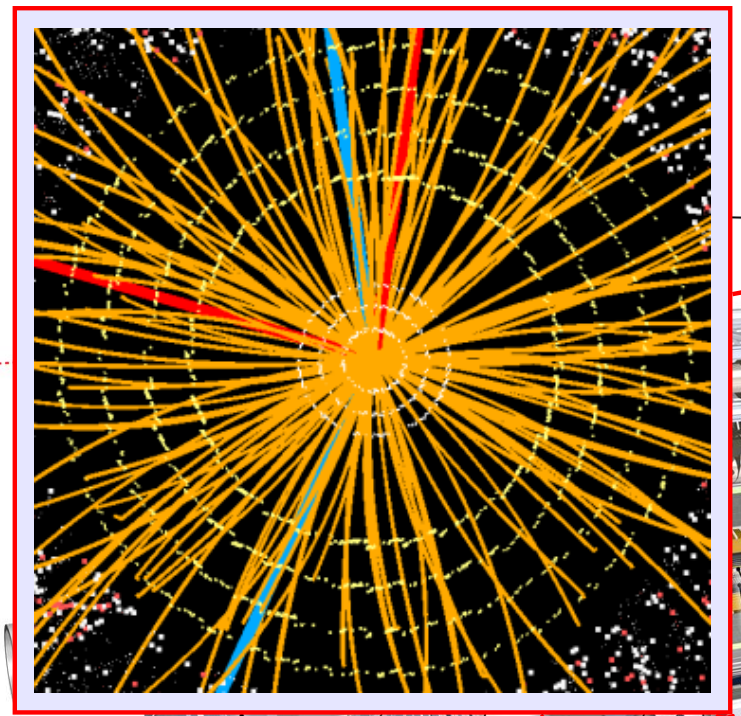
Development of Associative Memory ASIC in 28nm technology

Francesco Crescioli

LPNHE
PARIS

Journées VLSI - FPGA - PCB et Outils CAO de l'IN2P3
16-5-2018

Tracking



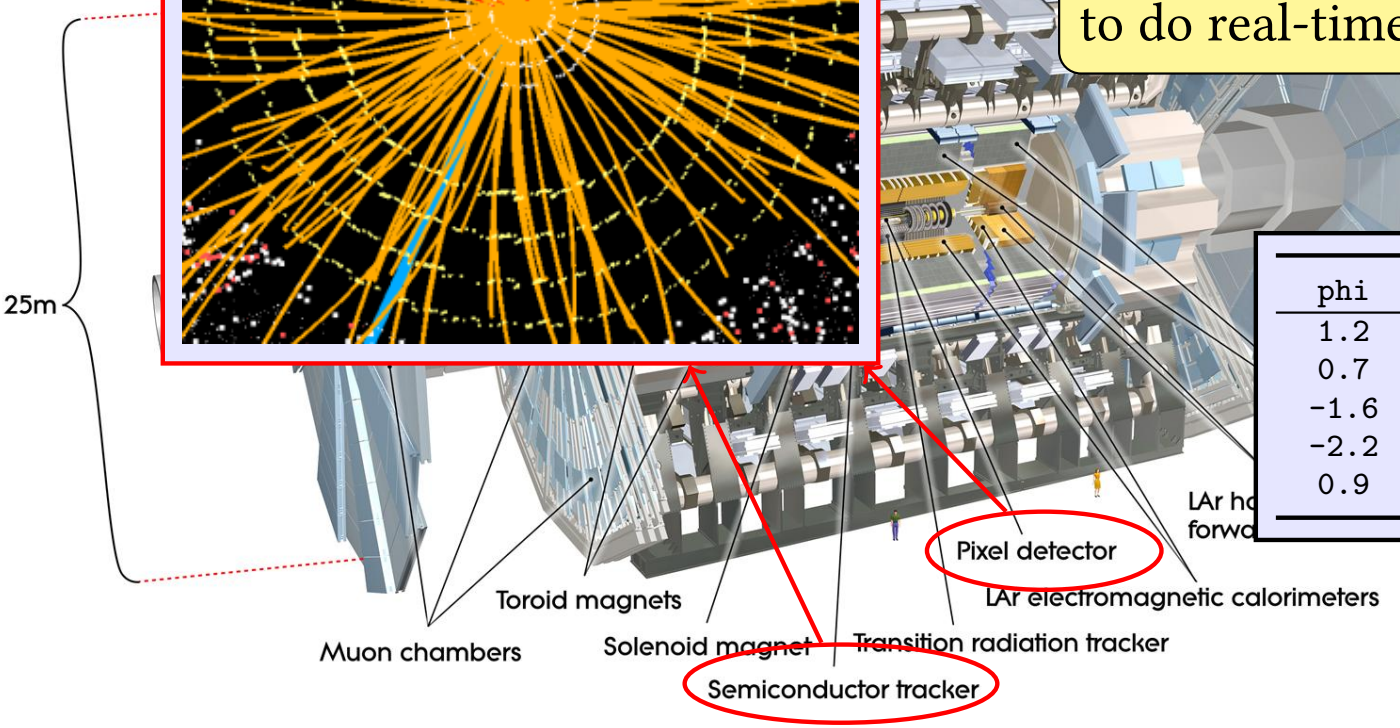
Hardware Tracking

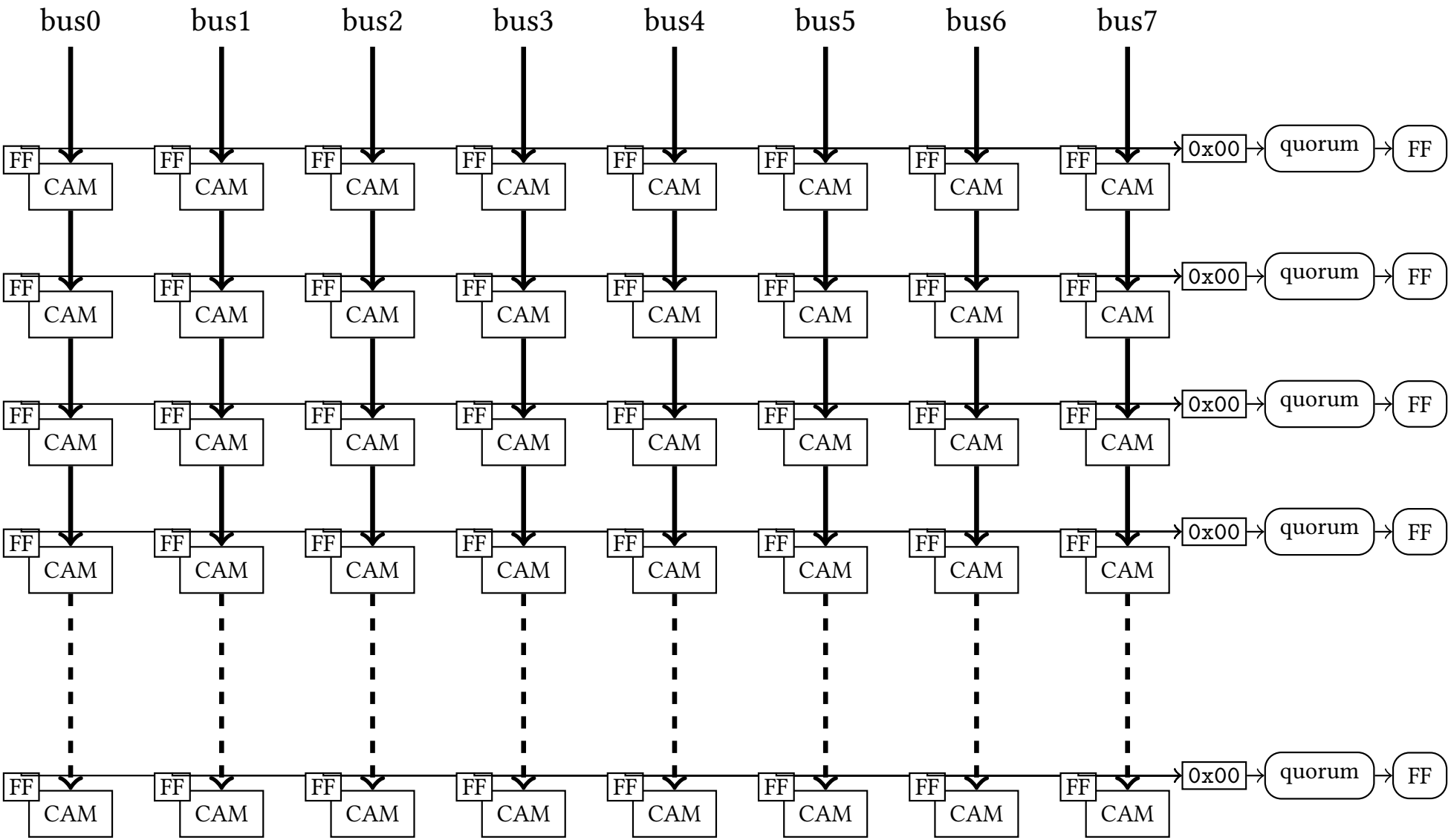
Data from the silicon detector
is analyzed in real-time.
The **Associative Memory**
is a powerful computing tool
to do real-time pattern recognition

