Attention to the strengths of physics interactions

Enhanced Deep Learning Event Classification for Particle Physics Experiments

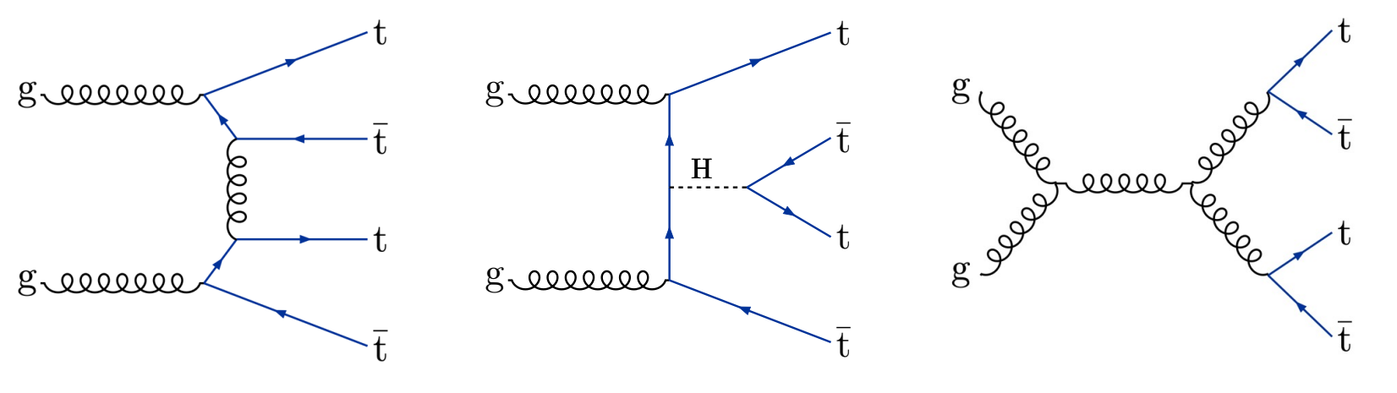
Polina Moskvitina
Sascha Caron, Clara Nellist, Roberto Ruiz de Austri, Rob Verheyen, Zhongyi Zhang

arXiv:2211.05143

The four-top-quarks and $t\bar{t}H$ production at LHC

Production of **four top quarks** is very rare

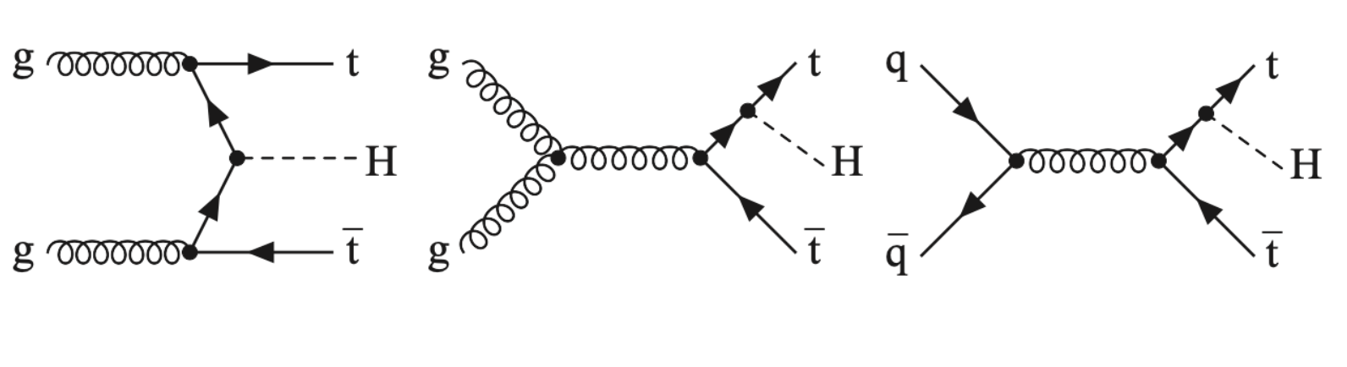
- **NLO QCD:** $\sigma(t\bar{t}t\bar{t}) = 12\ \text{fb} \pm 20\%$ [JHEP02(2018)031]
- **NLO+NLL:** $\sigma(t\bar{t}t\bar{t}) = 13.4\ \text{fb} \pm 11\%$ [arXiv:2212.03259]



