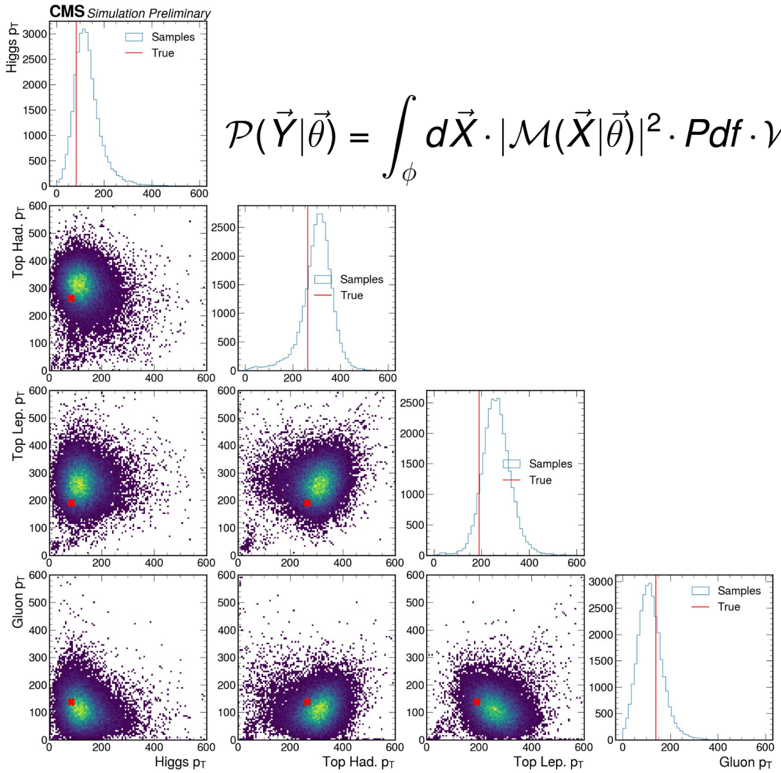
Computing optimal observables from the matrix element method with conditional normalizing flows

Davide Valsecchi (ETH Zurich)
for the CMS Collaboration

CMS-DP-2023-85

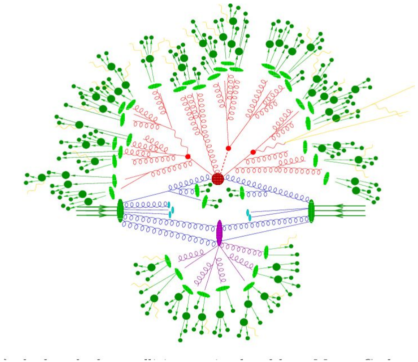
$$\mathcal{P}(\vec{Y}|\vec{\theta}) = \int_{\phi} d\vec{X} \cdot |\mathcal{M}(\vec{X}|\vec{\theta})|^2 \cdot Pdf \cdot \mathcal{W}(\vec{Y}|\vec{X})$$



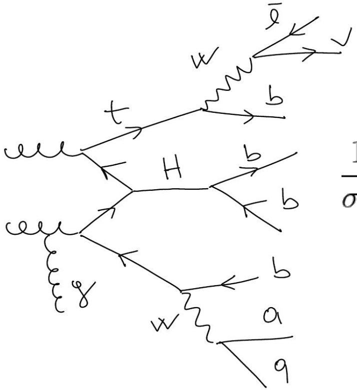
ETH zürich

Intro and motivation

Hadronization/showering

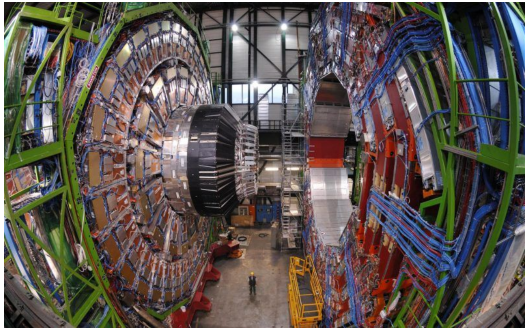


QFT

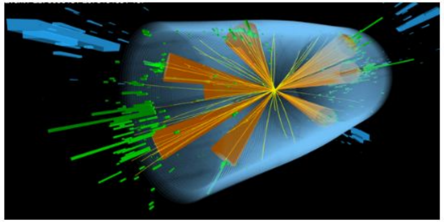


$$\frac{1}{\sigma_\theta}\frac{d\sigma_\theta}{dz_{parton}}$$

Detector



Reconstruction



Interpretation

- HEP relies on a complex chain of high fidelity simulators to model the signals in our detectors

  → full likelihood of an event *x* is **intractable** $p(x|z) = \int \mathrm{d}z_{\text{detector}} \int \mathrm{d}z_{\text{shower}}\, p(x|z_{\text{detector}})\, p(z_{\text{detector}}|z_{\text{shower}})\, p(z_{\text{shower}}|z)$

- **The Matrix Element Method is a way to link directly the detector-level info with the theory**

The matrix element method master formula

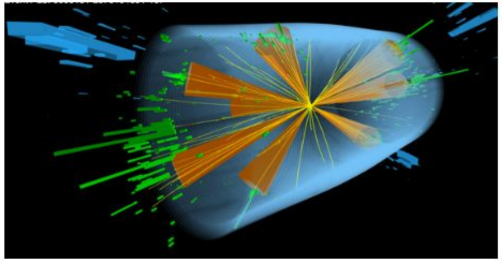
arxiv:1101.2259

Estimation of the probability that a reconstructed event **y,** is generated by a physical process defined by **θ** parameters.

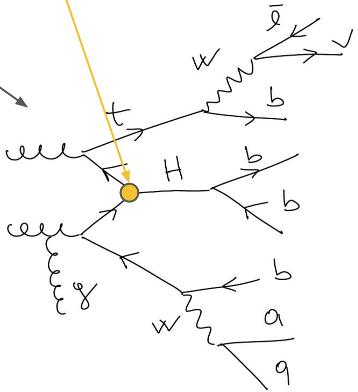
Prob. of the reco event hard-scattering transfer function

$$\mathcal{P}(\vec{Y}|\vec{\theta}) = \int_{\phi} d\vec{X} \cdot |\mathcal{M}(\vec{X}|\vec{\theta})|^2 \cdot Pdf \cdot \mathcal{W}(\vec{Y}|\vec{X})$$

Reconstructed level observables



Leptons, jets, MET



Integrate over all parton-level phasespace

A transfer function *W(Y|Z)* models the probability that the reconstructed event parton-level configuration

Example: MEM for EFT

$$\mathscr{P}(\vec{x}_{\text{reco}}\,|\,\vec{\theta}) \propto \mathscr{P}_0(\vec{x}_{\text{reco}}) + \sum_k \left(\frac{2\theta_k}{\theta_0}\right) \boxed{\mathscr{P}_{0k}(\vec{x}_{\text{reco}})} + \sum_k \left(\frac{\theta_k}{\theta_0}\right)^2 \mathscr{P}_k(\vec{x}_{\text{reco}}) + \sum_{i<j} \left(\frac{2\theta_i\theta_j}{\theta_0^2}\right) \mathscr{P}_{ij}(\vec{x}_{\text{reco}})$$

SMEFT quadratic parametrization of dim-6 Wilson coefficients θ effects

$$\mathcal{P}(\vec{Y}|\vec{\theta}) = \int_\phi d\vec{X} \cdot |\mathcal{M}(\vec{X}|\vec{\theta})|^2 \cdot Pdf \cdot \mathcal{W}(\vec{Y}|\vec{X})$$

Optimal observables for the linear and quadratic terms

$$\mathcal{R}_{\text{opt},1} = \frac{2\mathcal{P}_{01}(\vec{x}^{\text{full}}_{\text{reco}})}{\mathcal{P}_0(\vec{x}^{\text{full}}_{\text{reco}}) + c \cdot \mathcal{P}_1(\vec{x}^{\text{full}}_{\text{reco}})},$$

$$\mathcal{R}_{\text{opt},2} = \frac{\mathcal{P}_1(\vec{x}^{\text{full}}_{\text{reco}})}{\mathcal{P}_0(\vec{x}^{\text{full}}_{\text{reco}}) + c \cdot \mathcal{P}_1(\vec{x}^{\text{full}}_{\text{reco}})},$$

LHC EFT WG Group2 report: arxiv

- **Goal**: compute each term of the likelihood expansion with the MEM

  → Encapsulate the full event kinematic information in a single likelihood ratio

- Extract limit on Wilson coefficients with optimal observables with a unbinned or standard binned fit to CMS data

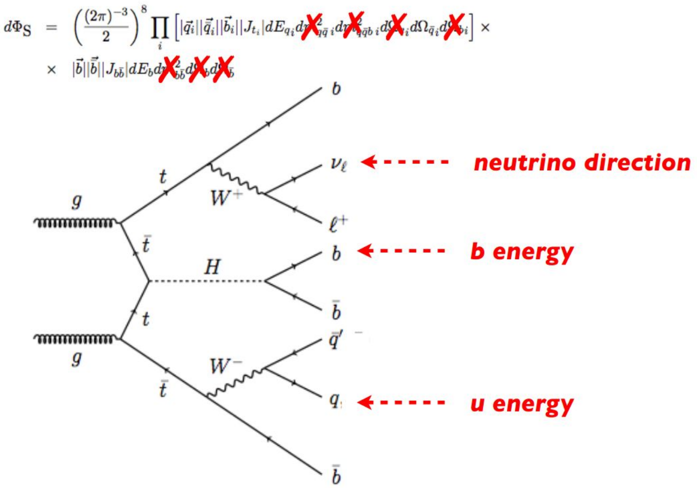
LHC EFT WG Group2 report: arxiv

Classical MEM computation

1. consider **generator-level particles as aligned to jets**
   a. > jets than needed → all possible permutations and sum
   b. < objects than needed → discard the event or integrate out the missing partons
2. Assume full reconstruction, particles mass on-shell
3. Compute the numerical integral with VEGAS over the free d.o.f.
4. **Permute** the jet-parton assignment → **extremely time consuming**

Example of MEM for ttH(bb) process

From Bianchini's talk at DataScience@LHC 2015



 # variables = 24 - 15 - 5 = 4

alignment

mass on-shell

A new approach based on generative ML

Model the conditional probability of parton-level events given a reconstructed event using **normalizing flows**

General **strategy** based on:

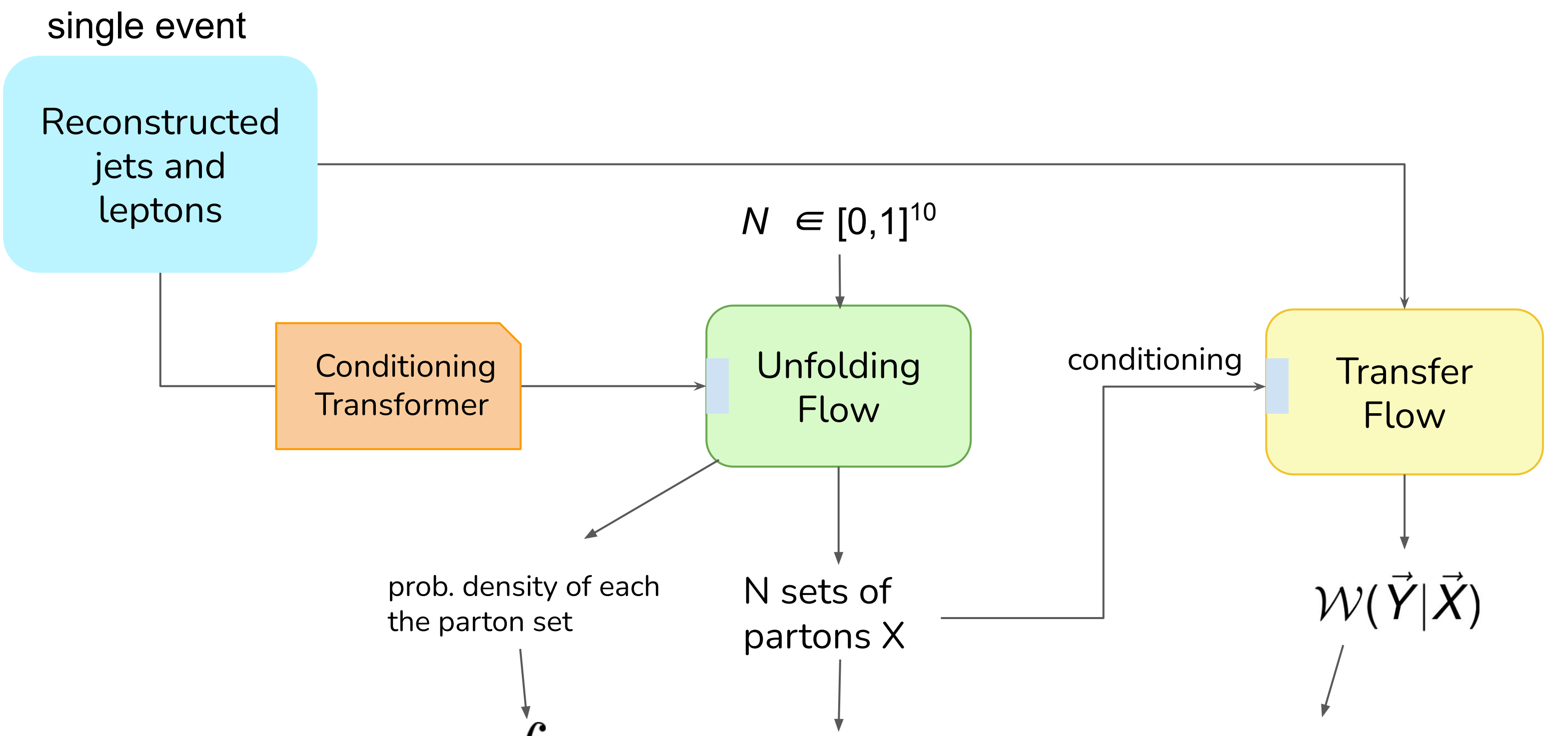
- *Precision-Machine Learning for the Matrix Element Method* 2310.07752 (see next talk!)
- *Two Invertible Networks for the Matrix Element Method* 2210.00019
- *Invertible Networks or Partons to Detector and Back Again* 2006.06685

