Tracking and Trigger Upgrades to CMS

19 November 2014

University of Bristol
Brunel University
Imperial College London
Rutherford Appleton Laboratory

Geoff Hall
CMS: Compact Muon Solenoid

- CMS: Compact Muon Solenoid
- ECAL
- Tracker
- HCAL
- Muon chambers
- 4T solenoid
- Total weight: 12,500 t
- Overall diameter: 15 m
- Overall length: 21.6 m
- Magnetic field: 4 T
CMS Tracker and its sub-systems

- Two main sub-systems: Silicon Strip Tracker and Pixels
  - pixels quickly removable for beam-pipe bake-out or replacement
  - SST not replaceable in reasonable time

<table>
<thead>
<tr>
<th>Microstrip tracker</th>
<th>Pixels</th>
</tr>
</thead>
<tbody>
<tr>
<td>~210 m² of silicon, 9.3M channels</td>
<td>~1 m² of silicon, 66M channels</td>
</tr>
<tr>
<td>73k APV25s, 38k optical links, 440 FEDs</td>
<td>16k ROCs, 2k olinks, 40 FEDs</td>
</tr>
<tr>
<td>27 module types</td>
<td>8 module types</td>
</tr>
<tr>
<td>~34kW</td>
<td>~3.6kW (post-rad)</td>
</tr>
</tbody>
</table>

Geoff Hall
Current CMS L1 Trigger

- Pipelined trigger at 40 MHz
- Mix of FPGAs and ASICs
- Many copper parallel links
- Internal bandwidth constraints e.g. jet finding

- Detectors designed to accommodate:
  - 100 kHz maximum L1 rate
  - Max latency 4µs
CMS $\gamma\gamma$ event

J. Incandela
UCSB/CERN

CMS $ZZ \rightarrow ee\mu\mu$ candidate
Yields for $m(4\ell) = 110..160$ GeV

<table>
<thead>
<tr>
<th>Channel</th>
<th>$4e$</th>
<th>$4\mu$</th>
<th>$2e2\mu$</th>
<th>$4\ell$</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZZ background</td>
<td>$2.65 \pm 0.31$</td>
<td>$5.65 \pm 0.59$</td>
<td>$7.17 \pm 0.76$</td>
<td>$15.48 \pm 1.01$</td>
</tr>
<tr>
<td>$Z+X$</td>
<td>$1.20 \pm 0.78$</td>
<td>$0.92^{+0.65}_{-0.55}$</td>
<td>$2.29^{+1.81}_{-1.36}$</td>
<td>$4.41^{+2.21}_{-1.66}$</td>
</tr>
<tr>
<td>All backgrounds</td>
<td>$3.85^{+1.12}_{-0.84}$</td>
<td>$6.58^{+0.88}_{-0.81}$</td>
<td>$9.46^{+1.96}_{-1.56}$</td>
<td>$19.88^{+2.43}_{-1.95}$</td>
</tr>
</tbody>
</table>

$m_H = 126$ GeV

164 events expected in [100, 800 GeV]
172 events observed in [100, 800 GeV]

Discovery!
What next?

• Detector upgrades planned around essential shutdowns
  – LS1 – currently underway – collisions in 2015
    • upgrade energy to 13-14 TeV
  – LS2 – 18 months 2018-2019
    • collimation, injector and cryogenic upgrades
  – LS3 - 30 months 2023-2025
    • prepare for levelled high-luminosity running

• Some extra activities possible in Year End Technical Stops
  – e.g. probable extended 2016 YETS for CMS pixel installation

• Major upgrade essential for Run 3 (post-2025)
HL-LHC - goals

- Prepare machine for operation beyond 2025 and up to 2035
- Devise beam parameters and operation scenarios for:
  - total integrated luminosity of $3000 \text{ fb}^{-1}$ in around 10-12 years
  - an integrated luminosity of $250 \text{ fb}^{-1}$ per year
  - $\mu \leq 140$ (peak luminosity of $5 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$)
2010 - 2035

Peak luminosity
Integrated luminosity

Luminosity [cm$^{-2}$s$^{-1}$]

Integrated luminosity [fb$^{-1}$]

Year

LS3: HL-LHC upgrade – machine and experiments

5.0E+34

M Lamont CERN
Nov 2014

Geoff Hall
Motives for upgrades

• Many big physics questions outstanding
  – perhaps improving the Higgs precision will be the only way?
    • experimentally challenging – both for detectors and analyses
  – if SUSY is discovered in Run 2, may expect some revision of plans
    • if discoveries are absent should expect further theoretical ideas

• probably unwise to assume today’s experimental and theoretical landscape will be static
  – LHC machine and experiments have been running only three years
  – gains from software and analyses have been impressive