Examples of Feynman diagrams for SM $t\bar{t}t\bar{t}$ production at leading order in QCD and via an off-shell Higgs boson mediator

First observation of $t\bar{t}t\bar{t}$ production with an observed (expected) significance of **6.1σ (4.3σ)** with **GNN** by **ATLAS** [Eur. Phys. J. C 83, 496 (2023)]
5.6σ (4.9σ) with **BDT** by **CMS** [Phys. Lett. B 847 (2023) 138290]

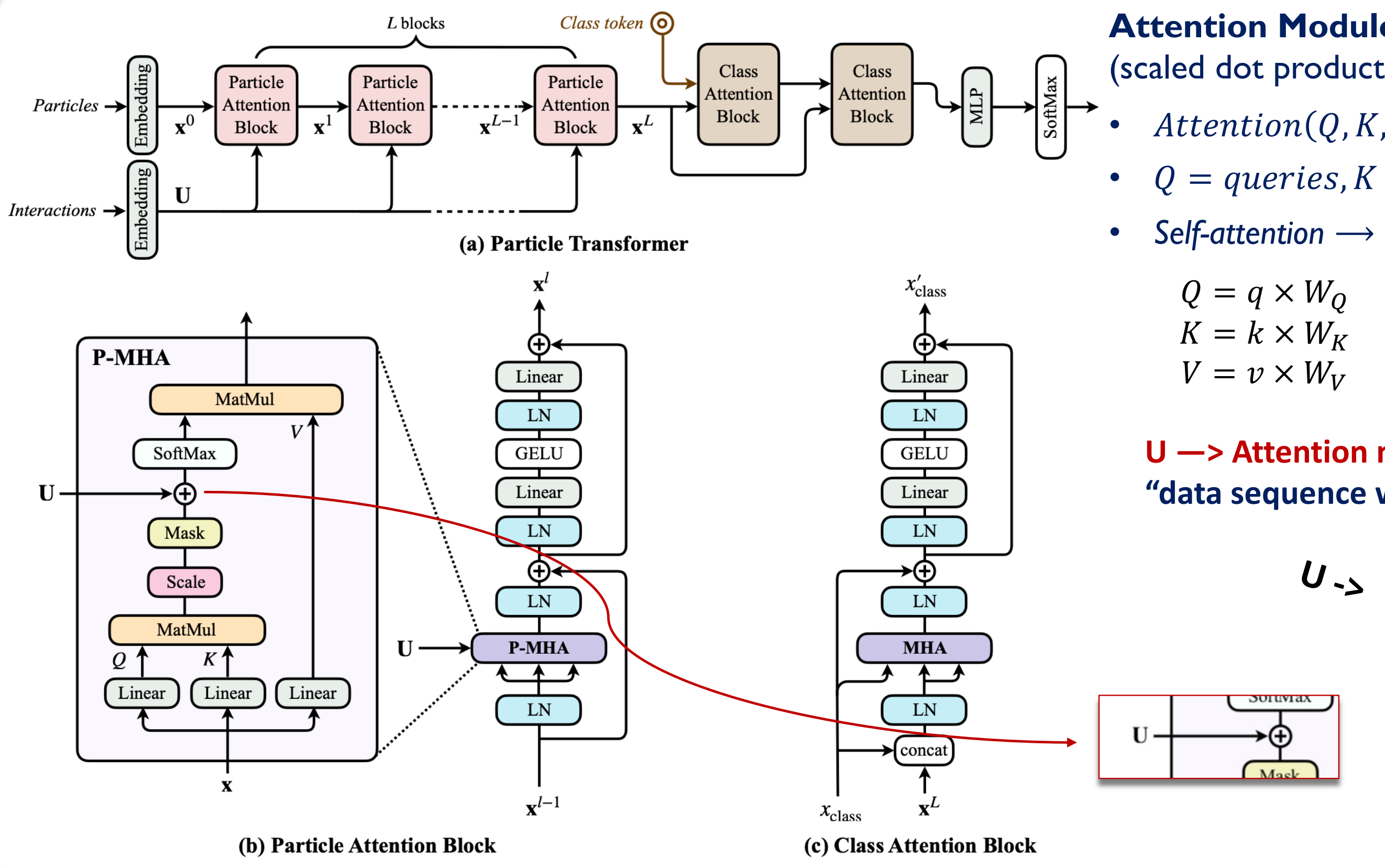
The **Top-top-Higgs**
has a small cross section (1/100 ggF)
$\sigma(t\bar{t}H) \sim 0.507\ \text{pb}$



