

Particle Transformer for Jet Tagging

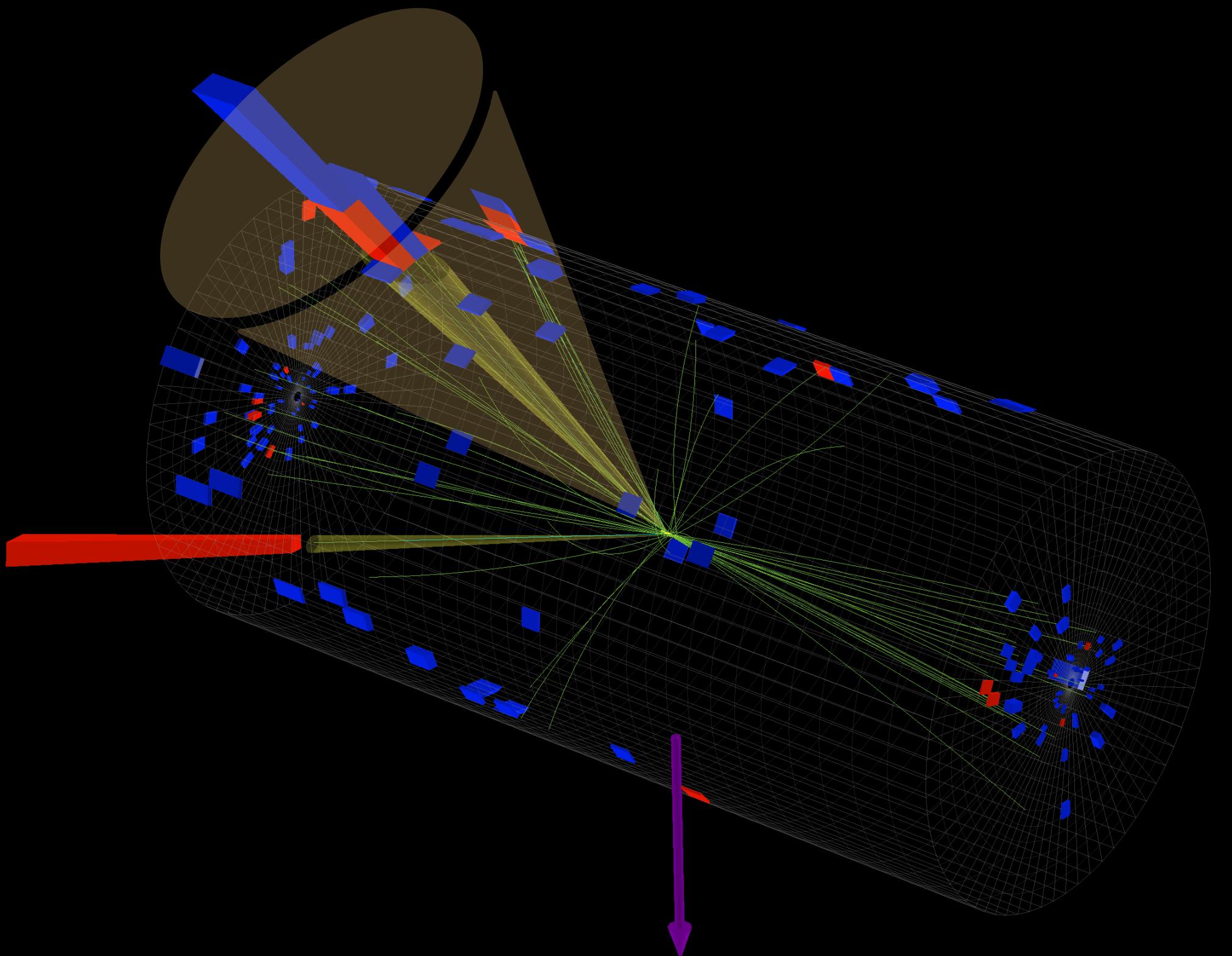
Huilin Qu

CERN

Congqiao Li

Peking University

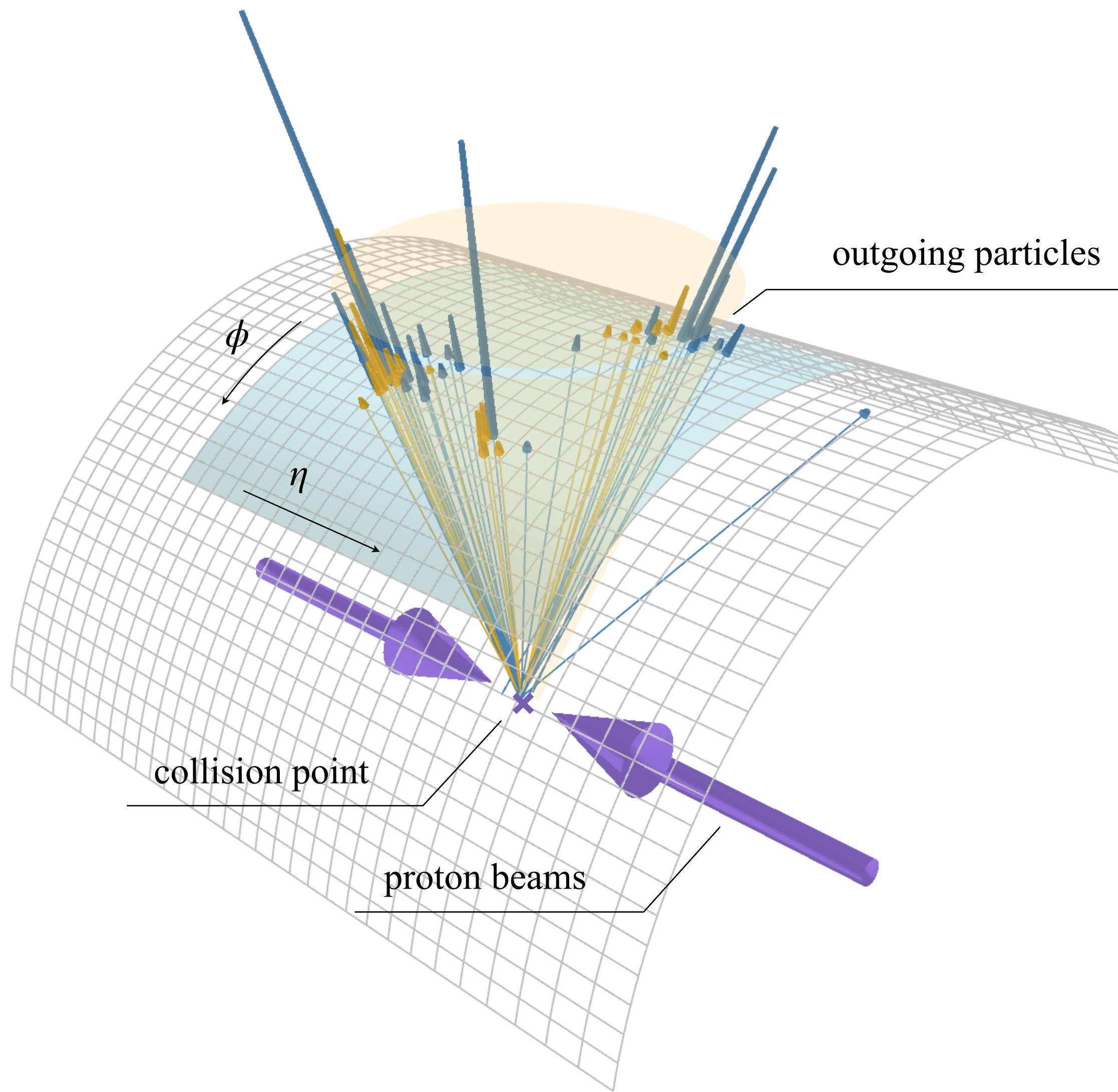
Sitian Qian



The Thirty-ninth International Conference on Machine Learning

