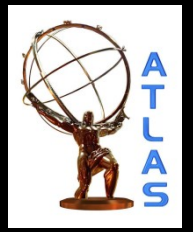


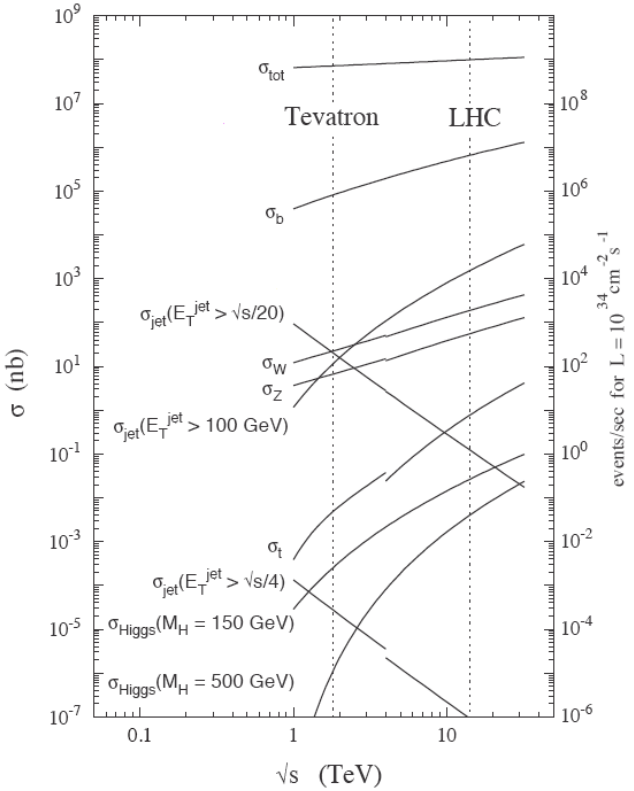
G. Volpi - INFN Frascati

ANIMMA 2011

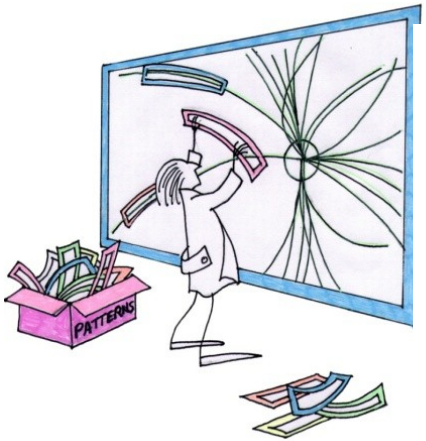
A new Variable Resolution Associative Memory for High Energy Physics

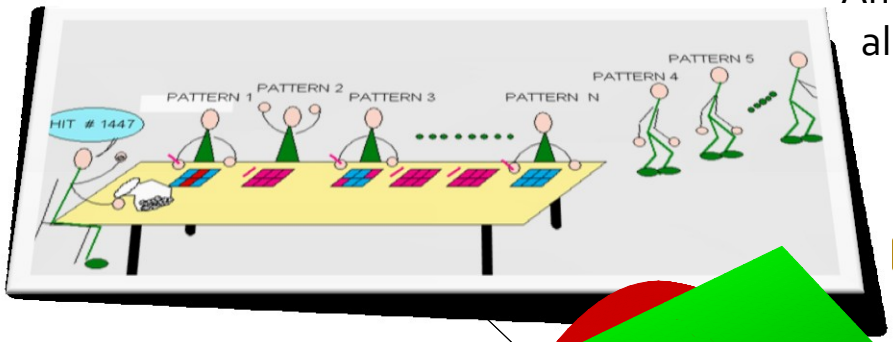
Data selection at hadronic colliders

- Search for rare SM or predicted BSM processes push the colliders intensity to new frontiers
- Rare processes are overwhelmed by well-known processes
  - 8/9 order of magnitudes between Higgs and total cross-section
  - Need to prioritize the physics output and leave flexibility for the unexpected
- Need sophisticated techniques to maximize the bandwidth for interesting events
- Several strategies are used, track trigger showed good results at CDF (SVT), ATLAS implementation (FTK) can have similar success

