Introduction to Online System

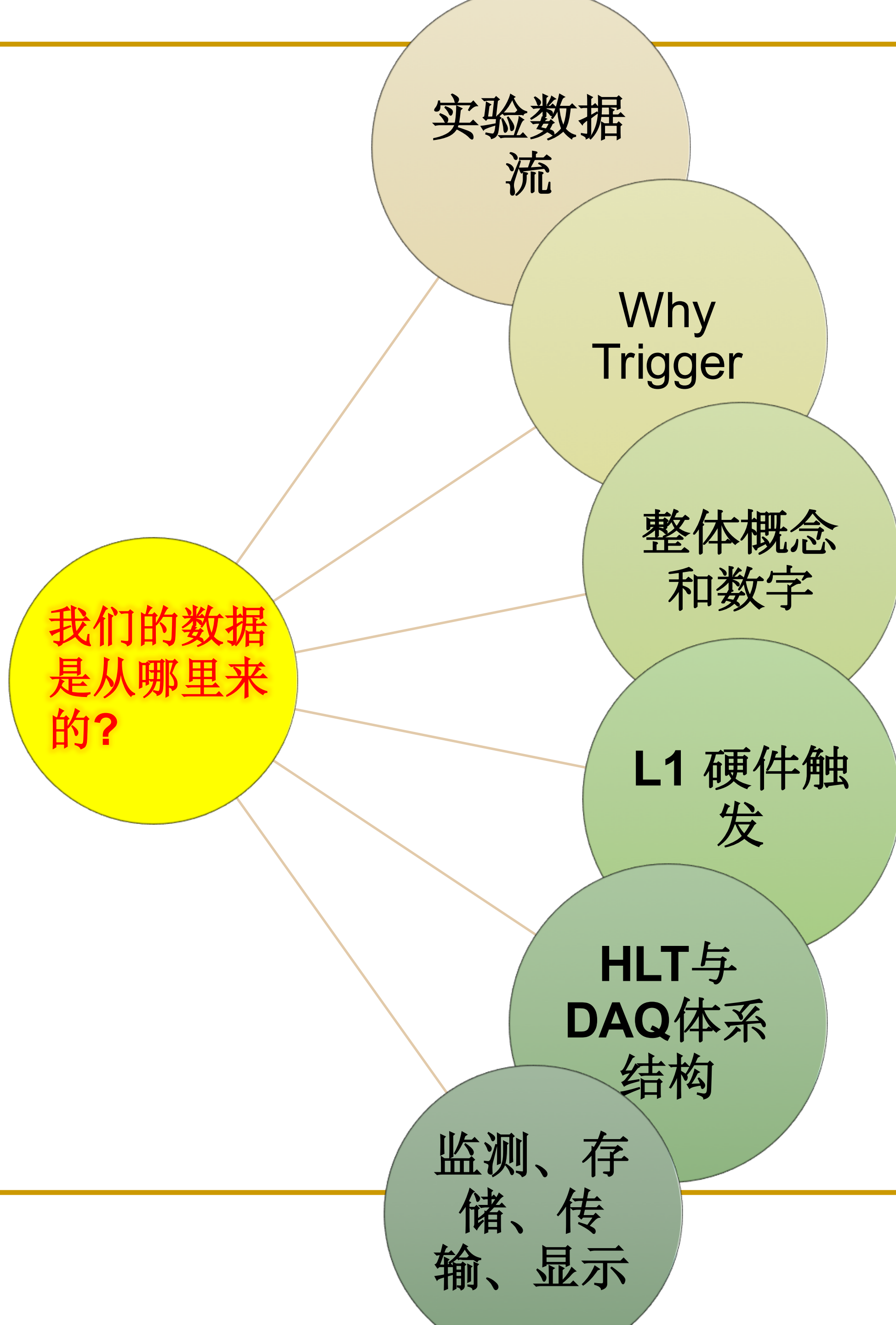
Trigger and DAQ (TriDAS) and beyond

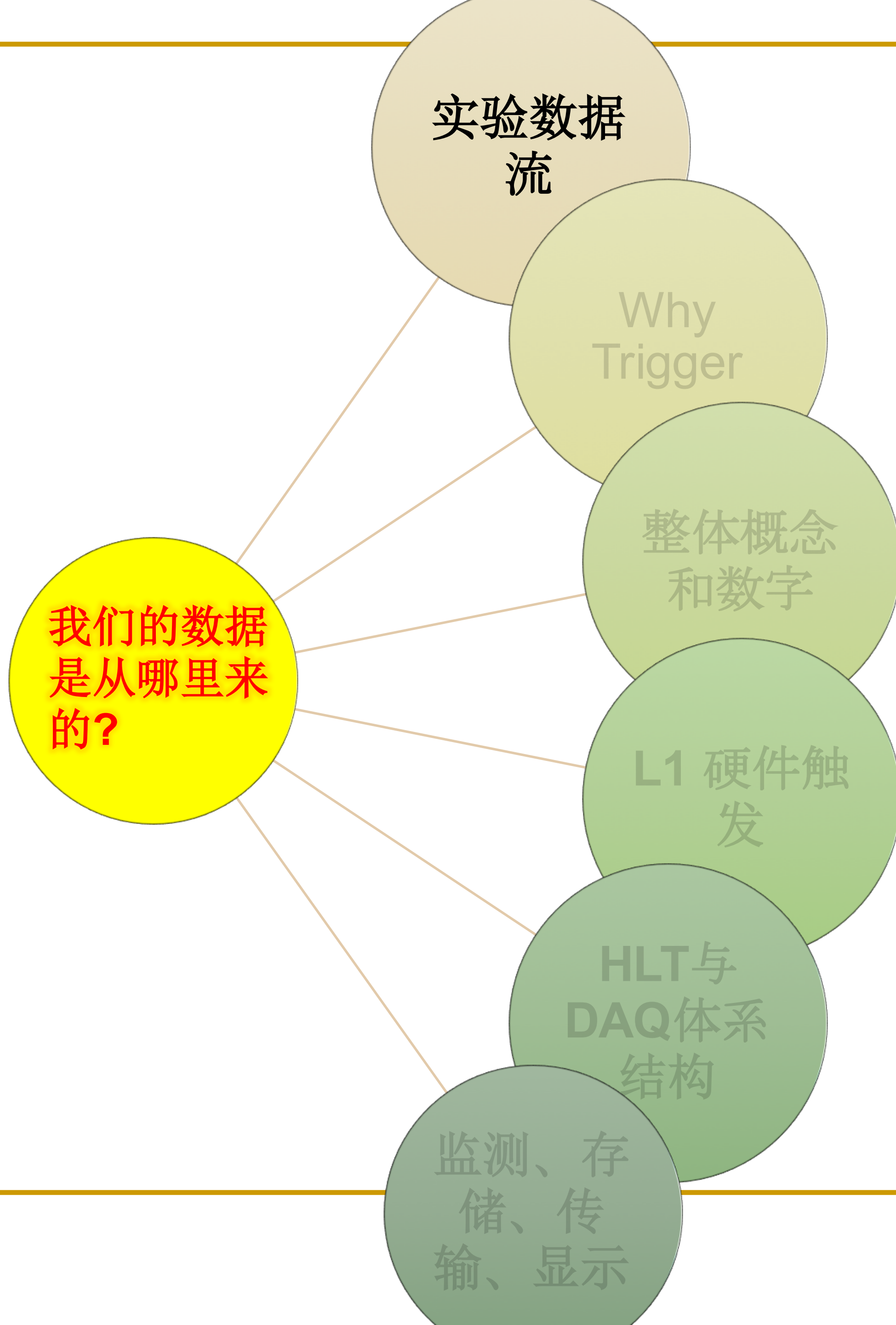
北京大学物理学院　王大勇
dayong.wang@pku.edu.cn

The 3rd International Summer school on TeV Experimental Physics
Tsinghua, July 17, 2016

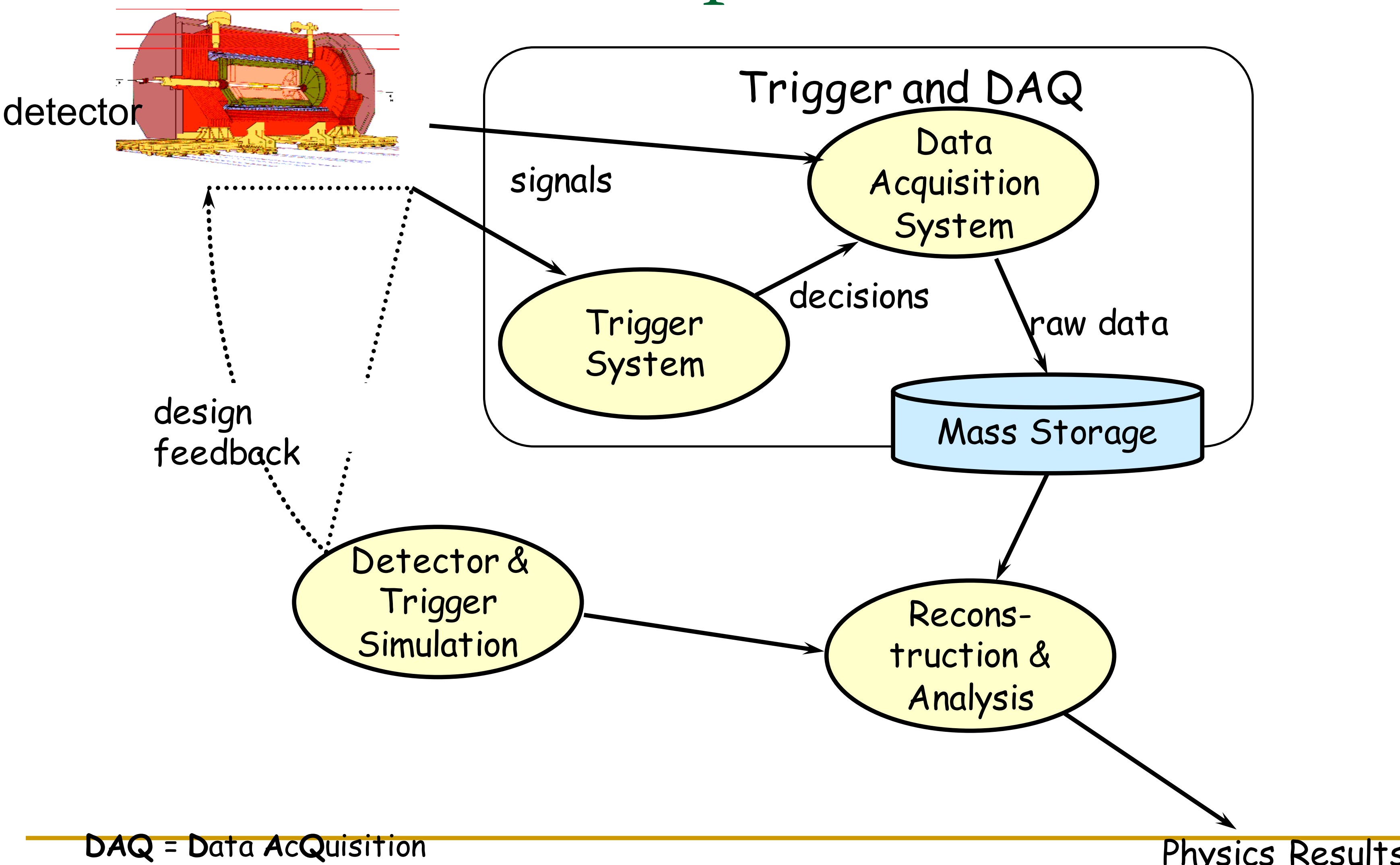
是从天上掉下来的吗？
… 不是。
是自己头（电）脑里固有的吗？
…. 不是。
… 只能从 …科学实验 …中来。

我们的数据是从哪里来的?


我们的数据是从哪里来的?
实验数据流
Why Trigger
整体概念和数字
L1 硬件触发
HLT与DAQ体系结构
监测、存储、传输、显示

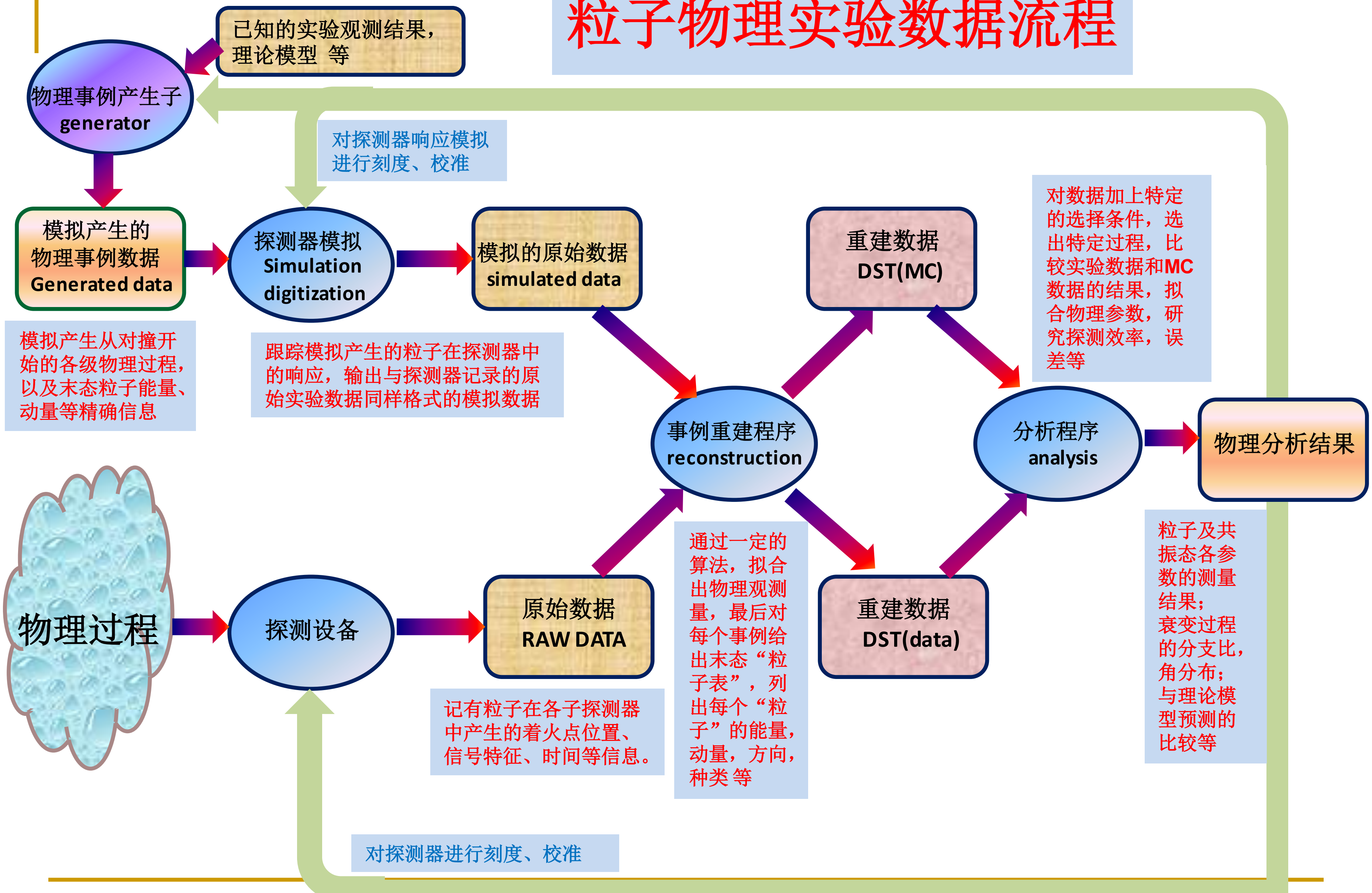

我们的数据是从哪里来的?
实验数据流
Why Trigger
整体概念和数字
L1 硬件触发
HLT与DAQ体系结构
监测、存储、传输、显示

A modern HEP experiment



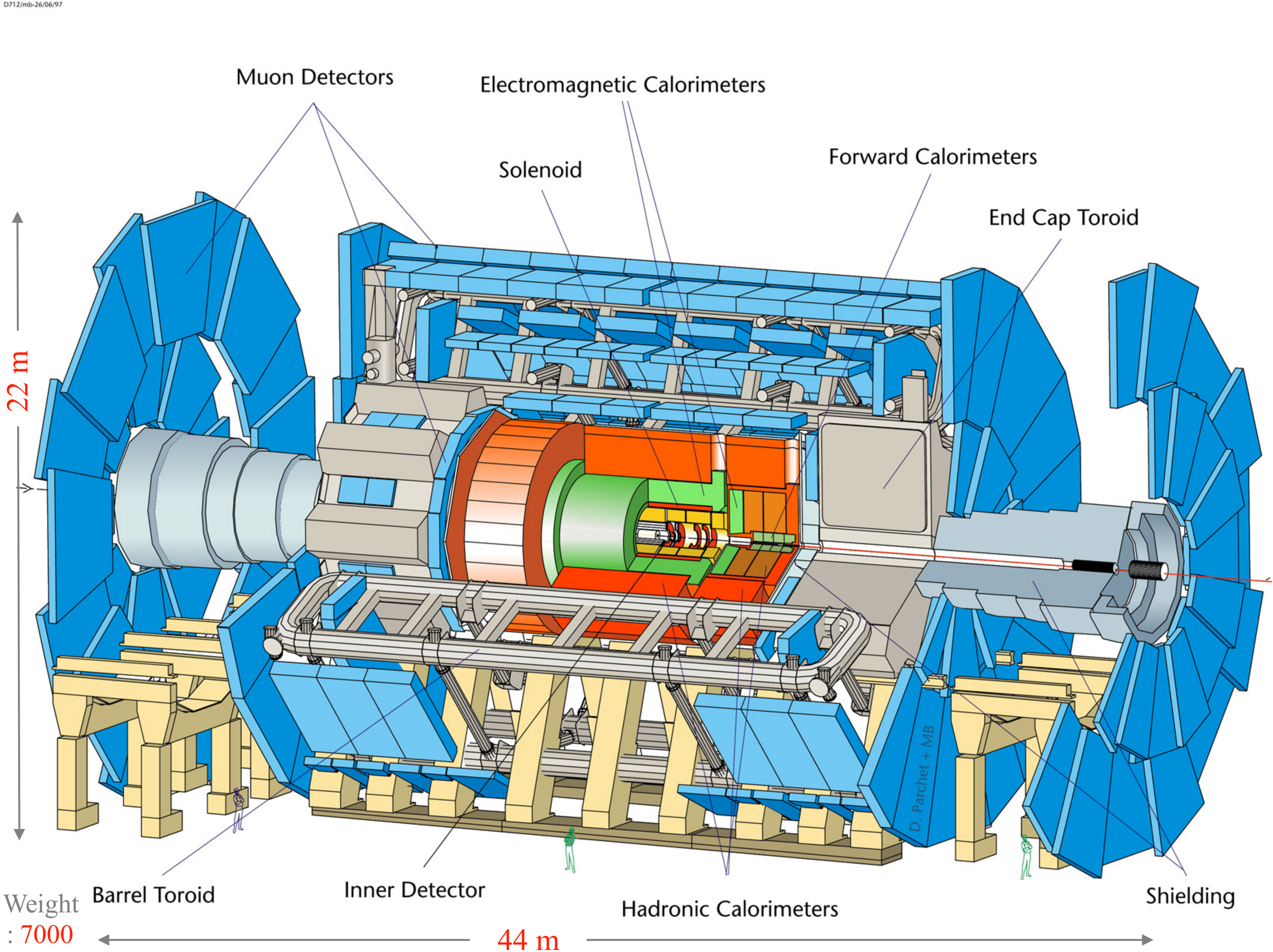
DAQ = Data AcQuisition

粒子物理实验数据流程



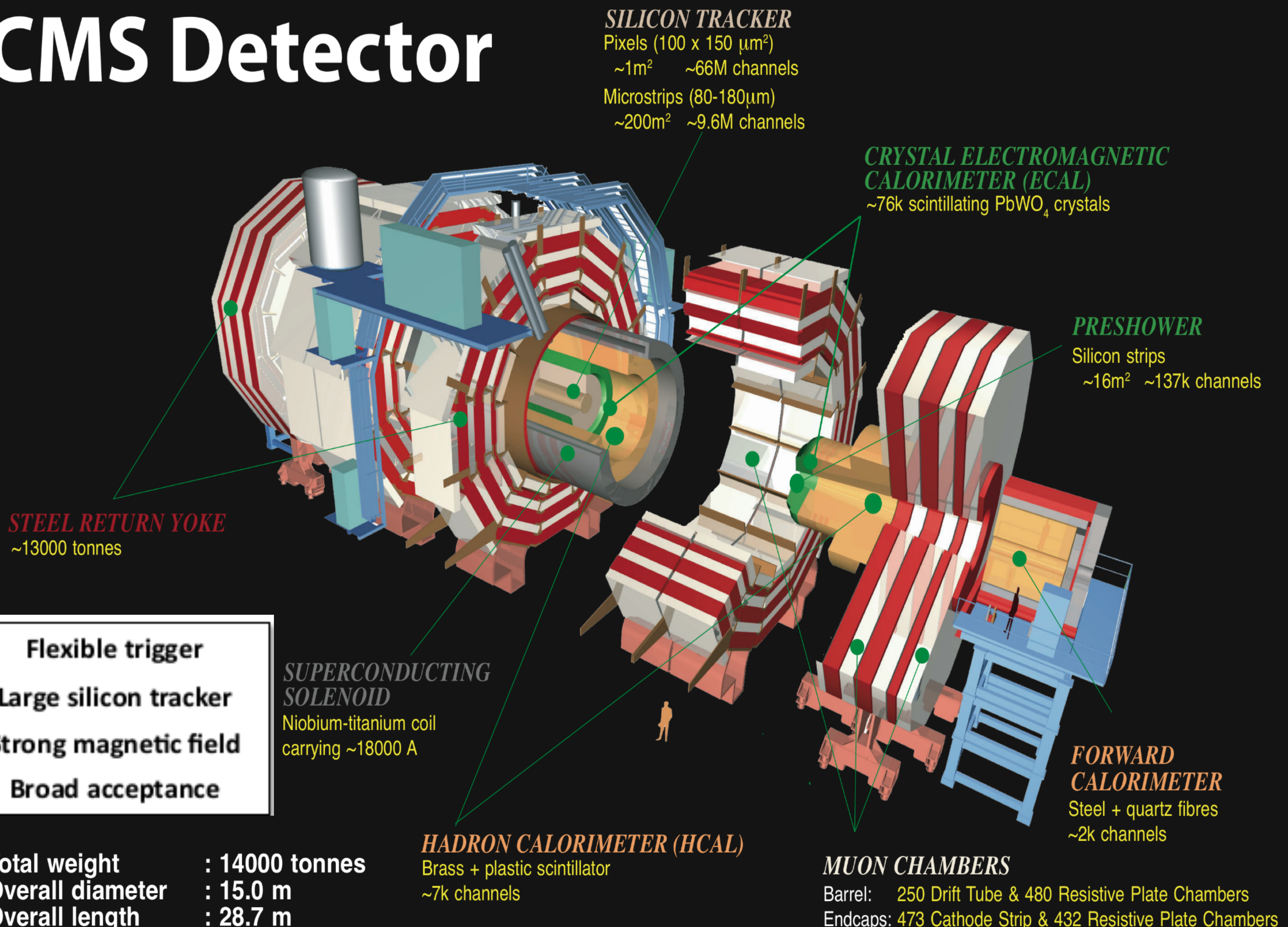
已知的实验观测结果，理论模型 等
物理事例产生子 generator
对探测器响应模拟进行刻度、校准
模拟产生的物理事例数据 Generated data
探测器模拟 Simulation digitization
模拟的原始数据 simulated data
重建数据 DST(MC)
对数据加上特定的选择条件，选出特定过程，比较实验数据和MC数据的结果，拟合物理参数，研究探测效率，误差等
模拟产生从对撞开始的各级物理过程，以及末态粒子能量、动量等精确信息
跟踪模拟产生的粒子在探测器中的响应，输出与探测器记录的原始实验数据同样格式的模拟数据
事例重建程序 reconstruction
分析程序 analysis
物理分析结果
物理过程
探测设备
原始数据 RAW DATA
重建数据 DST(data)
通过一定的算法，拟合出物理观测量，最后对每个事例给出末态“粒子表”，列出每个“粒子”的能量，动量，方向，种类 等
粒子及共振态各参数的测量结果；衰变过程的分支比，角分布；与理论模型预测的比较等
记有粒子在各子探测器中产生的着火点位置、信号特征、时间等信息。
对探测器进行刻度、校准

ATLAS (A Toroidal Lhc ApparatuS)



Weight : 7000

CMS Detector



Flexible trigger

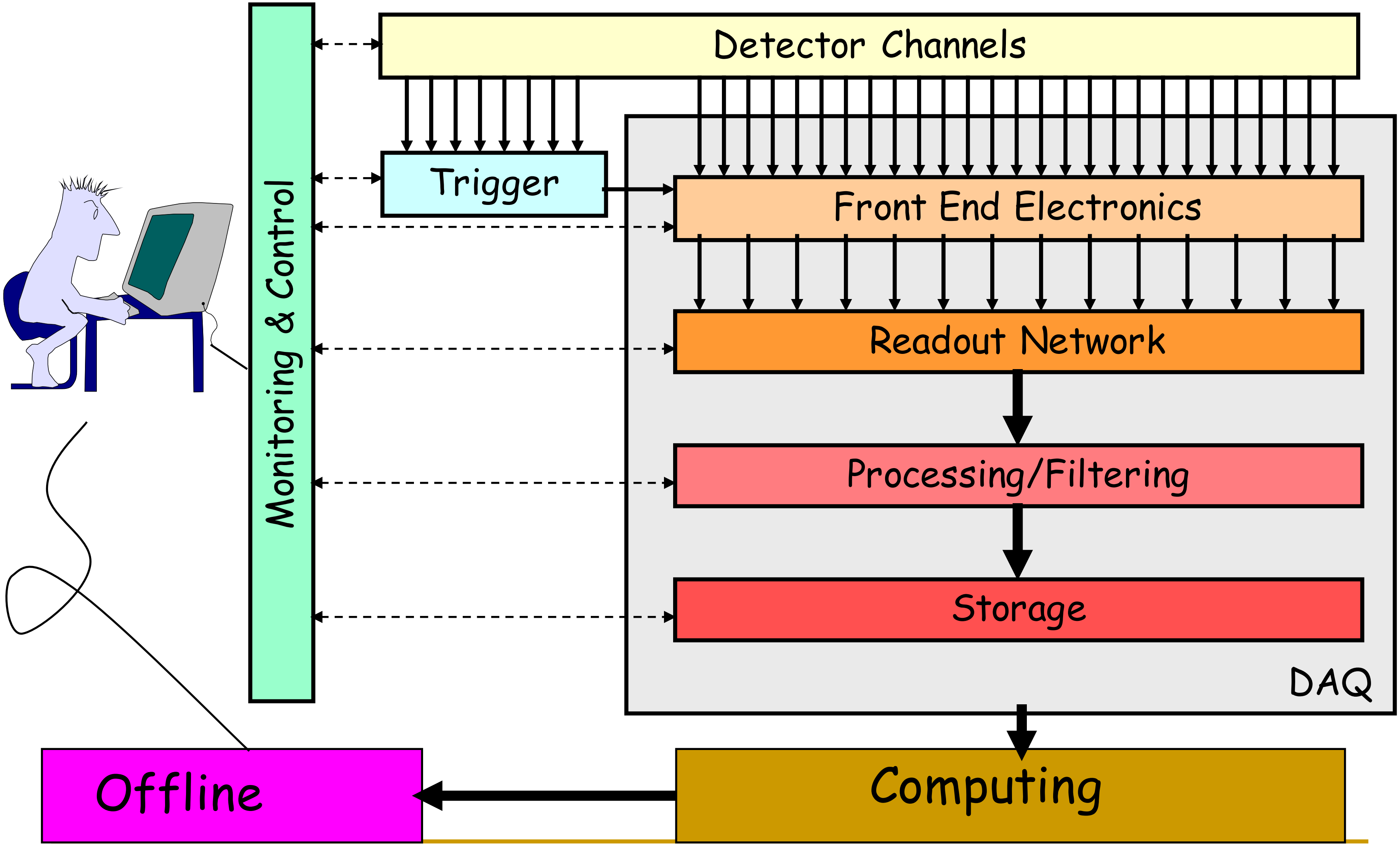
Large silicon tracker

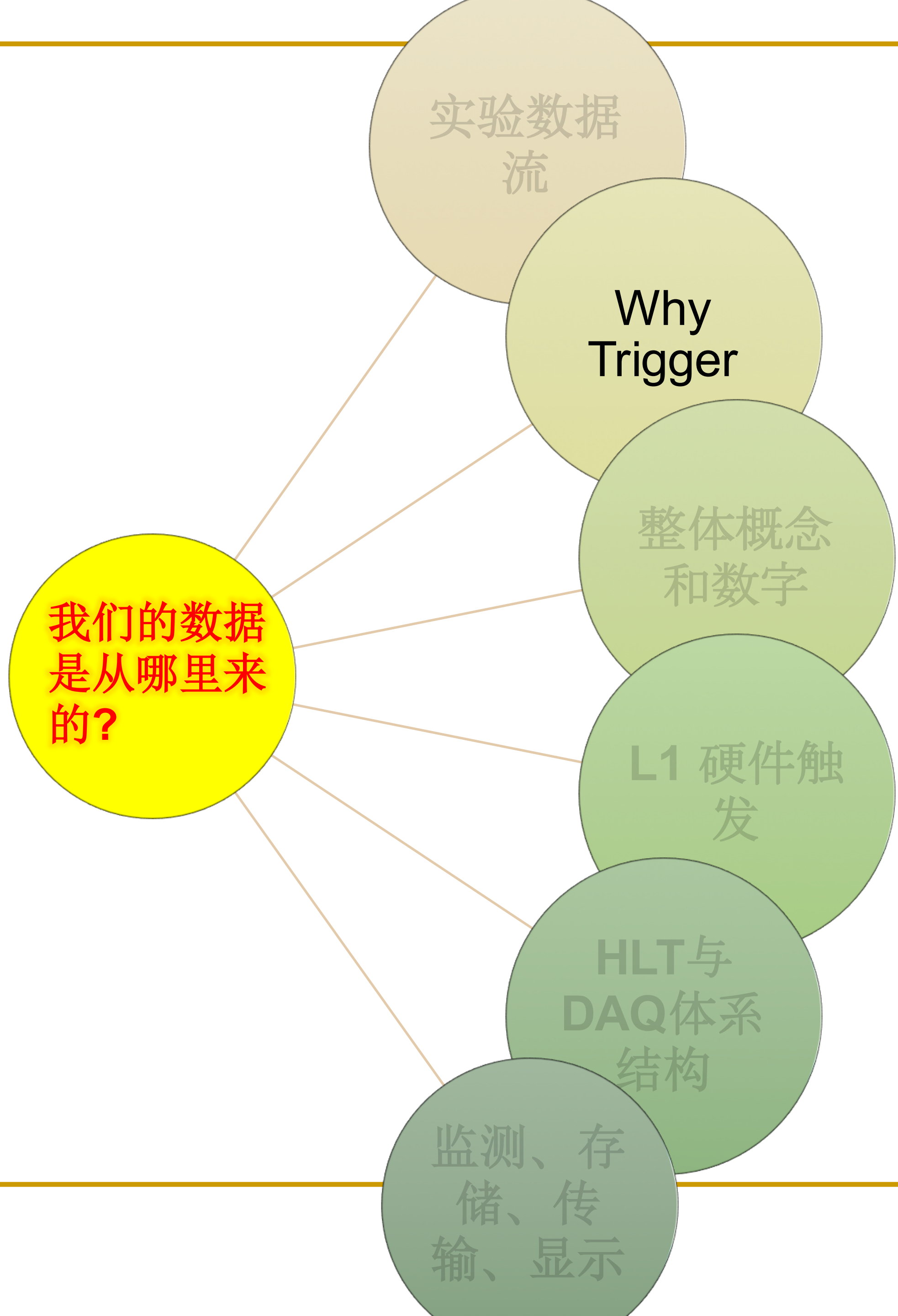
Strong magnetic field

Broad acceptance

Total weight	: 14000 tonnes
Overall diameter	: 15.0 m
Overall length	: 28.7 m
Magnetic field	: 3.8 T

