PhD work so far

$t\bar{t}H(H \to c\bar{c})$ analysis, ML and statistics methods, MC generators

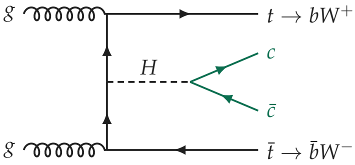
Lenart Jerala

Jožef Stefan Institute · University of Ljubljana · ATLAS

Working draft · September 2026

$t\bar{t}H(H \to c\bar{c})$: the charm Yukawa in top-associated production

Physics case



- Probes y_c directly in the decay: $\sigma \times \mathcal{B} \propto \kappa_t^2 \kappa_c^2 / \kappa_H^2$.
- Channels by $t\bar{t}$ decay: 0ℓ, 1ℓ, 2ℓ.
- Backgrounds: $t\bar{t}+\geq 1c$, $t\bar{t}+\geq 1b$, $t\bar{t}$+light, $t\bar{t}Z$.

My contributions (thanks to Jan)

- **Event classifier**: multiclass $t\bar{t}H$ / $t\bar{t}Z$ / $t\bar{t}$ transformer on reconstructed objects; methods blew others out of the water, now widely adopted in the ATLAS Higgs group.
- **Signed MC weights**: CMOK, a novel optimal observable construction for samples with negative weights.
- **Deployment**: PyTorch $\to$ ONNX $\to$ C++ ONNX Runtime, with Jan.
- **Likelihood fits**: novel one-histogram fit on the classifier's efficient score, and unbinned likelihood-ratio fits.

Small rate, irreducible $t\bar{t}$+HF, ambiguous jet–parton assignment: hard to classify.

CANDY: rank-3 attention for the $t\bar{t}H(H \to c\bar{c})$ classifier

First author · draft in preparation

Architecture

- Transformers correlate particles: attention learns how every object relates to every other.
- **E_T^{miss} pairwise bias** b_{ij}^{MET}: each pair's relation to E_T^{miss} enters the attention.
- **CANDY** (Cross-order Attention for N-body Dynamics): each query scores key pairs (j,k) jointly.

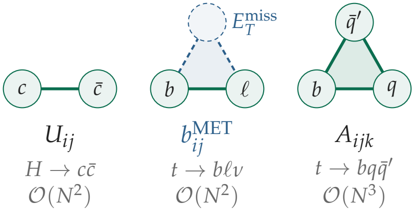
Attention scores

$$A_{ij} = \text{softmax}_j \left(q_i \cdot k_j / \sqrt{d} + U_{ij} + b_{ij}^{\text{MET}} \right)$$

$$U_{ij} = \text{MLP}\left(\ln \Delta R_{ij},\ \ln k_{T,ij},\ \ln z_{ij},\ \ln m_{ij}^2 \right)$$

$$A_{ijk} = \text{softmax}_{j<k} \left(q_i \cdot W(k_j \odot k_k) + B_{ijk} \right)$$

What each term sees



$$b_{ij}^{\text{MET}} = \text{MLP}\left(u_i + u_j,\ |u_i - u_j|\right)$$

$$u_i = W_p\, \text{MLP}(v_i)$$

$$v_i = \left(\cos \Delta\phi_i,\ \sin \Delta\phi_i,\ \ln m_{T,i}\right)$$

From large-scale ML: listening to industry (just a taste)

Attention block

QK-norm
Henry+ 2020; Dehghani+ 2023

$$\hat{q} = \mathrm{LN}_h(q), \quad \hat{k} = \mathrm{LN}_h(k), \quad \ell_{ij} = \hat{q}_i \cdot \hat{k}_j / \sqrt{d_h}$$

Norm gains fix the logit scale: no saturated, frozen heads.

Talking heads
Shazeer+ 2020

$$\tilde{\ell}^h = \textstyle\sum_{h'} P^{\ell}_{hh'}\, \ell^{h'}, \qquad \tilde{A}^h = \textstyle\sum_{h'} P^{w}_{hh'}\, \mathrm{softmax}(\tilde{\ell}^{h'})$$

Heads mix before and after the softmax.

GeGLU feed-forward
Shazeer 2020

$$\mathrm{FFN}(x) = W_2\left[\mathrm{GELU}(W_1 x) \odot W_3 x\right]$$

Gated MLP, as in current large language models.

Stochastic depth
Huang+ 2016

$$x_{l+1} = x_l + b_l\, F_l(x_l) / (1 - p_l)$$

$b_l \sim \mathrm{Bern}(1 - p_l)$; drop rate p_l rises linearly to 0.1.

Optimisation

Schedule-Free AdamW
Defazio+ 2024

$$\begin{aligned} y_t &= (1-\beta_1)\, z_t + \beta_1 x_t, \qquad g_t = \nabla \mathcal{L}(y_t) \\ z_{t+1} &= z_t - \gamma_t \big(g_t / (\sqrt{v_t} + \epsilon) + \lambda\, y_t\big) \\ x_{t+1} &= (1 - c_{t+1})\, x_t + c_{t+1} z_{t+1} \end{aligned}$$

$c_{t+1} = \gamma_t^2 / \sum_{i \le t} \gamma_i^2$; warmup only, no decay schedule.

