Performance Studies of the Associative Memory System of the ATLAS Fast Tracker

2018. Nov. 13 (Tue.)

Tomoya Iizawa
University of Geneva
for the ATLAS FTK team

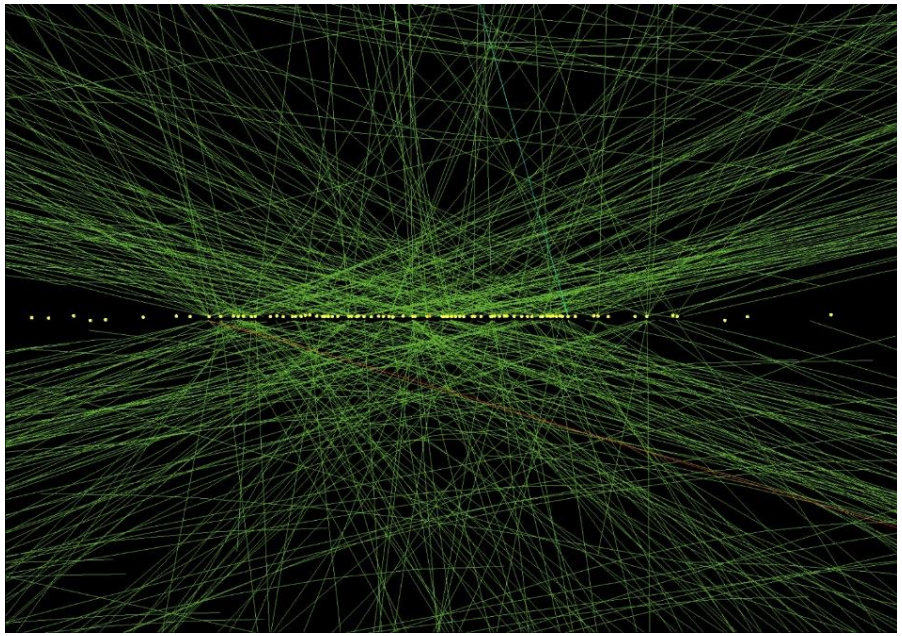
IEEE-NSS 2018
@ International Convention Center
Sydney, Australia

Challenge in High Luminosity Collider : Pile-up

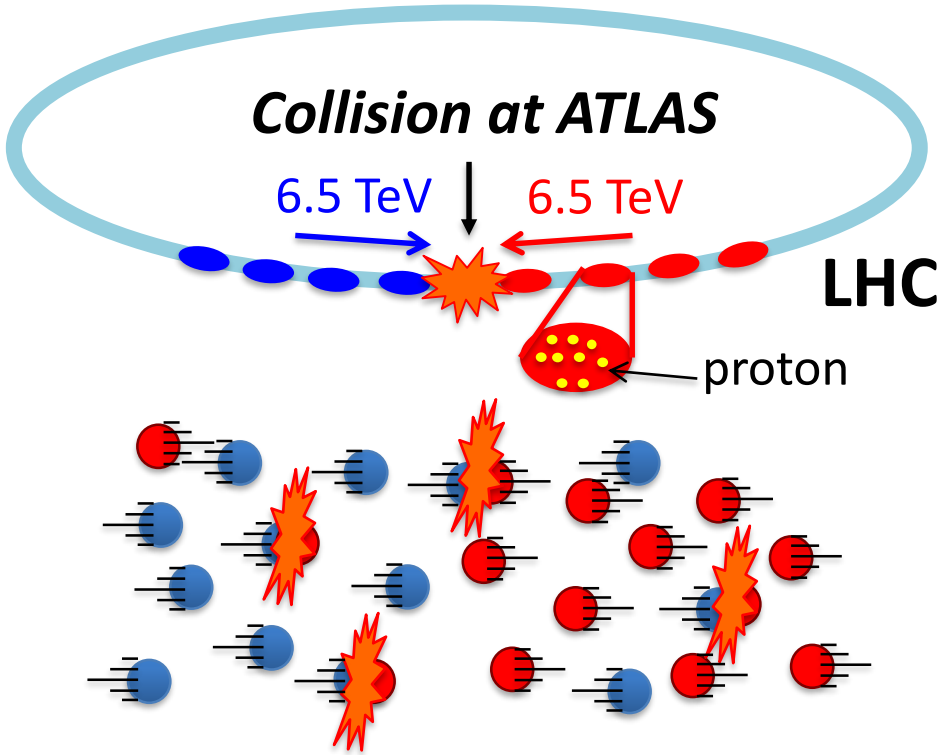
There are a large number of proton interactions per bunch crossing (pile-up).
The number of the pile-ups is expected to increase at LHC run3 starting from 2021.
→ Signal/background discrimination will become more and more difficult.

Track information is critical to distinguish primary vertices and pile-up.

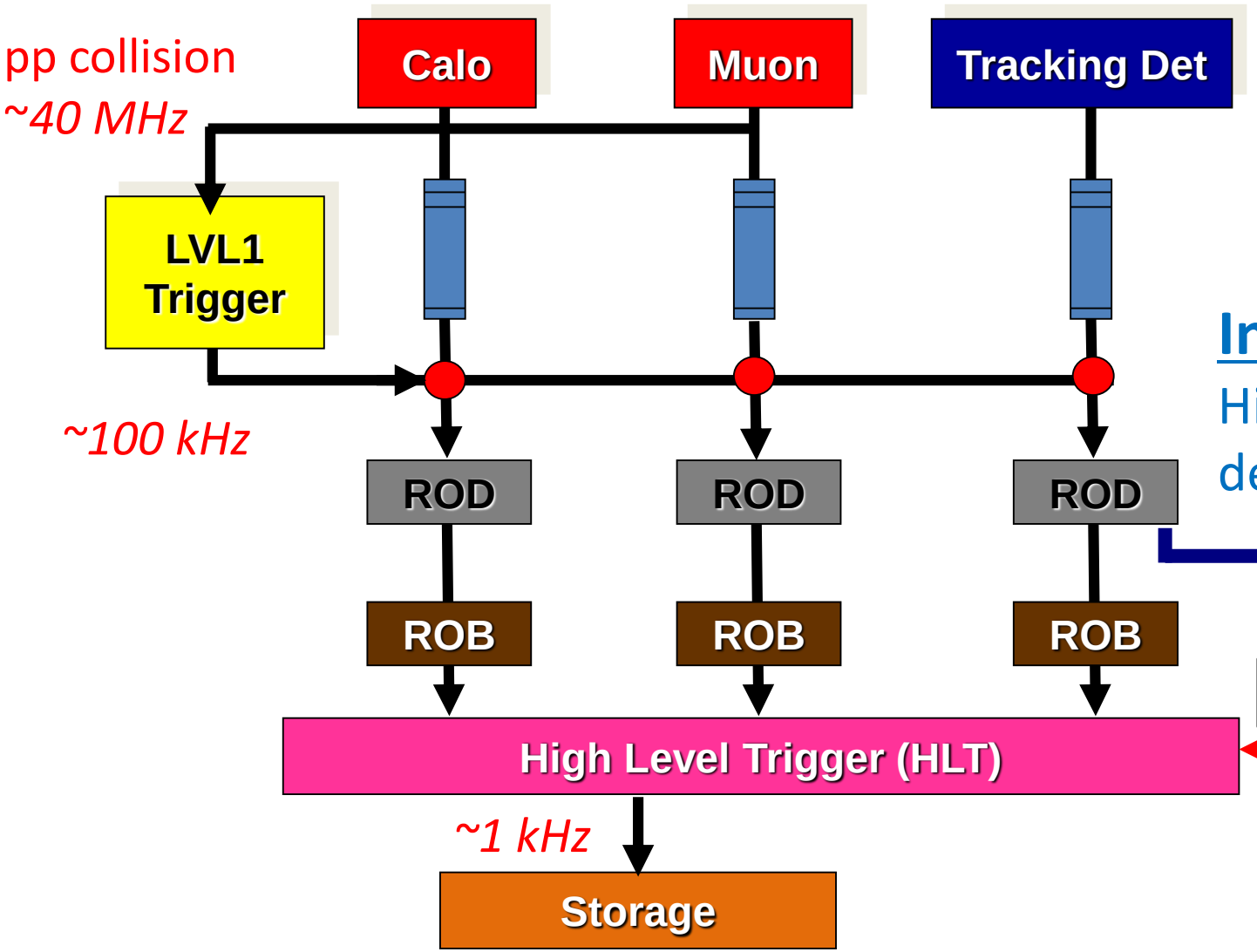
→ The Fast Tracker (FTK) is newly installed dedicated hardware system to reconstruct all tracks in this pile-up situation, which is not possible with currently existing software system.

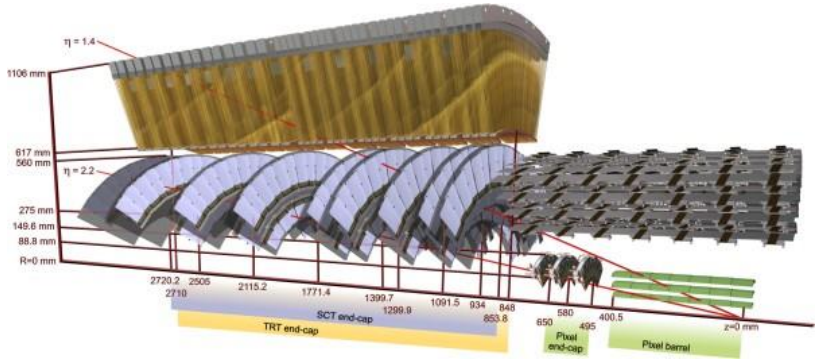


pile-up expected at the run in 2021



ATLAS Trigger System with FTK





Input:

Hit information from inner tracking detector (IBL/PIX/SCT, 12Layer) @ 100 kHz

Output:

p_T, η, ϕ, d_0, z_0

for all tracks with $p_T > 1$GeV

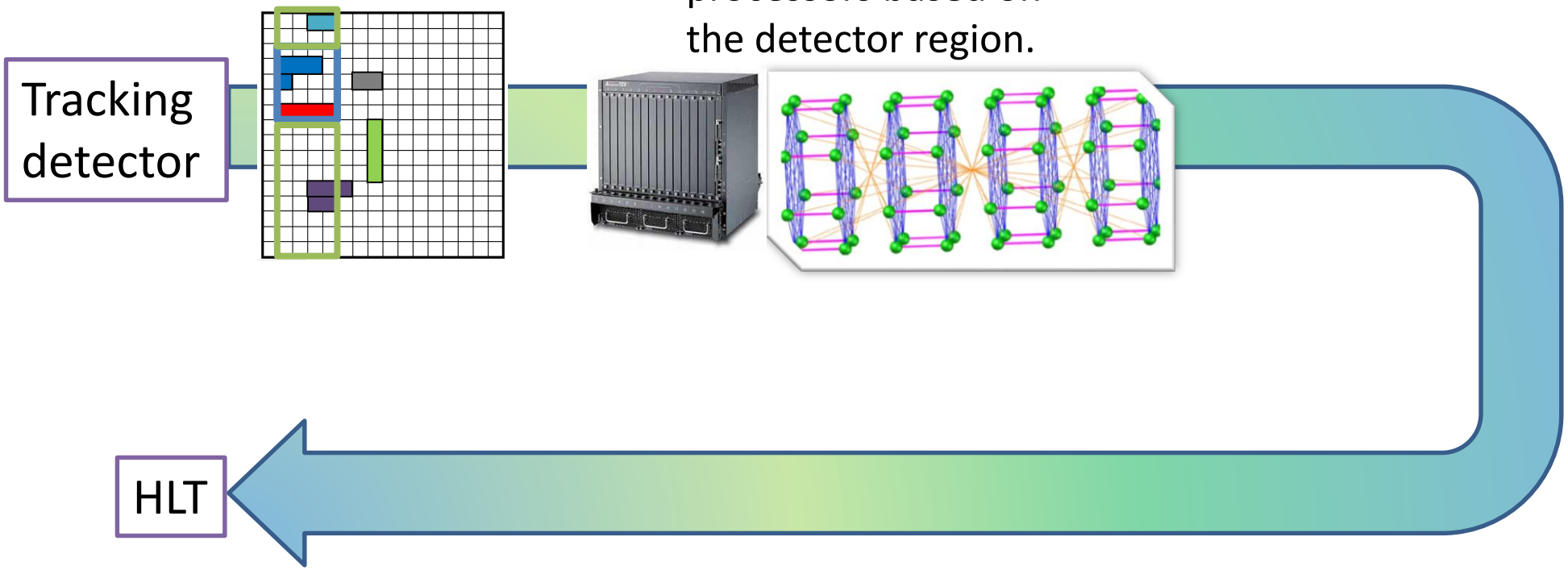
FTK reconstructs tracks and provides them to the HLT.

→ HLT can use more sophisticated selection with utilizing the tracks and extra processing time.

FTK Processing Schemes

Clustering
Format data to be fit with FTK processing.

