Jet Algorithms and Matrix Element Analysis

DOE Review of TTU HEP Program
Dec 9 2010

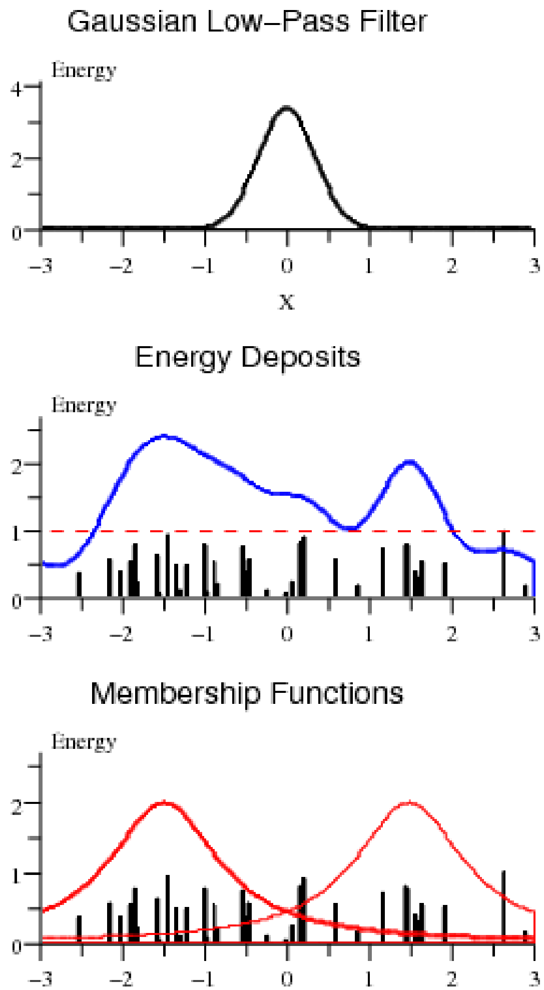
Igor Volobouev

Texas Tech University

FFTJet Overview

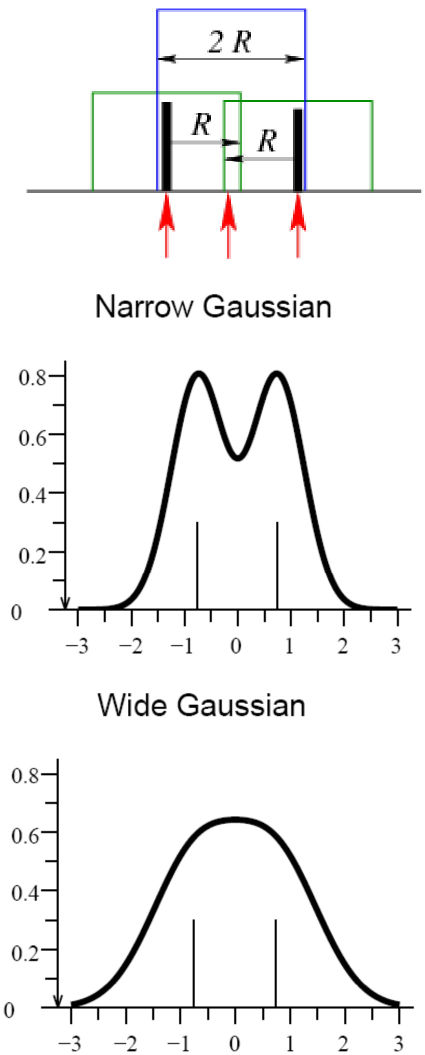
- FFTJet is a TTU-developed open source framework for reconstructing jets using multiresolution spatial filtering approach
  - Global algorithm, does not suffer from problems inherent in cone-based and sequential recombination techniques. Collinear and infrared safe.
  - Pattern recognition is performed first, utilizing Fourier-domain filtering. The code is fast and its computational complexity is independent from detector occupancy.
  - Potential jet locations are identified taking into account user-specified process assumptions
  - Jet energies are determined with probabilistic jet shape models. These models can take into account calorimeter response, magnetic field, detector noise, pile-up, underlying event, *etc.*
- Method ideas developed since ~2006. First public release of code and documentation in Summer 2009. The project is maintained at http://projects.hepforge.org/fftjet/
- Now ported to CMSSW, CMS-specific tuning is under way

