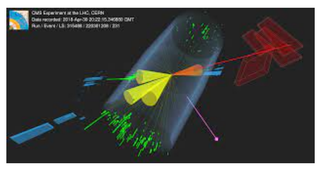
# Generating parton-level events from CMS reconstructed events with Conditional Normalizing Flows

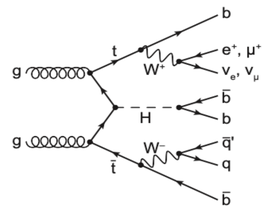
Antonio Petre on behalf of the CMS collaboration

## Matrix Element Method (MEM)

Matrix element method estimates the probability of a single reconstructed event $\overrightarrow{Y}$ to be generated by a physical process defined by $\theta$ parameters:

$$\mathscr{P}(\overrightarrow{Y}\,|\,\theta) \propto \int_{\phi} d\overrightarrow{X}\; M(\overrightarrow{X}\,|\,\theta)|^2 \cdot Pdf \cdot W(\overrightarrow{Y}\,|\,\overrightarrow{X})$$

$\overrightarrow{Y}$ = 

$\overrightarrow{X}$ = 


Pros & Cons:

- ✓ It can be used for hypothesis testing or parameter estimation
- ✓ Maximizes the amount of theoretical information for the discriminator
- ✓ It is not bound to a specific process
- x Integral computation is very CPU demanding due to jet-parton matching (combinatorial problem)
- x Many approximations used to speedup the computation e.g. jet-parton alignment

## New Method & Normalizing Flows

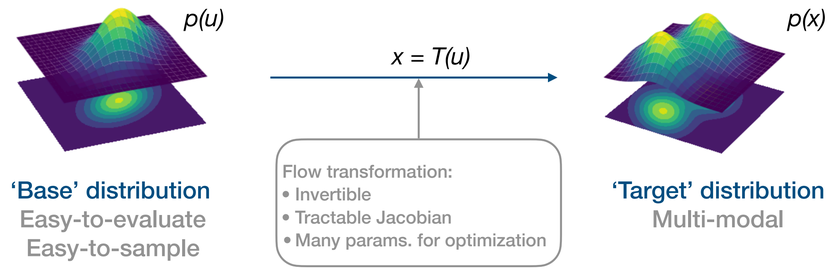
Our goal is to model the conditional probability of parton-level events given a reconstructed event using generative machine learning architectures, more specifically **normalizing flows:**

$$\int_{\phi} d\overrightarrow{X}\; M(\overrightarrow{X}\,|\,\theta)|^2 \cdot Pdf \cdot W(\overrightarrow{Y}\,|\,\overrightarrow{X})$$

Use importance sampling: $\overrightarrow{X} \sim \mathscr{P}(\overrightarrow{X}\,|\,\overrightarrow{Y},\theta)$

$\mathscr{P}(\overrightarrow{X}\,|\,\overrightarrow{Y},\theta)$ found using **normalizing flows**

Flow models: Machine-learned maps (transformations) between probability distributions




## Dataset and Legacy Performance

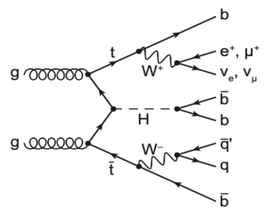
Process used: **single-lepton channel in the ttH(→bb) with an additional radiation**

Events details:

- Generated with a CM energy of 13 TeV
- Pileup profile of LHC Run II (~30-50 simultaneous pp collisions)
- Full CMS detector simulation, including standard RUN II reconstruction

Data selection:

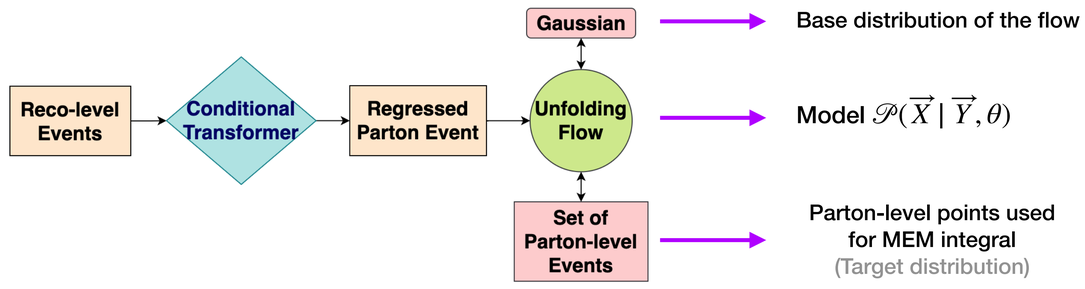
- At least 4 jets with $p_T > 30$ GeV and $|\eta| < 2.4$
- At least 3 jets tagged as b-quark flavour
- One prompt reconstructed lepton with $p_T > 30$ GeV
- MET > 20 GeV




