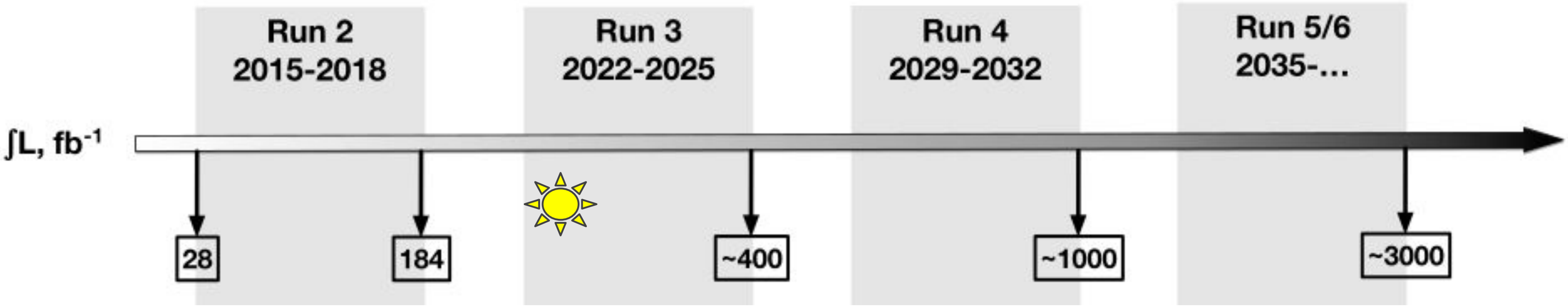
ATLAS towards HL-LHC: Physics, Methods and Tools

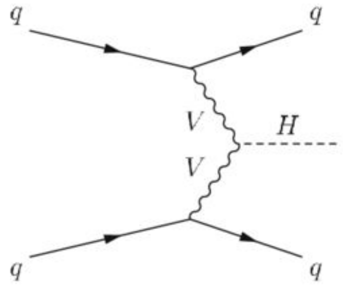
NorCC Activity 1 - Particle Physics

Physics Interests (Run 3 $\rightarrow$)

LHC timescale



Higgs Physics



<u>New sources of CP violation in Higgs to tau-tau</u>:

Our existence requires more CP violation than we presently see (**baryogenesis**)

Precision of 3-5^0 needed to test some of the baryogenesis models (10-5^0 expected after LHC Run 5)

<u>Precision Measurements of Higgs:</u>

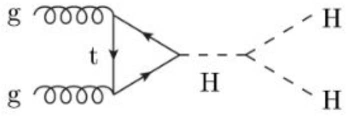
Improved background modelling

New approach to shape systematic uncertainties

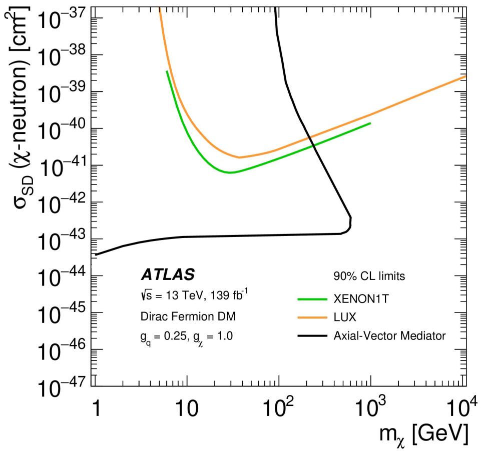
<u>Higgs self-coupling via double Higgs production</u>:

Higgs self-coupling through di-higgs production is basis to understand the Electroweak Phase Transition

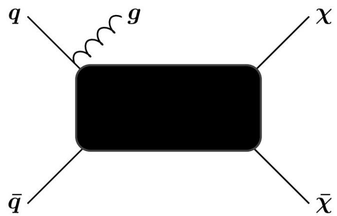
Present predictions: “Evidence” for hhh (if SM self-coupling) is end of LHC Run 5



Dark Matter Searches



Phys. Rev. D 103 (2021) 112006



Mono-jet signatures:

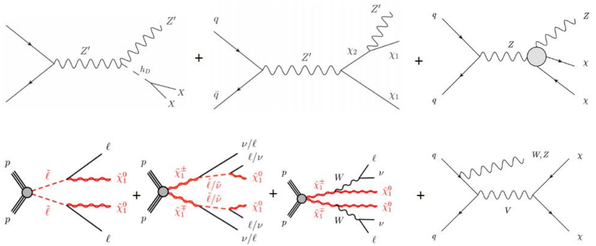
- Interpretation in SUSY models
- Use of ML

Tau signatures:

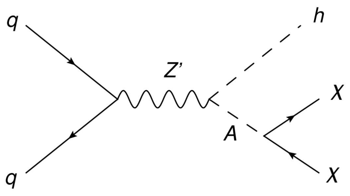
- supervised, semi-supervised and unsupervised ML techniques
- computer vision inspired techniques
- Lowering momentum threshold on taus [**improved reconstruction**]

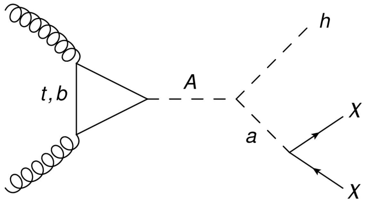
Electron/Muon (multi-lepton) signatures:

- Looking at several different DM models in di-lepton + missing transverse energy final state
- Developing more generic DL algorithm for (more) model independent searches



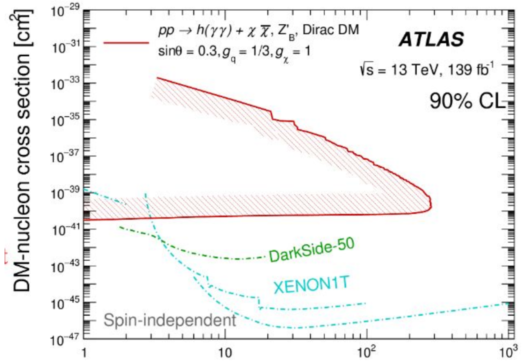
Dark Matter Searches cont’d



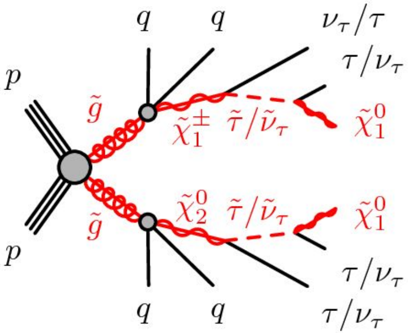


<u>Higgs and Dark Matter</u>:

- 2-Higgs Doublet Models (2HDM) + Z’ or A
- ATLAS published in H →γγ and H → bb-bar channels
- **Short term goal**: Publish on H →ττ [first draft in preparation]
- **Longer term**: combine all Higgs decay channels



New Physics with taus



Supersymmetry

Squarks and gluinos decaying to taus, jets, MET

Run 2 + 3 analysis underway

Scenario with m(q/g) ~ m(LSP) experimentally challenging but motivated by e.g. coannihilation

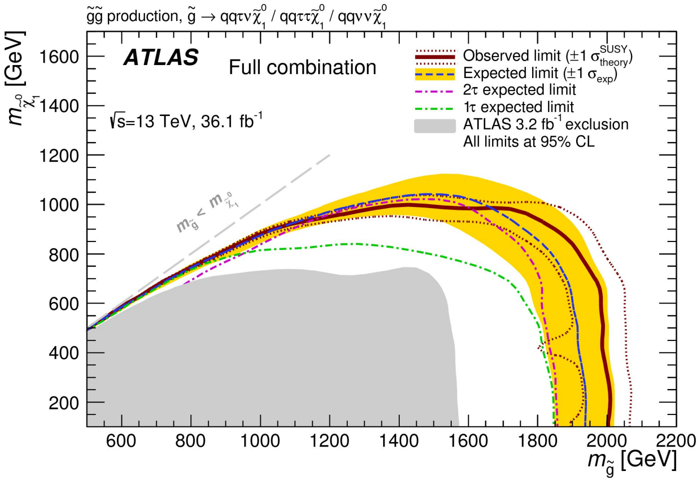
New resonance decaying to 2 collimated taus

Requires specific reconstruction when taus overlap - **algorithms being re-designed for Run3**

