Trigger & Analysis

Avi Yagil
UCSD

2nd Hadron Collider Physics Summer School
June 2007

Last time, we discussed and solved all the trigger issues at LHC.

In the next hour, we'll deal with the remaining issues related to:

Data Formats, Access, Computing, Analysis.

Fasten your seat belts…

Reminder - Trigger Concepts Discussed

- Signals may be as rare as $1:10^{10}$ --> Rejection
  - 10^4 at level1
  - Subsequent 10^3, prior to data logging
- You better have it in…
  - Does it do what you think it is doing - monitor, validate

- trigger table
- trigger path
- pre-requisites
- Volunteers
- efficiency, dead-time
- backup triggers
- trigger x-section
- pre-scales

Computing - What's the problem??

- Huge data volume
- Individual sample sizes are very large

==> Ability to:

- Locate & Access data (small part of relevant sample)
- Phrase and refine a question
- Get an answer (from full sample)

becomes a highly non-trivial task!

- Solution: increase sample granularity

==> The (only) relevant parameter: Trigger Path

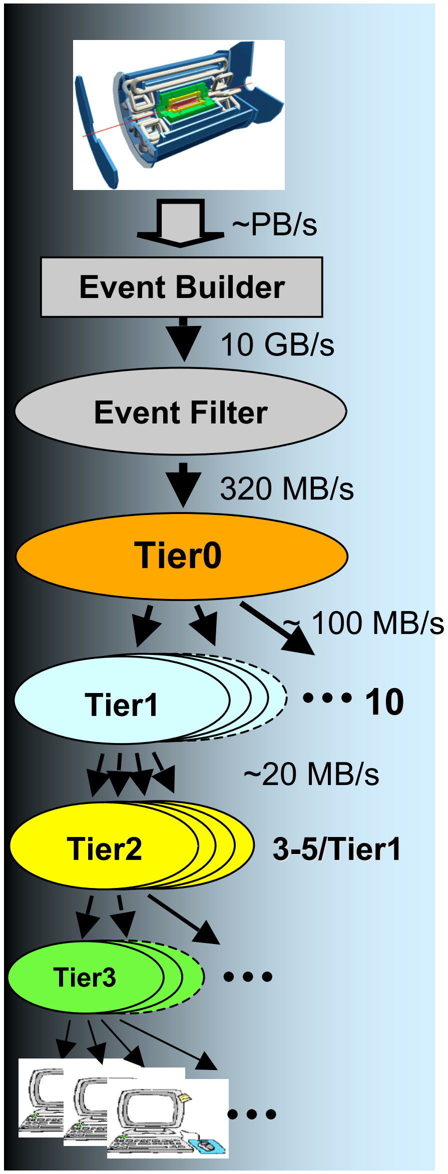
From Dario Barberis

Example: The ATLAS Event Data Model

- RAW:
  - "ByteStream" format, ~1.6 MB/event
- ESD (Event Summary Data):
  - Full output of reconstruction in object (POOL/ROOT) format:
    - Tracks (and their hits), Calo Clusters, Calo Cells, combined reconstruction objects etc.
  - Nominal size 1 MB/event initially, to decrease as the understanding of the detector improves
    - Compromise between "being able to do everything on the ESD" and "not enough disk space to store too large events"
- AOD (Analysis Object Data):
  - Summary of event reconstruction with "physics" objects:
    - electrons, muons, jets, etc.
  - Nominal size 100 kB/event
- TAG:
  - Database used to quickly select events in AOD and/or ESD files

From Dario Barberis

Example: ATLAS Data flow and distribution



In order to provide a reasonable level of data access for analysis, it is necessary to replicate the ESD, AOD and TAGs to Tier-1s and Tier-2s.

RAW:

- Original data at Tier-0
- Complete replica distributed among all Tier-1
  - Randomized dataset to make reprocessing more efficient

ESD:

- ESDs produced by primary reconstruction reside at Tier-0 and are exported to 2 Tier-1s
- Subsequent versions of ESDs, produced at Tier-1s (each one processing its own RAW), are stored locally and replicated to another Tier-1, to have globally 2 copies on disk

AOD:

- Completely replicated at each Tier-1
- Partially replicated to Tier-2s (~1/3 – 1/4 in each Tier-2) so as to have at least a complete set in the Tier-2s associated to each Tier-1
  - Every Tier-2 specifies which datasets are most interesting for their reference community; the rest are distributed according to capacity

TAG:

- TAG databases are replicated to all Tier-1s (Oracle)
- Partial replicas of the TAG will be distributed to Tier-2 as Root files
  - Each Tier-2 will have at least all Root files of the TAGs that correspond to the AODs stored there

Samples of events of all types can be stored anywhere, compatibly with available disk capacity, for particular analysis studies or for software (algorithm) development.

Table of Contents

"Analysis Model" Basics
or
Size matters!
or
How long does it take?

The problem: Size and Speed

- Turn-around time is key to develop, debug, tune and complete an analysis
  - To go fast, must have:
    - Compact event size or read-on-demand
    - Relevant info stored
    - Ability to quickly re-compute when needed (w/o going to a lower data tier)

  Obvious tension!

- Some reconstruction algorithms become un-affordable:
  - Tracking
  - DB/Geometry access

  Tension between storing info (size) and computing it (time)

- Pre-selection and reduction are crucial

  If done correctly - strong function of time!

  ==> validation, understanding of detector, SW…
  - Skimming, Thinning, Slimming…
  - Adding results of expensive algorithms

  This is where races can be won or lost!

Size & Speed - Example

- Data-set sizes (e.g. inclusive lepton) will get very large, very fast.
- Even clean signatures such as leptonic decays of W will grow very fast.
  - Consider a relatively low rate events like W-->$e\nu$, W-->$\mu\nu$.
    - At LHC luminosity of 10^{33} cm^{-2} s^{-1}, these will occur at ~1Hz
    - Over a "standard LHC year" of 10^7 sec period of data taking

    ==> Huge samples. Bigger than LEP & Tevatron!

