Challenges and opportunities for a First Level Trigger based on Silicon Trackers for the High Luminosity LHC



**Fabrizio Palla INFN Pisa**

### Run1 legacy



#### **Whiggs-like particle has now become a (the) Higgs boson**





### No evidence of New Physics (yet!) [INFN]



#### No evidence of New Physics (yet!) **INFN**



F. Palla INFN Pisa

Istituto Nazionale di Fisica Nucleare

## No evidence of New Physics (yet!)



F. Palla INFN Pisa

stituto Nazionale di Fisica Nucleare

#### No evidence of New Physics (yet!) Istituto Nazionale di Fisica Nucleare



F. Palla INFN Pisa

### No evidence of New Physics (yet!)



**CMS Exotica Physics Group Summary - March, 2014** 

 $n$ et= $\sigma$ 

niet=0

stituto Nazionale di Fisica Nucleare

### What could be done next?

#### **O**Run 2 (2015-18):

 $Q$ ~100 fb<sup>-1</sup> at 13 TeV, 10<sup>34</sup> cm<sup>-2</sup> s<sup>-1</sup>

#### **Gain depends on coupling (qq, gg, qg)**

 $-1$  year worth ~250 years of 8 TeV data for e.g. 4 TeV Z' or 2 TeV squarks big gains in sensitivity



stituto Nazionale di Fisica Nucleare



 $\sim$ 3000 fb<sup>-1</sup> expected in 10 years running

Excellent opportunity to search for (rare) and new phenomena

Need anyway still to trigger on "SM" objects (leptons, b, jets, MET)

### Implications for the Trigger



- $\bullet$  At 5x10<sup>34</sup> cm<sup>-2</sup> s<sup>-1</sup> up to ~140 interactions per bunch crossing
	- About 6k primary tracks per bunch crossing in the Tracker volume |η|<2.5 ...
		- ...plus any other coming from γ conversions and nuclear interactions
	- ~ one order of magnitude larger wrt LHC
	- **Severe Triggering conditions** 
		- Too many primary vertices, need to have smarter triggers combining information from several subdetectors
		- Need to maintain low thresholds for basic objects, even with an increase in the L1-Accept bandwidth (currently at 100 kHz)
	- Both ATLAS and CMS will replace their "inner trackers" to cope with the nasty environmental conditions
		- The usage of the Tracker would help to disentangle among those 140 pileup events





### Why a Track Trigger at L1



- **OHL-LHC physics goals require excellent Trigger selectivity on** basic objects (leptons, jets, taus, b-jets, MET)
	- This might be jeopardized by the increased level of pileup events (140 on average)
		- $\Theta$ Huge rate of  $\mu$  from heavy flavors  $\Rightarrow$  use better p<sub>T</sub> resolution from tracker **OPrompt electrons at L1 need to be separated from huge γ**  $\Rightarrow$  **Tracker tracks**  $\Theta$ High E<sub>T</sub> jets from (many) different primary vertices  $\Rightarrow$  jet-vertex association  $\bullet$ Photon isolation in Calorimeters compromised by large pileup  $\Rightarrow$  use tracks





F. Palla INFN Pisa

## The challenge and the way out

Cluster rate (in MHz/cm

#### Take data off the tracker

- $\odot$  ~ 4k primary tracks within  $ln|<2.5$ 
	- Large data rates (up to 25 MHz/cm2)
		- $\Theta$  huge contribution from nuclear interactions and photon conversions
	- $\bullet$  ~1.3 events/mm  $\times$  Gauss( $\sigma$ =4 cm)
	- $\odot$  Short L1A trigger latencies (10-20  $\mu$ s)
	- Cannot read all (~60 M strips) channels at 40 MHz
		- Even a 1% occupancy: 0.5 M channels x 40 MHz x 20 bit =  $400$  Tb/s

 $\sim$ 120k links at 3.25 Gb/s (GBT) - Current CMS Tracker has 40k links (320 Mb/s)

