Analysis at a Hadron Collider
Lecture 1: Introduction

Doug Glenzinski
Fermilab
January 2010

- Lecture 1: Introduction
- Lecture 2: Searches
- Lecture 3: Measurements

Preamble

- Aimed at mid-level graduate students
  - Assume general knowledge of Standard Model
  - Assume familiarity with nomenclature and terminology
  - Assume limited practical experience
- Will discuss some general analysis guidelines
  - Hopefully not overly pedantic
  - Aim to provide something useful
- Will review some important experimental methodologies
  - Begin with generalities and then concentrate on specifics
  - Real life examples lifted from CDF

Preliminaries

- I am an active member of the CDF collaboration
  – I take many examples from CDF
  – Performance, strategy, etc. very similar between CDF/D0 and LHC experiments
  – Differences noted when important
- I gratefully acknowledge the help of several people in preparing these lectures:

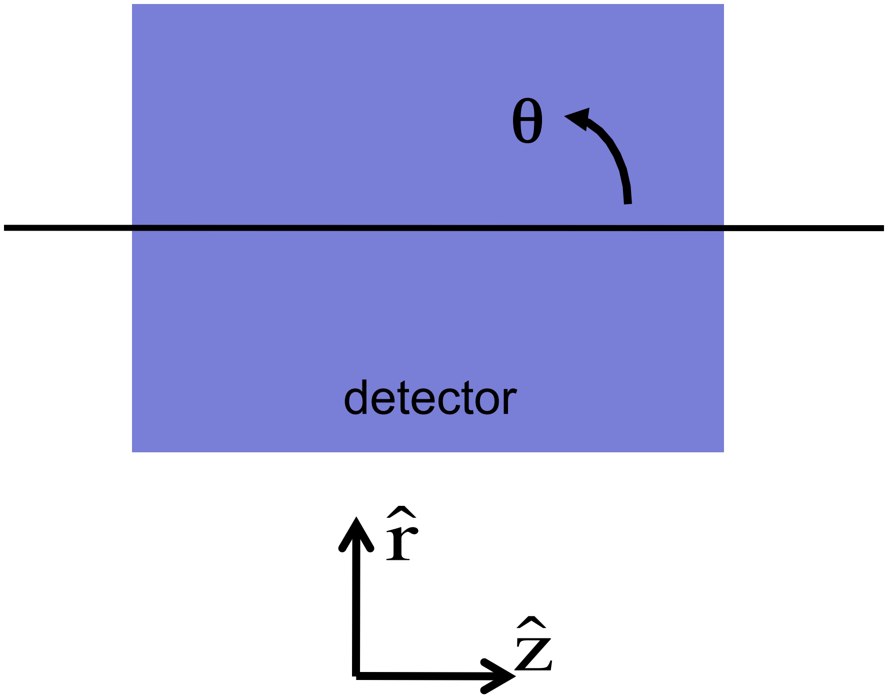
  D.Green, B.Heineman, J.Incandela, J.Konigsberg,

  K.Lannon, T.Lecompte, M.Mangano, C.Neu,

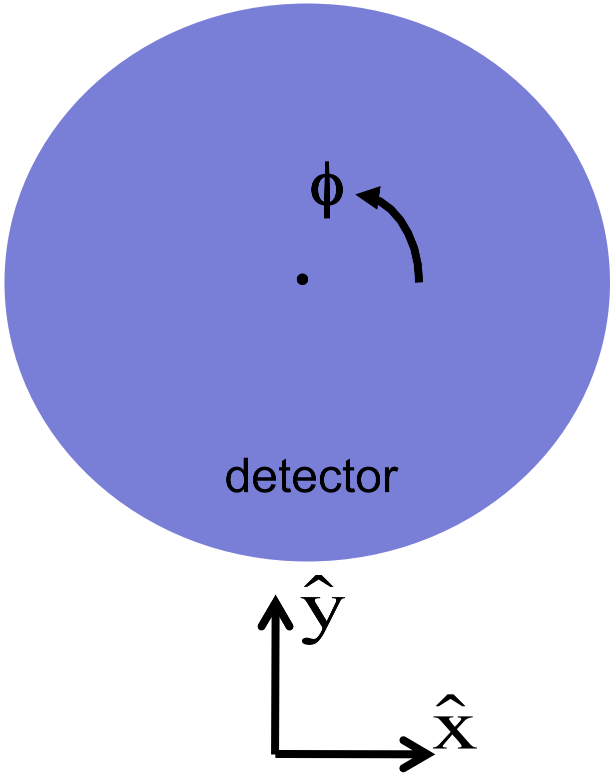
  A.de Roeck, R.Roser, I.Shipsey

Preliminaries

- All collider experiments employ a right-handed coordinate system with z along the beam line



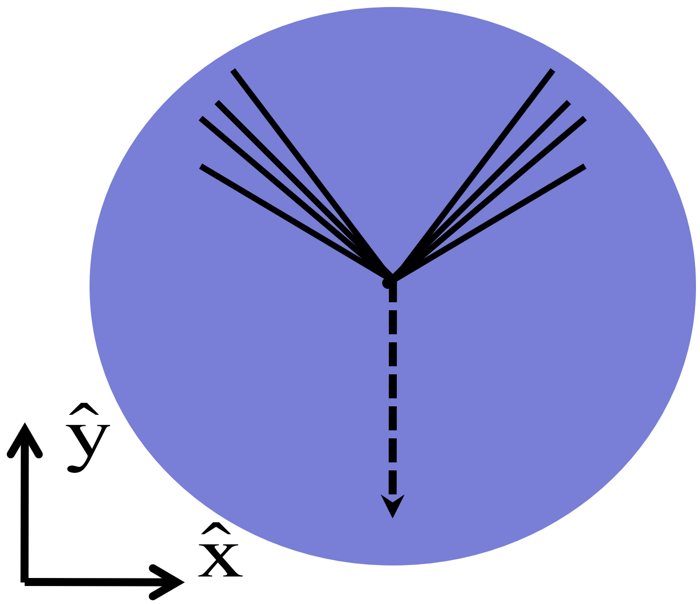
“r-z plane”



“x-y”, “r-ϕ”, “transverse plane”

