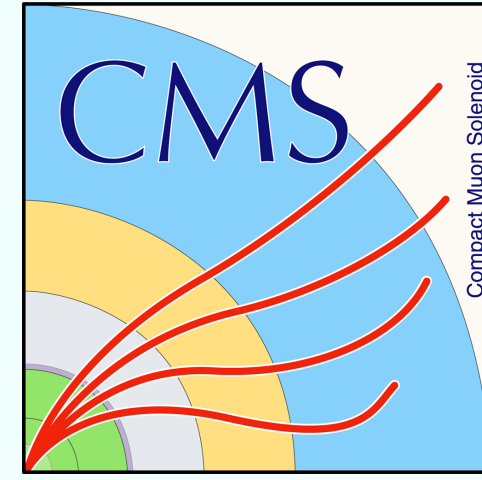




From Collisions to Predictions:

Search for Non-Resonant Higgs Boson Pair
Production in the Fully Leptonic $bbWW$ Final State
using CMS Run 3 Data

Content

1) Theory

- Higgs mechanism
- Di-Higgs production
- Effective Field Theory

$$\kappa_\lambda = \frac{\lambda_{hhh}}{\lambda_{hhh}^{SM}} = 1 + c_6 \frac{v^4}{m_H^2 \Lambda^2}$$

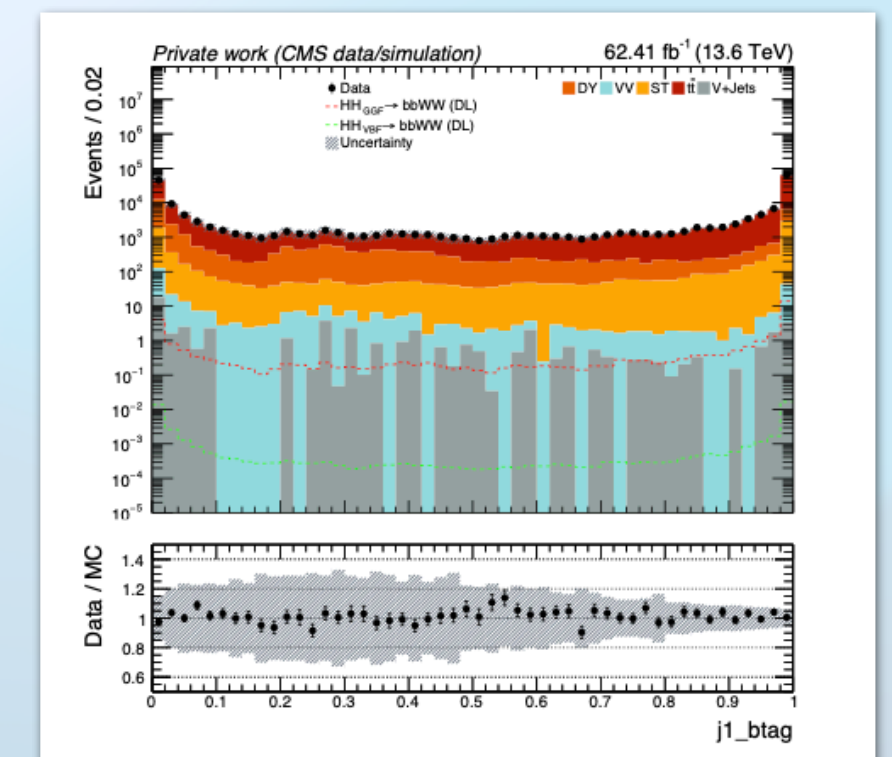
2) Machine learning

- General training procedures
- 3 ML models
- Comparison & choice



3) Data Analysis

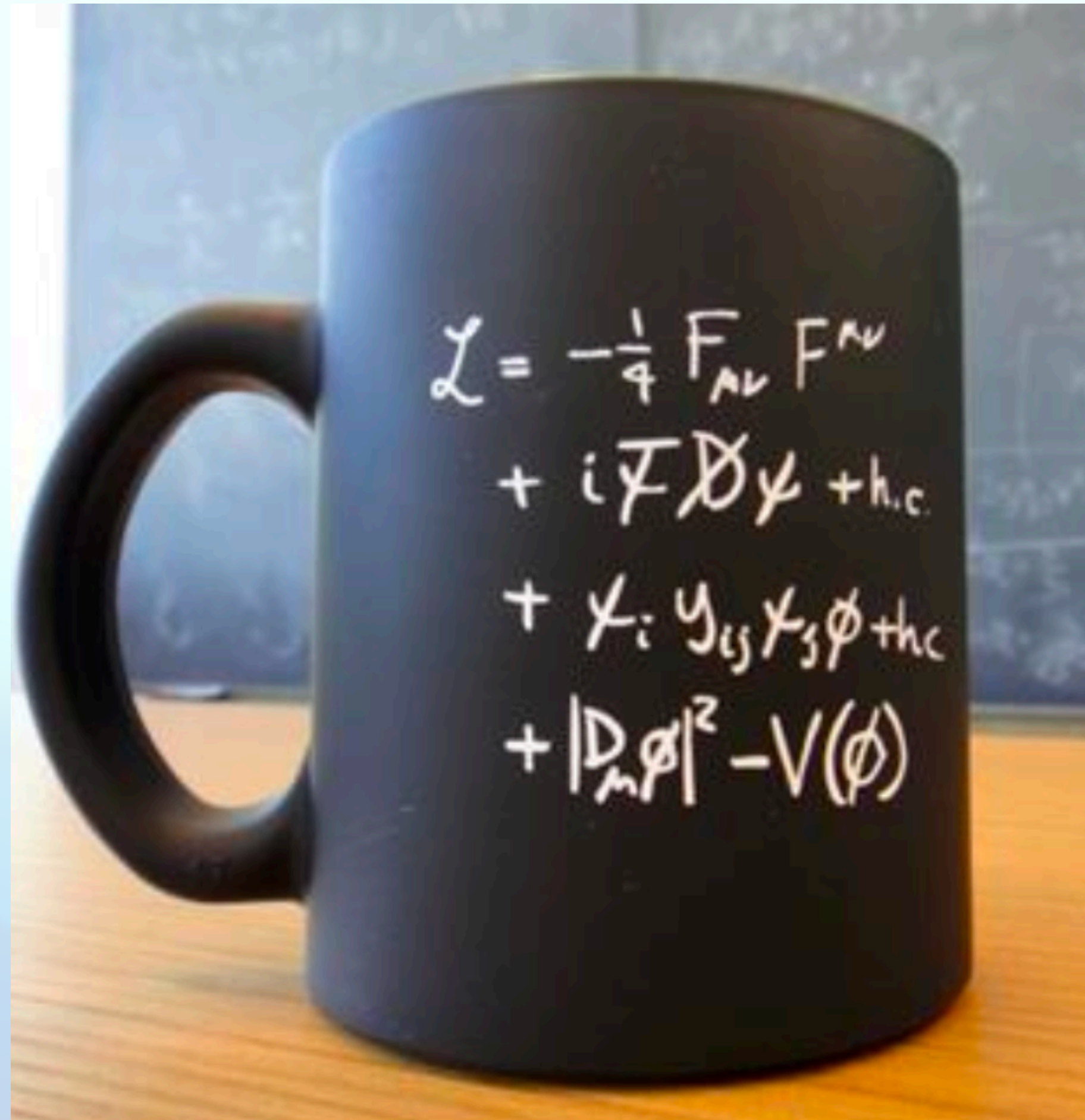
- Data & MC
- Object def. & event sel.
- Corrections & Uncertainties
- Results



Theory

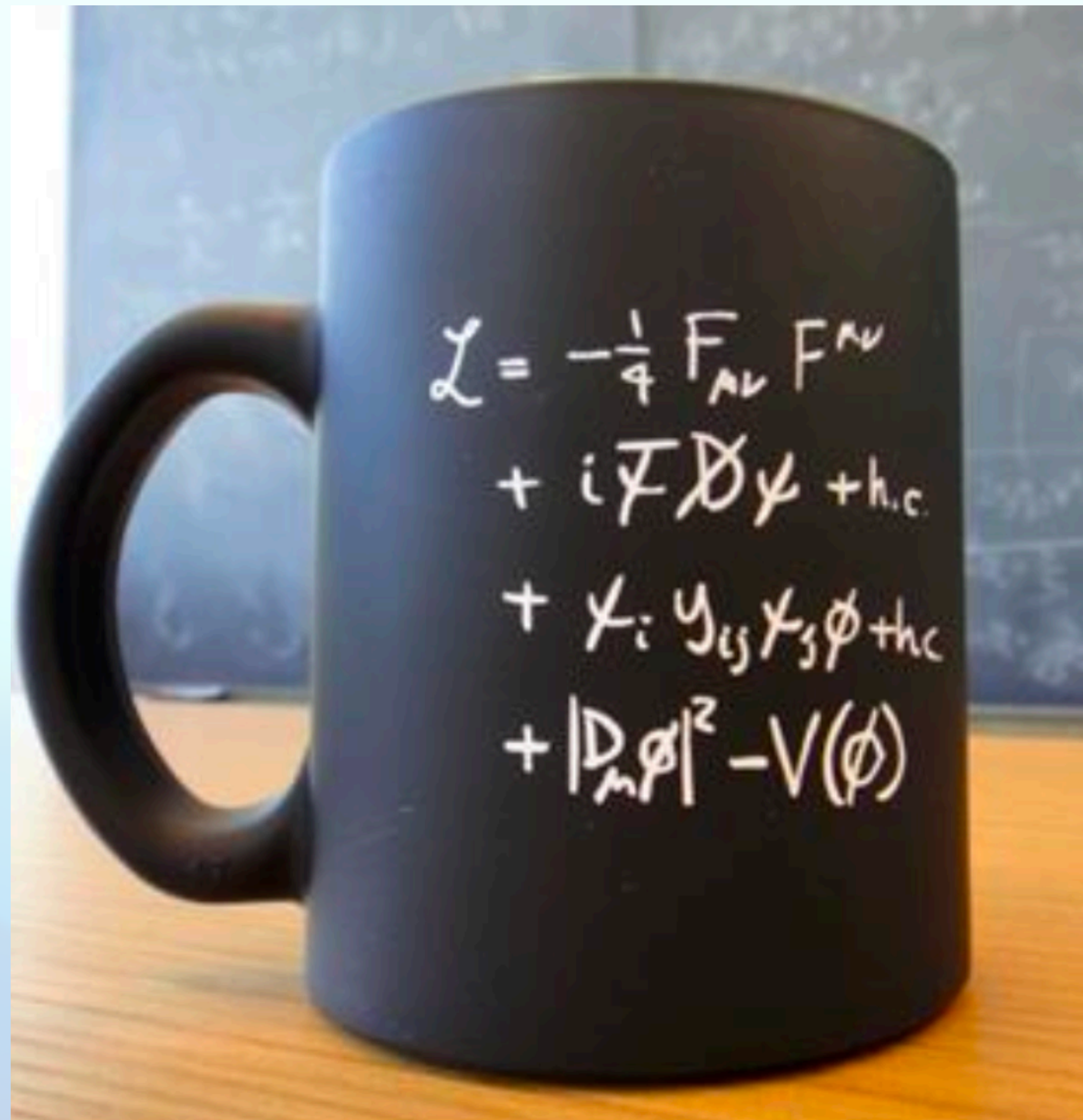
Theory

Standard Model of Particle Physics



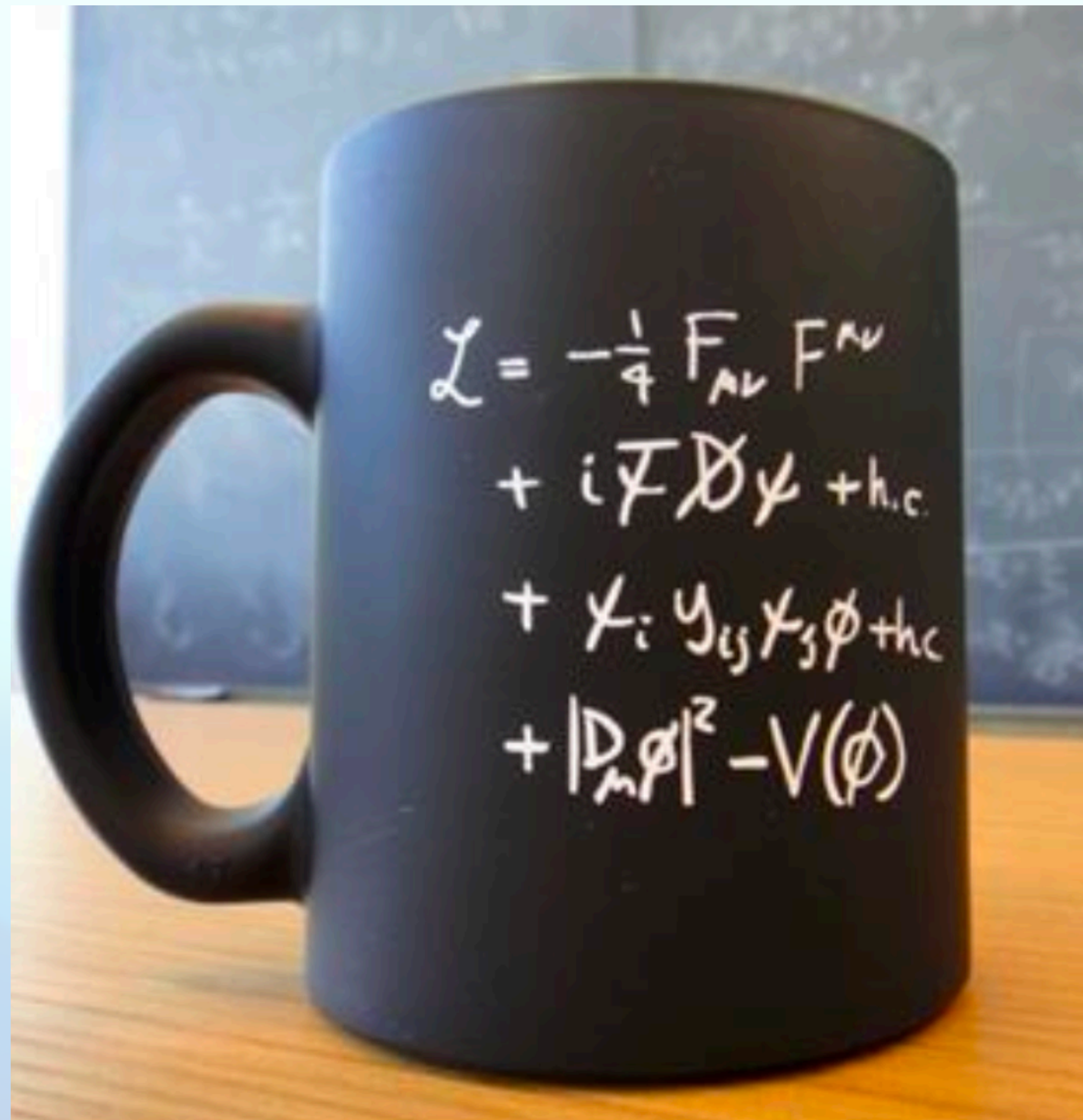
Theory

Standard Model of Particle Physics



Theory

Standard Model of Particle Physics



$$\mathcal{L}_{SM} = \mathcal{L}_{gauge} + \mathcal{L}_{fermion} + \mathcal{L}_{scalar}$$

$$\mathcal{L}_{gauge} = -\frac{1}{4} G_{\mu\nu}^a G_a^{\mu\nu} - \frac{1}{4} W_{\mu\nu}^i W_i^{\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu}$$