• Eventually important parts of detectors will be under great stress
  – radiation damage
  – data volumes and rates
  – performance improvement from technology evolution
GPDs: scope of Phase II detector upgrades

• Most sub-detectors are foreseen to survive to 3000 fb$^{-1}$
  – with on-going maintenance and refurbishment where possible

• Trackers must be completely replaced
  – radiation damage limits their lifetimes to <500 fb$^{-1}$

• New tracker readout systems are therefore also essential
  – based on more modern technologies, which improve performance
  – all sub-system readout systems must remain compatible

• Triggers must also be substantially upgraded
  – designed for $10^{34}$ cm$^{-2}$s$^{-1}$, $<N_{ev}>$~25
    • with safety factors – but exploited to maximise acceptance
Basic Requirements For ATLAS and CMS

- **Radiation hardness**
  - Ultimate integrated luminosity considered ~ 3000 fb\(^{-1}\) (original ~ 400 fb\(^{-1}\))
  - Radiation hard sensor material
  - New readout electronics required

- **Granularity**
  - Resolve 140-200 collisions per bunch crossing
  - Maintain occupancy below % level
  - Requires much higher granularity

- **Improve tracking performance**
  - Reduce material in the tracking volume
    - Improve performance at low \(p_t\)
    - Reduce rates of nuclear interaction, photon conversions, Bremsstrahlung…
  - Reduce average pitch
    - Improve performance at high \(p_t\)
CMS upgrade summary

**New Tracker**
- Radiation tolerant - high granularity - less material
- Tracks in hardware trigger (L1)
- Coverage up to $\eta \sim 4$

**Muons**
- Replace DT FE electronics
- Complete RPC coverage in forward region (new GEM/RPC technology)
- Investigate Muon-tagging up to $\eta \sim 3$

**Barrel ECAL**
- Replace FE electronics
- Cool detector/APDs

**Trigger/DAQ**
- L1 (hardware) with tracks and rate up $\sim 750$ kHz
- L1 Latency $<12.5$ $\mu$s
- HLT output rate 7.5 kHz

**Endcap Calorimeters**
- Radiation tolerant
- High granularity

**Other R&D**
- Fast-timing for in-time pileup suppression
- Pixel trigger
Need for Tracker replacement

Blue tracker modules are inactive after 1000 fb\(^{-1}\) due to very high leakage currents induced by hadron fluence.
Extensive R&D campaigns happened in all experiments defined to follow up.

- For ATLAS and CMS - Outer Tracker well defined
  - Common ATLAS & CMS Market Survey for Outer Tracker for AC-coupled sensors
- More studies necessary for inner pixel
  - Some common ATLAS/CMS wafer submissions planned

<table>
<thead>
<tr>
<th></th>
<th>Strips/strixel baseline</th>
<th>Pixel outer layers baseline / options</th>
<th>Pixel inner layers baseline / options</th>
<th>Special</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALICE</td>
<td></td>
<td>MAPS (Monolithic Active Pixels)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATLAS</td>
<td>• n-in-p planar</td>
<td>• n-in-p (n) planar and/or HR/HV-CMOS</td>
<td>• n-in-n planar 100-200μm active thickness and/or HR/HV-CMOS and/or 3D and/or diamonds</td>
<td></td>
</tr>
<tr>
<td></td>
<td>FZ 300μm thick AC-coupled and/or HV-CMOS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CMS</td>
<td>• n-in-p planar 200μm active thickness AC- and DC-coupled and/or MCz (pref) and/or 300 μm</td>
<td>• n-in-p planar 100-200μm active thickness</td>
<td>• n-in-p planar 100-200μm active thickness</td>
<td>HGCAL</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• p-in-n planar DC-coupled large PAD sensors 100-300μm active thickness (deep diffused)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• Or n-in-p (deep diffused)</td>
</tr>
<tr>
<td>LHCb</td>
<td>UT planar n-in-p or p-in-n</td>
<td>VELO planar n-in-p or n-in-n</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

With ~700m² needs to be factored in the global orders
New issues for trigger

• $L \sim 5 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$ (levelled) $\Rightarrow N_{\text{ev}}/\text{BX} \sim 140 - 200$

• Calorimeters
  – isolation of $e/\gamma/\tau$ degraded by pile-up from $\pi^0\gamma$s and hadrons
  – many more jets, which overlap

• Muon systems
  – increased combinatorial fakes, enhanced by multiple scattering

• Outcome: much higher rate of L1 triggers
  – usual response is to increase thresholds, which risks physics
  – even worse - raising thresholds does not look effective
Why tracker input to L1 trigger?