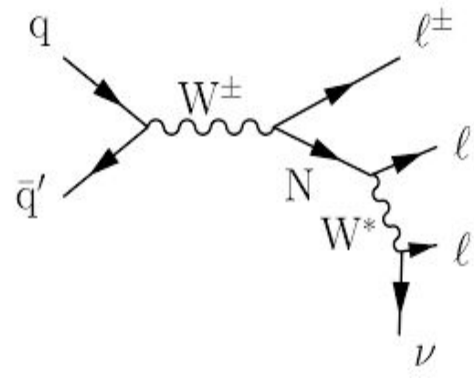
Low-mass resonance search (H→ aa, a→ tautau)

High-mass resonance search (X→ HH, H→ tautau)

Phys. Rev. D 99, 012009



New Physics with leptons



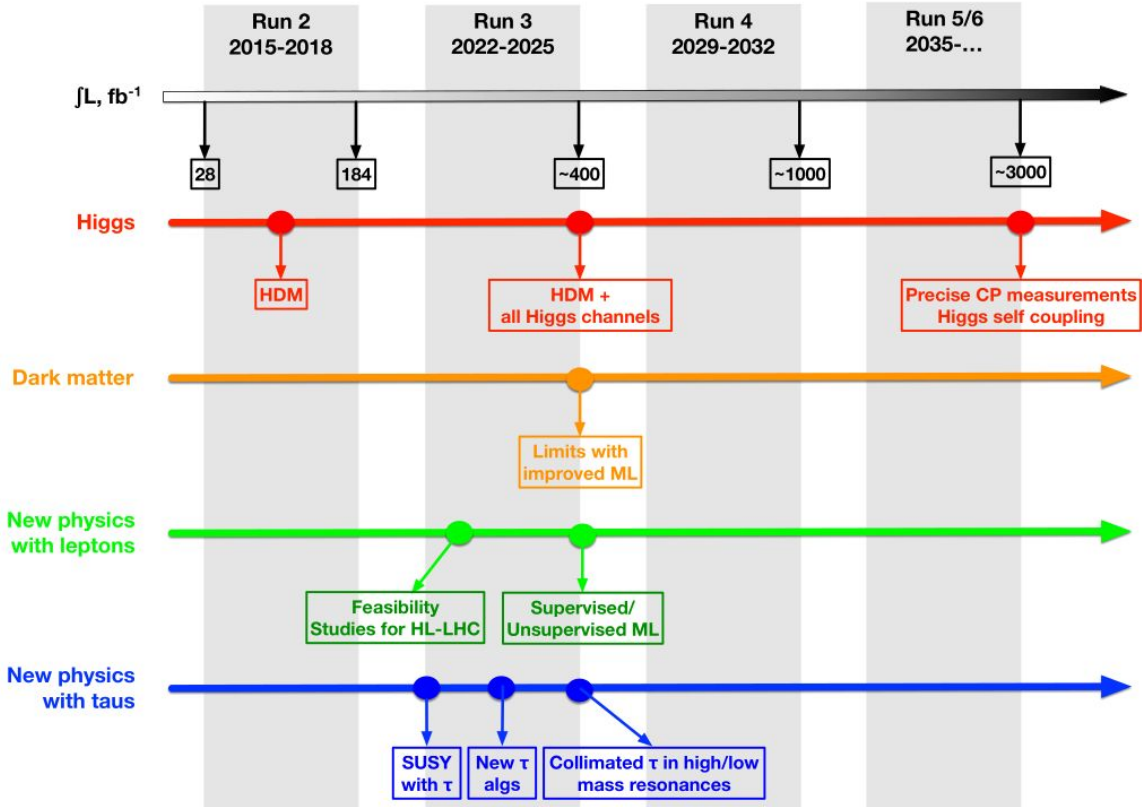
Searches for Heavy Neutrinos:

- Baryogenesis through leptogenesis
- Concentrating on the 3-lepton final state
- Comparison of supervised and semi-supervised (anomaly detection) DL a

Di-lepton studies in searches for new gauge bosons

- Run-3 and prospects for HL-LHC
- Study various beyond SM theories with new heavy gauge bosons

WIP: NorCC ATLAS Physics Roadmap



Common Approaches/Methods

Statistical Methods

Gaussian Process-based calculation of look-elsewhere trials factor: [arxiv:2206.12328]