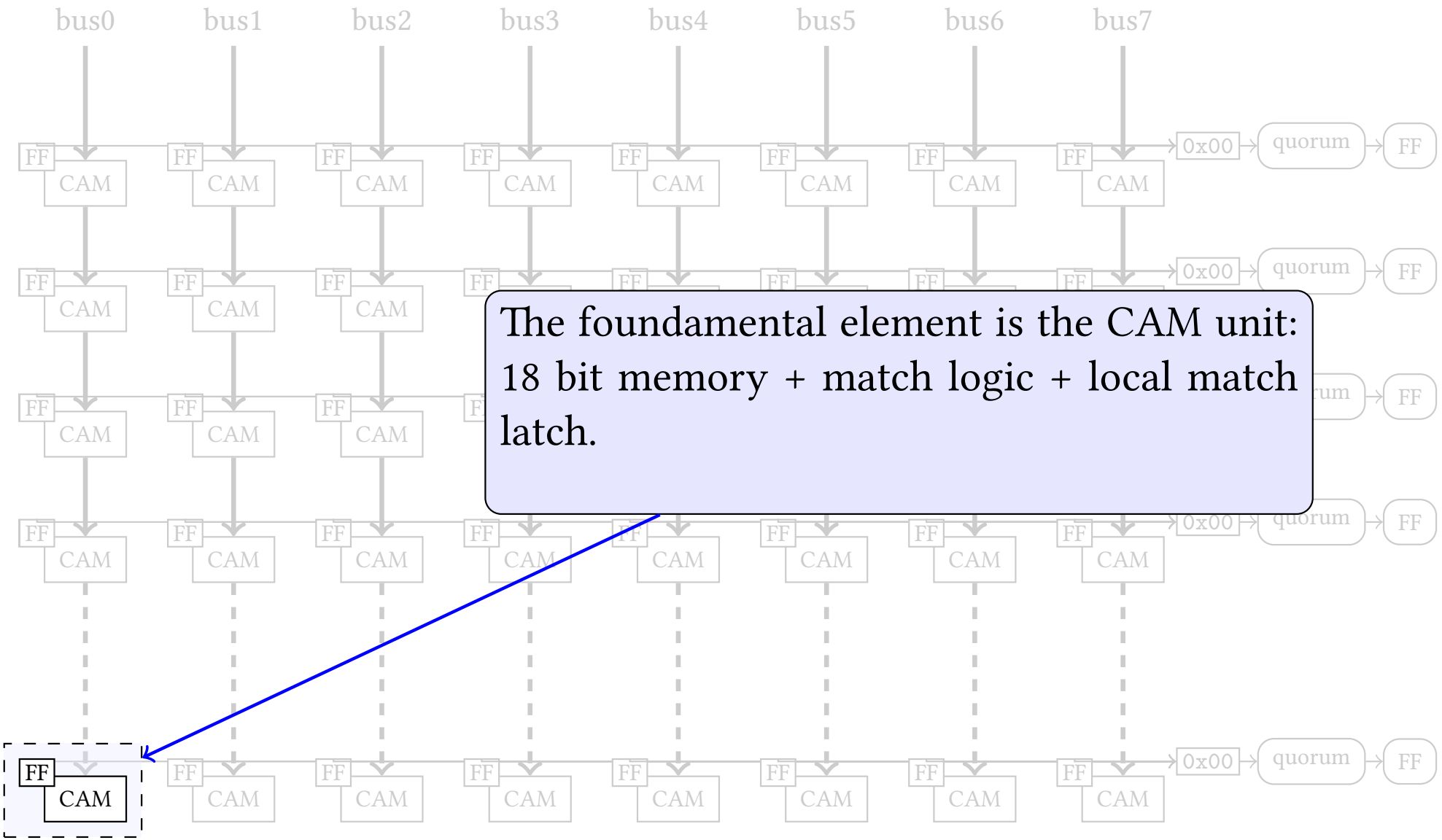
hits

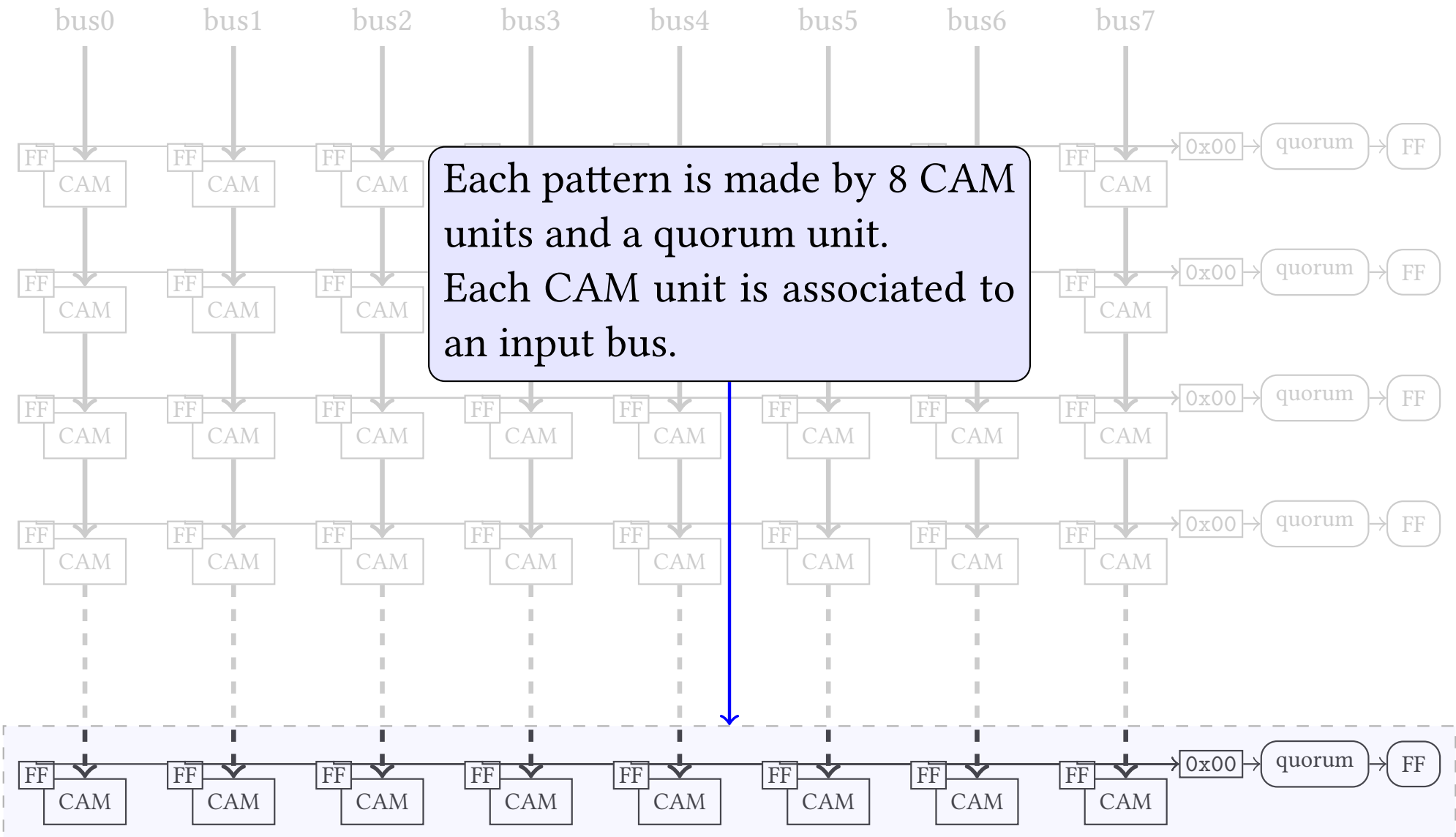
tracks

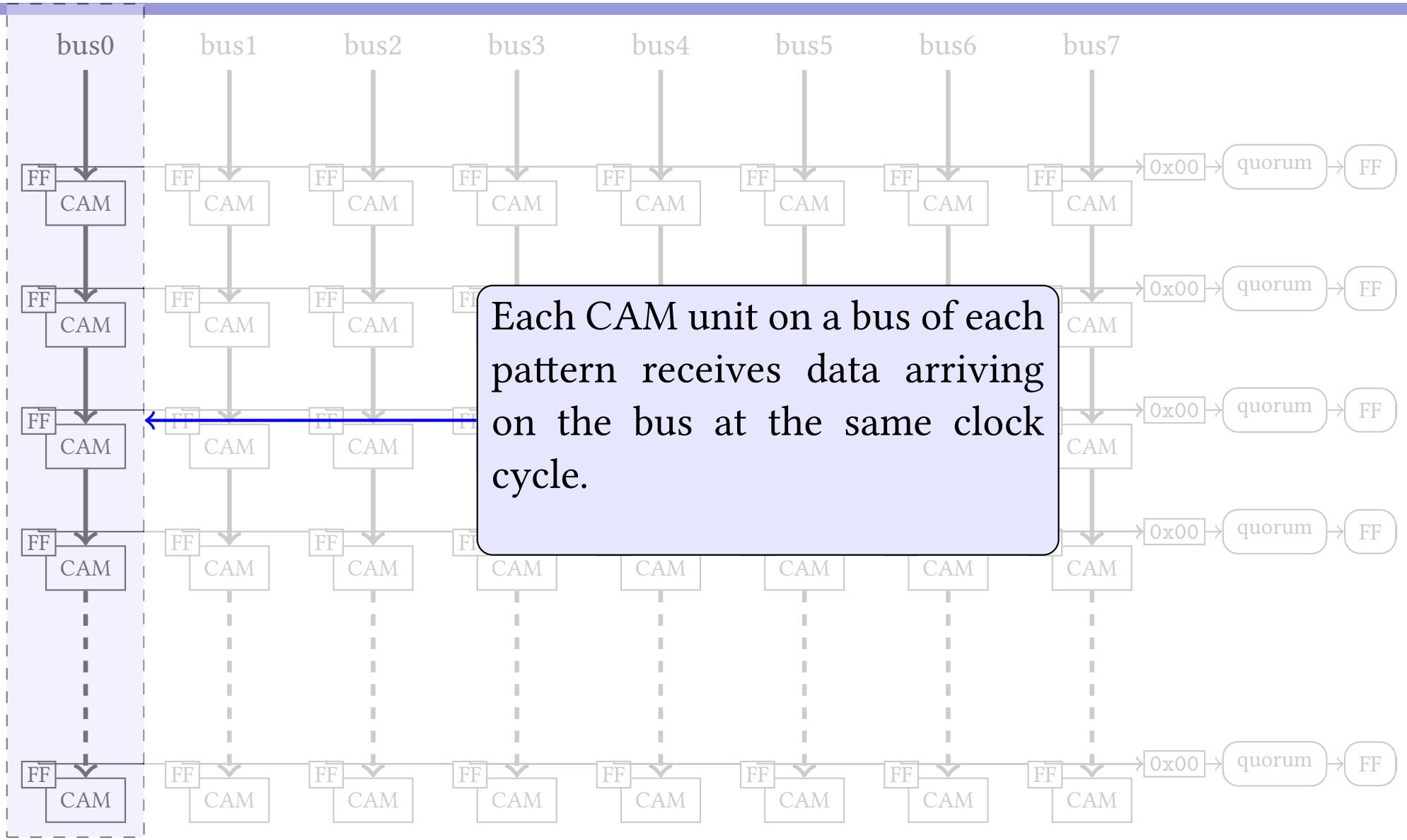
phi	η	z	p_T	d
1.2	0.2	1.3	10.5	0.03
0.7	1.1	1.1	4.3	0.14
-1.6	-1.8	1.2	6.6	0.32
-2.2	-0.5	1.3	12.2	0.28
0.9	0.9	1.2	1.9	0.83

