Likelihood-based Particle Flow Algorithm at CDF for Accurate Energy Measurement of Hadronically Decaying Tau Leptons

Andrey Elagin[*]
on behalf of the CDF collaboration
Texas A&M University
E-mail: elagin@tamu.edu

Alexei Safonov
Texas A&M University
E-mail: safonov@tamu.edu

We present a new technique for accurate energy measurement of hadronically decaying tau leptons. The technique was developed and tested at CDF experiment at the Tevatron. The technique employs a particle flow algorithm complemented with a likelihood-based method for separating contributions of overlapping energy depositions of spatially close particles. In addition to superior energy resolution provided by the method and improved discrimination against backgrounds, this technique provides a direct estimate of the uncertainty in the energy measurement of each individual hadronic tau jet. This new technique is now being deployed to improve sensitivity of the $H \to \tau\tau$ search at the Tevatron. With appropriate adjustments, the algorithm can be further extended to the case of generic (quark or gluon) jets and adopted at other experiments.

---

[*]Speaker.

1. Particle Flow Algorithm

Precise energy measurement of QCD jet is in demand for current and future collider experiments. Calorimeter is an essential part of the detector to measure energy of the particles which produce electromagnetic or hadron showers in the absorber material. Particle energy can be determined based on energy it deposited in a properly calibrated calorimeter. Each particle in a jet deposits energy in the calorimeter producing a broad shower. Therefore jet energy can be measured as energy of the calorimeter cluster associated with the shower produced by all the particles within jet. However the resolution of this calorimeter approach for jet energy measurement is limited by the resolution of the hadron calorimeter which is quite poor.

Hadrons carry about 70% of the total energy of the jet: 60% comes from charged and 10% from neutral hadrons. Photons account for the rest 30%. Typical energy resolution of the calorimeter is $\frac{\delta E}{E} \sim \frac{0.2}{\sqrt{E}}$ for the electromagnetic part and $\frac{\delta E}{E} \sim \frac{0.5}{\sqrt{E}}$ for the hadronic part. Following standard calorimeter approach one gets jet energy as $E_{jet} = E^{HAD} + E^{EM}$ where most of the energy depends on the resolution of hadron calorimeter.

The resolution can be improved by adding information from other sub-detectors to the calorimeter clusters. Calorimeter is usually preceded by a tracker for momentum measurement of charged particles. The resolution of the tracker is much better than the resolution of the hadron calorimeter. One can improve jet energy measurement significantly by utilizing tracker: $E_{jet} = E_{tracks} + E_n + E_\gamma$, where E_{tracks} is the energy of charged components of the jet extracted from the tracker, E_n - energy of neutral hadrons measured in the hadron calorimeter and E_γ - photon energy measured in the electromagnetic calorimeter. In this case only 10% of the jet energy relies on hadron calorimeter.

The concept of breaking jet components into different particle species and using the best suitable detector to measure the energy of each component is called Particle Flow Algorithm (PFA)

For experimental setup which includes tracker and calorimeter, jet energy resolution obtained with Particle Flow Algorithm consists of the following terms: $\sigma^2(E_{jet}) = \sigma^2(E_{tracks}) + \sigma^2(E_n) + \sigma^2(E_\gamma) + \sigma^2(conf.)$, where first three terms corresponds to the detectors resolution and the last term accounts for the uncertainties in assigning detector responses to particular particle species. This term is called "confusion term". When tracker is used for energy measurements of charged particles their energy deposition in the calorimeter should be subtracted from the calorimeter cluster before assigning energies to neutral components of the jet. In many cases this can only be done approximately which leads to additional smearing of the resolution on top of the combined resolution of the tracker and the calorimeter. In general the confusion term grows with the size of the calorimeter segmentation since it becomes difficult to separate overlapping energy depositions from particles in the jet. High particle density also deteriorates the resolution as calorimeter energy deposition cannot be assigned unambiguously to spatially close particles even in a highly segmented calorimeter.

Thus one of the challenges for the PFA is to minimize the confusion term which requires separation of charged and neutral jet components with proper energy assigned to each component.

