VBS at Snowmass:

MACHINE LEARNING FOR VBS

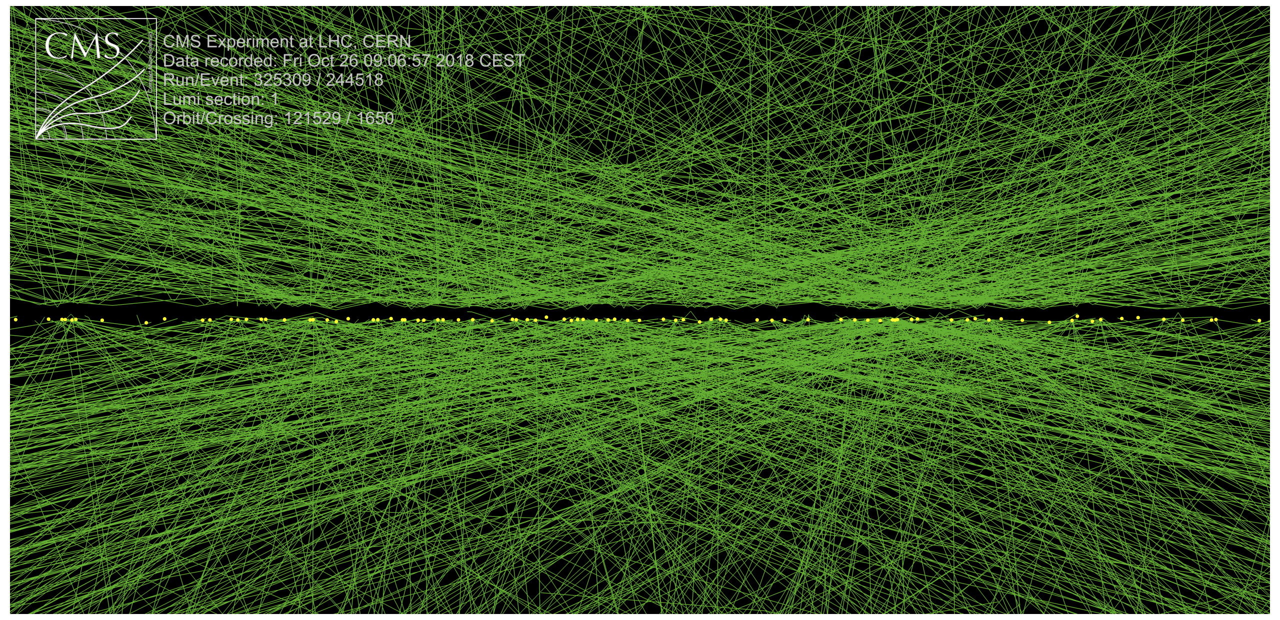
Thea Årrestad (CERN)

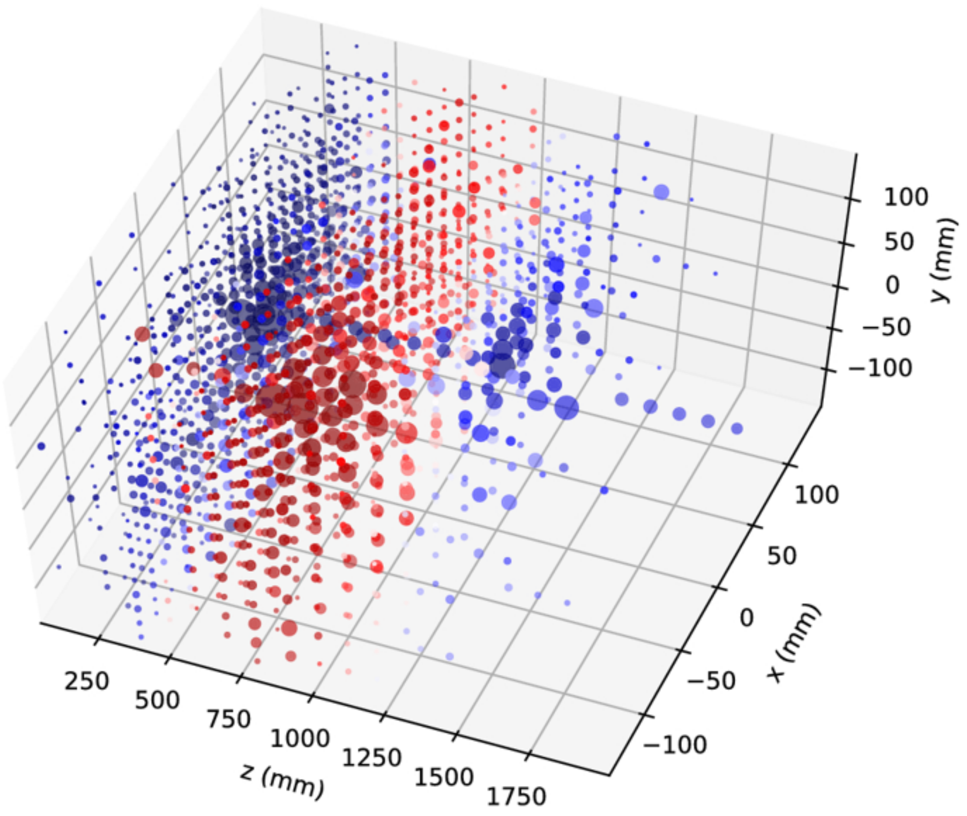
26.01.2021
Virtual

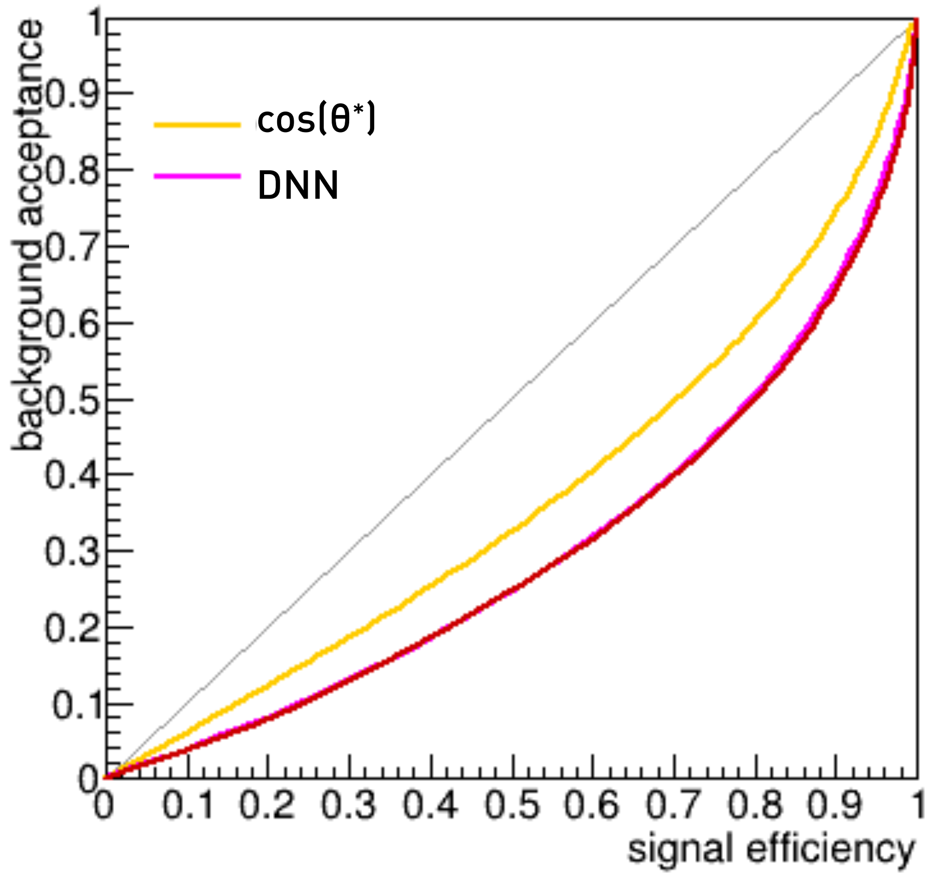
ML for VBS

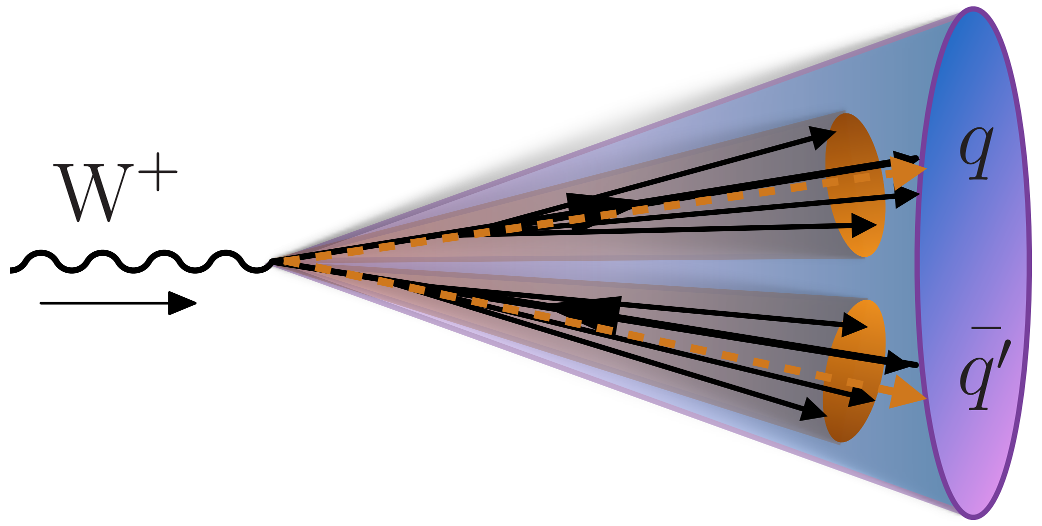
VBS in HL-LHC (and Run3!) extremely exciting

- Luminosity (x2 in Run3, x10 in HL-LHC): More data
- Detectors (e.g CMS High-Granularity Endcap Calorimeter): Better data
- Analysis: Squeeze the most out of bigger and better data!









ML for VBS

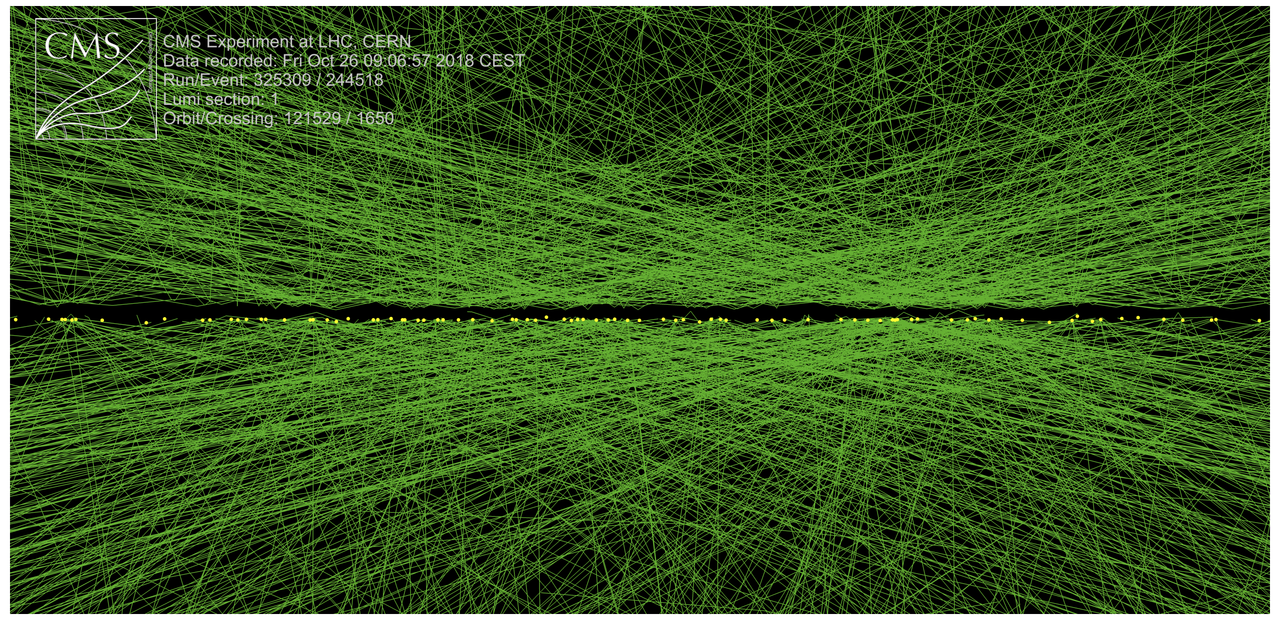
VBS in HL-LHC (and Run3!) extremely exciting

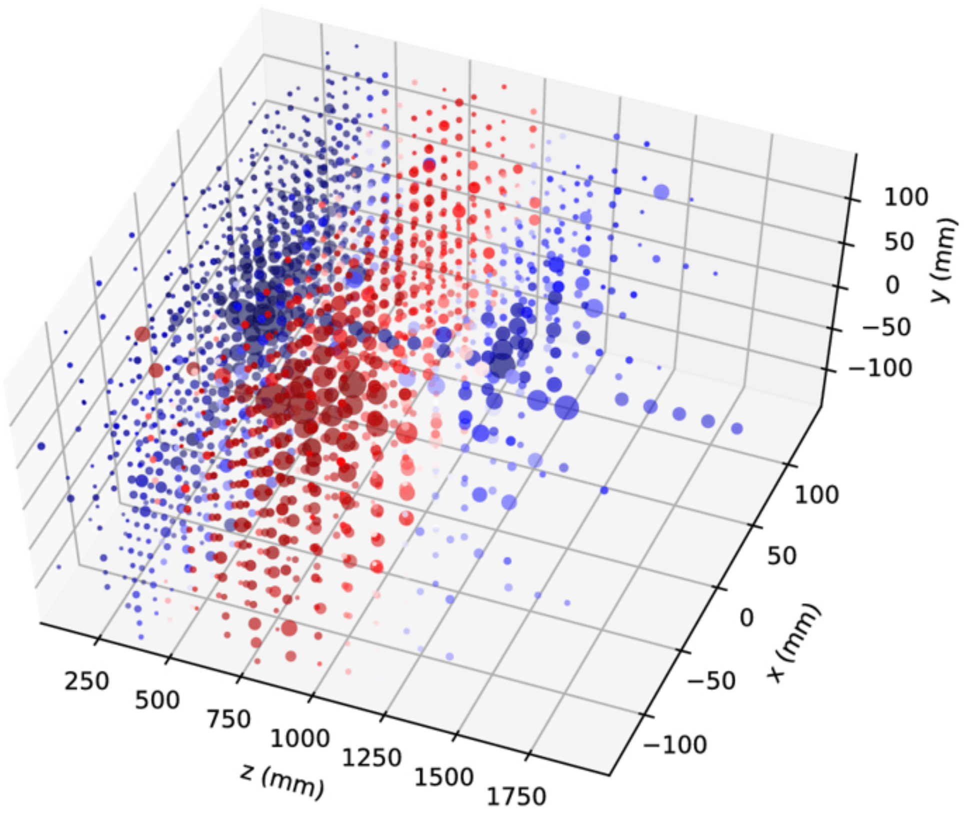
- Luminosity (x2 in Run3, x10 in HL-LHC): More data
- Detectors (e.g CMS High-Granularity Endcap Calorimeter): Better data
- Analysis: Squeeze the most out of bigger and better data!

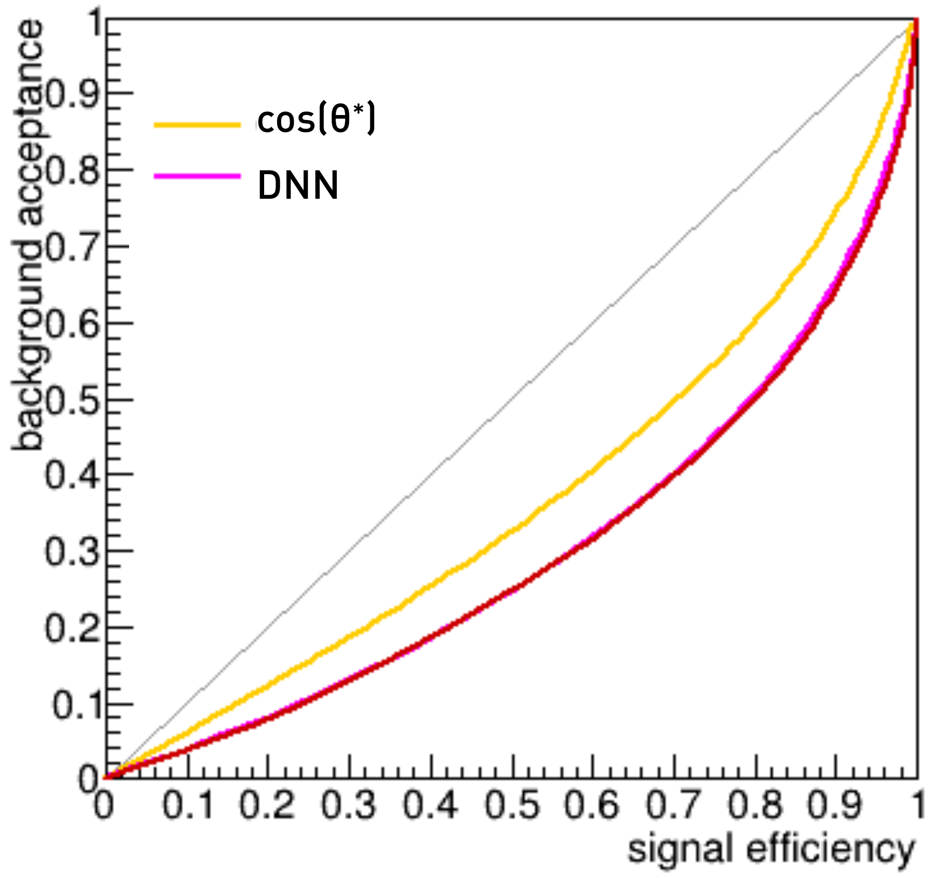
ML crucial part of all three

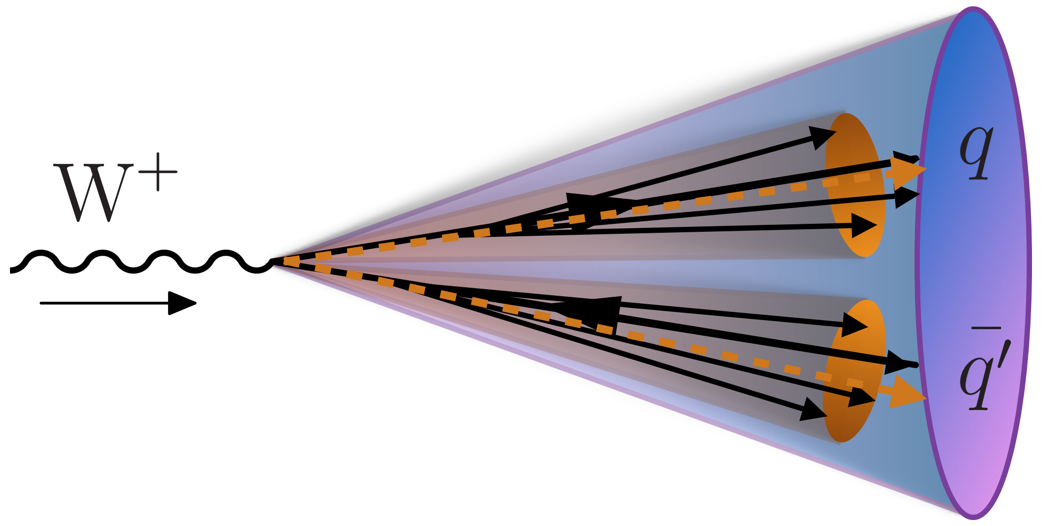
- Luminosity: Managing huge data rates
- Detectors: Reconstructing complex patterns in complex new detectors
- Analysis: Enhancing $S/\sqrt{B}$ to tackle highest-background channels

- Can we dream about all-hadronic VBS? Longitudinal/transverse?

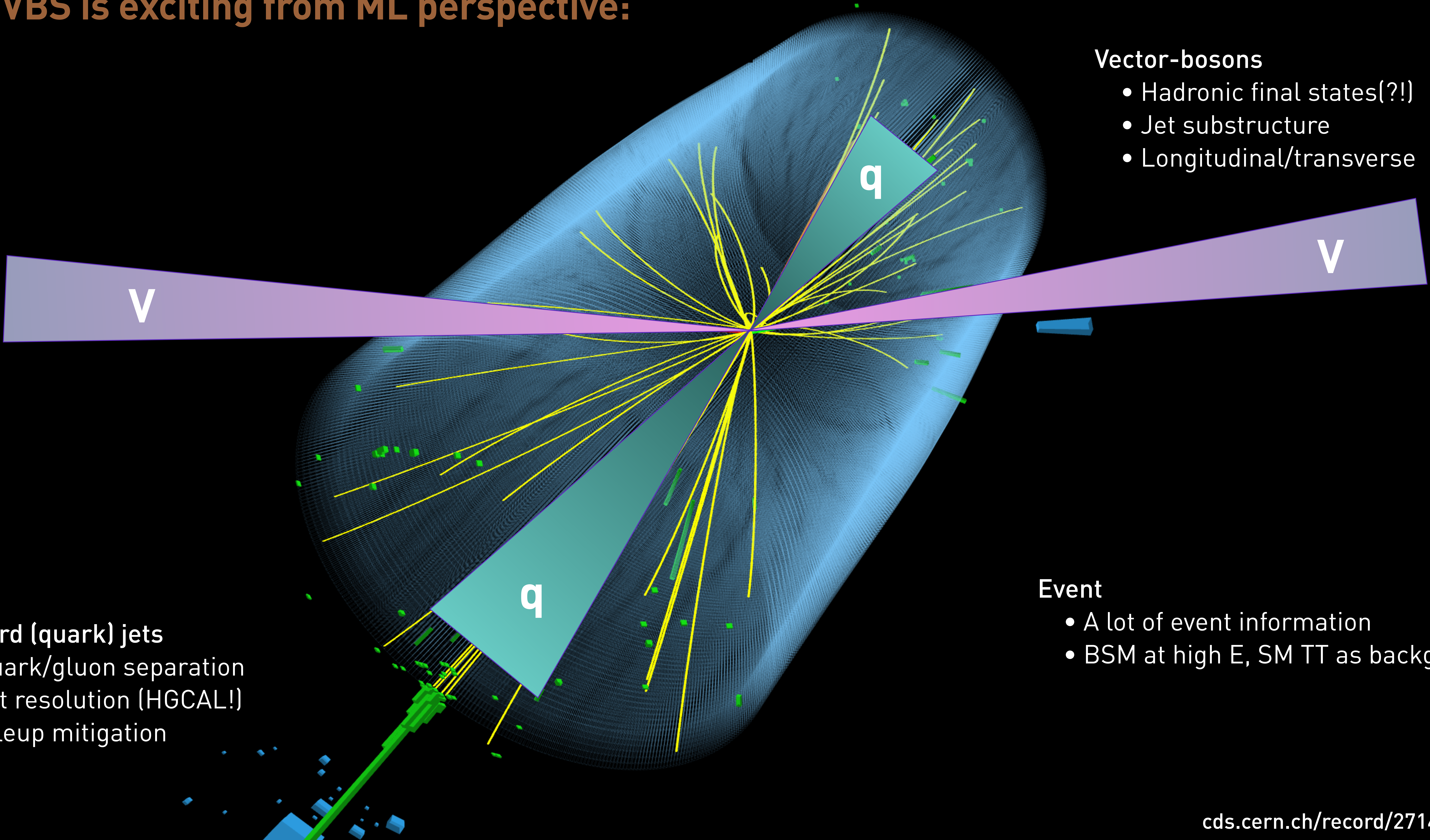








Why VBS is exciting from ML perspective:



Vector-bosons
- Hadronic final states(?!)
- Jet substructure
- Longitudinal/transverse

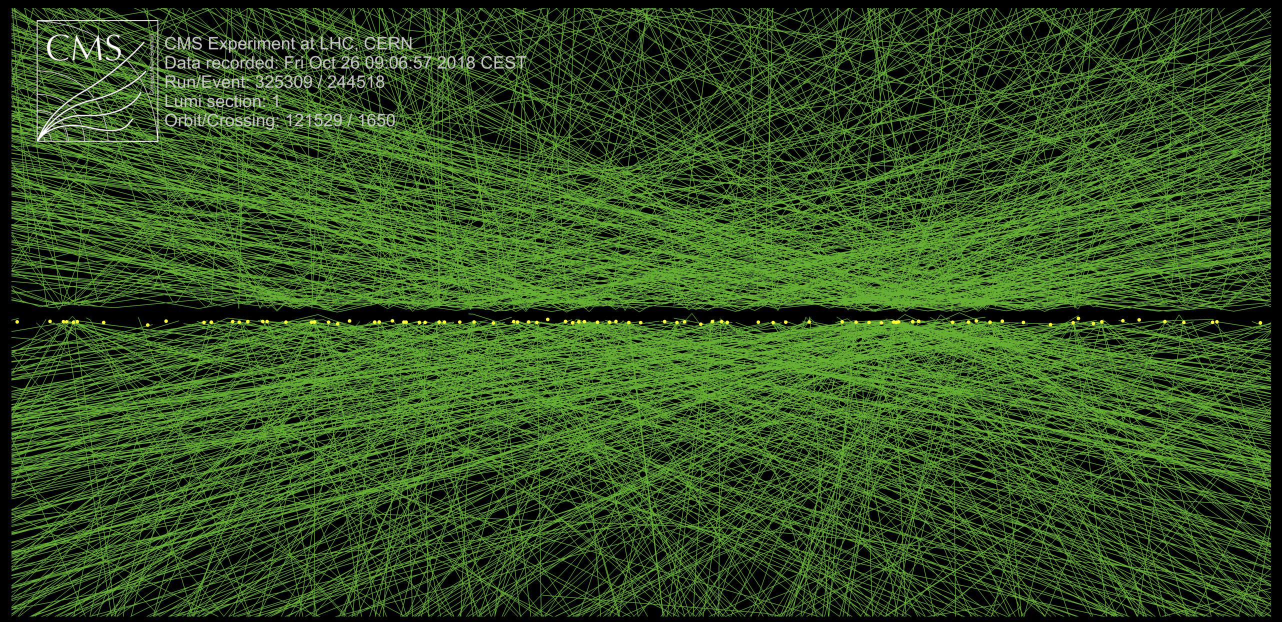
Event
- A lot of event information
- BSM at high E, SM TT as background?

