CMS-Data analysis school (Egypt)

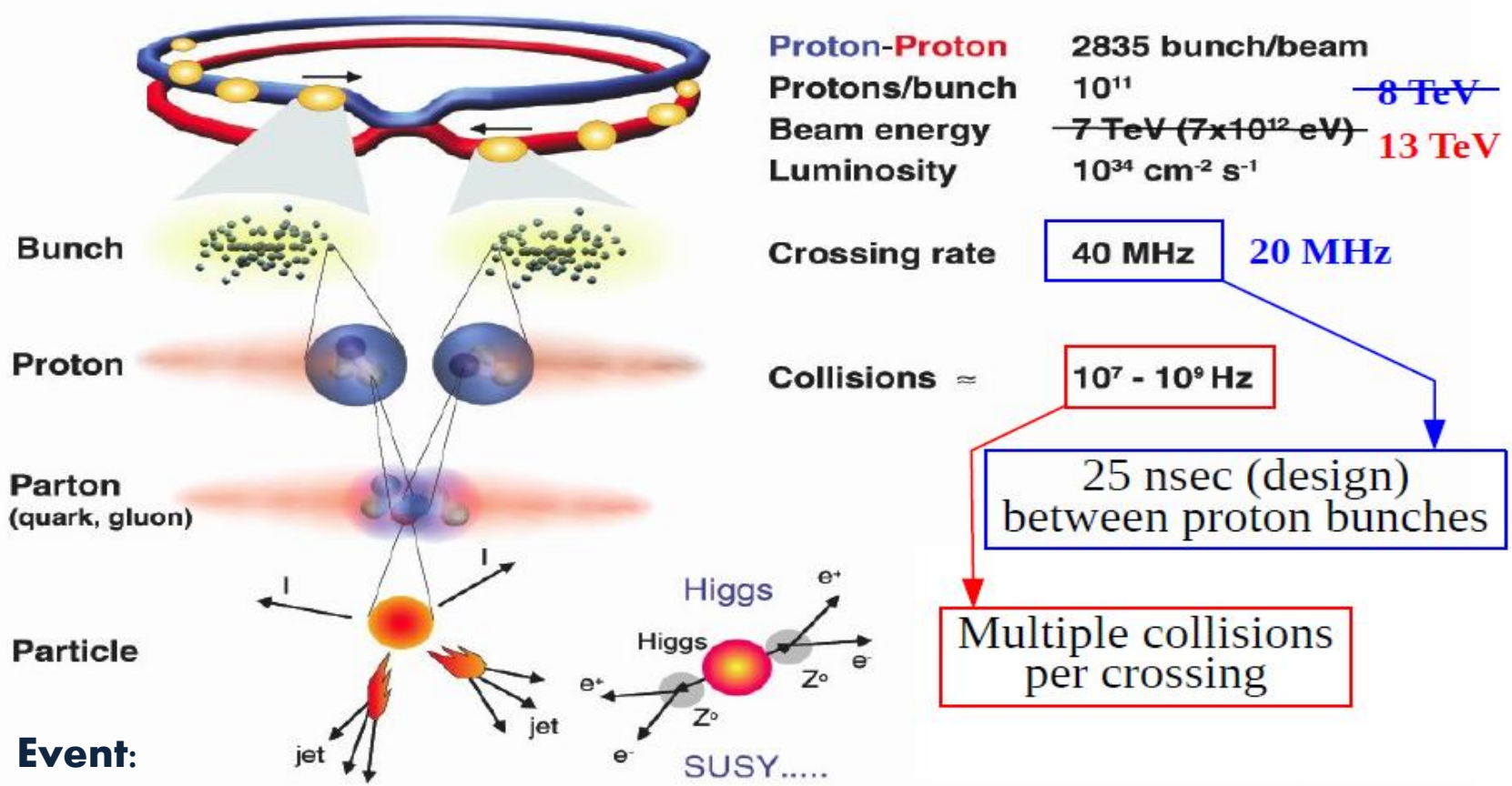
Using Trigger Technique in The analysis

Ahmed Qamesh

Centre for theoretical physics (BUE)

20/4/2015

How big is our Data?



Event:
the products of a given bunch-crossing or the products of a single interaction between the incident particles

Storage needed 40 TeraBytes/Sec

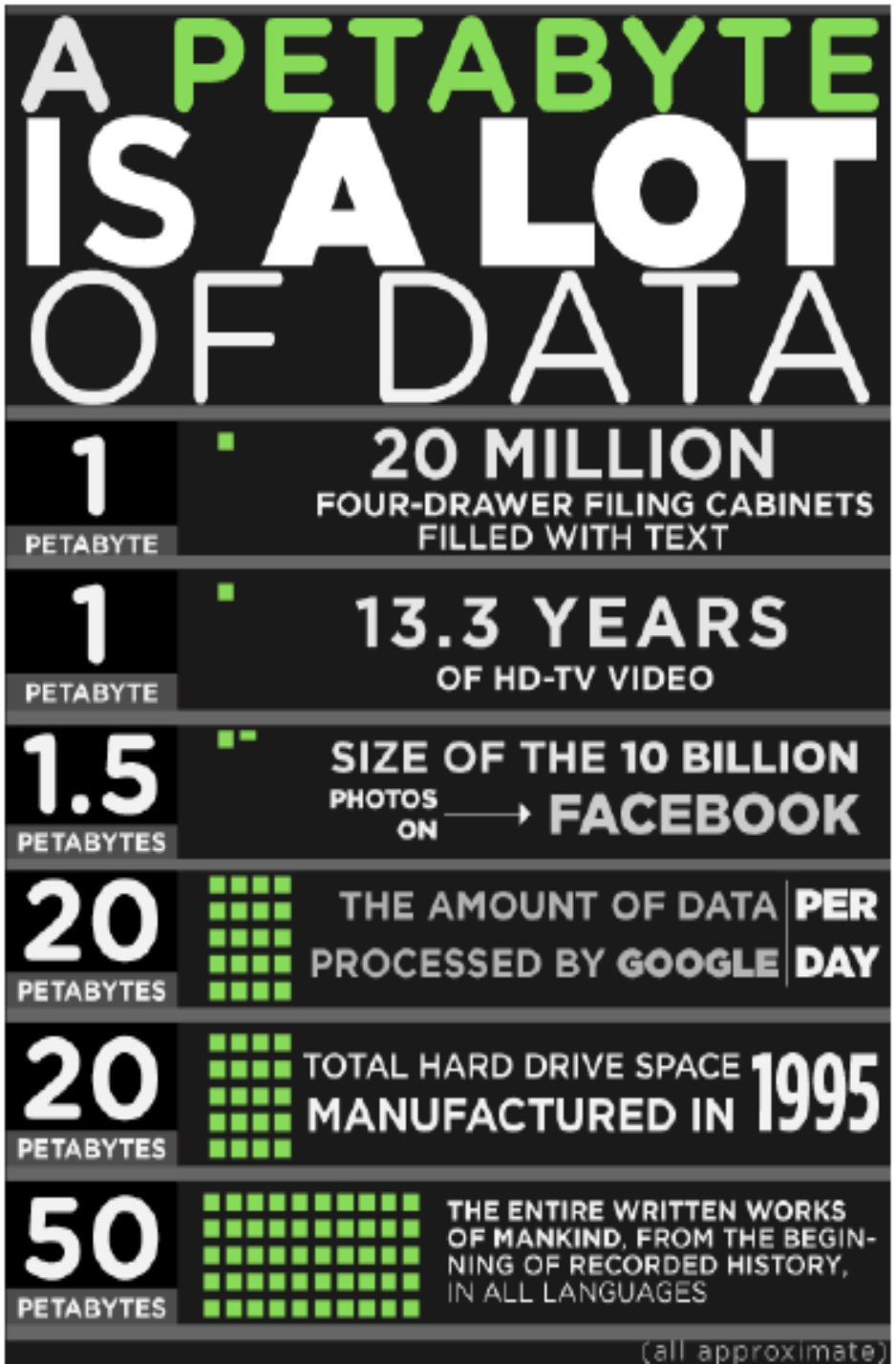
$$1.262276000\times10^{9}\text{TB}/\text{y}=1.262276000\times10^{4}\text{TB}/\text{y}$$

Not too Much?