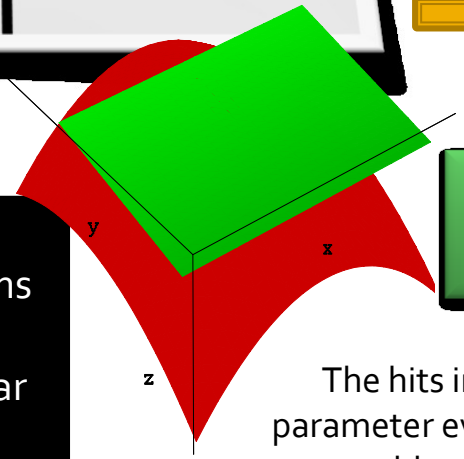


Online tracking with FTK





An associative memory (AM) chip allows a full parallel search in the pattern bank.
Similar to commercial CAMs.



$$p_i = \sum_j C_{ij} \cdot x_j + q_i$$

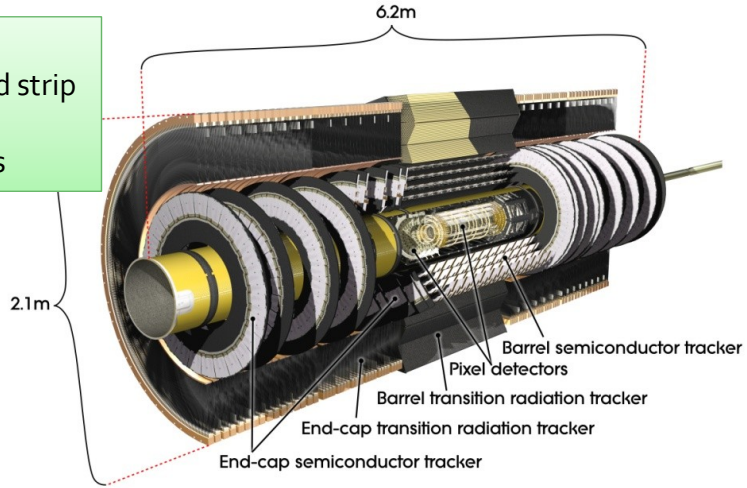
- In FTK tracking divided into two separated steps
- The pattern matching uses a large bank of patterns and special hardware
- The track's parameters are evaluated using a linear Principal Component Analysis algorithm

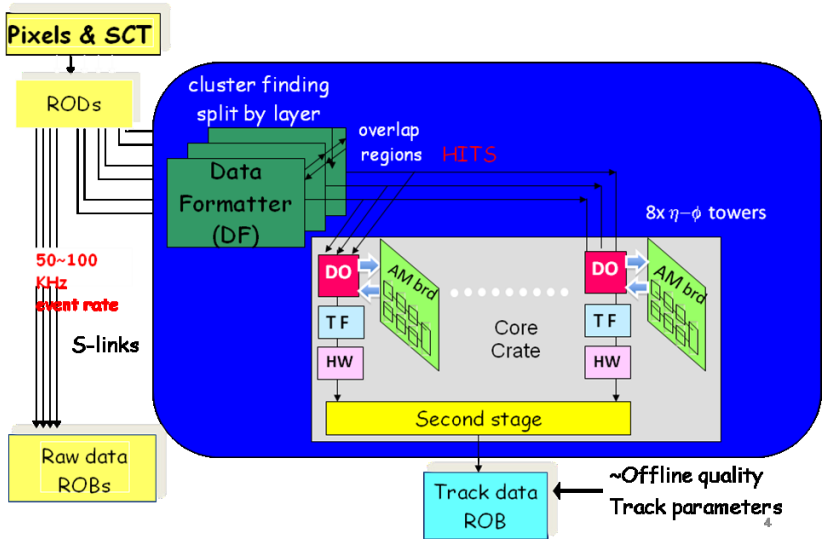
(j.nima.2003.11.078)

The hits in a road are combined and the parameter evaluation is reduced to a linear problem, efficiently solved in an FPGA

