Chapter 1

Introduction

The Large Hadron Collider (LHC) [1], at the CERN Laboratory, the European Laboratory for Particle Physics, outside Geneva, Switzerland, will be completed in 2007. The LHC will be a unique tool for fundamental physics research and will be the highest energy accelerator in the world for many years following its completion. The LHC will provide two proton beams, circulating in opposite directions, at an energy of 7 TeV each (centre-of-mass $\sqrt{s} = 14$ TeV). The CMS experiment [2, 3] is a general purpose detector at the LHC to explore physics at an unprecedented physics energy scale namely that at the TeV scale [4–6]. It is expected that the data produced at the LHC will elucidate the electroweak symmetry breaking mechanism (EWSB) and provide evidence of physics beyond the standard model. CMS will also be an instrument to perform precision measurements, e.g., of parameters of the Standard Model, mainly as a result of the very high event rates, as demonstrated for a few processes in Table 1.1 for a luminosity of $\mathcal{L} = 2 \times 10^{33}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$. The LHC will be a Z factory, a W factory, a b quark factory, a top quark factory and even a Higgs or SUSY sparticle factory if these new particles have TeV scale masses.

Table 1.1: Approximate event rates of some physics processes at the LHC for a luminosity of $\mathcal{L} = 2 \times 10^{33}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. For this table, one year is equivalent to 20 fb^{-1}.

Process	Events/s	Events/year
$W \to e\nu$	40	$4 \cdot 10^8$
$Z \to ee$	4	$4 \cdot 10^7$
$t\bar{t}$	1.6	$1.6 \cdot 10^7$
$b\bar{b}$	10^6	10^{13}
$\tilde{g}\tilde{g}$ (m = 1 TeV)	0.002	$2 \cdot 10^4$
Higgs (m= 120 GeV)	0.08	$8 \cdot 10^5$
Higgs (m= 120 GeV)	0.08	$8 \cdot 10^5$
Higgs (m= 800 GeV)	0.001	10^4
QCD jets $p_{\mathrm{T}} > 200$ GeV	10^2	10^9

The Physics Technical Design Report (PTDR) reports on detailed studies that have been performed with the CMS detector software and analysis tools. The CMS detector and its performance are described in detail in Volume 1 of the PTDR [7], while in the present Volume (Volume 2) the physics reach with the CMS detector is explored.

The CMS detector measures roughly 22 metres in length, 15 metres in diameter, and 12,500 metric tons in weight. Its central feature is a huge, high field (4 Tesla) solenoid, 13 metres in length, and 6 metres in diameter. Its “compact” design is large enough to contain the electromagnetic and hadron calorimetry surrounding a tracking system, and allows a superb muon detection system. All subsystems of CMS are bound by means of the data acquisition and trigger system.

This Volume is organised in two parts. In the first part a number of physics channels challenging for the detector are studied in detail. Each of these channels is associated with certain physics objects, such as electrons, photons, muons, jets, missing E_{T} and so on. The analyses are performed in a fully realistic environment as the one expected for real data. Methods on determining the backgrounds from the data as well as on evaluating the experimental systematic effects, e.g., due to miscalibration and misalignment, resolution and signal significance are developed. In short these analyses are performed imitating real data analyses to the maximum possible extent.

In the second part the physics reach is studied for a large number of physics process, for data samples mostly with luminosities in the range of 1 to 30 fb^{-1}, expected to be collected during the first years of operation at the LHC. Standard model measurements of, e.g., W and top quark mass determinations are studied; many production and decay mechanisms for the SM and MSSM Higgs are studied, and several models Beyond the Standard Model are explored.

1.1 The full analyses

In total 11 analyses were studied in full detail. All the studies were performed with detailed Geant4 based simulation of the CMS detector and reconstruction of the data, including event pile-up, and a detailed analysis of the systematics.