First application of the method on the **experiment side:**

- **Complex ttH(bb) channel** in the semileptonic final state (1 lepton, >6 jets):
  - MEM used for sig/bkg discrimination in CMS analyses: doi:10.1007/JHEP03(2019)026, CMS-PAS-HIG-19-011
- Full **CMS detector simulation** including pileup
- **All jet multiplicities** considered

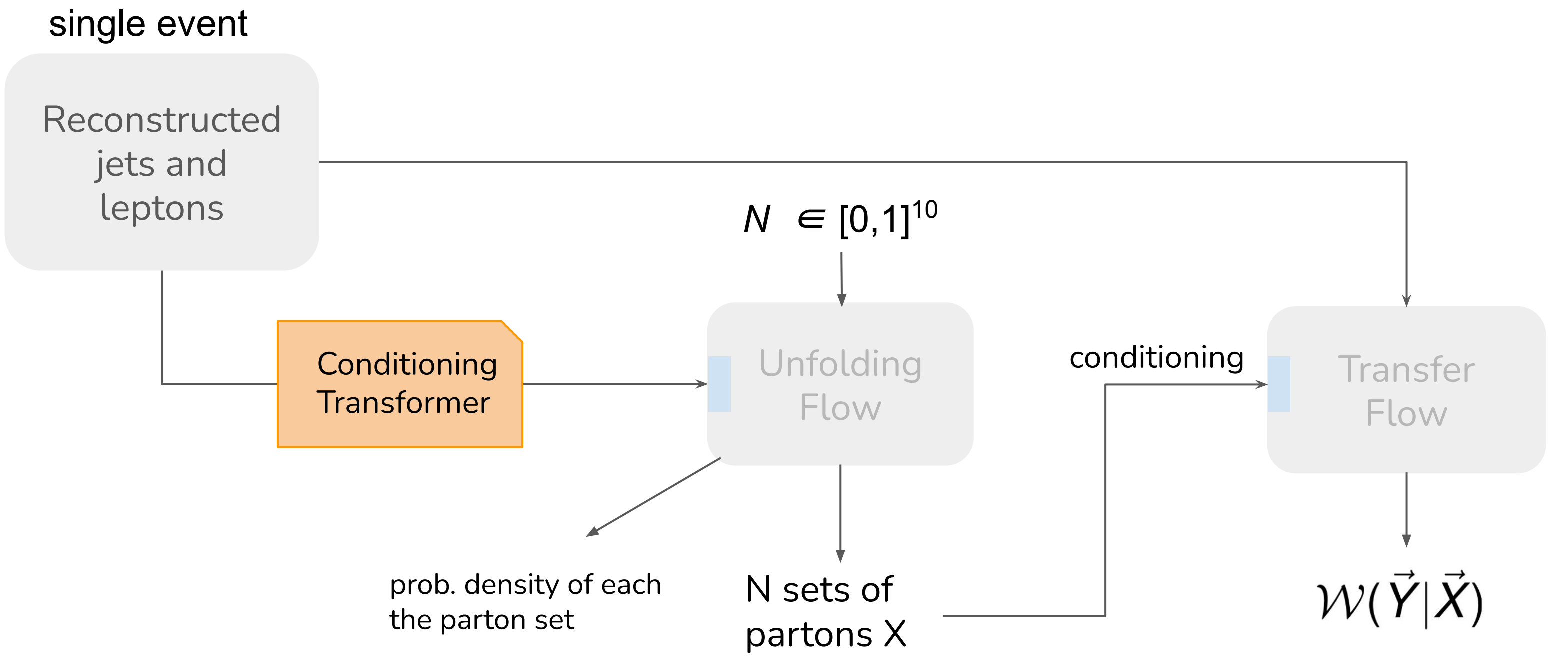
Today talk: generating generator-level events for MEM integration

MEM computation



$$\mathcal{P}(\vec{Y}|\vec{\theta}) = \int_{\phi} d\vec{X} \cdot |\mathcal{M}(\vec{X}|\vec{\theta})|^2 \cdot Pdf \cdot \mathcal{W}(\vec{Y}|\vec{X})$$

Conditioning transformer

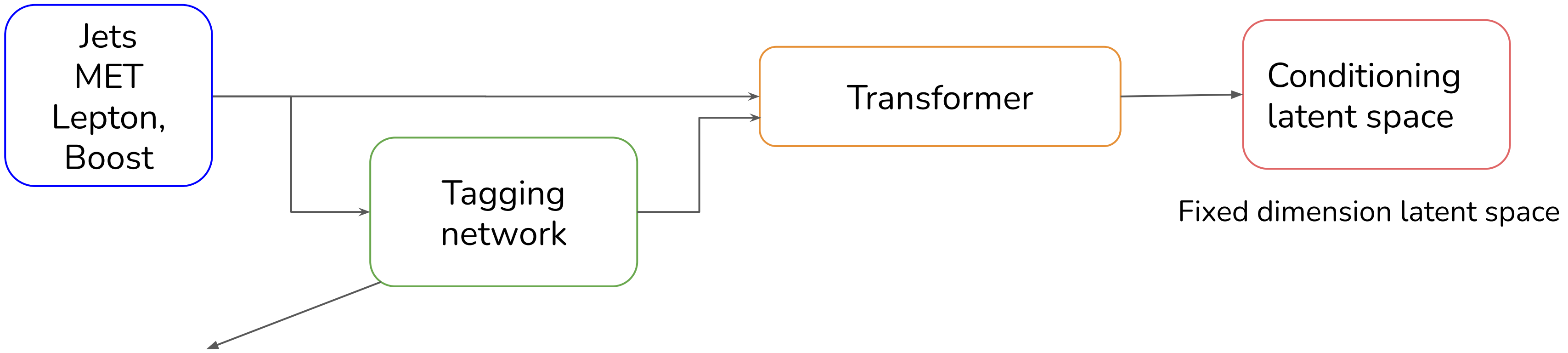


Conditioning on reconstructed events

- Sampled particle sets for the MEM integral computation **strongly depends** on the **reconstructed objects.**

- Use a **transformer** to extract a fixed-size conditioning latent space for the unfolding flow
  → can handle additional radiation and missing objects
  → avoids direct jet-parton combination

- The conditioning latent space should be correlated with the most probable partons



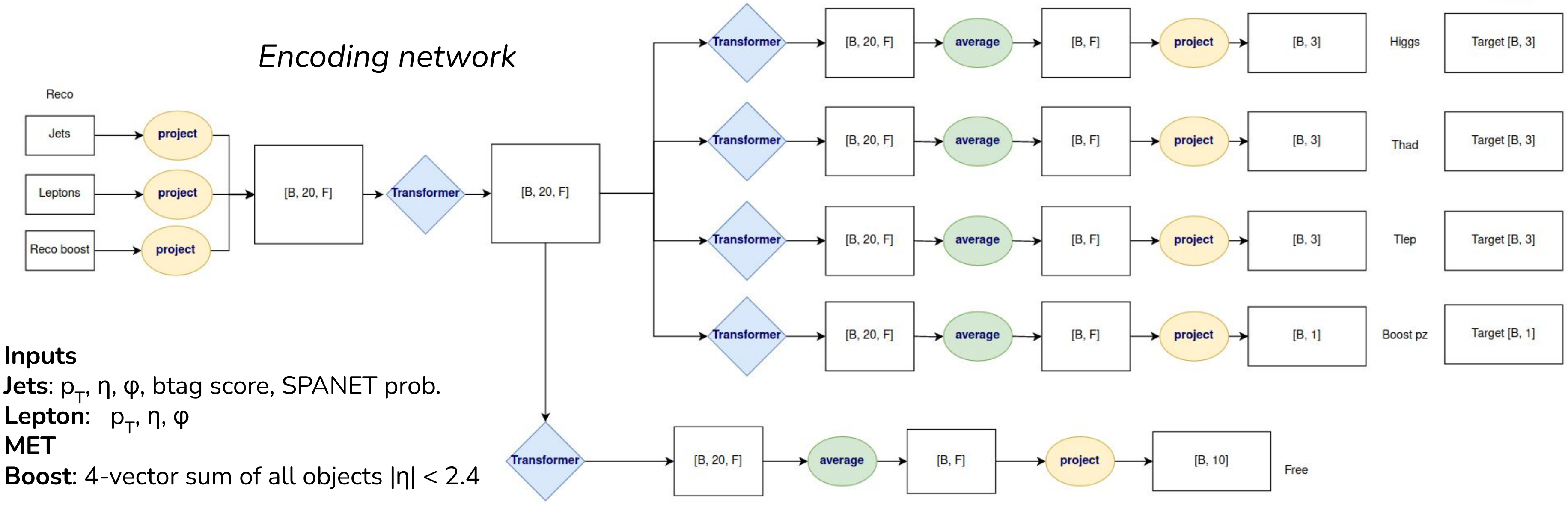
Using a pretrained **SPANET network** and adding to each jet the probability to be generated from H, top hadronic, top leptonic

Conditioner pre-training

- **Idea**: pretrain the conditioning transformer with a **regression** of the **generator-level particles**: higgs, top_{had}, top_{lep} (p_T, η, φ) + total event boost p_Z
  - → additional radiation (gluon) computed from momentum balance
  - → maximize the correlation with the target of the unfolding flow

Decoding network

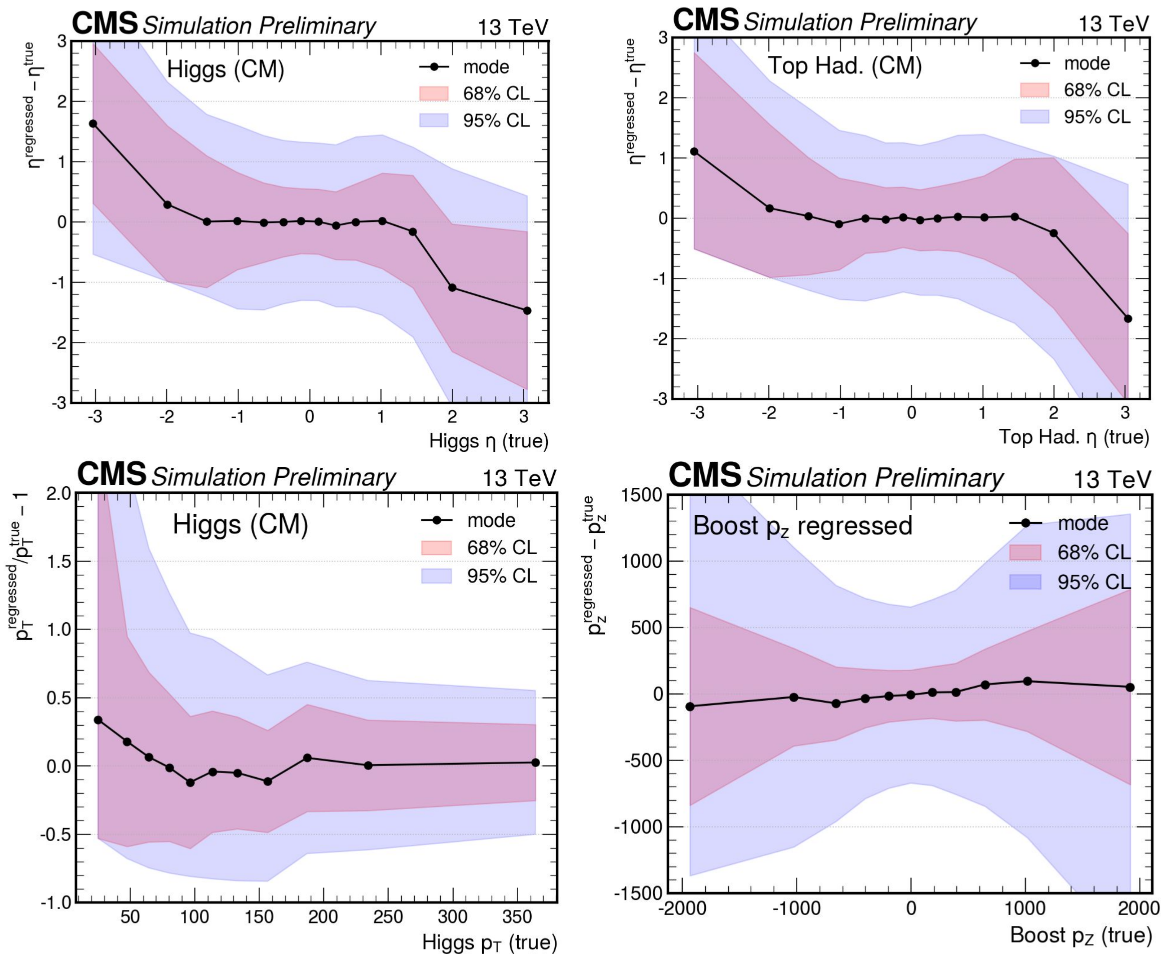
Encoding network



Inputs
Jets: p_T, η, φ, btag score, SPANET prob.
Lepton: p_T, η, φ
MET
Boost: 4-vector sum of all objects $|\eta| < 2.4$