Atlas and CMS can store 500 collisions per second!!
If correct , the Get ready to work
1-2 Higgs particles may(May not) be made at LHC

Big challenge!!!!

You need to pick only the interested one



But what is interesting?

But what is interested?

But what is interested?

But what is interested?

It depends on what you are looking for

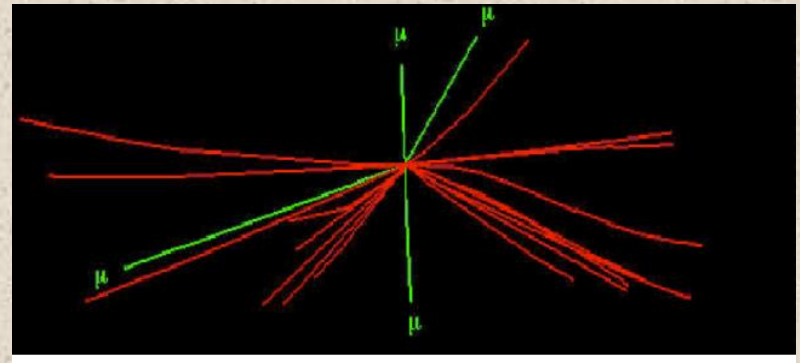
1. An electron or positron (anti-electron), even of low energy
2. A muon or anti-muon, even of low energy
3. A photon, even of low energy
4. A tau lepton or anti-lepton of moderate energy
5. Signs of invisible particles of moderate energy
6. Jets [manifestations of quarks, ant quarks and gluons] of very high energy
7. Many jets of moderate energy
8. Jets from bottom quarks of moderate energy
9. Multiples or combinations of the above

Is it Difficult?

A little bit…but not in *Cern* ☺

rare events one in 10,000,000,000,000 collisions [Exactly like looking for a single drop of water in a Jet d'Eau over 30 minutes]

Big Gap between what <u>you want</u>



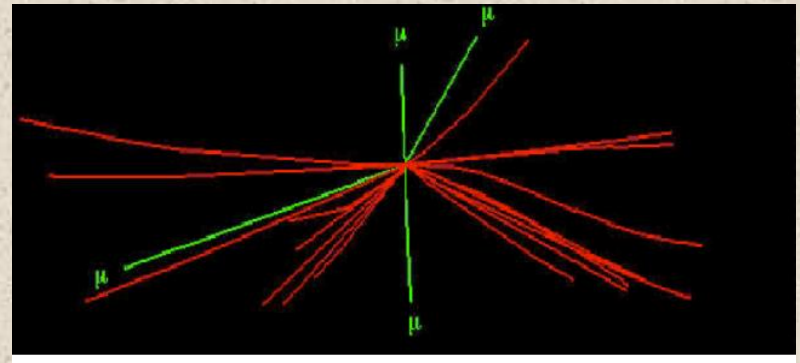
Higgs -> 4μ

Is it Difficult?

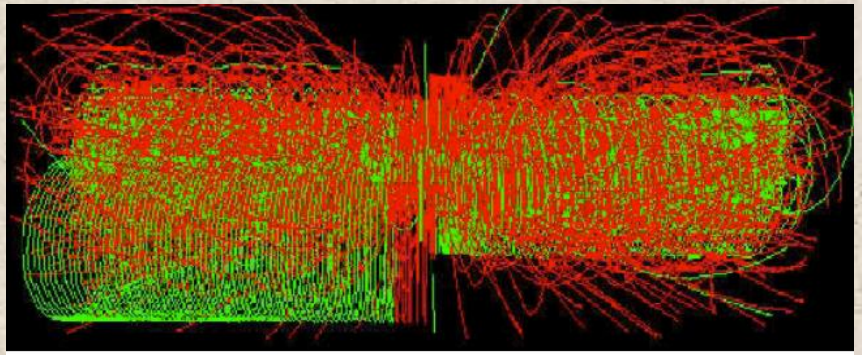
A little bit…but not in *Cern* ☺

rare events one in 10,000,000,000,000 collisions [Exactly like looking for a single drop of water in a Jet d'Eau over 30 minutes]

Big Gap between what <u>you want</u> and what <u>actually happened</u>



Higgs -> 4μ



99.999% to 99.9999% of all the data at the Large Hadron Collider is erased within a second of its being collected

Triggering

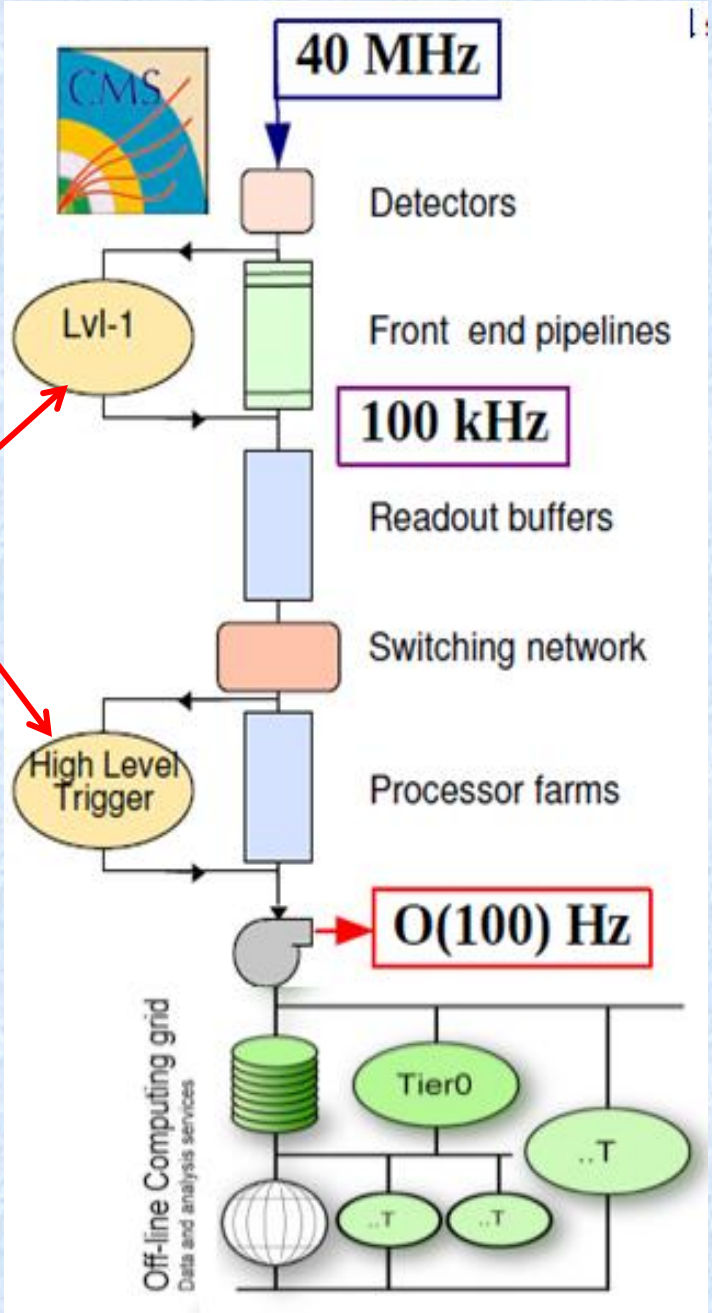
a database trigger is procedural code that is automatically executed in response to certain events on a particular table overview in a database

