CHAPTER 5

OBJECT RECONSTRUCTION

This chapter describes the reconstruction and selection of the final state objects (leptons, jets, and b-tagged jets) used in this analysis. The reconstruction algorithm is explained in Section 5.1. The identification of jets and b-tagged jets is discussed in Section 5.2. The lepton selection is described in Section 5.3, with a summary of the requirements for electron and muons provided in Tables 5.1 and 5.2, respectively.

5.1 Particle flow reconstruction

In order to identify and reconstruct the particles produced in each collision, CMS uses a holistic reconstruction technique to correlate the elements from each subdetector (tracks and clusters) and construct a global picture of each event. This approach is referred to as particle flow (PF) reconstruction [31]. Since a particle generally interacts with multiple subdetectors, there are expected to be several PF elements associated with a given particle. The first step in the PF reconstruction is to link together the PF elements from the different subdetectors into sets of elements referred to as PF blocks.

In each block, the identification and reconstruction of the particles is performed in a specific order; as each particle is identified, the associated elements are removed from the block. First, muons are identified based on tracks in the muon chambers, tracks in the inner tracker, and lack of energy deposits in the calorimeters. Next, electrons and isolated photons are reconstructed from energy clusters in the ECAL, lack of significant energy clusters in the HCAL, and tracks (or lack thereof) in the inner tracker. Finally, charged hadrons, neutral hadrons, and non-isolated photons are identified through several iterative steps using information from the inner tracker and both calorimeters. Despite this thorough reconstruction algorithm, there may still be cases in which particles are misreconstructed, so a post-processing step is performed to mitigate these potential issues. A schematic representation of PF reconstruction is shown in Figure 5.1, with example signatures displayed for each particle type (muon, electron, charged hadron, neutral hadron, and photon).

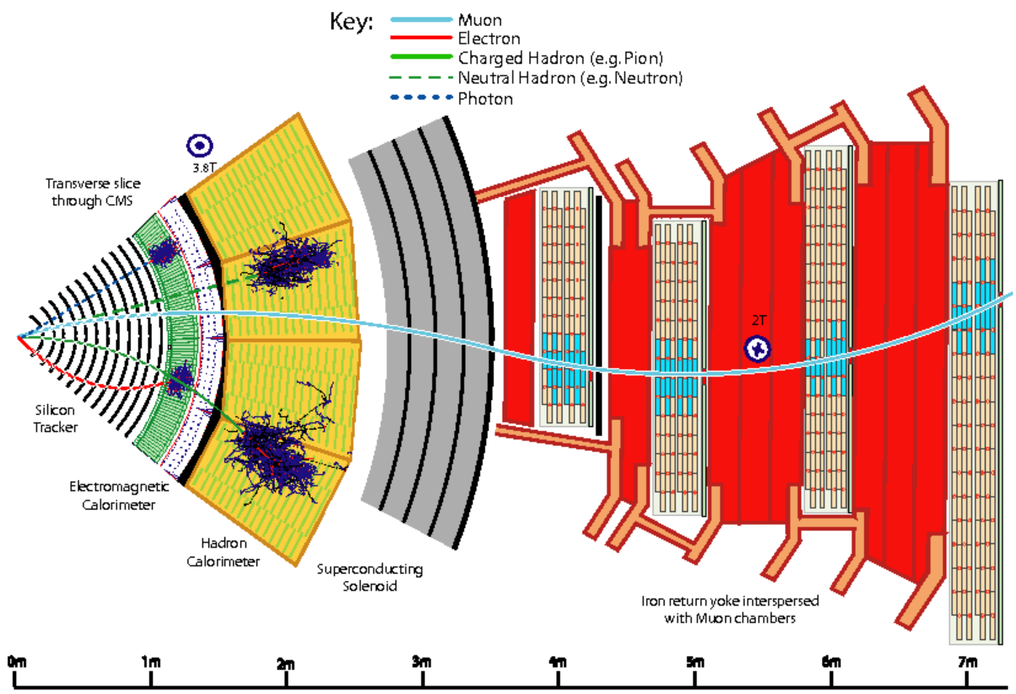

Figure 5.1. Schematic representation of the reconstruction of particles within CMS. An example signature for a muon, electron, charged hadron, neutral hadron, and a photon are shown. The PF algorithm uses the elements from each subdetector (tracks and clusters) to reconstruct and identify the particles, as described in Ref. [31] (from which this figure is taken) and summarized in Section 5.1.

