Trigger & Analysis

Avi Yagil
UCSD

2nd Hadron Collider Physics Summer School
June 2007

Table of Contents

Introduction
or
Its all about the Trigger

For more info, on trigger architecture, design choices etc., see recent talks for example:

- N. Ellis: Triggering in the LHC environement
- P. Sphicas: Trigger @LHC

Rates…

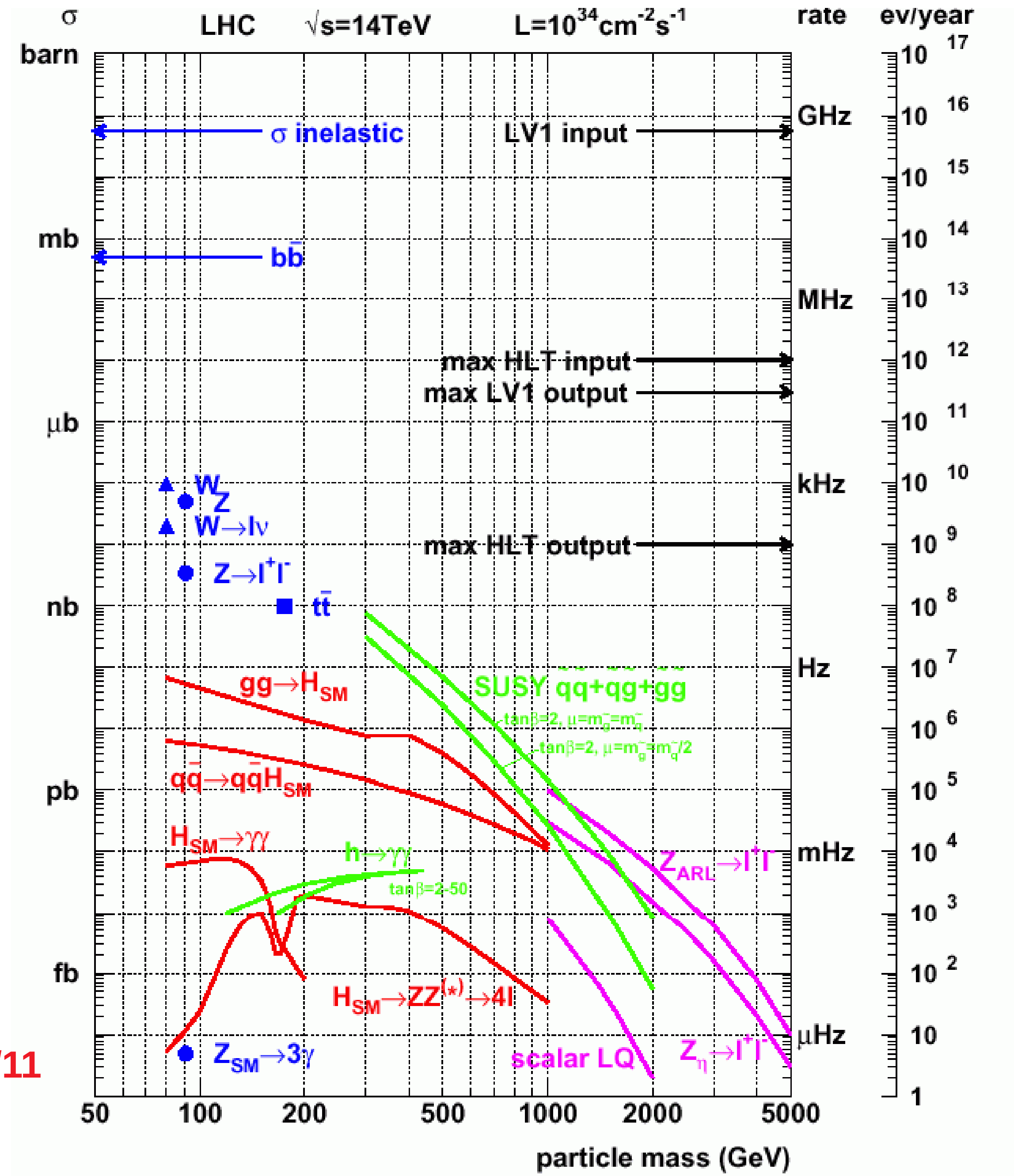
Cross Section: σ(pp) = 70 mb
Interaction Rate, R = $7x10^8$Hz
Bunch Spacing: Δt= 25 ns