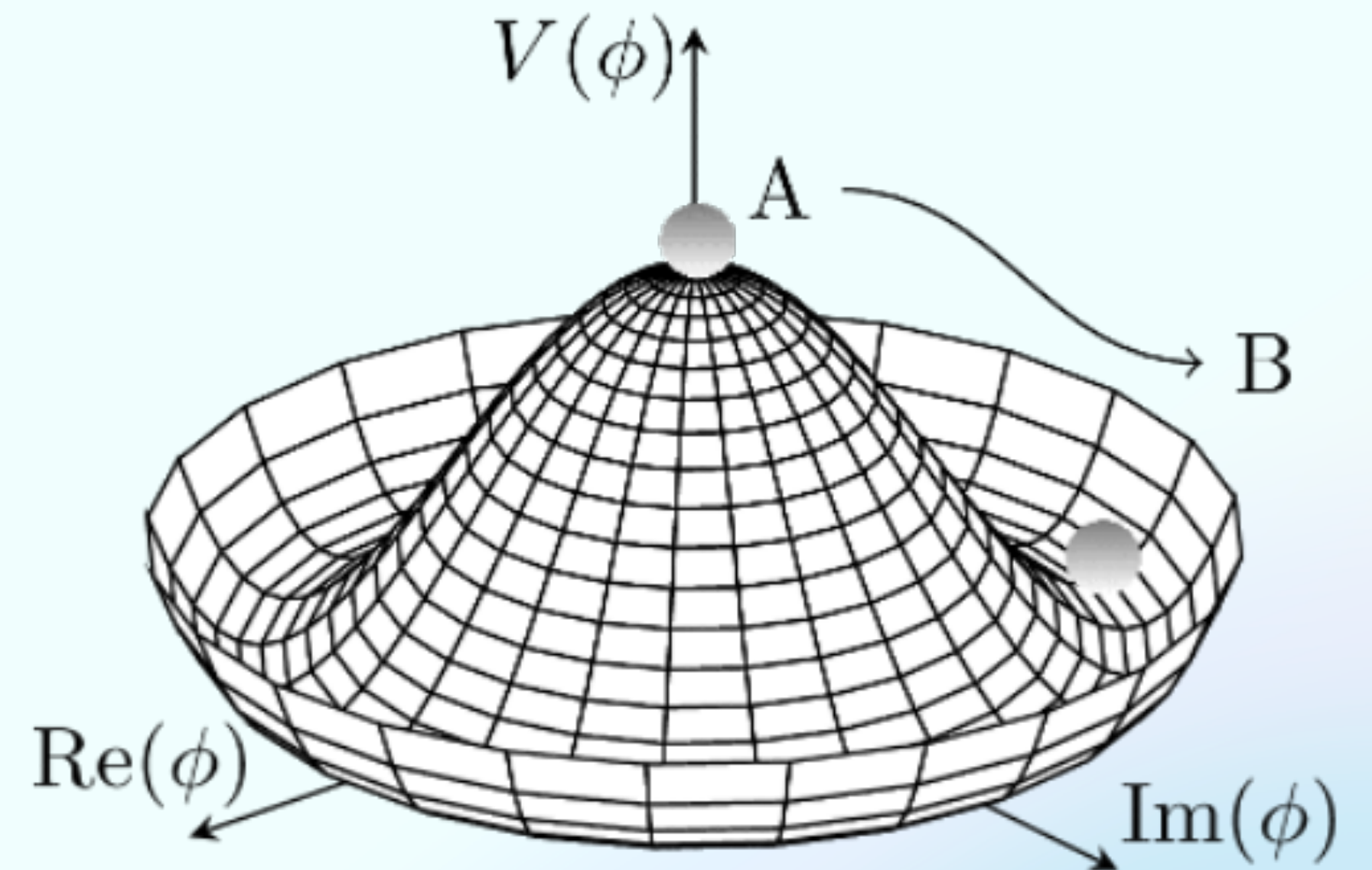
$$\mathcal{L}_{fermion} = \sum_{\psi, I} \bar{\psi}^I (i \bar{D}) \psi^I$$

Theory

The Higgs Potential & Self-Coupling

Brout-Englert-Higgs Mechanism ($\mathcal{L}_{scalar}$)

$$V(\phi^\dagger\phi) = -\mu^2(\phi^\dagger\phi) + \lambda(\phi^\dagger\phi)^2$$



Trilinear Higgs Self-Coupling

$$V \supset \underbrace{\frac{1}{2}m_h^2 h^2}_{\text{mass term}} + \underbrace{\lambda_{HHH} h^3}_{\text{trilinear coupling}} + \underbrace{\lambda_{HHHH} h^4}_{\text{quartic coupling}}$$

Core objective: Reconstruct shape of the Higgs potential

- testing the SM
- hints to the BSM (require EFT)

Theory

Probing new physics via EFTs - κ FR

To parametrize these effects: SMEFT and HEFT

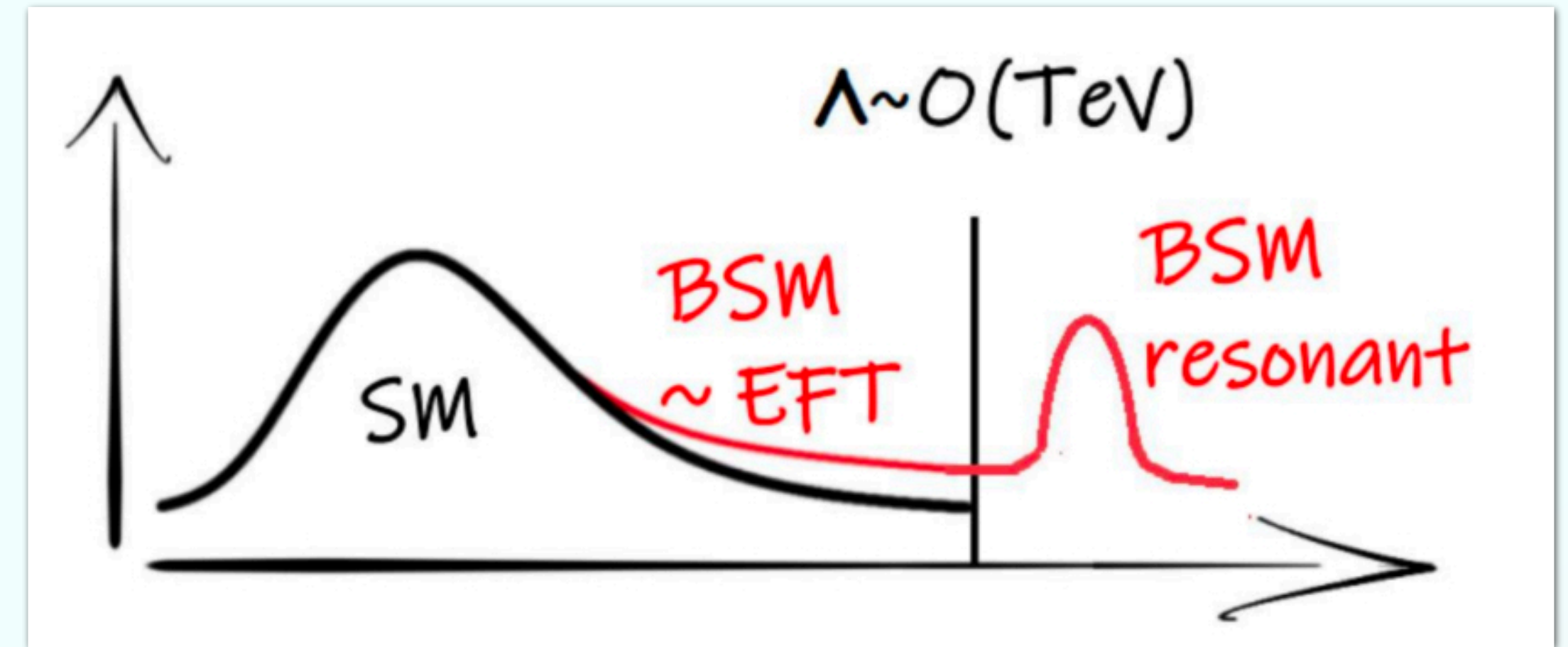
Expanding $\mathcal{L}_{SMEFT}$ after EWSB

$$\kappa_\lambda = \frac{\lambda_{hhh}}{\lambda_{hhh}^{SM}} = 1 + c_6 \frac{v^4}{m_H^2 \Lambda^2}$$

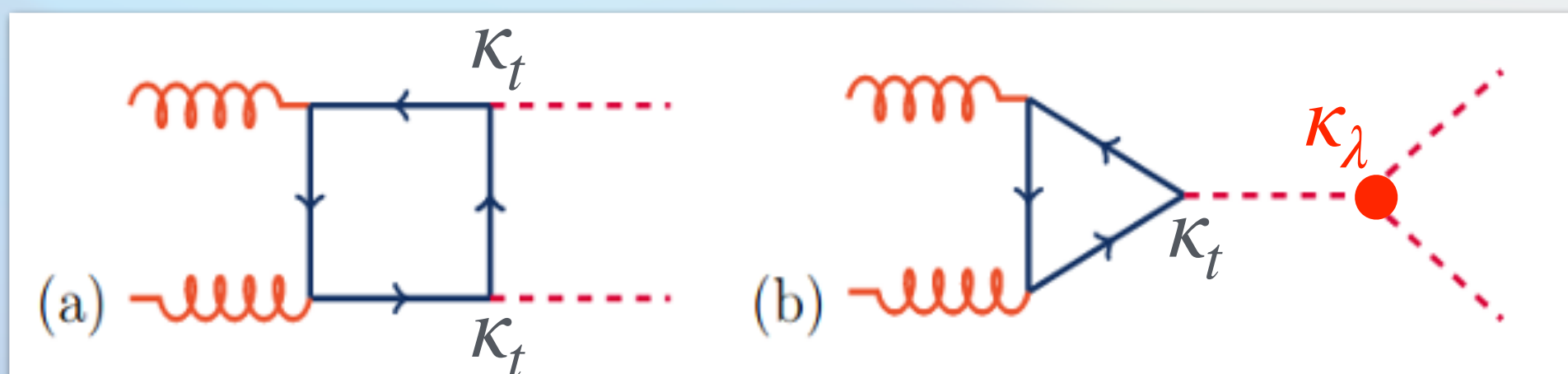
From EFT point of view, we vary the Wilson coefficients to see Λ 's effect

The κ view: we scale the entire vertex

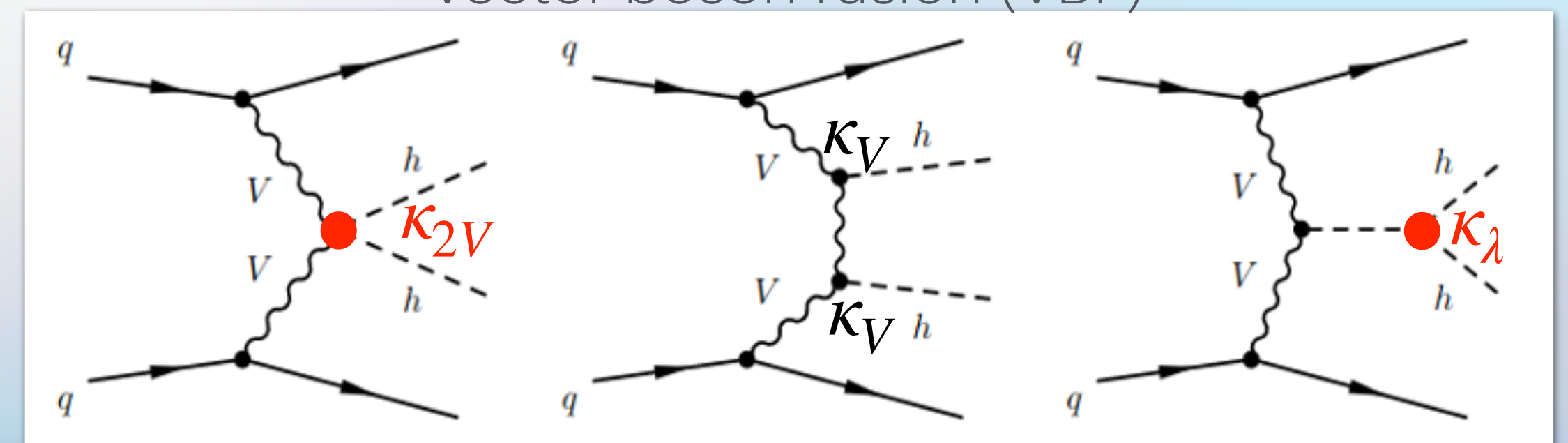
SMEFT cannot shift κ_{2V} without shifting the κ_V



gluon-gluon fusion (ggF)



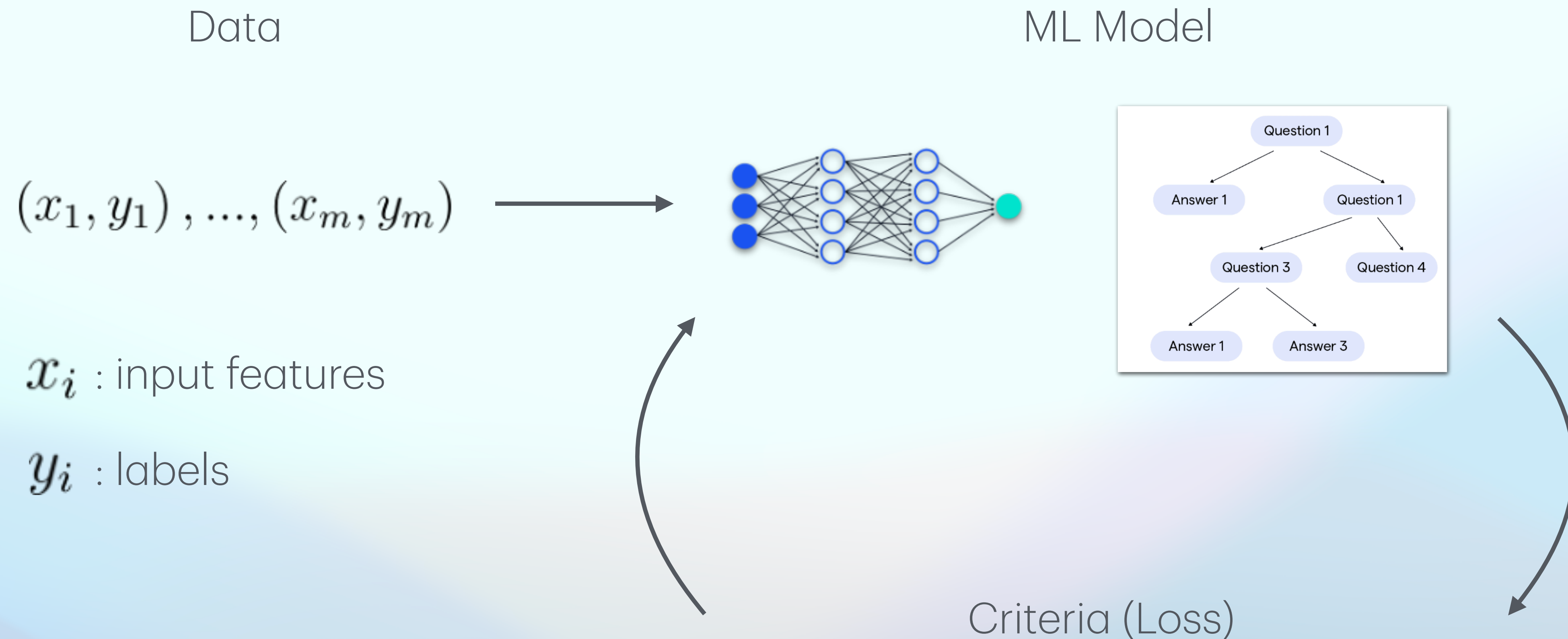
vector boson fusion (VBF)



