TAMPERE UNIVERSITY OF TECHNOLOGY

Bayesian Classification of Hadronic Diffraction in the Collider Detector at Fermilab

Mikael Mieskolainen
Master of Science Thesis

Examiners: Prof. Ulla Ruotsalainen, Prof. Risto Orava

Thesis subject and examiners approved by
Faculty Council on 4th Dec, 2013

Abstract

Tampere University of Technology
Department of Signal Processing

MIKAEL MIESKOLAINEN: Bayesian Classification of Hadronic Diffraction in the Collider Detector at Fermilab
Master of Science Thesis, 111 pages, Appendix 14 pages
1st of March 2013
Major: Signal Processing (TUT), Particle Physics (ETH Zürich)
Examiners: Prof. Ulla Ruotsalainen, Prof. Risto Orava
Supervisor: PhD. Tuula Mäki
Keywords: Standard Model, QCD, Proton Structure, High Energy Diffraction, Regge Theory, Good-Walker Image of Diffraction, Hadron Collider, Tevatron, CDF, Probabilistic Multivariate Analysis, Multinomial Logistic Regression, L1-Regularization

Diffraction is fundamentally a wide scale phenomena, and well understood from macroscopic mechanical waves up to quantum mechanical electron diffraction. However, hadronic diffraction is still missing a rigorous quantum field theoretical formulation, but it can be experimentally probed in high energy accelerators. Because diffraction is inherently a coherent process, it allows a unique perspective to probe partonic inner structure of protons (hadrons) and relativistic space-time evolution of high energy hadron-hadron collisions.

In this thesis a Bayesian, probabilistic multivariate approach is developed for experimentally classifying diffractive hadronic scattering events from non-diffractive. For each measured collision event, the algorithm assigns a finite probability to an event to belong to a diffractive or non-diffractive process class. By integrating these probabilities over the full data sample, the interaction probabilities, known as cross sections, are estimated for different processes. The approach is Bayesian because it partly relies on the theoretical prior knowledge of cross sections.

This probabilistic way is shown to be a sound approach, because hard event-by-event decisions are both theoretically and experimentally not uniquely definable. The reasons for this are thoroughly explained in this thesis. The underlying algorithm is based on ℓ_1-norm regularized multinomial logistic regression. This regularization is shown to provide a mathematical view to the de-facto experimental physical signature of hadronic diffraction, known as the large rapidity gap.

The experimental part of the thesis is done with proton-antiproton data collected in the CDF run II experiment at the center of mass collision energy $\sqrt{s} = 1.96$ TeV at Fermilab. For the first time major components of the proton-antiproton scattering total cross section are estimated using a multivariate algorithm. The obtained cross sections for single diffractive $\sigma_{SDL} = (4.87 \pm 1.06)$ mb, $\sigma_{SDR} = (4.83 \pm 1.04)$ mb, double diffractive $\sigma_{DD} = (6.16 \pm 1.93)$ mb and non-diffractive $\sigma_{ND} = (45.20 \pm 1.59)$ mb match the phenomenological theory predictions within errors. Results of the thesis indicate that the probabilistic approach is viable, and emphasize also the importance of experimental forward (small-angle) instrumentation that is limited at the CDF detector.

Acknowledgements

This thesis was written in professor Risto Orava's experimental particle physics group at University of Helsinki during September 2012 - February 2013. First steps were taken in the summer of 2011, when working as a summer student at CERN via Helsinki Institute of Physics. The algorithmic work continued a little during the summer of 2012 at CERN, and was paused for a while due to examination session at ETH Zürich in August.

I would like to thank professor Orava for the unique opportunity to work at CERN and the original idea of the "soft" multivariate classification. I was given free hands to design and do the analysis, which was a matching approach. Thanks to Erik Brücken for pre-processing the raw CDF data sample and the Monte Carlo simulations. Thanks also to Tuula Mäki for co-supervising the thesis writing and to professor Ulla Ruotsalainen for examining the thesis.

Finally, before one can get the degree out - man måste tala flytande svenska. Thanks to Veronica Wallängen from Stockholm for helping me out with that!

Contents

Frequent abbreviations and terms

Barn, mb	A unit of cross section (area), 1 millibarn (mb) = 10^{-27} cm^2
BFKL	Balitsky-Fadin-Kuraev-Lipatov parton evolution scheme
CD	Central Diffractive scattering
CDF	Collider Detector at Fermilab
CMS	Center of Mass Lorentz frame
CP(T)	Charge-Parity(-Time) reversal transformation
DD	Double Diffractive scattering
DDIS	Diffractive Deep Inelastic Scattering
DGLAP	Dokshitzer-Gribov-Lipatov-Alterelli-Parisi parton evolution scheme
DIS	Deep Inelastic Scattering
EL	Elastic scattering
EW	Electroweak
Gauge	Theory with local continuous redundant degrees of freedom
Geant	Detector geometry and electrical response modelling software
INEL	Inelastic scattering
Jet	Collimated "spray" of particles
LHC	Large Hadron Collider at CERN
MAP	Maximum a Posteriori estimation
MC	Monte Carlo, physics simulation software chain
ML	Maximum Likelihood estimation
MLR-ℓ_1	Multinomial Logistic Regression with ℓ_1-norm regularization
N(N)LO	Next (to Next) to Leading Order calculation
ND	Non-diffractive scattering
Pomeron	Vacuum exchange object in Regge theory
PU	Pile-up, i.e. multiple simultaneous hadron pair collisions
PYTHIA	Collision event physics simulation software
QCD	Quantum Chromodynamics
QED	Quantum Electrodynamics
QFT	Quantum Field Theory
Rapidity gap	A region in rapidity (or pseudorapidity) without any final states
Regge	Pre-QCD theory of strong interactions
SD(L/R)	Single Diffractive scattering left or right forward direction
SM	The Standard Model of Particle Physics
Tevatron	Proton-Antiproton collider complex at Fermilab
TOT	Total scattering (cross section)
UE	Underlying Event
0-bias	Bunch crossing triggered and collected data

Frequent symbols and operators

ξ	Fractional forward (longitudinal) momentum loss of the surviving proton
$A^\dagger$	Complex conjugate transpose, Hermitian adjoint
$\langle f\|$	Final state particle, bra-vector
$\|i\rangle$	Initial state particle, ket-vector
$\langle f\|i\rangle$	Inner product between initial state and final state vectors
$\langle x\rangle$	Expectation value of a variable x
x	Bjorken-x, momentum fraction carried by a parton, $0 \leq x \leq 1$
$\mathbf{x}$	Real valued vector, $\mathbf{x} \in \mathbb{R}^d$
p^μ	Contravariant 4-momentum vector
p_μ	Covariant 4-momentum vector
$\vec{p}$	3-momentum vector
$g_{\mu\nu}$	Metric tensor of the Minkowski space, $g_{\mu\nu} = g^{\mu\nu} = \mathrm{diag}(+1, -1, -1, -1)$
∂_μ	Partial derivative operating on 4-vectors
$W_{\mu\nu}$	Tensor operator with two lower (covariant) indices, type of (0,2)
$SU(N)$	Special unitary Lie group with real dimension $N^2 - 1$
$\mathfrak{su}(N)$	Lie algebra with the corresponding Lie group $SU(N)$
α	Electromagnetic coupling "constant", $\alpha(Q^2) \triangleq e^2(Q^2)/4\pi$
α_s	Strong force coupling "constant", $\alpha_s(Q^2) \triangleq g_s^2(Q^2)/4\pi$
σ_X	Integrated cross section of a process X
s	Mandelstam variable, center of mass energy squared
t	Mandelstam variable, 4-momentum transfer squared
Q^2	4-momentum transfer squared $\sim$ squared energy scale of the process
p_T	Transverse momentum (with respect to the beam axis)
E_T	Transverse energy
$F_{1,2}$	Proton structure functions
$f_{q,\bar{q},g}$	Parton distribution functions
$\mu^2_{F,R}$	Factorization scale (F), renormalization scale (R)
ϵ^{ijk}	Totally antisymmetric Levi-Civita permutation tensor of rank 3
$\mathcal{M}^4$	Minkowski 4-dimensional flat space-time
M_X^2	Invariant mass squared of a system X
η	Pseudorapidity
y	Rapidity
Δy	Rapidity gap, $\Delta y \triangleq \|y_1 - y_2\|$
$\int \mathcal{L}\, dt$	Integrated luminosity
$\triangleq$	"Defined as"
λ	ℓ_1-regularization strength parameter
$f_j(\mathbf{x})$	Likelihood function for a j-th class
P_j	Prior probability for a j-th class
ψ	Dirac 4-spinor fermion field
$\bar{\psi}$	Dirac adjoint 4-spinor fermion field, $\bar{\psi} = \psi^\dagger\gamma^0$
γ^μ	Dirac's gamma matrices (4×4), $\mu = 0, 1, 2, 3, (5)$
$\psi_{L/R}$	Left (right) handed spinors, $\psi_{L/R} = \frac{1}{2}(1 \mp \gamma^5)\psi$
$\mathcal{O}(\alpha_s^n)$	The order n of perturbation in α_s
$\{\cdot\}$	Set brackets
$(\cdot, \cdot)$	Tuplet brackets
$[A, B]$	Commutator of operators A and B

Chapter 1

Introduction

The physics of strong interactions is modelled using non-Abelian $SU(3)$ gauge field theory called Quantum Chromodynamics (QCD). Hadrons, like protons or neutrons, are composite objects made of quarks and interactions between these quarks are propagated by gluons. The so-called *confinement* property of QCD guarantees that quarks are never seen free. Out of hadrons, baryons are made of three valence quarks and mesons of two. However, the structure is actually much more complicated as inside the hadrons there are also so-called sea quarks and gluons which become evident with higher energies.

In this thesis, a probabilistic multivariate classification framework is developed for the difficult problem of classifying diffraction in high energy physics. That means, proton-antiproton $p\bar{p}$-collisions are classified into different major scattering processes, i.e. diffractive or non-diffractive. Unfortunately, in practise the classification is an ill-posed inverse problem, especially experimentally. The classes are not uniquely definable, thus makes it an ill-posed problem in the sense of classic Hadamard definition. This fact stands out as a need for an approach which takes this into account.

Understanding diffraction is a major part of the ultimate goal, which is to have a complete picture of hadronic dynamics and the structure of hadrons. Diffraction in high energy physics can be approximately categorized into *soft* and *hard* diffraction based on 4-momentum transfer. Diffraction is predominantly soft and thus perturbative QCD is not a solution for theoretical predictions due to divergent power series expansion. Classic phenomenological approaches are based on Regge theory where soft diffraction is described with a Pomeron exchange, which is an object with quantum numbers of vacuum. Monte Carlo parametrizations are however very different, and there is no real consensus of how hadronic diffraction should be really treated.

The key of our approach is in probabilistic output, which is used to weight physical variables according to their probabilities to belong to different classes, e.g to signal or background processes. This gives more accurate distributions of physical variables. The framework itself is general and transparent enough to be used with several kind of particle physics analysis where currently Boosted Decision Trees, Neural Networks and traditional hand-tuned variable histogram cuts dominate.

The most straightforward way of utilizing our classification framework is to calculate estimates of relative cross sections for different scattering processes by integrating obtained probabilities. Given a priori measured total inelastic cross section σ_{inel} or integrated luminosity $\int \mathcal{L}\, dt$, absolute cross sections can be estimated. The

first results of this approach with real data are presented, using the CDF run II experiment 0-bias $p\bar{p}$-run collected at $\sqrt{s} = 1.96$ TeV center of mass energy. The CDF run I experiment measured earlier the cross sections of single diffraction [34] and double diffraction [3] at $\sqrt{s} = 1.8$ TeV. However, comparison with the results from this thesis and CDF I results can be done using the higher energy predictions by MBR (Minimum Bias Rockefeller) event generator model in PYTHIA 8 [16] at $\sqrt{s} = 1.96$ TeV. This is because the MBR model is basically tuned to match the results of CDF run I measurements.

Our framework takes a pragmatic approach, i.e. the algorithms obey definition of diffraction described by the selected Monte Carlo generator. Thus, if there is a mismatch between classification results and the real world, allegations can be forwarded to the corresponding physics generator or detector response modelling. It should be emphasized that the framework developed in this thesis is fundamentally different way to measure diffraction than what has been done previously in traditional rapidity gap analyses [3,34]. This is because all the major inelastic scattering process classes are taken into account simultaneously and background corrections are obtained directly via probabilistic weighting.

The Standard Model is introduced in Chapter 2 and the basis of proton structure knowledge in Chapter 3. Then diffraction and diffraction as a quantum mechanical process are described in Chapters 4 and 5, respectively. A short hadron collider phenomenology round up is in Chapter 6 and the Tevatron hadron collider with the CDF experiment are introduced in Chapter 7. The mathematical basis for the multivariate analysis is given in Section 8 and the actual analysis algorithm, sparsity regularized multinomial logistic regression, is described in Chapter 9. Finally, the results with CDF data are given in Chapter 10, with relevant discussion and conclusions finally given in Chapter 11.

Chapter 2

The Standard Model in a nutshell

Particle physics is described by the Standard Model (SM) of particle physics, which is a gauge field theory. It can be considered as a triumph of theoretical and experimental physics, but it leaves several questions unanswered. The first building block of the SM was Quantum Electrodynamics (QED), which describes electromagnetic interactions at quantum level in a relativistic way. In theory, the classic Maxwell equations come out from QED as macroscopic, differential-geometrical descriptions of classic fields.

Next interaction included in the SM was the weak-interaction, which was combined together with the electrodynamic interaction to produce the theory of electroweak interactions (EW). The Higgs mechanism together with Yukawa interaction is used to describe how gauge bosons (force propagators) and fermions (matter particles) acquire mass. This involves the so-called spontaneous symmetry breaking of the electroweak theory.

Chronologically, the last thing to complete the SM was Quantum Chromodynamics (QCD), which describes strong interactions. The strong interactions are responsible for ordinary atomic nucleus not breaking apart, and describes the partonic structure of nucleons inside nucleus.

2.1 Symmetries and particles

Particle physics is based on an idea that interactions between particles and the conservations laws are coming from physical symmetries. Symmetries that are based on transforms can be continuous (like translation) or discrete (like parity inversion). The famous Noether's theorem says that for every continuous symmetry, there is a conserved quantity [25].

Space-time symmetries

For space-time continuous symmetries, this means that momentum conservation is a result of invariance under space translations. In a same way, energy conservation comes from invariance under time translations. Together, these can be represented mathematically as 4-momentum conservation and they form Abelian Lie group of translations in the Minkowski space-time $\mathcal{M}^4$. In practise, highly important relation is the dispersion relation [45]

$$E^2 = |\vec{p}|^2 + m^2 \tag{2.1}$$

where $\vec{p}$ is the relativistic 3-momentum of the particle, m is the invariant mass and E is the total energy. In the rest frame $\vec{p} = \vec{0}$ and thus $E = m$, which is the Einstein's famous $E = mc^2$ in natural units ($c = 1, \hbar = 1$) (see Appendix A.1). By definition, in the center of mass (CMS) frame of particles $\sum_i \vec{p}_i = \vec{0}$. Note that for a system X of particles, e.g. $\gamma\gamma$ pair, the invariant mass m_X is not equal to the sum of individual invariant masses, which is identically zero in this case.

In addition to translations, also crucial are Lorentz boosts (moving frames) and rotations in the Minkowski space which are transformations of hyperbolic geometry of special relativity. Direct analogies are the classic Galileo transformations and rotations of the Euclidean space.

Combined group theoretically, all these space-time symmetries form the non-Abelian Poincaré symmetry group $P(1,3) = \mathbf{R}^{1,3} \rtimes SO(1,3)$ with 10 generators, which a semi-direct product between the group of translations and the Lorentz group. The irreducible representations of the Poincaré group are linked with the particles, with non-negative mass and integer or half-integer spin as indices. [45]

Local gauge symmetries

In the SM, internal symmetries generate the interactions between particles, and these symmetries are described by the gauge group made of Lie groups

$$G_{SM} = SU(3)_C \times SU(2)_L \times U(1)_Y, \tag{2.2}$$

with the corresponding Lie algebras $\mathfrak{su}(3), \mathfrak{su}(2)$, $\mathfrak{u}(1)$. By Noether's theorem, the corresponding conservation laws are the conservation of color charge, weak isospin and electric charge. For more details see Appendix A.3. The full symmetry group of the SM is thus [45]

$$P(1,3) \times G_{SM}. \tag{2.3}$$

The gauge symmetry is *local*, which means it depends on the space-time point $x \in \mathcal{M}^4$. This need for a local gauge symmetry is sometimes seen as a fundamental philosophy or an aesthetic goal of modern particle theories. It means that by any

means, field theories are usually built in a way that the gauge symmetry must not be broken.

The non-Abelian Lie group $SU(3)_C$ describes QCD gauge symmetry, the non-Abelian $SU(2)_L$ the weak force and the Abelian $U(1)_Y$ describes the gauge symmetry of the electromagnetic interactions. Together, $SU(2)_L \times U(1)_Y$ form the celebrated theory of electroweak interactions. The color charge of a particle describes how it transforms under $SU(3)$, the weak isospin T_3 how it transforms under $SU(2)$ and the weak hypercharge Y describes the transformations under $U(1)$. The last two are related to the electric charge Q by formula [29]

$$Q = T_3 + Y/2. \tag{2.4}$$

The electroweak theory is a chiral gauge theory. In practise, chirality is mathematically reflected by the fact that there are both left-handed $SU(2)$ *doublets* and right-handed *singlets*, which are explained later in the context of Lagrangians. In practise, only left-handed fermions or right-handed antifermions can interact in the charged current (CC) weak interactions via $W^{\pm}$-bosons. In contrast, the weak neutral currents (NC) can interact with right-handed fermions and left-handed antifermions via Z^0-boson. See Figure 2.3 for Feynman diagrams of basic interaction vertices. One must remember that the antiparticle of a left-handed fermion is right-handed, and the antiparticle of a right-handed fermion is left-handed. [15, 45]

Table 2.1: Differences between chirality and helicity, where γ^μ are Dirac gamma matrices, $\vec{\sigma}$ is a spin vector and $\vec{p}$ is a 3-momentum vector. [29]

Mass	Property	Chirality operator $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$	Helicity $\vec{\sigma} \cdot \vec{p}$
$m = 0$	Constant of motion	yes	yes
	Lorentz invariant	yes	yes
$m \neq 0$	Constant of motion	no	yes
	Lorentz invariant	yes	no

The handedness of a particle is given by its chirality operator, and for massless particles it equals particles helicity which is a dot product between particles 3-momentum and spin. Note that spin is not affected by Lorentz boosts, but 3-momentum naturally is. This equivalence between chirality and helicity does not hold for massive particles $m \neq 0$, because helicity is not a Lorentz covariant quantity. See Table 2.1 for a comparison. Helicity is not a constant of motion for massive particles because it does not commute with the total energy operator, i.e. with Hamiltonian H_{Dirac}. Mathematically that is $[\gamma_5, H_{Dirac}] \neq 0$. [29]

Gell-Mann is well-known for his quark model based on $SU(3)$ symmetry, or the eightfold-way, of up, down and strange quarks. However, this symmetry is only an approximation like the isospin symmetry of proton and neutron. This approximative symmetry is sometimes wrongly understood to be the same thing as the exact $SU(3)_C$ gauge symmetry of QCD, which describes how quarks and gluons come in different color combinations and interact. For this reason, Gell-Mann's symmetry is correctly called as *flavour* $SU(3)_F$ symmetry.

Discrete transformations

The fundamental discrete transformations in particle physics are charge conjugation C, parity transformation P (spatial mirror reflection) and time reversal T. An interesting thing is that the weak interactions can violate C, P and CP symmetries, which means that the probability for a given interaction is different before and after the transformation. Electromagnetic and strong forces are not known to violate these discrete symmetries. Any interaction in the SM should not violate CPT transformation, given that the CPT theorem is correct [45]. Parity transformation and charge conjugation operator eigenvalues of the particle are often written together with spin[1] J, as J^{PC}. For example, let us have a meson $|q\bar{q}\rangle$. Now use the charge (C-parity) and parity (P-parity) inversion operators to the meson wavefunction $|q\bar{q}\rangle$ or quantum field configuration as

$$\hat{C}|q\bar{q}\rangle = \pm|\bar{q}q\rangle \tag{2.5}$$

$$\hat{P}|q\bar{q}\rangle = \pm|q\bar{q}\rangle, \tag{2.6}$$

where $+$ means that the meson is parity "even", and $-$ that the meson is parity "odd".

Then, there are some "accidental" symmetries in the SM which are known as baryon and lepton number conservations, which are not as fundamental as other symmetries.

Particles

The elementary particles of the SM are written in Table 2.2 and in a more theoretical format in Appendix in Table A.1, where $SU(3)_C$ color triplet and $SU(2)_L$ weak isospin doublet structure is represented. The matter particles quarks and leptons come in three mass generations and obey Fermi-Dirac spin-1/2 statistics, i.e. Pauli's exclusion principle holds. Of these, only quarks flavours of the first family, up and down, are stable and build the universe together with electrons. The reason for three families is unknown, and the symmetry groups of the SM do not give any answer for this. [45]

Most of the mysteries within the SM are related to neutrinos, those nearly zero mass, highly non-interacting particles only interacting via weak interactions. Currently it is not known if neutrinos are their own antiparticles (Majorana fermion) or not (Dirac fermion), and can there be even a fourth generation of neutrinos. Also, right handed neutrino singlets $(\nu_l)_R$ have not been observed experimentally, and they might be so-called *sterile neutrinos*, i.e. not taking part in the Standard Model EW interactions. Neutrinos propagate in free space as a linear combination of *mass eigenstates* (ν_1, ν_2, ν_3) and interact and are created as *weak eigenstates* $(\nu_e, \nu_\mu, \nu_\tau)$. This change of flavour in free space is called *oscillation*, and it requires that their mass is non-zero, with differing values to have phase difference in propagation. However, currently their absolute masses are unknown. The neutrino mixing is described by 3×3 Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix, but its components are not well-known and it is not known if it has a complex phase, which leads to CP-violation in neutrino oscillation. [44]

[1] e.g. for ordinary mesons, $J = L+S$ is the total angular momentum, i.e. sum of orbital angular momentum L and intrinsic total spin S.

The interactions are mediated by gauge bosons which obey Bose-Einstein spin-0 statistics. The peculiar non-Abelian feature of QCD (and weak interactions) is that the corresponding gauge bosons, gluons (and $W^{\pm}$) carry color (electric) charge, and can interact together. This non-linear behaviour does not happen in electrodynamics, which is seen also in macroscopic Maxwell equations, where values of fields can be calculated by linear superposition. In terms of QED, the classic static electric and magnetic fields are at the quantum level mediated by virtual photons. The electromagnetic waves are mediated by real photons.

Table 2.2: Elementary particles and their properties. DF = Dirac fermion, MF = Majorana fermion, B = Boson. Fermions are grouped in three families. [44]

Name	Type	Charge	Color	Spin	Mass
Quarks					
$u;\bar{u}$	DF	$\pm 2/3$	$r,g,b;\bar{r},\bar{g},b$	1/2	2.4 MeV/c^2
$d;\bar{d}$	DF	$\mp 1/3$	$r,g,b;\bar{r},\bar{g},\bar{b}$	1/2	4.8 MeV/c^2
$c;\bar{c}$	DF	$\pm 2/3$	$r,g,b;\bar{r},\bar{g},b$	1/2	1.27 GeV/c^2
$s;\bar{s}$	DF	$\mp 1/3$	$r,g,b;\bar{r},\bar{g},\bar{b}$	1/2	104 MeV/c^2
$t;\bar{t}$	DF	$\pm 2/3$	$r,g,b;\bar{r},\bar{g},b$	1/2	171.2 GeV/c^2
$b;\bar{b}$	DF	$\mp 1/3$	$r,g,b;\bar{r},\bar{g},\bar{b}$	1/2	4.2 GeV/c^2
Leptons					
$e^-;e^+$	DF	∓ 1	-	1/2	0.511 MeV/c^2
$\nu_e,(\bar{\nu_e})$	DF/MF ?	0	-	1/2	< 2.2 eV/c^2
$\mu^-;\mu^+$	DF	∓ 1	-	1/2	105.7 MeV/c^2
$\nu_\mu,(\bar{\nu_\mu})$	DF/MF ?	0	-	1/2	< 0.17 MeV/c^2
$\tau^-;\tau^+$	DF	∓ 1	-	1/2	1.777 GeV/c^2
$\nu_\tau,(\bar{\nu_\tau})$	DF/MF ?	0	-	1/2	< 15.5 MeV/c^2
G-Bosons					
$W^{\pm}$	B	± 1	-	1	80.4 GeV/c^2
Z^0	B	0	-	1	91.2 GeV/c^2
γ	B	0	-	1	0
g	B	0	#8 (octet)	1	0
H^0	B	0	-	0	~ 126 GeV/c^2

The conjectured property of QCD, *confinement*, says that the quarks are never seen free, i.e. they are always trapped inside color singlet bound states, hadrons, where the quark color charges combined a "colorless", $SU(3)_C$ singlet state. Hadrons as non-elementary particles are QCD bound states made of three quarks or three antiquarks, also known as baryons. Mesons are bound states of quark-antiquark pairs. Baryons obey Fermi-Dirac statistics and mesons Bose-Einstein statistics. These composite particles have a rich mass spectrum and currently the only stable hadron seems to be proton $|uud\rangle$. However, some grand unified theories (GUT) predict proton half-life at 10^{36} years. The mass of hadrons is not simply mass of quarks, because the binding energy of QCD builds up the most of the hadron's mass. [45] This property makes also the concept of quark mass a matter of definition.

The SM cannot answer currently the questions about dark, non-luminous matter which should be there based on gravitational lensing and galactic rotational curves. This leads to supersymmetric (SUSY) models where each fermion acquires a bosonic

superpartner particle and vice versa. SUSY models can provide a stable, natural dark matter candidate particle, namely WIMP (weakly interacting massive particle). Unfortunately, currently there is no experimental evidence for supersymmetry.

Also unanswered question is the matter-antimatter asymmetry of the universe, i.e. why there is more matter than antimatter in the current universe. Finally, the theory of gravity, general relativity, is not included at all in the Standard Model. This unification requires currently extra-dimensional (string theory) models, which have their own major obstacles.

2.2 Couplings

The strength of three different interactions are described with coupling constants, which are not actually constants, but run with energy. Electromagnetic interaction strength, is given at electron mass scale by [45]

$$\alpha_{em}(Q = m_e) = \frac{1}{137}, \tag{2.7}$$

with this energy scale it is also known as the fine structure constant.

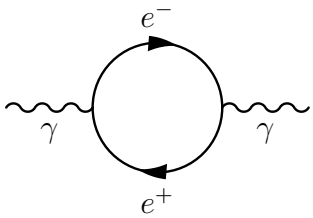


Figure 2.1: One loop Feynman diagram of QED vacuum polarization.

The running with energy is due to quantum scale effect known as *screening*, which is illustrated in Figure 2.1. In screening, the vacuum polarizes around an electrical charge with virtual dipole charges created "out of vacuum" and this makes interaction strength larger with smaller distances. This effect must not be mixed up with Coulomb potential (low energy, long distance) which goes as $V \propto 1/r$ over distance r from the charge, which comes purely from spherical geometry.

The weak interactions have much lower interaction strengths, which is logical given the name, and the *effective* strength at energies $Q \ll M_X$ is given by the Fermi theory constant [45]

$$G_F = 1.167 \times 10^{-5}\ \text{GeV}^{-2}, \tag{2.8}$$

where M_X is the mass of the massive vector bosons of weak interactions. The electroweak theory of the Standard Model uses dimensionless, renormalizable couplings g and g'.

The coupling of strong interactions $\alpha_s \triangleq g_s^2/4\pi$ is notoriously running with energy scale such as

$$\alpha_s(Q = m_{\text{hadrons}}) \sim 1 \ \text{ Confinement} \tag{2.9}$$

$$\alpha_s(Q \to \infty) \to 0 \ \text{ Asymptotic freedom} \tag{2.10}$$

This makes calculations based on perturbation theory impossible in QCD at low energy scales[2], i.e. series expansion is unstable/divergent. This feature of QCD is due to screening and *antiscreening*, and it turns out that at large distances the coupling α_s is large and vice versa. Like tension in a rubber band. Corresponding Feynman diagrams are shown in Figure 2.2.

The evolution of the coupling, both in QED and QCD, can be described with the so-called β-function expansion with renormalization group differential equation [29]

$$\mu_R^2 \frac{\partial(\alpha \text{ or } \alpha_s)}{\partial(\mu_R^2)} = \beta(\alpha \text{ or } \alpha_s), \tag{2.11}$$

[2]One must use e.g. numerical lattice QCD or phenomenological models.

where μ_R^2 is the renormalization energy scale $(\mathrm{GeV})^2$. It turns out that the leading order term β_0 in the perturbative power series is negative for QED and positive for QCD, which gives the opposite energy scale behavior of both forces.

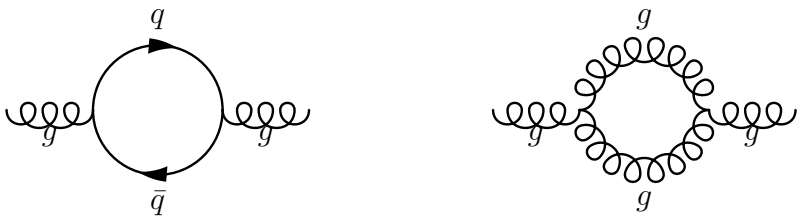


Figure 2.2: QCD screening (on the left) and antiscreening (on the right) at one loop level.

In grand unification theories, where the unified gauge group is e.g. $SU(5)$ or $SO(10)$, these couplings come together at some finite high energy scale. [5] At this scale, there would be only one force, including the gravity. However, usually these energy scales are beyond the realms of current collider experiments and the only hope to see some experimental evidence for these unifications comes from astrophysical processes.

2.3 Lagrangians

The mathematics of the Standard Model is written in a language of Lagrangians, which has roots in a variational formulation of classic mechanics. For a pedagogical introduction to the derivation of these equations, see [47]. The basic idea is that the corresponding Lagrangians for each interaction theory are derived under local gauge invariance. Algebraic generators of the gauge group give the number of gauge mediators (bosons), for general case of $SU(N)$, the number is given by N^2-1. Three for $SU(2)$ which are $W^{\pm}, Z^0$ and eight for $SU(3)$ which are gluons $g_\alpha, \alpha = 1, \ldots, 8$.

The Lagrangians can be written with highly compact or highly explicit formalism. In the following, a semi-compact notation is followed from [15, 24].

Electroweak

The Lagrangian density for the electroweak unification can be written in parts as [15]

$$\mathcal{L}_{EW} = \mathcal{L}_{gauge} + \mathcal{L}_{matter} + \mathcal{L}_{Higgs} + \mathcal{L}_{Yukawa}, \tag{2.12}$$

and it combines QED interactions together with weak interactions as the unification name suggests.

The gauge boson interactions are described by the gauge part [15]

$$\mathcal{L}_{gauge} = -\frac{1}{4}\vec{W}_{\mu\nu} \cdot \vec{W}^{\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} \tag{2.13}$$

with field strength tensors

$$W^i_{\mu\nu} = \partial_\mu W^i_\nu - \partial_\nu W^i_\mu - ig\epsilon^{ijk}W^j_\mu W^k_\nu, \quad i = 1, 2, 3 \tag{2.14}$$

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu, \tag{2.15}$$

where W^i_μ is the $SU(2)_L$ gauge field triplet with the weak isospin correspondence and B_μ is the $U(1)_Y$ singlet with the weak hypercharge correspondence and the partial derivative is defined as $\partial_\mu = (\partial_t, \nabla)$ and $\partial^\mu = (\partial_t, -\nabla)$. These are the *weak eigenstates* of gauge bosons, and the corresponding *mass eigenstates* are linear combinations of those. See vertices e.) and f.) in Figure 2.3 for the gauge interactions.

First, the massive charged current $W^{\pm\mu} \sim W^{\pm}$

$$W^{\pm\mu} = \frac{1}{\sqrt{2}}(W^{1\mu} \mp iW^{2\mu}) \tag{2.16}$$

and two neutral currents, massless $A^\mu \sim \gamma$ and massive $Z^\mu \sim Z^0$ are given by linear combination or an orthogonal rotation in $SO(2)$ by

$$\begin{pmatrix} A^\mu \\ Z^\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_w & \sin\theta_w \\ -\sin\theta_w & \cos\theta_w \end{pmatrix} \begin{pmatrix} B^\mu \\ W^{3\mu} \end{pmatrix}, \tag{2.17}$$

where θ_w is the Weinberg mixing angle and it can be expressed with the gauge boson masses and couplings as

$$\cos\theta_w = \frac{M_W}{M_Z} = \frac{g}{\sqrt{g^2 + g'^2}}, \tag{2.18}$$

where g and g' are the $SU(2)_L$ and $U(1)_Y$ running couplings, respectively. [15]

The matter Lagrangian, which includes fermion fields and their electroweak $V-A$ (vector - axial vector) interactions with gauge bosons are described by [15]

$$\mathcal{L}_{matter} = \sum_L \bar{L}\gamma^\mu \underbrace{\left(i\partial_\mu + g\frac{1}{2}\vec{\tau}\cdot\vec{W}_\mu + g'\frac{Y}{2}B_\mu\right)}_{\triangleq D_\mu} L + \sum_R \bar{R}\gamma^\mu \left(i\partial_\mu + g'\frac{Y}{2}B_\mu\right) R, \tag{2.19}$$

where L is a left-handed fermion doublet and R is a right-handed fermion singlet. See vertices a.) - d.) in Figure 2.3 for an illustration. The covariant derivative taking care of gauge invariance is denoted with D_μ.

The three Pauli matrices are denoted with $\vec{\tau}$ and γ^μ are the Dirac gamma matrices. Doublets have weak isospin $T_3 = (+1/2, -1/2)$ and singlets have $T_3 = 0$. The quark mixing and the probabilities of flavour changing decays are described by famous Cabibbo-Kobayashi-Maskawa (CKM), 3×3 unitary matrix, with three mixing angles and CP-violating complex phase (not shown here). The transition probability from up type quark i to down type quark j, is proportional to the CKM matrix element $|U_{ij}|^2$. [44]

The Higgs part and the corresponding potential term $V(\phi)$ with the complex $SU(2)$ doublet $\phi(x)$ (4 degrees of freedom) are [15]

$$\mathcal{L}_{Higgs} = (D_\mu\phi)^\dagger(D^\mu\phi) - V(\phi) = |D_\mu\phi|^2 - V(\phi) \tag{2.20}$$

$$V(\phi) = \mu^2\phi^\dagger\phi + \lambda(\phi^\dagger\phi)^2, \tag{2.21}$$

which gives the mass to massive gauge bosons, i.e. to Z^0 and $W^\pm$ and leaves photon massless. The three degrees of freedom of ϕ are "transmuted" by massive gauge bosons, and only one is left $\Rightarrow$ scalar Higgs field. The free parameters $\mu^2 < 0$ and $\lambda > 0$ control the shape of the Higgs "mexican hat" potential and λ works also as the quartic self coupling term. The doublet ϕ including the Higgs field $h(x)$ after this so-called *spontaneous symmetry breaking* (SSB) of EW theory is

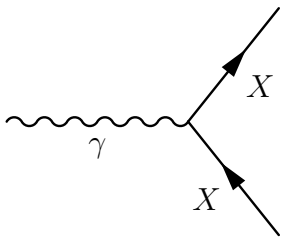
$$\phi(x) = \begin{pmatrix}\phi^+\\ \phi^0\end{pmatrix} \overset{\text{SSB}}{\mapsto} \frac{1}{\sqrt{2}}\begin{pmatrix}0\\ v + h(x)\end{pmatrix} \tag{2.22}$$

with the non-zero vacuum expectation value $v = 246$ GeV, which can be theoretically obtained by using the measured Fermi's constant G_F and the W mass. However, the mass of the Higgs boson H^0, the neutral, scalar (spin-0) excitation particle of the field itself with $J^{PC} = 0^{++}$, cannot be calculated from the theory but must be experimentally "scanned". Only some restrictions can be given for its mass.

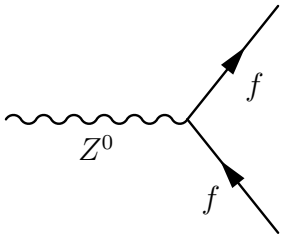
The final part is the Yukawa coupling part to give fermions their mass by interaction with the Higgs doublet [15]

$$\mathcal{L}_{Yukawa} = -\sum_{f-} G^f_-(\bar{L}\phi R + \bar{R}\bar{\phi}L) - \sum_{f+} G^f_+(\bar{L}\phi^c R + \bar{R}\bar{\phi}^c L), \tag{2.23}$$

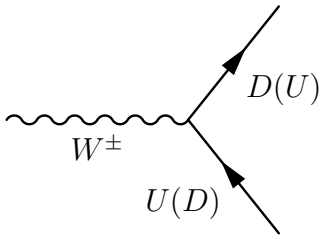
where ϕ^c is the charge conjugate of the Higgs doublet, i.e. $\phi^c = (\bar{\phi}^0, -\bar{\phi}^+)$, necessary such that the "up" type quarks are able to get mass. The Yakawa fermion coupling is denoted with $G^f_\pm$, where $\pm$ corresponds with $T_3 = \pm\frac{1}{2}$.



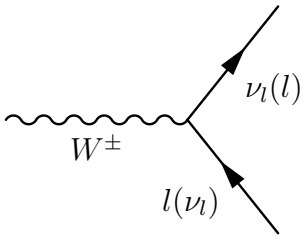
a.) QED photon interaction, X has charge.



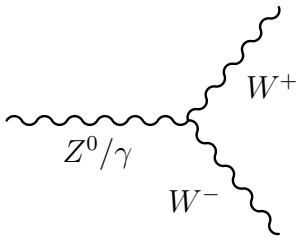
b.) EW neutral current, f is any fermion.



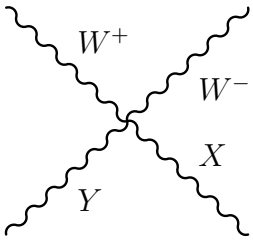
c.) EW flavour-changing current, U is "up" type, D is "down" type quark.



d.) EW charged current with a lepton-neutrino pair.



e.) EW non-Abelian 3-boson vertex.



f.) EW non-Abelian int., X and Y EW bosons obeying charge conservation.

Figure 2.3: EW interaction vertices excluding Higgs. Note that $W^{\pm}$ couples only to left-handed fermions or right-handed antifermions.

These relations about masses can be shown to hold at tree level [15]: $M_W = gv/2, M_Z = M_W/\cos\theta_w, M_f = vG^f/\sqrt{2}$ and $M_H = v\sqrt{2\lambda}$. So the last unknown parameter in the Standard Model (excluding neutrino oscillations) was the mass of the Higgs M_H (~ 126 GeV), and thus the value of λ.

After all this, as expected, $\mathcal{L}_{EW}$ is still invariant under local $SU(2)_L \times U(1)_Y$ transformations and particles have obtained masses in a gauge invariant way by "spontaneous symmetry breaking". Fixing the masses of weak bosons by hand in the Lagrangian would break the gauge invariance [15] and the driving principle (gauge invariance) of all modern field theories would be suspicious.

Quantum Chromodynamics

The full Lagrangian density of QCD can be written in parts as

$$\mathcal{L}_{QCD} + \mathcal{L}_{gauge\,fix} + \mathcal{L}_{ghosts}. \tag{2.24}$$

The first part without gauge fixing and ghosts at the classical level is [24]

$$\mathcal{L}_{QCD}(x) = -\frac{1}{4}F^a_{\lambda\rho}(x)F_a^{\lambda\rho}(x) + \sum_q \bar{q}(x)(i\gamma^\lambda D_\lambda - m_q)q(x), \tag{2.25}$$

where $q(x)$ $SU(3)_C$ triplet describes the quark matter fields in Minkowski space $x \in \mathcal{M}^4$ with masses m_q for the six flavours $q \in \{u, d, c, s, t, b\}$. The gluon field strength tensors are given by [24]

$$F^a_{\lambda\rho}(x) = \partial_\lambda A^a_\rho(x) - \lambda_\rho A^a_\lambda(x) - g_s f_{abc} A^b_\lambda(x) A^c_\rho(x) \tag{2.26}$$

with the eight different gluon potentials $A^a_\lambda(x), a = 1, \ldots, 8$. The strength of coupling is described by the bare coupling g_s, which is here a non-physical (non-renormalized) quantity due to screening. The gauge group $SU(3)$ structure constants (scalars) are f_{abc}, which are given in Appendix A.3.

Now the covariant derivative, in a similar fashion as in EW theory is [24]

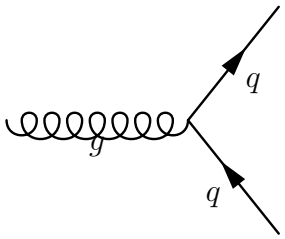
$$D_\lambda q(x) = (\partial_\lambda + i g_0 A^a_\lambda T_a) q(x), \tag{2.27}$$

where generators $T_a = \frac{1}{2}\lambda_a$ are given by the eight 3×3 Gell-Mann matrices $\lambda_a, a = 1, \ldots, 8$ of the $SU(3)$ group, again see Appendix A.3. Finally, the gluon potential and field-strength matrices are

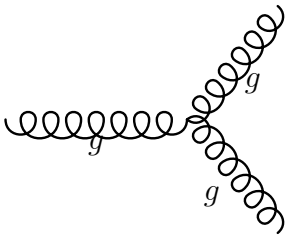
$$A_\lambda(x) = A^a_\lambda(x) T_a \tag{2.28}$$

$$F_{\lambda\rho}(x) = F^a_{\lambda\rho}(x) T_a, \tag{2.29}$$

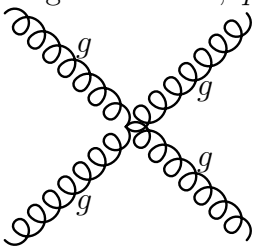
which are just potentials and field strengths contracted over all eight generators T_a, $a = 1, \ldots, 8$. Easily can be shown that $\mathcal{L}_{QCD}$ is invariant under local $SU(3)$ transformations which depend on $x \in \mathcal{M}^4$, also the mass term of quarks [24]. See Figure 2.4 for an illustration of QCD interaction vertices.



a.) QCD strong interaction, q is a quark.



b.) QCD 3-gluon non-Abelian strong interaction.



c.) QCD 4-gluon non-Abelian strong interaction.

