Chapter 6

Particle Flow Reconstruction

6.1 Introduction

Analyzing the experimental data from the modern day high energy collider experiments aim to obtain a description of the collisions in terms of the original partons of the underlying physics process. This is achieved by reconstructing the basic objects like jets, electrons, muons, taus and missing transverse energy. A particular signal is identified in data using these reconstructed particles by seeking a compatibility between the relevant final state and the particle content in the observed event. Hence identification and subsequent disentanglement of signal from the background is crucially dependent on the efficiency and purity with which we identify the parton species, accuracy with which we measure their energies and direction, and the relative level of the systematic uncertainties associated with these measurements. Throughout the remaining part of the thesis we use the particle flow based reconstruction [74], [75] which tries to optimally combine information from all the CMS sub-detectors in order to reconstruct and identify all the stable particles in the event i.e electrons, muons, photon, charged hadrons and neutral hadrons. Along with the momentum, their direction and energy are also measured with high resolution. This list of the particles is then used to create the higher physics objects just like a list of stable particles obtained from Monte Carlo event generator. The technical design of the CMS detector makes it ideally suited for the particle flow reconstruction. The silicon
tracker placed in uniform axial magnetic field (3.8) T allows a very efficient reconstruction of the charge particle tracks with reasonably low fake rate down to the transverse momentum of 150 MeV/c. The hermetic and highly granular electromagnetic calorimeter surrounding the tracker allows very efficient reconstruction of the photons. The particle flow is especially benefited by the granularity of the ECAL and the large magnetic field inside the tracker since this allows the photons to be separated from the charged particles. The granularity in the HCAL is about 25 times coarser than the ECAL, which means that it won't be possible to spatially resolve the neutral and charged particles inside a jet with momentum above 100 GeV/c. However since we have resolution of about 10% in combined HCAL+ECAL system, we can start by reconstructing the charged hadrons using superior angular and energy resolution of the tracker and then identify the neutral hadrons as an energy excess on the top of energy deposited by them. Reconstruction of the electrons occur by the combination of a track and many energy deposits in the ECAL. These deposits may either be due to the electron itself or due to the Bremsstrahlung photons radiated by the electrons as they traverse through the detector material. Muons are reconstructed with high purity and efficiency by using a combination of the information from the muon system and the tracker. Finally, the presence of the neutrinos and other weakly interacting particles is inferred by calculating the missing transverse energy in the event. Current chapter serves as an important link as we shift from the study of radiation environment and various safety issues, towards the studying the potential for an early SUSY discovery at CMS. Next section gives a detailed information about the creation of basic elements of the particle flow where as the selection and identification of the particles is presented in the section 6.3. Finally in section 6.6, we conclude this very small chapter with summary and plans for the coming chapters.
6.2 Fundamental Elements of Particle Flow

To understand the process of the particle flow reconstruction we need to have an understanding of the various building blocks of the algorithm like tracks, clusters and blocks.

6.2.1 Iterative Tracking

Tracker is the cornerstone of the particle flow event reconstruction. It provides the momentum measurement of the charged hadrons (which carry about two thirds of energy of a jet on average) with a resolution much superior than that of calorimeters up to several hundreds of GeV/c. It also provides a precise measurement of the direction of the charged particle at the production vertex before it gets deflected during the journey to ECAL. In order to have efficient track reconstruction with very low fake rates, we use an iterative strategy for the tracking. We describe the iterative tracking scheme as follows:

- Begin seeding and reconstructing the tracks with very tight criteria. This causes the efficiency to remain moderate but keeps the fake rate low as well.

- Remove the hits which were unambiguously assigned to a track in previous iteration, and restart seeding the tracks but with loosened criteria. This leads to higher efficiency at the same time the fake rate stays low because of removal of the hits.

- With in first three iterations, all the tracks originating within a thin cylinder about the beam axis are recovered with an efficiency of 95% for muons and 90% for the charged hadrons in the jets.

