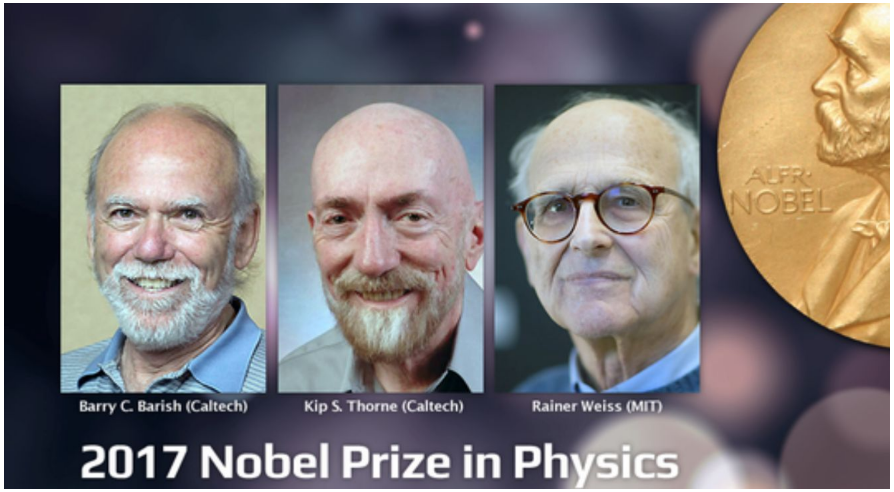
LIGO Tour

UW REU
July 9 2019

Laser Interferometer Gravitational-Wave Observatory



LIGO observatories in the Contiguous United States

LIGO

Vacuum tube

LIGO Hanford's 'arms'

LIGO Design

(V. GRAVITATION RESEARCH)

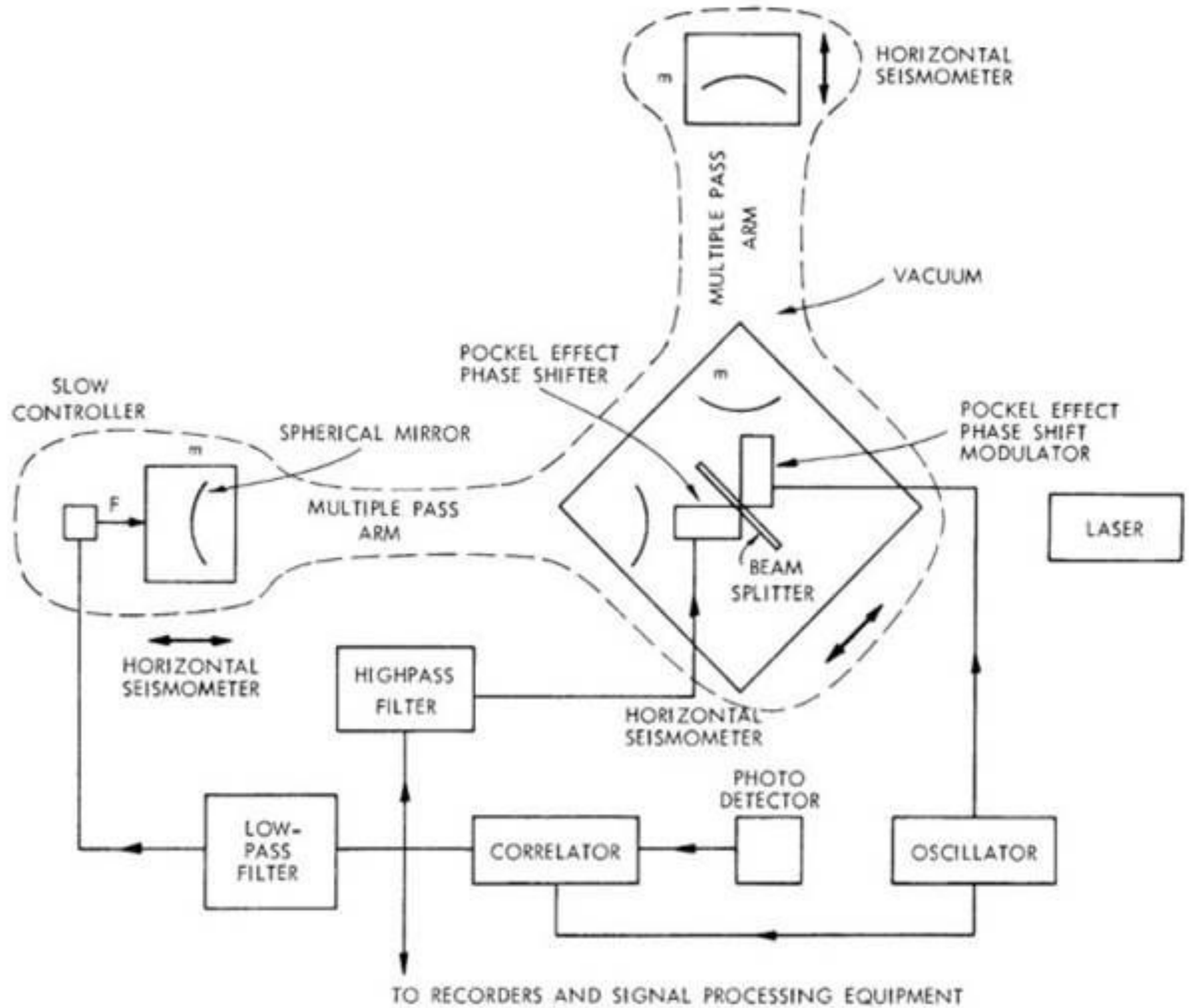


Fig. V-20. Proposed antenna.

Travel Plan

06:45am Gather at dorm

07:00am Departure

09:00am Short break, Safeway
400 N Ruby St, Ellensburg, WA 98926
Pickup food on the way to LIGO

10:45am arrive LIGO Hanford Observatory
127124 N Route 10, Richland, WA 99354

11:00am LIGO tour begin
Access Laser and Vacuum Equipment Area (LVEA)

2:00pm LIGO tour ends, return to Seattle

3:00pm Short break, Ginkgo Petrified Forest State Park
630 Ginkgo Ave, Vantage, WA 98950

5:30pm Dinner, The Attic at Salish Lodge (TBC)
6501 Railroad Ave #102, Snoqualmie, WA 98065
8:30pm Arrive dorm

Student drivers:
Gabriel Moreau,
Carlos Sevilla, Emilee Wurtz

Get your Breakfast using your meal card *TONIGHT*

We will cover lunch and dinner to budget announced tomorrow.

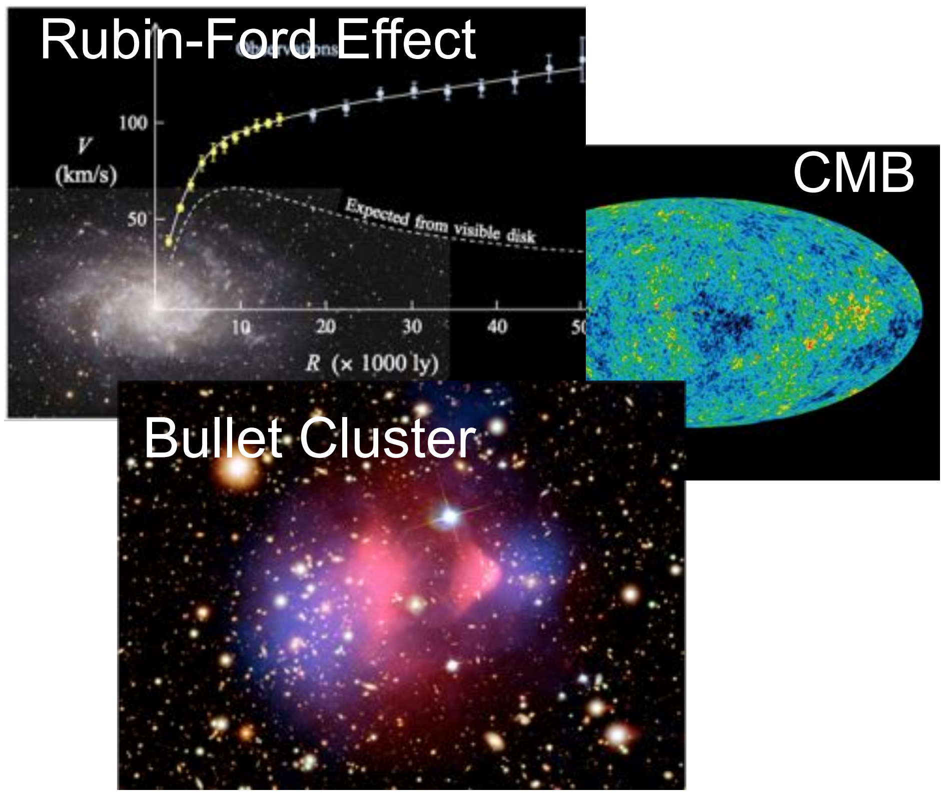
Accelerating Machine Learning for Dark Matter Search

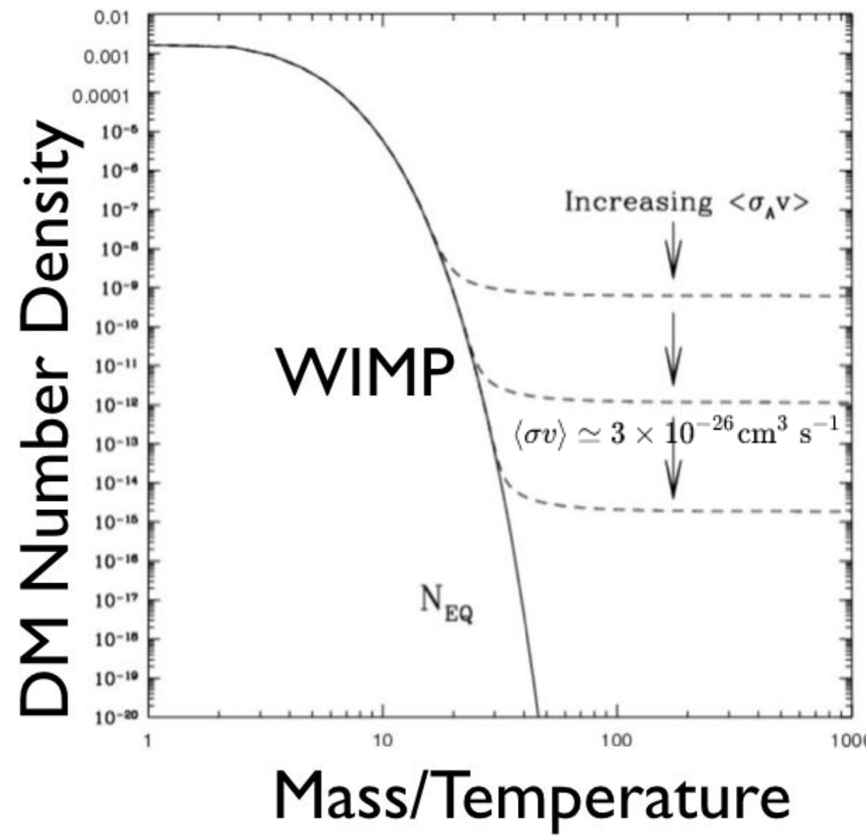
Shih-Chieh Hsu
University of Washington Seattle

UW REU
PAT C521, July 8 2019

Dark Matter (WIMPs)

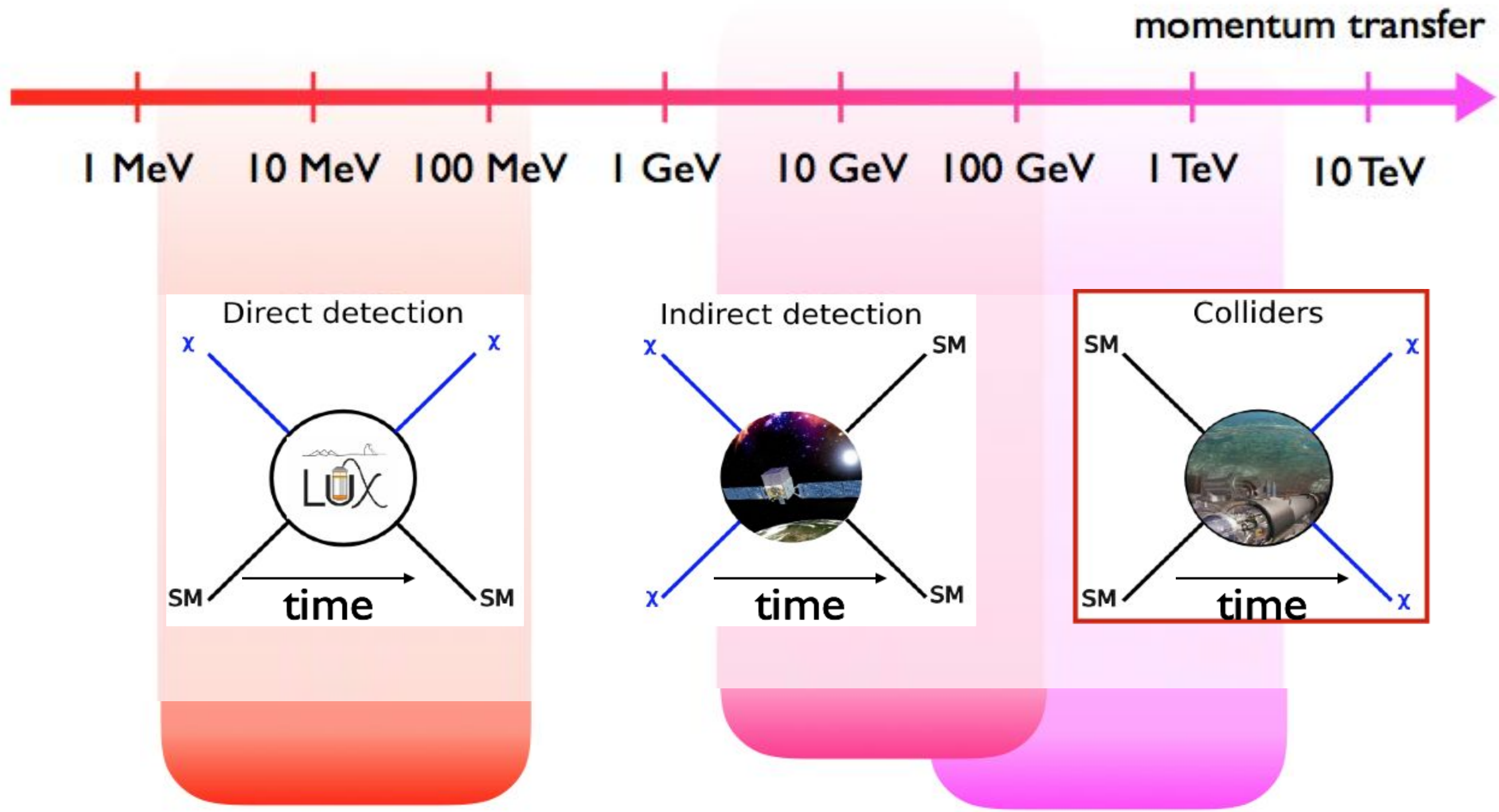
- Particle nature of dark matter strongly suggested by gravitational anomaly observations
- Weakly Interacting Massive Particles (WIMPs) with properties consistent to thermal relics (**WIMPs Miracle!**)





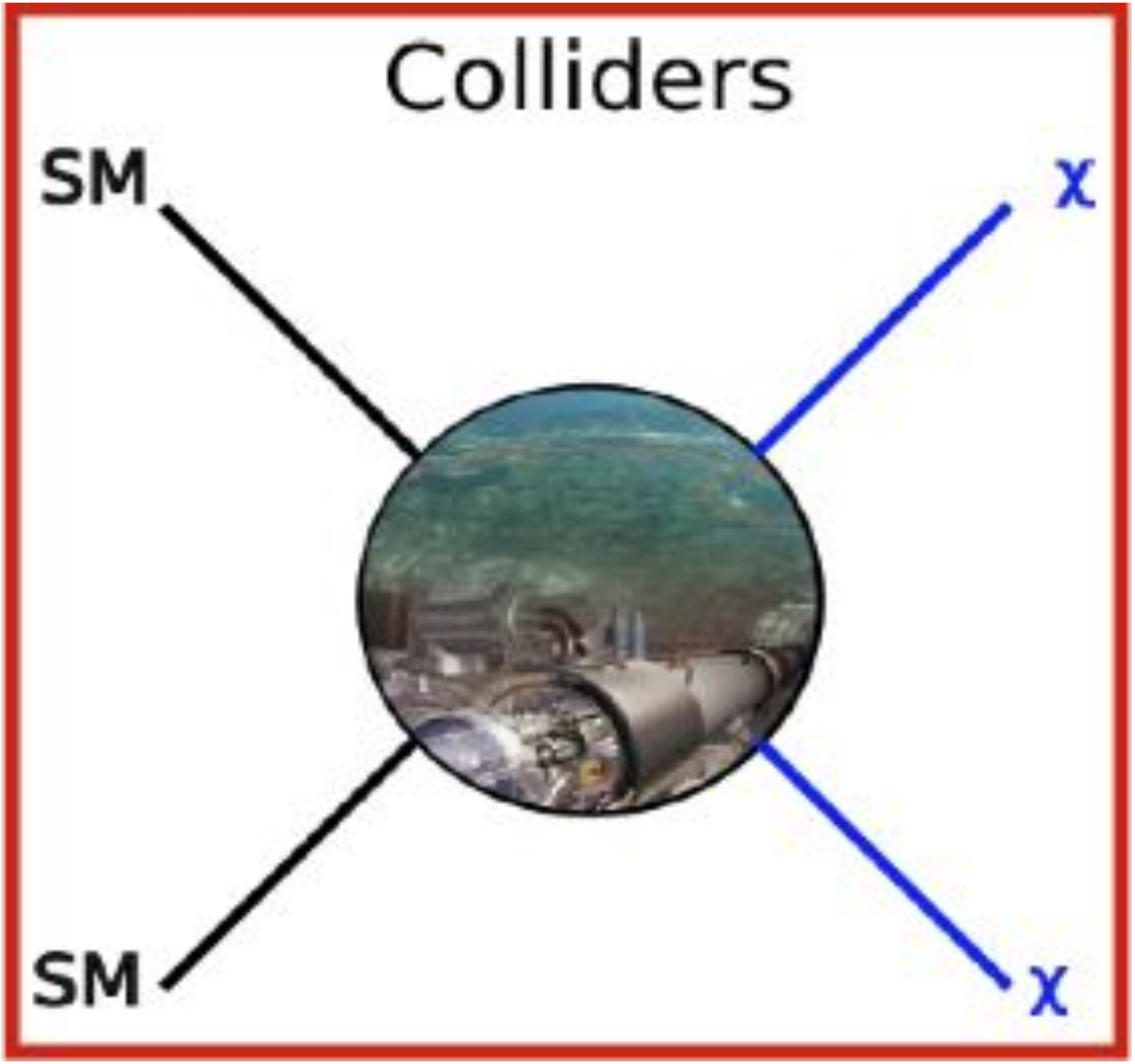
WIMPs Detections

WIMPs may be produced through proton-proton collisions at the **LHC**!