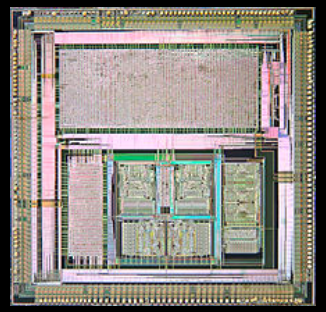
Forward (quark) jets
- Quark/gluon separation
- Jet resolution (HGCAL!)
- Pileup mitigation

cds.cern.ch/record/2714080

Detector

- Data rate **O(100)Tb/s**
- **O(10) ns** latency
- Better endcap calorimetry





Forward jet resolution

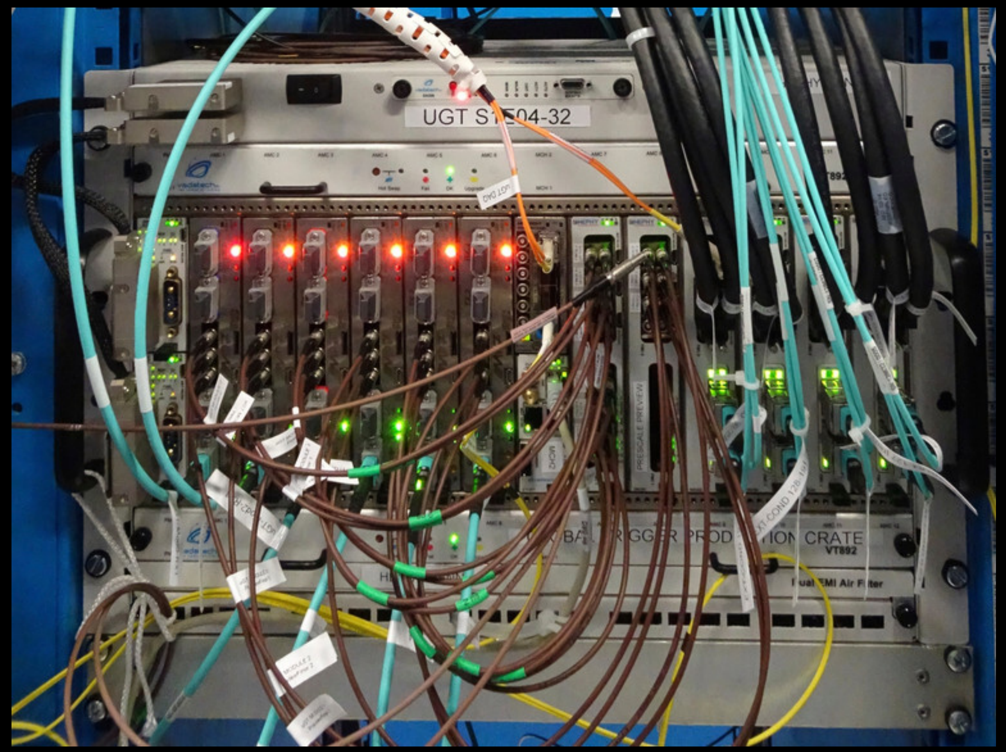
Level-1 hardware trigger

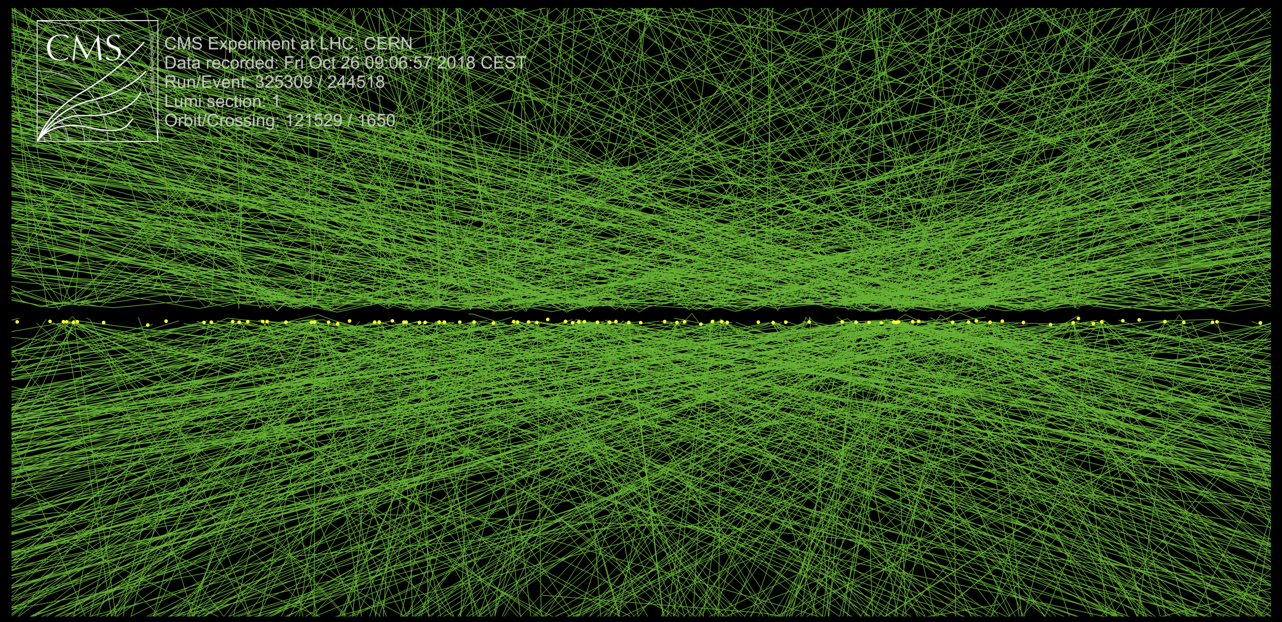
- **12.5 μs** latency
- Input bandwidth **54 Tb/s**
- Tigger cleanly+reconstruct narrow VBS jets!

Detector

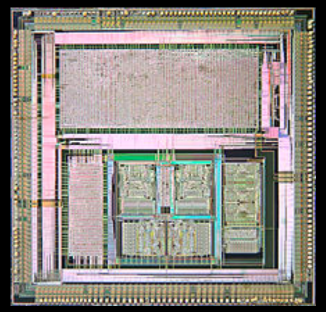
- Data rate **O(100)Tb/s**
- **O(10) ns** latency
- Better endcap calorimetry

40 MHz





VBS trigger



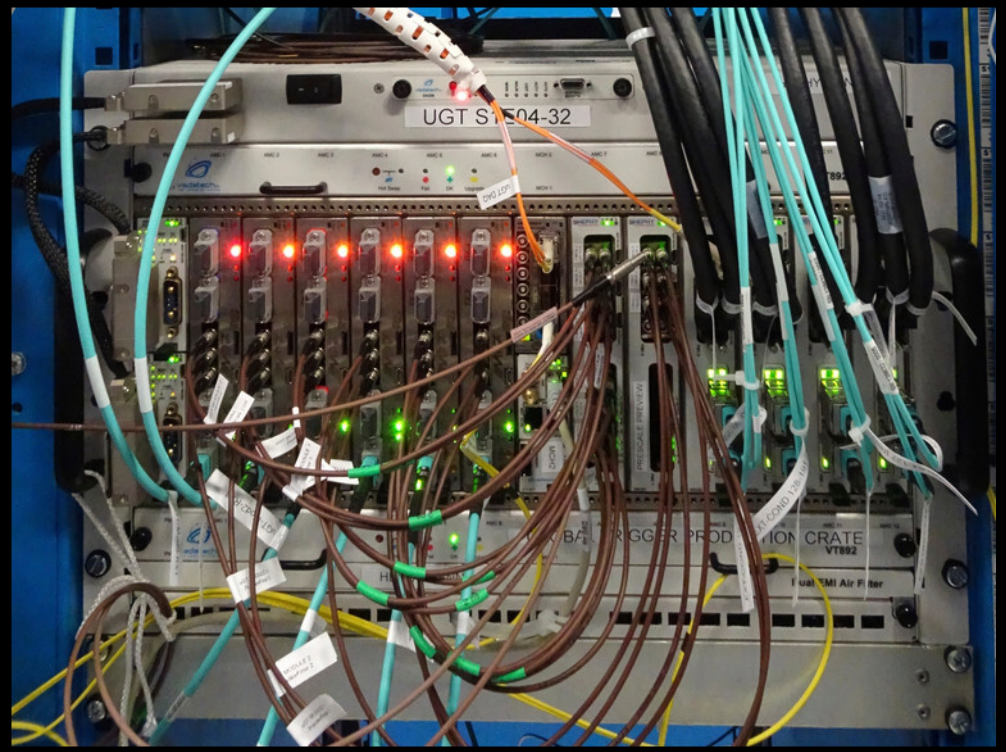
Forward jet resolution

Level-1 hardware trigger
- **12.5 μs** latency
- Input bandwidth **54 Tb/s**
- Tigger cleanly+reconstruct narrow VBS jets!

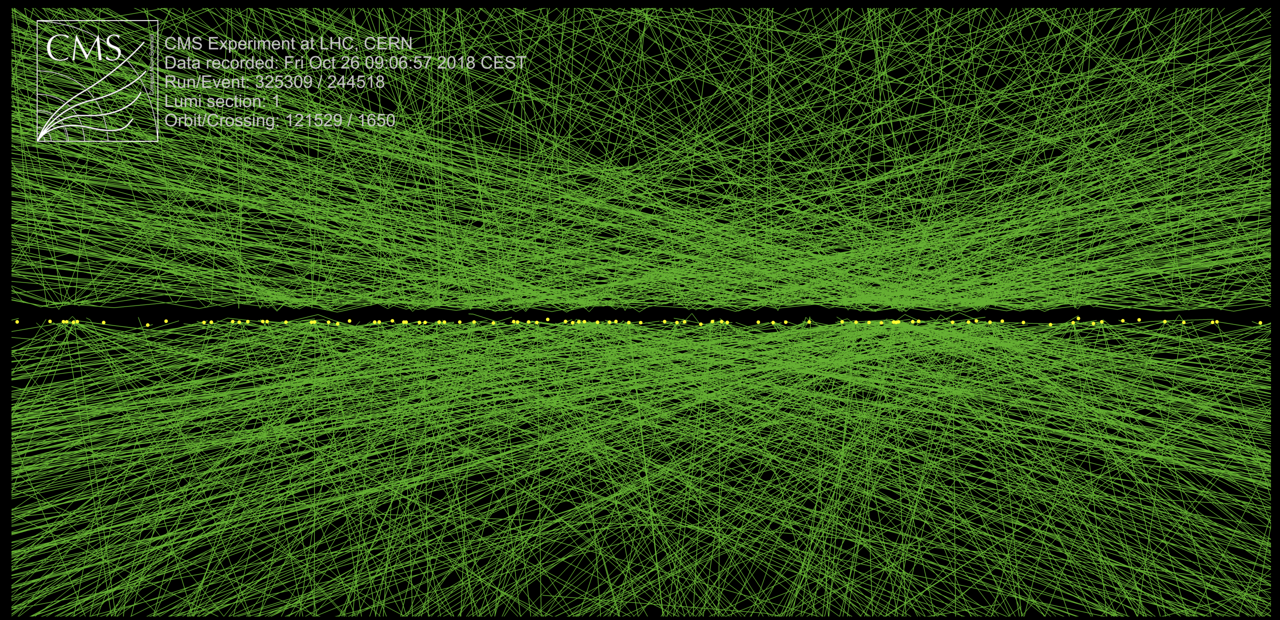
High Level Trigger
- O(100) latency
- q/g discrimination!

Detector
- Data rate **O(100)Tb/s**
- **O(10) ns** latency
- Better endcap calorimetry

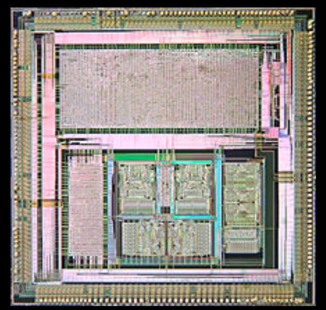
40 MHz



750 kHz



Dedicated ML-based VBS trigger paths

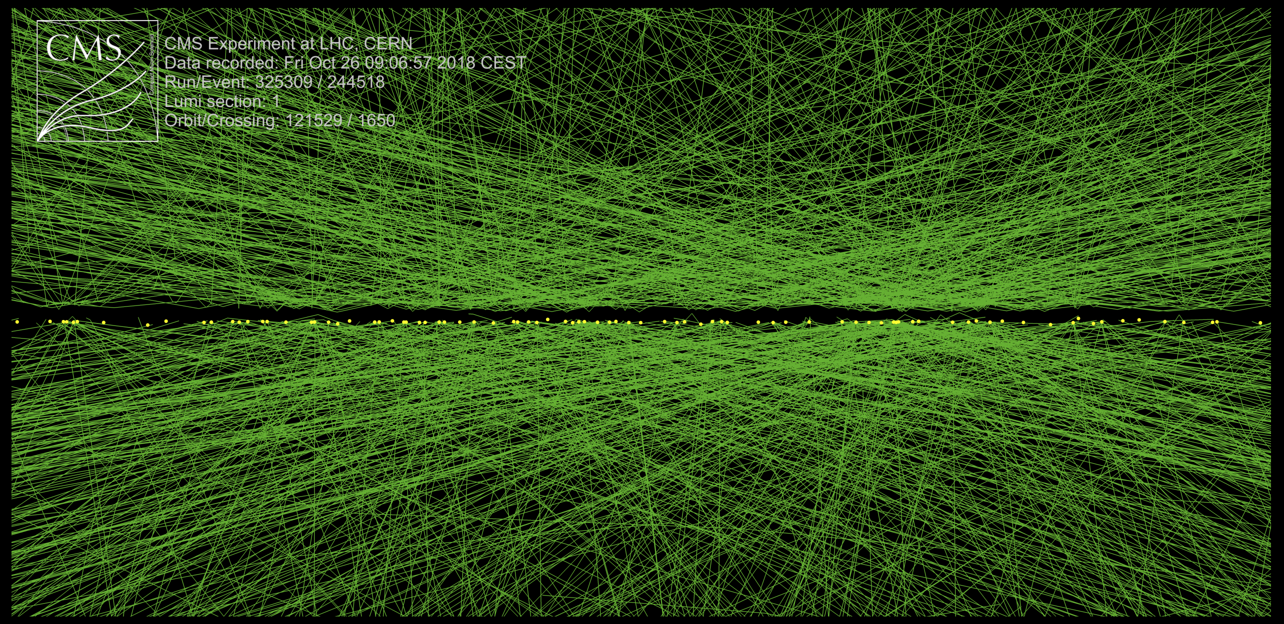


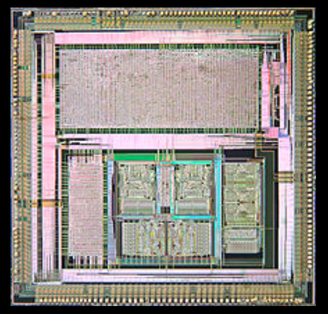
Forward jet resolution

VBS trigger

Detector

- Data rate **O(100)Tb/s**
- **O(10) ns** latency
- Better endcap calorimetry



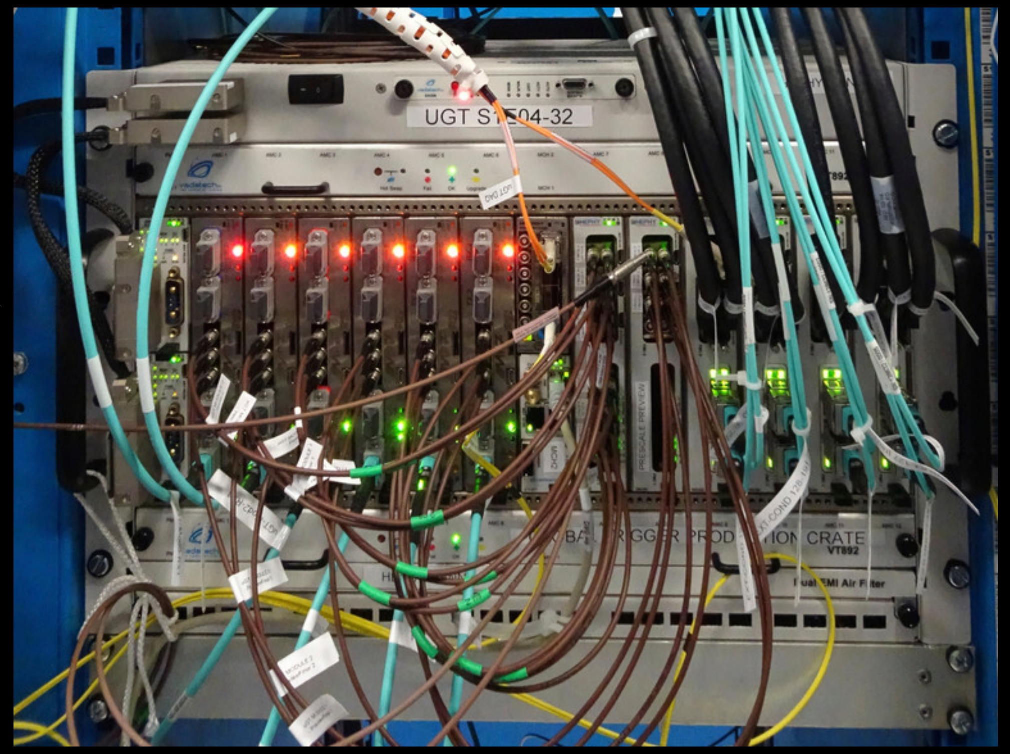


Forward jet resolution

40 MHz

Level-1 hardware trigger

- **12.5 μs** latency
- Input bandwidth **54 Tb/s**
- Tigger cleanly+reconstruct narrow VBS jets!



VBS trigger

750 kHz

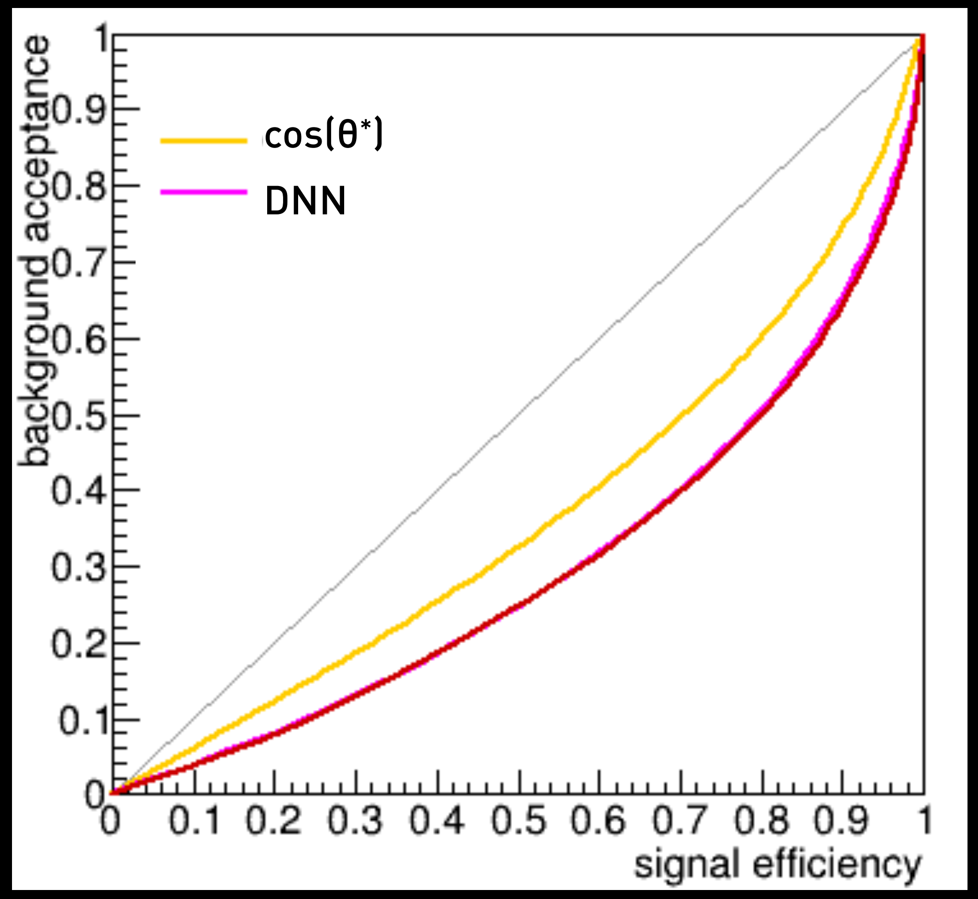
High Level Trigger

- O(100) latency
- q/g discrimination!

Dedicated ML-based VBS trigger paths

7.5 kHz

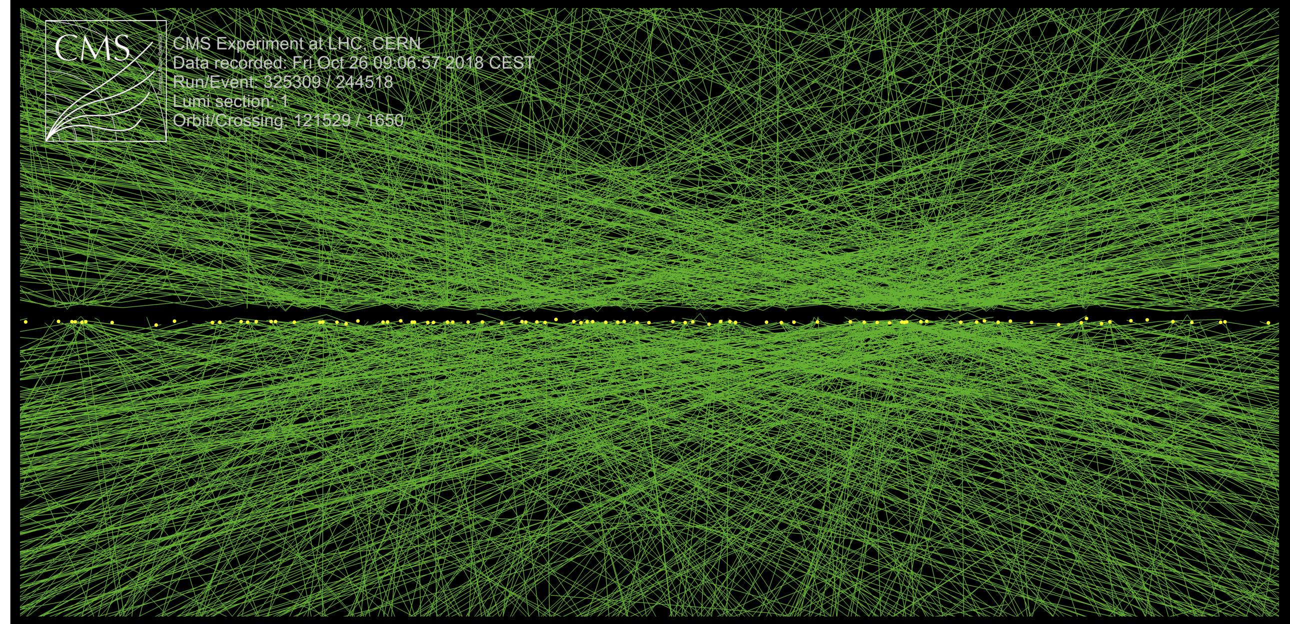
Offline reconstruction and analysis



Longitudinal vs. transverse?
Enhance BSM?