Example tree-level Feynman diagrams for the pp → $t\bar{t}H$

Observation of $t\bar{t}H$ production
6.3σ (5.1σ) with **BDT** by **ATLAS** [Phys. Lett. B 784 (2018) 173]
5.2σ (4.2σ) with **BDT** by **CMS** [Phys. Rev. Lett. 120, 231801]

Transformers



Attention Modules
(scaled dot product attention):

- $Attention(Q,K,V) = SoftMax\left(\frac{QK^T}{\sqrt{d}} + \boldsymbol{U}\right)V$
- $Q = queries, K = keys, V = values$
- *Self-attention* $\longrightarrow Q = K = V$

$$Q = q \times W_Q$$
$$K = k \times W_K$$
$$V = v \times W_V$$

Attention is All You Need!

U —> Attention matrix —> correlation of "data sequence with data sequence"

***U* ->** **Pairwise Features + SM interaction matrix** **(attention matrix)**

To show the interaction strength based on the SM coupling constants

Pairwise features

Include pairwise features in **Par**ticle **T**ransformer through a trainable embedding $\boldsymbol{U_{ij}}$ for particles i and j

ParT uses high level features for better performance

1. $\Delta = \sqrt{(y_a - y_b)^2 + (\phi_a + \phi_b)^2}$
2. $k_t = min(p_{T,a}, p_{T,b})\Delta$
3. $z = min(p_{T,a}, p_{T,b})/(p_{T,a}, p_{T,b})$
4. $m^2 = (E_a + E_b)^2 - \|p_a + p_b\|^2$

- These were also tested in **LightGBM**

We end up using :

$\boldsymbol{m_{ij}}, \Delta\boldsymbol{R_{ij}}$ and dynamically calculated **coupling constants** of interaction terms
(i.e. a feature that is coupling constant when i and j are components of a **SM** current, and 0 otherwise)

The four-top decays and Background composition

Simulated pp Collisions at $\sqrt{S} = 13$ TeV

The most sensitive channel for **four-top** is:

- **Multilepton final state:**
  2 Leptons Same Sign and 3 Leptons (2LSS/3L),
  13% branching ration, highest sensitivity – observation

	jets	b-jets	e^-	e^+	μ^-	μ^+	γ	$N_{\max}$
FCN, BDT	4	4	1	1	1	1		12
CNN, PN, ParT	no limits							18

$N_{\max}$ – the maximum number of objects in an event

$$\text{event ID; process ID; weight; } \not{E}_T;\ \phi_{\not{E}_T};$$
$$\text{obj}_1, E_1, p_{T_1}, \eta_1, \phi_1;\ \text{obj}_2, E_2, p_{T_2}, \eta_2, \phi_2; \ldots$$