July 17 - 23, 2022

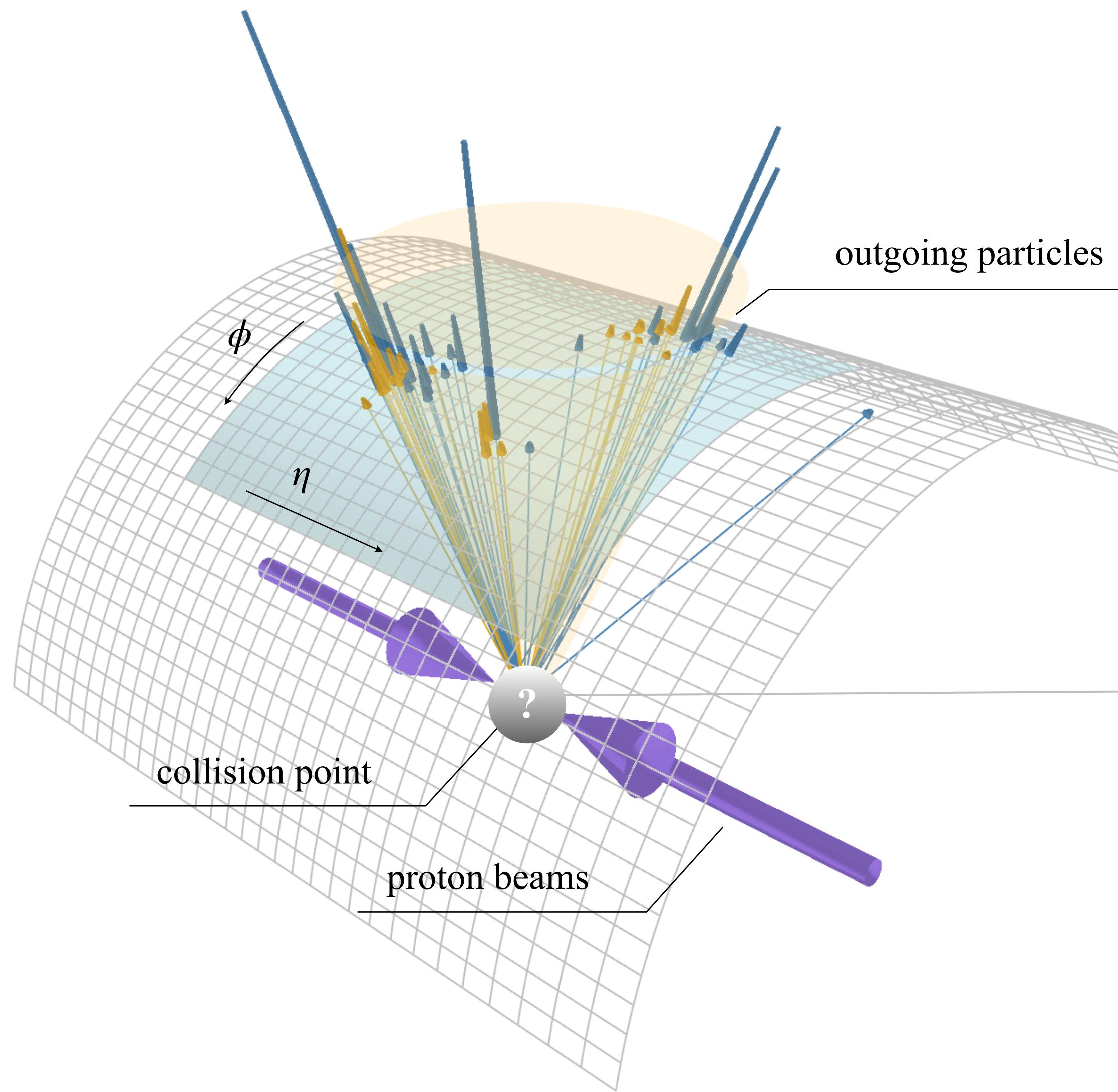
JET TAGGING — AN INTRODUCTION



What is a jet?

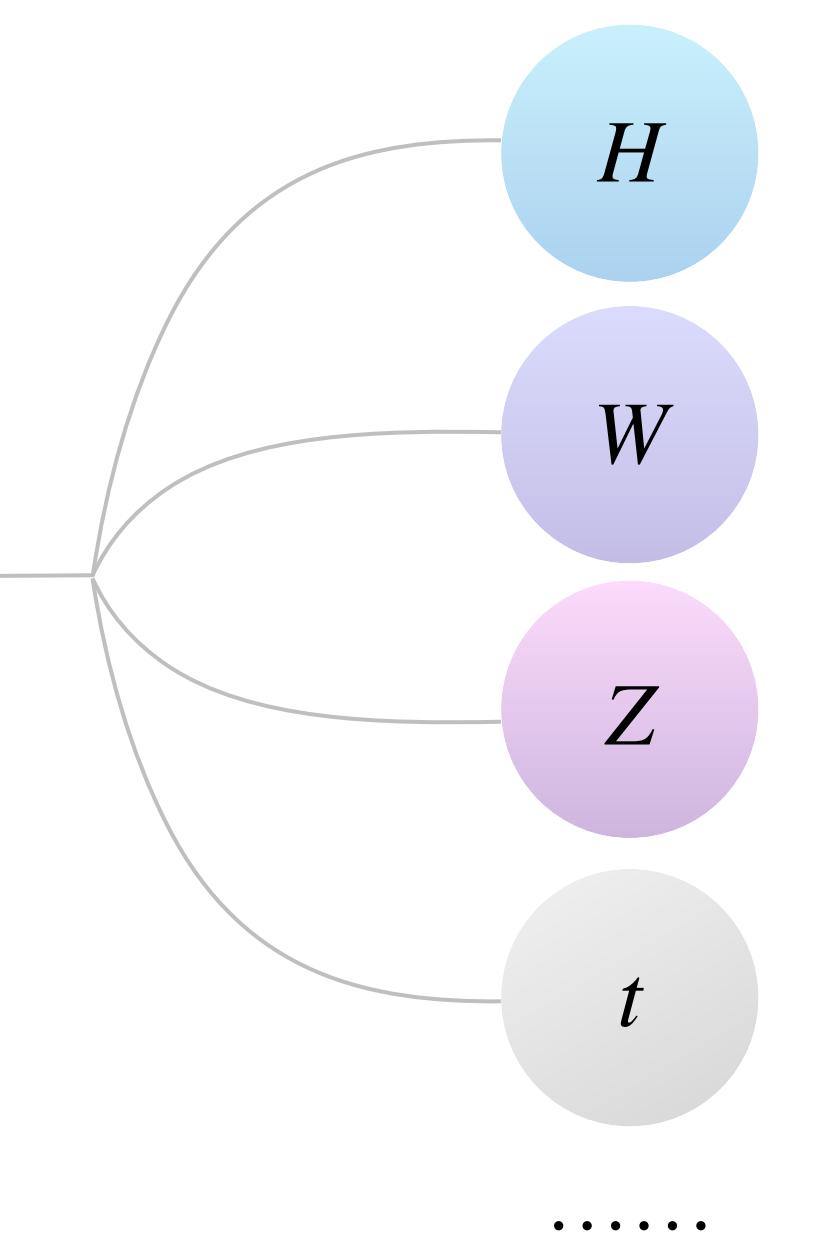
- a collimated spray of outgoing particles