- Fourth and fifth iterations have relaxed constraints on the origin vertex to allow for the reconstruction of secondary charged particles. They may have either come from the photon conversions and nuclear interactions in the tracker or from the decay of the long lived particles such as $K_\pi$'s.
This iterative technique allows for the reconstruction of the charged particles with as little as three hits and transverse momentum as small as 150 MeV/c and origin vertex more than 0 cm away from beam axis with a fake rate of lower than one percent.

6.2.2 Calorimeter Clustering

Clustering aims to measure the individual energy deposit from the charged and the neutral particles. Particle flow algorithm uses the high granularity of the detectors to separate the deposits which are close together as long as the particle showers don’t overlap.

- Rec-hits greater than seed threshold and representing local maxima in calorimeter are identified and marked to become the seeds. Clusters are grown around the seeds by connecting to them the remaining rec-hits with energies greater than the cell threshold. A cell may be shared by more than one cluster and although there is no restriction on size but in practice clusters don’t grow larger than 2-3 cells.

- Each cluster is assigned energy and position using an iterative procedure. Initially each cluster is assigned a position equal to that of its original seed, then each rec-hit contributes energy to each of its parent clusters with a weight,

\[ W = w_{ij} \times \exp \left( -d_{ij}^2 / R^2 \right) \]  

(6.1)

where \( d_{ij} \) is the distance between the i’th cluster and the j’th cell. R is a constant and \( w_{ij} \) is the normalization constant to prevent the double counting.

- The position is recalculated as the average of positions of constituent rec-hits weighed by a factor of \( \log(E_i / E_{cell}) \).

- Now the energy is re-evaluated with these new positions, and this iterates till the clusters position does not become stable with respect to the sub-detectors position resolution.
We must mention that in the ECAL the position of the cluster is corrected for the fact that the cells do not point to the origin of the coordinate system.

### 6.2.3 Linking Algorithm

A particle traversing through the CMS will give rise to several particle-flow elements as it traverses across the various sub-detectors. The link algorithm aims to connect various such elements to each other in order to fully reconstruct each single particle and at the same time avoiding any double counting from different detectors. We begin by forming a link between every pair of the particles in the event, whose quality is quantified by the distance between the elements. Then algorithm produces blocks of the elements that are linked directly or indirectly, and each block consists of not more than two-three elements. The blocks now act as the simple inputs for the particle reconstruction and identification algorithms. This is possible due to the granularity of the CMS detector, particularly in the tracker, preshower, ECAL and muon systems. The number of blocks in an event is proportional to the complexity of the event, and a jet from a typical QCD event will give rise to a large number of the blocks with few elements each, rather than large block with many elements.

### 6.3 Particle Reconstruction and Identification

For each block, the algorithm begins by promoting the global muons to a particle flow muon depending on the compatibility of its combined momentum with that determined from the sole tracker within three standard deviations. Electron reconstruction follows afterwards. Since the electrons tend to give shorter tracks and loose energy via Bremsstrahlung, the tracker can be used as preshower for the a pre-identification stage. Pre-identified electrons are refitted with the Gaussian Sum Filter in an attempt to follow their trajectories to ECAL. Finally a large number of the tracking and calorimeter variables are used to perform an identification, giving rise to an particle flow electron. The
corresponding track and ECAL clusters are removed from the further processing of the block. Tracks that remain in the block which have a poor quality and carry less $p_T$ than the corresponding calorimeter energy measurement, within calorimeter’s resolution, are discarded. Approximately 0.2% of the tracks in the hadron jets meet this fate, of which 90% are the fakes. Next in case of an HCAL cluster linked to a track, the calibrated HCAL cluster energy is compared to the track momentum. A charged hadron is created if they are found compatible, and the energy is determined from weighted average of the track momentum and the cluster energy. But if the difference between them is significant, a neutral hadron is created out of the excess cluster energy. In case where an ECAL cluster and HCAL cluster are linked together with a track, the calibrated combined energy of clusters is compared with the track momentum. If it is compatible, a charged hadron is created otherwise a neutral hadron is created out of the excess calorimeter energy. This decision is taken by a multivariate analysis of the track momentum, the relative energy deposits in the ECAL and HCAL, cluster-track link quality and the transverse cluster shapes. After removing these tracks and the clusters from the list of the unassociated objects, only cluster not linked to any track remain uncleared from the event. Any such ECAL clusters are assumed to be photons and any such HCAL clusters are assumed to be neutral hadrons. The complete list of the particles so formed is used to derive composite physics objects such as jets, MET and taus. Jets are obtained by clustering the reconstructed particles except isolated leptons as if they come from a monte carlo event generator.

