1

Introduction

There is a theory which states that if ever anyone discovers exactly what the Universe is for and why it is here, it will instantly disappear and be replaced by something even more bizarre and inexplicable. There is another theory which states that this has already happened.

– Douglas Adams, *The Hitchhiker's Guide to the Galaxy*

Nevertheless, humankind is still trying to understand the most fundamental aspects of our universe, by studying the fundamental particles matter is made of and the interactions between them. The story so far has been summarised in a theory called the Standard Model of Particle Physics. This theory has been extensively tested and has already predicted many experimental observations. However, it cannot explain the full story, and some pieces remain missing. Gravity, for example, is not incorporated in the Standard Model. Similarly, it cannot explain the observed neutrino masses or the matter-antimatter asymmetry.

Another mystery stems from a series of cosmological observations made during the last century. These observations are based on gravitational effects, such as measurements of the rotation curves of galaxies [1] and gravitational lensing [2], and on the analysis of the Cosmic Microwave Background (CMB) [3–5]. The collected evidence shows that there is matter in the universe, which is not visible from measurements at any wavelength of the electromagnetic spectrum. This so-called dark matter was found to constitute about 85% of the matter in the universe, which means that the ordinary matter described by the Standard Model only accounts for 15% of all matter. So far, only very little is known about the dark matter, as it does not interact through any of the forces included in the Standard Model, and has only been observed through gravitational interaction at large scales. Many theoretical models therefore exist, that try to model this unknown type of matter. These theories generally assume that this form of matter is composed of particles, just as the known matter, and describe a new type of interaction through which these dark matter particles interact with the Standard Model particles.

If the dark matter indeed interacts with ordinary matter through a new force, mediated by a new particle, it can be searched for, and many existing theories can be tested. A myriad of experiments are currently looking for dark matter, and can be divided into three categories. Firstly, direct detection experiments take advantage of the dark matter particles that should be present in a halo surrounding our galaxy and try to measure the recoil of nuclei generated by dark matter particles passing through the Earth and scattering off the ordinary matter. The detectors used for this type of experiment are mostly located underground and are well shielded from radiation, though a few are airborne or space experiments. Indirect detection experiments on the other hand look for particles or radiation coming from the annihilation of dark matter particles in dense regions such as the galactic centre. These searches are studying gamma rays, neutrinos, electrons and positrons, or radio emissions. Finally, dark matter particles could potentially also be produced and detected at particle colliders. One of the direct detection

permeating?

? which ones? curiosity

experiments observed evidence pointing to the existence dark matter particles, but so far no conclusive observations have been made.

At collider experiments, such as ATLAS and CMS, dark matter candidates are often looked for by focusing on missing energy. Indeed, if the dark dark matter is assumed to interact weakly with the ordinary matter, it will be able to leave the detector unnoticed. However, these so-called weakly interacting massive particles (WIMPs), can be detected when they recoil against another object. Some examples of such collider searches are the monophoton, monojet, monolepton, and mono-Higgs analyses, categorised based on the signature in the detector. Additionally, different signatures are obtained when the dark matter is for example produced in a cascade of decays. Also resonances in e.g. the dijet mass spectrum are looked for, as this could indicate the existence of a new dark matter mediator. In general, more and more analyses are adding dark matter interpretations to their results. This thesis describes two searches for dark matter performed using data from high-energy proton-proton collisions produced at a centre-of-mass energy of 13 TeV and recorded with the CMS detector. The first analysis is the so-called the monojet analysis, which investigates the existence of WIMPs as dark matter candidates. Conversely, the second search looks for dark matter in the form of strongly interacting massive particles (SIMPs).

In Chapter 2, an overview of the Standard Model is given, as well as a short description of a few of its shortcomings. Furthermore, a summary of the existing evidence for dark matter, together with a concise review of popular dark models and a brief description of the operational or developing dark matter experiments are given. As this thesis covers dark matter searches that are performed using the CMS detector, located at one of the collision points of the LHC at CERN, more details concerning this accelerator and particle detector are summarised in Chapter 3. In Chapter 4, the procedure to reconstruct the collisions occurring inside the detector is detailed, as well as the necessary simulations of the predicted signal, which are needed in order to design a search for a particular dark matter candidate and to tune the analysis to the expected signature in the detector. The required techniques for this are described, with more details on the specific simulations needed for the searches covered in this thesis. The two complementary dark matter searches are described in Chapters 5 and 6.

were ? :-)

The monojet analysis, covered in Chapter 5, is one of the flagship analyses which are expected to quickly detect potential dark matter candidates, for a broad range of models. I contributed to this analysis by improving the prediction of the main background, coming from the invisible decay of Z bosons into neutrinos, produced in association with one or more jets. The used strategy for the background estimation is detailed in Section 5.4 and the resulting impact on the sensitivity of the analysis is shown in Section 5.7. In the strongly interacting massive particle (SIMP) analysis, described in Chapter 6, the dark matter candidates and the Standard Model particles interact strongly through a new force ~~force,~~ carried by a new, light mediator. The signature therefore does not consist of missing energy, but instead trackless jets are created due to the interaction of these SIMPs in the dense material of the calorimeters in the detector. First, a phenomenological study of the dark matter model was performed and published [6], and subsequently the search was carried out using data collected by CMS in 2016. This work, together with my contribution to the monojet analysis are the main topics of my PhD research. The monojet search provides new, stronger limits on WIMP dark matter candidates, while the trackless jets search rules out a new dark matter model which ~~was not~~ tested at colliders yet.

well, only if DM was to be found. Maybe "designed"?

mention something about tracker work? (Kevin mentions b-tagging)

] Yes!

neutral

had not been

Either here or make a separate page with the previous paragraph's contents + tracker, like e.g. Dominic had done.