Legacy ttH analysis: **MEM used for discriminating between signal and background**

Computation performance: **~ 1 min/event → speedup needed**
**normalizing flow is a good candidate**

## Our Strategy



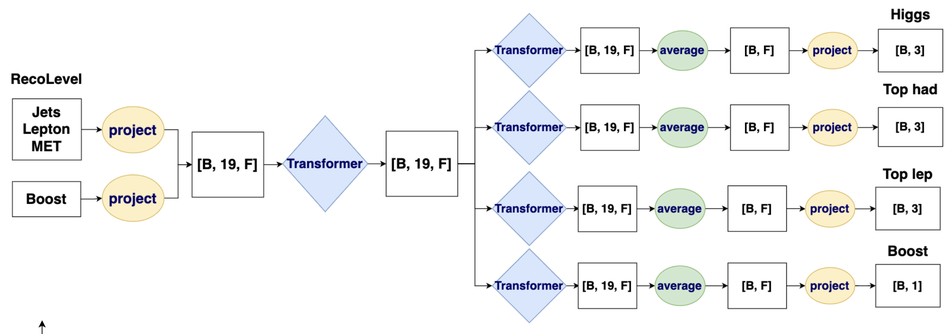

**Reco-level Events** → Jets + Lepton + MET : $p_T$, $\eta$, $\phi$, btag score, SPANET output

**Regressed Parton Event** → Higgs + two tops + additional radiation : $p_T$, $\eta$, $\phi$
In practice, RAMBO parametrization was used

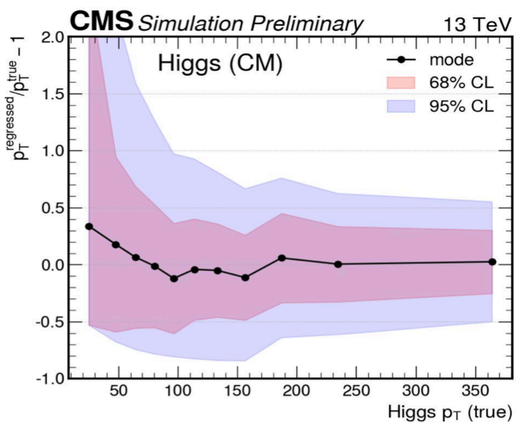
**Conditional Transformer** → Regress the **parton-level event** for a given **reco-level event**
Extracts a latent information vector which conditions the **Unfolding Flow**

**Unfolding Flow** →
- ✓ Generates plausible phase-space points compatible with reco-objects
- ✓ Reduces assumptions on partons' directions
- ✓ Handles events with out-of-acceptance final state objects and multiple jet multiplicities

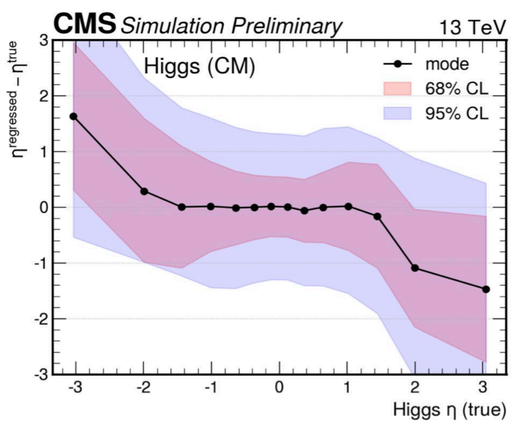
## Conditional Transformer




- The main building block is the **Transformer Encoder**
- The training data → partons or reco-objects in the laboratory frame
- The model was pretrained using modified differential multiplier method **(MDMM):**
  - **Main loss $L_0$** is the Huber loss for partons and boost pz
  - **Second loss $L_1$** is the maximum mean discrepancy (MMD) loss → keep distributions coherent
  - **MDMM:** minimize $L_0$ ensuring $L_1 \leq \epsilon$