Preliminaries

- All hadron collider experiments talk about "missing transverse energy" (really it's missing momentum)



$$\not{E}_T = -\left| \sum_{i=towers} E_i \cdot \hat{n}_T^i \right|$$

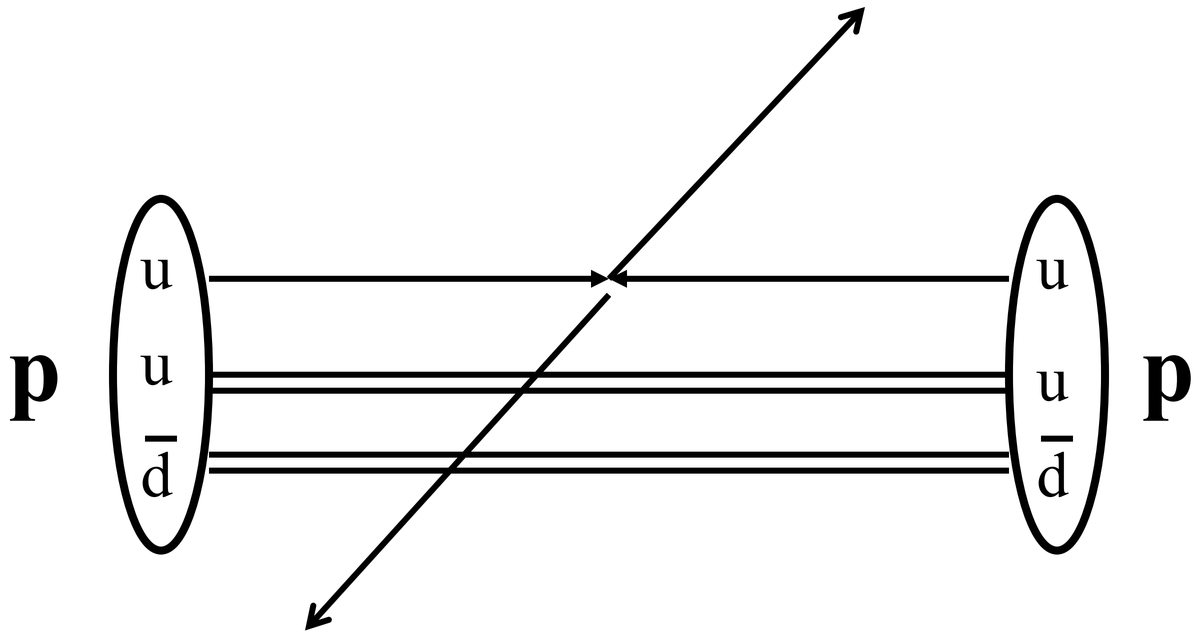
where $\hat{n}_T^i$ is the unit vector from the interaction vertex to the tower center in *xy* plane

"missing E_T" "MET" "$\not{E}_T$"

Today's Outline

- Hadron Collider Physics Introduction
  - Basic concepts
  - Their experimental consequences
- Hadron Collider Experiments
  - Basic concepts
  - Their strengths and limitations
- Analysis Strategy
  - General guidelines
  - Specific Examples in next lectures

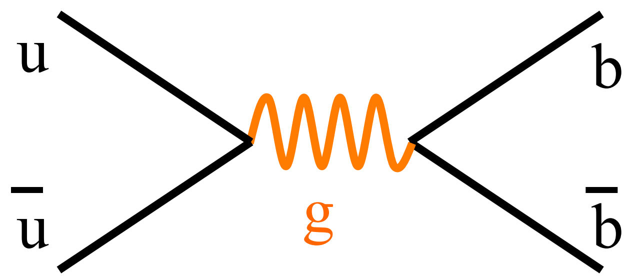
Hadronic Collisions



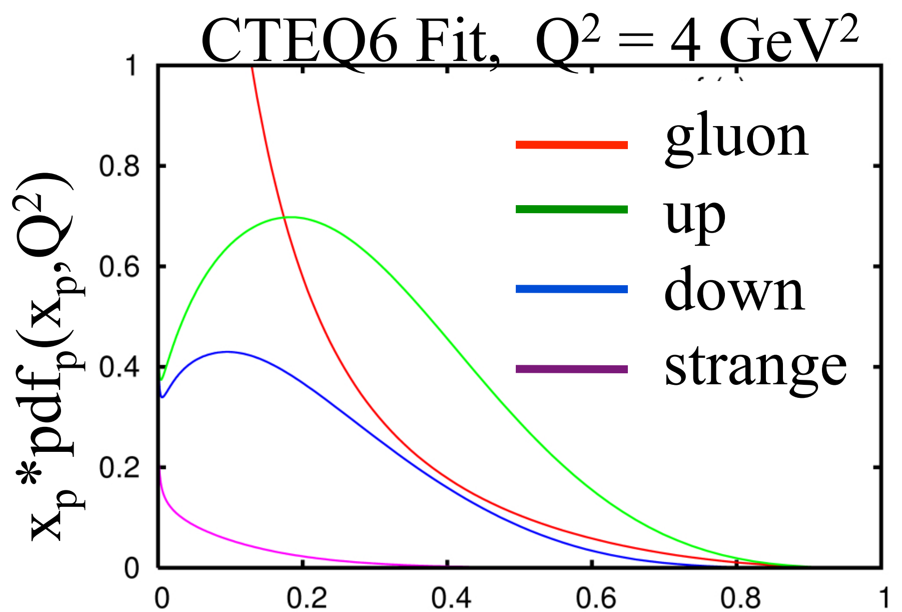
- The physics processes we are most interested in are interactions between the constituent partons and *not* between the protons themselves

Hadronic Collisions

- That protons have partonic constituents has some important experimental consequences
  1. To make use of theory calculated cross sections you need to integrate over the energy distribution of the constituent partons



Hadronic Collisions



x_p = fractional momentum carried by parton

- The parton distribution functions (pdf) estimated from experiment
- Available as software packages (e.g. CTEQ, MRST)
- Source of uncertainty for all calculations and measurements

Hadronic Collisions

- That protons have partonic constituents has some important experimental consequences

2. The initial state is only partially known. For a symmetric collider (e.g. Tevatron, LHC)

   - p_T(initial) = 0
   - p_Z(initial) = ?
     (and in general != 0)
   - $E_{total} < 2E_{beam} = E_{cm}$

Hadronic Collisions

- That protons have partonic constituents has some important experimental consequences

3. The rest of the proton contributes to the event
   - The remnants can radiate gluons
   - The remnants are color connected to the initial state of the hard scattering interaction
   - Give rise to the "Underlying Event", the set of tracks and energy deposits uniformly populating the detector but unassociated with the hard scatter interaction

