A Data-Directed Paradigm for Resonance Detection at the LHC

61st Winter Nuclear Particle Physics Conference

17 February 2024

Université de Montréal

Jean-François Arguin
Georges Azuelos
Émile Baril
Fannie Bilodeau
Ali El Moussaouy
Bruna Pascual
Muhammad Usman

מכון ויצמן למדע
WEIZMANN INSTITUTE OF SCIENCE

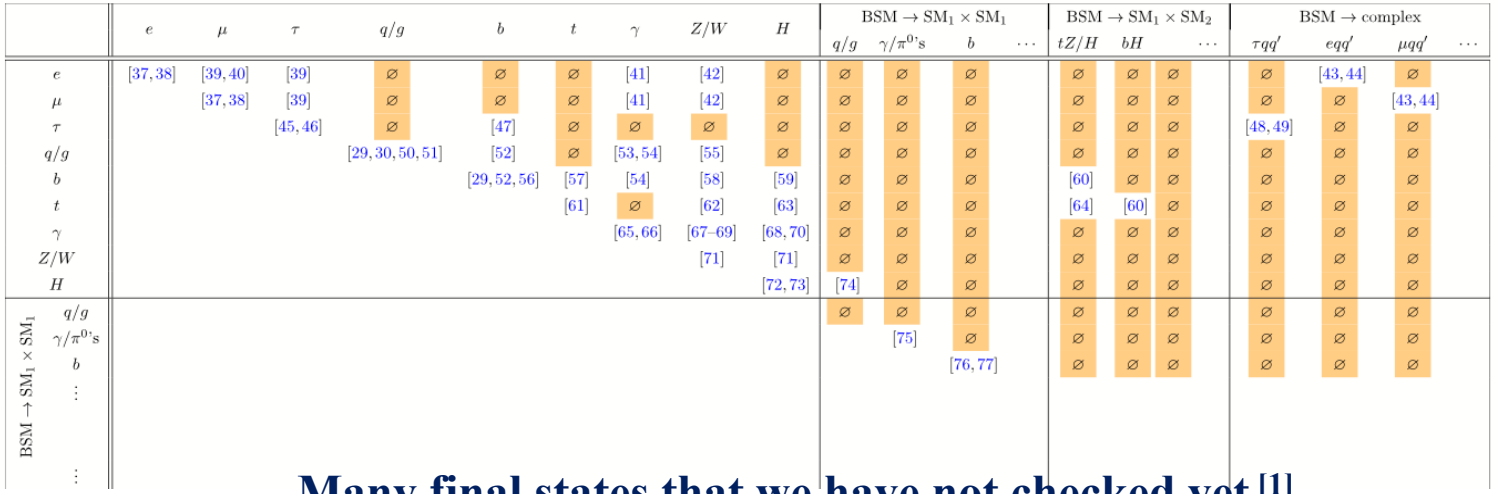
Shikma Bressler
Etienne Dreyer
Nilotpal Kakati
Amit Shkuri

LPC Particules pLasmas Univers applicationS
Laboratoire de Physique de Clermont

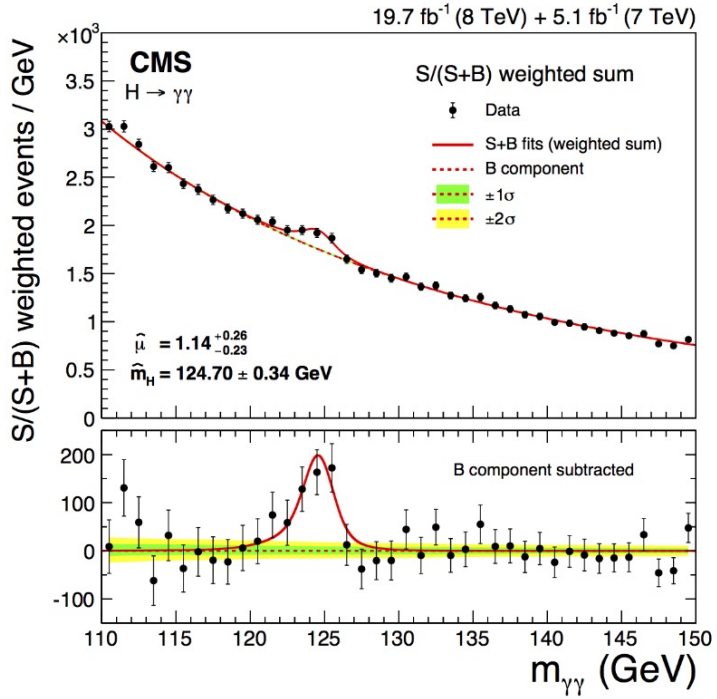
Samuel Calvet
Julien Noce Donini
Eva Mayer

Motivation

- **Search for new resonances** has historicallly been the main strategy for discovery in the experimental high energy physics.
- Explored limited region till date in the observable space.
- **Accelerated approach** is required to search for new resonance that will **cover vast observable space**.
- A search strategy that is capable of:
  1. **Identifying mass bumps** directly in the experimental data without the need of modeling the background.
  2. **Rapid scanning of many final states** in search for deviation.



