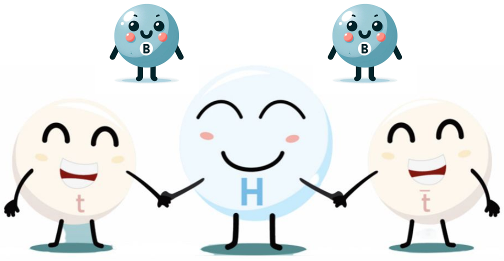


A Top Friendship

Measurement of ttH production in the H(bb) decay channel with Transformer Networks

Levi Evans | Pedro Teixeira-Dias | Sam Waterhouse

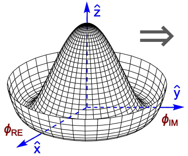
10.04.2024

Joint APP, HEPP and NP IOP conference

Introduction and motivation | Why $t\bar{t}H$?

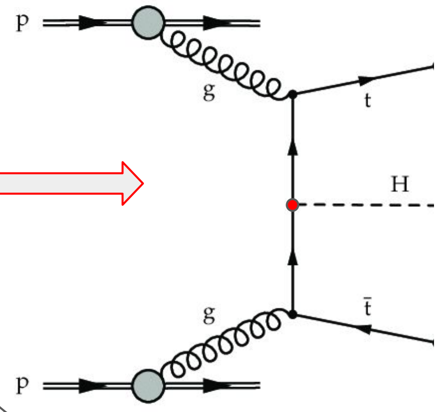
- Measuring $t\bar{t}H$ gives us a window into the top-Higgs coupling (largest in the SM)

$$\sigma_{t\bar{t}H} \propto |y_t|^2$$



⇒ Important test of the Standard Model and EW symmetry breaking

- Direct experimental probe of top Yukawa coupling at tree-level



- Opportunity to further probe Effective field theories

⇒ Probe Wilson coefficients as a function of Higgs p_T in Simplified Template cross-sections (STXS) framework

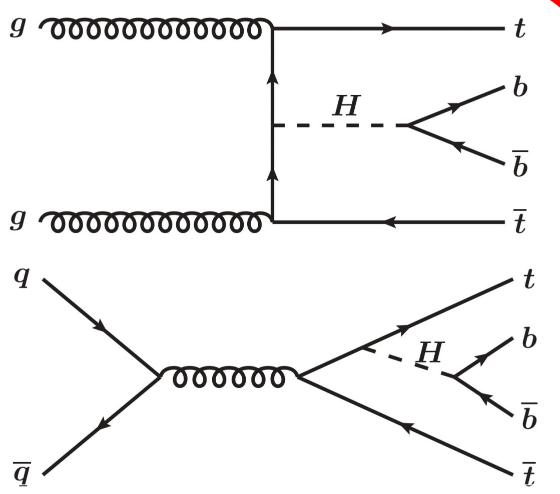
- Possible to probe CP structure of fermionic Higgs couplings

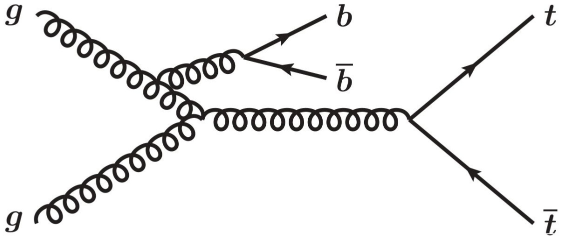
$$\mathcal{L}_{t\bar{t}H} = -\kappa_t' y_t \phi \bar{\psi}_t (\cos\alpha + i\gamma_5 \sin\alpha)\psi_t$$

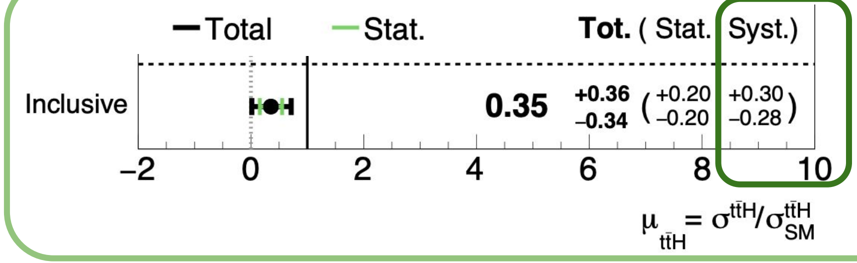
$$\mathcal{L}_{\mathrm{EFT}} = \mathcal{L}_{\mathrm{SM}} + \sum_d \frac{C_d}{\Lambda^{d-4}} \mathcal{O}_d$$

Introduction and motivation | Analysis challenges

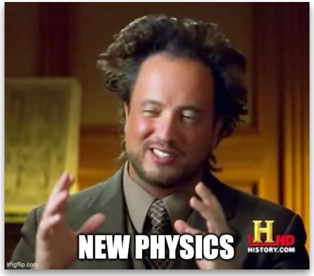
- The $H \to b\bar{b}$ decay channel offers good stats. in high p_T regimes
- $t\bar{t}H(H \to b\bar{b})$ challenging measurement to perform
- Large irreducible $t\bar{t} + b\bar{b}$ background
- Use of leptonic decays mitigates large QCD multi-jet backgrounds
- Last measurement limited by systematic uncertainties in modelling of dominant $t\bar{t}$ background (JHEP06(2022)097)
- Many improvements made to analysis design and modelling uncertainties



