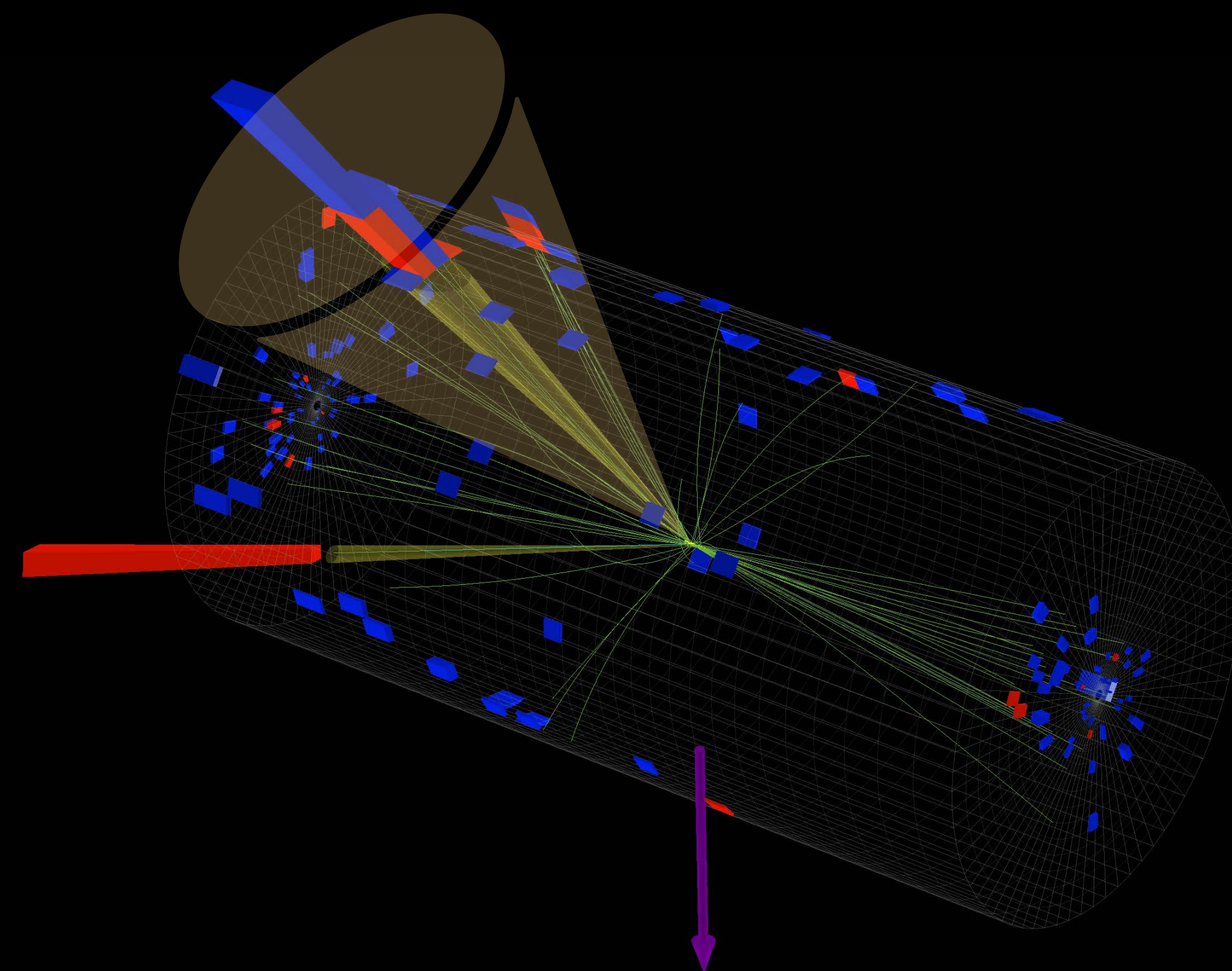


Jet Tagging in the Era of Deep Learning

Huilin Qu (CERN)

ODSL Forum

September 15, 2023

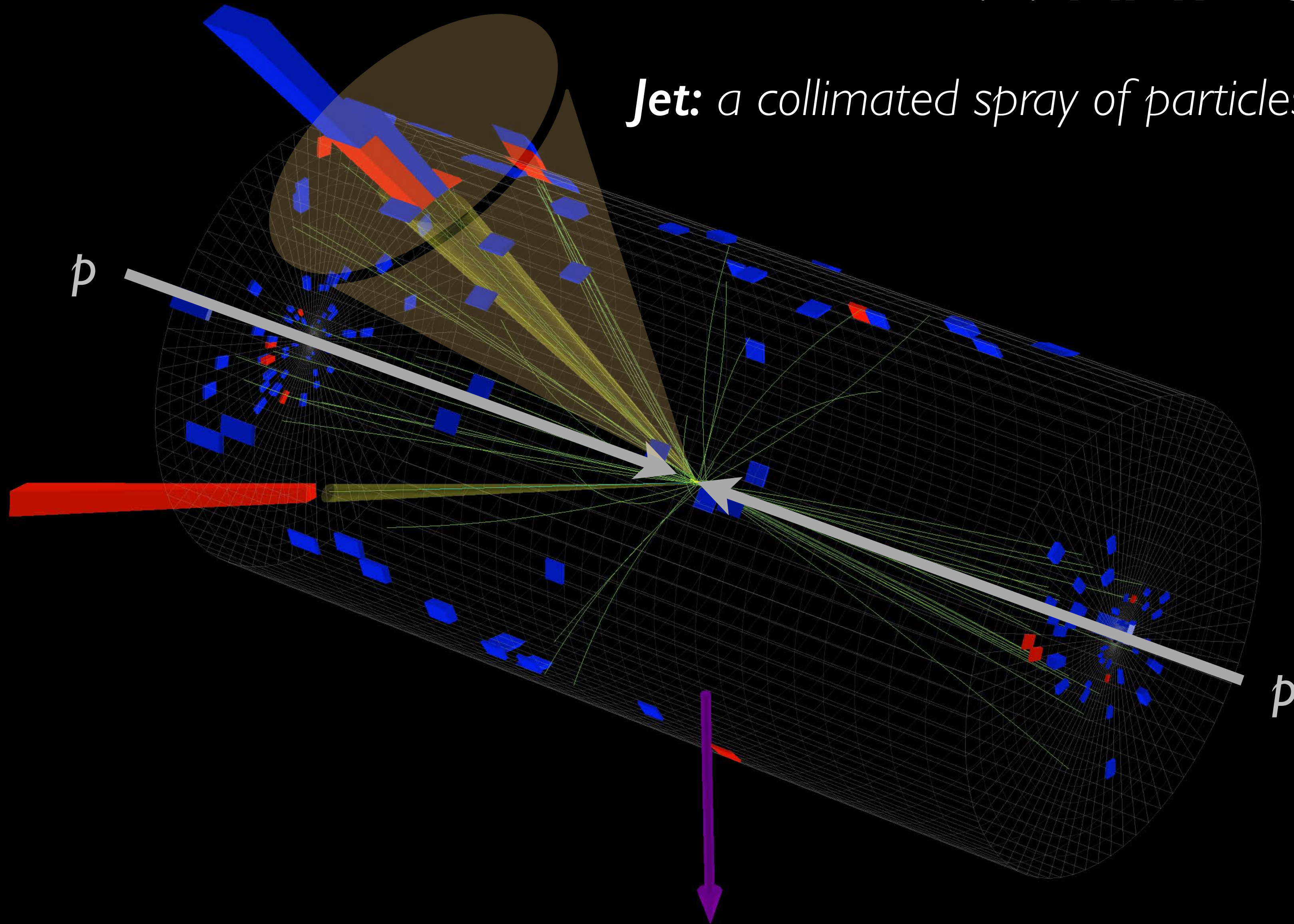




CMS Experiment at LHC, CERN
Data recorded: Sat Aug 5 15:32:22 2017 CEST
Run/Event: 300515 / 205888132

WHAT IS A JET?

Jet: a collimated spray of particles



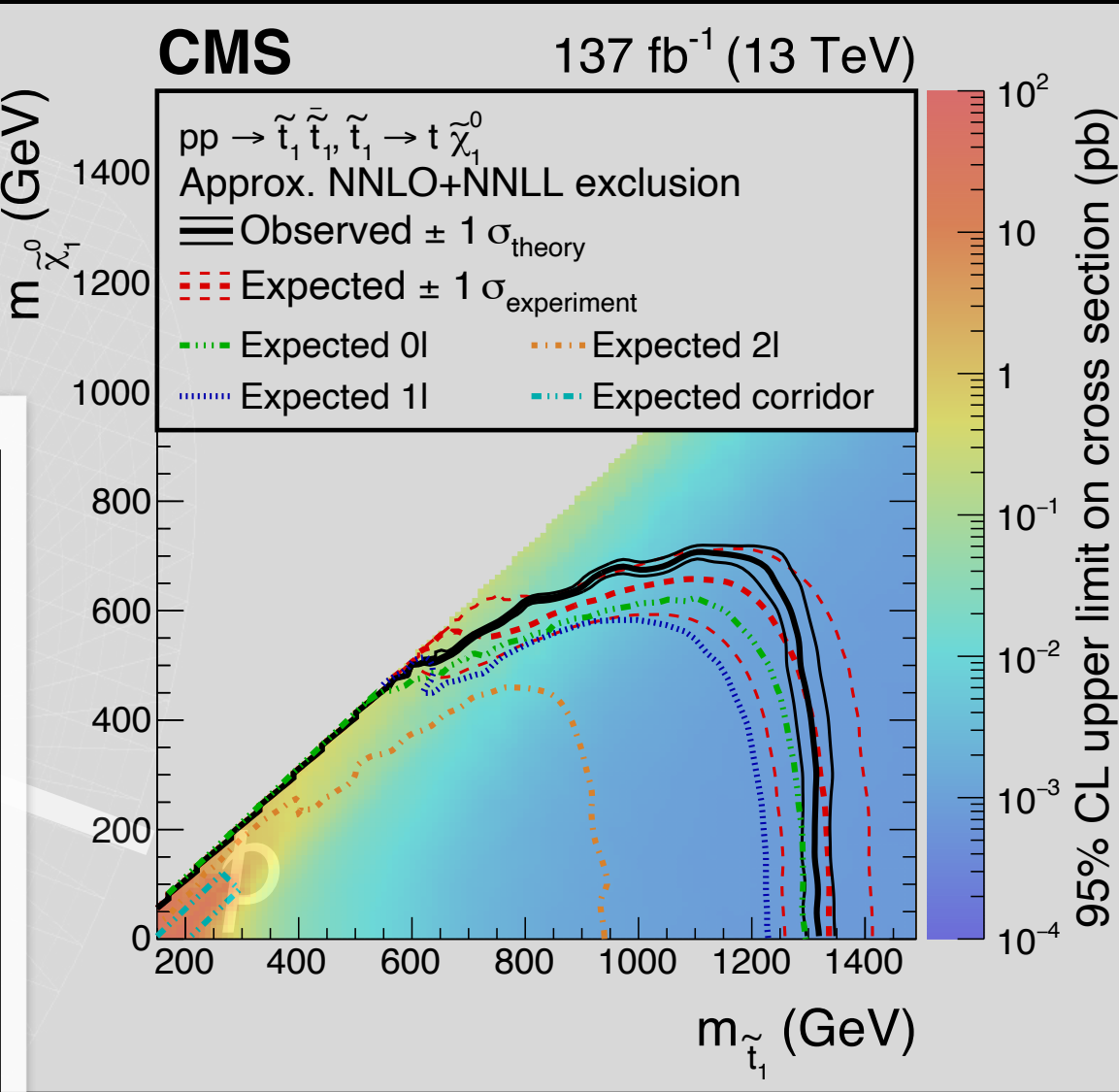
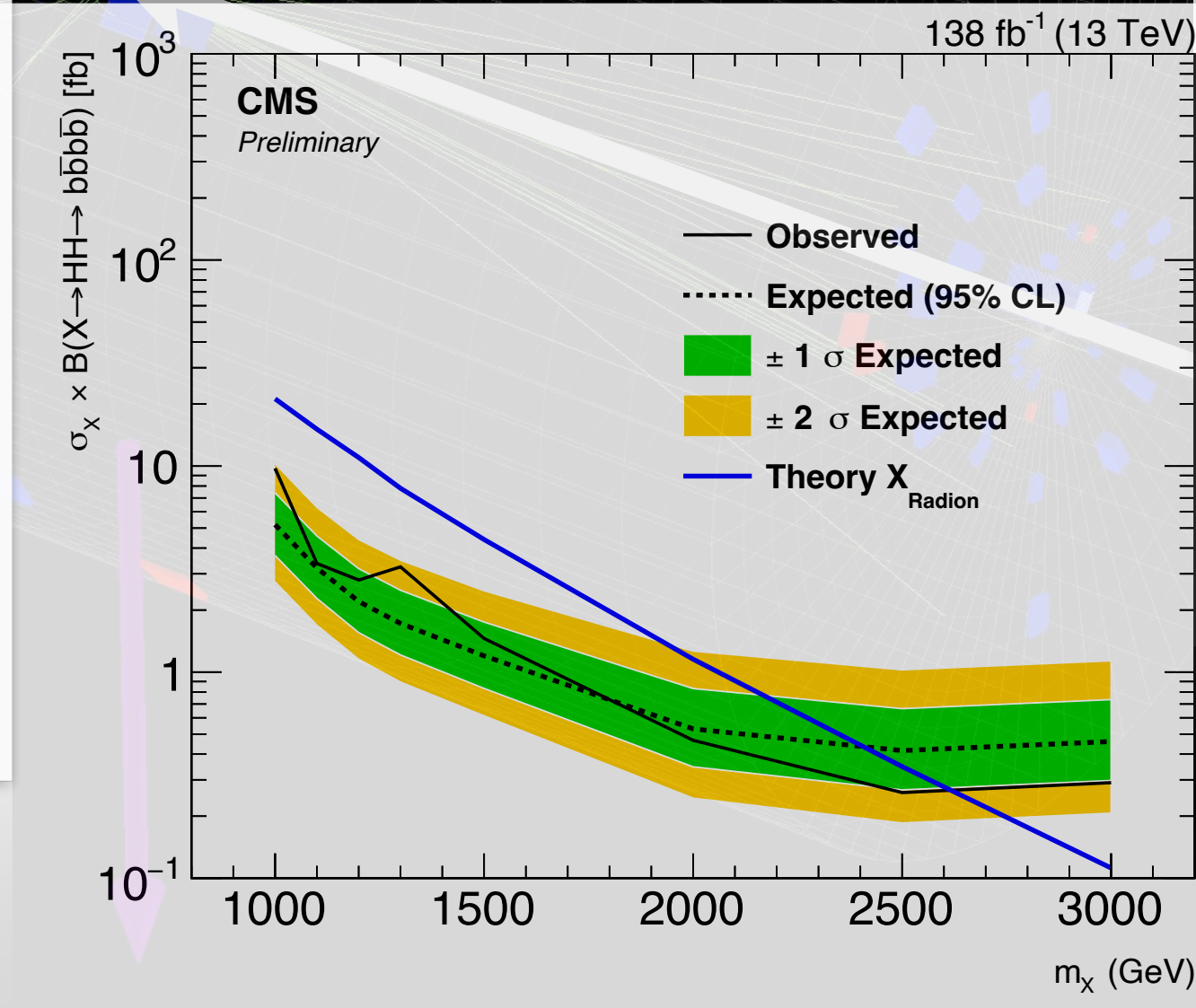
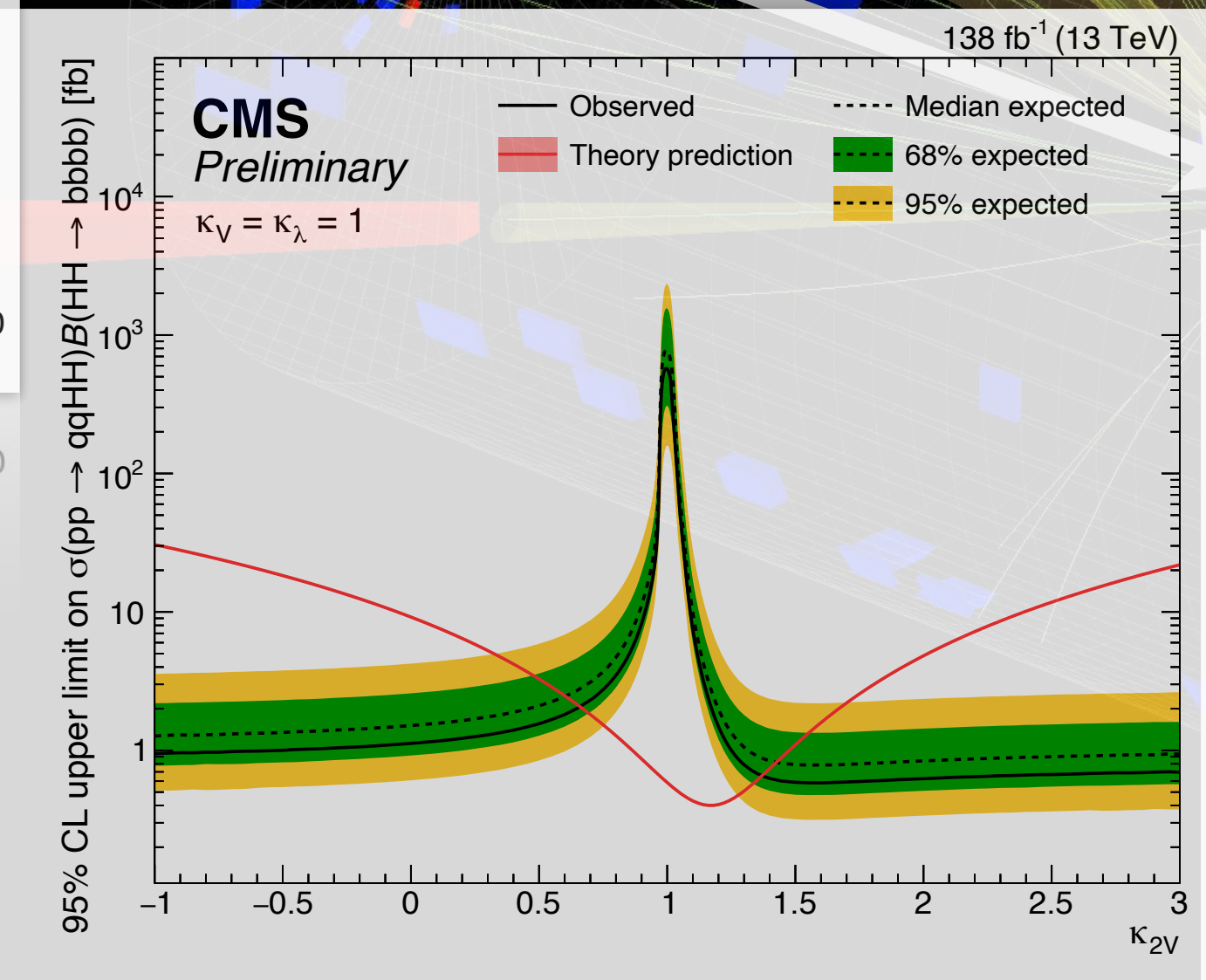
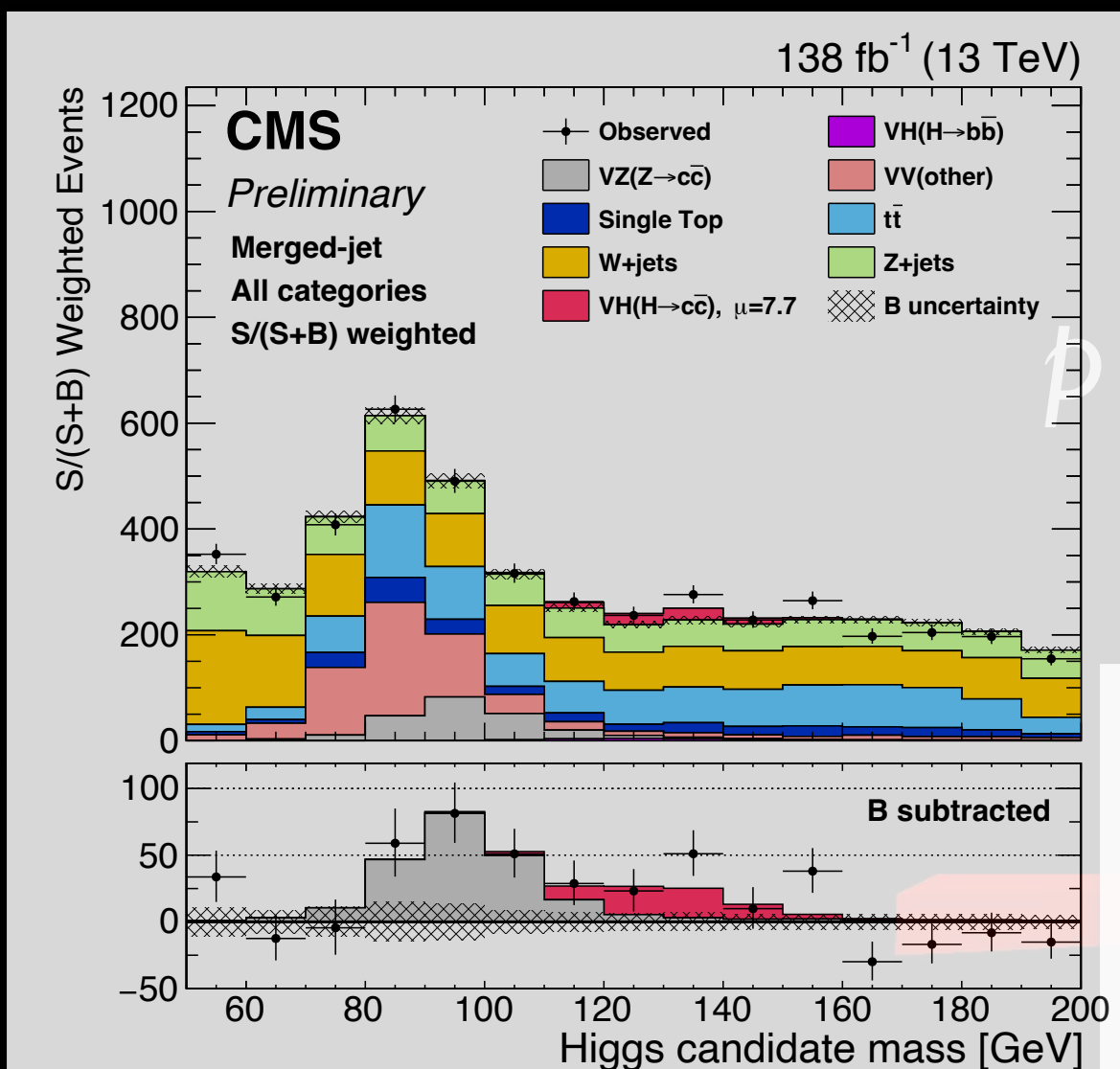
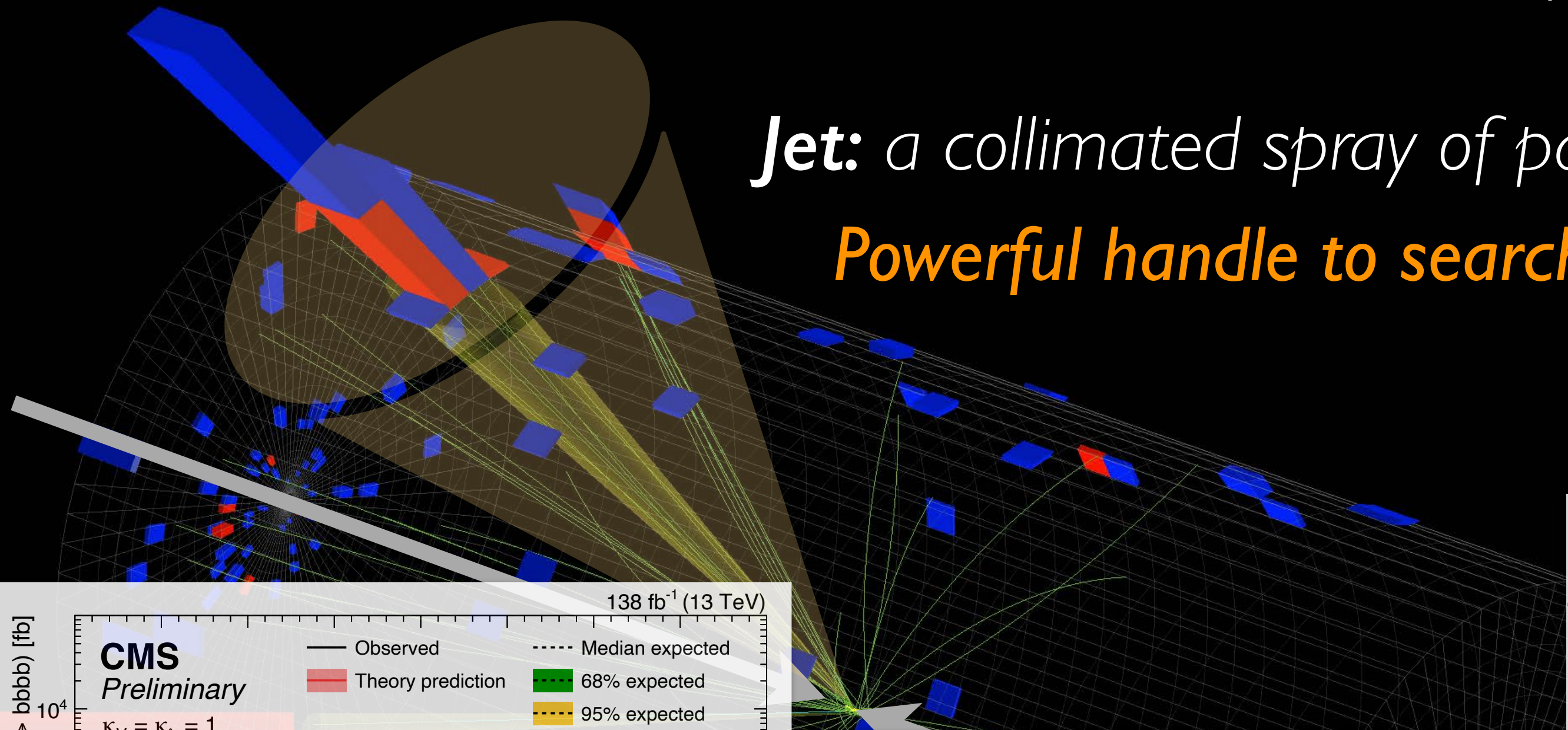


CMS Experiment at LHC, CERN
 Data recorded: Sat Aug 5 15:32:22 2017 CEST
 Run/Event: 300515 / 205888132

WHY JETS?

Jet: a collimated spray of particles

Powerful handle to search for new phenomena



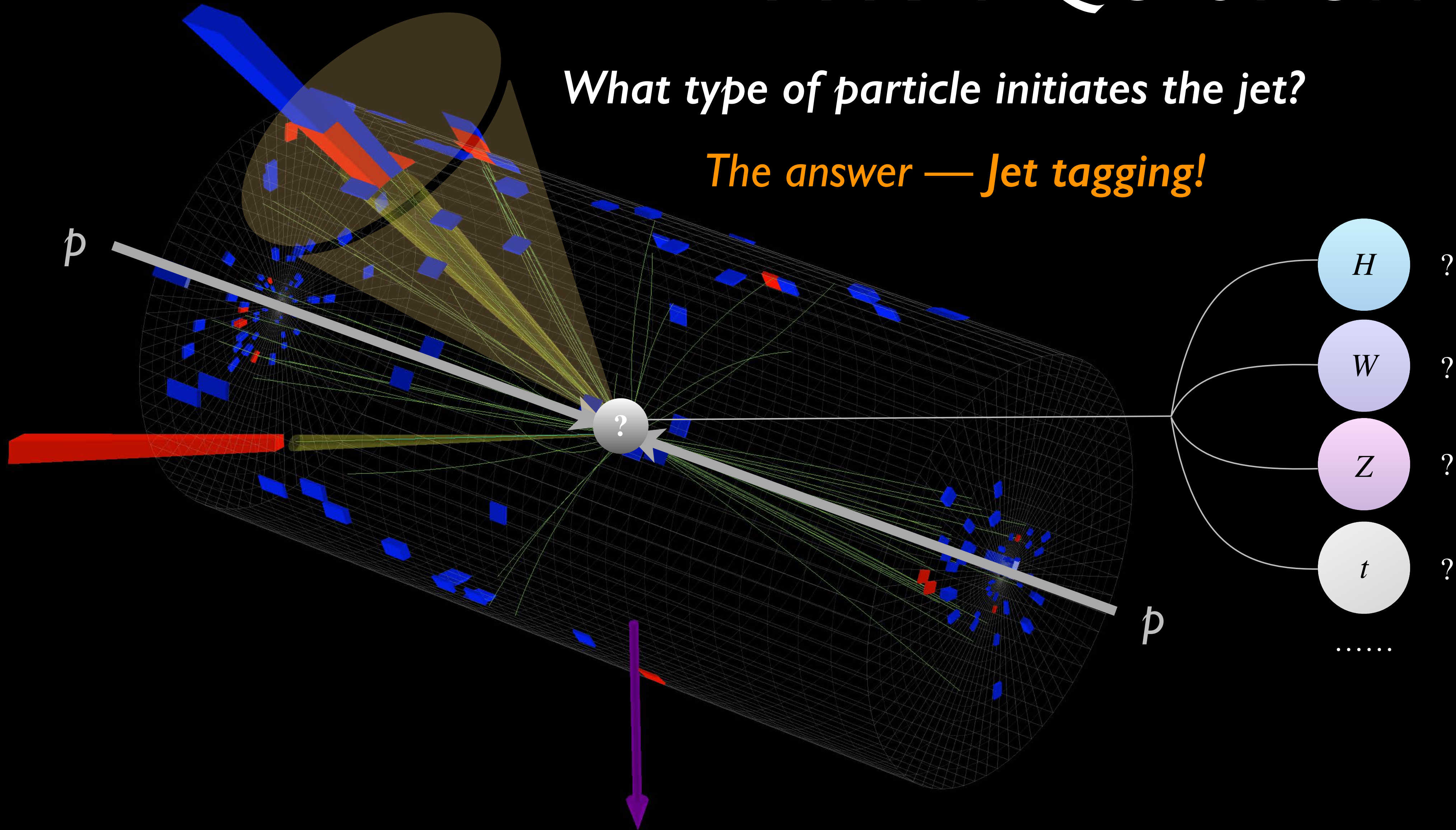


CMS Experiment at LHC, CERN
Data recorded: Sat Aug 5 15:32:22 2017 CEST
Run/Event: 300515 / 205888132

A KEY QUESTION

What type of particle initiates the jet?