ATLAS Inner detector layout

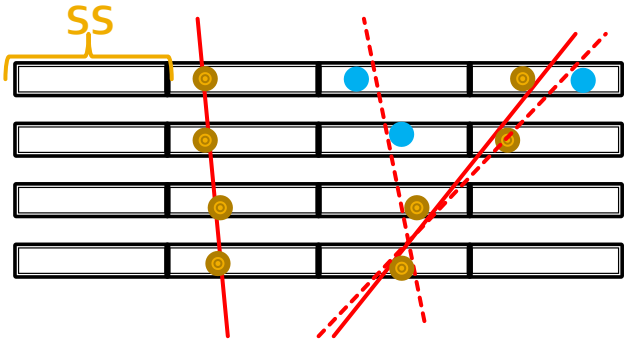
- 3 pixel layers
- 4 double sided strip layers
- Up to 11 layers



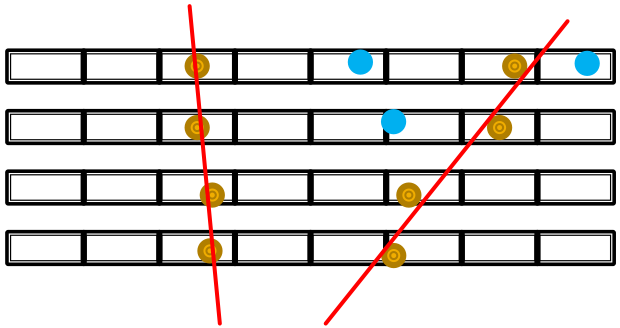


- FTK processor will read the ID for each Level 1 accept
- To improve the input bandwidth the ID is divided into 8 regions
  - Each region will have 16 processing units, all working in parallel
  - A total of about 1 Giga patterns
- The device is able to produce a complete list of tracks in tens of microseconds

Associative memory and granularity



3 patterns , 3 roads (1 fake), 4 tracks (2 fakes)



4 patterns , 2 roads, 2 tracks, no fakes

- A large segmentation (SS) improves the pattern recognition efficiency
  - More likely that noisy hits enter the road: fake tracks
  - Or random combinations seem to be aligned in pattern: fake roads
- Reducing the segmentation has benefits
  - Fake roads or tracks are less probable
  - But more hardware, more money

SS size, efficiency and bandwidth

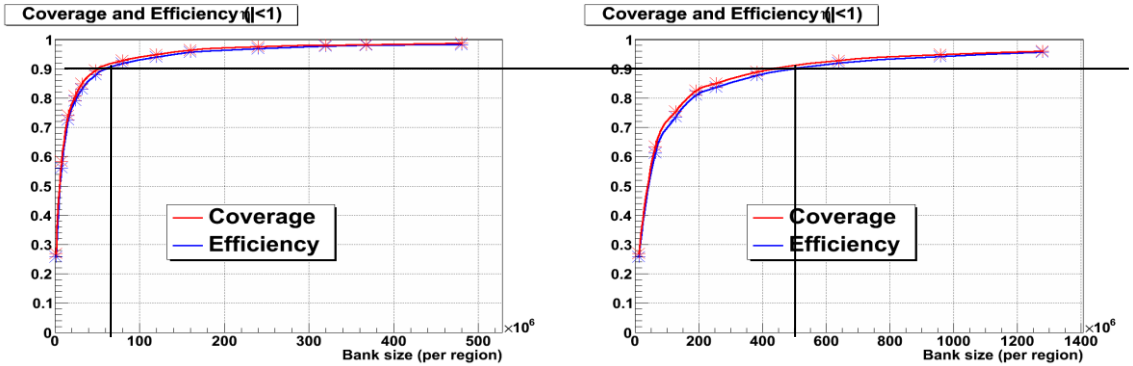
Pattern size r-ϕ: 24 pixel; 20 SCT; 36 pix z

(1.2 mm) (1.6 mm) (14 mm)

Pattern size r-ϕ: 12 pixel; 10 SCT; 36 pix z

(0.6 mm) (0.8 mm) (14 mm)

