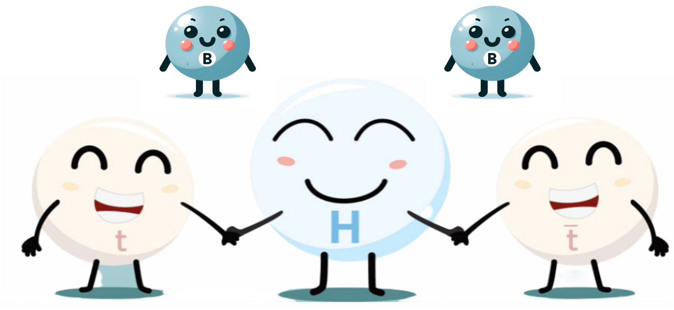
Transformer Neural Networks for the Measurement of $t\bar{t}H$ Production in the $H \to b\bar{b}$ Decay Channel with ATLAS

Chris Scheulen (he/him) on behalf of the ATLAS Collaboration

TOP 2024 Workshop – YSF

2024-09-27

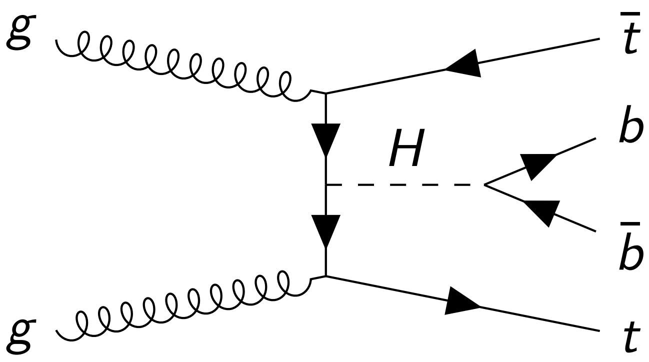


Introduction

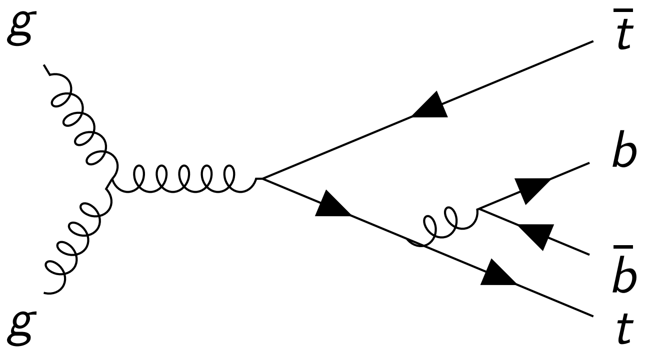
UNIVERSITÉ DE GENÈVE

GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN IN PUBLICA COMMODA SEIT 1737

$t\bar{t}H(b\bar{b})$ signal



$t\bar{t} + b\bar{b}$ background

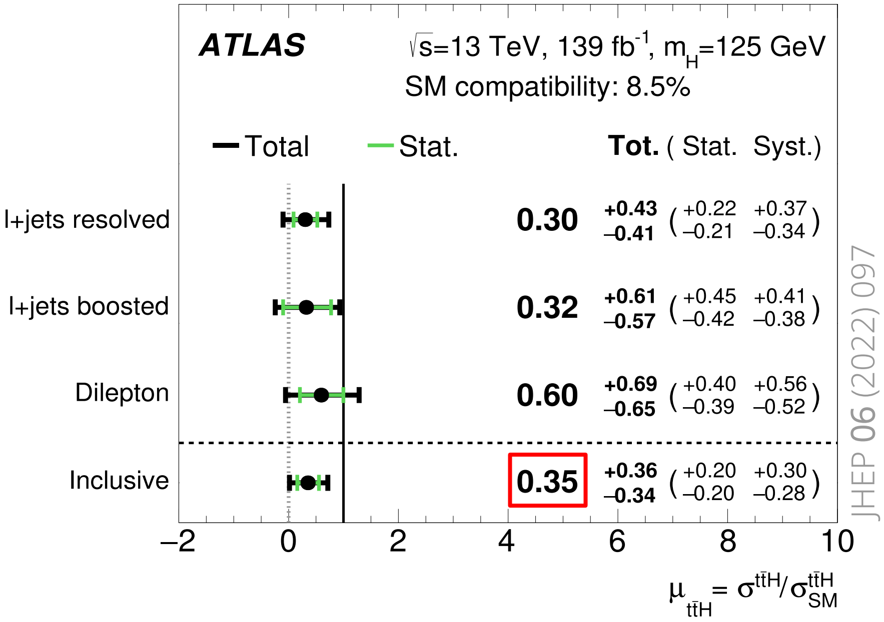


- $t\bar{t}H$ provides direct probe of top Yukawa coupling
- $H \rightarrow b\bar{b}$ adds sensitivity to high Higgs boson p_T region in differential measurements
- Challenging dominant irreducible $t\bar{t} + b\bar{b}$ background
  - ➥ Dedicated systematics studies for ATLAS legacy analysis (ATL-PHYS-PUB-2022-006)
  - ➥ Also dedicated measurements (see Egor's talk on Tuesday, Knut's talk for $t\bar{t} + c\bar{c}$ in ATLAS)

