Towards a search for Sphalerons in leptonic channels with CMS experiment

von Benjamin Burkic

Bachelorarbeit in Physik

vorgelegt der

Fakultät für Mathematik, Informatik und Naturwissenschaften
der RWTH Aachen

im Oktober 2021

angefertigt im

III. Physikalisches Institut A

bei

Prof. Dr. Thomas Hebbeker

Abstract

This thesis presents an analysis to search for electroweak Sphalerons in proton-proton collisions. The procedure with the generator, as well as with the reconstructed data, is examined with a centre-of-mass energy of $\sqrt{s}$=13 TeV. Therefore the work deals with the particle distributions and with the invariant mass of the leptons in final states eμ, eτ and $\mu\tau$ in different parton distribution function sets. In addition, the jets are also taken into consideration and analyzed to what extent they play a role in the final states. Furthermore, the thesis concerns itself with the sum of the transversal impulses and the sphericity. The result of the thesis is that the jets cannot simply be neglected.

Kurzdarstellung

Diese Bachelorarbeit präsentiert eine Analyse zur Suche nach elektroschwachen Sphalerons in Proton-Proton Kollisionen. Die Vorgehensweise mit dem Generator, sowie mit den rekonstruierten Daten, wird bei einer Schwerpunktsenergie von $\sqrt{s}$=13 TeV untersucht. Deshalb beschäftigt sich die Arbeit um die Teilchen Verteilungen und um die invariante Masse der Leptonen in Endzustände eμ, eτ und $\mu\tau$ in unterschiedlichen Parton-Verteilungsfunktionssätze. Dazu werden noch die Jets mitberücksichtigt und analysiert inwiefern sie eine Rolle für die Endzustände spielen. Des Weiteren beschäftigt sich die Arbeit um die Summe der transversale Impulse und die Sphärizität. Das Ergebnis der Bachelorarbeit ist, dass die Jets nicht einfach zu vernachlässigen sind.

Contents

1 Introduction

The Sphaleron is not "new physics", but a solution within the Standard Model and it plays a role in the calculation of the rate of processes that violate baryon and lepton numbers. It cannot be described in terms of perturbation theory within the framework of the quantum field theory of the Standard Model. The Sphaleron processes could not yet be confirmed experimentally. Now that the Higgs boson is known to have a mass of 125 GeV, the Sphaleron energy is estimated to be around 9 TeV. It is generally believed that the process of Sphaleron production is exponentially suppressed during particle collisions. There are ways to see Sphaleron at the processes in LHC, even though the required energies are relatively high.
A previous search for Sphalerons in CMS made use of the large predicted particle multiplicities and heavily relied on jets [1]. Other analyses make use of lepton flavour violation in final states with at least two leptons, but so far not in the context of Sphalerons [2], [3]. It is therefore interesting to look into leptonic channels, because one has the advantage that the leptons can be well reconstructed later with the help of the algorithmic flow of the CMS reconstruction software. The motivation for analyzing the invariant mass of different lepton pairs is that Sphaleron processes explicitly violate lepton flavour. The idea of the thesis is to do a study to see how the Sphalerons behave in this area and to look at the properties more closely on a fundamental level, what one would actually expect in a search.
The thesis is structured as follows: Section 2 briefly describes the Standard Model and the theory of Sphalerons, which are important for the analysis. Section 3 explains the experimental setup in detail. Section 4 deals with the Monte Carlo generator and the respective settings. The generated files are discussed as well as an explanation of what parton distribution functions (PDF) are and which are used for the analysis. The main part of the work is described in Section 5. The generator output is analyzed in terms of the particle distribution, the invariant mass and the sum of transverse momentum. Later the variables are examined for different PDF sets. Building on this, first preliminary distributions on the reconstructed level and finally the sphericity are examined.

2 Theoretical Introduction

In this chapter, the physical fundamentals that are needed for the present analysis are discussed. For this purpose, the first Section 2.1 provides a greatly reduced owerview of the standard model (SM) of particle physics. The second Section 2.2 describes in more detail the Sphaleron's theory.
An overview of the fundamental forces and the elementary particles of the SM are introduced. Natural units ($\hbar = c = 1$) are used for simplification [4]. The following chapter is mostly based on [4], [5] and [6].

2.1 Standard model of particle physics

The so-called standard model (SM) is the fundamental theory to describe all interactions in particle physics and it describes the elementary particles.
The fundamentals of SM are made up of fermions, which are divided into six different flavors of quarks and leptons, and made up of bosons, four different types of gauge bosons as force mediators, and the Higgs boson. The fermions are split into three generations. A basic overview on all known elementary particles is given in Figure 1.

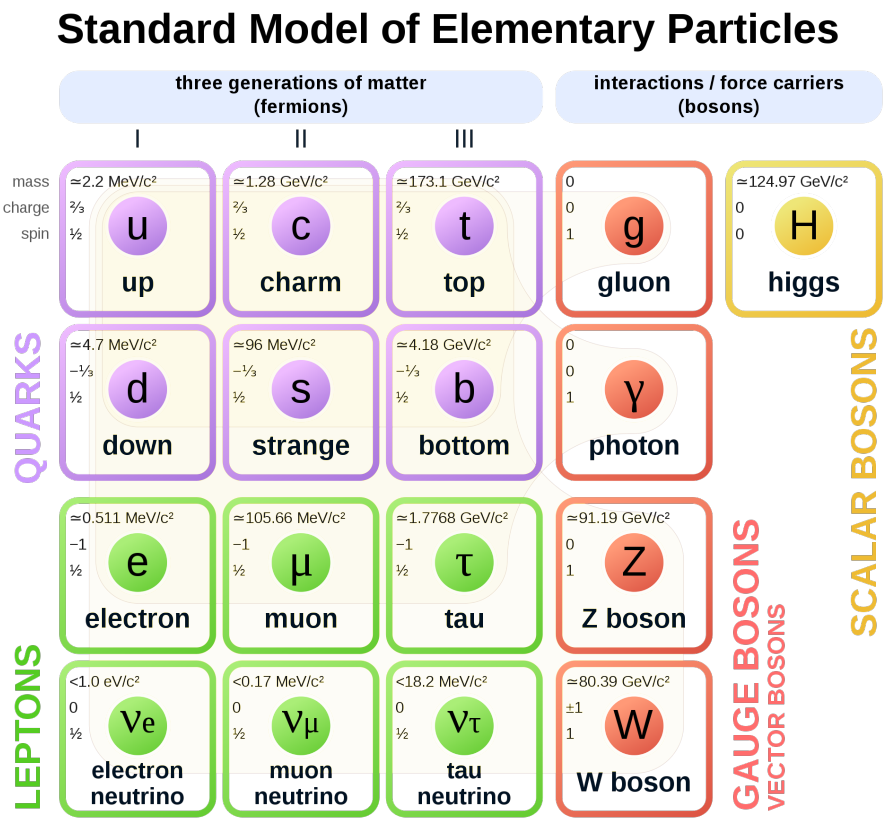


Figure 1: The standard model of particle physics [7]

2.1.1 Matter Particles

All matter particles can be seperated into two kinds of particles, fermions and bosons, as seen in Figure 1. Fermions are particles with a half odd integer spin ($s = \frac{1}{2}$) and can be separated into six leptons and six quarks. The six different flavors of quarks carry color and electric charge and participate in the strong interaction.
The first generation consists of up and down quarks, the second of the charm and strange quarks and the third of bottom and top quarks. The first quark in each generation has an electric charge $+\frac{2}{3}$, the counterpart has an electric charge of $-\frac{1}{3}$ [8]. The three lepton generations are electrons, muons and taus. Each lepton generation includes one neutrino. According to SM, neutrinos are massless. In the meantime there are clear indications that the mass of the neutrinos is very small [5].

2.1.2 Interactions

The gauge bosons of the SM are mediators for the three different types of particle interaction: strong, electromagnetic and weak. Gravitation is not part of SM. All gauge bosons are bosons with spin 1, whereas the Higgs boson is a scalar boson with spin 0. The strong interaction is transmitted by massless, colored gluons (g) and affects all colored particles, the quarks. There are a total of eight different gluons. The $W^{\pm}$-bosons and Z bosons mediate the weak interaction and all quarks and leptons are affected by it. Neutrinos only interact weakly. The electromagnetic interaction is mediated by the massless photon (γ), which has no electric charge. It is subject to all electrically charged particles, so quarks, charged leptons and the $W^{\pm}$-bosons. In contrast to the massless photons and gluons, $W^{\pm}$-bosons and Z bosons are massive.
The gauge field bosons are not classified into families, but can be sorted by their transmitted interaction, see Table 1.

Force	Gauge boson	Charge [e]	Mass [GeV]
strong	Gluons (g)	0	0
electromagnetic	Photon (γ)	0	0
weak	W bosons ($W^{\pm}$)	± 1	80,379
weak	Z bosons (Z)	0	91,187

Table 1: Presented are the force mediating gauge bosons. All values taken from [8]

2.1.3 Higgs-Mechanism

The Higgs boson is the latest addition in the Standard Model. It has a spin of 0 and is a scalar boson. The interaction of matter with the Higgs field, which can be formally defined by the so-called Higgs mechanism, gives matter its mass. In March 2013, the Higgs boson was confirmed by CERN [9]. Since then, there have been further efforts

to fathom the nature of the Higgs boson and its properties through targeted scattering experiments.

2.2 Sphaleron Processes

A Sphaleron is a high-energy and unstable, static solution to the electroweak interaction of the SM, which decays immediately. The Sphaleron process cannot be described by perturbation theory within the framework of the quantum field theory of the SM. That means a Sphaleron is non-pertubative and it cannot be illustrated trough a Feynman diagram. Sphalerons have not yet been confirmed experimentally.

A Sphaleron can be described as a transition between two vacuum states. The vacuum states of the electroweak theory are not unique in the SM and the potential has a periodic structure of minima. It leads to a nontrivial vacuum structure with an infinite number of ground states. These ground states can be enumerated by the so-called Chern-Simons number N_{CS} [10]. The Sphaleron solutions exist at the local maxima, where N_{CS} is a half-integer, while the local minima are at integer values of N_{CS}. If a system changes from one vacuum state to the other, this is done by Sphalerons and they take their way across the potential barrier ($E_{sph} \simeq 9\,\text{TeV}$) [10] between the vacuum states. The Figure 2 illustrates the periodic potential with the ground states.

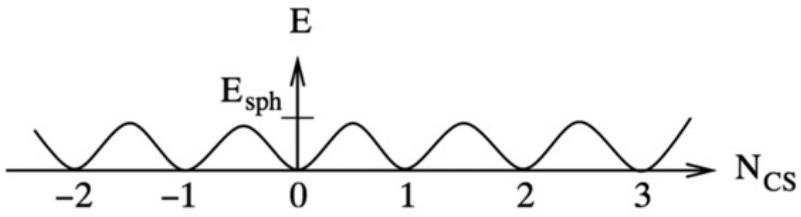


Figure 2: The periodic Sphaleron potential [11]

Another property of Sphaleron transitions is the violation of the baryon number B and lepton number L. Sphalerons preserve the difference between the baryon number and the lepton number (B-L), but violate (B+L). Sphaleron transitions are associated with a change in the Chern-Simons number N_{CS}, which changes both the baryon number B and the lepton number L by 3 ΔN_{CS}. This factor 3 is caused by a change in the lepton number for each lepton doublet by ΔN_{CS}. The same is true for quarks, as the baryon number changes by ΔN_{CS} for each of the three quark generations [10]. The following equations show the important relations:

$$\Delta(B - L) = 0 \tag{1}$$

$$\Delta B = 3 = \Delta L \tag{2}$$

$$\Delta(B+L) = 6 \cdot \Delta N_{CS} \tag{3}$$

Sphaleron-induced transitions in proton-proton collisions can be mediated such as in equations 4 and 5 below:

$$u + u = e^{+}\mu^{+}\tau^{+}\overline{ttt}\bar{c}\bar{c}\bar{c}\bar{d}\bar{d}\bar{d} + uu + X(\Delta N_{CS} = -1) \tag{4}$$

$$u + u = \nu_e\mu^{-}\nu_\tau tbbcscddu + uu + X(\Delta N_{CS} = +1) \tag{5}$$

where X is a set of particles that has B=L=0 [10]. The ΔN_{CS} =+1(-1) Sphaleron transitions are fundamentally composed of 12 (anti)fermions: three (anti)leptons, one from each generation, and nine (anti)quarks, corresponding to three colors and three generations.

