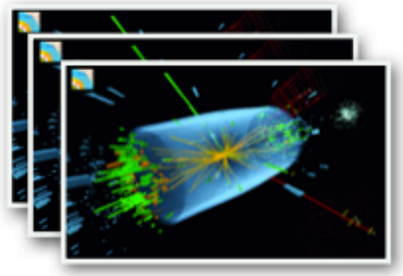
Real time AI

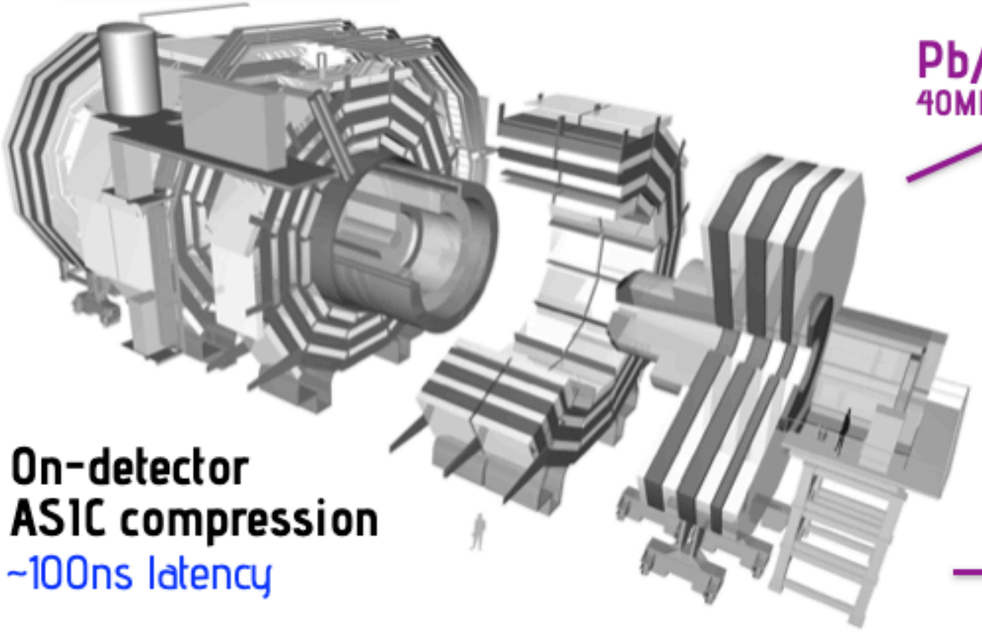
Ben Hawks, Sergo Jindariani, Jovan Mitrevski, Nhan Tran (Fermilab) for the hls4ml team (fastmachinelearning.org)



