ECFA

European Committee for Future Accelerators

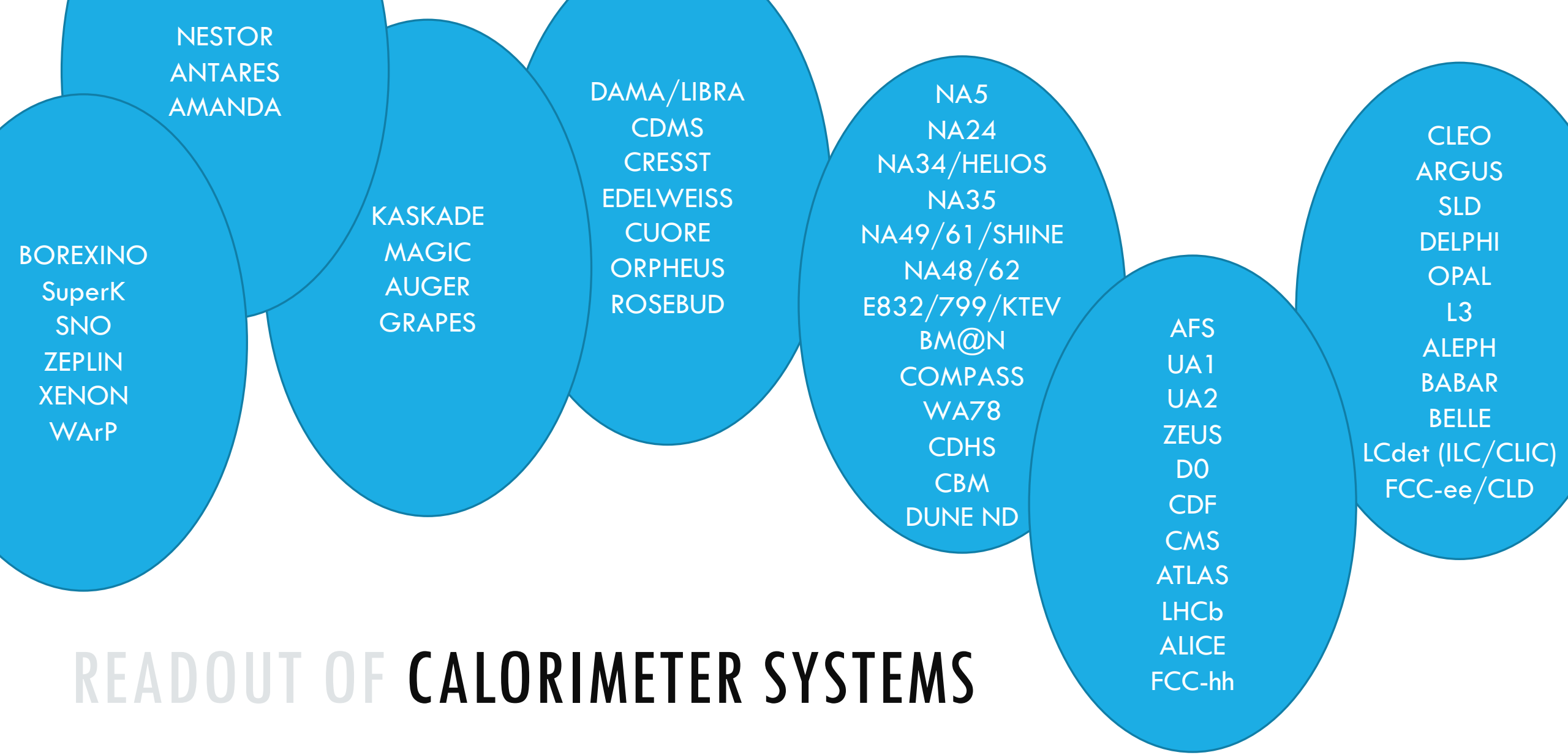
READOUT OF CALORIMETER SYSTEMS

A. David (CERN)

READOUT OF CALORIMETER SYSTEMS

A. David (CERN)

Beware omissions or inaccuracies in surveying such a cross-cutting topic.

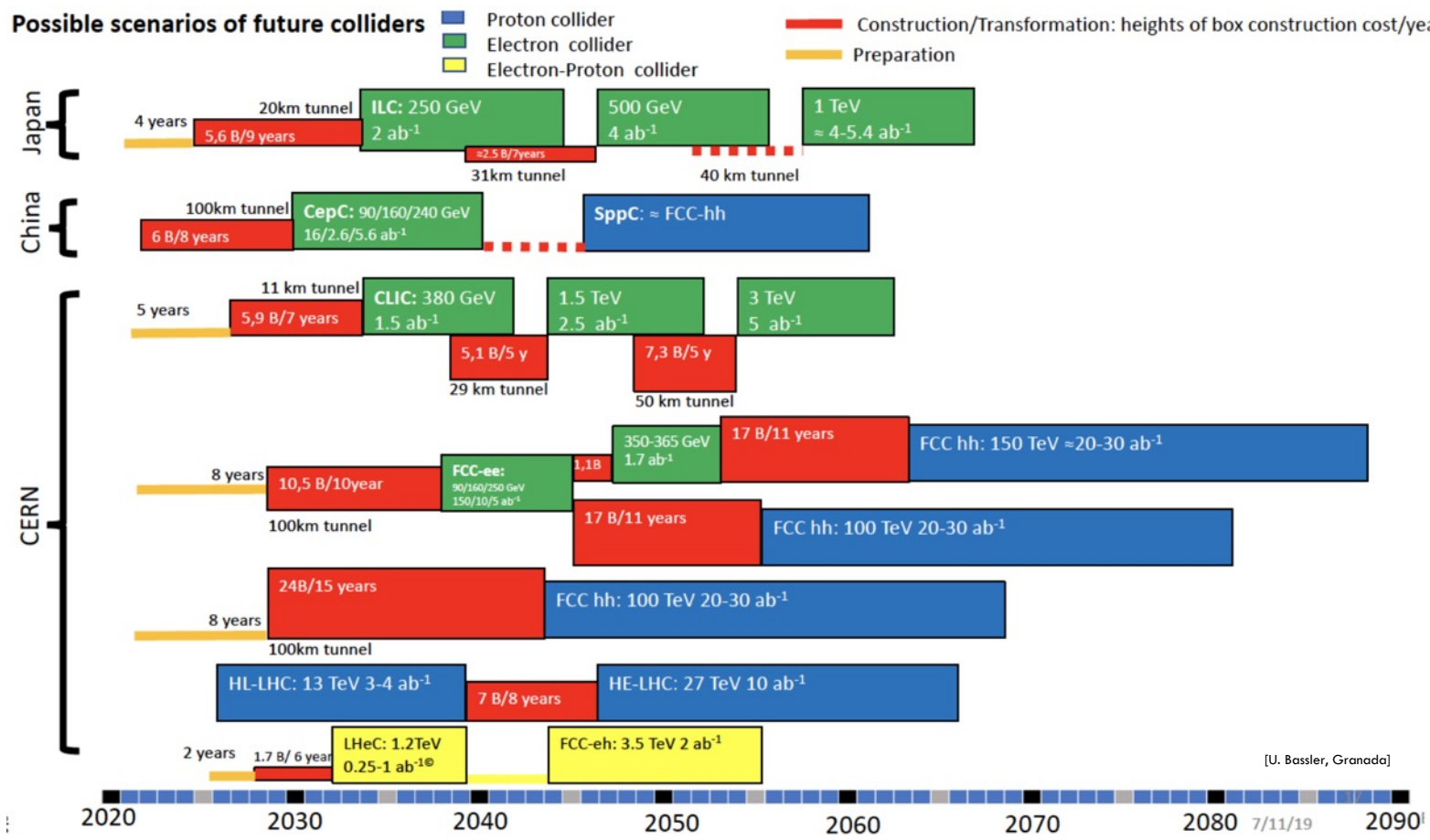


READOUT OF CALORIMETER SYSTEMS

SWITCHING **COLLIDER** GEARS

In 20 years

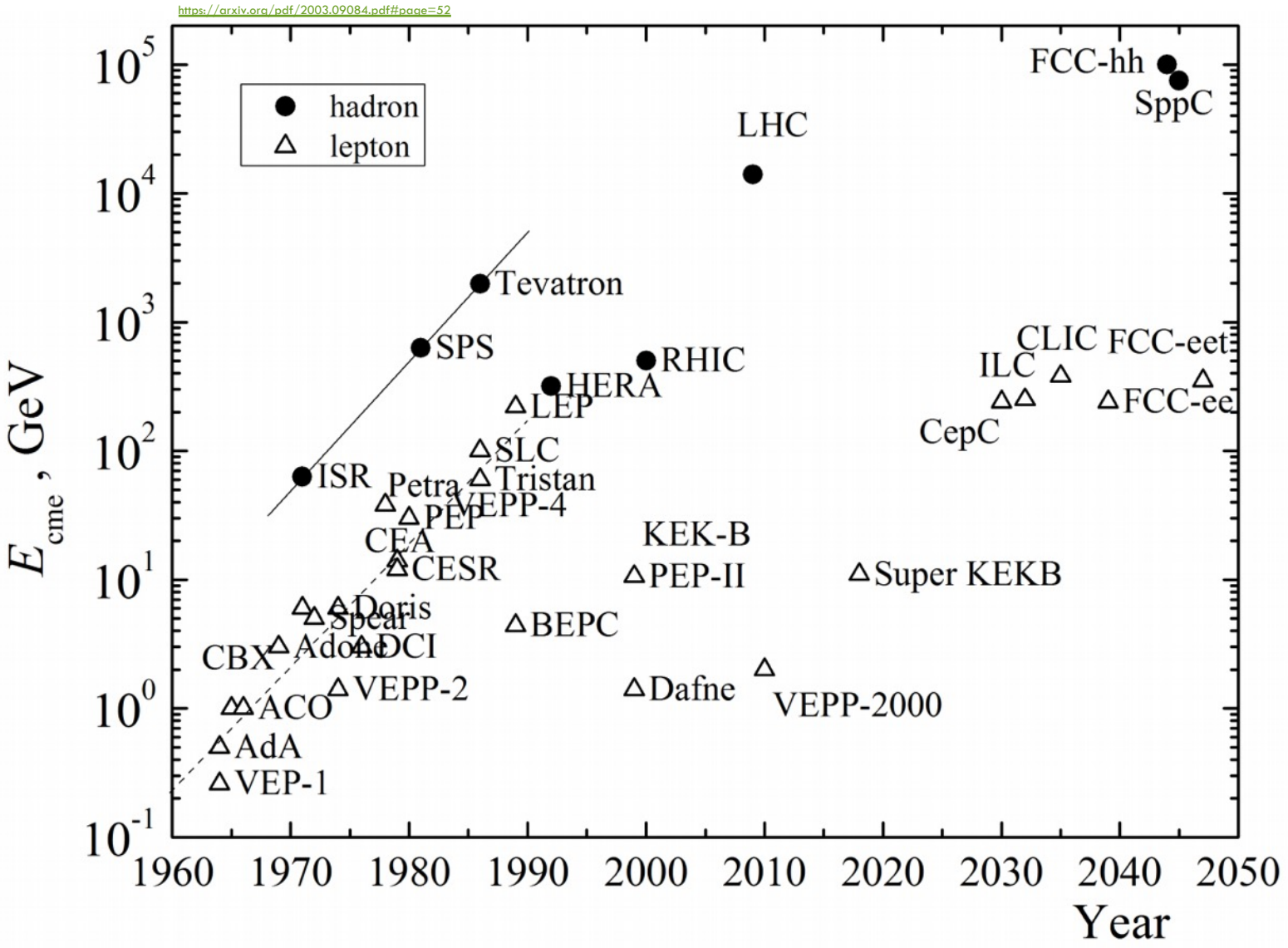
- HL-LHC is over, long live HL-LHC.
- Electron-positron Higgs factory starting/in operation.
- Hadron-hadron energy frontier in preparation.