Collider searches are complementary to Direct Detections and Indirect Detections.

How to detect WIMP?



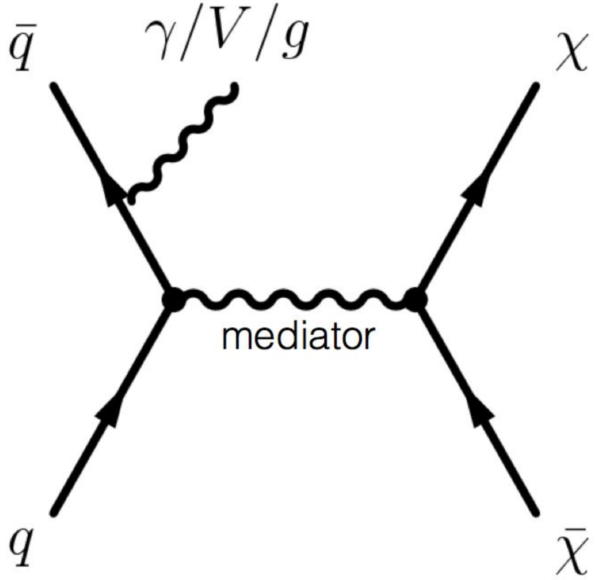
Key Experiment Observables

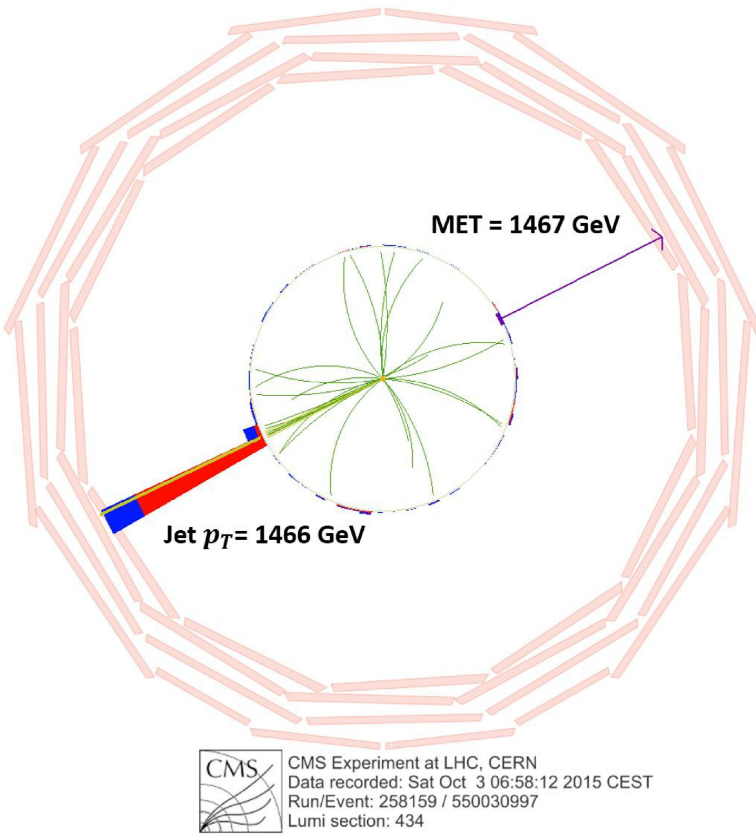
Missing transverse momentum (energy)

- Transverse momentum balanced with initial transverse momenta = 0!