Pixel and SCT unit is the number of adjacent channels



65M

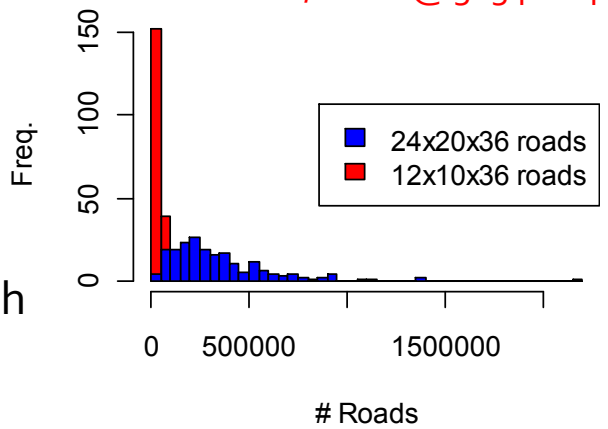
500M

90%

# of patterns in Amchips (barrel only, 45 ϕ degress)

<# roads/event @ 3E34> = 342k

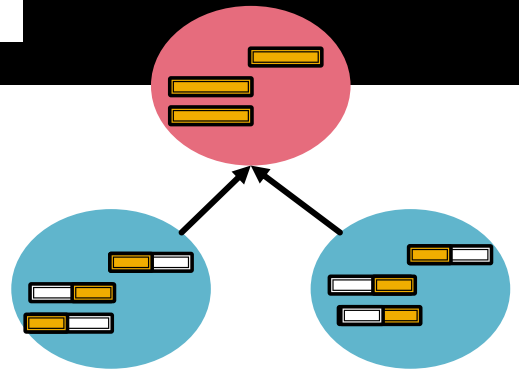
<# roads/event @ 3E34> = 40k

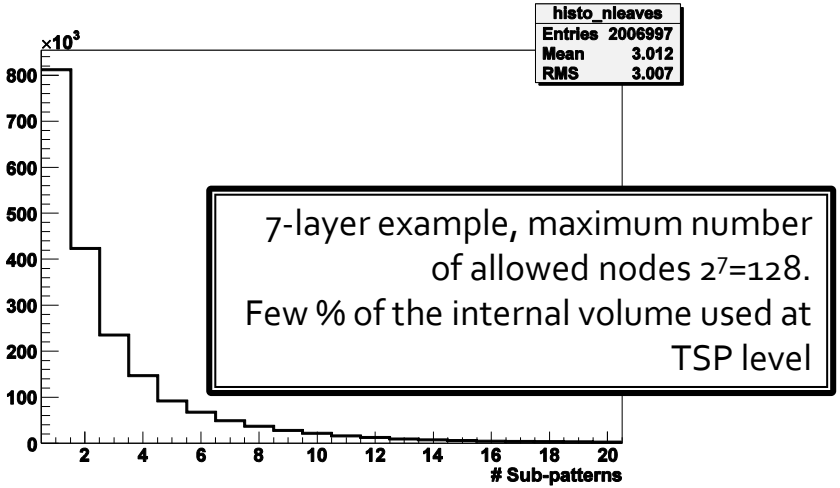


The bank with the larger SS size can be 90% efficient with a limited number of patterns but needs a greater output bandwidth, not possible in high luminosity events.

Tree search processor idea

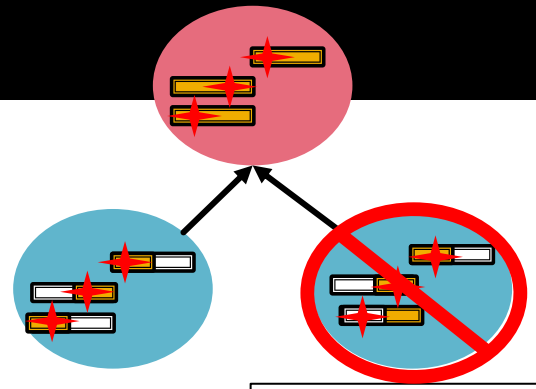
- Patterns can be generated at the finest resolution and grouped according to the low resolution granularity
  - A tree organization is natural
- AM chip size is a function of the low resolution granularity
  - Limited number of patterns
- Efficiency and fake rate driven by the granularity in the leaves
  - Final pattern recognition uses the full precision
- AM patterns usually have a limited number of leaves
  - Most of the internal volume unused

