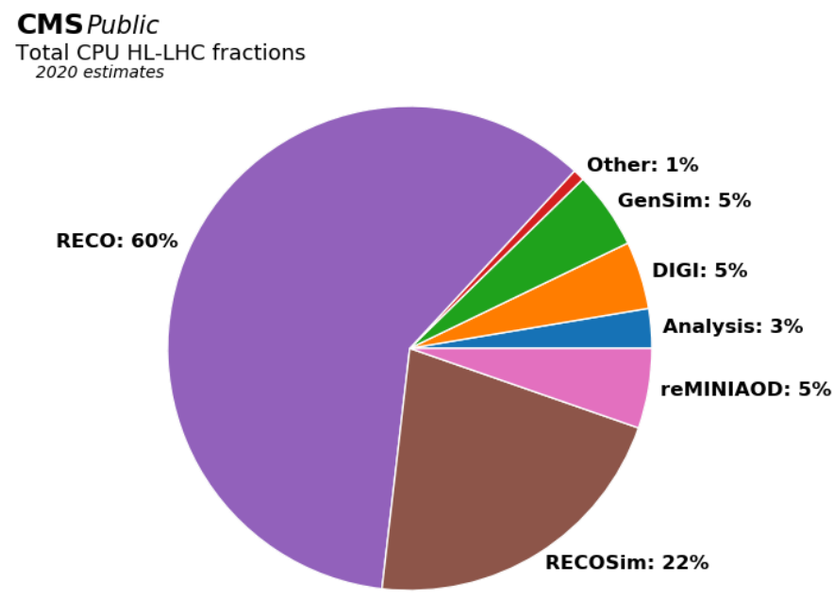
CaloChallenge: A Proposal

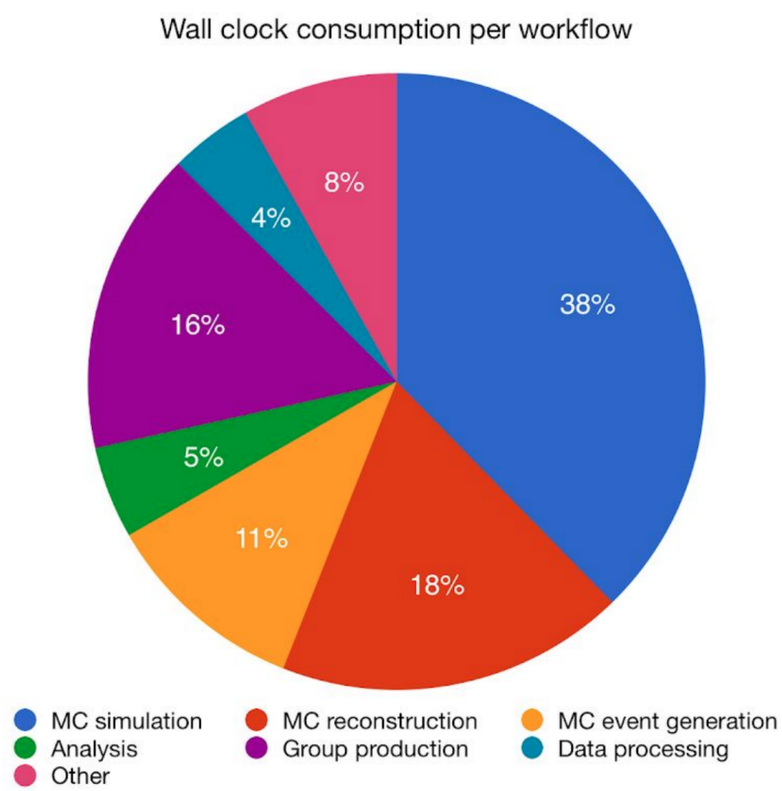
David Shih

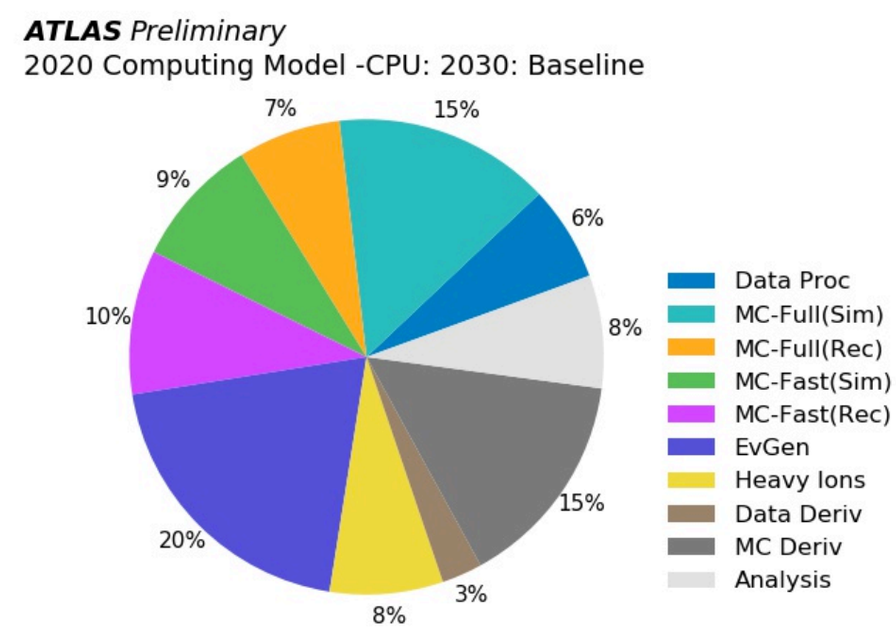
on behalf of Michele Faucci Giannelli, Gregor Kasieczka,
Ben Nachman & Dalilia Salamani

