ATLAS challenging trigger in the High Luminosity LHC

F. Pastore
past, present & future

1964

2013

~2030
The Higgs discovery at 125 GeV

- ATLAS and CMS verified two fingerprints of the Higgs Boson

- Its mass value at 125 GeV:
  - Opens access to many Higgs decay channels, and then to the measurements of its coupling constants
  - Push performance of our detectors since Higgs width is only few MeV
LHC for future frontiers of particle physics

- The existence of a Boson at low mass scale sets a limit to the validity of the SM (around the TeV scale?) and its extension must answer still open questions
  - mass hierarchy, cosmological questions (DM, DE, inflation), gravity and CP violation

- The key answers are hidden in the properties of the Higgs boson...
  - spin, self-interaction, multiplicity,... to be studied through the interactions with other particles

- Standard Model is completed!
  - We have no evidence of New Physics!

- ...and in the deep exploration of the TeV scale and of rare processes
  - investigating B-physics, top decays, gauge bosons scattering

- Precise measurements and rare process discovery need at least 3000 fb⁻¹ of data
  - ATLAS and CMS collected 30 fb⁻¹ so far

- European Council: “CERN is the strong European focal point for particle physics in next 20 years”
What can we do with the Higgs Factory?


![ATLAS Simulation](https://example.com/atlas_simulation.png)

**3000 fb⁻¹!**

More than 3M Higgs events for precise measurements (≥ ILC/CLIC/TLEP)

- SM energy scale
- Lepton signatures
- Increased importance of tau and b-quark selections

- Forward jets give clear signature (possible extension of trackers to |eta|<4)
- Measure Higgs self-coupling (giving access to lambda)
- Verify that the Higgs boson fixes the SM problems with W/Z scattering at high energy

- Measurement of as many Higgs coupling as possible
  - Increase precision on already observed
  - Access to rare processes (H->μμ, ttH→ttγγ)
LHC becoming impressively luminous: HL-LHC

- LHC plans for next 10 years are approved (LS1 and LS2): next HL-LHC starts in 2024
- New project to upgrade large part of the accelerator complex
  - Linac4, Booster, SPS, Interaction regions
- Collect 300 fb⁻¹/year, peak luminosity increases by factor 5 w.r.t. the design value (+ Luminosity leveling)

**LS1**: Shut down for interconnects to overcome energy limitation (LHC incident of Sept. 2008) and consolidation

**LS2**: Shut down to overcome beam intensity limitation (injectors, collimation and more...)

**LS3**: Full upgrade (new magnet technology for the IR, new bigger quadrupoles) to 5x10³⁴, to reach 3000 fb⁻¹ by 2030’s

**LS4**

**LS5**

**one of the latest upgrade plans...**
Agreed on Monday Dec. 2nd 2013 between LHC, experiments and CERN management based on input from ECFA meeting and RLIUP (Review of LHC & Injector Upgrade Plans Workshop)
Ls1 is on schedule: http://cern.ch/ls1dashboard
Well, a dirty Higgs factory!

- HL-LHC: 25 ns bunch crossing, $L=5 \times 10^{34}$ cm$^{-2}$s$^{-1}$

- Higher luminosity is reached by increasing the number of interactions/collision, but new future techniques (leveling, crab cavities, crab kissing...?) can modify the interaction region and help in maintaining the number of overlapping interactions low

  - Pessimistic view: **experiments must deal with <140> interactions per collision, with tails of 200**

- Detectors requirements will go **beyond the current design** specifications:
  - Higher **peak luminosity** means increased density of interactions in space and time and higher detector occupancy: need higher resolutions
  - Higher **integrated luminosity** pose limits of irradiation damage and activation of materials
What ATLAS will change for HL-LHC

Phase-II Lol: https://cds.cern.ch/record/1502664?ln=en

- **Inner detector components** (R<1 m) will suffer from radiation damage and high occupancy:
  - New silicon tracker, current one would not survive
  - New calorimeter FE electronics