Physics Cross Sections:

- Inelastic: 10^9Hz
- W→lν : 10^2Hz
- t t production: 10 Hz
- Higgs (100 GeV/c2): 0.1 Hz
- Higgs (600 GeV/c2): 10^{-2}Hz

- 250 GeV E_T Jets - 1kHz

Rejection needed - $10^{10/11}$



Time to think?
Bunch Spacing…

QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.

Less time to think.
MUCH more to think about…

Physics Content of collision-data

- In LEP or a B-factory, ~every event is "important".
- Hadron collider environment is "Physics poor", or "dirty", or "challenging",... :
  1. Must have a very sophisticated trigger
  2. Dangers of very complex, irrevocable online decision process

==>Major trigger challenge!

To make matters worse:

- There is not much time to "think" and "decide" between crossings (also detector size)
- Physics rate is very low with respect to collisions rate

Trigger - The big picture

- The role of the Level-1 trigger is to take an LHC experiment from the 25ns to the 10-25μs timescale.
  - Custom hardware, big switches, Gb/s rates
  - Simple, coarse fast algorithms
- Experiments differ on the implementation of the next level of filtering:
  - Commercial hardware, large networks, Gb/s rates..
    - Either: Multi level filtering (Level-2, 3…)
    - Or: A large software-based High Level Trigger
- Would like to make the High Level Trigger filtering as similar to the offline analysis software as possible.
  - Very large PC farms

Monitoring and error detection are crucial and non-trivial!

P. Sphicas, Trigger @LHC talk (06 SSI)

QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.

What do we trigger on?
Basics

You saw already that:

- Hadron colliders produce mostly low momentum hadrons
- Most interesting Physics has signatures involving large transverse energy (E_T) particles.

==> Require high E_T on reconstructed particles

- Basic objects used for trigger:
  - Electrons, Photons, Muons, Jets, Missing Et
- The more complexity one adds to a signature - the lower the rate, and hence the required threshold(s)
- The more complex the signature - the harder it is to understand (measure efficiency, monitor performance)

Efficiency and Dead-time

- Goal of trigger is to maximize collection of data for physics process of interest:
  - Aim for high efficiency !
- For each process, look for:

  ε trigger= Ngood(accepted)/Ngood(Produced)

  - And watch the dead-time !
- Trigger Dead-time:
  - Incoming rate is higher than processing rate

    ==> valid interactions are rejected due to system busy
- Buffering incoming data could reduce dead-time
- But dead-time always incurred if:
  - <incoming rate> larger than 1/<processing time> !

What do we trigger on?
Some numbers

- Objects used for trigger:
  - Electrons, Photons, Muons, Jets, Missing Et
- A few approximate thresholds and associated rates:
  - Single muon with P_T>20 GeV - 10kHz
    - Double muon P_T>6 GeV - 1kHz
  - Single em cluster with E_T>30 GeV - 10-20 kHz
    - Double em cluster with E_T>20 GeV - 5 kHz
  - Single jet with E_T>300 GeV - 0.2-0.4 kHz
  - Missing Transverse Energy, is another story…

Trigger Nomenclature
or
some buzz-words and concepts

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

What is a Trigger Table?

Trigger table is how our trigger menu is called.

- It is a list of selection criteria
- Each item on the menu:
  - Is called Trigger Path
  - Has a few "courses": L1, (L2) and HLT recipes:
    - Set of cuts, parameters, instructions specific for each level.
- An event is stored if one or more trigger paths criteria are met.
- Each time data taking starts (a new run), the whole content is communicated to the system (run control)
- For bookkeeping, all menus and recipes are stored in a specially designed database.
  - Understanding of the data (a.k.a. analysis) depend on it!

