Trigger Happy 2

The ongoing evolution of the CMS Detector and its L1-Trigger

Andy Rose, Imperial College, London

awr01@imperial.ac.uk

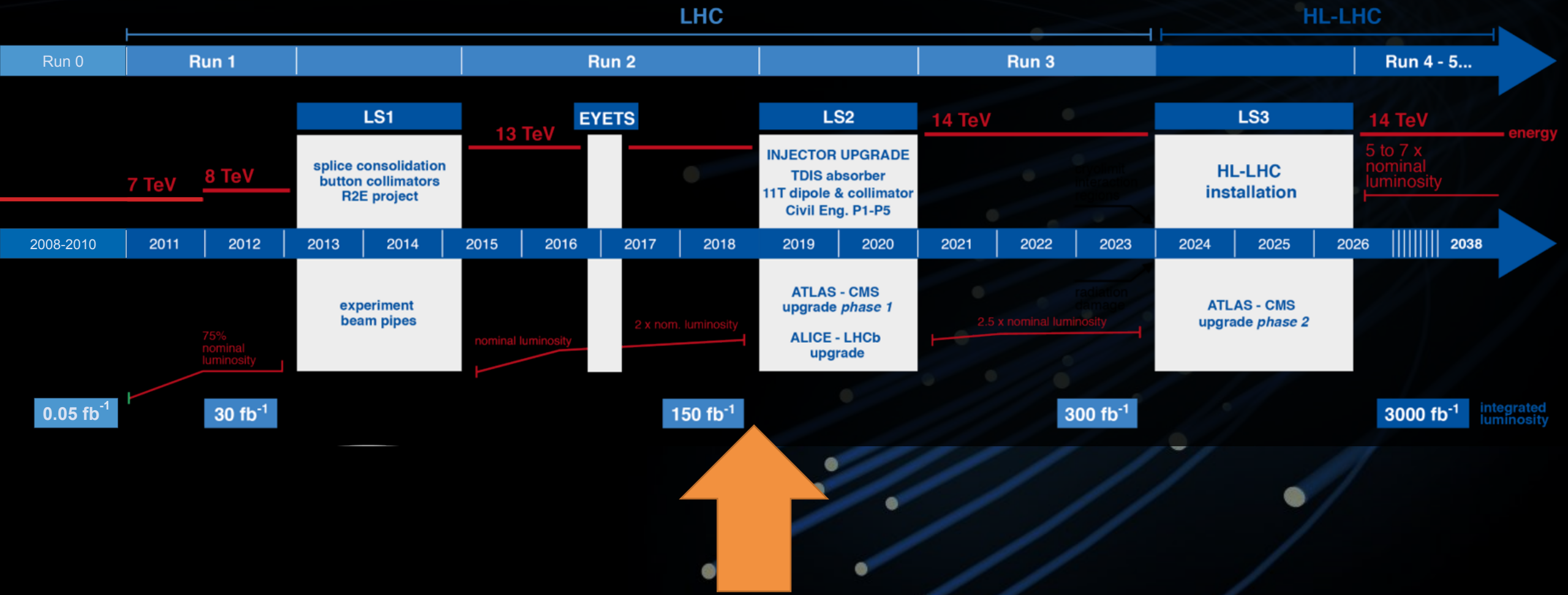
www.hep.ph.ic.ac.uk/~awr01

Introduction

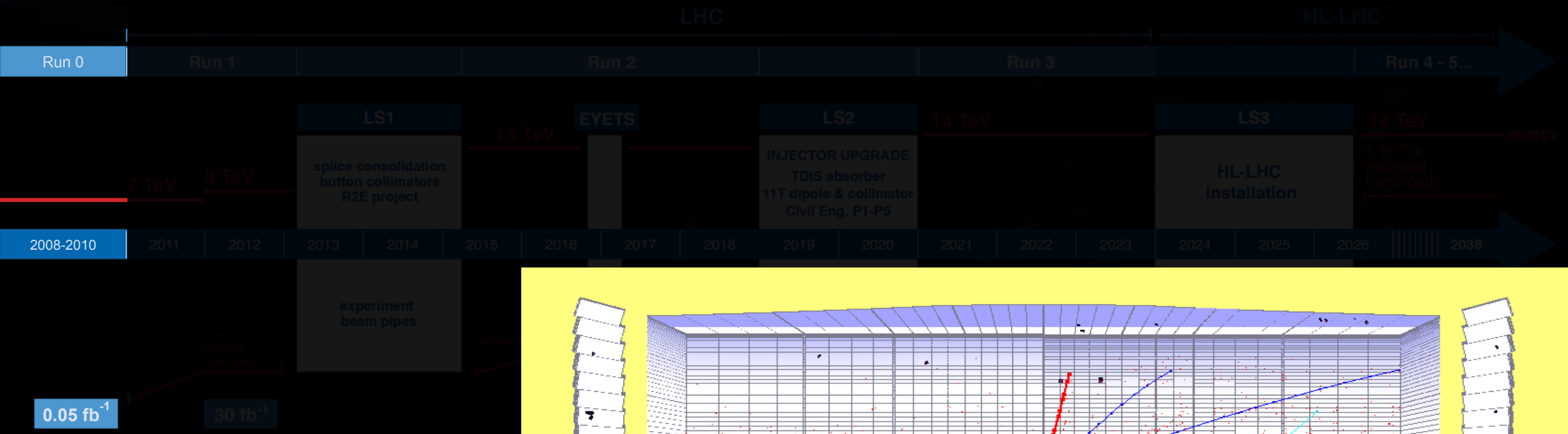
- This is a sequel to my 2011 seminar “Trigger Happy: A light-hearted look at the seriously hard business of hardware triggering (and how we are trying to make it easier)”
  - http://www.hep.ph.ic.ac.uk/~awr01/seminars/triggerhappy/TriggerHappy_v1.1.pptx
- That talk:
  - Honestly (brutally) assessed the shortcomings of the existing CMS trigger
  - Lay out my glorious vision for the Phase-I upgrade of CMS trigger electronics
    - Speaker’s prerogative when you gave me a seminar soapbox
  - Parts of that vision came to pass, some parts didn’t
- This talk will assess where CMS stands now, and what it faces after the Phase-II upgrades
  - And I’ll lay out my new glorious vision
  - Well… if you are a seminar soapbox again

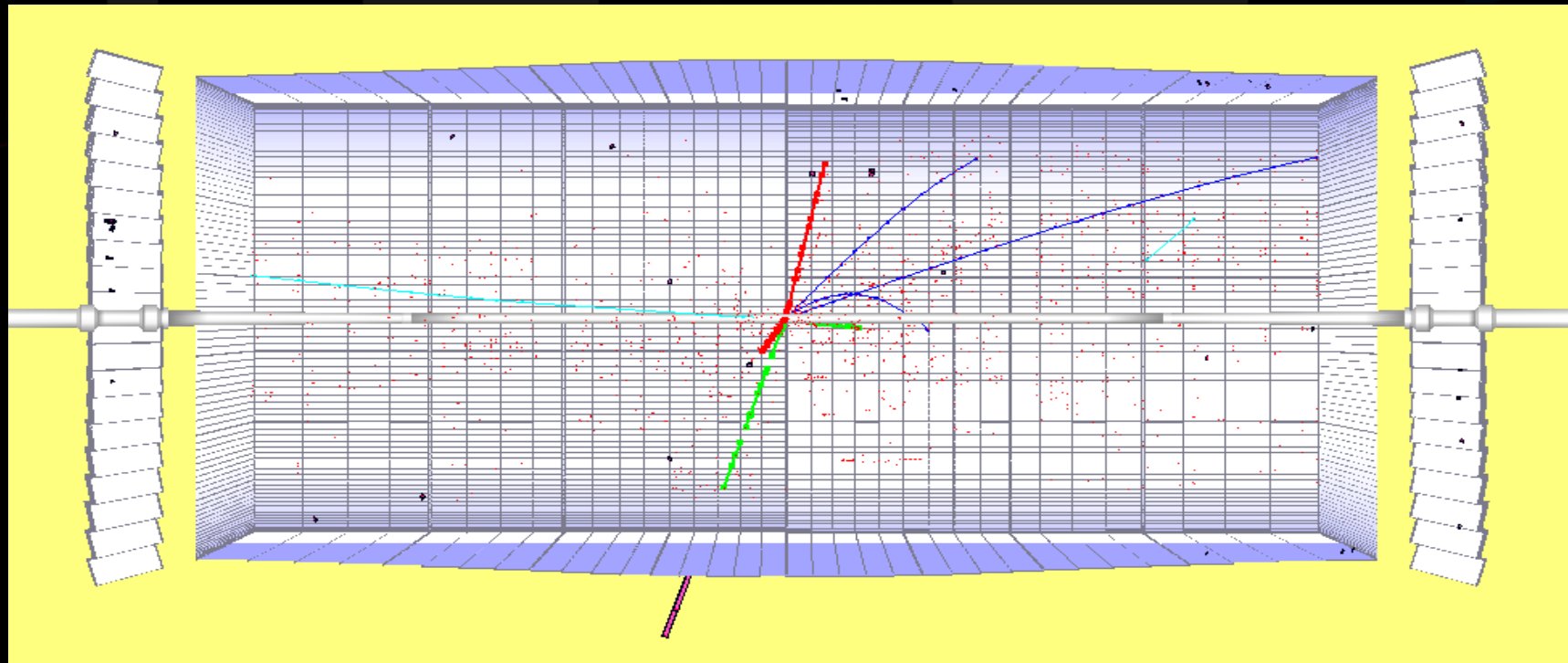
Where are we with the LHC again?

Where are we with the LHC again?

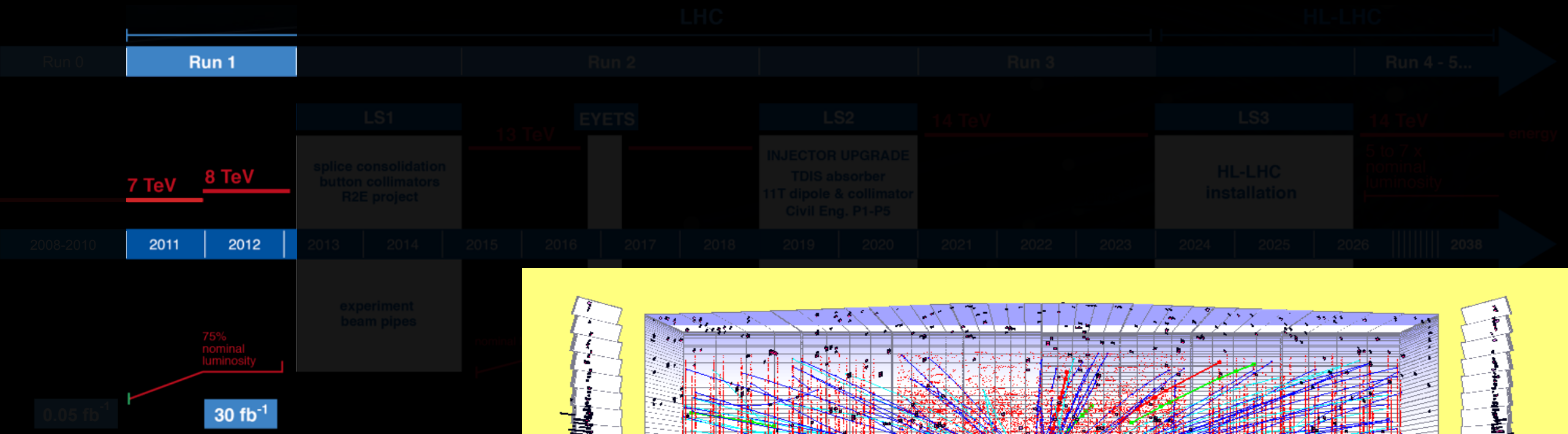


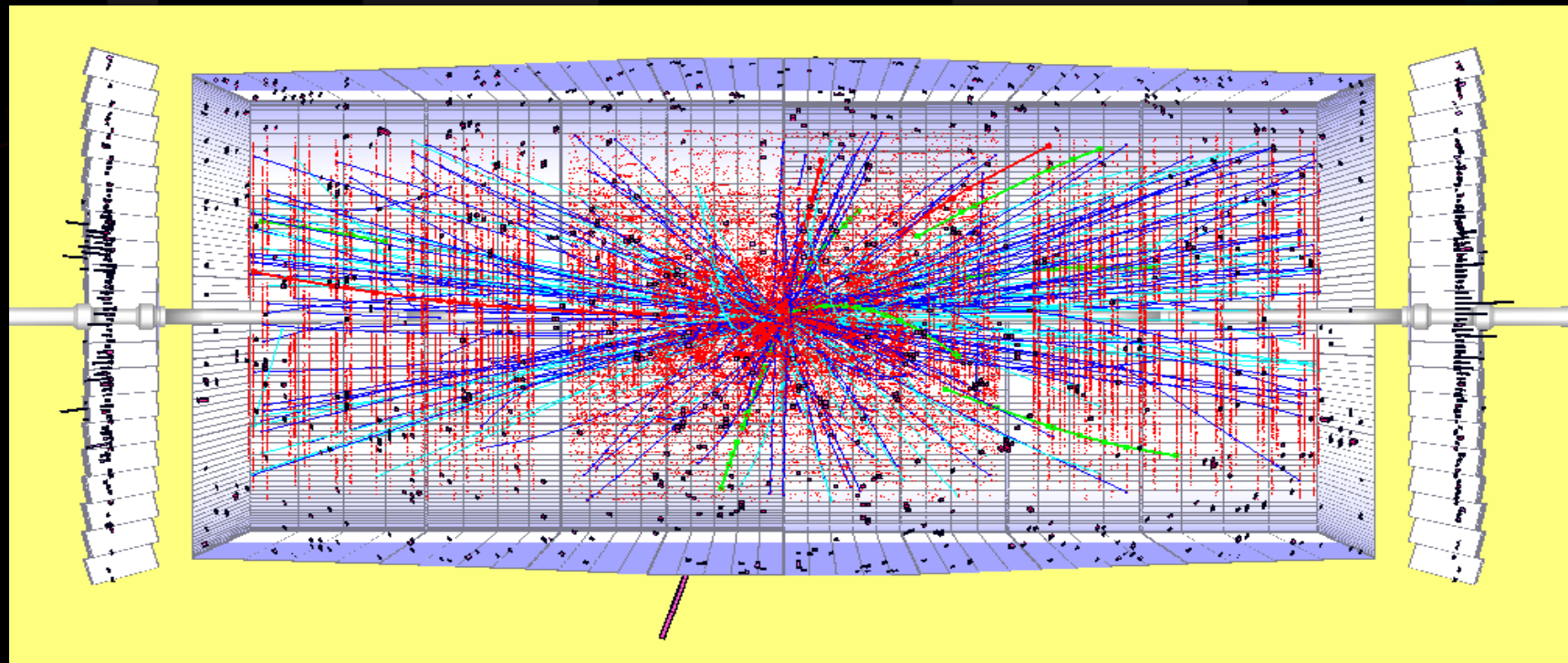
In the beginning…



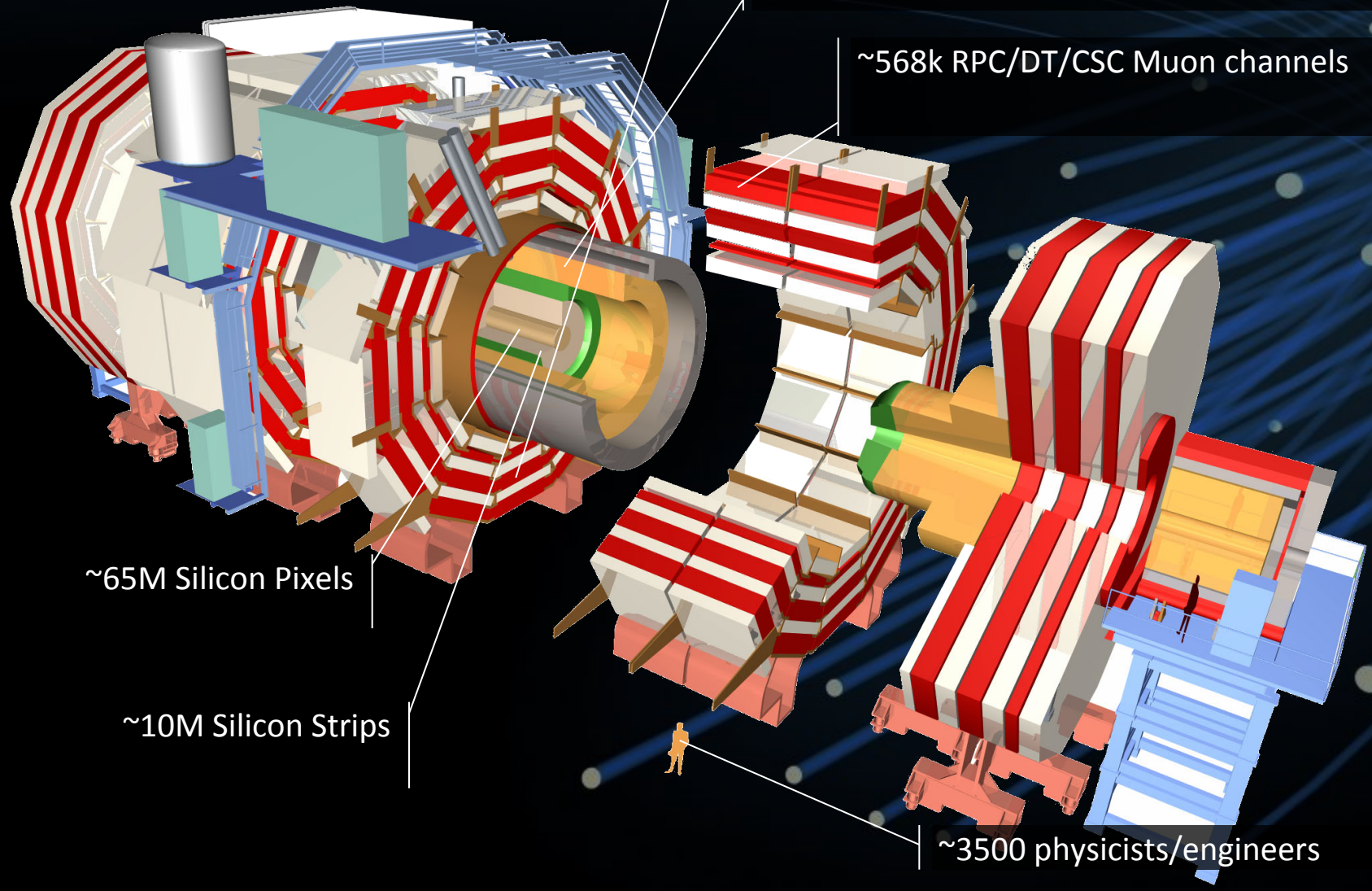


Run 1

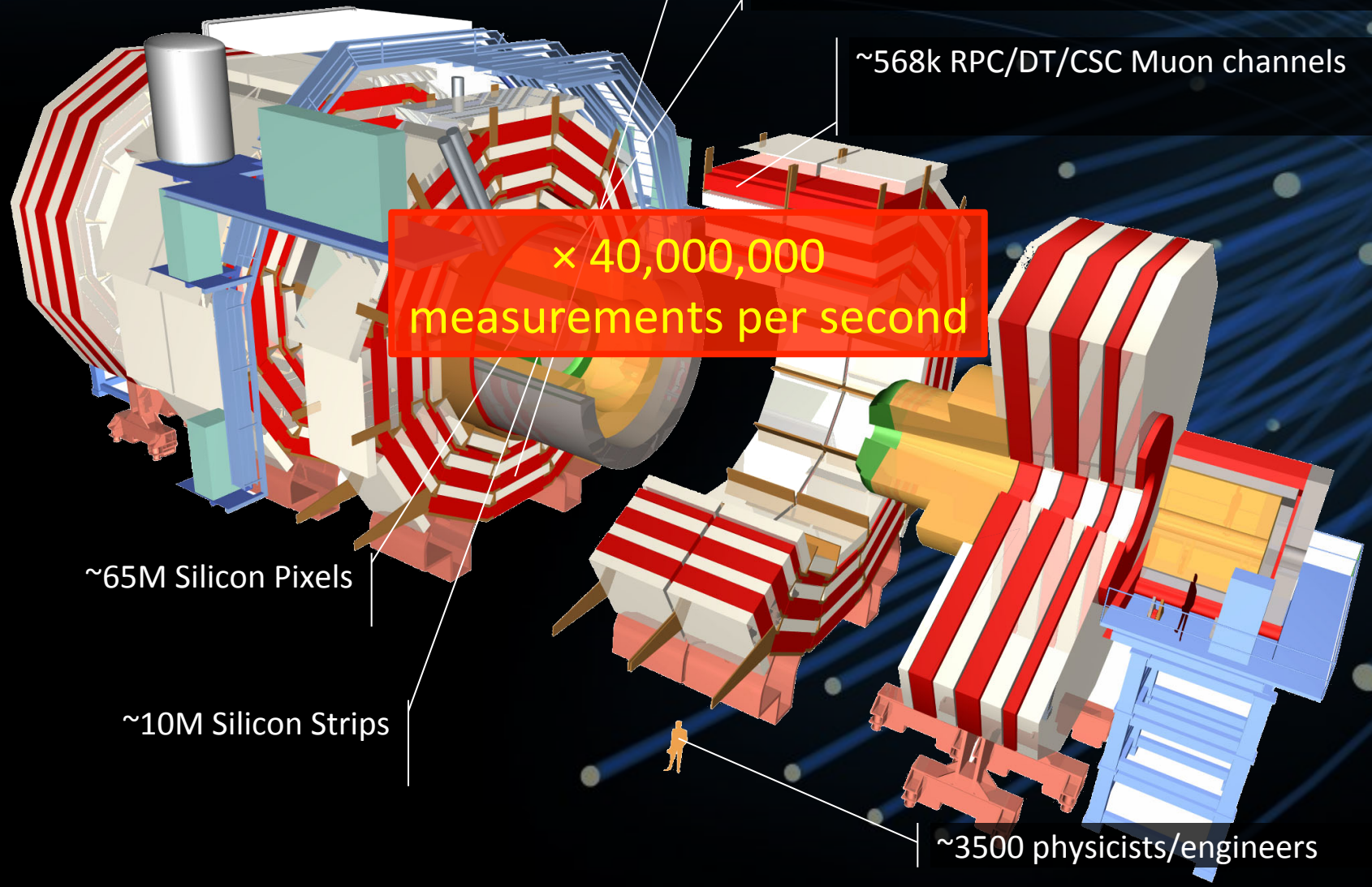

LHC
HL-LHC
Run 0
Run 1
Run 2
Run 3
Run 4 - 5...
LS1
EYETS
LS2
LS3
14 TeV
splice consolidation
button collimators
R2E project
INJECTOR UPGRADE
TDIS absorber
11T dipole & collimator
Civil Eng. P1-P5
HL-LHC
installation
7 TeV
8 TeV
13 TeV
2008-2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2038
experiment
beam pipes
75%
nominal
luminosity
0.05 fb-1
30 fb-1



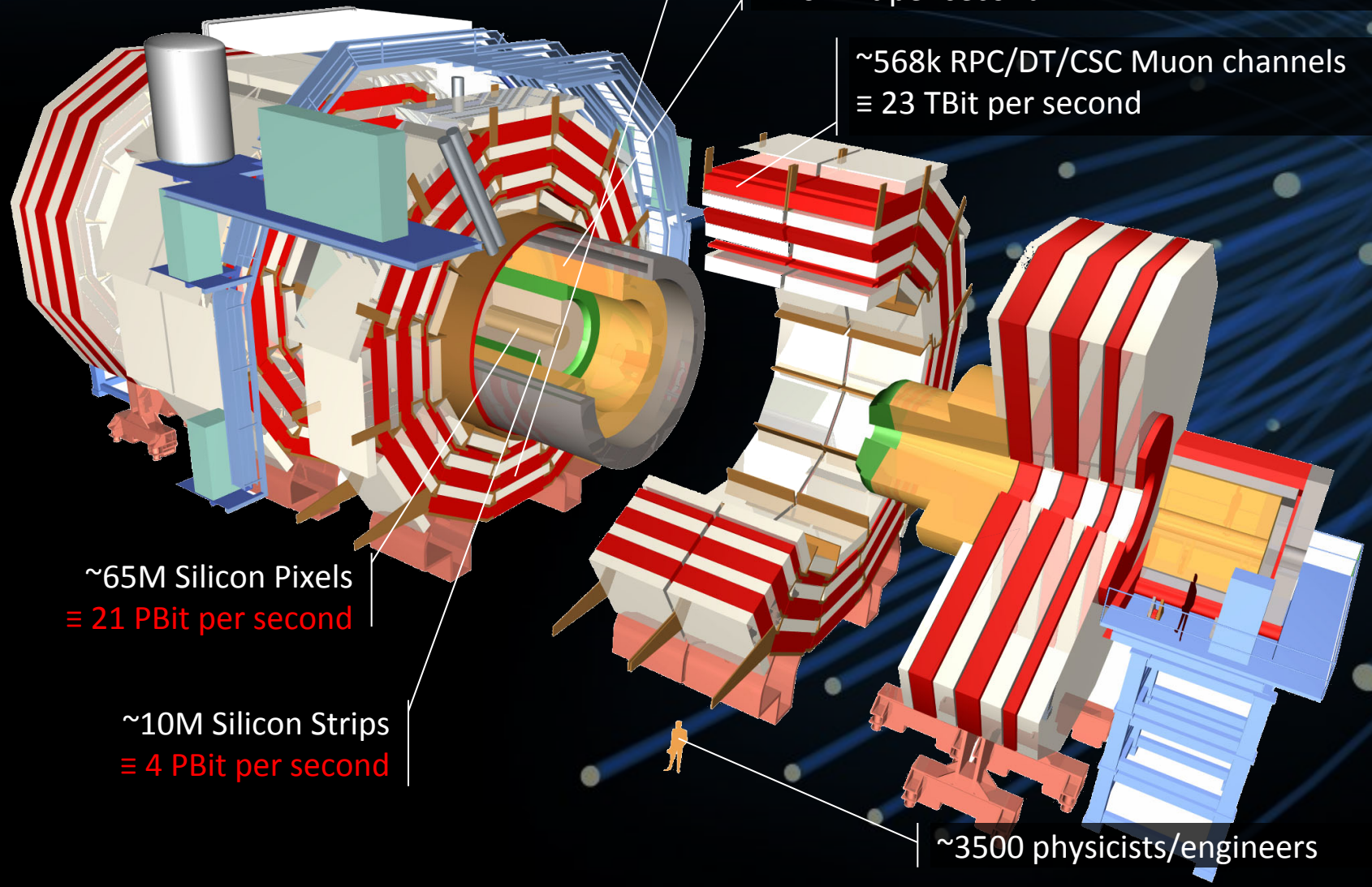
Recap


~80k $PbWO_4$ Ecal Crystals
~15k channel Brass/Plastic sampling HCAL
~568k RPC/DT/CSC Muon channels
~65M Silicon Pixels
~10M Silicon Strips
~3500 physicists/engineers

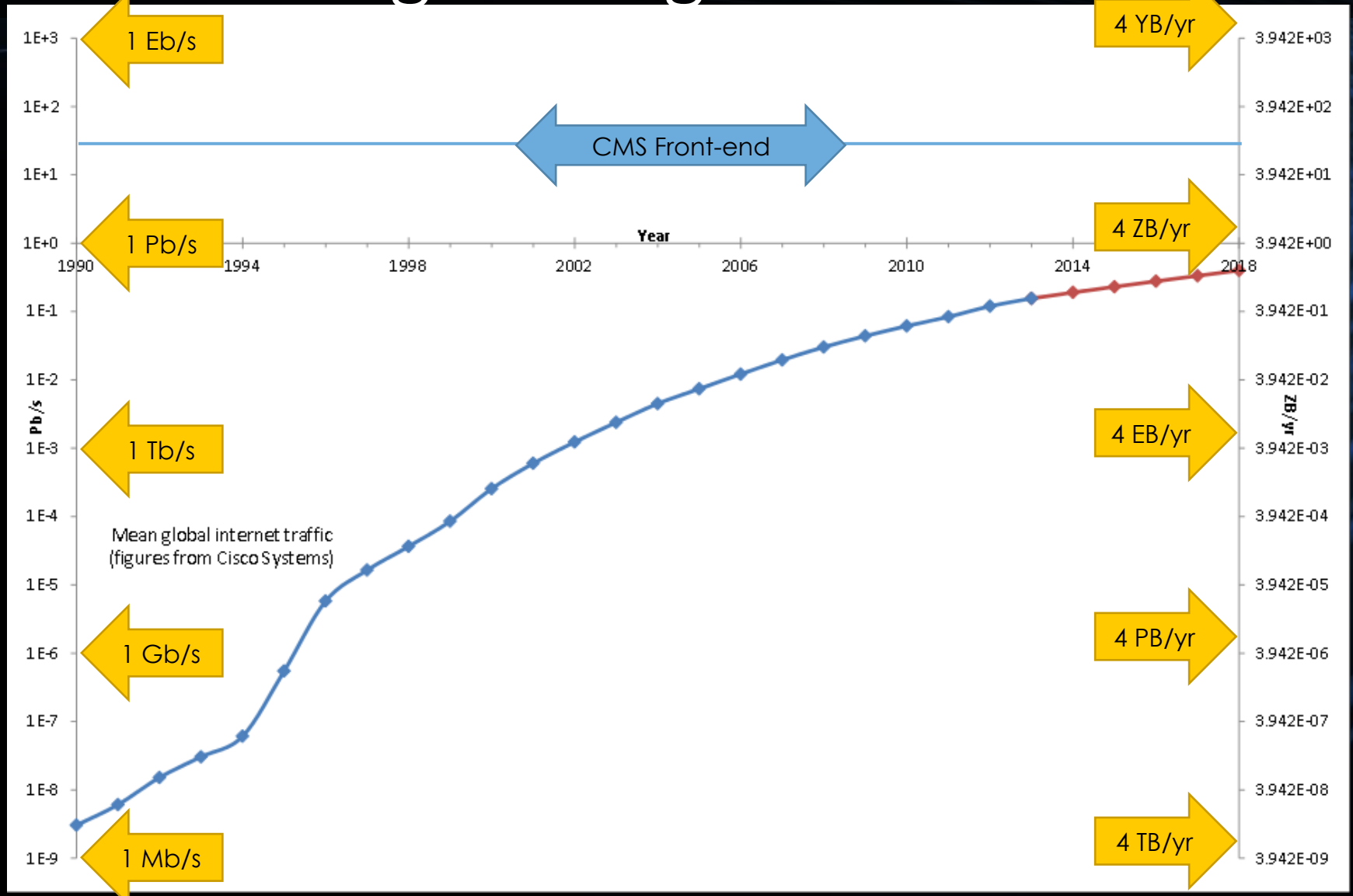
Recap


~80k $PbWO_4$ Ecal Crystals
~15k channel Brass/Plastic sampling HCAL
~568k RPC/DT/CSC Muon channels
× 40,000,000
measurements per second
~65M Silicon Pixels
~10M Silicon Strips
~3500 physicists/engineers