$$\Delta\phi\left(p_T^{miss}, X\right) \sim \pi$$

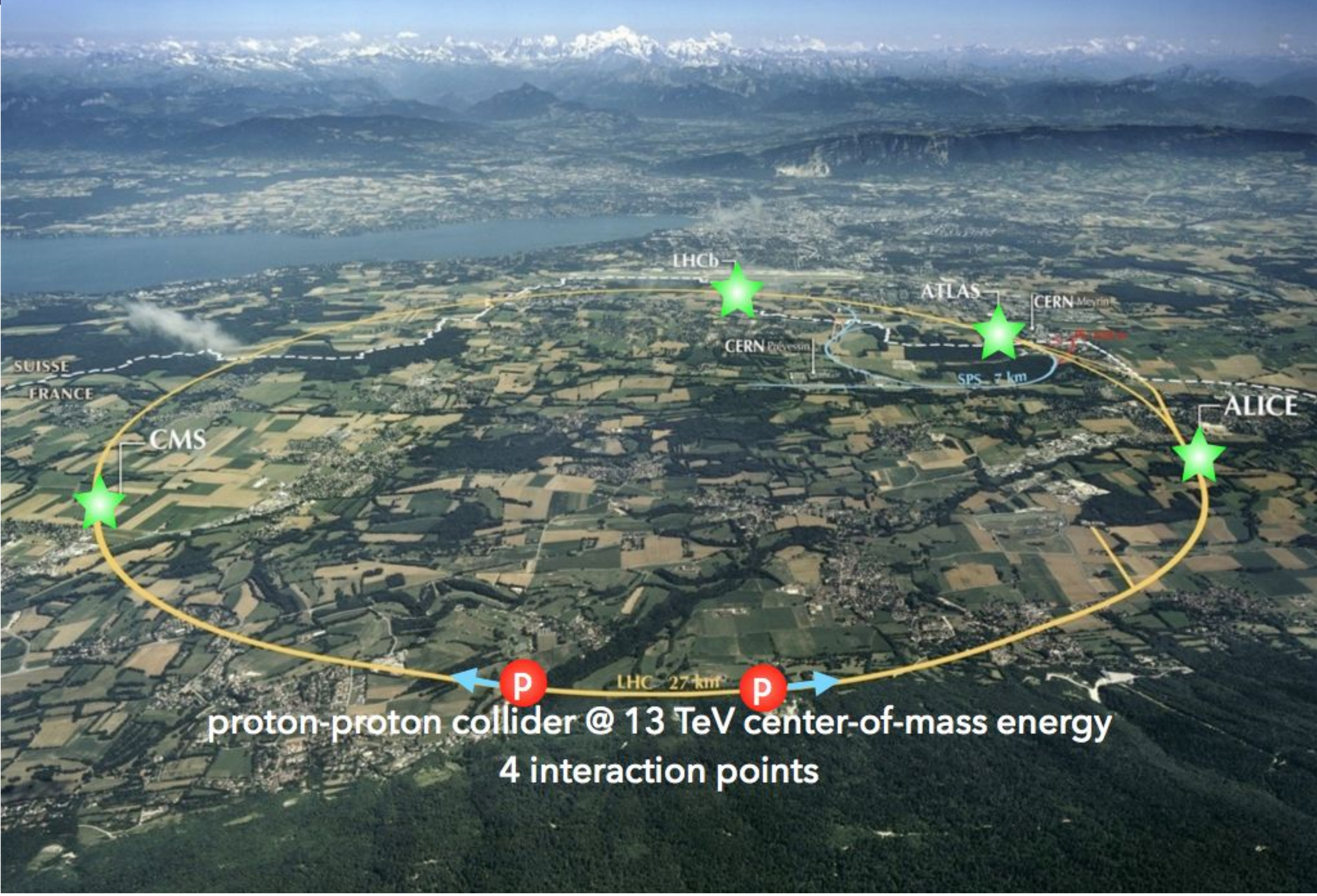
$$p_T^{miss} \sim p_T^X (ET^{miss})$$





THE LARGE HADRON COLLIDER

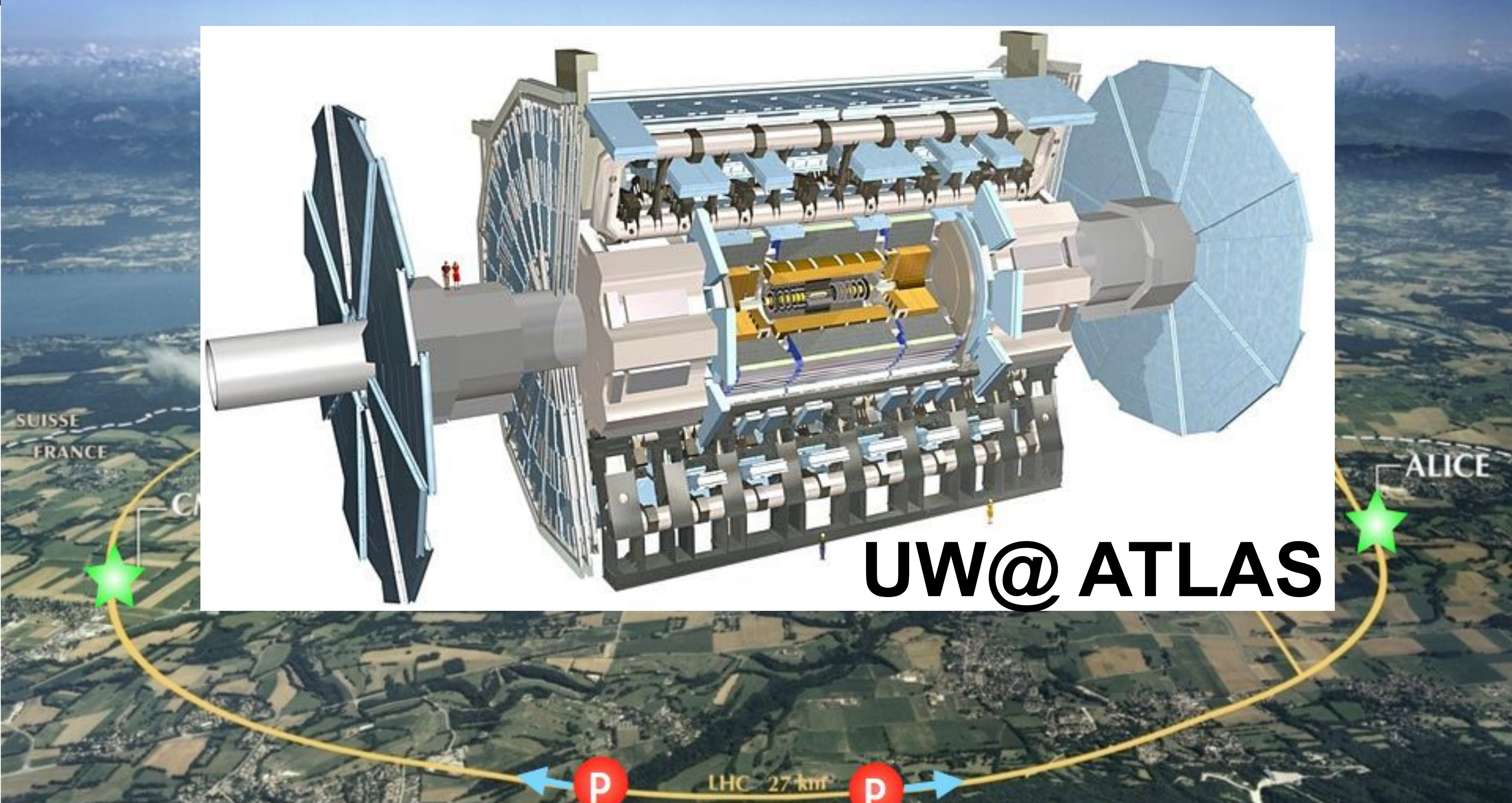
THE LARGE HADRON COLLIDER



proton-proton collider @ 13 TeV center-of-mass energy

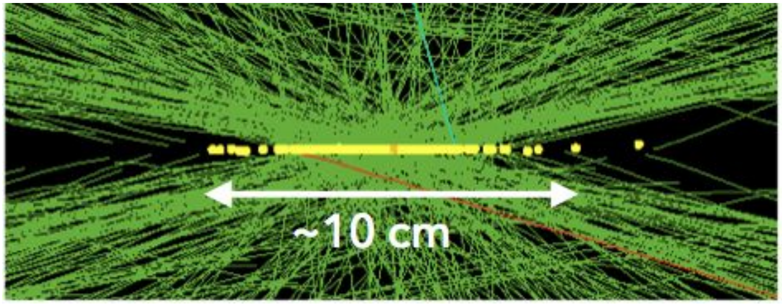
4 interaction points

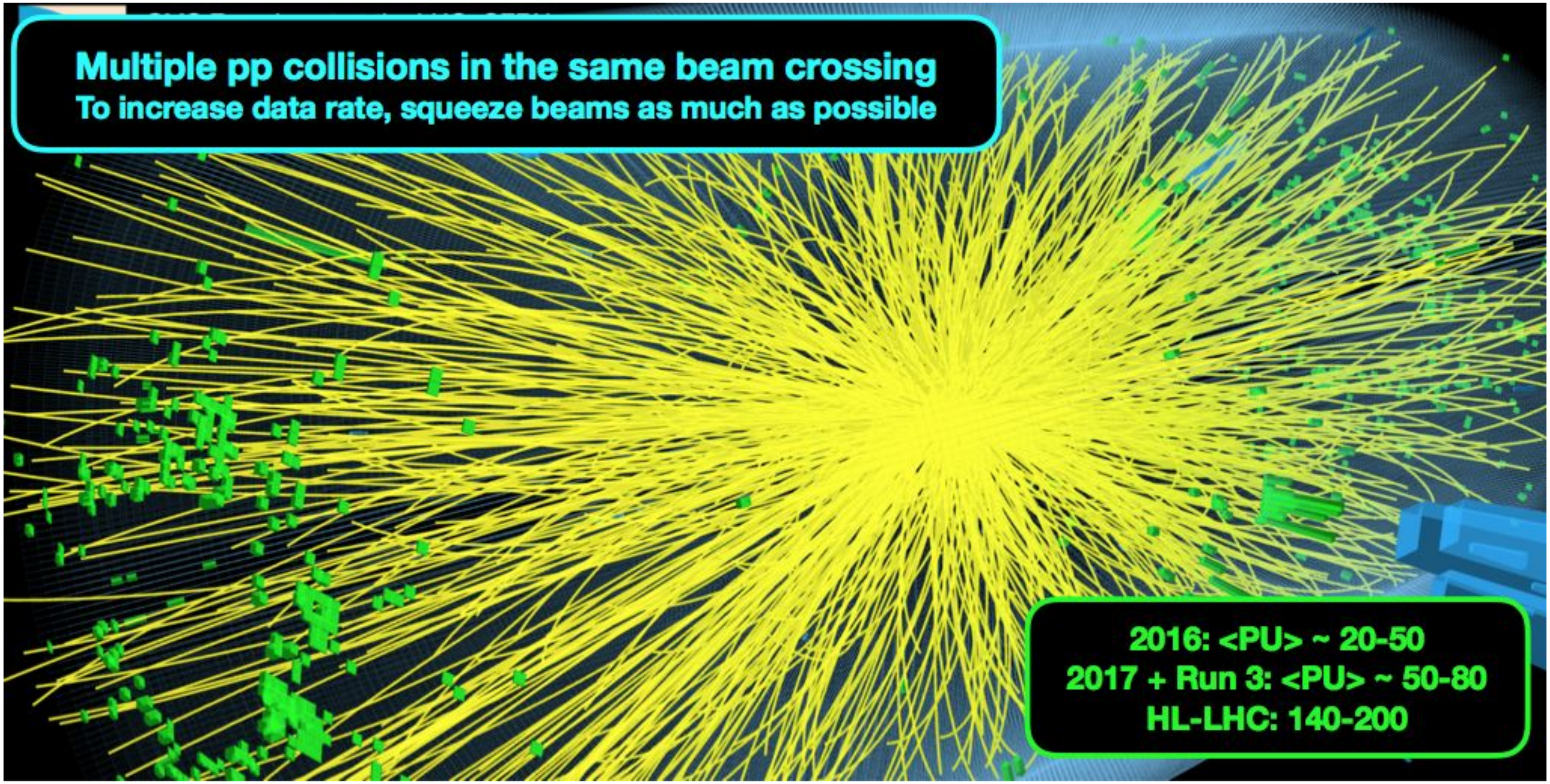
THE LARGE HADRON COLLIDER



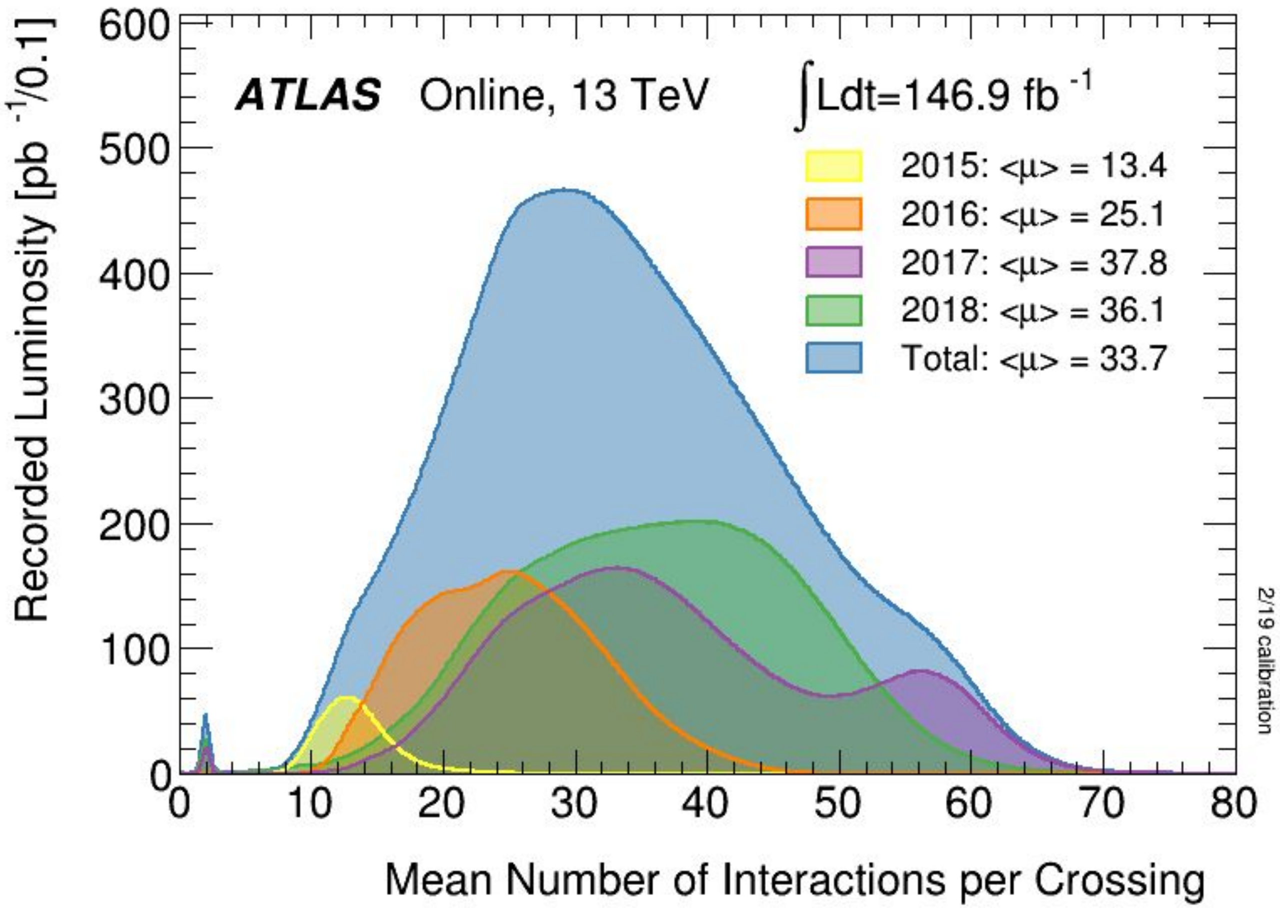
proton-proton collider @ 13 TeV center-of-mass energy
4 interaction points

Pile-up Challenge

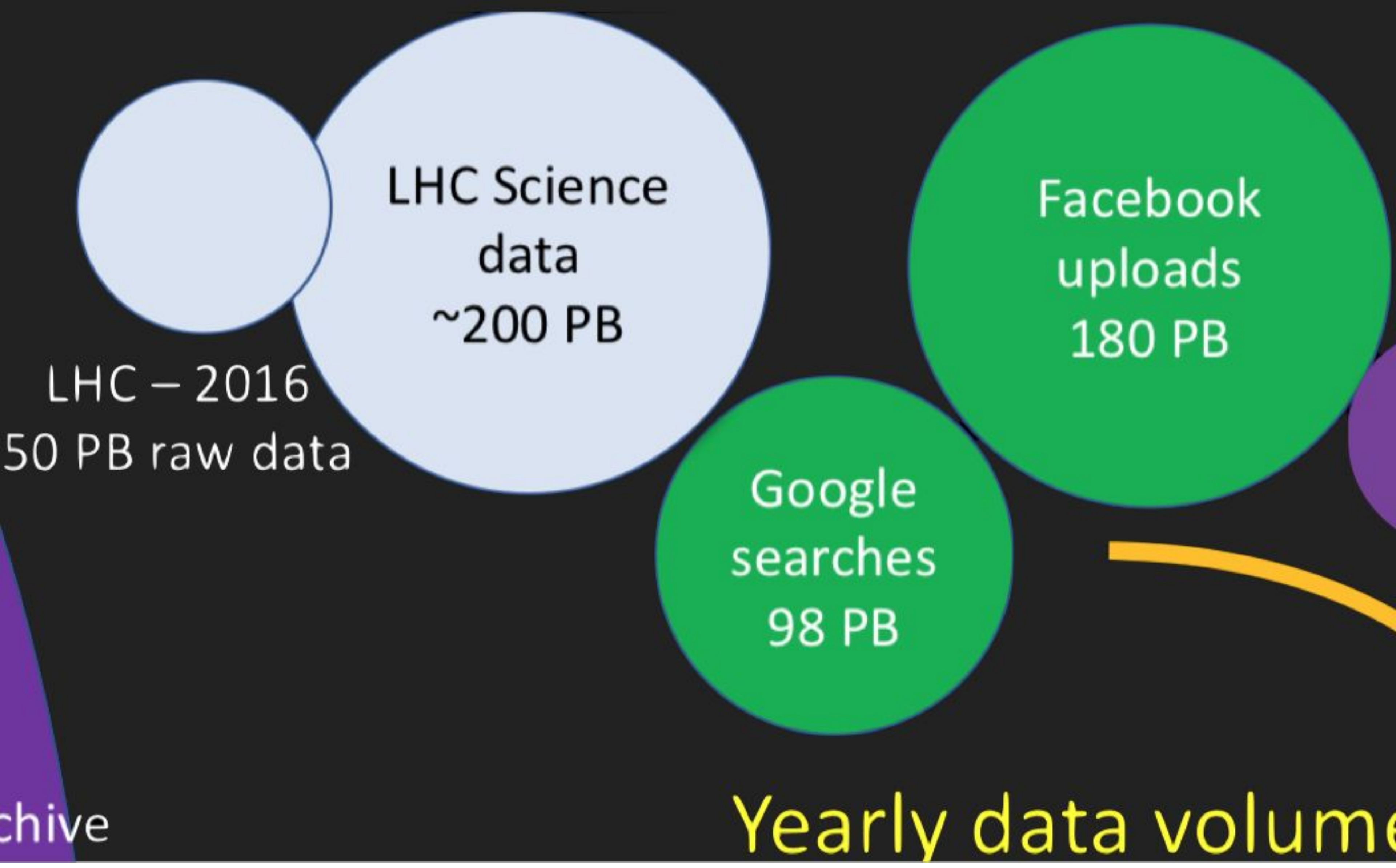




Challenge

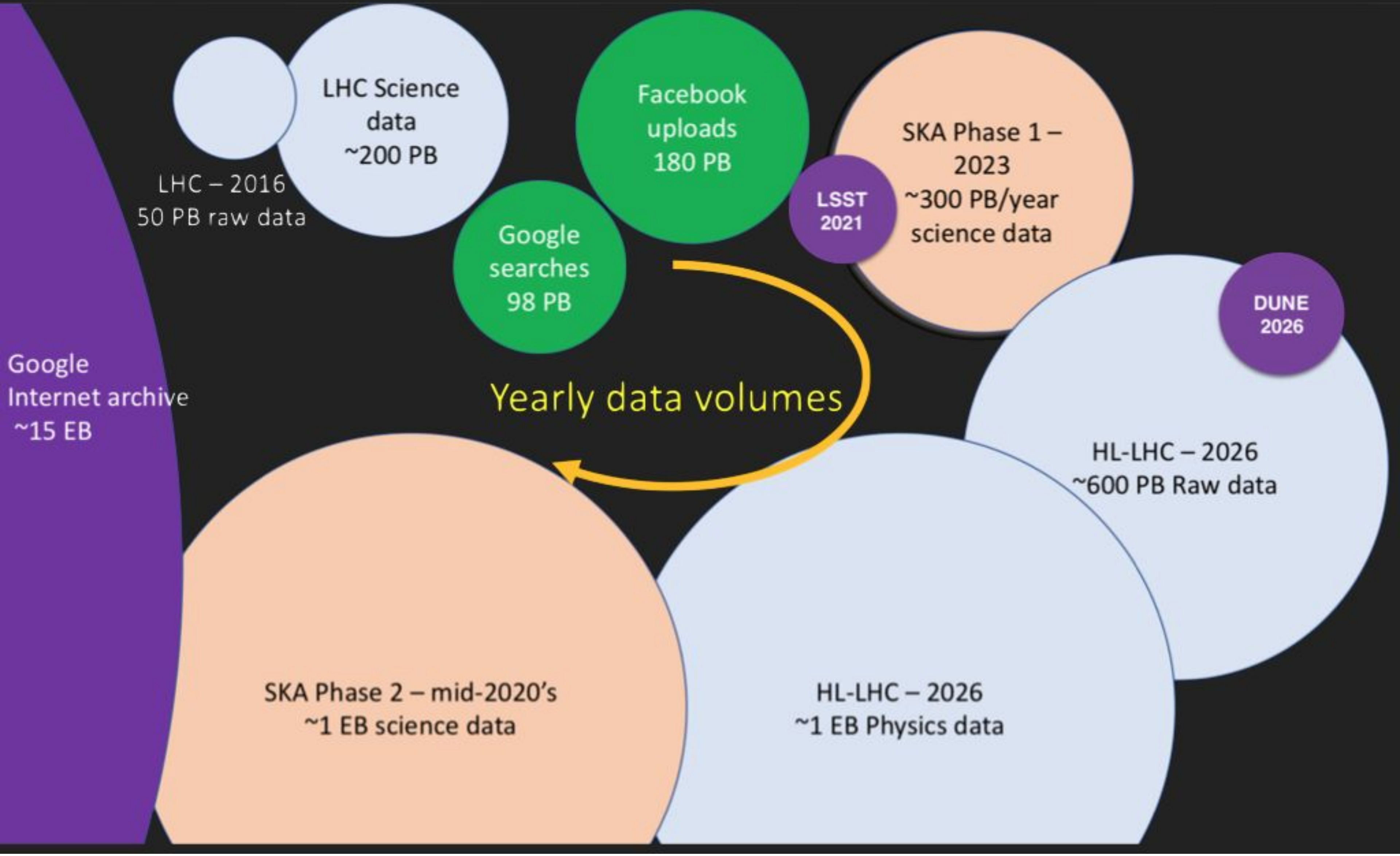

ATLAS Online, 13 TeV
∫Ldt=146.9 fb⁻¹
2015: <μ> = 13.4
2016: <μ> = 25.1
2017: <μ> = 37.8
2018: <μ> = 36.1
Total: <μ> = 33.7
2/19 calibration
Recorded Luminosity [pb⁻¹/0.1]
Mean Number of Interactions per Crossing
0
100
200
300
400
500
600
10
20
30
40
50
60
70
80

NEXT-GEN BIG DATA



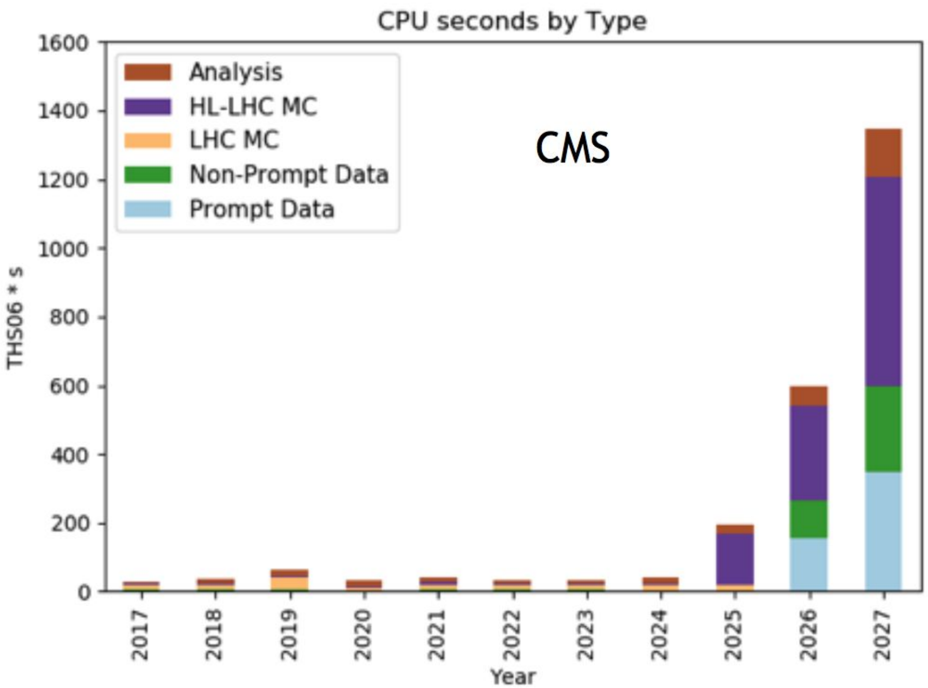
NEXT-GEN BIG DATA

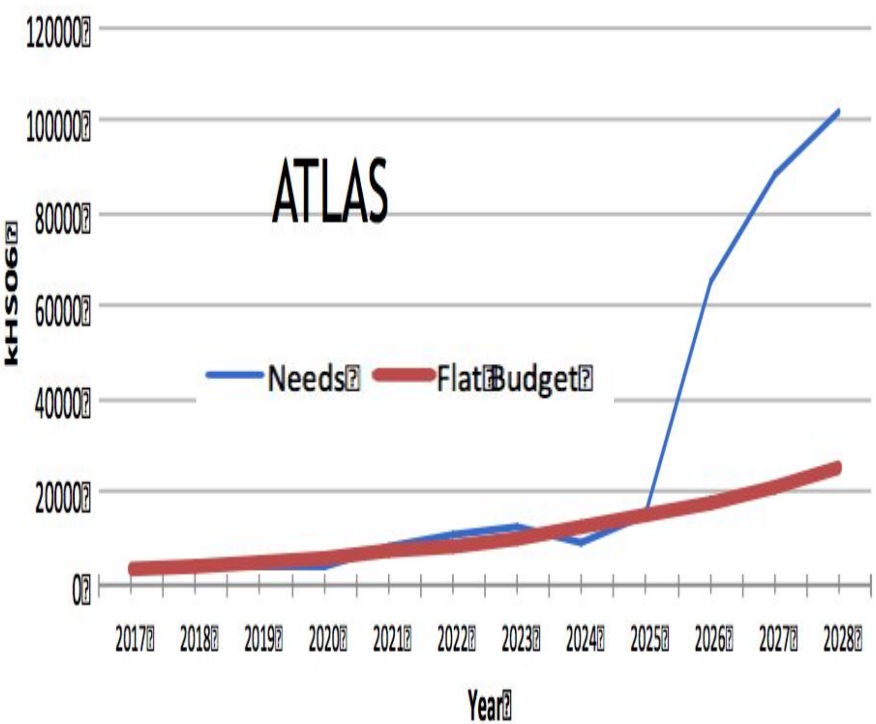
- HL-LHC will reach 1 exabyte of data per year



Computing Challenge at the LHC

CPU: If we stay with plain old CPUs (think of Intel Xeons), and assume more and more computing cores with roughly today's speed ~15M Cores needed per experiment
Disk: ~3 EB per experiment
Tape: ~10 EB per experiment





How to gain back 5x?

- Easy solution:
  - Tighter selection to reduce data volume
  - Run the risk to miss new physics discovery

How to gain back 5x?

- Easy solution:
  - Tighter selection to reduce data volume
  - Run the risk to miss new physics discovery
- Try alternative approaches which preserve physics
  - Be smarter: fewer reprocessings, less simulation, smaller data formats, …
    - IRIS-HEP: Prof. Gordon Watts
  - Cheaper technologies: GPUs, FPGAs seem to offer more “event throughput per $”
    - **HLS4ML: Prof. Shih-Chieh Hsu**
    - Rewrite our algorithms from the group-up to take advantage of High Performance Computers.
    - **Recast our physics problem as Machine Learning problem**

iris hep
Institute for Research & Innovation in Software for High Energy Physics

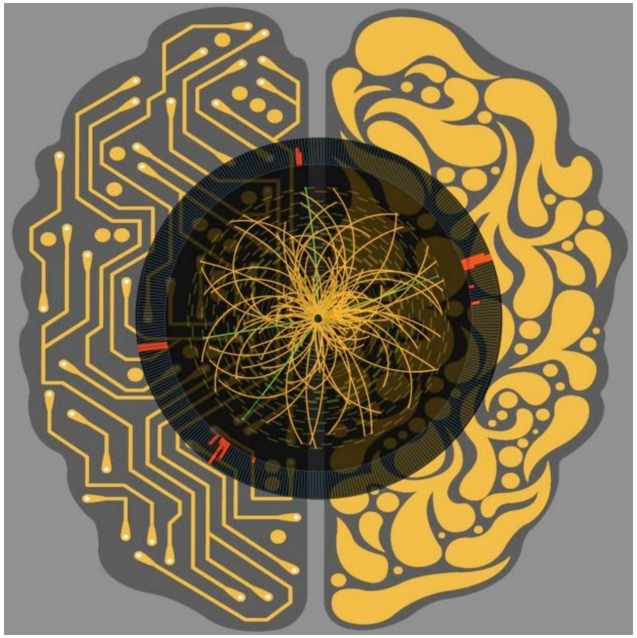
hls 4 ml

How to gain back 5x?