- All other kinematic variables can be calculated from four-vectors

Signal region:
≥ 6 jets ≥ 2b-jets and $\mathbf{H_T}$ ≥500 GeV

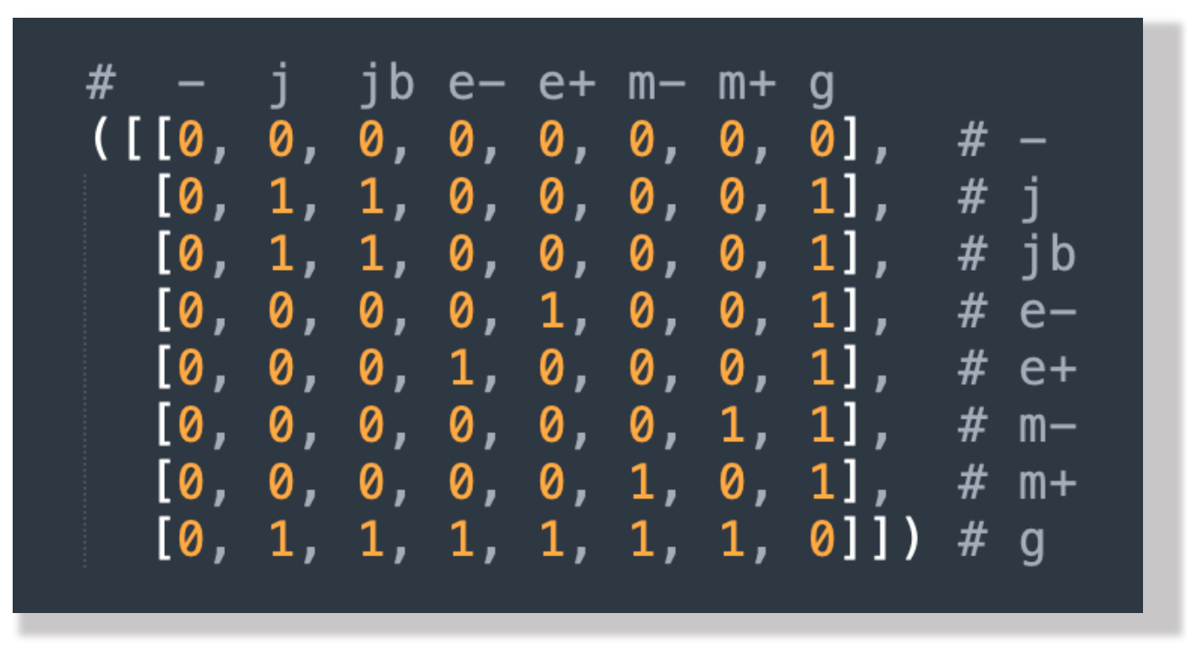
Signal process:
- $t\bar{t}t\bar{t}$

Physical backgrounds:
- $t\bar{t}Z$, $\boxed{t\bar{t}H,}$ $t\bar{t}W$, $t\bar{t}WW$

Used for a second analysis as a Signal

The energy dependence of the coupling constants

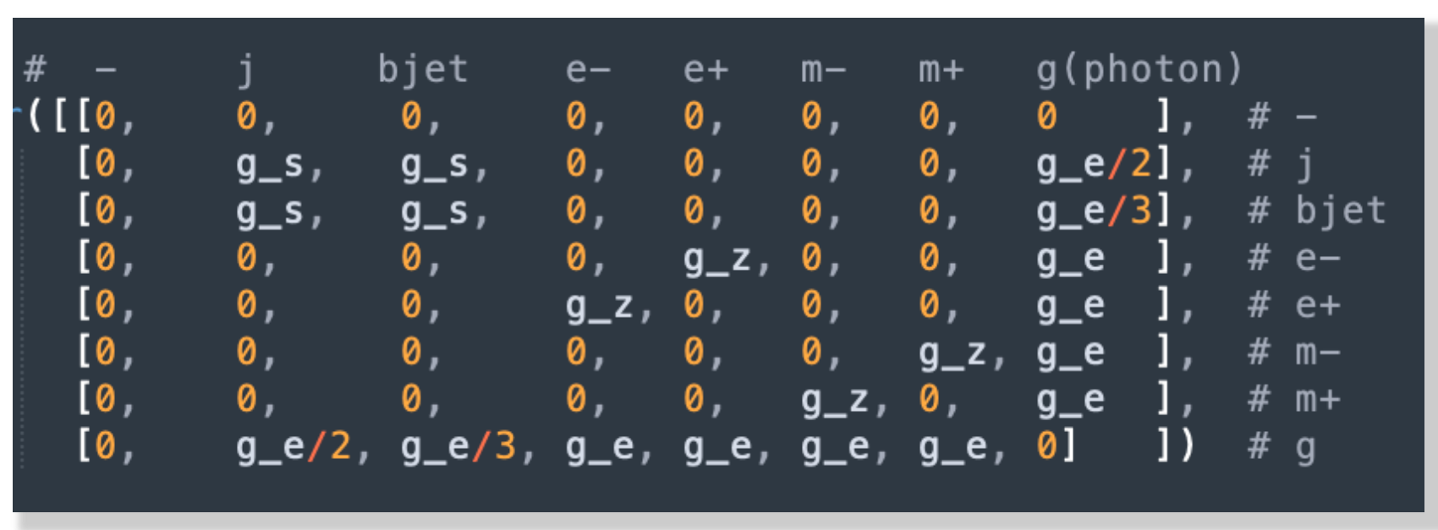
Matrix [1] – SM ids



```
#  -  j  jb e- e+ m- m+ g
([[0, 0, 0, 0, 0, 0, 0, 0],  # -
  [0, 1, 1, 0, 0, 0, 0, 1],  # j
  [0, 1, 1, 0, 0, 0, 0, 1],  # jb
  [0, 0, 0, 0, 1, 0, 0, 1],  # e-
  [0, 0, 0, 1, 0, 0, 0, 1],  # e+
  [0, 0, 0, 0, 0, 0, 1, 1],  # m-
  [0, 0, 0, 0, 0, 1, 0, 1],  # m+
  [0, 1, 1, 1, 1, 1, 1, 0]]) # g
```

