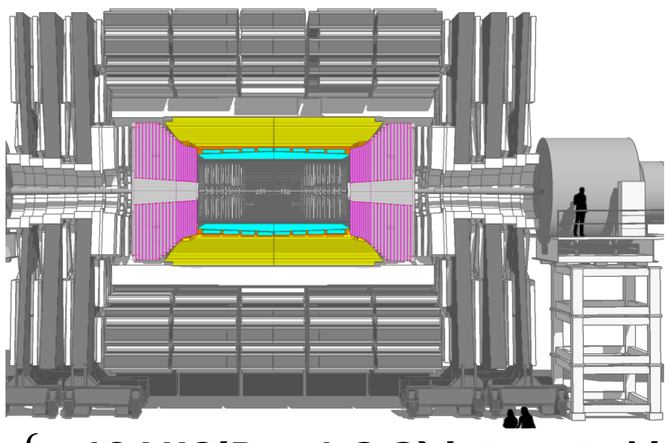
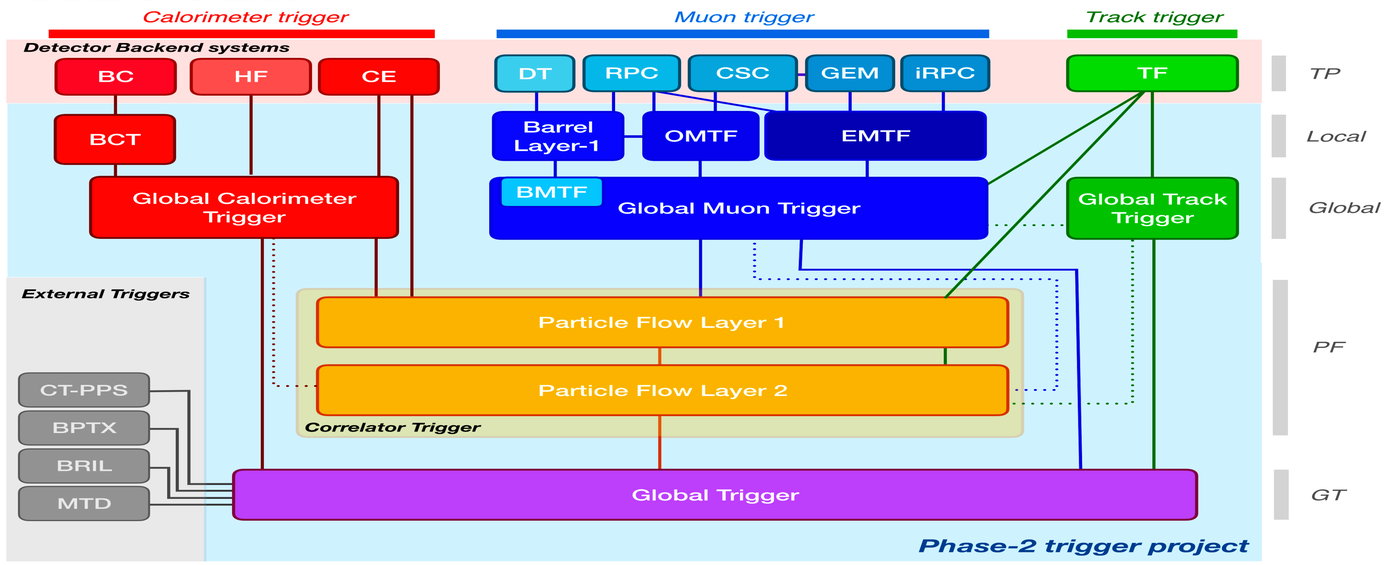
Development and firmware implementation of a machine learning based hadronic tau lepton Level-1 Trigger algorithm in CMS for the HL-LHC

Jona Motta*, on behalf of the CMS Collaboration - Poster Session at the 2022 November LHCC meeting

HL-LHC AND THE CMS PHASE 2 LEVEL-1 TRIGGER



40 MHz →



HL-LHC = { × 10 LHC(Run 1-2-3) integrated luminosity; × 2.5 LHC(Run 3) instantaneous luminosity; unprecedented radiation and pile-up (PU) }