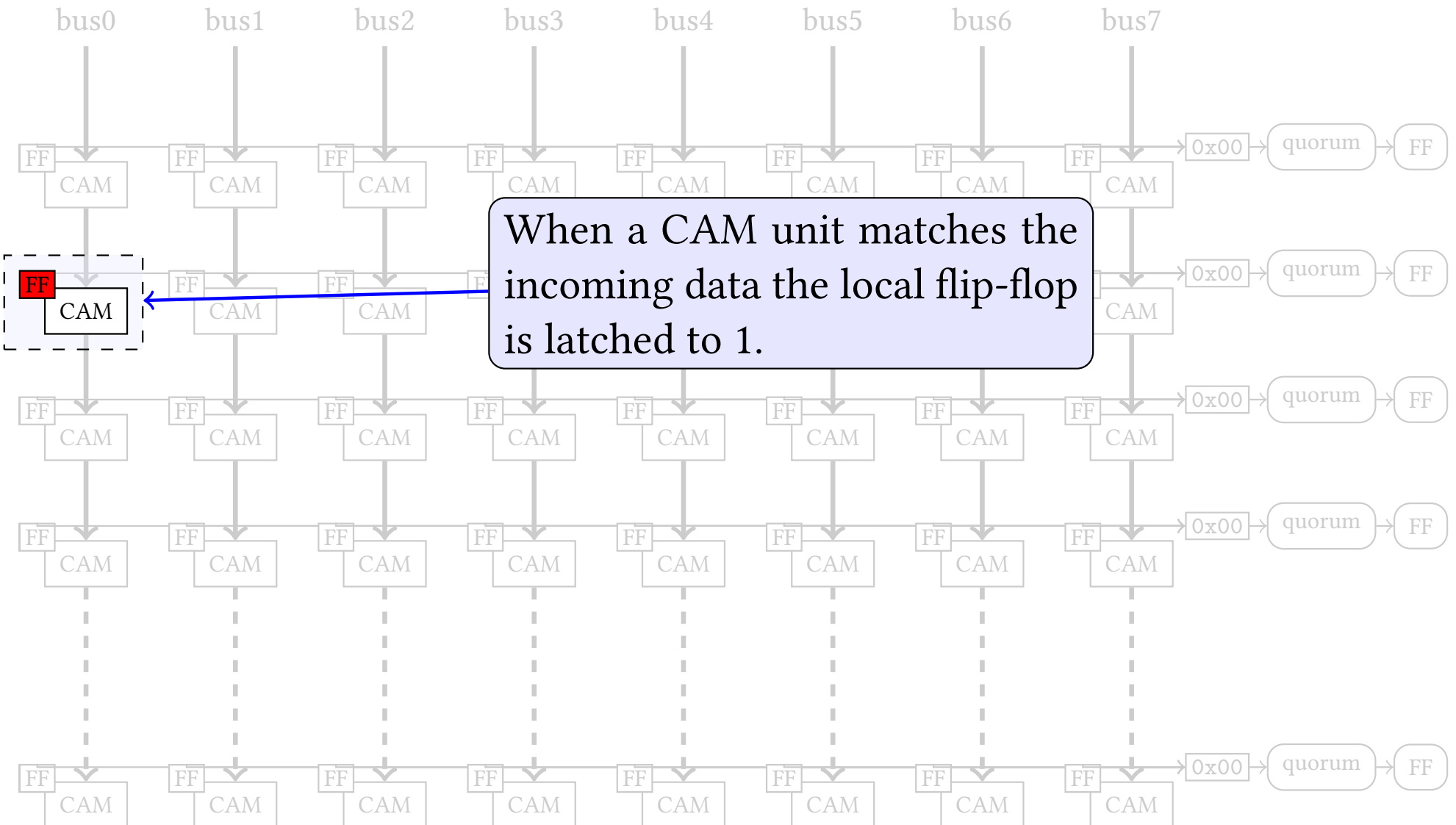


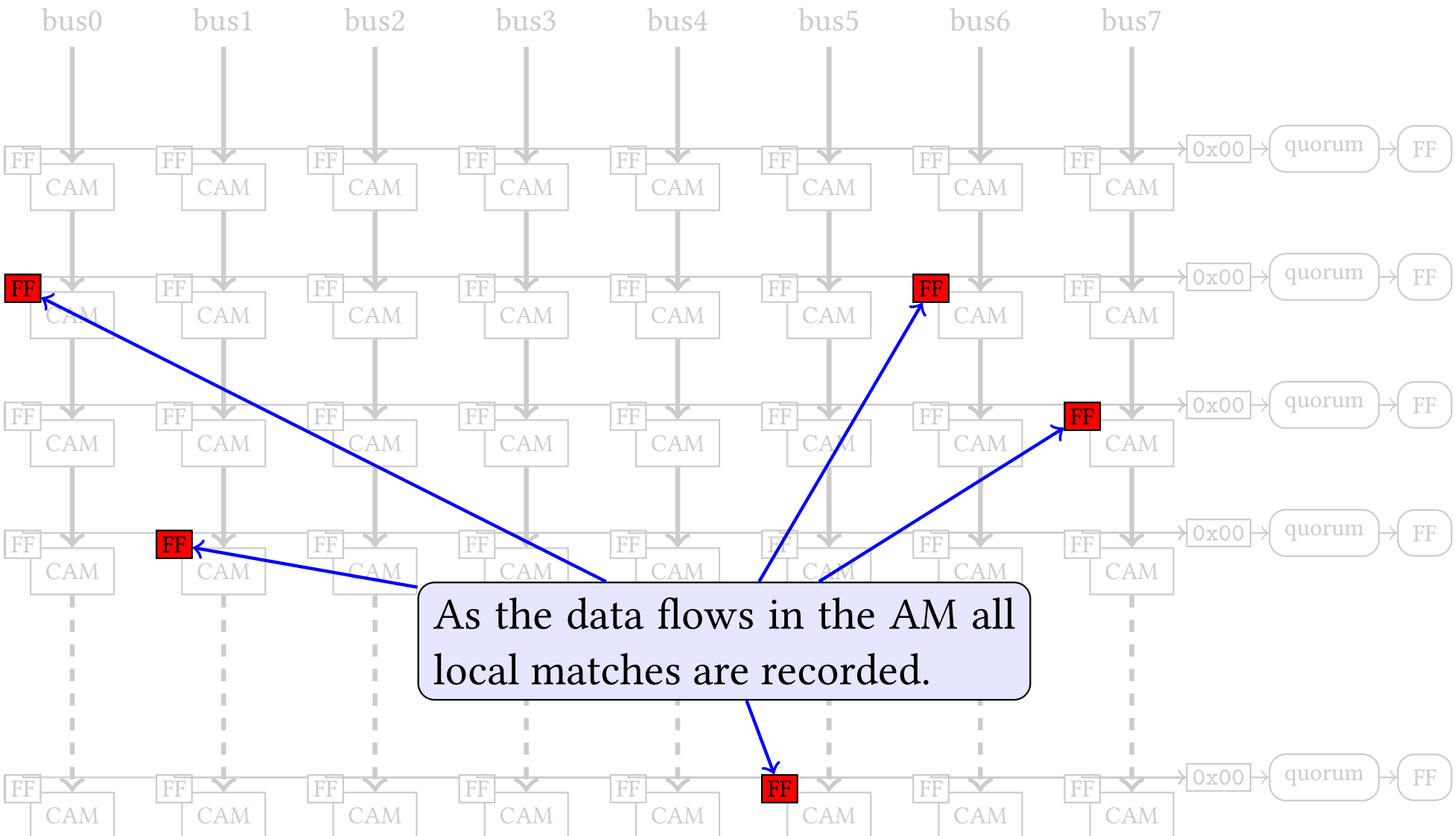

bus0
bus1
bus2
bus3
bus4
bus5
bus6
bus7
FF
CAM
0x00
quorum
FF

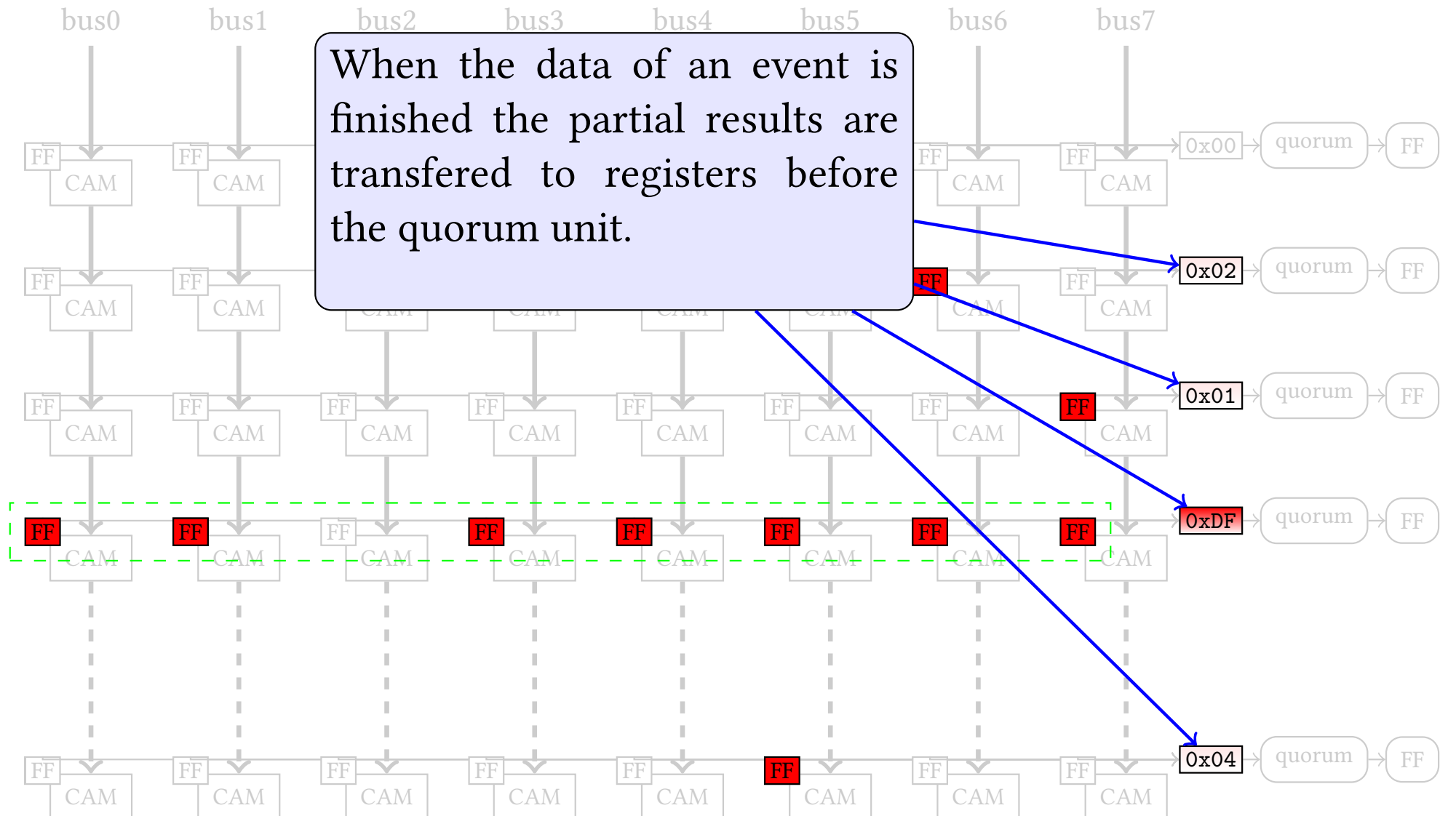

bus0
bus1
bus2
bus3
bus4
bus5
bus6
bus7
FF
CAM
0x00
quorum
FF
The foundamental element is the CAM unit: 18 bit memory + match logic + local match latch.

