0, 1.8 TeV (NSD)</td>
<td>X</td>
<td></td>
<td>g</td>
</tr>
<tr>
<td></td>
<td>[67]</td>
<td>0.3, 0.5, 1.0, 1.8 TeV (NSD)</td>
<td>X</td>
<td></td>
<td>h</td>
</tr>
<tr>
<td></td>
<td>[68]</td>
<td>1.8 TeV (NSD)</td>
<td>X</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a Error of cross-section included in multiplicity distribution.
b Comparison $p + \bar{p}$ vs. $p + p$; multiplicity only uncorrected.
c Comprehensive report.
d Multiplicity distribution forced to be of NBD shape.
e Method partially revised by [8].
f No systematic error assessment.
g Method description very limited; extrapolated from $|\eta| < 3.25$ to full phase space.
h Only in KNO variables; no systematic error assessment.

Although not explicitly mentioned in the publication, it is assumed for this review that the low-momentum cut-off correction is part of the acceptance correction. UA1 quotes the overall systematic error to be 15%: contributions are from strange-particle decays, photon conversions and secondary interactions (3%), as well as the uncertainty in the acceptance (4%). Other contributions arise from the selection criteria and uncertainties in the luminosity measurement (10%). (The luminosity measurement uncertainty only applies to the cross-section measurement, not to the normalized distribution. The
uncertainty due to the selection criteria is not quoted. Therefore, assuming that the
systematic uncertainties were summed in quadrature, this uncertainty is 10% and the
overall systematic error without the uncertainty on the luminosity is 11% which is the
value applicable to the normalized multiplicity distribution.)

UA1 measured the $dN_{ch}/d\eta$ distribution at $\sqrt{s} = 540$ GeV [60]. The analysis used
8,000 events that have been taken without magnetic field which reduced the amount
of particles lost at low-momenta to about 1%. The systematic error of the applied
corrections is estimated by the authors to be 5% without enumerating the contributions.

The UA5 (Underground Area 5) experiment has operated at the ISR and
the SpøS. A comparison of data taken in $p+p$ and $p+\bar{p}$ collisions at $\sqrt{s} = 53$ GeV was
made [61]. 3,600 $p+p$ events and 4,000 $p+\bar{p}$ events were used. Trigger and vertex finding
efficiencies as well as acceptance effects have been evaluated with a MC simulation that
was tuned to reproduce ISR data. The $dN_{ch}/d\eta$ distribution was measured for both
collision systems and compared. The comparison was done using the uncorrected data
and only for events with at least two tracks. In this way the authors attempted to achieve
lower systematic errors on the result. A ratio of $1.015 \pm 0.012$ ($p+\bar{p}$ over $p+p$) has
been found. Furthermore, the multiplicity distributions were compared. It is concluded
that these distributions agree within errors. The authors summarize that the differences
between $p+p$ and $p+\bar{p}$ collisions are less than 2%.

UA5 measured the $dN_{ch}/d\eta$ distribution at $\sqrt{s} = 200$ and 900 GeV for NSD events
[62, 63]. 3,500 (2,100) events have been used for the analysis at 900 (200) GeV. It
should be noted that the corrections are based on a MC simulation that has been
tuned to reproduce data measured at $\sqrt{s} = 546$ GeV. The results of the simulation
were parameterized and scaled to $\sqrt{s} = 200$ and 900 GeV in order to estimate the
corrections for acceptance and contamination by secondaries. Unfortunately, the authors
only mention statistical errors explicitly.

Measurements of the multiplicity distribution have been presented in [64, 22]
[23, 8, 63]. The distribution is measured in increasing $\eta$-regions (smallest: $|\eta| < 0.2$
for 540 GeV and $|\eta| < 0.5$ for 200 and 900 GeV) up to $|\eta| < 5.0$. Furthermore, the
result is presented extrapolated to full phase space. The analysis used 4,000 events
for 200 GeV and 7,000 events each for 540 and 900 GeV. In all cases the unfolding of
the measured spectrum was performed by minimizing a $\chi^2$ function. For the case of
$\sqrt{s} = 540$ GeV [22] it was required that the resulting function is an NBD which is
regarded as a strong constraint. This has to be taken into account when interpreting
the result at 540 GeV. The distributions at 200 and 900 GeV were unfolded using the
maximum-entropy method [8] which is considered to be a less restrictive assumption.
The assessment of the systematic errors is not very comprehensive and concludes an
uncertainty of about 2%.

A Forward Silicon Micro-Vertex detector that was tested in the context of a
proposed hadronic B-physics experiment (P238) measured the $dN_{ch}/d\eta$ distribution in

† [8] partially revised the method to obtain the distribution and thus the results. Therefore, the results
from [8] are used instead of [23].
forward rapidities at $\sqrt{s} = 630$ GeV [63]. A sample of 5 million events is corrected for
tracks from secondaries (2%) and SD events (0.5%). Acceptance and resolution effects
are corrected by using MC simulations tuned to UA5 data. Their magnitude as well
as the magnitude of the trigger- and vertex-efficiency correction are not mentioned. A
normalization error of 5% dominates the systematic error that stems from inconsistent
results when only the $x$ or $y$ tracking information is used compared to when both of
them are used. Other effects such as detector efficiency, misalignment, and the SD cross-
section are considered by the authors to not significantly contribute to the systematic
uncertainty.

The CDF (Collider Detector at Fermilab) experiment [70], a detector at
the Tevatron collider, measured the $dN_{ch}/d\eta$ distribution at $\sqrt{s} = 630$ GeV and 1.8 TeV
with their so-called vertex time-projection chambers (VTPCs) [66]. These VTPCs have
been replaced after years of operation by a silicon detector. Unfortunately, the authors
do not mention if their corrections correspond to NSD or inelastic events. However,
the trigger configuration requires a hit on both sides. This points to the fact that the
trigger is insensitive to the majority of SD events. Furthermore, the authors compare
their measurement to NSD data from UA5 which confirms that the CDF data is for NSD
events. 2800 (21000) events have been used for the analysis at 630 (1800) GeV. Only
events with at least 4 tracks are considered to reduce the beam-gas background. The
authors stated that they “do not correct for events missed by the trigger or selection
procedure” and estimated that the selection procedure misses $(13 \pm 6)$% of the events.
This is surprising because the normalization for $dN_{ch}/d\eta$ would be significantly distorted
if this correction was not applied. This is not the case shown in the comparison to
UA5 data. Tracks with $p_T < 50$ MeV/c are not found due to the magnetic field and a
correction of $(3 \pm 2)$% is applied to account for this loss. A systematic error assessment
is made; the error is dominated by uncertainties in the tracking efficiency and ranges
from 3% (at $\eta = 0$) to 15% (at $|\eta| = 3.25$).