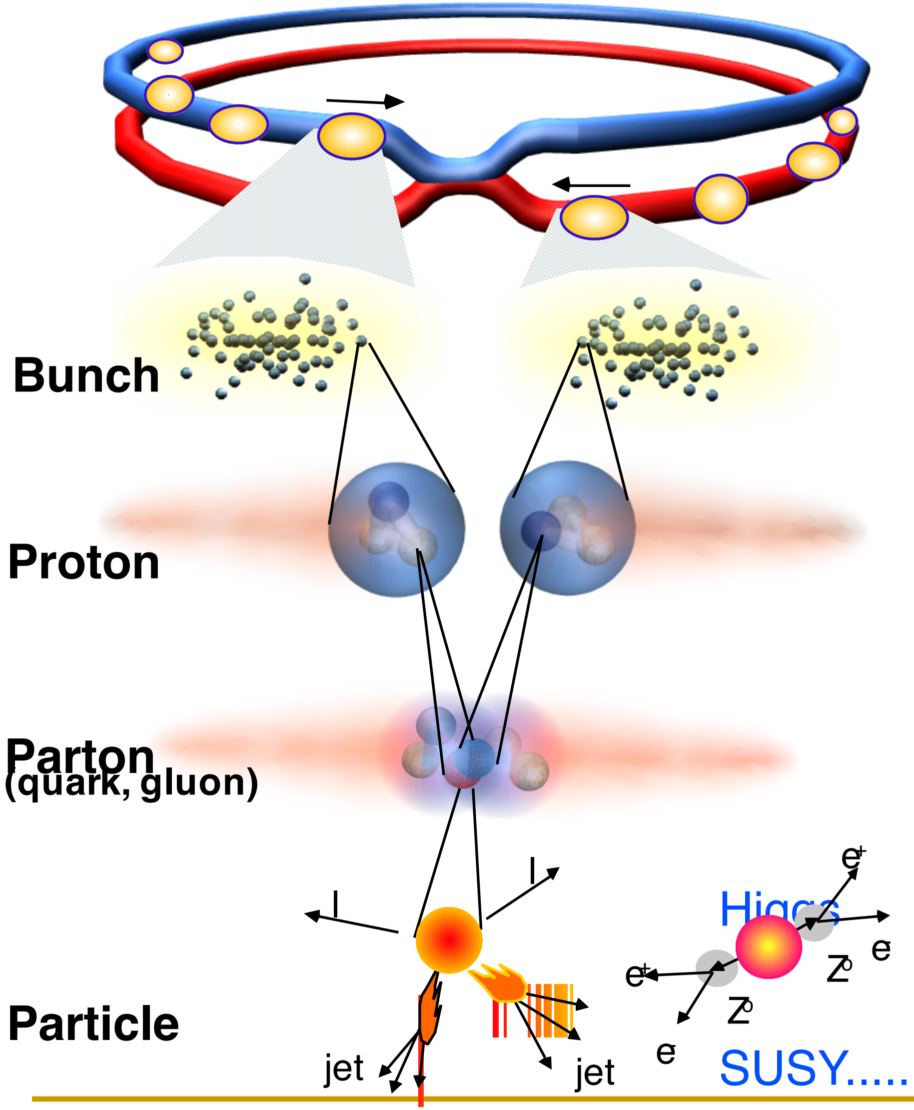
Online: Trigger, DAQ, Monitoring & Control …





实验数据流
Why Trigger
整体概念和数字
我们的数据是从哪里来的?
L1 硬件触发
HLT与DAQ体系结构
监测、存储、传输、显示

Collisions at the LHC:



Proton - Proton

	2804 bunch/beam
Protons/bunch	10^{11}
Beam energy	7 TeV (7×10^{12} eV)
Luminosity	$10^{34} cm^{-2}s^{-1}$
Crossing rate	40 MHz(every25ns)
Collision rate ≈	$7 \times 10^{8}\ s^{-1}$
data recording rate	300 Hz 200-300 MB/sec

New physics rate ≈ .00001 Hz

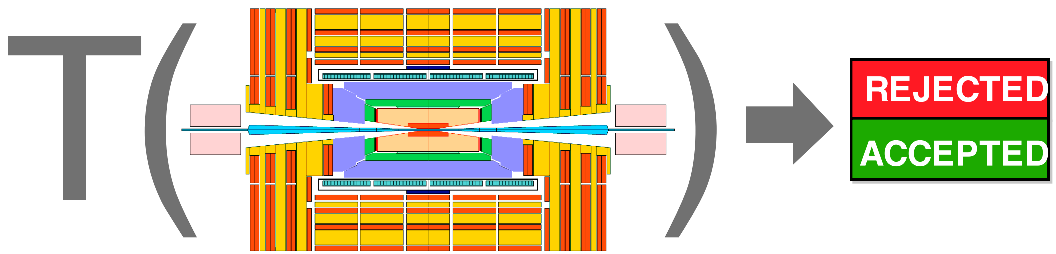
Event selection:

1 in 10,000,000,000,000

Triggering

- Task: inspect detector information and provide a first decision on whether to keep the event or throw it out

The trigger is a function of :



Event data & Apparatus
Physics channels & Parameters

- **Detector data not (all) promptly available**
- **Selection function highly complex**

⇒T(...) is evaluated by successive approximations, the

TRIGGER LEVELS

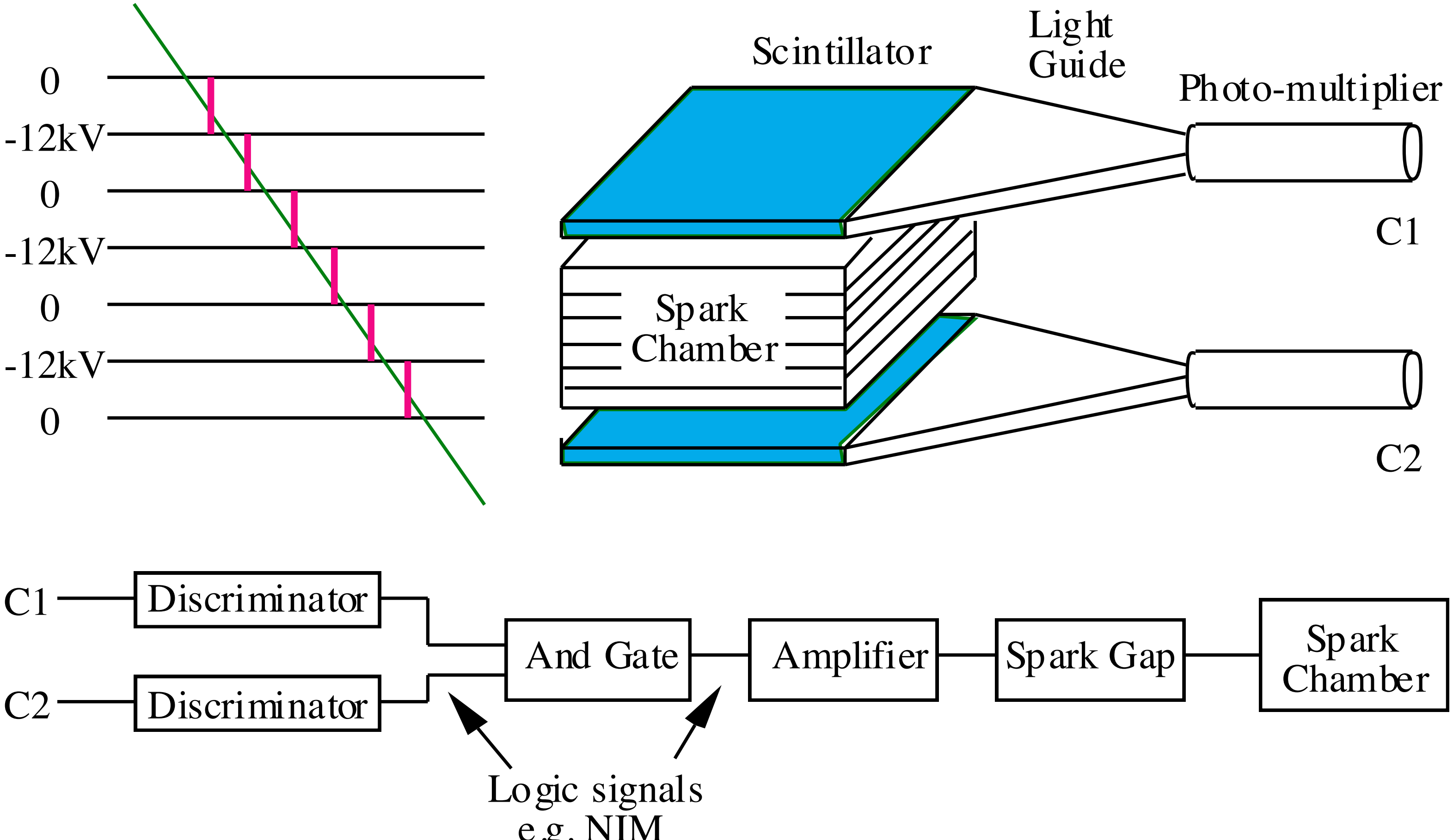
(possibly with zero dead time)

General trigger requirements

- The role of the **trigger** is to make the **online selection** of particle collisions potentially containing interesting physics
- Need **high efficiency** for selecting processes of interest for physics analysis
  - Efficiency should be precisely known
  - Selection should not have biases that affect physics results
- Need **large reduction of rate** from unwanted high-rate processes (capabilities of DAQ and also offline computers)
  - Instrumental background
  - High-rate physics processes that are not relevant for analysis
- System must be **affordable**
  - Limits complexity of algorithms that can be used
- Not easy to achieve all the above simultaneously!
- And never forget that **an event rejected by the Trigger is lost for ever!**

No Trigger, No Physics !

1960/70s:Simple trigger for spark chamber



Dead time

- Experiments frozen from trigger to end of readout
  - Trigger rate with no deadtime = R per sec.
  - Dead time / trigger = τ sec.
  - For 1 second of live time = 1 + Rτ seconds
  - Live time fraction = 1/(1 + Rτ)
  - Real trigger rate = R/(1 + Rτ) per sec.

Rate in Hz	Dead time ms.	Live time %	Trigger rate Hz
10	10	91	9.1
1000	10	9.1	91

Solution: multi-level triggers!
Exercise: the good event trigger rate of the two level trigger is

$$S = \frac{pfR}{1 + R(t + f\tau)}$$

where p is the purity of the event sample selected by the second-level trigger, R is the first-level trigger rate, t is the second-level decision time, f is the fraction of events that pass the second-level trigger, and τ is the readout time.

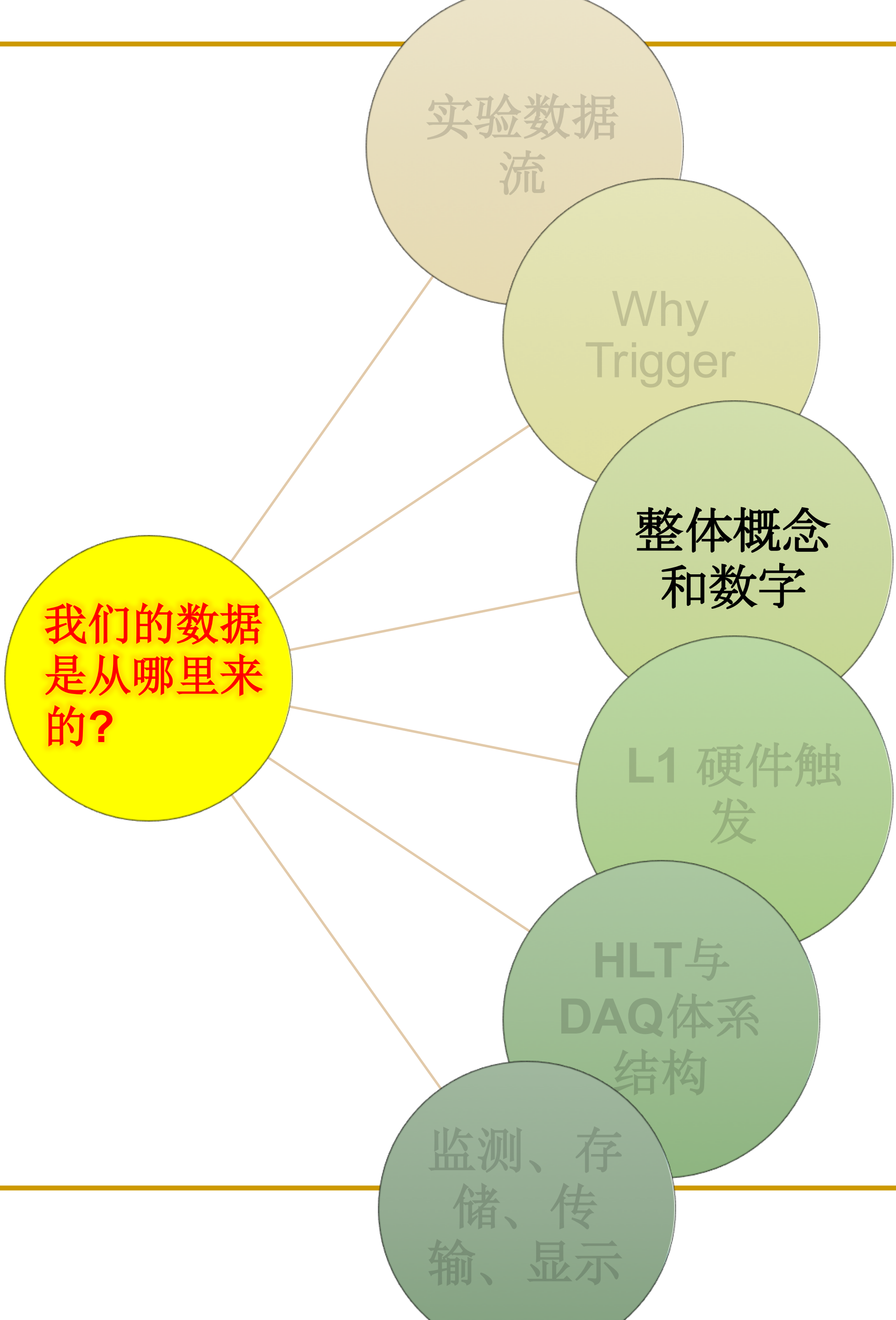
Trigger systems 1980's and 90's

- bigger experiments → more data per event
- higher luminosities → more triggers per second
  - both led to increased fractional deadtime

- use multi-level triggers to reduce dead-time
  - first level - fast detectors, fast algorithms
  - higher levels can use data from slower detectors and more complex algorithms to obtain better event selection/background rejection

Trigger systems 1990's and 2000's

- Dead-time was not the only problem
- Experiments focussed on rarer processes
  - Need large statistics of these rare events
  - But increasingly difficult to select the interesting events
  - DAQ system (and off-line analysis capability) under increasing strain - limiting useful event statistics
    - This is a major issue at hadron colliders, but will also be significant at ILC
- Use the High Level Trigger to reduce the requirements for
  - The DAQ system
  - Off-line data storage and off-line analysis

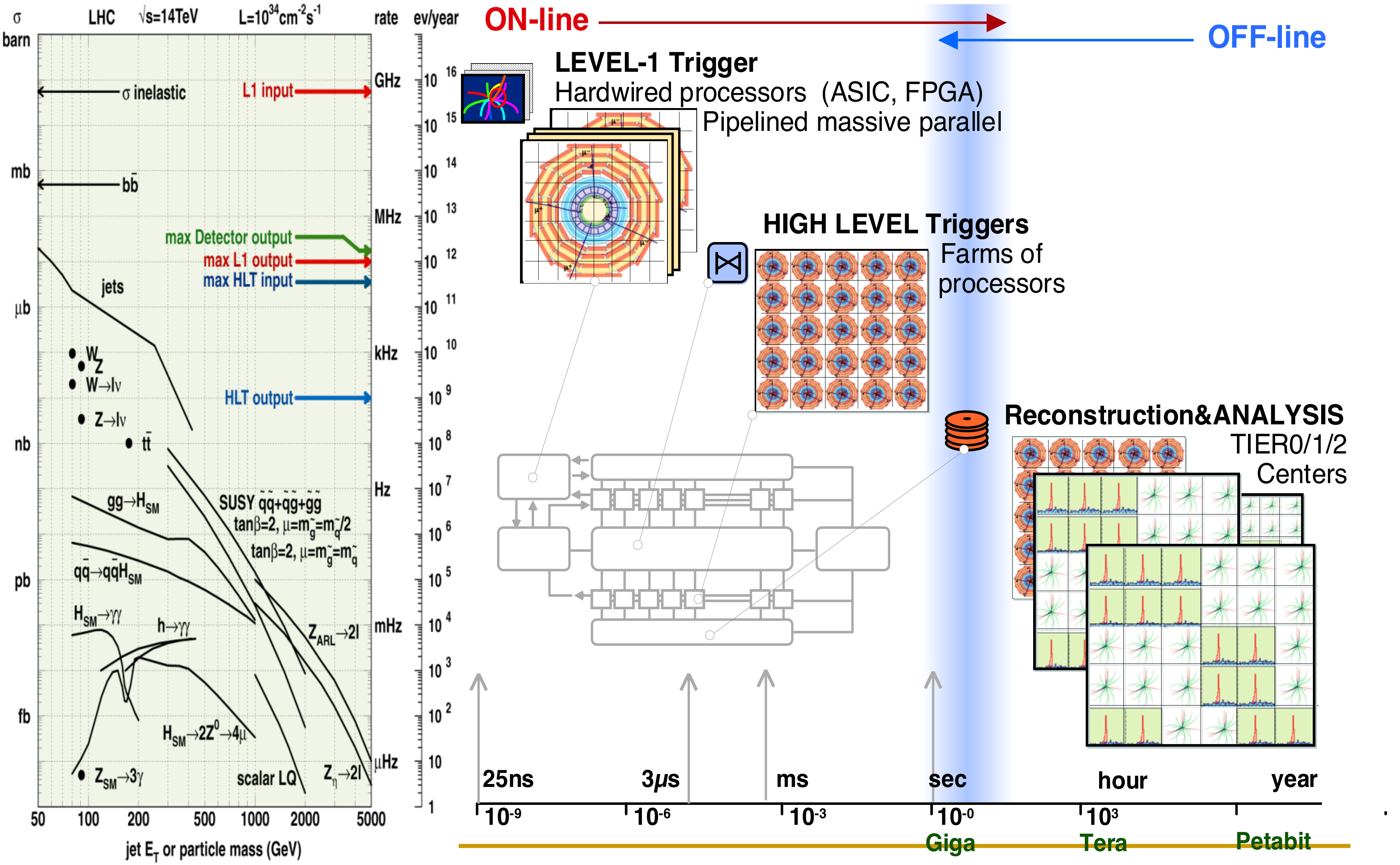

我们的数据是从哪里来的?
实验数据流
Why Trigger
整体概念和数字
L1 硬件触发
HLT与DAQ体系结构
监测、存储、传输、显示

实验的取数环境：加速器的时间结构，事例率和数据率，数据量，死时间

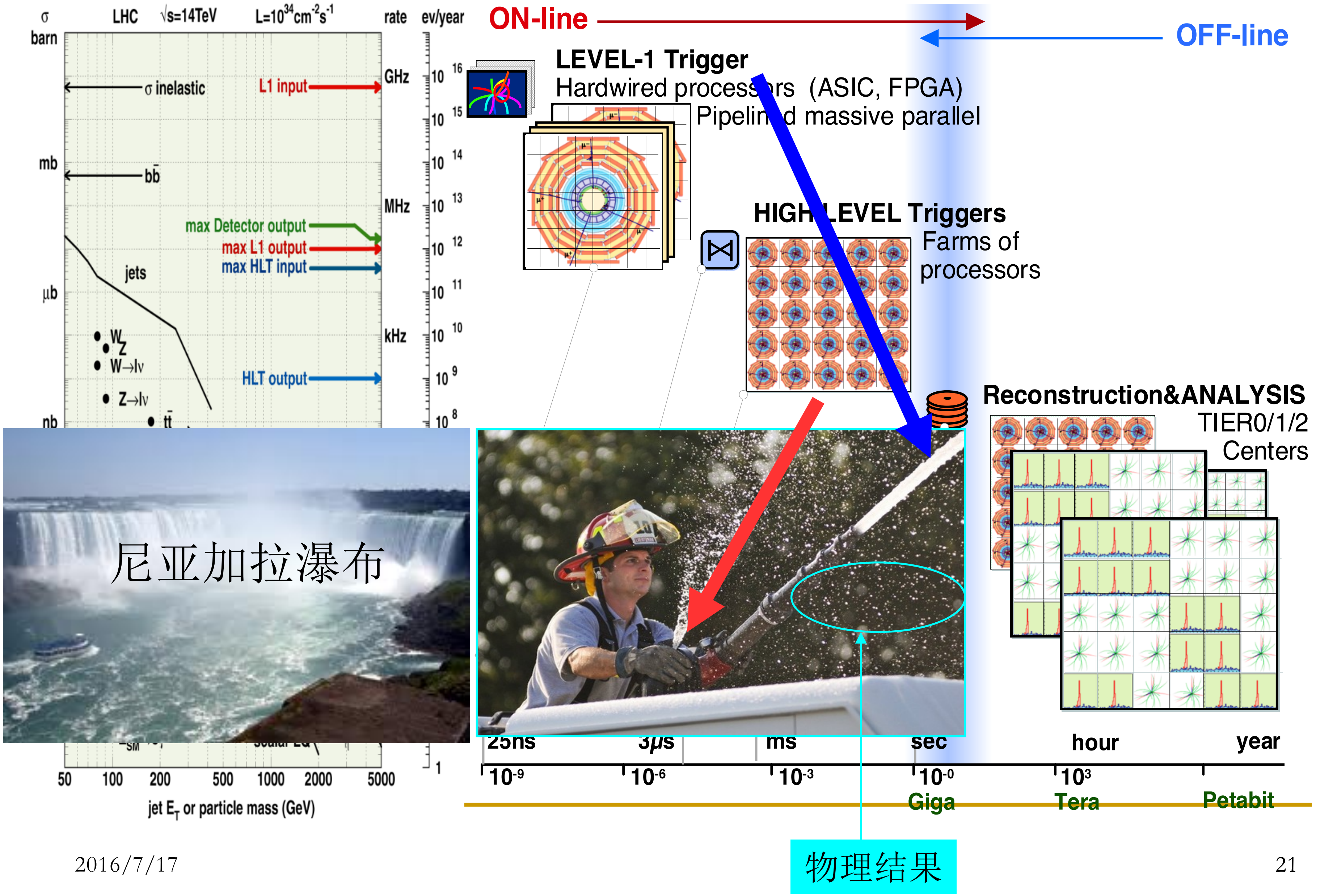
一些对撞机及相应探测器的参数

对撞机	最大束流能量/GeV	亮度/(10^{30} cm^{-2} s^{-1})	周长 /km	束团数	对撞周期 /ns	探测器	电子学道数
DAΦNE	0.75	5	0.0977	30~120	2.7 ~ 10.8	KLOE	23k
BEPC	2.2	5	0.2404	1	802	BES	20k
CESR	6	830	0.768	9X4	14-220	CLEO	400k
LEP	101	10000	26.66	5120	22000	ALEPH etc	100~ 300k
KEKB	8+3.5	3000	3.016	1658	2	BELLE	133k
HERA	e30+ p920	14	6.336	189+180	96	H1 etc	250k
Tevatron	1000	210	6.28	36	396	CDF etc	75~100k
LHC	7000	10000	26.66	2835	25	CMS etc	100M

