Variable resolution Associative Memory for the Fast Tracker ATLAS upgrade

Alberto Annovi

Istituto Nazionale di Fisica Nucleare
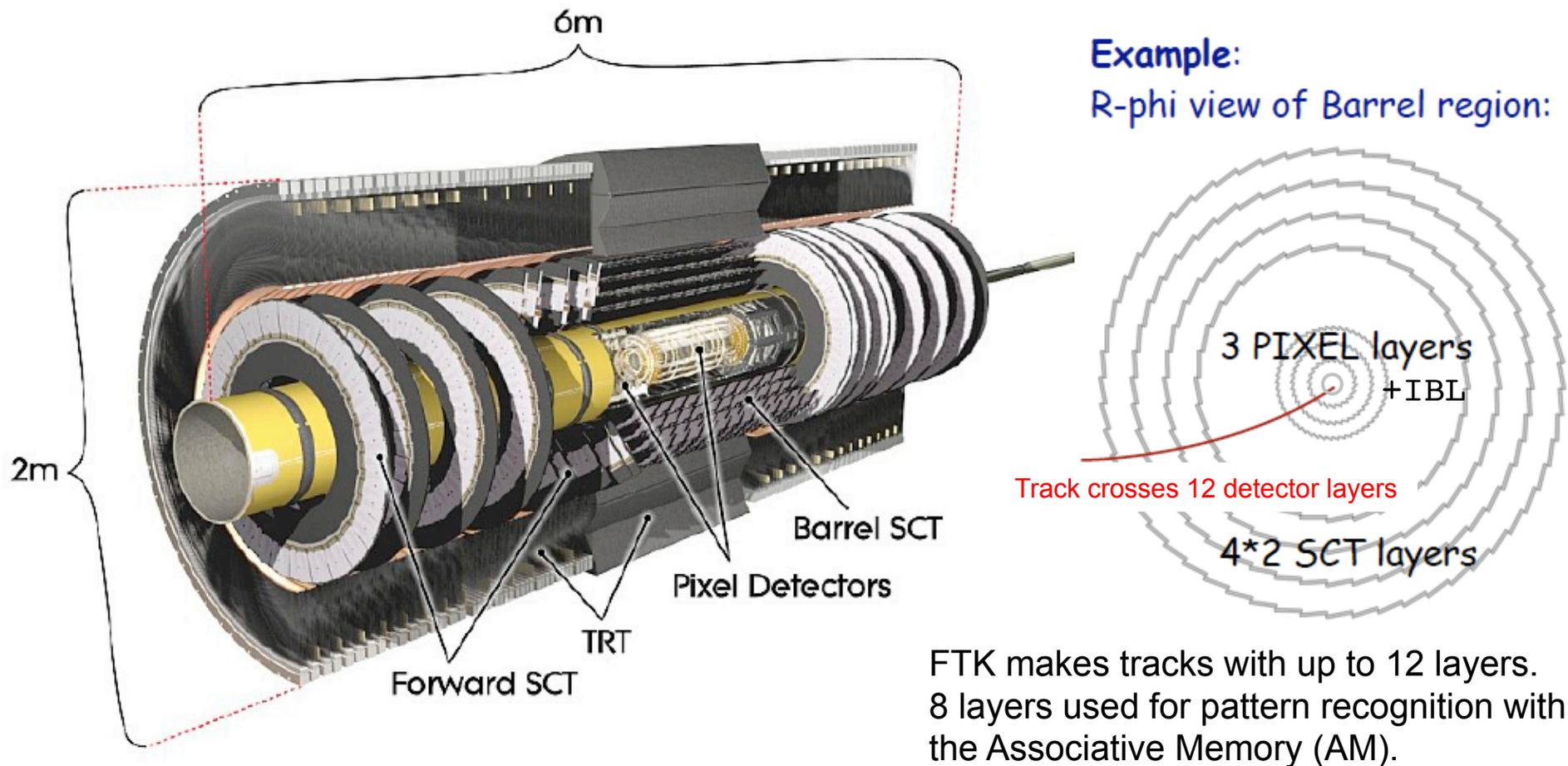Laboratori Nazionali di Frascati

Fast tracking in Pixel and SCT detectors

Detector and trigger coverage up to $|\eta| < 2.5$

FTK processes all level-1 accepted events (100kHz), it provides tracks for level-2 algorithms

Output: all tracks down to $p_T > 1$ GeV. Typical latency ~100µs

Advantages: high-bandwidth connection with detector & HW optimized for the specific tasks

6m

2m

Barrel SCT

Pixel Detectors

TRT

Forward SCT

Example:
R-phi view of Barrel region:

3 PIXEL layers
+IBL

Track crosses 12 detector layers

4*2 SCT layers

FTK makes tracks with up to 12 layers.
8 layers used for pattern recognition with
the Associative Memory (AM).

FTK part 1: Associative Memory

- **AM Pattern recognition** – find track candidates with enough Si hits
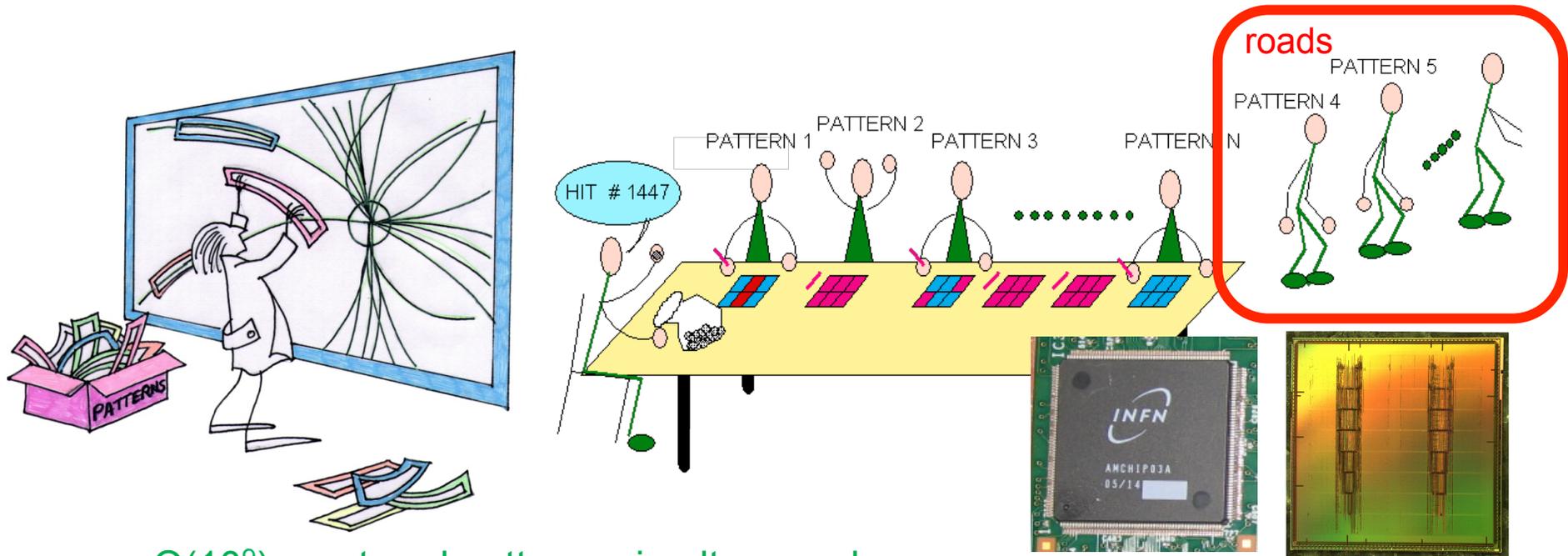
- O(10^9) prestored patterns simultaneously see the silicon hits leaving the detector at full speed.
- This pattern recognition step is essential to reduce the combinations for the following fit procedure.
- The AM outputs patterns that match 7 out-of 8 layers: called roads.

