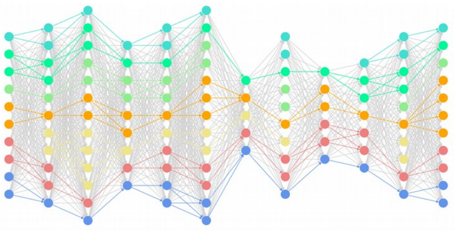
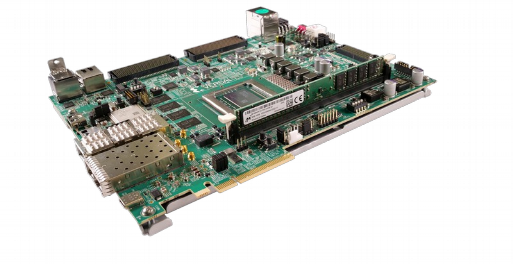
GNN live tracking and triggering on the Versal platform for Dark Matter searches

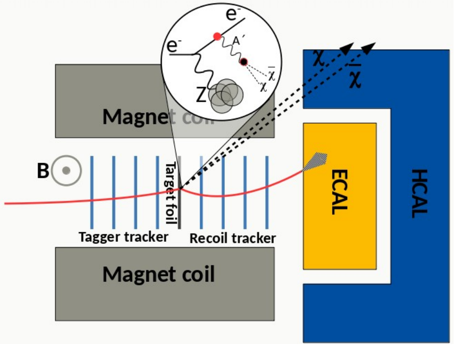
Workshop on Realtime Machine Learning

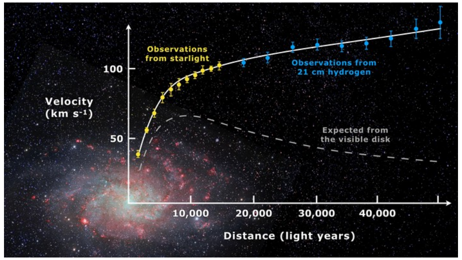
11.04.2024

Michael Lupberger, Patrick Schwäbig









FTD UNIVERSITÄT BONN

VISION & OUTLINE

Vision

Strategy Pyramid, from: DPG Leading for Tomorrow

Increase the knowledge of humankind and advance humanity with technological progress and fundamental findings

Sustainable smart-data acquisition and intelligent feature extraction

with

GNN live tracking and triggering on the Versal

to

e.g. enlighten the dark sector at the intensity frontier

The Why


The dream /
picture of the future
The Why /
purpose
Vision
The What
The How /
Next Steps
Mission
Strategic Goals
Strategic Actions

THE INTENSITY FRONTIER CHALLENGE

The Why

- Particle/Hadron Physics experiments as (HL-)LHC, FCC, … ➞ more intense beams
  $\Rightarrow$ find rare (*e.g. Beyond SM)* events in big data

THE INTENSITY FRONTIER CHALLENGE

The Why

- Particle/Hadron Physics experiments as (HL-)LHC, FCC, … ➞ more intense beams
  ⇒ find rare (*e.g. Beyond SM)* events in big data
  + Computing↗ and
  power consumption↗ relevant

THE INTENSITY FRONTIER CHALLENGE

The Why

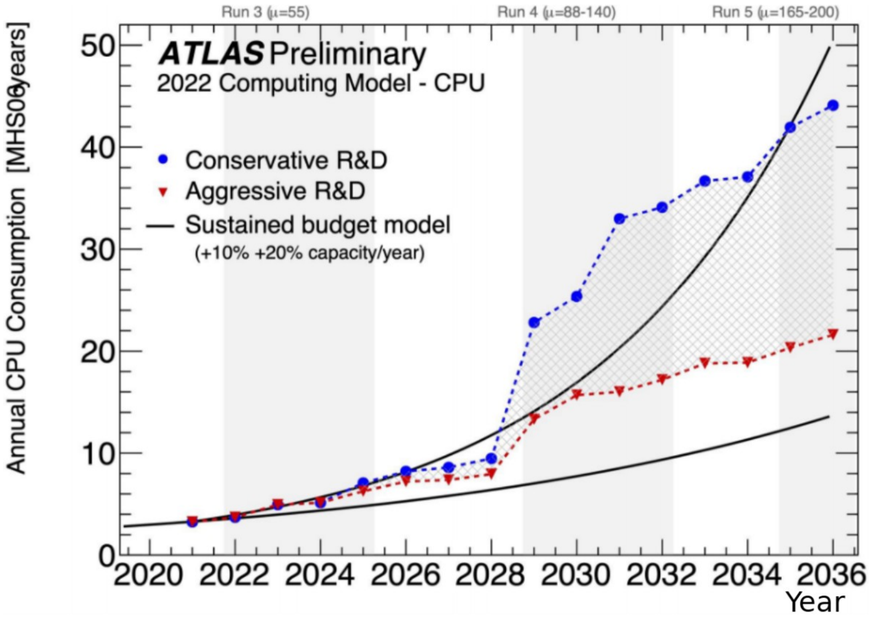
- Particle/Hadron Physics experiments as (HL-)LHC, FCC, … ➞ more intense beams
  ⇒ find rare (*e.g. Beyond SM*) events in big data
  + Computing➚ and
  power consumption➚ relevant

For the first time, electronics has become an enabling, but potentially also limiting, aspect.
ECFA Detector R&D Roadmap

THE INTENSITY FRONTIER CHALLENGE

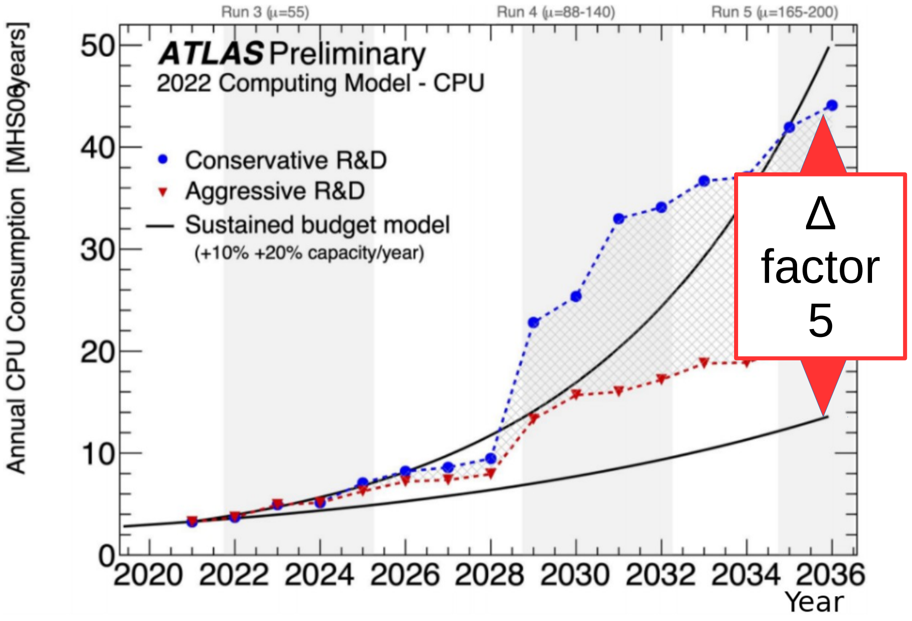
The Why

- Particle/Hadron Physics experiments as (HL-)LHC, FCC, … ➞ more intense beams
  $\Rightarrow$ find rare (*e.g. Beyond SM)* events in big data
  + Computing ➚ and
  power consumption ➚ relevant



THE INTENSITY FRONTIER CHALLENGE

- Particle/Hadron Physics experiments as (HL-)LHC, FCC, … ➙ more intense beams
  $\Rightarrow$ find rare (*e.g. Beyond SM)* events in big data
  + Computing ➚ and
  power consumption ➚ relevant

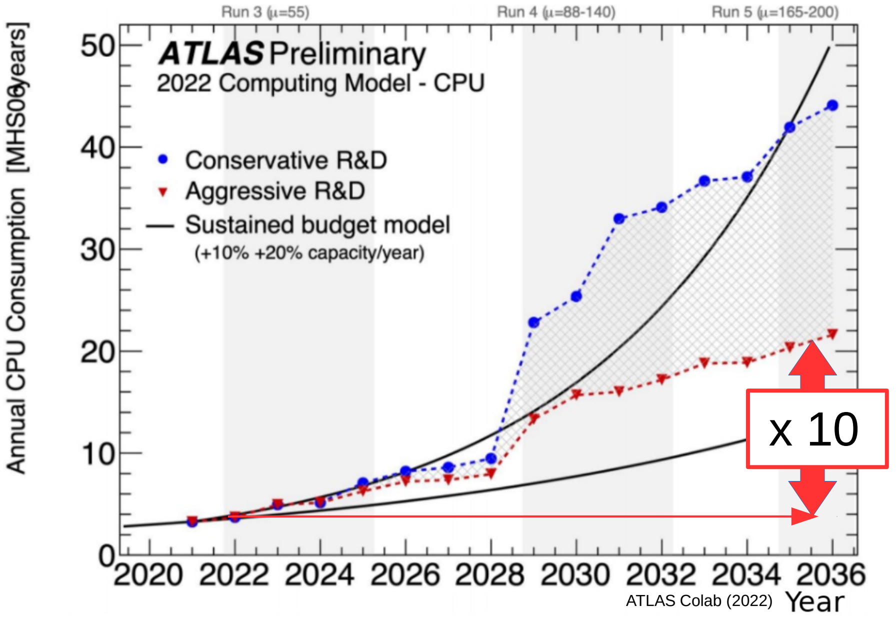


THE INTENSITY FRONTIER CHALLENGE

The Why

- Particle/Hadron Physics experiments as (HL-)LHC, FCC, … ➞ more intense beams
  ⇒ find rare (*e.g. Beyond SM)* events in big data
  + Computing↗ and
  power consumption↗ relevant

> *Data selection and reduction will increasingly be addressed by intelligent processors close to the front end, reducing data movement.*
> ECFA Detector R&D Roadmap
>
> DRDT 7.2: *Advanced data reduction techniques (ML/AI)* in FPGAs.

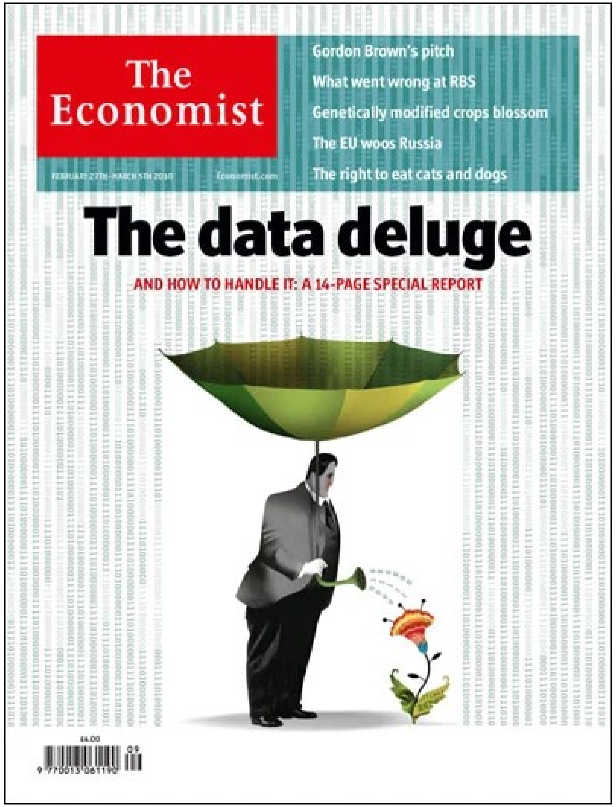


THE INTENSITY FRONTIER CHALLENGE

The Why

- Particle/Hadron Physics experiments as (HL-)LHC, FCC, … ➝ more intense beams
  ⇒ find rare (*e.g. Beyond SM)* events in big data
  + Computing↗ and
  power consumption↗ relevant
- Data deluge: general challenge of information-driven society

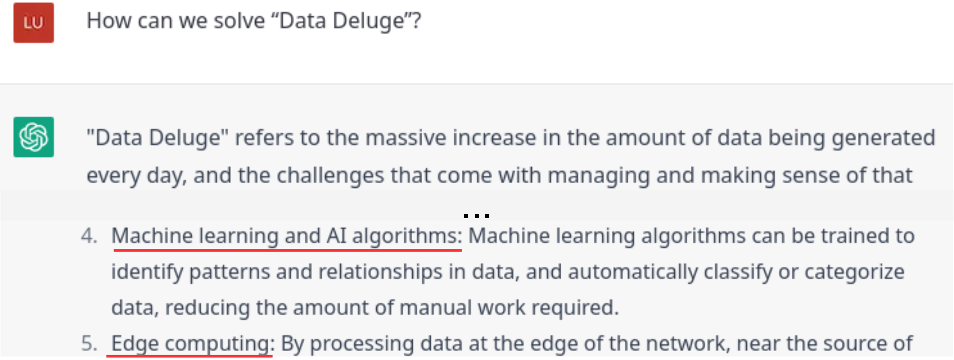
Data centres are responsible for 3.5 % of the global carbon emissions.
A. Andrae. "Total consumer power consumption forecast". In: Nordic Digital Business Summit 10 (2017), p. 69.
An increase to 5.5 % (20 % of all electricity) is expected until 2025.
J. M. Lima. "Data centres of the world will consume 1/5 of earth's power by 2025". In: Data Economy (2017).



THE INTENSITY FRONTIER CHALLENGE

The Why

- Particle/Hadron Physics experiments as (HL-)LHC, FCC, … ➞ more intense beams
  ⇒ find rare (*e.g. Beyond SM)* events in big data
  + Computing↗ and
  power consumption↗ relevant
- Data deluge: general challenge of information-driven society



- AI/ML and edgy computing are one part of a solution

THE INTENSITY FRONTIER CHALLENGE

The Why

- Particle/Hadron Physics experiments as (HL-)LHC, FCC, … ➞ more intense beams
  ⇒ find rare (*e.g. Beyond SM)* events in big data
  + Computing➚ and
  power consumption➚ relevant
- Data deluge: general challenge of
  information-driven society

- AI/ML and edgy computing are one part of a solution

Technological objective:
Development and investigation of power-efficient, next-generation data acquisition architecture with the goal of O(10 MHz − 100+ MHz) online pattern recognition and triggering at latencies O(1 − 10µs) using ML methods close to the detector.

➞ Future of (PP) data taking

MISSION STATEMENT

Vision:

Increase the knowledge of humankind and advance humanity with technological progress and fundamental findings

Mission statement:

Taming data deluge and reducing power consumption in PP instrumentation

to

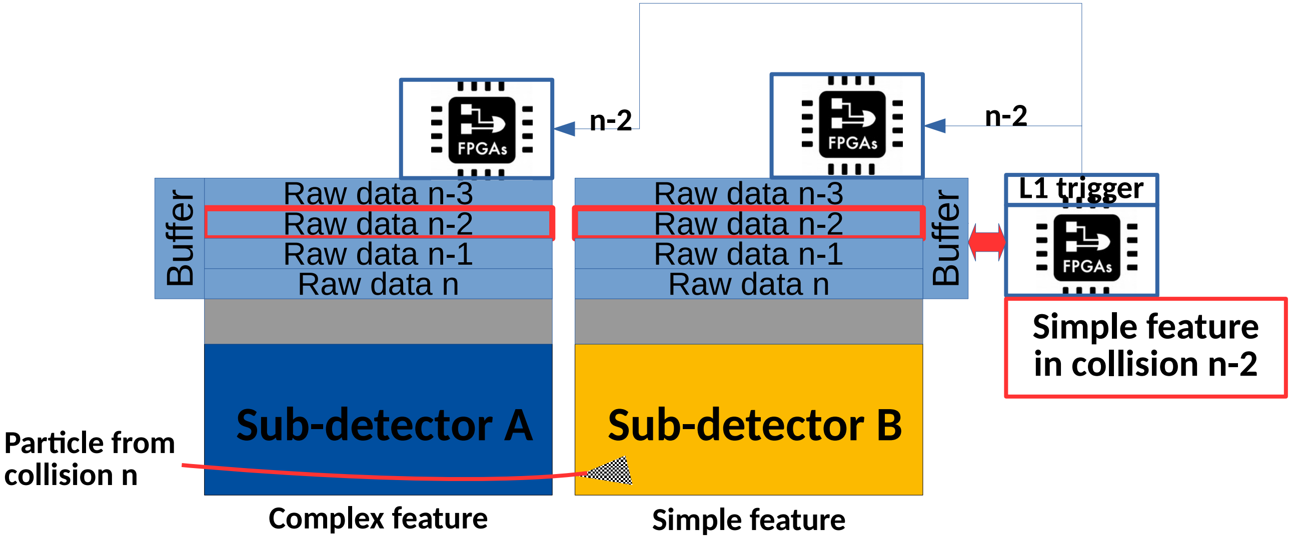
enable sustainable/economic physics searches at the intensity frontier


The dream /
picture of the future
The Why /
purpose
Vision
Mission
The What
The How /
Next Steps
Strategic Goals
Strategic Actions

