Kobayashi-Maskawa Institute
Nagoya University

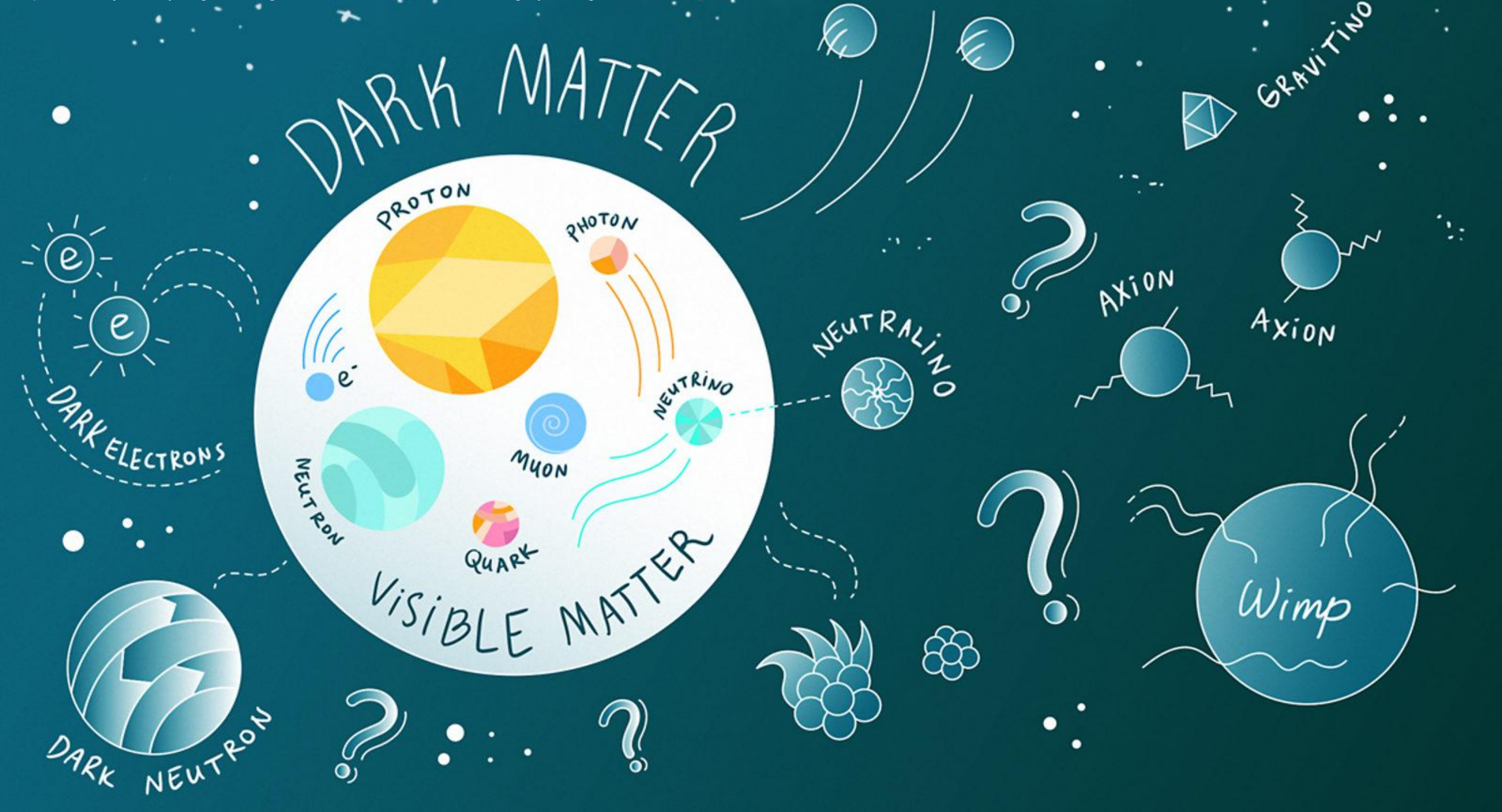
Searching For Anomalies with Foundation Models

vmikuni@nagoya-u.jp

vinicius-mikuni

Vinicius M. Mikuni

https://www.symmetrymagazine.org/article/december-2013/four-things-you-might-not-know-about-dark-matter



The Challenge

Q: How do you find a needle in a haystack?

The Challenge

Q: How do you find a needle in a haystack?

A: You use a magnet!

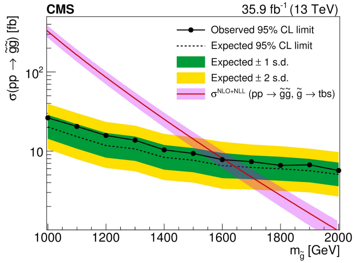
The Challenge

Plausible Theory: SUSY, WIMPs, LLPs

Verification: Confirm the theory using data

Theory was right!

Constrain the new theory



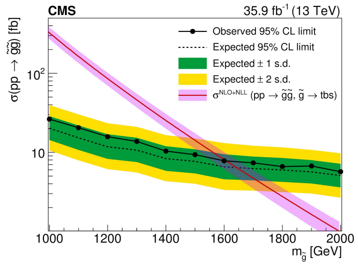
The Challenge

Plausible Theory:
SUSY, WIMPs, LLPs

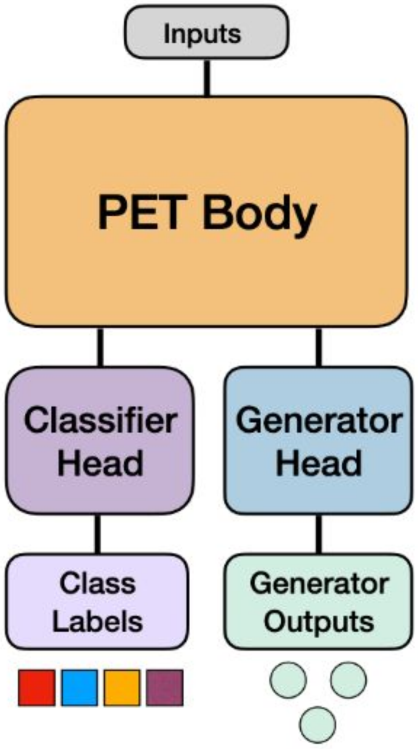
Search for anomalies!

Interpretation

Constrain many theories



OmniLearn



Create a neural network to accomplish 2 tasks:

- **Classify jets**: learns the difference in radiation signature between jet types
- **Generate jets**: implicitly learn the likelihood of jets for different initial partons

V. Mikuni, B. Nachman, Phys.Rev.D 111 (2025) 5, L051504
W. Bhimji, C. Harris, **V. Mikuni,** B. Nachman, *Phys.Rev.D* 113 (2026) 3, 3

1 Billion Dataset

Dataset	Training	Validation	Test
JetClass [14]	100M	20M	5M
JetClass2 [33]	200M	600k	600k
Aspen Open Jets [34]	125M	25.7M	26.6M
ATLAS Top Tagging [35]	178M	20M	2.2M
H1 DIS	42.2M	872k	255k
CMS QCD	239M	17.5M	16M
CMS BSM	173.5M	17M	17M
Total	1057.7M	101.8M	67.6M