• Single $\mu$ and $e$ L1 trigger rates will greatly exceed 100kHz
  – similar behaviour for jets

Single electron trigger rate

$<p_T> \approx$ few GeV/bx/trigger tower

Isolation criteria alone are insufficient to reduce rate at $\mathcal{L} = 10^{35}$ cm$^{-2}$s$^{-1}$

$\mathcal{L} = 2 \times 10^{33}$

$5$kHz @ $10^{35}$

$\mathcal{L} = 10^{34}$ muon L1 trigger rate
Trigger levels

• Not feasible to achieve sufficient data reduction in L1 single step
  – 100 kHz -> ~500 kHz in Run 3

• When decisions were made on L2, two points of view
  – (custom) hardware processors needed to reduce data volume in ~50ms
  – sufficient computing power would evolve to avoid intermediate level

• ATLAS and CMS therefore diverged, with future implications
  – CMS must always store data on-detector until L1 decision
    • hardware trigger latency limited by shortest buffer length
    • transfer large data volume quickly to HLT = large BW
  – ATLAS can transfer selected data to L2 buffers
    • potentially much longer trigger latency possible
    • much smaller fraction of data, but more complexity

• Latency is an important constraint: target < 12µs
Silicon tracker with trigger-stub capability

Strip/Strip Modules
90 µm pitch/5 cm length

Strip/Pixel Modules
100 µm x 1.5 mm “macropixels”

Inner Pixel
Covers up to $\eta=4.0$
Stacked-tracker principle

• Compare pattern of hits in contiguous sensor elements in closely spaced layers
  – $p_T$ cut set by angle of track in layer
  – primarily depends on layer separation
    • but increasing separation worsens fake combinations
  – details depend on
    • pitch
    • thickness
    • charge sharing
    • track impact point
    • ...

![Graphs showing efficiency vs. $p_T$]
CMS Tracker ASIC evolution

- **1999: APV25 0.25µm**
  - 7 mm x 8mm (128 chan)

- **2011: CBC 0.13µm**
  - 7mm x 4mm (128 chan)

- **2013: CBC2 0.13µm**
  - 11mm x 5mm (254 chan)

**Programmable settings (now standard)**

**Analogue data**
- ~4 µs latency
- Wire-bondable

**Pulse-shaping choice**
- Binary data, 6.4 µs latency
- Wire-bondable

**Beam profile**
- APV plane
- CBC sensor

**Cluster & correlation logic**
- Bump-bondable
• ~15000 modules transmitting
  – $p_T$-stubs to L1 trigger @ 40 MHz
  – full hit data to HLT @ 0.5-1 MHz

~8400 2S-modules

CMS Outer Tracker trigger

EXP. $p_T$ cut: 2.14 GeV/c
FIT: $p_T$ cut: 2.17 GeV/c
$\sigma$: 0.1 GeV/c

reconstructed $p_T$ cut of
r=75cm layer
Efficiency, resolution and fake rate

Pixel used in simulation results to date is identical to the Phase 1 Pixel detector with additional forward disks.
Further optimization of pixel parameters for b-tagging and forward track parameter resolution is planned.
Improvements To Current Trigger

Run 2 will already require an improved trigger because of energy and luminosity increases.

Have not yet exploited fully potential improvements from present detector

  e.g. 10 years ago technology limited data transfer rate which enforced some data suppression
  now 10 Gbps optical links are standard,
  along with fast, flexible, more powerful processing
Motivation and requirements

- **Motivation**
  - Trigger rates are driven by the increase in luminosity, the centre-of-mass energy, and by the higher PU (especially hadronic objects).
  - CMS detector electronics are limited to a L1 trigger rate of 100 kHz. It will be a major challenge to maintain physics acceptance within this rate.
    - Example: factor of 3 in inst. lumi x factor of 2 from CoM energy → 6 times the rate compared to 2012 → single lepton trigger thresholds ~50 GeV

- **Requirements**
  - Maintain sensitivity for electroweak scale physics and for TeV scale searches similar to what was achieved before LS1.
  - Making significant changes to the trigger system could put CMS data-taking at risk. The plan must allow parallel commissioning of the new trigger to mitigate this risk.
Calorimeter Algorithms

- **Electron/photon**
  - Large deposition of energy in small region, well separated from neighbour
  - pileup worsens the separation for lower $p_T$ objects

- **jets**
  - hadrons – large, likely overlapping objects
  - $\tau$ - isolated irregular, narrow energy deposits
  - simulations identify likely patterns to accept or veto
Table 2.1: Extrapolated trigger rates for several current Level-1 Triggers and thresholds, for a range of instantaneous luminosity (given in \( \times 10^{34} \, \text{cm}^{-2}\text{s}^{-1} \)) and pile-up \((\langle PU \rangle)\). The rate in the third column used a typical run in 2012, and the next two columns are determined using the high pile-up fills of 2012 at 8 TeV. There is no scaling for higher beam energy. The final column uses a Monte Carlo sample to account for the increased beam energy and out-of-time pile-up. The second column gives the offline \( E_T \) (or \( p_T \)) values at which the trigger reaches 95% efficiency, to enable a fair comparison with menus presented later.