FTK part 2: Linearized Track Fit

– 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.
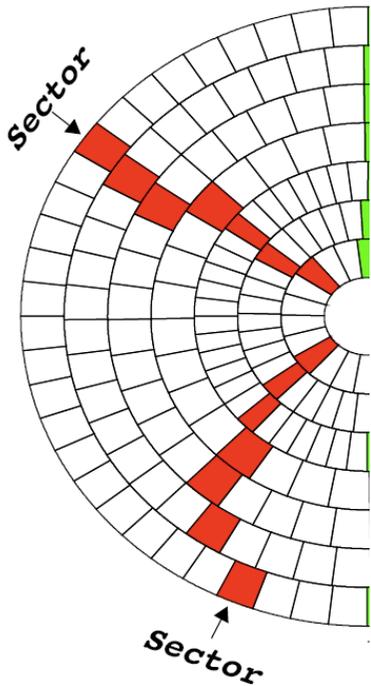
$$p_i = \sum_{j=1}^{16} a_{ij} x_j + b_i$$
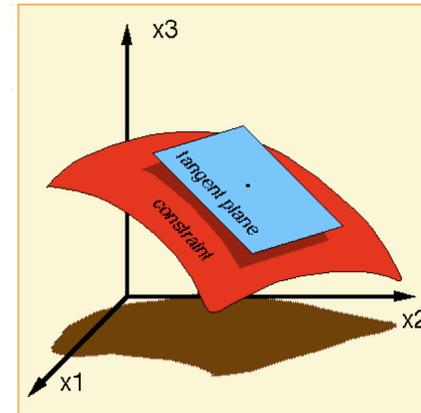
16D coord. space of hit combinations

5D track surface

– p_i's are the helix parameters and χ^2 components.
– x_j's are the hit coordinates in the silicon layers.
– a_{ij} & b_i are prestored 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 approx 1 Gfit/s/FPGA

• Based on Principal component analysis

j.nima.2003.11.078, and H. Wind, CERN-EP-INT-81-12-REV, 1982

Associative Memories

Extrapolate to inner silicon layers

- First AM for HEP idea
  - Search its entire memory at each clock cycle: fast pattern recognition
  - Inspired from Content Addressable Memories (CAM)
  - M. Dell'Orso, L. Ristori NIM A **278, 436 (1989)**
- First application: SVT @ CDF
  - Seeded by drift chamber tracks
  - Look for associated Silicon hits at radii 2.5-10.5cm
  - Started with 384k patterns
 - Upgraded to 6M patterns
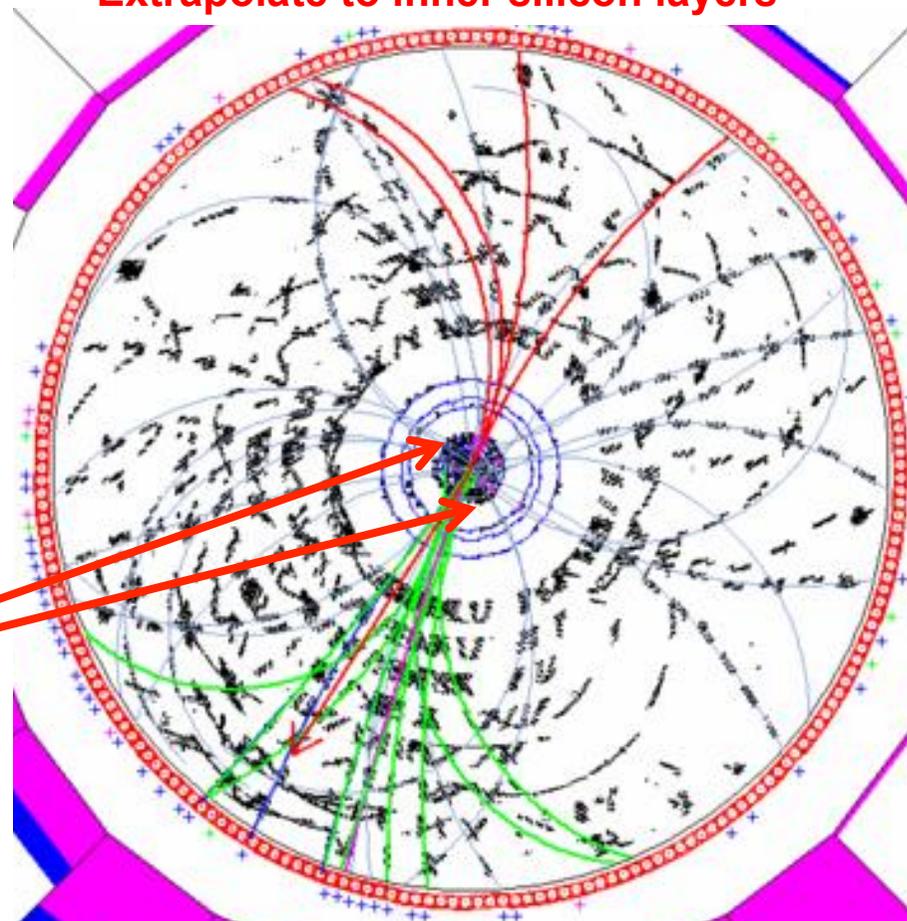
We discuss the architecture of a device based on the concept of *associative memory* designed to solve the track finding problem, typical of high energy physics experiments, in a time span of a few microseconds even for very high multiplicity events. This "machine" is implemented as a large array of custom VLSI chips. All the chips are equal and each of them stores a number of "patterns". All the patterns in all the chips are compared in parallel to the data coming from the detector while the detector is being read out.

AM for ATLAS

- Silicon only trackers
- High luminosity → high detector occupancy
- Thousand tracks / bunch crossing $<\mu>=20$

- For AM to reduce information
  - Needs very high resolution
  - Needs billions of patterns
  - Needs faster clock of 100MHz
  - Can profit from today electronics
  - Requires O(8k) of AM chips
  - Need also a new kind of Associative Memory!!