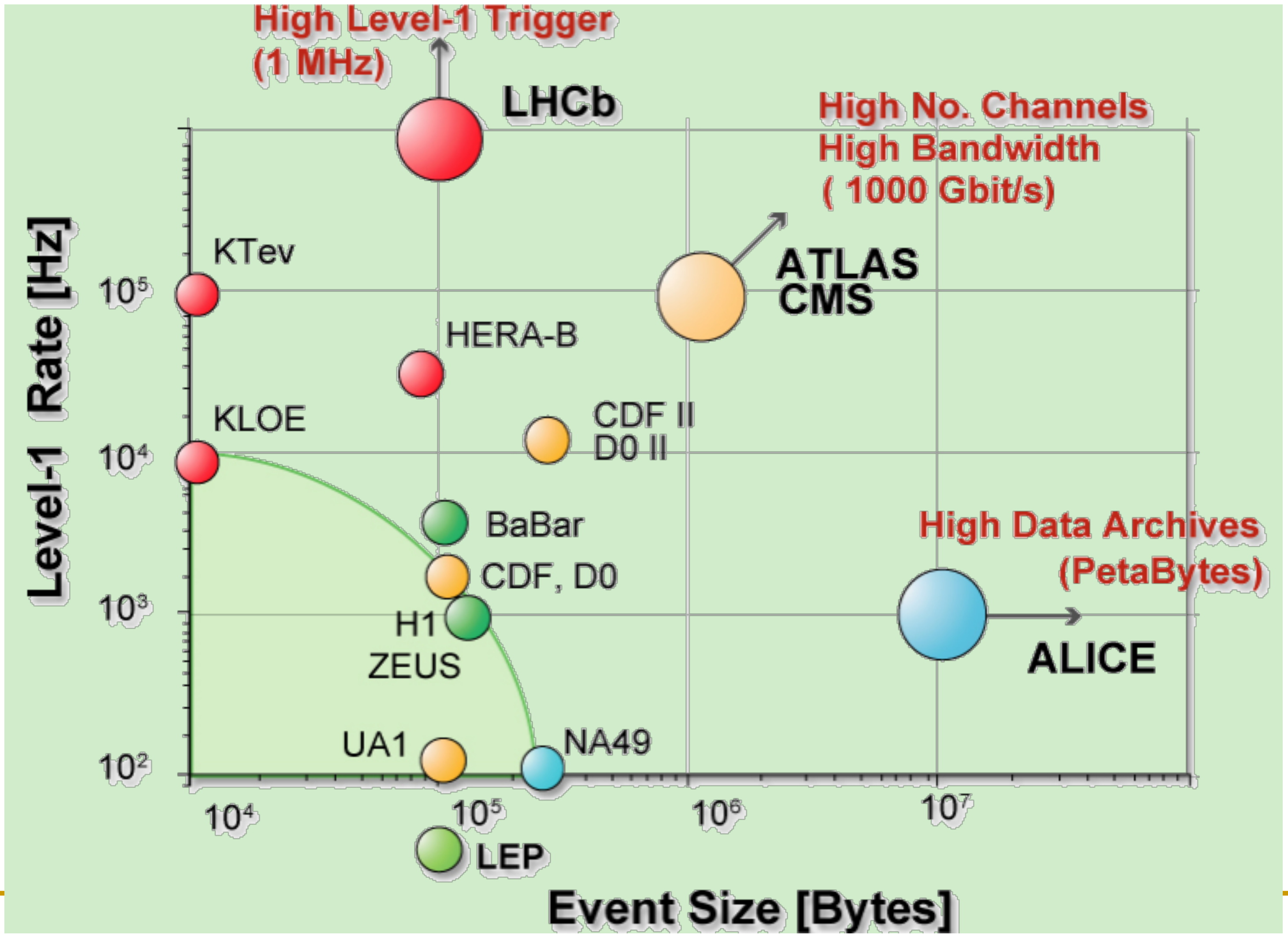
LHC实验数据流：物理信号的挑选



触发系统：现代对撞实验的大脑



Data Rate: TrigDAQ Comparisons



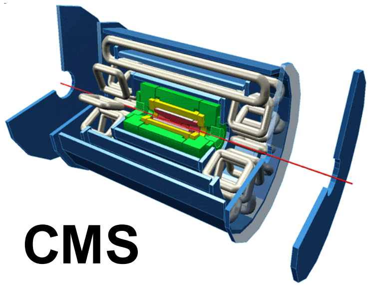
ATLAS/CMS Data Flow Rates

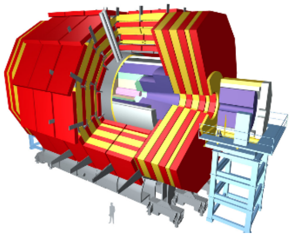
- From detectors> 10^{14} Bytes/sec
- After Level-1 accept~ 10^{11} Bytes/sec
- Into event builder~ 10^{9} Bytes/sec
- Onto permanent storage~ 10^{8} Bytes/sec

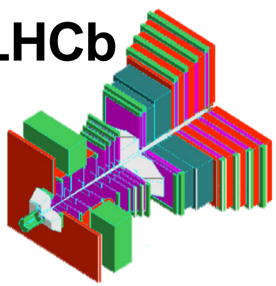
  $\rightarrow$ ~ 10^{15} Bytes/year

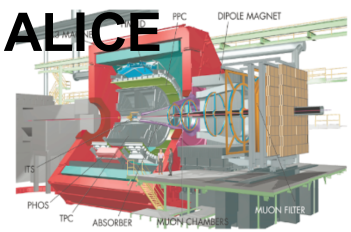
LHC Trigger and DAQ summary

	No.Levels Trigger	First Level Rate (Hz)	Event Size (Byte)	Readout Bandw.(GB/s)	Filter Out MB/s (Event/s)
ATLAS	**3**	**10^5** LV-2 **10^3**	**10^6**	**10**	**100** (10^2)
CMS	**2**	**10^5**	**10^6**	**100**	**100** (10^2)
LHCb	**3**	LV-0 **10^6** LV-1 **$4\ 10^4$**	**$2x10^5$**	**4**	**40** ($2x10^2$)
ALICE	**4**	Pp-Pp **500** p-p **10^3**	**$5x10^7$** **$2x10^6$**	**5**	**1250** (10^2) **200** (10^2)

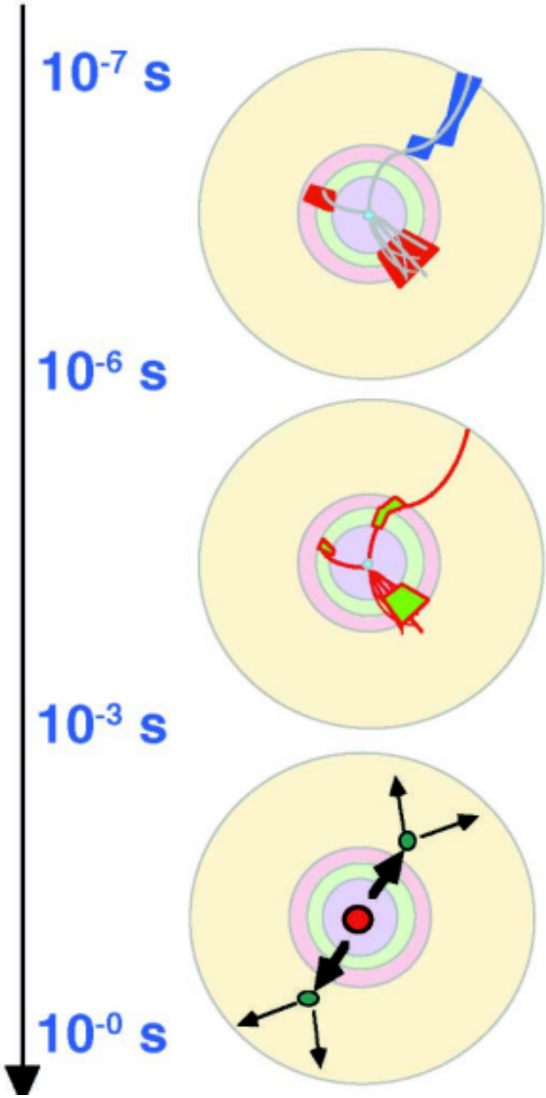








Trigger Levels



Collision rate 10^9 Hz

Channel data sampling at 40 MHz

Level-1 selected events 10^5 Hz

Particle identification (High p_T e, μ, jets, missing E_T)

- Local pattern recognition
- Energy evaluation on prompt macro-granular information

Level-2 selected events 10^3 Hz

Clean particle signature (Z, W, ..)

- Finer granularity precise measurement
- Kinematics. effective mass cuts and event topology
- Track reconstruction and detector matching

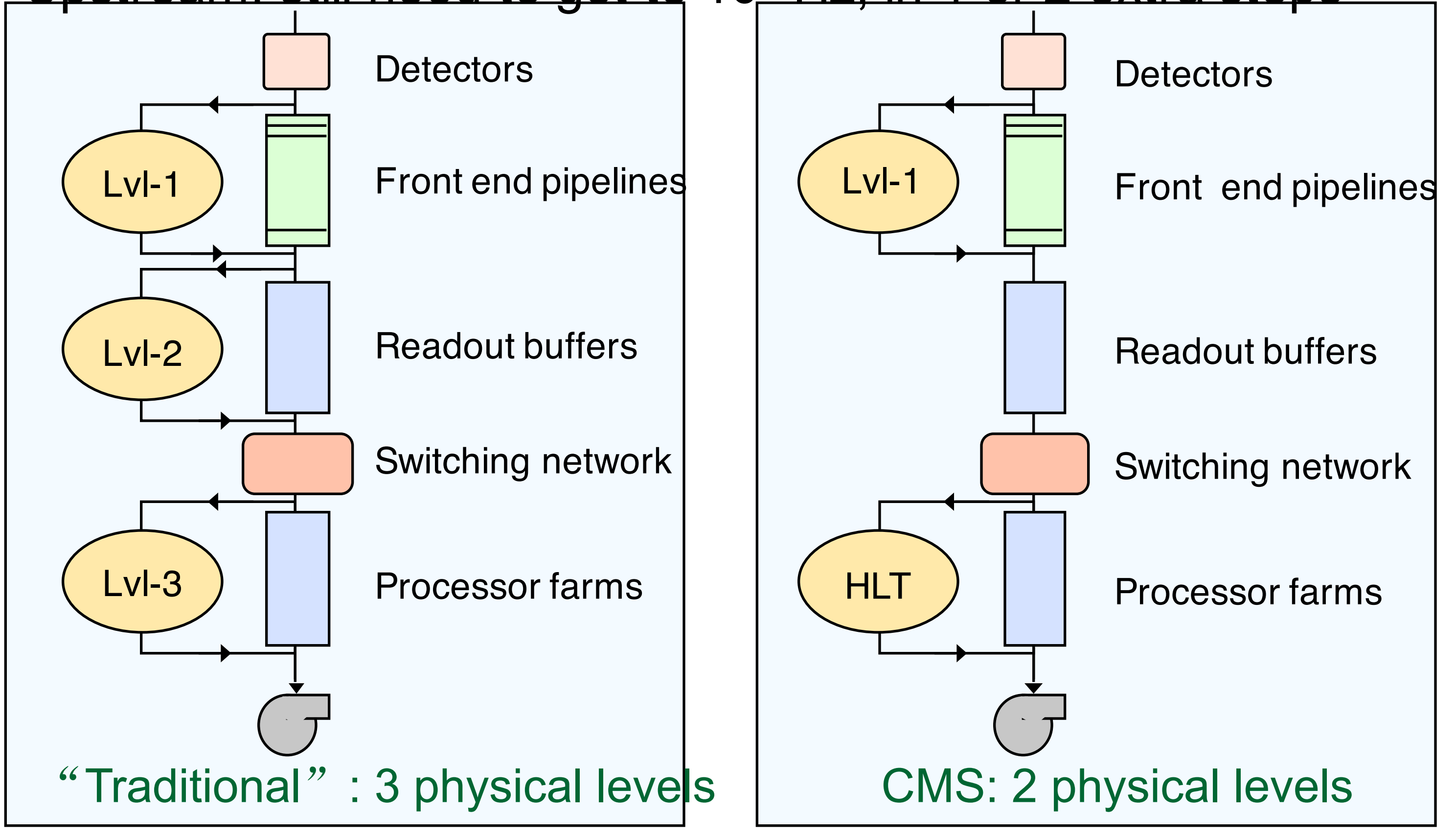
Level-3 events to tape 100- 400 Hz

Physics process identification

- Event reconstruction and analysis

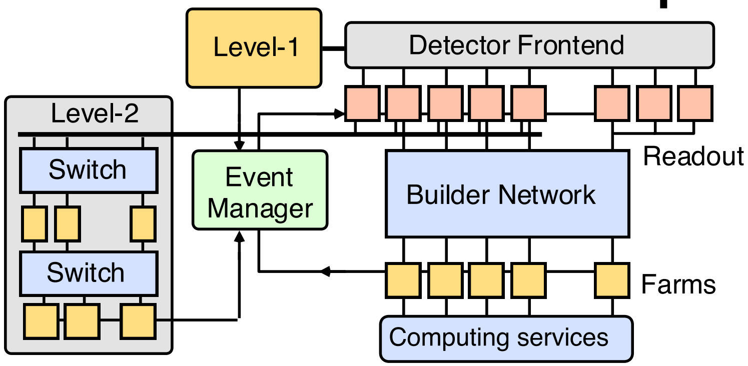
Level-1 Trigger

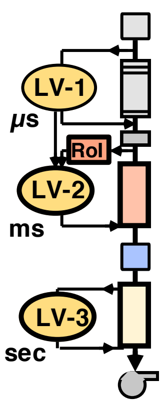
- Level-1 trigger: reduce 40 MHz to 10^5 Hz
  - This step is always there
  - Upstream: still need to get to 10^2 Hz; in 1 or 2 extra steps



Three physical entities(ATLAS)

- Additional processing in LV-2: reduce network bandwidth requirements





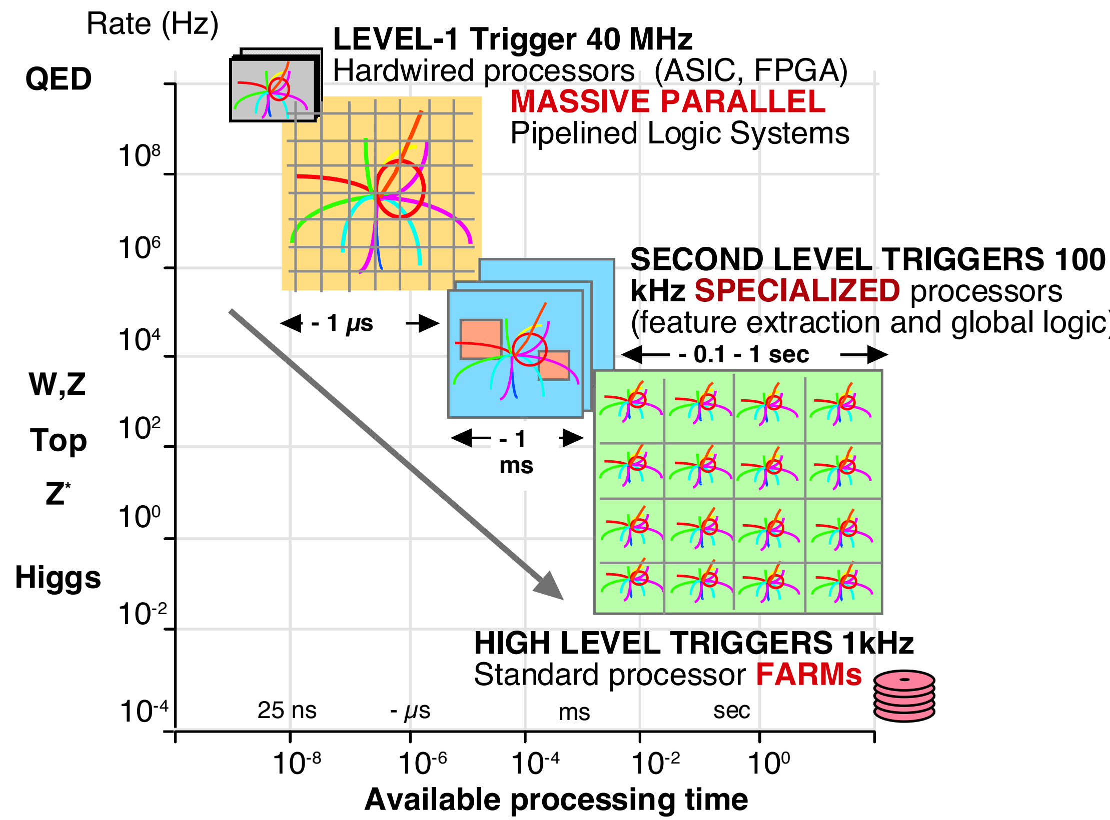
40 MHz

10^5 Hz

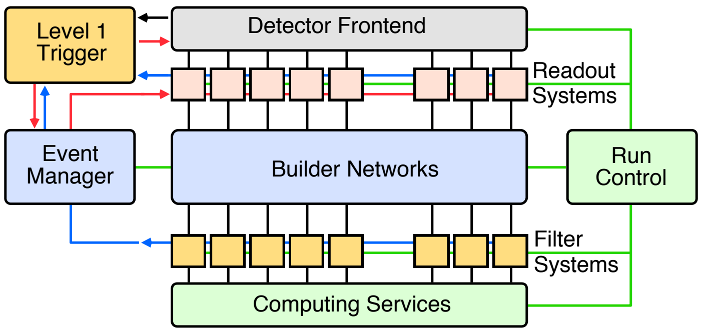
10^3 Hz

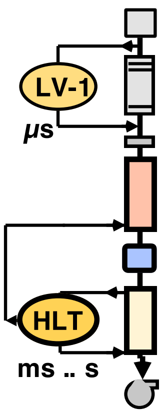
10 Gb/s

10^2 Hz



Two physical entities(CMS)



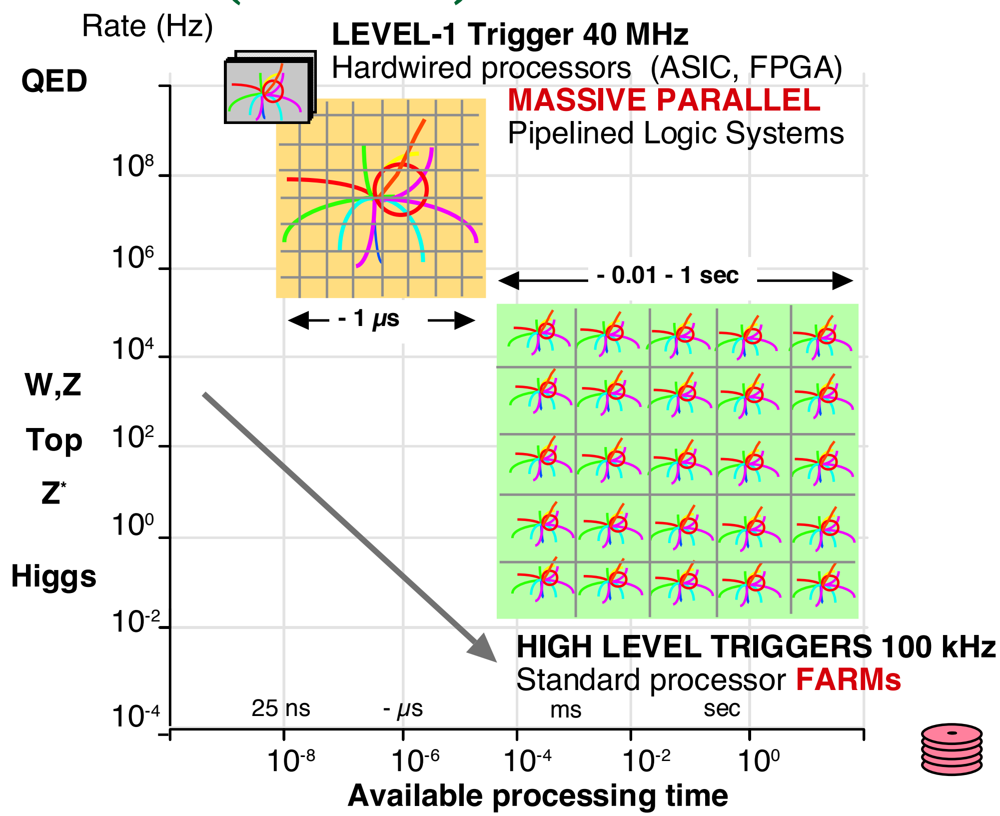


40 MHz

10^5 Hz

1000 Gb/s

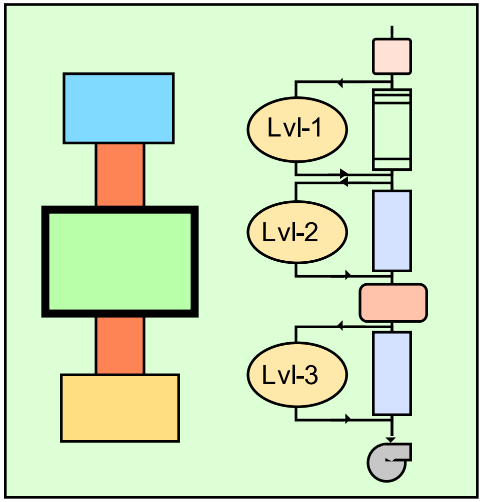
10^2 Hz



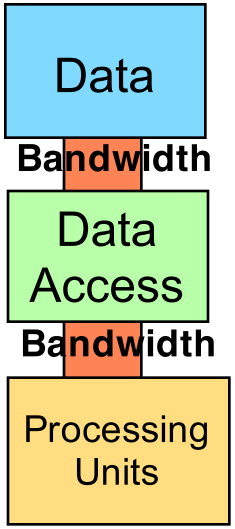
- Reduce number of building blocks
- Rely on commercial components (especially processing and communications)

Comparison of 2 vs 3 physical levels

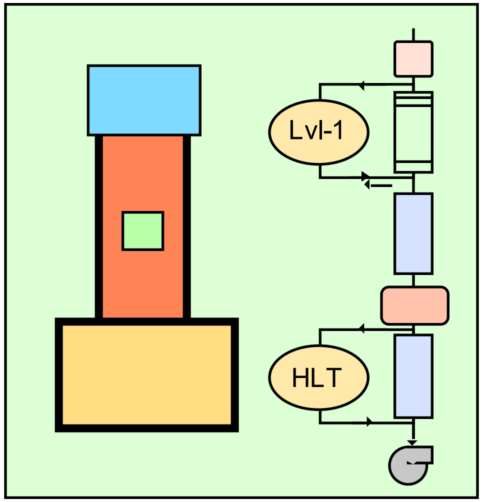
- Three Physical Levels
  - Investment in:
    - Control Logic
    - Specialized processors

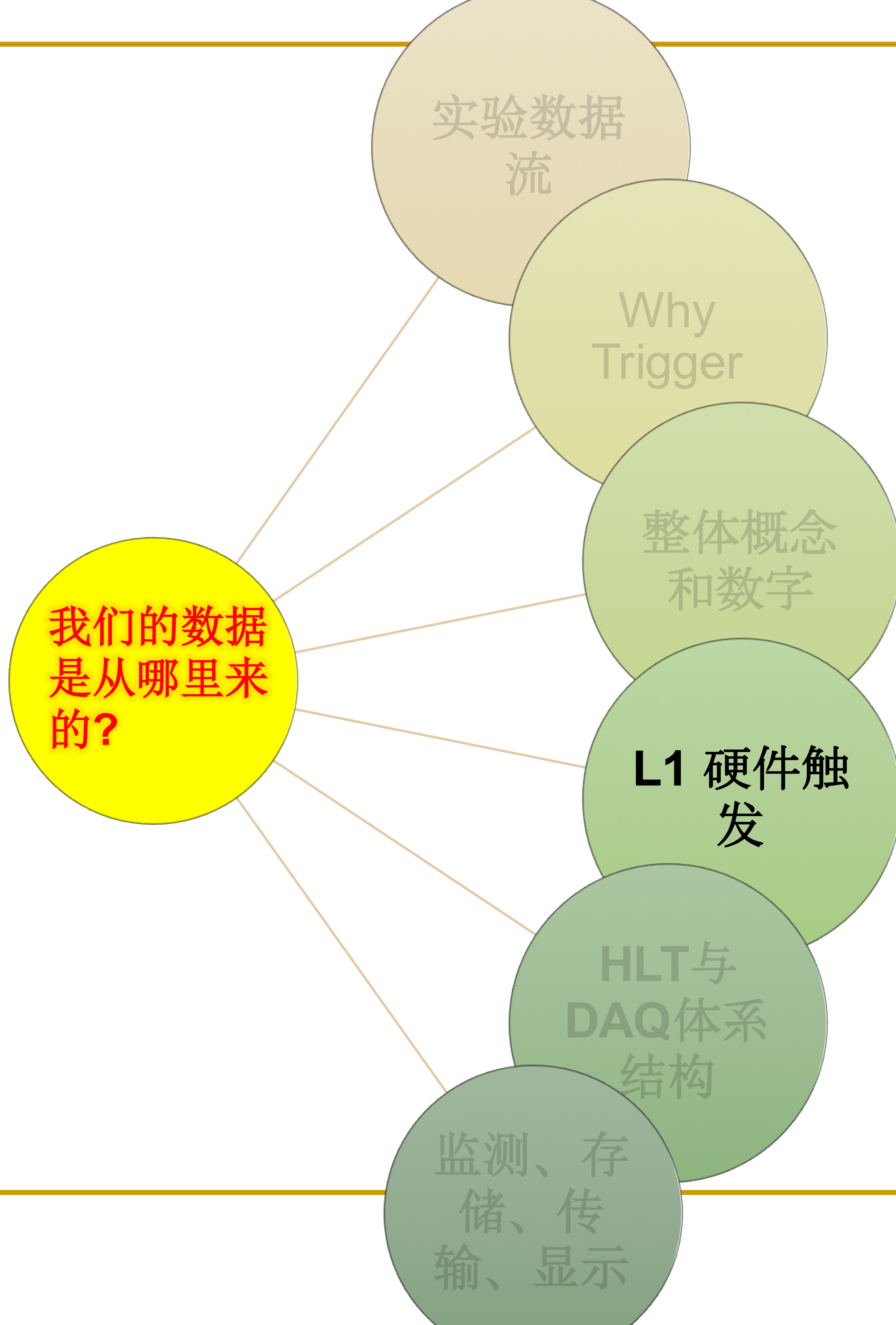


Model

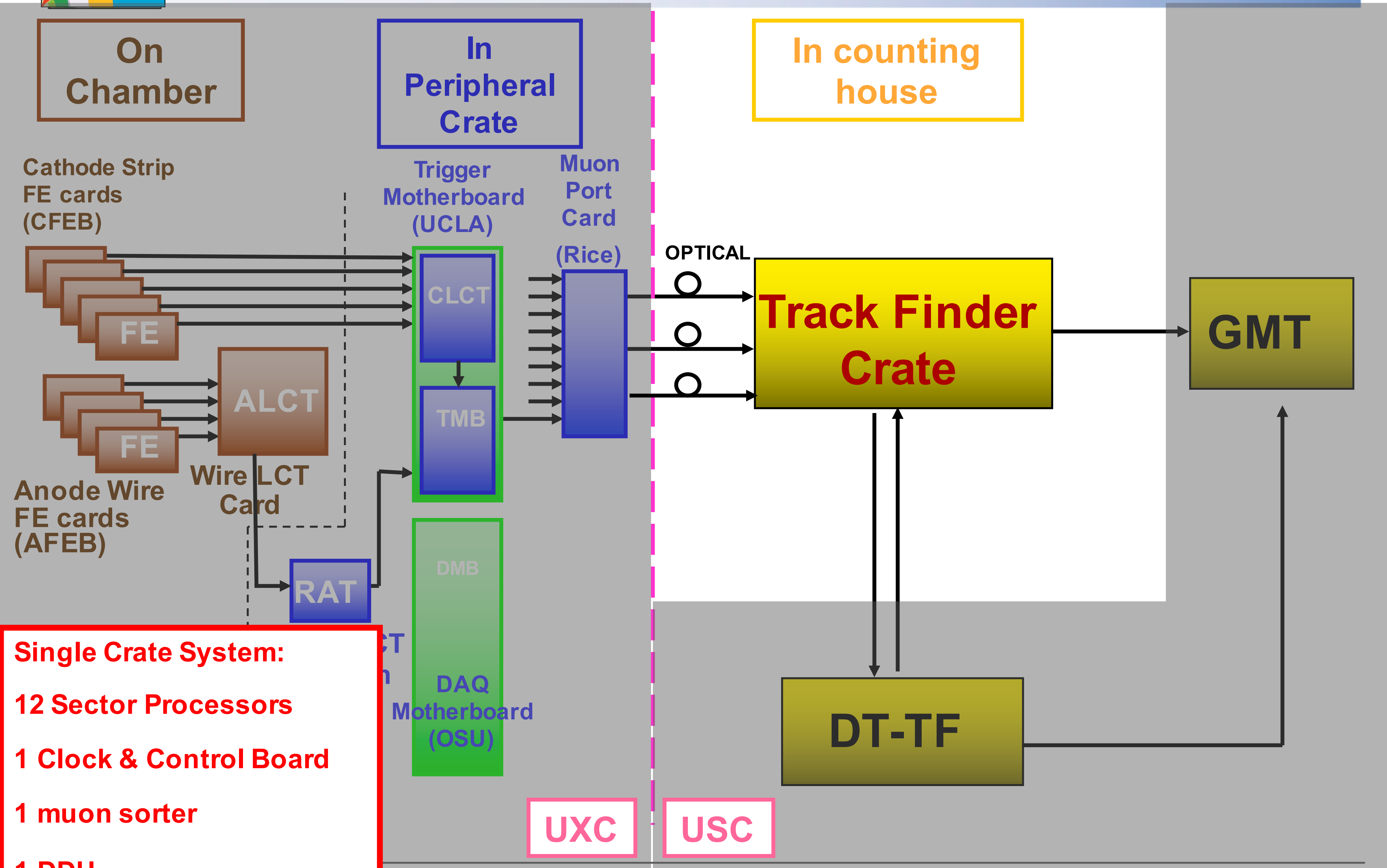


- Two Physical Levels
  - Investment in:
    - Bandwidth
    - Commercial Processors




我们的数据是从哪里来的?
实验数据流
Why Trigger
整体概念和数字
L1 硬件触发
HLT与DAQ体系结构
监测、存储、传输、显示

CMS CSC Track Finder



Single Crate System:

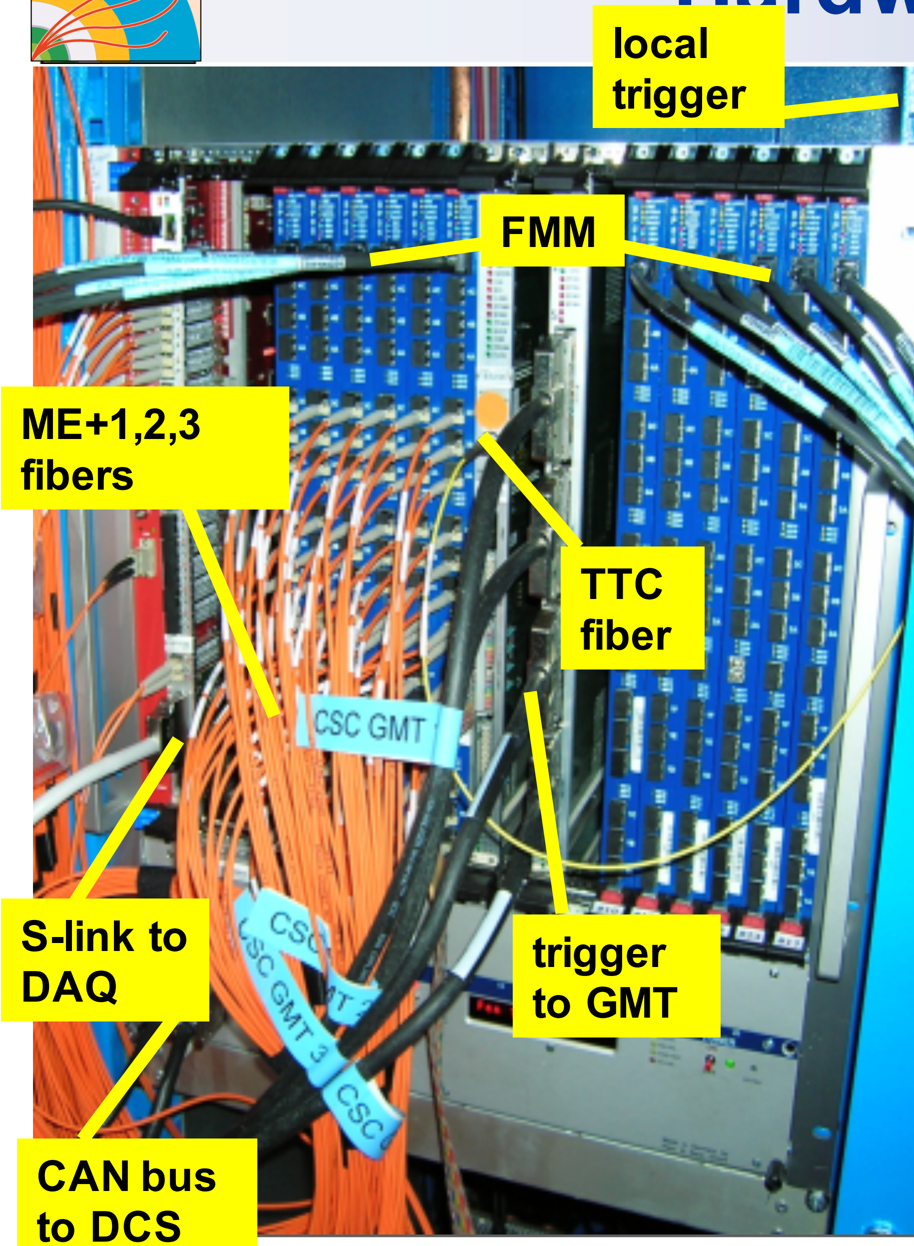
12 Sector Processors

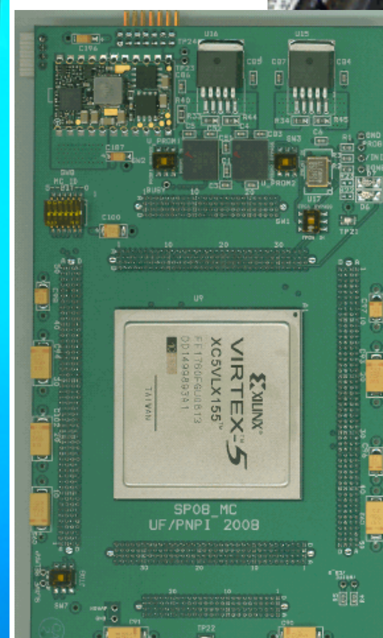
1 Clock & Control Board

1 muon sorter

1 DDU

Hardware Layout





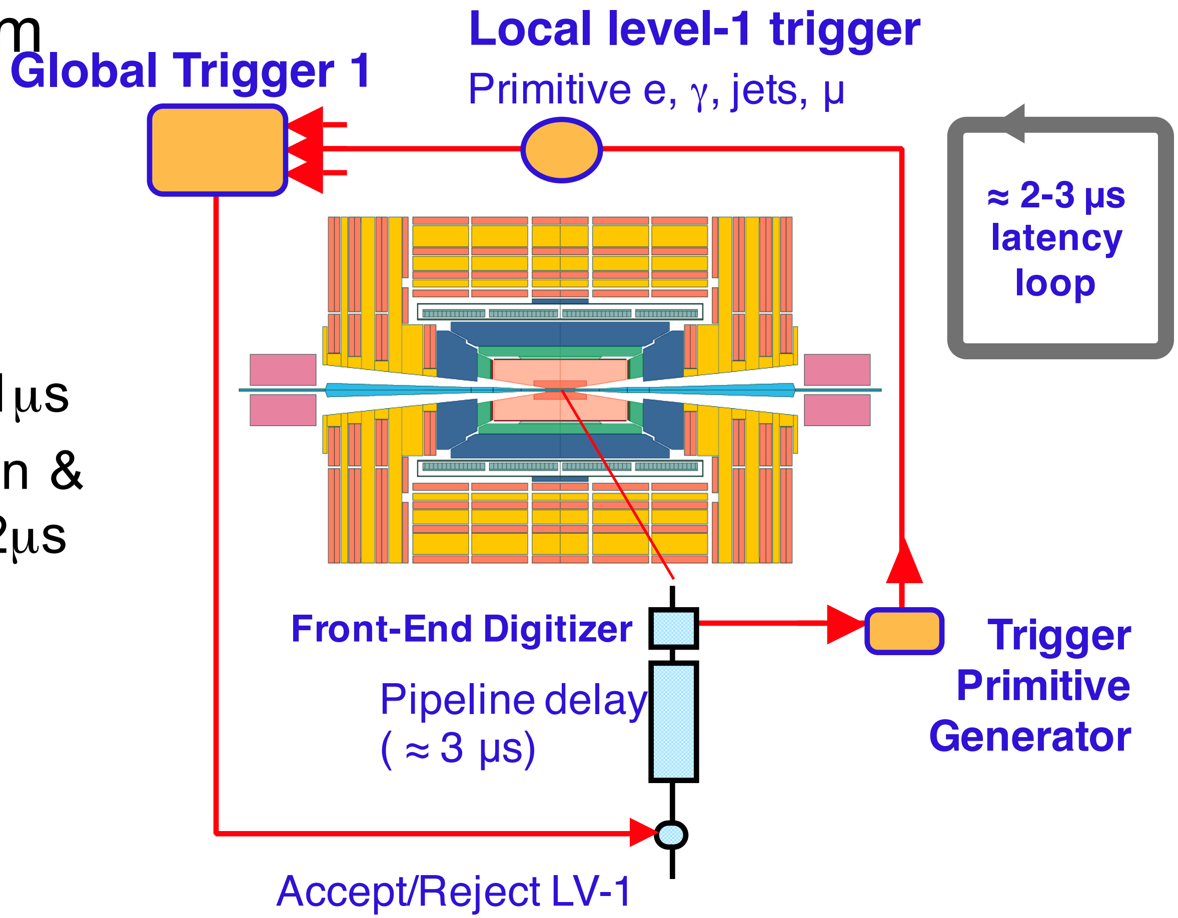
CSCTF transition board and connection cables to DTTF with Channel-Link LVDS transmission

Technologies in Level-1 systems

- ASIC(Application-Specific Integrated Circuit) used in some cases
  - Highest-performance option, better radiation tolerance and lower power consumption (a plus for on-detector electronics)
- FPGA(Field-Programmable Gate Array) used in all systems
  - Impressive evolution with time. Large gate counts and operating at 40 MHz (and beyond)
  - Biggest advantage: flexibility
    - Can modify algorithms (and their parameters) in situ
- Communication technologies
  - High-speed serial links (copper or fiber)
    - LVDS up to 10 m and 400 Mb/s; HP G-link, Vitesse for longer distances and Gb/s transmission
  - Backplanes
    - Very large number of connections, multiplexing data
      - operating at ~160 Mb/s
  - High speed optical links (fibers)
    - Up to 10Gb/s per link

Level-1 Trigger: decision loop

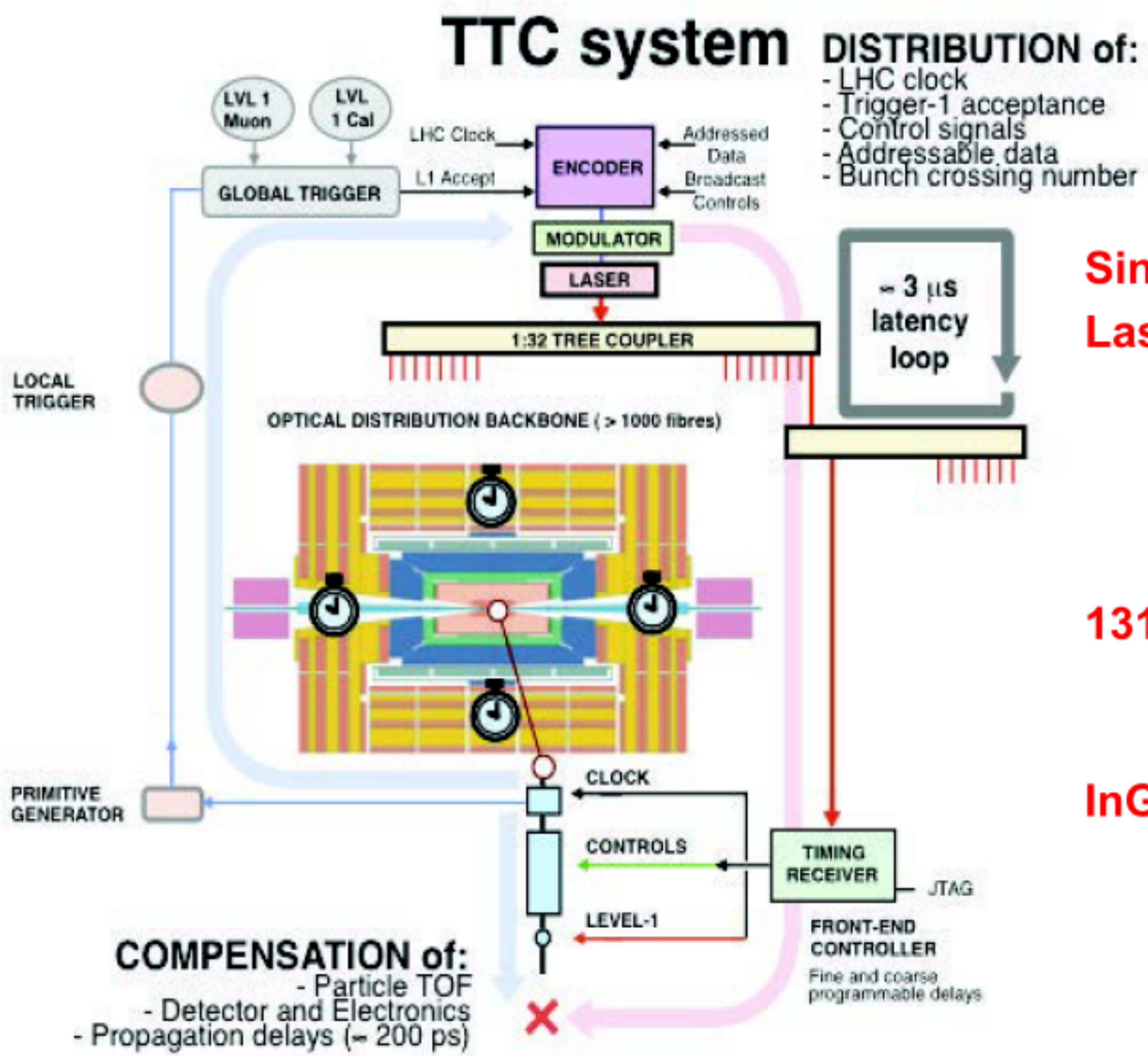
- Synchronous 40 MHz digital system
  - Typical: 160 MHz internal pipeline
  - Latencies:
    - Readout + processing: < 1μs
    - Signal collection & distribution: ≈ 2μs
- At Lvl-1: process only calo+μ info



Global Trigger

- A very large OR-AND network that allows for the specification of complex conditions:
  - 1 electron with P_T>20 GeV OR 2 electrons with P_T>14 GeV OR 1 electron with P_T>16 and one jet with P_T>40 GeV…
  - The top-level logic requirements (e.g. 2 electrons) constitute the "trigger-table" of the experiment
    - Allocating this rate is a complex process that involves the optimization of physics efficiencies vs backgrounds, rates and machine conditions

Trigger timing and control



Optical System:

Single High-Power Laser per zone

- **Reliability, transmitter upgrades**
- **Passive optical coupler fanout**

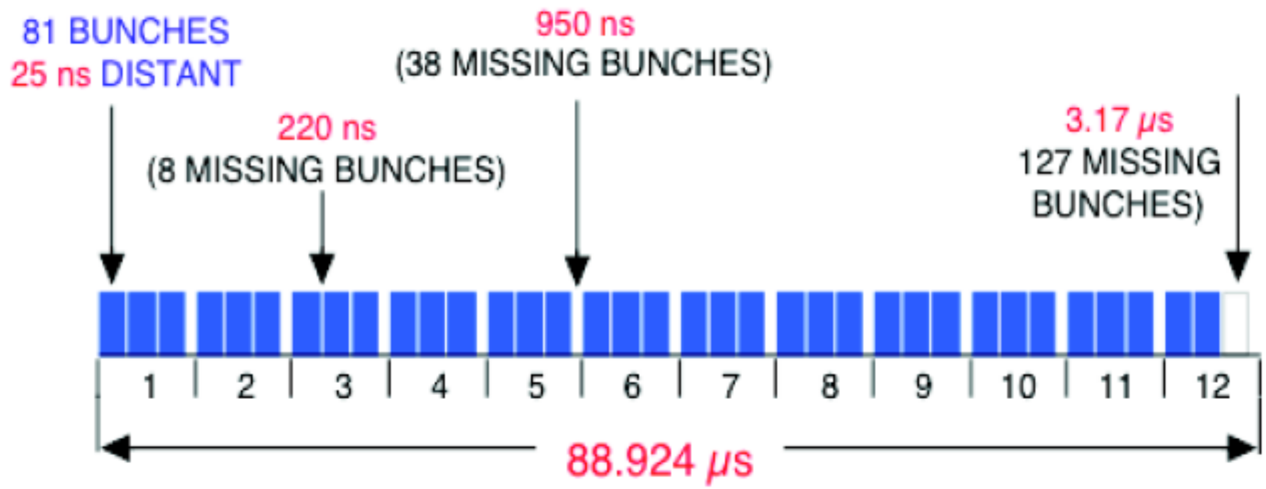
1310 nm Operation

- **Negligible chromatic dispersion**

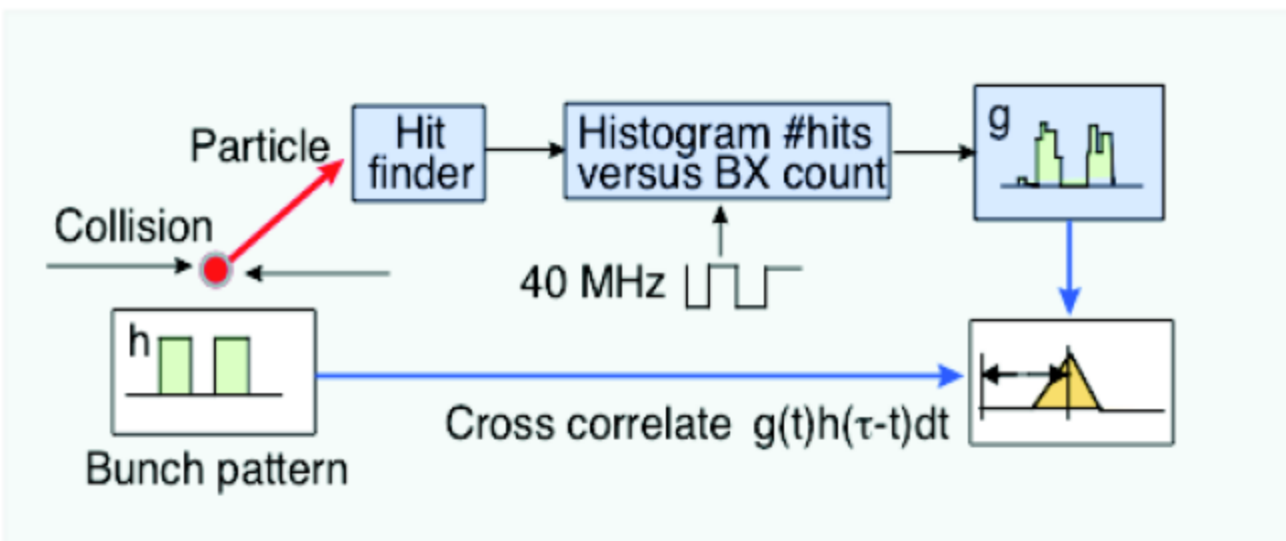
InGaAs photodiodes

- **Radiation resistance, low bias**

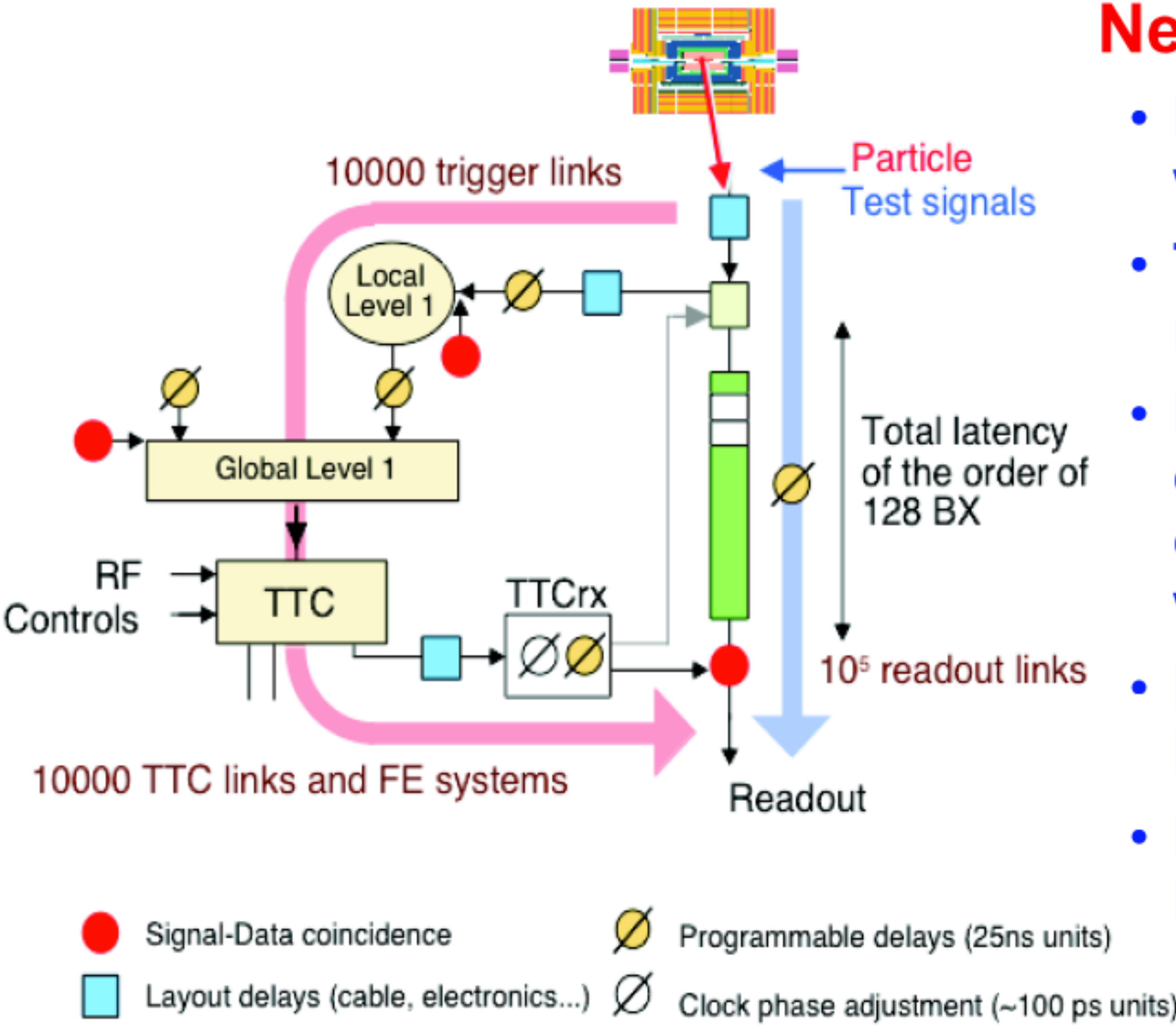
Synchronization



2835 out of 3564 p bunches are full, use this pattern:



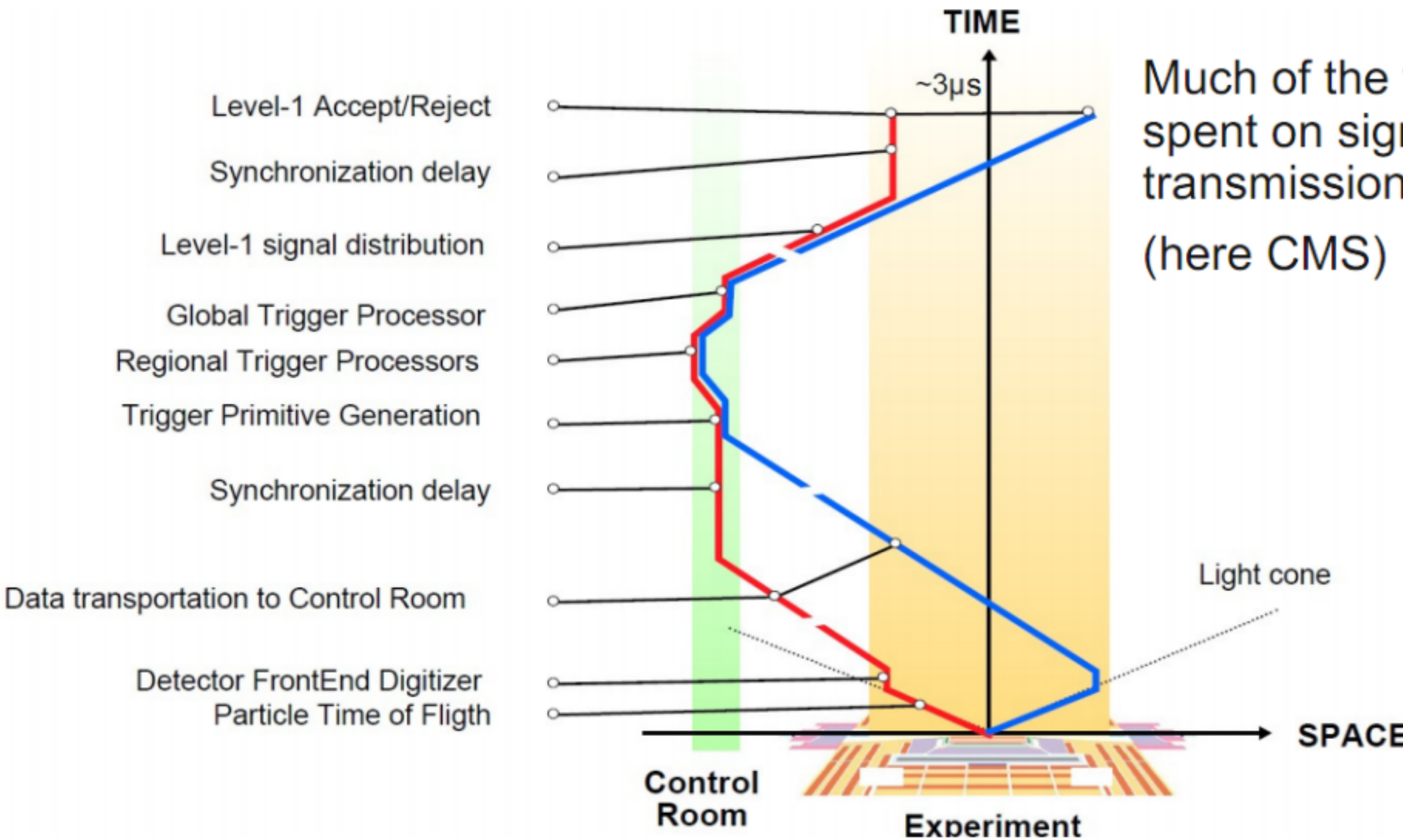
Align the signal



Need to Align:

- **Detector pulse w/collision at IP**
- **Trigger data w/ readout data**
- **Different detector trigger data w/each other**
- **Bunch Crossing Number**
- **Level 1 Accept Number**

Trigger Latency

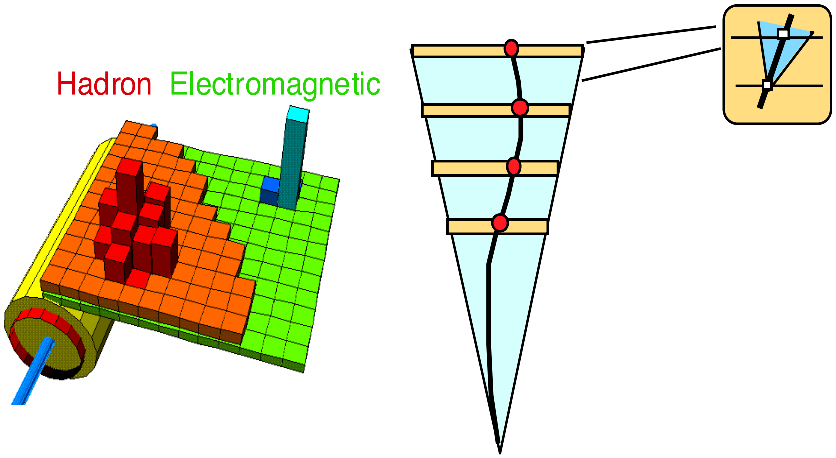


Much of the time spent on signal transmission

(here CMS)

