DIFFERENTIAL PROGRAMMING FOR DETECTOR OPTIMISATION

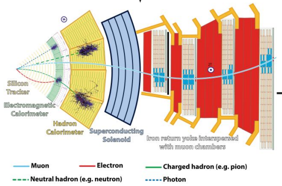
Giles Strong, on behalf of the MODE Collaboration*

Analysis Ecosystems II, IJCLab, France - 23/05/22

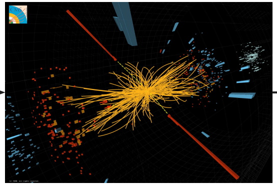
*https://mode-collaboration.github.io/#collaboration

TYPICAL LHC PROCESSING CHAIN

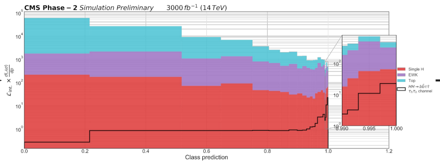
Particles
(real/simulated)



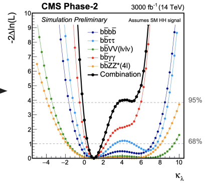
Detection
(real/simulated)
cds.cern.ch/record/2120661/



Reconstruction
cds.cern.ch/record/1406073



Analysis
CMS-FTR-18-019

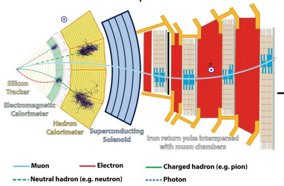


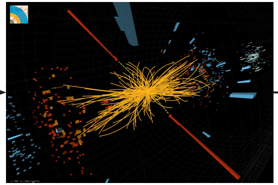
Measurement
CMS-FTR-18-019

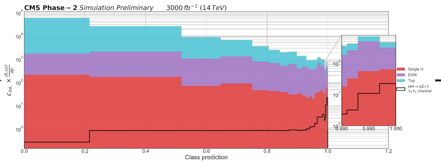
Each stage optimised separately

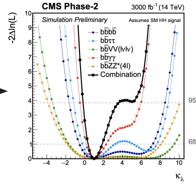
ISOLATED OPTIMISATION: PROXY OBJECTIVES

Particles (real/simulated)









Detection:

- Track first, destroy later
- Kinematic precision
- Dedicated sub-detectors
- Design convenience over analysis convenience

Reconstruction:

- Generic optimisation of algorithms
- Fixed working points
- Expert-interpretable data-representations (PID)

Analysis:

- Signal/background separation

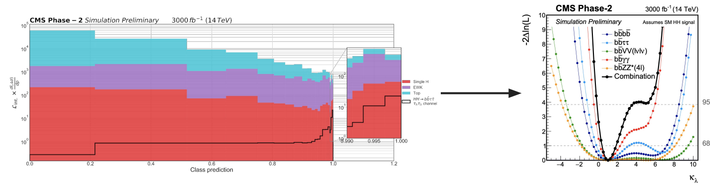
Measurement:

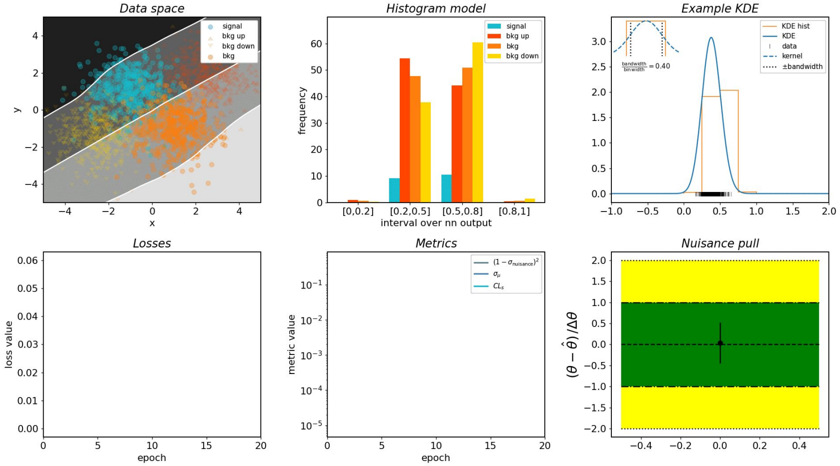
- Domain-driven categorisation
- Separate by decay channel, combine later

Many of these are “necessary evils” for HEP! Time, interpretation, MC corrections, etc.

PAIRED OPTIMISATION: ANALYSIS & MEASUREMENT

- Optimise analysis to directly optimise the measurement:
    - DNN training accounts for systematic uncertainties
    - Loss function monotonic w.r.t PoI uncertainty, CL_s limits, discovery significance, etc.
- INFERNO (de Castro & Dorigo, 2018)
- NEOS (Simpson & Heinrich, 2022)

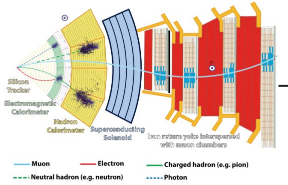


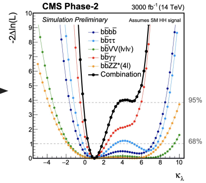


Animation: NEOS, Nathan Simpson

PAIRED OPTIMISATION: DETECTOR & MEASUREMENT

- CMS-FTR-18-019 projection study for HL-LHC di-Higgs sensitivity
- Analysis reused to test impact of new CMS timing detector in CMS-TDR-020
- But:
  - No reco. algo. re-optimisation: changes computed by simple rescaling
  - No analysis re-optimisation
  - Fixed test points