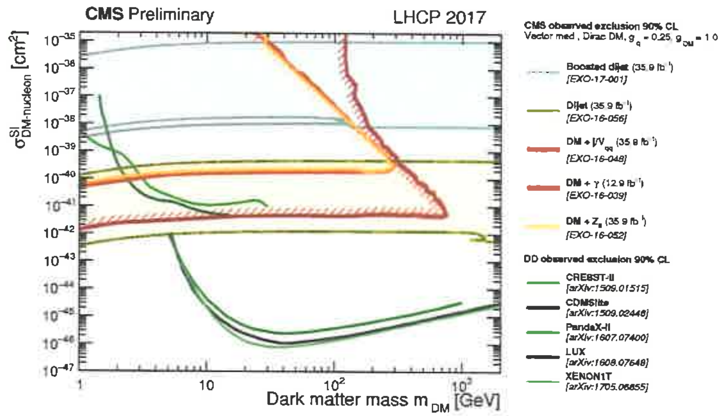

Figure 2.9: A comparison of CMS results to direct detection experiments in the $m_{DM} - \sigma_{SI}$ plane. The limits are shown at 90% CL. The shown CMS contours are for a vector mediator with Dirac dark matter and couplings $g_q = 0.25$ and $g_{DM} = 1.0$. The spin-independent exclusion contours are compared with the XENON1T 2017, LUX 2016, PandaX-II 2016, CDMSLite 2015 and CRESST-II 2015 limits, which constitutes the strongest documented constraints in the shown mass range. It should be noted that the CMS limits do not include a constraint on the relic density and also the absolute exclusion of the different CMS searches as well as their relative importance will strongly depend on the chosen coupling and model scenario. Therefore, the shown CMS exclusion regions in this plot are not applicable to other choices of coupling values or models. Figure taken from [78].

2.2.4 From EFTs to simplified models

In order to efficiently look for dark matter at colliders, effective field theories (EFTs) have been used extensively to model the dark matter signal. The EFT models assume the dark matter production can be described as a contact interaction defined by an effective mass scale and coupling structure. This contact interaction is for example illustrated in Figure 2.10 for the monojet final state where the dark matter pair is produced in association with an initial state radiation jet. The resulting signal models can then be classified by coupling structure, and the effective scale Λ can be extracted for a specific model, defining both the coupling strength and the scale of the theory. An EFT is characterized by a total of 3 parameters, the dark matter mass, the EFT scale, and the EFT coupling structure. However, this approach has several limitations [79–81]. First, it implicitly assumes that the dark matter production happens through a heavy mediator, which is not resonantly enhanced at the LHC. Additionally, for low enough effective scales, the EFT breaks down. Finally, the incompleteness of the EFT makes a comparison with direct detection experiments difficult or inconsistent. Due to the limitations of EFTs, there has been a trend in the past few years to instead use simplified models which allow for a fair comparison to low energy underground direct detection experiments. In a simplified model the effective scale is then replaced by a physical mediator. The resulting models contain six parameters that can be scanned to search for dark matter, namely the coupling structure, the dark matter mass, the mediator scale, the couplings to the Standard Model and the dark matter, and the mediator width. This transition has been overseen by the joint ATLAS/CMS dark matter forum [82] by establishing a well defined set of benchmark models to enable the combination of different channels and the recasting of dark matter models against direct and indirect detection searches.

add ref to DMWG doc

In the two dark matter searches covered in this thesis, the results have been interpreted in terms of simplified models. The monojet search described in Chapter 5 includes several simplified models recommended by the dark matter forum. Four types of mediators are considered, i.e. a vector, axial, scalar, and

not a free parameter, unless decays to other particles are allowed

usually "axial-vector" is used, "axial" being the jargon shorthand

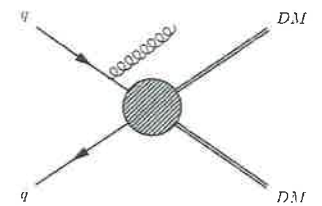

Figure 2.10: Illustration of EFT dark matter production in the monojet final state.

pseudoscalar mediator. In the case of a scalar or pseudoscalar coupling, the production mode is dominated by gluon fusion. As illustrated in the right diagram of Figure 2.11, the scalar is produced through a t or b quark loop. A Yukawa coupling is assumed for the coupling of the mediator to Standard Model particles, proportional to the mass of the particle. For a vector or axial mediator, the production happens through the fusion of two quarks into a heavy mediator, similarly to the Z and W boson production. The coupling to quarks and potentially leptons is taken to be unity, and universal for all flavours. For all mediator types, the coupling to the dark matter particles is assumed to be unity. In addition, the minimal width assumption is made, implying that the mediator couples to all Standard Model and the dark matter particle and no extra particles are introduced. If such particles would be present, the width would increase and the sensitivity of the analysis would be reduced. A scan is then performed over the mass of the dark matter candidate and the mass of the mediator.

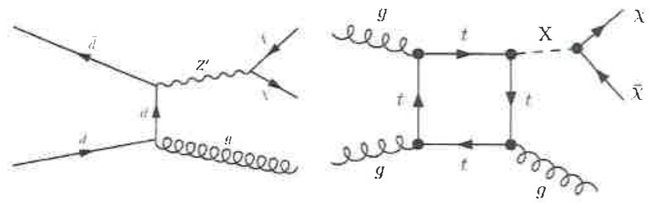

Figure 2.11: The vector (left) and scalar (right) production diagrams in the monojet final state.

not a problem

staat enkel in de AN, niet in de PAS. Furthermore, some non-standard dark matter models are investigated as ~~well in the monojet analysis~~, namely a complete simplified scalar model, known as the inert two Higgs doublet model and a baryon number violating dark matter model which can explain electroweak baryogenesis [83,84], known as non-thermal dark matter. In contrast to the simplified models, these theories are completed theories. The first consists of an extended scalar field theory, while the second consists of resonant production induced by flavour changing neutral currents.

you could leave this out if you don't show results

The SIMP simplified model on which the trackless jets analysis detailed in Chapter 6 is a specific simplified model which is not part of the models recommended by the dark matter forum. It is described in more detail in Section 2.3.