<table>
<thead>
<tr>
<th>Level-1 Trigger, 2012 Threshold</th>
<th>95% Threshold [GeV]</th>
<th>Rate [kHz] ( L = 0.4 \langle PU \rangle=15 )</th>
<th>Rate [kHz] ( L = 1.1 \langle PU \rangle=45 )</th>
<th>Rate [kHz] ( L = 1.6 \langle PU \rangle=66 )</th>
<th>Rate [kHz] ( L = 1.1 ) (14 TeV) ( \langle PU \rangle=50 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single iso e/( \gamma ), 18 GeV,</td>
<td>26</td>
<td>6.3</td>
<td>19</td>
<td>27</td>
<td>40</td>
</tr>
<tr>
<td>(</td>
<td>\eta</td>
<td>&lt; 2.1 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single Mu, 14 GeV,</td>
<td>19</td>
<td>4.1</td>
<td>11</td>
<td>15</td>
<td>27</td>
</tr>
<tr>
<td>(</td>
<td>\eta</td>
<td>&lt; 2.1 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single Jet, 128 GeV</td>
<td>150</td>
<td>1.0</td>
<td>3.6</td>
<td>6.6</td>
<td>14</td>
</tr>
<tr>
<td>( H_T ), 150 GeV</td>
<td>280</td>
<td>0.9</td>
<td>55</td>
<td>( \sim 10^3 )</td>
<td>110</td>
</tr>
<tr>
<td>Double e/( \gamma ), 13, 7 GeV</td>
<td>20,13</td>
<td>5.4</td>
<td>15.4</td>
<td>24</td>
<td>47</td>
</tr>
</tbody>
</table>
Requirements

- Identified the following areas for upgrade:
  - Improved electromagnetic object isolation using calorimeter energy distributions with pile-up subtraction
  - Improved jet finding with pile-up subtraction
  - Improved hadronic tau identification with a much narrower cone
  - Improved muon $p_T$ resolution
  - Isolation of muons using calorimeter energy distributions with pile-up subtraction
  - Improved global Level-1 trigger menu with a greater number of triggers and with more sophisticated relations involving the input objects
Improvements to e-$\gamma$-$\tau$
Calorimeter trigger: MP7

- IN
- 72 x 12.5 Gbps
  = 0.9 Tbps
- OUT
- 72 x 12.5 Gbps
  = 0.9 Tbps
- ~80W
New Trigger Architecture

Conventional Trigger

Time Multiplexed Trigger

TPG
Regional stages
Global stages
L1 decision

TPG
MUX
TMT processors
GT - ?

L1 decision
What is a Time Multiplexed Trigger?

• Multiple sources send to single destination for complete event processing
  – as used, e.g., in CMS High Level Trigger

• Requires two layers with passive switching network between them
  – can be “simple” optical fibre network
  – could involve data processing at both layers
  – could also be data organisation and formatting at Layer 1, followed by data transmission to Layer 2, with event processing at Layer 2

  – illustration on next slide
Time-multiplexing

All data for 1bx from all regions in a single card! Everything you need!

All data for 1bx from all regions in a single card! Everything you need!

All data for 1bx from all regions in a single card! Everything you need!

BX:7
What are advantages of TMT?

• “All” the data arrive at a single place for processing
  – in ideal case avoids boundaries and sharing between processors
  – however, does not preclude sub-division of detector into regions
• Architecture is naturally matched to FPGA processing
  – parallel streams with pipelined steps at data link speed
• Single type of processor, possibly for both layers
  – $L_1 = \text{PP: Pre-Processor}$  $L_2 = \text{MP: Main Processor}$
• One or two nodes can validate an entire trigger
  – spare nodes can be used for redundancy, or algorithm development
• Many conventional algorithms explode in a large FPGA
  – timing constraints or routing congestion for 2D algorithms
• Synchronisation is required only in a single node
  – not across entire trigger
Calorimeter trigger current status

• Currently being installed at LHC Point 5 in CMS for parallel operation with the existing trigger during 2015
  – following a series of demonstrators
    • time-multiplexed data transfer from layer-1 to layer-2
    • pseudo-random data and emulated physics data with pileup
    • implementation of realistic algorithms in firmware
    • bit-level confirmation of algorithm operation – software vs firmware
    • verification of latency estimates

• Huge FPGA + huge IO = all problems solved?
  – not quite so simple...
A note on Xilinx 7-series FPGAs