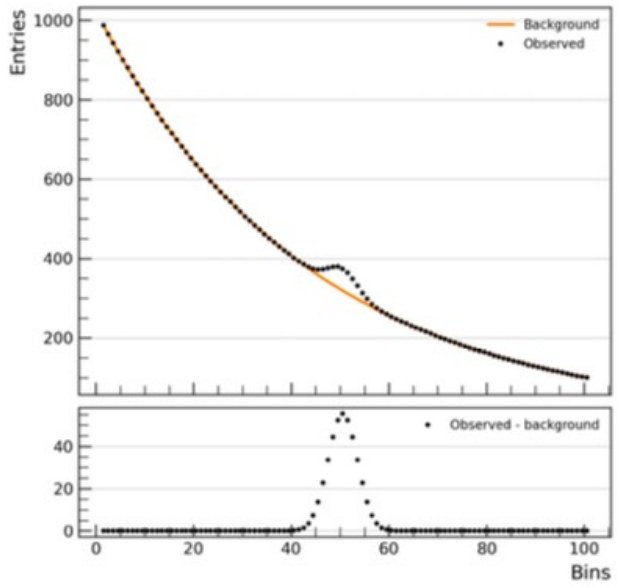
Many final states that we have not checked yet [1]

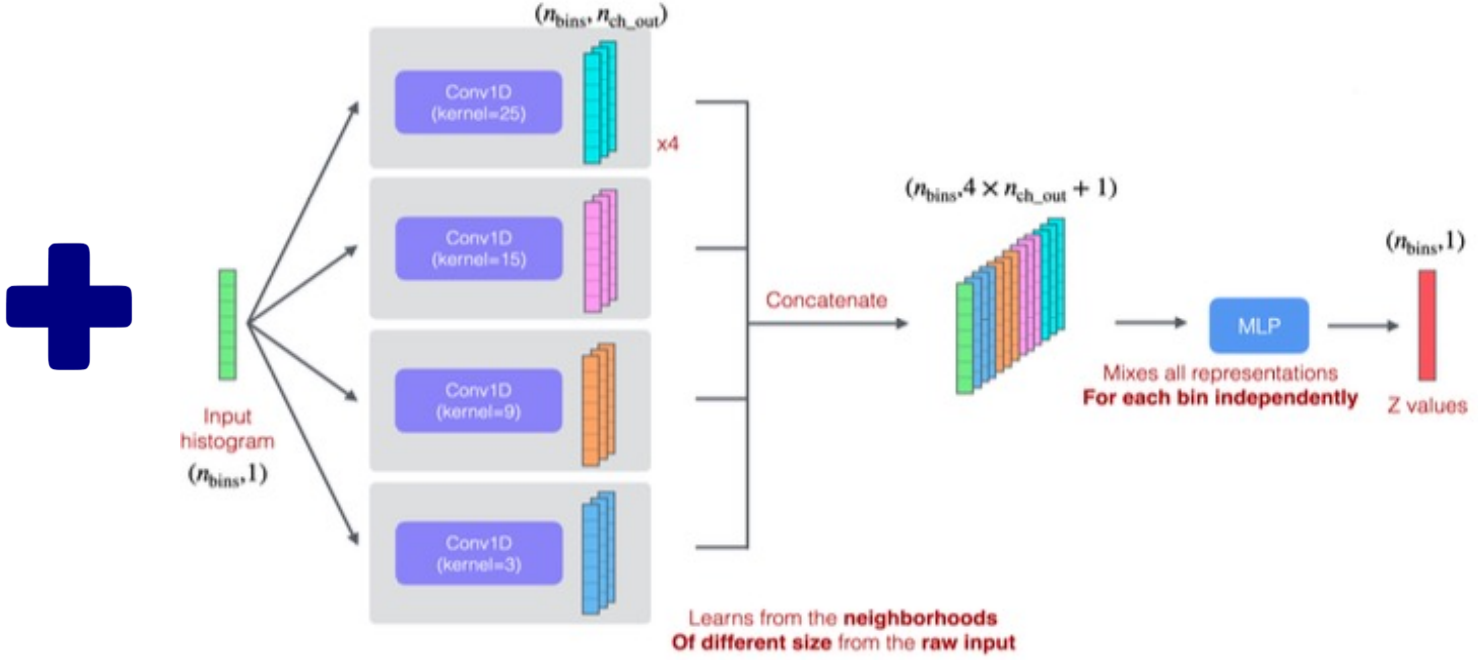


Higgs boson decaying to a pair of photons ($H \rightarrow \gamma\gamma$) [2]

Data-Directed Paradigm

- The **Data Directed Paradigm (DDP)** is a search strategy to efficiently identify regions of interest in the data. It requires two ingredients :
  1. Well established property of the SM with respect to which deviations can be searched for
     Invariant Mass Distribution (smoothly falling when no resonance)
  2. Efficient algorithm to scan the observable-space in search for deviations of this property
     Neural Network getting statistical significance for bumps

