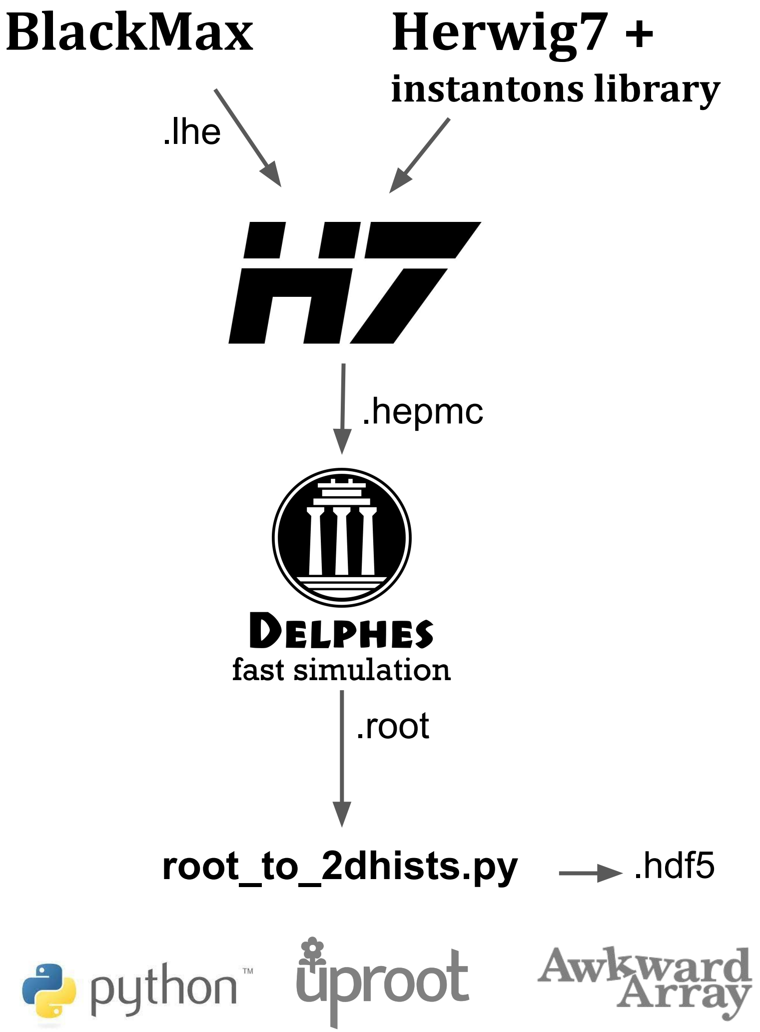
Sphalerons vs black holes

Datasets and method

Data pipeline

1. Use **BlackMax/Herwig7 instanton library** to generate parton level events. (ANDREAS)
2. **Herwig7** for hadronization
3. **Delphes** for detector response simulation
4. My own **python code** for transforming root files to images (2D histograms).



Datasets

X black hole models and Y sphaleron models.

≈ 10 000 events for each model

For now: 1000 events for each, for the data analysis

Black holes

- Parton level events generated using BlackMax (https://blackmax.hepforge.org/)
- Hadronized using Herwig7
- We explore the effect of number of extra dimensions and minimum mass

Model name	BH_n2_M8	BH_n4_M8	BH_n6_M8	BH_n2_M12	BH_n4_M12	BH_n6_M12
Number of extra dimensions	2	4	6	2	4	6
Minimum mass	8 TeV	8 TeV	8 TeV	12 TeV	12 TeV	12 TeV

Sphalerons

- Generated in Herwig7 using the instanton library provided by Herwig7 developer Andreas Papaefstathiou (https://gitlab.com/apapaefs/instantons)
- Set 0-boson final state
- Vary the sphaleron energy

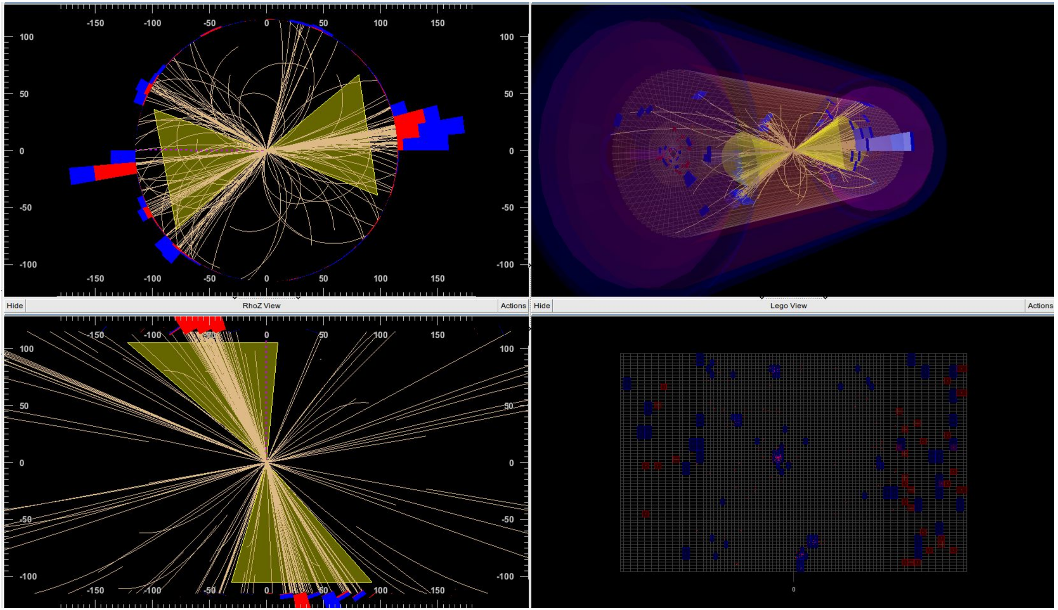
Model name	sph_8TeV	sph_9TeV
Sphaleron energy	8 TeV	9 TeV