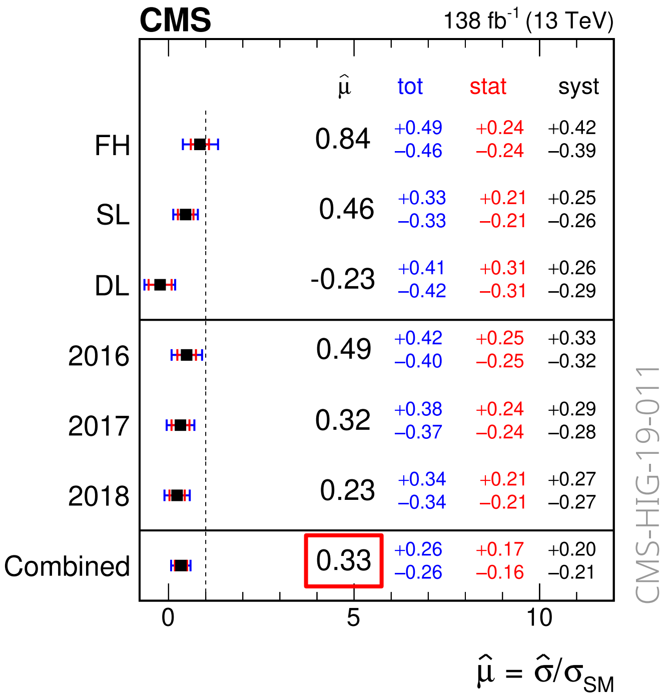
Previous Measurements

UNIVERSITÉ DE GENÈVE
GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN IN PUBLICA COMMODA SEIT 1737

First Full Run 2 ATLAS



Full Run 2 CMS



- Low signal strengths: Compatibility with Standard Model at 8.5% (ATLAS) and 2% (CMS)

ATLAS Run 2 Legacy Analysis

UNIVERSITÉ DE GENÈVE
GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN IN PUBLICA COMMODA SEIT 1737

(arXiv:2407.10904)

- Re-analysis of full ATLAS Run 2 dataset
- Inclusive & differential measurement
- Major updates incorporated into analysis:
  - Improved Object Model
  - Consistent 4FS $t\bar{t} + b\bar{b}$ systematics model
  - Loosened kinematic pre-selection
    - ➥ $t\bar{t}H(b\bar{b})$ acceptance of 6.3% (increase by factor 3)
  - Overhauled event classification & Higgs p_T reco:
    Attention-based Transformers (arXiv:1706.03762) using basic particle information

main focus of this talk

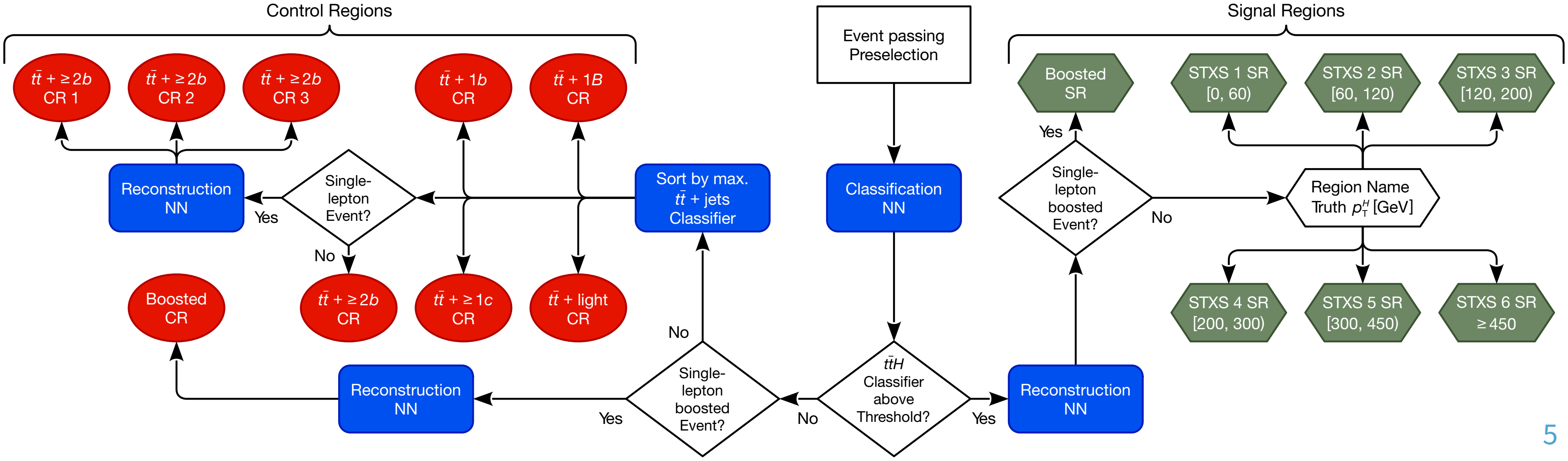
(Also see Matthias' talk on Monday for general top ML)

- Event Selection:
  - Single-lepton resolved:
    ≥ 5 jets, ≥ 3 b-tags
  - Single-lepton boosted:
    ≥ 4 jets, large-radius jet boosted Higgs boson candidate, ≥ 2 additional b-tags
  - Dilepton:
    ≥ 3 jets, ≥ 3 b-tags

Overall Analysis Strategy

UNIVERSITÉ DE GENÈVE
GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN IN PUBLICA COMMODA SEIT 1737

- $t\bar{t}$ + jets backgrounds split in 5 categories by event classification Transformer
  - ➥ Constraint of each category possible in simultaneous fit
- SRs split by reconstruction Transformer Higgs boson p_T in each channel
  - Additional split of single-lepton $t\bar{t} + \geq 2b$ component for better constraining power
- Fit respective Transformer classifiers (Higgs boson p_T in single-lepton boosted regions)



Transformer Architecture

UNIVERSITÉ DE GENÈVE

GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN IN PUBLICA COMMODA SEIT 1737

Jets, Leptons, MET as input objects

- Agnostic to object number & ordering
- Features: 4-vectors, b-tag scores, object types