Trigger Path

- The "name" of the sequence of requirements that results in an event being recorded:
  - L1 accept
  - L2 accept (if present)
  - HLT confirmation of:
    - Object type(s)
    - threshold(s)
    - ID cut(s)

    Example:

    emu_20_20_iso_met_50

    hipt_emu_met

185 such paths in CDF

- There has to be a bit for each one.
  - In a "global" trigger summary word(s).
- Defines how events are classified (immutably!)
  - Implication: complete exploration of all L1 accepts

HLT Trigger Table - Example (CMS)

Trigger	Threshold (GeV) for $L=2x10^{33}\ cm^{-2}s^{-1}$	Prescale	Rate [Hz]
Single Electron	26		23.5 +- 6.7
Double Electron	12, 12		1.0 +- 0.1
Relaxed Double Electron	19, 19		1.3 +- 0.1
Single Photon	80		3.1 +- 0.2
Double Photon	30,20		1.6+-0.7
Relaxed Double Photon	30, 20		1.2+-0.6
Single Photon Prescaled	23	400	0.3+-0.02
Double Photon Prescaled	12, 12	20	2.5+-1.4
Relaxed Double Photon Prescaled	19, 19	20	0.1+-0.03
Single Muon	19		25.8+-0.8
Relaxed Single Muon	37		11.9+-0.5
Double Muon	7, 7		4.8+-0.4
Relaxed Double Muon	10, 10		8.6+-0.6
Double Pixel TauJet	100 SingleTau@L1 or 66 DoubleTau@L1		4.1+-1.1
Double Tracker TauJet	100 SingleTau@L1 or 66 DoubleTau@L1		6.0+-1.1
Electron-TauJet	16, 52		~0
Muon-TauJet	15, 40		0.1+-0.06
TauJet-MET	93, 65		0.5 +- 0.1
Single jet	400		4.8 +- 0.02
Single jet Prescale 1	250	10	5.2+-0.02
Single jet Prescale 2	120	1000	1.6+-0.008
Single jet Prescale 3	60	100000	0.4+-0.002
Dijet	350		3. +- 0.02
Trijet	195		1.1 +- 0.01
Fourjet	80		8.8 +- 0.2
Acoplanar Dijet	200, DeltaPhi(dijets)<2.1		0.2+-0.008
Single jet - MET acoplanar	100, 80, DeltaPhi(jet,MET)<2.1		0.1+-0.02
Single jet - MET	180, 80		3.2+-0.07
Dijet - MET	155, 80		1.6+-0.03
Trijet - MET	85, 80		0.9+-0.07
Fourjet - MET	35, 80		1.7+-0.2
MET	91		2.5+-0.2
H_T - MET	350, 80		5.6+-0.2
H_T - Single Electron	350, 20		0.4+-0.1
B-jets (leading jet)	350, 150, 55 (1,3,4-jet event cuts)		10.2 +- 0.3
B-jets (second jet)	350, 150, 55 (1,3,4-jet event cuts)		8.5 +- 0.3
TOTAL			**129.8 +- 7.3**

Electron, photon

muon

~60%

tau

jets

jets,met, HT

Pre-requisites, Volunteers

- Pre-Requisite:
  - Only muons that have a L1 accept are pursued in the HLT.
  - Moreover, only that region may be even looked-at (reconstructed).

- Volunteer:
  - A muon "found" in the HLT, without a corresponding L1 accept
    - Possible Convention: such cannot be the cause of a trigger decision (CDF/CMS)
  - Cannot happen if only "seeded" (on L1 muon track) reconstruction is pursued in HLT
  - Can happen if global reconstruction is performed.

  - Very useful in understanding trigger efficiencies (more later).

Trigger Efficiency - How to?