Detector response simulation

Delphes

- Simplified, parameterised simulation
- Outputs root files with calorimeter information and reconstructed objects
- I have used the ATLAS card

https://cp3.irmp.ucl.ac.be/projects/delphes

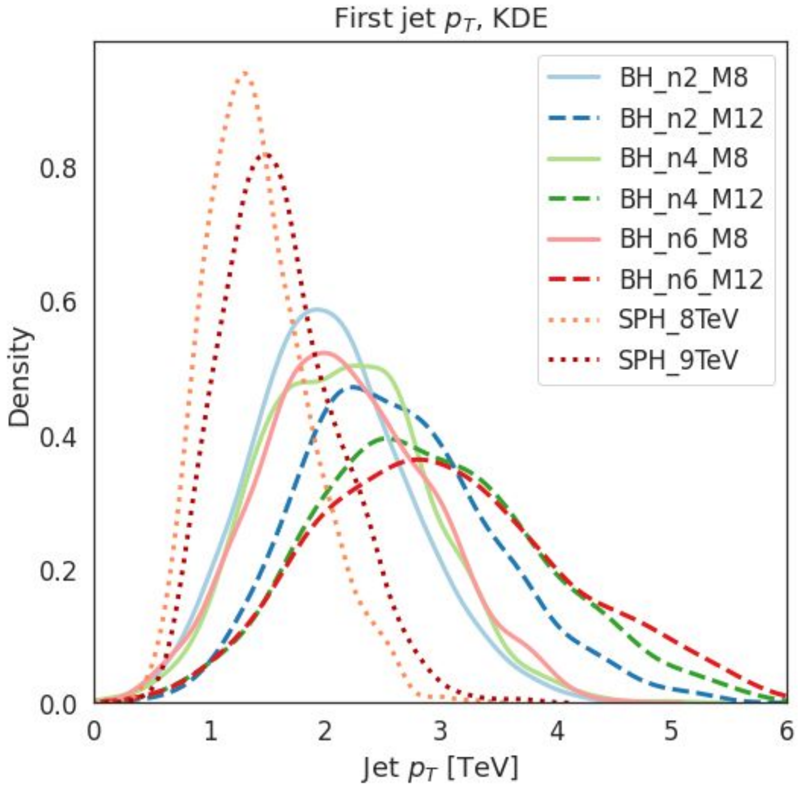


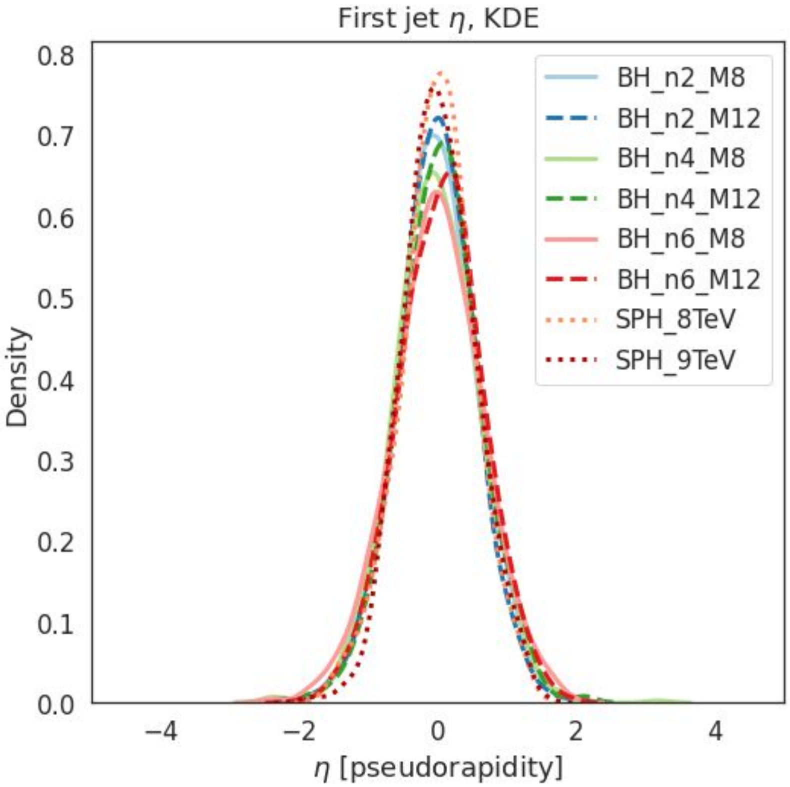
Event display of a black hole event from our dataset.