Fully-connected layers applied consistently to all objects

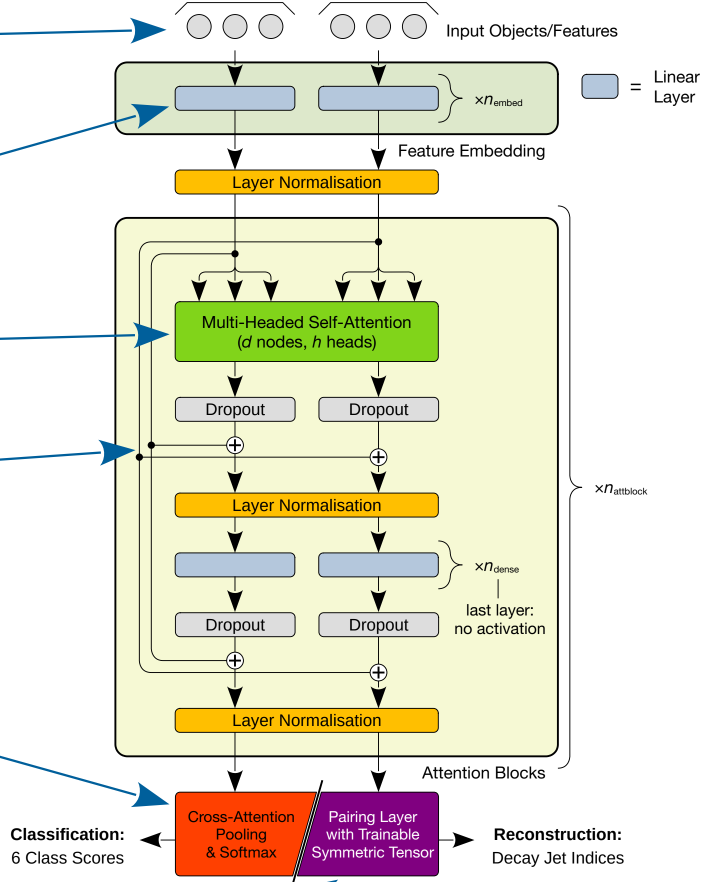
Self-Attention Mechanism for pair-wise object interactions (generates linear object combinations)

Residual connections to improve training

➥ Improved scheme with respect to original Transformer architecture

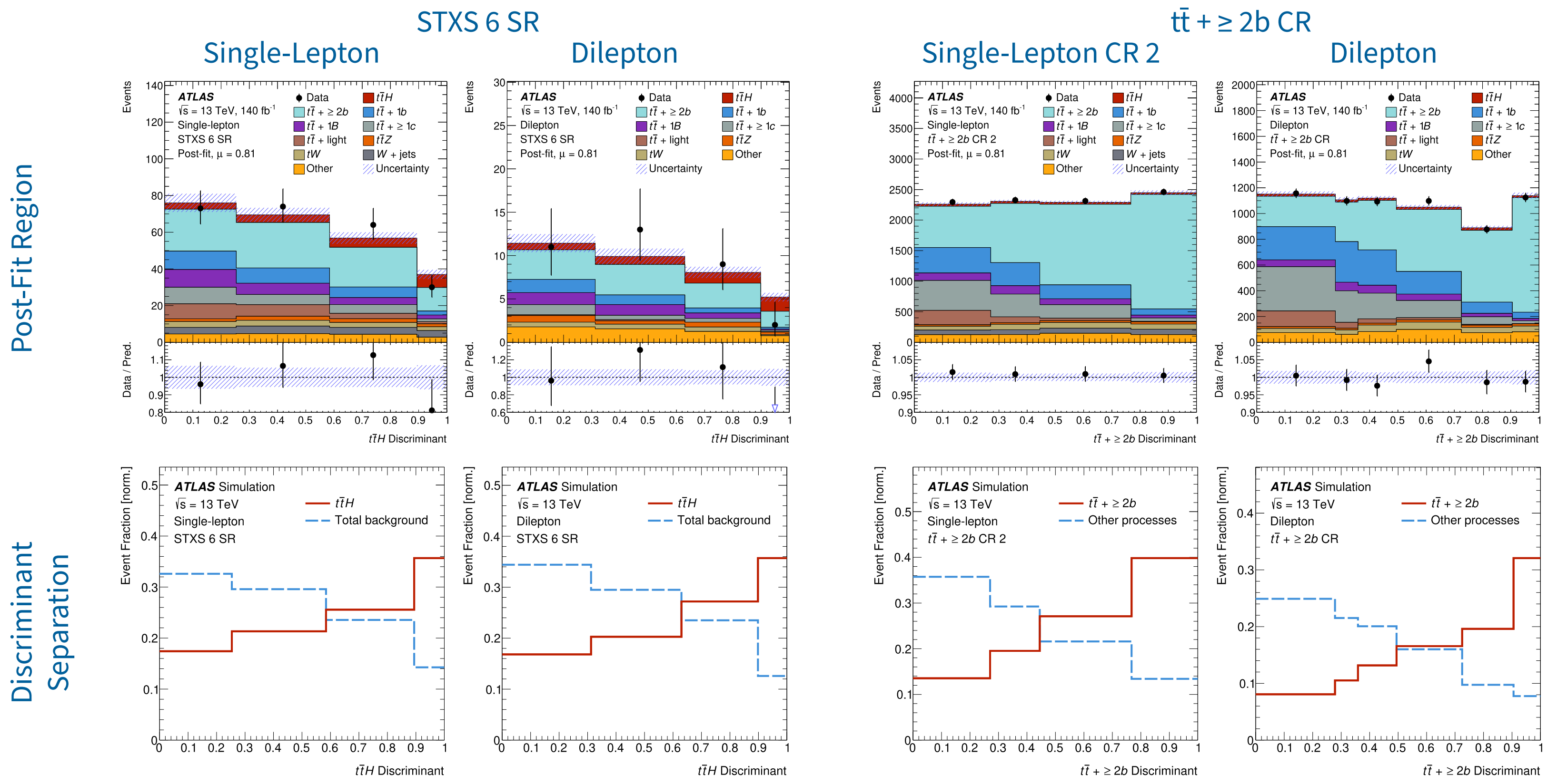
Classification layer via pooling of latent features