CMS Experiment

40MHz collision rate

~1B detector channels

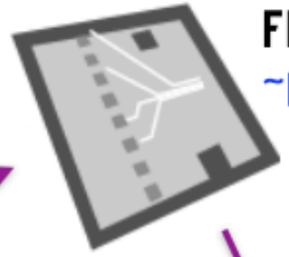


On-detector ASIC compression

~100ns latency

Pb/s

40MHz

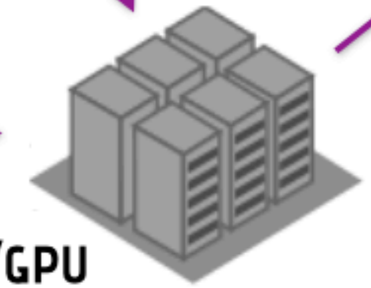


FPGA filter stack

~µs latency

10s Tb/s

100s kHz

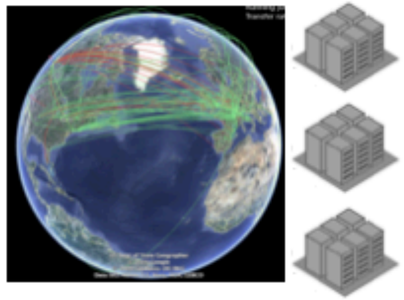


On-prem CPU/GPU filter farm