2.3 Strongly Interacting Massive Particles

As no observation of dark matter has been made so far, despite many searches probing the more popular models described in the previous section, many scenarios now venture beyond minimal models or give up basic assumptions for the WIMP. In the following model, which is studied in this thesis, the interaction cross section of the dark matter with normal matter is so high that the particles are no longer WIMPs,

4

Event Simulation and Reconstruction

In order to use the recorded data, the obtained signals coming from various parts of the detector must be reconstructed to be able to identify the particles in the event. Additionally, to compare the experimental results with theory, events are generated and the resulting signals in the detector are simulated, as detailed in Sections 4.1 and 4.2, respectively. The event reconstruction is detailed in Section 4.3. Finally, some details about the simulation of SIMPs are given in Section 4.4.

4.1 Event generation

the colliding

The event structure at the LHC is complicated by the composite nature of protons, as demonstrated in Figure 4.1. This sketch shows the hard interaction in red, with a tree-like structure surrounding it, representing the ensuing shower. In this hard scattering, the quark or gluon constituents of the protons, called partons, will interact according to a so-called parton distribution function (PDF), which is determined by the parton's momentum fraction and the momentum transfer. Due to their colour charge, the partons involved in the hard interaction will induce parton showers consisting of a cascade of radiation from QCD processes. This is shown in blue for the incoming partons and in red for the outgoing partons. The produced partons will also hadronize due to colour confinement, as illustrated in green, with hadron decays in dark green and radiated photons in yellow. Finally, the purple interaction represents a second interaction between the proton remnants. Next to interactions between the proton remnants, additional activity in the event can come from multiple parton interactions and pileup. All these aspects must be taken into account when generating events, as detailed below.

the same

Hard scattering

In the hard interaction, two partons of the colliding protons, will interact with a certain probability at a given momentum transfer. This is parametrized by the PDFs $f(x, Q^2)$, were x is the proton's momentum fraction and Q^2 is the momentum transfer scale. Experimentally determined PDFs are available from various groups, including e.g. CTEQ [111], MRST/MSTW [112], and NNPDF [113]. An example of such PDFs obtained by the NNPDF group is shown in Figure 4.2. The PDFs are then convoluted with the matrix element of the hard scattering, which is the process of interest where the two colliding partons create high-energetic final state particles. This is done using an event generator, such as MADGRAPH5_aMC@NLO [114] and POWHEG [115]. With MADGRAPH5_aMC@NLO the matrix element can be calculated at tree-level or leading order (LO), and since the addition of aMC@NLO at next-to-leading order (NLO) as well. This generator was used to produce most of the background processes for the Monojet analysis detailed in Chapter 5 and for the SIMP signal used in Chapter 6. POWHEG is able to generate events using

potentially underline the distinction between valence and sea quarks, and/or the growth of the gluon PDF at high Q^2

the ME for many processes

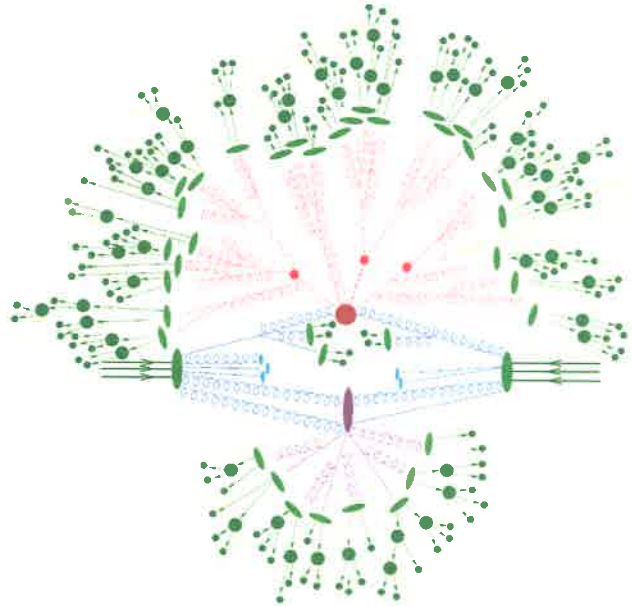

Figure 4.1: Illustration of an event showing the hard scattering, parton shower, hadronization, and underlying event. Figure taken from [110].

NLO computations, but only for a relatively limited number of physics processes. This generator was used to produce the monojet signal samples and the background processes from single-top production. Since NLO calculations are more time-consuming, one can instead use the less precise method of scaling a LO cross section to the NLO level by using a so-called k-factor, defined as the ratio of the NLO and LO cross sections. However, these k-factors often need to be determined as a function of the relevant kinematic variables as they depend on the kinematic phase space and the probed energy scale.

Parton showering

Since the colliding partons have a colour charge, the hard scattering will be accompanied by a cascade of radiation from QCD processes. The partons will for example radiate soft gluons or split into two collinear partons. This radiation can originate from the incoming partons, which is referred to as initial state radiation (ISR), or the outgoing partons in the final state, the so-called final state radiation (FSR). The perturbative evolution of the cascade can be modelled using the DGLAP (Dokshitzer-Gribov-Lipatov-Altarelli-Parisi) equations [116–118]. These equations describe the time evolution of the probability of a 'mother' parton to split into 'daughter' partons at an energy scale Q^2. The momentum of the mother is then divided among the daughter partons, which can in turn split into other partons at a lower Q^2 scale. The cascade continues down to an energy scale Λ_{QCD} where the strong coupling constant becomes unity. The resulting number of jets can vary depending on the modelled process.

I'm not sure that's the definition of Λ_{QCD}: unity, 4π...

you could say: where the strong interaction becomes non-perturbative.

