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

> Fabrizio Palla INFN Pisa

Run1 legacy



Weight and the set of the set









F. Palla INFN Pisa

Istituto Nazionale di Fisica Nucleare



F. Palla INFN Pisa

Istituto Nazionale di Fisica Nucleare



F. Palla INFN Pisa

lstituto Nazionale di Fisica Nucleare



CMS Exotica Physics Group Summary – March, 2014

njet=0

njet=0

lstituto Nazionale di Fisica Nucleare

What could be done next?

Run 2 (2015-18):

Gain depends on coupling (qq, gg, qg)

I 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



~3000 fb⁻¹ 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



- Sector States and the sector of the sector o
 - \bigcirc About 6k primary tracks per bunch crossing in the Tracker volume $|\eta| < 2.5 \dots$
 - \odot ...plus any other coming from γ conversions and nuclear interactions
 - <u>~ one order of magnitude larger wrt LHC</u>
 - 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



- HL-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)
 - Huge rate of μ from heavy flavors \Rightarrow use better p_T resolution from tracker
 - Prompt electrons at L1 need to be separated from huge $\gamma \Rightarrow$ Tracker tracks • High E_T jets from (many) different primary vertices \Rightarrow jet-vertex association
 - •Photon isolation in Calorimeters compromised by large pileup \Rightarrow use tracks





F. Palla INFN Pisa

The challenge and the way out

rate (in MHz/cm

Cluster I

Take data off the tracker

- \bigcirc ~ 4k primary tracks within $|\eta| < 2.5$
 - Large data rates (up to 25 MHz/cm²)
 - huge contribution from nuclear interactions and photon conversions
 - ~1.3 events/mm × Gauss(σ =4 cm)
 - Short L1A trigger latencies (10-20 μ 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

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

Need to

suppress hits from low $p_{\rm T}$ tracks read at smaller (affordable) rate

Once data are off-detector, find tracks and

formidable pattern recognition problem

• need latencies of ~5 μs

F. Palla INFN Pisa



tituto Naziona



Trigger architectures

PUSH path (CMS)

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

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

PULL path (ATLAS)

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

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



stituto Nazionale li Fisica Nucleare

Data reduction





The L0+L1 scheme

✓Level-0:

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

✓ Level-1:

- Tracker data only from ROIs
- Refined information from calo and muons
- Rate 500 kHz → 200 kHz
- Latency < 20 μ s

Issues for FTK to be used in Phase 2

- \bigcirc the larger pileup (x2.5), rate (x5) and granularity
 - increase in the number of patterns by ~one order of magnitude • no p_T filtering - rise p_T threshold
- \bigcirc need to cope with shorter latency (20 μ s instead of 200 μ 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 LO accept, copy data from primary to secondary buffer
- Identify "region of interest" (1-10% of the detector on each L0 accept)
- Generate a "Regional Regional Request" (R3)
 - 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





Istituto Nazionale di Fisica Nucleare

Simulation results - ATLAS





Select only hits from "high-p_T" tracks

 \bigcirc Select "high-p_" tracks (>2 GeV) by correlating hits in 2 nearby sensors (stub)



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

 $\Delta z = \Delta R / tg 9$





CMS 2S modules







2S module prototype









2S module prototype









Electronics







CMS PS modules



@P(ixel)S(strip) module

- \bigcirc strips = 100 μ m x 2.4 cm
- \bigcirc pixels = 100 μ m x 1.5 mm

• Pixels are logically OR-ed for finding coincidence in the r- ϕ plane, and the precise z-coordinate is retained in the pixel storage and provided to the trigger processors.





Data organization and dispatch



 \bigcirc 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



Send data to Track-finding processors

Full 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

• keep latency < 5 μ s, including pattern recognition and track fitting

Number of connections to trigger processors







Trigger Tower segmentation

Regional multiplexing => divide the detector into trigger towers



F. Palla INFN Pisa

lstituto Nazionale di Fisica Nucleare



Tower interconnections







40G full-mesh backplane







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.