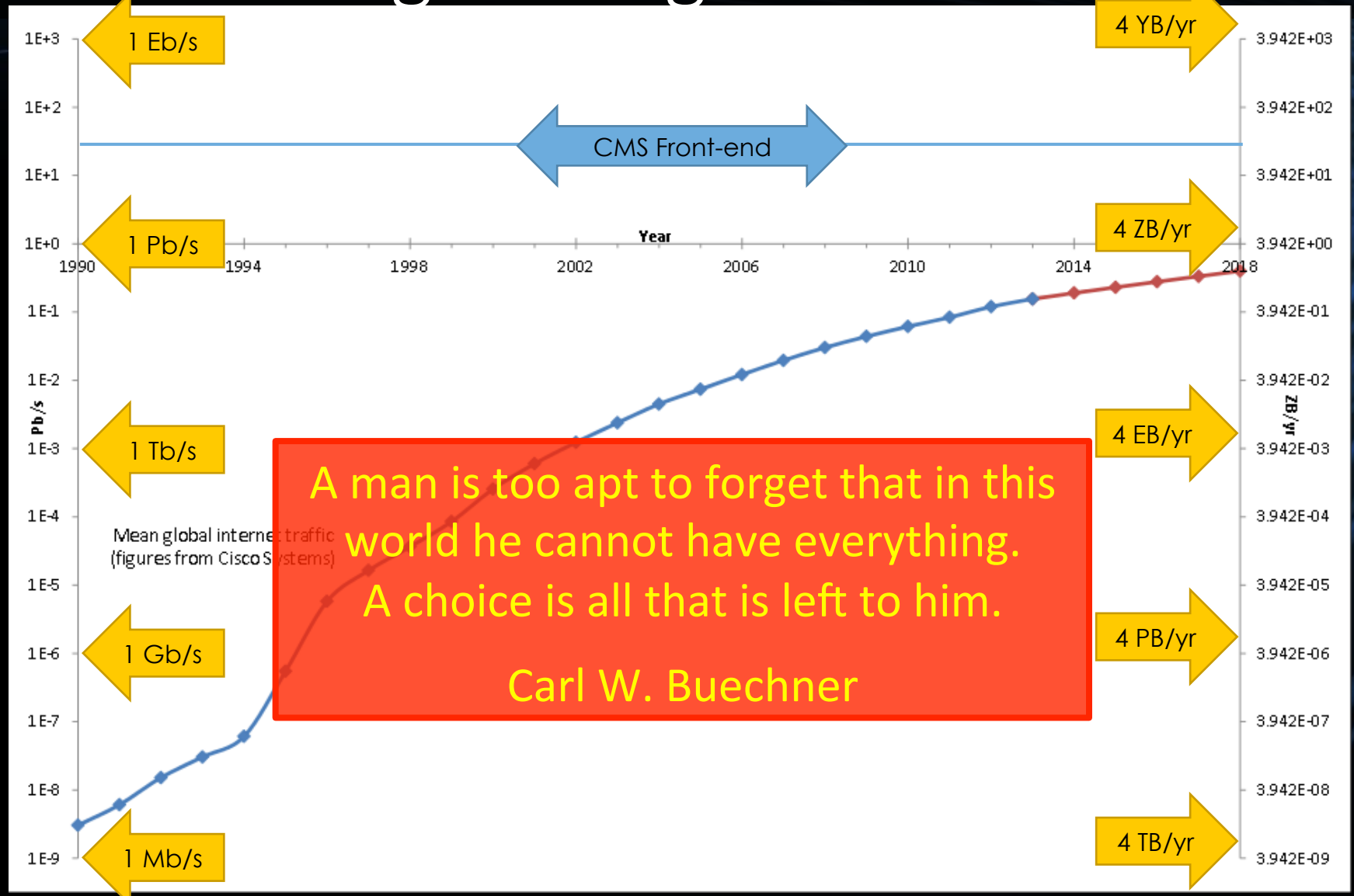
Recap


~80k $PbWO_4$ Ecal Crystals
≡ 40 TBit per second
~15k channel Brass/Plastic sampling HCAL
≡ 10 TBit per second
~568k RPC/DT/CSC Muon channels
≡ 23 TBit per second
~65M Silicon Pixels
≡ 21 PBit per second
~10M Silicon Strips
≡ 4 PBit per second
~3500 physicists/engineers

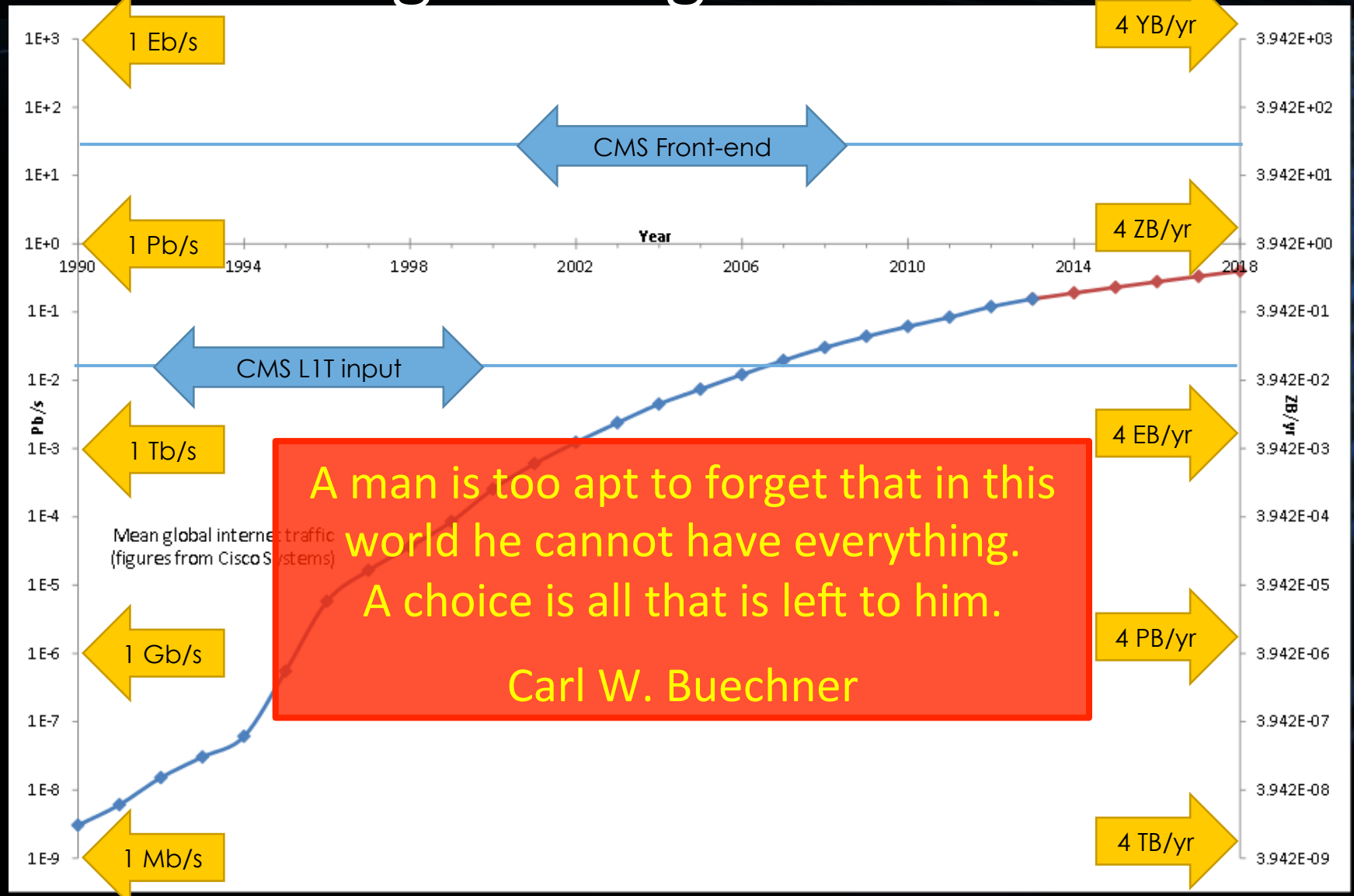
CMS: Visualizing the big numbers



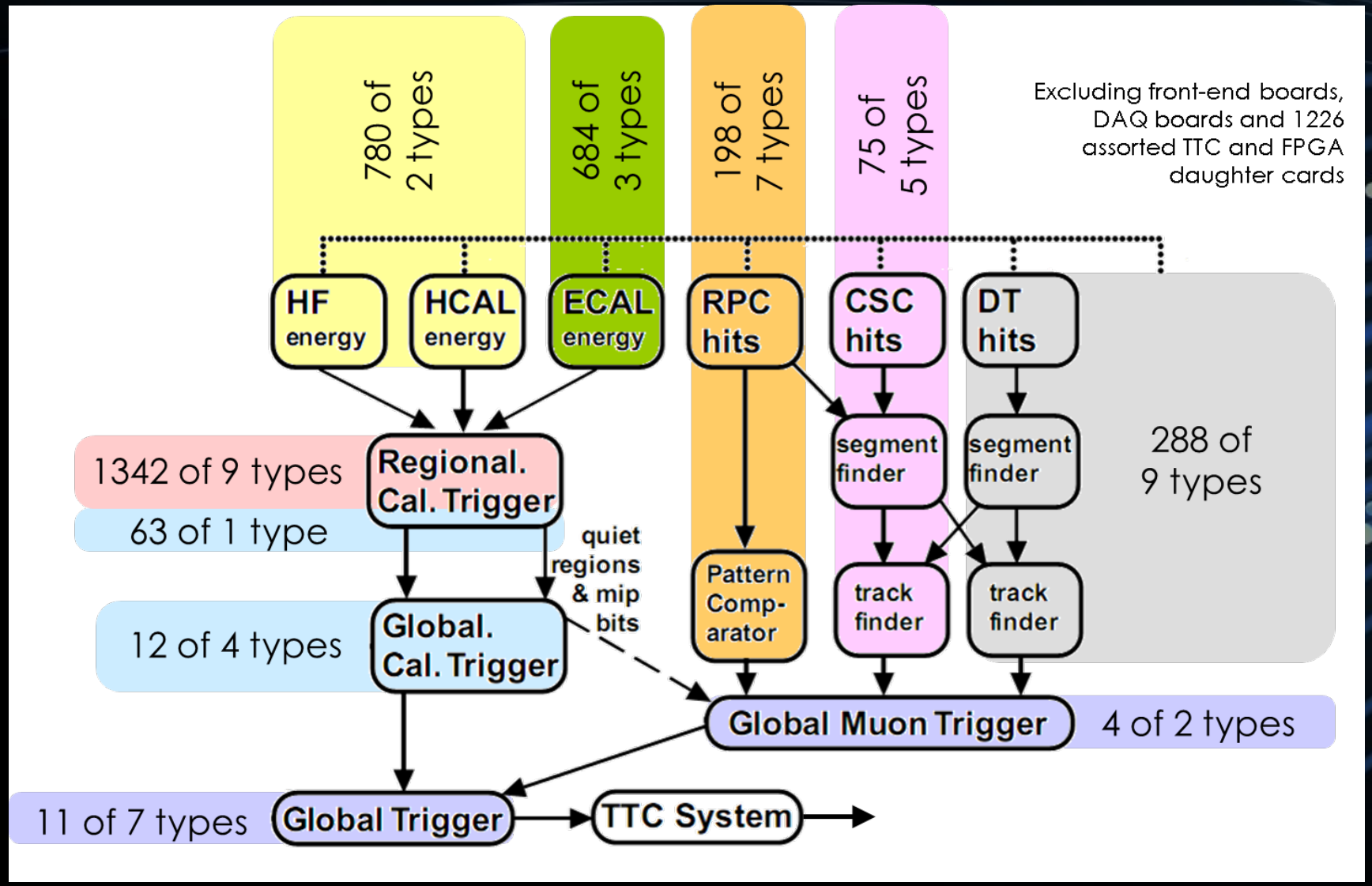
CMS: Visualizing the big numbers



CMS: Visualizing the big numbers

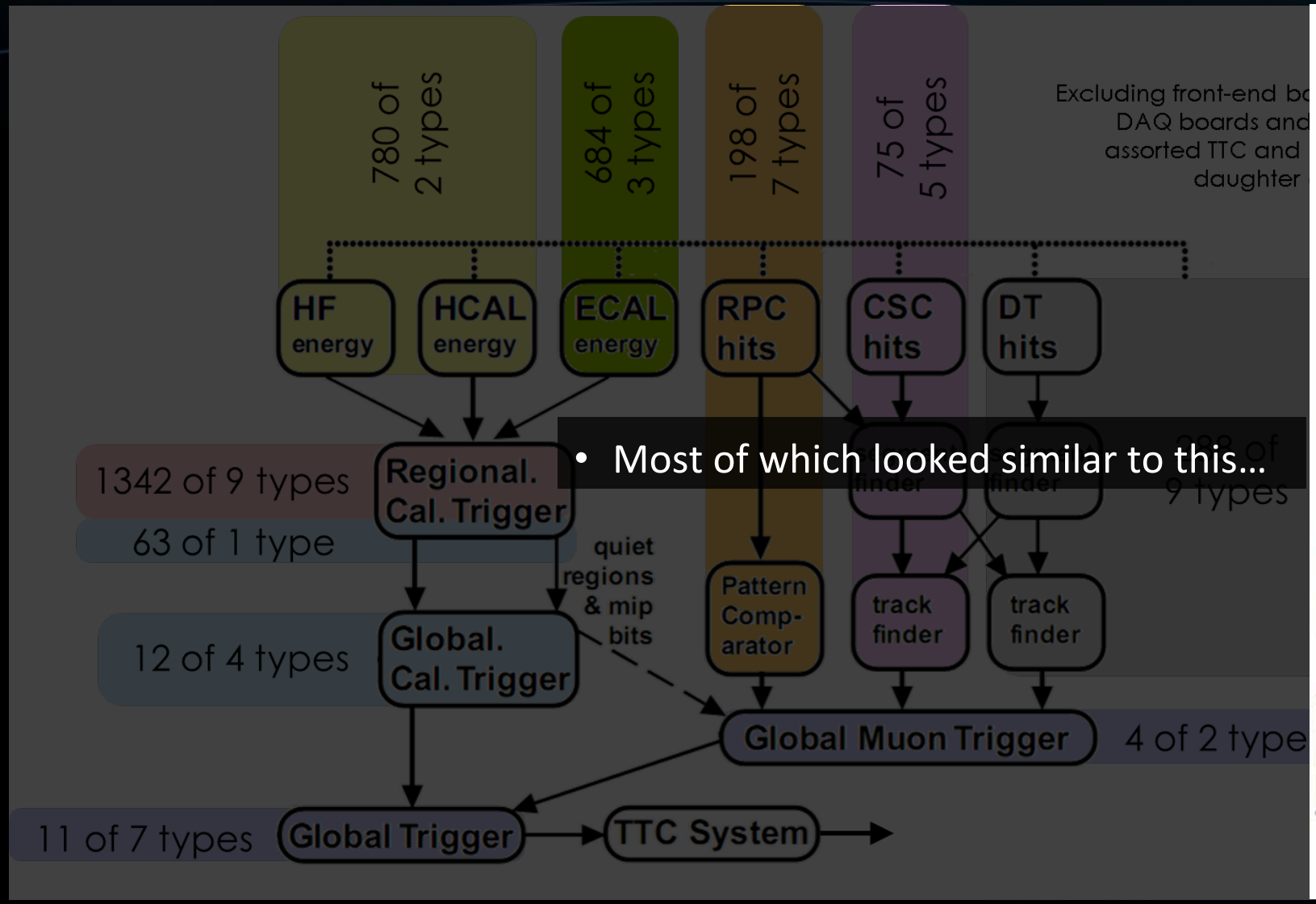


Runs 0 & 1: How we processed data

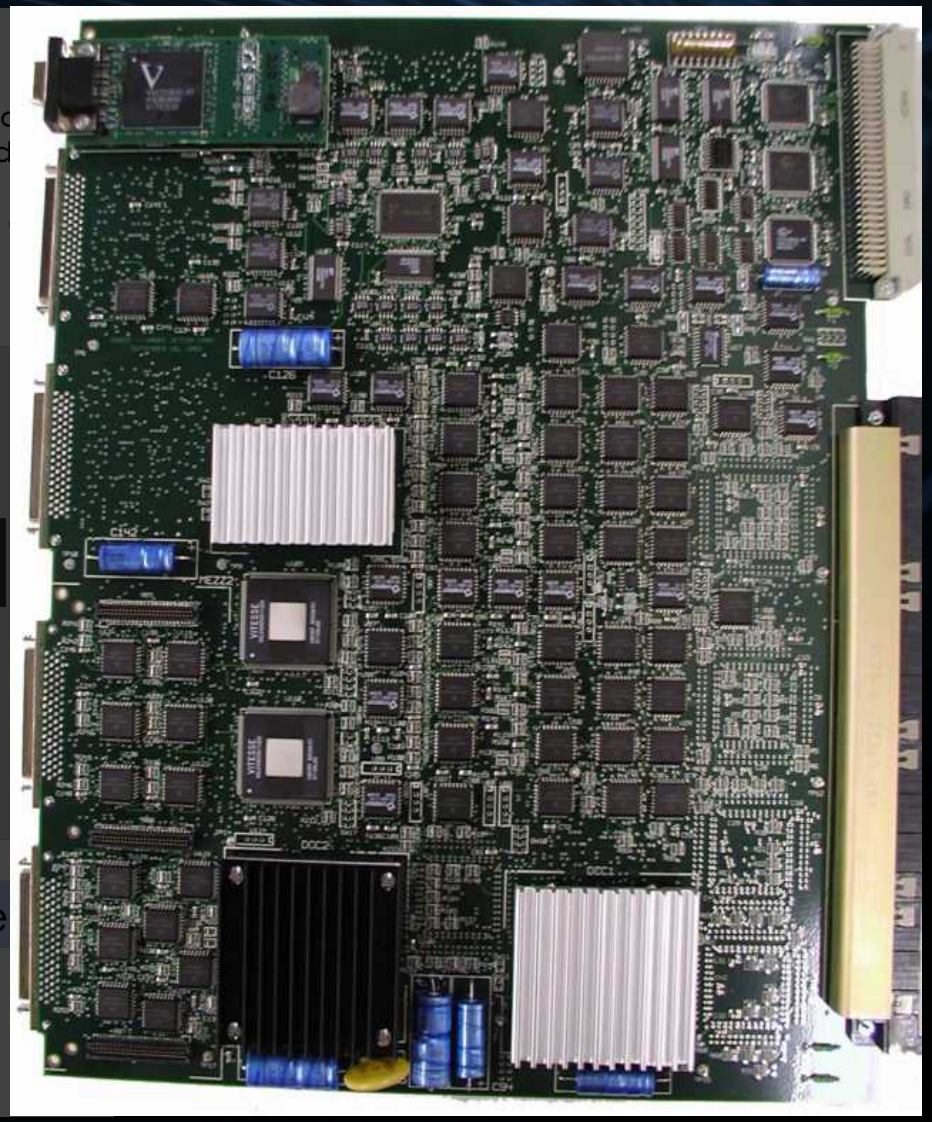


- Many boards of many different types

Runs 0 & 1: How we processed data

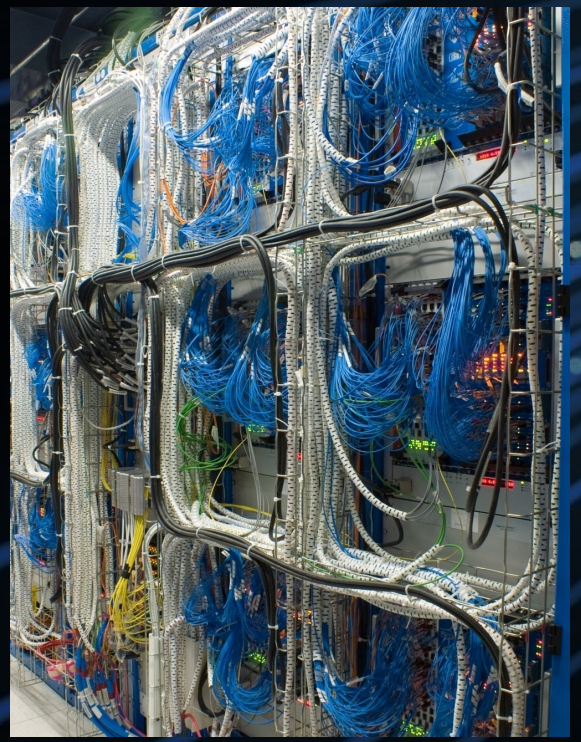


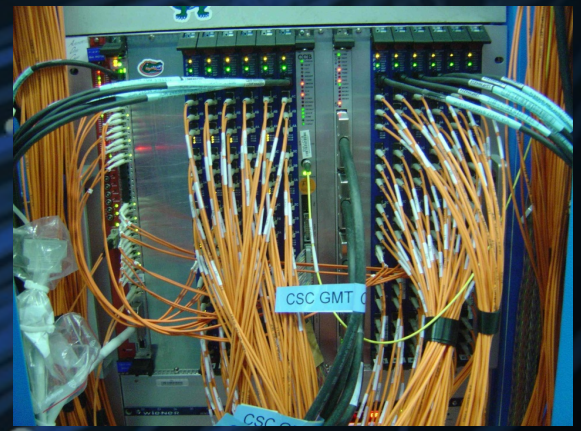
- Most of which looked similar to this…



Runs 0 & 1: How we processed data

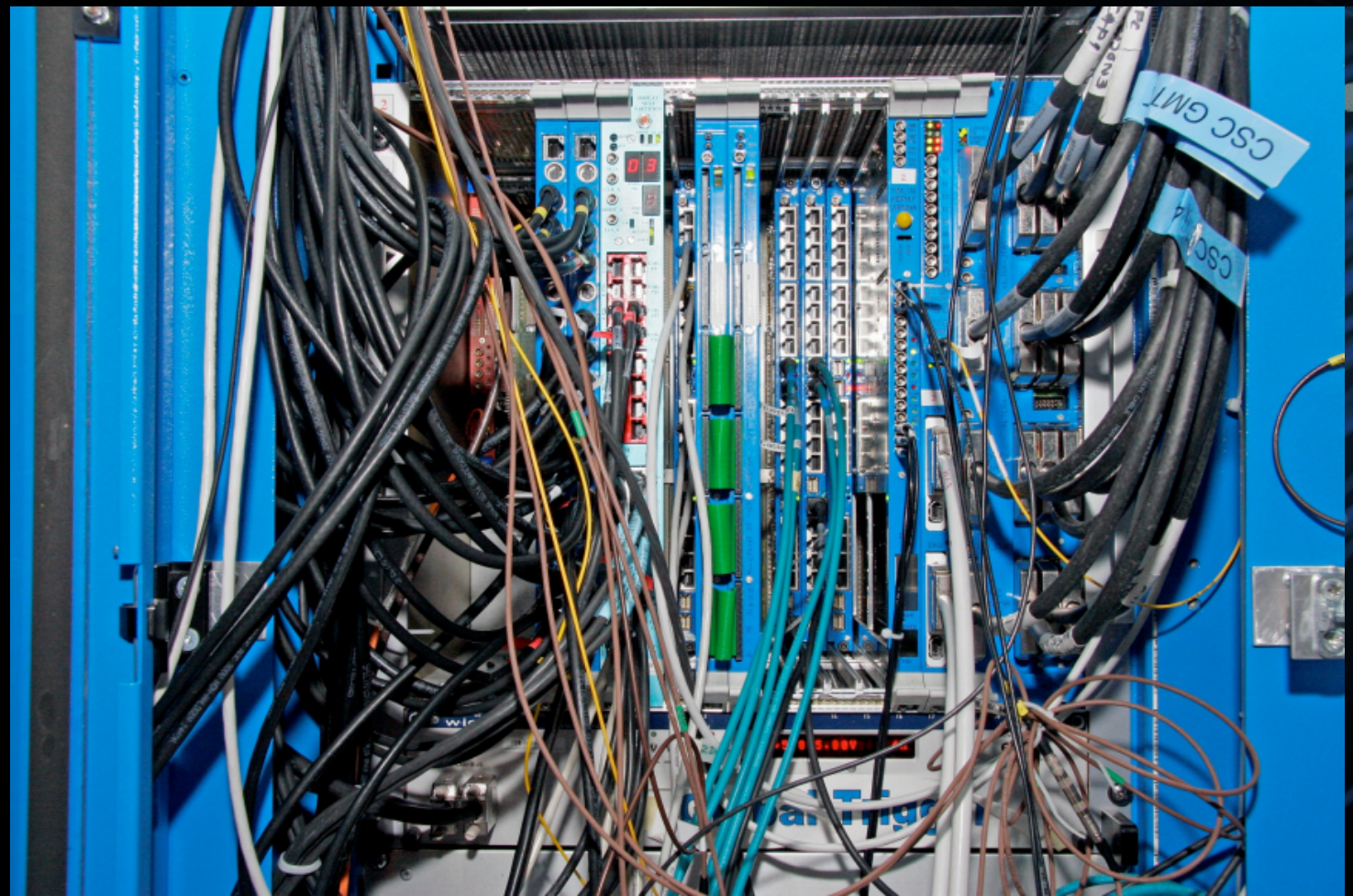
- In systems that looked like this…





Runs 0 & 1: How we processed data

- And like this…



Runs 0 & 1: How we processed data

- But sometimes it's OK to be hideous if it works

Runs 0 & 1: How we processed data

- But sometimes it's OK to be hideous if it works

Who am I kidding? Such an opinion is for Philistines. We should always be aiming to produce something <u>OF BEAUTY</u>

The Prophecy and the Vision

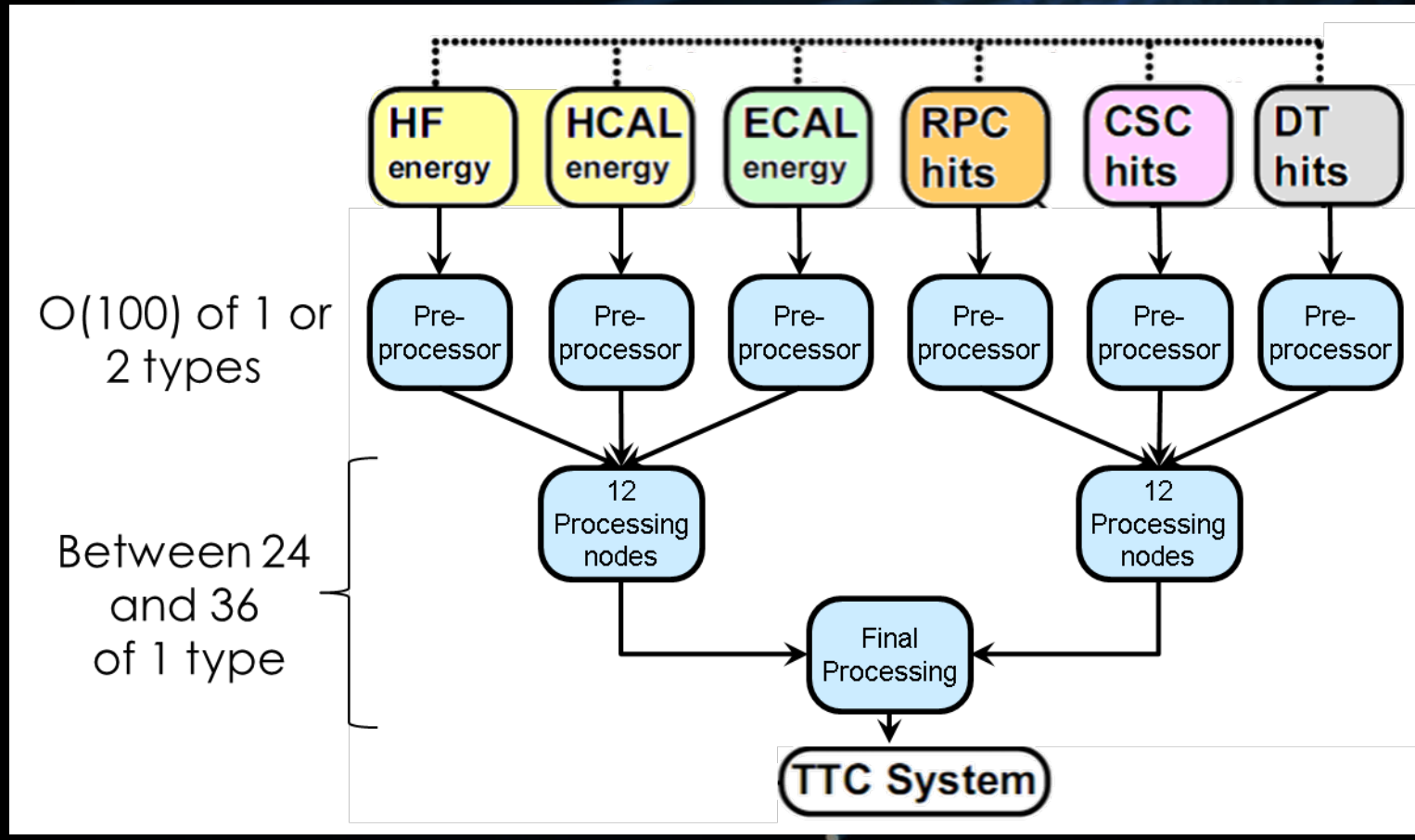
The Prophetic Witticism…

By 2015, every board in the [CMS] trigger will be identical and, after that, they will only get more similar

Andy Rose, 2010

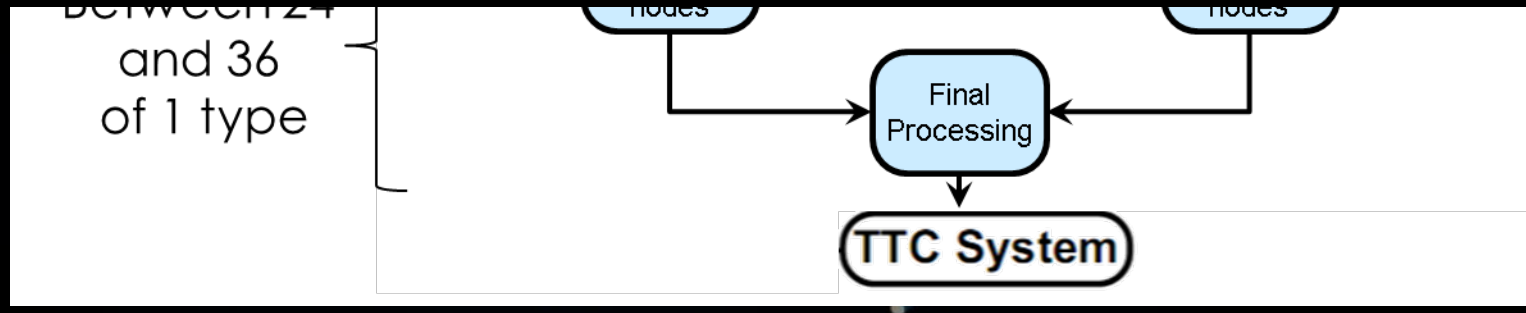
… which accompanied “The Vision”

- In 2012, I presented my glorious vision:
- Using time-multiplexing to achieve homogeneity by eliminating boundaries
- A beautiful, highly homogeneous system
  - Common hardware
  - Common link-standards
  - Common software for controlling everything



The Vision

One board to rule them all,
One link-standard to bind them,
One control-bus to rule them all,
And one software to mind them.