	Signal increase (%)		Expected significance	
Di-Higgs decay	BTL	BTL+ETL	No MTD	MTD
bbbb	13	17	0.88	0.95
bb$\tau\tau$	21	29	1.3	1.6
bb$\gamma\gamma$	13	17	1.7	1.9
bbWW			0.53	0.58
bbZZ			0.38	0.42
Combined			2.4	2.7

35ps (above) 50ps (below) timing resolution

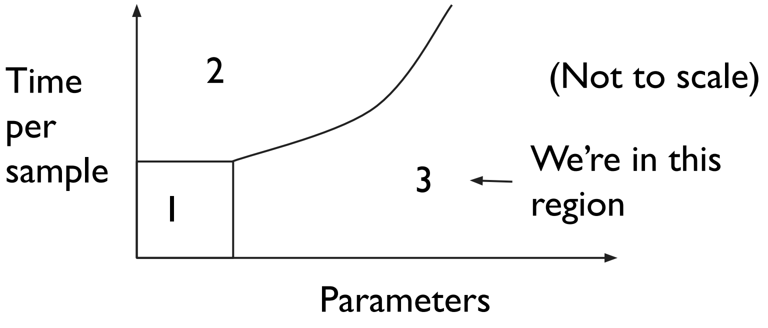
	Expected significance	
Di-Higgs decay	No MTD	MTD
bbbb	0.88	0.94
bb$\tau\tau$	1.3	1.48
bb$\gamma\gamma$	1.7	1.83
bbWW	0.53	0.58
bbZZ	0.38	0.42
Combined	2.4	2.63

Tables: CMS-TDR-020

MODE: WHAT IF…

- What if just like measurement-aware analysis-optimisation, we could go one step further:
- Measurement-aware detector-optimisation
- MODE mandate:
  - Make simulation & analysis chain differentiable
  - Specify physics goal as a loss function
  - Compute analytic dependence of performance on detector parameters
  - Design end-goal-optimal instruments
- Can it be achieved?
  - CERN LHC-style detectors = huge-parameter space + complicated simulation and analysis algorithms

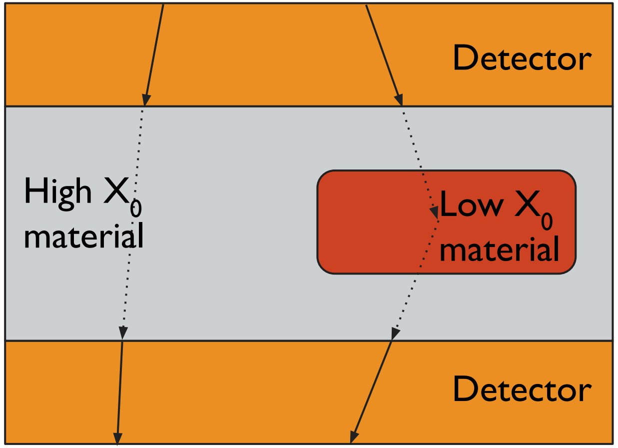
Let's start with a simple use-case: muon tomography



1. Grid/random search
2. Bayesian optimisation, Simulated annealing, genetic algorithm, particle swap optimisation, …
3. Gradient-based optimisation: Newtonian, gradient descent, BFGS, …

TOMOGRAPHY VIA MULTIPLE SCATTERING

- Consider a volume with unknown composition
  - E.g. Shipping container, archeological site, nuclear waste, industrial machinery
  - Want to infer properties of the volume:
    - E.g. build a 3D map of elemental composition
- Cosmic muons scattered by volume according to radiation-length (X_0 [m]) of elements in material
  - Measure muons above and below volume
  - Kinematic changes provide info on material composition



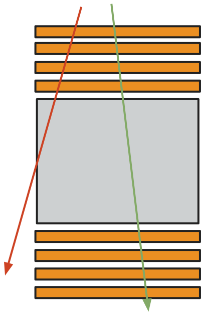
High X_0 = low scattering

Low X_0 = high scattering

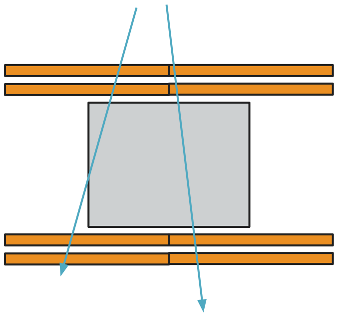
X_0 = average distance between scatterings

PROBLEM

- Each use-case likely to have a budget:
  - E.g. fiscal, heat, power, spatial, imaging time
- How should detectors be positioned to best function in each use case subject to constraints?
- Domain knowledge, experience, and intuition can help
  - But solutions likely to be based on heuristics and proxy objectives (e.g. lowest uncertainty on muon-path angles)



