Fast b-tagging at the ATLAS Trigger

Lucas Bezio & Stefano Franchellucci

Université de Genève

Working with: Claire Antel, Kostas Axiotis, Quentin Berthet and Anna Sfyrla

Thanks to Michael Kagan (SLAC) for the useful discussions

UNIVERSITÉ DE GENÈVE

FACULTÉ DES SCIENCES

ATLAS EXPERIMENT

OUTLINE

Trigger in ATLAS Experiment

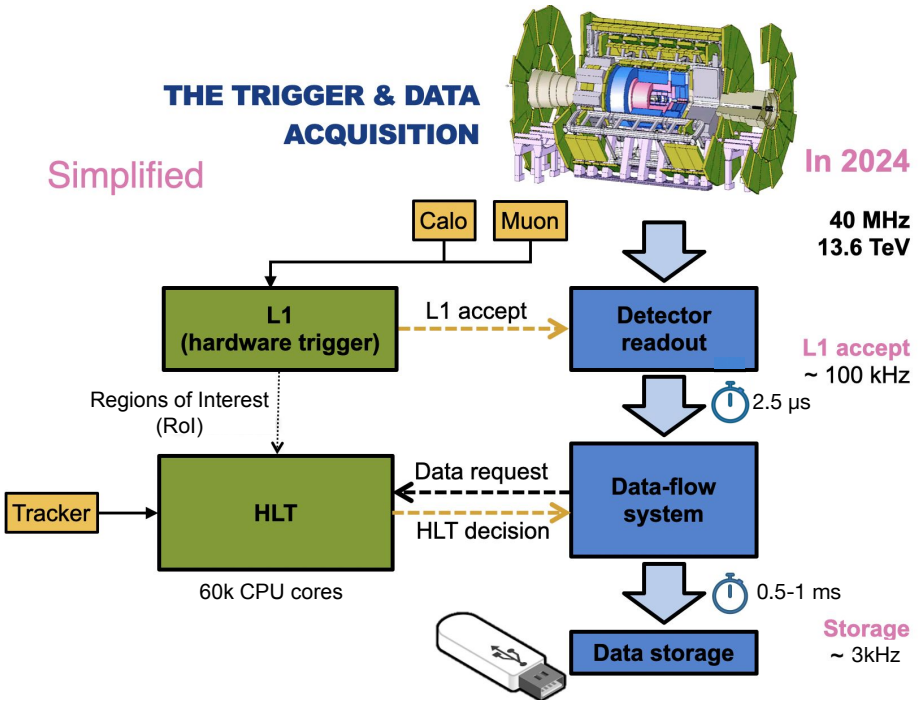
b-tagging

GNN for fast b-tagging

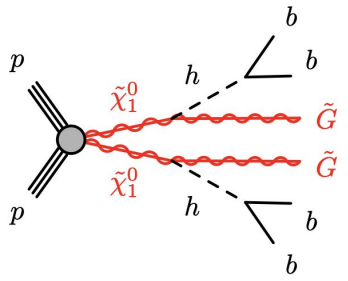
Perspectives for b-tagging in trigger

ATLAS Experiment & Trigger

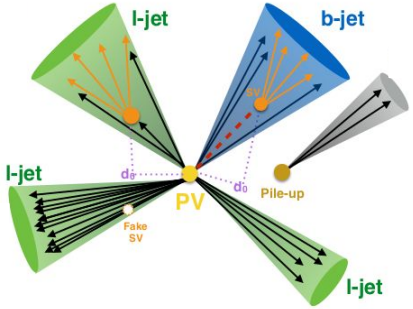
- ATLAS experiment at LHC has to deal with 40 MHz pp collisions, in high pile-up conditions
  - Large event size (~1.5MB)
  - Low rates for “interesting” physics
    - ➽ Need for a robust trigger system!
- ATLAS uses a two level trigger
  - L1: Low latency hardware trigger
  - High Level Trigger (HLT): large set of algorithms for reconstruction / tagging / selections
- Different reconstruction steps for each trigger object (muons, jets, e/gamma)