- Easy solution:
  - Tighter selection to reduce data volume
  - Run the risk to miss new physics discovery
- Try alternative approaches which preserve physics
  - Be smarter: fewer reprocessings, less simulation, smaller data formats, …
    - IRIS-HEP: Prof. Gordon Watts
  - Cheaper technologies: GPUs, FPGAs seem to offer more “event throughput per $”
    - **HLS4ML: Prof. Shih-Chieh Hsu**
    - Rewrite our algorithms from the group-up to take advantage of High Performance Computers.
    - **Recast our physics problem as Machine Learning problem**
- Exploring options at the edge of technology?
  - **Quantum Computer: Prof. Shih-Chieh Hsu**

iris hep
Institute for Research & Innovation
in Software for High Energy Physics

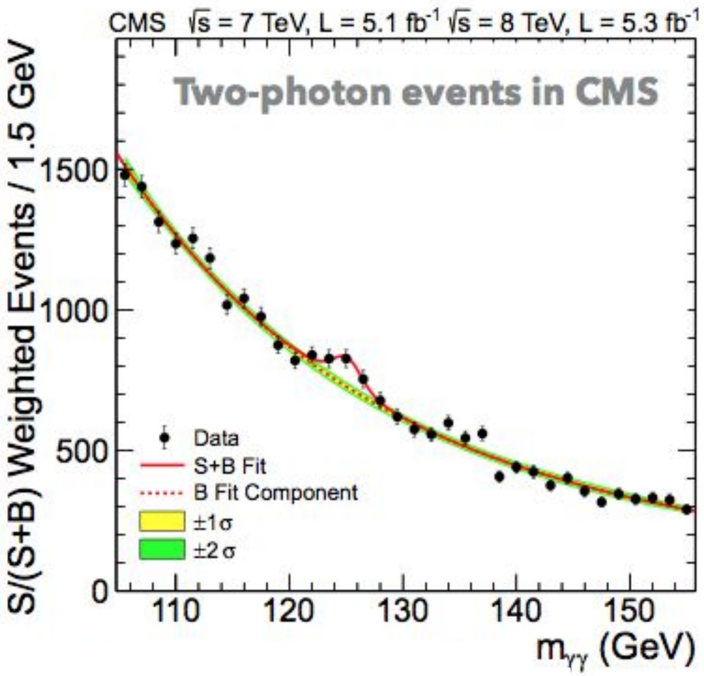
hls4ml



[Fermilab scientists help push AI to unprecedented speeds]()

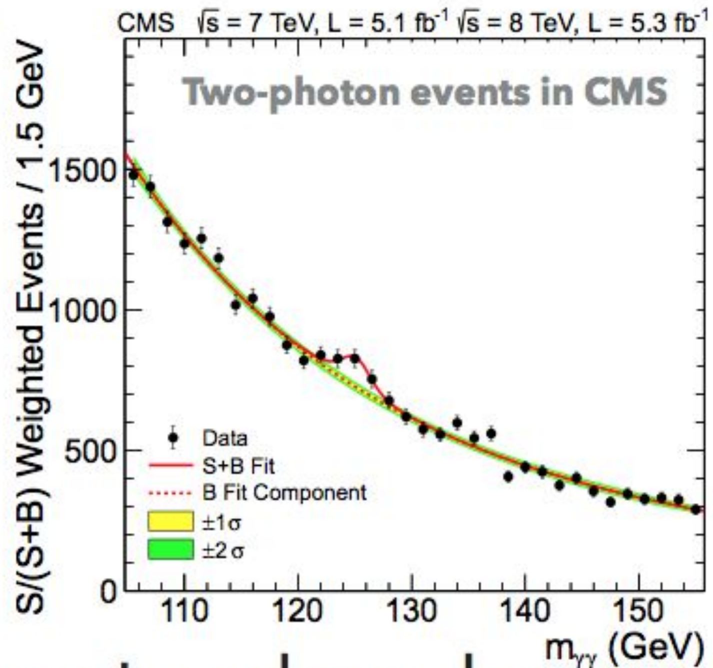
Machine Learning in High Energy Physics

- **Machine learning** was vital to make big discoveries like the Higgs boson on July 4, 2012



Machine Learning in High Energy Physics

- **Machine learning** was vital to make big discoveries like the Higgs boson on July 4, 2012

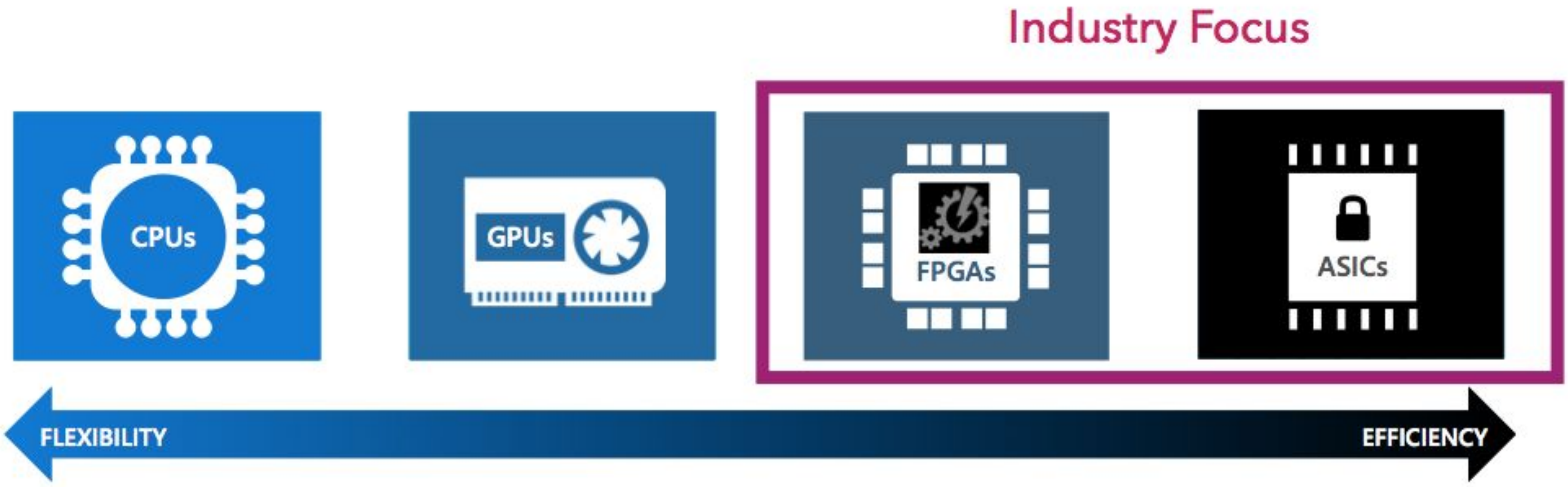


- Today, ML is **enabling** new measurements and searches never thought possible at the LHC
- At the same time, we must **plan** how we will overcome challenges in the **next generation of colliders**
- ML may be a way out

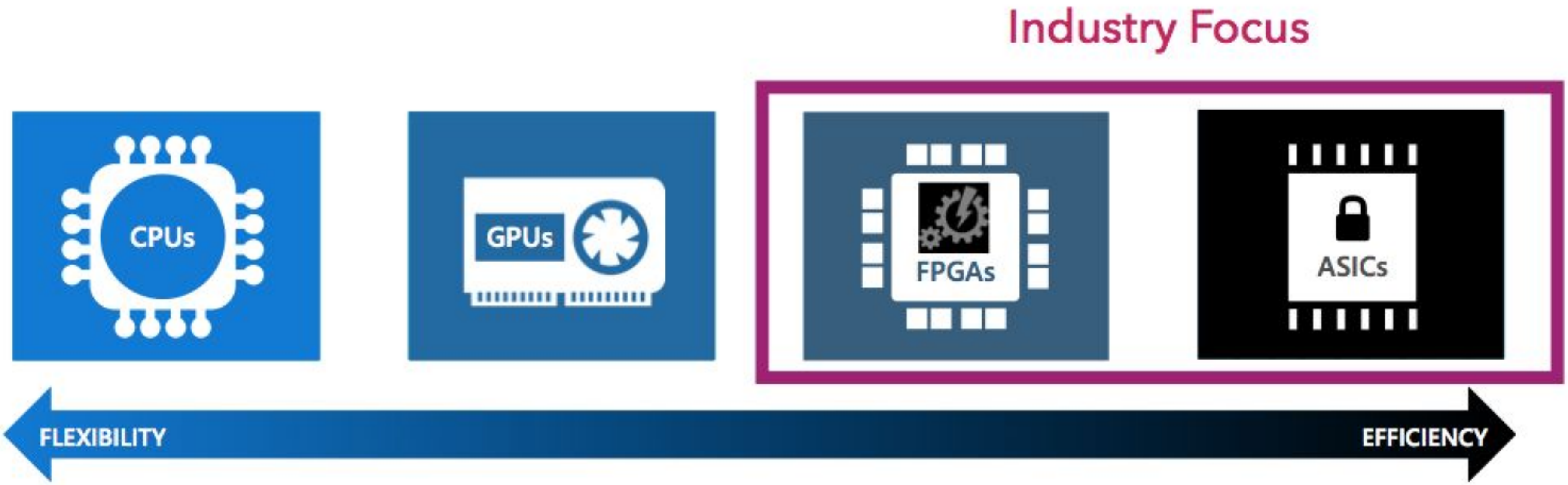
Hardware Alternatives

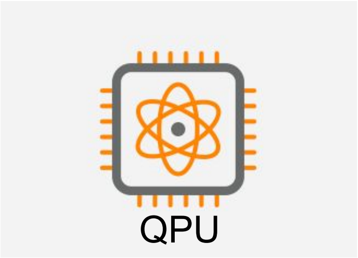
Hardware Alternatives

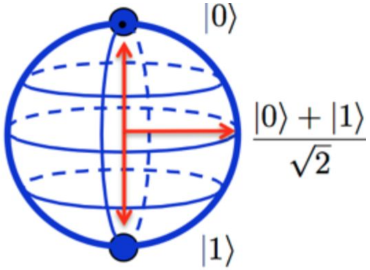
Hardware Alternatives



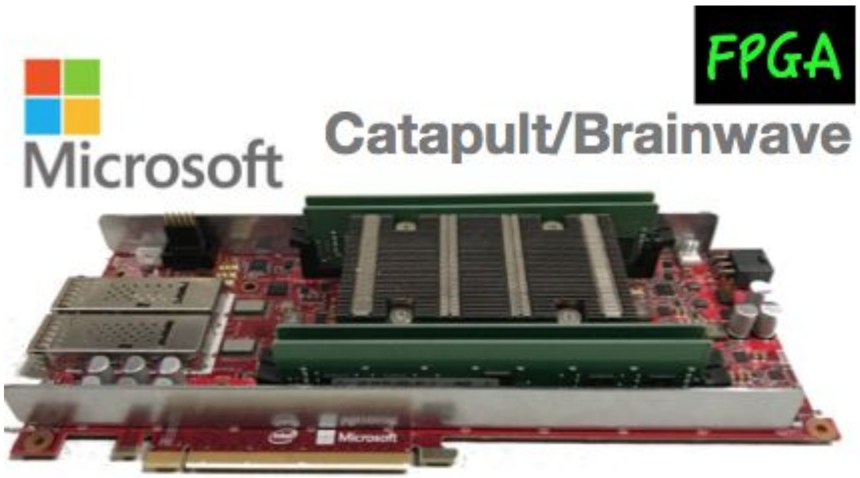
Hardware Alternatives







Trend in Industry



Specialized co-processor hardware for machine learning inference

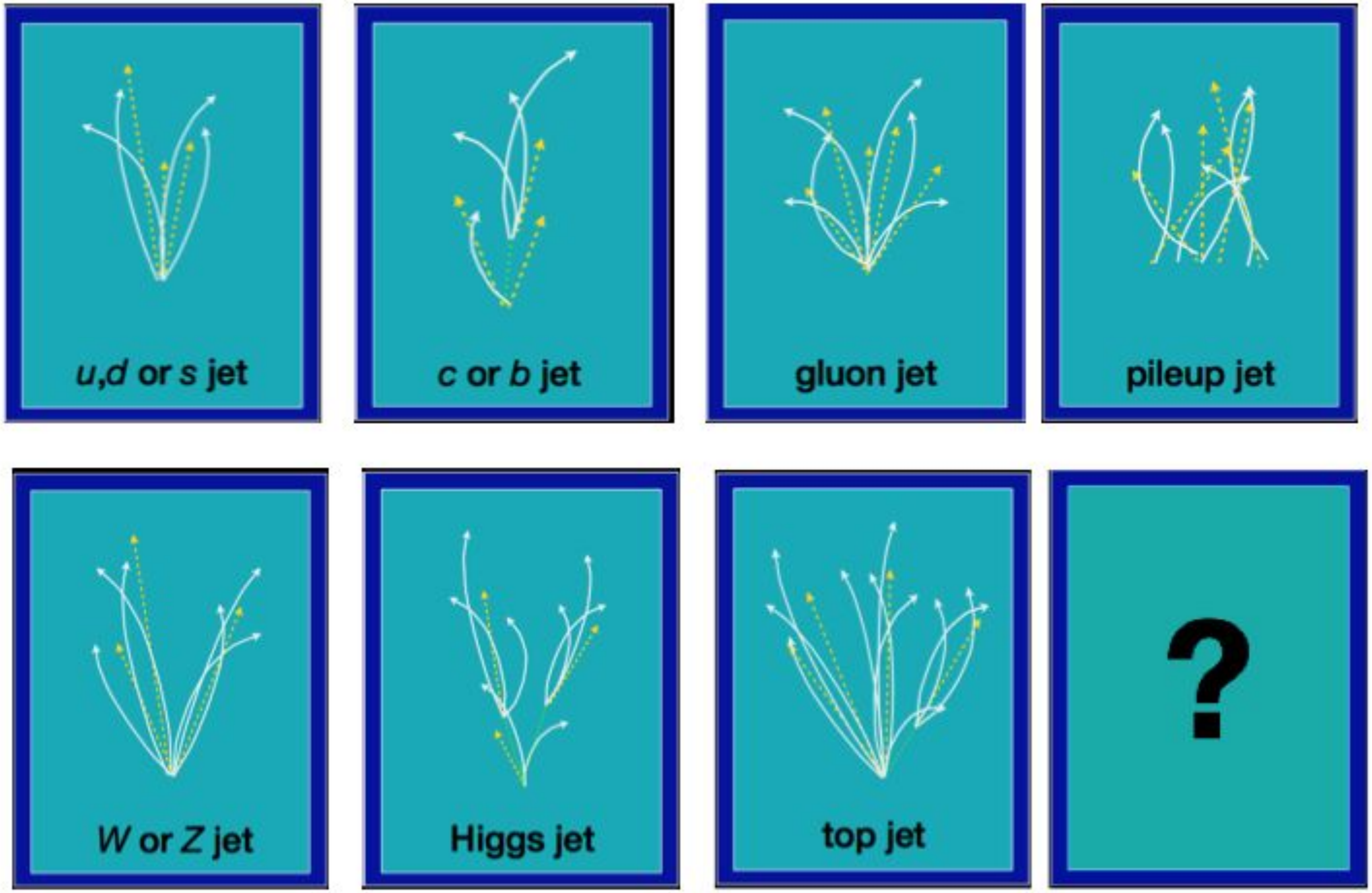
Noisy Intermediate Scale Quantum (NISQ)

Quantum Chip	Qubits	Announced	Qubit Archetype	Computing Model
D-Wave XX & 2000Q  	~5000	02/2019	Superconducting **flux** qubits	Quantum annealing
	2048	01/2017		
IBM 20Q and 50Q  	20	11/2017	Superconducting **transmon** qubits	Quantum circuits
	50	11/2017 (tests)		
Rigetti 19Q 	19	12/2017	Superconducting **transmon** qubits	Quantum circuits
Intel Tangle Lake 	49	01/2018 (tests)	Superconducting qubits	Quantum circuits
Google Bristlecone 	72	03/2018 (tests)	Superconducting **transmon** qubits	Quantum circuits
UC Berkeley QNL 	4 (8)	2017	Superconducting **transmon** qubits	Quantum circuits
	64	2022 ?		