CMS detector will undergo large upgrade campaign
→ relevant for this work are:
* the upgraded barrel backend electronics
* the new HGCAL endcap calorimeter

* state-of-the-art FPGA broad use
* high-speed optical links broad use
* highly modular architecture

* new Correlator and Track Trigger
* increased latency
* HGCAL 3D-clusters (CL3D) inclusion

The Phase-2 Level-1 Trigger upgraded system

THE TAUMINATOR ALGORITHM

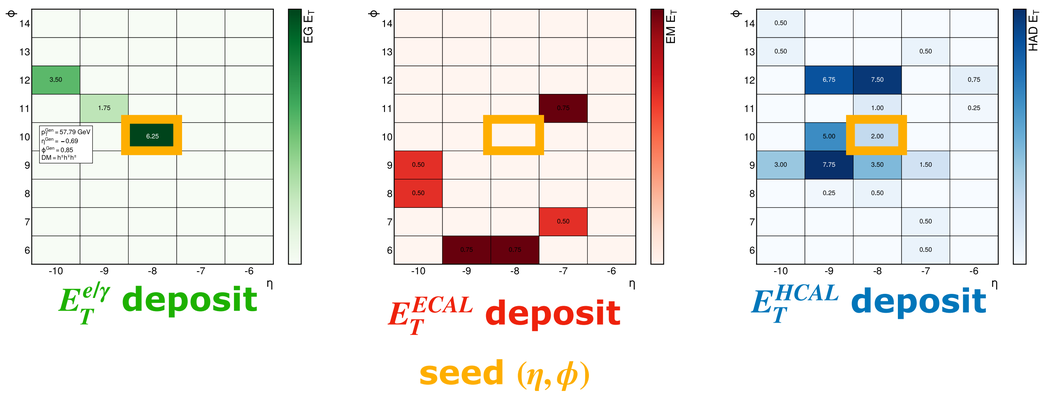
Algorithm development

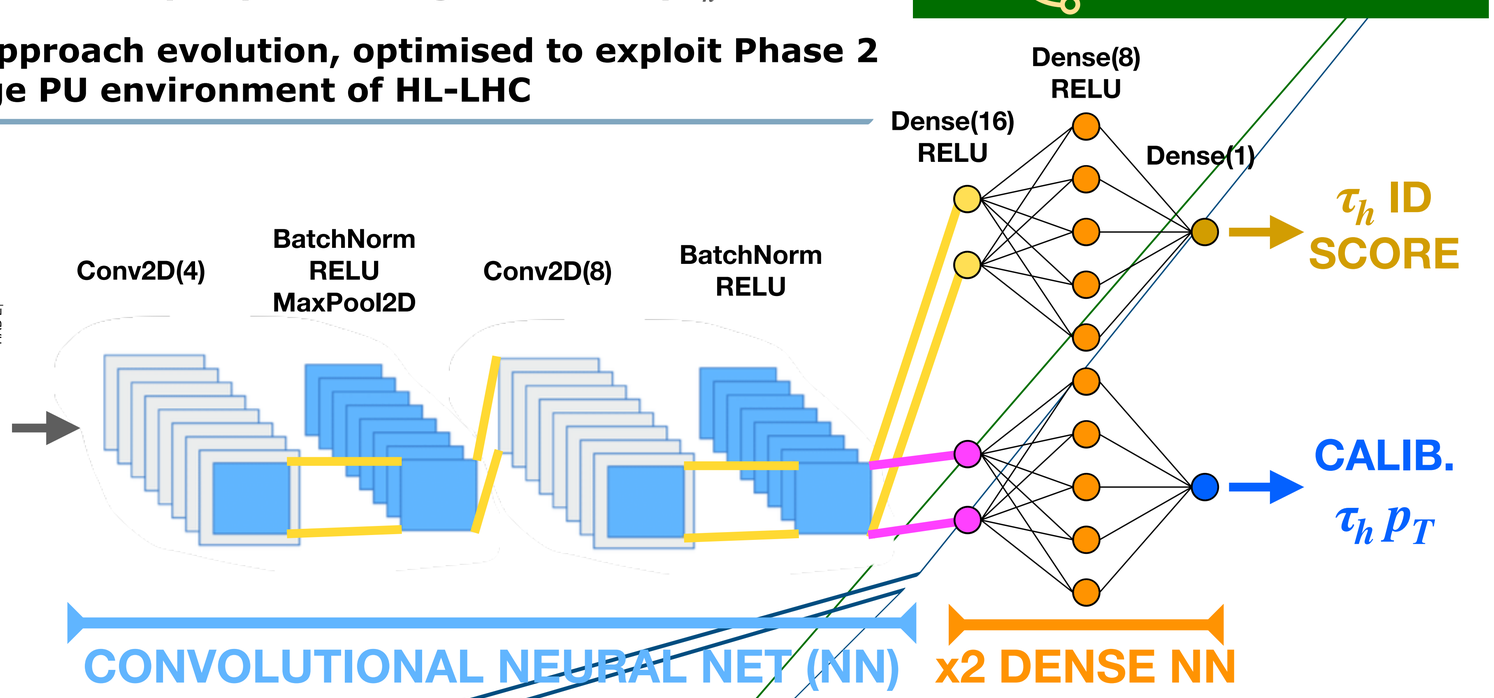
MINATOR

Hadronically decaying taus (τ_h) are challenging objects and need clever treatment of decay multiplicity
Run 2/3 reconstruction: dynamical Calorimeter Trigger Towers (TTs) clustering followed by τ_h isolation