SWITCHING **COLLIDER** GEARS

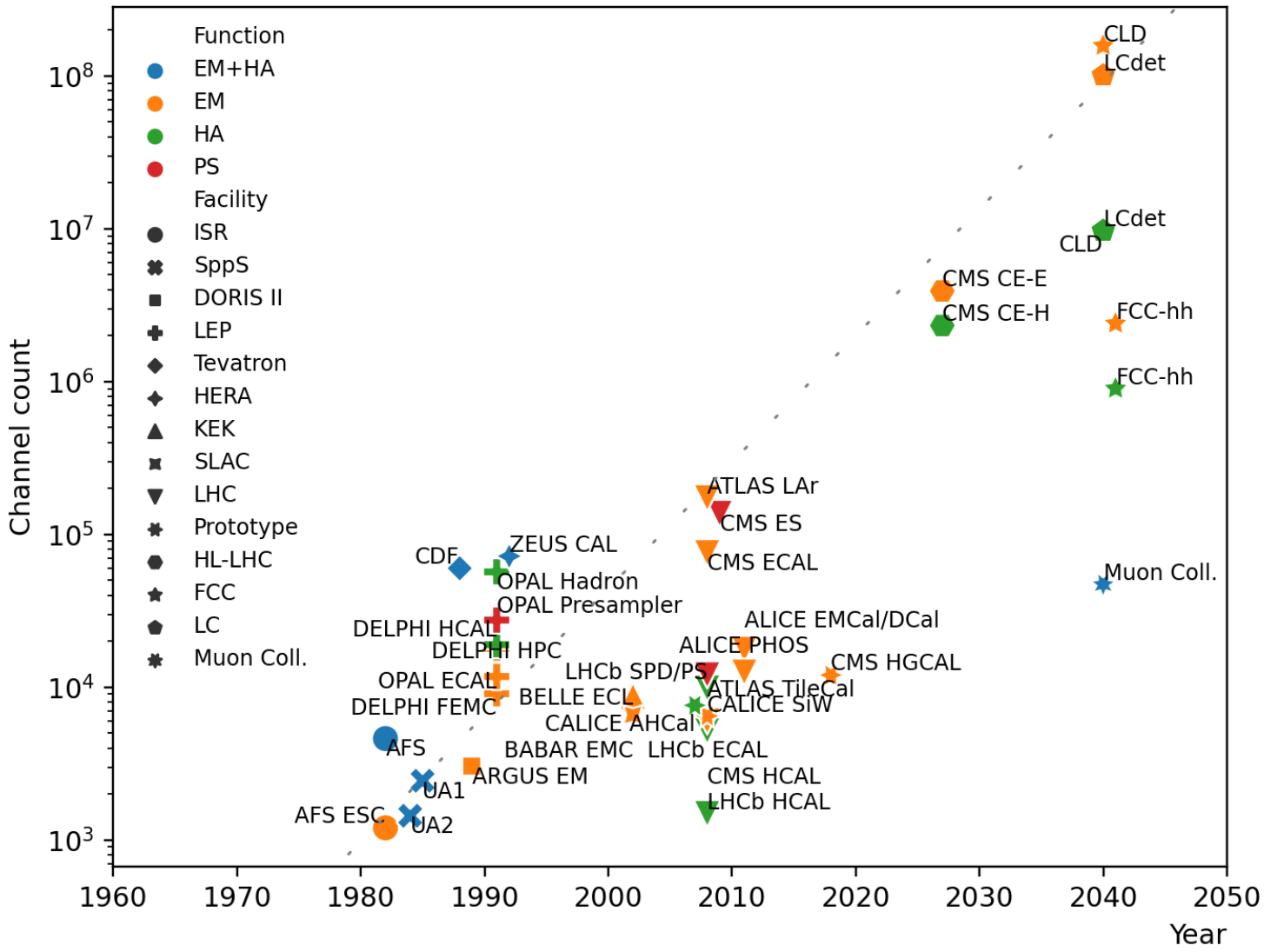
In 20 years

- HL-LHC is over, long live HL-LHC.
- Electron-positron Higgs factory starting/in operation.
- Hadron-hadron energy frontier in preparation.



SWITCHING **CALORIMETER** GEARS

In the last 40 years detectors **grew in size.**



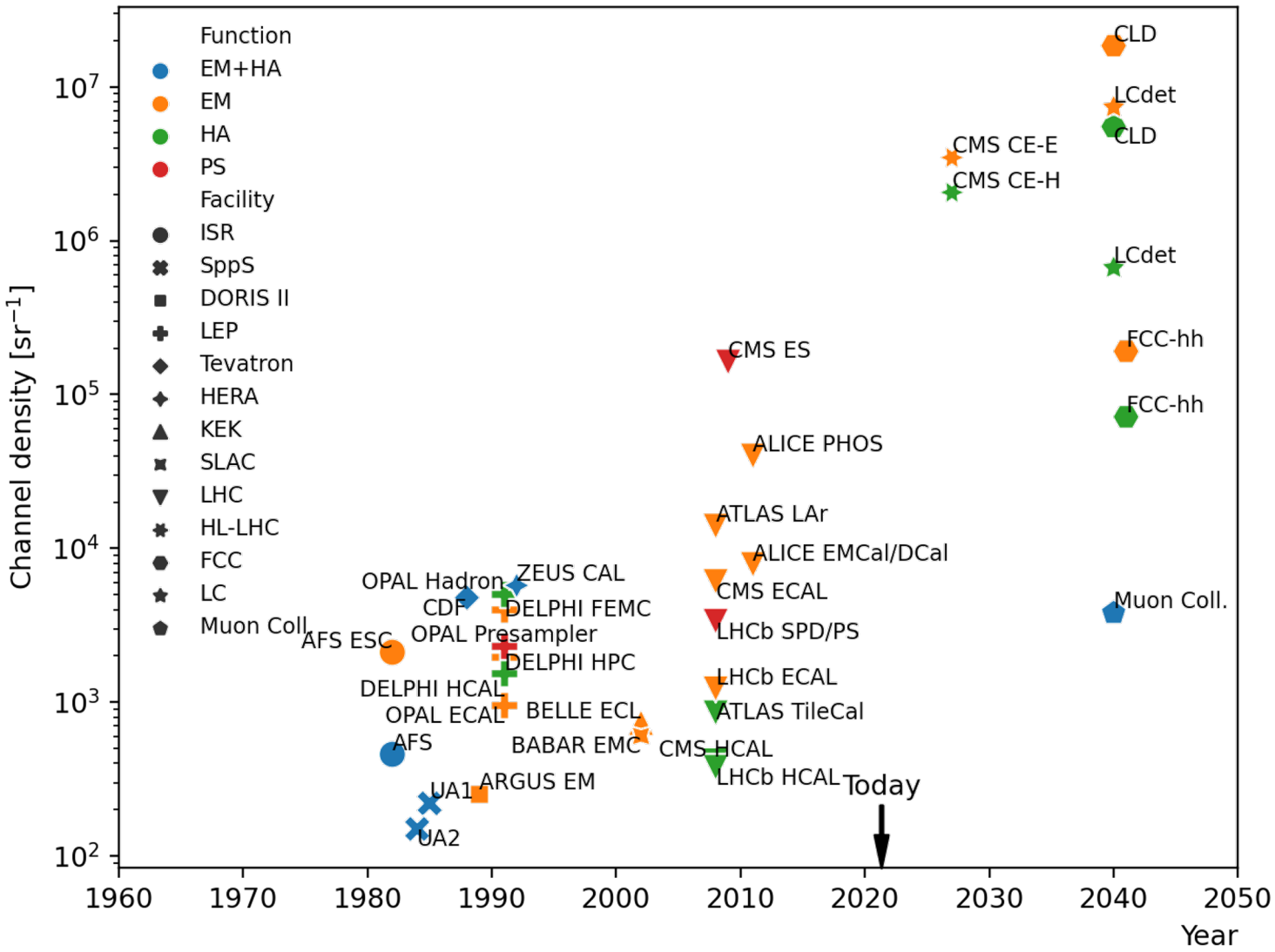
Beware omissions or inaccuracies in surveying such a cross-cutting topic.

SWITCHING **CALORIMETER** GEARS

In the last 40 years detectors grew in size.

In the next 20 years **they will grow in:**

- spatial density and/or
- timing resolution.

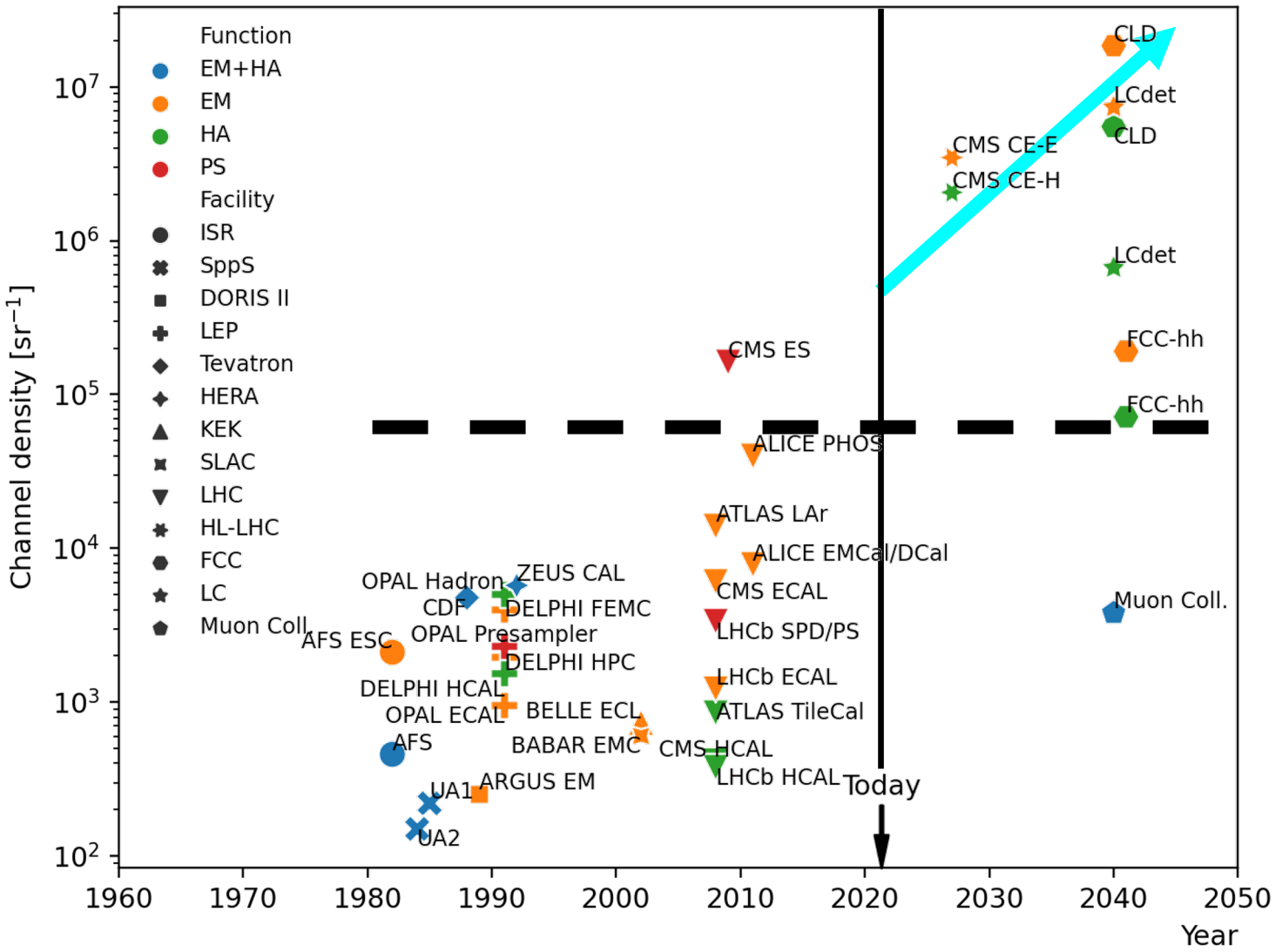


SWITCHING **CALORIMETER** GEARS

In the last 40 years detectors grew in size.

In the next 20 years **they will grow in:**

- **spatial density** and/or
- timing resolution.