3 Experimental Setup

This chapter describes the experimental setup which forms the basis for the presented analysis. It will explain how the Large Hadron Collider (LHC) and the Compact Muon Solenoid (CMS) work, as well as how collisions and detection are processed. The information in this section was taken from [12] and [13], if it is not indicated otherwise.

3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) is a hadron accelerator and collider. It is the world's largest particle collider. It has a radius of 27 km and was built in the already existing tunnel of the Large Electron-Positron collider (LEP) at CERN, the European Organization for Nuclear Research. The LHC is a particle-particle collider, therefore, it has two rings with beams that rotate in the opposite direction.
The particles are pre-accelerated by one linear and three circular accelerators before entering the LHC. Protons (or heavy ions) are accelerated as bunches in two separate beam pipes, one clockwise and one counterclockwise, up to the required beam energy of (in Run II) 6.5 TeV each. These protons collide at four different places around the LHC ring. The experiments are ATLAS, CMS, ALICE and LHCb [14]. The whole acceleration process is depicted in figure 3.

3.2 The Compact Muon Selonoid experiment

The task of the CMS detector is to record the physical dimensions of the collisions that the LHC delivers. With the help of the recorded data, the energy, momentum, type and charge of the particles can be determined after the electrical signals have been reconstructed. The detector is composed of various components with different tasks in an onion-shaped structure. Overall, the detector is cylindrical, has a length of 21 m, a mass of 14,000 t and a diameter of 15 m [16]. For this purpose, it is mounted rotationally

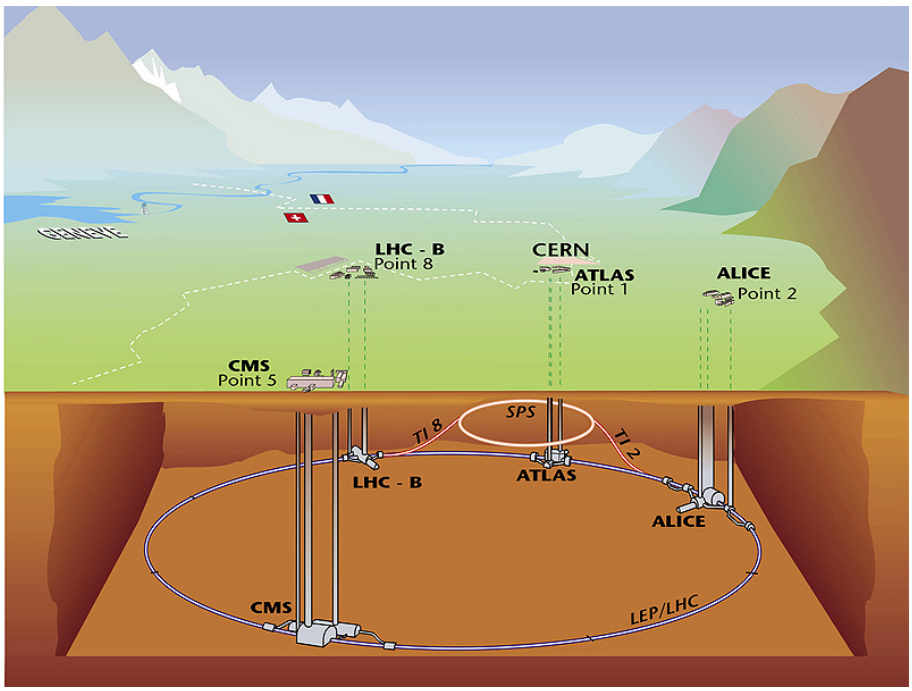


Figure 3: Overview: LHC and experiments [15]

symmetrical around the beam axis of the LHC. Figure 4 shows a schematic sectional view perpendicular to the beam axis of the detector. The most important components are presented in the following sections.

3.2.1 Solenoid

A high magnetic field can be used to determine transverse momentum because charged particle trajectories are curved as they pass through the magnetic field depending on their momentum, causing a Lorentz drift. Furthermore, the sign of the particle's charge becomes visible via the direction in which its trajectory is bent. A coil of superconducting wire forms the powerful solenoid magnet, which generates a homogeneous magnetic field designed for 4T [13].

3.2.2 Tracker

The tracker is the innermost component of the detector, positioned around the beamline. It is mainly composed of a pixel detector with three barrel layers and a silicon strip tracker with ten barrel layers. This subdetector's goal is to determine the primary interaction vertex and to measure their trajectory. The silicon tracker provides three-dimensional information on the trajectories of the particles via the tracker by combining signals from the pixel and the strip section.

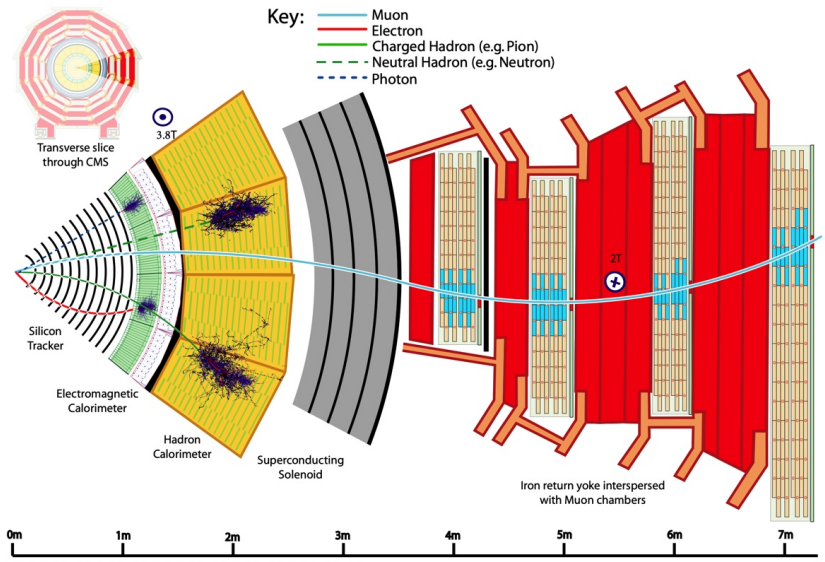


Figure 4: CMS detector [17]

3.2.3 Electromagnetic Calorimeter

The electromagnetic calorimeter (ECAL) is constructed around the tracker to measure the electromagnetic energy carried by the particles hitting the detector. It is a homogeneous calorimeter constructed of lead tungstate crystals ($PbWO_4$) that function as both an absorber and an active material. Electrons, positrons, and photons produced by collisions or subsequent decays are trapped in the crystal and deposit their energy there. Photodiodes in the barrel and -triodes in the endcaps detect the scintillation light.

3.2.4 Hadronic Calorimeter

The hadronic calorimeter (HCAL) is the next layer for particles to travel through after the ECAL for absorption of the hadronic energy of jets (a jet is a narrow cone of hadrons and other particles produced by the hadronization of a quark or gluon [18]) emerging from high energetic hadron collisions. The HCAL is a sample calorimeter made up of alternating layers of plastic scintillator to measure the energy of the particles and brass or steel to slow them down. After passing through the absorber material, the hadrons generate secondary particles which ultimately end up creating optical photons in the scintillator material, allowing the hadrons' energy to be measured. Accordingly, the HCAL in principle measures the strongly interacting hadrons such as the ECAL electrons and photons.

3.2.5 Muon System

Muons are able to traverse through the whole inner detector without being absorbed or depositing all of their energy. As a result, muon detection occurs in the solenoid's

iron return yoke's outermost subdetector. The muon system is made up of drift tubes (barrels), cathode strip chambers (endcaps), and resistive plate chambers. The bending angle and energy of muons are measured by this subdetector device.

3.3 Important Quantities/Formulas

In particle physics, the transverse momentum $p_T(=p_\perp)$, ϕ and y describe the kinematics of a particle, which is completely determined by these three independent variables with a given mass. ϕ is the azimuthal angle and y is the rapidity of the particle. The quantities p_T and ϕ are each determined in the plane perpendicular to the scattering axis. In the following, the scattering axis corresponds to the z-axis of the coordinate system. p_z is particle's momentum along the sphericity or thrust axis. The rapidity is defined as [19]:

$$y = \frac{1}{2}\ln(\frac{E + p_z}{E + p_z}) \tag{6}$$

For energies (E) much larger than the rest mass (m) of the particle (E $\gg$ m), y can be determined by the pseudorapidity

$$\eta = -\ln(\tan(\frac{\theta}{2})) \tag{7}$$

where θ is the scattering angle of the particle (in the laboratory system). In the high-energy approximation, the pseudorapidity is numerically roughly equal to the rapidity. M is called the invariant mass and occurs frequently in the analysis

$$M = \sqrt{(\Sigma_i E_i)^2 - (\Sigma_i \vec{p_i})^2} \tag{8}$$

The position vector or momentum vector is broken down into a component parallel to the direction of the transformation speed and one perpendicular to it. The vertical component remains unchanged and the parallel one transforms like the z-component

$$p_T = \sqrt{p_x^2 + p_y^2} = p_\perp \qquad p_z = p_\parallel \tag{9}$$

The angular distance is often specified in a form that is invariant under longitudinal boosts at hadron colliders.

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} \tag{10}$$

In a collision experiment, the number N of scattering events that can be measured is given by the cross section and the integrated luminosity [19]

$$N = \sigma \mathcal{L}_{int} \tag{11}$$

The cross-section characterizes the collision of two particles, while the luminosity describes how many particles actually meet on a certain cross-sectional area. For example, the luminosity of a circular accelerator is given by

$$\mathcal{L} = N_B f \frac{N_1 N_2}{4\pi\sigma_x\sigma_y} \tag{12}$$

Here σ_x and σ_y are the widths of the distribution of the particles transverse to the beam direction, which are approximately Gaussian distributions. The number of particles per bunch, $N_{1,2}$, multiplied by the number of bunches, N_B, and the orbital frequency of the bunches, f, is the flow of one of the particle beams in the machine.
The transverse sphericity S_T is defined in terms of the linearised version of the transverse momentum tensor's eigenvalues $\lambda_1 > \lambda_2$ [20]:

$$S_{XY}{}^{L} = \frac{1}{\Sigma_i p_T^{(i)}} \Sigma_i \frac{1}{p_T^{(i)}} \begin{pmatrix} {p_x^{(i)}}^2 & p_x^{(i)} p_y^{(i)} \\ p_y^{(i)} p_x^{(i)} & {p_y^{(i)}}^2 \end{pmatrix} \tag{13}$$

The transverse sphericity is given by:

$$S_T = \frac{2\lambda_2}{\lambda_2 + \lambda_1} \tag{14}$$

It is used to describe the geometric shape of collision events and, for example, to filter out jet events or to indicate the proportion of transverse impulses.

4 Signal Generation

This section is concerned with Monte Carlo simulations in order to make studies with new physics models to see what one expects and to compare it with measured data. The simulation calculations consist of algorithms that numerically solve complex problems with the help of probability theory.

For this study the BARYOGEN v1.0 generator [21] is used. How this generator works exactly is explained in the Section 4.2. In addition, the respective parameters that can be set are varied and the results are analyzed. The analysis environment CMSSW (CMS software) is available for investigations into the CMS experiment [22].

The BaryoGEN code is available online on a github repository [23]. BaryoGEN is an event generator for proton-proton collisions with Sphaleron-like transitions. BaryoGEN is relevant for any model that violates baryon and lepton numbers via a $\Delta N_{CS} = \pm 1$ vacuum transition caused by a chiral anomaly, including beyond the standard model (BSM) searches at LHC energies [21].