Figure 2.4: QCD interaction vertices.

Description above is valid at the level of classical fields and for non-Abelian theory, like QCD, the most convenient way to turn this into quantum field theory is by functional integration approach also known as path integral quantization. For more information, see e.g. [24]. Ultimately, this allows one to derive Feynman rules and

the perturbation theory as power series in the coupling. However, several technical difficulties (infinite quantities) are encountered when one tries to calculate higher than leading order predictions, and the need for renormalization arises which involves suitable regularization techniques. This is because every loop in a Feynman diagram is basically a divergent 4-momentum integral. Also, when the QCD coupling is large enough, the perturbation theory cannot be used. This in contrast with QED, which does not have similar problems with a large running coupling.

When selecting or fixing the "gauge", one needs to add the so-called Faddeev-Popov ghost fields which are necessary auxiliary fields in non-Abelian quantum field theories. In general, the Feynman rules are made of external (lines) and internal propagators of fermions and bosons and interaction vertex factors, which are derived from the Lagrangian of the theory. In QCD, these are the gluon propagator, the quark propagator and the ghost propagator. The coupling constant is included the vertex factors, which are the three-gluon vertex, four-gluon vertex, quark-gluon vertex and ghost-gluon vertex. The renormalization procedure can also add several counter-term rules to mitigate the infinities. [24]

In QED, as in QCD, the propagators are described by Green's functions usually in momentum space and they have the same role as in classic field theories and in non-relativistic quantum mechanics. In QED we can also use the Fourier decomposition to describe the solution to systems as a sum of individual plane-waves, because the theory is Abelian, i.e. superposition principle holds. However, in QCD, gluons can interact with themselves and thus, the theory is also non-linear with respect to gauge boson interactions (like EW) and no superposition cannot be used in general. [29] However, QED, EW and QCD are all non-linear theories in a sense that the only general way to calculate anything analytically is to do perturbative expansions. Finally, there is always the fact that QCD describes the fields in terms of quarks and gluons, not actual observables, which are hadrons.

2.4 S-matrix and cross sections

The golden era of the S-matrix (scattering matrix) theory was in the 1960s, and originally it was meant to be *the theory* of particle physics. Nowadays it is understood more abstraction than a suitable computational tool for every problem. It could explain bound states, i.e. hadrons, as poles in complex domain etc., but failed to answer questions in more physical terms, especially in comparison with the parton image and QCD. [6]

Let there be the initial state ket-vector $|i\rangle$ and the final state bra-vector $\langle f|$. The *probability* of this transition from $|i\rangle \to \langle f|$ is described by the S-matrix element, also known as *scattering amplitude* $\langle f|S|i\rangle \in \mathbb{C}$. The transition probability is [6]

$$P_{fi} = |\langle f|S|i\rangle|^2 = \langle i|S^\dagger|f\rangle\langle f|S|i\rangle \tag{2.30}$$

and the set of orthonormal states $|f\rangle$ obeys the completeness relation

$$\sum_f |f\rangle\langle f| = 1, \tag{2.31}$$

so the probability of producing any of the final states, given some initial state, must be equal to one

$$\sum_f |\langle f|S|i\rangle|^2 = \sum_f \langle i|S^\dagger|f\rangle\langle f|S|i\rangle = \langle i|S^\dagger S|i\rangle = \langle i|i\rangle = 1. \tag{2.32}$$

This must be true if want to conserve probability, which means that $S^\dagger S = 1$, thus it is a unitary operator.

Now the scattering matrix S can be related to the transition matrix T in process of 2-body to n-bodies [6]

$$\langle f|S|i\rangle = \langle p_1' p_2' \dots p_n'|S|p_1 p_2\rangle = \underbrace{\delta_{fi}}_{\text{no int.}} + \underbrace{i(2\pi)^4\delta^4(p^f - p^i)}_{\text{4-momentum cons.}} \underbrace{\langle f|T|i\rangle}_{\text{amplitude}}, \tag{2.33}$$

where p^i and p^f are the 4-momentum of the initial state and final state, respectively. These $2 \to n$ processes and their kinematics were much studied in the 1960s, before QCD which moved the interest to hard exclusive $2 \to 2$ processes.

Cross sections and decay widths

Basically, the most fundamental measurements done in particle physics are the measurements of cross sections and decay widths (particle lifetimes). By definition

$$\begin{aligned} \text{Cross section} &\triangleq \frac{\#\text{ of transitions } 12 \to n \text{ per unit time}}{\#\text{ of particles per unit surface and time}} \\ &\propto \int \text{Kinematics} \times \text{Dynamics} \\ &\sim \textit{Probabilities}. \end{aligned}$$

The total cross section for a scattering process $12 \to n$, given in Lorentz invariant form, is in general [24]

$$\begin{aligned}\sigma_{12\to n} &= \frac{1}{F}\sum (2\pi)^4\delta^4(p^f - p^i)|\langle f_n|T|i\rangle|^2 \\ &= \frac{1}{F}\int \left(\prod_{i=1}^{n}\frac{d^4p_i'}{(2\pi)^4}2\pi\delta^+(p_i'^2 - m_i^2)\right)(2\pi)^4\delta^4\left(\sum_{i=1}^{n}p_i' - p_1 - p_2\right) \\ &\times |\langle p_1' \cdots p_n'|T|p_1p_2\rangle|^2 \\ &= \frac{1}{F}\int \left(\prod_{i=1}^{n}\frac{d^3p_i'}{2E_i(2\pi)^3}\right)(2\pi)^4\delta^4\left(\sum_{i=1}^{n}p_i' - p_1 - p_2\right)|\mathcal{M}_{fi}|^2, \end{aligned} \tag{2.34}$$

where the kinematical, Lorentz invariant flux factor is

$$F = 4\sqrt{(p_1 \cdot p_2)^2 - m_1^2m_2^2} \to 2s, \text{ when } s \gg m_1^2, m_2^2 \tag{2.35}$$

and δ^4 is the Dirac's delta in Minkowski space used to keep track of 4-momentum conservation. So one needs a flux factor, an integral over the phase space $\sim$ kinematics, and the matrix element squared inside the phase space integral $\sim$ dynamics, to be able to answer questions about integrated cross sections.

In a similar way, the total decay width of a particle decay $a \to 1 + 2 + \cdots + n$ to n-particle final state is given by [15]

$$\Gamma_a = \sum_{i=1}^{n}\Gamma_i = \frac{1}{2E_a}\int \left(\prod_{i=1}^{n}\frac{d^3p_i'}{2E_i(2\pi)^3}\right)(2\pi)^4\delta^4\left(p^f - p^i\right)|\mathcal{M}_{fi}|^2 \tag{2.36}$$

which is clearly not a Lorentz invariant quantity due to E_a, thus particle lifetimes $\tau_a = 1/\Gamma_a$ are by definition expressed in the rest frame of the decaying particle a.

In a case of simple $2 \to 2$ scatterings, which are processes described by the given theory of dynamics like QED, QCD, EW, SUSY etc., the Lorentz invariant matrix element squared $|\mathcal{M}_{fi}|^2$ in the integrand of Equation 2.34 can be calculated with the Feynman rules of the perturbation theory at least on the tree level. The integration limits for phase-space integral are usually straightforward, but require care.

Chapter 3

Proton structure

First experimental measurements to probe hadronic structure were done in the 1960s at SLAC (Stanford Linear Accelerator) experiment using electron-proton scattering. Later, similar experiments were done in the 1990s at HERA (DESY, Hamburg) and these measurements about proton structure are used to predict scattering cross sections at LHC (Large Hadron Collider). This process of electron-proton scattering, where the proton breaks in parts in the collision and new particles are produced from the energy, is known as as deep inelastic scattering (DIS) [15]. SLAC experiments were a strong motivation for parton model of Feynman which describes proton as a composite object of point-like particles when probed with high spatial resolution.

Measurements in DIS showed partons to be *charged* spin-1/2 objects. From the naive parton model of Feynman, theoretical development lead finally to QCD partons which have the postulated *asymptotic freedom*, i.e. interaction between partons should vanish at small distances (large momentum transfers). These charged QCD partons are nowadays called quarks. However, electrically uncharged gluons inside protons play an important role especially at hadron colliders.

Probing the structure of proton using electron is different based on energy of the electron. 1. First at very low energies $\lambda \gg r_p$ the scattering happens from a point-like object with no spin. 2. At low energies $\lambda \sim r_p$ the scattering happens from an extended charge distribution and the resolution λ is close to the proton's classic[1] radius $r_p \sim 10^{-15}$ m. 3. At high energies $\lambda \sim 10^{-18}$ m $< r_p$ scattering happens from the valence quarks which dictate the proton quantum numbers. 4. Finally, at very high energies $\lambda \ll r_p$ the proton looks to be sea of quarks, and indirectly, sea of gluons. [15]

[1] Classic, because proton's radius is quantum mechanically a probability density.

3.1 Deep inelastic scattering

In deep inelastic scattering, a lepton is used to probe distributions of partons inside a hadron. Let the target proton with mass M be at rest in the lab frame, with 4-momentum by definition $p = (M, \vec{0})$. The electron has 4-momentum in the initial state $k_i = (\omega_i, \vec{k_i})$ and in the final state $k_f = (\omega_f, \vec{k_f})$. All coming derivations are based only on leading order quantum electrodynamic (QED) interactions between electron and proton, with single virtual photon γ^* exchange. In general, when using EW theory, 4-momentum in DIS can be also propagated by a Z^0 boson, or via charged current through $W^\pm$ bosons. [15]

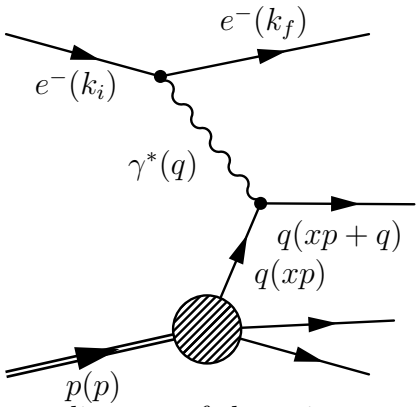


Figure 3.1: Feynman diagram of the naive parton model DIS.

The first Lorentz scalar, a frame independent quantity, is 4-momentum transfer Q^2 or *virtuality* of the exchanged boson, which can be understood as the resolution power of probing the hadron structure as $\lambda \sim 1/\sqrt{Q^2}$. It is defined as

$$Q^2 = -q^2 = (k_i - k_f)^2 = -(\nu, \vec{q})^2, \tag{3.1}$$

where k_f and k_i are the 4-momentum vectors of the lepton in the final state and the initial state, respectively. The energy ν transferred to the hadron by the scattering lepton is [15]

$$\nu = \frac{p \cdot q}{M}, \tag{3.2}$$

where p and M are the 4-momentum and the invariant mass of the hadron, respectively.

Now the very important, the so-called *Bjorken scaling* dimensionless x-variable is [15]

$$x = -\frac{q^2}{2p \cdot q} = \frac{Q^2}{2M\nu} = \frac{Q^2}{Q^2 + W^2},\ 0 \leq x \leq 1, \tag{3.3}$$

and the research field of low-x QCD has its name based on this. Another useful variable is the so-called *inelasticity* [15]

$$y = \frac{p \cdot q}{p \cdot k_i},\ 0 \leq y \leq 1, \tag{3.4}$$

which is not to be confused with rapidity y. The Bjorken[2] $x \in [0, 1]$ can be understood as the *momentum fraction* of the hadron carried by the parton in a frame

[2]Not to be confused with the kinematical Feynman x_F variable, often used in heavy ion physics.

where the hadron has very high energy, i.e. in the infinite momentum frame. Originally, Bjorken described these ideas in the abstract language of current algebra, which was a theoretical tool of the 1960s.

By using these variables and 4-momentum conservation, the invariant mass squared W^2 of the scattered hadronic system is described as

$$W^2 = (p+q)^2 = M^2 + 2M\nu - Q^2, \tag{3.5}$$

which is the same as hadronic center of mass energy squared. The elastic scattering case, where the initial and final states of the hadron are the same, can be easily derived by setting $x = 1$ by

$$\begin{aligned} W_{el}^2 &= M^2 + 2M\nu - Q^2 \\ &= M^2 + 2M\nu - x2M\nu \\ &= M^2, \end{aligned} \tag{3.6}$$

which means that the scattered proton has the same invariant mass as the original.

Now, the hadronic tensor to describe the proton structure is [15]

$$W_{(p)}^{\mu\nu} = -W_1 g^{\mu\nu} + \frac{W_2}{M^2} p^\mu p^\nu + \frac{W_4}{M^2} q^\mu q^\nu + \frac{W_5}{M^2}(p^\mu q^\nu + q^\mu p^\nu), \tag{3.7}$$

where $g^{\mu\nu}$ is the Minkowski metric tensor. The $W_{(p)}^{\mu\nu}$ is a rank-2 tensor, where the electroweak part W_3 is left out due simplicity and because of small EW interaction cross sections. Functions $W_i, i = 1, \ldots, 5$ are based on Lorentz invariant scalars and they describe the internal structure of the proton. By reformulation [15]

$$W_{(p)}^{\mu\nu} = W_1(-g^{\mu\nu} + \frac{q^\mu q^\nu}{q^2}) + \frac{W_2}{M^2}(p^\mu - \frac{p\cdot q}{q^2} q^\mu)(p^\nu - \frac{p\cdot q}{q^2} q^\nu) \tag{3.8}$$

Then, by using the leptonic tensor [15] and the hadronic tensor, the scattering matrix element squared is obtained by contraction [15]

$$|\mathcal{M}|^2 = L_{\mu\nu}^{(e^-)} W_{(p)}^{\mu\nu} = 4EE'(W_2(Q^2,\nu)\cos^2(\theta/2) + W_1(Q^2,\nu)\sin^2(\theta/2)). \tag{3.9}$$

Differential cross section in the lab-frame (proton at rest) is obtained by including flux F, differential phase-space factor $d\Pi$ and matrix element squared $|\mathcal{M}|^2$, formally

$$d\sigma = \frac{1}{F} d\Pi |\mathcal{M}|^2 \tag{3.10}$$

and integrating over the energy of the electron in the final state results in [15]

$$\left.\frac{d\sigma}{d\Omega}\right|_{\text{lab}} = \frac{\alpha^2}{4E^2\sin^4(\theta/2)} \frac{E'}{E}(W_2(Q^2,\nu)\cos^2(\theta/2) + W_1(Q^2,\nu)\sin^2(\theta/2)), \tag{3.11}$$

which is the differential cross section for deep inelastic electron-proton scattering.

Above purely magnetic W_1 and electro-magnetic W_2 are the so-called *structure functions*, which describe the momentum distribution of the quarks inside the proton. The *independence* of the structure functions on Q^2 is the famous *Bjorken scaling*, which states asymptotically [15]

$$F_1(x) = \lim_{Q^2\to\infty} M W_1(Q^2,\nu) \tag{3.12}$$

$$F_2(x) = \lim_{Q^2\to\infty} \nu W_2(Q^2,\nu). \tag{3.13}$$

However, Bjorken scaling is known to break down and scaling violations arise when QCD corrections are included into this naive parton image. It was still a major breakthrough and one of the most important things invented in the 1960s in particle physics, and it was a strong argument to support the parton model.

These Q^2 independent structure functions $F_{1,2}(x)$ obey also the so-called Gallan-Gross relation [15]

$$F_2(x) = 2xF_1(x), \tag{3.14}$$

which is expected for scattering from spin-1/2 quarks. In a case quarks would be spin-0 particles, the purely magnetic structure function $F_1(x)$ would be $F_1(x) = 0$. This is coming from the fact that the spin is a property which is magnetic by its nature.

3.2 Partons - quarks and gluons

In the proton $|uud\rangle$, the valence up quarks carry twice the momentum of the valence down quarks. However, experimentally can be seen that the quarks carry just a little over 50 percent of the total momentum [15]. The rest is carried by gluons and these as electrically neutral objects do not effect the electron-proton scattering. The probability $f_i(x)$ for the parton to carry a fraction x of the momentum is described by *parton distribution function* (PDF), and in order to conserve probability

$$\sum_i \int_0^1 dx\, x f_i(x) = 1, \tag{3.15}$$

where i runs over different quark (antiquark) flavour pdfs and a gluon pdf. This means that in total the parton distribution functions include the valence quarks, the sea quarks, which are virtual and produced by the gluons splitting $g \to q\bar{q}$, and gluons itself.

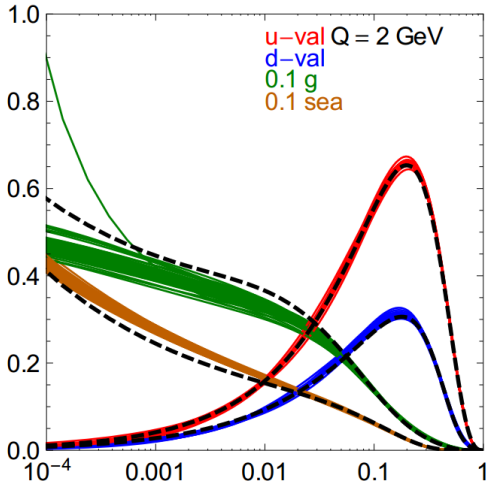


Figure 3.2: CT10-NNLO parton distributions for proton at $Q = 2$ GeV. Axis variables are $(x, xf_i(x, Q^2))$. [43]

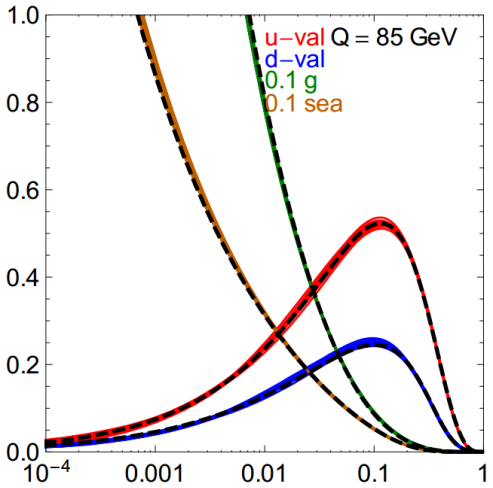


Figure 3.3: CT10-NNLO parton distributions for proton at $Q = 85$ GeV. Axis variables are $(x, xf_i(x, Q^2))$. [43]

Feynman's parton model uses the parton distribution functions to construct the proton's structure functions as [15]

$$F_2(x) = \sum_i e_i^2 x f_i(x) \tag{3.16}$$

$$F_1(x) = \frac{1}{2x} F_2(x), \tag{3.17}$$

where the second line uses Gallan-Gross relation and e_i is the fractional electric charge of the given quark flavour. In a case that the proton is probed at low energies, parton distribution functions only of up, down and strange quarks, and corresponding antiquarks can be used as an approximation. Figures 3.2 and 3.3 show a next-to-next-to-leading order (NNLO) parton distribution calculations at $Q = 2$ GeV and $Q = 85$ GeV, respectively. Gluon and sea quarks distributions are scaled down for visualization with a factor of 10. At higher energies and low-x, the

contribution especially from gluons, but also from sea quarks (due to $g \to q\bar{q}$) grows significantly larger than from the valence quarks. This means directly that colliders like the Large Hadron Collider (LHC), are effectively gluon colliders.

3.3 QCD corrections

When the contribution from gluons is included, i.e. naive parton model + gluons, at order $\mathcal{O}(\alpha\alpha_s)$ in QED and QCD, there is a change for processes where the virtual photon probes the quark, and this quark emits a gluon after interaction $\gamma^* q \to qg$. Or a gluon splits into quark-antiquark pair and virtual photon probes one of these quarks. These break down the Bjorken scaling, and the structure function must be modified $F_2(x) \to F_2(x, Q^2)$. Scaling violations become evident with small values of x, which were not experimentally available at the time in SLAC experiment. [15]

These QCD corrections are taken into account by redefining the quark pdf distribution $q(x)$ in the infrared region where soft gluons are emitted. This is called *mass factorization* at scale μ_F^2, and this scale is ultimately related with the value of Q^2. Mass factorization of quark density is expressed as [15]

$$f_q(x, \mu_F^2) = f_q(x) + \frac{\alpha_s}{2\pi} \int_x^1 \frac{d\xi}{\xi} f_q(\xi) P_{qq}(\frac{x}{\xi}) \log \frac{\mu_F^2}{\mu^2}, \tag{3.18}$$

where $f_q(x, \mu_F^2)$ is the actual measurable quark density, which due to QCD vacuum screening is different than the bare, non-physical $f_q(x)$. The function $P_{qq}(z)$ is a splitting kernel, which describes the probability for gluon emissions with a fraction z of the quark momentum. The integral is evaluated due to unknown and unmeasurable gluon emission with transverse momentum p_T obeying

$$\mu_F^2 \geq p_T^2 \geq \mu^2, \tag{3.19}$$

where μ^2 is an arbitrary infrared cutoff scale, and can be basically set always below what is experimentally achievable in resolution [15].

Finally, QCD corrected structure function with famous logarithmic scaling corrections is [15]

$$\begin{aligned} &F_2(x, Q^2) \\ &= x \sum_q \int_x^1 \frac{d\xi}{\xi} f_q(\xi, \mu_F^2) e_q^2 \left(\delta(1 - \frac{x}{\xi}) + \frac{\alpha_s}{2\pi} P_{qq}(\frac{x}{\xi}) \log \frac{Q^2}{\mu^2} - \frac{\alpha_s}{2\pi} P_{qq}(\frac{x}{\xi}) \log \frac{\mu_F^2}{\mu^2} \right) \\ &= x \sum_q \int_x^1 \frac{d\xi}{\xi} f_q(\xi, \mu_F^2) e_q^2 \left(\delta(1 - \frac{x}{\xi}) + \frac{\alpha_s}{2\pi} P_{qq}(\frac{x}{\xi}) \log \frac{Q^2}{\mu_F^2} \right) \end{aligned} \tag{3.20}$$

and after all, the arbitrary infrared cutoff scale μ vanished and only the factorization scale parameter μ_F is left. Now, if the factorization scale is set as $\mu_F^2 = Q^2$, which is a valid factorization scale in DIS, the structure function is

$$F_2(x, Q^2) = x \sum_q e_q^2 f_q(x, Q^2), \tag{3.21}$$

where sum runs over different flavours of quarks and antiquarks. Now there is a dependency on both x and Q^2, and the energy scale Q^2 dependency is ultimately due to gluons. Also can be seen that this whole procedure of mass factorization is closely related to defining the running strong coupling $\alpha_s(Q^2)$ using QCD β-function. The same screening plays here a major, physically same role.

3.4 DGLAP and BFKL evolution

The factorization of QCD means that the perturbative, hard short distance parton-parton scattering, and all soft long distance non-perturbative phenomena can be separated, i.e. factorized. However, there are some cases where factorization breaks down, but in general it gives very good predictions for many high energy processes. Factorization introduces an adjustable scale μ_F^2 and "absorbs" all non-perturbative, soft phenomena related to hard scattering below this scale into renormalised parton densities. In a collinear emission factorization scheme, this simple means a collinear cut off energy [15]. For a long introduction to different factorization approaches, see [19].

The famous DGLAP (Dokshitzer, Gribov, Lipatov, Altarelli, Parisi) or Altarelli-Parisi evolution equations describe how the parton distribution functions $f_i(x, Q^2)$ of quarks and antiquarks evolve when x is *fixed* and Q^2 is evolved. It is a very important part of the factorization of hadronic scatterings. There are several different splitting "kernels", which describe e.g. quark radiating a gluon before interaction. The gluon distribution function $f_g(x, Q^2)$ describes how the hadron can emit gluons which might turn into a quark-antiquark pairs before interaction. The DGLAP integro-differential equations are written as [15]

$$\frac{\partial}{\partial \log(\mu_F^2)}\mathbf{u}(x, \mu_F^2) = \frac{\alpha_s(\mu_F^2)}{2\pi}\int_x^1 \frac{dz}{z}\,\mathbf{P}(z)\mathbf{u}(\frac{x}{z}, \mu_F^2), \tag{3.22}$$

where $\mathbf{P}(z, \alpha_s(\mu_F^2))$ is the splitting kernel 2×2 matrix. At order $\mathcal{O}(\alpha_s)$, four different splitting kernels can be written in terms of perturbative QCD [15]

$$\mathbf{P}(z) = \begin{pmatrix} P_{qq}(z) & P_{qg}(z) \\ P_{gq}(z) & P_{gg}(z) \end{pmatrix}, \tag{3.23}$$

where each $P_{ji}(z)$ describes the probability for $i \to j$ splitting with transverse momentum fraction z. The quark and gluon distributions are written inside the vector

$$\mathbf{u}(x, \mu_F^2) = \begin{pmatrix} f_q(x, \mu_F^2) \\ f_g(x, \mu_F^2) \end{pmatrix}. \tag{3.24}$$

Using the DGLAP equation starts by choosing some initial value $\mu_F = \mu_0^2$ which is safely on the perturbative domain of QCD, where Q^2 should be approximately larger than 1 or 2 GeV^2 [24]. Then the parton distributions $\mathbf{u}(x, \mu_0^2)$ are fitted to the measurements for each fixed value of x. Finally, the Q^2 evolution of $\mathbf{u}(x, \mu_F^2)$ can be solved at any other scale μ_F^2 by applying analytically Mellin transformation to the DGLAP equations, or by using numerical integration techniques like Runge-Kutta. Unfortunately, there are some problems with low-x behaviour of DGLAP evolution, because it does not give safe predictions with very small values of x [24].

The BFKL (Balitsky, Fadin, Kuraev, Lipatov) is a similar kind of perturbative evolution scheme like the DGLAP. The main difference is that Q^2 is *fixed* and x is evolved. Unfortunately, BKKL is only applicable when x is small and Q^2 is large enough, and BFKL cannot handle $Q^2 \to \infty$ [24]. In general, phenomenology related to BFKL evolutions is much more controversial thing than of DGLAP, that is because BFKL is ultimately dealing closer to non-perturbative physics. Illustration of the parton evolution is shown in Figure 3.4. When using BFKL evolution with

fixed low value of Q^2, the number of partons increasing as a function of $1/x$, and finally leads to parton overlap in low-x saturation region, but their "size" stays the same. This saturation region is not yet experimentally well-probed. By using DGLAP evolution with fixed high value of x, the interpretation is that the number of partons increases but their size decreases as a function of Q^2. [21]

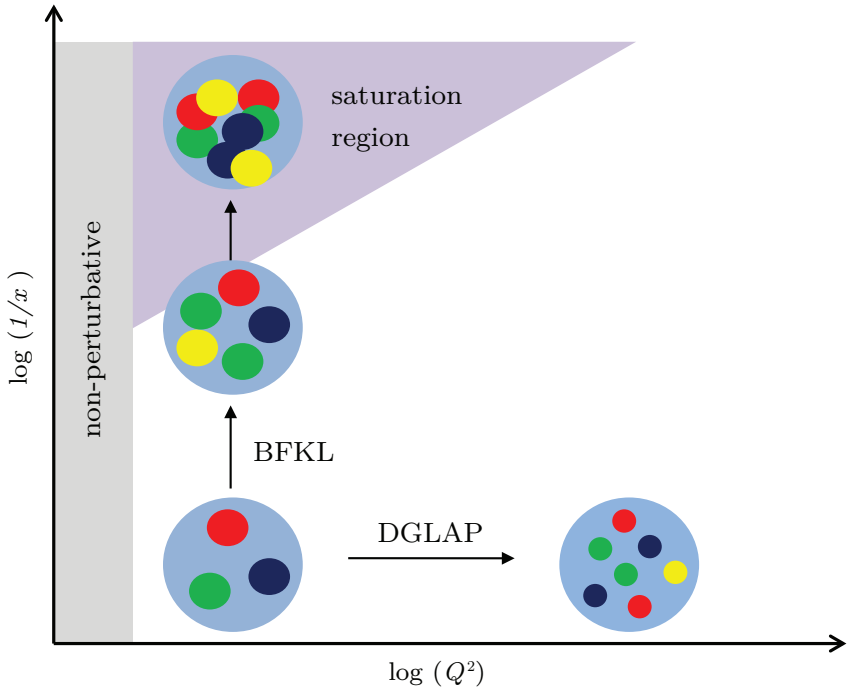


Figure 3.4: Perturbative parton evolution plane. Adapted from [21].

Other parton evolution related thing is the Sudakov suppression model, which describes the probability of a parton to evolve without branching. This is especially suitable for parton shower modelling in Monte Carlo. Also, there are interesting coherent phenomena related to soft gluon emission, which are angular ordering and coherent branching [25]. Angular ordering restricts soft gluon radiation by a heavy quark to be limited within a cone, which radius goes down as the evolution of *the jet* continues. Thus, the emitted gluons which split $g \to q\bar{q}$ are near in phase-space and the hadronization process where color-singlet objects are produced is assumed to happen locally. [25]

These QCD effects are analogous to phenomena which causes the Sudakov effect in QED, i.e. supression of soft bremsstrahlung photons from e^+e^- in electromagnetic cascades.

Chapter 4

Diffraction

The hadronic scattering processes are divided into soft and hard processes. The soft processes are processes where 4-momentum transfer squared between colliding particles is "small", $|t| \sim 0.5\,\mathrm{GeV}^2$, and the hard processes are the ones where the 4-momentum transfer squared is "large" $|t| \gtrsim 2\,\mathrm{GeV}^2$ [6]. Perturbative QCD can be used to describe these hard processes, but unfortunately the soft processes are beyond perturbative approaches. This is due to large interaction distances between hadrons which makes strong force coupling α_s too large.

The soft processes are elastic scattering, soft diffractive dissociation, and the most dominant, soft non-diffractive processes. These soft non-diffractive processes are also known as minimum bias processes, which is basically an experimental term coming from the detector trigger terminology. Hard processes produce e.g. high transverse momentum p_T-jets, electroweak vector bosons ($Z^0, W^{\pm}$) and possibly Higgs etc. Even if the hard processes can be treated under perturbation theory of QCD, there is a soft part of these processes which effect is included in the universal quark and gluon distribution functions.

The mass factorization theorem introduced in the previous chapter says that the soft and the hard part can be separated of each other, i.e. non-perturbative from perturbative, and the measured parton distribution functions can be nearly universally used in calculations of cross section estimates of the hard processes. However, the factorization in the context of hadronic *hard diffraction* is not well understood, and it is known to be a problematic concept. The diffractive-DIS and corresponding diffractive parton distributions functions, measured at HERA in $e^- p$-collisions, cannot be trivially applied to hadron-hadron hard diffraction, and it is known to lead to factorization break-down [18, 24].

4.1 Definition of diffraction

Def. I. Diffraction is a scattering process where quantum numbers of vacuum are exchanged between the colliding particles. Final states have a multiplicity spectrum narrow in the rapidity space. In a diffractive scattering process $A+B \to C+D$, quantum numbers are the same between A/C and B/D, because of vacuum exchange. [6]

Def. II. Diffractive scattering processes are characterized by large rapidity gaps (LRG), a region in the rapidity space with no final states, which is nearly an equivalent definition with the first one. [6]

Elastic (EL) scattering

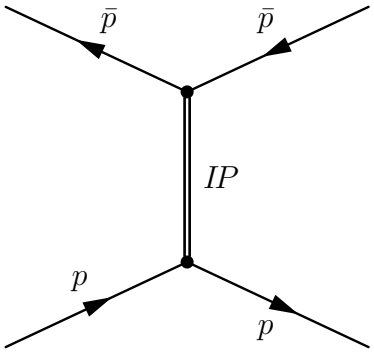


Figure 4.1: Feynman diagram of elastic scattering.

$$p\bar{p} \to p\dot{+}\bar{p}, \tag{4.1}$$

where both hadrons stay intact in collision and are possibly detected at small scattering angles using e.g. Roman pots. For now on, symbol $\dot{+}$ denotes "large" gap in rapidity Δy or experimentally, in pseudorapidity $\Delta\eta$. The usual measurement with elastic scattering is the differential elastic cross section as a function of $|t|$, where on low $|t|$, Coulomb (QED) interaction dominates and then QCD comes into image on higher $|t|$. The term *Pomeron IP*, a color singlet "vacuum object", perturbatively at leading order understood as a 2-gluon system, is used to describe the strong interaction between colliding protons as a t-channel exchange. [6]

Single Diffractive (SD) scattering

$$p\bar{p} \to p\dot{+}X, \tag{4.2}$$

where another proton stays intact and second one is excited into diffractive system denoted with X with the same quantum numbers as the incident particle. Mass of the system M_X depends on the fractional longitudinal momentum loss [6]

$$\xi = 1 - x_F = 1 - \frac{p_z^f}{p_z^i} = \frac{M_X^2}{s} \tag{4.3}$$

of the incident antiproton momentum, or proton, when the antiproton is disassociated into X. In Figure 4.2, the ξ is defined for antiproton.

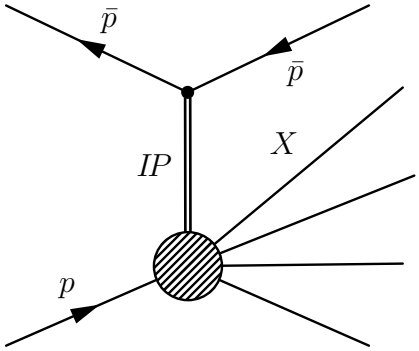


Figure 4.2: Feynman diagram of single diffractive scattering.

Experimentally, the double differential cross section spectrum of single diffractive scattering is approximately [6]

$$\frac{d^2\sigma_{SD}}{dM_X^2 dt} \propto \frac{1}{M_X^2} e^{-b|t|}, \tag{4.4}$$

where the slope parameter $b \in \mathbb{R}_+$ can in principle be a function of M_X^2 [48]. Low mass diffraction happens when the ξ is small, and the elastic scattering comes from the limit when $\xi \to 0$, then $M_X \to m_p$. The coherence condition [6] for single diffraction says that when $\xi \lesssim 0.05$, the rapidity gap Δy is approximately

$$\Delta y \simeq -\ln(\xi) \tag{4.5}$$

and substituting gives a minimum gap size of $\Delta y = -\ln(0.05) = 3$. Under coherence condition, also the maximum invariant mass of the diffractive system can be approximated to be

$$M_X = \sqrt{\xi}\sqrt{s} = \sqrt{0.05} \cdot 1960 \text{ GeV} \lesssim 430 \text{ GeV} \text{ (At Tevatron)}. \tag{4.6}$$

The fractional longitudinal momentum loss ξ of the surviving proton can be experimentally reconstructed within geometrical acceptances with at least two Roman Pot detector layer signals, to have a vector measurement, and by using accelerator optics transport equations.

Double Diffractive (DD) scattering

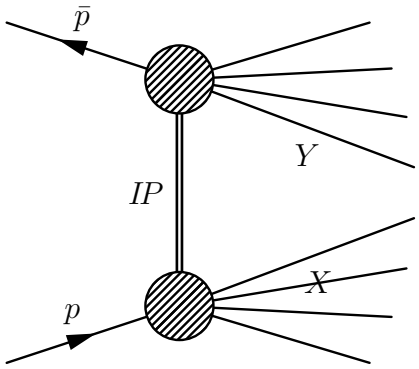


Figure 4.3: Feynman diagram of double diffractive scattering.

$$p\bar{p} \to X \dot{+} Y, \tag{4.7}$$

is a process where both incoming protons are excited into diffractive system in both forward and backward region, with same the quantum numbers as the incident particles. Thus, there is no fractional longitudinal momentum loss ξ variable to be measured experimentally in a way there was in single diffraction. [6]

Double diffractive events have a tendency to have experimental signature close to non-diffractive events, i.e. to have a vanishing rapidity gap when system X and Y overlap in the rapidity space. This happens when both diffractive systems have large mass which happens when we approach the coherence condition.

Central Diffractive (CD) scattering

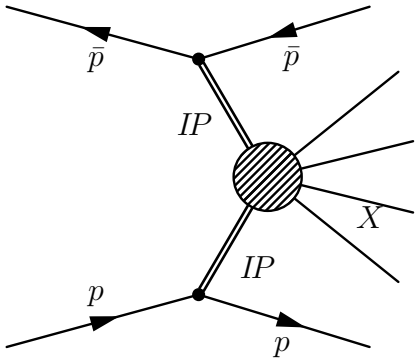


Figure 4.4: Feynman diagram of central diffractive scattering.

$$p\bar{p} \to p \dot{+} X \dot{+} \bar{p}, \tag{4.8}$$

also known as the double Pomeron exchange, which is a process with both forward and backward protons separated from the central system X. The X is a soft system or, in central exclusive production, for example a pair of $\gamma\gamma$, $\pi^+\pi^-$ or even hard scale objects like 2-jets ($q\bar{q}$) or Higgs H^0 [22, 23]. By using the mass-shell[1] condition, the

[1]Mass-shell, real particle with $q^2 = m^2$, not virtual off-shell with $q^2 \neq m^2$.

invariant mass of the central system is given by

$$[(p_i - p_f) + (\bar{p}_i - \bar{p}_f)]^2 = M_X^2. \tag{4.9}$$

which is the same as

$$M_X^2 = \xi_1 \xi_2 s, \tag{4.10}$$

where ξ_1, ξ_2 are the fractional longitudinal momentum losses. Note, that there is also a possibility that the forward protons are excited into a resonance N^*, e.g. $N(1440)$-baryon with decay $N \to \pi^0 p$.

One interesting exclusive state is the bound state of gluons, known as *glueball* [4], which is predicted by QCD but not yet observed. Based on lattice simulations, it should exist at very low masses of under 2 GeV/c^2.

Non-Diffractive (ND) scattering

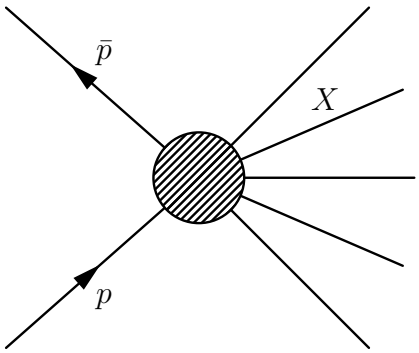


Figure 4.5: Feynman diagram of non-diffractive scattering.

$$p\bar{p} \to X, \tag{4.11}$$

where X is a system of soft particles nearly uniformly produced in rapidity with $dN/dy \cong constant$, and central hard scattering end products with high transverse momentum p_T. In contrast to diffractive interactions, "color is exchanged" between colliding hadrons, i.e. the strong interaction is propagated in a t-channel sense by color octet gluons. These hard scattering end products can be gluon or quark originated jets, or some energetic lepton pairs, like in the classic Drell-Yan process where quark and antiquark originating from protons annihilate via s-channel $q\bar{q} \to \gamma^*(Z^0) \to l^- l^+$.

The highest probability to have high-p_T objects is in non-diffractive events, and most of the new physics is believed to be discovered via high-p_T objects. However, each of the diffractive process classes (except elastic) can produce also hard objects, but with much lower cross sections than soft objects. This is known as hard or semi-hard diffraction. Hard diffractive jet events are measured at CDF to be in order of 1-2% of all jet events [1, 2].

Total cross section

These scattering processes contribute to the total $p\bar{p}$ cross section defined here as

$$\sigma_{tot} \triangleq \sigma_{el} + \sigma_{inel} \tag{4.12}$$

$$= \sigma_{el} + \sigma_{sd} + \sigma_{dd} + \sigma_{cd} + \sigma_{nd}. \tag{4.13}$$

These cross sections are in order of tens of millibarns at Tevatron $\sqrt{s} = 1.96$ TeV (see section 10.10), except the central diffraction which is believed to be below one millibarn.

Measuring the total cross section σ_{tot} is done in practise by utilization of optical theorem and Roman Pots or similar in-beam pipe detectors. The optical theorem gives a relation between the forward $\theta_{CM} = 0$ scattering amplitude and the total cross section. By using the definition of the T-matrix 2.33 and completeness relation 2.31, gives

$$\langle j|Ti\rangle - \langle j|T^\dagger|i\rangle = (2\pi)^4 i \sum_f \delta^4(p^f - p^i)\langle j|T^\dagger|f\rangle\langle f|T|i\rangle. \tag{4.14}$$

Then, taking the case $j = i$ of elastic scattering and rewriting the left hand side as

$$\text{Re }\langle i|T|i\rangle + i\,\text{Im }\langle i|T|i\rangle - (\text{Re }\langle i|T|i\rangle - i\,\text{Im }\langle i|T|i\rangle) = 2\,\text{Im }\langle i|T|i\rangle \tag{4.15}$$

results in an interesting formula

$$2\,\text{Im }\langle i|T|i\rangle = \sum_f (2\pi)^4 \delta^4(p^f - p^i)|\langle f|T|i\rangle|^2. \tag{4.16}$$

Because the right hand side is the total transition rate, the total cross section is obtained by dividing with the flux factor F (Equation 2.35) as

$$\frac{2\text{Im }\langle i|T|i\rangle}{F} = \frac{\sum_f (2\pi)^4 \delta^4(p^f - p^i)|\langle f|T|i\rangle|^2}{F}$$

$$\frac{2\text{Im }\langle i|T|i\rangle}{F} = \sigma_{tot}$$

$$\frac{2\text{Im }\langle i|T|i\rangle}{4\sqrt{(p_1 \cdot p_2)^2 - m_1^2 m_2^2}} = \sigma_{tot}. \tag{4.17}$$

The transition amplitude $\langle i|T|i\rangle$ corresponds to elastic scattering with 4-momentum transfer squared $t = 0$, when $m_3 = m_1$ and $m_4 = m_2$. Thus, the optical theorem can be written as

$$\sigma_{tot} = \frac{\text{Im } A(s, t = 0)}{2\sqrt{(p_1 \cdot p_2)^2 - m_1^2 m_2^2}}, \tag{4.18}$$

in terms of elastic scattering amplitude $A(s, t) \in \mathbb{C}$ [24]. Experimentally, a slightly modified version is used and the experimental data is extrapolated down to unmeasurable $t = 0$.

4.2 Survival of the large rapidity gap

Experimentally it is difficult to measure these diffractive processes because the defining signature, the large rapidity gap, can be missing. The probability that the gap is filled by some soft spectator partons or bremstrahlung gluons is not theoretically well known. It is believed that with the LHC center of mass energies $\sqrt{s}$, the gap survival probabilities $\langle|S|^2\rangle \in [0,1]$ are under 0.1 for SD, DD and CD [32]. This suppresses the theoretical event rate with gaps F_{gap} to measured

$$f_{gap} = \langle|S|^2\rangle F_{gap}, \tag{4.19}$$

which clearly makes the distinction between diffractive and non-diffractive events experimentally an ill-posed problem.