2. Tau Leptons at CDF

At CDF experiment at the Tevatron hadronically decaying tau leptons have signature in the

detector which is similar to quark or gluon jets. Visible products of tau decay consist of charged and neutral pions collimated into a narrow cone due to boost effect. Therefore PFA can be used to measure visible energy of the tau jet. Neutral pions are detected after their decay into a pair of photons which essentially leaves no neutral hadron component in a tau jet except for very rare admixture of kaons. Hadronic taus are then, less complicated objects than usual jets and therefore they can serve as a feasibility probe for Particle Flow Algorithms.

In addition to be a step forward toward better understanding of jets, hadronic taus play an important role at CDF [1] [2]. Higgs boson decay ratio into a pair of tau leptons is only about 10% of the decay ratio to a pair of b-quarks. However modes with one tau decaying hadronically and another decaying to leptons have advantage of low Drell-Yan background. In addition backgrounds from QCD jets can be controlled by isolation requirements applied to either of the taus. These "golden modes" accounts for 46% of $H \rightarrow \tau\tau$ channel. Efficient identification and precise energy measurement of hadronically decaying tau leptons is indispensable for the search of Higgs boson in this channel.

At CDF precision of Particle Flow Algorithm is limited by large calorimeter segmentation. The size of calorimeter towers is comparable to the size of a typical tau jet. We propose to use likelihood-based approach to separate hadrons from electromagnetic objects and properly assign energy to all visible decay products [5].

2.1 CDF detector overview

For detailed description of the CDF detector we refer to [3]. The following parts of the detector are used in tau reconstruction and identification. The silicon vertex detector (SVX) is the part of CDF detector closest to the beam line. It allows precise measurement of the vertex position in the transverse plane via $r - \phi$ tracking covering pseudorapidity $|\eta| < 2.0$. Central Outer Tracker (COT) is a drift chamber covering radii from 0.4 m to 1.37 m. The COT resolution is $\delta p_T / p_T^2 \approx 0.0015 (\text{GeV}/c)^{-1}$ covering pseudorapidity region $|\eta| < 1$. If available SVX hits are added to the COT information improving the resolution. Central calorimeter, electromagnetic (EM) and hadronic (HAD), covers pseudorapidity $|\eta| < 1.1$. EM is a lead-scintillator calorimeter with resolution $\delta E_T / E_T = 0.135 / \sqrt{E_T} \oplus 0.02$. HAD is an iron-scintillator with resolution $0.5 / \sqrt{E_T} \oplus 0.03$. The calorimeter consists of towers with azimuthal segmentation $\phi = 15°$ and pseudorapidity $\eta \approx 0.1$. Shower Maximum (CES) detector is a set of strip-wire chambers located inside the EM at the expected maximum of the electromagnetic shower profile (six radiation lengths). The clusters formed by strip and wires are primarily used to measure position of electromagnetic showers. While CES clusters pulse height is rarely used for energy measurement due to poor energy resolution of 23%, one can benefit from CES information by combining it with information from other parts of the detector.

2.2 Reconstruction and identification of tau leptons

There are several triggers at CDF which are used to select events with hadronic tau candidates in the final state. For the likelihood algorithm performance tests we utilize a set of *lepton plus track* triggers [4] which require at least one electron or muon candidate in the central part of the detector and one tau candidate. The *lepton plus track* triggers are highly efficient in selection of $Z \rightarrow \tau\tau$

events. Isolated track with $p_T > 5$ GeV/c becomes a hadronic tau candidate. Track is considered isolated if there are no other tracks with $p_T > 1.5$ GeV/c in the annulus $0.17 < \sqrt{\Delta\phi^2 + \Delta\eta^2} < 0.52$ around the seed track.