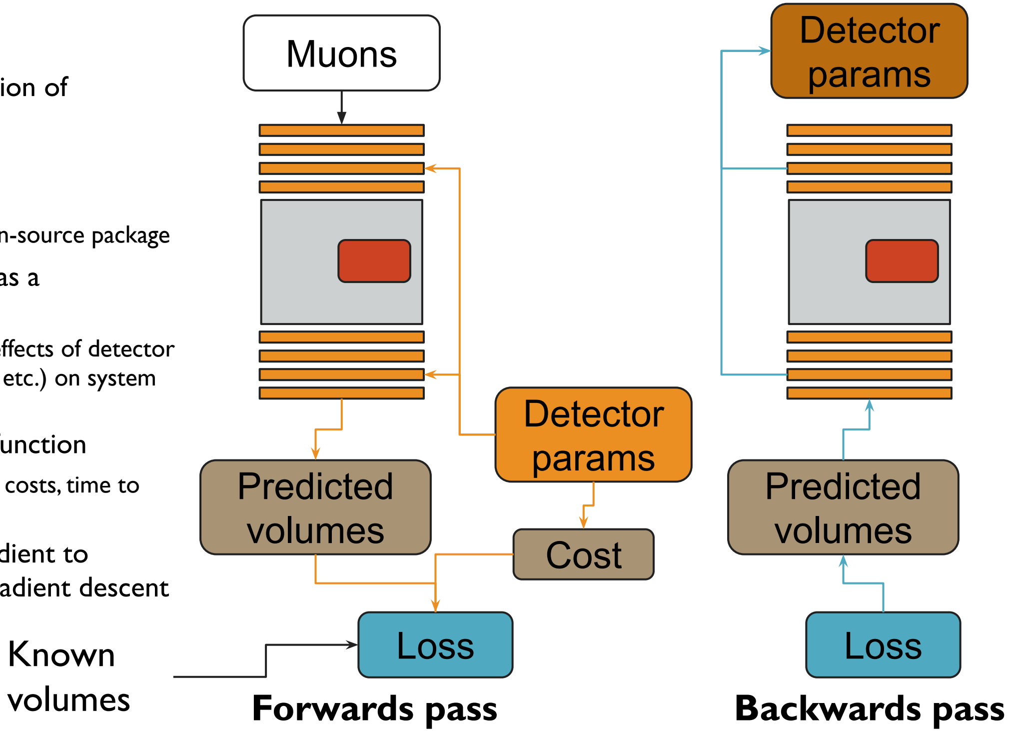
Example 1: Muons measured precisely but less efficiently



Example 2: Muons measured less precisely but more efficiently

TOMOPT

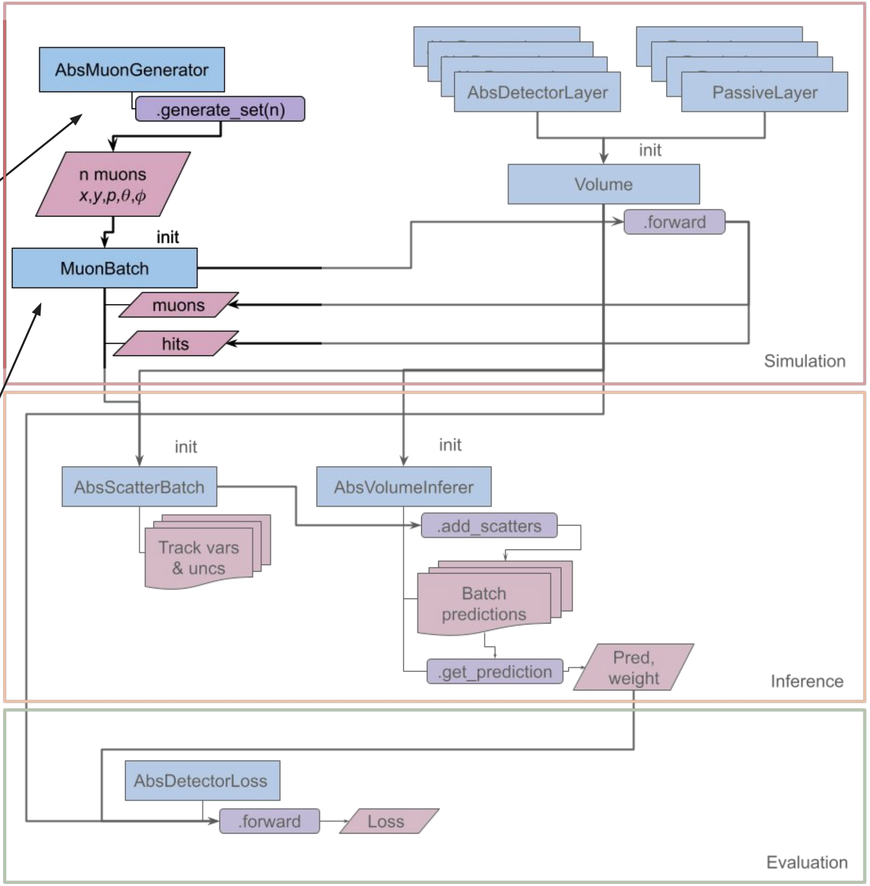
- Python package for differential optimisation of muon-tomography detectors
  - Modular design
  - PyTorch provides autodiff
  - Still underdevelopment; aim is an open-source package
- First, express the entire inference chain as a differentiable system
  - We can now compute the analytical effects of detector parameters (position, size, resolution, etc.) on system outputs
- Now express the desired task as a loss function
  - E.g. error on X_0 predictions, detector costs, time to achieve desired resolution
- We can now backpropagate the loss gradient to detector parameters and optimise via gradient descent
  - Just like a neural network