The What

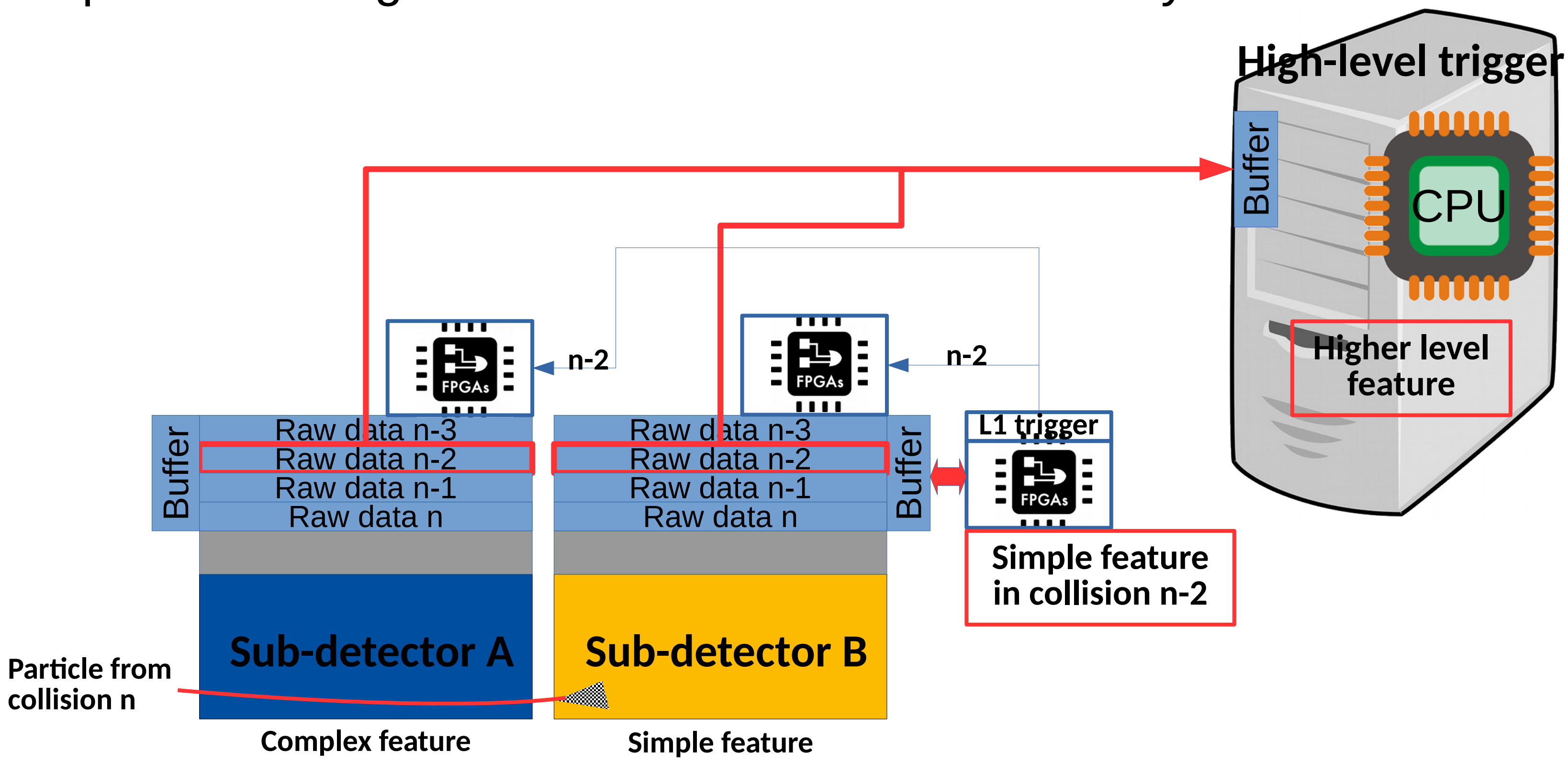
GENERAL APPROACH – TECHNOLOGICAL

Experimental big-data collection: the traditional way



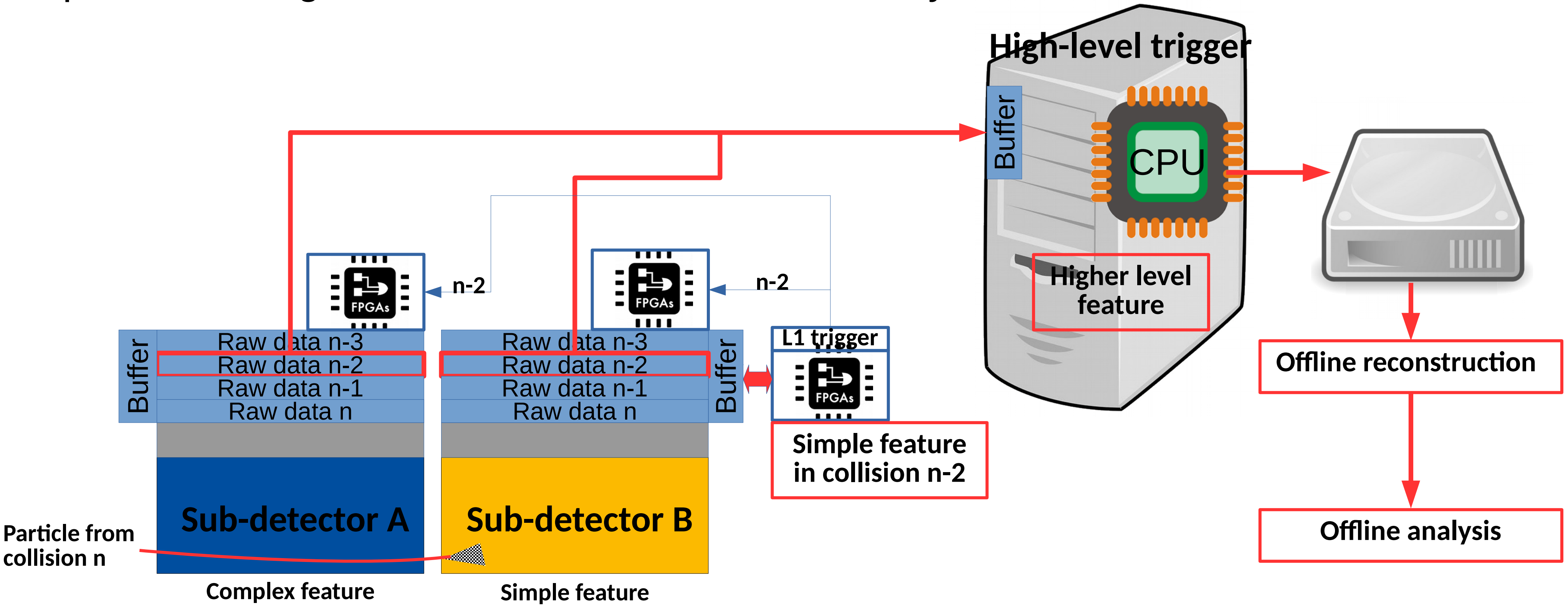
GENERAL APPROACH – TECHNOLOGICAL

Experimental big-data collection: the traditional way



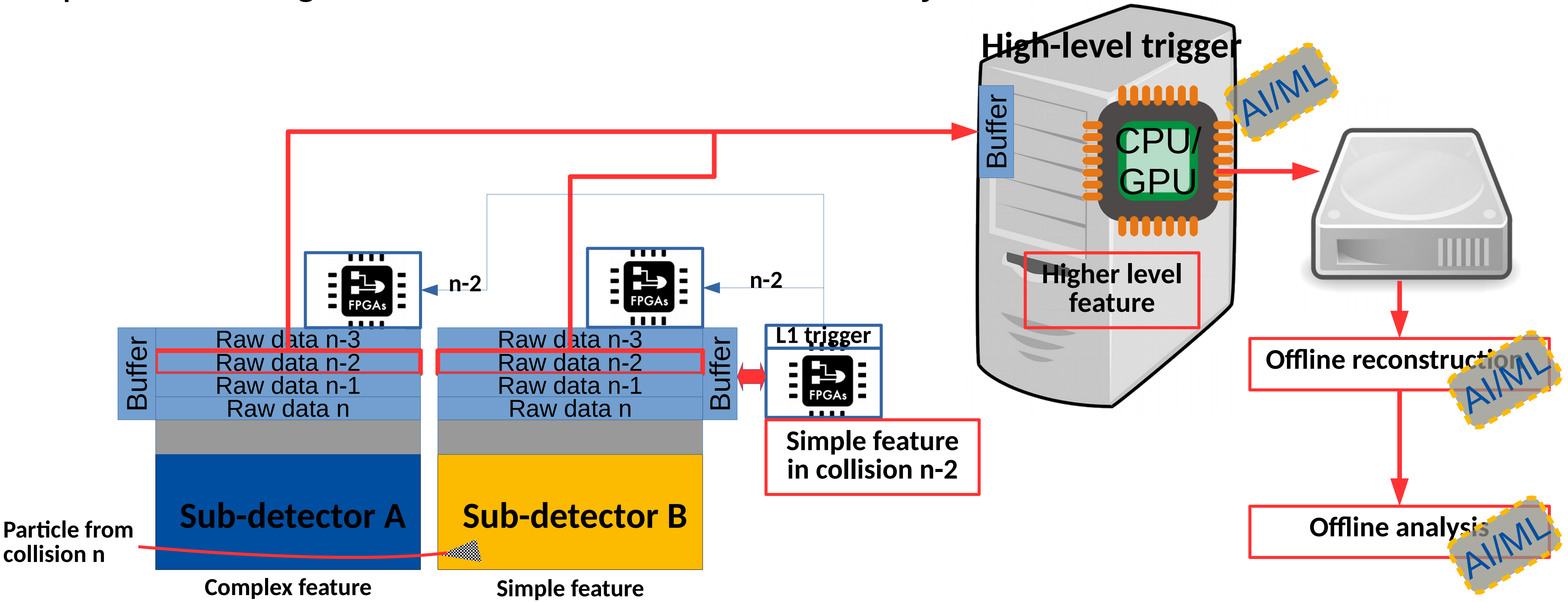
GENERAL APPROACH – TECHNOLOGICAL

Experimental big-data collection: the traditional way



GENERAL APPROACH – TECHNOLOGICAL

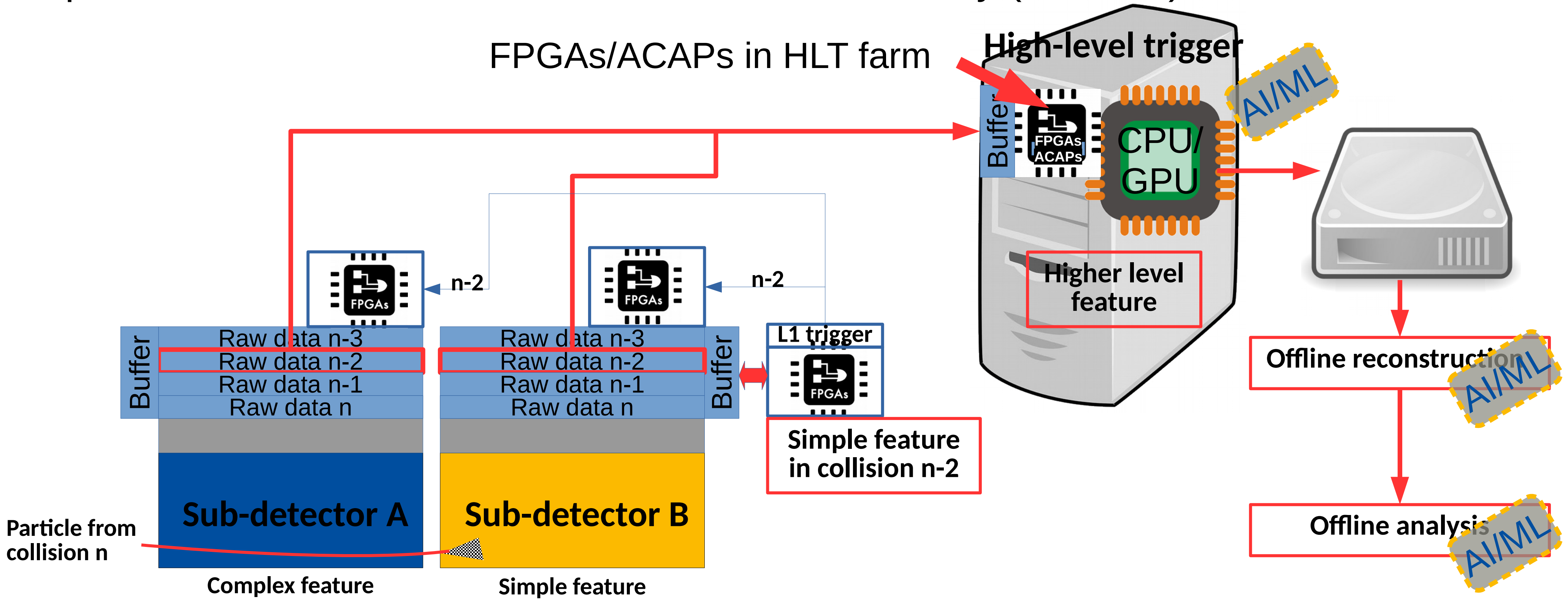
Experimental big-data collection: the traditional way



GENERAL APPROACH – TECHNOLOGICAL

The What

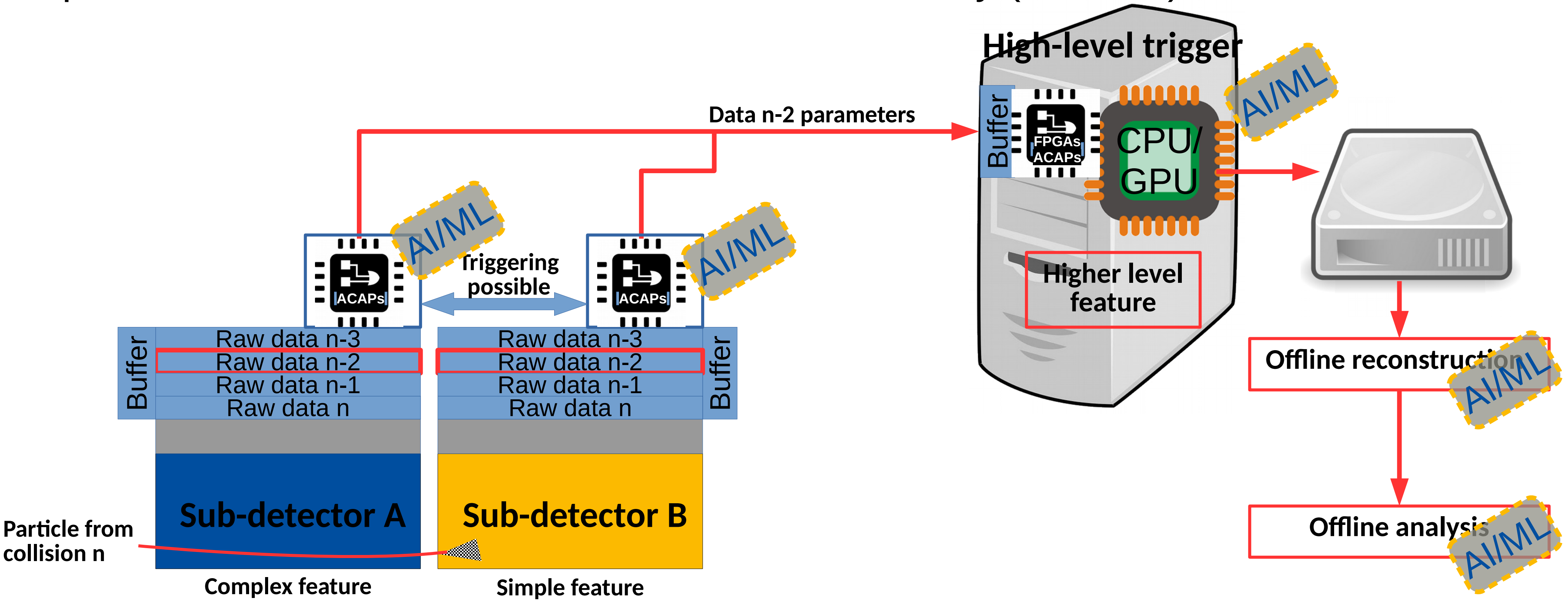
Experimental smart-data collection: the innovative way (extreme)



GENERAL APPROACH – TECHNOLOGICAL

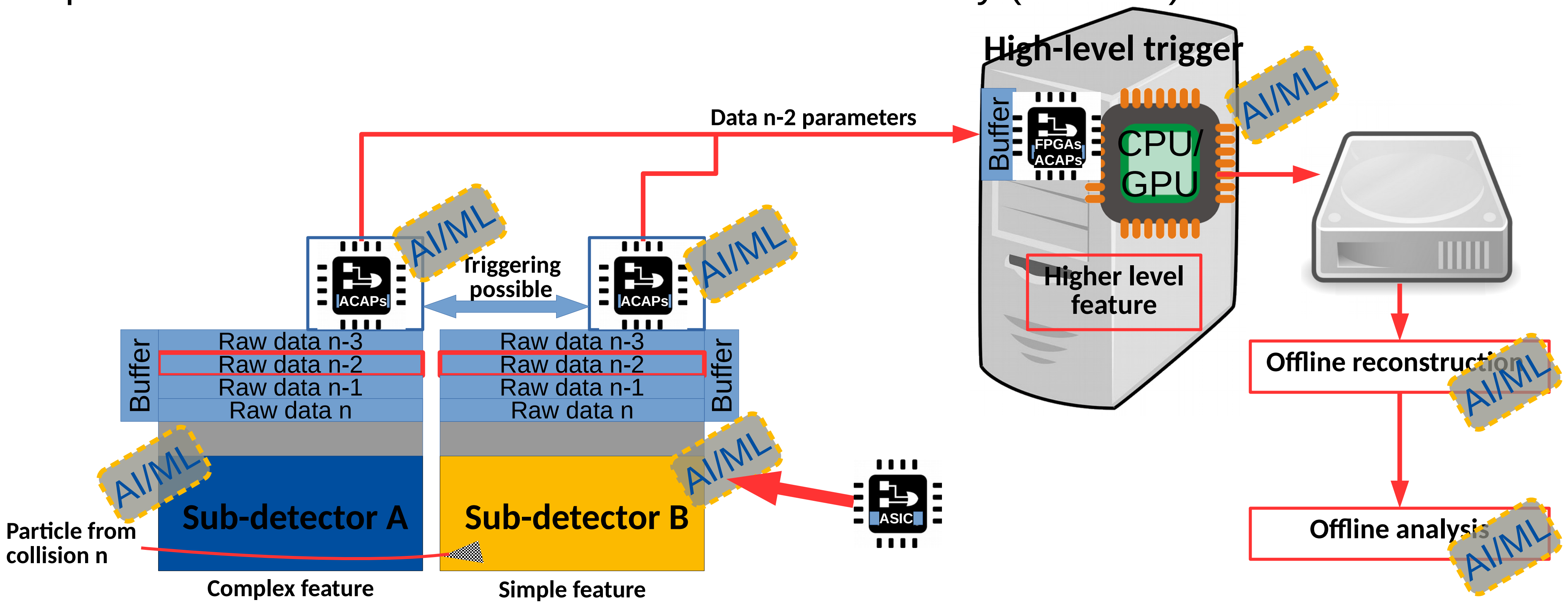
The What

Experimental smart-data collection: the innovative way (extreme)



GENERAL APPROACH – TECHNOLOGICAL

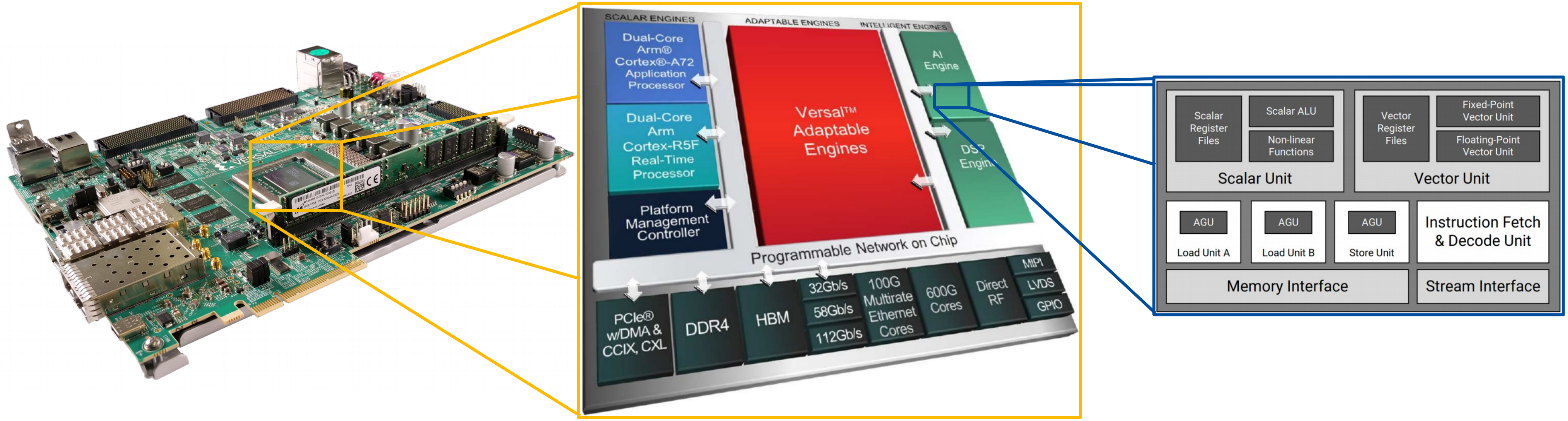
Experimental smart-data collection: the innovative way (extreme)



PROJECT TECHNICAL BASIS

The
What

Xilinx Advanced Common Application Platform (ACAP) technology in Versal AI



PROJECT TECHNICAL BASIS

Resources:

- Hardware

Funded by

UNIVERSITÄT BONN

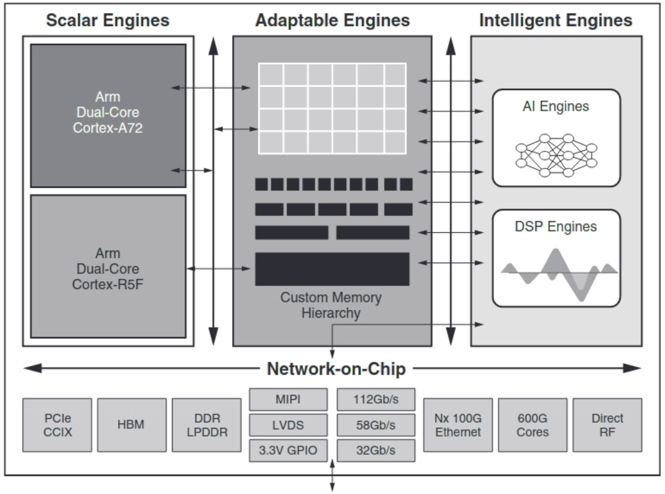
TRA MATTER

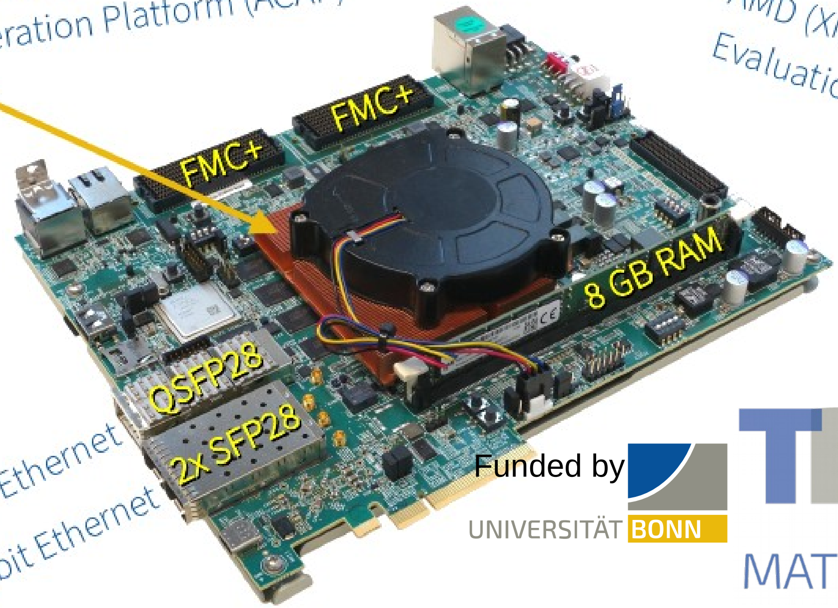
PROJECT TECHNICAL BASIS

The
What

Resources:

- Hardware
  - **400 AI processors (“AI engines”)**
  - FPGA (“Adaptable Engines”): **2k** DSPs, nearly **2M** logic cells
  - Arm CPU, Arm RPU (“Scalar Engines”)





Funded by
UNIVERSITÄT BONN
TRA MATTER

PROJECT TECHNICAL BASIS

Resources:

- Hardware

Funded by UNIVERSITÄT BONN Global cooperation

PROJECT TECHNICAL BASIS

The
What

Resources:

- Hardware
  - → Similar to 100 Gb/s Network server or our HRZ
  - → Mellanox MCX516A-CCAT 100 Gb/s Ethernet interface
  - → QSFP28 Transceiver
  - → 4xMTA36ASF8G72PZ-3G2B2 64 GB DDR4 RAM
  - → AMD Epyc 7352 24 Core CPU
  - → Gigabyte MZ32-AR0 Mainboard, Broadcom 9300-8i SAS
  - → Samsung 980 Pro 250GB SSD (System)
  - → Crucial P3 Plus SSD 1 TB NVMe memory
  - → (GPU)

(HLT) DAQ Server

Used for 100 Gb/s data reception and treatment
(and firmware compiling)

Funded by UNIVERSITÄT BONN Global cooperation

The
What

PROJECT TECHNICAL BASIS

Resources:

- Hardware
- International network

AMD XILINX
Research Lab

FastML collaboration

Contact for ACAP technology

cea

Özgür Sahin

Funded by UNIVERSITÄT BONN

Global cooperation

ETH zürich

Thea Arrestad

Funded by UNIVERSITÄT BONN

Strategic partnership

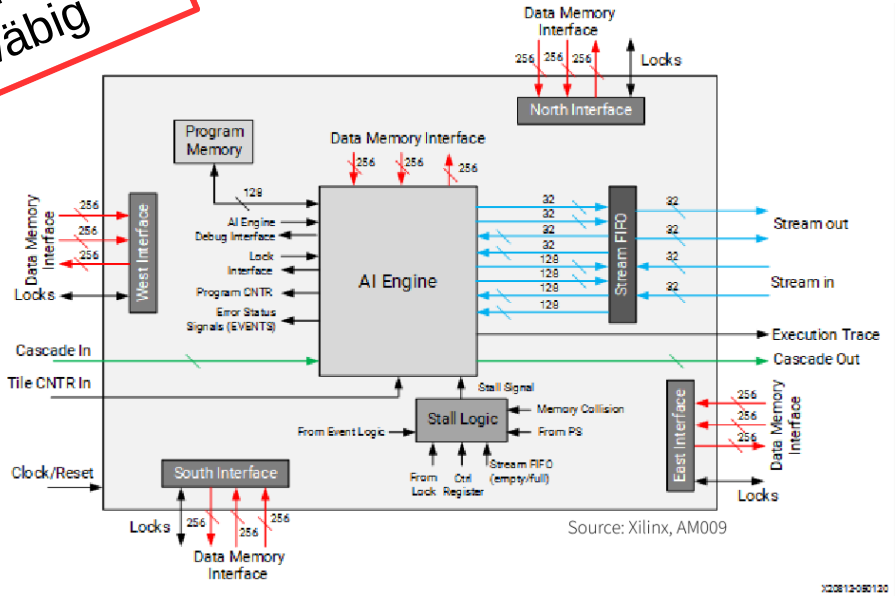
The
What

PROJECT TECHNICAL BASIS

Resources:

- Hardware
- International network
- ACAP technical know how

PhD student
P. Schwäbig



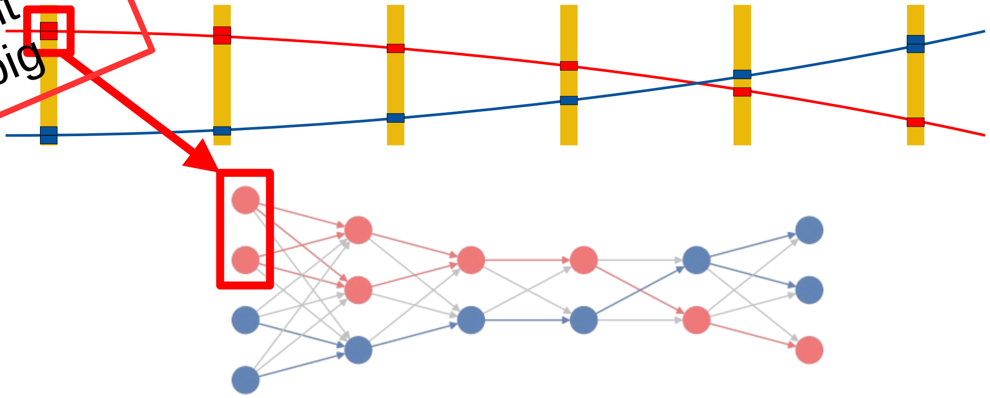
Source: Xilinx, AM009

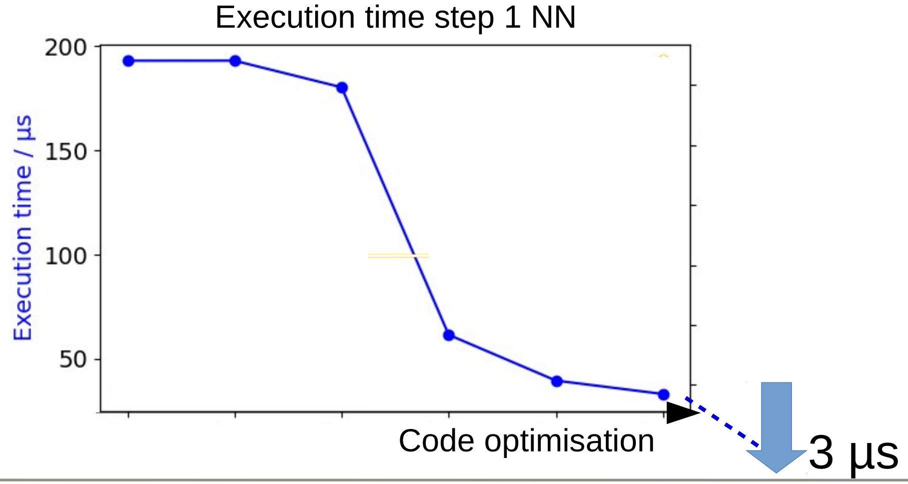
PROJECT TECHNICAL BASIS

Resources:

- Hardware
- International network
- ACAP technical know how
- Preliminary implementation of Graph Neural Network (GNN)

PhD student
P. Schwäbig





The How

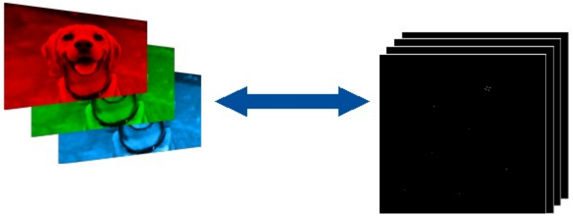
Technical details:
- GNN implementation
- How do we work with the ACAP/VCK190
- DAQ chain setup

The
How

GNN implementation

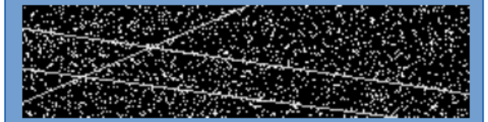
What is a GNN or more precisely a Message Passing Graph Neural Network (MPNN)

- “Generalization” of CNNs to graphs
  - CNN: Working on 2D data (or multiple levels of 2D data e.g. RGB images)
  - Detector data: 2D image but mainly zeros → sparse data
- Difficult for CNN:
  - weights tend to become zero → network learns zeros
  - “high-res” raw data (most of which we do not need: zero)



or projected

with noise?