JET TAGGING — AN INTRODUCTION

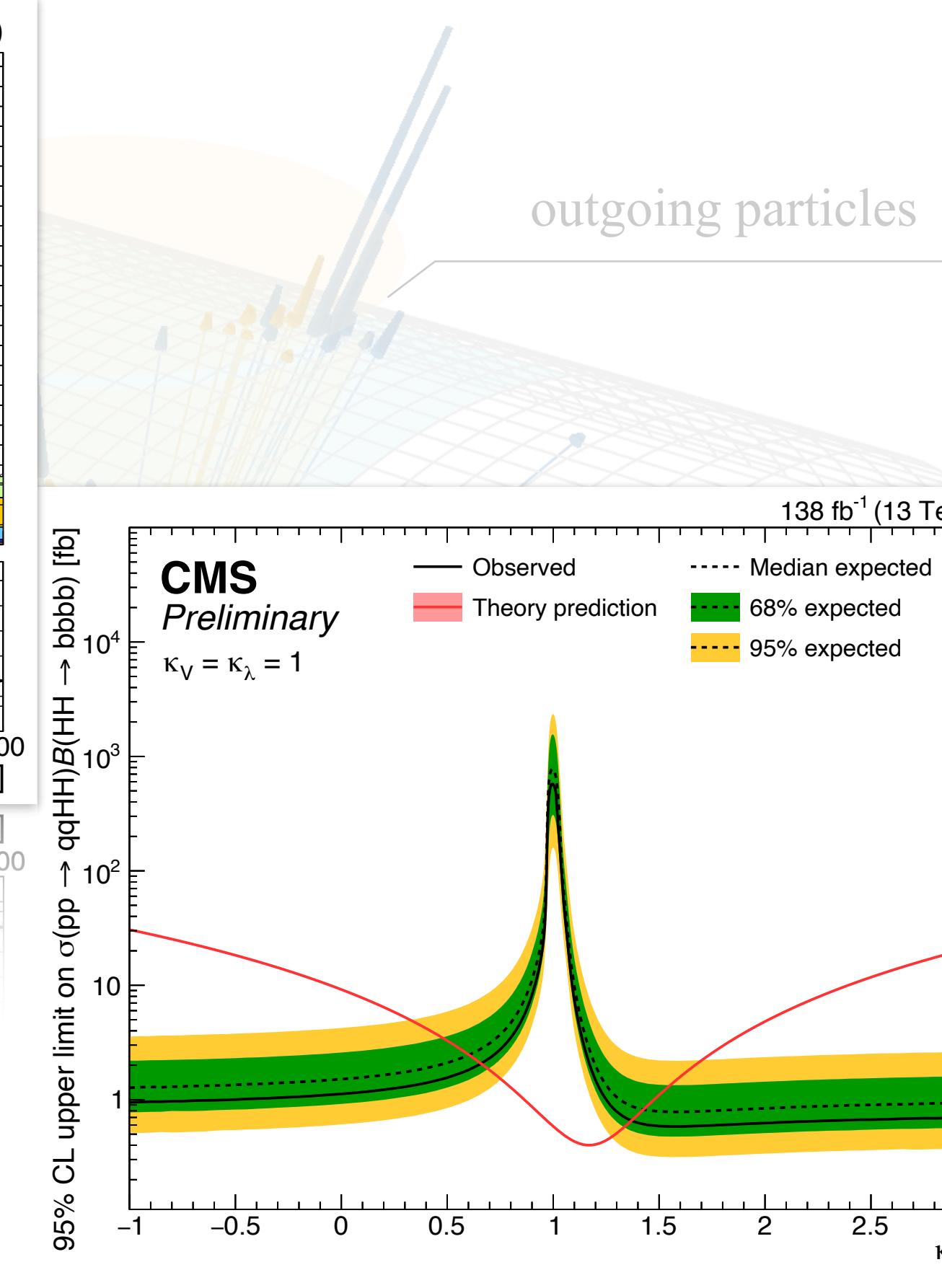
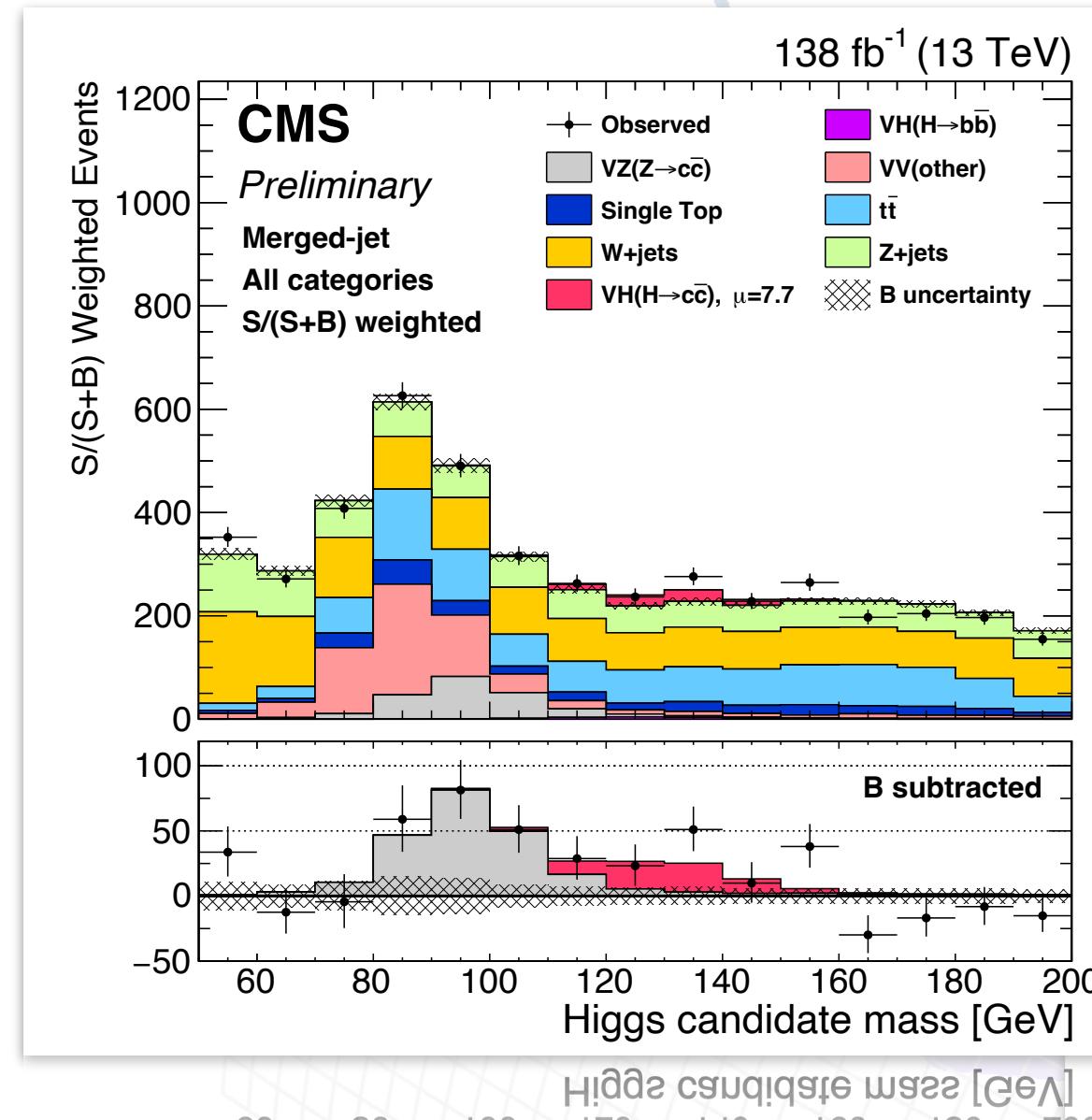