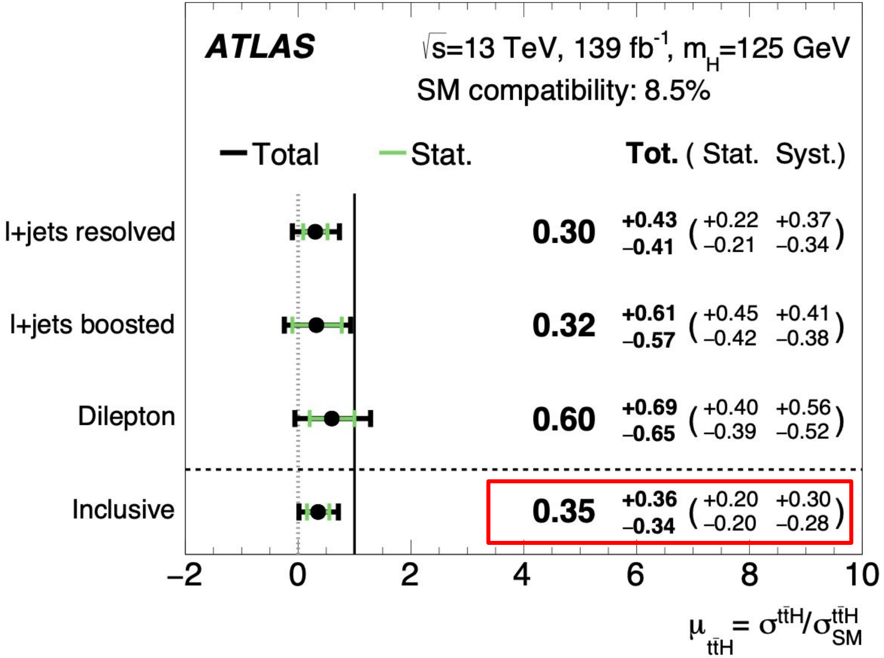


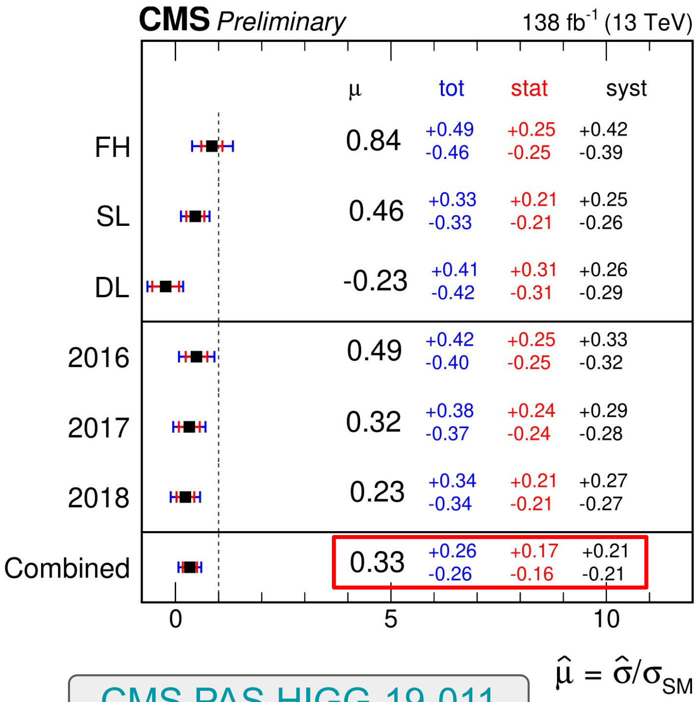


Introduction and motivation | Existing Measurements



- Talk today will show details on the **improved Full Run 2 analysis** (i.e a re-analysis)
- Increased interest in the improved measurement due to latest CMS PAS



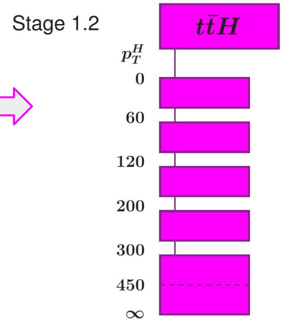


[1st Full Run 2 ATLAS Measurement]

[CMS PAS HIGG-19-011]

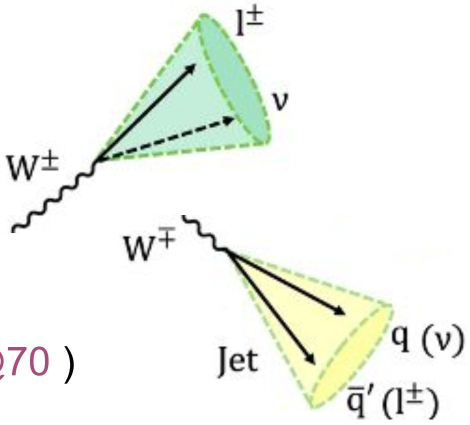
Analysis Overview | What has changed from the previous measurement?

- Based on Athena release 21, with improved object definitions and detector reconstruction
  - ⇒ PFlow jets (instead of EMTopo)
  - ⇒ DL1r b-tagger (instead of MV2c10)
  - ⇒ Full JER uncertainty model with pseudo-data smearing
- Loosened pre-selection to increase statistics
- State of the art $t\bar{t} + b\bar{b}$ nominal and systematics model developed for this analysis (ANA-PHYS-PUB-2022-006)
  - ⇒ Corresponding newly developed systematic approach to $t\bar{t} + \geq 1c$ and $t\bar{t} + \mathrm{light}$
- State of the art MVA approach using attention-based transformer encoders
- First STXS bin split into two bins to fully align with STXS v1.2
- Data-driven re-weighting to mitigate pre-fit data/MC disagreement: H_T^{all}



Analysis Overview | Strategy, Event selection and Channel definitions I

- Events recorded using lowest unprescaled single-lepton triggers
- Consider two channels, based on decays of W-boson in $t\bar{t}$ system
  - ⇒ = 2 leptons: Dilepton channel
  - ⇒ = 1 lepton: Single-lepton channel (resolved + boosted)
- Event selection loosened w.r.t previous analysis (≥3j,3b@70, ≥5j,4b@70)
  - ⇒ Allows for better control of each $t\bar{t}$ + jets background components



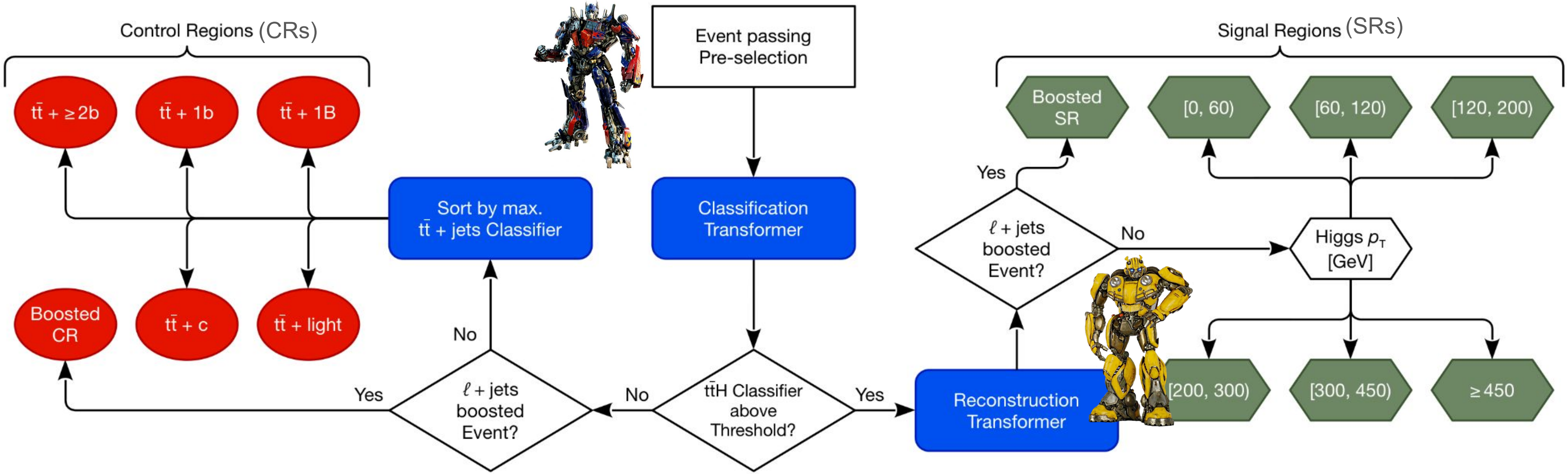
Channel	Jets	b-tags (DL1r) 70% WP	b-tags (DL1r) 85% WP	e/μ	τ	RC Jets*
dilepton	≥ 3	≥ 2	≥ 3	2	0	-
1 + jets resolved	≥ 5	≥ 3	≥ 3	1	≤ 1	-
1 + jets boosted	≥ 4	-	≥ 3	1	≤ 1	≥ 1

* Reclustered (RC) jets

$p_T \geq 300$ GeV

Analysis Overview | Strategy, Event selection and Channel definitions II

- Region definition performed via multi-class NN classifier (transformer encoders)
  - ⇒ Classification transformer classifies events according to *6 classes* (Signal Region (SR) and 5 Control Regions (CRs))
  - ⇒ Each CR enriched in relevant flavour component of $t\bar{t}$ + jets
- Events classified as ttH enter reconstruction transformer and assigned STXS bin