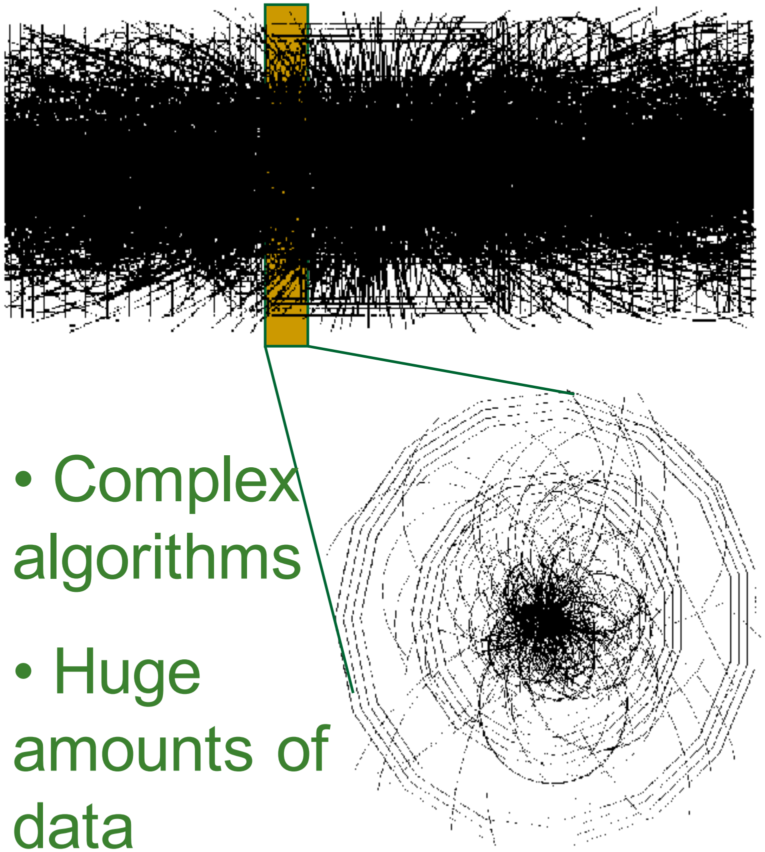
At Level-1: only calo and muon info

- Pattern recognition much faster/easier



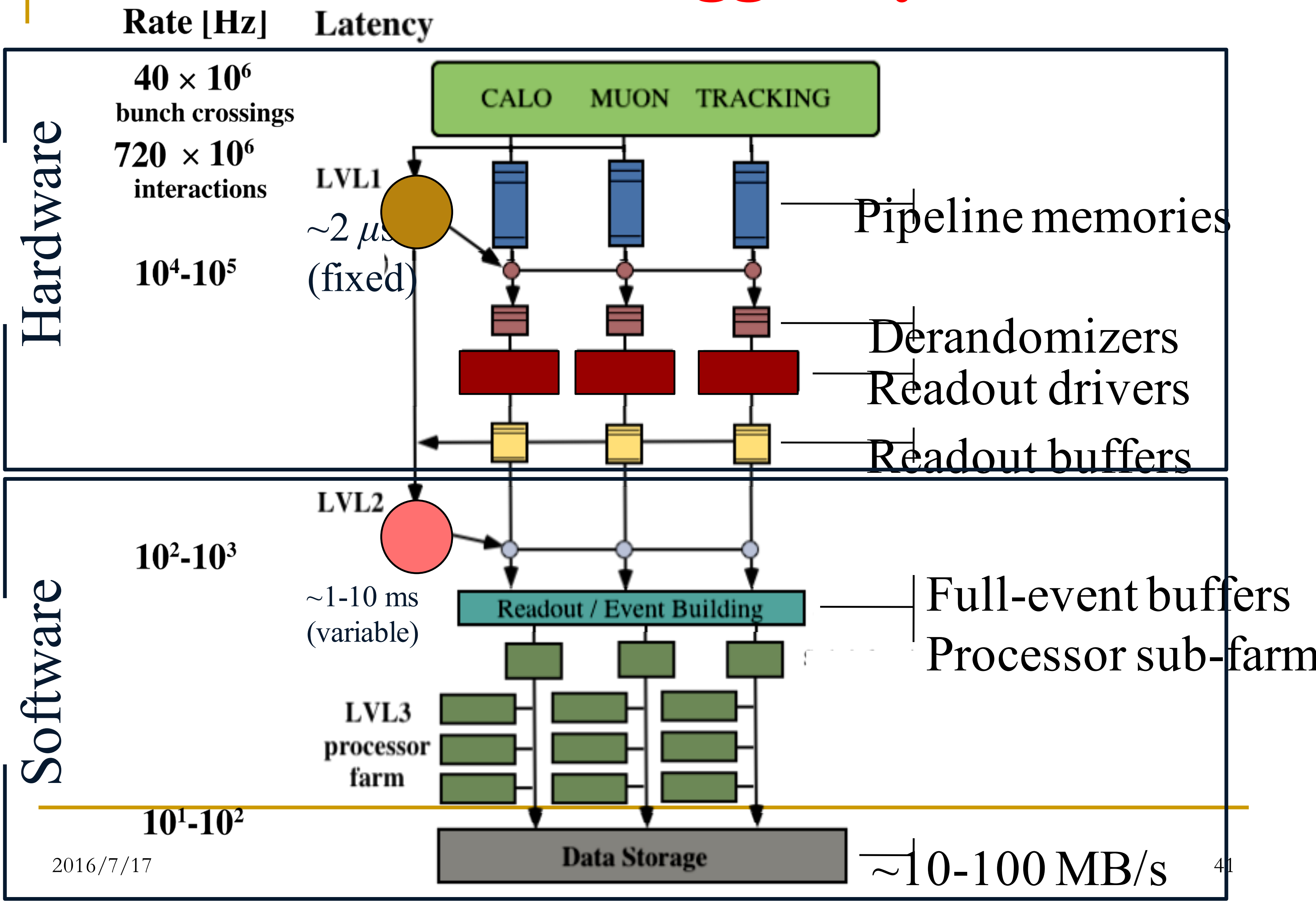
- **Simple algorithms**
- **Small amounts of data**
- **Local decisions**

- Compare to tracker info

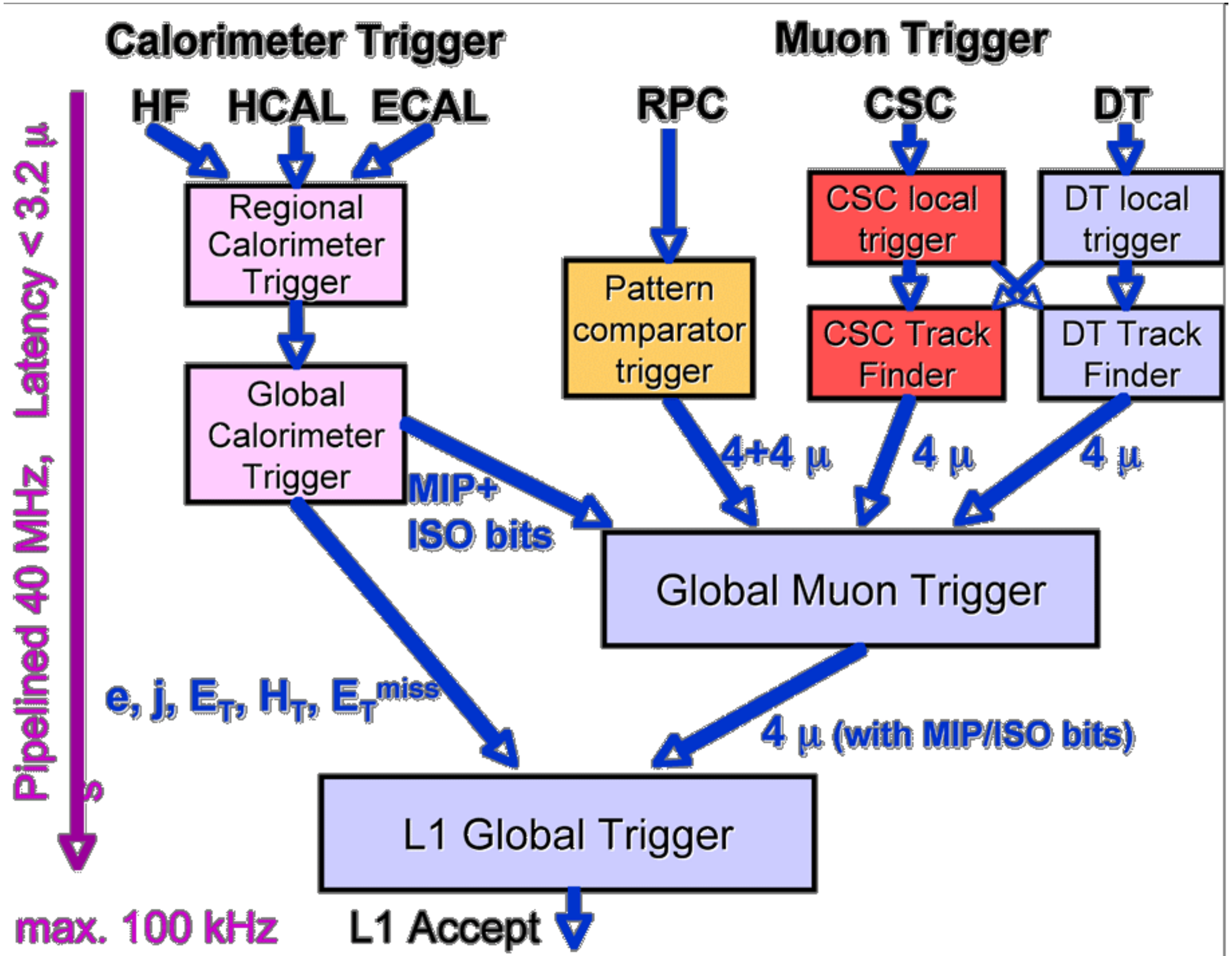


- Complex algorithms
- Huge amounts of data
- Need to link sub-detectors

The ATLAS Trigger System

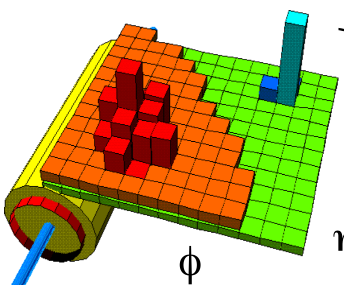


Overview of CMS L1 Trigger



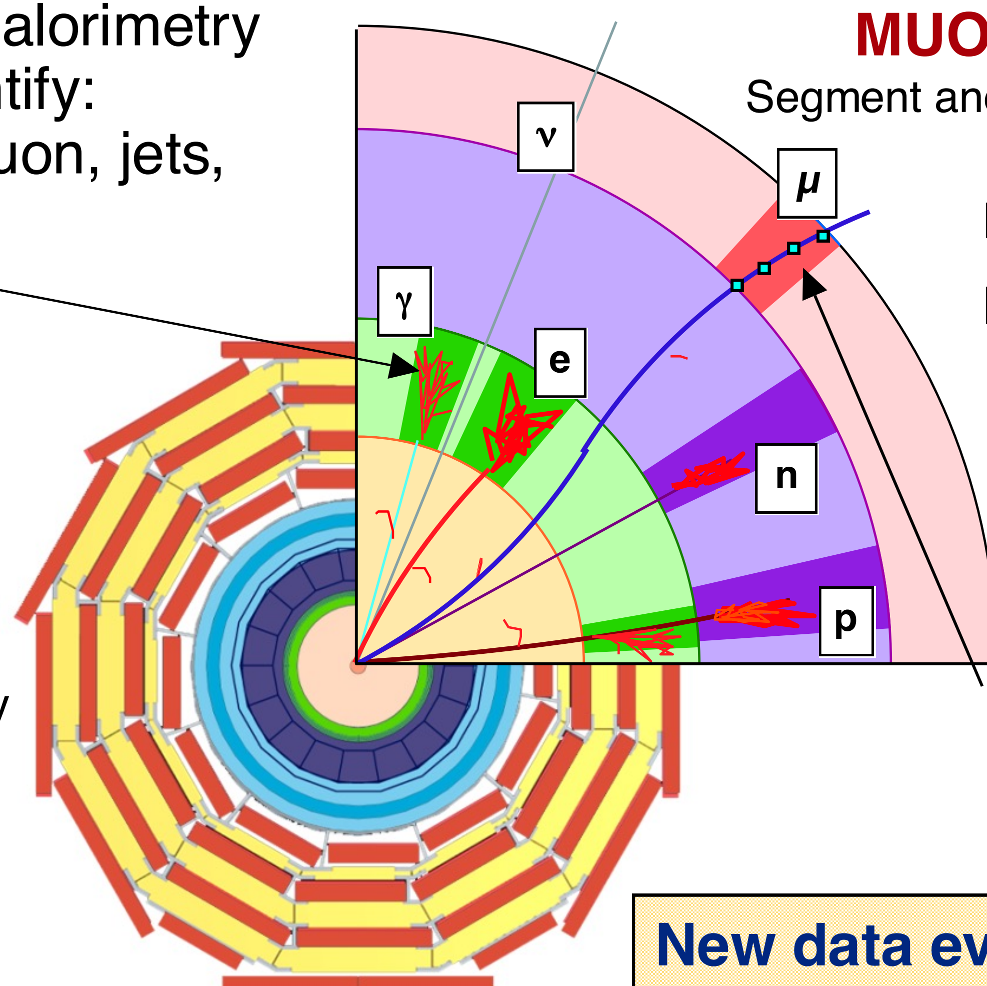
Trigger alg: based on Particle signatures

Use prompt data (calorimetry and muons) to identify:
High p_t electron, muon, jets, missing E_T



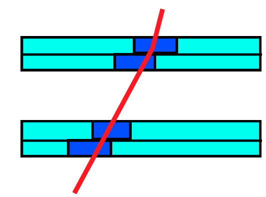
CALORIMETERs

Cluster finding and energy deposition evaluation



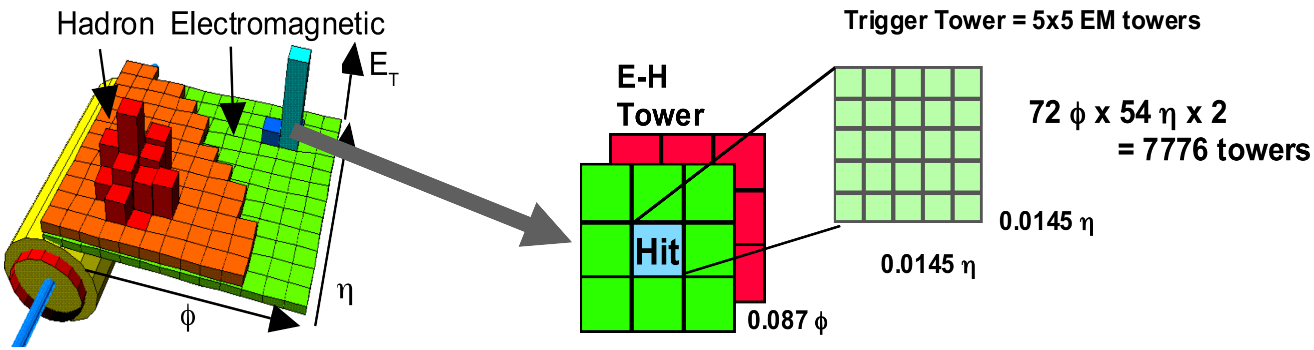
MUON System

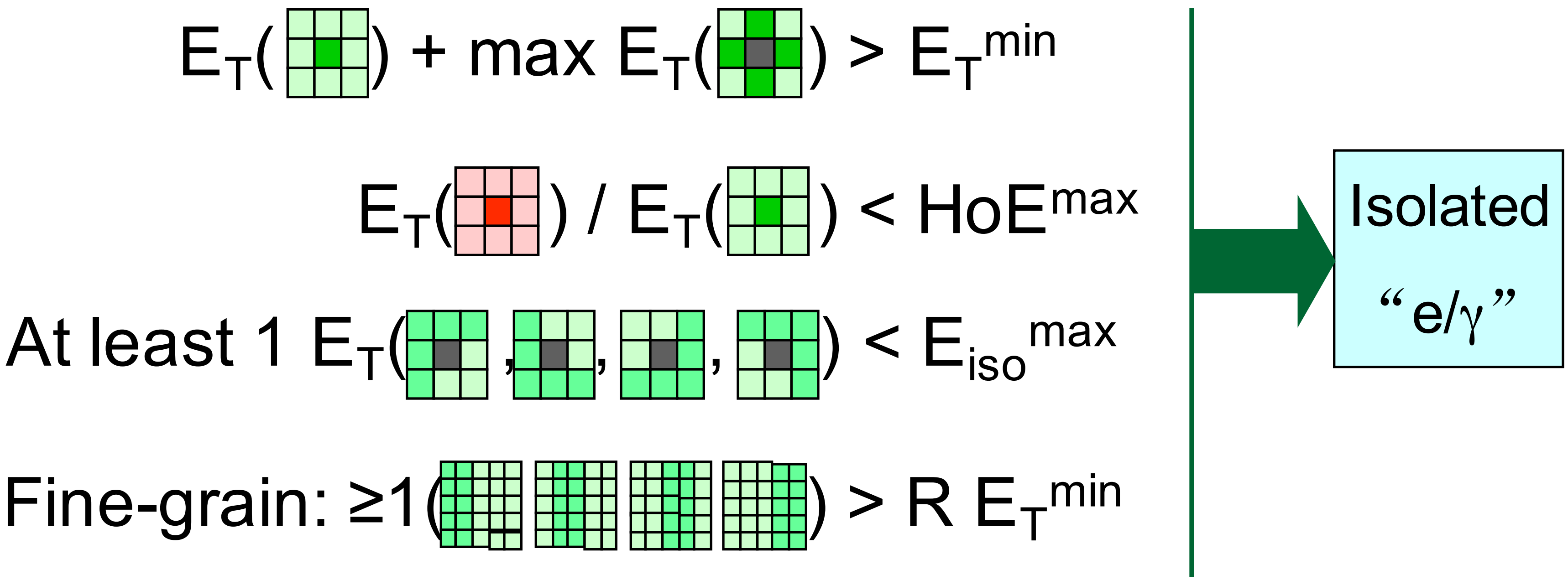
Segment and track finding



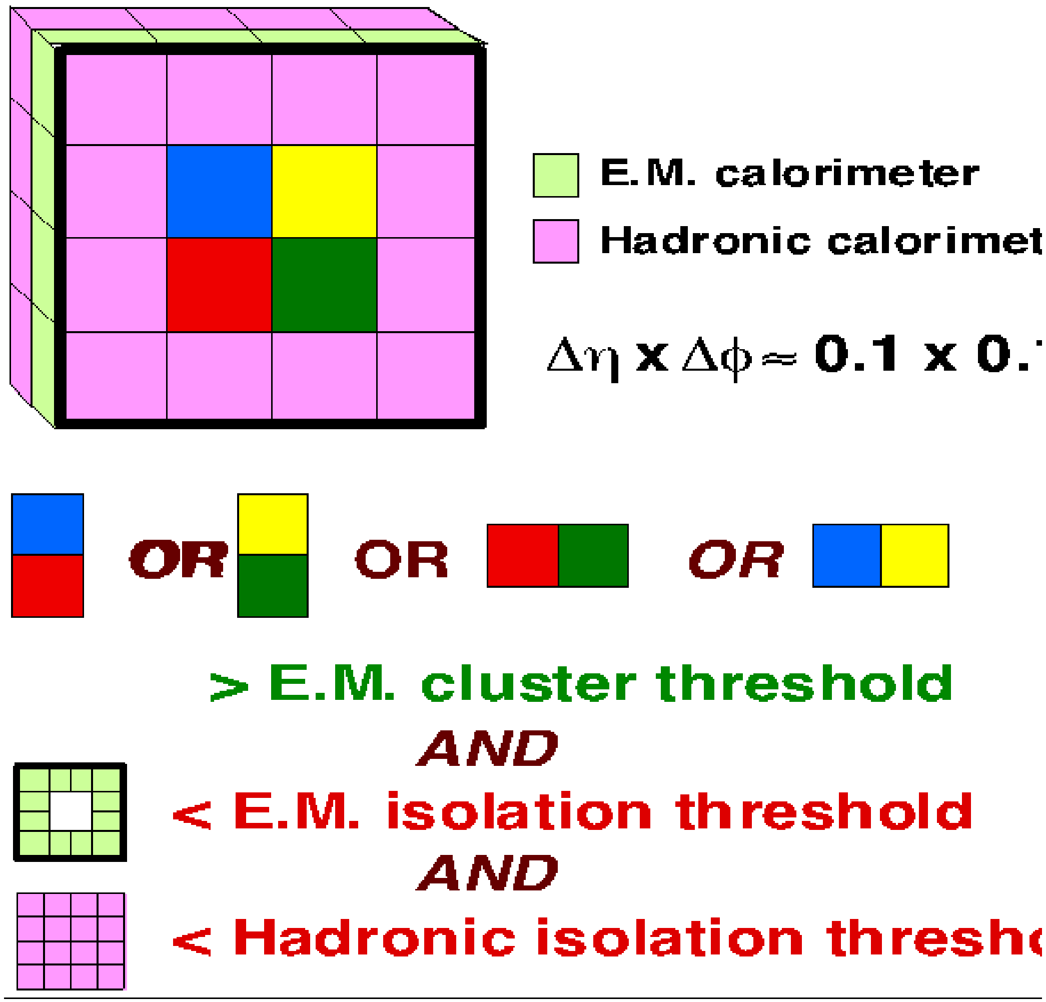
New data every 25 ns
Decision latency ~ μs

Lvl-1 Calo Trigger: e/γ algorithm (CMS)





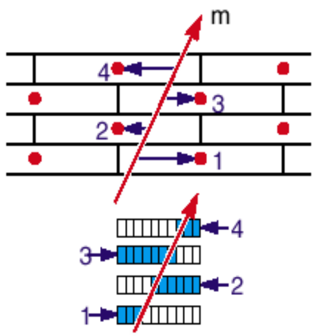
ATLAS em cluster trigger algorithm



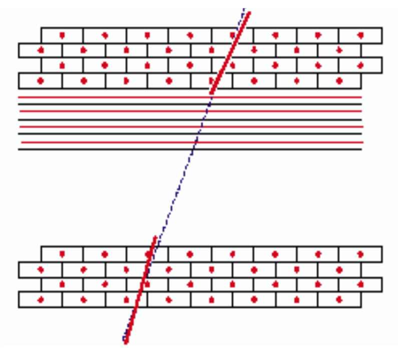
"Sliding window" algorithm repeated for each of ~4000 cells

Lvl-1 muon trigger (CMS)

Drift Tubes

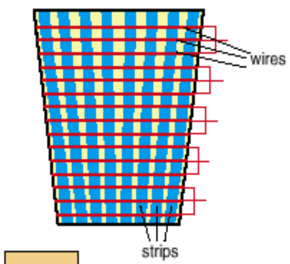


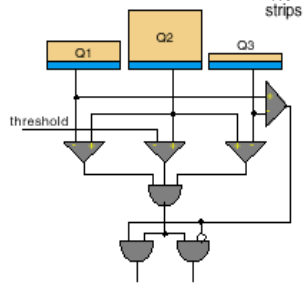
Meantimers recognize tracks and form vector / quartet.



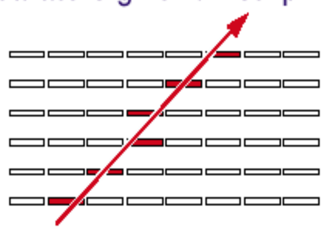
Correlator combines them into one vector / station.

CSC

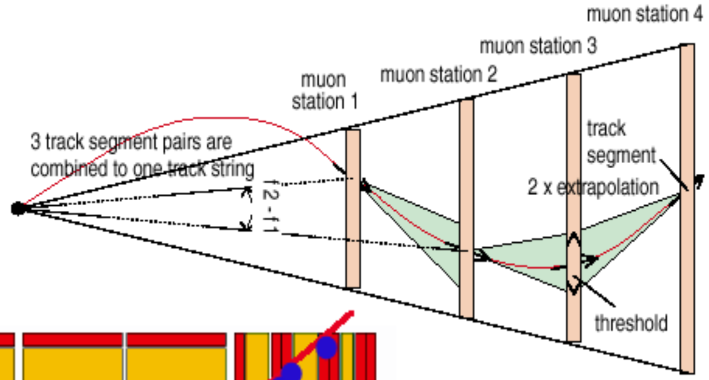


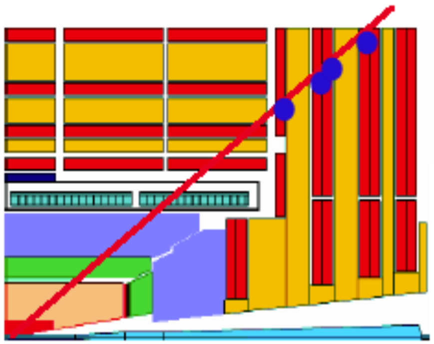


Comparators give 1/2-strip resol.



Hit strips of 6 layers form a vector.

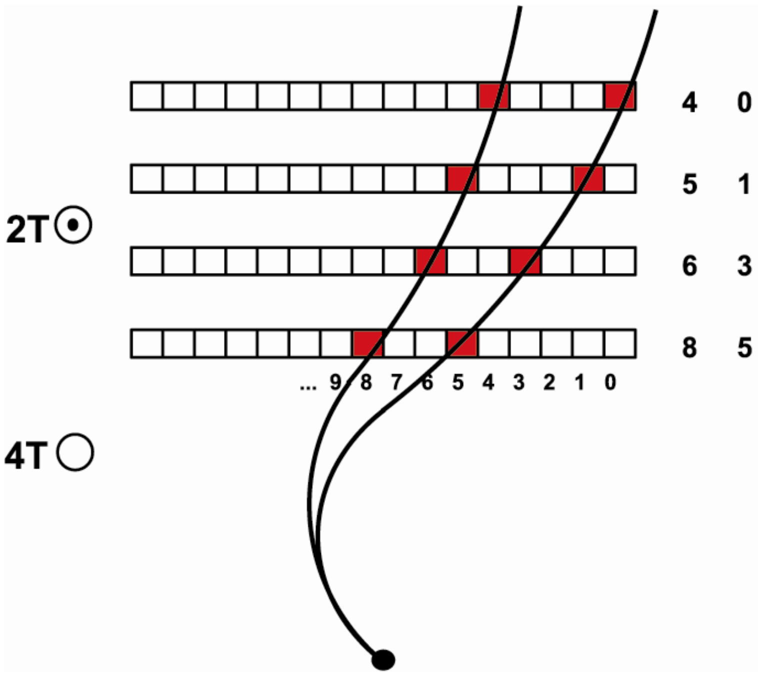


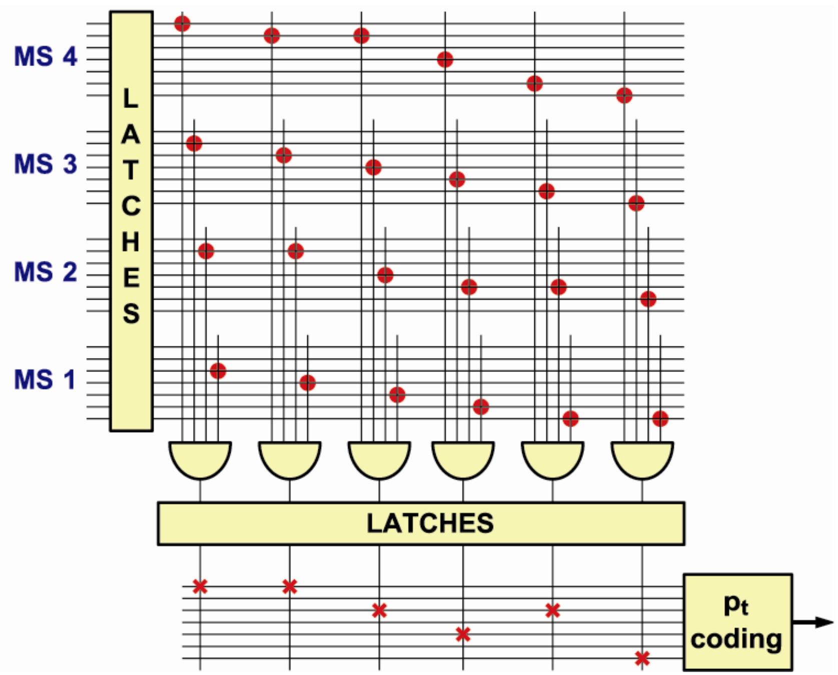


- Extrapolation: using look-up tables
- Track Assembler: link track segment-pairs to tracks, cancel fakes
- Assignment: P_T (5 bits), charge, η (6 bits) , φ(8 bits), quality (3 bits)

Hardware implementation:
ASICs for Trigger Primitive Generators
FPGAs for Track Finder processors

Lvl-1 muon trigger (CMS)