- **Outer detector components** (R>1 m) will suffer from pile-up and high occupancy -→ reduced single sensor size, with consequent higher granularity, increased redundancy and faster time response
  - Some new **muon chambers** (inner endcaps) will be installed already in LS2

**1 MeV neutron equivalent fluence**

- expected fluence at 14TeV
  - 3000 fb⁻¹
  - Inner Tracker Region

*Figure showing a 3D diagram of the ATLAS detector with highlighted changes for HL-LHC.*
Higher trigger rates will impose a new design of the trigger and DAQ system (TDAQ).
Real life.... another reason to “upgrade”

- A lot of hardware components become old
- System reliability decreases
  - It makes sense to replace PCs and network equipments every 5 years
  - Custom hardware is usually kept longer... by of course it also starts breaking

![Graph showing general behavior of hardware components with failure rates over time.](#)
Cost of all LHC upgrades

Phase-I: minor upgrade 36 MCHF
Phase-II: major upgrade 275 MCHF
The silicon trackers evolution
Inner Tracker: key issues for the Upgrades

- **Radiation damage** with more integrated Luminosity, observed in ATLAS, CMS and LHCb (RD50 R&D project)
  - Increase in leakage current and S/N degradation
  - Projections demonstrate that the tracker will survive 500 fb⁻¹ if operated at -20°C after LS1
  - Must replace the full tracker after LS3

- **Increased performance**
  - Higher granularity
  - Lower material budget

- **Control and minimize cost**
  - Large areas & stable/timely production
The quadrature of vertex detectors

In p-p environments, the high level of radiation and hit occupancy imposes struggling requirements.

Adding features and performance means increasing the number of chips, then power consumption and additional material.
The future all-silicon Inner TracKer at HL-LHC

- Full silicon tracker: barrel cylinders and endcap disks, with different granularity
- Baseline layout to maintain optimal tracking performance (and cost)

Expected material budget for baseline design

Robust tracking:
- total of 14 hits with full coverage to $\eta=2.5$
- Pixels to $\eta<2.7$ (forward muon ID)
- Expected hit occupancy: everywhere less than 1%
Expected performance of the baseline layout

- Average 14 hits per track
- Occupancy < 1%

But other layouts are under study

- Efficiency for low and high $p_T$ regimes
- Momentum resolution

Phase-II Letter of Intent
ATL-COM-UPGRADE-2012-040
All-silicon sensors evolution (few words)

- **Planar silicon-sensors**
  - n-in-p: Single-sided process (less expensive)
  - n+-in-n: Double-sided (more expensive)
  - Both can work at HL-LHC radiation levels
    - If carefully designed...
    - And if they are kept cold ~-20°C

- **3D sensors**
  - Very good performance at high fluences
  - Production time and complexity to be investigated for larger scale production
  - Used in ATLAS IBL (LS1 upgrade)

- **CMOS sensors**
  - Contain sensor and electronics combined in one chip
  - Standard CMOS processing (many foundries, lower cost/area)
  - Prominent advantage: high granularity, low material, high data throughput

<table>
<thead>
<tr>
<th>Upgrades</th>
<th>Area</th>
<th>Baseline sensor type</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALICE ITS</td>
<td>10.3m²</td>
<td>CMOS</td>
</tr>
<tr>
<td>ATLAS Pixel</td>
<td>8.2m²</td>
<td>tbd</td>
</tr>
<tr>
<td>ATLAS Strips</td>
<td>193m²</td>
<td>n-in-p</td>
</tr>
<tr>
<td>CMS Pixel</td>
<td>4.6m²</td>
<td>tbd</td>
</tr>
<tr>
<td>CMS Strips</td>
<td>218m²</td>
<td>n-in-p</td>
</tr>
<tr>
<td>LHCb VELO</td>
<td>0.15m²</td>
<td>tbd</td>
</tr>
<tr>
<td>LHCb UT</td>
<td>5m²</td>
<td>n-in-p</td>
</tr>
</tbody>
</table>
Time evolution of highly segmented silicon detectors
Robustness: detector modules are integrated and fully functional packages, called staves, that can be produced in parallel and fully tested before assembly.