6.4 From Particles to Jets

Phenomenologically speaking the Jets in high energy proton-proton collisions can be traced back to the initial hard scattering process of two incoming partons producing outgoing partons. The outgoing partons produce a parton shower, which is collimated along the direction of its motion. These partons hadronize into the colorless particles and
interact with the detector material as the propagate through it as conceptually sho

Figure 6.1. One possible way to retrieve this information is to combine the energy de

in calorimeters. The jets reconstructed this way are called CaloJets. This work pres
in this thesis use the jets reconstructed using the particle flow information, called p
With in the reach of ECAL and tracker, the pfJets are composed of individually iden
carged hadrons, photons, electrons, and neutral hadrons. In the forward region the
built of hadronic and electromagnetic energy deposits as only HF calorimeter exter
this region with its long and short fibres. The detailed reconstruction and commissi
are discussed in [74], and [75]. In this approach the 65% of the jet energy in tr
covered region is carried by charged hadrons, 25% by photons and about 10% by n
hadrons. Hence the influence of poor HCAL resolution is restricted only to the
contribution from neutral hadrons.

6.4.1 Jet Clustering

Regardless of type, the reconstruction proceed via so called jet-finding algorithms ·
provide a set of rules for grouping input objects into jets based on one or more paran
(eg. distance parameter). Input for caloJet creation are the calorimeter deposits, wh
for pfJets the full list of reconstructed particles is used.
Key requirements for jet-finding algorithms are that they should be insensitive to the addition of soft-radiation and collinear splitting of particles. In general there are two main types of the jet finding algorithms: Cone-type algorithms and sequential-clustering algorithms. Here is the short introduction to both these types:

**Cone-Type Algorithms** : The hadrons produced in fragmentation process following hard scattering form a collimated shower in the direction of initial partons. This suggests a geometry where a cone in direction same as that of initiating parton, can be used to contain the deposits/particles of the jet. One of the popular cone-type algorithms at CMS is the Iterative Cone (IC) algorithm. It starts with a $p_T$ ordered list of all input objects and use the highest $p_T$ object as seed. A trial cone with radius $R$ is formed and the four momenta of the all input objects within the radius are summed. Now a new trial cone is formed in direction of trial jet axis and this process is continued till the direction of the trial jets change by less than $\Delta R = 0.01$ and energy of the trial jet changes by less than $1\%$ between the iterations. IC algorithms is computationally fast and hence it is used for triggering. The flip side is that it is not infrared and collinear safe. This problem is better taken care in Seedless Infrared Safe (SIS) Cone algorithm. It does not use seeds but instead considers all possible jet cones and has a dedicated split and merge procedure for overlapping jets. However this comes for a price, that this algorithm is computationally very time consuming and hence not very popular in CMS collaboration.

**Sequential-Clustering Algorithms** : The jets formed this way do not necessarily have a regular shape as in the cone algorithms. These algorithms form a family of clustering algorithms ($k_T$, anti-$k_T$ and Cambridge/Aachen) which are based on a similarly defined distance measure between the objects that are to be clustered.

$$d_{ij} = \min(p_{T,i}^{2k}, p_{T,j}^{2k}) \frac{\Delta^2}{R^2}$$

$$d_{ij} = p_{T,i}^{2k}$$  \hspace{1cm} (6.2)
Here R is the distance parameter, $k = 1$ for the $k_T$ algorithm, $k = -1$ for anti-$k_T$ algorithm and $k = 0$ for Cambridge/Aachen-algorithm. And $\Delta_{ij}$ denotes the distance in $\eta - \phi$ space as
\[
\Delta_{ij}^2 = (\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2
\] (6.3)