arXiv:1801.00862

Slide credit: I. Shapoval

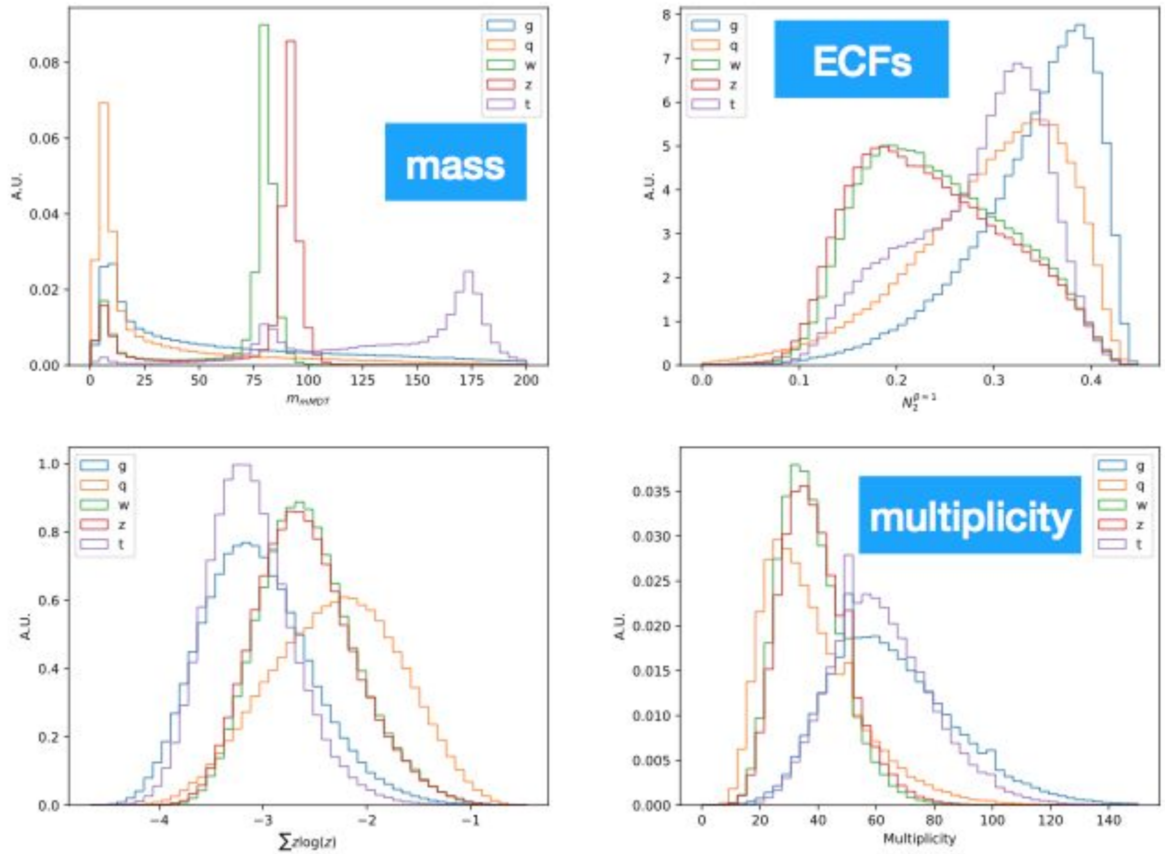
Jet Classification



Jet Classification Variables

Observables

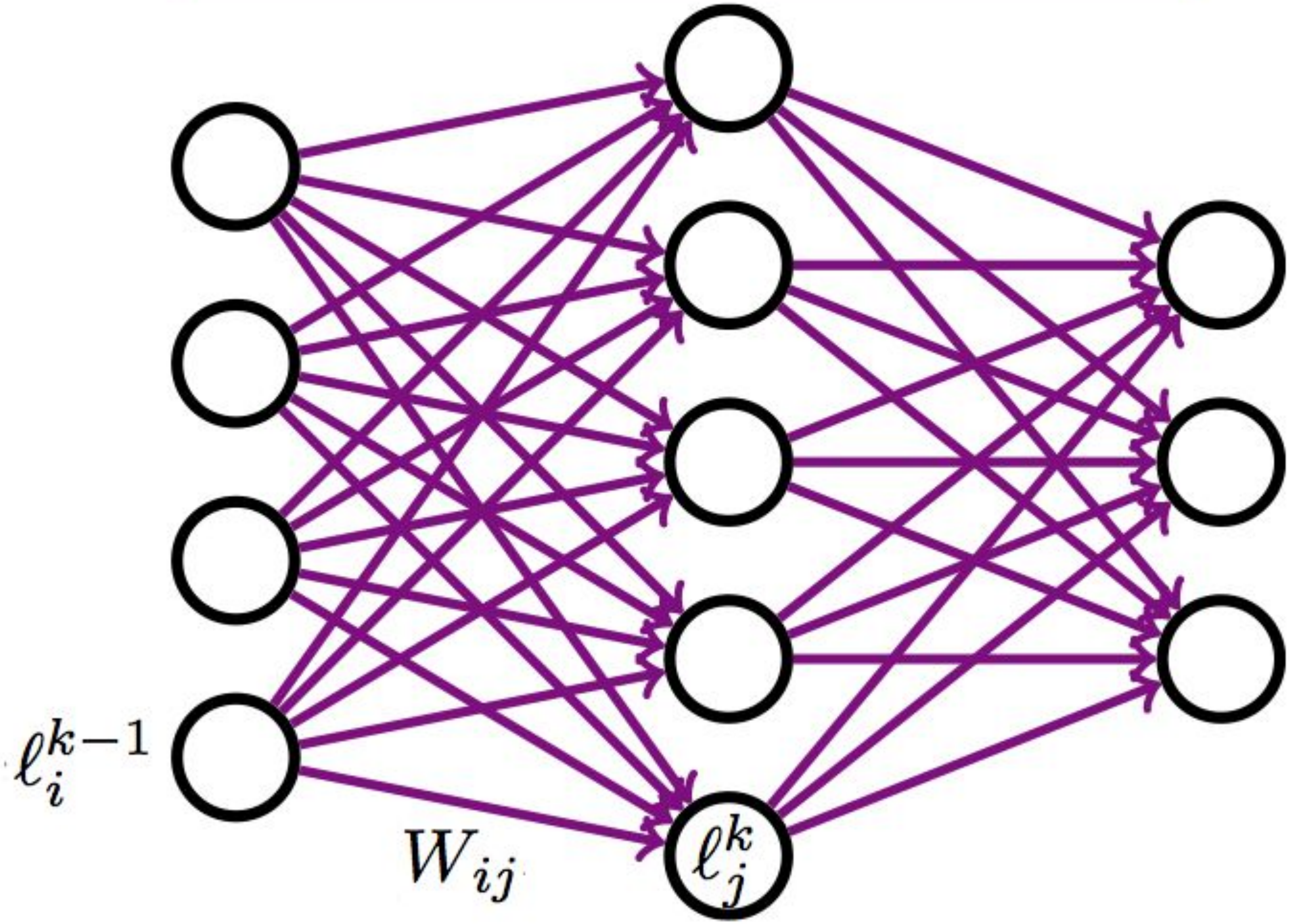
m_{mMDT}
$N_2^{\beta=1,2}$
$M_2^{\beta=1,2}$
$C_1^{\beta=0,1,2}$
$C_2^{\beta=1,2}$
$D_2^{\beta=1,2}$
$D_2^{(\alpha,\beta)=(1,1),(1,2)}$
$\sum z \log z$
Multiplicity



- 16 expert observables provide separation between top, W/Z, and quark/gluon

Neural Network Model

$$\boxed{\ell_j^k = \phi(W_{ij}\ell_i^{k-1} + b_j)}$$



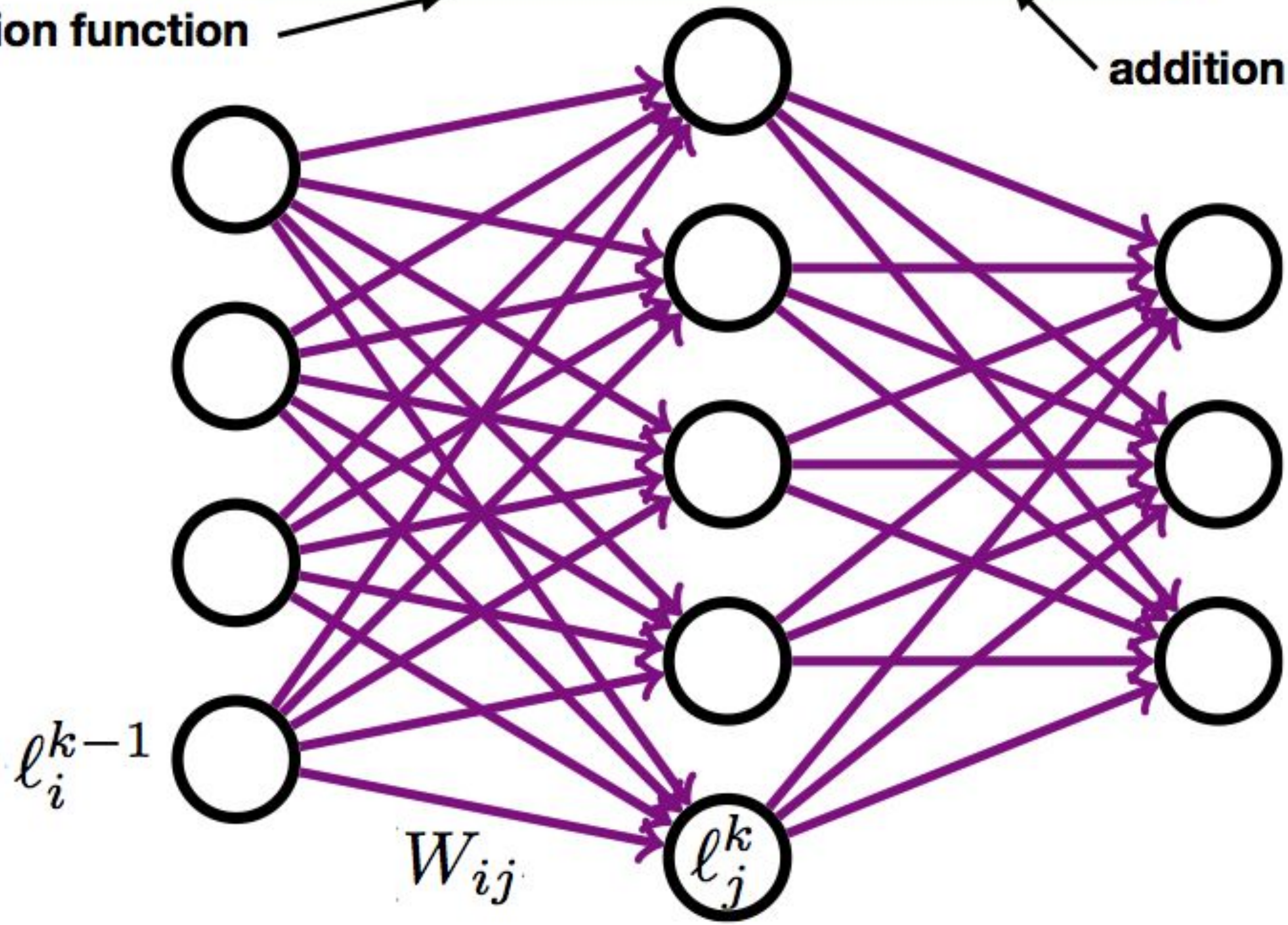
Neural Network Model

multiplication

$$\boxed{\ell_j^k = \phi(W_{ij}\ell_i^{k-1} + b_j)}$$

activation function

addition



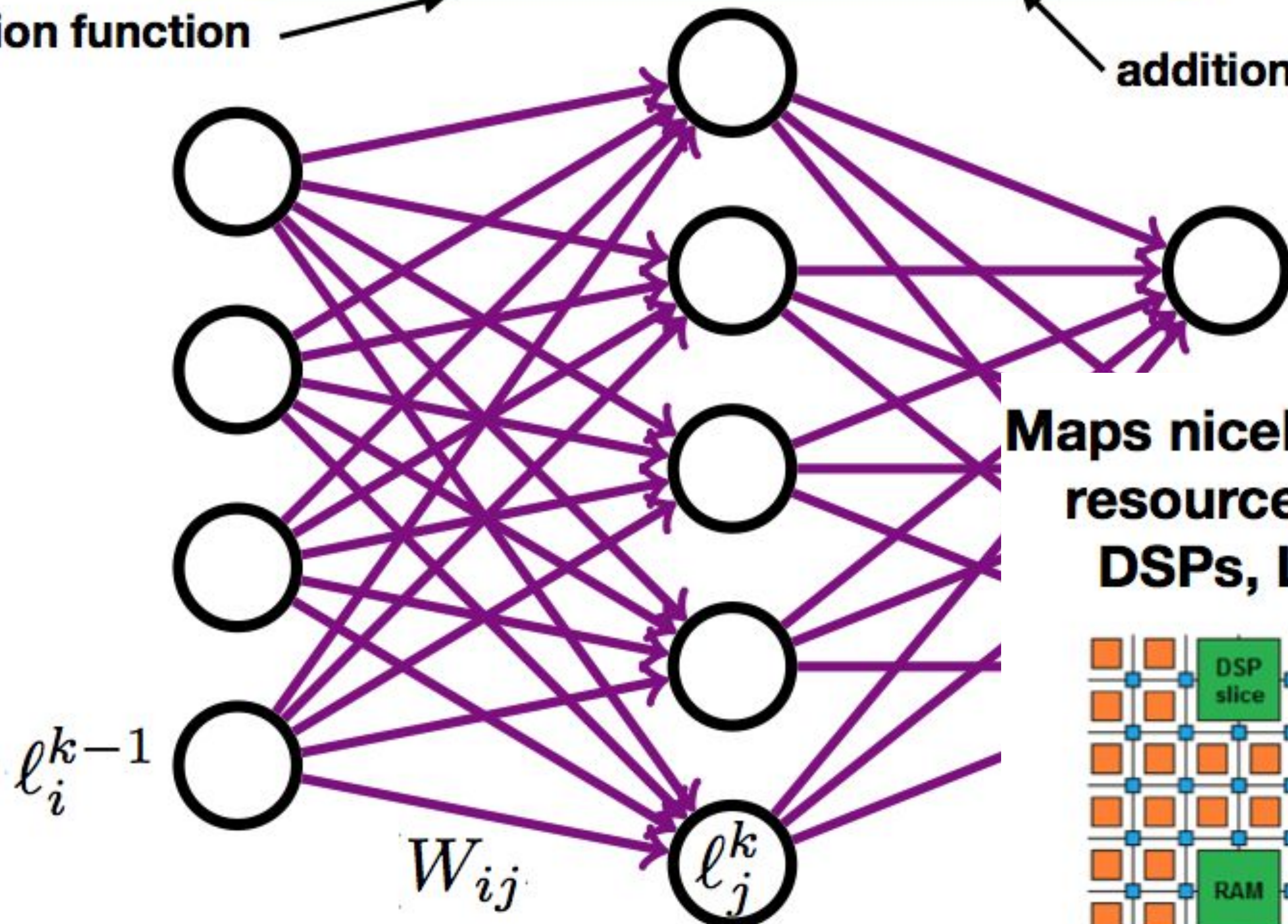
Neural Network Model

multiplication

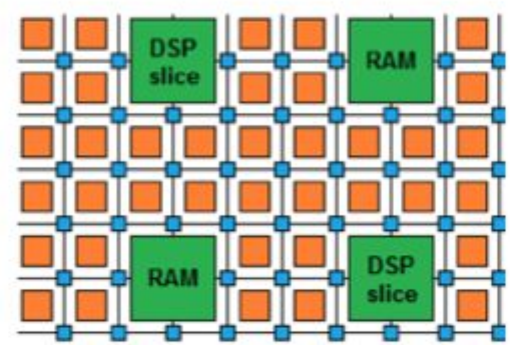
$$\boxed{\ell_j^k = \phi(W_{ij}\ell_i^{k-1} + b_j)}$$

activation function

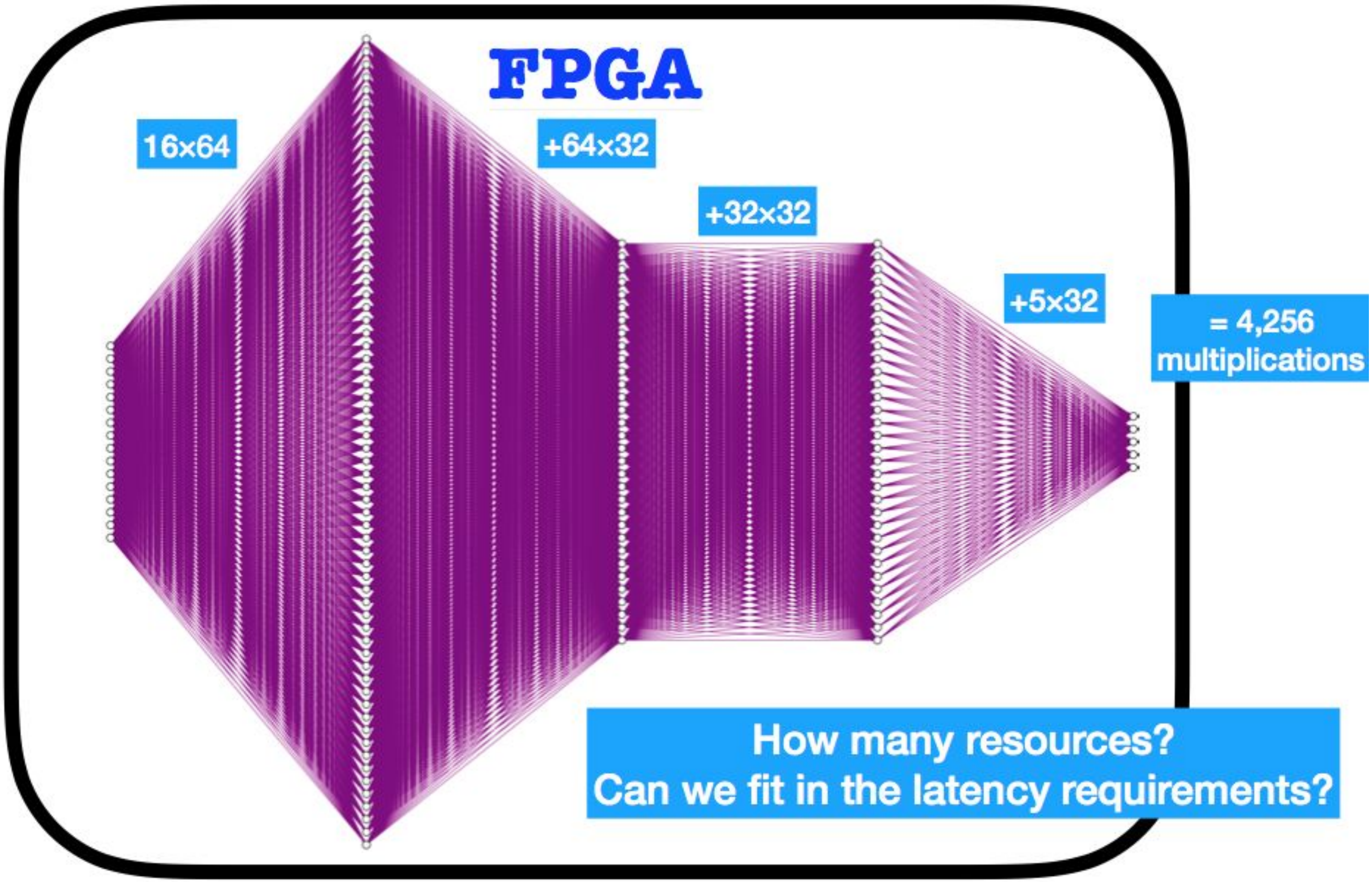
addition



Maps nicely onto FPGA resources: high IO, DSPs, LUTs, etc.