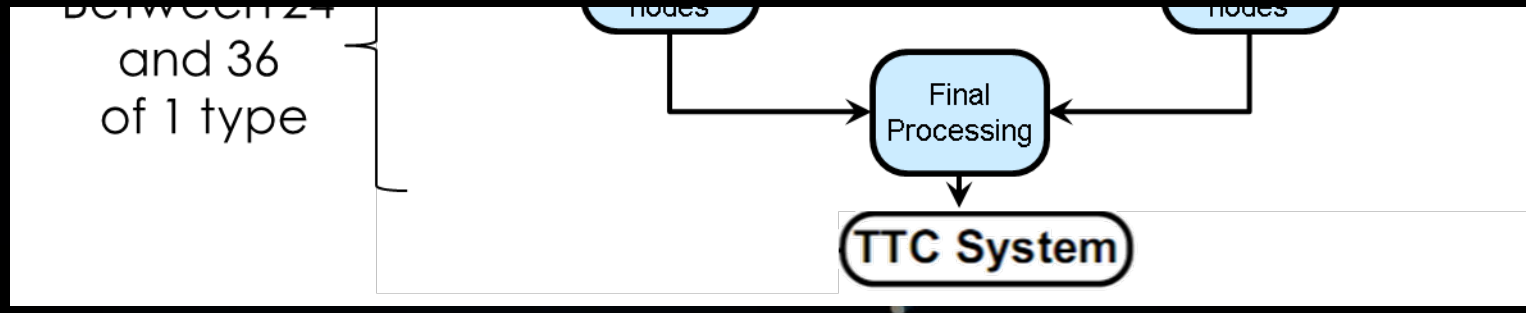
The Vision

One board to rule them all,

One link-standard to bind them,

One control-bus to rule them all,

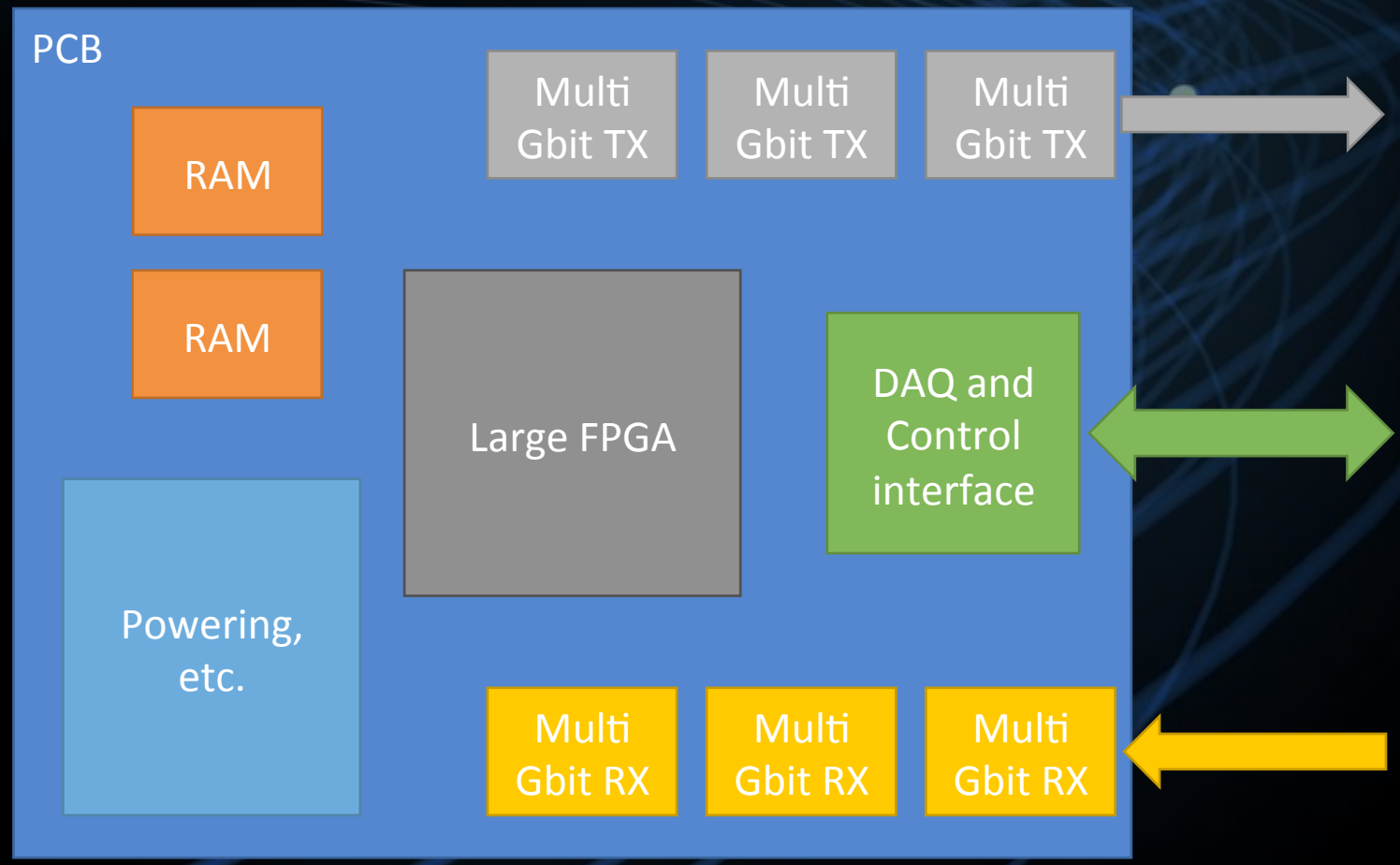
And one software to mind them.



The Prophetic Witticism

If you standardize the links and eliminate system-specific boundaries, then all boards

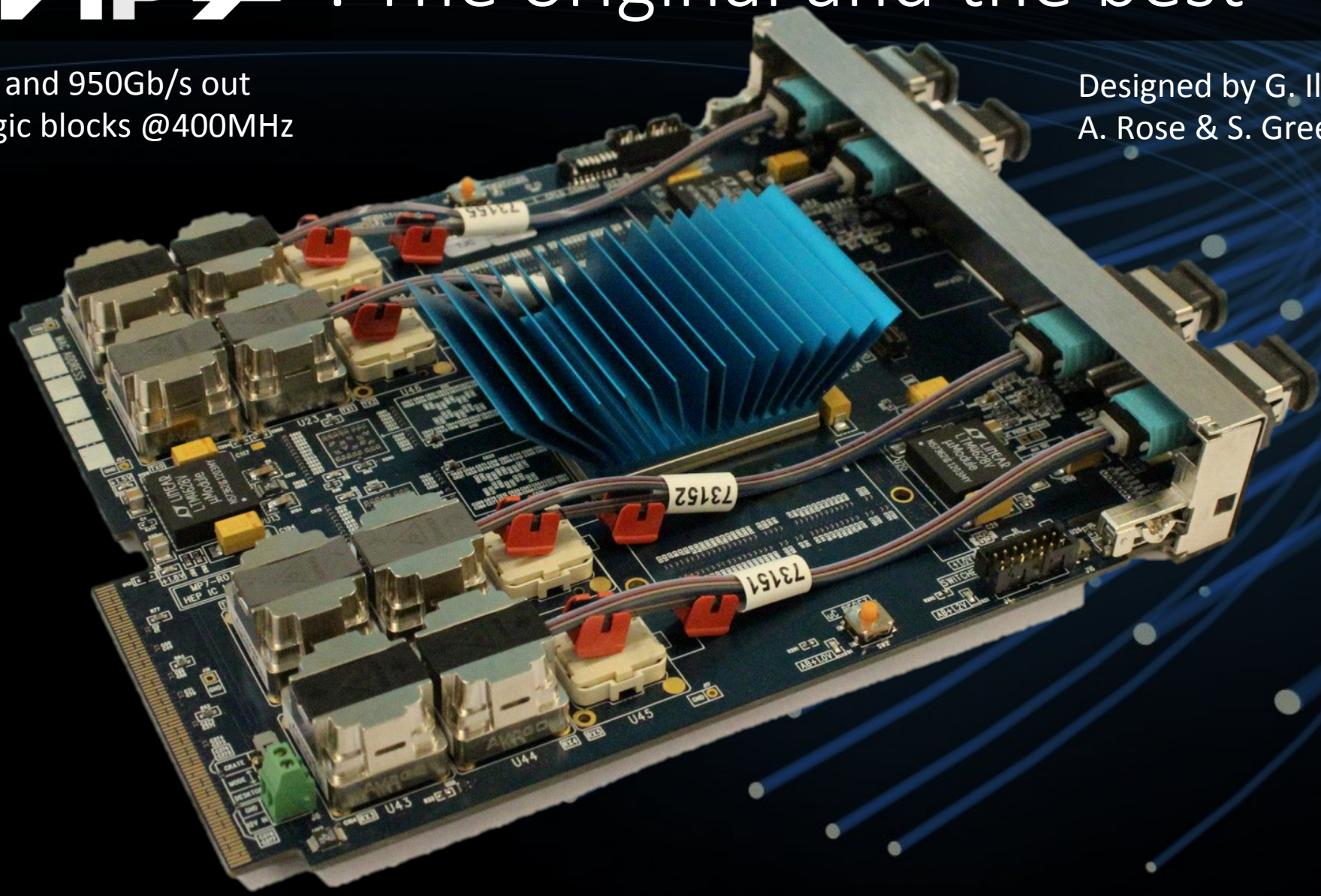
- receive parallel streams of data
- process parallel streams of data
- output parallel streams of data



MPZ : The original and the best

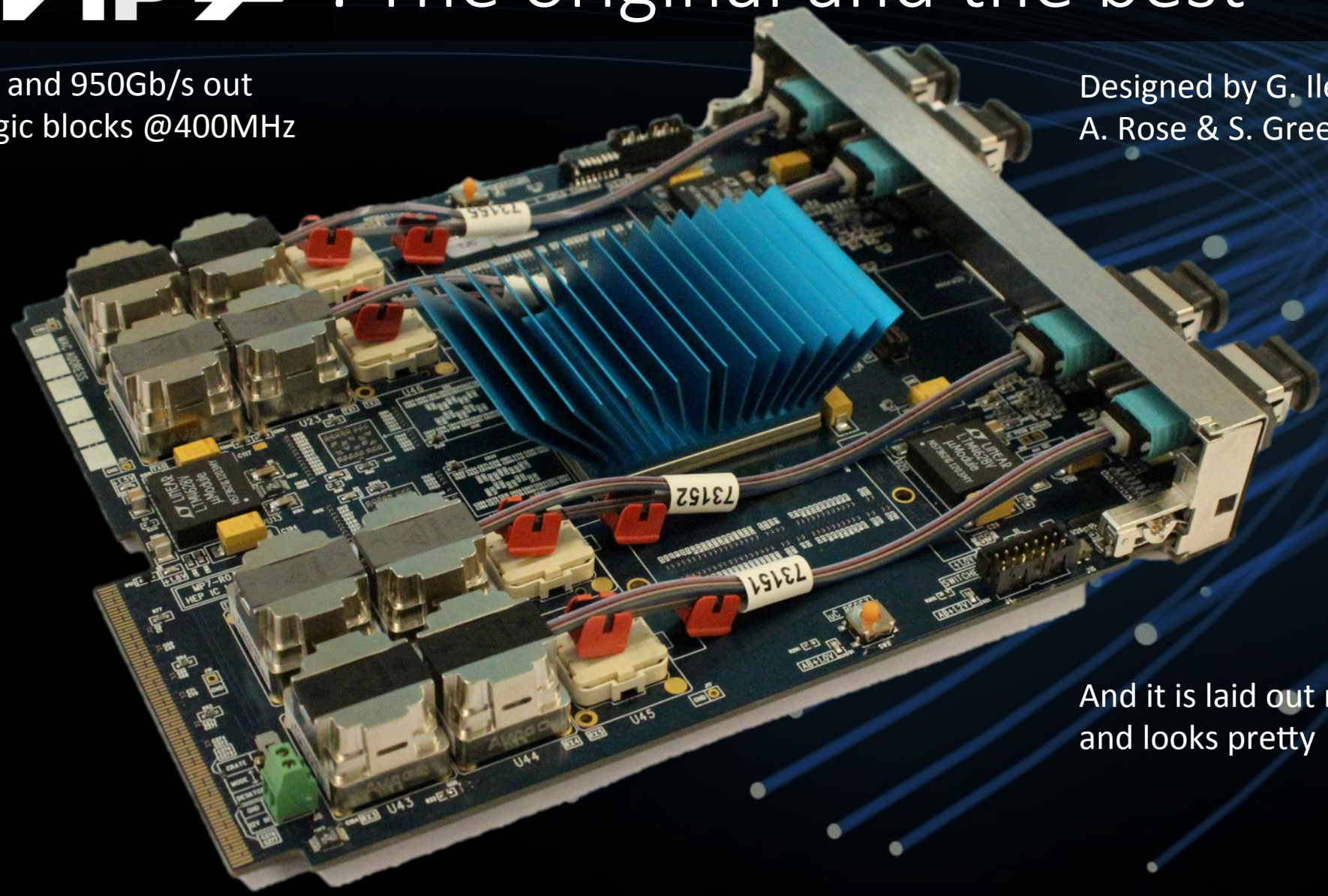
950Gb/s in and 950Gb/s out
700,000 logic blocks @400MHz

Designed by G. Iles, J. Jones,
A. Rose & S. Greenwood at IC



MPZ : The original and the best

950Gb/s in and 950Gb/s out
700,000 logic blocks @400MHz



Designed by G. Iles, J. Jones, A. Rose & S. Greenwood at IC

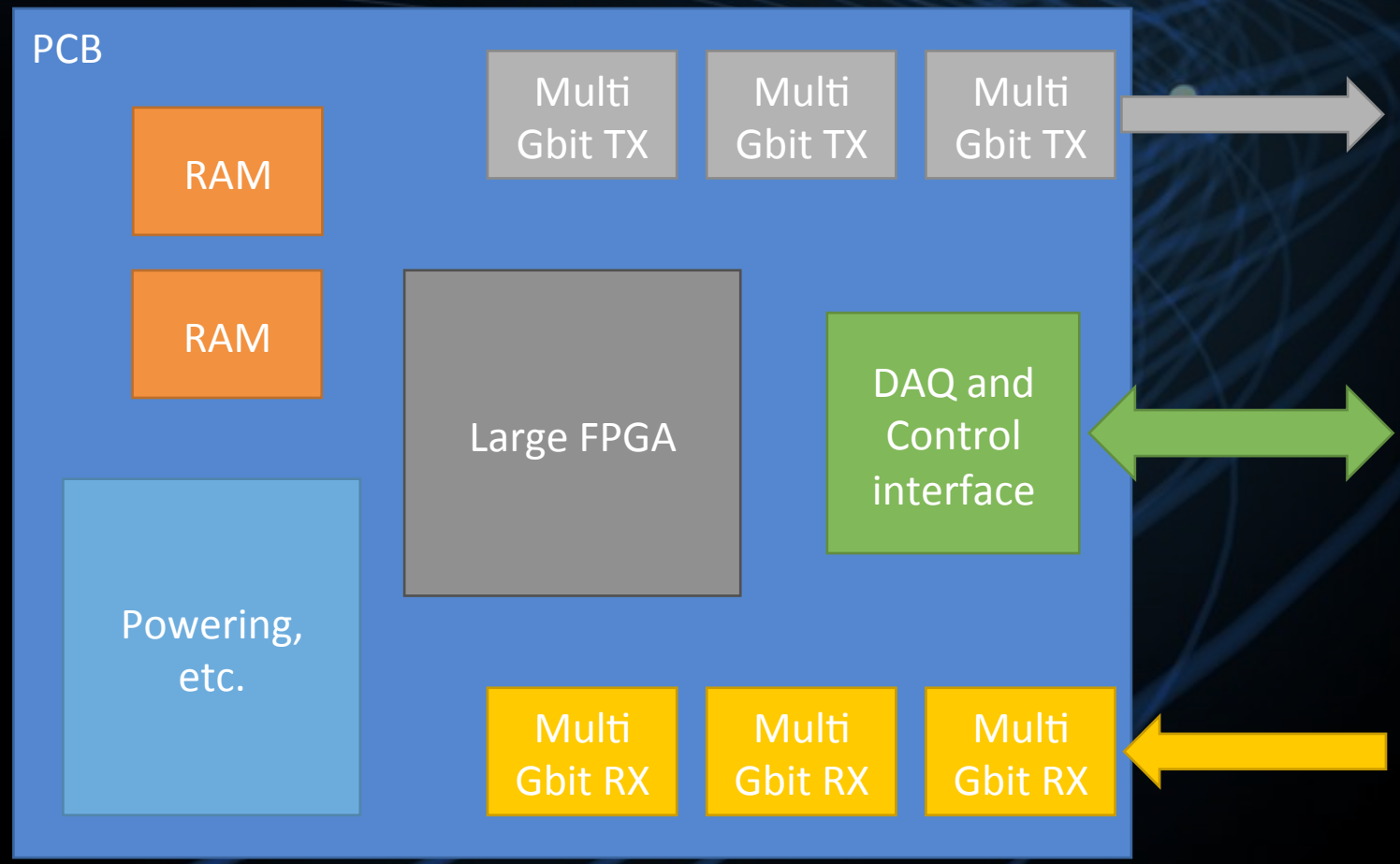
And it is laid out rationally and looks pretty

The Prophetic Witticism

Couple the limited variation in the tasks with the fact that:

- The reign of ASICs was over: Long live the FPGA
- The number of vendors making sufficiently powerful FPGAs was limited
- The number of vendors making sufficiently fast optics was also limited

And the logic behind the prophecy becomes obvious

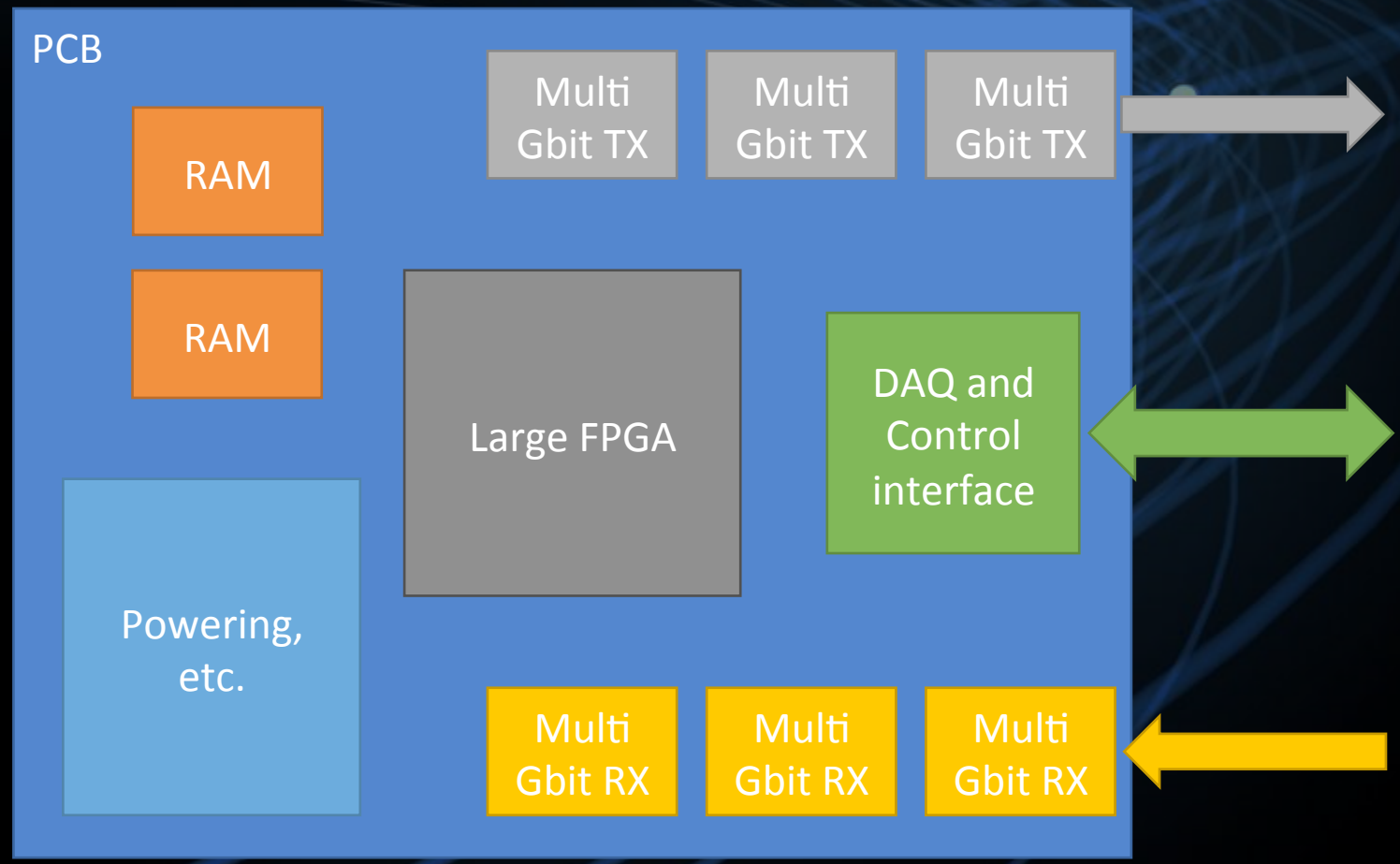


The Prophetic Witticism

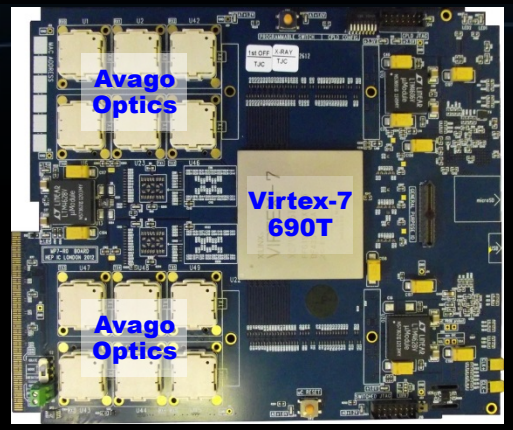
Couple the limited variation in the tasks with the fact that:

- The reign of ASICs was over: Long live the FPGA
- The number of vendors making sufficiently powerful FPGAs was limited
- The number of vendors making sufficiently fast optics was also limited

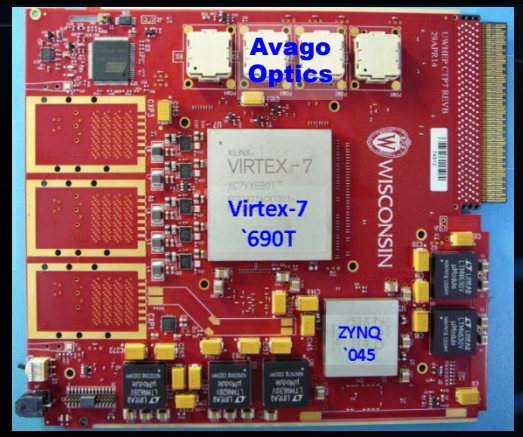
And the logic behind the prophecy becomes obvious, and so the prophecy came to pass…



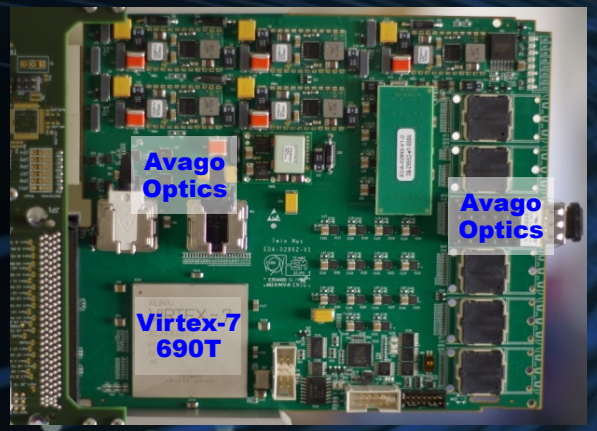
The Prophetic Witticism



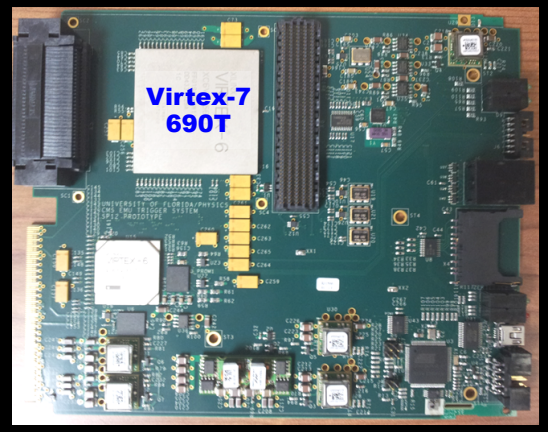
MP7

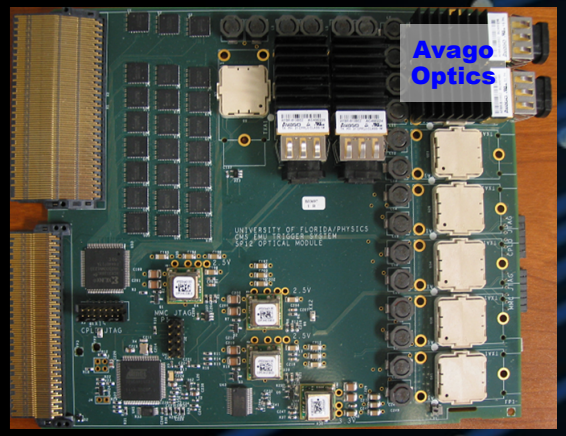


CTP7

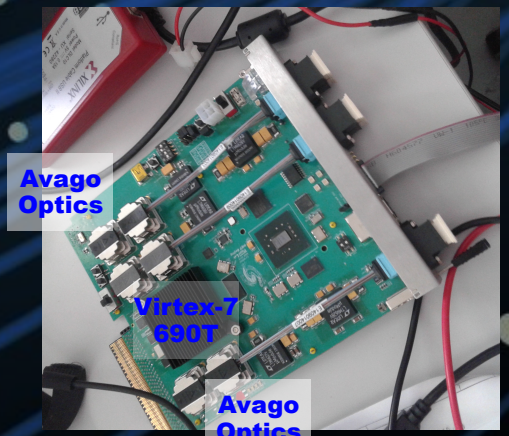


TwinMux





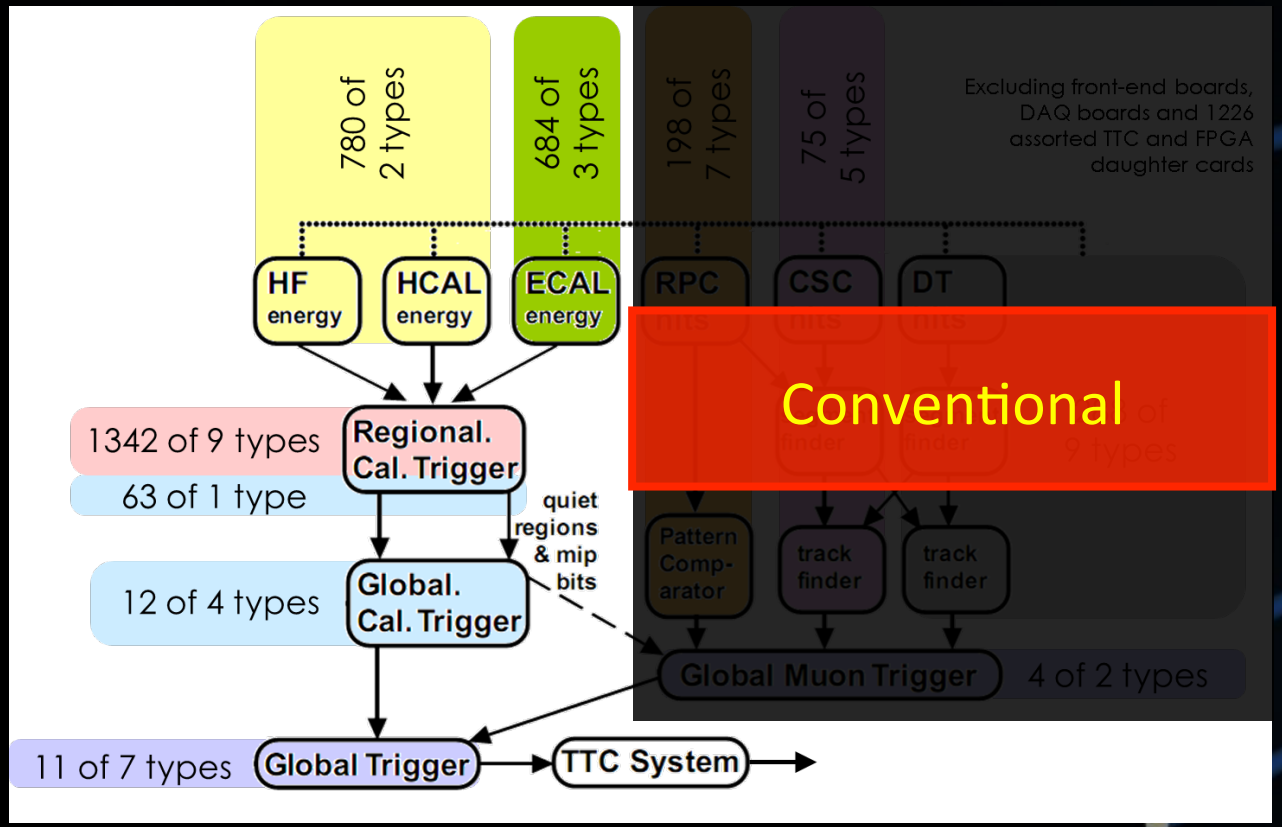
MTF7 (MTF6 shown)