25 reconstructed vertex + Z→2μ
Up to $<\mu>=80$ by 2019; $<\mu>=200$ by 2023 HL-LHC [CERN-LHCC-2012-022]
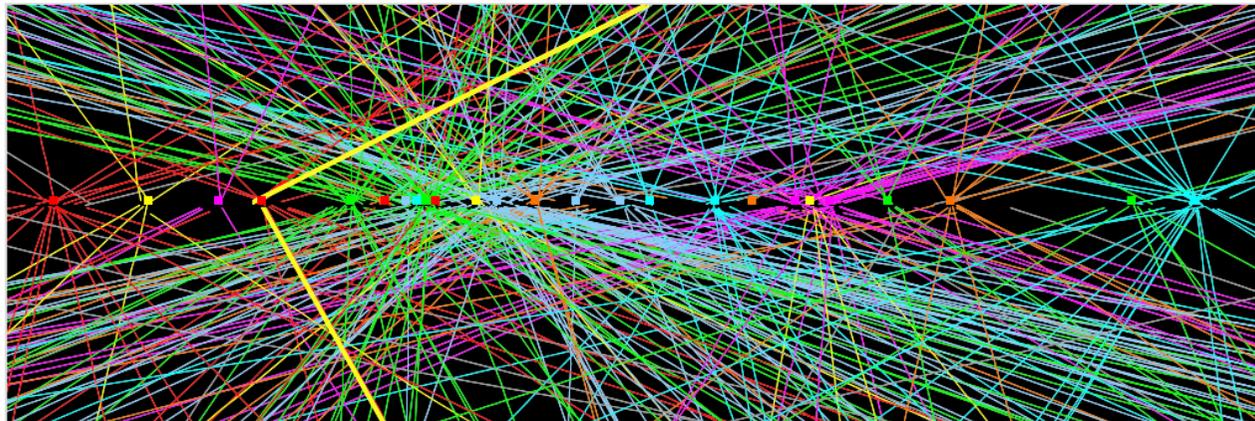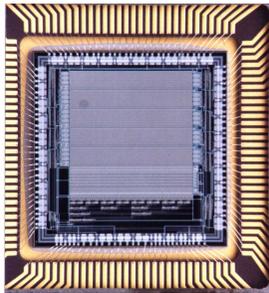
AM working principle
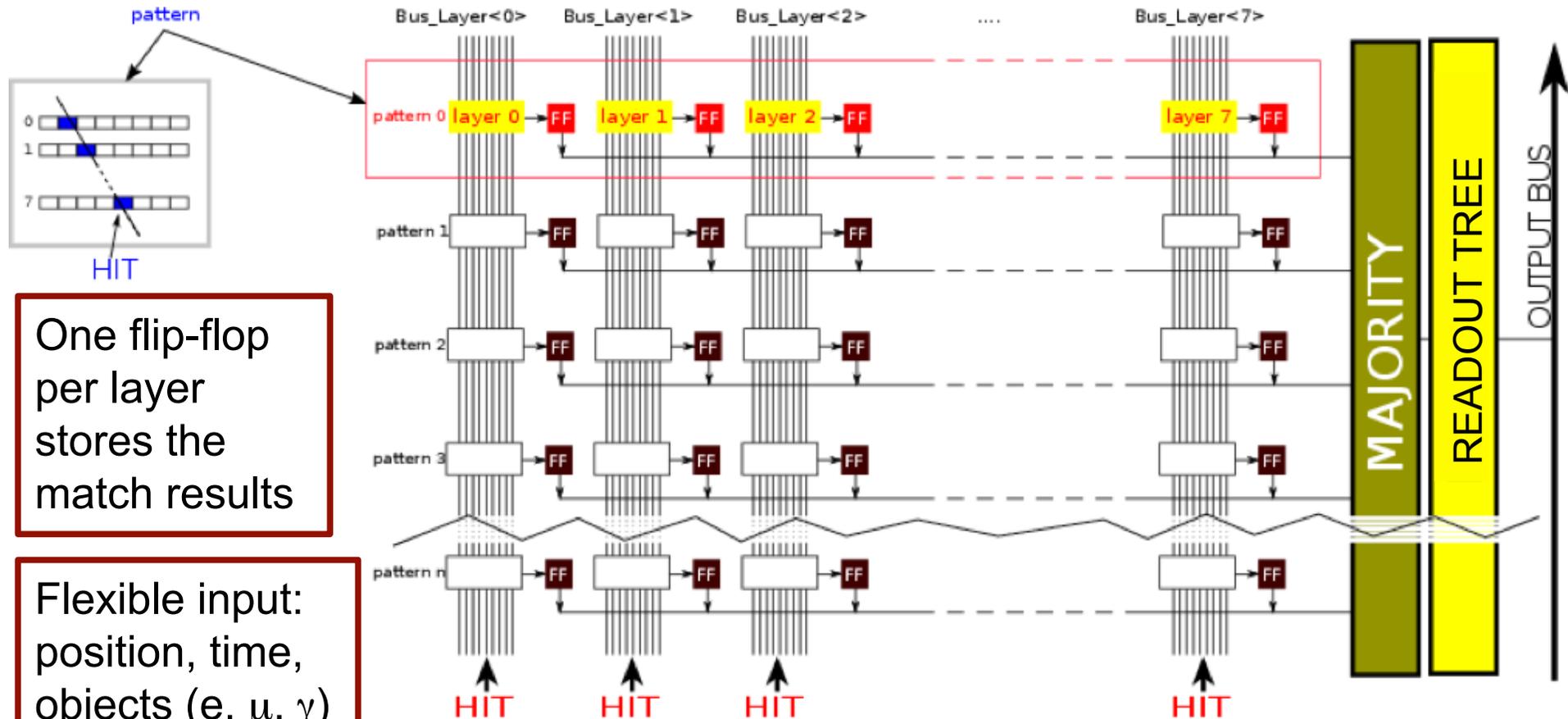
One flip-flop per layer stores the match results

Flexible input: position, time, objects (e, μ, γ)

Pattern matching is completed as soon as all hits are loaded.
Data arriving at different times is compared in parallel with all patterns.
Unique to AM chip: look for correlation of data received at different times.

AM technological evolution

SVT AM chip

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

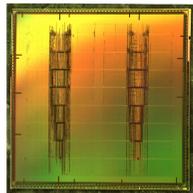F. Morsani et al., IEEE Trans. on Nucl. Sci., vol. 39 **(1992)**

Alternative **FPGA** implementation of SVT AM chip

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

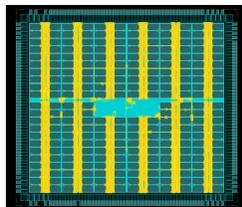G Magazzù, 1st std cell project presented @ LHCC (1999)

SVT upgrade

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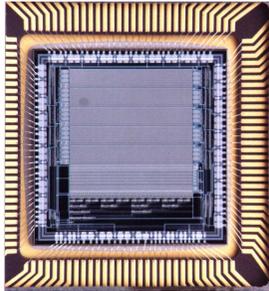A. Annovi et al., **IEEE TNS,** Vol 53, Issue 4, Part 2, **2006**

FTK R&D

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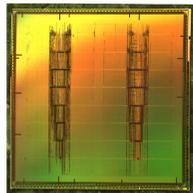
AM technological evolution

SVT AM chip

- (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)**
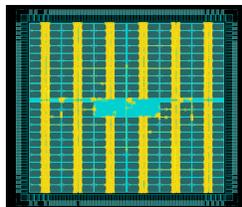
Alternat

P. Gian

G Maga

FTK R&D in progress:
AMchip05 prototype: switch to serialzed IO (11*2Gpbs)
          sumbission fall 2013
AMchip06 FTK final: increase area goal 128k patterns/chip
          submission spring 2014