Pairing Layer to reconstruct 2 most likely b-jets from Higgs boson (à la tensor attention in SPANet (SciPost Phys. 12 (2022) 178))



Transformer Performance

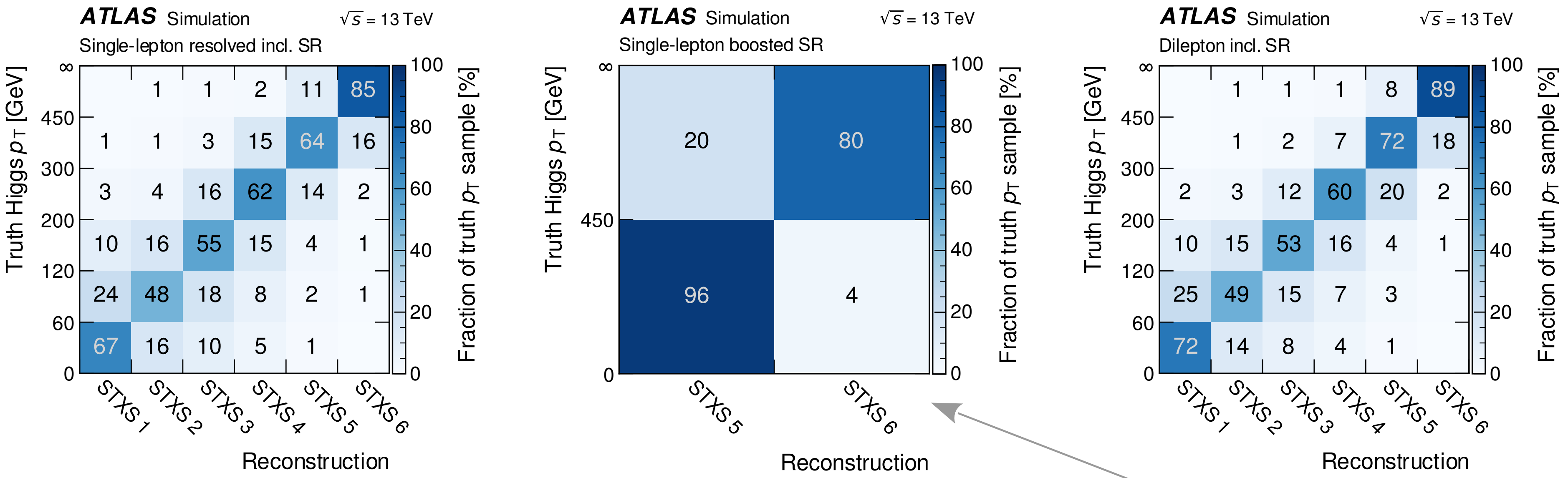
Classification



- Loosening of event selection possible due to good signal & background classification capabilities

Transformer Performance

Reconstruction



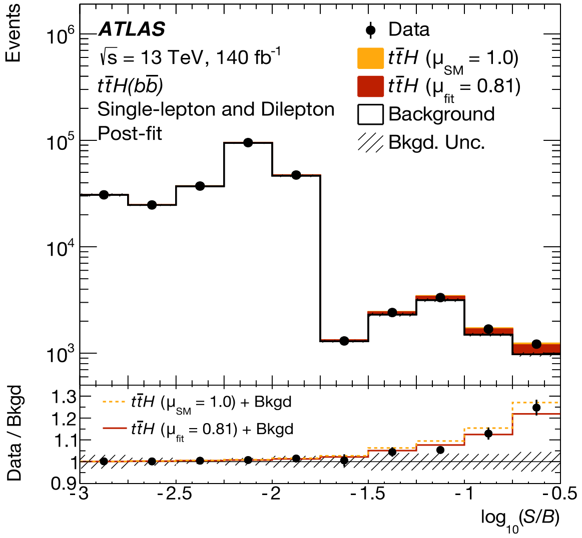
- Diagonal migration matrices for the Higgs boson p_T STXS bins
- Bin edges of SR optimised for responses of $t\bar{t}H$ truth Higgs boson p_T
  - ➥ Further reduction in migrations between SRs

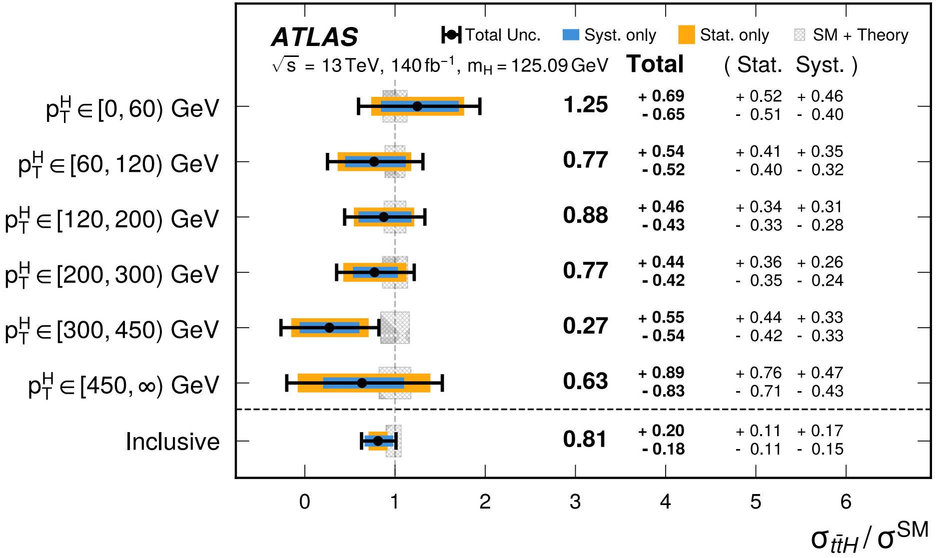
Specifically targets high Higgs boson p_T region via STXS 5 and STXS 6

Results & Conclusion

UNIVERSITÉ DE GENÈVE

GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN IN PUBLICA COMMODA SEIT 1737