TomOpt contributors: Giles Strong, Tommaso Dorigo, Andrea Giammanco, Pietro Vischia, Jan Kieseler, Maxime Lagrange, Federico Nardi, Haitham Zaraket, Max Lamparth, Federica Fanzago, Oleg Savchenko, Nitesh Sharma, Anna Bordignon

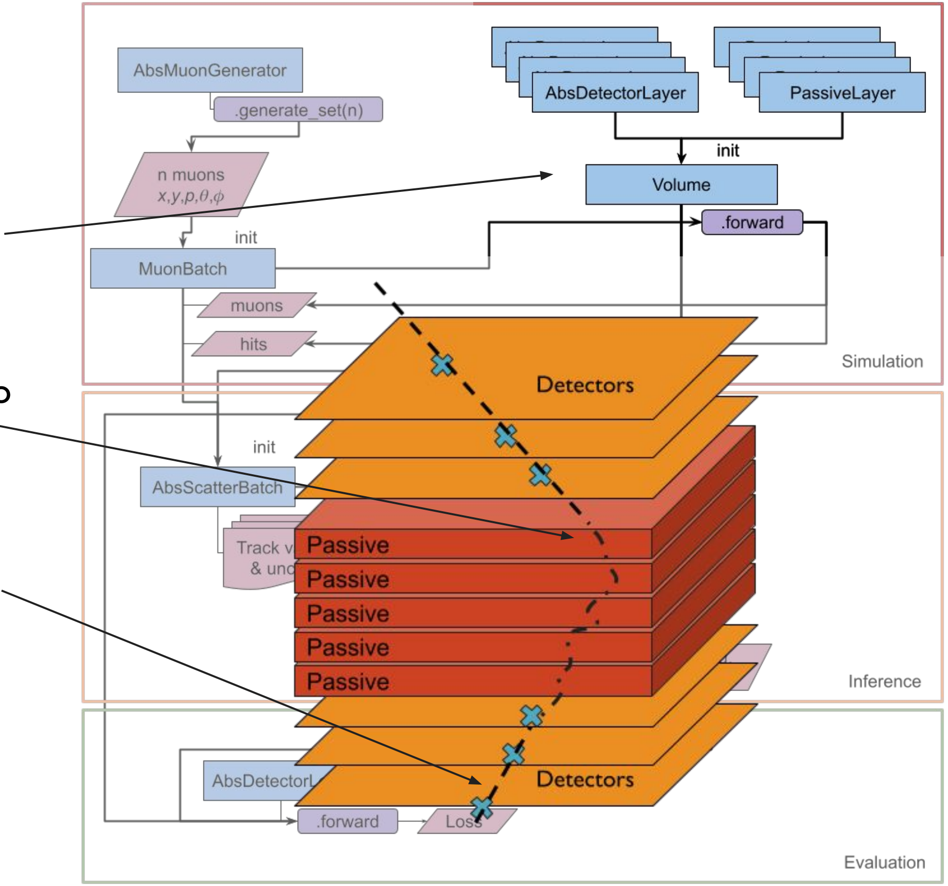
BASIC MODULES: MUON GENERATION

- Can generate muons by sampling literature models [2015, 2016]
- Sampling can provide realistic spectra for incoming angles and momenta
- Code designed to handle many muons at once



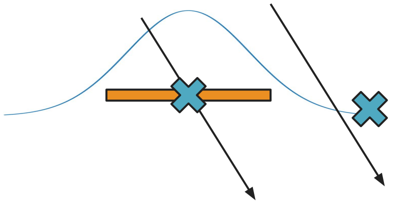
BASIC MODULES: VOLUME SPECIFICATION

- A volume consists of Layers in z stacked on top of each other
- Passive layers scatter muons according to material density (X_0)
- Detectors record muon positions (hits) with a certain resolution and efficiency

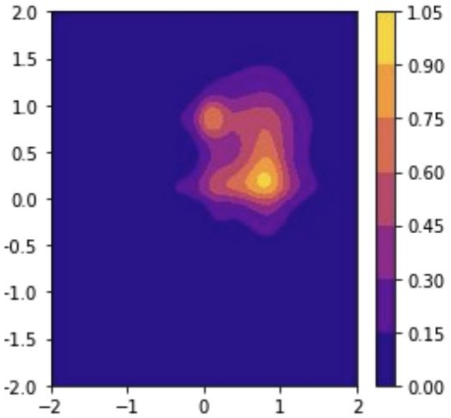


DETECTOR MODELLING

- Assume commercial detectors ⇒ fixed resolution, fixed efficiency, fixed cost per m^2
- Optimise XYZ position and XY span
- But, muons either hit or miss detectors. How can we make hits be differentiable w.r.t detector parameters?
- Instead, let resolution and efficiency be distributed, e.g. Gaussian centred on panel, with width set by panel span
  - The PDF at the muon position is now diff. w.r.t panel position and span
- Can further generalise by using Gaussian Mixture model