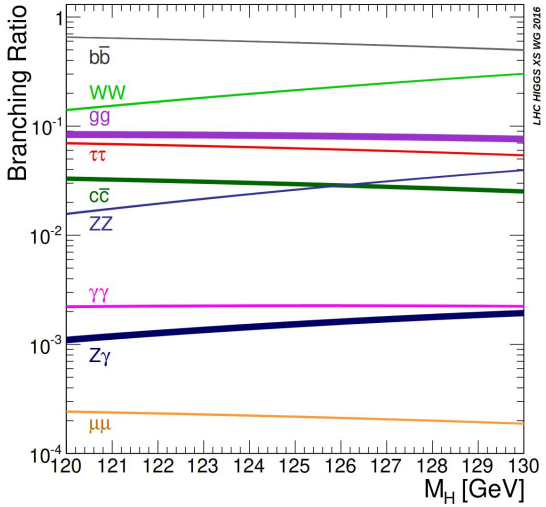
b-tagging



- Many interesting processes produces b-jets e.g. HH→ 4b or beyond SM physics
- Identifying events with b-jets in the final state at trigger level is crucial given the overwhelming QCD background
- A typical feature of b-jets is existence of secondary vertex and displaced tracks due to long lifetime of B hadrons



Credits: Universität Siegen



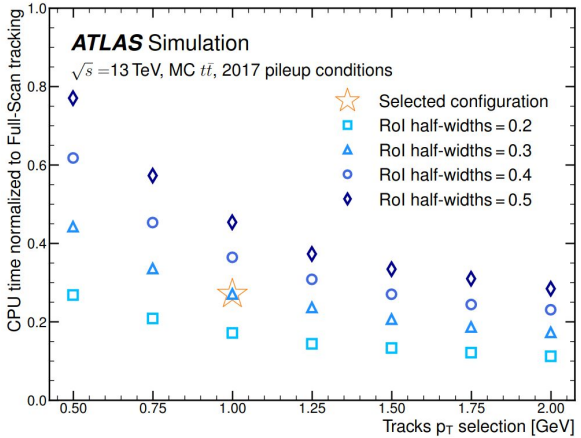
- b-tagging algorithms requires track reconstruction which is CPU intensive

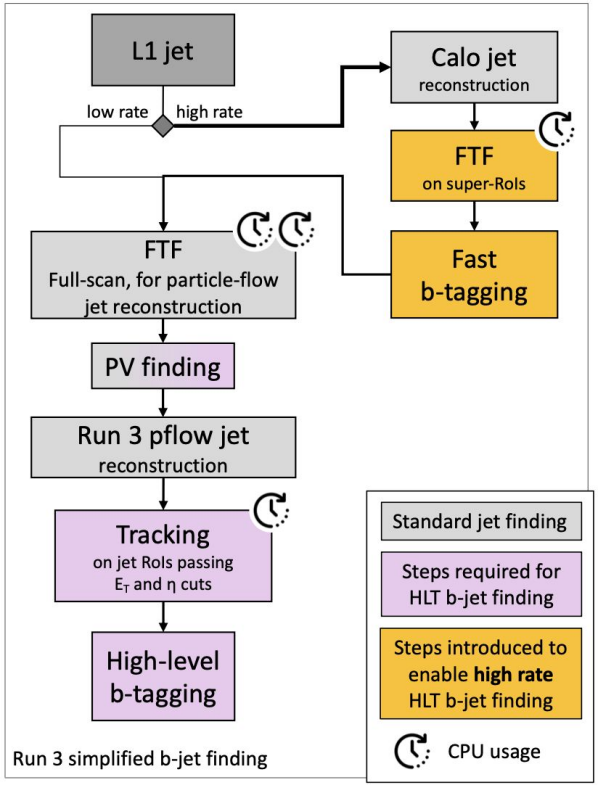
⚠ **Bottleneck for high rate trigger reconstruction**

Fast b-tagging at the HLT

- Reject background as early as possible
- Run b-tagging before Full detector acceptance tracking