Reduce material: services are included in the module (cooling, monitoring, control,...)
- outer pixel layers have 5mm thick staves, with modules on both sides (n-in-p)
- inner pixel layers have I-beam shape and clamshell design (n+-in-n sensors)

Data links challenge the high radiation level and the high data throughput (~1Gbps)
- pixels use twisted micro-cables to send LVDS data to a dedicated optical board
- strips use dedicated optical links (CERN Versatile) up to 5Gb/s

Commercial copper cables can transmit several Gb/s over tens of meters. However, the diameters of these cables are too large for the pixel detector.
The trigger upgrade strategy
Legacy from today’s trigger system

First-level trigger (L1)
- Synchronous at 40MHz, with fixed latency: \(2.5 \mu s\)
- Identifies Region-of-Interest (RoI) in the muon spectrometer and/or in the calorimeter, with coarse resolution
- No tracking information can be used due to limited latency

High-level trigger (HLT)
- Handles complexity with custom fast software on commercial CPUs
- Accessing the full resolution of all the detectors (both RoIs and full event)

Event display of a 2-tau event in the ATLAS detector. Run number: 204153, Event number: 35369265.
The taus decay into an electron (blue line) and a muon (red line).
Expected trigger rates at HL-LHC

\[ R = \sigma_{in} \times L \]

Event rate = 0.8GHz @ 40MHz

\[ \div 400 \]

Level-1 = 100 kHz

\[ \div 300 \]

On-tape = 300 Hz

2013 LHC

L = 1 \times 10^{34} / \text{cm}^2 / \text{s}

2024 HL-LHC

L = 5 \times 10^{34} / \text{cm}^2 / \text{s}

Event rate = 4GHz @ 40MHz

\[ \div 40 \]

Level-1 = 0.5/1 MHz

\[ \div 100 \]

On-tape = 5-10 kHz

- Change of FE buffer size
- Maintain ~same rejection factor on HLT
- Event size will increase: 1.5MB to 2MB
- Challenging storage: 10-20 GB/s
- Moore’s law can handle this!
But in one ATLAS event at High-Luminosity ($L=5\times10^{34} \text{ cm}^2/\text{s}$)

- 200 minimum bias collisions per 25 ns bunch crossing
- ~10 000 particles per event
- Mostly low momentum ($p_T$) particles due to low transfer energy interactions
The trigger selection will become harder and harder

**Strategy:** maintain adequately wide trigger selections at the Electroweak scale:
- Inclusive single leptons with thresholds ~LHC
- Exclusive / multi-object triggers

Higher occupancies in the detectors bring:
1. Increased fake rate
   - Jets mimicking electrons
   - High radiation in the forward regions
2. Reduced rejection power of the algorithms
   - Worse resolution in calorimeters
   - Less effective isolation and pattern recognition
Phases of the L1 trigger evolution: become more intelligent!

- **Phase-0**: be prepared for $L = 10^{34}/\text{cm}^2/\text{s}$ (PU~25)
  - Complete detector & consolidate operations
  - Allow L1 topological criteria / more exclusive selections

- **Phase-1**: be prepared for $L = 3 \times 10^{34}/\text{cm}^2/\text{s}$ (PU~40)
  - Add more flexibility, without major architectural changes:
    - Additional coincidence layers in the forward muon spectrometer
    - Increased granularity in the calorimeter

- **Phase-2**: be prepared for $L = 5 \times 10^{34}/\text{cm}^2/\text{s}$ (PU~140)
  - Major upgrade for HL-LHC era: ensure appropriate rejection
  - Expected L1 rates over the limit allowed by detector FE
  - A new tracker will be available...