In the offline reconstruction, a tau candidate is matched to a cluster of calorimeter towers. At least one tower with $E_T > 5$ GeV is required. All tracks that point to the calorimeter cluster are added to the tau candidate. The track with highest p_T becomes a seed track. CES clusters with no COT track nearby become π^0 candidates. To suppress large jet induced backgrounds, typical analyses demand tau candidates to be isolated, e.g. by requiring no tracks or π^0 candidates above a certain threshold in the isolation annulus. Separation of hadronic taus from electrons is typically achieved by applying cut on the ratio of energy deposition in HAD and EM calorimeters. Tau candidate is rejected if track associated with tau is identified as a muon by the muon system of the CDF detector. In order to further reject backgrounds the invariant mass of the tracks and π^0 associated with tau candidate is required to be consistent with the mass of tau lepton.

3. Likelihood-based method

To reconstruct tau energy we utilize likelihood function based on particle type-specific signatures in the detector sub-systems. Sub-detector responses associated with the tau candidate (tracks p_T in the COT, energy deposition in the EM and HAD towers and CES clusters energies) are used as inputs into the likelihood function. Parameters of the likelihood are represented by energies of the assumed particles inside the tau jet. For each candidate we maximize likelihood function by varying these parameters to obtain the best estimate on particle energies for given detector responses. Sum of the energies of the constituent particles gives an estimator of tau energy. Shape of the likelihood function can be used to obtain uncertainties on the energy estimator. After energy of each particle within tau jet is reconstructed p-value is used to test the assumption about particle content of tau decay products. Low p-value indicates that original assumed combination of particle is unlikely to produce observed response in the detector and we adjust the hypothesis about particle content.

Hadronic tau decay products nearly 100% of the time consist of charged pions which may be accompanied by neutral pions subsequently decaying to photons. Good calibration of the sub-detector response for pions and photons is essential. Reference [5] describes details on how calorimeter and CES response functions are obtained for charged pions and photons. Since position of the photon with respect to the calorimeter tower is known from CES, calibration of EM takes into account energy leakages from the photons which are close to the edges of the tower. Momenta of charged pions are fully determined by the high precision COT measurement and therefore can be fixed in the likelihood function. Therefore only photons energies are unknown parameters that need to be determined by likelihood maximization.

3.1 Construction of the Likelihood

For completeness we repeat here mathematical details from [5] on how likelihood function is constructed. Highest p_T track in a tau candidate is extrapolated to the CES and corresponding calorimeter tower becomes a seed tower. A grid of 3x3 towers is formed around the seed tower. Each track and CES cluster is associated to one tower on the grid. Each EM tower provides its

own measurement used in the likelihood. In the case of hadronic calorimeter, we sum energies of all nine towers into a single measurement, E^{HAD}, for the entire "super-cluster". In assumption that decay products of tau are charged tracks and photons, the likelihood function has the following form:

$$L = f^{CAL}(\mathbf{E^{EM}}, E^{HAD} | \mathbf{E}_{\pi^{\pm}}, \mathbf{E}_{\gamma}) \times \prod_{j}^{N} f^{CES}(E^{CES(j)} | E_{\gamma}^{(j)}). \tag{3.1}$$

Where $\mathbf{E^{EM}} = (E^{EM(1)}, ..., E^{EM(i)}, ..., E^{EM(9)})$ is set of energy depositions in each EM tower, $\mathbf{E}_{\gamma} = (E_{\gamma_1}, ..., E_{\gamma_j}, ..., E_{\gamma_N})$ - set of photons energies and $\mathbf{E}_{\pi^{\pm}}$ represents energies of charged pions determined by the COT measurement. $f^{CES}(E^{CES(j)} | E_{\gamma}^{(j)}) = PDF_{\gamma}^{CES}$ is the CES response function for isolated photons. $f^{CAL}(\mathbf{E^{EM}}, E^{HAD} | \mathbf{E}_{\pi^{\pm}}, \mathbf{E}_{\gamma})$ is calorimeter response function for a jet consisting of charged pions and photons with energies $\mathbf{E}_{\pi^{\pm}}$ and $\mathbf{E}_{\gamma}$. Note that f^{CAL} depends on the number of particles in the jet and on position relative to the seed tower. This function is derived from calorimeter response functions to isolated particles and generally can involve rather lengthy integral expressions, which fortunately can be relatively easily computed numerically.