The $H \rightarrow \gamma\gamma$ analysis covers one of the most promising channels for a low mass Higgs discovery and for precision Higgs mass measurement at the LHC. This channel has been an important motivation for the design of the electromagnetic calorimeter (ECAL) of CMS. It is used here as a benchmark channel for identifying photons with high purity and efficiency, and as a driver for optimising the ECAL energy resolution and calibration of the analyses. Furthermore new statistical techniques that make use of event kinematics and neural network event selection algorithms have been used to enhance the sensitivity in this channel.

The analysis $H \rightarrow ZZ \rightarrow 4$electrons covers electron identification and selection optimisation. In particular the classification of electron candidates according to quality criteria which depends on their passage through the material of the tracker was studied, and the impact on the Higgs search quantified.

The same process has been studied in the muon decay channel $H \rightarrow ZZ \rightarrow 4\mu$. This process is an important benchmark for optimising the muon analysis tools. It is one of the cleanest discovery channels for a Standard Model Higgs with a mass up to 600 GeV/c^2. Methods to minimise the systematics errors have been developed.

The channel $H \rightarrow WW \rightarrow 2\mu 2\nu$ is of particular importance if the mass of the Higgs is around 165 GeV/c^2, and is again an interesting muon benchmark channel. The challenge is to establish with confidence a dimuon excess, since this channel does not allow reconstruction of the Higgs mass on an event by event basis. The event statistics after reconstruction and selection is large enough for an early discovery, even with about 1 fb^{-1} of integrated luminosity, provided the systematic uncertainty on the background can be kept well under control.

The production of a new gauge boson with a mass in the TeV range is one of the possible early discoveries at the LHC. The clean final state for the decays into two high p_T leptons leads to a clearly detectable signal in CMS. The channel $Z' \rightarrow \mu\mu$ was selected as a benchmark to study muons with p_T in the TeV/c range. Dedicated reconstruction techniques were developed for TeV muons and the experimental systematics eg. due to misalignment effects were studied in detail.

Jets will be omnipresent in the LHC collisions. The analysis of dijets events and the dijet invariant mass has been studied in detail. A pre-scaling strategy of the jet threshold for the trigger, in order to allow a dijet mass measurement starting from approximately 300 GeV/c^2 has been developed. Calibration procedures, and experimental and theoretical systematics on the dijet mass distribution have been evaluated in detail. The results were interpreted as sensitivities to new physics scenarios.

The determination of the missing transverse momentum in collisions at a hadron collider is in general a difficult measurement, since it is very susceptible to detector inefficiencies, mis-measurements, backgrounds such a halo muons or cosmic muons, and instrumental backgrounds. On the other hand it is probably the most striking signature for new physics with escaping weakly interacting particles, such as the neutralinos in supersymmetry. A low mass mSUGRA SUSY benchmark point was selected to exercise a full analysis, including techniques to suppress spurious backgrounds as well as QCD residual contribution due to mis-measurements. Techniques to calibrate the $E_\mathrm{T}^\mathrm{miss}$ with known Standard Model processes have been also developed. Such a low mass SUSY scenario could already be detected with 0.1 fb^{-1} of data with a well understood detector and well controlled background.

The decay $B_s \rightarrow J/\psi\,\phi$ is chosen as a benchmark channel since it is representative of exclusive B-physics studies. It allows to study the capability of CMS to identify, select and reconstruct a fully reconstructed decay of the B_s, which presents a significant challenge due to its relatively low momentum and high background. In addition, the measurement is performed of the width difference $\Delta\Gamma$ on a sample of untagged $B_s \rightarrow J/\psi\,\phi \rightarrow \mu^+\mu^-K^+K^-$ candidates using a maximum likelihood fit of the time dependent angular distribution.

The detection of the τ particle will be very important at the LHC since a clear excess of τ production is also a sign of new physics. The τ selection and analysis tools have been used to search for and measure the A/H heavy Higgs bosons in the MSSM. Various decay channels of the τ have been considered, and τ tagging tools have been deployed and refined. A τ-trigger is very challenging but necessary for these physics studies, and has been studied in detail.

