Hard Scattering in Hadron-Hadron Collisions: Physics and Anatomy

Section 7: Acceptance & Efficiency

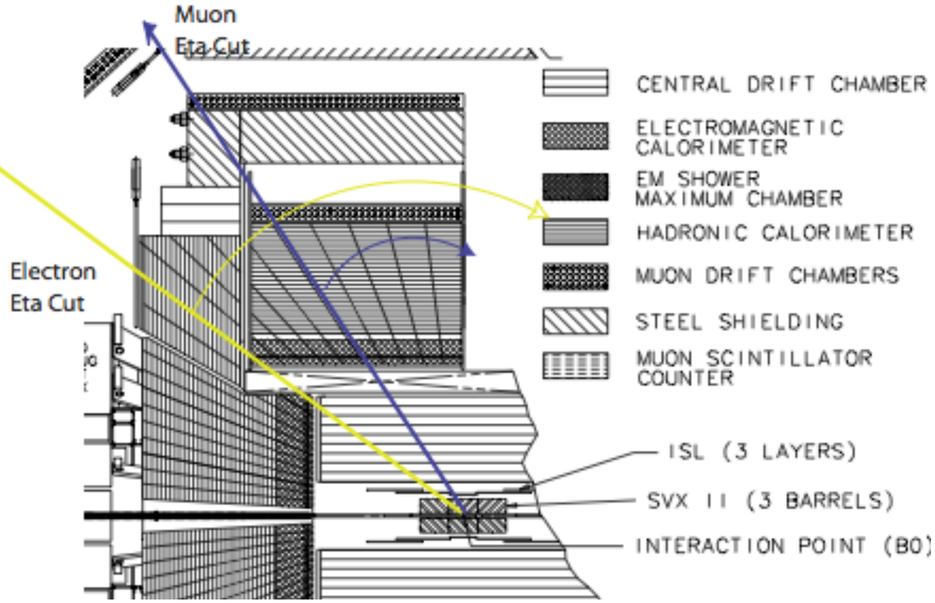
1. Strategy for Collider Experiments
2. Acceptance Calculations
3. Efficiency Measurements
4. Validating Calculations
5. Example: Optimizing Jet Energy Resolutions
6. Example: Loose vs Tight Leptons

Acceptance & Efficiency

- **Definitional issues**
  - **First, nomenclature not consistent!**
  - **Why worry?**
    - Focus on the issues you can control
    - Allow ready comparisons
- **Acceptance:**
  - **Geometry of detector**
  - **Major fiducial volumes**
    - Calorimeter
    - Tracking
  - **Identify those issues where "hard edges" can be defined and understood readily**
  - **Usually done with MC and detector simulation**
    - Uncertainties tend to arise primarily from kinematics of process
      - **Details of ME, PDFs, fragmentation and/or hadronization**
- **Efficiency:**
  - **Probability that identification/ reconstruction is successfully**
  - **Often requires clear definition of "fiducial" volume, e.g.**
    - Examples include
      - **"taggable jet" (for b-tagging)**
      - **Electron ID (avoiding cracks in calorimeter)**
      - **Muon ID (avoiding areas not fully instrumented)**
    - Driven often by defining regions of detector that are well-understood
  - **Measure efficiency using data or data-driven techniques**

Acceptance

- **Usually defined by**
  - **Set of geometrical/kinematic requirements, e.g.**
    - Charged lepton with
      - **$P_T > P_T^{lmin}$ & $|\eta| < \eta^{jmin}$**
    - N jets with
      - **$P_T > P_T^{jmin}$ and $|\eta| < \eta^{lmin}$**
  - **Usually reflects the maximum phase space in which object reconstruction is possible**
  - **Should include all processes that contribute to final state**
    - For example, in ttbar final states with e/μ, need to take into account $\tau \to e\nu\nu$/$\mu\nu\nu$
  - **Have to be careful about making cuts at "truth" level**
    - Often done to reduce # events that have to go through detector simulation & reconstruction
    - But don't throw away events that could ultimately make it into sample