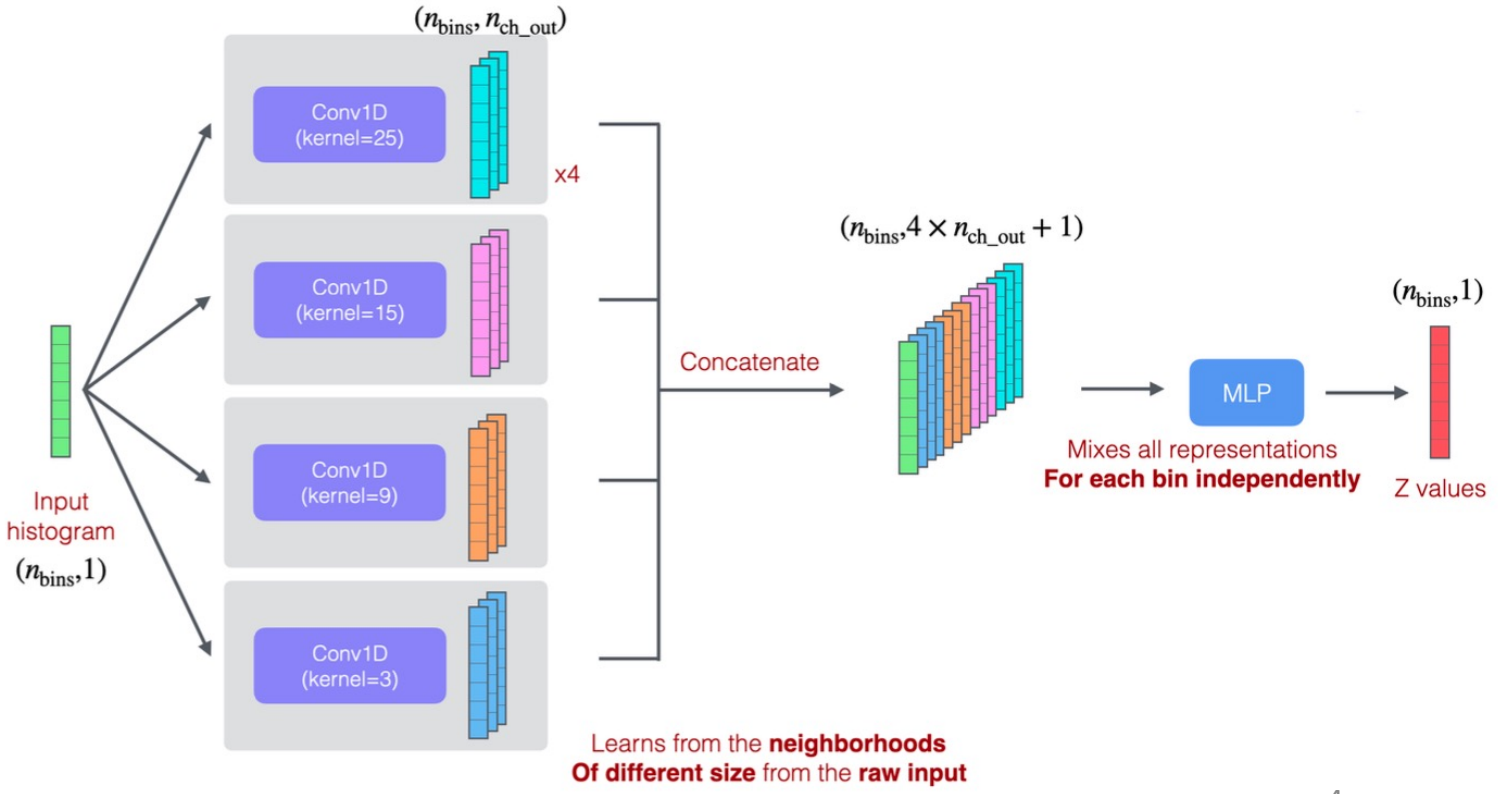




Neural Network Architecture

- **Deep neural network (NN)** to map any invariant mass distribution into a distribution of statistical significance (z distribution)
  - **Input :** Vector of bin entries from invariant mass histogram
  - **Output :** Vector of statistical significance z from likelihood-ratio test

- 1D convolutions with different kernel sizes followed by a dense layer
- Intuitive and agnostic to the number of bins in the histogram



Dark Machines Dataset

- Using the **Dark Machines (DM)[3] dataset** :
  - Dataset designed to test anomaly detection techniques
  - **Contains all of the highest cross-section processes at the LHC**
  - Generation with **Madgraph** and **Pythia**, including fast detector simulation using **Delphes**
  - **Events divided into signal regions/channels** e.g. channel 3, which is more inclusive with cuts on $E_{Tmiss} > 100$ GeV and $H_T \geq 600$ GeV
  - Dataset equivalent to **10 fb^{-1}**