TERMINOLOGY

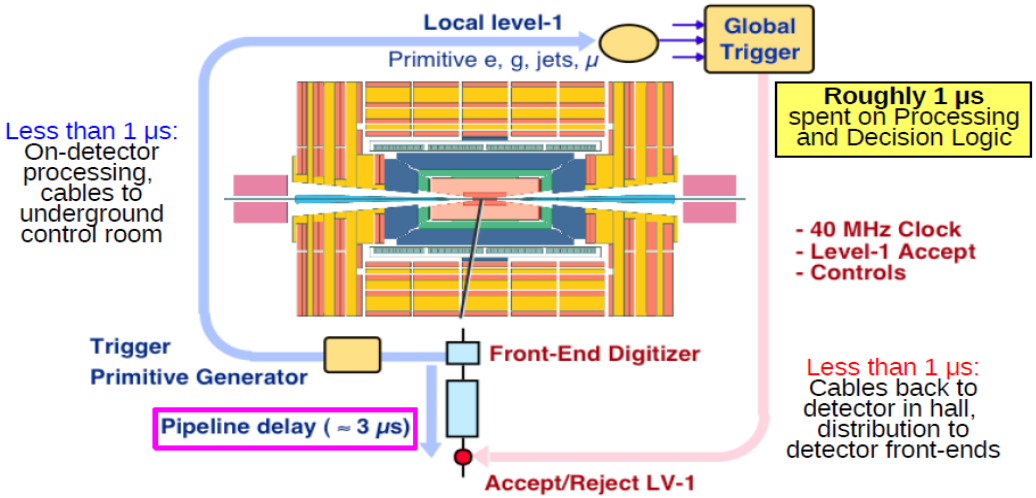
1. Data is collected online by each detector in cms
2. the trigger is a fast online filter that selects the useful events for offline analysis
3. Physicists analyze this data offline
4. Optimizing selection, estimating/modeling background, establishing limits, discovering New Physics, etc

The happens in two stages

1. Level 1 Trigger (400MHZ→100 KHZ)
2. High level Trigger (100HZ→100 HZ)



L1 Trigger



In the calorimeters the particles (e/photon, tau, jet, muon) despite energy in the detectors to produce a particle primitives

In the muon system Muons leaves a hits

1. The data will stay for a while (3.2μs) in a front-end pipelines till the trigger takes its discussion

Advantages of The pipelines

Process many events at once (without any **dead time**)

Store the data from the detectors for 3.2μsec

Makes a parallel processing of different inputs as much as possible

2. The data now will go through a **Trigger primitive generator**

(Produces a trigger primitive objects e, γ, μ, jets above a set of transverse energy and momentum for each event) Using a pure Electronic System (No soft ware involved) of integrated circuits include microprocessors, memory blocks

High level Trigger

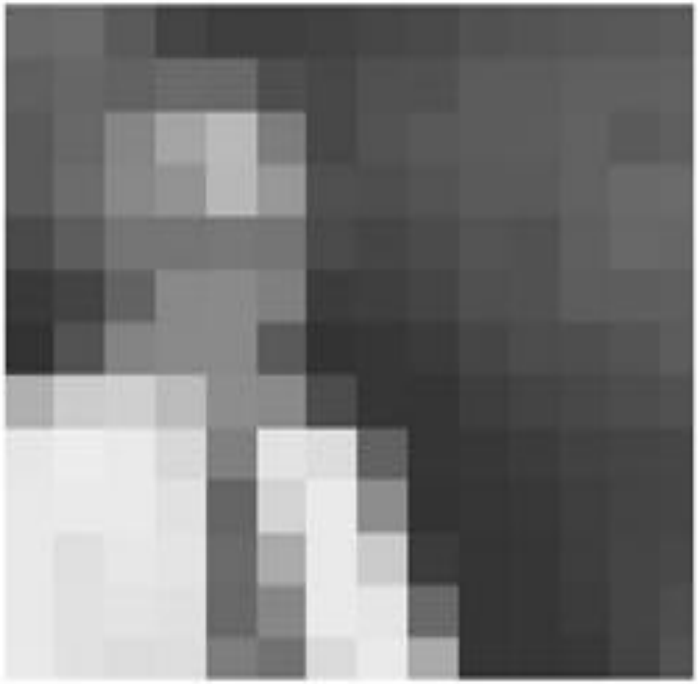
Very similar to the offline analysis (even with the same sofware)

Aim: calculate quantities with a better resolution, purity and efficiency by analyzing the detector data using software reconstruction and filter algorithms

It decide whether an event should be kept for an offline analysis

HLT algorithms Work only with Raw Data (Not RECO or AOD)

No Trigger



High level Trigger

Very similar to the offline analysis (even with the same sofware)

Aim: calculate quantities with a better resolution, purity and efficiency by analyzing the detector data using software reconstruction and filter algorithms

It decide whether an event should be kept for an offline analysis

HLT algorithms Work only with Raw Data (Not RECO or AOD)

L1 Trigger

High level Trigger

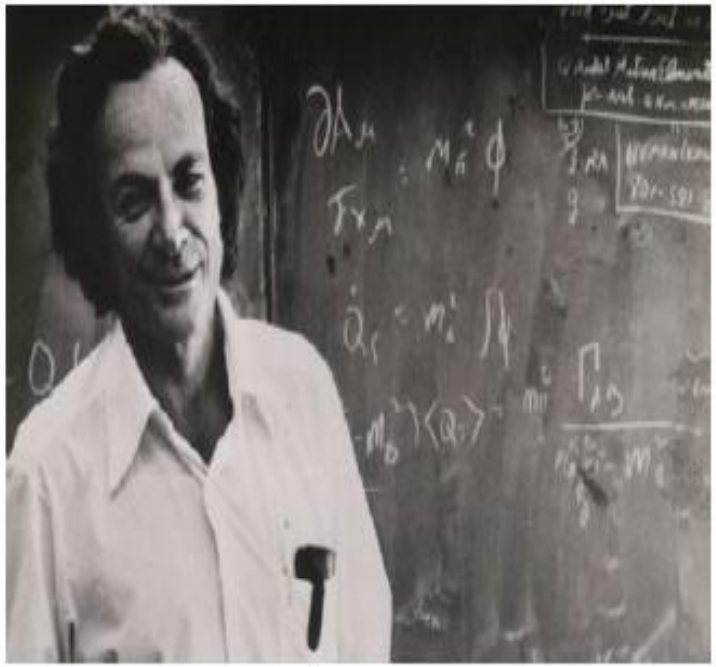
Very similar to the offline analysis (even with the same sofware)

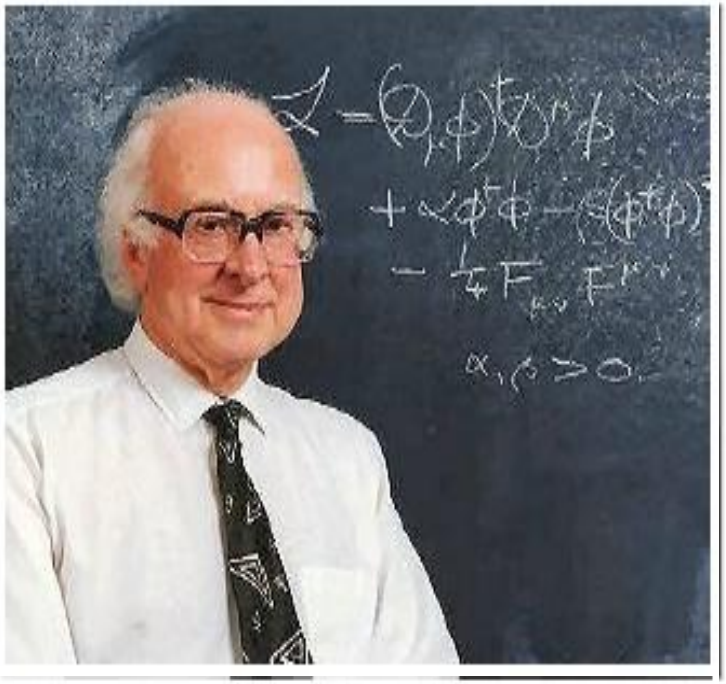
Aim: calculate quantities with a better resolution, purity and efficiency by analyzing the detector data using software reconstruction and filter algorithms