free latent space, not constrained in the pretraining

Parton regression performance

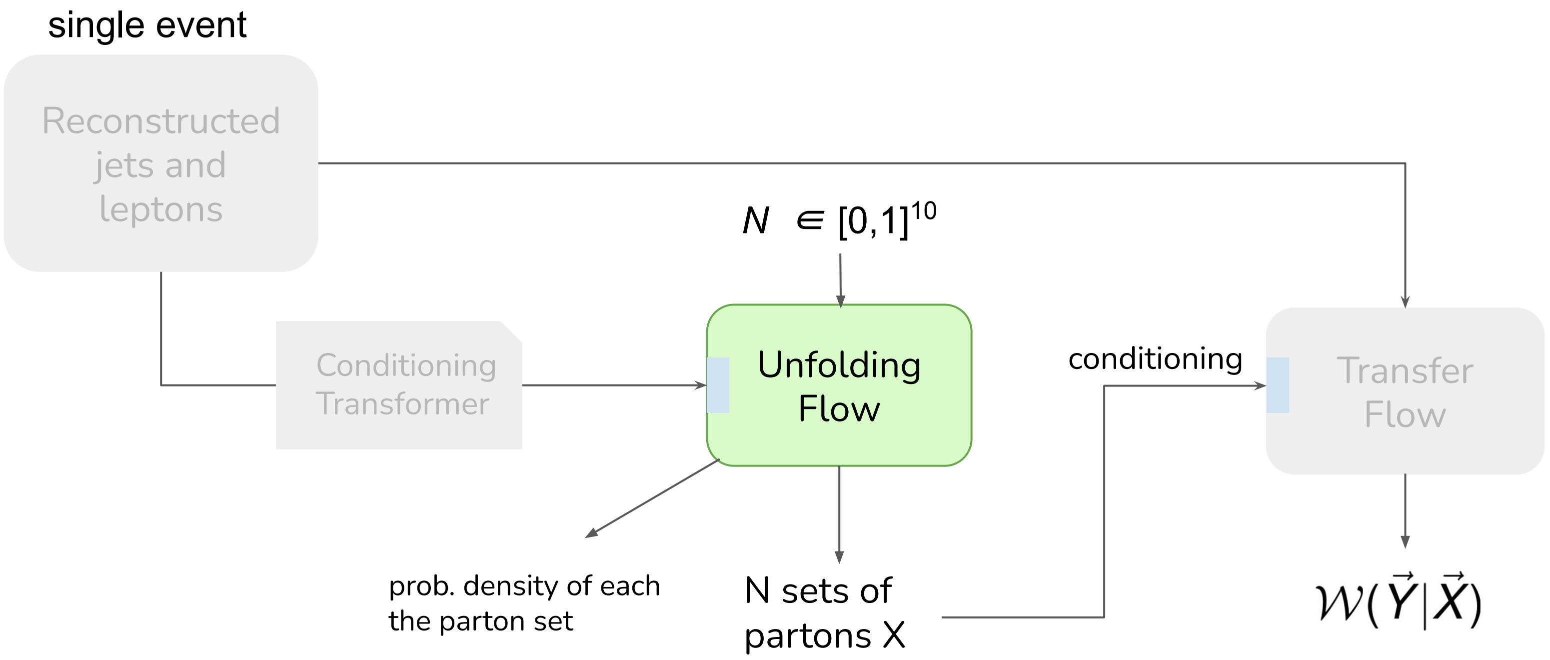


The regression of the generator-level particles is overall unbiased

Also the total p_Z of the event is well regressed
→ the particles can be boosted in the centre-of-mass (CM) correctly.

*More plots in backup

Unfolding flow



Phase space representation

Need to parametrize 4-momenta of particles in an event in a way suitable to be modeled by normalizing flows.

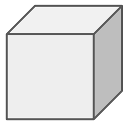
Using the **RamboOnDiet** 1308.2922 algorithm → analytical mapping, compact and fast

- parametrizes events in the **CM frame**
- **D = 3N-4 numbers ∈ [0,1]** to describe N particles final state
- almost flat phase-space density (for massless particles)
- **2 additional numbers** for the momentum fractions of the scattering partons x_1 x_2

Also look at *Enhanced latent spaces for improved collider simulations* 2305.07696

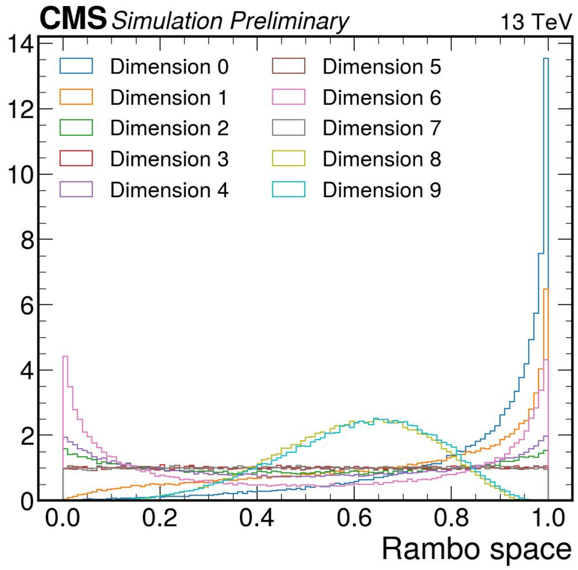
$8D + x_1 x_2$

10D

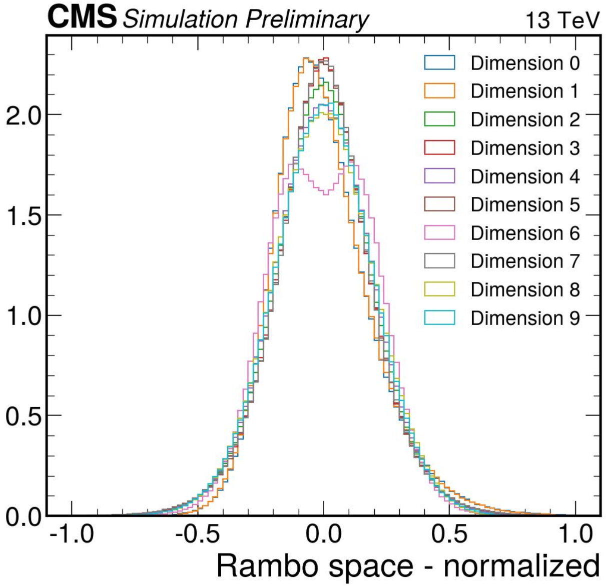


[0,1]

ttH(bb) + 1 gluon
semileptonic channel



logit + std. scaling

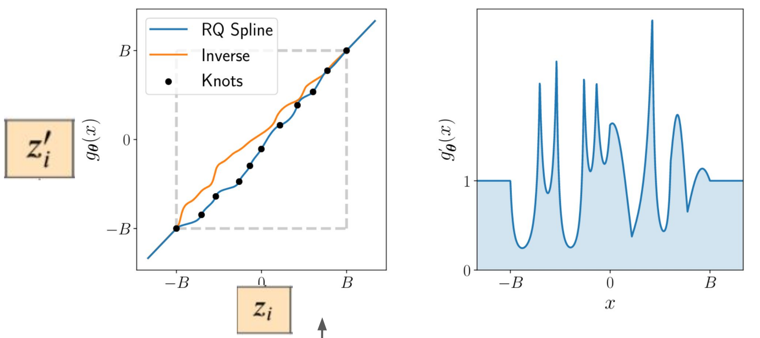


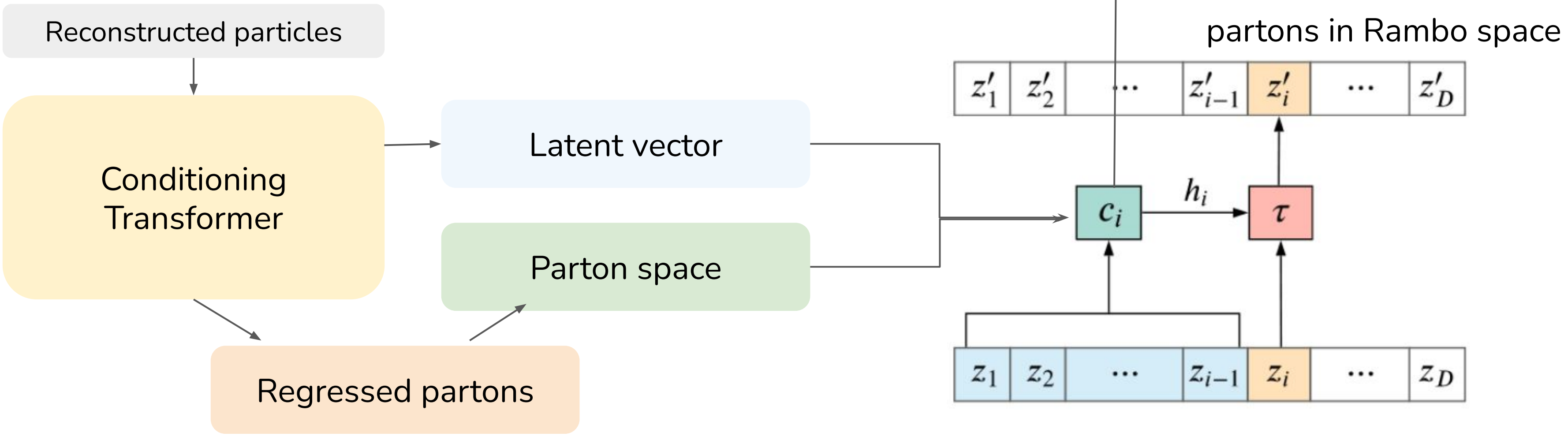
Flow architecture

Implemented the flow as Rational Quadratic Spline [1906.04032](https://arxiv.org/abs/1906.04032) flow with autoregressive blocks.

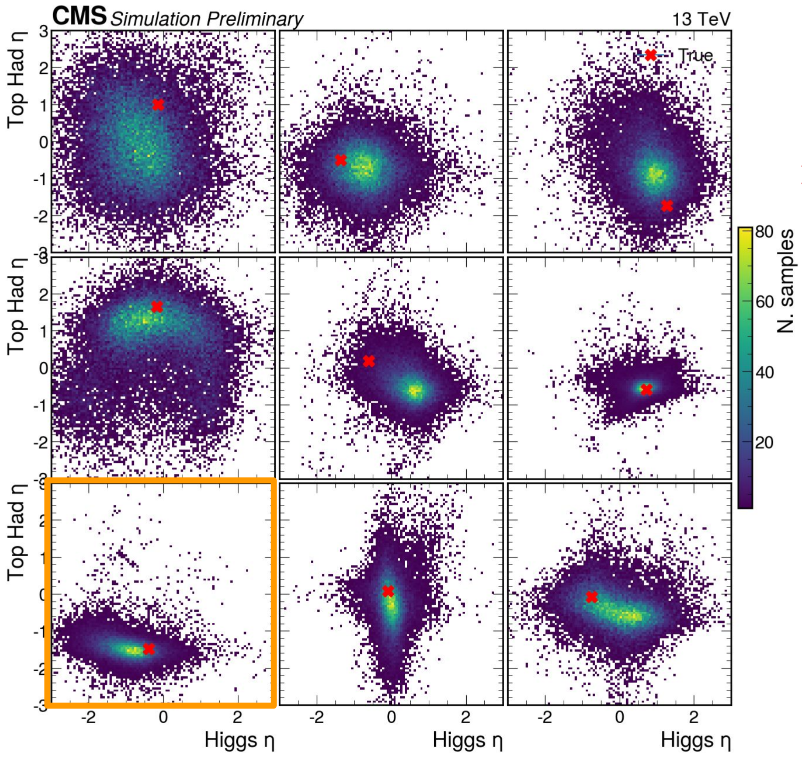
The conditioning vector is made up from the regressed partons converted to Rambo space and a free latent vector.