CDF measured the multiplicity distribution in various $\eta$-intervals for NSD events
at $\sqrt{s} = 1.8$ TeV [9]. The publication does not mention the number of events used in
the analysis. A systematic-error assessment is reported to be ongoing, but has not yet
been published. It is unclear if an unfolding method was used.

The E735 experiment [67] at the Tevatron collider measured the multiplicity
distribution of NSD events at energies of $\sqrt{s} = 0.3$, 0.5, 1.0, and 1.8 TeV [36].
The extrapolation to full phase space has been done by the authors based on
Pythia simulations. They provide no further information about the statistics used, the
corrections, and in particular the question whether an unfolding was used. This has to
be taken into account when the result is interpreted.

In [67] multiplicity distributions of NSD events are presented in limited intervals
of $|\eta| < 1.62$ and $|\eta| < 3.25$ as well as extrapolated to full phase space for the four
aforementioned energies. A total number of 25 million events is mentioned, however
only a subset is used for the multiplicity analysis whose size is not mentioned. The
results are only presented in KNO variables. The data has been unfolded using the
Figure 5. KNO scaling at ISR energies. The figure shows normalized multiplicity distributions for NSD events in full phase space vs. multiplicity (left panel) and using KNO variables (right panel). The data was measured by the SFM [3].

maximum-entropy method. A systematic error assessment is not performed.

[DS] presents multiplicity distributions, like before, in $|\eta| < 1.62$, $|\eta| < 3.25$, and in full phase space at $\sqrt{s} = 1.8$ TeV of NSD events. About 2.8 million events have been used and unfolded with an iterative method similar to the mentioned Bayesian unfolding. Systematic uncertainties have been evaluated concentrating on the effect of the cuts to reduce contamination by single-diffractive and beam-gas events. In [67] and [68] a correction for strange-particle decays is not mentioned explicitly, but it can be assumed to have been part of the MC simulation used to obtain the correction factors.

3.3. Multiplicity Distributions from $\sqrt{s} = 20$ to 1800 GeV

In the following sections the theoretical and phenomenological concepts introduced in the first part of the review are applied to selected multiplicity distributions. An example for KNO scaling as well as the fit with a NBD and a combination of two NBDs is shown. The available distributions in full phase space are shown together in multiplicity and KNO variables to assess the validity of KNO scaling, which is further discussed in Section 3.6.

Figure 5 shows multiplicity distributions in full phase space for NSD events taken at the ISR. The distribution is shown in multiplicity and KNO variables. The latter shows that KNO scaling is fulfilled at ISR energies (the moments of these distributions are analyzed further below).

Multiplicity distributions are described by NBDs up to $\sqrt{s} = 540$ GeV in full phase space as well as in different $\eta$-ranges. This behavior does not continue for $\sqrt{s} = 900$ GeV. Figure 6 shows multiplicity distributions together with NBD fits in
Figure 6. Normalized multiplicity distributions of NSD events at $\sqrt{s} = 900 \text{GeV}$ in various rapidity intervals are shown fitted with single NBDs (top left panel) or a combination of two NBDs (bottom left panel). The two contributing NBDs (dashed lines) are shown exemplarily for $|\eta| < 3.0$ and 5.0. The right panels show the normalized residuals with respect to the corresponding fits defined by $(1/e)(P(N_{ch}) - \text{fit})$ with $e$ being the error on $P(N_{ch})$. These are smoothed over four data points to reduce fluctuations. The data has been measured by UA5 [8].

increasing pseudorapidity ranges at 900 GeV (top left panel). The respective normalized residuals are also shown (top right panel). The NBD fit works very well for the interval $|\eta| < 0.5$, but it becomes more and more obvious with the increasing $\eta$-range that the region around the most probable multiplicity is not reproduced. The structure found around the peak gave rise to the two-component approach, discussed previously, in which the data are fit with a combination of two NBDs. The bottom left panel of Figure 6 shows these fits with Eq. (26) on page 11 (no constraints are put on the fit parameters),
and normalized residuals (bottom right panel) to the same data which yields good fit results for all pseudorapidity ranges.

To assess the validity of KNO scaling all available multiplicity distributions are drawn as function of the KNO variable $z = N_{ch}/\langle N_{ch} \rangle$. This is shown in Figure 7 for NSD events in full phase space from 30 to 1800 GeV. Although it is evident that the high-multiplicity tail does not coincide between the lowest and highest energy dataset, no detailed conclusion is possible for the data in the intermediate energy region. The study of the moments of these distributions allows to draw further conclusions which is performed in Section 3.6.

3.4. $dN_{ch}/d\eta$ and $\langle N_{ch} \rangle$ vs. $\sqrt{s}$

Figure 8 shows $dN_{ch}/d\eta$ at energies ranging over about two orders of magnitudes, from ISR ($\sqrt{s} = 23.6$ GeV) to CDF ($\sqrt{s} = 1.8$ TeV). Increasing the energy shows an increase in multiplicity. The multiplicity of the central plateau increases together with the variance of the distribution. Note that the data points at the lowest energy are for inelastic events, the other data points refer to NSD events. We recall that the dip around $\eta \approx 0$ is due to the transformation from rapidity $y$ to pseudorapidity $\eta$.

The left panel of Figure 9 shows $dN_{ch}/d\eta|_{\eta=0}$ as a function of $\sqrt{s}$. Filled symbols are data for inelastic events; open symbols for NSD events. $dN_{ch}/d\eta|_{\eta=0}$ increases with increasing $\sqrt{s}$, which is violating Feynman scaling. Two fits are shown for the NSD data: a fit with $a + b \ln s$ (solid black line) and $a + b \ln s + c \ln^2 s$ (dashed red line). Due to the fact that different published values include different errors, e.g., no systematic errors for the UA5 data, the errors are not used for the fit. The $\ln s$ dependence was used to describe the data at center-of-mass energies below 1 TeV. Data at a higher energy from...
Charged-Particle Multiplicity in Proton–Proton Collisions

Figure 8. \( dN_{\text{ch}}/d\eta \) at different \( \sqrt{s} \). Data points from [6, 60, 22, 8, 66, 65].

Figure 9. \( dN_{\text{ch}}/d\eta \mid_{\eta=0} \) (left panel) and \( \langle N_{\text{ch}} \rangle \) vs. \( \sqrt{s} \) in full phase space and \( |\eta| < 1.5 \) (right panel) as a function of \( \sqrt{s} \). Data points from [5, 71, 61, 64, 3, 22, 8, 36, 66, 69, 60].

CDF showed deviations from this fit [66]. The additional \( \ln^2 s \) term yields a better result (\( \chi^2/\text{ndf} \) about half, although the \( \chi^2 \) definition is not valid without using the errors) and shows that \( dN_{\text{ch}}/d\eta \mid_{\eta=0} \) increases faster than \( \ln s \). The functional fits are extrapolated up to the nominal LHC energy, \( \sqrt{s} = 14 \text{ TeV} \).