Detector

- Improve forward jet resolution and reducing pileup
- Limited bandwidth and time
- High radiation

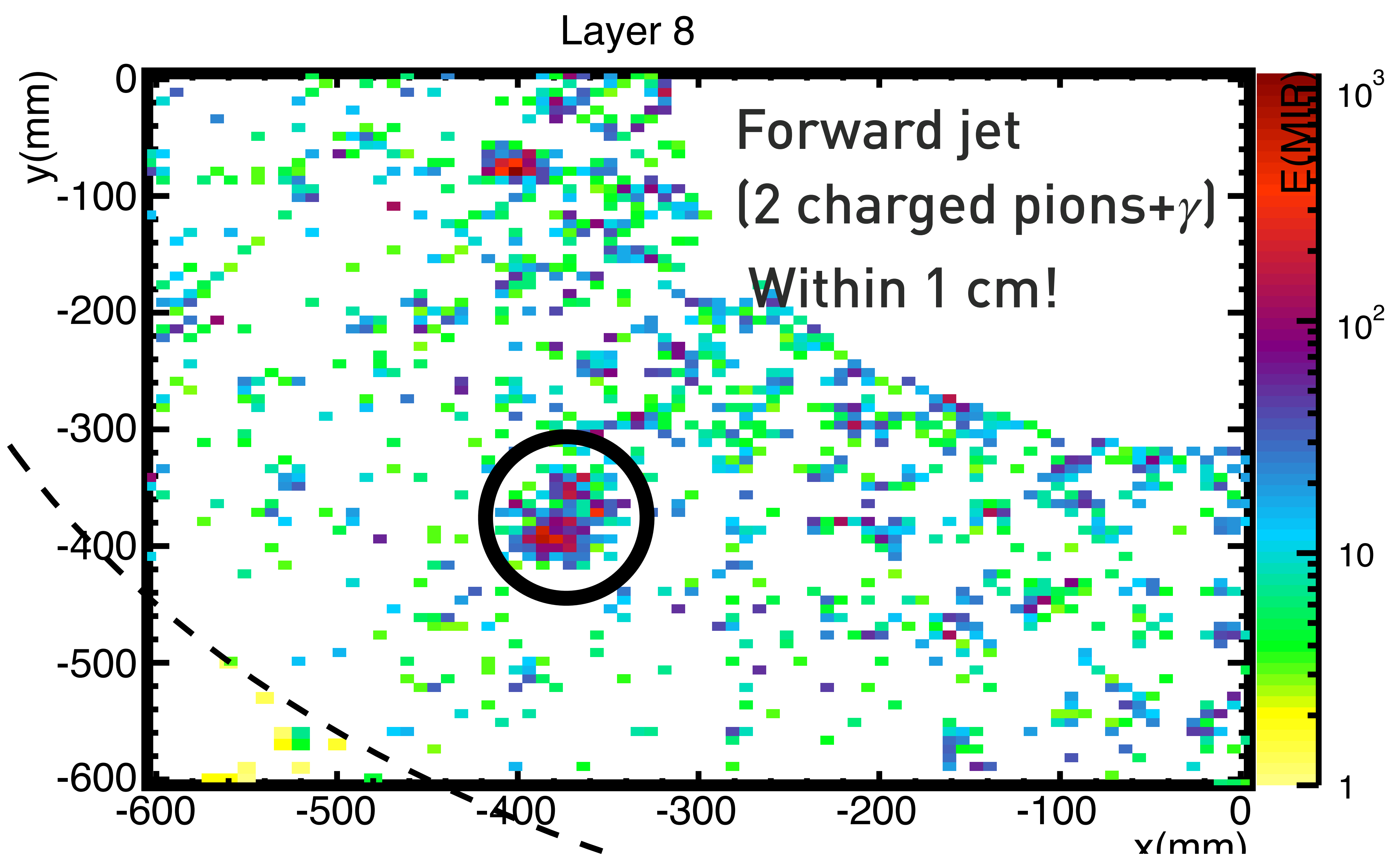


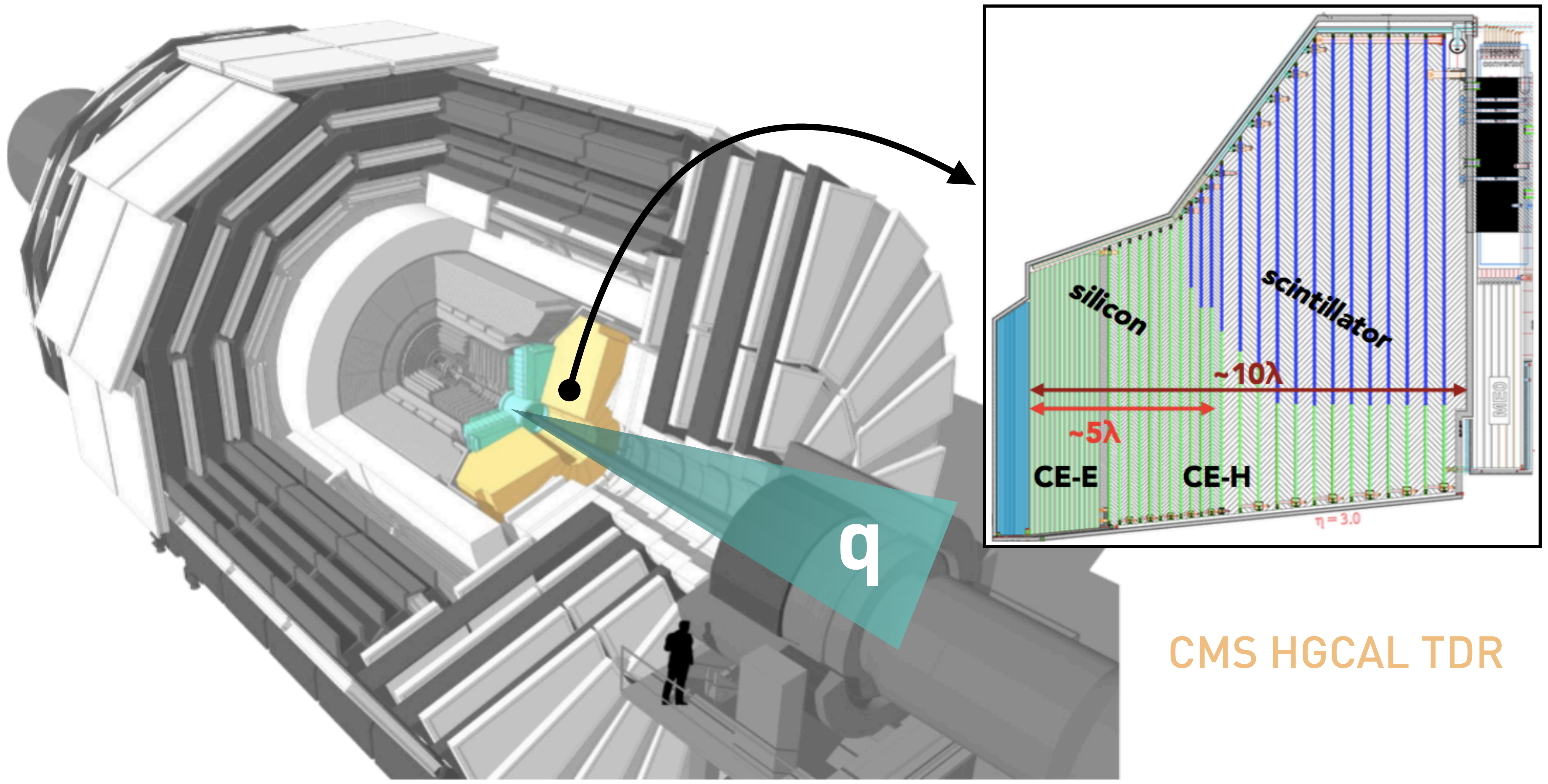
→ ML on ASICs

On-detector: HGCAL

CMS Endcap High-Granularity Calorimeter (1.5<η<3)

- Unprecedented transverse/longitudinal segmentation
- Shower development+narrowness of VBS jets



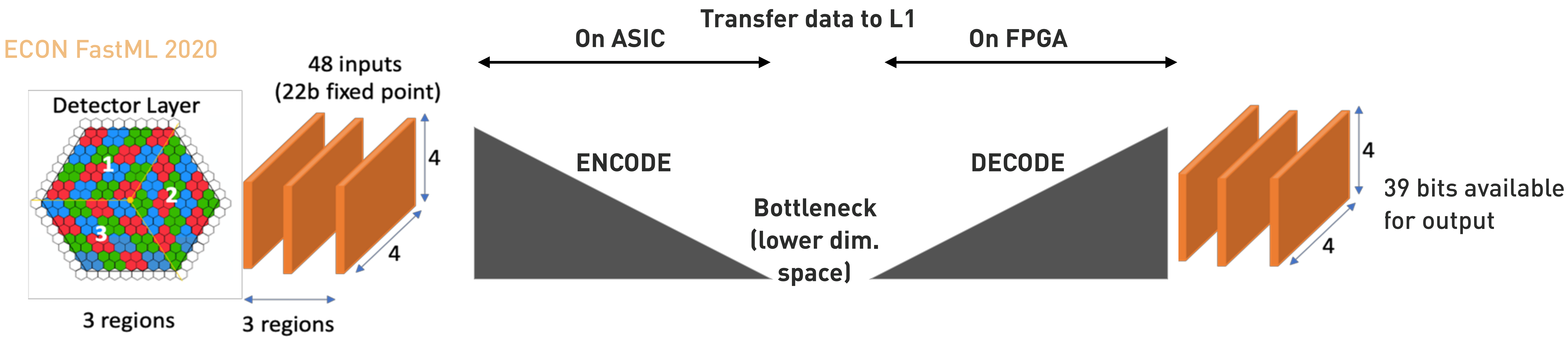


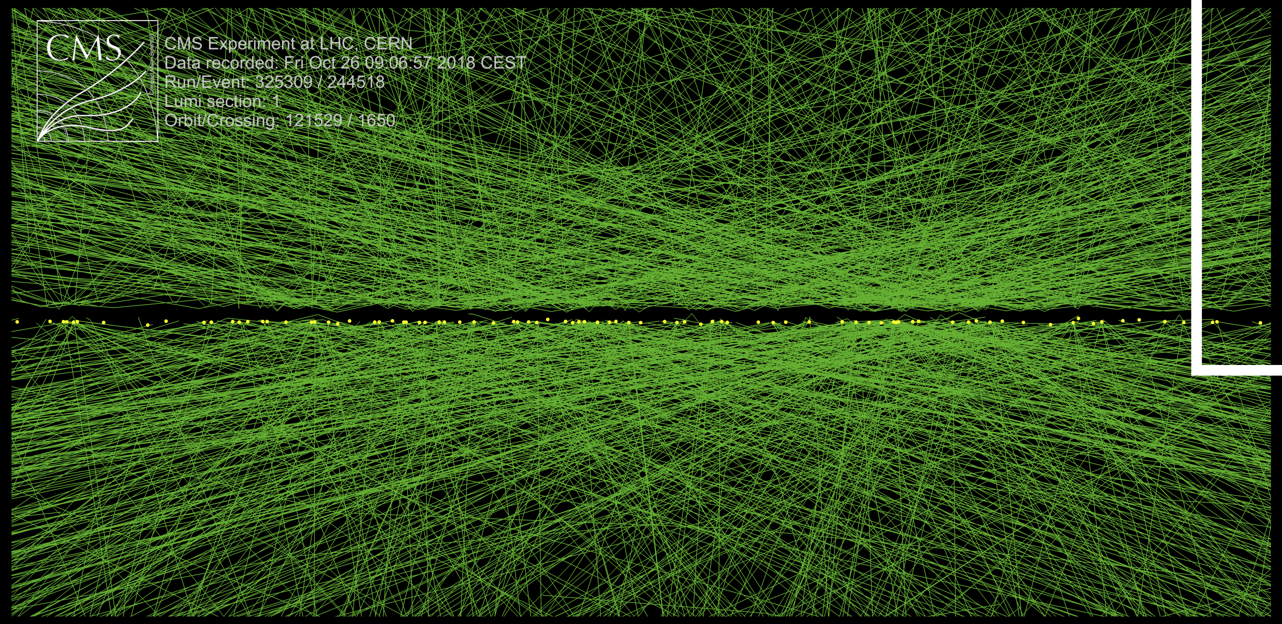
CMS HGCAL TDR

- 52 layers, 6 million silicon channels, limited output bandwidth
- Operate at -30℃ → need low-power on-ASIC preprocessing

On-detector: HGCAL

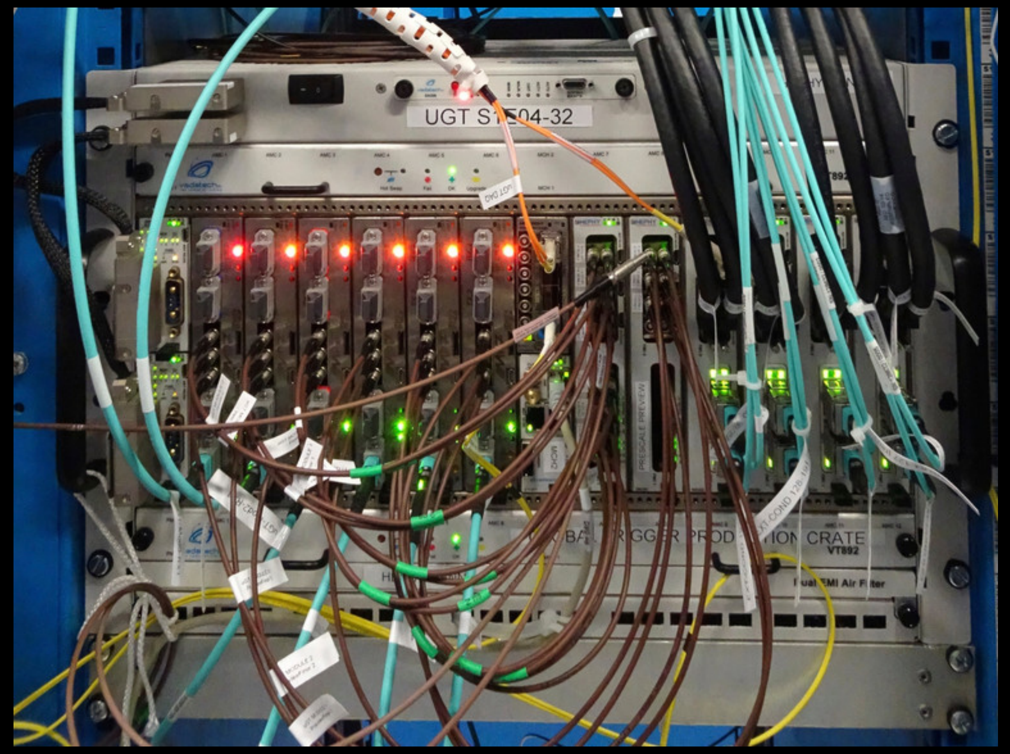
Optimise information output using ML! Maximise resolution on extremely low power.





L1 trigger

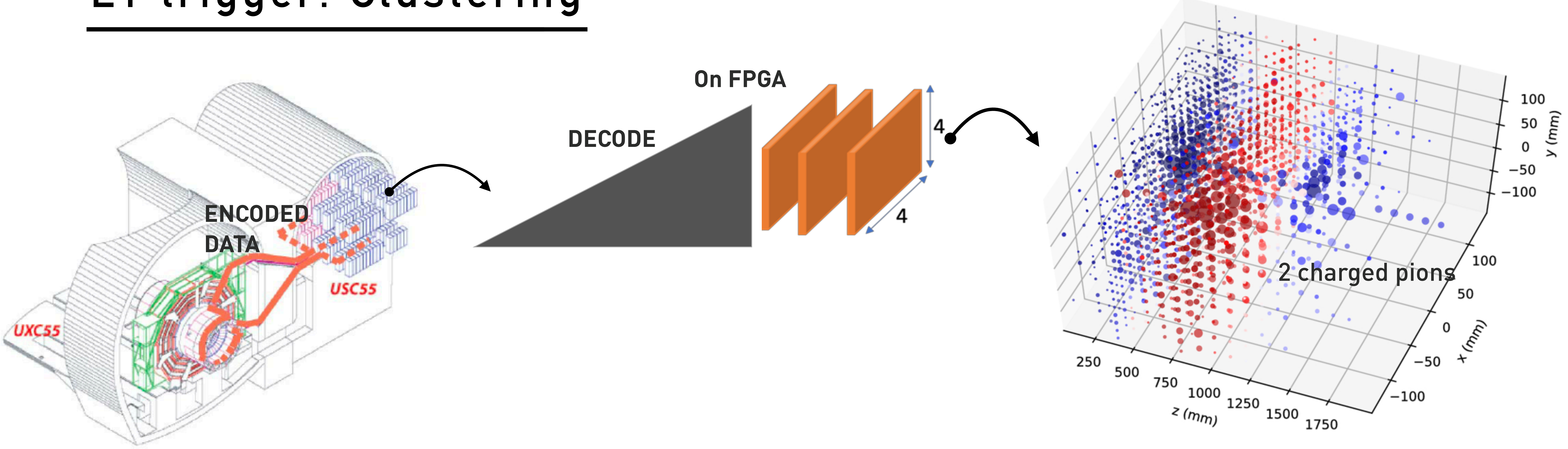
- Maximise VBS signal acceptance through dedicated triggers
- High accuracy + low latency



→ **ML on FPGA**

L1 trigger: Clustering

arxiv:2008.03601
DOI:10.1140



After energy deposits are decoded on FPGFA, need to cluster!

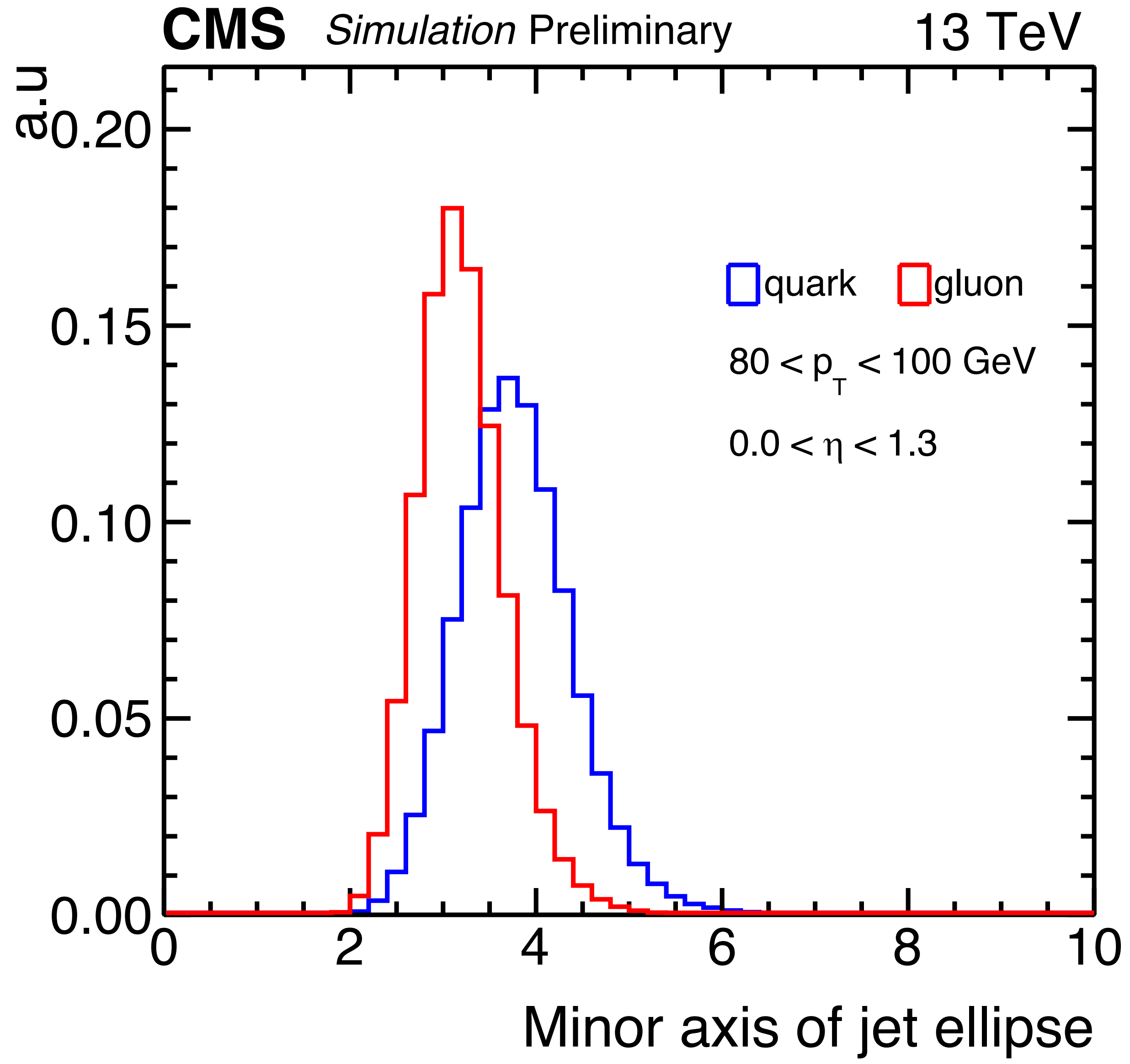
- Exploring Graph Neural Network to cluster energy deposits into disentangled showers from individual particles
- Difficult to achieve desired throughput (huge input!), but has been demonstrated on reduced input sizes

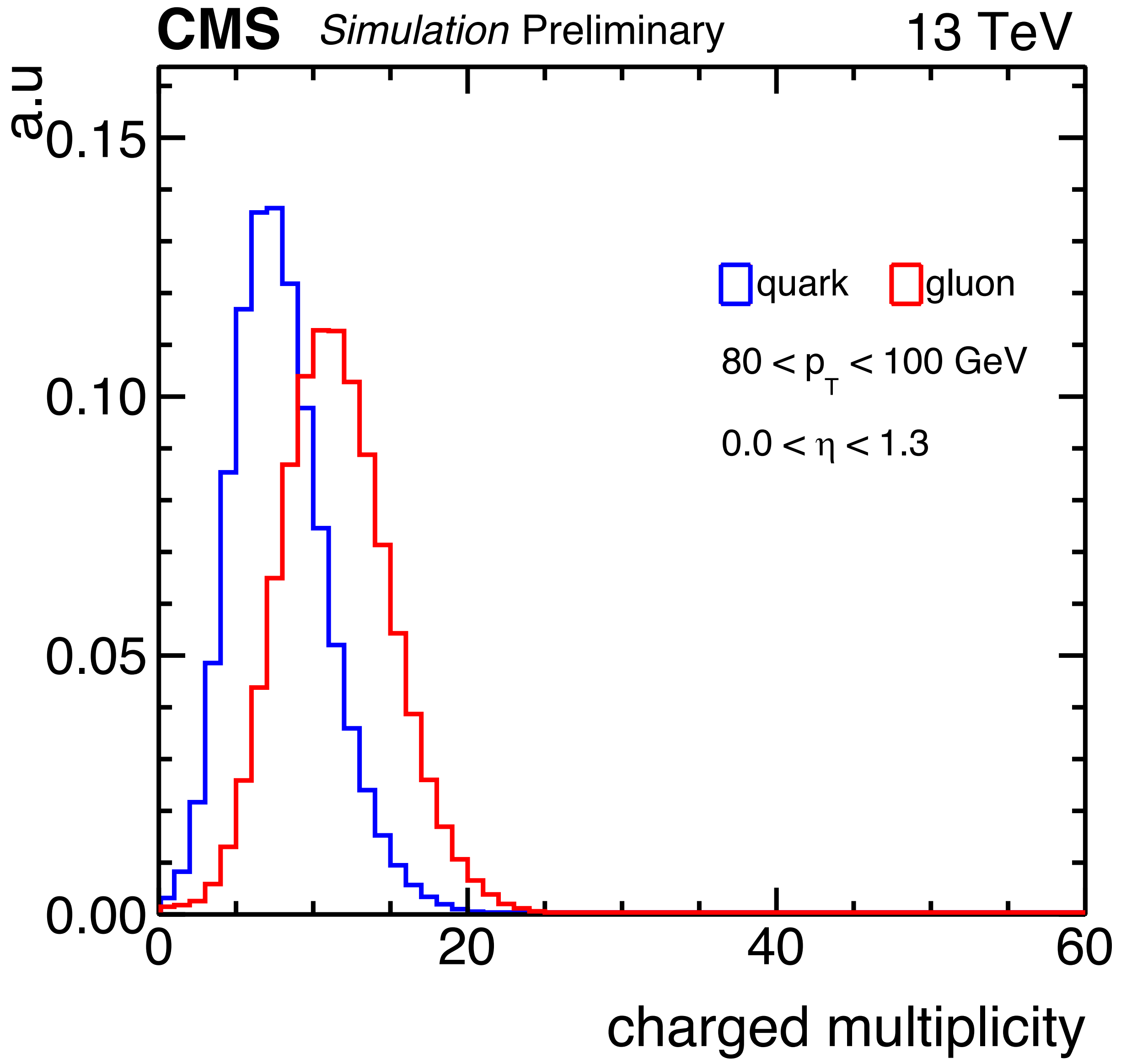
Good energy resolution and clustering important for VBS jets in 200 PU environment!

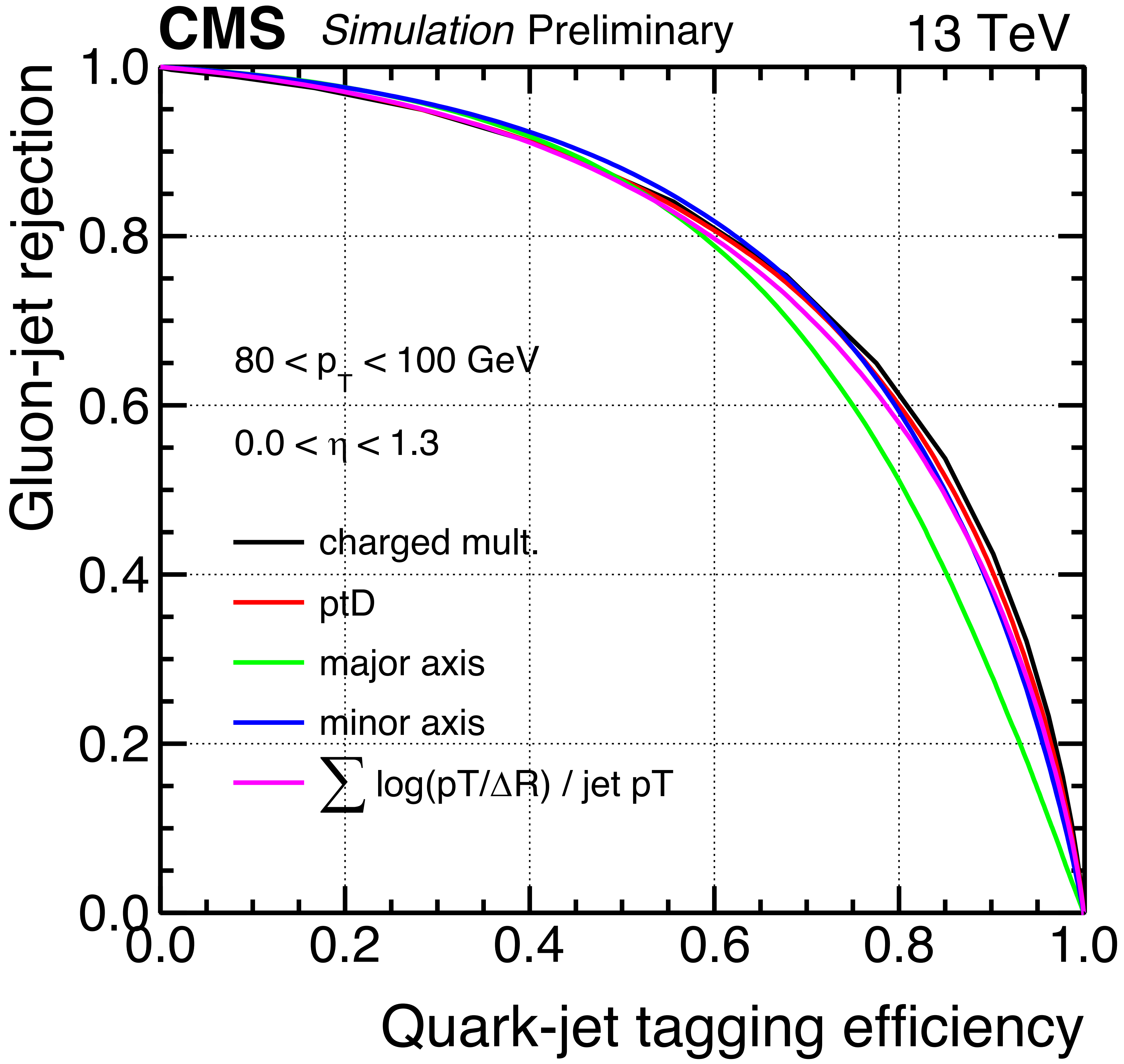
L1 trigger: q/g

With good endcap shower reconstruction:

- "Cheap" ways to get q/g separation at L1, eg. simple 5 input BDT (2017)
- High-resolution inputs + better DNN (Kallonen et. al.), q-tag at L1 @ HL-LHC