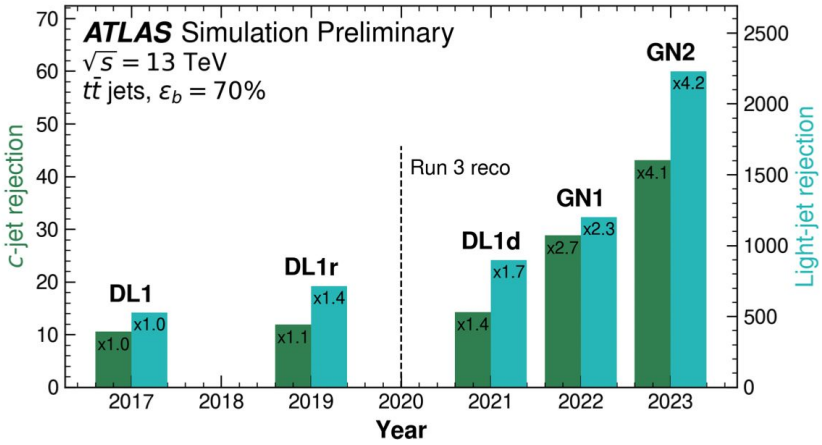
  → Saves CPU, enables complex reconstruction downstream
- Use ML models to infer the flavour of the jet using coarser inputs
  - "*Fast*" tracking only, in narrow regions of interest





Recent history of b-tagging in ATLAS

- First implementation of fast b-tagging was done using a Deep Sets architecture (DIPS)
- Offline analysis showed the success of GNNs for flavour tagging



Run 3 Taggers

DL1d is a combination of DNNs

GN1 is a Graph Attention Network

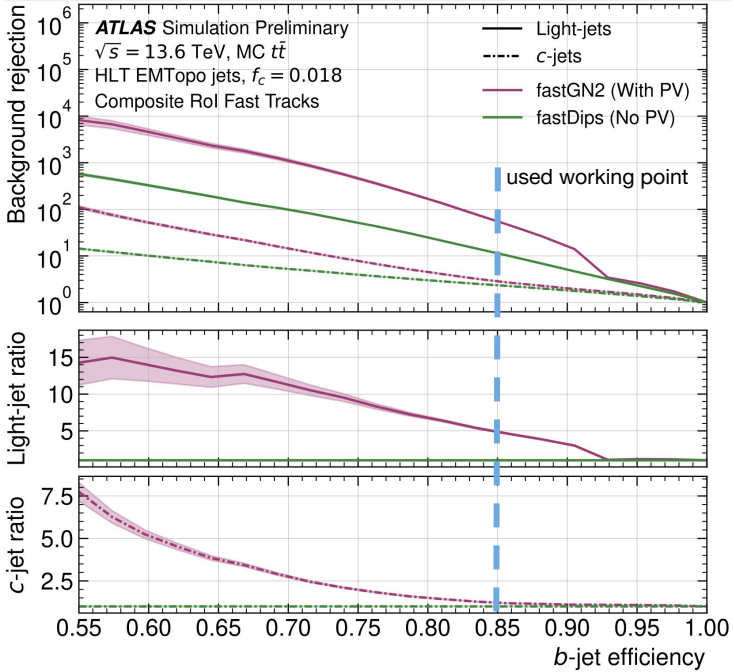
GN2 is based on a Transformer architecture with Multi-Head Attention

GNNs for fast b-tagging

- GN2 models have been trained for trigger b-tagging, currently running for 2024 physics data-taking
- **Factor ~5** improvement in background rejection with GN2 compared to previous implementation
  - PV information also helps to increase performance
- Huge impact on CPU resources
  - **- 20% CPU time** per event savings overall!