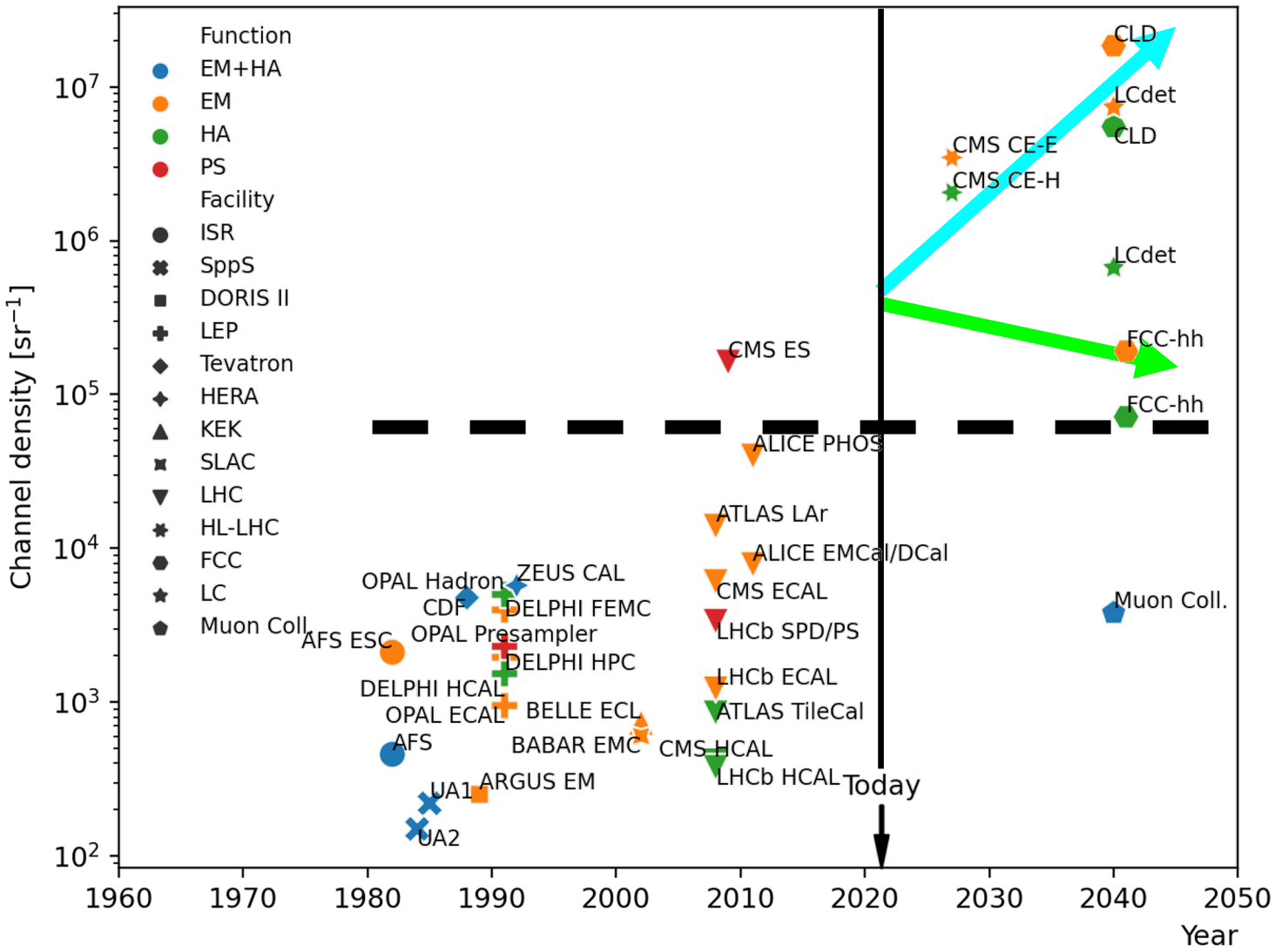
SWITCHING **CALORIMETER** GEARS

In the last 40 years detectors grew in size.

In the next 20 years **they will grow in:**

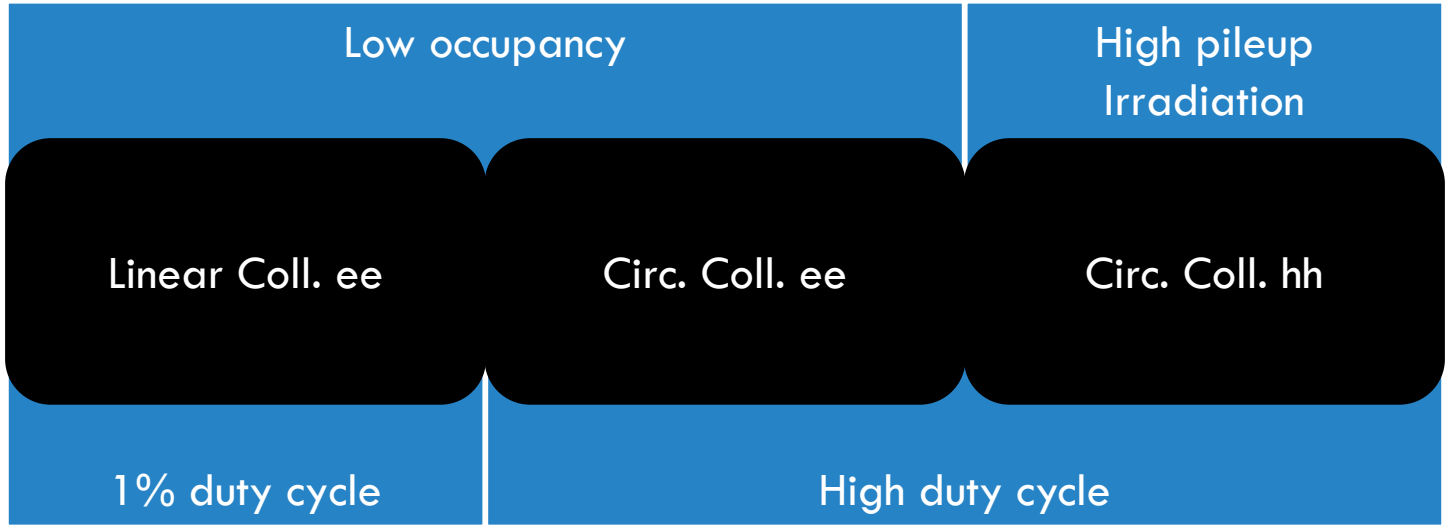
- **spatial density** and/or
- **timing resolution**.

Needed for **next-generation particle flow 5D reconstruction** in space, energy, and time.

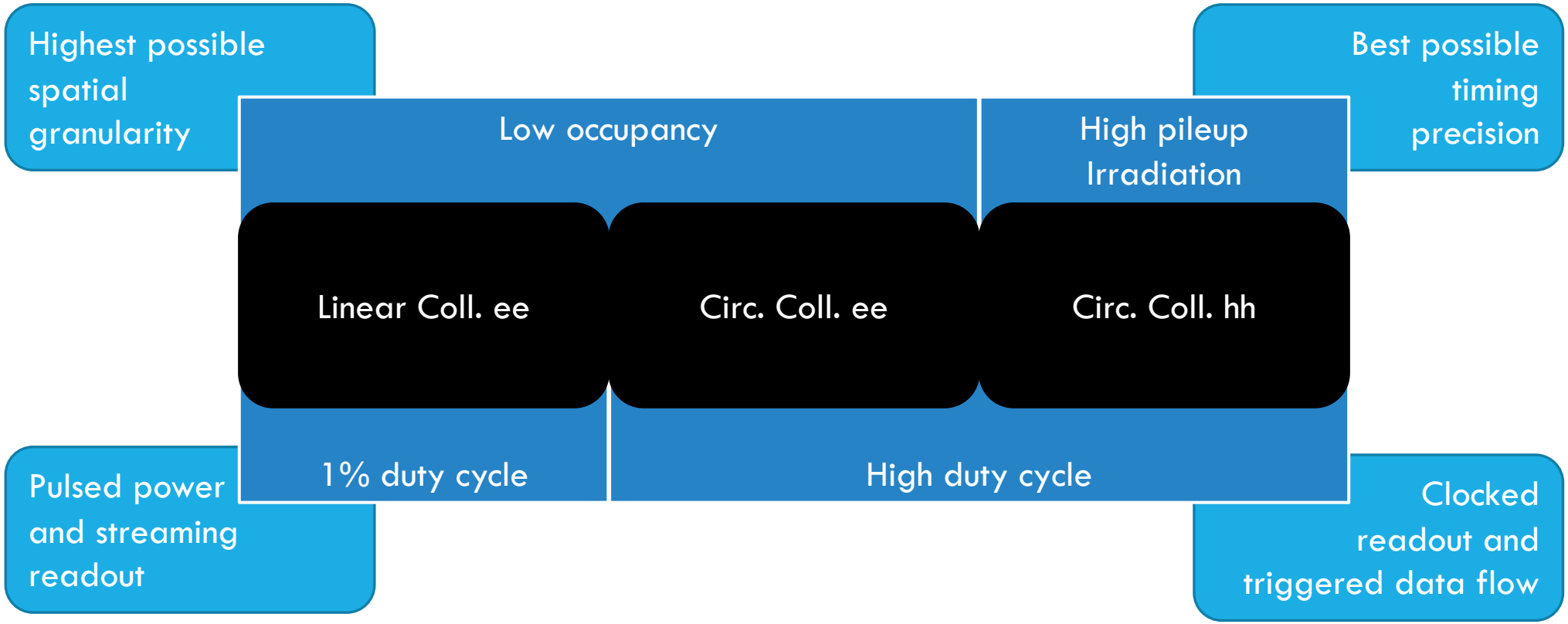


SPATIAL GRANULARITY **AND** PRECISE TIMING

MAIN DRIVERS



MAIN CONSEQUENCES



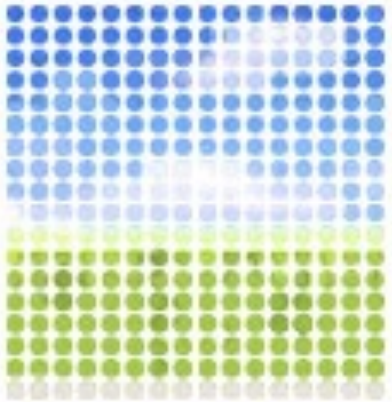
ONE TRACK, TWO TRACKS, . . . , A SHOWER

Low occupancy environment

https://www.pn-design.co.uk/guide-to-resolution-and-dpi-issues/

Highest **spatial** granularity down to individual tracks.

High pileup environment

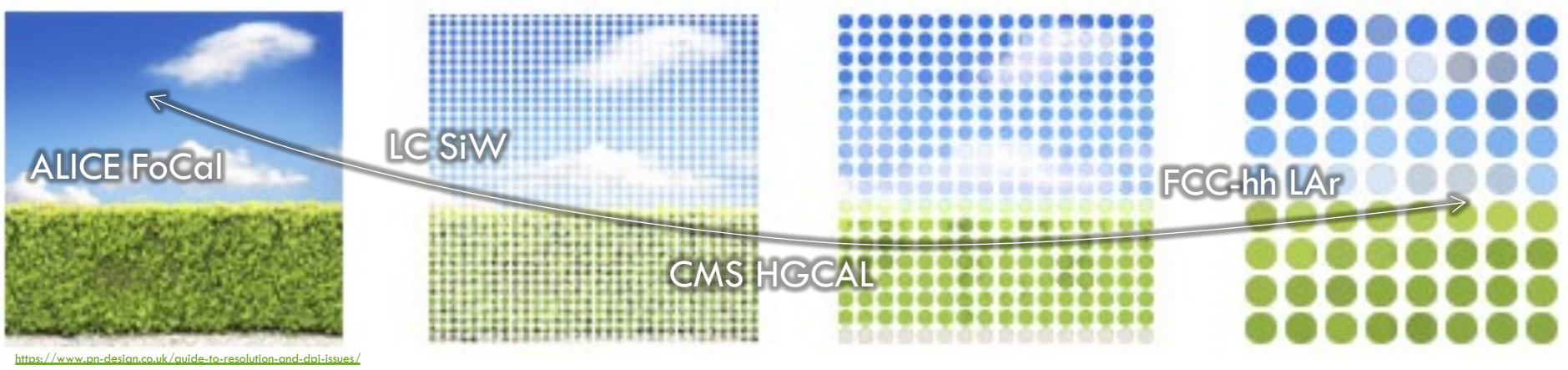


Energy from multiple tracks and mandatory precise **timing**.