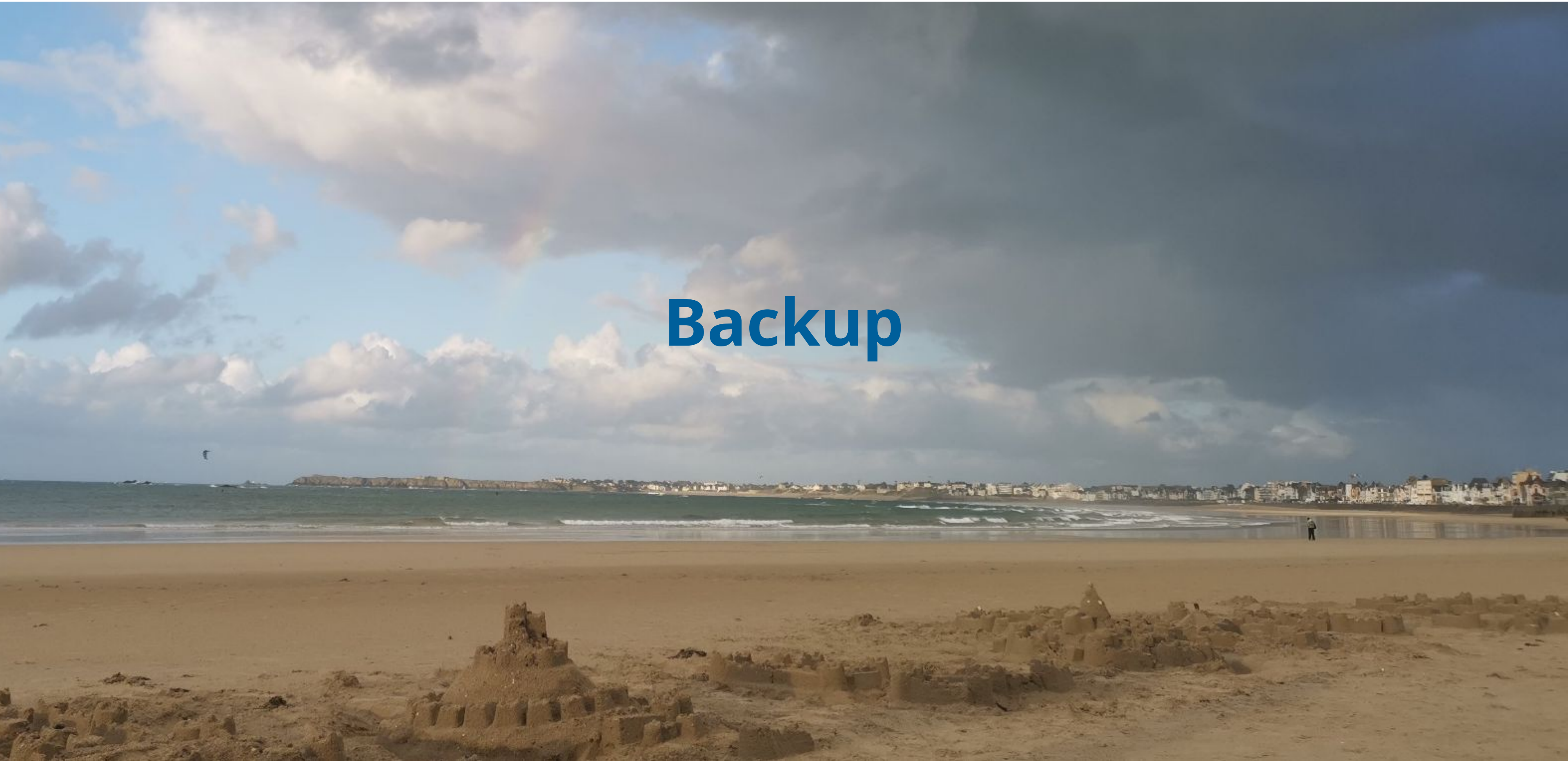


- Inclusive cross-section: $\sigma(t\bar{t}H) = 411 \pm 54\,(\text{stat.})\,^{+85}_{-75}\,(\text{syst.})$ fb (SM prediction: $507\,^{+35}_{-50}$ fb at NLO QCD+EW)
- Largest syst. uncertainties: modelling of signal and $t\bar{t} + \geq 2b$
- Significant part in improvement due to MVA developments

Observed (expected) significance: 4.6 (5.4) σ

Most precise single-channel $t\bar{t}H$ cross-section measurement to date (both inclusive & differential)!