  - How long does it take to produce a W transverse mass plot?
    - Assuming one has only to read the data, and not reconstruct anything.
    - Data read limited to ~2-3MB/sec.

      Too slow!!
    - Need to do I/O at rate of ~1kHz, using on-demand reading of needed event info.
      - it will take:
        - » ~1 days for just making a W transverse mass plot.
        - » When the data is on your disk…
    - For making a more sophisticated analysis of an inclusive lepton sample, may take weeks!
    - Interactive analysis will be a real challenge to establish

    (Note: No mention of access time to data, etc.)

This can be a real competitive edge or a big problem for an experiment!

Reminder - Data formats

- RECO: (output of production) ~500kB/ev
  - Objects from all reconstruction algorithms.
  - Limited distribution!
    - CERN analysis facility for reconstruction/calibration community (not for analysis, in general)
    - Fraction stored at each Tier 1s.
    - Access and usage - centrally managed
- AOD: (Analysis Object Data) ~100kB/ev
  - Allows some re-calibration and reapplication of algorithms (b-tagging, jet finding). Maybe refitting.
  - Available at Tier 2s.
  - Intended steady state analysis format
- "Ntuples":
  - User defined, analysis specific data structures

Evolution of Analyses

Detector Commissioning

Low level reconstruction

High Level Objects

Peak hunting

Counting experiments

Cross-sections…

My view of “day-1”, “week-1”, “year-1” and such

Evolution of Analysis goals, implications…

- Detector commissioning
  - Timing
  - Cabling
  - …

Raw Data
Latch-all (no 0-suppression)
Basic detector DQM software
Local/global DAQ

- Comish. Low level reco.
  - Val. hits, segments, clusters
  - Calibrate
  - Noise, 0-suppression

Raw and RecHits
Local geometry, DB access
Low level reconstruction software
Establish 0-suppression asap!

- Comish. Object reco.
  - Establish e, m, jet… IDs
  - Define basic eff. - Z signal
  - Isolation!

Raw, RecHits and Reco data
Detector geometry, DB, Ca/Al…
Full reconstruction software, also HLT!
Basic trigger, Access to data

- Peak Hunting

Relative calibration, uniformity, efficiency.

- Counting experiments

Absolute efficiencies (tracking…), MC scale factors

- Cross section measurements

Luminosity

Timing - detectors, L1 trigger…

from Darin Acosta

Clock

- Is the heartbeat of the experiment
- Most of the front-end detector electronics and the Level-1 trigger electronics march to its beat
- LHC bunch crossing frequency: 40.0788 MHz
  - Approximately 25 ns bunch crossing (BX) spacing
- Need ~100 crossings to make a Level-1 trigger decision.

==> Therefore: digital electronic systems are pipelined, with the clock synchronized to the LHC frequency

  - Each clock edge marks the arrival of data from the next crossing
- Catastrophic error if experiment clock disrupted, or frequency changes

Trigger Synchronization

- For a synchronous system, one needs to add delays and adjust phases to keep data synchronized when collecting data from multiple boards (e.g. at the Global Trigger)
  - If not, different events get mixed up!
- This can be tricky
  - There are a lot of boards! But some delays can be calculated.
- Need to send periodic pulses to check time alignment, and look at the data itself for coincidences

Timing - Mini Analysis

Outline:

- Look for coincidences
- Compare info from different time slices and different crossings
- Match: signals, events, detectors
- Luminosity effects - yes, even pileup!! (bunch spacing)
  - Need to read the small print… See slide from

- Need
  - RAW data (not AOD or such)
    - W/O zero-suppressed
    - Info from multiple crossings
  - Triggers
    - Tagging only (no rejection)
    - All intermediate info kept

- Implications
  - Huge event sizes
  - Dedicated data taking conditions
  - Low rate (best done at low lumi)

Beam commissioning

Jörg Wenninger

Beam commissioning will proceed in phases with increased complexity:

- Number of bunches and bunch intensity.
- Crossing angle (start without crossing angle !).
- Less focusing at the collision point (larger 'β*').
- It cannot be excluded that initially the LHC will operate at 6 TeV or so due to magnet 'stability'. Experience will tell…