The answer — Jet tagging!



JET TAGGING

- Jet tagging: identifying the origin of a jet, i.e., what kind of particle initiates the jet
 - essentially a classification task from the machine learning perspective

Jet flavor tagging

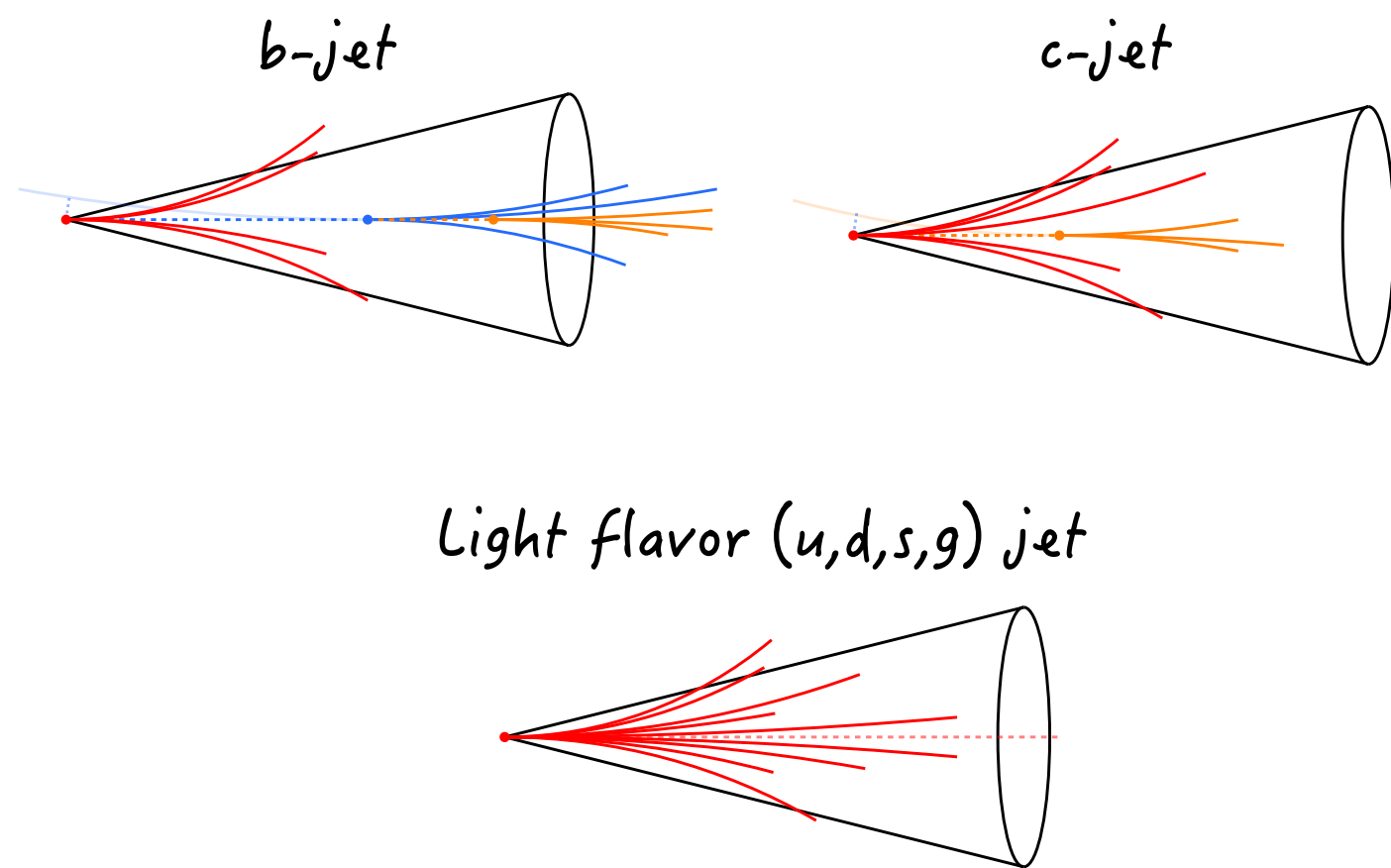


Image credit

Focus of today

Boosted jet tagging

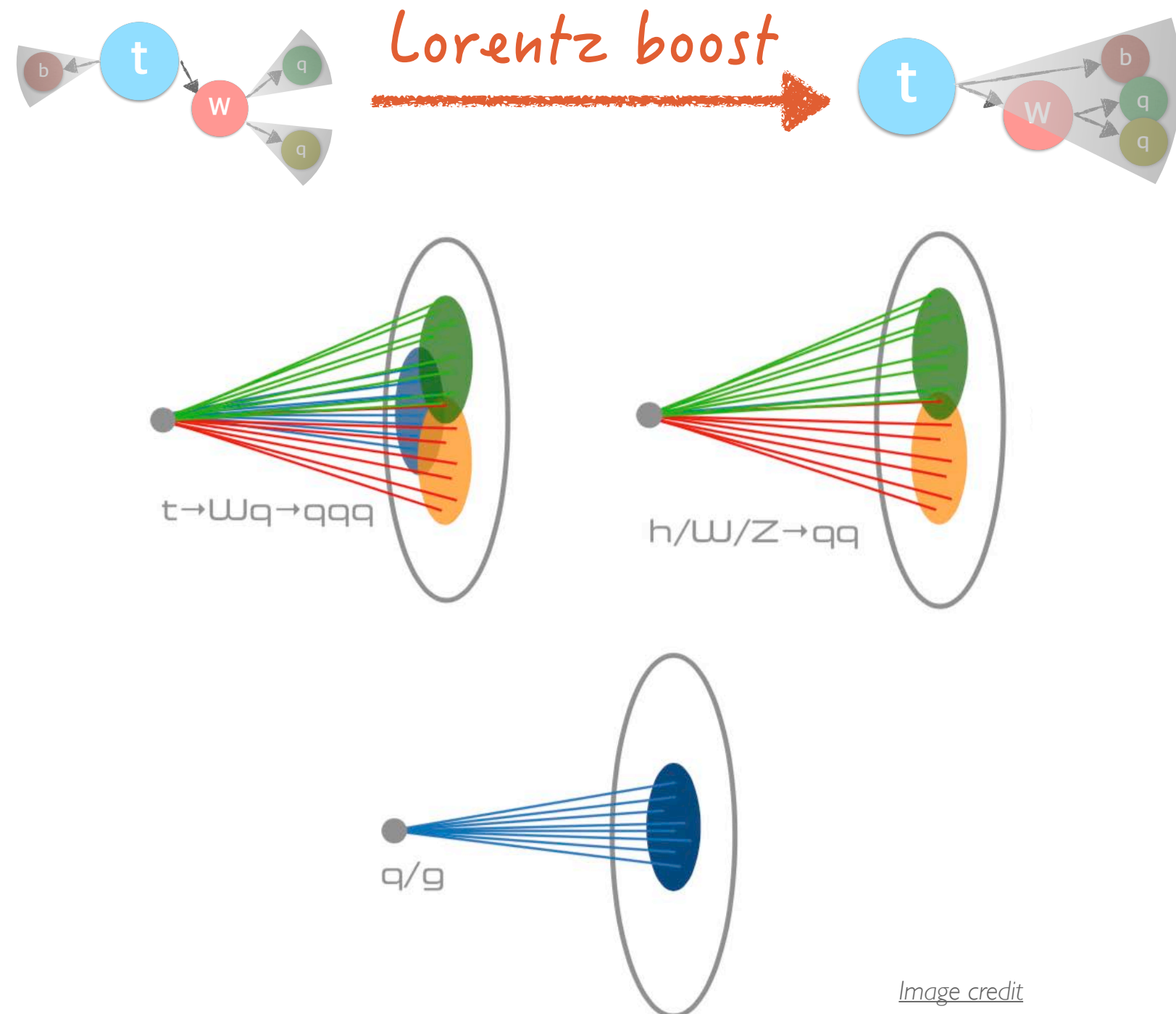


Image credit

Hadronic τ tagging

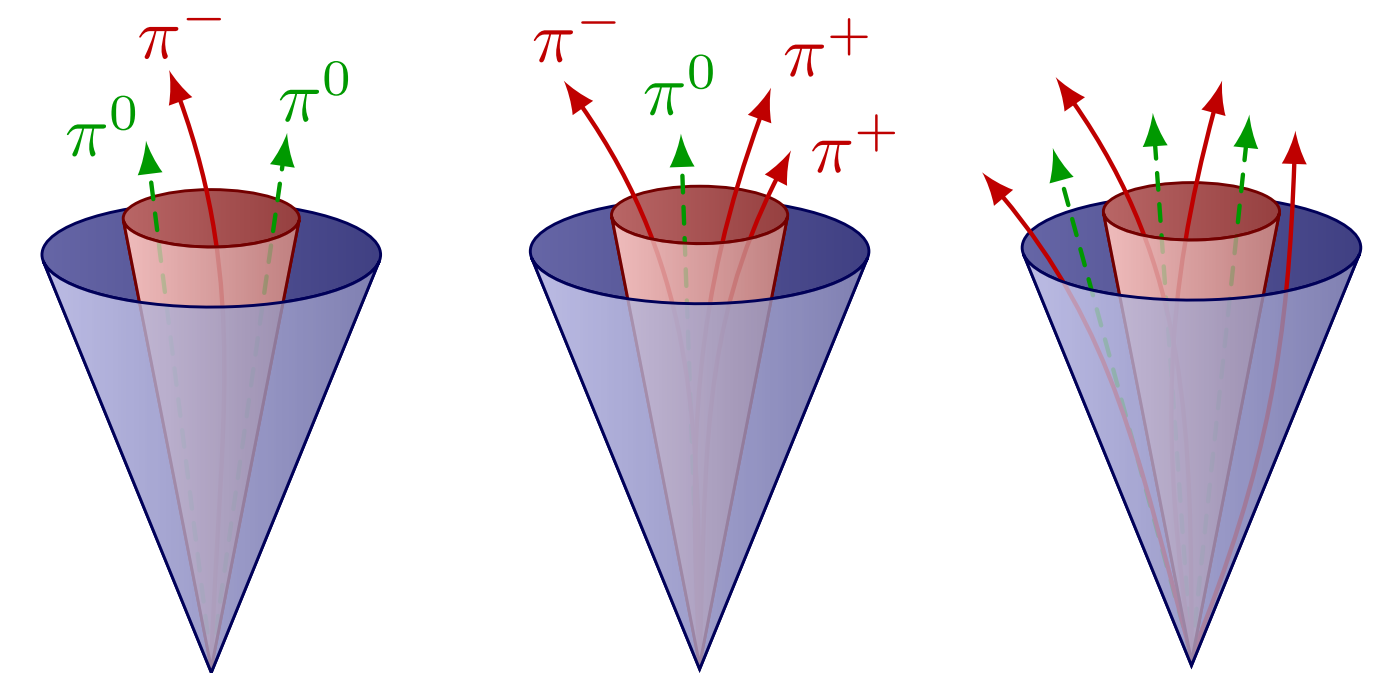
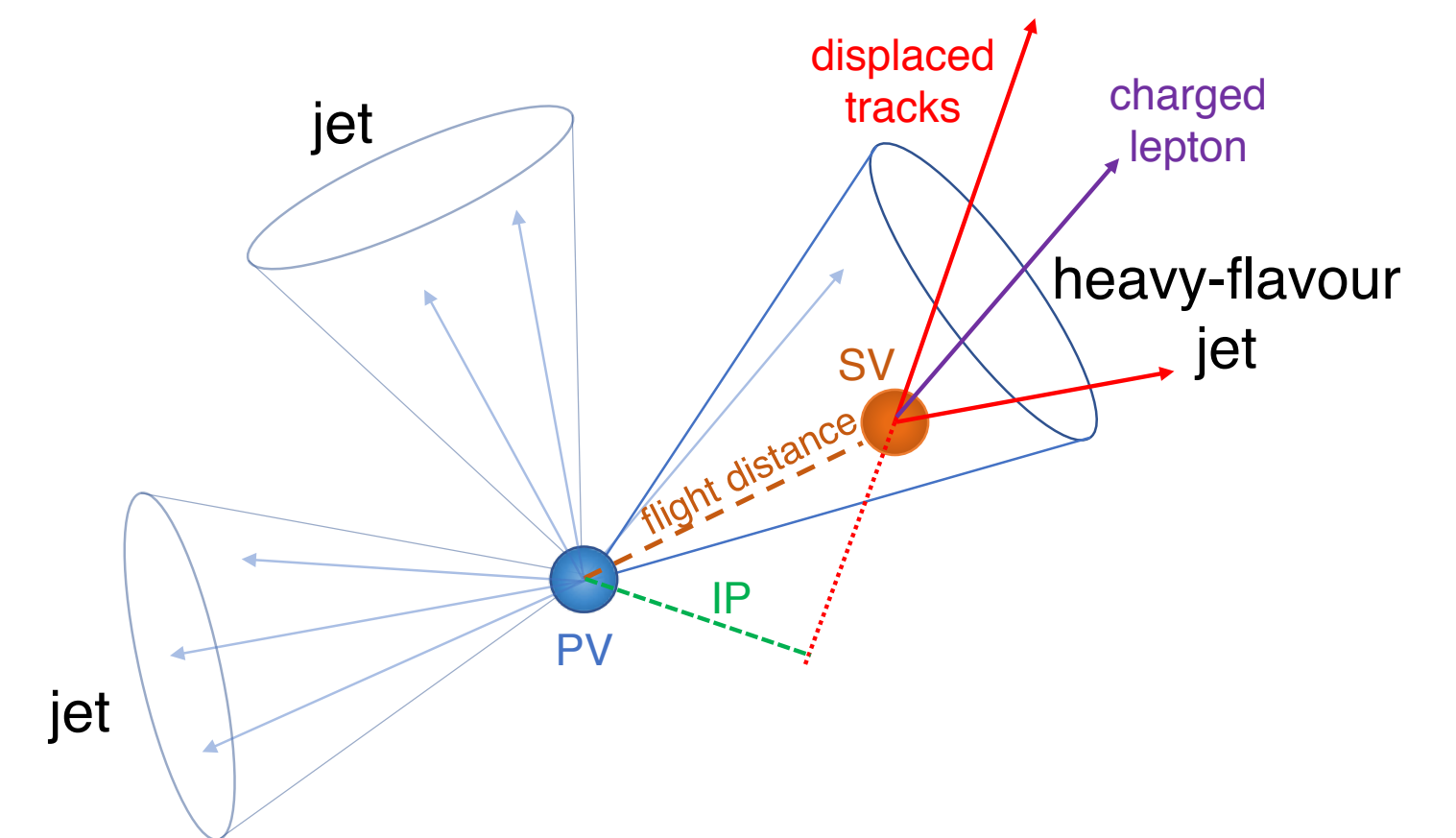
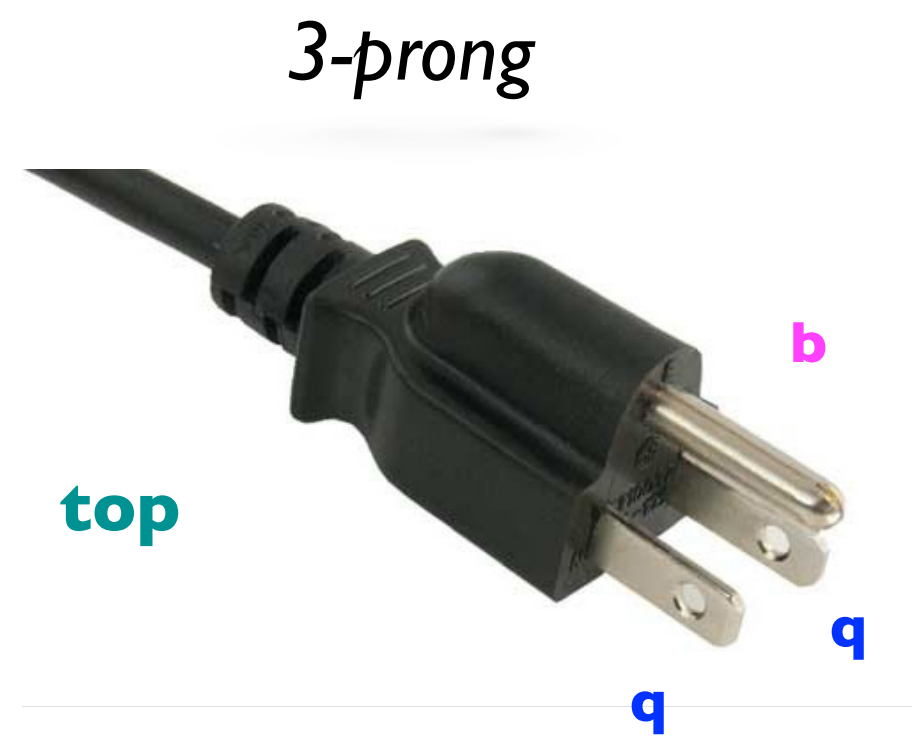


Image credit

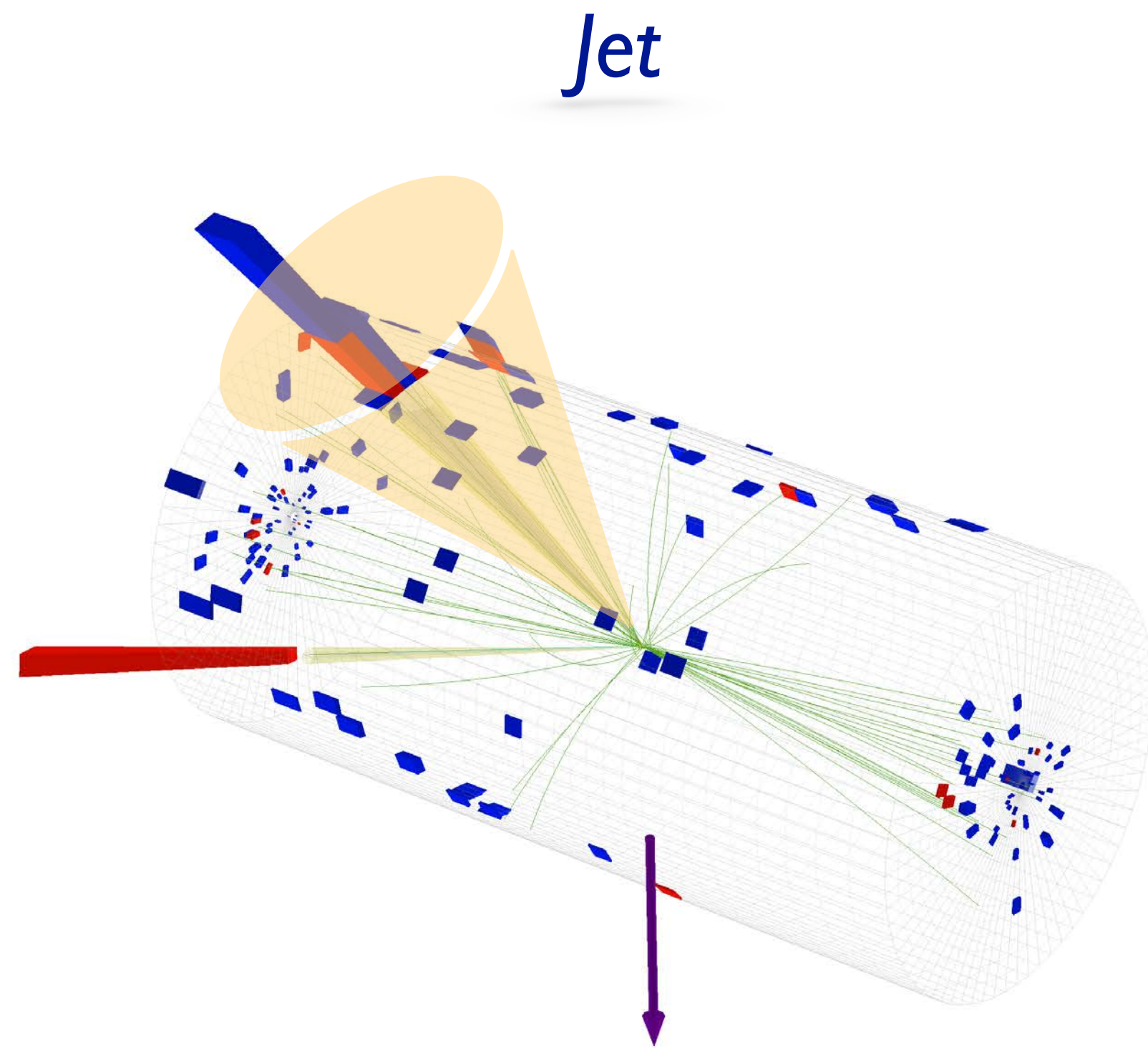
BOOSTED JET TAGGING

- Hadronic decays of highly Lorentz-boosted heavy particles (Higgs/W/Z/top) lead to large-radius jets with distinctive characteristics:
 - different radiation patterns (“**substructure**”)
 - 3-prong (top), 2-prong (W/Z/H) vs 1-prong (gluon/light quark jet)
 - different **flavor** content: existence of one or more b-/c-quarks
- Boosted jet tagging:
 - simultaneously exploiting both **substructure** and **flavor** to maximize the performance
 - significant performance leap thanks to deep learning techniques