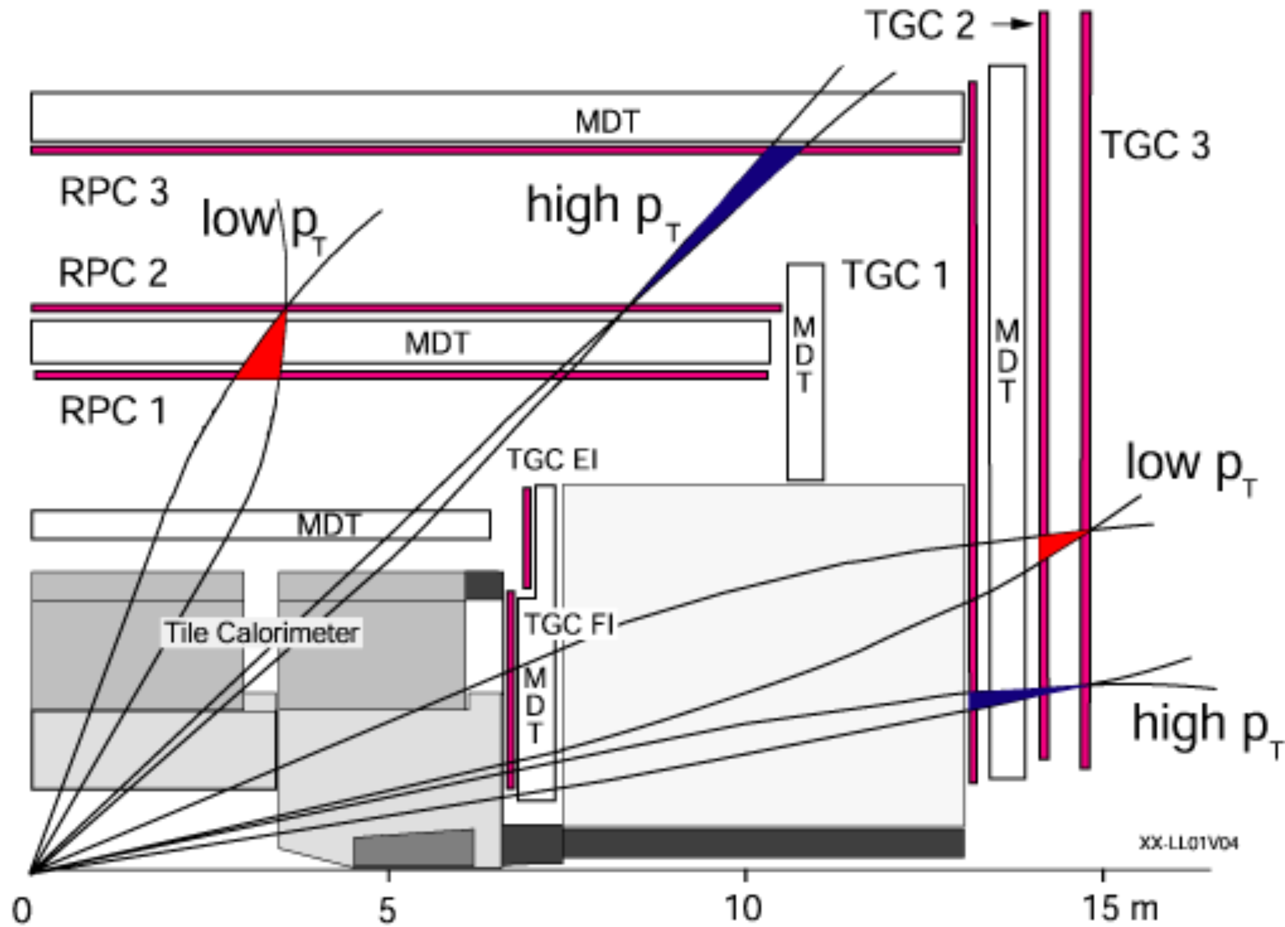
Pattern of strips hit:
Compared to predefined patterns corresponding to various p_T

Implemented in FPGAs

The LVL1 Muon Trigger（ATLAS）

- Safe Bunch Crossing Identification
- Wide p_T-threshold range
- Strong rejection of fake muons (induced by noise and physics background)

→Fast and high redundancy system

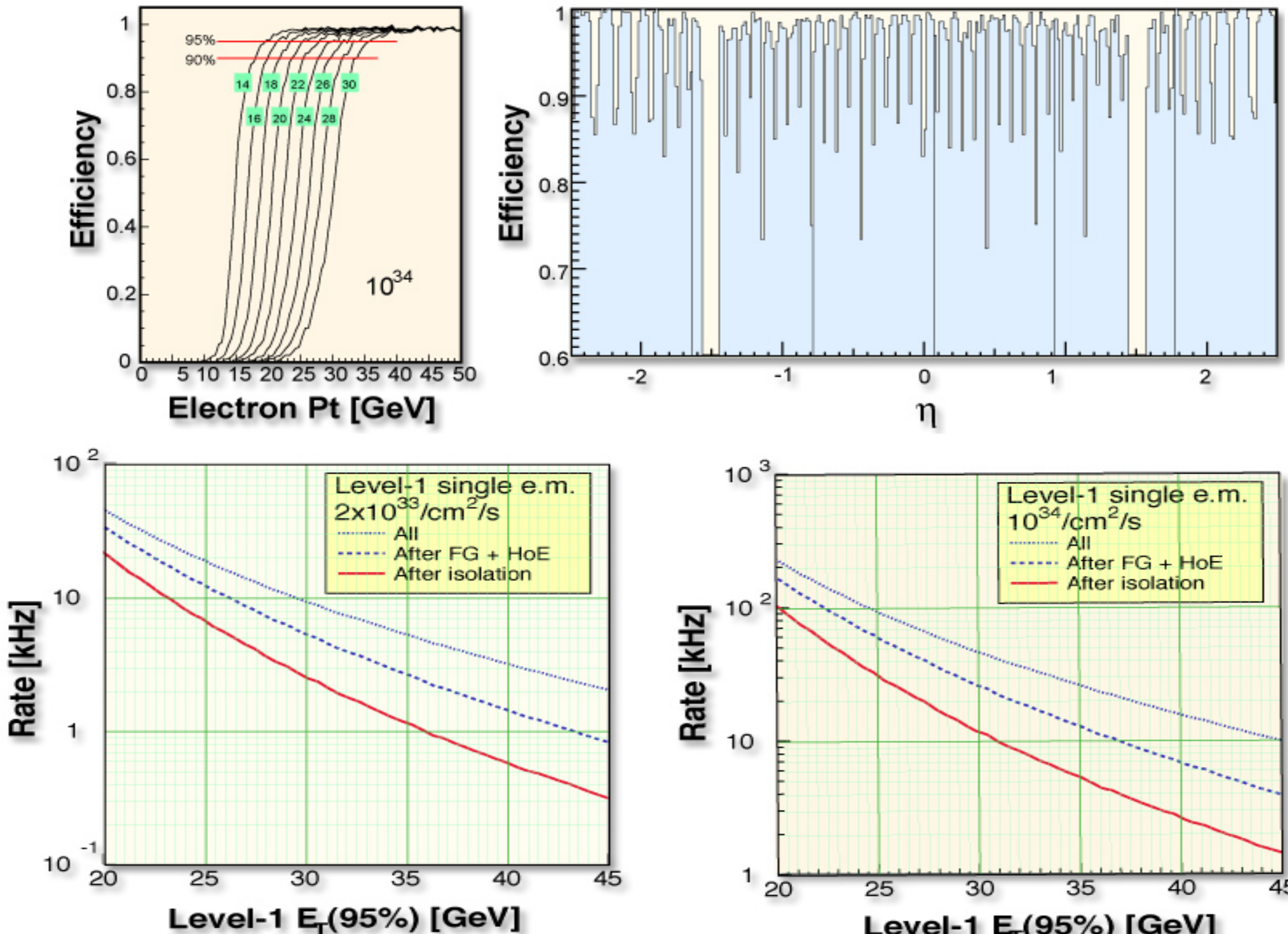


However this system:

1. Looks only for tracks coming from the pp collision point
2. Looks only for ultrarelativistic tracks

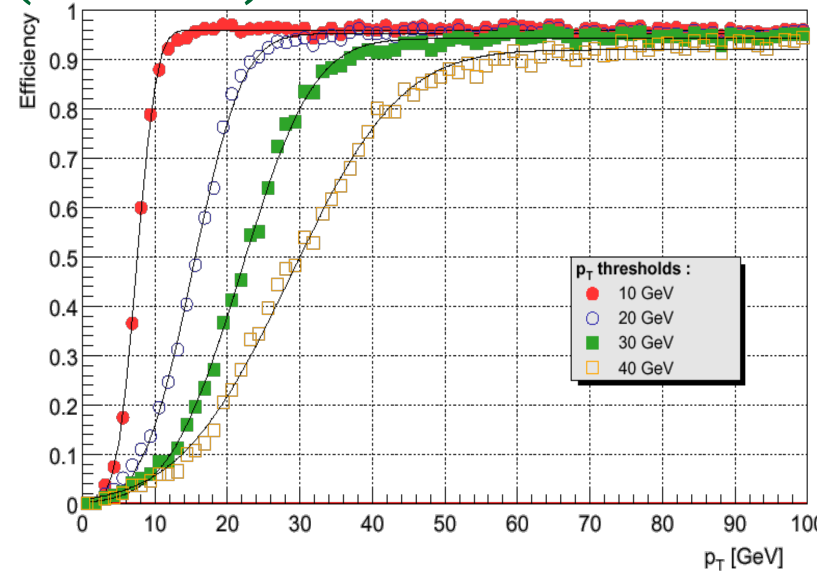
Lvl-1 Calo e/γ trigger: performance

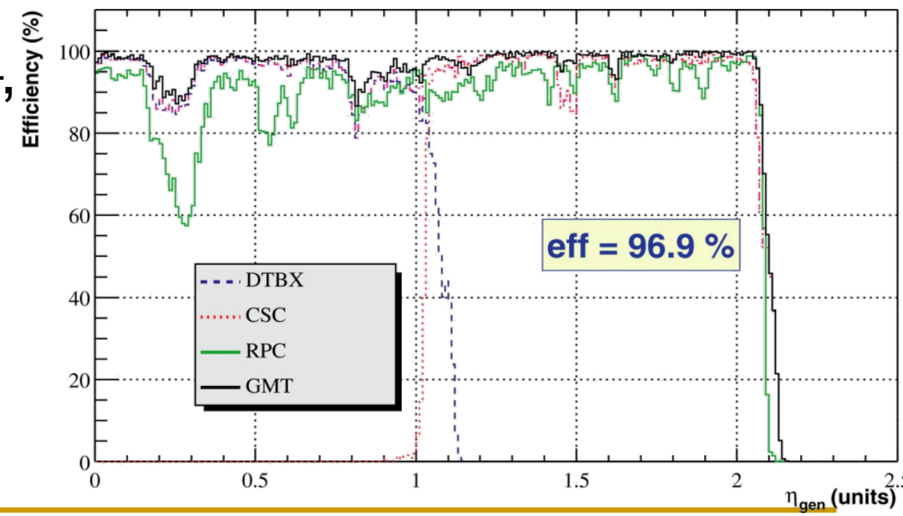
- Efficiencies and Trigger Rates

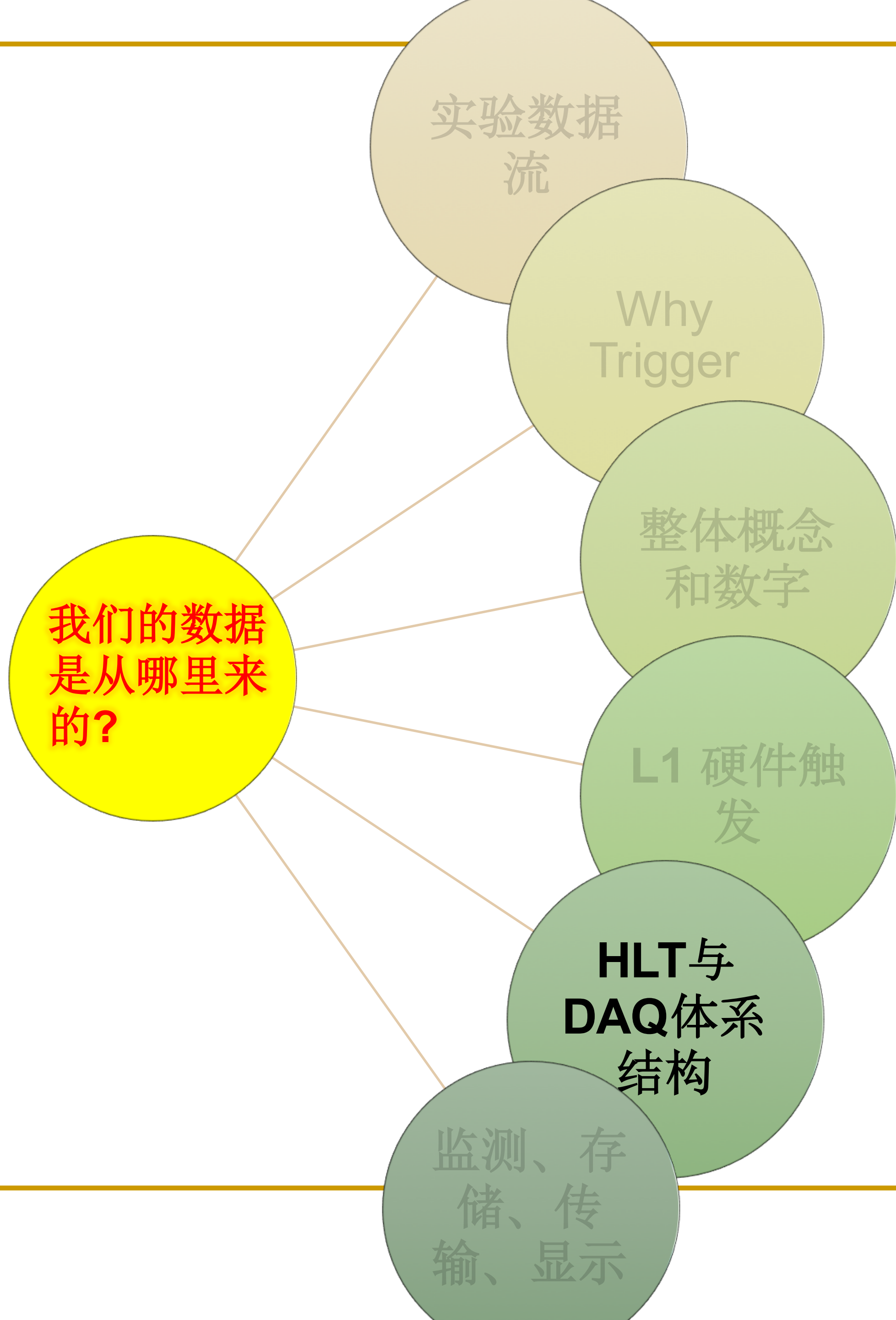


Global muon trigger (CMS)

- Combine results from RPC, CSC and DT triggers
- Match muon candidates from different trigger systems; use complementarity of detectors
- improve efficiency and rate
- assign muon isolation
- deliver the 4 best (highest P_T, highest-quality) muons to Global Trigger
- Pt resolution:
  - 18% barrel
  - 35% endcaps
- Efficiency: ~ 97%



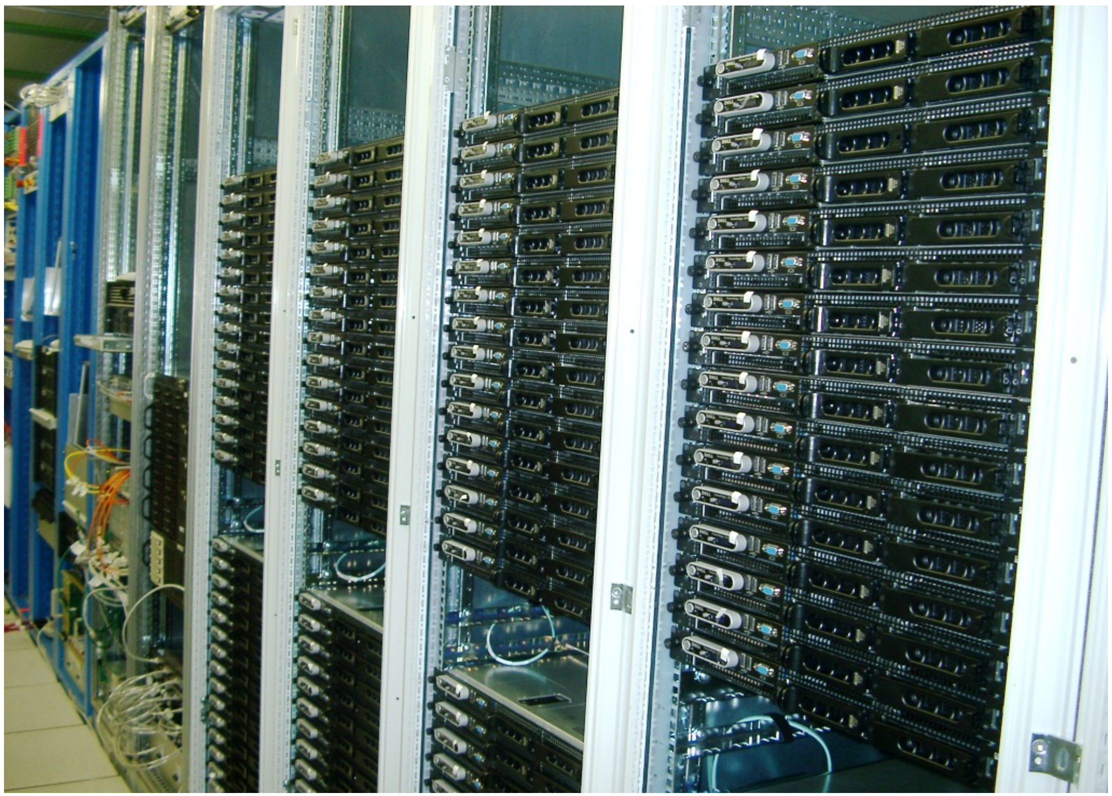


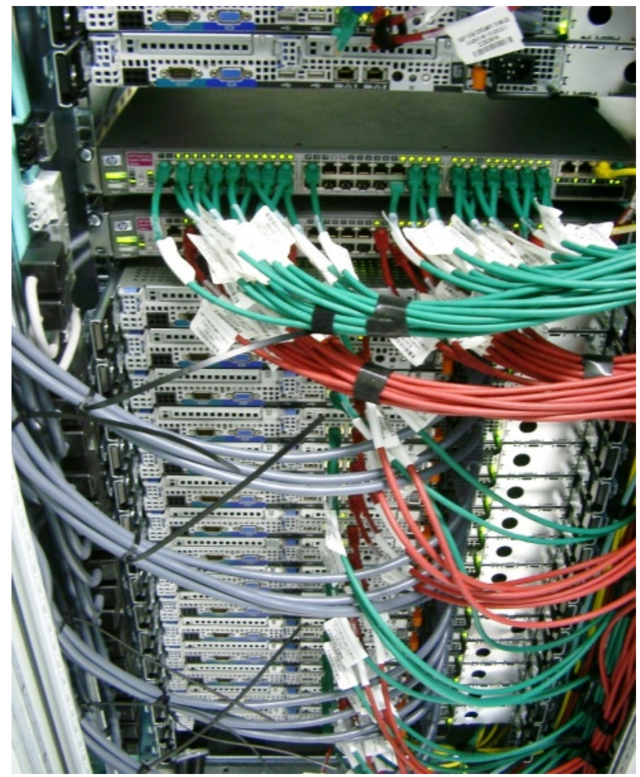

我们的数据是从哪里来的?
实验数据流
Why Trigger
整体概念和数字
L1 硬件触发
HLT与DAQ体系结构
监测、存储、传输、显示

HLT/DAQ Hardware(ATLAS)

First 4 racks of HLT processors, each rack contains

- ~30 HLT PC's (PC's very similar to Tier-0/1 compute nodes)
- 2 Gigabit Ethernet Switches
- a dedicated Local File Server





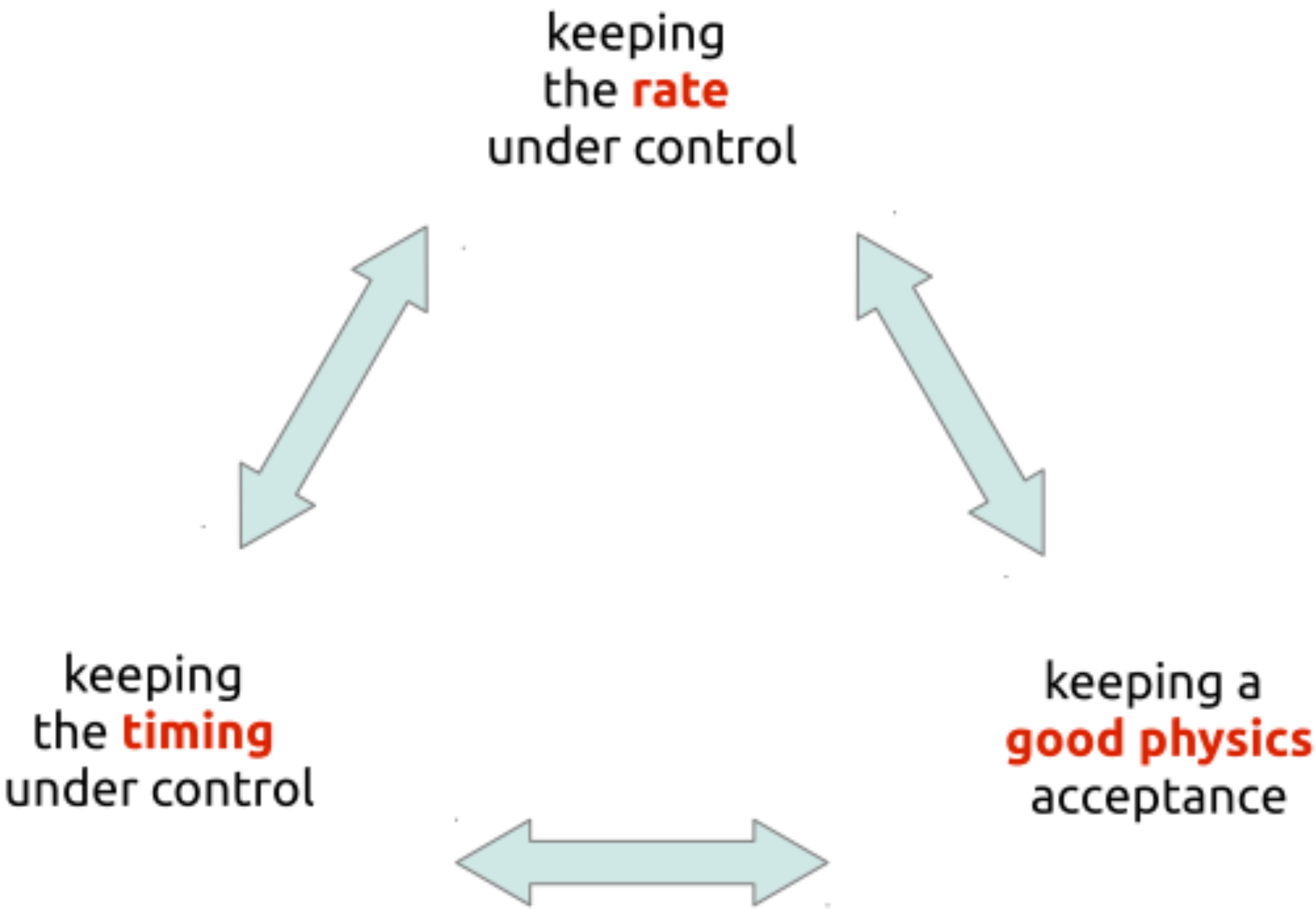
High Level Triggers (HLT)

- **Run on farm of commercial CPUs: a single processor analyzes one event at a time and comes up with a decision**
- **Has access to full granularity information**
- **Freedom to implement sophisticated reco algorithms, complex selection requirements, exclusive triggers …**

Constraints:

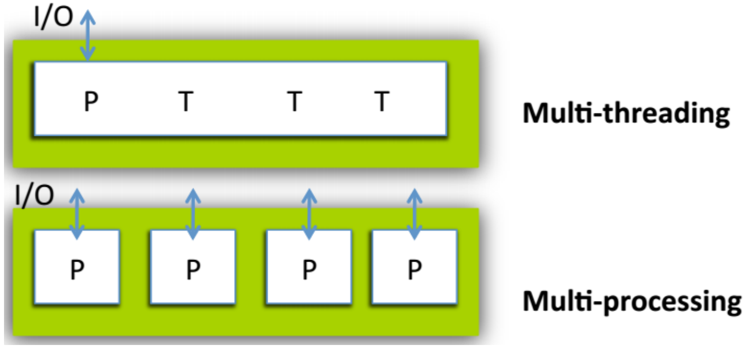
- CPU time (Cost of filter farm)
  - Reject events ASAP: set up internal “logical” selection steps
    - L2: muon+ calorimeter only
    - L3: use full information including tracking
- Must be able to measure efficiency from data
  - Use inclusive selction whenever possible
    - Single/double object above pT/ET, etc.
  - Define HLT selection paths from the L1
- Keep output rate limited (obvious…)

HLT Challenge: Compromise



HLT design principles

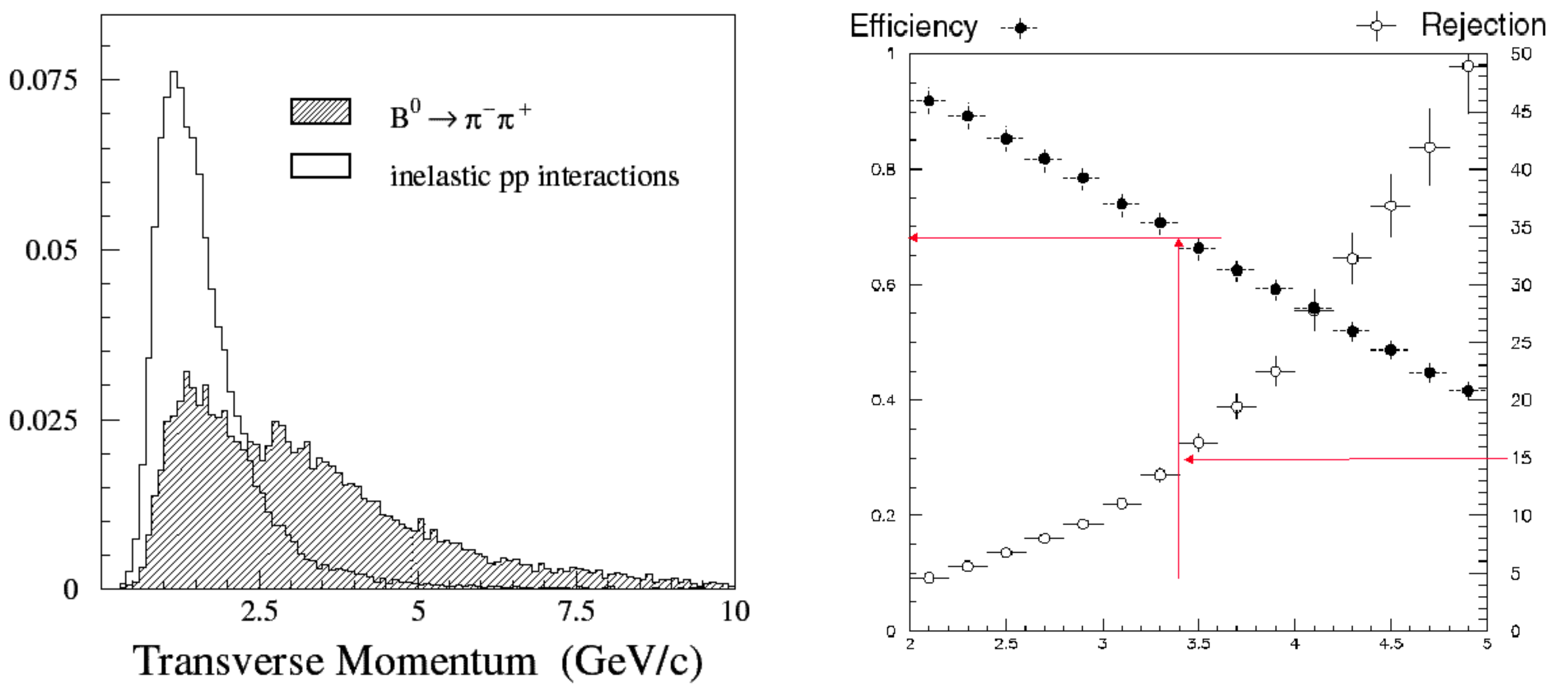
- **Early rejection**
  - Alternate steps of feature extraction with hypothesis testing: events can be rejected at any step with a complex algorithm scheduling
- **Event-level parallelism**
  - Process more events in parallel, with multiple processors
  - Multi-processing or/and multi-threading



  - Queuing of the shared memory buffer within processors
- **Algorithms are developed and optimized offline, often software is common to the offline reconstruction**