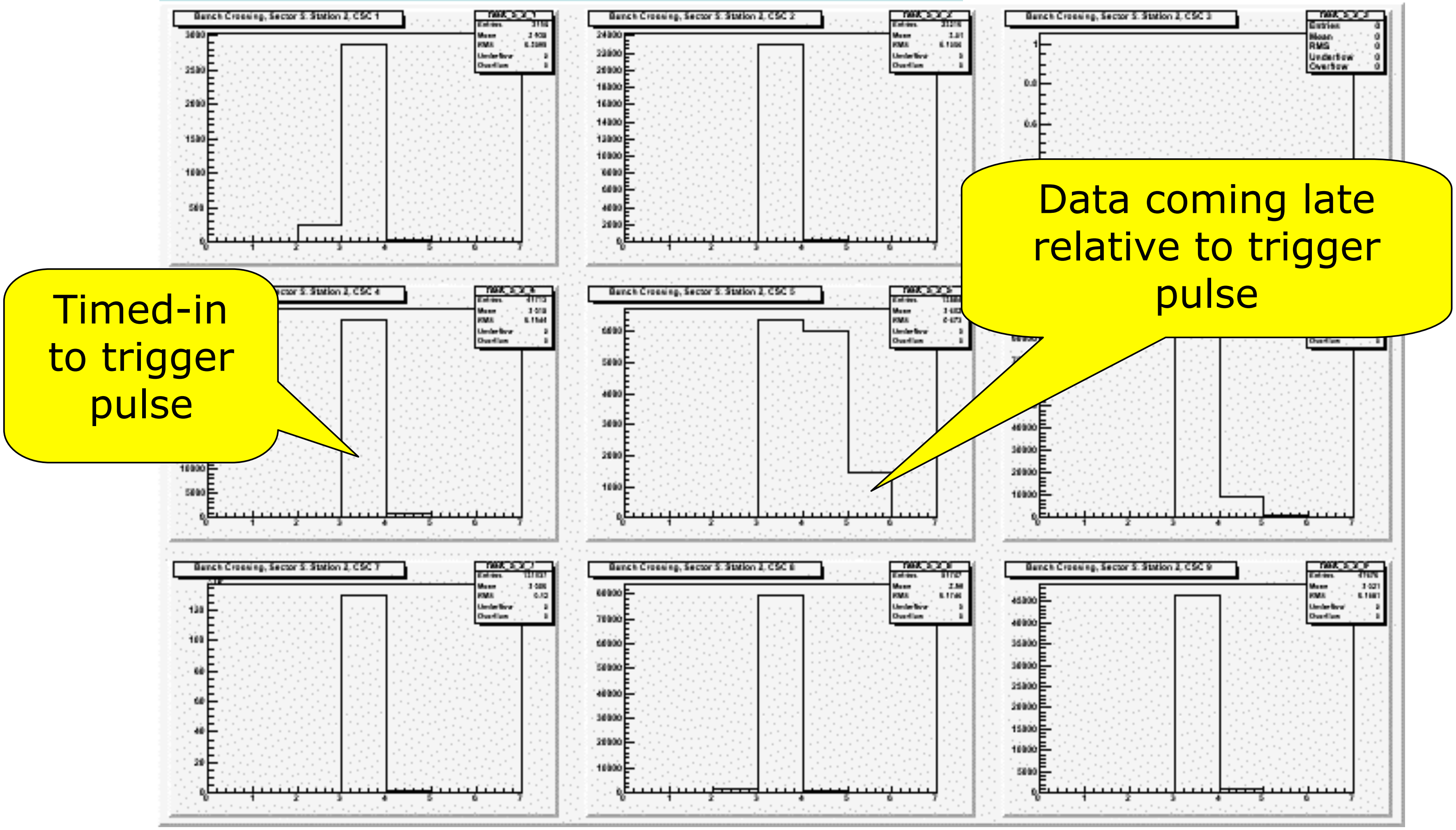
It will most likely take YEARS to reach design luminosity !!!

Parameter	Phase A	Phase B	Phase C	Nominal
k / no. bunches	43-156	936	2808	2808
Bunch spacing (ns)	2021-566	75	25	25
N (10^{11} protons)	0.4-0.9	0.4-0.9	0.5	1.15
Crossing angle (μrad)	0	250	280	280
$\sqrt{(\beta^*/\beta^*_{nom})}$	2	$\sqrt{2}$	1	1
σ* (μm, IR1&5)	32	22	16	16
L ($cm^{-2}s^{-1}$)	6×10^{30}-10^{32}	10^{32}-10^{33}	$(1-2)\times10^{33}$	10^{34}

detectors & trigger timing

from Darin Acosta

Different muon detector sections:



Trigger Signals, Relative to BX

Timing - High Level Trigger…

Event "assembly" and HLT

- The High Level Trigger of an experiment is typically asynchronous, as is the data acquisition system
- Thus instead of data marching synchronously, each fragment has a time marker attached to it:

  ==>event number
- Nevertheless, one still must make sure the marker assignment is correct so as to assemble fragments from the same collision

from Darin Acosta

Selective readout, Zero suppression…

- LHC detectors are:
  - Very granular
  - V

  ==> l
- Exam
  - C
  - A

Cannot p

Most are
in rec

- Issue
  - P
  - C
  - C
  - ..

  Wou
  s

How

Must perform many mini-analyses:

s

tions

e

ed,

with

t

aster

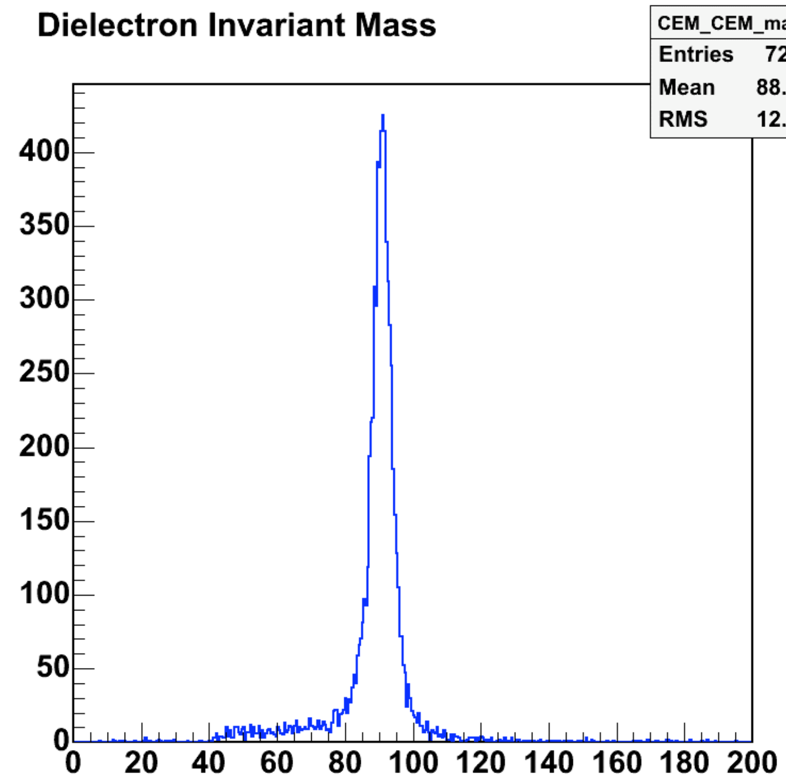
you turn-on your Physics menu!