- '1' indicates an interaction possible at **LO** in the **SM**
- '0' indicates interactions that only appear at higher orders

Matrix [2] – SM const



```
#  -     j       bjet    e-    e+    m-    m+    g(photon)
([[0,    0,      0,      0,    0,    0,    0,    0    ],  # -
  [0,    g_s,    g_s,    0,    0,    0,    0,    g_e/2],  # j
  [0,    g_s,    g_s,    0,    0,    0,    0,    g_e/3],  # bjet
  [0,    0,      0,      0,    g_z,  0,    0,    g_e  ],  # e-
  [0,    0,      0,      g_z,  0,    0,    0,    g_e  ],  # e+
  [0,    0,      0,      0,    0,    0,    g_z,  g_e  ],  # m-
  [0,    0,      0,      0,    0,    g_z,  0,    g_e  ],  # m+
  [0,    g_e/2,  g_e/3,  g_e,  g_e,  g_e,  g_e,  0]    ]) # g
```

- $g_Z = 0.758$ for the weak force for leptons
- $g_s = 1.22$ for the strong force in jet interactions
- $g_e = 0.31$ for the electromagnetic force in photon interactions

Matrix [3] – SM

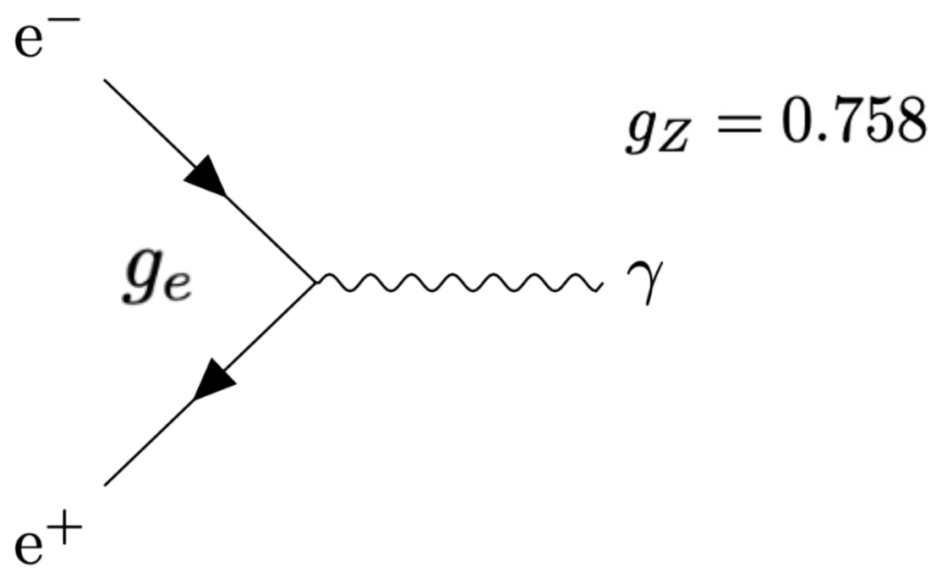
Dynamically calculated **coupling constants** of interaction terms !