(current) future applications:
ATLAS & CMS L1 Track triggers for High Luminosty LHC
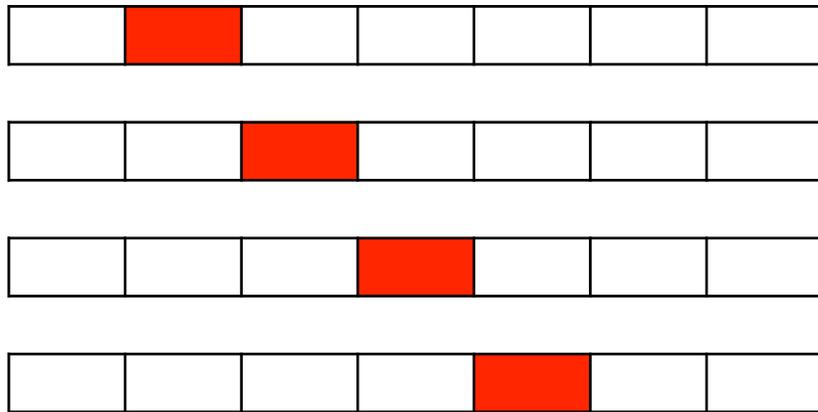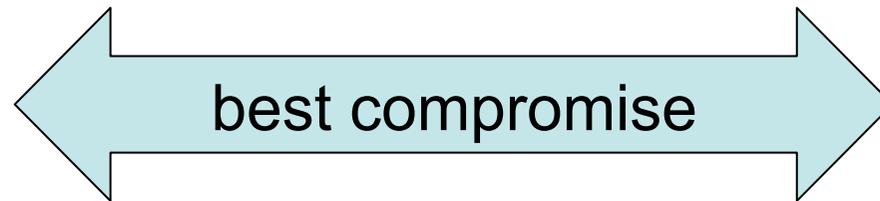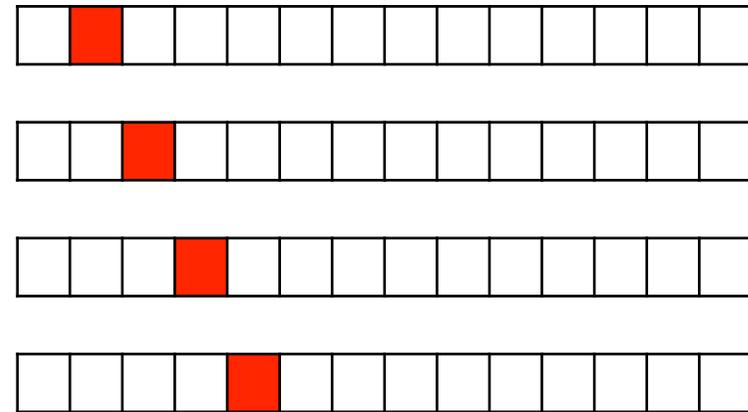Mark Sutton's talk later today

SVT upgrade

FTK R&D

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

Generatig the pattern bank

Wide patterns   **Thin patterns**

best compromise

High efficiency
with less patterns (hardware)
BUT more fakes

More patterns (hardware)
for same efficiency less fakes
Fakes are workload for track fitter

Pattern bank size and efficiency

Pattern size
r-ϕ: 24 pixels, 20 SCT strips
z: 36 pixels

Pattern size (half size)
r-ϕ: 12 pixels, 10 SCT strips
z: 36 pixels

Coverage: pattern efficiency
Efficiency: track efficiency

65M

500M
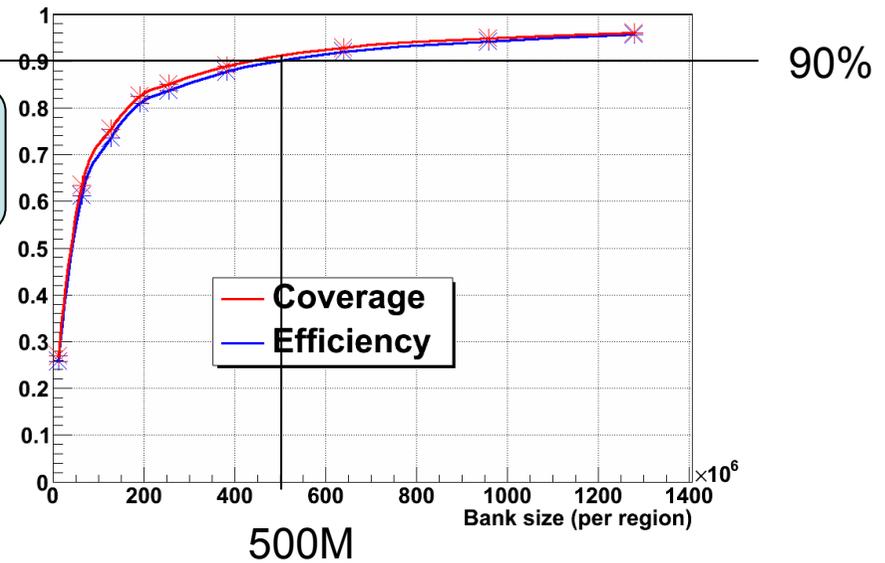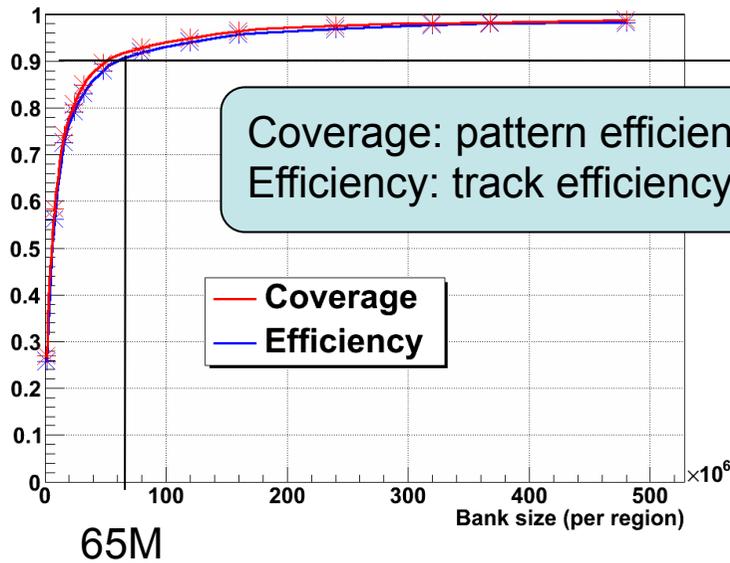
# of patterns in AM chips (barrel only, 45 ϕ degress)
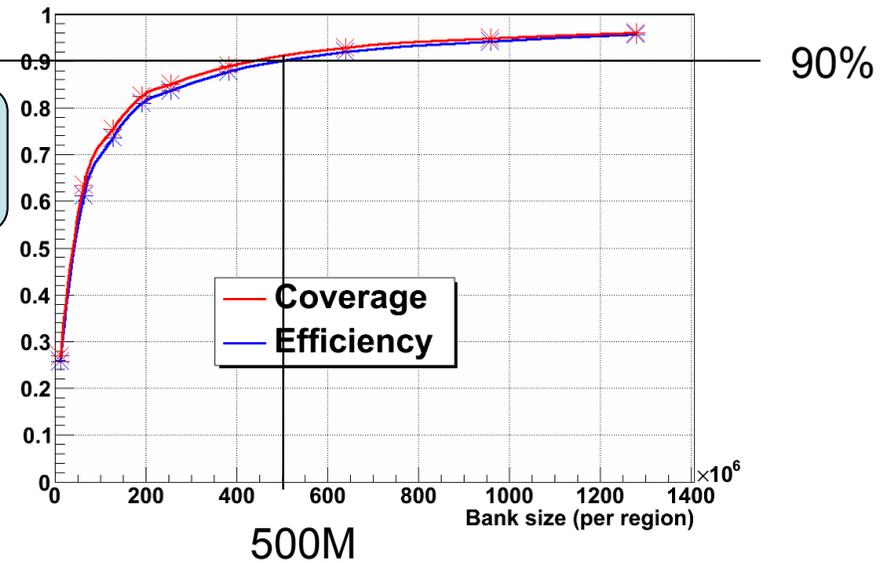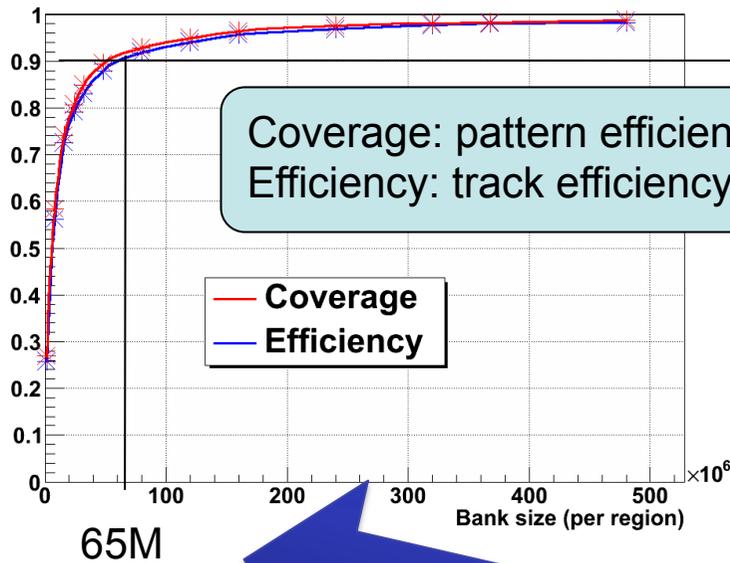
<# matched patterns/event @ 3E34> = 342k   <# matched patterns/event @ 3E34> = 40k