SM processes			
Physics process	Process ID	σ (pb)	N_{tot} ($N_{10\,fb^{-1}}$)
$pp \to jj(+2j)$	njets	$19718_{H_T>600GeV}$	415331302 (197179140)
$pp \to l^{\pm}\nu_l(+2j)$	w_jets	$10537_{H_T>100GeV}$	135692164 (105366237)
$pp \to \gamma j(+2j)$	gam_jets	$7927_{H_T>100GeV}$	123709226 (79268824)
$pp \to l^+l^-(+2j)$	z_jets	$3753_{H_T>100GeV}$	60076409 (37529592)
$pp \to t\bar{t}(+2j)$	ttbar	541	13590811 (5412187)
$pp \to t + \text{jets}(+2j)$	single_top	130	7223883 (1297142)
$pp \to \bar{t} + \text{jets}(+2j)$	single_topbar	112	7179922 (1116396)
$pp \to W^+W^-(+2j)$	ww	82.1	17740278 (821354)
$pp \to W^{\pm}t(+2j)$	wtop	57.8	5252172 (577541)
$pp \to W^{\pm}\bar{t}(+2j)$	wtopbar	57.8	4723206 (577541)
$pp \to \gamma\gamma(+2j)$	2gam	47.1	17464818 (470656)
$pp \to W^{\pm}\gamma(+2j)$	Wgam	45.1	18633683 (450672)
$pp \to ZW^{\pm}(+2j)$	zw	31.6	13847321 (315781)
$pp \to Z\gamma(+2j)$	Zgam	29.9	15909980 (299439)
$pp \to ZZ(+2j)$	zz	9.91	7118820 (99092)
$pp \to h(+2j)$	single_higgs	1.94	2596158 (19383)
$pp \to t\bar{t}\gamma(+2j)$	ttbarGam	1.55	95217 (15471)
$pp \to t\bar{t}Z$	ttbarZ	0.59	300000 (5874)
$pp \to t\bar{t}h(+1j)$	ttbarHiggs	0.46	200476 (4568)
$pp \to \gamma t(+2j)$	atop	0.39	2776166 (3947)
$pp \to t\bar{t}W^{\pm}$	ttbarW	0.35	279365 (3495)
$pp \to \gamma\bar{t}(+2j)$	atopbar	0.27	4770857 (2707)
$pp \to Zt(+2j)$	ztop	0.26	3213475 (2554)
$pp \to Z\bar{t}(+2j)$	ztopbar	0.15	2741276 (1524)
$pp \to t\bar{t}t\bar{t}$	4top	0.0097	399999 (96)
$pp \to t\bar{t}W^+W^-$	ttbarWW	0.0085	150000 (85)

Dark Machines Sample Processing

- **Get all the possible combinations of the selected objects :**
  - Electron
  - Muon
  - Photon
  - Jet
  - Reconstructed leptonic Z: lepton pair compatible with leptonic Z decay
  - Boosted top: jet compatible with hadronic decay of top
  - Boosted hadronic W/Z boson: jet compatible with hadronic decay of W/Z
  - High mass jet: jets with mass > 200 GeV

- **Additional kinematic cuts:**
  - E_{Tmiss} > 200, 500 GeV ; leading object p_T > 100, 200, 400 GeV ...

- **Split the sub-dataset according to jet multiplicity**
  - Number of jets in final state : 0 jet, 1 jet, 2 jets, ... , ≥ 6 jets (depends on the available stat)
  - Motivation: reduce look-elsewhere effect by requiring a signal to be present in neighboring jet bins

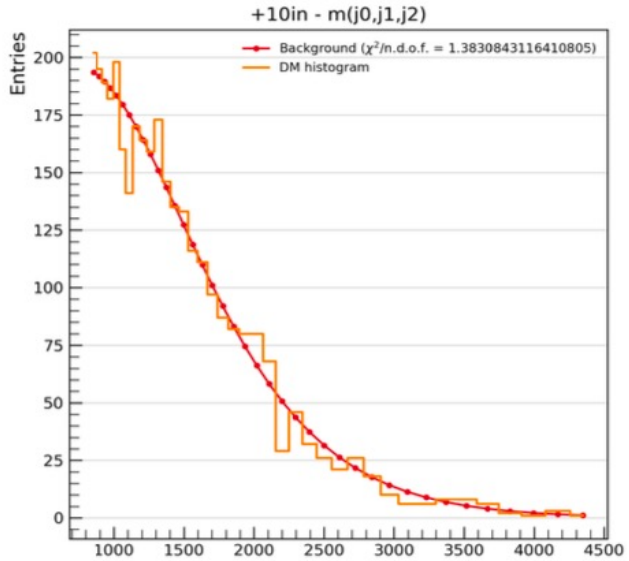
- **Build the variables**
  - Mass distributions of the objects and their combinations
  - Transverse mass distributions including E_{Tmiss}