Jet Reconstruction as Spatial Filtering



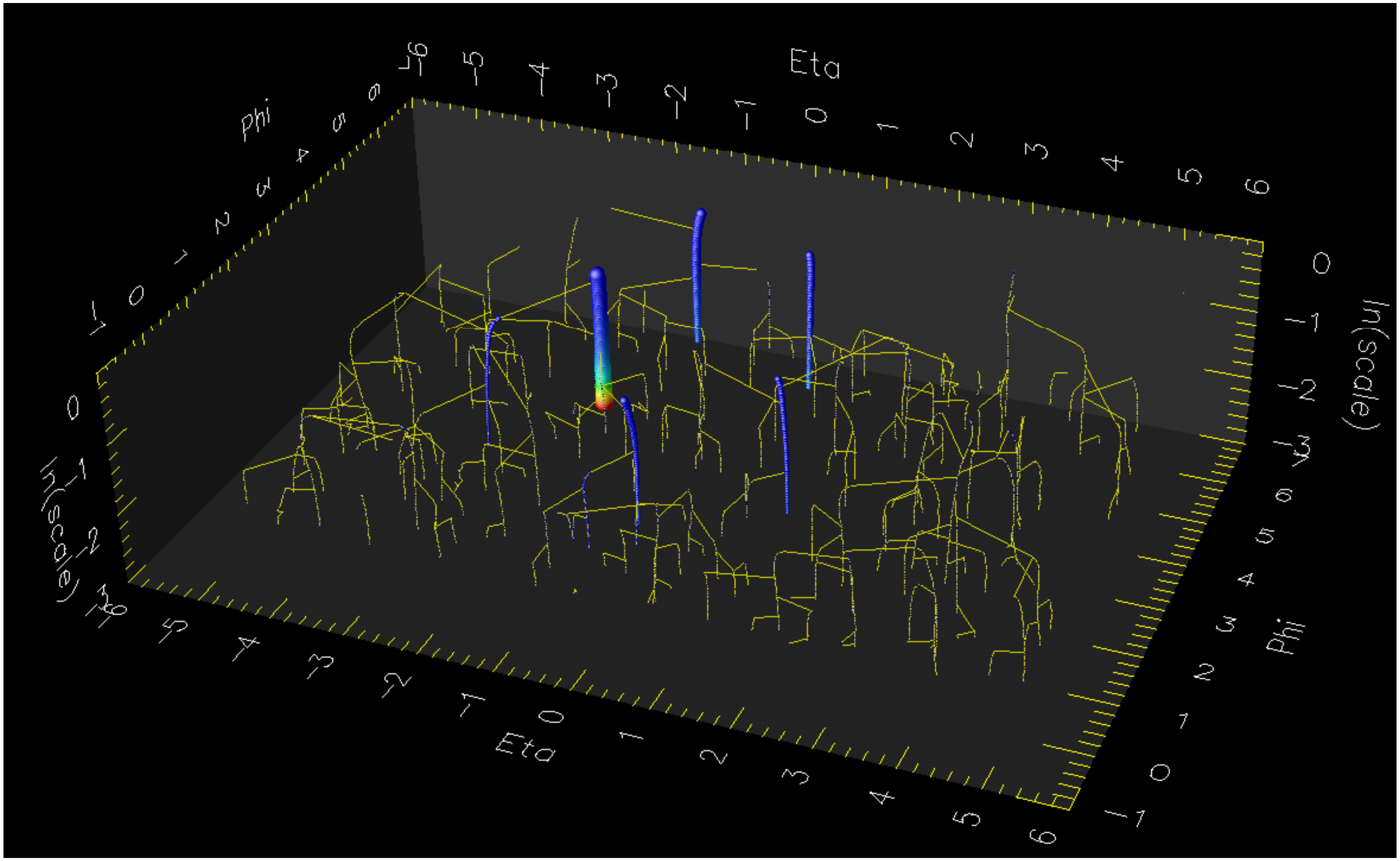
- Convolute the energy deposition picture (as a function of η and φ) with a low-pass filter. This can be done quickly by DFFT.
- Find the intensity peaks. These are the “precluster” locations.
- Apply a threshold to the peak magnitudes
- Assign a cluster membership function to each surviving precluster and to the pileup/noise
- Distribute calorimeter towers among jets and pileup/noise with weights generated by the membership functions (fuzzy clustering) or assign each tower to the jet with the highest function value (crisp clustering)