Machine learning

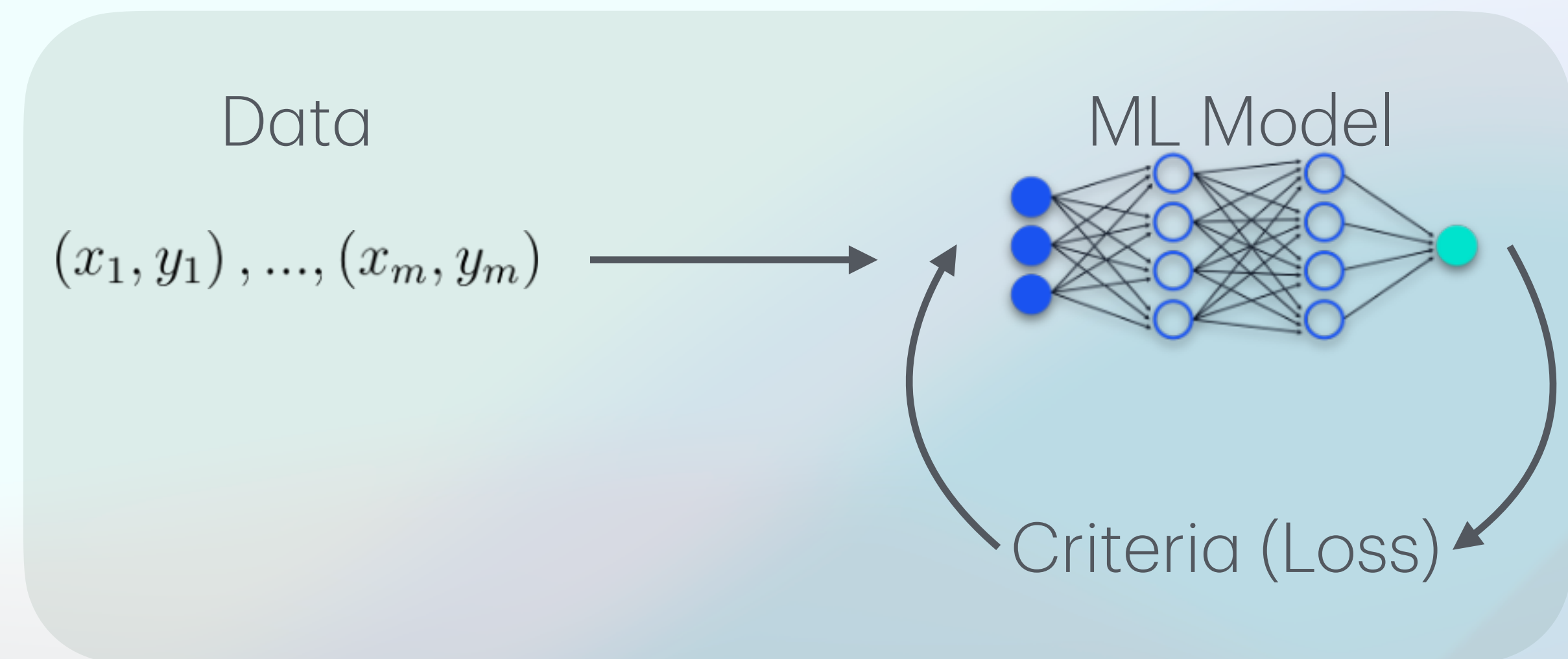
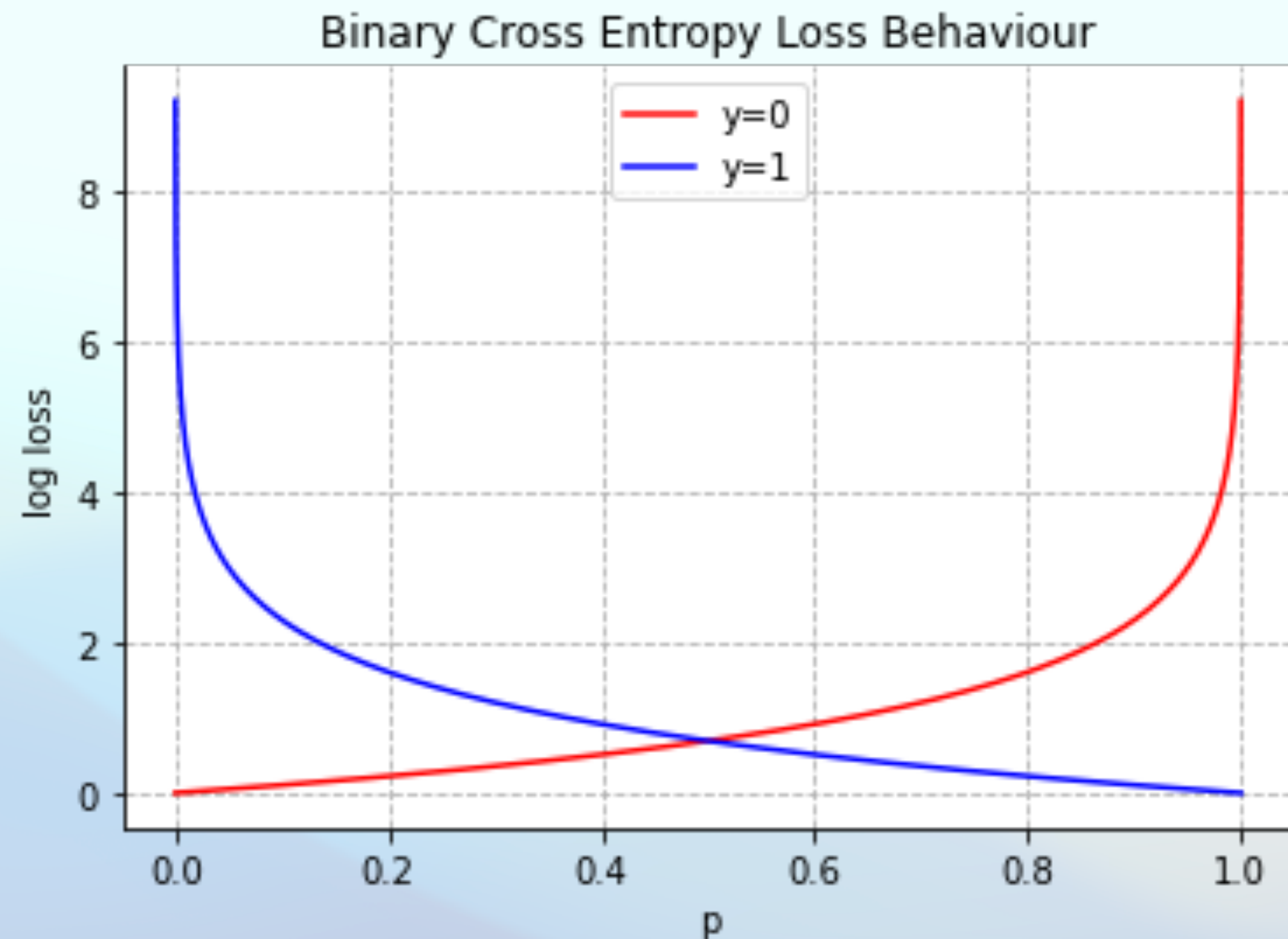
Machine Learning

Supervised learning



Machine Learning

General training procedure - Loss



Machine Learning

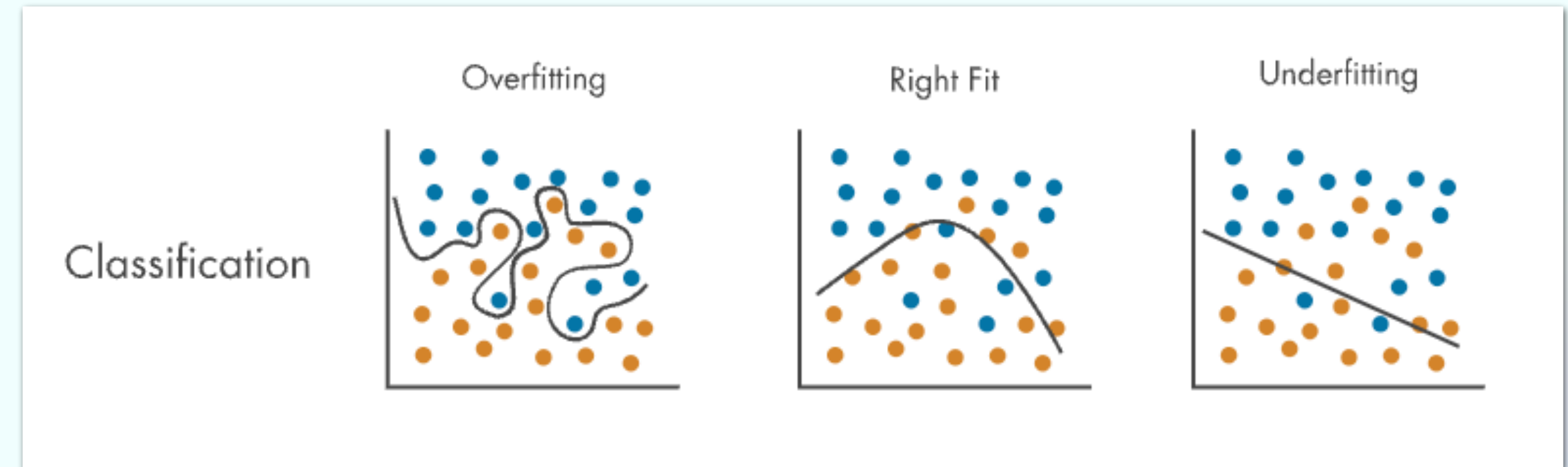
Optimization and validation

Preventing overfitting:

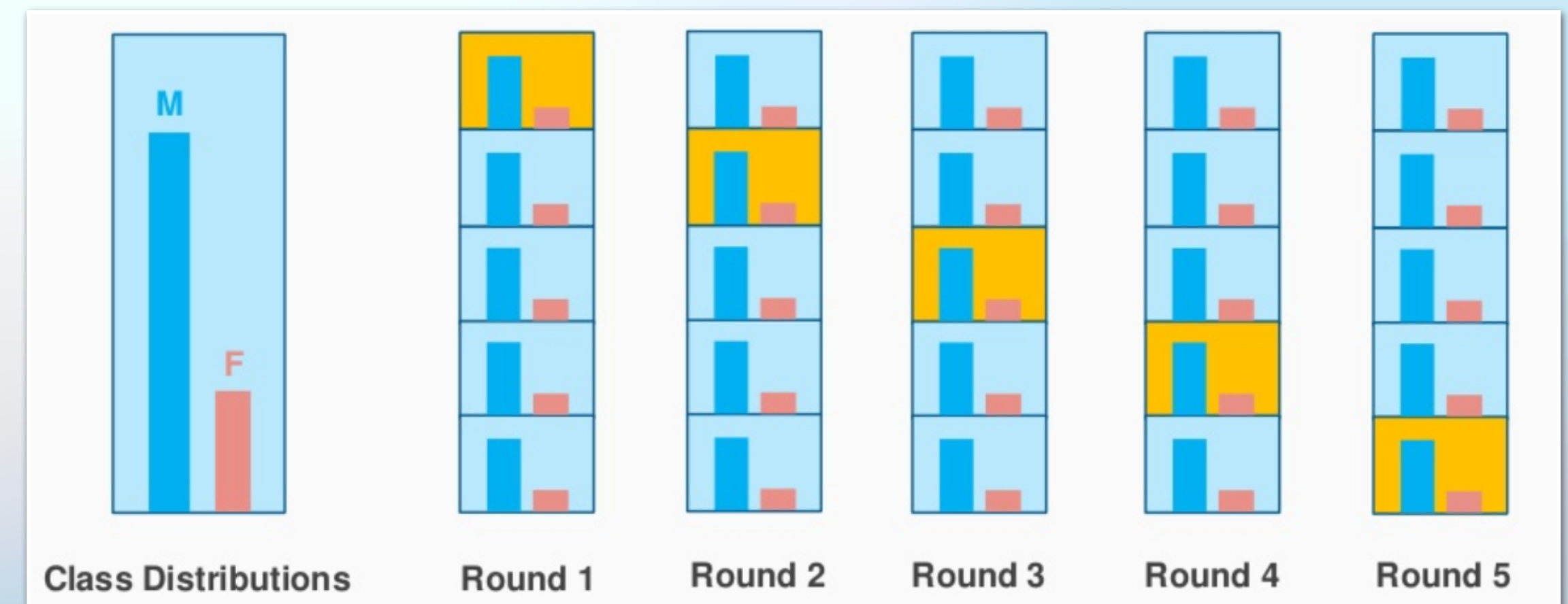
- decreasing model complexity
- add noise
- early stopping
- dropout...

Preventing underfitting:

- increasing model complexity
- engineer better or more complex features
- train for more epochs
- adjust LR



Stratified cross-validation



Machine Learning

Hyperparameters and performance

Hyperparameters

learning rate

batch size

number of epochs

architectural complexity

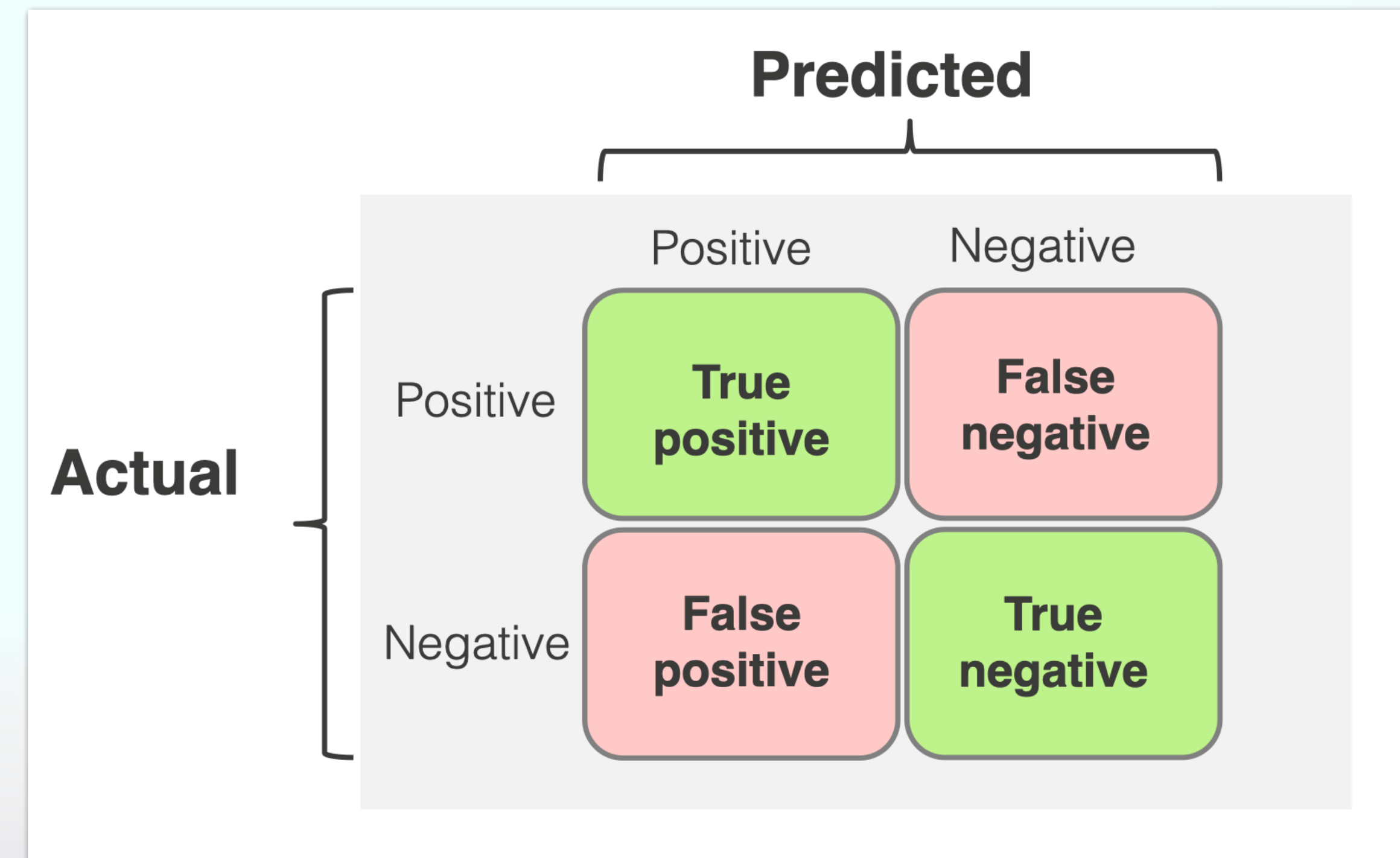
Performance metrics

Confusion matrix

Accuracy

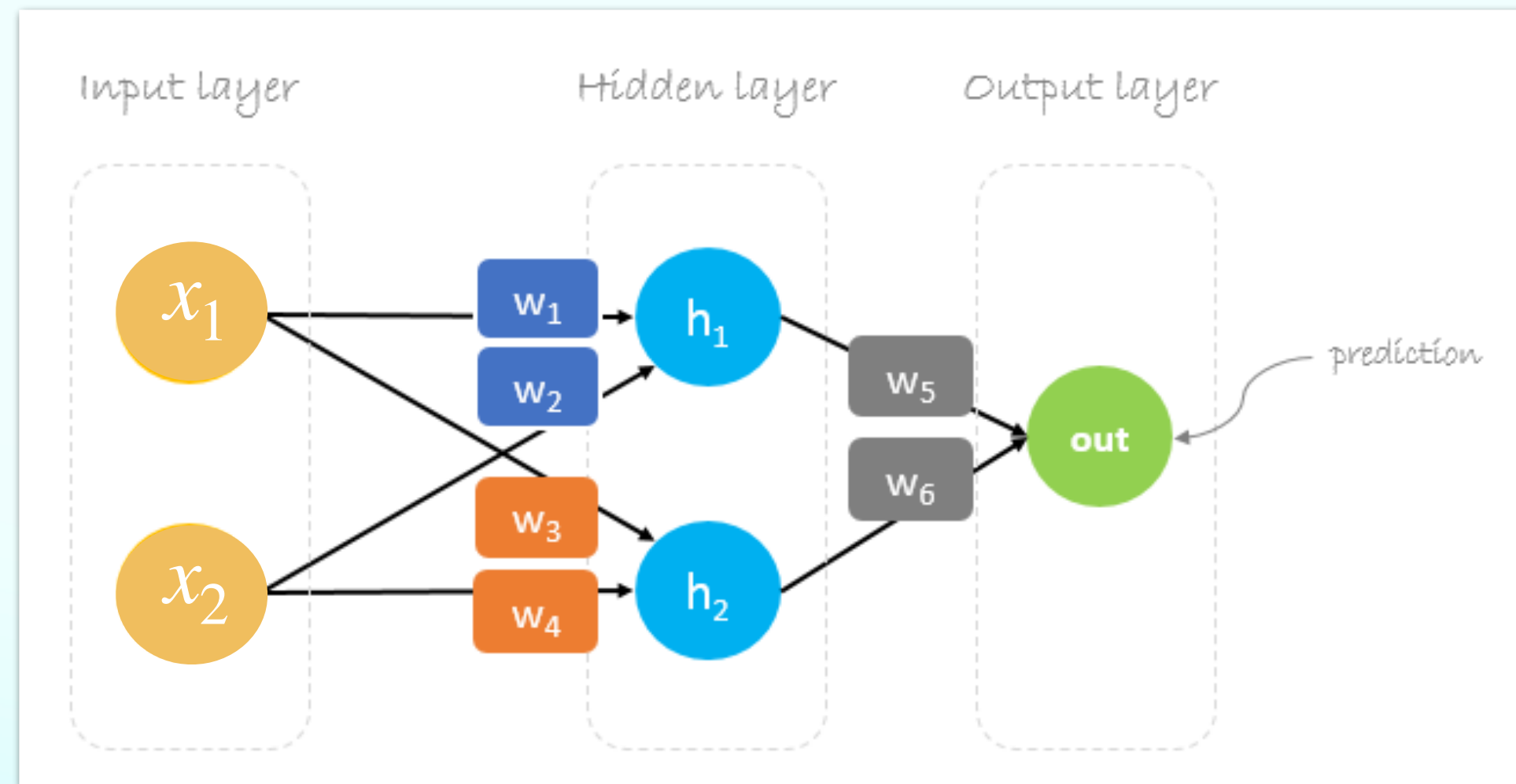
ROC and AUC

confusion matrix



Machine Learning

Deep Neural Networks



a : output of a neuron

w : weight

x : input

b : bias

σ : activation function

$$a = \sigma \left(\sum_{i=1}^n w_i x_i + b \right)$$

Approximation for any continuous function

Automated feature engineering

Scalability with data

Flexibility of architecture

Considerations in HEP:

- Input pre-processing (scale of the inputs)
- Negative event weights

Machine Learning

Transformers

- an alternative to DNNs to solve vanishing gradient problem

- core innovation: **self-attention mechanism**

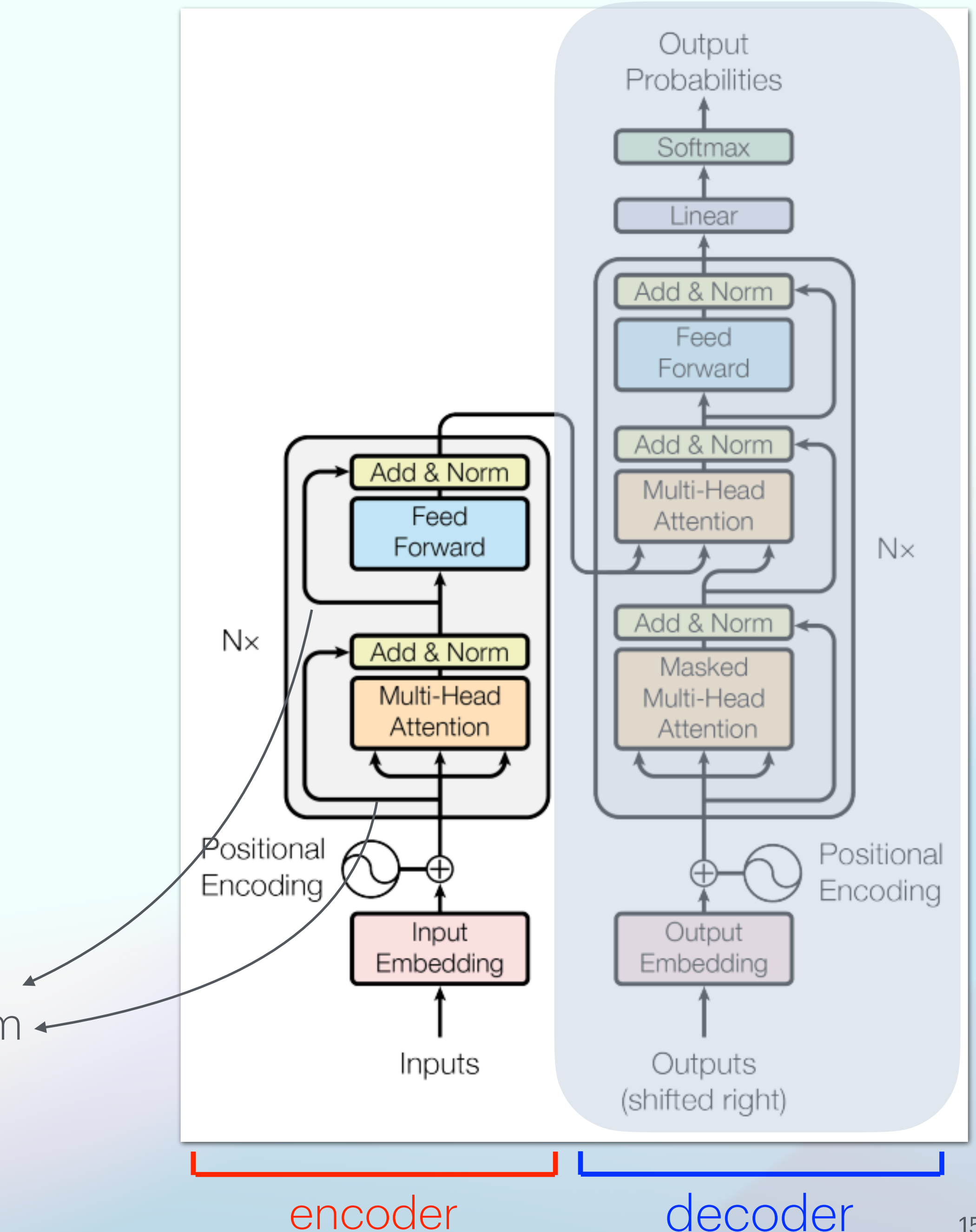
Q(query), K (key), V(value)

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right) \mathbf{V}$$

$$\text{MultiHead}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Concat}(\text{head}_1, \dots, \text{head}_h) \mathbf{W}^O$$

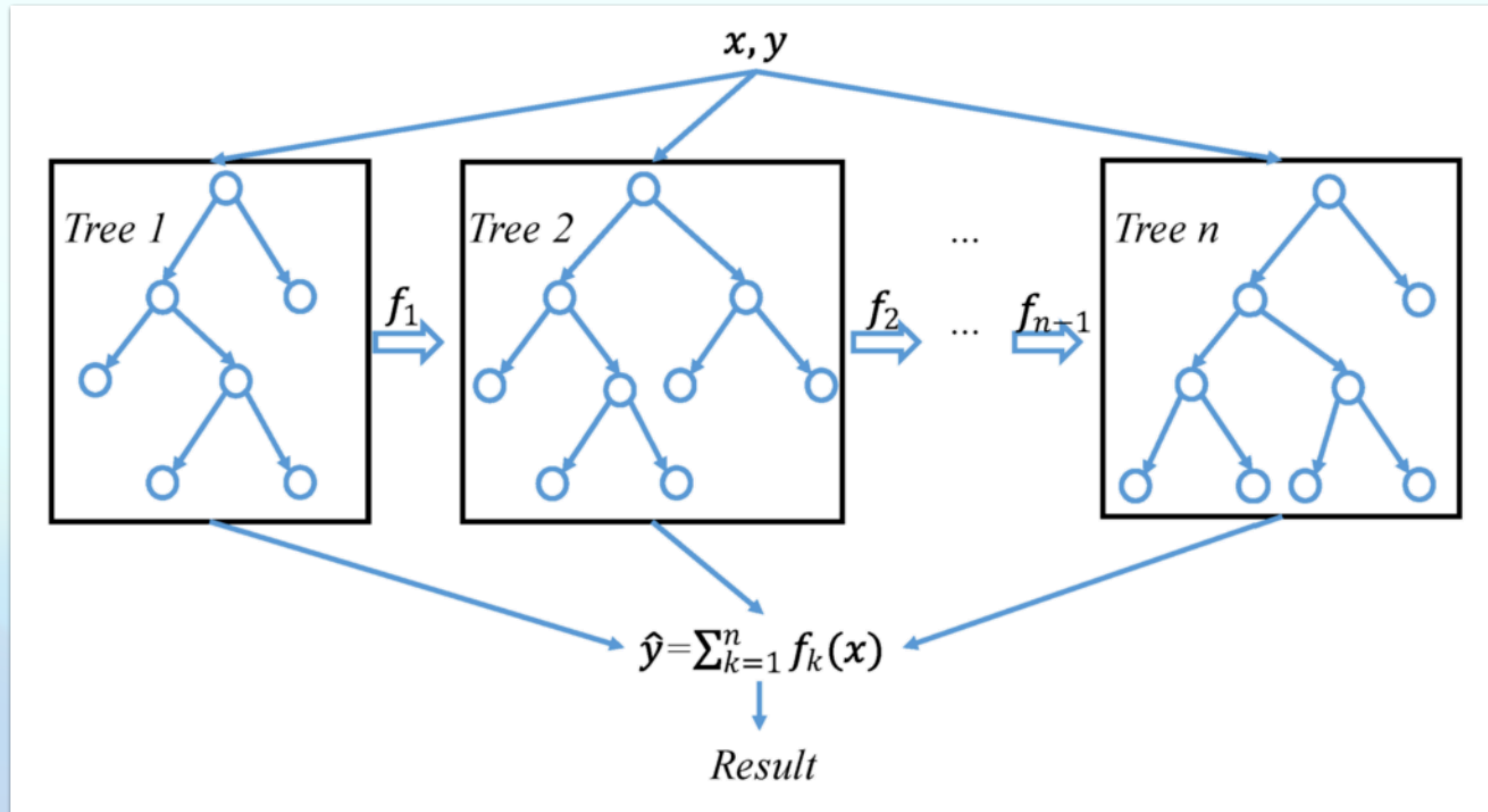
- residual connections for stable training and vanishing problem

- decoder for generative applications



Machine Learning

Boosted Decision Trees



Hyperparameters:

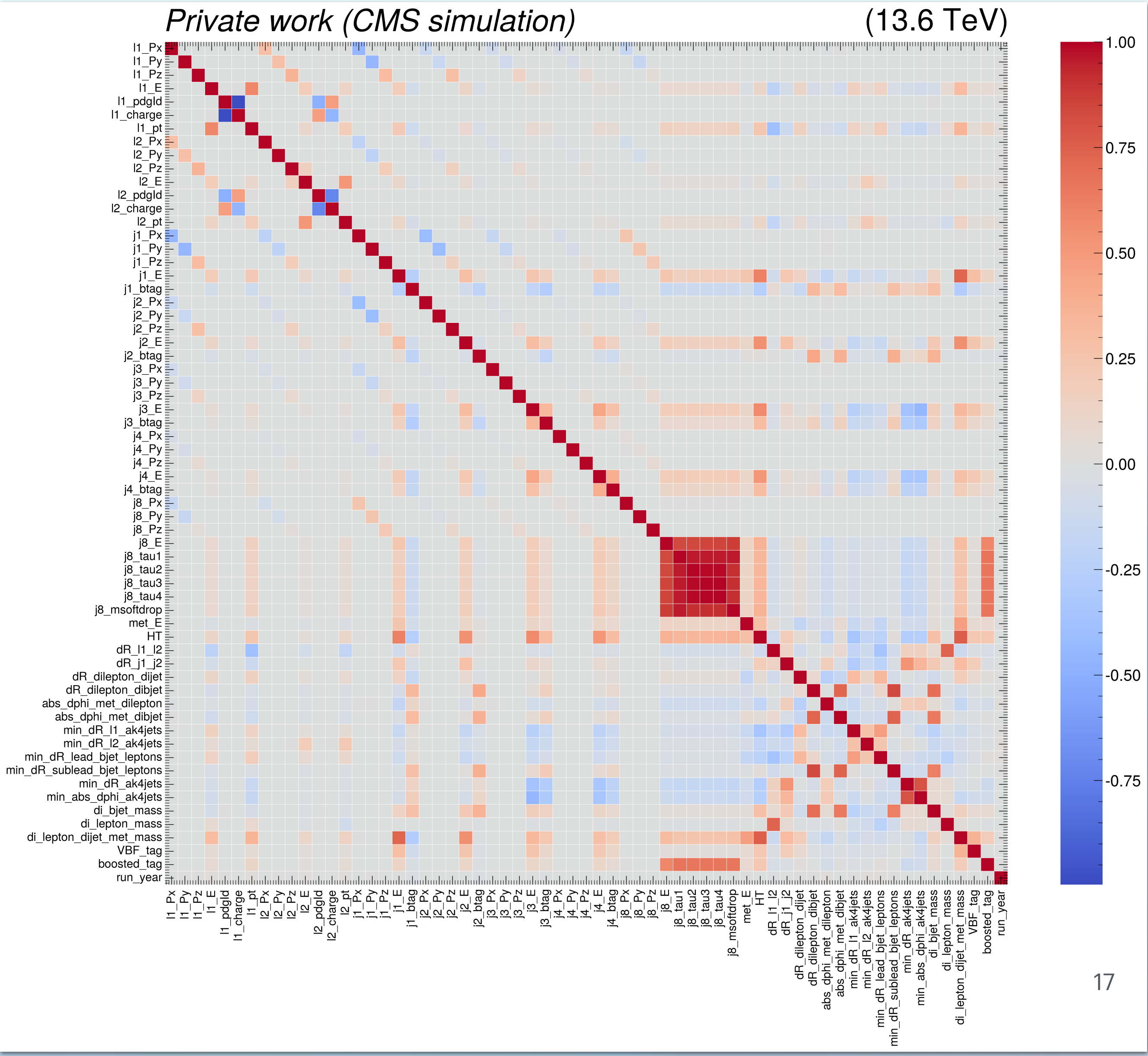
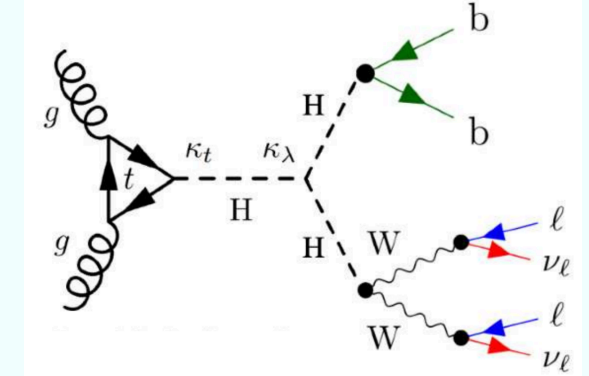
Number of boosting rounds

Tree depth (splits)