One of the first theoretical calculations of these probabilities was done by Bjorken [8]. He did these calculations in the context of electroweak virtual vector boson WW-fusion events hypothetically characterized with LRGs, with event topology highly similar to double diffraction where the interaction is propagated via strong interaction (Pomeron).

Bjorken defined the gap survival probability for hard WW-fusion as a ratio

$$\langle|S|^2\rangle = \frac{\int d^2B d^2b\, \rho(b)\rho(B-b)|S(B)|^2}{\int d^2B d^2b\, \rho(b)\rho(B-b)}, \tag{4.20}$$

based on convolution of parton densities $\rho(b)$ of two incoming hadrons over the transverse impact parameter B space, where B is defined as the distance between centres of two incoming hadrons and b is the convolution variable. This convolution is basically representing the interaction cross section for the hard scattering up to some factor σ_0.

The term $|S(B)|^2$ is the probability that both hadrons pass through each other without generating any inelastic activity on the rapidity interval of the gap. It is often theoretically calculated in the Eikonal picture, where the hadrons are described with some profile function in the impact parameter space [8, 32, 33]. See end of paper [32] for some discussion of how Monte Carlo event generators usually deal with these.

In this thesis, the methods developed for measuring the diffractive cross sections defined in equation 4.13, are *not* relying on rapidity gaps and thus are robust in that sense.

4.3 Space-time image of hadronic collision

Feynman diagrams are covariant, i.e. invariant under space-time Lorentz (Poincaré) transformations and they give no information about space-time image of scattering processes. This is computationally extremely good property, because this allows one to calculate Lorentz invariant cross sections. These cross sections, interaction probabilities, are not however everything in physical understanding.

In the Feynman/Bjorken naive parton picture, the hadron is moving infinitely fast and consists of infinitely many partons each with momentum fraction x. When one looks from the laboratory frame, colliding hadrons are highly Lorentz contracted discs and boosted. Thus, the time scale over the interactions between hadron constituents is much longer. At low energies the partonic fluctuation is small and the valence quarks dominate the content. Then, by parton evolution, low-x gluons dominate the image with high center of mass energies $\sqrt{s}$.

When studying more the naive parton image, one can say that the fast moving partons are localized on the Lorentz contracted disc. The slower moving partons, with smaller momentum fraction are more spread out in the longitudinal direction. Experimentally, this means that when two partons collide with $x_1 \neq x_2$, the laboratory (detector) frame and center of mass frame of the collision are not the same, which is the case always in lepton colliders with $\vec{p}(e^-) + \vec{p}(e^+) \equiv 0$.

Already in 1952, before any parton picture, Heisenberg [36] had an idea that if two Lorentz contracted hadrons collide, they produce a shock wave of mesons filling up all the kinematically allowed longitudinal momentum space. Mathematically, this is equivalent of taking Fourier transform of the spatially localized Dirac's delta-like distribution and ending up in the momentum space with nearly flat distribution. This picture is quite intuitive for non-diffractive soft QCD processes, and it explains why soft neutral and charged pions or kaons are produced uniformly in rapidity $\sim$ longitudinal momentum. Even the broading of dN/dy in rapidity as a function of $\sqrt{s}$ can be easily seen from this Fourier transform perspective, i.e. purely from kinematics of space-time. Even though this view does not say anything about the real QCD dynamics, and basically assumes a constant matrix element.

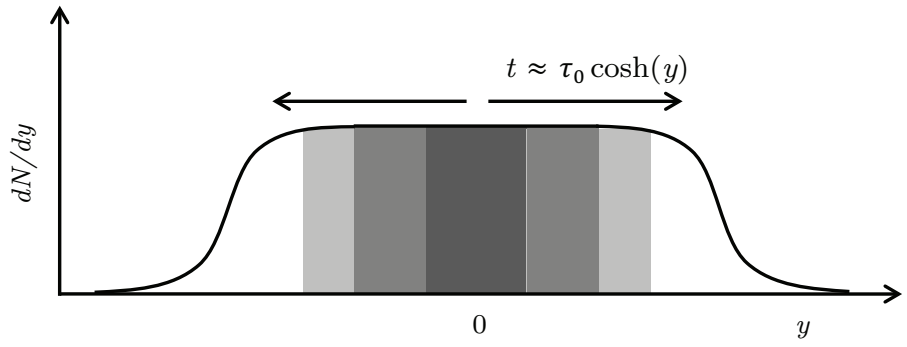


Figure 4.6: Bjorken chain-reaction view of space-time evolution of an ordinary non-diffractive soft process.

More complicated picture of space-time evolution was developed by Bjorken and Gribov [9,35]. In their picture, the hadronic collision is a chain-reaction started first

by the interaction of *wee*[2] partons [28]. At first, only a small domain of partons inside $|\Delta y| \approx 1$, centered around $y = 0$ in the CMS frame, is excited. After the excitation, de-excitation or "cooling" happens within a time $\tau_0 \approx 1$ fm/c by hadron emissions which can excite the neighbouring rapidity domains. This happens at time $t \approx \tau_0 \cosh(y)$ [3] in the rapidity space around $\pm|y|$, thus large rapidity (fast) partons are excited last. The production of hadrons happens e.g. by gluons turning into quark-antiquark pairs during the excitation interval, and these can hadronize or annihilate to produce something else, like leptons. This image of space-time evolution is visualized in Figures 4.6 and 4.7.

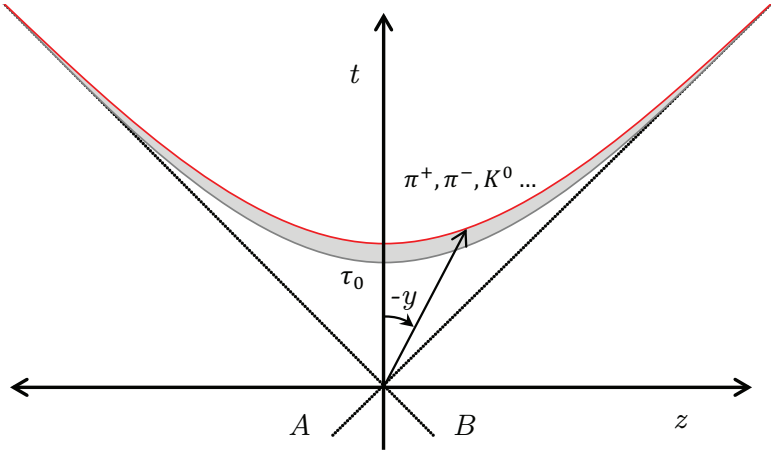


Figure 4.7: Minkowski space-time (t, z) diagram with z as a beam axis in the CMS frame of colliding hadrons A and B at $(0, 0)$, which are moving almost on the light cone near the speed of light. The grey part of hyperbola illustrates hadronization process. The rapidity y of the produced particle on the red boundary shown as the hyperbolic angle.

Diffractive processes in particle physics have some familiarity with classical diffraction, especially the elastic scattering resembles optical diffraction. The picture is however much more complicated in diffractive dissociation than in elastic scattering, due to complicated inner structure of the hadrons under dissociation. The complete space-time image and QCD dynamics picture for coherent processes, like low-Q^2 diffractive scattering is highly non-trivial and not well-known in any sense. A quite intuitive image of diffraction by Good and Walker is represented later in Chapter 5.

[2] Partons with Bjorken $x \sim 0$.

[3] Directly from a general Lorentz boost along z-axis: $\begin{pmatrix} ct' \\ z' \end{pmatrix} = \begin{pmatrix} \cosh y & -\sinh y \\ -\sinh y & \cosh y \end{pmatrix} \begin{pmatrix} ct \\ z \end{pmatrix}$.

4.4 Regge theory

Regge theory, named after Tulio Regge's complex angular momentum ideas, is a theory for strong interactions from 1960s, before QCD, and it is based on sound and general mathematical principles from quantum mechanics. Later on, Gribov's Reggeon calculus and the so-called cutting rules of Abramovskii, Gribov and Kancheli (AGK) extended these basic principles much further. Here, only the leading principles are presented and the three most important are the following. [6]

Unitary, which also means conservation of probability and leads for example to the optical theorem for elastic scattering. The optical theorem gives a relationship between the total cross section and the forward scattering amplitude at $t = 0$.

Duality, which in Regge theory context usually means that scattering amplitudes can be expressed as a sum of poles *produced* in the s-channel *or* equivalently as a sum of poles *exchanged* in the t-channel.

Analyticity, which is the concept from mathematical complex analysis giving integral and continuity relations, a crossing symmetries between particle and antiparticle scattering amplitudes and is related to causality in quantum field theories.

Spins and masses

Soft, low-Q^2 interactions, are described in Regge theory as scatterings which are modelled as sums of different exchanges. These exchange objects are given by particle-families with different spin. Meson families demonstrate a surprising linear (affine) relationship between spin J and squared masses m^2 of the particles [6]

$$J \equiv \alpha(m^2) \simeq \alpha_0 + \alpha' m^2, \tag{4.21}$$

where α_0 is the intercept parameter and the slope α' is sometimes called *inverse string tension*.

In Regge theory the scattering amplitude is modelled as a linear combination of *partial-waves* with spin J and mass m, i.e. sum of orthogonal Legendre polynomials P_J with phase shifts ϕ_j

$$A(s,t) \sim \sum_J \phi_J P_J(z_t = \cos\theta_t), \tag{4.22}$$

where $z_t \sim 1 + \frac{2s}{t-4m^2}$. For scattering of initial states with equal invariant masses $m_1 = m_2 = m$, the "physical" region of t-channel is given by requiring t-channel scattering angle $|z_t = \cos\theta_t| \leq 1$ which leads to $t > 4m^2$ and $s < 0$ [26]. This partial-wave approach in is already familiar from the ordinary quantum mechanics scattering problems and thus it was naturally extended to S-matrix Regge formalism.

Meson family exchanges

Analyticity gives a property that the partial-wave representation of $A(s,t)$ can be complex analytically continued to "unphysical" region of t-channel which states $t < 0$, and $s \gg 4m^2$, i.e. to physical high energy s-channel energies. Shortly, using Legendre polynomials and their properties with respect to analyticity gives [26]

$$z_t^J \sim \left(\frac{s}{|t-4m^2|}\right)^J \sim s^J \sim s^{\alpha(t)}. \tag{4.23}$$

This leads to *Regge trajectories* exchanged in t-channel, instead of the partial-waves representation in s-channel. The scattering amplitudes in the crossed t-channel are then [26]

$$A(s,t) \sim \sum_{\alpha_i} \beta_i(t) s^{\alpha_i(t)}, \tag{4.24}$$

where $\beta_i(t) : \mathbb{R} \to \mathbb{R}$ are some unknown functions. From this one can deduce differential cross section as a function of t as

$$\frac{d\sigma}{dt} \sim s^{2\alpha(t)-2} \tag{4.25}$$

and by using optical theorem, one gets

$$\sigma_{tot} \sim s^{\alpha(0)-1} \sim s^{-\frac{1}{2}} \quad \text{when} \quad \sqrt{s} \lesssim 10\,\text{GeV} \tag{4.26}$$

for exchange families of ρ, ω, f_2 and a_2 mesons. These are the leading exchange particle families, *Reggeons*, in the Regge theory sense at very low center of mass energies. [6, 24, 26]

Rise of the Pomeron

When $\sqrt{s} \gtrsim 20$ GeV, the total and elastic cross sections rise experimentally as $\sim s^{0.08}$. This triggers the need for a new pole, the vacuum pole Pomeron with vacuum quantum numbers $J^{PC} = 0^{++}$ and a trajectory [6]

$$\alpha_P(t) = \alpha_P(0) + \alpha'_P t \quad \text{when} \quad \sqrt{s} \gtrsim 20\,\text{GeV}. \tag{4.27}$$

The total cross section rises as $\sigma_{tot} \sim s^{\alpha_P(0)-1}$, the exponent is experimentally measured to be $\epsilon = \alpha_P(0) - 1 \approx 0.08$ and the slope $\alpha'_P \approx 0.25\,\text{GeV}^{-2}$. So at high $\sqrt{s}$, the total hadron-hadron interactions are phenomenologically described by exchange of Pomeron trajectory in t-channel. The dips and peaks observed experimentally as a function of $|t|$ in elastic scattering are result of interference of real and imaginary part of forward scattering amplitude. [24]

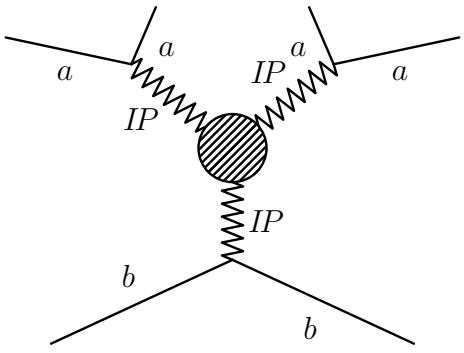


Figure 4.8: Triple Regge diagram with $IP\,IP\,IP$-vertex $\sim$ squared amplitude of Figure 4.2 summed over the system X, in the high mass limit of single diffraction $ab \to a\dot{+}X$.

The diffractive cross sections are described usually under the triple Regge formalism, where eight different (2^3) combinations of Pomerons and Reggeons "interacting" in the triple vertex are possible. At very high masses of diffractive system,

the (*IPIPIP*) vertex is dominating. See Figure 4.8 for an illustration. In the lower mass region a *second* triple coupling term is needed such that the lower Pomeron is replaced with a Reggeon family, for example with f_2 or ω family. Also, in the intermediate range when ξ is not close to zero, triple vertices with one or both upper Pomerons are replaced with Reggeons. [24]

Especially the low mass region and its differential cross section as a function of diffractive mass is not well known, because experimentally it requires very forward instrumentation. Approximately single diffractive cross section behaves like $d\sigma_{SD}/dM_X^2 \propto 1/M_X^2$ but lower mass region ($M_X \lesssim 5$ GeV) can be steeper and also the basic Regge theory is not theoretically valid anymore [6, 24].

Even if QCD is widely regarded as the right theory for the strong interactions, it is unable to give quantitative predictions for soft and diffractive processes. QCD gives only good predictions perturbatively for high-Q^2 processes and by using lattice QCD, meson and baryon mass spectrums can be approximately reproduced. Currently, the predictions for the total cross section for hadronic collisions are based on the phenomenological Regge approach. Regge theory lives alongside with QCD, and much effort is still needed to understand these two theories together. At leading order, the Pomeron can be perturbatively described by a 2-gluon color-singlet state, which is of practical use only in high-Q^2 diffraction, when production of hard objects is triggered diffractively. Higher order description of hard diffraction is done in a gluon ladder picture, and often the term BFKL Pomeron is used [6, 24].

These mysterious, approximately linear Regge trajectories have been one of the inspirations for development of classic string theories [26]. In classic string theory, one could think these resonance mesons following a Regge trajectory as a different excitations of a string, which is rotating. Extremely popular AdS/CFT (Anti de Sitter/Conformal Field Theory) duality, which is a conjectured equivalence between a chosen string theory and gravity on one space and QFT without gravity in a lower dimensional space, has triggered also ideas about Graviton-Pomeron Regge trajectories [17].

Martin-Froissart bound

Most of the phenomenological models for the total cross section as a function of s are based on Regge theory fits and utilization of Martin-Froissart bound. The Martin-Froissart upper-bound is [6]

$$\sigma_{tot} < \frac{\pi}{m_\pi^2} \ln^2 s, \tag{4.28}$$

which is derived using mathematical tools similar to Regge theory and not in terms of QCD. Experimentally $\sigma_{tot} \sim (0.3 \text{ mb}) \ln^2 s$, but $\pi/m_\pi^2 \approx 64$ mb. This indicates that the bound could need a modification. Taking account the chiral perturbation theory, where $m_\pi \to 0$, i.e. when up and down quark masses vanish and chiral symmetry is obtained, the total cross section might diverge. The saturation of the bound is also unsolved and it's context to gluons and their impact parameter b picture.

4.5 Regge theory in Monte Carlo

What follows, a phenomenological model [16] for predicting the total cross section and diffractive cross sections. It is introduced because of its implementation in the latest version of PYTHIA 8 event generator and it is somewhat similar to other Regge theory based soft diffraction models. However, it does not represent any theoretical superiority over other phenomenologies.

Total cross section formula

The Minimum Bias Rockefeller (MBR) model [16] predicts the total $p-(\bar{p})p$ cross section with

$$\sigma_{tot}^{p\pm p}(s) = \begin{cases} 16.79s^{0.104} + 60.81s^{-0.32} \mp 31.68s^{-0.54} & \text{when} \quad \sqrt{s} < 1.8 \text{ TeV} \\ \sigma_{tot}^{CDF} + \frac{\pi}{s_0}\left[\left(\ln\frac{s}{s_F}\right)^2 - \left(\ln\frac{s_{CDF}}{s_F}\right)^2\right] & \text{when} \quad \sqrt{s} \geq 1.8 \text{ TeV}, \end{cases} \tag{4.29}$$

where the upper term is obtained by a global Regge-theory fit to data and the lower term is a prediction of a saturated Martin-Froissart bound model.

One can see that on higher energies there is no difference between particle-(anti)particle scattering cross sections. Intuitively, this reflects similar parton distributions (full of small-x gluons) on the asymptotic energies between colliding proton and anti-proton. This asymptotic equivalence between particle-(anti)particle scattering from the pre-QCD era is known as the Pomeranchuk theorem (1958) [6].

In Equation 4.29, the free parameters are the saturation energy $\sqrt{s_F} = 22$ GeV and the energy scale $s_0 = 3.7 \pm 1.5$ GeV^2. To obtain values in millibarns, the s_0 is divided by $(\hbar c)^2 \approx 0.389$ GeV^{2}mb. The normalization is done using the measured cross section of $\sigma_{tot}^{CDF}(\sqrt{s} = 1.8 \text{ TeV}) = 80.03 \pm 2.24$ mb. [16]

Diffractive cross section formulas

The proton fractional longitudinal (forward) momentum losses $\xi = \Delta p_z/p_z$ of the intact proton for the single diffraction, double diffraction and central diffraction (double Pomeron exchange) are

$$\xi_{SD} = \frac{M^2}{s} \tag{4.30}$$

$$\xi_{DD} = \frac{M_1^2 M_2^2}{s \cdot s_0} \tag{4.31}$$

$$\xi_{CD} = \xi_1\xi_2 = \frac{M^2}{s}, \tag{4.32}$$

where M^2 is the invariant mass squared of the dissociated system(s) and the energy scale parameter is s_0. Clearly the definition of ξ_{DD} is not really physically the same as ξ_{SD} and ξ_{CD}, because both protons have dissociated.

This forward momentum is carried away by the interchanged "particle", by Pomeron, and the fractional momentum loss has an inverse exponential relation with the rapidity gap width Δy, given by

$$\xi = e^{-\Delta y} \Leftrightarrow \Delta y = -\ln(\xi). \tag{4.33}$$

In the central diffraction $\Delta y = \Delta y_1 + \Delta y_2$, which is a sound definition because $\xi_{CD} = e^{-\Delta y_1} e^{-\Delta y_2} = e^{-(\Delta y_1 + \Delta y_2)} = e^{-\Delta y}$. In the double diffraction, the center of the rapidity gap y_0 can be calculated with

$$y_0 = \frac{1}{2} \ln \left(\frac{M_2^2}{M_1^2} \right), \tag{4.34}$$

which means that when $M_1 \neq M_2$, it shifts from zero.

The differential cross sections are modelled using the Pomeron trajectory $\alpha(t)$, the Pomeron-proton coupling $\beta(t)$ and the ratio κ of the triple Pomeron to Pomeron coupling [16]

$$\alpha_P(t) = \overbrace{1 + \epsilon}^{\alpha_P(0)} + \alpha'_P t = 1.104 + 0.25(\mathrm{GeV}^{-2}) \cdot t \tag{4.35}$$

$$\beta^2(t) = \beta^2(0) F^2(t) \tag{4.36}$$

$$\kappa = g(t)/\beta(0), \tag{4.37}$$

where $F(t)$ is the proton form factor [16], a concept already used in nuclear physics (Fourier transform of the charge density). The value for the Pomeron-proton cross section is fixed by $\sigma_0 := \kappa\beta^2(0) = 2.82$ mb at $s_0 = 1$ GeV^2.

Now, the differential cross section formulas are expressed as a function of kinematical terms as [16]

$$\frac{d^2\sigma_{SD}}{dt d\Delta y} = \frac{1}{N_G(s)} \left[\frac{\beta^2(t)}{16\pi} e^{2[\alpha_P(t)-1]\Delta y} \right] \left(\kappa\beta^2(0) \left(\frac{s'}{s_0} \right)^{\epsilon} \right), \tag{4.38}$$

$$\frac{d^3\sigma_{DD}}{dt d\Delta y dy_0} = \frac{1}{N_G(s)} \left[\frac{\kappa\beta^2(0)}{16\pi} e^{2[\alpha_P(t)-1]\Delta y} \right] \left(\kappa\beta^2(0) \left(\frac{s'}{s_0} \right)^{\epsilon} \right), \tag{4.39}$$

$$\frac{d^4\sigma_{CD}}{dt_1 dt_2 d\Delta y dy_c} = \frac{1}{N_G(s)} \left[\prod_i \left(\frac{\beta^2(t_i)}{16\pi} e^{2[\alpha_P(t_i)-1]\Delta y_i} \right) \right] \kappa \left(\kappa\beta^2(0) \left(\frac{s'}{s_0} \right)^{\epsilon} \right), \tag{4.40}$$

where y_c is the center of the central system in the rapidity space. Above, the expression in normal $(\cdot)$ brackets encapsulates the total Pomeron-proton cross section at the reduced squared collision energy, where $s' = e^{-\Delta y} s = \xi s = M_X^2$. The expression in square brackets $[\cdot]$ is understood as the Pomeron flux $\sim$ probability of Pomeron exchange. [16]

These expressions are said to hold when rapidity gaps are $\Delta y \gtrsim 3$ ($\xi \lesssim 0.05$), i.e. when Pomeron exchange is dominating and coherence condition holds. The $N_G(s) = \min(1, f)$ is renormalization factor and f is the integral of the square bracket term in equations 4.38, 4.39 and 4.40. This integral is calculated over phase space as [16]

$$f_{SD} = \int dt \, [\,\cdot\,] \tag{4.41}$$

$$f_{DD} = \int dt dy_0 \, [\,\cdot\,] \tag{4.42}$$

$$f_{CD} = \int dt_1 dt_2 dy_c \, [\,\cdot\,] \tag{4.43}$$

for $\Delta y_{\min} > 2.3$. The authors of the model say that this "renormalization"[4] corresponds to interpretation of the Pomeron flux as a rapidity-gap formation probability. Mainly, this renormalization procedure makes the MBR model different from standard Regge phenomenologies, and it is not theoretically on rigorous basis. Its purpose is to fix the break of unitary along the rise of diffractive cross sections. This rise seems to start with the standard triple Pomeron Regge approach to diffraction, with the center of mass energy larger than 20 GeV [37].

The Equations 4.38, 4.39 and 4.40 are assumed to be valid only at large rapidity gaps Δy, as said earlier. A somewhat arbitrary approach is then needed at small rapidity gaps. The cross sections obtained at small rapidity gaps values are suppressed with a multiplicative factor

$$S = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{\Delta y - \Delta y_S}{\sigma_S}\right)\right], \tag{4.44}$$

where $\mathrm{erf}(x)$[5] is the sigmoidal, Gaussian error function with parameters $\Delta y_S = 2$ (coherence limit set $\xi \leq 0.135$) and $\sigma_S = 0.5$. These parameters are problematic, and especially with the DD events, the Δy_S controls how large fraction of the DD events in total look like the ND events, i.e. have a vanishing rapidity gap. [16]

Refer to the original reference [16], how the events are generated in practise. To summarize, the free parameters of the model are: Pomeron trajectory $\epsilon = 0.104$ and $\alpha' = 0.25$, the Pomeron-proton coupling $\beta(0) = 6.556$ (GeV^{-1}), ratio $\kappa = 0.17$, the lower bound of the dissociated system mass-squared $M_0^2 = 1.5$ (GeV^2), the minimum widths of the rapidity gap $\Delta y_{\min,SD}, \Delta y_{\min,DD}, \Delta y_{\min,CD} = 2.3$ used in flux renormalization and the parameters $\Delta y_{S,SD}, \Delta y_{S,DD}, \Delta y_{S,CD} = 2.0$ and $\sigma_{S,SD}, \sigma_{S,DD}, \sigma_{S,CD} = 0.5$ for low Δy cross section suppression. [16]

[4] Which is not to be mixed with the concept of renormalization in quantum field theories.

[5] $\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$

Chapter 5

Diffraction as a quantum mechanical process

In general, describing hadronic processes in particle physics requires much more mathematical machinery than in plain-old quantum mechanics. However, in the so-called Good-Walker image [31] of diffractive scattering, a purely quantum mechanical S-matrix formalism is used. The whole approach is based on unitarity of the scattering matrix.

5.1 Good-Walker image

The moving hadron $|B\rangle$ is described as *a coherent*, complex valued superposition of eigenstates $|\psi_i\rangle$ of the scattering operator [31]

$$|B\rangle = \sum_k C_k |\psi_k\rangle, \tag{5.1}$$

with probability amplitudes $C_k \in \mathbb{C}$. The kets $|\psi_k\rangle$ are the eigenstates of the scattering amplitude operator Im $T = 1 -$ Re S, with an eigenvalue equation

$$\text{Im } T|\psi_k\rangle = t_k(\vec{b})|\psi_k\rangle. \tag{5.2}$$

From the unitarity condition $0 \leq t_k(\vec{b}) \leq 1$, and the imaginary operator is basically coming from the definition of T-matrix with respect to the S-matrix, in Equation 2.33 . These eigenvalues are the probabilities for the corresponding state to interact or "absorb" with the target, and are a function of the impact parameter $\vec{b} \in \mathbb{R}^2$.

The usual normalization is used

$$\langle B|B\rangle = \sum_k C_k^* C_k \langle \psi_k|\psi_k\rangle = \sum_k |C_k|^2 = 1, \tag{5.3}$$

with an orthonormal set of eigenstates obeying $\langle \psi_i|\psi_j\rangle = \delta_{ij}$. After collision, the state of hadron is not anymore $|B\rangle$, this means that different eigenstates are absorbed by different intensities.

Directly using these, the imaginary part of elastic amplitude is given by the expectation over absorption coefficients

$$\langle B|\text{Im } T|B\rangle = \sum_k |C_k|^2 t_k = \langle t_k\rangle, \tag{5.4}$$

which is simply weighted average of absorption coefficients t_k weighted by the probability amplitudes $|C_k|^2$ of states.

The differential total and elastic cross sections with respect to the impact parameter $\vec{b}$, by using only imaginary part of the elastic amplitude, are described by [31]

$$\frac{d\sigma_{tot}}{d^2\vec{b}} = 2\langle t_k\rangle \tag{5.5}$$

and

$$\frac{d\sigma_{el}}{d^2\vec{b}} = \langle t_k\rangle^2. \tag{5.6}$$

The differential cross section of diffraction is the dispersion of absorption coefficients [31]

$$\frac{d\sigma_{diff}}{d^2\vec{b}} = \sum_k |\langle \psi_k|\text{Im } T|B\rangle|^2 - \frac{d\sigma_{el}}{d^2\vec{b}} \tag{5.7}$$

$$= \sum_k |C_k|^2 t_k^2 - \left(\sum_k |C_k|^2 t_k\right)^2 \tag{5.8}$$

$$= \langle t_k^2\rangle - \langle t_k\rangle^2, \tag{5.9}$$

and the dispersion is equal to variance of the absorption coefficients. Utilization of two-dimensional Fourier transform is needed to express these equations in the momentum space instead of the impact parameter space.

This general, a quite simple picture of diffraction by Good-Walker does not really directly tell us much about diffraction quantitatively, which is a common property to theories utilizing S-matrix formalism. However, one interesting, a simple model suitable for soft diffraction and based on Good-Walker formalism is by Miettinen and Pumplin.

5.2 Miettinen-Pumplin model

Under the Good-Walker image, Miettinen and Pumplin (MP) [42] formulated a reasonable simple model of soft diffraction. In their model, hadrons are described as composite objects. These constituents make up the hadron as a quantum-mechanical superposition of different states. In diffraction, these states are *absorbed* unequally by the target. This process produces inelastic states which have the same internal quantum numbers as the incident hadron.

Most important fact about the MP model is that diffraction is understood arising from the *differences* in absorption probabilities of hadron's component states. Under the usually understood model of probabilities (Kolmogorov's axioms), i.e. no negative values, these absorption probabilities must be $t_k \in [0, 1]$. From there, the Pumplin s-channel unitary upper bound can be derived as [42]

$$\sigma_{diff}(b) + \sigma_{el}(b) \leq \frac{1}{2}\sigma_{tot}(b), \tag{5.10}$$

under given $b = |\vec{b}|$, which is the impact parameter of the collision. However, this bound is strictly valid only under the assumptions of this model. Using the Pumplin bound, non-diffractive cross section is bounded by

$$\sigma_{nd}(b) \geq \frac{1}{2}\sigma_{tot}(b). \tag{5.11}$$

The impact parameter has a major effect on production of soft diffraction, and thus it is a main constituent of the MP model.

The first assumption of the MP model is that if the interactions of states are described just by using real valued absorption coefficients, these states must *diagonalize* the corresponding diffractive part of the S-matrix. Diagonalization means simply that the corresponding matrix valued operator eigenvalue equation can be expressed in a form where eigenvalues are on the diagonal of the diagonal matrix.

In the MP model, the diagonal states of the S-matrix are derived from the parton model. Every state is described by a number of $N \in \mathbb{N}$ partons, which each are characterised with a pair $(\vec{b}, y)$ of impact parameter and energy and longitudinal momentum (as rapidity). Thus, a set of parton characterizations is $\{(\vec{b}_i, y_i)\}_{i=1}^{N}$. These partons are assumed to be structureless, point-like constituents, but no quark model or similar hypothesis is made. Under QCD, these must be valence quarks, sea quarks or gluons.

The MP model generates the global properties of diffraction on distribution and interactions of *wee* partons only. The parton states are understood as the eigenbasis for diffraction, i.e. states which diagonalize the S-matrix. These parton states can be well populated with wee partons which leads to high interaction probability and some states have no or just few wee partons, and these generate the transparent channels of diffraction. Main feature of the MP model is that the fluctuations in interaction probabilities which are responsible for diffraction, are due to fluctuations in the number of wee partons N and their characterizations of impact parameters and rapidities. Here is the direct connection to probabilistic classification, and the multidimensional likelihood distributions introduced in Chapter 8 describe the information embedded in the final state multiplicity fluctuations over rapidity.

Probabilities

Now, under the optical MP model, the hadron is modelled as a coherent sum over N parton states which are assumed to be the eigenstates of diffraction [42]

$$|B\rangle = \sum_N \prod_{i=1}^{N} \int d^2\vec{b}_i dy_i \; C_N(\vec{b}_i, \ldots, \vec{b}_N; y_1, \ldots, y_N) \, |\vec{b}_1, \ldots, \vec{b}_N; y_1, \ldots, y_N\rangle, \quad (5.12)$$

where coefficients C_N are the wee parton configurations (impact parameters and rapidities). This model clearly assumes no dependency e.g. on spin. Also no correlation between wee partons is assumed, thus product rule of probability calculus can be used. The total probability $|C_N|^2$ of N wee partons is modelled as a random process with a Poisson distribution [42]

$$|C_N(\vec{b}_1, \ldots, \vec{b}_N; y1, \ldots, y_N)|^2 = e^{-G^2} \frac{G^{2N}}{N!} \prod_{i=1}^{N} |C_i(\vec{b}_i, y_i)|^2, \quad (5.13)$$

where G^2 is the mean of Poisson distribution and the probabilities are normalized to

$$\int d^2\vec{b}_i dy_i |C(\vec{b}_i, y_i)|^2 = 1. \quad (5.14)$$

The interaction probability of a parton to interact with the target is denoted with τ_i, and due to Kolmogorov's axioms, probability to *not* to interact is $1 - \tau_i$. Thus, the probability for N independent partons to interact is

$$1 - \prod_{i=1}^{N} (1 - \tau_i). \quad (5.15)$$

This result is said to depend directly on s-channel unitarity [42]. Now the whole MP-model is characterized by independent wee-parton states

$$|\psi_k\rangle = |\vec{b}_1, \ldots, \vec{b}_N; y_1, \ldots, y_N\rangle \quad (5.16)$$

with likelihoods to exist in the hadron as

$$e^{-G^2} \frac{G^{2N}}{N!} \prod_{i=1}^{N} |C(\vec{b}_i, y_i)|^2 \; d^2\vec{b}_i dy_i. \quad (5.17)$$

These partons interact with the target hadron with probabilities given by

$$t(\vec{b}_1, \ldots, \vec{b}_N; y_1, \ldots, y_N; \vec{b}) = 1 - \prod_{i=1}^{N} [1 - \tau(\vec{b}_i + \vec{b}, y_i)], \quad (5.18)$$

with parton's impact parameters translated with the hadron's impact parameter $\vec{b}$. [42]

Differential cross sections

Now the cross sections can be described as a function of the impact parameter under the Good-Walker image using moments of the absorption coefficients t_k. These first-order and second-order (central) moment are given in the Miettinen and Pumplin model by Eikonal forms [42]

$$\langle t \rangle = 1 - e^{-G^2\langle \tau \rangle} \tag{5.19}$$

$$\langle t^2 \rangle - \langle t \rangle^2 = e^{-2G^2\langle \tau \rangle}(e^{+G^2\langle \tau^2 \rangle} - 1), \tag{5.20}$$

where the independency assumption of wee-parton distributions was used. In general, the n-th moment of the single-parton interaction is

$$\langle \tau^n \rangle = \int d^2\vec{b}_1 dy_1 \, |C(\vec{b}_1, y_1)|^2 [\tau(\vec{b}_1 + \vec{b}, y_1)]^n, \tag{5.21}$$

with the given impact parameter $\vec{b}$.

Differential cross sections with respect to the transverse momentum vector $\vec{q}$ are obtained by two-dimensional Fourier transform of the impact parameter amplitudes [42]

$$\frac{d\sigma_{el}}{d^2\vec{q}} = \frac{1}{4\pi^2}\langle \tilde{t} \rangle^2 \tag{5.22}$$

$$\frac{d\sigma_{diff}}{d^2\vec{q}} = \frac{1}{4\pi^2}\left(\langle \tilde{t}^2 \rangle - \langle \tilde{t} \rangle^2\right), \tag{5.23}$$

where

$$\begin{aligned}\langle \tilde{t}^j \rangle = &\sum_{N=0}^{\infty} e^{-G^2}\frac{G^{2N}}{N!}\prod_{i=1}^{N} d^2\vec{b}_i dy_i |C(\vec{b}_i, y_i)|^2 \\ &\times \left\{ \int d^2\vec{b} e^{i\vec{q}\cdot\vec{b}}[1 - \prod_{k=1}^{N}(1 - \tau(\vec{b_k} + \vec{b}, y_k))] \right\}^j .\end{aligned} \tag{5.24}$$

Fourier transform can be used due to well-known fact from quantum mechanics, which says that position and momentum are conjugate variables[1], thus their operator representations do not commute, are constrained by Heisenberg uncertainty principle $\Delta x \Delta p \geq \hbar/2$ and are coupled via Fourier transform. Also, using Fourier transform is possible due to the fact that the parton states are the same for all global impact parameter values $\vec{b}$.

The expectation values in differential cross section equations 5.22 and 5.23 are calculated using 2-dimensional Fourier transform as [42]

$$\langle \tilde{t} \rangle = \int d^2\vec{b}\, e^{i\vec{q}\cdot\vec{b}}\langle t \rangle \tag{5.25}$$

$$\langle \tilde{t}^2 \rangle - \langle \tilde{t} \rangle^2 = \int d^2\vec{b}\, d^2\vec{b'} e^{i\vec{q}\cdot(\vec{b'} - \vec{b})} e^{-G^2[\langle \tilde{\tau}(\vec{b}) \rangle + \langle \tilde{\tau}(b') \rangle]}(e^{G^2\langle \tau(\vec{b})\tau(\vec{b'}) \rangle} - 1), \tag{5.26}$$

where

$$\langle \tau(\vec{b})\tau(\vec{b'}) \rangle = \int d^2\vec{b}_1 dy_1 |C(\vec{b}_1, y_1)|^2 \tau(\vec{b}_1 + \vec{b}, y_1)\tau(\vec{b}_1 + \vec{b'}, y_1). \tag{5.27}$$

[1] like time and energy.

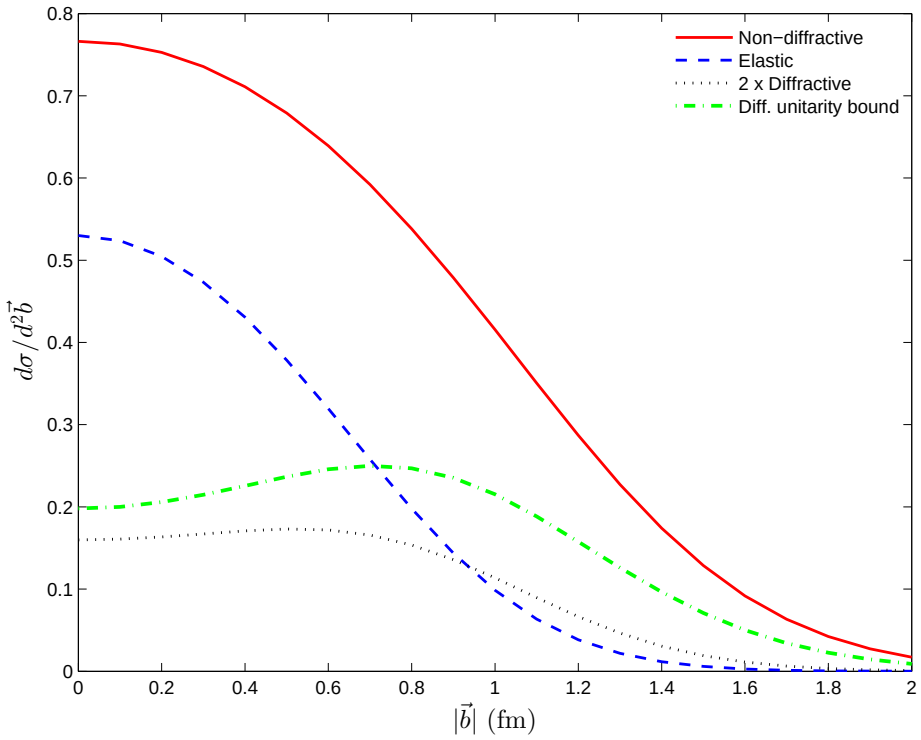


Figure 5.1: Differential cross section as a function of the impact parameter using the MP model parametrization at $\sqrt{s} = 53$ GeV, as described in [42].

Basically, the MP model is described with three factors. First G^2 defines the average number of wee partons in a Poissonian way, $|C(\vec{b}, y)|^2$ defines the single-parton probability distribution and $\tau(\vec{b}, y)$ defines the single-parton interaction probability.

However, easily can be seen that all the values in the model are integrated and an event-by-event Monte Carlo model would need a modification in the model. In the original paper [42], parametrizations were used to give predictions. See Figure 5.1 for an illustration of the MP model predictions at $\sqrt{s} = 53$ GeV. The figure is done with $\sigma_{diff} = \sigma_{sdl} + \sigma_{sdr}$, so the double diffraction is not explicitly included, because it was not really discussed in the original paper. Diffractive cross section peaks at non-zero impact parameter value, which is intuitive if one thinks diffraction in optical terms.

Chapter 6

Hadron colliders

Hadron colliders are the high energy new physics search tools. In comparison with lepton colliders, like the pre-LHC collider LEP (large electron-positron collider) at CERN, hadron colliders can achieve much higher center of mass energies due to much smaller synchrotron radiation which has proportionality $\propto 1/m^4$ to the particle mass. As a downside, hadron colliders are colliding composite objects which have a non-deterministic initial state $|i\rangle$, i.e. colliding particles are actually quarks or gluons with some probability. This compositeness creates the so-called underlying event which makes precision physics a great challenge in hadron colliders. Naturally, as a positive side, hadron colliders allow the research of hadronic structure over large kinematical regions.

6.1 Master equation

In an inelastic hard collision in a hadron collider, the center of mass energy of two interacting partons is

$$\sqrt{\hat{s}} = \sqrt{x_1 x_2 s} = Q, \tag{6.1}$$

where Bjorken variables x_1 and x_2 are the fractional momentums of the two partons colliding, and they have a kinematical relation with rapidity y of the heavy end product as [15]

$$x_{1,2} = \frac{Q}{\sqrt{s}} e^{\pm y}. \tag{6.2}$$

By using this, one can see easily that a centrally produced $y \sim 0$ hypothetical massive particle requires $x_1 \backsimeq x_2$. Then the mass of the particle $M = Q$ dictates is it the sea (low x) or valence (higher x) partons which produce the particle most probably.

Equations 6.1 and 6.2 demonstrate a dramatic difference in comparison to lepton colliders where $\sqrt{\hat{s}} = \sqrt{s}$ is fixed due to non-composite colliding particles, and the only way to change it is to vary beam energy. The value of $\sqrt{\hat{s}}$ is the threshold for particle production, i.e $\sqrt{\hat{s}} \geq M_X$, where M_X is the invariant mass of the system produced. Statistically, x_1 and x_2 are described by the same parton distribution functions than measured in DIS, and at the LHC energies, the highest probability is actually to have hard gg-collisions.

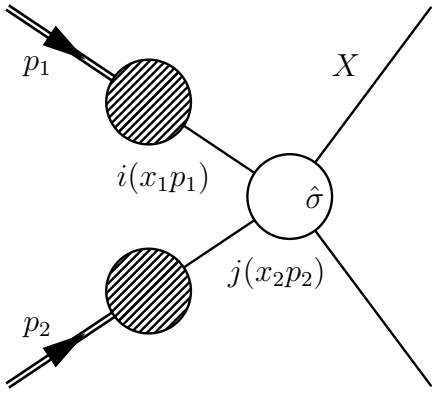


Figure 6.1: Illustration of the hard collision master equation.