- How does one measure trigger efficiency?
  - Especially L1 trigger is hard. Why?
  - How does one make a plot like this:
  - What is the main challenge?
  - The solutions:
    - Special backup triggers
    - Usage of volunteers
    - MUST be thought out in advance!

QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.

Trigger Cross sections

- For any process:
  rate $R = L\sigma$
  (L = instantaneous luminosity, σ= cross section.)
- For a physics process, σ is independent of L.
- For trigger cross sections, we observe:
  $$\sigma = A/L + B + CL + DL^2$$
  - A,B,C,D are constants depending upon trigger.
  - High purity triggers typically have C~D~0.
  - Two effects cause extra powers of L:
    - Overlapping objects from different interactions.
    - Fakes that are luminosity dependent.

- Rates: $R=L\sigma = A + BL + CL^2 + DL^3$

“Backup” triggers

- Triggers are not only used to look for that special signature (signal) one is interested in
- They are also used for calibration/efficiencies/background studies
- Term backup is misleading
- For example, for top analyses, need to:
  - Measure L1/L2/L3 signal trigger efficiencies
  - Develop and tune soft lepton taggers
  - Calibrate b-tagging efficiency
  - Calibrate jet energy scale
  - …

Pre-scales

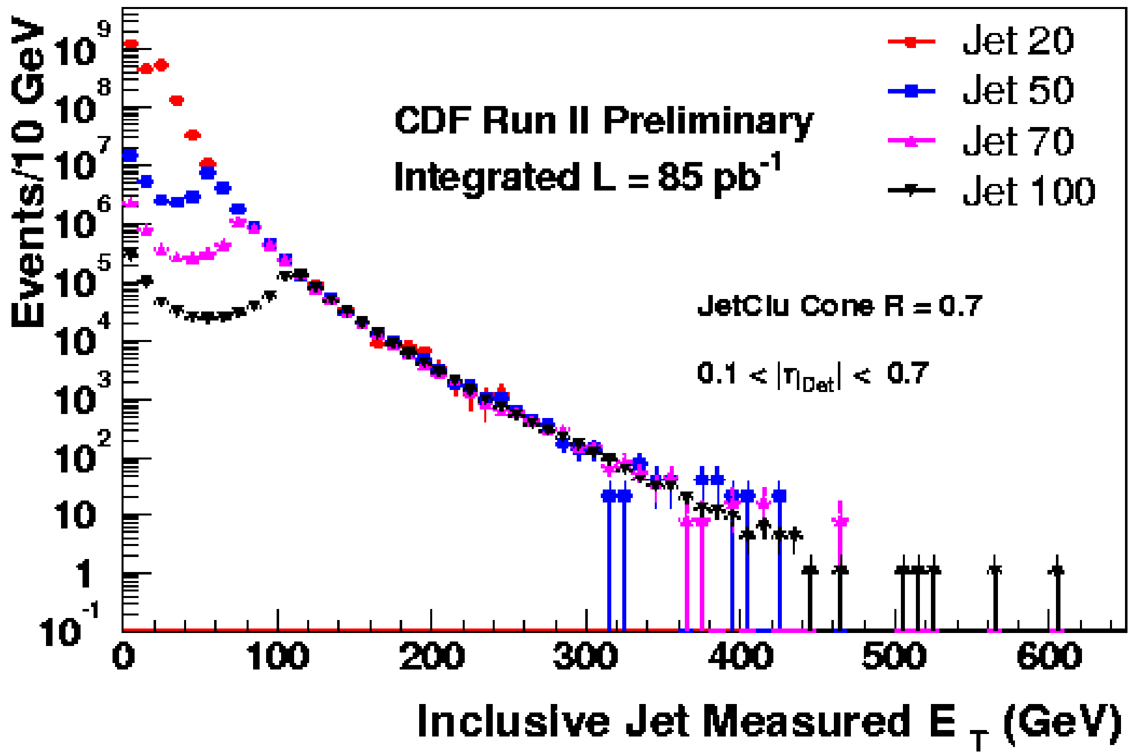
- Inclusive jet triggers - rate too high to take them all

Sample

- Lower threshold -- higher pre-scale
- Lower threshold at lower trigger level
  - "sharp" cut at HLT
- "assemble" spectrum

- Fancy:

  dynamic prescales



“Computing Model” Basics

or

Data Logistics

or

Where the @#$% is my sample?!