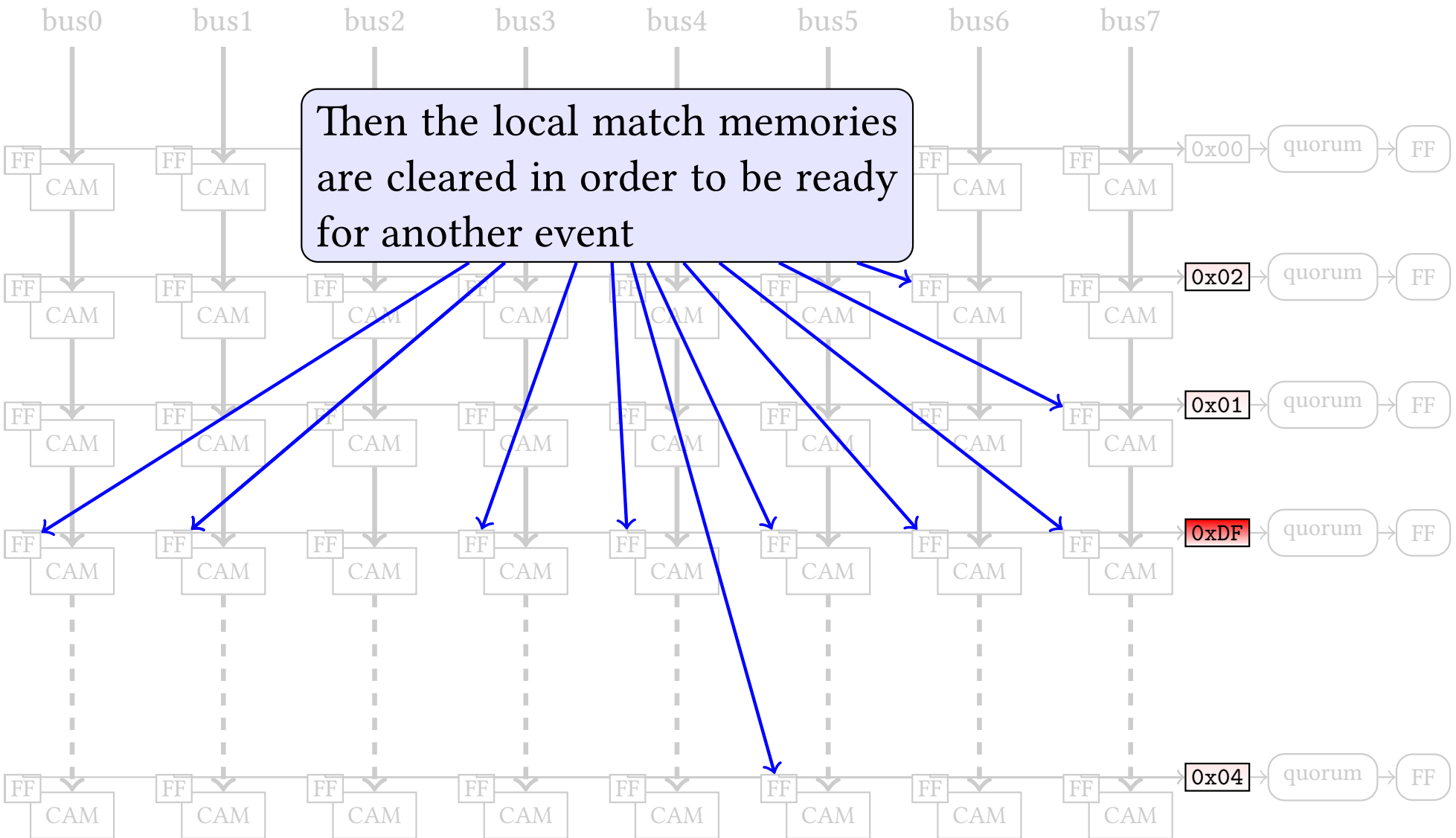

bus0
bus1
bus2
bus3
bus4
bus5
bus6
bus7
FF
CAM
0x00
quorum
FF
Each pattern is made by 8 CAM units and a quorum unit.
Each CAM unit is associated to an input bus.

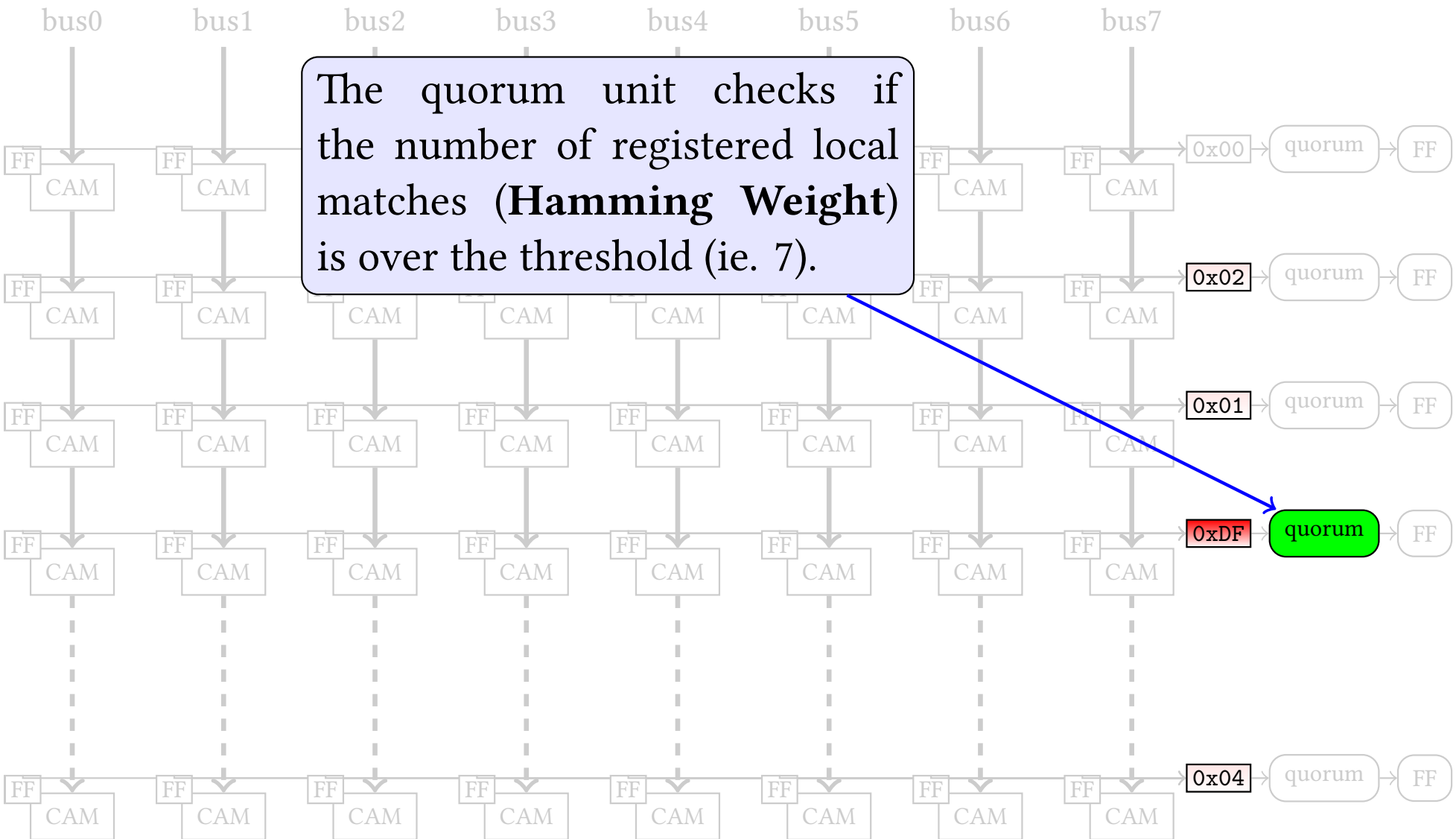

bus0
bus1
bus2
bus3
bus4
bus5
bus6
bus7
FF
CAM
0x00
quorum
FF
Each CAM unit on a bus of each pattern receives data arriving on the bus at the same clock cycle.


bus0
bus1
bus2
bus3
bus4
bus5
bus6
bus7
FF
CAM
0x00
quorum
FF
When a CAM unit matches the incoming data the local flip-flop is latched to 1.