Neural Network Model

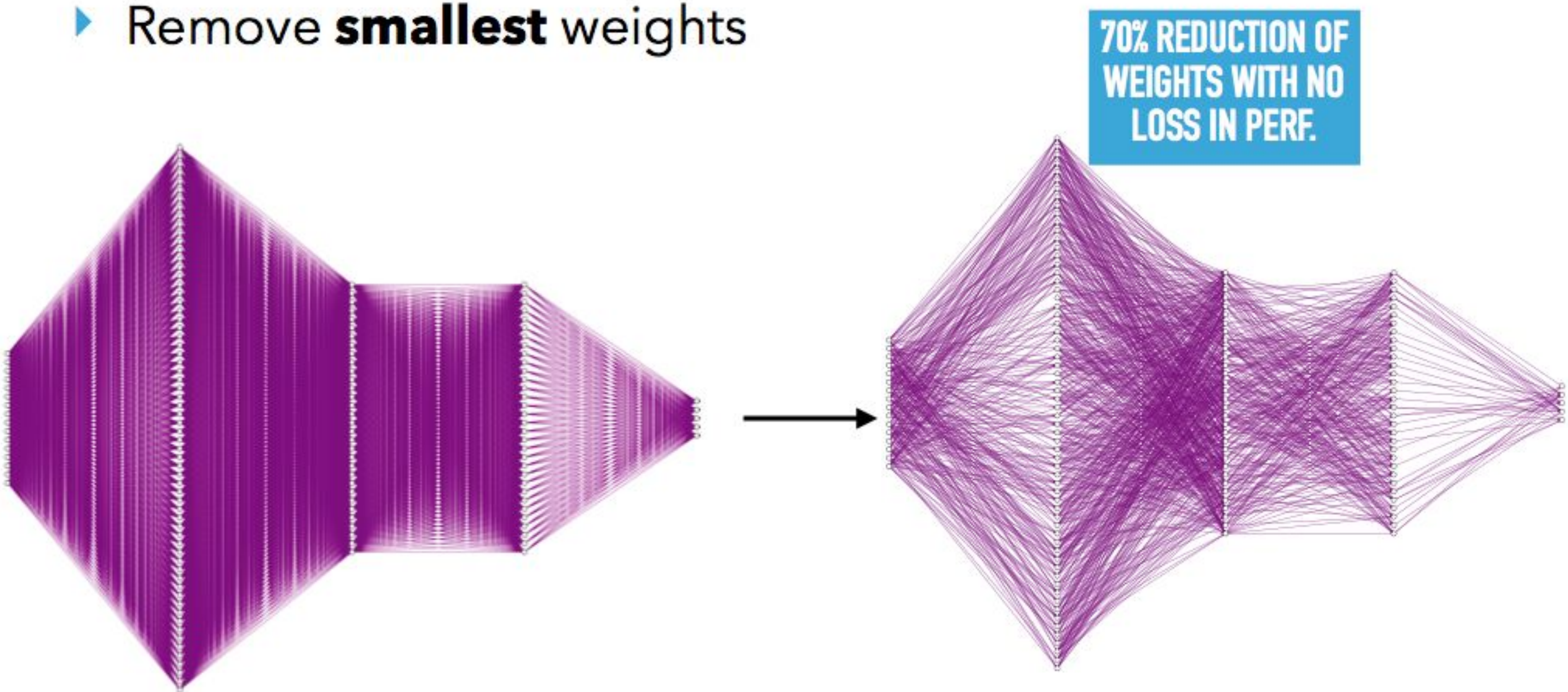


Neural Network Tuning: Compression

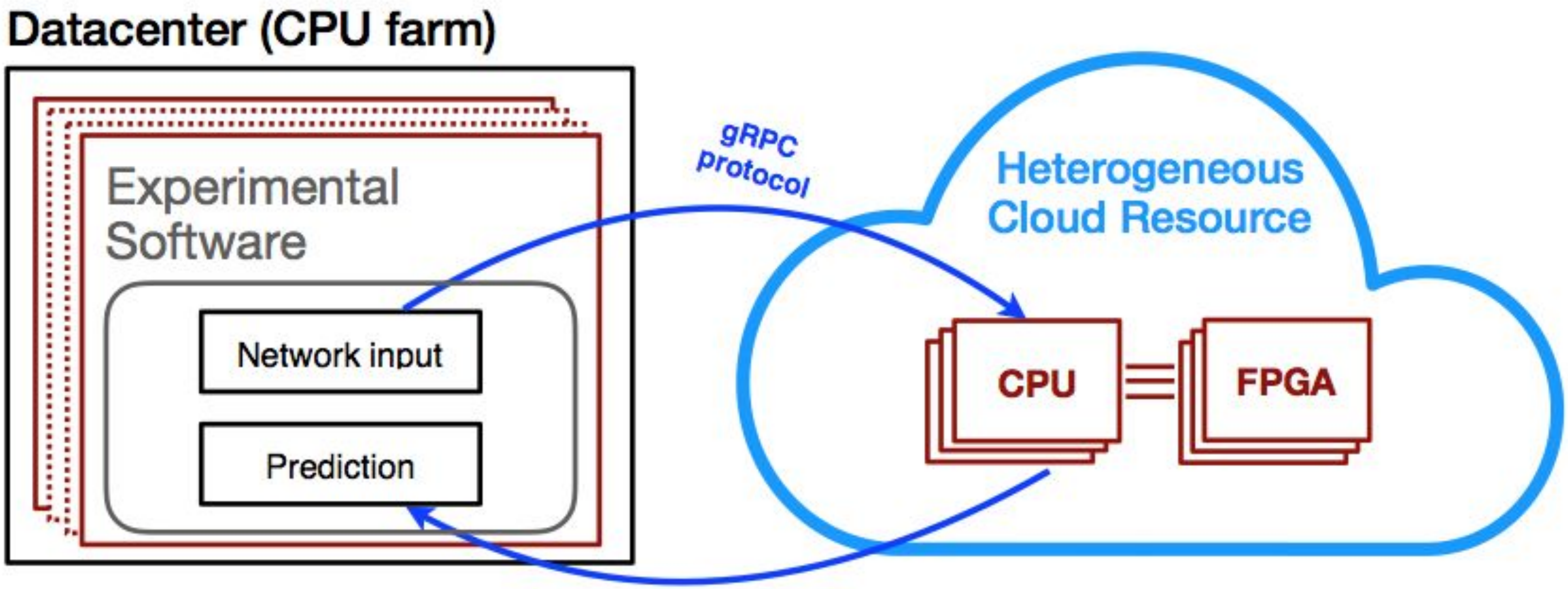
- Train with **L_1 regularization** (down-weights unimportant synapses)

$$L_\lambda(\boldsymbol{w}) = L(\boldsymbol{w}) + \lambda \|\boldsymbol{w}\|_1 \qquad \|\boldsymbol{w}\|_1 = \sum_i |w_i|$$

- Remove **smallest** weights

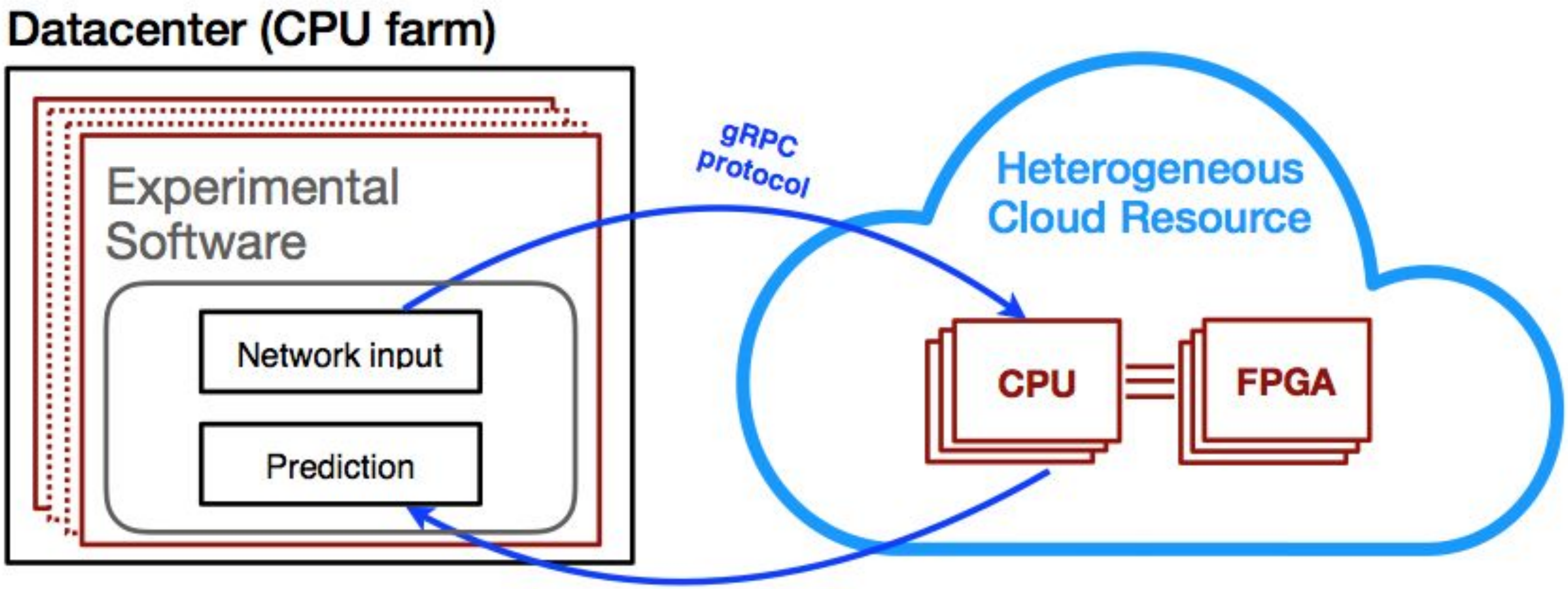


Cloud Service vs Edge Service

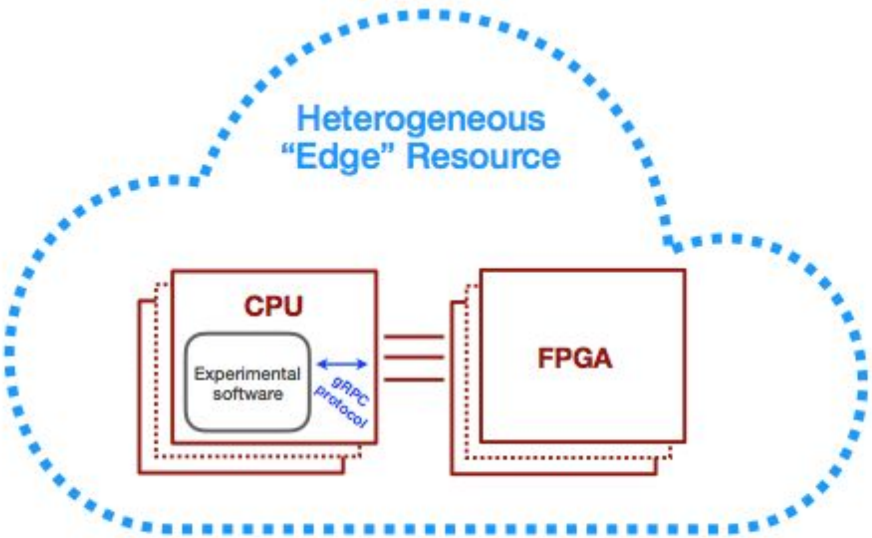


- Cloud service has latency

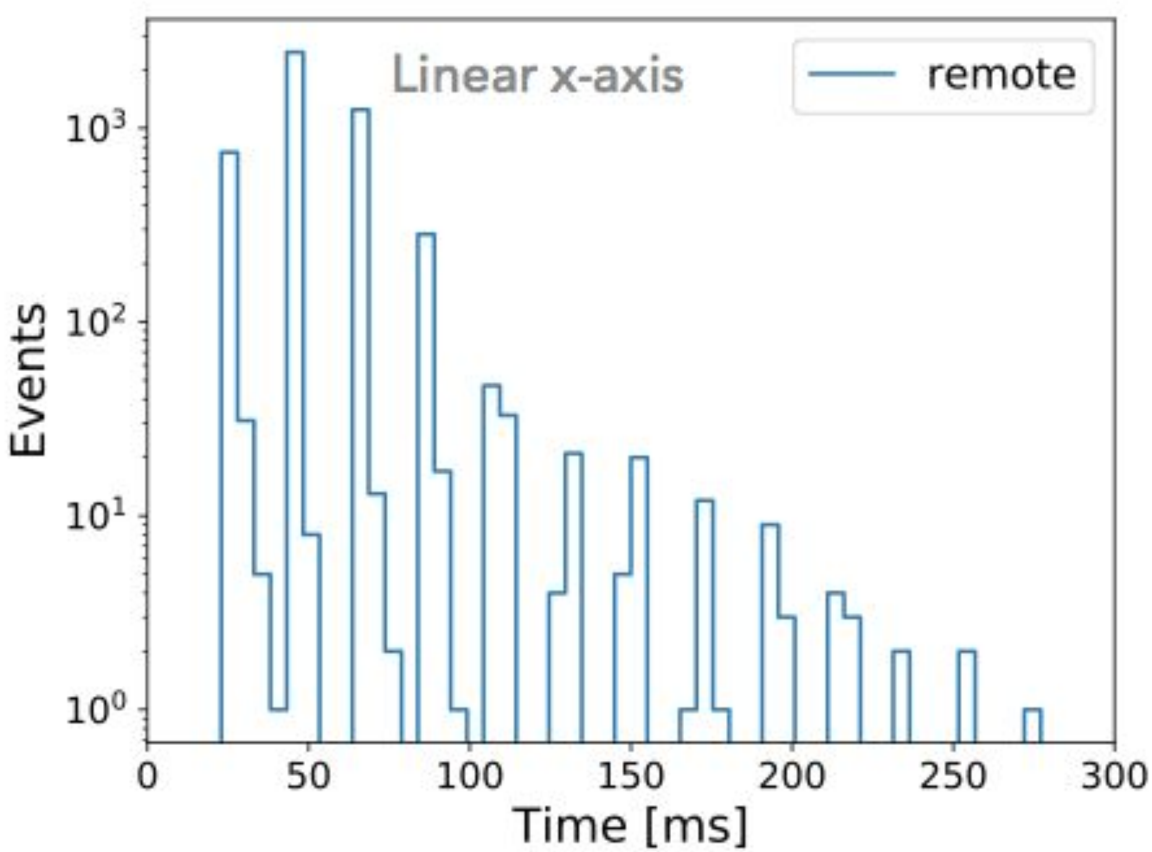
Cloud Service vs Edge Service



- Cloud service has latency
- local installation of FPGAs ("on-prem" or "edge")

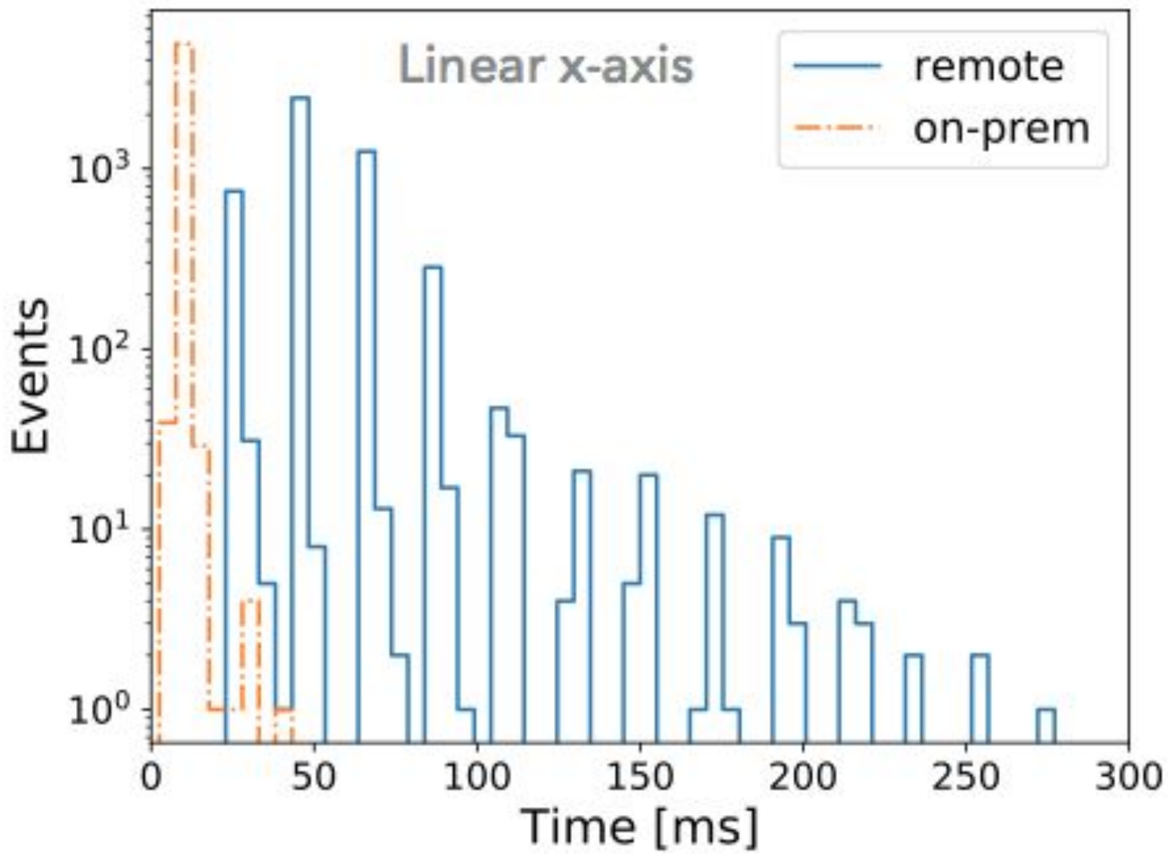


Fast Inference as Service

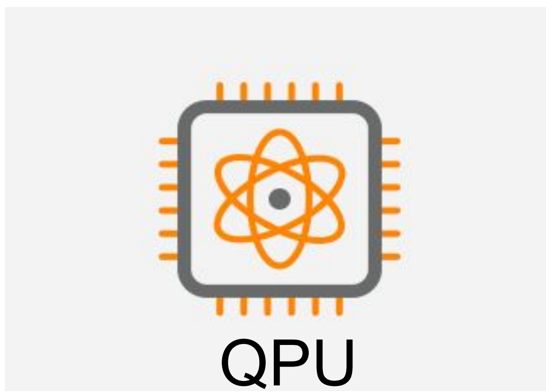


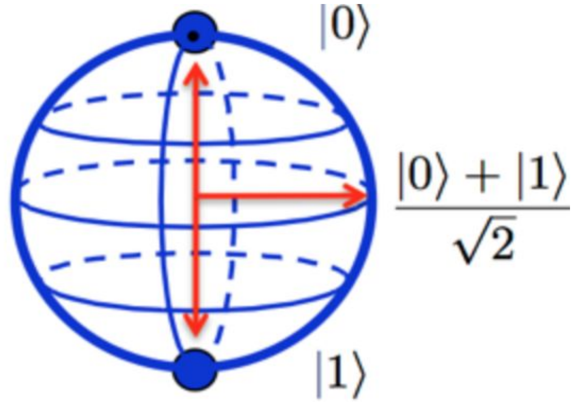
- Remote: FNAL (IL) to Azure (VA) ‹time› = 60 ms
  - Highly dependent on network conditions