It decide whether an event should be kept for an offline analysis

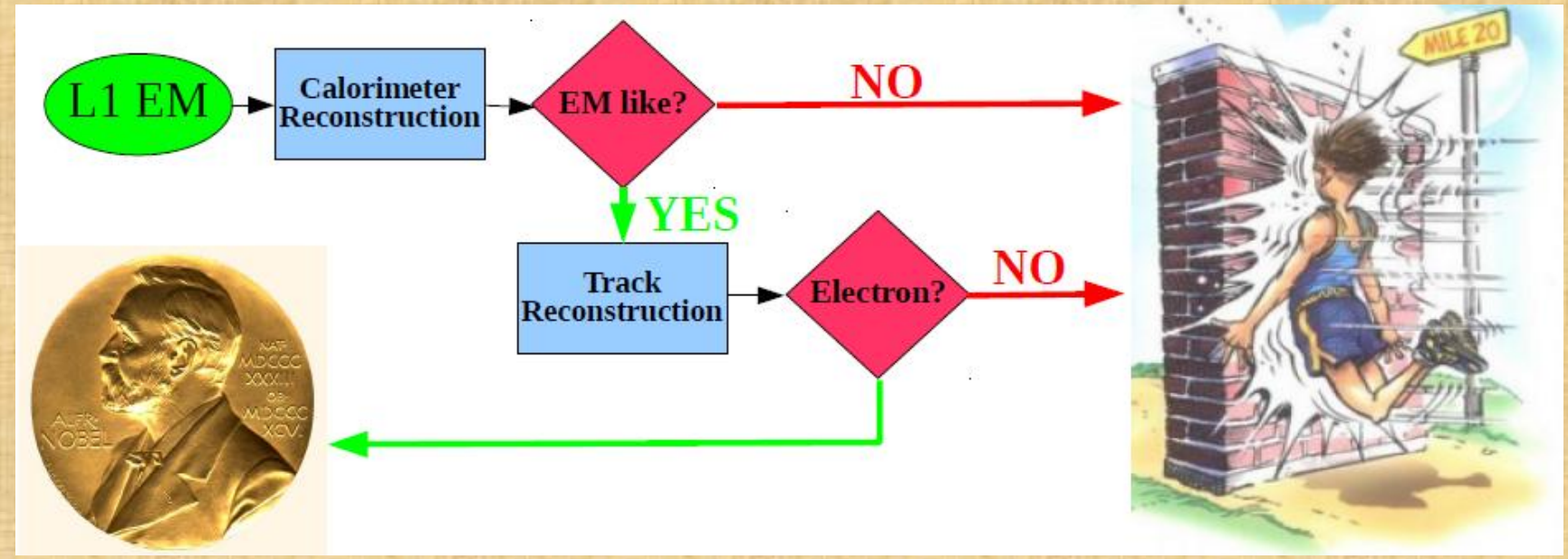
HLT algorithms Work only with Raw Data (Not RECO or AOD)

HL Trigger



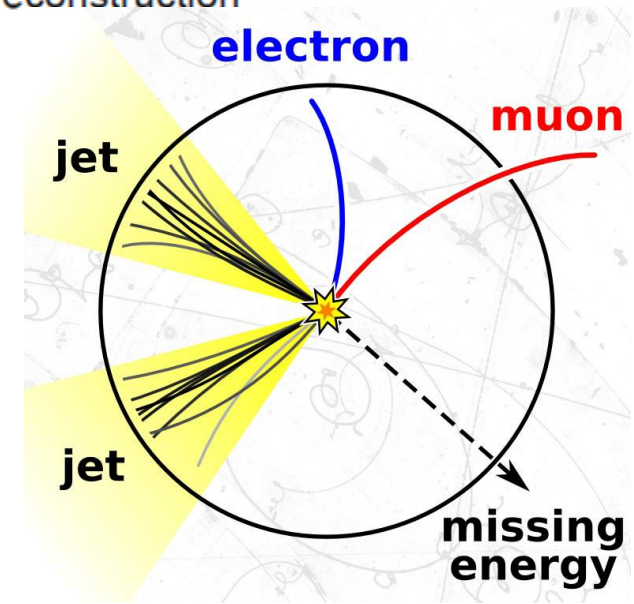


Structure of HLT: pure software algorithms (Hundreds) each designed with a specific physics signature ,Every algorithm includes a several Trigger paths
(the time of processing depends mainly on the trigger paths)
Electron candidates defined by combining information from pixel and silicon strip detectors (for High E_T)
Make sure that it is not a Hadron **(By checking HCAL)**
Muon candidates defined by combining information from tracker hits with the Muon system (for High P_T)



HLT objects

- muons
  - "L2" stand alone muons
  - "L3" global muons
  - tracker-based isolation
- photons
  - based on ECAL superclusters
  - calorimeter-based id and isolation, tracker-based isolation
- electrons
  - match ECAL superclusters, pixel tracks, and full tracking
  - calorimeter-based id and isolation, tracker-based id and isolation
- taus
  - particle flow reconstruction
- jets, MET, HT
  - calorimteric jets and MET
  - particle flow-based jets and MET
- b-tagging
  - jets, full tracking
  - secondary vertex reconstruction



At the end

- These datasets are sent to the CMS Tier2 centers
- Accessible via the GRID by everyone in the collaboration

A further step to the analysis

- Physicists start with an analysis Idea

  (*Ex: I want to look for the Z prime boson*)
- Figure out how to select the data
- (*Ex: I am looking for two high P_T leptons with opposite charge*)

  Do your analysis and Try to trigger your events
- **What If: Your interesting event already fails the trigger**

  Then: you need to build a trigger that has loose requirements (many tutorials about this using confDB)

What do I need to get an HLT path?

The trigger is coordinated by the Trigger Studies Group (TSG)

Their main task is to make sure that CMS has a working and efficient trigger

Then, simply your path has to be presented and approved in the TSG meetings (Thursdays at 14,00h)