Analysis Overview | Strategy, Event selection and Channel definitions III

- Transformer utilises Softmax function in final layer
  - ⇒ Allows for category outputs to be interpreted as probabilities for each class
  - ⇒ These are used in region construction, where maximal scores employed for region assignment
- "Discriminant" formula used to integrate outputs for more physical classification scores, i.e. weighted to theoretical cross-sections

Raw output scores

Number of events for category j in training sample

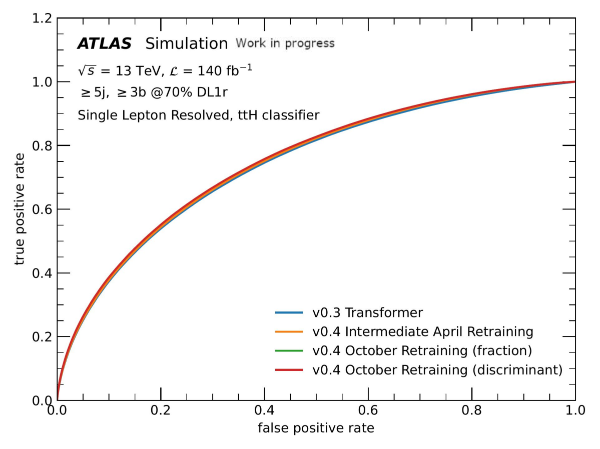
$$d_i = \frac{p_i}{\sum_{j \neq i} p_j \cdot \hat{N}_{ij}}, \quad \text{where } \hat{N}_{ij} = \frac{N_j}{\sum_{k \neq i} N_k}$$

Normalised weight term

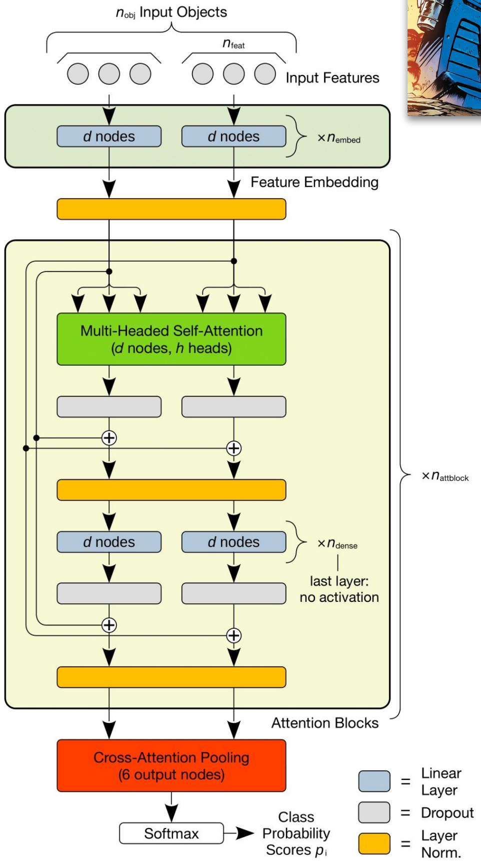
Consider all other classes against which we want to discriminate class i

Analysis Overview | MVA Algorithms | Classification Transformer

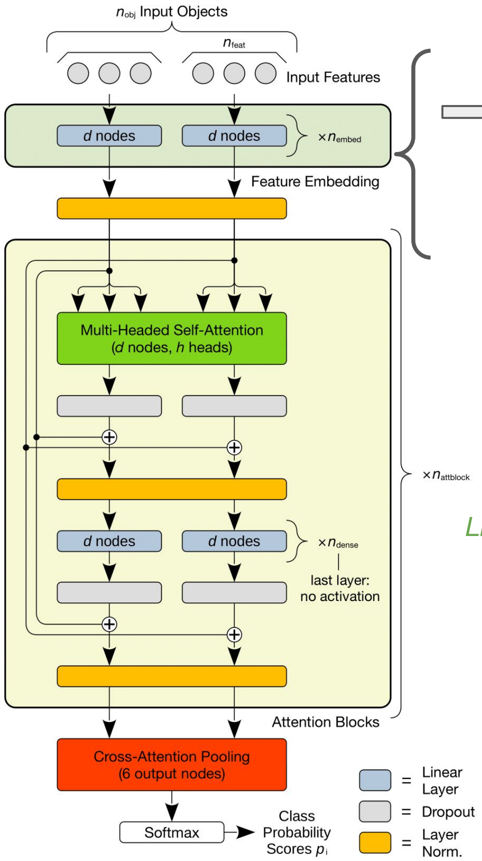
- Transformers central to the analysis design
- Trained on low-level object feature variables, as opposed to higher-level event variables used in prev. analysis iterations (reduces feature engineering task)
  - ⇒ 4-vectors of all final state objects, PCBT scores, MET, lep. type/charge
- Utilises permutation-invariant architectures & attention mechanisms
- Five-fold cross-validation employed for improved training statistics

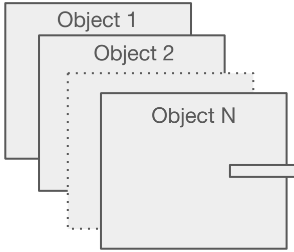


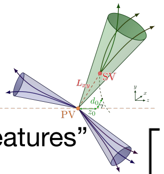
- Trained independently for each channel in pre-selection region of phase-space
- AUC scores:
  - ⇒ Single-lepton: 0.753
  - ⇒ Dilepton: 0.774



Analysis Overview | MVA Algorithms | Attention







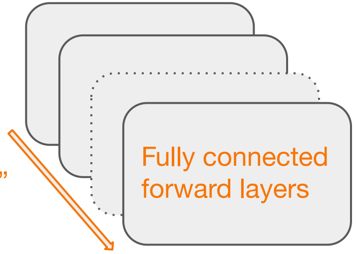
"Features"

$$\vec{P} = \begin{bmatrix} E \\ p_x c \\ p_y c \\ p_z c \end{bmatrix} = \begin{bmatrix} E \\ \vec{p}c \end{bmatrix}$$

$$p_x = p_T \cos(\phi)$$
$$p_y = p_T \sin(\phi)$$
$$p_z = p_T \sinh(\eta)$$

"Embedding"

Transform inputs into higher dimension, *D*

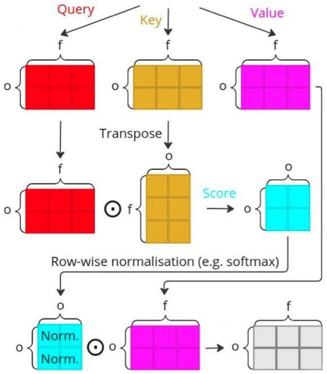


D Latent features

OBJ x FEAT matrix

"Self-attention"

Linear projection of input sequence



Query: What's important?

Key: These are important!