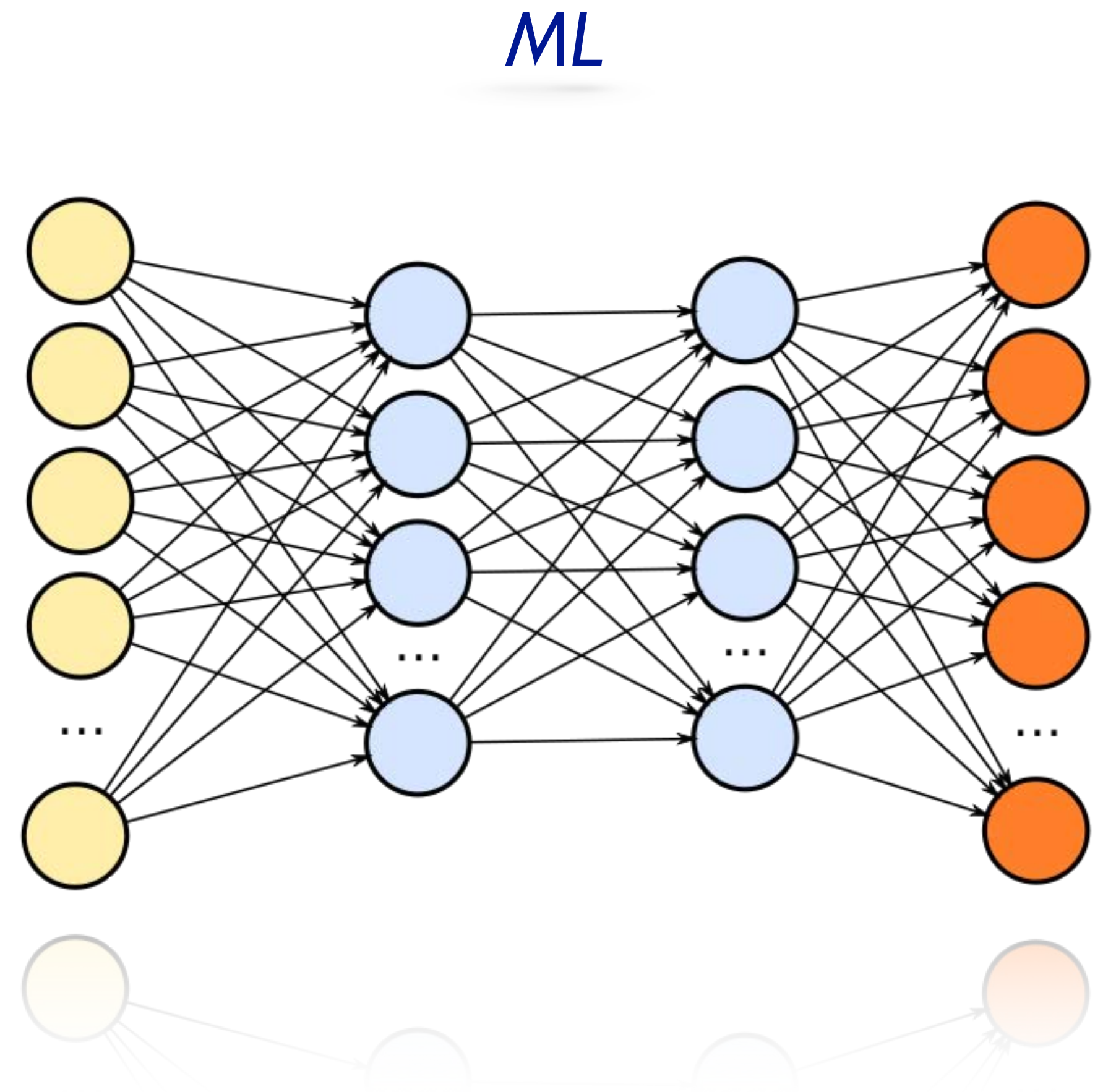


JET REPRESENTATION FOR DEEP LEARNING

JET REPRESENTATION

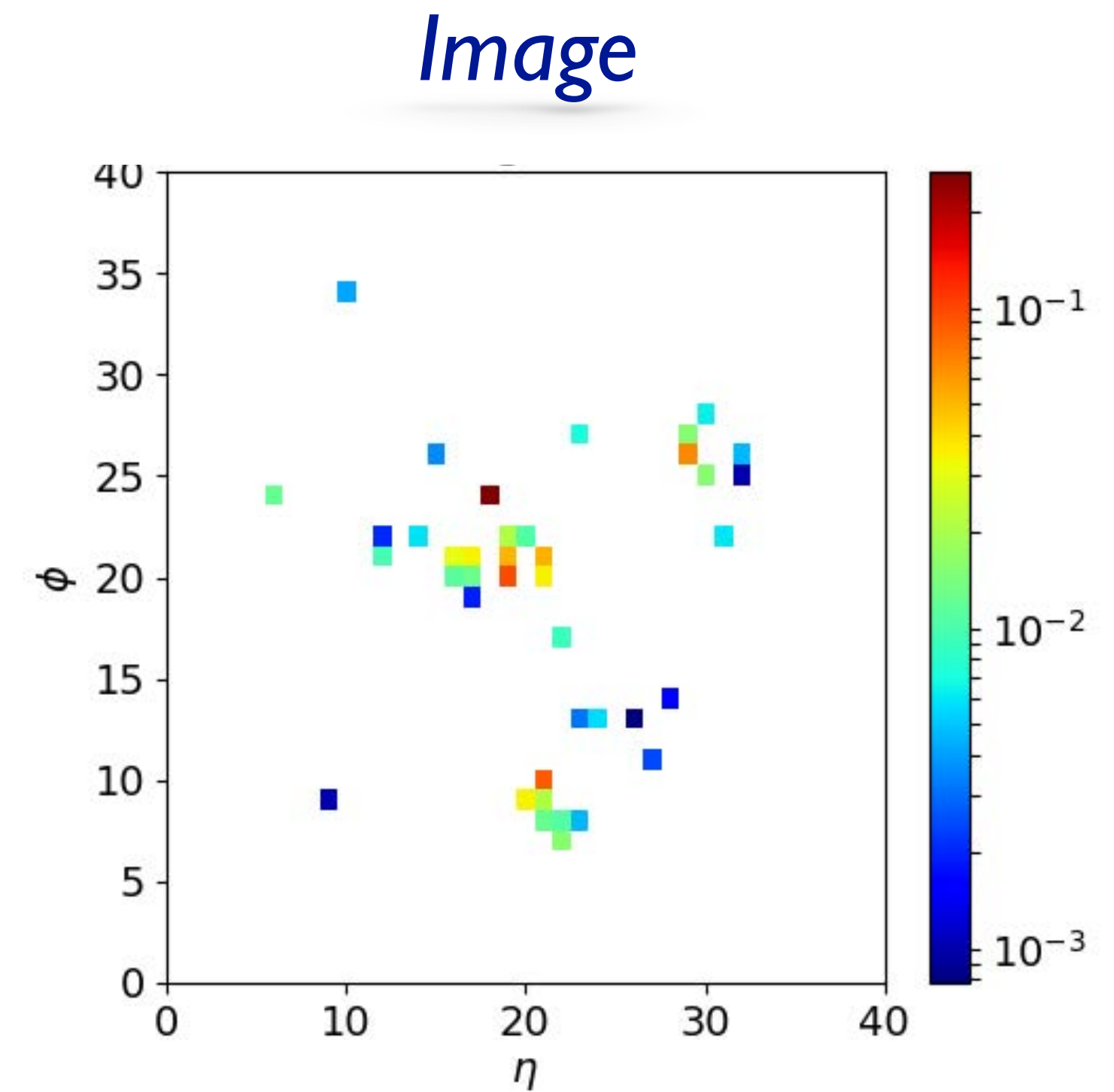
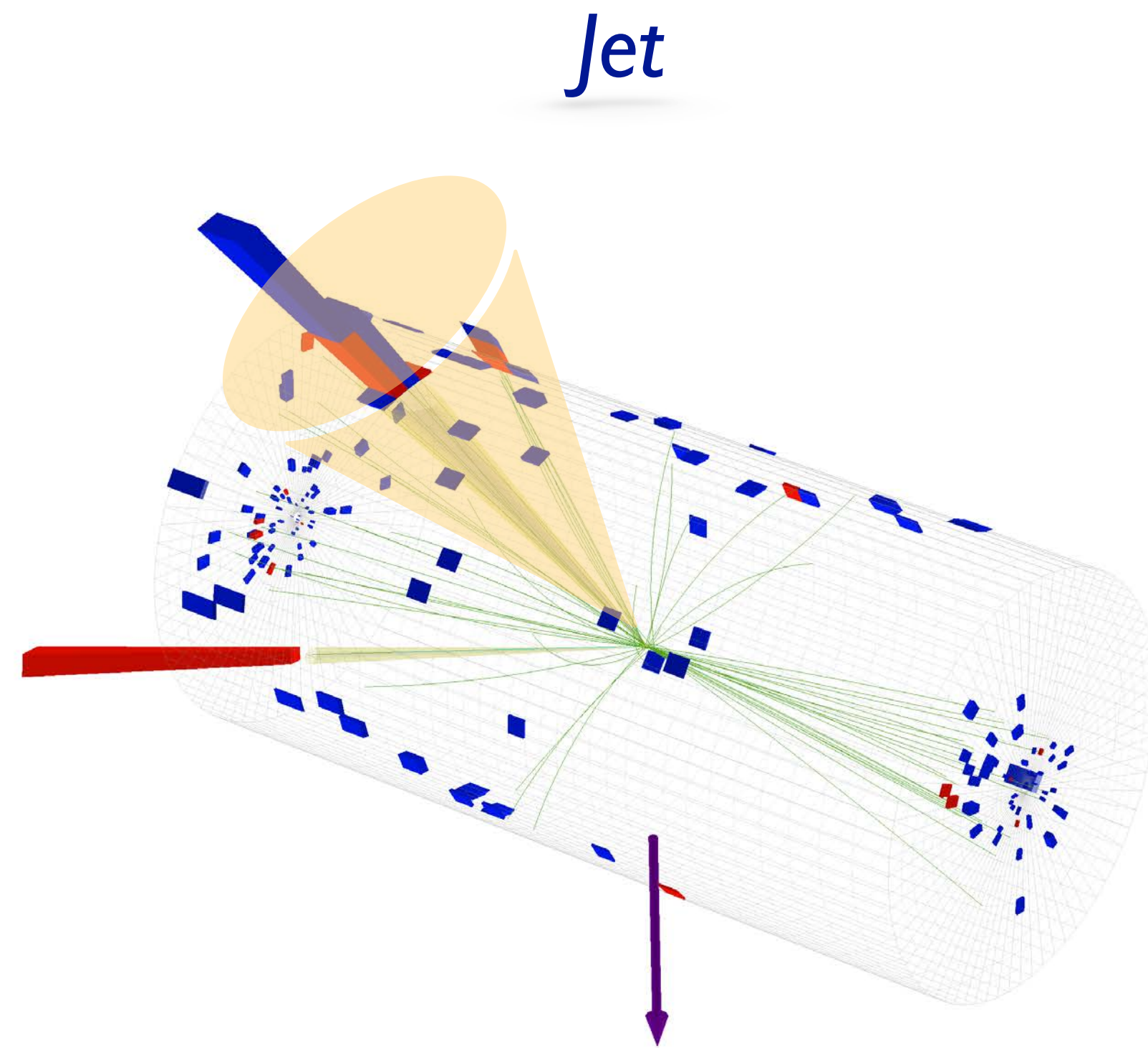


×



First and foremost:
How to represent the data?

JET REPRESENTATION: IMAGE

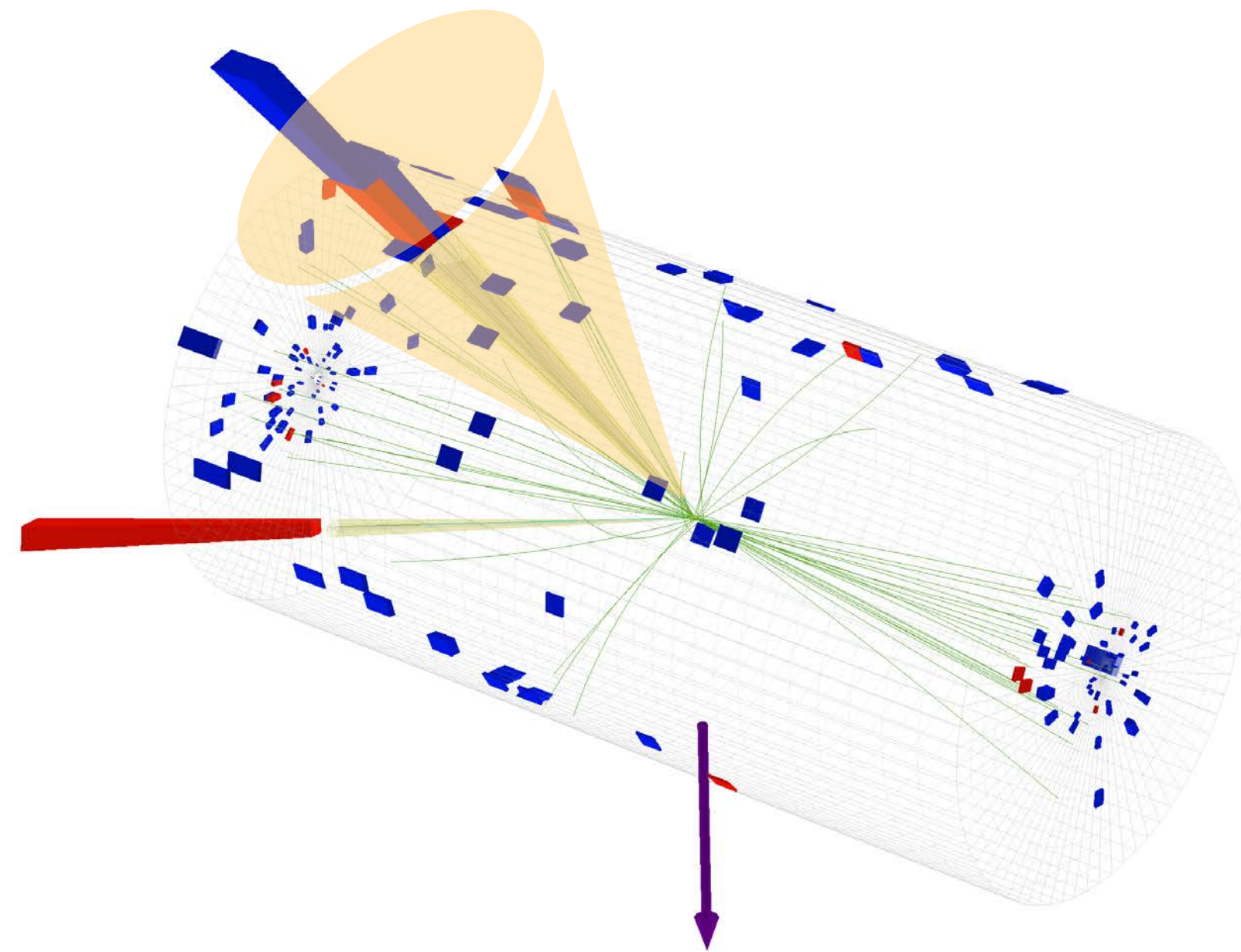


e.g., review in Kagan, arXiv:2012.09719

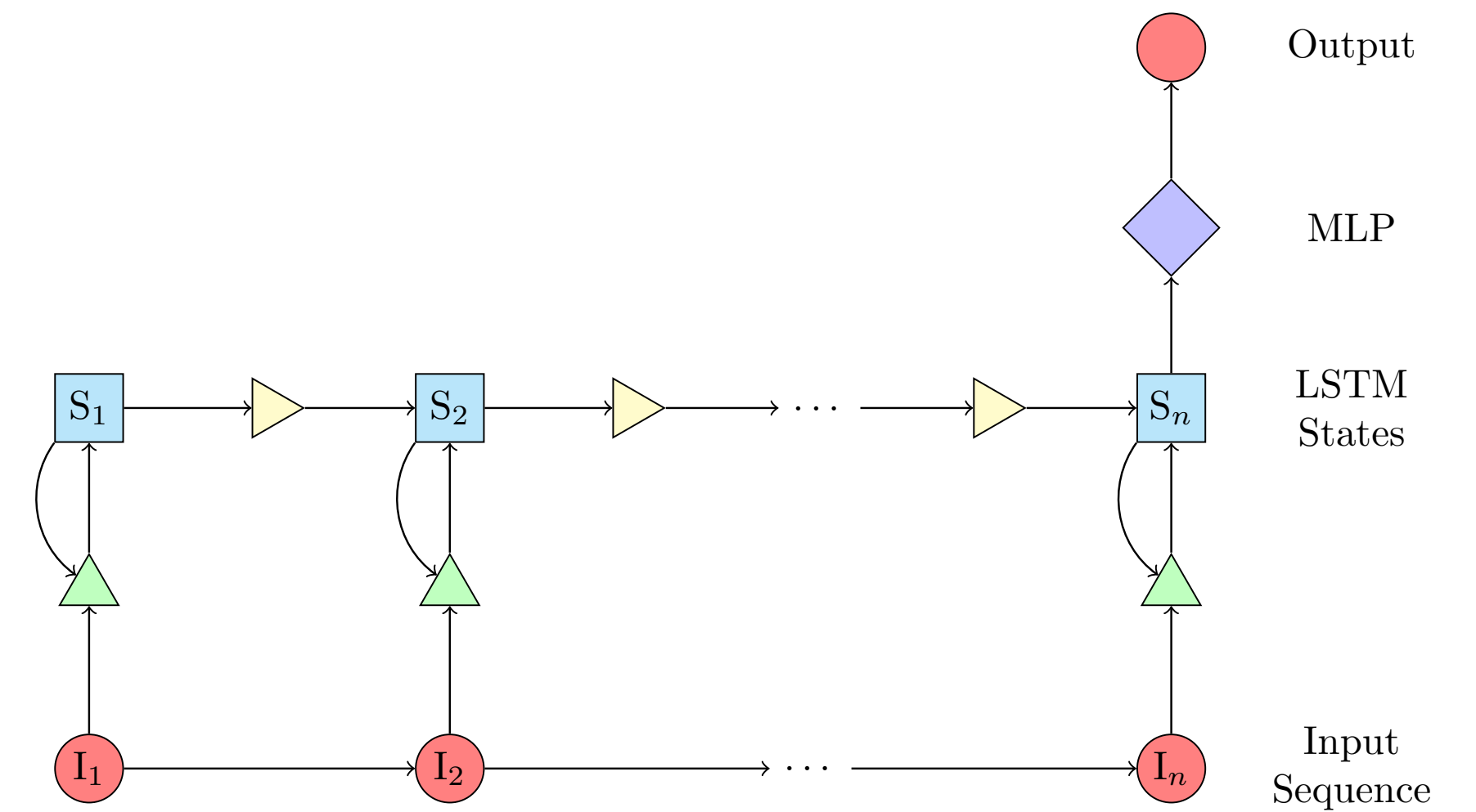
- Convert to 2D/3D image => **Computer vision**
 - then use convolutional neural networks (CNNs)
 - but:
 - inhomogeneous geometry, high sparsity, ...

JET REPRESENTATION: SEQUENCE

Jet



Sequence

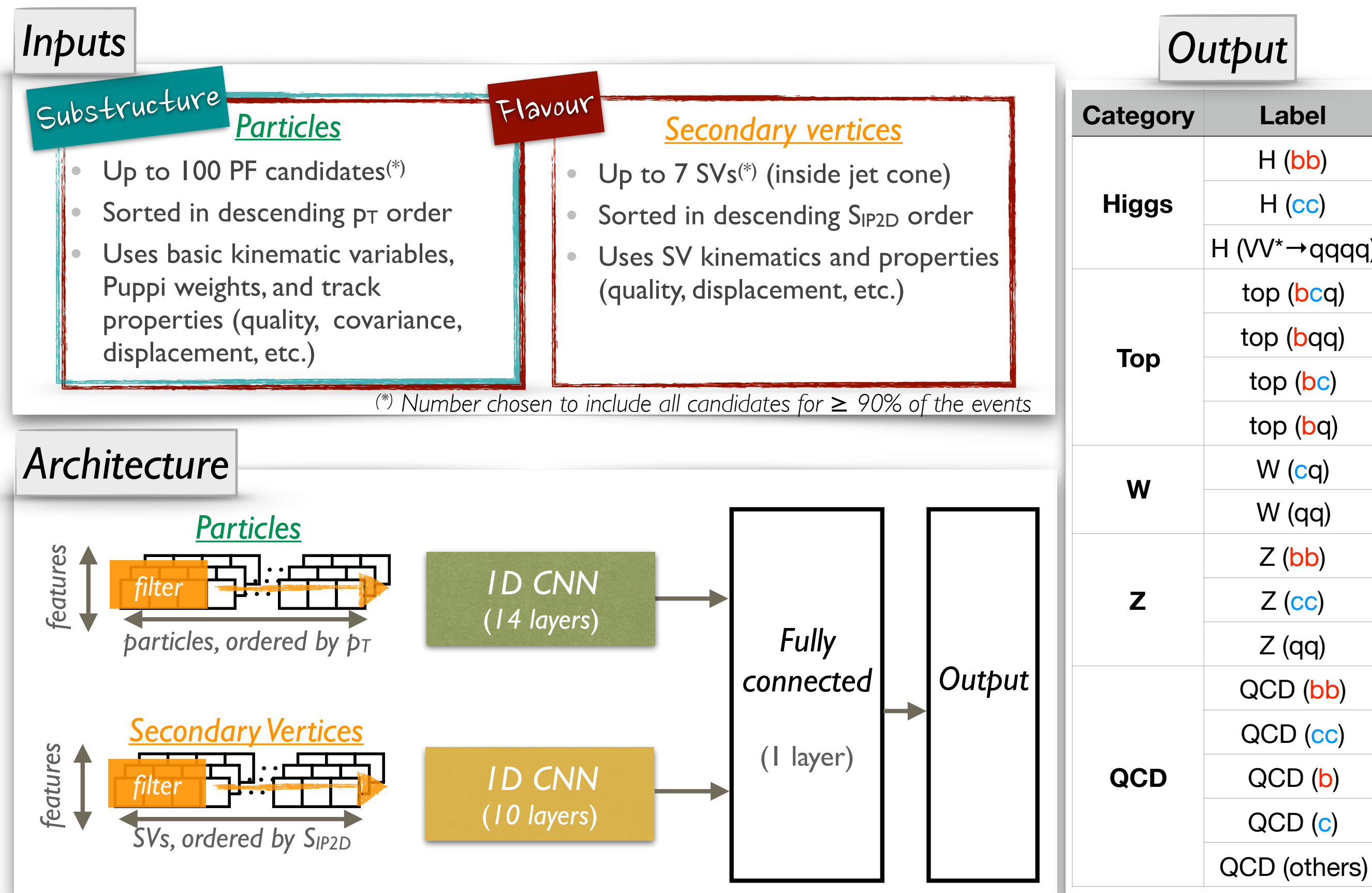


e.g., Guest, Collado, Baldi, Hsu, Urban, Whiteson
arXiv: 1607.08633

- Convert to a sequence => **Natural language processing (NLP)**
 - recurrent neural network (RNN), e.g., GRU/LSTM; 1D CNNs; etc.

DEEPAK8

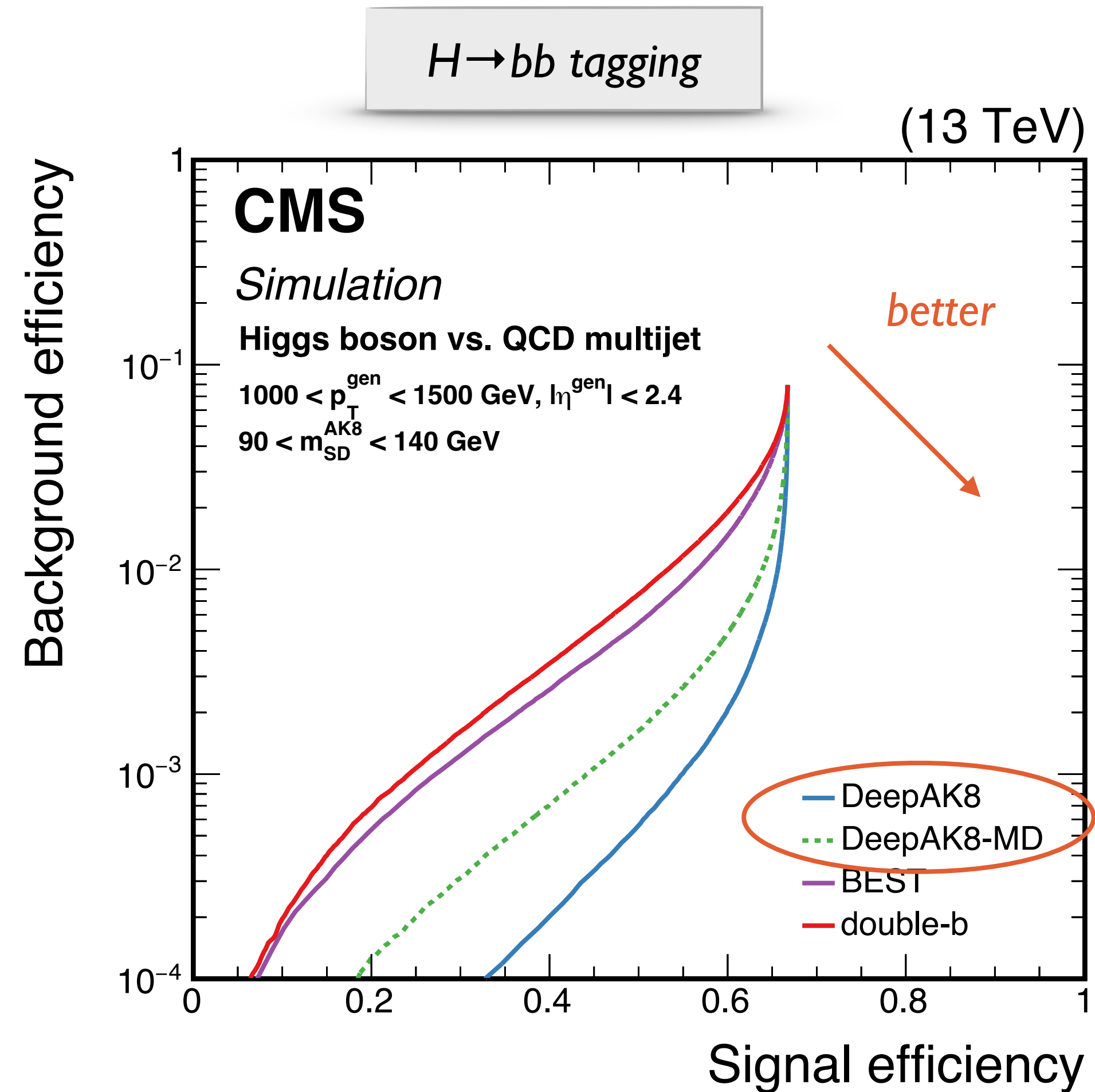
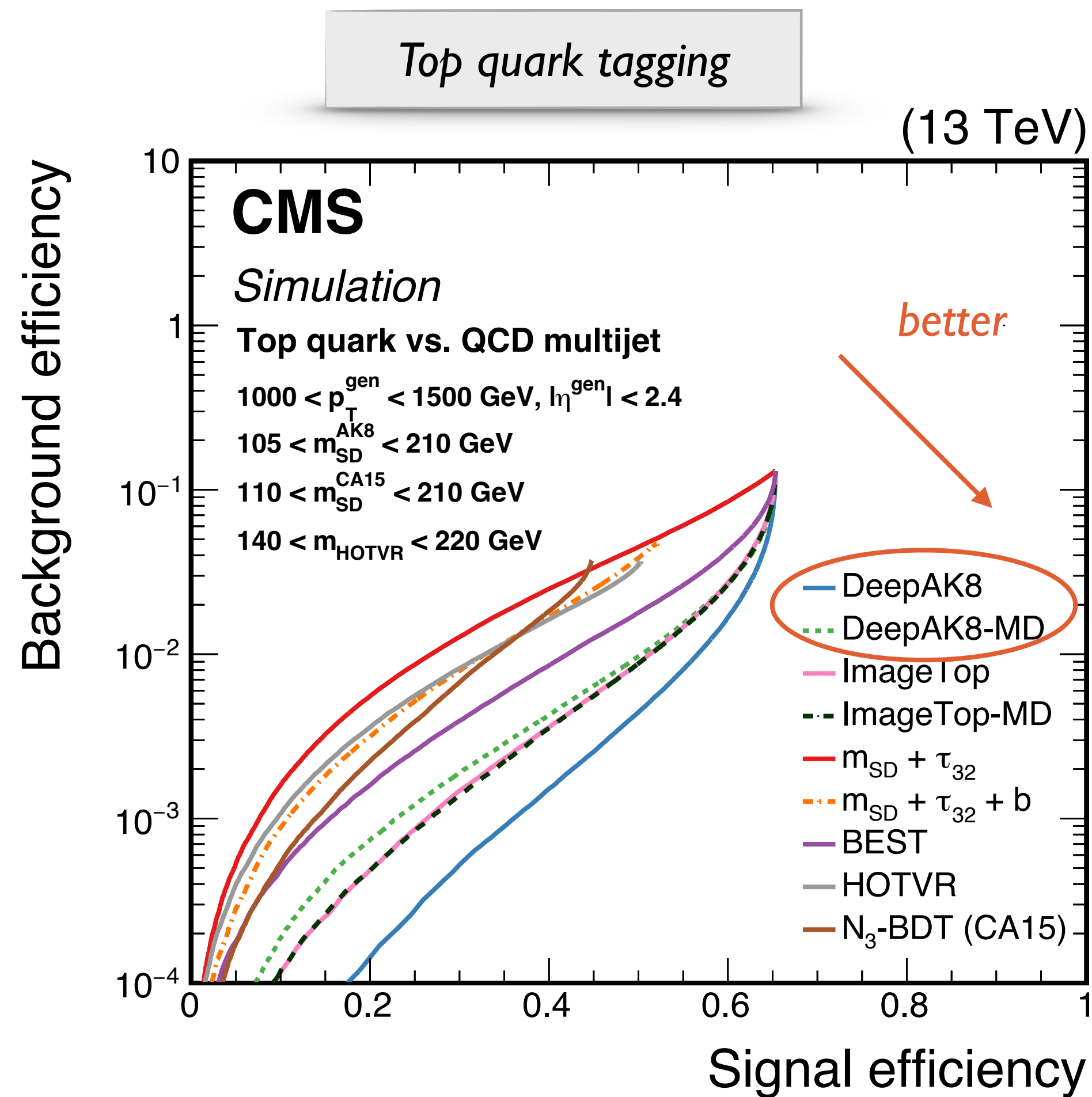
- Advanced deep learning-based algorithm for boosted jet tagging, using AK8 (anti- k_T $R=0.8$) jets
 - multi-class classifier for top quark and W, Z, Higgs boson tagging
 - directly uses jet constituents (particle-flow candidates / secondary vertices)
 - 1D convolutional neural network (CNN), based on the ResNet [arXiv: 1512.03385] architecture