5.2 Jets and b-tagging

After the PF reconstruction has identified all particles in the event, jets (collimated sprays of particles representing the experimental signature of quarks and gluons) are reconstructed using the anti-k_{T} jet clustering algorithm [13]. As explained in [13], the algorithm first defines a distance measure between the objects and between the objects and the beam; if the smallest distance measure is between two objects, they are combined into a single object, while if the smallest distance measure is between an object and the beam, the object is called a jet and removed from the list of objects. The distances are then recalculated and the process continues until all objects have been clustered into jets. The the anti-k_{T} clustering algorithm is implemented with the FASTJET package [25]. The distance parameter ΔR in the anti-k_{T} algorithm is set to 0.4. Charged particles arising from pileup interactions are excluded from the jet clustering algorithm using the charged hadron subtraction (CHS) technique [33].

In this analysis, we require all jets to have $p_T > 30$GeV and $|\eta| < 2.4$. The jets are cleaned using the loose (or better) leptons with a ΔR requirement of greater than 0.4. Jets that overlap with objects in the fakeable lepton collection (electrons and muons) are also removed. The definitions of loose and fakeable leptons are provided in Section 5.3.

In order to identify jets originating from b quarks, the DeepJet b-tagging algorithm is used [32]. The relatively long lifetimes of b quarks can lead to displaced secondary vertices, and the relatively large mass can lead to jets with relatively large p_T; these distinguishing characteristics are leveraged in the identification of b jets with the DeepJet algorithm. The efficiency and mistag rates depend on the working point utilized (i.e. the cut on the discriminant output, which ranges from zero to one). For this analysis, we make use of the loose and medium working points of the DeepJet algorithm, as defined by the CMS BTV POG for the UL datasets [1–4].

5.3 Lepton object selection

The lepton object aims to identify signal leptons arising from the hard scatter process. In synchronization with the $t\bar{t}H$ multilepton analysis [37], the selection is performed in three stages, referred to as the loose, fakeable, and tight selections. A general description of each stage is provided in Section 5.3.1, the requirements for which are summarized in Tables 5.1 and 5.2. The details of the observables that are used in each the selection requirements are defined and described in Section 5.3.2.

5.3.1 Overview of lepton object selection stages

With the goal of separating leptons from jets, the first stage of the lepton object selection makes use of a boosted decision tree (BDT) multivariate algorithm, trained by the CMS EGamma POG [5, 20, 22]. For this loose selection, we use the loose working point (which has an effeciency of 98% [5]). The loose selection requirements for electrons and muons also include cuts on the isolation of the lepton and the impact parameter of the lepton's track with respect to the primary vertex (PV). Since signal leptons are expected to be relatively isolated from hadronic activity and to originate from the hard scatter process, these requirements can help to distinguish the leptons of interest from background objects. The requirements for the loose selection are are summarized in the "Loose" columns of Tables 5.1 and 5.2. The loose leptons are used to identify and veto events with a low-mass resonance (as described in Section 6.2), and are also used for training the prompt-e and prompt-μ MVAs described below.

The next stage in the object selection is referred to as the fakeable selection. Building on the loose collection, the objects that pass the fakeable requirements are a subset of the objects that pass the loose selection. Listed in the "Fakeable" columns of Tables 5.1 and 5.2, these selection requirements result in a collection of leptons that are very close to the final selection of signal leptons; this collection is used for the estimation of the nonprompt background contribution (as described in Section 8.1).