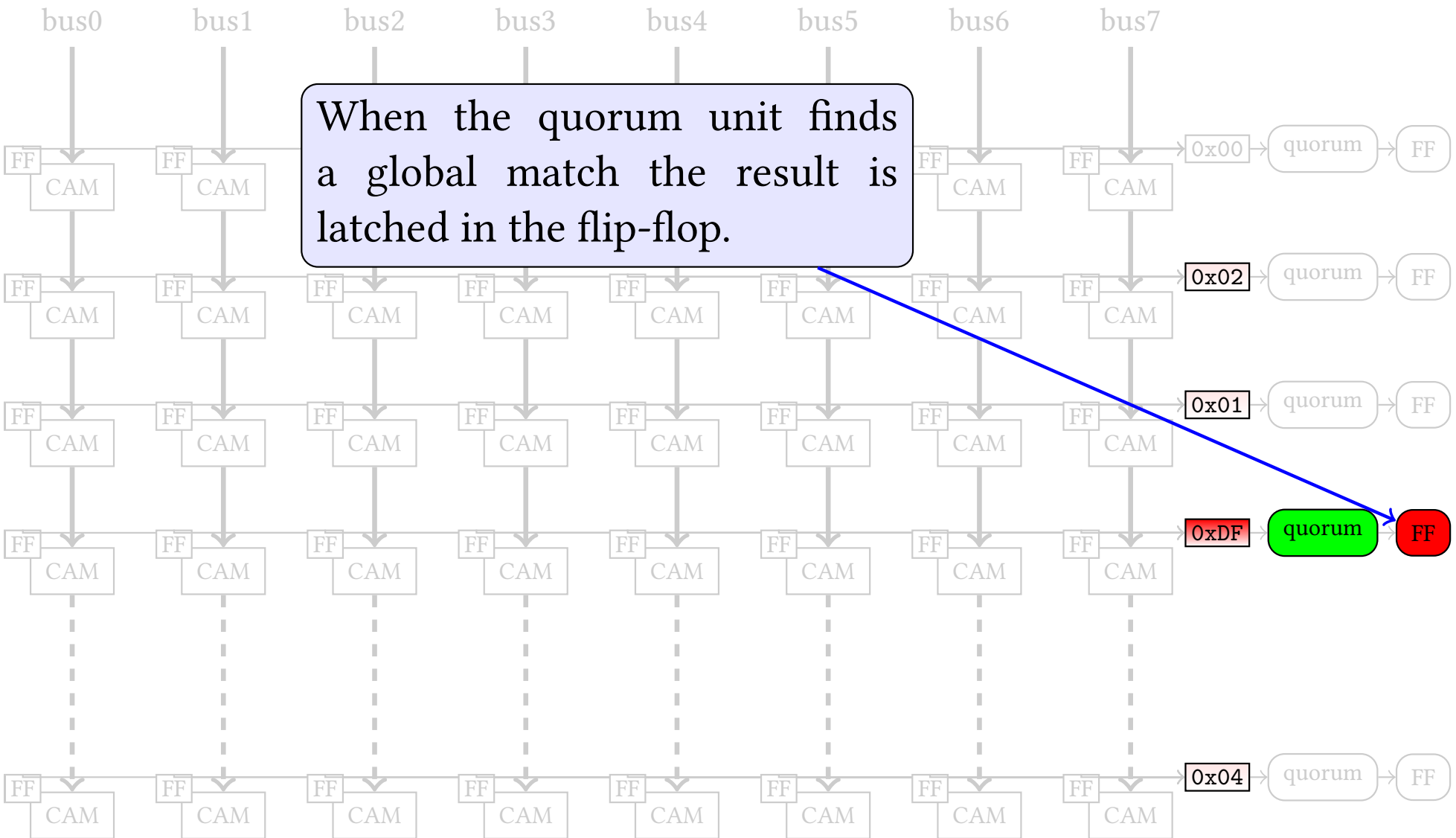

bus0
bus1
bus2
bus3
bus4
bus5
bus6
bus7
FF
CAM
0x00
quorum
As the data flows in the AM all local matches are recorded.

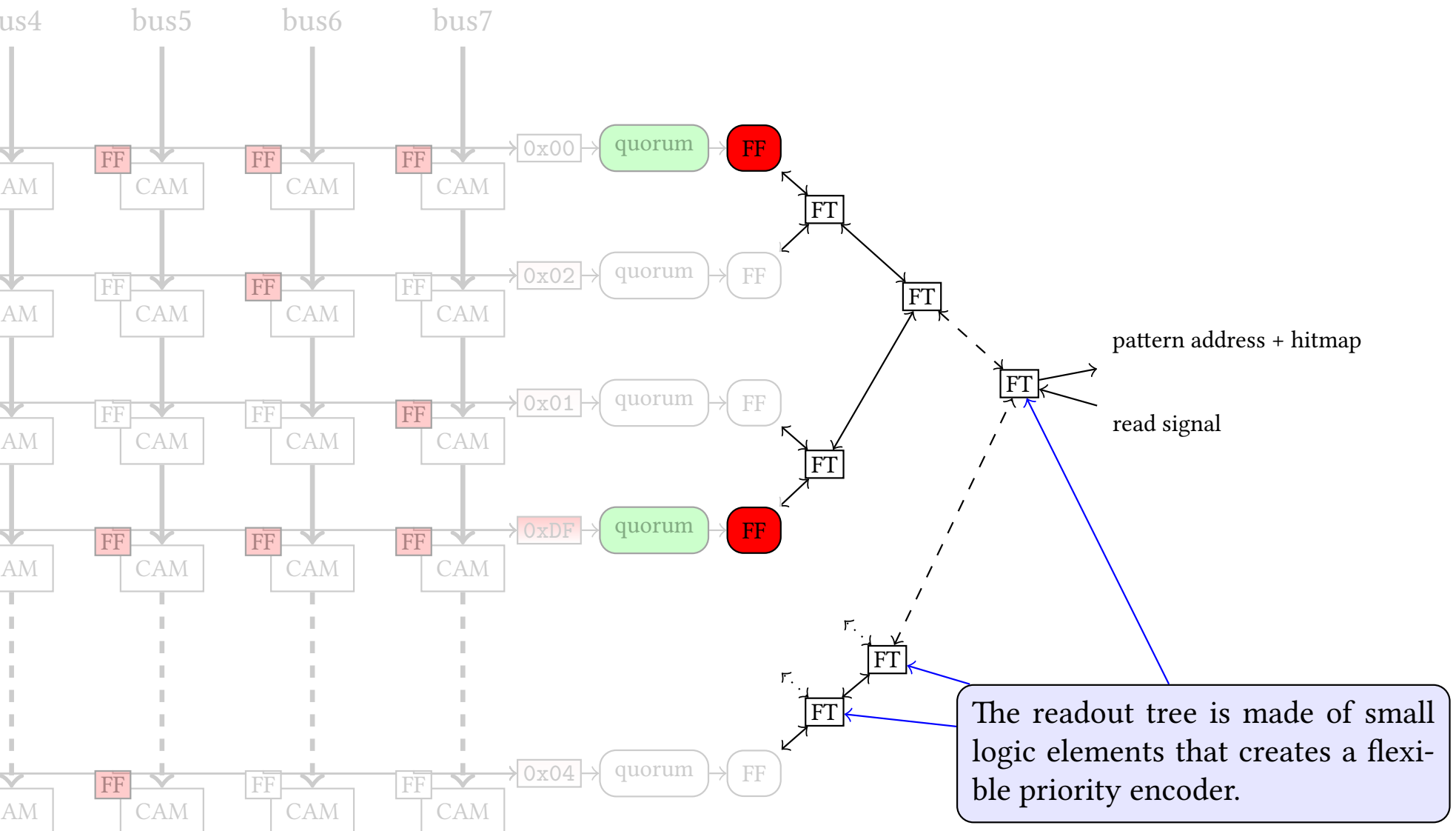

When the data of an event is finished the partial results are transfered to registers before the quorum unit.
bus0
bus1
bus2
bus3
bus4
bus5
bus6
bus7
FF
CAM
0x00
0x02
0x01
0xDF
0x04
quorum

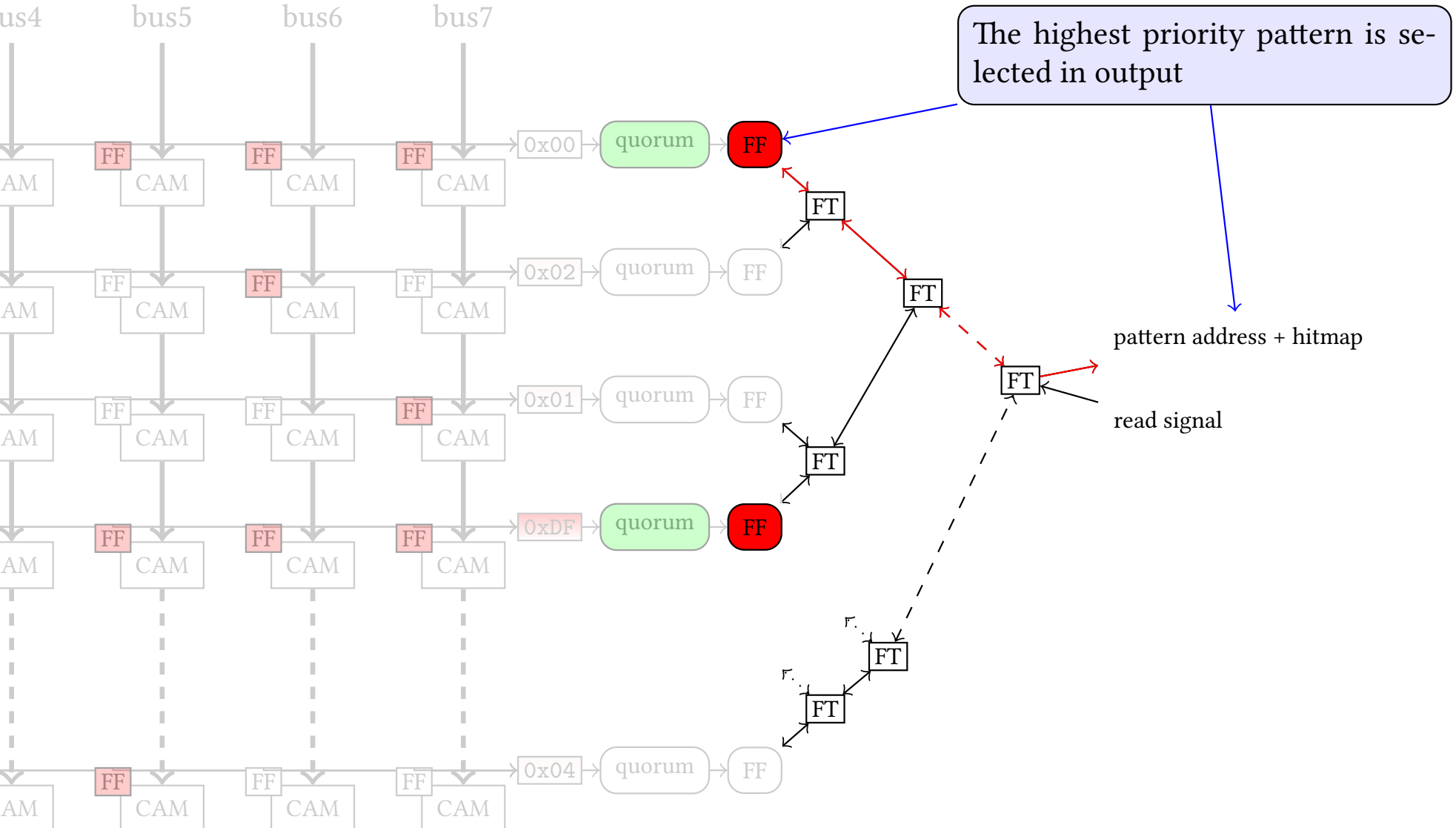

bus0
bus1
bus2
bus3
bus4
bus5
bus6
bus7
Then the local match memories are cleared in order to be ready for another event
FF
CAM
0x00
0x02
0x01
0xDF
0x04
quorum
FF