4.1 Les Houches Event Files

BaryoGEN generates Les Houches Event (LHE) files and the generator interfaces to general-purpose tools such as PYTHIA [24] via LHE-files. It is commonly used to transfer information from matrix-element-based generators to general-purpose generators in order to generate complete events for a multitude of processes [25]. In the following source code one can see how such an LHE-file is structured. Only one event was shown here as an example out of several thousand events. The number of events can be set in the generator, further details follow in the Section 4.2.

```
<LesHouchesEvents version="1.0">
<header>
</header>
<init>
  2212 2212 6.500000e+03 6.500000e+03 292200 292200 292200 292200 3 1
  7.3 1.0 1.0 7000
</init>
<event>
  19 7000 1 9.294677e+03 7.3e-03 0.118
  2 -1  0 0 501 0 0.000000e+00  0.000000e+00  4.175184e+03  4.175184e+03  2.299171e-03  0 9
  1 -1  0 0 502 0 0.000000e+00  0.000000e+00  -5.172887e+03 5.172887e+03  4.800484e-03  0 9
  1000022 2 1 2 0 0 -6.112151e+02 9.740626e+02  -1.939754e+03 2.391965e+03  7.977900e+02  0 9
  6 1 3 3 506 0 -3.109065e+01 5.059065e+02  -3.640676e+02 6.476882e+02  1.733400e+02  0 9
  5 1 3 3 508 0 -1.438135e+01 2.230745e+00  -3.566010e+01 3.874164e+01  4.180000e+00  0 9
  5 1 3 3 510 0 -5.657431e+02 4.659254e+02  -1.540026e+03 1.705535e+03  4.180000e+00  0 9
  1006213 2 1 2 0 0 -1.816062e+02 5.087116e+02  -6.043770e+02 1.565162e+03  1.338915e+03  0 9
  3 1 7 7 504 0 1.587300e+02  5.570749e+01  -7.209208e+02 7.402874e+02  9.500000e-02  0 9
  4 1 7 7 507 0 -3.158152e+02 5.879946e+02  1.490259e+02  6.838766e+02  1.275000e+00  0 9
  4 1 7 7 509 0 -2.452099e+01 -1.349905e+02 -3.248199e+01 1.409980e+02  1.275000e+00  0 9
  1006213 2 1 2 0 0 4.116484e+02  -9.329842e+02 9.883565e+02  2.948771e+03  2.584277e+03  0 9
  2 1 11  11  503 0 1.340318e+03  -3.161213e+02 -7.615813e+01 1.379197e+03  2.299880e-03  0 9
  1 1 11  11  505 0 -8.753997e+02 -6.501954e+02 1.001045e+03  1.480260e+03  4.800338e-03  0 9
  2 1 11  11  511 0 -5.326941e+01 3.333255e+01  6.346974e+01  8.931459e+01  2.300001e-03  0 9
  15  1 1 2 0 0 -4.058610e+02 1.657342e+00  -1.298864e+01 4.060761e+02  1.776800e+00  0 9
  14  1 1 2 0 0 1.299738e+02  4.343532e+02  -3.889085e+01 4.550477e+02  -1.078959e-05 0 9
  11  1 1 2 0 0 4.573028e+02  -7.140506e+02 8.223821e+02  1.181231e+03  5.120228e-04  0 9
  2 1 1 2 501 0 1.980306e+02  -2.707418e+02 -2.134332e+02 3.975814e+02  2.300019e-03  0 9
  1 1 1 2 502 0 1.726752e+00  -1.008163e+00 1.001733e+00  2.236416e+00  4.800000e-03  0 9
</event>
```

The generator creates several events that are subdivided between the lines of code <event> and < /event>, as can be seen in the source code above. Color flow lines must be drawn and enumerated in order to use PYTHIA to decay and hadronize the hard generated processes. PYTHIA can only handle one baryon number violation at a given vertex, and because Sphaleron-induced transitions violate the baryon number by 3, the transition is factorized by introducing placeholder mediators in the LHE-file [10]. The particle numbers 1-8 represent the quarks, 11-18 the leptons [26]. See also Tables 3 and 4.
After cancellations, the outgoing quarks are first categorized by generation. Figure 5 shows an example of a graph representing an event in an outgoing LHE file. This transition vertex has two incoming quarks, outgoing leptons, outgoing uncanceled incoming quarks, and a fake mediator for each generation of outgoing quarks.

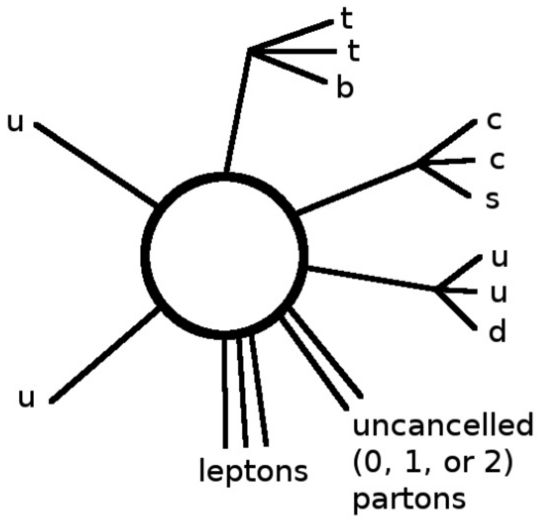


Figure 5: Sphaleron-induced transition

The graphic in Figure 5 is a representation of the event as written to the LHE file to allow PYTHIA to correctly determine color flow for decay and hadronization [10].

4.2 Using the Generator

The following syntax is used to run the program of the generator [10]:

./BaryoGEN sqrtS threshold Nevents pNCS bCancel Filename

If the number of parameters is not entered correctly, an error message will result. The following points explain the individual arguments.

- **sqrtS**: the proton-proton center-of-mass energy in TeV

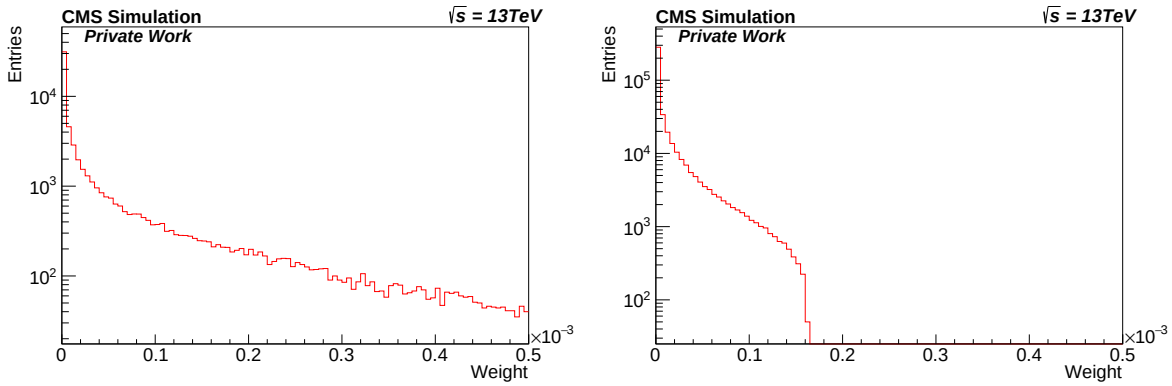
$\sqrt{s}$=13 TeV is used for the whole analysis. This parameter is not changed.

- **threshold**: the minimal energy required for a transition in GeV

The minimum Sphaleron energy E_{Sph} is kept at 9 TeV at the beginning of the analysis and changed to study in different masses at 8 TeV and 10 TeV.

- **maxweight**: is a parameter with a value that is greater than any possible probability given by the Parton Distribution Function (PDF) set used

PDF sets are introduced and discussed in more detail in the Section 4.3. Some Monte Carlo generators generate events with associated weights, and the model estimate should be the sum of the weights rather than the number of events found [27]. The maximum weight should be high enough for the sample produced. $5{\cdot}10^{-4}$ was used as the value for the generated samples for any settings. However, if the Sphaleron energy was reduced to 8 TeV, then the maxweight had to be corrected. It does not matter for the maxweight which PDF set one chooses or **pNCS** varies. This is shown in Figure 6 as an example for the energies 8 TeV and 9 TeV. In the left Figure 6 in a) for 8 TeV one can see that one runs out of the range and not for 9 TeV. Increasing the maxweight value leads to an increased calculation time of the generator.



(a) E_{Sph}=8 TeV (b) E_{Sph}=9 TeV

Figure 6: All variables are identical only the E_{Sph} was changed

- **Nevents**: Number of events produced by the generator

The number of analyzed events was between 10000 and 100000.

The next two points below are options for configuring the types of output events.

- **pNCS**: the probability of events to have $\Delta N_{CS} = +1$
- **pCancel**: parton cancellations

The parameter **pNCS** indicates the probability of events to have $\Delta N_{CS} = +1$. The probability of getting $\Delta N_{CS} = +1$ is one minus this probability. More details in the Section 5.1.1.
The generator also allows to disable parton cancellations by setting **pCancel** to 0 or enable them by setting it to 1. The configurations that correspond to Ellis and Sakurai [28] are with **pNCS** set to 0.0 or 1.0 and parton cancellations activated. In addition, the **pNCS**=0.5 is also considered in the analysis.

4.3 Parton Distribution Function Set

The interpretation of experiments at the Large Hadron Collider necessitates the development of high-precision statistical analysis methods. One example is the determination of Parton Distribution Functions (PDF), which is done through a worldwide analysis of existing data sets [29].
A proton consists of two up quarks and one down valence quark. The Gluon mediates between the quarks and is responsible for the stability of the atomic nucleus. For each parton in the proton a structure function $f(x, Q^2)$ is introduced, which describes the probability to scatter off a quark or gluon in the proton with the momentum fraction x that the parton carries at the energy scale Q^2. This is why all structure functions that are very important for Monte Carlo simulation are saved in the PDF.
These PDFs, along with their associated uncertainties, should be free of theoretical prejudice and should be accompanied by a true statistical confidence level.
The NNPDF collaboration [30] has developed an approach based on a Monte Carlo estimate of uncertainties that allows one to transmit the uncertainty in the experimental data to the fitted PDFs and any other quantity that depends on them.
Neural networks (stands for NN in NNPDF) are used to parameterize the PDFs. These networks have a large number of free parameters, which are sufficient to assure that the resultant ensembles of customized PDFs are free of bias caused by assumptions about the underlying functional form.
CT10 is a PDF Set of CTEQ collaboration [31] and it is based on Hessian method. The different to NNPDF is that CT10 does not function with neural network.
The PDF Set NNPDF3.1 is the latest version and is currently in use.

This analysis uses 4 different PDF sets [21], which are listed in Table 2.

LHAPDF ID	PDF Set name
10800	CT10
232000	NNPDF23_nnlo_as_0118
261000	NNPDF30_nnlo_as_0118
303600	NNPDF31_nnlo_as_0118

Table 2: PDF Sets [32]

LO, NLO, NNLO are the accuracy levels of the calculation and stands for leading order, next-to-leading order and next-to-next-to-leading order, which means that there are more and more corrections in the feynman graphs. CT10 has NLO as accuracy level and the NNPDF2.3/NNPDF3.0/NNPDF3.1 all have NNLO.

5 Analysis

This chapter covers the analysis of events on the generator level and the reconstructed, simulated events, which are obtained from the generator and processed further. Simulation was conducted at the center of mass energy of 13 TeV. All following plots are normalized to the integral of the entries.
Only the outgoing particles are observed for the entire analysis.

5.1 Generator Level

The following 3 paragraphs are about 3 variables that are introduced in each case and in Section 5.2 they are compared with each other in different PDF sets.

5.1.1 Particle Distribution

This section examines the Particle Data Group ID (PDG ID) distribution, which means the distribution of the individual particles. All particles between -16 and 16 are considered and include all leptons and quarks. The negative numbers are respective antiparticles. The following Tables 3 and 4 shows the particles with their respective particle numbers. There are also Gauge and Higgs bosons and many other particles that are not discuss here. Only the particles from 1 to 6 and 11 to 16 and their antiparticles are used for this analysis. The other particles do not appear.