Data analysis

Jet energies

First jet: Jet with the highest p_T

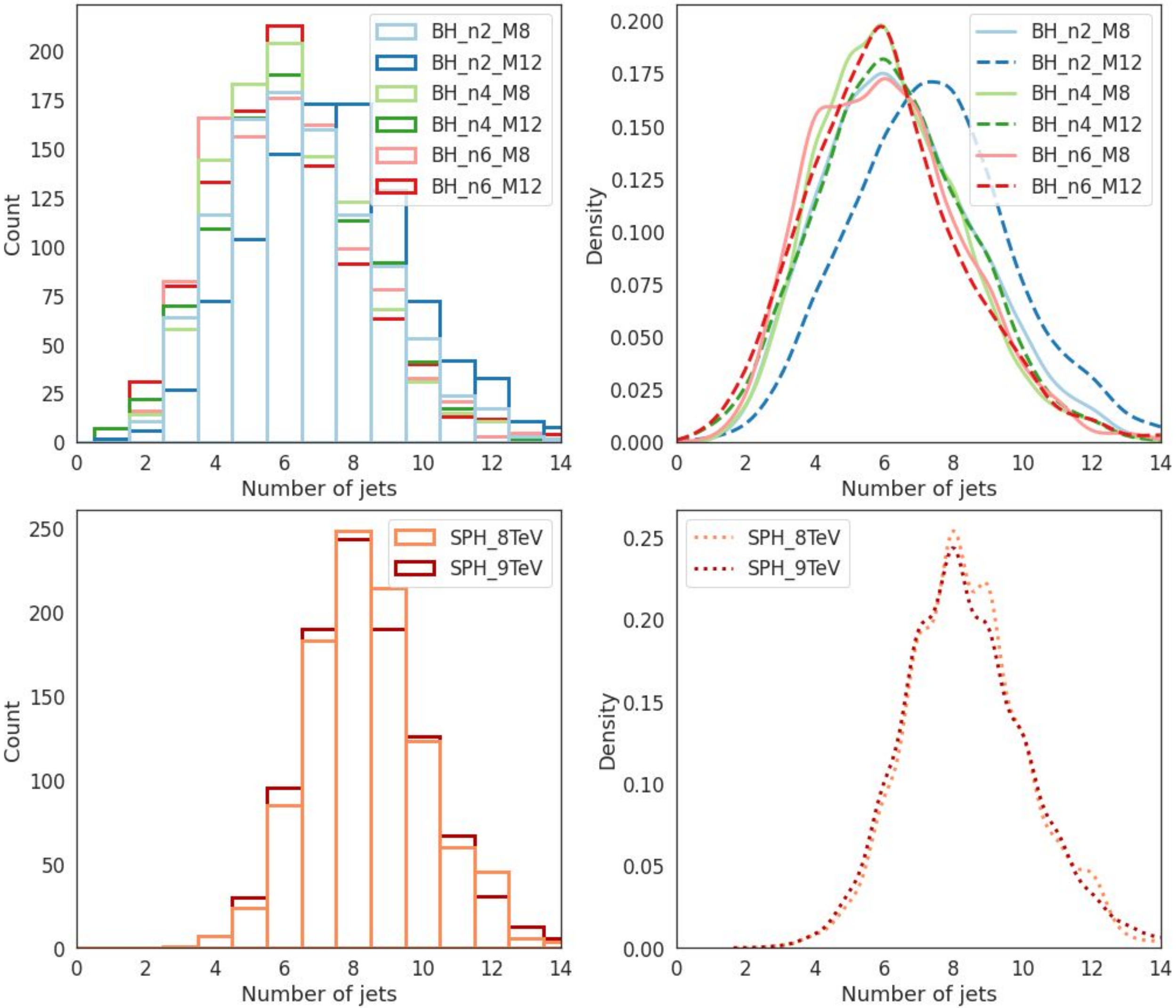




Jet multiplicity

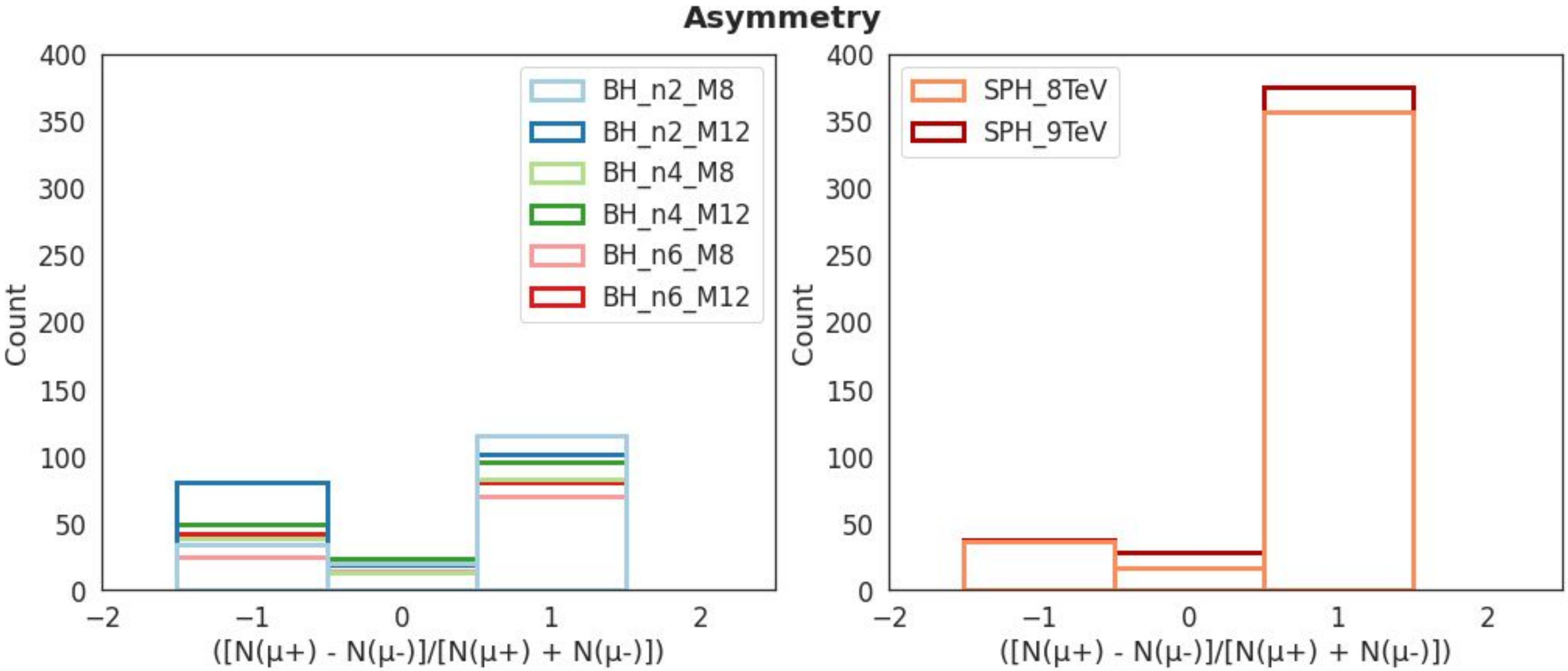
Sphalerons: higher number of jets than BH in general.

BH: Fewer extra dimensions and high mass leads to higher number of jets.