Combine multiple open datasets in HEP

More than **1 Billion Jets** are accessible directly from the software package

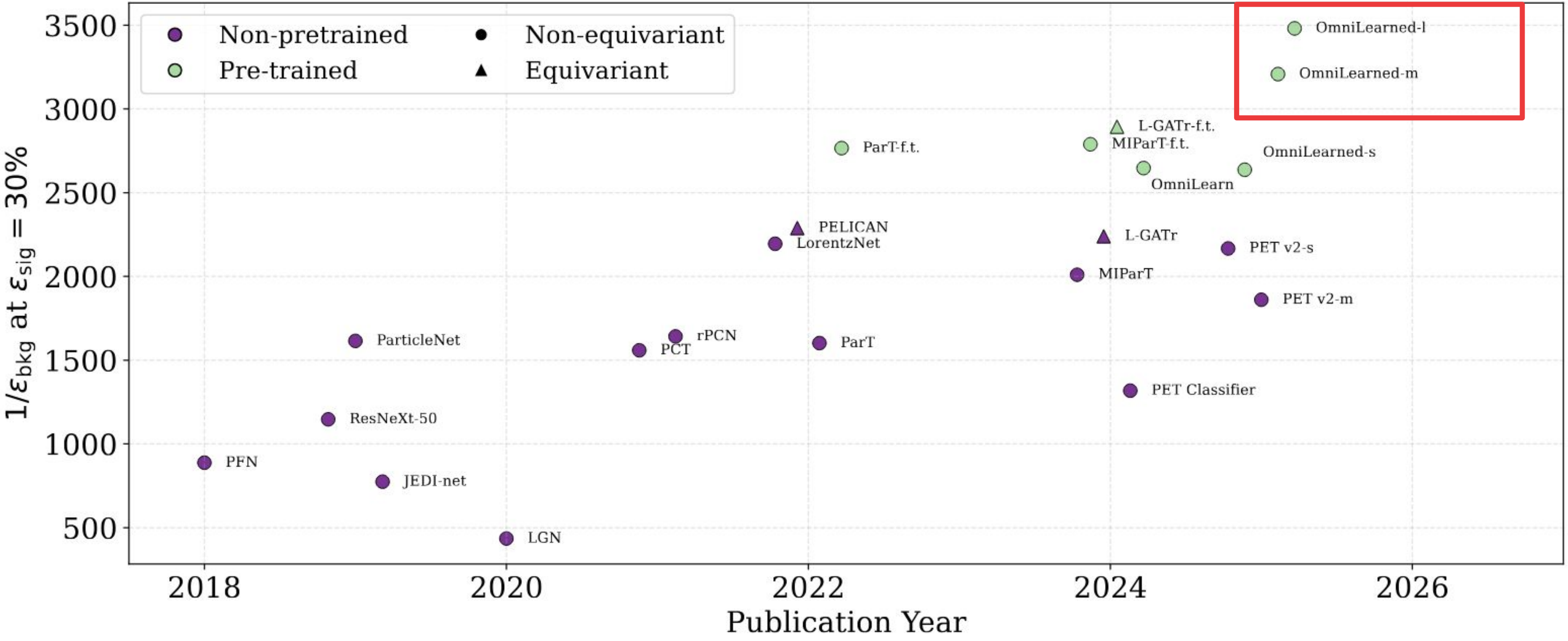
- 210 Jet Classes
- Mixture of simulations with real experimental data

Jet Classification: Top Tagging

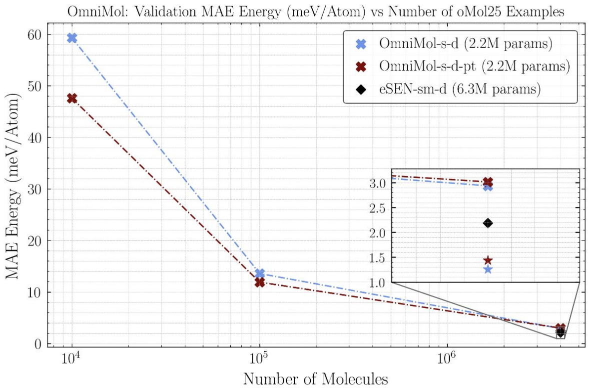
	Acc	AUC	$1/\epsilon_B$	
			$\epsilon_S = 0.5$	$\epsilon_S = 0.3$
ResNeXt-50 [19]	0.936	0.9837	302 ± 5	1147 ± 58
P-CNN [19]	0.930	0.9803	201 ± 4	759 ± 24
PFN [16]	-	0.9819	247 ± 3	888 ± 17
ParticleNet [19]	0.940	0.9858	397 ± 7	1615 ± 93
JEDI-net [18]	0.9300	0.9807	-	774.6
PCT [22]	0.940	0.9855	392 ± 11	1559 ± 98
LGN [63]	0.929	0.964	-	435 ± 95
rPCN [20]	-	0.9845	364 ± 9	1642 ± 93
LorentzNet [64]	0.942	0.9868	498 ± 18	2195 ± 173
PELICAN [65]	0.9425	0.9869	-	2289 ± 204
ParT [14]	0.940	0.9858	413 ± 16	1602 ± 81
ParT-f.t. [14]	**0.944**	0.9877	$\mathbf{691 \pm 15}$	2766 ± 130
Mixer [66]	-	0.9859	416	-
MIParT [26]	0.942	0.9868	505 ± 8	2010 ± 97
MIParT-f.t. [26]	0.944	0.9878	640 ± 10	2789 ± 133
L-GATr [67]	0.9423	0.9870	540 ± 20	2240 ± 70
L-GATr-f.t. [67]	**0.9446**	0.9879	651 ± 11	2894 ± 84
PET [11, 12]	0.938	0.9848	340 ± 12	1318 ± 39
OmniLearn [11, 12]	0.942	0.9872	568 ± 9	2647 ± 192
PET v2-s	0.9427	0.987	505 ± 14	2167 ± 153
OmniLearned-s	0.944	0.9875	565 ± 12	2637 ± 128
PET v2-m	0.9423	0.987	482 ± 11	1861 ± 61
OmniLearned-m	**0.944**	**0.9880**	656 ± 12	3208 ± 176
OmniLearned-l	**0.944**	**0.9880**	$\mathbf{688 \pm 9}$	$\mathbf{3486 \pm 157}$

- **Fine-tune the pre-trained model** on the Top Tagging Community Dataset, a traditional benchmark dataset in HEP