**O** Need to

suppress hits from low  $p_T$  tracks read at smaller (affordable) rate

#### Conce data are off-detector, find tracks and

 $\heartsuit$  formidable pattern recognition problem

 $\circ$  need latencies of ~5  $\mu$ s

F. Palla INFN Pisa



stituto Nazionale



## Trigger architectures

#### **OPUSH path (CMS)**

**Q Reduced Tracker information readout at** 40 MHz and then combined with calorimeter & muon at L1

**OTrigger objects made from tracking,** calorimeter & muon inside a Global Trigger module

#### **OPULL path (ATLAS)**

**QUse calorimeter & muon detectors to** produce a "Level-0" to request tracking information in specific regions

**OTracker sends out information from** regions of interest to form a new combined L1 trigger



stituto Nazionale li Fisica Nucleare

#### Data reduction



### Track Trigger with pull architecture (ATLAS)

tituto Nazionale i Fisica Nuclean

#### The LO+L1 scheme

Level-0:

- Coarse calo and muon data
- $\odot$  Rate 40 MHz  $\rightarrow$  500 kHz
- $\odot$  Latency  $<$  6.4  $\mu$ s
- Defines Region of Interest (ROIs) for L1

Level-1:

- **O Tracker data only from ROIs**
- Refined information from calo and muons
- $\odot$  Rate 500 kHz  $\rightarrow$  200 kHz
- $\Omega$  Latency < 20  $\mu$ s

#### **Olssues for FTK to be used in Phase 2**

- Level-1: the larger pileup (x2.5), rate (x5) and granularity
	- rder of magnitude and calorimeters and music increase in the number of patterns by ~one order of magnitude
		- $\bigcirc$ no p<sub>T</sub> filtering rise p<sub>T</sub> threshold
- $3r$ d May 2012  $\mu$ need to cope with shorter latency (20  $\mu$ s instead of 200  $\mu$ s)

F. Palla INFN Pisa







### ATLAS readout



ATLAS Tracker for HL-LHC



 $z(m)$ 



# ATLAS L0 and Regional Readout Requests (R3) implementations



L0 Trigger accept rate 500 kHz

- On a L0 accept, copy data from primary to secondary buffer
- $\odot$  Identify "region of interest" (1-10% of the detector on each L0 accept)
- Generate a "Regional Regional Request" (R3)
	- $\odot$  Reading only ~10% of the Tracker data, the total bandwidth is only 50% more with the Track Trigger than without.
- To reduce the latency, a prioritization scheme is envisaged, by using a dedicated R3 buffer **HCC**





stituto Nazionale di Fisica Nucleare

### Simulation results - ATLAS





 $\mathcal{L}_{\mathcal{A}}$  accept rate and the R3 rate (occupancy) for the discrete event simulation of the Phase III rate  $\mathcal{L}_{\mathcal{A}}$ 

# Select only hits from "high-p<sub>T</sub>" tracks

 $\bullet$  Select "high-p<sub>T</sub>" tracks (>2 GeV) by correlating hits in 2 nearby sensors (stub)



- In the end-cap, it depends on the location of the detector
	- $\rightarrow$  End-cap configuration typically requires wider spacing (up to  $\sim$  4 mm)

 $\Delta z = \Delta R /$  tg  $\theta$ 





### CMS 2S modules







### 2S module prototype









### 2S module prototype









#### **Electronics**







### CMS PS modules



#### P(ixel)S(strip) module  $\odot$ strips = 100  $\mu$ m x 2.4 cm  $\Theta$ pixels = 100  $\mu$ m x 1.5 mm  $\bullet$  Pixels are logically OR-ed for finding coincidence in the r- $\phi$  plane, and the precise z-coordinate is retained in the pixel storage and provided to the power converter trigger processors. 2 W **SSA Short Strip ASIC Strip sensor Pixel Sensor MPA Carbon Fiber** C4 pad dimension  $\sim$  90 um **Macro Pixel ASIC** concentrator **Flex Board** TSV diameter ~ 75um 200 mW 2 x 8 SSA chips 512 mW **L1 Data Pipeline** @ BX (40 MHz) GBT & **Pixel/Strip Memory &** @ L1 (1 MHz) 2 x 8 MPA opto package Up to 60 bit Data compression 12 bits 3004 mW 800 mW each row Output @160 MHz Interface To the **Trigger Logic for Stub SSA** @ BX (40 MHz) @ BX (40 MHz) Concentrator Up to 80 bit **Finding** concentrator chip 200 mW**Trigger Logic**



F. Palla INFN Pisa

## Data organization and dispatch



Example CMS: 8(r-ϕ)x6(r-z) trigger sectors (some 10% overlapping)