The right panel of Figure 9 shows the average multiplicity \( \langle N_{\text{ch}} \rangle \) as a function of...
\sqrt{s}$. Data is shown for full phase space and for a limited rapidity range of $|\eta| < 1.5$. In publications two different approaches are found to obtain average values in a limited $\eta$-range. The first uses a normalization to all events having at least one track in the considered phase space. The second approach uses a normalization to the total considered cross-section (inelastic or NSD) including events without any particle in the considered range (data shown here). While the latter is the more evident physical observable, the former is not dependent on the efficiency to measure the total cross-section. Thus the former is less dependent on model assumptions used in the evaluation of the trigger efficiency. Data from bubble chambers at low $\sqrt{s}$ is included in Figure 9: from the Mirabelle chamber at Serpukhov, Russia [5] and from several bubble chambers at FNAL [71]. Both sets of NSD data are fitted, as before, ln and ln$^2$ dependent. For full phase space the logarithmic dependence does not reproduce the data and is only shown to demonstrate the violation of Feynman scaling; the ln$^2$ dependence fits the data well. For limited phase space both fits are reasonable. The functional fits are extrapolated to $\sqrt{s} = 14$ TeV.

3.5. Universality of Multiplicities in $p + p(\bar{p})$ and $e^+e^-$

Charged-particle multiplicities in $e^+e^-$ collisions are found to be higher than the multiplicity in $p + p(\bar{p})$ collisions at the same center-of-mass energy (see Figure 11). The multiplicities in $e^+e^-$ and $p + p(\bar{p})$ collisions become strikingly similar if the $p + p(\bar{p})$ points are plotted at $\sim 1/2$ of their collision energy [72, 73]. This leads to the concept of an "effective energy" in $p + p(\bar{p})$ collisions that is available for particle production [74, 75, 76, 77]. This concept emerged already in 1970s in the study of high-energy cosmic rays [78]. In this picture the remaining energy is associated with the two leading baryons which emerge at small angles with respect to the beam direction:

$$E_{\text{eff}} = \sqrt{s} - (E_{\text{lead,1}} + E_{\text{lead,2}}); \quad \langle E_{\text{eff}} \rangle = \sqrt{s} - 2\langle E_{\text{leading}} \rangle.$$ (35)

It has been speculated that $E_{\text{eff}}$ or correspondingly the inelasticity $K = E_{\text{eff}}/\sqrt{s}$ are related to the 3-quark structure of the nucleon [79, 80, 81]. In this simple picture the interaction of one of the three valence quarks in each nucleon would correspond to an average inelasticity $\langle K \rangle \approx 1/3$.

Here we estimate the coefficient of inelasticity $K$ of a $p + p(\bar{p})$ collisions by comparing $p + p(\bar{p})$ with $e^+e^-$ collisions. Given a parametrization $f_{ee}(\sqrt{s})$ of the $\sqrt{s}$ dependence of $N_{\text{ch}}$ in $e^+e^-$ collisions one can fit the $p + p(\bar{p})$ data with the form [79]

$$f_{pp}(\sqrt{s}) = f_{ee}(K \cdot \sqrt{s}) + n_0.$$ (36)

The parameter $n_0$ corresponds to the contribution from the two leading protons to the total multiplicity and is expected to be close to 2.

As a parametrization of multiplicity data in $e^+e^-$ we use the analytic QCD expressions of Eq. (27). The strong coupling constant $\alpha_s$ was fixed at the $Z$ mass to $\alpha_s(M_Z^2) = 0.118$ leaving $A_0$ and $A_{LPHD}$ as fit parameters. The second form is from a 3NLO calculation [14, 45] where the normalization and the $\Lambda$ parameter in the expression
Charged-Particle Multiplicity in Proton–Proton Collisions

Figure 10. Left panel: Comparison of charged particle multiplicities in \( p + p(\bar{p}) \) and \( e^+ e^- \) collisions. Right panel: The inelasticity in \( p + p(\bar{p}) \) calculated for three different assumptions. The \( \sqrt{s} \) dependence of the inelasticity assumed in the theoretical study [82] is shown for comparison.

for \( \alpha_s \) were taken as fit parameters. Both forms yield excellent fits of the \( e^+ e^- \) data and essentially provide the same extrapolation for \( \sqrt{s} > 206 \text{ GeV} \) where no data are available.

A fit with Eq. (36) describes the \( p + p(\bar{p}) \) well (\( \chi^2 / n_{\text{df}} = 12.0 / 6 \)) and yields \( K = 0.35 \pm 0.01 \) and \( n_0 = 2.2 \pm 0.19 \). The fraction of the effective energy, the inelasticity, is studied in more detail in the right panel of Figure 10. The inelasticity \( K \) is determined for each \( p + p(\bar{p}) \) point by solving

\[
f_{ee}(K \sqrt{s_{pp}} - \Delta m) = \langle N_{\text{ch}} \rangle_{pp} - n_0 .
\]

In the simple quark-scattering picture the offset \( \Delta m \) takes the contribution of the masses of the two participating constituent quarks to the center-of-mass energy into account.

Depending on the values for \( \Delta m \) and the leading particle multiplicity \( n_0 \) different inelasticities can be defined. In Figure 10 the three cases \( K_1 \) \((n_0 = 0, \Delta m = 0)\), \( K_2 \) \((n_0 = 2.2, \Delta m = 0)\), and \( K_3 \) \((n_0 = 2.2, \Delta m = 2/3 m_{\text{proton}})\) are shown. The inelasticity \( k_1 \) decreases from \( \sim 0.55 - 0.6 \) at ISR energies to 0.4 at \( \sqrt{s} = 1.8 \text{ TeV} \). The inelasticities \( K_2 \) and \( K_3 \) appear to be energy independent at \( \sim 0.35 \), in remarkable agreement with the expectation of 1/3 in the simple quark-scattering picture.