Value: Here's how to combine them

"Multi-head" extends by selecting multiple interesting linear combinations from different representation subspaces

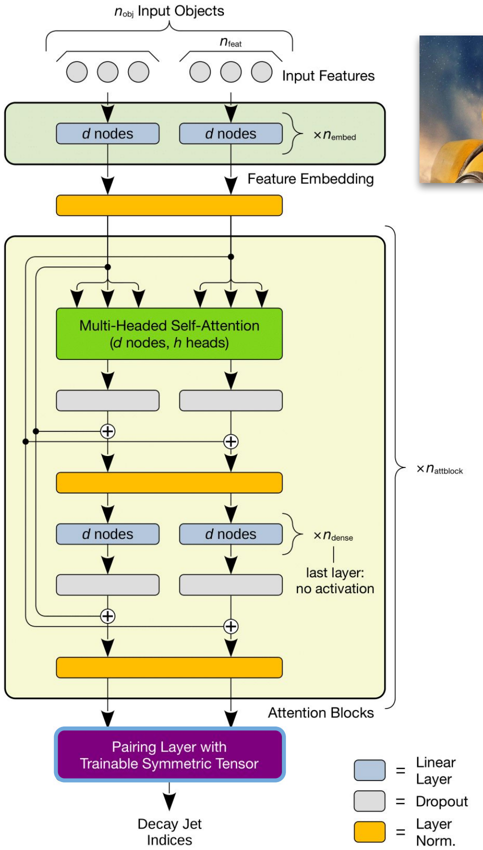
Output = score ◦ value for new object x feature matrix

"Pooling"

Fold in features from other layers. Aggregate over latent features for each object to get fixed-sized.

Analysis Overview | MVA Algorithms | Reconstruction Transformer

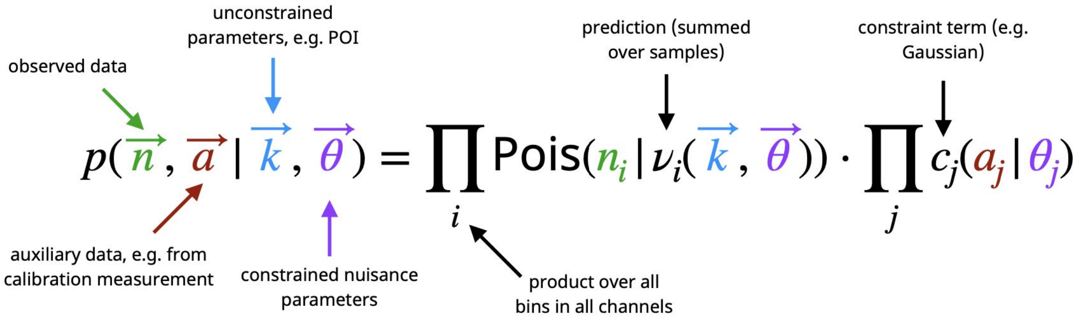
- Reconstruction transformer identifies jets originating from Higgs decay from kinematic event information
  - ⇒ Pairing score S_{ij} calculated via pairwise dot product of jet latent features, for each pair of jets i and j (à la SPANET)
- Makes full Higgs-boson 4-vector reconstruction possible
  - ⇒ Extract reconstructed Higgs p_T
- Improvements in Higgs reconstruction for STXS measurement
  - ⇒ More diagonal migration matrices
  - ⇒ Comparison shown in backup slides



Statistical Analysis | Brief overview

More information on improved systematics model in [backups]

- Binned profile likelihood fit to all bins in analysis phase-space performed



$$p(\vec{n}, \vec{a} \mid \vec{k}, \vec{\theta}) = \prod_i \mathrm{Pois}(n_i \mid \nu_i(\vec{k}, \vec{\theta})) \cdot \prod_j c_j(a_j \mid \theta_j)$$

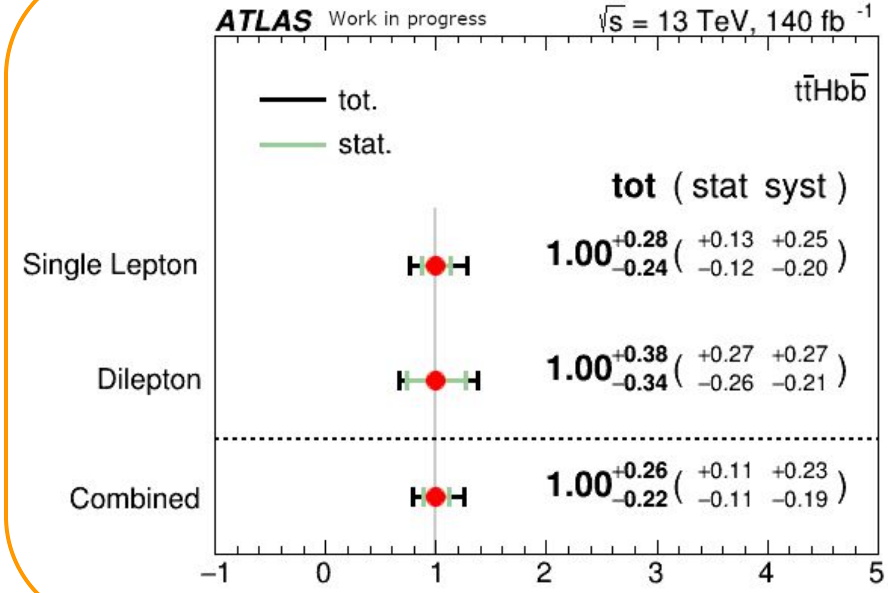
- Respective discriminant observables used as template variables
  - ⇒ Boosted regions use reconstructed Higgs pT as fitting variable
  - ⇒ **Fits to data** (in unblinded bins, under the background-only hypothesis)
    - 5 free-floating normalisation parameters for each $t\bar{t}$ + jets flavour component
  - ⇒ **Fits to Asimov dataset** (including blinded bins, under S+B hypothesis)
    - 6 (11) free-floating parameters in inclusive (stxs) fits, with 1 (6) POI(s)

Statistical Analysis | Fit to the Asimov Dataset | Inclusive Measurement

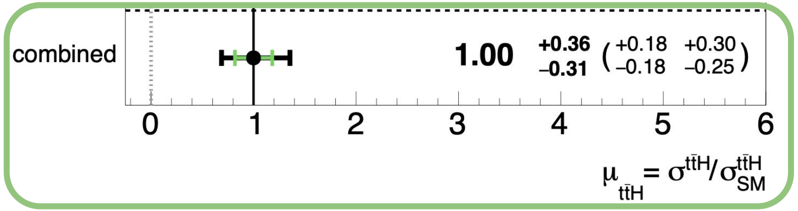
- Increased precision w.r.t last measurement
- Inclusive measurement dominated by systematic uncertainty, but with clear reduction w.r.t prev. analysis
- Expected discovery significance

  ⇒ $\mathcal{Z} = 5.5\sigma$

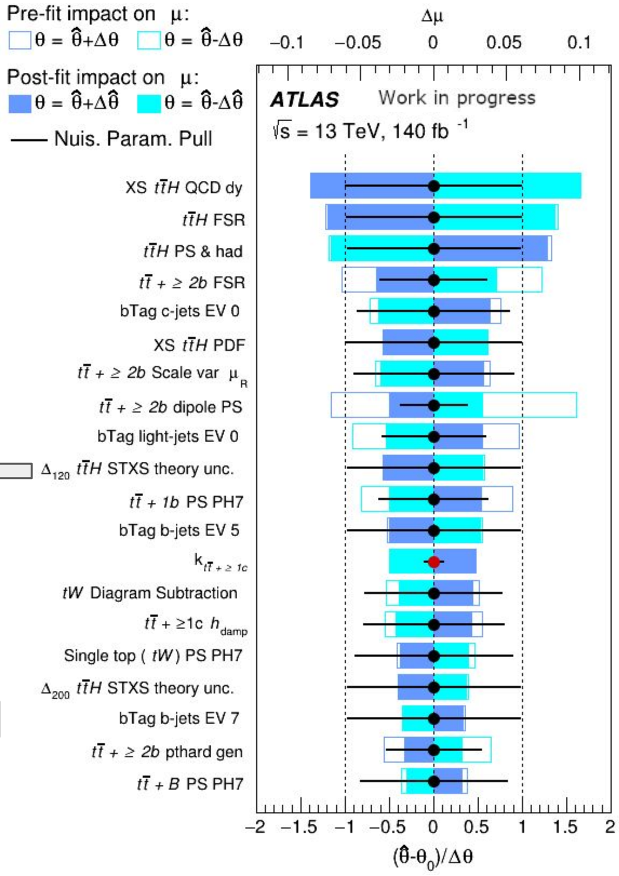
  ⇒ Above boundary for discovery!