- Large 7-series FPGAs are beasts!
- Scale of routing problem has outstripped performance of PCs
- Xilinx has invested heavily in improving tools
- Still have to work a lot harder at optimising designs than we used to

![Diagram showing Bandwidth and Logic trends from 2006 to 2012]
CMS Calo TMT demonstrator (Sep 2013)

MP7 used as PP & MP

- TPG input to PP not part of test
- oSLB
- uHTR

Test set-up @ 904

Simulating half of the PP cards with a single MP7

Simulating half of the PP cards with a single MP7
Status of TMT jet algorithms

- **Jets**
  - 9×9 sum of trigger towers at every site
  - Fully asymmetric jet veto calculation
  - Local (“Donut”) or Global pile-up estimation
  - Full overlap filtering
  - Pile-up subtraction
  - Pipelined sort of candidates in $\phi$
  - Accumulating pipelined sort of candidates in $\eta$

- **Ring sums**
  - Scalar and Vector (“Missing”) ET
  - Scalar and Vector (“Missing”) HT

9×9 jet at tower-level resolution

50% LUT utilization **INCLUDING** links, buffers, control, DAQ, etc. Runs at 240 MHz
Results (from September test)

- Implementation of an algorithm and successful transmission of data through it
- Random data passed through an emulator was used in the testing of the algorithms

Data injected into PP → Time-multiplexed → Optical Fibre → Circle jet algorithm (8x8) → Sort → Capture

Compared emulated results (solid line) with those from the MP7 (markers)

C++ emulator and hardware match precisely
**Results – Latency Measurement**

- Verification of latency and how it compares to TDR value - in particular the SerDes link

<table>
<thead>
<tr>
<th>Source of Latency</th>
<th>BX (TDR)</th>
<th>BX - measured in Sept 2013</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1 processing + TM</td>
<td>10</td>
<td>7</td>
</tr>
<tr>
<td>L1/L2 SerDes (Tx+Rx) @ 10Gbps</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>L1/L2 SerDes Align Data</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>L1/L2 cable (20m)</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>L2 Processing</td>
<td>8</td>
<td>5.5 (clustering, jets, ring sums)</td>
</tr>
<tr>
<td>L2/GT SerDes (Tx+Rx)</td>
<td>5.5</td>
<td>5 (identical link to L1/L2 above)</td>
</tr>
<tr>
<td>L2/GT SerDes Align Data</td>
<td>1</td>
<td>1 (identical link to L1/L2 above)</td>
</tr>
<tr>
<td>L2/GT cable</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>De-multiplex</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>41</strong></td>
<td><strong>36</strong></td>
</tr>
</tbody>
</table>
A simple example of Routing Congestion: 1

• (G Iles) Created simple design to find routing limit
  – 30x36 2x2 tower clusters (“electrons”) with 10bit energy
  – 432 Gb/s (without 8B/10B)
  • Approximately ¾ of CMS
  – Sum 16 clusters to create “pseudojets”
  – No other firmware (e.g. no sort, no transceivers, no DAQ, etc)
  – XC7VX485T – Place & Route fails even though LUT usage only at 29%

but number of LUTs is not the whole story…

A bigger FPGA may not solve all the problems…

Route:463 - The router has detected a very dense, congested design. It is extremely unlikely the router will be able to find an optimal solution within the excessive run time the router will exit with a partially routed design. This behavior will allow you to identify difficult design constraints, putting too much logic into this device, or an issue with the implementation. If you would prefer a more optimal solution, you may want to change some logic from the design, or change the placement before running router again. If you are willing to accept a long run time behavior.

Route:543 - This design is experiencing routing congestion. Please review the Xilinx Routing Optimization White Paper, in resolving this issue.
• (G Iles) Implemented a proposed circular isolation algorithm
  – using pipelined design
• Searches every tower location in 56 x 72 region
  – 4032 sites
• Counts the number of objects above threshold within a circular ring of diameter 9 towers or clusters
  – Result passed into LUT with the energy to determine object status

Operates up to 400MHz
Compact: < 1% of the FPGA
Low latency - 9 clks (no overlap)
1.5 BX @ 240 MHz

* Only synthesised 36 towers in eta, rather than 56, but in the small FPGA
Why was Time Multiplexed Trigger not already used?