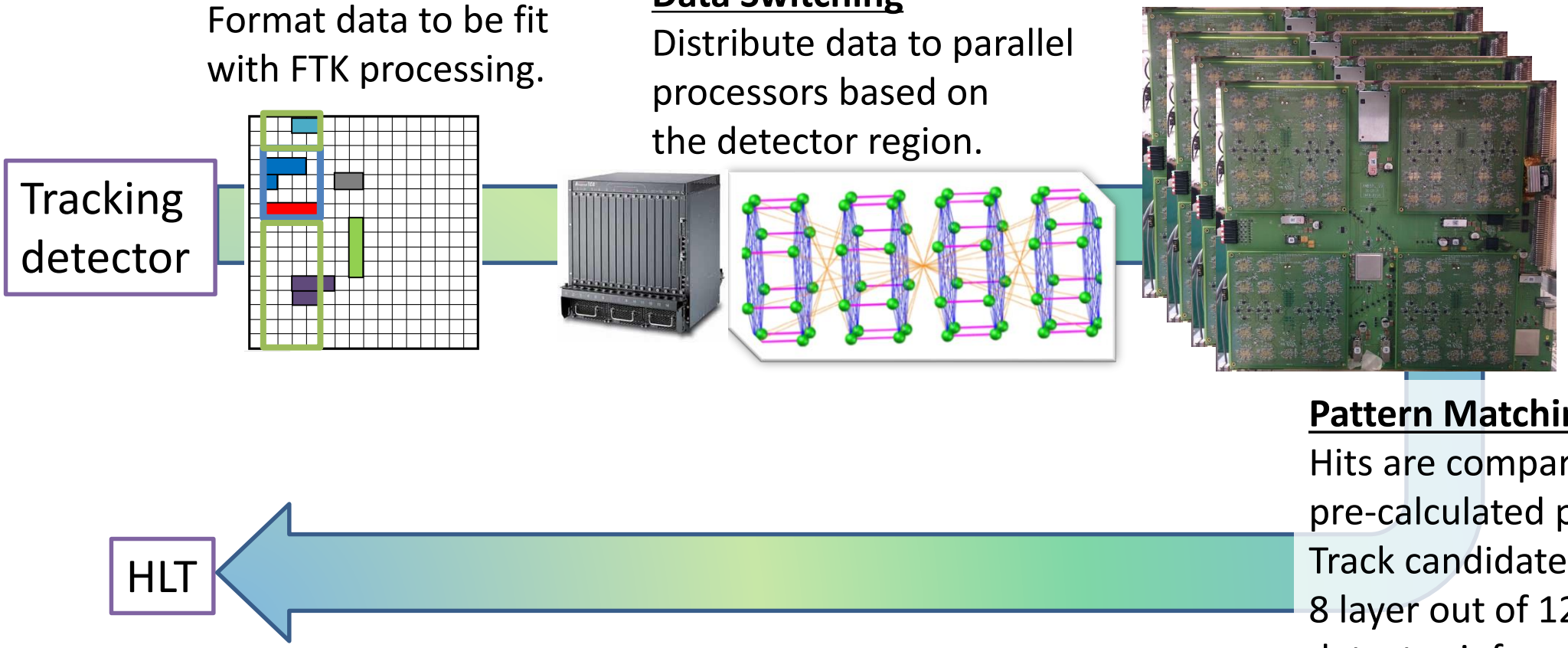
Data Switching
Distribute data to parallel processors based on the detector region.



FTK Processing Schemes

Clustering
Format data to be fit with FTK processing.

Data Switching
Distribute data to parallel processors based on the detector region.

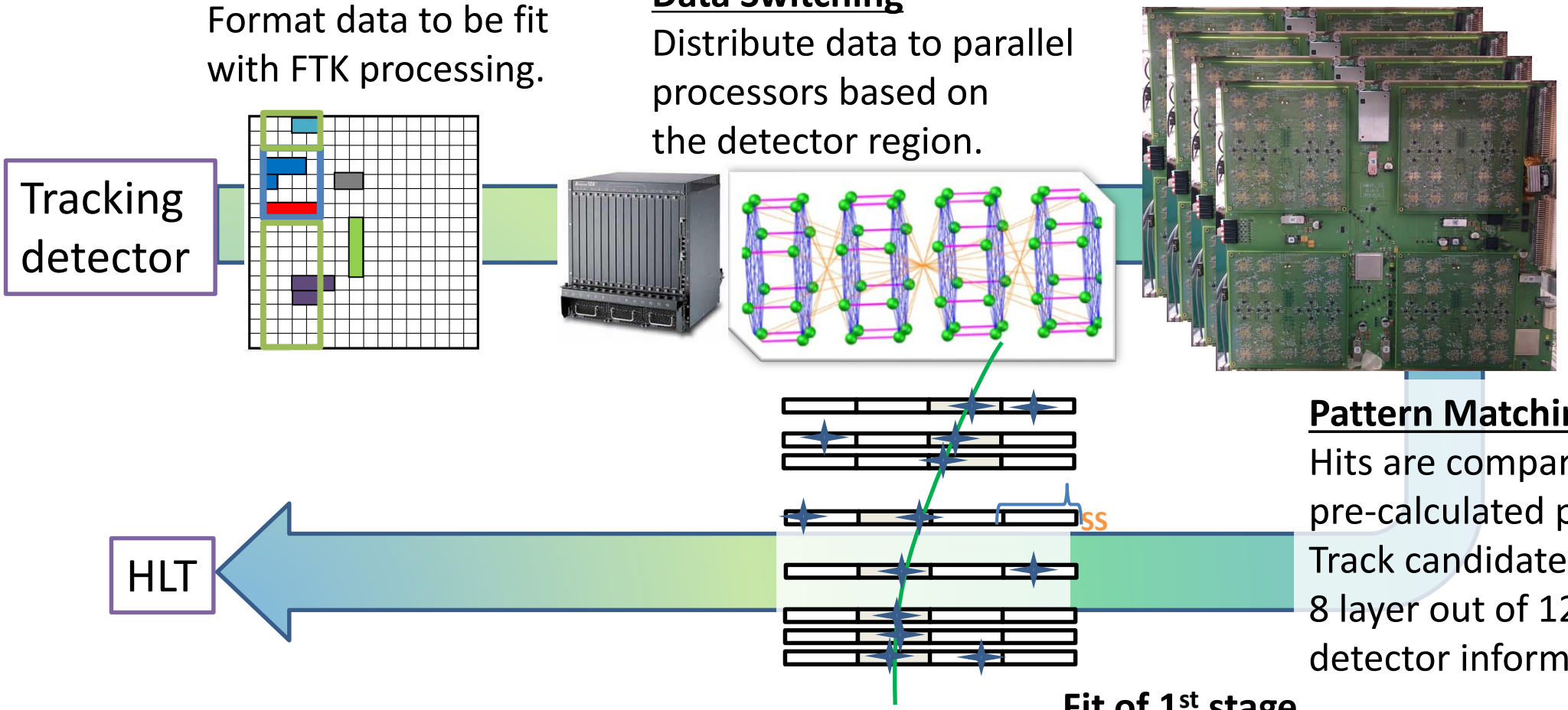


Pattern Matching
Hits are compared with pre-calculated patterns.
Track candidates are found.
8 layer out of 12 layer detector information used.

FTK Processing Schemes

Clustering
Format data to be fit with FTK processing.

Data Switching
Distribute data to parallel processors based on the detector region.



Pattern Matching
Hits are compared with pre-calculated patterns.
Track candidates are found.
8 layer out of 12 layer detector information used.

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

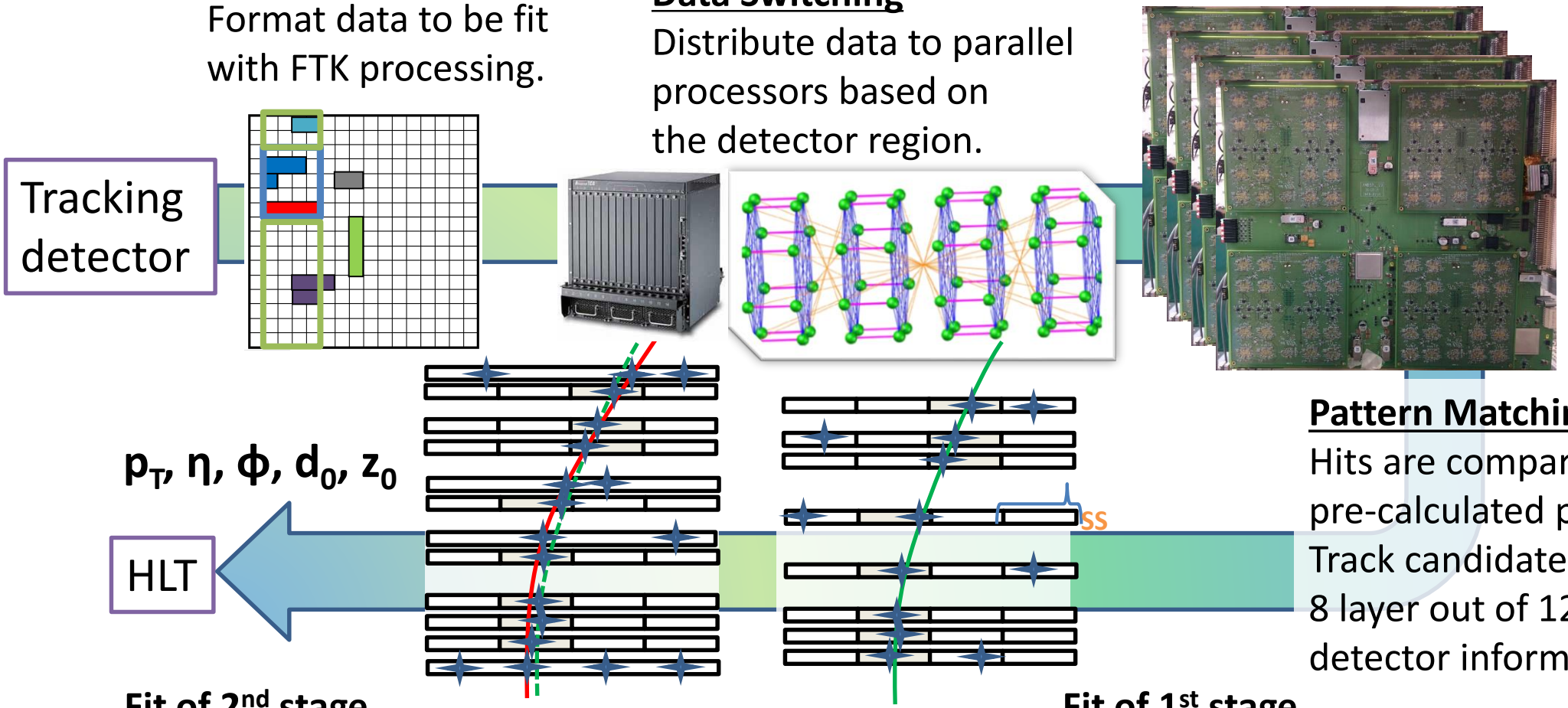
Fit of 1st stage
Linear approximation are performed with 8-layers.
χ^2 component are calculated.

FTK Processing Schemes

Massive parallel processing hardware

Clustering
Format data to be fit with FTK processing.

Data Switching
Distribute data to parallel processors based on the detector region.



Pattern Matching
Hits are compared with pre-calculated patterns.
Track candidates are found.
8 layer out of 12 layer detector information used.

Fit of 2nd stage
Tracks are refit with full 12-layer hit information.
→ Reduce fakes, improve resolution.

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

Fit of 1st stage
Linear approximation are performed with 8-layers.
χ^2 component are calculated.

FTK System Overview

Data Formatter (DF)

Clustering
Data sharing

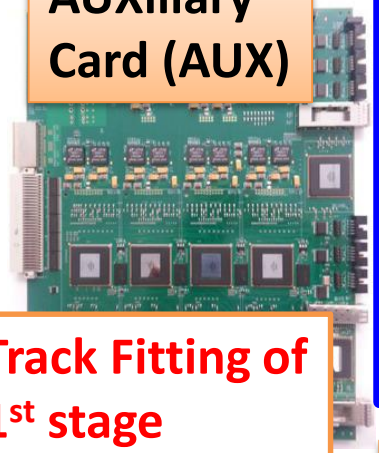
AUXiliary
Card (AUX)

Associative Memory
(AM)

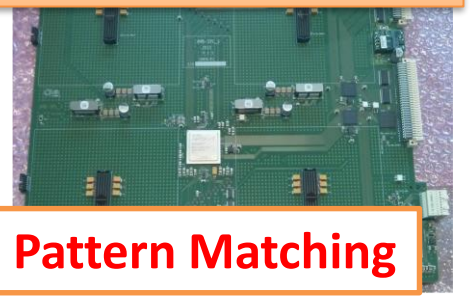
Today's
talk

Dual-output HOLA

Copy info. from ROD,
send it to FTK.

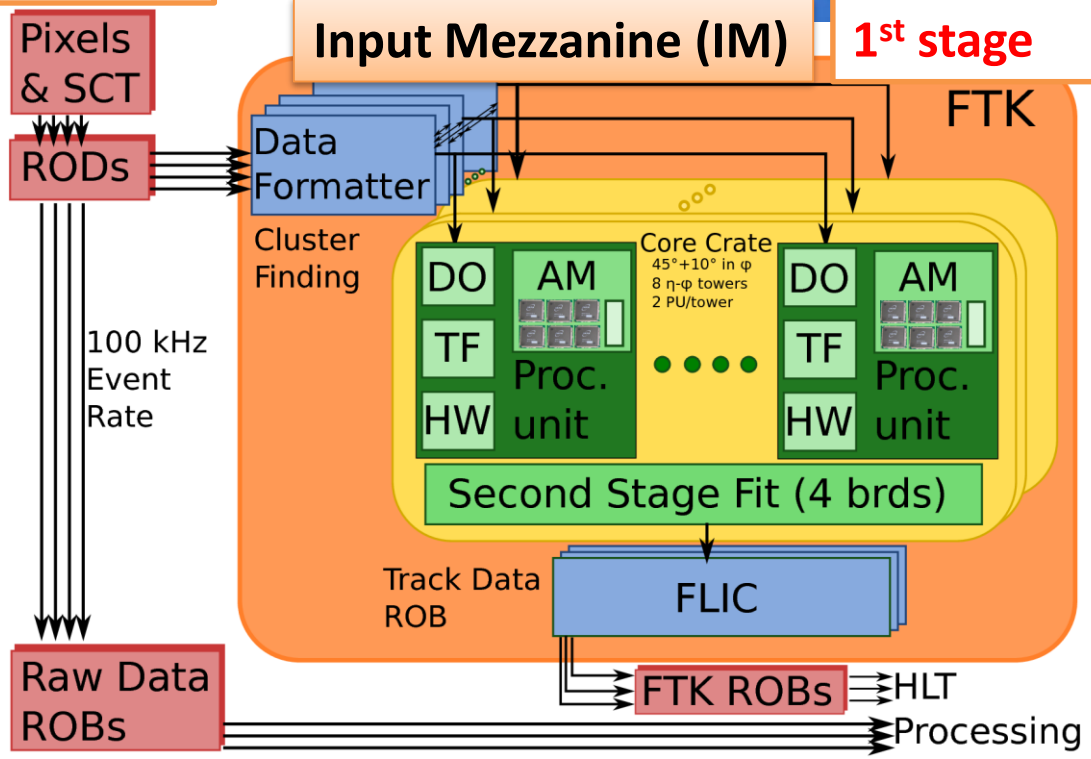


Track Fitting of
1st stage

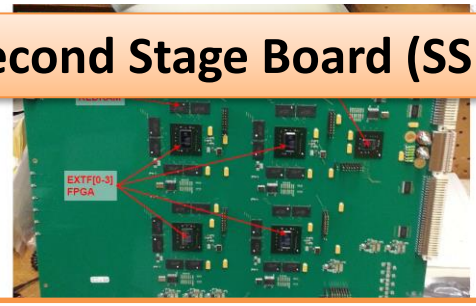


Pattern Matching

Input Mezzanine (IM)



Second Stage Board (SSB)