Machine Learning

Input features

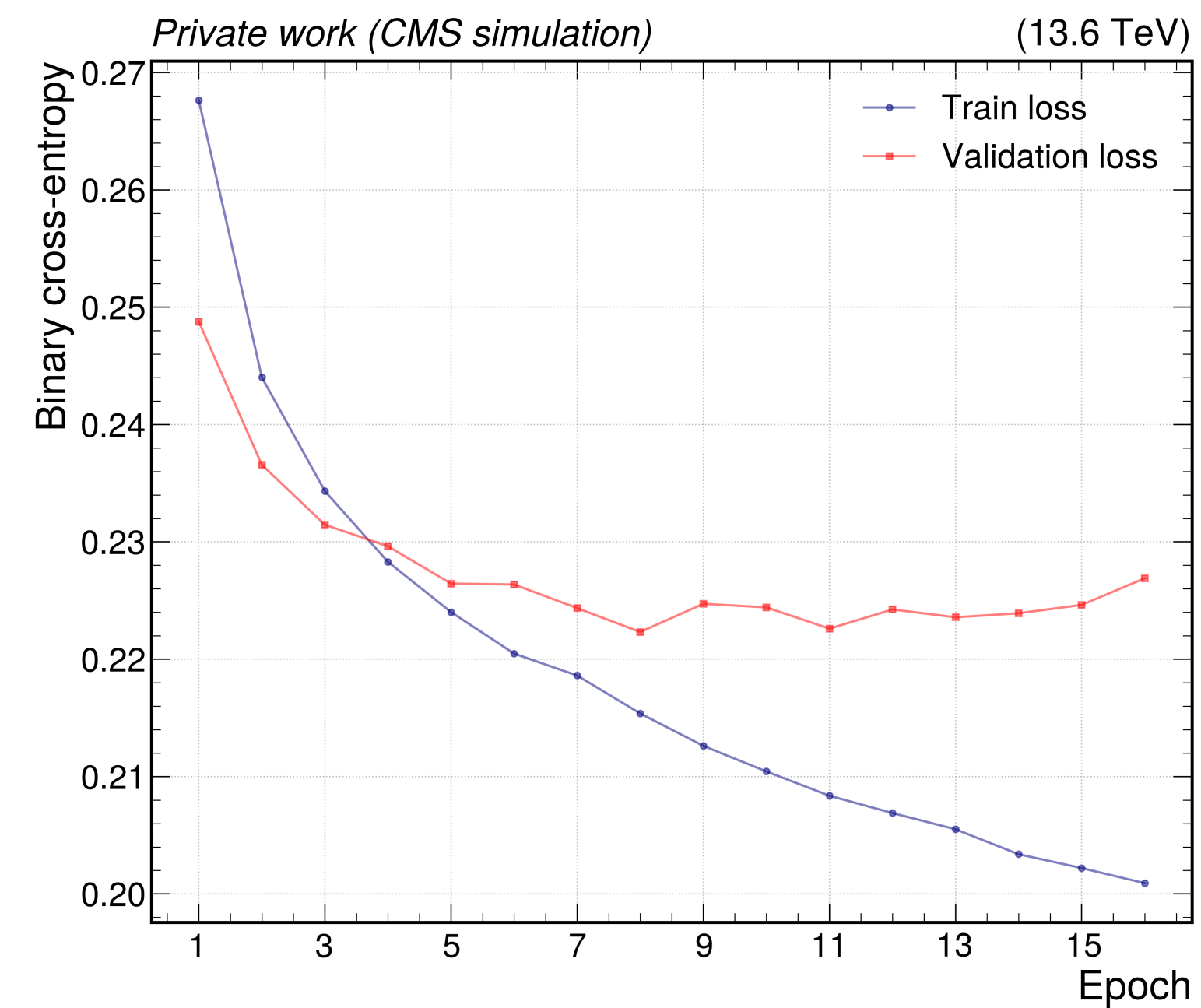
<i>l1(2)_Px</i>	x component of the (sub)leading lepton's p_T
<i>l1(2)_Py</i>	y component of the (sub)leading lepton's p_T
<i>l1(2)_Pz</i>	z component of the (sub)leading lepton's p_T
<i>l1(2)_E</i>	energy of the (sub)leading lepton
<i>l1(2)_pdgId</i>	PDG ID of the (sub)leading lepton
<i>l1(2)_charge</i>	charge of the (sub)leading lepton
<i>l1(2)_pt</i>	p_T of the (sub)leading lepton
<i>j(i)_Px</i>	x component of the i^{th} leading jet's p_T ($i = 1, 2, 3, 4$)
<i>j(i)_Py</i>	y component of the i^{th} leading jet's p_T ($i = 1, 2, 3, 4$)
<i>j(i)_Pz</i>	z component of the i^{th} leading jet's p_T ($i = 1, 2, 3, 4$)
<i>j(i)_E</i>	energy of the i^{th} leading jet ($i = 1, 2, 3, 4$)
<i>j(i)_btag</i>	b-tagging score of the i^{th} leading jet ($i = 1, 2, 3, 4$)
<i>j8_Px</i>	x component of the leading AK8 jet's p_T
<i>j8_Py</i>	y component of the leading AK8 jet's p_T
<i>j8_Pz</i>	z component of the leading AK8 jet's p_T
<i>j8_E</i>	energy of the leading AK8 jet
<i>j8_tau(N)</i>	N-subjettiness of the leading AK8 jet ($N = 1, 2, 3, 4$)
<i>j8_msoftdrop</i>	soft drop mass of the leading AK8 jet
<i>met_E</i>	magnitude of the missing transverse energy
<i>HT</i>	scalar sum of the p_T of all jets
<i>dR_l1_l2</i>	ΔR between the two leptons
<i>dR_j1_j2</i>	ΔR between the two leading jets
<i>dR_dilepton_dijet</i>	ΔR between the dilepton and dijet systems
<i>dR_dilepton_dibjet</i>	ΔR between the dilepton and dibjet systems
<i>abs_dphi_met_dilepton</i>	$ \Delta\phi $ between the missing transverse energy and the dilepton system
<i>abs_dphi_met_dibjet</i>	$ \Delta\phi $ between the missing transverse energy and the dibjet system
<i>min_dR_l1_ak4jets</i>	Minimum ΔR between the leading lepton and all AK4 jets
<i>min_dR_l2_ak4jets</i>	Minimum ΔR between the subleading lepton and all AK4 jets
<i>min_dR_lead_bjet_leptons</i>	Minimum ΔR between the leading b-tagged jet and both leptons
<i>min_dR_sublead_bjet_leptons</i>	Minimum ΔR between the subleading b-tagged jet and both leptons
<i>min_abs_dphi_ak4jets</i>	Minimum $\Delta\phi$ between all AK4 jets and the missing transverse energy
<i>di_bjet_mass</i>	Mass of the dijet system formed by the two leading b-tagged jets
<i>di_lepton_mass</i>	Mass of the dilepton system
<i>di_lepton_dijet_met_mass</i>	Mass of the dilepton + dijet + missing transverse energy system
<i>VBF_tag</i>	Binary VBF tagging variable
<i>boosted_tag</i>	Binary boosted topology tagging variable
<i>run_year</i>	Run year of the event (2022 or 2023)