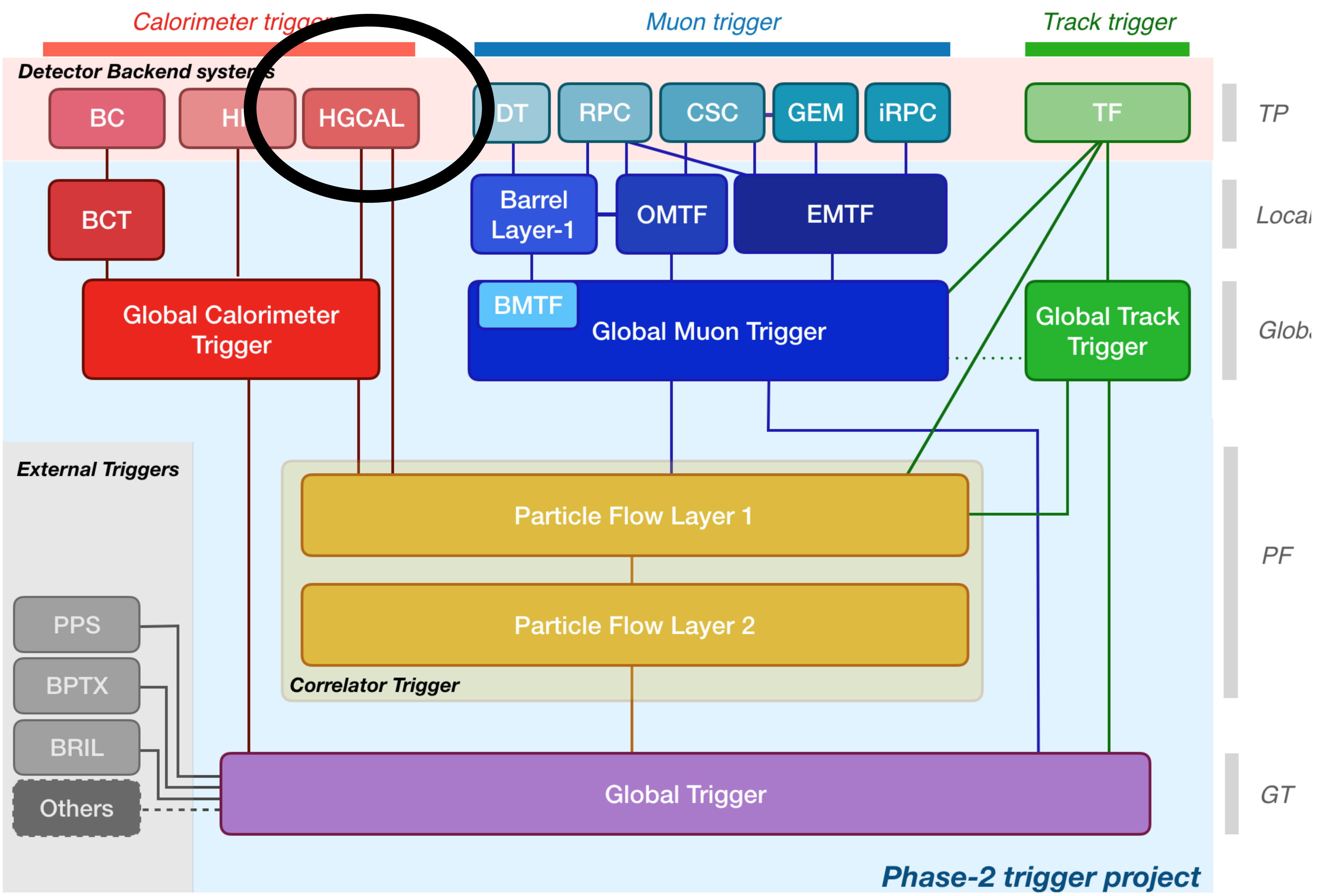




L1 trigger: Global tag

CMS Level-1 in HL-LHC has a lot to offer VBS

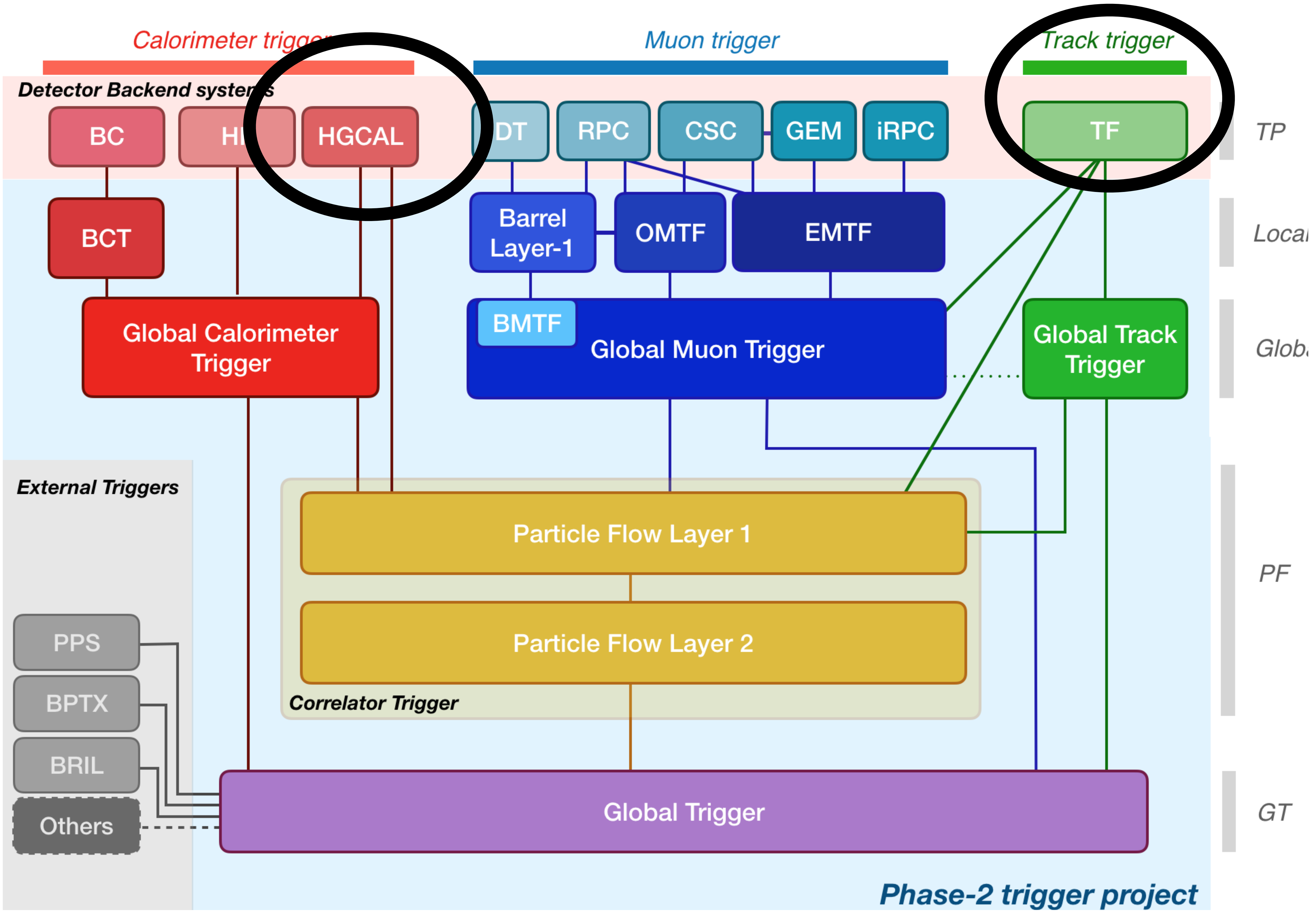
- High resolution forward calorimeter
  → lower $m_{q\bar{q}}$ thresholds? q/g-tag?



L1 trigger: Global tag

CMS Level-1 in HL-LHC has a lot to offer VBS

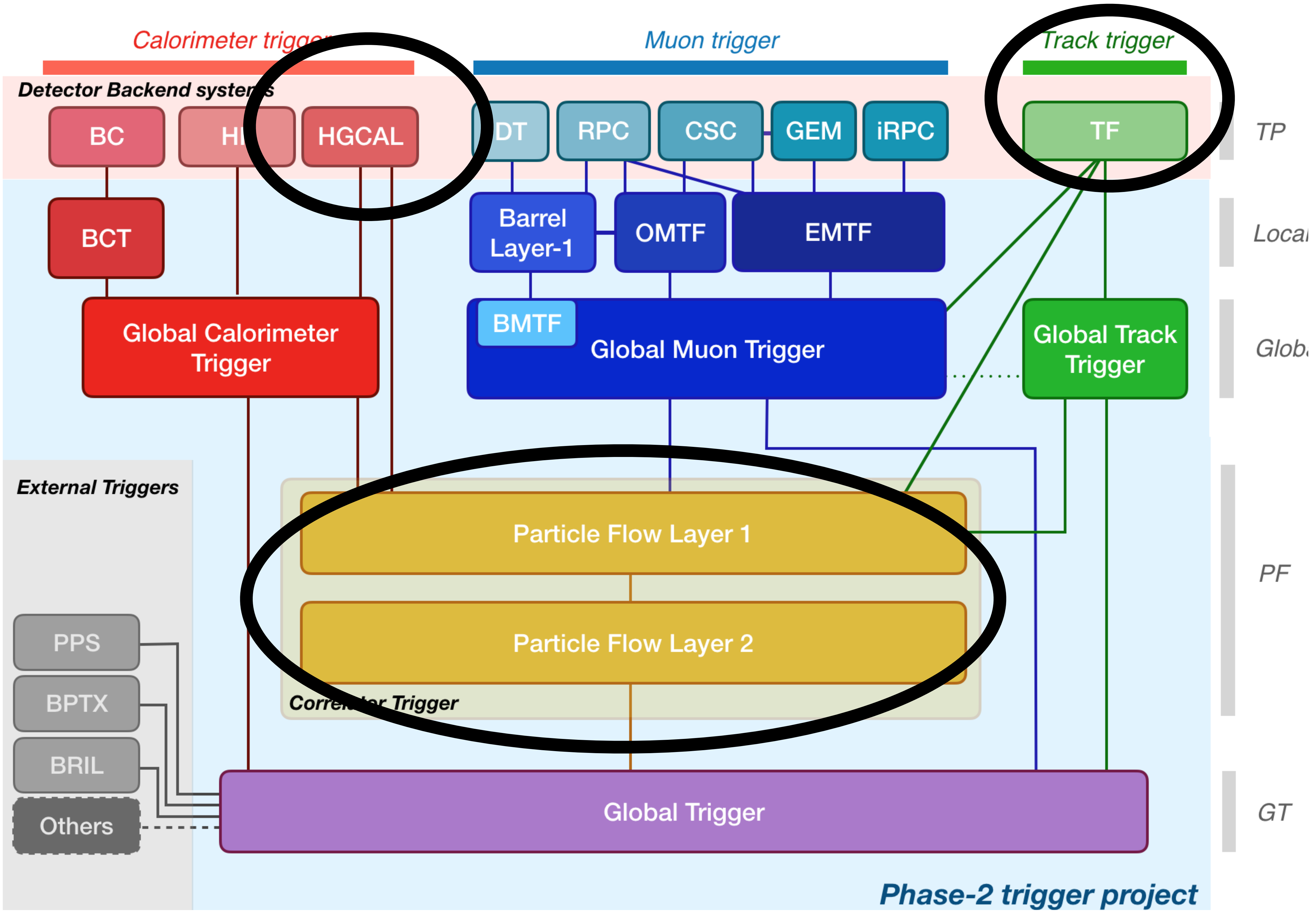
- High resolution forward calorimeter
  → lower $m_{q\bar{q}}$ thresholds? q/g-tag?
- Tracking (η < 2.4)
  → charge multiplicity for q/g ++?



L1 trigger: Global tag

CMS Level-1 in HL-LHC has a lot to offer VBS

- High resolution forward calorimeter
  → lower $m_{q\bar{q}}$ thresholds? q/g-tag?
- Tracking (η < 2.4)
  → charge multiplicity for q/g ++?
- ParticleFlow at Level 1
  → substructure algorithms for $V \to q\bar{q}$ and q/g?



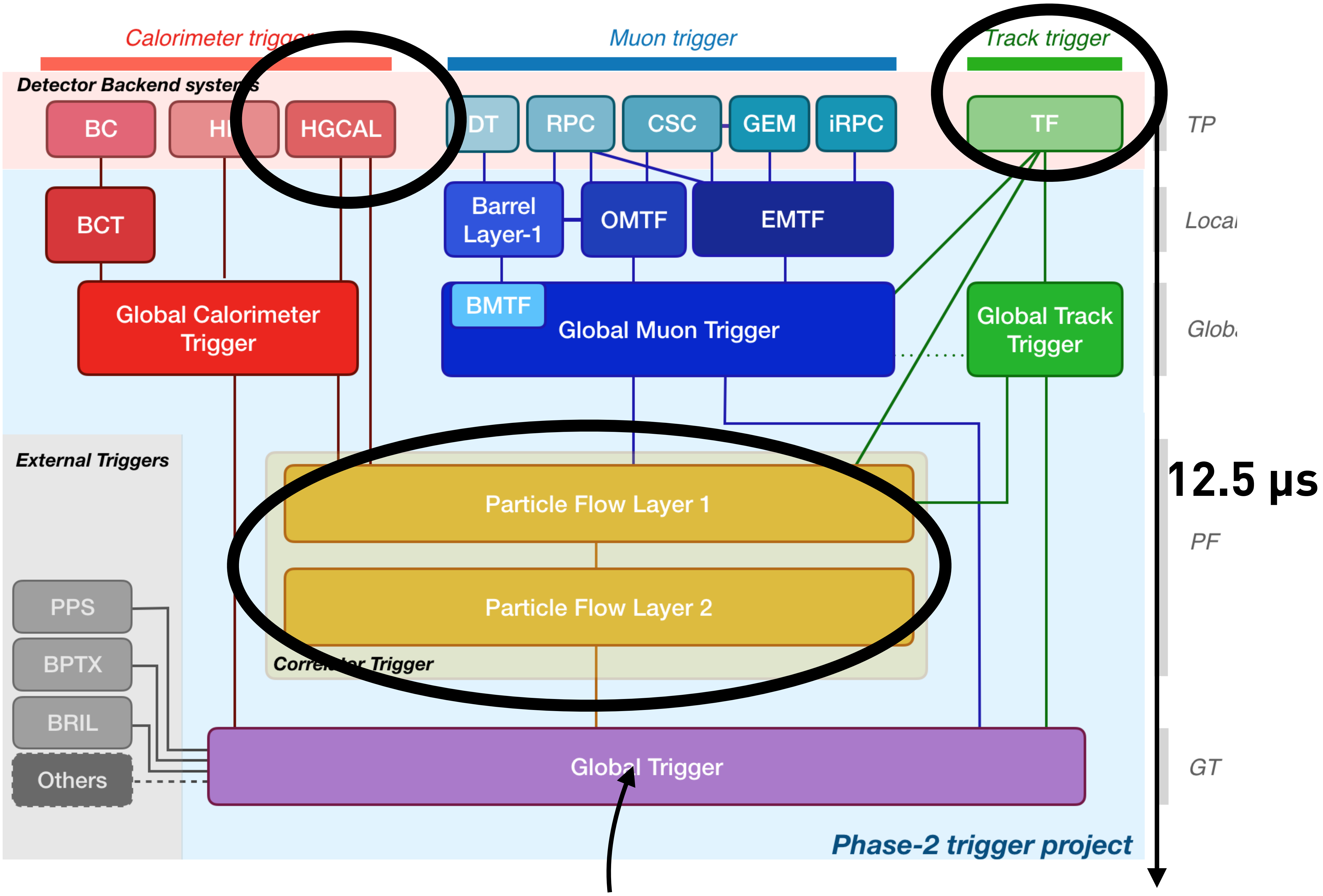
L1 trigger: Global tag

CMS Level-1 in HL-LHC has a lot to offer VBS

- High resolution forward calorimeter
  → lower $m_{q\bar{q}}$ thresholds? q/g-tag?
- Tracking ($\eta < 2.4$)
  → charge multiplicity for q/g ++?
- ParticleFlow at Level 1
  → substructure algorithms for $V \to q\bar{q}$ and q/g?

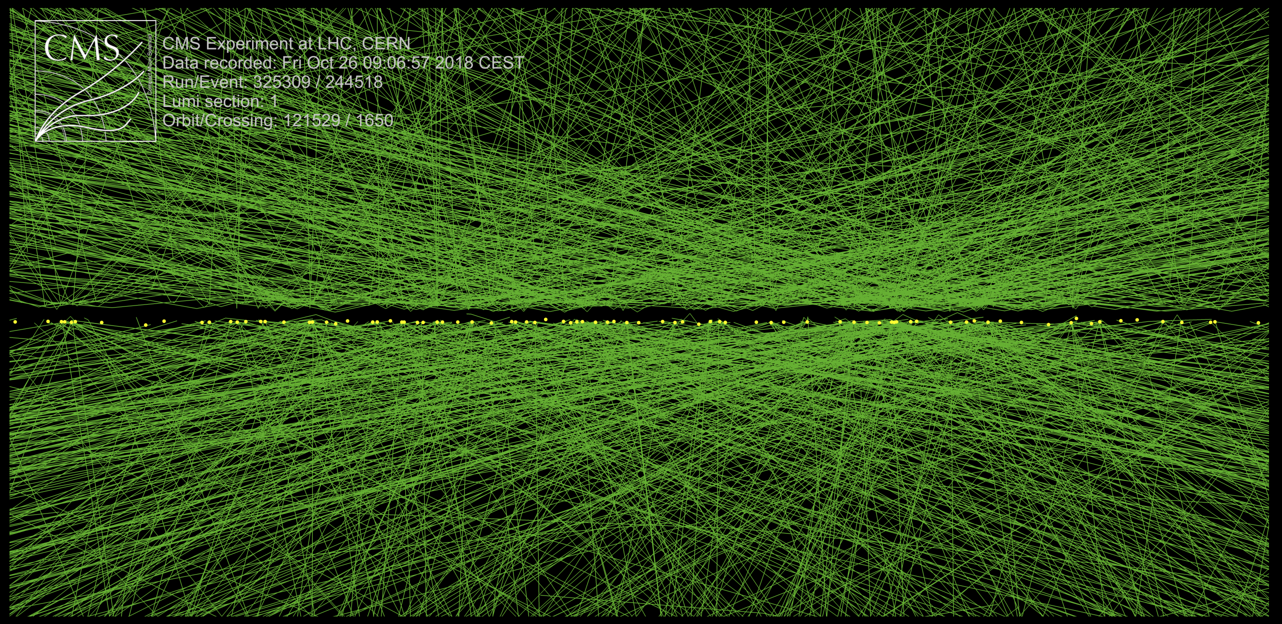
- Study DNN for jet substructure tagging at L1

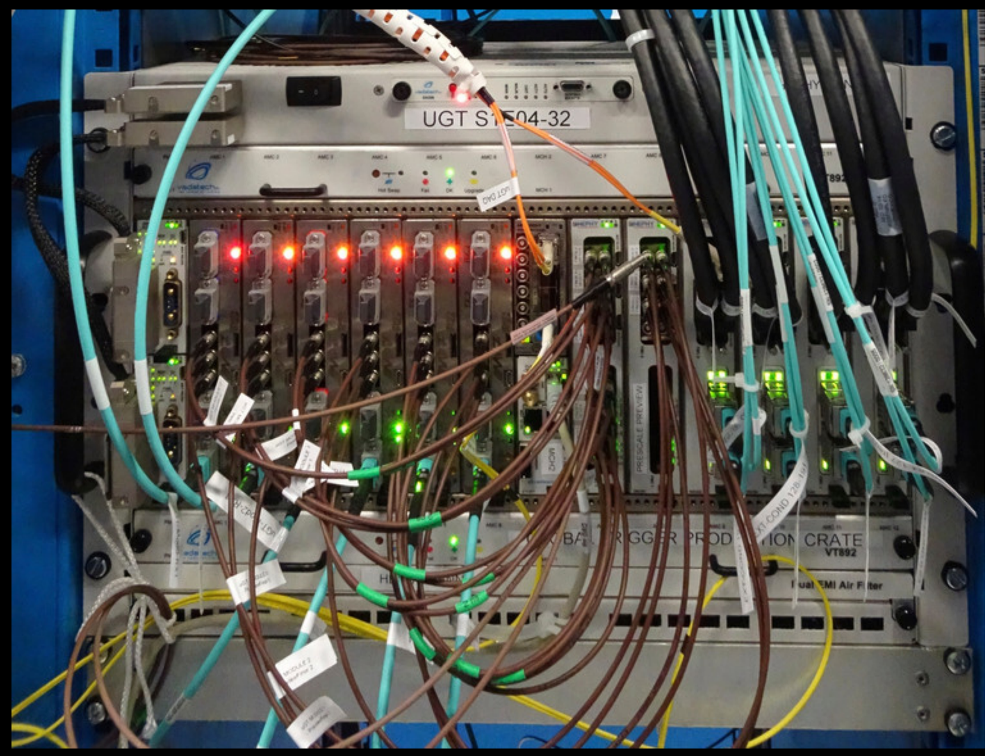
- AI can provide highly efficient VBS tags!
  Boils down to latency, resources and bandwidth



O(1000) algorithms
in parallel on **~10 FPGAs**

Trigger accept/reject





High Level Trigger

- Maximize signal acceptance with all the things we couldn't afford to do at L1 (e.g. jet substructure)
- More time, limited bandwidth

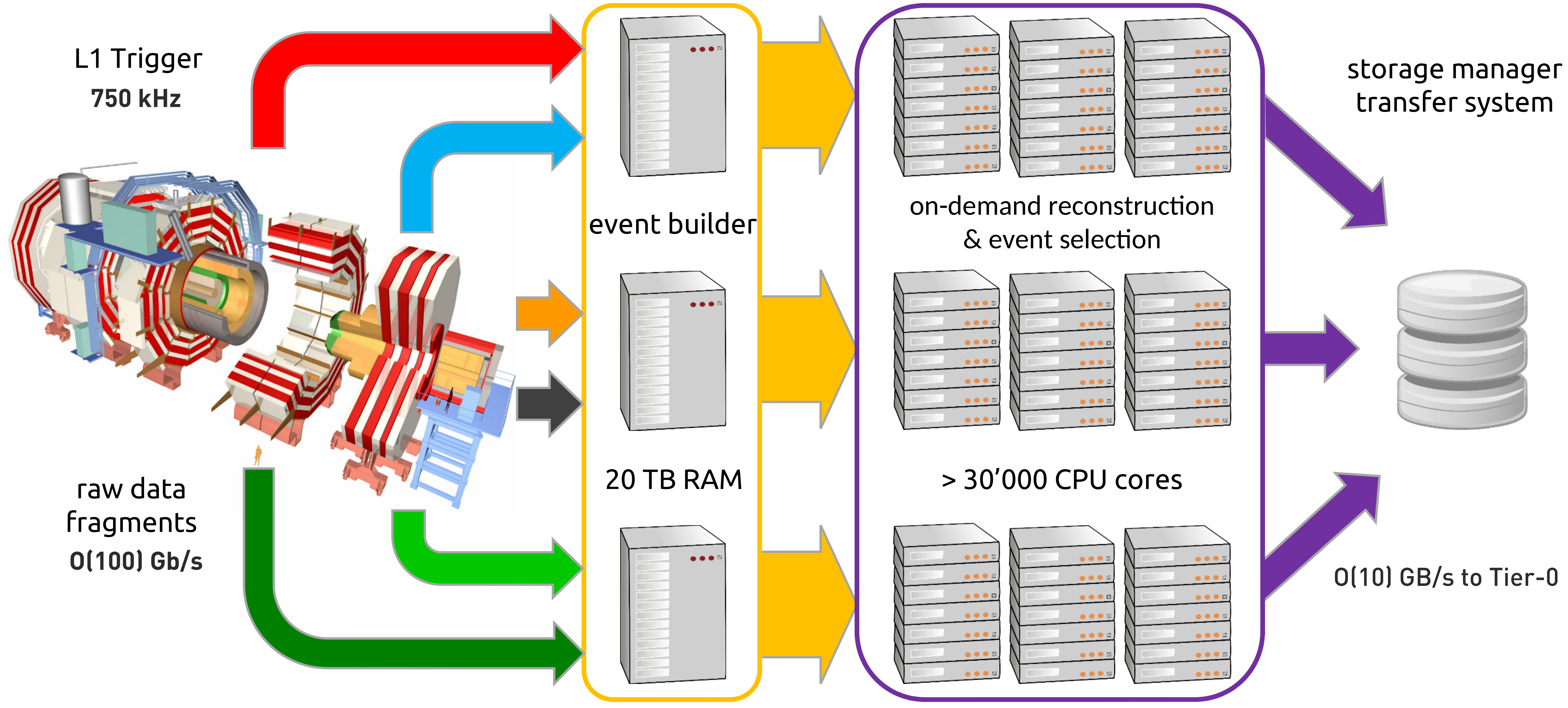
→ **ML on GPU**

High Level Trigger

Amount of data we can store for use in analysis limited by bandwidth, O(10) GB/s to Tier-0

- 300 ms to decide keep/reject
- Running thousands of “modules” on many collision events in parallel

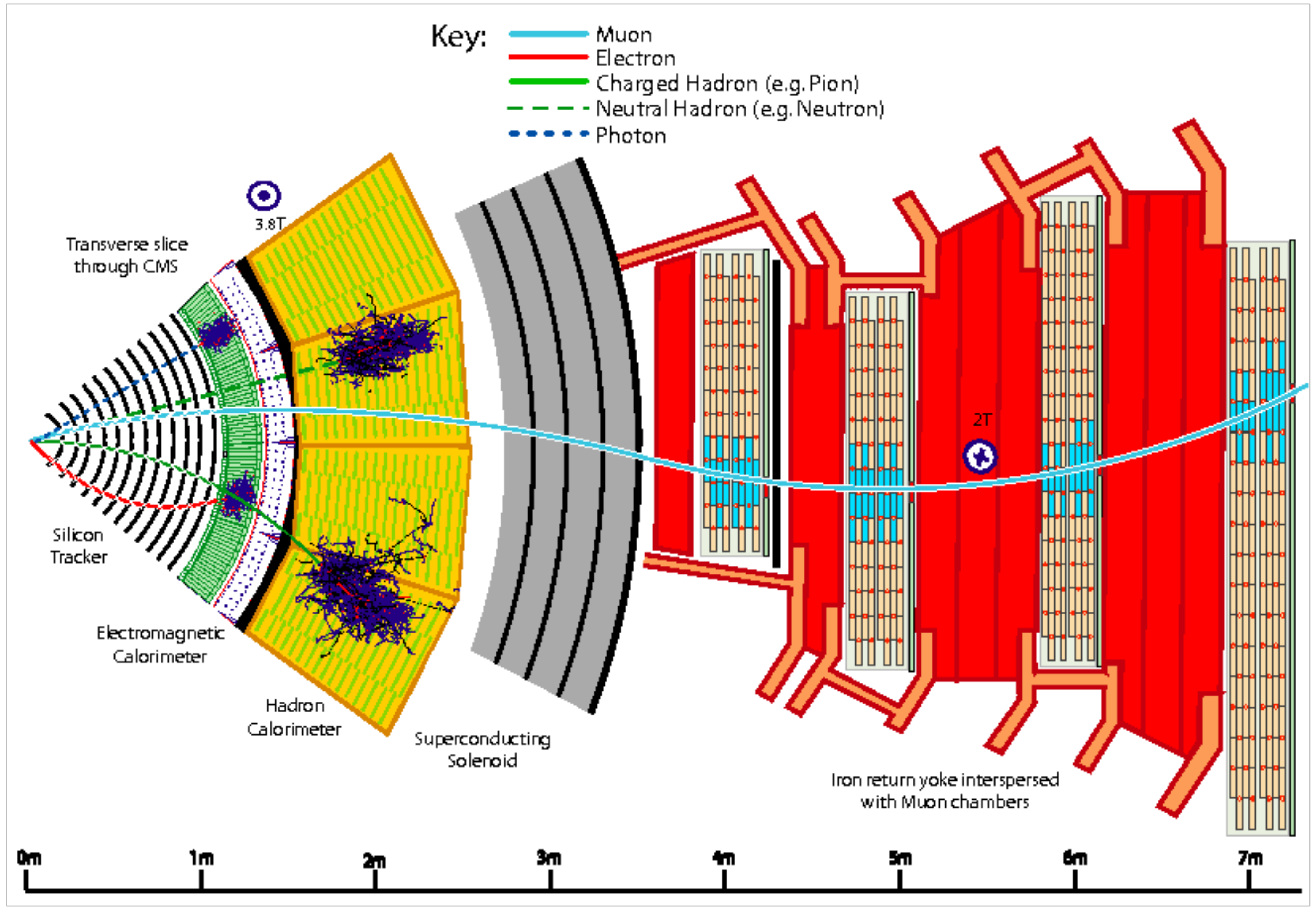
Bandwidth (kB/s) = Event rate (kHz) x Event size (kB)