Hadronization

The next step after the showering is the hadronization of the coloured particles produced in the

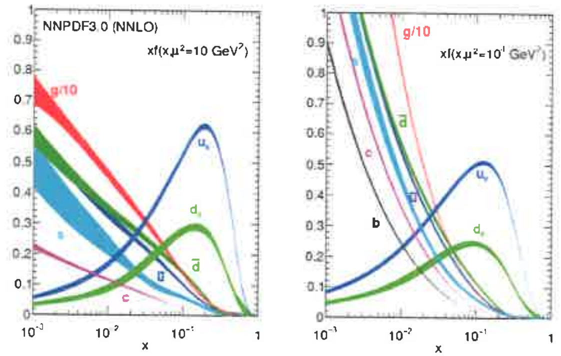

Figure 4.2: The parton distribution functions times the momentum fraction x at energy scales $10\,\mathrm{GeV}^2$ (left) and $10\,000\,\mathrm{GeV}^2$ (right), obtained in NNLO NNPDF3.0 global analysis. Figures taken from [113].

parton shower, transforming them into colour-neutral hadrons. Since this happens at low energy scales where the perturbative approach of QCD is not valid, phenomenological models have to be used. For most of the processes considered in this thesis, the showering and hadronization is done with PYTHIA 8 [119], using a standard set of parameters which were tuned to reproduce the experimental data. In PYTHIA, the string Lund model [120] is used, based on string fragmentation. This model starts from the idea of a string connecting a quark q and an antiquark $\bar{q}$, following the assumption of linear confinement. As the two quarks move away from each other, the string stretches and the potential energy stored in the string increases. The increase in potential energy is assumed to be proportional to the distance between the quarks. When the energy becomes sufficient to produce a new pair of quarks $q'\bar{q}'$ with mass m, the string breaks and the original quark pair is split into two new pairs, $q\bar{q}'$ and $q'\bar{q}$. If the invariant mass of the new strings is large enough, the same process is repeated, leading to a new break-up. This procedure continues until only colour-neutral hadrons with an on-shell mass remain.

Additional activity in the event

In addition to ISR and FSR, also beam remnants and multiple parton interactions give rise to additional activity in the event, referred to as the underlying event. After the partons participating in the hard scattering are extracted, the remainder of the protons have a non-zero colour charge. The creation of additional hadrons during the hadronization is therefore possible. Multiple parton interactions represent additional interactions which can take place between other incoming partons. As the probability for an additional hard interaction to occur is rather small, the activity from multiple parton interaction is typically much less energetic than the hard interaction, producing mostly low energetic hadrons. Finally, additional collisions between other protons in the same bunch crossing or from a previous bunch crossing, respectively referred to as in-time and out-of-time pileup, add extra activity in the event. The pileup distribution is for example shown in Figure 4.3 for QCD dijet events recorded in 2016, and is compared to simulated QCD events. This shows that there were about 20 collisions per bunch crossing on average. Typically, the simulation does not completely agree with the data and needs to be reweighted in order to match the data.

also from future bunch crossing.

4.1.1 Simulation of the monojet signals

axial-vector

In the monojet analysis, the simplified models described in Section 2.2.4 are considered. The used signal samples were generated with POWHEG, which can generate NLO vector and axial mediator production and LO scalar and pseudoscalar production. The samples were also produced at LO with MCFM [121]

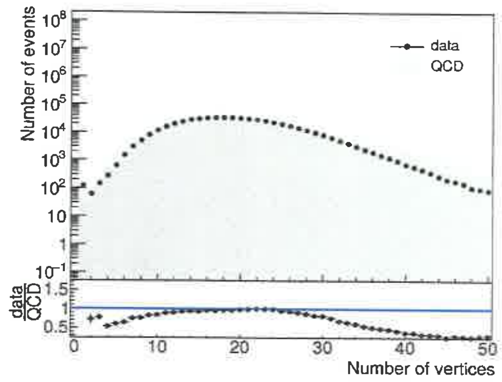

Figure 4.3: The ~~pileup~~ distribution of QCD dijet events recorded in 2016 compared to simulated QCD events.

number of vertex

as a cross check. The scanned mediator masses are $m_\phi = 10, 20, 50, 100, 200, 300, 500, 1000, 2000, 10\,000\,\mathrm{GeV}$, for dark matter masses of $m_\chi = 1, 10, 50, 100, 150, 500, 1000\,\mathrm{GeV}$, with $m_\chi \leq m_\phi$.

4.2 Detector simulation

After being generated, the collision events are passed on to the CMS detector simulation, which is based on the GEANT 4 [122] simulation toolkit. This toolkit provides a description of the interaction between particles and the detector material, including effects such as bremsstrahlung of charged particles, photon conversions, energy loss of charged particles by ionization, and the showering of electrons, photons and hadrons in the calorimeters due to interaction with the material. The CMS simulation package contains the geometry of the detector with all the sensitive layers designed to detect the traversing particles, as well as the dead material regions consisting of e.g. support structures, cables and cooling pipes. A precise map of the magnetic field is also included in order to simulate the curvature of the charged particles correctly.

Next, the impact of the detector, coming from the electronic response produced by the hits in the active detector material, the digitization, the data transmission, and any reconstruction performed in the electronics such as zero-suppression or cluster reconstruction, is simulated. In this way, an event content similar to the output of the real detector is obtained. At this point the effect of pileup is also included by adding detector hits of generated proton-proton interactions on top of the hits resulting from the main interaction. Most of the simulated event samples used in this thesis are processed using this detector simulation. However, the interaction of new particles that can arise from specific theory models is not always readily described in GEANT. This is the case for the signal samples used in the analysis described in Chapter 6, so an additional step was needed in order to simulate strongly interacting massive particles (SIMPs) in the CMS detector, described in Section 4.4.

4.3 Event reconstruction

Once the detector response has been simulated, the obtained events can be reconstructed. The same method is applied for these simulated events and for data coming from the detector. First, the reconstruction of tracks is performed, with a specific track reconstruction for electrons and muons. Furthermore, described in Section 4.3.6

the calorimeter deposits, generated by electrons, photons, and hadrons, are grouped into clusters. Additionally, the reconstruction is further improved by using the so-called particle flow (PF) algorithm. This algorithm greatly improves the performance for jet and hadronic τ decay reconstruction, missing transverse energy momentum determination, as well as electron and muon identification. Finally, the obtained particle flow (PF) candidates are clustered into jets, and the missing transverse energy can be derived.