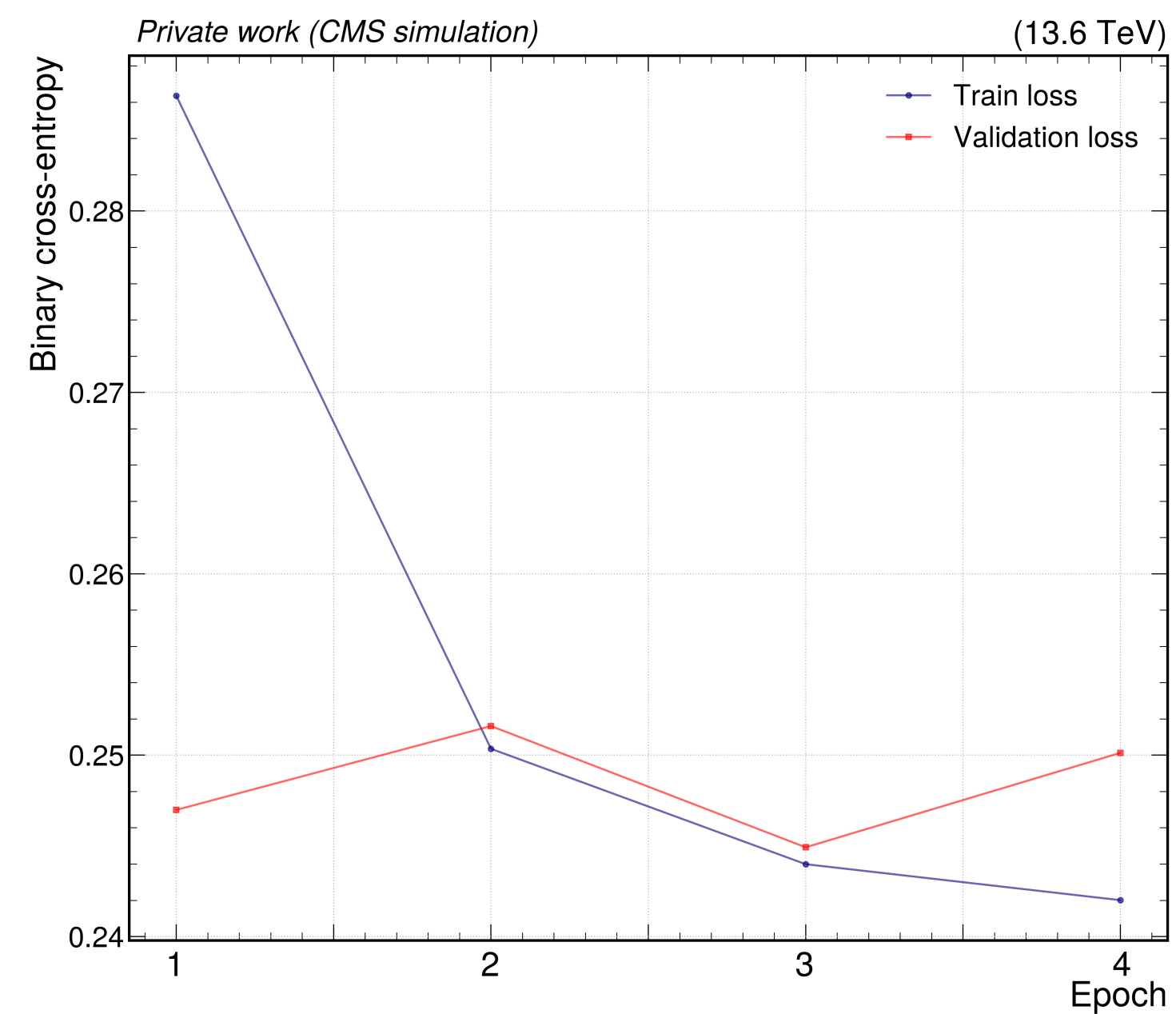
Machine Learning

Comparison of models - Loss

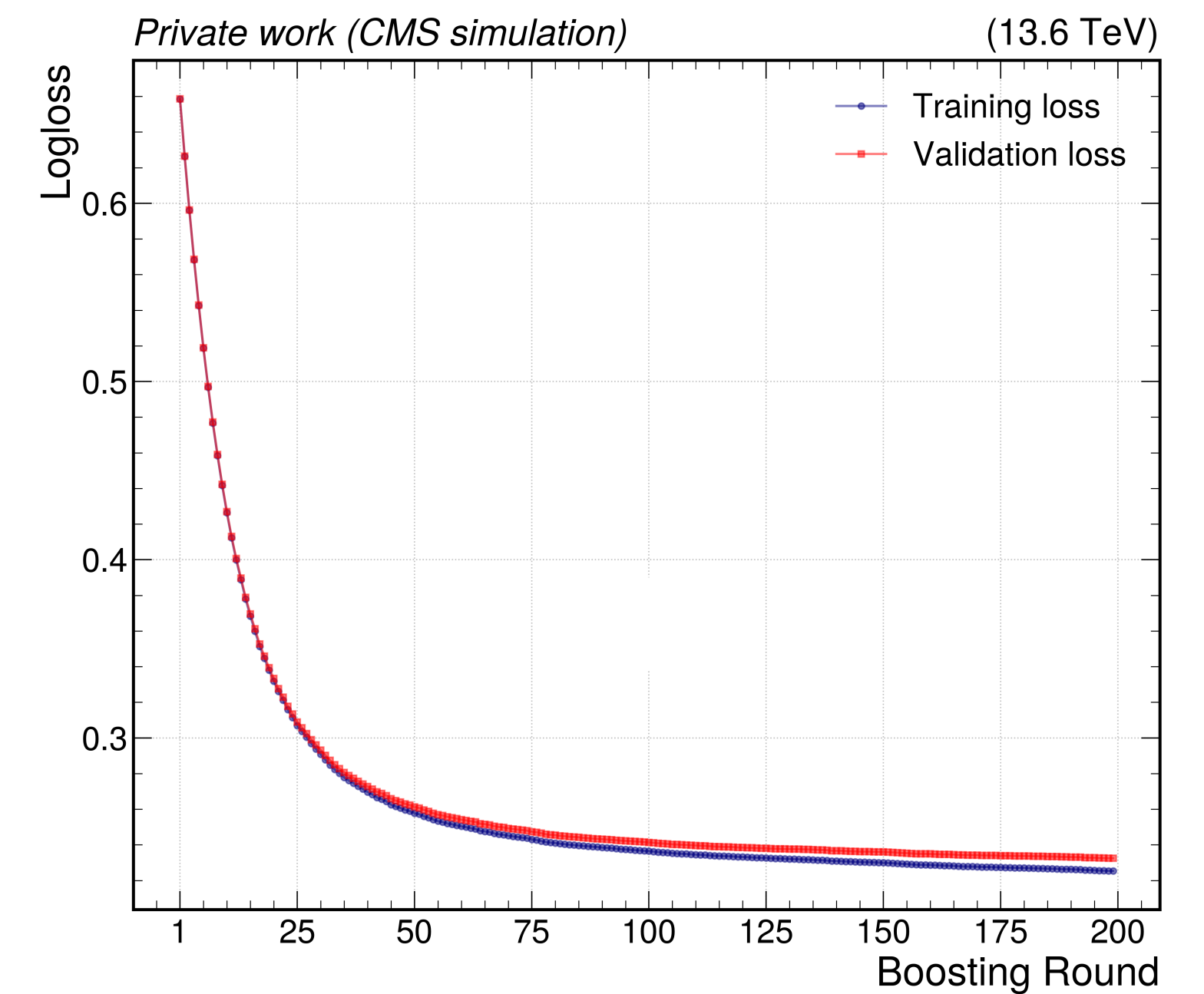
DNN



Transformer



BDT



Machine Learning

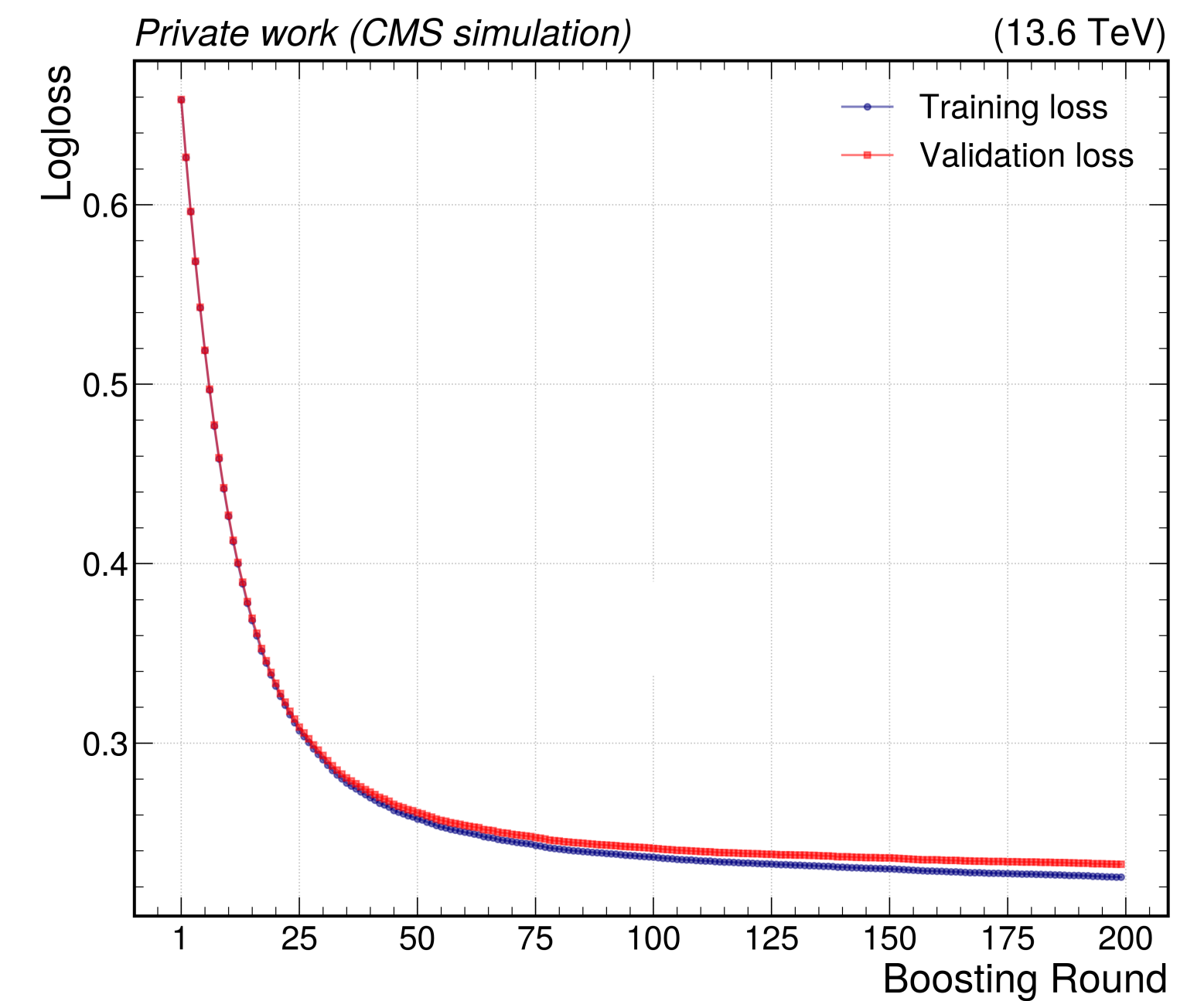
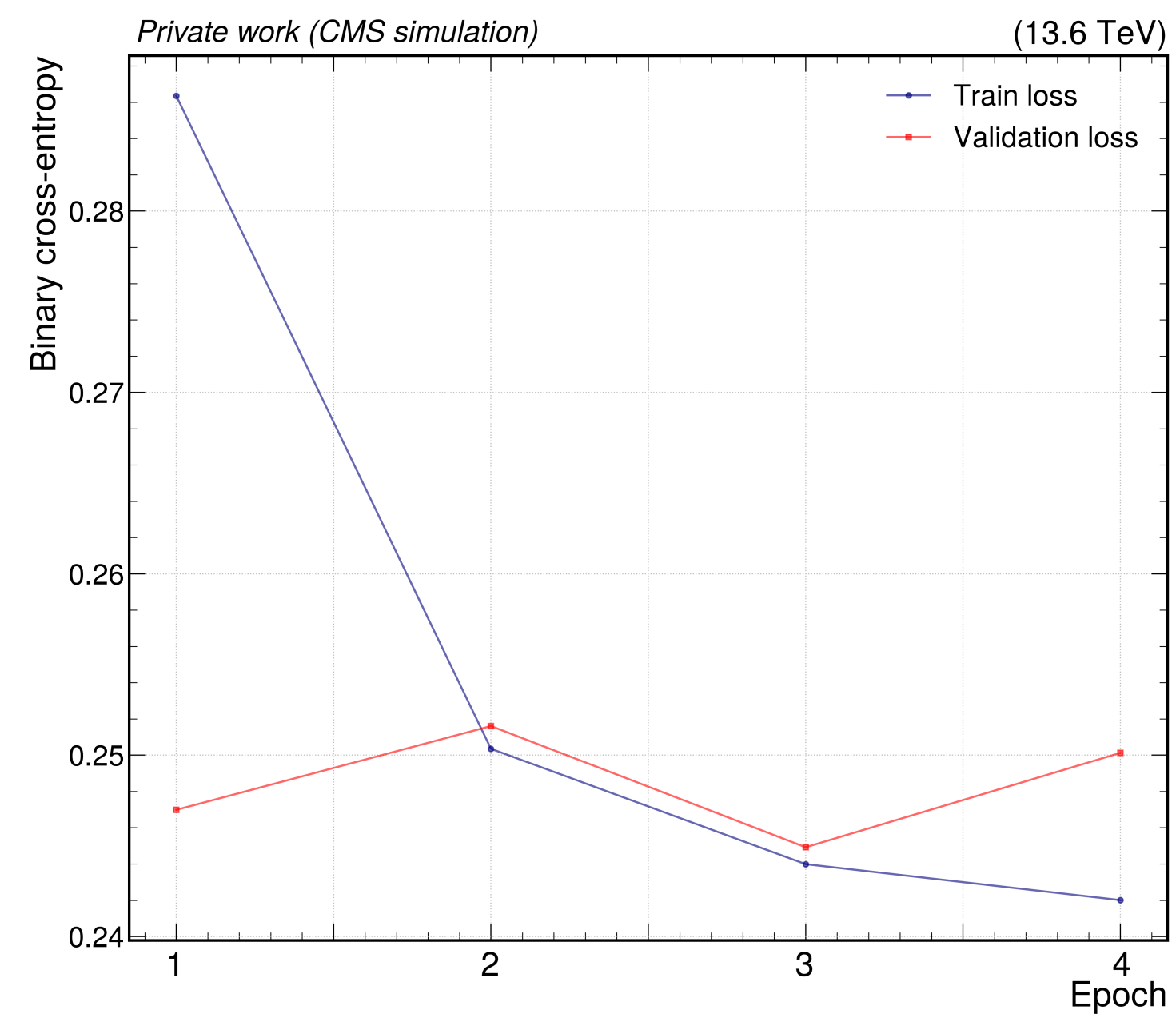
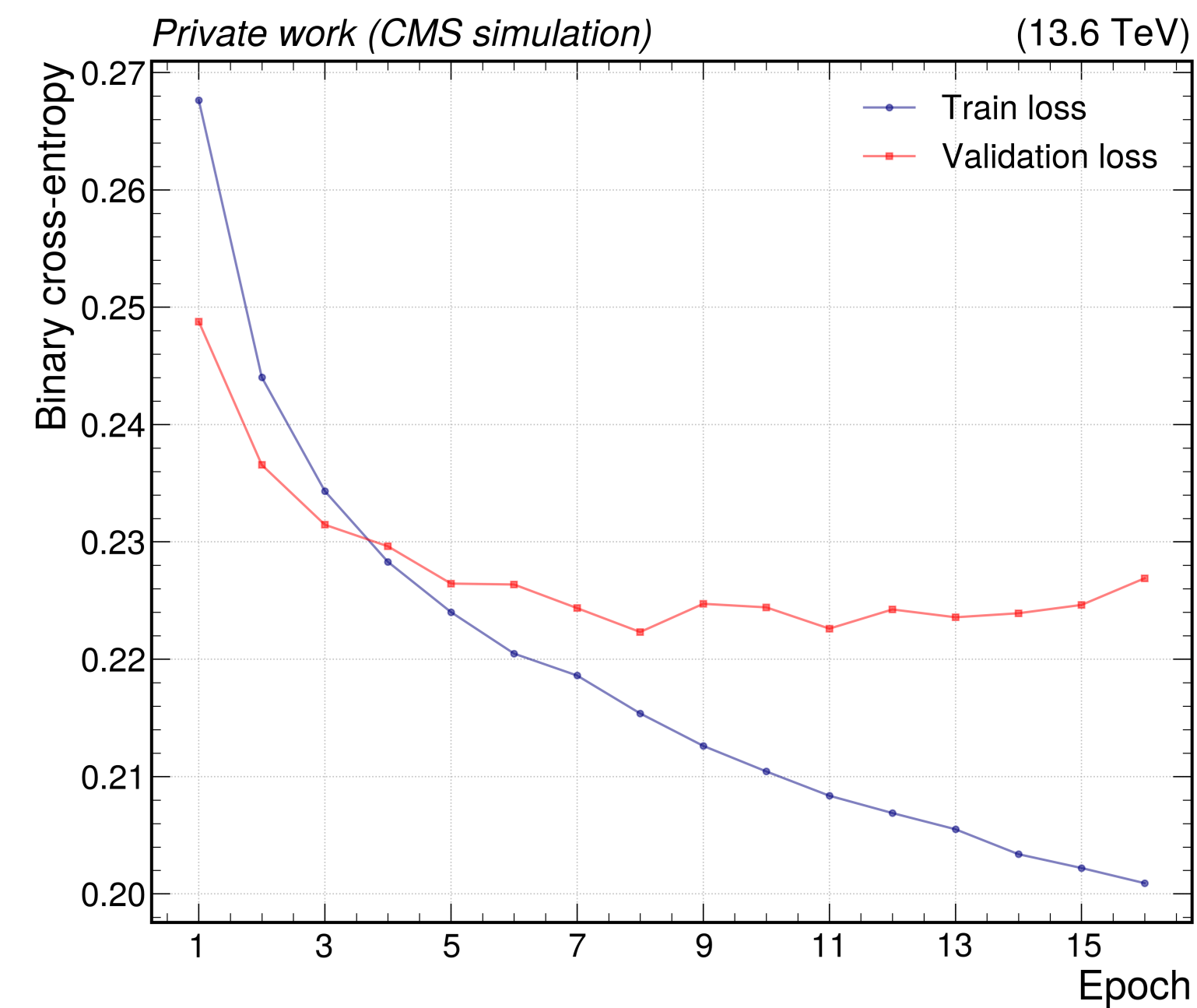
Comparison of models - Loss

DNN : 0
Transformer : 0
BDT : 1

DNN

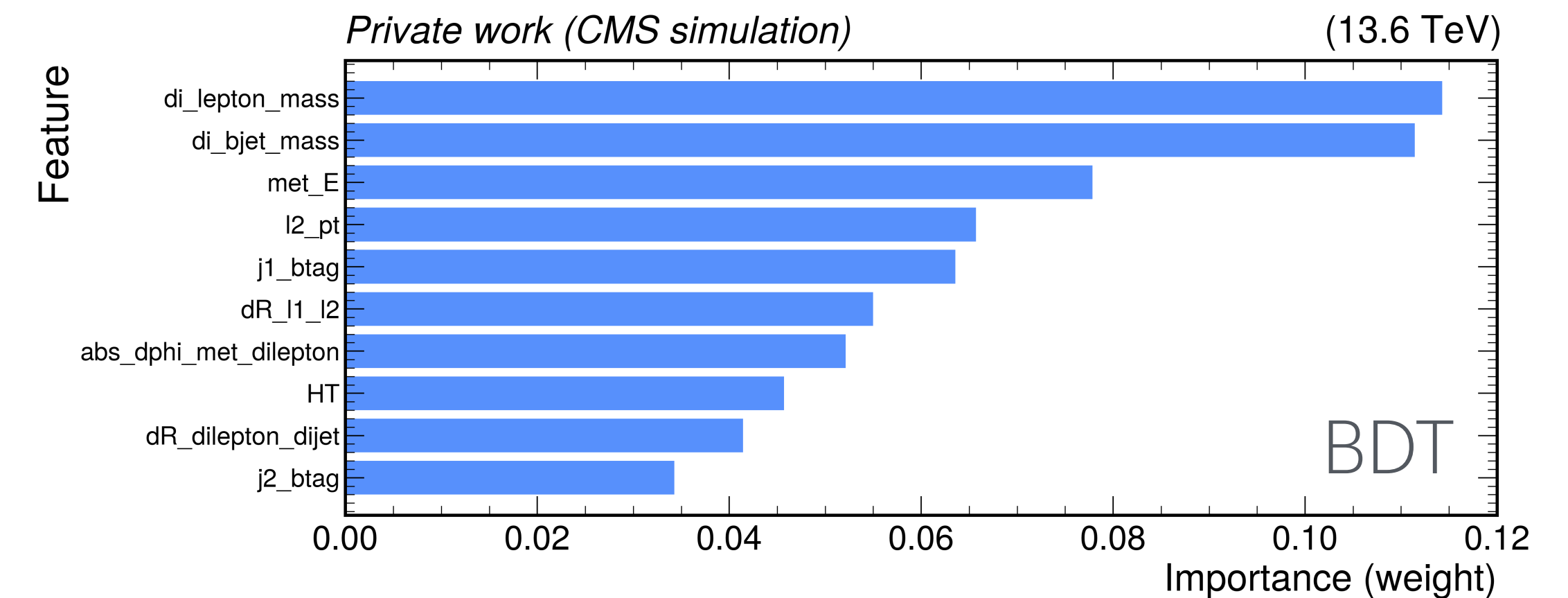
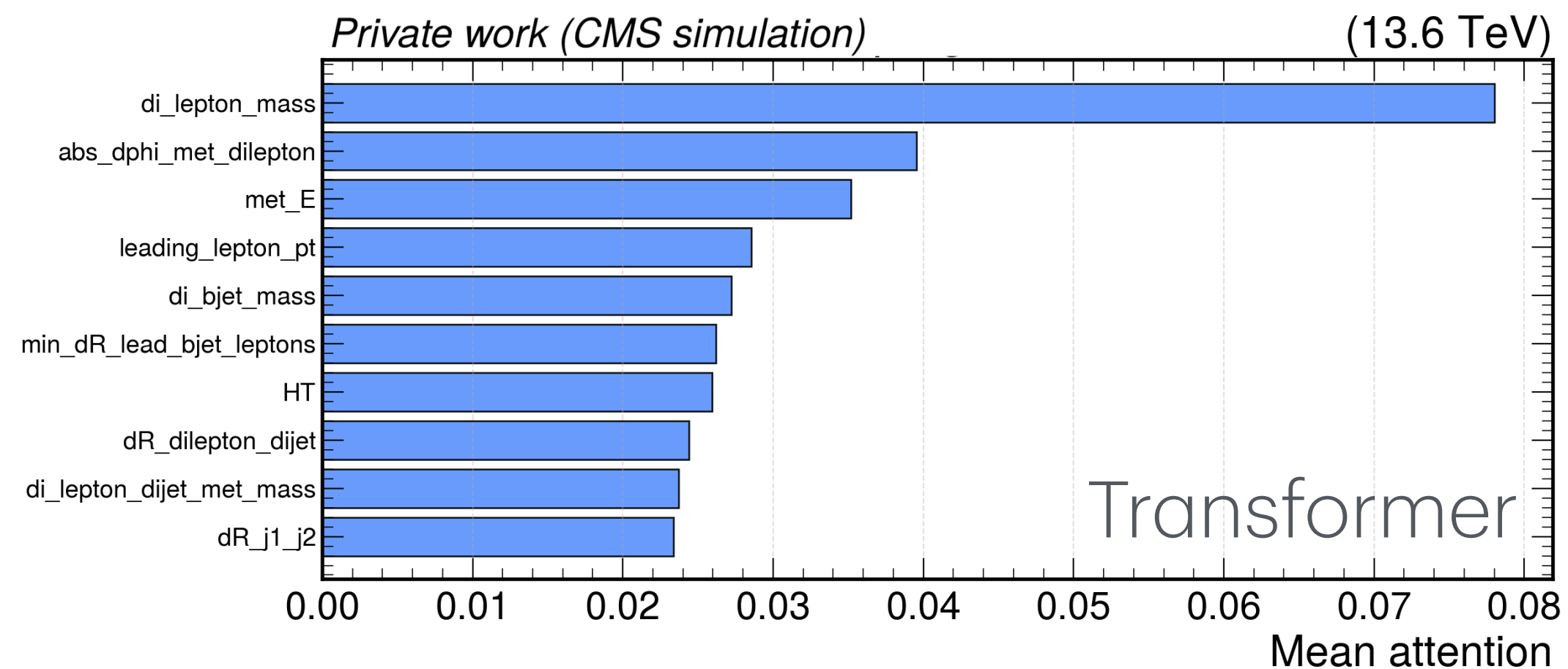
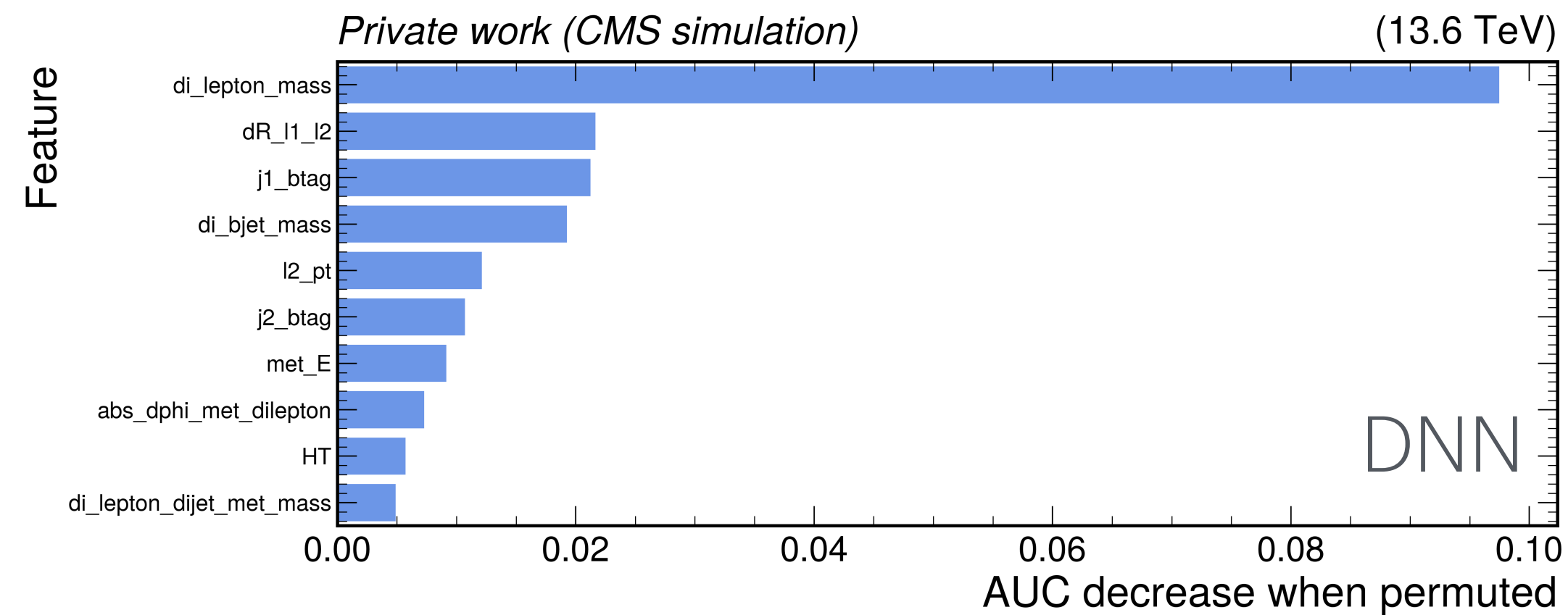
Transformer