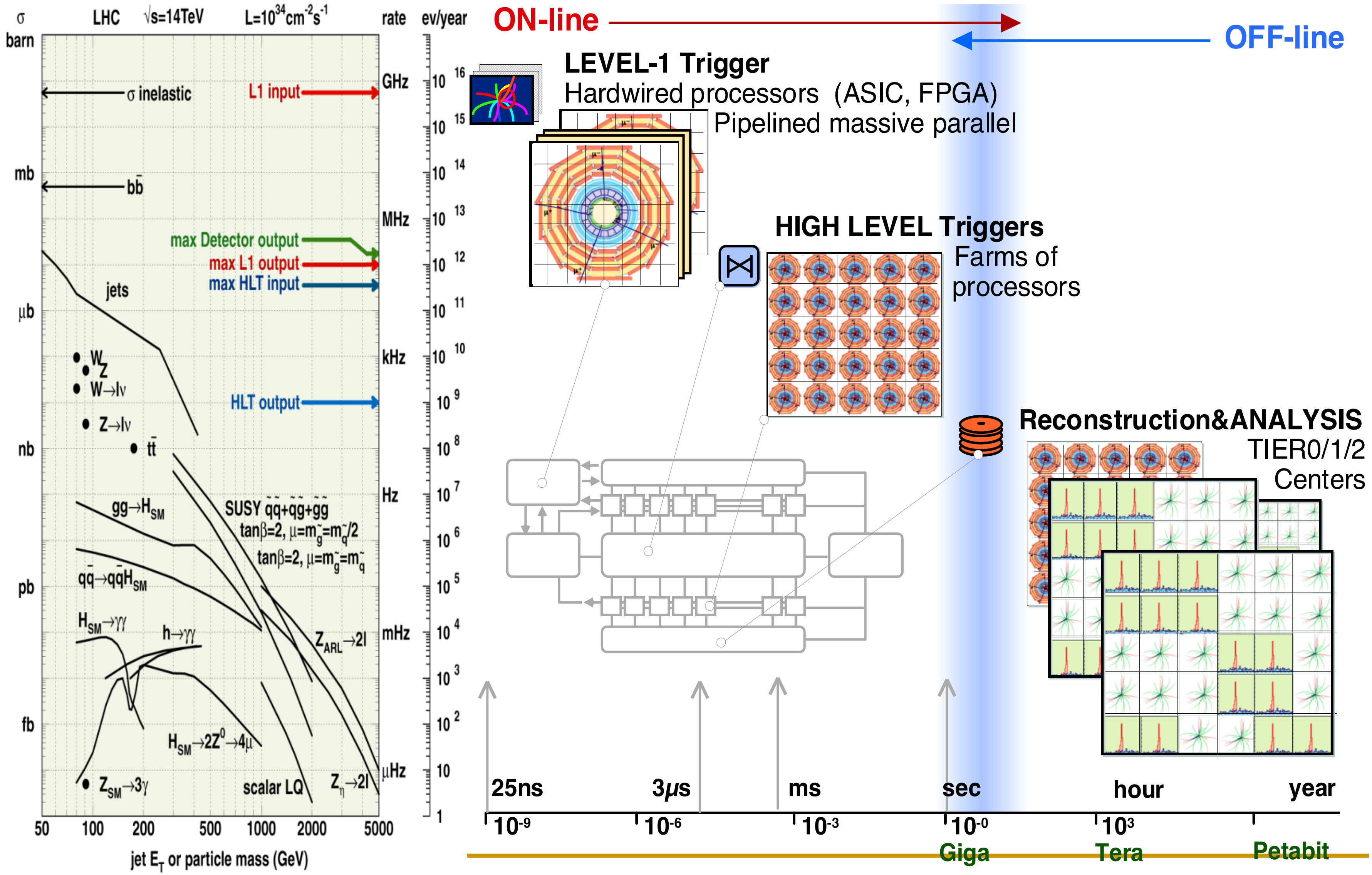
55

Trigger performance: Efficiency vs background rejection

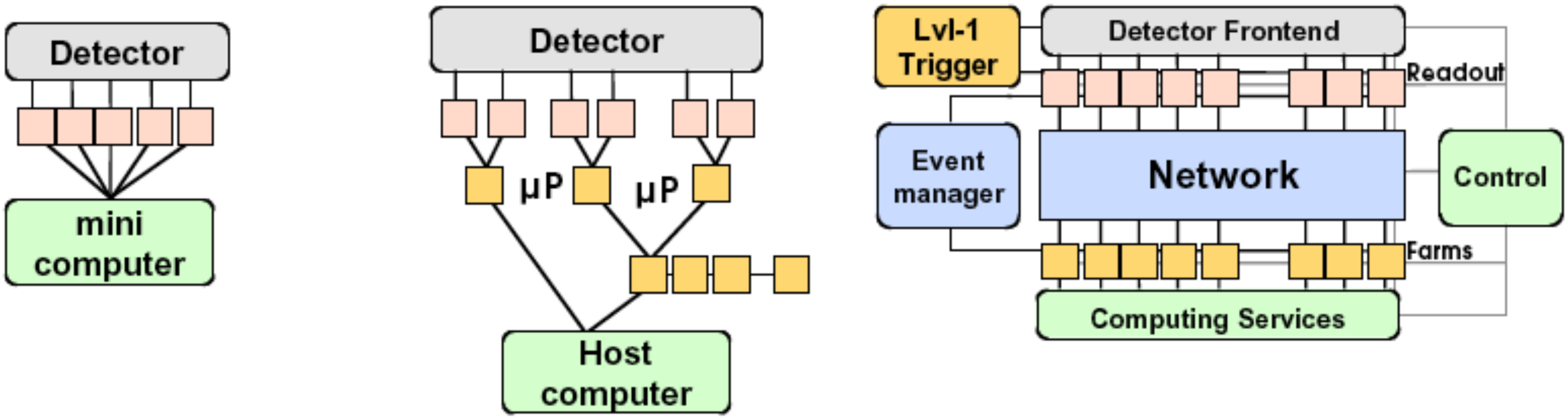


- Example:B meson trigger in LHCb
- **Discriminating variable:** Transverse momentum (P_T)

LHC实验数据流：物理信号的挑选



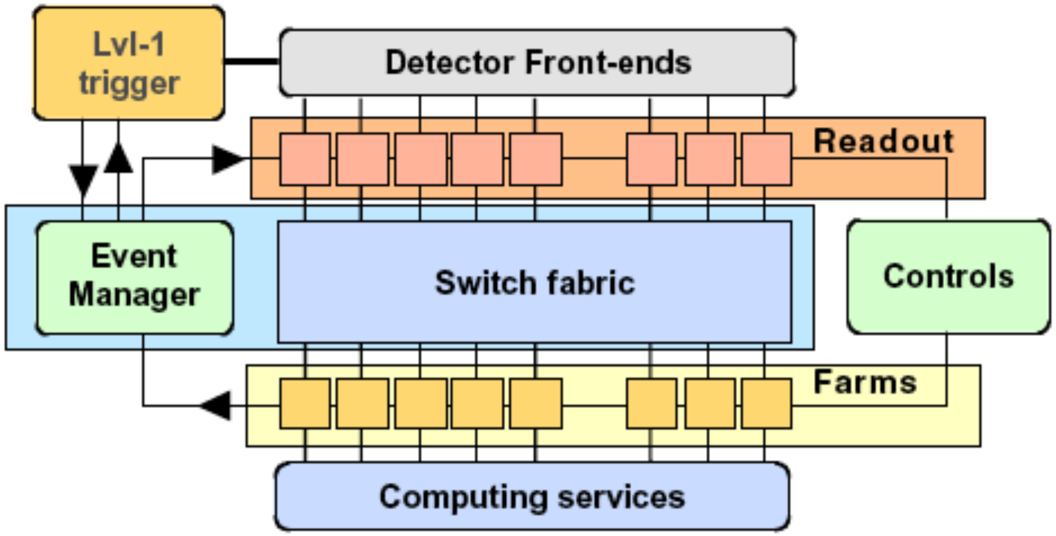
The evolution of DAQ systems



1970-80	1980-90	1990-2000+
MiniComputers	Microprocessors	Communications networks
first standard: CAMAC	Distributed systems	Control & Data networks Embedded processors
•kByte/s	•MByte/s	•GByte/s

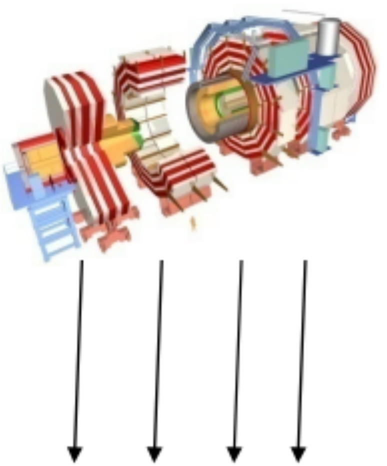
Typical architecture 2000+

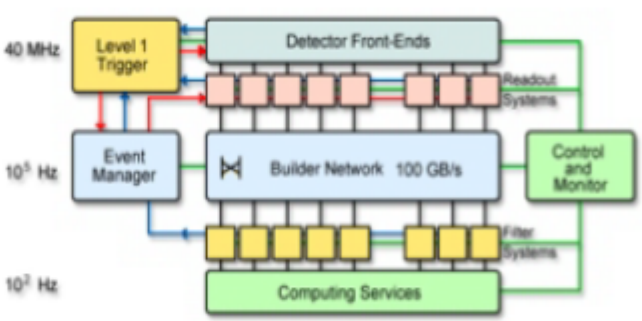
- Basic Architecture: ~ same for most experiments



- Readout (units/drivers/buffers/…)
- Switching network
- Processor Farm
- Control & Monitor System

Overview of the CMS DAQ and useful terminology





Storage Manager

- Detector signals are collected through individual data acquisition systems (cables and boards) that end up at the FEDs: the first element of Global Data Acquisition system (DAQ)

- **FED (detector FrontEnd boards):** multiple FEDs per detector collect event fragments that are sent to the online event processing farm

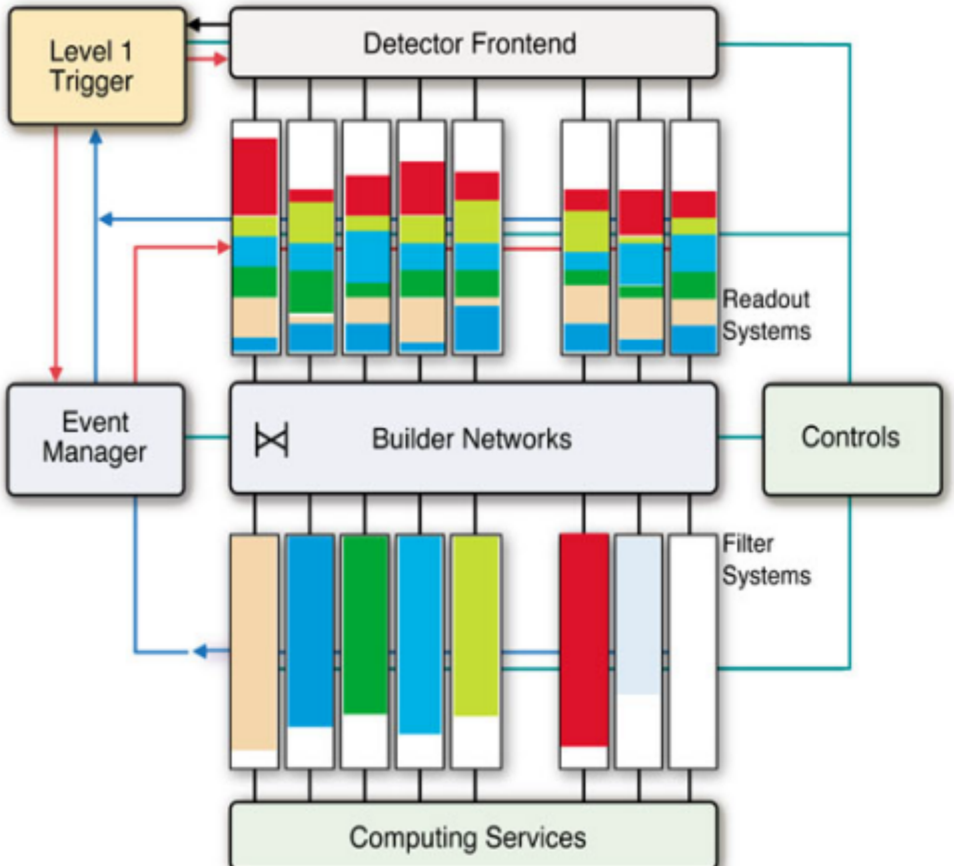
- **Builder Units:** Computing farm that collects event fragments from all FEDs and merge them to produce full event information

- **Filter Units:** Computing farm where the High Level Trigger (HLT) is run to filter interesting events

- **Storage Manager:** application that saves to local disks events selected by the HLT

Event Building

Event builder :
Physical system interconnecting data sources with data destinations. It has to move each event data fragments into a same destination



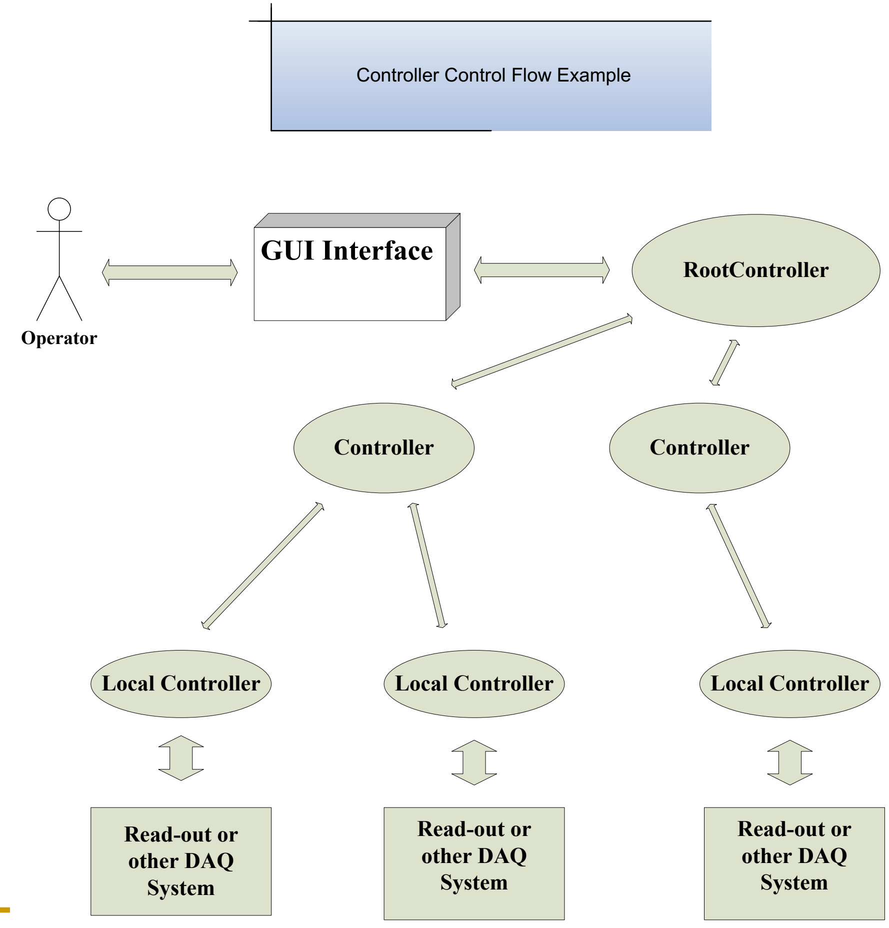
Event fragments :
Event data fragments are stored in separated physical memory systems

Full events :
Full event data are stored into one physical memory system associated to a processing unit

Hardware:
Fabric of switches for builder networks
PC motherboards for data Source/Destination nodes

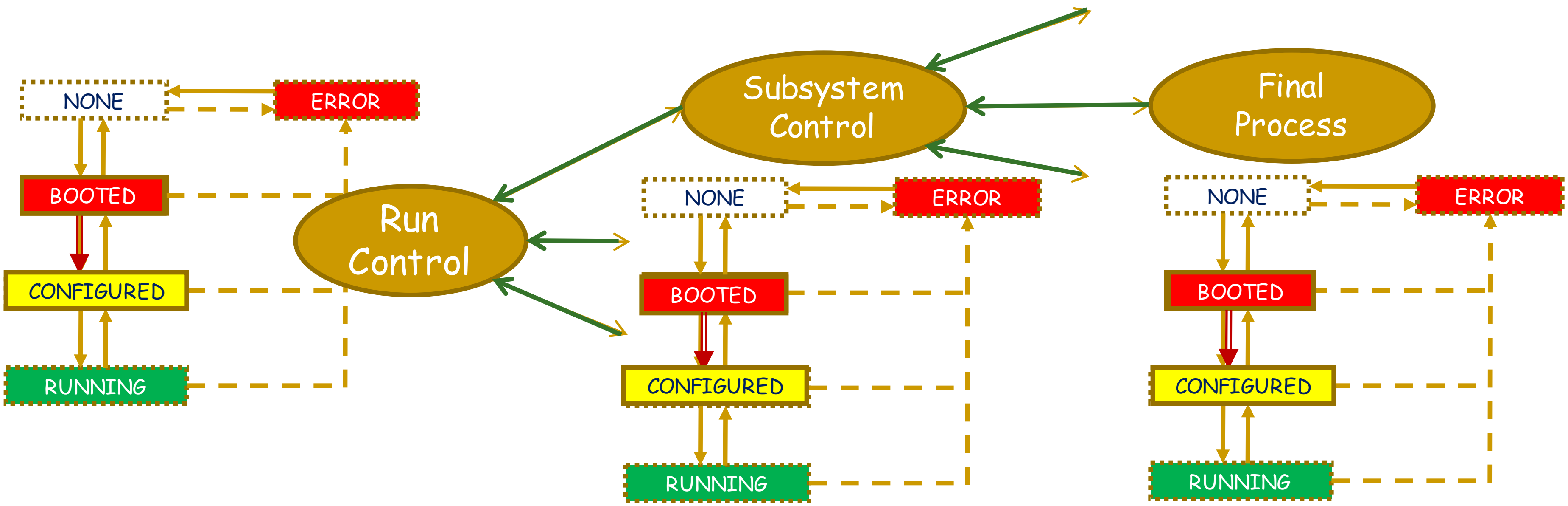
运行控制（RunControl）

- 负责系统运行控制，提供DAQ系统的状态管理，控制数据获取的动作行为
- 遵循有限状态机模型的控制器分级系统，组织成树型层次结构，避免单一控制结点产生消息瓶颈



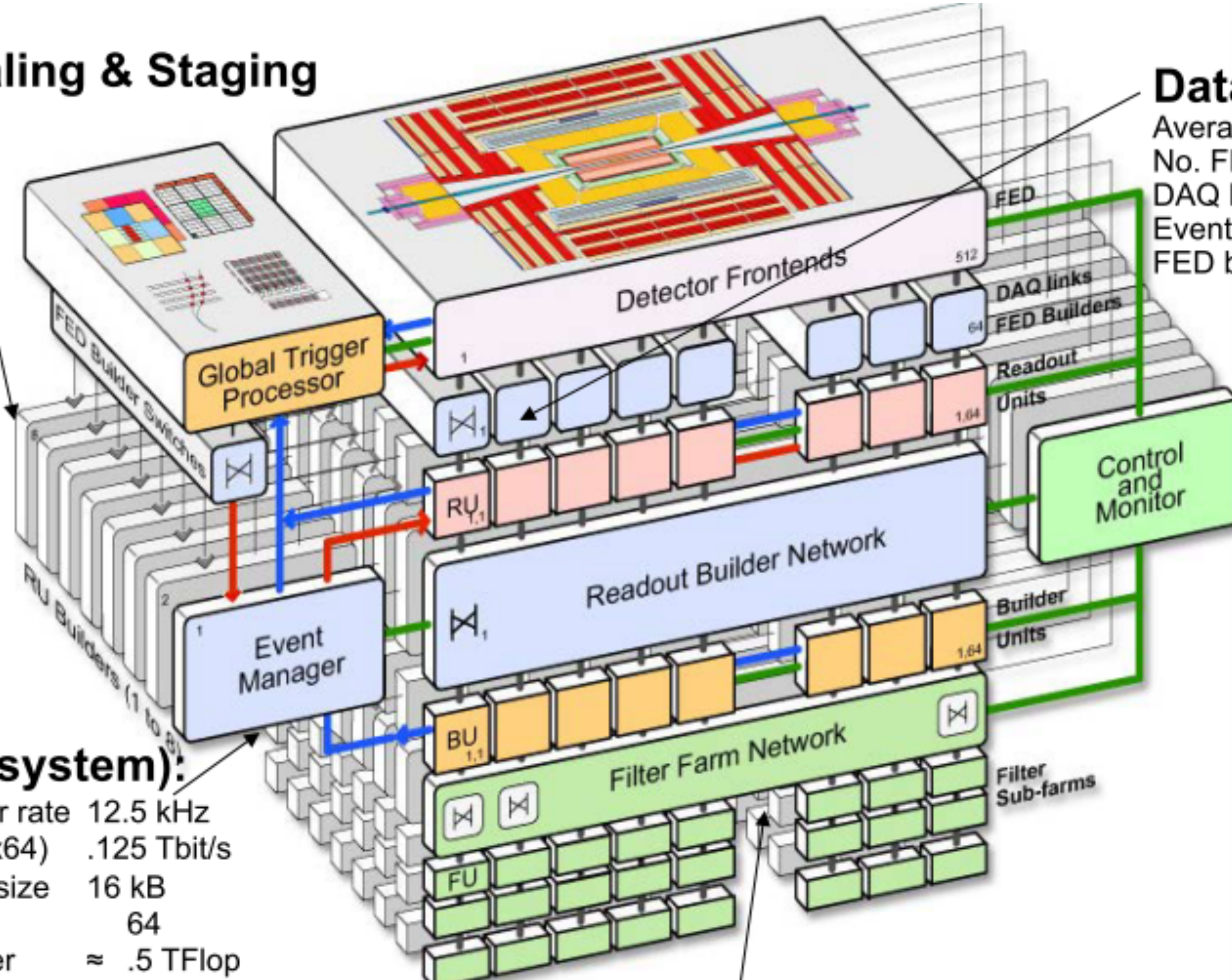
Propagating transitions

- Each component or sub-system is modeled as a FSM
  - The state transition of a component is completed only if all its sub-components completed their own transition
  - State transitions are triggered by commands sent through a *message system*



CMS DAQ

DAQ Scaling & Staging



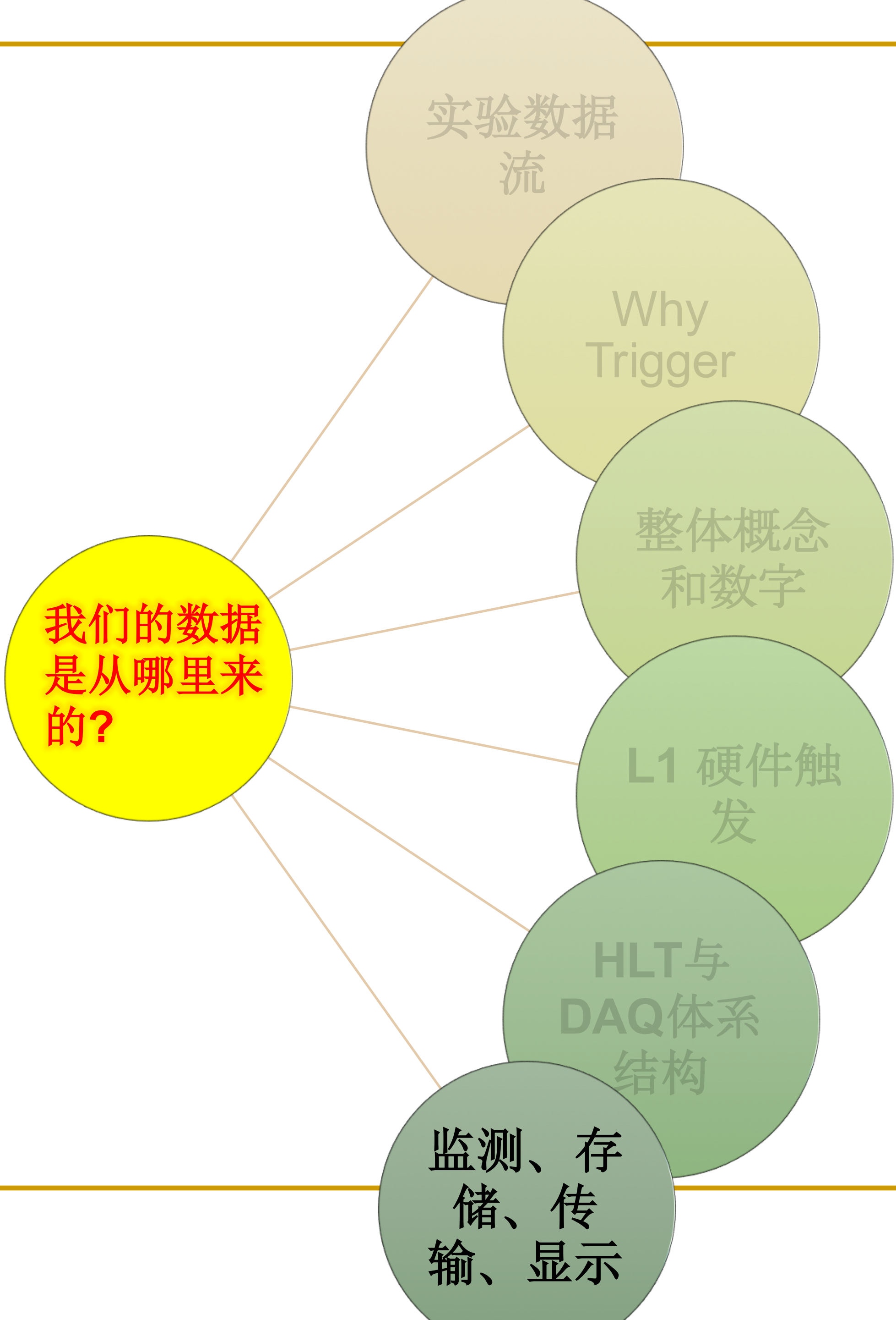
Data to surface:

Average event size		1 Mbyte
No. FED s-link64 ports	>	512
DAQ links (2.5 Gb/s)		512+51
Event fragment size		2 kB
FED builders (8x8)	≈	64+64

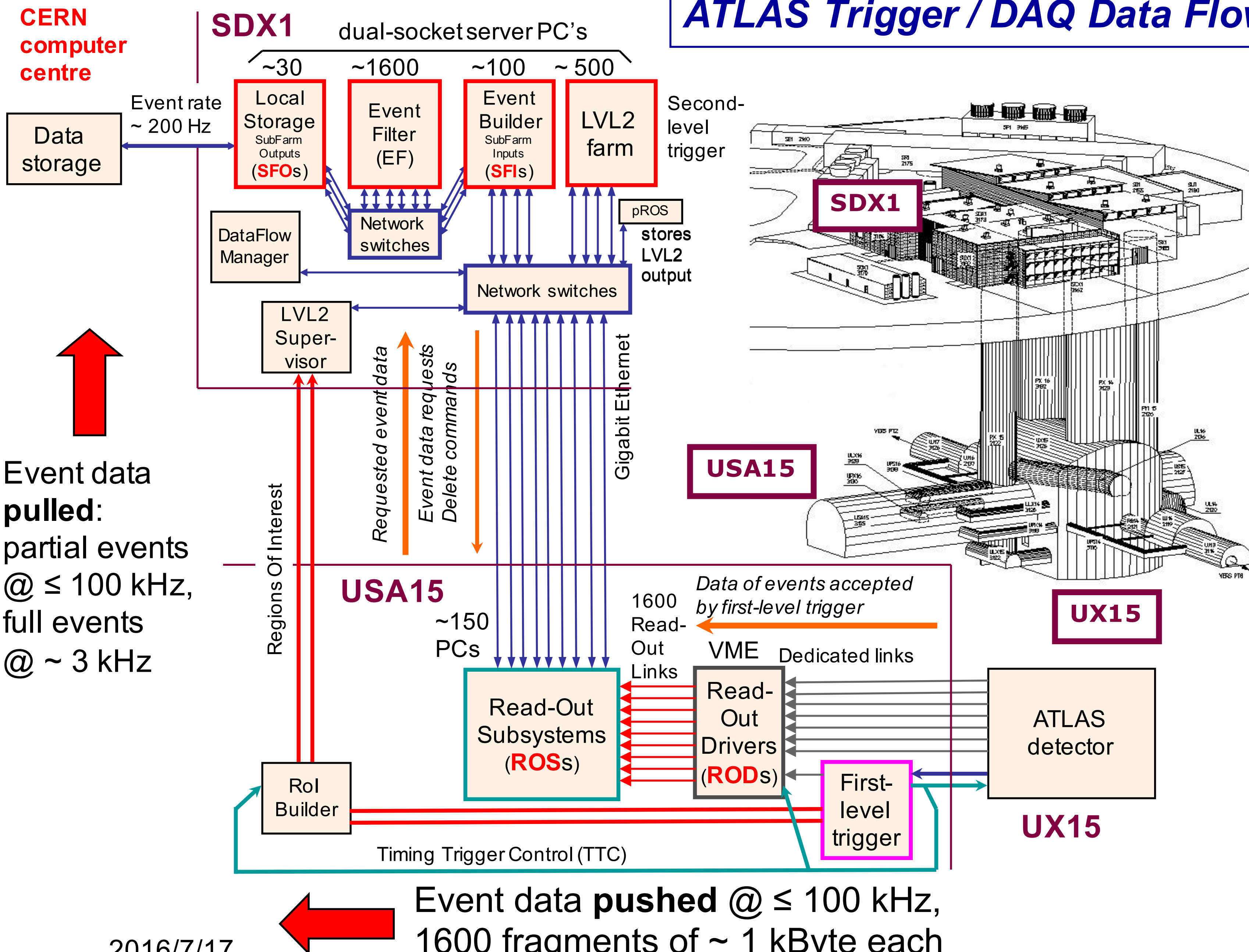
DAQ unit (1/8th full system):