4.3.1 Track and vertex reconstruction

The tracks of charged particles going through the CMS tracker are reconstructed with an iterative tracking approach. This is used to cope with the high occupancy and consequently high combinatorics. Additionally, the first iterations search for tracks with less possible combinations, such as tracks with many pixel hits or a high momentum. After every iteration, the hits associated with the found track are removed to reduce the combinatorics. Each iteration consists of four steps:

1. **Seed generation.** In this first step hits are combined into seeds for the subsequent track finding. In the initial iterations pixel triplets are used, then pixel pairs, in order to take gaps or non-working modules into account. Next, mixed pixel/strip triplets are taken, and finally strip-only seeds are used. These additional iterations improve the acceptance in p_T and in displacement with respect to the primary vertex.

2. **Track finding.** The seeds are used as starting point for a Kalman filter algorithm. This method extrapolates the seed trajectory outward to the next layer, taking into account potential energy loss and multiple scattering. If compatible hits are found in the next layer, the parameters of the trajectory are updated. This process continues until the outermost layer of the tracking system. Using this method, a given seed can generate multiple tracks, or different tracks can share hits. A trajectory cleaner therefore determines the fraction of hits the tracks have in common and discards the track with the lowest number of hits when there are too many shared hits. If both tracks have the same number of hits, the track with the largest χ^2 value is removed.

3. **Track fitting.** The track parameters are then refitted using a Kalman filter and smoother, taking all hits determined in the track finding step into account.

4. **Track selection.** Finally, the tracks are selected based on quality requirements, such as the number of layers that have hits, the χ^2/dof, and the distance to a primary vertex. This greatly reduces the fraction of reconstructed tracks that are fake.

The procedure has changed, but I would advise you to keep as is, since the newest description is very confusing.

at very high momentum

The performance of the track reconstruction is excellent, and a high track-finding efficiency is obtained [123] while keeping the rate of fake tracks negligible. The highest tracking efficiency is obtained for muons, which traverse the full detector volume and have an improved momentum resolution due to tracking information from the muon detectors giving a long lever arm. For isolated muons with p_T between 1 and 100 GeV the tracking efficiency is higher than 99% for the entire η coverage of the tracker, as can be seen from the left plot in Figure 4.4. The p_T resolution is about 2-3% for a muon with p_T = 100 GeV up to $|\eta| < 1.6$, but worsens for higher pseudorapidities. Different types of particles interact differently with the detector material. Charged hadrons, for example, are also subject to elastic and inelastic nuclear interactions and have a tracking efficiency of 80-95% depending on pseudorapidity and transverse ~~momentum~~, as shown in the right plot of Figure 4.4.

Finally, the primary vertex is reconstructed from the tracks. Since the collisions happen between bunches of protons, multiple protons will be colliding at the same time. The extra collisions, next to the potentially interesting collision, are referred to as pile-up interactions. The particles generated in these collisions are all detected simultaneously and form a ~~challenge to disentangle~~ them from the particles coming from the to be studied interaction.

plural?

The reconstruction is done in 2 steps: first the tracks that appear to originate from the same interaction vertex are clustered, then a fitting procedure computes the vertex parameters and assigns a weight to each associated track, reflecting the probability that it corresponds to the considered vertex. Figure 4.5 shows the reconstruction efficiency and the resolution of the primary vertex. The more tracks, the better the vertex is constrained and thus the better the resolution.

you have discussed pileup before, so only define first time round

cosmetic note: consistency pileup ↔ pile-up

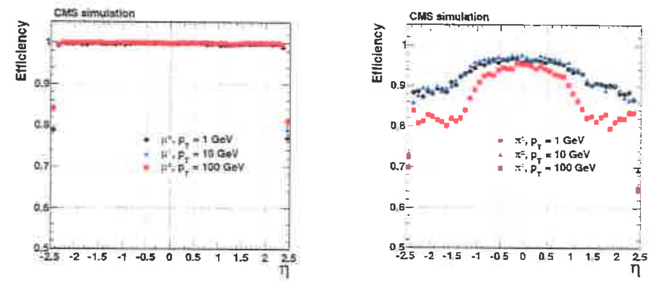

Figure 4.4: The muon efficiency (left) and pion efficiency (right) as a function of pseudorapidity, for multiple transverse momenta. Figures taken from [123]

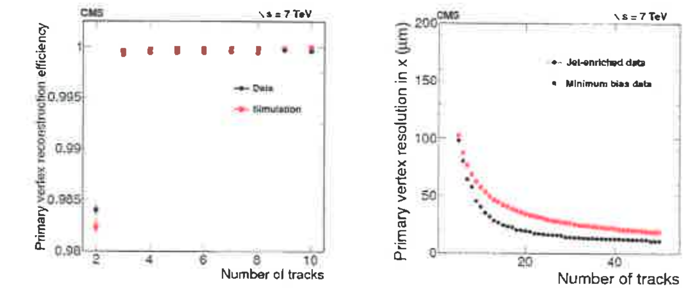

Figure 4.5: The primary vertex reconstruction efficiency (left) and resolution (right) as a function of the number of tracks associated to it. Figures taken from [123]

The vertex with the highest $\sum p_T^2$ is chosen as primary vertex, where the sum runs over the tracks associated to the vertex following the application of a deterministic annealing filter which assigns weights to sufficiently high-quality tracks that enter the vertex fit [123]. While in events with jets many tens of high-momentum tracks can usually be associated to a primary vertex, thus making primary vertex finding almost fully efficient and pure, in the case of a pair of neutral jets, produced for example by SIMPs, this is not the case any more. The underlying event and potentially initial state QCD radiation can still provide some tracks, but in extreme cases a wrong vertex is chosen, arising from a hard pileup collision.

4.3.2 Electron and isolated photon reconstruction

Electrons are reconstructed using information from both the tracker and the calorimeters. Due to the large amount of material present in the tracker, electrons will emit bremsstrahlung photons, and photons will often convert into e^+e^- pairs, which can again radiate bremsstrahlung photons.