Backup



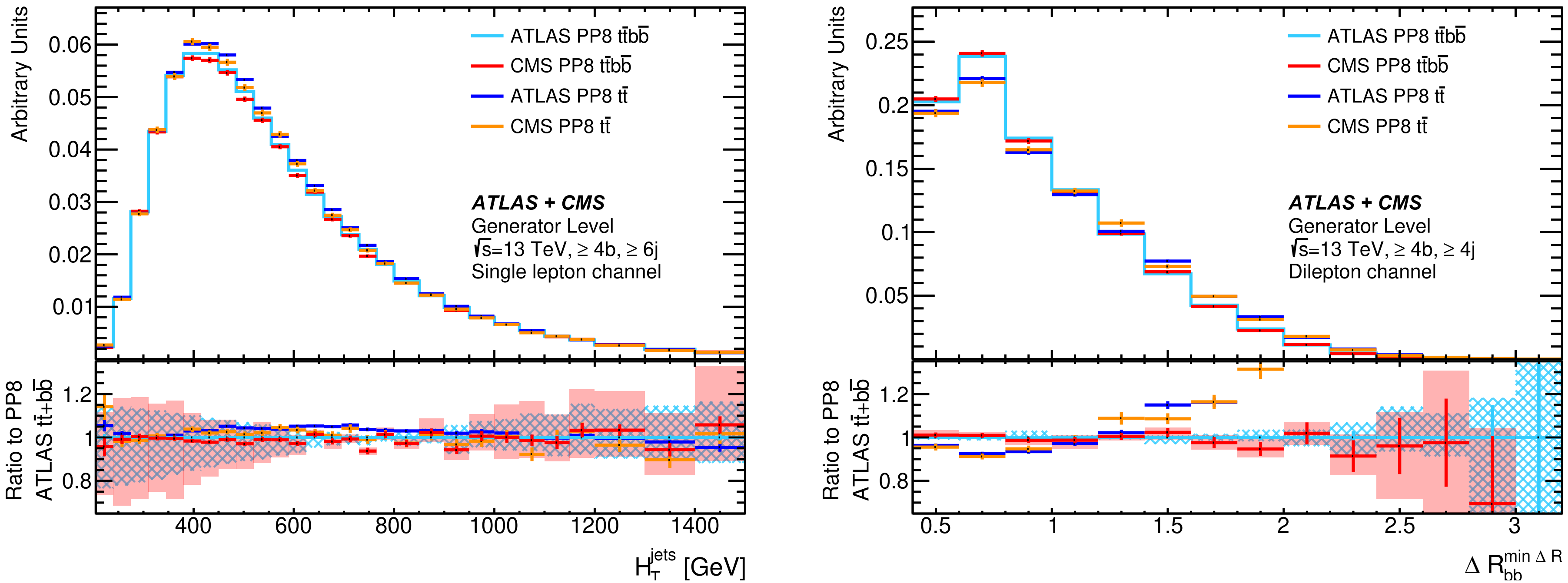
$t\bar{t}b\bar{b}$ Samples – A Comparison

Parameter	ATLAS First Full Run 2	ATLAS Run 2 Legacy	CMS Full Run 2
μ_F	$\frac{1}{2} \Sigma_{i=t,\bar{t},b,\bar{b},g}\, m_T(i)$		$\frac{1}{4} \Sigma_{i=t,\bar{t},b,\bar{b},g}\, m_T(i)$
μ_R	$\sqrt[4]{(\Pi_{i=t,\bar{t},b,\bar{b}}\, m_T(i))}$	$\frac{1}{2} \sqrt[4]{(\Pi_{i=t,\bar{t},b,\bar{b}}\, m_T(i))}$	
h_{damp}	$H_T / 2$		$1.379\, m_t$
$h_{bornzerodamp}$	2	5	2

- ATLAS legacy & CMS parameters follow dedicated studies summarised in ATL-PHYS-PUB-2022-006, LHCHWG-2022-003
- Nominal $t\bar{t}b\bar{b}$ process simulated with POWHEGBOXRES+PYTHIA8 in 4-flavour scheme (m_t = 172.5 GeV, m_b = 4.75 GeV for CMS, and m_b = 4.95 GeV for ATLAS)
- h_{damp} & $h_{bornzerodamp}$ control NLO gluon emission in POWHEG, negligible impact observed in studies

$t\bar{t}b\bar{b}$ Samples – A Comparison

UNIVERSITÉ DE GENÈVE
GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN IN PUBLICA COMMODA SEIT 1737



- Comparison of samples used in ATLAS Run 2 legacy & CMS in LHCHWG-2022-003
  - Fiducial volume of ≥ 4 b-jets, ≥ 6 jets, 1 lepton and ≥ 4 b-jets, ≥ 4 jets, 2 leptons
- Comparison also to 5-flavour scheme $t\bar{t}$ + jets samples
- Uncertainty bands from stats & QCD scale variations

$t\bar{t}$ + Jets Systematics Model

UNIVERSITÉ DE GENÈVE

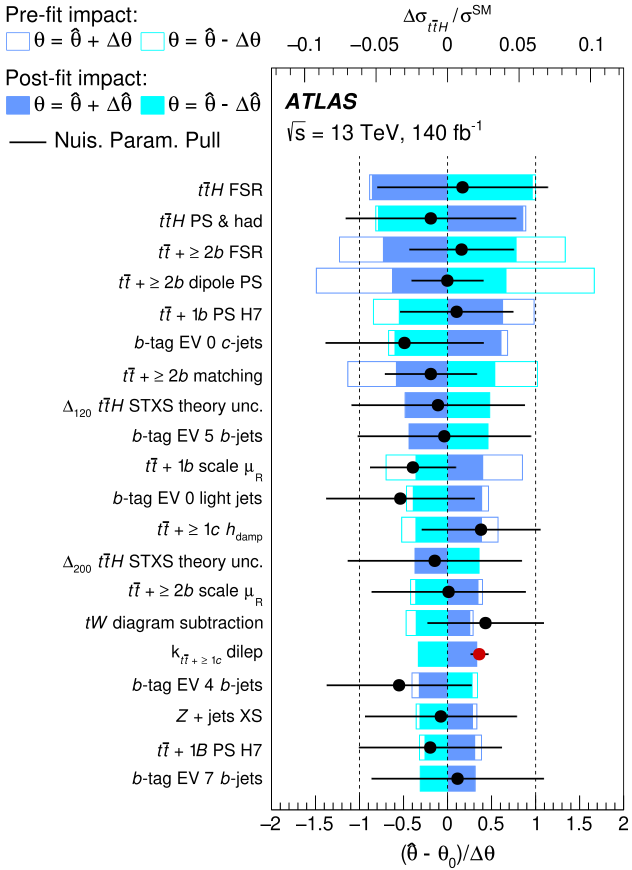
GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN IN PUBLICA COMMODA SEIT 1737