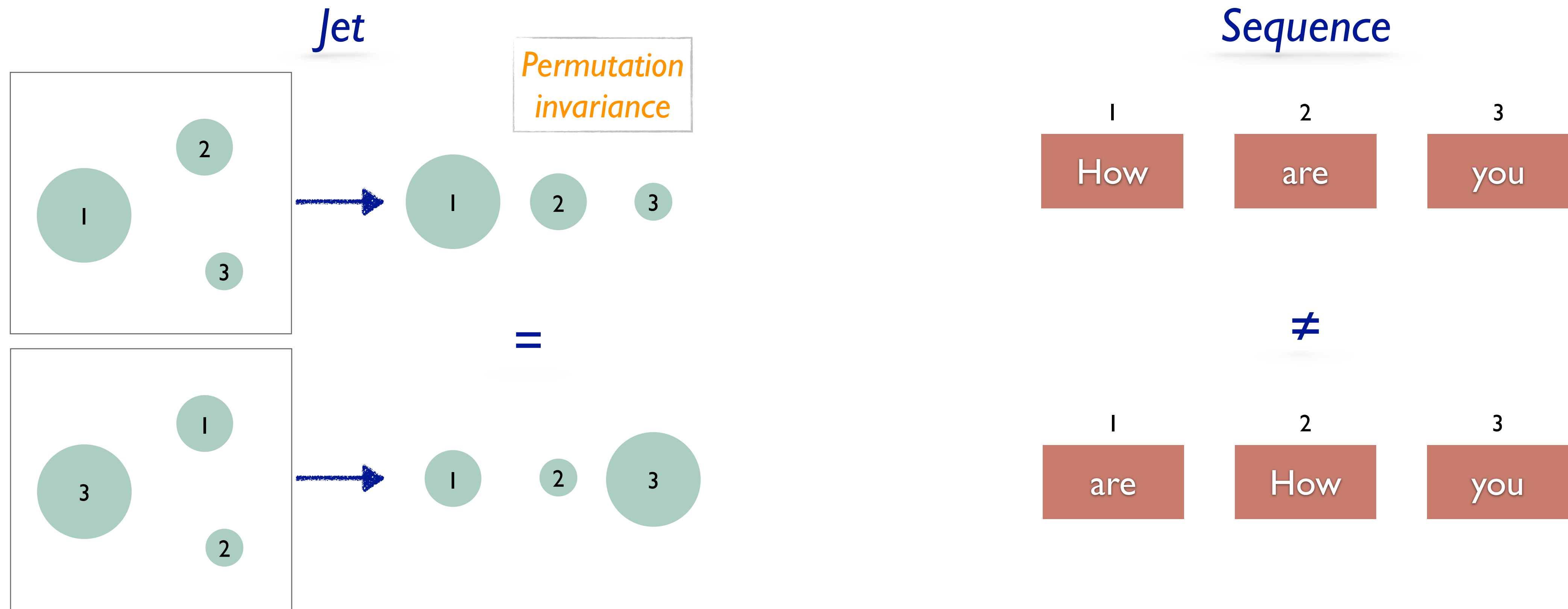
DEEPAK8 PERFORMANCE

CMS, JINST 15 (2020) P06005

- Significant performance improvement compared to traditional approaches



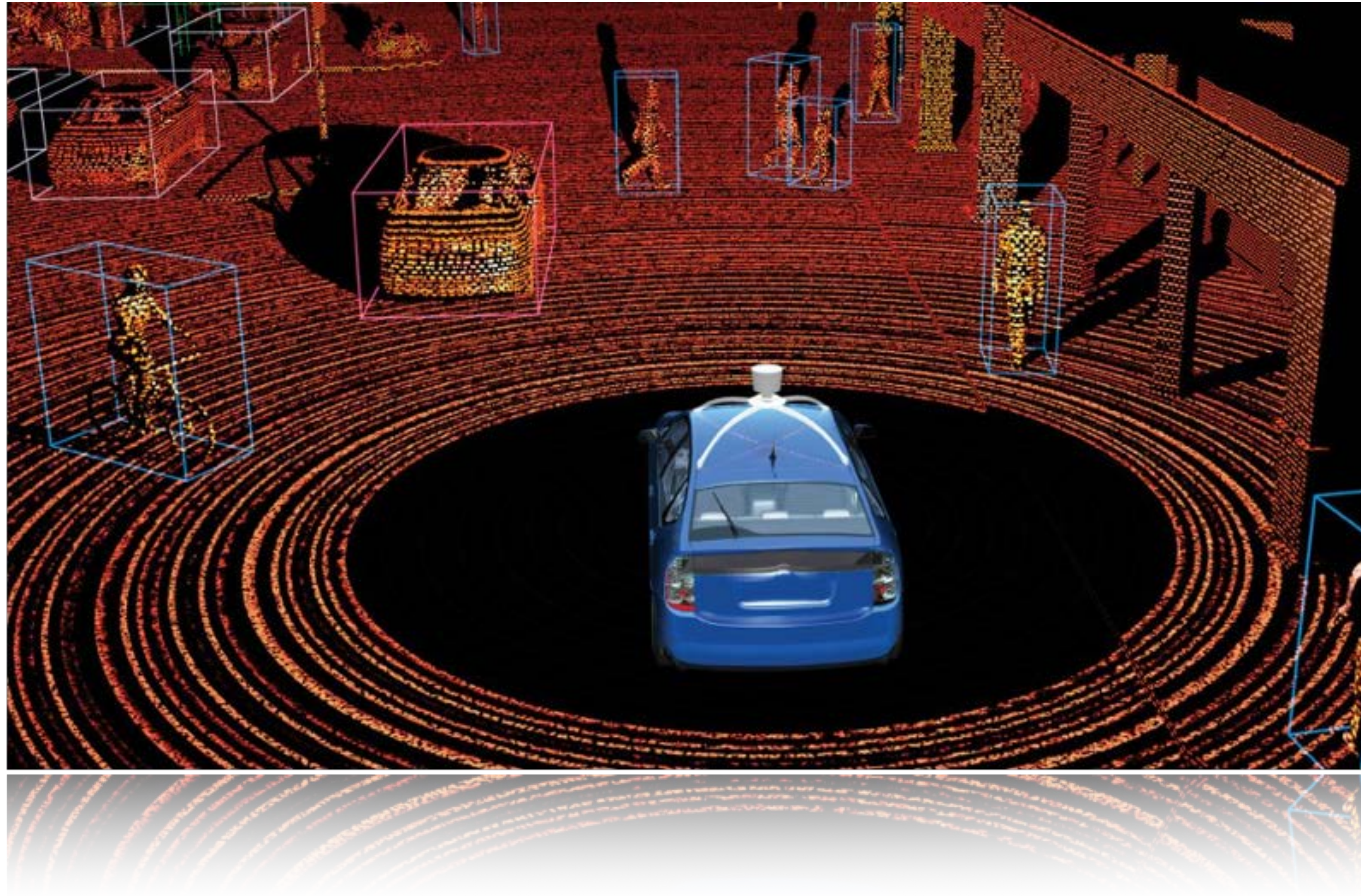
JET REPRESENTATION: SEQUENCE



- Limitations

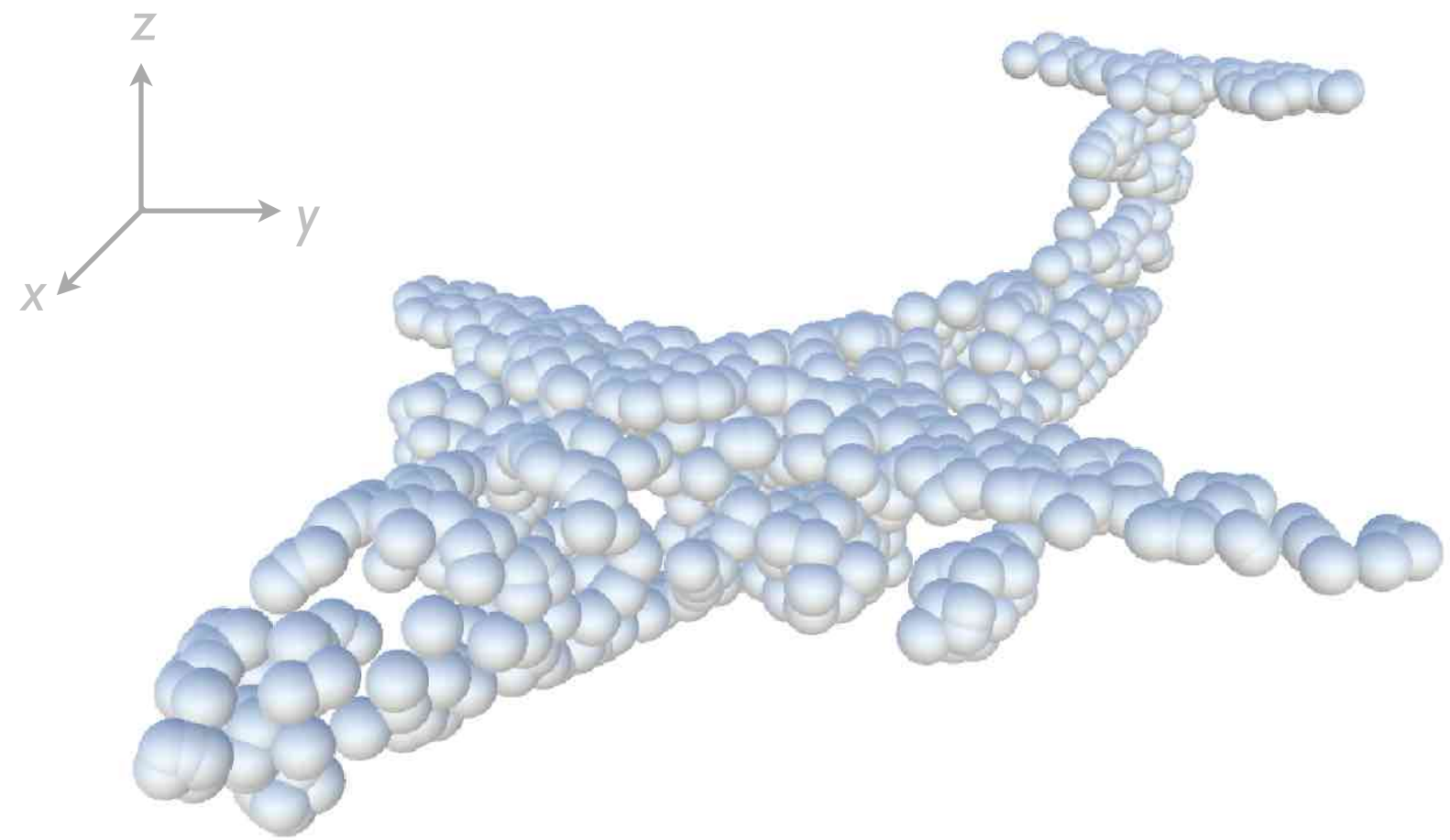
- while words are naturally ordered in natural languages, particles are intrinsically **unordered** in a collision event
 - an ordering has to be *imposed* (p_T , distance, ...), which can limit the learning performance

POINT CLOUD

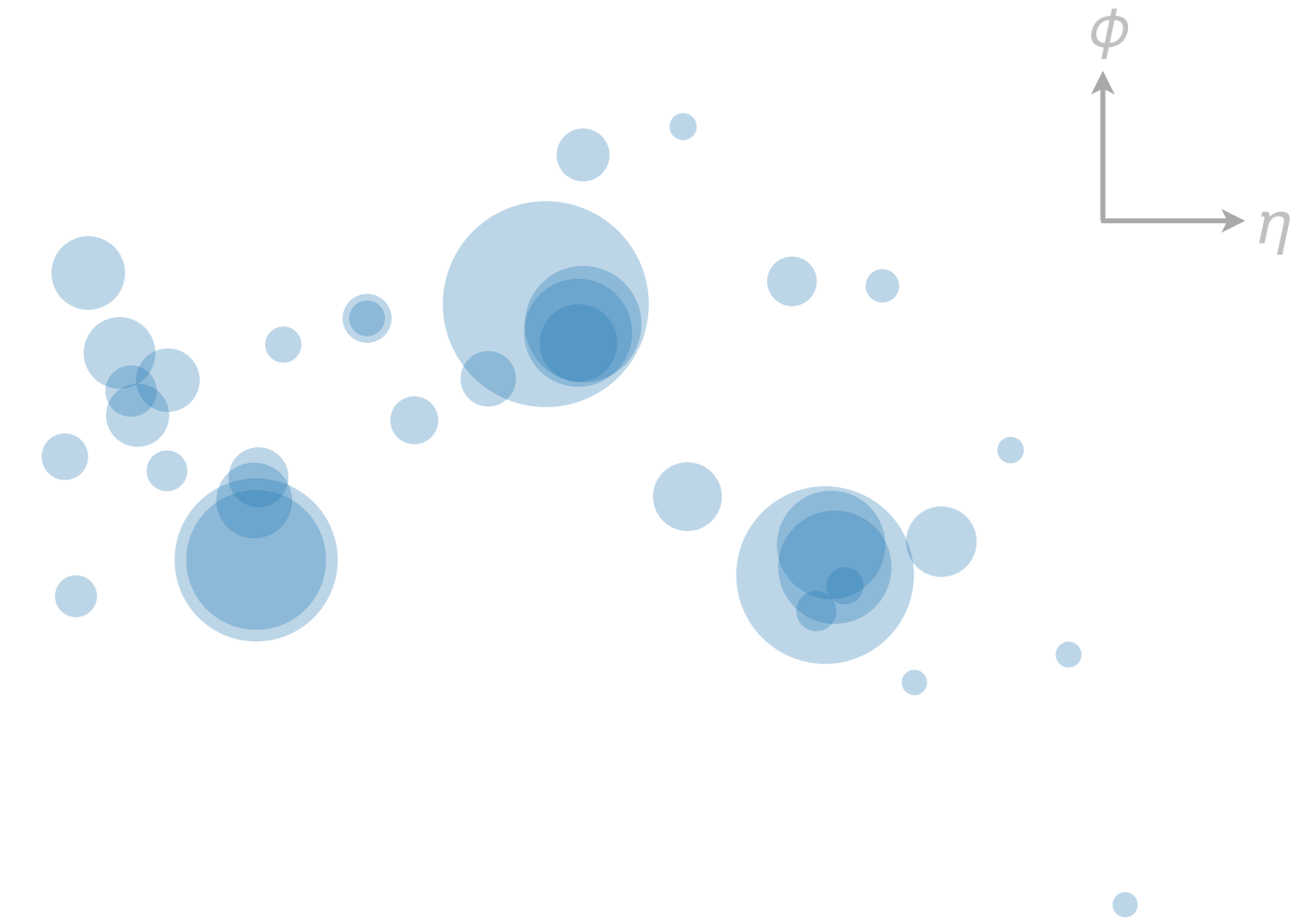


- Point cloud
 - an **unordered** set of points in space
 - typically produced by a LiDAR / 3D scanner
 - spatial distribution of points
 - geometric structure of the objects

JET REPRESENTATION: PARTICLE CLOUD



Permutation invariance





Point cloud

From Wikipedia, the free encyclopedia

A **point cloud** is a set of data points in [space](#). Point clouds are generally produced by [3D scanners](#), which measure a large number of points on the external surfaces of objects around them.

Jet (Particle cloud)

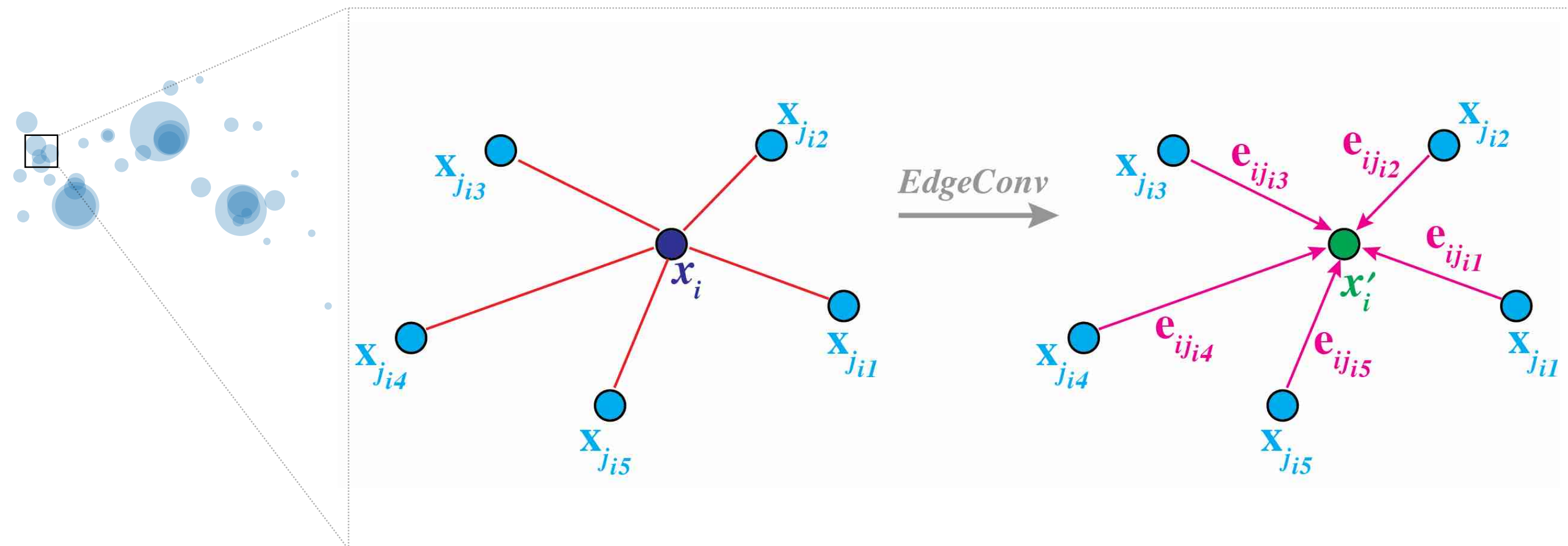
From Wikipedia, the free encyclopedia

A **jet (particle cloud)** is a set of particles in [space](#). Particle clouds are generally created by clustering a large number of particles measured by [particle detectors](#), e.g.,  **ATLAS** and  **CMS**.

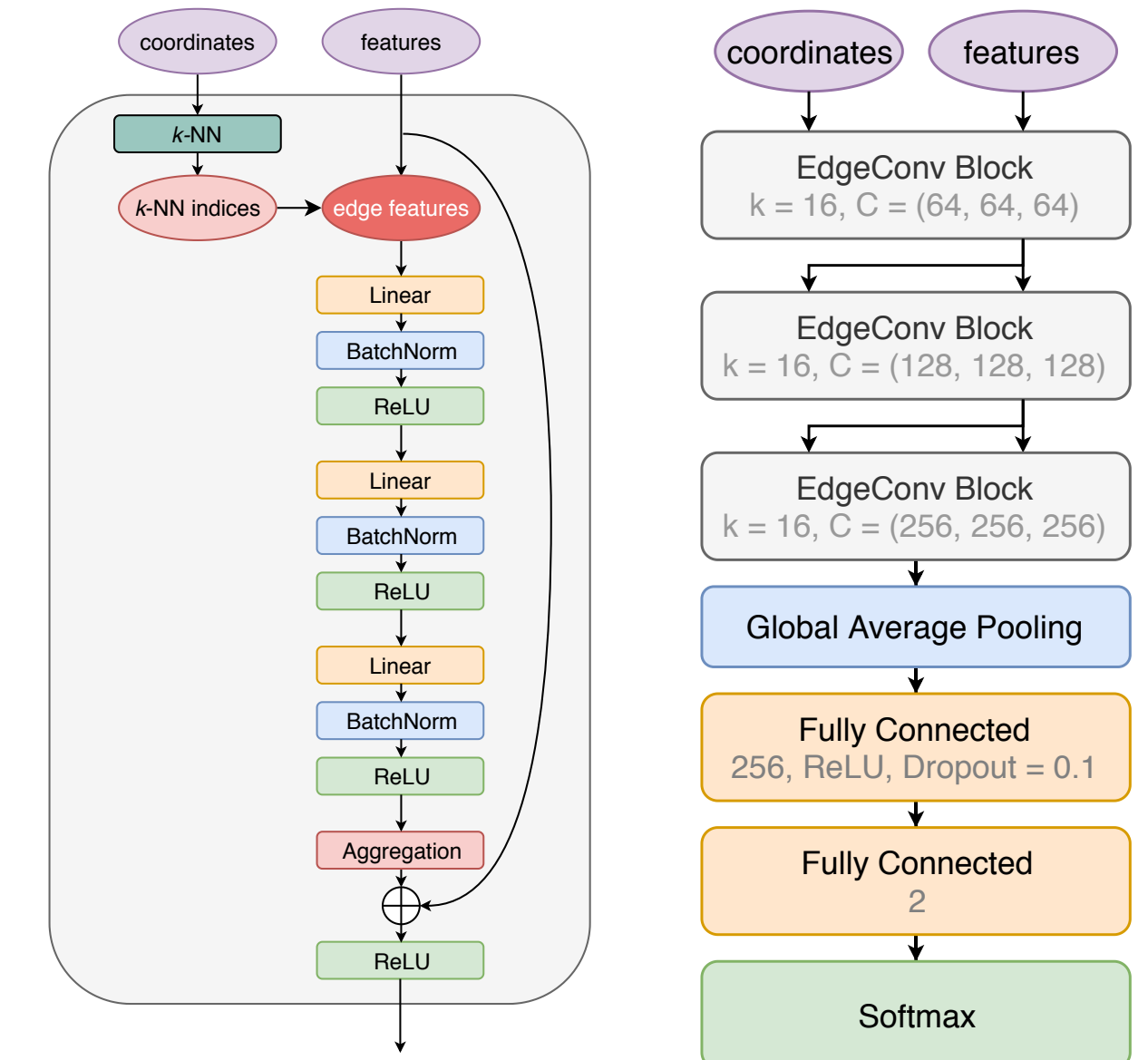
PARTICLENET

H. Qu and L. Gouskos
Phys.Rev.D 101 (2020) 5, 056019

- ParticleNet: jet tagging via particle clouds
 - treating a jet as an **unordered set of particles**, distributed in the $\eta - \phi$ space
 - **graph neural network architecture**, adapted from Dynamic Graph CNN [arXiv:1801.07829]
 - treating a point cloud as a graph: each point is a vertex
 - for each point, a local patch is defined by finding its k-nearest neighbors
 - designing a permutation-invariant “convolution” function
 - define “edge feature” for each center-neighbor pair: $e_{ij} = h_{\theta}(x_i, x_j)$
 - aggregate the edge features in a symmetric way: $x'_i = \text{mean}_j e_{ij}$



ParticleNet architecture



cf. P.T. Komiske, E. M. Metodiev and J. Thaler, *JHEP 01 (2019) 121*;
 V. Mikuni and F. Canelli, *Eur. Phys. J. Plus 135, 463 (2020)*; *Mach.Learn.Sci.Tech. 2 (2021) 3, 035027*.

PARTICLENET: PERFORMANCE

G. Kasieczka et al.
SciPost Phys. 7 (2019) 014

- Top performance among a variety of deep learning taggers on the community-wide top tagging benchmark

	AUC	Acc	$1/\epsilon_B$ ($\epsilon_S = 0.3$)			#Param
			single	mean	median	
CNN [16]	0.981	0.930	914±14	995±15	975±18	610k
ResNeXt [30]	0.984	0.936	1122±47	1270±28	1286±31	1.46M
TopoDNN [18]	0.972	0.916	295±5	382±5	378±8	59k
Multi-body N -subjettiness 6 [24]	0.979	0.922	792±18	798±12	808±13	57k
Multi-body N -subjettiness 8 [24]	0.981	0.929	867±15	918±20	926±18	58k
TreeNiN [43]	0.982	0.933	1025±11	1202±23	1188±24	34k
P-CNN	0.980	0.930	732±24	845±13	834±14	348k
ParticleNet [47] (Preliminary ver.)	0.985	0.938	1298±46	1412±45	1393±41	498k
LBN [19]	0.981	0.931	836±17	859±67	966±20	705k
LoLa [22]	0.980	0.929	722±17	768±11	765±11	127k
Energy Flow Polynomials [21]	0.980	0.932	384			1k
Energy Flow Network [23]	0.979	0.927	633±31	729±13	726±11	82k
Particle Flow Network [23]	0.982	0.932	891±18	1063±21	1052±29	82k
GoaT	0.985	0.939	1368±140		1549±208	35k
<i>ParticleNet-Lite</i>	0.984	0.937	1262±49			26k
ParticleNet	0.986	0.940	1615±93			366k

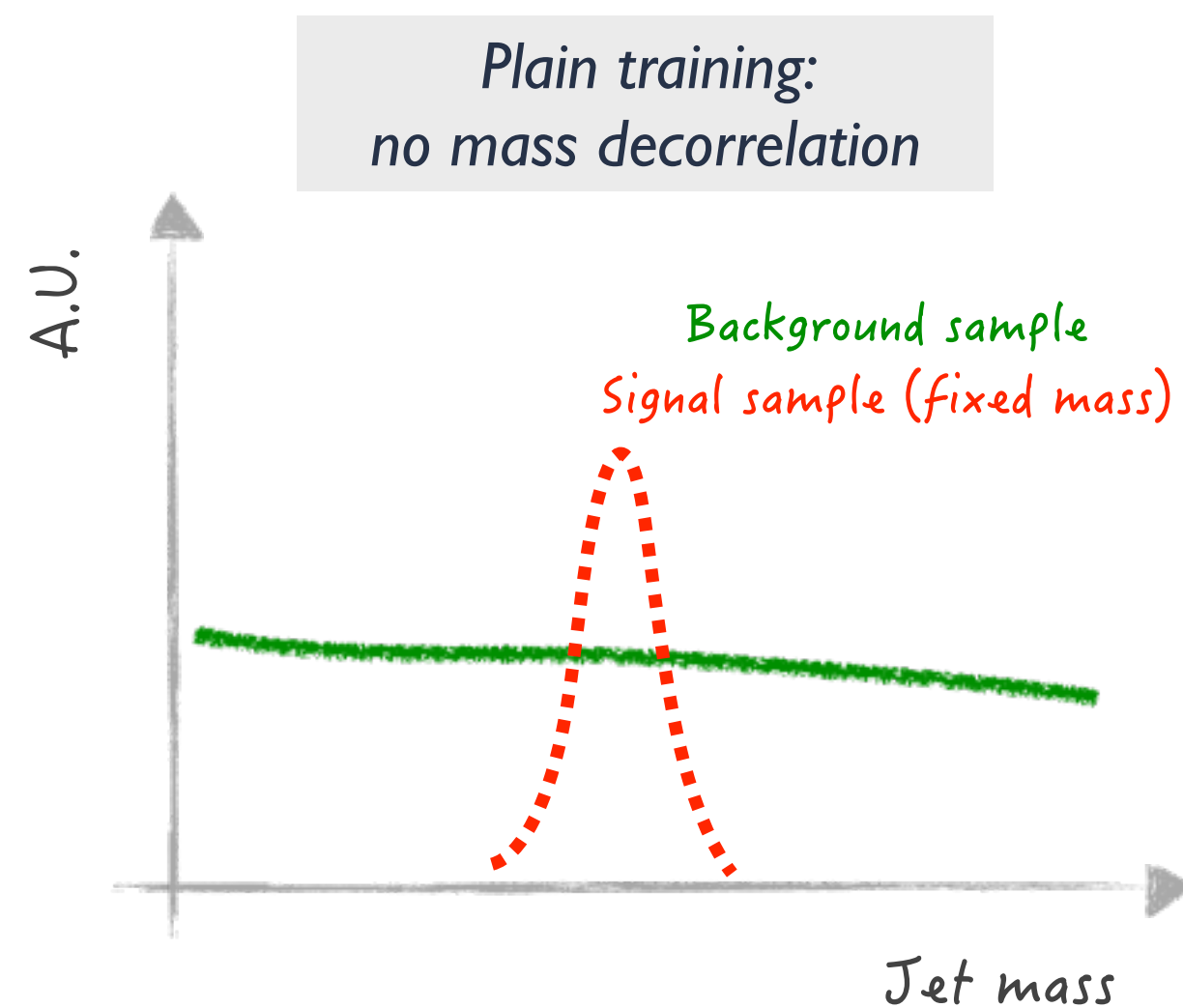
Architecture used by DeepAK8

Ensemble of all taggers

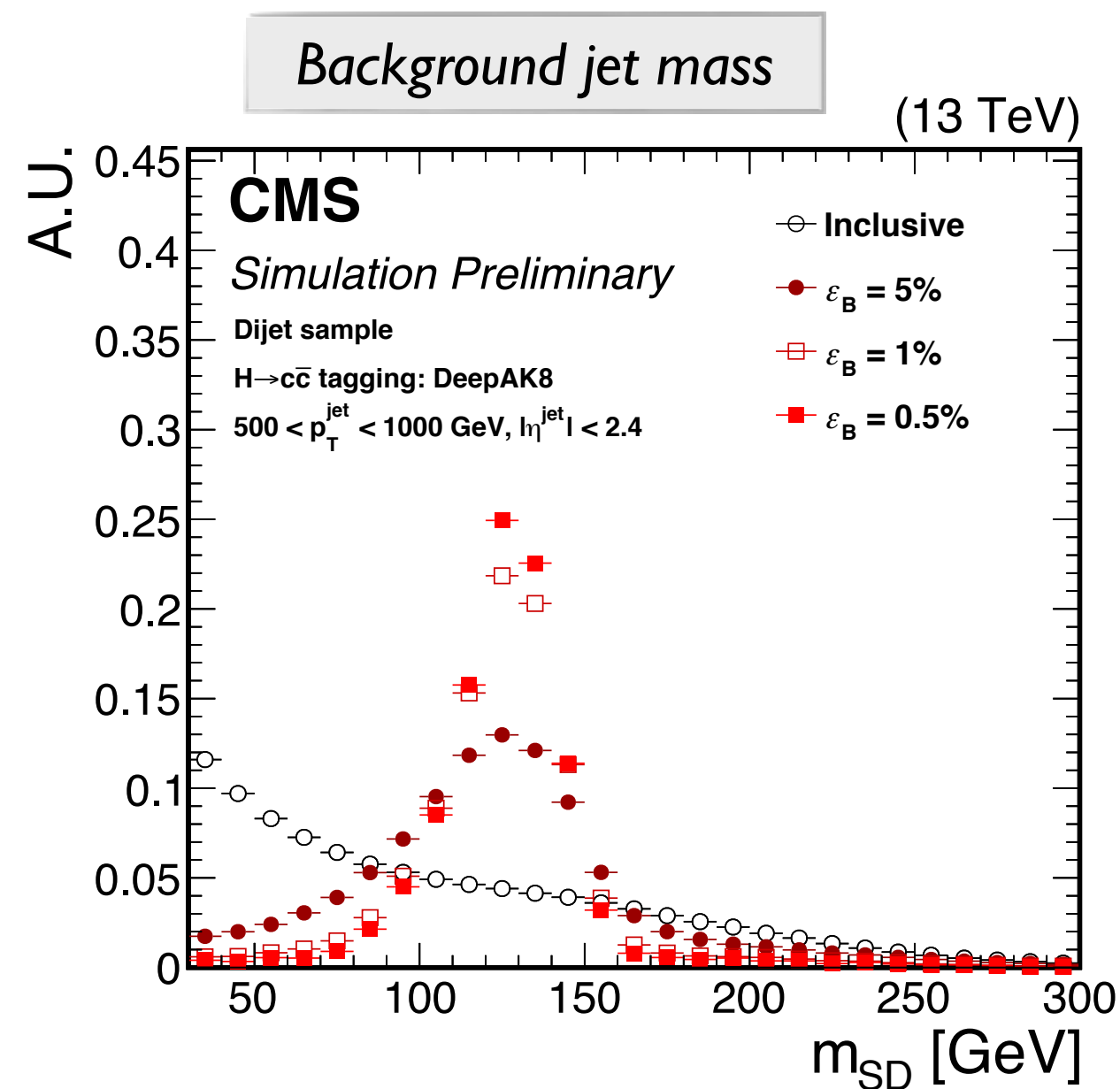
BOOSTED JET TAGGING IN ACTION

CORRELATION WITH THE JET MASS

CMS DP-2020/002



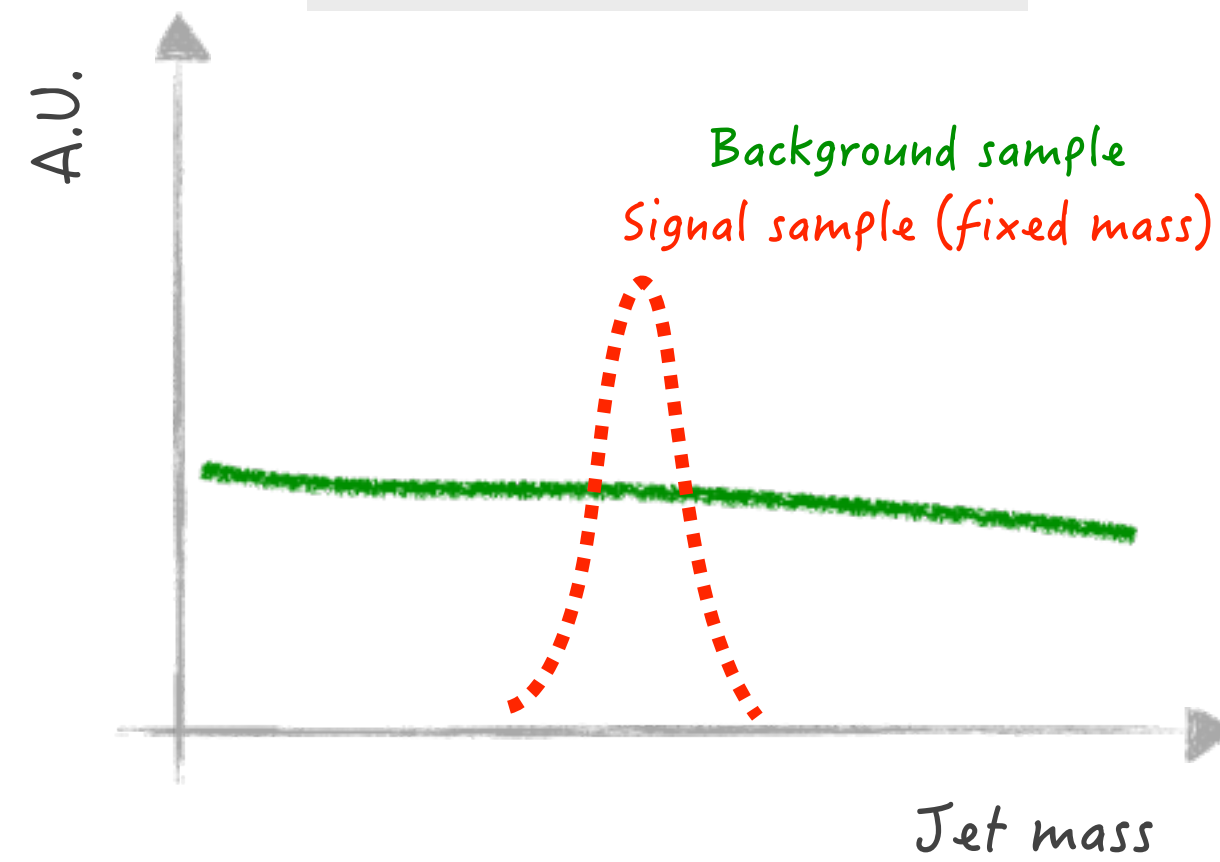
- One feature of these taggers is the correlation with the jet mass
 - jet mass shape of the background becomes similar to that of the signal after selection with the tagger: “**mass sculpting**”
 - not necessarily a problem, but a mass-independent tagger is often more desirable:
 - allows to use the mass variable to further separate signal and background
 - enables tagging signal jets with an unknown mass
 - ...