Fast Inference as Service



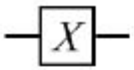
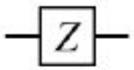
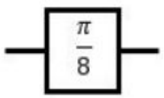
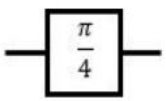
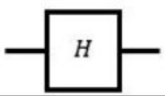
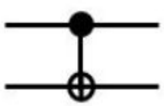
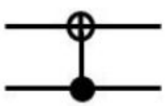
- Remote: FNAL (IL) to Azure (VA) ‹time› = 60 ms
  - Highly dependent on network conditions
- On-prem: run CMSSW on Azure ‹time› = 10 ms
  - on FPGA: 1.8 ms for inference
  - Remaining time used for classifying and I/O





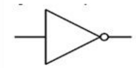
Jakub Kruper (undergrad)
Alessandro Roggero (postdoc)
Nathan Wieber (scientist)

Logic Gate: Quantum vs Classical

Gate	Notation	Matrix
NOT (Pauli-X)		$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$
Pauli-Z		$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$
$\frac{\pi}{8}$ gate		$\begin{pmatrix} 1 & 0 \\ 0 & e^{i\pi/4} \end{pmatrix}$
$\frac{\pi}{4}$ gate		$\begin{pmatrix} 1 & 0 \\ 0 & e^{i\pi/2} \end{pmatrix}$
Hadamard		$\frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$
Controlled C^1NOT_2		$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$
C^2NOT_1		$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$

Universal Quantum Gate

NOT Gate

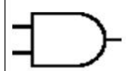


A	Y
0	1
1	0

Boolean Expression

Y = A'

AND Gate



A	B	Y
0	0	0
0	1	0
1	0	0
1	1	1

Boolean Expression

Y = A . B

OR Gate

A	B	Y
0	0	0
0	1	1
1	0	1
1	1	1

Boolean Expression

Y = A + B

XNOR Gate

A	B	Y
0	0	1
0	1	0
1	0	0
1	1	1

Boolean Expression

Y = A⊙B

XOR Gate

A	B	Y
0	0	0
0	1	1
1	0	1
1	1	0

Boolean Expression

Y = A ⊕ B

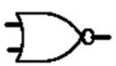
NAND Gate

A	B	Y
0	0	1
0	1	1
1	0	1
1	1	0

Boolean Expression

Y = (A . B)' = A' + B'

NOR Gate



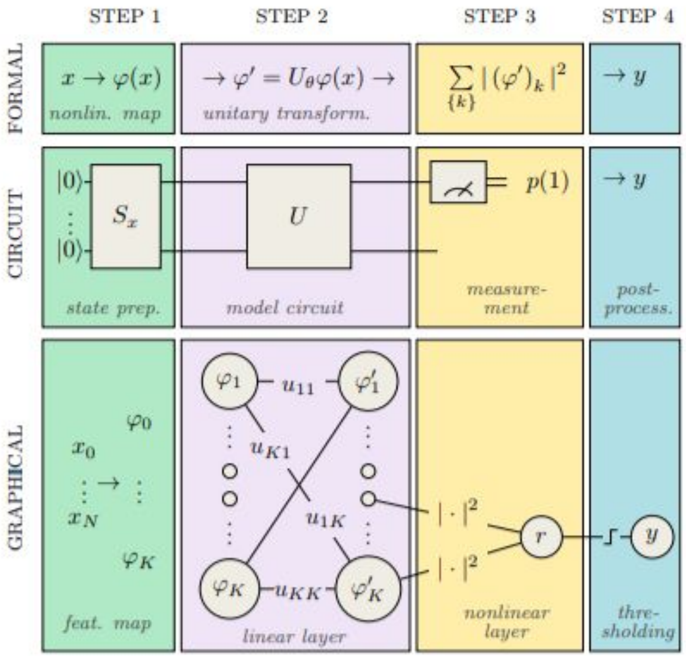
A	B	Y
0	0	1
0	1	0
1	0	0
1	1	0

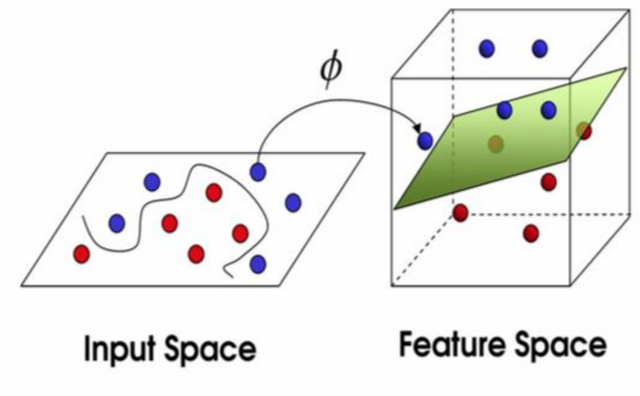
Boolean Expression

Y = (A + B)'

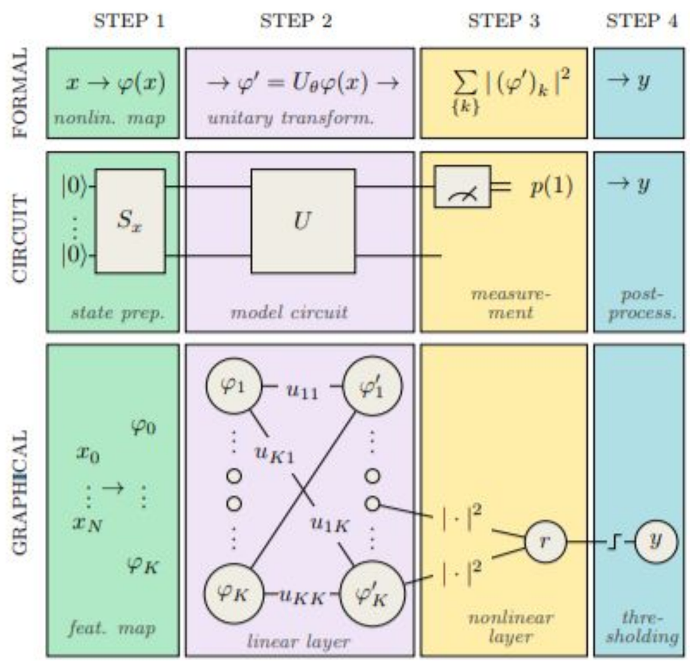
Universal Gate

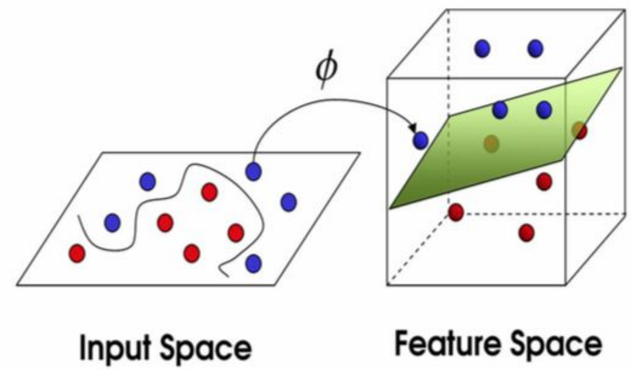
Quantum Higgs Classifier

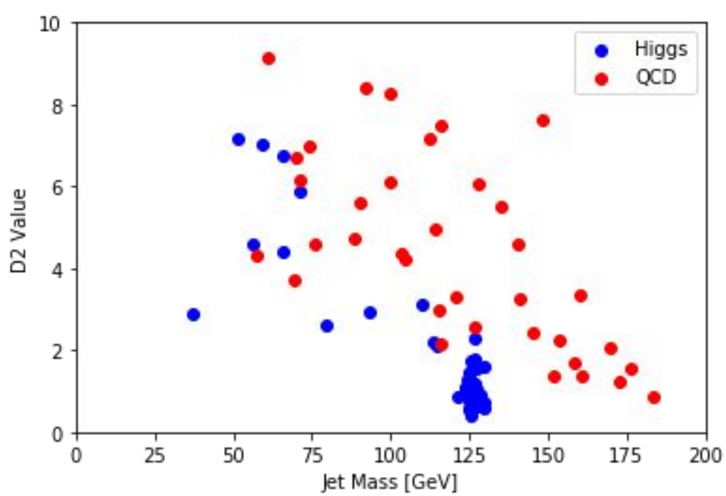


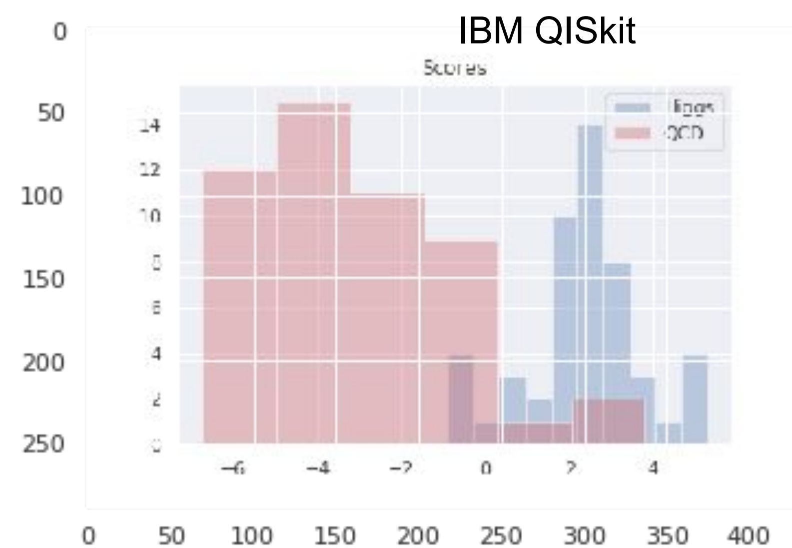


Quantum Higgs Classifier









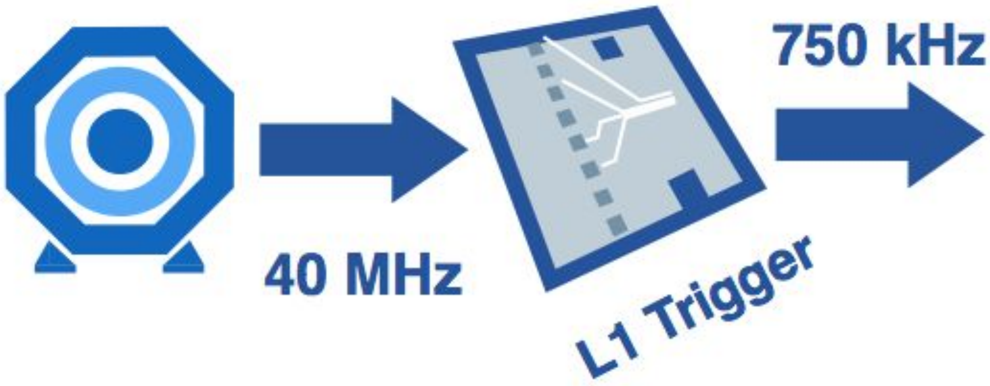
Summary

- Particle Physics confronted Big Data challenge in coming decades
  - CPU consumptions
  - Data storage
- Frontier approach - recasting physics problem to machine learning problem
- Accelerating Machine Learning with cutting edge processors
  - Taking advantage from Industry trend of co-processor development
- Exploration of Quantum Machine Learning started
- Contact me (schsu@uw.ed) to learn more details!

Field-programmable gate array

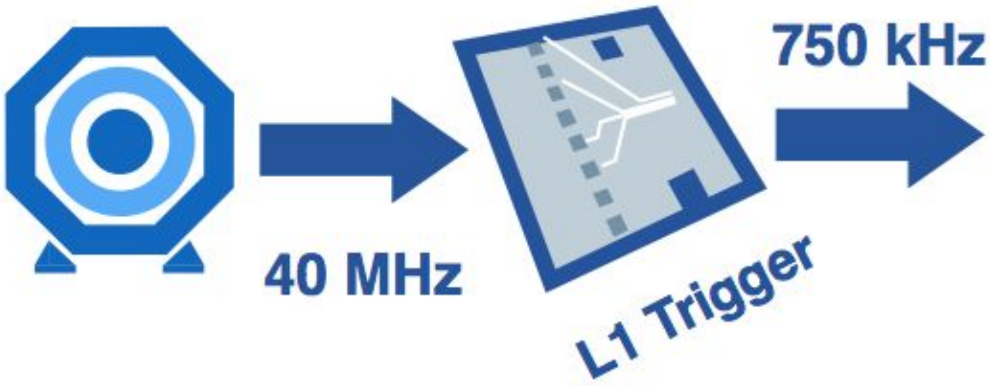
an integrated circuit designed to be configured by a customer or a designer after manufacturing

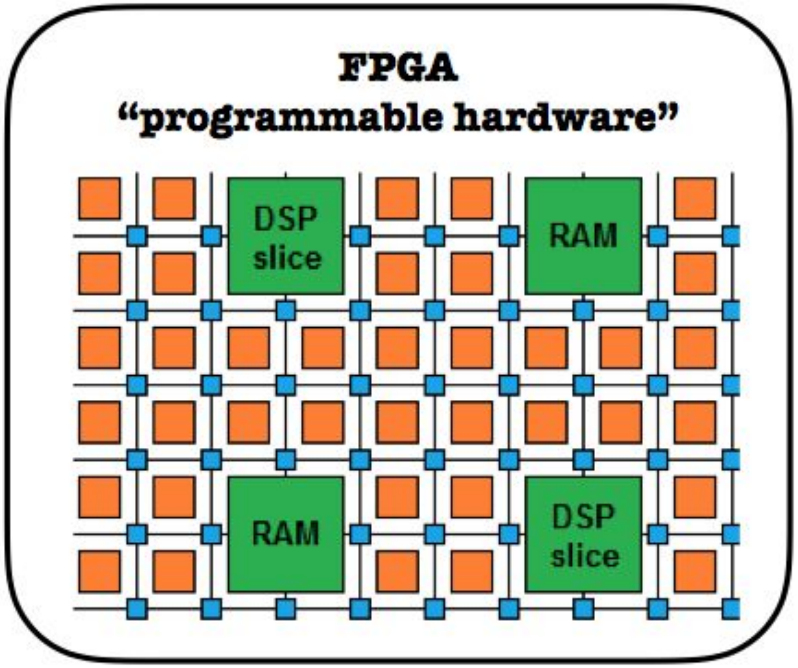
Level-1 Trigger



- Level-1 Trigger:
  40 MHz → 750 kHz
- Reconstruct and filter
  2% of events in ~12 μs

Level-1 Trigger

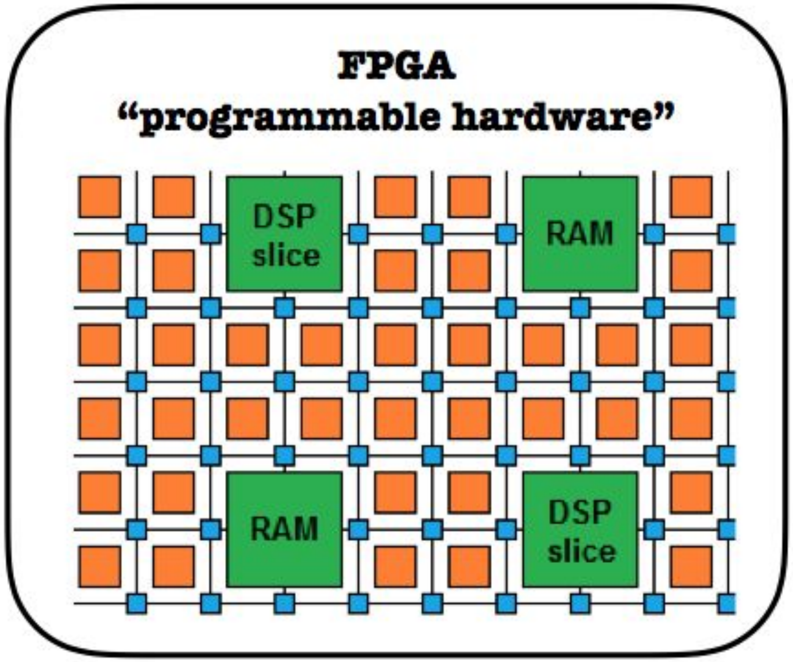




- Level-1 Trigger:
  40 MHz → 750 kHz
- Reconstruct and filter
  2% of events in ~12 μs
- Latency necessitates all
  FPGA design

Level-1 Trigger

- **Pros:**
  - **Reprogrammable** interconnects between **embedded components** that perform multiplication (**DSPs**), apply logical functions (**LUTs**), or store memory (**BRAM**)
  - **High throughput:** O(100) optical transceivers running at O(15) Gbs
  - **Massively parallel**
  - **Low power**



- **Cons:**
  - **Requires domain knowledge to program** (using VHDL/Verilog)