Illustration of flow splines from 1906.04032





Unfolding flow samples

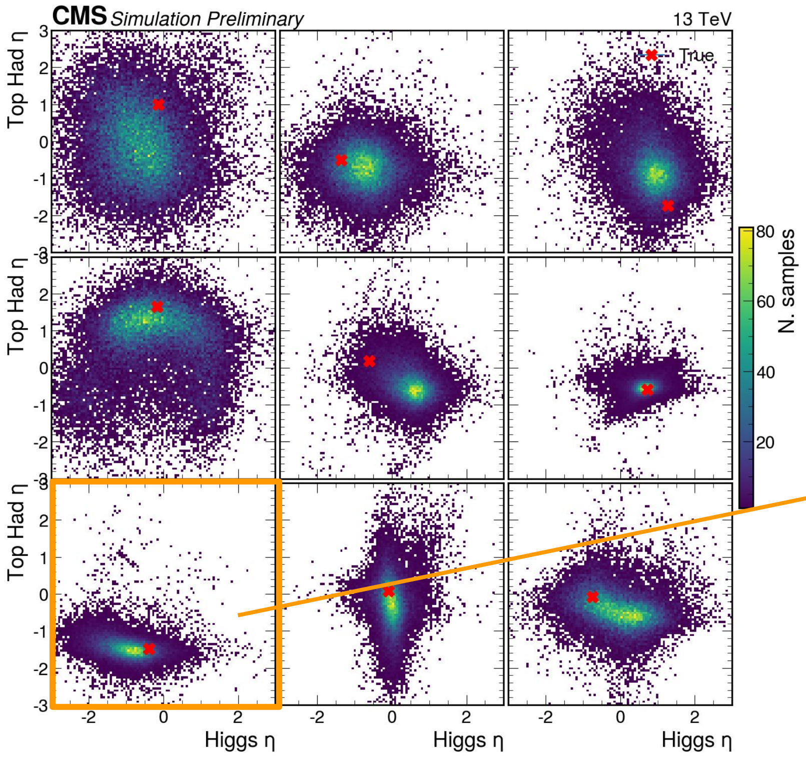


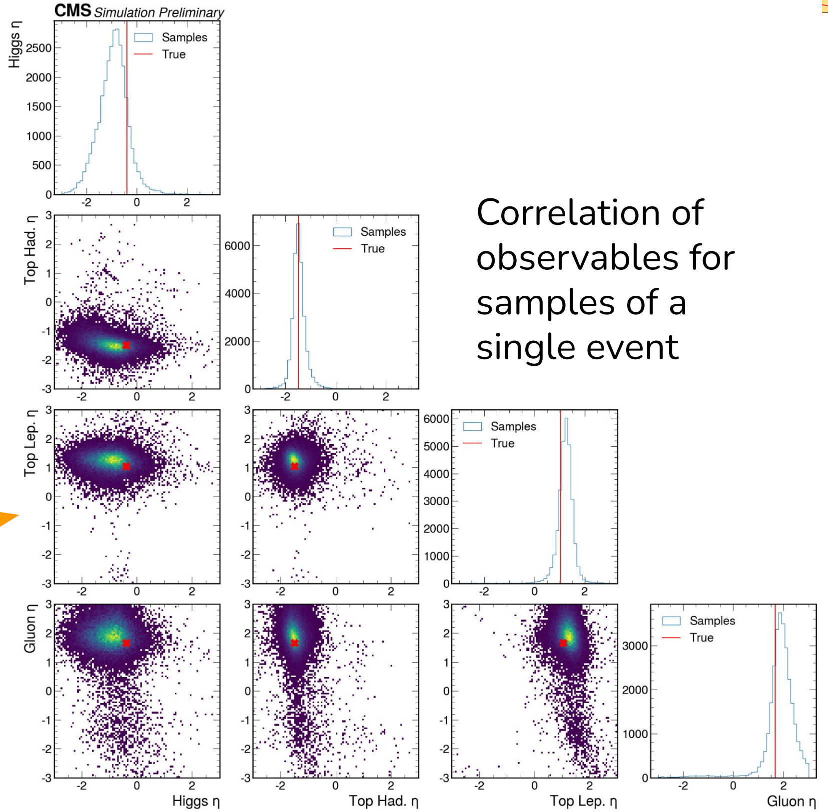
1 event in each box
20k samples

True gen-level particles

The unfolding flow learns the conditional probability ***P(partons|reconstructed event)*** and generates partons in the most probable configurations.

Unfolding flow samples

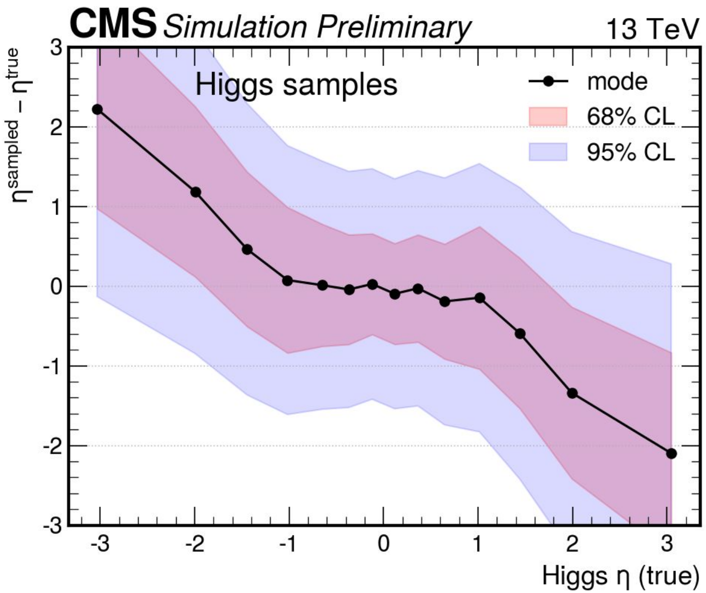


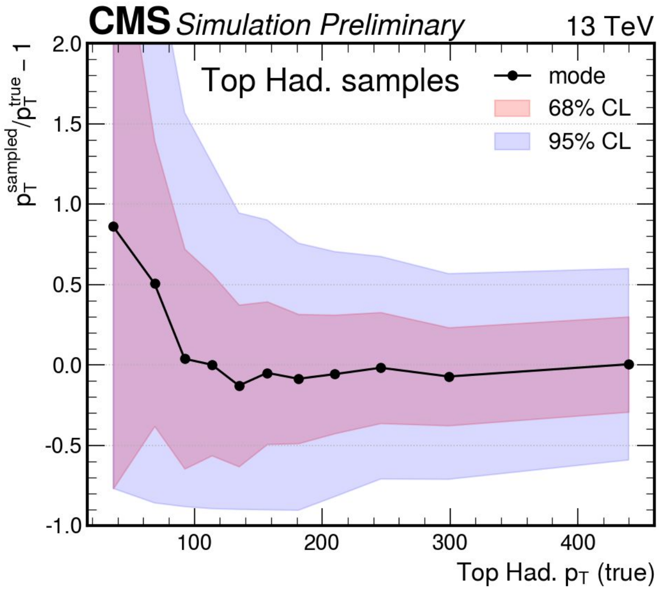


Correlation of observables for samples of a single event

The unfolding flow learns the conditional probability ***P(partons|reconstructed event)*** and generates partons in the most probable configurations.

Quality of the sampled partons





The quality of the sampled partons configuration is evaluated by sampling 30 sets per 1.5M events and analyzing the bias w.r.t of the target true partons.

The bias of the sampled particles is on average small in the bulk of the distributions and increases in the tails.

Overall the performance is very correlated with the one of the **regression** for the conditioning network.

Conclusions

→ First steps for the full-fledged application on **CMS simulation** of the novel strategies (2210.00019,2006.06685) for the Matrix Element Method computation with generative models on a **complex final state**, like ttH(bb)

→ The **full reconstructed event information** is analyzed by a **transformer** to regress the 4-momenta of the generator-level particles. This information is then used to generate set of particles with a **normalizing flow**.

→ The quality of the generative model is promising and it can be further optimized

→ The next step is to model the transfer function W(reco|hard)

→ All plots and public results on CMS-DP-2023-85

Questions?

Backup

References

1. Matrix Element Method in HEP: Transfer Functions, Efficiencies, and Likelihood Normalization 1101.2259
2. Phase Space Sampling and Inference from Weighted Events with Autoregressive Flows 2011.13445
3. i-flow: High-dimensional Integration and Sampling with Normalizing Flows 2001.05486
4. MadNIS – Neural Multi-Channel Importance Sampling 2212.06172
5. Normalizing Flows: An Introduction and Review of Current Methods 1908.09257
6. Normalizing Flows for Probabilistic Modeling and Inference 1912.02762
7. Masked Autoregressive Flow for Density Estimation 1705.07057
8. Neural Spline Flows 1906.04032
9. RAMBO on diet 1308.2922
10. LHC EFT WG Report: Experimental Measurements and Observables 2211.08353
11. Invertible Networks or Partons to Detector and Back Again 2006.06685
12. Two Invertible Networks for the Matrix Element Method 2210.00019
13. Modified Differential Multiplier Method paper
14. Measurement of the ttH and tH production rates in the H→bb decay channel with 138fb−1 of proton-proton collision data at $\sqrt{s}$=13 TeV CMS-PAS-HIG-19-011
15. Attention is all you need 1706.03762
16. SPANet: Generalized Permutationless Set Assignment for Particle Physics using Symmetry Preserving Attention 2106.03898
17. POWHEG: NLO calculations in shower Monte Carlo P. Nason, *JHEP* 0411 (2004) 040, hep-ph/0409146 [paper]
18. An Introduction to PYTHIA 8.2 1410.3012

Conditioner transformer training

The pretraining regression is performed with multiple components:

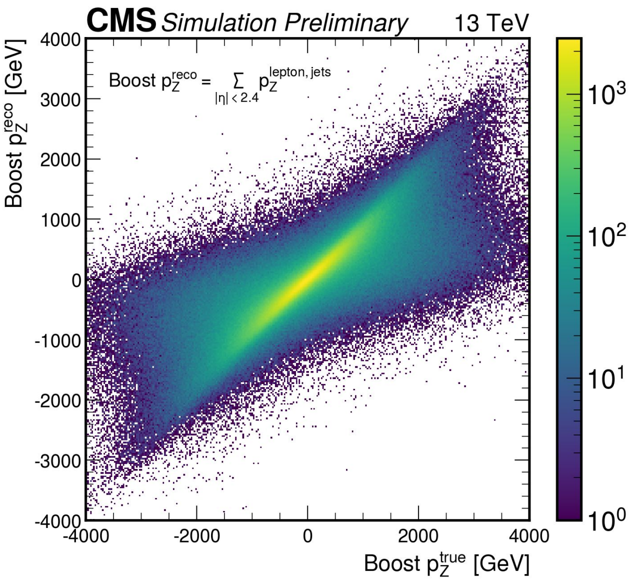
- **Huber** loss for single particles + event boost p_z + computed gluon
- **MMD** loss for single particles + full 13D space → keeps distributions coherent

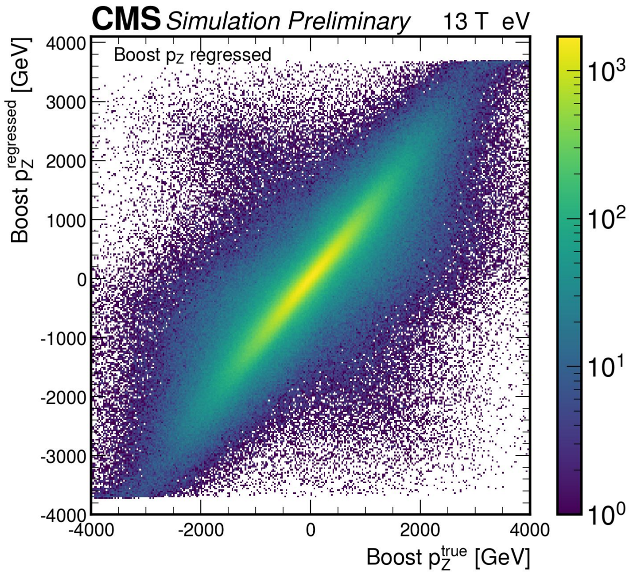
Trained using the Modified Differential Multiplier Method (MDMM) technique to balance the difference losses (paper)