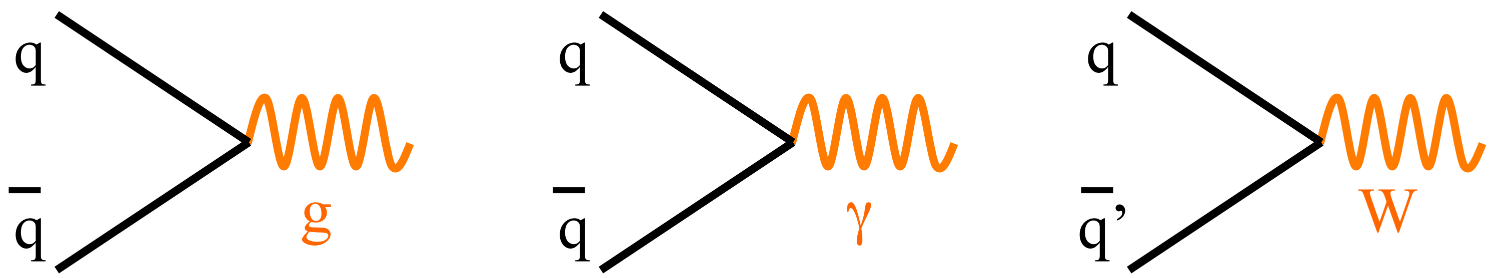
(Note: not to be confused with Multiple Interactions)

Hadronic Collisions

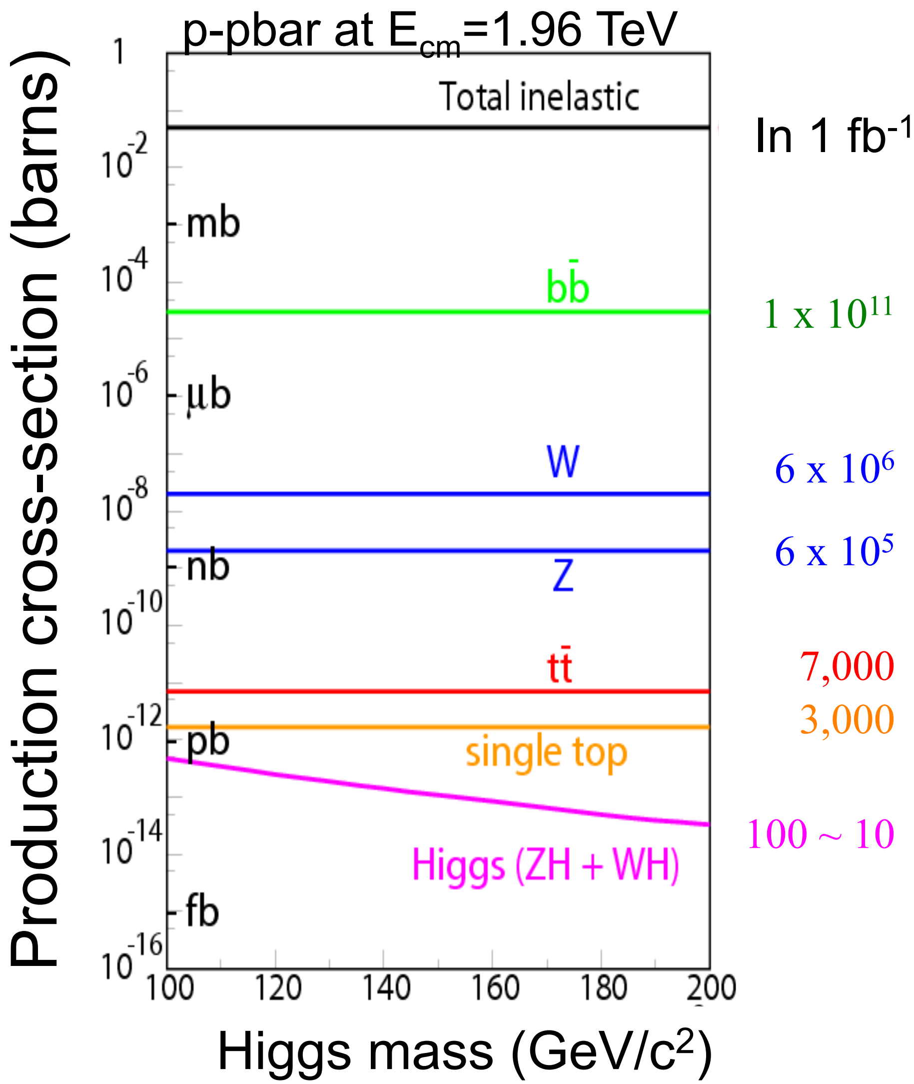
- That protons have partonic constituents has some important experimental consequences

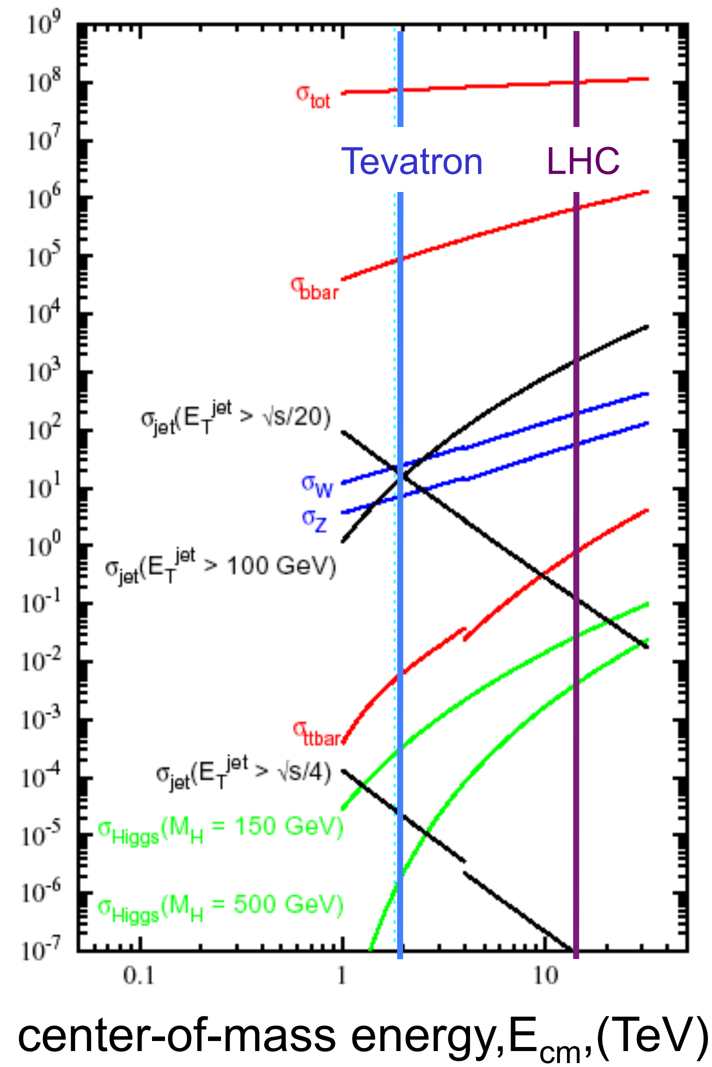
    4. The protons partonic constituents carry color and charge

        - Quarks participate in Strong, Weak, and Electromagnetic interactions



Hadronic Cross Sections





Hadronic Cross Sections

- Some observations:
  - Wide variety of processes produced
    - Enables rich physics program
    - Enables (largely) data driven analysis techniques via suitably defined control samples
  - Production XS span 12-13 orders of magnitude
    - Collision rate dominated by the mundane
    - Background discrimination crucial
    - Multiple interactions per crossing possible

Hadronic Cross Sections

- The scale and breadth of hadronic interactions have some important experimental consequences.