For electrons, a Gaussian-sum filter (GSF) [124] candidate is taken as starting point. This GSF candidate is obtained using 2 different methods to reconstruct the electron track from the hits in the tracker, which should gather all radiated energy from the electron. First, the ECAL-based approach is used, grouping ECAL clusters into superclusters. These superclusters collect the energy of the electron and the bremsstrahlung photons in a small η window and a large ϕ window, taking the bending of the electron track in the magnetic field into account. The supercluster energy and position is then used to estimate the position of the corresponding hits in the tracker layers. Subsequently, the tracker-based approach is used to find electrons missed by the ECAL-based method. In this case, the tracks from the iterative tracking with transverse momentum larger than 2 GeV are used. Additional requirements are placed on the number of hits and the χ^2 of the fit, and the specific electron tracking is performed, using a GSF fit, which is more adapted to electrons than the Kalman filter used in the iterative tracking, as it describes the energy loss in each tracker layer. The electron seeds obtained with both methods are merged and used as input for the full electron tracking. The obtained electron tracks are then linked to ECAL clusters by the PF algorithm, as described in Section 4.3.6. In the case of isolated photons, a candidate is seeded from an ECAL supercluster with transverse energy larger than 10 GeV which is not linked to a GSF track.

The total energy of the accumulated ECAL clusters is corrected for the energy that was lost in the process of reconstruction, using analytical functions of the energy and pseudorapidity. The applied corrections can be as large as 25%, at low transverse momentum and at $|\eta| = 1.5$, where the material density in the tracker is largest. The energy of the electron is then obtained from a combination of the corrected energy and the momentum of the GSF track, while the direction of the electron is taken from the GSF track. For photons, the corrected energy and the direction of the supercluster are used.

4.3.3 Electron and photon identification

multivariate

In general, the electron and photon candidates must satisfy identification criteria to be retained. In the case of electrons two methods for identification are available: a cut-based identification or a boosted decision tree (BDT) combining fourteen variables including the amount of energy radiated and the ratio between the energies gathered in HCAL and ECAL. In the monojet analysis described in Chapter 5, the former is used. In method, four different working points are defined, denoted as "tight", "medium", "loose", and "veto", with varying signal efficiency and background rejection. For the electron veto the loose selection is used, while a tight identification is required on one electron to select the events in the dielectron and single electron control regions.

Is this ok English? Did you mean: "In this method"?

add units

variable	loose		tight	
	barrel	endcaps	barrel	endcaps
full 5x5 $\sigma_{i\eta i\eta}$	< 0.0114	< 0.0352	< 0.0101	< 0.0279
$\lvert\Delta\eta_{in}\rvert$	< 0.0152	< 0.0113	< 0.00926	< 0.00724
$\lvert\Delta\phi_{in}\rvert$	< 0.216	< 0.237	< 0.0336	< 0.0918
H/E	< 0.181	< 0.116	< 0.0597	< 0.0615
relative isolation	< 0.126	< 0.144	< 0.0354	< 0.0646
1/E - 1/p	< 0.207	< 0.174	< 0.012	< 0.00999
$\lvert d_{xy}(vtx)\rvert$	< 0.0564	< 0.222	< 0.0111	< 0.0351
$\lvert d_z(vtx)\rvert$	< 0.472	< 0.921	< 0.0466	< 0.417
expected inner missing hits	≤ 2	≤ 3	≤ 2	≤ 1
pass conversion veto	yes	yes	yes	yes

Table 4.1: Loose and tight electron identification criteria. The isolation is computed in a cone of $\Delta R < 0.3$ around the electron.

Similarly, for the photons, both a cut-based identification and a multivariate analysis can be used. Both for the monojet and the SIMP analysis, described in Chapters 5 and 6, the cut-based identification is used. Three standard working points are provided, denoted as "loose", "medium", and "tight", with an average efficiency of 70%, 80%, and 90%, respectively. In both analyses, the loose identification is used for the photon veto. In the SIMP analysis, the event is only rejected when the identified photon is within a cone of $\Delta R < 0.1$ of one of the two leading jets. Additionally, the photon veto is extended to reject events containing jets with a large photon energy fraction and unidentified photons. This happens for instance when there is a photon conversion in the tracker. The event is therefore rejected when the jet photon energy fraction is larger than 0.8, the photon is not identified by the loose criteria, and ~~the~~ a conversion is matched to the photon within $\Delta R < 0.2$ and has $p_{T,conv}/p_{T,\gamma} > 0.3$. Lastly, the two jets are also required to have a neutral electromagnetic energy fraction lower than 0.9, corresponding to one of the standard tight jet identification requirements mentioned in Section 4.3.7. The full jet ID is not applied, since the requirements on e.g. the neutral hadronic energy fraction and the charged multiplicity

do you need to mention this here? It doesn't relate to photons and you repeat it later with the jets → ah maybe not but it would fit better on p45, no? (end of 4.3.7)

would reject the signal events. Finally, the monojet analysis uses a photon + jets control region as well. These events are selected by applying the tight photon identification.

units

variable	loose		tight
	barrel	endcaps	barrel
full 5x5 $\sigma_{i\eta i\eta}$	< 0.0102	< 0.0274	< 0.0102
H/E	< 0.05	< 0.05	< 0.05
charged hadron isolation	< 3.32	< 1.97	< 1.37
neutral hadron isolation	$< 1.92 + 0.014 \times p_T + 1.9 \times 10^{-5} \times {p_T}^2$	$< 11.86 + 0.0139 \times p_T + 2.5 \times 10^{-5} \times {p_T}^2$	$< 1.06 + 0.014 \times p_T + 1.9 \times 10^{-5} \times {p_T}^2$
photon isolation	$< 0.81 + 0.0053 \times p_T$	$< 0.83 + 0.0034 \times p_T$	$< 0.28 + 0.0053 \times p_T$
conversion safe electron veto	yes	yes	yes

Table 4.2: Loose and tight photon identification criteria. The isolation is computed in a cone of $\Delta R < 0.3$ around the photon.

4.3.4 Muon reconstruction

Muon tracking is performed using 2 complementary approaches. The first method starts from standalone muons, which are reconstructed from hits in the muon detectors only using pattern recognition. The standalone muons are then matched to tracks in the tracker, and the hits are combined to form a global muon track. This global muon fit improves the momentum resolution compared to the tracker-only fit at muon momenta larger than 200 GeV.