As an example, we provide a calorimeter response function for a tau jet which consist of one charged pion and two merged photons from neutral pion all located in the seed calorimeter tower. Here we denote calorimeter response function to isolated particle by $PDF^{CAL}(x, y | E)$, where x and y are energy deposition in EM and HAD respectively.

$$f^{CAL}(E^{EM}, E^{HAD} | E_{\pi^{\pm}}, E_{\gamma}) = \int_{0}^{E^{EM}} dx_1 \int_{0}^{E^{EM}} dx_2 \int_{0}^{E^{HAD}} dy_1 \int_{0}^{E^{HAD}} dy_2 \times$$
$$\times PDF_{\gamma}^{CAL}(x_1, y_1 | E_{\gamma}) \times PDF_{\pi^{\pm}}^{CAL}(x_2, y_2 | E_{\pi^{\pm}}) \times$$
$$\times \delta(E^{EM} - x_1 - x_2) \times \delta(E^{HAD} - y_1 - y_2) \tag{3.2}$$

3.2 Correction for misidentified particles

Values of $\mathbf{E}_{\pi^{\pm}}$ and $\mathbf{E}_{\gamma}$ which maximize 3.1 represents the best estimate on energies of particles in tau decay. However if original hypothesis about particle content were incorrect the visible energy of the tau will be wrongly determined. The number of photon candidates in a tau jet comes from CES clustering algorithm described in [5]. If photon candidates are left unreconstructed, the energy of the tau will be underestimated since in this case the energy is defined as track energy only. We can correct for this inefficiency of CES clustering by looking at p-value for such events:

$$p = \int_{f^{CAL}(\mathbf{x}, y | \mathbf{E}_{\pi^{\pm}}) < f^{CAL}(\mathbf{E^{EM}}, E^{HAD} | \mathbf{E}_{\pi^{\pm}})} f^{CAL}(\mathbf{x}, y | \mathbf{E}_{\pi^{\pm}}) d\mathbf{x} dy, \tag{3.3}$$

where $\mathbf{E}_{\pi^{\pm}}$ - tracks energies reconstructed by COT and $\mathbf{E^{EM}}$, E^{HAD} - observed calorimeter response. For small p-value ($p < 0.01$) we insert a photon into the likelihood function and change it from $L = f^{CAL}(\mathbf{E^{EM}}, E^{HAD} | \mathbf{E}_{\pi^{\pm}})$ with tau energy determined as $E_{\tau} = \sum E_{\pi^{\pm}}$ to $L = f^{CAL}(\mathbf{E^{EM}}, E^{HAD} | \mathbf{E}_{\pi^{\pm}}, E_{\gamma})$ with $E_{\tau} = \sum E_{\pi^{\pm}} + E_{\gamma}$. Apart from CES inefficiency in photon reconstruction, neutral kaons can also lead to small p-values calculated using 3.3 because energy deposition in the hadron calorimeter will be larger than it is expected for the tracks reconstructed in the COT. Due to excess of

hadron energy p-value will not improve after a photon is added to the likelihood. Consequently if p-value remains small ($p < 0.03$) we remove the photon and add kaon instead and maximize $L = f^{CAL}(\mathbf{E^{EM}}, E^{HAD} | \mathbf{E}_{\pi^\pm}, E_{K^0})$ with tau energy determined as $E_\tau = \sum E_{\pi^\pm} + E_{K^0}$. For the purpose of tau energy reconstruction we assume that kaon response in the calorimeter can be adequately described by charged pion response.

4. Algorithm Performance

We test performance of the algorithm using $Z \to \tau\tau$ events. We select a sample, in which one tau decays hadronically ($\tau \to \tau_h \nu_\tau$) and the other decays to leptons ($\tau \to e \nu_\tau \nu_e$ or $\tau \to \mu \nu_\tau \nu_\mu$). Presence of a muon or an electron with tight isolation cuts significantly reduces backgrounds coming from QCD jets. Events with lepton and tau candidates having electric charge of the same sign allow simple estimation of backgrounds directly from data.