TABLE 5.1

ELECTRON OBJECT SELECTION REQUIREMENTS SUMMARY

Observable	Loose	Fakeable	Tight
p_T	> 7 GeV	> 10 GeV[5]	> 10 GeV
$\lvert\eta\rvert$	< 2.5	< 2.5	< 2.5
$\lvert d_{xy}\rvert$	< 0.05 cm	< 0.05 cm	< 0.05 cm
$\lvert d_z\rvert$	< 0.1 cm	< 0.1 cm	< 0.1 cm
d/σ_d	< 8	< 8	< 8
I_e	$< 0.4 \times p_T$	$< 0.4 \times p_T$	$< 0.4 \times p_T$
$\sigma_{i\eta i\eta}$	–	$<$ { 0.011 / 0.030 }[1]	$<$ { 0.011 / 0.030 }[1]
H/E	–	< 0.10	< 0.10
1/E - 1/p	–	> -0.04	> -0.04
Conversion rejection	–	✓	✓
Missing hits	≤ 1	$= 0$	$= 0$
EGamma POG MVA	>WP-loose[2]	>WP-90 (>WP-loose)[2]†	>WP-loose[2]
Deep Jet of nearby jet	–	<WP-interp. (<WP-medium)[2]	<WP-medium[2]
Jet relative isolation[4]	–	< 1.0 (–) †	–
Prompt-e MVA	–	< 0.90 (> 0.90)	> 0.90

[1] Barrel / endcaps.
[2] WPs as defined by EGamma POG (see Section 5.3.1).
[3] WPs as defined by BTV POG (see Section 5.2).
[4] Defined as $1/p_T^{ratio}$-1 if the electron is matched to a jet within $\Delta R < 0.4$ or as the PF relative isolation with ΔR=0.4 otherwise.
[5] Here cone-p_T is used.
† Fails (passes) the requirement prompt-e MVA > 0.80.

TABLE 5.2

MUON OBJECT SELECTION REQUIREMENTS SUMMARY

Observable	Loose	Fakeable	Tight
p_T	> 5 GeV	> 10 GeV[4]	> 10 GeV
$\|\eta\|$	< 2.4	< 2.4	< 2.4
$\|d_{xy}\|$	< 0.05 cm	< 0.05 cm	< 0.05 cm
$\|d_z\|$	< 0.1 cm	< 0.1 cm	< 0.1 cm
d/σ_d	< 8	< 8	< 8
I_μ	$< 0.4 \times p_T$	$< 0.4 \times p_T$	$< 0.4 \times p_T$
PF muon	>WP-loose[1]	>WP-loose[1]	>WP-medium[1]
Deep Jet of nearby jet	–	<WP-interp. (<WP-medium)[2]	<WP-medium[2]
Jet relative isolation[3]	–	<0.5 (–) †	–
Prompt-μ MVA	–	< 0.85 (> 0.85)	> 0.85

[1] WPs as defined by Muon POG (see Section 5.3.1).
[2] Upper cut on the Deep Jet score defined with a linear interpolation from Deep Jet WP-medium at cone-p_T 20 GeV to Deep Jet WP-loose at cone-p_T 45 GeV, taking the Deep Jet WPs as defined by JetMET POG (see Section 5.2).
[3] Defined as 1/`jetPtRatio`-1 if the muon is matched to a jet within $\Delta R < 0.4$ or as the PF relative isolation with ΔR=0.4 otherwise.
[4] Here cone-p_T is used.
† Fails (passes) the requirement prompt-μ MVA > 0.85.

The final stage in the lepton object selection is the tight selection, which is used to identify the set of leptons that are considered to be signal leptons. Building on the fakeable and loose selections described above, the objects that pass the tight selection are a subset of the objects that pass the fakeable selection. The purpose of the tight selection is to identify leptons that arise promptly from the hard scatter, e.g. a lepton that is produced in the decay of a W boson in a $t\bar{t}W$ event. These leptons are considered to be the signal leptons for this analysis, and are referred to as "prompt" leptons. The tight selection criteria likewise aims to reject leptons that are produced in other ways (e.g. leptons arising from the decays of hadrons produced in the hadronization of a b jet), which are referred to as "nonprompt" leptons. This separation of prompt from nonprompt leptons is accomplished with a BDT that is trained by the $t\bar{t}H$ multilepton group. The BDT is described in detail in [12, 36, 37], a brief overview of which is provided below.