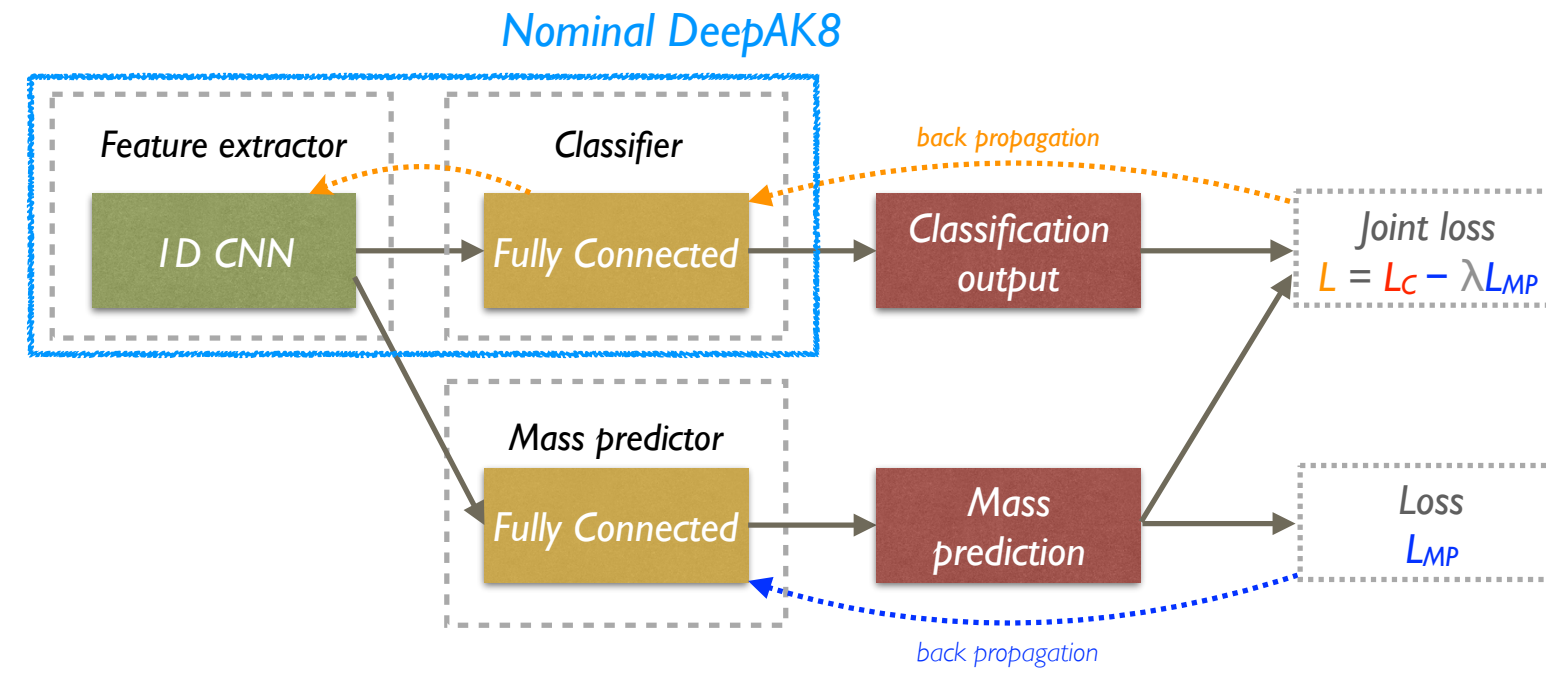
DECORRELATION WITH THE JET MASS

CMS DP-2020/002

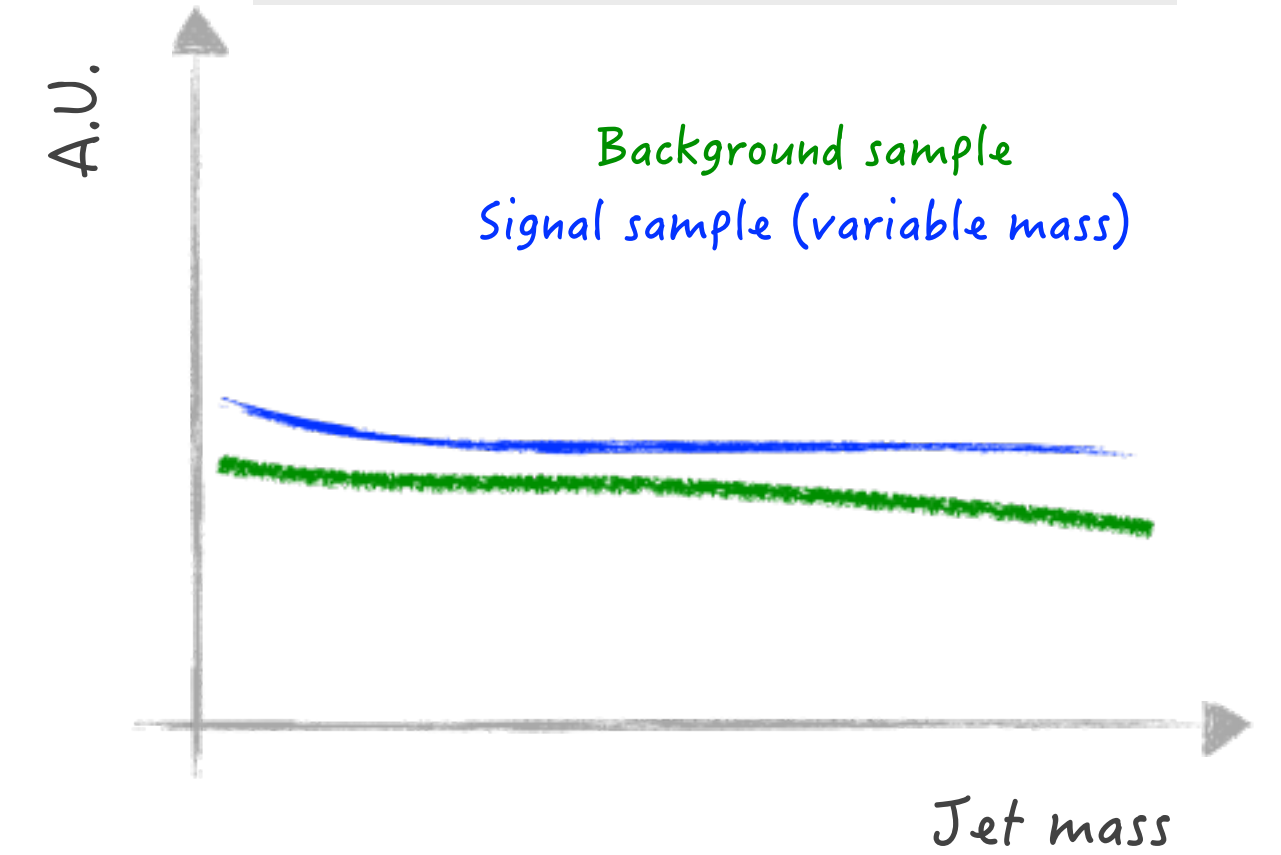
Plain training:
no mass decorrelation



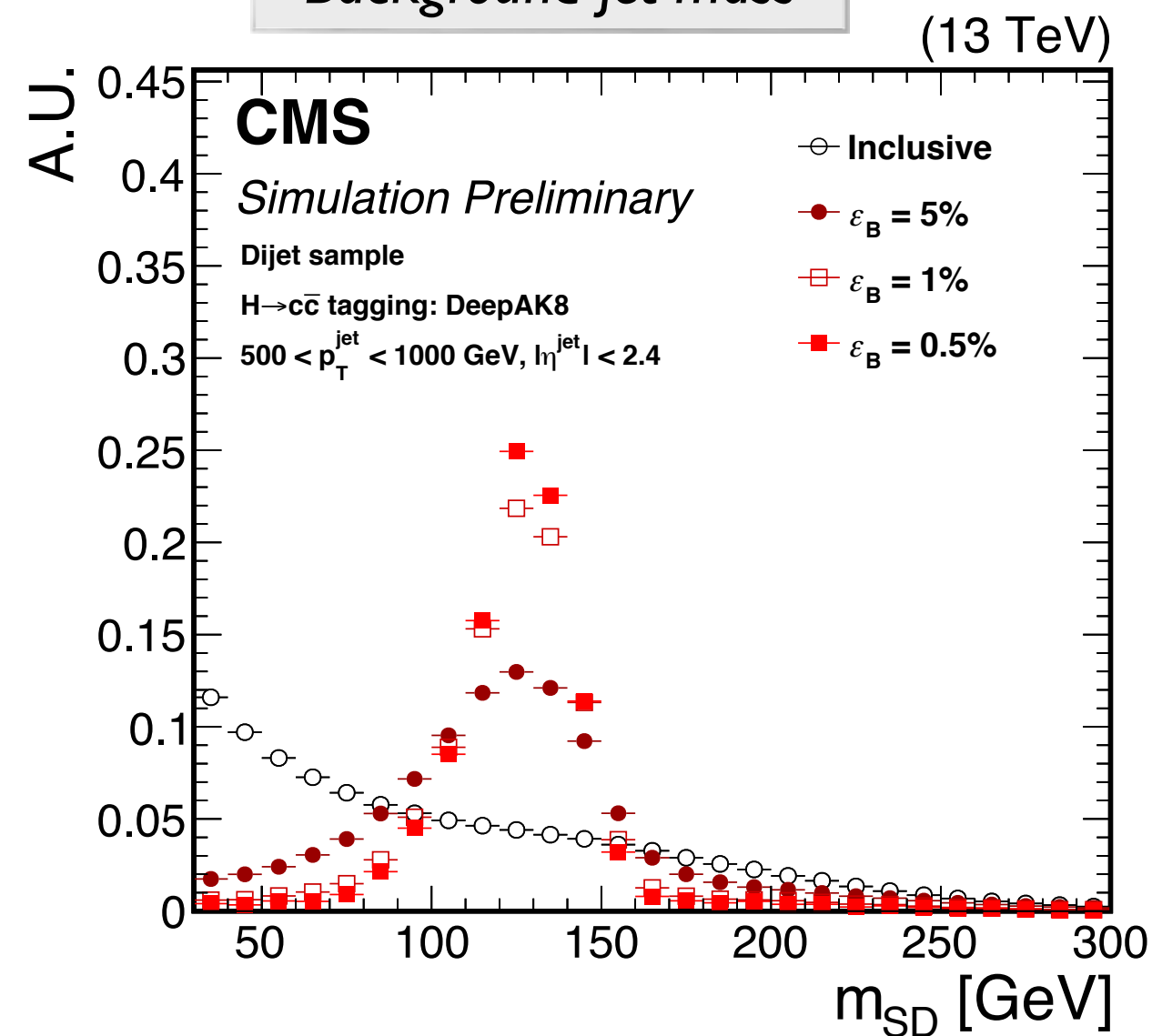
Mass-decorrelated DeepAK8:
“adversarial training”



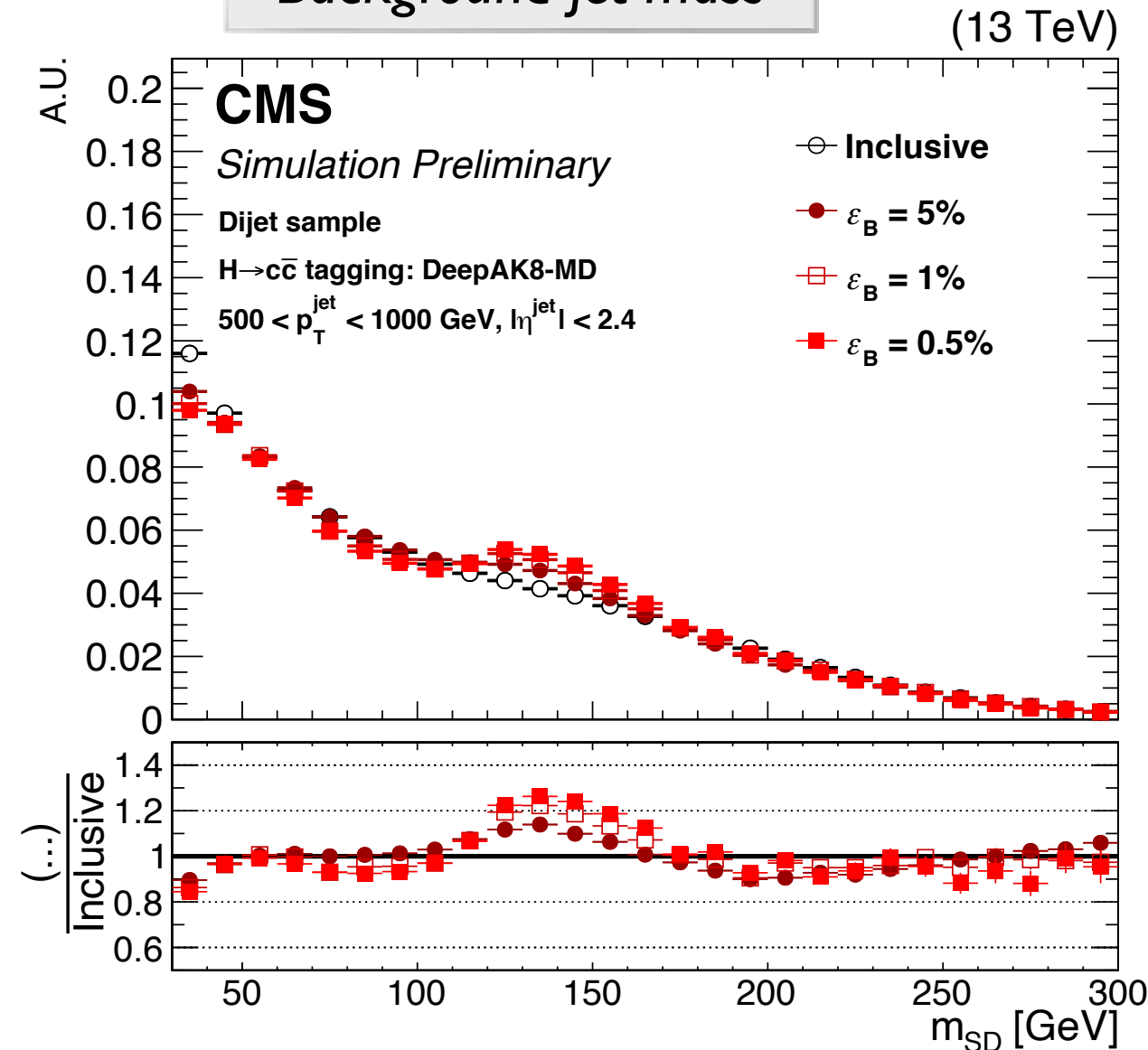
Mass-decorrelated ParticleNet:
training with variable-mass signal



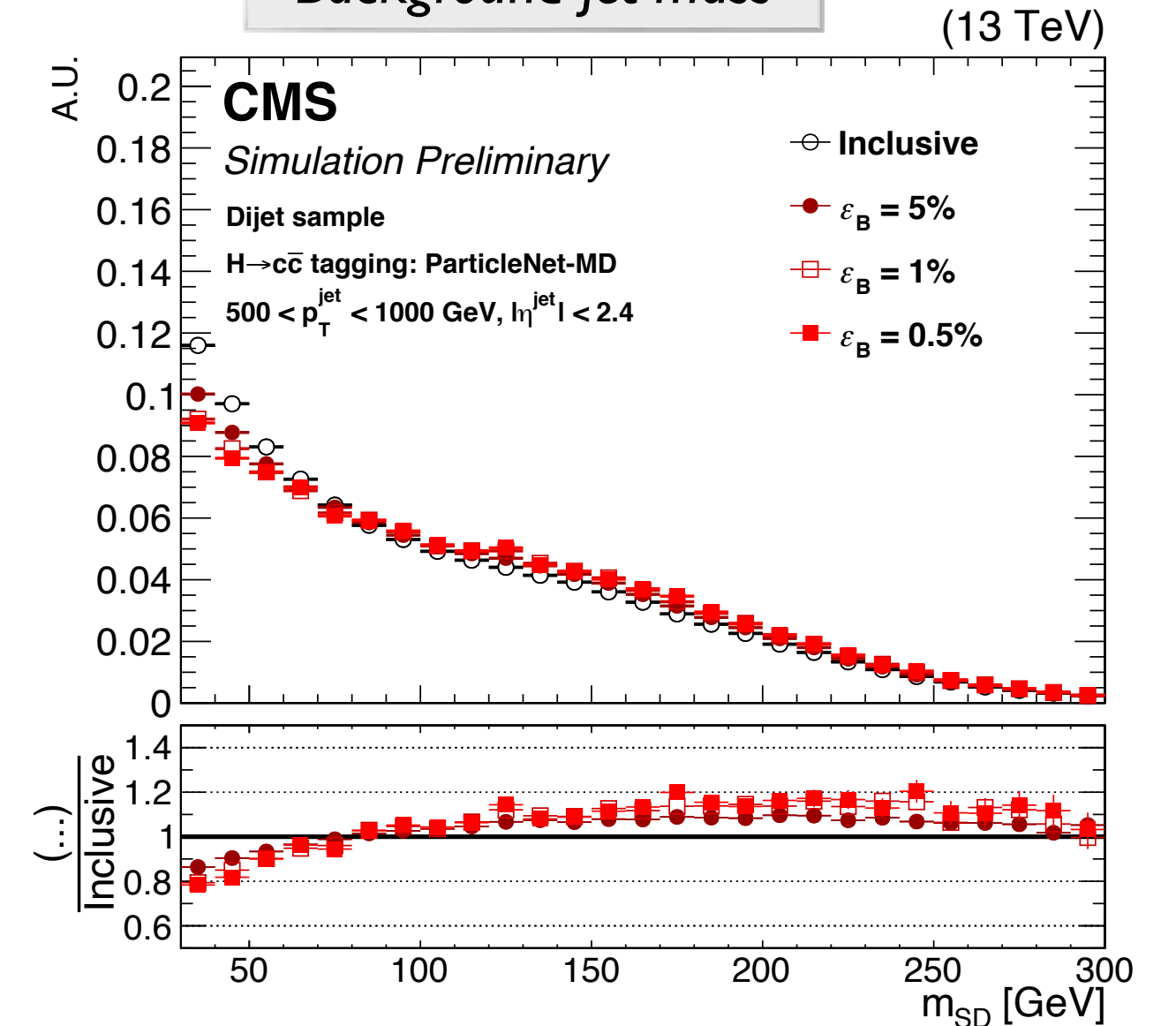
Background jet mass



Background jet mass

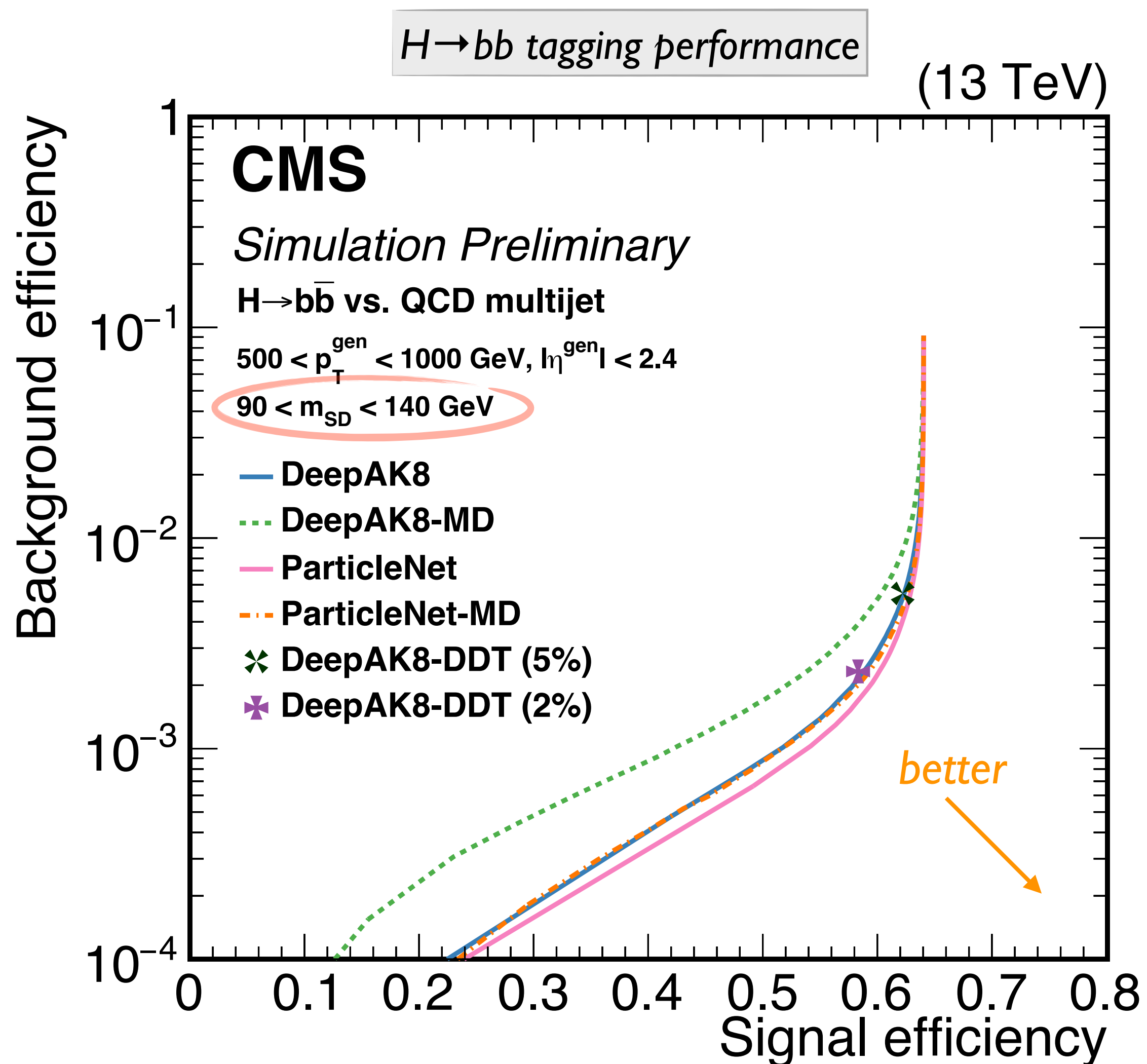


Background jet mass



PERFORMANCE COMPARISON

CMS DP-2020/002



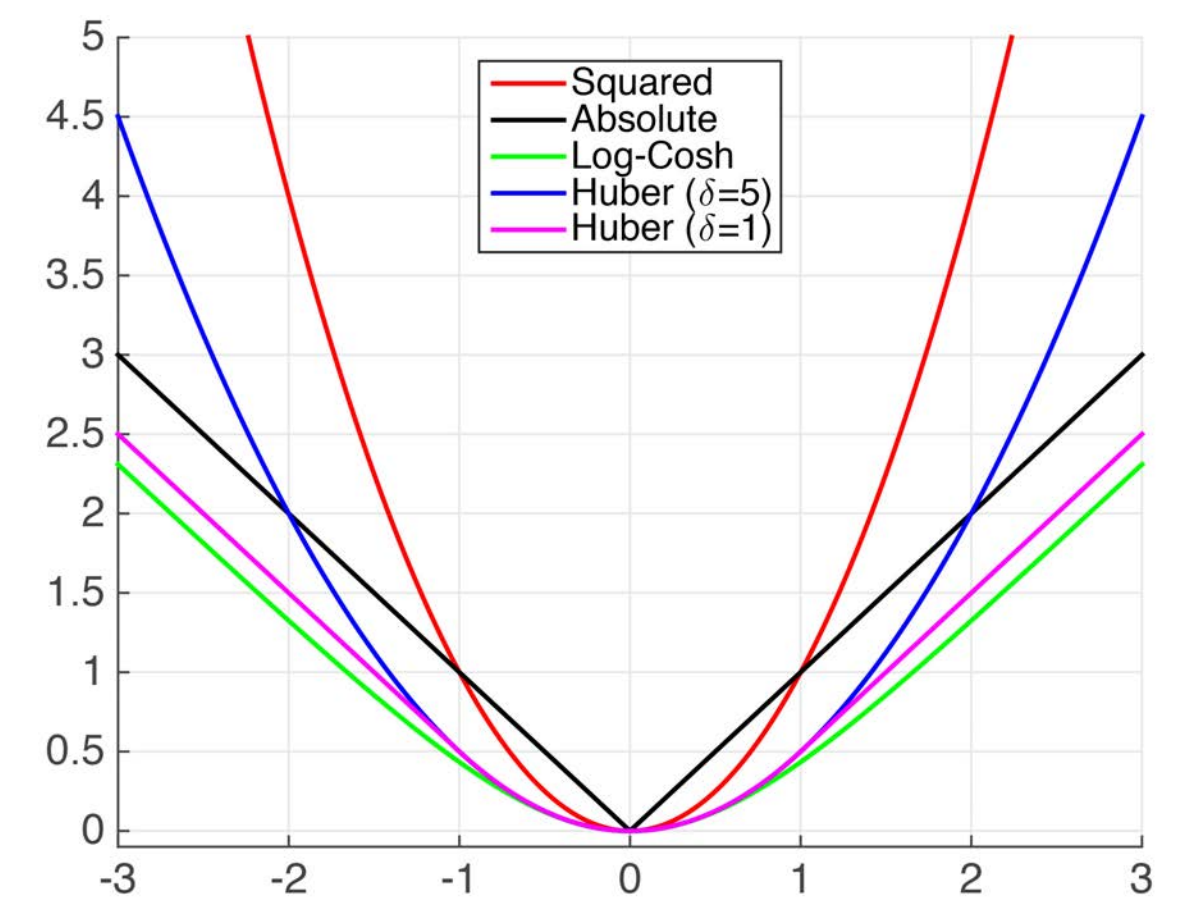
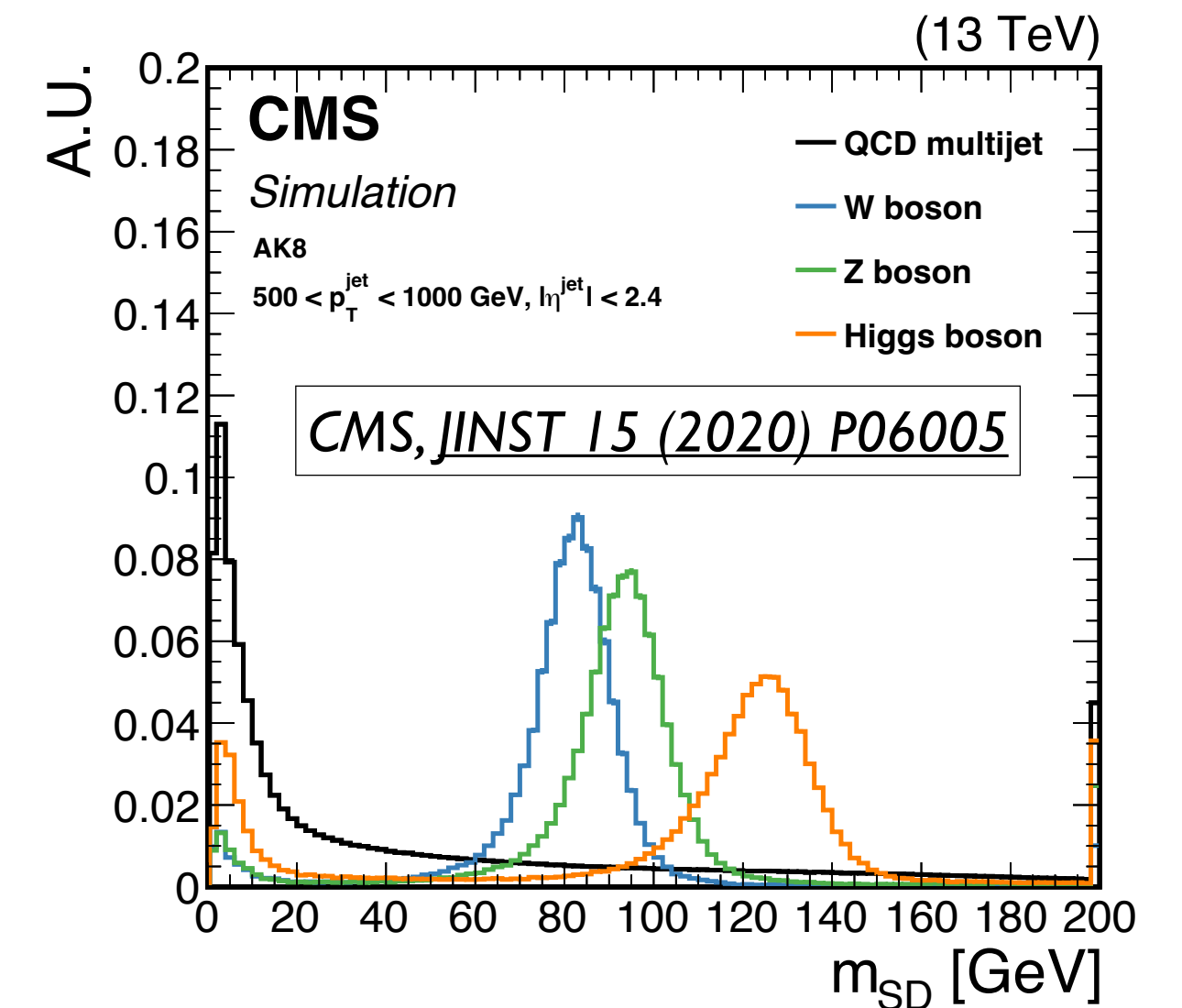
- ParticleNet-MD
 - using a special signal sample for training
 - hadronic decays of a spin-0 particle *X*
 - $X \rightarrow bb, X \rightarrow cc, X \rightarrow qq$
 - not a fixed mass, but a flat mass spectrum
 - $m(X) \in [15, 250] \text{ GeV}$
 - allows to easily reweight both signal and background to a \sim flat 2D distribution in (p_T , mass) for the training
- ParticleNet-MD shows the best performance
 - \sim 3-4x better background rejection compared to DeepAK8-MD (based on “adversarial training”)
 - only slight performance loss compared to the nominal version w/o mass decorrelation