Any component installed in Phase-I must be fully operational also through Phase-II
First-level trigger in Phase-2: deriving ideas from HLT...

- **Tracking information at L1**: *add flexibility*
  - Combines calorimeter/muon with tracks, to remove mis-reconstructed or fake objects
  - Provides track isolation and multiplicity for \( \tau \), impact parameter for b-tagging
  - Vertex information for multi-object triggers (multi-jet)
  - ....

**EM rate reduction when applying the track match @14TeV, L=3e34, <mu>=70**

![Graph showing EM rate reduction](image)

**Tau rate reduction and efficiency with different selections based on track multiplicity and momentum thresholds**

![Graph showing tau rate and efficiency](image)
The ATLAS L1Track project

So far....

- Simulation studies to define upgrade requirements and evaluate detector and physics performance at high rates and PU
  - Even modest resolution tracking information ($p_T$, $\eta$, $\phi$) can provide sufficient rejection
  - Factors of $x_3$ for muons and $x_{10}$ for electrons, with only small efficiency losses
  - Double-lepton signatures are under control
  - Minimum track $p_T$ can be $\sim 17$ GeV for single leptons, few GeV for double signatures and taus

Next steps.....

- Development of conceptual design and technical solutions during next coming years, in connections with the Tracker upgrade (tracker construction will start in 2016)
  - Good view of the L1Track system design for the Initial Design Review in 2015
  - Document the overall scope in a Technical Design Proposal around 2016 (same time ITK TDR)

$L1Track$ is effective in reducing the rates in two momentum regimes: high-$p_T$ single leptons, low-$p_T$ double leptons and tau.

<table>
<thead>
<tr>
<th>Object(s)</th>
<th>Trigger</th>
<th>Estimated Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>no L1Track</td>
<td>with L1Track</td>
</tr>
<tr>
<td>$e$</td>
<td>EM20</td>
<td>200 kHz</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>EM40</td>
<td>20 kHz</td>
</tr>
<tr>
<td>$\mu$</td>
<td>MU20</td>
<td>$&gt; 40$ kHz</td>
</tr>
<tr>
<td>$\tau$</td>
<td>TAU50</td>
<td>50 kHz</td>
</tr>
<tr>
<td>$ee$</td>
<td>2EM10</td>
<td>40 kHz</td>
</tr>
<tr>
<td>$\gamma\gamma$</td>
<td>2EM10</td>
<td>as above</td>
</tr>
<tr>
<td>$ee$</td>
<td>EM10_MU6</td>
<td>30 kHz</td>
</tr>
<tr>
<td>$\mu\mu$</td>
<td>2MU10</td>
<td>4 kHz</td>
</tr>
<tr>
<td>$\tau\tau$</td>
<td>2TAU15I</td>
<td>40 kHz</td>
</tr>
<tr>
<td>Other</td>
<td>JET + MET</td>
<td>$\sim 100$ kHz</td>
</tr>
</tbody>
</table>
Triggering on tracks at L1 maybe be difficult!

- Cannot readout the full tracker at 40 MHz
- Already \(\sim 10-20 \text{ Gbs/cm}^2\) per layer at L1

Incercept = hard scattering! negligible compared to pileup

Reconstruction complexity/timing naively scale with the number of tracks....

Tracking is huge combinatorial problem, not linear with the number of interactions

Data reduction/reformatting

Longer latencies/larger Front End buffers

Faster data transmission and processing (Increase parallelism, network and trigger CPU needs)
CMS approach: low-\(p_T\) filtering

- CMS is designing tracker and FE modules with \(p_T\) discrimination capability
  - Reject low-\(p_T\) tracks, reducing data volume by one order of magnitude (40 MHz to \(~\)MHz)
- Correlate signals in two closely-spaced sensors, exploiting the strong magnetic field of CMS, with two steps:
  - **Cluster width approach**: preselection of hits according to their cluster width
  - **Stacked tracker**: correlation between preselected hits in nearby sensors