Track Fitting of 2nd stage

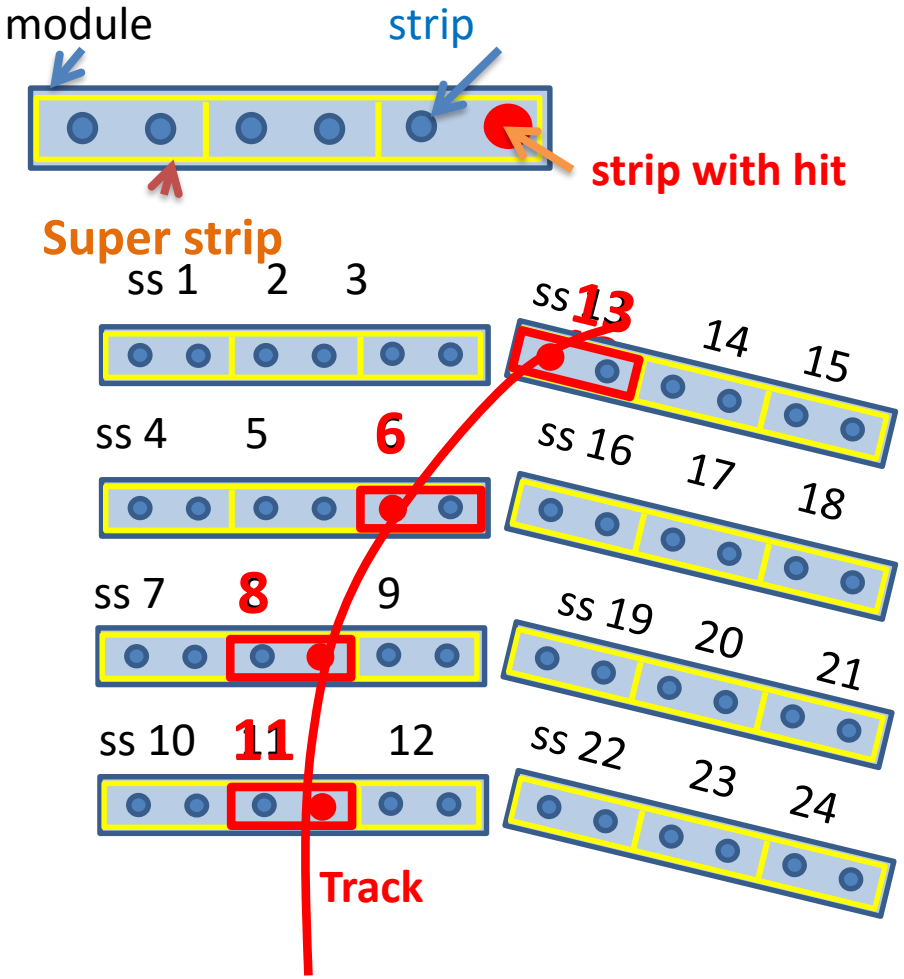
FTK to L2 Interface Crate (FLIC)

Send track information

>8000 65nm ASIC chips, >2000 FPGAs

Key Algorithm: Pattern Matching

To achieve fast tracking, pattern matching is performed. Hits are compared with pre-calculated patterns based on coarse resolution hits (**Super-Strip: SS**). The matched pattern is called as **Road**.



Actually 1 billion patterns are used

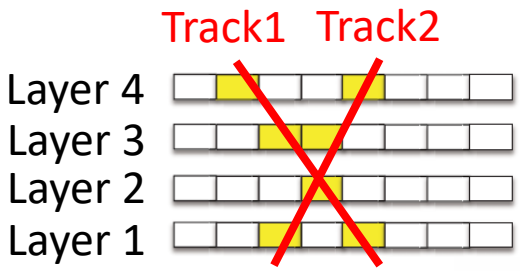
Pattern Bank

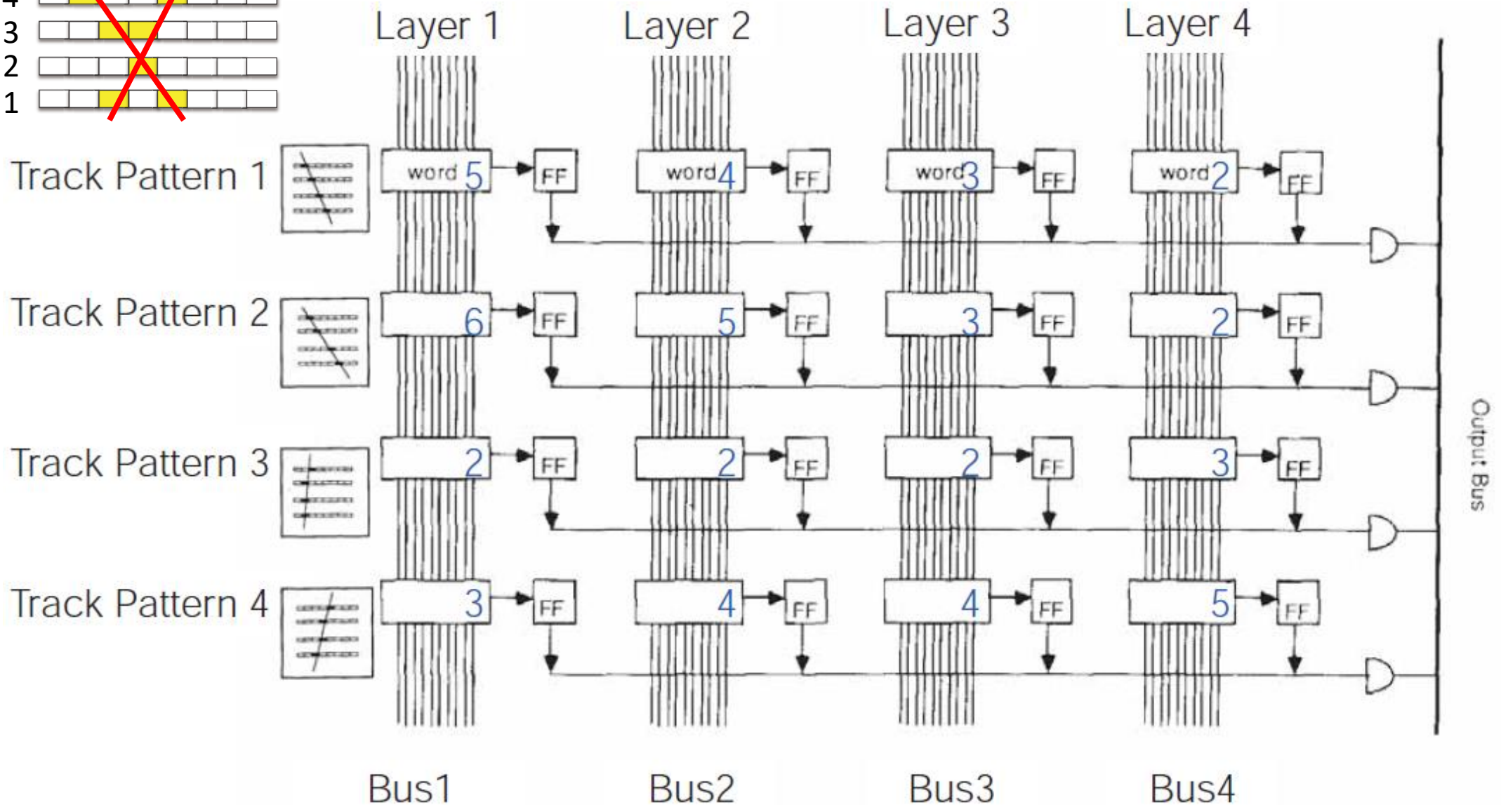
Road 1	2	3	4	5
SS 1	SS 2	SS 3	SS 13	SS 3
SS 4	SS 5	SS 6	SS 6	SS 5
SS 7	SS 8	SS 8	SS 8	SS 8
SS 10	SS 11	SS 10	SS 11	SS 10

✓

(Toy detector figure assuming 4 layers. Actually, 8 layers are used for pattern matching.)

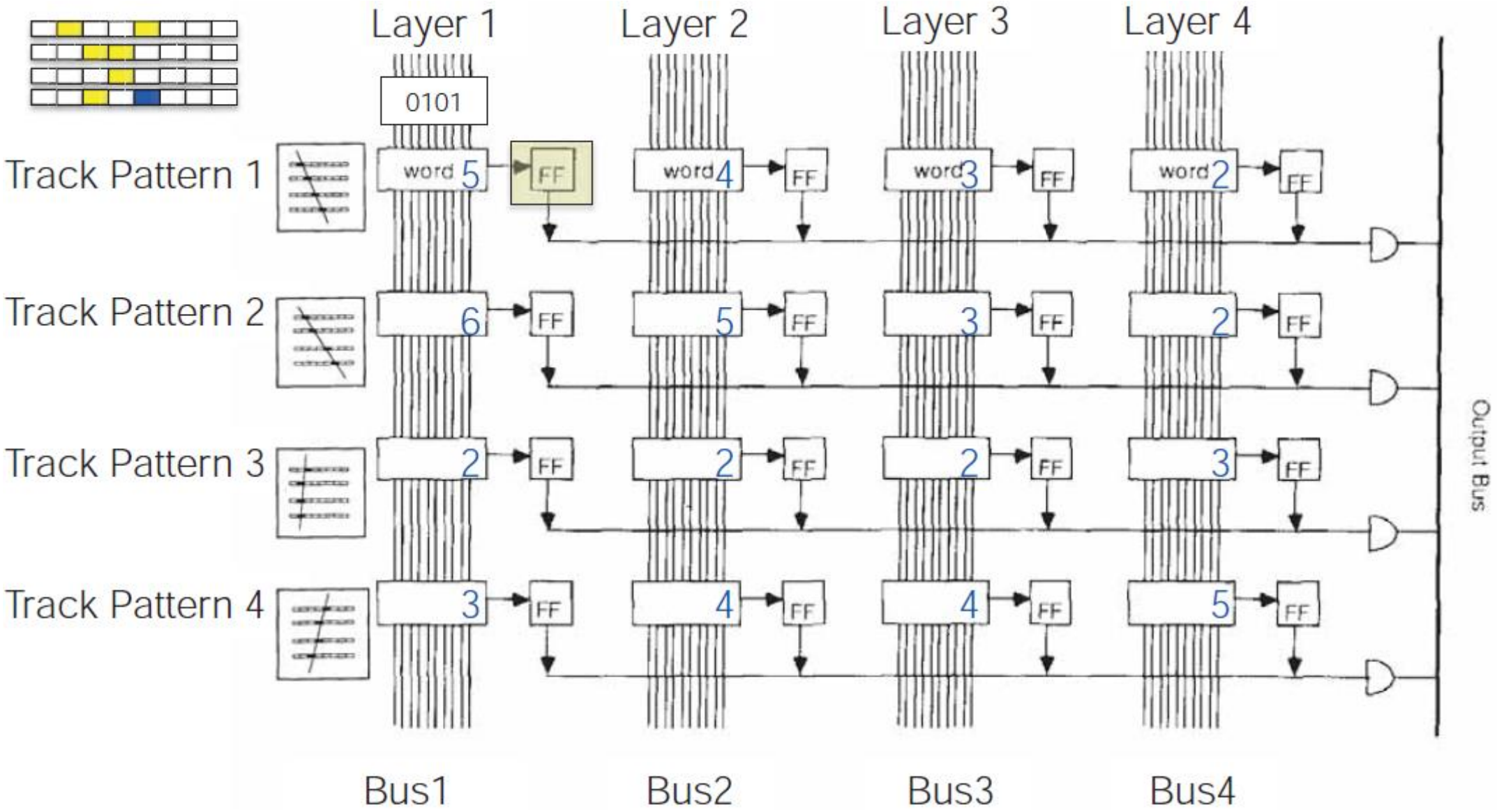
Pattern Matching



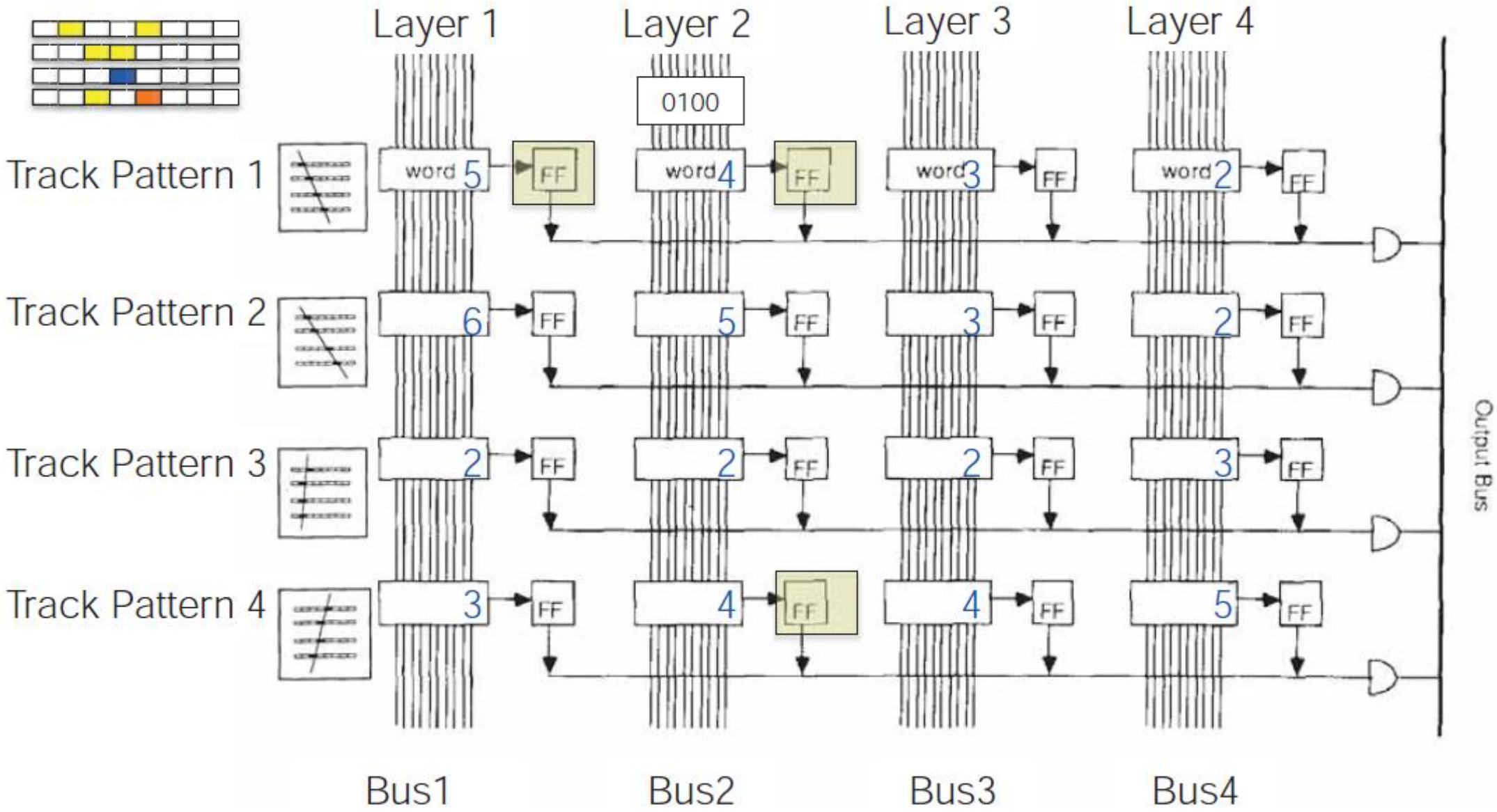


Real FTK uses 8 layers for pattern matching. This figure has 4 layers and 4 patterns as an example.