$$\alpha(Q^2) = \frac{\alpha(\mu_0^2)}{1 - \frac{n\alpha(\mu_0^2)}{3\pi} \cdot \ln\left(\frac{Q^2}{\mu_0^2}\right)},$$
$$g_e = \sqrt{4\pi\alpha}$$

$$\alpha_s(Q^2) = \frac{\alpha_s(\mu_0^2)}{1 + \frac{\alpha_s(\mu_0^2)(33-2n_f)}{12\pi} \ln\left(\frac{Q^2}{\mu_0^2}\right)},$$
$$g_s = \sqrt{4\pi\alpha_s}$$

$$\mu_0 = 91.1876\ \text{GeV},\ \alpha(\mu_0) = \frac{1}{127.5},$$
$$\alpha_s(\mu_0) = 0.118,\ n_f = 6$$

$$Q^2 = \bar{p}_t^2 = \left(\frac{p_t^i + p_t^j}{2}\right)^2$$



Results for the $t\bar{t}t\bar{t}$ and $t\bar{t}H$ signals

The AUC for both 4 top and top-top-Higgs signal detection

The models containing both the **pairwise features** and the **SM interaction matrix** performs **best.** The **background** can be significantly **reduced** by about **30%** compared to a **PN** (**GNN**)

		PN	$\text{PN}_{\text{int.}}$	$\text{PN}_{\text{int. SMids}}$	$\text{PN}_{\text{int. SM const}}$	$\text{PN}_{\text{int. SM}}$
$t\bar{t}t\bar{t}$	AUC	0.8471(1)	0.8729(0)	0.8725(0)	0.8727(0)	**0.8739(0)**
	$\epsilon_B(\epsilon_S = 0.7)$	0.1758(3)	0.1387(1)	0.1377(0)	0.1384(0)	**0.1369(1)**
	$\epsilon_B(\epsilon_S = 0.3)$	0.0207(0)	0.0182(0)	0.0178(0)	0.0178(0)	**0.0176(0)**
		ParT	$\text{ParT}_{\text{int.}}$	$\text{ParT}_{\text{int. SMids}}$	$\text{ParT}_{\text{int. SM const}}$	$\text{ParT}_{\text{int. SM}}$
$t\bar{t}t\bar{t}$	AUC	0.8404(0)	0.8708(0)	0.8715(0)	0.8717(0)	**0.8732(0)**
	$\epsilon_B(\epsilon_S = 0.7)$	0.1842(3)	0.1394(0)	0.1389(2)	0.1372(1)	**0.1366(0)**
	$\epsilon_B(\epsilon_S = 0.3)$	0.0230(0)	0.0172(0)	0.0180(0)	**0.0167(0)**	0.0169(0)