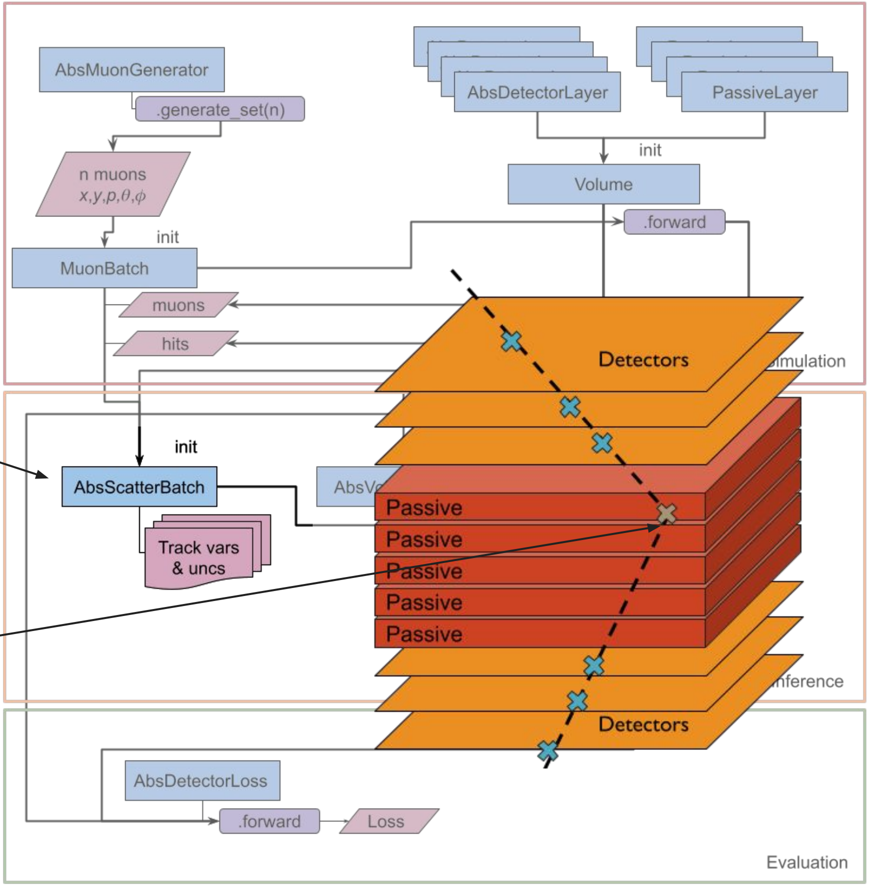
Both muons recorded, but with different resolutions



Plot: Max Lamparth

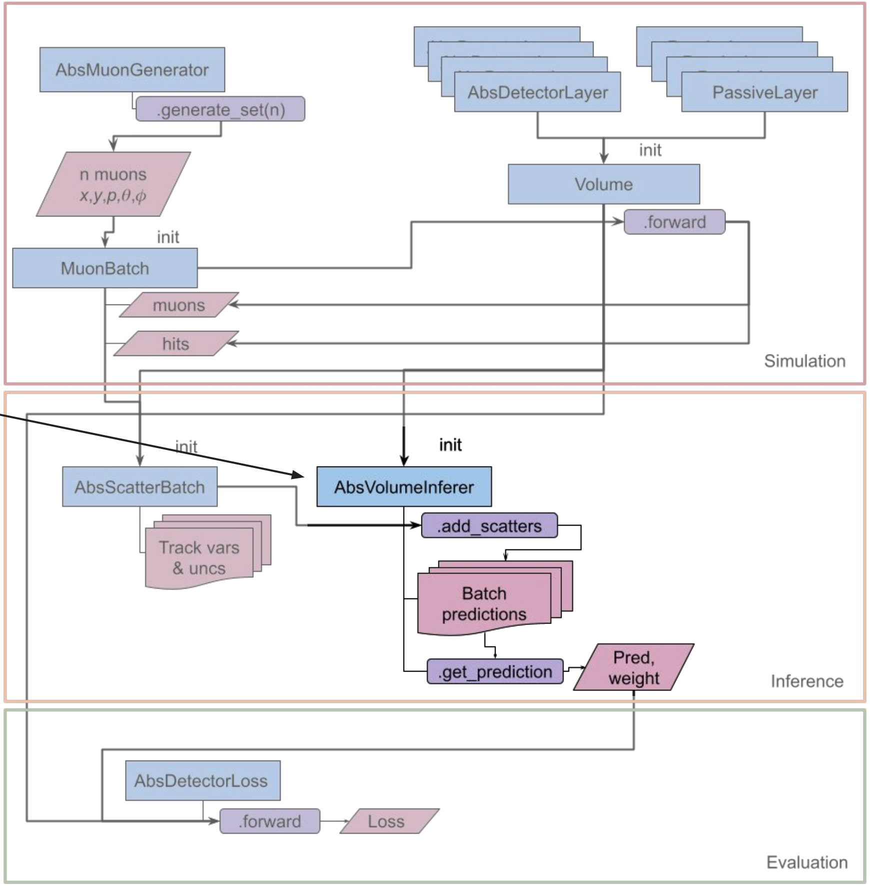
BASIC MODULES: SCATTER INFERENCE

- Next, need to fit tracks to the detector hits
- Fit uses analytic maximum likelihood considering hits and their uncertainties
  - Is fully differentiable w.r.t detector parameters
- Can then compute track parameters and their uncertainties for each muon
  - Uncertainties computed via autograd
  - Also provides the Point of Closest Approach between the tracks