ML4Jets 2021

Detector Simulation is Expensive





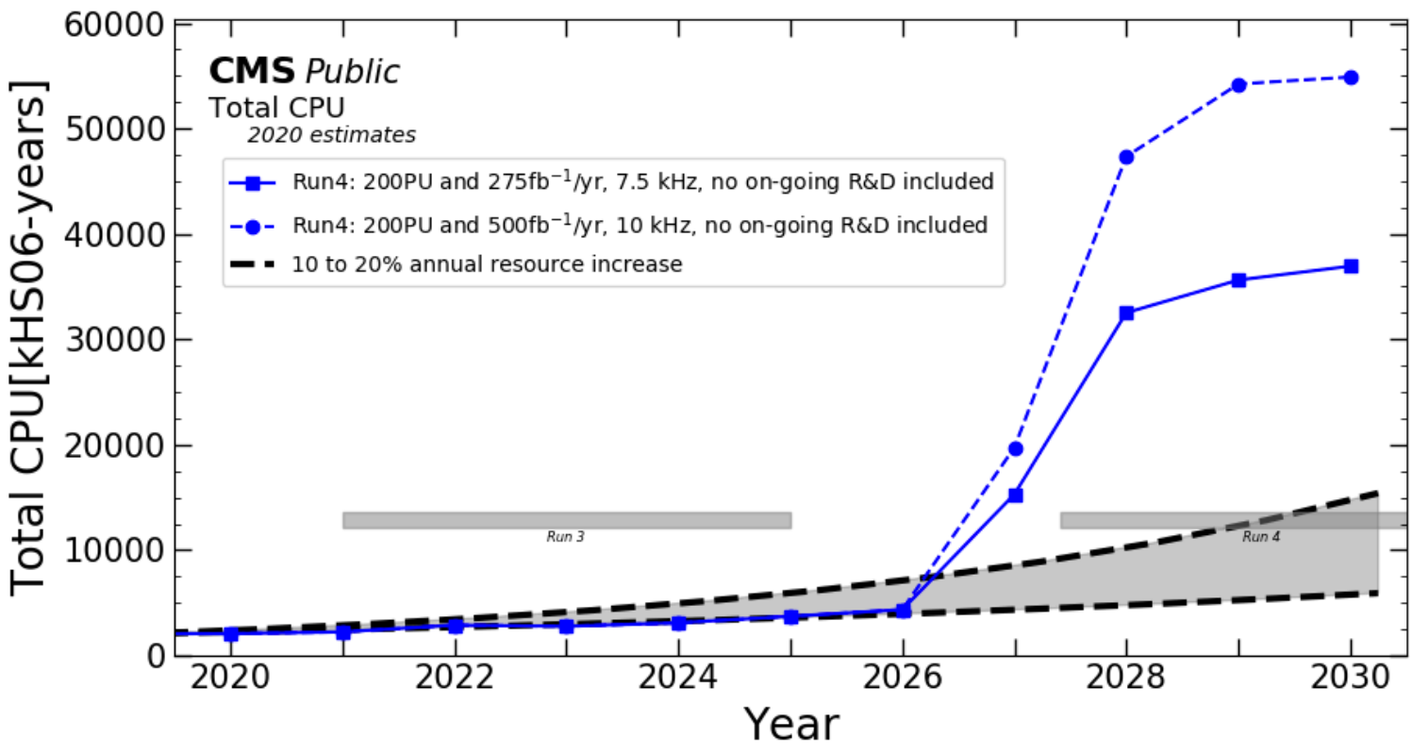


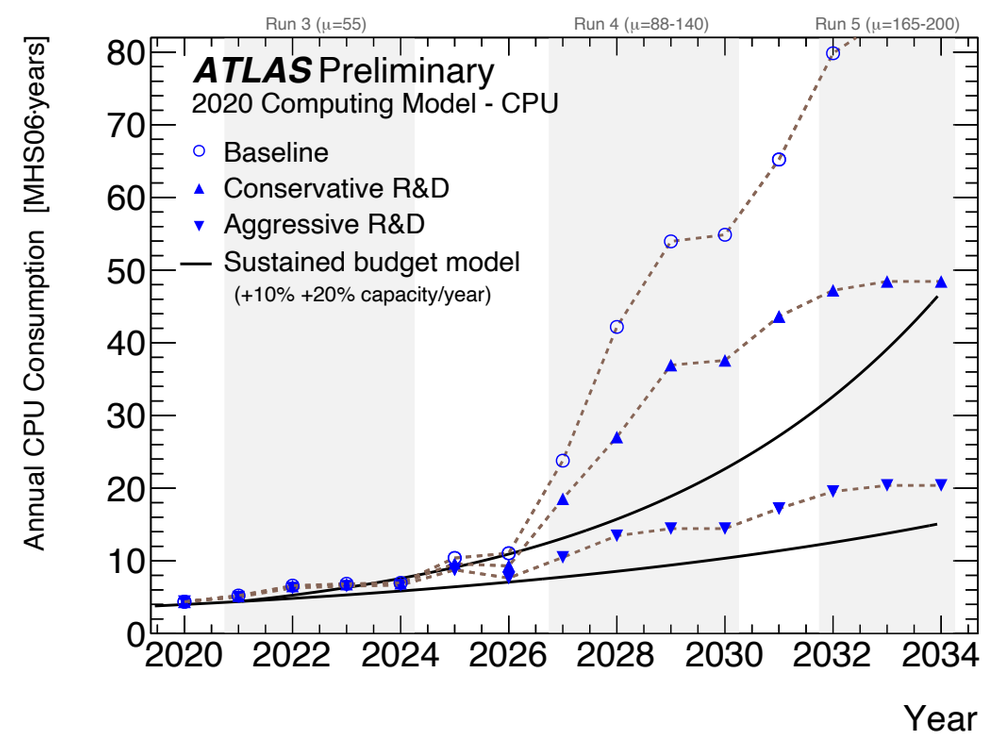
CERN-LHCC-2020-015

https://twiki.cern.ch/twiki/bin/view/CMSPublic/CMSOfflineComputingResults

https://twiki.cern.ch/twiki/bin/view/AtlasPublic/ComputingandSoftwarePublicResults

Without major R&D, computing needs at HL-LHC will far outstrip available resources