Fast b-tagger	Time per event [ms]
fastDips	1164
fastGN2	929,2

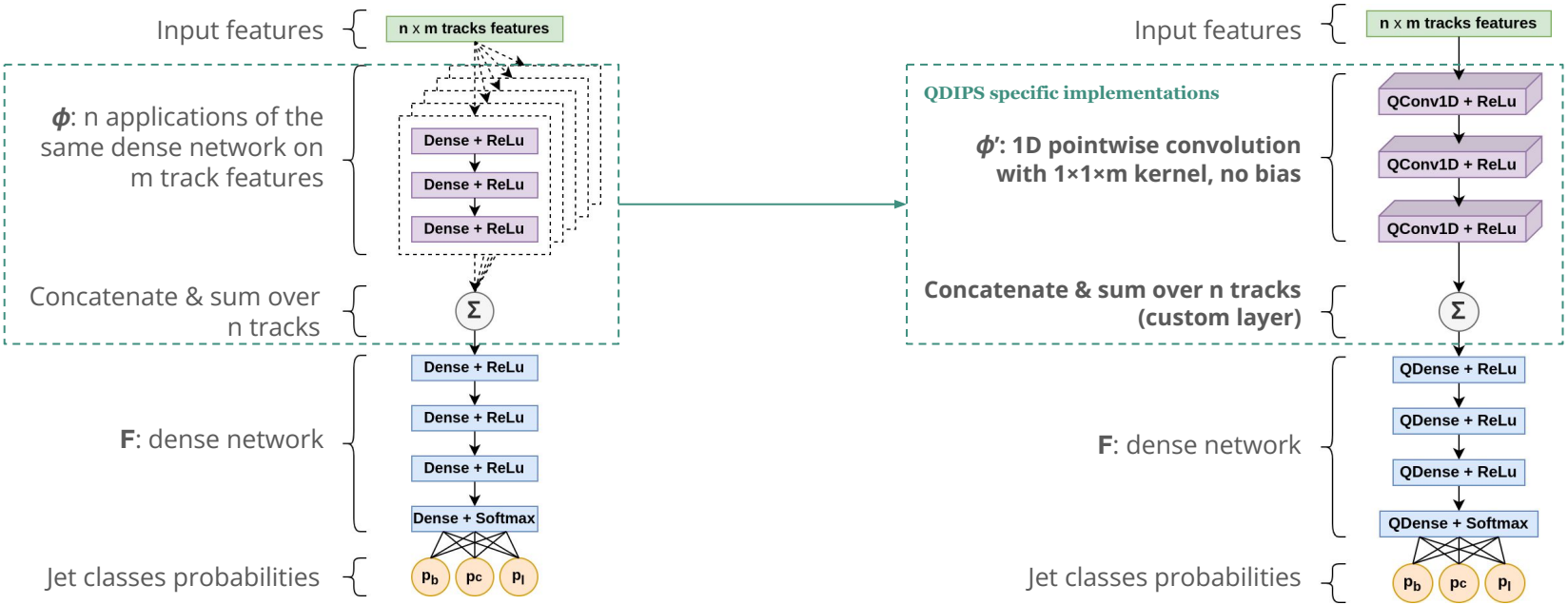
testing same trigger chain, but different taggers



fastDips is the Deep Sets model used for 2022 and 2023

Fast b-tagging for upgrade trigger

- Big restructuring of the trigger system for HL-LHC
- Explorations for fast b-tagging, accelerate with implementation in FPGA
  - Need to quantize weights and slightly adapt the architecture ⇒ **QDIPS**

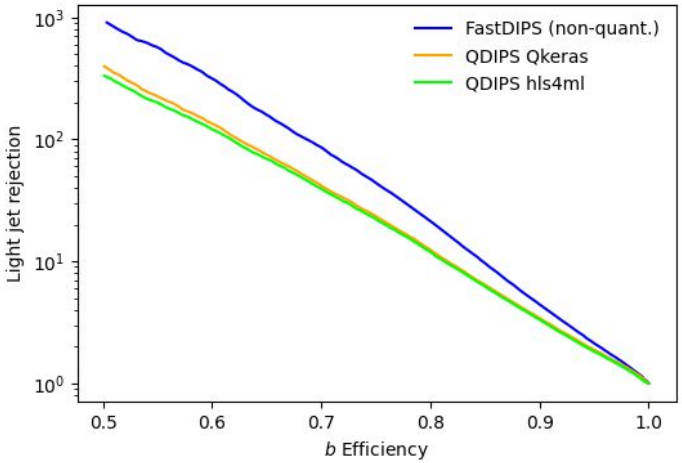


Fast b-tagging for upgrade trigger

- Training with same dataset of fastDIPS
  - Fully implemented and tested model on FPGA