Muon multiplicity

N(**μ**) = number of **μ** in an event

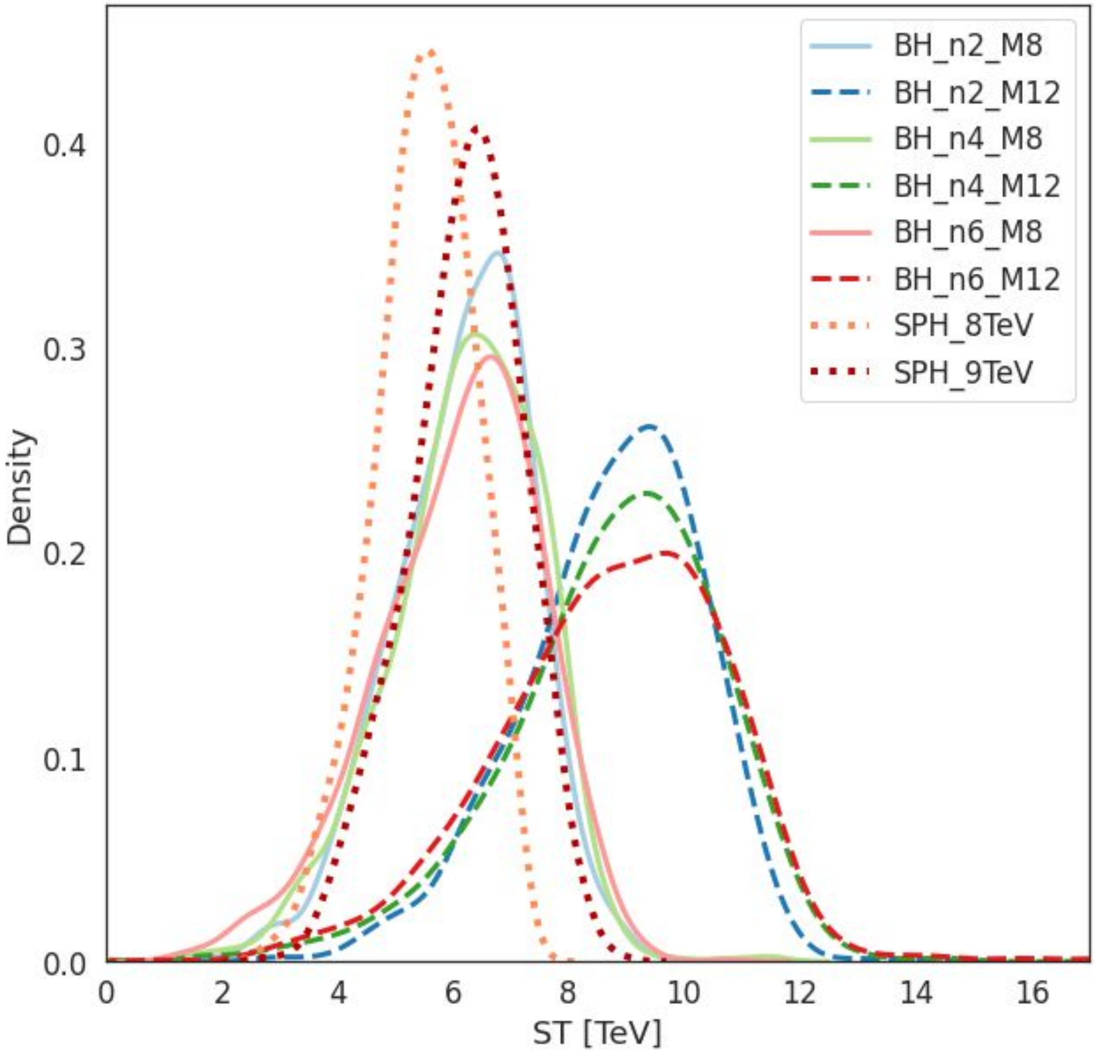


ST (HT)

ST is the scalar sum of the transverse momentum of all final state physics objects recorded for the event.

- Jets, leptons, photons, MET

Sphalerons are more similar to low mass black holes, but these are the ones most different in multiplicity.