If the factorization of short and long distance parts of QCD interactions can be assumed, the cross section for different processes is expressed as [15]

$$d\sigma_{p_1 p_2 \to X} = \sum_{i,j \in \{q,g\}} \int_0^1 dx_1\, dx_2 f_{i/p_1}(x_1, \mu_F^2) f_{j/p_2}(x_2, \mu_F^2) d\hat{\sigma}_{ij \to X}(Q^2; \mu_F^2, \mu_R^2), \tag{6.3}$$

where μ_F^2 is the factorization scale and μ_R^2 is the renormalization scale. $Q^2 \gtrsim 2$ GeV^2 is the scale of the hard parton-parton process, i.e. 4-momentum transfer $\hat{t}$ or $Q = \sqrt{\hat{s}}$ in an s-channel process. This means that two partons (quarks or gluons) do any of possible hard collisions which can produce the system X under study, and the sum is done incoherently over all the possibilities. The master equation is illustrated in Figure 6.1. One must still remember that the factorization is not theoretically rigorous for hadron colliders, in contrast with DIS, where it can be

proven with operator product (OPE) Laurent series expansion. The soft gluon ($\sim$ long wavelength) color connections can spoil the factorization in hadron-hadron collisions, but the effects can be shown to be power suppressed as a function Q^2.

The parton-parton cross section is calculated perturbatively as power series [15, 29]

$$d\hat{\sigma}_{ij\to X} = \alpha_s^k \sum_n d\hat{\sigma}_{ij}^{(n)} \left(\frac{\alpha_s}{\pi}\right)^n, \tag{6.4}$$

where $n \in \mathbb{N}$ is the order of perturbation series, with leading order (LO) also known as tree level or Born level. The tree level is the lowest possible order diagram in the coupling g_s for a given process. The α_s^k denotes the leading power of the hard process, e.g. for jets $k = 2$ and vector boson production $k = 0$ [29]. Increasing the order of perturbation in QCD is computationally very complicated and there are no universal techniques to calculate at arbitrarily high order. At each order of perturbation theory, both parton-parton cross section and pdfs have a residual factorization μ_F scale dependency. This is supposed to be smaller with higher order calculations. After all, the measurable cross sections $d\sigma$ must be independent of any (arbitrary) renormalization μ_R and factorization scale μ_F, i.e. they obey renormalization group equations (RGE) condition[1].

The calculation techniques must handle two kind of divergences: the ultraviolet (UV) and infrared (IR) divergences. The UV divergences are coming from the small scale loop corrections, from higher resolution, and the IR divergences from large scale, soft effects. The former is usually handled by renormalization via dimensional regularization, i.e. calculations are done in space-time of dimension $d = 4 - \epsilon_{UV}$, where ϵ_{UV} is the regularization term. This is done to avoid singularities (poles) encountered with complex integrals. Shortly expressed, the UV divergences are absorbed into bare strong coupling $\sim$ renormalization scale μ_R dependency. The latter, IR divergences are usually tailed by cancellation of virtual and real particle contributions at all order of perturbation theory together with a suitable pdf factorization scheme, which can absorb collinear and soft gluon emissions $\sim$ factorization μ_F scale dependency.

[1] $\mu_{R,F} \frac{\partial \sigma}{\partial \mu_{R,F}} = \frac{\partial \sigma}{\partial \ln(\mu_{R,F})} = 0.$

6.2 Jets and algorithms

Jets are collimated, experimentally measured sprays of particles in a "restricted" space angle and close in phase-space, and highly important in searching for new physics and in the Standard Model measurements. Like $b\bar{b}$-jets which is one of the decay signatures for Higgs H^0, but also signature for standard QCD processes. As discussed earlier, they can be produced also in hard diffraction.

Due to QCD confinement, energetic quarks or gluons produced in hard high-Q^2 scatterings are never seen free, but they produce sprays of hadrons (mesons, baryons) which are detected experimentally by their tracks (if charged), and by their energy deposits. Unfortunately, *hadronization*, the process of how these quarks or gluons turn into jets can be only phenomenologically modelled.

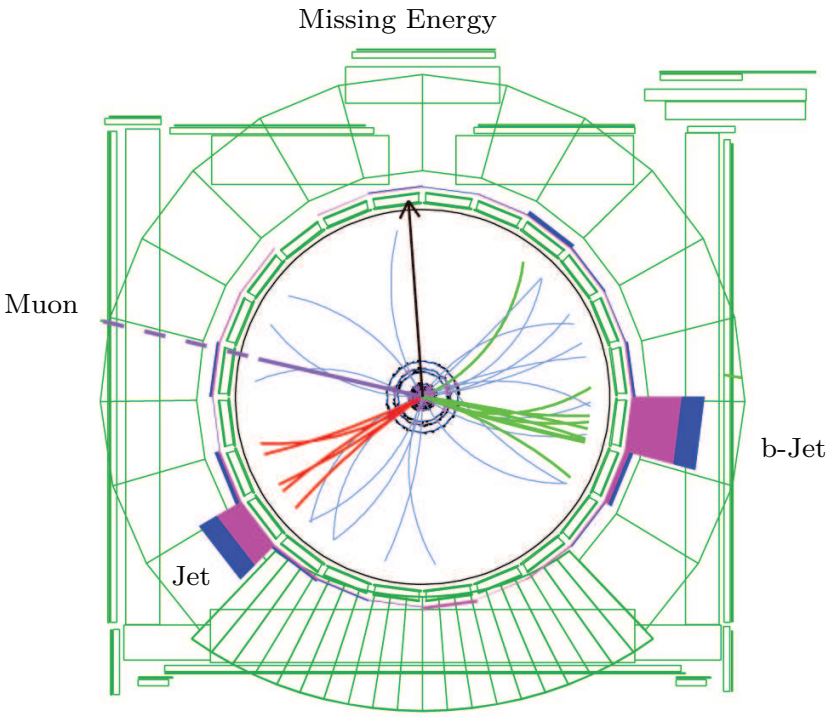


Figure 6.2: Single top quark event reconstructed in transverse (x, y) plane at CDF. Figure shows reconstructed tracks and transverse energies of jets (bars). [13]

In practise, the definition of a jet is not unique and is basically an algorithmic definition. First, the jet clustering algorithms use some "metric", which is not a real metric in mathematics sense, but a QCD motivated way to measure distances between hadrons in an experimental (η, ϕ) space. Together with the metric, algorithm uses some *combinatorial* process with usually at least one free parameter, the maximum jet cone radius, to cluster hadrons together. To define the 4-momentum of two particles clustered together, most usually the sum is taken. Jet algorithm must be both experimentally relatively easy to apply and give a correspondance between QCD final states, which are quarks or gluons, and the real end products which are the jets of hadrons. Naturally jet composition also consists of leptons and photons via EM interactions. Figure 6.2 shows an event reconstruction of a rare single top quark decay event at the CDF experiment, where missing energy is due to muon neutrino ν_μ.

Jet algorithms must be both infrared and collinear safe, thus they must be able to handle the soft QCD properties of parton evolution. The algorithmically found number of jets should stay the same even if partons radiates a soft gluon or one parton is replaced with a collinear pair of partons. This property is a requirement from QCD perturbation theory, which executes an exact cancellation of real and virtual diagrams of soft and collinear emissions in all orders of α_s [29]. This is a requirement from unitary, i.e. $P(\text{resolvable emission}) + P(\text{virtual + unresolvable}) \equiv 1$. Thus, the jet algorithm must obey this requirement if one wants to be able to compare experimental jet rates with theoretical parton level calculations. The most commonly used jet algorithm nowadays in hadron colliders is the anti-k_T [14].

6.3 Underlying event

Because the colliding hadrons are composite particles, it is usual that even in hard, high-p_T interactions, there is soft particle production in addition to the hard process. This is the so-called *underlying event* (UE), which ultimately makes precision QCD measurements in hadron colliders extremely difficult. The case is different with lepton colliders, which have a well understood non-composite initial state such as $|e^+e^-\rangle$. Formally, after one beam cross in a hadron collider, the set of final states is

$$\bigcup_i^H \{|\text{Hard FS}\rangle\}_i \cup \{|\text{UE FS}\rangle\}_i \cup \bigcup_j^S \{|\text{Soft FS}\rangle\}_j, \tag{6.5}$$

which is a union over H hard scattering final state (FS) events and their underlying events, and a union over S soft diffractive or non-diffractive events. The total number of simultaneous hadron-hadron pair collisions is $H + S$, the number of pile-up. One must notice that the underlying event is not exactly the same thing to model as soft non-diffractive events, even if the phenomenology common to both is highly similar.

Now assume a hard gluon-gluon collision. All the rest of the partons, i.e. proton remnants can and usually will have, soft interactions which create unwanted background activity in measurements. This must be corrected e.g. in jet energy scale analysis by statistical subtraction techniques. Currently the modelling of underlying event depends on phenomenological models [48] implemented in the Monte Carlo event generators, which of course can lead easily to circular reasoning by tune-fit-tune procedure.

Usually these models for UE are based on the impact parameter b pictures with some matter distribution function, like double Gaussian shape, and some hard scale p_T cut-off, which depends on $\sqrt{s}$. This p_T cut-off regularizes the hard scattering $2 \to 2$ processes which have divergent cross sections when $p_T \to 0$, which would otherwise grow to infinity. Below this cut-off, in UE modelling sense soft gluons are exchanged and stretching the corresponding color "strings" between partons explains phenomenologically the soft hadron production. [48]

The proton remnants can have multiple interactions (MPI) even with the hard process and there can be even a secondary hard scattering involved. Also initial state and final state radiation are part of this UE scheme. Tuning the underlying event is such a difficult subject, that models tuned at the center-of-mass energy $\sqrt{s} = 1.96$ TeV of the Tevatron do not easily extrapolate to the energies $\sqrt{s} = 7$ or 8 TeV of the LHC.

6.4 Luminosity

The instantaneous luminosity $\mathcal{L}$ of a collider describes the rate of how often the colliding particle bunches have an interaction, and has units of $[\text{cm}^2\ \text{s}]^{-1}$. One must remember that luminosity is purely collider related quantity, like a particle flux. Physics processes interaction probabilities do not depend on luminosity, but they depend on cross sections σ. Model of instantaneous luminosity multiplied with the number of N_B bunches circulating, is described in a simple form with [29]

$$\mathcal{L} = \frac{N_B f_0 N_p N_{\bar{p}}}{4\pi\rho^2} \approx 10^{31-32}\ \text{cm}^{-2}\text{s}^{-1} \quad \text{(Max at Tevatron)} \tag{6.6}$$

$$\approx 10^{34}\ \text{cm}^{-2}\text{s}^{-1} \quad \text{(Max at LHC)} \tag{6.7}$$

where f_0 is the bunch revolution frequency and N_p and $N_{\bar{p}}$ are the number of protons (antiprotons) per particle bunch. The bunch crossing frequency $\sim$ MHz ($\rightsquigarrow$ trigger rate) is given by $f_X = N_B f_0$. The parameter ρ describes the Gaussian transverse width of the beam. It can be expressed as $\pi\rho^2 = \epsilon\beta^*$, where ϵ is the beam emittance, which is the average spread or volume of beam particles in position-momentum phase space and β^* the accelerator optics parameter, both with units of length. The integrated luminosity $\int \mathcal{L}dt$ describes the amount of data collected over a period of time.

Increasing the luminosity increases the probability for simultaneous $p\bar{p}$ collisions, and this pile-up is usually modelled to be Poisson distributed with a mean $\nu \in \mathbb{R}$. This can be estimated based on the measured instantaneous luminosity $\mathcal{L}$, integrated over all N_B circulating bunches per revolution. Thus, the expected number ν of simultaneous inelastic collisions per bunch cross is

$$\nu = \frac{\mathcal{L}\sigma_{inel}}{N_B f_0}, \tag{6.8}$$

where e.g. $N_B = 36$ and the bunch revolution frequency $f_0 \sim 48$ kHz at the Tevatron run II, and the formula assumes that the instantaneous luminosity $\mathcal{L}$ is summed over N_B bunches.

The probability to have $k \in \mathbb{N}$ simultaneous (or none) collisions per bunch cross is then given by discrete Poisson distribution

$$P(X = k; \nu) = \nu^k \frac{e^{-\nu}}{k!}. \tag{6.9}$$

At the LHC, the mean pile-up can be as high as $\nu = 30$. It makes data collection faster, but requires very sophisticated pile-up detection (vertex reconstruction) and correction algorithms in the triggering and offline data-analysis.

6.5 Counting experiment

In a counting experiment, the cross section for a given physics process at the given $\sqrt{s}$ is experimentally defined as [29]

$$\sigma(\sqrt{s}) \triangleq \frac{\#\text{Number of observed events}}{\text{Integrated luminosity}} = \frac{N_{obs}}{\int_{\Delta t} \mathcal{L} dt}, \tag{6.10}$$

where $\int L\, dt$ is the integrated luminosity [cm^{-2}] with usually Gaussian uncertainty and N_{obs} is the number of observed signal event candidates with Poisson uncertainty $\sim \sqrt{N_{obs}}$. However, when taking experimental effects, i.e. detector acceptance probability and analysis efficiencies ϵ_i in account, the formula is

$$\sigma = \frac{N_{obs} - \langle B \rangle}{\epsilon_1 \cdot \epsilon_2 \cdots \epsilon_n \int L\, dt} = \frac{N_{obs} - \langle B \rangle}{A_{tot} \int L\, dt}, \tag{6.11}$$

where $A_{tot} \in [0, 1]$ is the total acceptance usually following log-normal distribution [10] and $\langle B \rangle$ is the expected number of unwanted, background events passing through the event selection. In a case of rare signal search, the irreducible background contamination by ordinary decays can be easily larger than the event count of rare signal itself. Systematic uncertainties are usually coming from selection efficiencies, background subtraction, integrated luminosity measurement, acceptance corrections, detector modelling and from theoretical models for a given signal process and background processes.

The cross section formula 6.11 can be easily generalized as binned cross section via matrix formalism to differential cross section $d\sigma/dx$ estimation, i.e. in practise to histogram vector $\mathbf{x}$ estimation of some variable x. This leads to a statistical, discretized (via histogram) inverse problem known as *unfolding*, naively solved as

$$\mathbf{x} = A^{-1}(\mathbf{y} - \mathbf{b}), \tag{6.12}$$

where $\mathbf{y} \in \mathbb{R}^n_+$ describes the folded measurements of each n histogram bin and $\mathbf{b}$ is the background estimate. However, applying this directly usually results in highly unstable inversion and requires suitable regularization techniques. Classic regularization tactics are Tikhonov regularization and truncated Singular Value Decomposition (SVD). The non-negativity constraint should be also included in the regularization, and it is a requirement what closed form estimators do not usually guarantee. The mixing between bins, known also as bin migration, is due to non-diagonal measurement matrix A_{ij}. This measurement matrix ultimately defines the probabilities of migrations from a histogram bin $i \to j$, and it must be very well understood, thus the measurement calibration and Monte Carlo modelling must be done with care. Also the background vector $\mathbf{b}$ must be known with low uncertainty.

Acceptance and efficiency

Above, the total acceptance A_{tot} was defined as the factor, which incorporates all the acceptance and efficiency corrections. However, the terminology has some differences, and sometimes the acceptance is defined as

$$A(\text{process; detector}) \triangleq \frac{\#\text{Low-level reconstructed}}{\#\text{Generated}} \in [0, 1], \tag{6.13}$$

which can be purely geometrical limiting factor, known as *geometrical acceptance*. A detector with a space angle reach of $\Omega = 4\pi$ steradians would be the optimal. Unfortunately, in practise the geometrical acceptances are far from that and for example in elastic scattering processes with no Roman Pots or similar very forward detectors, the acceptance is zero.

All the inefficiencies, dead times etc. of the detectors are part of the acceptance characterization, and they are not necessarily constant over time, e.g. due to detector ageing and different luminosity conditions.

An important factor to analysis comes from the online triggering efficiency

$$\epsilon_{trig}(\text{process; online chain}) \triangleq \frac{\#\text{Triggered}}{\#\text{Low-level reconstructed}} \in [0,1] \tag{6.14}$$

and highly important is also the offline selection efficiency

$$\epsilon_{sel}(\text{process; offline chain}) \triangleq \frac{\#\text{Selected}}{\#\text{Triggered}} \in [0,1]. \tag{6.15}$$

All these can be combined into same total acceptance A_{tot} correction by multiplication, assuming mutual independence of each.

Purity

The expected purity π at the end of the analysis chain is

$$\pi \triangleq \frac{\langle S \rangle}{\langle S \rangle + \langle B \rangle} \tag{6.16}$$

and when $\pi \to 1$, then $\langle B \rangle \to 0$. One can see that the purity factor is *not* directly used in evaluating the cross section in equation 6.11. However, it is related to the term $\langle B \rangle$ and estimation of background. Why this is so? Well, if one incorporates the effect of background as a multiplicative factor π in the numerator, one wrongly scales the background rate together with the measured signal rate, if the signal rate (cross section) turns out the be higher than expected.

Sometimes the background can be estimated without MC by using a *control sample*, which is a data sample only of background events, and then extrapolating the background rate from this sample.

Offline selection rules are traditional single-variate cuts, or cuts in the output of multivariate algorithms, which are often tuned to optimize quantity $\epsilon \cdot \pi \propto S/\sqrt{S+B}$, which is optimal only in a case when the estimated background $\langle B \rangle$ has vanishing uncertainty [40]. For rare signal searches, one usually wants to optimize roughly the ratio $S/\sqrt{B}$, i.e. maximize significance.

Based on the estimated efficiencies, acceptances and background estimates, one is able to do statistical correction when doing the counting measurement. The case with a well understood background class (small uncertainty), and a one signal class, the problem is well-posed. In a multi-class case, as in this work, with not well understood theoretical cross sections of each class, the problem is more or less ill-posed.

Chapter 7

Tevatron and CDF experiment

The Tevatron accelerator at Fermilab (Fermi National Accelerator Laboratory), USA, was the highest energy particle accelerator in the world before the LHC (Large Hadron Collider), CERN, Switzerland. It started its operation in 1983 as the first superconducting synchrotron, with two experiments, CDF and D∅. After the last major upgrade in 2001, the Tevatron accelerated protons and antiprotons in a ring with a circumference of 6.28 km and achieved a center of mass energy of $\sqrt{s} = 1.96$ TeV, with energy of 0.98 TeV per beam particle. The final shutdown of the Tevatron beam was on September 30, 2011.

7.1 Tevatron accelerator

The acceleration process at the Tevatron was done in several phases, which is a similar approach as in the LHC. Figure 7.1 shows a schematics of the whole acceleration chain. The MiniBoone and NuMI denote neutrino experiment baselines, which use the beam from the booster and the Main Injector to create a beam of neutrinos travelling under the ground, respectively.

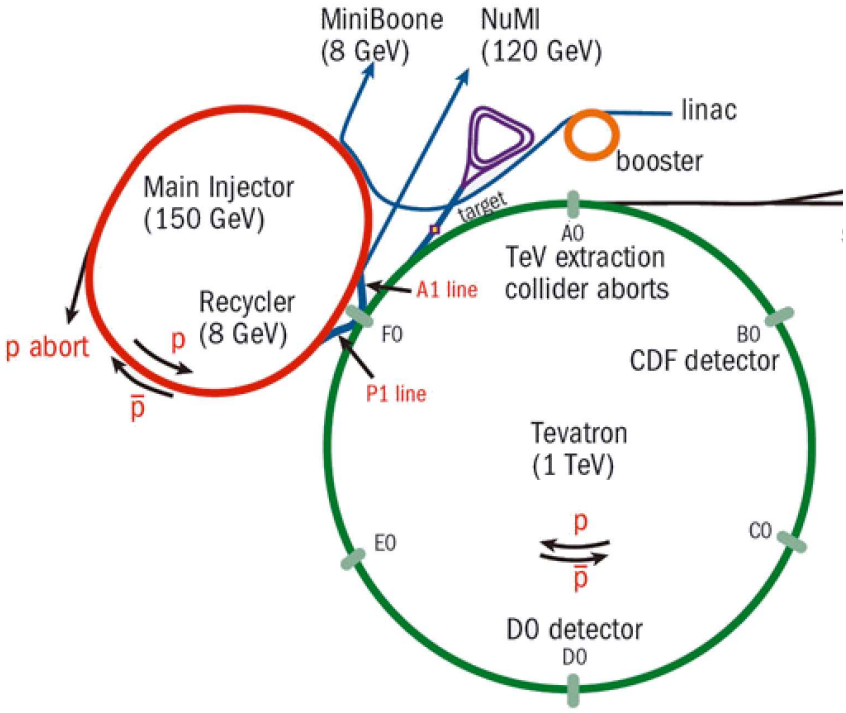


Figure 7.1: The Tevatron accelerator complex. [49]

The collection of protons starts by ionizing hydrogen gas which leads to ions with two electrons and one proton. Then the 750 keV Cockroft-Walton (CW) preaccelerator is used, which is a diode-capacitor ladder type voltage multiplier, similar in principle as in electronics. These negatively charged ions with charge q are accelerated using the CW with a positive voltage potential ϕ, which creates an accelerating electric field $\vec{E} = -\nabla\phi$ with force $\vec{F} = q\vec{E}$, and then passed to the 150 meter long linear accelerator (linac), where the ions achieve energy of 400 MeV. [27]

In general, radio frequency (RF) cavities are used to accelerate protons in each of the accelerators.

The creation of protons is made by shooting these ions through a carbon foil, which removes electrons. These protons are then transferred into the Booster, which is a synchrotron that accelerates protons into energy of 8 GeV. After the Booster follows the Main Injector which accelerates protons up to 150 GeV and produces 120 GeV protons for antiproton creations process.

The antiproton creation process is highly non-trivial, which is one of the reasons why the LHC uses only protons. However, using antiprotons allows both antiproton and proton bunches to travel in the same beam pipe. At the Tevatron, the Antiproton source creates antiprotons by colliding 120 GeV protons with a nickel target

which produces a small amount of antiprotons, which are then trapped, collected and stored in the accumulator ring. These are then passed into the Main Injector, which passes proton and antiproton bunches which travel in opposite directions in the same beam pipe, to the main accelerator.

The main accelerator keeps the charged particles on a circular track by using 774 Niobium-Titanium (NbTi) superconducting dipole magnets with liquid helium cooling which generate 4.2 Tesla bending magnetic fields $\vec{B}$, which apply an orthogonal force to charge q moving with velocity $\vec{v} = (v_x, v_y, v_x)$, towards the centre of ring. This can be expressed with relativistic, covariant Lorentz force law[1] as derivative of the 4-momentum P^μ with respect to the proper time τ of the particle as

$$F^\mu = \frac{dp^\mu}{d\tau} = qF^{\mu\nu}U_\mu, \tag{7.1}$$

where the 4-velocity and the 4-momentum of the particle are

$$U_\mu = (\gamma c, -\gamma v_x, -\gamma v_y, -\gamma v_z) \tag{7.2}$$
$$p^\mu = (E/c, \gamma m\vec{v}) = (\gamma mc, \gamma mv_x, \gamma mv_y, \gamma mv_z), \tag{7.3}$$

obeying $p^\mu = mU^\mu = mg^{\mu\nu}U_\nu$, where the factor $\gamma = 1/\sqrt{1 - \frac{|\vec{v}|^2}{c^2}}$. The usual antisymmetric electromagnetic field rank-2 tensor, encapsulating Maxwell's equations in space-time is

$$F^{\mu\nu} = \begin{pmatrix} 0 & -E_x/c & -E_y/c & -E_z/c \\ E_x/c & 0 & -B_z & B_y \\ E_y/c & B_z & 0 & -B_x \\ E_z/c & -B_y & B_x & 0 \end{pmatrix}. \tag{7.4}$$

Finally, the beam transverse size is controlled with focussing and defocussing magnetic "lenses" by using 240 NbTi quadrupole magnets. The modelling of these magnets are done classically by linear matrix approaches, where the magnets are modelled as thin lenses, i.e. Fourier optics analogy, and the full optics transport over the ring is modelled by chaining these matrices together as $M_n M_{n-1} \cdots M_1$. Unfortunately, the actual problem is non-linear, even chaotic in the sense of Lyapunov when it comes to beam stability, and modern computational software (like MAD-X [41]) try to use more flexible approaches.

[1]Non-relativistic version is $\vec{F} = q(\vec{E} + \vec{v} \times \vec{B})$.

7.2 CDF experiment

The Collider Detector at Fermilab (CDF) is an experiment, which collected proton-antiproton collision data from year 1985 (run 0), initiated the first upgrade in 1989 (run I) and the second upgrade in 2001 (run II). The multipurpose CDF detector weights around 5500 tons, and its diameters are around 12 meters in each space dimension. The detector is illustrated in Figure 7.2.

The most famous physics result of the CDF experiment was the discovery of the top quark, which was announced in February 1995. The experiment was also able to exclude a wide region of possible Higgs boson's masses and announced, together with DØ, evidence for a Higgs like particle in July 2, 2012. Two days later, the discovery announcement of a Higgs like particle was announced by the CMS and ATLAS experiments at CERN, with statistical significance over 5σ.

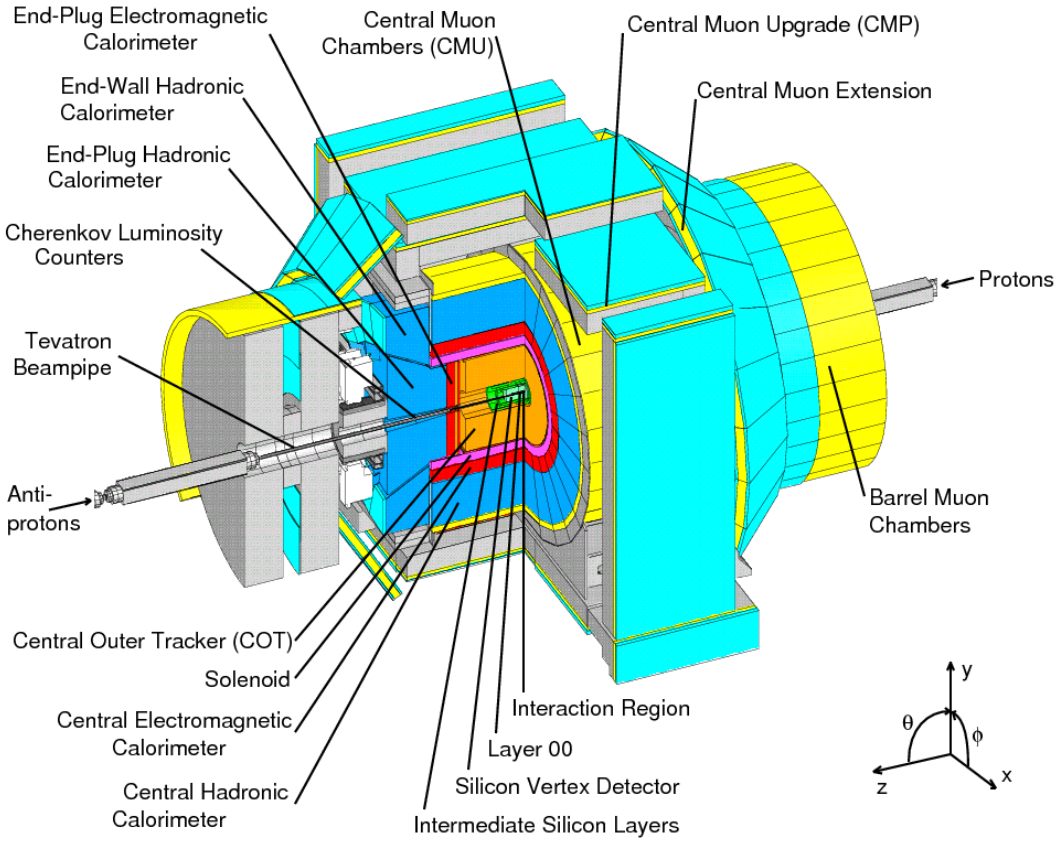


Figure 7.2: The CDF run II detector structure with subdetectors marked, and the followed coordinate system where positive pseudorapidities ($\theta \in [0, \pi/2]$) are in proton direction. [12]

SVX II - Silicon Vertex Tracker

The SVX II detector measures the position of charged particles as they transverse through the detector layers of silicon. The first layer 00 is especially radiation hard single sided silicon with 1.5 cm away from beam line, and the rest 5 layers of the SVX II are double sided silicon strip detectors. The outermost intermediate silicon layers

(ISL) have two double sided silicon layers. The total number of signal channels is about ~ 0.72 million. [11]

The measurements of charged particle hits are done in silicon *pn*-junctions with suitable amplification. Finally, the hits are algorithmically reconstructed as tracks by a fitting procedure. The spatial accuracy is in order of tens of micrometers, which is enough to discriminate between a primary $p\bar{p}$-collision vertex and a secondary vertex which is due to the decay of a B-meson or a D-meson, for example.

COT - Central Outer Tracker

The COT tracker is argon and ethane gas filled tracking drift chamber, which detects charged particles via ionization and drifting. The so-called sense wires collect free electrons left by ionization. The field wires with negative voltage collect the signal from positive ions in gas, but only the electrons are actually *drifting* in the gas. The COT has eight "superlayers" with 12 wire-layers each. [11]

In the COT tracker, the positions of the charged particles are measured and due to solenoidal 1.5 Tesla magnetic field, the trajectories of charged particles bend in helices. From the bending, 3-momentum of the particle can be measured via a fitting procedure. Higher the momentum of the particle, less bending and vice versa.

Electromagnetic calorimeter

The electromagnetic calorimetry is based on lead (Pb) sheets sandwiched with a scintillator material. The charged particle creates an EM shower by interacting with the lead, and the shower creates visible light by ionization in the scintillator, which is transferred by light guides to photomultiplier tubes. [11]

There is nearly linear correlation with the amount of light observed and the energy of the original incoming particle (photon or electron/positron). The discrimination between a photon and electron can be done by combining the information from tracking, i.e. photon does not leave a track.

Hadronic calorimeter

The hadronic calorimetry is based on iron (Fe) sheets sandwiched with a scintillator material. The leading idea is that the EM shower should be finished before the hadronic calorimeter. Hadrons should leave only a small energy deposit in the EM calorimeter, and thus most of the energy should be deposited in the hadronic part. The hadronic calorimeter has physical size much larger than the EM calorimeter. This is due to the fact that hadron are more massive, do not start showering as early as photons or electrons/positrons, and need more interaction length to have a nuclear collision. After the hadronic calorimeter, there is an iron absorber which is used to absorb the hadronic shower generated in the hadronic calorimeter. [11]

Both electromagnetic and hadronic calorimeter are actually made of separate calorimeters over the pseudorapidity η: central, end-wall and end-plug as shown in Figures 7.2 and 7.3. This means that they are not absolutely homogenic, and this fact must be taken account in characterization and calibration.

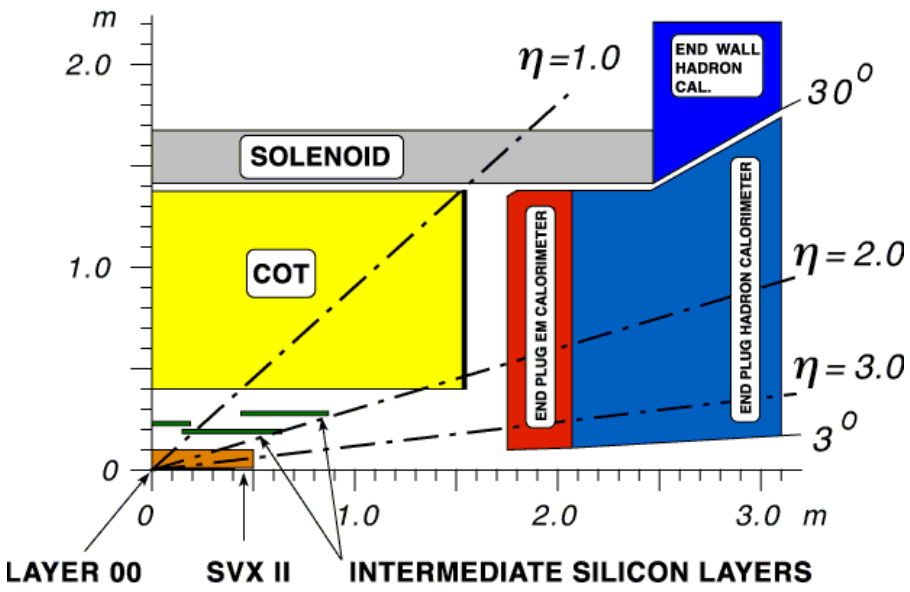


Figure 7.3: The CDF run II detector cutaway view exposing the pseudorapidity η-span. [11]

Muon chambers

The muon chambers are the last detectors (in radial direction), designed to tag the muons, which go through all the other detectors without any significant deposit in the calorimeters. However, as charged particles, they can be tracked and thus their momentum can be measured in the COT. This momentum can then be compared with the muon chamber track signal. The muon chambers are gas filled drift chambers, as the COT. [11]

Neutrinos, which do not carry either electric or strong (color) charge, do not leave any signal in the CDF detector. They only interact via electro-weak interaction, with extremely small cross sections. The presence of neutrinos can be thus known in this kind of general purpose detector only by utilization of conservation laws, e.g. by calculating missing transverse energy $E_T = E \sin\theta_{CM}$ or momentum $p_T = |\vec{p}| \sin\theta_{CM}$. Transverse quantities are favoured in hadron colliders, because before collision, the colliding particles (partons) have zero energy or momentum sum only in transverse plane.

Cherenkov luminosity counters

Cherenkov luminosity counters are there to measure the instantaneous luminosity. Cherenkov light production is initiated by charged particles, which move faster than light in dielectric, transparent medium. The after effect is visible light in a cone shape, dominated by blue end of the spectrum, which can be measured with photomultiplier tubes. This effect is analogous to breaking the speed limit of mechanical sound waves, e.g. when a bullet travels faster than sound waves in the air.

7.3 Monte Carlo chain

Because modelling real physical measurables in particle physics is far from closed form algebraic expressions with analytical solutions, Monte Carlo simulations must be used. The term Monte Carlo here means simply numerical integration techniques. The most obvious targets are phase space integrals in the cross section formulas and statistical models of detector responses.

Table 7.1: A typical simulation and event reconstruction chain.

1. Event "physics" generation (PYTHIA, SHERPA...)
Input parameters: Initial state (e.g. $\|p\bar{p}\rangle$), $\sqrt{s}$, list of processes
High and low-p_T processes, underlying event,
ISR, FSR, hadronization
Output: Final state 4-momentum vectors with particle IDs
→ 2. Detector simulation (Geant)
Detector geometry, magnetic fields, material response (calorimeter showering) and electrical (sensor) response
Output: Low level signals with $\mathcal{O}(10^8)$ of channels
→ 3. Offline event reconstruction (custom software)
Reconstruction of high level objects (like jets, tracks)
from low-level signals
Output: High level physics objects for physics analysis algorithms

Monte Carlo event generators, like PYTHIA, are programs designed to generate high-energy collision simulations. These generators simulate both theoretically well understood hard scattering events and phenomenologically understood soft scattering, diffraction and hadronization. For the hard scattering matrix elements are needed, which are nowadays often calculated at the given order using computer algebra software like Feyncalc or MADGraph. Finally numerical integration is used together with the matrix elements to randomly sample phase space, to obtain (integrated) cross sections.

The simulation of a hadronic collision goes as follows. It starts with two initial states, e.g. two protons with the given $\sqrt{s}$ energy. Next, the generator generates the hard sub-process, e.g. $q\bar{q} \to Z^0 \to e^+e-$. After this, the generator simulates initial state radiation (ISR) and final state radiation (FSR), i.e. gluon (or photon) emissions. Gluons can emit new gluons, or split into quark-antiquark pairs. The evolution of this process is known as *parton showering*, which can be treated reasonably well under perturbative QCD. After evolving the parton shower, partons must be combined into colourless hadrons, this is called *hadronization*. The hadronization process is based purely on phenomenological, QCD inspired models, like PYTHIA's Lund color string *stretch-and-split* model or HERWIG's clustering approach. After hadronization, the generator adds hadronic decays. Even more complicated is to handle the underlying event, i.e. dealing with proton remnants and their collisions and multi-parton interactions (MPI). [48]

The parton shower is evolved as long as all partons have energy low enough, then they are combined into hadrons. The generation of hadrons stops after the generation of "stable" final state particles like pions, kaons etc.

The detector geometry and electrical response simulation is done usually using the software called Geant. This takes the event generator output, which includes the particle IDs (i.e. charge, lifetime etc.) and 4-momentum vectors of each, and evolves these particles through the layers of sensors. It does the simulation of interaction of particles in the detector, which includes particle ionisation in trackers, energy deposition in calorimeters and the decays of intermediate particles and radiation. The output of Geant is the raw result of all the electric signals of each detector, in a comparable and same format to real measurements.

These raw signals can be passed through the event reconstruction algorithms, which reconstruct e.g. particle trajectories (using Kalman filter or similar iterative estimation), primary and secondary vertices, time of flight (TOF) information, particle transverse momentum, transverse energy deposits, particle identification and jet reconstruction. The whole simulation chain is summarized in Table 7.1, where the event reconstruction phase works naturally with simulated (Geant) or real data.

In this work, PYTHIA 6.4x was used to generate diffractive and non-diffractive (minimum-bias) events, which were then simulated with Geant 4 CDF simulation chain. Actual variables used were calorimeter deposits and number of reconstructed tracks. The non-diffractive events included both perturbative QCD and soft, non-perturbative events. See Appendix A.4 for details about the PYTHIA simulation.

Chapter 8

Multivariate analysis

The experimental target of this study is to classify diffractive $p\bar{p}$-collisions on an collision-by-collision basis, and to obtain integrated cross sections. Classic way of doing this, or any other high energy physics analysis, is to introduce 1-dimensional reductive cuts on the measured variable histograms to select events. In the case of diffraction, this basically means some definition of a rapidity gap, which needs some threshold on particle production or energy inside the gap and so on. These gap defining rules are usually more of an ad-hoc nature than mathematically or physically well defined. When this problem is understood as a statistical inverse problem, 1-dimensional reductive analysis is not usually the most precise approach.

8.1 Probabilistic classification

Now when classification is understood as a statistical *inverse problem*, there must be the corresponding *forward problem*. In this case, the forward problem means a way to calculate estimates of physically measurable variables. This is usually solved by using the Monte Carlo simulator (like PYTHIA) together with detector geometry and electrical response modelling (like Geant), and the event reconstruction algorithm chain. Monte Carlo generator can be basically either based on sound physical models, like perturbative QCD expansion, or purely parametric fits to reproduce distributions and cross sections measured in experiments. The forward problem parametrization defines our *supervised* classification scheme, and the inverse solution estimate is as good as the simulation model.

Let us now define our probabilistic multivariate classification scheme. Probability calculus starts with tuplet $(\Omega, \mathcal{A}, P)$, which is called a *probability space*. Here Ω is a sample space, which is the set of all possible outcomes. $\mathcal{A}$ is a σ-algebra with corresponding properties and it is a set of subsets of Ω. Every $A \in \mathcal{A}$ is called an *event*. An event with only one element is called *elementary*. Finally, $P : \mathcal{A} \to [0, 1]$ is a function which maps each event $A \in \mathcal{A}$ with probability $P(A)$. Probability of disjoint events is $P(A_1 \cup A_2 \cup \dots) = P(A_1) + P(A_2) + \dots$, with $A_i \cap A_j = \emptyset$ if $i \neq j$. Naturally P is our *probability measure* with $P(\Omega) = 1$ and $P(\emptyset) = 0$. [38]

Now the formal structure of the probability space is given. Let us also define a real-valued d-dimensional *continuous random vector* $\mathbf{X} \in \mathbb{R}^d$. Especially, it must be assumed that there is a function $f_{\mathbf{X}} : \mathbb{R}^d \to [0, \infty)$ such that there exists probability [38]

$$P(\mathbf{X} \in A) = \int_A f_{\mathbf{X}}(\mathbf{x})\, d\mathbf{x}, \tag{8.1}$$

where the event $A \subset \mathbb{R}^d$ is a set with interesting vector values. The function $f_{\mathbf{X}}$ is known as *likelihood* function or multidimensional *density* function[1]. One must note that the value $f_{\mathbf{X}}(\mathbf{x})$ is not a probability, but the integral over Ω must be equal to one.

Here one must not get confused with the terminology, because often one talks about "events" i.e. collisions and not probability space events, which can be sets. But for now on, the term event is used only in physics sense to denote a realization of the random variable $\mathbf{X}$.

Now, each collision is modelled as a pair of continuous and discrete random variables $(\mathbf{X}, C)$, where C is the discrete random variable representing scattering process "class". Let us number each scattering process with the number $C \in \mathcal{C}$, where $\mathcal{C} = \{1, 2, \dots, |\mathcal{C}|\}$ is the set of possible class labels. Each class represents different diffractive/non-diffractive processes, e.g. 1 can be double-diffractive processes, 2 central-diffractive and so on.

First of all, the *a priori* probabilities of each class are denoted with

$$P_j := P(C = j), \quad \sum_j P_j = 1, \quad j = 1, \dots, |\mathcal{C}|. \tag{8.2}$$

These probabilities should come from theory, or from previous experiments. A class

[1] Likelihood has also a slightly different meaning in a case of parameter estimation, but here density and likelihood are synonyms.

likelihood or density function for each class is

$$f_j(\mathbf{x}) = f_{\mathbf{X}}(\mathbf{x}|C = j), \quad j = 1, \ldots, |\mathcal{C}|. \tag{8.3}$$

This means that the total density function of a random variable $\mathbf{X}$ is here

$$f_{\mathbf{X}}(\mathbf{x}) = \sum_{j=1}^{|\mathcal{C}|} f_j(\mathbf{x}) P_j, \tag{8.4}$$

which can be seen by integrating over $\mathbf{x}$ and summing over j, such that the total probability is

$$P(\mathbf{X} \in A) = \sum_j P(\mathbf{X} \in A|C = j)P(C = j) \tag{8.5}$$

$$= \int_A \left(\sum_j f_{\mathbf{X}}(\mathbf{x}|C = j)P(C = j) \right) d\mathbf{x}, \tag{8.6}$$

where $A \subset \mathbb{R}^d$.