- Up from $\mathcal{Z} = 2.7\sigma$ for previous measurement



Statistical Analysis | Fit to the Asimov Dataset | Inclusive Ranking



Higgs pT bin migration uncertainties

ATL-PHYS-PUB-2023-031

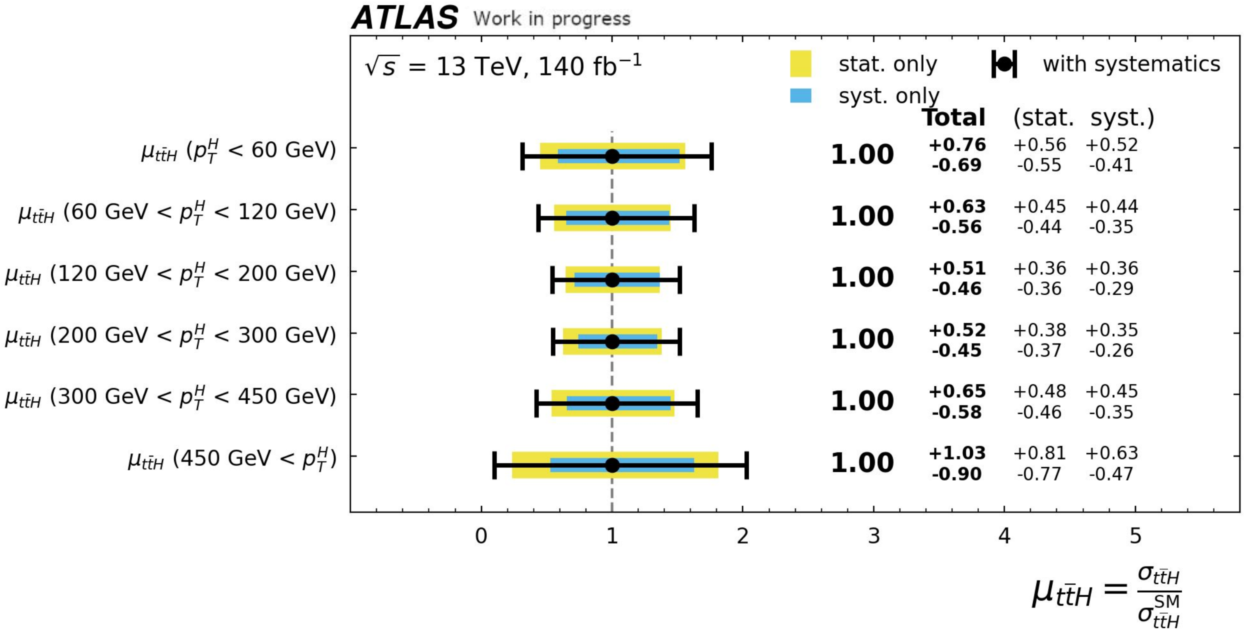
- Dominant uncertainties ttH modelling
  - ⇒ Cross-section, FSR, PS

tt + ≥ 2b modelling dominant in prev. iteration!

- tt + ≥ 2b modelling uncertainties rank highly
  - ⇒ FSR, dipole PS
  - ⇒ Renorm. scale variations
- Contributions also from
  - ⇒ tt + 1b PH7 PS
  - ⇒ B-tagging efficiency and mis-tags (light EV0)

Statistical Analysis | Fit to the Asimov Dataset | STXS Measurement

- Perform differential measurement w.r.t reconstructed Higgs pT, within STXS framework
- Further explore Higgs properties and better highlight deviations



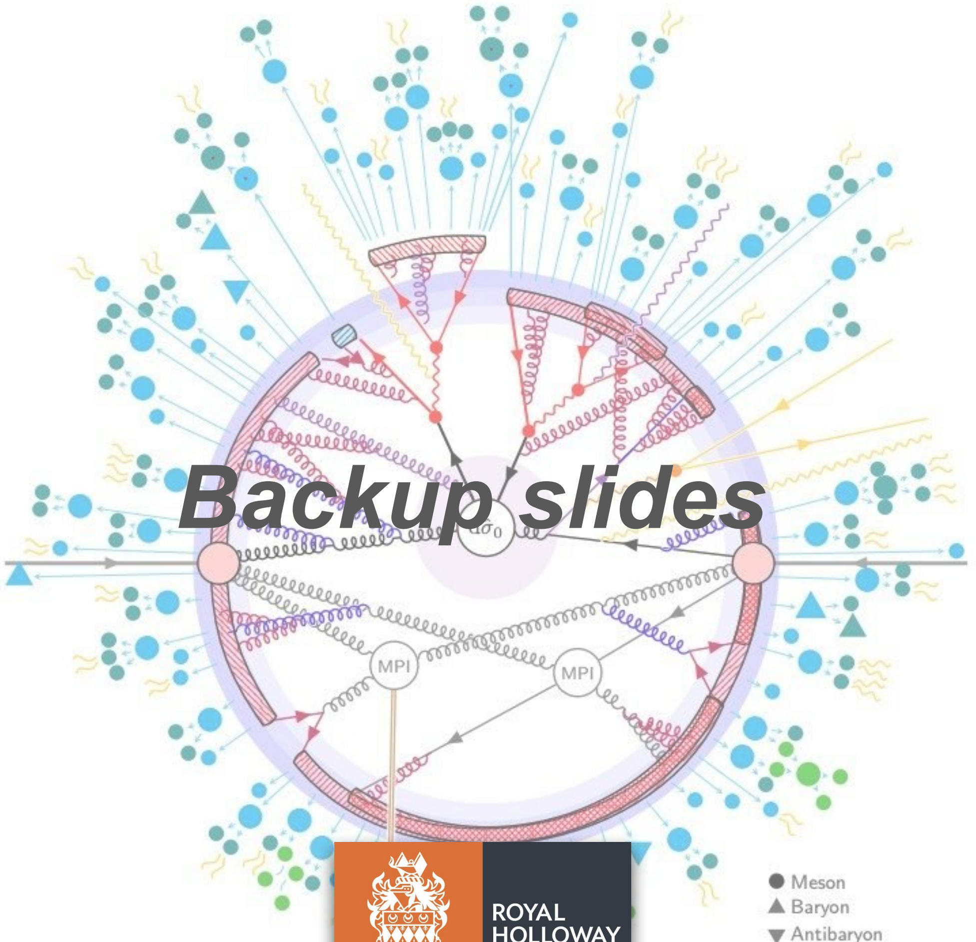
- Improved reconstruction performance and statistics allows for split in lower STXS bins to fully align with STXS v1.2
- Statistically limited
- Improvements in expected sensitivity
  - ⇒ STXS 1/2: ~ 20%
  - ⇒ STXS 3: ~ 45%
  - ⇒ STXS 4: ~ 40%
  - ⇒ STXS 5: ~ 15%
  - ⇒ STXS 6: ~ 40%