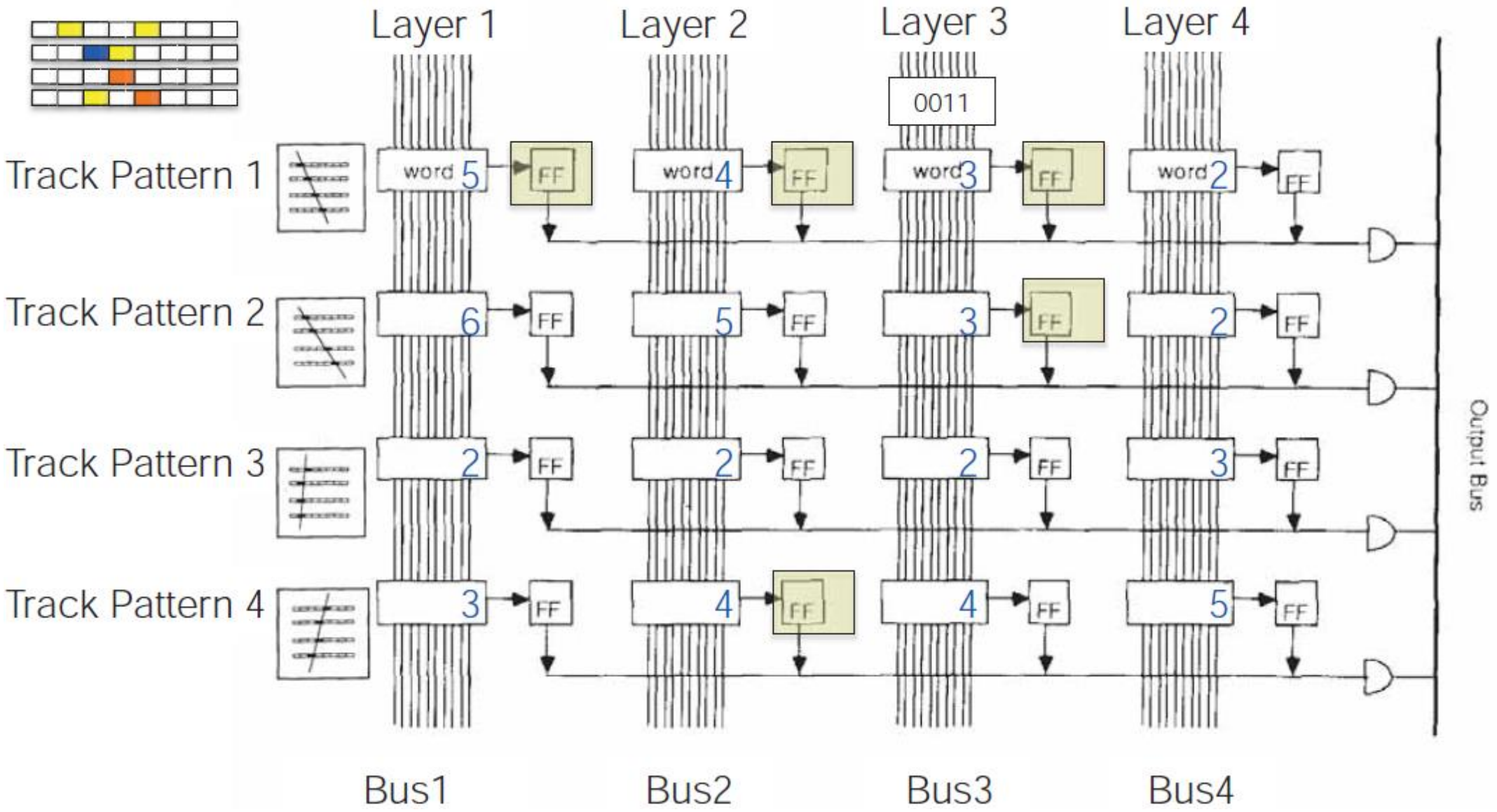
Pattern Matching



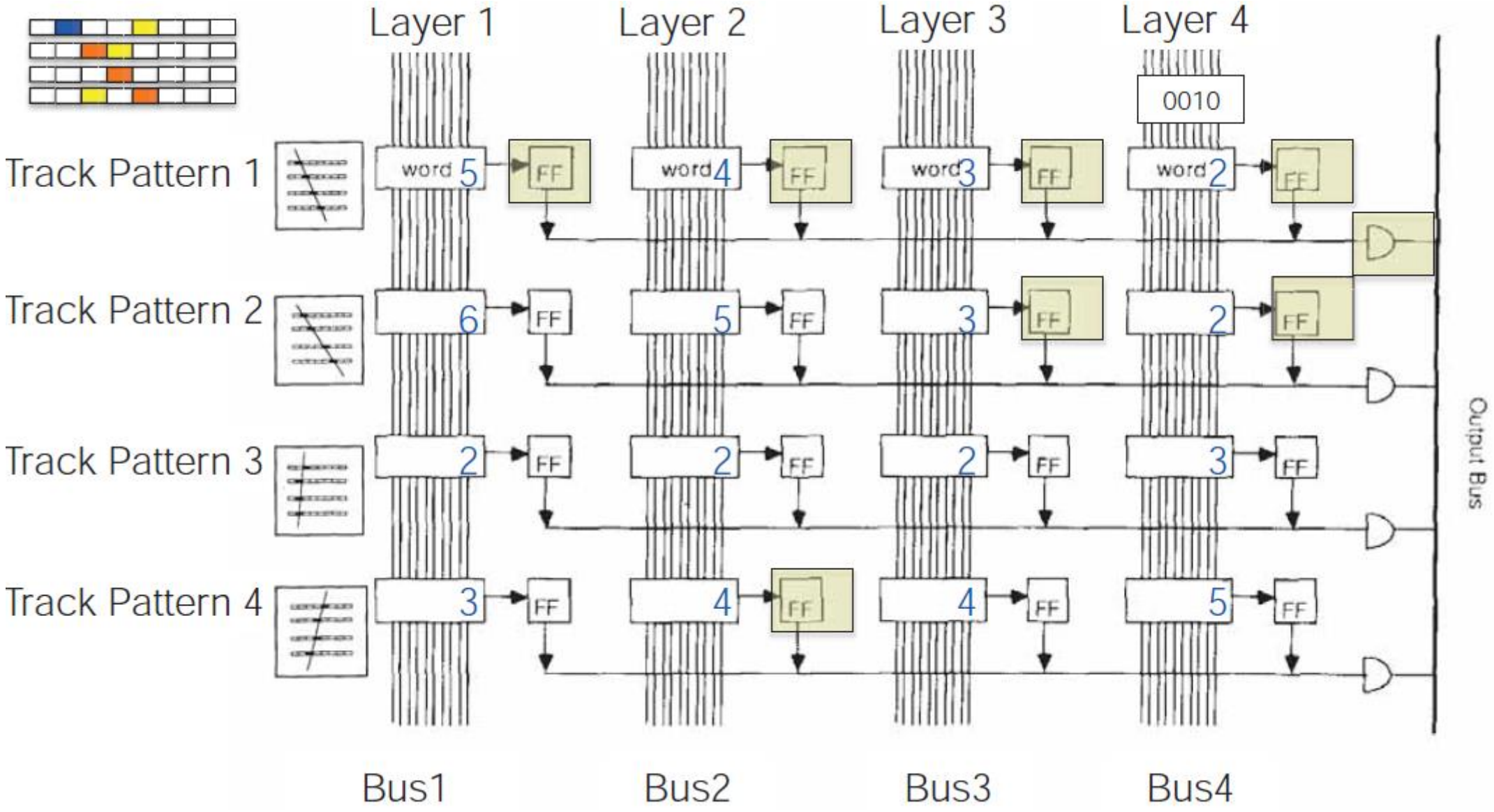
Pattern Matching



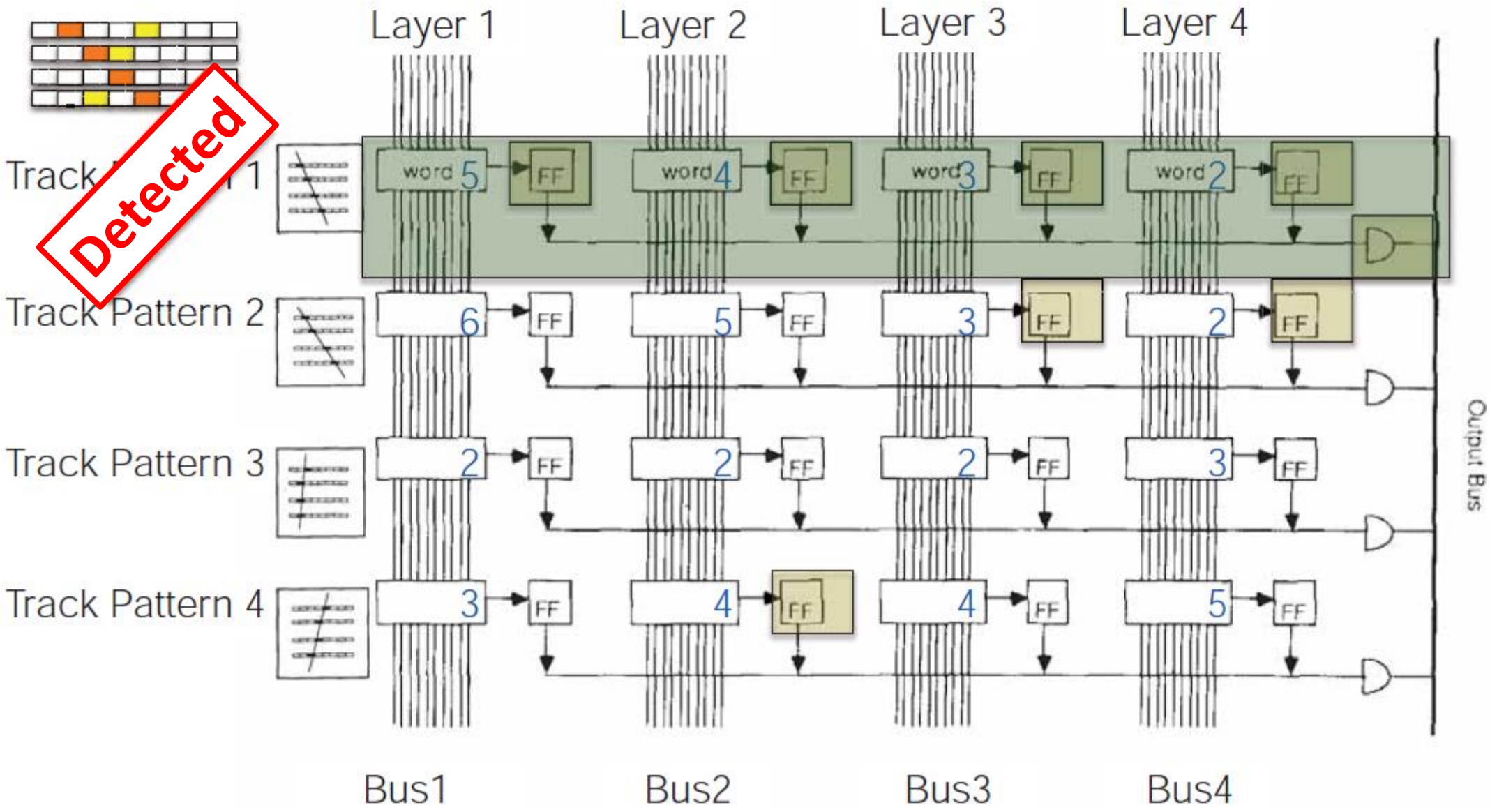
Pattern Matching



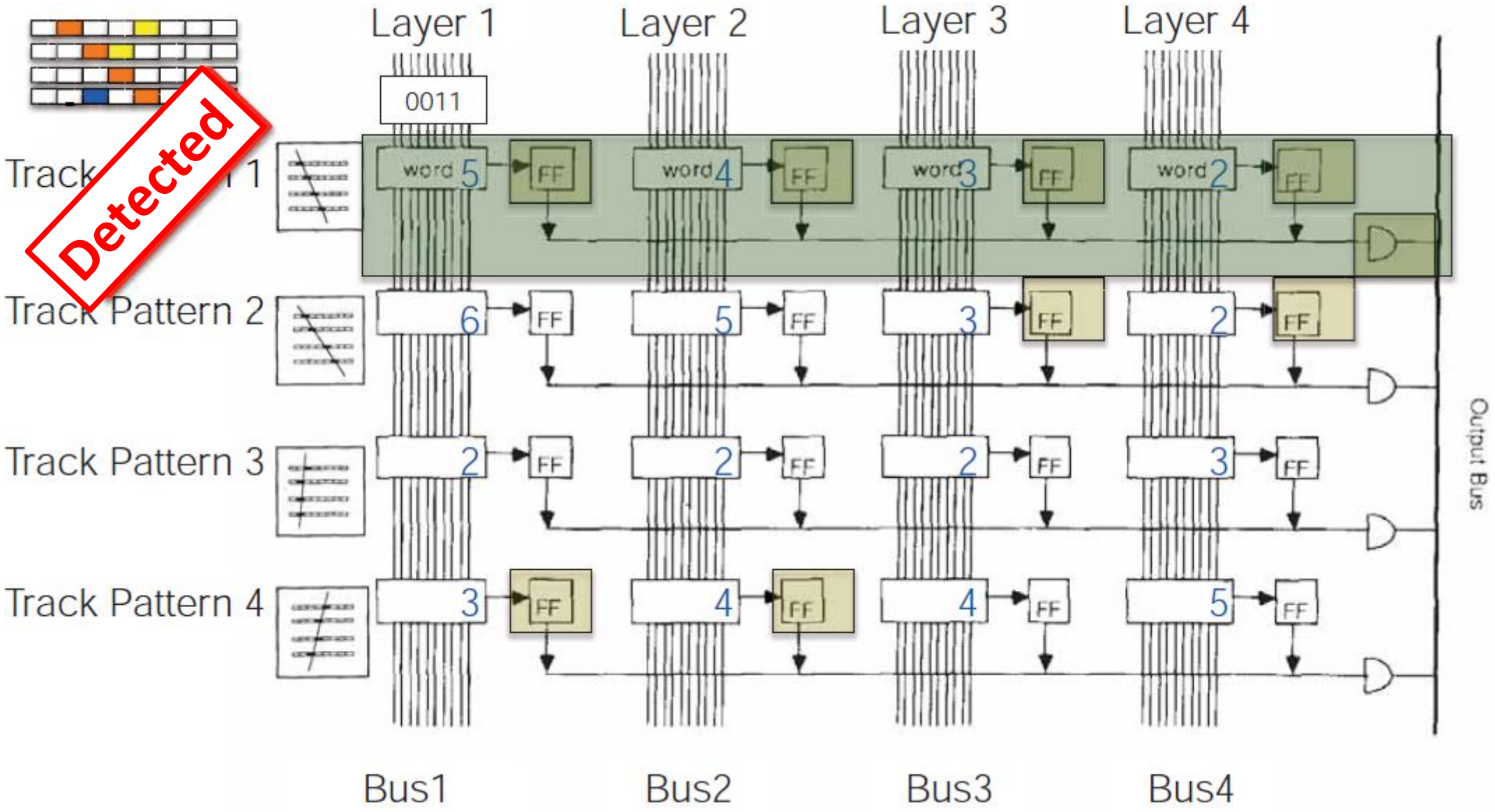
Pattern Matching



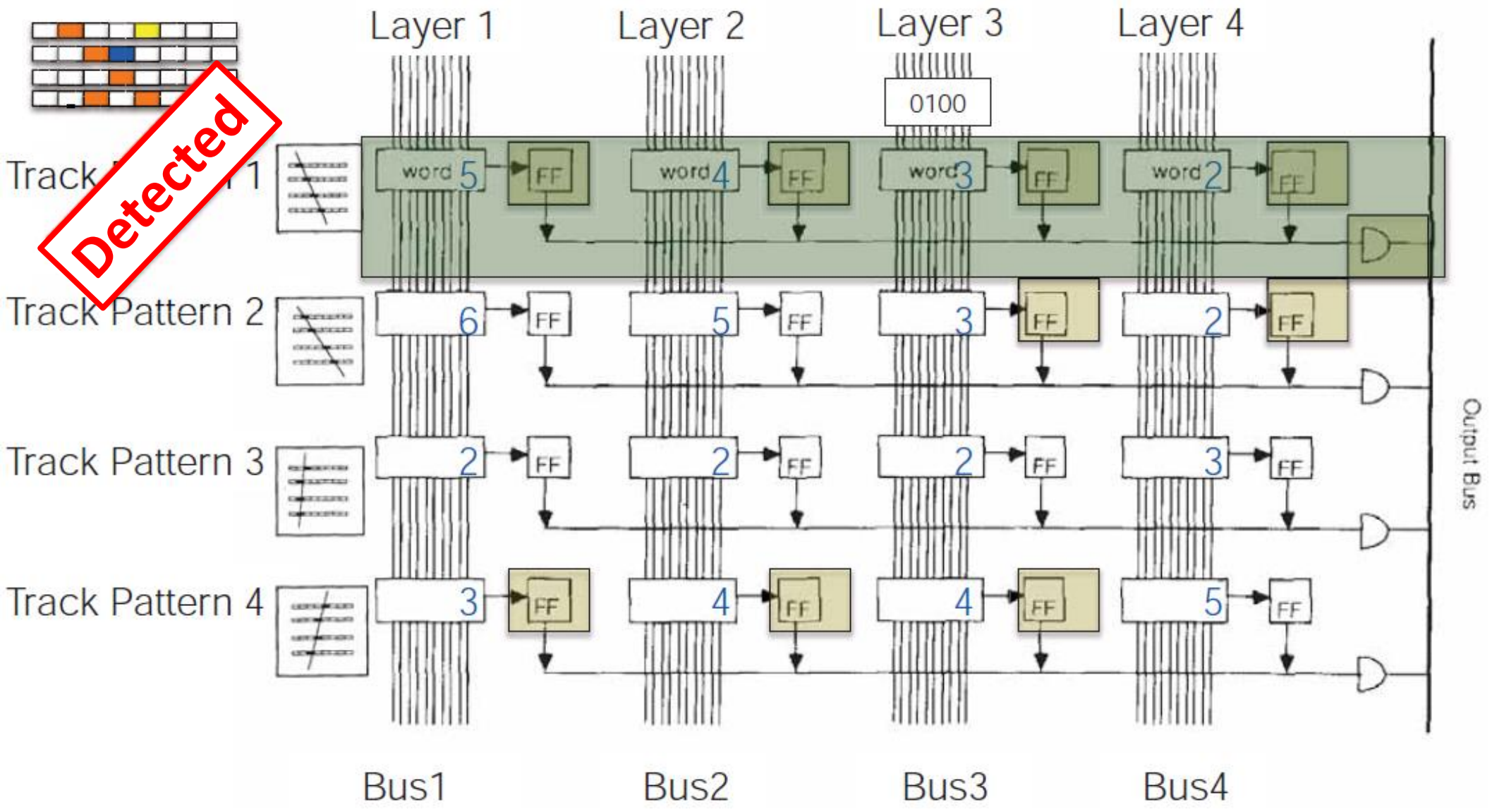
Pattern Matching



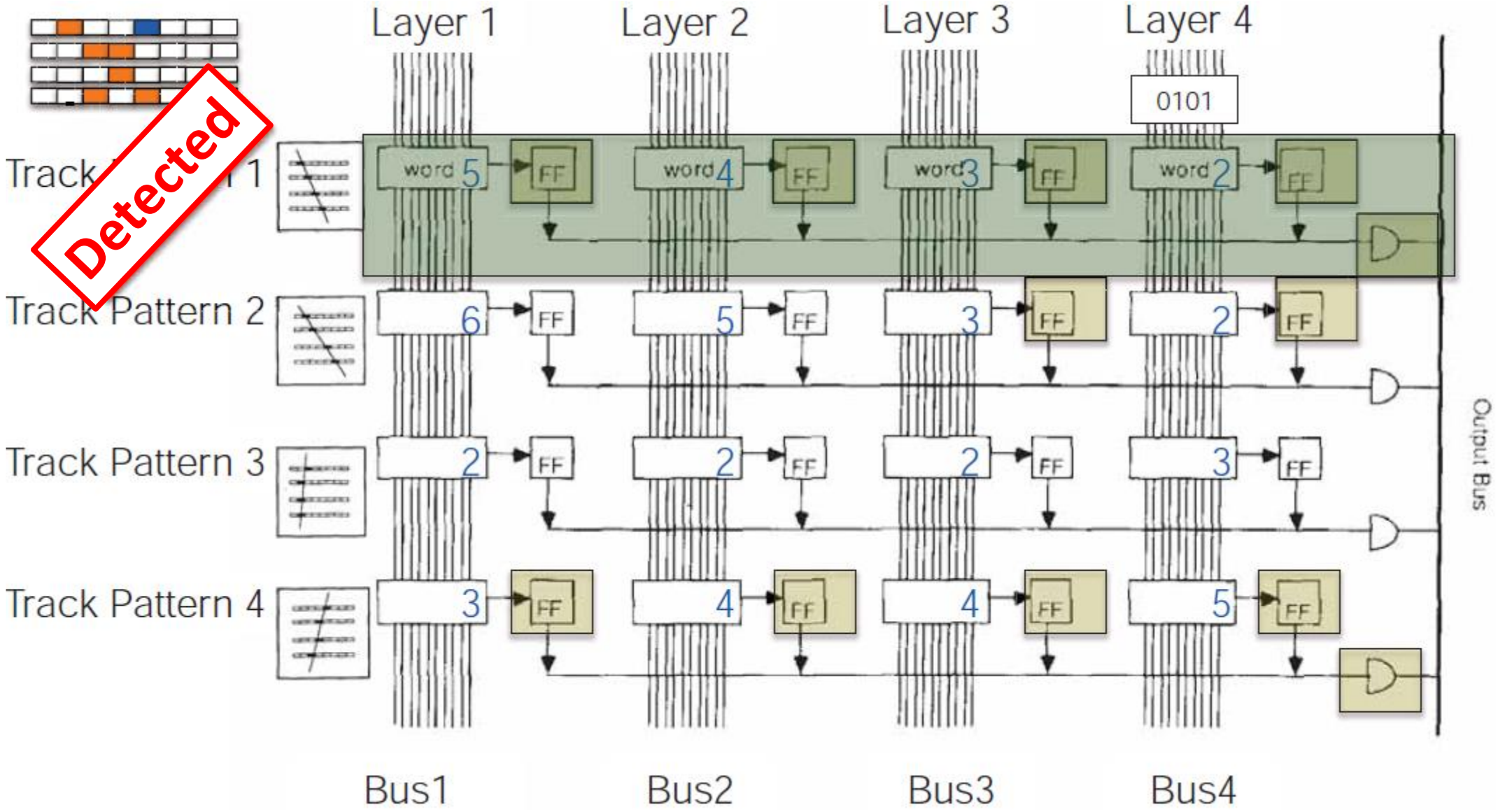
Pattern Matching



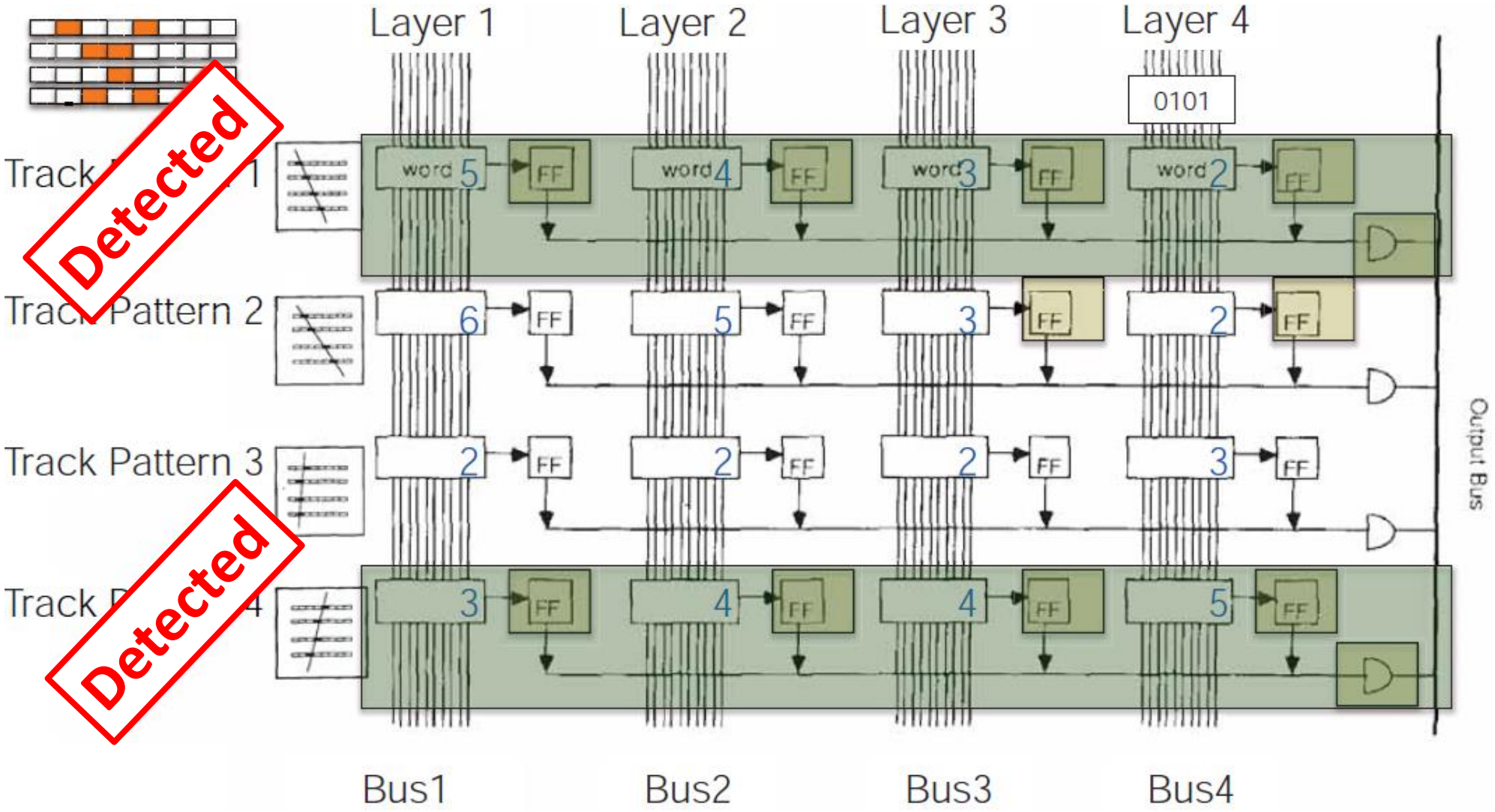
Pattern Matching



Pattern Matching



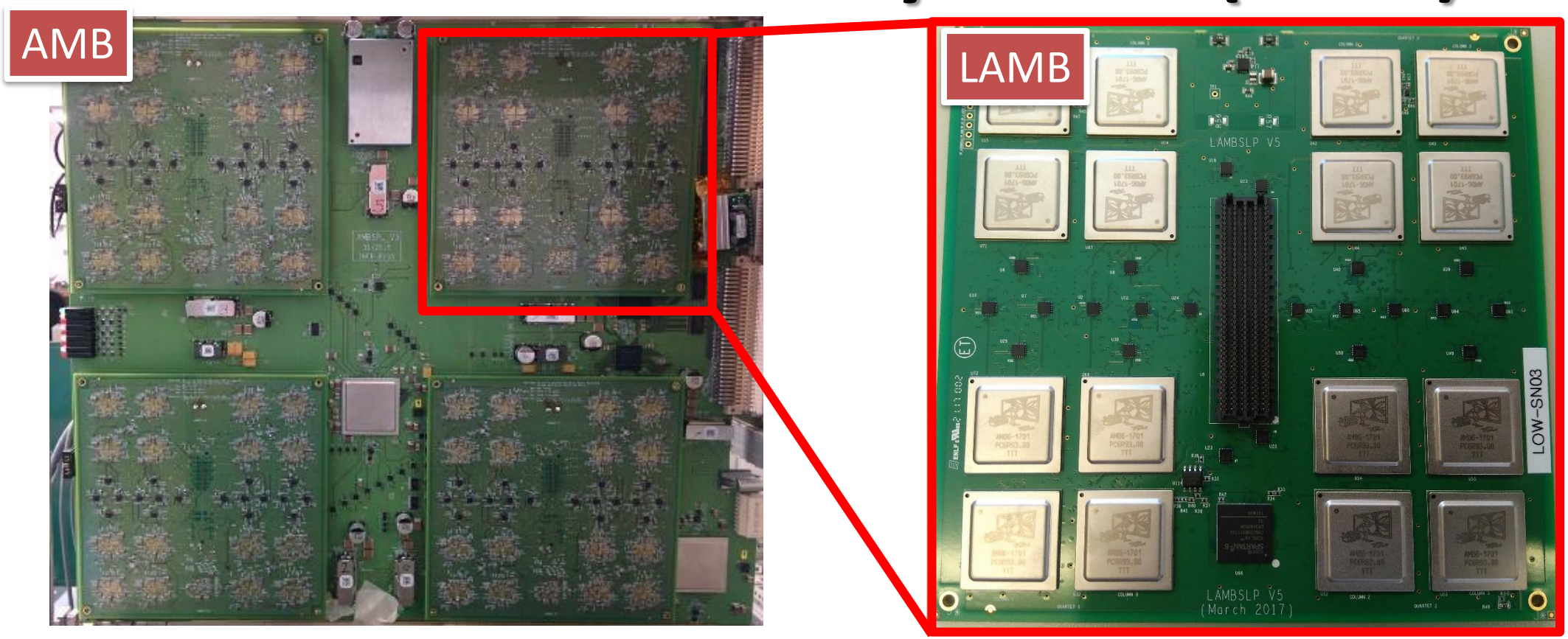
Pattern Matching



All patterns are checked in a single clock cycle.

→ Pattern matching of the event finishes when all the hits are loaded.

Associative Memory Board (AMB)

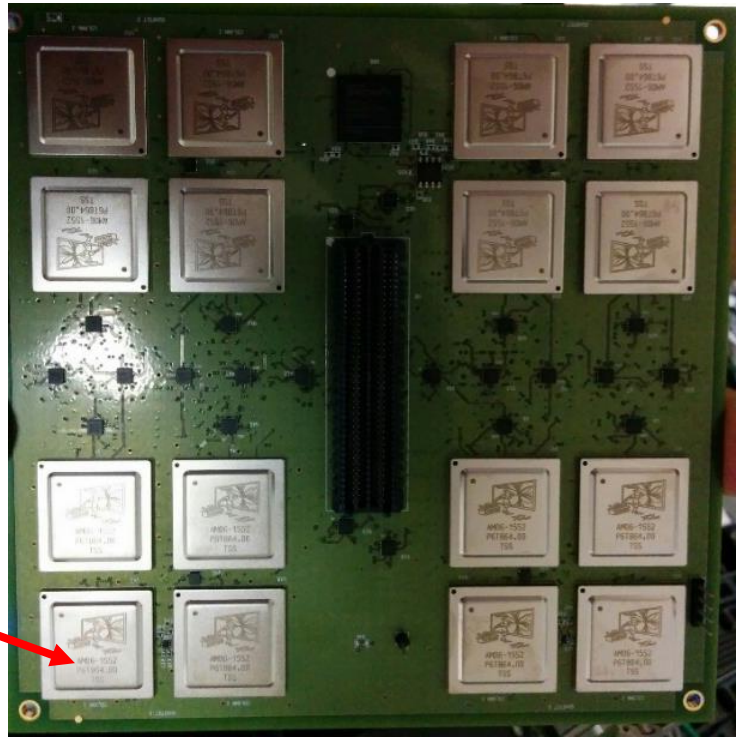


- ✓ AMB is 9U VME board which performs pattern matching to find track candidates. (A board has 4 FPGAs each has different function to control dataflow.)
- ✓ Each AMB has 4 Local AMBs (LAMBs).
- ✓ Each LAMB has 16 ASIC chips (AMchips) which store pre-calculated track patterns.
- ✓ 128 AMBs are used for FTK full system.

AM Chip Specification

	Technology	Area	#Patterns	Speed (MHz)	For
CDF SVT AM			128	30	CDF SVT (1992)