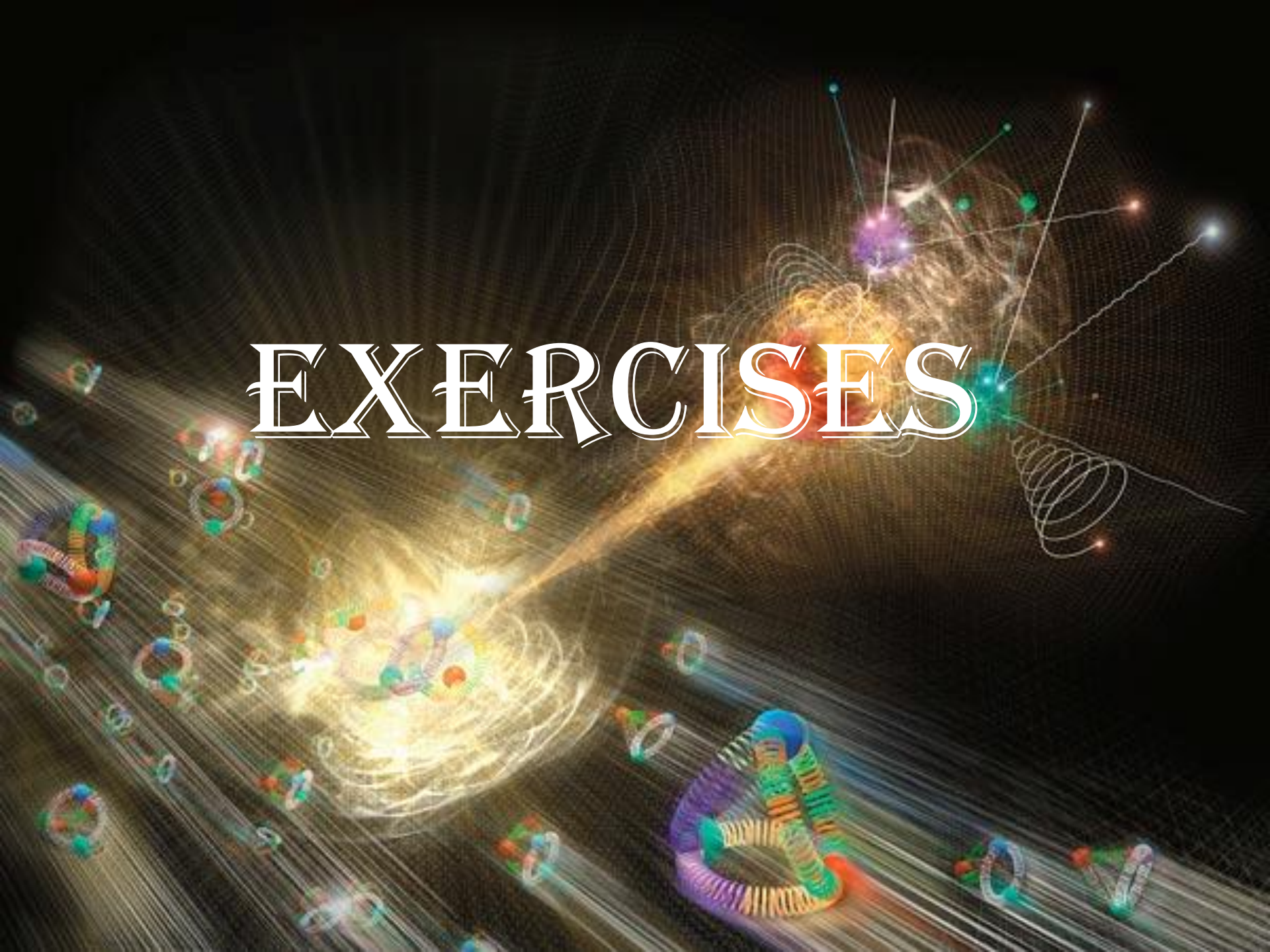
https://cms.web.cern.ch/org/trigger-coordination

- You should provide the following elements
- Physics motivation: explain your physics case clearly
- Implementation of your HLT path
- Measurement of the rate consumed by your HLT path
- Measurement of the timing of your HLT path
- The efficiency of your HLT path: a plan to measure it
- A proper DQM validation for your path

What happens if your trigger has a large rate (>100 Hz)?

Hopefully many physics analysis beside yours could use the same Trigger
(Ex: Standard model Z, W, top, SUSY, Exotic signatures all are using the same Trigger)

EXERCISES



Preliminaries

CMS Computing Concepts:

- **Operating Systems**: Linux
- **Programming Language**: C++
- **Analysis Packages**: Root
- A "release" is a set of software
- Releases indicated by the series of numbers such as 5_3_14.
- The first number in series indicates a major cycle
- The second number a major release with new features with respect to the proceeding release,
- The third number a release with some updates and bug fixes to the proceeding release.
- You can find out the existing releases installed on the local user interface by typing
- scramv1 list CMSSW
- MC data is not compatible between major release series
- So you can't analyze Monte Carlo created for a much earlier release with code from a later release.
- **7_3_X is the currently recommended Analysis Release for Phys14 exercise on MC.**
- **5_3_X is the currently recommended Analysis Release for 2012 Data/MC.**
- Any CMSSW release contain several HLT menus
- **Example**: proton-proton collisions (called GRun), lead-lead heavy-ion collision (called HIon), proton-lead collisions (called PIon)

Exercise (1)

CMSSW_5_2_7_onlpatch2_ONLINE /cdaq/physics/Run2012/8e33/v2.0/HLT/V2

(runs 206744 – 206745)

RUN	LUMI_NB_LIVE_DELIV	SEQUENCE	TRIGGER_MODE	L1_KEY	HLT_KEY	STARTTIME	STOPTIME	TRIGGERS	BFIELD	TIER0	COMPONENTS
206745	138.507645 140.085063	GLOBAL-RUN toppro	l1_hlt_collisions_2012/v140	TSC_20121102_002949_collisions_BASE	/cdaq/physics /Run2012 /8e33/v2.0 /HLT/V2	2012.11.06 00:09:11	2012.11.06 13:04:50	1744260262	3.801	1	CSC DAQ DCS DQM DT ECAL ES HCAL HFLUMI PIXEL RPC SCAL TRACKER TRG
206744	53.761434 55.781328	GLOBAL-RUN toppro	l1_hlt_collisions_2012/v140	TSC_20121102_002949_collisions_BASE	/cdaq/physics /Run2012 /8e33/v2.0 /HLT/V2	2012.11.05 20:58:58	2012.11.06 00:04:33	663305367	3.801	1	CSC DAQ DCS DQM DT ECAL ES HCAL HFLUMI PIXEL RPC SCAL TRACKER TRG

You have a physics menu called (**CMSSW_5_2_7_onlpatch2_ONLINE**)

This define the release created this trigger

Trigger Mode called (l1_hlt_collisions_2012/v140)

Defines the used paths from L1 and HLT for this Run

1. For L1_KEY TSC_20121102_002949_collisions_BASE)

Includes all the paths, the average trigger rates for L1 Trigger for this Run

---

L1Summary Algorithm Triggers									
Bit	Name	Pre-DT Counts	Pre-DT Rate, Hz	Pre-DT RMS Rate, Hz	Post-DT Counts	Post-DT Rate, Hz	Post-DT RMS Rate, Hz	InitialPrescale	FinalPrescale
0	L1_ZeroBias	71767915	1,541.71	1,541.71	71211778	1,530.53	1,530.75	9973	0
1	L1_AlwaysTrue	23875001	512.88	512.88	23738696	510.21	510.28	241	0
2	L1_BeamGas_Hf_BptxPlusPostQuiet	112992	2.43	2.75	112097	2.41	2.73	1000	0
3	L1_DoubleJet20	19548661489	419,941.75	470,445.44	818141243	17,584.01	18,505.19	1	0
4	L1_BeamGas_Hf_BptxMinusPostQuiet	34844	0.75	0.84	34618	0.74	0.83	5000	0
5		0	0.00	0.00	0	0.00	0.00	1	0

- For HLT_KEY (/cdaq/physics/Run2012/8e33/v2.0/HLT/V2)

Menu includes all the paths and streams for HLT Trigger for this Run

Sequence	Path	Stream	L1 Prerequisite
0	HLTriggerFirstPath (100164)	null	null
1	HLT_Activity_Ecal_SC7_v13 (100165)	A	"L1_ZeroBias"
		DQM	
		HLTMON	

The two runs are different in Time!!
Lots of changes happened in HLT and L1 Trigger
Ex: Larger run time, different modes