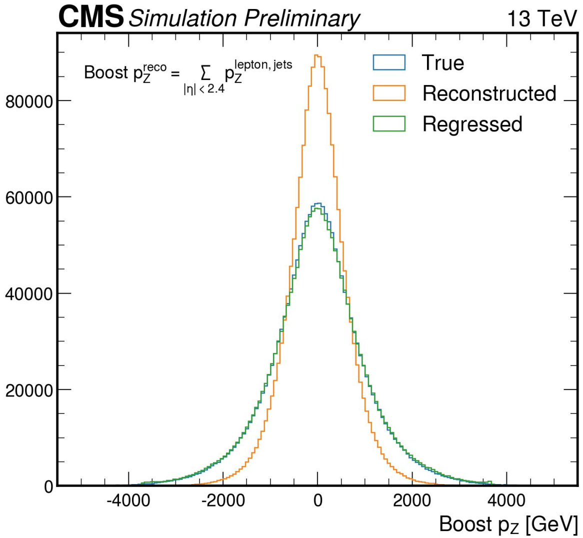
Conditioner transformer configuration:

- Encoder: 4 layers, 8 heads, 64D latent vector, 512 hidden layer,
- Decoders: 4 layers, 4 heads, 64D latent vector, 512 hidden layer
- Total weights: 1.8M

Parton regression performance - Boost



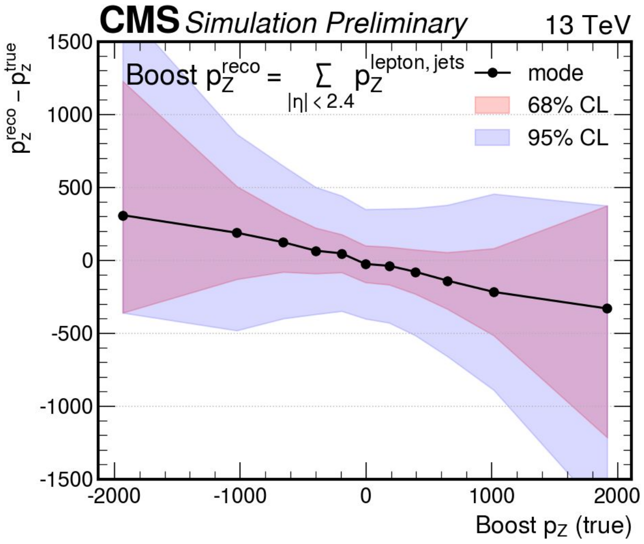


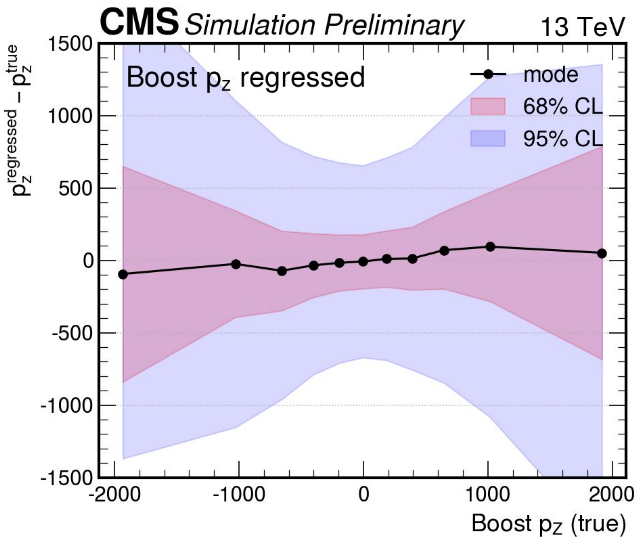


The conditioning transformer regresses the event boost in order to be able to bring the partons to the center-of-mass.

The left figure shows the correlation between true boost p_z and the boost p_z estimated by summing all the reconstructed objects with $|\eta| < 2.4$. The figure in the middle shows the regressed boost p_z which improved a lot the correlation. The plot on the right shows the 1D profile of the boost p_z for the truth level, the reconstructed estimation and the regression.

Parton regression performance - Boost

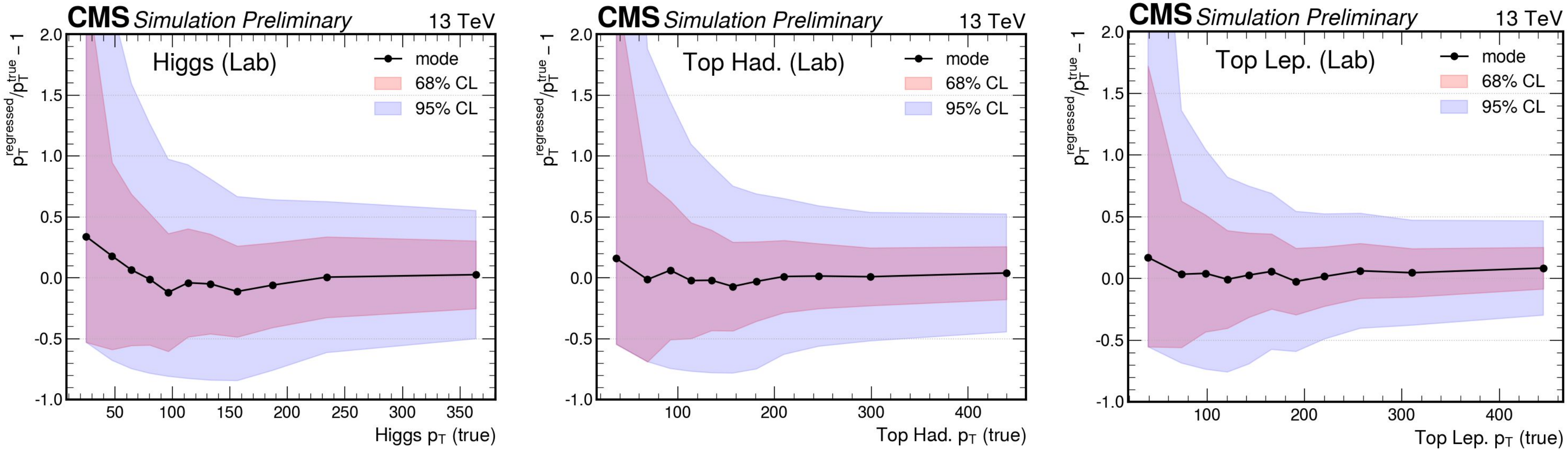




The conditioning transformer regresses the event boost in order to be able to bring the partons to the center-of-mass.

The left figure shows the bias in the estimation of the true boost p_Z computed by summing all the reconstructed objects with $|\eta| < 2.4$. The figure on the right shows the bias in the estimation of the true boost p_Z using the boost regressed by the conditioning transformer. The bias at high p_Z is reduced by the regression.

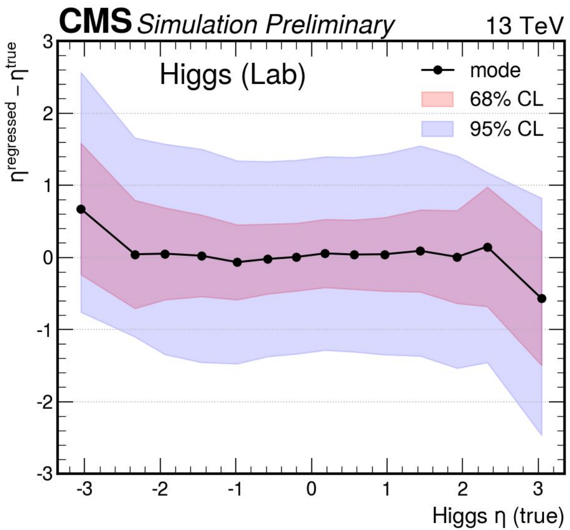
Parton regression performance - Lab Frame

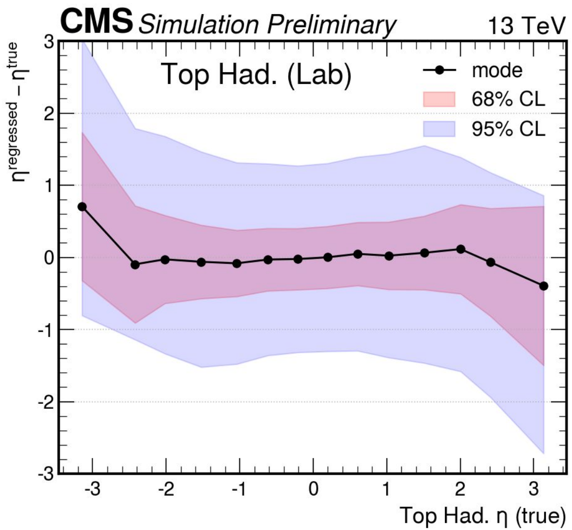


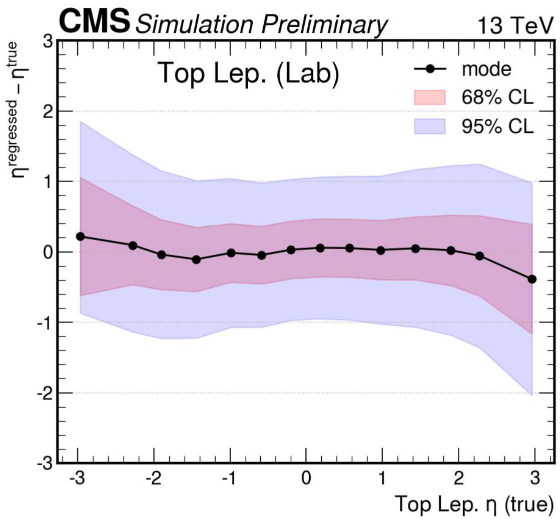
The plots show the performance of the regression performed by the conditioning transformer. The bias in the regression of the p_T of the higgs, hadronically decaying top quark (Top Had), and leptonically decaying top quark (Top Lep), is shown in bins of the true p_T of the particles in the lab frame.

The regression is overall unbiased: at high p_T the 68% confidence level interval reaches an uncertainty in the regressed p_T of ~30%, whereas at low p_T (< 50 GeV) the uncertainty is larger.

Parton regression performance - Lab Frame



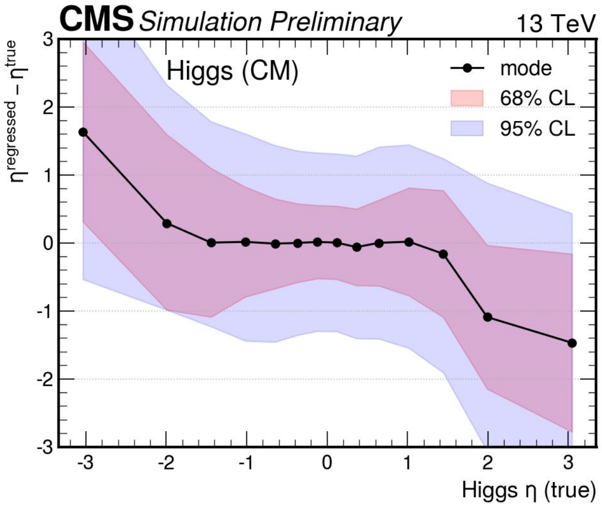


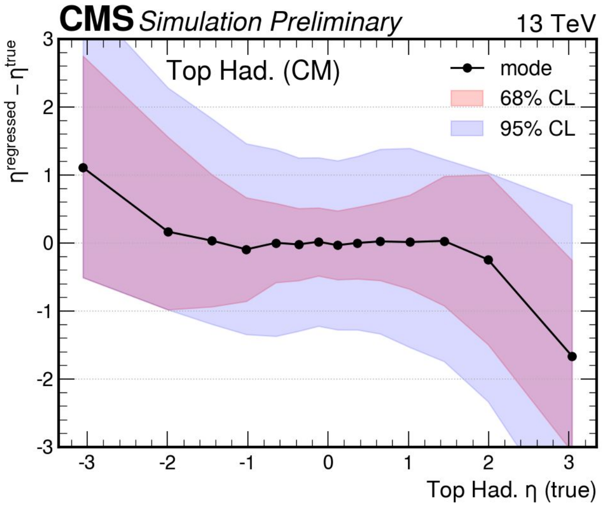


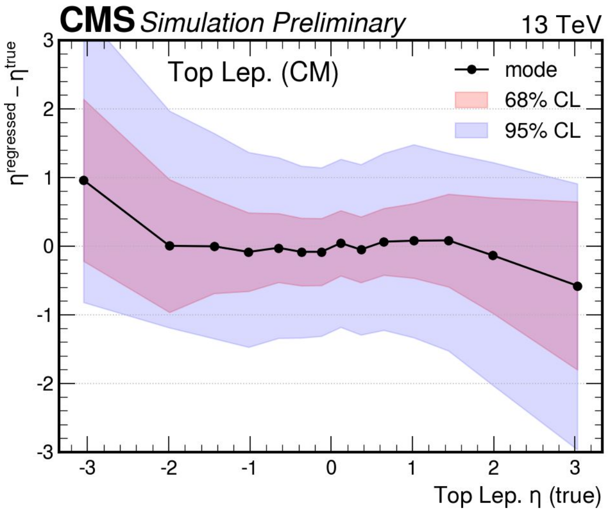
The plots show the performance of the regression performed by the conditioning transformer. The bias in the regression of the η of the higgs, hadronically decaying top quark (Top Had), and leptonically decaying top quark (Top Lep), is shown in bins of the true η of the particles in the lab frame.