HLT: More Particle Flow

Particle Flow is highest resolution reconstruction at HLT. Slow, can't run on all events! Currently only PF on 17% of total

- High resolution, but small rate



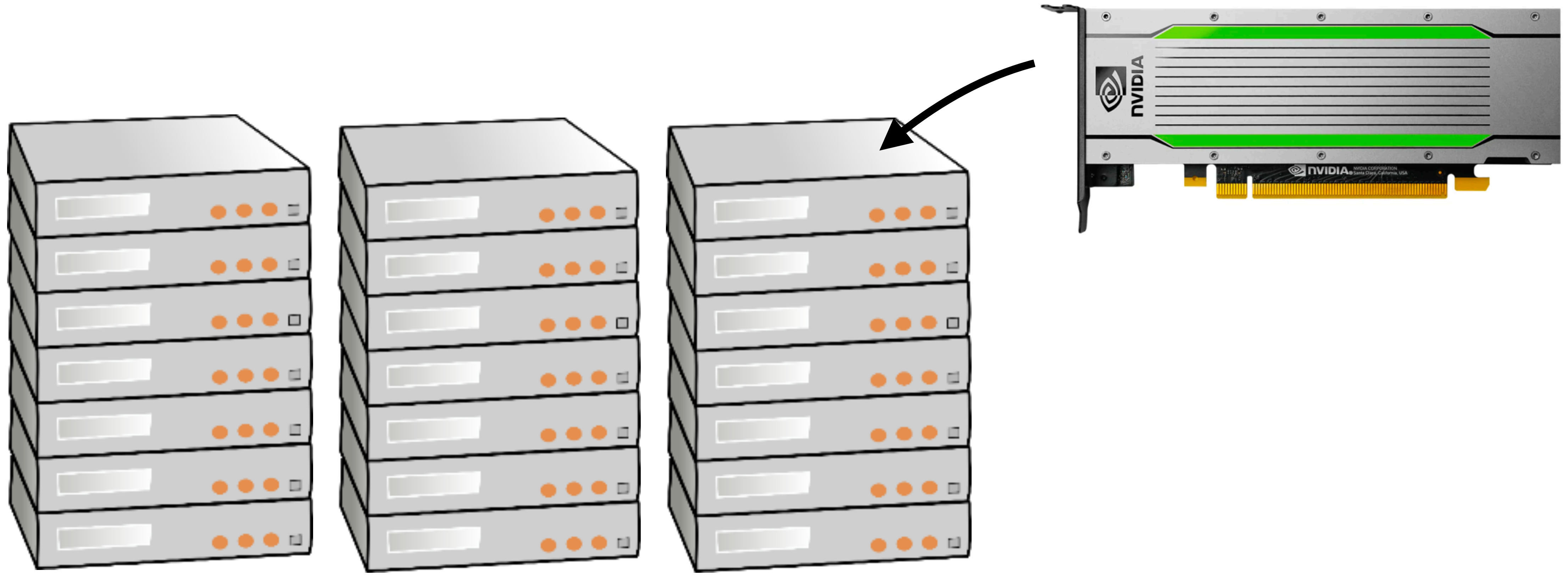
HLT: More Particle Flow

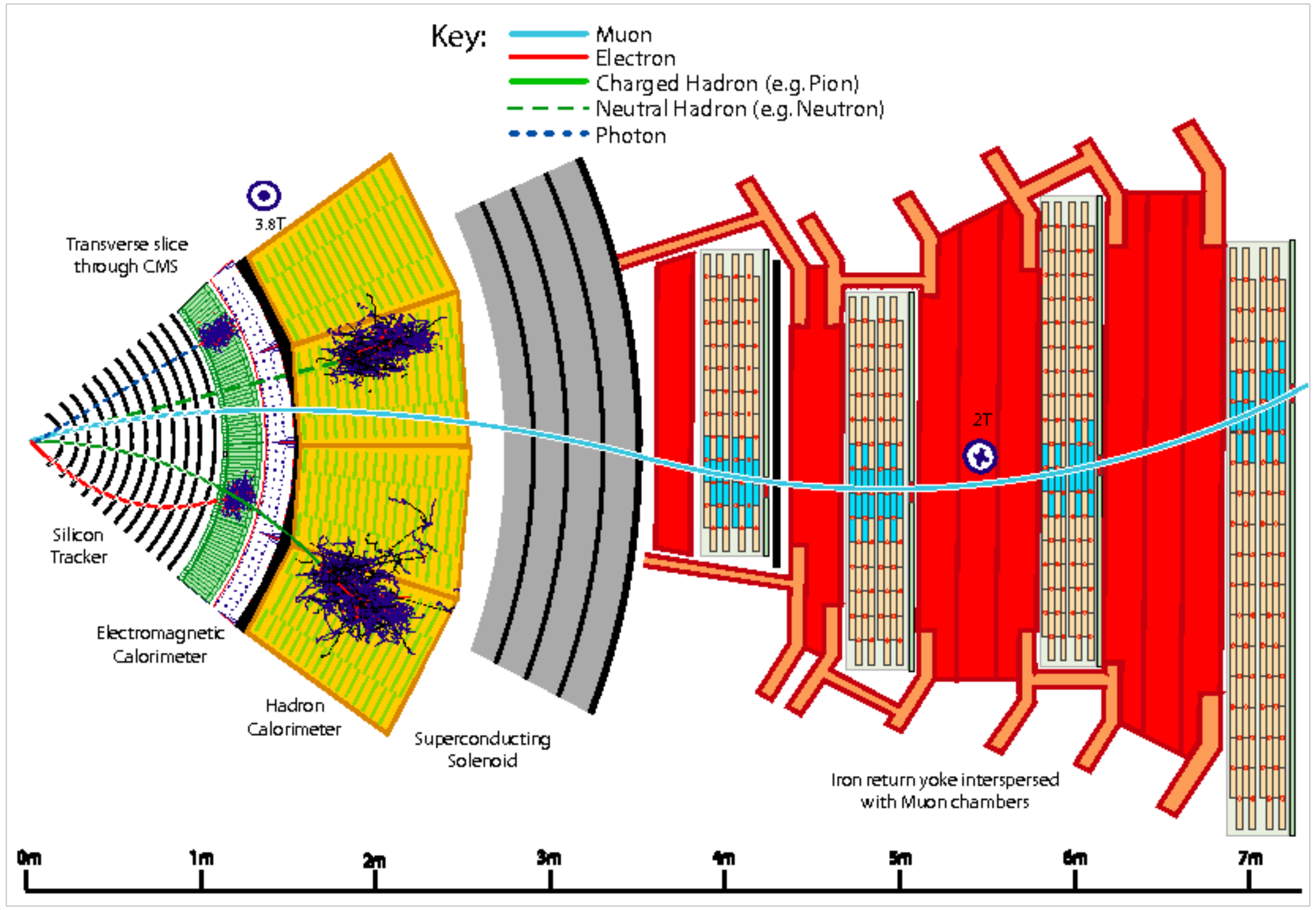
Particle Flow is highest resolution reconstruction at HLT. Slow, can't run on all events! Currently only PF on 17% of total

- High resolution, but small rate

To handle HL-LHC data rates

- Offload resource-intensive computations to GPU
- Can achieve speed-ups ~x3
- More compute to run PF!





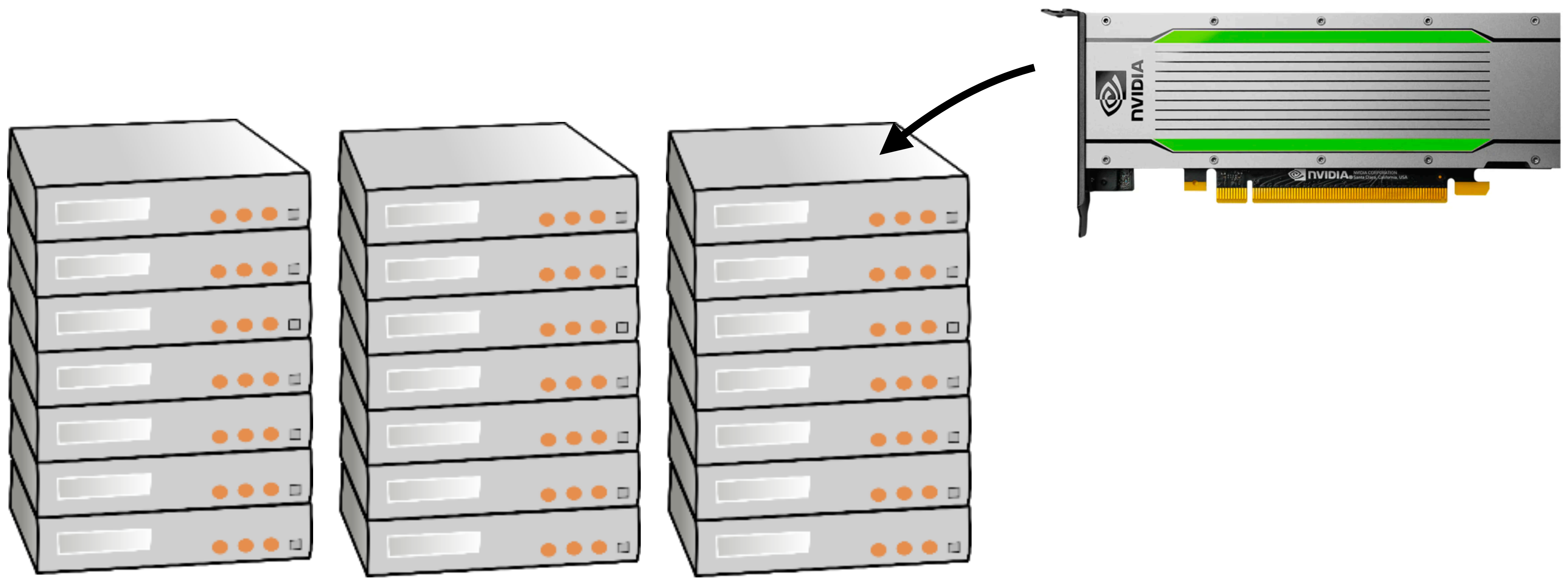
HLT: More Particle Flow

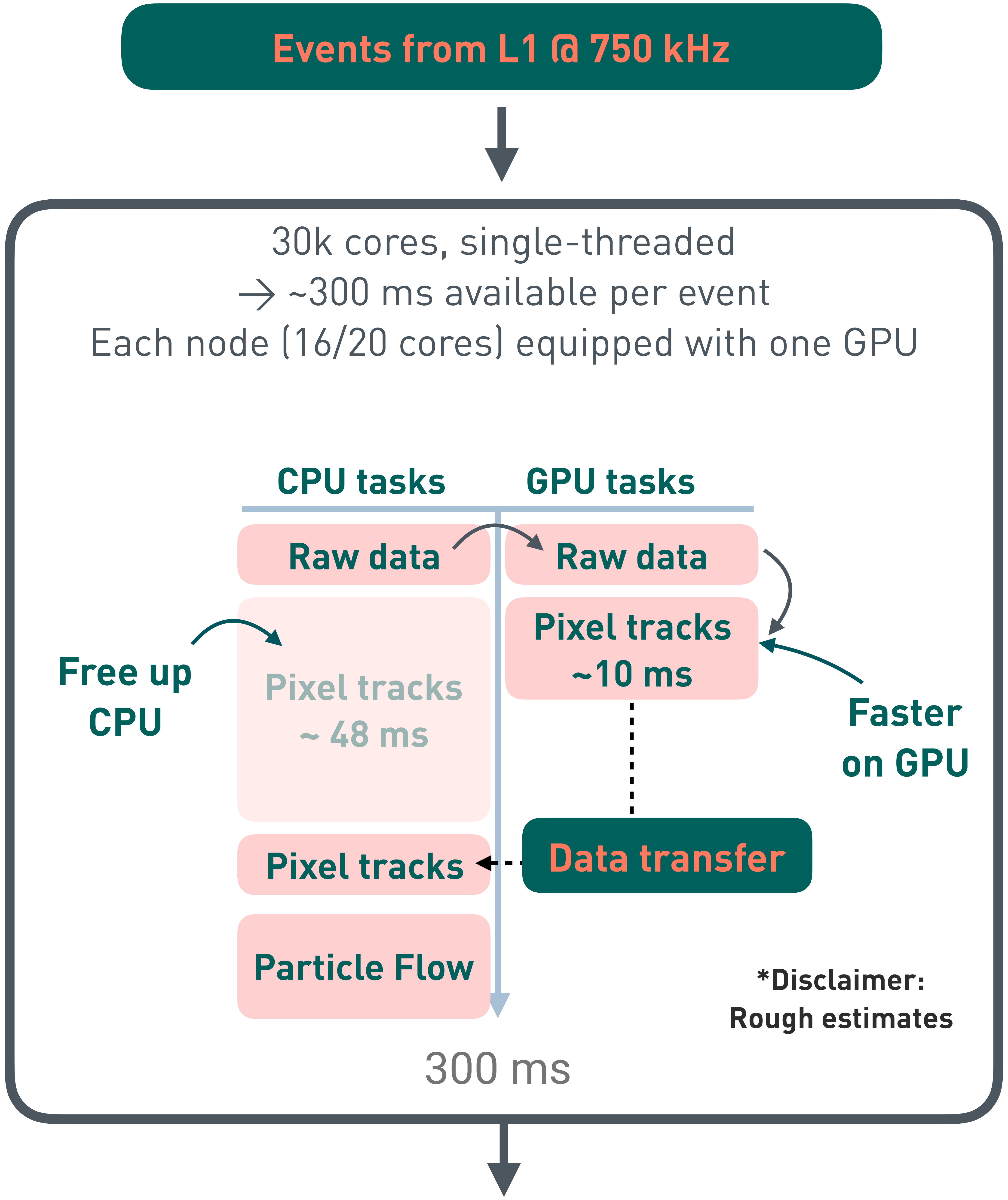
Particle Flow is highest resolution reconstruction at HLT. Slow, can't run on all events! Currently only PF on 17% of total

- High resolution, but small rate

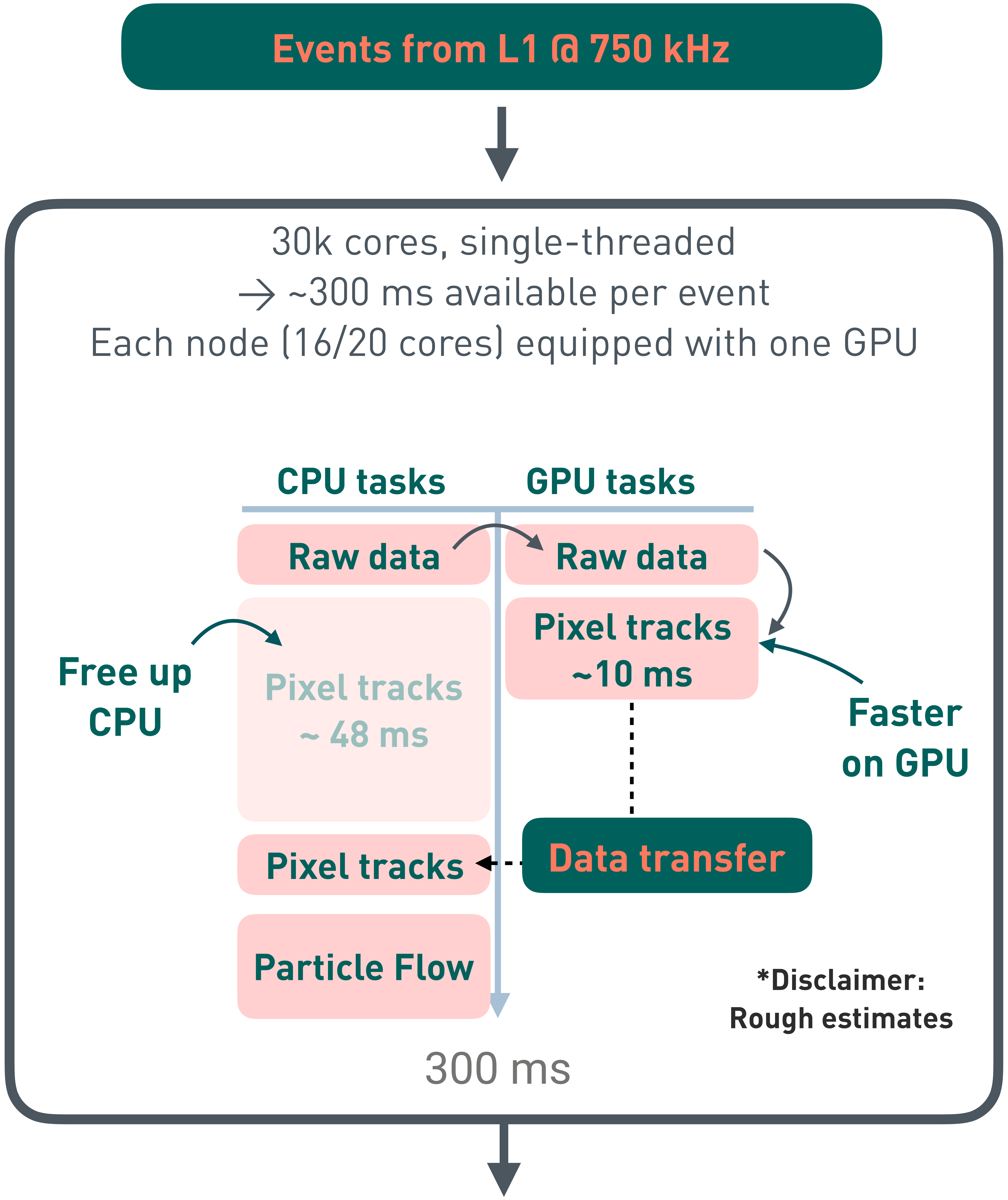
To handle HL-LHC data rates

- Offload resource-intensive computations to GPU
- Can achieve speed-ups ~x3
- More compute to run PF!





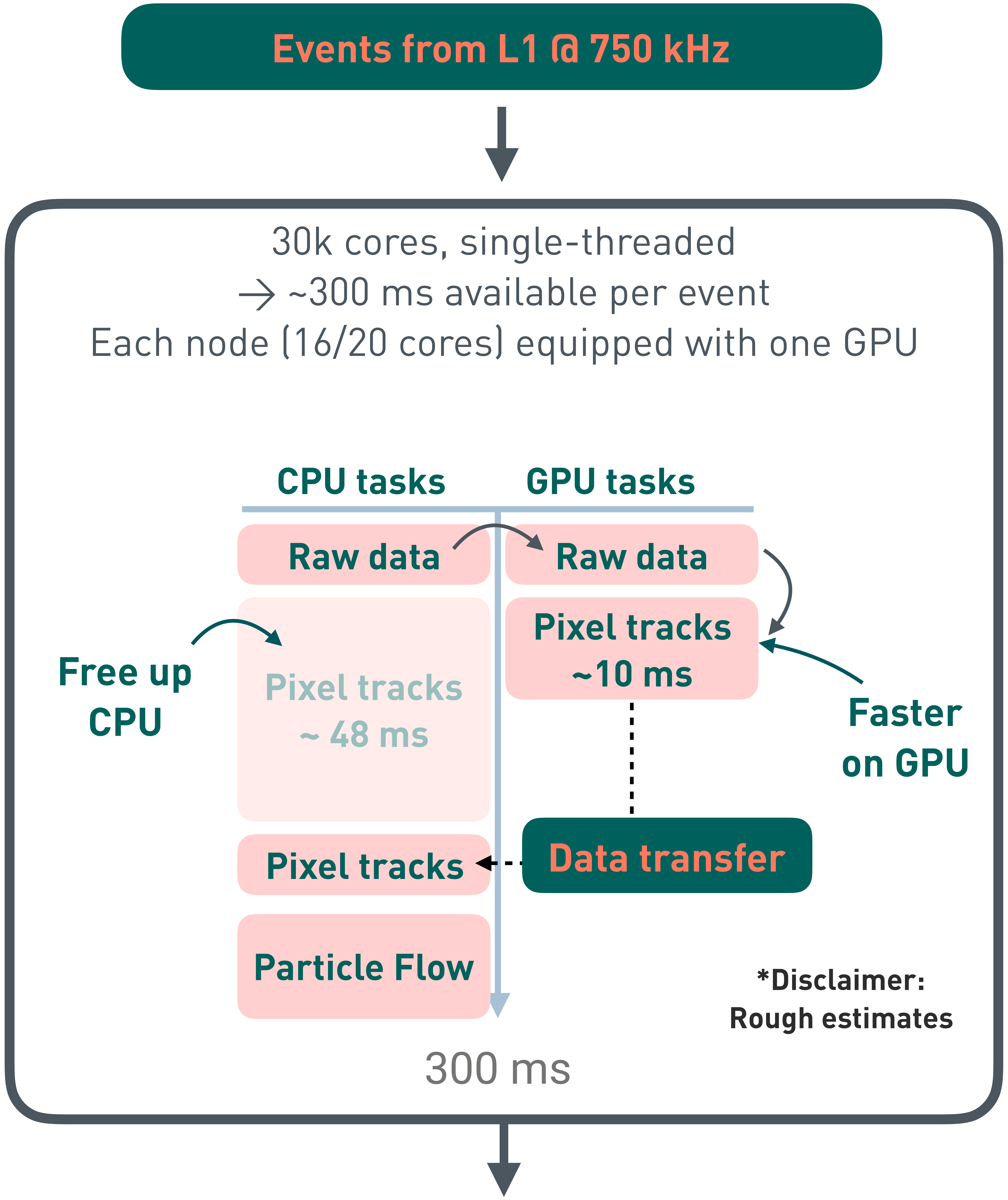
HLT: More Particle Flow



HLT: More Particle Flow

More PF means cleaner VBS triggering!
- Especially for all-hadronic VBS, jet substructure at HLT important!

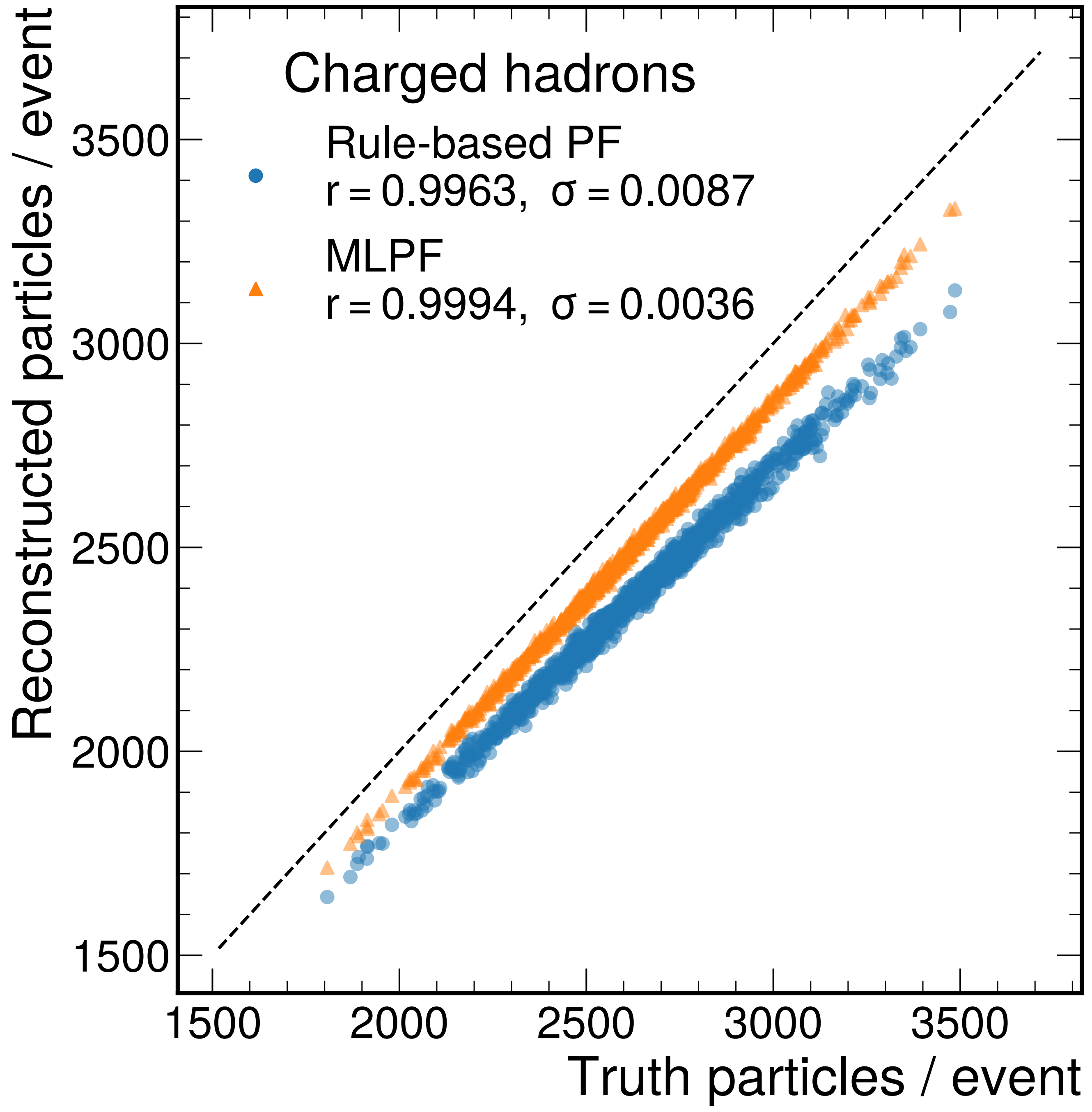
Transfer data GPU → CPU expensive, can we avoid it by doing PF on GPU?