Jet Tagging

- What type of particle initiates the jet?

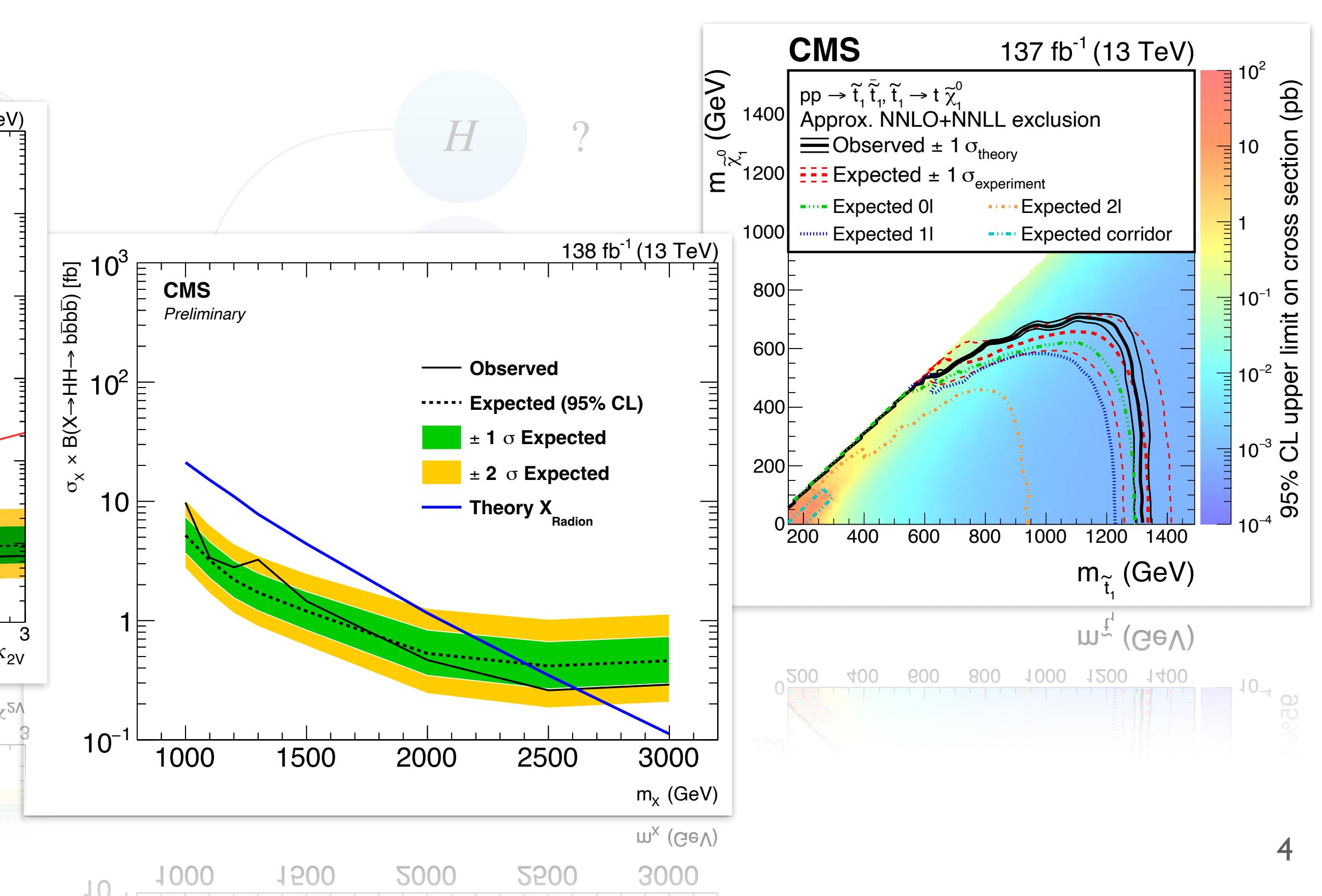


JET TAGGING — AN INTRODUCTION



Why Jets?

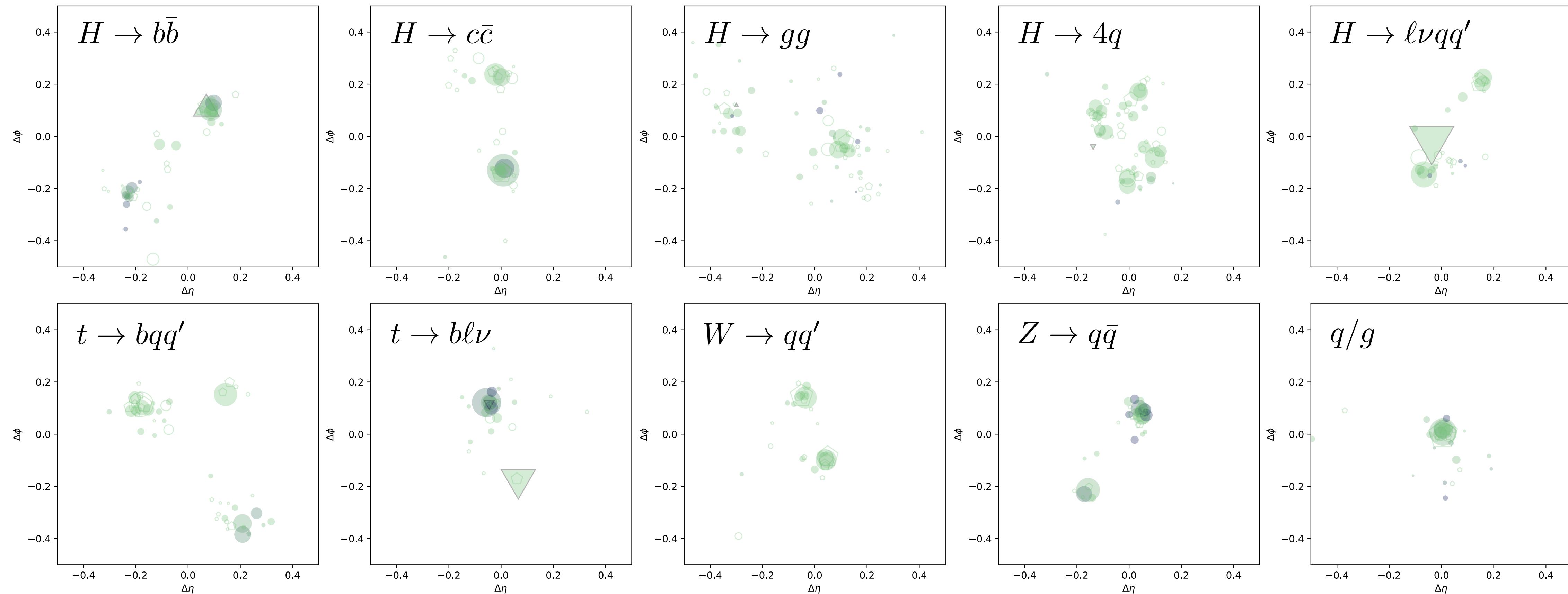
- Powerful handle to search for new phenomena