Multiresolution Jet Reconstruction



- Jets can be reconstructed using many angular resolution scales. Cone algorithm with many values of *R* can't be used for this purpose (inherent pattern recognition defect is a killer). Gaussian filtering works much better.
- FFTJet builds a "clustering tree": preclusters formed at different resolution scales are related to each other by parent-daughter relationships. Using the parent and the "closest daughter", precluster characteristics can be determined as a function of the resolution scale.
- Studies of precluster behavior in the scale space result in
  - Advanced pattern recognition capabilities
  - Optimal determination of jet properties
  - Control over bifurcation points

Example Clustering Tree



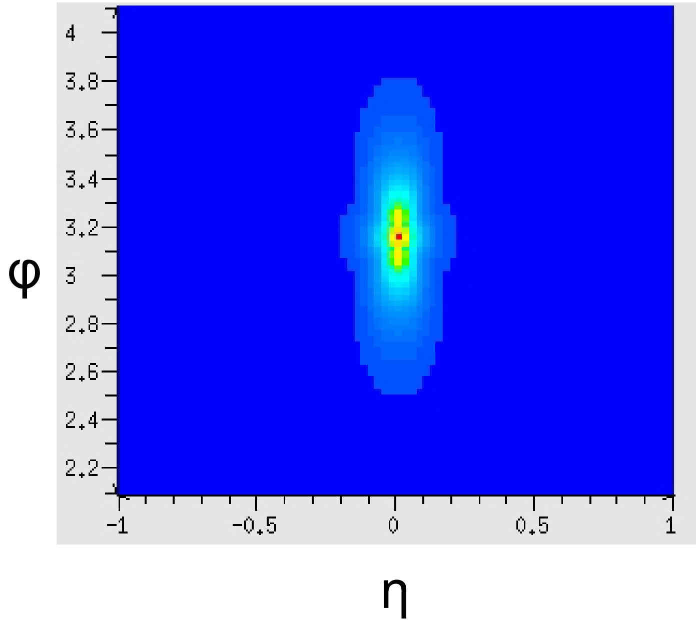
Clustering tree for a CMS all hadronic $t\bar{t}$ event (MC)