The data set: /DYJetsToLL_M-50_13TeV-madgraph-pythia8

Data Aggregation System (DAS): Home | Services | Keys | Bug report | Status | CLI | FAQ | Help

results format: list , 10 results/page, dbs instance prod/global , autocompletion disable Search Reset

dataset dataset=/DYJetsToLL_M-50_13TeV-madgraph-pythia8*/*/*

Show DAS keys description

mongoDB

Showing 1—10 records out of 31. <first | prev | next | last>

Add filter/aggregator function to the query: grep dataset.created_by + Clear

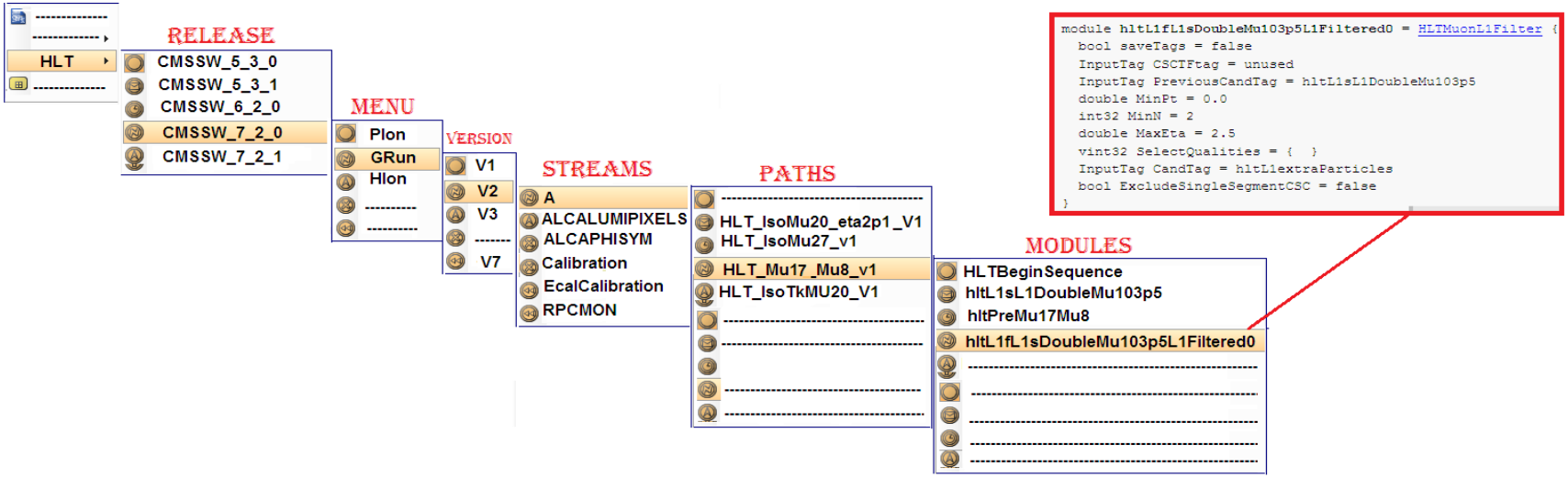
Dataset: /DYJetsToLL_M-50_13TeV-madgraph-pythia8/Fall13-POSTLS162_V1-v2/GEN-SIM
Creation time: 2014-01-27 22:36:49, Physics group: NoGroup, Status: VALID, Type: mc
Release, Blocks, Files, Runs, Configs, Parents, Children, Sites, Physics Groups, py , Subscribe to PhEDEx Sources: dbs3 show

Dataset: /DYJetsToLL_M-50_13TeV-madgraph-pythia8/Phys14DR-StreamALCACombined-PU20bx25_PHYS14_25_V1-v1/ALCARECO
Creation time: 2014-11-14 06:57:32, Physics group: NoGroup, Status: VALID, Type: mc
Release, Blocks, Files, Runs, Configs, Parents, Children, Sites, Physics Groups, py , Subscribe to PhEDEx Sources: dbs3 show

Dataset: /DYJetsToLL_M-50_13TeV-madgraph-pythia8/Spring14miniaod-141029_PU40bx50_PLS170_V6AN2-v1/MINIAODSIM
Creation time: 2014-11-09 08:30:13, Physics group: NoGroup, Status: VALID, Type: mc
Release, Blocks, Files, Runs, Configs, Parents, Children, Sites, Physics Groups, py , Subscribe to PhEDEx Sources: dbs3 show

Dataset: /DYJetsToLL_M-50_13TeV-madgraph-pythia8/Spring14miniaod-PU20bx25_POSTLS170_V5-v1/MINIAODSIM
Creation time: 2014-07-09 17:57:48, Physics group: NoGroup, Status: VALID, Type: mc
Release, Blocks, Files, Runs, Configs, Parents, Children, Sites, Physics Groups, py , Subscribe to PhEDEx Sources: dbs3 show

Exercise [2]



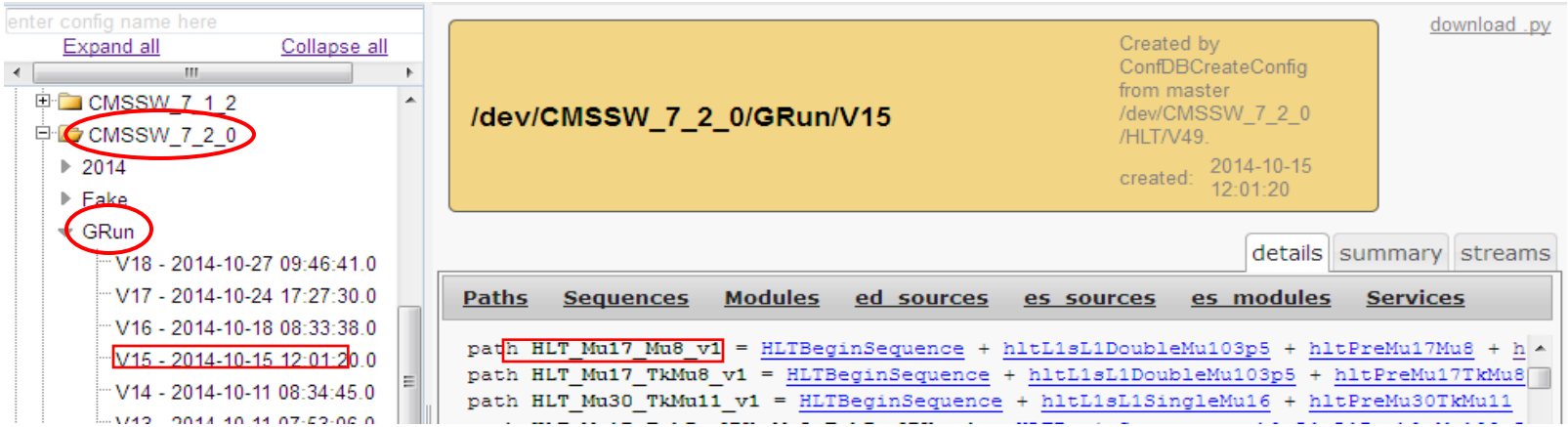
From HLT config browser: https://j2eeps.cern.ch/cms-confdb/browser/

You can get your trigger object (Muon for example)

In the menu GRun: /dev/CMSSW_7_2_0/GRun/V15

In the Menu (GRun): Related to p-p collision

HLT classify the Trigger tasks between streams according to (purposes, event content and rates)



- Each stream makes a specific task physics, calibrations,
- Our menu belongs to Stream A (Responsible for the physics of primary data set)
- In stream A : You can find lots of paths,
- The Name of the path (convention: should contain the major selection as pt and eta cut)

Exercise (3)

- In the trigger path **HLT_Mu17_Mu8**
- Any path consists of sequences and modules

  I. For the sequence (HLTBeginSequence, HLTEndSequence): Standard HLT sequence which define input and output of path, should **ALWAYS** be present in a path