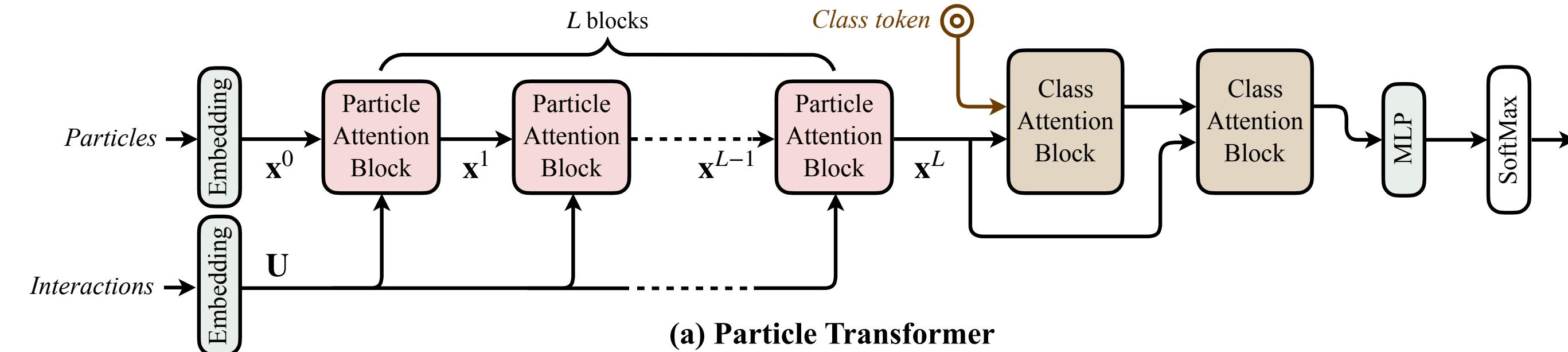
NEED FOR A LARGE DATASET

- **JETCLASS**: a new large and comprehensive jet dataset
 - 100M jets for training: ~two orders of magnitude larger than existing public datasets
 - 10 classes: several unexplored scenarios (e.g., $H \rightarrow WW^* \rightarrow 4q$, $H \rightarrow WW^* \rightarrow \ell\nu qq$, etc.)

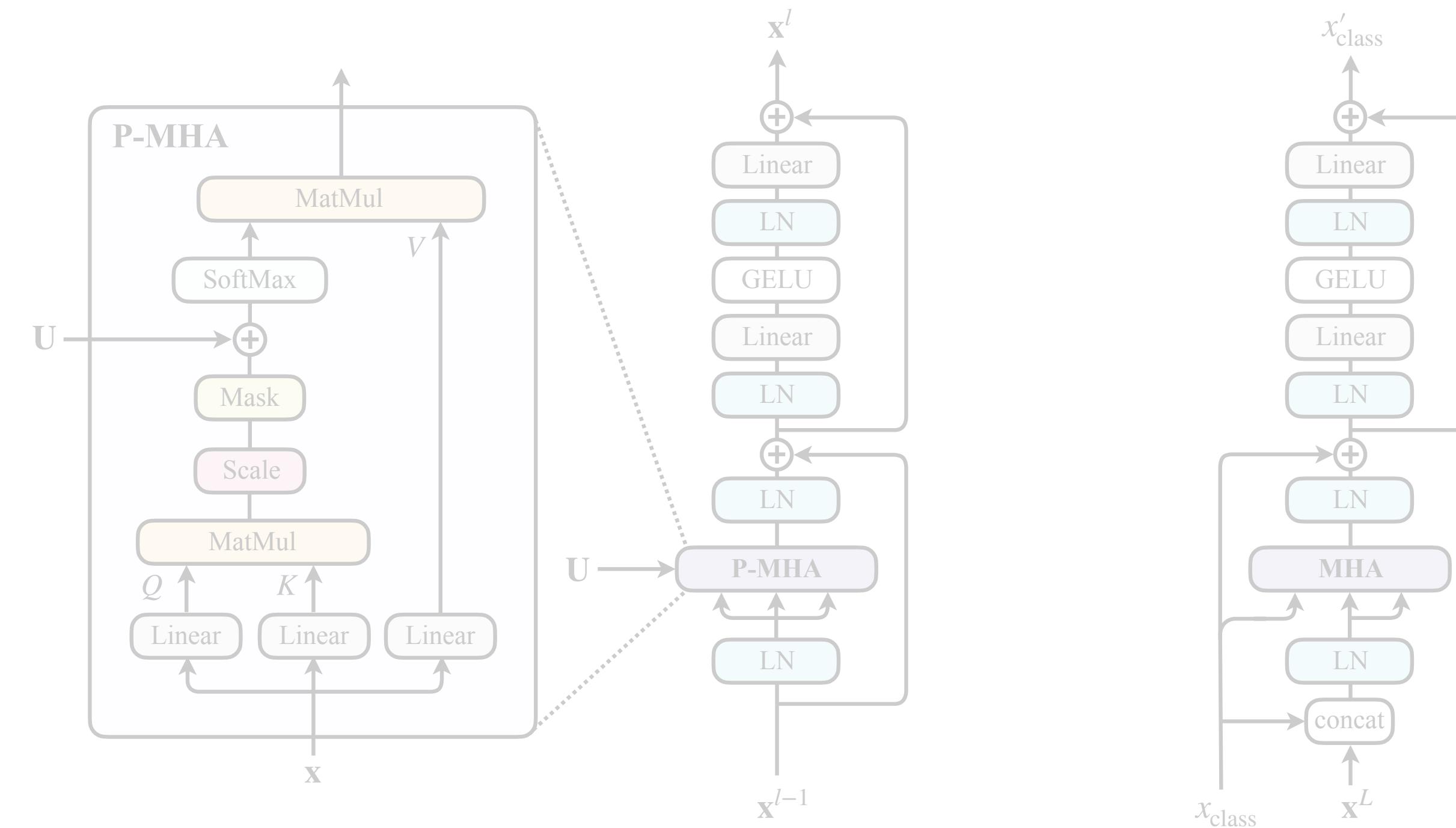
Simulated w/ *MadGraph* +
Pythia + *Delphes*