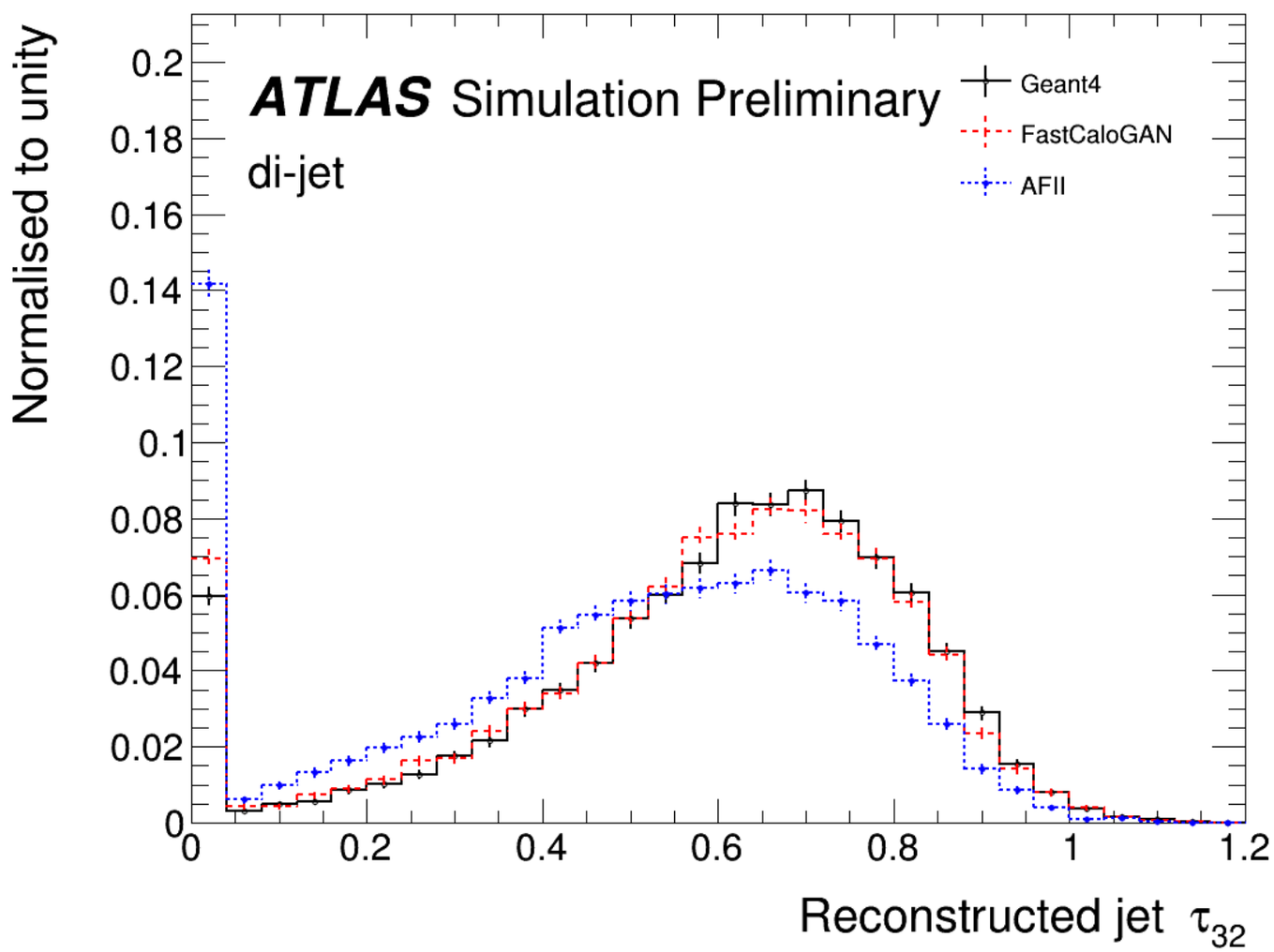


There is a major need for fast and accurate alternatives to GEANT4 calorimeter simulation!

Fast CaloSim with Deep Learning

Actually, regardless of computing resources, there is a lot of potential for improving fast calorimeter simulation with deep learning

from ATL-SOFT-PUB-2020-006



Fast CaloSim with Deep Learning

Classification | Datasets | Regression, Calibration, and Fast | Simulation and Generative Models

14:00	**CaloFlow: Fast and Accurate Generation of Calorimeter S…** *Dr Claudius Krause*	**Machine learning based Particle Flow algorithm and appl…** *Sanmay Ganguly*
	Multi-detector geomotery modeling and Geant4 Integration *Dalila Salamani*	**Using Machine Learning for Heavy-Ion Jet $p_{\rm T}$ R…** *Hannah Bossi*
	Angular Conditioning of Deep Generative Models for Fast… *Peter McKeown*	**Learning Uncertainties the Frequentist Way: Calibration …** *Rikab Gambhir*
15:00	**Fast and Accurate Electromagnetic and Hadronic Shower…** *Engin Eren*	**ML in jet physics beyond classification** *CMS Collaboration* 15:00 - 15:20