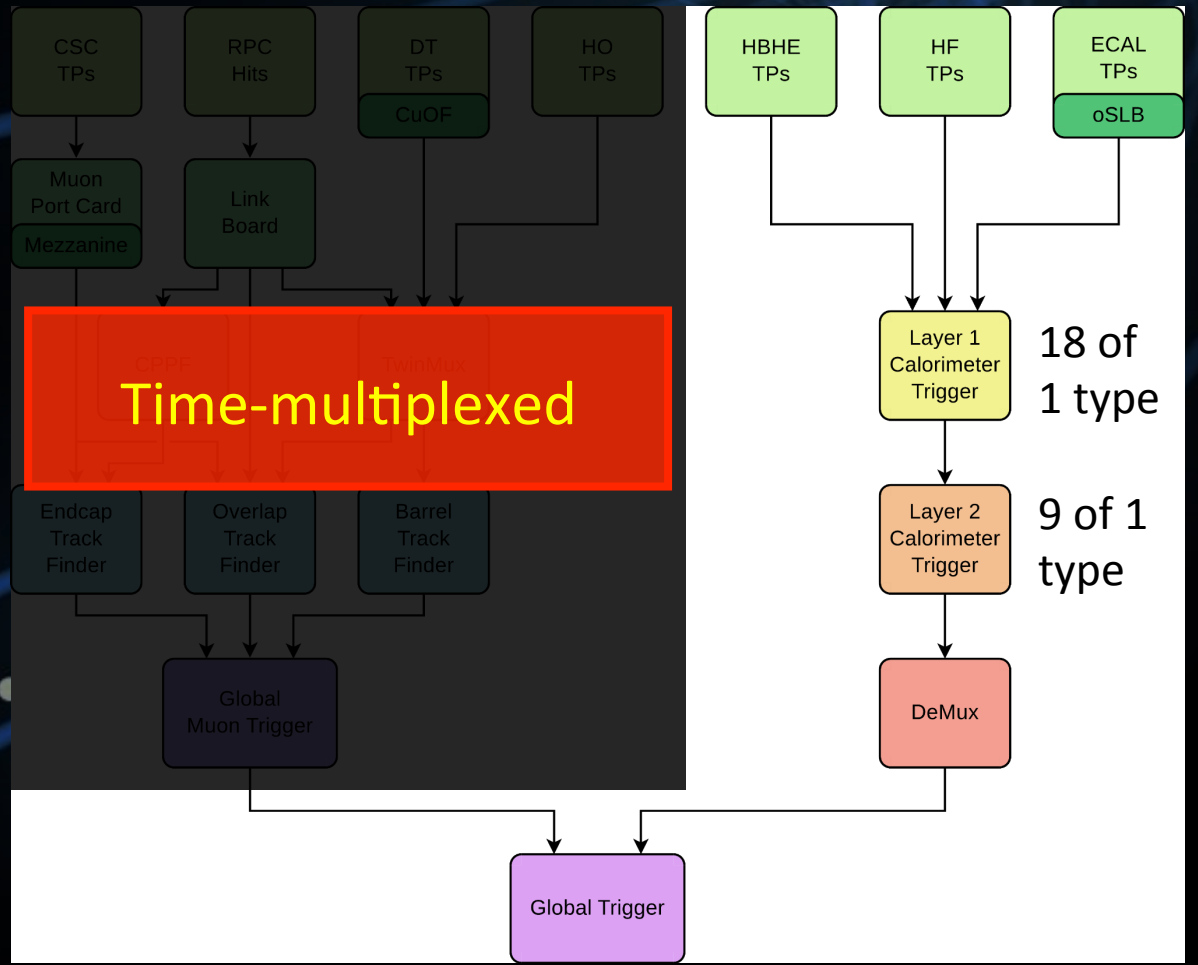
CPPF

The Vision vs. Reality

- Runs 0 & 1

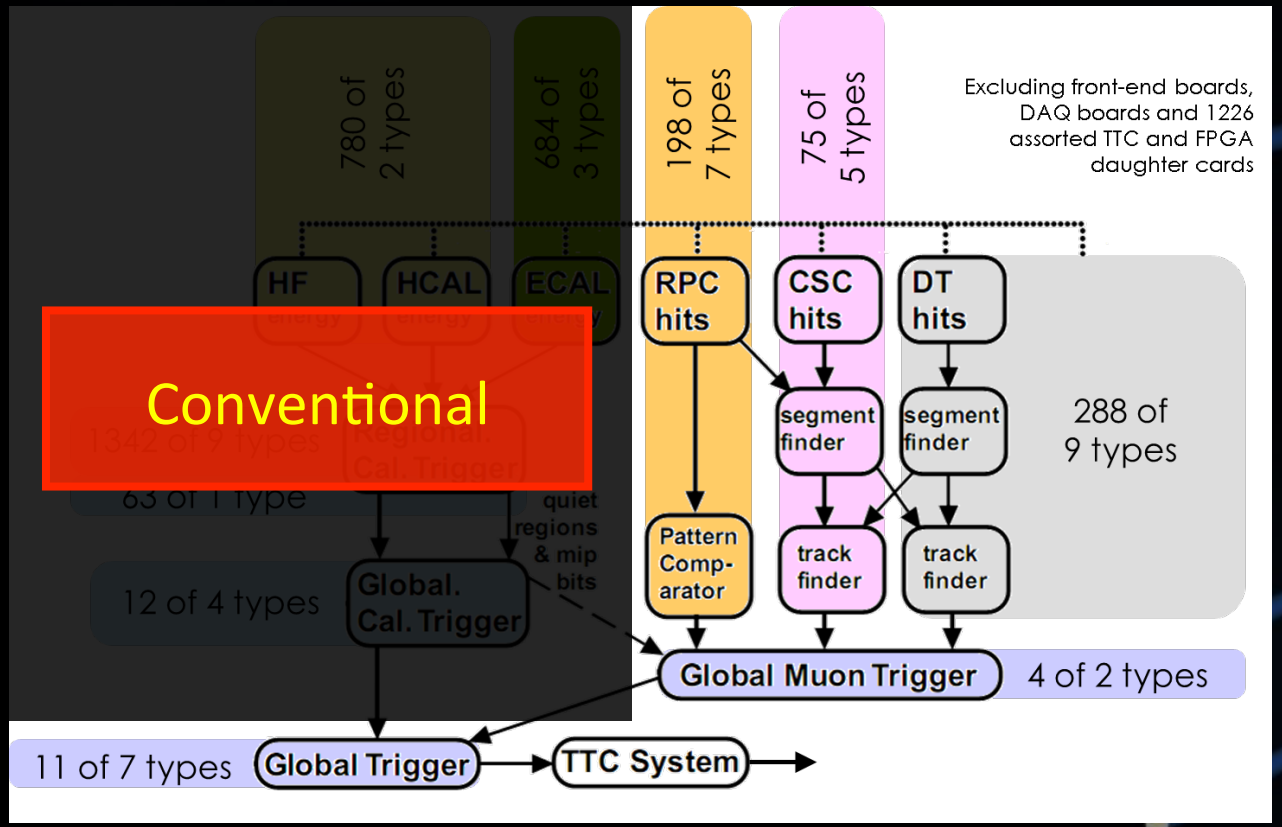


- Run 2



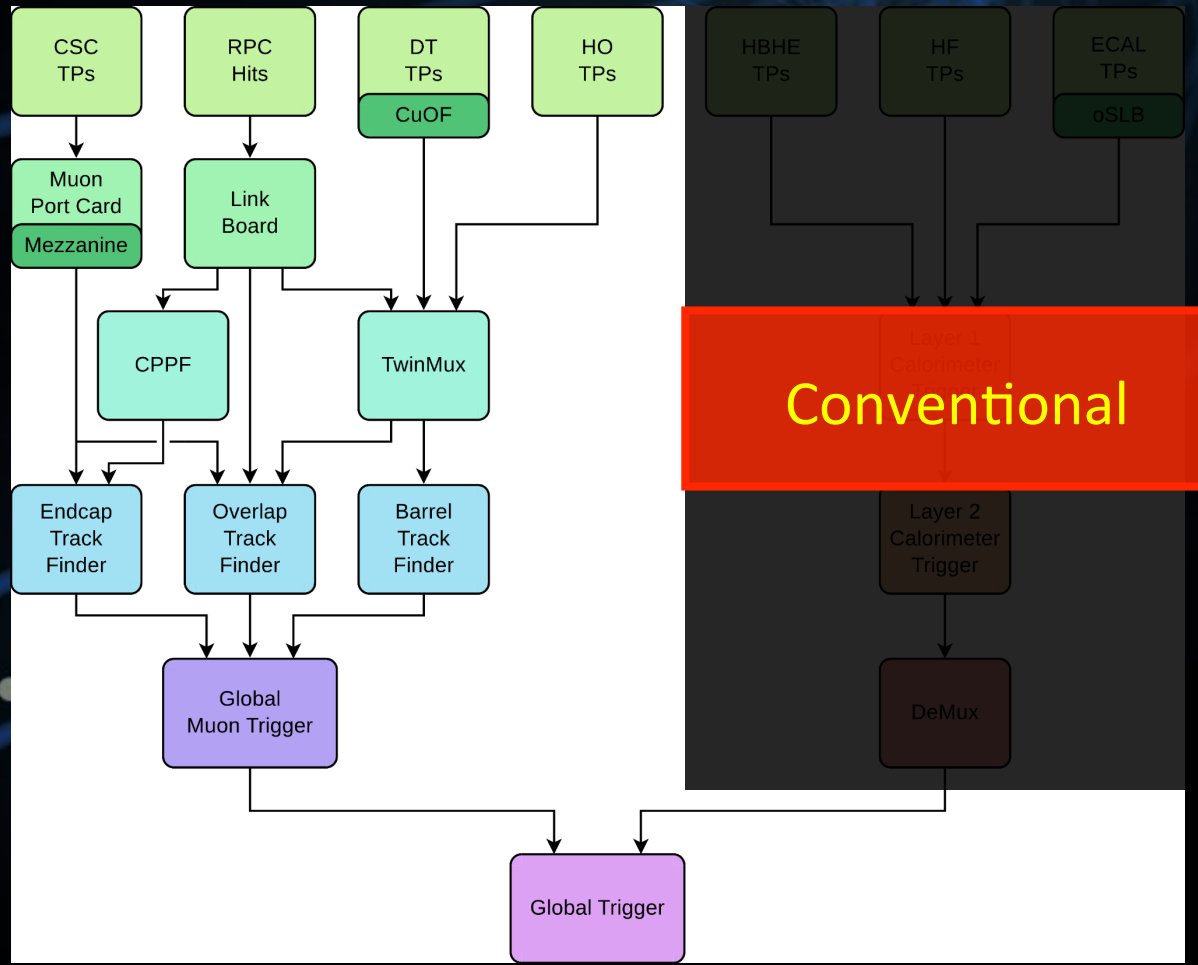
The Vision vs. Reality

- Runs 0 & 1



Conventional

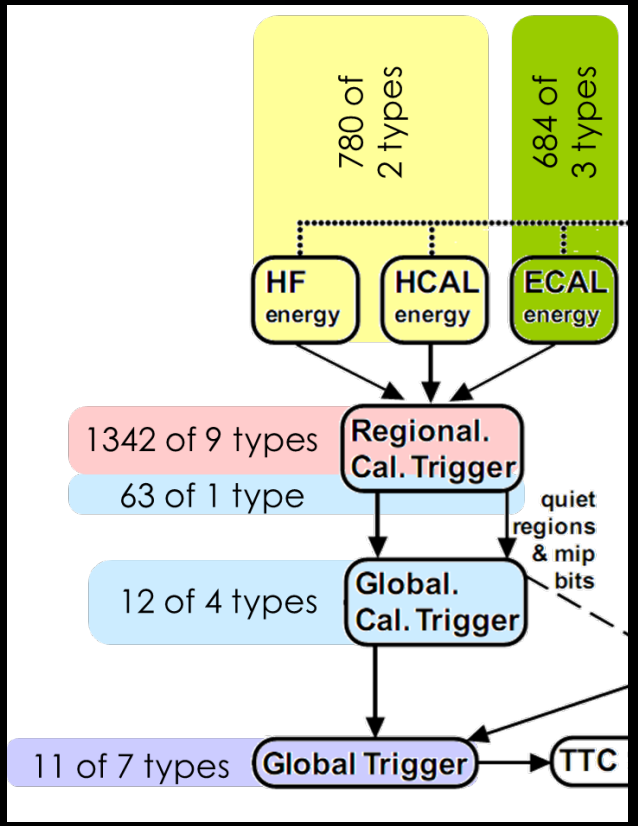
- Run 2

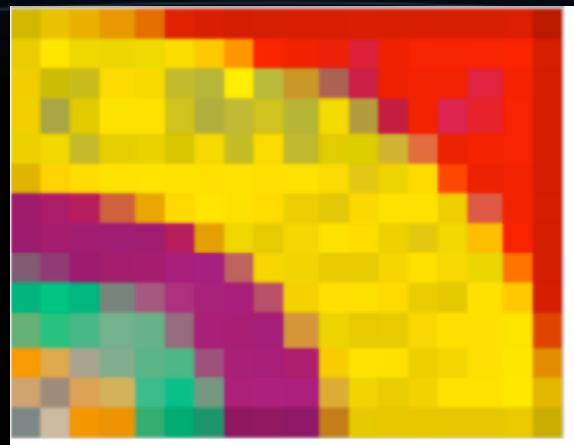


Conventional

The Vision vs. Reality

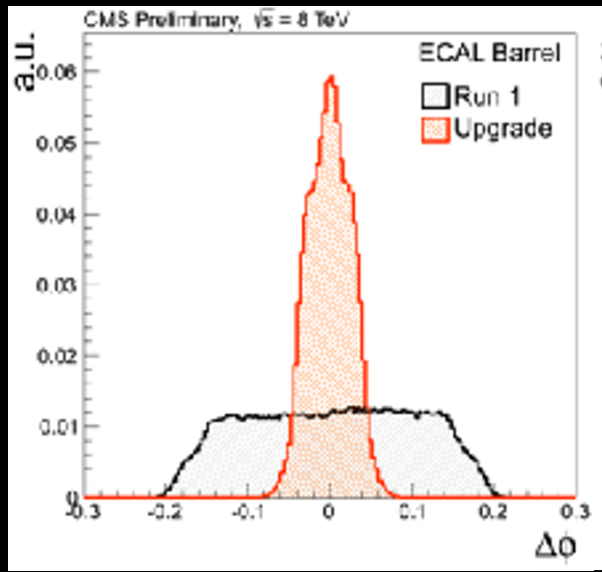
- Runs 0 & 1

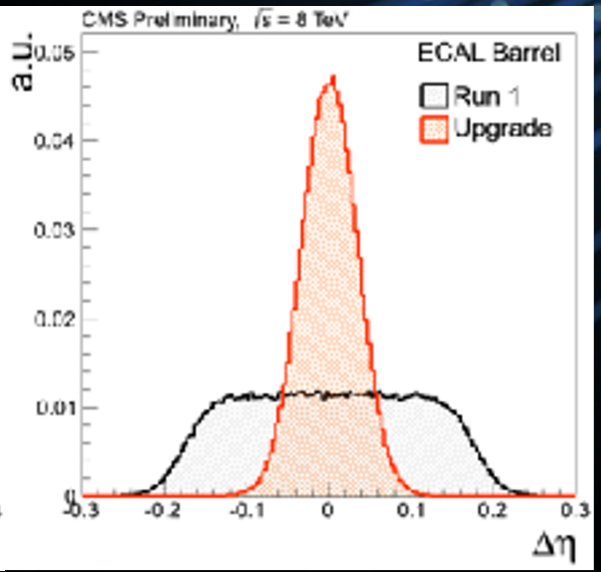




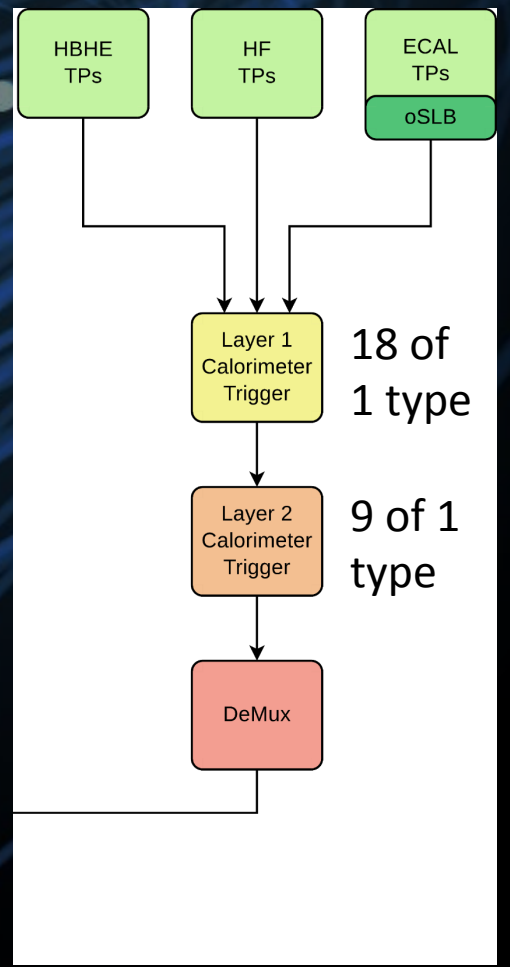
14 (η) x 18 (φ)

56 (η) x 72 (φ)

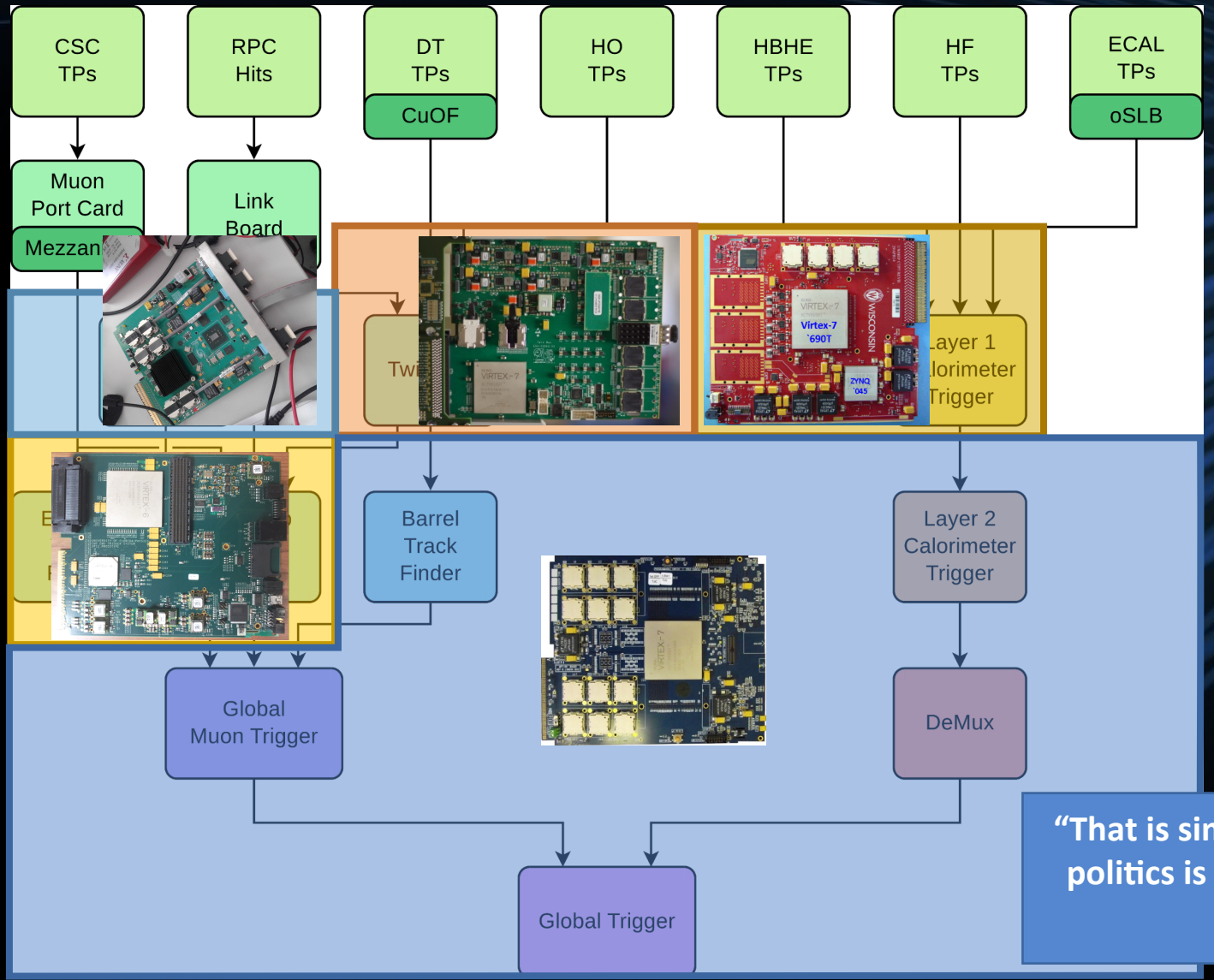




- Run 2



The Vision vs. Reality: Definitely better

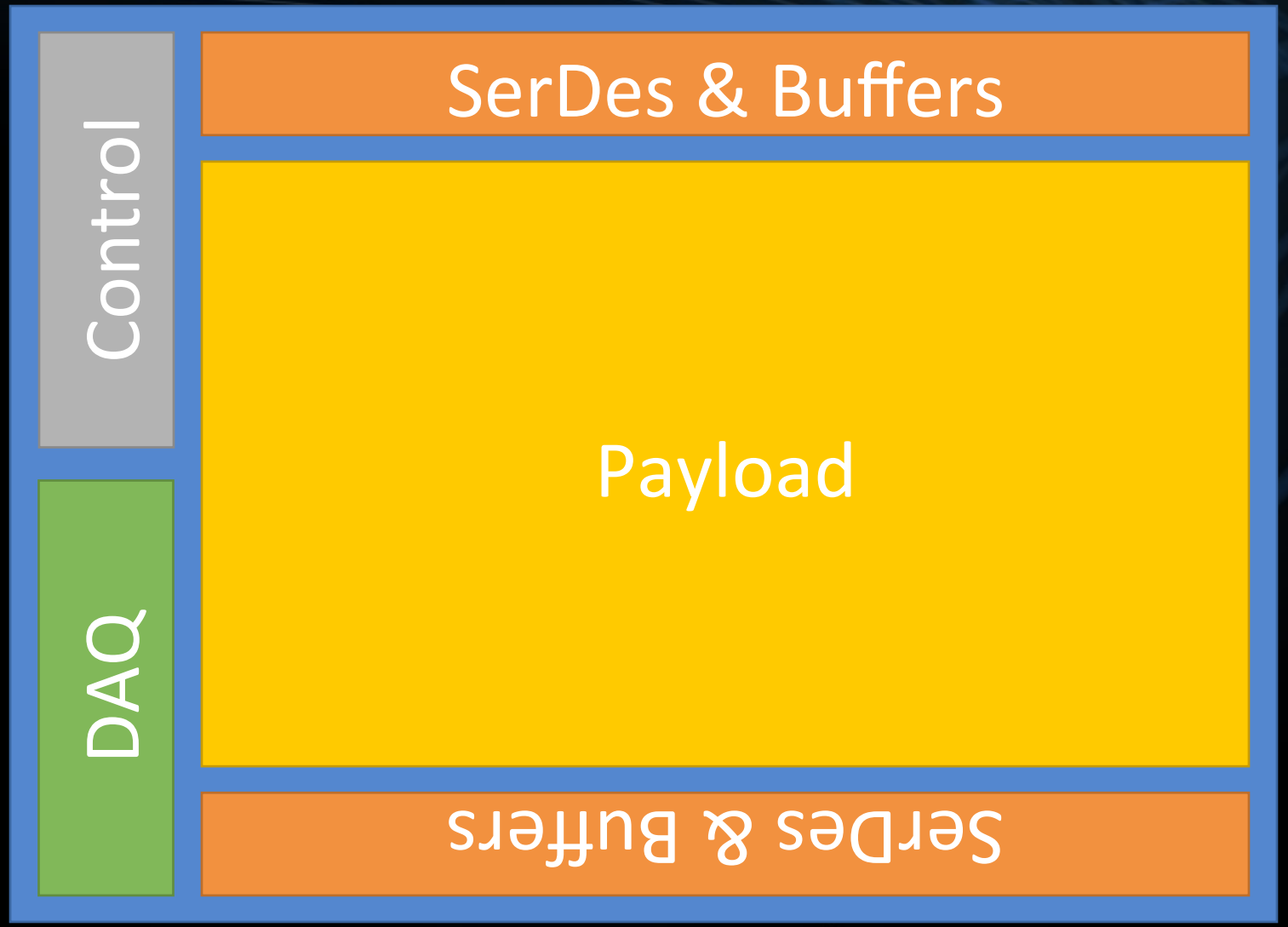


"That is simple, my friend. It is because politics is more difficult than physics."

Albert Einstein 1946

An aside: The Firmware Operating System

- MP7 achieved this by
  - Separating the payload from the infrastructure
  - Standardizing the interfaces
  - Producing a build-tool to allow anyone to build their design in "unfamiliar hardware"
- A victory for standardization!



The Vision vs. Reality: Much, much better!

The Vision vs. Reality: Ever so nearly perfect!

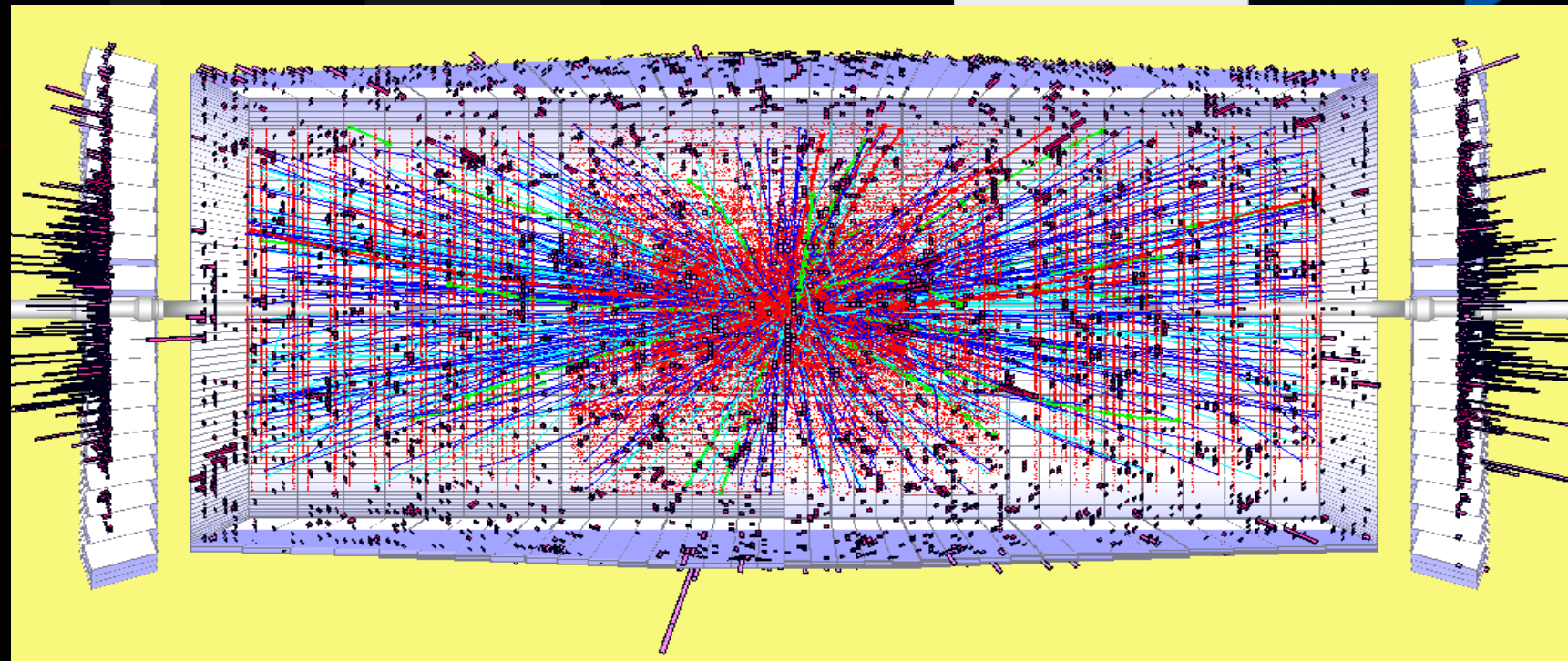
- IPbus designed out the US
  - Picked up at IC and bugs fixed
  - V2 produced at IC/RAL/Bristol
  - Adopted as standard across CMS
  - Then in the LHC accelerator group, ATLAS, LHCb, Alice, XFEL, SoLid, ProtoDune…
- uHAL software library written at IC
  - Goes hand-in-hand with IPbus
- Hurrah! A major victory for standardization!

Phase 1: Conclusion

- Time-multiplexing works
  - Has been shown to offers all the benefits we said it would!
  - Now well established in the CMS mindset
- The prophecy that all off-detector electronics would be “identical” in 2015 came to pass
- The run-2 trigger was certainly more elegant and robust than runs 0 & 1
- Many UK-originated ideas gained traction outside the UK:
  - Common hardware
  - Firmware operating system
  - Standardization of link protocols and control infrastructure


Phase 2: A Brave New World

Brave New World



The CMS Detector after Phase-II

Full granularity Ecal included in trigger

New CSC electronics
New RPC/GEM layers

New pixel layers
Tracking included in trigger

Complete replacement of endcap Ecal & Hcal
30ps timing resolution

Replace HPD with SiPM in Barrel HCAL
Depth info into trigger

MIP timing layer
30ps timing resolution

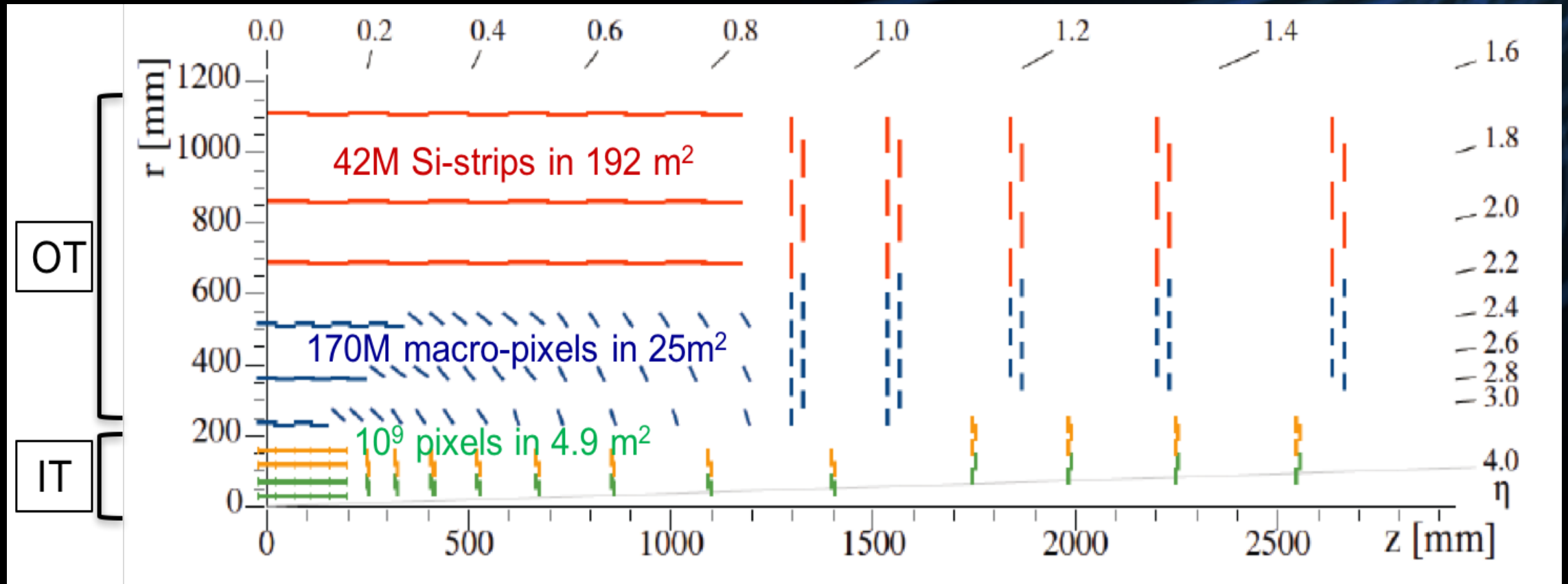
New Beampipe
Modified Magnet Yolk
…

What's the UK doing?

- UK has led, and continues to lead, the development of the Stacked-tracker concept and the tracking-trigger
  - UK delivering outer-tracker trigger and readout ASIC
  - UK delivering outer-tracker readout-electronics and developing track-finding algorithms

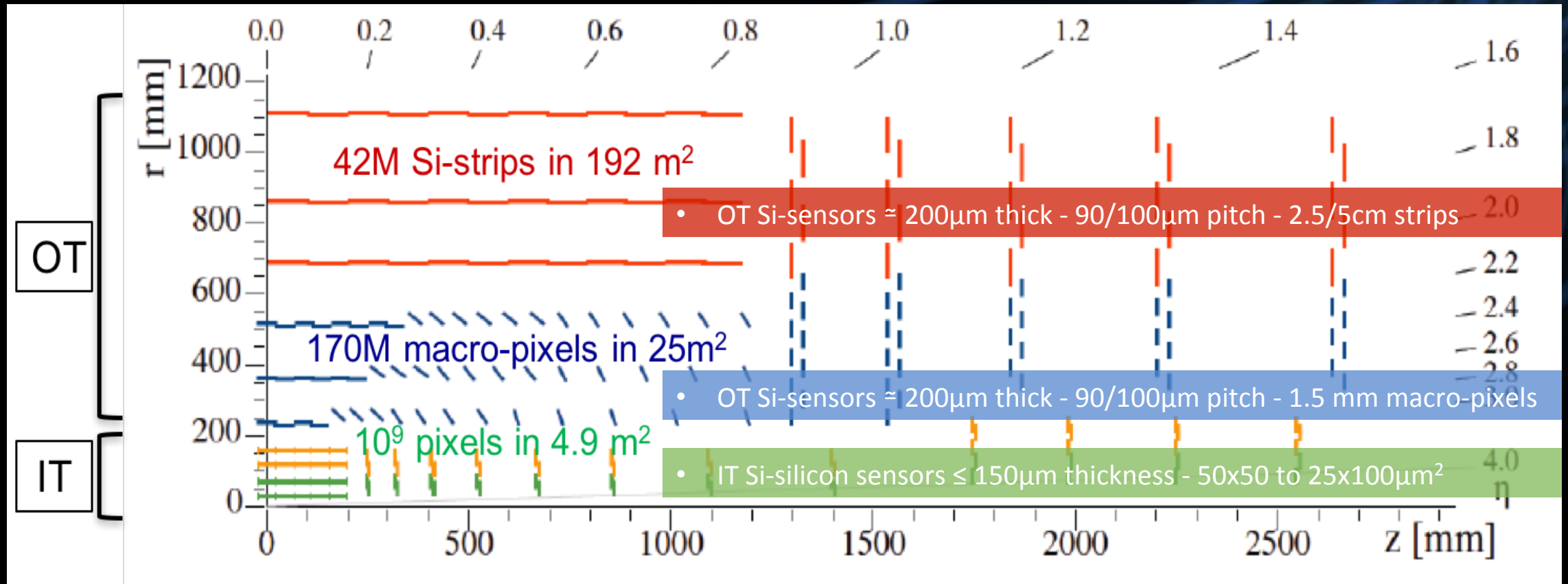
Tracker

- Inner Tracker (pixel) design to extend coverage to η ≃ 3.8



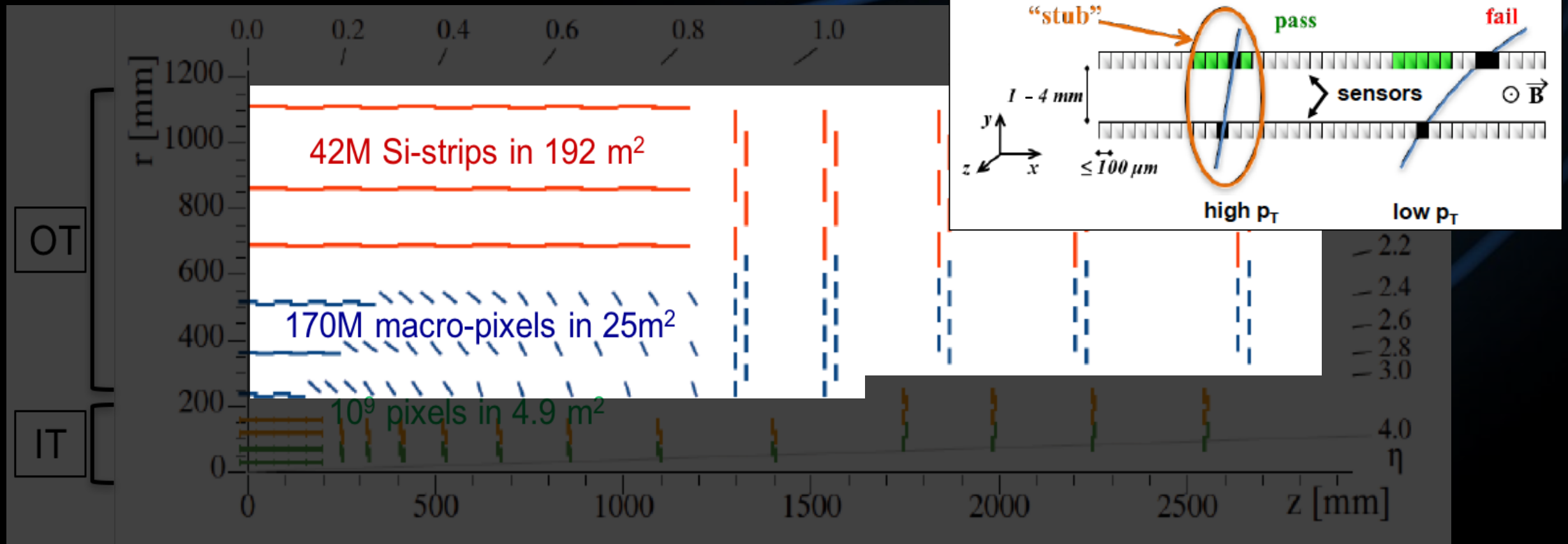
Tracker

- Inner Tracker (pixel) design to extend coverage to η ≃ 3.8

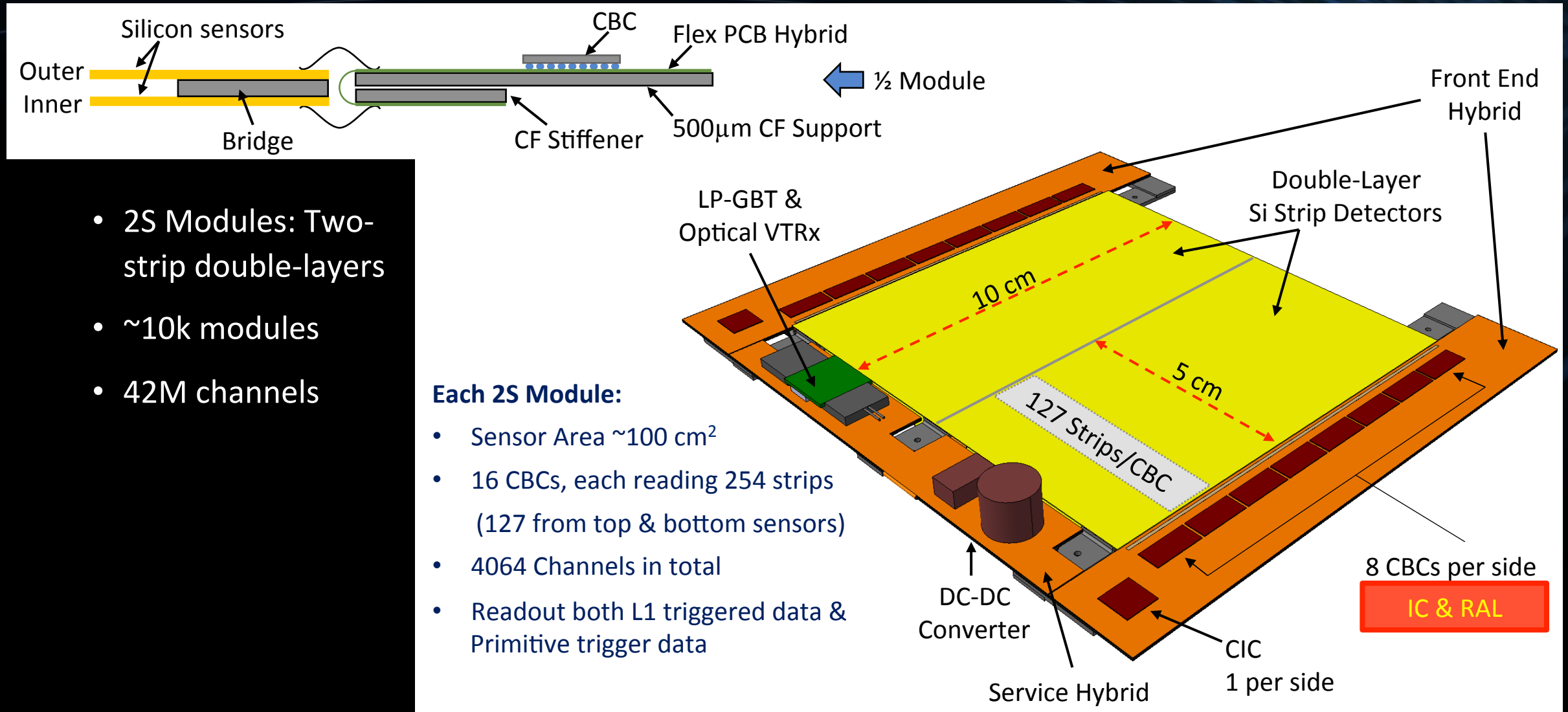


Tracker

- Inner Tracker (pixel) design to extend coverage to η ≃ 3.8
- Outer Tracker design driven by ability to provide tracks at 40 MHz to L1-trigger