Superstrip enco	ding	Generic superstrip address definition		
Z module	Z inner	φ module φ inner		
		Strip t	racker module example	
Z segment (1bit) Z module (5bits)		φ for 3.2mm pitch (5bits)		
V	→ ↓	V	V	
2 4 8 16	32 2 2 4	8 2 4	8 16 32 64	
	1		^	
	φ module	(3bits)	6mm pitch <i>(6bits)</i>	

 \rightarrow The encoding is divided into 4 parts, giving module and intra-module SS position in Z and ϕ 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





FPGA





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



| N F N

Istituto Nazionale di Fisica Nucleare



Associative Memory for pattern matching







Track fitting - high quality helix parameters and χ^2



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

$$\boldsymbol{p}_i = \sum_{j=1}^{14} \boldsymbol{a}_{ij} \boldsymbol{x}_j + \boldsymbol{b}_i$$



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

 $\boldsymbol{\cdot} \boldsymbol{x}_{j} \boldsymbol{\cdot} \boldsymbol{s}$ are the hit coordinates in the silicon layers.

•a_{ij} & b_i 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.

~few hundred fitting engines/trigger sector for CMS

Time multiplexing and data formatting

INFN 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.



AM technological evolution

XILINX[®]

3200E[™] 156AFS0025

- (90's) Full custom VLSI chip 0.7μm (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, 1st std cell project presented @ LHCC (1999)

Standard Cell 0.18 µm → 5000 pattern/AM chip SVT upgrade total: 6M pattern, 40MHz A. Annovi et al., IEEE TNS, Vol 53, Issue 4, Part 2, 2006

A. Annovi -2013

14th ICATPP Conference

System dimensioning

Pattern matching

Optimization on-going

- fixed super-strip (32 strips each) size for all layers: ~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 (μ , electrons)~99%

TIB 1 Before AM

 \bigcirc Purity of stubs after AM filtering ~60%

Further ~30% gain from stub pT info

		Dallk	Λ (in %)	
		Baseline	Combo	
Endcap	Muon gun efficiency (Nhits≥5)	98,6	98,4	-0,2
	Electron gun efficiency (Nhits≥5)	96,0	93,6	-2,5
	PU140 road rate	119,0	24,0	-79,8
	PU140 fake road proportion	52,8	31,7	-40,0
	PU200 road rate	216,0	38,0	-82,4
	PU200 fake road proportion	62,0	41,7	-32,7
	Muon gun efficiency (Nhits≥5)	98,5	97,5	-1,0
σ	Electron gun efficiency (Nhits≥5)	97,6	95,6	-2,0
Hybri	PU140 road rate	248,0	35,0	-85,9
	PU140 fake road proportion	73,7	48,4	-34,3
	PU200 road rate	534,0	61,0	-88,6
	PU200 fake road proportion	83,7	65,6	-21,6
	Muon gun efficiency (Nhits≥5)	99,0	98,6	-0,4
Barrel	Electron gun efficiency (Nhits≥5)	98,5	97,3	-1,2
	PU140 road rate	201,0	36,0	-82,1
	PU140 fake road proportion	62,4	24,9	-60,1
	PU200 road rate	416,0	51,0	-87,7
	PU200 fake road proportion	75,0	41,8	-44,3

stituto Nazionale di Fisica Nucleare

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

- AMchip04 power consumption / # of bit / MHz decreases of a factor ~28 w.r.t. AMchip03
- AMchip04 memory density (patterns*layers/area) increase of a factor ~18 w.r.t. AMchip03
- 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 mm²