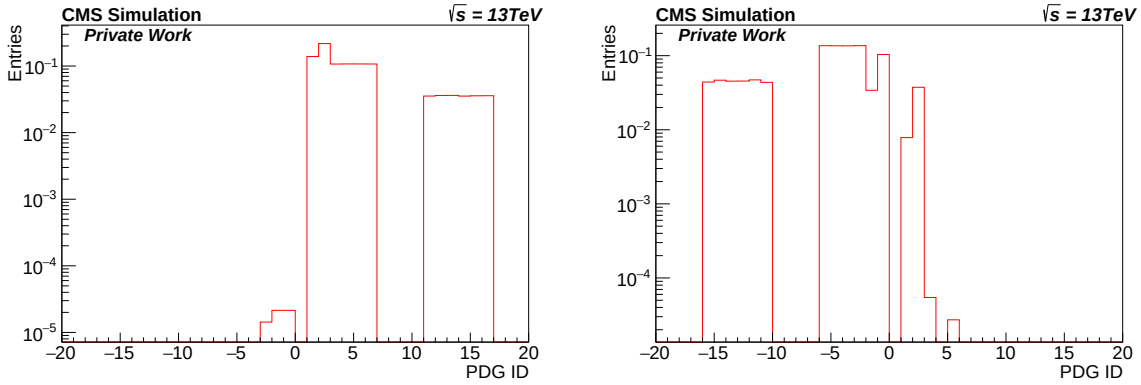
Quarks	
d	1
u	2
s	3
c	4
b	5
t	6

Table 3: PDG IDs Quarks

Leptons	
e^-	11
ν_e	12
μ^-	13
ν_μ	14
τ^-	15
ν_τ	16

Table 4: PDG IDs Leptons

For the following three plots, the PDF set CT10 was used at E_{Sph}=9 TeV.
If the particles obtained are analyzed for **pNCS**=1 and **pNCS**=0 in the Figure 7 a) and b) it can be stated that for **pNCS**=1 one mainly gets quarks and leptons. For **pNCS**=0 one gets anti-leptons and anti-quarks, but also a lot of up and strange quarks. Leptons are obtained for **pNCS**=1 and antileptons for **pNCS**=0. It should be noted that there is a factor of 3 more quarks than leptons. The additional factor could be explained by the theory prediction of additional particles (see Equations 4 and 5), which could explain the appearance of a factor of 3 between leptons and quarks.



(a) **pNCS**=1 (b) **pNCS**=0

Figure 7: PDG ID distribution

Comparing the leptons and antileptons in the Figure 8 with **pNCS**=0.5, one can see that they are evenly distributed, which resembles approximately an uniform distribution. Looking at the quarks in the three figures, one can see that there is a peak for **pNCS**=0.5. More quarks appear than in the cases **pNCS**=1 and **pNCS**=0. Looking at the exact explanation of **pNCS**, it indicates the probability of whether one has events with ΔNCS=+1 or none where the probability is set to 0.
(The generator works fine if one considers at the Equations 4, 5 as described in the theory and compares them with the plots.)

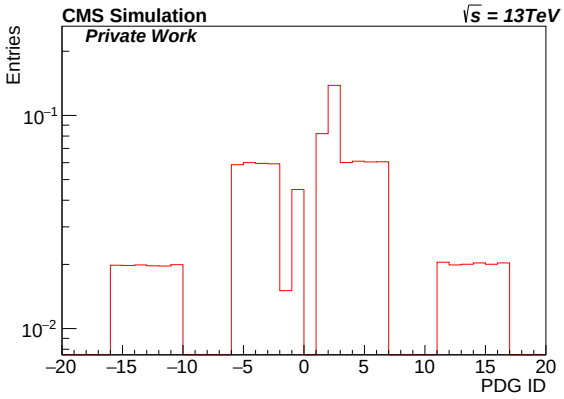


Figure 8: **pNCS**=0.5

The number of jets per event is also interesting variable that can be analysed as one expects a lot of jets. Analyzing the number of jets in the following Figure 9, one can see that one is expecting at least 9, just like the decay in theory predicts. From the three plots one mainly expects 9, 10, 11, 12, 13 or 14 jets in the final state. In comparison to paper [10] one can see that the outcome of the number of jets is comparable to the 2016 CMS analysis. As possible final states are 3 leptons with 11 quarks, 3 leptons with 9 quarks or 3 leptons with 7 quarks listed.

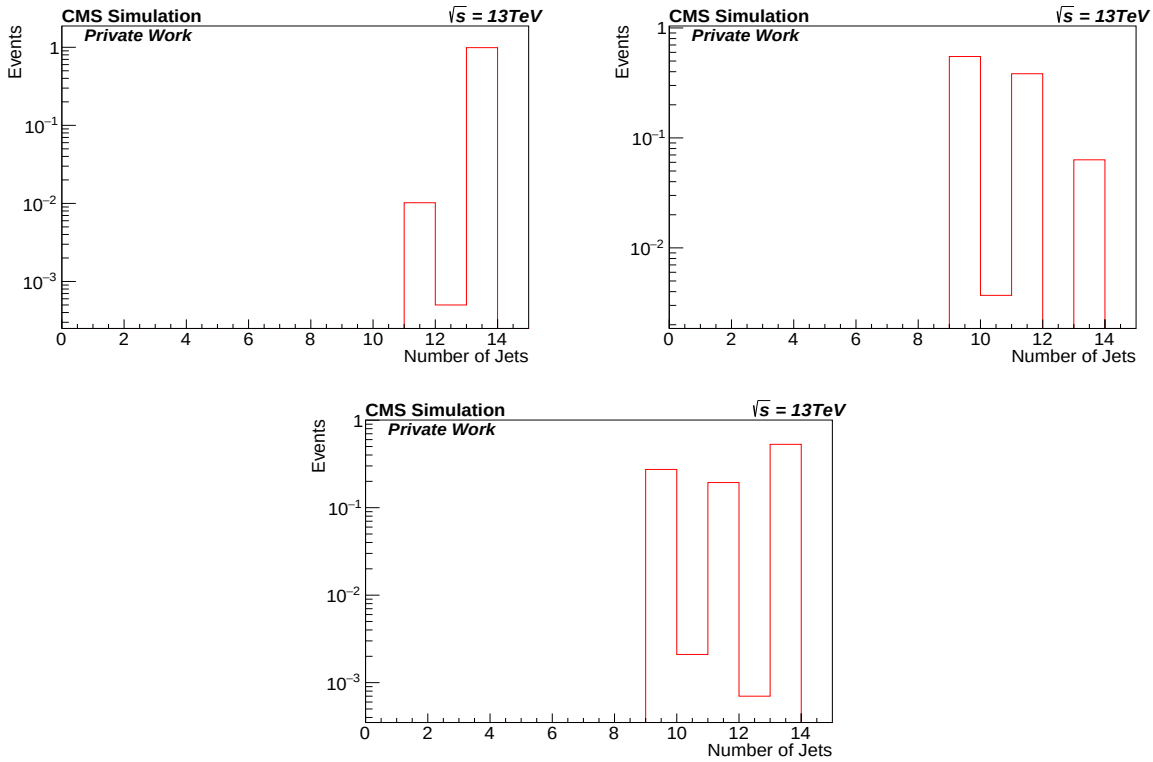


Figure 9: Top left with **pNCS**=1, top right **pNCS**=0 and bottom with **pNCS**=0.5, at E_{Sph}=9 TeV

5.1.2 Invariant Mass Distribution

In this section the invariant mass of two lepton pairs in final states are considered and determined. The next plots show how the invariant masses behave in the respective lepton channels e+μ, e+τ and μ+τ. The leptons can be measured better than jets, so one wants to try whether one can ignore jets and not include the uncertainty caused by jets in the analysis. Here the jets were not considered. The invariant mass $M_{e\mu}$, $M_{e\tau}$ and $M_{\mu\tau}$ with $\sqrt{s}$=13 TeV with different Sphaleron energy at 8, 9 and 10 TeV are shown in the following figure.

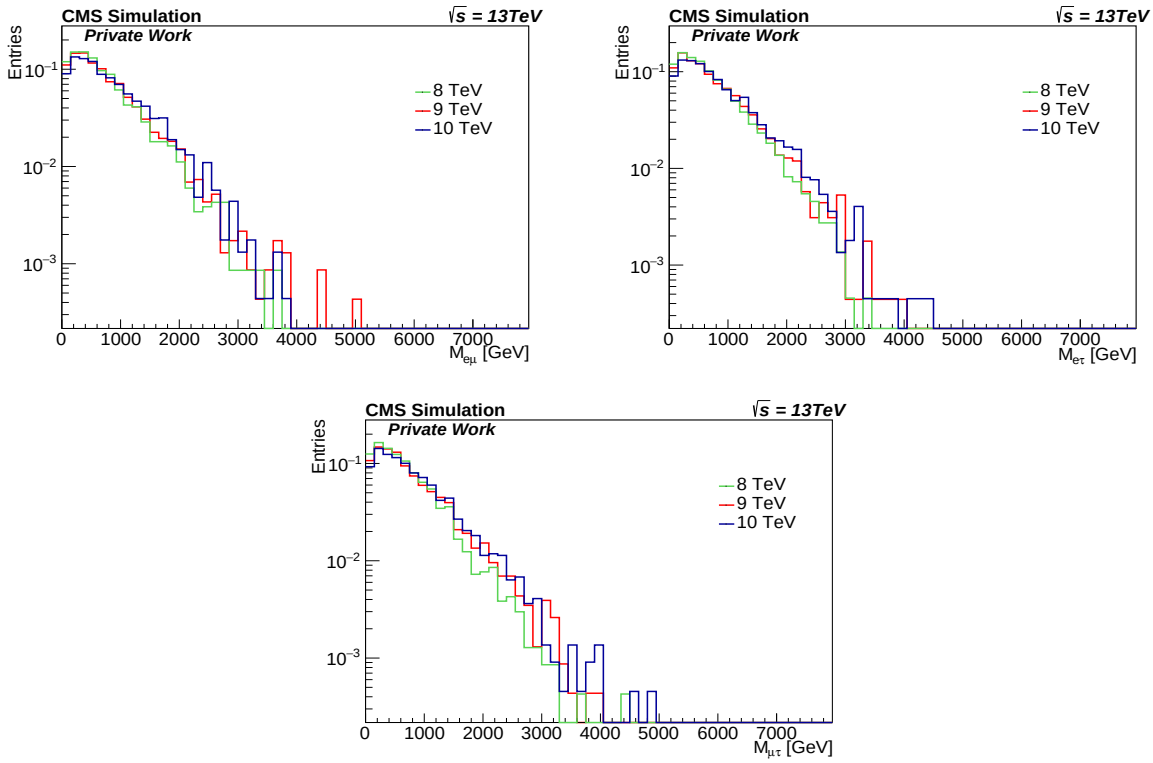


Figure 10: The invariant masses $M_{e\mu}$ (top left), $M_{e\tau}$ (top right) and $M_{\mu\tau}$ (bottom)

The signal shapes do not show any resonance points, here the shapes are relatively smooth and do not contain any peaks.
If one fixes the Sphaleron energy and changes **pNCS**, one does not get much differences. For higher Sphaleron energy the invariant mass per event increases. One realizes in the plots that the curve becomes larger with higher E_{Sph} on the x-axis.
However, adding the jets, the invariant mass does not change much, like in Figure 11.