Detector and Process-Specific Tuning

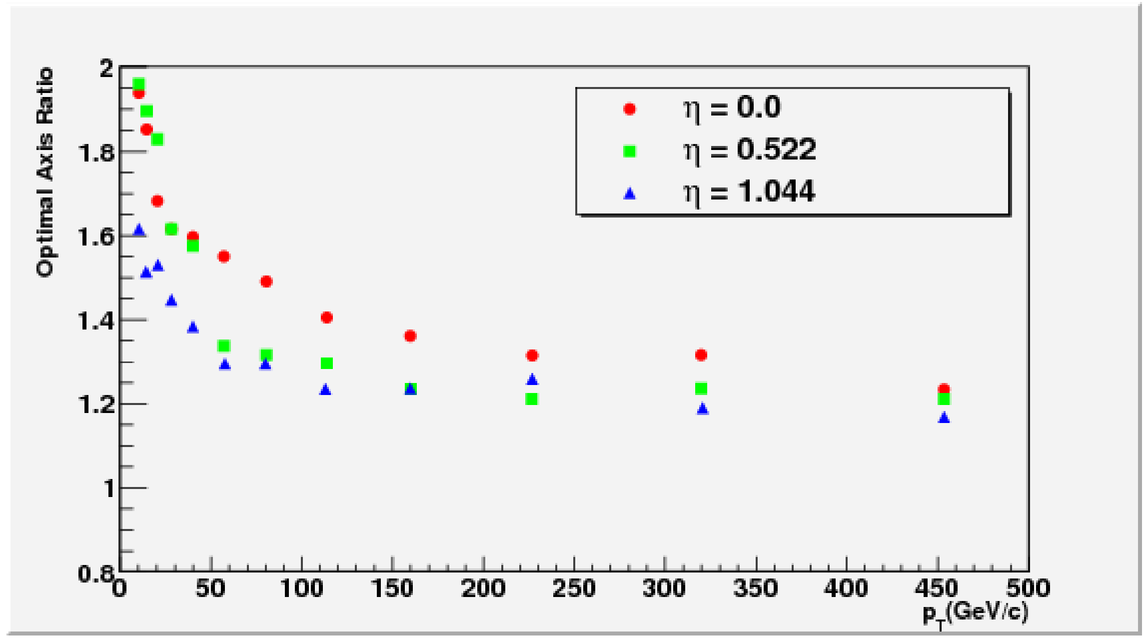
- FFTJet has quite a few tunable parameters because it recognizes more facets of jet reconstruction process than any other algorithm. Don't want to make arbitrary choices.

Tuning Goals	Handles
Suppression of showering fluctuations and detector noise	Spatial filter shape / spectrum. Peak thresholds.
Optimal process-specific clustering	Multiscale analysis of energy flow structure and clustering stability
Optimal energy resolution	Jet size and shape / background model

Jet Shapes in the CMS Calorimeter



Effect of 3.8 T magnetic field on p_T = 20 GeV/c jets



Shape of optimal elliptical jets with full detector simulation

- Optimal elliptical jets result in ~15% better jet energy resolution for the same jet area compared to calorimeter-based jet reconstruction method currently employed by CMS (anti-k_T)

Jet Reconstruction Goals

- Near term:
  - Jet energy scale determination for FFTJet elliptical jets
  - Jet cross section measurement
  - Jet reconstruction optimization for $t\bar{t}$ events (great calibration sample, interesting physics)
- Long term:
  - Detailed understanding of jet shapes at CMS. Construction of detector-level jet fragmentation functions:

$$M_i(\eta, \varphi, \varepsilon, s) = \left\langle \frac{\partial^3 N(p_T)}{\partial\eta\, \partial\varphi\, \partial\varepsilon} \right\rangle$$

  - Development of multiresolution pattern recognition strategies (analysis of jet substructure, *etc*)

Matrix Element Analysis

- The most powerful way to analyze HEP data
- Probability to see an event in the detector:

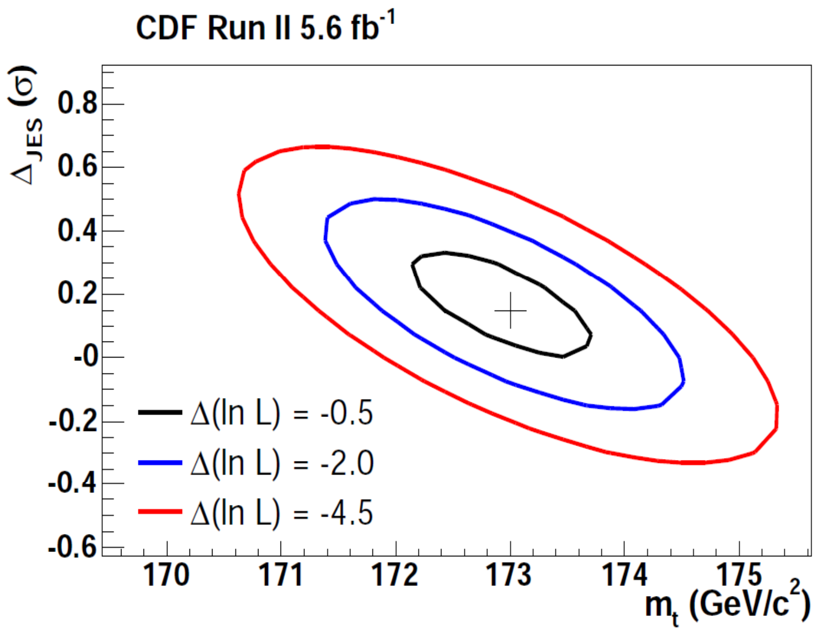
$$P_{ev}(\mathbf{y}|\mathbf{a}) = \sum_i f_i P_i(\mathbf{y}|\mathbf{a})$$