Conclusions and closing remarks

- The $t\bar{t}H(H \rightarrow b\bar{b})$ full Run 2 re-analysis greatly improves on the previous measurement
  - ⇒ Improved tt + bb systematics model
  - ⇒ Improved MVA approach, benefitting from latest advances in machine-learning
  - ⇒ Split of background categories to better constrain various modelling aspects
- A more than two-fold increase in expected discovery significance
- Expected performance in STXS-based differential measurement improved
- Important contribution to Global Higgs combination measurements
- Updated result important verification of SM, or exciting hints of something beyond…

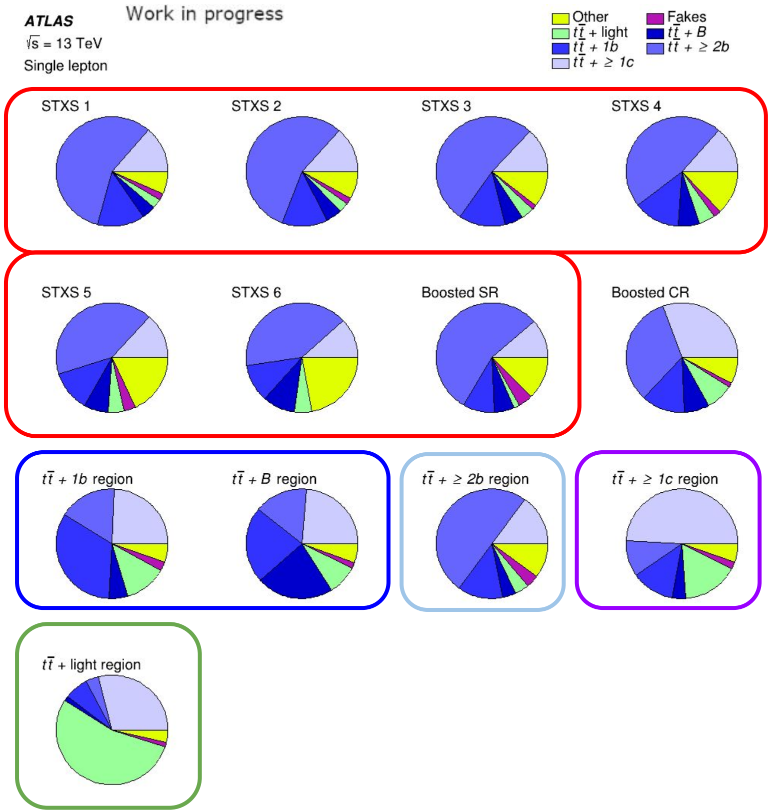
Thank you for your attention!

Backup slides



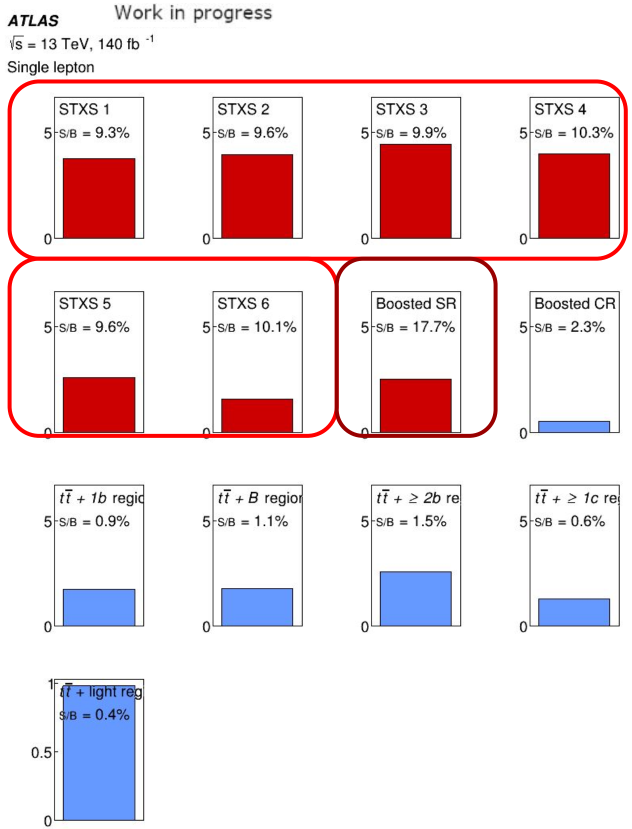
Backup Slides | Single-lepton Channel | Background Composition

- Dominant background in signal regions is tt + ≥ 2b
- Other backgrounds large in high Higgs pT bins
  - ⇒ Boosted SR better discriminates against non-ttbar
- tt + light region very pure (> 50%)
- tt + ≥ 1c and tt + ≥ 2b good purity (~ 50%)
- tt + 1b and 1B regions slightly less pure
- Compositional changes between SR and CRs warrant sep. scale factors



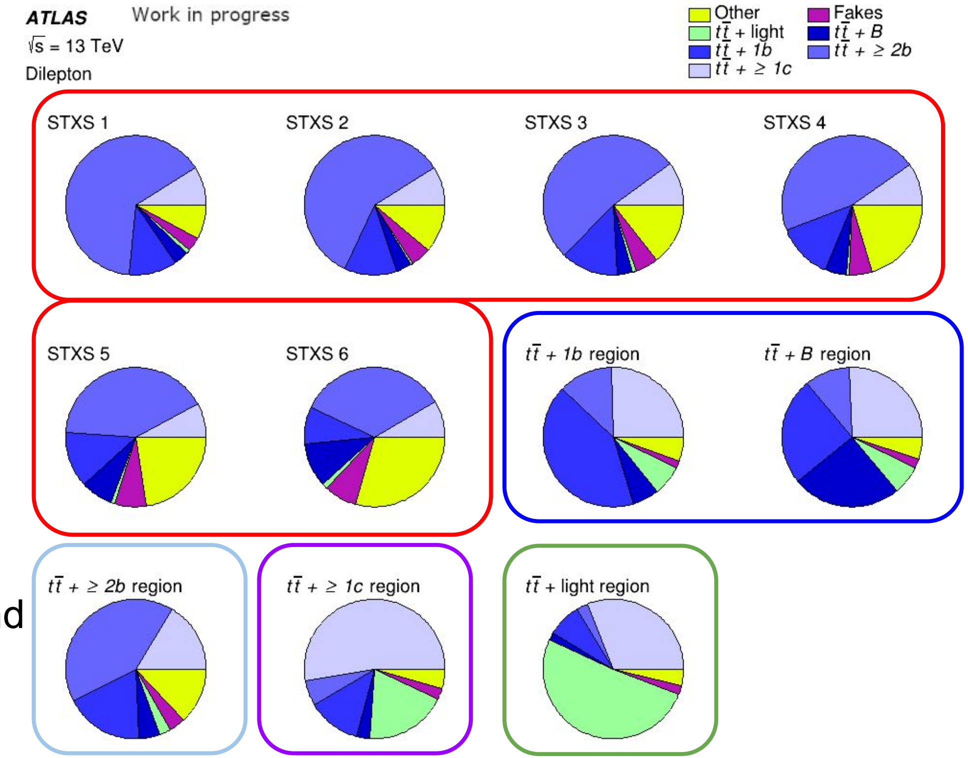
Backup Slides | Single-lepton Channel | Signal Composition