TauMinator algorithm presented here: is the Run 2/3 approach evolution, optimised to exploit Phase 2 Level-1 Trigger features optimally to cope with the large PU environment of HL-LHC

Over the full calorimetric $|\eta| < 3.0$ coverage
τ_h reconstructed as $(\eta, \phi) = 5 \times 9$ TTs clusters (CL5X9)
build from a seed TT with energy deposit $E_T > 2.5GeV$

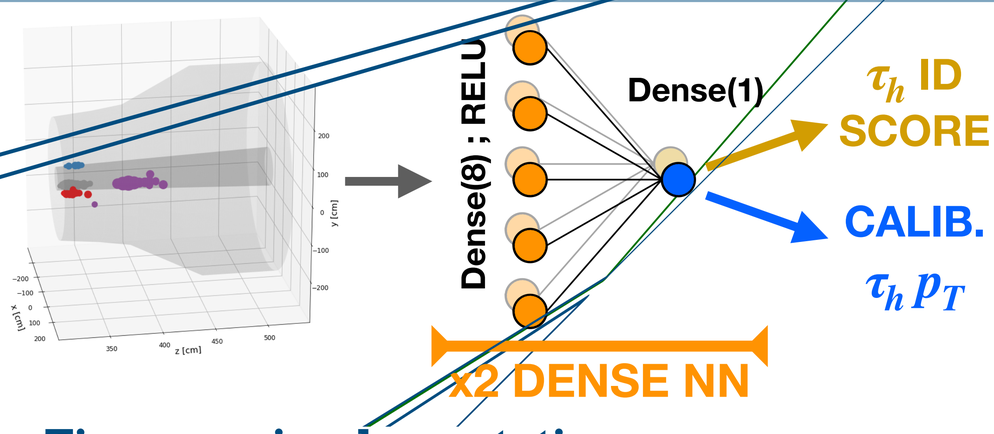




In the HGCAL detector $1.5 < |\eta| < 3.0$
τ_h also reconstructed as single CL3D with $E_T > 4GeV$

Several shape features computed for each CL3D in HGCAL backend and sent to Level-1 Trigger

Optimal set of variables for DNN is selected via random backward skimming strategy



+

Use the HGCAL detector in a *tracker-like* manner: geometrically match CL3D cluster to CL5x9 after passing ID selection