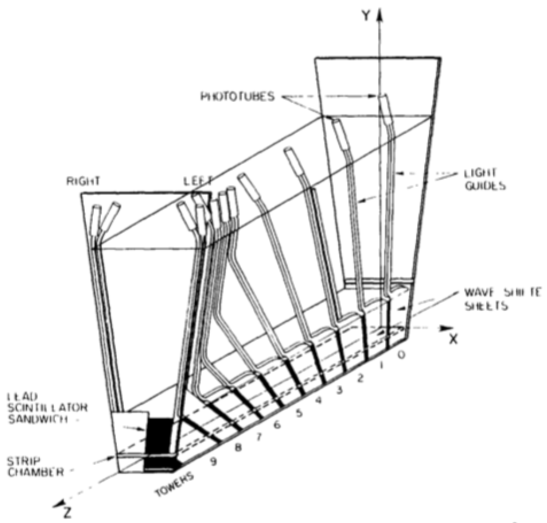


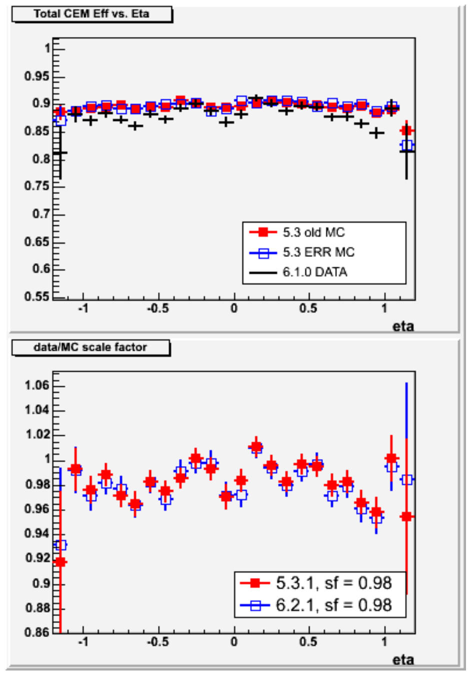
ATLAS ttbar dilepton analysis

2 Electrons	Total	2 W	1W 1b	1W 1c	1W 1Tau	1W 1Other
# Events	1,494	1,246	38	1	176	7
rate	100.0	83.4	2.5	0.1	11.8	0.5
2 Muons	**Total**	**2 W**	**1W 1b**	**1W 1c**	**1W 1Tau**	**1W 1Other**
# Events	2,831	2,203	313	6	258	3
rate	100.0	77.8	11.1	0.2	9.1	0.1
1 E 1Mu	**Total**	**2 W**	**1W 1b**	**1W 1c**	**1W 1Tau**	**1W 1Other**
# Events	4,167	3,293	320	5	453	18
rate	100.0	79.0	7.7	0.1	10.9	0.4

Efficiency Calculations

- **Efficiency measurements perhaps most challenging**
  - **Require excellent knowledge of detector response**
    - Usually define a "fiducial" region in which detector response is well understood
    - Measure efficiency in fiducial
  - **Usually can find physics process that allows measurement**
  - **Electron efficiency**
    - $Z \to e^+e^-$ where one triggers and selects first electron
    - Then look for second electron leg
  - **Need to worry that you have included the correct correlations in data**

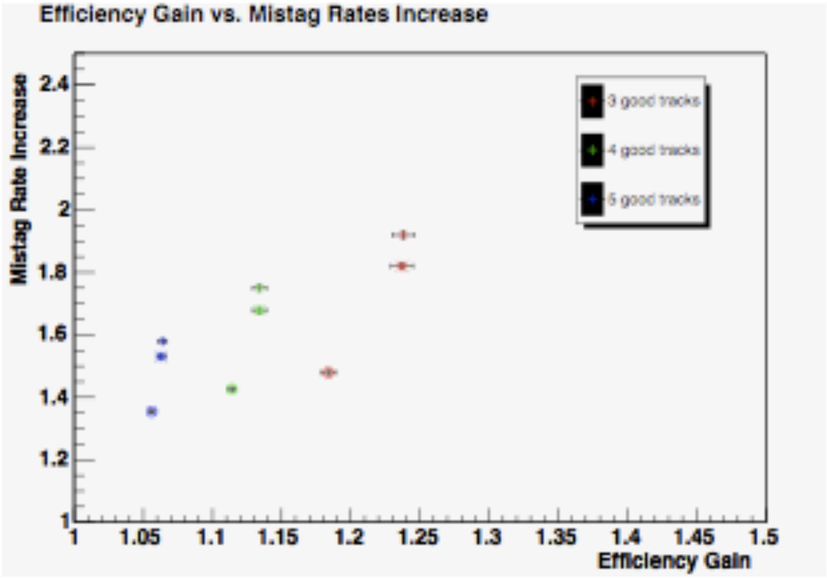




- **Key fiducial cut is to restrict to region away from edge of shower-max detector**
  - **-10% reduction in acceptance!**
    - Use +-22.5 cm out of 24.5 cm
  - **Use the data to adjust MC efficiency**