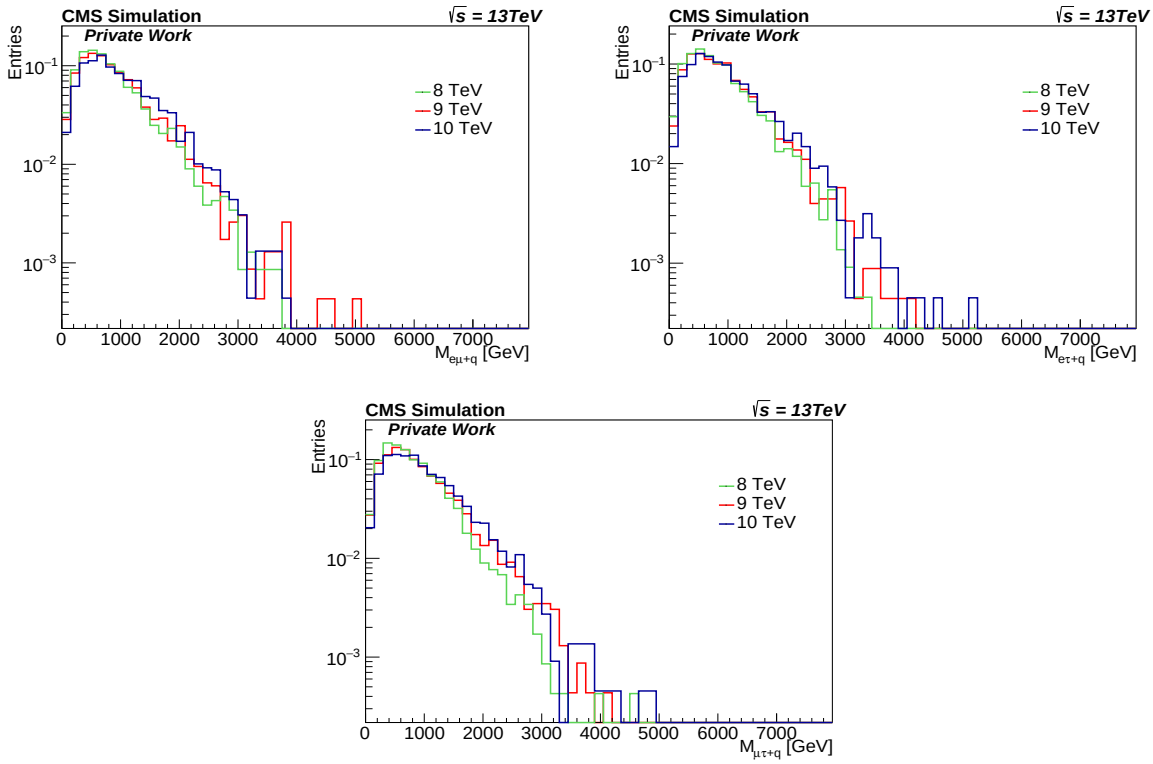


Figure 11: The invariant masses $M_{e\mu+q}$ (top left), $M_{e\tau+q}$ (top right) and $M_{\mu\tau+q}$ (bottom)

When comparing all outgoing particles with the only two outgoing leptons, one finds that it does not lead to a great difference if one compares their invariant mass.

5.1.3 Sum of transverse momentum

A single discriminating variable S_T is used, defined as the scalar sum of the p_T of all N energetic objects in an event (defined as jets, electrons, muons, and photons with p_T more than a specified threshold), plus p_T^{miss} in the event if it exceeds the same threshold:

$$S_T = p_T^{miss} + \Sigma_i p_T^i \qquad \text{with} \qquad p_T^{miss} = 0 \tag{15}$$

$p_T^{miss} = 0$ because on generator level all particles including neutrinos are included. This variable is interesting because many particles are created during the production of the Sphalerons. In the previous section the invariant mass was considered, where one cannot see a large difference between two leptons in final state and all outgoing particles. The advantage of the S_T distribution is that one uses more information, but the disadvantage is that this included object information may not be as easy to reconstruct. The following figure shows the sum of transverse momentum being peaked at 9 TeV, using the PDF Set CT10 with the variables **pNCS**=1 and a Sphaleron energy of 9 TeV. If one changes the parameter **pNCS**, one does not see any large deviations. The particle content changes, but the energy that is distributed over them does not change. In the next section 5.2 one examines S_T in different PDF sets.

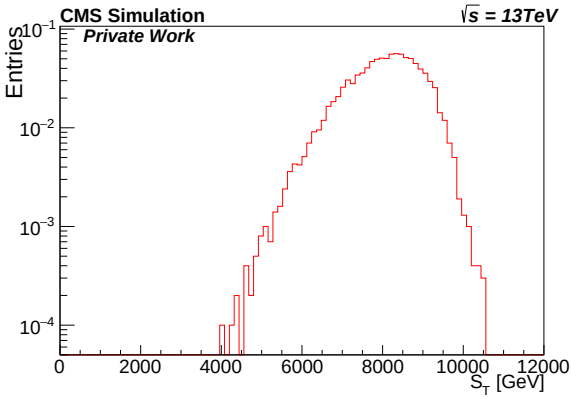


Figure 12: S_T with CT10 at E_{Sph}=9 TeV and **pNCS**=1

5.2 Comparison of different PDF sets

In this section, the ratio of predictions using two different PDF sets is analyzed. The results obtained from the previous sections are used for more detailed studies of the PDF Set. Since one is still on the generator level, the LHE files always are produced with the same parameters, such as the number of events, the Spaheleron energy, etc. only with different PDF sets. It is important that the files are comparable in order to keep the differences between the simulated points at a minimum. For the following analysis, two different data files are placed in one plot. One creates a second, smaller plot window under the histogram. This little plot window is prepared for the ratio of the one data file divided by the other.
First, the PDG ID distribution is considered. One expects the same number of jets for the respective PDF sets. The first Figure 13 shows the LHE-files generated in NNPDF3.1 and NNPDF3.0 with 10k events, **pNCS**=0.5 and 9 TeV. Here the range in the ratio was not changed, errors and outliers are visible in one plot.

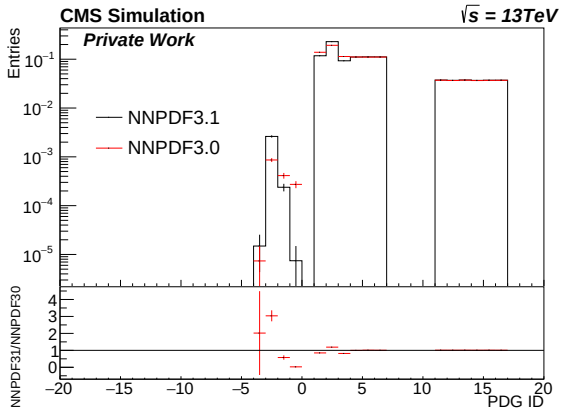


Figure 13: E_{Sph}=9TeV and **pNCS**=1

In the following three plots, the range has been reduced to 0 to 2 here. One can see that

there are minor deviations between the different 2 PDF sets. Slightly larger deviations are observed for the number of quarks. For example, in Figure 14 for E_{Sph}= 9 TeV the predictions for the quarks deviate by 10-20%.

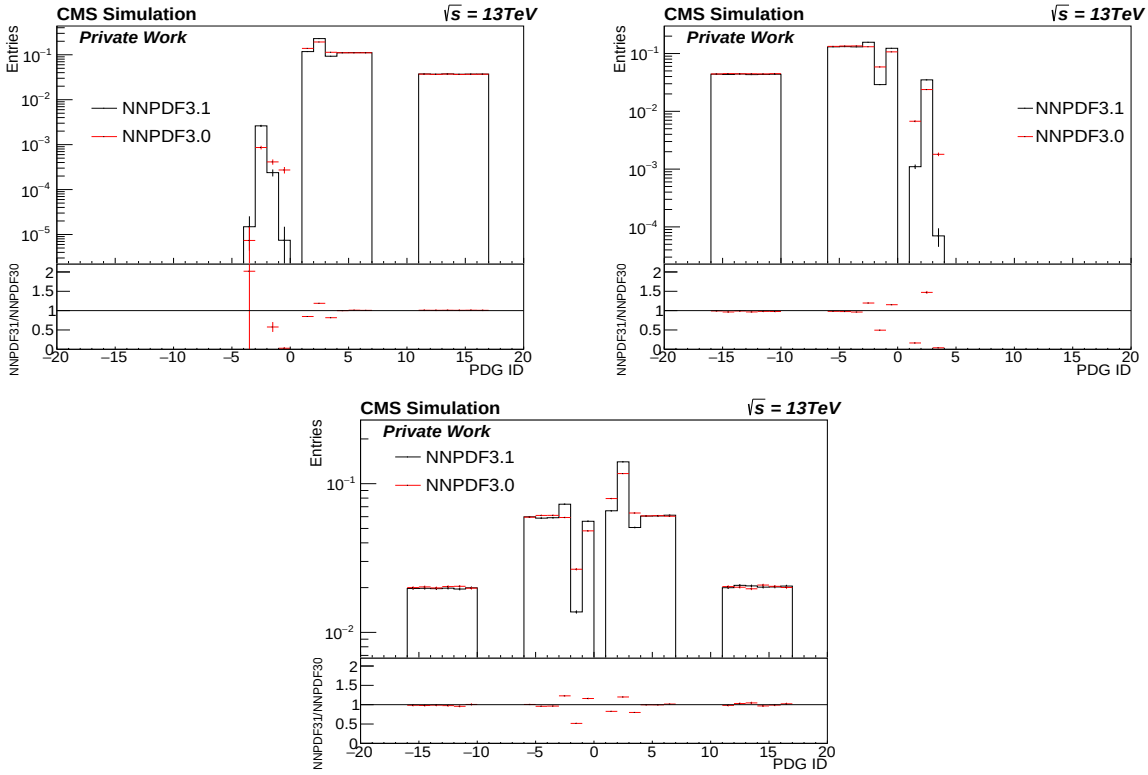


Figure 14: Top left with pNCS = 1, top right pNCS = 0 and bottom with pNCS = 0.5, at E_{Sph}=9 TeV

A small outlier can be seen in the top left figure for anti charm. For the case **pNCS**=0.5, the errors are near line 1. Mainly large deviations occur for quarks.
As these differences occur for the quarks, one might need to consider these in an additional uncertainty. These differences should be included as systematic uncertainties in jet-based analyses, while leptons are largely unaffected.

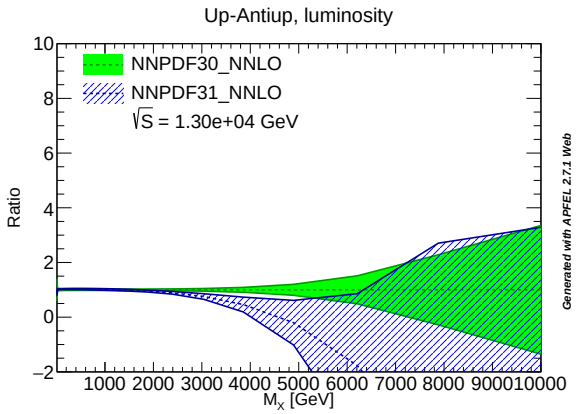


Figure 15: Comparing PDF Set between NNPDF3.1 and NNPDF3.0

The Figure 15 was generated with the help of the APFEL WEB [33] page, where one can choose different PDF sets and quarks as one likes. It shows the comparison between the two PDF sets NNPDF3.1 and NNPDF3.0 by only looking at the up and anti-up quarks. Analyzing the ratio between the two sets, one can see large variations for some TeV. From 3 TeV, the higher the energy, the greater the factor. From around 4 TeV one can see that the predictions become negative, which is unphysical. In other words, for larger masses, the up-anti-up densities are very insecure, that means different predictions are expected for different PDF sets. Therefore, as in the previous plots, one gets a deviation for quarks.

Next, the latest PDF set NNPDF3.1 is compared with NNPDF3.0 by varying the Sphaleron energy at **pNCS**=0.5.

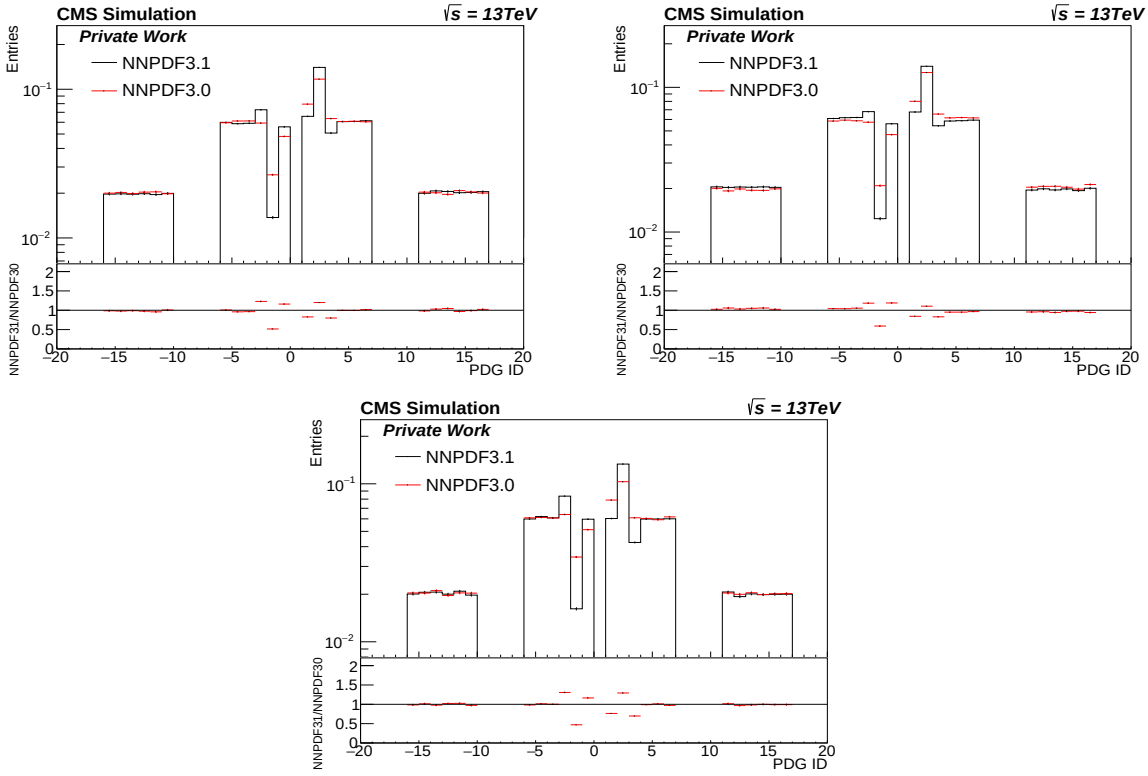


Figure 16: Top left with E_{Sph}=9 TeV, top right E_{Sph}=8 TeV and bottom with E_{Sph}=10 TeV with **pNCS**=0.5

Afterwards the PDF Set CT10 is compared with the NNPDF3.1, NNPDF3.0 and NNPDF2.3. The Sphaleron energy is kept constant at 9 TeV and **pNCS**=1. Here it is interesting to see if there is any significant deviation between the sets. Looking at the next 3 plots there are very large differences between the PDF sets for the quarks. A same effect as before with the ratio between NNPDF3.1 and NNPDF3.0 only more distinct can be seen here.