• Mainly technology limitations
  – It is reliant on high performance hardware
    • large & powerful FPGAs
    • many high speed (optical) links

• More recent objections to latency penalty in L1-L2 transmission
  – but this is mostly a myth!
  – If properly organised, data processing does not need to wait for entire event data.
  – It can begin as soon as first cycle’s worth of data arrive
The track-trigger challenge

- Impossible to transfer all data off-detector for decision logic so on-detector data reduction (or selective readout) essential
  - The hit density means high combinatorial background
  - 99% tracks < 2 GeV/c
  - after ~2 GeV/c cut,
    ~60-70% hits associated with real track

- What are track-trigger requirements?
  - original assumption: a few points would help
  - but simulations did not show big rate reductions
  - hence quasi-full track reconstruction for $p_T >$~3 GeV
  - separation of primary vertices is required

- application in trigger based on plausible simulations
  - i.e. 2025 trigger not well defined
  - isolation, pileup subtraction, vertex association, object pointing,....
Possible layout of CMS TM Track-Trigger
split tracker into **phi** regions

constrained problem by looking at **minimum** number of trigger regions (TRs) required, and imposed constraint that **one module cannot be shared across more than two TRs**

- 5 TRs in phi only
- 1 GeV/c boundary region assumed
- e.g. could allow for better reco @ 2GeV/c in case of $e^+/e^-$, bremsstrahlung, low $p_T$, multiple scattering etc.
the time multiplex period is not a completely free parameter

<table>
<thead>
<tr>
<th>small TM period</th>
<th>large TM period</th>
</tr>
</thead>
<tbody>
<tr>
<td>full event must be quickly assembled into one MP</td>
<td>could allow more efficient processing of pipelined data into MP</td>
</tr>
<tr>
<td>reduces data volume per event from PP to MP (or requires increased number of links)</td>
<td>increases data volume per event from PP to MP (or reduces number of links)</td>
</tr>
<tr>
<td>reduces latency</td>
<td>increases latency</td>
</tr>
<tr>
<td>reduces number of MPs</td>
<td>increases number of MPs</td>
</tr>
<tr>
<td>min ~15bx (PP output bandwidth without more Trigger Regions)</td>
<td>max ~34bx (68 links/2 Trigger Regions)</td>
</tr>
</tbody>
</table>

**TM period of 24BX chosen for case study (could be optimised in future)**
from non-shared modules

68 FE links
3.2Gbps per link

4 bidirectional DAQ links
10Gbps per link

to one TR

24 TRG links
10Gbps per link

from shared (boundary) modules

68 FE links
3.2Gbps per link

4 bidirectional DAQ links
10Gbps per link

to two TRs - 24 TRG links
to each

48 TRG links
10Gbps per link

4 DAQ links per PP/FED allows a maximum bandwidth of 40Gbps (~588Mbps available per tracker module)
Organisation for 5 TRs in Phi & 24BX TM Period

<table>
<thead>
<tr>
<th># FE links</th>
<th>1865</th>
<th>1194</th>
<th>1930</th>
<th>1156</th>
<th>1944</th>
<th>1102</th>
<th>1954</th>
<th>1170</th>
<th>1865</th>
<th>1328</th>
</tr>
</thead>
<tbody>
<tr>
<td># PP/FEDs</td>
<td>28</td>
<td>18</td>
<td>29</td>
<td>17</td>
<td>29</td>
<td>17</td>
<td>29</td>
<td>18</td>
<td>28</td>
<td>20</td>
</tr>
<tr>
<td># PP-&gt;MP links</td>
<td>672</td>
<td>432</td>
<td>432</td>
<td>696</td>
<td>408</td>
<td>408</td>
<td>408</td>
<td>696</td>
<td>432</td>
<td>480</td>
</tr>
<tr>
<td># PP links/MP</td>
<td>66</td>
<td>64</td>
<td>63</td>
<td>64</td>
<td>66</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td># PP-&gt;MP links total</td>
<td>1584</td>
<td>1536</td>
<td>1512</td>
<td>1536</td>
<td>1584</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td># TM nodes</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(numbers from tkLayout...)

Geoff Hall
full tracker L1 and trigger data can be read out with a total of 353 MP7s

after 24BX, all trigger data from tracker from first event are assembled into 5 MPs  
   each MP corresponds to a trigger region in phi  
   But processing should have started earlier than BX 25

subsequent events are available in next set of 5 MPs after one extra BX

no boundary sharing is required after duplication in PP;  
   no post-TM removal of duplicates necessary

What processing is possible in MPs?  
   track-finding? track fitting? data processing before an AM stage?
2 MP7s emulate event data from 1 out of 5 regions, one out of every 24BX

- This demonstrator already exists
  - just need to program source data to be ready to try algorithms
Firmware design

- Still to establish the best way of finding tracks at HL-LHC
  - latency and efficiency, as well as (firmware) programming challenges
  - we know from MP7 that large FPGAs present serious challenges, e.g.:
    - exceeding RAM resources
    - logic fails to synthesize within timing constraints after many hours