OmniLearned achieves SOTA performance

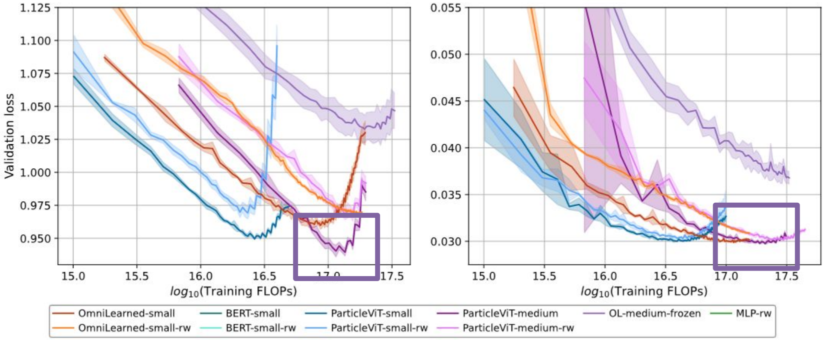


OmniLearned Generalizes across scientific fields



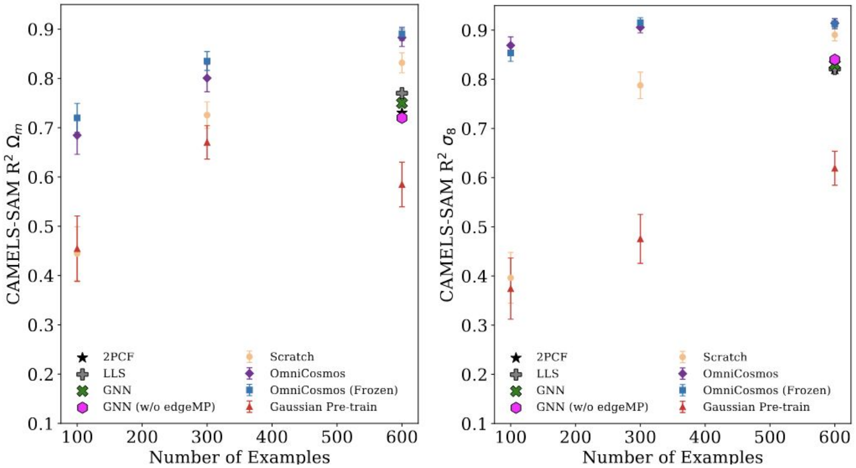
Interatomic potentials from molecules

I. Elsharkawy, **V. Mikuni**, B. Nachman, W. Bhimji [arXiv:2601.10791](https://arxiv.org/abs/2601.10791)



Pion Classification and energy regression ar Minerva

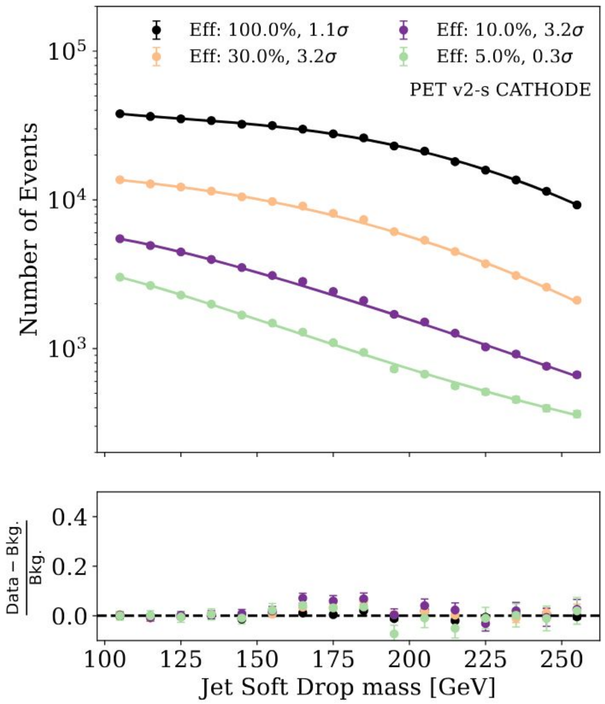
G. Krzmanc, **V. Mikuni**, B. Nachman, C. Wilkinson [arXiv:2604.12364](https://arxiv.org/abs/2604.12364)