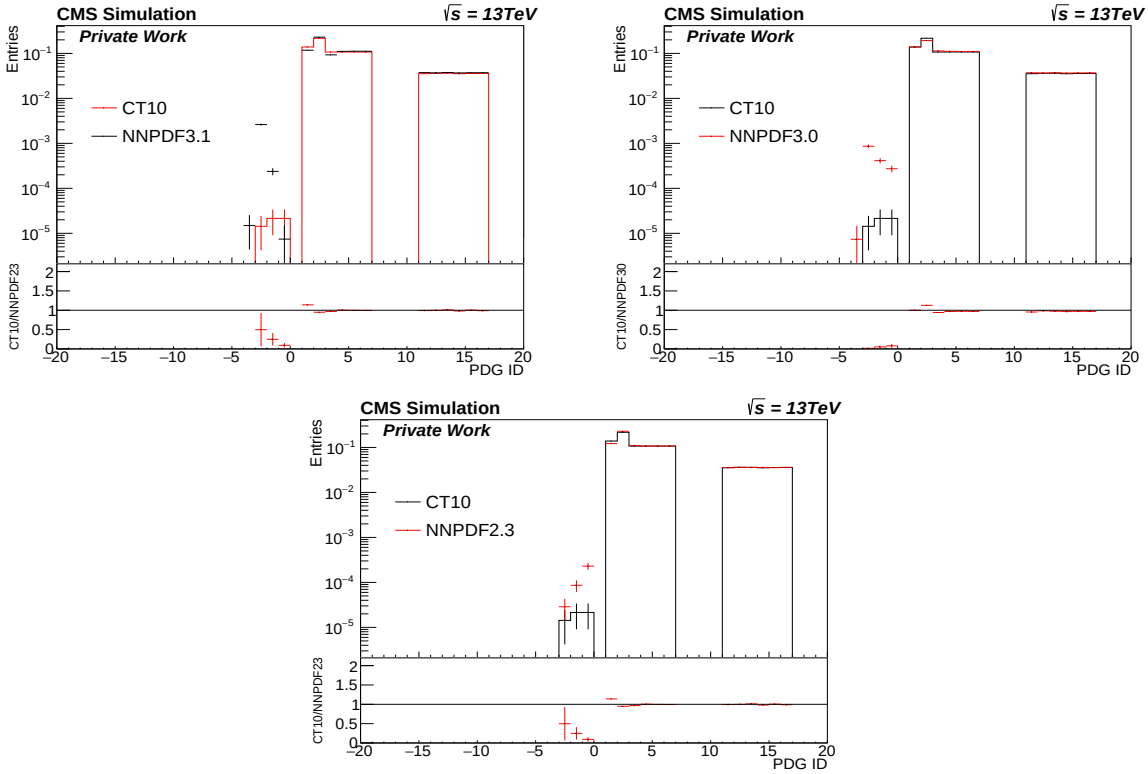


Figure 17: All with E_{Sph}=9 TeV and **pNCS**=1

It is consistent that the quarks differ extremely between the respective PDF sets. Analyzing the generated plot 18, one can recognize that there are very strong deviations for higher energies between the two PDF sets CT10 and NNPDF3.1. Up- and Anti-Up was also taken here as an example. Comparing this plot with the plot 15 where NNPDF3.1 and NNPDF3.0 have been compared, one will notice that there are very highly deviations from 6 TeV, which explains the deviations in the Figure 17.

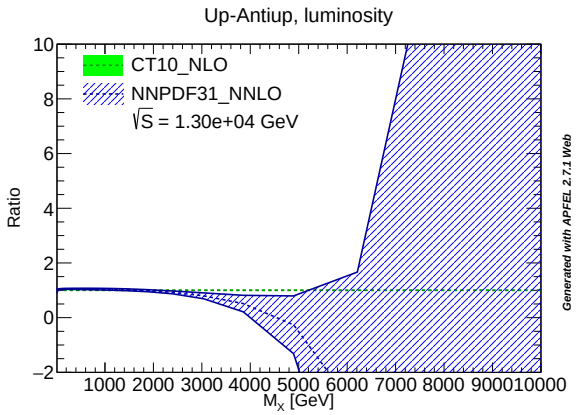


Figure 18: Comparing PDF Set between NNPDF3.1 and CT10

Next, the invariant mass of electron and muon by 9 TeV and with pNCS=0.5 is analyzed.

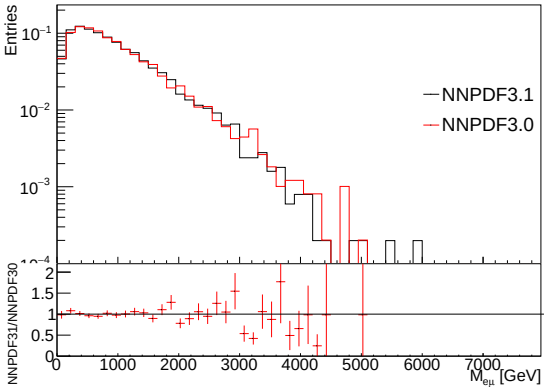


Figure 19: E_{Sph}=9TeV and **pNCS**=0.5

If one compares the two invariant masses of both PDF sets, one finds that within the statistical uncertainty the shapes of the two predictions agree.
Figure 20 shows the invariant mass for the other two final states.

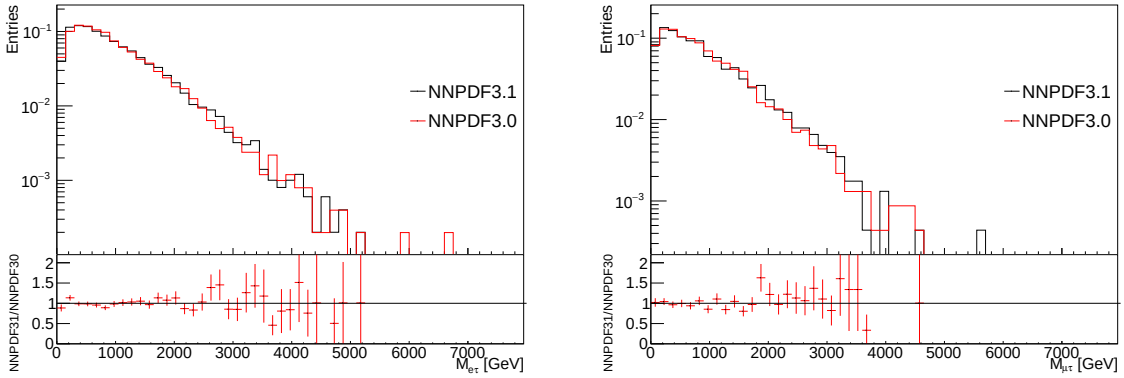
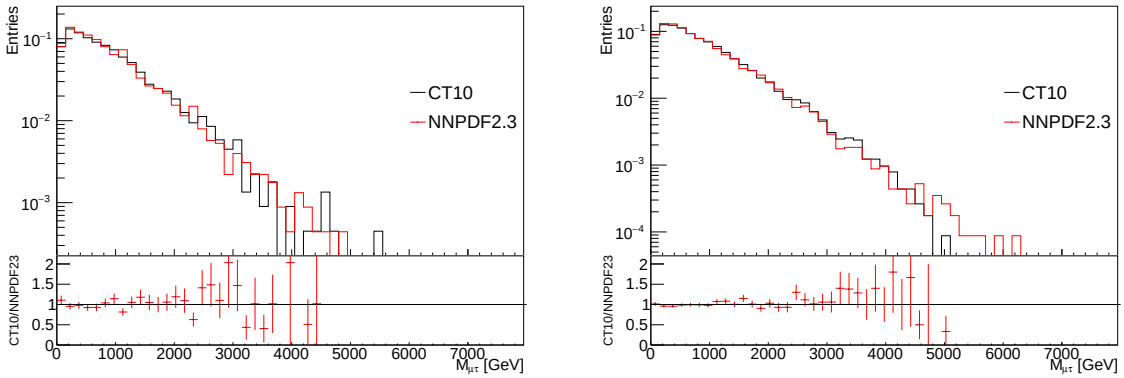


Figure 20: E_{Sph}=9 TeV and **pNCS**=0.5

Increasing the number of events helps to better distinguish the predictions at high masses. An example can bee seen in the Figure 21, the plot on the left has 10k events and on the right 50k. As an example, the invariant mass of a muon and a tau is chosen, using CT10 and NNPDF2.3 as the PDF set.



(a) 20k (b) 50k

Figure 21: Left with 10k and right 50k by E_{Sph}=9 TeV and **pNCS**=0.5

The Figure 22 shows the sum of transvere momenta for the respective PDF set. If the Sphaleron energy is increased, the sum of transverse momentum also increases. The uncertainty becomes smaller at the peak and the further one moves towards the edge, the greater the differences in the tails, but still within the statistical uncertainty.