● Each sector ~200 stubs on average; tails up to ~500 stubs/event in 140 evts pileup+ttbar (to be compared with ATLAS-Phase 1 ~2000)

About 600 Gb/s per one trigger tower



G Send data to Track-finding processors

#### **OFull mesh ATCA shelfs**

- Capable of "40G" full-mesh backplane on 14 slots = 7.2 Tb/s
- Several options being investigated, all include time multiplexing data transfer from a set of receiving processors boards to pattern recognition and track finding engines
- $O(10)$  time multiplexed at the shelf level

 $\odot$  keep latency < 5  $\mu$ s, including pattern recognition and track fitting

Number of connections to trigger processors

23



### Trigger Tower segmentation

#### Regional multiplexing => divide the detector into trigger towers



F. Palla INFN Pisa

stituto Nazionale di Fisica Nucleare



### Tower interconnections













#### Pattern matching in CDF





F. Palla INFN Pisa



A pattern



#### $\rightarrow$  Superstrip definition:



 $\rightarrow$  A superstrip is simply a bunch of strips in one module of the tracking detector.

 $\rightarrow$  The superstrip address is the info sent to the AM board. Is is coded on a certain number of bits, depending on the superstrip resolution.



 $\rightarrow$  The encoding is divided into 4 parts, giving module and intra-module SS position in Z and  $\phi$ direction (R is not necessary)

 $\rightarrow$  We are not using pixel info yet, so our Z intramodule encoding is very basic for the moment.





FPGA  $EP($ 











### The AM chip at work



**The event hit positions are received over 8 input buses of 15 bits each. All the hits are then compared with the data stored inside each layer block, as soon as they are loaded into the chip, each one in the corresponding bus.** If a layer block is matched, the corresponding Flip-Flop (FF) is set. It should be noted that each hit is fed into the memory only once. In fact the bus line transmits the information to all the layer blocks, and, if matched, all the corresponding FF are set simultaneously. **Finally, a given pattern is matched with a logic that counts the number of FF set to 1 within a row, using a majority logic: that means that one could ask a minimum number of FF set**



#### Usage in ATLAS @L1.5



**NFN** 

Istituto Nazionale<br>di Fisica Nucleare



# Associative Memory for pattern matching





F. Palla INFN Pisa



# Track fitting - high quality helix parameters<br>and  $x^2$



#### **O** Principal component analysis

Over a narrow region in the detector, equations linear in the local silicon hit coordinates give resolution nearly as good as a time-consuming helical fit.

Nucl.Instrum.Meth.A623:540-542,2010 doi:10.1016/j.nima.2010.03.063

$$
p_i = \sum_{j=1}^{14} a_{ij} x_j + b_i
$$



•pi 's are the helix parameters and 2 components.

•xj 's are the hit coordinates in the silicon layers.

•a<sub>ij</sub> & b<sub>i</sub> are pre-stored constants determined from full simulation or real data tracks.

•The range of the linear fit is a "sector" which consists of a single silicon module in each detector layer.

•This is VERY fast in FPGA DSPs.

#### $\odot$  ~few hundred fitting engines/trigger sector for CMS

## Time multiplexing and data formatting

Istituto Nazionale di Fisica Nucleare

Ten processors send data to target processor blade in round robin scheme. Each blade will have a few mezzanines to handle multiple events. Does not need to wait last stub inputed to start track finding.



F. Palla INFN Pisa



### AM technological evolution



**XILINX®** 200E™<br>56AFS002!

- **(90's) Full custom VLSI chip 0.7um (INFN-Pisa)**
- **128 patterns, 6x12bit words each, 30MHz**
- F. Morsani et al., IEEE Trans. on Nucl. Sci., vol. 39 (1992)

Alternative FPGA implementation of SVT AM chip

P. Giannetti et al., Nucl. Intsr. and Meth., vol. A413/2-3, (1998)

G Magazzù,  $1^{st}$  std cell project presented @ LHCC (1999)



j **Standard Cell 0.18**  $\mu$ m  $\rightarrow$  5000 pattern/AM chip SVT upgrade total: 6M pattern, 40MHz A. Annovi et al., **IEEE TNS,** Vol 53, Issue 4, Part 2, 2006