The process of associated production of a Higgs particle with top quarks, and with the Higgs decaying into b-quarks, is no doubt one of the most challenging channels studied in this part of the TDR. The physics interest is high since this channel gives access to a measurement of the $H \rightarrow bb$ decay and thus to the Yukawa coupling of the Higgs to the b quark. The inclusive $H \rightarrow bb$ production channel cannot be used due to a too large QCD $b\bar{b}$ background. This analysis uses techniques to tag b quarks and calibration methods to reconstruct top quarks from multi-jet decays. Furthermore the backgrounds such as $t\bar{t}$ jet-jet have been carefully examined. The results demonstrate that this will be a very challenging measurement even with the highest luminosity in the first phase of the LHC operation.

Finally a benchmark channel for heavy ions collisions was studied. Quarkonia $(J/\psi, \Upsilon)$ were reconstructed and measured via the two muon decay modes. The particular challenge is an efficient track reconstruction in an environment of 2000 to perhaps even 5000 tracks produced per unit of rapidity. The analysis shows that the detection of the quarkonia is possible with reasonable efficiencies and leads to a good event statistics for detailed studies of the "melting" of these resonances in a hot dense region.

In general these detailed studies in this first part of the PTDR have demonstrated that the CMS experiment is up and ready to meet the challenge, and can deliver measurements with the quality and precision as anticipated from its detector design.

1.2 The physics reach

The physics reach of the Report contains three main parts: Standard Model processes, Higgs searches and measurements and searches beyond the Standard Model.

The Standard Model sections contain a study of the strong interactions, top quark physics and electroweak physics. Jet production is revisited but this time to measure inclusive single jet p_{T} spectra, with emphasis placed on the experimental uncertainties related to such a measurement. The underlying event is still enigmatic, and procedures are outlined to get better insight with the first LHC data. B-hadrons will be copiously produced at the LHC and inclusive B production and B_c production have been studied. At the LHC about one top quark pair is produced per second. Such a huge sample of top quarks allows for detailed measurements of the top quark properties such as cross sections and mass, spin properties, single top production, and searches for new physics in top decays. A detailed study on the mass measurement precision, limited by the systematics errors, is reported. In the electroweak part of this chapter the production of W and Z bosons is discussed, as well as multi-boson production, and a precise measurement of the Drell-Yan process. The precision with which the mass of the W boson can be determined is analysed.

One of the main missions of the LHC is the discovery of the origin of the electroweak symmetry breaking mechanism. Therefore the search for the Higgs particle is a major task for the experiments. The Higgs particle search is studied for the SM and MSSM Higgs(es) in the full mass range starting from the LEP exclusion limits. Detailed systematic studies were included in the estimates for the integrated luminosity needed for a 5σ discovery. The methods used to calculate the 5σ discovery limit are detailed in Appendix A. Over a large range of Higgs boson masses, a discovery is possible with a few fb^{-1}, but for the interesting mass region below 130 $\mathrm{GeV/c^2}$, 10 fb^{-1} will be needed. MSSM Higgs discoveries are studied both for neutral and charged Higgs particles, and discovery regions are presented. Finally the Higgs Chapter also contains studies of other scalar particles such as the radion that emerges in models with warped extra dimensions, and a double charged Higgs that may be produced in Little Higgs scenarios.

The LHC will probe the TeV energy scale and is expected to break new ground. A important part of the CMS program will be to search for new physics. If low mass supersymmetry exists it will be within the reach of the LHC. The studies in this Report are mainly signature based, to test the discovery potential in as many channels as possible, using a number of chosen benchmark points covering a large part of different signatures. The discovery reach for scenarios with extra dimensions, and new vector bosons high mass states are analysed using several different experimental signals. The methods used to calculate the 5σ discovery limit are detailed in Appendix A. Finally alternative signatures for new physics such as technicolour, contact interactions, heavy Majorana neutrinos, heavy top in Little Higgs models, and same sign top quarks have been analysed.