MASS REGRESSION

- **Jet mass: one of the most powerful observables for boosted jet tagging**
 - characteristic mass peak for top/W/Z/H jets v.s. continuum for QCD jets
- **Mass regression:**
 - exploit deep learning to reconstruct jet mass with the highest possible resolution
 - training setup similar to the ParticleNet tagger
 - but: predict the jet mass directly from the jet constituents
- **Regression target:**
 - signal ($X \rightarrow bb/cc/qq$): generated particle mass of X [flat spectrum in 15 – 250 GeV]
 - background (QCD) jets: soft drop mass of the generated particle-level jet

Loss function

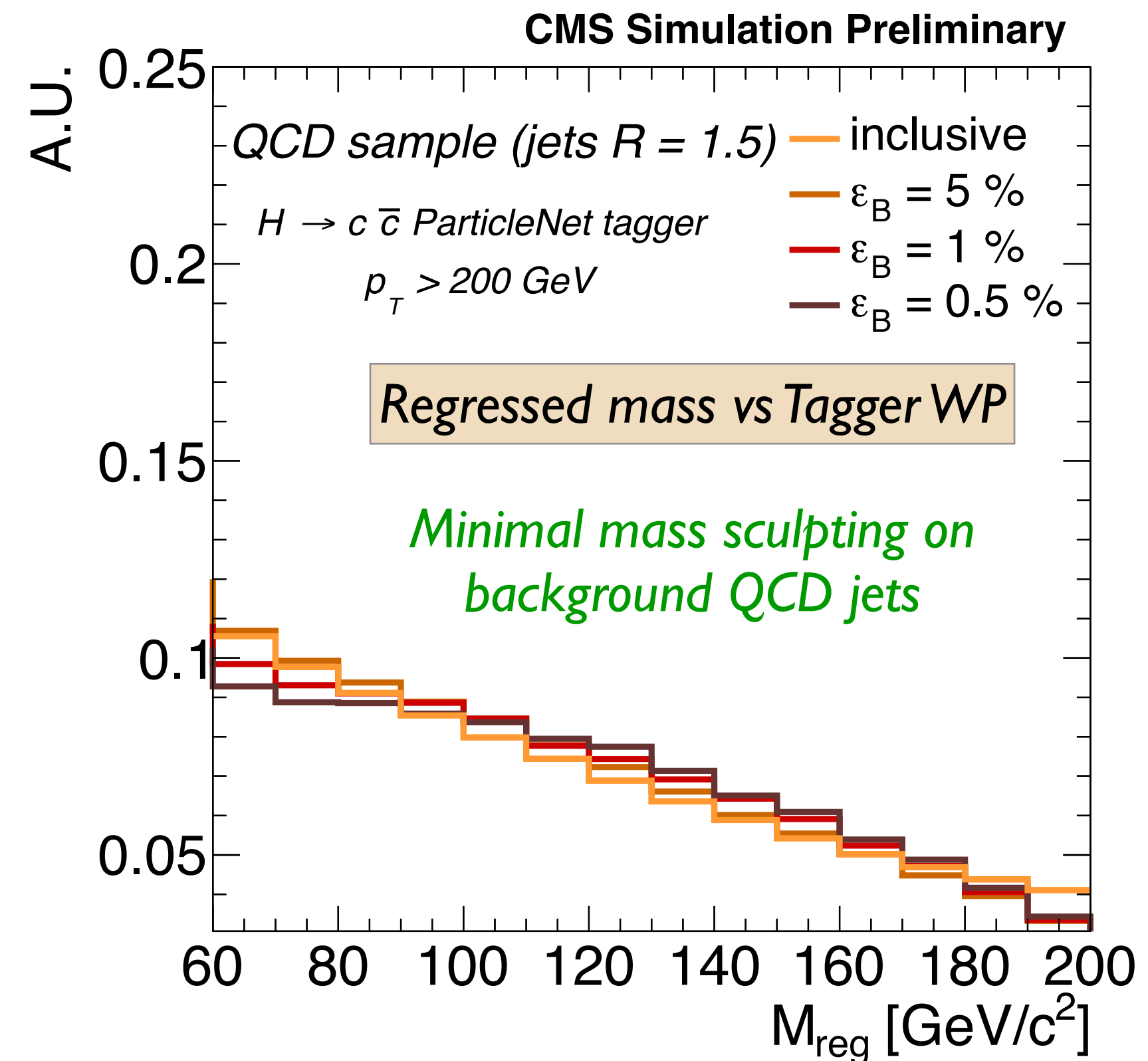
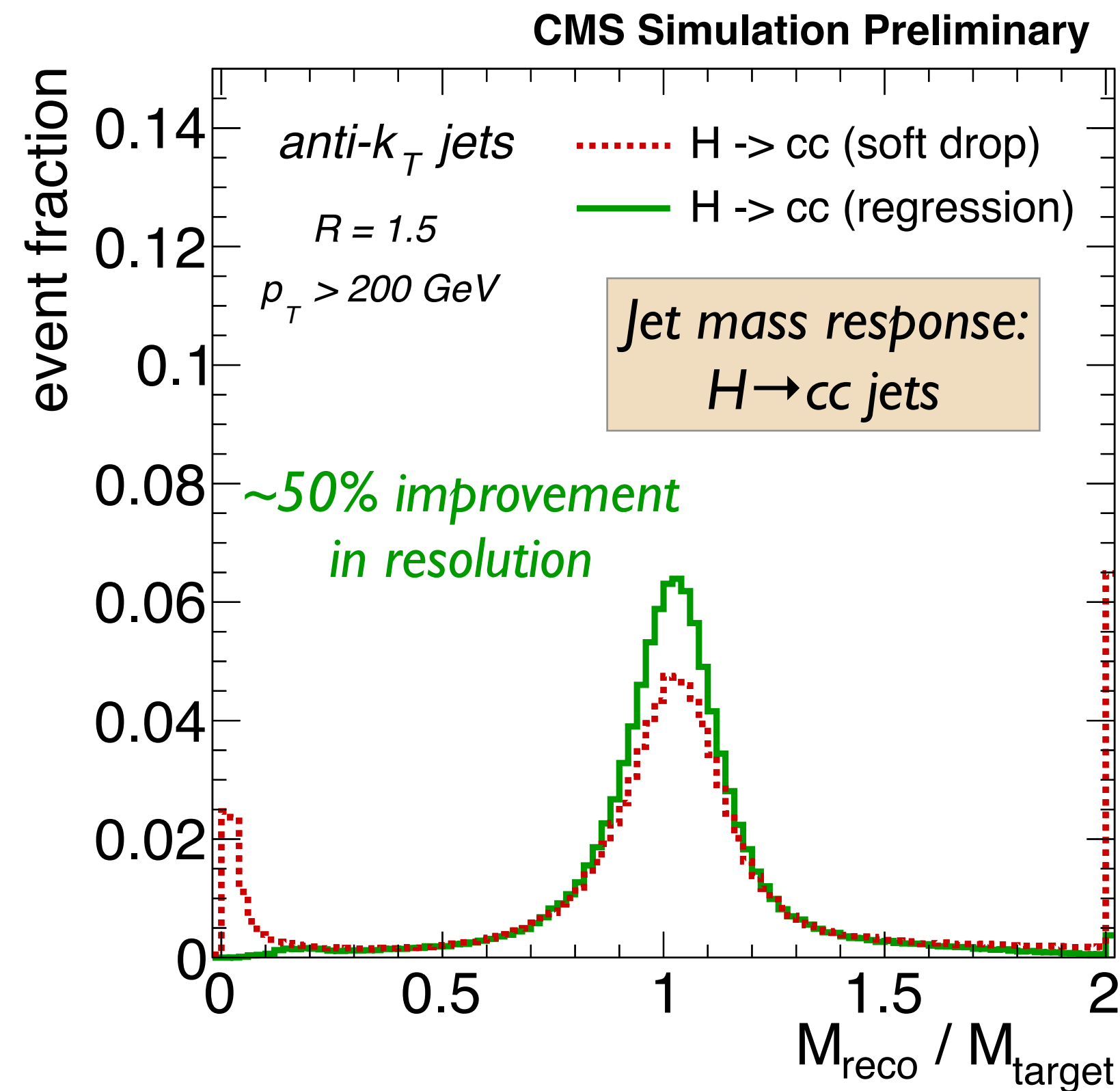
- LogCosh:
$$L(y, y^p) = \sum_{i=1}^n \log(\cosh(y_i^p - y_i))$$



<https://www.cs.cornell.edu/courses/cs4780/2015fa/web/lecturenotes/lecturenote10.html>

MASS REGRESSION: PERFORMANCE

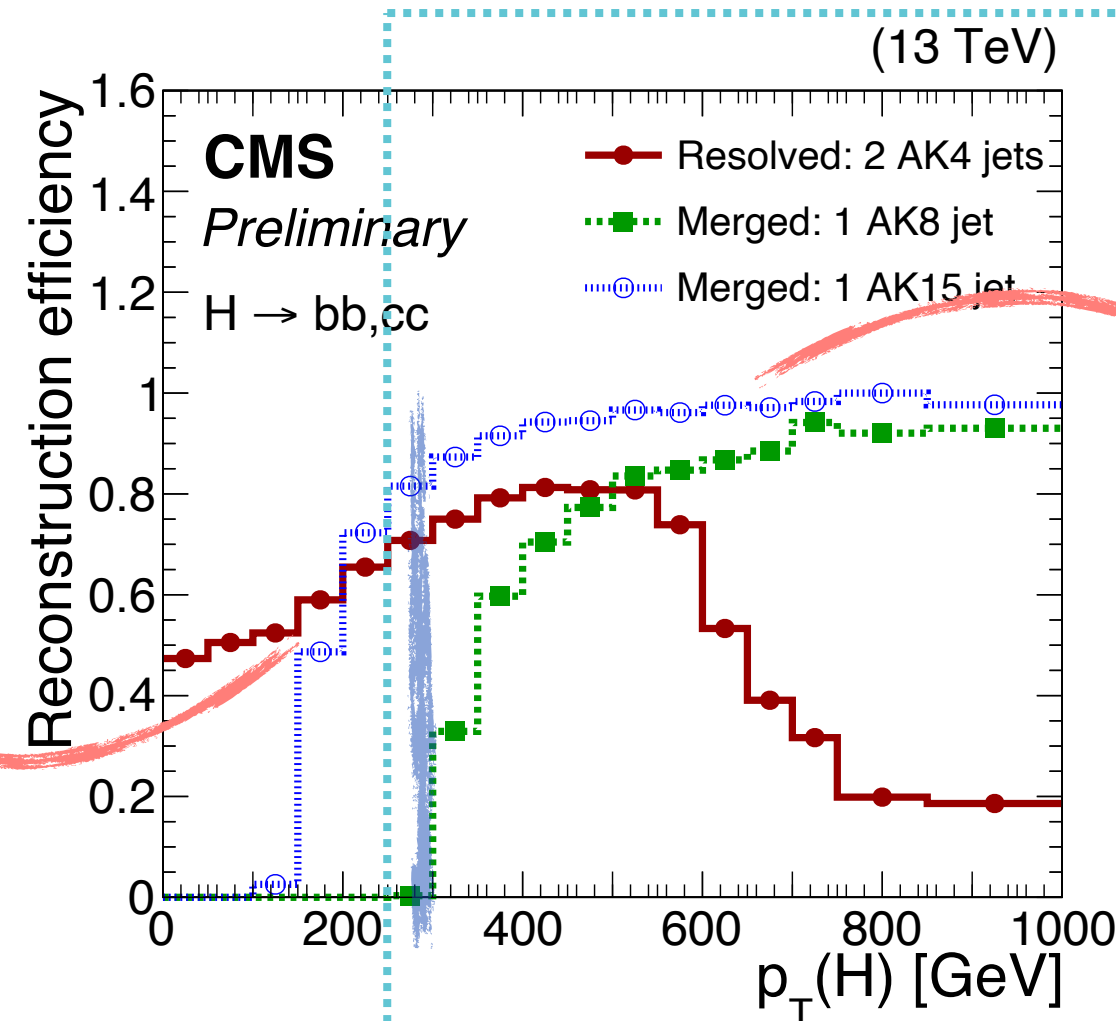
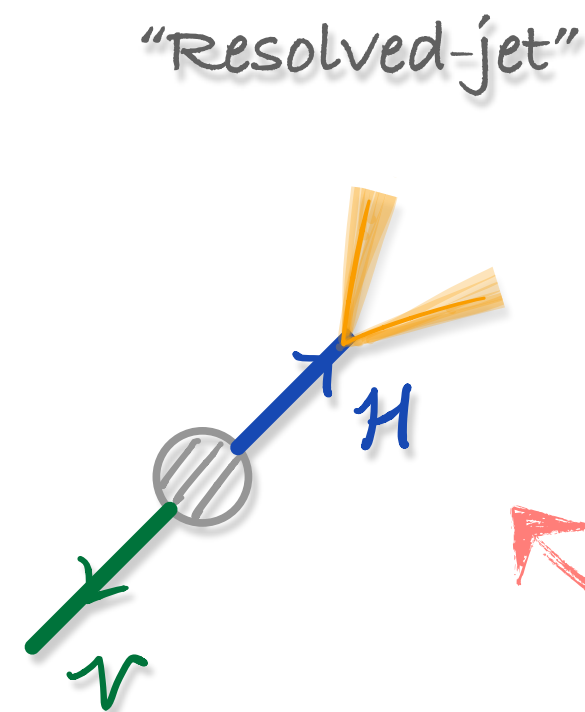
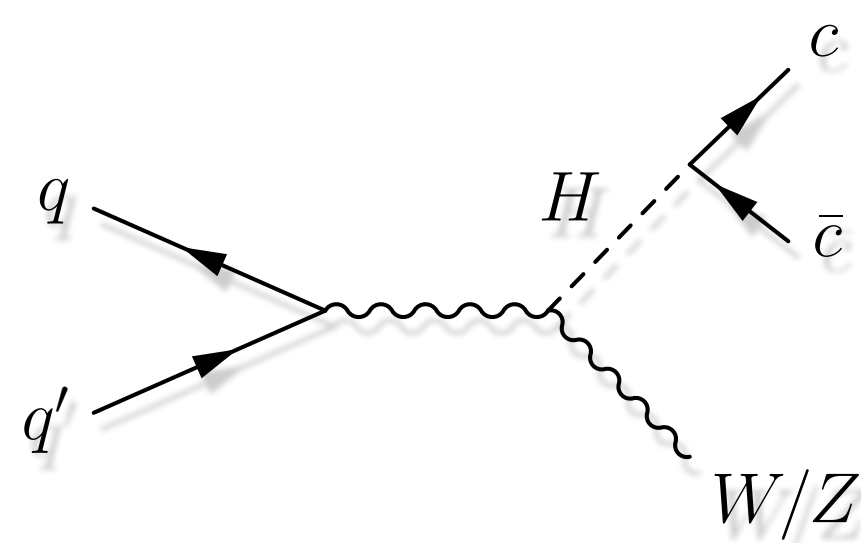
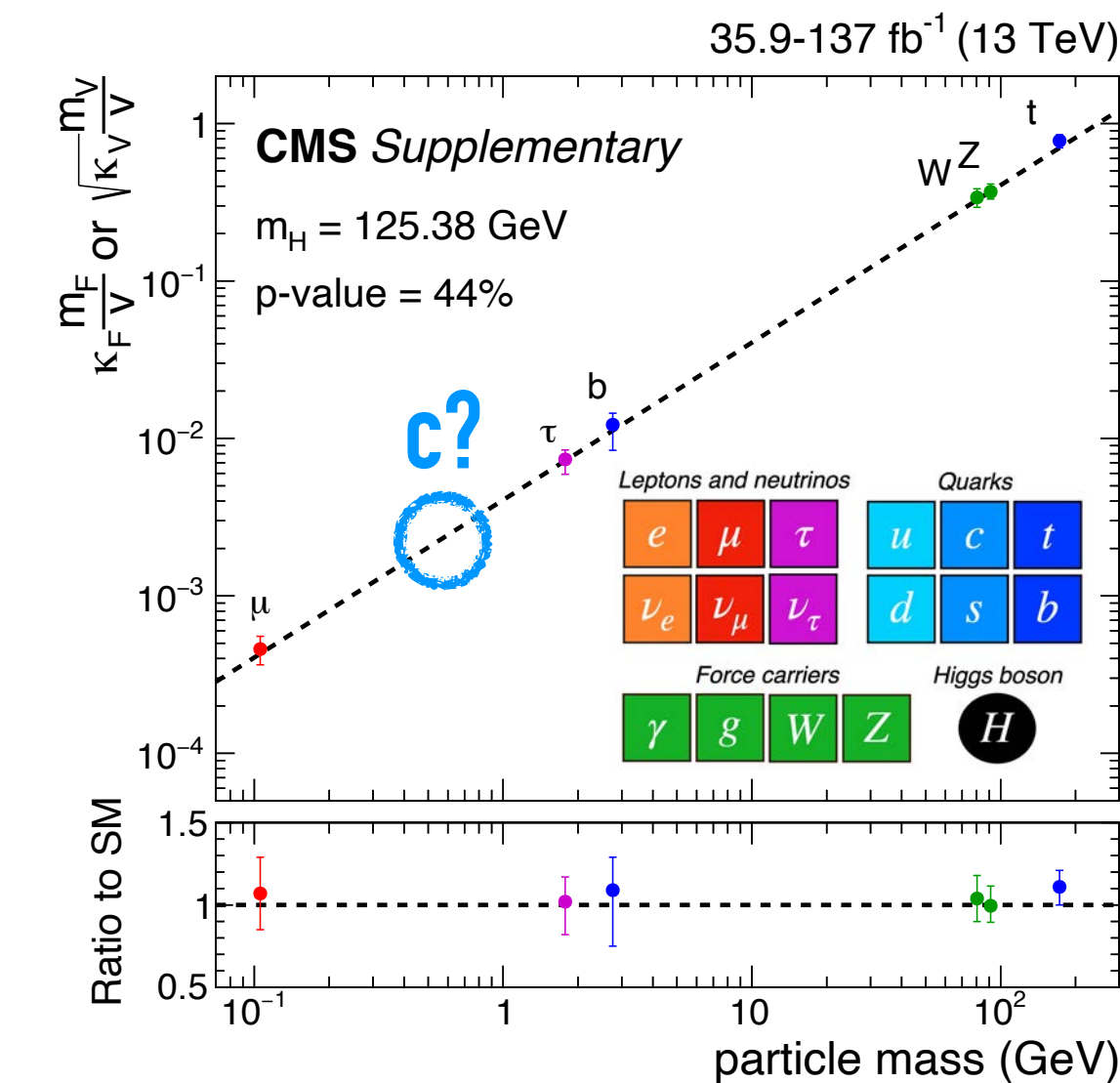
CMS DP-2021/017



$\sim 20\text{-}25\%$ improvement in the final sensitivity for H \rightarrow bb / H \rightarrow cc analyses

SEARCH FOR $H \rightarrow CC$

- Search for the Higgs boson decay to a pair of charm quarks ($H \rightarrow cc$)
 - next milestone in Higgs physics – couplings to second generation quarks
 - extremely challenging at the LHC:
 - small branching fraction ($\sim 3\%$) vs enormous backgrounds; difficulty in charm tagging
- $H \rightarrow cc$ search at CMS
 - targets WH/ZH production, with 3 channels: $Z \rightarrow \nu\nu$ (0L), $W \rightarrow \ell\nu$ (1L), $Z \rightarrow \ell\ell$ (2L) ($\ell = e, \mu$)
 - two complementary approaches to fully explore the $H \rightarrow cc$ decay topologies



$$\Delta R(c, c) \sim 2m(H)/p_T(H)$$

Powered by boosted jet tagging

Resolved-jet topology

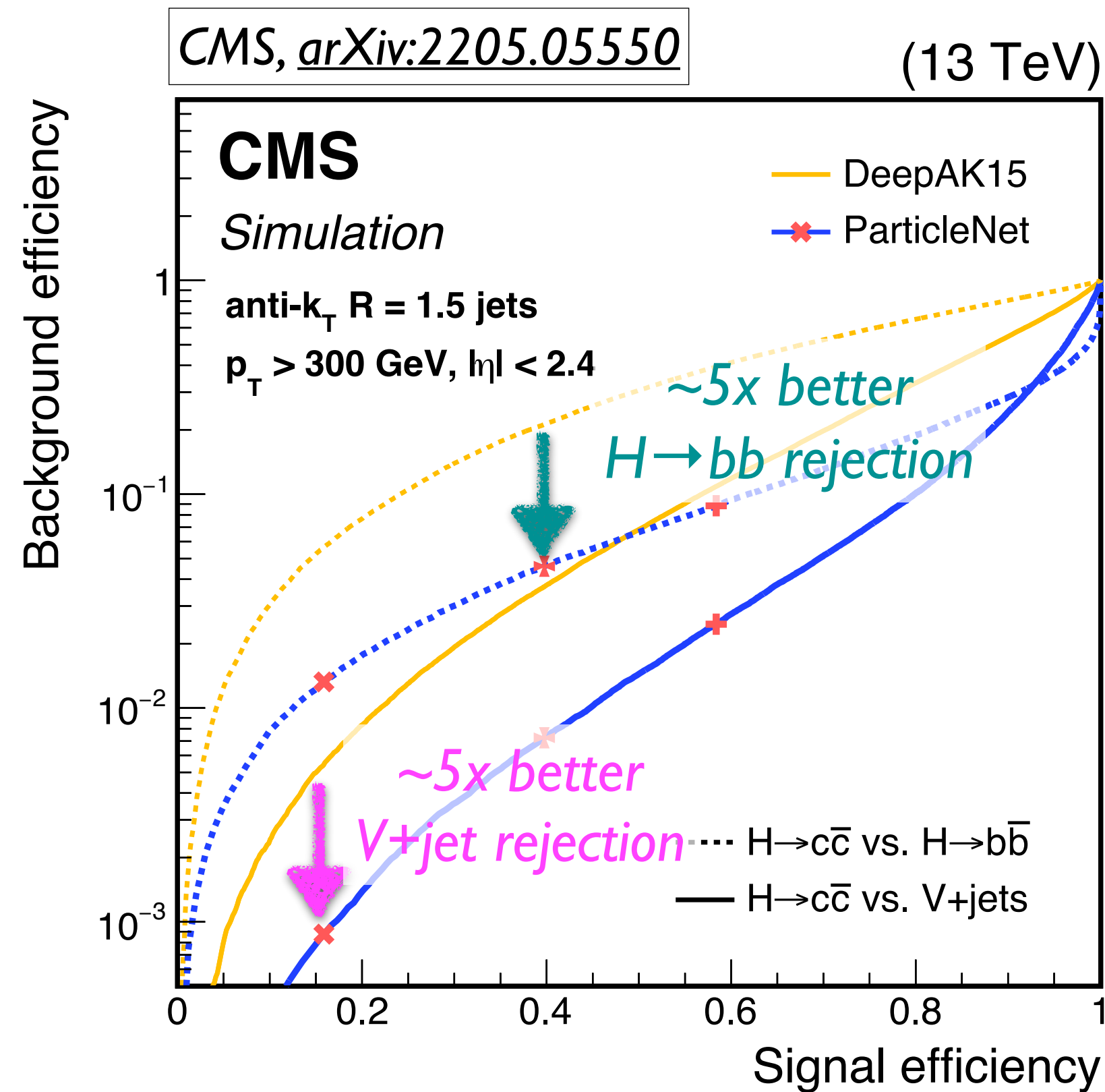
- reconstructs $H \rightarrow cc$ decay with two resolved jets ($R=0.4$)
- probes the bulk of the phase space

Merged-jet topology

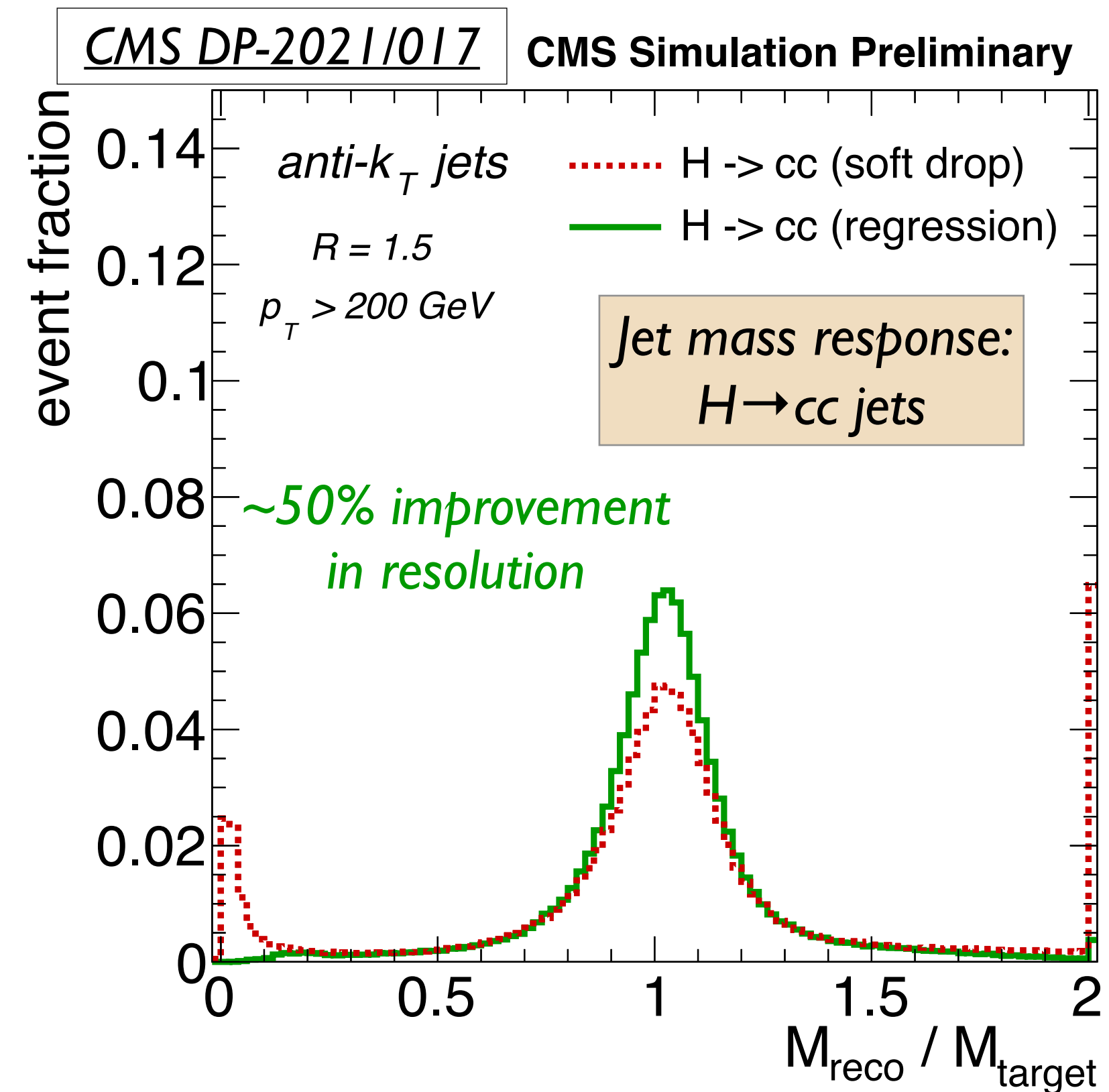
- reconstructs $H \rightarrow cc$ decay with one large-R jets ($R=1.5$)
- higher purity, but lower acceptance