**Main challenges:**

- **L1 latency < 10 \(\mu\)s
- **L1 requirements affect the design of the tracker**
  - Different geometries are under study, to have coherent \(p_T\) threshold over the entire volume
  - Material may affect resolution at low-\(p_T\) due to MS
Increase latency: a new trigger scheme for ATLAS Phase-II

- Exploit the Region-of-interest mechanism!
- Add one trigger level, with extended latency (20 μs) to include the tracker information at L1
- Scale down current L1 to become L0, with extended latency (from 2.5 to 6 μs) and increased accept rate (0.5MHz, maximum 1MHz option under study)
Double-buffer readout strategy

**ABCn130 FE chip**: Analog Binary Chip, 130nm CMOS ASICs, with 256 readout channels and double-buffer architecture

Buffer1: synch, with pipelines
Buffer2: asynch

- Reduced readout rate at L0
  - For trigger purpose, only chips identified in the Region-of-Interest (Regional Readout R3) are readout: ~10% of the detector
  - So any specific part has a reduced data request rate: reads only at 50 kHz (10% of 500 kHz)

- Increased latency: Everything must be completed within <20 μs
**L1Track latency budget**

- To stay within 20 μs latency, crucial limits on the readout (6 μs) and on the L1TT algorithm (6 μs) timings are imposed.

- **Can we handle this?**

- Readout data size contributes hugely to latency, but tracking doesn’t need complete data.

- If necessary, data size can be reduced.

- Different formatting strategies under study.

<table>
<thead>
<tr>
<th>Event</th>
<th>Latency [μs]</th>
<th>cum. Latency [μs]</th>
</tr>
</thead>
<tbody>
<tr>
<td>formation of L0A</td>
<td>3.0</td>
<td></td>
</tr>
<tr>
<td>map RoI-ITK and send RoIs to ITK</td>
<td>1.25</td>
<td>4.25</td>
</tr>
<tr>
<td>ITK readout in RoI regions</td>
<td>6.00</td>
<td>10.25</td>
</tr>
<tr>
<td>transmit to L1TT</td>
<td>2.00</td>
<td>12.25</td>
</tr>
<tr>
<td>L1TT algorithm</td>
<td>6.00</td>
<td>18.25</td>
</tr>
<tr>
<td>L1A formation from track+L1MU+L1Calo</td>
<td>1.00</td>
<td>19.25</td>
</tr>
</tbody>
</table>

**Different deadlines for decisions**

- All aspects connected to the actual detector (front-end, communication etc.) fixed before construction of ITK starts.

- **Data formatting** studies to be completed in parallel.

- **Trigger processor technology** can be decided later.
L1Track trigger readout challenge
ABC130 chip prototype already prepared

- Hybrid Chip Controller (HCC) still under design
- Different strategies to control traffic are under study, without undue additional material (and power)
  - **HCC prioritisation**: prioritize the R3 data flow w.r.t L1 in the chips daisy-chains
  - **Increase HCC FIFO depth**: to absorb fluctuations and to fast clear the daisy-chain
  - **Increase the number of daisy-chains links** (from 2 to 4)
  - **Increase the HCC output bandwidth** from 160 Mbps to 320 Mbps

To keep latency low, regional data from the front-end cannot be queued.
R3 latency on all modules with different layouts

- If it is possible to use fast 320Mbps, all the 95% latencies would be under 5 μs
- A mixed layout can be adopted, with increased bandwidth only on some hybrids
  - Higher bandwidth is expected in the **Endcaps** due to higher occupancy and lower density of chips
What about robustness?