- Resolved signal regions have signal purity of ~10%
- Boosted signal region has signal purity of ~18%
- Minimal signal contribution in control regions of analysis (< 2%)



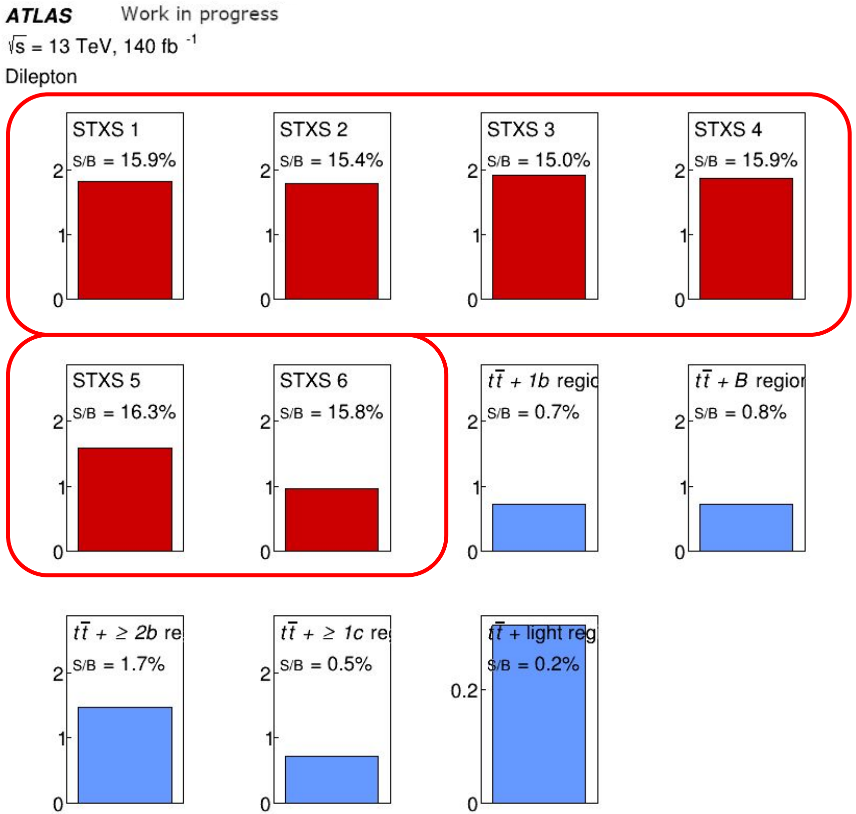
Backup Slides | Dilepton Channel | Background Composition

- Dominant background in signal regions is tt + ≥ 2b
- Other backgrounds large in high Higgs pT bins
- tt + light region very pure (> 50%)
- tt + ≥ 1c very good purity (> 50%)
- tt + ≥ 2b, tt + 1b and tt + 1B regions slightly less pure
- Compositional changes between SR and CR warrant sep. scale factors



Backup Slides | Dilepton Channel | Signal Composition

- Dilepton signal regions have signal purity of ~16%
  - ⇒ Owing to better discrimination power
- Minimal signal contribution in control regions of analysis (< 2%)



Backup Slides | Transformer Input Features

Feature	Description	Feature Transformations
p_x	Object momentum in x-direction.	Re-scaled to $\mu = 0, \sigma = 1$.
p_y	Object momentum in y-direction.	Re-scaled to $\mu = 0, \sigma = 1$.
p_z	Object momentum in z-direction.	Re-scaled to $\mu = 0, \sigma = 1$.
energy	Object energy.	Re-scaled to $\mu = 0, \sigma = 1$.
p_{T}	Object transverse momentum.	Re-scaled to $\mu = 0, \sigma = 1$.
mass	Object mass.	Re-scaled to $\mu = 0, \sigma = 1$.
η	Object pseudo-rapidity.	Re-scaled to $\mu = 0, \sigma = 1$.
ϕ	Object azimuthal angle.	Re-scaled to $\mu = 0, \sigma = 1$.
$\cos\phi$	Sine of object azimuthal angle.	Re-scaled to $\mu = 0, \sigma = 1$.
$\sin\phi$	Cosine of object azimuthal angle.	Re-scaled to $\mu = 0, \sigma = 1$.
PCBT bin	DL1r pseudo-continuous b-tagging bin assigned to jets in the following manner. Set to 0 for leptons and $E_{\mathrm{T}}^{\mathrm{miss}}$. $$\text{feature} = \begin{cases} 1, & \text{if un-tagged} \\ 2, & \text{if tagged at } [85\%, 77\%) \\ 3, & \text{if tagged at } [77\%, 70\%) \\ 4, & \text{if tagged at } [70\%, 60\%) \\ 5, & \text{if tagged at } 60\%. \end{cases}$$	None.
lepton type	Lepton type of input objects. Set to 1 for electrons, 2 for muons, and 0 for jets and $E_{\mathrm{T}}^{\mathrm{miss}}$.	None.
lepton charge	Charge of lepton objects in units of e. Set to 0 for jets and $E_{\mathrm{T}}^{\mathrm{miss}}$.	Re-scaled to $\mu = 0, \sigma = 1$.
$E_{\mathrm{T}}^{\mathrm{miss}}$ flag	Whether input object is $E_{\mathrm{T}}^{\mathrm{miss}}$ (value of 1) or not (value of 0).	None.

- Legacy transformers trained on low-level object features
  - ⇒ Object 4-vectors
  - ⇒ PCBT b-tagging scores
  - ⇒ Lepton type
  - ⇒ Lepton charge
  - ⇒ MET
- Full event information available for sequence processing!

Backup Slides | Transformer Training

Training hyper-parameters

Parameter	Classification Single Lepton	Classification Dilepton	Reconstruction Single Lepton	Reconstruction Dilepton
Max. number of jet objects	11	10	11	10
Dropout Rate	0.04			
Cross-attention pooling query regulariser	linear, 1×10^{-6}		n/a	
Pairing layer trainable tensor regulariser	n/a		linear, 1×10^{-6}	
Training batch size	2048			
Validation batch size	2048			
AdamW weight decay	5×10^{-6}		1×10^{-6}	
AdamW first moment decay rate	0.9			
AdamW second moment decay rate	0.999			
Label smoothing	0.1			
Max. number of training epochs	300			
Early stopping min. loss decrease	0			
Early stopping patience	10			
Initial learning rate	1×10^{-8}			
Max. learning rate	3.5×10^{-5}			
Final learning rate	1×10^{-7}			
Learning rate burn-in duration	10			
Learning rate ramp-up duration	10			
Learning rate plateau duration	15			
Learning rate ramp-down duration	100			

