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 $\sim 100\mu$s

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

Example:
R-phi view of Barrel region:

- 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.

\[ p_i = \sum_{j=1}^{16} a_{ij} x_j + b_i \]

- \( p_i \)'s are the helix parameters and \( \chi^2 \) components.
- \( x_j \)'s are the hit coordinates in the silicon layers.
- \( a_{ij} \) & \( b_i \) are pre-stored constants determined from full simulation or real data tracks.

» The range of the linear fit is a “sector” which consists of a single silicon module in each detector layer.

- This is VERY fast in FPGA DSPs approx 1 Gfit/s/FPGA

- Based on Principal component analysis
• 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 $\rightarrow$ 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 + $Z\rightarrow 2\mu$
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**

- (90’s) **Full custom VLSI chip** - 0.7 µm (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 x12.
  **First variable resolution implementation!**
  
  F. Alberti *et al* 2013 JINST 8 C01040, doi:10.1088/1748-0221/8/01/C01040
AM technological evolution

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


14th ICATPP Conference
Generating the pattern bank

Wide patterns

Thin patterns

High efficiency with less patterns (hardware)
BUT more fakes

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

best compromise
Pattern bank size and efficiency

Pattern size
- r-\(\phi\): 24 pixels, 20 SCT strips
- z: 36 pixels

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

# of patterns in AM chips (barrel only, 45 \(\phi\) degrees)

\(<#\ \text{matched patterns/event @ } 3E34> = 342k\>

\(<#\ \text{matched patterns/event @ } 3E34> = 40k\>

# roads (large fake fraction) represents the workload for the track fitter
Pattern bank size and efficiency

<table>
<thead>
<tr>
<th>Pattern size</th>
<th>Pattern size (half size)</th>
</tr>
</thead>
<tbody>
<tr>
<td>r-(\phi): 24 pixels, 20 SCT strips</td>
<td>r-(\phi): 12 pixels, 10 SCT strips</td>
</tr>
<tr>
<td>(z): 36 pixels</td>
<td>(z): 36 pixels</td>
</tr>
</tbody>
</table>

**Coverage**: pattern efficiency

**Efficiency**: track efficiency

We want this!

\[ \text{\# matched patterns/event @ 3E34} = 342k \]

\[ \text{\# matched patterns/event @ 3E34} = 40k \]

\# roads (large fake fraction) represents the workload for the track fitter
Variable resolution with “don’t care” (DC) bits

- 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:

<table>
<thead>
<tr>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td>15</td>
</tr>
<tr>
<td>16</td>
<td>17</td>
<td>18</td>
<td>19</td>
<td>20</td>
<td>21</td>
<td>22</td>
<td>23</td>
</tr>
<tr>
<td>24</td>
<td>25</td>
<td>26</td>
<td>27</td>
<td>28</td>
<td>29</td>
<td>30</td>
<td>31</td>
</tr>
</tbody>
</table>

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
A new “Variable Resolution Associative Memory” for High Energy Physics
ATL-UPGRADE-PROC-2011-004
doi:10.1109/ANIMMA.2011.6172856

**AMCHIP04: VARIABLE RESOLUTION**

<table>
<thead>
<tr>
<th>Fixed resolution</th>
<th>Variable resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 pattern</td>
<td>1 pattern</td>
</tr>
<tr>
<td>Volume: Thin x 2^7</td>
<td>Volume: Thin x 4</td>
</tr>
<tr>
<td>3 patterns</td>
<td></td>
</tr>
<tr>
<td>Volume: Thin x 3</td>
<td></td>
</tr>
<tr>
<td>Low S/N</td>
<td></td>
</tr>
<tr>
<td>High S/N</td>
<td></td>
</tr>
</tbody>
</table>

Good rejection and occupy only one pattern location.
Per-pattern choice of optimal resolution.

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

A. Annovi - September 24th, 2013
14th ICATPP Conference
The patterns: a different point of view

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

Variable resolution patterns

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

Large pattern

Thin pattern

Fixed resolution patterns → fixed aspect ratio

5D track manifold
Many-bits variable resolution

We can use multiple DC bits to increase the compression factor (up to 6 per layer in an 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 an 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 pattern
Volume 4*
Volume $2^{(7*2)*4*}$ = 2^16
1/16 less volume → less fakes!!!

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

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 1/4*
Volume $2^{^12}$
Performance (max 1 DC/layer)

AM thin channel grouping:
- Pixels: 12 along $\phi$, 36 along $\eta$
- Strips: 10 strips

<table>
<thead>
<tr>
<th>Pileup events</th>
<th>config</th>
<th>Max # DC bits / layer</th>
<th># roads / 45°</th>
<th># patterns</th>
</tr>
</thead>
<tbody>
<tr>
<td>75</td>
<td>AM large patterns</td>
<td>0</td>
<td>53500</td>
<td>138M</td>
</tr>
<tr>
<td>75</td>
<td>AM w/ DC</td>
<td>1</td>
<td>8250</td>
<td>138M</td>
</tr>
<tr>
<td>75</td>
<td>AM thin patterns</td>
<td>0</td>
<td>5950</td>
<td>384M</td>
</tr>
</tbody>
</table>

- 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

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

# of patterns (cost) \( \rightarrow \)
Total volume (fakes) \( \rightarrow \)

\( \text{90\% efficiency} \)

Factor 2

Max DC
Pixel 2/layer
Strip 1/layer

AM small
AM large

AM thin
AM w/ DC
Factor ~20

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)
A working configuration for the Fast Tracker

- High resolution patterns: \((15\times36)_{\text{pix}} \times 16_{\text{sct}}\)
  - Pixels: 15 channels along \(\phi\), 36 ch. along \(\eta\)
  - Strips: 16 strips
- Background events with 69 superimposed pp collisions
  - Instantaneous luminosity \(3 \times 10^{34} \text{ Hz/cm}^2\)
- Hardware constraints (for each of 64 \(\eta\)-\(\phi\) towers)
  - \# AM patterns < \(16.8 \times 10^6\)
  - \# roads/event < \(16 \times 10^3\)
  - \# fits/event < \(80 \times 10^3\)

<table>
<thead>
<tr>
<th>Coarse resolution roads</th>
<th>Max # DC bits / layer</th>
<th># AM pattern (\times 10^6)</th>
<th>Efficiency %</th>
<th>roads / evt (\times 10^3)</th>
<th>fits / evt (\times 10^3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barrel</td>
<td>((30\times72)<em>{\text{pix}} \times 32</em>{\text{sct}})</td>
<td>(2_{\text{pix}} \times 1_{\text{sct}})</td>
<td>16.8</td>
<td>3.2</td>
<td>26</td>
</tr>
<tr>
<td>Endcap</td>
<td>((30\times72)<em>{\text{pix}} \times 32</em>{\text{sct}})</td>
<td>(2_{\text{pix}} \times 1_{\text{sct}})</td>
<td>16.8</td>
<td>6.9</td>
<td>55</td>
</tr>
</tbody>
</table>
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