~100 ms latency

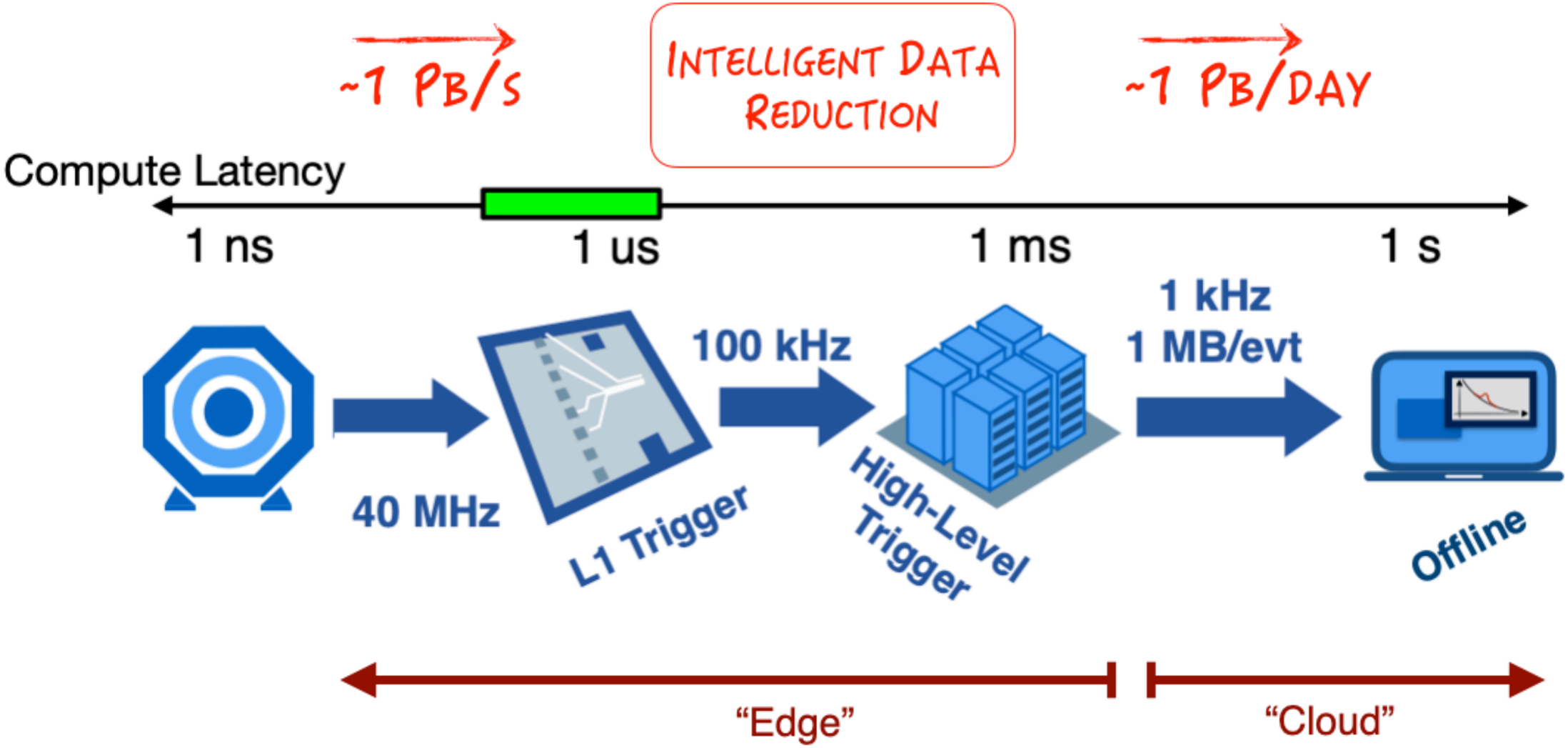
10s Gb/s

~5 kHz



Worldwide computing grid

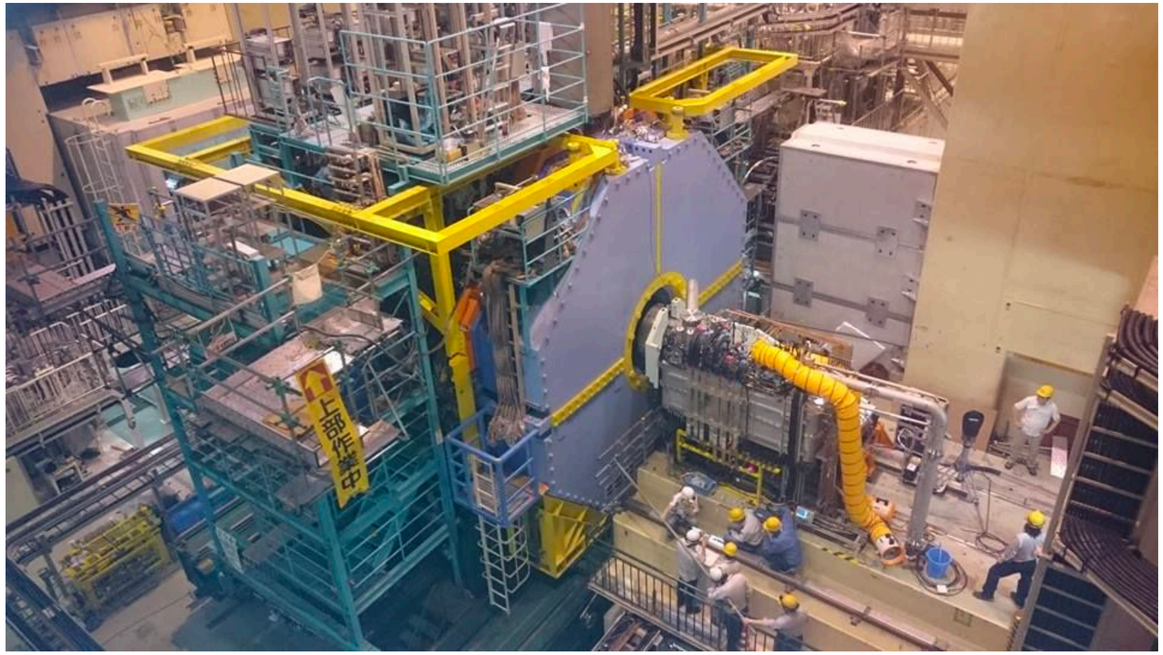
Exabyte-scale datasets


~1 PB/s
Intelligent Data Reduction
~1 PB/day
Compute Latency
1 ns
1 us
1 ms
1 s
40 MHz
L1 Trigger
100 kHz
High-Level Trigger
1 kHz
1 MB/evt
Offline
"Edge"
"Cloud"