The similarity between \( \langle N_{\text{ch}} \rangle \) in \( e^+ e^- \) and \( p + p(\bar{p}) \) collisions when the effective energy is taken into account raises the question whether these similarities still persist in more differential observables. Here we concentrate on rapidity distributions. Note that a remarkable similarity was observed between \( dN_{\text{ch}} / d\eta \) per participating nucleon pair in central \( \text{Au+Au} \) collisions at \( \sqrt{s_{NN}} = 200 \text{ GeV} \) and in \( e^+ e^- \) collisions at \( \sqrt{s} = 200 \text{ GeV} \) [72]. This suggests that the effective energy in central \( \text{Au+Au} \) collisions is close to 100%
of the beam energy, most likely due to the multiple interactions of the nucleons. In the left panel of Figure 11 $dN_{ch}/d\eta$ distributions from $p + p(\bar{p})$ collisions are compared with rapidity distributions $dN_{ch}/dy_T$ with respect to the thrust axis from $e^+e^-$ collisions. Datasets are compared for which $\sqrt{s_{pp}} \approx 2 - 3\sqrt{s_{ee}}$. For the shown cases the $dN_{ch}/d\eta$ distribution in $p + p(\bar{p})$ are broader than the $dN_{ch}/dy_T$ distributions. This might indicate the contribution from beam particle fragmentation in $p + p(\bar{p})$. The width $\lambda$ of the distribution defined as $\lambda = \langle N_{ch} \rangle / dN_{ch}/d\eta \rangle |_{\eta=0}$ and $\lambda = \langle N_{ch} \rangle / dN_{ch}/dy_T \rangle |_{y_T=0}$, respectively, are shown in the right panel of Figure 11. Based on the QCD calculation in [83] $\lambda$ is expected to scale as $\sqrt{\ln s}$. As shown in Figure 11 this form does not describe the $p+p(\bar{p})$ data which are well parameterized with $\lambda = a + b \ln \sqrt{s}$. The Landau hydrodynamic model also predicts a $\sqrt{\ln s}$ dependence of $\lambda$ and hence also fails to describe the $p+p(\bar{p})$ data.

It will be interesting to see whether this universality of multiplicities in $e^+e^-$ and $p + p(\bar{p})$ collisions also holds at LHC energies. This is guided by the fact the universality appears to be valid at least up to Tevatron energies and not so much by theory as discussed in Section 2.6. Under the assumptions that $k_2$ remains constant at about 0.35 also at LHC energies and that the extrapolation of the $e^+e^-$ data with the 3NLO QCD form is still reliable at $\sqrt{s} \approx 5$ TeV one can use the fit of $p + p(\bar{p})$ data to predict the multiplicities at the LHC. This yields $\langle N_{ch} \rangle \approx 79.7$ at 10 TeV and $\langle N_{ch} \rangle \approx 88.9$ at 14 TeV. Extrapolating the ratio $\lambda = \langle N_{ch} \rangle / (dN_{ch}/d\eta) |_{\eta=0}$ with the form $\lambda = a + b \ln \sqrt{s}$ (see Figure 11) these multiplicities correspond to $dN_{ch}/d\eta |_{\eta=0} \approx 5.9$ at 10 TeV and $dN_{ch}/d\eta |_{\eta=0} \approx 6.4$ at 14 TeV.

Figure 11. Left panel: The width $\lambda$ of the $\eta$ distributions $(p+p(\bar{p}))$ and $y_T$ distributions $(e^+e^-)$ as a function of $\sqrt{s}$. $y_T$ is the rapidity with respect to the thrust axis of the $e^+e^-$ collisions. Right panel: Comparison of $\eta$ $(p + p(\bar{p}))$ and $y_T$ distributions $(e^+e^-)$ at different energies.
3.6. Moments

The study of the moments of the multiplicity distribution allows to identify general trends as function of $\sqrt{s}$ and to study the validity of KNO scaling. The moments analyzed here have been defined in Section 2.2.

First the reduced $C$-moments, Eq. (6), are studied. The left panel of Figure 12 shows $C_2$ to $C_5$ from $\sqrt{s} = 30$ to 1800 GeV. These have been calculated from the available multiplicity distributions and are consistent with published values where available. At lower energies data from bubble-chamber experiments shows that the moments are constant (see e.g. [23] for a compilation). In the right panel a constant is fitted to the data points from ISR which shows that the moments are not constant. For $\sqrt{s}$ larger than at ISR, they increase with energy.

However, as argued in Section 2.3 the conclusion about constant $C$-moments follows from KNO scaling only in an approximation. Therefore the behavior of factorial moments is analyzed. Exemplarily $F_2$ and $F_4$ are shown in the left panel of Figure 13 compared to their $C$-moments counterparts. Also these increase with increasing $\sqrt{s}$. Therefore for conclusions about KNO scaling it is not crucial if $C$- or $F$-moments are studied; both show an increase with $\sqrt{s}$.

Important to note is the influence of the tail of the distribution, i.e., of bins at high multiplicity, on the moments; especially on the higher ones. The right panel of Figure 14 compares $C_4$ and $C_5$ calculated from a subset of bins, excluding the ones that are below 0.01 (i.e. less than 1% of events occur at this multiplicity) with the values calculated with all bins. The value of 0.01 is approximately the smallest bin content in the data from
Charged-Particle Multiplicity in Proton–Proton Collisions

 ISR. The difference is significant which shows that the value of the moments will change if more events are collected at a given energy. Nevertheless, although less pronounced, the moments increase with $\sqrt{s}$. One may ask of course why the moment calculated with all bins and the moment calculated from the subset do not agree within uncertainties. This is due to the fact that for all bins without entries an uncertainty of 0 is assumed which is incorrect. Assuming a Poissonian distribution in each bin (with an unknown mean), a bin with no measured entries has an upper limit of 2.3 at 90% confidence level (see e.g. [84]). However, following this strictly would mean to assign this error for all bins without entries up to infinity. Consequently, also the uncertainty on the moments goes to infinity.

In Figure 13 the ratio of the average multiplicity $\langle n \rangle$ and the dispersion $D$ is shown that is supposed to be constant when KNO scaling holds [10]. Additionally to the $p+p$ data also $e^+e^-$ results are shown. For $p+p$ the ratio is clearly not constant, while it is approximately constant for $e^+e^-$ where, however, the data points feature significantly larger errors. At the same $\sqrt{s}$ the multiplicity distribution in $p+p$ is significantly broader than in $e^+e^-$. In summary, the $C$- and $F$-moments increase with $\sqrt{s}$, even considering the influence of high-multiplicity bins. Furthermore, $\langle n \rangle/D$ is not constant. These facts confirm that KNO scaling is broken.

CDF has addressed the question whether the violation of KNO scaling is related to a special class of events [93]. They use events at 1.8 TeV and only tracks with a $p_T$ above

Figure 13. Left panel: $C$- and $F$-moments at different $\sqrt{s}$. The lines are constant functions fitted to the low-energy datapoints from the ISR. Right panel: Influence of high-multiplicity bins on $C$-moments. The moments $C_4$ and $C_5$ are shown once using all bins for the calculation and once only bins that contain at least 1% of the topological cross-section. The lines are constant functions fitted to the low-energy data points from the ISR. Data from [3, 8, 22, 36].
Charged-Particle Multiplicity in Proton–Proton Collisions

Figure 14. Ratio of dispersion and average multiplicity as function of $\sqrt{s}$ for $p + p$ and $e^+e^-$ data. Data from [3, 8, 22, 30]($p + p(\bar{p})$) and [85, 2, 86, 87, 10, 88, 89, 90, 91]($e^+e^-$).