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

Triggers & Primary Data Sets

- It is easy to foresee about 50 primary datasets to be identified during the HLT processing, each following a strict trigger path (L1 and the appropriate HLT confirmation of it).
  - How many will be needed?
  - Serious Physics Coordination issue!

- As a Data Management artifact, these 50 primary datasets may be grouped into about 10 streams constructed to be roughly of similar sizes and contain more or less related primary datasets.

- These streams have:
  - no (physics) significance outside of the Online-to-Offline data management context.
  - May have significant logistical importance in prioritizing processing and access to relevant info.

Streams (online, intermediate)

- One can imagine the following example of streams definition:
  - Express
  - Jet data stream
  - electron/photon
  - Muon, tau/MET
  - calibration
  - min bias
  - …
- For example the ele/photon stream can contain the primary datasets for inclusive electron as well as the di-electron triggers of various thresholds (some may be pre-scaled) and the closely related photon triggers.

- These streams should also contain back-up triggers to assist in understanding the efficiency and rejection of the main ones (they will be usually looser and/or pre-scaled).

Overlaps

- A frequently asked question is:
  - What is the overlap between Primary Data Sets?
- Recall that the overlap between data sets is a result of the overlap between trigger paths.

How big is it? How big should it be?

- Up front overlap - express stream.
  - An experiment decides how big it is (~10%)
  - Completely controlled
- Additional duplication occurs due to Physics but is very small in comparison.
    - An e-mu-met event will (hopefully) go to three data sets and we want them to! (trig ε)
    - But (sadly) there are very few of these…
  - Can have huge overlaps - need good design of trigger table
    - One can have as big an overlap as one can afford.

The 4 experiments - Data Flow Estimates

QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.

50 days (low energy) running in 2007 - **canceled**
10^7 seconds/year pp from 2008 on ==> ~10^9 events/experiment
10^6 seconds/year heavy ion

Hierarchical Computing Model

- Tier-0 at CERN
  - Record and store RAW data
  - Distribute second copy to Tier-1s
  - Determine calibration & alignment required for prompt reconstruction
  - First-pass reconstruction
  - Distribute reconstruction output to Tier-1s
- Tier-1 centers (11 defined)
  - Manage permanent storage - RAW, RECO (custodial fraction)
  - Compute (more) precise calibration alignment
  - Reprocessing, bulk analysis (skimming…)
- Tier-2 centers (>~ 100 identified)
  - Monte Carlo event simulation
  - End-user analysis
- Tier-3 centers
  - Facilities at universities and laboratories
  - Access to data and processing in Tier-2s, Tier-1s

LCG web site

LHC Pledged Tier1 Centers

Centre	*Experiments served with priority*			
	ALICE	*ATLAS*	*CMS*	*LHCb*
Canada, TRIUMF		X		
France, CC-IN2P3	X	X	X	X
Germany, FZK-GridKA	X	X	X	X
Italy, CNAF	X	X	X	X
Netherlands LHC/Tier1	X	X		X
Nordic Data Grid Facility (NDGF)	X	X	X	
Spain, PIC		X	X	X
Taipei, ASGC		X	X	
UK, RAL	X	X	X	X
USA, BNL		X		
USA, FNAL			X	

From Oli Gutsche

QuickTime™ and a
TIFF (Uncompressed) decompressor
are needed to see this picture.