ONE TRACK, TWO TRACKS, . . . , A SHOWER

Low occupancy environment

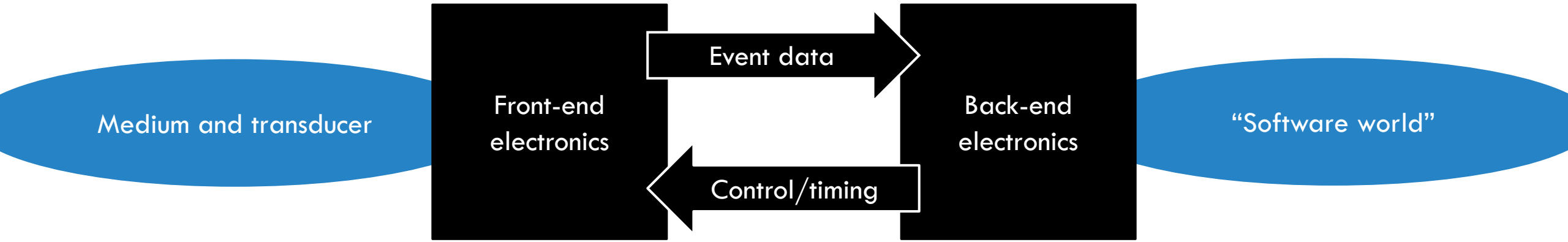
High pileup environment



https://www.pn-design.co.uk/guide-to-resolution-and-dpi-issues/

Highest **spatial** granularity down to individual tracks.

Energy from multiple tracks and mandatory precise **timing**.



READOUT OF CALORIMETER SYSTEMS

LARGE DYNAMIC RANGES

Channel density variations

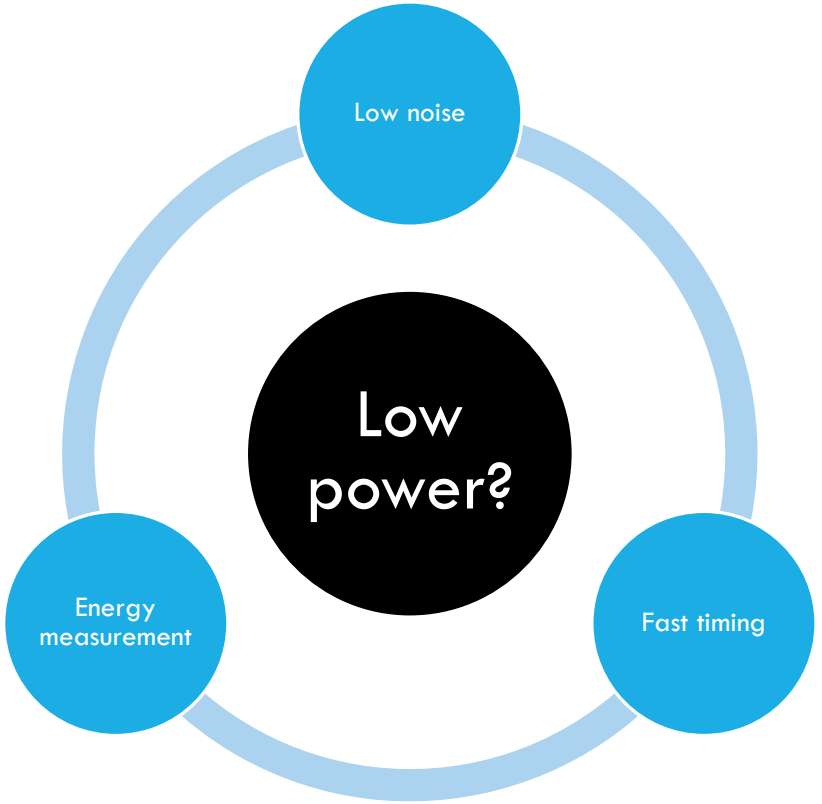
- E.g., EM vs Hadr. segmentation.
- **Challenge**: data load-balancing, same ASIC for different environments (in same detector)?

Energy measurement with tracker-like noise

- E.g., <1 MIP to O(10^3) MIPs.
- **Challenge**: analogue amplifiers and (fast) shapers?

Can it be kept within the **power budget?**

Multiple common key IPs in ASICs. [TF7]



LARGE DYNAMIC RANGES

Channel density variations

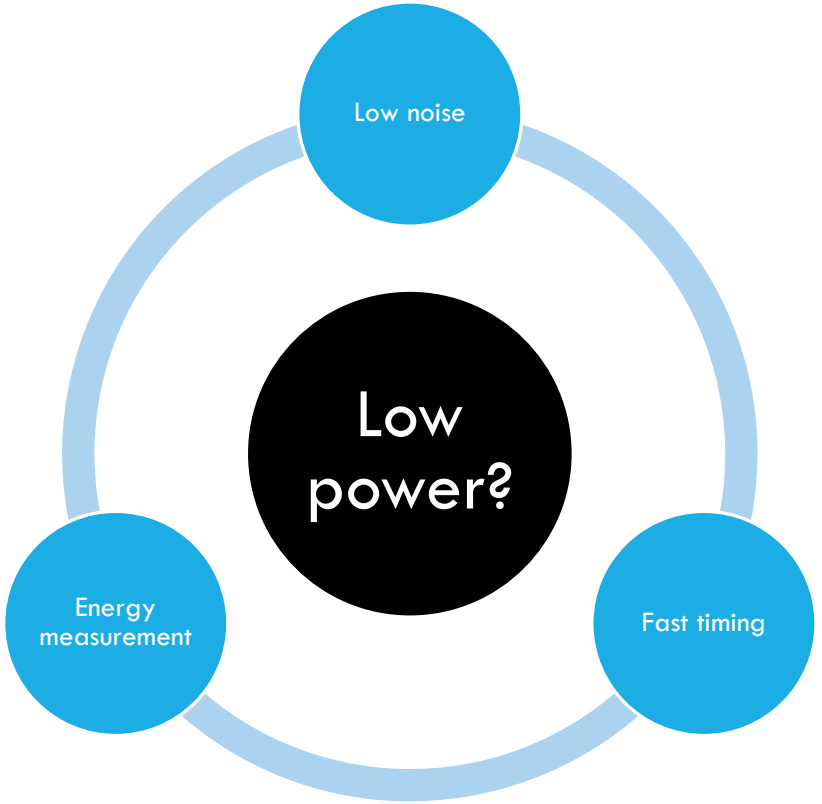
- E.g., EM vs Hadr. segmentation.
- **Challenge**: data load-balancing, same ASIC for different environments (in same detector)?

Energy measurement with tracker-like noise

- E.g., <1 MIP to O(10^3) MIPs.
- **Challenge**: analogue amplifiers and (fast) shapers?

Can it be kept within the **power budget?**

Multiple common key IPs in ASICs. [TF7] →



HEP ASICs are very specific systems fabricated in mainstream technologies built out of very common critical blocks (ADCs, TDCs, PLLs, DLLs, Power converters, ser-des, etc..)

https://indico.cern.ch/event/1001692/contributions/4215289/attachments/2215392/3750382/ECFA_ASIC_Rivetti.pdf

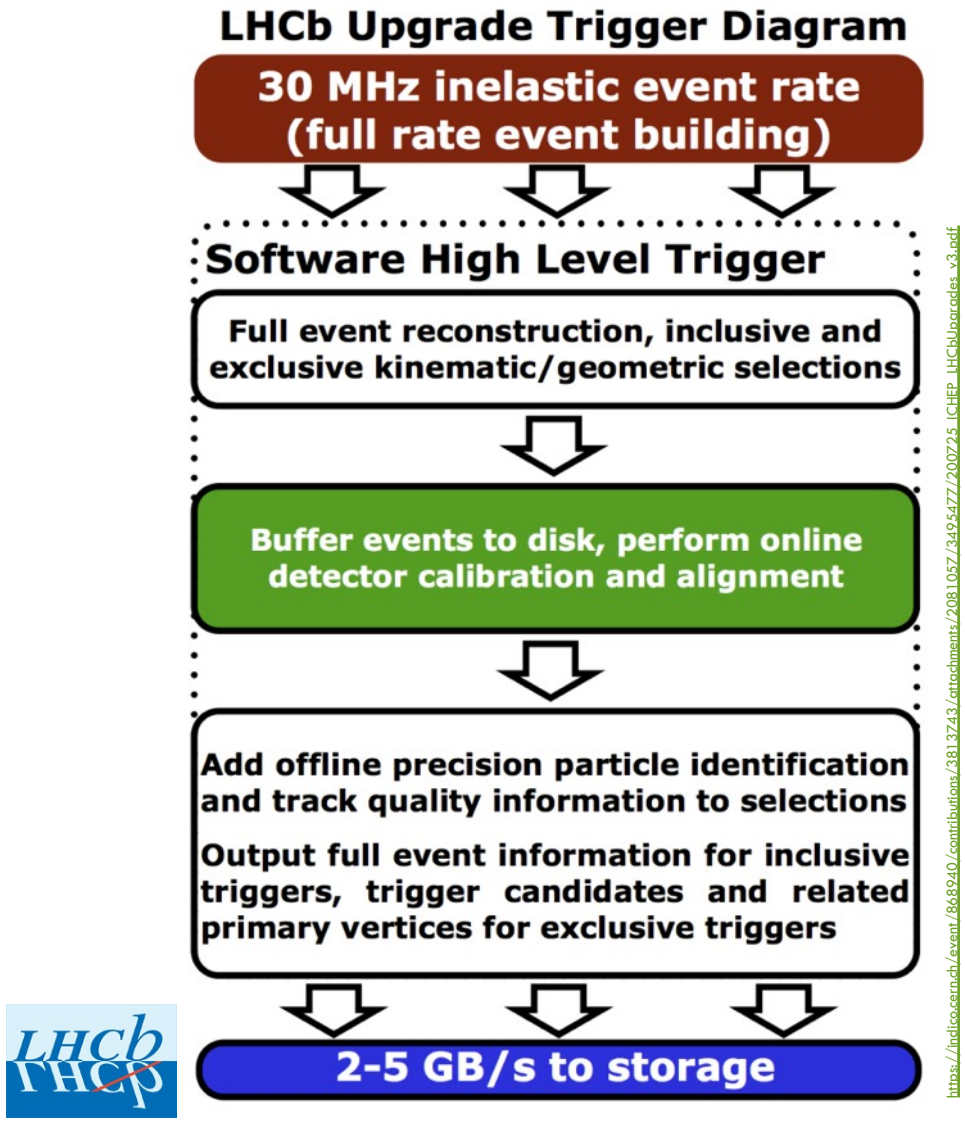
COLLIDER DUTY CYCLE

Fundamental impact on readout systems

- ILC L upgrade: 366 ns (at 5 Hz).
- FCC-ee: 20 ns (Z^0) to 3.4 μs ($t\bar{t}$).

Deep consequences

- **Data flow**: triggered vs streaming.
  - Feasibility depends heavily on:
    - Occupancy/size and
    - Information processing/compression capabilities.
  - LHCb putting streaming to work. →
- **Powering**: pulsed vs continuous.
  - Important integration constraints.



COLLIDER DUTY CYCLE

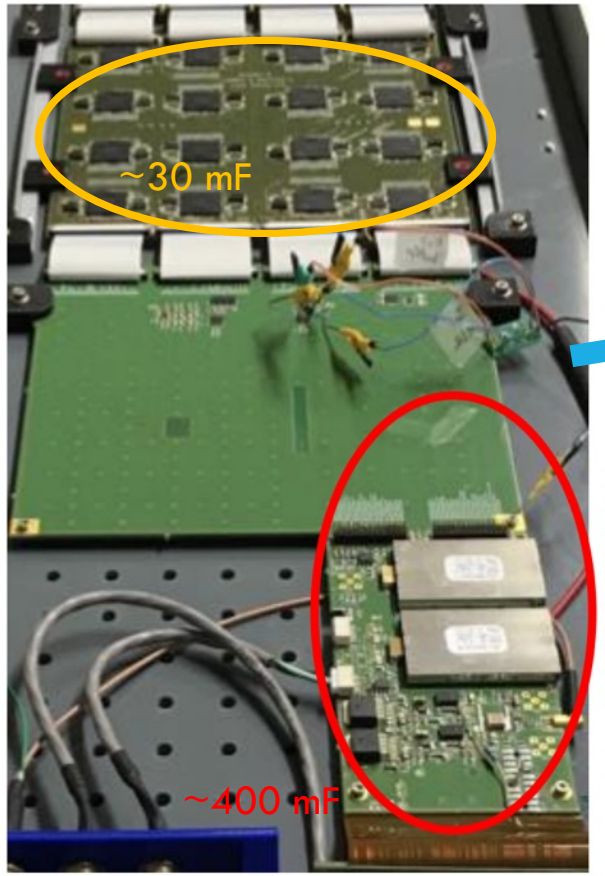
Fundamental impact on readout systems

- ILC L upgrade: 366 ns (at 5 Hz).
- FCC-ee: 20 ns (Z^0) to 3.4 µs ($t\bar{t}$).