Machine learning

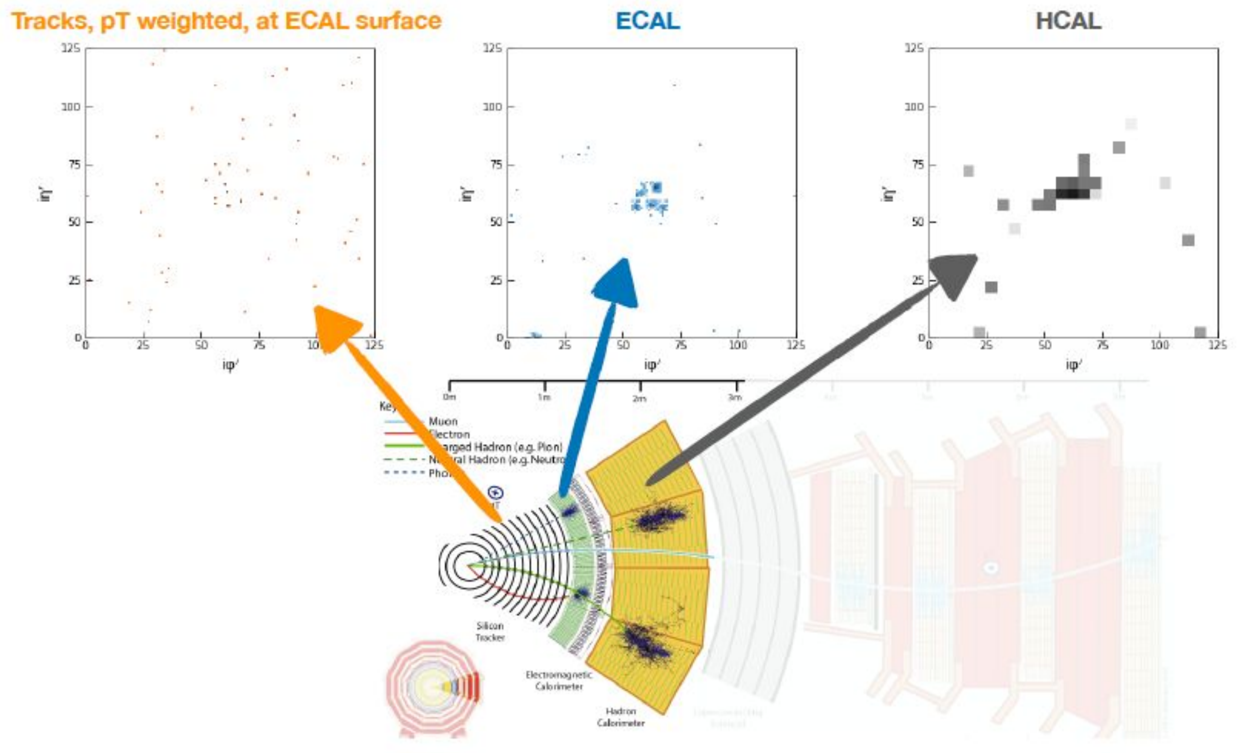
Images

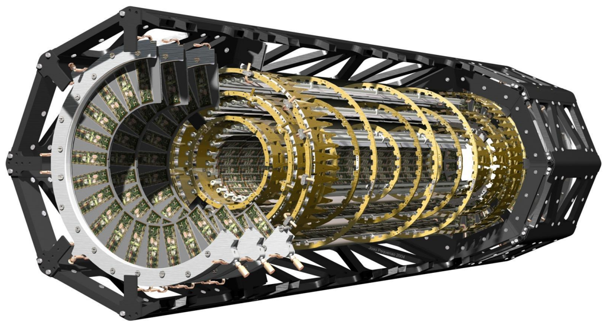
(R, G, B) = (EMCal, HCal, tracks)

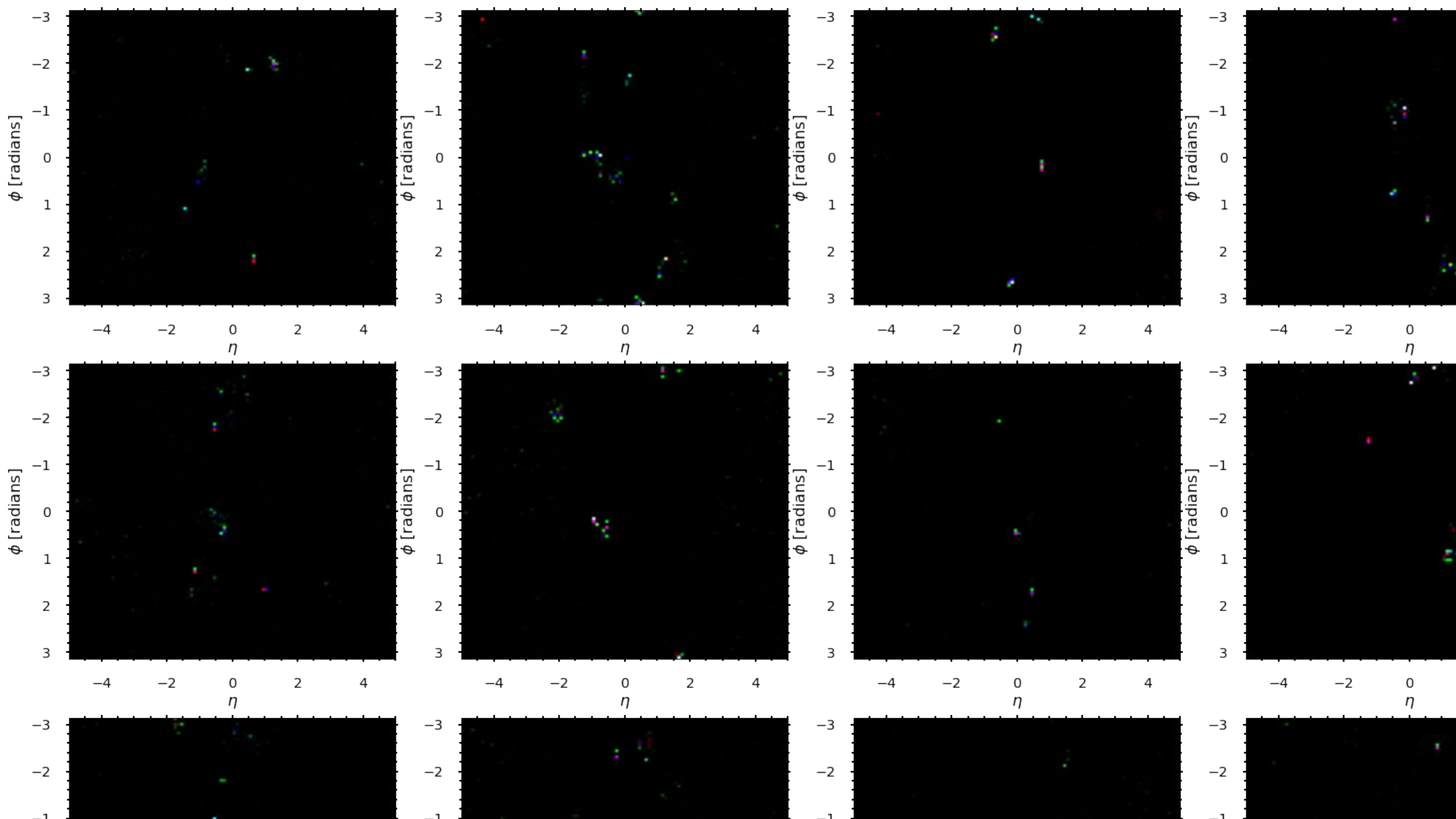
Intensity ∝ Energy deposit

Process based on this paper:

https://arxiv.org/abs/1807.11916



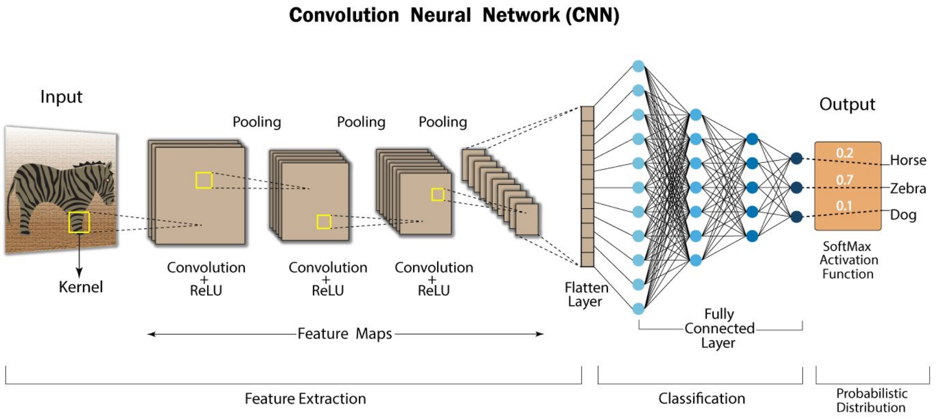


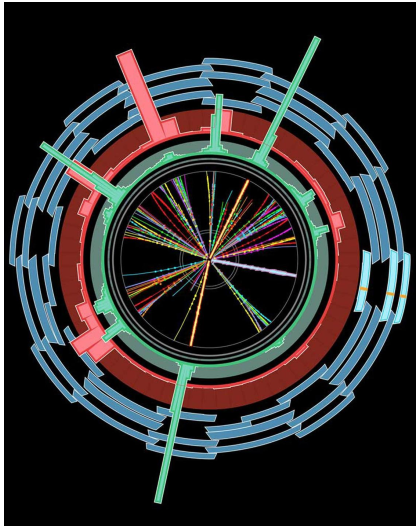


ϕ [radians]
η
−3
−2
−1
0
1
2
3
−4
−2
0
2
4

Machine learning process

1. Create a Convolutional Neural Network
   a. Preferably a circular CNN, for panoramic images.
2. From the labeled 3-channel images we train the network
   a. This step requires fine tuning of hyper parameters, and is where a lot of the machine learning expertise can be used.
   b. Exploit symmetries to perform data augmentation
3. We test on a subset of images that were not used in training
4. Compare to kinematic-cut based separation





Example

https://github.com/choisant/imcalML/blob/main/notebooks/CNN_simple_classifier.ipynb

You can learn more about the project at my github page:

https://github.com/choisant/imcalML

Still a work in progress :)

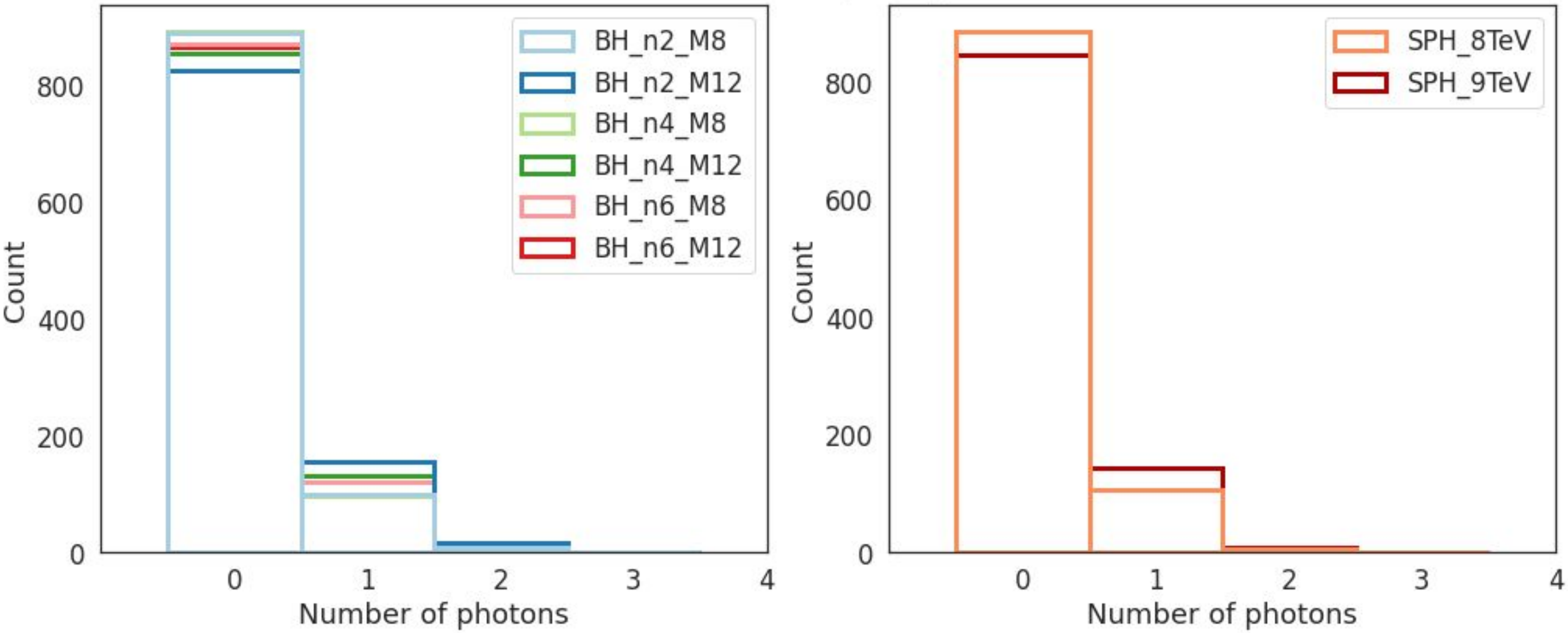
Project plans

- [ ] Write introduction and theory
- [x] Generate datasets
- [x] Select datasets for analysis
- [x] Create images
- [ ] Do machine learning training and testing
- [ ] Compare results to kinematic based cut.
- [ ] Finish first draft by **December**

Questions?

Backup slides

Photon multiplicity

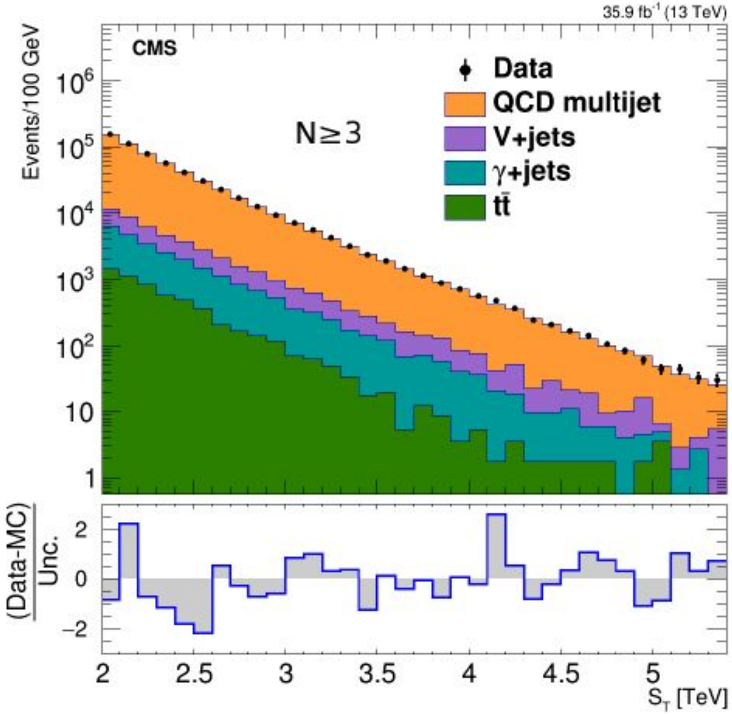


CMS-EXO-17-023

CERN-EP-2018-093
2018/11/16

Search for black holes and sphalerons in high-multiplicity final states in proton-proton collisions at $\sqrt{s} = 13$ TeV[†]