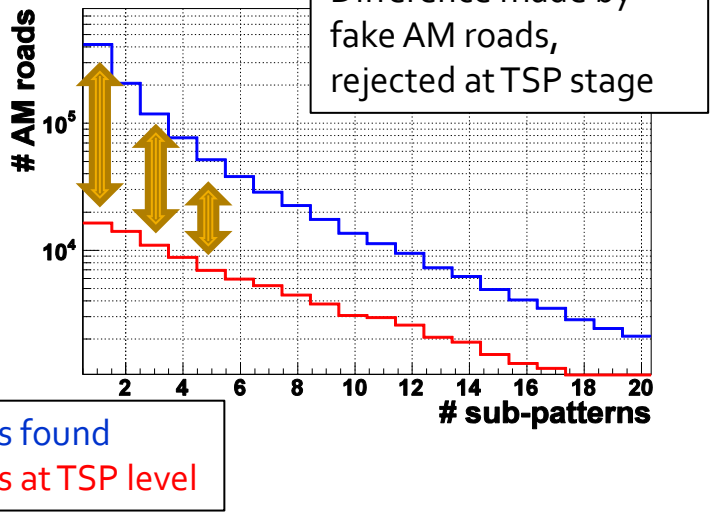




TSP rejection power

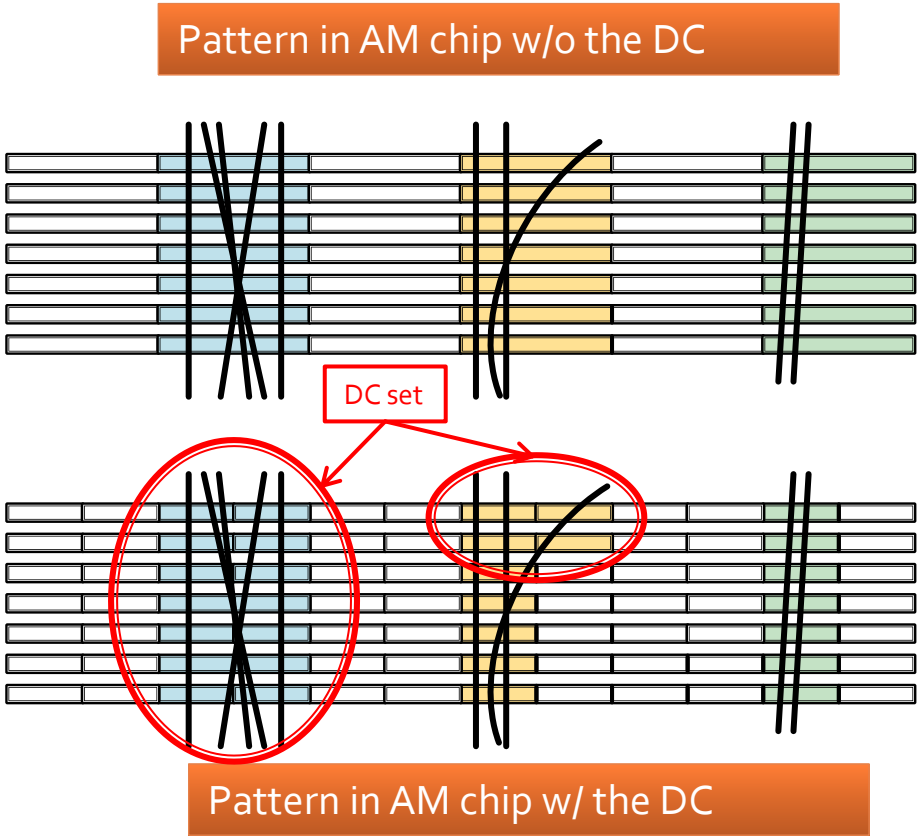
- TSP simulation implemented
  - The AM chip simulation finds the low resolution roads
  - For each AM road all the nodes are compared looking for matches
- WH events with 75 pileup-event simulation used
- Simulation results confirm TSP can reject large number of roads
  - Most of the fired roads have few sub-patterns
  - Special attention to this category is justified