AMchip04 -65nm technology, std cell & full custom, 100MHz Power/pattern/MHz ~30 times less. Pattern density x12. First variable resolution implementation! F. Alberti et al 2013 JINST 8 **C01040**, doi:10.1088/1748-0221/8/01/C01040

2013

### System dimensioning

#### **OPattern matching**

#### Optimization on-going

- $\bullet$  fixed super-strip (32 strips each) size for all layers:  $\sim$ 4 M patterns
- projective (8/16/32 ... ) sizable reduction (up to 2); ~same (or even better) performance

Unique roads fired per trigger ~<50 @ PU 200 and 3 GeV threshold

 $\bigcirc$  Efficiency ( $\mu$ , electrons)~99%

 $\bigcirc$  Purity of stubs after AM filtering ~60%

#### Further ~30% gain from stub pT info















F. Palla INFN Pisa



#### **Hardware**





Pulsar 2b in hand Fully qualified for 10Gb/s speed

Mezzanine in hand being qualified for AM05/06

Other version in production for ProtoVipram







#### State of the art and immediate R&D

INFN/IN2P3 65 nm AM05 (3k patterns) in hand to produce 4 mezzanines x 16 chips and AM06 (128k patterns) procurements Fall 2015.

https://indico.cern.ch/event/354340/contribution/0/material/slides/0.pdf

●8 input, 1 output serial lines @ 2 Gbps, 100 MHz

Sufficient to test latency and projections - small ratio matched roads to input stubs

Started R&D for 28 nm, target 0.5M pattern, 200+MHz speed - not for demonstrator



Also a dozen of FNAL protoVipram 4k pattern, 130 nm

Started R&D on a 40 nm chip, 0.5M patterns, 200+MHz speed - not for demonstrator

F. Palla INFN Pisa



### AM05, its Test Stand, and AM06



#### *Design: Stabile (MI) – Crescioli (LPNHE) – Beretta (LNF)*

#### **Big improvements in AMchip design**

- $\circ$  AMchip04 power consumption / # of bit / MHz decreases of a factor ~28 w.r.t. AMchip03
- $\circ$  AMchip04 memory density (patterns\*layers/area) increase of a factor ~18 w.r.t. AMchip03
- $\circ$  High speed serial links (11 times 2 Gb/s)

23x23 mm2 BGA for AM05: 3k patterns and AM06: 128k patt.





#### AM06 New layout 128 kpatt

(Stabile work, Stabile/Liberali cell) **14.6 mm x 10.8 mm<sup>2</sup>** 



See for full information: https://agenda.infn.it/getFile.py/access?contribId=10&sessionId=1&resId=0&materialId=slides&confId=8420

F. Palla INFN Pisa



## The AM chip up to 2020 – R&D





- The AM2020 chip features
	- $\odot$  Assuming technology scaling would allow  $(65/28)^2 \sim 5.4$  more density  $$
	- Need to optimize the design to decrease power consumption and area of memory arrays

Target ~3 W total power @100 MHz

- Optimize latency reduction: inputs from simulations and demonstrator
	- $\bullet$  Is 200 MHz is required to speed up I/O ? could double the power
	- O Need more output buses ?
- <sup>◎</sup> In collaboration with LPNHE Paris (FTK) and Lyon under ANR project
- Starting the design as soon as AM06 chip is submitted



### Tracklet approach



#### **Tracklet Algorithm: Road Search**



**28 regions in ϕ**



#### Expected performances (tracklet) Tracking Efficiency



- L1 tracking efficiency as function of  $\eta$  &  $p_T$  for single  $\mu$ ,  $\pi$ , e with  $\langle \text{P} U \rangle = 140$ 
	- **Muons** Sharp turn-on at 2 GeV & high efficiency across all n
	- **Pions** Somewhat lower efficiency due to higher interaction rate
	- **Electrons** Slower turn-on curve, efficiency reduced from bremsstrahlung
- For  $|n| < 1.0$  &  $p_T > 2$  GeV, efficiency for  $\mu$ ,  $\pi$ , e is  $>99\%$ ,  $95\%$ ,  $87\%$







F. Palla INFN Pisa

F N

Istituto Nazionale di Fisica Nucleare



### Usage of L1 Tracks - CMS





Matching Drift Tube trigger primitives with L1Tracks: **large rate reduction:**

 **> 10 at threshold > ~ 14 GeV.** Normalized to present trigger at 10 GeV. Removes flattening at high  $P_t$ 



**Rate reduction** brought by matching L1 e/γ to L1Track stubs for  $|\eta|$  < 1. Red: with current (5x5 xtal) L1Cal granularity.