Outer tracker 2S modules

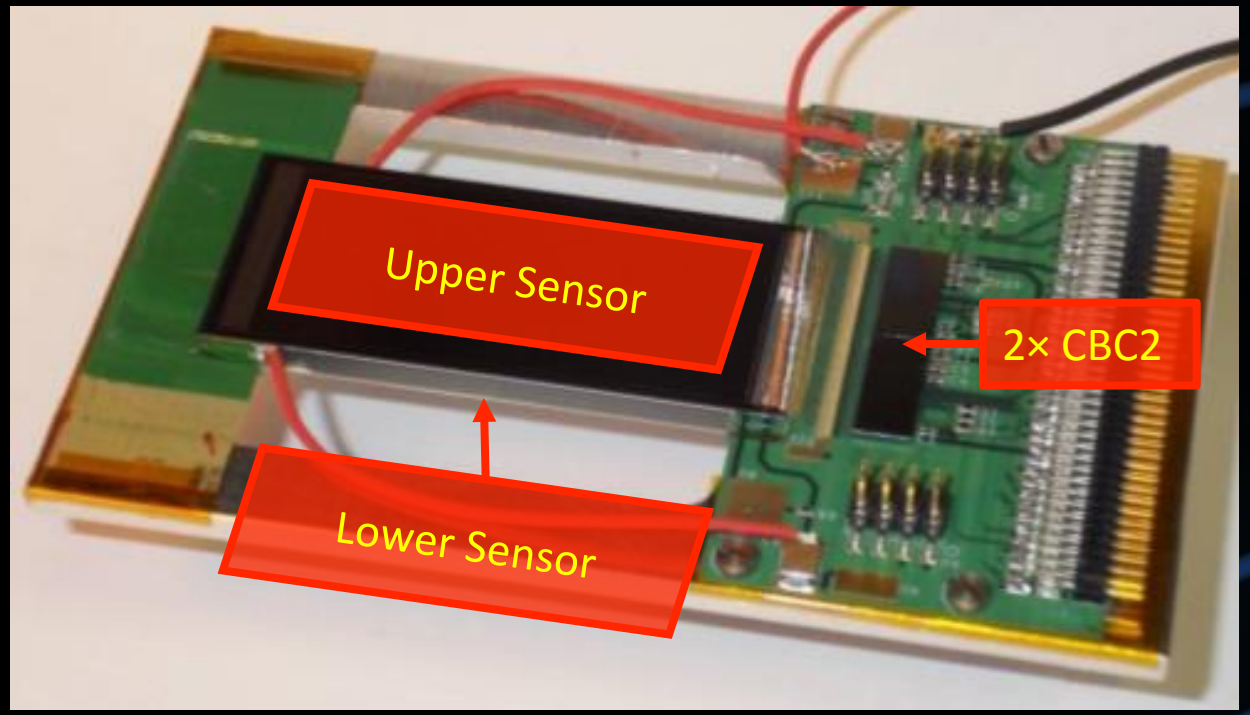


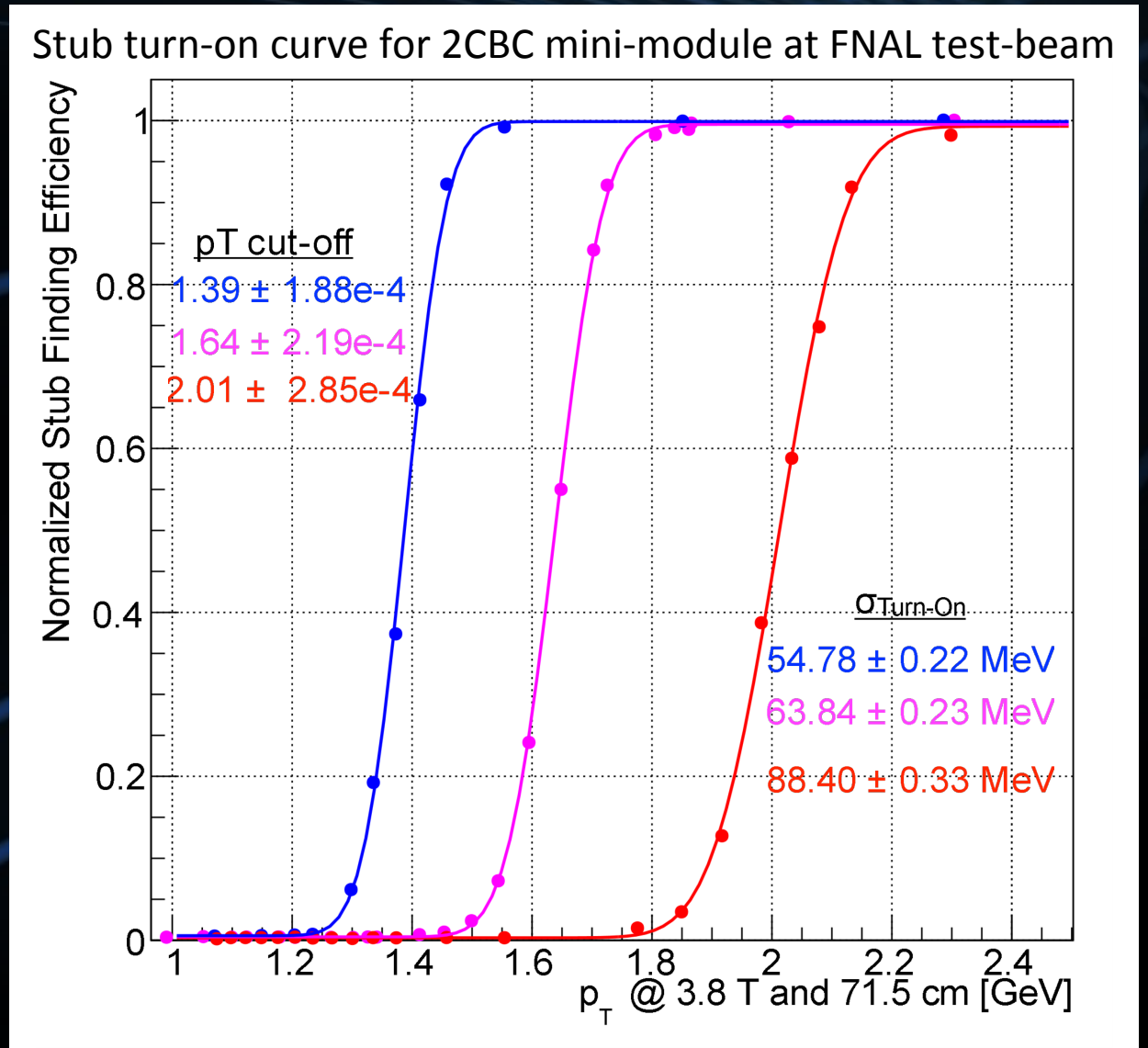
- 2S Modules: Two-strip double-layers
- ~10k modules
- 42M channels

Each 2S Module:

- Sensor Area ~100 cm^2
- 16 CBCs, each reading 254 strips (127 from top & bottom sensors)
- 4064 Channels in total
- Readout both L1 triggered data & Primitive trigger data

Outer tracker 2S modules: Do they work?



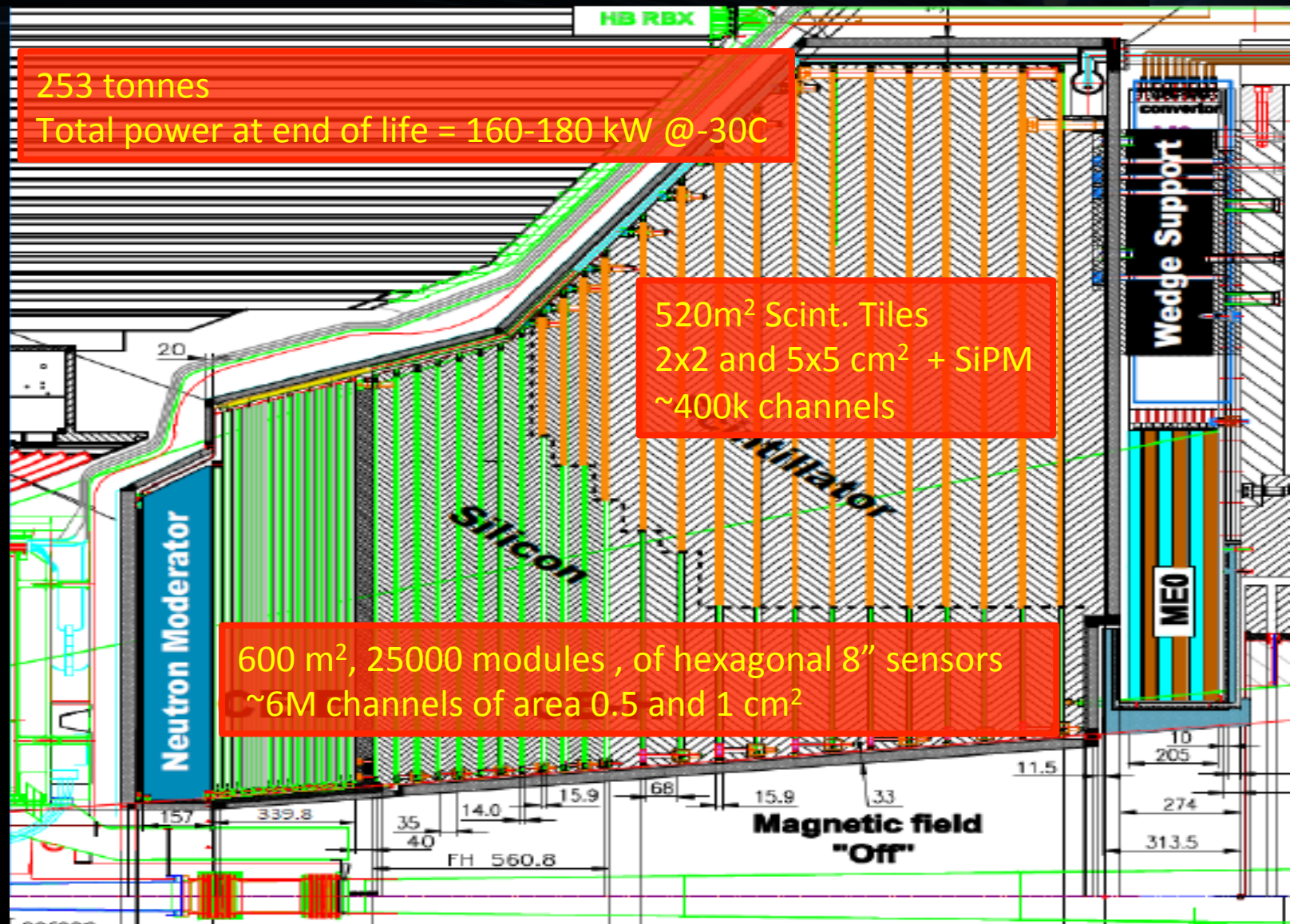


What's the UK doing?

- UK has led, and continues to lead, the development of the Stacked-tracker concept and the tracking-trigger
  - UK delivering outer-tracker trigger and readout ASIC
  - UK delivering outer-tracker readout-electronics and developing track-finding algorithms
- UK has led, and continues to lead, the development of the HGC concept
  - Off-detector electronics and algorithms primarily a UK responsibility

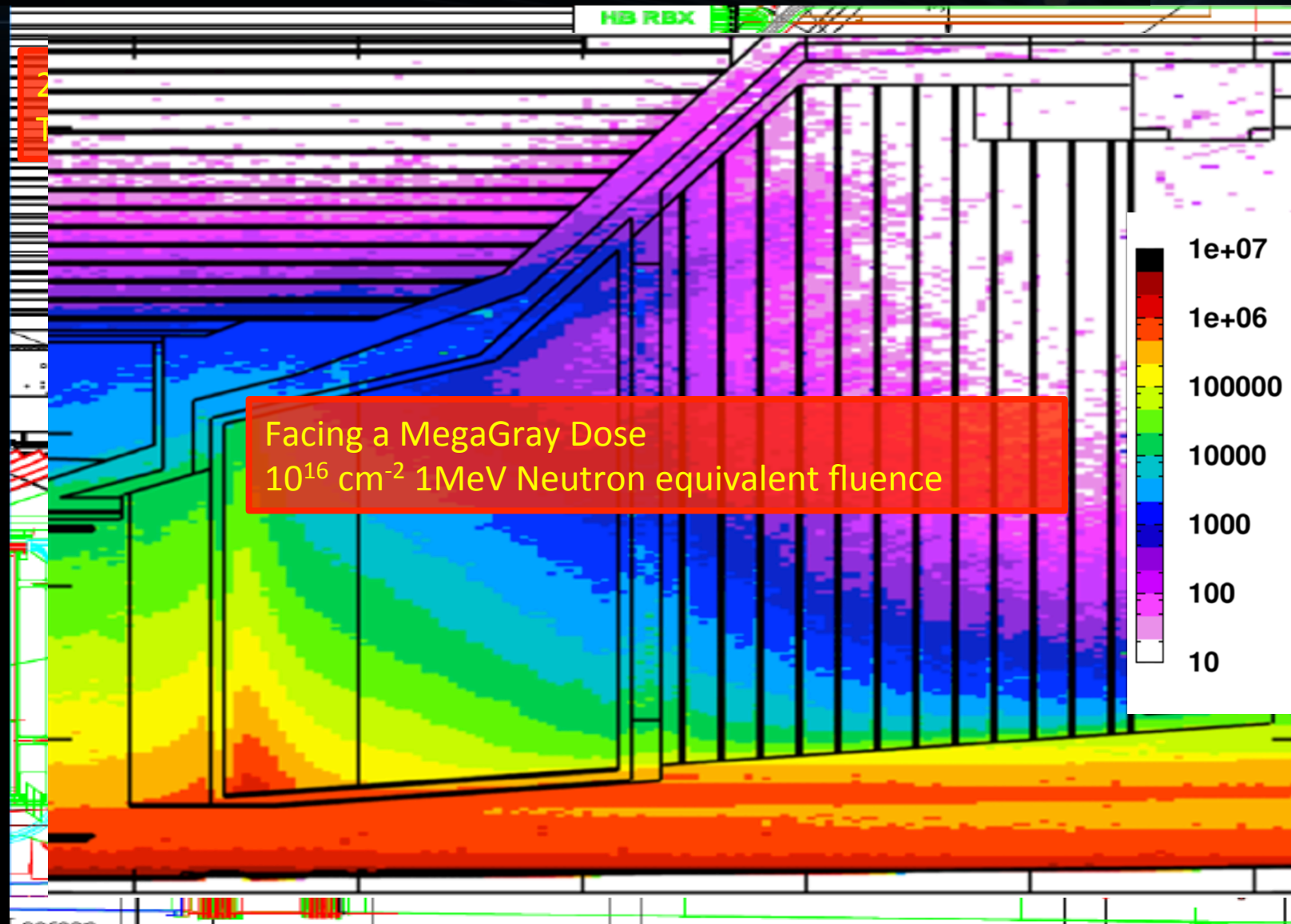
Calorimeter Endcap design

- 3D shower topology and time resolution of ~30ps
- Electromagnetic Endcap (EE)
  - 28 layers of Silicon sensors in W/Pb absorber (25 X_0, 1.7λ)
- Hadronic Endcap (EH)
  - 24 layers: 8 silicon + 16 silicon/scint. tiles at high/low η in stainless steel absorber (9λ)



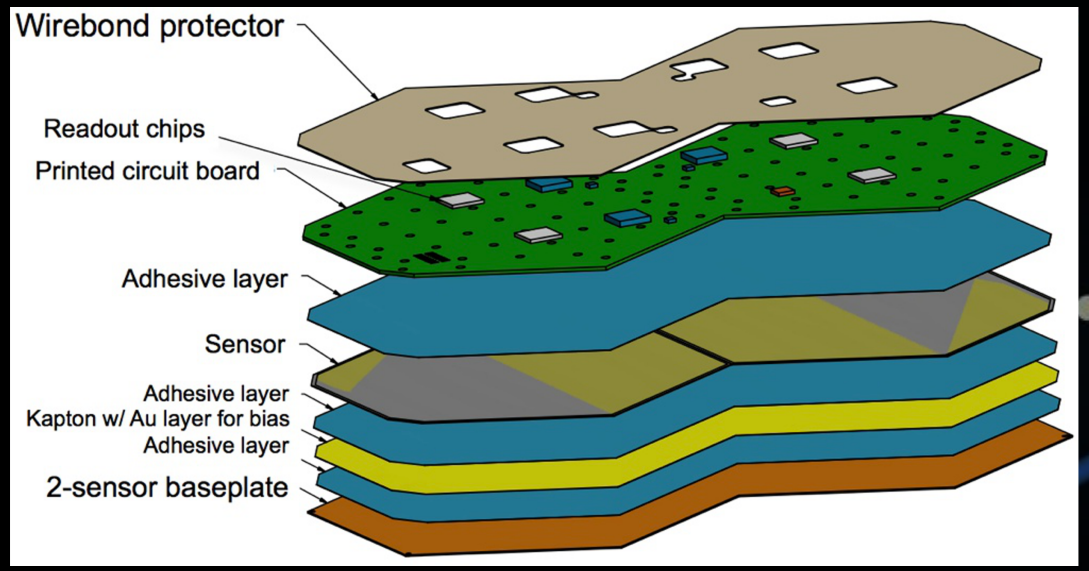
Calorimeter Endcap design

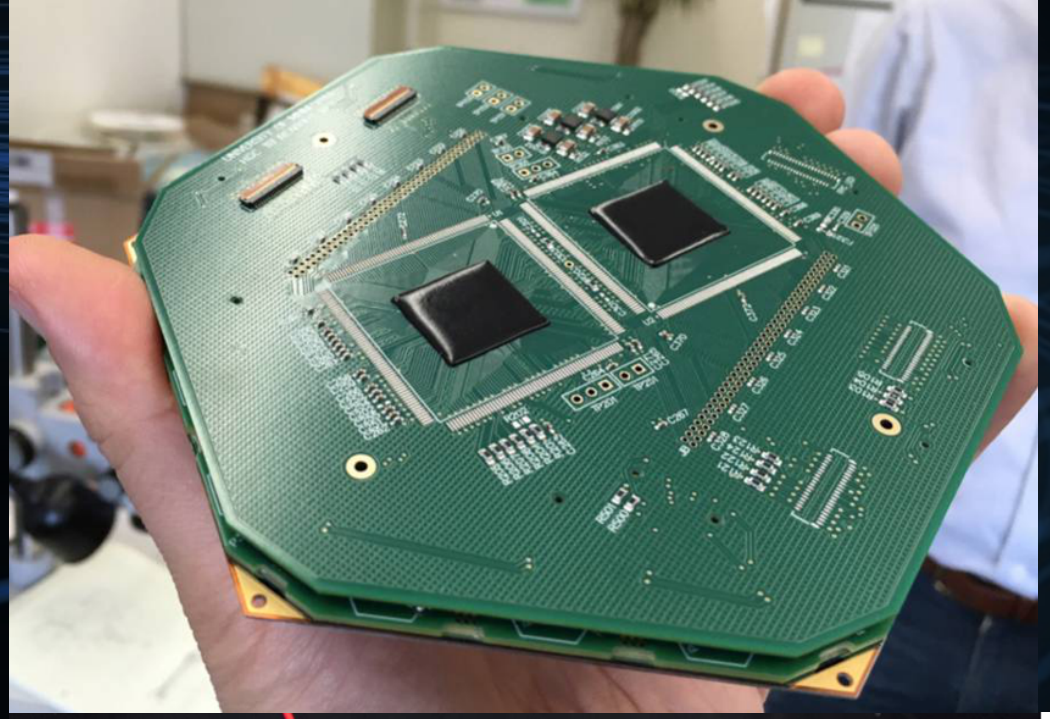
- 3D shower topology and time resolution of ~30ps
- Electromagnetic Endcap (EE)
  - 28 layers of Silicon sensors in W/Pb absorber (25 X_0, 1.7λ)
- Hadronic Endcap (EH)
  - 24 layers: 8 silicon + 16 silicon/scint. tiles at high/low η in stainless steel absorber (9λ)

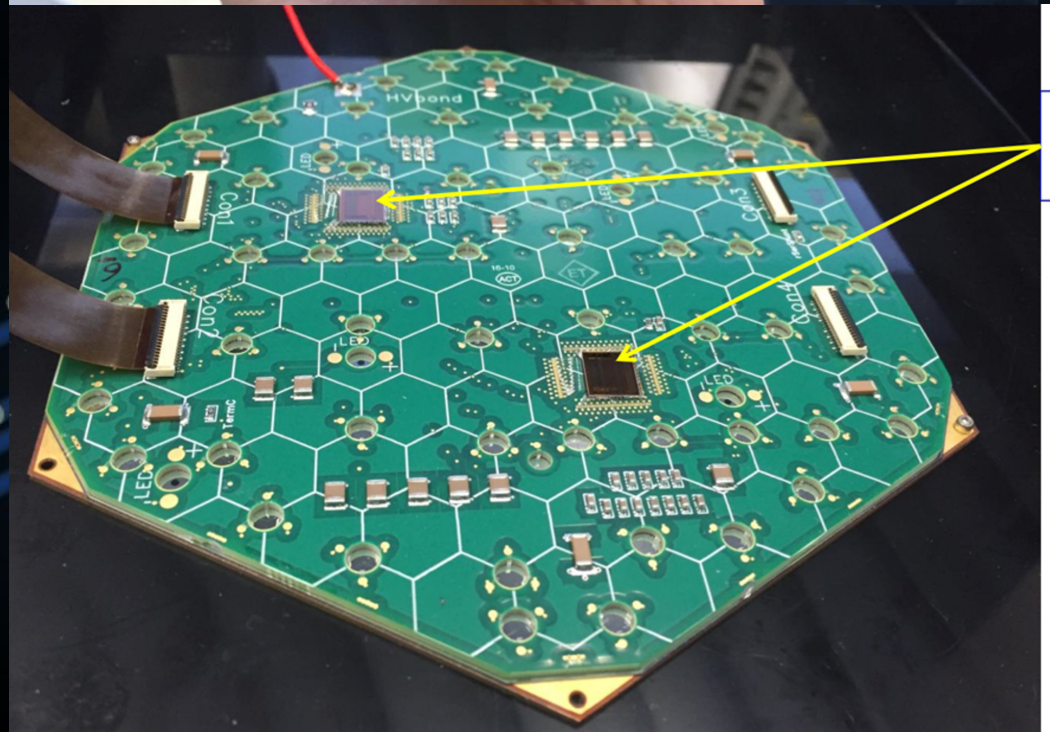


Calorimeter Endcap modules

- 593 m2 of silicon
- 6M ch, 0.5 or 1 cm2 cell-size
- 21,660 modules (8” or 2x6” sensors)
- 92,000 front-end ASICS



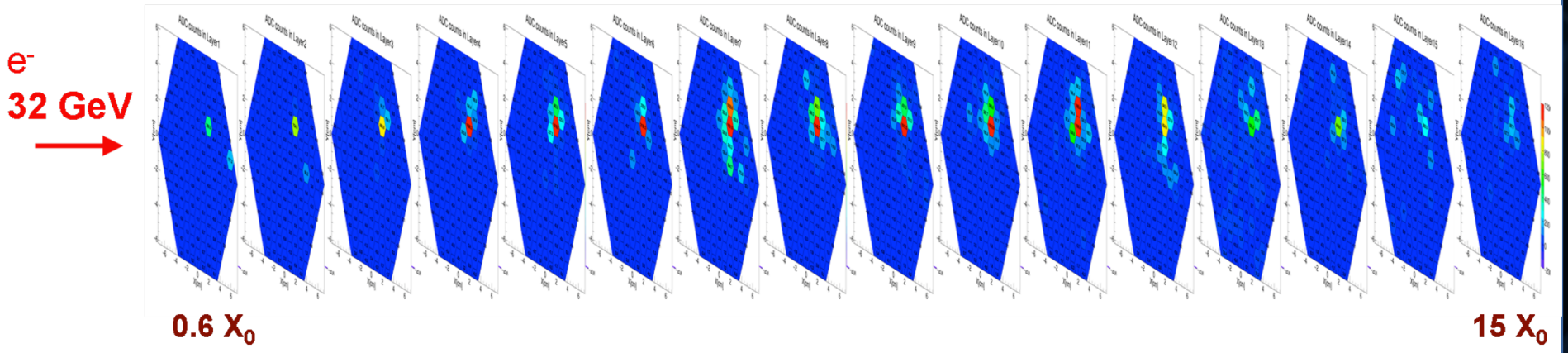




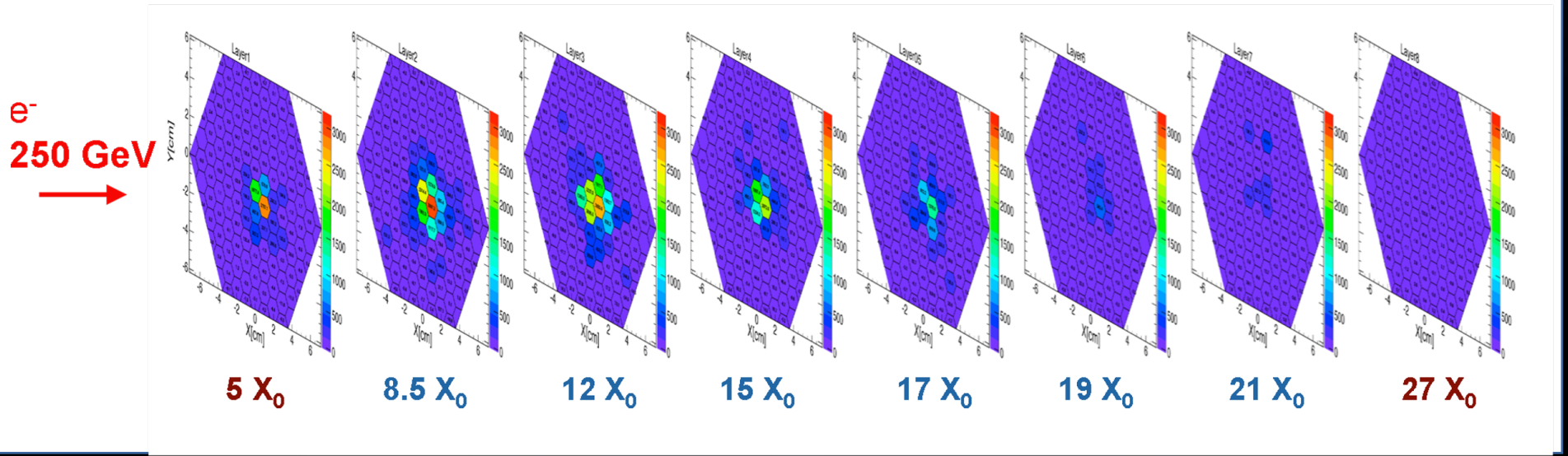
SKIROC2 ASIC

Calorimeter Endcap modules: Do they work?

Fermilab: **32 GeV electrons** passing through **15 X_0**.



CERN: **250 GeV electrons** passing through **27 X_0**.



What's the UK doing?

- UK has led, and continues to lead, the development of the Stacked-tracker concept and the tracking-trigger
  - UK delivering outer-tracker trigger and readout ASIC
  - UK delivering outer-tracker readout-electronics and developing track-finding algorithms
- UK has led, and continues to lead, the development of the HGC concept
  - Off-detector electronics and algorithms primarily a UK responsibility
- UK has played leading role in the trigger electronics and algos, would like to continue to do so
- Barrel-Muon trigger currently using UK hardware and keen to continue doing so in future
- Global trigger currently using UK hardware (future TBD)
- Good relationship with those responsible for designing the MTD back-end, they are proposing to use UK hardware

What's the UK doing?