All distances $d_{ij}$ and $d_{ij}$ are calculated and the minimal among all distances is determined. If minimum is $d_{ij}$, the objects $i$ and $j$ are recombined into a single new object by addition of their four momenta and all the distances are recalculated. And if minimum is $d_{ij}$, this object is regarded as jet and it’s constituents are removed from input list. This process continues until all input objects are clustered. The anti-$k_T$ has specialty that the resulting jets have a regular shape comparable to cone algorithms. It is also infrared and collinear safe and has become standard at CMS. Current work uses anti-$k_T$ jets reconstructed with a distance parameter of $r=0.5$.

6.4.2 Jet-energy Corrections

When studies are performed using the simulated MC samples, an additional type of the jet is used. The so called GenJets are clustered from the list of stable particles as defined in Monte Carlo event generators. They are a useful tool since they are independent of detector response and hence can be regarded as a reference for calibrating the energy of measured jets. We need to calibrate the jet energy scale because the measured jet energy is typically different from actual energy carried by initial parton. This inconsistency arise mainly because of non uniform and non-linear response of CMS calorimeters. In addition the electronic noise and event pile up can cause extra unwanted energy. The process of correcting the detector level jet energies to corresponding GenJet level is called 

JetEnergyCorrection (JEC). The strategy deployed within CMS for jet energy corrections is discussed in [76], where a factorized approach has been proposed. A very brief summary various correction terms is provided below:

- Offset: Required correction for pileup and electronic noise.
• Relative: Required for jet response versus rapidity relative to a control region.
• Absolute: Required for jet response versus $\rho_T$ in a control region.
• EMF: Option correction for jet response versus electromagnetic fraction.
• Flavor: Option correction to particle level.
• Underlying Event: Optional correction for UE energy.
• Parton: Optional correction for parton level.

The current status of JEC at CMS is presented in [77].

6.5 Missing Transverse Energy

An accurate determination of the missing transverse energy both for the events with missing energy (typical of standard model backgrounds) and for the events with missing energy (often characterizing new physics) is a major asset in separation of two types of events. Particle flow algorithm allows a simple deduction of the missing $E_T$ variance; one simply has to perform a transverse momentum-vector sum over all the reconstructed particles in the event and then take opposite of this azimuthal, momentum two-vector. The performance of particle flow missing $E_T$ (also referred to as pfMET) in MC have been reported in [71]. It was observed that the response for pfMET was very close to unbiased even without applying any posteriori corrections. While working with CMS data from 2010A we observed [79] that the pfMET distributions agreed well with the simulation shown in figure 6.2. It was stated in the reference [78] that early phase of data taking CMS will be marked by poor resolution and poor understanding of the inferred variance especially the missing $E_T$ at CMS. This is a natural expectation since there are several sources that may contribute to poor missing $E_T$ resolution e.g. the energy loss in cracks and un-instrumented regions of the detector, energy loss from dead cells, hot in the calorimeter, mis-identified cosmic rays etc. While working on the current th
we began by exploration of the possibility of SUSY searches in early CMS data with using the missing transverse $E_T$. But this approach was re-evaluated after the success commissioning of pfMET variable with real data. This was done to obtain maximum sensitivity for the current low integrated luminosity used for these searches. Now we have briefly summarized the basics of particle flow based particle reconstruction identification, we move to next part of the thesis. In the chapter 7 we present a simple analysis to measure the jet multiplicity ratios in $Z(\rightarrow \mu\mu)+X$ production. The aim is to gain an insight into the physics characteristics of this channel, since it is a major background for the SUSY searches presented in chapter 8.

6.6 Summary

Particle flow combines the information from the various sub detectors in order to identify and reconstruct each stable particle in the event including the ones that are within jets. Because of the use of full redundancy of the CMS subsystems accurate measurements of the energy, position and the direction are possible. The particle list yielded by particle flow is then used to reconstruct composite objects like jets, taus and MET, and all of them benefit from the improved energy and the angular resolution. In the next chapter...
present details of the $Z/\gamma^* + Jet$ analysis performed using the particle flow.