Inference of Cosmological Parameters from galaxy halos

V. Mikuni, I. Elsharkawy, B. Nachman [arXiv:2512.24422](https://arxiv.org/abs/2512.24422)

Anomaly Detection

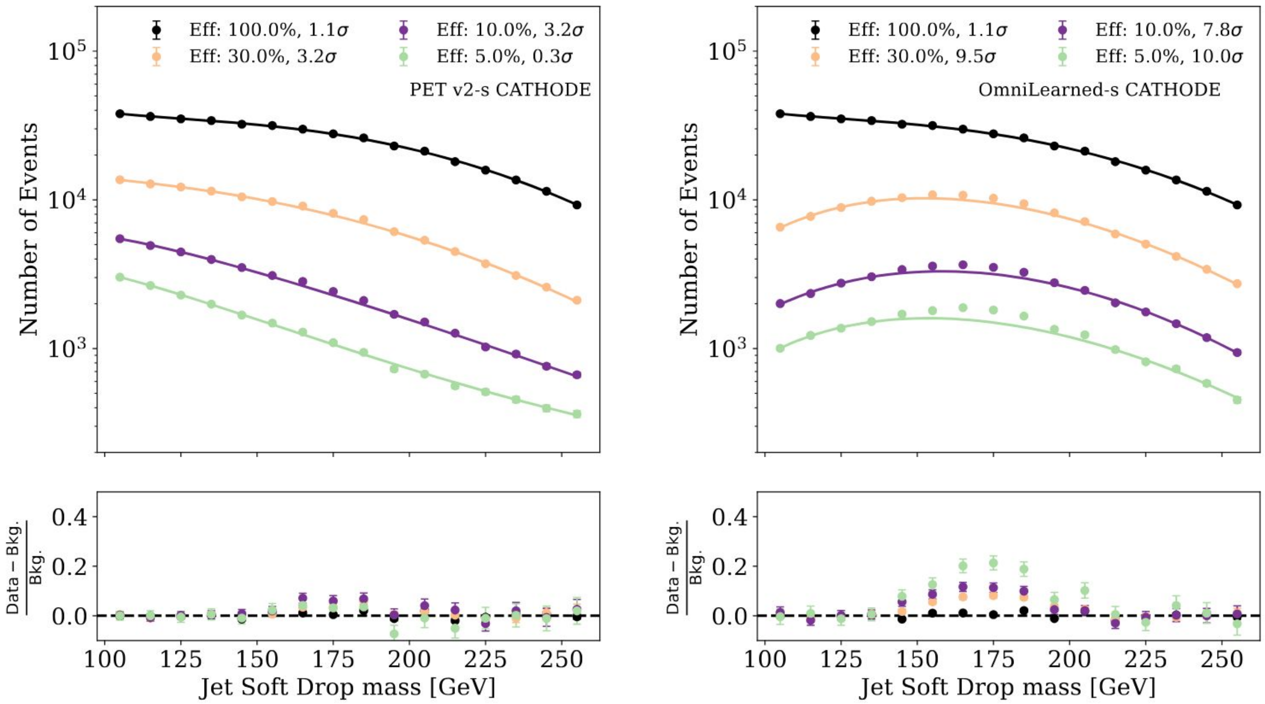


Use **experimental data**

Signal: top quarks, corresponding to 0.1% of the data

- Model trained from scratch **unable to find the anomaly**

Anomaly Detection

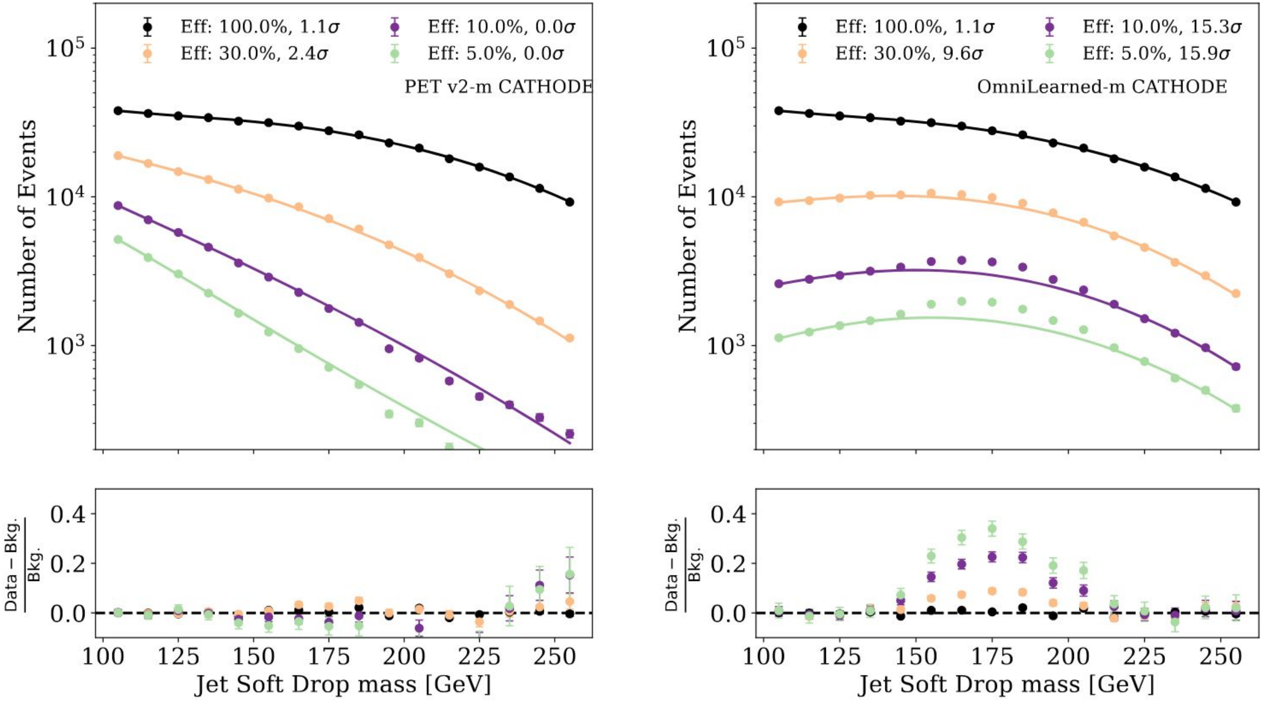


Use **experimental data**

Signal: top quarks, corresponding to 0.1% of the data

- Model trained from scratch **unable to find the anomaly**

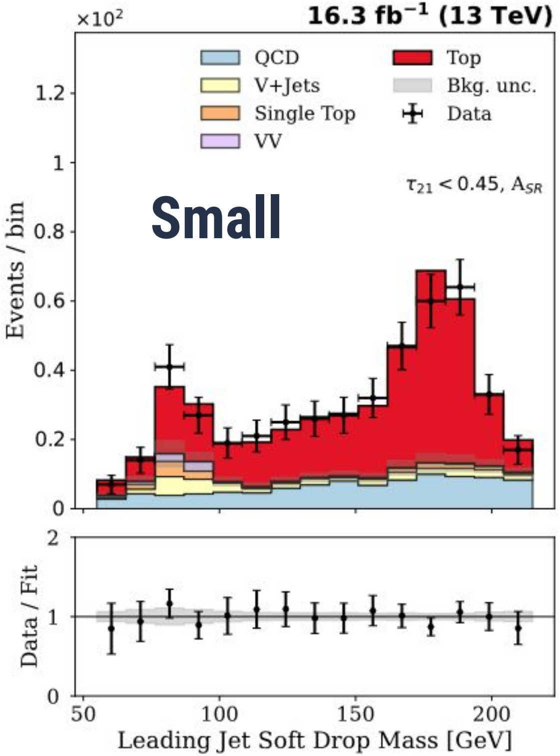
Anomaly Detection

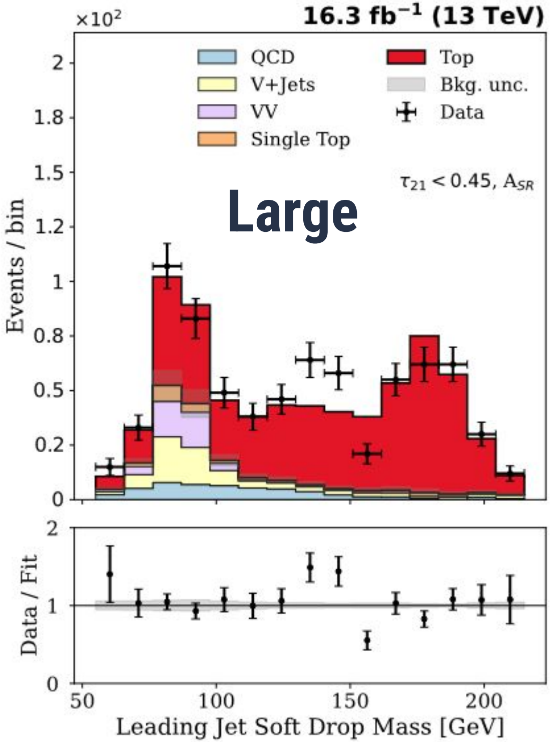


Use **experimental data**
Signal: top quarks, corresponding to 0.1% of the data

- Model trained from scratch **unable to find the anomaly**

Anomaly Detection

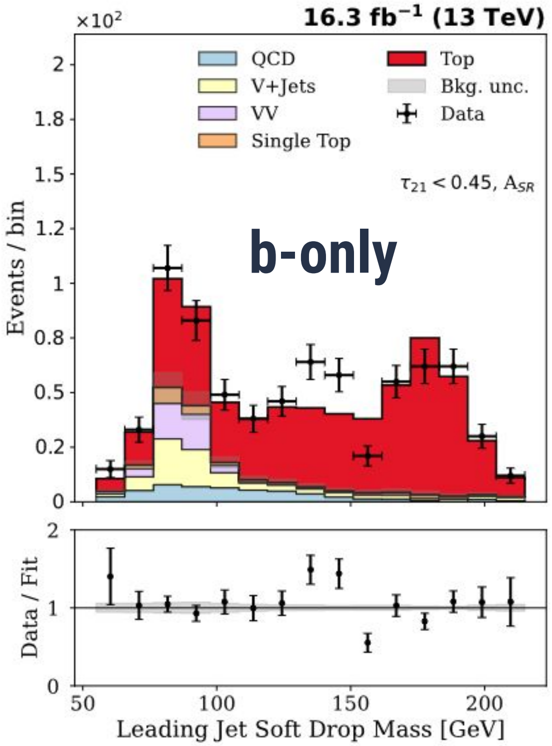


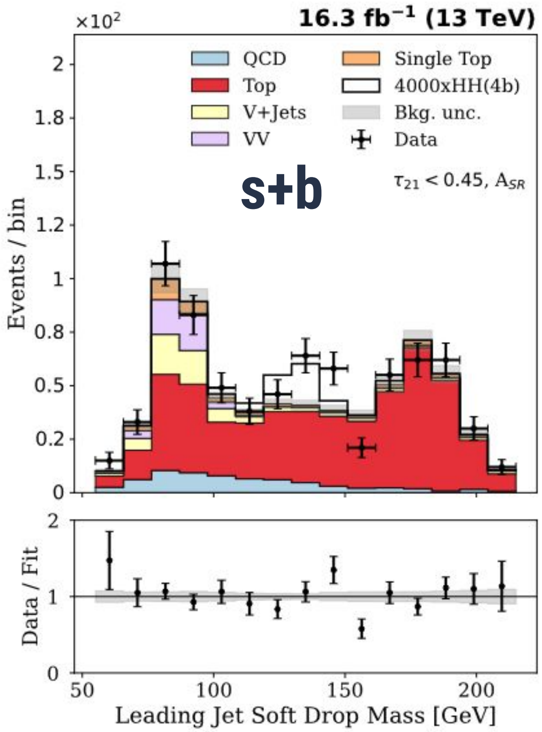


We perform a **full analysis** of data events flagged as anomalous in a dijet system

- Dominated by top quark pair production

Anomaly Detection

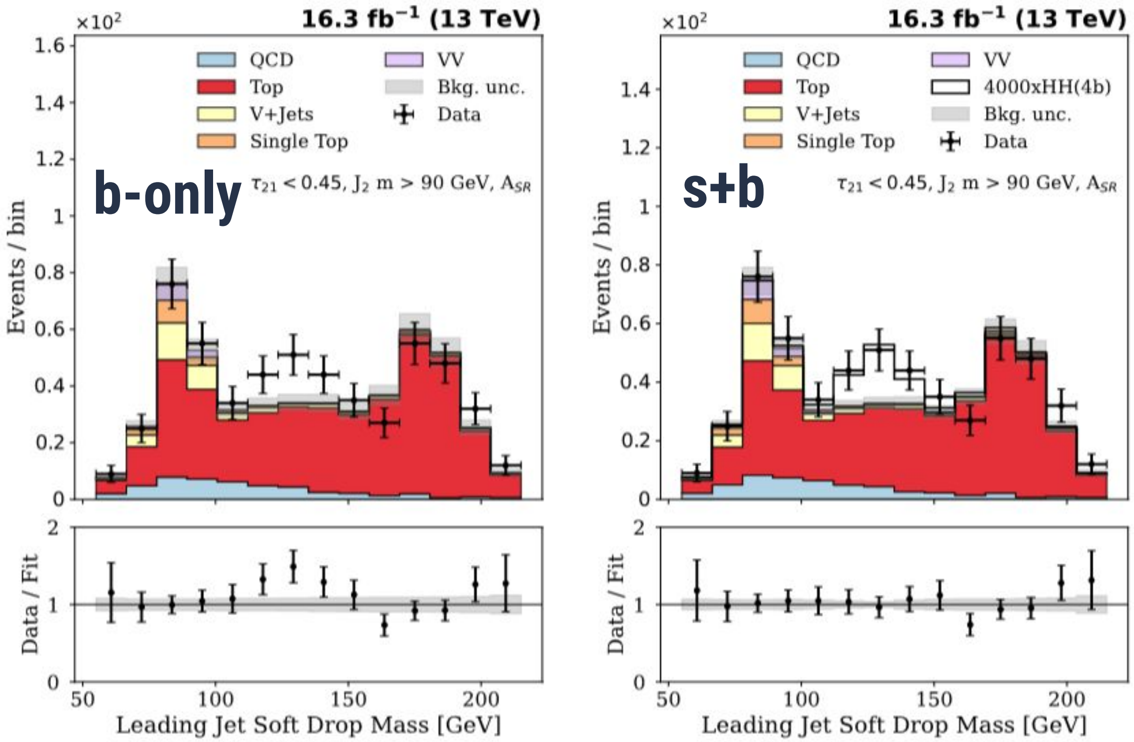




We perform a **full analysis** of data events flagged as anomalous in a dijet system