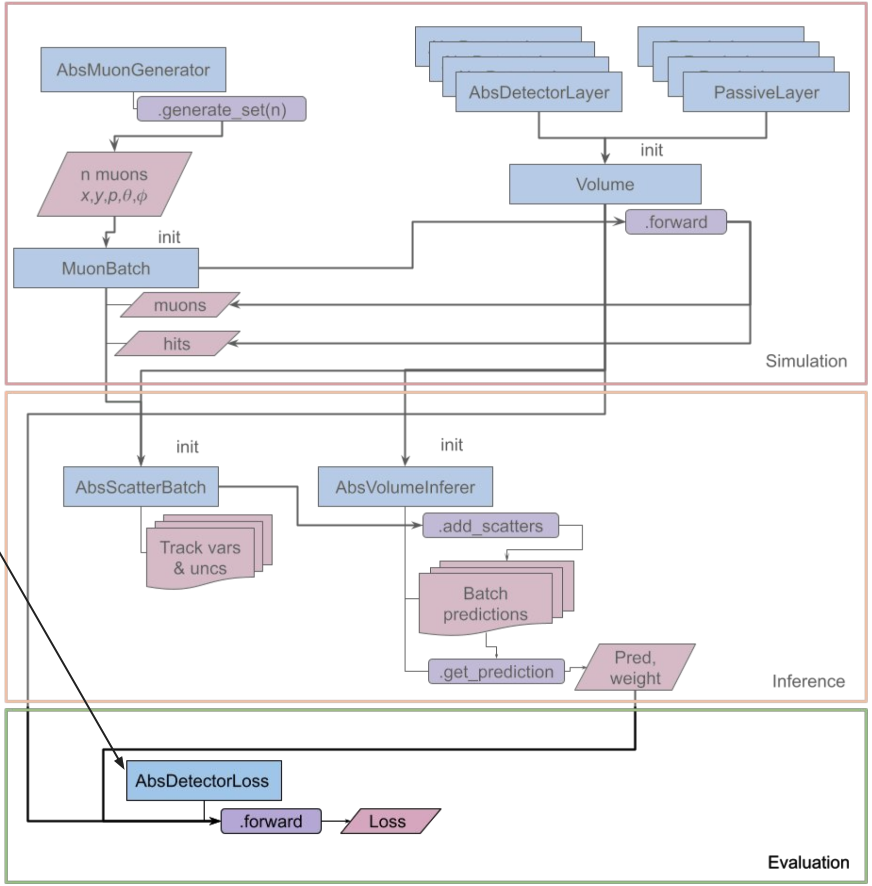
BASIC MODULES: VOLUME INFERENCE

- Next, use muon track information to infer properties of the volume
- Can run a range of classical and ML/DL algorithms here to obtain predictions
  - Must be fully differentiable
- Basic approach: Invert scatter model using track delta-angle to compute X_0
  - Highly biased
- Better: construct a task-specific summary statistic from X_0 predictions



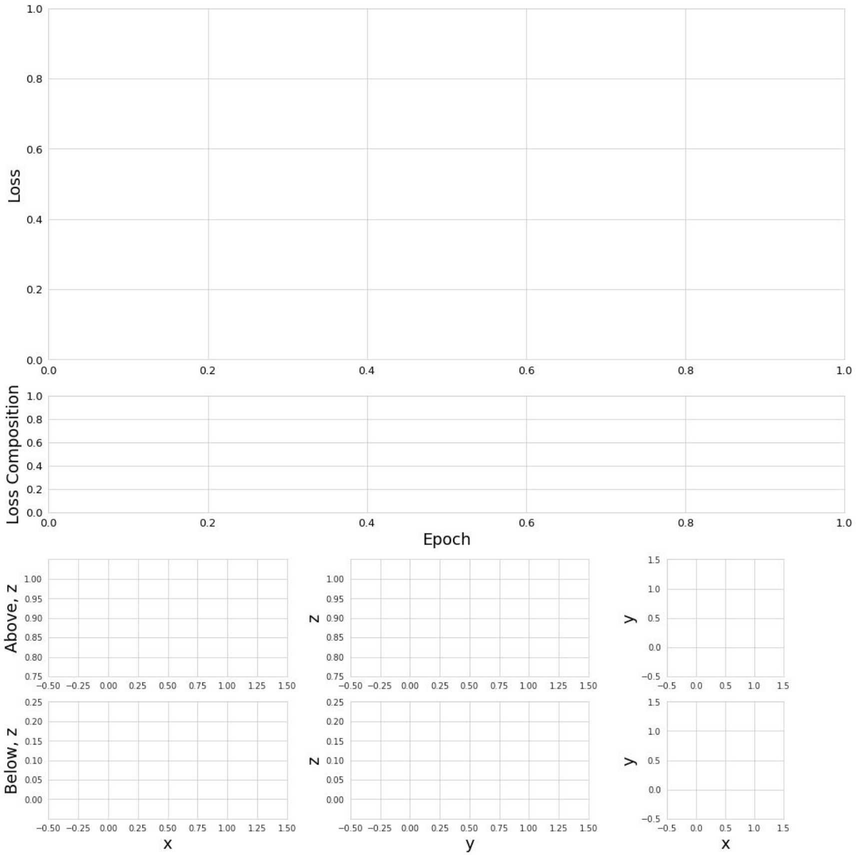
BASIC MODULES: OPTIMISATION

- Finally, compare prediction to target in a loss function
  - Suitable loss depends on the task
- The loss can also account for the cost of the detector
- Standard optimisers (SGD, Adam, etc.) can be used to update the detector parameters.