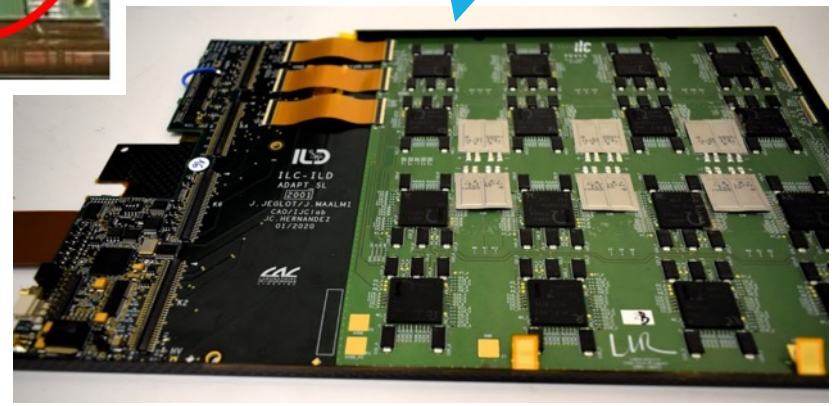
Deep consequences

- **Data flow**: triggered vs streaming.
  - Feasibility depends heavily on:
    - Occupancy/size and
    - Information processing/compression capabilities.
  - LHCb putting streaming to work.
- **Powering**: pulsed vs continuous.
  - Important integration constraints. ↗

https://indico.cern.ch/event/818783/contributions/3598490/attachments/1952892/3242642/CHEF_Novembre_2019.pdf



Capacitor type and location impact ASIC performance



https://indico.cern.ch/event/932973/contributions/4062109/attachments/2140570/3606706/CALICETechnologies.pdf

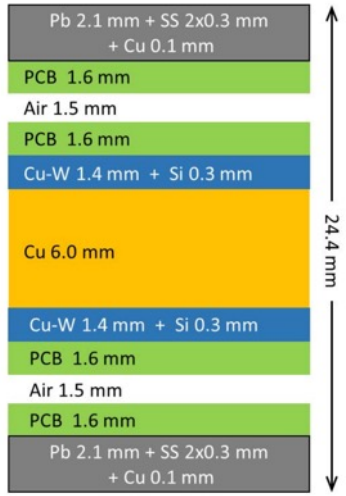
ON-DETECTOR INTEGRATION

Showering material takes space

- Connectors, coils, capacitors, boards. →
- Power distribution, heat management.
- For Si: electronics in sensor. [TF3]

Layer-to-layer information/power flow?

https://cds.cern.ch/record/2293646/files/CMS-TDR-019.pdf



System challenges

- **Simulation for validation.**
  - In-ASIC, across-ASICs, signals in PCBs.
- Operational issues, recovery processes.
- Challenge: recruit and **retain people**!

CMS HGCAL: *How it started*

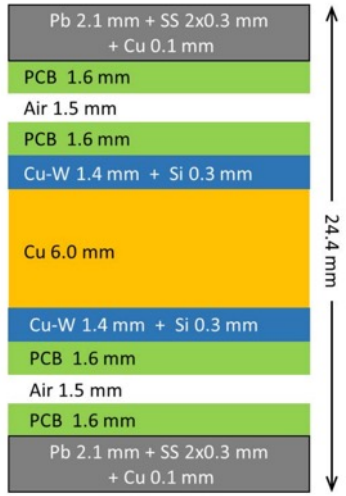
ON-DETECTOR INTEGRATION

Showering material takes space

- Connectors, coils, capacitors, boards. →
- Power distribution, heat management.
- For Si: electronics in sensor. [TF3]

Layer-to-layer information/power flow?

https://cds.cern.ch/record/2293646/files/CMS-TDR-019.pdf

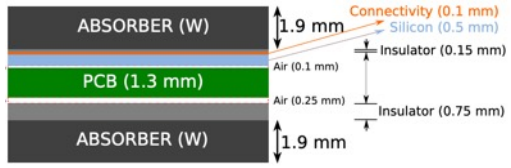


CMS HGCAL: *How it started*

System challenges

- **Simulation for validation.**
  - In-ASIC, across-ASICs, signals in PCBs.
- Operational issues, recovery processes.
- Challenge: recruit and **retain people**!

CLD (2019)



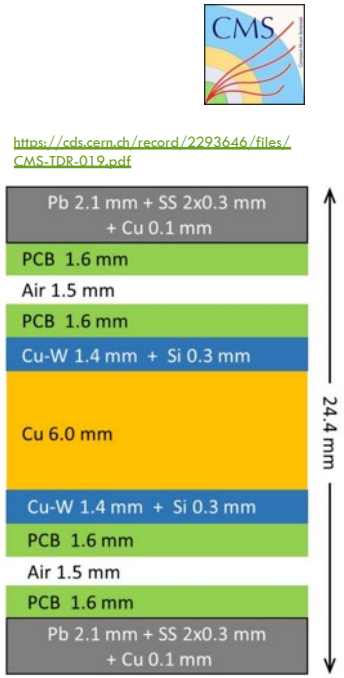
https://arxiv.org/pdf/1911.12230.pdf

ON-DETECTOR INTEGRATION

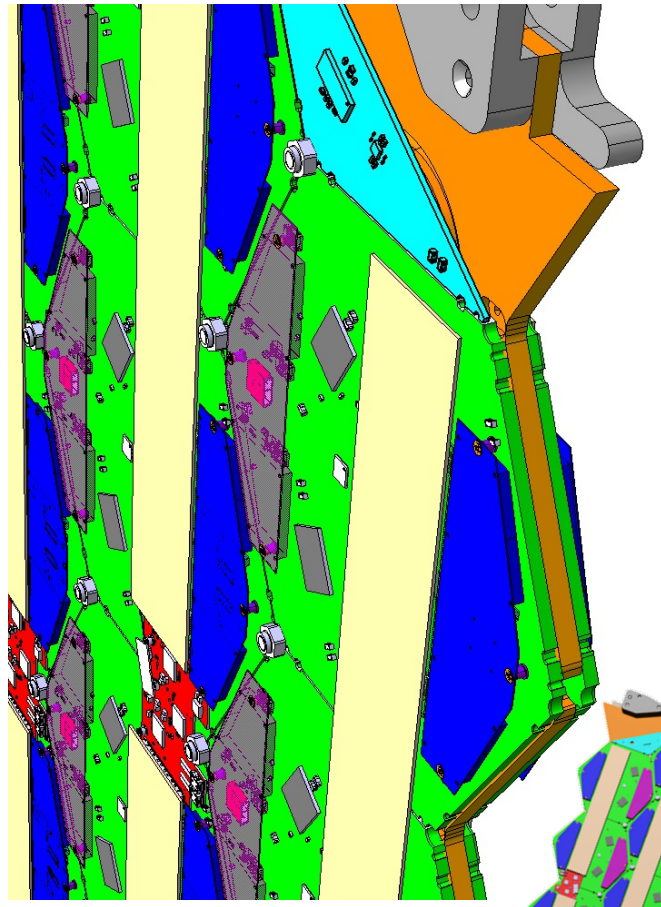
Showering material takes space

- Connectors, coils, capacitors, boards. →
- Power distribution, heat management.
- For Si: electronics in sensor. [TF3]

Layer-to-layer information/power flow?

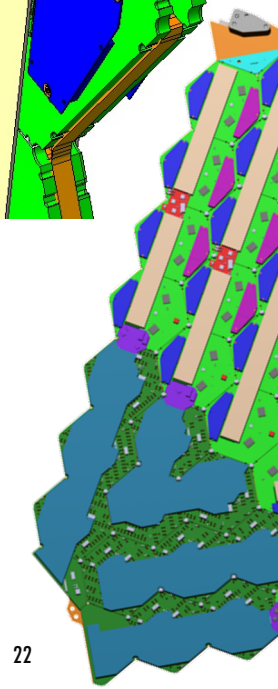


[Courtesy K. Rapacz]



CMS HGCAL: *How it started*

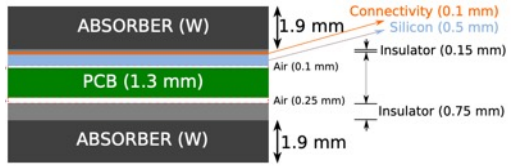
How it is going



System challenges

- **Simulation for validation.**
  - In-ASIC, across-ASICs, signals in PCBs.
- Operational issues, recovery processes.
- Challenge: recruit and **retain people**!

CLD (2019)

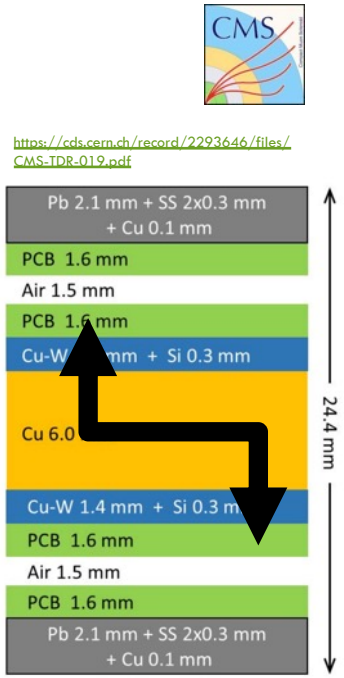


ON-DETECTOR INTEGRATION

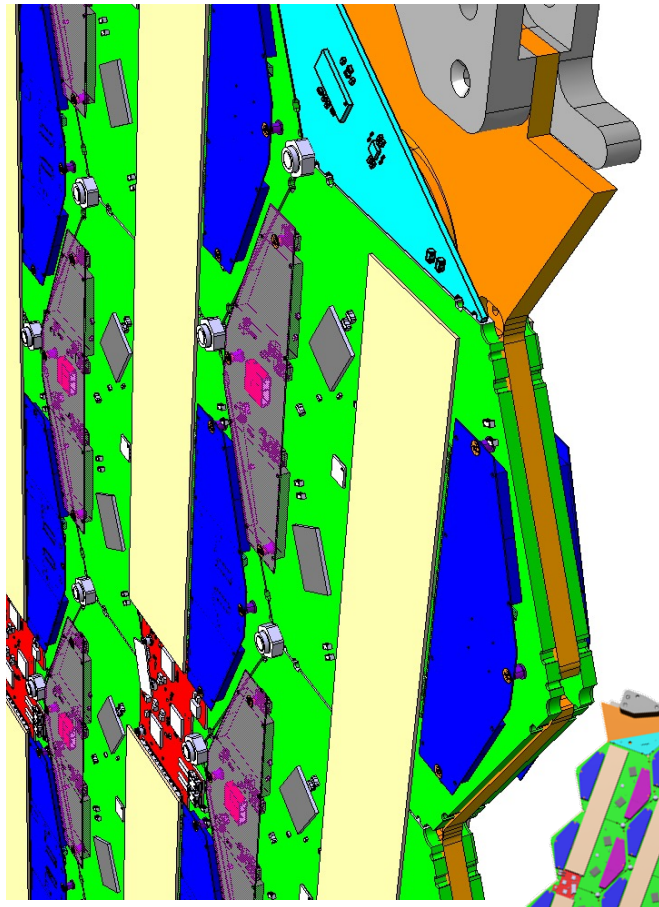
Showering material takes space

- Connectors, coils, capacitors, boards. →
- Power distribution, heat management.
- For Si: electronics in sensor. [TF3]

Layer-to-layer information/power flow?