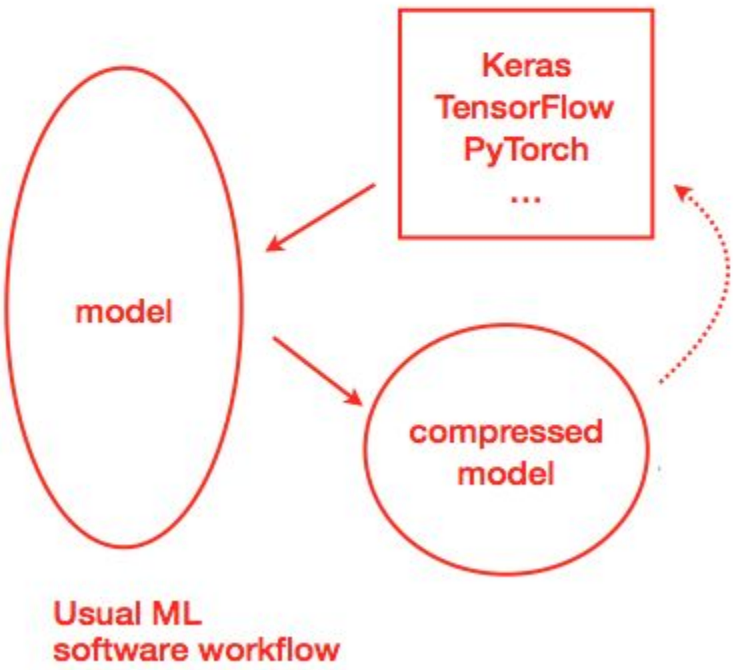
hardware description language (HDL)

Performance Comparison

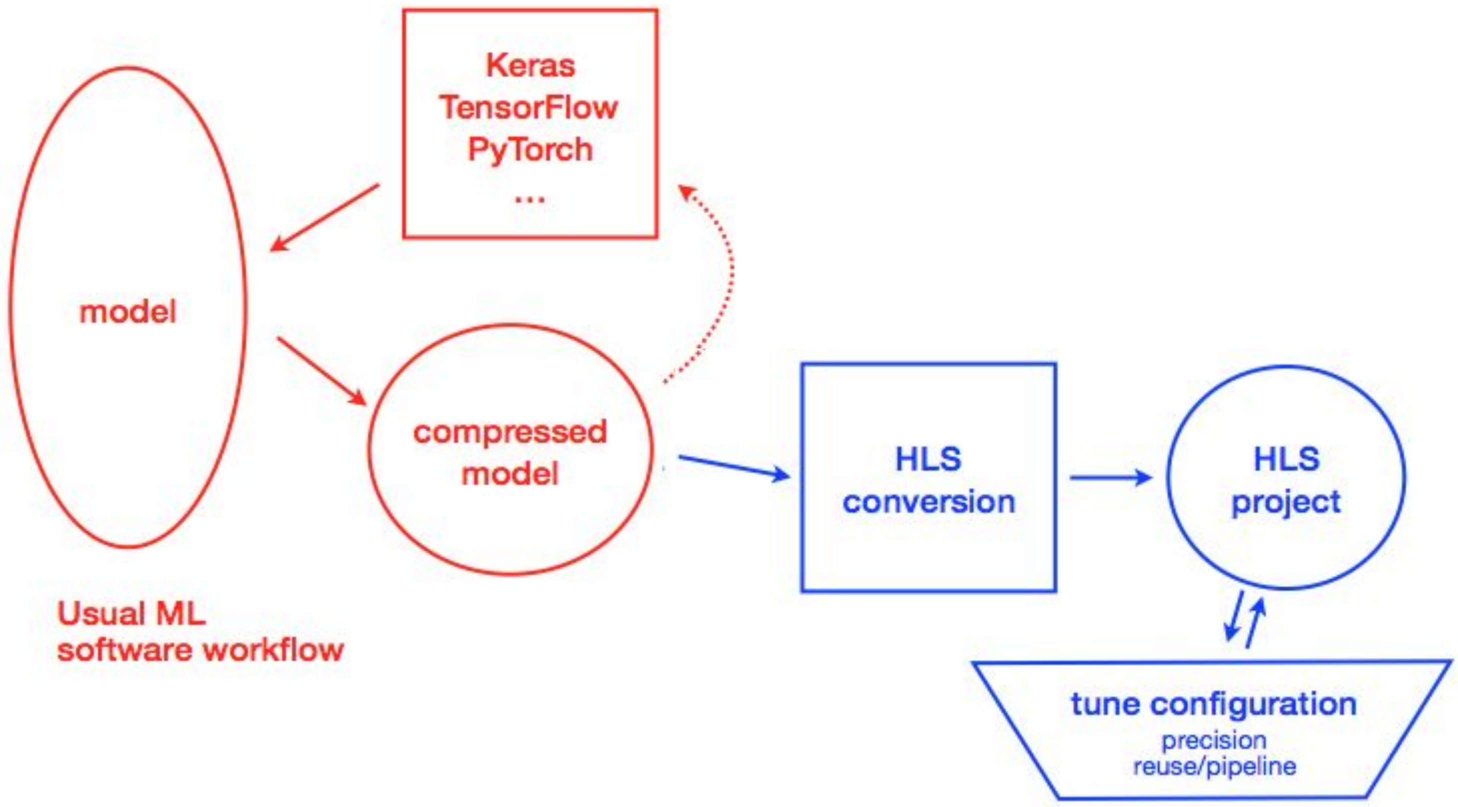
Type	Note	Latency [ms]	Throughput [inf./s]
CPU*	Xeon 2.6 GHz	1750	0.6
	i7 3.6 GHz	500	2
GPU**	batch = 1	7	143
	batch = 32	1.5	667
Brainwave	remote	60	660
	on-prem	10 (1.8 on FPGA)	660

- **Brainwave** is 175x (30x) on-prem (remote) faster than **CPU**
- **Brainwave** has competitive throughput vs **GPU** with single-image as service
- Comparison with other co-processors, e.g. **AWS FPGA**, **Google TPU** are on-going active research activities.

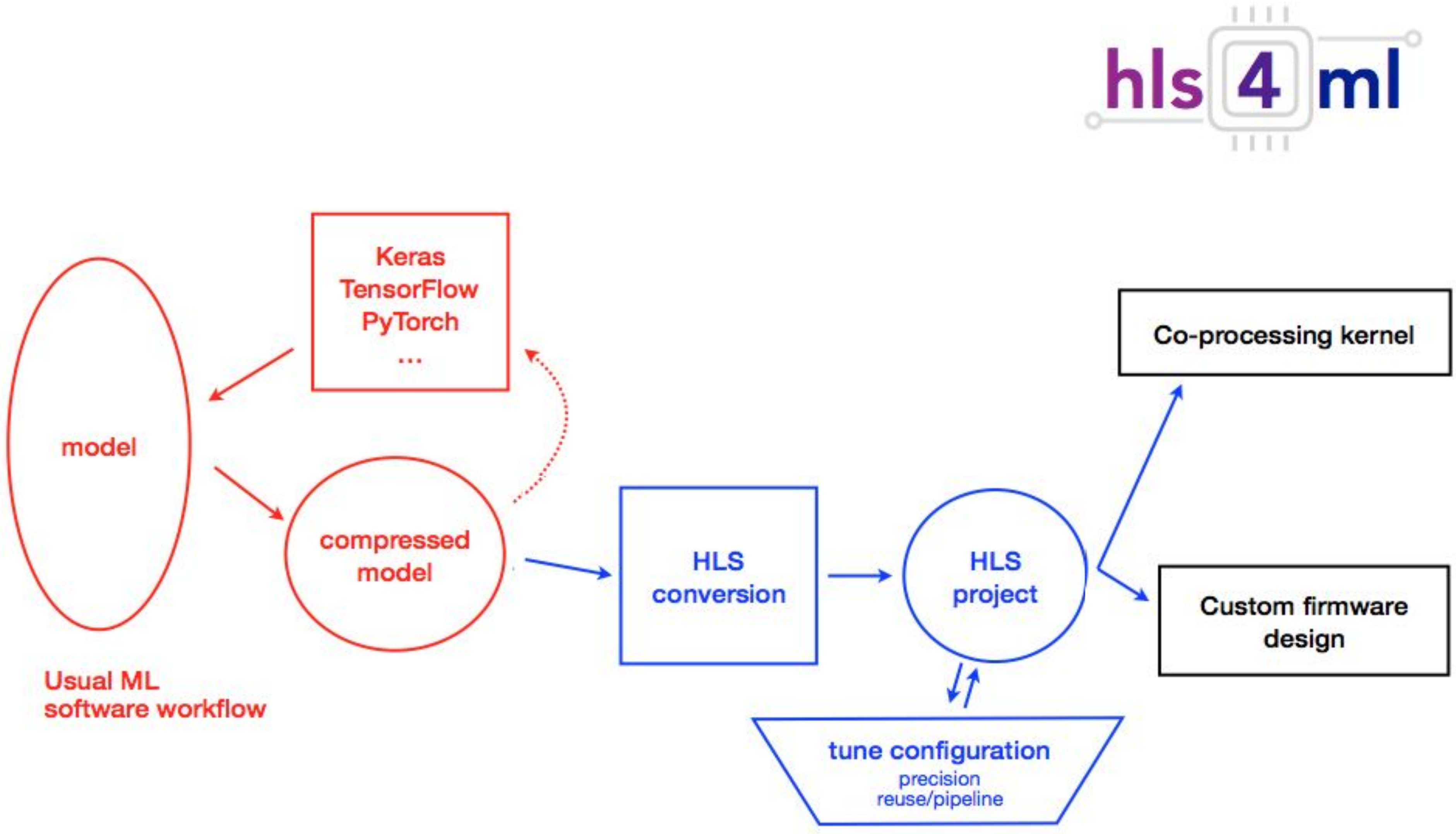
HLS4ML Workflow



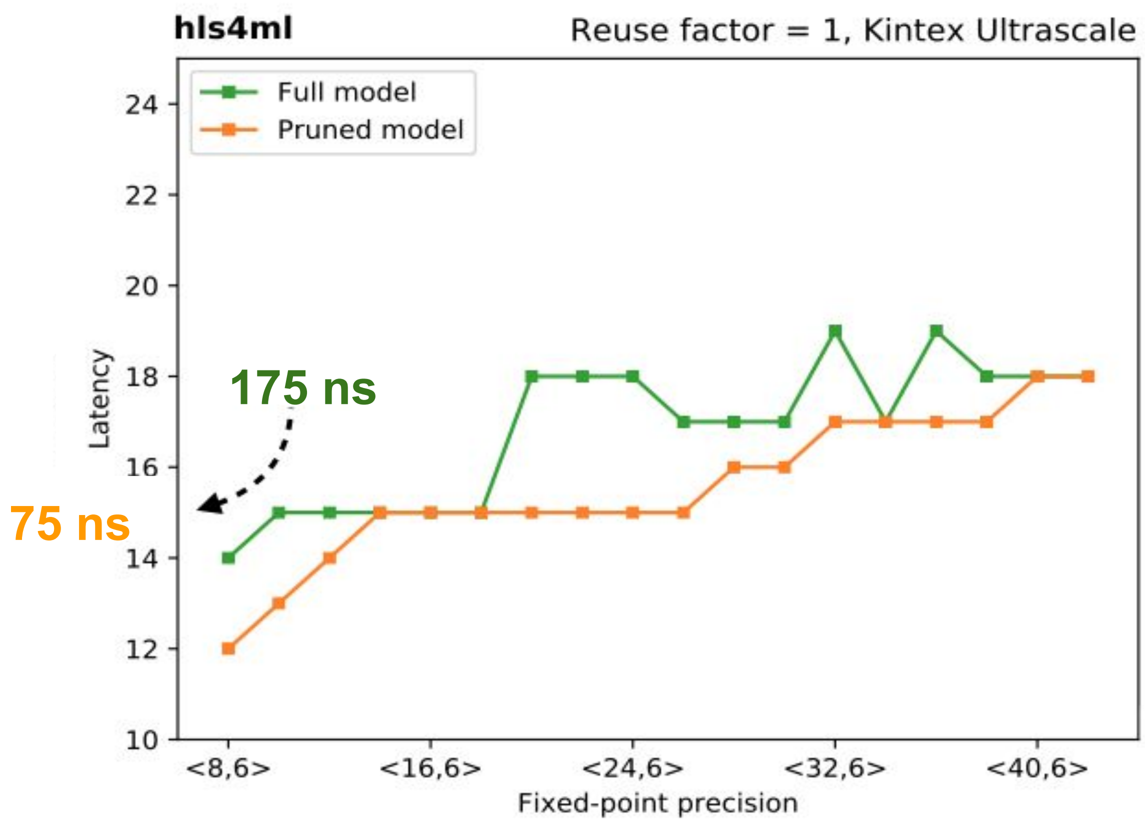
HLS4ML Workflow



HLS4ML Workflow



Network Tuning: Faster Inference



~35 clocks
@ 200 MHz
= 175 ns

~15 clocks
@ 200 MHz
= 75 ns

ap_fixed<width,integer>

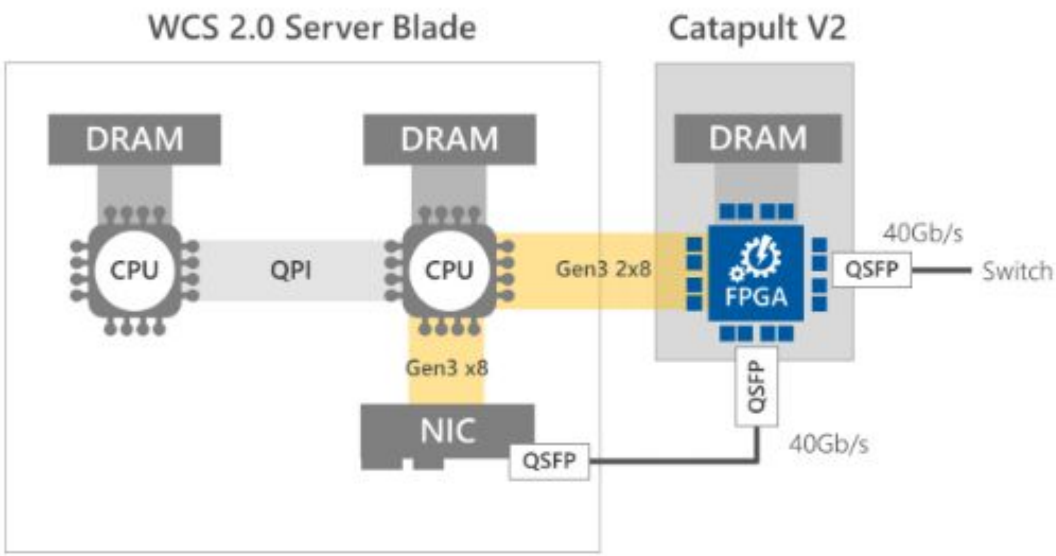
0101.1011101010

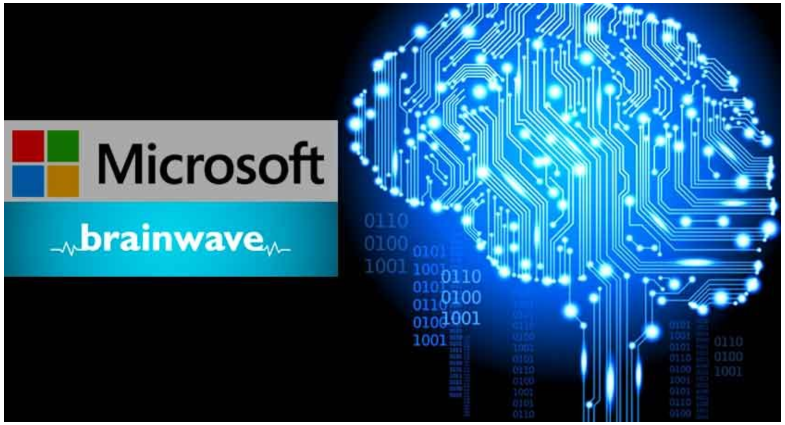
integer fractional

width

Xilinx Vivado 2017.2
Clock frequency: 200 MHz
FPGA: Xilinx Kintex Ultrascale
(XCKU115-FLVB2104)

FPGA as Service



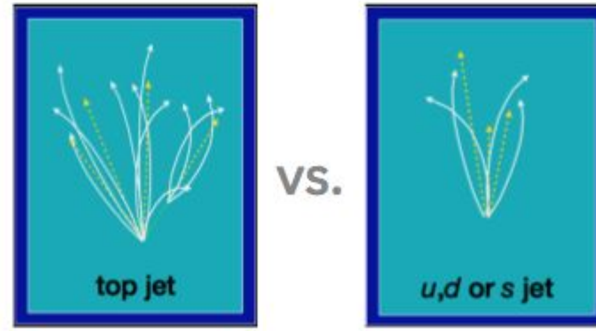


Prof. Scott Hauck (ECE)
Matthew Trahms (undergrad)
Kylie Lim (undergrad)
Donovan Erickson (undergrad)
Jessica Lan (undergrad)

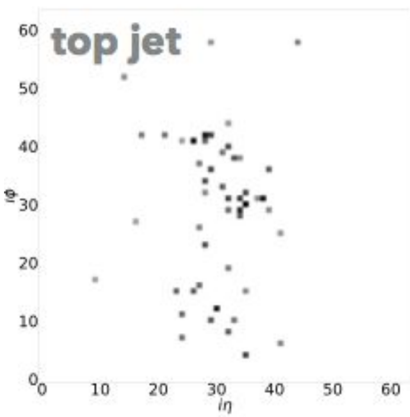
Alex Schuy (grad)
Haoran Zhao (grad)

Jet Vision with ResNet-50

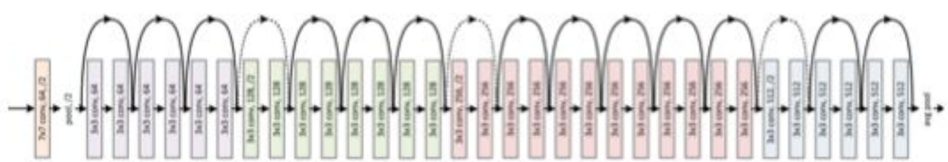
- Re-train ResNet-50 to identify the origin of jets
- Inputs are *jet images* = pixelated versions of calorimeter hits in 2D (η, Φ)



input



ResNet-50



output

$$\begin{pmatrix} \text{top jet} & 0.9757 \\ \text{QCD jet} & 0.0243 \end{pmatrix}$$