- With different unexpected conditions (increase of L0/L1 rates, occupancy)?
  - on some critical rings, latency can change rapidly if the bandwidth is not appropriately increased
  - see the latency maps for Endcap ring 6 at largest z

- Case LoA rate changes 0.5MHz → 1MHz
  - **ABC130 buffers**: can be easily increased
  - **Latency**: increase is negligible with extra links
  - **L1Track trigger logic**: more work can be required to double the rejection power and have the same L1A rate
L1Track trigger logic challenge
Level-1 trigger processors
(a small excursion on technology trends)

Follow the evolution of digital integrated circuits on a single chip (SoC)
- Request of higher complexity ⇒ higher chip density ⇒ smaller
  structure size (for transistors and memory size): 32 nm ⇒ 10 nm
- Custom ASICs or Off-the-Shelf component (COTS):
  - Specific microprocessors (CPUs, DSPs=Digital Signal Processors,..)
  - Programmable logic devices (FPGAs)
    - high throughput, flexible, parallel
    - development requires more effort

Over recent years, the latency and performance gap between multicore processors has been closing to the point that many of the functions that required the specialized hardware properties of DSPs and FPGAs can now be done in software in General Purpose Processors (GPP).
The right choice can be to combine the best of both worlds by analyzing which strengths of FPGA, GPU, and CPU best fit the different demands of the application.

Nvidia GPUs: 3.5 B transistors
Virtex-7 FPGA: 6.8 B transistors

can implement multiple DSP algorithms
ATLAS Upgrade: take full advantages of modern real-time technology

The current technology using fiber data transfer, FPGAs, custom chips and modern PCs could not be scaled in a simple manner to accommodate all the tracking trigger demands. Significant improvements, or breakthroughs, will be probably needed. In other words: 

aggressive R&D

The golden time for “easy” digital electronics is over

- Very high clock frequency (20 MHz to 20 GHz and beyond)
- Analog interference on digital electronics becomes important (noise, cross-talk, signal reflection)
  - Major challenges for system design, from power distribution, PCB layout, ....
- Cannot just buy some FPGAs, write some VHL code and claim to have an electronics board!
  - see High-Speed Digital Design: A Handbook of Black Magic
Past, present and future of tracking trigger systems

**CDF- SVX II**

- peak $L = 3 \times 10^{32}$, 10 PU
- BC = 396 ns
- $L_1 = 30$ kHz
- $L_2 = 750$ Hz

**ATLAS FTK Run2-Run3**

*FastTracKer built over current detector*

- peak $L = 3 \times 10^{34}$, 69 PU
- BC = 25 ns
- $L_1 = 100$ kHz
- $L_2 = 10$ kHz

**ATLAS L1Track Run4**

- peak $L = 5 \times 10^{34}$, 200 PU
- BC = 25 ns
- $L_0 = 0.5/1$ MHz
- $L_1 = 200$ kHz

**Fast tracking for L2**
An evolution of methods and technologies for fast tracking

**CDF-SVX II**

- ~ 0.2 millions channels
- L2 decision: ~ 20 μs, 30 kHz
- Tracks with offline-like resolution: i.e. 35 μm on the impact parameter

**ATLAS FTK Run2-Run3**

- FastTracKer built over current detector
- 80 M (Pixel) + 6 M (SCT) channels
- L2 decision: ~ 10 μs, 100 kHz
- ~offline quality tracks with $p_T > 1$ GeV

**ATLAS L1Track Run4**

- 638 M (pixel)+ 74 M (strip) channels
- L1 decision: ~ 20 μs, 1 MHz
- ~offline quality tracks with $p_T >$ few GeV

**Other relevant aspects:**
- Symmetrical design or not
- Materials
- Cabling map
Tracking trigger approach from the past

**Same requirements**

- Reconstruct tracks as close to the offline performance as possible
- Highly parallelism
- Reduce combinatorics by use of multiple step processing

1. Find low resolution track candidates called “roads”
   - Solve most of the combinatorial problems

2. Then fit tracks inside roads
   - Thanks to 1st step, this is much easier
   - A linear approximation gives near ideal precision