[Courtesy K. Rapacz]

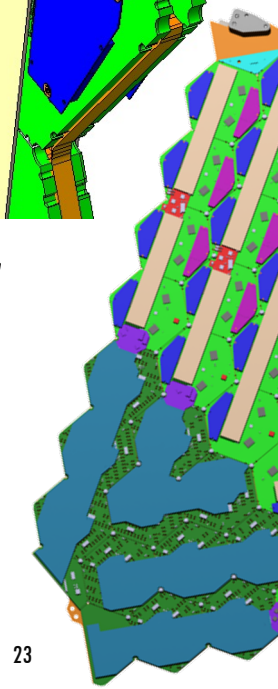


System challenges

- **Simulation for validation.**
  - In-ASIC, across-ASICs, signals in PCBs.
- Operational issues, recovery processes.
- Challenge: recruit and **retain people**!

CMS HGCAL: *How it started*

How it is going



5D RECONSTRUCTION AND CALIBRATION

Space, time, and energy information contribute to the overall picture.

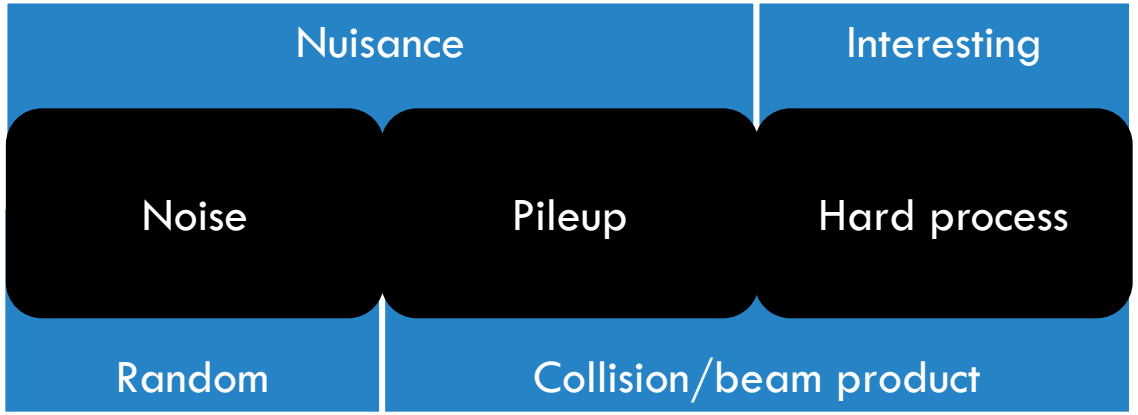
- **Novel algorithms** needed.

Pileup is not noise, just physics we're not interested in

- Getting the most information about the interesting process requires identifying all three components.
- Opportunities for more processing on-detector, beyond noise rejection.

Calibration also a processing algorithm

- Rates and data formats crucial.
- Streaming skips reprocessing, saves offline computing.



5D RECONSTRUCTION AND CALIBRATION

Space, time, and energy information contribute to the overall picture.

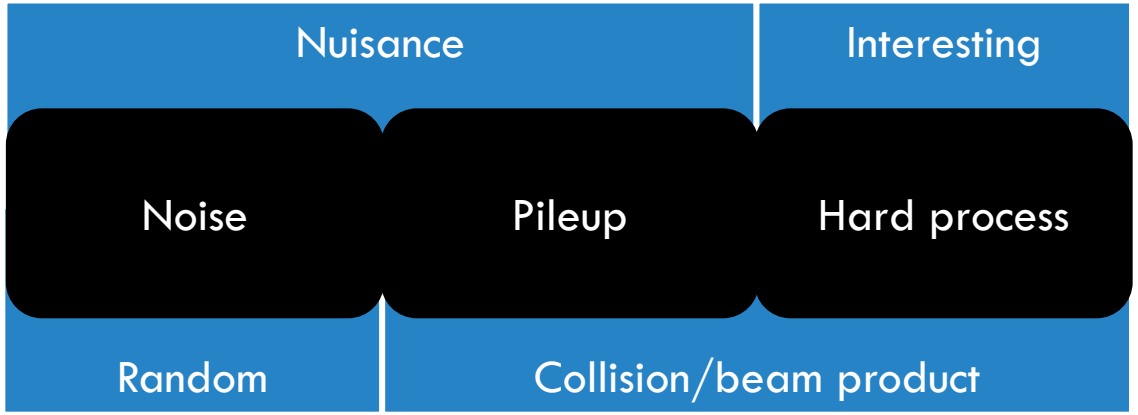
- **Novel algorithms** needed.

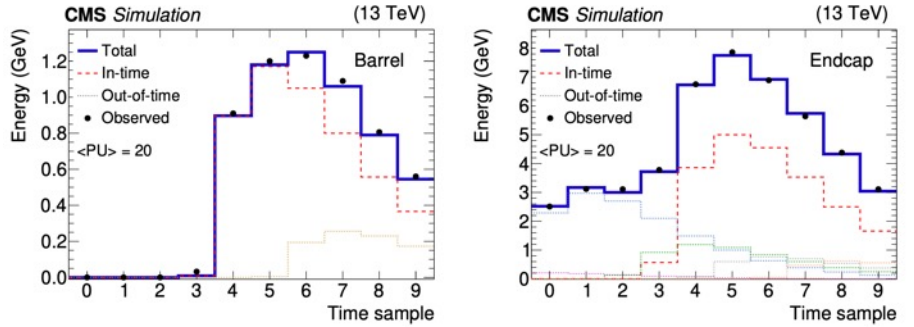
Pileup is not noise, just physics we're not interested in

- Getting the most information about the interesting process requires identifying all three components.
- Opportunities for more processing on-detector, beyond noise rejection.

Calibration also a processing algorithm

- Rates and data formats crucial.
- Streaming skips reprocessing, saves offline computing.





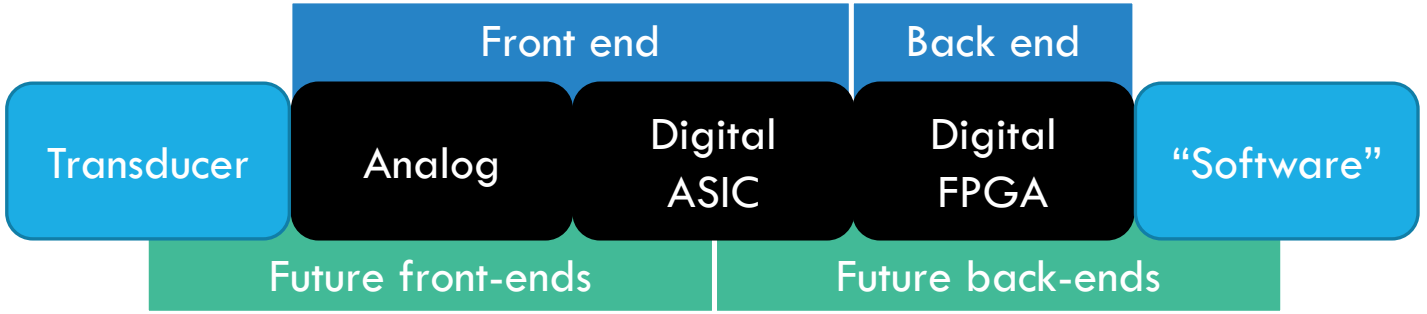
https://arxiv.org/abs/2006.14359

BUT WHAT PROCESSING? WHERE? ON WHAT?

Before:
- CPU-FPGA-ASIC

Now also:
- GPU, TPU (matrices), DPU (network)
- "Software" migrating now from CPU to GPU!

The "dial" position will move. ↗



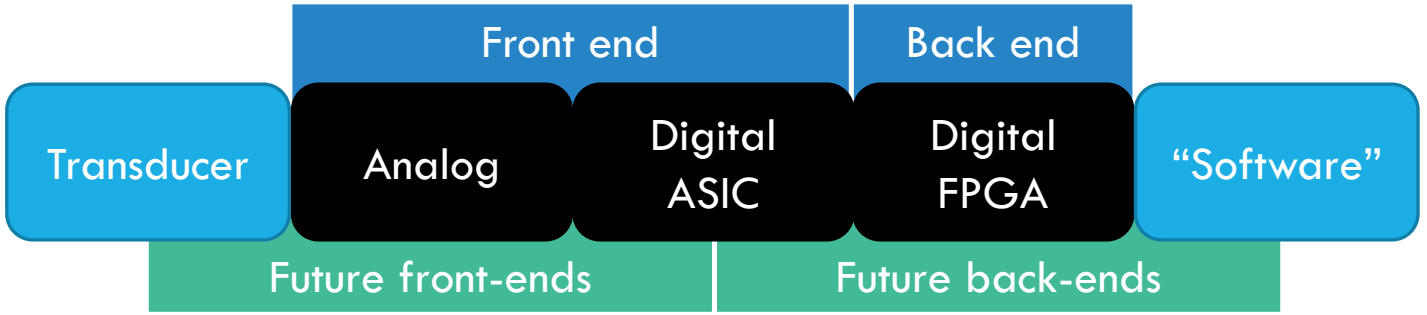
Analysis/compression
- Lossy neural networks. [TF7]
- Novel lossless compression that works on FE?
- Accelerator units, embedding, composition.

BUT WHAT PROCESSING? WHERE? ON WHAT?

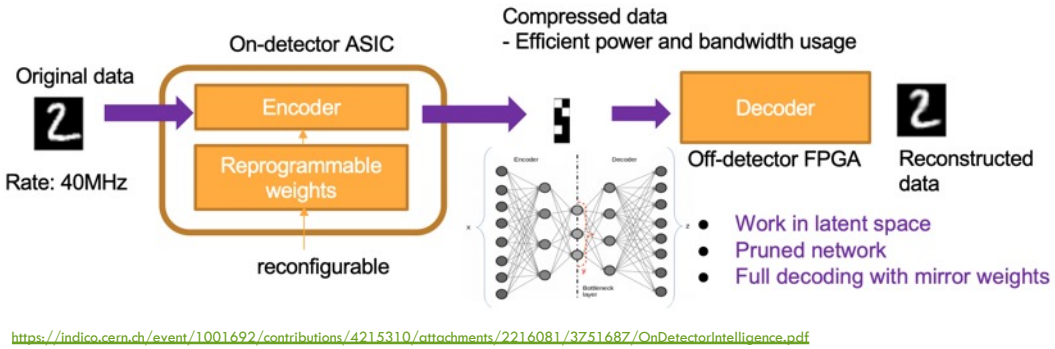
Before:
- CPU-FPGA-ASIC

Now also:
- GPU, TPU (matrices), DPU (network)
- "Software" migrating now from CPU to GPU!

The "dial" position will move. ↗



Analysis/compression
- Lossy neural networks. [TF7] →
- Novel lossless compression that works on FE?
- Accelerator units, embedding, composition.



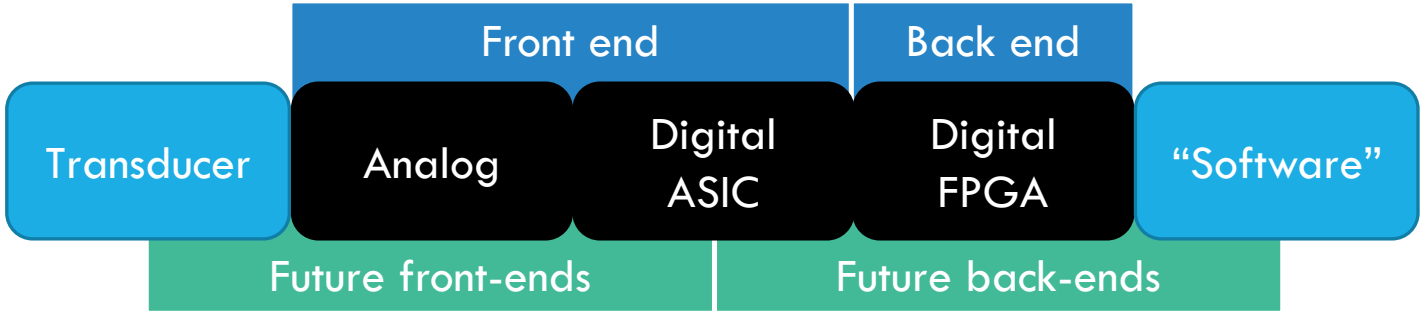
https://indico.cern.ch/event/1001692/contributions/4215310/attachments/2216081/3751687/OnDetectorIntelligence.pdf

BUT WHAT PROCESSING? WHERE? ON WHAT?