  1. Can’t save every event to tape

     - Need to identify most interesting events in real time and toss the rest (event rate dominated by the mundane)
     - Ideally, will keep some events from all processes (to provide physics breadth and control samples)

Hadronic Cross Sections

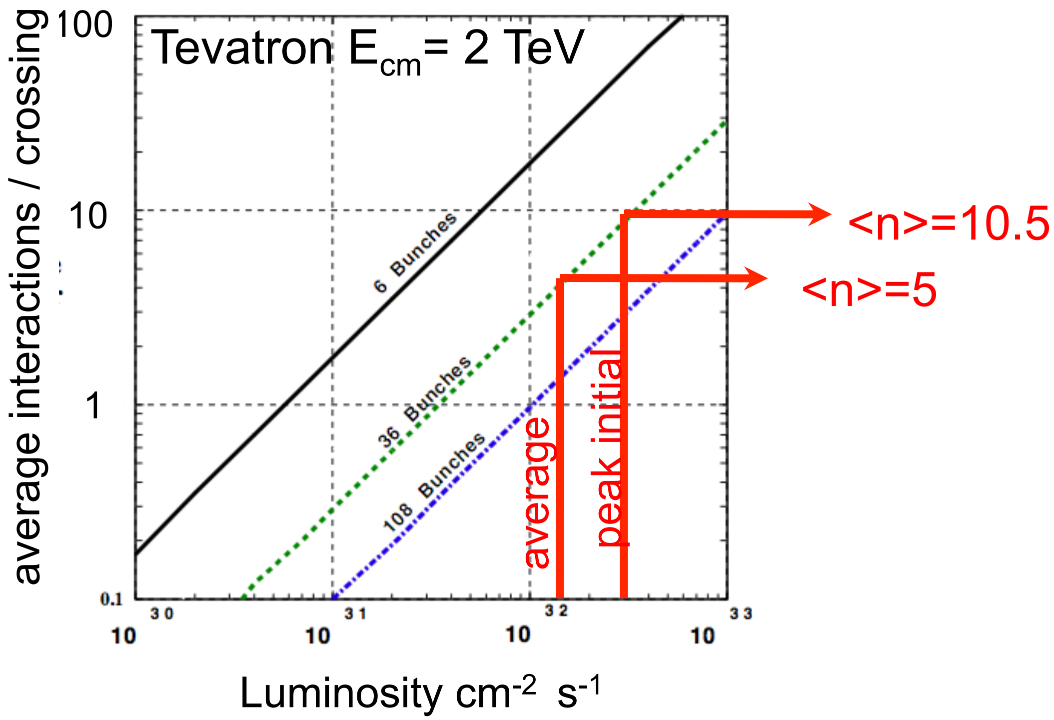
- The scale and breadth of hadronic interactions have some important experimental consequences.

2. Usually, the (mundane) Backgrounds have rates much larger than signal.
   - Need to identify characteristics which can suppress the background
   - Need to demonstrate solid understanding of background rate and shapes

Hadronic Cross Sections

- The scale and breadth of hadronic interactions have some important experimental consequences.

3. There can be multiple interactions per crossing



Detector Implications

Hadron Collider "Facts of Life"

- Initial state largely unconstrained
  - Only know that initial state $p_T = 0$
  - Often work in the transverse plane
  - Usually choose B-field parallel with beam line
- Cross sections large, many processes contribute
  - Broad physics program possible – must choose which evts to keep
  - Drives design of Trigger and DAQ
  - Backgrounds often large and varied - particle ID important
- Each event has contributions beyond the hard sctr
  - Underlying event (always; from proton remnants)
  - Multiple interactions (only when lumi high enough)
  - Precision vertexing and fine segmentation helpful

Hadron Collider Detectors

Main objectives:

1. Provide, fast, flexible, and efficient triggering

   (choose which events to keep)

2. Provide precision vertexing

   (identify the hard scatter vertex)

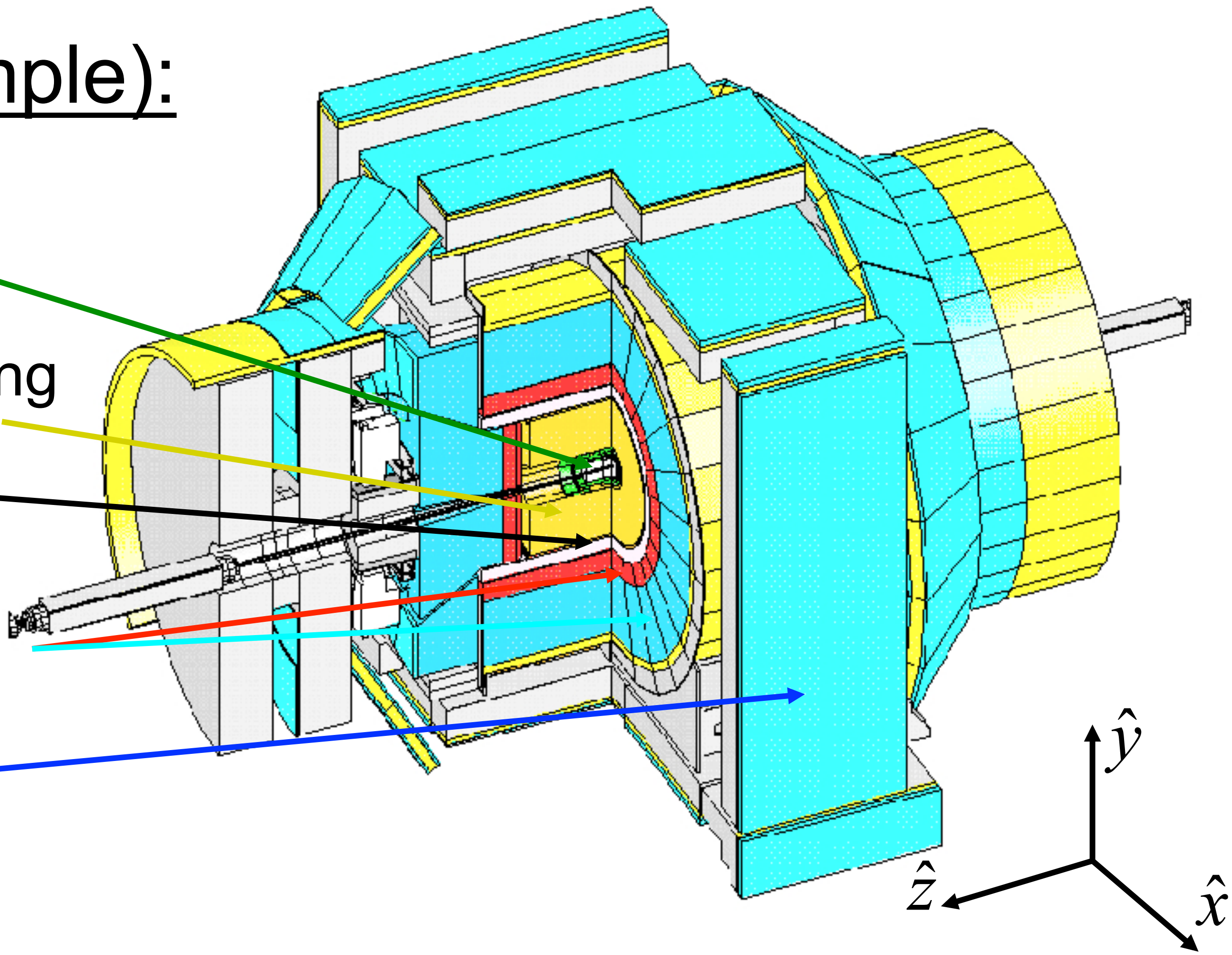
3. Measure charged particle trajectories and p_T
4. Precisely measure EM and Hadronic Energies
5. Identify muons