PARTICLE TRANSFORMER (PART)



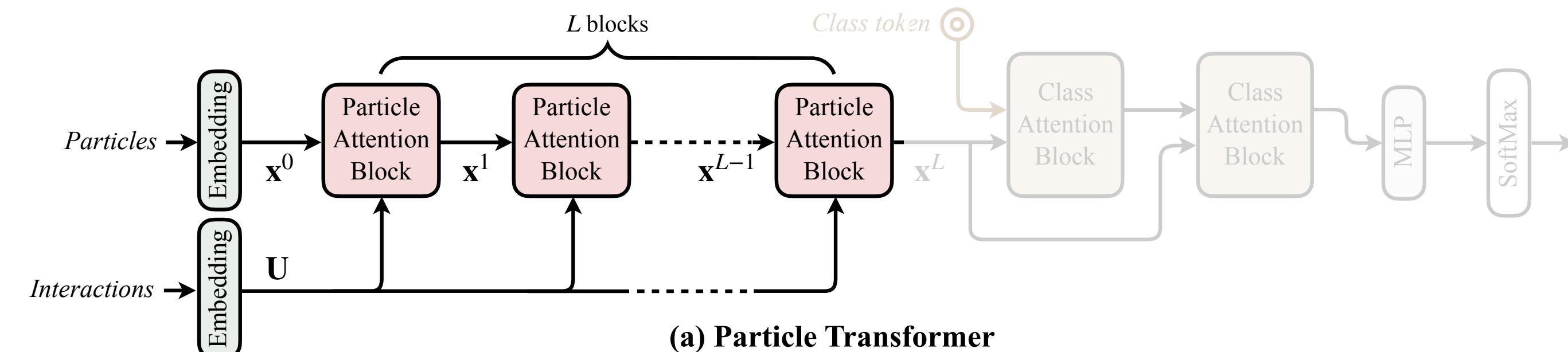
(a) Particle Transformer



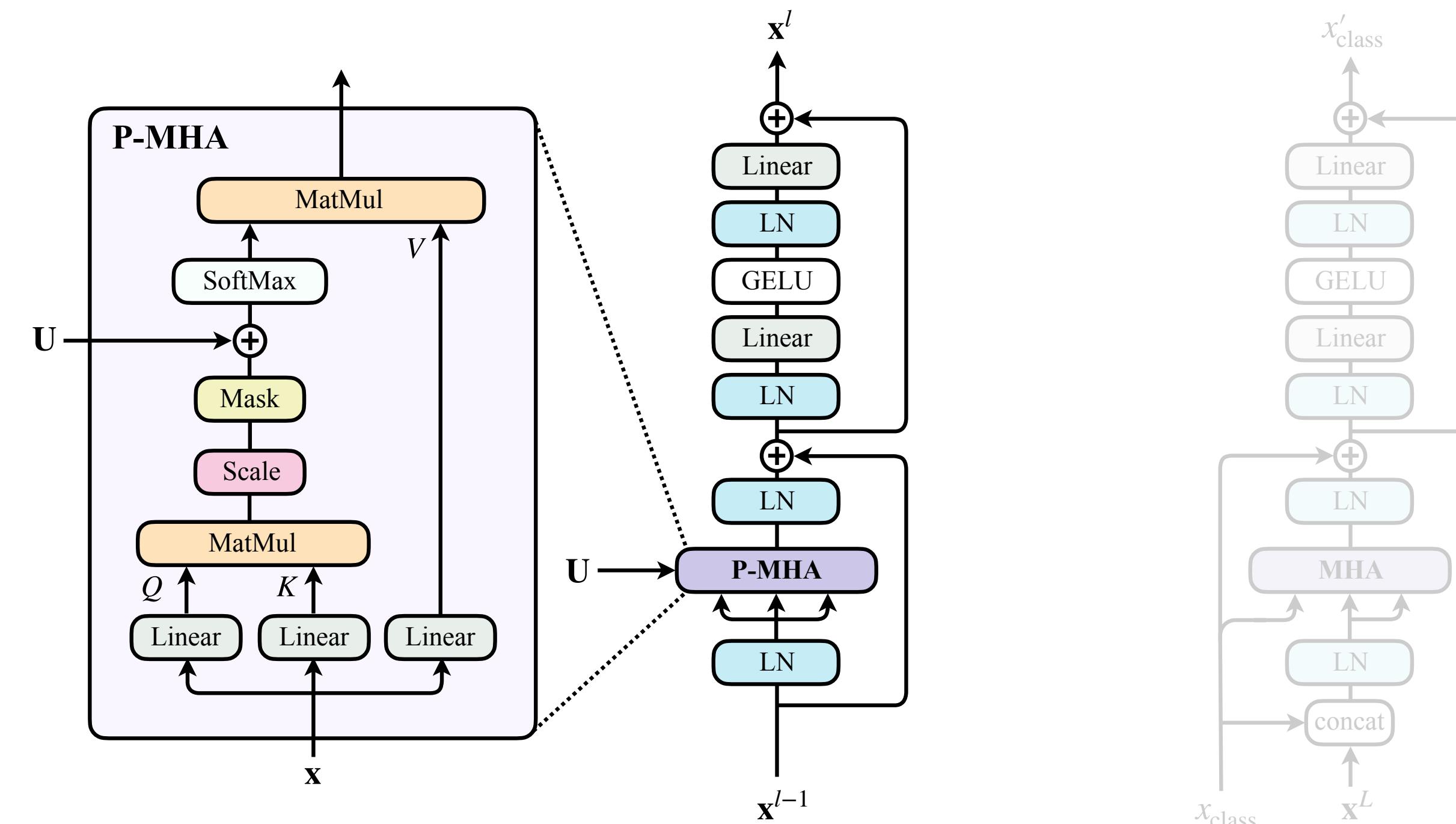
(b) Particle Attention Block

(c) Class Attention Block

PARTICLE TRANSFORMER (PART)