Before:
- CPU-FPGA-ASIC

Now also:
- GPU, TPU (matrices), DPU (network)
- "Software" migrating now from CPU to GPU!

The "dial" position will move. ↗



Analysis/compression
- Lossy neural networks. [TF7]
- Novel lossless compression that works on FE?
- Accelerator units, embedding, composition. →

The key issue is to have enough engineers to work out the best of it to our needs

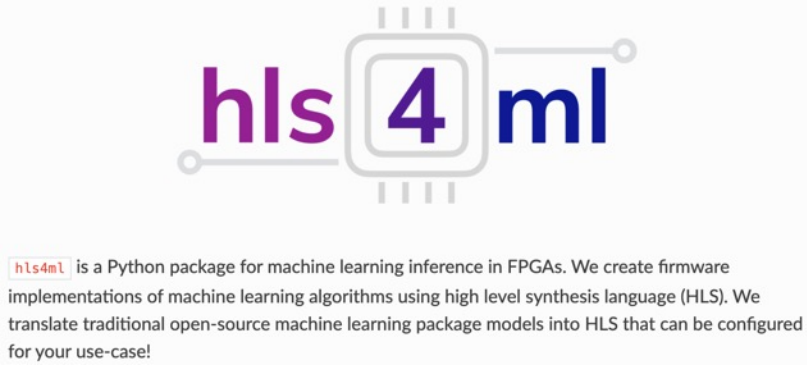
E.g. 9 mm2 in 12 nm = 234 k euros= 20.000.0000 transistors. What do you do with that!?

https://indico.cern.ch/event/1001692/contributions/4215289/attachments/2215392/3750382/ECFA_ASIC_Rivetti.pdf

AN ERA WITH FEWER ~~SUPPLIERS~~ PARTNERS?

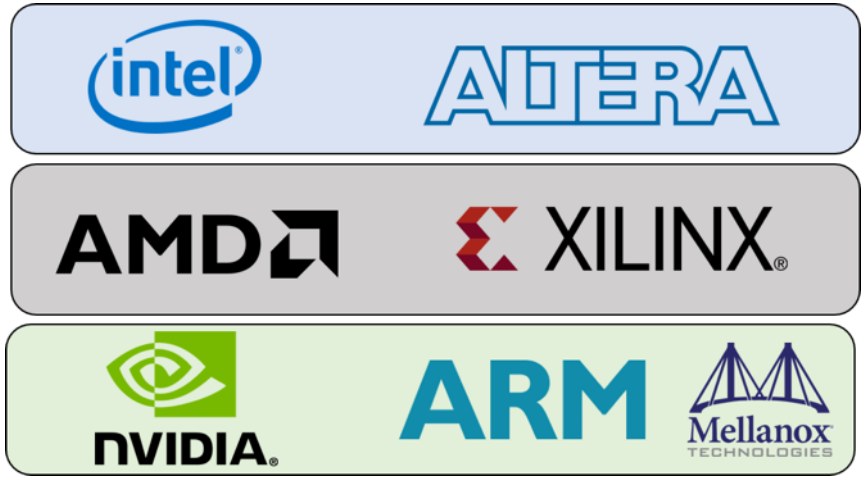
Community sharing challenge

- **Firmware is the next software.**
  - High-level synthesis ⇒ physicists programming FPGAs! →
- IP blocks developed in HEP can be shared if **licensing** is sorted out on day zero.



Industry engagement challenge

- Train with/through/for partners/suppliers.
- Usually very hard to "get" people from industry.
- What is our added value to multi-billion giants? →



https://semiwiki.com/fpga/292823-the-future-of-fpgas/

QUO VADIS REPROGRAMMABLE LOGIC?

Existing HEP-like applications ever smaller fish in ever larger ("AI") ocean.

- Unclear FPGAs will exist how we know them. →
- Unlikely able to drive, must adapt.

A future of networked accelerator units.

- Many opportunities if we have the resources.

Similar challenges for ASIC nodes. →

This poses a dilemma for FPGA providers. To effectively support FPGA prototype platforms, FPGAs should trade off DSPs, SoCs and other embedded blocks and processors for LUT area. To serve the DLA market, however, FPGAs should trade off generic logic area for specialized silicon area. The two requirements are not reconcilable, leaving the FPGA vendor in quandary.

Recently, Trimberger shared with me his thinking about the future of FPGAs, and elected to name it the "Age of Computation," in expectation that software processing will play a fundamental role in them. In his words, "In today's devices I/O communications infrastructure can consume over 50% of silicon area, in tomorrow's over 50% could be consumed by software computation units."

https://www.eetasia.com/the-future-of-fpga/

In May 2019 MOSIS dropped support for research institutions leaving US HEP foundry support in limbo as MOSIS turned their attention to very expensive 22nm and smaller technology nodes.

https://indico.cern.ch/event/870453/contributions/3671193/attachments/1960641/3258991/BRN_DC_ASICsreadout-2.pdf

POSSIBLY NEW INDUSTRIAL PARTNERS IN EUROPE

This decade will be marked by the post-COVID recovery.

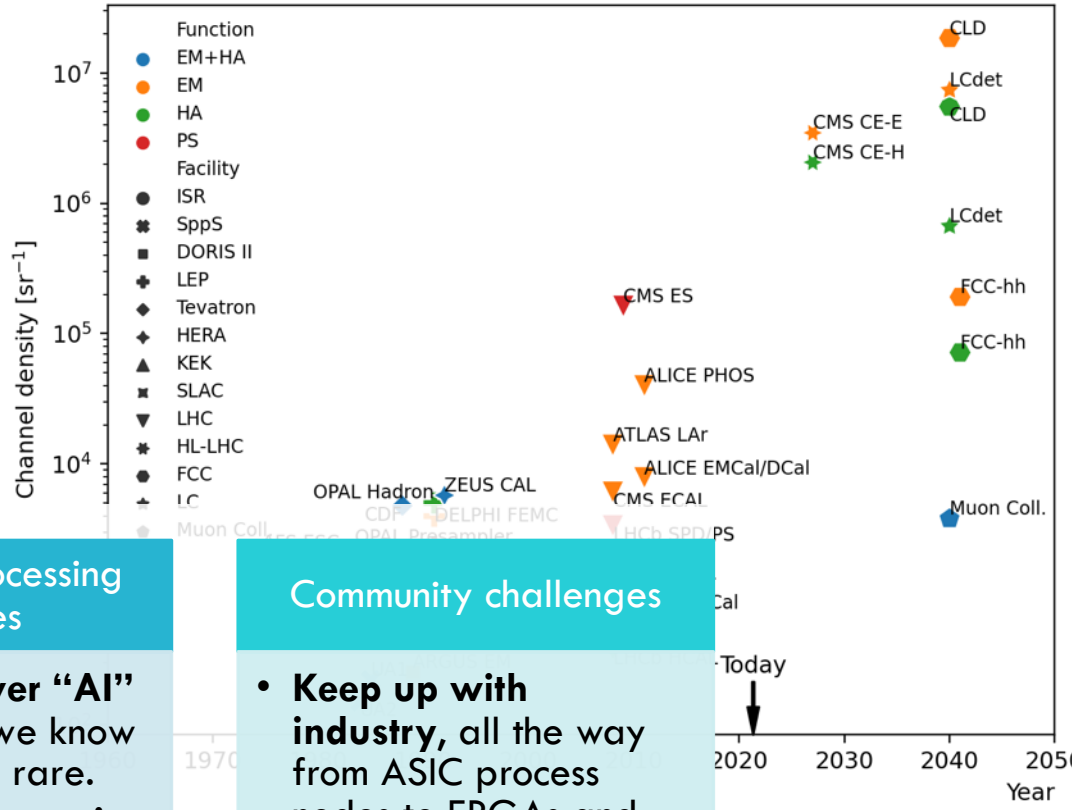
EU states committing funds to silicon technologies.

The European Union wants to double its chip manufacturing output to 20 percent of the global market by 2030. The goal is part of its new Digital Compass plan, announced yesterday, which aims to boost "digital sovereignty" by funding various high-tech initiatives.

As well as doubling chip output, the EU also wants all households to have 5G access and gigabit internet connectivity by 2030; for "all key public services" to be available online in every member state; and for the bloc to have its first quantum computer. Funding for these and other projects will come from the EU's €672.5 billion ($800 billion) coronavirus response fund, with 20 percent of this money ($160 billion) earmarked for tech investment.

https://www.theverge.com/2021/3/10/22322860/eu-semiconductor-chip-supply-double-output-2030-global-compass-investment

READOUT ⟶ PROCESSING



Application challenges

- **LC vs CC & ee vs hh** different boundary conditions.
- **Complementary research lines,** esp. timing/pileup and low-occupancy.
- And it's more than the "parts", **bringing systems together.**

Integration challenges

- **Little space** for components; mobile phone technology?
- Integrate across absorbers for **on-detector PFA**?
- Low-power, low-noise, high dynamic-range, and precise timing: all-in-one **free lunch**?
- Logical integration and physics **simulation – system aspects** and calibration.

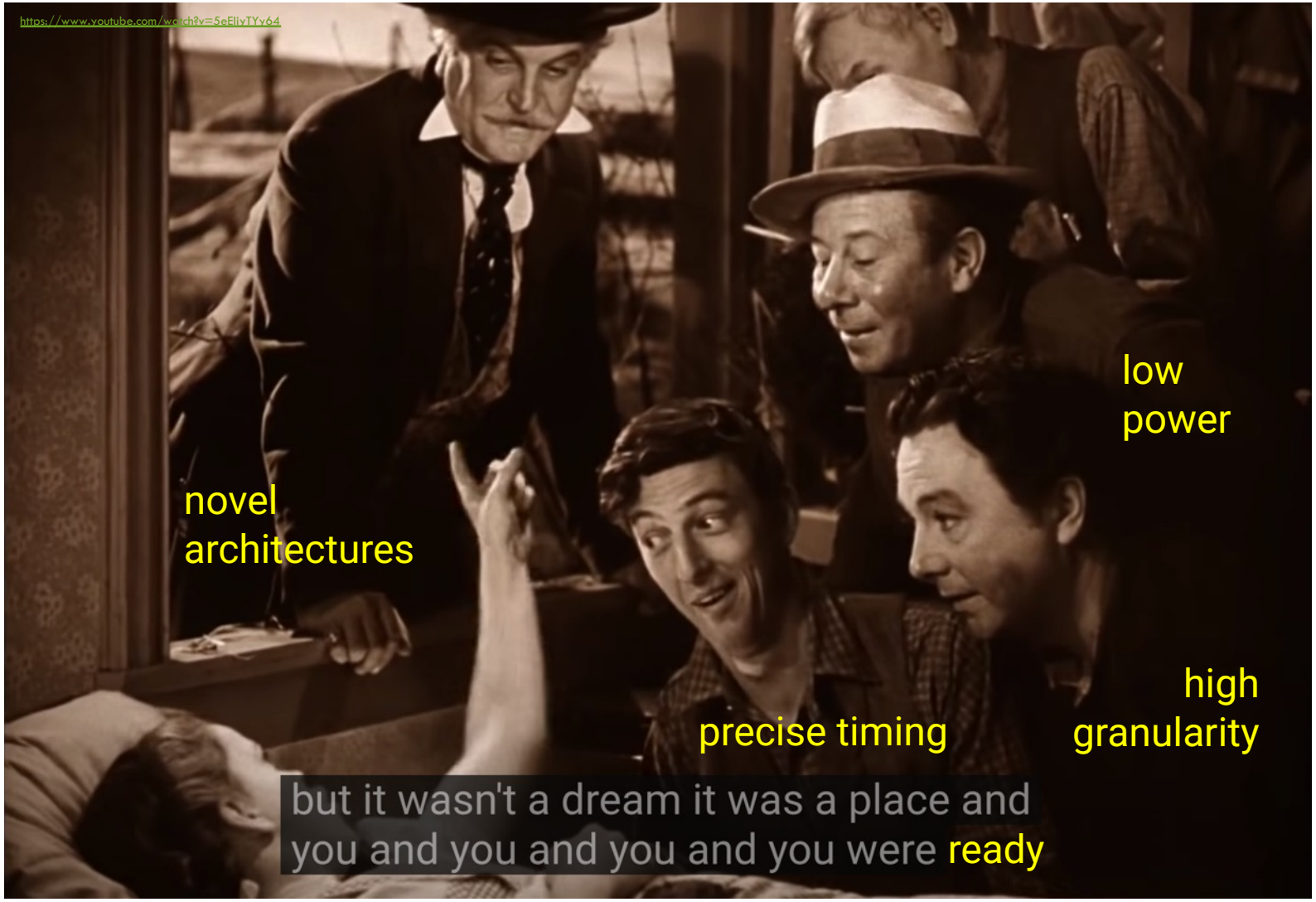
Information processing challenges

- **Industry all over "AI"** and FPGA as we know it may become rare.
- **New accelerator units** on the horizon (TPU, DPU).
- **Processing distributed** along the front-end to back-end continuum.
- Novel, robust, **lossless and lossy compression** algorithms.