After the realization (measurement) $\mathbf{x}$ of the random variable $\mathbf{X}$, the discrete distribution with point probabilities $P(C = j|\mathbf{X} = \mathbf{x})$ exists. These are the so-called *a posteriori* probabilities of a $p\bar{p}$-collision to be generated by any of the process classes. These can be interpreted as updating our prior knowledge with new data by

$$\textit{posterior} \propto \textit{likelihood} \times \textit{prior}. \tag{8.7}$$

Formally, a posteriori probabilities of the measurement vector $\mathbf{x}$ are obtained from the Bayes' rule [7]

$$P(C = j|\mathbf{X} = \mathbf{x}) = \frac{f_{\mathbf{X}}(\mathbf{x}|j)P(j)}{f_{\mathbf{X}}(\mathbf{x})} = \frac{f_j(\mathbf{x})P_j}{\sum_{j'=1}^{|\mathcal{C}|} f_{j'}(\mathbf{x})P_{j'}}, \tag{8.8}$$

where $j = 1, \ldots, |\mathcal{C}|$. It is the task of a classifier algorithm to estimate these probabilities.

Training and classification

The classifier is trained to estimate the posterior probabilities using a *training set* of random vector and class label pairs

$$\mathcal{T} = \{(\mathbf{X} = \mathbf{x}, C = j)\}, \tag{8.9}$$

where each i-th pair is generated using Monte Carlo, hand classified measurements or any other way. This training sample is run through an optimization algorithm which constructs the multidimensional class likelihoods, or directly generates the discrimination mappings which define the classification *decision boundaries* of the multivariate space.

In practise, most of the classifiers do *not* estimate posterior probabilities (likelihoods) but are based on defining decision boundaries in the space of $\mathbf{x}$, which can be proven to be equivalent way as far as hard classifications are considered. These can be seen as mappings

$$g : \mathbf{x} \mapsto \{1, 2, \ldots, |\mathcal{C}|\}. \tag{8.10}$$

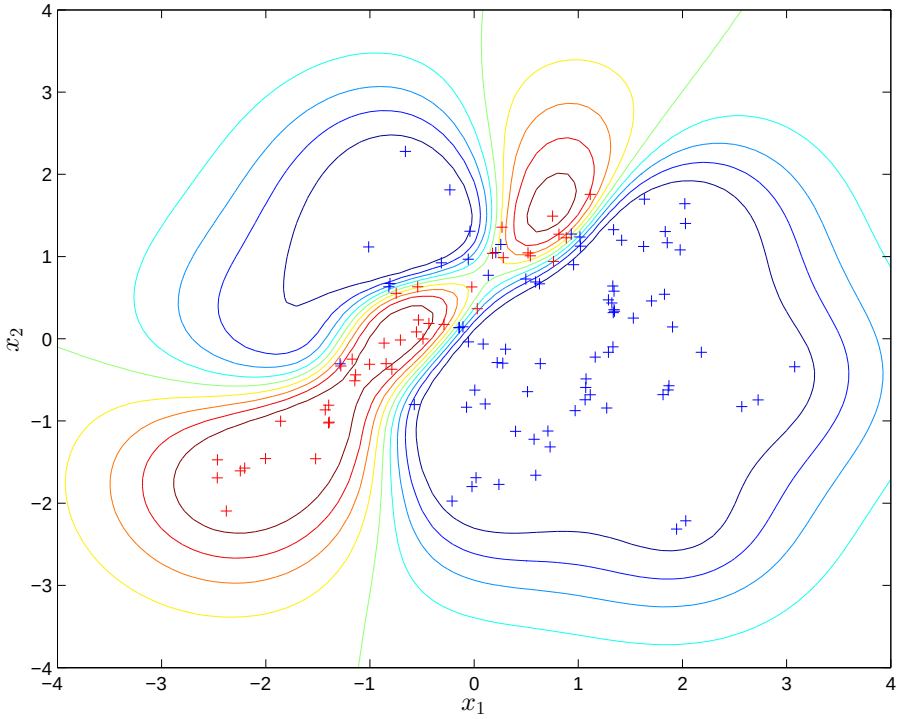


Figure 8.1: Decision boundary illustration (green lines) of an arbitrary two class problem and a non-linear classifier in the multivariate space $\mathbb{R}^2$. Red and blue markers and line-contours denote class vectors and their likelihood distributions $f_j(\mathbf{x})$, respectively.

Decision rule mappings g define *decision regions* as

$$\mathcal{R}_j = \{\mathbf{x} \in \mathbb{R}^d : g(\mathbf{x}) = (C = j)\}, \tag{8.11}$$

and thus $\mathcal{R}_j$ is the region in $\mathbb{R}^d$ where the posterior of class j is the highest. These decision regions can be defined by affine hyperplanes or in general, by nonlinear manifolds (or surfaces). Classic linear algorithms are Fisher's linear discriminant, support vector machines (SVM) and logistic regression. Nonlinear classifiers are usually neural networks and decision trees, but nowadays different type of *kernel machines* can be used to define nonlinear decisions, e.g. SVM with kernels. See Figure 8.1 for a demonstration of decision boundaries and class densities.

Bayes' minimum error classifier, optimal in Bayesian sense [7], does the *hard classification* according to

$$g^\star(\mathbf{x}) = \underset{j=1,\ldots,|\mathcal{C}|}{\arg\max}\, P(j|\mathbf{x}) = \underset{j=1,\ldots,|\mathcal{C}|}{\arg\max}\, f_j(\mathbf{x})P_j, \tag{8.12}$$

because the denominator in equation 8.8 does not depend on C. This is also known as Maximum a Posteriori (MAP) decision rule.

Theoretically, the lower bound for classification error is the Bayes error rate, given by [7]

$$e(g^\star) = 1 - \sum_{j=1}^{|\mathcal{C}|} \int_{\mathcal{R}_j} f_j(\mathbf{x})P_j \, d\mathbf{x}, \tag{8.13}$$

which is always non-zero for a problem with overlapping class densities.

In many particle physics problems where the interesting "signal class" is very rare, it is not usually possible to use the Bayes' minimum error criteria. By using the criteria directly, not a single signal event would be classified right, and the efficiency would be zero. Thus, different kind of criteria are being used like minimum efficiency.

However, in this work, a classifier with probabilistic output is used. This has many obvious benefits. First, one can see how certainly the vector to be classified belongs to each class, not just binary yes-no answer. Second, as it is shown later, class fractions $\sim$ cross sections can be estimated in a coherent way by averaging a posteriori probabilities.

8.2 Cross sections via probabilities

It is well-known that conditional expectation values obey the so-called *iterated expectation* relation [38]

$$\mathbb{E}[h(\mathbf{X},\mathbf{Y})] = \mathbb{E}[\mathbb{E}[h(\mathbf{X},\mathbf{Y})|\mathbf{Y}]] = \mathbb{E}[\mathbb{E}[h(\mathbf{X},\mathbf{Y})|\mathbf{X}]], \tag{8.14}$$

where $\mathbf{X},\mathbf{Y}$ are random vectors and $h(\mathbf{X},\mathbf{Y})$ some arbitrary function of those. This means that the prior mean of a pair $(\mathbf{X},\mathbf{Y})$ is equal to the average of all posteriori means, averaged over all possible $\mathbf{X}$ or $\mathbf{Y}$.

This is easily proved, using for example the case where pair $(\mathbf{X},\mathbf{Y})$ has continuous distribution

$$\mathbb{E}[\mathbb{E}[h(\mathbf{X},\mathbf{Y})|\mathbf{X}]] = \int \mathbb{E}[h(\mathbf{X},\mathbf{Y})|\mathbf{X}=\mathbf{x}] f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \tag{8.15}$$

$$= \int\int h(\mathbf{X},\mathbf{Y}) f_{\mathbf{Y}}(\mathbf{y}|\mathbf{X}=\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} d\mathbf{y} \tag{8.16}$$

$$= \int\int h(\mathbf{x},\mathbf{y}) f_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y}) d\mathbf{x} d\mathbf{y} \tag{8.17}$$

$$= \mathbb{E}[h(\mathbf{X},\mathbf{Y})]. \tag{8.18}$$

By definition, the conditional expectation values for $\mathbf{X}=\mathbf{x}$ or $C=j$ are given by [38]

$$\mathbb{E}[h(\mathbf{X},C)|\mathbf{X}=\mathbf{x}] = \sum_{j=1}^{|\mathcal{C}|} h(\mathbf{x},j) P(j|\mathbf{x}) \tag{8.19}$$

$$\mathbb{E}[h(\mathbf{X},C)|C=j] = \int_{\mathbb{R}^d} h(\mathbf{x},j) f_j(\mathbf{x}) d\mathbf{x}, \tag{8.20}$$

and the corresponding expectation values are given by the iterated expectation relation as

$$\mathbb{E}[h(\mathbf{X},C)] = \mathbb{E}[\mathbb{E}[h(\mathbf{X},C)|\mathbf{X}]] = \int_{\mathbb{R}^d} \sum_{j=1}^{|\mathcal{C}|} h(\mathbf{x},j) P(j|\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \tag{8.21}$$

$$= \mathbb{E}[\mathbb{E}[h(\mathbf{X},C)|C] = \sum_{j=1}^{|\mathcal{C}|} \int_{\mathbb{R}^d} h(\mathbf{x},j) f_j(\mathbf{x}) d\mathbf{x} P_j. \tag{8.22}$$

After the a posteriori probabilities $P(j|\mathbf{x}_i)$ are estimated using the multivariate algorithm for each measurement vector $\mathbf{x}_i$ and for each class j, the empirical estimates of the expectation values can be calculated. In this application, these are the relative cross section estimates of the scattering processes. The selection of class is done by using an indicator function which has no dependence on $\mathbf{X}$

$$h_I(j;k) = \begin{cases} 1, & \text{when } j = k \\ 0, & \text{when } j \neq k, \end{cases} \tag{8.23}$$

where the parameter $k = 1,\ldots,|\mathcal{C}|$ is used to encode the selected scattering process class. Now using Equation 8.21, and a priori measured total inelastic cross section σ_{inel} for normalization

$$\frac{\sigma_k}{\sigma_{inel}} \triangleq \mathbb{E}[\mathbb{E}[h_I(C;k)|\mathbf{X}]] \tag{8.24}$$

and by using Kolmogorov's strong law of large numbers which states for an independent, identically distributed random vectors $\mathbf{X}_i \sim f_{\mathbf{X}}(\mathbf{x})$ with $\mathbb{E}\|\mathbf{X}_1\| < \infty$, that

$$\frac{1}{N}\sum_{i=1}^{N}\mathbf{X}_i \to \mathbb{E}[\mathbf{X}_1] \tag{8.25}$$

holds almost certainly, when $N \to \infty$. Using this, the expectation operator over $\mathbf{X}$ is replaced with summation, and by using equation 8.24 and 8.19 results in

$$\frac{\sigma_k}{\sigma_{inel}} \cong \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}[h_I(C;k)|\mathbf{X}=\mathbf{x}_i] \tag{8.26}$$

$$= \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{|\mathcal{C}|}h_I(j;k)P(j|\mathbf{x}_i) \quad \square \tag{8.27}$$

for all k, where $N \in \mathbb{N}$ is the number of collision events in the event sample. In a case where our priors are correct and likelihood functions are correctly estimated i.e. data matches the MC modeling and the multivariate algorithm is working, expression 8.27 returns relative cross sections for each class. Discrepancies between a priori expected and a posteriori obtained results indicate that something is poorly understood either in experiment or in theory.

Also can be seen that this probabilistic way of estimating cross sections does implicitly efficiency ϵ and purity π corrections, which must be done explicitly when using either simple cut and count, or hard classifiers. However, this assumes that all the events belong to any of the classes of $\mathcal{C}$. The events which do not belong in $\mathcal{C}$, should be removed from the sample somehow before the classification procedure, or their effect should be estimated some way and then subtracted from the final results (see analysis Equation 10.3).

In a similar way as above, by using a posteriori probabilities as weights, expectation value estimates for many other physical variables can be calculated. For example, the inclusive distribution $dN/d\eta$ can be estimated for any given scattering process $C = j$. This means that a classifier with probabilistic output is a general tool for different physics analysis purposes.

8.3 Classifier performance metrics

In general, estimating the performance of a classifier is based on dividing the Monte Carlo sample into train and test sample(s). Of course, this is the most practical approach, not always the most optimal.. Theoretically, the generalization of these empirical estimates can be discussed under Vapnik-Chervonenkis (VC) theory.

So, to be able to calculate the performance metrics, one must know the class label of each event vector in the test sample, which is the case with Monte Carlo.

The most important metrics can be read from the so-called *confusion matrix*, here denoted with $A \in \mathbb{N}^{|\mathcal{C}|\times|\mathcal{C}|}$. By convention, rows are the right classes and columns are the hard classification estimates. It has the following properties

$$\sum_{i,j}[A]_{ij} = N, \quad \text{Total number of test vectors} \tag{8.28}$$

$$\sum_{j}[A]_{ij}/N = f_i, \quad \text{Fraction of each } i\text{-th class} \tag{8.29}$$

$$[A]_{ii}/\sum_{j}[A]_{ij} \triangleq \epsilon_i, \quad \text{Classification efficiencies for each class} \tag{8.30}$$

$$[A]_{ii}/\sum_{j}[A]_{ji} \triangleq \pi_i, \quad \text{Classification purities for each class} \tag{8.31}$$

$$\text{trace}(A)/N, \quad \text{Total classification accuracy} \tag{8.32}$$

$$1 - \text{trace}(A)/N, \quad \text{Total classification error} \tag{8.33}$$

and often its row normalized version $A_n \in [0,1]^{|\mathcal{C}|\times|\mathcal{C}|}$ is used, which has class efficiencies on its diagonal, or similarly column normalized, which has purities on its diagonal. Thus, each empirically estimated element

$$[A_n]_{ij} \triangleq \mathbb{E}_{\mathbf{x}|C=i}\left[h_I(g(\mathbf{x}); j)\right] = P(g(\mathbf{x}) = j|C = i) \tag{8.34}$$

gives the conditional probability of classifying a vector originating from the i-th class to the j-th class. Note that using confusion matrix is not limited to multivariate algorithms, it can be also calculated when estimating performance of a cut and count analysis.

8.4 Inversion based on the hard confusion matrix

In hard classification, the class fraction estimates by directly counting the classifications, are biased, if the efficiencies and purities are not both nearly perfect for each class. There are at least two ways to try to correct the class fraction estimates $\sim$ cross sections by using the Monte Carlo estimated confusion matrix A. The reason why this confusion matrix was denoted with same symbol as the bin migration measurement matrix A in the context of unfolding, is because they have an analogous role (bins $\leftrightarrow$ classes). In the unfolding, we have a well-known background, which is at least on formal level subtracted from the measurement and then the unfolding of bin migration is solved. However, because the bin migration is caused by detector effects it is fundamentally different kind of "mixing" than what is encountered in this classification problem. Here the mixing is driven by non-unique physical signatures of events, not just by detector (resolution) effects.

In the classification, in principle, we could use a similar approach, estimate the normalized confusion matrix A and solve "the true" class fractions by multiplying the measured class fractions vector from left with A^{-1}. Assuming that all events belong to one of the classes. However, this again needs specific regularization techniques with problem specific prior distributions. General smoothness ℓ_2-prior, like Tikhonov induced Gaussian, has no use. That is because there is no general reason for class fractions to be smooth when represented in a vector form. Inverting the confusion matrix was experimented, and it yielded highly unstable results with several kinds of classic regularization techniques when the confusion matrix was not exactly correct.

The second approach is well-known also in unfolding context, bin-by-bin correction factors, which cannot be really considered as a real inversion method. The use of Monte Carlo estimated efficiencies ϵ_j and purities π_j to correct "class migration" by dividing each obtained class-fraction with ϵ_j and by multiplying with π_j, is a kind of "correction by expectation" technique. Also, it is basically as Monte Carlo driven method as it can be, using only integrated class fractions, class-by-class and only in a scalar way. Doing the purity correction as a multiplicative factor has also the problems explained in section 6.5.

The approach of this work is to use empirical event-by-event probabilities to correct the "class migration", and not to use the integrated information from the confusion matrix. This event-by-event correction can be seen as the fundamental difference in comparison to other techniques. It has natural non-negativity constraint via probabilities, and does the correction based on both likelihood $f_j(\mathbf{x})$ (data) and prior information P_j, instead of looking at integrated quantities in matrix A.

Ultimately, this discussion about correction methods is application specific, what is the background uncertainty, how many process classes, can the corrections be based on multiplicative factors ($\sim$ stable ratios) or should there be subtractive or additive corrections, how good the chosen inversion approach is found in simulation[2] and so on.

[2]Is the simulation based on *inverse crime*, i.e. does it naively use exactly the same sampling and parameters in both forward problem and inversion.

8.5 Data pre-processing

For algorithmic purposes, the components of data vectors are often normalized using some component by component mapping $f : \mathbb{R} \to \mathbb{R}$. Usually this is done to avoid biasing the classifier training, e.g. to make sure that not a single component can dominate the regularization procedure. Standard way of doing this is by mean-variance normalization

$$f_{MV} : x_i \mapsto (x_i - \langle x_i \rangle) \,/ \sqrt{\mathbb{V}\mathrm{ar}(x_i)}, \quad \forall i \in [1, d] \tag{8.35}$$

where means $\langle x_i \rangle$ and variances $\mathbb{V}\mathrm{ar}(x_i)$ are estimated from the training set. These values are then used also for normalizing the real data sample.

Logarithmic normalization

$$f_{log} : x_i \mapsto \ln\,(\epsilon + x_i), \quad \forall i \in [1, d], \quad \epsilon \in \mathbb{R}_+ \tag{8.36}$$

is often done when the vector components are exponentially distributed, i.e. positive with long tailed distribution and often $\epsilon = 1$. Empirically, calorimeter deposits and track multiplicities are exponentially distributed. This transformation can be done in combination with the mean-variance normalization, by first applying the logarithmic transformation. Effectively, the logarithmic transformation makes exponentially distributed variables more Gaussian distributed.

Normalization can be done also as a vector valued mapping $f : \mathbb{R}^d \to \mathbb{R}^d$. To normalize each vector on the surface of hypersphere S^{d-1} in Euclidean space, a unit-length mapping

$$f_{UL} : \mathbf{x} \mapsto \frac{\mathbf{x}}{\|\mathbf{x}\|_2} \tag{8.37}$$

can be done. This has different kind of properties, and it preserves the orientation of each vector but length differences vanish. This mapping is not always suitable, because it can amplify noisy vectors to look like signal vectors.

More usual vector valued mapping is to do d-dimensional rotation to decorrelate components. This is usually called whitening and can be done using the Principal Component Analysis (PCA), also known as the Karhunen-Loève transform. In practise, this means calculating the covariance matrix and finding numerically the corresponding eigenvectors as new orthogonal basis vectors for the data representation, and it can be shown to be optimal way in variance maximizing sense. At the same time, dimension reduction is possible using less than d basis vectors when synthesizing data as a linear combination of the new basis vectors. However, whitening is not in general necessary with multivariate algorithms and one also loses physical interpretation of each variable after whitening, due to rotation.

Chapter 9

Regularized Multinomial Logistic Regression

Multinomial Logistic Regression (MLR) is a multi class extension of the classic logistic regression. Logistic regression is a bivariate response model which is used to predict probabilities of two classes. Multinomial logistic regression does not assume independence of components of vector $\mathbf{X}$, and does not make any assumptions on likelihood distributions $f_j(\mathbf{X})$, like Gaussianity etc., which is a clear advantage over other simple classifiers.

9.1 Discriminative model

MLR is a so-called *discriminative* classifier, which means it does not model likelihood functions $f(\mathbf{x})$ directly like *generative* classifiers. The purpose of MLR is to predict probabilities $P(C = j|\mathbf{X} = \mathbf{x})$. MLR assumes a parametric form $t_j(\mathbf{x}, \theta)$ for the distribution as [38]

$$P(C = j|\mathbf{X} = \mathbf{x}) = t_j(\mathbf{x}, \theta) \tag{9.1}$$

and estimates the parameters θ of this distribution from the training data. In this work, a linear ℓ_2 inner product model MLR was used

$$P(C = j|\mathbf{X} = \mathbf{x}; \mathbf{w}) = \frac{\exp(\langle \mathbf{w}_j, \mathbf{x} \rangle)}{\sum_{i=1}^{|\mathcal{C}|} \exp(\langle \mathbf{w}_i, \mathbf{x} \rangle)}, \quad j = 1, \ldots, |\mathcal{C}| \tag{9.2}$$

which can be interpreted as "squishing" a linear model through a non-linear response function. The denominator normalizes this model, and it is not necessary to estimate one of the weight vectors $\mathbf{w}_j \sim \theta$. The model can easily include an intercept term w_0, i.e. $\langle \mathbf{w}_j, \mathbf{x} \rangle + w_0 = \langle [w_0, \mathbf{w}_j], [1, \mathbf{x}] \rangle$. For now on, the notation assumes the intercept to be included for each class.

In general, the parametric functional form of the logistic regression can be anything which obeys [38]

$$0 \leq t_j(\mathbf{x}, \theta) \leq 1, \ \sum_{j=1}^{|\mathcal{C}|} t_j(\mathbf{x}, \theta) = 1, \forall \mathbf{x}, \forall \theta. \tag{9.3}$$

One way to derive the model in equation 9.2 is by setting

$$\frac{P(j|\mathbf{x})}{P(|\mathcal{C}||\mathbf{x})} = \exp(\langle \mathbf{w}_j, \mathbf{x} \rangle), \ \ j = 1, \ldots, |\mathcal{C}| - 1, \tag{9.4}$$

then writing out with all possible values of j and solving for the wanted posterior. It can be also easily shown to obey conditions of Equation 9.3.

Equation 9.5 is valid when the class priors are implicitly incorporated in the training data class fractions $f_j \sim P_j$. However, when using training data with balanced, equal fractions for each class, one needs to be explicit with priors in the classification phase and use Bayes' theorem to correct the posterior probabilities as

$$P(C = j|\mathbf{X} = \mathbf{x}; \mathbf{w}) = \frac{\exp(\langle \mathbf{w}_j, \mathbf{x} \rangle) P_j}{\sum_{i=1}^{|\mathcal{C}|} \exp(\langle \mathbf{w}_i, \mathbf{x} \rangle) P_i}, \tag{9.5}$$

which was the approach used in this work, and also suggested in [7].

The parameter (weight) vectors $\mathbf{w}_j$ Maximum Likelihood (ML) estimates can be calculated using iterative mathematical optimization routines, such as IRLS. Formally, conditional ML estimates are obtained by maximizing concave cost function $l : \mathbb{R}^{d|\mathcal{C}|} \to \mathbb{R}$

$$l(\mathbf{w}) = \sum_{j=1}^{n} \ln \ P(\mathbf{y}_j|\mathbf{x}_j, \mathbf{w}) = \sum_{j=1}^{n} \left(\sum_{i=1}^{|\mathcal{C}|} \mathbf{y}_j^{(i)} \langle \mathbf{w}_i, \mathbf{x}_j \rangle - \ln \ \sum_{i=1}^{|\mathcal{C}|} \exp(\langle \mathbf{w}_i, \mathbf{x}_j \rangle) \right), \tag{9.6}$$

where n is the number of training vector-class label pairs and the binary vector $\mathbf{y}_j \in \{0, 1\}^{|\mathcal{C}|}$ is used to encode the classification of the j-th training vector $\mathbf{x}_j$.

In practise, the problem must be *regularized* by inducing a prior distribution over parameter vectors. This prior is used to avoid *overfitting* and having degenerate solutions to the optimization problem, and makes parameter estimates Bayesian Maximum a Posteriori (MAP). Degenerate solutions can arise when classes are linearly separable and $l(\mathbf{w}) \to \infty$. The problem with regularization is in a functional form

$$\hat{\mathbf{w}}_{MAP} = \arg\max_{\mathbf{w}} L(w) = \arg\max_{\mathbf{w}} \left[l(\mathbf{w}) + \log p(\mathbf{w}) \right], \tag{9.7}$$

with regularization distribution denoted by $p(\mathbf{w})$.

Classic regularization is done by using a Gaussian prior distribution which is usually known as *Tikhonov regularization* or ℓ_2-norm regularization. This regularization works well when induction of smoothness is wanted into solution. Another way is to induce sparsity into solution. This is done by using ℓ_1-norm[1] regularization or a corresponding Laplace distribution as the prior

$$p(\mathbf{w}) \propto \exp(-\lambda \|\mathbf{w}\|_{\ell_1}), \tag{9.8}$$

where $\lambda \in \mathbb{R}_+$ is the regularization strength parameter.

The problem with the Laplace distribution is, that it creates an optimization problem which cannot be calculated e.g. with IRLS (iteratively reweighted least squares) due to non-differentiability of the ℓ_1-norm at $\mathbf{x} = 0$, but it is still a convex[2] minimization problem with unique, global optimum guarantee. [39]

Formally, a function $f : \mathbb{R}^n \to \mathbb{R}$ defined as

$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta) f(y), \tag{9.9}$$

is convex if its domain $\mathcal{D}(f)$ is a convex set for all $x, y \in \mathcal{D}(f)$ with $\theta \in [0, 1]$. The definition of a convex set C with $x, y \in C$ is that $\theta x + (1-\theta)y \in C$, i.e. all points on the line between any x and y which belong to C, belongs also to C. It is well-known that e.g. Neural Networks are not convex optimization problems, and thus, they have several local minima solutions.

[1] $\|\mathbf{x}\|_{\ell_1} = \sum_{i=1}^{d} |x_i|, \ \ \mathbf{x} \in \mathbb{R}^d$

[2] If f is a concave function, $-f$ is called a convex function.

9.2 Optimization algorithm

The use of ℓ_1-norm requires a state-of-the-art optimization approach, and in this work, a fast algorithm from [39] was implemented. Now a simple description of the algorithm follows. Notation assumes that the weight vector $\mathbf{w}$, without any lower indices, encapsulates weight vectors for each class, i.e. $\dim(\mathbf{w}) = d(|\mathcal{C}| - 1)$. Except for one, which does need not to be estimated due to normalization of probabilities.

First start with the ML optimization problem by defining

$$p_j^{(i)} = P(y_j^{(i)} = 1|\mathbf{x}_j, \mathbf{w}), \tag{9.10}$$

where lower index j denotes now the j-th training vector out of n and upper index i denotes the class, with the corresponding vector and the diagonal matrix representations

$$\mathbf{p}_j(\mathbf{w}) = [p_j^{(1)}(\mathbf{w}), \ldots, p_j^{(|\mathcal{C}|-1)}(\mathbf{w})]^T \in \mathbb{R}^{|\mathcal{C}|-1} \tag{9.11}$$

$$\mathbf{P}_j(\mathbf{w}) = \text{diag } \left(p_j^{(1)}(\mathbf{w}), \ldots, p_j^{(|\mathcal{C}|-1)}(\mathbf{w})\right) \in \mathbb{R}^{(|\mathcal{C}|-1)\times(|\mathcal{C}|-1)}. \tag{9.12}$$

The maximum likelihood cost function $l(\mathbf{w})$ is a concave function and its Hessian square matrix of size $(|\mathcal{C}| - 1)d$ is

$$\mathbf{H}(\mathbf{w}) = -\sum_{j=1}^{n} \left(\mathbf{P}_j(\mathbf{w}) - \mathbf{p}_j(\mathbf{w})\mathbf{p}_j^T(\mathbf{w})\right) \otimes \left(\mathbf{x}_j\mathbf{x}_j^T\right), \tag{9.13}$$

expressed using the Kronecker matrix product operator $\otimes$, a special case of tensor product. This is defined as $\mathbf{Q} \otimes \mathbf{W} \in \mathbb{R}^{mp\times nq}$, if $\mathbf{Q} \in \mathbb{R}^{m\times n}$ and $\mathbf{W} \in \mathbb{R}^{p\times q}$.

This Hessian is lower bounded [39] by a negative definite matrix $\mathbf{B}$ which has no dependency on $\mathbf{w}$. This is written as

$$\mathbf{H}(\mathbf{w}) \succeq -\frac{1}{2}[\mathbf{I} - \frac{\mathbf{1}\mathbf{1}^T}{|\mathcal{C}|}] \otimes \sum_{j=1}^{n} \mathbf{x}_j\mathbf{x}_j^T =: \mathbf{B}, \tag{9.14}$$

where $\mathbf{1}$ is a vector of ones with $\dim(\mathbf{1}) = |\mathcal{C}| - 1$. Together with the Hessian, we need the gradient of $l(\mathbf{w})$, given by

$$\nabla l(\mathbf{w}) = \sum_{j=1}^{n} \left(\mathbf{y}'_j - \mathbf{p}_j(\mathbf{w})\right) \otimes \mathbf{x}_j, \tag{9.15}$$

with the target class encoding vector $\mathbf{y}'_j = [y_j^{(1)}, \ldots, y_j^{(|\mathcal{C}|-1)}]^T$.

Finally, the iterative update rule for ML estimates without any regularization is

$$\hat{\mathbf{w}}^{(t+1)} = \hat{\mathbf{w}}^{(t)} - \mathbf{B}^{-1}\nabla l(\hat{\mathbf{w}}^{(t)}), \tag{9.16}$$

where $t \in \mathbb{N}$ is the iteration counter index.

Regularized case

The optimization rule of the ℓ_1-regularized cost function, that is the MAP case, is given by maximizing [39]

$$\mathbf{w}^T \left(\nabla(l(\hat{\mathbf{w}}^{(t)}) - \mathbf{B}\hat{\mathbf{w}}^{(t)}\right) + \frac{1}{2}\mathbf{w}^T(\mathbf{B} - \lambda\Lambda^{(t)})\mathbf{w}, \tag{9.17}$$

where the matrix $\Lambda^{(t)} = \text{diag}\ \left(|\hat{w}_1^{(t)}|^{-1}, \ldots, |\hat{w}_{d(|\mathcal{C}|-1)}^{(t)}|^{-1}\right)$.

The actual optimization step of the sparsity promoting MAP ℓ_1 optimization algorithm is then expressed as

$$\hat{\mathbf{w}}^{(t+1)} = \left(\mathbf{B} - \lambda\Lambda^{(t)}\right)^{-1}\left(\mathbf{B}\hat{\mathbf{w}}^{(t)} - \nabla l(\hat{\mathbf{w}}^{(t)})\right), \tag{9.18}$$

where the matrix $\mathbf{B}$ must be computed only once. This is the most straightforward way to maximize function 9.17, but in practise numerically more optimal approaches are available. For those, see [39]. Smoothness promoting ℓ_2 algorithm can be easily obtained by changing the $\Lambda^{(t)} \mapsto \mathbf{I}$ [39], which is a little surprising, given the technicality needed for the ℓ_1-norm in the first place. The initialization of $\hat{\mathbf{w}}^{(0)}$ is not very important, due to global optimum guarantee of this problem.

9.3 Adjusting regularization

In Figure 9.1, the regularization parameter λ-paths are calculated using the MC training sample. When the regularization parameter $\lambda \to \infty$, the most important variables (vector coefficients) are remained and the rest are suppressed to zero. Mathematically speaking, regularization with ℓ_1 norm applies a Laplacian distribution prior over coefficients of $\mathbf{w}_j$, which induces sparsity into solution and has an effect of variable selection.

The physical interpretation is clear, e.g. with single diffraction the most important detector components are those with high $|\eta|$. This can be seen as *a large rapidity gap - multivariate space correspondence*, in the limit where only a subspace $S \subset \mathbb{R}^d$ of the multivariate space $\mathbb{R}^d$ is used with $\dim(S) \ll d$. Classification metrics in Figures 9.2 and 9.3 show that using only the subspace information, i.e. classifying events based on large rapidity gaps, is suboptimal in terms of classification performance. Information theoretically it means that the entropy of diffraction is distributed over all components, even if the most of the entropy belongs to components corresponding to high $|\eta|$. All this is consistent with the rapidity gap survival probability discussion, where pure gaps of diffraction are destroyed by some spectator partons.

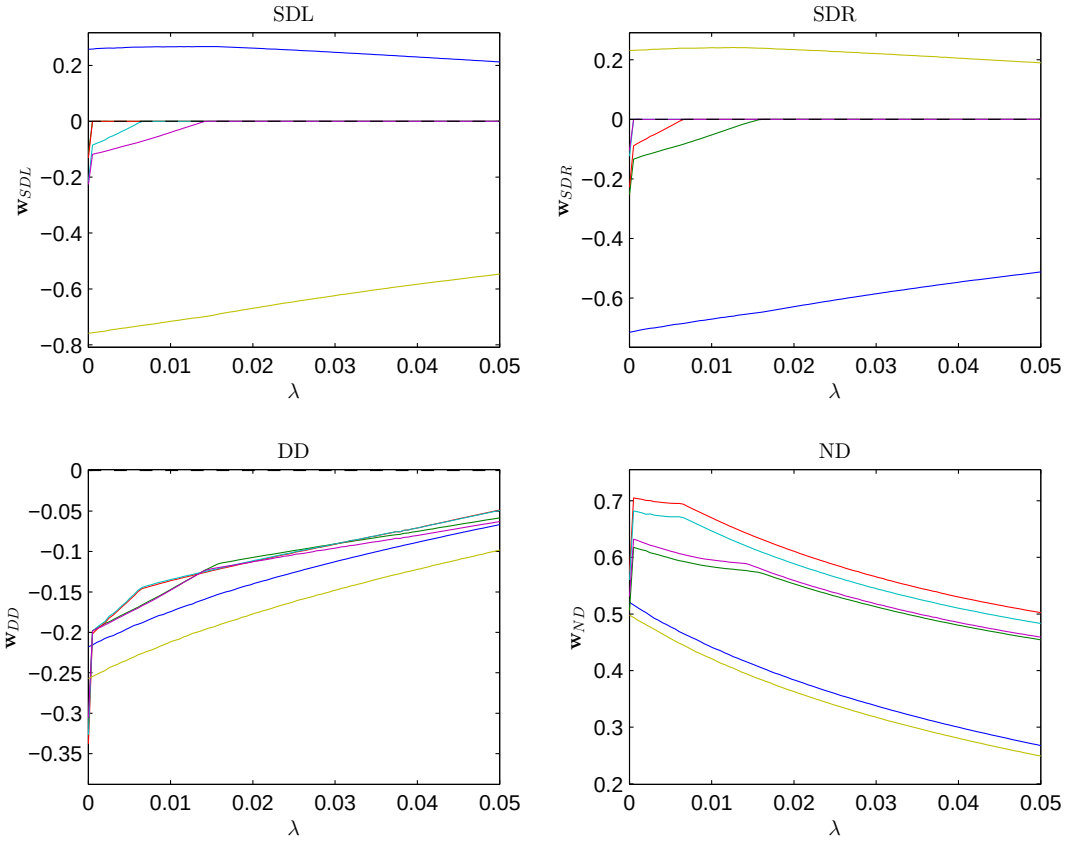


Figure 9.1: Regularization λ-paths using MLR-ℓ_1 with the MC training sample. On y-axis the coefficients of $\mathbf{w}_j$ in order: $w_i :=$ (blue, green, red, light blue, purple, yellow), with discrete binning $\mathbf{d}_\eta = (-3.64, -1.78, -0.88, 0, 0.88, 1.78, 3.64)$, such that $\eta_{\min,\max}(w_i) \in [d_i, d_{i+1}]$. All 6 variables are log transformed, centred (0-mean) and variance normalized calorimeter deposits.

The interpretation of the coefficients $\{w_i\}_{i=1}^{d}$ of the weight vector $\mathbf{w}_j$ is relatively easy, due to ℓ_2 inner product $\langle \mathbf{w}_j, \mathbf{x} \rangle$ which is evaluated between the weight vector and the measurement vector $\mathbf{x}$ in logistic Equation 9.5. This inner product operation is a correlator, and a positive coefficient $w_i > 0$ with large $x_i > 0$ means strong correlation with the corresponding j-th scattering class and e.g. a negative coefficient $w_i < 0$ but $x_i > 0$ is understood as anti-correlation. This interpretation makes this multivariate algorithm more transparent.

When looking at Figure 9.1, there is expected, natural antisymmetry between the coefficients of left and right single diffraction. There is also slight discrepancy between the corresponding antisymmetric components of single and right diffraction, and between the high $|\eta|$ components of double diffraction. This is probably reflecting the differences in the detector modelling and characterization in the Geant model, i.e. detectors in backward and forward regions are not physically perfectly identical. The signs of the coefficients are dominated by non-diffractive events, which have the largest deposit energies in the given η range and thus its the coefficients are all positive.

9.4 Classification performance

Effect of the regularization parameter λ on the classifier performance must be evaluated numerically. This can be done using cross-validation, by doing a scan over different λ values and inspecting the classification metrics. This is done in Figure 9.2, where efficiency, purity, purity $\times$ efficiency and total accuracy are calculated for each class using a MC test sample. Clearly, one sees that at least with a linear hyperplane classifier, as MLR-ℓ_1 is, the classification efficiencies of double and single diffraction are anti-correlated with non-diffractive class. Which means that the efficiency of non-diffractive class rises when the efficiency of diffractive classes go down, and vice versa with the purities. This is one of the good points of a linear classifier over non-linear ones, it has predictable and stable behaviour.

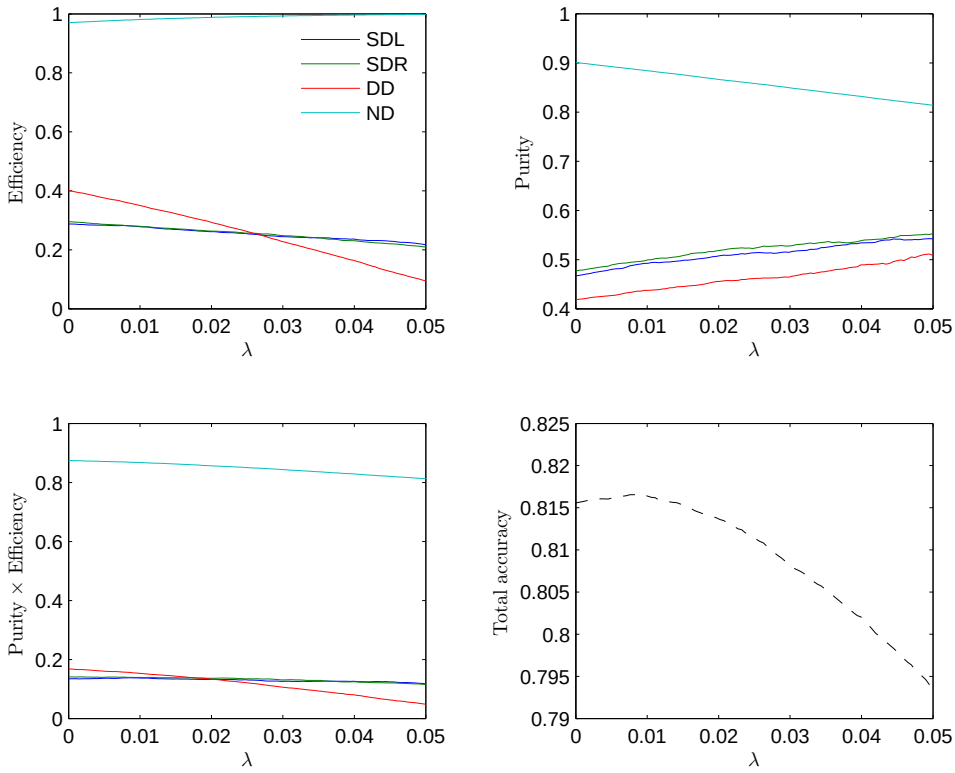


Figure 9.2: Hard (Bayes' minimum error rule) event-by-event classification metrics obtained using MLR-ℓ_1 with the MC test sample.

Clearly, the regularization can be seen here as a mathematical tool for adjusting the sensitivity of classification toward different event classes. Usually that is done by adjusting the cut value on the output of the 2-class classifier (like neural net) and by plotting signal efficiency vs. background rejection ROC-curve. This approach is much more complicated for the multi-class classification case, but MLR-ℓ_1 posterior probability outputs allow this kind of optimization when used as a hard classifier. Choosing optimal working point(s) should be always based on the nature of the analysis, i.e. are we searching for some new physics signal over known background etc.

So, there is always a trade-off between purity and efficiency of individual classes. However, the total accuracy can be seen as the optimal criteria in this application, because it minimizes the empirical Bayes error rate. The optimal total classification

accuracy is achieved with regularization around $\lambda \approx 0.003$. The biggest difference between hard and probabilistic classification can be seen from Figure 9.3, where the relative differences between estimated and ground truth class cross sections are evaluated. The minimum error point of the probabilistic classification (sum of absolute relative errors ~ 0.01) corresponds to the maximum total accuracy plateau in Figure 9.2.

This is an important result. First of all, it gives numerical backup for the analytical derivation of the integrated probabilities done in Section 8.2. Secondly, it shows that the optimal probabilistic integrated cross section estimates coincide with the optimal total hard classification accuracy point, obtained using Bayes' minimum error classifier rule in equation 8.12. Bayes' minimum error rule assumes always that unbiased priors $P_j \sim \sigma_j/\sigma_{inel}$ are being used, and the probabilistic approach does no exception.

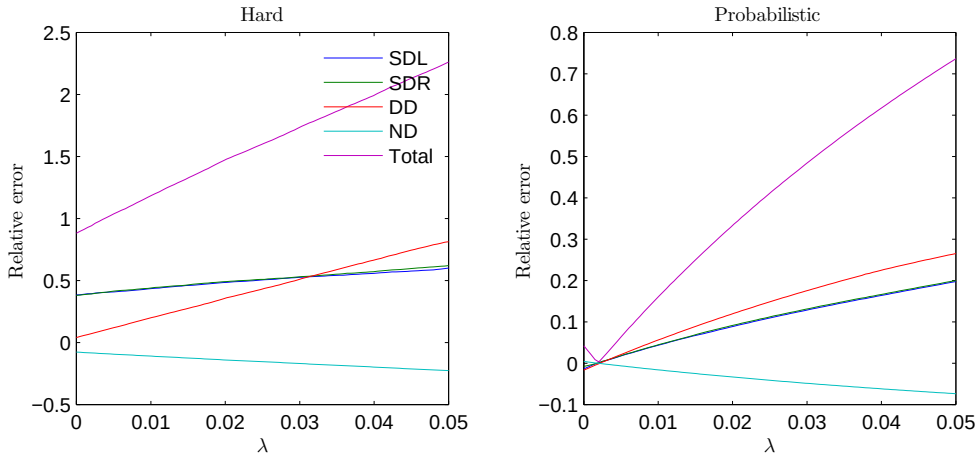


Figure 9.3: The relative errors of hard (Bayes' minimum error rule) and probabilistic MLR-ℓ_1 classification, when estimating class fractions $\sim$ cross sections of the MC test sample.