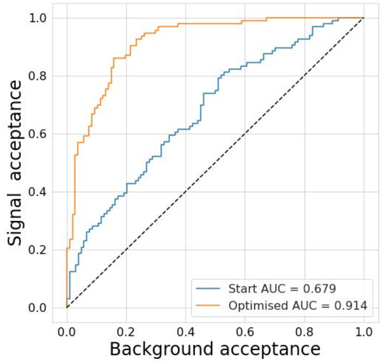


EXAMPLE

- Task is to infer presence of uranium block in lorry filled with scrap metal
  - Inference uses a dedicated summary statistic
  - The U block can be anywhere in the volume, so intuitively expect the detectors should be placed centrally in XY over the volume
- Detectors start in corner of volume and optimisation does indeed move them to cover the volume



- Optimised detector provides large improvement to ROC AUC



SUMMARY

- Measurement-aware detector-optimisation = challenging but rewarding task
  - Doesn't aim to replace detector experts; provide tools to make more informed design choices
  - Currently testing on a simplified case: muon tomography
- TomOpt indicates this is possible, and is under rapid development
  - Publications and open-source package this year

GETTING INVOLVED

- MODE involved in several other projects:
  - ECal, hybrid HCal, Cherenkov arrays, …
  - Recent whitepaper arXiv:2203.13818
  - Open to new members (contact)
  - TomOpt also welcoming new contributors: giles.strong@outlook.com
- Second MODE workshop on differentiable programming
  - 12-16 September, Crete & online
  - https://indico.cern.ch/event/1145124/

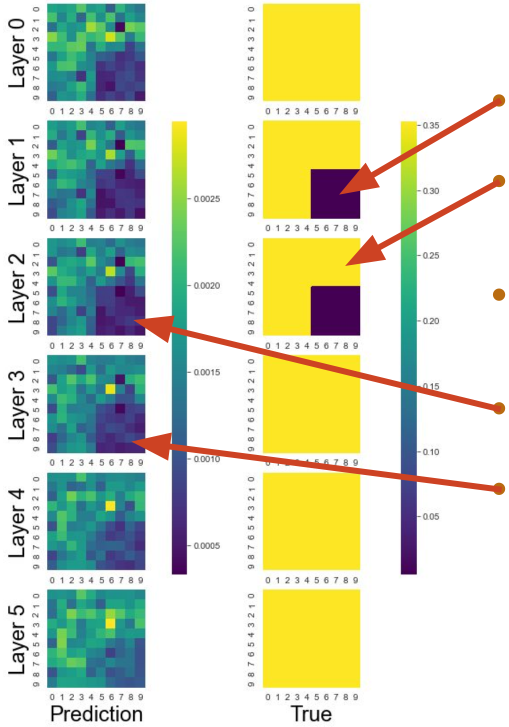
Overview of the sessions:

- Confirmed keynote speakers
  - Adam Paszke (Google Brain): DEX
- Lectures and tutorials:
  - Differentiable Programming (Pietro Vischia, UCLouvain)
  - Hackathon (Giles Strong, INFN Padova)
- Applications in muon tomography
- Progress in Computer Science
- Applications and requirements for particle physics
- Applications and requirements in astro-HEP
- Applications and requirements for neutrino detectors
- Applications and requirements in nuclear physics experiments
- Discussion on the status and needs of the discipline (one parallel session per each of the other sessions)

BACKUPS

VOLUME INFERENCE: POCA

- Point of Closest Approach: Assign entirety of muon scattering to single point
  - Invert analytic scattering model to compute X_0
  - Average X_0 predictions in each voxel
- We know, though, that the muon scattering results from multiple interactions throughout the volume
  - Assigning the whole scattering to a single point inherently leads to underestimating the X_0
  - Can slightly improve by weighting muon predictions by their X_0 uncertainty
  - Can also allow muons to predict in multiple voxels according to their PoCA uncertainty

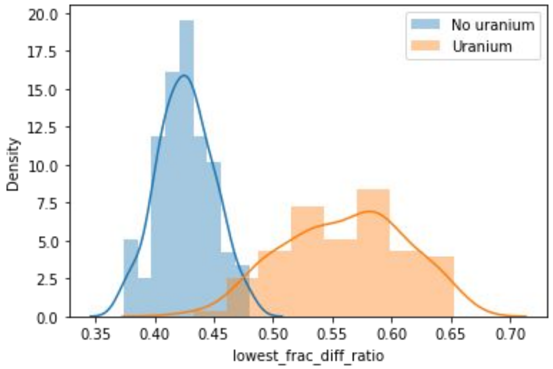


- Block of lead (X_0=0.005612m)
- Surrounded by beryllium (X_0=0.3528m)
- Predictions highly biased to underestimate X_0
- Lead block clearly visible
- but high z uncertainty in scatter location causes 'ghosting' above and below