While many signals and processes have been studied, it was not the goal of this PTDR to study and to include all possible channels to give a full physics review. Besides what is contained here in this Report there are other ongoing analyses nearing completion on topics such as GMSB SUSY, UED extra dimensions, split SUSY scenarios, invisible Higgs production, TGC sensitivity of dibosons, strongly interacting vector boson scattering, and others. The channels included in this Report have however been very instrumental to test and deploy the tools and techniques for performing physics studies with CMS at the LHC.

1.3 Tools used in the studies for the PTDR

1.3.1 Detector simulation and reconstruction

For the studies presented in this TDR, the CMS detector response was simulated using the package OSCAR [8]. It is an application of the Geant4[9] toolkit for detector description and simulation. OSCAR is used to describe the detector geometry and materials. It also includes and uses information about the magnetic field. OSCAR reads the individual generated events and simulates the effects of energy loss, multiple scattering and showering in the detector materials with Geant4. The digitisation (simulation of the electronic response), the emulation of the Level-1 and High-Level Triggers (HLT), and the offline reconstruction of physics objects were performed with the CMS full-reconstruction ORCA package [10].

A number of analyses for the physics reach studies were performed with the fast parameterised simulation FAMOS [11]. FAMOS has been tuned to the detailed simulation and reconstruction and is roughly about a factor 1000 faster. FAMOS allows to perform, e.g., accurate sensitivity scans in a large parameter space of a model for new physics.

In the CMS coordinate system the origin coincides with the nominal collision point at the geometrical centre of the detector. The z direction is given by the beam axis. The rest frame of the hard collision is generally boosted relative to the lab frame along the beam direction, θ is the polar angle with respect to the z axis and ϕ the azimuthal angle with respect to the LHC plane. The detector solid angle segmentation is designed to be invariant under boosts along the z direction. The *pseudorapidity* η, is related to the polar angle θ and defined as $\eta \equiv -\ln(\tan(\theta/2))$. The transverse momentum component z-axis is given by $p_{\mathrm{T}} = p\sin\theta$ and similarly $E_{\mathrm{T}} = E\sin\theta$ is the transverse energy of a physics object.

1.3.2 Pile-Up Treatment

The total inelastic cross section at the LHC is assumed to be $\sigma_{\mathrm{T}} \sim 80\,\mathrm{mb}$. The LHC will operate at a bunch crossing rate of 40 MHz. Only 80 % of the bunches will be filled , resulting in an effective bunch crossing rate of 32 MHz. The instantaneous luminosity in the first two years after start–up is expected to be $\mathcal{L} = 2 \times 10^{33}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ and subsequently upgraded to $\mathcal{L} = 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ in a second phase. The average number of inelastic non-diffractive interactions per bunch crossing μ is $\mu = 25$ at high and $\mu = 5$ at low luminosity.

Both the detailed simulation and reconstruction chain OSCAR/ORCA and FAMOS allow the overlay of pile-up events, according to a Poisson distribution with average μ, on top of real signal events, exactly as for real data. These events were sampled from a data base of 600K minimum bias events, generated with parameters discussed in Appendix C.

All the studies reported in this TDR include the effects of pile-up on the signal. For all studies with luminosities up to $60\,\mathrm{fb}^{-1}$ $\mu = 5$ was used. Several techniques have been developed to minimise the effect of pile-up, and have been used in the studies reported in this TDR. Both in-time and out-of-time pile-up has been included.

1.3.3 Systematic effects on measurements

The results of the PTDR Volume 1 were used to form the baseline for all systematic studies in this Volume. Systematic effects include energy scale uncertainties for the calorimeters, effects of misalignment, uncertainties in the background estimation either from theory or from techniques to estimate these backgrounds from data. Misalignments of the tracker and of the muon system expected at the initial and at the well-advanced stages of the data taking have been taken into account by using two misalignment scenarios developed in the framework of the CMS reconstruction.

A comprehensive review on the experimental and theoretical systematics used in this PTDR is presented in Appendix B.

1.3.4 Event generators

The studies for this physics TDR have been performed with a variety of event generators, suitably chosen for each processes studied. The main work-horse was PYTHIA, the general multi-purpose generator, and in some case checks have been performed with HERWIG. More specialised generators which include a more complete description of the relevant matrix elements, have been used for a number processes, as detailed in the analysis reports. A list of generators used in this TDR is given in Appendix C.