- If we consider a HH signal, the observed significance is **3.3**

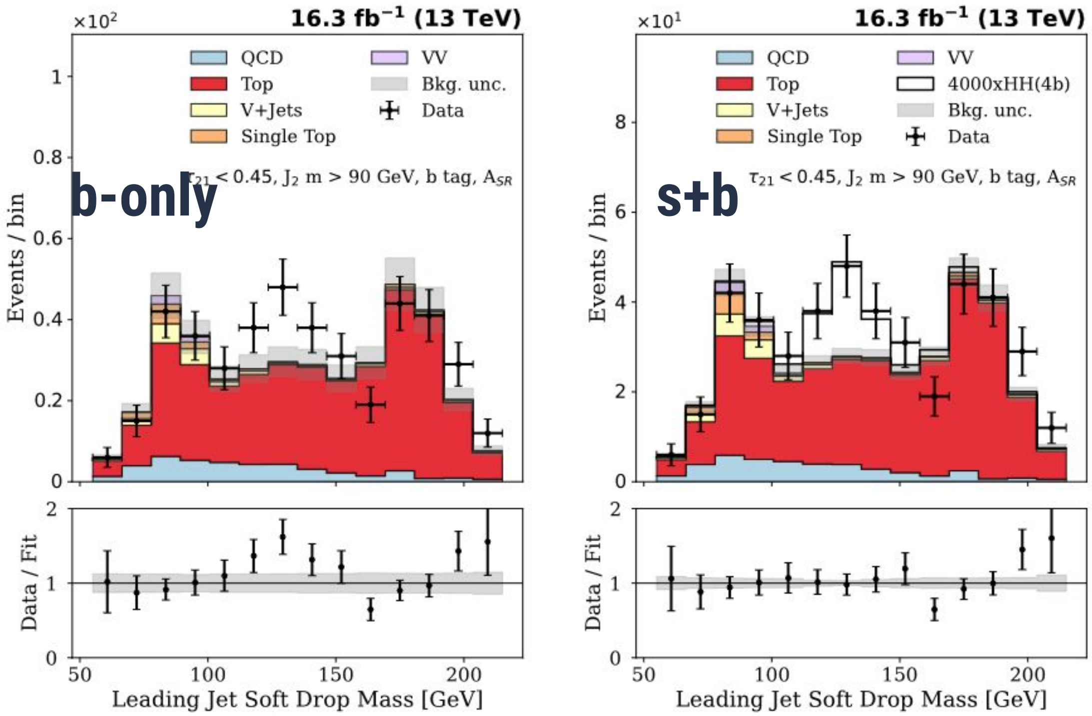
Anomaly Detection



We perform a **full analysis** of data events flagged as anomalous in a dijet system

- If we consider a HH signal, and subleading jet mass > 90 GeV, the observed significance is **3.84**

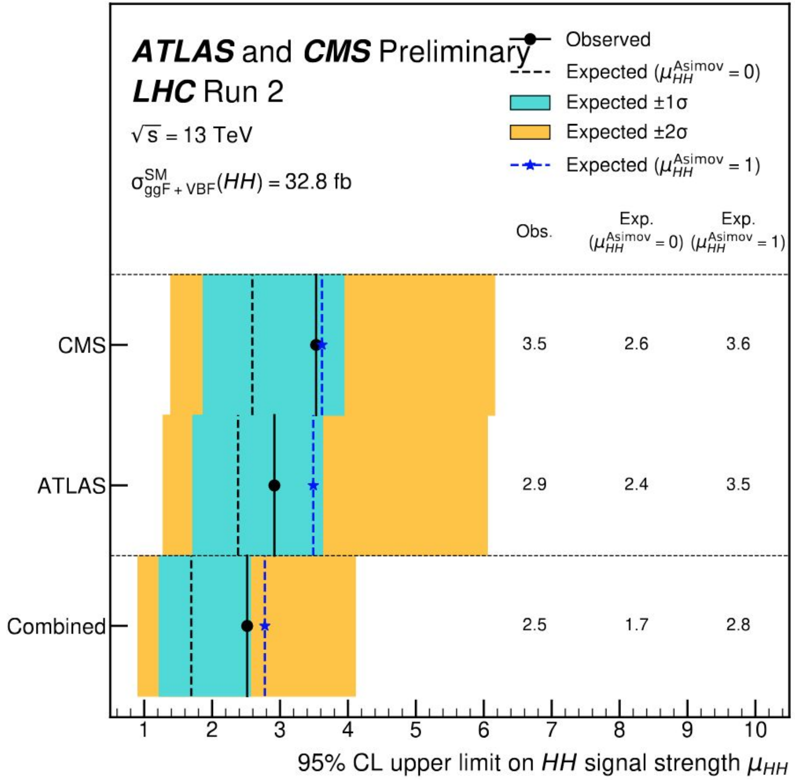
Anomaly Detection



We perform a **full analysis** of data events flagged as anomalous in a dijet system

- If we consider a HH signal, subleading jet mass > 90 GeV, and 1 b-tagged jet, the observed significance is **4.25**

Anomaly Detection

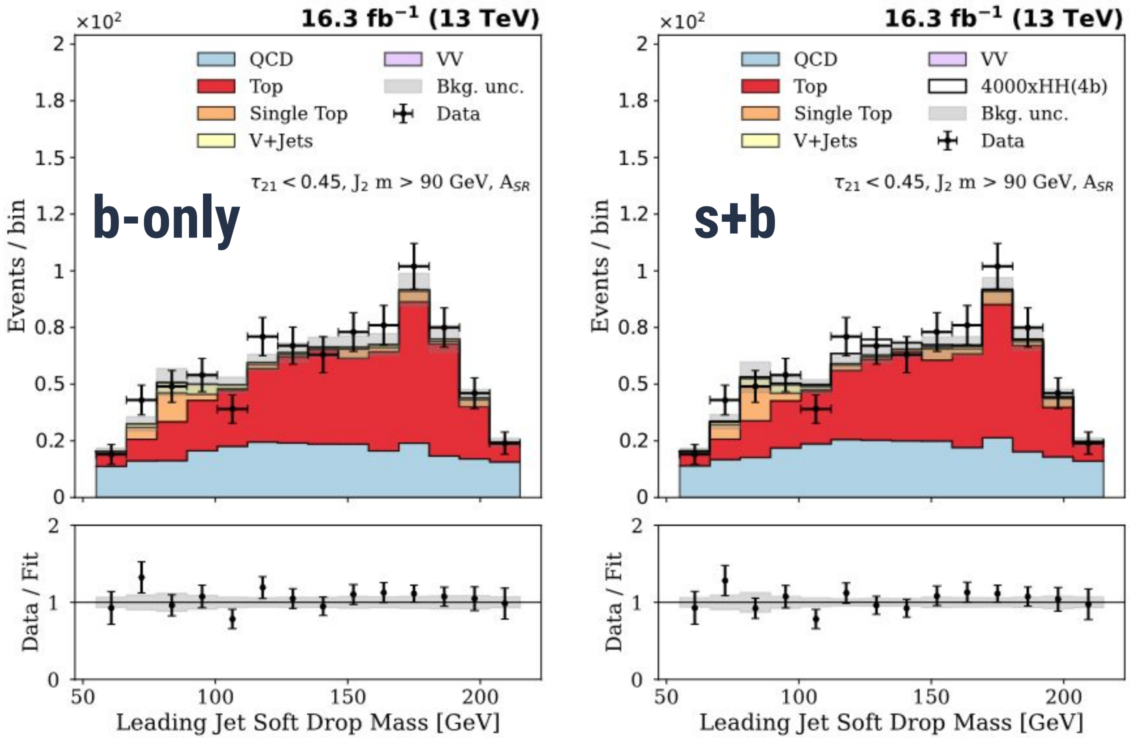


While compatible with the signal template, it is **highly unlikely** that the excess comes from SM HH