BDT



Machine Learning

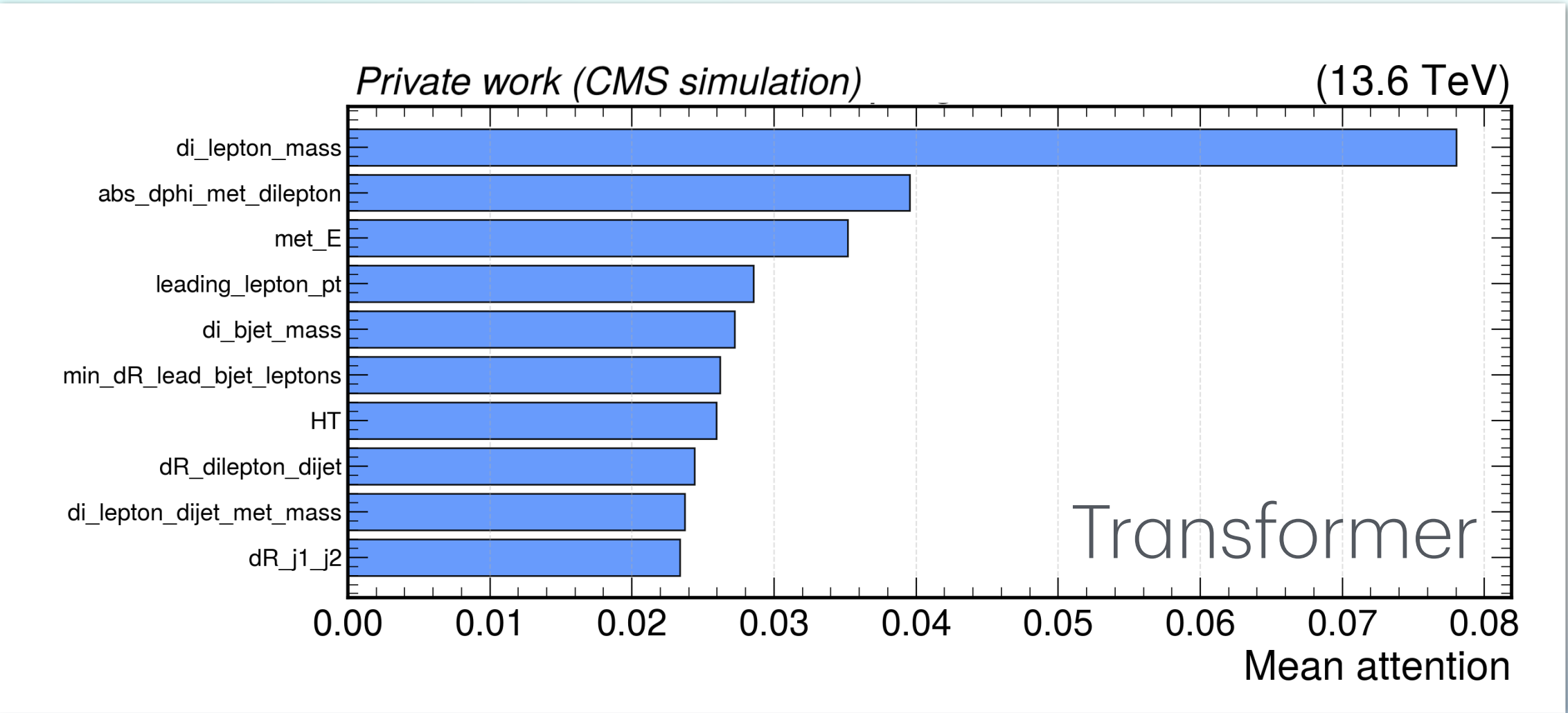
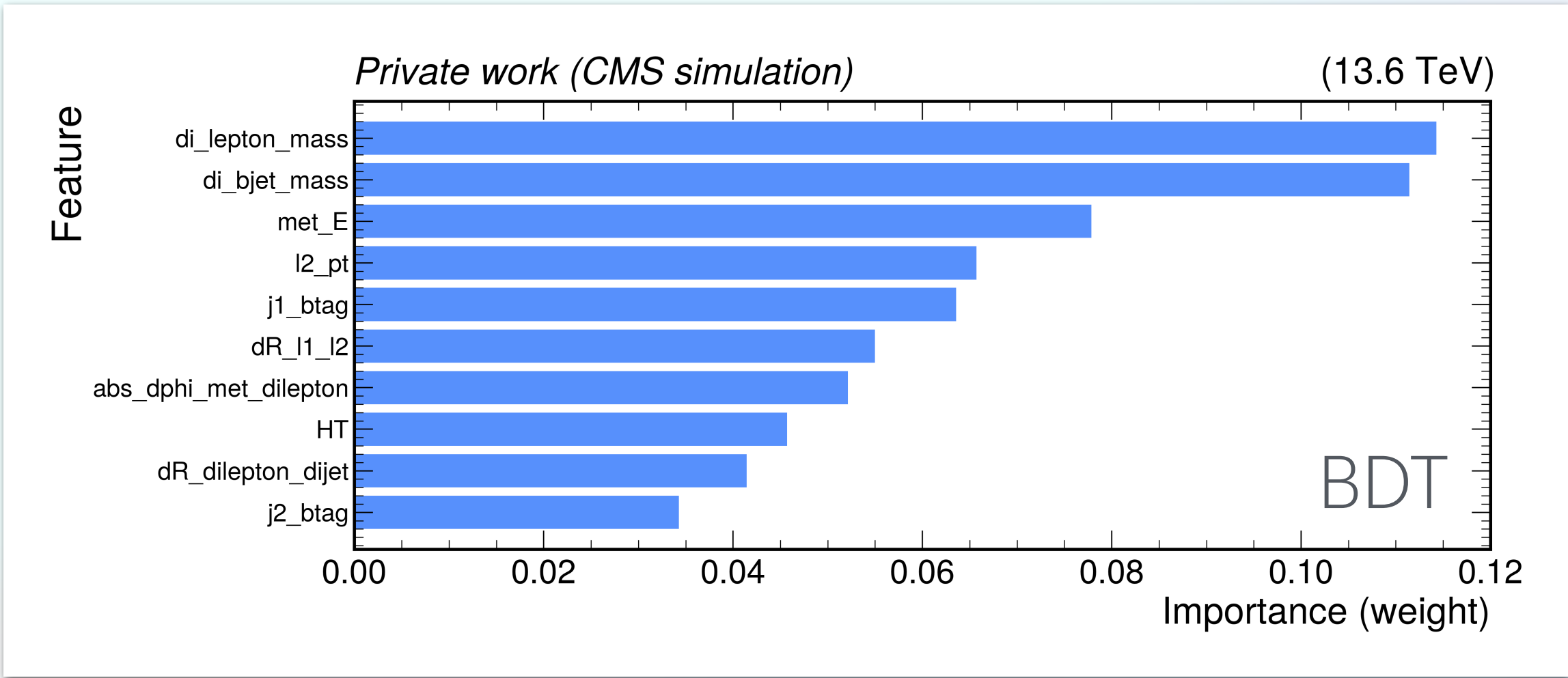
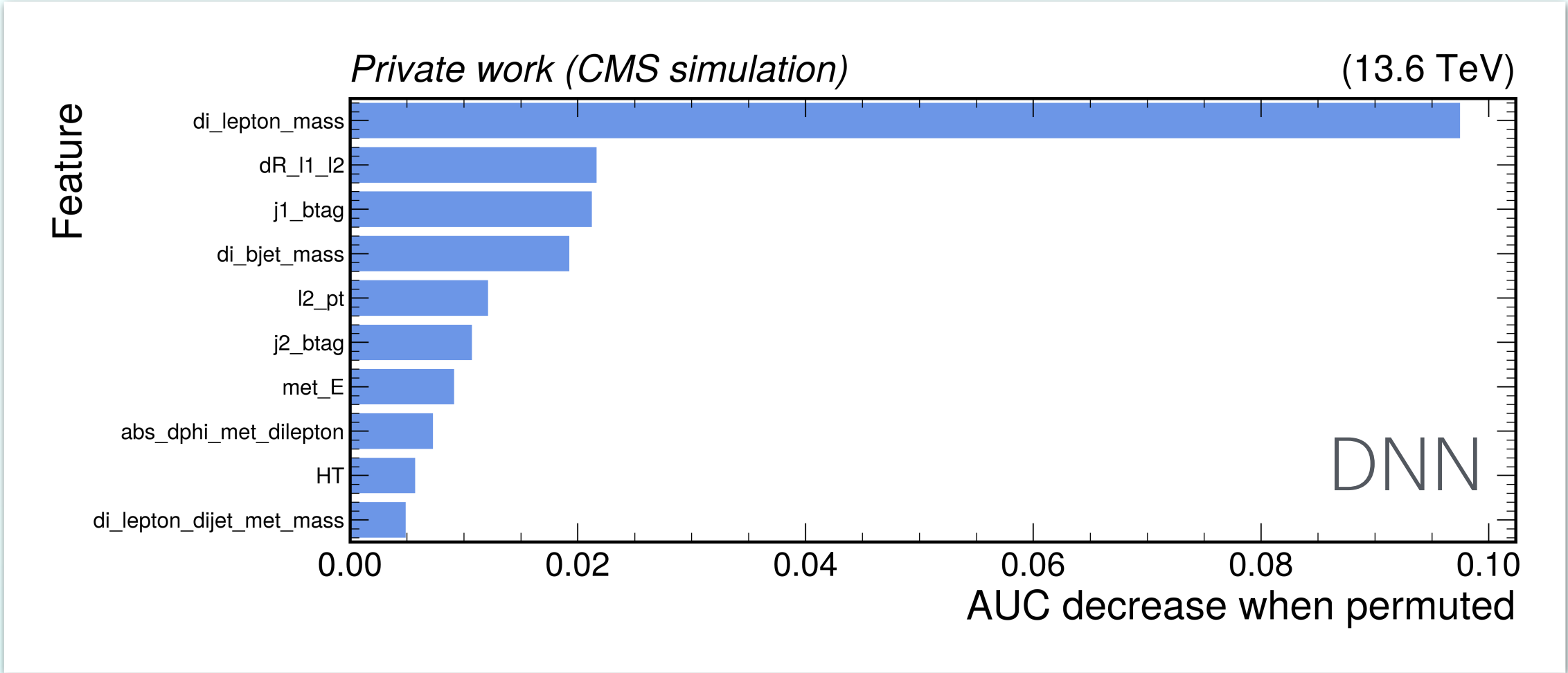
Comparison of models - Interpretability