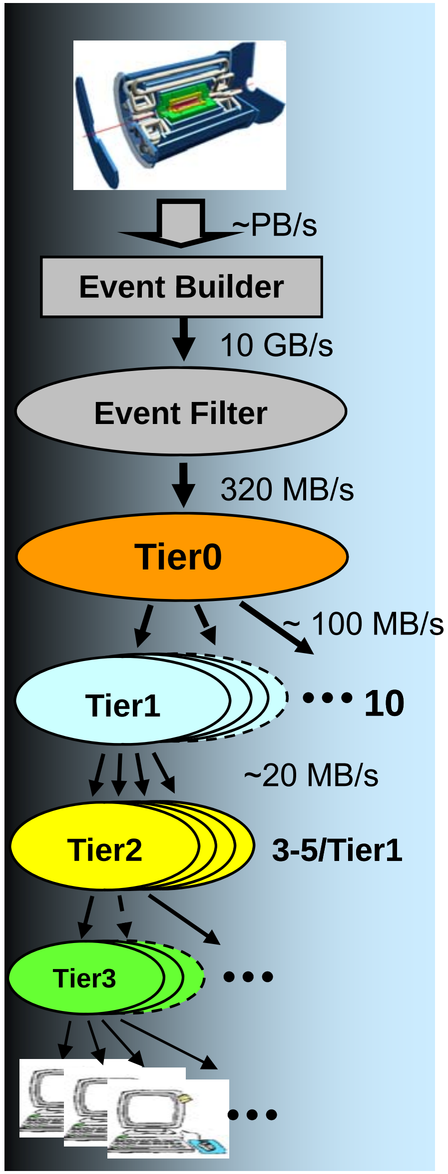
From Dario Barberis

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" (POOL/ROOT) 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.

Computing Resources

LCG web site

Summary of Regional Centre Capacities
25/04/2007

Tier-1 Planning for 2008		ALICE	ATLAS	CMS	LHCb	SUM 2008
CPU - MSI2K	Offered	6.3	21.6	11.3	3.7	42.9
	TDR Requirements	10.2	18.1	12.4	1.8	42.5
	Balance	-38%	19%	-9%	108%	1%
Disk - PBytes	Offered	2.6	12.1	5.3	2.0	22
	TDR Requirements	5.2	9.9	5.6	1.0	21.7
	Balance	-50%	22%	-5%	91%	1%
Tape - PBytes	Offered	3.3	8.2	9.2	1.5	22.2
	TDR Requirements	7.0	7.7	13.1	0.9	28.7
	Balance	-54%	7%	-30%	74%	-23%

Includes current planning for all Tier-1 centres

Tier-2 Planning for 2008		ALICE	ATLAS	CMS	LHCb	SUM 2008
CPU - MSI2K	Offered	5.7	18.4	17.1	3.6	44.8
	TDR Requirements	9.6	17.5	15.2	4.6	46.9
	Balance	-41%	5%	12%	-21%	-4%
Disk - PBytes	Offered	1.3	5.8	4.3	0.2	11.6
	TDR Requirements	2.5	7.7	4.2	n/a	14.4
	Balance	-46%	-25%	1%	n/a	-19%
# Tier-2 federations - included(expected)		17 (19)	31 (34)	30 (34)	11 (12)	52 (57)

Next time: Table of Contents

- Introduction
  - Rates & cross sections
  - Beam Crossings
  - What do we trigger on?
  - Trigger Table (example)
- Some trigger jargon
  - Trigger paths
  - Prerequisites
  - Volunteers
  - Efficiency/Dead-time
  - "backup" triggers
- Computing Model Basics
  - Hierarchy: Multi-Tiers
  - Data Tiers: Event Data Model
  - Data flow
  - Triggers & Primary Data Sets
  - Online streams
  - Overlaps

- "Analysis Model" Basics
  - Size & Speed
  - Data formats
- Starting an Analysis
  - Physics signature
  - What is my trigger?
- Evolution of Analysis
  - Detector Commissioning
  - Low level reconstruction
  - High Level Objects
  - Peak hunting
  - Counting experiments
  - Cross-sections…