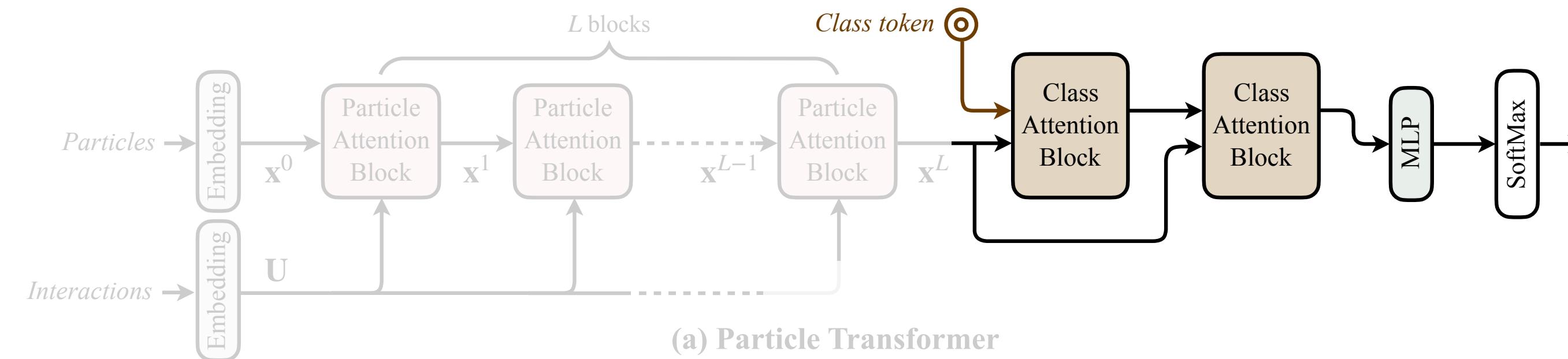
(a) Particle Transformer



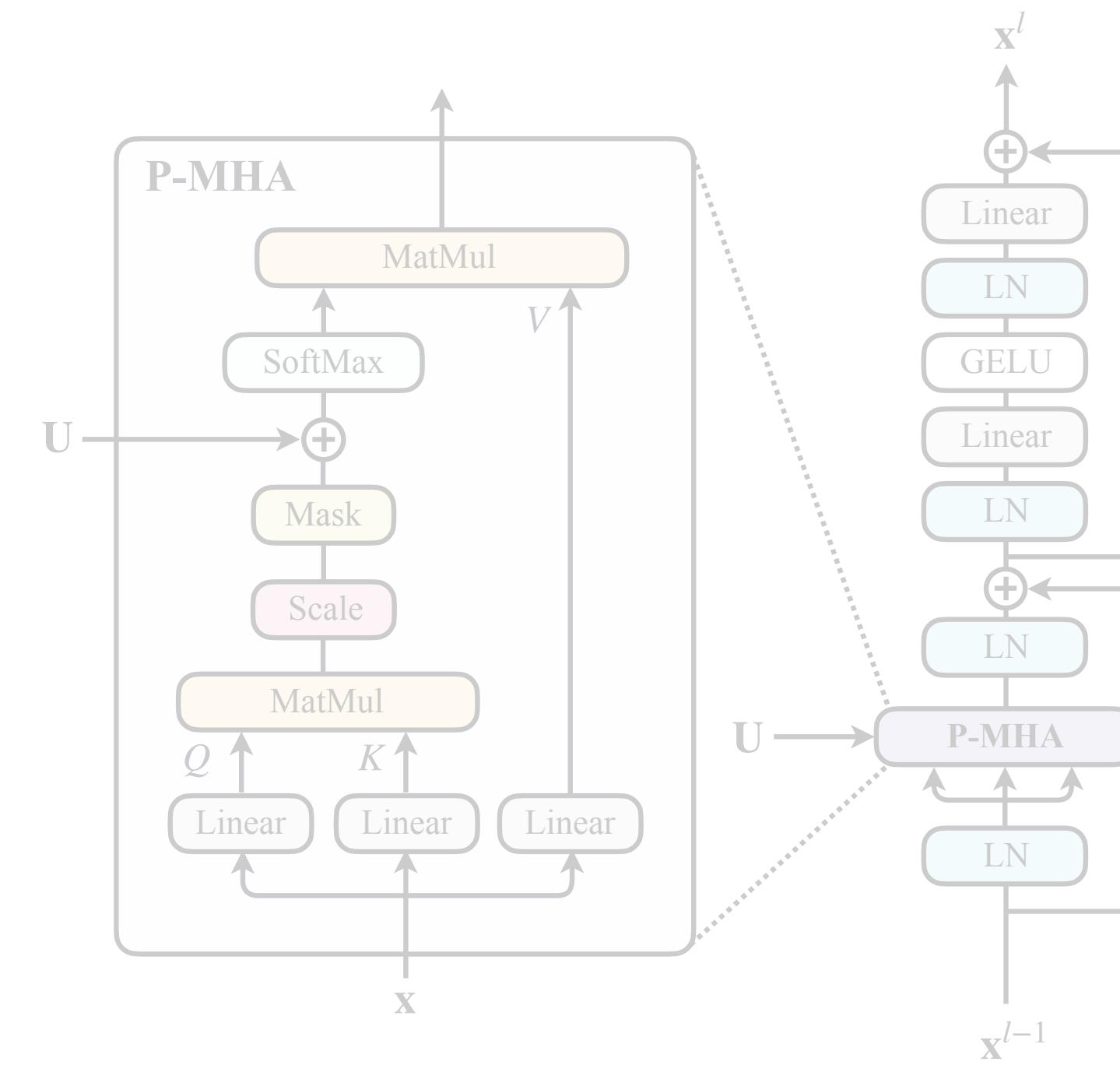
(b) Particle Attention Block

(c) Class Attention Block

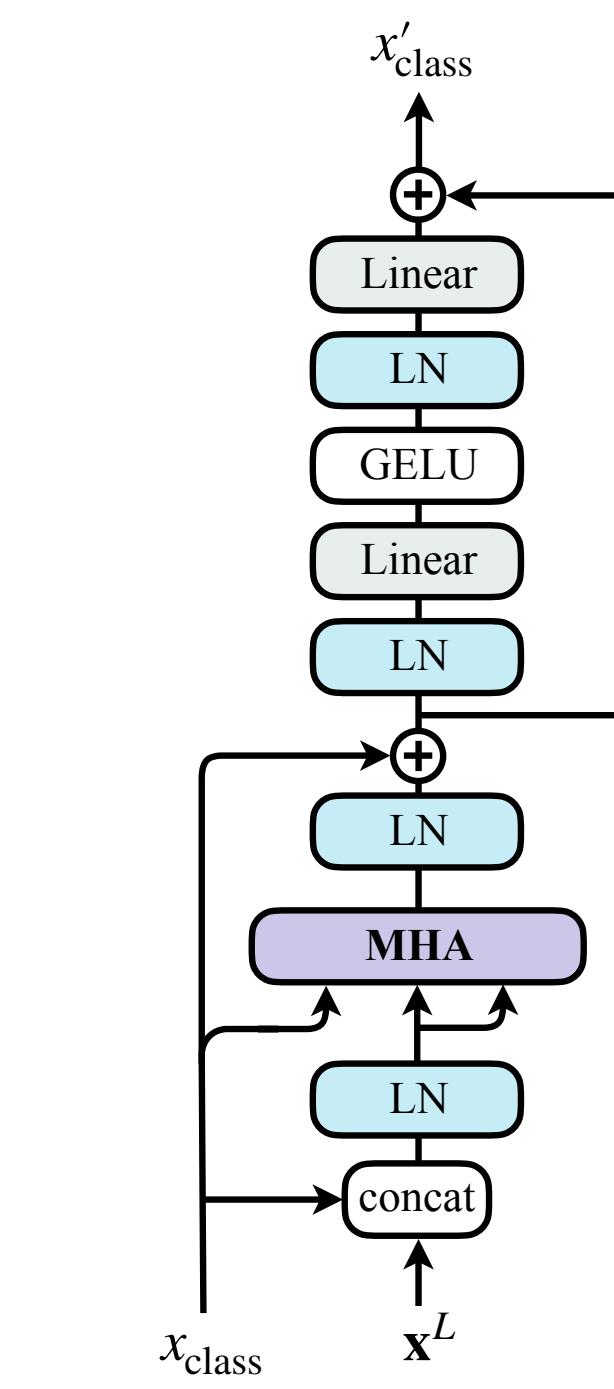
PARTICLE TRANSFORMER (PART)