				3	
/bank2k_60	. /bank2k_61			/bank2k_62	. /bank2k_63
/bank2k_54	/bank2k_55	/bank2k_56	. /bank2k_57	/bank2k_58	_/bank2k_59
Abank2k_48	hank2k_49	. /bank2k_50	/bank2k_51	/bank2k_52	Mank2k_53
., /bank2k_42	_/bank2k_43	"/bank2k_44	, /bank2k_45	/bank2k_46	_/bank2k_47
/bank2k_36	. /bank2k_37	. /bank2k_38	. /bank2k_39	Mank2k_40	M./bank2k_41
Mank2k_30	N/bank2k_31	N./bank2k_32	1./bank2k_33	M/bank2k_34	Mank2k_35
			4		
ManKZK_Z4	N/DanKZK_Z5	N/banKZK_Zb	/bank2k_2/	N/banKZK_28	X./banKZK_Z9
/bank2k_18	/bank2k_19	/bank2k_20	. /bank2k_21	_/bank2k_22	/bank2k_23
Abank2k_12	/bank2k_13	Mank2k_14	"/bank2k_15	M/bank2k_16	Mank2k_17
/bank2k_6	/bank2k_7	N./bank2k_8	. /bank2k_9	Mank2k_10	N/bank2k_11
.,∕bank2k_0	Mank2k_1	Mank2k_2	k,/bank2k_3	X./bank2k_4	Mank2k_5

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

AM Chip	Year	Density (No. Mbits)	Working Frequency (MHz)	Power (W)	Voltage (V)	Technology	Area (cm²)
AM03	2004	0.5	40	1.26	1.8	180 nm	1
AM04	2012	1.18	100	3.70	1.2	65 nm	0.12
AM06	2014	18.9	100	2-3	1.0/0.8	65 nm	1.6
AM2020	2020?	76?	200+	~3 @100 MHz	0.8	28 nm	?

- The AM2020 chip features
 - Assuming technology scaling would allow (65/28)² ~5.4 more density
 Conservatively target a factor 4x patterns ⇒ 0.5 M patterns
 - 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
 - Is 200 MHz is required to speed up I/O ? could double the power
 - Need more output buses ?
- 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

Expected performances (tracklet)

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

lstituto 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 & Track MET Primary Vertex and Track MET

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

- Track "MET"
 - Define L1 track-based missing transverse momentum from L1 tracks coming from primary vertex

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".

Trigger, Threshold	Algorithm	Rate reduction	Full eff. at the plateau	Comments
Single Muon, 20 GeV	Improved Pt, via track matching	~ 13 (η < 1)	~ 90 %	Tracker isolation may help further.
Single Electron, 20 GeV	Match with cluster	> 6 (current granularity) >10 (crystal granularity) (η < 1)	90 %	Tracker isolation can bring an additional factor of up to 2.
Single Tau, 40 GeV	CaloTau – track matching + tracker isolation	O(5)	O(50 %) (for 3-prong decays)	
Single Photon, 20 GeV	Tracker isolation	40 %	90 %	Probably hard to do much better.
Multi-jets, HT	Require that jets come from the same vertex			Performances depend a lot on the trigger & threshold.

Tracker 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
 - ●Full readout @40 MHz, on-detector data reduction using p_T-modules
 - Implications on Tracker detector layouts ongoing
 - Some demonstrators being built to validate the full chain
 - Large gains in combining tracking with other subdetectors
 ©Electrons, 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 LO accept, copy data from primary to secondary buffer
- Identify "region of interest" (1-10% of the detector on each L0 accept)
- Generate a "Regional Regional Request" (R3)
 - 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

Istituto Nazionale di Fisica Nucleare

Simulation results - ATLAS

Anatomy of a PRAM (Pattern Recognition Associative Memory)

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

Increasing the pattern density

AMCHIP04: VARIABLE RESOLUTION

ATLAS FTK

Inner Tracker Overview

- To deal with data flow designed as highly parallel system
 - 8 'core crate' with own pattern recognition and track fitter
 - Detector subdivided in 64 trigger tower
- PIX (3 layers) & SCT (4 double layers)
- Fit posses combinatorics problem, executed in two sequential steps:
 - Use 8 layer for patter recognition and 8 layer fit
 - Refit track found using all 11 layer