More real-time AI

NA62, CERN

Belle-II, Japan

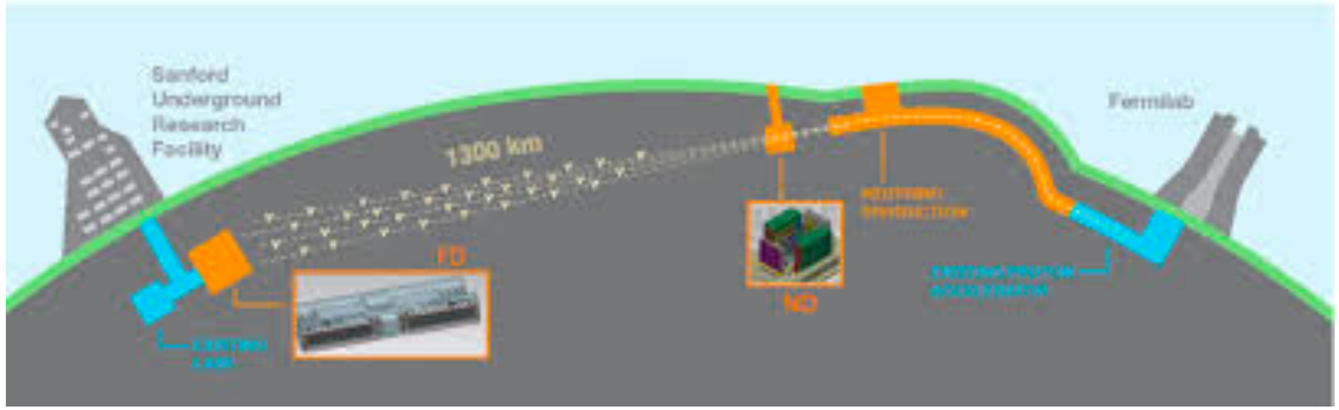


More real-time AI

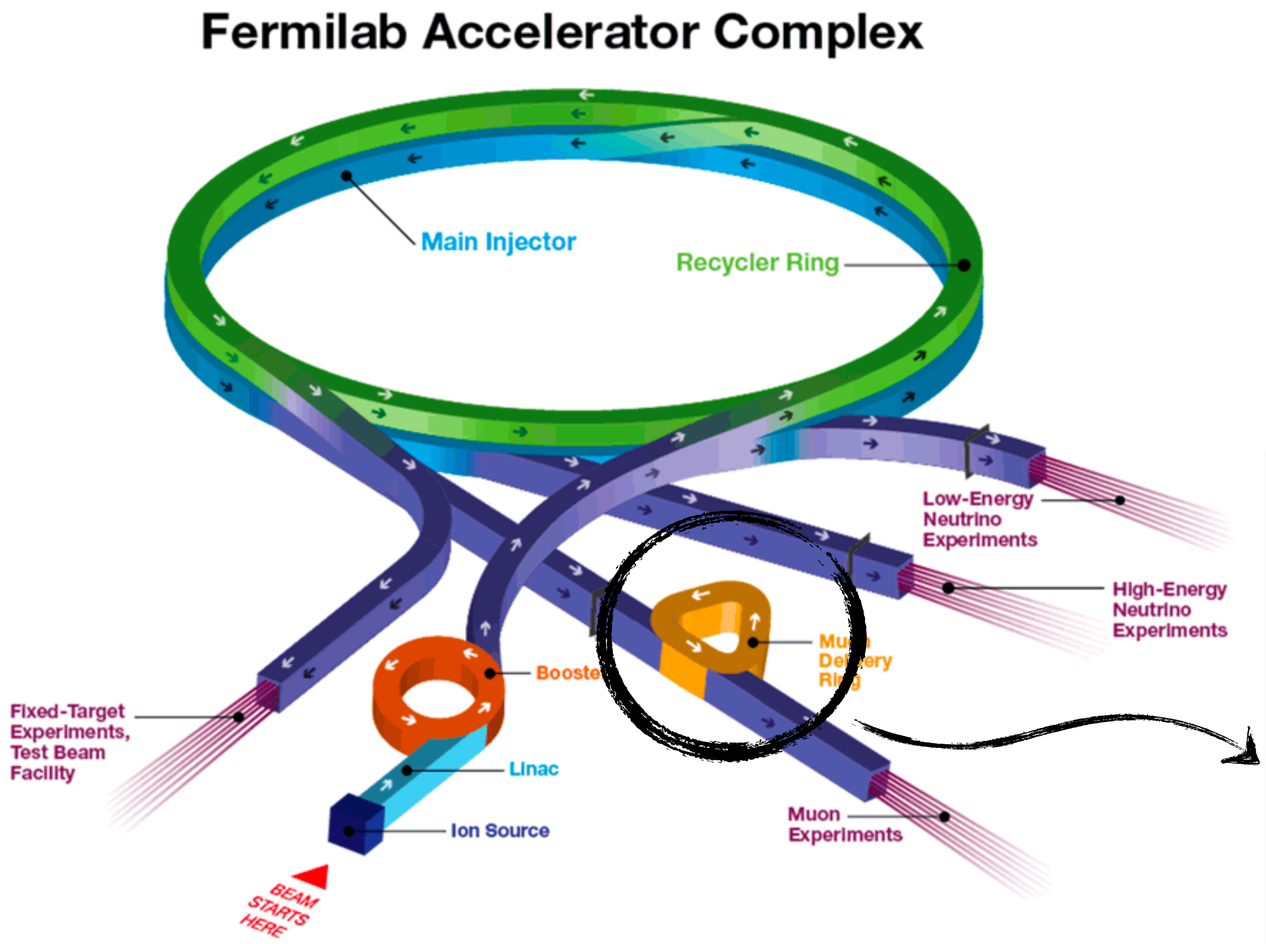
5. Researchers expect DUNE's data system to catch about 10 neutrinos per day—but must be able to catch thousands in seconds if a star goes supernova nearby.

A supernova could produce thousands of neutrino events within seconds.