(a) Particle Transformer

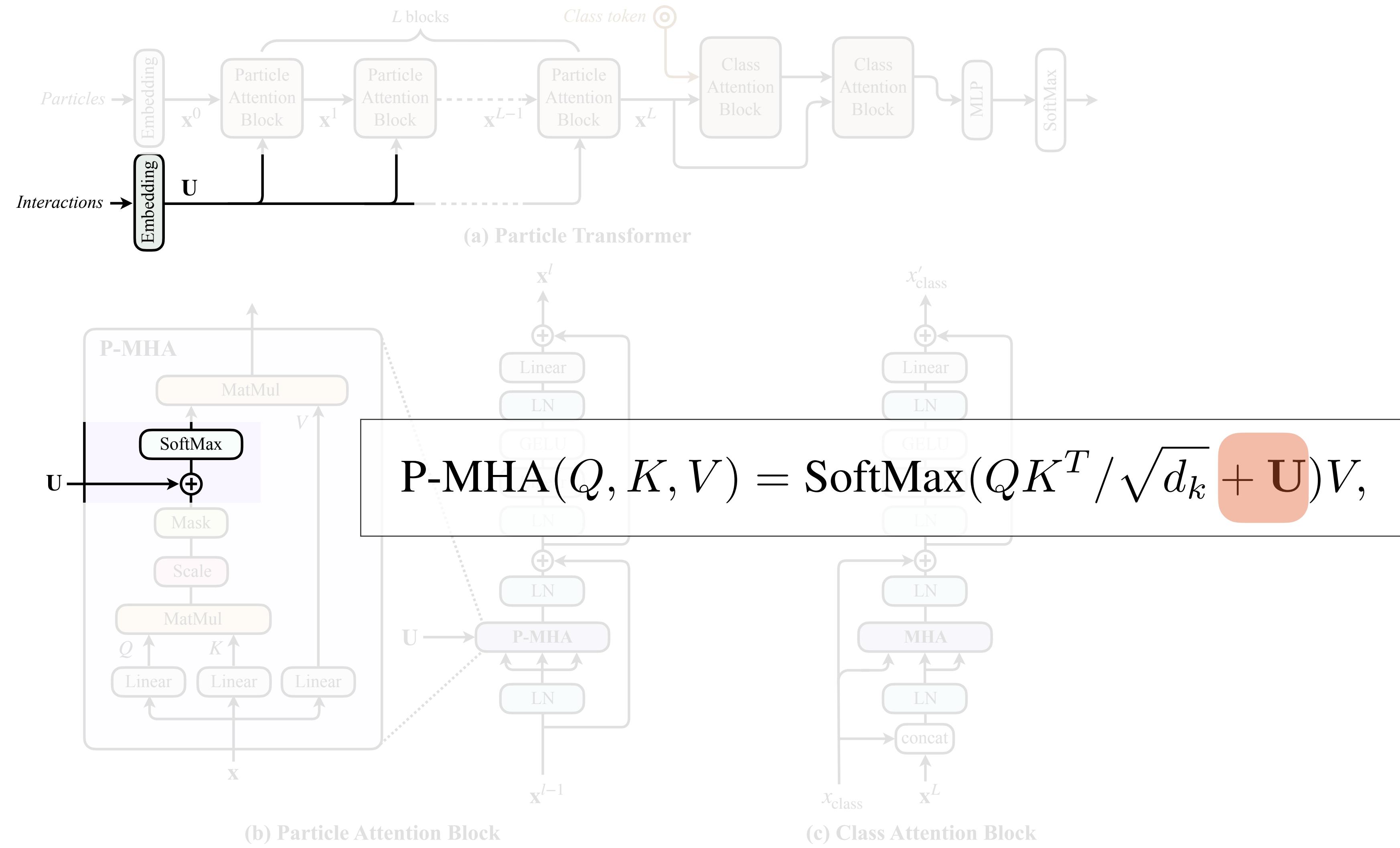


(b) Particle Attention Block



(c) Class Attention Block

PARTICLE TRANSFORMER (PART)



PERFORMANCE ON JETCLASS DATASET

	All classes		$H \rightarrow b\bar{b}$	$H \rightarrow c\bar{c}$	$H \rightarrow gg$	$H \rightarrow 4q$	$H \rightarrow \ell\nu qq'$	$t \rightarrow bqq'$	$t \rightarrow b\ell\nu$	$W \rightarrow qq'$	$Z \rightarrow q\bar{q}$
	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

- 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%})

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

PERFORMANCE ON JETCLASS DATASET

	All classes		$H \rightarrow b\bar{b}$	$H \rightarrow c\bar{c}$	$H \rightarrow gg$	$H \rightarrow 4q$	$H \rightarrow \ell\nu qq'$	$t \rightarrow bqq'$	$t \rightarrow b\ell\nu$	$W \rightarrow qq'$	$Z \rightarrow q\bar{q}$
	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

- 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%})

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

- 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!**

PRE-TRAINING + FINE-TUNING

- The large-scale JETCLASS dataset enables new training paradigm
 - (supervised) pre-training on JETCLASS & fine-tuning on downstream tasks
 - significantly outperforms existing models

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

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

Quark-gluon tagging benchmark (~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 \pm 0.4	—
ParticleNet _{exp}	0.840	0.9116	39.8 \pm 0.2	98.6 \pm 1.3
rPCN _{exp}	—	0.9081	38.6 \pm 0.5	—
ParT _{exp}	0.840	0.9121	41.3 \pm 0.3	101.2 \pm 1.1
ParticleNet-f.t. _{exp}	0.839	0.9115	40.1 \pm 0.2	100.3 \pm 1.0
ParT-f.t._{exp}	0.843	0.9151	42.4 \pm 0.2	107.9 \pm 0.5
PFN _{full}	—	0.9052	37.4 \pm 0.7	—
ABCNet _{full}	0.840	0.9126	42.6 \pm 0.4	118.4 \pm 1.5
PCT _{full}	0.841	0.9140	43.2 \pm 0.7	118.0 \pm 2.2
LorentzNet _{full}	0.844	0.9156	42.4 \pm 0.4	110.2 \pm 1.3
ParT _{full}	0.849	0.9203	47.9 \pm 0.5	129.5 \pm 0.9
ParT-f.t._{full}	0.852	0.9230	50.6 \pm 0.2	138.7 \pm 1.3

SUMMARY

- **JETCLASS**: large-scale open dataset for deep-learning research in particle physics
- **Particle Transformer**: new architecture for jet tagging with substantially improved performance

JETCLASS: More possibilities ahead

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