VOLUME INFERENCE: SUMMARY STATISTIC

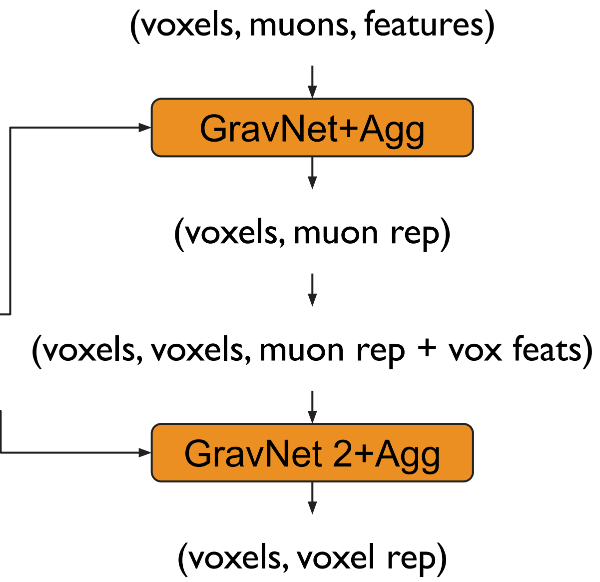
- In some cases, we don't care about predicting voxel X_0 values, but instead determining some higher-level property of the volume
  - E.g. is there uranium located anywhere in the volume?
- For this we can try to construct a summary statistic based on the X_0 predictions
- Statistics must be fully differentiable
  - Ideally, should also be invariant to scale X0 predictions, to mitigate PoCA bias

- E.g. for a uranium-block search, compare the mean of the lowest estimated to X_0 voxels to the mean of the rest
  - No block => small difference
  - Block => bimodal X_0 distribution => large difference



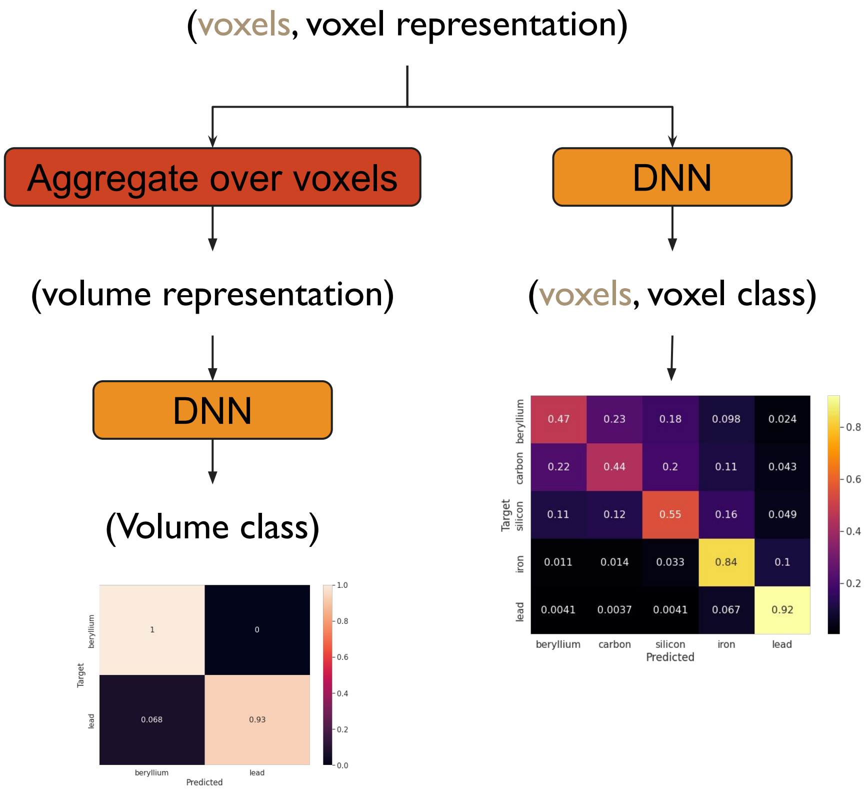
VOLUME INFERENCE: GNN

- Can use a deep learning approach
- Consider two-stage graph:
  - Each voxel has a graph built from muons
    - GNN+aggregation learns a representation of the muons specific to each voxel, by sharing features between muons
  - Each volume has a graph built from voxels
    - Second GNN+aggregation learns a representation of the voxels specific to each voxel, by sharing muon-representations between voxels.



VOLUME INFERENCE: GNN

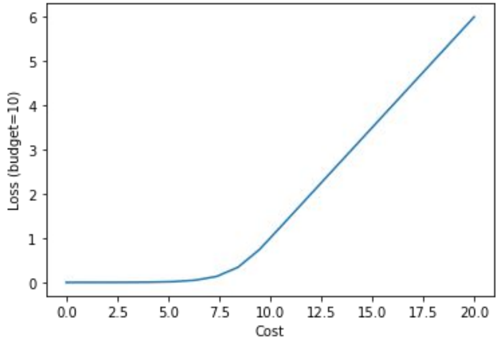
- At this point, we have a representation per voxel.
- We can transform these into X_0 predictions (class/value) with a DNN
- We can easily aggregate over the voxels to produce a volume representation.
  - This can then be further transformed into the appropriate prediction shape
- Further details in my IML talk



LOSSES AND COST

- The loss of the system should contain two components:
  - The error on the predictions
    - E.g. MSE for voxel X_0, or cross-entropy for class predictions
  - The cost of the detectors
    - Cost component smoothly "turns on" near target budget
      - Heavily penalises over-budget detectors
    - Loss scaled according to error loss

$$\mathcal{L}_{\text{Error}} = \frac{1}{N_{\text{voxels}}} \sum_{i=1}^{N_{\text{voxels}}} \frac{\left(X_{0,i,\text{True}} - X_{0,i,\text{Pred.}}\right)^2}{w_i}$$



$$\mathcal{L} = \mathcal{L}_{\text{Error}} + \alpha \mathcal{L}_{\text{Cost}}$$