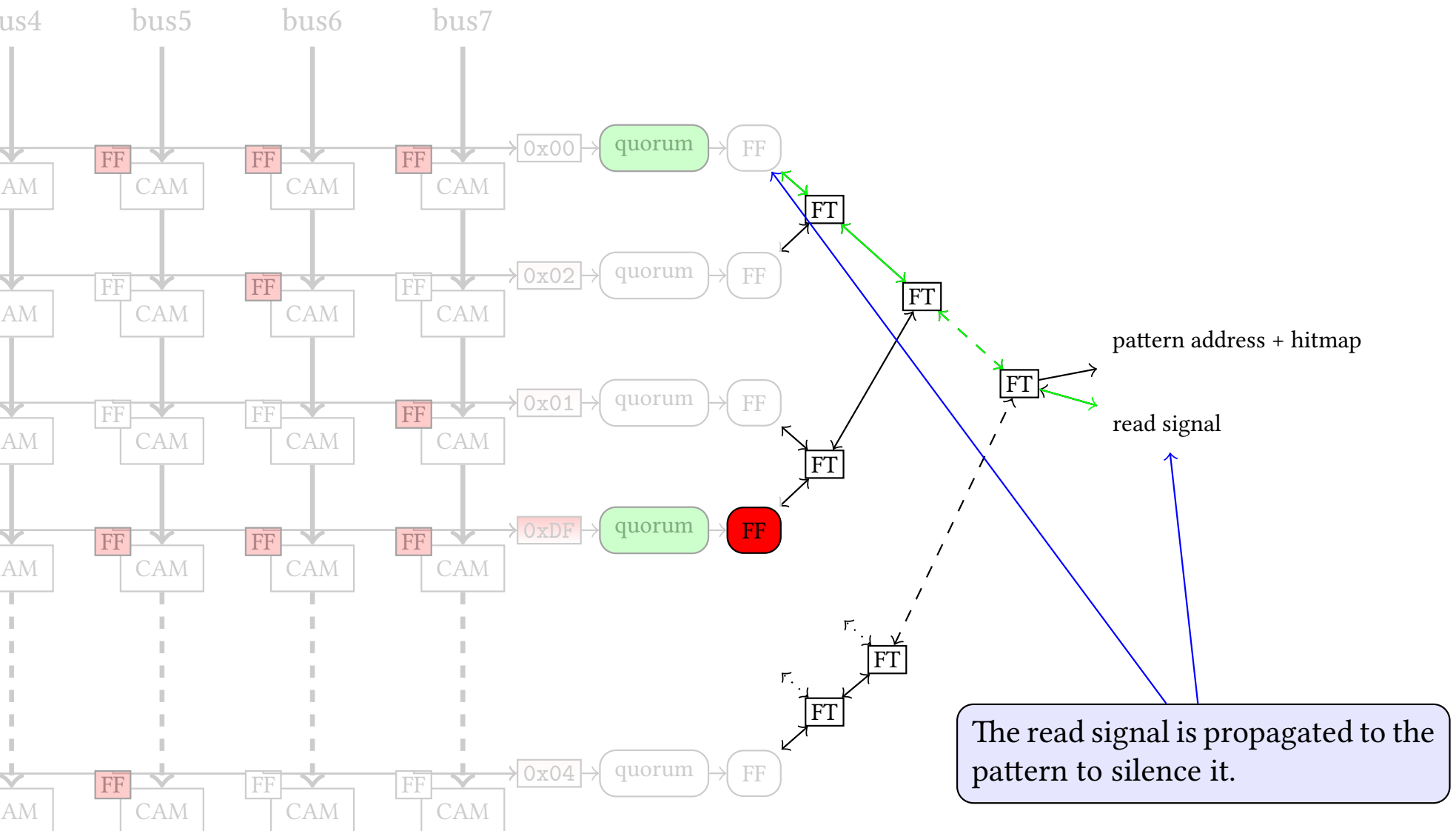


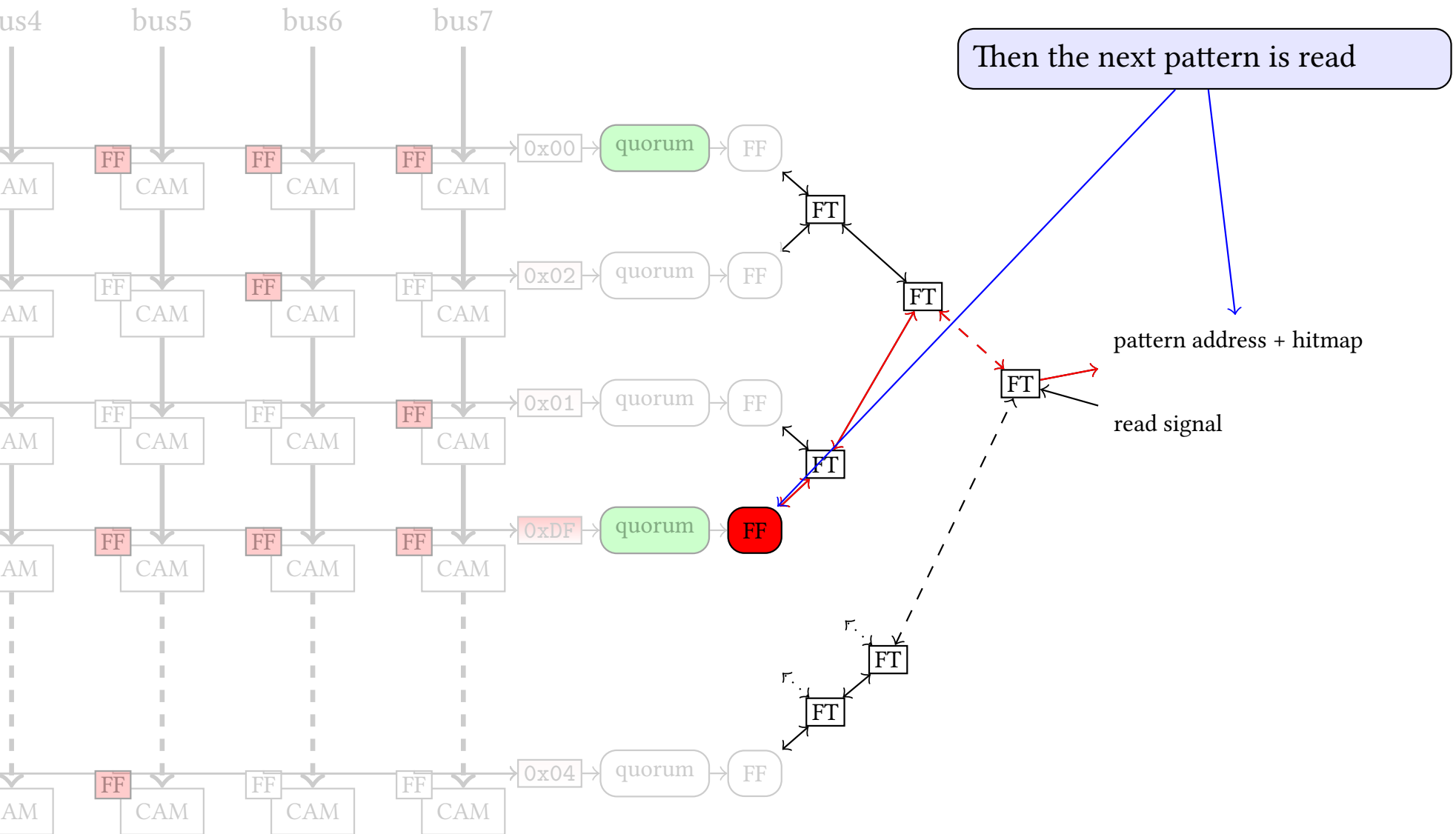
The quorum unit checks if the number of registered local matches (**Hamming Weight**) is over the threshold (ie. 7).