4.1 Reconstruction of Tau Energy in MC

Figure 1 shows difference between reconstructed tau candidate energy and true visible energy obtained at generator level. Likelihood-based technique is compared with standard CDF tau energy reconstruction, which utilizes Particle Flow Algorithm principles but suffer from insufficient separation of contribution from overlapping particles. For quantitative comparison between the two algorithms we use fraction of events which reconstructed energy falls within 10% of the true energy. Improvements in the resolution are visible the most in 1-prong tau events where the contribution from neutral pions to the total energy is larger. Most of 3-prong taus consist of only charged tracks which energies are perfectly determined by the COT and can not be improved by likelihood algorithm. If photons do accompany three charged pions their overall contribution to the total energy is smaller than in the case of 1-prongs and the resolution of the photons energy measurement will contribute less to the tau energy resolution.
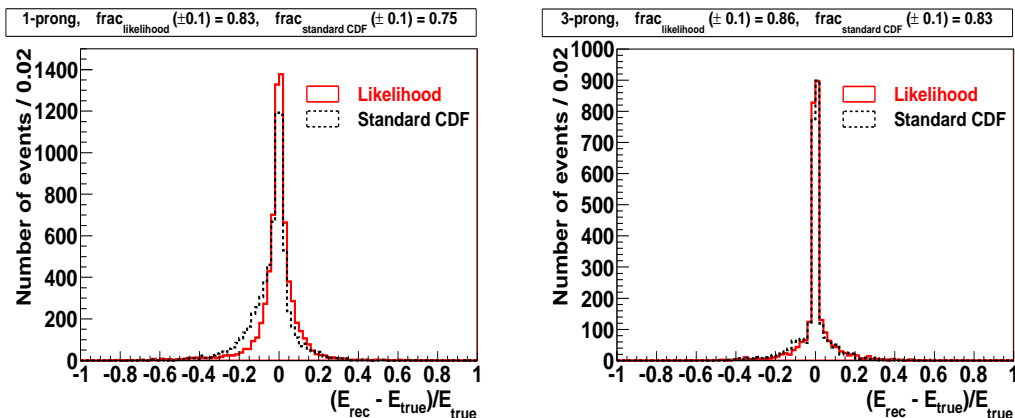

Figure 1: Comparison between reconstructed tau energy and true tau energy for $Z \to \tau\tau$ events in CDF II detector simulation. The red solid line corresponds to the likelihood method, the black dashed line corresponds to standard CDF tau reconstruction. Left: events with 1-prong hadronic taus, right: 3-prong taus.

4.2 $\rho(770)$ **Invariant Mass in MC and Data**

A large fraction ($\approx 40\%$) of all hadronic tau decays goes via $\rho(770)$ meson: $\tau_h^\pm \to \nu_\tau \rho^\pm \to \nu_\tau \pi^\pm \pi^0$. Measurement of the invariant mass of the $\pi^\pm \pi^0$ system provides an excellent test of the algorithm as it is sensitive to relatively small perturbations in 4-momentum assigned to π^0. To select a sample of tau candidates containing a single $\rho(770)$, we require exactly one COT track associated with tau candidate.
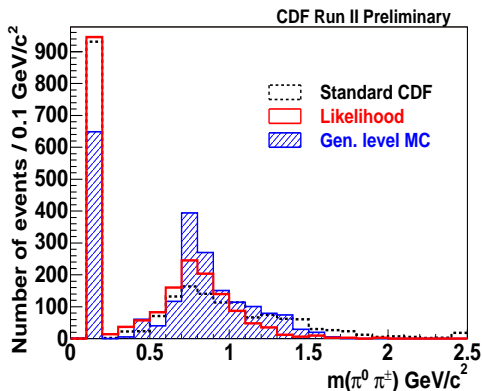

Figure 2: The invariant mass of $\pi^\pm$ and π^0 candidates. $Z \to \tau\tau$ events in CDF II detector simulation. The red solid line corresponds to likelihood method, the black dashed line corresponds to standard CDF tau reconstruction, the blue shaded histogram corresponds to the invariant mass reconstructed at generator level.