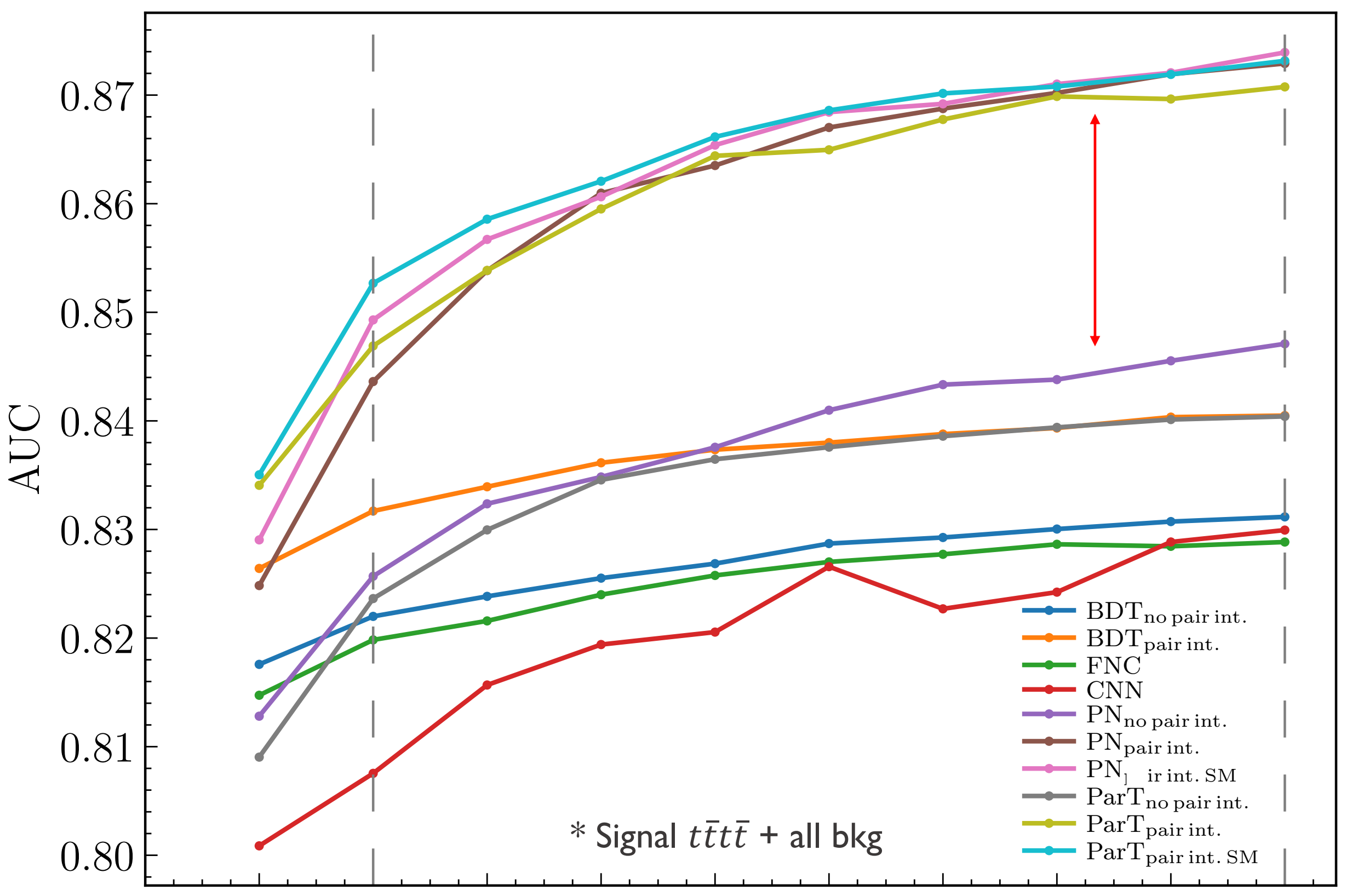
		PN	$\text{PN}_{\text{int.}}$	$\text{PN}_{\text{int. SMids}}$	$\text{PN}_{\text{int. SM const}}$	$\text{PN}_{\text{int. SM}}$
$t\bar{t}+h$	AUC	0.8146(2)	0.8505(0)	0.8489(1)	0.8505(0)	**0.8523(0)**
	$\epsilon_B(\epsilon_S = 0.7)$	0.2292(1)	0.1787(0)	0.1785(1)	0.1764(3)	**0.1733(1)**
	$\epsilon_B(\epsilon_S = 0.3)$	0.0471(1)	0.0345(0)	0.0343(1)	0.0350(0)	**0.0340(0)**
		ParT	$\text{ParT}_{\text{int.}}$	$\text{ParT}_{\text{int. SMids}}$	$\text{ParT}_{\text{int. SM const}}$	$\text{ParT}_{\text{int. SM}}$
$t\bar{t}+h$	AUC	0.8058(1)	0.8507(0)	0.8473(0)	0.8497(0)	**0.8532(0)**
	$\epsilon_B(\epsilon_S = 0.7)$	0.2399(2)	0.1794(1)	0.1836(3)	0.1801(1)	**0.1748(1)**
	$\epsilon_B(\epsilon_S = 0.3)$	0.0502(0)	0.0357(0)	0.0355(1)	0.0367(0)	**0.0351(0)**

We asked the question: → "**Do the models saturate ?**"

List of Models Used

BDT
$\text{BDT}_{\text{int.}}$
FCN
CNN
PN
$\text{PN}_{\text{int.}}$
$\text{PN}_{\text{int. SMids}}$
$\text{PN}_{\text{int. SM const}}$
$\text{PN}_{\text{int. SM}}$
ParT
$\text{ParT}_{\text{int.}}$
$\text{ParT}_{\text{int. SM (FL)}}$
$\text{ParT}_{\text{int. SMids}}$
$\text{ParT}_{\text{int. SM const}}$
$\text{ParT}_{\text{int. SM}}$
$\text{SetT}_{\text{int. SM}}$

The AUC scores as a function of training size



PN and ParT Models
(with the pairwise features + the SM coupling constants)