Variable resolution and don't-care bit

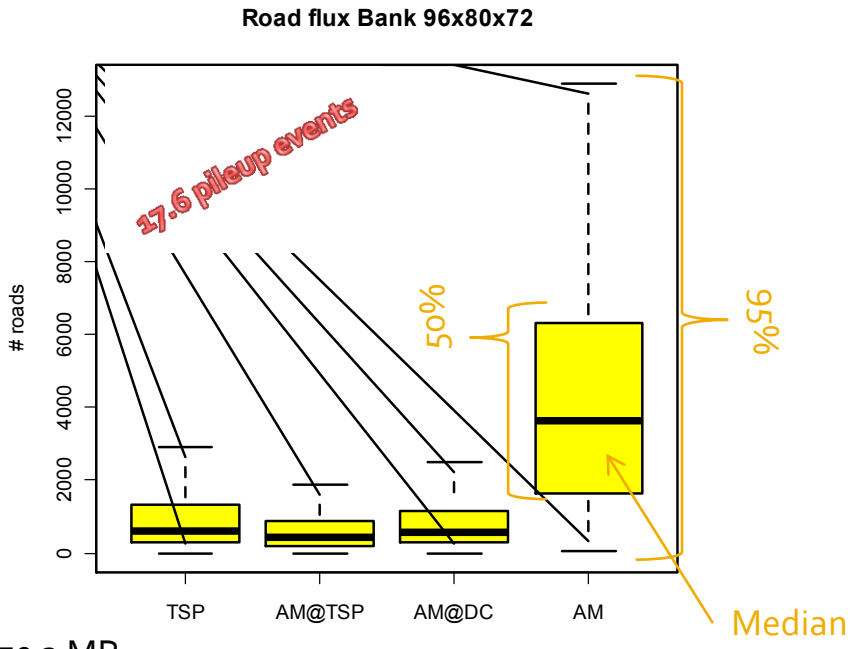
- To have a more flexible AM bank the SS are stored at maximum precision
  - Each additional bit halves the size of the SS
  - An additional bit is set if the match stops before the least significant bits, not using the full precision: "don't care" bit like modern CAM memories (*K. Pagiamtzis and A. Sheikholeslami, Solid-State Circuits, IEEE Journal of, vol. 41, no. 3, 2006*)
  - Where the full precision is used (no DC) a portion of the SS is vetoed
  - Where the DC is set the whole SS is active
- The use of DC allows a reduction in the active region of the patterns
  - This reduces the probability of fake roads and tracks
  - Reducing the number of roads in the output



Complete simulation studies

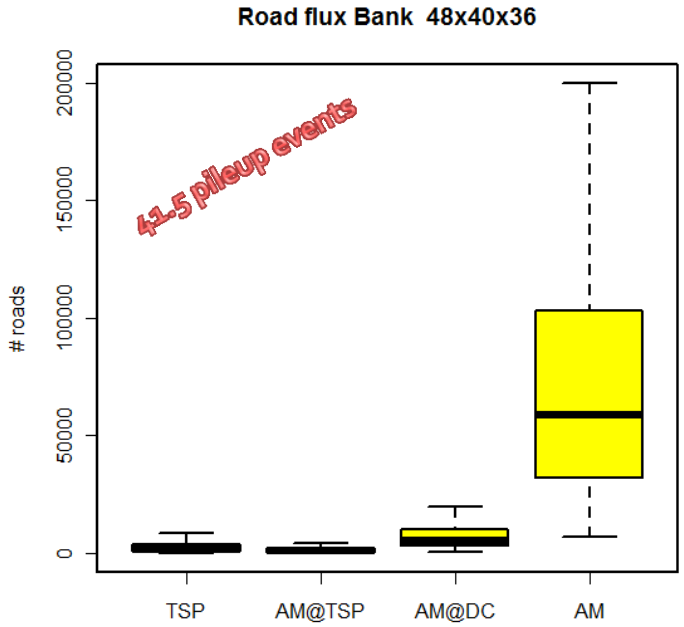
- FTK simulation improved to include DC and TSP
- TSP bank produced using 7 inner detector layers
- Thin patterns grouped within the large patterns doubling the SS size: 1 DC bit
  - Working point fixed at 90% TSP efficiency
  - Typical TSP→AM reduction 1/3
  - Increment of area used in the AM chip 1/15=7%
- Physics sample used: WH with full ATLAS simulation
  - Multi-jet events with high number of tracks
- 3 pileup scenarios at 25 ns bunch spacing are evaluated:
  - 17.6, 40.5 and 75 pileup events
- For each pileup scenario the number of roads per processing unit is important
  - Number extrapolated from barrel only simulation