But need enough layers for a good projection

The How

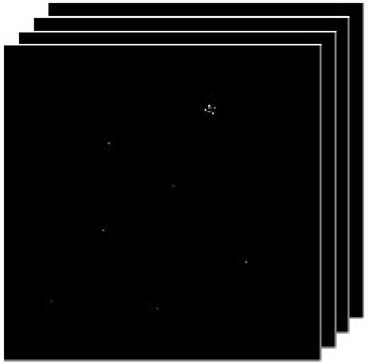
GNN implementation

Input to GNN: Graph = Representation of data

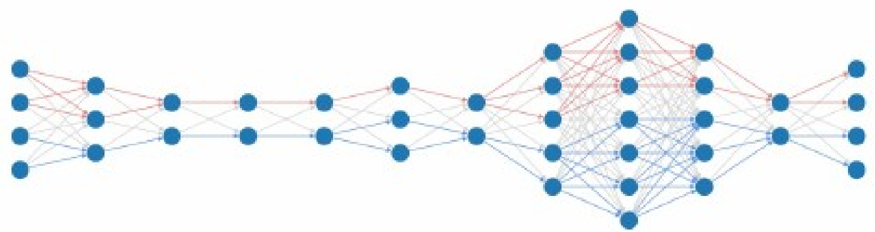
Build graph from data:

- Detector hits → nodes
- Interconnection (particle trajectory) → edges

both optionally with features e.g. time, signal amplitude, pixel ID (position) ,…



Draw all possible edges
(between hits of adjacent layers)



This is not the NN but example data

Huge variety of networks: Node classification, Graph classification,

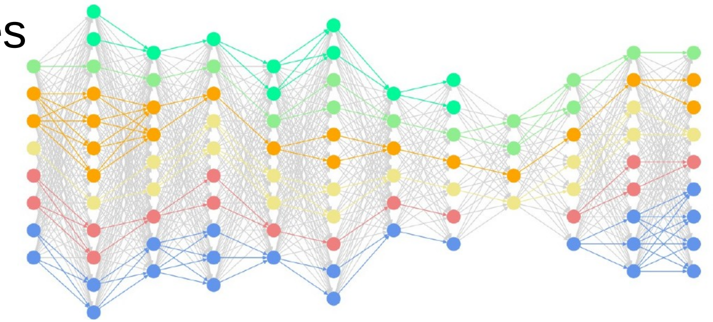
Edge classification: is from real track (1) or not (0)

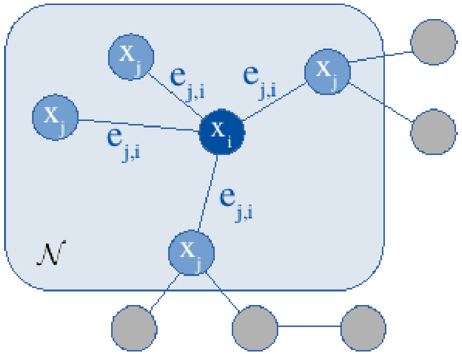
GNN implementation

The How

Our Message Passing Graph Neural Network (MPNN)

- Currently using interaction network (IN) which is a MPNN, described in 2103:16701*
  - Training: NN (1 hidden layer) works through all nodes/edges
  - Each node has a neighbourhood
  - Receives information (“messages”) from neighbouring nodes





Previous time step

$$\mathbf{x}_i^{(k)} = \gamma^{(k)} \left(\mathbf{x}_i^{(k-1)}, \square_{j \in \mathcal{N}(i)} \phi^{(k)} \left(\mathbf{x}_i^{(k-1)}, \mathbf{x}_j^{(k-1)}, \mathbf{e}_{j,i} \right) \right)$$

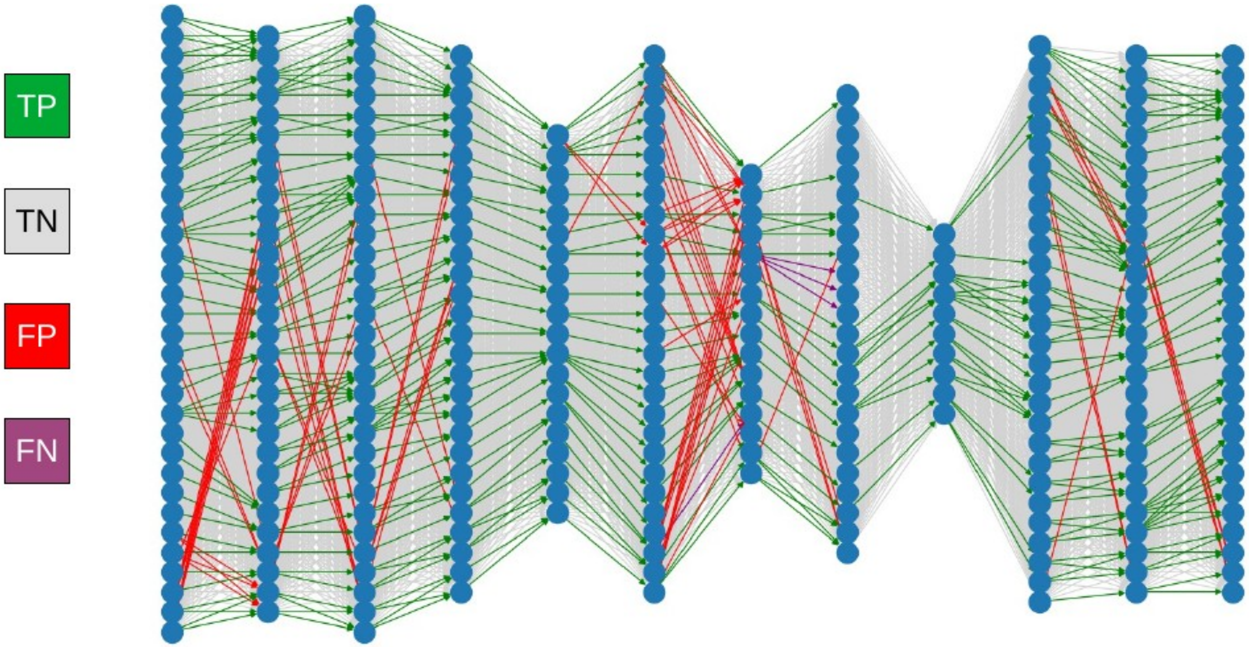
Update Aggregate Message

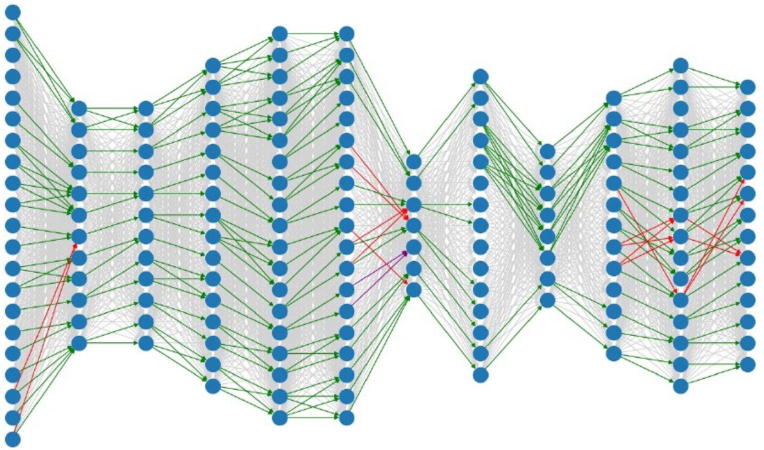
* DeZoort, G., Thais, S., Duarte, J. et al. Charged Particle Tracking via Edge-Classifying Interaction Networks. Comput Softw Big Sci 5, 26 (2021). https://doi.org/10.1007/s41781-021-00073-z

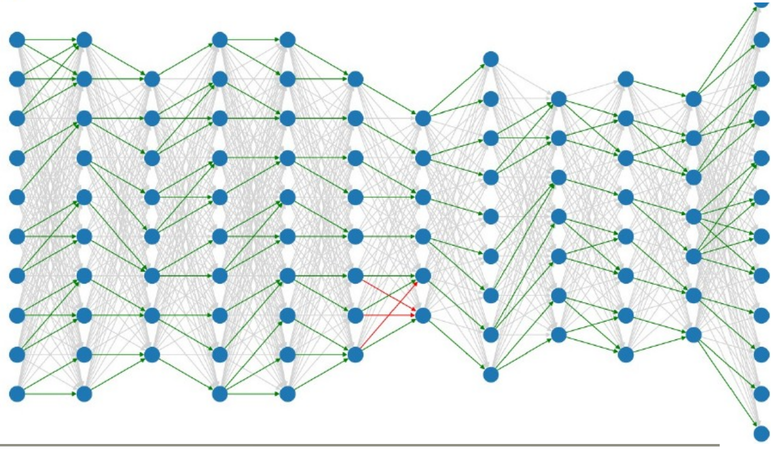
The How

GNN implementation

Examples with different simulated #tracks / graph on fixed detector size



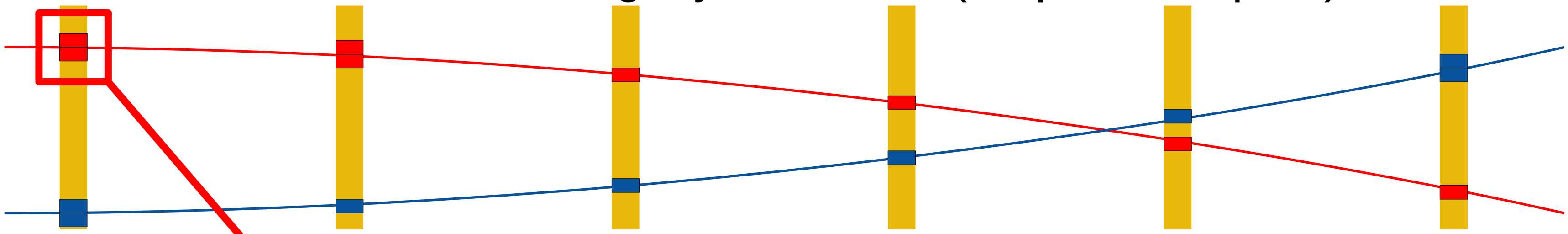




GNN implementation

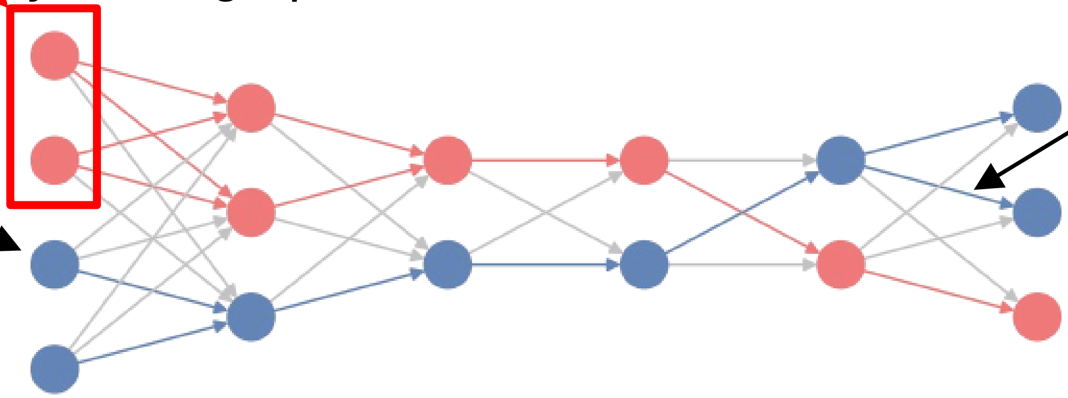
The How

Start with simulated data: Tracking layers w/ hits (ampl., time, pos.), certain #tracks



So far in software only: build graphs

Node = Hit with features: Ampl., time, pos.

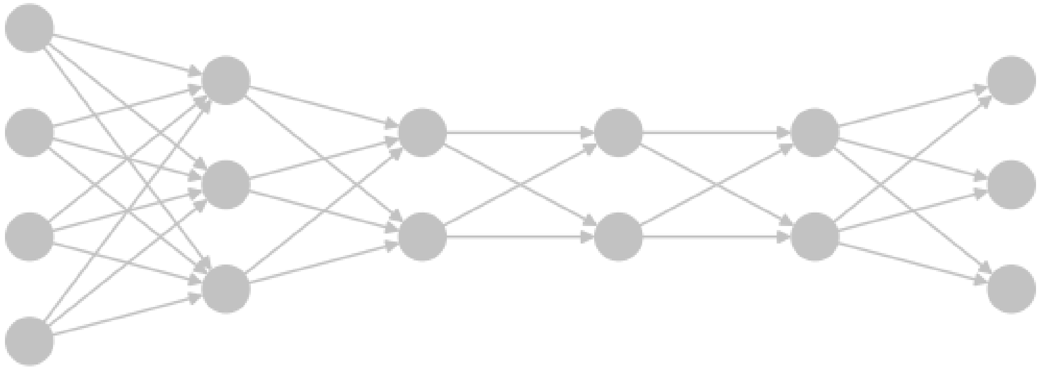


Edges = **possible** connections (hits between two adjacent layers)

The How

GNN implementation

Execution both in software and in hardware: No true edge information given



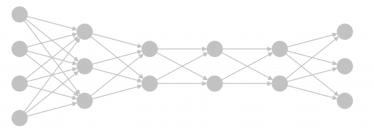
Let GNN work:
- Number of **hidden units decreased to 16** compared to reference
- Classify edges: true edge or false edge
- → Retrieve original track (pattern recognition)
- Uses graph (hits and possible connections) as input
- Data is not 2D or 1D anymore

The How

GNN implementation

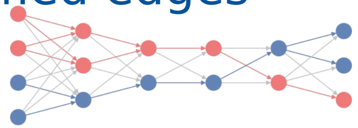
Structure/work flow of the GNN

Detector/Timepix3 → Readout & build graph (already on ACAP)



Classically: Executed sequentially

1) Loop edges:
   - Run neural network **“R1”**
2) Loop nodes:
   - Rearrange output of “R1” → Run neural network **“O”**
3) Loop edges:
   - Rearrange output of “O” → Run neural network **“R2”** & Sigmoid/apply threshold → Classified edges



Each NN a small MLP: lin. Layer → ReLU →lin. Layer → ReLU → lin. Layer with 16 hidden units

ACAP implementation

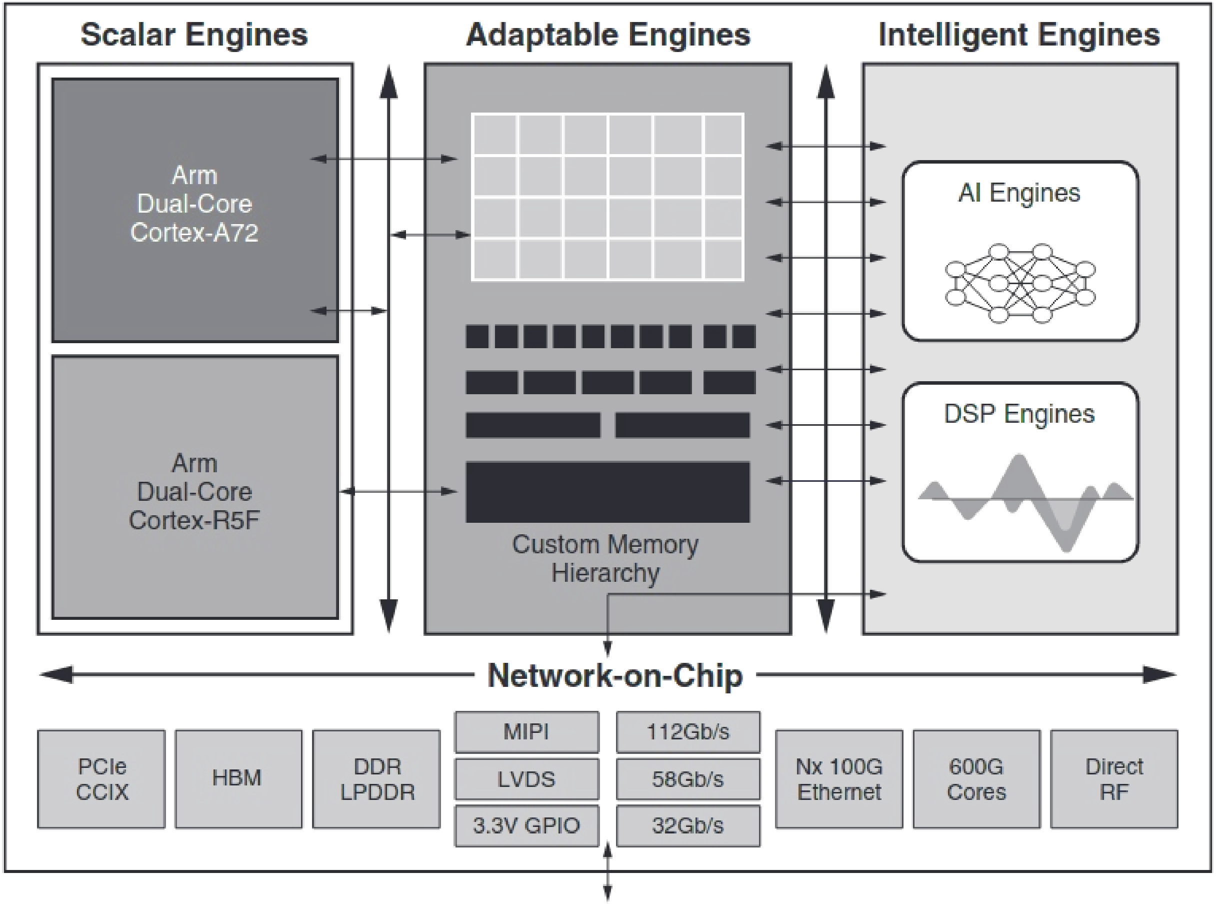
Process flow:

- GNN in PyTorch (from paper) in software as reference
- Transformed to pure **C/C++** without libraries→ compiled to machine code
- Existing MLIR comiler fragments (Xilinx work in progress)
- By hand conversion to assembler-like code for Xilinx Vitis
  No VHDL, no Verilog, no HLS → AIEs are **processors**, not FPGAs
- Supported bit-widths: 8b, 16b, 32b (also complex)
- First implementation: **32b floating-point (no quantization by then)**
- Quantisation: Down to 8b if reasonable by AIE data width
- **How to get the NN onto the ACAP?**
- → Remember: a) 400 processors, b) VLIW, c) SIMD
- Compiler helps but:
  - Each of the three has to be addressed during programming
  - Use of FPGA logic for some processes in evaluation

ACAP implementation

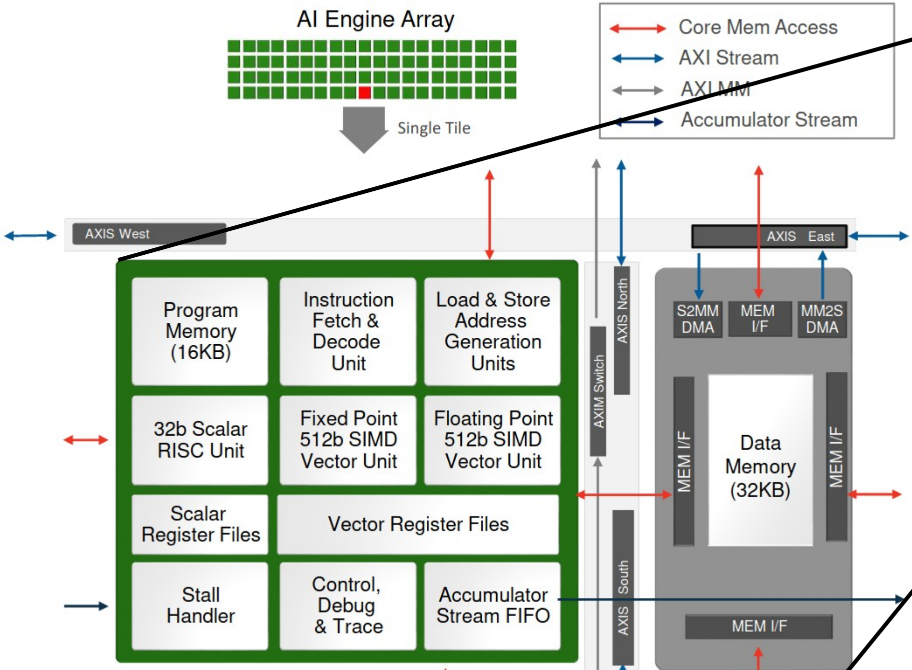
The How

ACAP overview

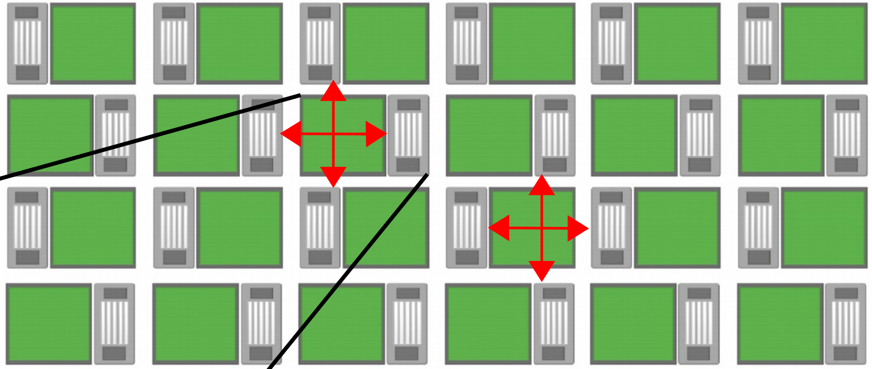


ACAP implementation

Deeper look into ACAP AI engines



Source: Xilinx, UG1079



Source: Xilinx

Access to “own” memory (32 KB)
+ North, South, East or West memory (3x 32KB)
→ 128 KB, seen as one contiguous memory

ACAP implementation

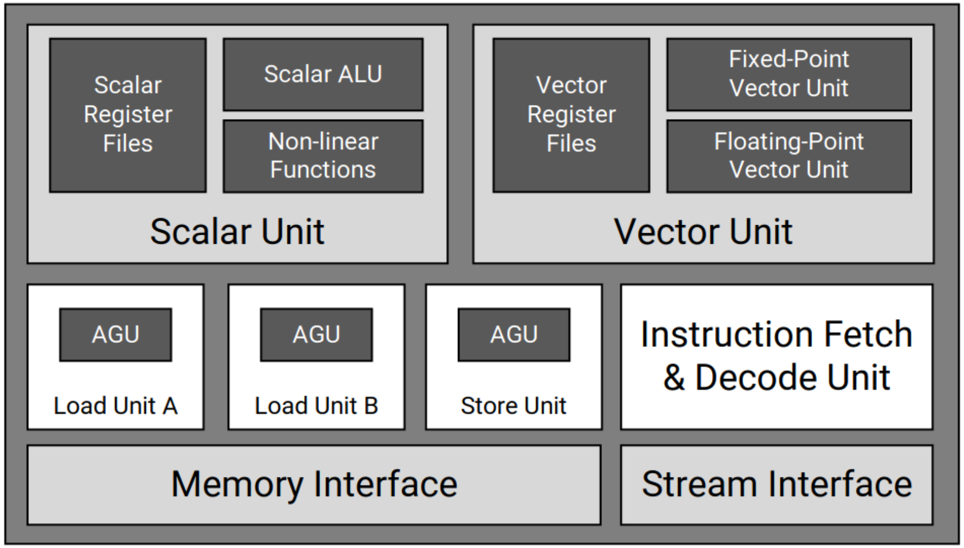
The How

Deeper look into ACAP AI engines

- Running at 1 GHz (VCK190)
- Highly parallelized: **400 processors, VLIW, SIMD**
- **VLIW (Very long instruction word)**
  → Simultaneous execution of:
  - 2x data load, 1x data store
  - 2x data move, 1x scalar operation
  - **1x vector operation → SIMD (Single instruction, multiple data)**
    with multiple accumulators:
    8 bit x 8 bit: 128 MAC/instruction
    16 bit x 16 bit: 32 MAC/instruction
    32 bit x 32 bit: 8 MAC/instruction

→ ILP (Instruction level parallelism)

Source: Xilinx UG1079



The How

GNN implementation

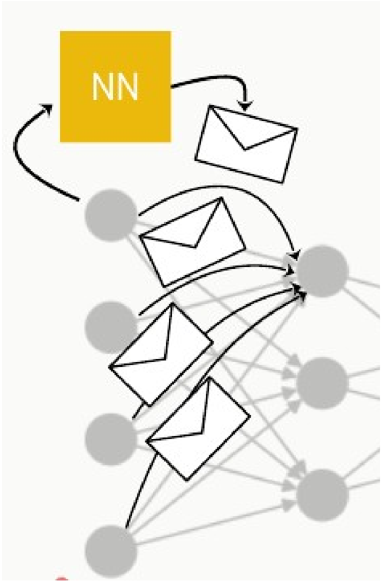
Implementation for the ACAP - Pipelining

Detector/Timepix3 → Readout & build graph (already on ACAP)