- Observed yield: 4000 x HH
- Current limits at 2.5 x HH

CMS-PAS-HIG-25-014

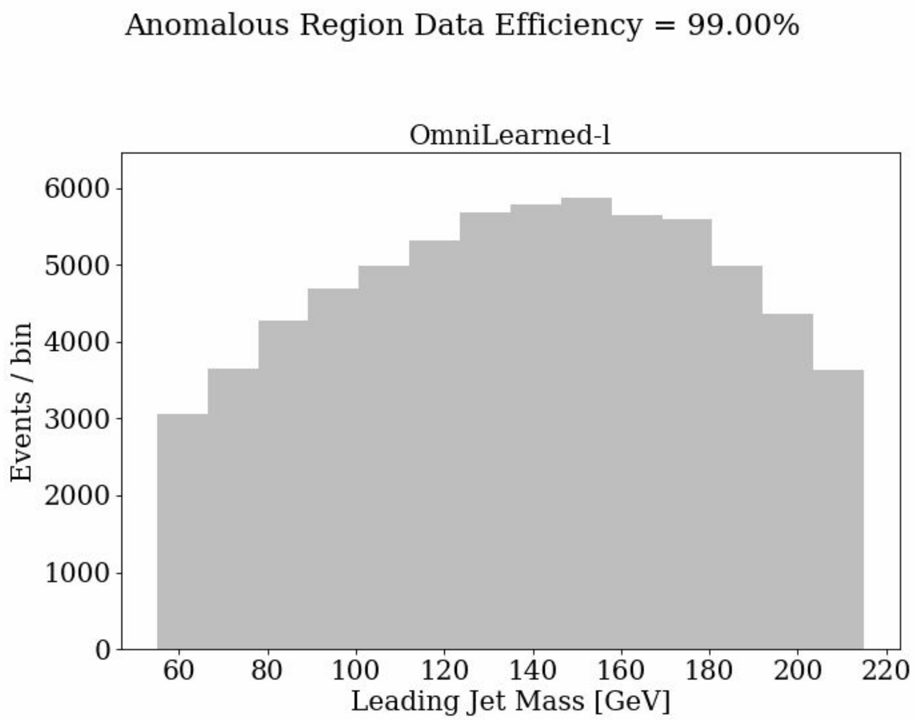
Anomaly Detection



If we change the anomaly score for a X(bb) tagger, **no excess is observed**

- Possibly **not a bb** final state

Conclusion



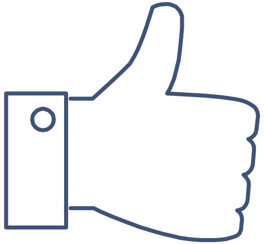
Anomaly Detection aims to find new physics with less assumptions

Foundation Models improve sensitivity to anomalous events

Intriguing excess observed, stay tuned for more developments

Read more:

https://arxiv.org/pdf/2603.23593

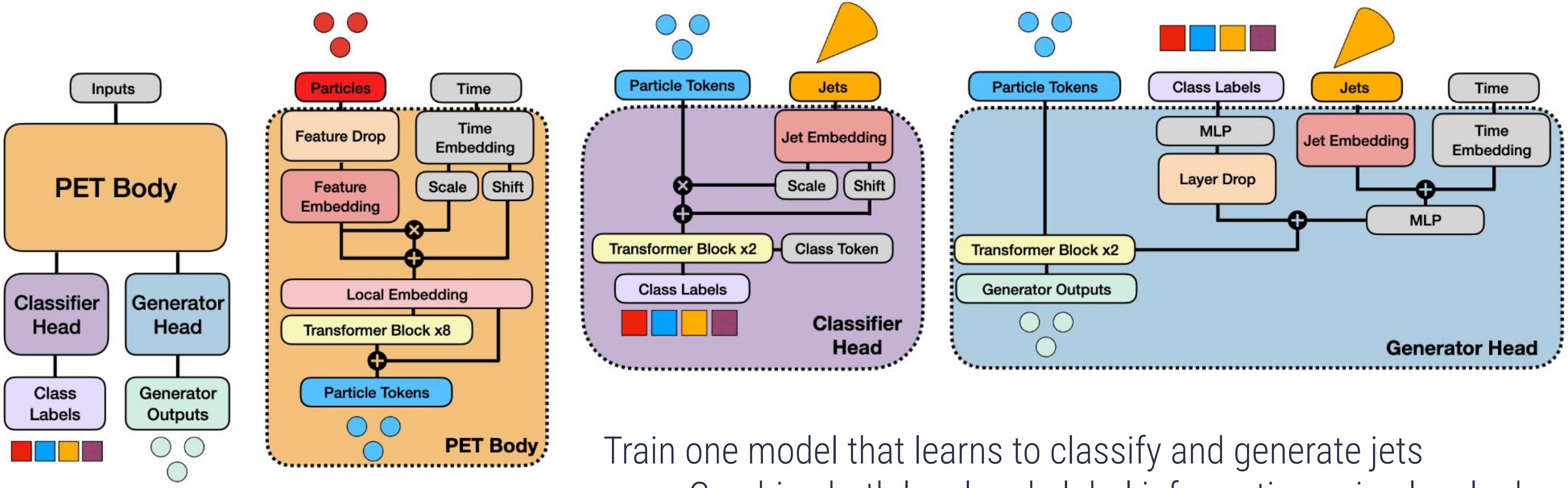


THANKS!

Any questions?