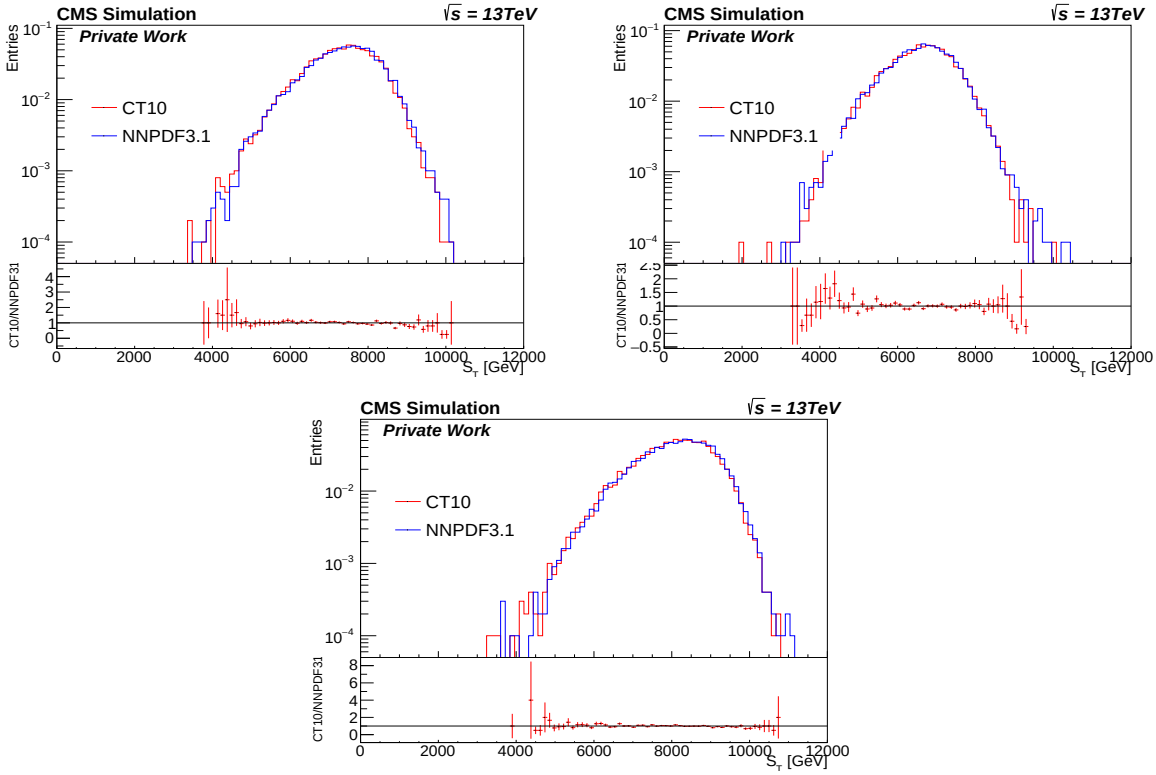


Figure 22: Top left with E_{Sph}=9 TeV, top right E_{Sph}=8 TeV and bottom with E_{Sph}=10 TeV with **pNCS**=0.5

5.3 RECO/Reconstructed-Level

The typical data analysis flow in CMS and other LHC experiments includes many phases of data processing and reduction. While collisions occur at a 40 MHz rate and hundreds of millions of channels are readout in the experiment, just a few observables from a very limited selection of relevant occurrences are included [34].
The data reduction process begins with zero suppression algorithms and trigger mechanisms in the detector hardware. The former reduces the quantity of data in each event (event content), whereas the latter decreases the number of events to be processed. In the reconstruction and analysis chain, additional data reduction processes are then used.
The Particle Flow (PF) algorithm aims to identify and reconstruct all of the particles from the collision by optimally integrating the data from different subdetectors. As a result, it relies on track reconstruction and clustering to separate different overlapping showers, as well as a combination of the responses of each sub-detector [35].

5.3.1 NANOAOD-Format

NanoAOD is a type of event data format. It contains high-level physics object information and is around 20 times smaller than the MiniAOD format, that was the main dataformat used in CMS before the development of NanoAOD. NanoAOD enables automated data analysis workflows and may be readily customized for development operations [36].
Naming conventions for collections of objects in NanoAOD format are shown in the Table 5.

Branch name	Type	Data type	Example
nObject	Scalar	Unsigned integer	nMuon, nElectron, nGenPart, nJet
Object_var[i]	Array	Any	Muon_pt[i], Tau_mass[i], GenPart_pdgId[i]

Table 5: Object collection naming conventions [37]

The rows correspond to the TTree branches associated with the event content. Object attributes can be of any data type, including signed or unsigned integers, floats, and booleans.
This section is about analyzing the generated data in NanoAOD format. NanoAOD accomplishes such a high level of data reduction by preserving just high level information on physics objects like jets and leptons, eliminating their individual constituents, and lowering the precision of recorded variables.
The analysis framework used to analyze NanoAOD is the RootDataFrame, which is based on the C++ programming language, which means that it had to be adapted accordingly in Python. Each method has to be predefined in the gInterpreter in C++, which can then be used later in Python code.
One looks at the reconstructed samples, where one can match the particle to generator level particle. Otherwise nothing else was done, e.g. no trigger, filter, etc.
The PDF Set CT10 was used for the next results. The number of events is set to 50k

and for **pNCS**=0.5. In the next three pictures one can see the invariant mass of eμ, eτ and $\mu\tau$.

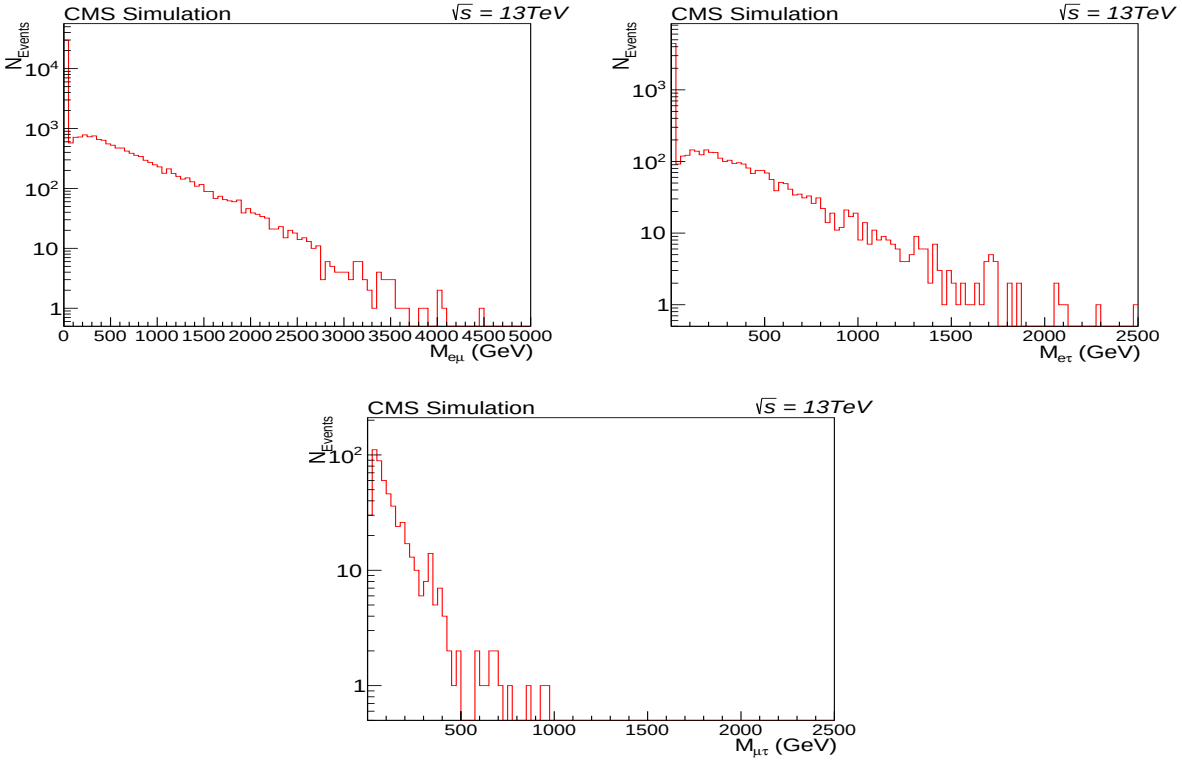


Figure 23: Top left invariant mass $M_{e\mu}$, top right $M_{e\tau}$ and bottom $M_{\mu\tau}$, at E_{Sph}=9 TeV and with **pNCS**=0.5

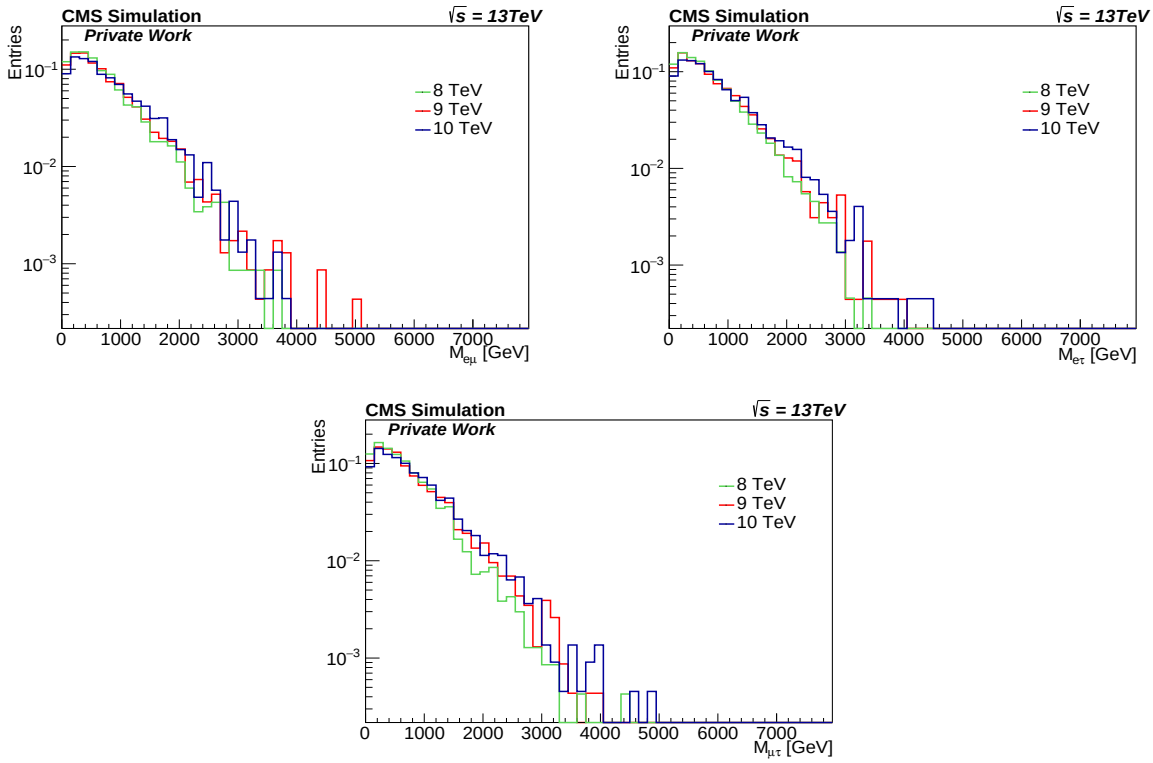
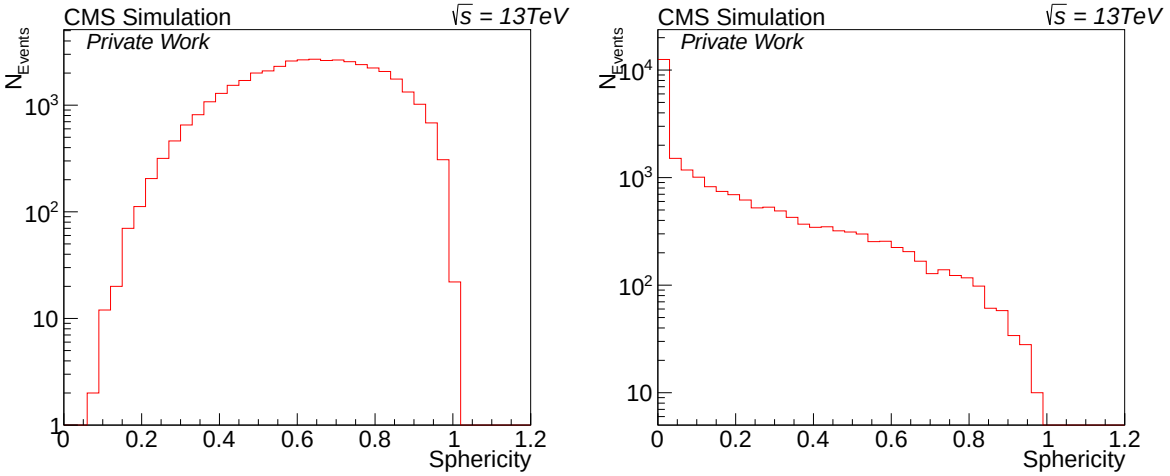


Figure 24: The invariant masses $M_{e\mu}$ (top left), $M_{e\tau}$ (top right) and $M_{\mu\tau}$ (bottom)

Comparing it with the generator level Figure 24, one can see that the shape is similar. Looking at the peaks, one can notice that they have pretty much the same values around 300-400 GeV. There are fewer events for eτ and $\mu\tau$ than at generator level, that could be due to Tau. The Taus are not well reconstructed. They decay quickly, already in the detector, because their main lifetime is very short and must therefore be reconstructed using the particle decay. Since not all particles are correctly reconstructed, a part of the energy is lost, which can also be seen in the plot. The invariant mass only goes up to about 1000 MeV. This could be due, for example, to missing neutrinos that have not been reconstructed.
Since no quality analysis was made here either (e.g. without a trigger) but only generator-level- and reco-level particles, which belong together, were looked at on an imprecise qualitative level, one also gets the one bin at 0 GeV.