1) Loop edges:
- Run neural network **“R1”** → AIE 1

2) Loop nodes:
- Rearrange output of “R1” → AIE 2
  → Run neural network **“O”** → AIE 3

3) Loop edges:
- Rearrange output of “O” → AIE 4
  → Run neural network **“R2”** → AIE 5
  & Sigmoid/apply threshold →

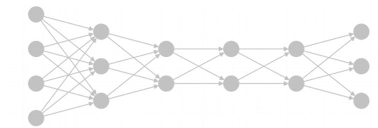


GNN implementation

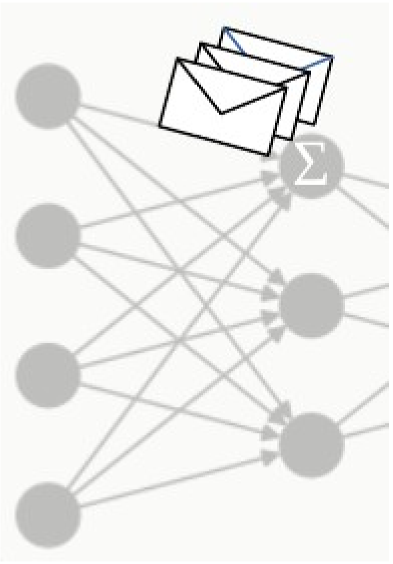
The How

Implementation for the ACAP - Pipelining

Detector/Timepix3 → Readout & build graph (already on ACAP)

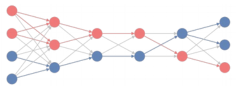


1)Loop edges:
- Run neural network **“R1”** → AIE 1



2)Loop nodes:
- Rearrange output of “R1” → AIE 2
  → Run neural network **“O”** → AIE 3

3)Loop edges:
- Rearrange output of “O” → AIE 4
  → Run neural network **“R2”** → AIE 5
  & Sigmoid/apply threshold → Classified edges

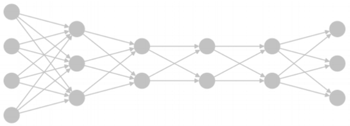


The How

GNN implementation

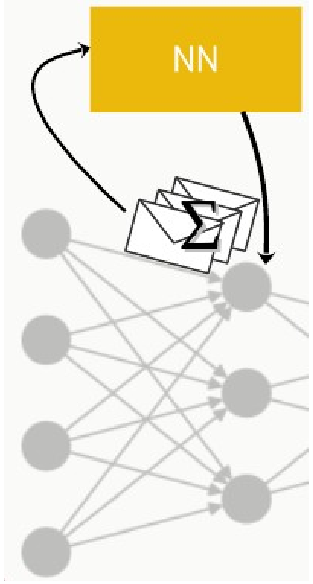
Implementation for the ACAP - Pipelining

Detector/Timepix3 → Readout & build graph (already on ACAP)

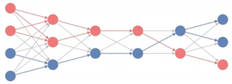


1)Loop edges:
- Run neural network **“R1”** → AIE 1

2)Loop nodes:
- Rearrange output of “R1” → AIE 2
  → Run neural network **“O”** → AIE 3



3)Loop edges:
- Rearrange output of “O” → AIE 4
  → Run neural network **“R2”** → AIE 5
  & Sigmoid/apply threshold → Classified edges

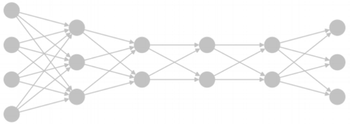


The How

GNN implementation

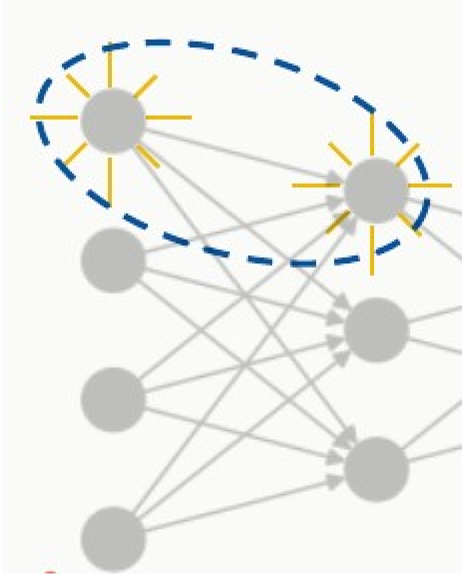
Implementation for the ACAP - Pipelining

Detector/Timepix3 → Readout & build graph (already on ACAP)

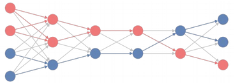


1)Loop edges:
- Run neural network **“R1”** → AIE 1

2)Loop nodes:
- Rearrange output of “R1” → AIE 2
  → Run neural network **“O”** → AIE 3



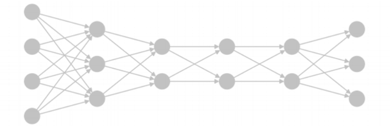
3)Loop edges:
- Rearrange output of “O” → AIE 4
  → Run neural network **“R2”** → AIE 5
  & Sigmoid/apply threshold → Classified edges



GNN implementation

Implementation for the ACAP - Pipelining

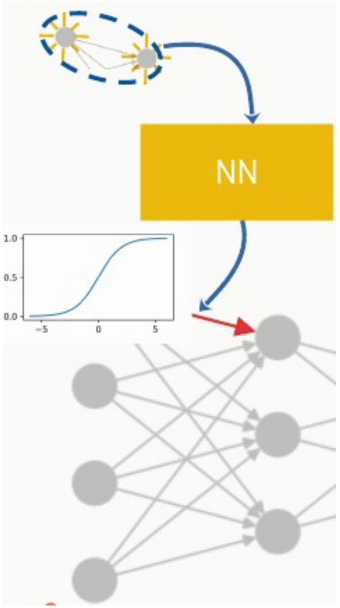
Detector/Timepix3 → Readout & build graph (already on ACAP)

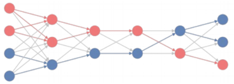


1)Loop edges:
- Run neural network **“R1”** → AIE 1

2)Loop nodes:
- Rearrange output of “R1” → AIE 2
  → Run neural network **“O”** → AIE 3

3)Loop edges:
- Rearrange output of “O” → AIE 4
  → Run neural network **“R2”** → AIE 5
  & Sigmoid/apply threshold → Classified edges

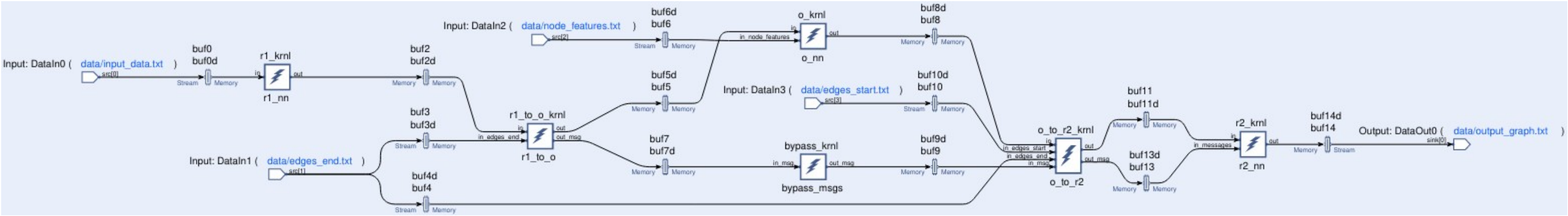




GNN implementation

The How

Implementation for the ACAP - Pipelining



⇒ AI engines not used efficiently, could do much more simultaneously

Load A	Load B	Store	Scalar Op.	Move A	Move B	Vector Op.
NOP	NOP	VST.SPIL wd1, [sp,	NOP	NOP	NOP	NOP
NOP	NOP	VST wc0, [sp, #-736	NOP	VMOV wr0, wr3	NOP	VSEL xc, ya.s32, r8, c5, r4, c5, c4, r8
NOP	NOP	VST wc1, [sp, #-768	NOP	VMOV wr1, wr3	NOP	NOP
NOP	NOP	VST.SPIL wd0, [sp,	NOP	NOP	NOP	NOP
NOP	NOP	VST wd1, [sp, #-83	NOP	VMOV wd0, wc0	NOP	NOP
VLDA wd0, [sp, #-5	NOP	NOP	NOP	NOP	NOP	VFPMAC wr3, r5, wr2, yd, r8, cl0, wc0, #0, cl5, #0, cl2

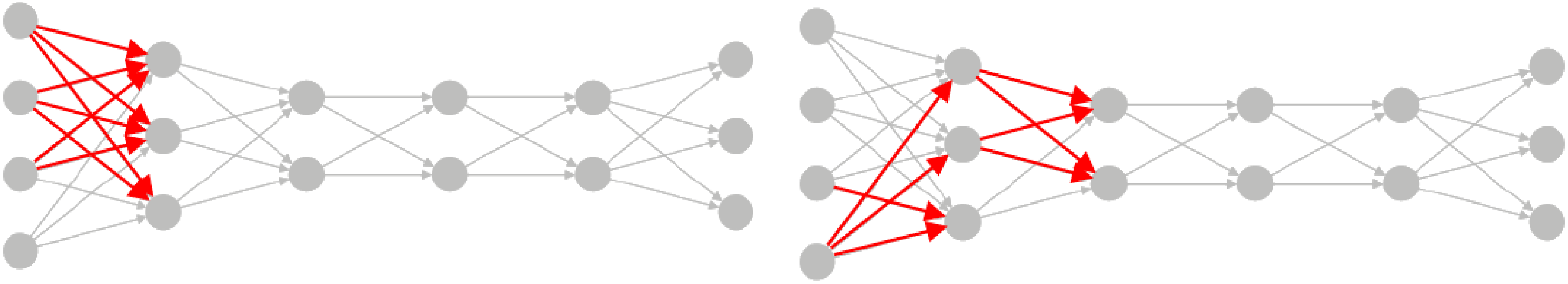
Helper tool developed to visualise AI engine use

→ one line = one VLIW → takes 1 ns to execute (but single units might take more time until finished)

GNN implementation

The How

Implementation for the ACAP - SIMD



1) Loop edges:

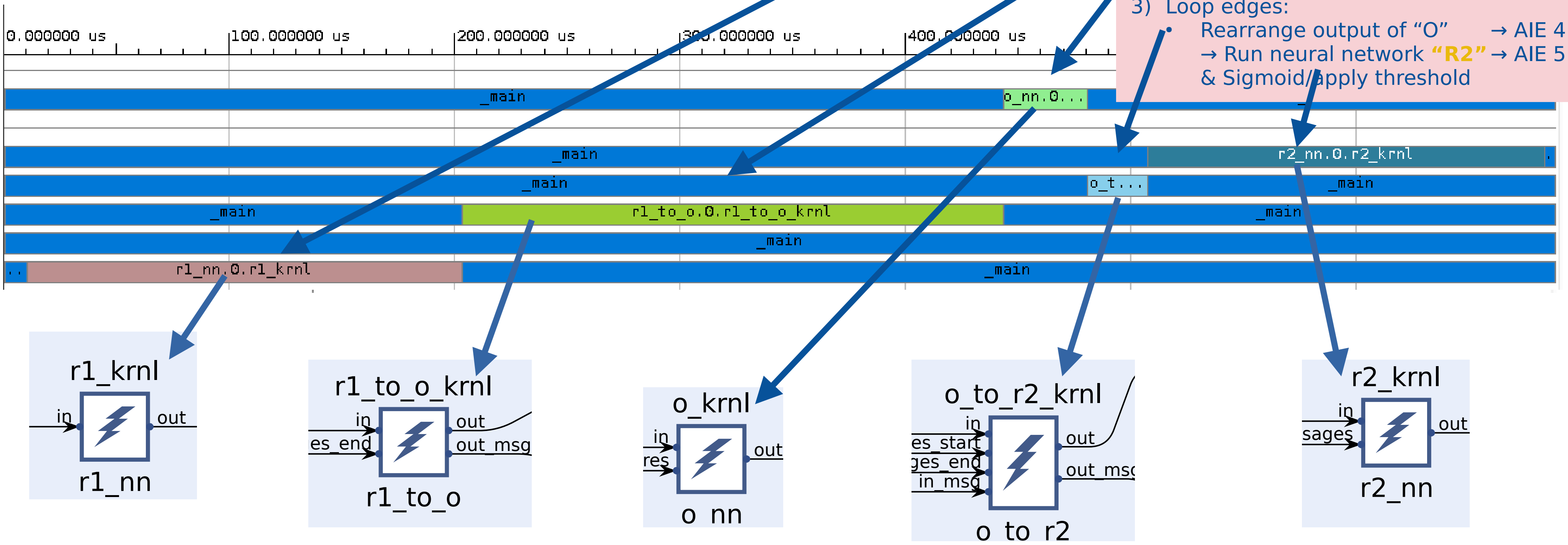
- Run neural network "R1" → AIE 1

...with 8 accumulators

GNN implementation

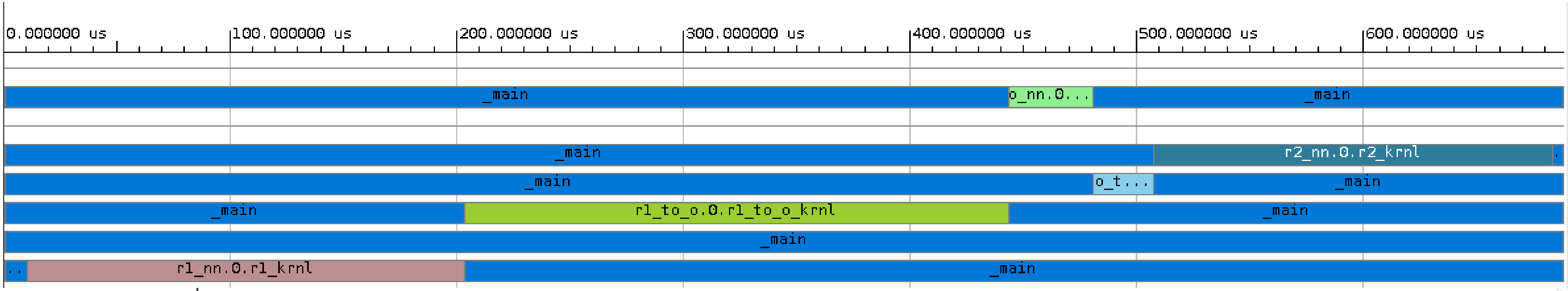
1) Loop edges:
   - Run neural network **"R1"** → AIE 1
2) Loop nodes:
   - Rearrange output of "R1" → AIE 2
     → Run neural network **"O"** → AIE 3
3) Loop edges:
   - Rearrange output of "O" → AIE 4
     → Run neural network **"R2"** → AIE 5
     & Sigmoid/apply threshold

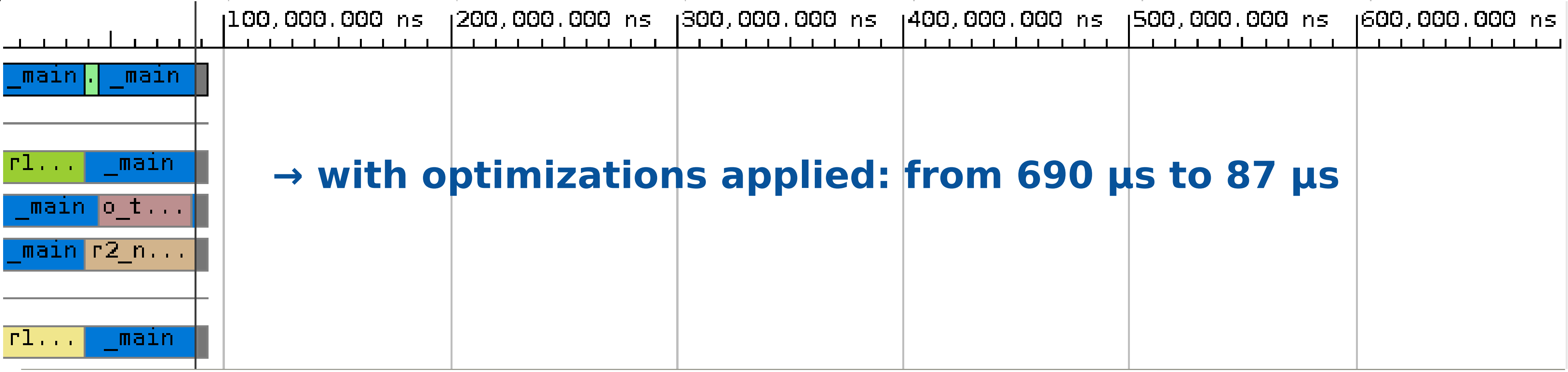
Inference time, for a graph with 384 edges and 82 nodes:



GNN optimisation: VLIW

Inference time, for a graph with 384 edges and 82 nodes:



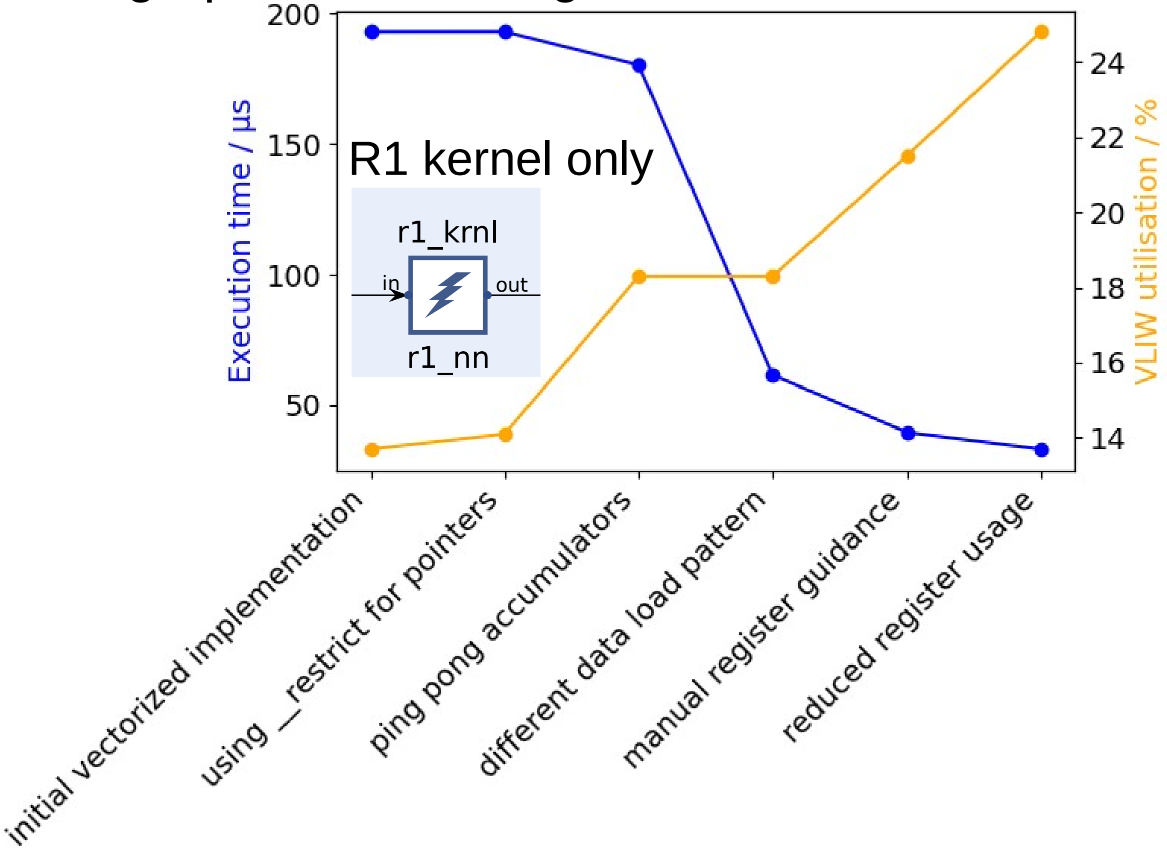


→ **with optimizations applied: from 690 µs to 87 µs**

GNN optimisation to AIE architecture

The How

Inference time, for a graph with 384 edges and 82 nodes:



GNN optimisation: Quantisation

The How

We want to go faster: Use more AI Engines and quantize

Quantization

- Limited by:

➔ Supported integer sizes: 32b, 16b, 8b (⇒16x more MACs/s)

➔ Vector register sizes: 128b, 256b, 512b

➔ 16 accumulators → e.g. compute 16 edges in parallel

- Can be mixed (8b x 16b vector multiplications, vectors in 128b and 256b registers)
- But can be challenging to implement
- 2 in-/output streams, can r/w 32 bit per instruction
- Additional Cascade stream between accumulators

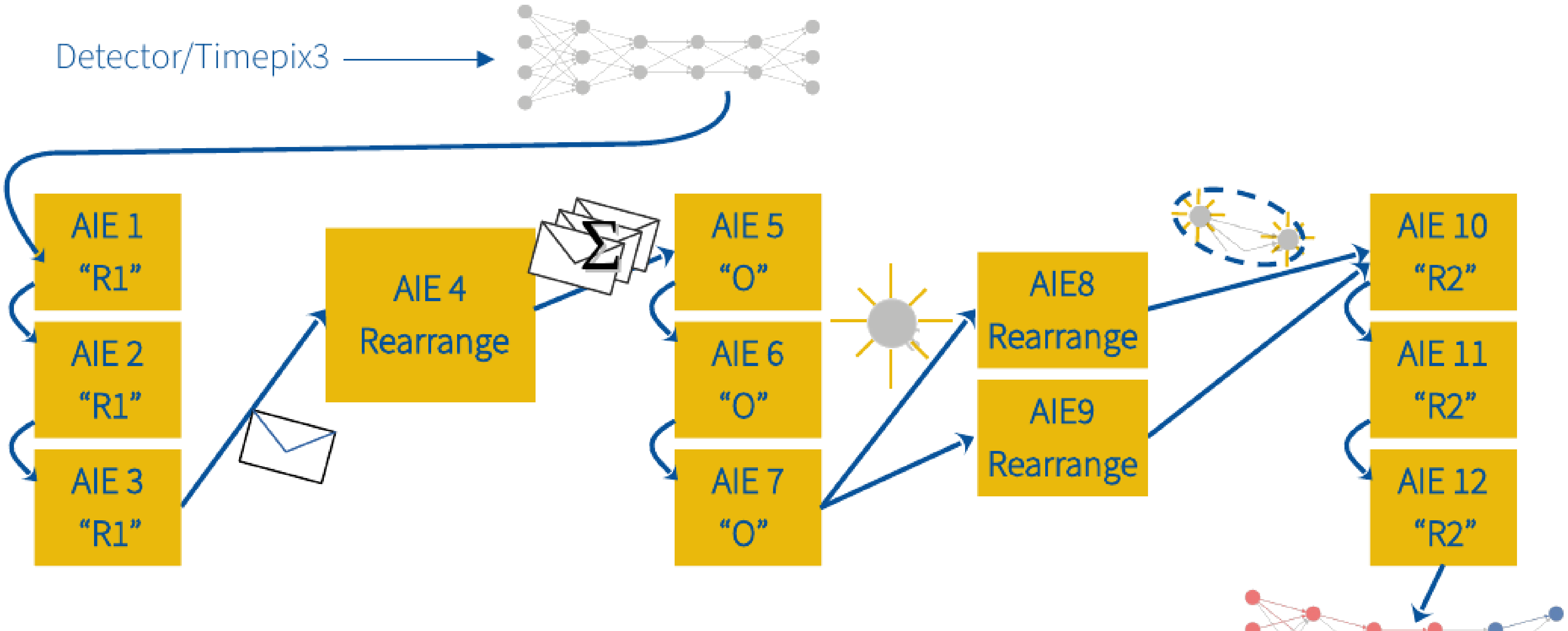
More AI Engines

1) Loop edges:
   - Run neural network "R1" → AIE 1,2,3

2) Loop nodes:
   - Rearrange output of "R1" → AIE 4
     → Run neural network "O" → AIE 5,6,7

3) Loop edges:
   - Rearrange output of "O" → AIE 8
     → Run neural network "R2" → AIE 9,10,11
     & Sigmoid/apply threshold