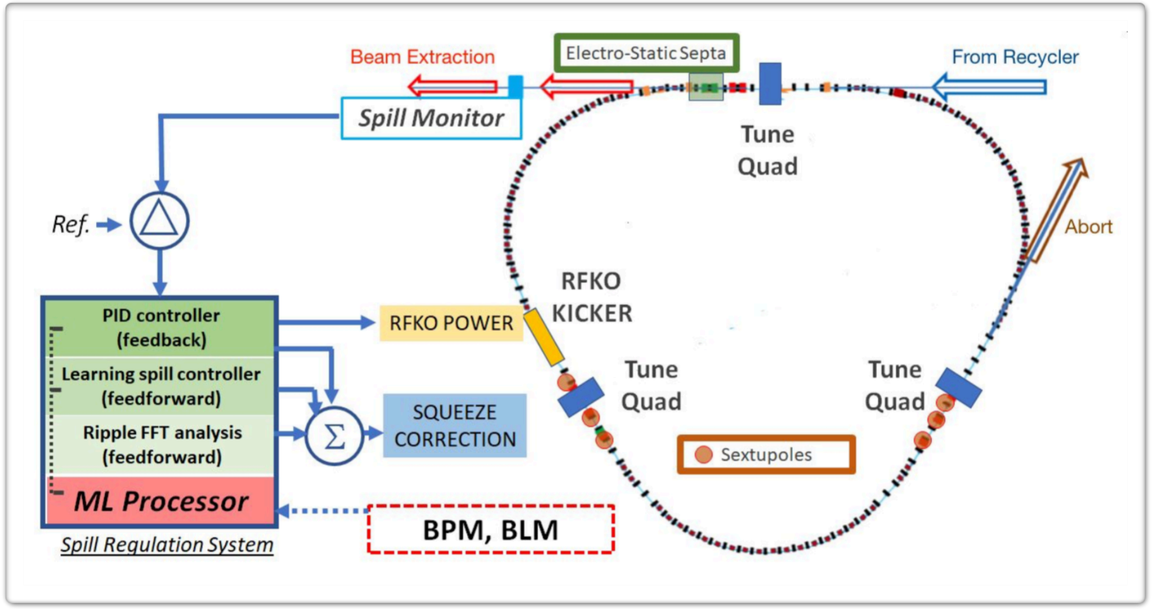
Real-time filtering at the sub-millisecond scale required for next generation neutrino detectors to identify neutrinos from supernovae, P ~ $(25\ \text{years})^{-1}$