Efficiency Measurements

- Lepton I.D., tracking efficiencies
  - Challenge: denominator!!
- Define a Z sample relying on:
  - One "tight" lepton selection
    - "anchor" leg
  - 2nd leg, unbiased
    - "probe" leg
  - Use Z mass spectrum
    - window, fit…
- Example:
- Can measure isolation, as well as reconstruc efficiencies

$$eff(Global\,\mu) = eff(track) \times eff(\mu - stub) \times eff(link)$$

- Need unbiased sample (inclusive)
- Statistics can be an issue
- Other examples:
  - W-notrak-trig
  - Min bias
  - Prescaled, lower threshold



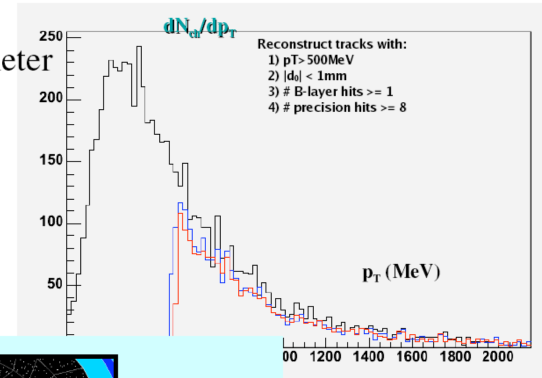
Underlying Event & Pile-Up

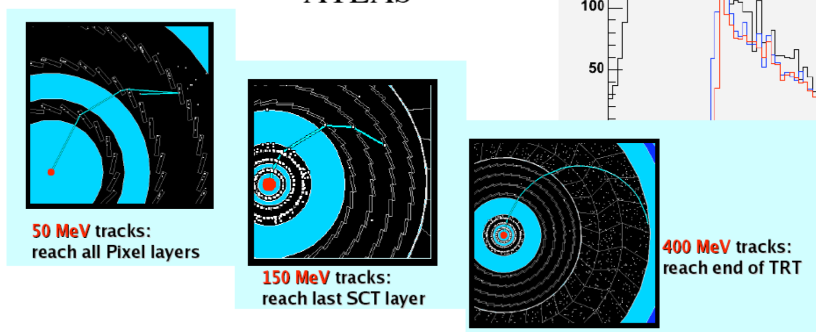
What is the difference?

Soft Tracks do not make it to calorimeter

from Ian Hinchliffe

ATLAS





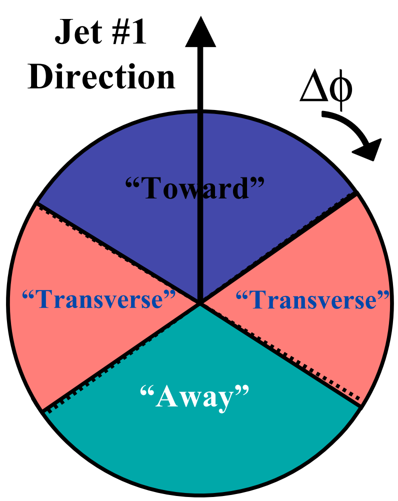
CMS has smaller radius, higher field

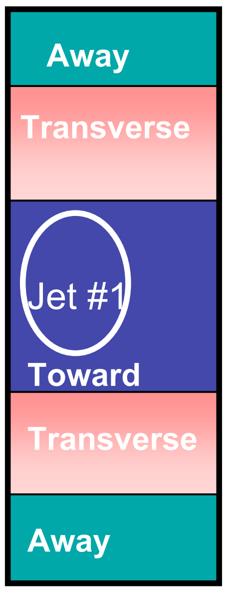
- Least understood
  - Must be measured
  - Used as input to MC
- Dominates occupancies of inner detectors
- Implications:
  - Data size
  - Tracking efficiency
  - Pattern recognition speed
  - fakes
  - …

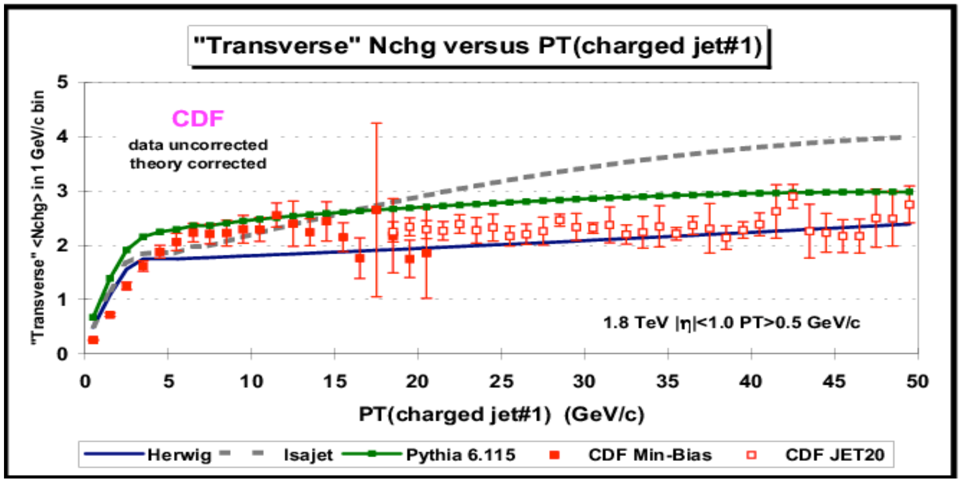
Ways to Measure U.E.