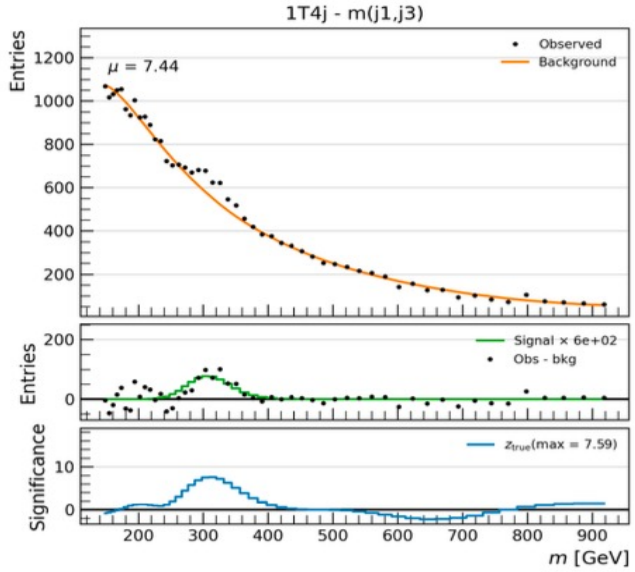
Preparation of Training Data

- **Background**
  - Smoothly descending analytical functions

$$be^{-ax},\ ax+b,\ \frac{1}{ax}+b,\ \frac{1}{ax^2}+b,\ \frac{1}{ax^3}+b,$$
$$\frac{1}{ax^4}+b,\ a\,(x-x_2)^2+y_2,\ -a\cdot\ln(x)+b,$$
$$(y_1-y_2)\cos(a\,(x-b))+y_2,\ \cosh(a\,(x-x_2))+b$$

  - Fits to Dark Machine simulation data
- **Inject signal**
  - Inject a Gaussian signal
  - Combine background and signal
  - Poisson fluctuate the histogram
  - Calculate the true significance using the likelihood ratio
- **Training Data**
  - Histograms with 30 to 100 bins
  - Broad dynamic range, 10 to 100k entries per bin
  - Different strengths of the signals (1 to 10σ)



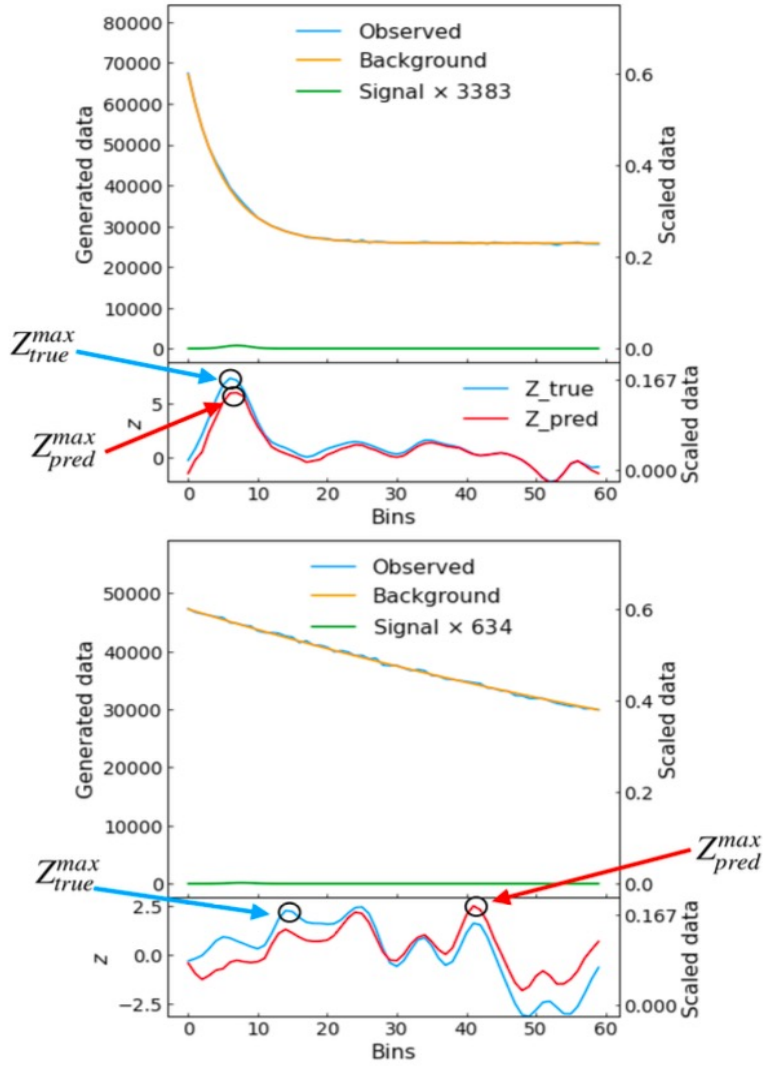


Performance Evaluation of Neural Network

- **Quantifying Performance** in terms of difference between Z^{max}_{true} and Z^{max}_{pred} maximum significance.
  - $\mathbf{Z^{max}_{true}}$: maximal significance calculated via the likelihood ratio test true
  - $\mathbf{Z^{max}_{pred}}$: maximal predicted significance