Optimization

- **One of the steps of any analysis is an optimization**
  - **Increase acceptance and/or efficiency**
  - **Usually some trade-offs**
    - Increased uncertainty in efficiency
    - Perhaps poorer S/N -> more background
  - **A very useful check:**
    - Once you have acceptable S/N
      - **Relax a cut individually and see what happens**
      - **In some cases, find out that correlations between criteria make some redundant**
  - **Example of b-tagging at CDF**
    - Combined two algorithms
      - **SECVTX and JETPRB**
    - Realized some improvement in efficiencies -- ~+15%
      - **But increased backgrounds**
    - Helpful to formulate "figure of merit"
      - **S/sqrt(S+B) often used**



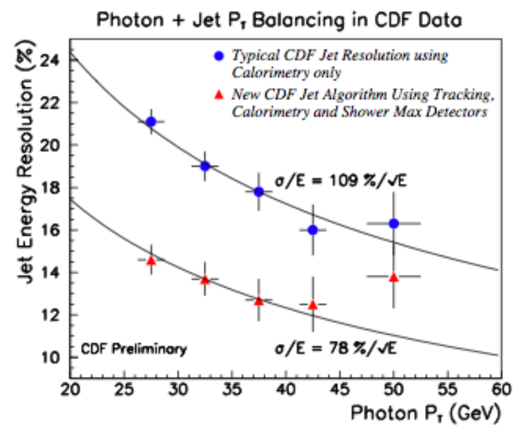
- **Combined algorithms AND improved track selection criteria**
  - **Measured efficiency gains in ttbar MC**
  - **Measured mistag rates using multijet data**
  - **Could afford some increase in background, especially because we were requiring >=3 b-tagged jets**

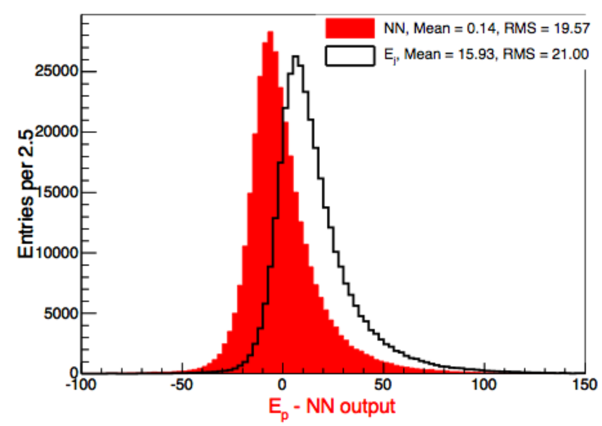
Example of “Wall”: DiJet Mass

- **Much work has gone into improving dijet invariant mass**
  - **Key to H-> bb**
    - Improve S/N given one has large background
  - **Precision M_{top} measurement**
- **Program set up from Run II to do this**
  - **Use tracking**
  - **Use shower structure**
    - Shower max detectors
    - Preshower detectors
  - **New jet algorithms**

Report of the Tevatron Higgs Working Group, Carena et al., arXiv::hep-ph/0010338, Dec 2000

- **Initial results in 1999/2000 were encouraging**
  - **Follow-on studies in Run II have not been as optimistic**
- **“Best” result has been using NN**
  - **Combine**
    - Jet P_T ΔR=.04
    - Track P_T in cone
    - Raw jet E_T and E_T in ΔR=0.7
    - EM fraction
  - **Train on MC**
    - See ~10% improvement