Backup

PET

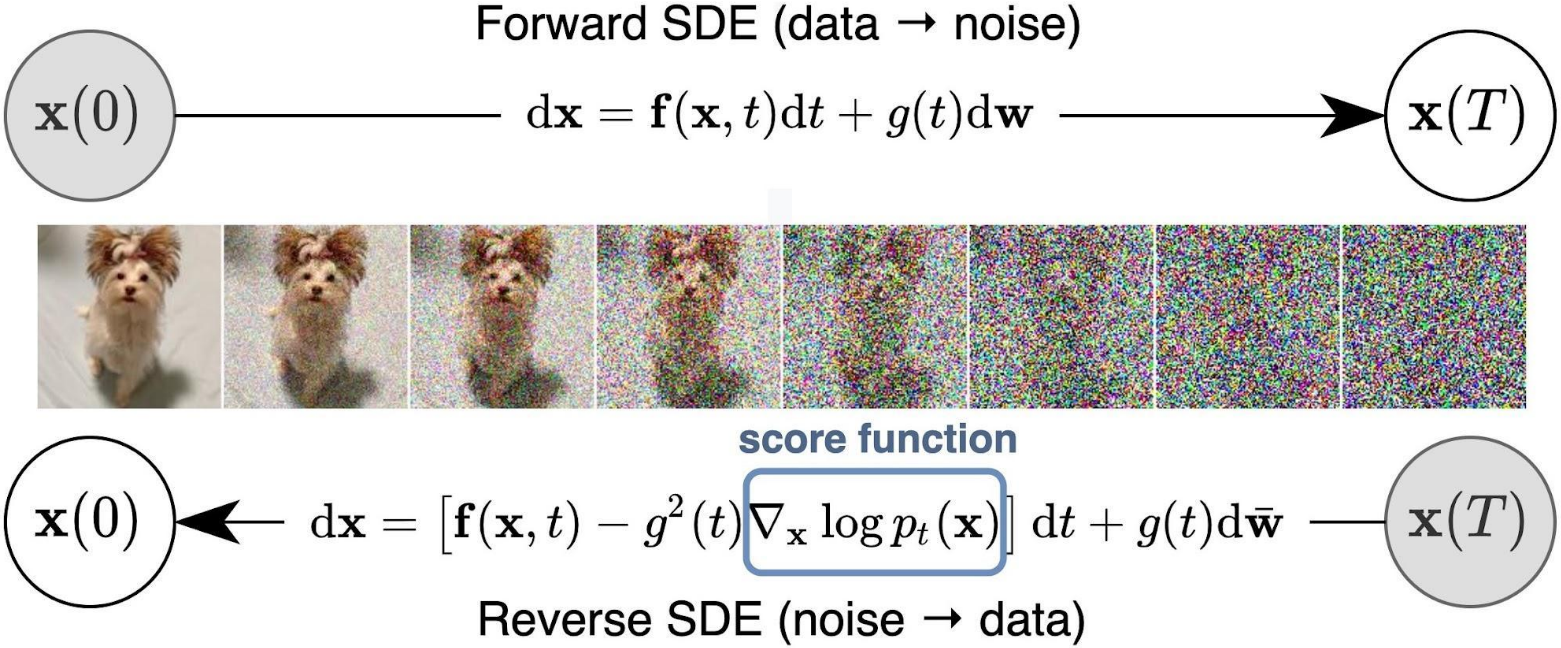


Train one model that learns to classify and generate jets

- Combine both local and global information using local edges and a transformer: **P**oint-**E**dge **T**ransformer

More details at: https://arxiv.org/abs/2404.16091

Diffusion Generative Models



Source: https://yang-song.net/blog/2021/score/

Loss function

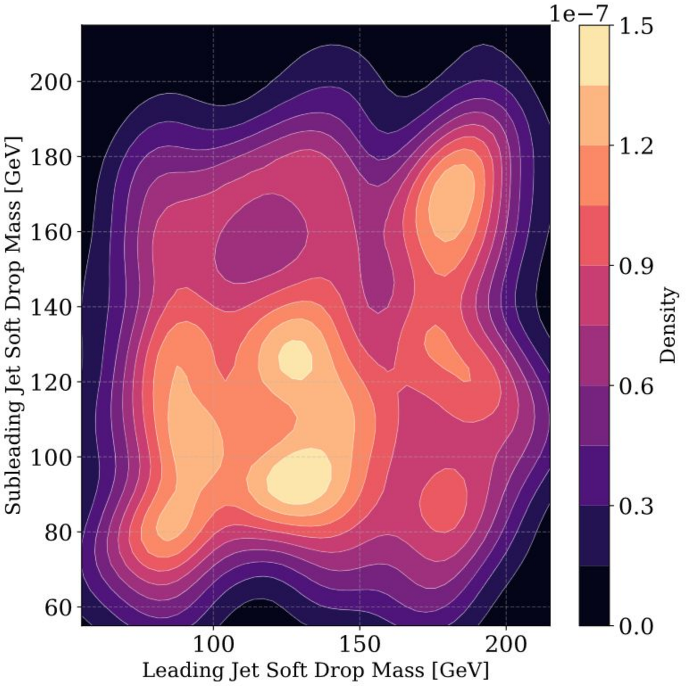
$$\mathcal{L} = \mathcal{L}_{\text{class}} + \mathcal{L}_{\text{gen}} + \mathcal{L}_{\text{class smear}}$$
$$= \text{CE}(y, y_{\text{pred}}) + \|\mathbf{v} - \mathbf{v}_{\text{pred}}\|^2 + \alpha^2 \text{CE}(y, \hat{y}_{\text{pred}})$$

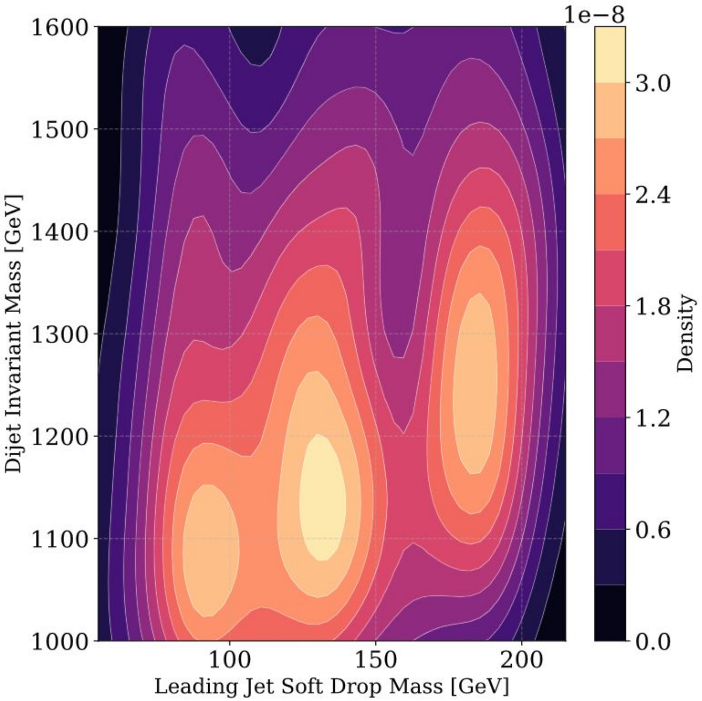
Straightforward loss function:

- **Cross entropy** for each class
- Perturbed data prediction from the **diffusion loss**
- Classification over perturbed inputs: **data augmentation**!