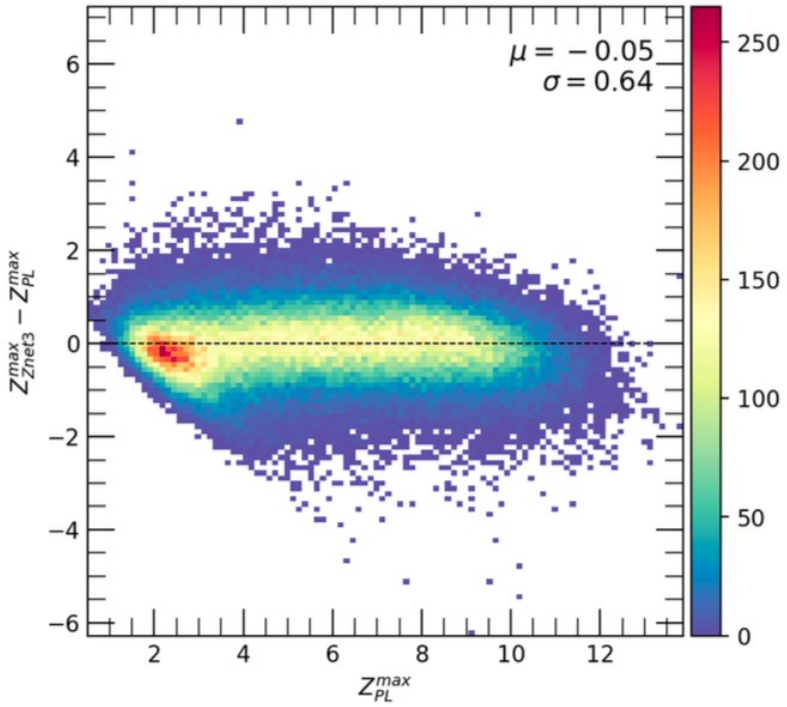
- Majority of entries should have $\mathbf{Z^{max}_{true}}$ - $\mathbf{Z^{max}_{pred}}$ close to 0 with the smallest variance possible

- **Non-zero significance** for background-only histograms
  - Due to look-elsewhere-effect
  - Could artificially bias the performance at low significance

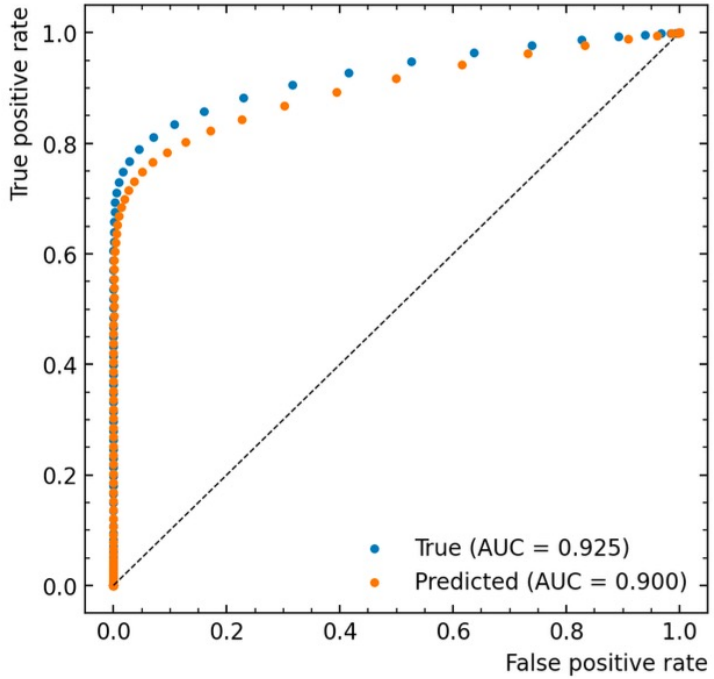


Performance on the Test Sample

- A mean (μ) of −0.05 indicates a true negligible bias in the prediction
- 0.64 standard deviation (σ) measures its precision