CDF Public Note 9463, Jan 2009

Higgs -> WW

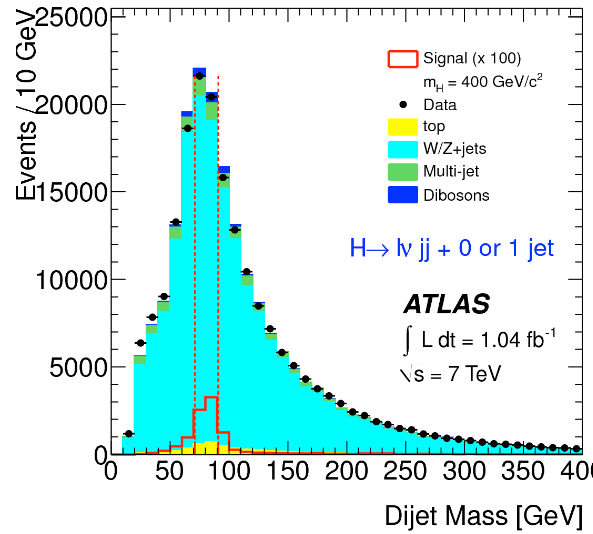
ATLAS Collaboration, 1109.3617

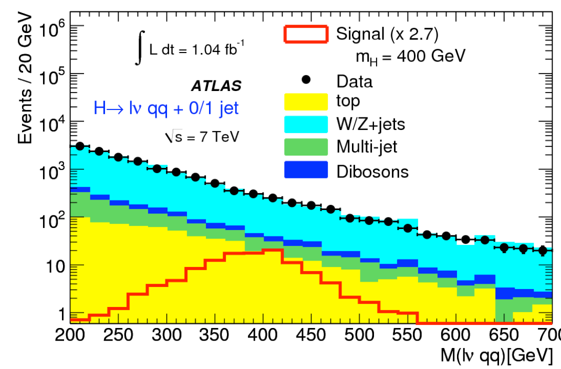
- **Another example comes from ATLAS Heavy Higgs search**

$$pp \rightarrow H + X \rightarrow WW + X \rightarrow l\nu_l jj + X$$

- **Require**
  - **Missing $E_T > 30$ GeV**
  - **Charged e, μ $p_T > 30$ GeV/c**
  - **Two or three reconstructed jets constructed with antikT R=0.4**
    - $p_T > 25$ GeV/c
    - $|\eta| < 2.5$
  - **Reject event with a b-tagged jet (why?)**
  - **Select jet pair closest to M_W with**
    - $71 < M_{jj}$ 91 GeV/c^2
  - **Analyze 1.04 fb-1**
- **Look for signal in M(*lνjj*) distribution**

- **In this case, see that the background also "peaks" in signal region**
  - **Mass resolution not great**
- **Signal (M_H=400 GeV/c^2)**
  - **σ = 2.16 pb**
  - **BR = 0.17**
  - **N_{exp} = 58 events**
  - **Acc*Eff?**
    - **15 %!**
  - Where are the primary losses?





	H($e\nu jj$) + 0j	H($\mu\nu jj$) + 0j	H($e\nu jj$) + 1j	H($\mu\nu jj$) + 1j	H + 0j or 1j
W/Z+jets	10780 ± 290	13380 ± 870	6510 ± 250	7410 ± 670	38080 ± 1160
Multi-jet	890 ± 24	256 ± 17	669 ± 25	212 ± 19	2027 ± 43
Top	170 ± 34	164 ± 33	489 ± 98	500 ± 100	1320 ± 270
Dibosons	397 ± 79	414 ± 83	161 ± 32	204 ± 41	1180 ± 240
Expected Background	12240 ± 300	14210 ± 870	7830 ± 270	8330 ± 680	42600 ± 1200
Data	11988	13906	7543	8250	41687
Expected Signal (m_H = 400 GeV)	14 ± 3.6	12 ± 3.1	18 ± 4.7	14 ± 3.6	58 ± 15

Example: Isolated track in Top Dileptons

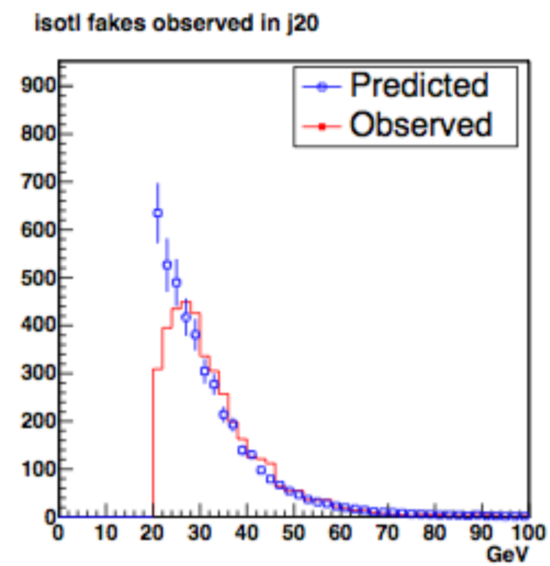
- **Initial Run II top quark studies focused on increasing signal acceptance & efficiency**
  - **Run I studies assumed that one required two well-identified leptons**
    - Effort was put into seeing how one could increase overall rate
  - **Strategy taken:**
    - Look for one well-identified lepton candidate
    - Ask for second lepton, where only "hint" of lepton was required
      - **Include leptons at higher-eta**
        - **PHX candidates**
      - **Became known as "isolated track" analysis**
  - **Resulted in the first top quark publication from Run II data**

CDF Collaboration, PRL 93, 142001(2004).