Weight EMA
Polyak & Juditsky 1992

$$\bar{\theta}_t = \beta_t\, \bar{\theta}_{t-1} + (1 - \beta_t)\, \theta_t$$

$\beta_t \nearrow \beta$; smoothed weights for validation and export.

Focal loss, label smoothing
Lin+ 2017; Szegedy+ 2016

$$\mathcal{L} = -\textstyle\sum_c \tilde{y}_c\, (1 - p_c)^{\gamma} \ln p_c$$

$\tilde{y} = (1 - \varepsilon)\, y + \varepsilon / K$, $\varepsilon = 0.05$; easy events count less.

Gradient clipping
Pascanu+ 2013

$$g \leftarrow g\, \min\big(1,\; c / \|g\|_2\big)$$

$c = 1$; guards against rare large-weight batches.

Adopted in $t\bar{t}H(c\bar{c})$, $t\bar{t}H(b\bar{b})$ and $VH(c\bar{c}/b\bar{b})$: ~50% gains, much more stable training.

CMOK: classifiers from signed Monte Carlo weights

First author · arXiv soon | Invited talk: Higgs workshop

Problem

- Signed weights, common in NLO MC, make the weighted CE training loss unbounded below: one isolatable negative-weight event suffices.

Conditional Mean on Observed Kinematics

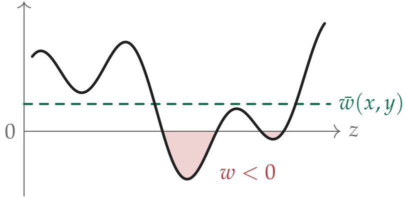
- Factorise the problem: learn $q_\theta(y \mid x)$ with *unweighted* CE and $\hat{w}_\phi$ by MSE on the signed w, then compose:

$$p(y \mid x) = \frac{\bar{w}(x,y)\, q(y \mid x)}{\sum_{y'} \bar{w}(x,y')\, q(y' \mid x)}$$

- Proven in the paper by marginalising over hidden variables: for the posterior, the mean weight is all you need.

Signed weights, positive mean

$w(x,y,z) = f/q$ at fixed (x,y)



Second analysis: $Z{+}c$ and $\sin^2\theta_W$ reweighting

$Z{+}c$ measurement, with M. Muškinja

– A second analysis, independent of $t\bar{t}H(H\to c\bar{c})$.

Reweighting, not regenerating: response to $\sin^2\theta_W$

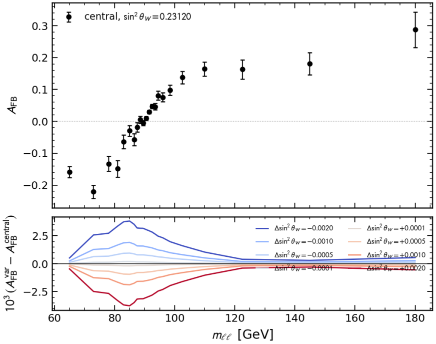
– **Matrix element**: same phase-space point, new coupling,

$$w_i(\theta) = w_i\,|\mathcal{M}_\theta(\Phi_i)|^2/|\mathcal{M}_{\theta_0}(\Phi_i)|^2.$$

– **ML at reconstruction level**: a classifier learns the density ratio, standing in for a new detector simulation,

$$\hat{r}(x;\theta) \simeq f_\theta(x)/f_{\theta_0}(x), \qquad w_i \to w_i\,\hat{r}(x_i;\theta).$$

No rerun of the detector simulation; made possible by my signed-weight method (CMOK).



Top: $A_{\mathrm{FB}}(m_{\ell\ell})$ at $\sin^2\theta_W = 0.23120$. Bottom: reweighted shifts; $\Delta\sin^2\theta_W = \pm0.002$ moves A_{FB} by $\mp4\times10^{-3}$ near 85 GeV.

ATLAS service work in PMG: MadGraph expert, central samples, bug fixes

Talk: PMG workshop | Talk: ATLAS Week

Production-facing impact

- Established **MadGraph expert status** inside ATLAS.
- Main technical work on **V+jets central samples**; also several Higgs signal-sample workflows.
- Fixed and remade broken **V+jets τ-samples** and Higgs samples.
- Generator–Athena maintenance and infrastructure improvements, including ARM compilation support.