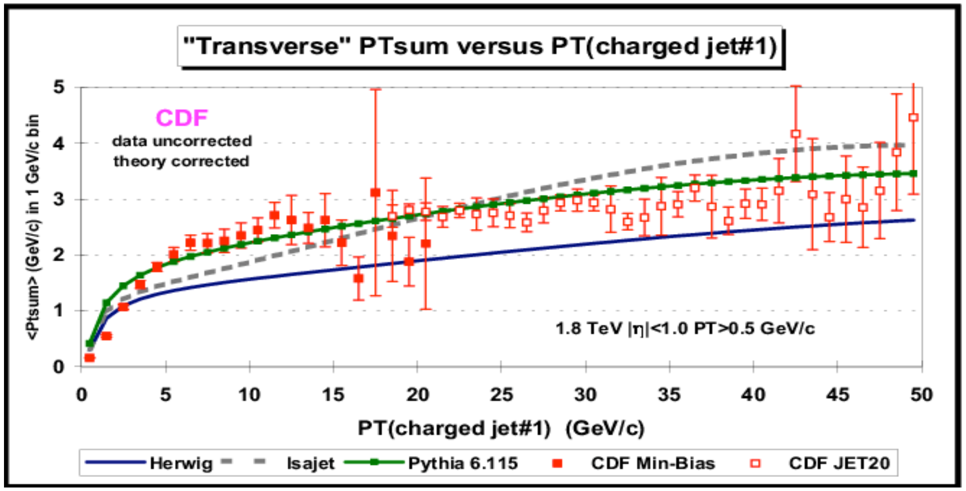
from Darin Acosta

- Look in regions transverse to jets in dijet events
  - Only slow growth with scale of hard scattering









Need: eff, uniform tracking

Background Estimates

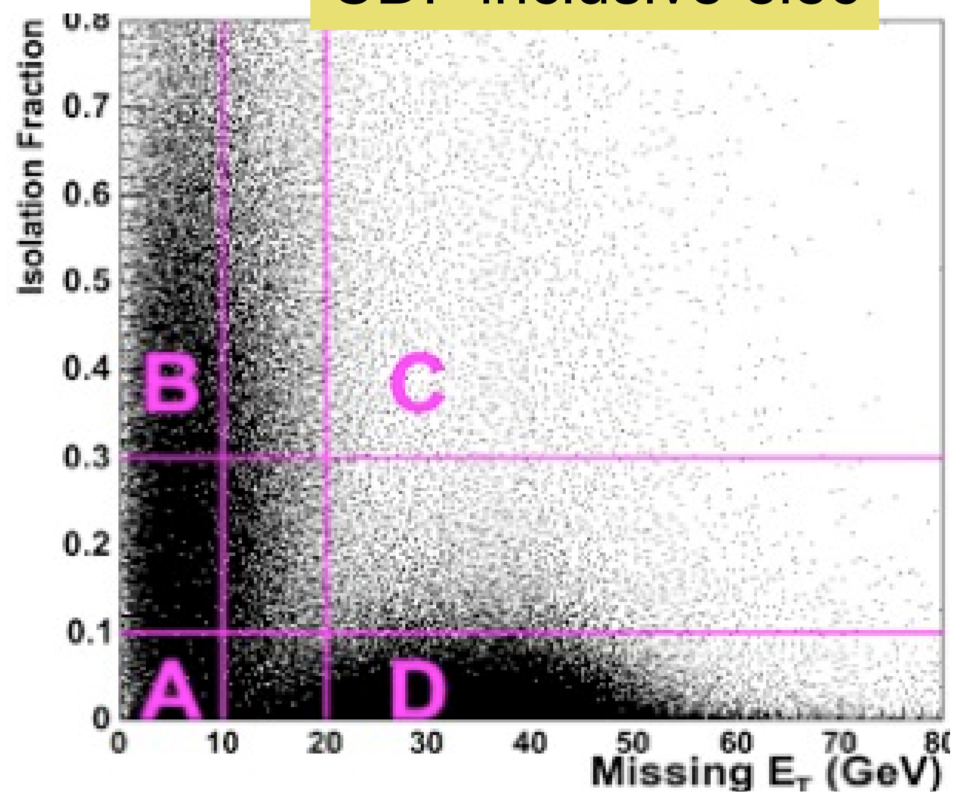
Do we have the tools? The data?

- Non-W background

- Consider a sample of W-->lν+X
- How do we estimate the QCD background?
  ==> Missing E_T Vs. Iso

- Since:
  1. Very different distributions of Missing E_T
  2. The Missing E_T is uncorrelated with the Isolation for QCD events.
- Areas:
  - A, B & C are QCD dominated
  - D is W-->lν dominated region

CDF inclusive elec



$$\frac{N_D}{N_C} = \frac{N_A}{N_B}$$

What do we do if Iso is part of trigger selection?!

A few analysis examples and some issues

Physics- what does one look for? - the signature
What is my trigger(s)?
backup triggers, samples

Physics, Signatures and Triggers

Some physics and their experimental signatures:

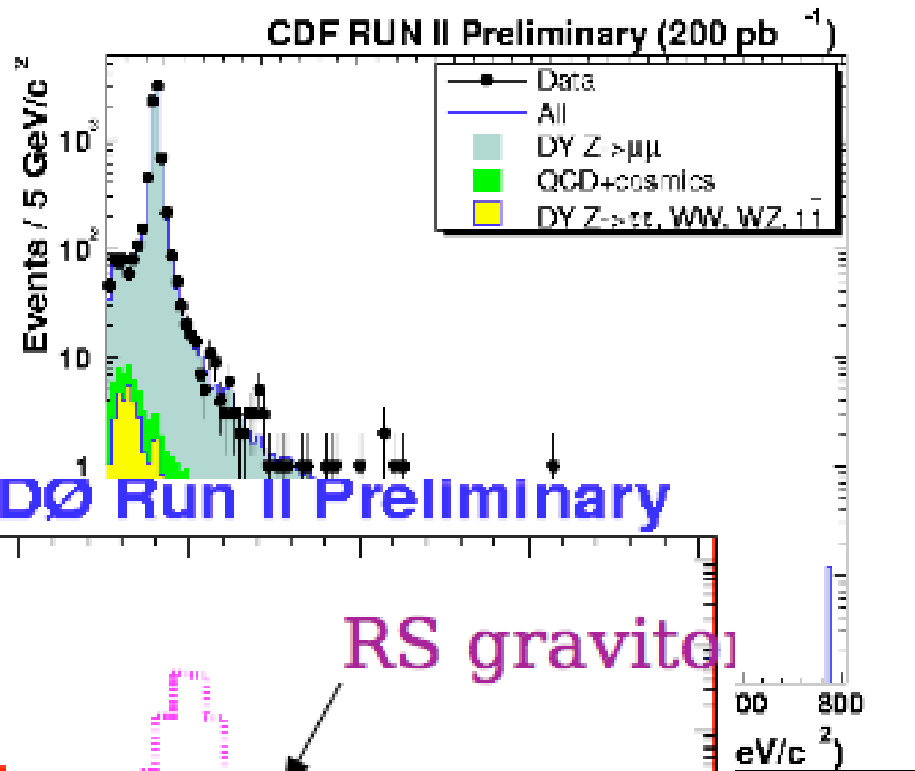
- Higgs
  - $\gamma\gamma$, bbar, WW, ZZ (peak)
- Supersymmetry
  - multi-leptons or same-sign lepton pairs
  - jets and Missing E_T
- Z'
  - di-electron, di-muon (peak)
- W'
  - electron or muon and Missing E_T
- Large Extra dimensions
  - jet + Missing E_T (mono-jet)
  - di-fermion, di-boson
- Compositeness
  - di-jet (hi mass tail)
  - lepton and jet (LeptoQuark)

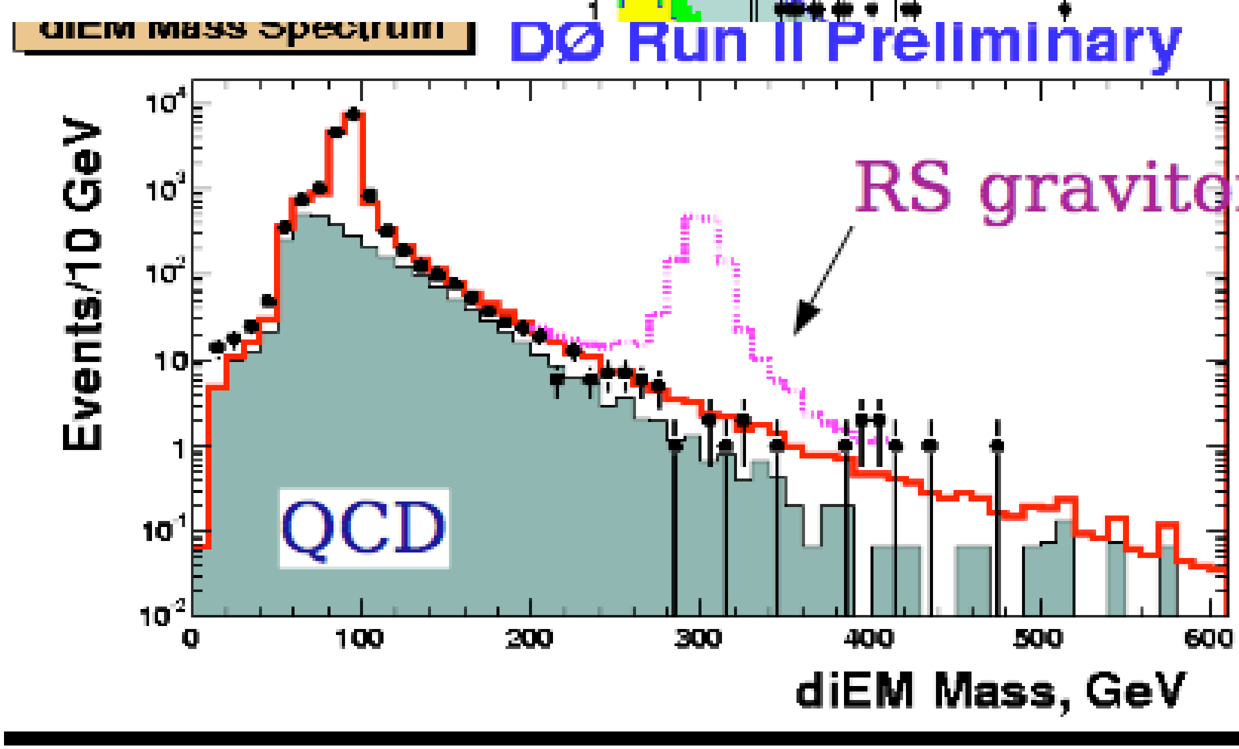
Corresponding "primary" triggers:

- di-photon
- di-electron
- di-muon
- di-jet
  - with b-tagging
- Inclusive leptons, either:
  - higher threshold
  - Isolation cut
  - Pre-scale applied
- Missing E_T .
  - Jet(s)
  - leptons
- "mixed" or "composite" triggers
  - …

High mass resonances

- Issues:
  - "self calibrating", "easy"…
  - cosmic background rejection
  - muons misbehave
  - …