Community challenges

- **Keep up with industry,** all the way from ASIC process nodes to FPGAs and networking.
- **Open-source** licensed blocks from day zero avoid collaboration issues.
- **Training people** {with, for, through} industry?

THANK YOU!


https://www.youtube.com/watch?v=5eEIivTYy64
novel architectures
low power
precise timing
high granularity
but it wasn't a dream it was a place and you and you and you and you were ready

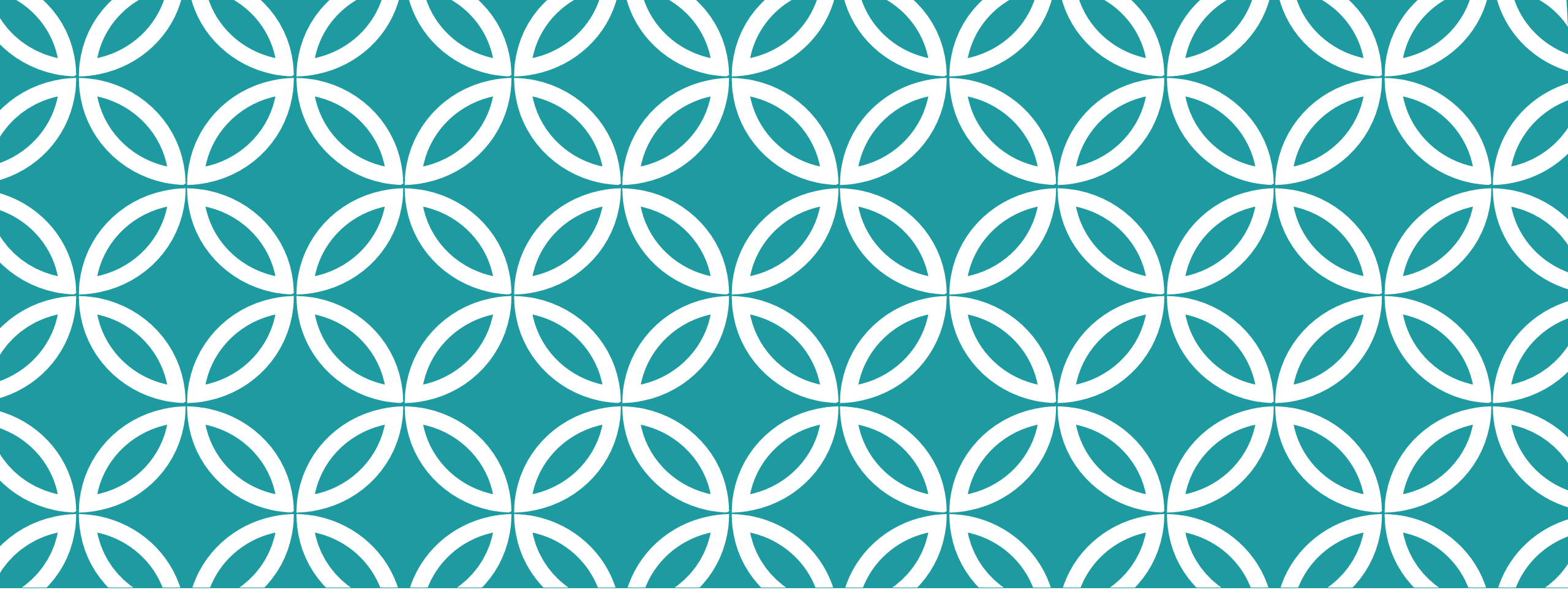
FURTHER INFORMATION

Other ECFA Detector R&D Symposia

- TF3 Solid State Detectors.
- TF7 Electronics and On-detector Processing.

2019 Report of the US DOE Office of Science Workshop on Basic Research Needs for HEP Detector R&D.

CHEP, CHEF, and other calorimetry conferences.



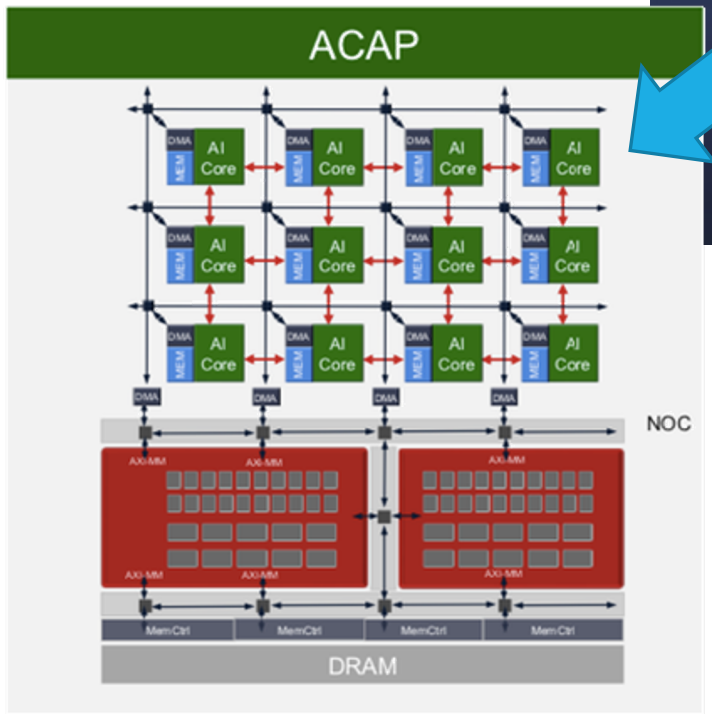
FOR DISCUSSION

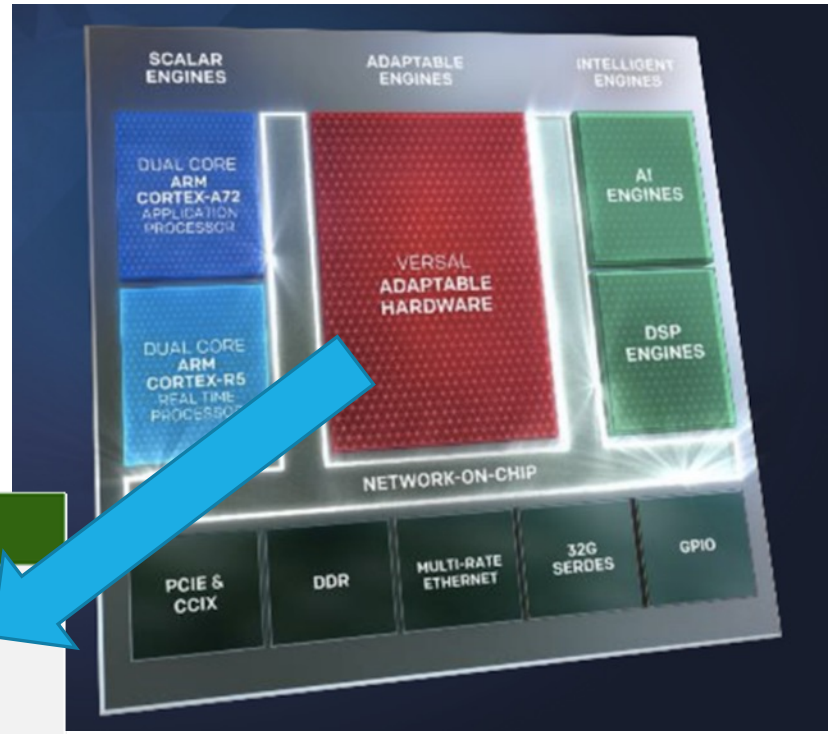
THE NEXT FPGAS ARE HERE

Adaptable Multi-Core Platform

- **2D Array** of SW Programmable Cores
- Vector Architecture (**Improves MAC density**)
- **Distributed Memory (TCM)**, No Caches
- DMAs For **Dataflow Processing**
- **Flexible Interconnect** Topologies

- **Adaptable** On-Chip **Memory**
- **Configurable NOC** Data Movement Backbone





6

XILINX.

https://moorinsightsstrategy.com/xilinx-reveals-more-versal-details/

VARIETY AND CONVERGENCE – TPU

CPU
- Small models
- Small datasets
- Useful for design space exploration

GPU
- Medium-to-large models, datasets
- Image, video processing
- Application on CUDA or OpenCL

TPU
- Matrix computations
- Dense vector processing
- No custom TensorFlow operations

FPGA
- Large datasets, models
- Compute intensive applications
- High performance, high perf./cost ratio

https://inaccel.com/cpu-gpu-fpga-or-tpu-which-one-to-choose-for-my-machine-learning-training/

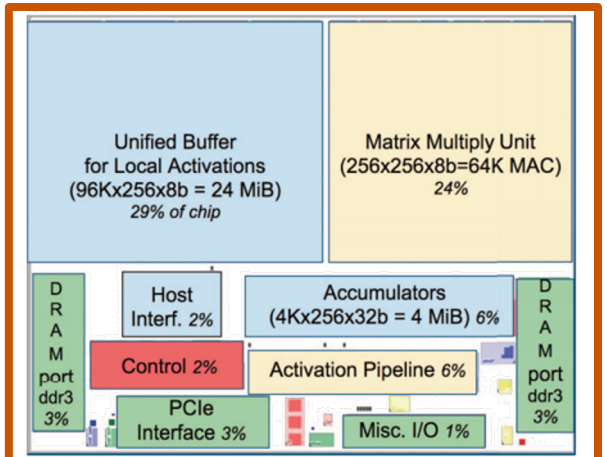


Figure 2. Floorplan of TPU die. The shading follows Figure 1. The light (blue) datapath is 67%, the medium (green) I/O is 10%, and the dark (red) control is just 2% of the die. Control is much larger (and much harder to design) in a CPU or GPU.

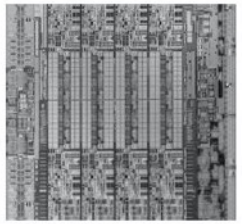
https://medium.com/@lightworld/a-survey-paper-comparing-modern-cpu-gpu-tpu-hardware-in-relation-to-neural-network-training-and-255c8626c168

VARIETY AND CONVERGENCE – DPU

A New Category of Microprocessor

Purpose-built for the data-centric era

CPU



General-purpose

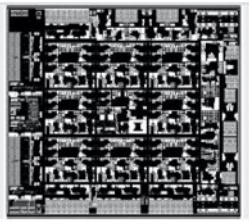
Multi-core, MIMD
High IPC for single threads
Fine-grain memory sharing
Classical cache coherency
Based on locality of reference
Ideal for low to medium I/O

GPU

Vector floating point

Multi-core, SIMD
High throughput for vector processing
Coarse-grain memory sharing
Relaxed coherency
Based on data >> instructions
Ideal for graphics, ML training

Fungible DPU™



Data-centric

Multi-core, MIMD + tightly-coupled accelerators
High throughput for multiplexed workloads
TrueFabric™ enables disaggregation and pooling
Specialized memory system and on-chip fabric
Ideal for network, storage, security, virtualization
Data-centric computations run >10X more efficiently

FUNGIBLE

https://www.servethehome.com/what-is-a-dpu-a-data-processing-unit-quick-primer/

"The purpose of computing is insight, not numbers."

Richard Hamming (1962)

"The purpose of readout is insight, not data."