roads (large fake fraction) represents the workload for the track fitter

Pattern bank size and efficiency

Pattern size
r-ϕ: 24 pixels, 20 SCT strips
z: 36 pixels

Pattern size (half size)
r-ϕ: 12 pixels, 10 SCT strips
z: 36 pixels

Coverage: pattern efficiency
Efficiency: track efficiency

65M

500M

90%

# of patterns (barrel only, 45 ϕ degress)

We want this !

<# matched patterns/event @ 3E34> = 342k   <# matched patterns/event @ 3E34> = 40k
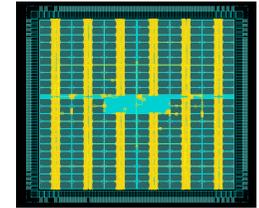
roads (large fake fraction) represents the workload for the track fitter

Variable resolution with "don't care" (DC) bits

AMchip04

- For each layer: a "bin" is identified by a number with DC bits (X)
- Least significant bits of "bin" number can use 3 states (0, 1, X)
- The "bin" number is stored in the Associative Memory
- The DC bits can be used to OR neighborhood high-resolution bins, which differ by few bits, without increasing the number of patterns

Pixels:

0	1	2	3	4	5	6	7
8	9	10	11	12	13	14	15
16	17	18	19	20	21	22	23
24	25	26	27	28	29	30	31

Using binary format
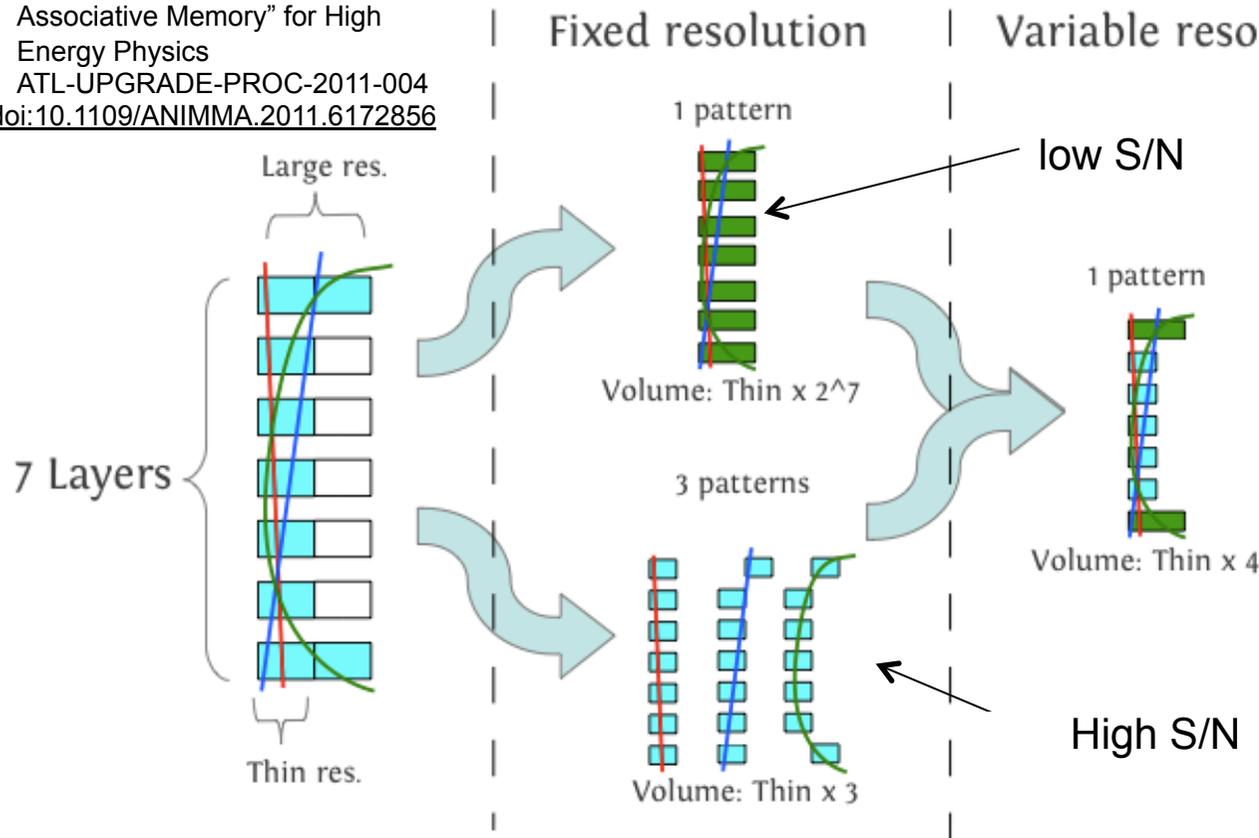**"01010" selects bin 10**
"0001x" selects bins 2 or 3
"1x000" selects bins 16 or 24
"0x11x" selects bins 6,7,14, or 15
"111xx" selects bins 28 to 31

AMCHIP04: VARIABLE RESOLUTION

Fixed resolution

Variable resolution

low S/N

Large res.