Confusion matrix estimates

The amount of overlap of class densities in the multivariate space can be estimated as follows. First, calculate the confusion matrix of a test sample which has uniform class fractions $f_i = f_j \ \forall i, j$. Then do a realistic sample of class vectors, and use class fractions according to theoretical relative cross sections $f_j \sim \sigma_j/\sigma_{inel}$. Then compare the confusion matrix of this sample to the first. This can be done with Tables 9.1 and 9.2, where in both cases the test data was nearly identical but only the fractions were different.

The classification efficiencies transform completely when the class fractions (and priors) are different. When realistic fractions are being used, the efficiencies of single diffraction go down and the efficiency of double diffraction rises. However, the purities stay nearly the same, thus the flip-flop of efficiencies happens between $SD \leftrightarrow DD$, which can be read from the confusion matrix elements. It seems that the signature of diffraction with given η-span of CDF is highly ill-posed, and the priors have a defining role in classification results.

Table 9.1: Hard confusion matrix estimate for the case with uniform priors $P_j = 1/|\mathcal{C}| \; \forall j$ with uniform test sample fractions $f_j \sim P_j$. Rows are true classes, columns are estimates and *Acc* denotes total accuracy.

	SDL	SDR	DD	ND	ϵ_j
SDL	0.62	0.13	0.15	0.09	0.62
SDR	0.13	0.62	0.15	0.09	0.62
DD	0.38	0.38	0.20	0.05	0.20
ND	0.08	0.08	0.02	0.82	0.82
π_j	0.52	0.52	0.38	0.78	*Acc* 0.57

Table 9.2: Hard confusion matrix estimate for the case with realistic priors $P_j = (0.07, 0.07, 0.10, 0.76)$, which are based on the product of theoretical cross sections and 0-bias cut efficiencies in table 10.3 and test sample has fractions $f_j \sim P_j$.

	SDL	SDR	DD	ND	ϵ_j
SDL	0.24	0.02	0.35	0.39	0.24
SDR	0.02	0.23	0.37	0.39	0.23
DD	0.13	0.13	0.43	0.31	0.43
ND	0.00	0.00	0.02	0.98	0.98
π_j	0.48	0.47	0.41	0.90	*Acc* 0.82

Hard classification efficiencies are anyway much higher than for a pure random classifier $\sim P_j$, which is the sanity check lower bound for accuracy. Also, non-linear multilayer perceptron (MLP) neural networks with different configurations and non-linear boosted decision trees (BDT) were tried out as a comparison, and they did not achieve any higher hard classification results than the MLR-ℓ_1. Also, they do *not* give probabilistic output without major modifications, like output histogram techniques etc. See Appendix A.5 for a Monte Carlo consistency comparison between MLR-ℓ_1 and BDT, obtained by classifying a different MC sample than what was used in training.

In addition, non-linear stochastic Gaussian processes (GP) [46] based classifiers with probabilistic output were studied and experimented. Unfortunately, the classification results with GP classifiers were not any better than with linear MLR-ℓ_1 in this problem, and they were computationally extremely heavy. Also, a non-linear modification of MLR was tried out, by using a *kernelized* [7] version of it. Also this trial did not prove to be useful. However, with a different physical detector setup, the case might be different.

These results reflect the fact that the class densities are not uniquely separable, given the physical η-span of the detectors. Most importantly, most of the other classifiers do not enable to provide as clear physical interpretation such as explained with MLR-ℓ_1 in section 9.3.

Chapter 10

Results with a CDF 0-bias sample

The primary target of the experimental part of this study was to obtain cross sections for each main inelastic scattering process defined here

$$\sigma_{inel} \triangleq \sigma_{SDL} + \sigma_{SDR} + \sigma_{DD} + \sigma_{ND}. \tag{10.1}$$

The central diffraction was excluded from the study because it was not available in the used Monte Carlo generator. Also, theoretically it has such a small, under one mb cross section that it should have a minor effect to integrated cross sections estimates.

Table 10.1: CDF 0-bias sample parameters.

Parameter	Details
CMS energy	$\sqrt{s} = 1.96$ TeV
Bunches	$N_B = 36$
Bunch revolution frequency	$f_0 = 46500$ Hz
Inst. luminosity over N_B bunches	$\langle \mathcal{L}_{inst} \rangle = 27.2 \cdot 10^{30}$ cm^{-2}s^{-1}

The accelerator parameters regarding the 0-bias sample collected at CDF in are given in Table 10.1.

10.1 Analysis parameters

The data sample used in the analysis was a CDF 0-bias $\sqrt{s} = 1.96$ TeV run triggered by a "random trigger", which triggered and collected the data based on the bunch crossing information. The variables used in the multivariate analysis were calorimeter deposits and charged particle multiplicities obtained from tracking, both summed over $\phi \in [0, 2\pi]$. These variables and their number are given in Table 10.2.

Table 10.2: Variables used in multivariate analysis.

Variable	#	Pseudorapidity η
Calorimeters	22	$[-3.64, 3.64]$
Tracks	19	$[-2.11, 2.11]$

Different binning over η were studied, and more discrete binning, i.e. higher dimensional multivariate space, gave always better results with the MLR-ℓ_1 algorithm. So no problem with a high dimensional multivariate space was experienced. This is probably due to fact that there were always enough training data used, around $\sim$ 50000 vectors for each class. Logarithmic pre-processing was used with the analysis classifier, because based on Monte Carlo, the classification efficiencies of diffractive classes were improved.

Table 10.3: Analysis parameters.

Parameter	Details (SDL, SDR, DD, ND)
Inelastic [16]	$\sigma_{inel} = 61.06$ mb
Priors (PYTHIA 6) [48]	$P_j = \sigma_j/\sigma_{inel} = (0.10, 0.10, 0.12, 0.68)$
Mean pile-up	$\nu = \langle \mathcal{L}_{inst} \rangle \sigma_{inel}/(N_B f_0) = 0.98$
Pre-selection cut	$E_c = 2.2$ GeV, $N_c = 1$
Pre-cut efficiencies	$\epsilon_j = (0.63, 0.63, 0.77, 1.00)$
CDF data sample size	$\lvert\mathcal{S}\rvert = 1563562$ (#)
After f_{cut}	809099 (#)
After pile-up discriminator	575482 (#)
Pile-up discriminator:	
Variable normalization	Mean-Variance
MLR-ℓ_1	Regularization $\lambda = 0.0025$
Analysis classifier:	
Variable normalization	Logarithmic ($\epsilon = 1$) and Mean-Variance
MLR-ℓ_1	Regularization $\lambda = 0.0025$

The parameters regarding the whole classification chain are given in Table 10.3. Note that the priors $\sim$ relative cross sections are before taking into account any cuts before the classification algorithm. These cuts alter the bare priors and this effect is explained in the following sections. The full analysis algorithm without pile-up filtering is explained in what follows.

Analysis algorithm: no pile-up case

INPUT: MC $\mathcal{T}$, Data $\mathcal{S}$, Priors $\{P_j\}$, Luminosity $\int \mathcal{L}\, dt$ or Inelastic σ_{inel}

A. Pre-cut ($\sim$ trigger) efficiency estimation
1. Estimate the 0-bias cut efficiencies $\{\epsilon_j\}$ based on applying the cut on MC samples $\tau_j \subset \mathcal{T}$ of the each j-th scattering process class separately and calculate efficiencies with $\epsilon_j = |\{\mathbf{x} \in \tau_j : f_{cut}(\mathbf{x}) = 1\}| \,/\, |\tau_j|$.

B. Classifier training
1. Normalize the training vectors using the mean-variance normalization, where means and variances are based on $\mathcal{T}$ used to train the classifier.
2. Train the MLR-ℓ_1 classifier with uniform number of training vectors for each class, i.e $|\tau_i| = |\tau_j|\ \forall ij$, to avoid biasing the classifier[1]

C. Estimating probabilities
1. Apply the 0-bias cut f_{cut} to the data sample $\mathcal{S}$.
2. Normalize the remained data vectors using means and variances calculated in the training phase.
3. Obtain the posteriori probabilities with the MLR-ℓ_1 using priors which have been re-weighted as $P_j \mapsto P_j \cdot \epsilon_j$, and normalized to sum up to one.
4. Calculate the mean values $\langle P(C = j|\mathbf{X})\rangle$ of a posteriori probabilities for each j-th class.

I Obtaining cross sections (using integrated luminosity)
1. Obtain the cross sections for each j-th class by

$$\sigma_j = \frac{\langle P(C = j|\mathbf{X})\rangle N_{tot} - \langle B_j\rangle}{\epsilon_j \int \mathcal{L}\, dt}, \tag{10.2}$$

where N_{tot} is the number of events in the data sample after 0-bias cut and the obtained (visible) event counts are extrapolated via efficiencies ϵ_j. The background events B_j are possibly propagating from the "non-interaction" event set, but not a priori from other classes, due to probabilistic event-by-event weighting.

II Obtaining cross sections (using inelastic cross section)
1. Obtain the cross section for each j-th class by

$$\sigma_j = \frac{N_j/\epsilon_j}{\sum_k^{|\mathcal{C}|} N_k/\epsilon_k}\sigma_{inel}, \tag{10.3}$$

[1] In general, this is a classifier dependent procedure and it depends on how the classifier incorporates prior probabilities. If priors are explicit, then training must be done with uniform samples, to avoid "double priors".

where $N_j = \langle P(C = j|\mathbf{X})\rangle N_{tot} - \langle B_j \rangle$, and the variables N_{tot}, ϵ_j and B_j are the same as above. This algorithm II is formally an equivalent formulation as the algorithm I, which is easy to proof by using Definition 10.1 and Equation 10.2.

OUTPUT: Cross sections $\{\sigma_j\}$

10.2 Selection of interaction events

Because the 0-bias sample was bunch crossing triggered and low luminosity, nearly half of the events are "empty" or "non-interaction". Actually, the elastic and low-mass diffractive interactions are part of this "non-interaction" subset, but their cross sections cannot be measured here experimentally due to missing very forward instrumentation.

To discriminate out the "non-interaction" events, a selection cut $f_{cut} : \mathbb{R}^n \rightarrow \{0, 1\}$ was used

$$f_{cut} = \begin{cases} 1, \text{ if } \sum_i E_i \geq E_c \ \vee \ \sum_i N_i \geq N_c \ \text{ (Interaction event)} \\ 0, \text{ otherwise } \text{ ("Non-interaction" event)} \end{cases} \tag{10.4}$$

where E_c is the energy threshold and N_c is the track multiplicity threshold in Table 10.3, and E_i is the i-th calorimeter variable and N_i is the i-th track count variable, over η. See section 10.5 for how these thresholds were chosen. This mapping is actually a multivariate projector with two linear operators (sums) combined with algebraic OR.

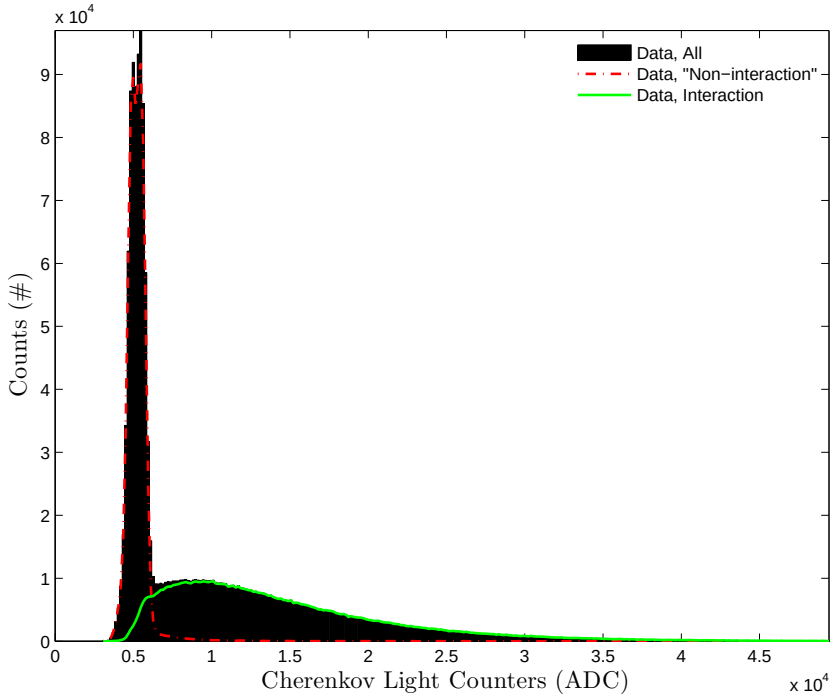


Figure 10.1: Cherenkov light counter distributions.

The calorimeter noise is filtered once before the cut rule, due to fact that every variable E_i is a sum of calorimeter towers over ϕ angle, i.e. over the detector cylinder. This summation over ϕ is a linear low-pass filter, with a Fourier spectrum interpretation, and the optimal linear filter when noise amplitude is Gaussian distributed.

As a cross-check, in Figure 10.1 are the distributions of Cherenkov Light Counters (CLC) for the full 0-bias event sample, interaction and "non-interaction" event

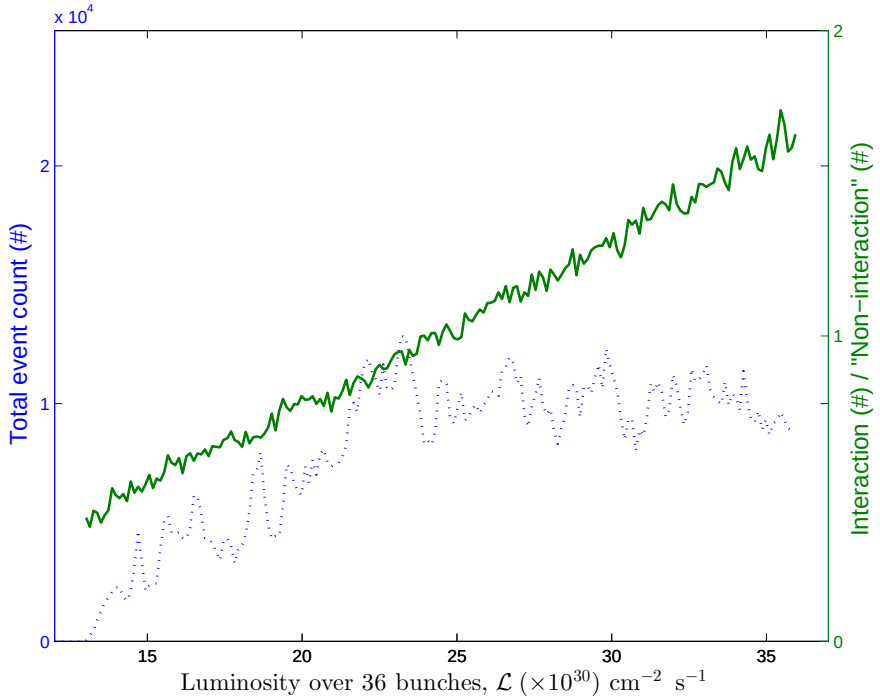


Figure 10.2: Instantaneous luminosity (dashed blue) and the ratio between interaction and "non-interaction" events (green).

samples. Clearly the "non-interaction" sample is Gaussian distributed peak, which indicates that the filtering rule 10.4 is working as designed. The CLC counters itself were not used in the selection rule, because they were not available in the MC simulation and thus the cut efficiency estimation would have been problematic. If one would like to make a mathematically sound 1-dimensional discrimination cut based on the CLC histogram, a one possible, well-known criteria in statistics and signal processing, is to minimize the intra-class variance (and maximize the inter-class variance). This cut value can be easily found by brute force search.

The selection efficiencies in Table 10.3 for each class were estimated using Monte Carlo, by applying cut-and-count to a Monte Carlo sample. Clearly the non-diffractive events, usually with much higher measured deposit energies than diffractive events and with several tracks in the central region of the detector, are selected with nearly perfect efficiency. The efficiencies of diffractive events are thus much lower. This is basically due to very forward events, low mass M_x events, which cannot be selected this way without very forward detectors.

The scaling of the discrimination rule f_{cut}, measured as the ratio between interaction and "non-interaction" as a function of measured instantaneous luminosity is shown in Figure 10.2. The rule seems to scale linearly with luminosity, and no bias with respect to that is observed.

After doing the 0-bias pre-cut, the multivariate classification was done. The details of the analysis procedures and results are given in the following sections.

10.3 Pile-up discrimination

Because the data was collected with instantaneous luminosity, which is too high for data to be completely pile-up free, the data sample was hard classified into pile-up (PU) and no pile-up (N-PU) by training an additional MLR-ℓ_1 as a 2-class discriminator. The training was done by simulating the pile-up conditions with Poisson mean ν in Table 10.3, and by summing the event vectors according to their cross sections in Table 10.3. This summation was done after the Geant simulation and event reconstruction, and implicitly it assumes perfect linearity (additivity and homogeneity) of the pile-up. After the simulation, there were two samples for training and testing, pile-up class and no pile-up class. The obtained efficiencies and purities are in Table 10.4.

If the primary vertex reconstruction information was available, it would have been straightforward to use only bunch crossings where at maximum one primary vertex is reconstructed. However, this information was not available for this analysis.

Table 10.4: Pile-up discriminator confusion matrix estimate based on MC.

	PU	N-PU	ϵ_j
PU	0.65	0.35	0.65
N-PU	0.15	0.85	0.85
π_j	0.77	0.76	*Acc* 0.76

The total accuracy of this discriminator using Bayes' minimum error rule $g^\star$ was 76 percent. By testing this discriminator with pile-up simulated Monte Carlo sample and then classifying the pile-up filtered data using the analysis algorithm in section 10.1, the performance was adequate. The integrated cross sections were estimated nearly as well as without pile-up ($\sim$ 5% error). This was reasonably good, giving that in simulation, based on purity from Table 10.4, the pile-up was 24 percent after discrimination. Thus, no additional corrections with respect to pile-up was added into the analysis, only the discrimination was performed.

10.4 Systematic uncertainties

Due to time scale and a proof-of-concept nature of this thesis, not all the systematic uncertainties were evaluated. Only the effect of prior probabilities was calculated. The prior probabilities were perturbed for each class, and then the whole classification chain was run with real data. The results are shown in Appendix A.6. There is a clear dependency over the priors and this reflects the physical facts already discussed. Note the Δ_{RMS} error behaviour between estimated and expected cross sections, there is a minimum solution with a typical quadratic error behaviour around the minimum. However, the minimum error solution can be biased due to any systematic source of error, for example due to pile-up.

Table 10.5: Possible sources of systematic uncertainties.

Source
Pre-cut efficiencies/purity (MC)
Pile-up discrimination (MC/MVA)
Instantaneous luminosity
Class likelihood models and priors (MC)
Posterior estimates (MVA)
Inelastic cross section (Pre-measured)

The possible source of systematic uncertainties are listed in Table 10.5. Most of the uncertainties are coming from Monte Carlo, like the pre-cut efficiencies which were estimated using MC. The pre-cut purity (leakage of non-interaction to interaction) cannot be answered directly by Monte Carlo, and in this work the purity was assumed to be perfect. The pre-measured inelastic cross section, which was used as a normalization instead of integrated luminosity, has also its own uncertainty.

The most important factors are the class likelihoods, because they incorporate the physics embedded in the chosen Monte Carlo generator (PYTHIA), the detector characterizations and corrections (Geant) and algorithmic reconstructions. Class priors are given by the used Monte Carlo model, or they can be also given by some other theory. These two make up the most of the total systematic uncertainty.

Finally, the a posteriori probabilities estimated by the multivariate algorithm also include a source of systematic uncertainty. They are not perfect because that would usually require an infinite amount of training data and high complexity from the algorithm.

Usually all the estimated individual systematic errors would be added in quadrature, by assuming mutual independence and orthogonality, and finally taking the square root to obtain total errors for each class. To take (linear) correlations in account, one would also need to estimate the covariance matrix between error sources.

10.5 A posteriori probabilities

In Figure 10.3 the classifier outputs, a posteriori probabilities, are given in normalized histograms for the 0-bias event sample. If the 0-bias pre-cut E_c was too loose, the pure noise non-interaction events were easily seen from this plot where a Gaussian peak was noticeable at $P(C = \mathrm{SD}|\mathbf{x}) \approx 0.24$ for both single diffraction classes. The pre-selection cut value of E_c was tuned on the threshold so that the Gaussian peaks vanished.

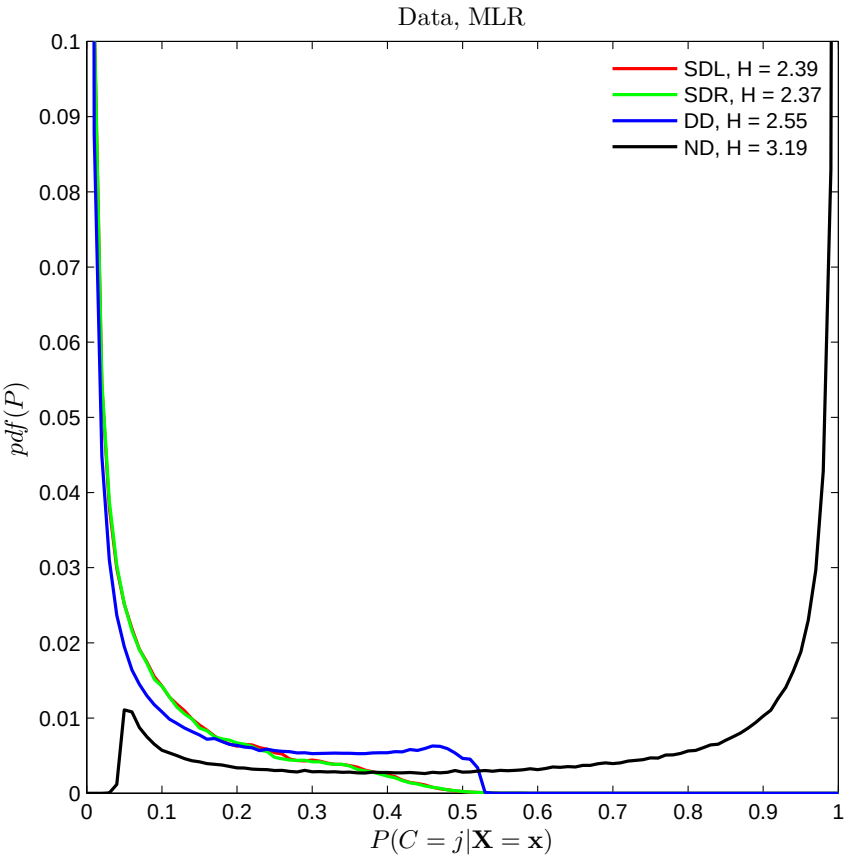


Figure 10.3: Histograms of the a posteriori probabilities for each class.

Quite an interesting fact is that the a posteriori probabilities of the diffractive classes are on the low end of $[0, 1]$, and the spectrum is highly similar for all three diffractive classes. Except the slight rise for the double diffraction after $P \simeq 0.3$. This means that with the given CDF detector η-span, there are really never highly distinctive diffractive event signatures available. Only non-diffractive events can be hard classified on an event-by-event basis into the right class with high confidence.

The information entropy[2] in natural units (nats) is shown in the legend, which was calculated using the binned and normalized histograms. It has an interpretation of being the average information $H(Y) = \mathbb{E}[-\ln P(Y)]$ or uncertainty measure of the random variable Y, embedded in the distribution $P(Y)$. Here, the non-diffractive posteriori distribution has the most entropy. This is physically reasonable, given that the non-diffractive collisions are more rich in phase-space and dynamics than the more restricted diffractive collisions. Theoretically, the maximum entropy is

[2] $H(Y) = -\sum_i P(y_i) \ln P(y_i)$, where $P(Y)$ is the pdf of a discrete random variable Y.

obtained with a uniform density and minimum with Kronecker's delta like distribution.

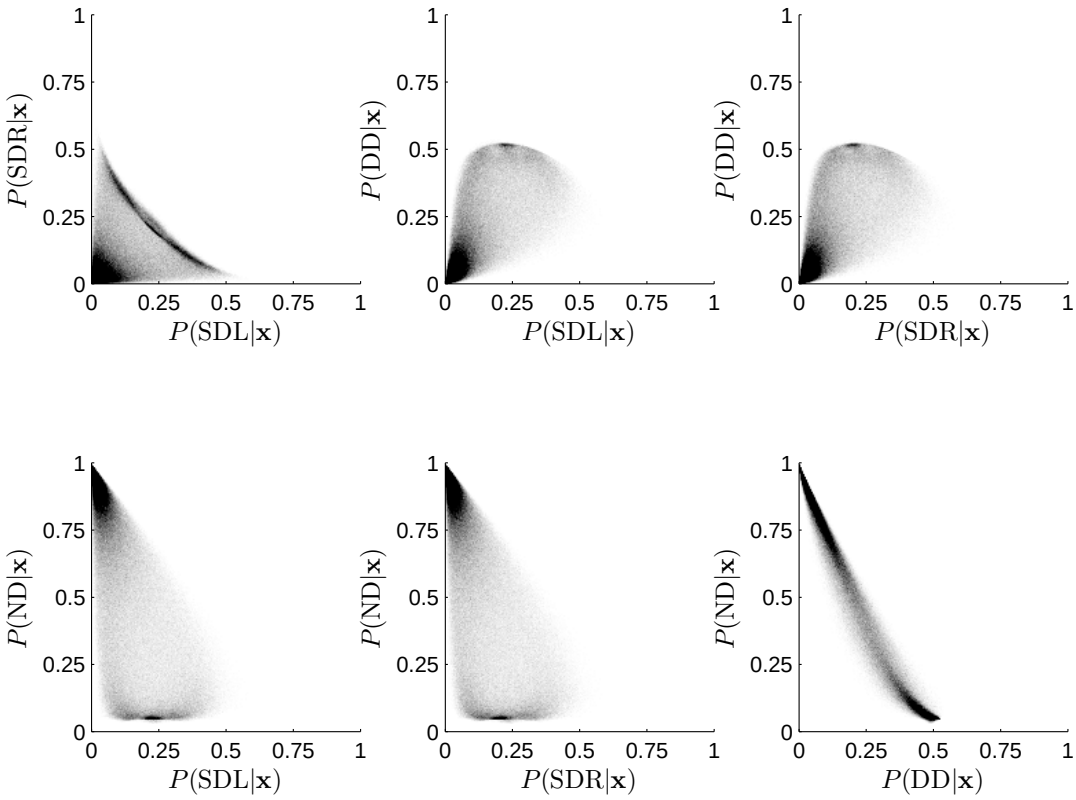


Figure 10.4: Two dimensional distributions of the pairwise a posteriori probabilities, where lighter intensity denotes lower density.

Figure 10.4 shows the pairwise distributions of the a posteriori probabilities, in a Dalitz-plot fashion. Naturally, all distributions have a property to be limited in "probability plane", due to constraint that the probabilities over all classes must sum up to one. The first plot indicates expected symmetry between left and right single diffraction. Also, no asymmetry between single diffractive left and right with respect to double diffractive or non-diffractive class is observed. Perhaps the most interesting distribution is the bottom right, between double diffractive and non-diffractive, which is the most constrained in area and demonstrates linear correlation. Also, each of the distributions are non-uniform, with higher densities on the boundaries. The small peaks near $P(\mathrm{SD}|\mathbf{x}) \approx 0.24$ are most probably detector noise "non-interaction" events, as discussed earlier.

10.6 Inclusive $dE/d\eta$ distributions

The inclusive distributions were calculated using a posteriori probabilities as weights from MLR-ℓ_1 algorithm, and for comparison, using a Boosted Decision Tree (BDT) with AdaBoost training algorithm as a hard classifier without any efficiency/purity corrections, to demonstrate the difference. Boosted Decision Trees are non-linear, widely used[3] out-of-the-box classifiers equipped with very good general performance.

The $dE/d\eta$ distributions are in Figures 10.5, 10.6, 10.7 and 10.8 for classes SDL, SDR, DD and ND, respectively. No major surprises with respect to Monte Carlo are seen, except there is a dip with every class around $|\eta| = 1.25$ in data, which is probably due to defect in detector modelling in Geant simulation. In this analysis, no run specific modelling was available and many different corrections and selections should be applied in order to get correct distributions. In forward regions where $|\eta| > 2$, the data shows ascending trends with ND class respect to the Monte Carlo. Probably PYTHIA's non-diffractive minimum-bias modelling needs tuning in forward regions, or there is some problem again in Geant simulation. Also the central region data seems to have higher energies in both DD and ND with respect to Monte Carlo. It can be mainly due to non-perfect pile-up discrimination and partially due to modelling differences.

When comparing the performance of MLR and BDT, there is similar kind of tendency than in purely simulated results. BDT uses hard cuts, which bias the estimated distributions at high $|\eta|$. These biases should then be compensated based on Monte Carlo or some other way. As expected, the performance of probabilistic MLR is much more consistent.

[3] e.g. in CMS Higgs search.

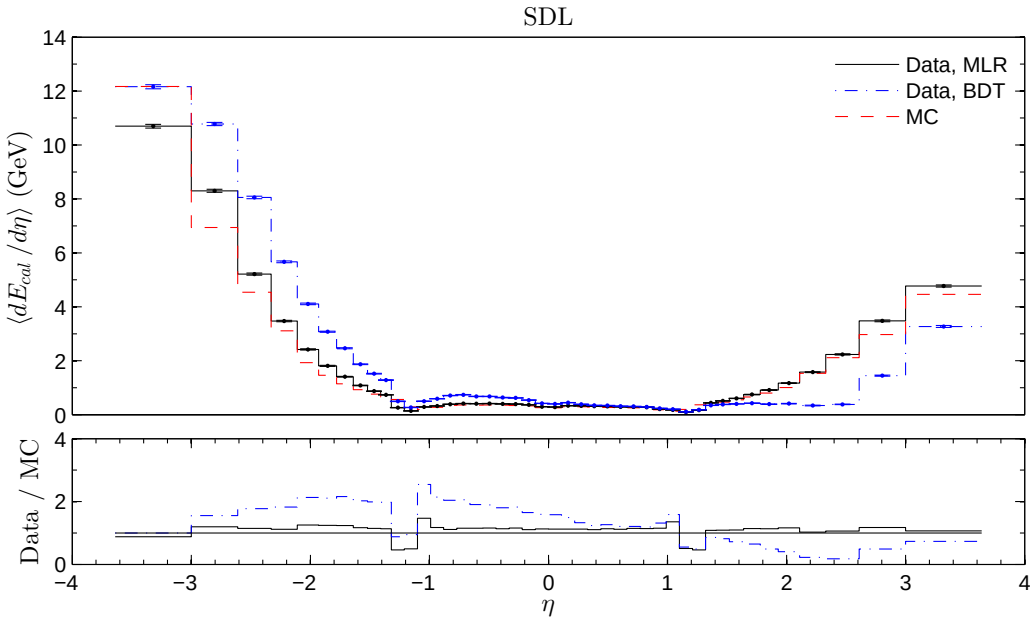


Figure 10.5: Calorimeter deposits E_{cal} as a function of pseudorapidity.

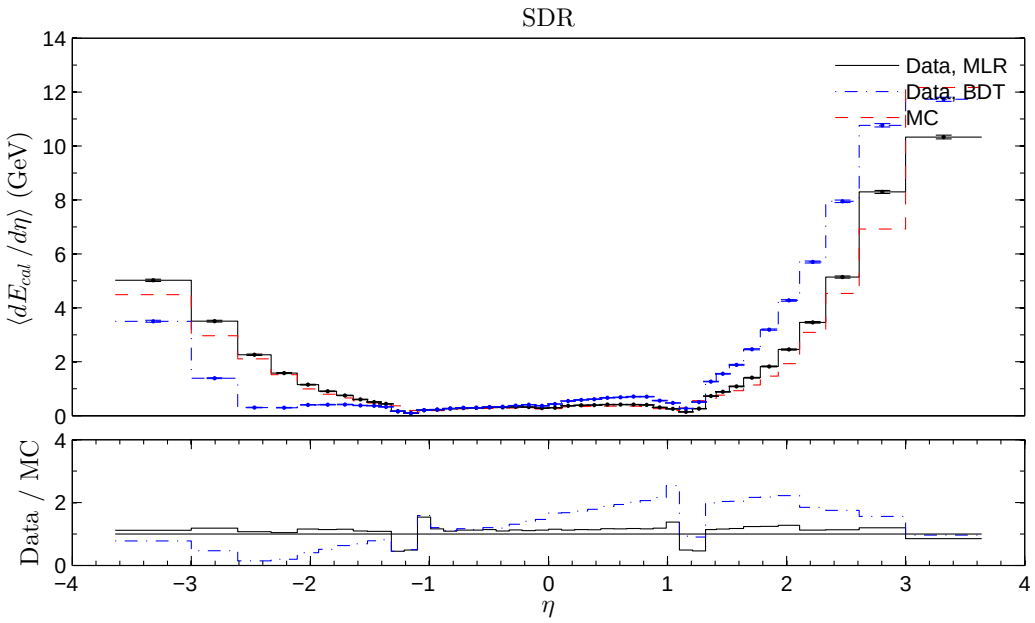


Figure 10.6: Calorimeter deposits E_{cal} as a function of pseudorapidity.

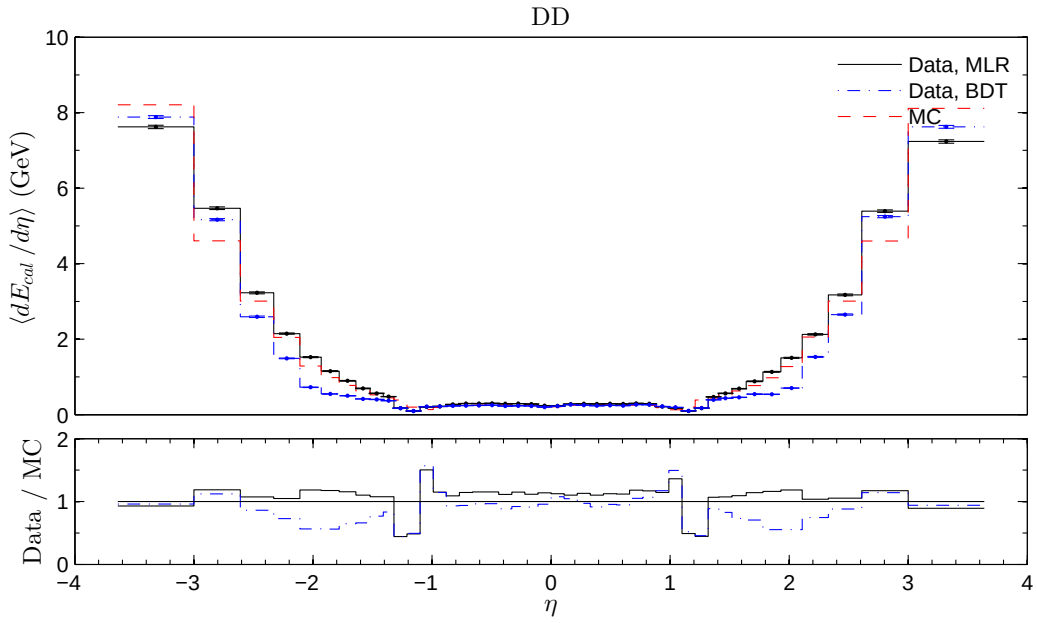


Figure 10.7: Calorimeter deposits E_{cal} as a function of pseudorapidity.

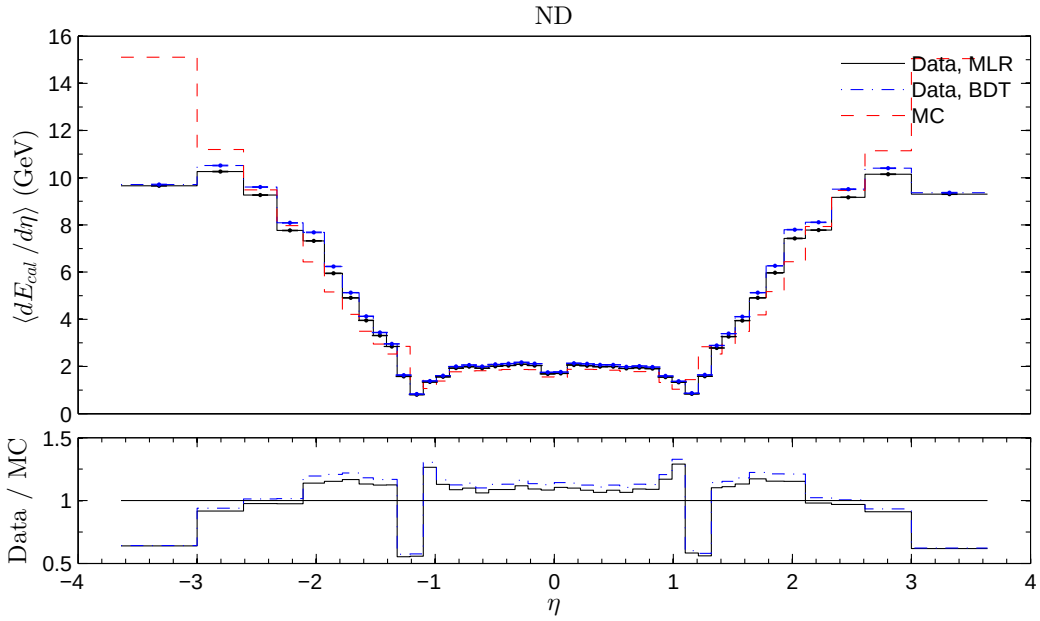


Figure 10.8: Calorimeter deposits E_{cal} as a function of pseudorapidity.

10.7 Calorimeter deposit E_{cal} histograms

The calorimeter deposit histograms are overlaid in Figure 10.9 and each process class is compared with the Monte Carlo predictions in Figure 10.10. Non-diffractive events are on much higher energies, which is expected because most of the diffractive energy is over the rapidity span of the calorimeters.

From Figure 10.9 can be seen that both single diffractive classes are nearly perfectly symmetric, and double diffraction has slightly higher cross section. The distributions for each process class are matching reasonably well the Monte Carlo predictions. Non-diffractive class has slight higher average than MC, which is probably due to non-perfect pile-up discrimination.

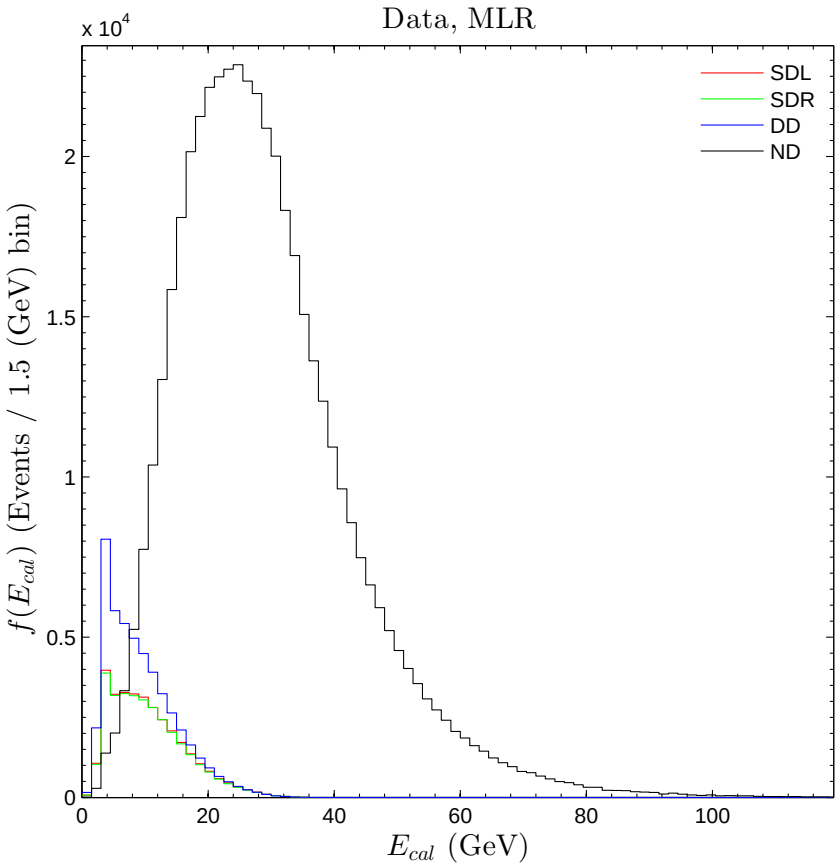


Figure 10.9: Calorimeter deposit histograms.

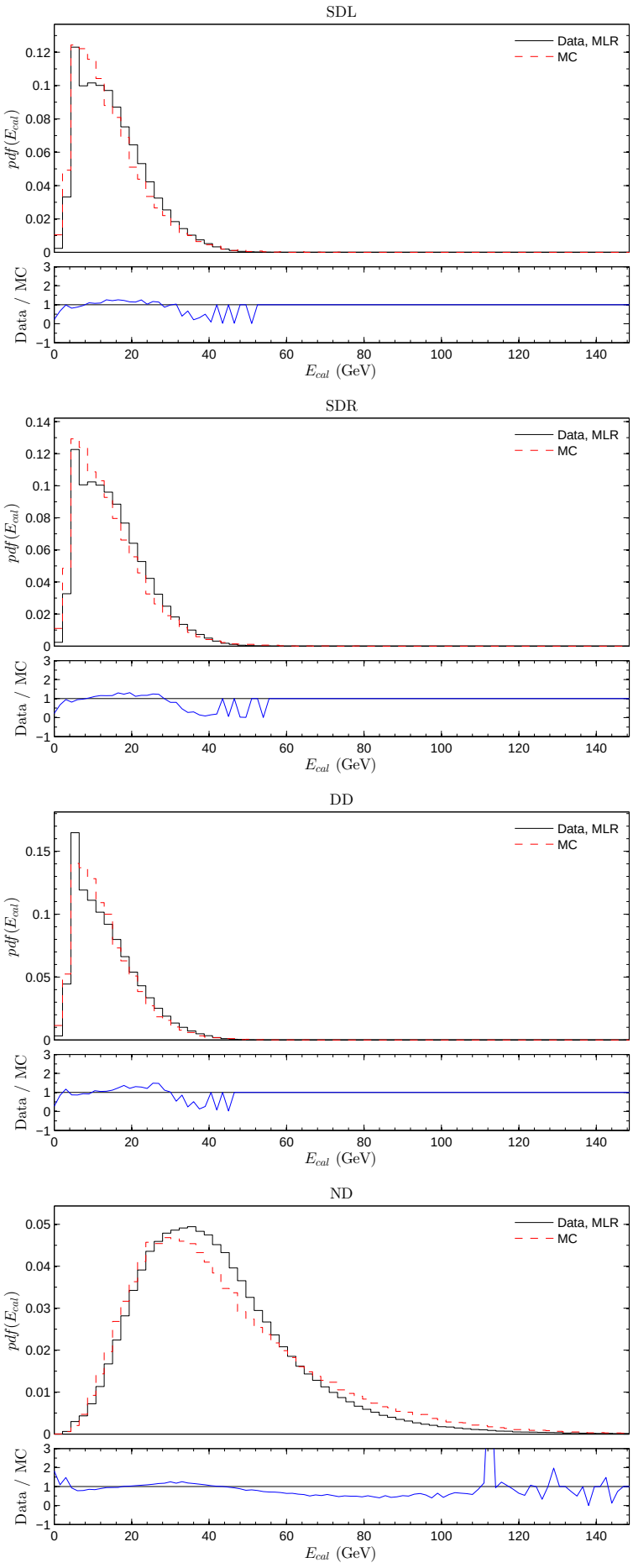


Figure 10.10: Normalized histograms of calorimeter deposits.