An important aspect for the LHC, is the QCD multi-jet production in various physics channels, and a correct and thorough understanding of Standard Model processes such as W+jets, Z+jets and $t\bar{t}$ + jet production will be paramount before discoveries can be claimed in channels such as jets + $E_{\mathrm{T}}^{\mathrm{miss}}$ and jets + leptons. CMS will measure these Standard Model processes in an early phase of the experiment, to reduce the impact of inherent uncertainties in the Monte Carlo models on searches and discoveries, using methods demonstrated in this TDR. These will allow estimation of the expected backgrounds directly or will allow to tune the generators in order to use these with increased confidence in regions of phase space not directly accessible with measurements from the data.

Generators with multi-parton final states are available at Leading Order (LO) for most Standard Model processes. Recently NLO generators have become available as well, be it for a more restricted number of processes. Sophisticated algorithms that match the hard jets generated by the matrix elements, with the softer parton jets, have become available. An example is the ALPGEN generator, which has been used for some studies and comparisons in this Report. For some of the detailed analyses, such as the $E_{\mathrm{T}}^{\mathrm{miss}}$ low mass SUSY search, it was shown that the effect of using ALPGEN instead of PYTHIA did not lead to different result, while for other analyses, such as background to ttH production, the difference was important.

Another difficulty in the estimation of the background to processes is the rate of QCD multi-jet events. Typically samples of events of more than 10^8 or 10^9 events would be needed to cover possible tails. Detailed simulation of such background samples cannot be easily done, and therefor other approaches were taken in this TDR. These include pre-selections at the generator level, fast simulation of large samples and factorising the efficiencies of independent selections cuts.

Hence one has to keep in mind that the exact results presented in this TDR could depend on the generators. They should therefore be taken as an indication albeit a good indication of what can be expected at the LHC.

1.3.5 Parton distributions and higher order corrections

One of the key differences between a hadron and an e^+e^- collider is that for hadrons the partons collide with a strongly varying incident energy, given by the distribution of the longitudinal momentum fraction x of the parton in the proton. These parton densities are determined from data, in particular from deep inelastic scattering data and other measurements of hard scattering processes. Several groups have fitted parton distribution functions (PDFs) to these data, e.g., the CTEQ [12] and MRST [13] groups.

For the studies in this report, the simulated event samples were generated with CTEQ5L but CTEQ6 was used to normalise cross sections and to study the PDF uncertainties. CTEQ 6.1 has 40 different error PDFs, 20 PDFs at positive error, and 20 PDFs at negative error. We use the CTEQ6.1M eigenvector PDF sets [12] and the "master" equations as detailed in Appendix B to evaluate the uncertainties characterising current knowledge of the parton distributions.

The precise knowledge of the parton distributions will remain an extremely important subject for the physics at the LHC. Currently a study group in the framework of the HERA-LHC workshop is tackling this topic in order to get as good knowledge as possible of the PDFs[14] and their uncertainties at the time of the startup of the LHC. Once the LHC starts data collection, several QCD process can be used to help to constrain the PDFs, as has been shown, e.g., using W production with studies at the HERA-LHC workshop.

1.4 Outlook

The work detailed in this Volume of the PTDR constitutes the pedestal for the physics studies that the experiment will pursue both at the start-up and the longer term running. In the process of carrying out these studies CMS has gained valuable experience in all aspects, both technical and strategic, in executing a high performance physics program. Of great value is also the identification of shortcomings and challenges that emerged in the context of completing these analyses.

As a follow-up of this work CMS is planning an elaborate program for the start-up studies and physics commissioning from the combined magnet test effort (MTCC) as well as the experience of the upcoming computing, software and analysis challenge (CSA06) that incorporates the full calibration and alignment framework in combination with the full-trigger path exercise. The whole edifice for data collecting and analysis is expected to be complete and tested by the turn-on of the LHC in 2007.

Part I

Complete Analyses