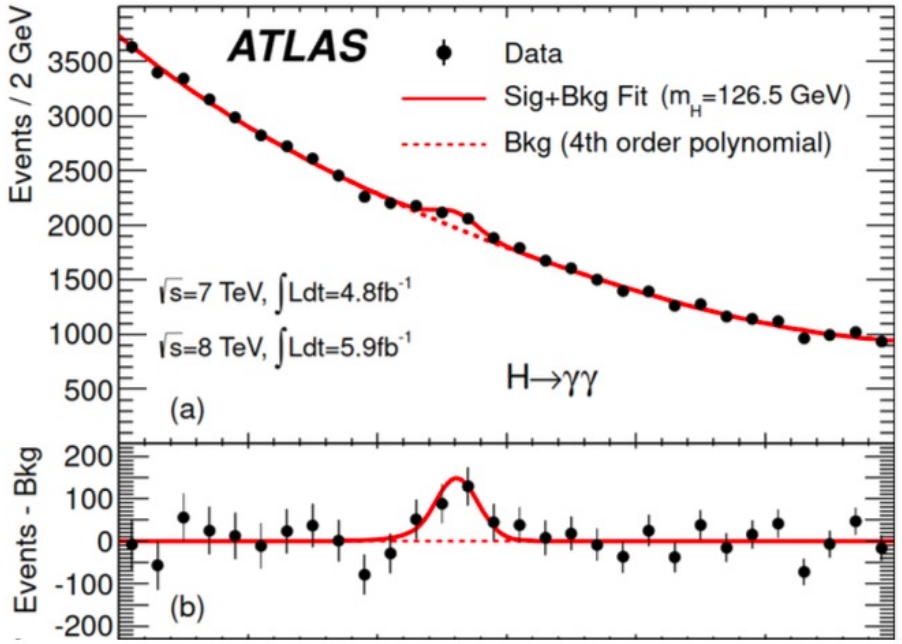


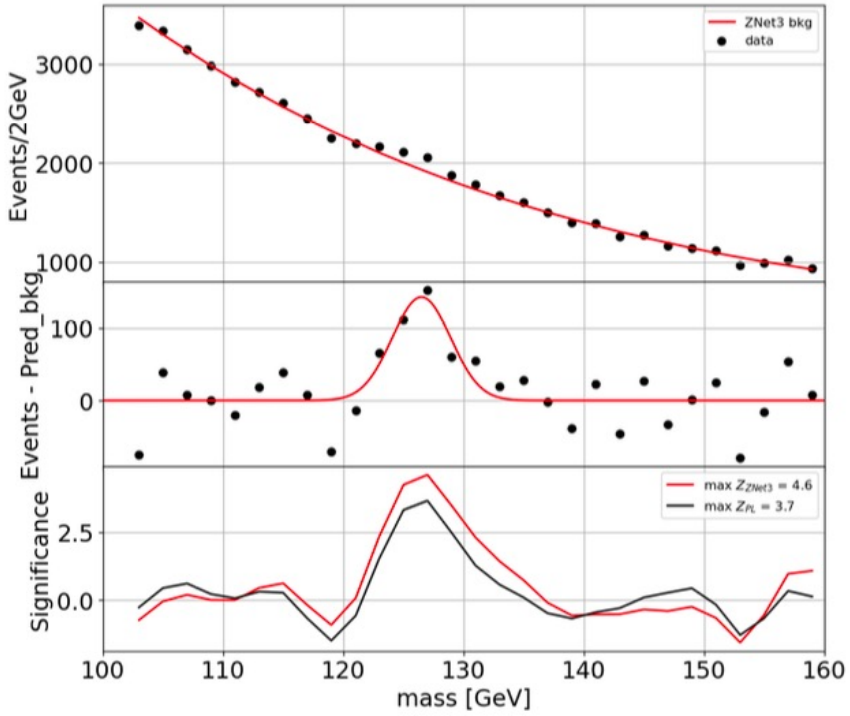
- Excellent discriminating performance of signal and background with an AUC of 0.900



Finding Higgs with DDP

- Predicted the Higgs in the correct mass position
- Predicted significance of 4.6σ whereas the ATLAS significance is 3.7σ





BSM Signal Datasets

- Using the Dark Machines dataset to construct BSM signal datasets

- **BSM signal datasets:**
  - Simulated new physics events added on top of the backgrounds
  - Used to test the network if it can find the new physics particles
  - Different levels of difficulty (e.g. cross-section, mass values, etc)

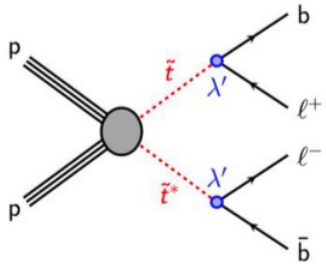
- Some of the new physics models we have available include:

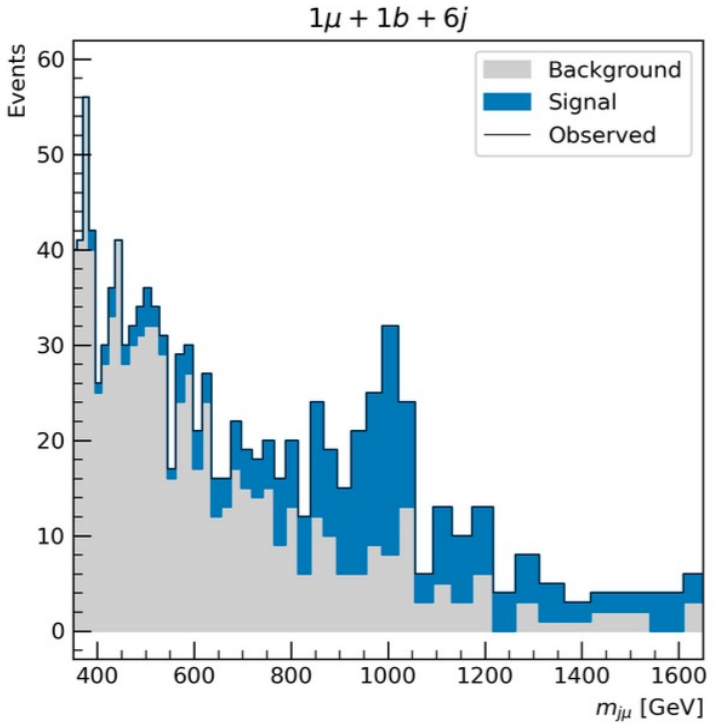
  RPV stop $\rightarrow b\ell$

  $W' \rightarrow WZ \rightarrow \ell\nu qq, qq\nu\nu$

  $LQ \rightarrow beb\mu, bebe, t\nu t\nu$

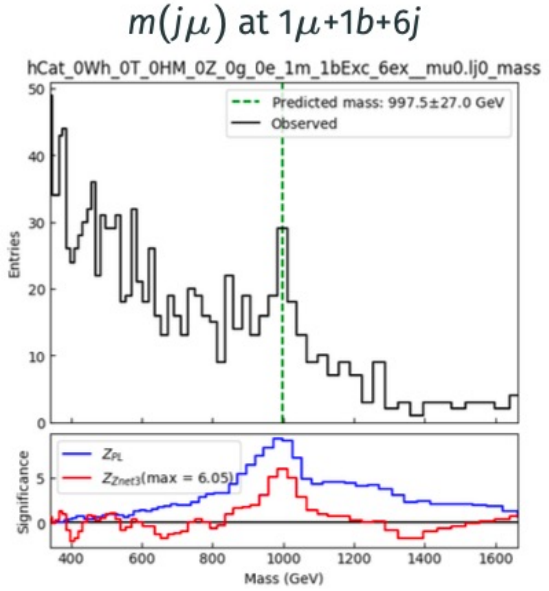
  $Z' \rightarrow 3l$

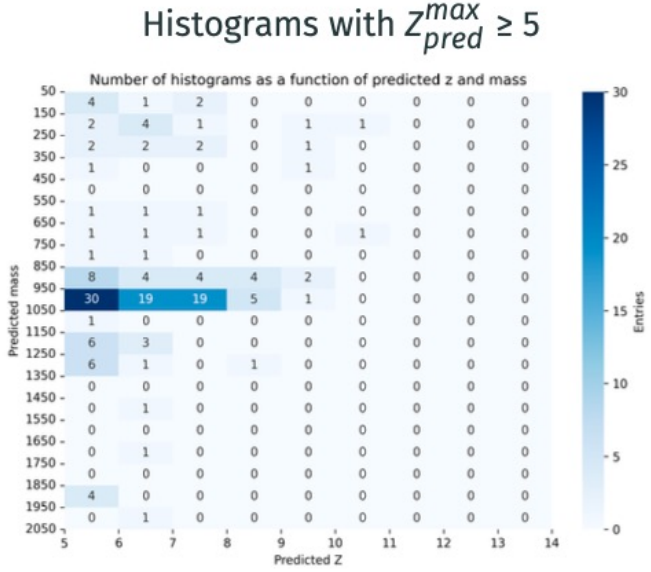


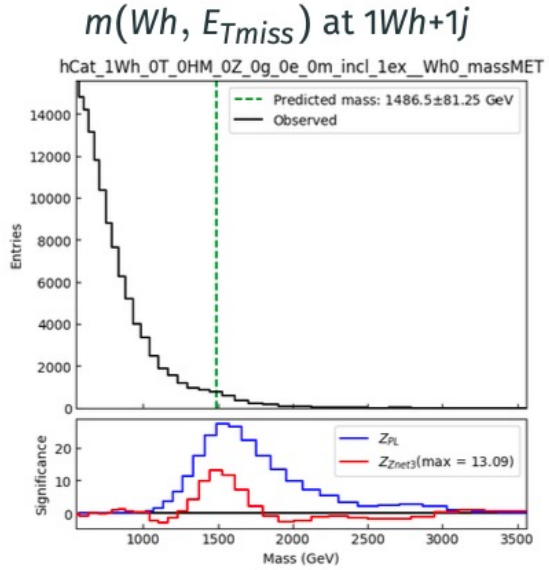


Finding BSM Signals

- Tested over simulated BSM signals added to the Dark Machines background
- Various BSM signals tested and successfully found
- False-positive rate of 0.1% when tested over background-only sample







Successfully finds an excess at the expected mass of the stop at 1 TeV

Successfully finds bump in W'

Conclusion

- DDP is an accelerated approach to search for new resonance in the vast observable space
- Successfully tested on Higgs bump and various BSM signals such as RPV stop and W′

- **Future developments**
  - Application to real experimental data, focusing on Run 2
  - Use full MC simulation data with basic selections
  - First iteration using single-lepton trigger and same objects
  - Eventually adding more objects, such as large-R jets

Thanks

References

1. J. H. Kim et al., J. High Energ. Phys. 2020, 30 (2020), arXiv:1907.06659 [hep-ph]
2. The CMS Collaboration, “Observation of the diphoton decay of the Higgs boson and measurement of its properties”, arXiv:1407.0558, submitted to Eur. Phys. J. C.
3. T. Aarrestad et al., SciPost Phys. 12, 043 (2022), arXiv:2105.14027 [hep-ph]