IMPROVEMENTS IN $H \rightarrow CC$ RECONSTRUCTION

- The ParticleNet $H \rightarrow cc$ tagger and mass regression bring substantial improvements to the analysis



ParticleNet tagger for $H \rightarrow cc$ tagging
>2x improvement in final sensitivity

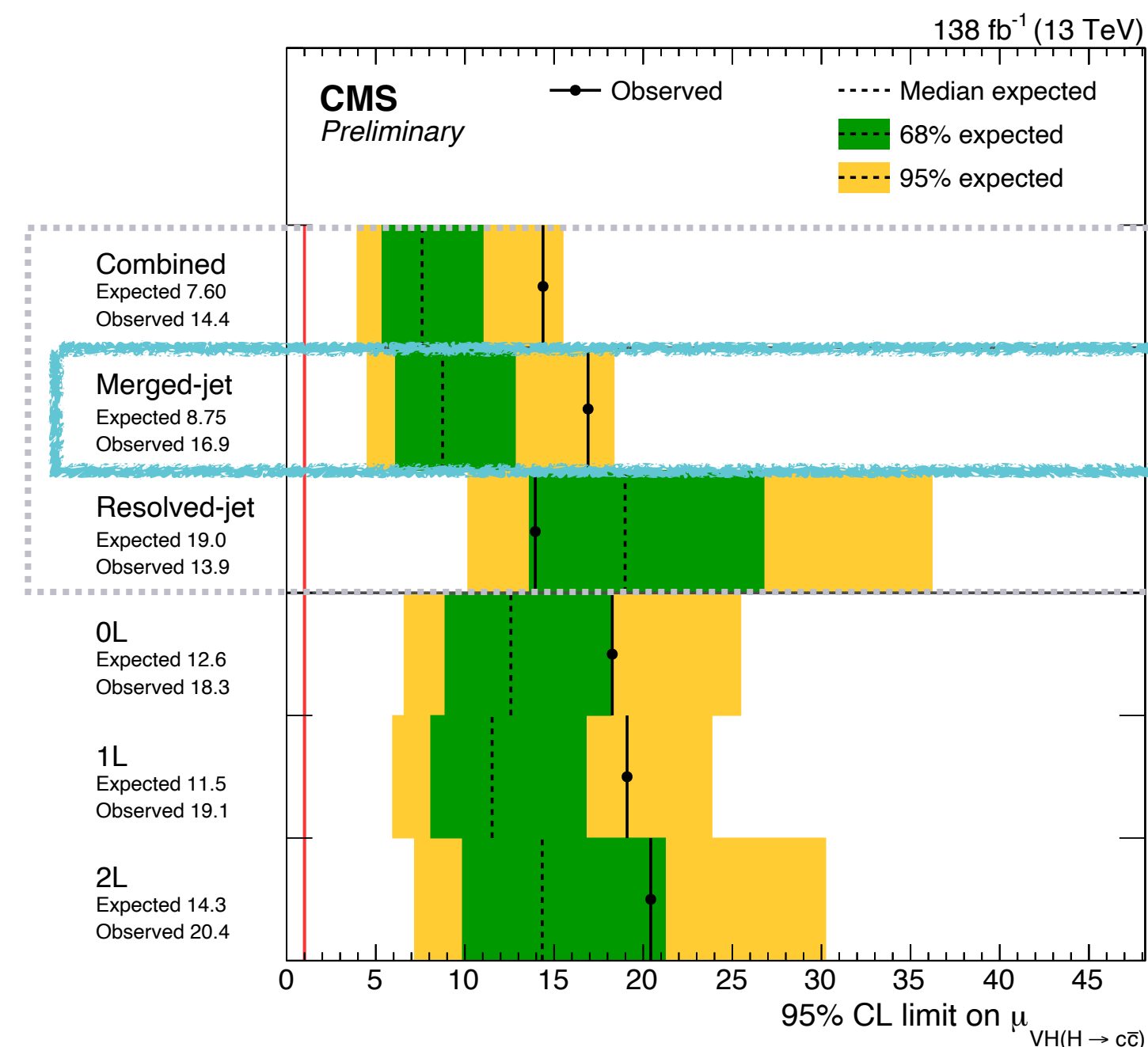
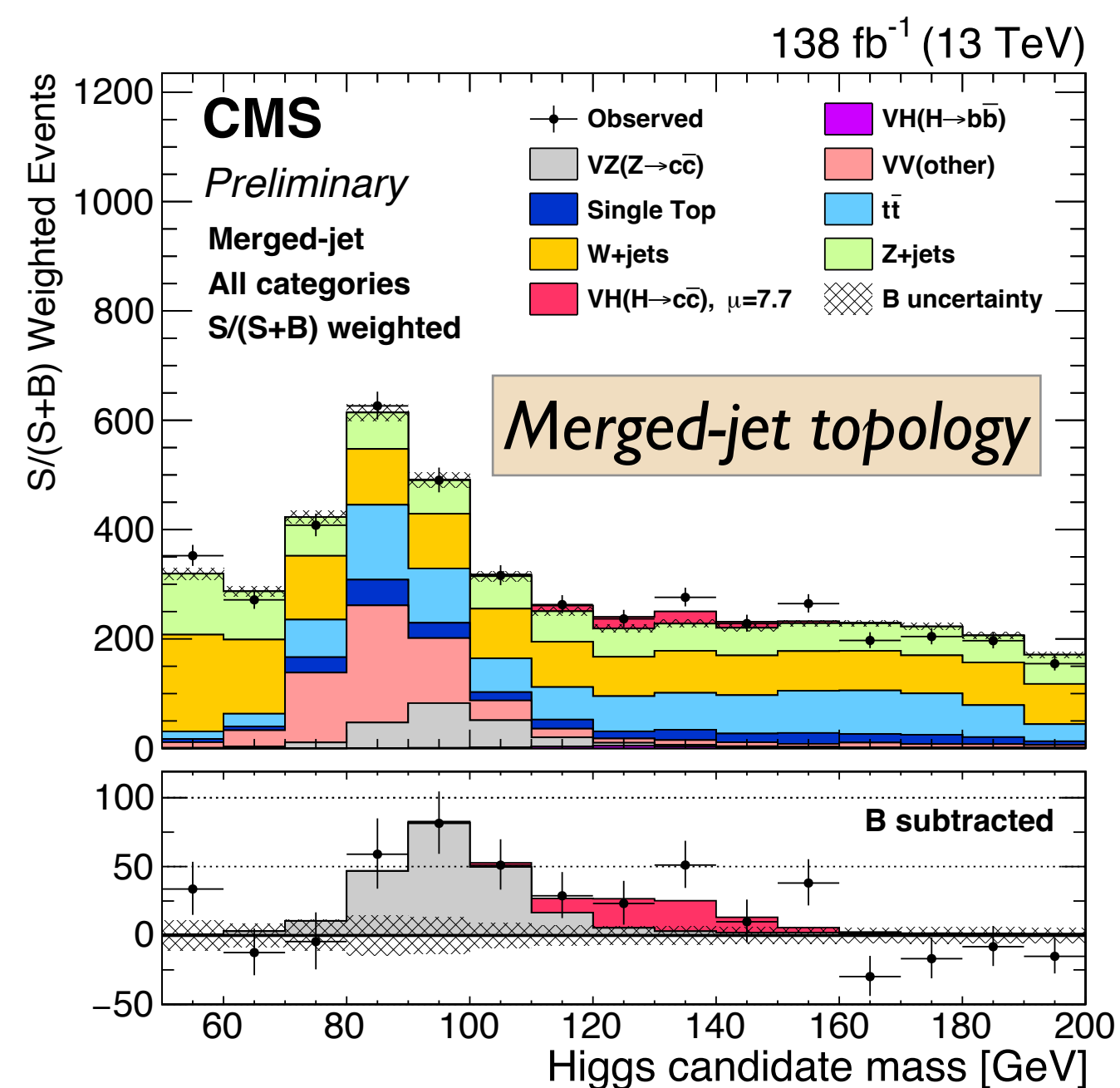


ParticleNet-based jet mass regression
~20-25% improvement in final sensitivity

H → CC RESULTS

CMS, [arXiv:2205.05550](https://arxiv.org/abs/2205.05550)

- VH(H → cc) results with the full Run-2 data set (138 fb⁻¹)
 - $\mu_{VH(H \rightarrow c\bar{c})} < 14$ (7.6) observed (expected)
 - substantially stronger than the ATLAS full Run-2 result: $\mu_{VH(H \rightarrow c\bar{c})} < 26$ (31) obs. (exp.) [arXiv:2201.11428]
 - expected sensitivity already comparable to the previous projection for HL-LHC w/ 3000 fb⁻¹: $\mu < 6.4$ [ATL-PHYS-PUB-2021-039]
- Analysis validated by measuring VZ(Z → cc): $\mu_{VZ(Z \rightarrow c\bar{c})} = 1.01^{+0.23}_{-0.21}$
- **First observation of Z → cc at a hadron collider, with a significance of 5.7σ**



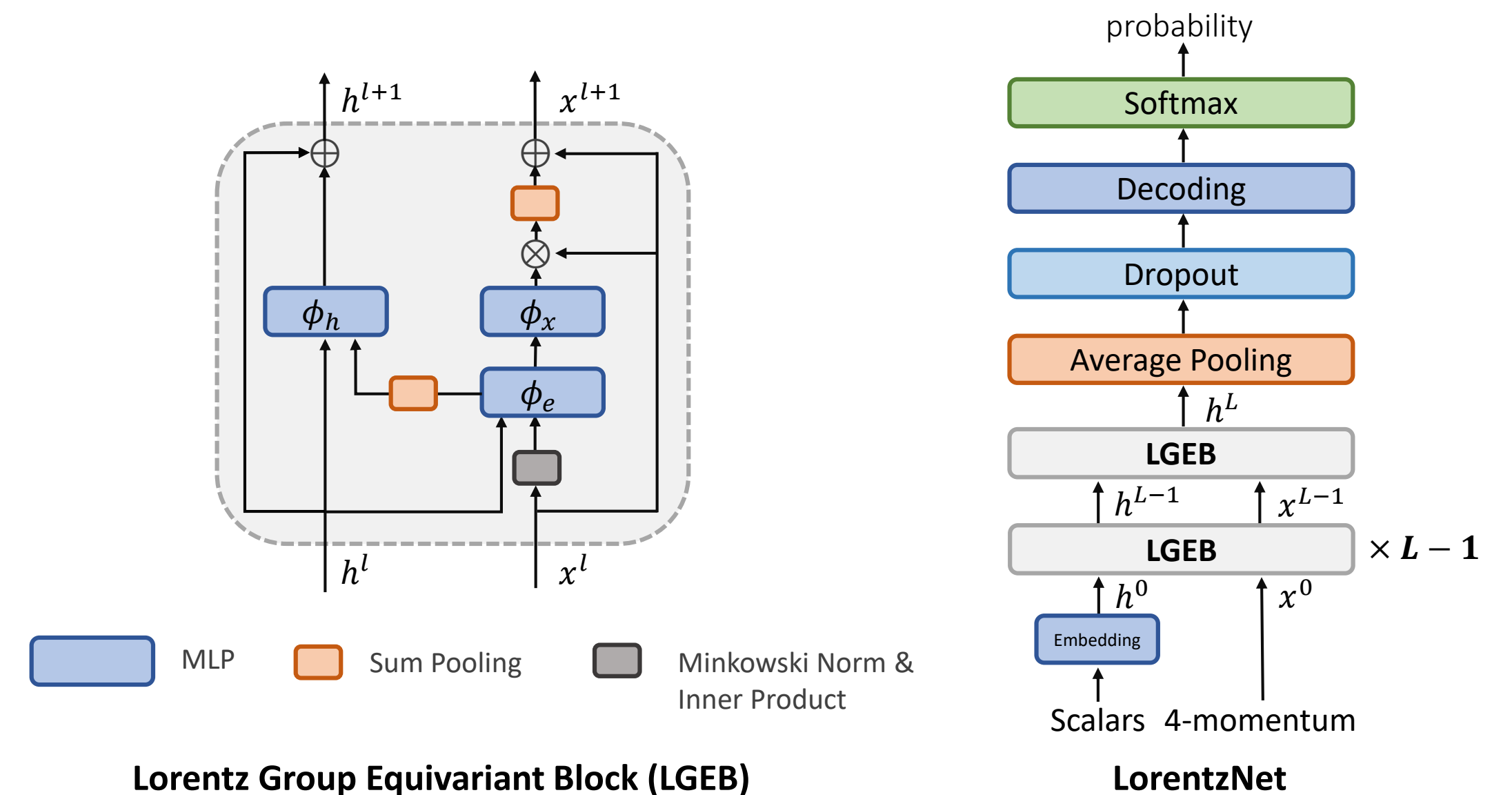
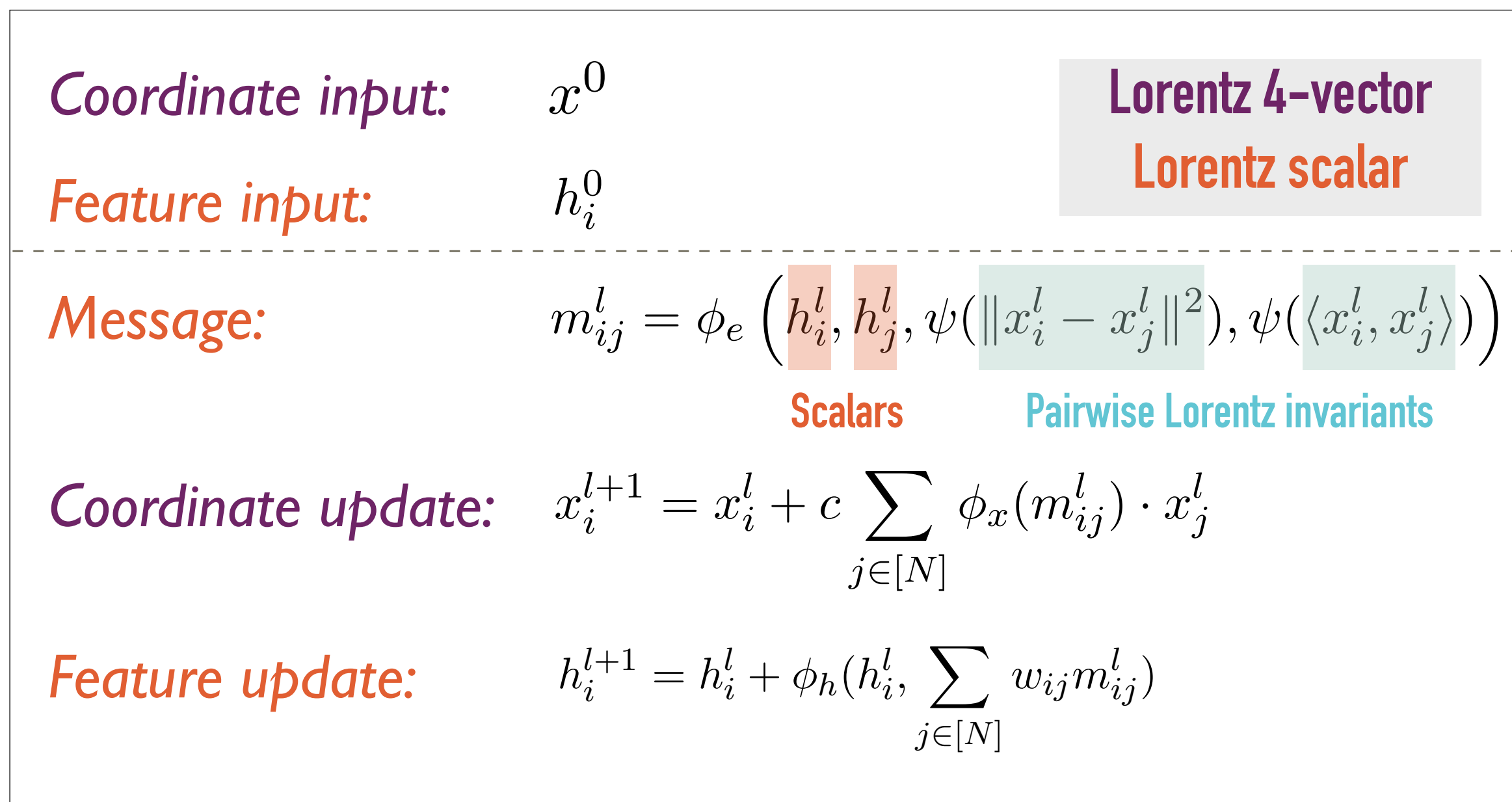
GOING BEYOND

LorentzNet

LORENTZNET

- Incorporating Lorentz symmetry into graph neural network architecture

S. Gong, Q. Meng, J. Zhang, HQ, C. Li, S. Qian,
W. Du, Z. M. Ma and T.Y. Liu,
JHEP 07 (2022) 030



cf. A. Bogatskiy, B. Anderson, J. Offermann, M. Roussi, D. Miller and R. Kondor,
arXiv: 2006.04780

LORENTZNET: PERFORMANCE

- Significant performance improvement, with fewer trainable parameters

S. Gong, Q. Meng, J. Zhang, HQ, C. Li, S. Qian,
W. Du, Z. M. Ma and T.Y. Liu,
JHEP 07 (2022) 030

Performance on top-tagging benchmark [SciPost Phys. 7 (2019) 014]

Model	Accuracy	AUC	$1/\epsilon_B$ ($\epsilon_S = 0.5$)	$1/\epsilon_B$ ($\epsilon_S = 0.3$)
ResNeXt	0.936	0.9837	302 ± 5	1147 ± 58
P-CNN	0.930	0.9803	201 ± 4	759 ± 24
PFN	0.932	0.9819	247 ± 3	888 ± 17
ParticleNet	0.940	0.9858	397 ± 7	1615 ± 93
EGNN	0.922	0.9760	148 ± 8	540 ± 49
LGN	0.929	0.9640	124 ± 20	435 ± 95
LorentzNet	0.942	0.9868	498 ± 18	2195 ± 173

Model complexity

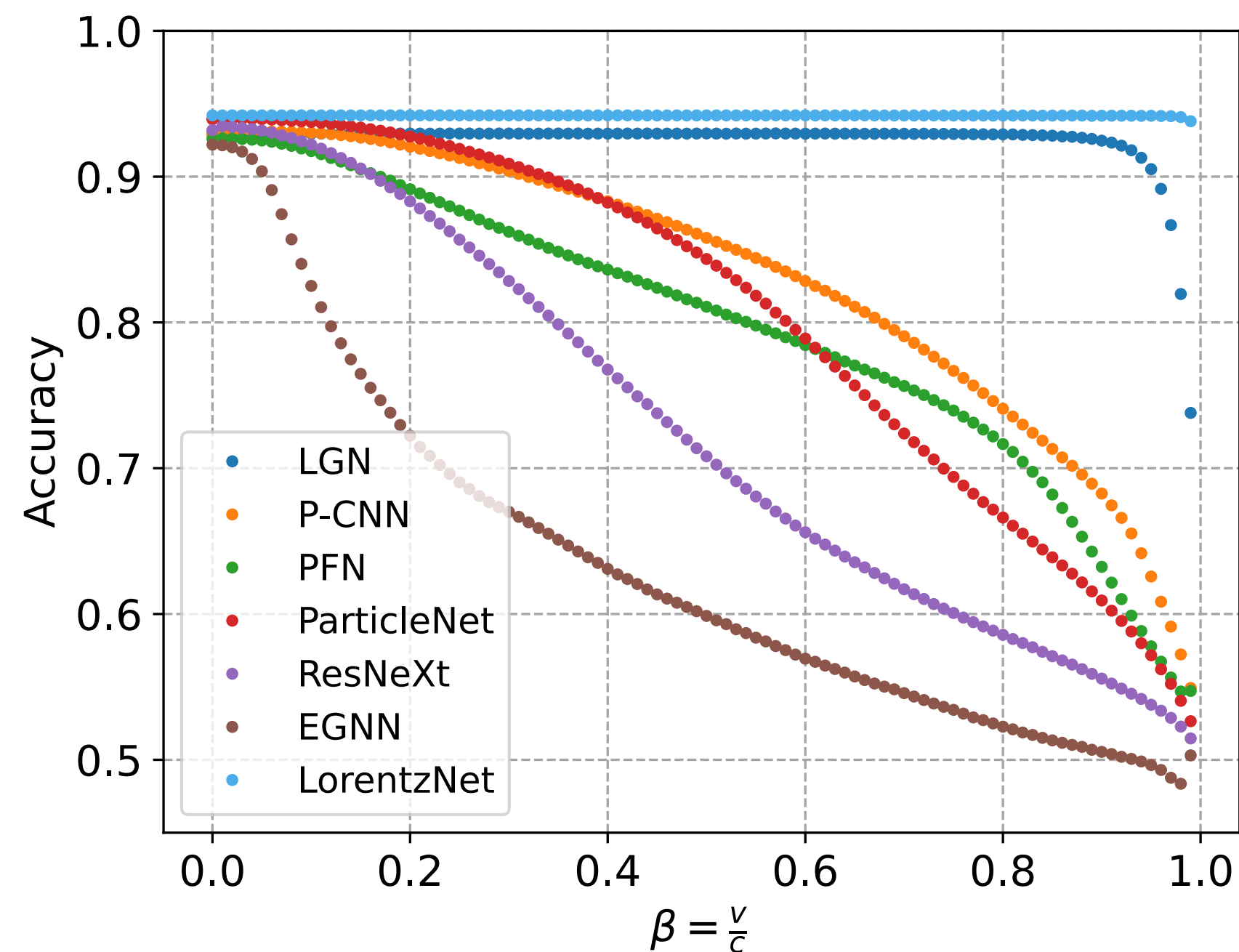
Model	Equivariance	Time on CPU (ms/batch)	Time on GPU (ms/batch)	#Params
ResNeXt	\times	5.5	0.34	1.46M
P-CNN	\times	0.6	0.11	348k
PFN	\times	0.6	0.12	82k
ParticleNet	\times	11.0	0.19	366k
EGNN	E(4)	30.0	0.30	222k
LGN	SO ⁺ (1,3)	51.4	1.66	4.5k
LorentzNet	SO ⁺ (1,3)	32.9	0.34	224k

LORENTZNET: BENEFITS FROM SYMMETRY

S. Gong, Q. Meng, J. Zhang, HQ, C. Li, S. Qian,
W. Du, Z. M. Ma and T.Y. Liu,
JHEP 07 (2022) 030

- Benefits from the symmetry preservation
 - model response invariant under Lorentz transformation
 - sample efficiency: incorporation of Lorentz symmetry allows to train with very few samples

Model stability under Lorentz boost



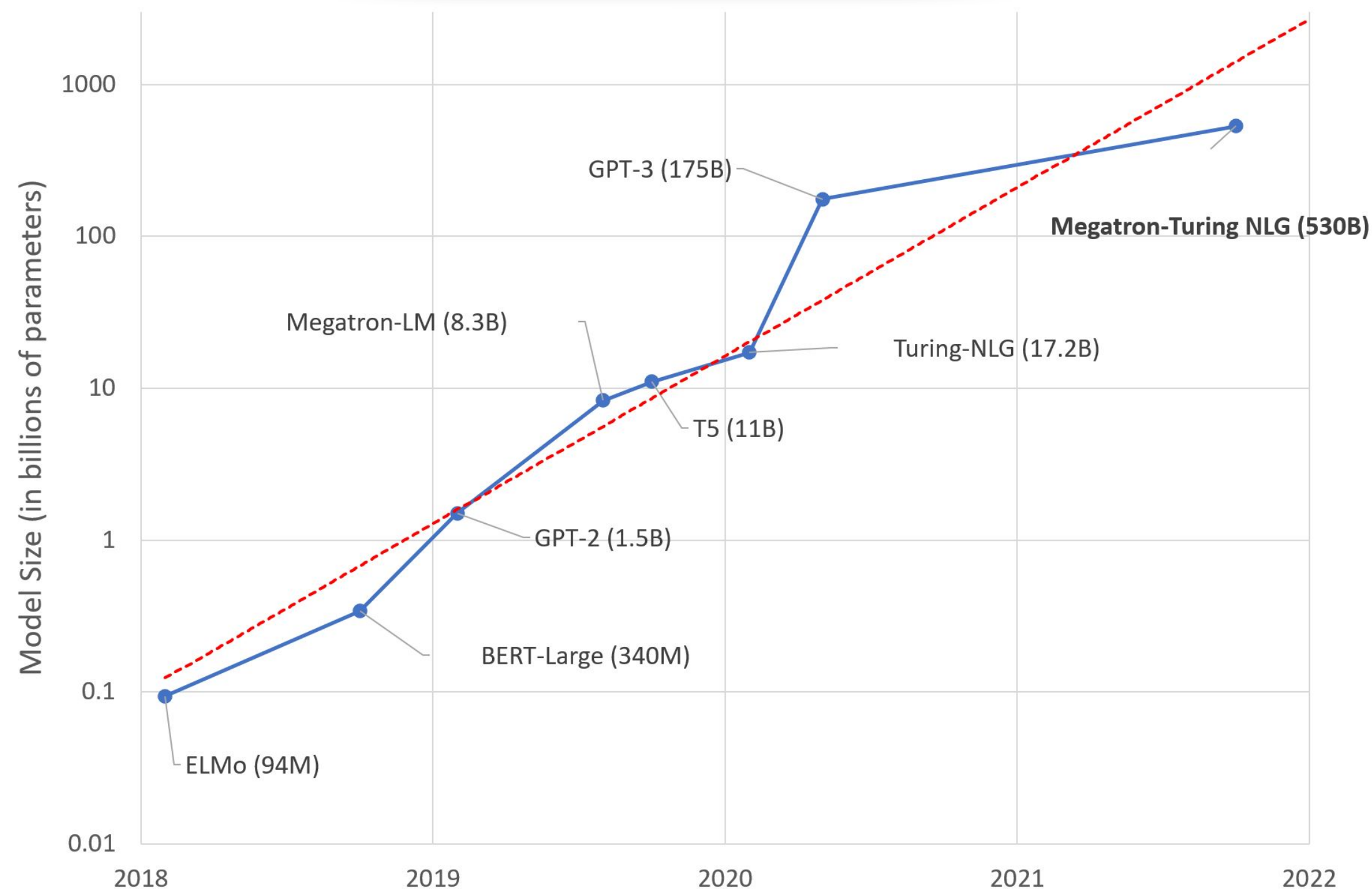
Performance when trained on a fraction of the top-tagging dataset

Training Fraction	Model	Accuracy	AUC	$1/\epsilon_B$ ($\epsilon_S = 0.5$)	$1/\epsilon_B$ ($\epsilon_S = 0.3$)
0.5% (~6k jets)	ParticleNet	0.913	0.9687	77 ± 4	199 ± 14
	LorentzNet	0.929	0.9793	176 ± 14	562 ± 72
1%	ParticleNet	0.919	0.9734	103 ± 5	287 ± 19
	LorentzNet	0.932	0.9812	209 ± 5	697 ± 58
5%	ParticleNet	0.931	0.9807	195 ± 4	609 ± 35
	LorentzNet	0.937	0.9839	293 ± 12	1108 ± 84

Particle Transformer

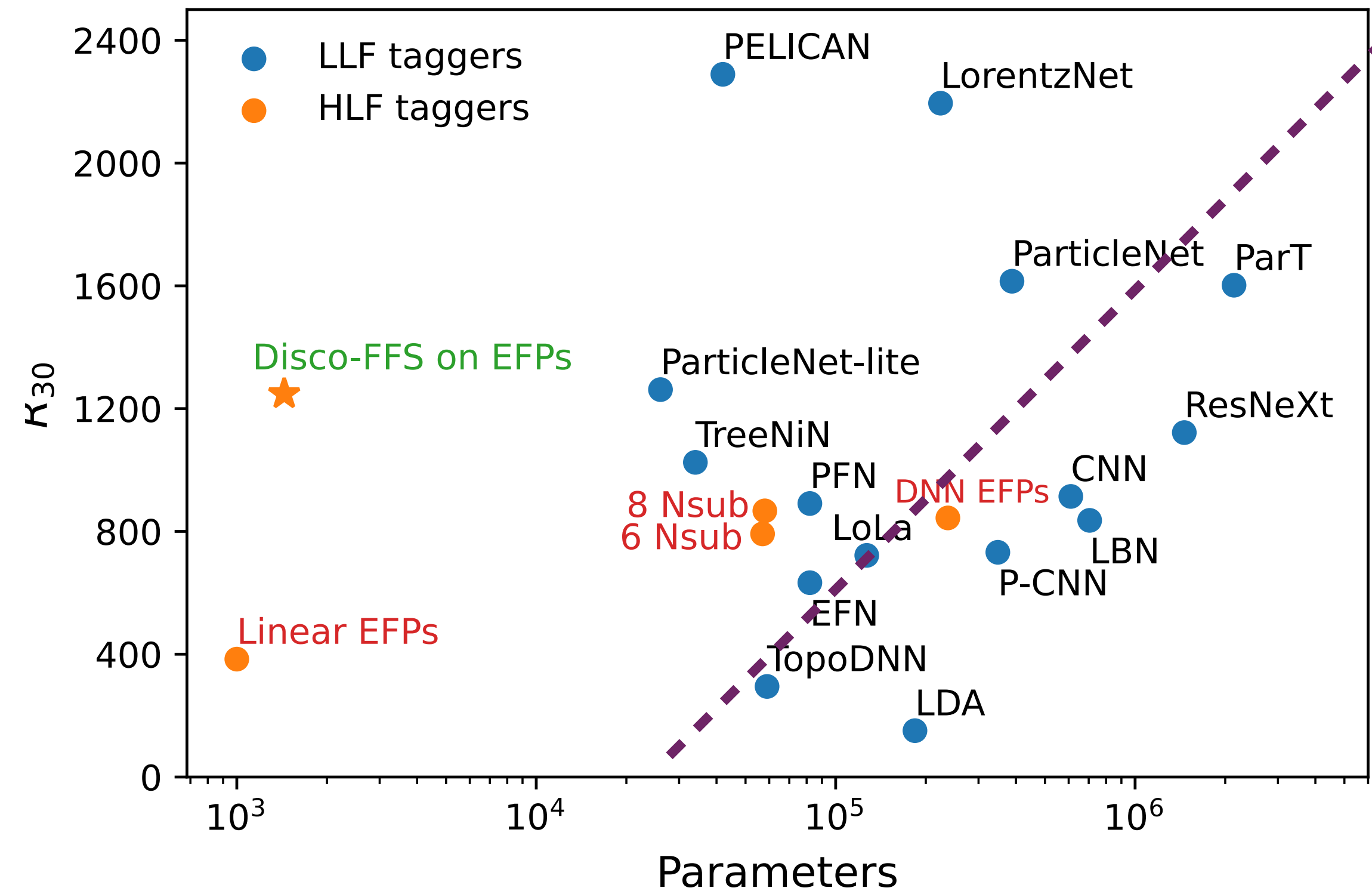
LARGE PHYSICS MODEL?