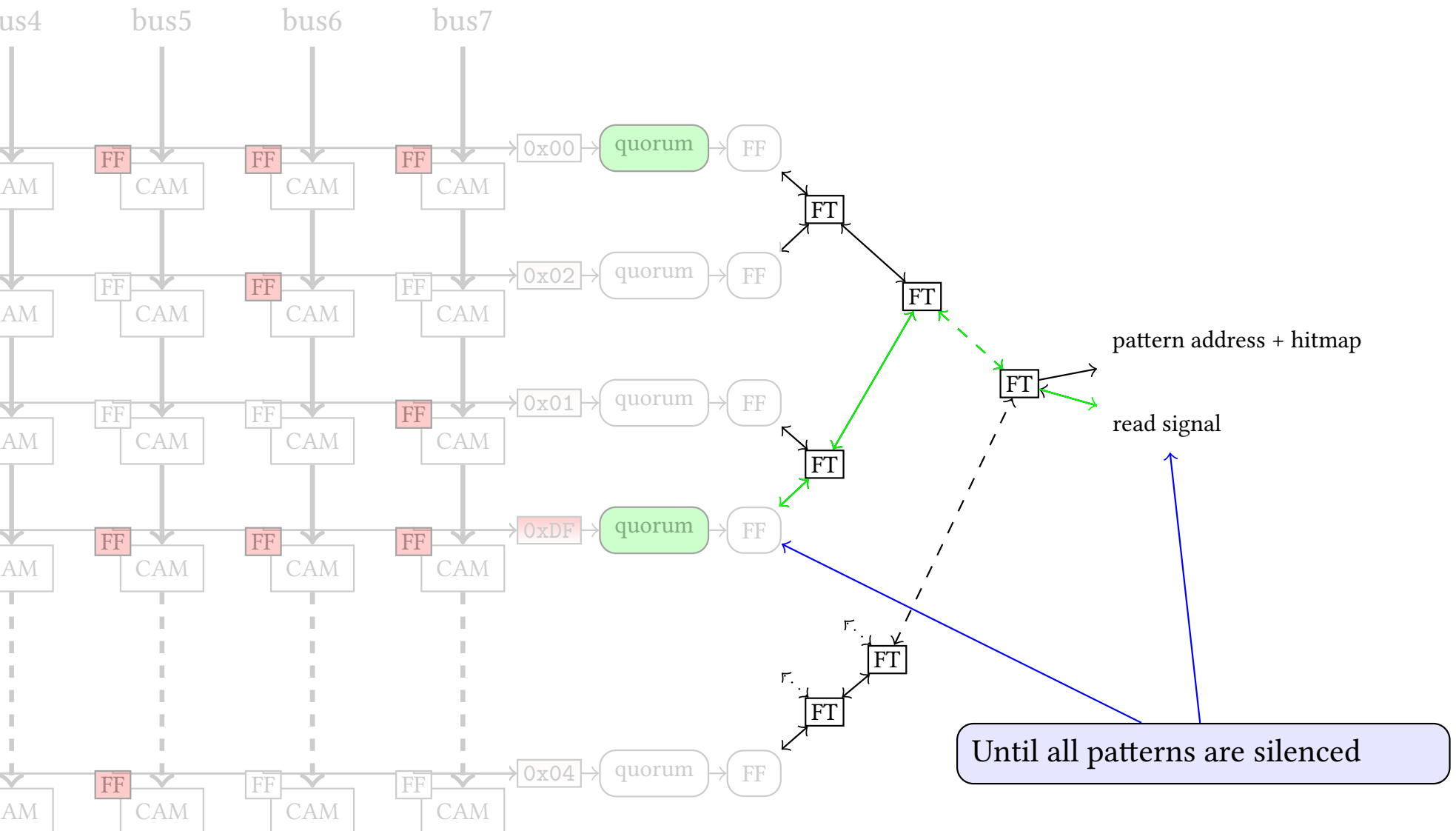

bus0
bus1
bus2
bus3
bus4
bus5
bus6
bus7
When the quorum unit finds a global match the result is latched in the flip-flop.
FF
CAM
0x00
0x02
0x01
0xDF
0x04
quorum
FF


bus4
bus5
bus6
bus7
FF
CAM
0x00
0x02
0x01
0xDF
0x04
quorum
FF
FT
pattern address + hitmap
read signal
The readout tree is made of small logic elements that creates a flexible priority encoder.

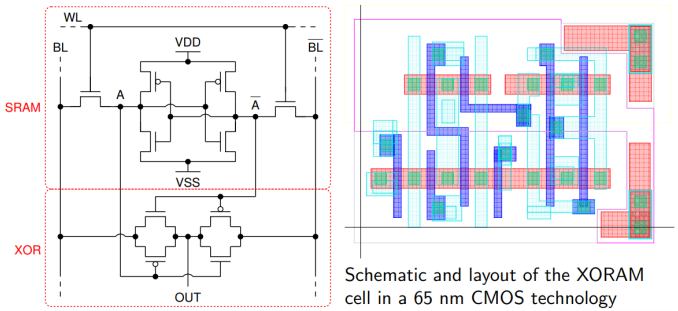

bus4
bus5
bus6
bus7
FF
CAM
0x00
0x02
0x01
0xDF
0x04
quorum
FT
The highest priority pattern is selected in output
pattern address + hitmap
read signal


bus4
bus5
bus6
bus7
0x00
0x02
0x01
0xDF
0x04
quorum
FF
CAM
FT
pattern address + hitmap
read signal
The read signal is propagated to the pattern to silence it.


bus4
bus5
bus6
bus7
Then the next pattern is read
FF
CAM
0x00
0x02
0x01
0xDF
0x04
quorum
FT
pattern address + hitmap
read signal


bus4
bus5
bus6
bus7
FF
CAM
0x00
0x02
0x01
0xDF
0x04
quorum
FT
pattern address + hitmap
read signal
Until all patterns are silenced

AMchip06



Schematic and layout of the XORAM cell in a 65 nm CMOS technology

Produced 25 wafers (8000 chips)

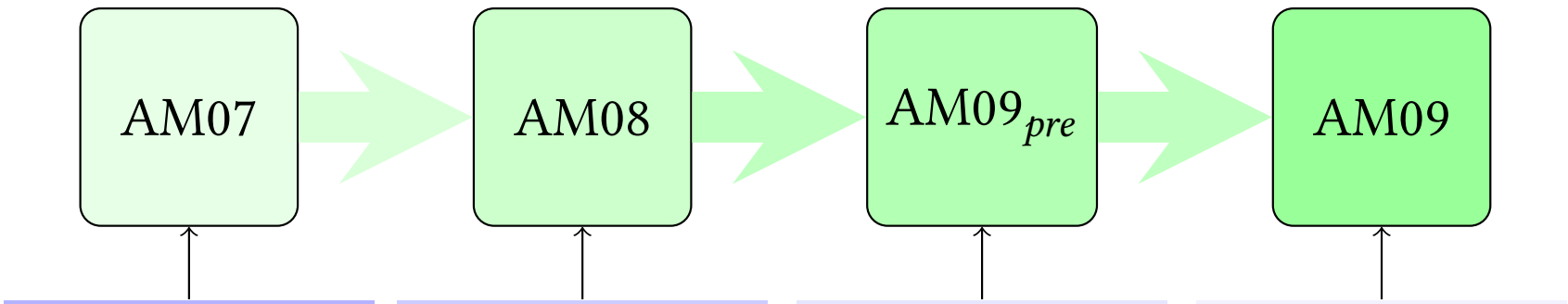
+25 wafers in 2018

- Digital ASIC
  - **65 nm TSMC**
  - 100 MHz
  - 168 mm^2
  - 128k 8x16 bit patterns
    - 2 ternary bits + 14 binary bits
  - Flip-chip BGA
  - MGT I/O at 2/2.4 Gbps
- Full-custom CAM cell
  - XORAM technology
    - doi:10.1109/ICECS.2012.6463629
  - Optimized for low power
- Power consumption during operations avg 3 W

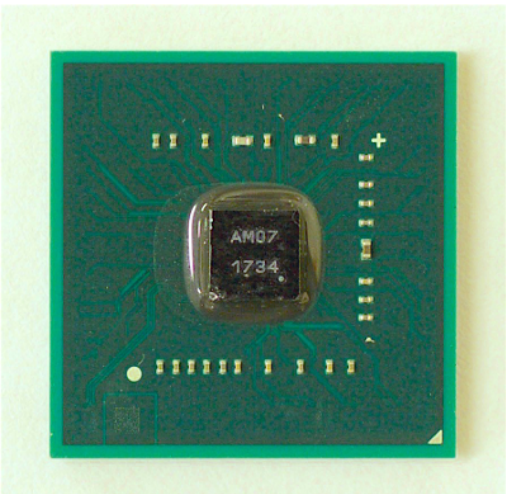
Power consumption figure of merit:
2.3 fJ/bit/comparison @ 1.15 V 1.8 fJ/bit/comparison @ 1.0 V

Goals for the ATLAS/CMS Phase-II chip

- 384k patterns in $\simeq 150$ mm^2
  - 16 bit * 8 busses
  - 2 ternary bits + 14 binary bits
- 250 MHz comparison clock
- LVDS DDR IOs @ 500 MHz to reduce pinout, but to avoid MGT complexity
- At least 1 fJ/bit/comparison
- **28 nm**



AM07

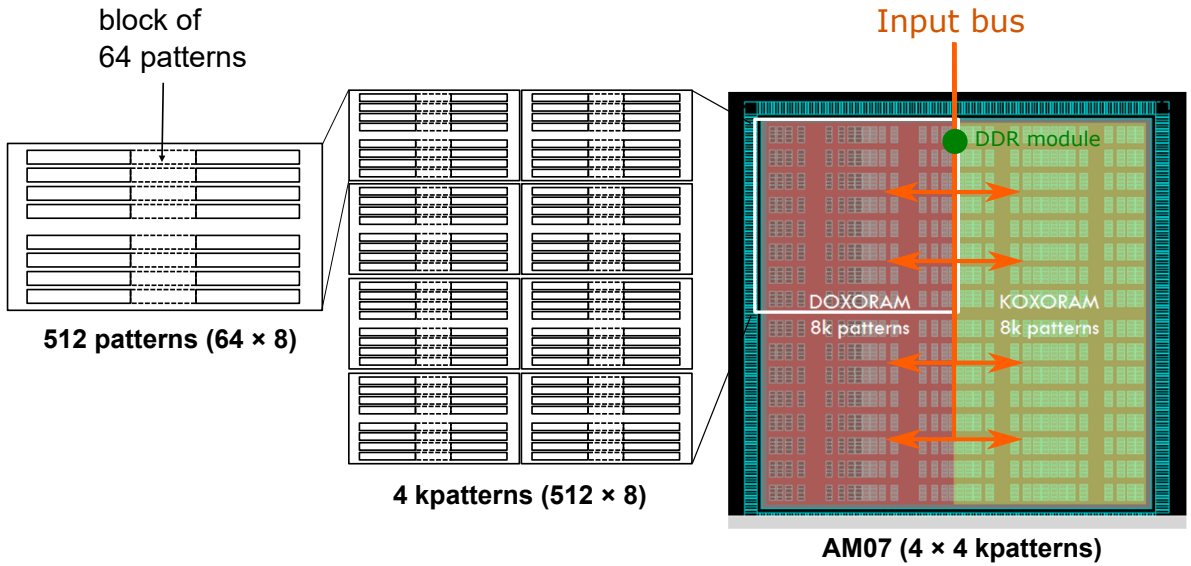


17 × 17 BGA

- 10 mm^2 28 nm
- $4 \times 4k$ patterns organized in independent cores
  - 2x DOXORAM, 2x KOXORAM
  - Both are evolutions of the XORAM of AM06. Patent pending in Italy.
- 1 bus using 9x LVDS pairs in DDR mode (18 bit)
- 7 busses using 18 bit LVCMOS
- 4 outputs LVCMOS (12 bit address + 8 bit hitmap)
- Designed to run internally at 200 MHz
- RX/TX LVDS test drivers