For momenta below 10 GeV, muons often fail the global muon conditions which require the muon to penetrate through more than one muon detector plane, due to the large multiple scattering in the return yoke. In this case, tracker-only muon reconstruction is more efficient since it only requires one muon segment. Each track in the tracker with a transverse momentum larger than 0.5 GeV and a total momentum larger than 2.5 GeV is therefore extrapolated to the muon system and if at least one matching track segment is found, it is retained as muon candidate.

Within the geometrical acceptance of the muon system about 99% of the muons are reconstructed, either as global muon or as tracker muon and frequently as both. Global and tracker muons that share the same track inside the tracker are merged into a single candidate. Muons that are only reconstructed as standalone muons have a worse momentum resolution compared to the global and tracker muons. These standalone muons are however only considered in the further reconstruction when the fit is of high quality and is associated with a large number of hits in the muon system.

Charged hadrons can be misreconstructed as muons if e.g. a part of the hadron shower reaches the muon system. In order to improve the muon identification, the PF muon identification algorithm described in Section 4.3.6 also matches energy deposits in the ECAL and HCAL with the muon track.

4.3.5 Muon identification

standard

When using muons for physics analysis, some identification criteria are generally applied in order to ensure the quality of the muons. There are several levels of identification, denoted as "tight", "medium", and "loose", which provide a trade-off between muon identification efficiency and misidentification. In general, the tight and loose identification are the most widely used identification criteria.

The loose identification only requires the muons to be either global or tracker-only muons, and to be identified as a PF muon. As a result, it is highly efficient for both prompt muons and muons from quark decays. In analyses with prompt muons, this identification is therefore often complemented by an impact parameter cut. — associating the muon to the primary vertex.

not clear what this means: just PFreco or are there implicit cuts? maybe comment this in some way.

For the tight identification, the muon is required to be a global muon and to pass the PF muon identification. The normalized χ^2 of the global muon track fit should be smaller than 10 to suppress hadronic punch through and muons from decays in flight. To further suppress these contributions at least one muon chamber hit should be included in the global muon track fit and muon segments should be found in at least two muon stations. Cosmic muons and tracks from pileup are suppressed by requiring the tracker track to have $|d_{xy}| < 2$ mm and $|d_z| < 5$ mm, with d_{xy} the traverse impact parameter and d_z the longitudinal

hadron

Table 4.1 has the abs. value

momentum then corresponds to all undetected particles in the event, and can be calculated from the vectorial sum of the transverse momenta of all the observed final state particles:

$$\vec{E}_T^{miss} = -\sum \vec{p}_T, \tag{4.4}$$

where the sum runs over all reconstructed PF particles. The variable that is generally used in particle physics analyses is the norm of the missing transverse momentum:

$$E_T^{miss} = |\vec{E}_T^{miss}|. \tag{4.5}$$

A notable example of particles leaving no hits or energy deposits behind are neutrinos, as they are neutral and weakly interacting and will therefore traverse the entire detector unhindered. Other hypothetical neutral weakly interacting particles, which are being searched for in many physics analyses, would escape the detector without producing hits as well.

4.4 Simulation and reconstruction of the SIMP signal

For the generation of the SIMP signal, the model Lagrangian given in equation 2.17 is implemented in FEYNRULES 2.0 [129]. The matrix element is then calculated at LO and events are generated using MADGRAPH 5. The subsequent parton shower and hadronization is done with PYTHIA 8, using tune CUEP8M1. Several samples were produced, with SIMP mass $m_\chi = 1, 10, 100, 200, 400, 700, 1000$ GeV. The corresponding productions cross sections are given in Table 4.3. Next, the events are then simulated in the CMS detector using GEANT. However, the SIMPs are not included in the simulation, as these new particles are unknown in GEANT and their interaction with matter has not been implemented yet. In order to simulate the new dark matter candidates in the CMS detector two new approaches were implemented.

you could merge this into 4.1.1. This feels a bit lost here now. It would not break the flow of this section.

m_χ [GeV]	$\sigma_{\bar{\chi}\chi}$ [pb]
1	4.46
10	4.40
100	2.55
200	0.790
400	0.0743
700	0.00485
1000	0.000571

Table 4.3: Production cross section for each SIMP mass, after $|\eta_\chi| < 2.5$ and $p_T^\chi > 200$ GeV generator level cuts.

In the first approach, the SIMPs were incorporated by adding an additional step to the standard reconstruction described in Section 4.3. In this additional step the SIMPs are directly converted to neutral PF candidates and merged with the rest of the PF candidates. Additionally, the new PF candidates are smeared with jet energy resolution (JER) distributions obtained from a sample produced using neutrons instead of SIMPs. Neutrons were chosen because of their resemblance to the SIMPs as single neutral particles generating a hadronic shower. what sample? In Order to produce this sample, the same additional custom step is applied, but in this case the neutrons will also be correctly reconstructed by the standard reconstruction. The reconstructed PF candidates that are matched to the generated neutrons are therefore removed before injecting the converted generated neutrons to the collection of PF candidates. The applied JER distributions are derived by comparing the resulting uncorrected PF jets with the corresponding neutrons in sample produced with the standard reconstruction using the the full GEANT simulation. The resolution is computed in bins of η and p_T, and an example is shown in Figure 4.7 for central neutrons with low and high transverse momentum.

??? I can't follow this.

repetition

After applying this smearing, the P2PF jets are processed with the standard sequence of CHS, jet clustering, L1FastJet, and L2/L3 corrections described in Section 4.3.7.