- UK has led, and continues to lead, the development of the Stacked-tracker concept and the tracking-trigger
  - UK delivering outer-tracker trigger and readout ASIC
  - UK delivering outer-tracker readout-electronics and developing track-finding algorithms
- UK has led, and continues to lead, the development of the HGC concept
  - Off-detector electronics and algorithms primarily a UK responsibility
- UK has played leading role in the trigger, electronics and algos would like to continue to do so
- Barrel-Muon trigger currently using UK hardware and keen to continue doing so in future
- Global trigger currently using UK hardware (future TBD)
- Good relationship with those responsible for designing the MTD back-end, they are proposing to use UK hardware

So… looking like UK has major hardware role to play in every system except

- Barrel Calorimeter
- Endcap Muons
- Pixels

What's the UK doing?

- UK has led, and continues to lead, the development of the Stacked-tracker concept and the tracking-trigger
  - UK delivering outer-tracker trigger and readout ASIC
  - UK delivering outer-tracker readout-electronics and developing track-finding algorithms
- UK has led, and continues to lead, the development of the HGC concept
  - Barrel Calorimeter
  - Endcap Muons
  - Pixels
- UK has played leading role in the trigger electronics and algos, would like to continue to do so
- Barrel-Muon trigger currently using UK hardware and keen to continue doing so in future
- Global trigger currently using UK hardware (future TBD)
- Good relationship with those responsible for designing the MTD back-end, they are proposing to use UK hardware

So... looking like UK has major hardware role to play in every system account

Not bad going for a collaboration of only 4 institutes

What's the UK doing?

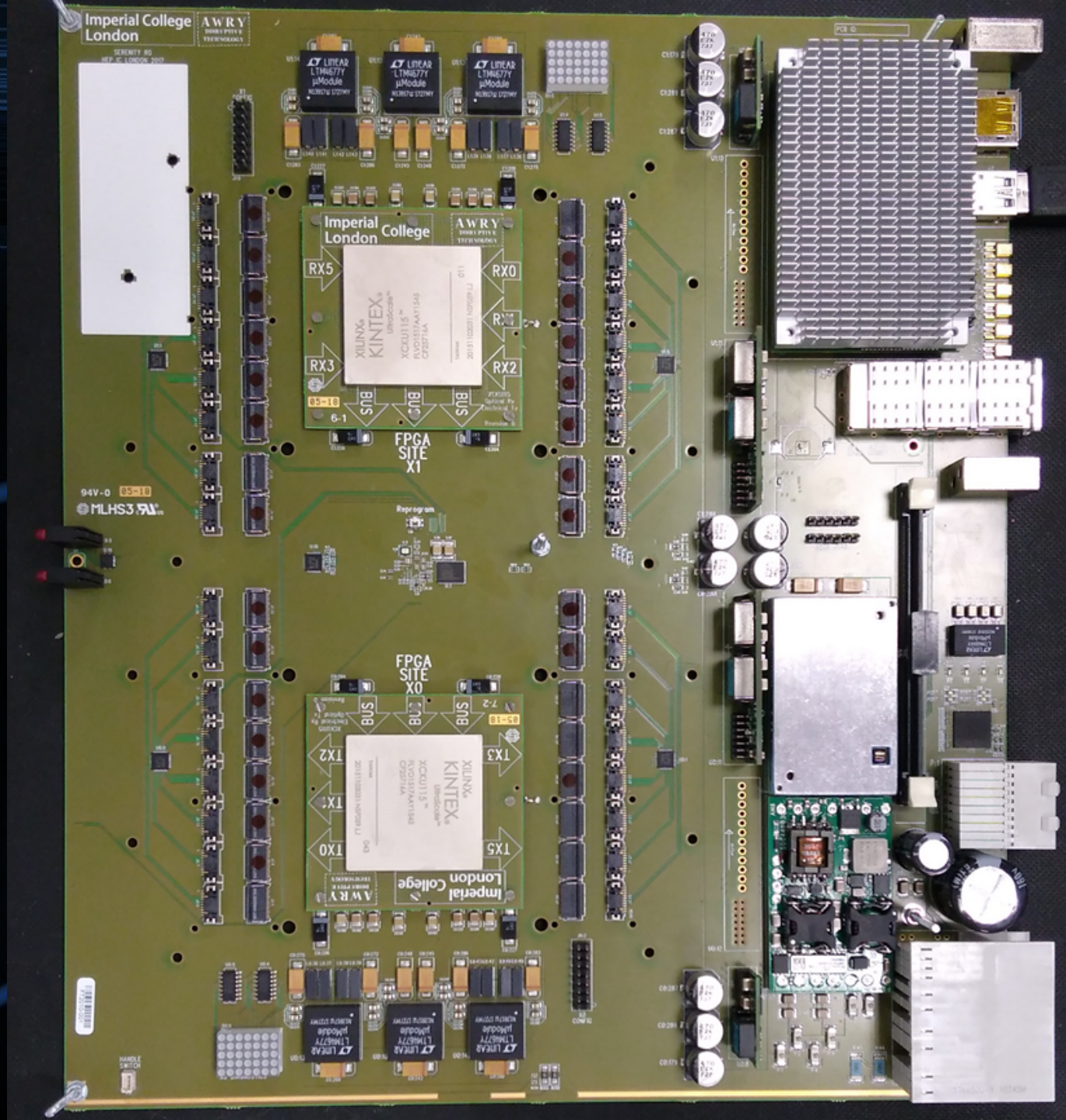
- UK has led, and continues to lead, the development of the Stacked-tracker concept and the tracking-trigger
  - UK delivering outer-tracker trigger and readout ASIC
  - UK delivering outer-tracker readout-electronics and developing track-finding algorithms
- UK has led, and continues to lead, the development of the HGC concept
  - Off-detector electronics and algorithms primarily a UK responsibility
- UK has played leading role in the trigger electronics and algos, would like to continue to do so
- Barrel-Muon trigger currently using UK hardware and keen to continue doing so in future
- Global trigger currently using UK hardware (future TBD)
- Good relationship with those responsible for designing the MTD back-end, they are proposing to use UK hardware

So who is stupid enough to take on the task of designing hardware for so many disparate systems?

So who is stupid enough to take on the task of designing hardware for so many disparate systems?

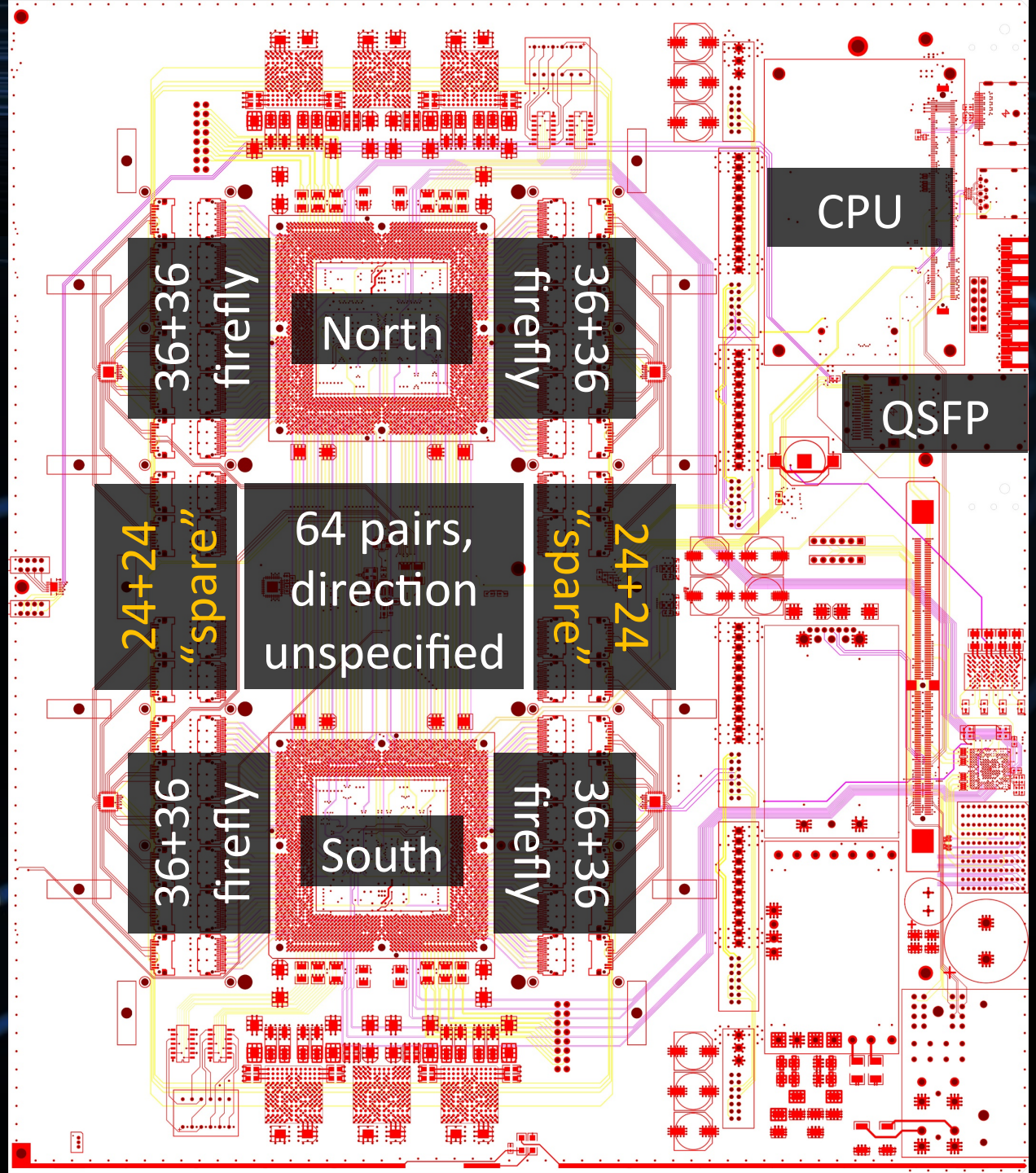
Serenity

- ATCA Development Platform
- Next generation optical stream processor
- Carrier Card provides the board services
- Daughter Cards host data-processing FPGAs



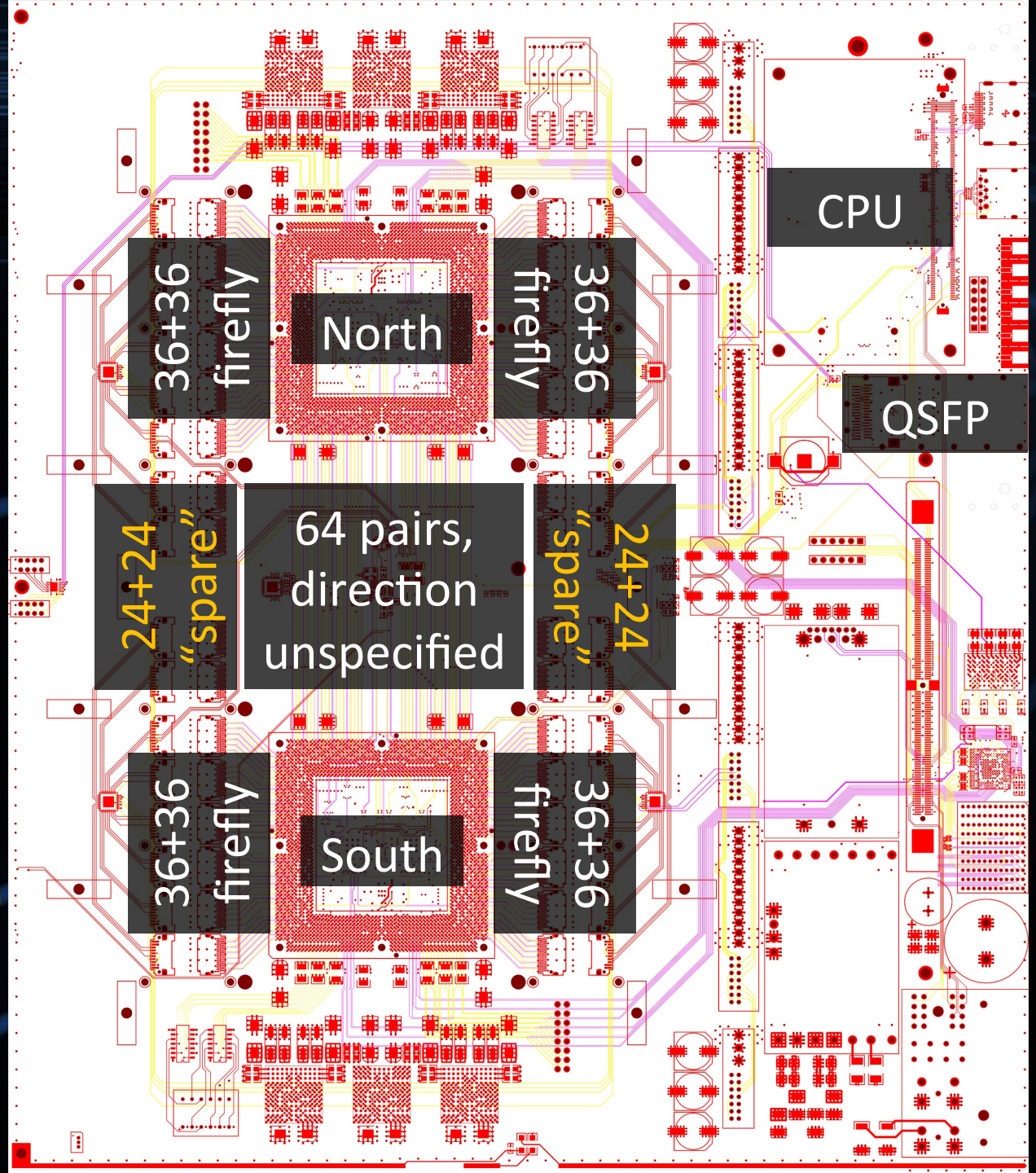
Serenity

- Standard configuration. Provision for:
  - 2 × 72+72 links @ 25Gbps optical link
    - 3.6 + 3.6 Tbps
  - 64 links @ 25Gbps between DCs
    - 1.6 Tbps
- Optional optical expansion:
  - 2 × 96+96 links @ 25Gbps optical link
    - 4.8 + 4.8 Tbps
  - 16 links @ 25Gbps between DCs
    - 0.4 Tbps

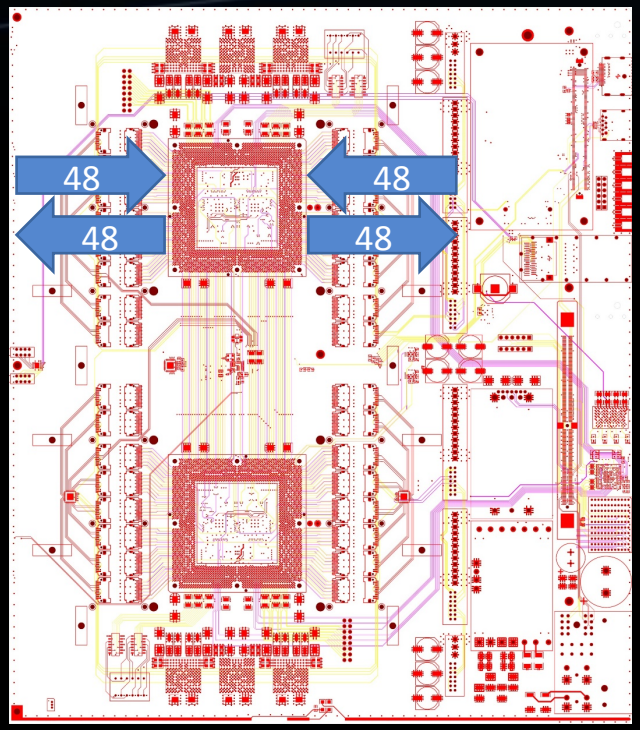


Serenity

- Freedom to choose your preferred family, package, generation, (vendor?)
- Freedom to choose your balance of optical and electrical connectivity
- Reduces financial risk since carrier (bulk of potential failure-modes) qualified before FPGAs (bulk of the cost) are fitted!

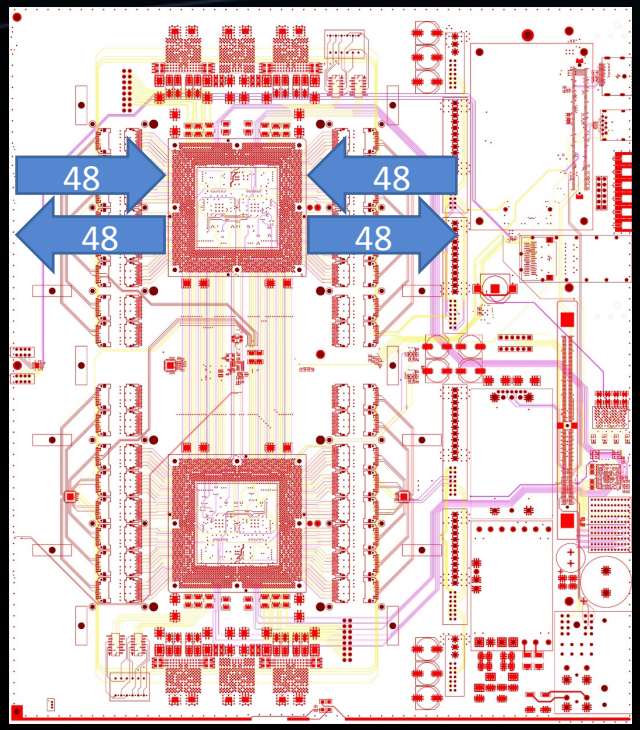


Serenity: One board to rule them all?

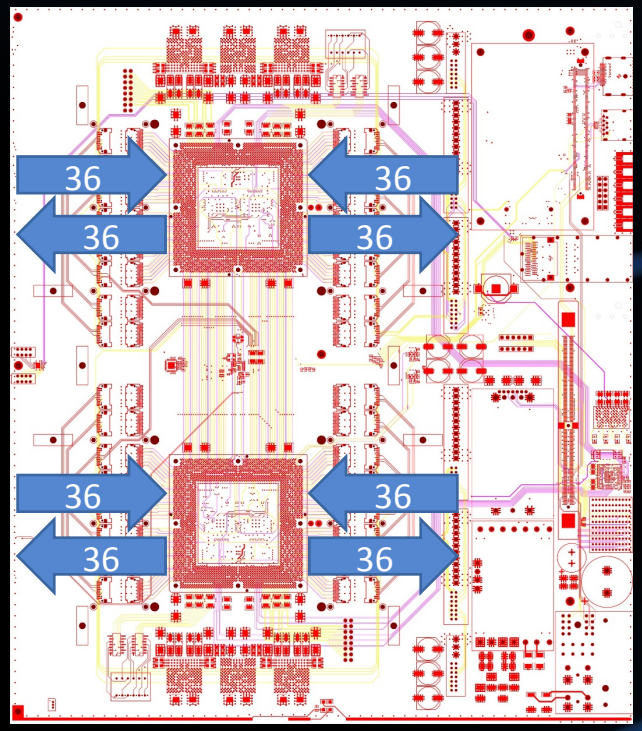


HGC Trigger v1
(now disfavoured)

Serenity: One board to rule them all?

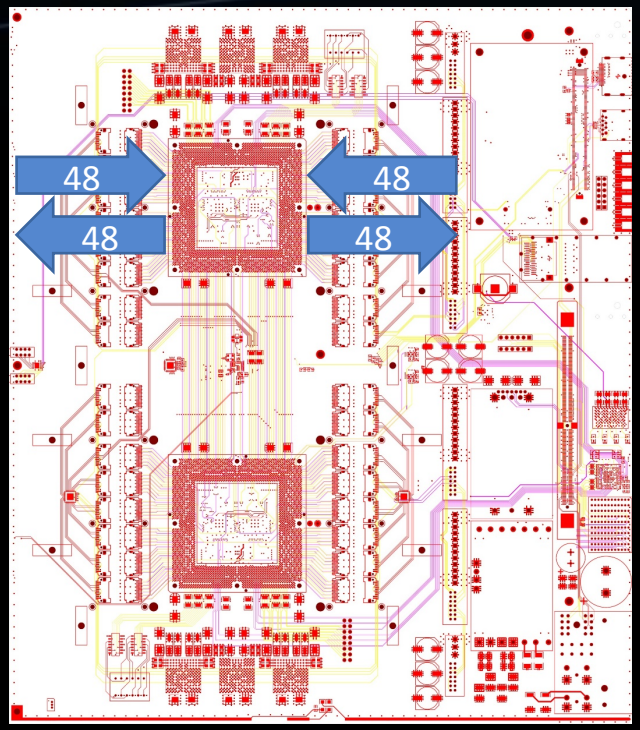


HGC Trigger v1
(now disfavoured)

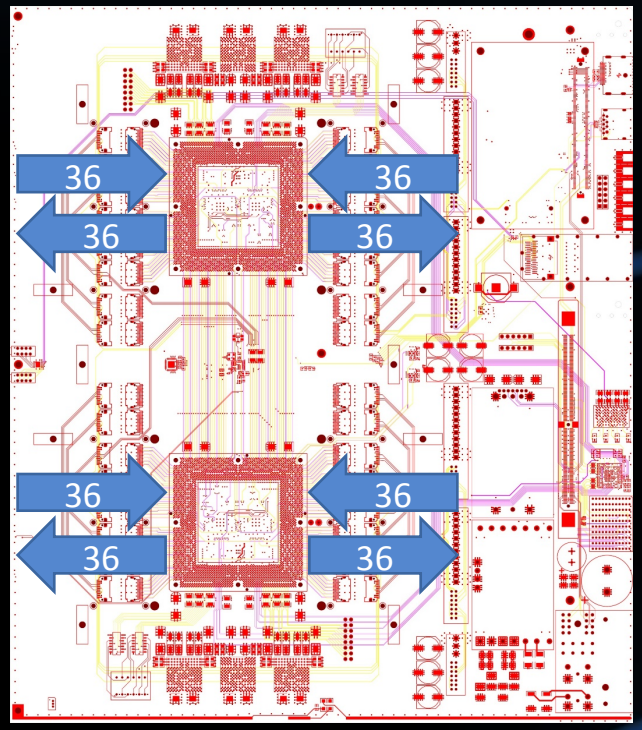


HGC DAQ
HGC Trigger v2
Barrel Muon Processor?

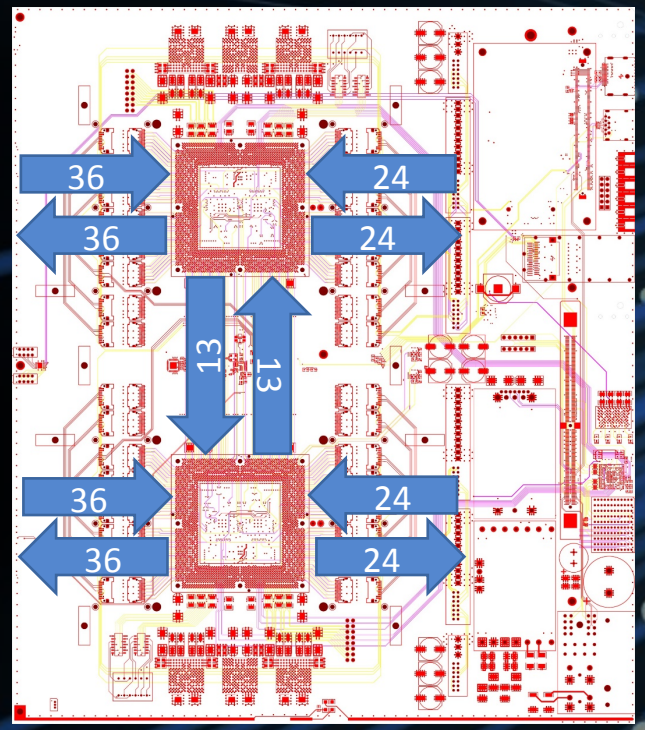
Serenity: One board to rule them all?



HGC Trigger v1
(now disfavoured)

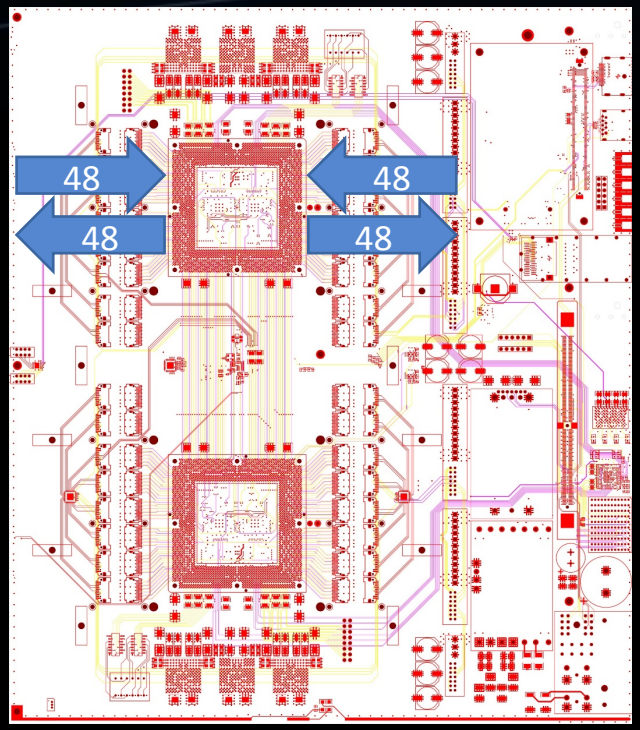


HGC DAQ
HGC Trigger v2
Barrel Muon Processor?

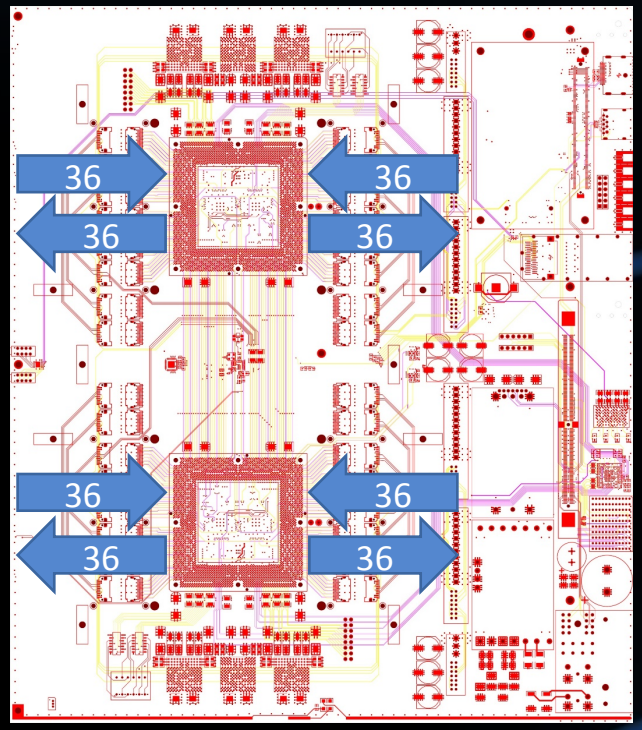


Outer Tracker DTC
MTD DTC?
The Pixel DTC we
are not building for
political reasons

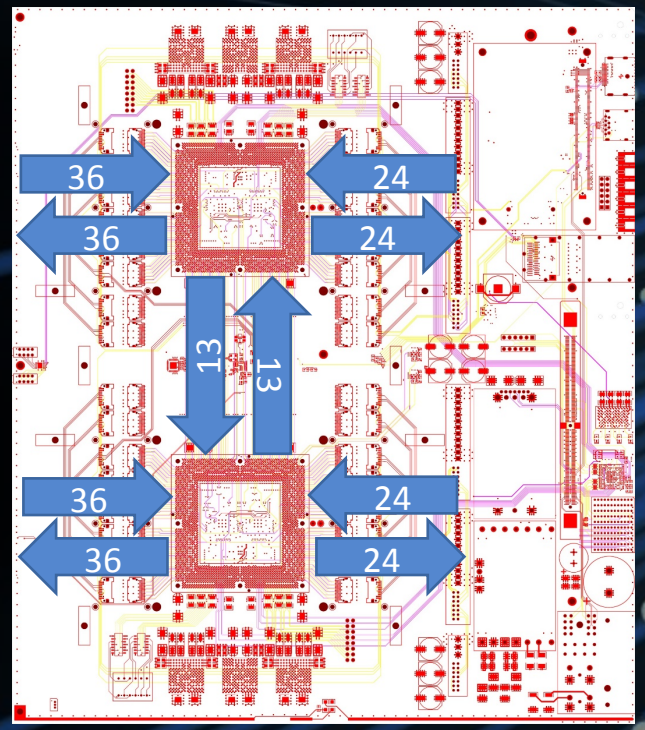
Serenity: One board to rule them all?



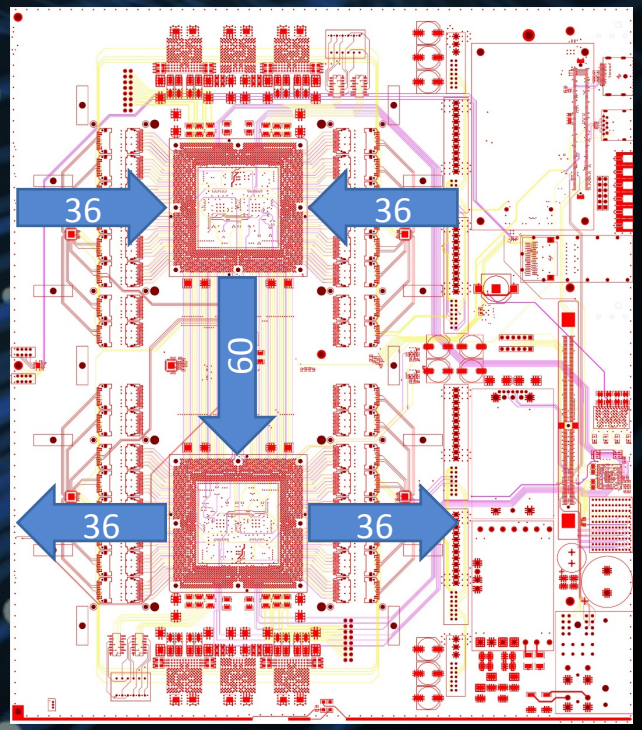
HGC Trigger v1
(now disfavoured)



HGC DAQ
HGC Trigger v2
Barrel Muon Processor?