GNN optimisation: Quantisation

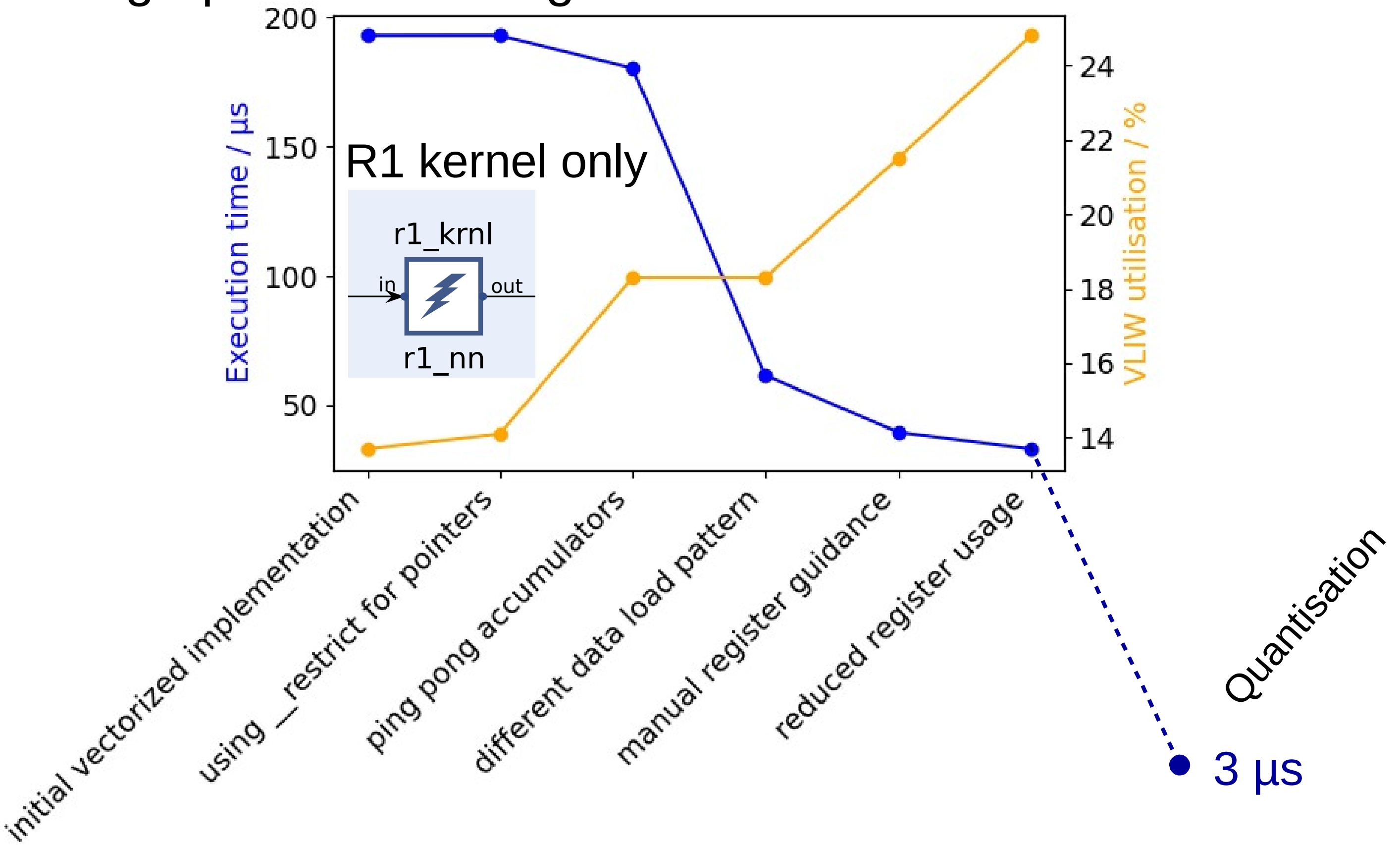


226 ns → 35 ns per classified edge

GNN optimisation: Quantisation

The How

Inference time, for a graph with 384 edges and 82 nodes:



GNN optimisation: Quantisation

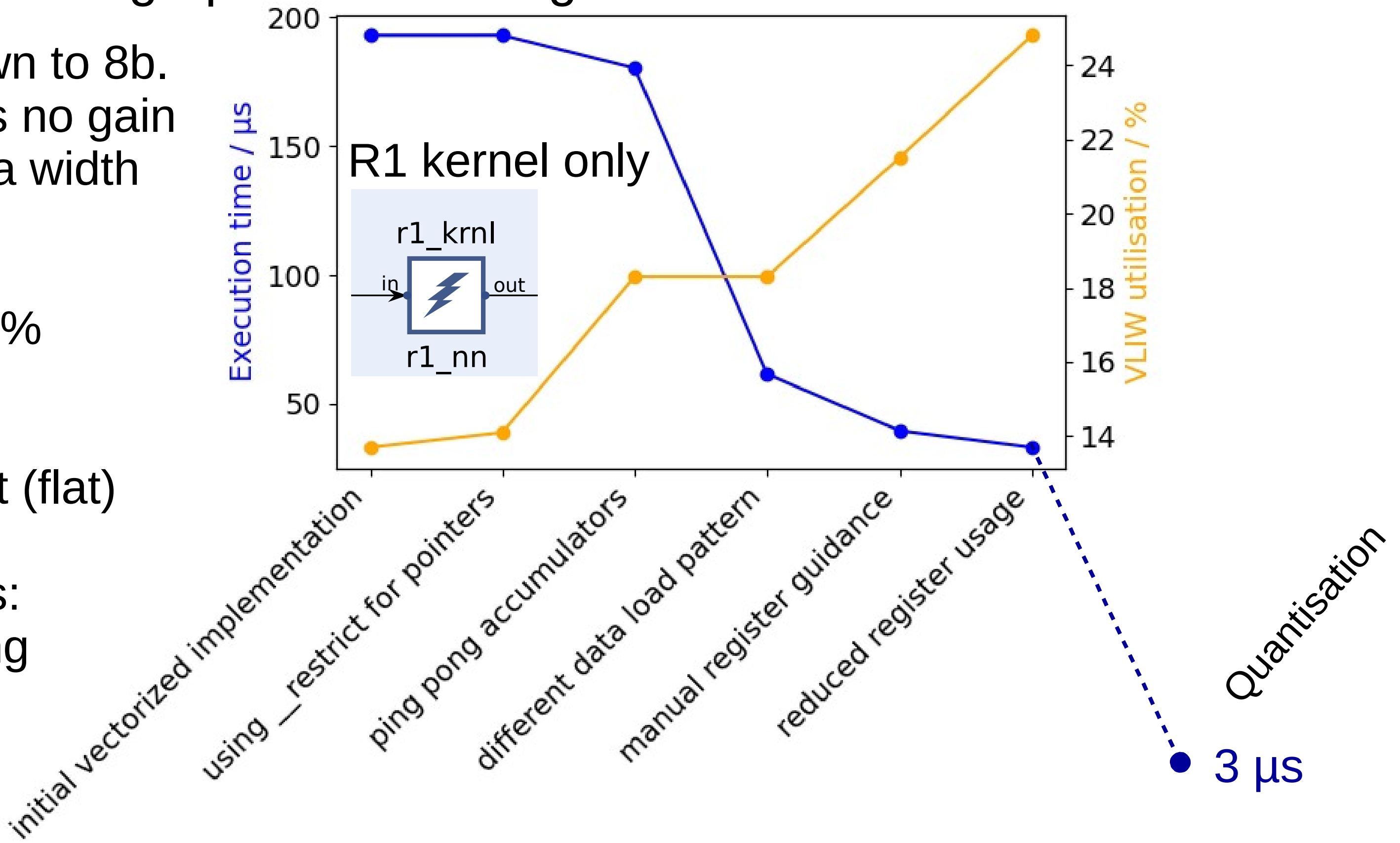
The How

Inference time, for a graph with 384 edges and 82 nodes:

Quantisation down to 8b.
Usually higher as no gain
by AI Engine data width

AUC: > 0.998,
Accuracy > 99.3 %
for edges
Data sample:
4-10 tracks/event (flat)

Work in Progress:
Quantised training
(Brevitas)

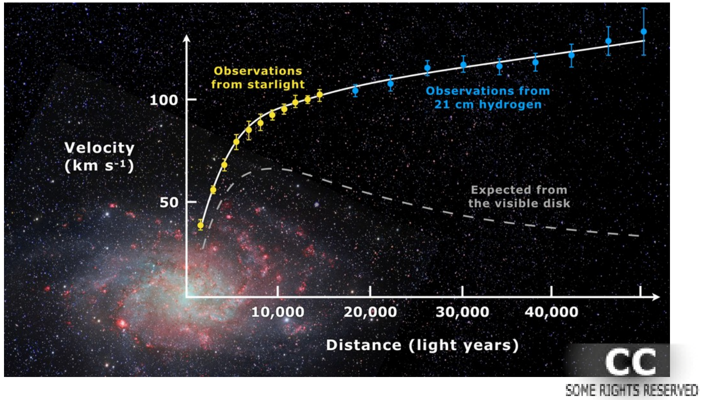


The How

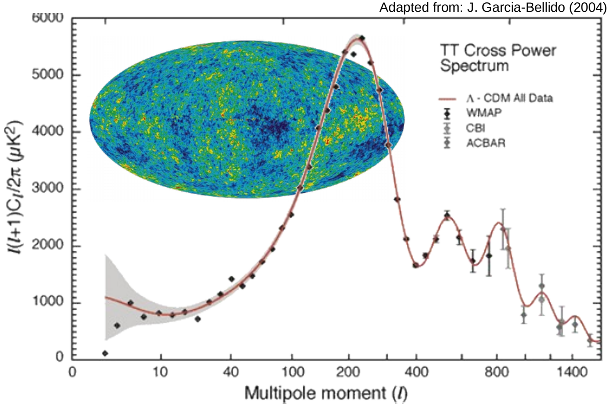
Fundamental physics driver example: Contribute to the discovery of LDM or its exclusion at world-best limits with the Lohengrin concept

DARK MATTER

- Clear evidence for a non-luminous form of matter



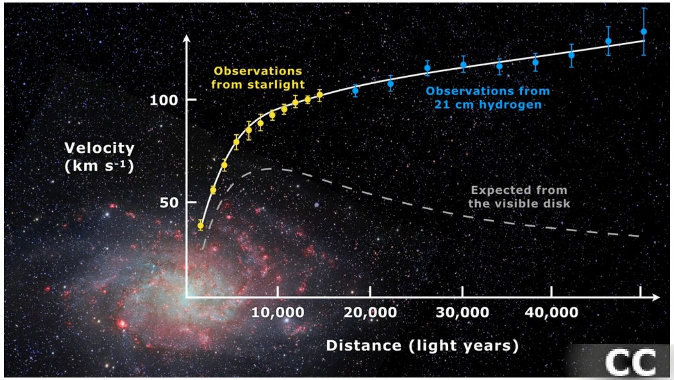
Adapted from: J. Garcia-Bellido (2004)



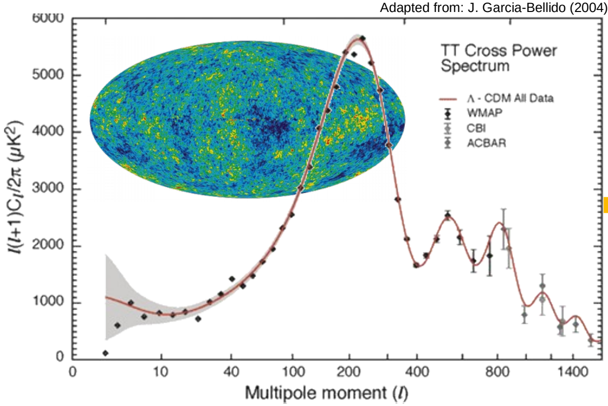
The How

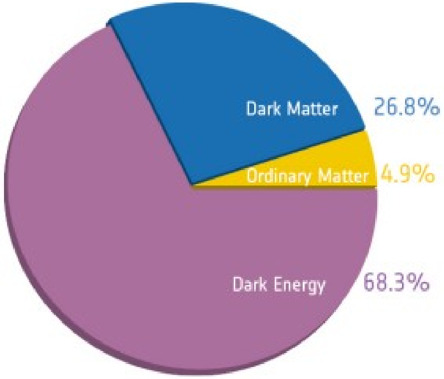
DARK MATTER

- Clear evidence for a non-luminous form of matter



Adapted from: J. Garcia-Bellido (2004)



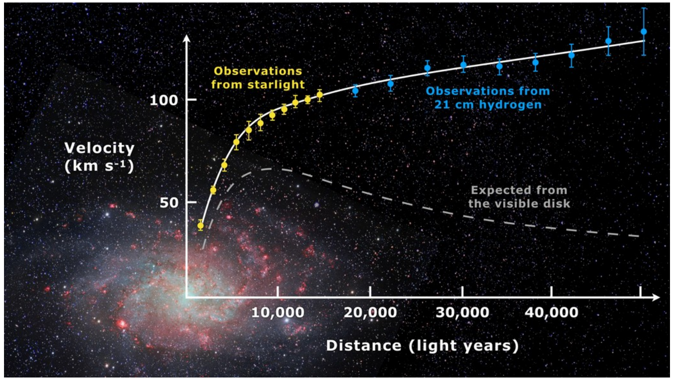


https://sci.esa.int/s/8o9qJ0w

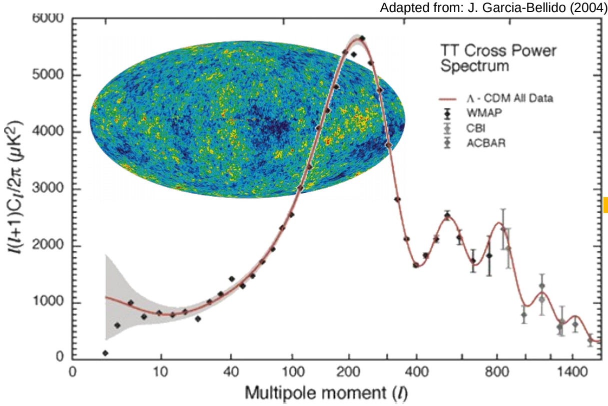
The How

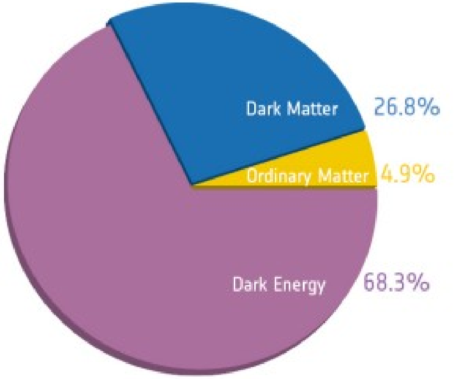
DARK MATTER

- Clear evidence for a non-luminous form of matter



Adapted from: J. Garcia-Bellido (2004)





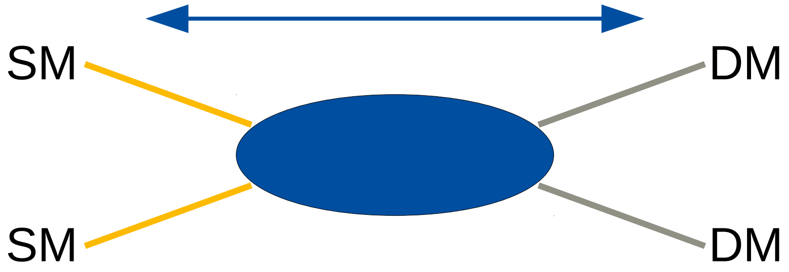
https://sci.esa.int/s/8o9qJ0w

- Dark Matter: one of the fundamental puzzles of particle physics
- Standard Model of Particle physics cannot explain Dark Matter
- What is the nature of Dark Matter? What is its origin?

THERMAL RELIC DARK MATTER

The How

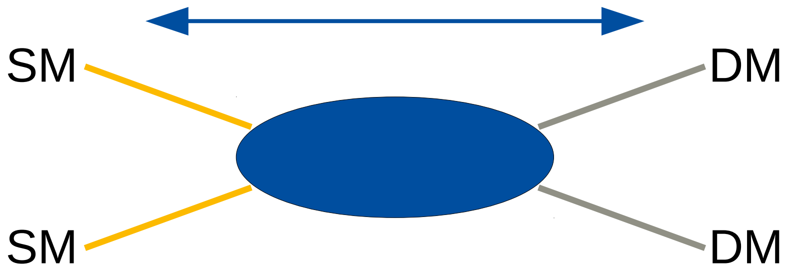
- Early universe: Dark Matter in thermal equilibrium (=interaction)



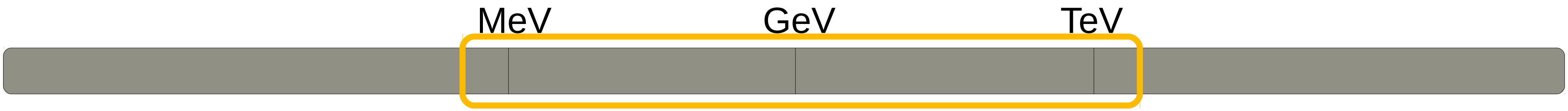
THERMAL RELIC DARK MATTER

The How

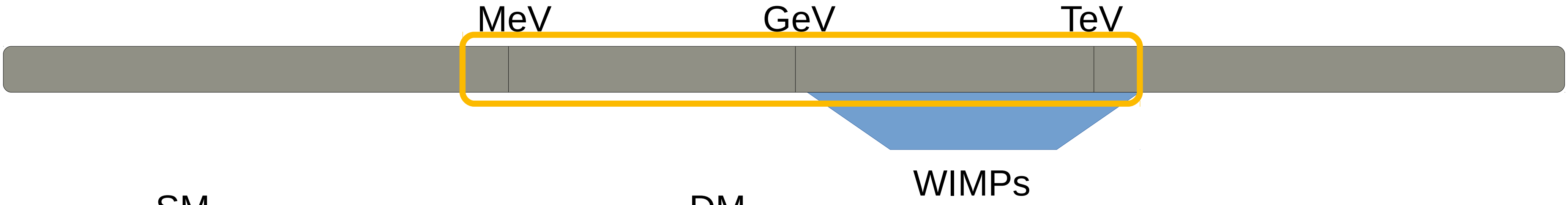
- Early universe: Dark Matter in thermal equilibrium (=interaction)

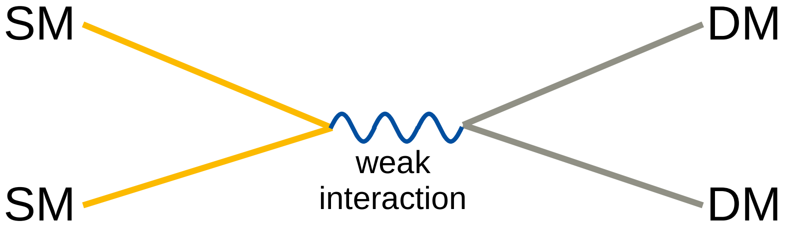


- Expansion ⇒ Freeze-out, Dark Matter abundance fixed ➞ Constraints on mass



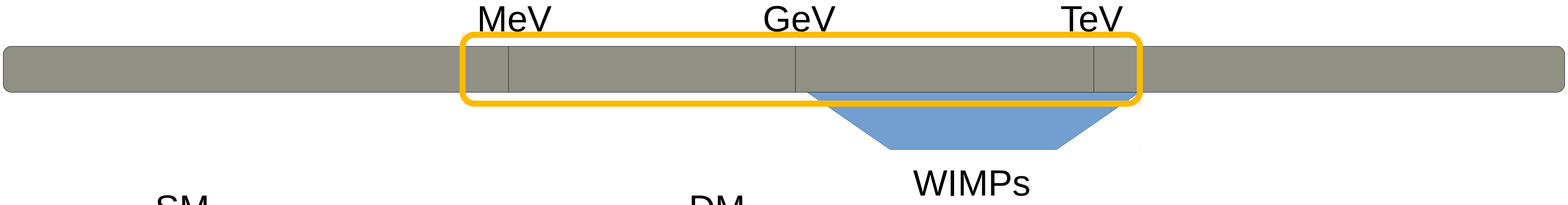
THERMAL RELIC DARK MATTER

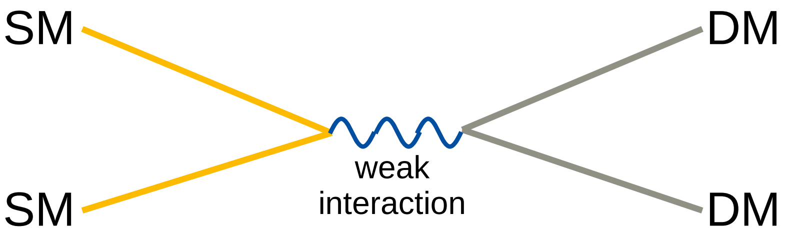




- Weakly interacting massive particles (WIMPs):
  - Expected properties fit ➞ popular

THERMAL RELIC DARK MATTER





- Weakly interacting massive particles (WIMPs):
  - Expected properties fit ➞ popular
  - Parameter space almost closed

nature

Explore content ⌄ About the journal ⌄ Publish with us ⌄ Subscribe

nature > news > article

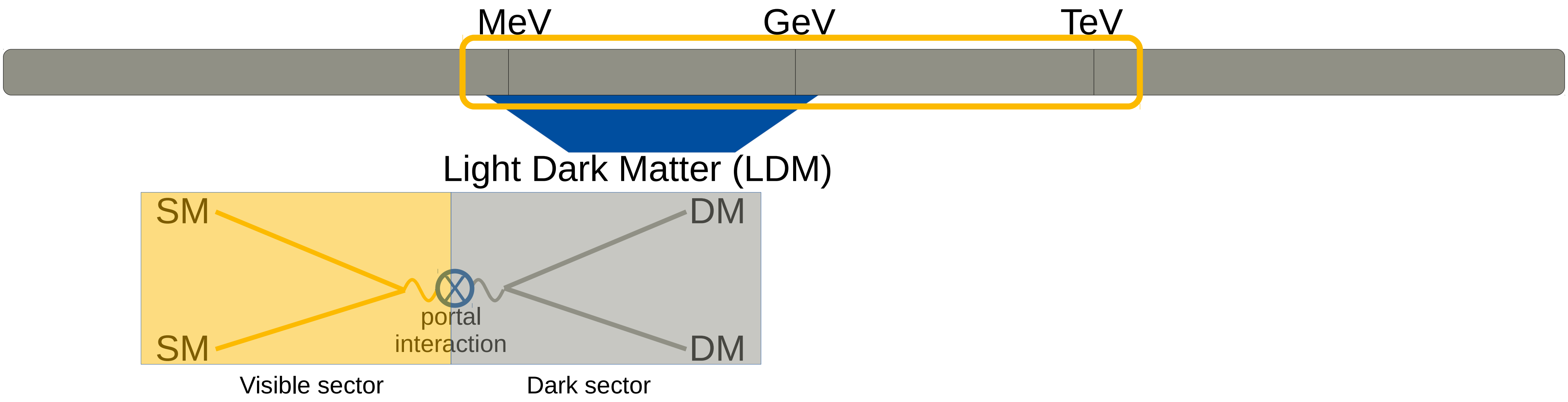
NEWS | 02 October 2020

Last chance for WIMPs: physicists launch all-out hunt for dark-matter candidate

Researchers have spent decades searching for the elusive particles – a final generation of detectors should leave them no place to hide.