  II. For the modules
- The modules in any path should be called in this order

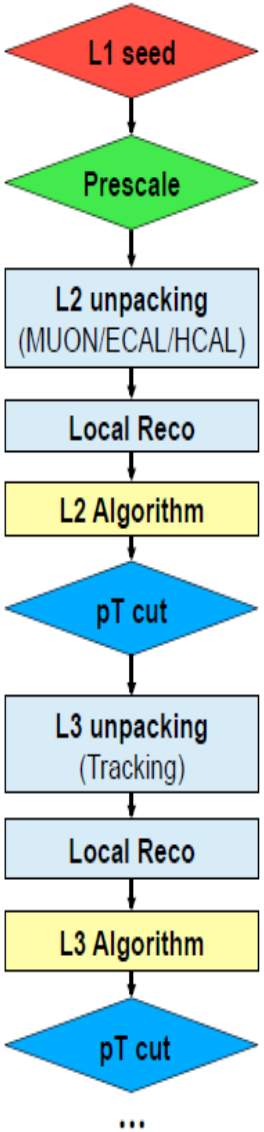
  1. Unpacking modules (raw to digi)

  2.Reconstruction of physics objects (electrons, muons, jets, MET, etc) [EDProducer]

  3.Filter modules (Include some cuts)

  4. All paths includes

  5. the L1 seed filter (which also provides the L1 seeds possibly needed later on in the path);




```
module hltL1sL1DoubleMu103p5 = HLTLevel1GTSeed {
  string L1SeedsLogicalExpression = "L1_DoubleMu_10_3p5"
  bool saveTags = true
```

L1 seed Name

Make sure that L1 object is saved in the event

2. the presale filter (for efficiency calculation or background studies)

```
module hltPreMu17Mu8 = HLTPrescaler {
  InputTag L1GtReadoutRecordTag = hltGtDigis
```

Exactly after L1 selection

Exercise (4)

Definitions of the parameters in the HLT Report heading

A numerical value corresponding to the path name. It can range from 0 to 499

Total number of events (100 events in this case)

The number of remaining events after L1 seed of this HLT path

Path name

The number remaining after the HLT

```
HLT-Report ---------- Event  Summary ------------
HLT-Report Summary for Job
HLT-Report Events total = 100 wanrun = 100 passed = 100 errors = 0

HLT-Report ---------- HLTrig Summary ------------
HLT-Report   HLT #   MaxRun    L1S    Pre    HLT   %L1sPre    Rate  RateNL  Errors Name
HLT-Report       0      100    100    100    100 100.00000   100.0     0.0       0 generation_step
HLT-Report       1      100    100    100    100 100.00000   100.0     0.0       0 digitization_step
HLT-Report       2      100    100    100    100 100.00000   100.0     0.0       0 L1simulation_step
HLT-Report       3      100    100    100    100 100.00000   100.0     0.0       0 digi2raw_step
HLT-Report       4      100    100    100      0   0.00000     0.0     1.8       0 HLTriggerFirstPath
HLT-Report       5      100     99     99     66  66.66666    66.0    71.1       0 HLT_Mu40_v1
HLT-Report       6      100     97     97     94  96.90722    94.0    96.4       0 HLT_IsoMu20_eta2p1_IterTrk02_v1
HLT-Report       7      100     97     97     92  94.84536    92.0    94.7       0 HLT_IsoTkMu20_eta2p1_IterTrk02_v1
HLT-Report       8      100     96     96     92  95.83334    92.0    94.7       0 HLT_IsoMu24_eta2p1_IterTrk02_v1
```

In line 0531

hlAccept_[i]: If the number accepted by HLT

An attempt to approximate the upper bound on the percentage in Rate given the limited available statistics prescale.

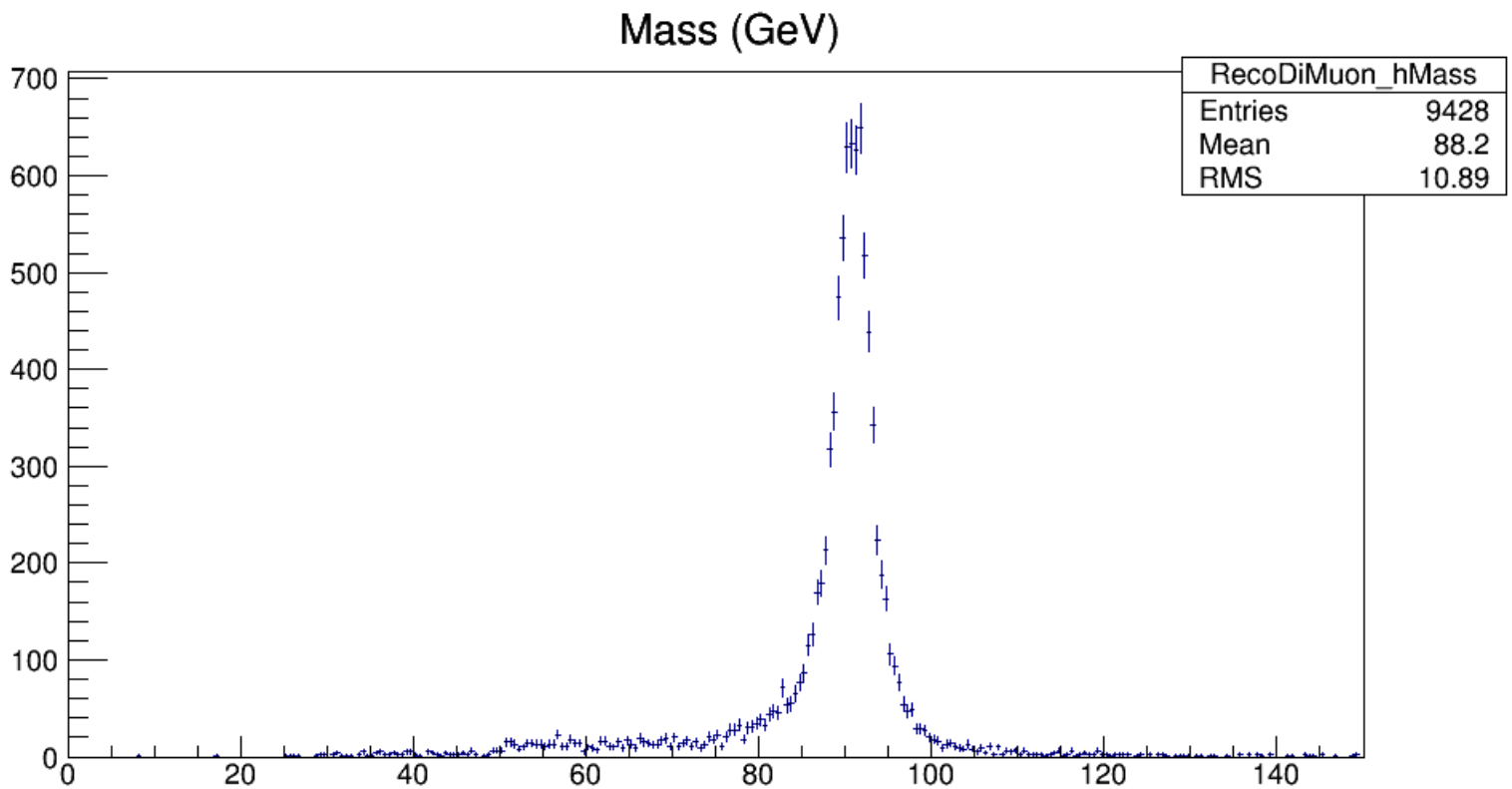
The Accepted HLT expressed as a percentage.

The percentage of L1 accepts that also pass the HLT $= HLT / L_1 \times 100$

The number accepted by the HLT for this path.