Systematic	ATLAS First Full Run 2	ATLAS Run 2 Legacy	$t\bar{t}$ + Jets Components
ME Scale	–	independent ME μ_R, μ_F variations	All
ISR	Var3c and ME μ_R, μ_F variations	A14 tune Var3c variations	All
FSR	PS FSR μ_R variations		All
Parton Shower & Hadronisation	Powheg + Herwig7 alternative (5FS only)	Powheg + Herwig7 alternative (5FS and 4FS)	All
NLO Matching	MG5_aMC@NLO+Pythia8 (5FS only)	PP8 p_T-hard = 1 alternative	All
ISR Recoil	–	PP8 dipole recoil alternative	$t\bar{t} + \geq 1b$
h_{damp} variation	–	h_{damp} up-variation alternative	$t\bar{t}$ + c/light
$t\bar{t} + \geq 1b$ Fractions	PP8 vs. Powheg + Herwig7	–	$t\bar{t} + \geq 1b$
Cross-section	6% ($t\bar{t}$ + light) and 100% ($t\bar{t}$ + c)	–	$t\bar{t}$ + c/light

Systematics Rankings & Grouped Impact

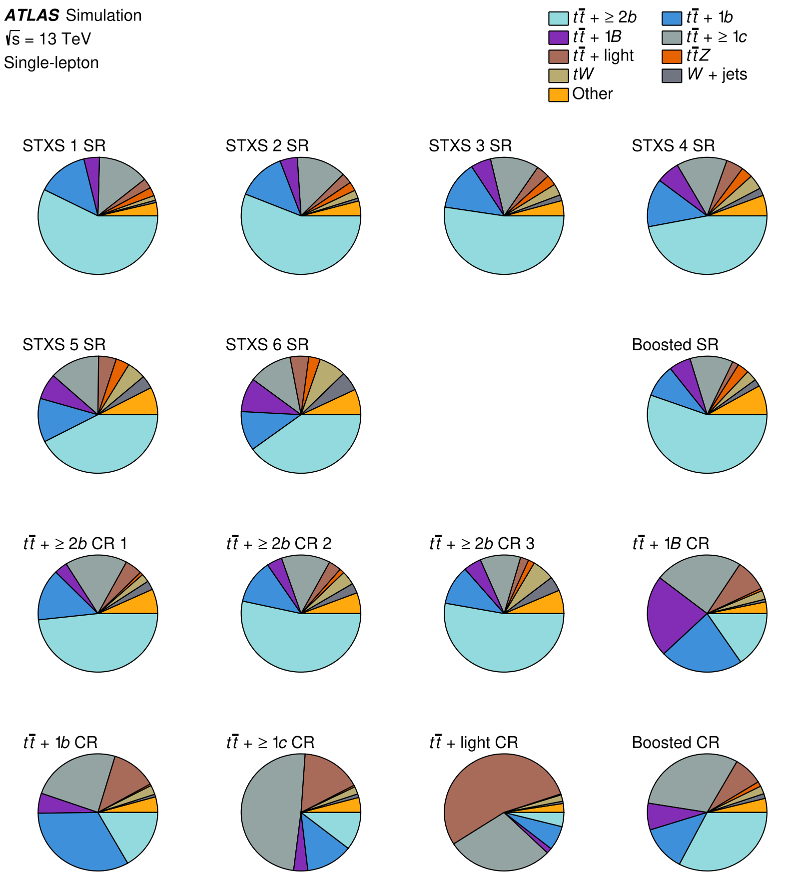
UNIVERSITÉ DE GENÈVE

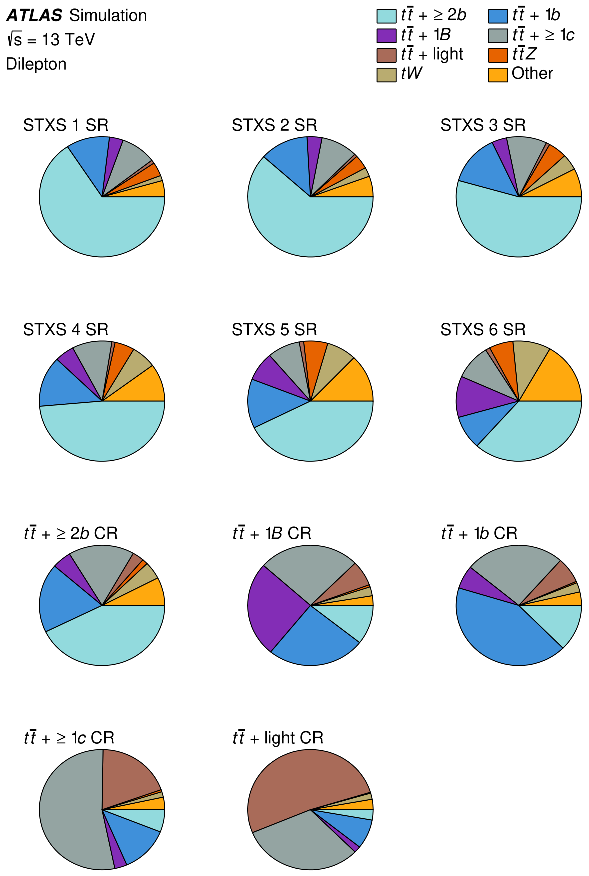
GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN IN PUBLICA COMMODA SEIT 1737



Uncertainty source	$\Delta\sigma_{t\bar{t}H}$ (fb)		$\Delta\sigma_{t\bar{t}H}/\sigma_{t\bar{t}H}$ (%)	
Process modelling				
$t\bar{t}H$ modelling				
$t\bar{t}H$ radiation	+35	−21	+9	−5
$t\bar{t}H$ parton shower	+32	−19	+8	−5
$t\bar{t}H$ matching	<0.1	−0.3	<0.1	−0.1
$t\bar{t}H$ theory	+25	−17	+6	−4
$t\bar{t} + \geq 1b$ modelling				
$t\bar{t} + \geq 1b$ radiation	±31		±8	
$t\bar{t} + \geq 1b$ parton shower	±29		±7	
$t\bar{t} + \geq 1b$ matching	±19		±5	
$t\bar{t} + \geq 1c$ modelling	±18		±4	
$t\bar{t}$ + light modelling	±5		±1	
tW modelling	±16		±4	
Minor background modelling	±19		±5	
Flavour tagging	±36		±9	
Jet modelling	±22		±5	
Monte-Carlo statistics	±17		±4	
Other instrumental	±10		±2	
Total systematic uncertainty	+85	−75	+21	−18
Normalisation factors	±21		±5	
Total statistical uncertainty	±54		±13	
Total uncertainty	+101	−92	+25	−22