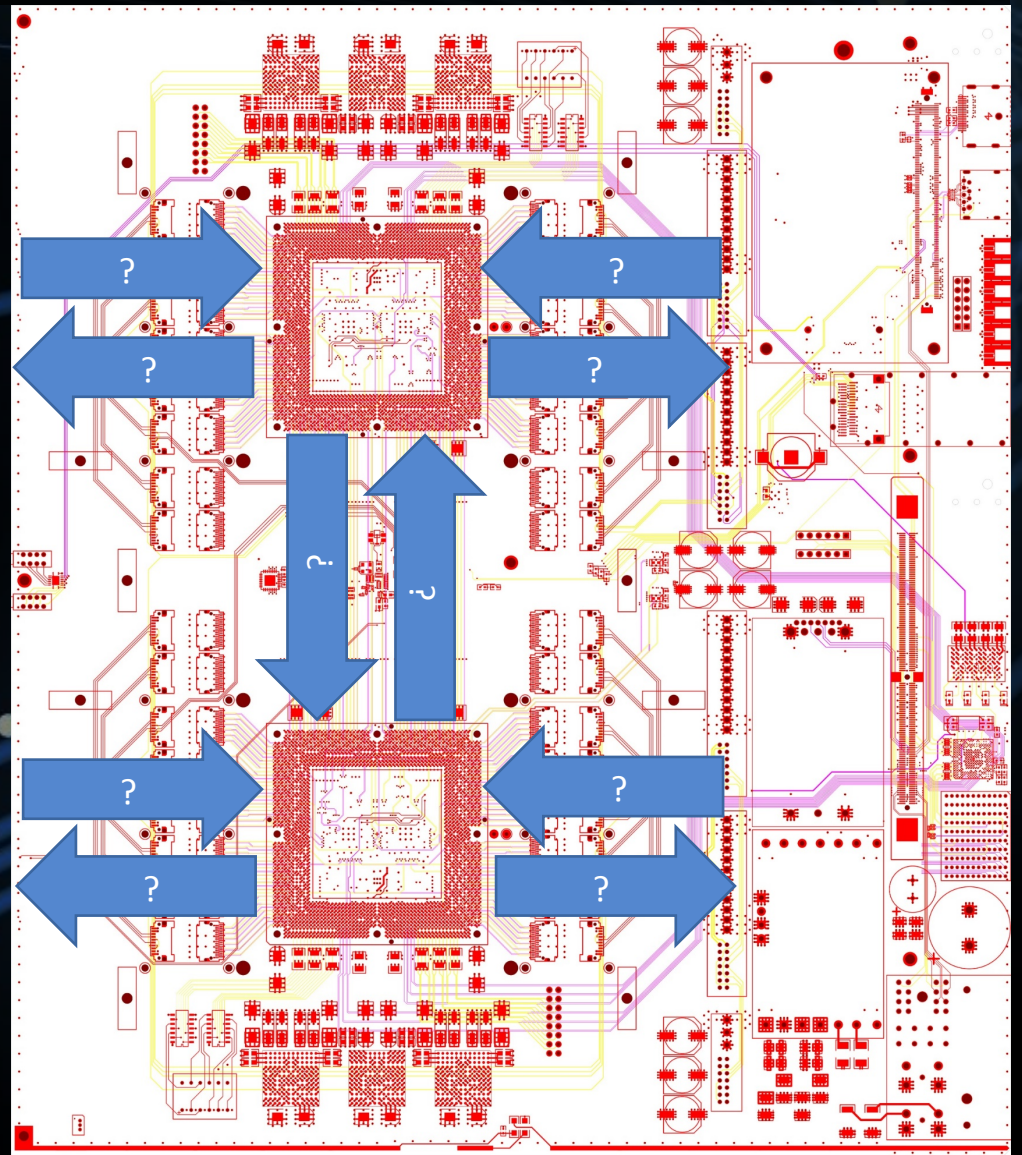
Outer Tracker DTC
MTD DTC?
The Pixel DTC we are not building for political reasons



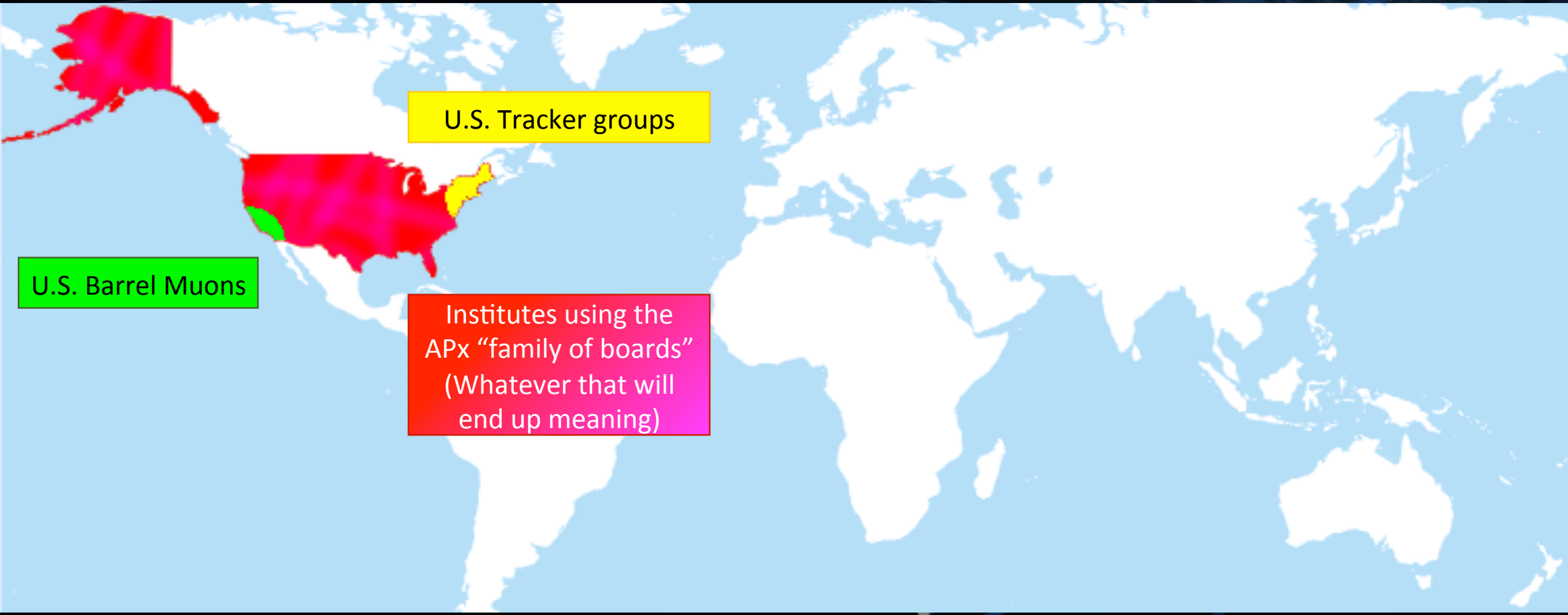
The track-finder processor we are not building for political reasons. Ooops.

Serenity: One board to rule them all?

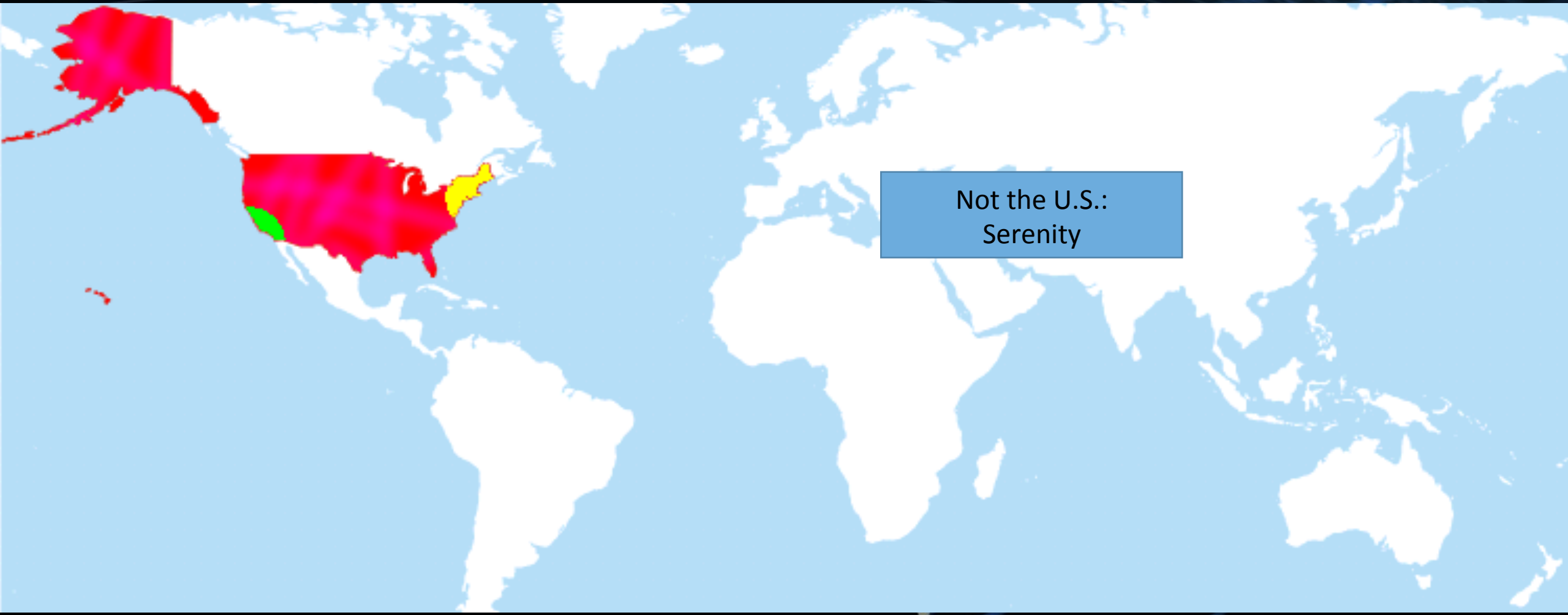
- I would show how it fulfils the role of the trigger correlator
  - If anyone could tell me what they require of the trigger correlator
- I'll be darned if they can come up with a scenario it can't handle



So is the prophecy coming true?

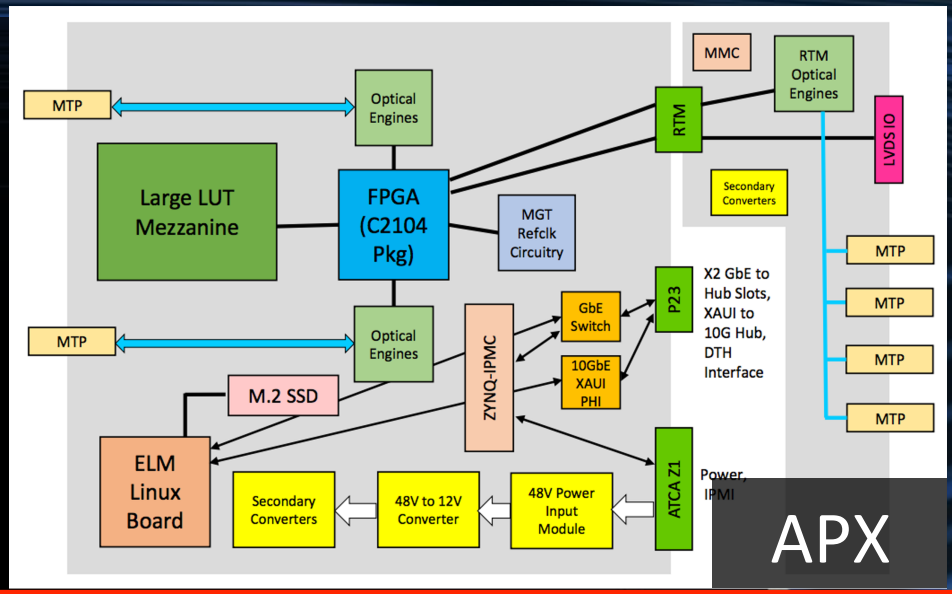


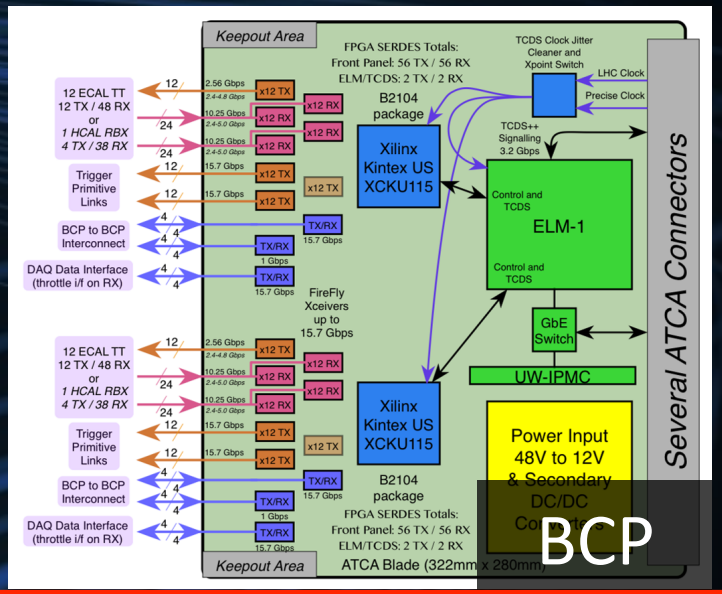
So is the prophecy coming true?

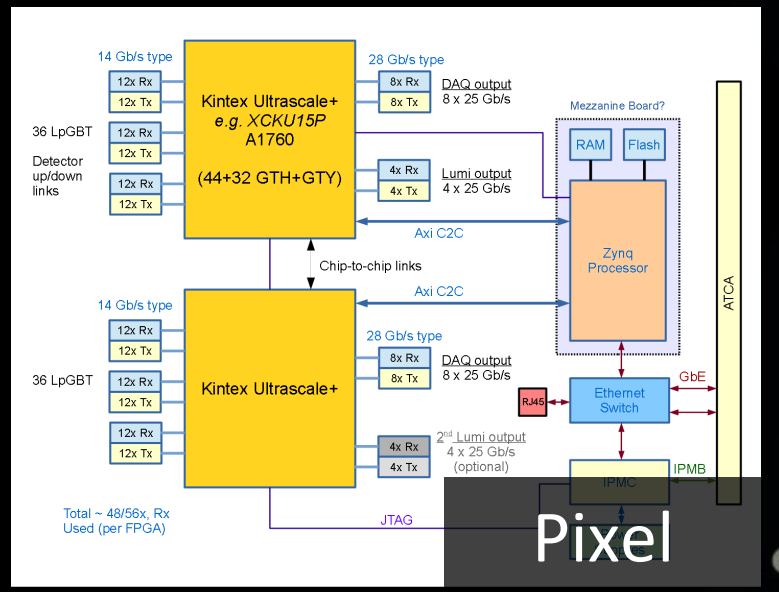


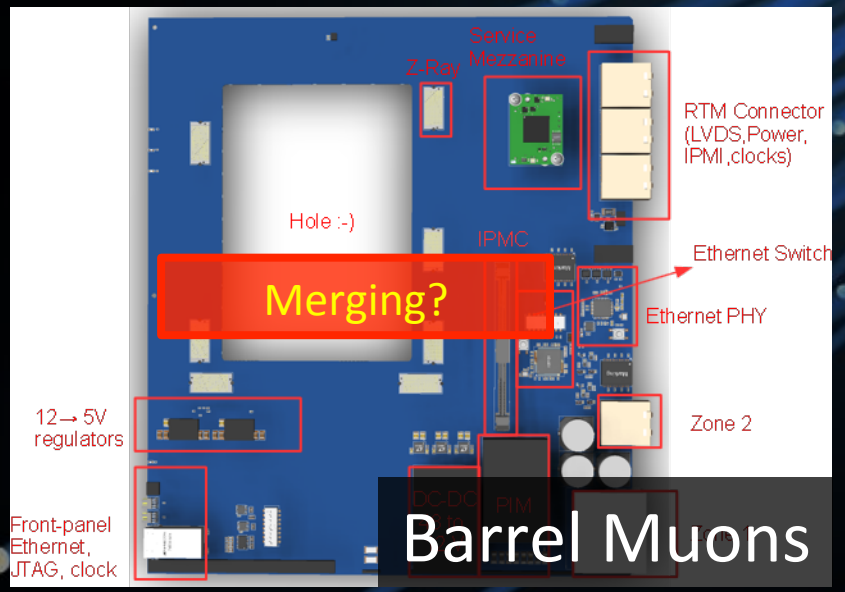
So is the prophecy coming true?

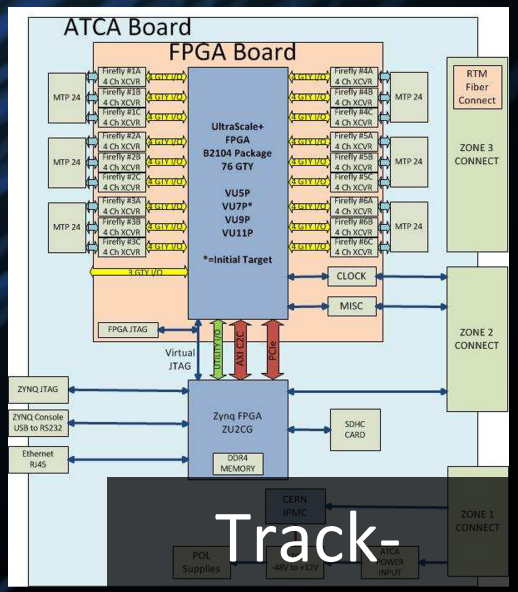
Or: How many boards does the US actually need?





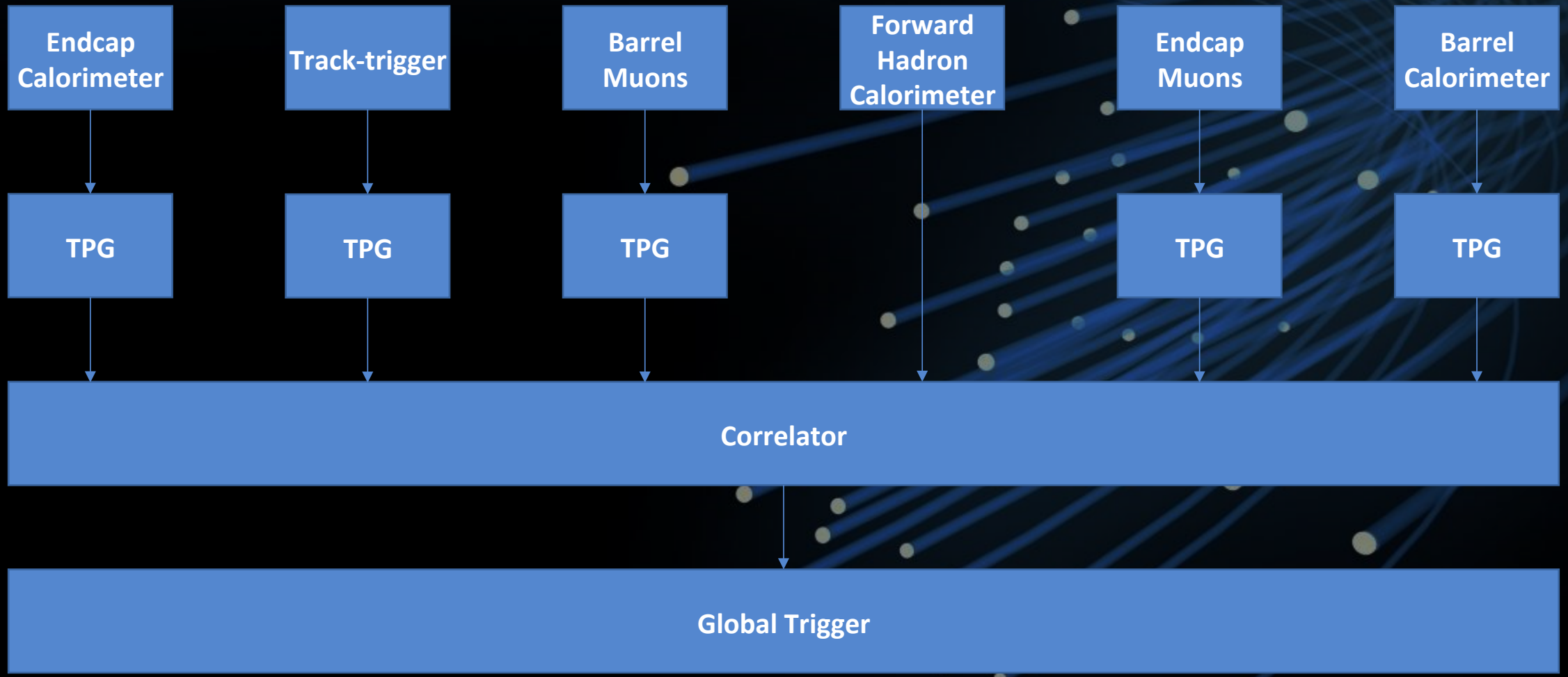




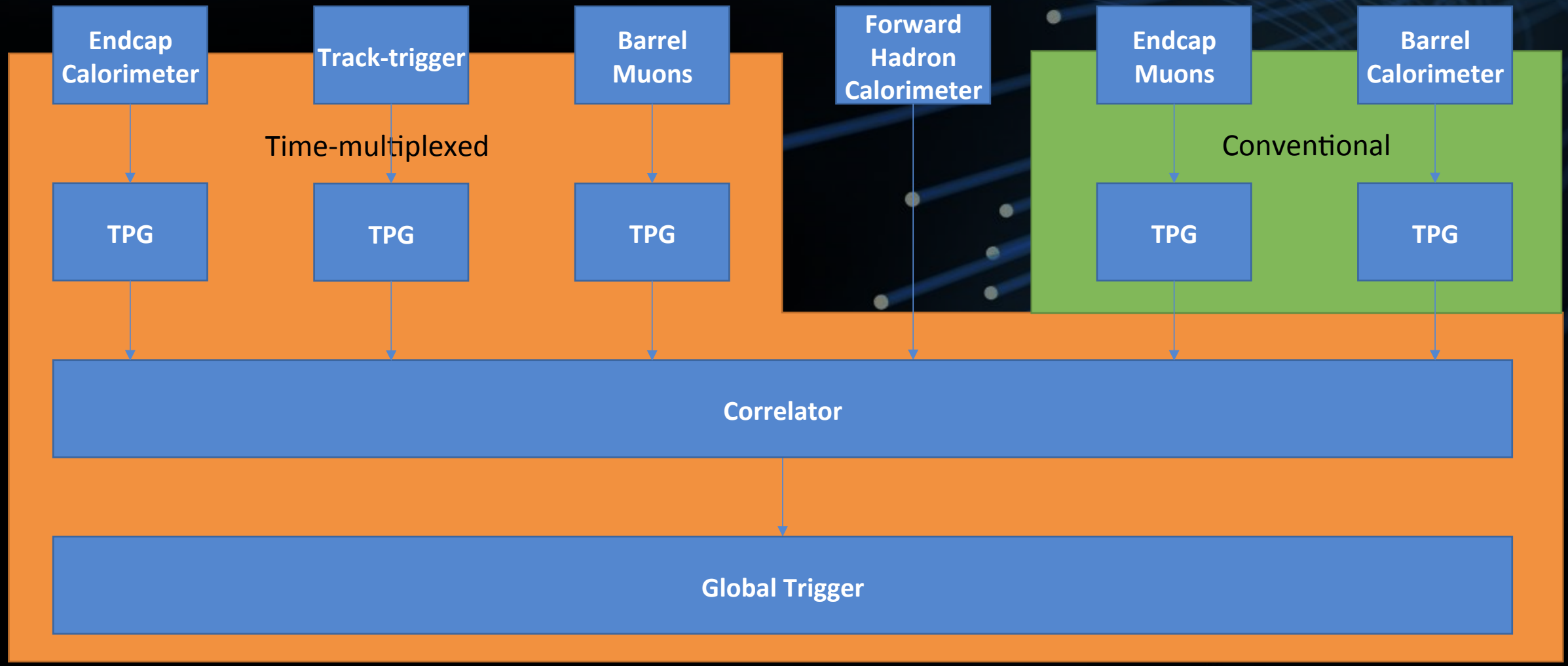


DTC

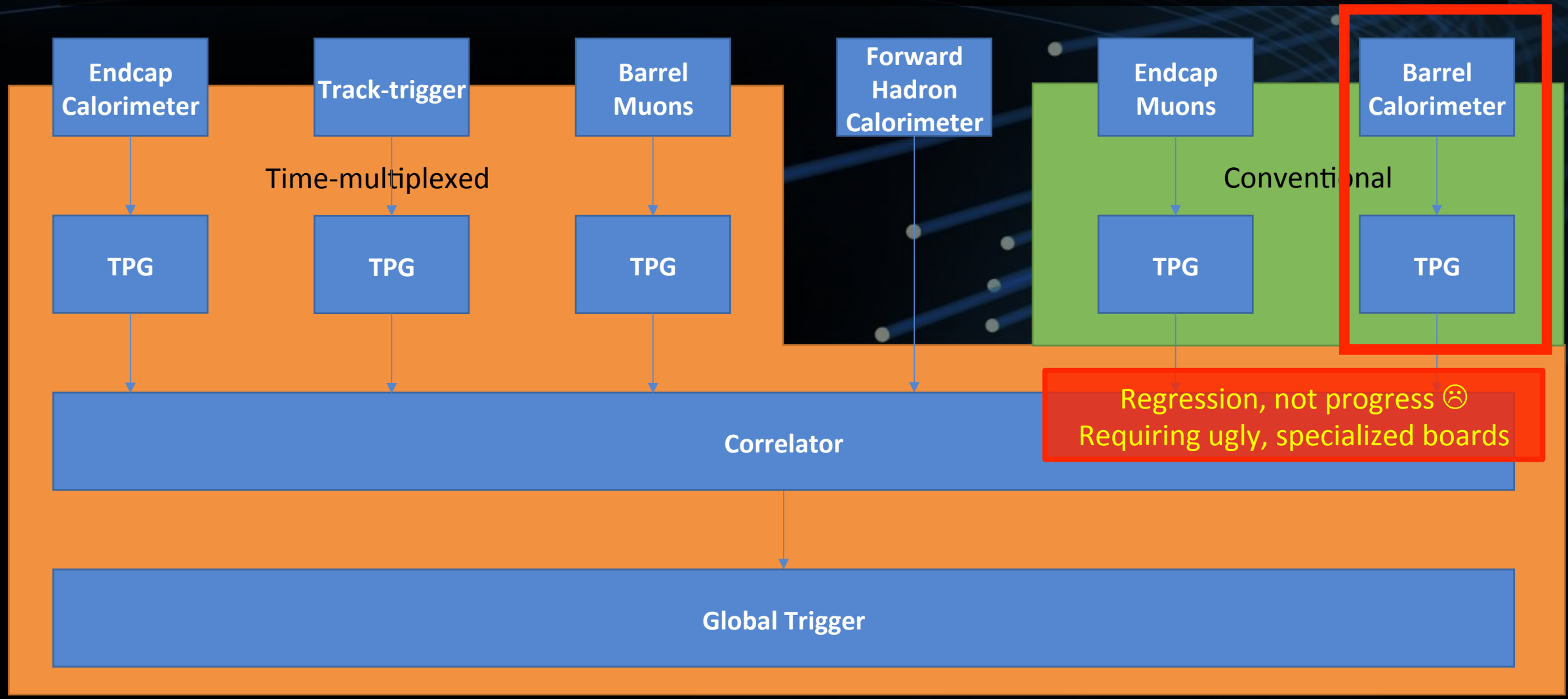
And have you settled the architecture question yet?



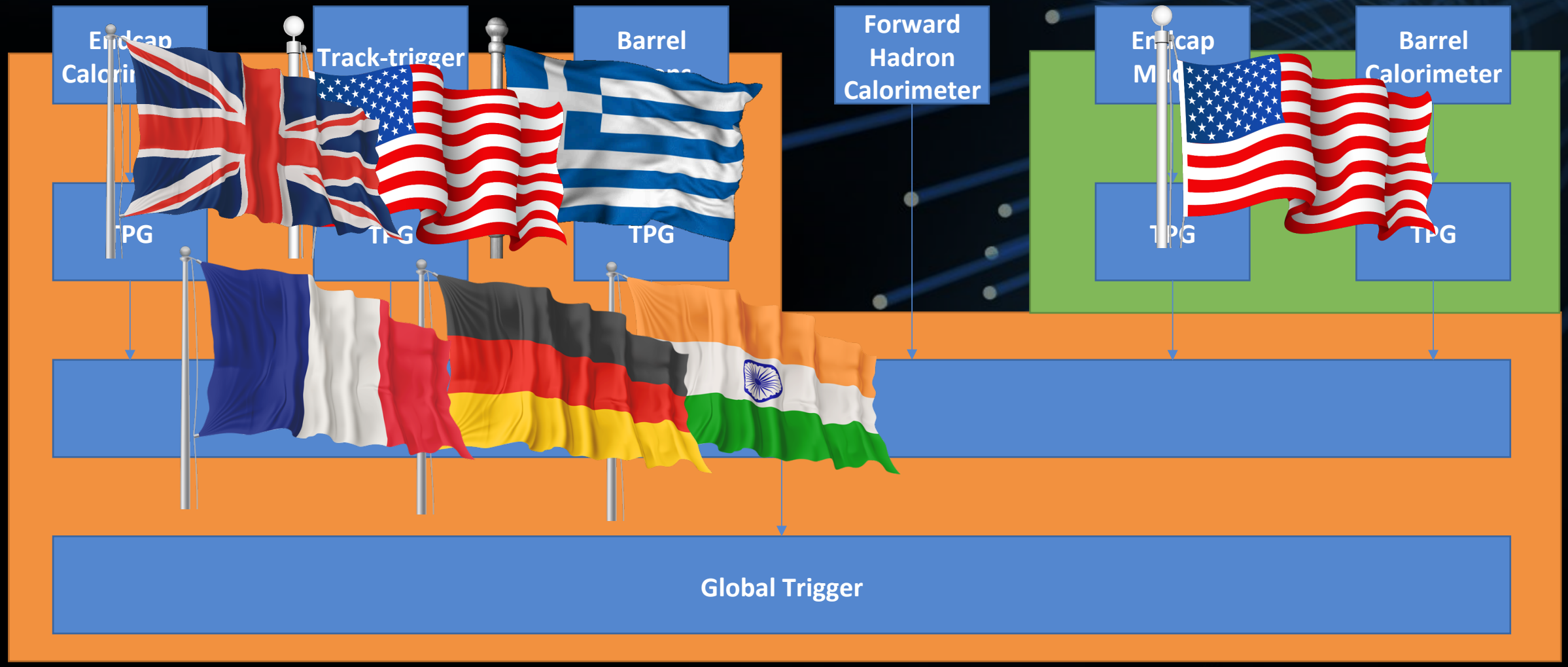
And have you settled the architecture question yet? Most likely architecture



And have you settled the architecture question yet? Most likely architecture



And have you settled the architecture question yet? Is this starting to look familiar?



Anyway…

- Enough of the politics
- Back to the interesting stuff!

What are the firmware challenges at Phase-II?

What are the firmware challenges at Phase-II?

- You mean, apart from the small matter of 300Tb/s of data?

What are the firmware challenges at Phase-II?

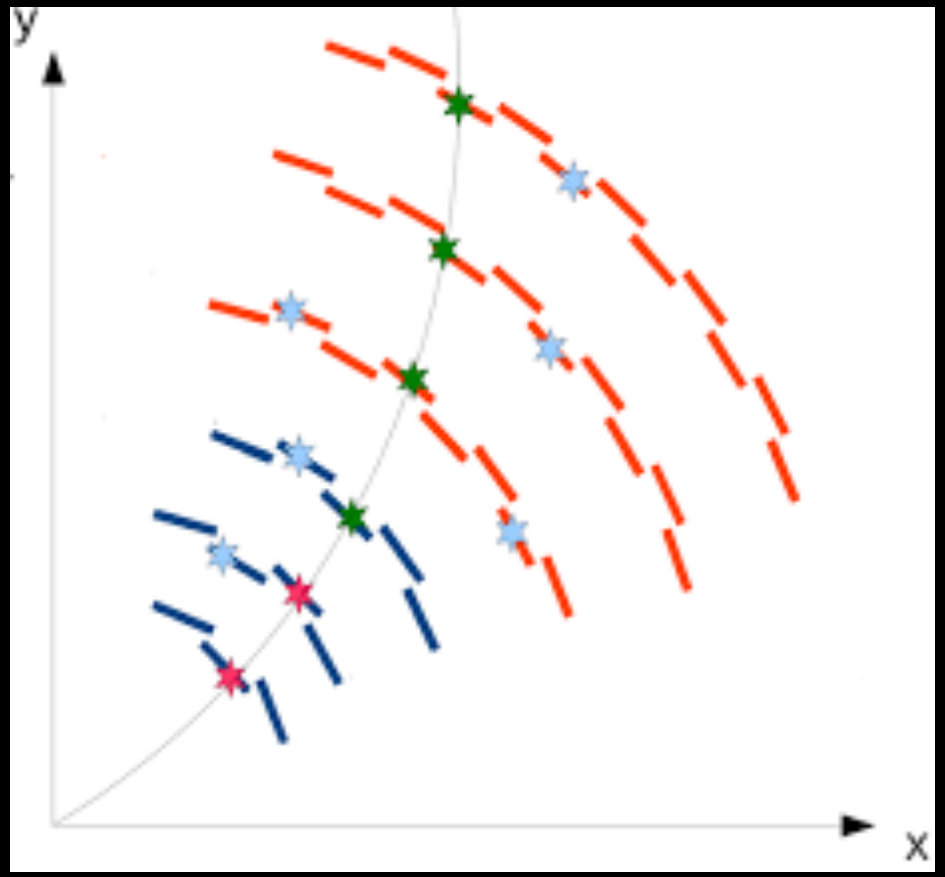
- You mean, apart from the small matter of 300Tb/s of data?
- So much data it has to be zero-suppressed
  - No (or, at least, limited) geometric timing which can be utilized
  - Variable data-volume
    - Do you handle the worst case? Very inefficient
    - Do you handle the average? How do you handle overflows?