0.4 GeV/$c$. Here, a weak KNO scaling violation is reported in $|\eta| < 1.0$. Furthermore, when they divide their data sample into two parts, they can confirm KNO scaling for the soft part of their events and at the same time rule it out for the hard part. In [93] soft events are defined as events without clusters of tracks with a total transverse energy above 1.1 GeV, regarded as jets.

Two further interesting features are observed together with the onset of KNO scaling violations [59]: the average transverse momentum that was about 360 MeV/$c$ at ISR energies starts to increase. Furthermore, a $\sqrt{s}$ dependent correlation between the average-$p_T$ and the multiplicity is measured. Both observations point to the fact that the influence of hard scattering becomes important at these energies.

It was mentioned earlier that higher-order QCD calculations predict oscillations of the $H$-moments as function of the rank. The uncertainties of moments increase with the rank (see e.g. Figure 12 for $C$-moments); this fact applies also to the $H$-moments. The search for oscillations requires the calculation of moments up to ranks of 10 – 20. The data studied in this review allows to calculate these moments only with large uncertainties. Oscillations are indicated but definite conclusions require a deeper study and distributions with large statistics that can hopefully be obtained at the LHC.

3.7. NBD Parameters $\langle N_{ch} \rangle$ and $k^{-1}$

Fitting multiplicity distributions with NBDs is successful up to about 540 GeV; at 900 GeV deviations are visible. Distributions at larger $\sqrt{s}$ can be successfully fitted with a combination of two NBDs (discussed in the subsequent section).

Figure 15 shows exemplarily multiplicity distributions from ISR and UA5 fitted
Charged-Particle Multiplicity in Proton–Proton Collisions

**Figure 15.** Multiplicity distribution at $\sqrt{s} = 30.4$ GeV measured at the ISR [3] and $\sqrt{s} = 900$ GeV measured by UA5 [8] fitted with a NBD (both panels) and a combination of two NBDs (right panel).

**Figure 16.** Left panel: Parameters of a single NBD fit and corresponding $\chi^2/\text{ndf}$. $k^{-1}$ and $\chi^2/\text{ndf}$ are scaled for visibility. Right panel: Parameters of a single NBD fit compared between $p+p$ (fits performed here) and $e^+e^-$ (from [85, 2, 87, 10]). The area between the dashed lines corresponds to the predictions for $1/k$ from [10].

with a NBD. While in the former the NBD reproduces the shape very well, in the latter structures are visible that are not reproduced by the fit.

In Figure 16 (left panel) the obtained fit parameters $\langle n \rangle$ and $k^{-1}$ are shown for datasets in full phase space at different $\sqrt{s}$ as well as the $\chi^2/\text{ndf}$ of the fits. $\langle n \rangle$ increases with $\ln \sqrt{s}$ like it was already discussed in Section 3.4. $k^{-1}$ increases with $\sqrt{s}$ and can be
fitted with a function of the form $a + b \ln \sqrt{s}$. KNO scaling corresponds to a constant, energy-independent $k$. Figure 16 (right panel) compares $k^{-1}$ from $p + p$ and $e^+e^-$ data. Both can be fitted with the same functional form, but the values for $e^+e^-$ are generally lower, indicating a narrower distribution. An extensive compilation of $k^{-1}$ of $p + p$ and $e^+e^-$ data can be found in [20, Figure 2.5]. For $e^+e^-$, this compilation includes further $k^{-1}$ at lower energies and shows that $k^{-1}$ flattens for larger $\sqrt{s}$, i.e., approaches KNO scaling. This behavior is not found for $p + p$. Interesting is the fact that the fit at $\sqrt{s} = 900$ GeV which does not reproduce the measured distribution very well (right panel of Figure 15) is not reflected in the $\chi^2/\text{ndf}$ (it is anyway about 1).

3.8. Two NBD Fit

Deviations between the multiplicity distribution and the fit with a single NBD are found at higher SppS energies. The combination of two NBDs (Eq. (26)) fits the data better. Both fit attempts are shown in the right panel of Figure 15 for $\sqrt{s} = 900$ GeV. Fits with two NBDs can be performed unconstrained or following an approach that constrains the parameters following some assumptions like it has been suggested in [29]. These two ways are compared in the following.

In [29] first the average multiplicity of the soft component $\langle n \rangle_{\text{soft}}$ using only data below $\sqrt{s} = 60$ GeV and the total average multiplicity $\langle n \rangle_{\text{total}}$ using available data up to $\sqrt{s}$ of 900 GeV are fitted. A ln-dependence is assumed for $\langle n \rangle_{\text{soft}}$, while additionally for $\langle n \rangle_{\text{total}}$ a ln²-term is added.

Following the assumption that the semi-hard component has about twice the average multiplicity than the soft component, which is based in [29] on a minijet-analysis by UA1, $\alpha$ can be calculated from $\langle n \rangle_{\text{soft}}$ and $\langle n \rangle_{\text{total}}$. Two scenarios are considered, scenario A in which $\langle n \rangle_{\text{semi-hard}} = 2 \langle n \rangle_{\text{soft}}$, and scenario B with $\langle n \rangle_{\text{semi-hard}} = 2 \langle n \rangle_{\text{soft}} + 0.1 \ln^2 \sqrt{s}$.

$k_{\text{soft}}$ is found rather constant between 200 and 900 GeV and thus set to 7. Three scenarios are presented then for the extrapolation to higher energies. The first considers that KNO scaling is valid above 900 GeV ($k_{\text{semi-hard}} \approx 13$). Scenario 2 fits $k_{\text{total}}$ with:

$$k_{\text{total}}^{-1} = a + b \ln \sqrt{s}. \quad (38)$$

Scenario 3 fits a next-to-leading order QCD prediction to $k_{\text{semi-hard}}$:

$$k_{\text{semi-hard}}^{-1} \approx a - \sqrt{b/\ln(\sqrt{s}/Q_0)}. \quad (39)$$

The free parameters are then found by fitting the data. These three scenarios (1–3) can be combined with the aforementioned scenario A and B, resulting in a total of six possibilities. In subsequent comparisons with data we restrict ourselves to only three of them (1–3 combined with A).

Figure 17 shows the functional forms found in [29]. These are compared to the unconstrained fitting results which we obtained by fitting the distributions with Eq. (26). Note that only the data from $\sqrt{s} = 200$ GeV to 900 GeV was used to fit the functional forms in [29]. There are clear differences, e.g., at 200 GeV for $\langle n \rangle_{\text{semi-hard}}$ and at
540 GeV for $k_{\text{semi-hard}}^{-1}$. One observes also large errors for certain fits showing that several solutions with similarly small $\chi^2/\text{ndf}$ exist. The $\chi^2/\text{ndf}$ is as expected generally better for unconstrained fitting. In several cases the $\chi^2/\text{ndf}$ is significantly lower than 1 which is unexpected and might be attributed to the requirement of smoothness in the unfolding procedure.