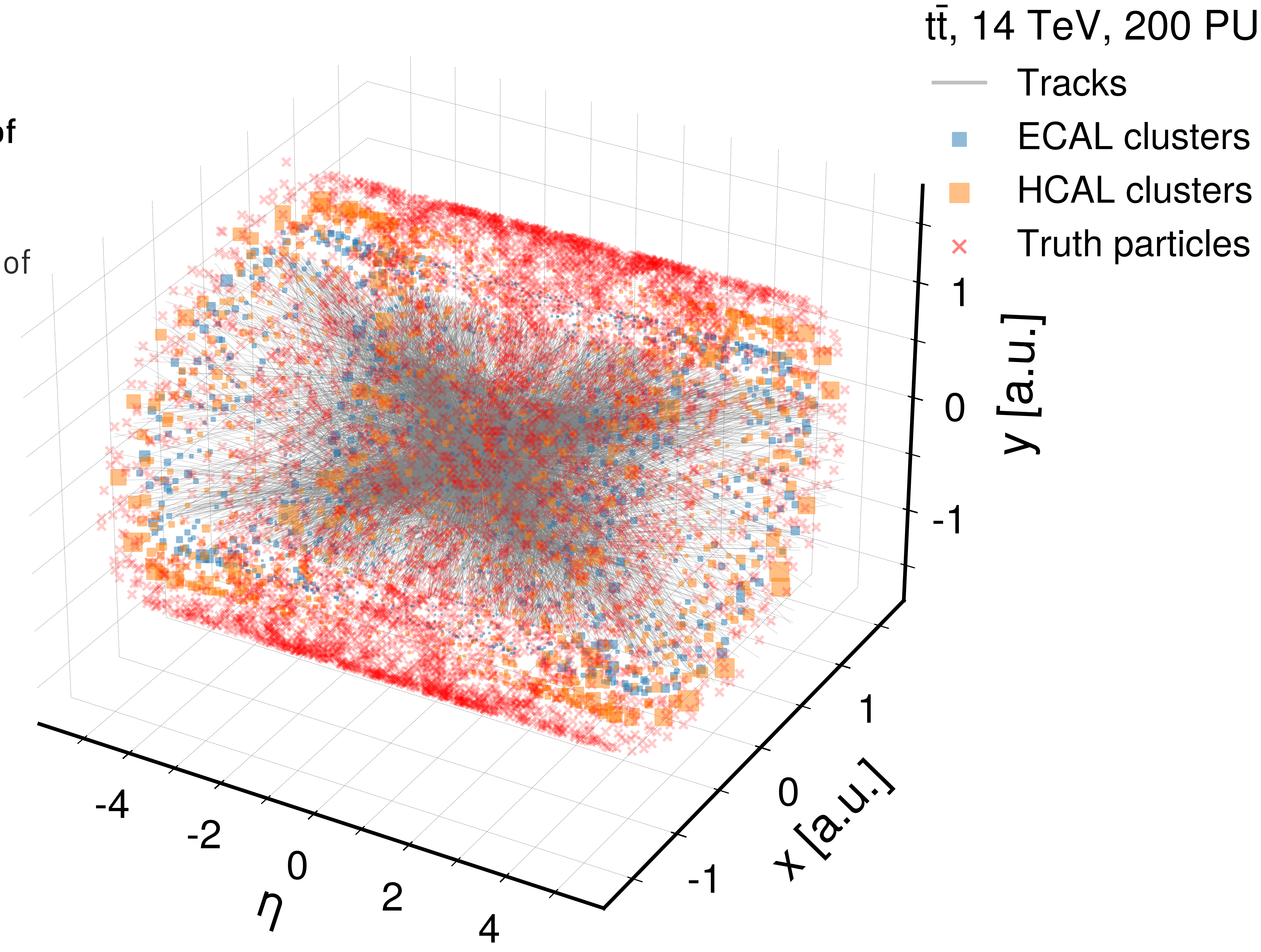


HLT: More Particle Flow

Deep Neural Networks as "fast" approximations of classical ParticleFlow

- In CUDA for free (naturally runs on GPU)
- Inherently parallelizeable, can take advantage of GPU acceleration
- High accuracy in high PU environment



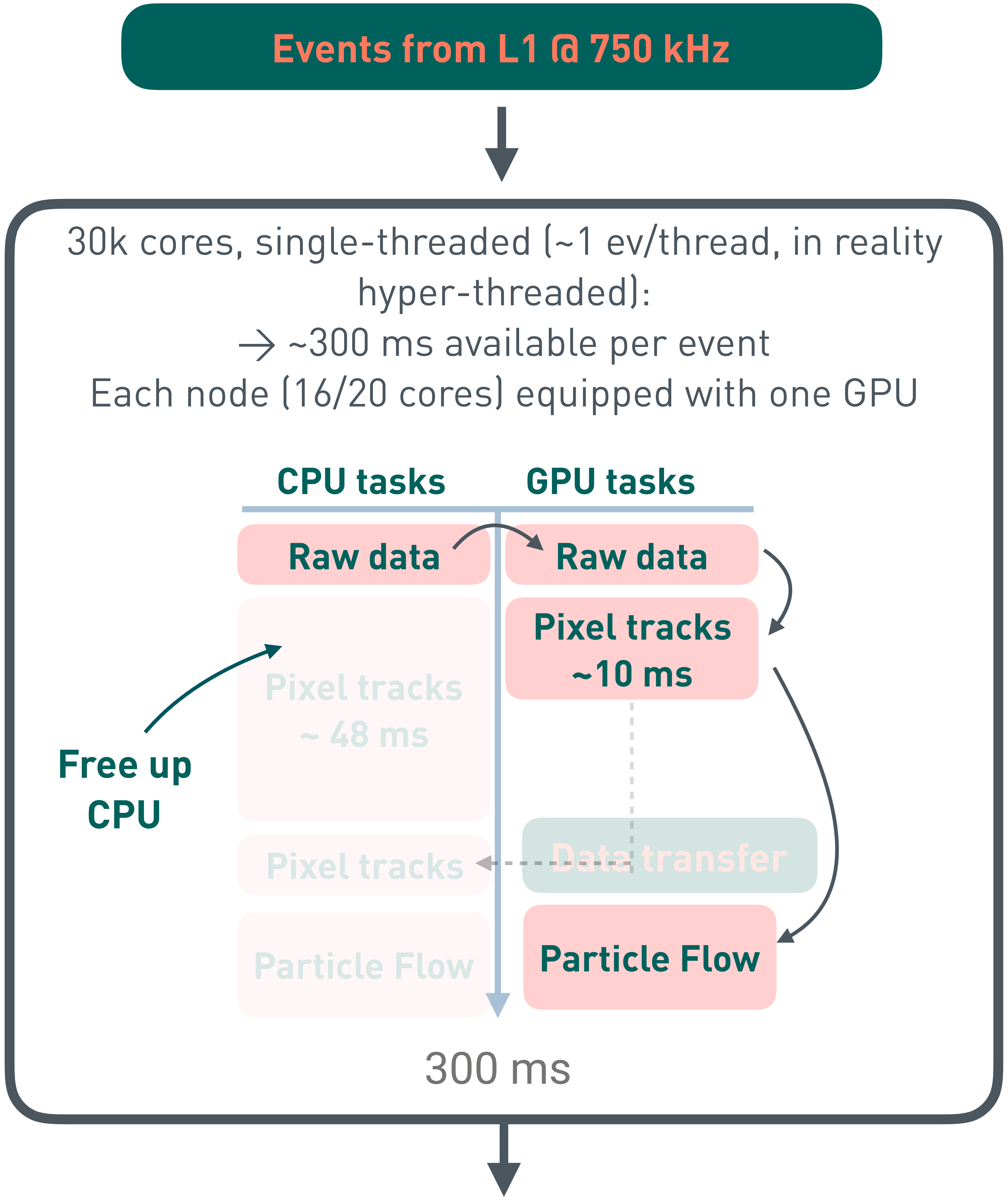


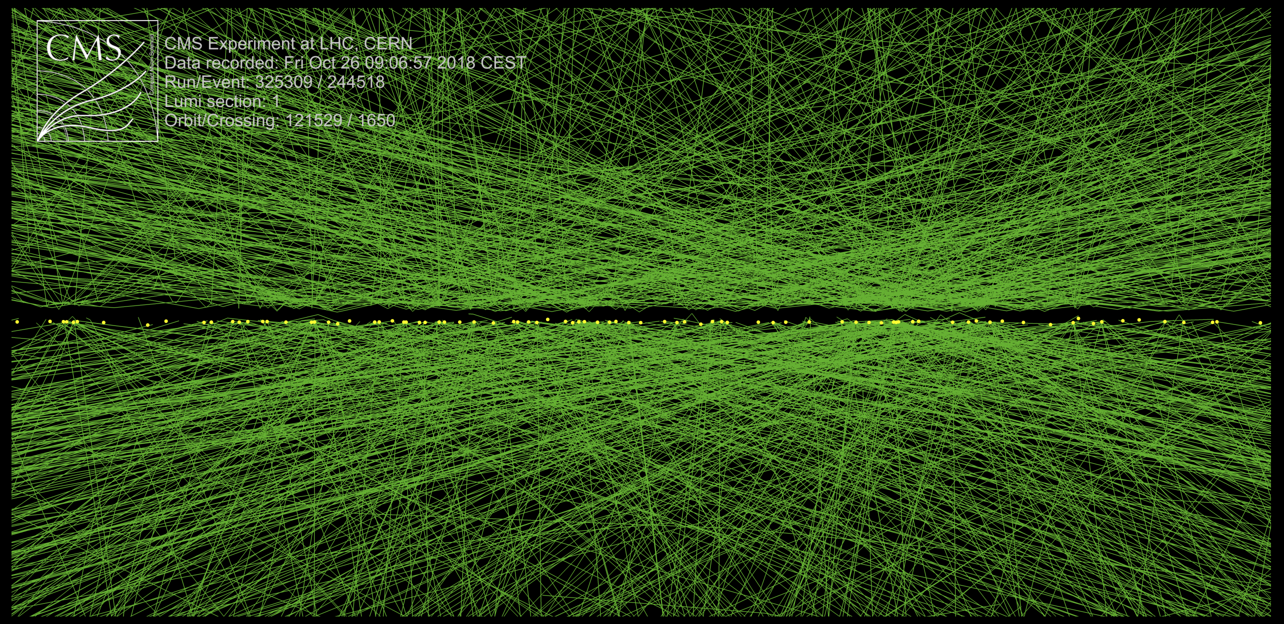
arXiv:2101.08578

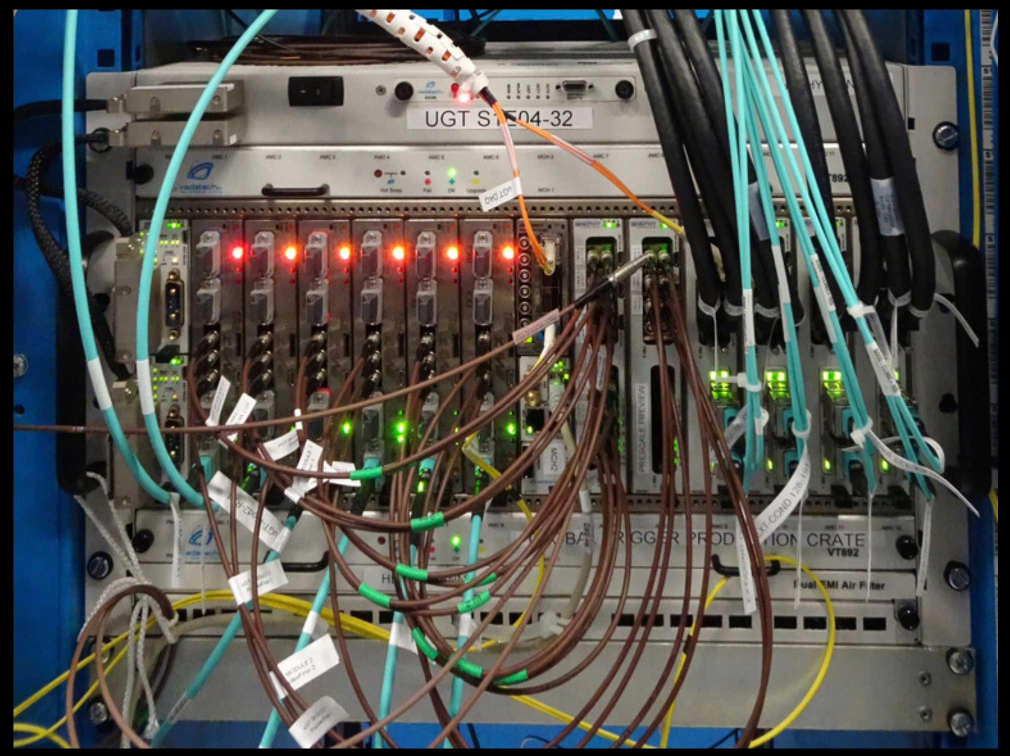
HLT: More Particle Flow

Dedicated VBS PF-based triggers

- q/g is an obvious one
- Jet substructure for V→qq

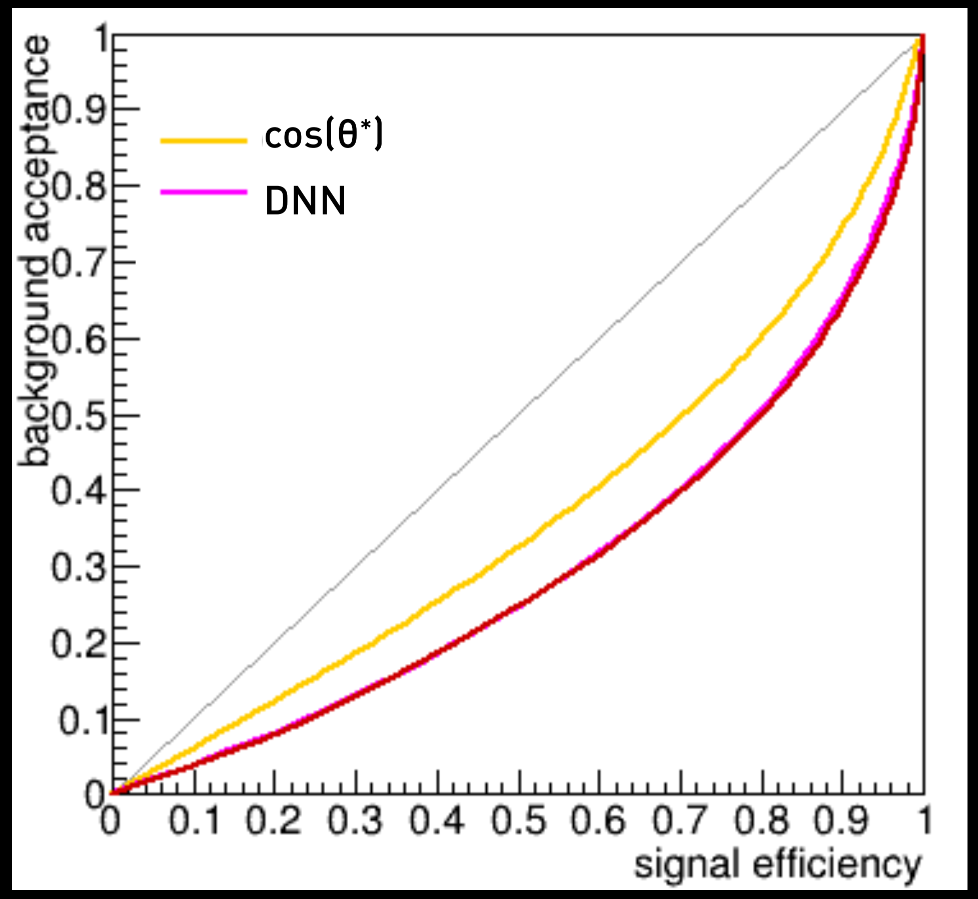






Offline analysis

- Maximize S/B for precision measurements



→ **"Classical" ML on CPU/GPU**

Polarization in VBS

Run 3/HL-LHC: Huge lumi increase at ~same C-O-M energy

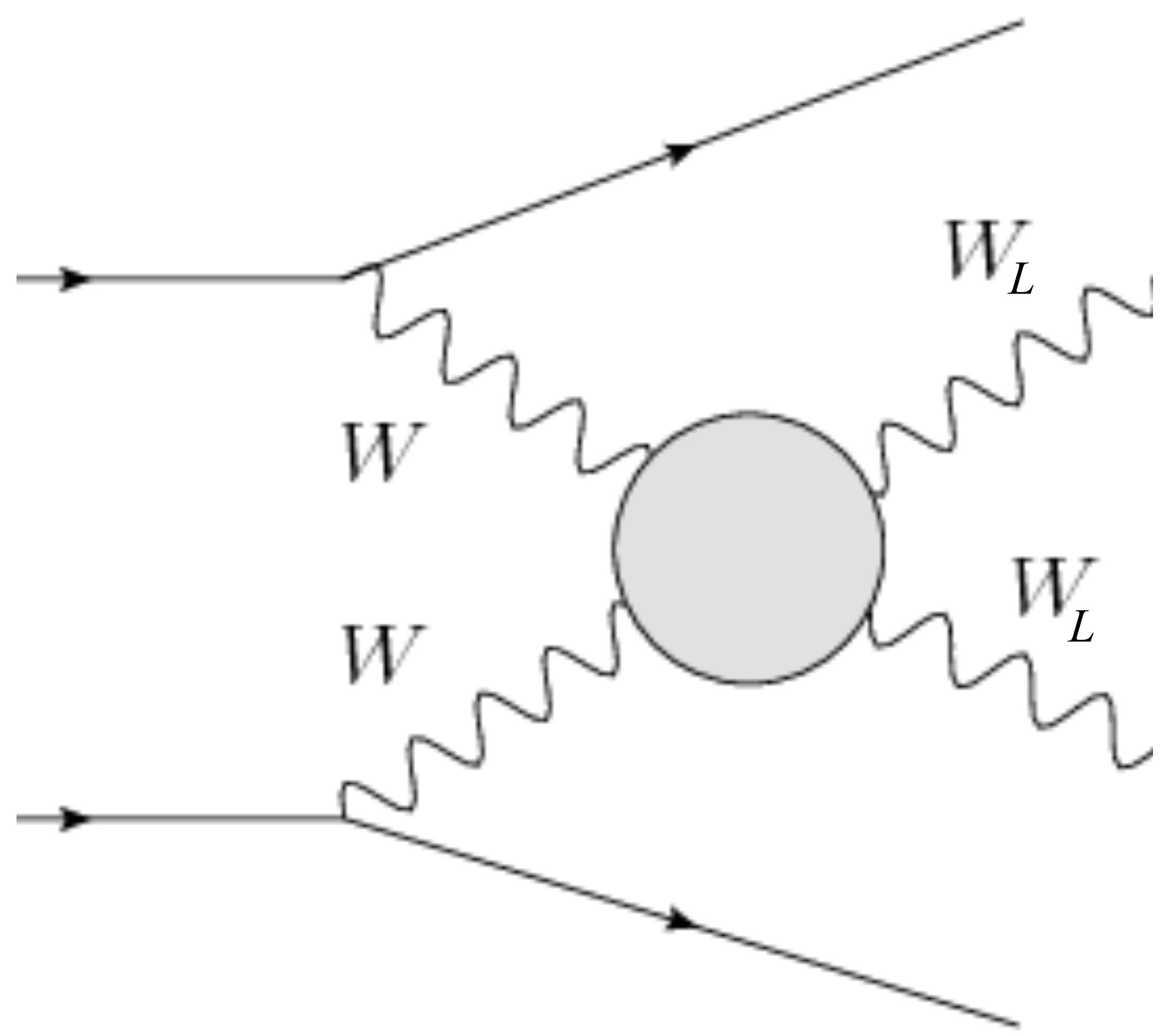
BSM bump hunts → precision measurements targeting BSM.
E.g interference in 2→2 scattering

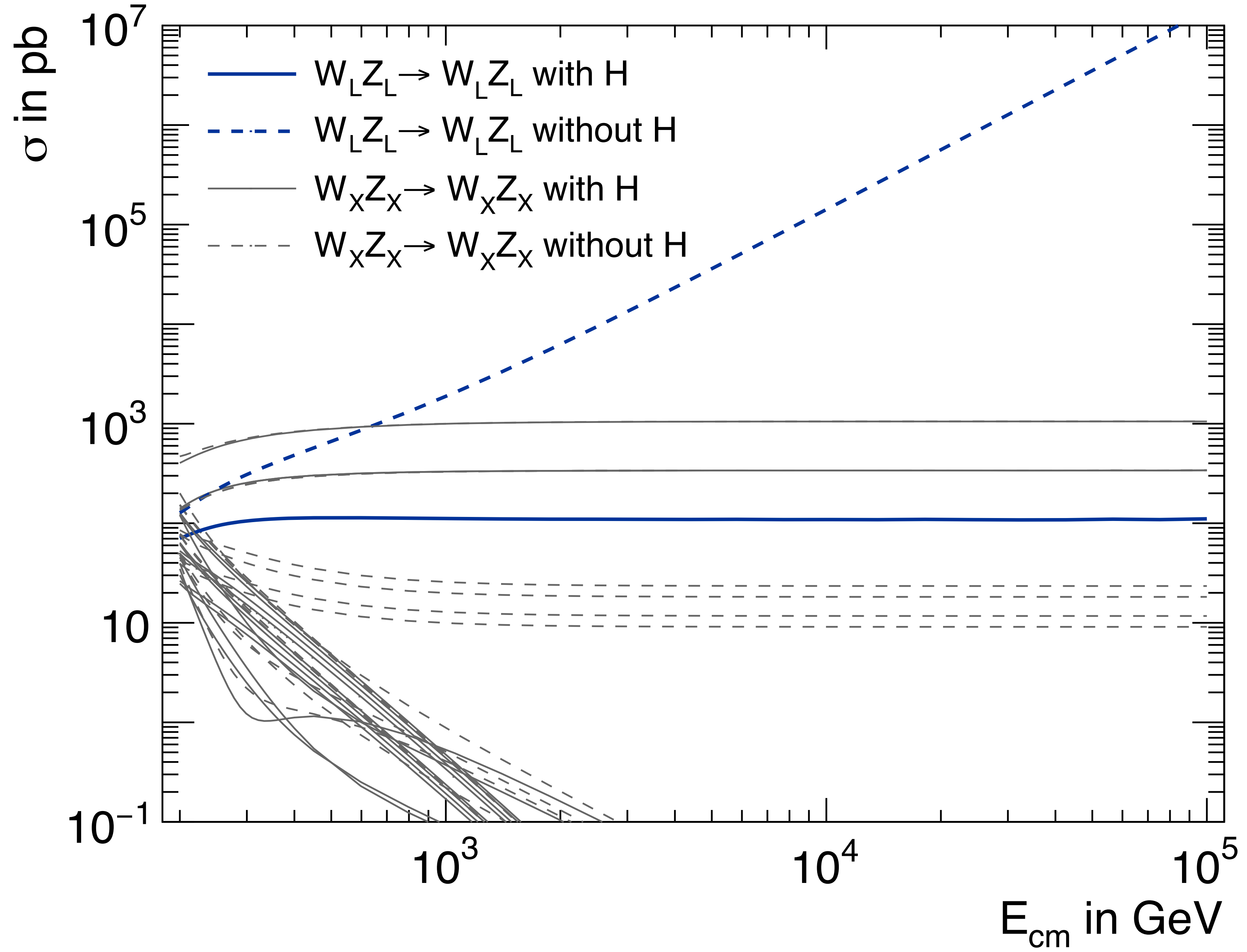
- Quartic coupling
- Higgs without Higgs
- Polarisation fractions (vs E)
- BSM enhancements

At $E \gg m_V$, BSM mainly couples to W_L

- 90% SM is TT (m_{VV} > 250 GeV), irreducible background!

Can we do W_T vs. W_L? (See Kenneth's talk)

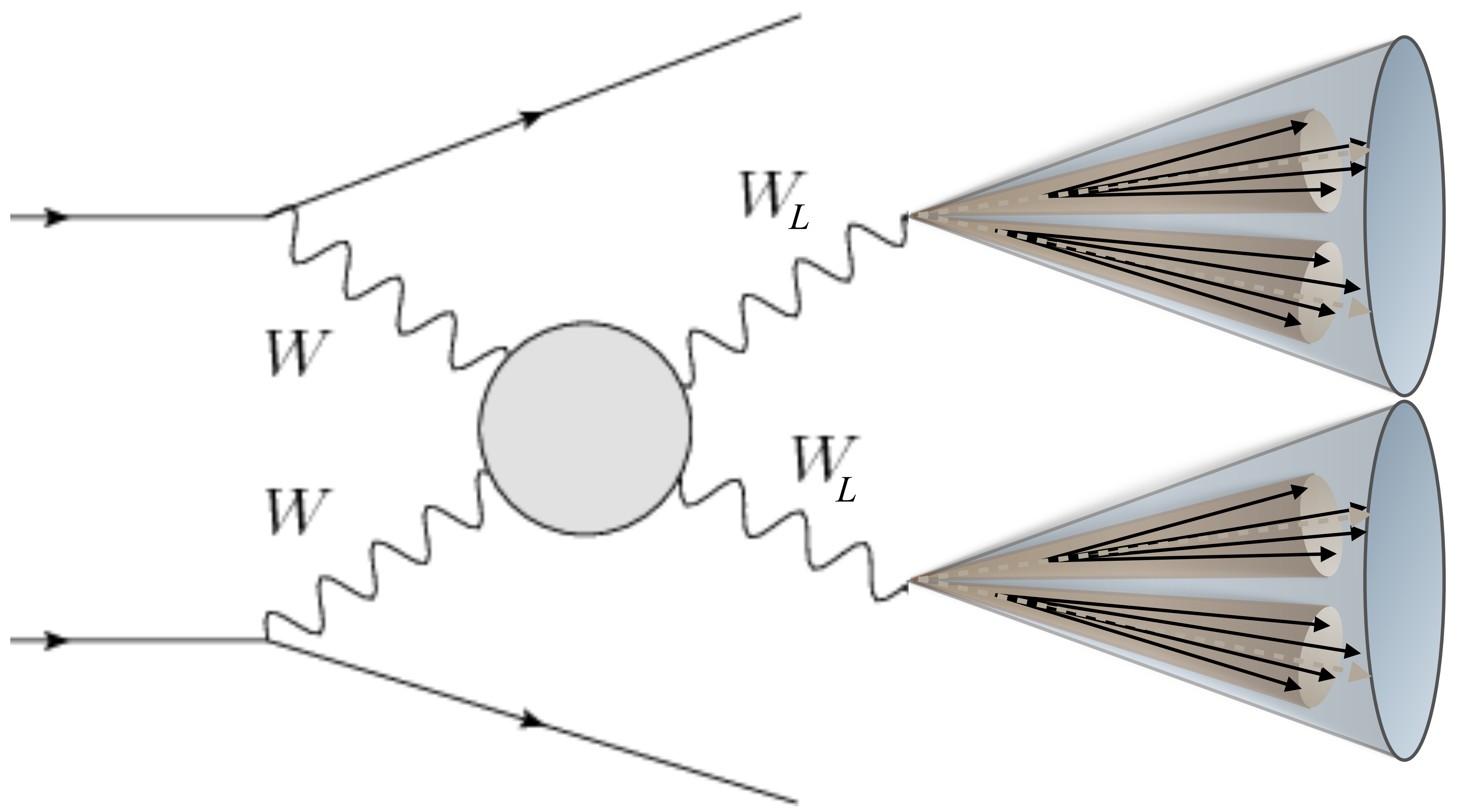




Polarization in VBS

All-hadronic VBS as a “golden channel” for polarization (don’t laugh)?