- Probability for each channel:

$$P_i(\mathbf{y}|\mathbf{a}) = \frac{1}{\sigma_i(\mathbf{a}) A_i(\mathbf{a})} \sum_k \int_X W_k(\mathbf{y}|\mathbf{x},\mathbf{a})\, \epsilon(\mathbf{x},\mathbf{a})\, |M_i(\mathbf{x},\mathbf{a})|^2\, T_i(\mathbf{x},\mathbf{a})\, d\mathbf{x}$$

- $\sigma_i(\mathbf{a})$: cross section
- $A_i(\mathbf{a})$: acceptance
- $\sum_k$: combinatorics
- $\int_X$: phase space
- $W_k(\mathbf{y}|\mathbf{x},\mathbf{a})$: transfer function
- $\epsilon(\mathbf{x},\mathbf{a})$: efficiency
- $|M_i(\mathbf{x},\mathbf{a})|^2$: matrix element
- $T_i(\mathbf{x},\mathbf{a})$: PDFs, flux

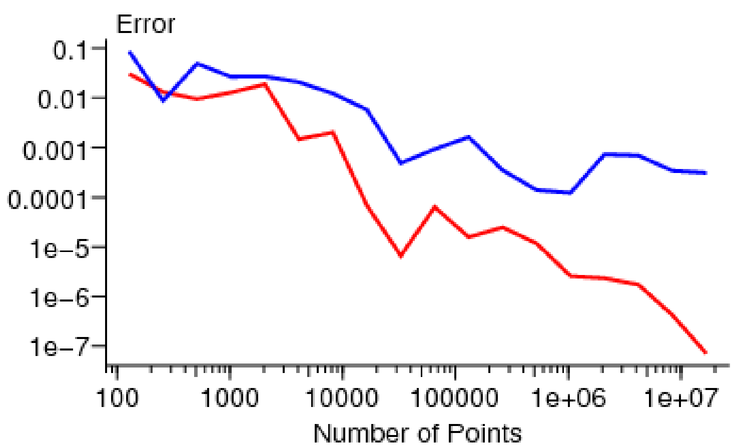
Recent Results



$m_t = 173.0 \pm 1.2$ GeV/c^2

- The flagship application of the method – the top quark mass measurement at CDF – has reached 0.7% precision (performed in collaboration with the LBNL CDF group). Fermilab Result of the Week 11/18/2010.
- World's most precise single measurement of the top quark mass
- Limited by systematics – will be difficult to beat at LHC

Generic ME Analysis Framework



Convergence of MC (blue) and QMC (red) integration methods for a 3-body phase space integral

- Features added since last year:
  - Advanced statistical techniques for transfer function modeling (decomposition of multivariate TFs into nonparametric copula and marginals). Automatic bandwidth selection methods.
  - Special interpolation schemes for TFs
  - New multidimensional integration method (higher order randomized Quasi-Monte Carlo)
  - Empirical integration convergence checker
  - Object serialization subsystem

Other ME-Related Work and Plans

- The "RK" relativistic kinematics software package has been released at http://projects.hepforge.org/rk/ (almost as a byproduct of FFTJet / ME software testing)
- Near future: construction of jet transfer functions for CMS
- Next step: a simple CMS data analysis (dijet resonances, top quark processes, Higgs search). Need to construct an interface to a grid-enabled workload distribution tool (such as Pegasus/Condor, BOINC, or PROOF) in order to give the system a serious test drive.
- The final frontier: automatic generation of efficient phase space sampling schemes. An important step in this direction was taken by developers of MadWeight software (generalization of phase space splitting relations).