At 1.8 TeV the fraction of soft events is much larger in the unconstrained fit; also the other fit parameters do not follow the extrapolations. Consequently, for the constrained fit, the $\chi^2/\text{ndf}$ is very large at this energy, the fit is not very good. This is further discussed below.

Figure 17 presents a comparison of this model’s predictions (using the values from the authors) with data from CDF in $|\eta| < 1$ at $\sqrt{s} = 1.8$ TeV. Scenario 1 and 3 reproduce the spectrum reasonably well.

Figure 18 shows the comparison for data from E735 in full phase space at $\sqrt{s} = 1.8$ TeV. Only scenario 3 follows the general trend of the distribution. However, none of the curves reproduces the fine structures.

We conclude that unconstrained fitting with two NBDs works successfully with a reasonable low $\chi^2/\text{ndf}$ for all distributions considered here. However, general trends as a function of $\sqrt{s}$ cannot always be identified. As an alternative in [29] parameters are fixed following certain assumptions resulting in more systematic fit results. However, the results are partly significantly different from the parameters obtained with unconstrained fitting.
Figure 18. Comparison between the predictions of the two-component model [94] with the CDF measurement in $|\eta| < 1$ at $\sqrt{s} = 1.8$ TeV [9]. The right panel shows the ratio between data and the predictions.

Figure 19. Comparison between the predictions of the two-component model [29] with the E735 measurement in full phase space at $\sqrt{s} = 1.8$ TeV [30]. The right panel shows normalized residuals between data and the predictions.

3.9. Open Experimental Issues

This section addresses open experimental issues and present some comparison plots between data of the disagreeing experiments.

A direct comparison between UA1 and UA5 at $\sqrt{s} = 540$ GeV in limited regions and in KNO variables shows that the two experiments agree with their confirmation of KNO scaling in the interval $|\eta| < 0.5$ and disagree in the interval $|\eta| < 1.5$, but the
violation of KNO scaling in the UA5 data is only due to an excess of events with $z > 3.5$, i.e., events that have more than 3.5 times the average multiplicity. This comparison has been performed in [59] and is shown for $|\eta| < 1.5$ in Figure 20. It also includes E735 data at $\sqrt{s} = 1.8$ TeV ($|\eta| < 1.62$) which agrees more with the UA5 data in the tail. Nevertheless, E735 confirmed KNO scaling in $|\eta| < 1.62$ based on their data; however, only by comparing the distributions in KNO variables and not by studying the moments [67]. A final conclusion about the slight KNO violation in $|\eta| < 1.5$ cannot be made at present.

Ref. [36] compares multiplicity distributions in full phase space from E735 and UA5 at three different energies (see the top left panel of Figure 21). The distributions disagree especially in their tails. This inconsistence has been frequently quoted. However, it is important to consider that the data from E735 is extrapolated from $|\eta| < 3.25$ to full phase space which may imply a significant systematic uncertainty. Also the data from UA5 are extrapolated, however, starting from $|\eta| < 5$. This is less a problem as, e.g., an estimation based on Pythia at $\sqrt{s} = 900$ GeV shows that 86% (64%) of the particles are in $|\eta| < 5$ (3.25). A direct comparison of data from E735 and UA5 in an $\eta$-region where both detectors are sensitive is therefore very interesting. The top right and bottom left panel of Figure 21 show such a comparison for approximately equivalent $\eta$-regions: $|\eta| < 1.5$ (1.62) and $|\eta| < 3.0$ (3.25) for UA5 (E735). The bottom right panel of Figure 21 shows the comparison in full phase space. Due to the fact that the E735 data is only available in KNO variables, the UA5 data is shown superimposed in KNO variables, too. No scaling correction has to be applied due to the different $\eta$-regions because the data is shown in KNO variables and thus already scaled with the average multiplicity.
The shown band corresponds to the region where data points are in the original figure which is not of the best quality [67, Figure 2]. Note also that this band corresponds to data points from $\sqrt{s} = 300$ GeV to 1.8 TeV. The band is most likely overestimating the error bars of the single points. It can be seen that in both $\eta$-intervals the distributions agree within errors (apart from very low multiplicities in the smallest $\eta$-region). Going to full phase space the discrepancy occurs. Hence, a systematic effect in the extrapolation procedure may be suspected as the cause of the discrepancy.

Furthermore, it should be noted that CDF and E735 do not agree too well in (approximately) same phase space regions, see Figure 22. Looking at the distribution

Figure 21. Top left panel: Comparison of UA5 and E735 data in full phase space at three different energies. Data from [8, 63, 36]. Other panels: Comparison between UA5 [8] at $\sqrt{s} = 900$ GeV and E735 [67] in approximately equivalent $\eta$-regions (top right and bottom left panels) and in full phase space (bottom right panel). See the text for an explanation of the E735 band.
Charged-Particle Multiplicity in Proton–Proton Collisions

Figure 22. Data from CDF and E735 at $\sqrt{s} = 1.8$ TeV are compared. The right panel shows normalized residuals using the error of the E735 data (these have been smoothed over 5 data points to reduce fluctuations); the black curve compares the experiments directly; the dashed red curve considers in a very basic way the difference in $\eta$-range by moving the CDF data points by a factor $(1.57/1.5)$.

in KNO variables shows that CDF at 1.8 TeV agrees better with UA5 at 900 GeV than with E735 (plot not shown).

In summary, there are various experimental inconsistencies, especially in the tail of the distributions which has a significant influence on, e.g., the calculation of moments of higher rank. It will be interesting to compare data taken at the LHC with the already present distributions.

4. Predictions

The measurement of the charged-particle multiplicity at the LHC has the potential to improve our understanding of multi-particle productions mechanisms by rejecting incorrect models. In Figure 23 several predictions for $dN_{ch}/d\eta|_{\eta=0}$ and $\langle N_{ch} \rangle$ in full phase space are summarized. Predictions for $dN_{ch}/d\eta|_{\eta=0}$ range between about 4.0 – 6.5 at $\sqrt{s} = 7$ TeV and 4.5 – 8 at $\sqrt{s} = 14$ TeV. For the charged multiplicity in full phase space the range is about 55 – 75 at $\sqrt{s} = 7$ TeV and 65 – 95 at $\sqrt{s} = 14$ TeV. Measured values outside these ranges would be a big surprise.