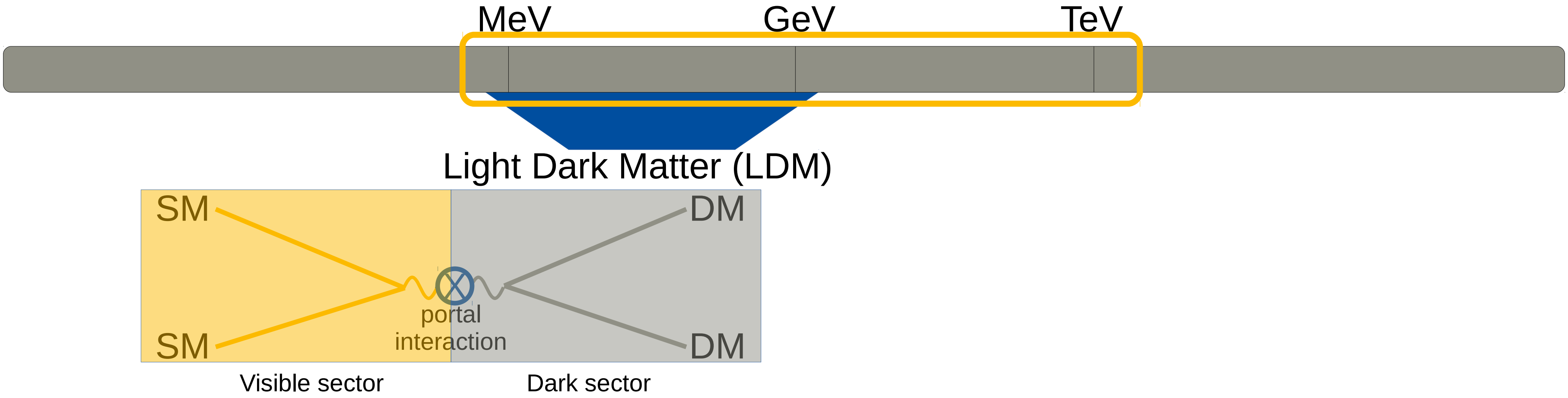
THERMAL RELIC DARK MATTER

The How



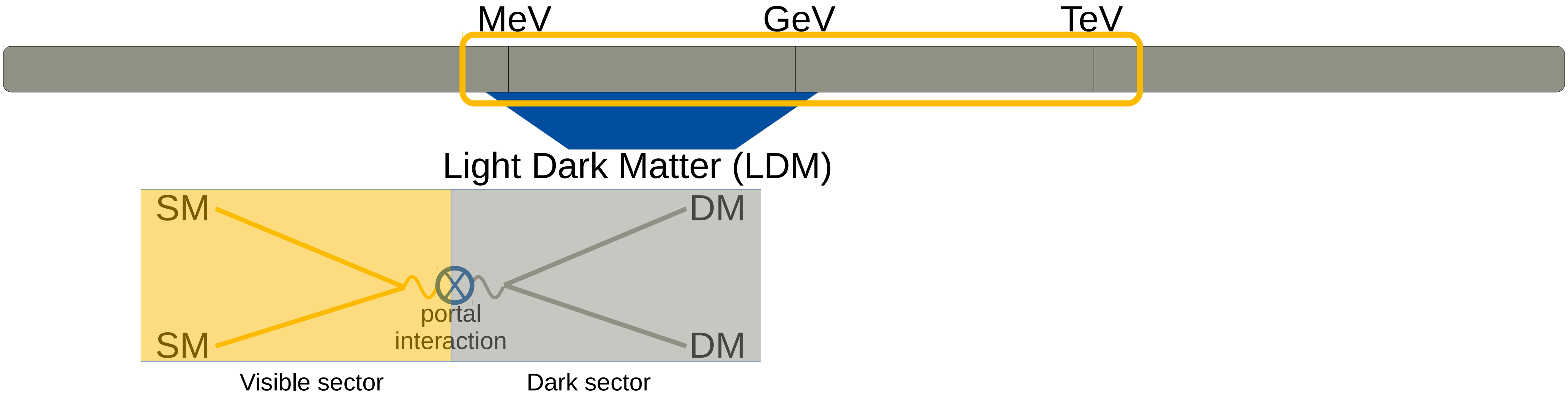
The How

THERMAL RELIC DARK MATTER



- Dark Sector extension of the Standard Model (SM)
  - Introduce new SM-DM mediator particle (portals) → Dark Photon (DP) *A'*

THERMAL RELIC DARK MATTER



- Dark Sector extension of the Standard Model (SM)
  - Introduce new SM-DM mediator particle (portals) → Dark Photon (DP) *A'*
  - Kinetic mixing of DP and SM photon

LIGHT DARK MATTER PRODUCTION

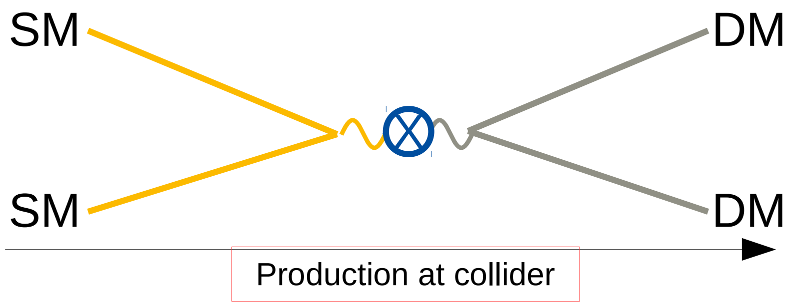
The How

- Dark Matter particles χ can be introduced
  - Coupling to DP *A'* ➞ *portal*

LIGHT DARK MATTER PRODUCTION

The How

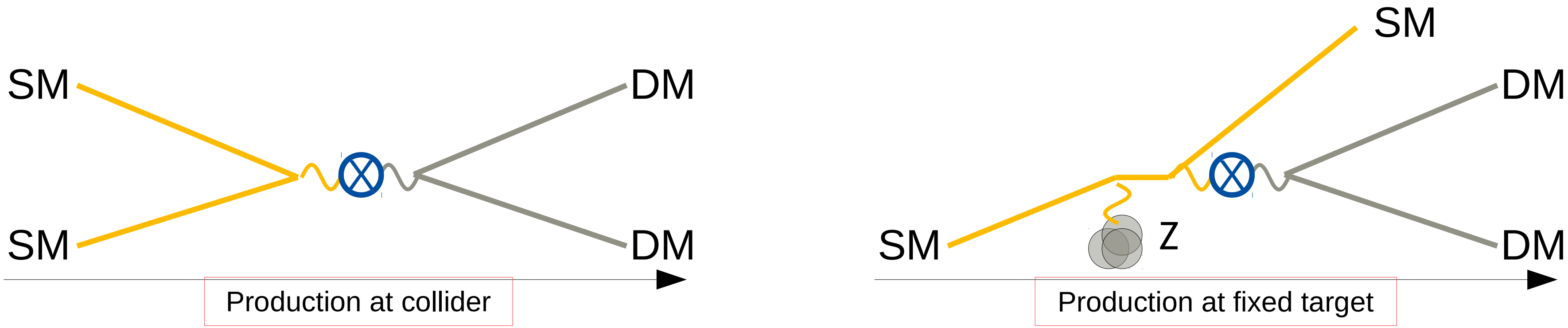
- Dark Matter particles χ can be introduced
  - Coupling to DP *A'* ➞ *portal*
  - At accelerator, process $f\,f$ ➞ *A'** ➞ $\chi\,\chi$ allows DM production



LIGHT DARK MATTER PRODUCTION

The How

- Dark Matter particles χ can be introduced
  - Coupling to DP *A'* → *portal*
  - At accelerator, process $f\,f \rightarrow A'^{*} \rightarrow \chi\,\chi$ allows DM production
  - Fixed-target experiments beneficial: Dark bremsstrahlung of fermion f



LDMX AND LOHENGRIN

The
How

- Design of an LDM experiment
  - Fermion beam ➝ electrons e
  - Fixed target
  - Tag dark bremsstrahlung

    ➝ e, missing momentum/energy

    ⇒ Tracker for incoming and outgoing e

    ⇒ Calorimeter for scattered e

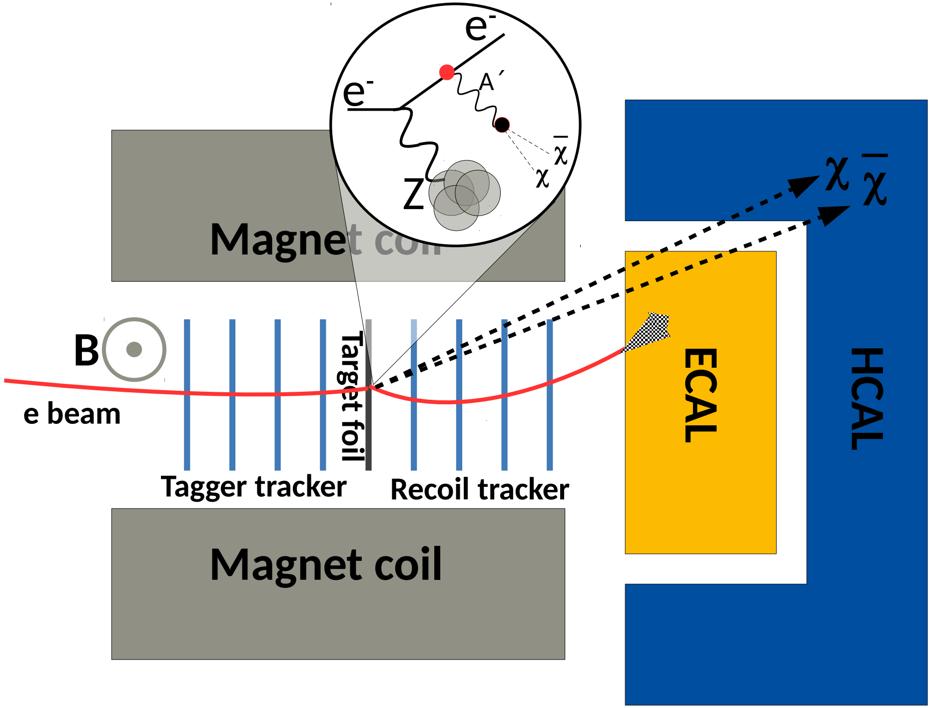
LDMX AND LOHENGRIN

- Design of an LDM experiment
  - Fermion beam ➝ electrons e
  - Fixed target
  - Tag dark bremsstrahlung

    ➝ e, missing momentum/energy

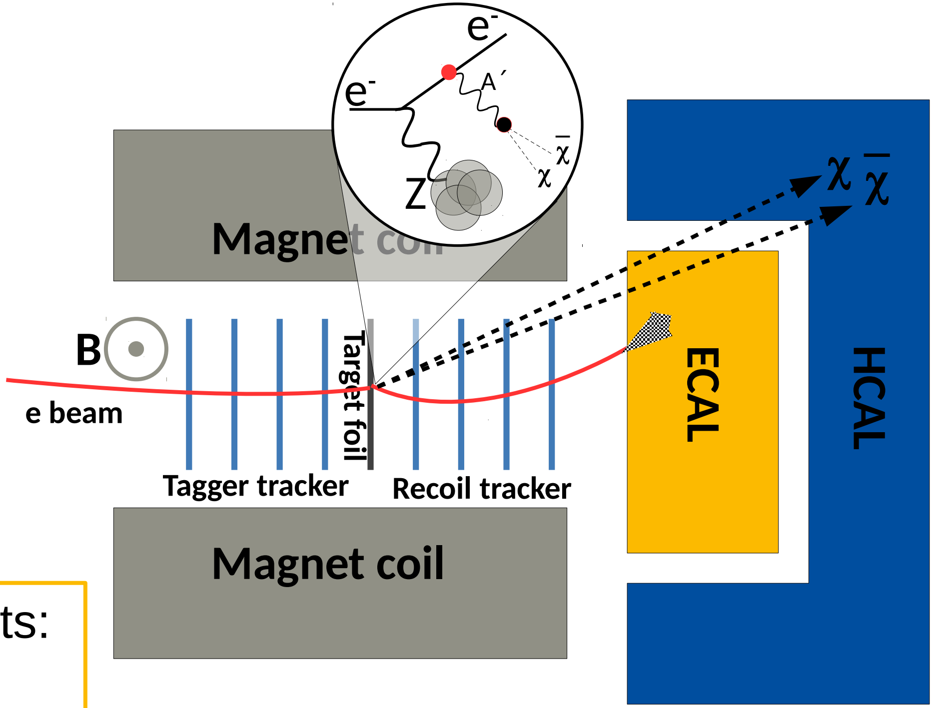
    ⇒ Tracker for incoming and outgoing e

    ⇒ Calorimeter for scattered e



LDMX AND LOHENGRIN

- Design of an LDM experiment
  - Fermion beam ➞ electrons e
  - Fixed target
  - Tag dark bremsstrahlung
    ➞ e, missing momentum/energy
    ⇒ Tracker for incoming and outgoing e
    ⇒ Calorimeter for scattered e



Similar detector layout – different readout concepts:

- LDMX @ SLAC, Stanford U, USA
- Lohengrin @ ELSA, U Bonn, Germany

The How

LDMX AND LOHENGRIN

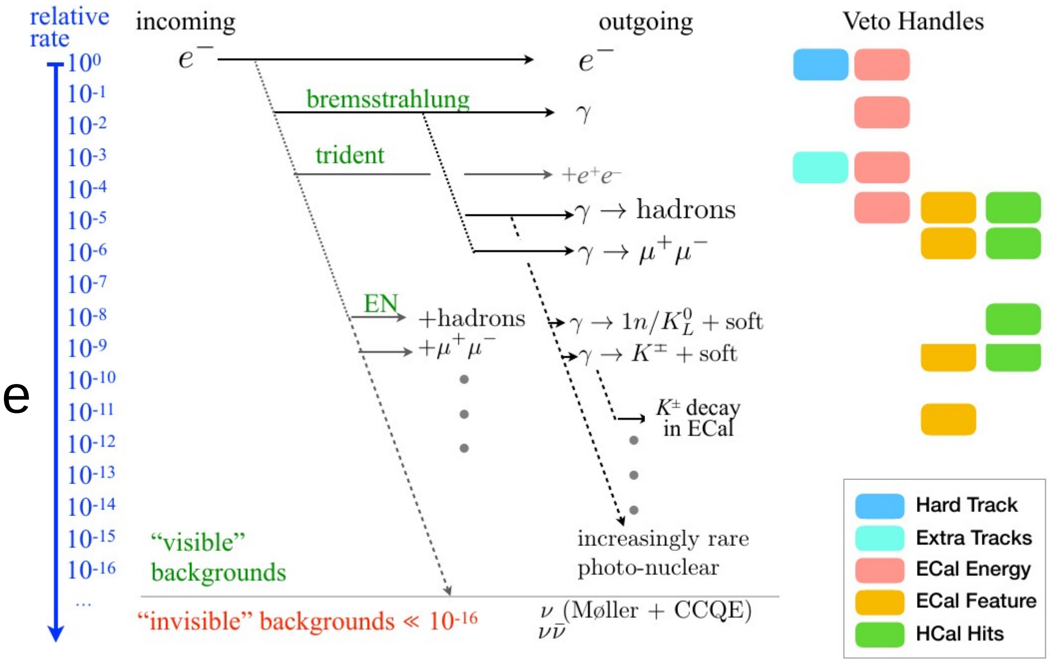
- Design of an LDM experiment
  - Fermion beam ➝ electrons e
  - Fixed target
  - Tag dark bremsstrahlung

    ➝ e, missing momentum/energy

    ⇒ Tracker for incoming and outgoing e

    ⇒ Calorimeter for scattered e
  - Suppress SM backgrounds

    ➝ Experimental layout: 10^{-16} rel. rate

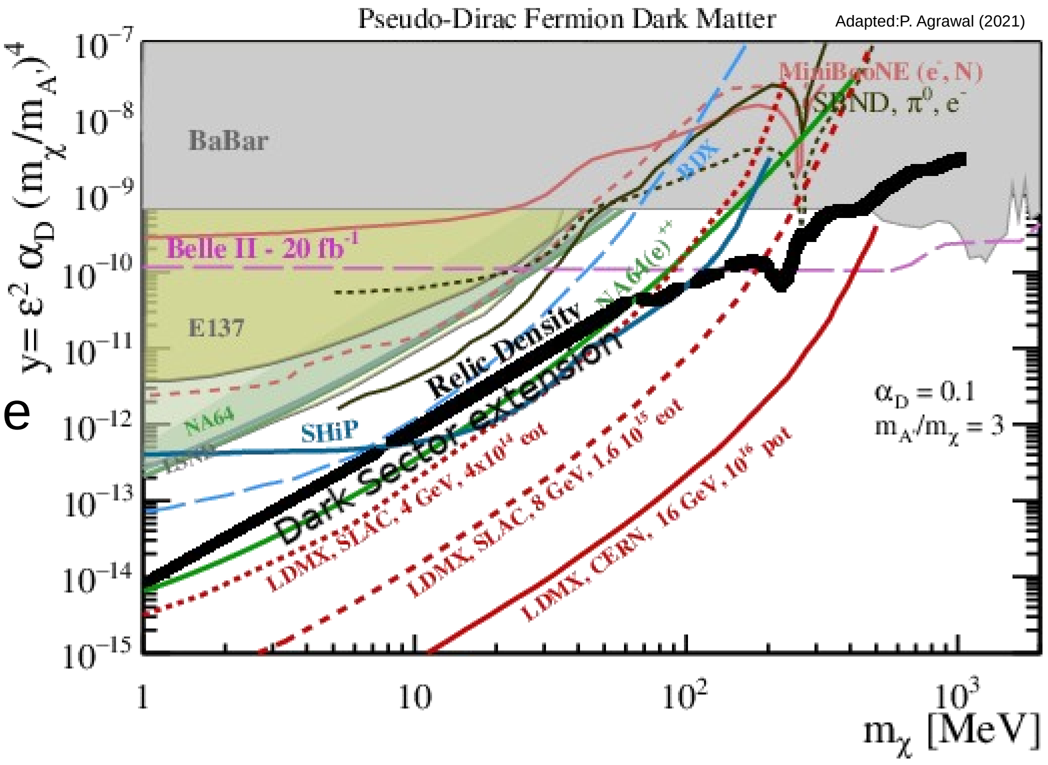


R. Pöttgen (2022)

LDMX AND LOHENGRIN

The How

- Design of an LDM experiment
  - Fermion beam ➝ electrons e
  - Fixed target
  - Tag dark bremsstrahlung
    ➝ e, missing momentum/energy
    ⇒ Tracker for incoming and outgoing e
    ⇒ Calorimeter for scattered e
  - Suppress SM backgrounds
    ➝ Experimental layout: 10^{-16} rel. rate



Adapted:P. Agrawal (2021)

LDMX AND LOHENGRIN

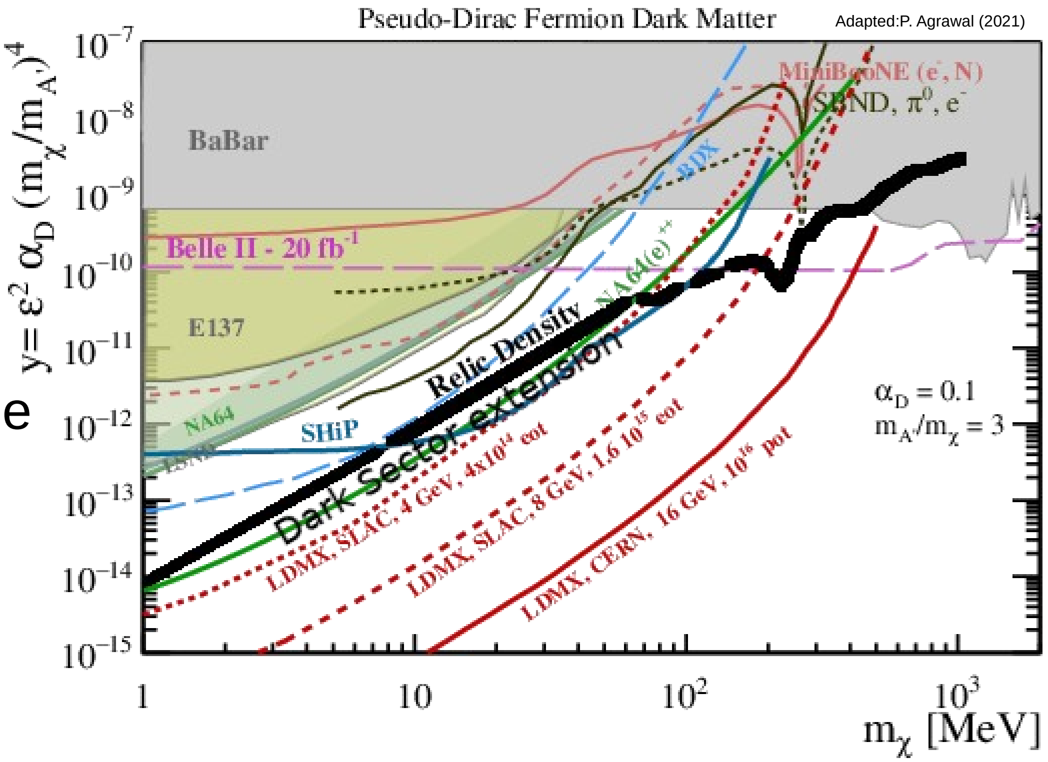
- Design of an LDM experiment
  - Fermion beam ➝ electrons e
  - Fixed target
  - Tag dark bremsstrahlung

    ➝ e, missing momentum/energy

    ⇒ Tracker for incoming and outgoing e

    ⇒ Calorimeter for scattered e
  - Suppress SM backgrounds

    ➝ Experimental layout: 10^{-16} rel. rate



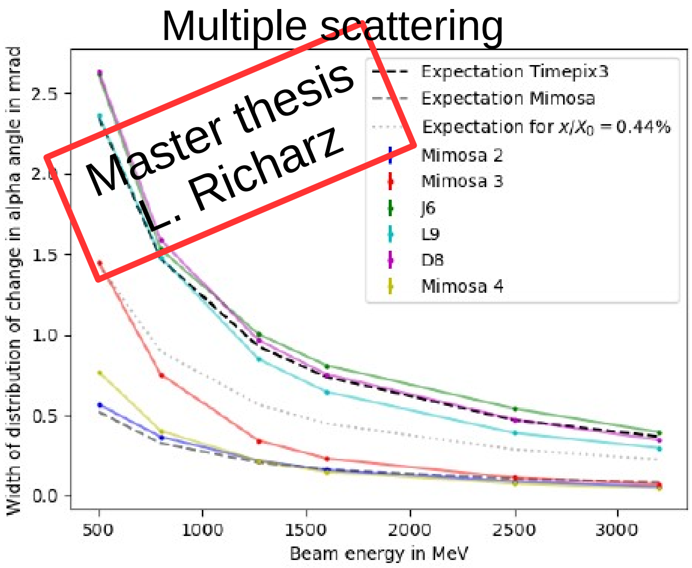
- Physics reach below theoretical relic density over wide mass range

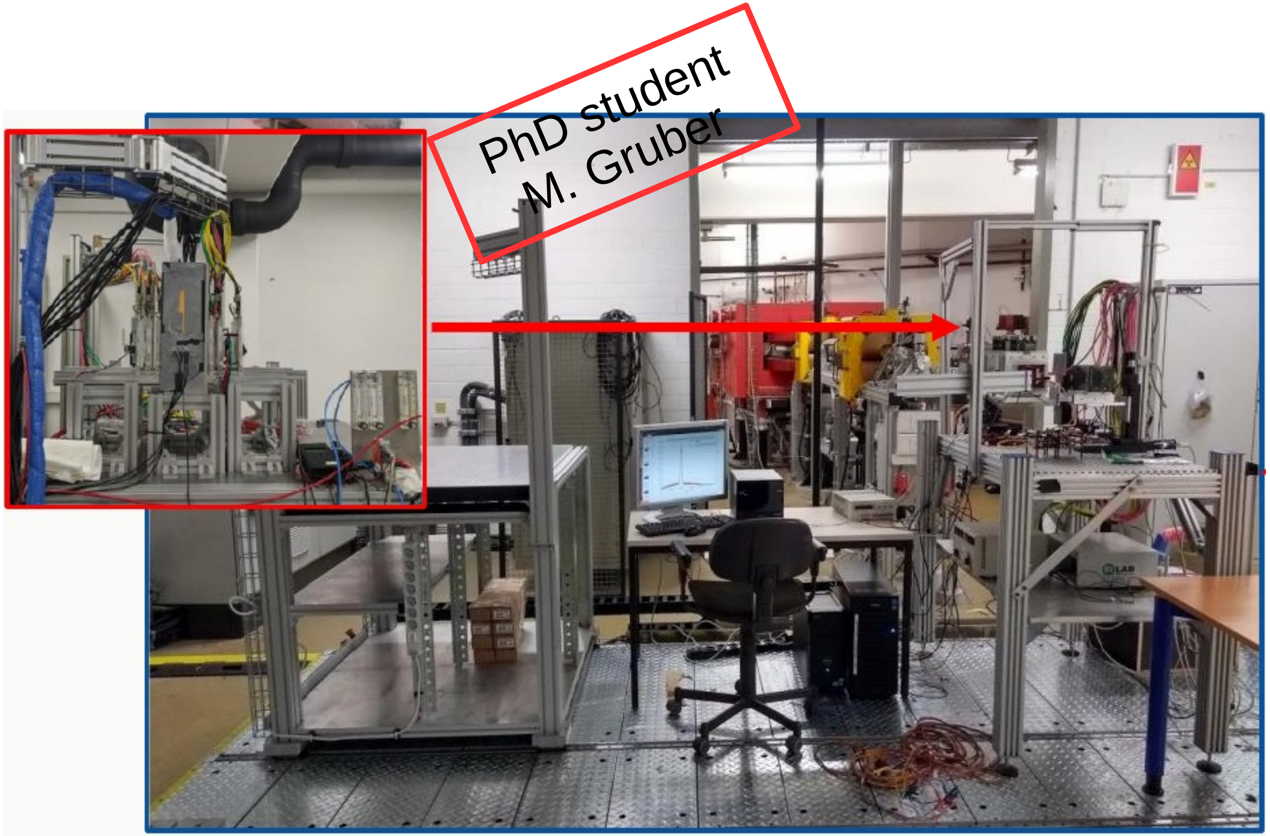
PRELIMINARY WORK

The How

Lohengrin efforts:

- Timepix3-based tracking detector as starting point

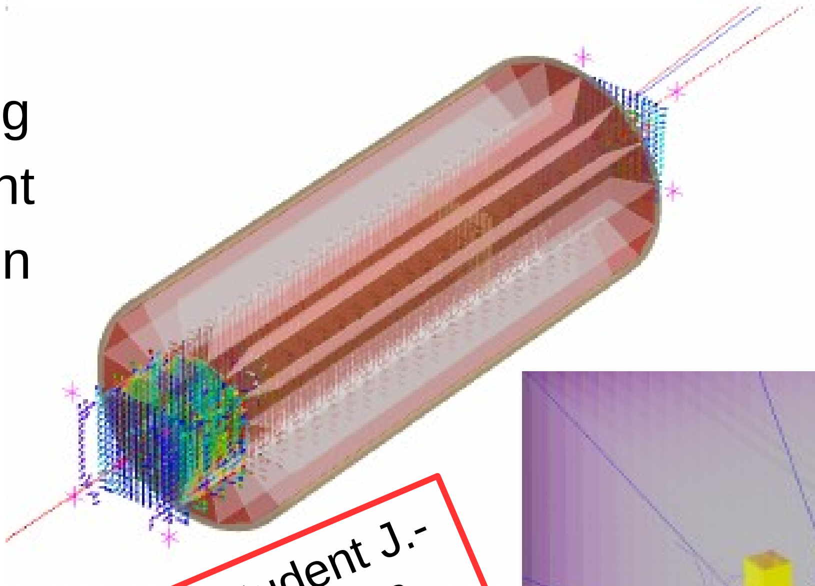




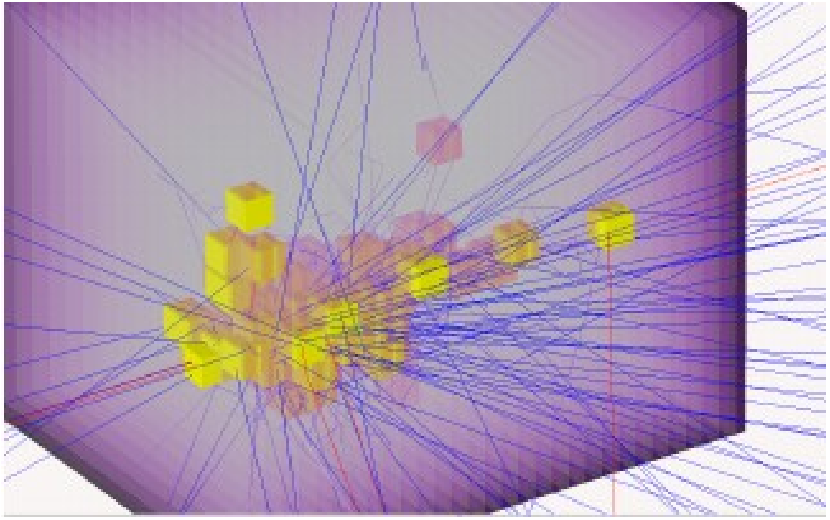
PRELIMINARY WORK

Lohengrin efforts:

- Timepix3-based tracking detector as starting point
- Experiment simulation in accelerator framework



PhD student J.-E. Heinrichs



Master Thesis/PhD Jan-Eric Heinrichs (Simulation/Bonn)

THE TECHNOLOGY CHALLENGE

The
How

- 10^{14} (first competitive results) – 10^{16} (physics reach) Electrons on Target (EoT) needed in meaningful time O(1) year $\Rightarrow$ O(3 – 300) MHz continuous beam rate

THE TECHNOLOGY CHALLENGE

The How

- 10^{14} (first competitive results) – 10^{16} (physics reach) Electrons on Target (EoT) needed in meaningful time O(1) year $\Rightarrow$ O(3 – 300) MHz continuous beam rate
- Lohengrin strategy to reach 10^{16} EoT earlier than competitors: track each e

THE TECHNOLOGY CHALLENGE

The How

- 10^{14} (first competitive results) – 10^{16} (physics reach) Electrons on Target (EoT) needed in meaningful time O(1) year $\Rightarrow$ O(3 – 300) MHz continuous beam rate
- Lohengrin strategy to reach 10^{16} EoT earlier than competitors: track each e
  - High rate detectors ➝ latest technology, ongoing R&D

THE TECHNOLOGY CHALLENGE

The
How

- 10^{14} (first competitive results) – 10^{16} (physics reach) Electrons on Target (EoT) needed in meaningful time O(1) year $\Rightarrow$ O(3 – 300) MHz continuous beam rate
- Lohengrin strategy to reach 10^{16} EoT earlier than competitors: track each e
  - High rate detectors → latest technology, ongoing R&D
  - Instantaneous data rate $\gg$ bandwidth & processing power
  - Complex background processes at still high relative rates
  - Interesting event rate at extremely low relative rates

THE TECHNOLOGY CHALLENGE

The How

- 10^{14} (first competitive results) – 10^{16} (physics reach) Electrons on Target (EoT) needed in meaningful time O(1) year $\Rightarrow$ O(3 – 300) MHz continuous beam rate
- Lohengrin strategy to reach 10^{16} EoT earlier than competitors: track each e
  - High rate detectors ➝ latest technology, ongoing R&D
  - Instantaneous data rate $\gg$ bandwidth & processing power
  - Complex background processes at still high relative rates
  - Interesting event rate at extremely low relative rates

Requires novel technologies

THE TECHNOLOGY CHALLENGE

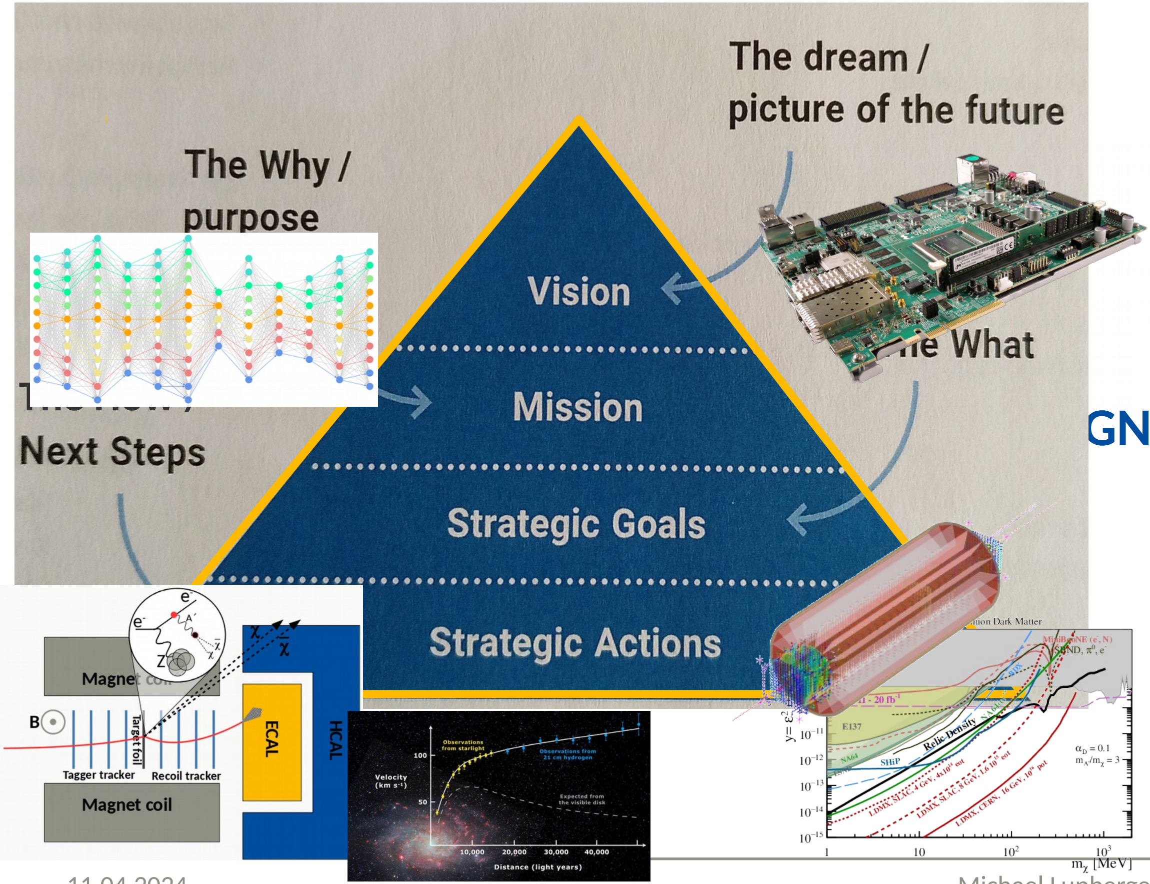
The How

- 10^{14} (first competitive results) – 10^{16} (physics reach) Electrons on Target (EoT) needed in meaningful time O(1) year $\Rightarrow$ O(3 – 300) MHz continuous beam rate
- Lohengrin strategy to reach 10^{16} EoT earlier than competitors: track each e
  - High rate detectors ➝ latest technology, ongoing R&D
  - Instantaneous data rate $\gg$ bandwidth & processing power
  - Complex background processes at still high relative rates
  - Interesting event rate at extremely low relative rates

Requires novel technologies

General challenge of particle physics: Intensity frontier

SUMMARY



Sustainable smart-data acquisition and intelligent feature extraction

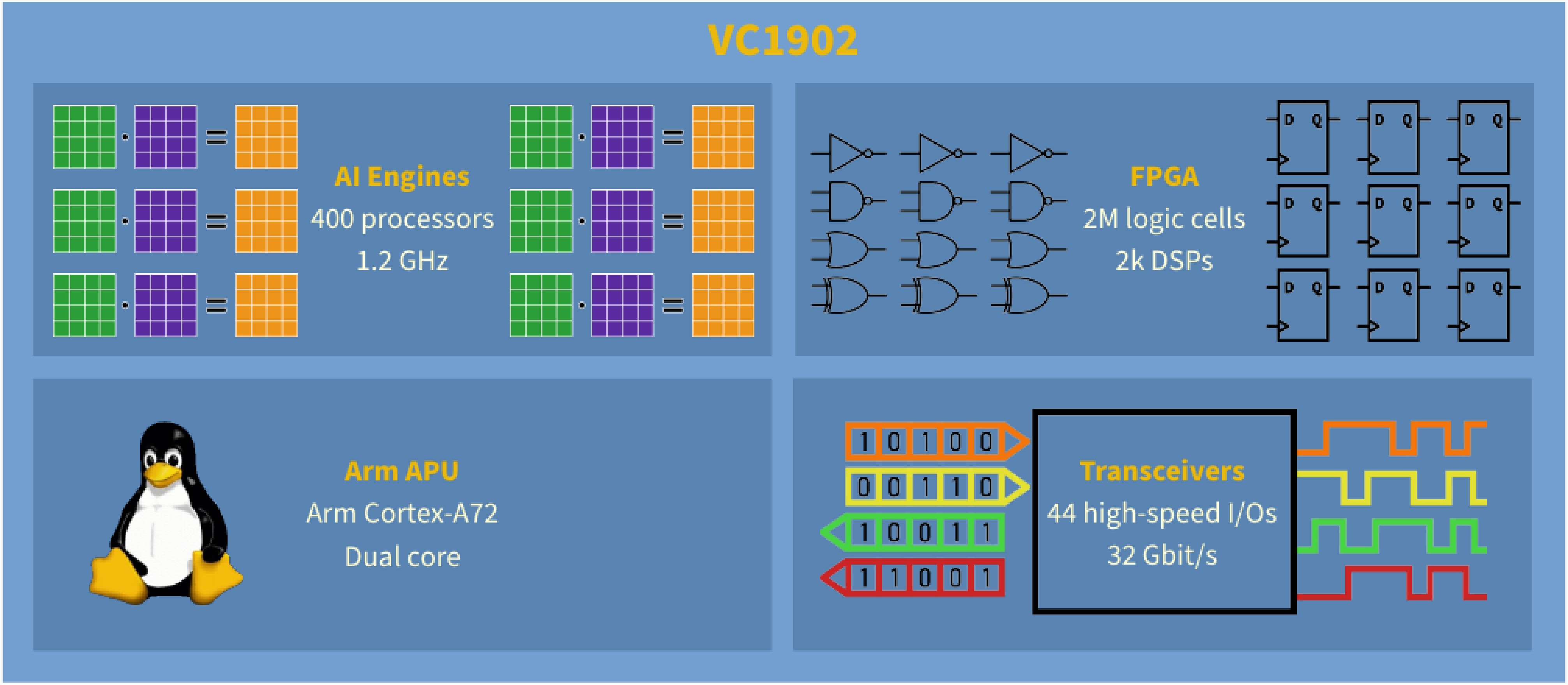
with

GNN live tracking and triggering on the Versal

to

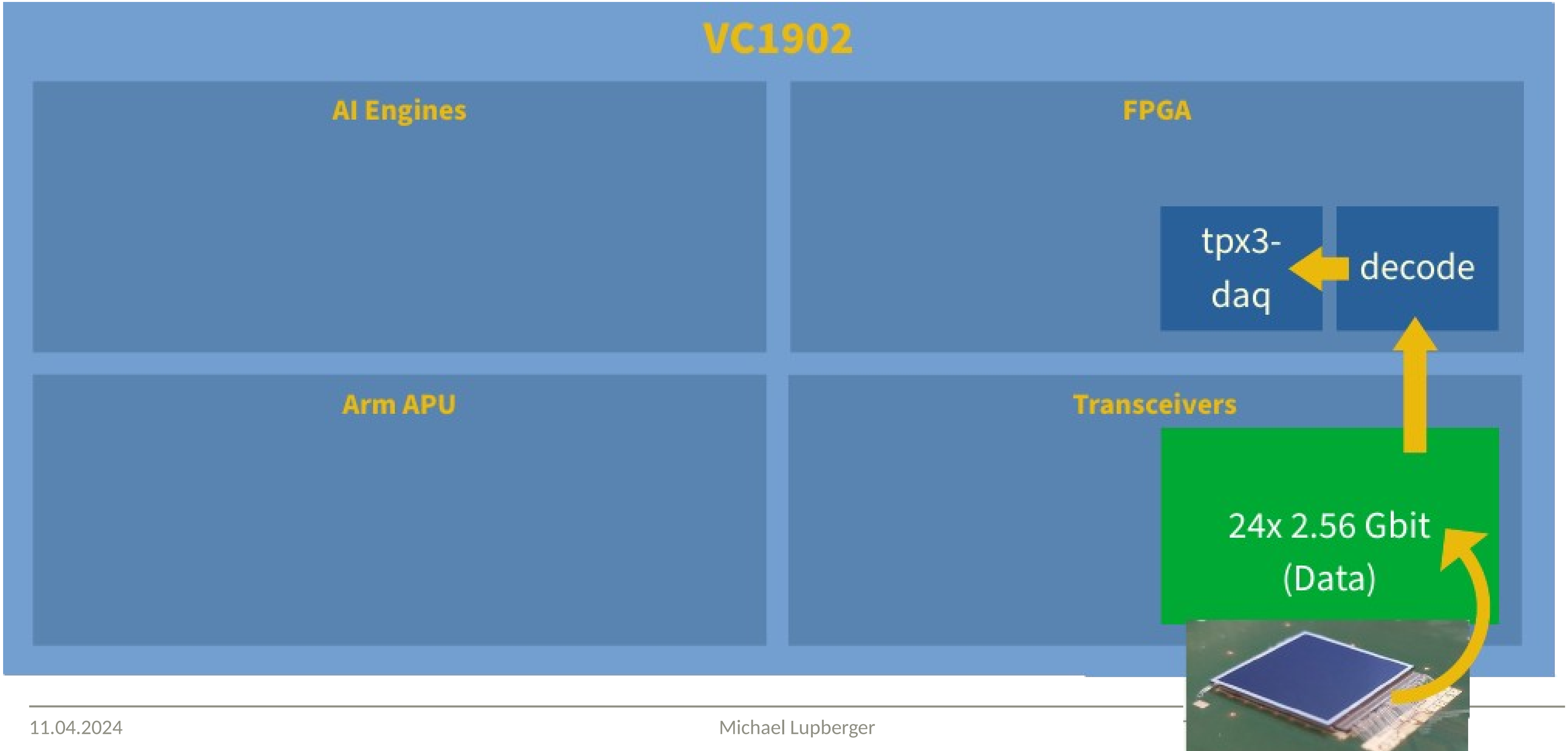
e.g. enlighten the dark sector at the intensity frontier