The regression is overall unbiased, apart from the region |η|>2.5 which covers the very tail <3% of the particle distribution.

Parton regression performance - CM Frame





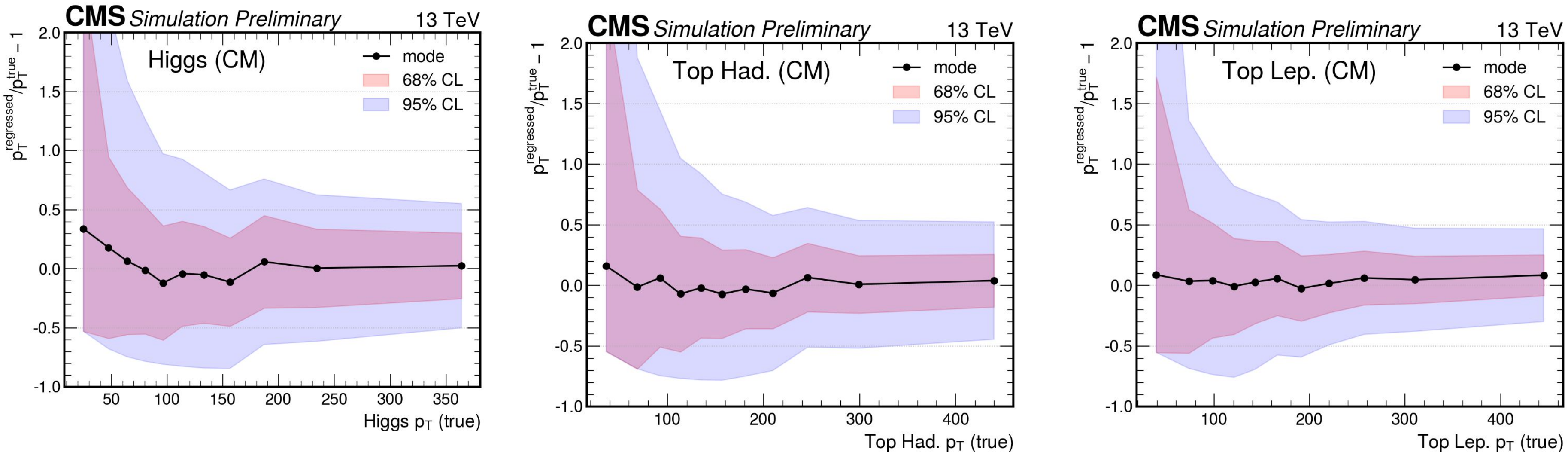


The plots show the performance of the regression performed by the conditioning transformer, after bringing the regressed partons in the center-of-mass (CM) using the regressed event boost. The bias in the regression of the η of the higgs, hadronically decaying top quark (Top Had), and leptonically decaying top quark (Top Lep), is shown in bins of the true η of the particles in the CM frame.

The regression is overall unbiased, apart from the region |η|>1.5 which covers the tail <~10% of the particle distribution. The performance is slightly worse than in the lab frame due to the imperfect regression of the total boost p_Z, used for the transformation.

Parton regression performance - CM Frame

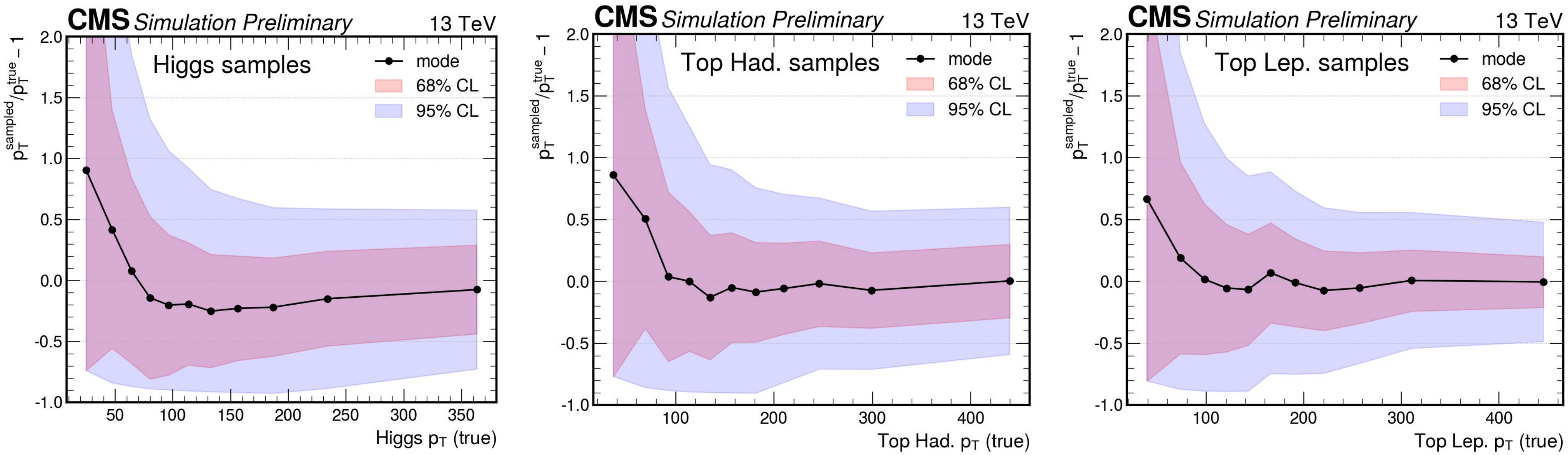
ETH zürich



The plots show the performance of the regression performed by the conditioning transformer, after bringing the regressed partons in the center-of-mass (CM) using the regressed event boost. The bias in the regression of the p_T of the higgs, hadronically decaying top quark (Top Had), and leptonically decaying top quark (Top Lep), is shown in bins of the true p_T of the particles in the CM frame.

The regression is overall unbiased: at high p_T the 68% confidence level interval reaches an uncertainty in the regressed p_T of ~30%, whereas at low p_T (< 50 GeV) the uncertainty is larger. The performance is overall unchanged w.r.t of the regression result in the lab frame.

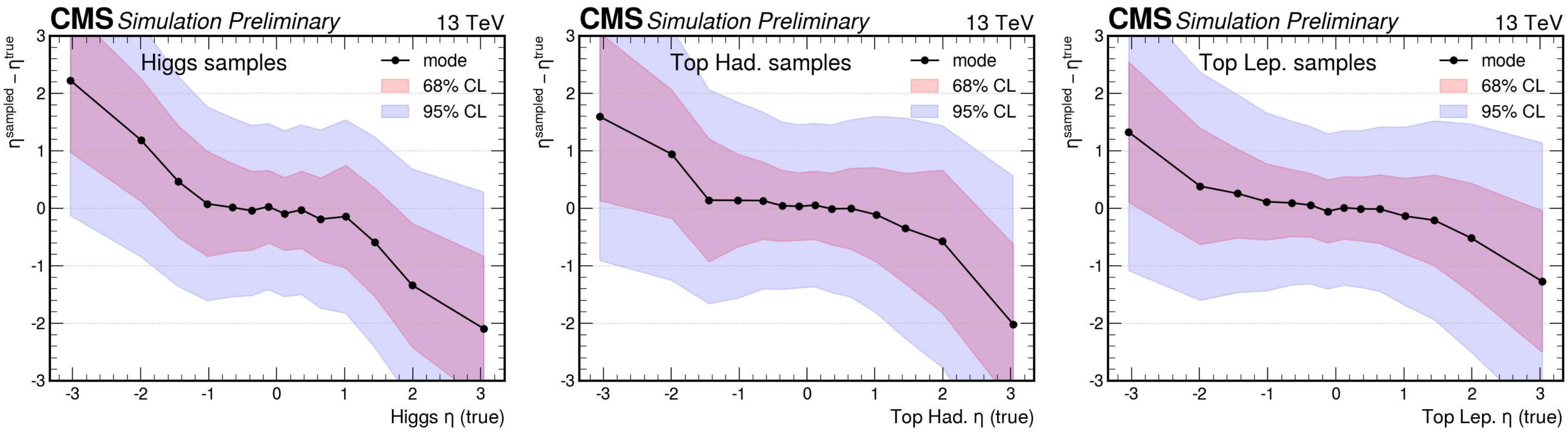
Quality of the sampled partons



The quality of the sampled partons configuration is evaluated by sampling 30 sets per 1.5M events and analyzing the bias w.r.t of the target true partons. The plots show the bias of the sampled partons in bins of the true p_T of the particles in the CM frame.

The average bias of the samples reaches 20% in the bulk of the distribution (100 GeV) for the higgs parton, while it is very close to 0% for the top quarks. The bias is very small at high pT, while at very low pt (<50GeV), it increases up to 100%.

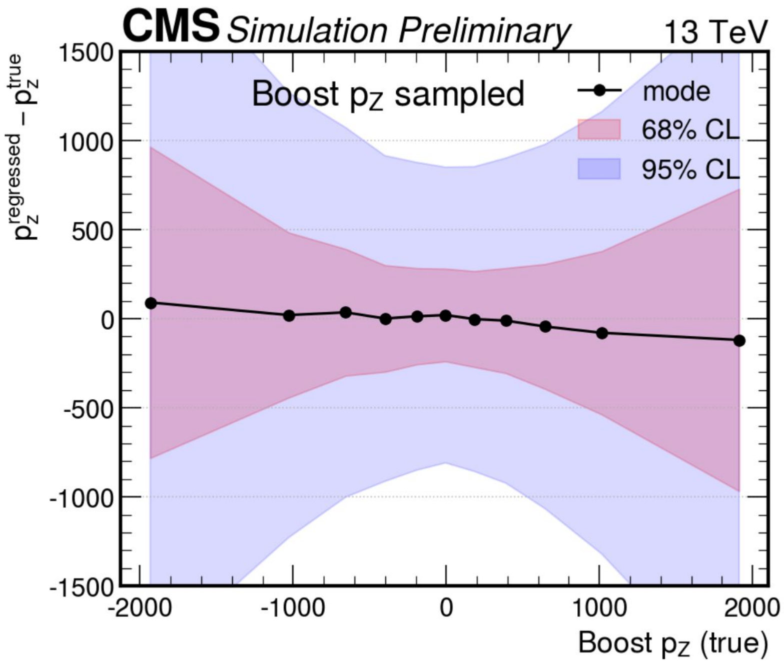
Quality of the sampled partons

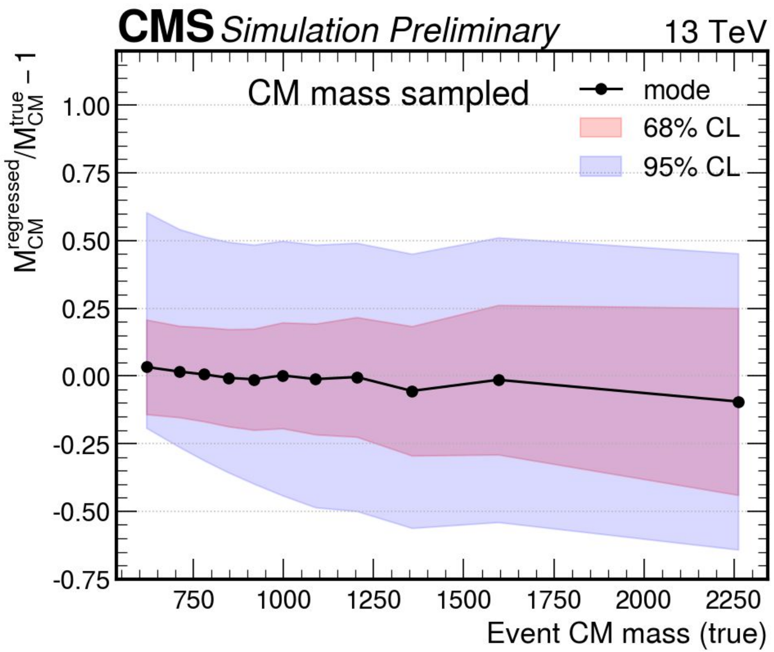


The quality of the sampled partons configuration is evaluated by sampling 30 sets per 1.5M events and analyzing the bias w.r.t of the target true partons. The plots show the bias of the sampled partons in bins of the true η of the particles in the CM frame.

The average bias of the samples is close to 0 in the bulk of the distribution, while it reached Δη~1 in the tails (only 10% of the events have $|\eta_{CM}| > 1.5$). In general, the performance of the sampling is highly correlated with the one of the regression, as expected, since the unfolding flow is conditioned with the regressed partons.