More details at: https://arxiv.org/abs/2404.16091

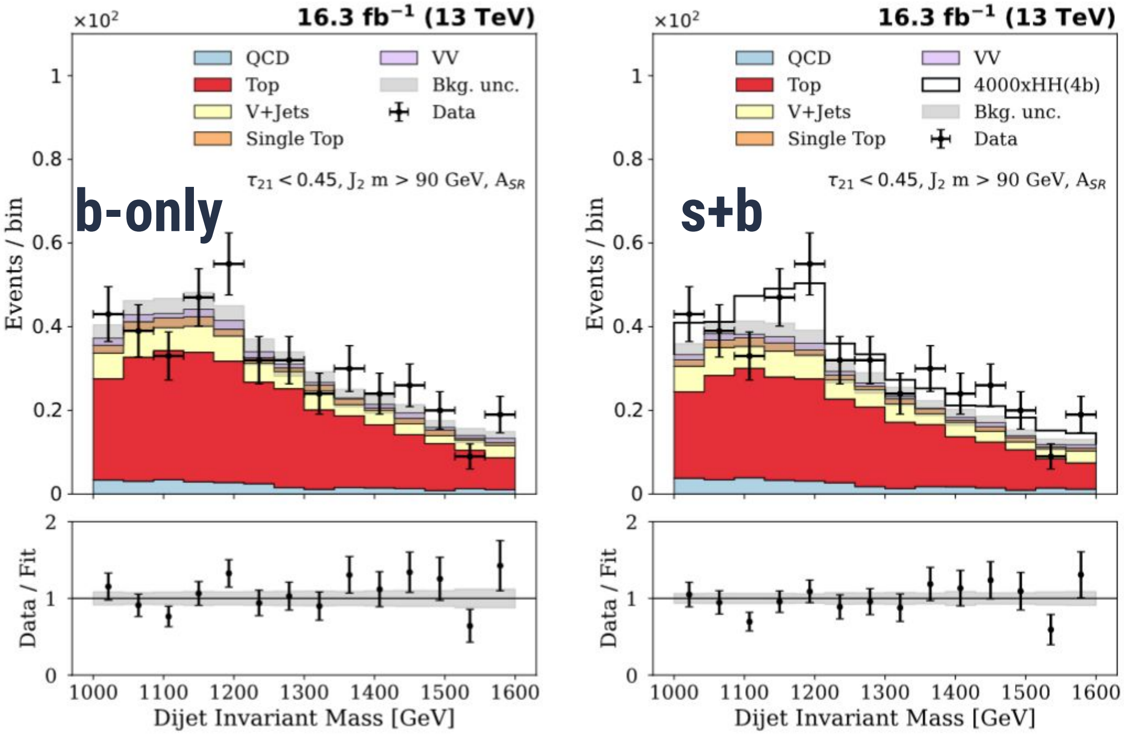
Anomaly Detection





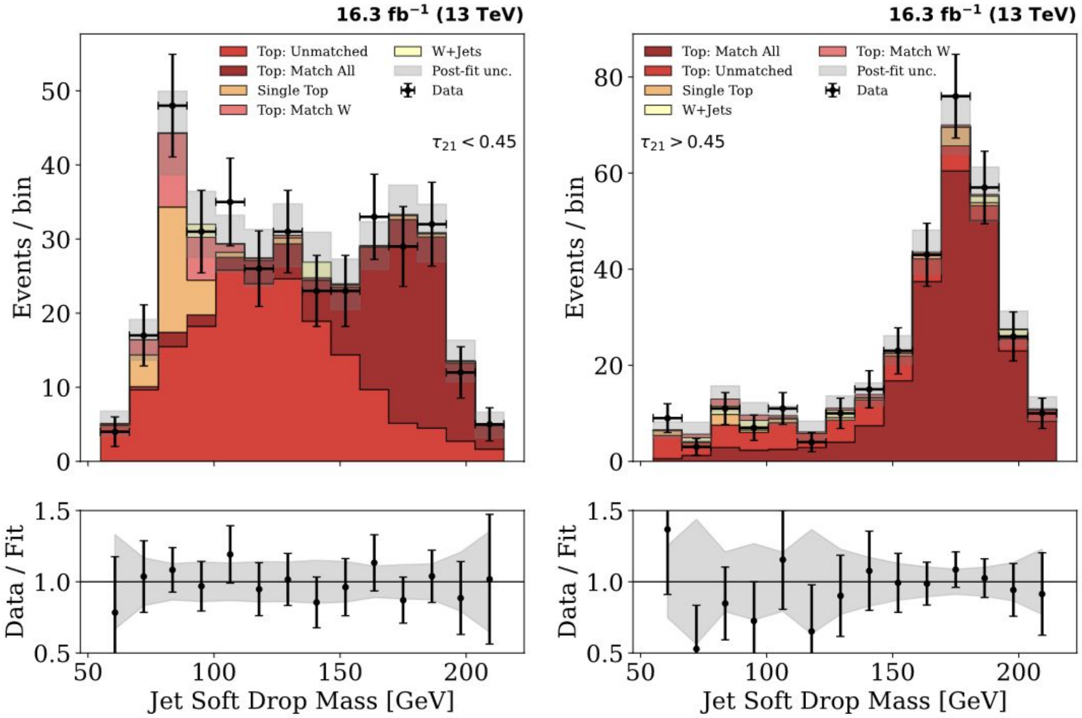
Use Kernel Density Estimation (**KDE**) to estimate the joint pdf

Anomaly Detection



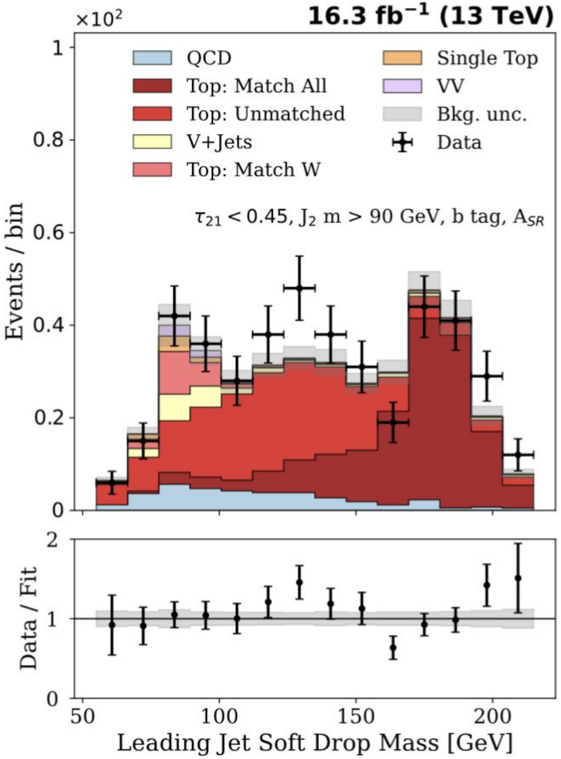
Dijet mass is not very smooth either

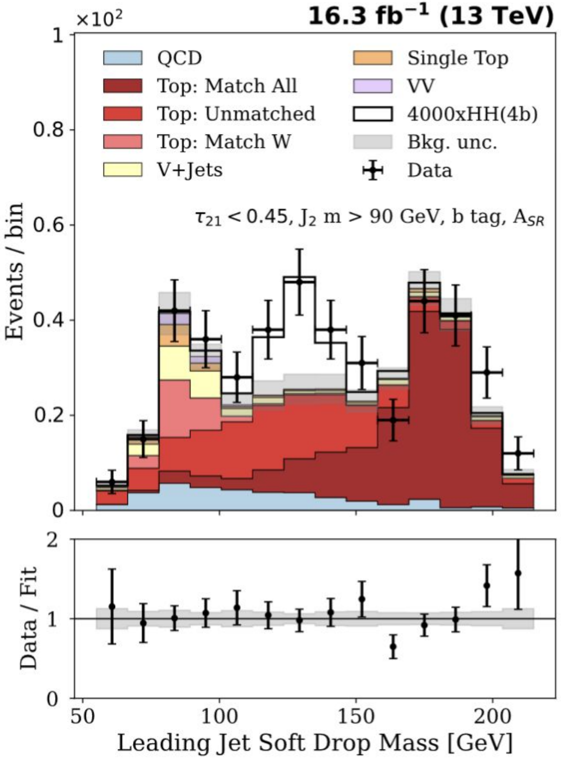
Additional Checks



Single Muon control region used to verify the ttbar agreement

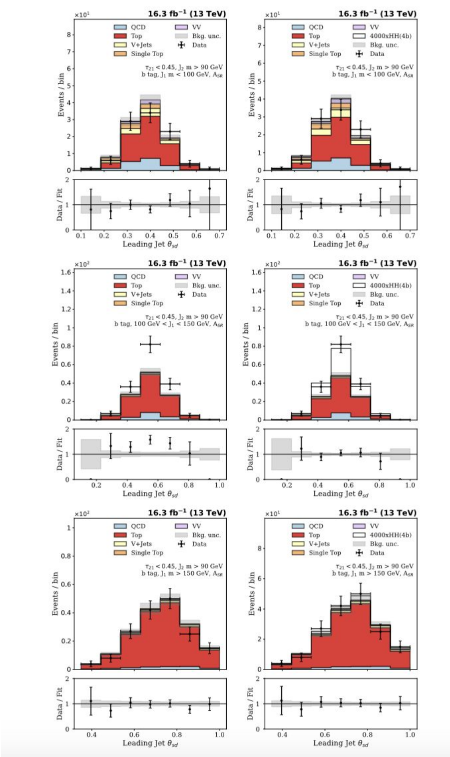
Additional Checks

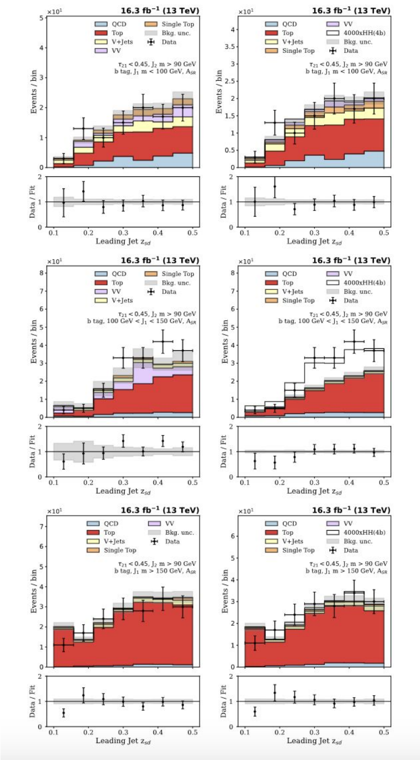




Split ttbar contribution based on matched gen particles

Additional Checks





Compatibility with 2-prong distributions for both theta and z distributions used to calculate the softdrop mass