- Possible approach based on Hough transform
  - locate series of hits on a trajectory, for which \( y = mx + c \)
    - For fixed \((m,c)\): every “y” corresponds to single “x”
    - For fixed \((x,y)\): every “c” corresponds to single “m”
  - point \((m,c)\) -> line \((x,y)\)  point \((x,y)\) -> line \((m,c)\)
    - All hits from a real track have same \((m,c)\)
  - For each data point \((x,y)\), hypothesize “m” and calculate “c”
    - When multiple hits have same \((m,c)\), send for fitting
    - identify by histogramming entries into array
Possible implementation

pipelined systolic array as proposed by Andy Rose

stub enters array, close to entry row and finds its way
Summary

• The TMT is now a proven architecture in CMS
  – which will operate in the CMS calorimeter trigger from 2016

• The hardware is very flexible and can be deployed for a TMTT
  – only a fraction of the system is required to validate the concept
  – installing and building should only require replicating identical nodes
    • a track-trigger could be built using present technology
      – safe to assume further technological progress in next decade

• The next challenge is to prove algorithms can be implemented
  – under way
Backup
<table>
<thead>
<tr>
<th>Trigger Algorithm</th>
<th>Current Level-1 ( L = 1.1 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1} )</th>
<th>Current Level-1 ( L = 2.2 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1} )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rate [kHz] 95% Threshold [GeV] Plateau Efficiency</td>
<td>Rate [kHz] 95% Threshold [GeV] Plateau Efficiency</td>
</tr>
<tr>
<td>Single e/( \gamma )</td>
<td>12 46 1.0</td>
<td>10 67 1.0</td>
</tr>
<tr>
<td>Single iso e/( \gamma )</td>
<td>10 38 0.9</td>
<td>9.4 52 0.9</td>
</tr>
<tr>
<td>Single Mu</td>
<td>12 23 0.95</td>
<td>11 42 0.95</td>
</tr>
<tr>
<td>Single isoTau</td>
<td>10 65 0.3</td>
<td>9.2 72 0.3</td>
</tr>
<tr>
<td>iso e/( \gamma ) + e/( \gamma )</td>
<td>10 24 15 0.9</td>
<td>16 26 16 0.9</td>
</tr>
<tr>
<td>Mu + Mu</td>
<td>6.3 18 0.9</td>
<td>7.4 20 12 0.9</td>
</tr>
<tr>
<td>Tau + Tau</td>
<td>7.5 36 0.1</td>
<td>8.2 36 36 0.1</td>
</tr>
<tr>
<td>iso e/( \gamma ) + Mu</td>
<td>9.6 21 0.85</td>
<td>6.2 24 12 0.85</td>
</tr>
<tr>
<td>Mu + e/( \gamma )</td>
<td>3.3 18 0.95</td>
<td>5.0 20 15 0.95</td>
</tr>
<tr>
<td>Single Jet</td>
<td>6.4 170 1.0</td>
<td>5.4 205 1.0</td>
</tr>
<tr>
<td>Double Jet</td>
<td>4.6 140 140 1.0</td>
<td>5.8 170 170 1.0</td>
</tr>
<tr>
<td>Quad Jet</td>
<td>9.4 4@71 1.0</td>
<td>4.8 4@96 1.0</td>
</tr>
<tr>
<td>Single iso e/( \gamma ) + Jet</td>
<td>7.5 32 68 0.9</td>
<td>8.5 38 82 0.9</td>
</tr>
<tr>
<td>Single Mu + Jet</td>
<td>8.6 22 0.9</td>
<td>7.5 27 54 0.9</td>
</tr>
<tr>
<td>Single iso e/( \gamma ) + ( H_T )miss</td>
<td>10 29 110 0.9</td>
<td>8.2 38 120 0.9</td>
</tr>
<tr>
<td>Single Mu + ( H_T )miss</td>
<td>4.6 18 89 0.95</td>
<td>9.8 20 93 0.95</td>
</tr>
<tr>
<td>( H_T )</td>
<td>3.9 500 1.0</td>
<td>5.4 580 1.0</td>
</tr>
<tr>
<td>Total Rate</td>
<td>94</td>
<td>92</td>
</tr>
<tr>
<td>Trigger Algorithm</td>
<td>Rate [kHz]</td>
<td>95% Threshold [GeV]</td>
</tr>
<tr>
<td>---------------------------</td>
<td>------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>Single e/γ</td>
<td>10</td>
<td>67</td>
</tr>
<tr>
<td>Single iso e/γ</td>
<td>9.4</td>
<td>52</td>
</tr>
<tr>
<td>Single Mu</td>
<td>11</td>
<td>42</td>
</tr>
<tr>
<td>Single iso Mu</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Single Tau</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Single iso Tau</td>
<td>9.2</td>
<td>72</td>
</tr>
<tr>
<td>iso e/γ + e/γ</td>
<td>16</td>
<td>26 16</td>
</tr>
<tr>
<td>(iso)Mu + Mu</td>
<td>7.4</td>
<td>20 12</td>
</tr>
<tr>
<td>(iso)Tau + Tau</td>
<td>8.2</td>
<td>36 36</td>
</tr>
<tr>
<td>iso e/γ + Mu</td>
<td>6.2</td>
<td>24 12</td>
</tr>
<tr>
<td>(iso)Mu + e/γ</td>
<td>5.0</td>
<td>20 15</td>
</tr>
<tr>
<td>iso e/γ + Tau</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>isoMu + Tau</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Single Jet</td>
<td>5.4</td>
<td>205</td>
</tr>
<tr>
<td>Double Jet</td>
<td>5.8</td>
<td>170 170</td>
</tr>
<tr>
<td>Quad Jet</td>
<td>4.8</td>
<td>4@96</td>
</tr>
<tr>
<td>Single iso e/γ + Jet</td>
<td>8.5</td>
<td>38 82</td>
</tr>
<tr>
<td>Single Mu + Jet</td>
<td>7.5</td>
<td>27 54</td>
</tr>
<tr>
<td>Single iso e/γ + H_T^{miss}</td>
<td>8.2</td>
<td>38 120</td>
</tr>
<tr>
<td>Single Mu + H_T^{miss}</td>
<td>9.8</td>
<td>20 93</td>
</tr>
<tr>
<td>H_T</td>
<td>5.4</td>
<td>580</td>
</tr>
<tr>
<td>Total Rate</td>
<td>92</td>
<td></td>
</tr>
</tbody>
</table>
Synchronisation & Structure