- **Isolated track lepton (tl) requirements:**
  - **Well-reconstructed charged track with P_T>20 GeV/c, |η|<1.1**
  - **Isolation requirement**
    - E_T(Cone R=0.2)/P_T < 0.1
    - If minimum-ionizing particle in calorimeter, identify as μ
      - **MET correction trickier**
  - **Also required higher quality tracking information**
    - χ^2 cut on track fit to coordinates
    - Illuminated an issue with the size of the decay-in-flight background
  - **Compare with selection where required two well-identified leptons**
    - Look at effect of requiring additional jets
      - **Use 0,1 jet samples as controls**
    - Looked at different kinematic regions

Cost for Increased Acceptance?

- **Had to pay closer attention to backgrounds**
  - **In practice, "fake" background will be larger**
  - **Also, background from DY with MET and jets needed to be evaluated carefully**
    - Some concerns that this was not well understood
    - Reduced jet and lepton P_T cuts to check behaviour
- **Fake background to "tl" estimated using dijet samples**
  - **Calculated "fake" probability per jet**
  - **Applied it to W+jets sample**



	$n_j = 0$	$n_j = 1$	$n_j \geq 2$
Top-dilepton	0.19 ± 0.02	4.22 ± 0.09	23.57 ± 0.21
WW	33.29 ± 0.58	7.32 ± 0.27	2.53 ± 0.16
WZ	3.09 ± 0.10	2.02 ± 0.08	0.72 ± 0.05
ZZ	0.85 ± 0.03	0.27 ± 0.02	0.18 ± 0.02
DYee	44.48 ± 10.08	27.25 ± 5.50	10.16 ± 2.38
DYmm	8.47 ± 2.79	6.20 ± 2.00	3.12 ± 1.15
DYtautau	8.21 ± 0.28	12.21 ± 0.34	4.80 ± 0.21
Fakes	72.13 ± 2.86	25.40 ± 1.06	10.06 ± 0.46
Total bkgd	170.71 ± 10.86	80.68 ± 5.97	31.57 ± 2.70
Total Predicted	170.71 ± 10.86	84.90 ± 5.97	55.14 ± 2.71
Observed	173	111	61

Table 21: Details of the contributions in the lepton+track analysis using kinematic thresholds on both the track lepton and jet of 15 GeV. Errors are statistical only, for systematic contributions, see Section 10. The opposite charge requirement is applied.

Comparison of Old and New

- **The direct comparison of the efficiency and acceptance informative**
  - **Real improvement came from**
    - avoiding electron/muon ID requirements
    - Adding very high-angle leptons (PHX)
  - **Overall increase of almost 100%**
- **Challenge was to understand background sources**
  - **Had to develop new strategy to calculate fake lepton background**
    - Used low-energy jet samples (Jet 20 and Jet 50)
  - **Forced us to confront the uncertainties from Drell-Yan background**
  - **Overall learned a great deal about backgrounds in dilepton final states**

channel	acceptance	data candidates	cross-section (pb^{-1})
CEM-tl	.120	10 123	260 ± 3
CMUP-tl	.125	7997	237 ± 4
CMX-tl	.064	4457	240 ± 4
PHX-tl	.101	620	235 ± 12
CEM-CEM	.068	6453	258 ± 4
CEM-PHX	.093	7250	247 ± 8
CMUP-CMUP	.022	1460	261 ± 8
CMUP-CMX	.020	1477	250 ± 8
CMX-CMX	.005	440	270 ± 14

Table 7: Measured Z cross-sections along with selected inputs (acceptance is raw acceptance from the Monte Carlo). Uncertainties are drawn from the number of candidates in the data, the number of events used to calculate the acceptance in the Monte Carlo, and the uncertainties on the lepton identification scale factors.

Event count per jet bin