SVT upgrade (AM03)	180 nm	100 mm^2	5k	40	CDF SVT (2005)
AM06	65 nm	160 mm^2	128k	100	**FTK**
AM09	28 nm		3 x 128k	250	Next Generation Hardware Track Trigger (HTT)

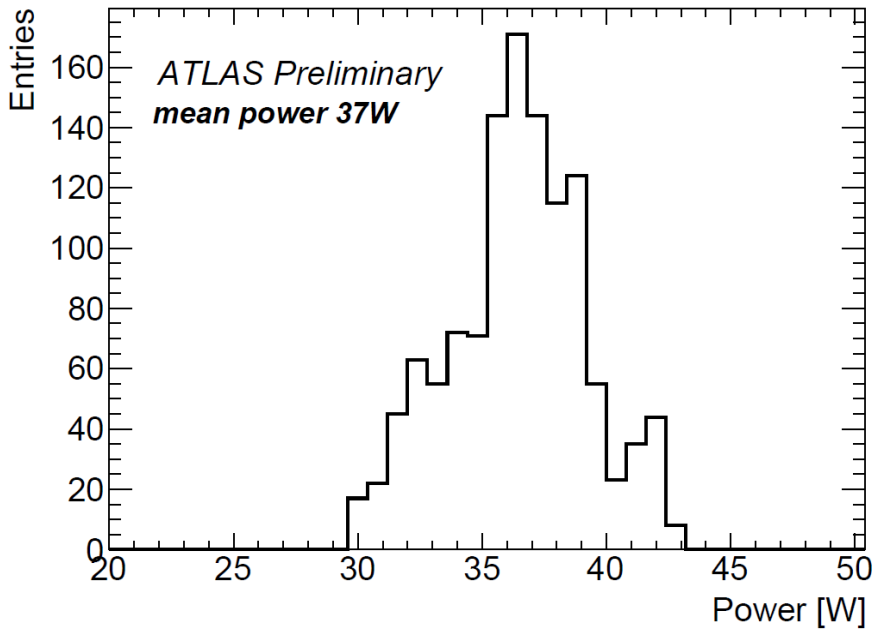
- ✓ Patterns are stored in special ASIC which are called “AM chips”.
- ✓ Significant improvement has been realized for recent ten years.
  - ~25 times more patterns can be stored.
  - Operating clock improved. (40 MHz → 100 MHz)
- ✓ FTK will use 8k chips in total.
  → 10^9 patterns are used for pattern matching to cover whole detector coverage.
- ✓ Development of the next generation AM chip ongoing.

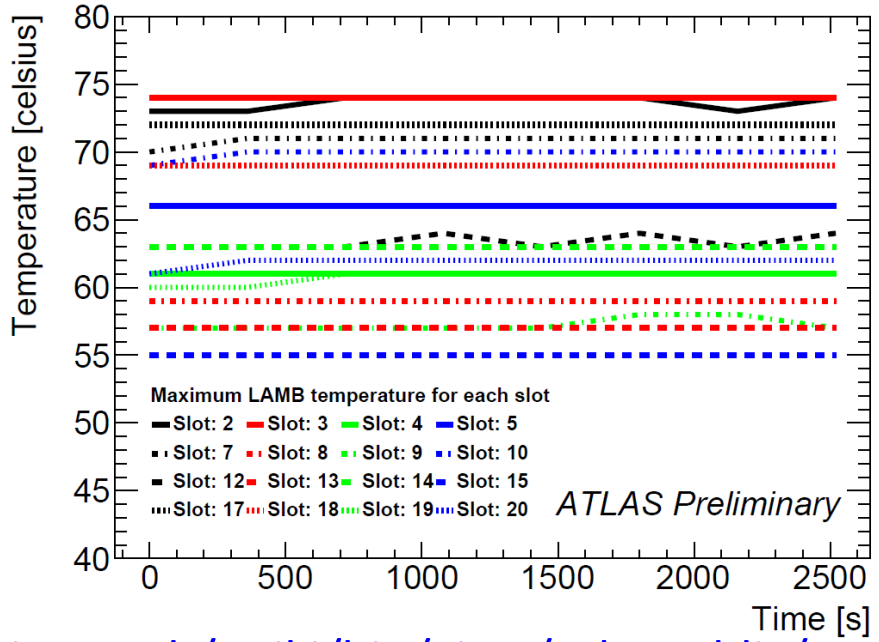


Production & Installation Status

- ✓ 64 AMBs + 256 LAMBs + a few spares are already produced and installed in ATLAS counting room underground.
  - Corresponds to half of the full system (which achieves a goal in 2018).
  - Possible to perform hardware test in realistic environment.
- ✓ The rest will be produced before the start of FTK operation in 2021.

Power Consumption and Temperature



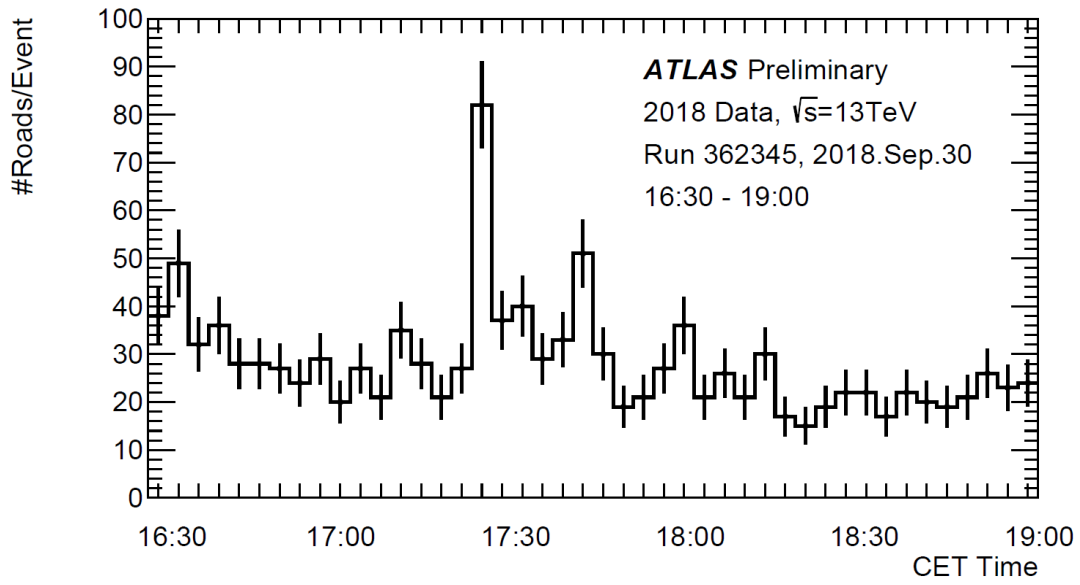


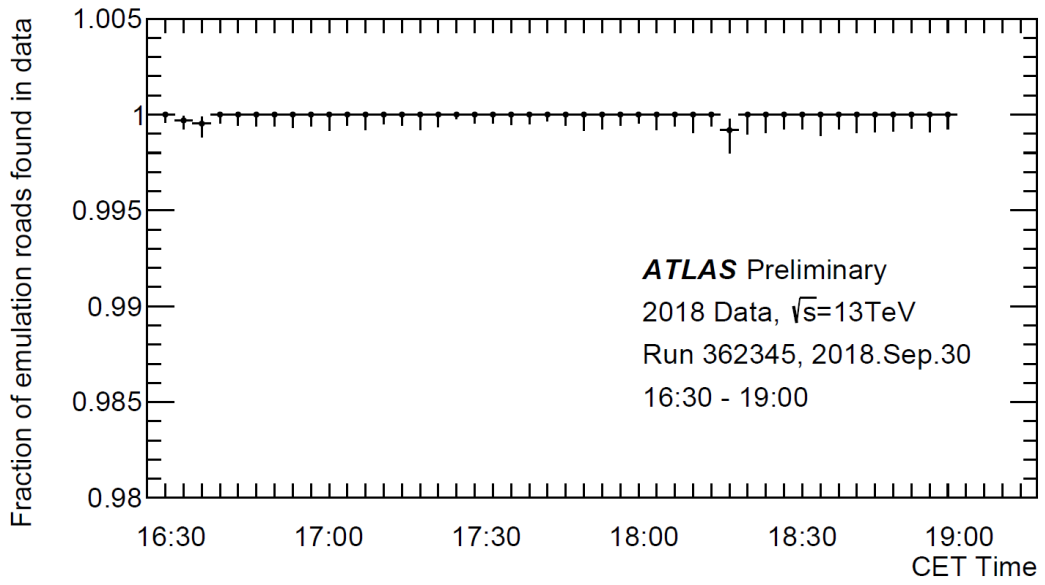
https://twiki.cern.ch/twiki/bin/view/AtlasPublic/FTKPublicResults

- ✓ AMB power consumption and board temperature depend on how heavy the input data is and how many AMBs are working in a single VME crate.
- ✓ Power consumption and temperature were measured with 16 AMBs populating a VME rack which is compatible with real operation in 2021, and with data of expected heaviness.
  - 37 W/LAMB → AMB power = 100 (AMB) + 4 x LAMB = **248 W**.
  - Temperature is kept below **80 degC** which is well below its limitation of 100 degC.

Power consumption and temperature are under control.

Dataflow Stability Test with Collisions





Example of improvements

https://twiki.cern.ch/twiki/bin/view/AtlasPublic/FTKPublicResults

- ✓ Firmware updates to avoid clock crossing the clock domain.
- ✓ Monitoring system being enriched in the software to perform smooth development.
- ✓ Avoid conflict between monitoring system and VME control by implementing “Semaphore”.

Current results of bit-level check

- ✓ Current level of AMB data quality with collision data is good. Error rate is $< 10^{-4}$.
- ✓ Same results are achieved with standalone processing on separate 32 AMBs.