(2-5 work together to provide particle id, which is crucial to achieving necessary background suppression)

A Collider Detector

<u>CDF (for example):</u>

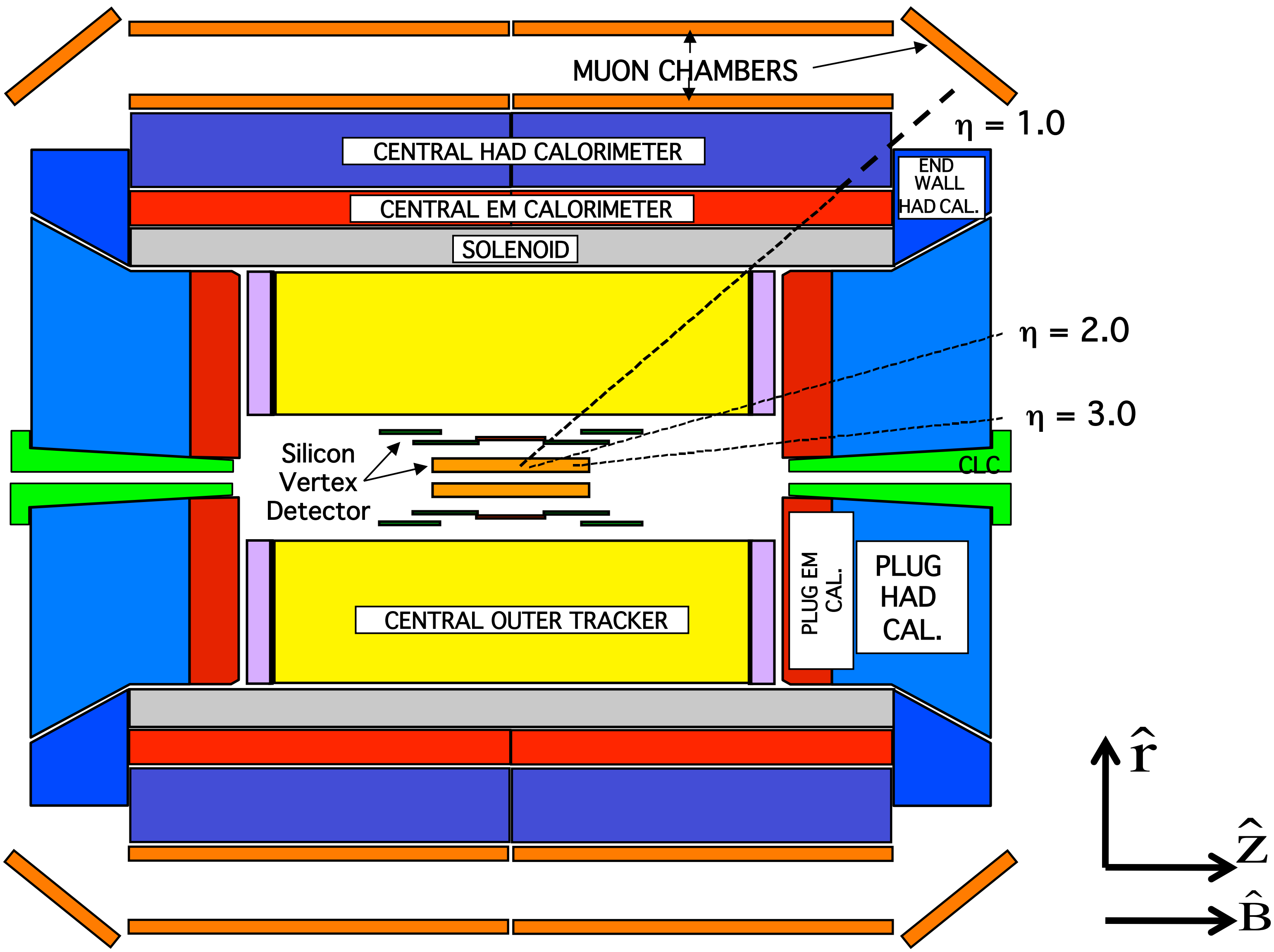
- precision silicon vertexing
- large radius tracking
- solenoid
- EM and Hadronic calorimetry ($|\eta|<3$)
- muon chambers