QKeras + hls4ml

- Reducing model size by roughly one order of magnitude at the cost of small loss in performances
- Still work in progress, many possible improvements
  - hope to achieve performances closer to fastDIPS keeping the model small and hardware friendly

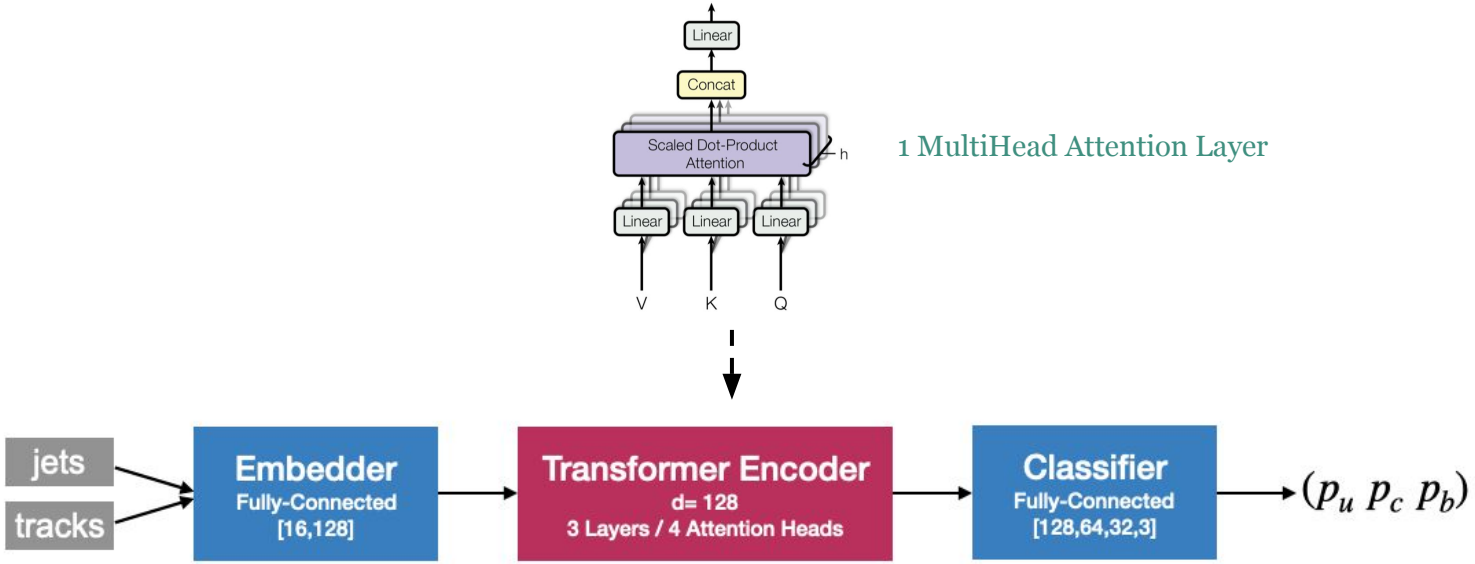


(more on FPGAs in Kostas talk)

Conclusion

- Rejecting background in trigger as early as possible is essential for data-taking nowadays and even more for HL-LHC
    - Fast b-tagging enabled to collect di-Higgs events at a higher rate than ever before!
- More performant taggers were successfully deployed in trigger, following closely offline b-tagging improvements
- The success of fast b-tagging in LHC Run 3 encourages the studies for improving these strategies for HL-LHC, e.g. leveraging hardware acceleration

Backup: GN2 in HLT



- Trained with 180M jets from simulated ttbar samples and 18M from Z' sample