Natural language models



<https://huggingface.co/blog/large-language-models>

HEP models (jet tagging)



R. Das, G. Kasieczka and D. Shih, arXiv: 2212.00046

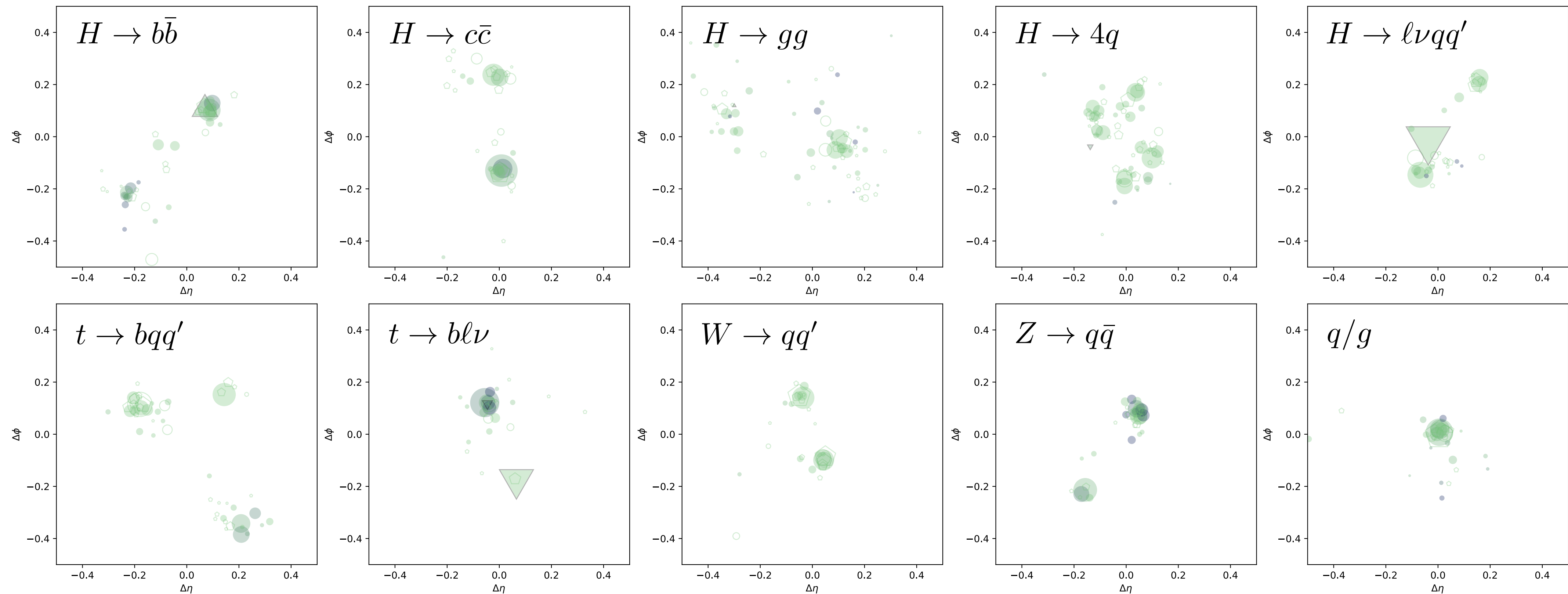
**Large Language Models (like GPT) has transformed NLP.
What if a Large Physics Model?**

A FIRST STEP



HQ, C. Li, S. Qian,
ICML 2022

- **JETCLASS**: a new large and comprehensive jet simulation dataset
 - 100M jets in 10 classes: ~two orders of magnitude larger than existing public datasets

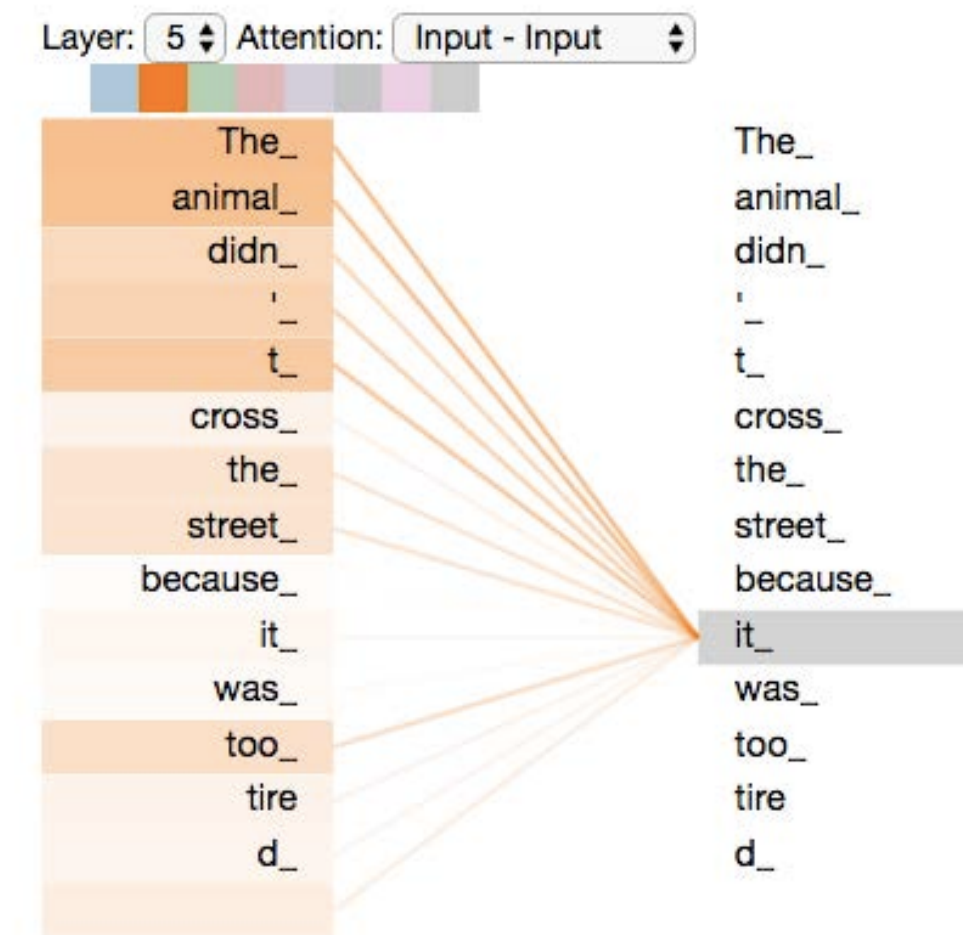
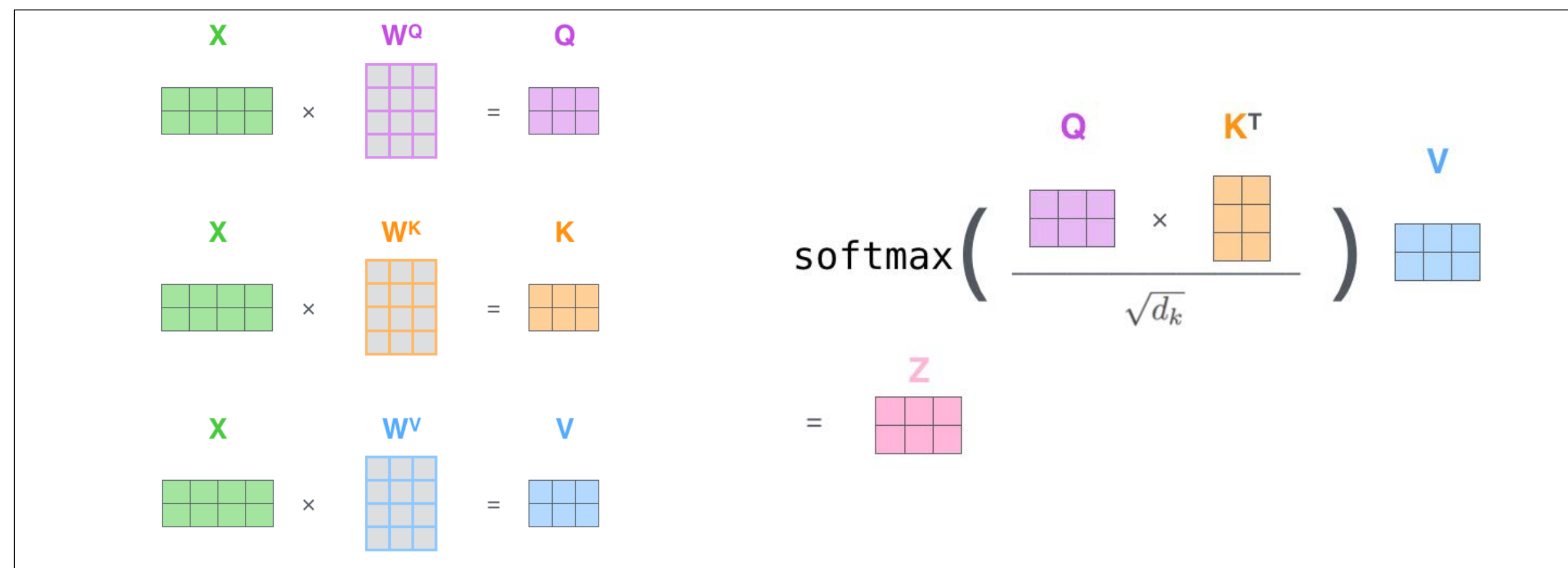


We invite the community to explore and experiment with this dataset and extend the boundary of deep learning and HEP even further.

PARTICLE TRANSFORMER

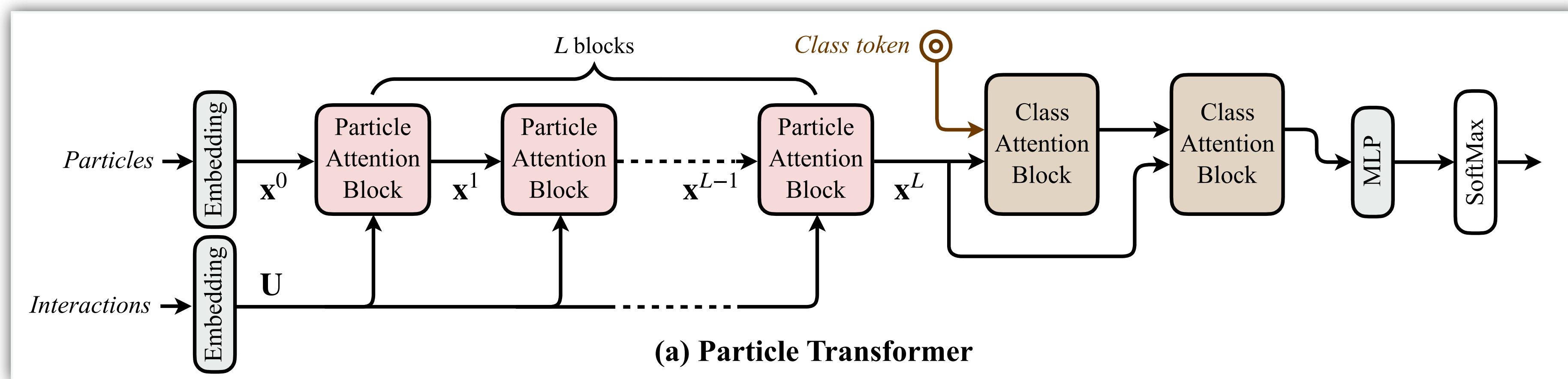
- **Transformers:** the new state-of-the-art architecture in ML – foundation of LLM like BERT/GPT

- core concept: self-attention mechanism



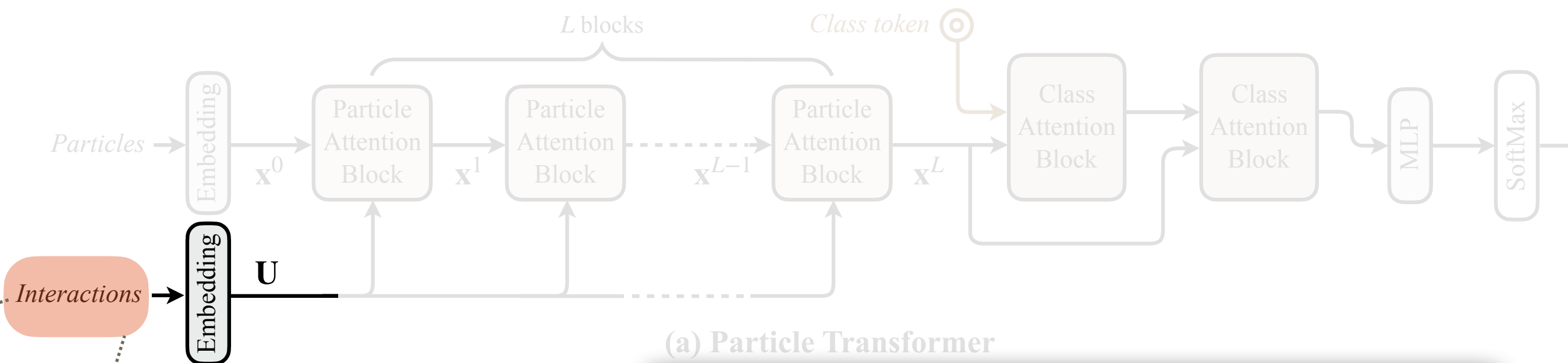
- **Particle Transformer (ParT):** Transformer model **tailored for particle physics**

HQ, C. Li, S. Qian,
ICML 2022



PARTICLE TRANSFORMER: ARCHITECTURE

HQ, C. Li, S. Qian,
ICML 2022



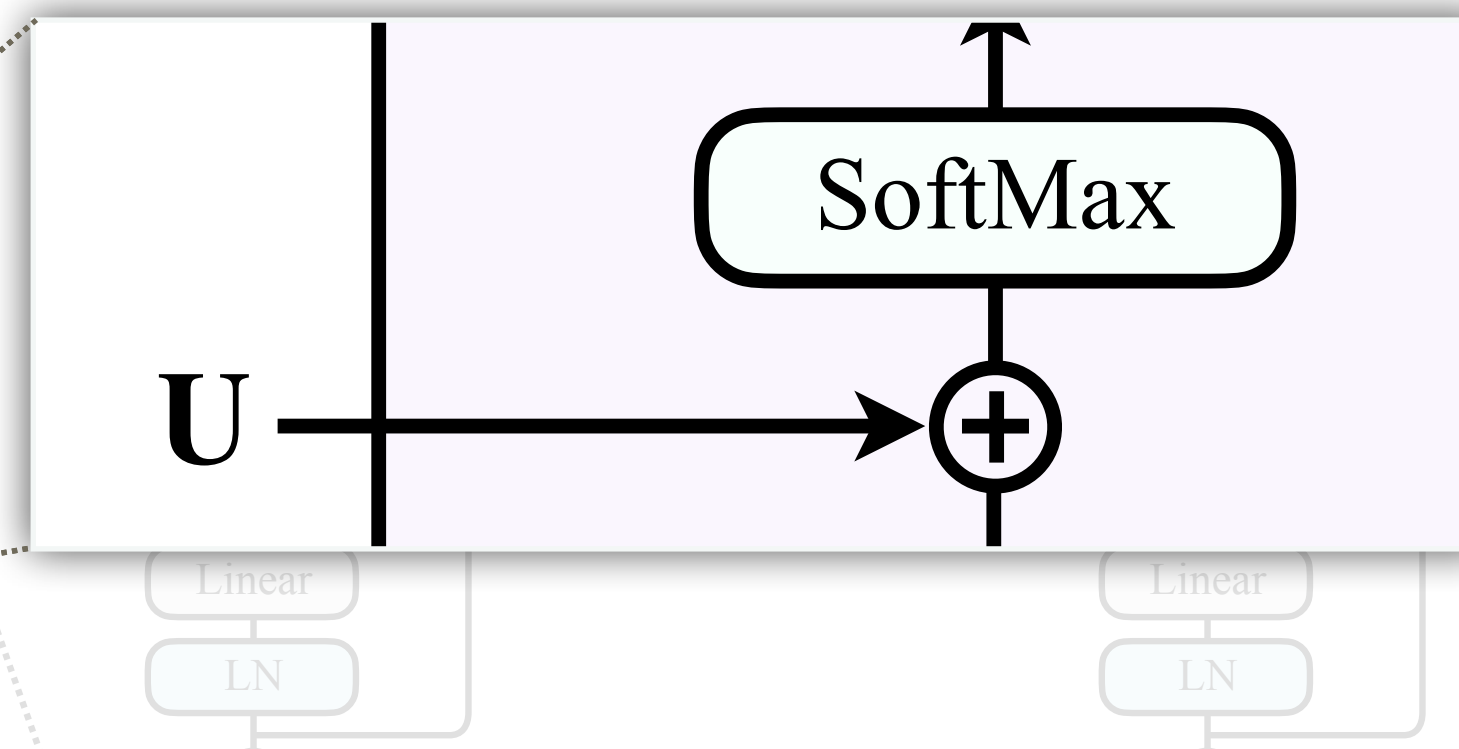
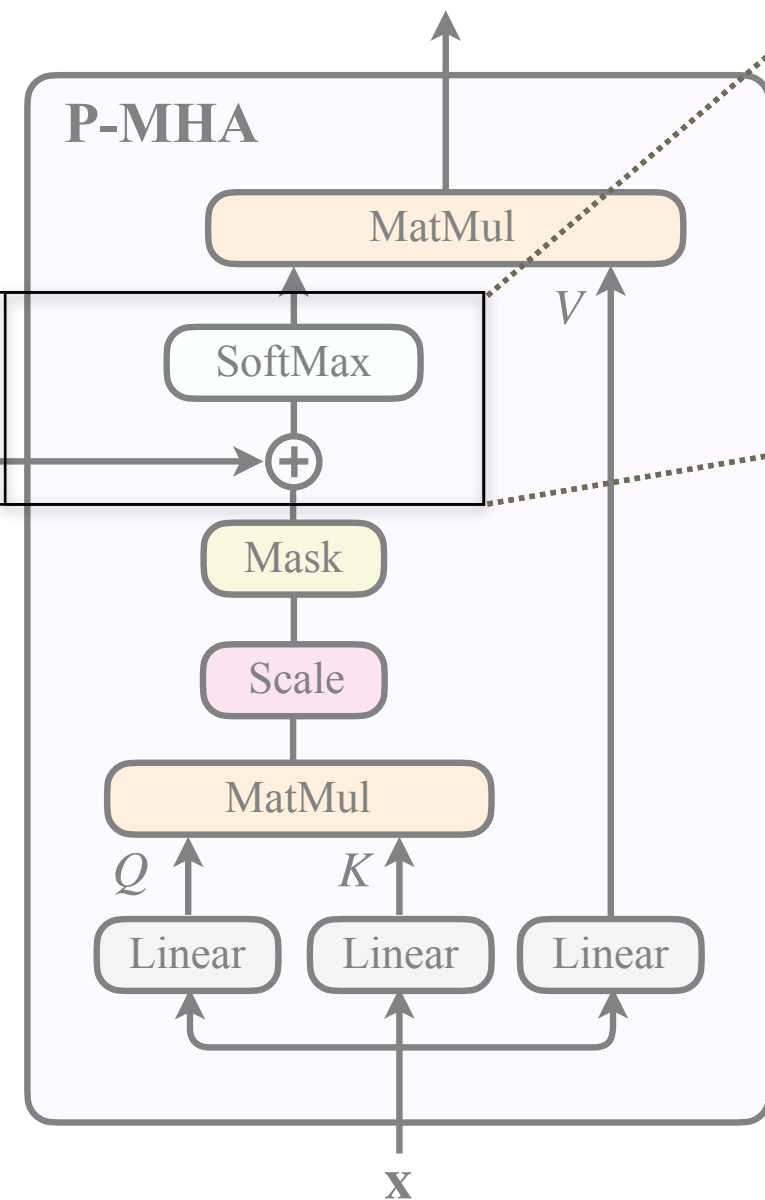
$$\Delta = \sqrt{(y_a - y_b)^2 + (\phi_a - \phi_b)^2},$$

$$k_T = \min(p_{T,a}, p_{T,b}) \Delta,$$

$$z = \min(p_{T,a}, p_{T,b}) / (p_{T,a} + p_{T,b}),$$

$$m^2 = (E_a + E_b)^2 - \|\mathbf{p}_a + \mathbf{p}_b\|^2,$$

and many other possible pairwise features...



$$\text{P-MHA}(Q, K, V) = \text{SoftMax}(QK^T / \sqrt{d_k} + \mathbf{U})V,$$

Injection of (physics-inspired) pairwise features to "bias" the dot-product self-attention

PARTICLE TRANSFORMER: PERFORMANCE

	All classes		$H \rightarrow b\bar{b}$	$H \rightarrow c\bar{c}$	$H \rightarrow gg$	$H \rightarrow 4q$	$H \rightarrow \ell\nu qq'$	$t \rightarrow bqq'$	$t \rightarrow bl\nu$	$W \rightarrow qq'$	$Z \rightarrow q\bar{q}$
	Accuracy	AUC	Rej _{50%}	Rej _{50%}	Rej _{50%}	Rej _{50%}	Rej _{99%}	Rej _{50%}	Rej _{99.5%}	Rej _{50%}	Rej _{50%}
PFN	0.772	0.9714	2924	841	75	198	265	797	721	189	159
P-CNN	0.809	0.9789	4890	1276	88	474	947	2907	2304	241	204
ParticleNet	0.844	0.9849	7634	2475	104	954	3339	10526	11173	347	283
ParT	0.861	0.9877	10638	4149	123	1864	5479	32787	15873	543	402
ParT (plain)	0.849	0.9859	9569	2911	112	1185	3868	17699	12987	384	311

JETCLASS dataset (100M jets)

- Particle Transformer (ParT): significant performance improvement!
 - compared to the existing state-of-the-art, ParticleNet
 - 1.7% increase in accuracy
 - up to 3x increase in background rejection (Rej_{X%})

$$\text{Rej}_{X\%} \equiv 1/\text{FPR at TPR} = X\%,$$

PARTICLE TRANSFORMER: PERFORMANCE

	All classes		$H \rightarrow b\bar{b}$	$H \rightarrow c\bar{c}$	$H \rightarrow gg$	$H \rightarrow 4q$	$H \rightarrow \ell\nu qq'$	$t \rightarrow bqq'$	$t \rightarrow bl\nu$	$W \rightarrow qq'$	$Z \rightarrow q\bar{q}$
	Accuracy	AUC	Rej _{50%}	Rej _{50%}	Rej _{50%}	Rej _{50%}	Rej _{99%}	Rej _{50%}	Rej _{99.5%}	Rej _{50%}	Rej _{50%}
PFN	0.772	0.9714	2924	841	75	198	265	797	721	189	159
P-CNN	0.809	0.9789	4890	1276	88	474	947	2907	2304	241	204
ParticleNet	0.844	0.9849	7634	2475	104	954	3339	10526	11173	347	283
ParT	0.861	0.9877	10638	4149	123	1864	5479	32787	15873	543	402
ParT (plain)	0.849	0.9859	9569	2911	112	1185	3868	17699	12987	384	311

JETCLASS dataset (100M jets)

- Particle Transformer (ParT): significant performance improvement!
 - compared to the existing state-of-the-art, ParticleNet
 - 1.7% increase in accuracy
 - **up to 3x increase in background rejection (Rej_{x%})**
- ParT (plain): plain Transformer w/o interaction features
 - 1.2% drop in accuracy compared to full ParT
 - **Physics-driven modification of self-attention plays a key role!**

Model complexity

	Accuracy	# params	FLOPs
PFN	0.772	86.1 k	4.62 M
P-CNN	0.809	354 k	15.5 M
ParticleNet	0.844	370 k	540 M
ParT	0.861	2.14 M	340 M
ParT (plain)	0.849	2.13 M	260 M

PARTICLE TRANSFORMER: PRE-TRAINING + FINE-TUNING

- The large Transformer-based model enables new training paradigm
 - (supervised) pre-training on a large dataset (e.g., JETCLASS) & fine-tuning to downstream tasks
 - significantly outperforms existing models

Top quark tagging benchmark ($\sim 2M$ jets) [SciPost Phys. 7 (2019) 014]

	Accuracy	AUC	Rej _{50%}	Rej _{30%}
P-CNN	0.930	0.9803	201 ± 4	759 ± 24
PFN	—	0.9819	247 ± 3	888 ± 17
ParticleNet	0.940	0.9858	397 ± 7	1615 ± 93
JEDI-net (w/ $\sum O$)	0.930	0.9807	—	774.6
PCT	0.940	0.9855	392 ± 7	1533 ± 101
LGN	0.929	0.964	—	435 ± 95
rPCN	—	0.9845	364 ± 9	1642 ± 93
LorentzNet	0.942	0.9868	498 ± 18	2195 ± 173
ParT	0.940	0.9858	413 ± 16	1602 ± 81
ParticleNet-f.t.	0.942	0.9866	487 ± 9	1771 ± 80
ParT-f.t.	0.944	0.9877	691 ± 15	2766 ± 130

Quark-gluon tagging benchmark ($\sim 2M$ jets) [JHEP 01 (2019) 121]

	Accuracy	AUC	Rej _{50%}	Rej _{30%}
P-CNN _{exp}	0.827	0.9002	34.7	91.0
PFN _{exp}	—	0.9005	34.7 ± 0.4	—
ParticleNet _{exp}	0.840	0.9116	39.8 ± 0.2	98.6 ± 1.3
rPCN _{exp}	—	0.9081	38.6 ± 0.5	—
ParT _{exp}	0.840	0.9121	41.3 ± 0.3	101.2 ± 1.1
ParticleNet-f.t. _{exp}	0.839	0.9115	40.1 ± 0.2	100.3 ± 1.0
ParT-f.t._{exp}	0.843	0.9151	42.4 ± 0.2	107.9 ± 0.5
PFN _{full}	—	0.9052	37.4 ± 0.7	—
ABCNet _{full}	0.840	0.9126	42.6 ± 0.4	118.4 ± 1.5
PCT _{full}	0.841	0.9140	43.2 ± 0.7	118.0 ± 2.2
LorentzNet _{full}	0.844	0.9156	42.4 ± 0.4	110.2 ± 1.3
ParT _{full}	0.849	0.9203	47.9 ± 0.5	129.5 ± 0.9
ParT-f.t._{full}	0.852	0.9230	50.6 ± 0.2	138.7 ± 1.3

SUMMARY & OUTLOOK

SUMMARY & OUTLOOK

- The rise of deep learning has brought lots of progress in jet physics
 - new approaches, particularly graph neural networks, significantly improved the jet tagging performance
 - leads to substantial increase in the physics reach at the LHC
- Towards the future
 - pushing the performance even further
 - new (physics-inspired) architectures: graph networks, Transformers, ...
 - training strategy: end-to-end training => supervised pre-training => un-/semi-/self-supervised training (on real data)?
 - increasing the robustness and controlling the systematics
 - robust architectures and training schemes
 - improvements in the simulation
 - beyond classification:
 - representation learning? anomaly detection? ...
 - JetClass: a large-scale open dataset to explore
- **Your innovation and creativity can make a big difference!**