Summary and Plan

- ✓ AMB performs pattern matching which is a key algorithm for fast tracking of the FTK.
- ✓ Half number of the AMBs/LAMBs produced and installed (achieved goal of 2018), the rest will be produced, tested and installed during LS2 for use in run3 starting from 2021.
- ✓ Power consumption and temperature of AMB and LAMBs are under control.
- ✓ Excellent dataflow and processing performance obtained even for real data.
- ✓ FTK will start operation in 2021 and run entire LHC run3 (~ end of 2023).

AM system shows excellent performance and being ready for the operation!
It is also important demonstration of the key technology at next generation Hardware Tracking for Trigger (HTT) at HL-LHC and also future fast tracking!

Backup

LHC and ATLAS Detector

➢ LHC

Large circular p-p collider at CERN, Geneva.
Its main parameters are

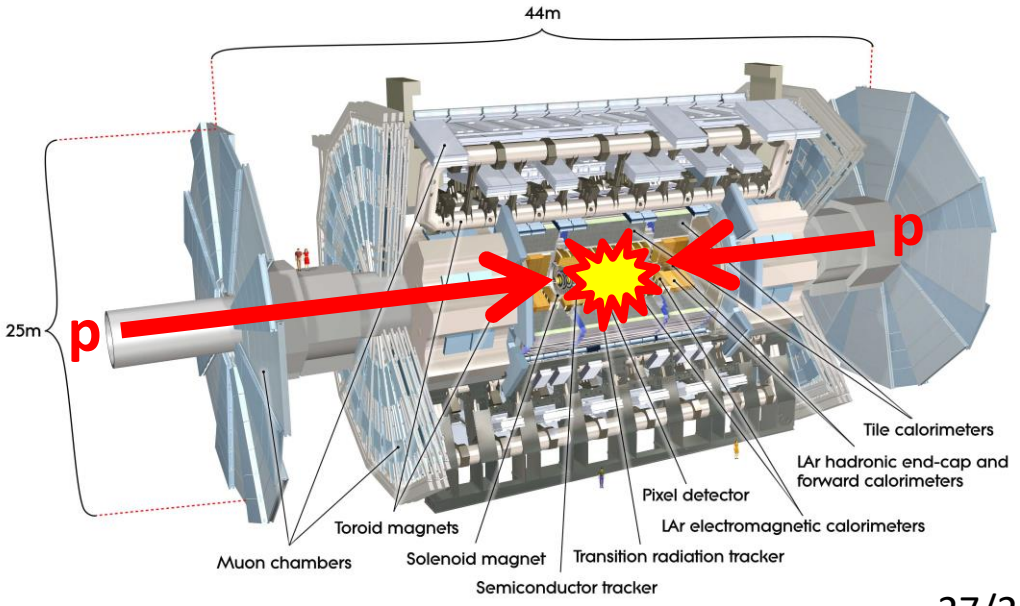
- ✓ √s = 13 TeV
- ✓ Max Instantaneous Luminosity = 2.14 x 10^{34} $cm^{-2}s^{-1}$
- ✓ 62.2 fb^{-1} is recorded in 2018, and 149 fb^{-1} during run2 total (2015 - 2018).

➢ ATLAS Detector

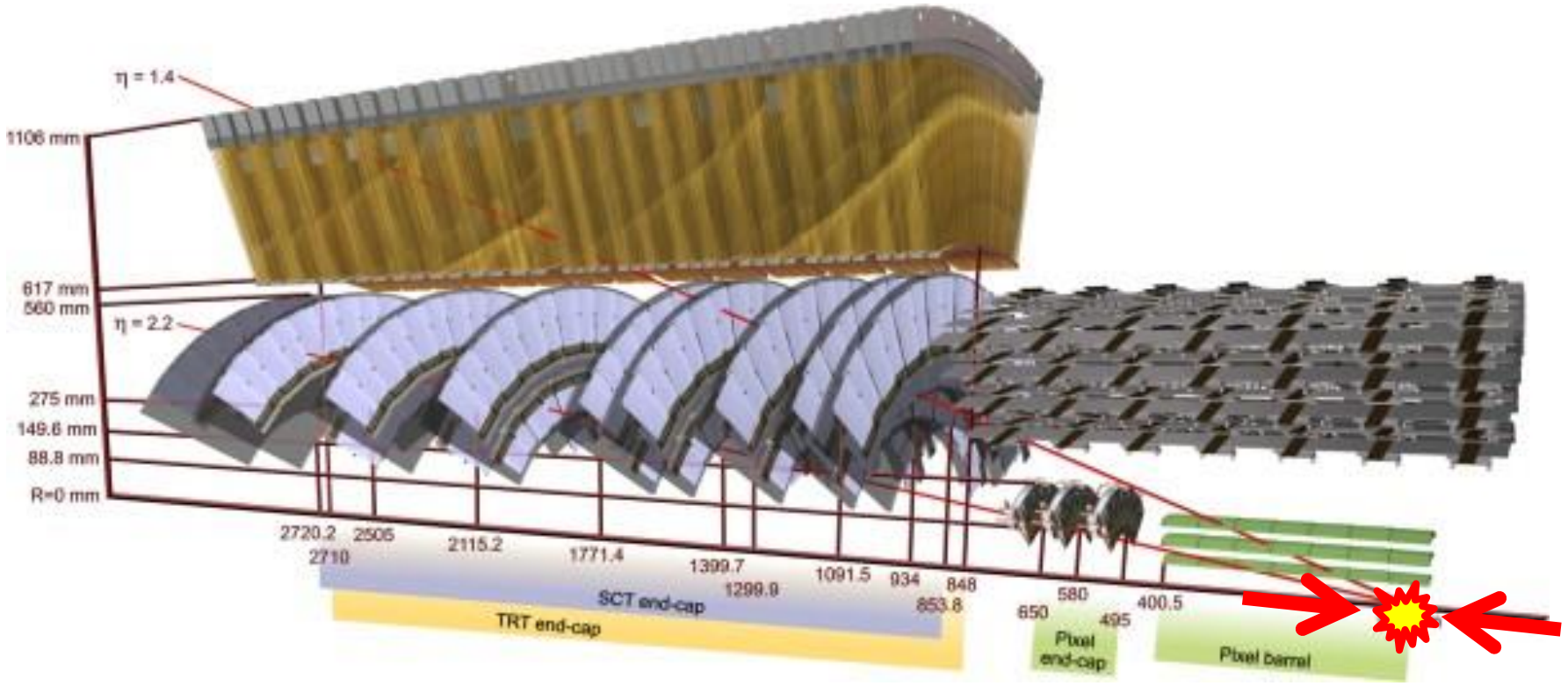
General purpose detector placed in one of the interaction points. It consists of

- ✓ Tracking Detector (IBL/PIX/SCT, in total 12 layers)
- ✓ Calorimeter
- ✓ Muon Detector

It measures position, momentum, energy, and types of the particles.



FTK input: Inner Detector



- ✓ Placed most inner part, measure position, momentum of charged particles.
- ✓ **The 12 layers of IBL/Pixel/SCT, in total ~100 million channels,** are the input of FTK.

	#layers	Readout	size/unit	#channels
IBL	1	2-dim	50μm x 250μm	12M
Pixel	3	2-dim	50μm x 400μm	80.4M
SCT	8	1-dim	80μm	6.3M

Key Algorithm: Track Fitting

Linear approximation are performed to reconstruct track parameters with using pre-calculated constants. Here full resolution hits are used.

☐ Linear approximation

(Toy detector figure assuming 4 layers.
In actually, the output is calculated with 12 layers).

Track parameters are calculated immediately hits are coming.

Constants

Parameter

$$\tilde{p}_i = \sum_{l=1}^{N} C_{il} x_l + q_i$$

$\tilde{p}_i$: Observed Track Parameter (i=0~4)

x_i : Hit Coordinate $\quad \vec{C}_i \; q_i$: Constants

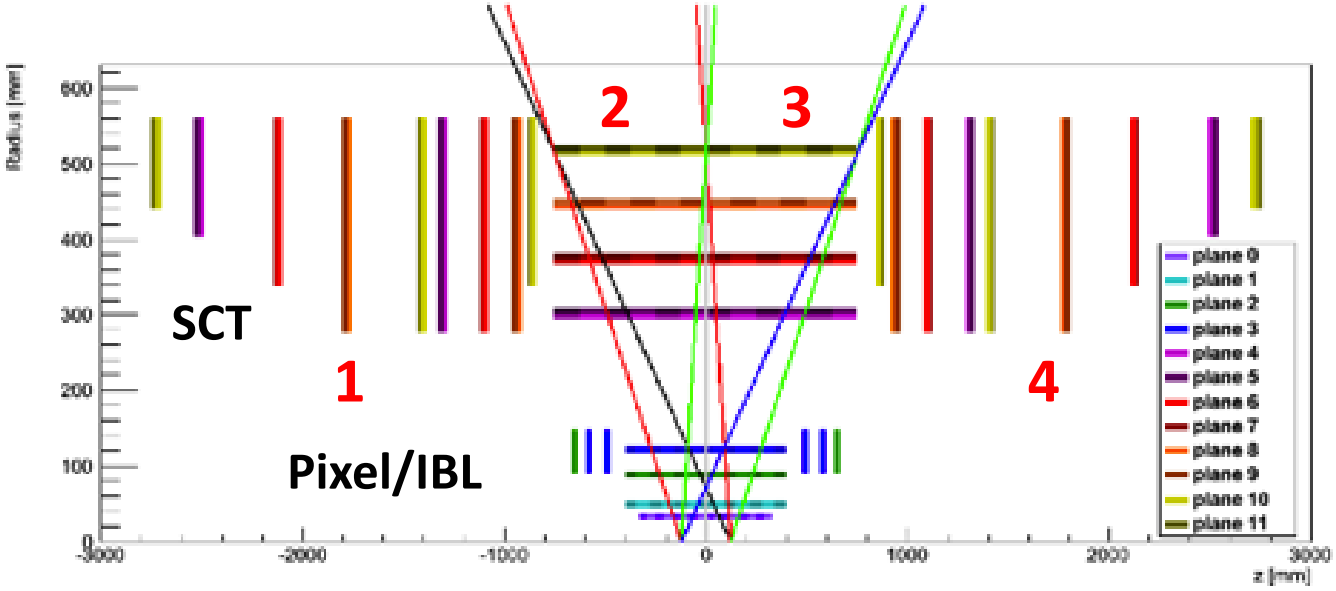
5 helix parameters
d0, cotθ, ϕ, c, z0

FTK patterns and constants are determined from simulation.

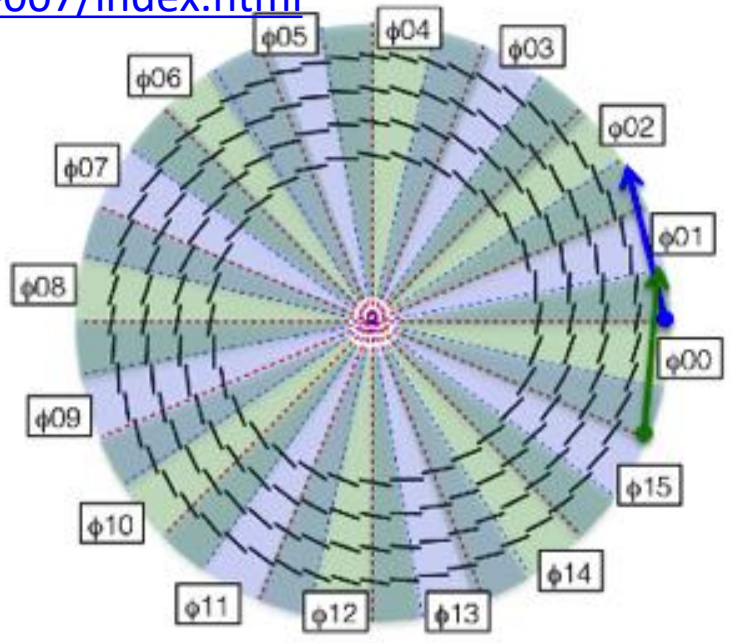
Parallel Processing

- ✓ Pattern matching and track fitting will run in parallel for further fast processing.
- ✓ Parallelization are performed based on detector region.
  - → 4 x η, 16 x ϕ, in total 64 regions.

https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/UPGRADE/CERN-LHCC-2013-007/index.html



4 x η



16 x ϕ

Each region has overlap to avoid inefficiency at tower boundaries due to

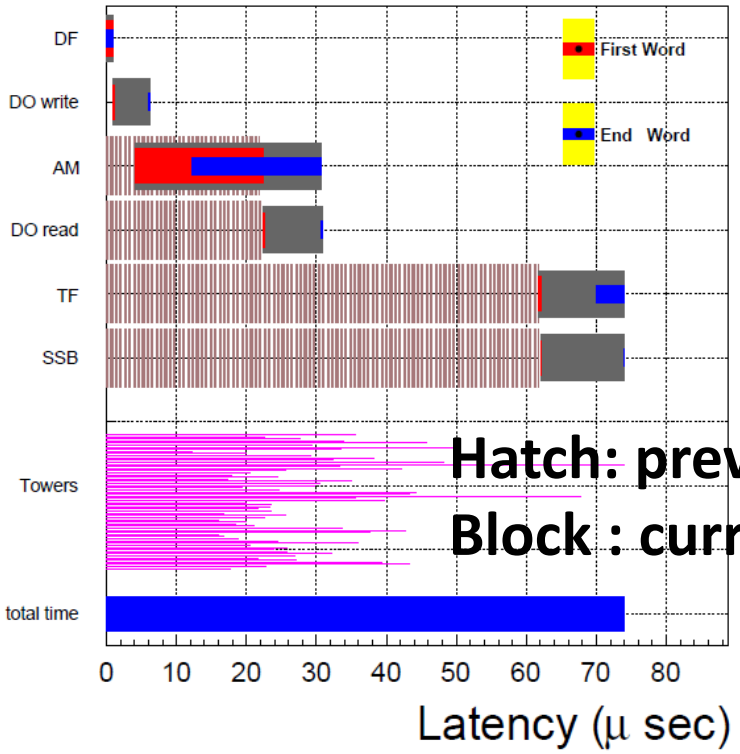
- ✓ The finite size of the beam luminous region in the z coordinate.
- ✓ The finite curvature of charged particles in the magnetic field.

Simulation of Processing Time

Since FTK runs in the trigger system, processing time is one of the important factors to be validated and optimized.