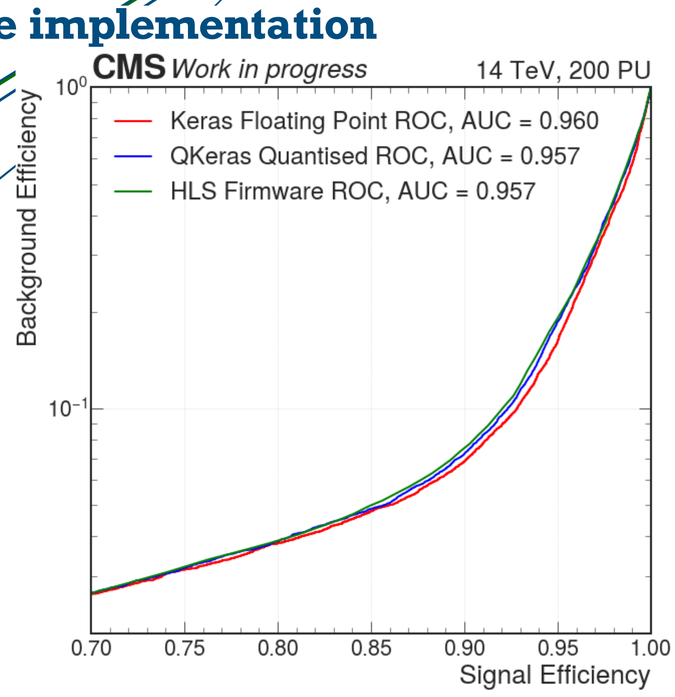
Firmware implementation

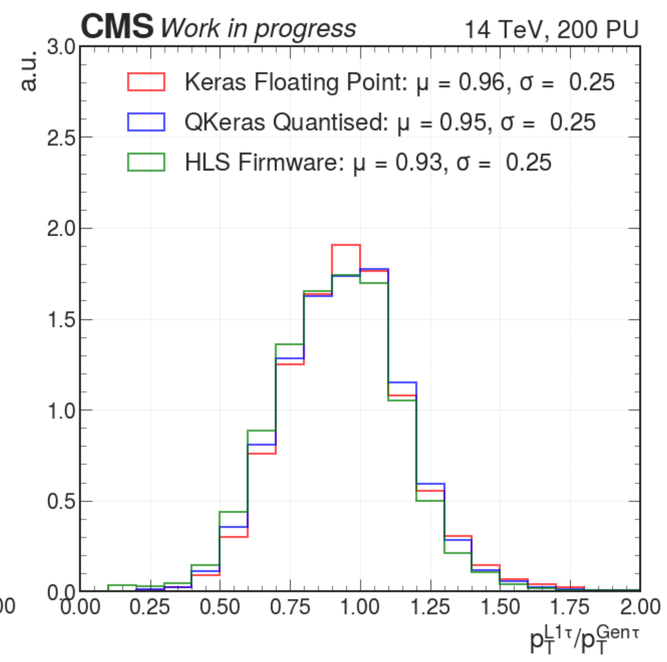
Foreseeing FPGA firmware (FW) implementation all models are developed with quantisation-aware training and pruned

CNN FW implementation aims for lowest inference latency
DNN FW implementation aims for lowest FPGA resources usage

Resources usage evaluated on Xilinx VU13P FPGA board (board foreseen for Phase 2 Global Calorimeter Trigger)

	LUT [%]	CLB [%]	DSP [%]	Latency [ns]	Interval [ns]
CNN	1.08	0.40	0.00	100	55
CL5x9 ID DNN	0.43	0.05	0.17	60	15
CL5x9 CALIBRATION DNN	0.45	0.06	0.61	40	15
CL3D ID DNN	0.20	0.03	0.22	40	13
CL3D CALIBRATION DNN	0.26	0.04	0.76	30	20





No significant performance loss from Keras floating point model, through QKeras quantised model, to HLS FW model for both τ_h ID and calib.

CONCLUSIONS AND OUTLOOK

* τ_h challenging object: this innovative approach exploits combination of NNs for optimal τ_h reconstruction, identification and calibration
* The Phase 2 Level-1 Trigger permits complex NNs FPGA FW implementation while ensuring modest resources usage and inference latency
* The algorithm showcases satisfactory performance (such as promising rate reduction) w.r.t to the current calorimeter-based algorithm
* The ongoing and next steps include: deeper physics and rate performance studies, full firmware implementation, and Vivado simulation

* jona.motta@cern.ch - Laboratoire Leprince-Ringuet, École Polytechnique, Institut Polytechnique de Paris, Rue de Fresnel, 91128 Palaiseau (France)