Figure 2 shows distribution of the invariant mass of 1-prong taus for $Z \to \tau_h \tau_e$ and $Z \to \tau_h \tau_\mu$ events in CDF II detector simulation. Events in the second bin correspond to 1-prong taus with no photon candidates and have invariant mass of charged pion which is 0.14 GeV/c^2. Figure 3 compares invariant mass distribution using standard CDF method and the likelihood based approach in data and in simulation. The total number of events is relatively small due to tight isolation requirements applied to leptonicaly decaying taus in $Z \to \tau_h \tau_e$ and $Z \to \tau_h \tau_\mu$ event candidates. The final tau sample shown on the Fig. 3 has QCD jet contamination less than 5%. The purity of the sample was estimated by counting the number of events in the data which pass the same selection criteria as *Ztt* events but have both tau candidates with the same electric charge. It is clear that the likelihood based technique allows a significant improvement compared to the standard CDF algorithm. Apart from serving as evidence of better energy resolution of the likelihood approach, the improved mass resolution allows more stringent jet background suppression while preserving high efficiency of tau identification.

5. Conclusions

It has been shown that the proposed likelihood based technique allows a substantial improvement in energy measurement for hadronically decaying tau leptons. The method provides better energy scale and resolution. The main improvement comes from the accurate treatment of cases with overlapping particle depositions that are difficult for conventional methods. The power of the likelihood based method is demonstrated in data using tau jet invariant mass measurement that
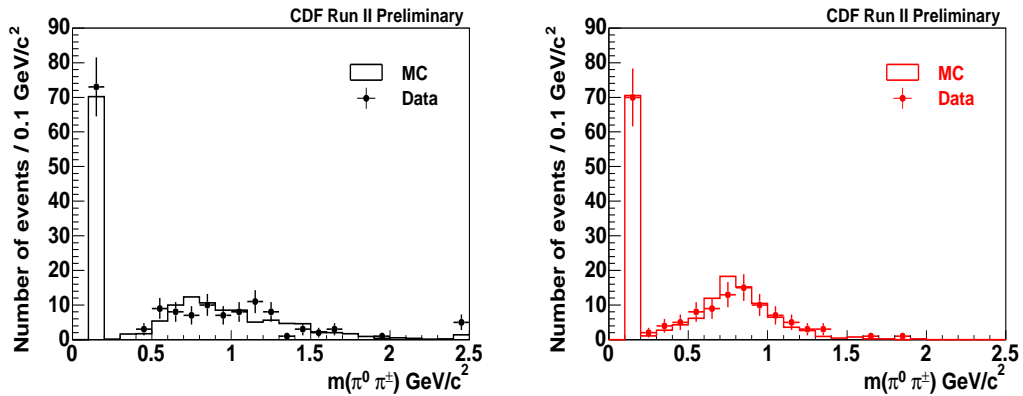
Figure 3: The invariant mass of $\pi^{\pm}$ and π^0 candidates for $Z \to \tau_h \tau_e$ and $Z \to \tau_h \tau_\mu$ events reconstructed in data. Left : standard CDF tau reconstruction algorithm. Right: likelihood technique. Solid line represents events in CDF II detector simulation normalized to the number of events in data.

indicates a substantial improvement in energy measurement resolution compared to existing methods. Likelihood approach allows to cross check the hypothesis about particle content within tau jet which further improves energy resolution.

References

[1] A. Abulencia et al. (CDF Collaboration), PRD 75, 092004 (2007).

[2] A. Aaltonen et al., (CDF Collaboration), Phys. Rev. Lett. 103, 201801 (2009).

[3] A. Abulencia et al. (CDF Collaboration), J. Phys. G 34, 2457 (2007).

[4] S.Baroiant et al., Nucl. Instrum. Methods A 518, (2004)

[5] A. Elagin, A. Safonov Nucl. Phys. B (Proc. Suppl.) 189 (2009) 332-337