5.4 Sphericity

The transverse sphericity distribution given by Eq. 14 is illustrated in figure for eμ. The data is the same as before with CT10, pNCS = 0.5 and number of events at 50k.

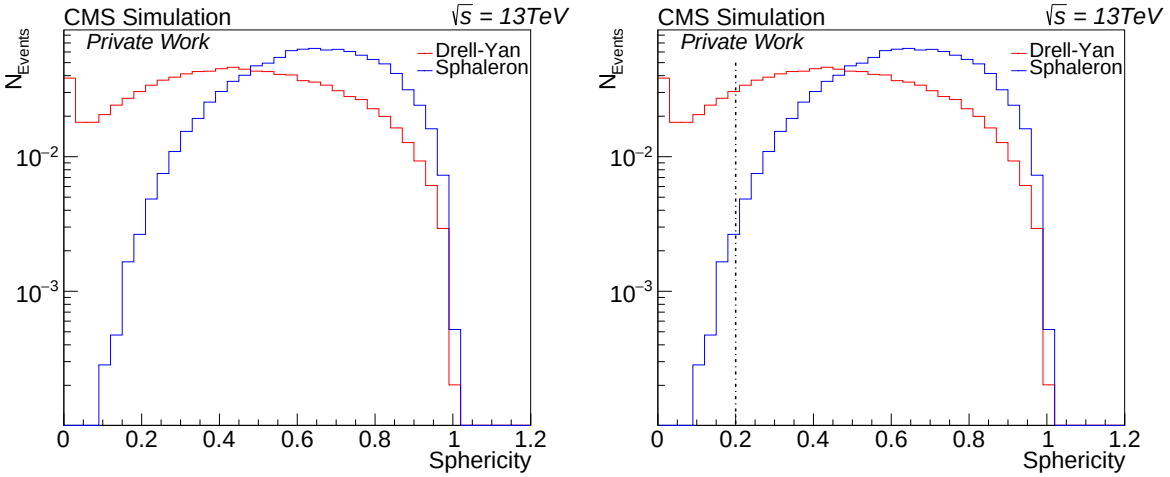


(a) eμ with jets

(b) only eμ

Figure 25: 9 TeV at 50k, eμ left with jets and right without jets

On the left side in Figure 25 there are leptons (electron and muon) with jets and on the right plot only the two leptons, there is a big difference here. One can also see quite well here that one leaves out a lot of information by only observing two leptons. There are relatively many entries for the sphericity between 0 and 0.2



(a) Drell-Yan background and Sphaleron Signal

(b) Cut at 0.2

Figure 26: Comparison of sphericity

Here the crosssection was normalized to 1, since the interest lies in the course in order to compare the shapes with one another. The red curve here in the Figure 26 is the Drell-Yan background and the blue curve is the Sphaleron signal, which is not at all at the very low, but mostly at the higher entries. Here one could, for example, cut away

a lot of background at 0.2 without losing a lot of Sphaleron signal. If the cut is carried out as shown in the Figure 26 b), about 16% of the Drell-Yan background would be omitted. On the other hand, not even 1% would be cut off for the Sphaleron signal, as can also be seen from the figure, there are not many events between 0 and 0.2.

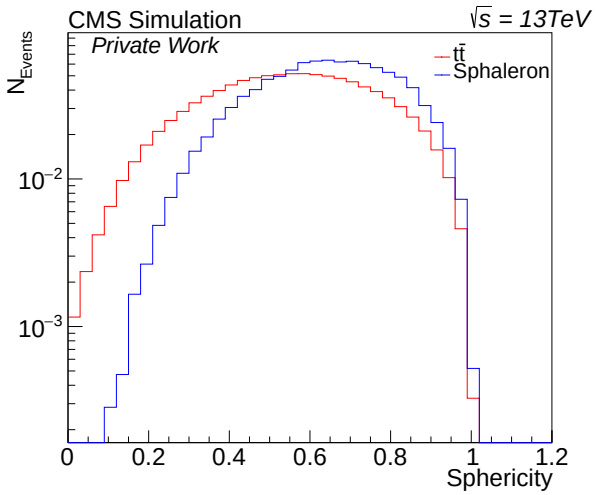


Figure 27: $t\bar{t}$ background and Sphaleron signal

The Figure 27 shows the $t\bar{t}$ background, one of the main backgrounds, where in the final state there can be two leptons with different flavors. $t\bar{t}$ is more similar to Sphaleron signal than Drell-Yan background. Cutting again at 0.2, the advantage is that only 5% of $t\bar{t}$ background is lost.

6 Conclusion and Outlook

The search for Sphalerons in three different lepton channels eμ, eτ and $\mu\tau$ were analyzed for different variables with the help of the BaryoGEN generator. It is very important to mention that if one only observes two leptons in their final states, a substantial amount of other information is lost. However, if one considers the invariant mass then the quarks can be ignored, that means they do not play an essential role here. Looking at the number of quarks, the distribution of the individual quarks, there are large deviations between the different PDF sets. An additional uncertainty should be included here. The distribution of S_T, the sum of transverse momentum of all the individual particles (here leptons and quarks) in the different PDF sets are similar and do not differ greatly from one another. In the second part of the analysis the data generated by the generator were used to reproduce the reconstruction of the invariant mass and the sphericity of two leptons in their final states. If one compares the distributions of the invariant mass of the BaryoGen and the reconstructed data, one finds that the top quarks have not been correctly reconstructed. Analyzing the transverse sphericity, a lot of information is lost without consideration the jets by only having two leptons in the final state. Without the jets the geometric shape changes from the collision event and this results in a different sphericity distribution.
It should also be noted here that the jets cannot simply be neglected. It was interesting to see what difference one gets if one only looks at the two leptons in the final state or all outgoing particles. It is possible to observe a pair of leptons in their final states, but there is however an elevated risk to lose a substantial amount of additional information. In the case of sphericity, it would be interesting to do another study, a qualitative analysis with filter, trigger,... a quality application to search further for Sphalerons. The continuation with $t\bar{t}$ background processes would be interesting, as one would not even have to cut off 1% for the Sphaleron signal or to analyze with other background samples.

7 Bibliography

[1] Albert M Sirunyan et al. "Search for black holes and sphalerons in high-multiplicity final states in proton-proton collisions at $\sqrt{s}$=13 TeV". In: *Journal of High Energy Physics* 2018.11 (2018), pp. 1–49.

[2] Morad Aaboud et al. "Search for lepton-flavor violation in different-flavor, high-mass final states in p p collisions at $\sqrt{s}$=13 TeV with the ATLAS detector". In: *Physical Review D* 98.9 (2018), p. 092008.

[3] CMS collaboration et al. "Search for lepton-flavor violating decays of heavy resonances and quantum black holes to eμ final states in proton-proton collisions at $\sqrt{s}$=13 TeV". In: *arXiv preprint arXiv:1802.01122* (2018).

[4] Dieter Meschede. *Gerthsen physik*. Springer-Verlag, 2015.

[5] PAUL A TIPLER and GENE MOSCA. *Physik Für Wissenschaftler und Ingenieure, Zweite deutsche Auflage*. 2006.

[6] David Griffiths. *Introduction to elementary particles*. John Wiley & Sons, 2020.

[7] Wikipedia. *Standard Model — Wikipedia, The Free Encyclopedia*. 2021.

[8] K Nakamura, C Amsler, Particle Data Group, et al. "Particle physics booklet". In: *Journal of Physics G: Nuclear and Particle Physics* 37.7A (2010), p. 075021.

[9] Wikipedia. *Higgs boson — Wikipedia, The Free Encyclopedia*. `http://en.wikipedia.org/w/index.php?title=Higgs%20boson&oldid=1048567943`. [Online; accessed 07-October-2021]. 2021.

[10] Cameron Bravo and Jay Hauser. "BaryoGEN, a Monte Carlo generator for sphaleron-induced transitions in proton-proton collisions". In: *arXiv preprint arXiv:1805.02786* (2018).

[11] Graham Albert White. *A pedagogical introduction to electroweak baryogenesis*. Morgan & Claypool Publishers, 2016.

[12] CMS CERN. `https://cms.cern/detector`.

[13] Roman Adolphi et al. "The CMS experiment at the CERN LHC". In: *Jinst* 803 (2008), S08004.

[14] Wikipedia. *Large Hadron Collider — Wikipedia, The Free Encyclopedia*. `http://en.wikipedia.org/w/index.php?title=Large%20Hadron%20Collider&oldid=1035089064`. [Online; accessed 25-July-2021]. 2021.

[15] Welt der Physik. *Large Hadron Collider – LHC*. `https://www.weltderphysik.de/gebiet/teilchen/experimente/teilchenbeschleuniger/lhc/`.

[16] Wikipedia. *Compact Muon Solenoid — Wikipedia, The Free Encyclopedia*. `http://en.wikipedia.org/w/index.php?title=Compact%20Muon%20Solenoid&oldid=1029594122`. [Online; accessed 25-July-2021]. 2021.

[17] Albert M Sirunyan et al. "Particle-flow reconstruction and global event description with the CMS detector". In: *Journal of Instrumentation* (2017), p. 86.

[18] Wikipedia. *Jet (particle physics) — Wikipedia, The Free Encyclopedia*. `http://en.wikipedia.org/w/index.php?title=Jet%20(particle%20physics)&oldid=997701177`. [Online; accessed 03-August-2021]. 2021.

[19] Christoph Berger. *Elementarteilchenphysik*. Springer, 2002.

[20] M Fernández, E Cuautle, and G Paić. "Event shape variables, a tool for QCD study". In: *Journal of Physics: Conference Series*. Vol. 761. 1. IOP Publishing. 2016, p. 012031.

[21] Andy Buckley et al. "LHAPDF6: parton density access in the LHC precision era". In: *The European Physical Journal C* 75.3 (2015), pp. 1–20.

[22] *CMS Offline Guide*. `https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBook`.

[23] C. Bravo. `https://github.com/cbravo135/BaryoGEN`.

[24] Torbjörn Sjöstrand et al. "An introduction to PYTHIA 8.2". In: *Computer physics communications* 191 (2015), pp. 159–177.

[25] Johan Alwall et al. "A standard format for Les Houches event files". In: *Computer Physics Communications* 176.4 (2007), pp. 300–304.

[26] F Krauss et al. *42. MONTE CARLO PARTICLE NUMBERING SCHEME*.

[27] Glen Cowan. "Statistical tests with weighted Monte Carlo events". In: *Physics Department, Royal Holloway, University of London, Egham, TW20 0EX, UK* (2012).

[28] Daria Satco. *Search for Sphalerons in Proton-Proton Collisions*. Tech. rep. 2017.

[29] Richard D Ball et al. "Fitting parton distribution data with multiplicative normalization uncertainties". In: *Journal of High Energy Physics* 2010.5 (2010), pp. 1–34.

[30] Richard D Ball et al. "NNPDF". In: *arXiv preprint arXiv:1410.8849* (2014).

[31] Marco Guzzi et al. "CT10 parton distributions and other developments in the global QCD analysis". In: *arXiv preprint arXiv:1101.0561* (2011).

[32] LHAPDF6. `https://lhapdf.hepforge.org/pdfsets.htmlN`.

[33] Stefano Carrazza et al. "APFEL Web: a web-based application for the graphical visualization of parton distribution functions". In: *Journal of Physics G: Nuclear and Particle Physics* 42.5 (2015), p. 057001.

[34] Andrea Rizzi, Giovanni Petrucciani, and Marco Peruzzi. "A further reduction in CMS event data for analysis: the NANOAOD format". In: *EPJ Web of Conferences*. Vol. 214. EDP Sciences. 2019, p. 06021.

[35] Florian Beaudette. "The CMS particle flow algorithm". In: *arXiv preprint arXiv:1401.8155* (2014).

[36] Marco Peruzzi et al. "The NanoAOD event data format in CMS". In: *Journal of Physics: Conference Series*. Vol. 1525. 1. IOP Publishing. 2020, p. 012038.

[37] Karl Ehatäht. “NANOAOD: a new compact event data format in CMS”. In: *EPJ Web of Conferences*. Vol. 245. EDP Sciences. 2020, p. 06002.