$$\eta = -\ln(\tan(\theta/2))$$

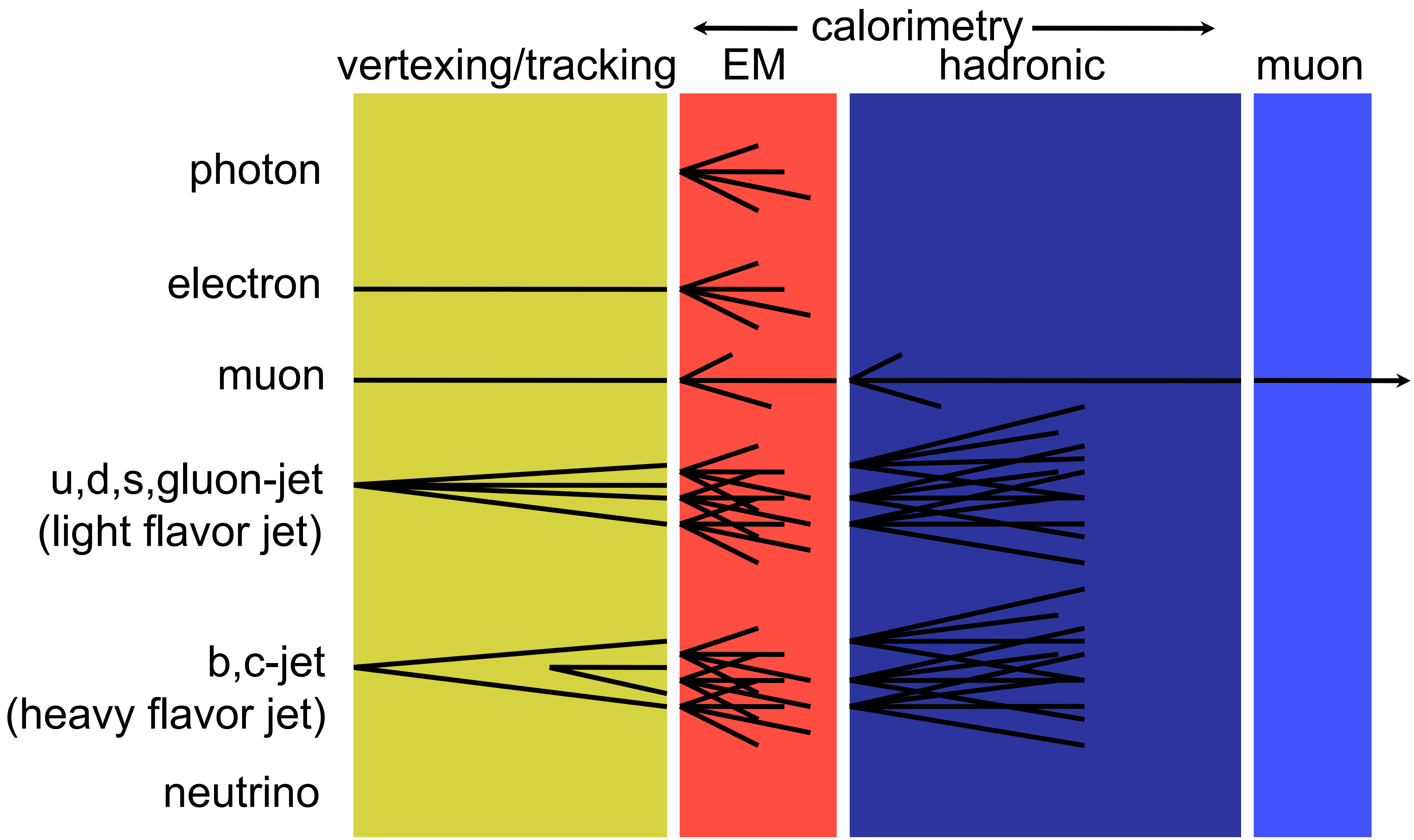
A Collider Detector

- CDF (for example)



Particle ID in 1000 Words

- Schematically, e, γ, μ, ν, b-jets identified like this:

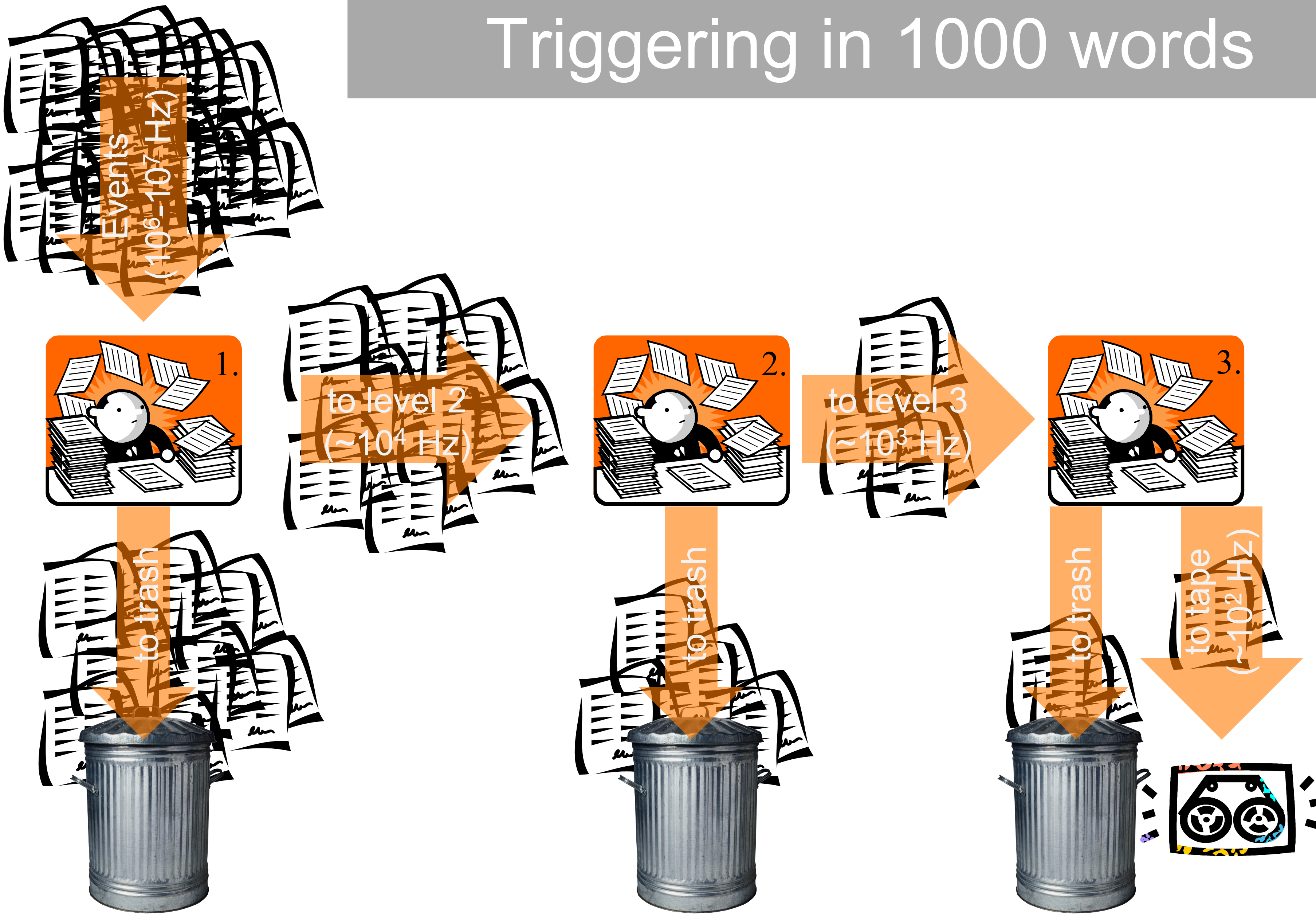


Resolutions

- Useful expressions
  - Momentum: $\frac{\delta P}{P} = \alpha \cdot P \oplus \beta$
  - Energy: $\frac{\delta E}{E} = \frac{\alpha}{\sqrt{E}} \oplus \beta$
  - Impact parameter: $(\delta d_0)^2 = \left(\frac{\alpha}{P}\right)^2 (\sin\vartheta)^{-3} + \beta^2$
- Some numbers