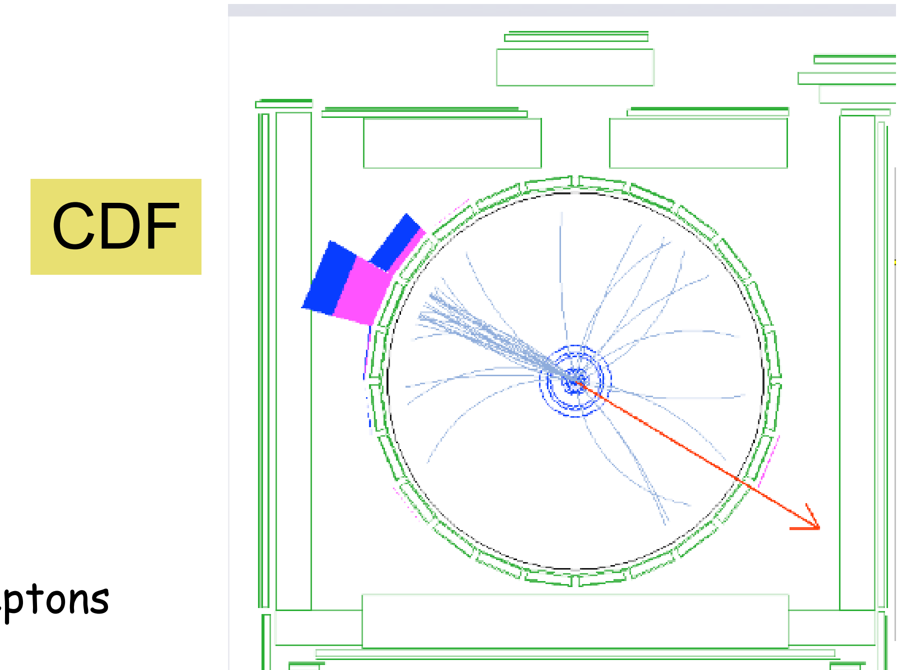


- Primary sam
  - di-muon

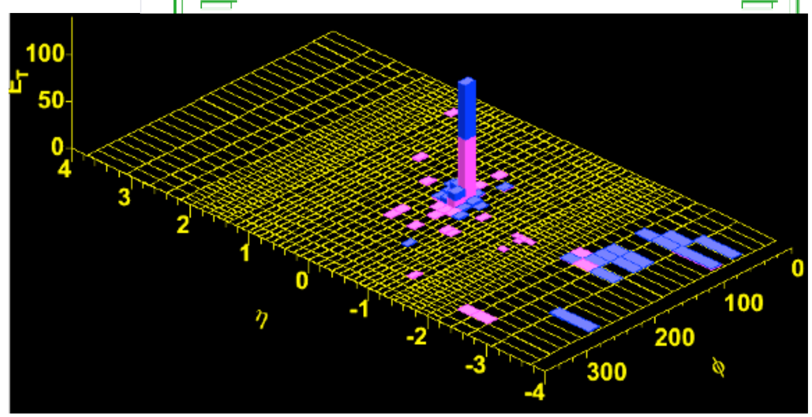
- Backup samp
  - inclusive n
  - lower thre
  - non-isolat

jet and Missing E_T (e.g. monojet)

- Issues:
  - Sizable SM backgrounds
  - Understand Missing E_T
  - Tricky counting experiment!

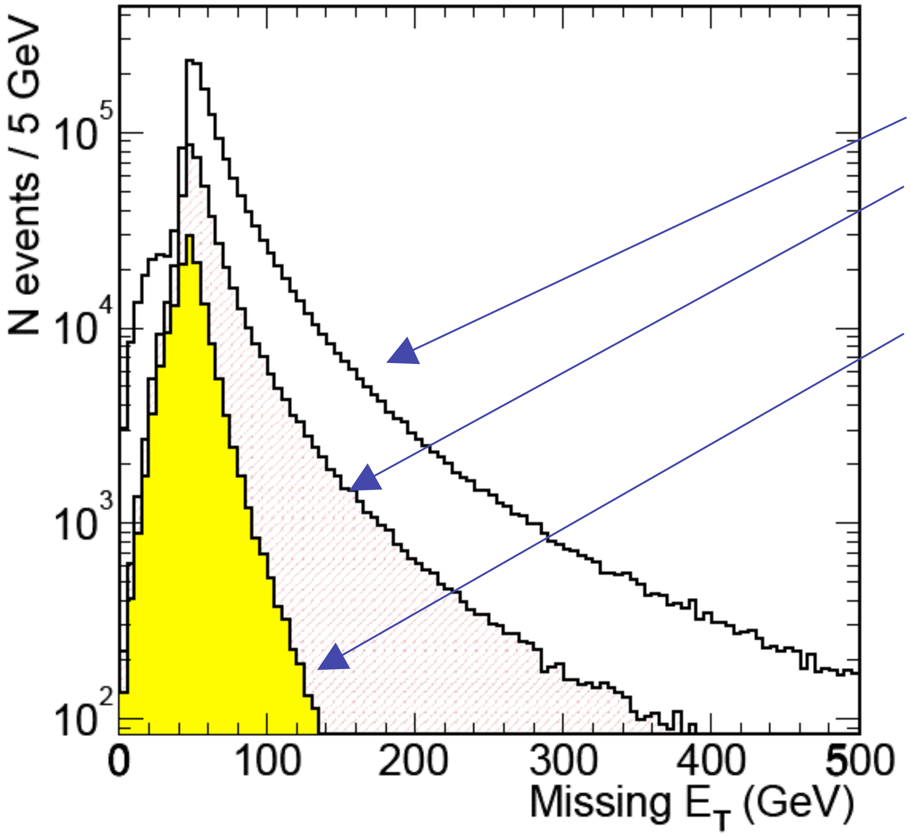


- "Primary" sample: Inc. jet 100
- Main SM backgrounds:
  - Z-->νν+jet - Irreducible
    - Use Z-->μμ sample
  - W-->lν+jet (lose l) - Real Met, veto leptons
    - Use W-->e/μν samples
  - QCD - Fake missing Et
    - Use di-jet samples
  - Non-Collision
    - Cosmic, beam