The predictions can be classified in several classes. First there are the simple extrapolations of trends observed at lower $\sqrt{s}$ (CDF [65], Busza [95, 96]). The predictions based on the logarithmic extrapolations performed in Section 3.4 (only those that fit the data reasonably are shown) as well as those based on the $e^+e^-/p + p$ universality discussed in Section 3.5 also belong to this class. Then there are model predictions based on the assumption of gluon saturation (Armesto, Salgado, Wiedemann [97] and
Kharzeev, Levin, Nardi [98].) The QGSM model is a representative of a class of models for soft scattering which are based on Regge pole theory and the parton structure of hadrons [99]. In these models proton-proton interactions are described in terms of the exchange of color-neutral objects called Pomerons. This leads to the production of strings which fragment into the observed particles.

In many cases it is more practical to implement theoretical ideas in terms of Monte Carlo event generators. Phojet is such a generator which is based on the Dual Parton Model [100]. Based on the Pomeron picture Phojet accounts both for soft and hard interactions. Epos is another event generator that aims at consistently treating soft and hard interactions [101, 95]. This model has been compared and tested with data from high-energy cosmic rays. Epos can be run in a mode which allows the formation of a quark-gluon plasma in $p + p$ collisions. In the Pythia event generator [34] the picture of individual parton-parton scatterings, which successfully describes high-$p_T$ phenomena, is extrapolated to low $p_T$. Pythia has many parameters and several Pythia tunes exist which, e.g., nicely describe Tevatron data. The shown predictions are based on the default Pythia settings and three frequently-used tunes: A, D6T, and ATLAS-CSC which are the main tunes used by the CDF, CMS, and ATLAS experiments, respectively. Pythia 6.4.14 has been used with the structure functions CTEQ6L [102]. The large jump in $\sqrt{s}$ from the Tevatron to the LHC will unveil whether certain Pythia tunes really capture the underlying physics or whether they are just ad hoc descriptions at specific energies.

Predictions for the multiplicity distribution in full phase space and in a limited range of $|\eta| < 1$ are shown in Figure 24 for $\sqrt{s} = 7$ TeV and in Figures 25 and 26 for 14 TeV. Shown are the extrapolations of the two component model with NBDs for the three scenarios (from [29]), Pythia with the four aforementioned tunes, and Phojet. Furthermore, and only for 14 TeV in full phase space a prediction from QGSM [] and one found in the framework of a multiple-parton interpretation of the collision [38] (see Section 2.5.2).

The predictions differ significantly, which is most pronounced in the tails of the distributions where the deviation is more than an order of magnitude. This applies in full phase space as well as in limited $\eta$-ranges. However, also in the low-multiplicity region there are clear differences (see the right panels of Figure 25 and 26). This difference is less pronounced in $|\eta| < 1$. From $\sqrt{s} = 7$ TeV to 14 TeV the predicted differences increase further.

Multiplicity distribution measurements at larger $\sqrt{s}$ will allow to decide which models and MCs describe the data best. For the specific case of Pythia it is clear that the parameter space is very large and several combinations of parameters may describe the data equally well. Nevertheless the measurement of the multiplicity distribution (together with the $p_T$ spectrum and the correlation of $\langle p_T \rangle$ and the multiplicity) will allow to learn about the color correlations in the final-state [104].
Charged-Particle Multiplicity in Proton–Proton Collisions

Figure 23. Predictions for $dN_{ch}/d\eta|_{\eta=0}$ and $\langle N_{ch} \rangle$ in full phase space in $p+p$ collisions at $\sqrt{s} = 7$, 10, and 14 TeV. For $dN_{ch}/d\eta|_{\eta=0}$ predictions 6 and 7 [97, 98] it is not explicitly stated whether the predictions are for inelastic, NSD, or non-diffractive collisions; all other predictions [66, 95, 96, 99, 100, 103, 34] are for NSD events.

Figure 24. Predictions for the multiplicity distribution of NSD events at 7 TeV are shown in full phase space (left panel) and $|\eta| < 1$ (right panel). The 2NBD prediction is from [29], for details see text.
5. Summary

This review has summarized measurements of charged-particle multiplicity distributions and pseudorapidity densities in high-energy $p + p(\bar{p})$ collisions. Moreover, related theoretical concepts have been presented on a basic level. The multiplicity being a simple observable in itself cannot be easily described within QCD because it is related to soft interactions for which the strong coupling constant is large and perturbative methods
cannot be applied. The validity of the available theoretical descriptions has been assessed using data from collider experiments at center-of-mass energies over about two orders of magnitude, from 23.6 GeV to 1.8 TeV. The energy dependence of \( \langle N_{\text{ch}} \rangle \) and \( dN_{\text{ch}}/d\eta \big|_{\eta=0} \) shows that Feynman scaling is broken at available energies; the moment analysis shows that KNO scaling does not hold except possibly for very central and small regions of phase space for which a definitive conclusion is not possible. For large energies the single NBD does not fit the data anymore; a combination of two NBDs is more successful; however, there are caveats identifying general trends as function of \( \sqrt{s} \). Although rapidity and multiplicity distributions differ between \( p + p \) and \( e^+e^- \) collisions, their average multiplicities as function of \( \sqrt{s} \) show similar trends that can be unified using the concepts of effective energy and inelasticity. Open experimental issues have been discussed and predictions for the LHC energy realm have been enumerated and briefly described. Interestingly, models that all more or less describe average multiplicities and multiplicity distributions up to Tevatron energies make significantly different prediction for the LHC.

At the LHC multiplicity measurements together with other global event properties give input to distinguish between the wealth of different model predictions including those from popular Monte Carlo event generators. This will allow to reduce the amount of possible interpretations of the underlying physics. In particular it will deepen the understanding of multiple-parton interactions and hadronisation as LHC’s energy will allow for the first time to probe \( p + p \) collisions where multiple hard parton interactions are present in the average event. Understanding the underlying dynamics of multiparticle production is not only an interesting research topic in itself because the characterization of the underlying event is also prerequisite for more specialized studies of exotic and rare channels which LHC is aiming at. Interesting measurements lie ahead of us.

Appendix A. Derivation of Feynman scaling

In his paper [13], Feynman concludes that the mean number of particles rises logarithmically, but does not give a mathematic proof. However, one can assess the asymptotic behavior by rewriting Eq. (3) in the form of the invariant cross-section:

\[
\frac{1}{\sigma} E \frac{d^3\sigma}{dp_T d^2p_T} = f_i(p_T, x).
\]

\( \frac{1}{\sigma} \) The definition of the Feynman function is different in some publications (e.g. [105]), not considering the \( 1/\sigma \) term in Eq. (A.1). This approach, however, results in conclusions that are not confirmed by experiment. In detail compared to the results of the calculation presented in the following, the left sides of Eqs. (A.8) and (A.9) have to be multiplied by \( \sigma \).
$f_i$ factorizes approximately (found experimentally) and a normalization of $g_i$ is chosen such that
\[
\int f_i(p_T, x) d^2 p_T = f_i(x) \int g_i(p_T) d^2 p_T = f_i(x). \tag{A.2}
\]
Integration of Eq. (A.1) and application of Eq. (A.2) yields:
\[
\int \frac{1}{E} \frac{d^3 \sigma}{dp_x dp_T} d\tau = \left\langle N \right\rangle = \int f_i(p_T, x) \frac{d^3 p}{E} = \int f_i(x) \frac{dp_z}{\sqrt{W^2 x^2 + m_T^2}} \tag{A.3}
\]
where on the left side the definition of the invariant cross-section is used with the average particle multiplicity $\left\langle N \right\rangle$, and for $m_T$ an effective average-$p_T$ is used.