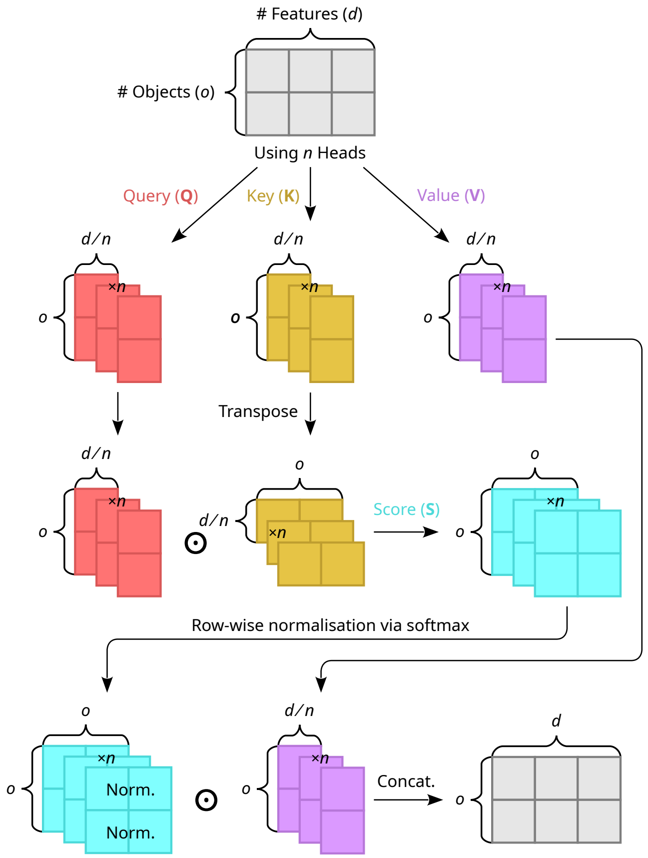
Pre-Fit Background Compositions

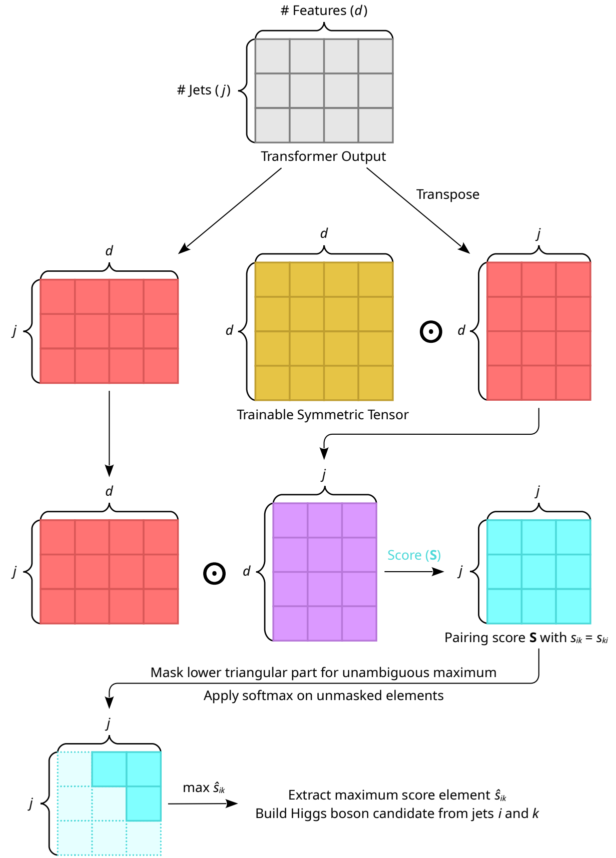




Self-Attention & Pairing Layer

UNIVERSITÉ DE GENÈVE
GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN IN PUBLICA COMMODA SEIT 1737





Input Features

UNIVERSITÉ DE GENÈVE

GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN IN PUBLICA COMMODA SEIT 1737

Feature	Description
p_x	Object momentum in x-direction.
p_y	Object momentum in y-direction.
p_z	Object momentum in z-direction.
energy	Object energy.
p_{T}	Object transverse momentum.
mass	Object mass.
η	Object pseudo-rapidity.
ϕ	Object azimuthal angle.
$\cos\phi$	Sine of object azimuthal angle.
$\sin\phi$	Cosine of object azimuthal angle.
PCBT bin	DL1r pseudo-continuous b-tagging bin assigned to jets in the following manner. Set to 0 for leptons and $E_{\mathrm{T}}^{\mathrm{miss}}$. $\text{feature} = \begin{cases} 1, & \text{if un-tagged} \\ 2, & \text{if tagged at } [85\%, 77\%) \\ 3, & \text{if tagged at } [77\%, 70\%) \\ 4, & \text{if tagged at } [70\%, 60\%) \\ 5, & \text{if tagged at } 60\%. \end{cases}$
lepton type	Lepton type of input objects. Set to 1 for electrons, 2 for muons, and 0 for jets and $E_{\mathrm{T}}^{\mathrm{miss}}$.
lepton charge	Charge of lepton objects in units of e. Set to 0 for jets and $E_{\mathrm{T}}^{\mathrm{miss}}$.
$E_{\mathrm{T}}^{\mathrm{miss}}$ flag	Whether input object is $E_{\mathrm{T}}^{\mathrm{miss}}$ (value of 1) or not (value of 0).

Some redundancy in input features, as seen to improve Transformer performance