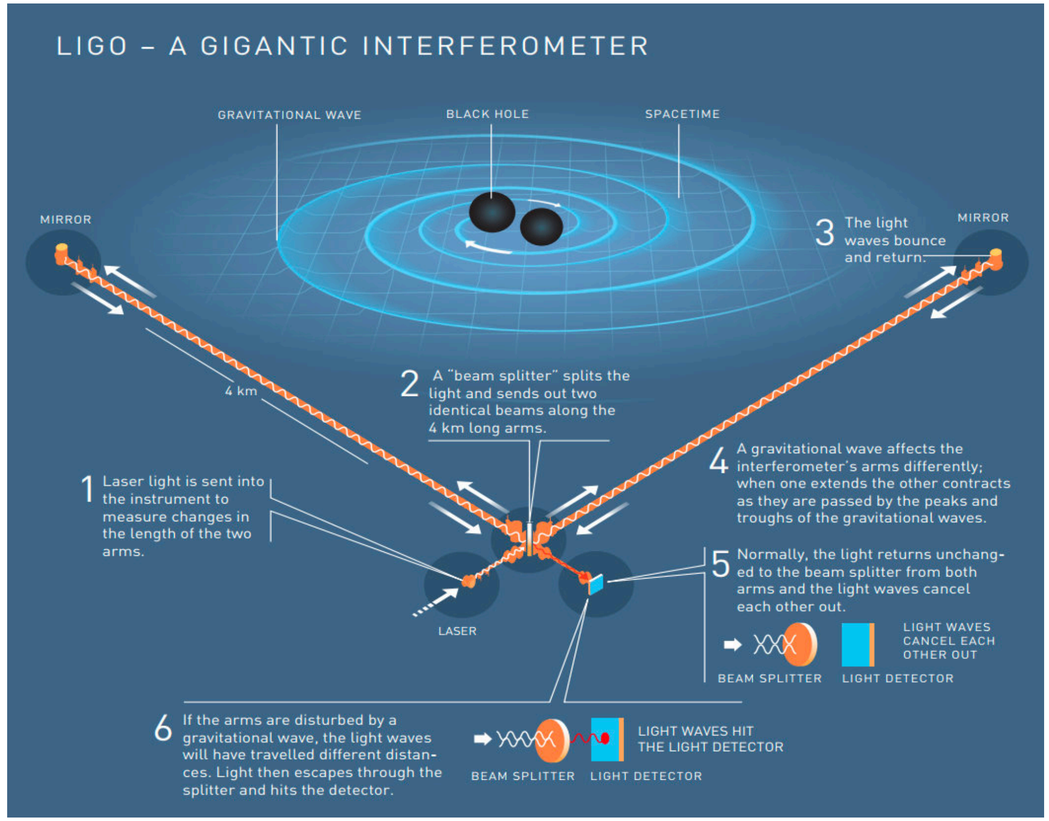
More real-time AI

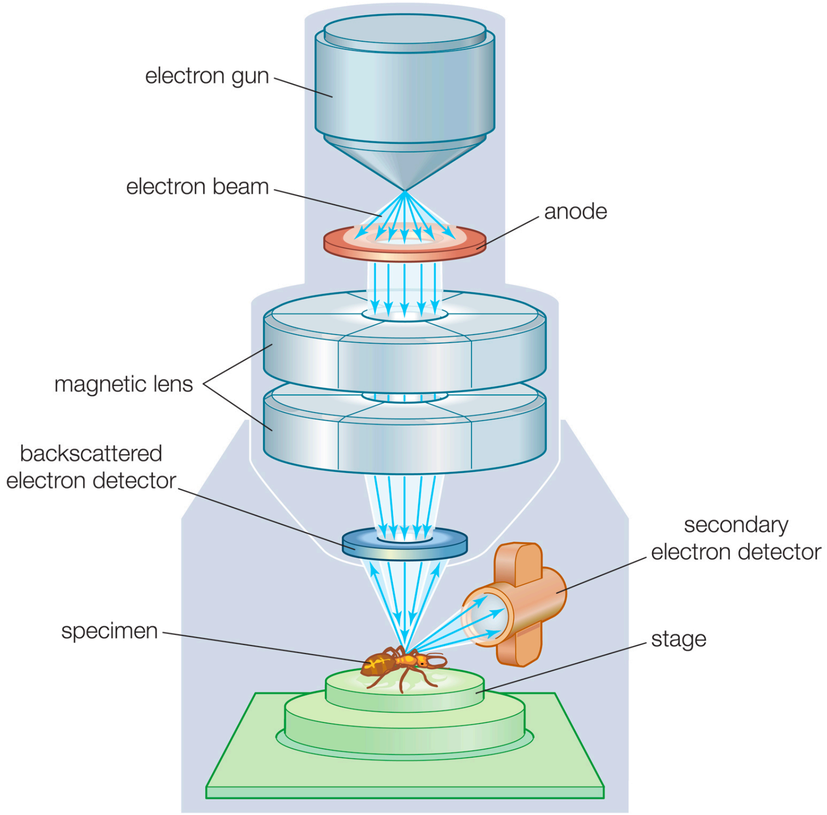


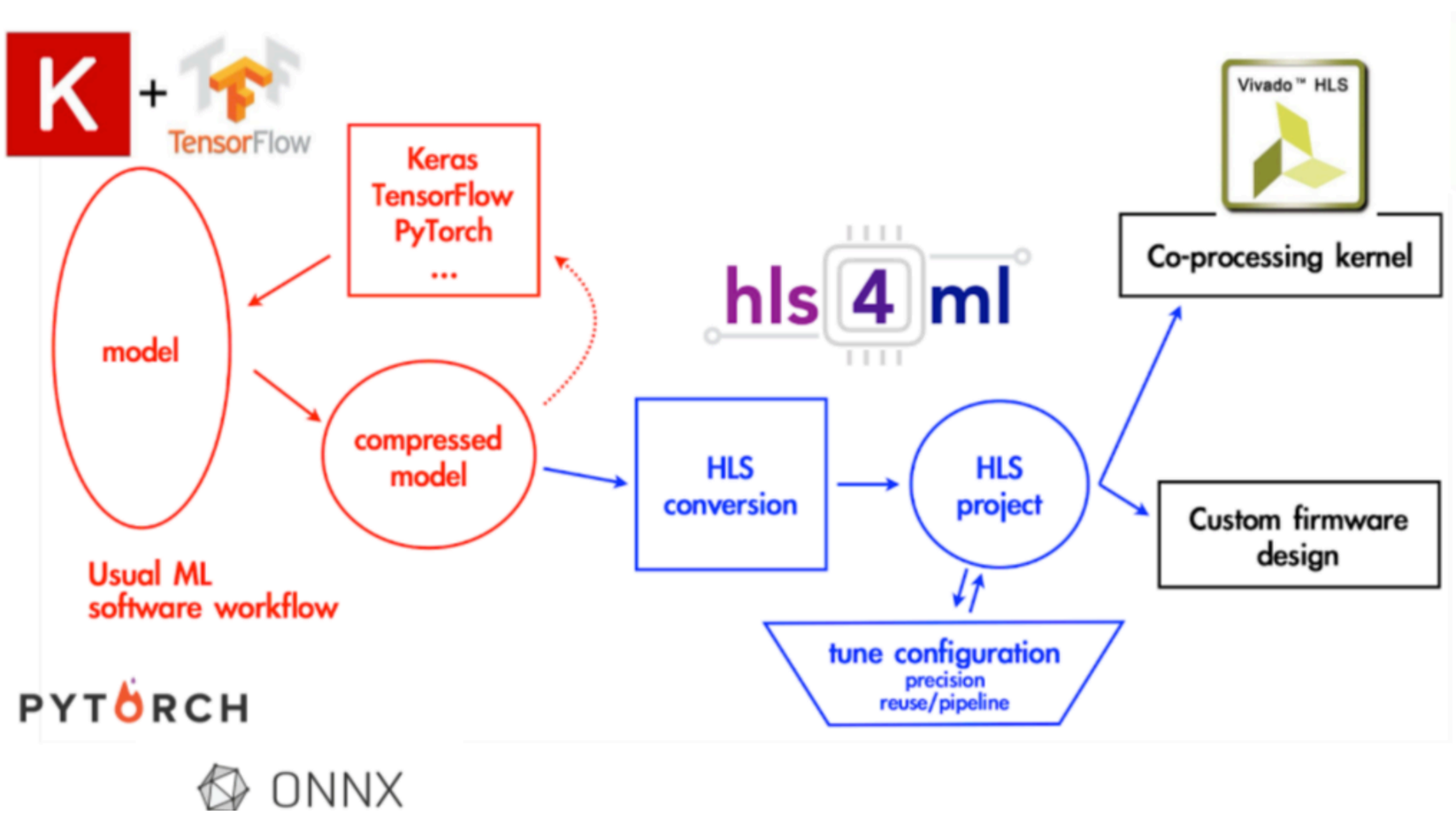
Reinforcement learning for millisecond-scale magnet current tuning required



And more…


LIGO – A GIGANTIC INTERFEROMETER
GRAVITATIONAL WAVE
BLACK HOLE
SPACETIME
MIRROR
MIRROR
3 The light waves bounce and return.
4 km
2 A "beam splitter" splits the light and sends out two identical beams along the 4 km long arms.
1 Laser light is sent into the instrument to measure changes in the length of the two arms.
4 A gravitational wave affects the interferometer's arms differently; when one extends the other contracts as they are passed by the peaks and troughs of the gravitational waves.
5 Normally, the light returns unchanged to the beam splitter from both arms and the light waves cancel each other out.
LASER
LIGHT WAVES CANCEL EACH OTHER OUT
BEAM SPLITTER
LIGHT DETECTOR
6 If the arms are disturbed by a gravitational wave, the light waves will have travelled different distances. Light then escapes through the splitter and hits the detector.
LIGHT WAVES HIT THE LIGHT DETECTOR
BEAM SPLITTER
LIGHT DETECTOR


electron gun
electron beam
anode
magnetic lens
backscattered electron detector
secondary electron detector
specimen
stage



Mentor
Catapult
Coming Soon

Today's hls4ml hands on

- Part 1:
  - Get started with hls4ml: train a basic model and run the conversion, simulation & c-synthesis steps
- Part 2:
  - Learn how to tune inference performance with quantization & ReuseFactor
- Part 3:
  - Perform model compression and observe its effect on the FPGA resources/latency
- Part 4:
  - Train using QKeras "quantization aware training" and study impact on FPGA metrics