10.8 Inclusive $dN_{ch}/d\eta$

The $dN_{ch}/d\eta$ distributions are in Figures 10.11, 10.12, 10.13 and 10.14 for classes SDL, SDR, DD and ND, respectively. Again, no major discrepancies with respect to Monte Carlo are seen. The smooth downward slope as a function of $|\eta|$ is due to decreasing tracking efficiency. Especially the diffractive distributions match well the Monte Carlo predictions. Also, now there is no dip around $|\eta| = 1.25$ as there was with $dE/d\eta$ distributions.

There seems to be a clear left-right asymmetry in MC with ND class, which is not seen as strong in data. The reason for this is not perfectly clear, probably some differences in the tracking system characterization. Because the asymmetry is much more strong in MC and nearly vanishing in data, a detector characterization or calibration mismatch is more likely.

When comparing the performance of MLR and BDT, the results are similar as in $dN/d\eta$. Again, the performance of probabilistic MLR is much more consistent. The distributions of SDL, SDR and DD match quite well the MC expectations, which is a little surprising. The distributions match much better overall than the $dE/d\eta$, probably due to better detector modelling. The characterization of the calorimeter responses is usually a quite demanding task, and in this analysis, out of the box characterizations were used.

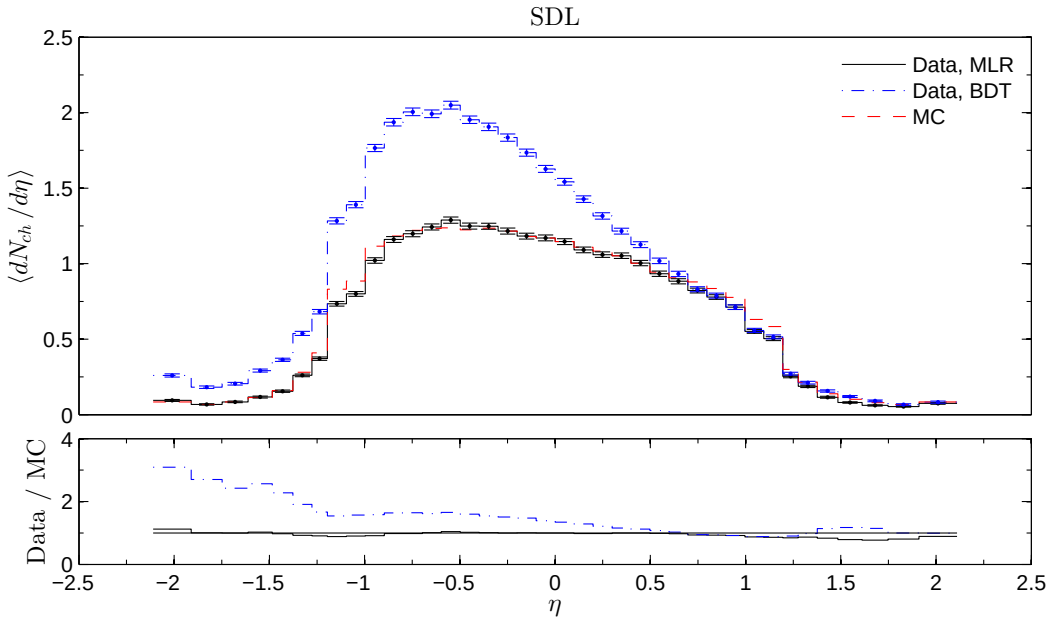


Figure 10.11: Track multiplicities N_{ch} as a function of pseudorapidity.

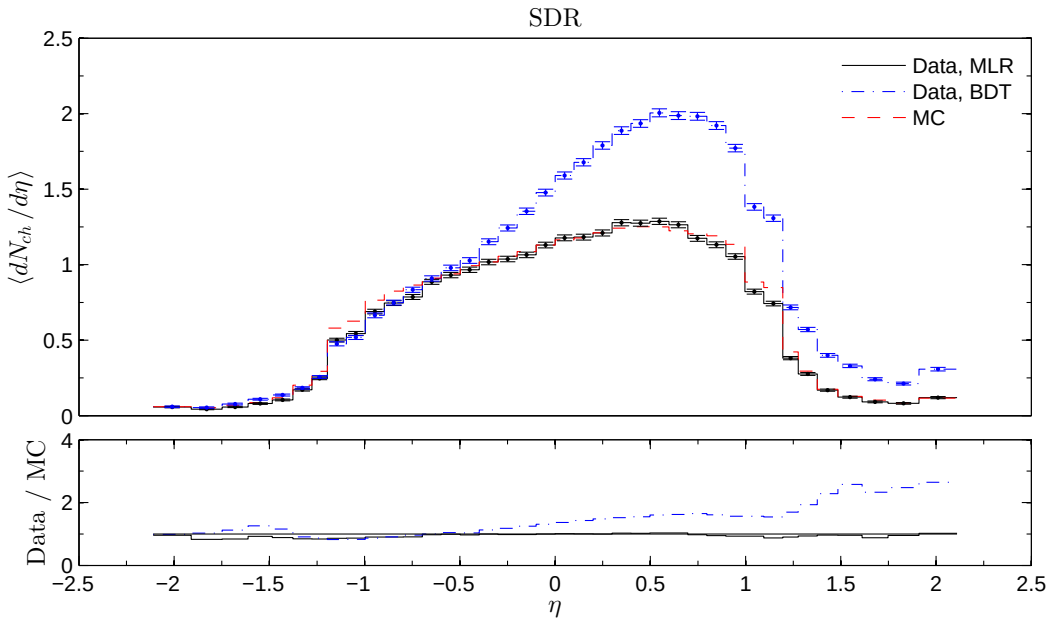


Figure 10.12: Track multiplicities N_{ch} as a function of pseudorapidity.

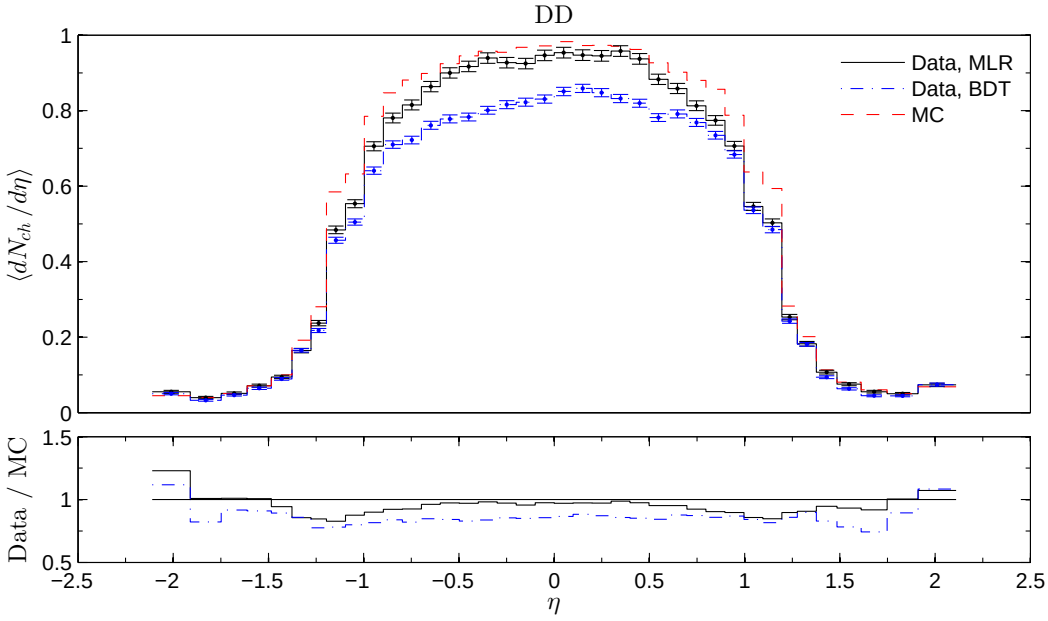


Figure 10.13: Track multiplicities N_{ch} as a function of pseudorapidity.

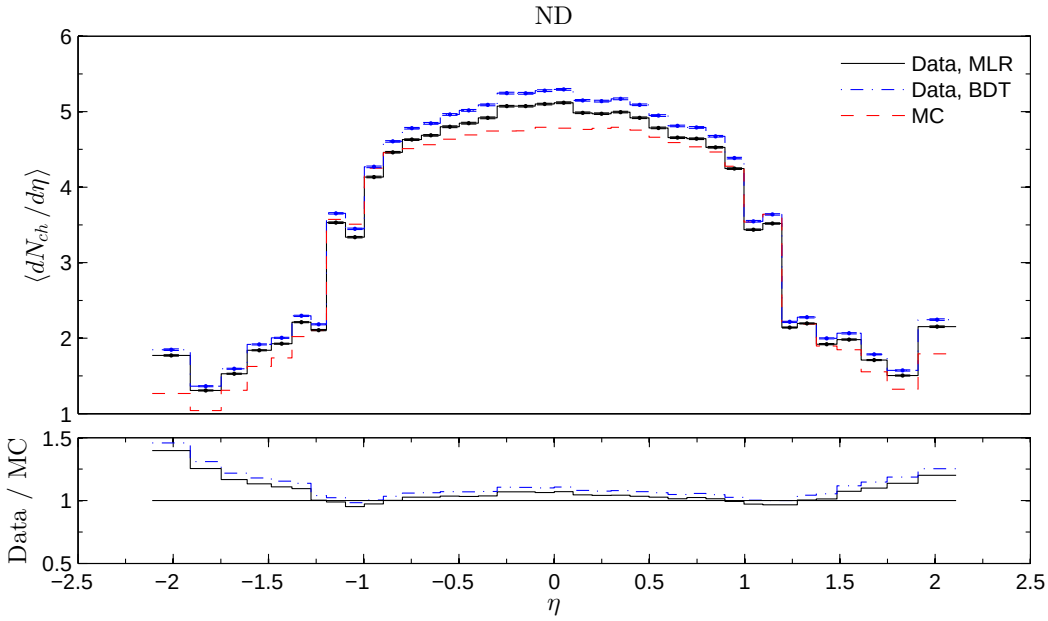


Figure 10.14: Track multiplicities N_{ch} as a function of pseudorapidity.

10.9 Multiplicity N_{ch} histograms

The track multiplicity histograms are overlaid in Figure 10.15 and each process class is compared with the Monte Carlo predictions in Figure 10.16. Non-diffractive events are on much higher multiplicities, which is expected because most of the diffraction goes over the rapidity span of the tracking systems.

From Figure 10.15 can be again seen that both single diffractive classes are nearly perfectly symmetric, and double diffraction has nearly similar distribution than both single diffractive. The distributions for each process class are again matching reasonably well the Monte Carlo predictions.

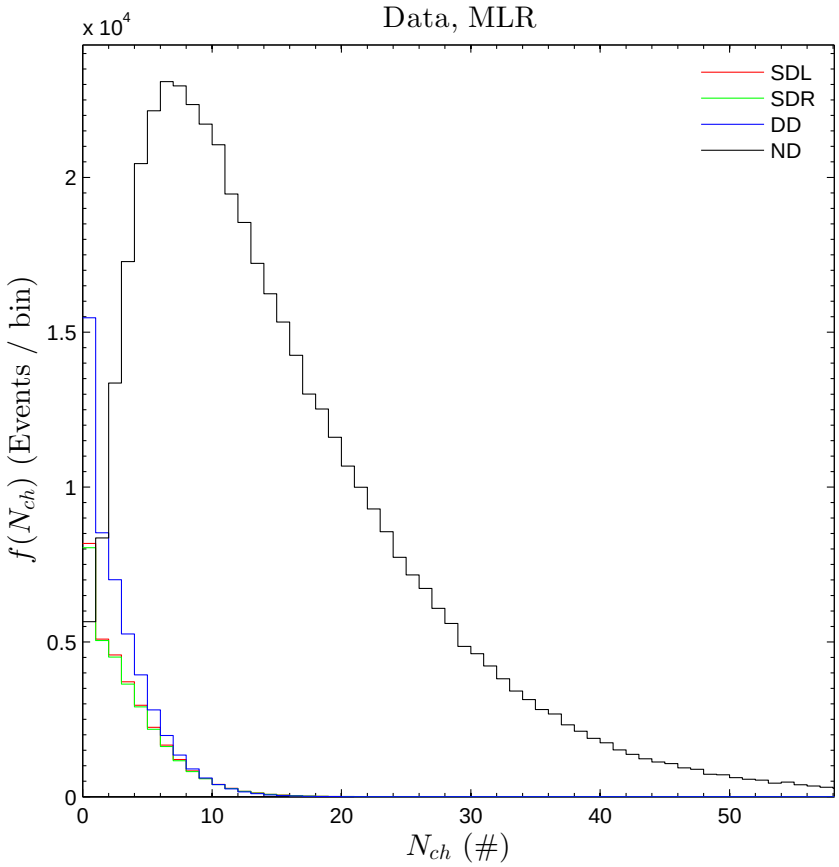


Figure 10.15: Track multiplicity histograms.

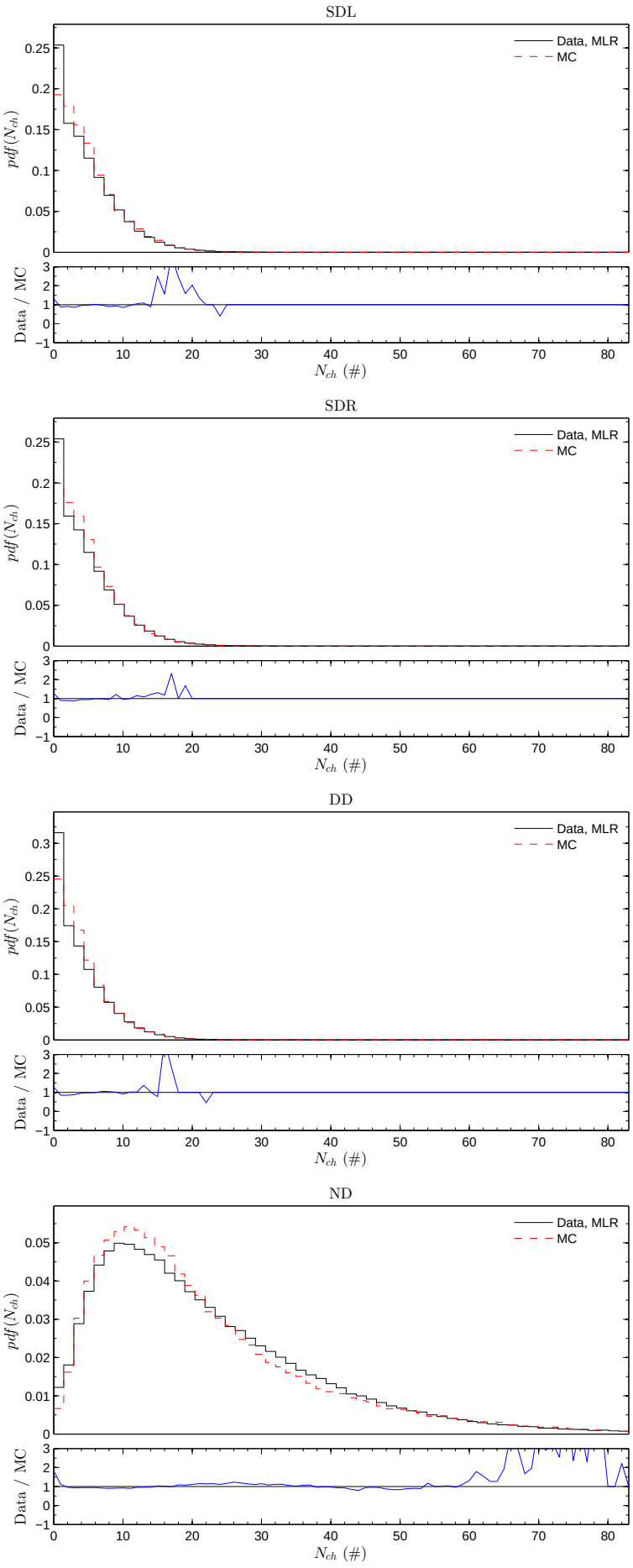


Figure 10.16: Normalized histograms of track multiplicities.

10.10 Cross section measurements

The cross sections were estimated using the algorithmic chain explained in section 10.1 with the pile-up discrimination as an extra procedure. The pile-up discrimination was applied after the pre-selection cut was applied to the 0-bias sample. The pre-selection cut was expected to be hard enough, i.e. with perfect purities, so that the "non-interaction" background correction was not necessary.

Finally, after obtaining the mean values of a posteriori probabilities, compensating the pre-cut efficiencies and by using the pre-measured total inelastic cross section σ_{inel} given in Table 10.3 as the final multiplicative normalization factor, the integrated cross section estimates[4] for each scattering process class are shown in Table 10.6. These values are calculated using the priors given in Table 10.3 and alternatively "fully Bayesian", marginalized (integrated) over a distribution of priors are given in Appendix A.6.

Table 10.6: Cross sections with statistical (Poisson) and systematic (1σ) errors for the CDF 0-bias run at $\sqrt{s} = 1.96$ TeV.

SDL [mb]	SDR [mb]	DD [mb]	ND [mb]
$4.87 \pm 0.03 \pm 1.06$	$4.83 \pm 0.03 \pm 1.04$	$6.16 \pm 0.03 \pm 1.93$	$45.20 \pm 0.07 \pm 1.59$

When comparing the results of Table 10.6 with the phenomenological theory predictions in Table 10.7, the measured values of diffractive cross sections are surprisingly close. The predictions by the MBR model are tuned to earlier CDF run I measurements at $\sqrt{s} = 1.8$ TeV of single [34] and double diffraction [3]. Of course, one must be cautions making any strong claims because the prior probabilities have a strong effect in this physics problem. Mathematically speaking this is due to heavy overlap of the class densities in the multivariate space.

Table 10.7: Theory estimates [mb] at $\sqrt{s} = 1.96$ TeV normalized to the same σ_{inel}.

Model	SDL	SDR	DD	ND	CD	σ_{inel}
PYTHIA 6 [48]	6.46	6.46	7.20	40.94	-	61.06
PYTHIA 8 (MBR) [16]	5.11	5.11	7.67	42.41	0.78	61.06

The prior dependency is not a shortcoming of the multivariate approach, it affects the simple cut-and-count approach as much. Instead, probabilistic multivariate algorithms makes this prior dependency very transparent and easily evaluated. Ultimately, this dependency on priors comes from a limited η-span of CDF and the physical nature of diffractive events. Especially double diffraction is a difficult process to measure due to experimental signature which is imitating ND and SD processes.

[4]Systematic error estimates include only the effect of prior probabilities.

Chapter 11

Discussion and conclusions

By using probabilistic multivariate classification algorithms, not only hard classifications are obtained, but also probabilities. These probabilities gives us *confidence* on an event-by-event basis and can be used also as *weights* in later physics analysis and estimation of physical distributions, as was done in this work. Most important, it was shown both analytically and numerically with a Monte Carlo sample that integration of a posteriori probabilities is a mathematically consistent way to estimate relative cross sections under the Bayesian framework. This is natural, because scattering cross sections are ultimately interaction probabilities. The demonstrated multivariate techniques are new to high energy physics, and can be used as a replacement or alternative for Boosted Decision Trees and Neural Networks in some analyses.

The classification accuracy is upper bounded by the chosen Monte Carlo generator. However, not bounded in a trivial way. The most important is how the differential cross sections are described on the boundaries between different scattering processes, i.e. when the processes' experimental signatures start to overlap in the multivariate space. As important is where these boundaries are modelled to be. Less important are the regions of the multivariate space, where no overlap between class densities happens. Thus, any advances in theoretical models or actual practical implementations of theory in the Monte Carlo event generators have a major impact.

The algorithmic ℓ_1-regularization is a mathematically sound variable selection tool and enables the utilization of high dimensional multivariate spaces. The multivariate space and rapidity gap correspondence was demonstrated using the ℓ_1-regularization, based on Monte Carlo, for the first time. The conclusion was that the gaps are defining signatures of diffraction in the limit when full information entropy available in the multivariate space is approximated with only few, high pseudorapidity $|\eta|$ components. This multivariate space - rapidity gap correspondence is an example of gaining a physical, i.e. kinematical or dynamical interpretation of the multivariate space. Not just using the multivariate algorithm as a black-box, as often happens.

The effectiveness of this multivariate approach was evaluated with a 0-bias sample measured at the CDF run II experiment by obtaining cross section measurements of diffraction, which were compared with theoretical predictions. The measured values matched the phenomenological theory expectations within estimated errors. This kind of simultaneous measurement of diffractive and non-diffractive scatter-

ing cross sections by using multi-class multivariate algorithms has never been done before using experimental data. Based on estimated cross sections, data seems to favour slightly smaller cross sections for diffractive classes than what is given in Monte Carlo event generators. However, because data was not pile-up free, this might be also an artefact coming from non-perfect pile-up correction.

Unfortunately hadronic diffraction is still theoretically poorly understood. Especially the diffractive mass $1/M_X^2$ and exponential t dependency of the cross sections are open questions. The multidimensional densities for diffractive events depend ultimately on differential cross sections which are driven by functional dependency of M_X^2, t and Regge parameters. The hadronization process modelling with particle multiplicities creates also some effects. The priors depend on integrated cross sections and what fraction of events can be seen within detector pseudorapidity range (acceptance), which are both driven by differential cross section formulas. However, the 4-momentum conservation should be well understood and because signature of diffraction is mainly kinematical, this should be a strong argument for the multivariate approach.

The usual factorization of Regge formulas into Pomeron flux and Pomeron-proton cross-section is also challenging, because effectively only their product can be measured. The triple-Pomeron coupling, which gives the $1/M_X^2$ dependency of the differential cross sections in the high mass limit, is also one of the most crucial parameters of the Regge phenomenology. New insight into these questions could be possibly acquired by developing multivariate techniques which allow estimation of diffractive mass and 4-momentum transfer event-by-event.

Also problematic fact is that there is no rigorous way to theoretically estimate cross section of soft non-diffractive events. Thus, it is often set to a value which is obtained by substracting diffractive and elastic contribution from the total cross section value. On the other hand, the question of low mass single and double diffraction cannot be answered without instrumentation in very forward direction, that is, it cannot be solved purely algorithmically. However, low mass central diffraction and its resonance structure could be an interesting subject for the multivariate analysis techniques.

Finally, the work at LHC energies has been already started with the experiments at CERN, and the work of this thesis is an extensive proof of concept for those upcoming multivariate analyses of diffraction.

Bibliography

[1] F. Abe et al. (CDF Collaboration), Phys. Rev. Lett. 79, 2636, 1997.

[2] T. Affolder et al. (CDF Collaboration), Phys. Rev. Lett. 85, 4215, 2000.

[3] T. Affolder et al. (CDF Collaboration), Phys. Rev. Lett. 87, 131802, 2001.

[4] M. Albrow, A. De Roeck, V. Khoze, J. Lämsa, E. Norbeck, Y. Onel, R. Orava, M.G. Ryskin, Forward Physics with Rapidity Gaps at the LHC, arxiv.org/abs/0811.0120, 2009.

[5] J. Baez and J. Huerta, The Algebra of Grand Unified Theories, Bull. Amer. Math. Soc. 47, 483-552, 2010.

[6] V. Barone, E. Predazzi, *High-Energy Particle Diffraction*, Springer, 2002.

[7] C. Bishop, *Pattern Recognition and Machine Learning*, Cambridge University Press, 2007.

[8] J.D. Bjorken, Rapidity gaps and jets as a new physics signature in very high-energy hadron hadron collisions, Int. J. Mod. Phys. A7 (1992) 4189; Phys. Rev. D47 101, 1993.

[9] J.D. Bjorken, Current Induced Reactions, Lecture Notes in Physics, Vol. 56, p. 93, Springer Verlag, 1976.

[10] V. Blobel, Data unfolding, Terascale Statistics Tools School, Spring 2010, www.desy.de/∼blobel/school_march10.pdf.

[11] CDF II technical design report, www-cdf.fnal.gov/upgrades/tdr/tdr.html.

[12] CDF II detector image, inspirehep.net/record/1082043/files/figures_p28_cdf_cdf_gerva.png.

[13] CDF II experiment images, www.fnal.gov/pub/presspass/images.

[14] M. Cacciari, G. P. Salam, G. Soyez, The anti-k_t jet clustering algorithm, JHEP 0804:063, 2008.

[15] V. Chiochia, G. Dissertori, T. Gehrmann, *Particle Physics Phenomenology II*, lecture notes script, ETH Zürich, 2012.

[16] R. Ciesielski, K. Goulianos, MBR Monte Carlo Simulation in Pythia 8, 2012, arxiv.org/pdf/1205.1446v2.pdf.

[17] M.S. Costa, V. Goncalves, J. Penedones, Conformal Regge theory, 2013, arxiv.org/abs/1209.4355.

[18] J.C. Collins, Factorization in hard diffraction, 2002, arxiv.org/pdf/hep-ph/0107252v2.pdf.

[19] J.C. Collins, D.E. Soper, G. Sterman, Factorization of Hard Processes in QCD, 2004, arxiv.org/pdf/hep-ph/0409313v1.pdf.

[20] J.C. Collins, Light-cone Variables, Rapidity and All That, 1997, arxiv.org/abs/hep-ph/9705393.

[21] CERN Courier, QCD scattering: from DGLAP to BFKL, July 20, 2010, cerncourier.com/cws/article/cern/43194.

[22] A. De Roeck, V.A. Khoze, A.D. Martin, R. Orava and M.G. Ryskin, Ways to detect a light Higgs boson at the LHC, Eur. Phys. J.C25, 391, 2002, arxiv.org/abs/hep-ph/0207042.

[23] Yu. L. Dokshitzher, V.A. Khoze, T. Sjöstrand, Rapidity gaps in Higgs production, Phys. Lett. B274 116, 1992.

[24] S. Donnachie, G. Dosch, P. Landshoff, O. Nachtmann, *Pomeron Physics and QCD*, Cambridge University Press, 2002.

[25] R.K. Ellis, W.J Stirling, B.R Webber, *QCD and collider physics*, Cambridge, 1996.

[26] Stephen D. Ellis, Particle Physics course notes, courses.washington.edu/phys55x/Physics%20557_lec11.htm.

[27] Fermilab accelerator chain, www.fnal.gov/pub/inquiring/physics/accelerators/chainaccel.html.

[28] R.P. Feynman, *Photon-Hadron Interactions*, W. A. Benjamin, 1972.

[29] T. Gehrmann, R. Wallny, *Particle Physics Phenomenology I*, lecture notes script, ETH Zürich, 2012.

[30] F. Gelis, School on QCD, low-x physics, saturation and diffraction, lecture slides, www-d0.fnal.gov/Run2Physics/qcd/qcd_school/gelis1.pdf.

[31] M.L. Good, W.D. Walker, Phys. Rev. 120, 1857, 1960.

[32] E. Gotsman, E. Levin, U. Maor, E. Naftali, A. Prygarin, Survival Probability of Large Rapidity Gaps, 2005, arxiv.org/abs/hep-ph/0511060.

[33] E. Gotsman, E. Levin, U. Maor, Large rapidity gaps in pp collisions, Phys.Rev.B39, 199, 1993.

[34] K. Goulianos, Phys. Lett. B 358, 379, 1995.

[35] V.N Gribov, *Elementary Particles*, Vol. I, p. 65.

[36] W Heisenberg, Zeitschrift fur Physik 133, 1952, page 65.

[37] G. Ingelman, P.E. Schlein, Phys. Lett. B152, 256, 1985.

[38] P. Koistinen, *Statistical pattern recognition*, Lecture notes (in Finnish), University of Helsinki, 2002.

[39] B. Krishnapuram, L. Carin, M.A.T. Figueiredo, A.J. Hartemink, Sparse Multinomial Logistic Regression: Fast Algorithms and Generalization Bounds, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 27, No. 6, June 2005.

[40] B. List, Why and When to Optimize Efficiency Times Purity, ETH Zurich, Institute for Particle Physics, 2002,
www.desy.de/∼blist/notes/whyeffpur.ps.gz.

[41] MAD-X, A program for accelerator design and simulation,
frs.home.cern.ch/frs/Xdoc/mad-X.html.

[42] H.I. Miettinen, J. Pumplin, Phys. Rev. D18, 1696, 1978.

[43] P. Nadolsky, J. Gao, M. Guzzi, J. Huston, H.-L. Lai, Z. Li, J. Pumplin, D. Stump, C.-P. Yuan, Progress in CTEQ-TEA PDF analysis, arxiv.org/abs/1206.3321.

[44] Particle Data Group (PDG), 2013, pdg.lbl.gov.

[45] M. E. Peskin, D. V. Schroeder, *An Introduction to Quantum Field Theory*, Addison-Wesley Publishing Company, 1995.

[46] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*, Carl Edward Rasmussen and Christopher K. I. Williams, The MIT Press, 2006.

[47] A. Signer, The Standard Model, Lecture notes given at Rutherford Appleton Laboratory, 2002.

[48] T. Sjöstrand, PYTHIA6 Manual,
home.thep.lu.se/∼torbjorn/pythia/lutp0613man2.pdf.

[49] Tevatron accelerator complex picture,
images.iop.org/objects/ccr/cern/51/9/22/CCtev2_09_11.jpg.

Appendix A

Appendix

A.1 Four vectors in Minkowski space

This is basic kinematics and natural units are used, i.e. $c = \hbar = 1$. Thus, energy (GeV), mass (GeV/c^2), momentum (GeV/c) can be all expressed in GeV, time $(\text{GeV}/\hbar)^{-1}$ and length $(\text{GeV}/\hbar c)^{-1}$ in 1/GeV and area $(\text{GeV}/\hbar c)^{-2}$ in GeV^{-2}. The important relativistic relations are then $\gamma = E/mc^2 \mathrel{\mathring{=}} E/m$ and $\beta = |\vec{p}|c/E \mathrel{\mathring{=}} |\vec{p}|/E$ with $\gamma = 1/\sqrt{1-\beta^2}$. [44] Highly important is the Lorentz invariant mass formula

$$m = \sqrt{\left(\sum_i E_i\right)^2 - \left|\sum_i \vec{p}_i\right|^2}, \tag{A.1}$$

where sum runs over a system of particles. This is coming from the energy-momentum relation $E^2 = m^2c^4 + |\vec{p}|^2c^2$. Modern notation treats mass always as constant, i.e. the concept of relative mass is not favoured.

The 4-vector notation is used, here for example the 4-momentum vector and the 4-position

$$p^\mu = (p^0, p^1, p^2, p^3) = (\gamma mc, \gamma m\vec{v}) = (\frac{E}{c}, \vec{p}) \mathrel{\mathring{=}} (E, \vec{p}) \tag{A.2}$$

$$x^\mu = (ct, x, y, z) \mathrel{\mathring{=}} (t, \vec{x}). \tag{A.3}$$

The 4-vectors belongs to space of special relativity, i.e. to flat Minkowski space $x^\mu \in \mathcal{M}^4$, and are naturally contravariant, upper index vectors. A covariant, lower index vector is obtained by contraction with the metric tensor. The Lorentz frame independent quantity, scalar product between two 4-vectors a and b is

$$a \cdot b = g_{\mu\nu}a^\nu b^\mu = a_\mu b^\mu = a^0b^0 - \vec{a} \cdot \vec{b}, \tag{A.4}$$

with the metric tensor $g_{\mu\nu} = g^{\mu\nu} = \text{diag}(1, -1, -1, -1)$ of Minkowski space, which uses the usual sign notation in particle physics. If the particle is *on-shell*, i.e. real particle, $p^2 = p^\mu p_\mu = m^2$ by energy-momentum relation. Einstein's summation convention is used over corresponding upper and lower indices. However, often the metric tensor and greek letter indices are left out for notational simplicity, when working with scalar products of 4-vectors.

Now, let us have a two body scattering process $2 \to 2$ with the initial state 4-vectors p_1, p_2 and the corresponding final state 4-vectors p_3, p_4. Commonly used

Lorentz scalars, the so-called *Mandelstam variables* are [29]

$$s = (p_1 + p_2)^2 = (p_3 + p_4)^2 \quad \text{Center of mass energy squared} \tag{A.5}$$
$$t = (p_1 - p_3)^2 = (p_2 - p_4)^2 \quad \text{4-momentum transfer squared} \tag{A.6}$$
$$u = (p_1 - p_4)^2 = (p_2 - p_3)^2 \quad \text{Final states crossed t-channel} \tag{A.7}$$

which have the relation

$$s + t + u = \sum_{i=1}^{4} m_i^2, \tag{A.8}$$

with particle masses m_i.

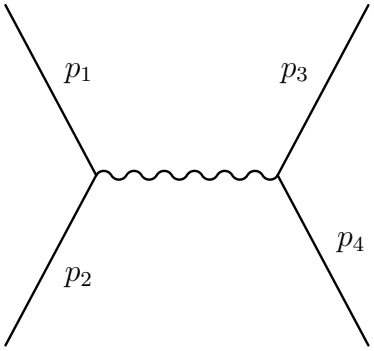


Figure A.1: s-channel.

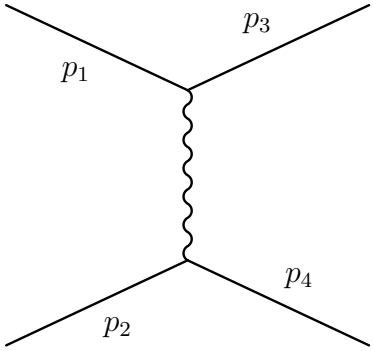


Figure A.2: t-channel.

The s-channel in Figure A.1 has an interpretation of an annihilation process, t-channel in Figure A.2 as a scattering where an intermediate particle is exchanged and u-channel as t-channel where final states are crossed $p_3 \leftrightarrow p_4$. Commonly seen probability amplitudes are often defined as a function of s and t as $A(s,t) \in \mathbb{C}$.

A.2 Rapidity

By definition, rapidity in particle physics for a given particle with energy E, mass m, longitudinal momentum $p_z = |\vec{p}|\cos\theta$ and transverse momentum $p_T = |\vec{p}|\sin\theta$ with respect to the beam axis is [29]

$$y = \frac{1}{2}\ln\left(\frac{E+p_z}{E-p_z}\right) = \frac{1}{2}\ln\left(\frac{(E+p_z)^2}{m^2+p_T^2}\right) \tag{A.9}$$

$$\Leftrightarrow \tanh(y) = \frac{p_z}{E}, \tag{A.10}$$

where the hyperbolic function gives rapidity an interpretation as a hyperbolic angle. It is easy to show that rapidity is constrained by

$$y \leq \frac{1}{2}\ln\left(\frac{2E^2}{m^2}\right) = y_{\max} \tag{A.11}$$

with symmetry $y_{\min} = -y_{\max}$.

Rapidity is in the massless limit, or for highly relativistic ($|\vec{p}| \gg m$) particles [20]

$$\eta = y|_{m=0} = \ln\left(\frac{E+p_z}{p_T}\right) = -\ln\left(\tan\frac{\theta}{2}\right), \tag{A.12}$$

which is known as *pseudorapidity*, and it runs from minus infinity to plus infinity. This is experimentally easy to measure in experiments, because it depends only on the angle $\theta \in [0,\pi]$ over the beam axis.

So rapidity is a variable which describes together both energy E and longitudinal momentum p_z of particles, and $\partial y/\partial p_z = 1/E$. Also it is sometimes useful to write the 4-momentum as [29]

$$p^\mu = (m_T\cosh y, p_T\cos\phi, p_T\sin\phi, m_T\sinh y), \tag{A.13}$$

where the transverse mass is $m_T = \sqrt{m^2+p_T^2}$ and the transverse angle $\phi \in [0,2\pi]$ is around the beam axis. Finally, the most important fact is that differences between rapidities are Lorentz boost invariant under longitudinal boosts, i.e. $y_1 - y_2$ or $\eta_1 - \eta_2$ are frame independent. This is the reason to use pseudorapidity η instead of θ.

A.3 Lie algebras of the Standard Model

QCD $\mathfrak{su}(3)$

The Lie algebra $\mathfrak{su}(3)$ is spanned by traceless hermitian 3×3 matrices with determinant 1, equipped with a basis represented by eight matrices. The eight generators are

$$T_a = \frac{1}{2}\lambda_a, \tag{A.14}$$

where λ_a is [24]

$$\begin{aligned}
\lambda_1 &= \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \lambda_2 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \lambda_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \\
\lambda_4 &= \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \quad \lambda_5 = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix} \quad \lambda_6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \\
\lambda_7 &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix} \lambda_8 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}
\end{aligned} \tag{A.15}$$

These are known as the Gell-Mann matrices, used to model $SU(3)_C$ gauge symmetry of QCD and in an approximate way in flavour symmetry invented by Gell-Mann. The totally antisymmetric structure constants f_{abc} are given by [29]

$$f_{abc} = \frac{1}{4i}\mathrm{Tr}([\lambda_a, \lambda_b]\lambda_c). \tag{A.16}$$

Using these, the fundamental commutation relation

$$[T_a, T_b] = i \sum_c f_{abc} T_c \tag{A.17}$$

can be proven to hold.

Weak force $\mathfrak{su}(2)$

The Lie algebra $\mathfrak{su}(2)$ is spanned by antihermitian 2×2 matrices with determinant 1, equipped with a basis represented by three matrices. The three generators are

$$T_j = \frac{1}{2}\tau_j, \tag{A.18}$$

where τ_j is [29]

$$\tau_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \tau_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \tau_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{A.19}$$

These are the famous Pauli matrices and the corresponding generators obey the fundamental commutation relation

$$[T_i, T_j] = i \sum_k \epsilon_{ijk} T_k, \tag{A.20}$$

where $\epsilon_{ijk} = \epsilon^{ijk}$ is the totally antisymmetric Levi-Civita permutation tensor, with $\epsilon_{123} = +1$.

Pauli matrices are used also in ordinary quantum mechanics to represent angular momentum and intrinsic angular momentum (spin) operator components. This is natural because mathematically $SU(2)$ group is the double-cover of $SO(3)$, which is the usual rotation group. Also, when looking closely, one can see that the Pauli matrices are actually embedded inside the first three Gell-Mann matrices.

Later on Heisenberg used this same mathematics to describe strong nuclear force symmetry between protons and neutrons, which is known to be only an approximation. The concept is known as isotopic spin or shortly as *isospin*. The idea is that both proton and neutron would be different states of the same particle, nucleon, transformed using the spin-$\frac{1}{2}$ representation of $SU(2)$.

In the context of weak force group $SU(2)_L$, the Pauli matrices correspond to three components of the weak isospin I_j. This idea was due to Glashow, Weinberg and Salam, the developers of the electroweak unification.

QED $\mathfrak{u}(1)$

Finally, the simplest Lie algebra $\mathfrak{u}(1)$ in the Standard Model is made by antihermitian 1×1 complex matrices, that means, imaginary numbers. Common choice as a basis vector is $\frac{i}{3}$, due to fractional charges of quarks.