MC: Z → μμ, 69PU (L = $3x10^{34}$ $cm^{-2}s^{-1}$)
Input rate: 100 kHz

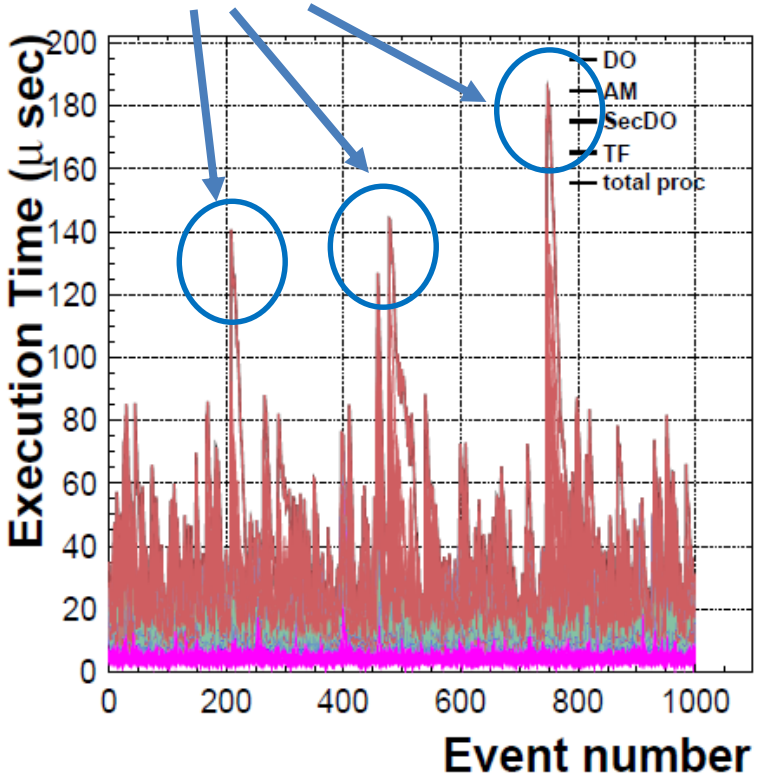
https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/UPGRADE/CERN-LHCC-2013-007/index.html



Hatch: previous event
Block : current event

Processing time for 1 event

Some events takes much time,
but it is soon recovered.



operate in time!

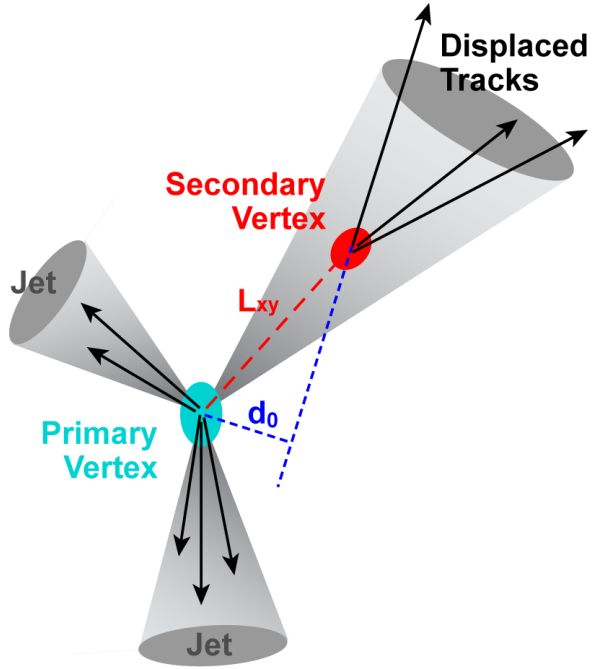
Processing times for 1000 events

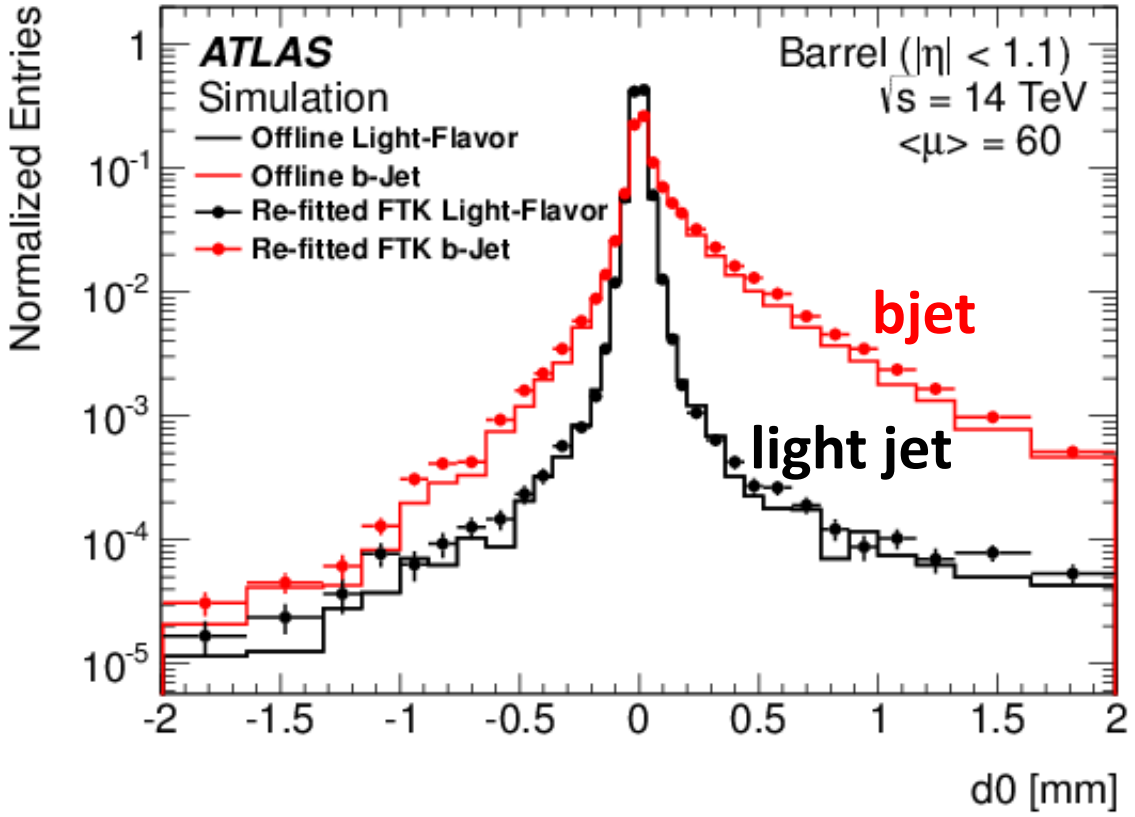
What is the benefit to introduce FTK ?

- ✓ HLT can use FTK tracks directly or quickly refit them. (→ HLT can have an extra time.)
- ✓ Better identification of the tracking objects such as b-jet, tau, track MET at the trigger level.
- ✓ It is also possible to reconstruct all the vertices in an event.

https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/UPGRADE/CERN-LHCC-2013-007/index.html

Ex.) Transverse Impact Parameter





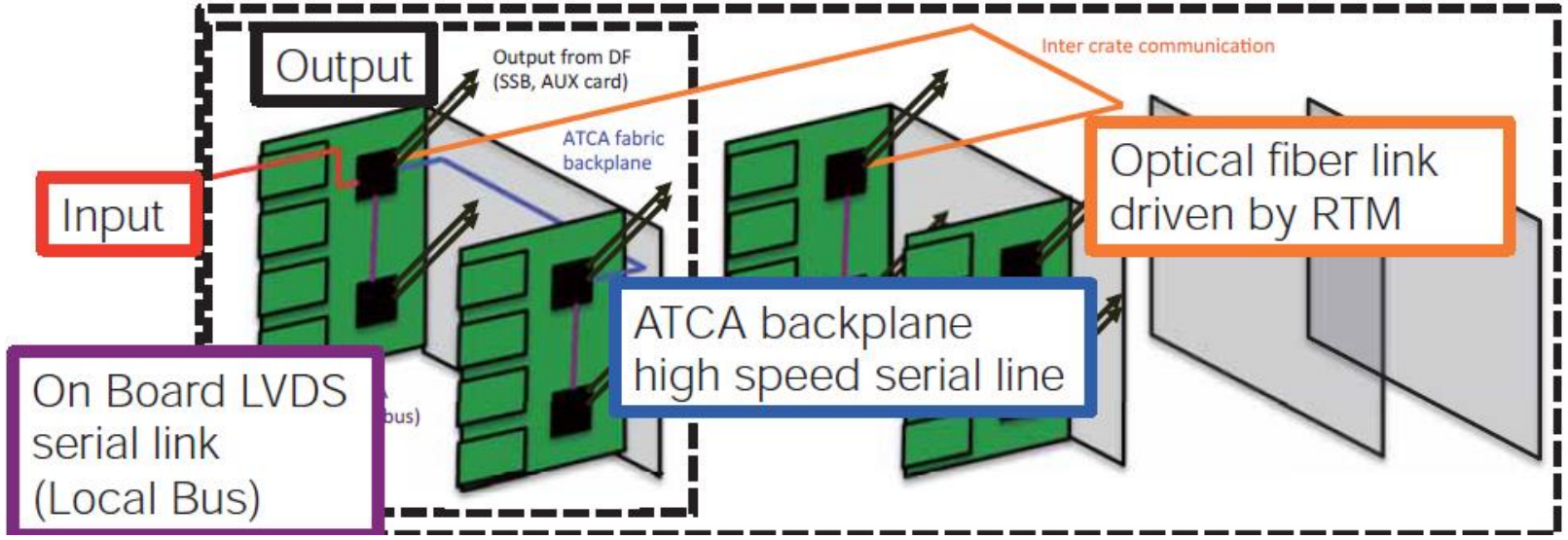
B-hadron flies O(mm) in the detector.
→ Large impact parameter

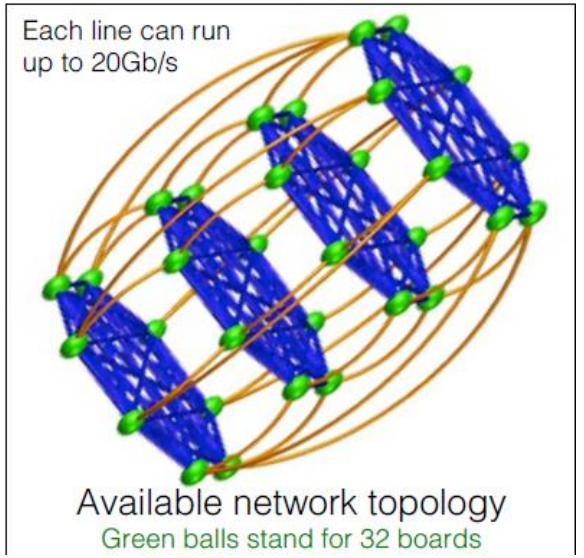
Similar quality w.r.t. offline analysis!

Data Sharing by ATCA

- ✓ Data Sharing for parallelization are done by Advanced Telecommunication Computing Architecture (ATCA) which can perform serial board communication on backplane.
- ✓ 1 crate holds 8 boards.
- ✓ 1 board process 2 regions.
  → 32 boards, 4 crates are used to process 64 regions.
- ✓ Data sharing between crate is done by fibers.

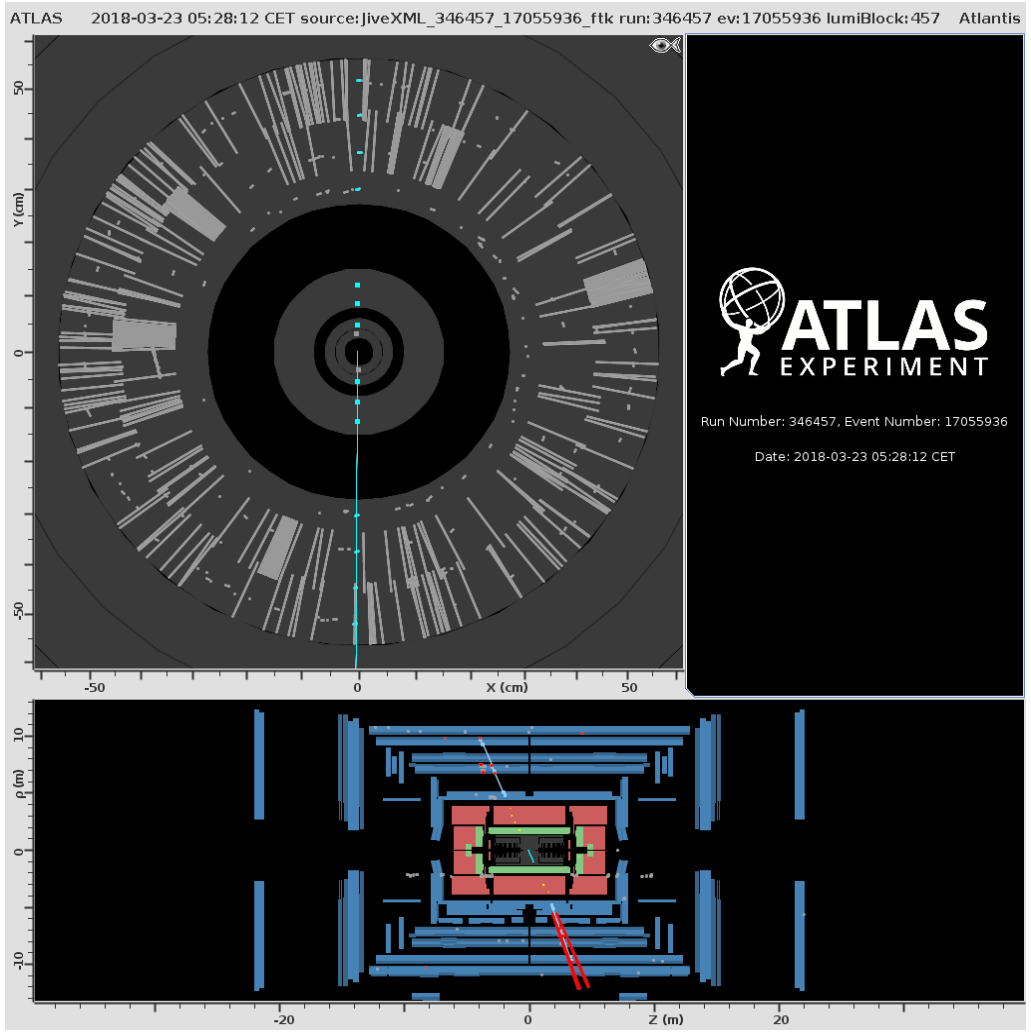
https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/UPGRADE/CERN-LHCC-2013-007/index.html





First FTK Real Track Output

LHC run number : 346457 (2018. Mar. 22)



https://twiki.cern.ch/twiki/bin/view/AtlasPublic/FTKPublicResults