Performance (1)



- AM size 2 MP
- TSP size 5 MP

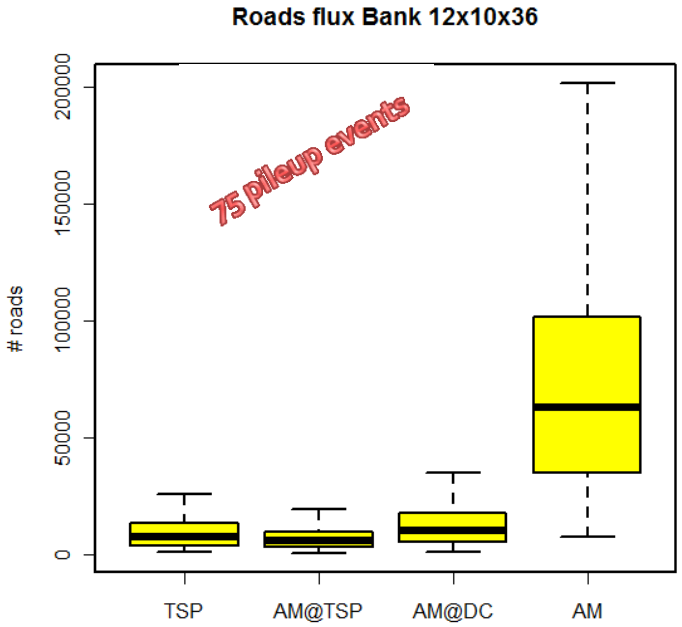
TSP bank 48x40x72



TSP bank 48x40x72

- AM size 32 MP
- TSP size 100 MP

Performance (2)



- AM size 138 MP
- TSP size 384 MP

	Number of output roads				
Pileup events	**AM**	**AM w/ DC**	**AM w/ TSP**	**TSP**	**AM bank size**
17.6	5040	**959**	709	1060	2 M
40	37000	**6500**	3880	5720	32 M
75	53500	**8250**	4430	5950	138 M

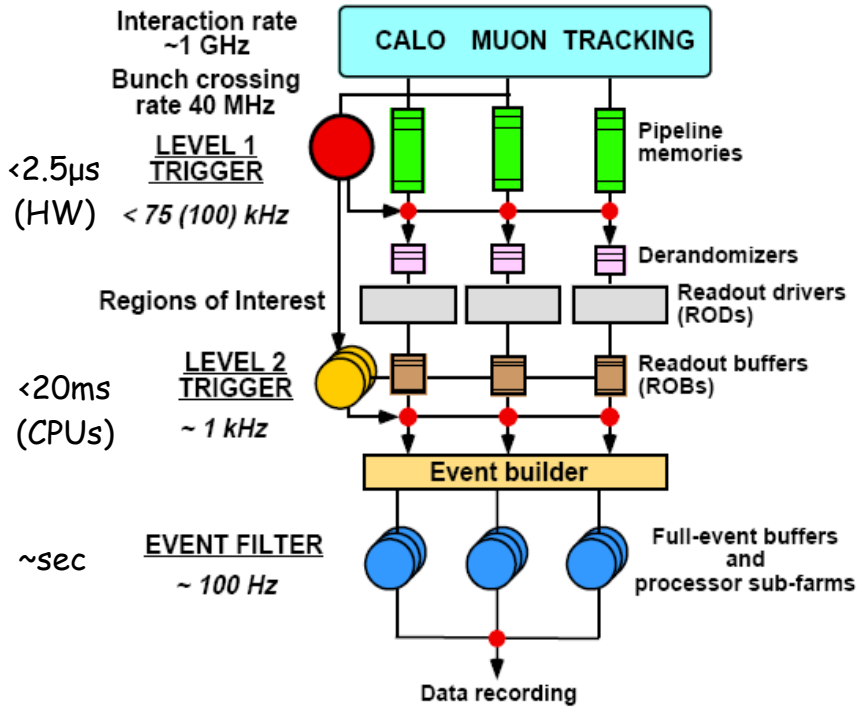
- Pattern bank reduction factor: 2.5 - 3
- A bank with DC capability reduces the fakes by a large factor: 5-7
- Good performance without adding further hardware