MadGraph expert team

> *"In light of your broad contributions to MadGraph work, we would like to formally invite you to join the ATLAS team of MadGraph experts."*
>
> — ATLAS PMG, invitation

Critical FxFx bug

> *"Fixed critical bug in FxFx merging, introduced in version 3.3.1. This resulted in too many negatively weighted events, and too many high-p_T jets. Particularly relevant for merging with up to 2 or 3 jets at NLO. Thanks to Lenart Jerala for pointing this out and testing the fixes."*
>
> — MG5_aMC@NLO release notes

EW Sudakov reweighting – now HEP theory

First author · arXiv by end of week | Invited talk: MadGraph developers meeting | Talk: PMG workshop

Problem

- State-of-the-art MC is NLO QCD (NNLO for colour singlets); no general NLO EW matching to the parton shower exists.
- Better predictions are needed for HL-LHC precision and the FCC.
- ME+PS generators face an uncertain future, yet they are a cornerstone of every HEP experiment.

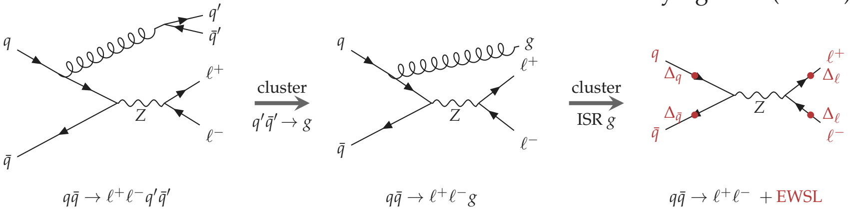
Solution: EWSL reweighting

- Above the EW scale, Sudakov logarithms dominate NLO EW; evaluate them at tree level, event by event:

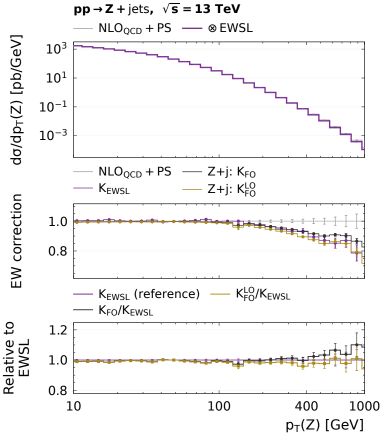
$$w = 1 + 2\,\mathrm{Re}\big(\mathcal{M}_B^*\,\delta\mathcal{M}_B\big)/|\mathcal{M}_B|^2,$$

$$\delta \sim \frac{\alpha}{4\pi}\left(\ln^2\frac{|r_{ij}|}{M_W^2} + \ln\frac{|r_{ij}|}{M_W^2}\right).$$

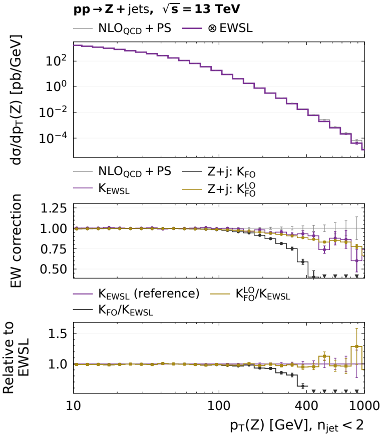
- They need $|r_{ij}| > M_W^2$: cluster the diagram back to its underlying Born (below).



EW Sudakov reweighting: FxFx $Z{+}{\leq}2j$ validation



Inclusive $p_T(Z)$



$p_T(Z)$ with a second-jet veto

Results

- $p_T(Z)$: -5% at 300 GeV, -15% at 1 TeV, following the LO-normalised fixed order (gold).
- With the veto, EWSL tracks $K_{\mathrm{FO}}^{\mathrm{LO}}$; the additive K_{FO} breaks down as its fixed-order denominator goes to zero.
- Q_{cut} 20–40 GeV: few %; ξ_{clust} 1–2: $< 1\%$.

Resonance-aware EW Sudakov weight: spin density matrices

Why

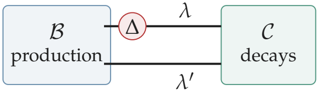
– A clustered resonance (W, Z, top) has a helicity-dependent Sudakov correction, which a spin-summed factor averages away.

How: spin density matrices

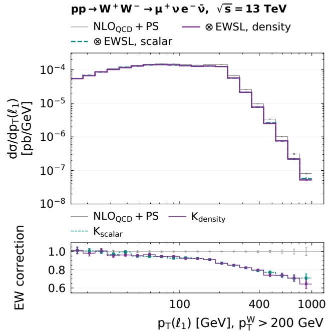
– Correct the production amplitude per helicity, $\mathcal{M}_\lambda \to (1+\Delta_\lambda)\,\mathcal{M}_\lambda$; keep the decays as generated.

– Contract production and decay density matrices: the event's spin correlations are reweighted consistently.

$$w = \frac{\mathrm{Tr}\left[\mathcal{B}^{\mathrm{EWSL}}\,\mathcal{C}\right]}{\mathrm{Tr}\left[\mathcal{B}\,\mathcal{C}\right]}$$



Validation: $W^+W^- \to \mu^+\nu\, e^-\bar{\nu}$



$p_T(\ell_1)$, $p_T^W > 200\,\mathrm{GeV}$: density (purple) and scalar (teal) agree within errors, also in φ^*: the extra helicity dependence is small here.