	AtlFast3: The next generation of fast simulation in ATLAS *Joshua Falco Beirer*	**Pileup mitigation in CMS** *CMS Collaboration* 15:20 - 15:40
	Fast Simulation of Jets with VAEs *Mary Touranakou* 15:40 - 16:00	**Lightweight Jet Reconstruction as an Object Detection T…** *Adrian Alan Pol*
16:00	**Foundations of a Fast, Data-Driven, Machine-Learned Sim…** *Jessica N. Howard et al.*	**Measurement of Muon Energy From Radiative Losses in …** *Giles Chatham Strong*
	Particle Cloud Generation with Message Passing GANs *Raghav Kansal*	**Deep learning jet modifications in heavy-ion collisions** *Dr Yilun Du*
	White Box AI for parton shower development *Felix Ringer* 16:40 - 17:00	**Matrix Element Calculations on the GPU** *Joshua Isaacson* 16:40 - 17:00
17:00	**How to GAN Event Unweighting** *Mr Mathias Backes* 17:00 - 17:20	**Jet Identification in L1 Trigger at HL-LHC based on DNN i…** *Dr Andre Sznajder*
	Sparse Data Generation with Convolutional VAE *Breno Orzari*	**OnlineFlow: Trigger Free Analysis Using Online Learned …** *Sascha Daniel Diefenbacher*
	Super-Resolution for QCD and Top Jets *Lukas Blecher* 17:40 - 18:00	

A lot of exciting work on this topic!