- $W_L W_L$ small → BR(V→qq)
- Access to both fermions (caveat: charge)
- Inference grows with energy → boosted jets
- q-tag on forward jets + two tagged V-jets
- Access polarisation both through forward q-jets + vector bosons



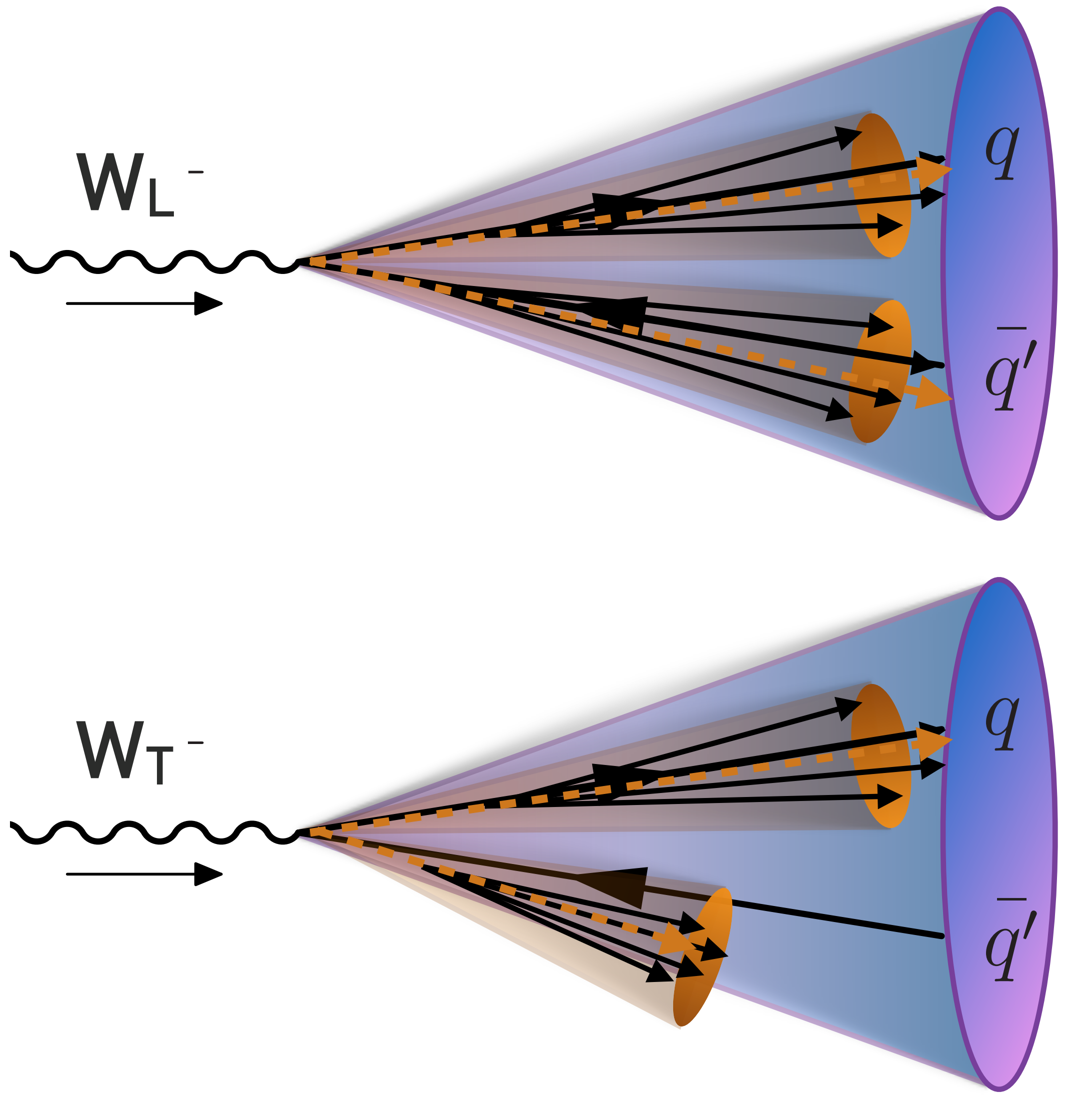
Two-step problem

- Discriminate VBS from QCD VV
- Discriminate W_T from W_L

Two powerful tests of SM that can be made feasible by ML

- Crossection measurement in W_L enriched phase space
- Measurement of helicity fractions

Polarization in VBS



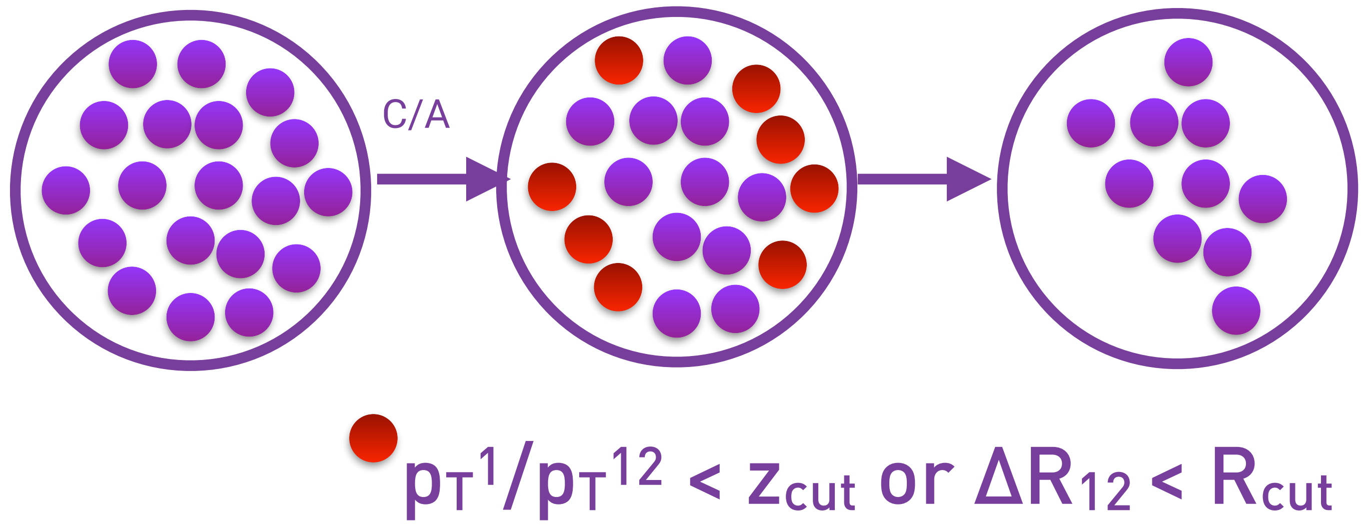
W_T vs. W_L discriminating power observed in jet substructure variables

- W_T decay products preferentially (anti-)parallel to W momentum
  → asymmetric p_T between subjects or overlapping partons (lab frame)

Jet substructure techniques

- Grooming: Remove soft+wide angle radiation (bad for asymmetric p_T)
- N-subjettiness: Probability for 2 axes within jet (bad for overlapping partons)

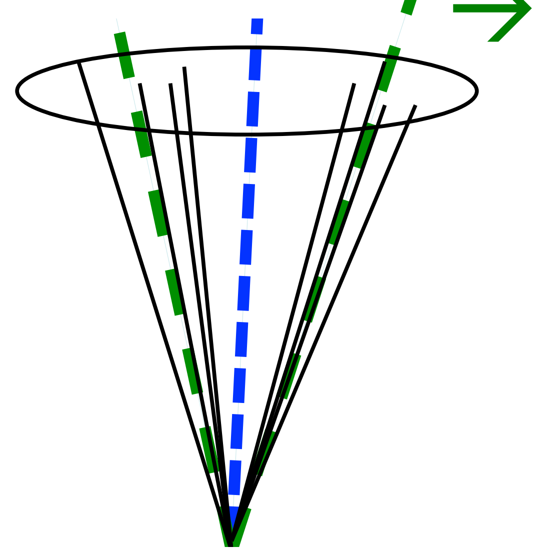
Jet mass resolution → Pruning



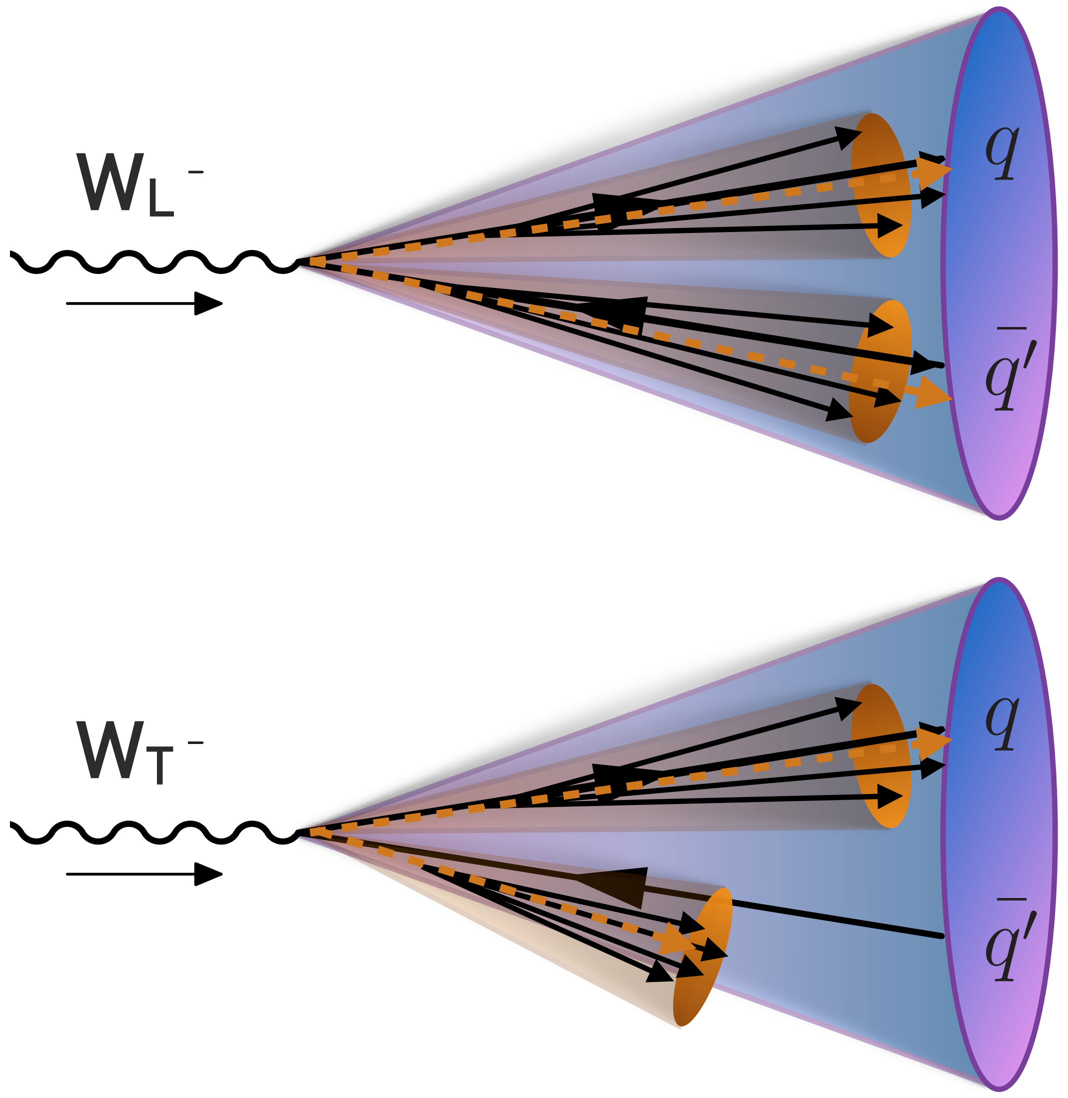
Are there "subjets" → n-subjettiness

Probability of jet having N subjets, τ_N

-use ratio τ_2/τ_1



Polarization in VBS

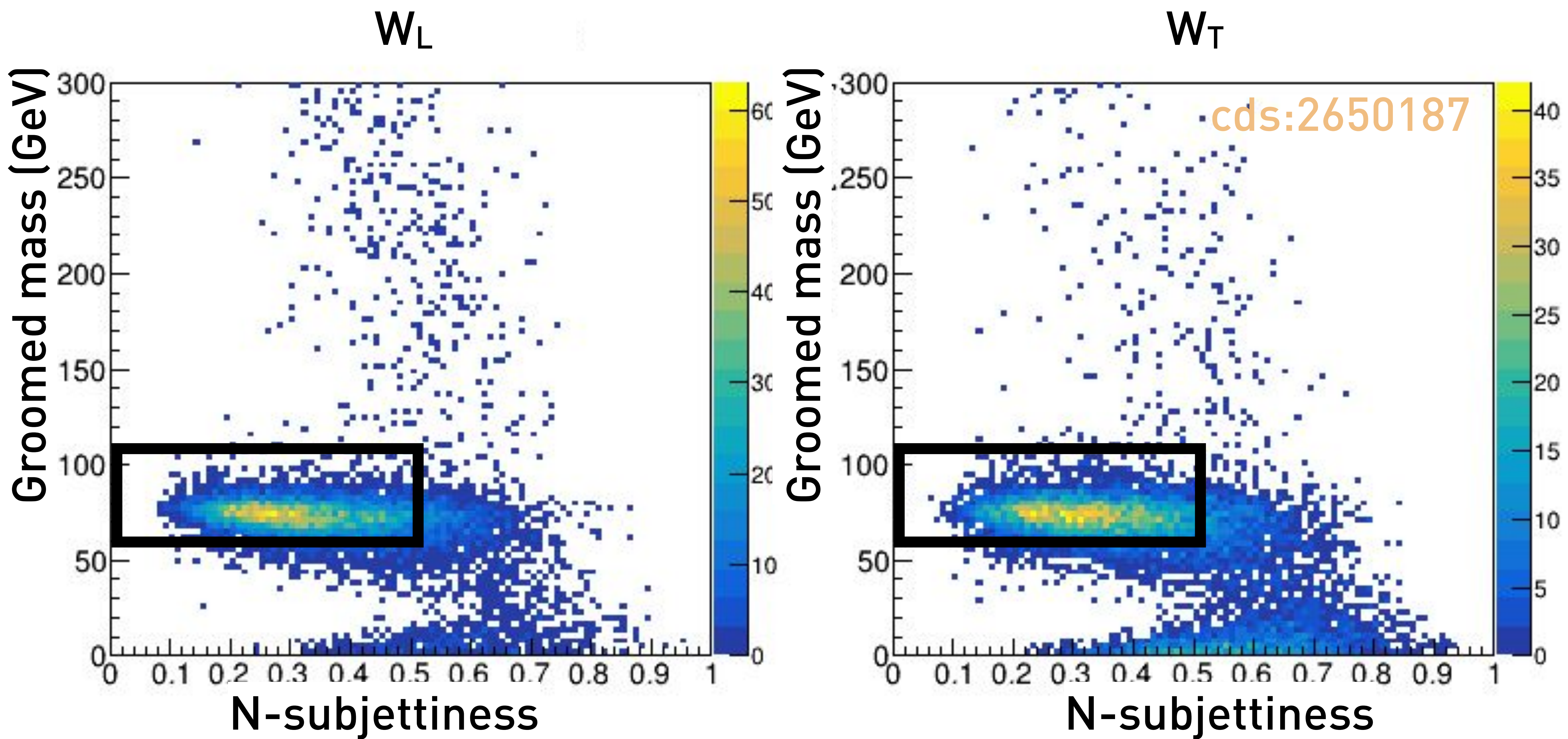


W_T vs. W_L discriminating power observed in jet substructure variables

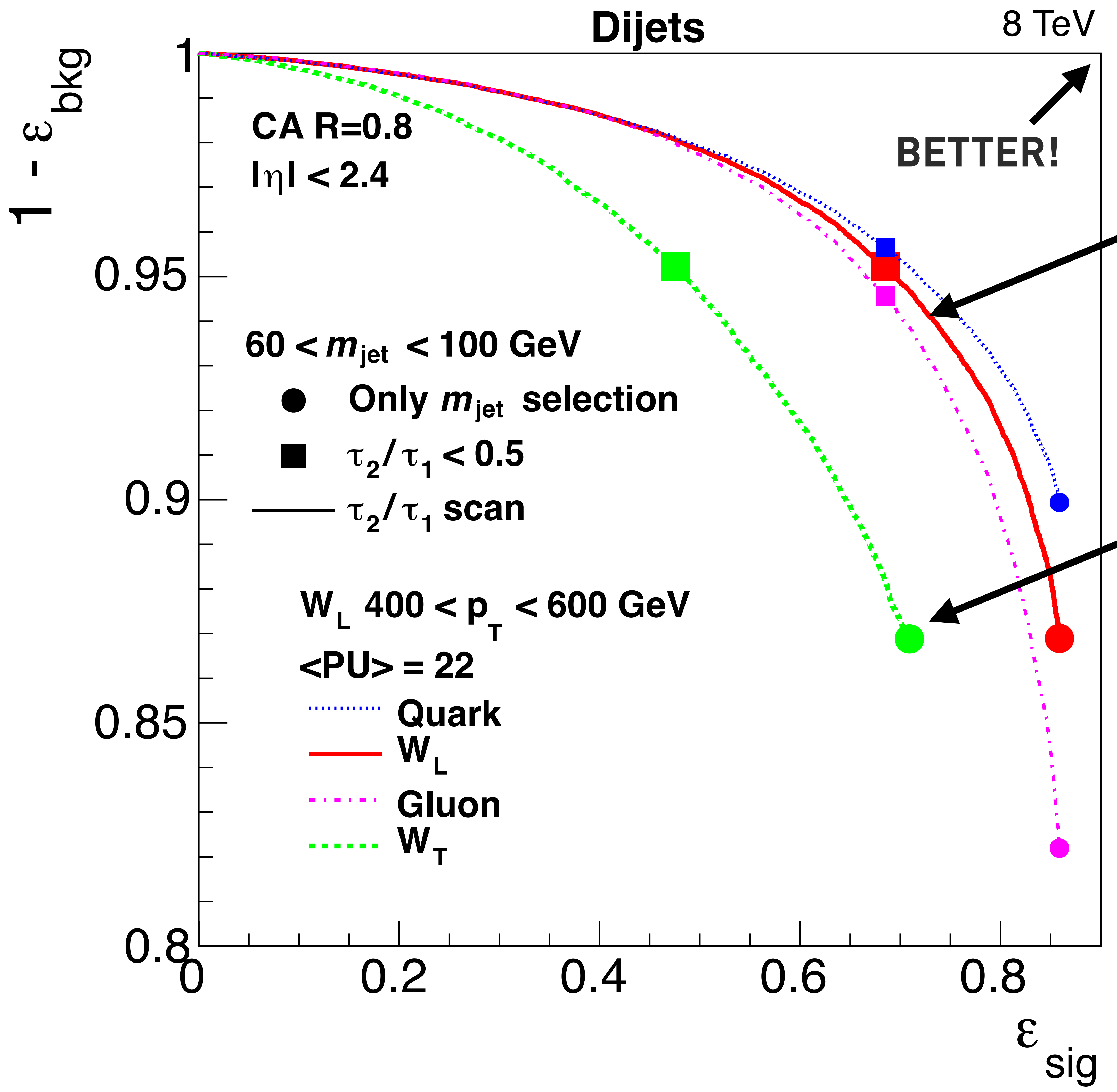
- W_T decay products preferentially (anti-)parallel to W momentum
  → asymmetric p_T between subjects or overlapping partons (lab frame)

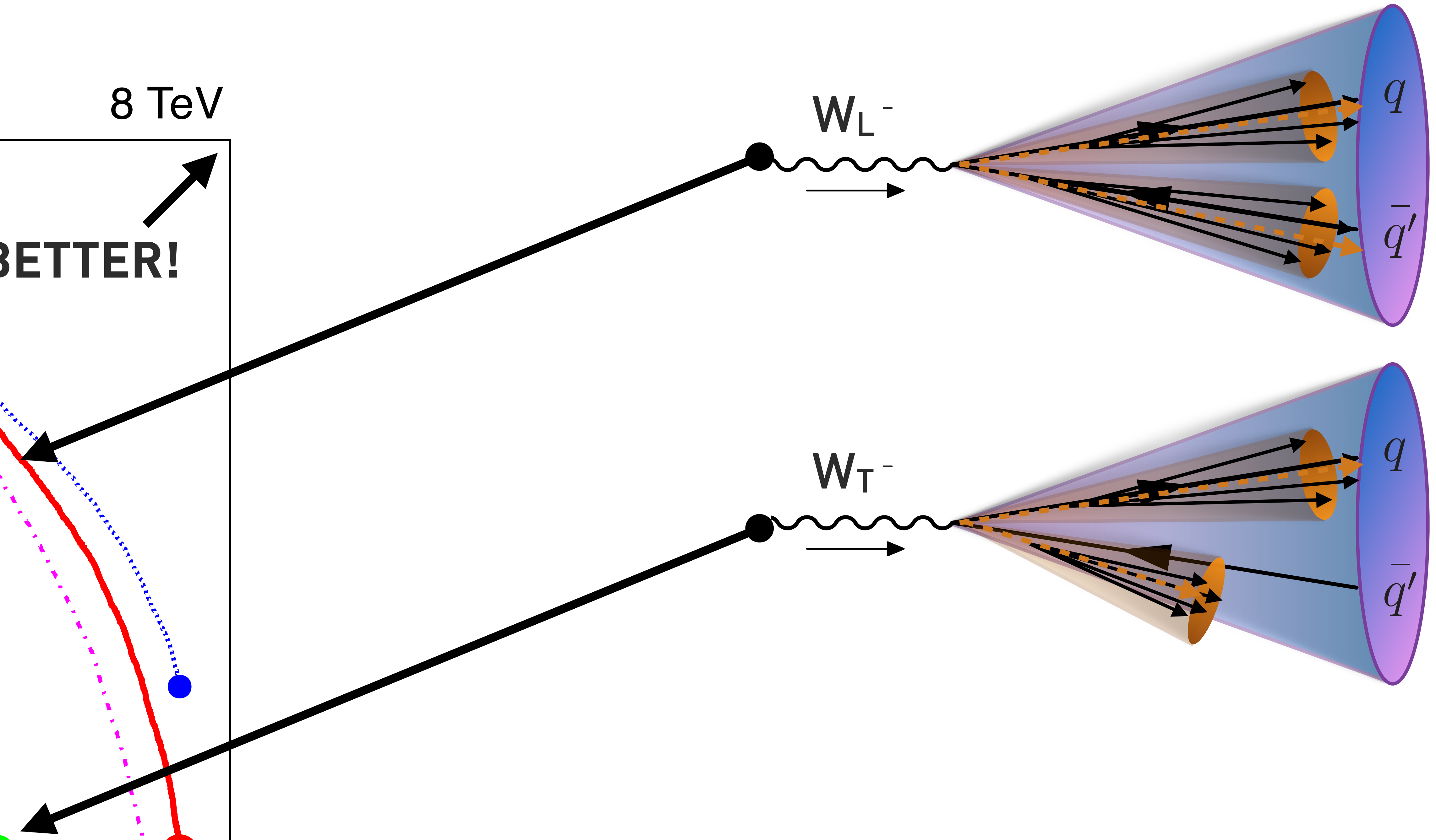
Jet substructure techniques

- Grooming: Remove soft+wide angle radiation (bad for asymmetric p_T)
- N-subjettiness: Probability for 2 axes within jet (bad for overlapping partons)



Polarization in VBS





W_T looks 1-prong, hard to tell apart from q/g!