Lv-1 max. trigger rate	12.5 kHz
RU Builder (64x64)	.125 Tbit/s
Event fragment size	16 kB
RU/BU systems	64
Event filter power	≈ .5 TFlop

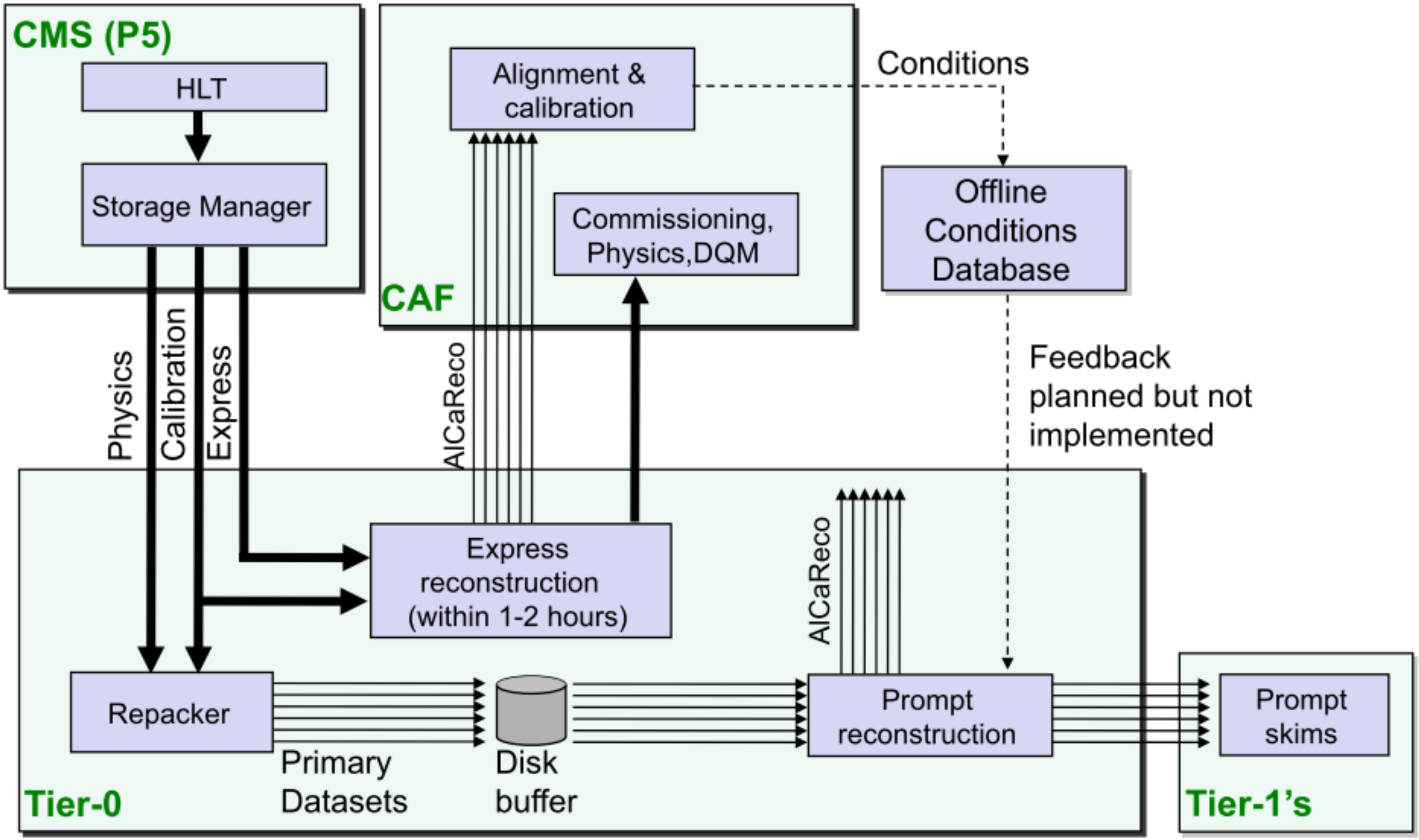
HLT: All processing beyond Level-1 performed in the Filter Farm
Partial event reconstruction "on demand" using full detector resolution


我们的数据是从哪里来的?
实验数据流
Why Trigger
整体概念和数字
L1 硬件触发
HLT与DAQ体系结构
监测、存储、传输、显示

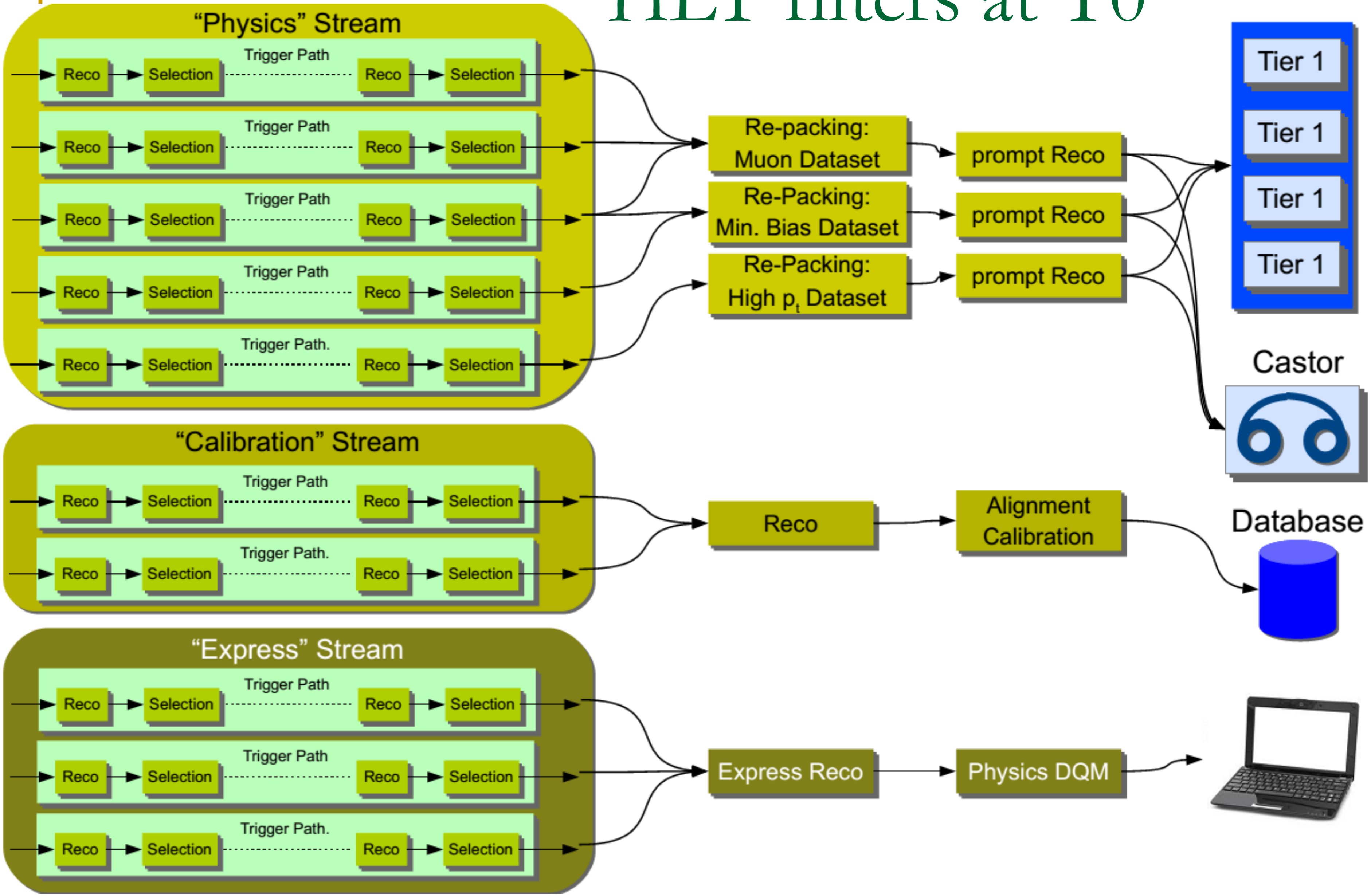
ATLAS Trigger / DAQ Data Flow



Central Processing @ CERN

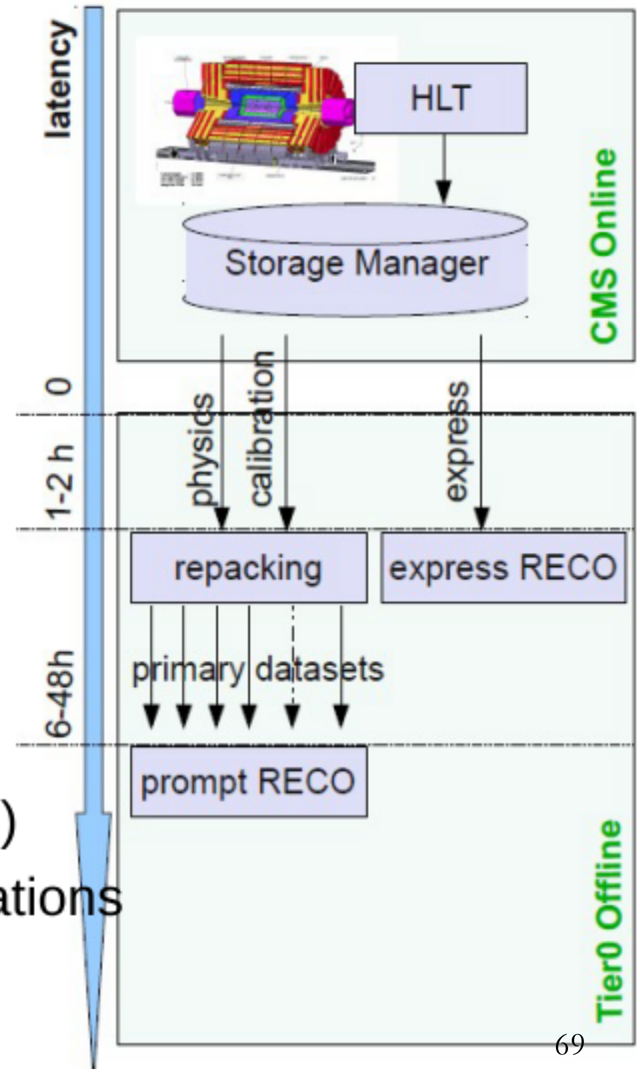


HLT filters at T0



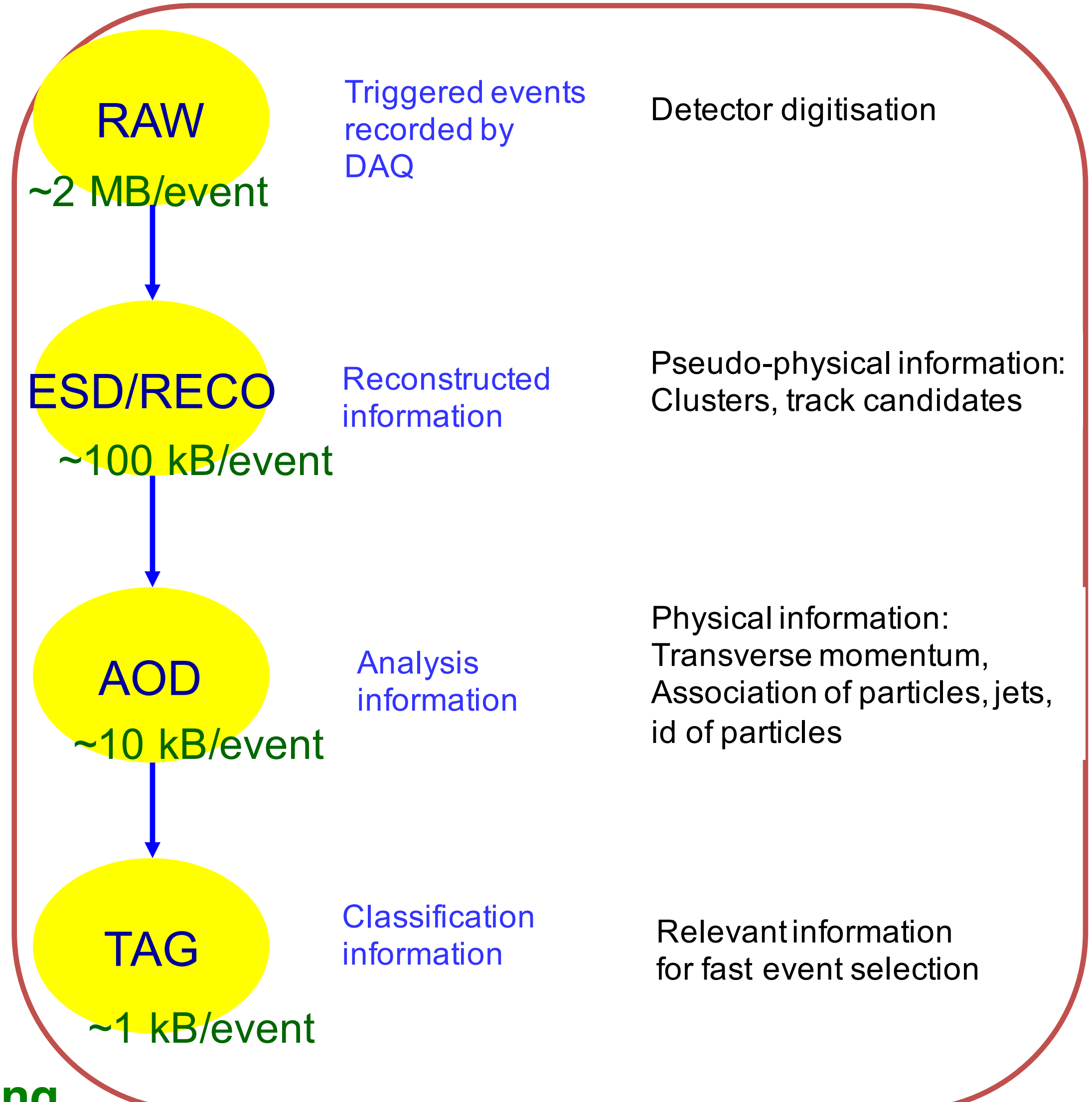
Data Streams and Tier0 workflow

- Data streams & Tier0 workflows → specialized for different tasks
- Depending on the latency
  - express → prompt feedback & calibrations
    - short latency: 1-2 hours
    - ~40Hz bandwidth shared by:
      - calibration (½)
      - detector monitoring (¼)
      - physics monitoring (¼)
  - Alignment & Calibration (AlCa) streams
  - bulk data → sample for physics analysis (prompt reconstruction)
    - split in Primary Datasets (using High Level Trigger (HLT) decision)
    - will be delayed of 48h → get latest calibrations
    - writing ~300Hz



Data Tiers and Algorithms

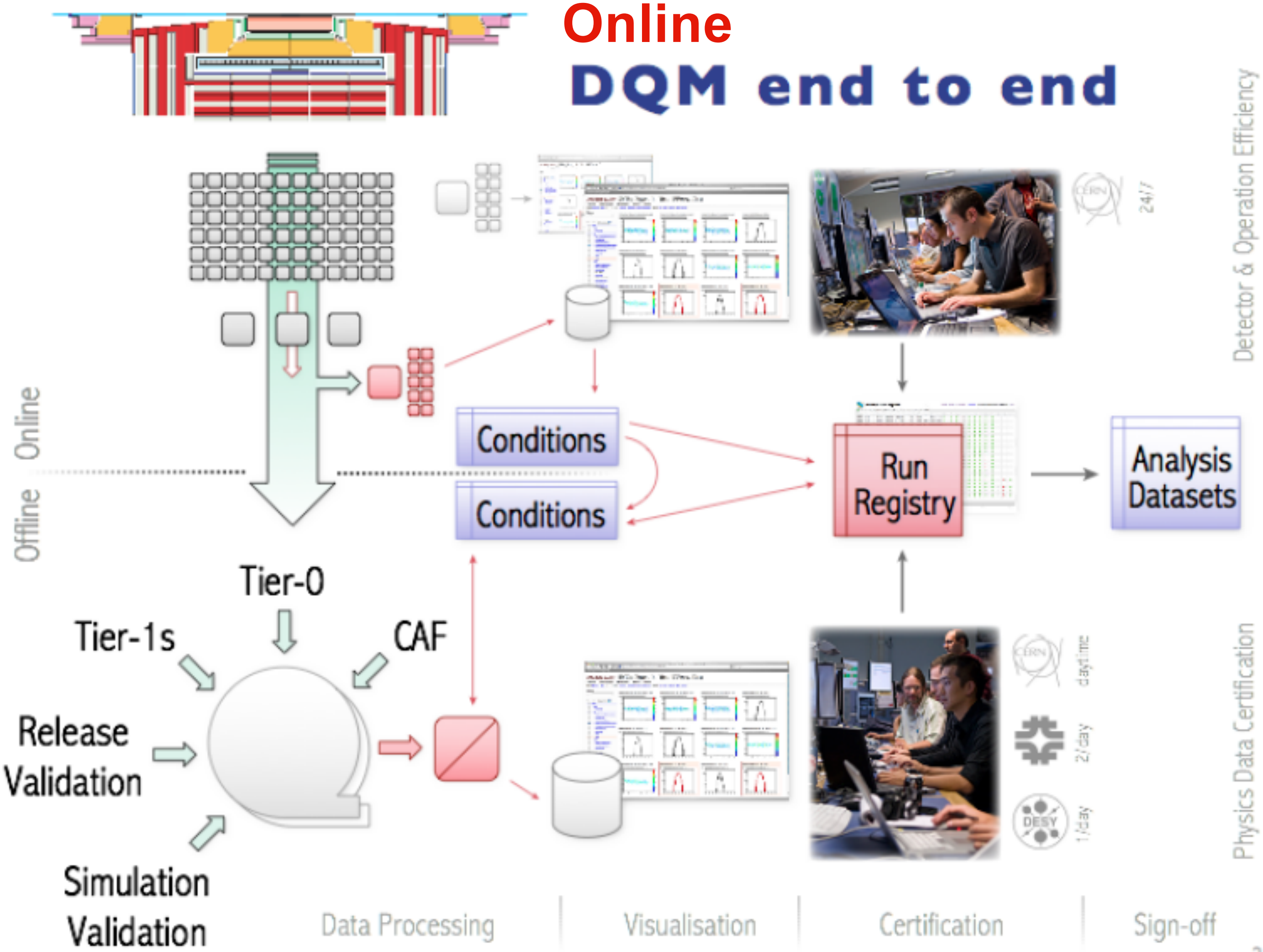
- HEP data are organized as *Events* (particle collisions)
- Simulation, Reconstruction and Analysis programs process "one event at a time"
  - Events are fairly independent → Trivial parallel processing
- Event processing programs are composed of a number of Algorithms selecting and transforming "raw" event data into "processed" (reconstructed) event data and statistics



Data Certification

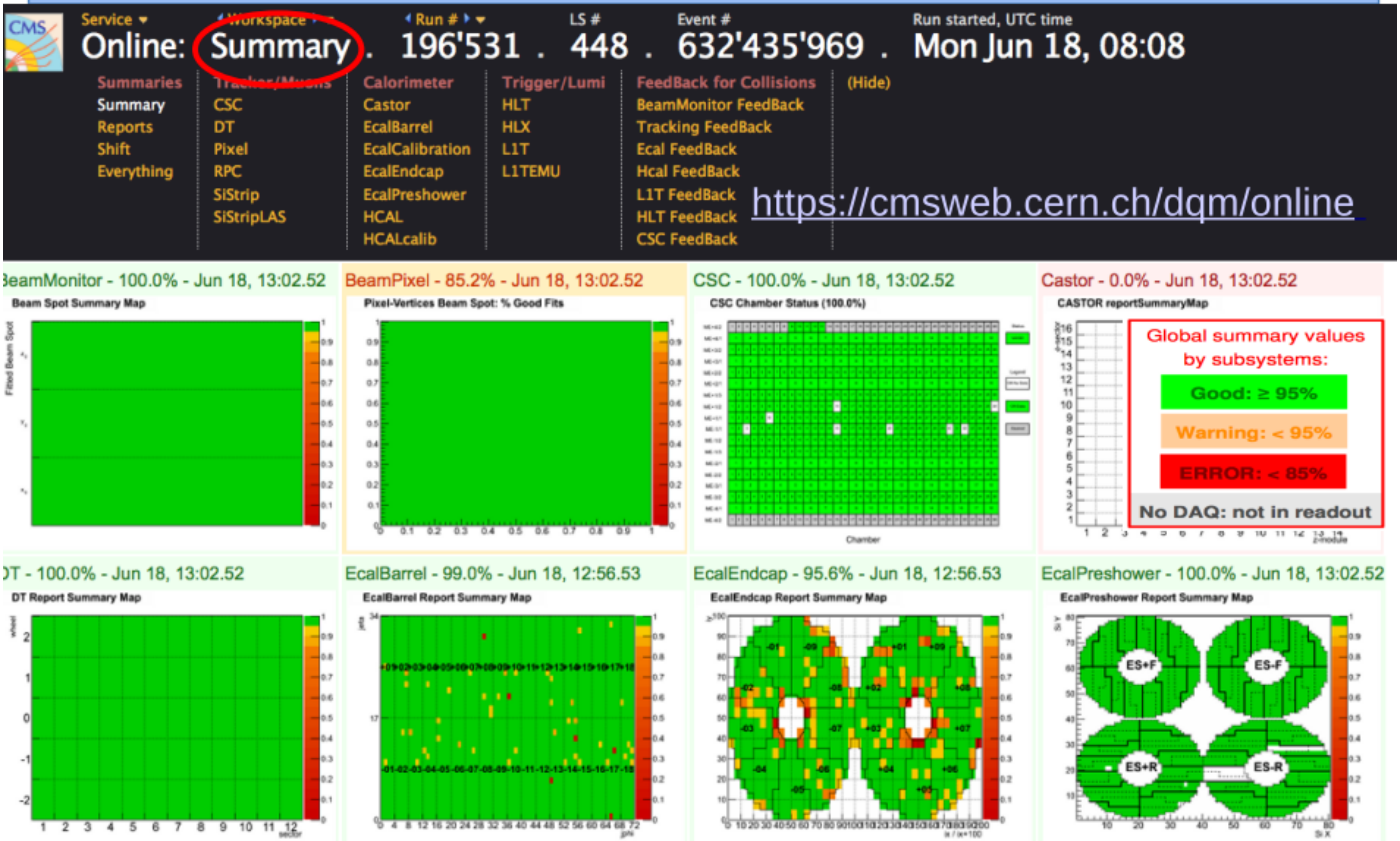
- The complexity of the offline workflows requires robust validation
- Several stages of Data Quality Monitoring (DQM):
  - online DQM → monitor detector performance during data-taking
    - dedicate event stream (sampling)
  - offline DQM → monitor performance of physics objects
    - runs on full statistics available for analysis:
      - express reco → fast feedback
      - prompt-reco → continuous monitor
      - offline re-reco → validation of software and condition updates
- Physics Validation Team → coordinates the validation activity. Feedback from:
  - groups responsible for physics objects
  - detector performance groups
  - analysis group

Online
DQM end to end

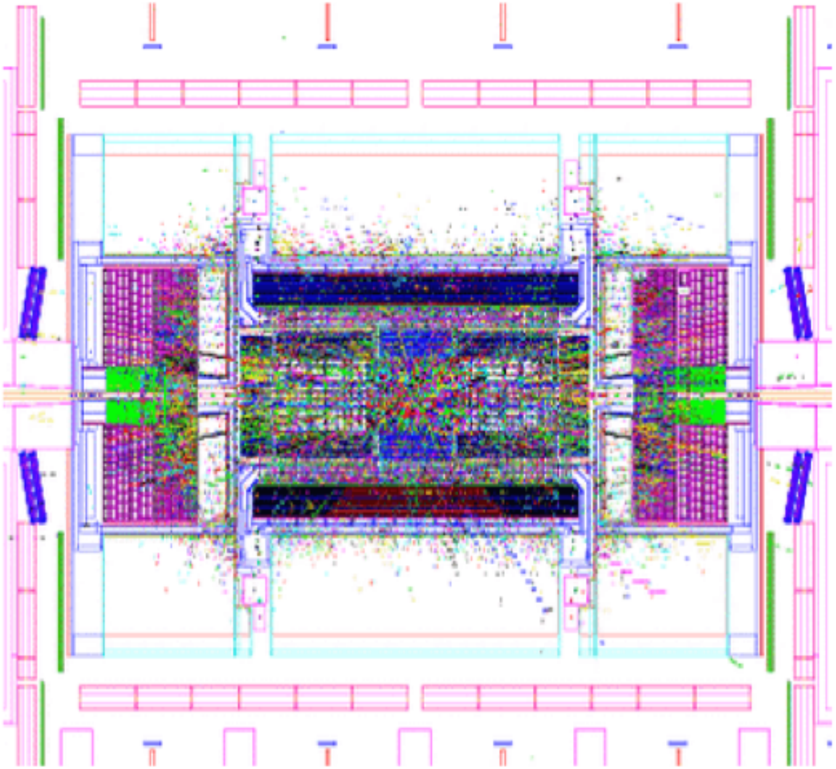


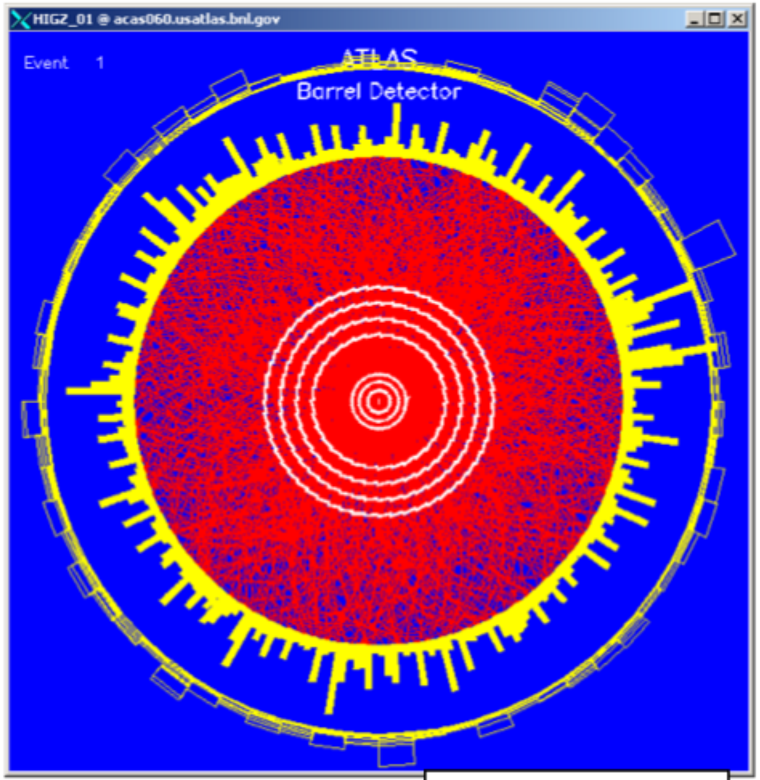
Scope of Online DQM Shifts:
Identify problems with detector performance or data integrity during the run
➔ SPOT PROBLEMS QUICKLY FOR OPTIMAL OPERATION EFFICIENCY 2

GUI – Summary Workspace



TriDAS Challenges: HL-LHC@10^{35}





$N_{ch}(|y|\leq 0.5)$

- 230 min.bias collisions per 25 ns. crossing
- ~ 10000 particles in $|\eta| \leq 3.2$
- mostly low p_T tracks
- requires upgrades to detectors

2023/24:在座诸位大有可为!

Further References:

- Bi-annual CHEP:
  - http://chep2015.kek.jp/
  - http://www.chep2013.org/
- Annual TWEPP conference:
  - http://www.lip.pt/events/2015/TWEPP/ 2015(in Sep)
  - https://indico.cern.ch/event/299180/overview 2014
- CMS TriDAS TDR.
  - V1: CERN-LHCC-2000-038 ; CMS-TDR-6-1
  - V2: CERN-LHCC-2002-026 ; CMS-TDR-6
- ATLAS TriDAS TDR
  - V1: CERN-LHCC-1998-014, ATLAS-TDR-12
  - V2: CERN-LHCC-2003-022, ATLAS-TRD-016
- ISOTDAQ: the international school of trigger and data acquisition
  - http://isotdaq.web.cern.ch/isotdaq/isotdaq/Home.html

关于实验，几句多余的话

- 物理学是实验科学
- 以大搏小，见微知著

$$dp \cdot dx \sim \hbar \xrightarrow{p \sim E} 1\ \text{TeV} \propto \frac{1}{2 \times 10^{-19}\ \text{meters}}$$

- 以今日之物理探究明日之物理
  - 技术支撑
  - 标准烛光
- 研无定法，无所不用其极