Transformer hyper-parameters

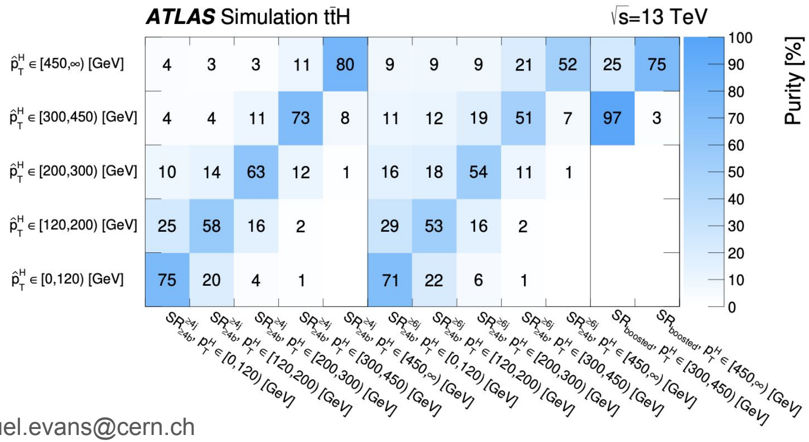
Parameter		Classification Single Lepton	Classification Dilepton	Reconstruction Single Lepton	Reconstruction Dilepton
d	number of latent features	256	256	128	128
h	number of attention heads	4	4	8	8
n_{embed}	number of feature embedding layers	2	2	2	2
n_{attblock}	number of attention blocks	11	10	10	8
n_{dense}	number of linear layers per attention block	5	4	3	3

Training Samples

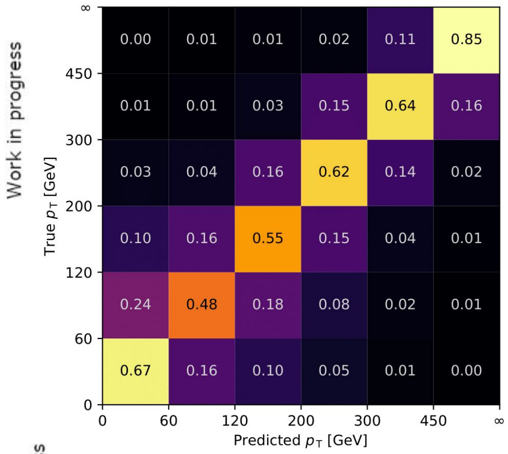
Process	Combined MC Statistics Single Lepton	Combined MC Statistics Dilepton	Samples
$t\bar{t}H$	4 438 478	2 511 844	PowhegBox+Pythia8 PowhegBox+Herwig7
$t\bar{t} + \geq 2b$	2 761 118	996 653	PowhegBox+Pythia8
$t\bar{t} + 1b$	2 390 870	1 332 277	PowhegBox+Herwig7
$t\bar{t} + 1B$	837 097	429 334	PowhegBox+Pythia8 , p_T-hard=1 variation
$t\bar{t} + \geq 1c$	696 998	908 461	
$t\bar{t}$ + light	1 431 191	1 121 241	

Backup Slides | Higgs pT reconstruction migration matrices comparison

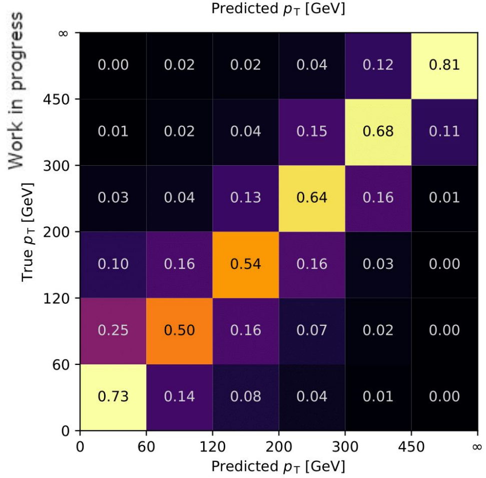
- Reconstruction transformer identifies jets originating from Higgs decay from kinematic event information
  - ⇒ Pairing score S_{ij} calculated via pairwise dot product of jet latent features for each pair of jets i and j (à la SPANET)
- Makes full Higgs-boson 4-vector reconstruction possible
  - ⇒ Extract reconstructed Higgs



Previous Analysis



Single-lepton Channel



Dilepton Channel

Backup Slides | Nominal MC Samples

Process	Sample
$t\bar{t}H$	Powheg + Pythia8
$t\bar{t}$ + light, $t\bar{t} + \geq 1c$	Powheg + Pythia8 $t\bar{t}$ @NLO 5FS
$t\bar{t} + \geq 1b$	Powheg + Pythia8 $t\bar{t} + b\bar{b}$ @NLO 4FS
tH, $t\bar{t}V$, $t\bar{t}t\bar{t}^{*}$	aMC@NLO + Pythia8
V + jets, diboson	Sherpa
Single top	Powheg + Pythia8 (dynamic scale for tW)

* AFII Simulated

Backup Slides | $t\bar{t}$ + jets normalisation

- Owing to enlarged phase space and improved MVA techniques
  - ⇒ ability to separately constrain each $t\bar{t}$ background component
- Each flavour component normalisation estimated via binned profile likelihood fit (i.e via 5 free-parameters)
  - ⇒ tt + 1b: one additional particle jet matched to one b-hadron
  - ⇒ tt + ≥ 2b: two or more additional particle jets
  - ⇒ tt + 1B: one additional particle jet matched to 2 or more b-hadrons
- Data-driven scaling factors derived from a previous fit of background-only model to bins below blinding threshold
- Subsequent normalisation factors used in kinematic correction derivations, fake lepton estimation and discriminants
  - ⇒ Significant $t\bar{t} + \geq 1c$ correction can affect H_T^{all} distribution at low values of HT
- Scaling corrections also applied to all pre-fit distributions

Channel	$t\bar{t}$ + X component	normalization factor value
Single Lepton	$t\bar{t} + c$	1.79
	$t\bar{t}$ + light	0.74
	$t\bar{t} + 1b$	1.08
	$t\bar{t} + 1B$	1.08
	$t\bar{t} + \geq 2b$	0.96
Dilepton	$t\bar{t} + c$	1.59
	$t\bar{t}$ + light	0.85
	$t\bar{t} + 1b$	1.29
	$t\bar{t} + 1B$	1.29
	$t\bar{t} + \geq 2b$	0.93

Backup Slides | $t\bar{t}$ + jets systematics model

- All systematic samples normalised to the same nominal cross-sections in analysis phase-space
- Comparison of improved systematic model for $t\bar{t}$ + jets background

Uncertainty	Previous analysis	Legacy analysis	$t\bar{t}$ + jets components
ISR	Var3c (PS) and μ_R/μ_F (ME)	Var3c (PS)	All
FSR	μ_R FSR (PS)	μ_R FSR (PS)	All
ME scale	-	Independent μ_R/μ_F	All
NLO matching	aMC@NLO + Pythia8	PP8 4FS pthard = 1 4FS	$t\bar{t} + \geq 1b$
		PP8 4FS pthard = 1 5FS	$t\bar{t}$ + light, $t\bar{t} + \geq 1c$
PS & Hadronisation	Powheg + Herwig7 (PH7)	PH7 4FS	$t\bar{t} + \geq 1b$
		PH7 5FS	$t\bar{t}$ + light, $t\bar{t} + \geq 1c$
Parton shower	-	PP8 4FS dipole recoil	$t\bar{t} + \geq 1b$