A very successful approach at CDF for RunII: SVT (Silicon Vertex Trigger) based on Associative Memory, in turn made of CAM

- APS Panofsky Prize to Aldo Menzione and Luciano Ristori

![Pattern recognition w/ Associative Memory](http://www.pi.infn.it/~orso/ftk/IEEECNF2007_2115.pdf)

Originally: M. Dell’Orso, L. Ristori, NIM A 278, 436 (1989)
Stories on some technological innovations at CDF in the 1980s-1990s

in the ‘80s the technology of VLSI design becomes available to the universities and to small research projects

A slide from Luciano Ristori at TIPP 2011 conference
Stories on some technological innovations at CDF in the 1980s-1990s
Main ingredient: the Associative memory (AM)

- **RAM (Random-Access-Memory):** memory address → the data word stored at that address.

- **CAM (Content-Addressable-Memory):** data word → searches its entire memory in one single operation and return the address:
  - much faster than RAM
  - commonly used for networking and computing: transform IP address, data compression, cache tag (parallel RAM access)

- **AM or PRAM (Pattern Recognition Associative Memory) CAM based**
  - Pattern recognition finishes as soon as all hits arrive
  - Use majority logic
  - Can be Ternary CAM: 3 states (1/0/x) with the addition of a “don’t care” bit

**Example: CAM Output**

- Address: 0x3f
- Input: 00111011010
- Output (Majority): 00111011011

**RAM Example**

- Address: 0x3f
- Data: 001110110110

**AM/PRAM Example**

- Flag whether a pattern is matched

- 001110110110

- Majority logic
AM based tracking system

- Dedicated device: maximum parallelism
- Each pattern with private comparator
- Track search during detector readout

When a pattern is matched, the corresponding hits are selected for the following step

Pattern-matching done in few 10 ns!
data readout, data distribution and data formatting takes longer.....

AM inputs are hits from different layers
Finite number of patterns (pattern-bank): given finite resolution, different tracks generate the same pattern.

- Higher resolution and rejection, if more patterns can be stored (and if more CAM cells/chips are available)
  - Preferred approach: 90% efficiency in a low fake scenario (to control the workload to the fitting step)
  - To add flexibility, the resolution can be variable - with the use of “don’t care” bits (Ternary CAMs)
AM evolution, to increase pattern density

- (90's) Full custom VLSI chip - 0.7μm (INFN-Pisa)
- 128 patterns, 6x12bit words each, 30MHz

Alternative FPGA implementation of SVT AM chip
G Magazzù, 1st std cell project presented @ LHCC (1999)

- Standard Cell 0.18 μm → 5000 pattern/AM chip
  SVT upgrade total: 6M pattern, 40MHz

- AMchip04 - 65nm technology, std cell & full custom, 100MHz
  Power/pattern/MHz ~30 times less. Pattern density ×12.
  First variable resolution implementation!
  F. Alberti et al 2013 JINST 8 C01040, doi:10.1088/1748-0221/8/01/C01040

FTK R&D in progress:
- AMchip05: switched to serialized IO (11*2Gb/s)
- AMchip06 prototype: the FTK AM chip with 128k patterns/chip

New technologies for L1Track?

Successfull! High pattern density, high speed and low power consumption

design with state-of-the-art technology

SVT upgrade ready for LHC performance
Limits of the AM approach

- Performance fundamentally limited by Moore’s Law
- AMChip near limit of conventional associative memory densities
  - Earlier studies demonstrated that ternary CAMs can be used with 10 billion patterns or more, doing a pattern lookup in < 200 ns

A challenge for HL-LHC

Increase the patterns density by 2 orders of magnitude

Increase the speed by a factor of \( \geq 3 \)

while keeping similar power consumption

or

go to higher dimensions
AM evolution: 3D approach

VIPRAM: Fermilab project using 3D vertical integration technology
*(TIPP 2011 pre-print)*