- Need dedicated substructure algorithms or jet-constituent based DNN to pick up all correlations

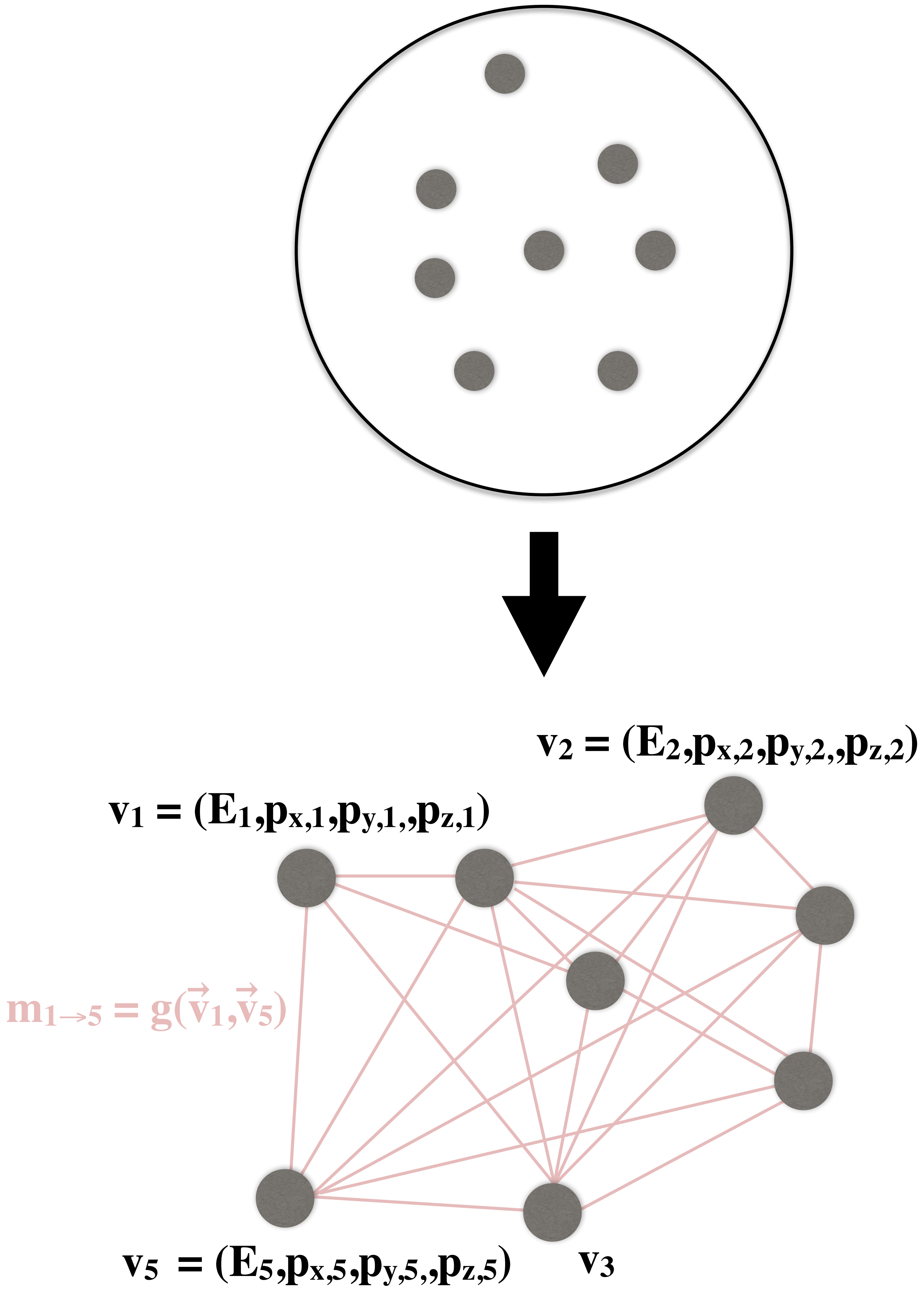
Also q/g discrimination with jet substructure,

- Same architecture for q/g?

V polarisation:GNNs

Graph Neural Network optimal architecture for jet constituents (sparse, unordered)

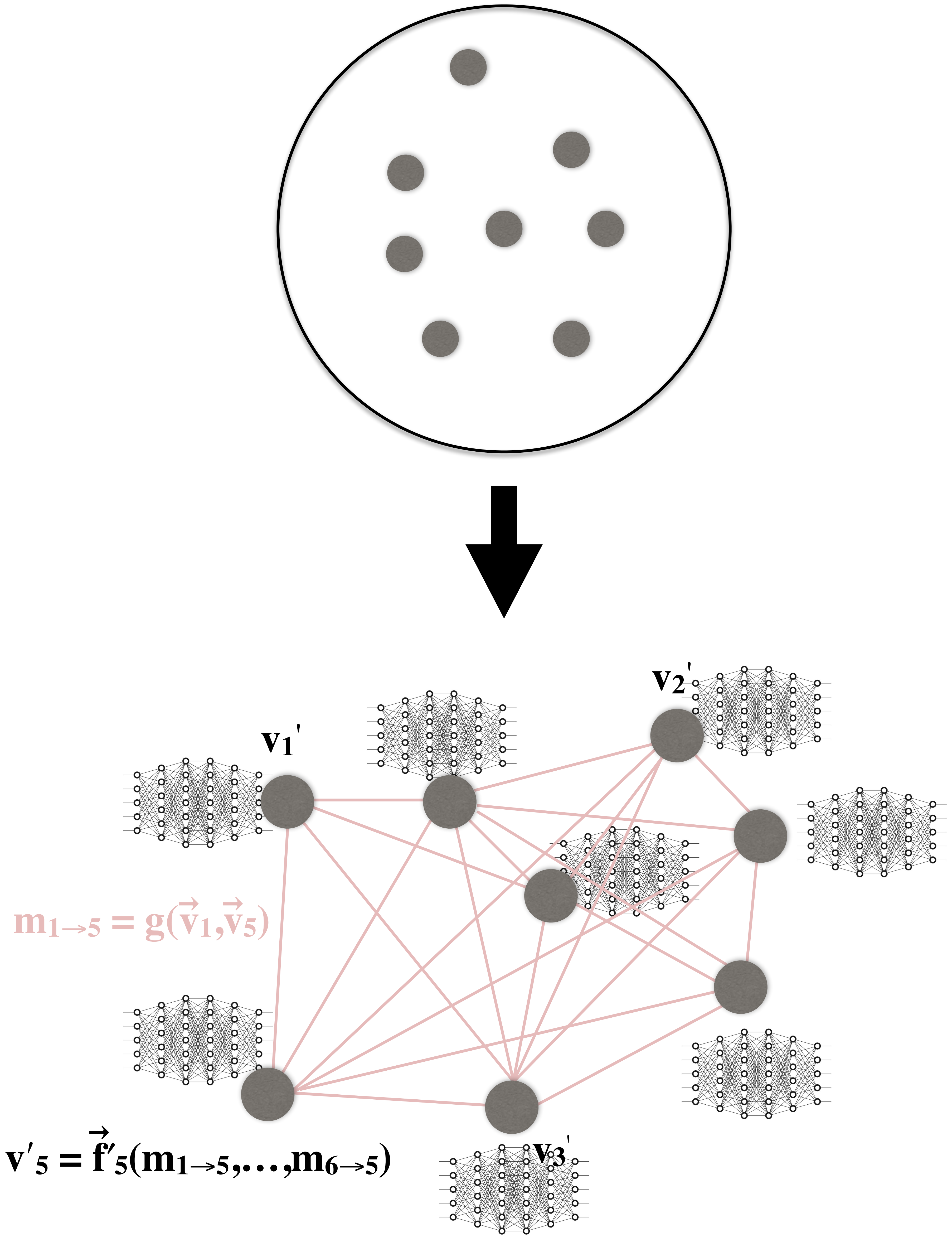
- Each node is feature vector (e.g constituent 4-vector)
- Connected through edges where message passing can occur



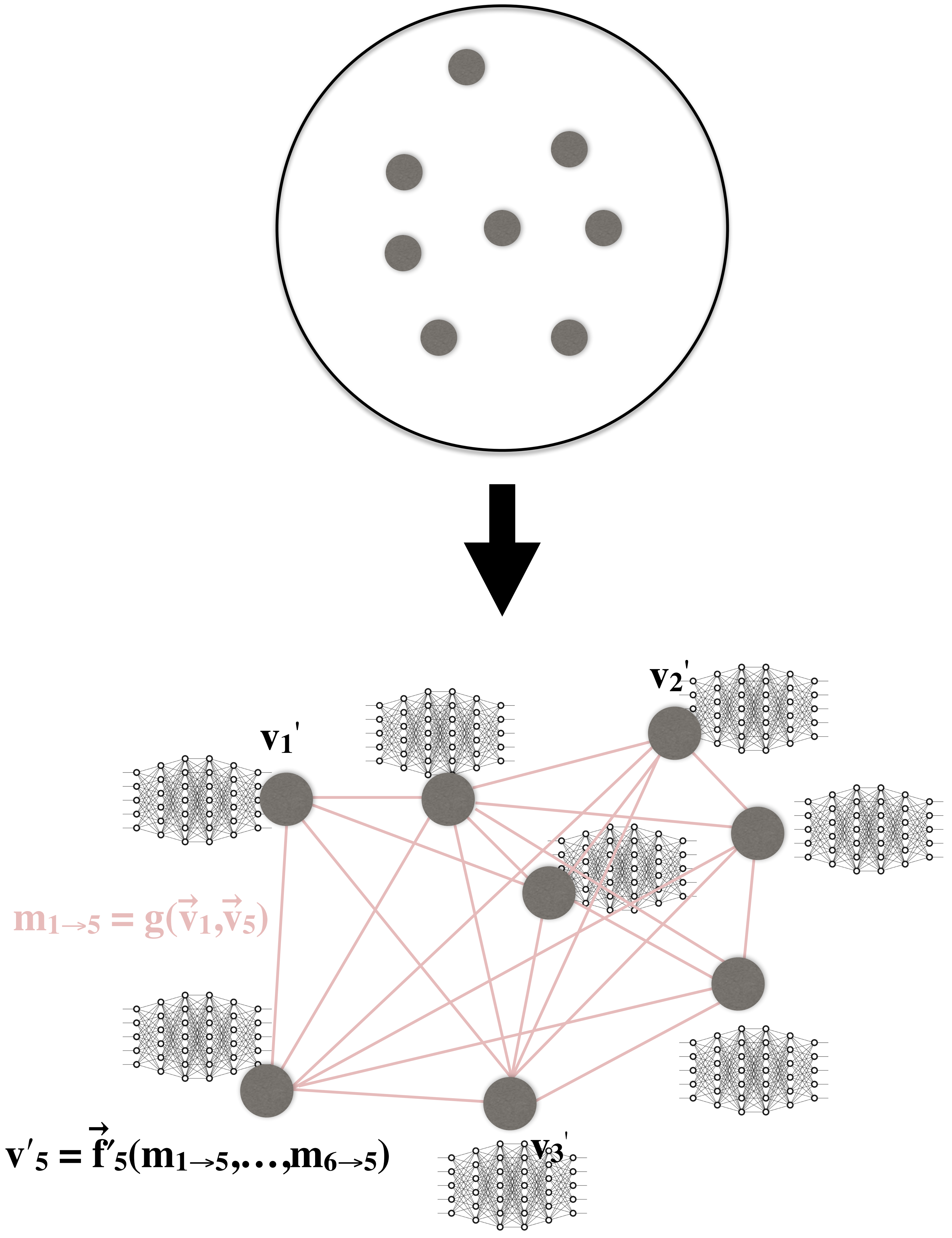
V polarisation:GNNs

Graph Neural Network optimal architecture for jet constituents (sparse, unordered)

- Each node is feature vector (e.g constituent 4-vector)
- Connected through edges where message passing can occur
- New representation of node created based on information gathered across graph
- Inference at each node using input features and learned representation



V polarisation: GNNs



Graph Neural Network optimal architecture for jet constituents (sparse, unordered)

- Each node is feature vector (e.g constituent 4-vector)
- Connected through edges where message passing can occur
- New representation of node created based on information gathered across graph
- Inference at each node using input features and learned representation

Can be defined as N-output multi-classifier

- W_T, W_L, q, g...

Jet constituents good to pick up correlations, but almost always good to extend base by adding high-level features

- Energy correlation functions, τ_N, groomed masses

Polarization in VBS

Proof-of-principle, simple DNN 80% W_L signal efficiency at ~40% W_T mistag rate

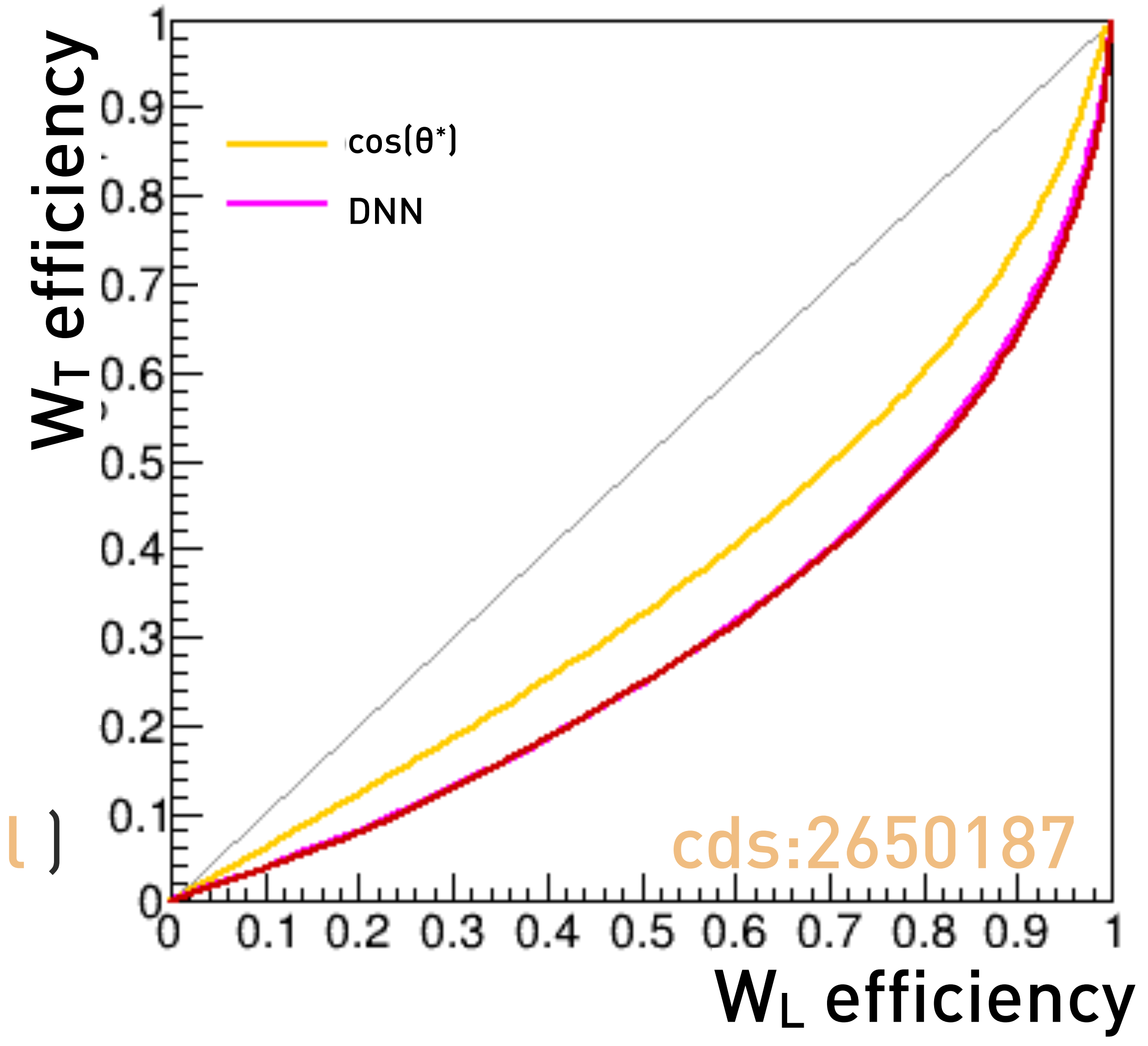
- Can we do better? Using more advanced architectures for W_T vs W_L, like GNNs, under study (H. Kirschenmann et. al.)

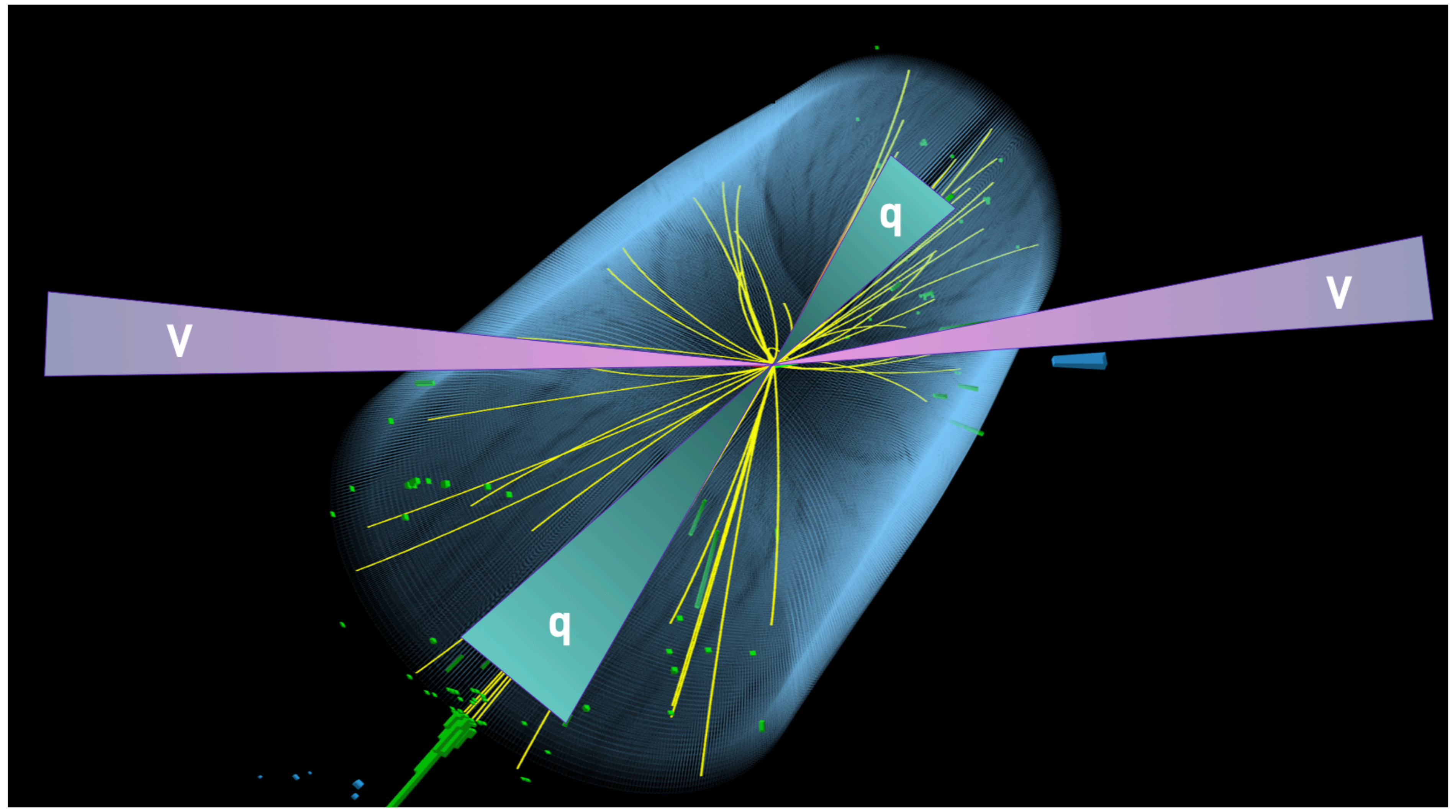
Single-object tag might not reach desired sensitivity, need to include full event information

- Di-fat jet correlations, forward jet correlations etc. (e.g Grossi. et al., Channon et. al)

Systematic uncertainties and data corrections will be extremely difficult!

- Calibration objects? Standard candles? Energy-dependence?





Polarisation in VBS

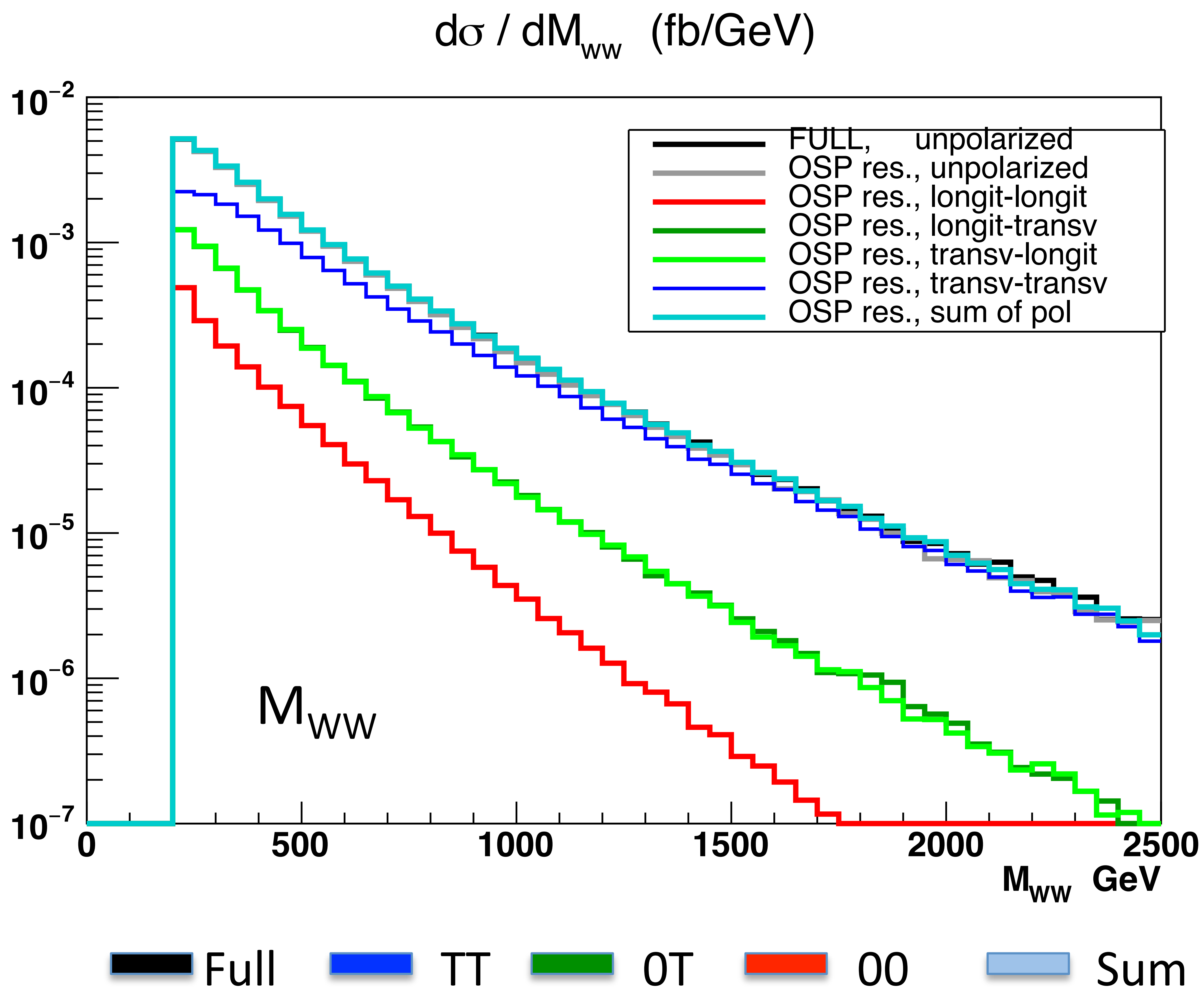
Ultimately want to extract components of polarisation density matrix

- Differential measurement of polarization fraction vs. E
- Simultaneous fit to extract amplitudes
- Look for deviations from SM in tail

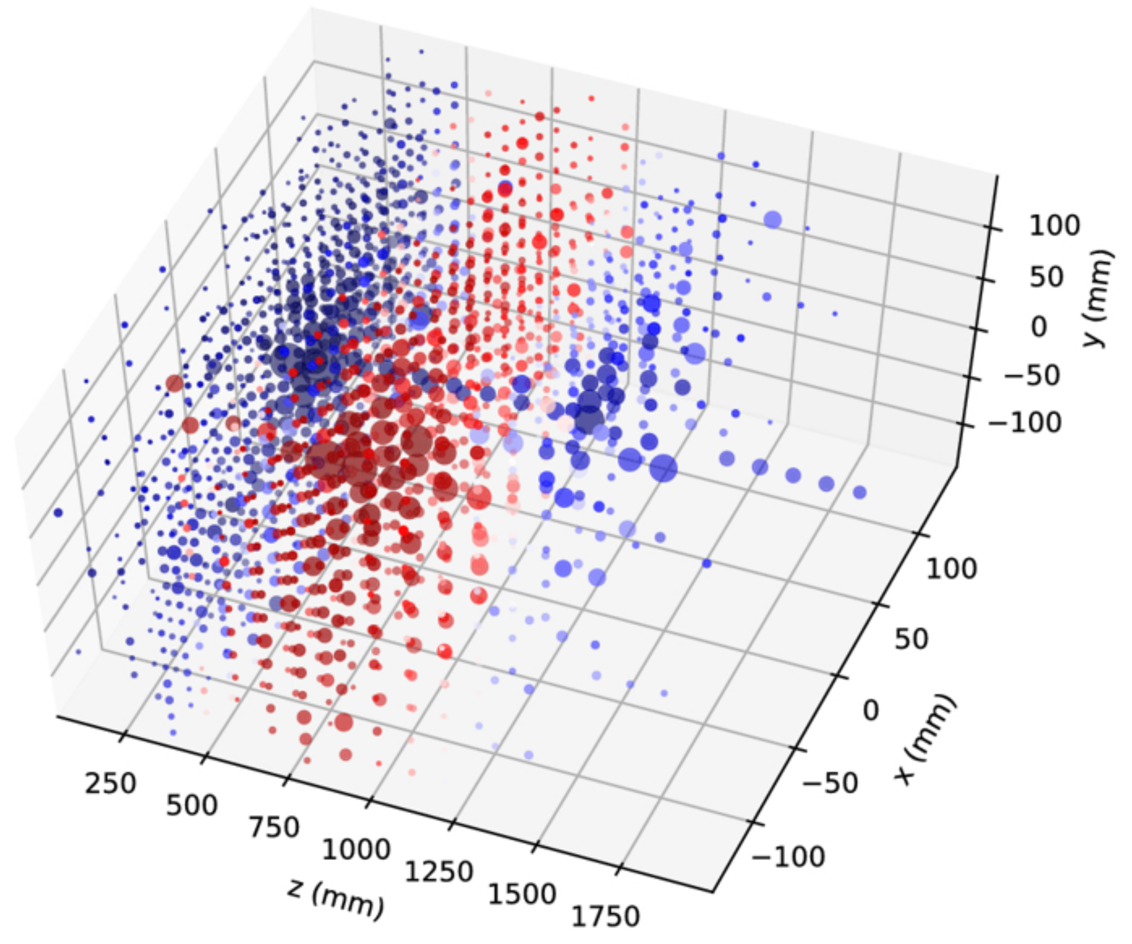
For all-hadronic, need "polarization un-biased" tagger (which current substructure algorithms are not)

- Design GNN as generic W tag (polarization unbiased)?

Need extremely high accuracy and good control over systematics, (energy dependence)



Summary



(b) Reconstructed