7 Layers

Thin res.

1 pattern

Volume: Thin x 2^7

3 patterns

Volume: Thin x 3

1 pattern

Volume: Thin x 4

High S/N

Good rejection and occupy only one pattern location.

Per-pattern choice of optimal resolution.

AMchip04
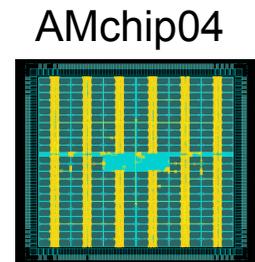
Implemented with the "don't care" feature: inspired by the Ternary CAMs
• Increases the width of a pattern only when needed
• Fully programmable
• Wider patterns can be used in high occupancy regions, smaller patterns in low coverage regions (where the number of trajectories is low, thus reducing the fakes)
• The choice of wider or narrower width patterns is made layer by layer with simulation
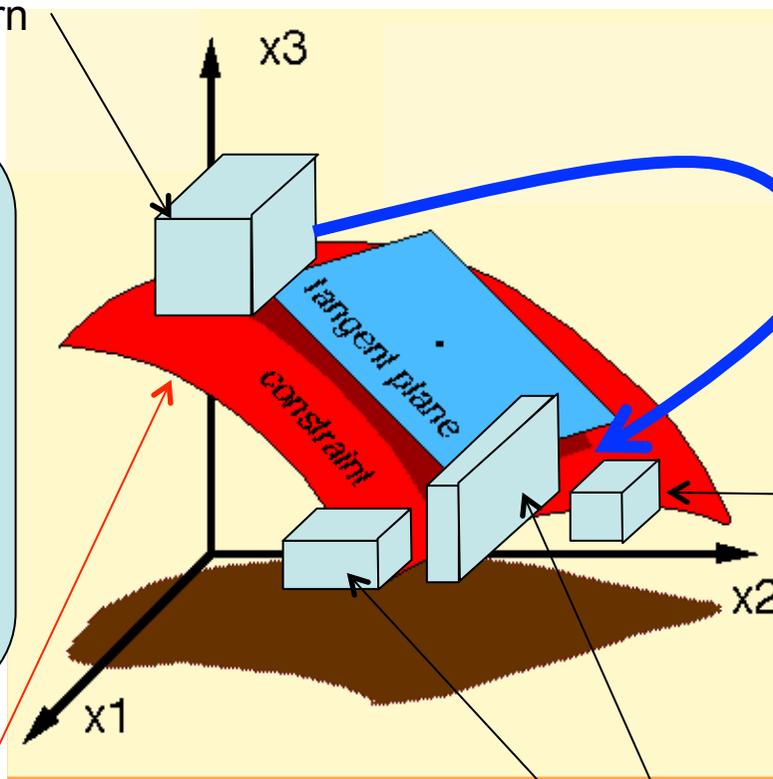
The patterns: a different point of view

Large pattern

5 strip + 3 pixel layers
→11 coordinates
→11D hit coord. space

The pattern bank:
• covers the track manifold with patterns.
• covered space outside manifold → acceptance for fakes
• with variable resolution → dramatically improves S/N

For example: a factor of 2 better resolution on each coordinate
→ a factor 2^11 less volume

Thin pattern

Fixed resolution patterns → fixed aspect ratio

Variable resolution patterns

5D track manifold

Many-bits variable resolution

We can use multiple DC bits to increase the compression factor (up to 6 per layer in a AMchip04)
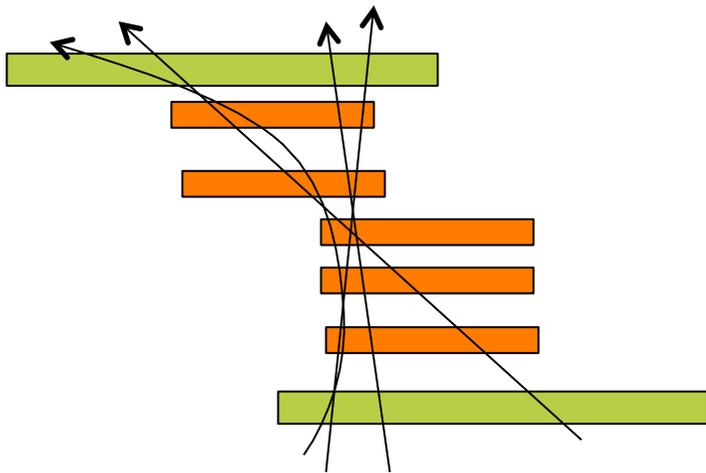
1-bit variable resolution

3-bit variable resolution

1 pattern
Volume 4*
Volume $2^{(7*2)}*4*$ = 2^{16}

1/16 less volume
→ less fakes!!!

1 pattern
Volume 1/4*
Volume 2^{12}

Many-bits variable resolution

We can use multiple DC bits to increase the compression factor (up to 6 per layer in a AMchip04)
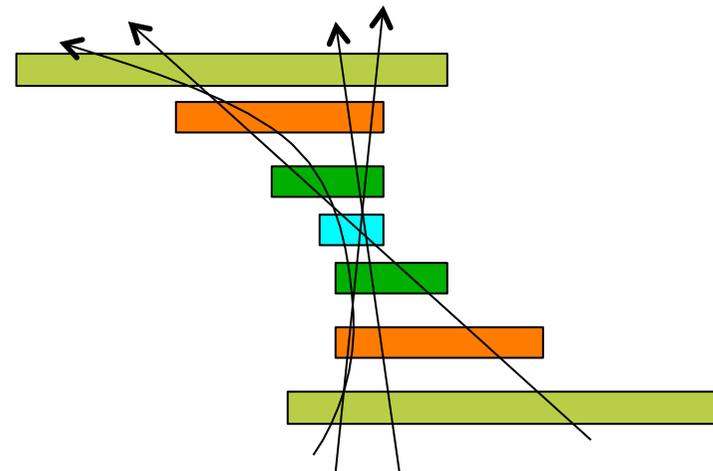
1-bit variable resolution

3-bit variable resolution

Any coincidence based trigger can exploit this technique!!!!

1 pattern
Volume 4*
Volume $2^{(7*2)}*4*$ = 2^{16}

1/16 less volume
→ less fakes!!!