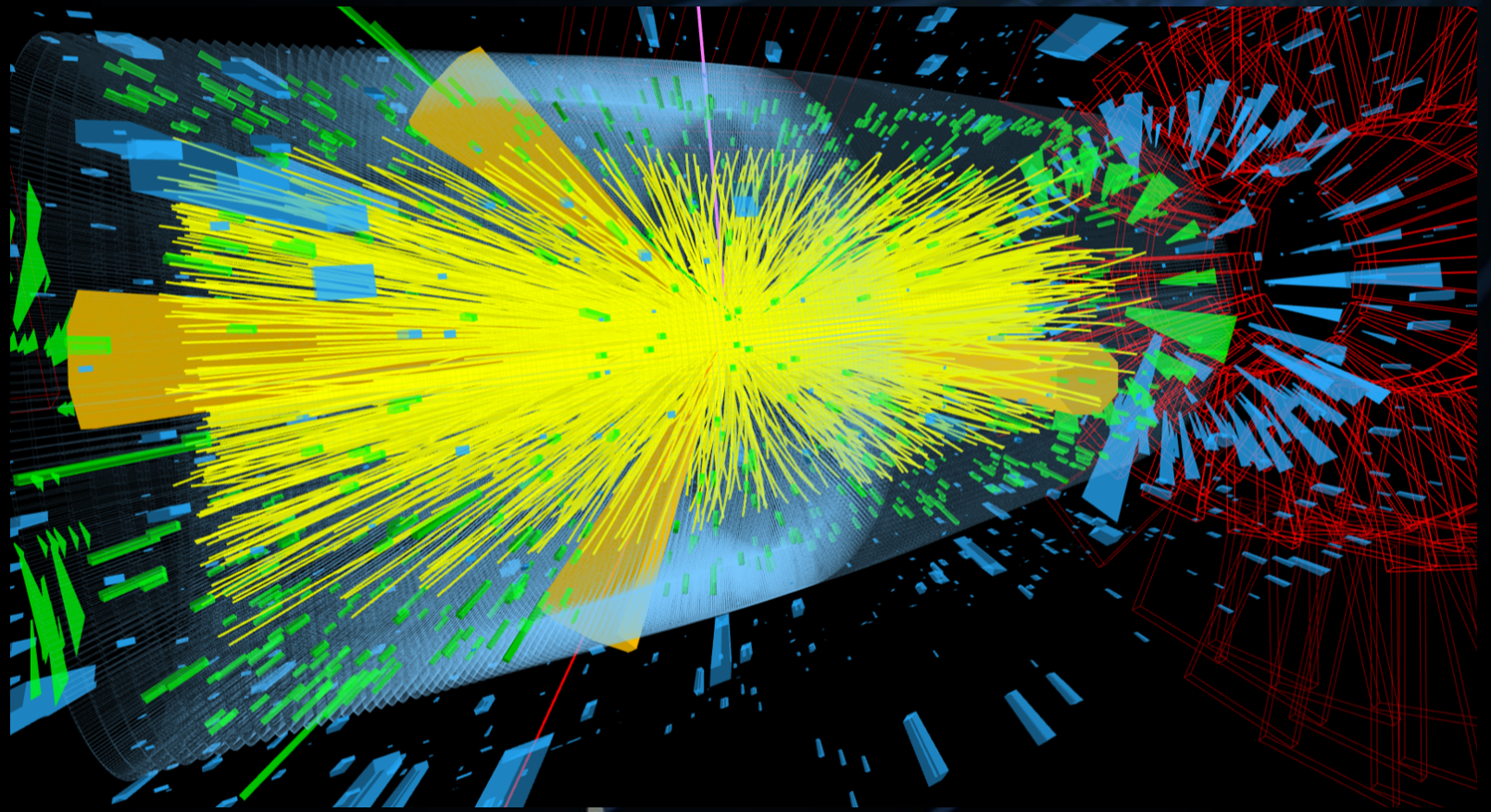
What are the firmware challenges at Phase-II?

- You mean, apart from the small matter of 300Tb/s of data?
- So much data it has to be zero-suppressed
  - No (or, at least, limited) geometric timing which can be utilized
  - Variable data-volume
    - Do you handle the worst case? Very inefficient
    - Do you handle the average? How do you handle overflows?
- We did such a good job at Phase-I, people have very high expectations…

What are the firmware challenges at Phase-II?

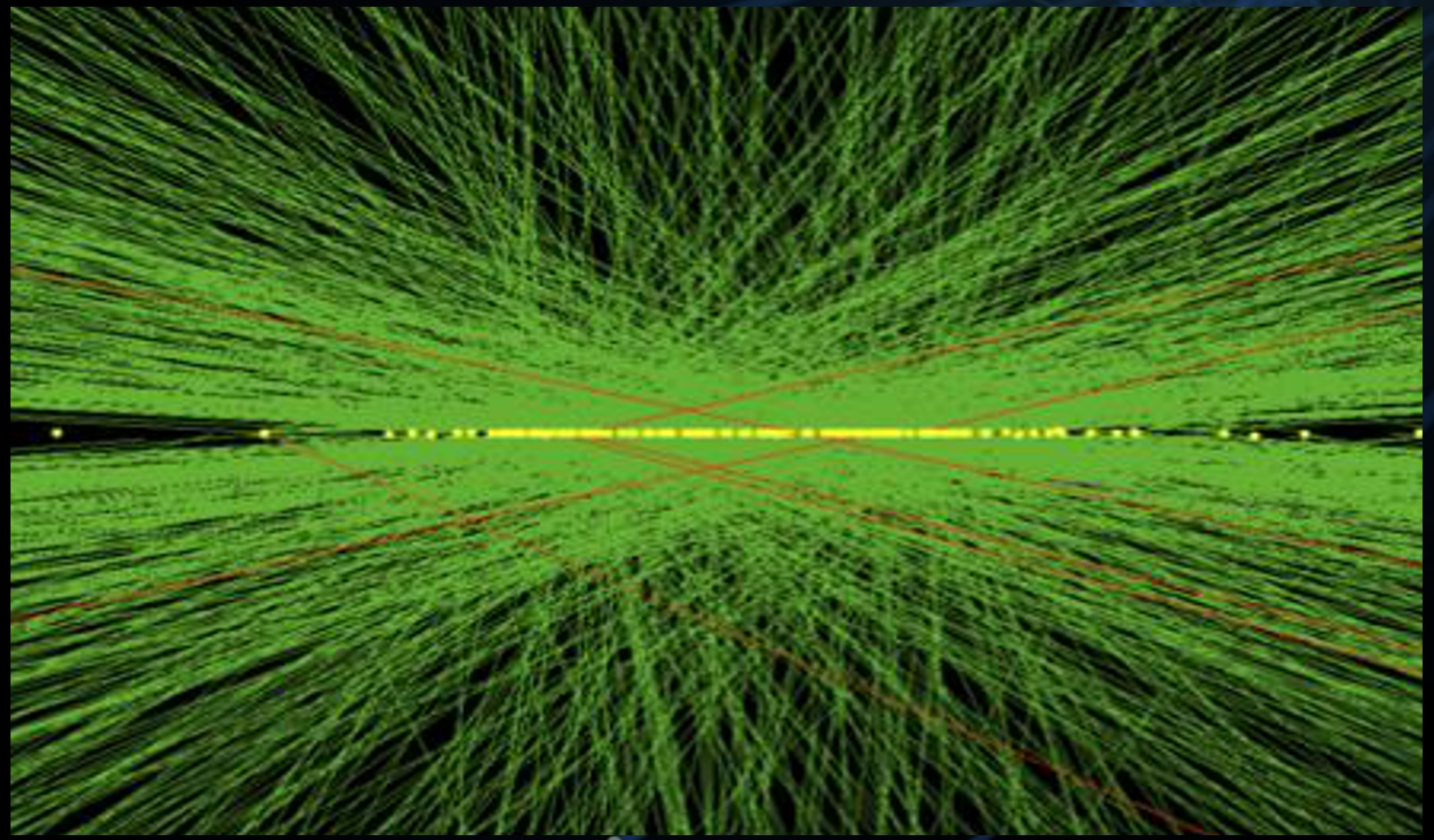
- Real-time track-finding and fitting





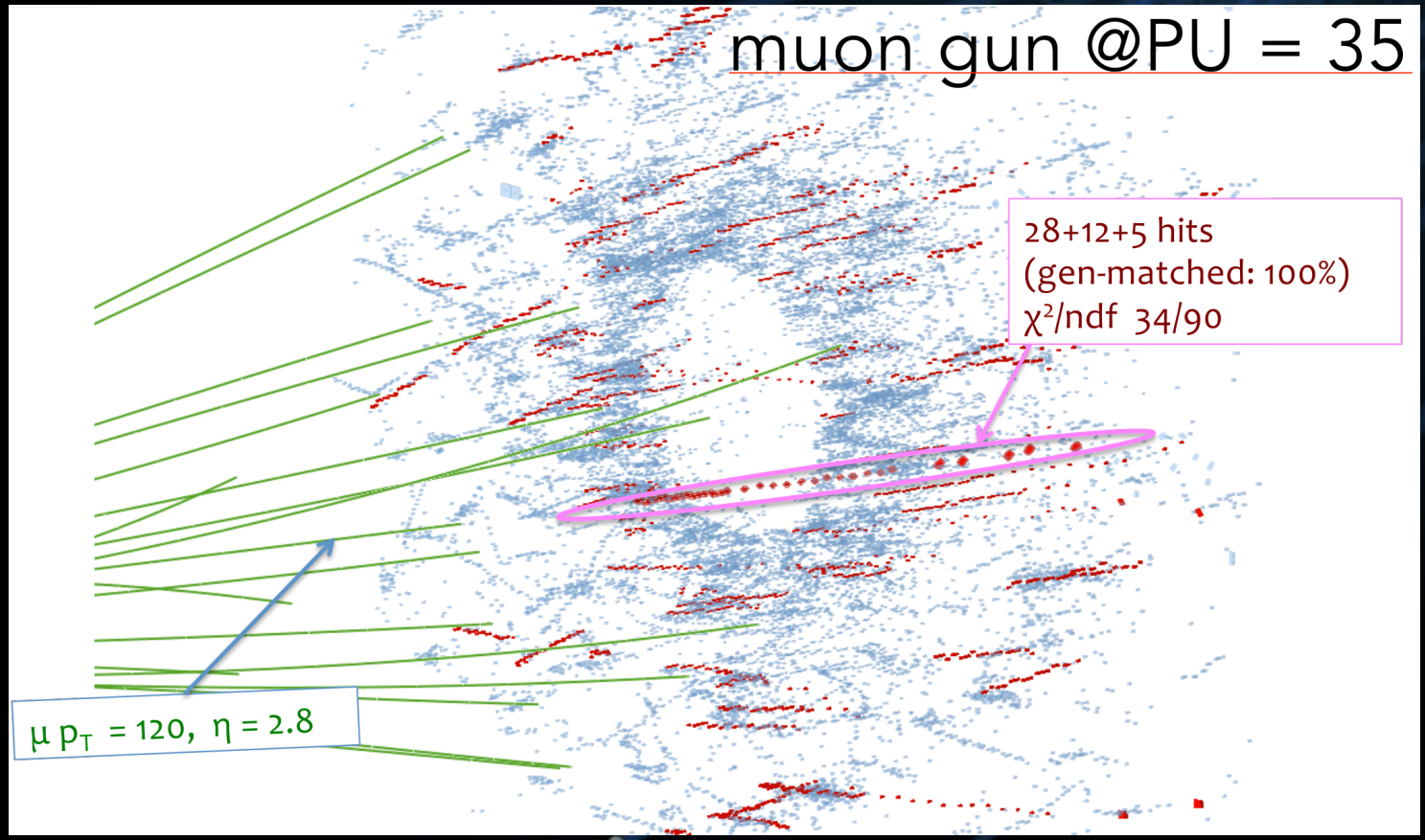
What are the firmware challenges at Phase-II?

- Real-time track-finding and fitting
- Real-time vertex-finding



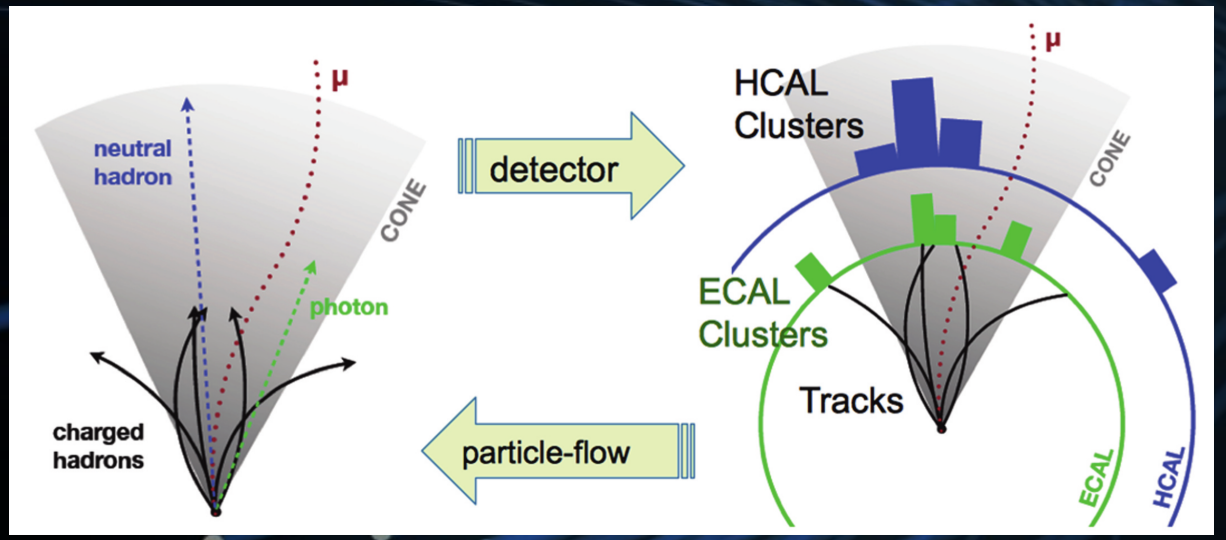
What are the firmware challenges at Phase-II?

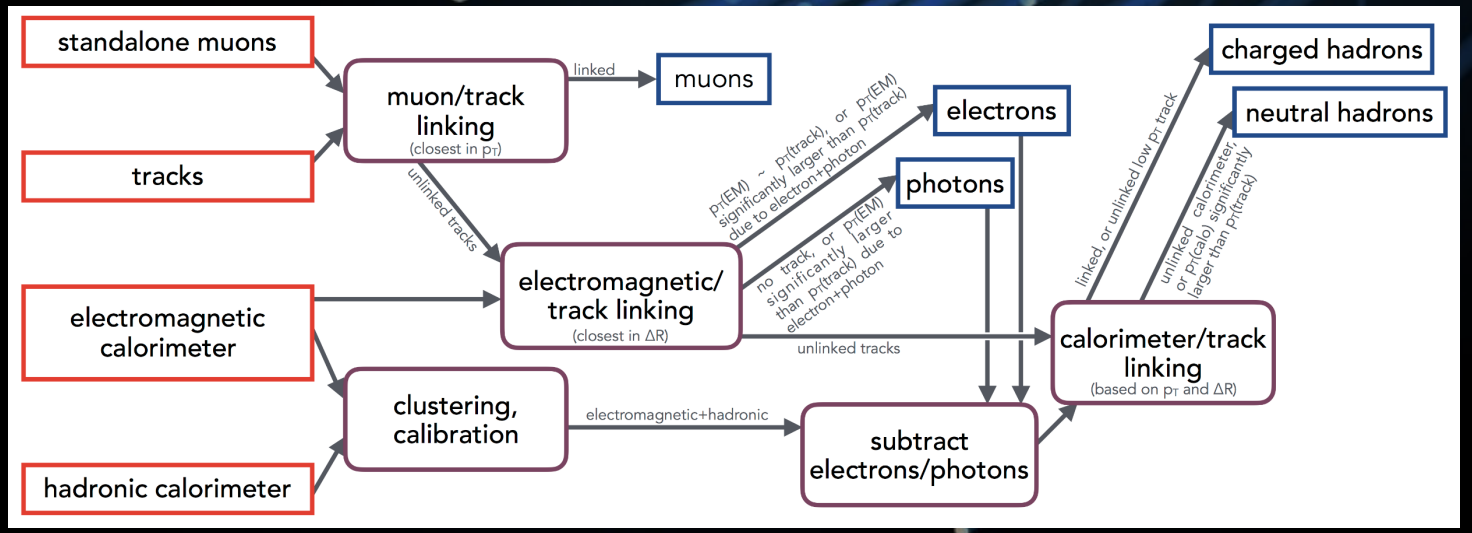
- Real-time track-finding and fitting
- Real-time vertex-finding
- 3D cluster-finding in endcap



What are the firmware challenges at Phase-II?

- Real-time track-finding and fitting
- Real-time vertex-finding
- 3D cluster-finding in endcap
- Particle-flow





What are the firmware challenges at Phase-II?

- Real-time track-finding and fitting
- Real-time vertex-finding
- 3D cluster-finding in endcap
- Particle-flow
- So, basically, they want event reconstruction

What are the firmware challenges at Phase-II?

- Real-time track-finding and fitting
- Real-time vertex-finding
- 3D cluster-finding in endcap
- Particle-flow
- So, basically, they want event reconstruction
  - in under 10µs

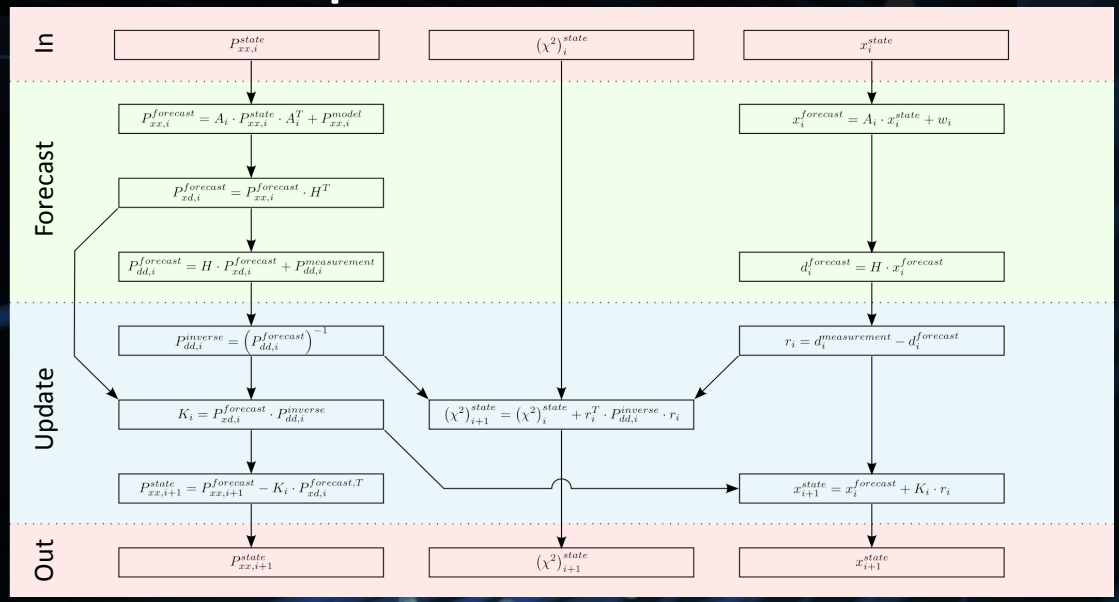
L1 cache reference/hit	1.5	ns		
Floating-point add/mult/FMA operation	1.5	ns		
L2 cache reference/hit	5	ns		
Branch mispredict	6	ns		
L3 cache hit (unshared cache line)	16	ns		
L3 cache hit (shared line in another core)	25	ns		
Mutex lock/unlock	25	ns		
L3 cache hit (modified in another core)	29	ns		
L3 cache hit (on a remote CPU socket)	40	ns		
QPI hop to a another CPU (time per hop)	40	ns		
64MB main memory reference (local CPU)	46	ns		
64MB main memory reference (remote CPU)	70	ns		
256MB main memory reference (local CPU)	75	ns		
256MB main memory reference (remote CPU)	120	ns		
Send 4KB over 100 Gbps HPC fabric	1,040	ns	1 us	
Compress 1KB with Google Snappy	3,000	ns	3 us	
Send 4KB over 10 Gbps ethernet	10,000	ns	10 us	
Write 4KB randomly to NVMe SSD	30,000	ns	30 us	
Transfer 1MB to/from NVLink GPU	30,000	ns	30 us	
Transfer 1MB to/from PCI-E GPU	80,000	ns	80 us	
Read 4KB randomly from NVMe SSD	120,000	ns	120 us	
Read 1MB sequentially from NVMe SSD	208,000	ns	208 us	
Write 4KB randomly to SATA SSD	500,000	ns	500 us	
Read 4KB randomly from SATA SSD	500,000	ns	500 us	
Round trip within same datacenter	500,000	ns	500 us	
Read 1MB sequentially from SATA SSD	1,818,000	ns	1,818 us	2 ms
Read 1MB sequentially from disk	5,000,000	ns	5,000 us	5 ms
Random Disk Access (seek+rotation)	10,000,000	ns	10,000 us	10 ms
Send packet CA->Netherlands->CA	150,000,000	ns	150,000 us	150 ms

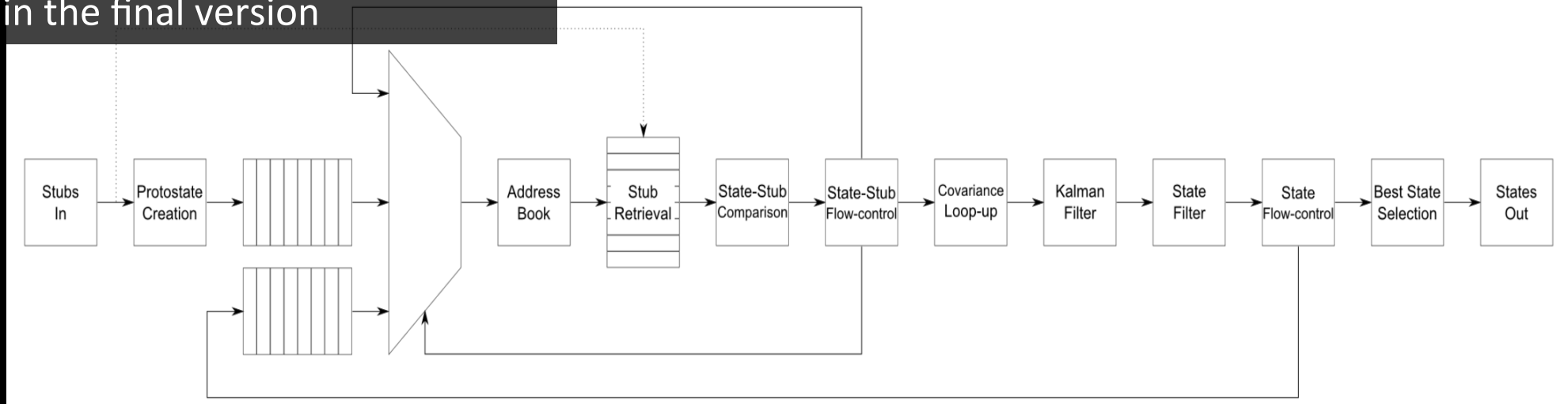
What are the firmware challenges at Phase-II?

- I used to give the following firmware advice to the students on preserving your sanity:
  - Avoid iterative algorithms
  - Avoid combinatorics
  - Make the data-flow deterministic
  - Division is hard, resource hungry and latency intensive
  - Floating-point is hard, resource hungry and latency intensive

And we have stepped up to the plate!

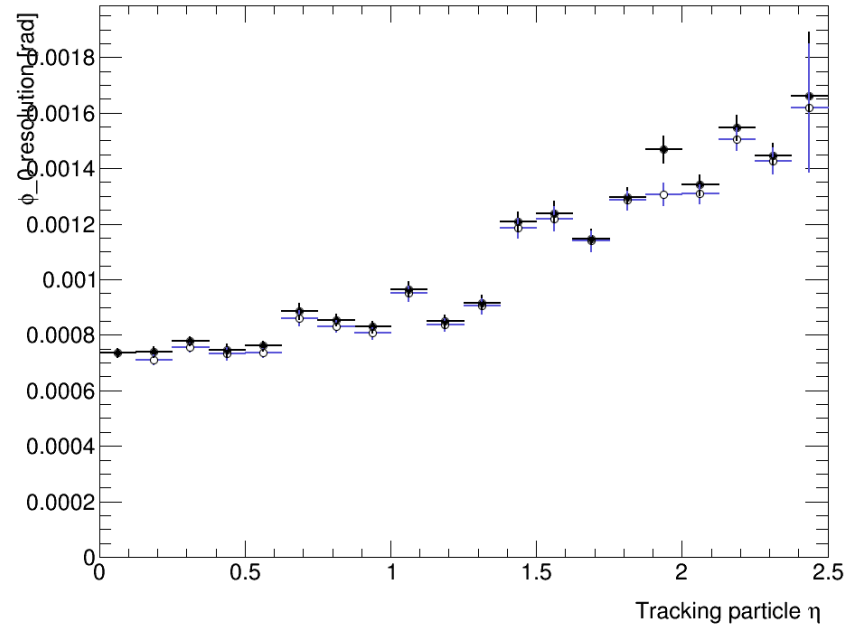
- For track fitter we wrote a full Kalman Filter
  - Which is iterative
  - Results in combinatorics
  - Is data-dependent (pseudo-deterministic)
  - Requires a division
  - We also played with floating-point but dropped it in the final version

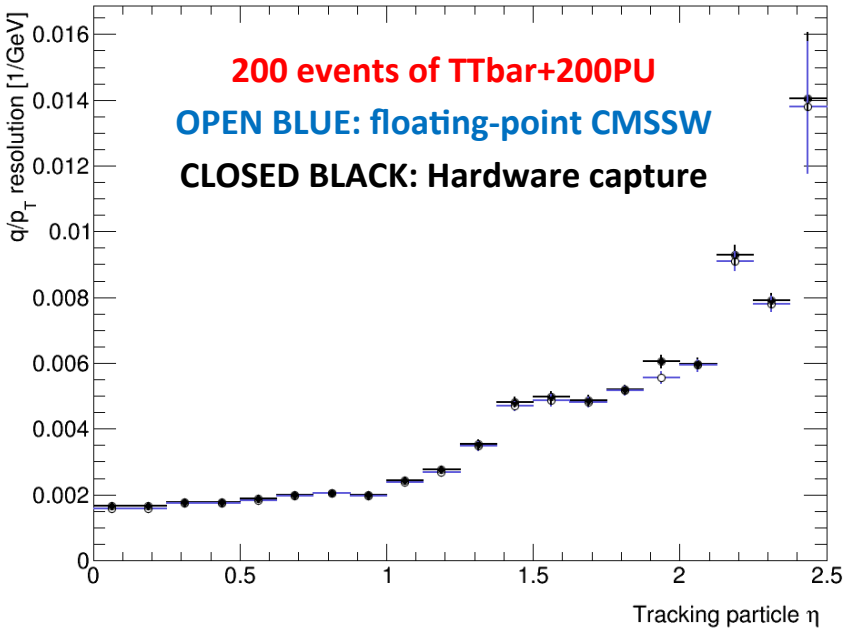


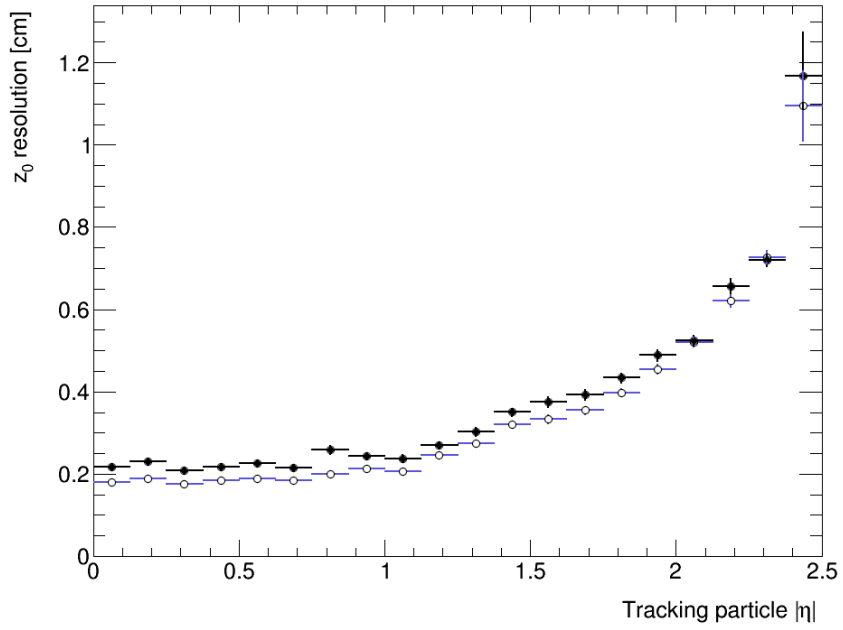


And we have stepped up to the plate!

- Fits a 4-point track in 1.5µs
- Matches offline floating-point resolution

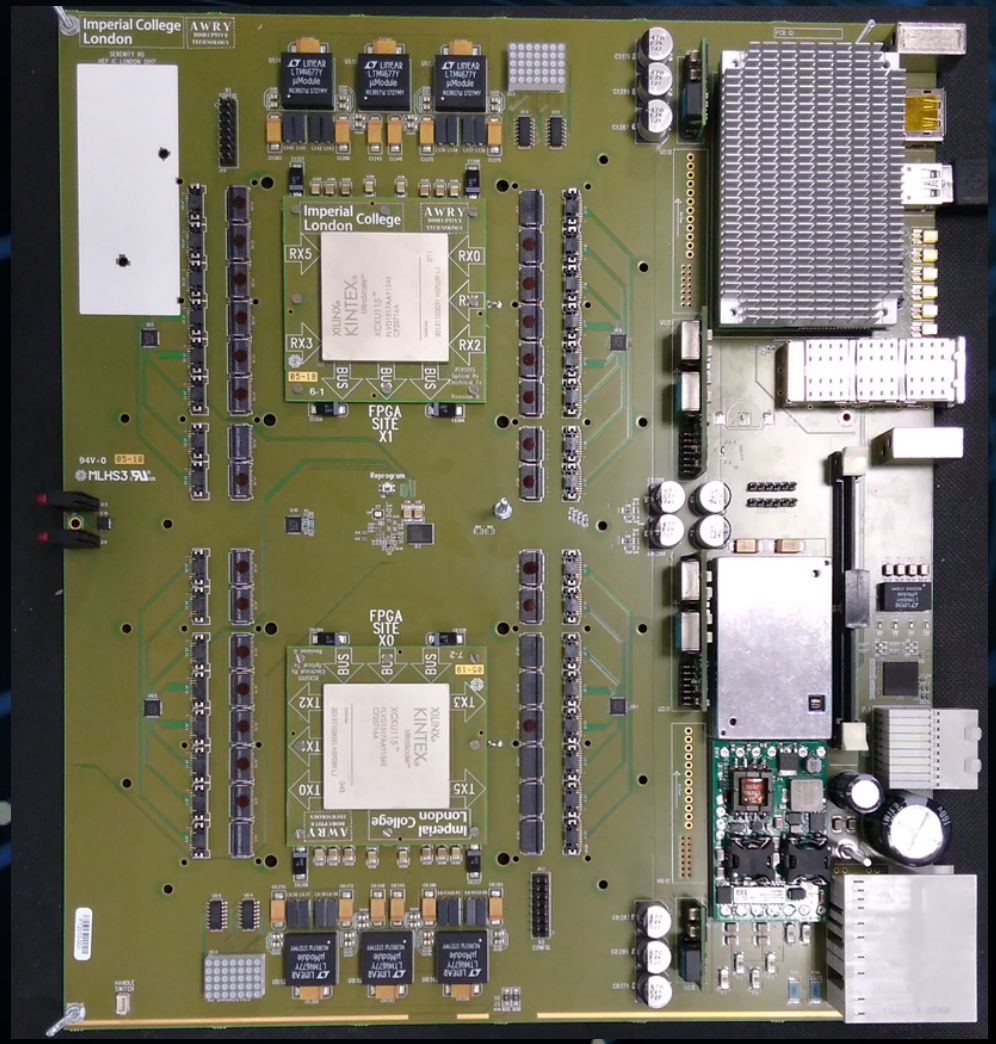






STOP PRESS!

- UK claims crown for 1st Phase-2 algorithm in Phase-2 hardware
- Real-time vertex-finding
  - 4% LUTS, ~2% RAM & DSPs in KU115 FPGA
  - 240MHZ clock
  - TM-period + 20clk latency
- Running in Serenity hardware



Conclusion

- UK has been very dominant in the phase-1 changes to the CMS off-detector electronics
  - Hardware, System Architecture, Firmware & Control Software
  - More elegant systems which are more robust, more performant and easier to maintain
- UK-originated ideas gaining traction for phase 2
  - Time-multiplexing
  - Common hardware
  - Infrastructure firmware
- Phase-2 upgrades introduce a whole new set of detector and trigger challenges
  - UK stepping up with state-of-the-art hardware and firmware solutions

Bring on the future!

Thanks for listening!

Any questions?