Conclusions

- Rapid tracking is an opportunity for the ATLAS experiment
- FTK based on a successful technology (SVT)
  - But moving from 2 TeV@10^{32} to 14 TeV@10^{34} needs new ideas
- Variable resolution is a general idea
  - The patterns are able to change in shape and matching volume
  - Improves the precision only where needed
- Just using a single DC bit has a significant effect
- The number of roads out of the AM chip is reduced greatly by the DC
  - It has almost the same effect as a complete TSP
- 7% more information in the bank has the same power as a bank 3 times bigger
  - Use of several DC bits possible in the chip and a study is ongoing

Backup

FTK and ATLAS DAQ



ROD output sent to FTK at L1 rate

FTK:

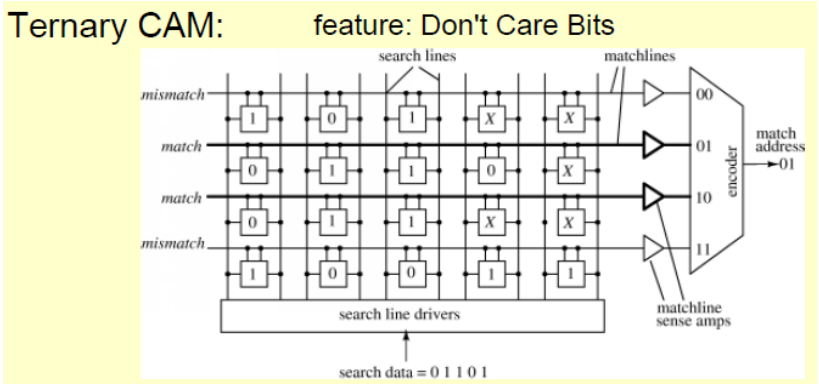
- Operates in parallel with silicon readout after each L1 trigger
- Reconstructs tracks over entire inner detector ($|\eta|<2.5$) in ~25μs (for $3x10^{34}$ cm^2 s^{-1} Luminosity)
- Provides high-quality tracks by the beginning of L2 trigger processing

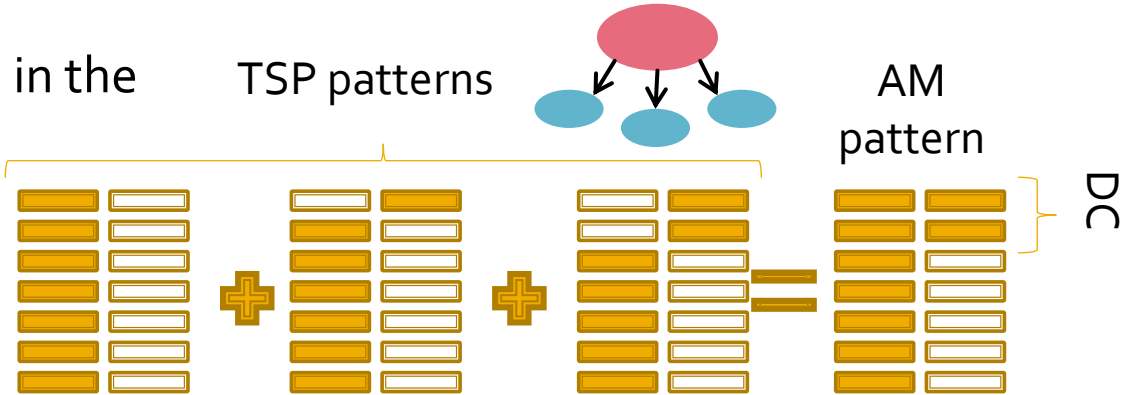
Both trigger levels benefit:

- L1: events needing L2 tracking could have a higher fraction of the L1 rate
- L2: with tracking already performed, the extra CPU time can be used for more advanced algorithms

DC bit content construction

- After the construction of the tree the content of the AM patterns is used
- If in one layer all the nodes shares the same portion, the DC is not set
  - Shrinking the active part of the AM pattern
- If all the portions are used in the nodes, the DC is set





The smaller the number of TSP nodes, smaller the use of DC