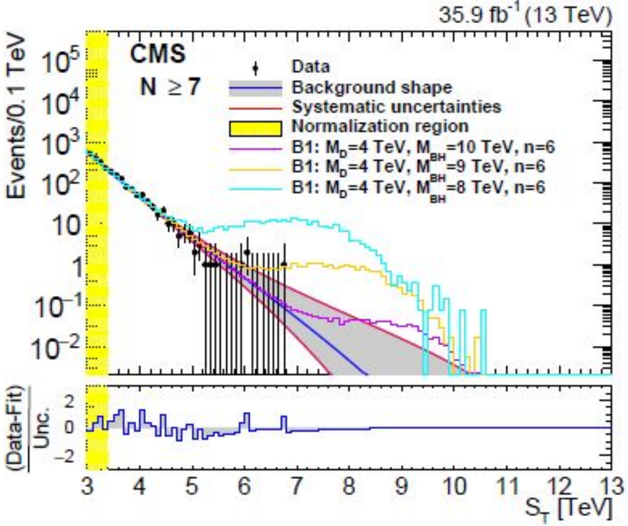
The CMS Collaboration[*]

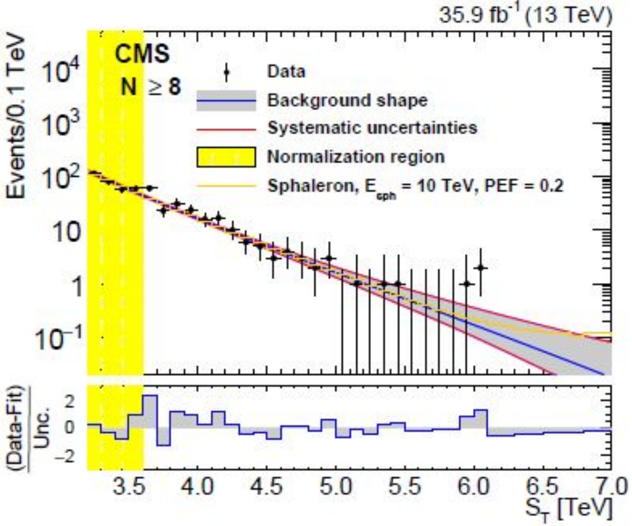


6 Background estimate

6.1 Background composition

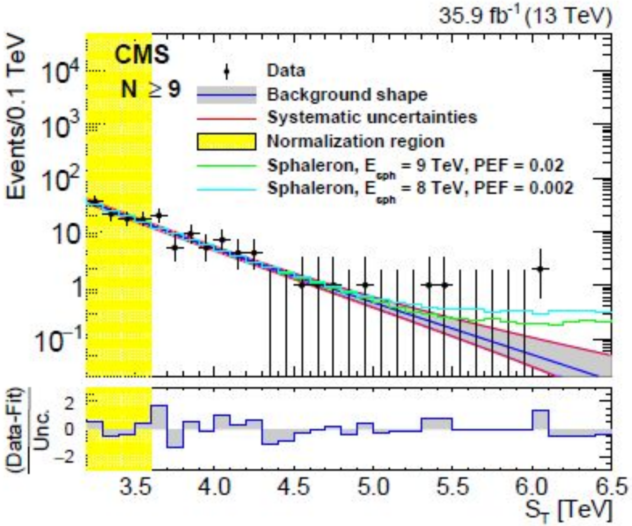
The main backgrounds in the analyzed multi-object final states are: QCD multijet, V+jets (where V = W, Z), γ+jets, and $t\bar{t}$ production, with the QCD multijet background being by far the most dominant. Figure 2 illustrates the relative importance of these backgrounds for the