- Backup samples:
  - Inclusive jet triggers
  - Inclusive Missing E_T trigger

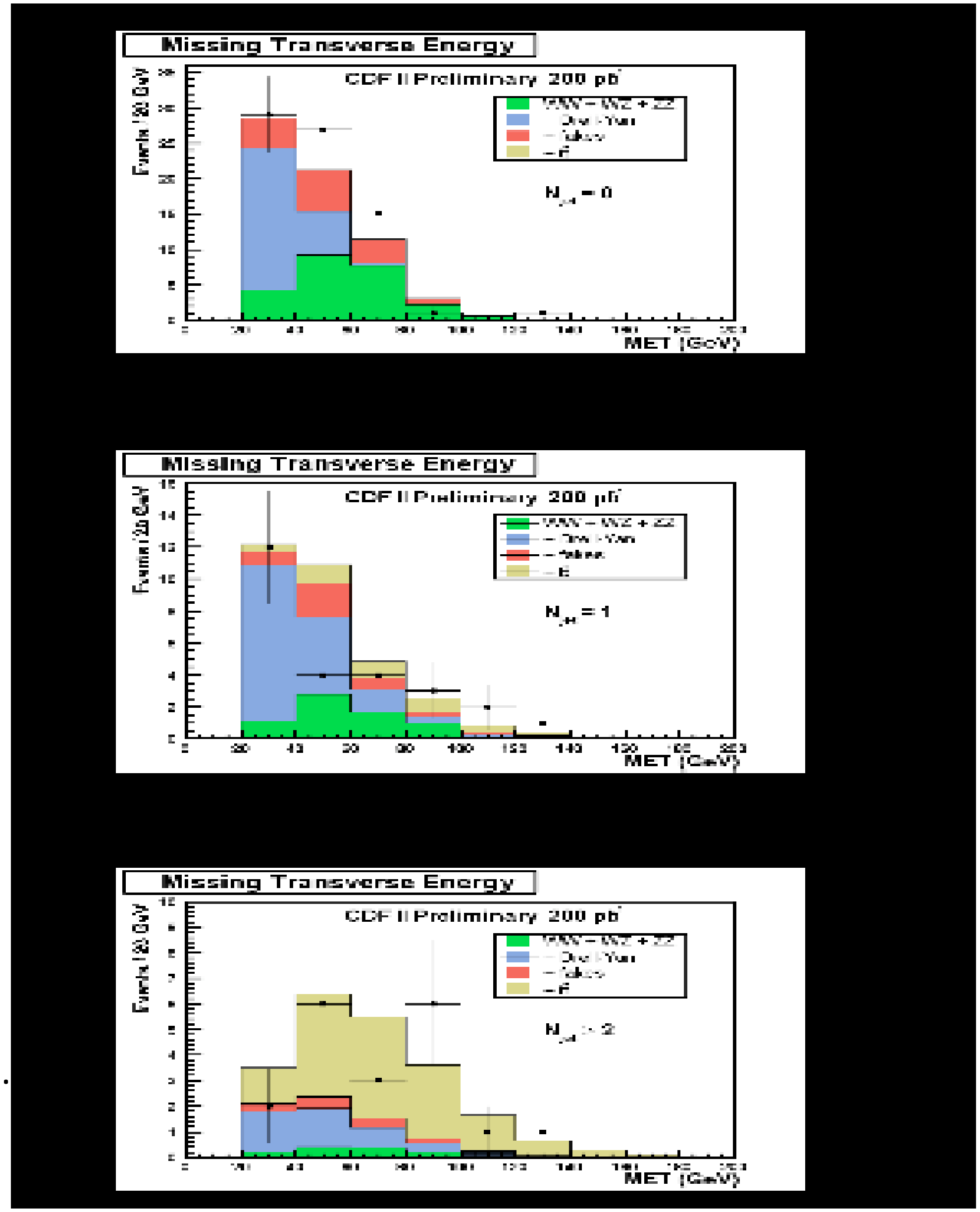
Effect of MET Clean-up (CDF, Run 2)



- After good run selection from DQM
- After a vertex requirement in tracking detectors
- After remaining preselection cuts
  - Event EM fraction > 0.1
  - Event charged fraction > 0.1
  - At least one central jet $E_T > 10$ GeV
  - Total calorimeter energy < $\sqrt{s}$

di-leptons and Missing E_T

- A tough counting experiment!
- Need to study and understand:
  - WW, WZ, ZZ, DY, and top
    - Of course, fakes of all sorts
  - As a function of:
    - Leptons Pt
    - Jet Et
    - Missing Et
- Look for deviations in rates
  - Rediscover top
- Look for deviations in shapes
  - Missing Et
  - HT
  - Leptons pt's
  - …
- First published hi-PT Runii analysis from Tevatron
  - 3 yrs after nominal startup date..



Trivial observation from the top search

Triggers & Data sets used

- Relied primarily on inclusive electron and inclusive muon triggers
- Many backup triggers needed:
  - Various inclusive jet samples for jet corrections
  - Gamma-jet sample for E-scale
  - W-notrack trigger for tracking efficiency
  - Jet50 sample for b-tagging MC scale factor
  - Many many more

A "family" of analyses

- Three main classes
  - Di-lepton
  - Lepton+jets
  - All hadronic

  All predicated on the basic trigger path, or primary data set.
- Finer split within each
  - Electron, Muon or jets as "anchor leg" (trigger requirement)
  - additional features (e.g. vertex b-tagging, SLT…)

To Re-cap:
Evolution of Analysis goals, implications…

- Detector commissioning
  - Timing
  - Cabling
  - …
- Comish. Low level reco.
  - Val. hits, segments, clusters
  - Calibrate
  - Noise, 0-suppression
- Comish. Object reco.
  - Establish e, m, jet… IDs
  - Define basic eff. - Z signal
  - Isolation!
- Peak Hunting
- Counting experiments
- Cross section measurements

Raw Data
Latch-all (no 0-suppression)
Basic detector DQM software
Local/global DAQ

Raw and RecHits
Local geometry, DB access
Low level reconstruction software
Establish 0-suppression asap!

Raw, RecHits and Reco data
Detector geometry, DB, Ca/Al…
Full reconstruction software, also HLT!
Basic trigger, Access to data

Relative calibration, uniformity, efficiency.

Absolute efficiencies (tracking…), MC scale factors

Luminosity

Thank You and Good Hunting!