Exercise (5)

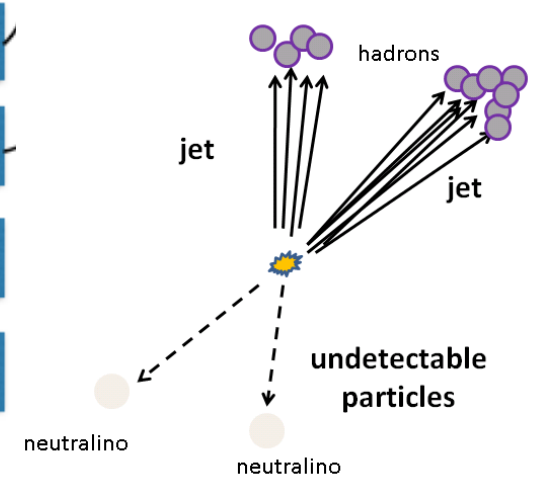
- **A TriggerResults product** containing all the usual decision bits
- **A TriggerEvent product**, summarizing the "L3" trigger collections and "L3" filter decisions
- **A TriggerEventWithRefs product**, summarizing the Refs to all trigger collections and all trigger filter decisions
- Tree for the skimed file Skim_Mu17Mu8_PHYS14.root



Exercise (6)

- HLT_DiCentralPFJet70_PFMET120_NoiseCleaned_v1 (17)
  - HLTBeginSequence (3)
  - hltL1sL1ETM60ORETM70 ← **L1 seed (MET)**
  - hltPreDiCentralPFJet70PFMET120NoiseCleaned
  - HLTRecoMETSequence (2)
  - hltMET70
  - HLTHBHENoiseCleanerSequence (2)
  - hltMetClean
  - hltMETClean60
  - HLTAK4CaloJetsSequence (2)
  - hltMetCleanUsingJetID
  - hltMETCleanUsingJetID60 ← **Calo MET filter**
  - hltDiCentralCaloJet50 ← **Calo Jets filter**
  - HLTAK4PFJetsSequence (3)
  - hltDiCentralPFJet70 ← **PF Jets filter**
  - hltPFMETProducer
  - hltPFMET120Filter ← **PF MET filter**
  - HLTEndSequence (1)

Running Particle Flow is expensive! Use calo filters to reduce rates first



read the information from the TTree:

```
// trigger decision
unsigned int passtrig = 0;  chain->SetBranchAddress("HLT_DiCentralPFJet70_PFMET120_NoiseCleaned_v1", &passtrig);
// auxiliary trigger decision
unsigned int passaux = 0;   chain->SetBranchAddress("HLT_IsoMu24_eta2p1_v1", &passaux);
// GenMET
float genmet = 0.0;         chain->SetBranchAddress("genmet", &genmet);
float genmetphi = 0.0;      chain->SetBranchAddress("genmetphi", &genmetphi);
// GenJets
int ngenjets = 0;           chain->SetBranchAddress("ngenjets", &ngenjets);
float genjetpt[maxsize];    chain->SetBranchAddress("genjetpt", &genjetpt);
float genjeteta[maxsize];   chain->SetBranchAddress("genjeteta", &genjeteta);
float genjetphi[maxsize];   chain->SetBranchAddress("genjetphi", &genjetphi);
```

declare denominator/numerator histograms:

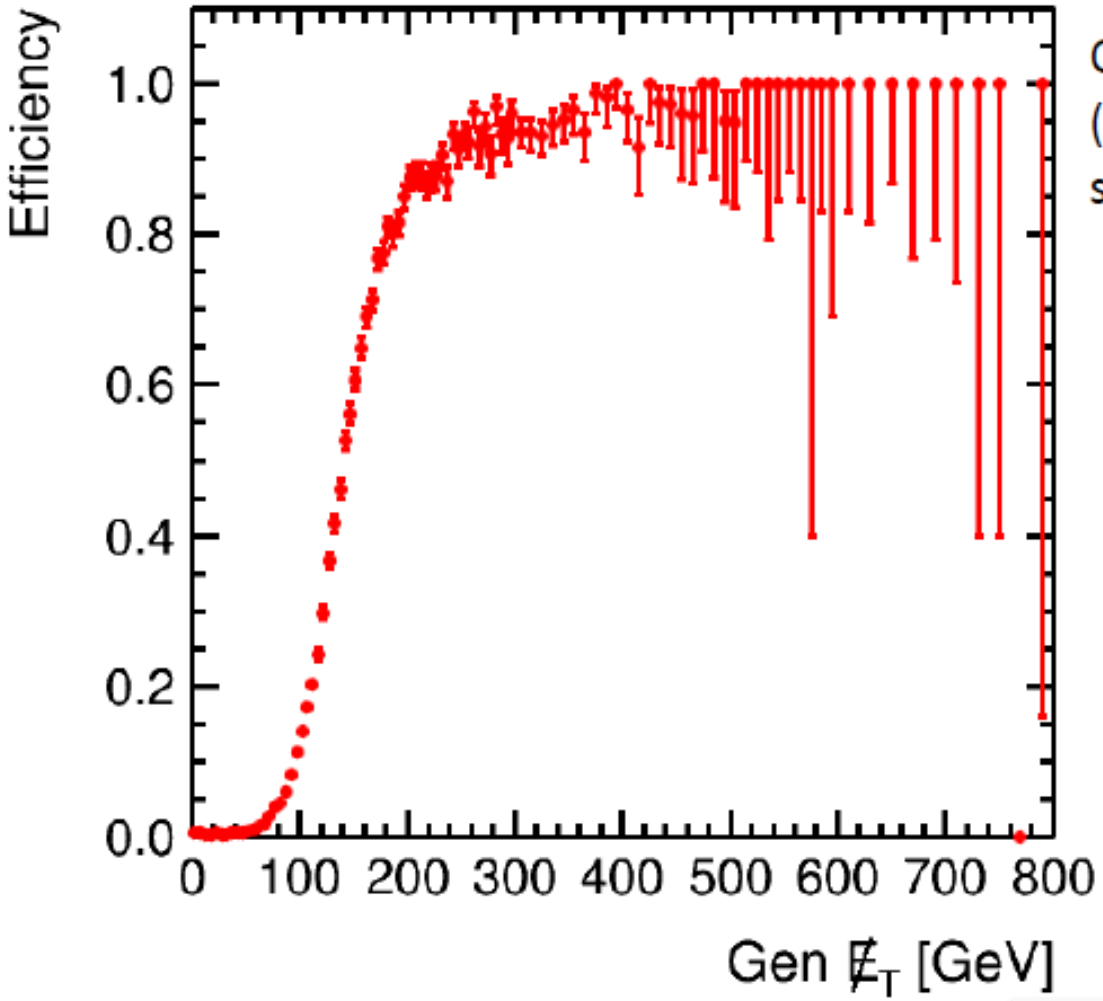
```
// histograms to store denominators and numerators
// vs MET
TH1F* hGenMet = new TH1F("hGenMet","",nbins,xbins);                    // denominator
TH1F* hGenMet_PassTrig = new TH1F("hGenMet_PassTrig","",nbins,xbins);  // numerator for full trigger
```

do some analysis, then fill histograms:

```
if(passaux && ncenjets30 > 4) { // require auxiliary trigger + offline selection
  // vs met
  hGenMet->Fill(genmet);
  if(passtrig) hGenMet_PassTrig->Fill(genmet);
  [...]
}
```

compute efficiencies (the **makeEffGraph** function divides the numerator and denominator histograms to produce the turn-on graph):

```
// make efficiency graphs
TGraphAsymmErrors *effvsGenMet = makeEffGraph(hGenMet_PassTrig, hGenMet);
```



Offline selection (baseline for hadronic stop analysis):

>= 5 jets with pT > 30 GeV (|eta| < 2.4)

eff_trig_vs_genmet.png


Thank You