Setting up the system

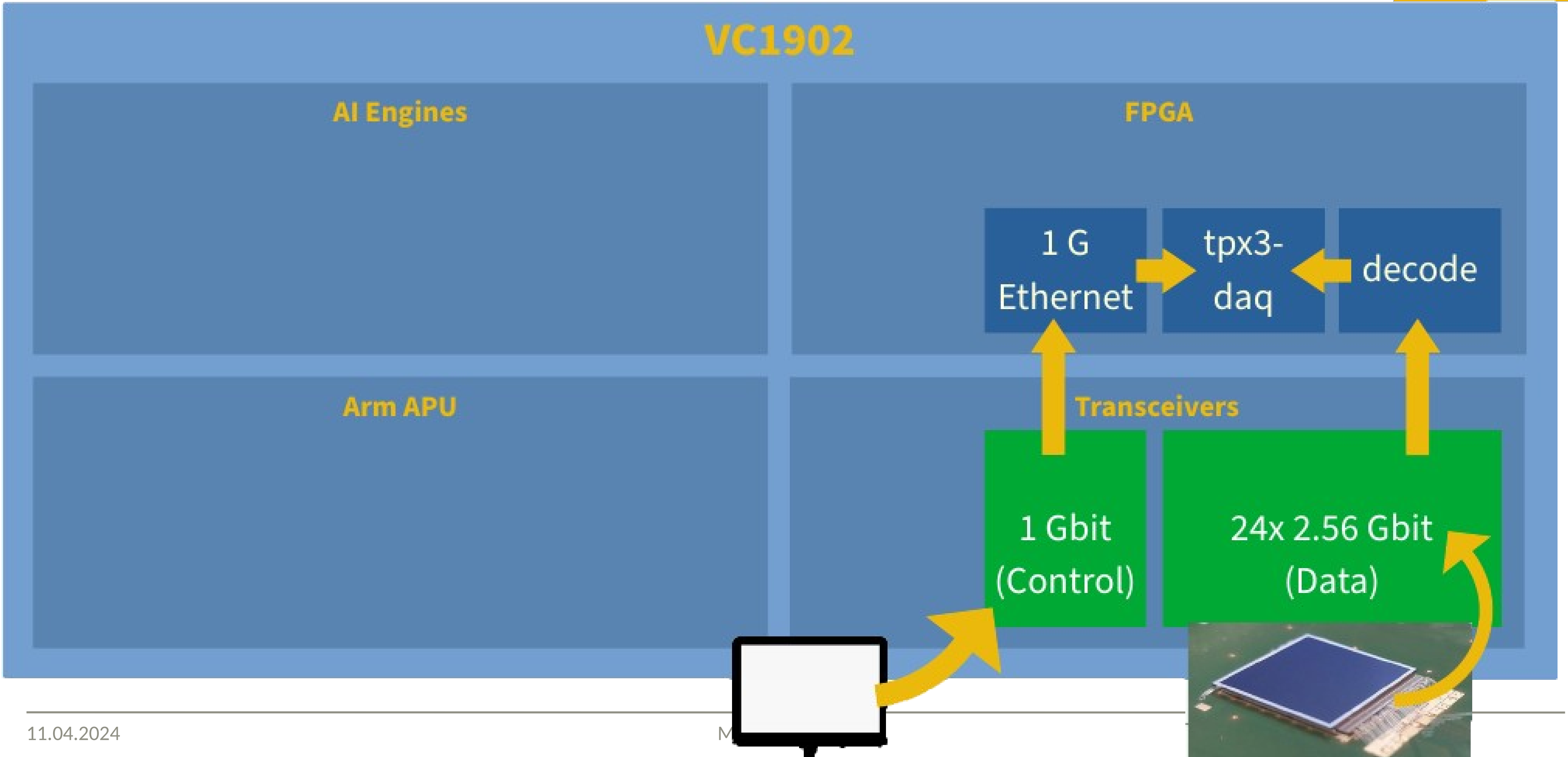


The How

Setting up the system

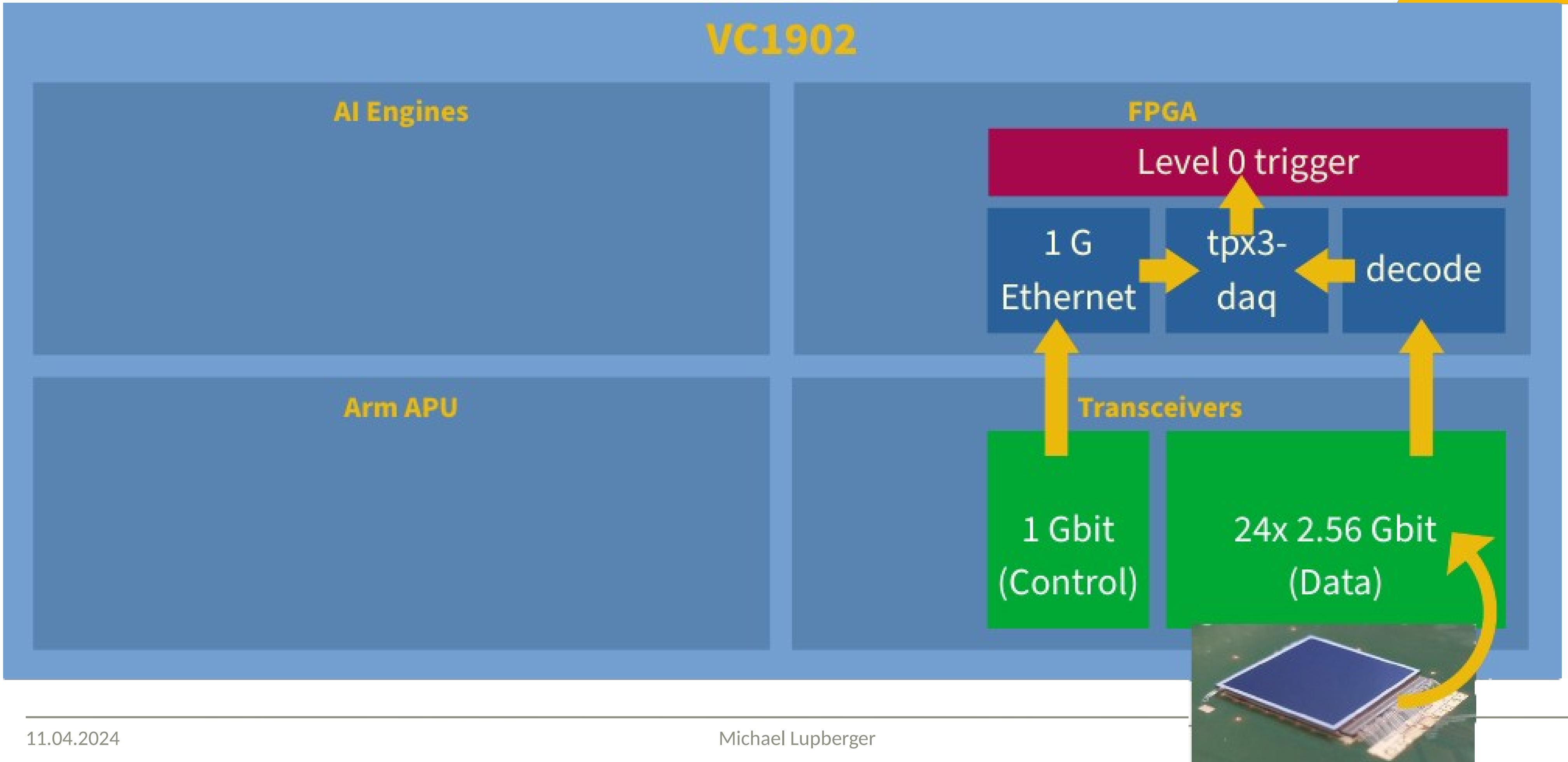


Setting up the system



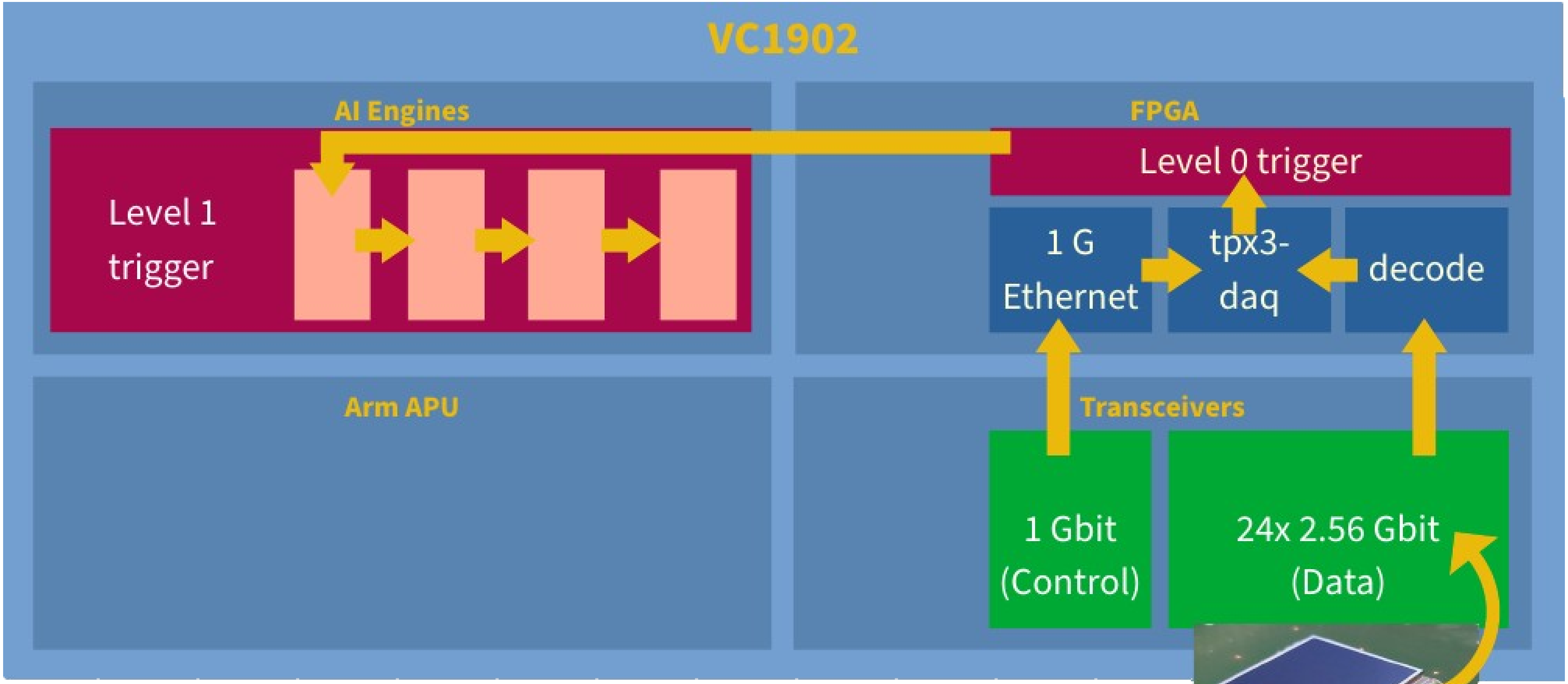
The How

Setting up the system

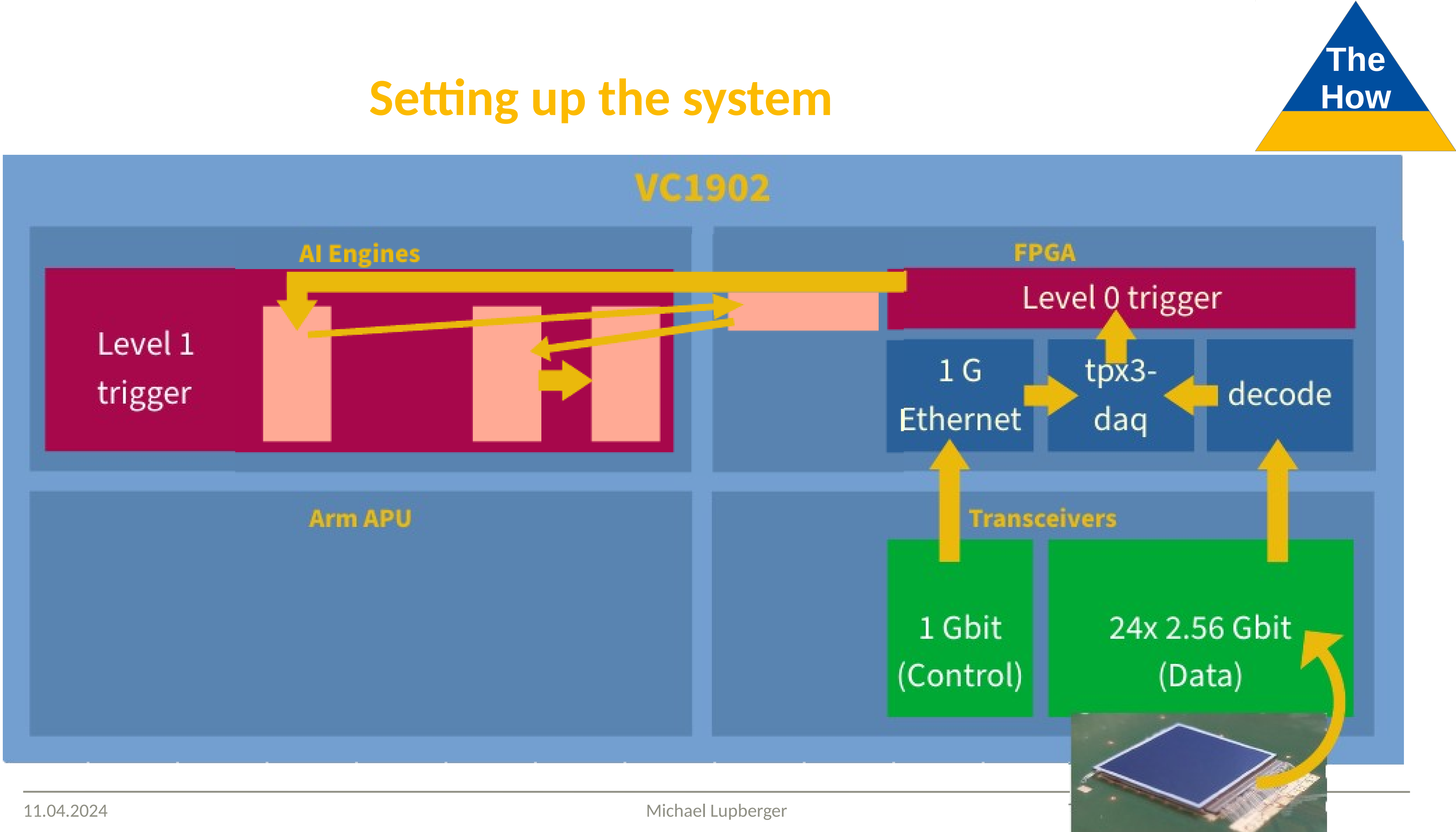


The How

Setting up the system



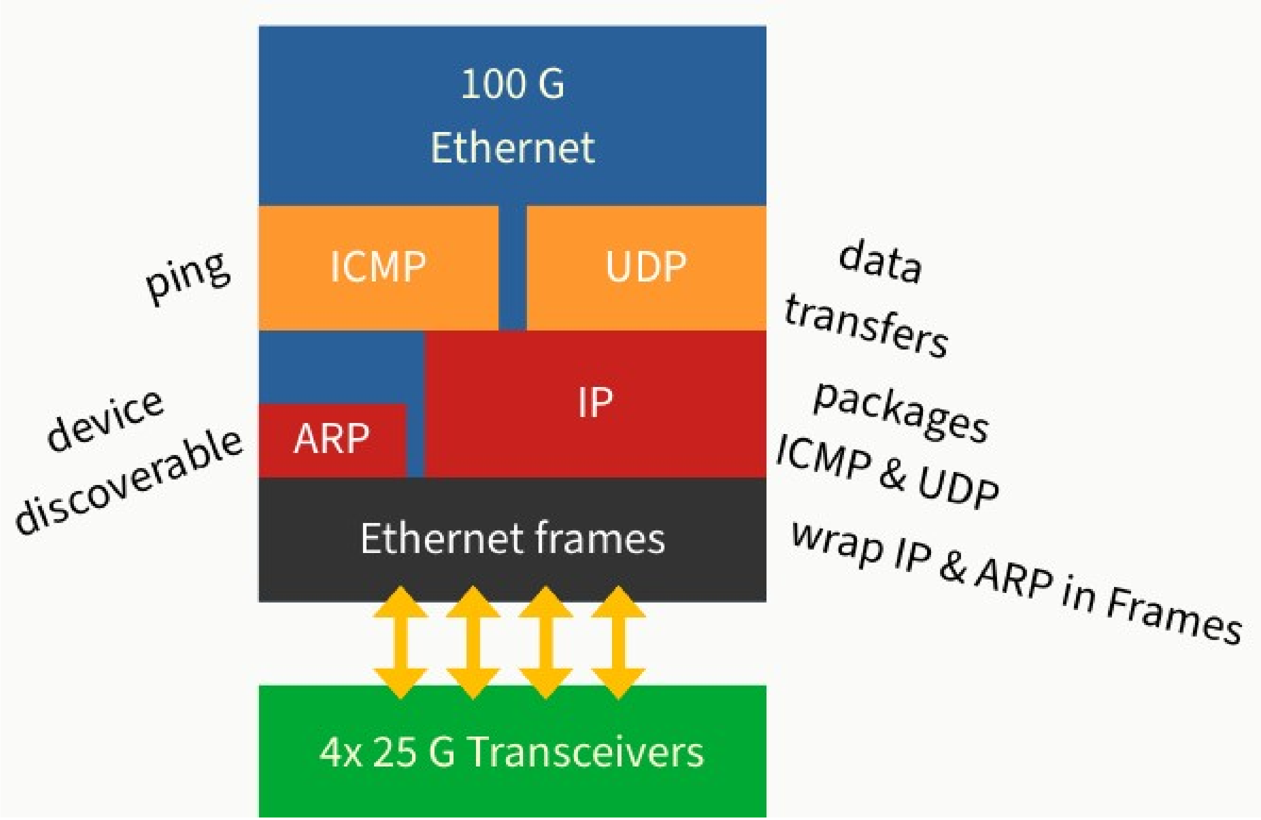
Setting up the system



100 Gbps Ethernet stack

The How

Needed to develop own stack controllable by Linux on Arm RPU



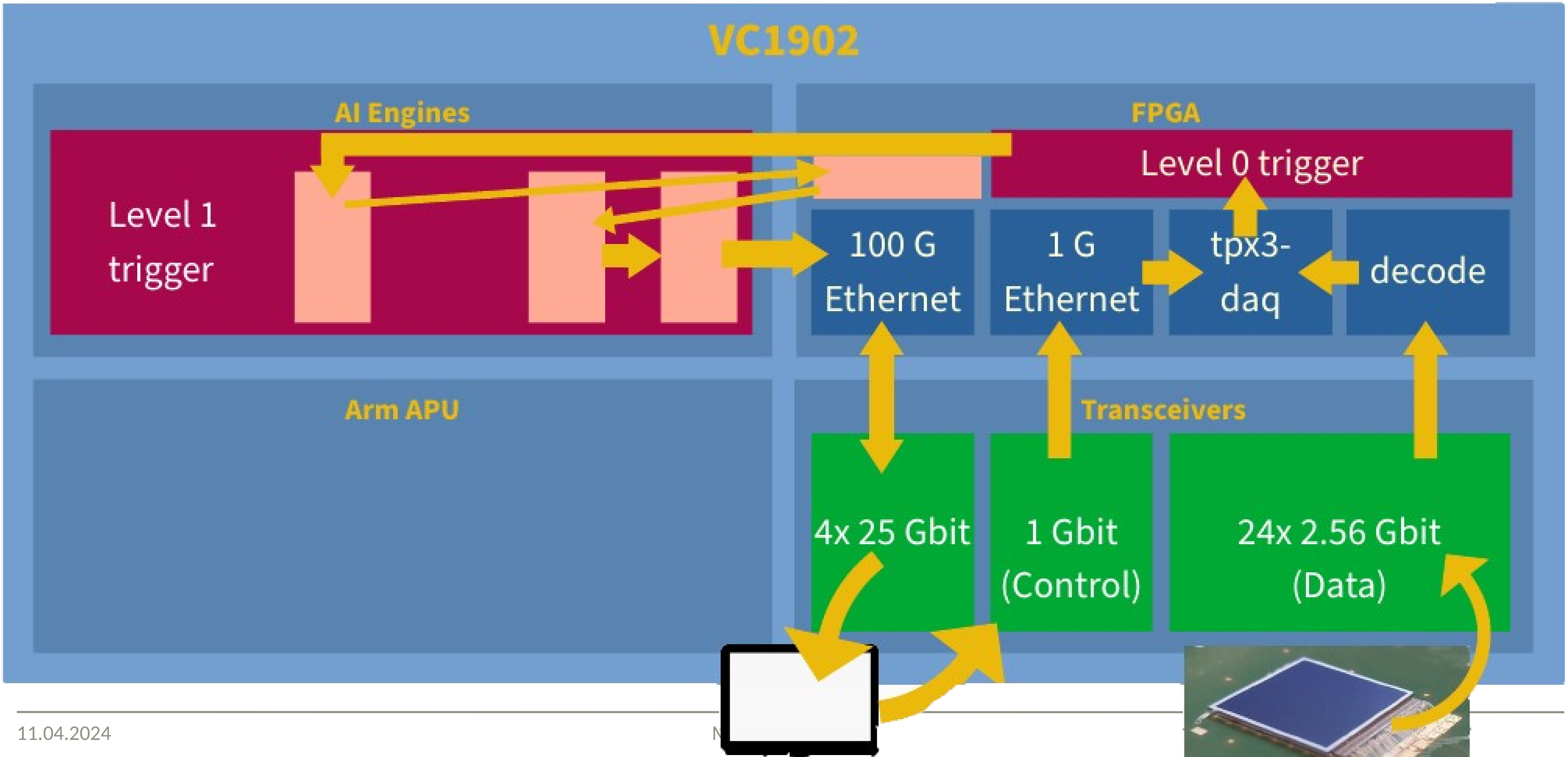
Issue with existing core:

Hard MRMAC core can be connected to CPU by AXI but driver supports only 25 GB/s

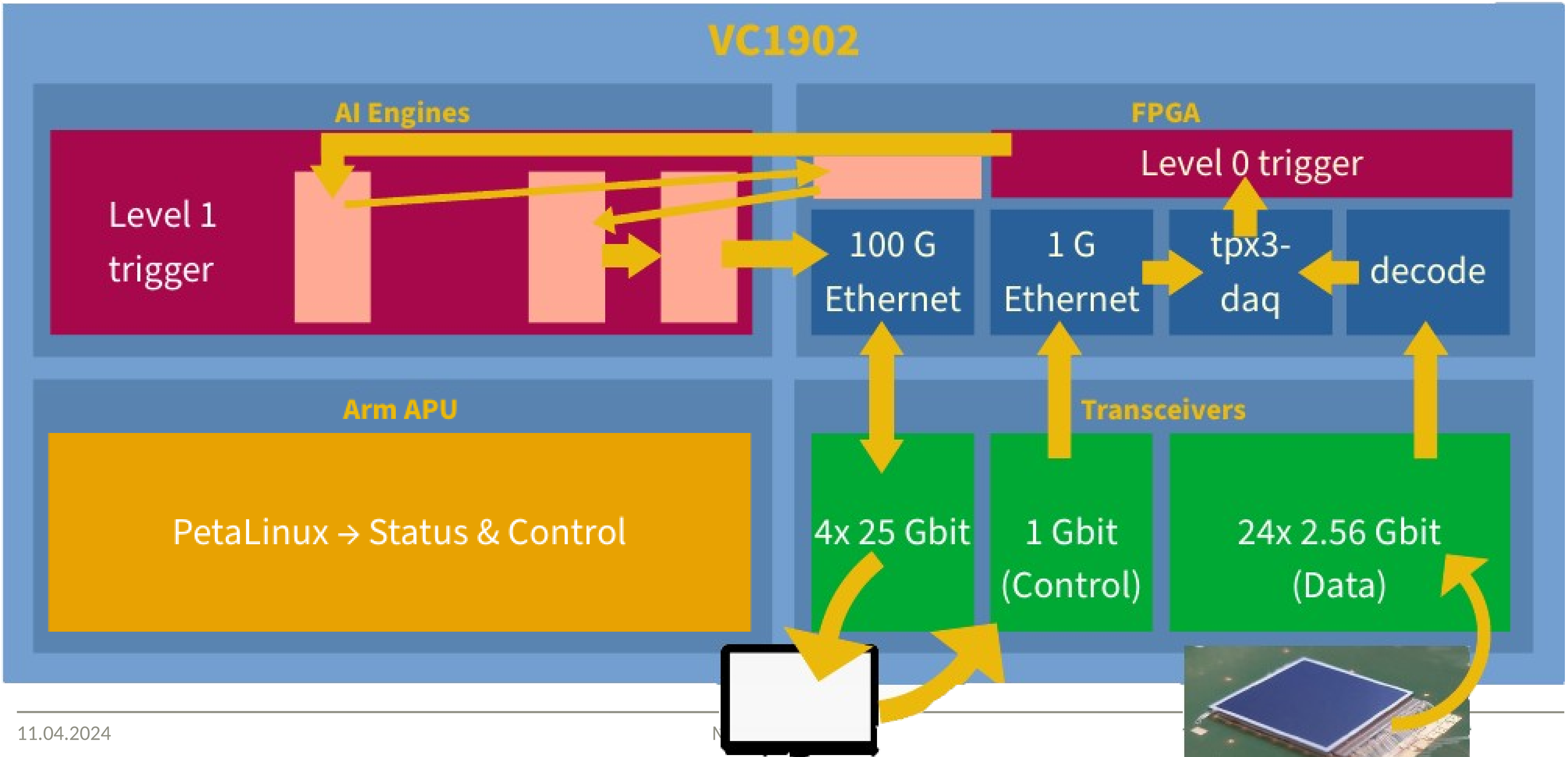
No really working 100 GB/s stack existing

=> develop own

Setting up the system



Setting up the system



ANALYSIS

The How

Fundamental physics objective: find LDM/best exclusion limits

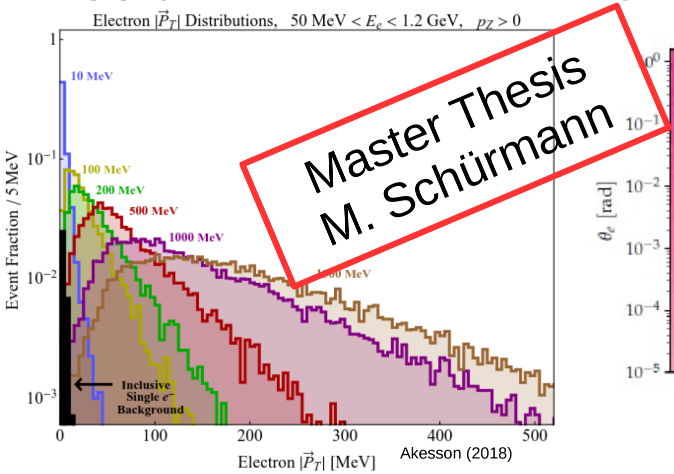
- Prepare data analysis through simulation / Monte-Carlo studies / framework
- Full involvement within collaboration, focus on electron kinematics
- Apply ML-based tracking in full precision in offline reconstruction → electron $\theta, \vec{p}_T$

ANALYSIS

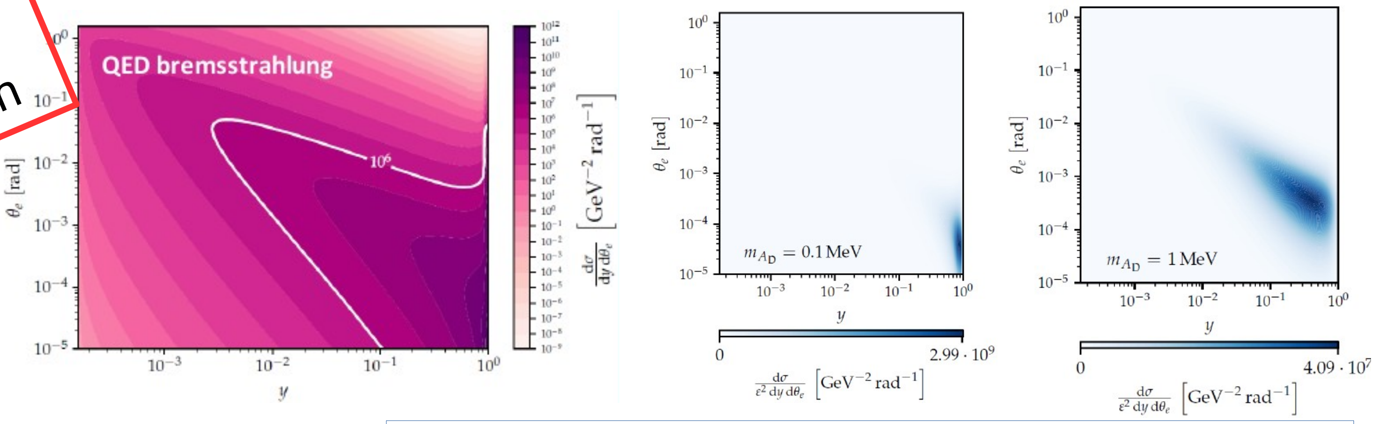
The
How

Fundamental physics objective: find LDM/best exclusion limits

- Prepare data analysis through simulation / Monte-Carlo studies / framework
- Full involvement within collaboration, focus on electron kinematics
- Apply ML-based tracking in full precision in offline reconstruction → electron $\theta, \vec{p}_T$



Akesson (2018)



Master Thesis Martin Schürmann (Theory/Bonn)

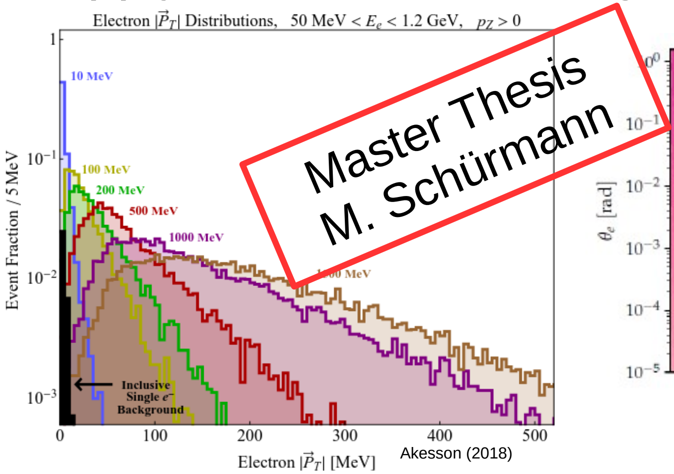
y: ratio recoil electron energy to incident electron energy, y = E'/E
θ_e: scattering angle of electron with respect to incident beam axis

ANALYSIS

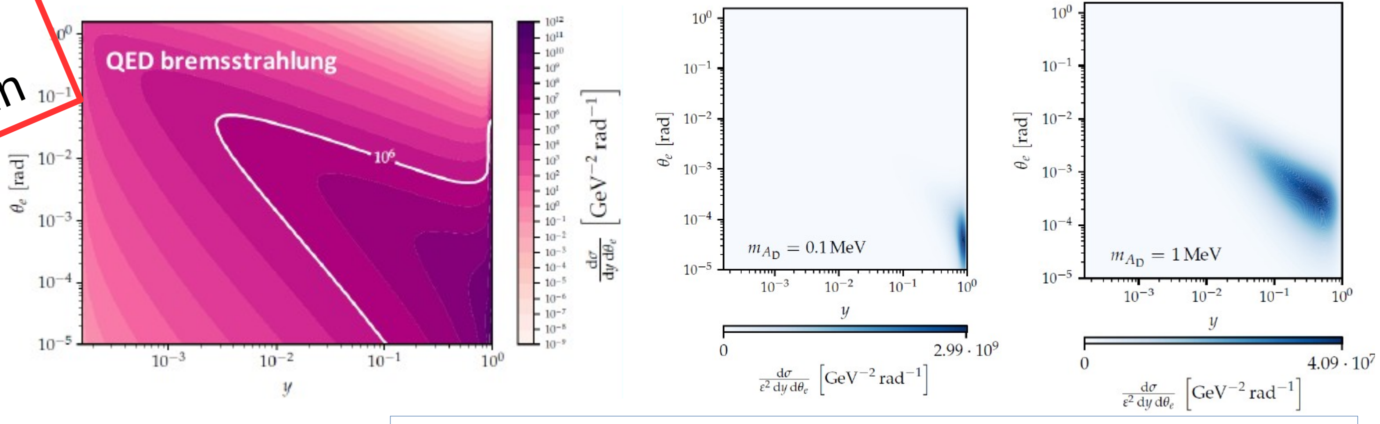
The How

Fundamental physics objective: find LDM/best exclusion limits

- Prepare data analysis through simulation / Monte-Carlo studies / framework
- Full involvement within collaboration, focus on electron kinematics
- Apply ML-based tracking in full precision in offline reconstruction ➞ electron $\theta, \vec{p}_T$



Akesson (2018)



Master Thesis Martin Schürmann (Theory/Bonn)

y: ratio recoil electron energy to incident electron energy, y = E'/E
θ_e: scattering angle of electron with respect to incident beam axis

⇒ Key: Capture SM Backgrounds

⇒ Detailed detector, data acquisition & trigger studies