Green : using single crystal-level position resolution improves matching



#### Primary Vertex and Track MET Primary Vertex & Track MET



- L1 tracks can also be used to reconstruct primary vertex of event
- Resolution of primary vertex using L1 tracks with  $p_T > 2$  GeV or 5 GeV
	- <1mm for events with large track multiplicity
		- *• Here: ttbar <PU>=140*
	- Similar performance with the higher track  $p<sub>T</sub>$  threshold Unimar purfuritured with the ingrid

- **Track "MET"** 
	- Define L1 track-based missing transverse momentum from L1 tracks coming from primary vertex  $Dofino$ ,  $\overline{1}$  track because prioring Deilig Li





#### SUSY Signal point SUSY Signal Point



- Rate reductions using L1 tracks for SUSY signal
	- Stop pair production with hadronic top decays (stop=775 GeV, LSP=550 GeV)
	- Signal defined by genMET > 100 GeV



- Missing  $H_T$  determined with/without vertex association
	- Algo1 & Algo2: Calorimeter-based L1 jet algorithms with different PU subtraction methods
- Sizable rate reductions achieved with tracking information!





## CMS Gains from Track Trigger



Preliminary simulation studies demonstrate addition of L1 tracking trigger provides significant gains in rate reduction with good efficiency for physics objects. Note these results are "work in progress".







**OT** racker information helps reducing drastically the rate of uninterested events

- This will become a new "must" for all future detectors
- **@HL-LHC detectors will make use of tracking information in** the Level-1 Triggers
	- Several trigger architectures exploited
		- $\bullet$  Full readout @40 MHz, on-detector data reduction using p $\uparrow$ -modules
		- **OImplications on Tracker detector layouts ongoing**
	- Some demonstrators being built to validate the full chain
	- Large gains in combining tracking with other subdetectors OElectrons, Muons, Jets and MET

High statistics of useful events for precision physics available **■ Stay tuned!** 







# ATLAS L0 and Regional Readout Requests (R3) implementations



L0 Trigger accept rate 500 kHz

- On a L0 accept, copy data from primary to secondary buffer
- $\odot$  Identify "region of interest" (1-10% of the detector on each L0 accept)



- $\odot$  Reading only ~10% of the Tracker data, the total bandwidth is only 50% more with the Track Trigger than without.
- To reduce the latency, a prioritization scheme is envisaged, by using a dedicated R3 buffer **HCC**





stituto Nazionale di Fisica Nucleare

### Simulation results - ATLAS





 $\mathcal{L}_{\mathcal{A}}$  accept rate and the R3 rate (occupancy) for the discrete event simulation of the Phase III rate  $\mathcal{L}_{\mathcal{A}}$ 

# Anatomy of a PRAM (Pattern Recognition Associative Memory)



F. Palla INFN Pisa Trace Length -> Capacitance -> Power Consumption or Reduced Speed More detector layers, or more bits involved, design more spread out in 2D  $\rightarrow$  less pattern density, higher power consumption ...

# Increasing the pattern density



#### AMCHIP04: VARIABLE RESOLUTION





### ATLAS FTK



#### **Inner Tracker Overview**



- To deal with data flow designed pas highly narallel system. Such the detector would include the effects of multiple scattering, and multiple s as highly parallel system
- is included, the size of the pattern bank becomes extremely large. Consequently we accept some **- 8** 'core crate' with own pattern recognition and track fitter
- range of track rapidity. We configure FTK to reconstruct any track leaving at least *M*-1 hits out **-** Detector subdivided in 64 **trigger tower** that can be crossed by a sector (see figure a sector (see figure a sector (see figure a sector  $\mathbf{r}$
- **• PIX (3 layers) & SCT (4 double layers)**
- **•** Fit posses combinatorics problem, **executed in two sequential steps:**  $\bullet$  5  $\bullet$ 
	- **- Use 8 layer for patter recognition and 8 layer fit**
	- **- Refit track found using all 11 layer**