Quality of the sampled boost





The unfolding flow models also the incoming parton energy fractions alongside the final state partons in the CM frame.

The bias of the sampled incoming parton energies is shown by plotting the event total CM mass and the event total boost $p_{Z.}$ Both the distributions are very well modelled by the flow.

Unfolding flow architecture and training

Unfolding flow configuration:

- Conditioning vector dimension: 16
- Number of transforms: 6
- Flow type: RQS autoregressive
- Number of spline bins: 40
- Spline network: 64 hidden features, 2x512 feedforward networks
- Number of trainable parameters: 5.3M (flow)+ 1.8M (conditioner) = 7.1M

The flow is trained **simultaneously my maximum likelihood** and by **sampling** parton configurations and comparing them with the target.

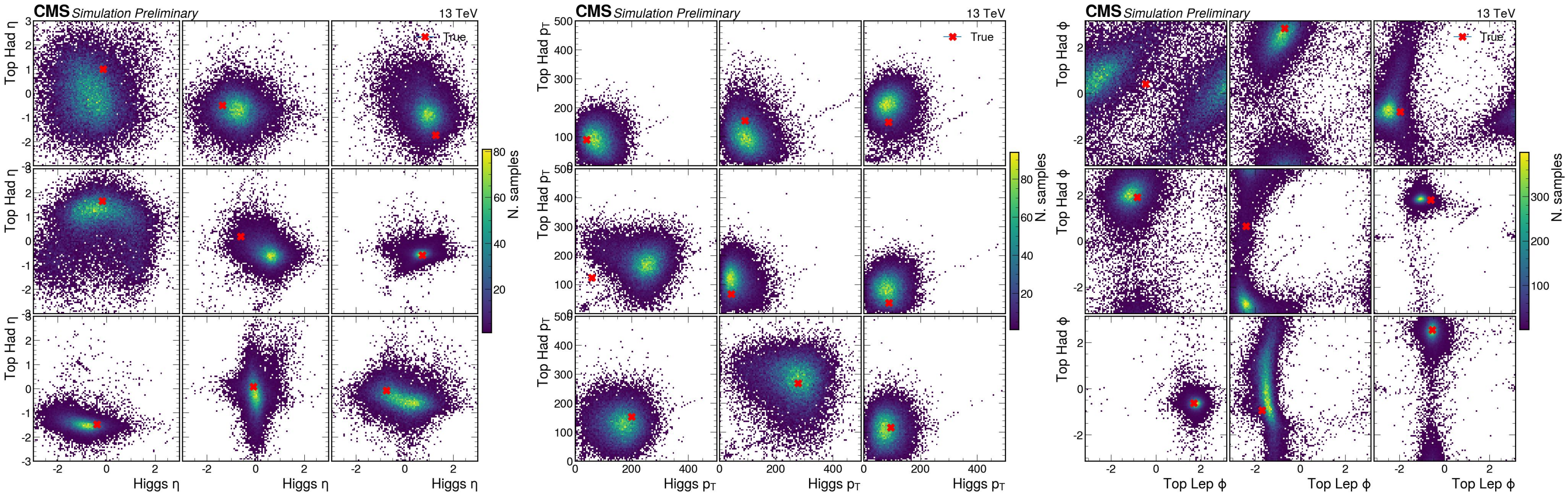
The training took 20 hours on a A100 Nvidia GPU over a 1.5M events training dataset.

2 approaches to evaluate the quality of the flow:

- **sample 20k parton sets** for a few random events and **inspect their distributions**
- **samples 20 partons** sets for 1.5M of testing events and analyze the bias w.r.t of the true partons

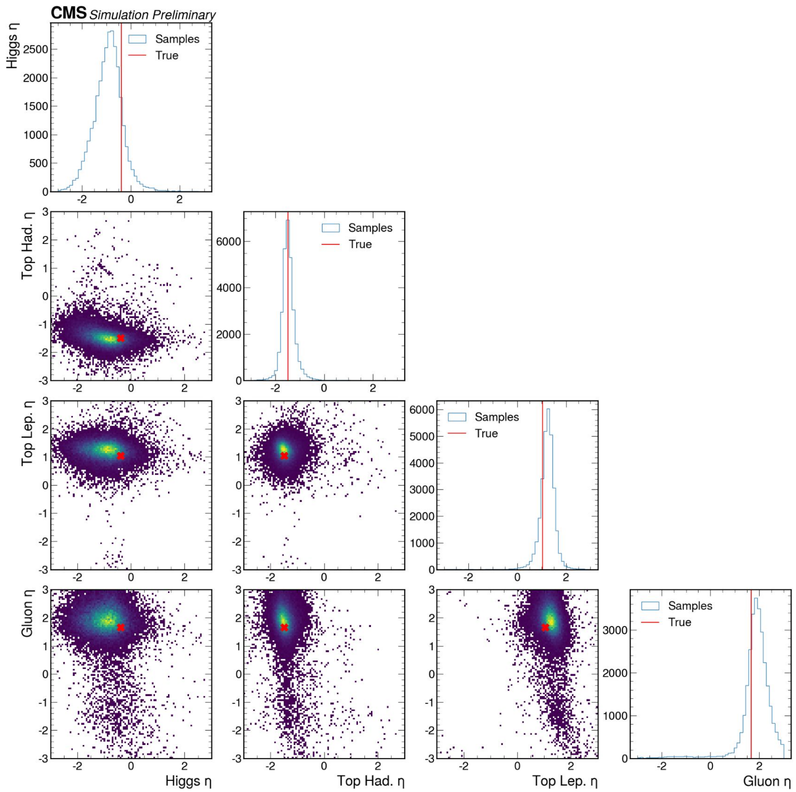
Goal for **importance sampling:**   minimize bias, sample points with the same probability of the integrand function.

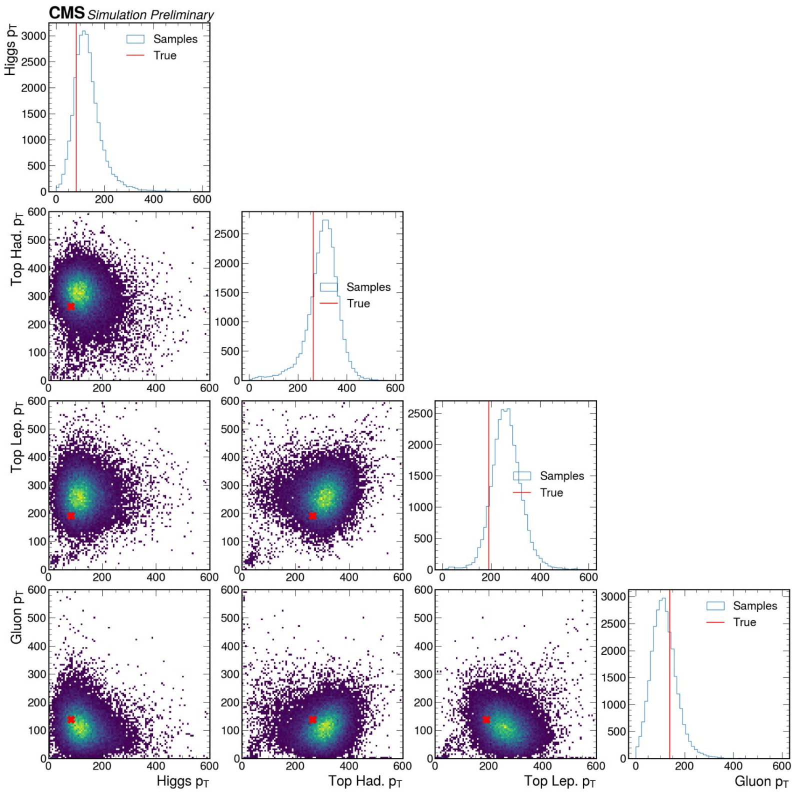
Unfolding flow samples



- The unfolding flow learns the conditional probability ***P(partons|reconstructed event)*** and generates partons in the most probable configurations.
- **Each box** in the figures shows 20k sampled partons for a **given event.**
- The red cross shows the true value of the partons position/energy in each event.

Unfolded flow samples: single event correlations





Correlation between η (left) and p_T (right) of the sampled generator-level particle for a single event.

The red cross/line shows the true value of the partons position/energy in each event.

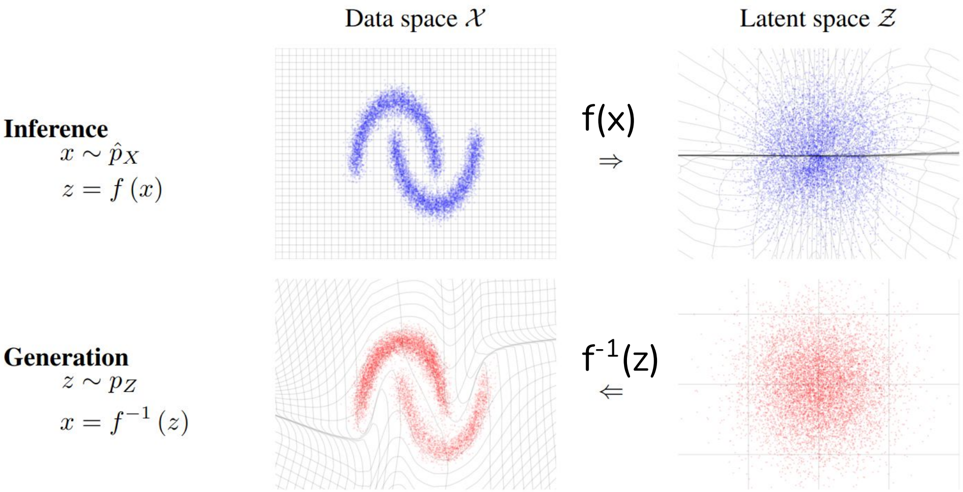
Normalizing flows : More formally

From the rules of change of integration variables

$$p_X(x) = p_Z(f(x)) \left|\det\left(\frac{\partial f(x)}{\partial x^T}\right)\right|$$

$$\log\left(p_X(x)\right) = \log\Big(p_Z(f(x))\Big) + \log\left(\left|\det\left(\frac{\partial f(x)}{\partial x^T}\right)\right|\right),$$

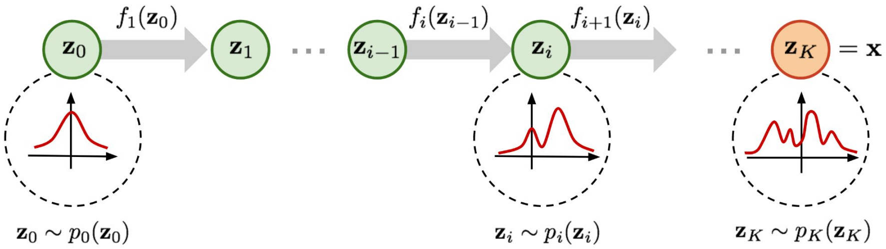
where f(x) goes in the “normalizing” direction to the z latent space.



We can both **sample** and evaluate the **density**

- If the p.d.f in the **latent space is tractable** (multidim gaussian, uniform)
- if the transformation is **invertible**

Requirement: the jacobian of the transformation must be computed in an efficient way
→ this defines the possible implementation of the flows



Expressiveness: transformations are composable!

$$(T_2 \circ T_1)^{-1} = T_1^{-1} \circ T_2^{-1}$$

$$\det J_{T_2 \circ T_1}(\mathbf{u}) = \det J_{T_2}(T_1(\mathbf{u})) \cdot \det J_{T_1}(\mathbf{u}).$$

MEM: the hard scatter

arxiv:1101.2259

Prob. of the reco event — hard-scattering — transfer function

$$\mathcal{P}(\vec{Y}|\vec{\theta}) = \int_{\phi} d\vec{X} \cdot \boxed{|\mathcal{M}(\vec{X}|\vec{\theta})|^2 \cdot Pdf} \cdot \mathcal{W}(\vec{Y}|\vec{X})$$

- From first principles: MadGraph, OpenLoop
- Only component depending on the **parameters θ**
- It can become the **slowest** part in the evaluation of the MEM
- It can be “multi-channel”: non trivial dependency on the different Feynman diagrams
- Parton distribution functions need to be convoluted as:

$$\int_{\phi} d\vec{X} \sum_{a,b} \int_{x_1,x_2} dx_1 dx_2\, f_a(x_1, Q^2) \cdot f_b(x_2, Q^2) \cdot |\mathcal{M}(x_1, x_2, \vec{X}|Q^2, \vec{\theta})| \cdot \mathcal{W}(\vec{Y}|\vec{X})$$

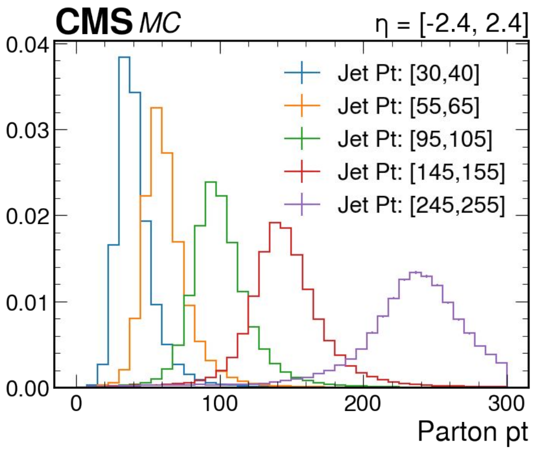
MEM: the transfer function

arxiv:1101.2259

Prob. of the reco event — hard-scattering — transfer function

$$\mathcal{P}(\vec{Y}|\vec{\theta}) = \int_{\phi} d\vec{X} \cdot |\mathcal{M}(\vec{X}|\vec{\theta})|^2 \cdot Pdf \cdot \boxed{\mathcal{W}(\vec{Y}|\vec{X})}$$

- "Core" of the MEM method: models the **probability to get a reconstructed event given a parton configuration**
- It is the main factor driving the power and precision of the method
- δ for leptons, more complex for jets
- Usually it is **factorized by object** taking many assumptions:
  - objects matching and geometrical alignment
  - ignore out-of-acceptance objects
  - additional term for MET and initial state radiation.