1 pattern
Volume 1/4*
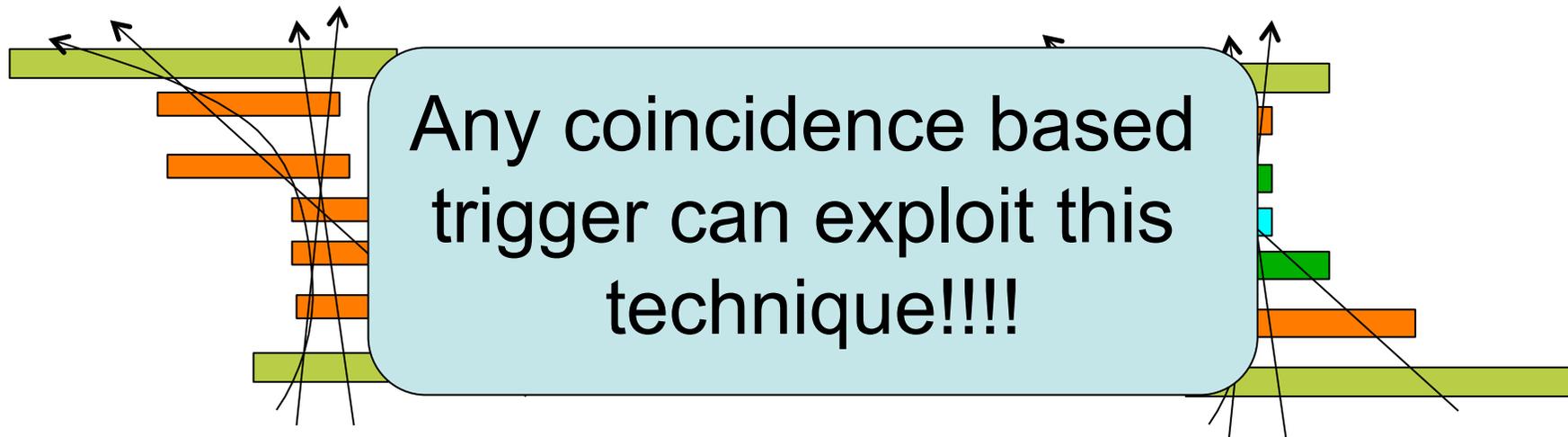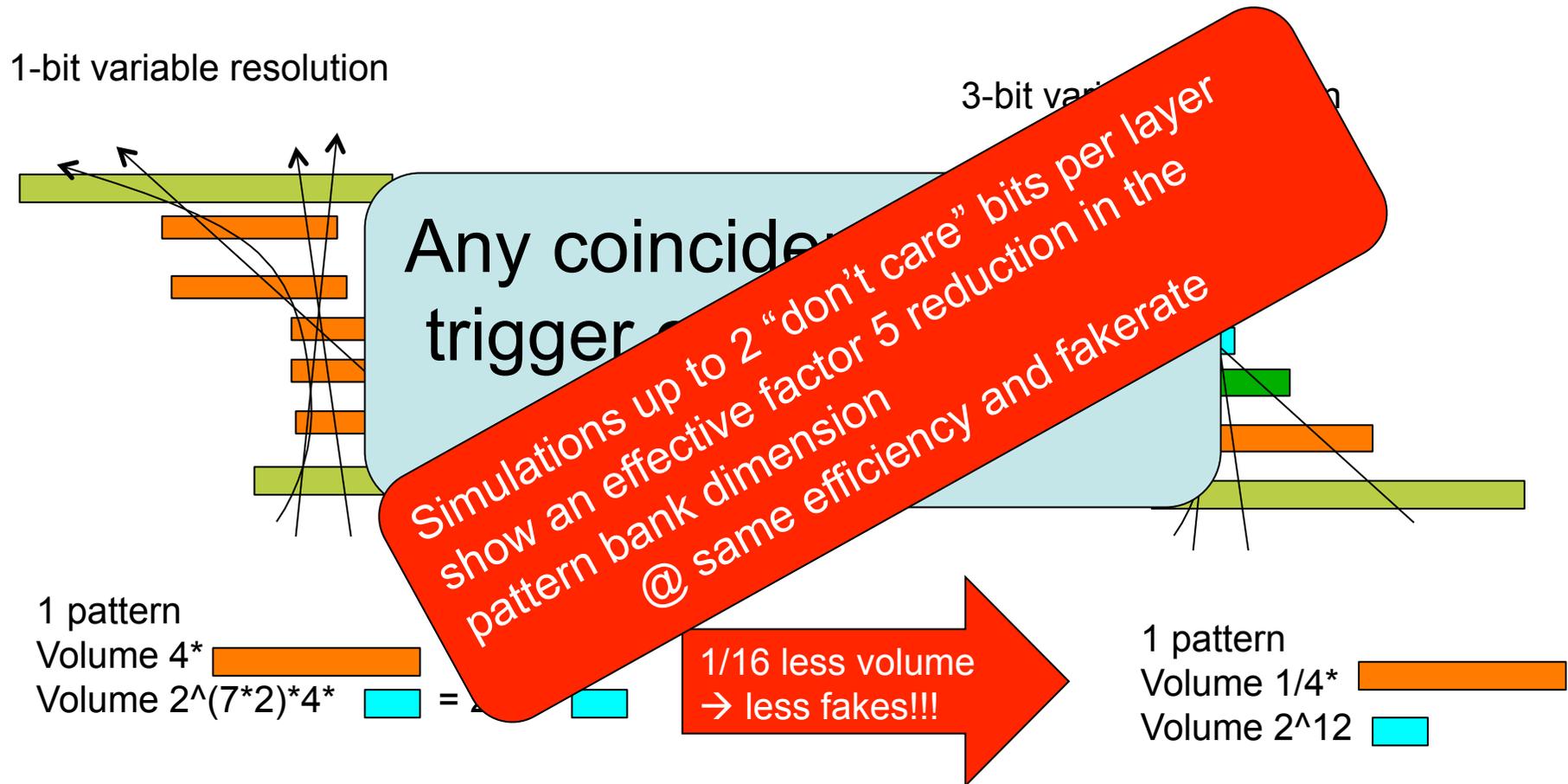Volume 2^{12}

Many-bits variable resolution

We can use multiple DC bits to increase the compression factor (up to 6 per layer in a AMchip04)

1-bit variable resolution

3-bit variable resolution

Any coincidence
trigger

Simulations up to 2 "don't care" bits per layer show an effective factor 5 reduction in the pattern bank dimension @ same efficiency and fakerate

1 pattern
Volume 4*
Volume $2^{(7*2)}*4*$ =

1/16 less volume
→ less fakes!!!

1 pattern
Volume 1/4*
Volume 2^{12}

Performance (max 1 DC/layer)

AM thin channel grouping:
Pixels: 12 along ϕ, 36 along η
Strips: 10 strips
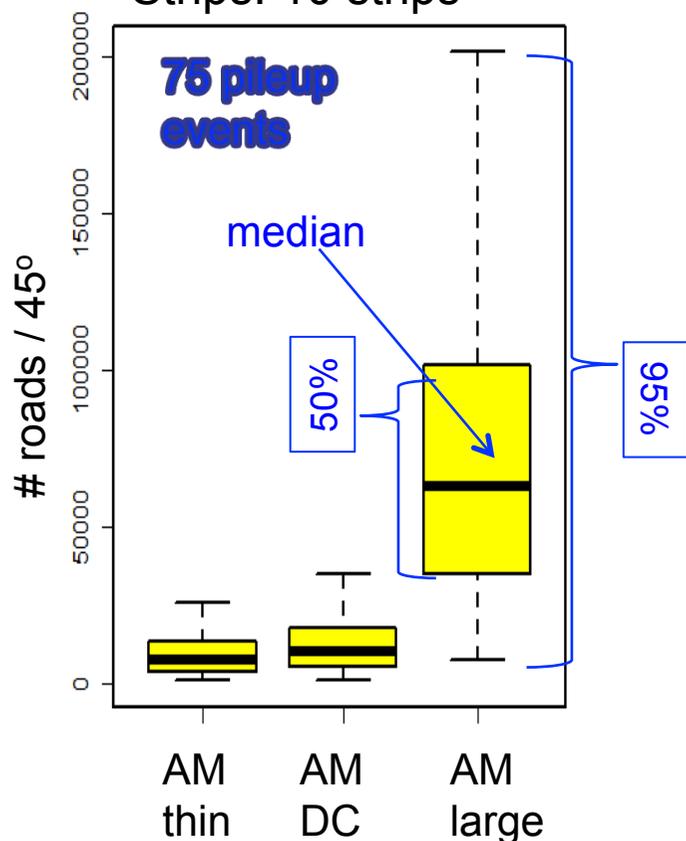
Pileup events	config	Max # DC bits / layer	# roads / 45°	# patterns
75	AM large patterns	0	53500	138M
75	**AM w/ DC**	**1**	**8250**	**138M**
75	AM thin patterns	**0**	5950	384M