- One cell can process N layers in about one CAM cell size ⇒ density increased by N
  - 2D with 65 nm: ~50K patterns/cm² (AMchip04)
  - 3D with 130 nm: ~200K patterns/cm²
- Reduced connections ⇒ higher speed and less power density
- More flexible design

*VIPRAM (Vertically Integrated Pattern Recognition Associative Memory)*

*From 2D to 3D*

*Physical detector layers ←→ silicon layers*
Track fitting techniques

- Simple algorithm performed on any good combination of hits ⇒ can be massively parallelized
- Linear approximation on a limited region: get a set of linear equations (instead of solving helix) → fast multiplications with pre-computed constants
  - Use of Look-up-tables (LUTs) with precalculated values (5 track parameters and the $\chi^2$) stored in a table and interpolated

Due to short latencies and huge number of inputs, use of more complex algorithms, like Kalman Filters and Hugh Trasforms, (used for image processing) is limited
**Track fitting technology evolution**

- **CDF-SVX II**
- **ATLAS FTK Run2-Run3**
  - **GPUs** is promising candidate: constant performance with increasing # of fits
  - Little is known about GPU performance, both in terms of speed and latency overheads, in low-latency environments
  - FERMILAB-CONF-11-710-PPD (2012)
- **ATLAS L1Track Run4**
  - Dedicated AM hardware combined with a dual-processor PC running an optimized Linux quasi-realtime kernel
  - **FPGAs** w/ many Digital Signal Processors (DSPs): \( \Rightarrow \sim 1 \text{ fit/ns} \)
  - Constraints due to limited bandwidth and processing power
    - #AM patterns < \( 16.8 \times 10^6 \)
    - #fits/event < \( 80 \times 10^3 \)
Future: miniaturize, larger AM chips, integrate!

- If the AM stage and the Trak Fitting can be integrated
  - latency is reduced
  - bandwidth is under control
- 3D Technology could help here (in the future)
- New generation of FPGAs with stacked silicon interconnect (SSI) technology: break through the limitations of Moore’s law
  - Xilinx SSI technology
Data sharing technique

- Jumper cables
  - Flexible, but ugly and difficult to maintain
  - Still requires custom backplane

- Dedicated traces on the backplane
  - Custom backplane
  - Each crate may be different
  - Inflexible design

- Modern ATCA with full-mesh

- ATLAS FTK Run2-Run3
- Patterns ~ Billion / crate/shelf

- ATLAS L1Track Run4

CDF- SVX II

data formatting FPGA Costellation for FTK
How could the L1Track may appear?

CDF - SVX II

ATLAS FTK Run2-Run3

ATLAS L1Track Run4

- CDF original SVT system had ~400K patterns total: 128 patterns per AMchip
- Test state-of-the-art CAD tools
- Commissioned around ~2001

16400 AM chips + 2000 FPGAs @ 100 MHz for 16-bit words (2 Gbs)

#AM patterns < 16.8 millions, with variable resolution

Schedule:
- Integration with limited coverage in Run2 (2015)
- 2016: full coverage

Aim to reach ~500K patterns/cm² for VIPRAM chip

Or other technology? GPU?

Schedule: ready for 2022
Conclusions

- I’m tempted to say: There are no conclusions, future is open
- Old stories from the past can help us in seeing how it could be, if much effort is concentrated in understanding the requirements
- L1Track project will deal with the possibility of triggering in HL-LHC
- We must maintain wide open sight to what we can steal from the technology market, that has somehow similar demands on large data-processing, in short time, on large systems…. Steal from you cell-phone!
References

- ECFA workshop in Aix-les-Bains
- Review of LHC & Injector Upgrade Plans Workshop (RLIUP)
- References for tracker evolution
  - Strip CMOS task force
  - High-performance Signal and Data Processing Workshop 2014
- WIT workshop on intelligent trackers
  - 2012: https://indico.cern.ch/conferenceTimeTable.py?confId=154525#all.detailed