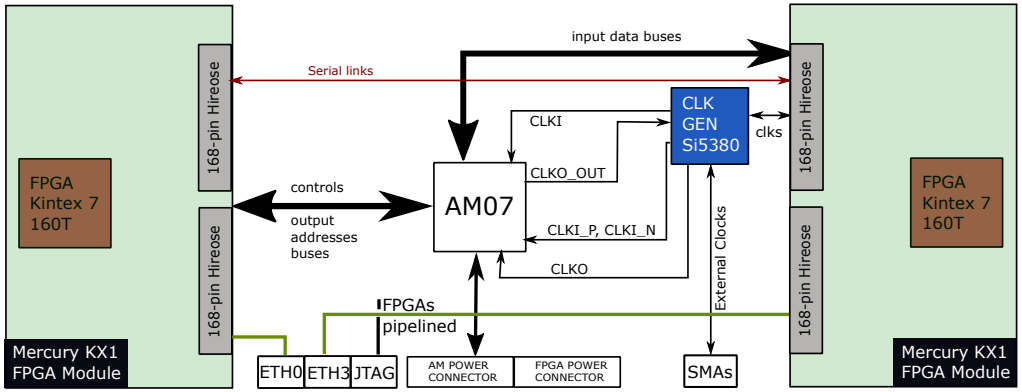
Financed by ANR FTK (LPNHE/IPNL) and INFN IMPART and RD_PHASE2 (ATLAS/CMS) projects

AM07 Internals

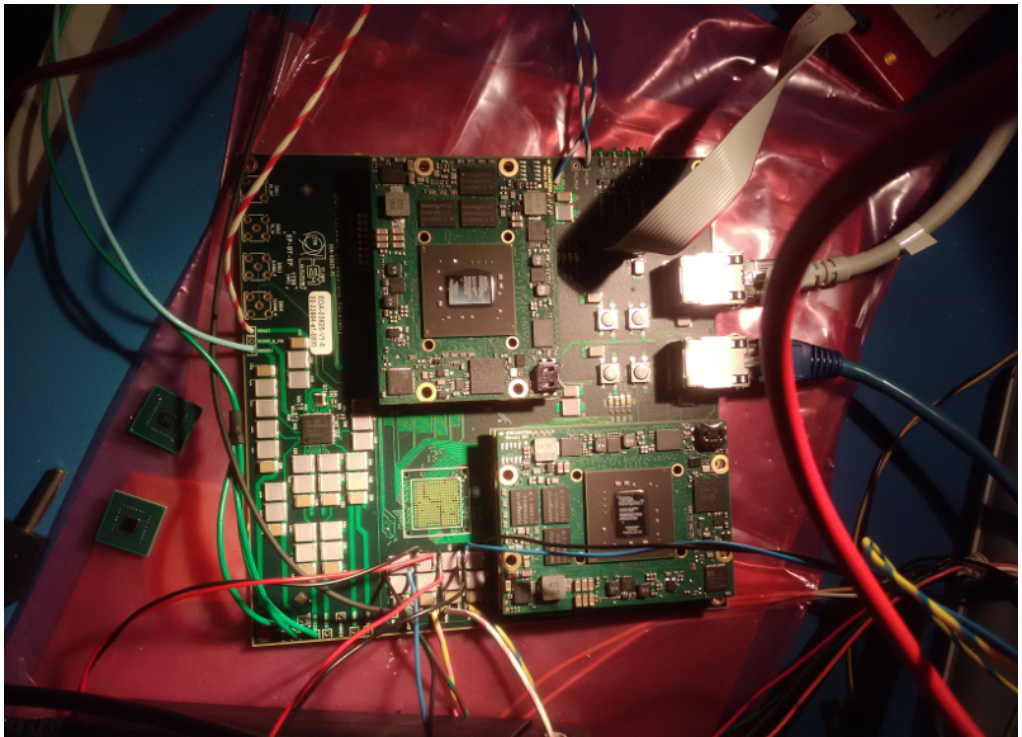


- Full-custom blocks (CAM, LVDS) designed with Cadence Virtuoso
- Standard cell logic synthesized from VHDL using Cadence Genus
- Top level integration and P&R using Cadence Innovus
- Static timing analysis using Cadence Tempus
- IR drop analysis using Cadence Voltus
- Functional simulation in UVM using Cadence Incisiv

AM07 Test Board



To test the AM07 we need two FPGAs on the test board to drive all the LVCMOS and LVDS signals.



2x Xilinx Kintex 160t
Si5380 low jitter clock manager (LTE freq optimized, not all freq available)
Firmware based on IPBus + python scripts

Functional Tests

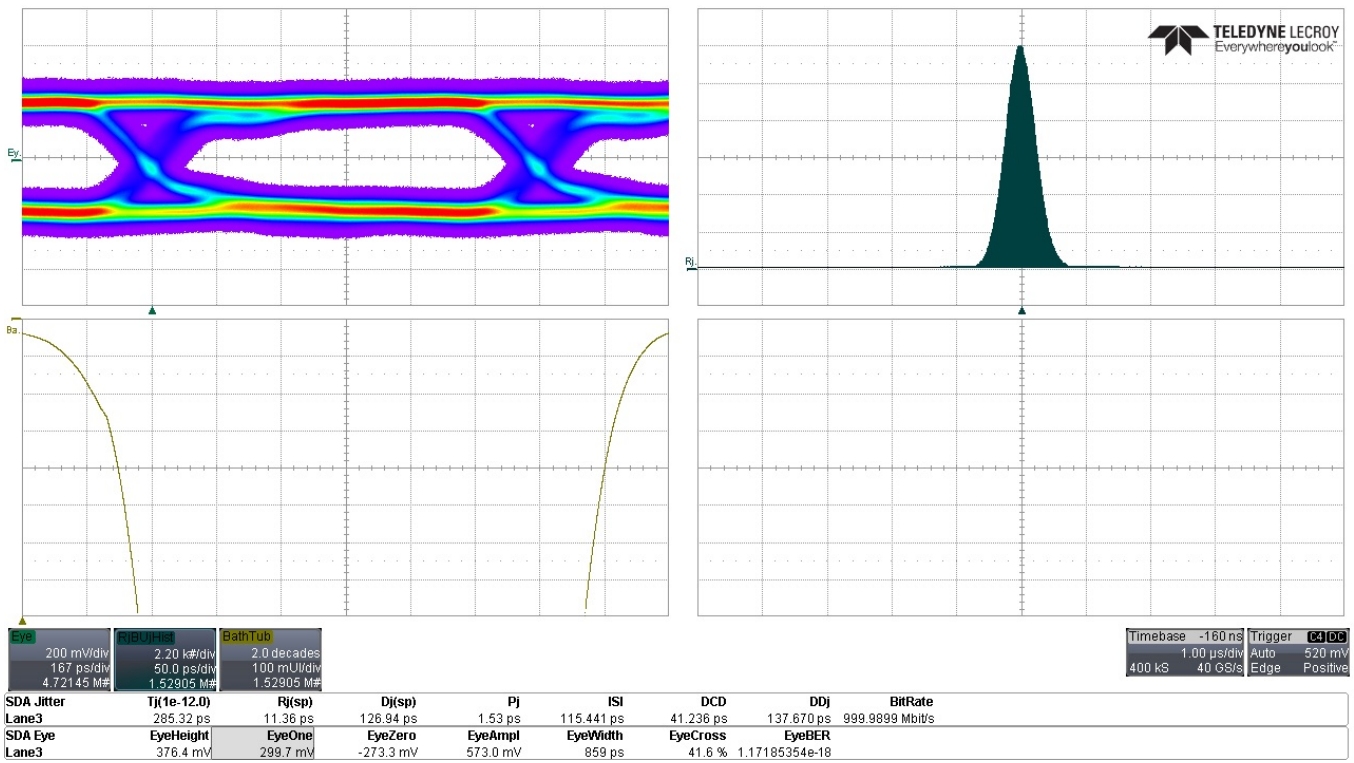
- AM07 has an internal feature to scan the pattern bank and produce an unique CRC value
  - Scan is successful up to 245 MHz (chip was designed for 200 MHz)
- LVDS RX/TX tested up to 1.1 gbps
- In the FPGA we can tune all clk relative phases and individual delays for each IO to deskew the busses
  - Zero errors up to 150 MHz using all busses, CLK via LVCMOS
  - Zero errors up to 180 MHz using LVDS bus and CLK via LVDS
- Tests still ongoing (FPGA fw more difficult than expected)

BIST & LVDS

BIST Max core frequency

0.65	0.75	0.85	0.95	1.00	1.05	1.10
fail	fail	105	184	245	245	245

Core supply [V]



Power consumption

Precise measurements are still ongoing, as the power consumption depends on the data and the matching and it must be verified in functionality tests.
However preliminary results show a good agreement with the Calibre simulations:

Tech.	Meas (fJ/bit/comp)	Sim (fJ/bit/comp)
KOXORAM	0.748	0.69
DOXORAM	0.851	0.91

AM08

The purpose of this small area prototype (10 mm^2) is to finalize all the features and interfaces that will be used by the final chip (AM09).

- AM08 development already in an advanced state
  - Submission foreseen in the fall
- New features:
  - All LVDS DDR IOs @ 500 MHz
    - Control words encoded in the data stream
  - Configuration via SPI
  - 250 MHz core freq
  - Internal DPLL
    - To generate 8x 250 MHz clocks 45° apart to be used in different cores and spread the power consumption over the clk period

Five papers accepted at IEEE's International Symposium on Circuits and Systems (ISCAS):

- Characterization of an Associative Memory chip in 28 nm CMOS technology
- Characterization of an LVDS link in 28 nm CMOS for multi-purpose pattern recognition chip
- Design and characterization of new Content Addressable Memory cells
- A fully-digital delay locked loop in 28 nm CMOS
- Temperature sensor with process and mismatch auto-compensation technique in 28 nm CMOS

Conclusions

- The Associative Memory is a device designed for real-time pattern recognition in high performance computing applications
  - It has been used in tracking hardware processor at hadron collider experiments (CDF SVT, ATLAS FTK)
  - It will be used in future HL LHC experiments upgrades (ATLAS HTT)
  - To meet new requirements we pursued the development of the AM in 28 nm
- AM07, the first fully functional prototype in 28 nm, has been produced in 2017 and it's under test
  - Functionality has been verified (at least up to 150 MHz, internally up to 245 MHz)
  - Power consumption is compatible with the simulation expectations and in track with our roadmap to the final chip
- AM08 is under design, it will be submitted in fall 2018
  - Features & interface of the final chip