Table A.1: Elementary particles in the Standard Model (without antiparticles for fermions) with corresponding irreducible representation algebras. The irreducible representation of the weak hypercharge $\hat{Y}$ operator is marked with $\mathbb{C}_y$, from $\hat{Y}|\psi\rangle = y|\psi\rangle$. [5]

Type	$SU(3)$ irrep	$SU(2)$ irrep	$U(1)$ irrep
1st gen. Fermions			
Quarks			
$(u_r, u_g, u_b, d'_r, d'_g, d'_b)_L$	$\mathbb{C}^3$	$\mathbb{C}^2$	$\mathbb{C}_{\frac{1}{3}}$
$(u_r, u_g, u_b)_R$	$\mathbb{C}^3$	$\mathbb{C}$	$\mathbb{C}_{\frac{4}{3}}$
$(d_r, d_g, d_b)_R$	$\mathbb{C}^3$	$\mathbb{C}$	$\mathbb{C}_{-\frac{2}{3}}$
Leptons			
$(\nu_e, e^-)_L$	$\mathbb{C}$	$\mathbb{C}^2$	$\mathbb{C}_{-1}$
$(e^-)_R$	$\mathbb{C}$	$\mathbb{C}$	$\mathbb{C}_{-2}$
$(\nu_e)_R$ not observed!	$\mathbb{C}$	$\mathbb{C}$	$\mathbb{C}_0$
2nd gen. Fermions			
Quarks			
$(c_r, c_g, c_b, s'_r, s'_g, s'_b)_L$	$\mathbb{C}^3$	$\mathbb{C}^2$	$\mathbb{C}_{\frac{1}{3}}$
$(c_r, c_g, c_b)_R$	$\mathbb{C}^3$	$\mathbb{C}$	$\mathbb{C}_{\frac{4}{3}}$
$(s_r, s_g, s_b)_R$	$\mathbb{C}^3$	$\mathbb{C}$	$\mathbb{C}_{-\frac{2}{3}}$
Leptons			
$(\nu_\mu, \mu^-)_L$	$\mathbb{C}$	$\mathbb{C}^2$	$\mathbb{C}_{-1}$
$(\mu^-)_R$	$\mathbb{C}$	$\mathbb{C}$	$\mathbb{C}_{-2}$
$(\nu_\mu)_R$ not observed!	$\mathbb{C}$	$\mathbb{C}$	$\mathbb{C}_0$
3rd gen. Fermions			
Quarks			
$(t_r, t_g, t_b, b'_r, b'_g, b'_b)_L$	$\mathbb{C}^3$	$\mathbb{C}^2$	$\mathbb{C}_{\frac{1}{3}}$
$(t_r, t_g, t_b)_R$	$\mathbb{C}^3$	$\mathbb{C}$	$\mathbb{C}_{\frac{4}{3}}$
$(b_r, b_g, b_b)_R$	$\mathbb{C}^3$	$\mathbb{C}$	$\mathbb{C}_{-\frac{2}{3}}$
Leptons			
$(\nu_\tau, \tau^-)_L$	$\mathbb{C}$	$\mathbb{C}^2$	$\mathbb{C}_{-1}$
$(\tau^-)_R$	$\mathbb{C}$	$\mathbb{C}$	$\mathbb{C}_{-2}$
$(\nu_\tau)_R$ not observed!	$\mathbb{C}$	$\mathbb{C}$	$\mathbb{C}_0$
Gauge Bosons			
$(g_{rg}, g_{rb}, g_{gr}, g_{gb}, g_{br}, g_{bg}, g_{rr} - g_{bb}, g_{bb} - g_{gg})$	$\mathfrak{su}(3)$	$\mathbb{R}$	$\mathbb{R}$
(W_1, W_2, W_3)	$\mathbb{R}$	$\mathfrak{su}(2)$	$\mathbb{R}$
(B_0)	$\mathbb{R}$	$\mathbb{R}$	$\mathfrak{u}(1)$
(H^0)	$\mathbb{C}$	$\mathbb{C}^2$	$\mathbb{C}_1$

A.4 PYTHIA 6.4x setup

Non-diffractive (Minimum-bias)

The tree-level $2 \to 2$ hard QCD processes (ISUB $= 11, \ldots, 68$). Soft, non-perturbative QCD (ISUB $= 95$).

Subprocess number	Process
ISUB $= 11$	$q_i q_j \to q_i q_j$
ISUB $= 12$	$q_i \bar{q}_i \to q_k \bar{q}_k$
ISUB $= 13$	$q_i q_i \to gg$
ISUB $= 28$	$q_i g \to q_i g$
ISUB $= 53$	$gg \to q_k \bar{q}_k$
ISUB $= 68$	$gg \to gg$
ISUB $= 95$	low p_T

CTEQ5L parton density functions. Standard underlying event modelling CDF-II tune A.

Diffractive

Subprocess number	Process
ISUB $= 92$	Single diffraction $AB \to AX$
ISUB $= 93$	Single diffraction $AB \to XB$
ISUB $= 94$	Double diffraction $AB \to XY$

A.5 Multivariate method consistency

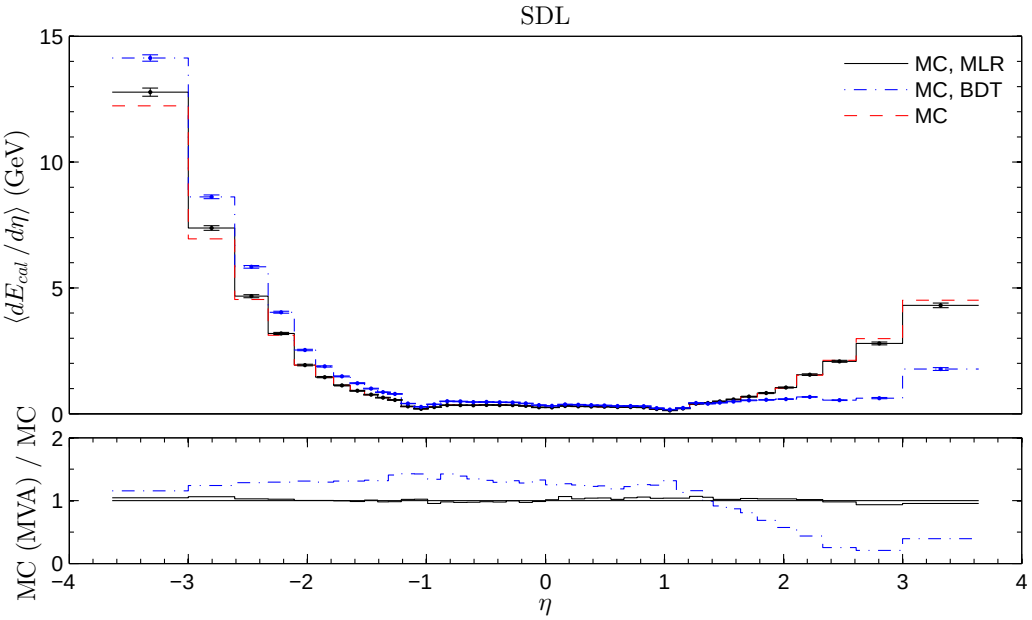


Figure A.3: Calorimeter deposits E_{cal} as a function of pseudorapidity.

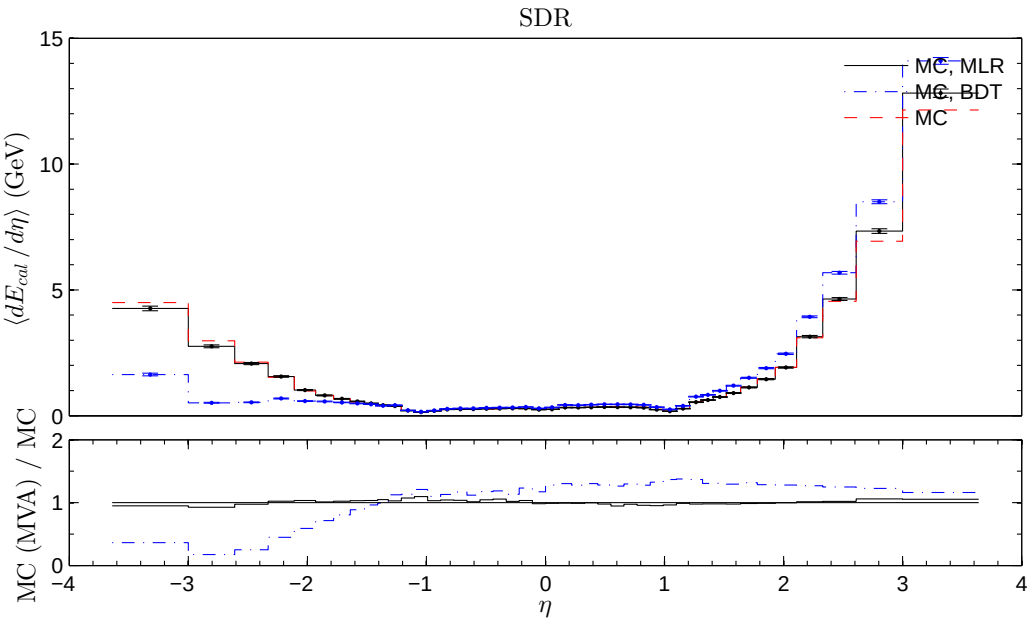


Figure A.4: Calorimeter deposits E_{cal} as a function of pseudorapidity.

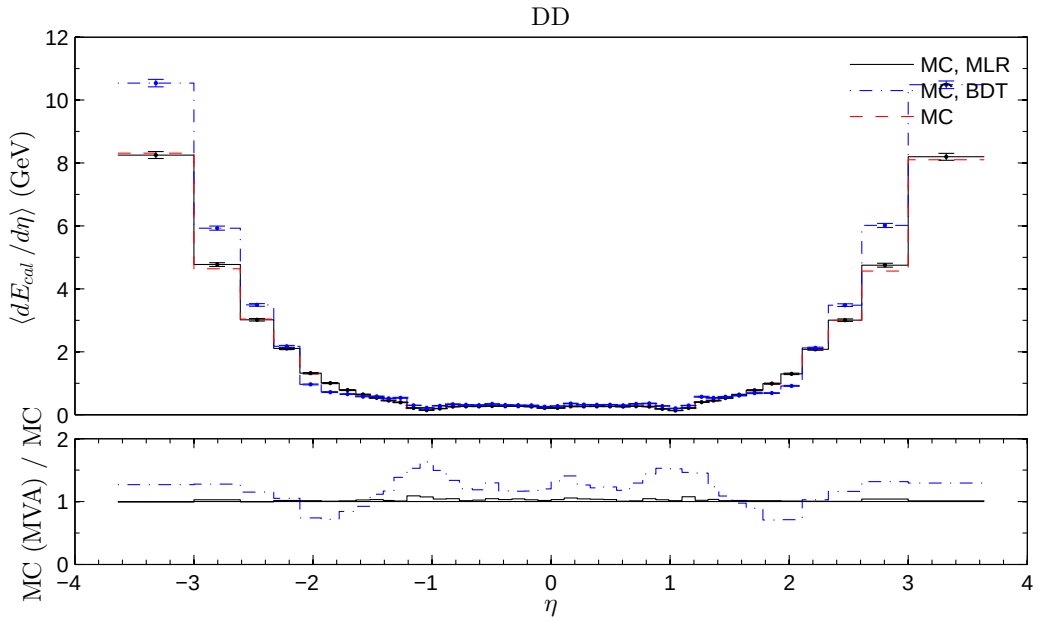


Figure A.5: Calorimeter deposits E_{cal} as a function of pseudorapidity.

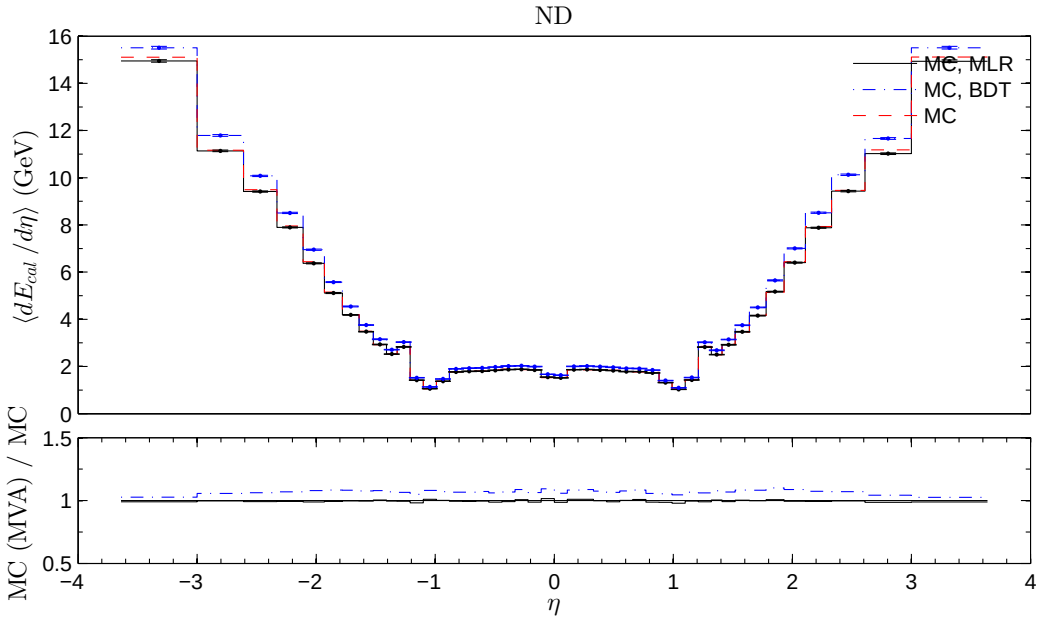


Figure A.6: Calorimeter deposits E_{cal} as a function of pseudorapidity.

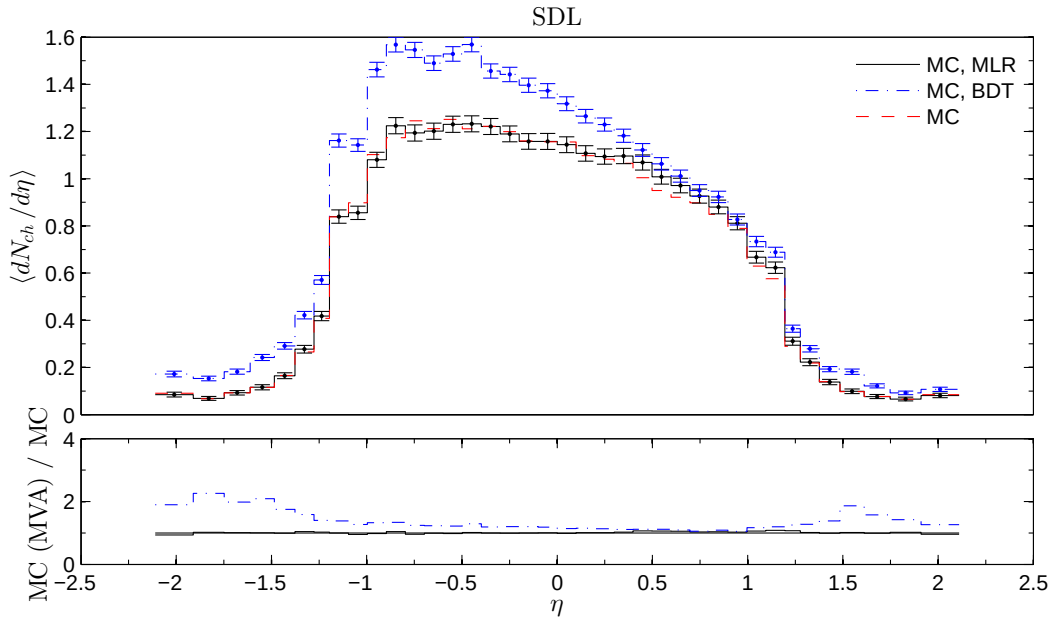


Figure A.7: Track multiplicities N_{ch} as a function of pseudorapidity.

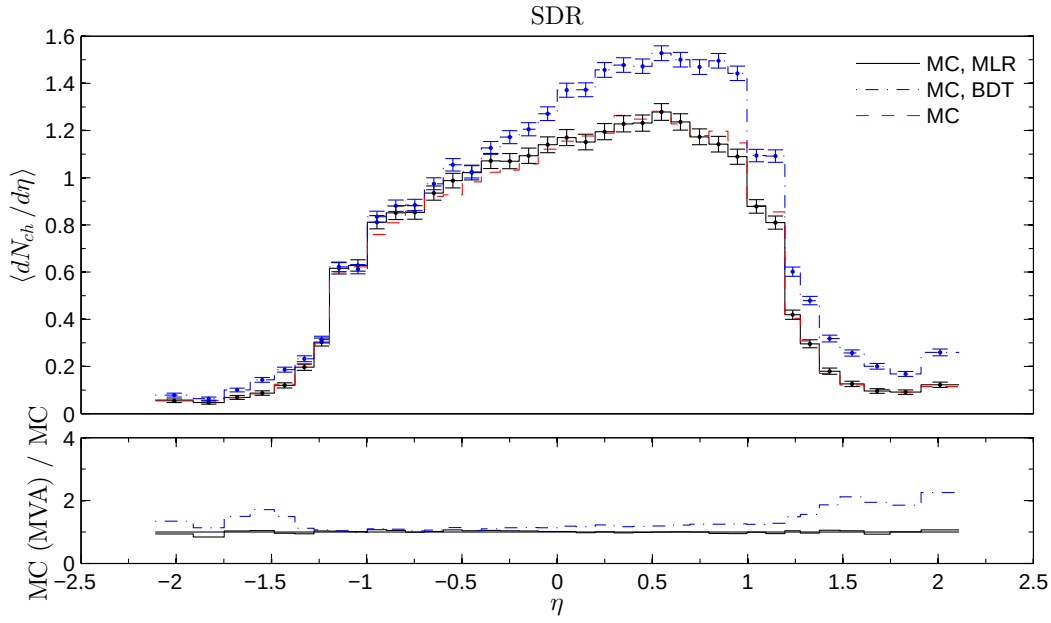


Figure A.8: Track multiplicities N_{ch} as a function of pseudorapidity.

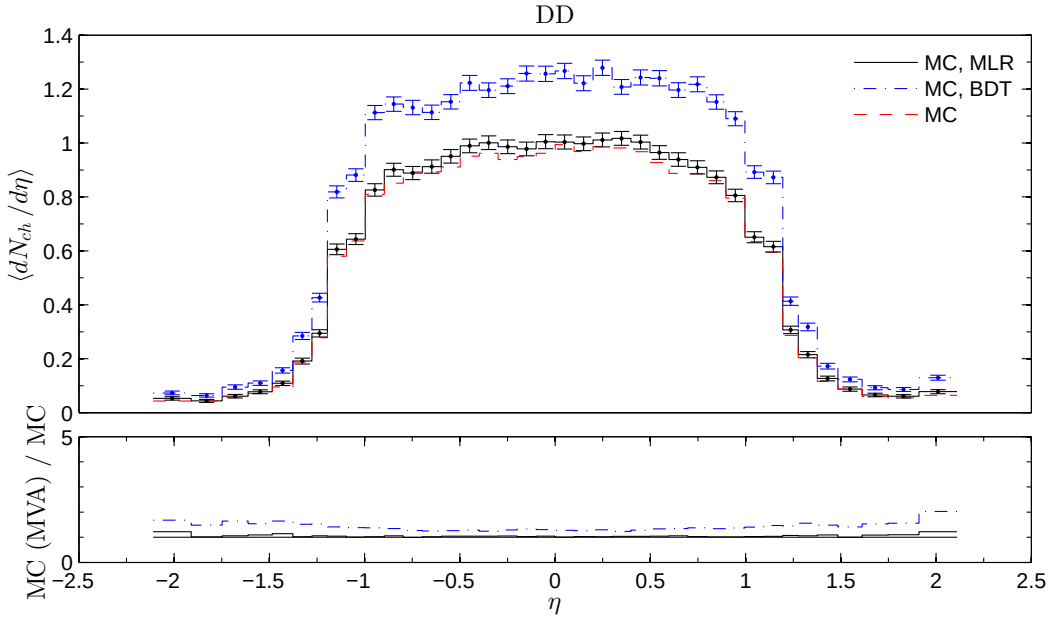


Figure A.9: Track multiplicities N_{ch} as a function of pseudorapidity.

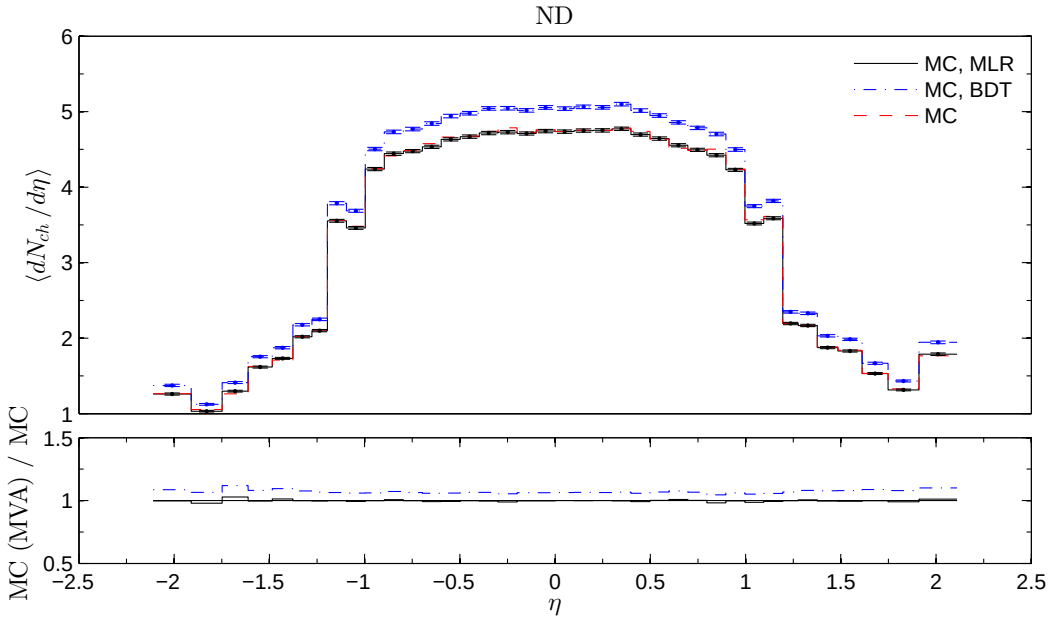


Figure A.10: Track multiplicities N_{ch} as a function of pseudorapidity.

A.6 Prior probabilities systematics

Cross section estimates for the CDF 0-bias sample by using different theory priors P_j are given in Table A.2. The algorithmic chain is the same as explained in Section 10.1. The RMS error between a posteriori estimated and a priori expected cross sections is denoted with $\Delta_{RMS} = \sqrt{\frac{1}{|\mathcal{C}|}\sum_j (\hat{\sigma}_j - \sigma_j)^2}$.

Table A.2: Prior probabilities systematics. Prior combinations with $P_{ND} < 0.66$ do not satisfy Pumplin s-channel unitarity bound! The last line includes marginalized, "fully Bayesian" estimates with 1σ standard deviations.

P_{SDL}	P_{SDR}	P_{DD}	P_{ND}	σ_{SDL} (mb)	σ_{SDR} (mb)	σ_{DD} (mb)	σ_{ND} (mb)	Δ_{RMS}
0.07	0.07	0.04	0.82	4.75 ± 0.03	4.66 ± 0.03	2.97 ± 0.02	48.68 ± 0.07	0.80
0.07	0.07	0.05	0.82	4.36 ± 0.03	4.28 ± 0.03	3.66 ± 0.02	48.75 ± 0.07	0.77
0.06	0.06	0.06	0.82	3.99 ± 0.02	3.92 ± 0.02	4.34 ± 0.02	48.82 ± 0.07	0.74
0.06	0.06	0.07	0.82	3.62 ± 0.02	3.56 ± 0.02	5.00 ± 0.03	48.89 ± 0.07	0.71
0.05	0.05	0.08	0.82	3.26 ± 0.02	3.20 ± 0.02	5.64 ± 0.03	48.95 ± 0.07	0.69
0.04	0.04	0.09	0.82	2.91 ± 0.02	2.86 ± 0.02	6.28 ± 0.03	49.02 ± 0.07	0.66
0.04	0.04	0.10	0.82	2.56 ± 0.02	2.52 ± 0.02	6.89 ± 0.03	49.08 ± 0.07	0.64
0.03	0.03	0.11	0.82	2.22 ± 0.02	2.19 ± 0.02	7.50 ± 0.03	49.15 ± 0.07	0.61
0.03	0.03	0.12	0.82	1.89 ± 0.02	1.86 ± 0.02	8.10 ± 0.03	49.21 ± 0.07	0.58
0.08	0.08	0.05	0.80	4.92 ± 0.03	4.83 ± 0.03	3.25 ± 0.02	48.05 ± 0.07	0.47
0.07	0.07	0.06	0.80	4.56 ± 0.03	4.48 ± 0.03	3.91 ± 0.02	48.12 ± 0.07	0.44
0.07	0.07	0.07	0.80	4.20 ± 0.03	4.13 ± 0.03	4.54 ± 0.02	48.19 ± 0.07	0.41
0.06	0.06	0.08	0.80	3.86 ± 0.02	3.79 ± 0.02	5.16 ± 0.03	48.25 ± 0.07	0.39
0.06	0.06	0.09	0.80	3.51 ± 0.02	3.45 ± 0.02	5.78 ± 0.03	48.32 ± 0.07	0.36
0.05	0.05	0.10	0.80	3.18 ± 0.02	3.12 ± 0.02	6.38 ± 0.03	48.38 ± 0.07	0.33
0.05	0.05	0.11	0.80	2.85 ± 0.02	2.80 ± 0.02	6.96 ± 0.03	48.45 ± 0.07	0.30
0.04	0.04	0.12	0.80	2.53 ± 0.02	2.48 ± 0.02	7.54 ± 0.03	48.51 ± 0.07	0.27
0.04	0.04	0.13	0.80	2.21 ± 0.02	2.17 ± 0.02	8.11 ± 0.03	48.57 ± 0.07	0.24
0.08	0.08	0.05	0.78	5.09 ± 0.03	5.00 ± 0.03	3.52 ± 0.02	47.46 ± 0.07	0.16
0.08	0.08	0.06	0.78	4.75 ± 0.03	4.66 ± 0.03	4.13 ± 0.02	47.52 ± 0.07	0.15
0.07	0.07	0.07	0.78	4.41 ± 0.03	4.33 ± 0.03	4.73 ± 0.02	47.59 ± 0.07	0.15
0.07	0.07	0.08	0.78	4.08 ± 0.03	4.01 ± 0.03	5.32 ± 0.03	47.65 ± 0.07	0.15
0.06	0.06	0.09	0.78	3.75 ± 0.02	3.69 ± 0.02	5.90 ± 0.03	47.72 ± 0.07	0.16
0.06	0.06	0.10	0.78	3.43 ± 0.02	3.37 ± 0.02	6.48 ± 0.03	47.78 ± 0.07	0.16
0.05	0.05	0.11	0.78	3.12 ± 0.02	3.06 ± 0.02	7.04 ± 0.03	47.84 ± 0.07	0.17
0.05	0.05	0.12	0.78	2.80 ± 0.02	2.76 ± 0.02	7.59 ± 0.03	47.91 ± 0.07	0.19
0.04	0.04	0.13	0.78	2.50 ± 0.02	2.46 ± 0.02	8.13 ± 0.03	47.97 ± 0.07	0.21
0.04	0.04	0.14	0.78	2.20 ± 0.02	2.16 ± 0.02	8.67 ± 0.03	48.03 ± 0.07	0.24
0.09	0.09	0.06	0.76	5.25 ± 0.03	5.16 ± 0.03	3.76 ± 0.02	46.89 ± 0.07	0.32
0.09	0.09	0.07	0.76	4.93 ± 0.03	4.84 ± 0.03	4.34 ± 0.02	46.96 ± 0.07	0.35
0.08	0.08	0.08	0.76	4.60 ± 0.03	4.52 ± 0.03	4.91 ± 0.03	47.02 ± 0.07	0.38
0.07	0.07	0.09	0.76	4.29 ± 0.03	4.21 ± 0.03	5.48 ± 0.03	47.08 ± 0.07	0.41
0.07	0.07	0.10	0.76	3.98 ± 0.02	3.91 ± 0.02	6.03 ± 0.03	47.15 ± 0.07	0.44
0.07	0.07	0.11	0.76	3.67 ± 0.02	3.61 ± 0.02	6.58 ± 0.03	47.21 ± 0.07	0.47
0.06	0.06	0.12	0.76	3.36 ± 0.02	3.31 ± 0.02	7.11 ± 0.03	47.27 ± 0.07	0.50
0.06	0.06	0.13	0.76	3.07 ± 0.02	3.01 ± 0.02	7.64 ± 0.03	47.34 ± 0.07	0.54
0.05	0.05	0.14	0.76	2.77 ± 0.02	2.73 ± 0.02	8.17 ± 0.03	47.40 ± 0.07	0.57
0.05	0.05	0.15	0.76	2.48 ± 0.02	2.44 ± 0.02	8.68 ± 0.03	47.46 ± 0.07	0.61
0.04	0.04	0.16	0.76	2.19 ± 0.02	2.16 ± 0.02	9.19 ± 0.03	47.52 ± 0.07	0.66
0.10	0.10	0.07	0.74	5.41 ± 0.03	5.32 ± 0.03	3.98 ± 0.02	46.35 ± 0.07	0.70
0.09	0.09	0.08	0.74	5.10 ± 0.03	5.01 ± 0.03	4.53 ± 0.02	46.41 ± 0.07	0.73
0.09	0.09	0.09	0.74	4.79 ± 0.03	4.71 ± 0.03	5.08 ± 0.03	46.48 ± 0.07	0.76
0.08	0.08	0.10	0.74	4.49 ± 0.03	4.41 ± 0.03	5.62 ± 0.03	46.54 ± 0.07	0.79

Continued on next page

Table A.2 – *Continued from the previous page*

P_{SDL}	P_{SDR}	P_{DD}	P_{ND}	σ_{SDL} (mb)	σ_{SDR} (mb)	σ_{DD} (mb)	σ_{ND} (mb)	Δ_{RMS}
0.08	0.08	0.11	0.74	4.19 ± 0.03	4.12 ± 0.03	6.15 ± 0.03	46.60 ± 0.07	0.82
0.07	0.07	0.12	0.74	3.89 ± 0.02	3.83 ± 0.02	6.68 ± 0.03	46.67 ± 0.07	0.86
0.07	0.07	0.13	0.74	3.60 ± 0.02	3.54 ± 0.02	7.19 ± 0.03	46.73 ± 0.07	0.89
0.06	0.06	0.14	0.74	3.31 ± 0.02	3.26 ± 0.02	7.70 ± 0.03	46.79 ± 0.07	0.93
0.06	0.06	0.15	0.74	3.03 ± 0.02	2.98 ± 0.02	8.20 ± 0.03	46.85 ± 0.07	0.97
0.05	0.05	0.16	0.74	2.75 ± 0.02	2.70 ± 0.02	8.70 ± 0.03	46.91 ± 0.07	1.01
0.05	0.05	0.17	0.74	2.47 ± 0.02	2.43 ± 0.02	9.19 ± 0.03	46.97 ± 0.07	1.06
0.10	0.10	0.07	0.72	5.57 ± 0.03	5.47 ± 0.03	4.19 ± 0.02	45.83 ± 0.07	1.10
0.10	0.10	0.08	0.72	5.27 ± 0.03	5.18 ± 0.03	4.72 ± 0.02	45.89 ± 0.07	1.13
0.09	0.09	0.09	0.72	4.97 ± 0.03	4.89 ± 0.03	5.24 ± 0.03	45.96 ± 0.07	1.16
0.09	0.09	0.10	0.72	4.68 ± 0.03	4.60 ± 0.03	5.76 ± 0.03	46.02 ± 0.07	1.19
0.08	0.08	0.11	0.72	4.39 ± 0.03	4.32 ± 0.03	6.27 ± 0.03	46.08 ± 0.07	1.22
0.08	0.08	0.12	0.72	4.11 ± 0.03	4.04 ± 0.03	6.78 ± 0.03	46.14 ± 0.07	1.26
0.07	0.07	0.13	0.72	3.82 ± 0.02	3.76 ± 0.02	7.27 ± 0.03	46.20 ± 0.07	1.29
0.07	0.07	0.14	0.72	3.55 ± 0.02	3.49 ± 0.02	7.76 ± 0.03	46.26 ± 0.07	1.33
0.06	0.06	0.15	0.72	3.27 ± 0.02	3.22 ± 0.02	8.25 ± 0.03	46.32 ± 0.07	1.38
0.06	0.06	0.16	0.72	3.00 ± 0.02	2.95 ± 0.02	8.73 ± 0.03	46.38 ± 0.07	1.42
0.05	0.05	0.17	0.72	2.73 ± 0.02	2.68 ± 0.02	9.20 ± 0.03	46.44 ± 0.07	1.47
0.05	0.05	0.18	0.72	2.46 ± 0.02	2.42 ± 0.02	9.67 ± 0.04	46.50 ± 0.07	1.52
0.11	0.11	0.08	0.70	5.72 ± 0.03	5.62 ± 0.03	4.38 ± 0.02	45.33 ± 0.07	1.51
0.11	0.11	0.09	0.70	5.43 ± 0.03	5.34 ± 0.03	4.90 ± 0.02	45.39 ± 0.07	1.54
0.10	0.10	0.10	0.70	5.15 ± 0.03	5.06 ± 0.03	5.40 ± 0.03	45.45 ± 0.07	1.57
0.09	0.09	0.11	0.70	4.87 ± 0.03	4.78 ± 0.03	5.90 ± 0.03	45.51 ± 0.07	1.60
0.09	0.09	0.12	0.70	4.59 ± 0.03	4.51 ± 0.03	6.39 ± 0.03	45.57 ± 0.07	1.63
0.08	0.08	0.13	0.70	4.31 ± 0.03	4.24 ± 0.03	6.88 ± 0.03	45.63 ± 0.07	1.67
0.08	0.08	0.14	0.70	4.04 ± 0.02	3.97 ± 0.02	7.36 ± 0.03	45.69 ± 0.07	1.71
0.07	0.07	0.15	0.70	3.77 ± 0.02	3.71 ± 0.02	7.83 ± 0.03	45.75 ± 0.07	1.75
0.07	0.07	0.16	0.70	3.50 ± 0.02	3.44 ± 0.02	8.30 ± 0.03	45.81 ± 0.07	1.79
0.07	0.07	0.17	0.70	3.24 ± 0.02	3.19 ± 0.02	8.77 ± 0.03	45.87 ± 0.07	1.84
0.06	0.06	0.18	0.70	2.98 ± 0.02	2.93 ± 0.02	9.23 ± 0.03	45.93 ± 0.07	1.89
0.05	0.05	0.19	0.70	2.72 ± 0.02	2.67 ± 0.02	9.68 ± 0.04	45.99 ± 0.07	1.94
0.05	0.05	0.20	0.70	2.46 ± 0.02	2.42 ± 0.02	10.13 ± 0.04	46.05 ± 0.07	2.00
0.12	0.12	0.09	0.68	5.87 ± 0.03	5.77 ± 0.03	4.57 ± 0.02	44.84 ± 0.07	1.93
0.11	0.11	0.10	0.68	5.60 ± 0.03	5.50 ± 0.03	5.06 ± 0.03	44.90 ± 0.07	1.96
0.11	0.11	0.11	0.68	5.32 ± 0.03	5.23 ± 0.03	5.55 ± 0.03	44.96 ± 0.07	1.99
0.10	0.10	0.12	0.68	5.05 ± 0.03	4.96 ± 0.03	6.03 ± 0.03	45.02 ± 0.07	2.02
0.10	0.10	0.13	0.68	4.78 ± 0.03	4.70 ± 0.03	6.51 ± 0.03	45.08 ± 0.07	2.06
0.09	0.09	0.14	0.68	4.51 ± 0.03	4.43 ± 0.03	6.98 ± 0.03	45.14 ± 0.07	2.09
0.09	0.09	0.15	0.68	4.25 ± 0.03	4.17 ± 0.03	7.44 ± 0.03	45.20 ± 0.07	2.13
0.08	0.08	0.16	0.68	3.98 ± 0.02	3.92 ± 0.02	7.90 ± 0.03	45.26 ± 0.07	2.17
0.08	0.08	0.17	0.68	3.72 ± 0.02	3.66 ± 0.02	8.36 ± 0.03	45.32 ± 0.07	2.22
0.07	0.07	0.18	0.68	3.47 ± 0.02	3.41 ± 0.02	8.81 ± 0.03	45.37 ± 0.07	2.27
0.07	0.07	0.19	0.68	3.21 ± 0.02	3.16 ± 0.02	9.25 ± 0.03	45.43 ± 0.07	2.32
0.06	0.06	0.20	0.68	2.96 ± 0.02	2.91 ± 0.02	9.69 ± 0.04	45.49 ± 0.07	2.37
0.06	0.06	0.21	0.68	2.71 ± 0.02	2.67 ± 0.02	10.13 ± 0.04	45.55 ± 0.07	2.43
0.12	0.12	0.09	0.66	6.02 ± 0.03	5.92 ± 0.03	4.75 ± 0.02	44.37 ± 0.07	2.36
0.12	0.12	0.10	0.66	5.75 ± 0.03	5.65 ± 0.03	5.23 ± 0.03	44.43 ± 0.07	2.39
0.11	0.11	0.11	0.66	5.49 ± 0.03	5.39 ± 0.03	5.70 ± 0.03	44.48 ± 0.07	2.42
0.11	0.11	0.12	0.66	5.22 ± 0.03	5.13 ± 0.03	6.16 ± 0.03	44.54 ± 0.07	2.45
0.10	0.10	0.13	0.66	4.96 ± 0.03	4.88 ± 0.03	6.62 ± 0.03	44.60 ± 0.07	2.48
0.10	0.10	0.14	0.66	4.70 ± 0.03	4.62 ± 0.03	7.08 ± 0.03	44.66 ± 0.07	2.52
0.09	0.09	0.15	0.66	4.44 ± 0.03	4.37 ± 0.03	7.53 ± 0.03	44.72 ± 0.07	2.56
0.09	0.09	0.16	0.66	4.19 ± 0.03	4.12 ± 0.03	7.97 ± 0.03	44.78 ± 0.07	2.60
0.08	0.08	0.17	0.66	3.94 ± 0.02	3.87 ± 0.02	8.41 ± 0.03	44.83 ± 0.07	2.65
0.08	0.08	0.18	0.66	3.69 ± 0.02	3.63 ± 0.02	8.85 ± 0.03	44.89 ± 0.07	2.70
0.07	0.07	0.19	0.66	3.44 ± 0.02	3.38 ± 0.02	9.29 ± 0.03	44.95 ± 0.07	2.75

Continued on next page

Table A.2 – *Continued from the previous page*

P_{SDL}	P_{SDR}	P_{DD}	P_{ND}	σ_{SDL} (mb)	σ_{SDR} (mb)	σ_{DD} (mb)	σ_{ND} (mb)	Δ_{RMS}
0.07	0.07	0.20	0.66	3.20 ± 0.02	3.14 ± 0.02	9.72 ± 0.04	45.00 ± 0.07	2.80
0.06	0.06	0.21	0.66	2.95 ± 0.02	2.90 ± 0.02	10.14 ± 0.04	45.06 ± 0.07	2.86
0.06	0.06	0.22	0.66	2.71 ± 0.02	2.67 ± 0.02	10.57 ± 0.04	45.12 ± 0.07	2.92
0.13	0.13	0.10	0.64	6.17 ± 0.03	6.06 ± 0.03	4.92 ± 0.02	43.90 ± 0.07	2.80
0.13	0.13	0.11	0.64	5.91 ± 0.03	5.81 ± 0.03	5.38 ± 0.03	43.96 ± 0.07	2.82
0.12	0.12	0.12	0.64	5.65 ± 0.03	5.55 ± 0.03	5.84 ± 0.03	44.02 ± 0.07	2.85
0.11	0.11	0.13	0.64	5.39 ± 0.03	5.30 ± 0.03	6.29 ± 0.03	44.08 ± 0.07	2.89
0.11	0.11	0.14	0.64	5.14 ± 0.03	5.05 ± 0.03	6.74 ± 0.03	44.13 ± 0.07	2.92
0.11	0.11	0.15	0.64	4.89 ± 0.03	4.81 ± 0.03	7.18 ± 0.03	44.19 ± 0.07	2.96
0.10	0.10	0.16	0.64	4.64 ± 0.03	4.56 ± 0.03	7.61 ± 0.03	44.25 ± 0.07	3.00
0.10	0.10	0.17	0.64	4.39 ± 0.03	4.32 ± 0.03	8.05 ± 0.03	44.30 ± 0.07	3.04
0.09	0.09	0.18	0.64	4.15 ± 0.03	4.08 ± 0.03	8.48 ± 0.03	44.36 ± 0.07	3.08
0.08	0.08	0.19	0.64	3.90 ± 0.02	3.84 ± 0.02	8.90 ± 0.03	44.42 ± 0.07	3.13
0.08	0.08	0.20	0.64	3.66 ± 0.02	3.60 ± 0.02	9.33 ± 0.03	44.47 ± 0.07	3.18
0.07	0.07	0.21	0.64	3.42 ± 0.02	3.36 ± 0.02	9.75 ± 0.04	44.53 ± 0.07	3.24
0.07	0.07	0.22	0.64	3.18 ± 0.02	3.13 ± 0.02	10.16 ± 0.04	44.58 ± 0.07	3.30
0.07	0.07	0.23	0.64	2.95 ± 0.02	2.90 ± 0.02	10.57 ± 0.04	44.64 ± 0.07	3.36
0.06	0.06	0.24	0.64	2.71 ± 0.02	2.67 ± 0.02	10.98 ± 0.04	44.70 ± 0.07	3.43
0.14	0.14	0.11	0.62	6.32 ± 0.03	6.21 ± 0.03	5.09 ± 0.03	43.44 ± 0.06	3.24
0.13	0.13	0.12	0.62	6.06 ± 0.03	5.96 ± 0.03	5.54 ± 0.03	43.50 ± 0.06	3.26
0.13	0.13	0.13	0.62	5.81 ± 0.03	5.71 ± 0.03	5.98 ± 0.03	43.56 ± 0.06	3.29
0.12	0.12	0.14	0.62	5.56 ± 0.03	5.47 ± 0.03	6.41 ± 0.03	43.62 ± 0.06	3.32
0.12	0.12	0.15	0.62	5.32 ± 0.03	5.22 ± 0.03	6.85 ± 0.03	43.67 ± 0.06	3.36
0.11	0.11	0.16	0.62	5.07 ± 0.03	4.98 ± 0.03	7.28 ± 0.03	43.73 ± 0.07	3.40
0.11	0.11	0.17	0.62	4.83 ± 0.03	4.75 ± 0.03	7.70 ± 0.03	43.78 ± 0.07	3.44
0.10	0.10	0.18	0.62	4.59 ± 0.03	4.51 ± 0.03	8.13 ± 0.03	43.84 ± 0.07	3.48
0.10	0.10	0.19	0.62	4.35 ± 0.03	4.27 ± 0.03	8.54 ± 0.03	43.90 ± 0.07	3.52
0.09	0.09	0.20	0.62	4.11 ± 0.03	4.04 ± 0.02	8.96 ± 0.03	43.95 ± 0.07	3.57
0.09	0.09	0.21	0.62	3.87 ± 0.02	3.81 ± 0.02	9.37 ± 0.03	44.01 ± 0.07	3.63
0.08	0.08	0.22	0.62	3.64 ± 0.02	3.58 ± 0.02	9.78 ± 0.04	44.06 ± 0.07	3.68
0.08	0.08	0.23	0.62	3.41 ± 0.02	3.35 ± 0.02	10.19 ± 0.04	44.12 ± 0.07	3.74
0.07	0.07	0.24	0.62	3.17 ± 0.02	3.12 ± 0.02	10.59 ± 0.04	44.17 ± 0.07	3.80
0.07	0.07	0.25	0.62	2.95 ± 0.02	2.90 ± 0.02	10.99 ± 0.04	44.23 ± 0.07	3.87
0.08	0.08	0.13	0.71	4.04 ± 1.06	3.97 ± 1.04	7.07 ± 1.93	45.98 ± 1.59	