The training of the BDT is performed with simulated $t\bar{t}H$ and $t\bar{t}$ samples; the electrons and muons used in the training are required to pass the loose selection criteria defined in Tables 5.1 and 5.2. A separate BDT is trained for electrons and for muons. For electrons, MC samples with detector conditions corresponding to each each ultra-legacy (UL) period (UL16APV, UL16, UL17, and UL18) were used. For muons, the training from Ref. [36] (which was performed with end-of-year samples) was used, as there was no improvement in performance observed with the ultra-legacy datasets. The variables used in the training include the p_T of the lepton, the η of the lepton, the impact parameters, the isolation of the lepton with respect to other charged particles, and the output of BDT trained by the CMS EGamma POG. The BDTs for electrons and for muons are referred to as the "prompt-e MVA" and "prompt-μ MVA", respectively.

5.3.2 Definitions of variables used in lepton object selection

The loose, fakeable, and tight lepton selection is summarized in Tables 5.1 and 5.2. This section will define and describe the observables that are used in each stage of the selection.

- p_T: The transverse momentum of the lepton. For the fakeable selection, the cone-p_T is used. The cone-p_T is designed to provide a characterization of the p_T of the parton that led to the nonprompt lepton. The lepton isolation and the p_T of the nearest jet are incorporated into the cone-p_T definition, which is provided in [12]. For fakeable leptons, the cone-p_T generally exceeds the reconstructed p_T of the lepton; for leptons that pass the tight selection criteria, the p_T and cone-p_T are equal.
- η: The lepton's `eta` (where η has been defined in 4.2.1).
- $|d_{xy}|$: The lepton's `dxy` property. This corresponds to the lepton track's transverse impact parameter with respect to the PV.
- $|d_z|$: The lepton's `dx` property. This corresponds to the lepton track's longitudinal impact parameter with respect to the PV.
- d/σ_d: The lepton's `sip3d` property. This refers to the signed 3-dimensional impact parameter (with respect to the PV) divided by its uncertainty.
- I_e, I_μ: The lepton's `miniPFRelIso_all` property. This is a measure of the isolation of the lepton (corresponding to the sum of the p_T of the objects reconstructed within a cone centered on the lepton direction, where cone size is scaled inversely with the p_T, which also helps to mitigate the effects of PU).
- $\sigma_{i\eta i\eta}$: The lepton's `sieie` property. The $\sigma_{i\eta i\eta}$ of the supercluster in the ECAL, a measure of the energy distribution within the crystal cluster.
- H/E: The lepton's `hoe` property. A measure of the energy deposited in the HCAL to the energy deposited in the ECAL.
- 1/E - 1/p: The electron's `eInvMinusPInv` property. This corresponds to the difference between the reciprocal of the electron cluster energy and the reciprocal of its track momentum.
- PF muon: The muon's `looseId` property. This requires the muon to pass the loose requirements specified by the Muon POG [6].
- Conversion rejection: Requires the electron's `convVeto` property to be `True`.
- Missing hits: The electron's `lostHits` property. The number of missing hits in the tracker.

- EGamma POG MVA: The output of the BDT trained by the EGamma POG. As described in Section 5.3.1, this helps to distinguish real electrons from jets.

- DeepJet of nearby jet: The output of the DeepJet discriminant (`btagDeepFlavB`) for the nearest jet (`matched_jet`).

- Jet relative isolation: The lepton's `jetRelIso` property. In the case where there is a matched jet, this corresponds to the relative isolation, defined as the difference between the matched jet p_T and the lepton p_T, with respect to the lepton-p_T. In the case where there is not a matched jet, the `jetRelIso` is equal to `pfRelIso04_all`.

- Prompt-e MVA: The output of the prompt lepton MVAs trained by the $t\bar{t}H$ multilepton team, as discussed in Section 5.3.1.