	CDF	D0	
$\delta M(J/\psi \rightarrow \mu\mu)$:	12	60	(MeV/c^2)
$\delta M(Z \rightarrow \mu^+\mu^-)$:	2.5	6.0	(GeV/c^2)
$\delta M(Z \rightarrow e^+e^-)$:	3.0	3.0	(GeV/c^2)
$\delta E_{jet}/E_{jet}$:	16	14	(%)
δd_0(at $p_T \sim 1$ GeV/c) :	30	30	(μm)

Triggering in 1000 words



- Multi-leveled approach usually employed

Triggering

The Trigger is crucial to your success

- If done well, the trigger menu
  - Enables a broad physics program
  - Provides control samples for data-driven analysis
- The details of a given trigger path effect the resulting data set by
  - Affecting which production processes contribute
  - Which in turn affects the flavor composition

Triggering Strategy (from CDF)

Broadly speaking, two prong strategy:

- Inclusive triggers designed to collect your signal samples (mostly un-prescaled) e.g.
  - High-pT e/μ/γ (pT>20 GeV), jets (pT>100 GeV)
  - Multi-object events: e-e, e-μ, μ-μ, e-τ, e-γ, μ-γ, etc.
- Back-up triggers designed to spot problems, provide control samples (often pre-scaled)
  - Jets (pT>8, 20, 50, 70 GeV)
  - Inclusive leptons (pT>4, 8 GeV)
  - Lepton+jet

Trigger Example (from CDF)

Inclusive High-p_T Central Electron Trigger Path:

Level 1

- EM Cluster $E_T > 8$ GeV
- $r\phi$ Track $p_T > 8$ GeV

Level 2

- EM Cluster $E_T > 16$ GeV
- Matched Track $p_T > 8$ GeV
- Hadronic / EM energy < 0.125

Level 3

- EM Cluster $E_T > 18$ GeV
- Matched Track $p_T > 9$ GeV
- Shower profile consistent with e^-

Will efficiently collect
W, Z, tt, tb, WW, WZ, ZZ, $W\gamma$, $Z\gamma$, W', Z', $\chi^0\chi^+$, etc.

Only one of these needs to measure trig effncy
can use SM candle (e.g. W or Z XS) then use with confidence in searches

Trigger Example (from CDF)

Back-up Triggers for CENTRAL-ELE-18 path:

W_NOTRACK

- L1: EM E_T > 8 GeV && MET > 15 GeV
- L2: EM E_T >16 GeV && MET > 15 GeV
- L3: EM E_T >25 GeV && MET > 25 GeV

NO_L2

- L1: EM E_T > 8 GeV && rϕ Track p_T > 8 GeV
- L2: AUTO_ACCEPT
- L3: EM E_T > 18 GeV && Track p_T > 9 GeV
  && Shower Profile Consistent with e^-

NO_L3

- L1: EM E_T > 8 GeV && rϕ Track p_T > 8 GeV
- L2: EM E_T > 8 GeV && Track p_T > 8 GeV
  && Energy at Shower Max > 3 GeV
- L3: AUTO_ACCEPT

Modular inputs allow factorization
- efficiency for track and EM inputs determined separately

Use resolution at L2/L3 to improve purity
- only really care about L1 efficiency near L2 threshold

Trigger Example (from CDF)

Inclusive, Redundant Inputs helpful

L1_EM8_PT8 feeds

- Inclusive high-pT central electron paths
- Dilepton paths (ee, eμ, eτ)
- Several back-up triggers
- 15 separate L3 trigger paths in total

A ttbar cross section analysis uses

- Inclusive high-pT central e^- paths
- Inclusive high-pT forward e^- paths
- MET+jet paths
- Muon paths

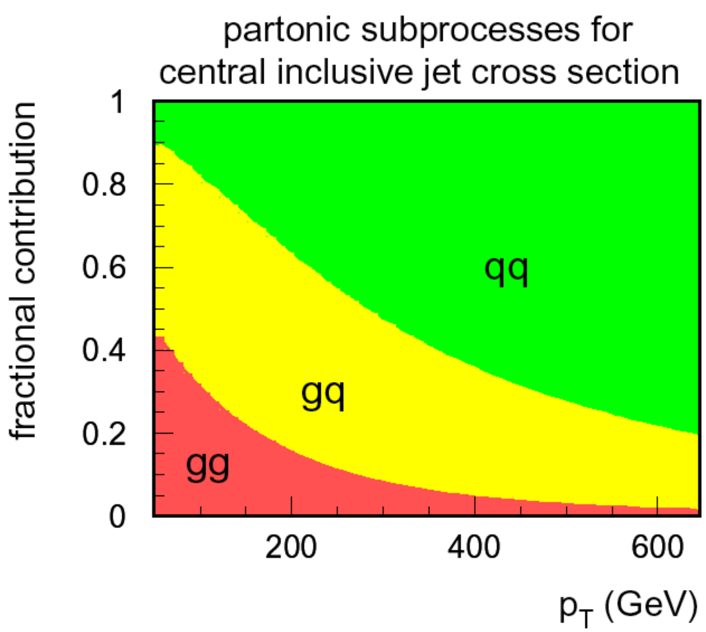
Inclusive → Analysis
once eff characterized
many analyses benefit

Inclusive → Improvemnt
one analysis picks up
where other left off

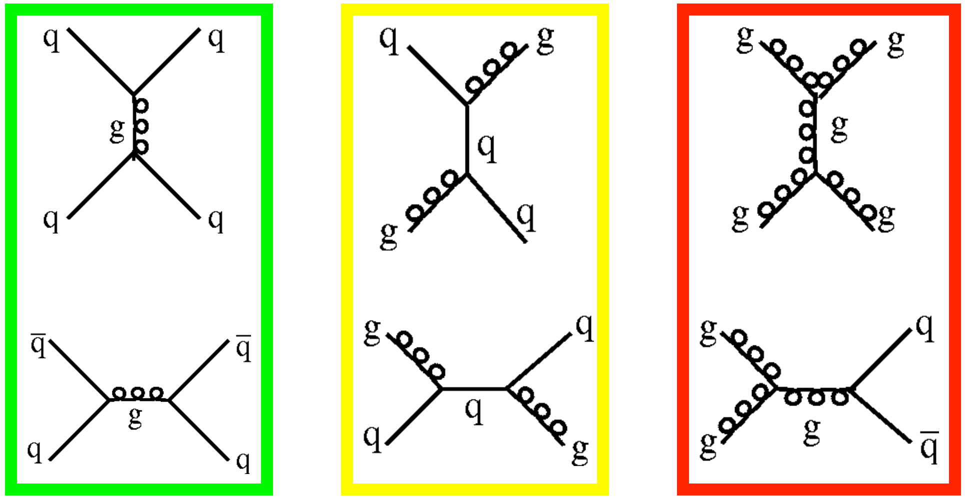
Inclusive → Monitoring
a problem shows-up
many places