- a recurring challenge to efficiently and precisely calculate the global significance of a potential new resonance
- Commonly used method not accurate for low significances (i.e. $\sigma < 3$)

New approach to shape systematic uncertainties

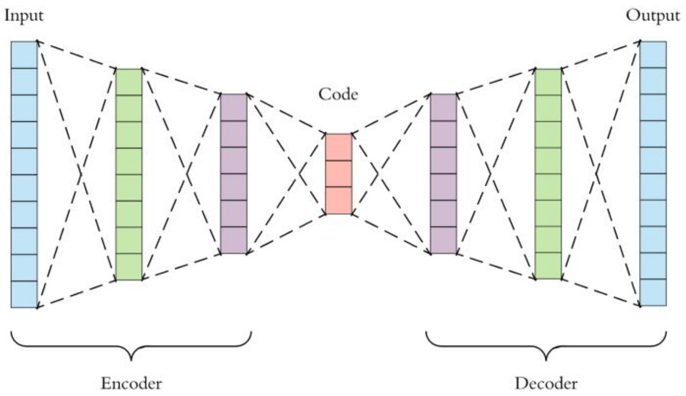
Reliable and Interpretable ML

More on this in the Computing, ML & AI Session on Thursday

Supervised, semi-supervised and unsupervised learning

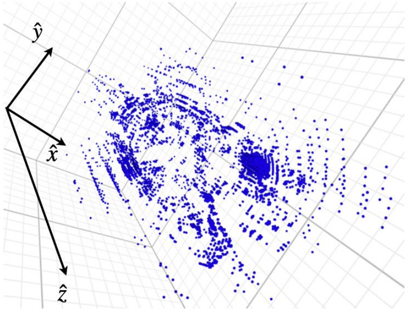
Build supervised and semi-supervised (anomaly detection) DL algorithms

- Compare performance by seeing how efficient the anomaly detector can identify “fake” signal in data compared with algorithms which have been trained on the same signal
- If the anomaly detector is efficient in picking out several different BSM signal - model independent approach
- Currently investigating Auto Encoders for anomaly detection



Credit: https://towardsdatascience.com/applied-deep-learning-part-3-autoencoders-1c083af4d798

Improved Tau Reconstruction



Jet performance group develops advanced techniques ("point cloud" methods) to reconstruct pions using calorimeter **and tracking** information:

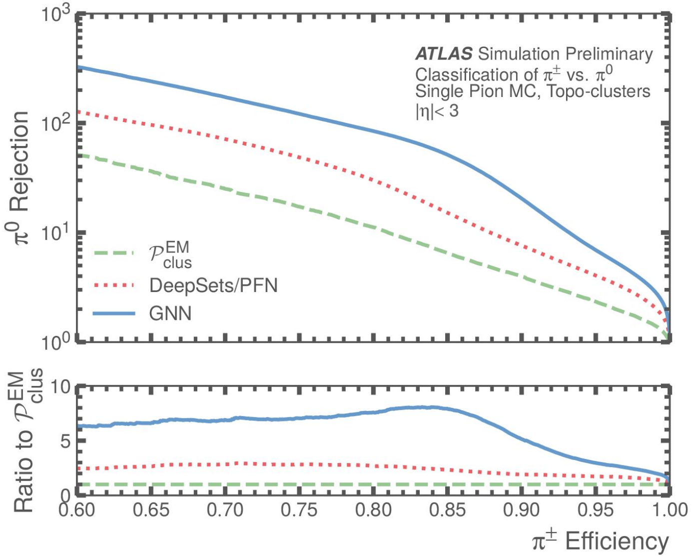
- classification: $\boldsymbol{\pi}^0$ vs $\boldsymbol{\pi}^{\pm}$
- regression: pion energy calibration

Graph Neural Networks outperform other methods (likelihood, deep set NN, convolutional NN), especially when combining clusters + cells + tracks.

Currently tested in "clean" environment (no pileup).

Promising for tau reconstruction (calorimeter cells available)!