Existing approaches

Many different approaches have been tried for fast GEANT4 emulation with deep learning:

- GANs
  Paganini, de Oliveira & Nachman 1712.10321
  Belayneh et al 1912.06794

- WGANs
  Erdmann et al 1802.03325, 1807.01954
  ATL-SOFT-PUB-2018-001, ATL-SOFT-PUB-2020-006

- Bib-AE
  Buhmann et al 2005.05334, 2102.12491

- VAEs
  ATL-SOFT-PUB-2018-001

- Normalizing Flows
  Krause & DS 2106.05285

- … plus many others that focus on other generative modeling tasks, e.g. jet images, cf HEP-ML living review

Existing approaches

Many different approaches have been tried for fast GEANT4 emulation with deep learning:

- GANs
  Paganini, de Oliveira & Nachman 1712.10321
  Belayneh et al 1912.06794

- WGANs
  Erdmann et al 1802.03325, 1807.01954
  ATL-SOFT-PUB-2018-001, ATL-SOFT-PUB-2020-006

- Bib-AE
  Buhmann et al 2005.05334, 2102.12491

- VAEs
  ATL-SOFT-PUB-2018-001

- Normalizing Flows
  Krause & DS 2106.05285

- … plus many others that focus on other generative modeling tasks, e.g. jet images, cf HEP-ML living review

But for the most part, each study uses its own dataset, making it hard to benchmark and compare different approaches

Proposal

A community challenge, based on a common dataset, for developing and benchmarking different approaches to fast calorimeter simulation

Modeled after:

- Top tagging community study
  Kasieczka, Plehn (eds) et al [https://arxiv.org/abs/1902.09914]

- LHC Olympics 2020
  Kasieczka, Nachman & DS (eds) et al [https://arxiv.org/abs/2101.08320]

Dataset

Dataset

Dataset should be

Dataset

Dataset should be

- Realistic
  - ECAL and HCAL
  - Realistic resolution
  - Multiple particles (eg e+ and pi+)
  - Voxels and cells?

Dataset

Dataset should be

- Realistic
  - ECAL and HCAL
  - Realistic resolution
  - Multiple particles (eg e+ and pi+)
  - Voxels and cells?
- Manageable
  - Not so high dimensional that one needs TBs of storage or a GPU cluster or weeks-long training
  - Maybe just the central eta slice of the detector?

Dataset

Dataset should be

- Realistic
  - ECAL and HCAL
  - Realistic resolution
  - Multiple particles (eg e+ and pi+)
  - Voxels and cells?

Something like the dataset used in ATL-SOFT-PUB-2020-006?

- Manageable
  - Not so high dimensional that one needs TBs of storage or a GPU cluster or weeks-long training
  - Maybe just the central eta slice of the detector?

Dataset

Dataset should be

- Realistic
  - ECAL and HCAL
  - Realistic resolution
  - Multiple particles (eg e+ and pi+)
  - Voxels and cells?

Something like the dataset used in ATL-SOFT-PUB-2020-006?

- Manageable
  - Not so high dimensional that one needs TBs of storage or a GPU cluster or weeks-long training
  - Maybe just the central eta slice of the detector?
- Get buy-in from CMS and ATLAS (and other experiments? eg LHCb?)

Dataset

Dataset should be

- Realistic
  - ECAL and HCAL
  - Realistic resolution
  - Multiple particles (eg e+ and pi+)
  - Voxels and cells?

→ Something like the dataset used in ATL-SOFT-PUB-2020-006?

- Manageable
  - Not so high dimensional that one needs TBs of storage or a GPU cluster or weeks-long training
  - Maybe just the central eta slice of the detector?
- Get buy-in from CMS and ATLAS (and other experiments? eg LHCb?)
- Something which outside ML people could also participate in

Metrics

Challenge should have clearly defined metrics from the outset (and not too many).

Some ideas include:

- Ultimate classifier metric: classifier between real and generated samples
  see Krause & DS [https://arxiv.org/abs/2106.05285]
- Histograms
- Average Images
- Nearest Neighbor Tests
- Other metrics (Frechet distance, Wasserstein Distance…)
  see Kansal et al [https://arxiv.org/abs/2106.11535]
- Training time
- Evaluation time
- Memory usage

Timeline

Dataset in: ~ 2-3 months

Challenge duration: ~ 1 year

Results deadline: ~ 1 month before ML4Jets2022

Presentation of results at ML4Jets2022

Community input welcome!!