Bias mode of the regression for Higgs $p_T$



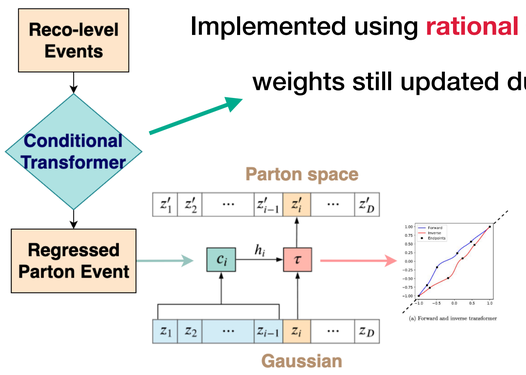

Bias mode of the regression for Higgs $\eta$

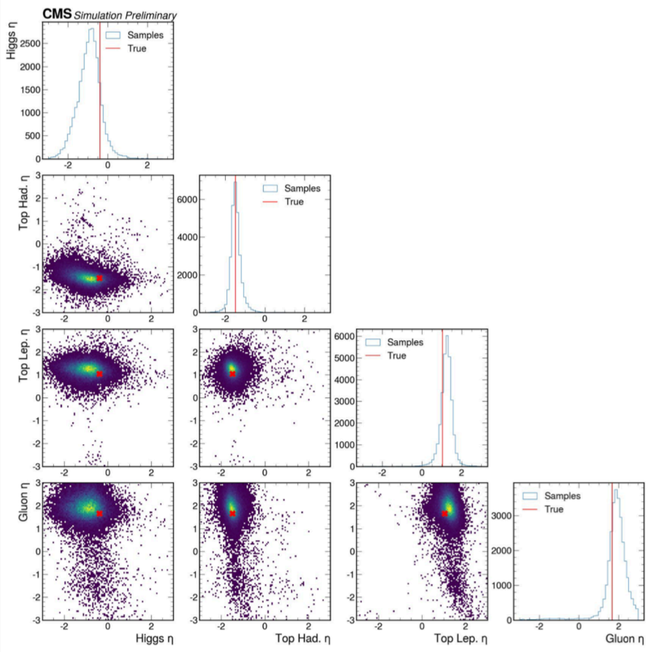
## Unfolding Flow

Reco-level Events → Conditional Transformer → Regressed Parton Event → Unfolding Flow

Implemented using **rational quadratic splines (RQS)** with **autoregressive blocks**

weights still updated during **Unfolding Flow** training

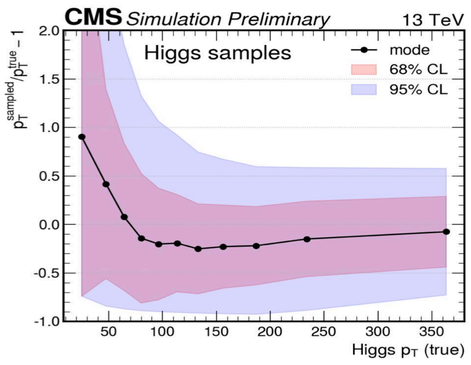


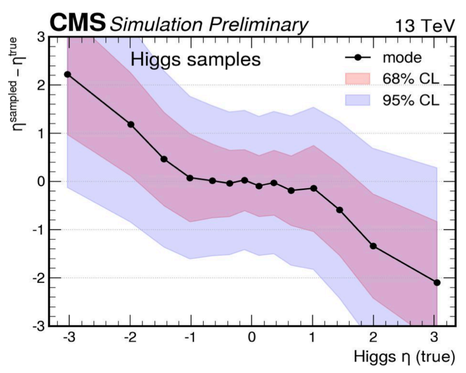




Sample 20k parton-level events for one reco-event
**Super fast (less than 1 second)**

Training:

- **maximum likelihood:** evaluate the density of the true partons from the signal MC
- **sampling parton-level events** from the flow and comparing them with the target




Sample 30 parton events for 1.5M reco-events
**Check the quality of the sampled partons**

**CMS Public Note DP-2023-085**

Contact: Antonio Petre (adrian.antonio.petre@cern.ch)