- Synchronization is only required per-time-node, not across the whole trigger
  - De-synchronization of a node only affects that node.
  - A timing glitch in a CT disrupts the whole trigger.

- The efficient pipelined logic should lead to lower latency, and the eventual clock speed can be as fast as the FPGA allows.

- Firmware build times should be significantly shorter due to the pipelined, as opposed to combinatorial, nature of the architecture.
first wafer probed manually
<table>
<thead>
<tr>
<th>Layer</th>
<th>Total</th>
<th>Genuine</th>
<th>Combin.</th>
<th>Unknown</th>
</tr>
</thead>
<tbody>
<tr>
<td>Layer B1</td>
<td>2500</td>
<td>1573 (62.9%)</td>
<td>788 (31.5%)</td>
<td>139 (5.6%)</td>
</tr>
<tr>
<td>Layer B2</td>
<td>1778</td>
<td>1350 (75.9%)</td>
<td>243 (13.7%)</td>
<td>185 (10.4%)</td>
</tr>
<tr>
<td>Layer B3</td>
<td>1263</td>
<td>947 (75.0%)</td>
<td>110 (8.7%)</td>
<td>206 (16.3%)</td>
</tr>
<tr>
<td>Layer B4</td>
<td>818</td>
<td>557 (68.1%)</td>
<td>68 (8.3%)</td>
<td>193 (23.6%)</td>
</tr>
<tr>
<td>Layer B5</td>
<td>642</td>
<td>430 (70.0%)</td>
<td>27 (4.2%)</td>
<td>185 (28.8%)</td>
</tr>
<tr>
<td>Layer B6</td>
<td>520</td>
<td>327 (62.9%)</td>
<td>14 (2.7%)</td>
<td>179 (34.4%)</td>
</tr>
<tr>
<td>Layer E1</td>
<td>2326</td>
<td>1795 (77.2%)</td>
<td>468 (20.1%)</td>
<td>63 (2.7%)</td>
</tr>
<tr>
<td>Layer E2</td>
<td>899</td>
<td>669 (74.4%)</td>
<td>178 (19.8%)</td>
<td>52 (5.8%)</td>
</tr>
<tr>
<td>Layer E3</td>
<td>493</td>
<td>351 (71.2%)</td>
<td>97 (19.7%)</td>
<td>45 (9.1%)</td>
</tr>
<tr>
<td>Layer E4</td>
<td>469</td>
<td>297 (63.4%)</td>
<td>109 (23.2%)</td>
<td>63 (13.4%)</td>
</tr>
<tr>
<td>Layer E5</td>
<td>342</td>
<td>239 (69.9%)</td>
<td>37 (10.8%)</td>
<td>66 (19.3%)</td>
</tr>
<tr>
<td>Layer E6</td>
<td>265</td>
<td>179 (67.6%)</td>
<td>17 (6.4%)</td>
<td>69 (26.0%)</td>
</tr>
</tbody>
</table>

**Table 3.7:** Average total number of stubs produced per event in each barrel and endcap layer, as well as the number and proportion of the total of each class of stub, averaged over 100 events using the single seed stub stub producer.