Rewriting in $x$ yields the expression used to prove Feynman’s hypothesis:
\[
\left\langle N \right\rangle = \int_{-1}^1 f_i(x) \frac{dx}{\sqrt{x^2 + \frac{m_T^2}{W^2}}}. \tag{A.4}
\]
The integral is symmetric because $f_i(x)$ is symmetric for collisions of identical particles. For other collision systems the integration can be performed separately for negative and positive $x$ and yields the same result. $f_i(x) \leq B$ is finite and bounded due to energy conservation. Furthermore, Feynman assumes that for $x = 0$ a finite limit is reached. Therefore:
\[
2 \int_0^1 f_i(x) \frac{dx}{\sqrt{x^2 + \frac{m_T^2}{W^2}}} \leq 2 \int_0^1 B \frac{dx}{\sqrt{x^2 + \frac{m_T^2}{W^2}}} \tag{A.5}
\]
\[
= 2B \ln \left(x + \sqrt{x^2 + \frac{m_T^2}{W^2}}\right) \bigg|_0^1 \tag{A.6}
\]
\[
= 2B \ln \left(1 + \sqrt{1 + \frac{m_T^2}{W^2}}\right) - 2B \ln \frac{m_T}{W} \tag{A.7}
\]
The first term can be shown to be constant for $W \to \infty$ and the second is proportional to $\ln W$.

In consequence, Feynman scaling implies that the average total multiplicity scales as
\[
\left\langle N \right\rangle \propto \ln W \propto \ln \sqrt{s}. \tag{A.8}
\]
Considering that the maximum reachable rapidity in a collisions increases also with $\ln \sqrt{s}$, and under the further assumption that the particles are evenly distributed in rapidity, it follows that:
\[
\frac{dN}{dy} = \text{const.} \tag{A.9}
\]
Appendix B. Relation of NBD and BD

This section shows that a NBD becomes binomial when \( k \) is a negative integer. To start with the NBD and the binomial distribution (BD) are recalled. The NBD is:

\[
P_{(n),k}(n) = \binom{n + k - 1}{n} \left( \frac{\langle n \rangle/k}{1 + \langle n \rangle/k} \right)^n \frac{1}{(1 + \langle n \rangle/k)^k} \tag{B.1}
\]

with \( n \) failures and \( k \) successes. The BD is

\[
P_{p,m}(n) = \binom{m}{n} p^n (1-p)^{m-n} \tag{B.2}
\]

with \( m \) trials, \( n \) successes, and success probability \( p \). Important is that for both, the NBD and the BD, the running variable is \( n \). Using Eq. (17), the NBD is rewritten as

\[
P_{(n),k}(n) = \prod_{i=k}^{n+k-1} \frac{i}{n!} \left( \frac{\langle n \rangle/k}{1 + \langle n \rangle/k} \right)^n (1 + \langle n \rangle/k)^{-k-n}. \tag{B.3}
\]

If we identify \( m \) with \(-k\) and \( p \) with \( \langle n \rangle/k \) we find:

\[
\prod_{i=-m}^{-m-1} \frac{i}{n!} (-p)^n (1-p)^{m-n}. \tag{B.4}
\]

Assuming that \( n - m < 0 \) and \( -m < 0 \), all terms in the product are negative and the following relation holds:

\[
\prod_{i=-m}^{n-m-1} i = (1)^{n-m-1} \prod_{i=-m}^{n-m-1} |i| = (1)^n \prod_{i=1}^{n-m-1} i \tag{B.5}
\]

When \( m \) is positive and \( n \leq m \) the limits of the product become:

\[
| - m | = m \quad \text{and} \quad | n - m - 1 | = | -(m - n + 1) | = m - n + 1 \tag{B.6}
\]

and Eq. (B.5) results in:

\[
= (-1)^n \prod_{i=m-n+1}^{m} i. \tag{B.7}
\]

Eq. (B.4) is then:

\[
= (-1)^n \prod_{i=m-n+1}^{m} i (-p)^n (1-p)^{m-n} = \frac{\prod_{i=m-n+1}^{m} i}{n!} p^n (1-p)^{m-n}. \tag{B.8}
\]

Applying Eq. (17) the first term can be identified as the binomial term of the BD:

\[
\binom{m}{n} = \frac{\prod_{i=m-n+1}^{m} i}{n!}. \tag{B.9}
\]

Thus the NBD with negative integer \( k \) is a BD with the Bernoulli probability \( p = -\langle n \rangle/k \) and the number of trials \( m = -k \). For such a BD the assumptions made above are indeed fulfilled: \(-m < k < 0 \) \((m > 0)\) follows trivially; the number of successes \( n \) is smaller or equal than the number of trials \( m \) and therefore also \( n - m - 1 < 0 \). It is required that \( 0 < p < 1 \), thus \( 0 < \langle n \rangle < -k \).
Appendix C. Uncertainties on Moments

Given a distribution $P(n)$ which is normalized to 1 with an uncertainty $e_n$, and assuming that the errors on the individual bins are uncorrelated (which may not be the case after an unfolding procedure is applied, see also Section 3.1.2) their errors can be calculated using the partial derivatives:

\[
\frac{\partial C_q}{\partial P(n)} = \frac{n^q \langle n \rangle - \langle n^q \rangle q n}{\langle n \rangle^{q+1}}, \tag{C.1}
\]

\[
\frac{\partial F_q}{\partial P(n)} = \frac{n(n-1)...(n-q+1)\langle n \rangle - \langle n(n-1)...(n-q+1) \rangle q n}{\langle n \rangle^{q+1}}, \tag{C.2}
\]

\[
\frac{\partial D_q}{\partial P(n)} = \frac{\langle (n-\langle n \rangle)^{q-1} \rangle}{q} \left[ \frac{q n \langle n \rangle^{q-1}}{\langle n \rangle^{q+1}} + (n-\langle n \rangle)^q \right]. \tag{C.3}
\]

The total is then

\[
E_q^2 = \sum_n \left( \frac{\partial X_q}{\partial P(n)} e_n \right)^2 \tag{C.4}
\]

where $X_q$ is $C_q$, $F_q$, or $D_q$.

Acknowledgements

References

Charged-Particle Multiplicity in Proton–Proton Collisions