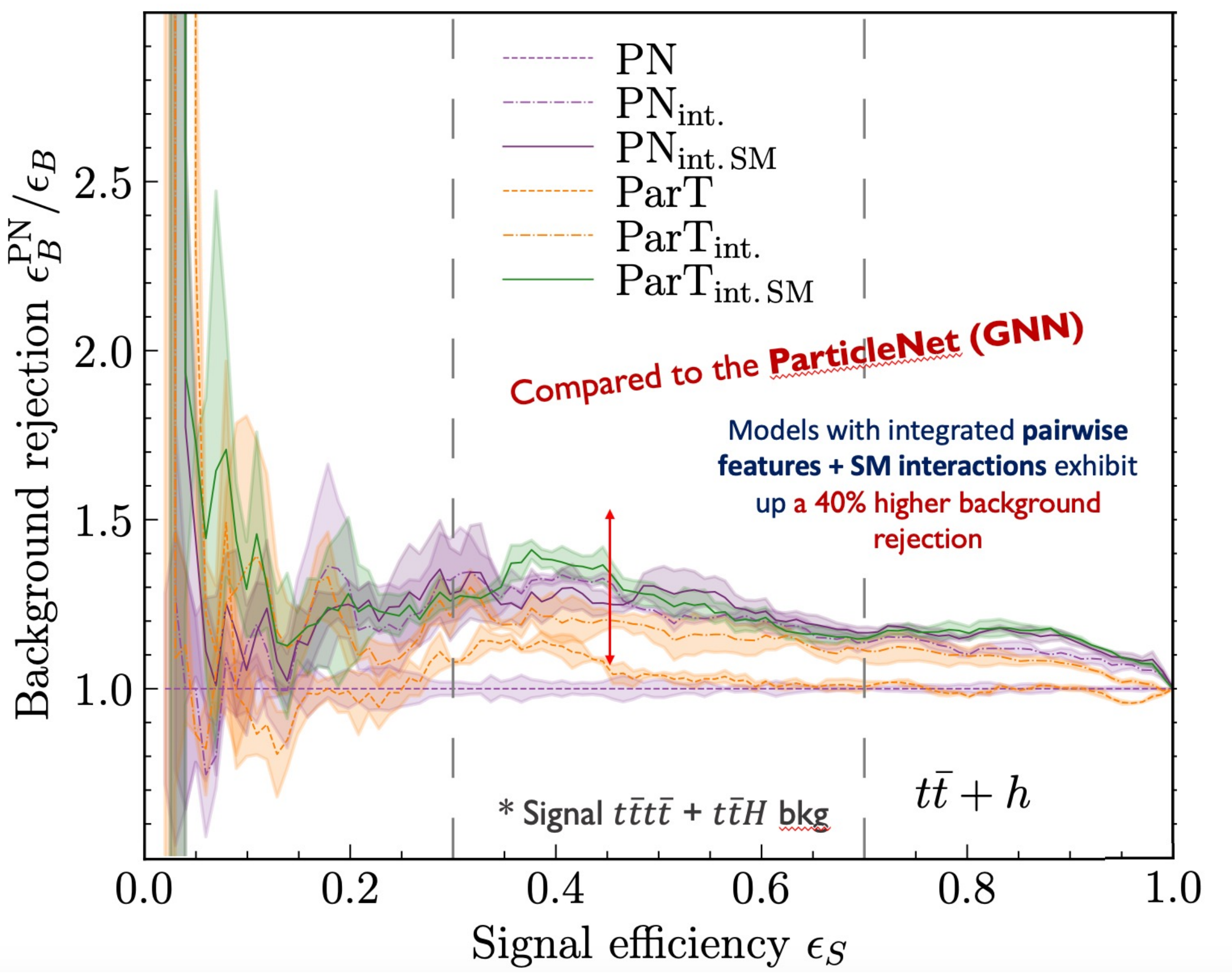
- Shows a steeper increase in **AUC** with fewer data
- Indicate higher data efficiency → less data needed for strong performance

Other Models

- **AUC** scores improve more gradually
- Suggest a requirement for larger datasets to match **PN** and **ParT** performance

… and also: → "**What is the best classifier ?**"

The Signal efficiency VS background rejection



Conclusions

Embedding SM interactions as physical information in NN structures is an important avenue in this field that could lead to more accurate and efficient event classification in particle physics!

- ➢ Enhanced background suppression by **10-40%** compared to baseline **PN** (**GNN**) models
- ➢ Approximately **10%** of this improvement is due to the **SM interaction matrix**
- ➢ ML models show up to **30%** increase in significance vs. baseline

6th Inter-experiment Machine Learning Workshop, 31 January 2024