Redundancy → Robust
limited physics impact
due to a problem in one
set of inputs

Triggering and Sample Composition (I)



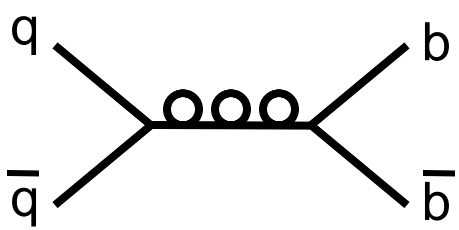
Inclusive jet processes: qq, qg, gg



- The relative contributions depend on jet-pT… affects quark/gluon ratio in final state

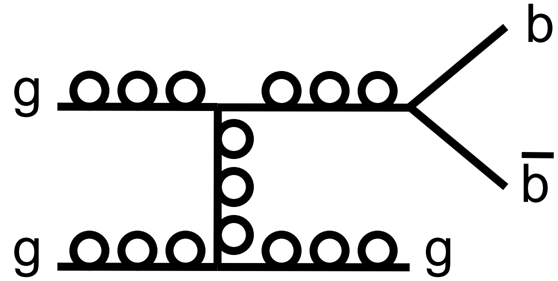
Triggering and Sample Composition (II)

Inclusive b-quark production:



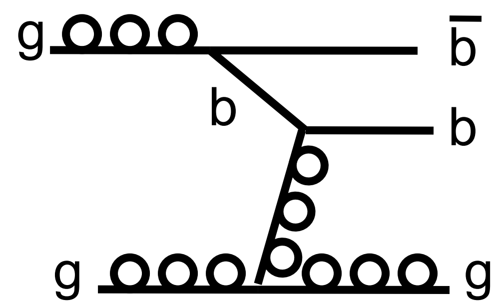
Direct Production

($\Delta\phi_{bb} \sim \pi$)



Gluon Splitting

($\Delta\phi_{bb} \sim 0$)



Flavor Excitation

($\Delta\phi_{bb} \sim$ flat)

- Heavy Flavor (hf) composition and kinematics depend on jet-pT and angular requirements

Implications for Analysis

Analysis Challenges

- In parent trigger sample, for most all analyses, Background orders of magnitude > Signal
  - Necessary to employ particle ID to suppress Bgd
  - Necessary to demonstrate thorough understanding of the Bgd using control samples
- Choice of trigger path affects sample composition
  - Takes time to characterize triggered sample
  - Custom Monte Carlo samples often required
  - Extrapolations from control samples uncertain

Analysis Challenges

- *a priori* uncertainties arise from pdfs
  - Introduce uncertainty in theory predictions
  - Introduce uncertainty in experimental acceptances
- *a priori* uncertainties arise from “other” contributions in the events
  - Must account for underlying event
  - Must discriminate and correct multiple interactions
- A complete analysis requires a range of expertise (e.g. theory, detector hardware, analysis software)
  - No one has expertise across *all* of these things

Analysis Strategy

Analysis Strategy

- There is no one way to perform data analysis
  - Depends a bit on what the analysis aims to do
    - Cross section measurement?
    - Determination of a particle property?
    - Search for something new?
  - Depends a bit on your style/taste
- Some practices and implementations are better than others
  - I'll aim to summarize the better ones
  - Offered as general guidelines rather
  - My opinion based on my experiences

Analysis Strategy

Q: What constitutes a complete analysis?

A: A suite of studies which together provide a coherent and thorough description of a particular set of data events

- Should cover all aspects necessary to understand and characterize these events
- Should be well documented via internal notes
- Should be subjected to peer review

Rules of Thumb

- Look before you leap
  - Plan your analysis strategy carefully
- Trust but verify
  - Always ask yourself, “Does this make sense?”
- A stitch in time saves nine
  - Sweat the (relevant) details, it will save time in the long run

Look before you leap

- While analyses are in general more iterative than linear, there are a few things that are quite helpful to do from the start
- Spend time thinking about the measurement with the goal of identifying those aspects which will drive the sensitivity
  - Analytic error propagation often a good start
  - Toy MC or MC truth level studies also very helpful in
    - What are the important physics effects?
    - What geometric and kinematic limitations do the detector and/or trigger introduce?
    - What are the most important instrumental effects?
  - The goal is to emerge with an understanding that helps prioritize which things need to be precisely understood and at what scale (1% or 10%?) and which don't

Look before you leap

- With the above information in hand, spend some time thinking about a plan of attack
  - What plots, figures, and tables will be important?
  - What data sets will you need?
  - What triggers do these data sets use?
  - What Monte Carlo (MC) samples will you need?

Trust but verify

- Unlikely you'll do everything for your analysis, but good to know where to find more detailed information if necessary
  - Dataset definitions
  - Trigger requirements and thresholds
  - Location and access to (raw-ish) data
  - Variable definitions in the ntuple
  - Location and access to source code and alignment and calibration details

Trust But Verify

- Probably inefficient to know all of that *a priori*, a more pragmatic approach is to learn those things as you need them. How do you know when you need them?
- Because at every step you're making plots and calculating ratios, efficiencies, etc. and asking yourself, "Does that make sense? Is that what I expect?"
  - e.g. d_0 vs ϕ, MET vs ϕ, muon eta, trigger track eta, p_T spectra, vs instaneous Lumi, vs #reconstructed vertices
  - e.g. cross-checks using intelligently chosen background control samples
- As first generation analyzers of a new experiment, this is particularly important

A stitch in time saves nine

- Be systematic/thorough/redundant in your approach… it will save you time in the long run
  - Follow your plan as best you can
  - When you spot a problem, take the time to understand it
  - Don't skip steps to get to the "answer"… it'll be inconclusive until you've demonstrated that all the inputs make sense

Concluding Remarks

- Today we
  - Reviewed the basic Phenomenology of the physics at a Hadron Collider
  - Discussed how that Phenomenology affects Detector Design and Analysis Challenges
  - Pontificated about how to design an Analysis Strategy that addresses those Challenges
- Over the next two lectures I'll discuss examples of these strategies in action using CDF analyses