MEM: integration space

arxiv:1101.2259

Prob. of the reco event

hard-scattering

transfer function

$$\mathcal{P}(\vec{Y}|\vec{\theta}) = \boxed{\int_{\phi} d\vec{X}} \cdot |\mathcal{M}(\vec{X}|\vec{\theta})|^2 \cdot Pdf \cdot \mathcal{W}(\vec{Y}|\vec{X})$$

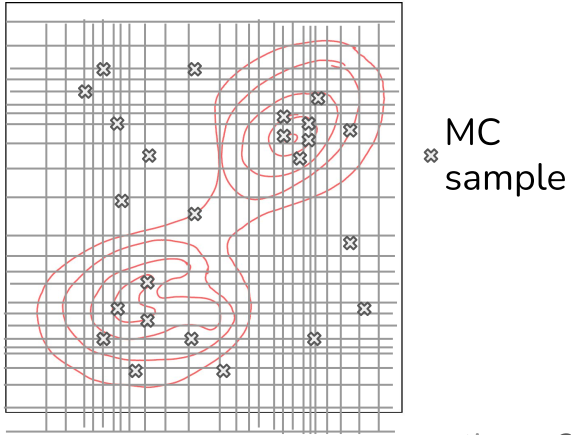
→ The integration is over the **partons (+pdfs) complete phase-space**

→ The integral is computed numerically with MC sampling

→ Very high dimensional → needs approximations

→ Strong dependence on the coordinate choices

- needs to be “aligned” with propagators
- invariant mass constraints
- Jet-parton alignments
- additional radiation complex to handle



ttH(bb) channel

The MEM has been applied to the problem of signal discrimination for the ttH(bb) CMS analysis.

→ Complex process with 2 → 8 topology

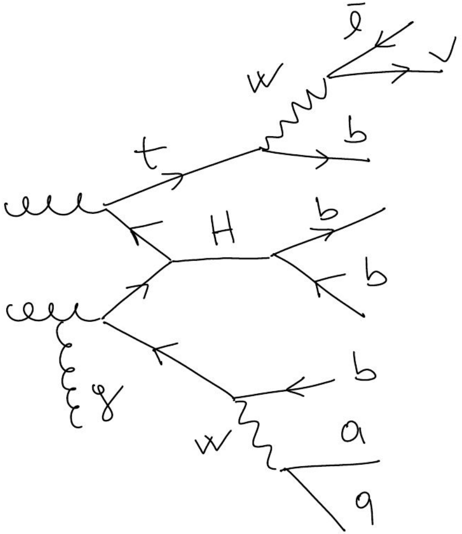
→ 3 channels:

- single lepton
- 2 leptons
- fully hadronic

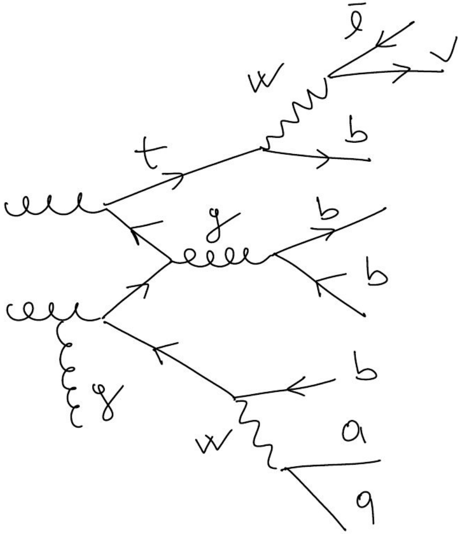
→ Large irreducible background from tt+bb SM process.

→ CMS Analyses:

- 2016: doi:10.1007/JHEP03(2019)026
- under review: HIG-19-001 (full Run2)



ttH(bb)
semileptonic



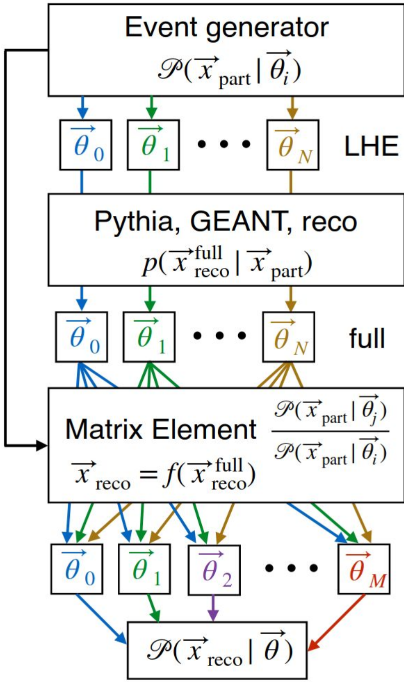
tt+bb

EFT interpretation

$$\mathcal{P}(\vec{x}_{\text{reco}}\,|\,\vec{\theta}) \propto \mathcal{P}_0(\vec{x}_{\text{reco}}) + \sum_k \left(\frac{2\theta_k}{\theta_0}\right)\mathcal{P}_{0k}(\vec{x}_{\text{reco}}) + \sum_k \left(\frac{\theta_k}{\theta_0}\right)^2 \mathcal{P}_k(\vec{x}_{\text{reco}}) + \sum_{i<j}\left(\frac{2\theta_i\theta_j}{\theta_0^2}\right)\mathcal{P}_{ij}(\vec{x}_{\text{reco}})$$

- **Goal**: full Run2 measurement of dim6 EFT operators parameters (Wilson coeff.) relevant for ttHbb channel.

- There are 2 possible ways to perform EFT studies in general
  - **one-step**, “direct” measurement: model EFT effects at reco level, build a template analysis, use likelihood profiling to get limits on θ
    - Useful to restrict the number of parameters
    - More complex bookkeeping of MC samples
    - it can reach optimal sensitivity
    - CMS internal, difficult for reinterpretation
  - **two-step**, “unfolding”, STXS:
    - fiducial measurements usable for re-interpretation outside of CMS collaboration
    - less optimal, more general



LHC EFT WG Group2 report: arxiv

EFT optimal observables

SMEFT quadratic parametrization with dim-6 Wilson coefficients

$$\mathscr{P}(\vec{x}_{\text{reco}}\,|\,\vec{\theta}) \propto \mathscr{P}_0(\vec{x}_{\text{reco}}) + \sum_k \left(\frac{2\theta_k}{\theta_0}\right)\mathscr{P}_{0k}(\vec{x}_{\text{reco}}) + \sum_k \left(\frac{\theta_k}{\theta_0}\right)^2 \mathscr{P}_k(\vec{x}_{\text{reco}}) + \sum_{i<j}\left(\frac{2\theta_i\theta_j}{\theta_0^2}\right)\mathscr{P}_{ij}(\vec{x}_{\text{reco}})$$

Using Neyman-Pearson lemma one can extract the maximum information from two optimal observables (likelihood ratios).

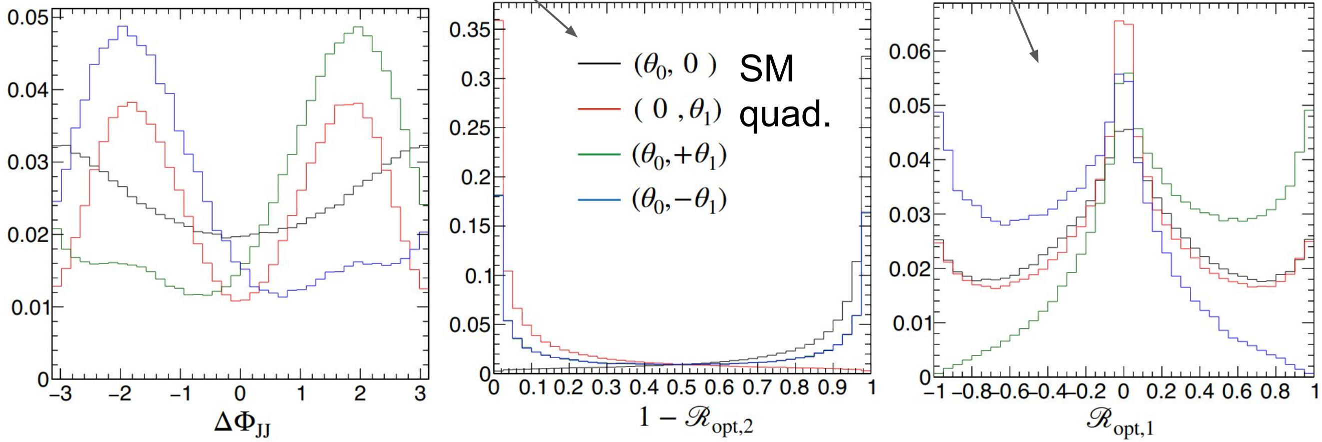
quadratic term

$$\mathcal{R}_{\text{opt},2} = \frac{\mathcal{P}_1(\vec{x}^{\,\text{full}}_{\text{reco}})}{\mathcal{P}_0(\vec{x}^{\,\text{full}}_{\text{reco}}) + c\cdot\mathcal{P}_1(\vec{x}^{\,\text{full}}_{\text{reco}})},$$

interference term

$$\mathcal{R}_{\text{opt},1} = \frac{2\mathcal{P}_{01}(\vec{x}^{\,\text{full}}_{\text{reco}})}{\mathcal{P}_0(\vec{x}^{\,\text{full}}_{\text{reco}}) + c\cdot\mathcal{P}_1(\vec{x}^{\,\text{full}}_{\text{reco}})},$$

Example: on-shell Higgs s-channel to VV. Δϕ between the two associated jets. CP-odd operator effect.



LHC EFT WG Group2 report: [arxiv]

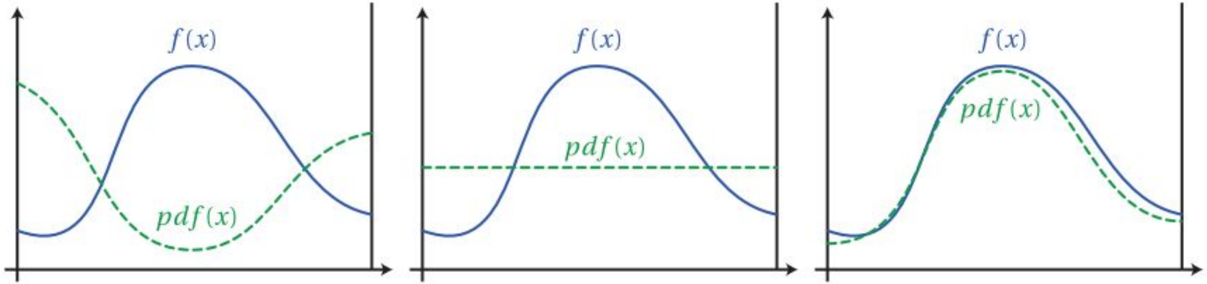
Importance sampling

An multi-dim integral can be estimated via MC sampling

$$I \approx \frac{V}{N} \sum_{i=1}^{N} f(x_i) \equiv V \langle f \rangle_x \,, \qquad \sigma_I = \sqrt{\mathrm{Var}} \approx V \sqrt{\frac{\langle f^2 \rangle_x - \langle f \rangle_x^2}{N-1}} \,,$$

If we can find a function g(x) which has a similar shape as f(x) we can improve the variance -> importance sampling

$$I = \int_\Omega \frac{f(x)}{g(x)} \, \mathrm{d}G(x) = V \langle f/g \rangle_G \,, \qquad \sigma_I = V \sqrt{\frac{\langle (f/g)^2 \rangle_G - \langle f/g \rangle_G^2}{N-1}} \,.$$



bad choice | initial uniform prior | very good choice