- Pattern bank reduction factor: ~ 3
- AM with DC capability reduces the fakes by a large factor: ~ 7
- Good performance with almost same HW

Pattern shape optimization

SVT @ CDF
shape optimization
with fixed resolution AM

CDF NOTE 5201 (1999)
G. Chlachidze, V. Glagolev, G. Punzi
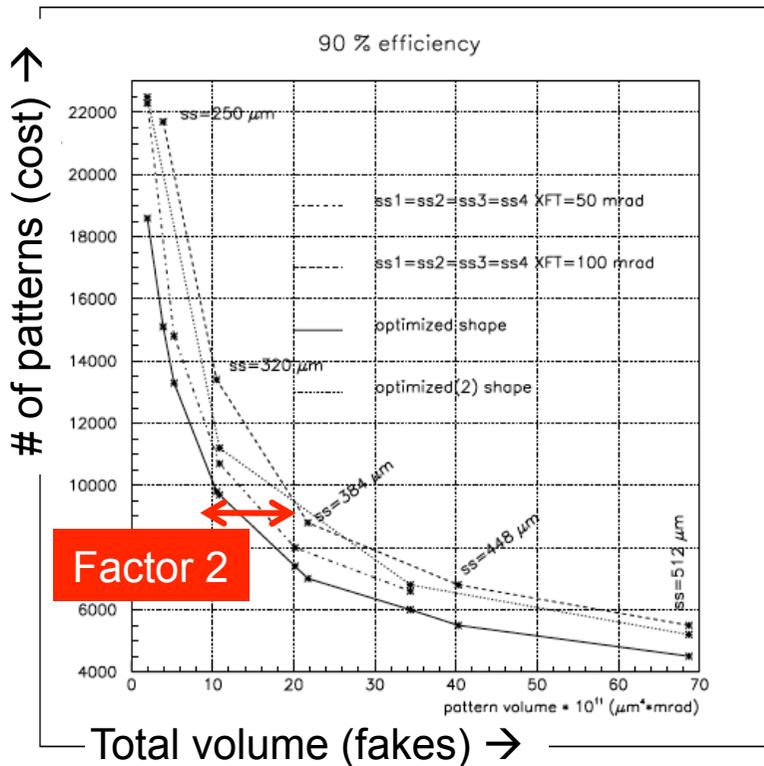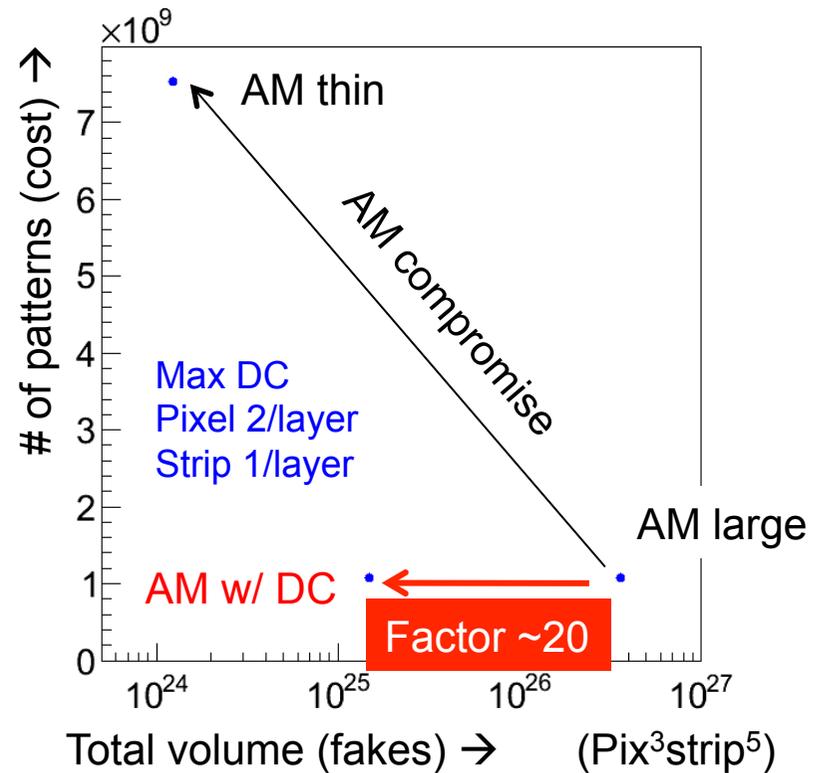
# of patterns (cost) →

Total volume (fakes) →

Figure 3: Effect of different choices of pattern shape: at fixed number of patterns the shape obtained with the formula described in the text yield the smallest volume (continuous line)

FTK @ ATLAS
with "variable resolution AM"
allows a significant improvement
(14 years later)

# of patterns (cost) →

AM thin

AM compromise

Max DC
Pixel 2/layer
Strip 1/layer

AM large

AM w/ DC

Factor ~20

Total volume (fakes) →   (Pix^{3}strip^5)

A working configuration for the Fast Tracker

- High resolution patterns: $(15 \times 36)_{pix} \times 16_{sct}$
  - Pixels: 15 channels along ϕ, 36 ch. along η
  - Strips: 16 strips

- Background events with 69 superimposed pp collisions
  - Instantaneous luminosity 3×10^{34} Hz/cm^2

- Hardware constraints (for each of 64 η-ϕ towers)
  - # AM patterns $< 16.8 \times 10^6$
  - # roads/event $< 16 \times 10^3$
  - # fits/event $< 80 \times 10^3$

> DC bits group detector channels together and increase the pattern resolution

> Work load for track fitter

	Coarse resolution roads	Max # DC bits / layer	# AM pattern * 10^6	Efficiency %	roads / evt * 10^3	fits / evt * 10^3
Barrel	$(30 \times 72)_{pix} \times 32_{sct}$	$2_{pix} \times 1_{sct}$	16.8	93.3%	3.2	26
Endcap	$(30 \times 72)_{pix} \times 32_{sct}$	$2_{pix} \times 1_{sct}$	16.8	91.2%	6.9	55

Summary

- An innovative algorithm has been introduced in the new FTK AM chips
  - The pattern resolution can be configured layer-by-layer and pattern-by-pattern
  - The use of DC bits increases the resolution only where needed
  - High rejection of fake coincidences → the number of roads out of AM is reduced greatly

- Limited "cost": AM chip area ~+15%; power ~+5%

- Equivalent to a factor 3-5 (or more) extra patterns
  - Not fully exploited yet

- Any coincidence based trigger can profit from this feature