In order to validate this method, the custom and standard neutron samples are used to compare the two leading generator-level jets to the new jets from the custom sample, denoted as P2PF jets, and the PF

undefined

comes too late

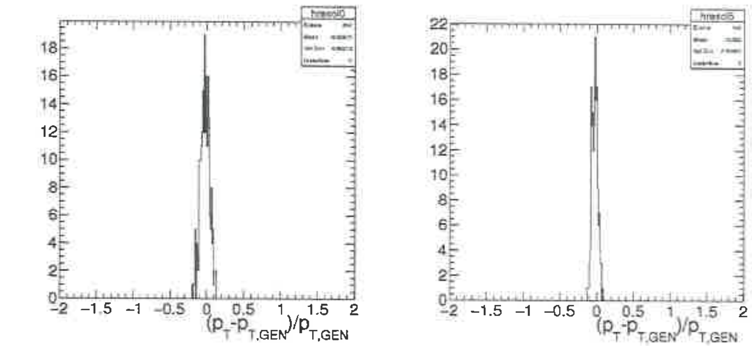

Figure 4.7: The jet energy resolution of neutrons with $0 < |\eta| < 0.5$ and 200 GeV$< p_T <$300 GeV (left) or 700 GeV$< p_T <$800 GeV (right).

jets from the standard sample. This is illustrated in Figure 4.8, where the p_T of the generated neutrons is shown on the horizontal axis, and the p_T of the reconstructed jet is shown on the vertical axis. The left plot shows the standard neutron sample produced with the full GEANT simulation, while the right plot shows the custom neutron sample where the neutron was directly converted into a neutral PF candidate. The JER distributions are also compared in Figure 4.9 and fitted with a Crystal Ball function, showing compatible parameters.

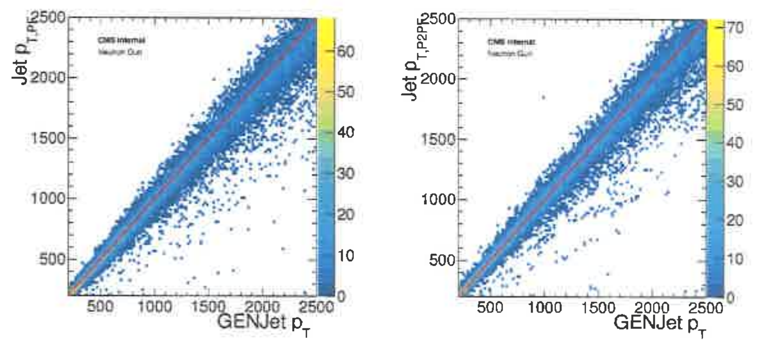

Figure 4.8: Comparison of the transverse momentum of the generator-level jets to the PF jets (left) and P2PF jets (right) in the region $0 < |\eta| < 0.5$ with jet energy resolution smearing, using a neutron sample.

This demonstrates that the procedure, where the JER distributions derived from a neutron sample are used to smear the PF candidates from generator-level SIMPs, can sufficiently accurately simulate SIMPs in a realistic detector, assuming SIMPs are neutron-like. However, since this procedure directly converts the generated SIMPs into PF candidates, the SIMPs do not interact in the Tracker and the resulting jets have a very steeply falling charged hadron energy fraction (CHF) distribution. This gives an optimistic image, which translates in a maximal signal efficiency.

may not be directly obvious to casual readers

The SIMP signal simulation and reconstruction was therefore further improved by moving to the second approach. In this method, the generated SIMP particles are not converted into neutral PF candidates, but they are instead replaced by neutrons, keeping the SIMP kinematics. The standard reconstruction and

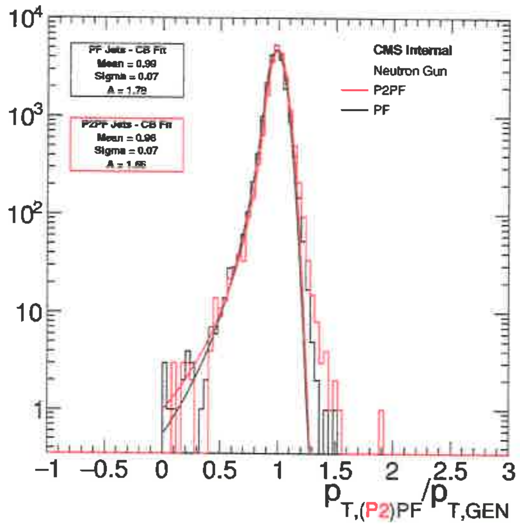

Figure 4.9: The jet energy resolution of the corrected P2PF jets (red) and PF jets (black), fitted with a Crystal Ball function.

full GEANT simulation can then be applied, since the neutrons are correctly recognized and simulated. In this case, interactions will happen inside the Tracker as well and the resulting jets will contain a larger CHF, as is shown in Figure 4.10.

This method gives a good approximation of a SIMP signal, since the shower generated by the SIMP is in principle contained inside the calorimeters, as the model described in Section 2.3 is constructed so that for a specific choice of couplings the SIMPs may be detected as regular hadrons. Although the considered SIMP-nucleon interaction is repulsive, this does not differ considerably from known attractive interactions at the probed high energies. The incoming SIMP hits a nucleon at rest in the calorimeter, breaking it up, and because of the large incoming momentum, there is a boost forward into the calorimeter and the shower starts. The cross section would therefore be identical for a repulsive or attractive interaction and the effect on the shower is negligible since the scattering angle is very small due to the momentum boost. Furthermore, the higher the momentum of the SIMP, the shorter the distance it travels to deposit its characteristic momentum. With the considered couplings, the depth containing a SIMP with 500 GeV momentum is below 1 m, within the calorimeter. Most of the energy will therefore be deposited in the first interaction with the material. Given the expected forward energy flow in the calorimeter shower, and the shower containment achieved by the choice of couplings in the simplified model, the shower induced by the SIMP interaction can to first order be modelled by the interaction of a high-momentum neutral hadron, like a neutron.

need citation to B. Zaldivar, ~~[illegible]~~ / private communication.

you can add that this 2nd method is what will be used in the analysis in Chapter 6.

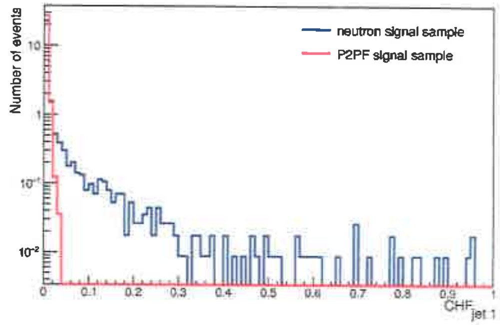

Figure 4.10: CHF distribution of the leading jet, for a signal sample produced with the first approach (red) and the corresponding sample produced using the second method (blue).