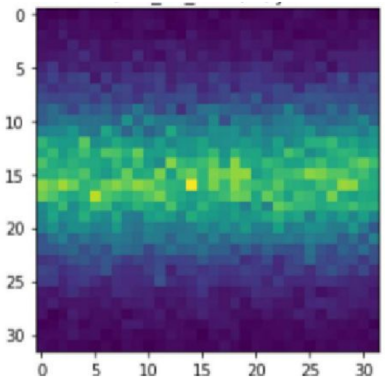
ATL-PHYS-PUB-2022-040



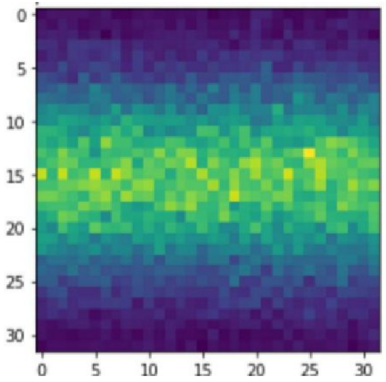
Computer Vision Inspired Techniques

Use of low level data of calorimetric deposits to distinguish between Black Holes and Sphalerons

Images of eta versus phi for jets



Black Hole



Spahleron

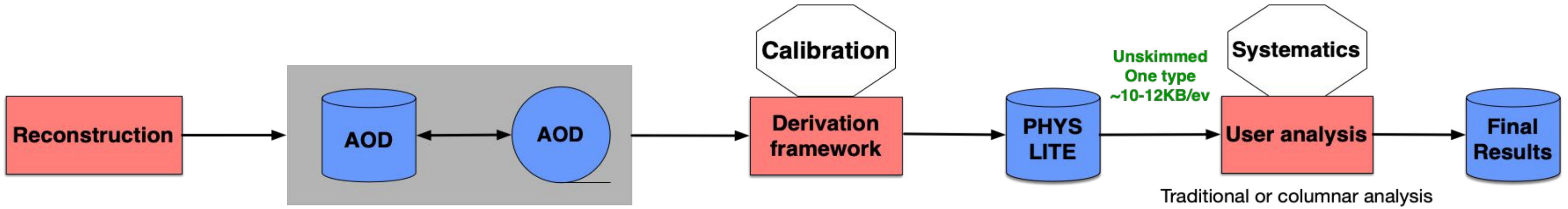
Tau reconstruction?

Dark Matter model indep. searches?

Common Tools and Infrastructures

Data formats for run 4

- DAOD_PHYSLITE is at the heart of the computing and analysis models for run 4. Almost all (~80%) of physics analyses will use these files as input
  - Written in the ATLAS xAOD data structure
  - Contains calibrated quantities and the variables required to evaluate instrumental systematic uncertainties
  - Target size: 10-12 KB/event
  - Significant development and maintenance input from NorCC personnel
- Allows more optimal use of disk due to its smaller size
- User analysis runs faster than current data formats due to lower I/O and pre-calibrated quantities
- xAOD format allows its use as an input for columnar analysis → next slide
- Commissioning will take place during run 3



From ROOT to Python Ecosystems

Most analysers using a common ATLAS format (xAOD PHYSLITE), converted into ROOT nTuples

With increasing use of ML there is a need for efficient parsing of data structure from **ROOT nTuples** to **Pandas DataFrames**

- Use of **RDataFrame**
  - parallelization comes for free
  - easy conversion into **numpy arrays** and **Pandas DataFrames**
  - Write to ML-friendly output formats like **hdf5**

Towards HL-LHC (i.e. Run 4 and onwards)

New data format:

The common ATLAS data format in Run 4 will be more light-weight, containing pre-calibrated objects only

- Less need for each analyser to run calibration tools (simplified workflow)
- Can be used directly from "the shelf"
  - Columnar analysis
  - Interactive Analysis (e.g. jupyter notebooks)
  - Direct access to data on local analysis facility

Expected updates to ROOT:

- A new nTuple format called **RNTuple** (optimized for RDataFrame)
- Moving to **ROOT7**

Exploitation of GPUs:

- Already being used in training of ML algorithms
- Use of novel data structures (CuPy)

Roadmap

LHC Run-3 will be a preparation towards Run-4 (HL-LHC)

Several interesting physics analyses

Where can we focus?

Particular development of new tau reconstruction algorithms

New workflows

Columnar and interactive analyses