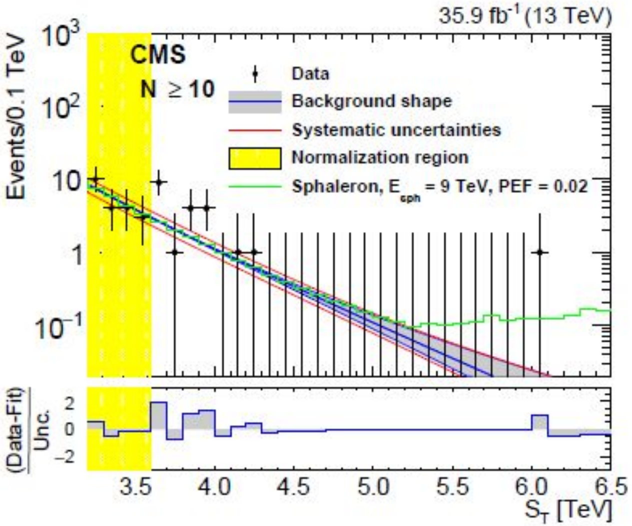


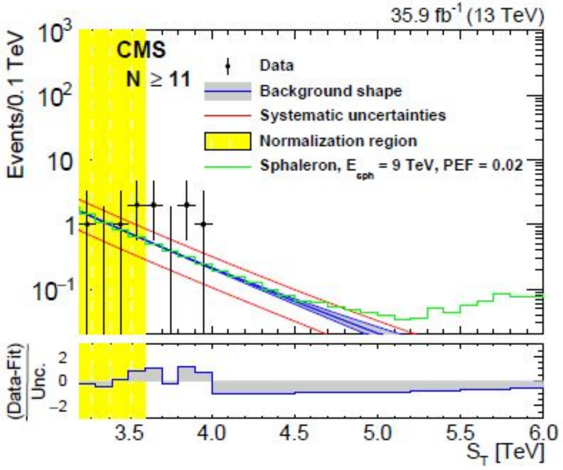


Page 14

https://arxiv.org/abs/1805.06013







End-to-End Physics Event Classification with CMS Open Data

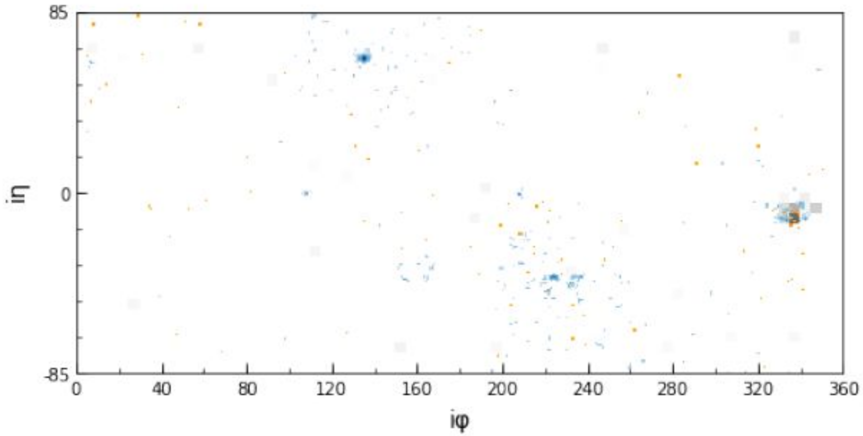
Applying Image-Based Deep Learning to Detector Data for the Direct Classification of Collision Events at the LHC

M. Andrews, M. Paulini, S. Gleyzer, B. Poczos

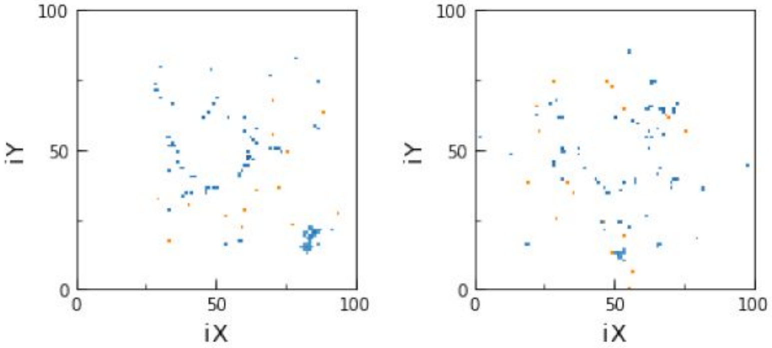
October 28, 2020

https://arxiv.org/abs/1807.11916

“the fully end-to-end event classification approach describes a general framework that can be applied to arbitrarily complex physics processes, as are found in some searches for physics beyond the standard model (BSM)“



(a) Barrel section of composite image in ECAL-centric geometry. Image resolution: 170 × 360.

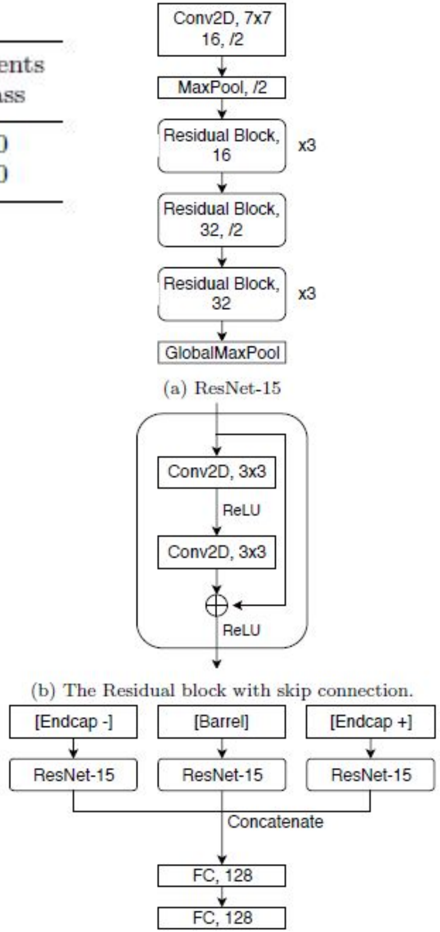


(b) Endcap sections of composite image in ECAL-centric geometry. Image resolution: 100 × 100.

ML results

Category	Training Events per class	Test Events per class
Central	51200	11800
Central+forward	120000	15600

- The image-based classifier performed better than a FCN using the 4 momentum of the two reconstructed photons as input.
- CNN’s need more training data than other models.
- Best suited for complex decays
- Robust against underlying event and pile up.



(a) ResNet-15

(b) The Residual block with skip connection.

(c) Concatenation of multiple ResNet-15 networks from separate barrel and endcap inputs.

Fig. 3: The Residual Net (ResNet) architecture, as used for single (3a) and multiple (3c) image inputs.

Table 6: Multi-class Event Classification Results, central $|\eta| < 1.44$ region. Uncertainties are on the last digit.

Metric	4-mom	EB, mass-aware	EB	CMS-B
1-vs-Rest: ROC AUC / FPR@TPR=0.7				
$H \to \gamma\gamma$	0.71/0.41	0.93/0.08	0.80/0.27	0.81/0.26
$\gamma\gamma$	0.81/0.25	0.92/0.06	0.83/0.24	0.84/0.22
$\gamma + \text{jet}$	0.81/0.22	0.95/0.01	0.94/0.02	0.96/0.02
1-vs-1: ROC AUC / FPR@TPR=0.7				
$H \to \gamma\gamma$ vs $\gamma\gamma$	0.77/0.32	0.91/0.11	0.72/0.40	0.72/0.40
$H \to \gamma\gamma$ vs $\gamma + \text{jet}$	0.78/0.28	0.97/0.02	0.94/0.07	0.96/0.04
CVM	0.002	0.080	0.002	0.002