Machine Learning

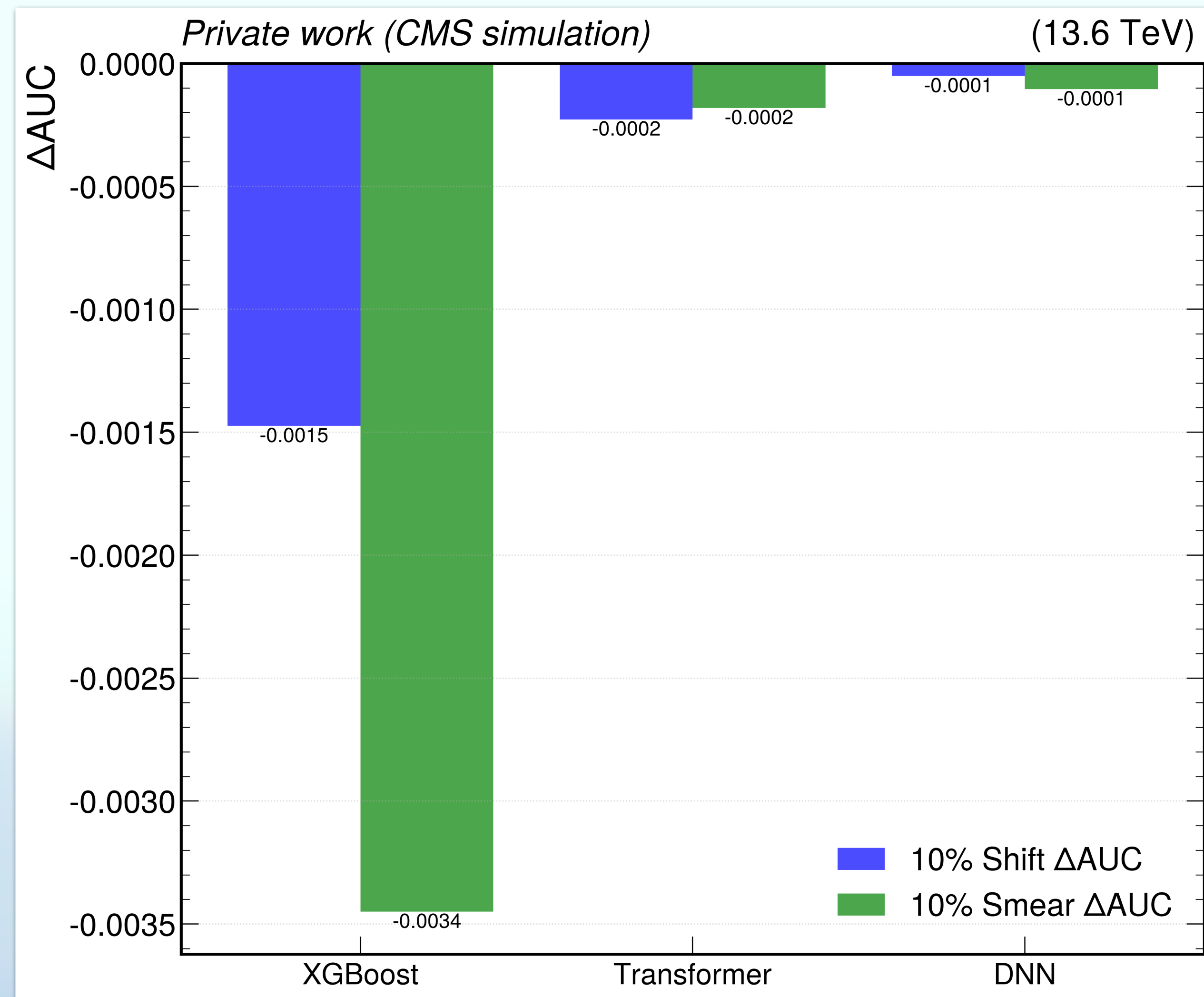
Comparison of models - Interpretability

DNN	: 0
Transformer	: 0
BDT	: 2



Machine Learning

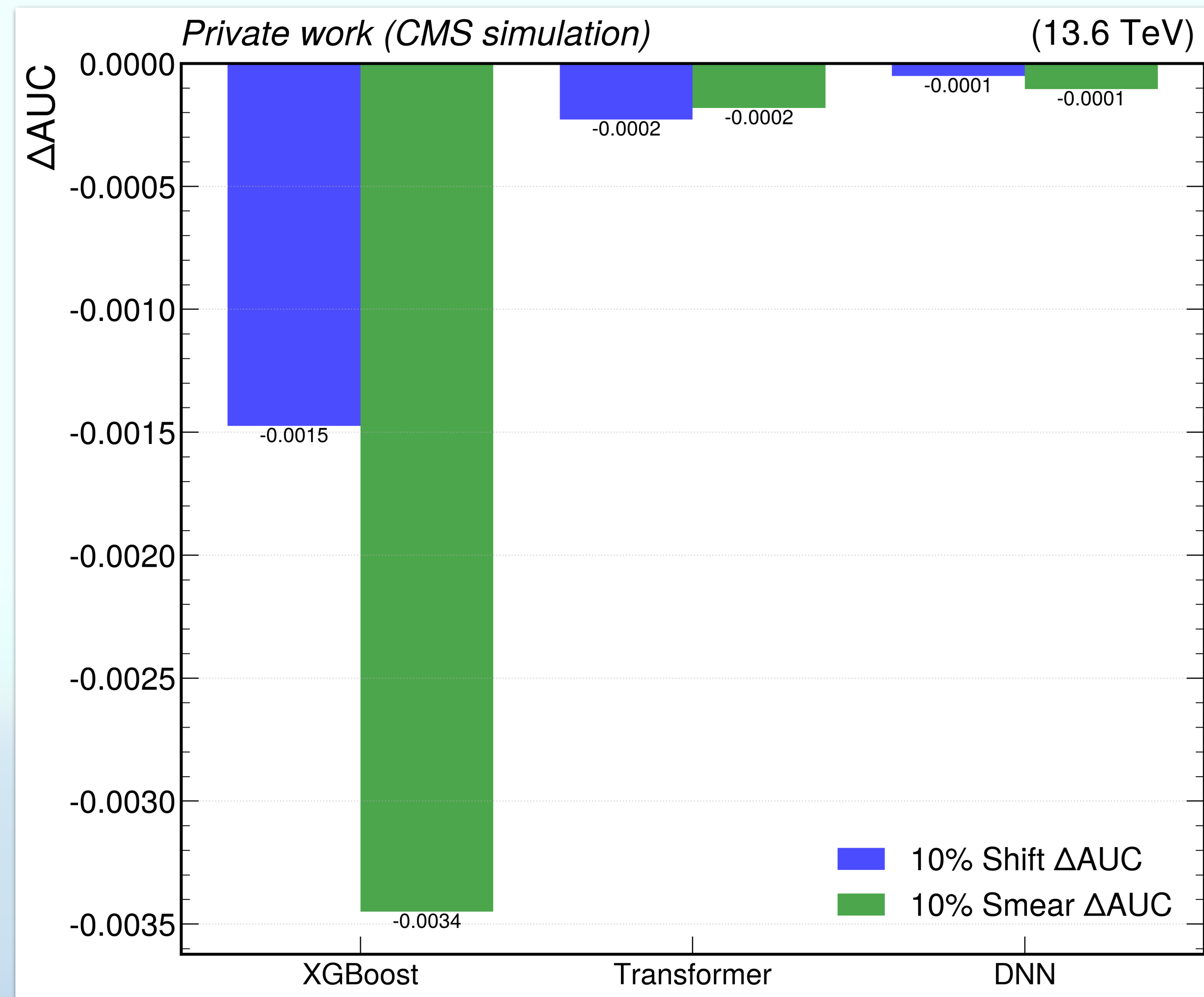
Comparison of models - Robustness



Machine Learning

Comparison of models - Robustness

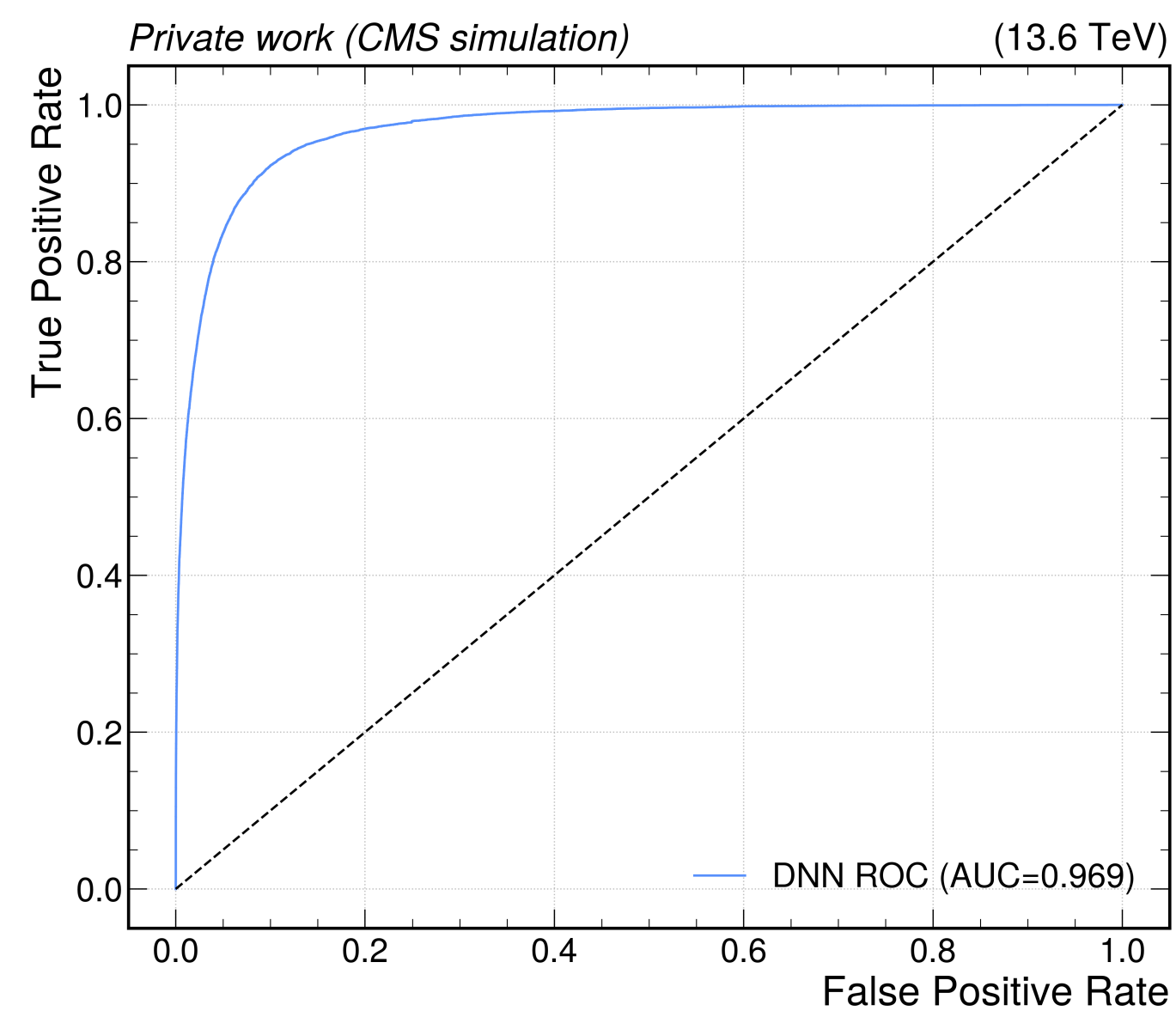
DNN : 1
Transformer : 1
BDT : 2



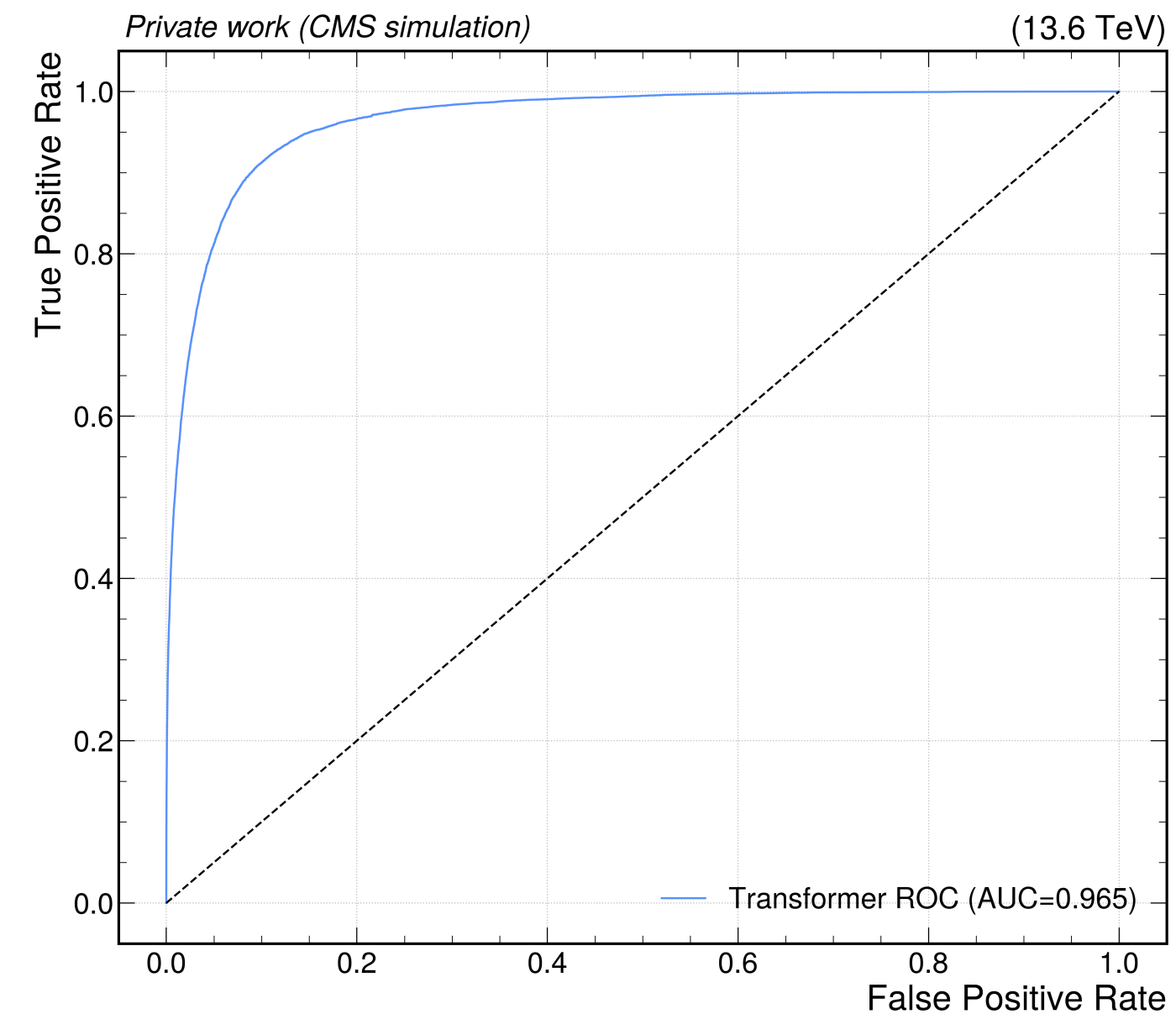
Machine Learning

Comparison of models - Performance (ROC curves and training time)

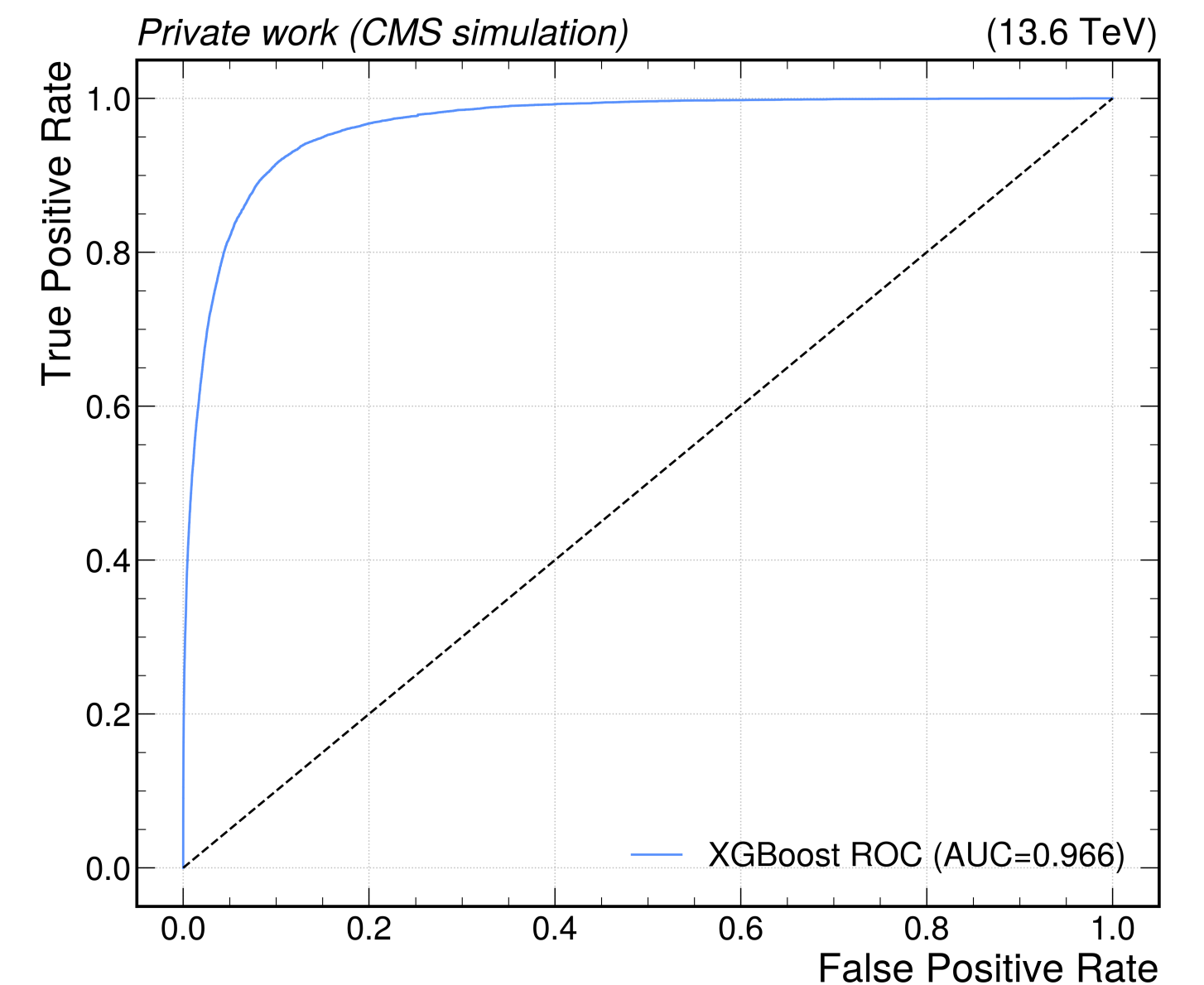
DNN



Transformer



BDT



Model	DNN	Transformer	BDT
AUC	0.965	0.969	0.966

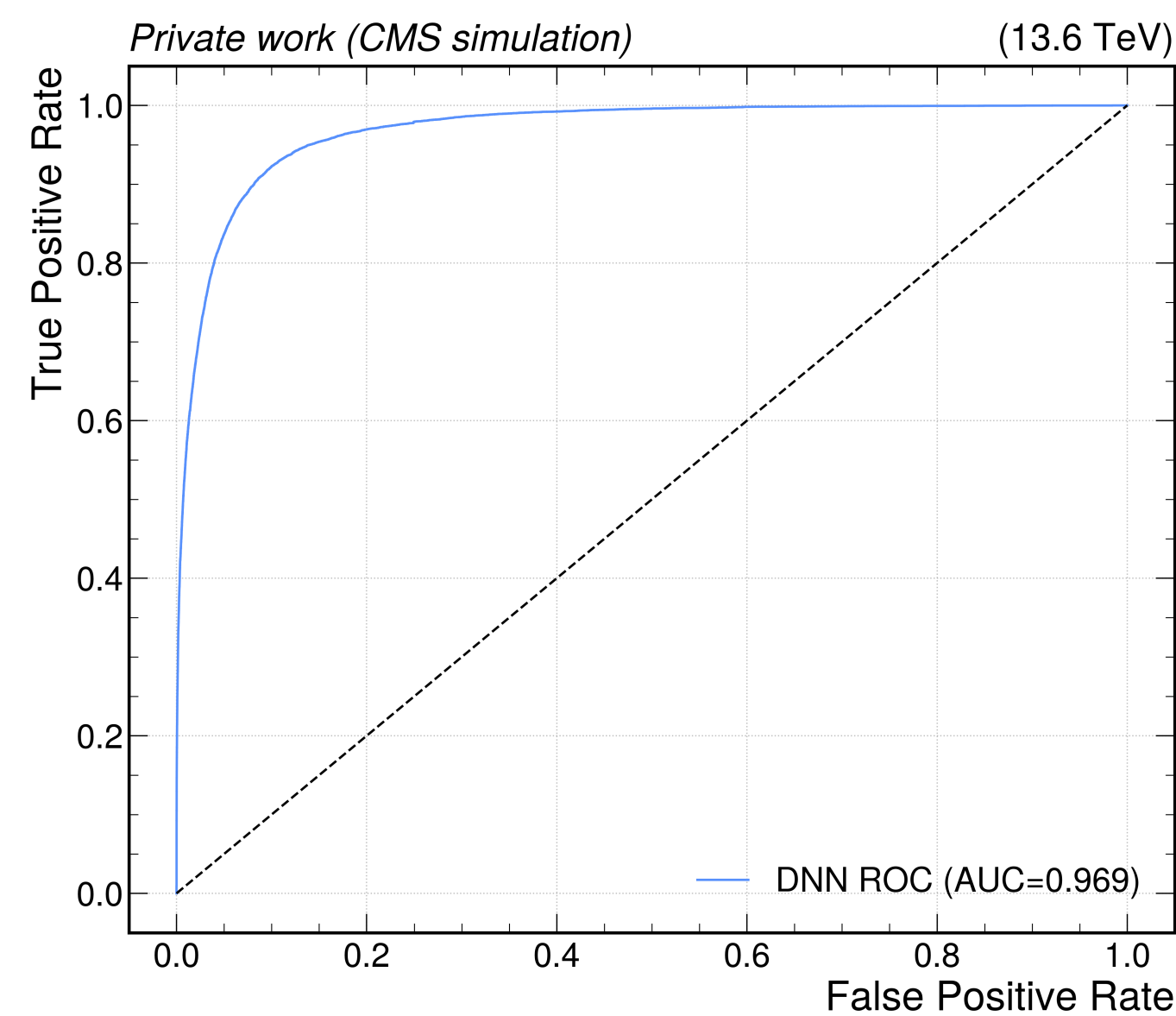
Model	Total Time	Epochs [*]	Time / Epoch	Early Stopping
Transformer	90 min	4	22.5 min	No
DNN	28 min	16	106 s	Yes
BDT	45 s	200	0.2 s	No

Machine Learning

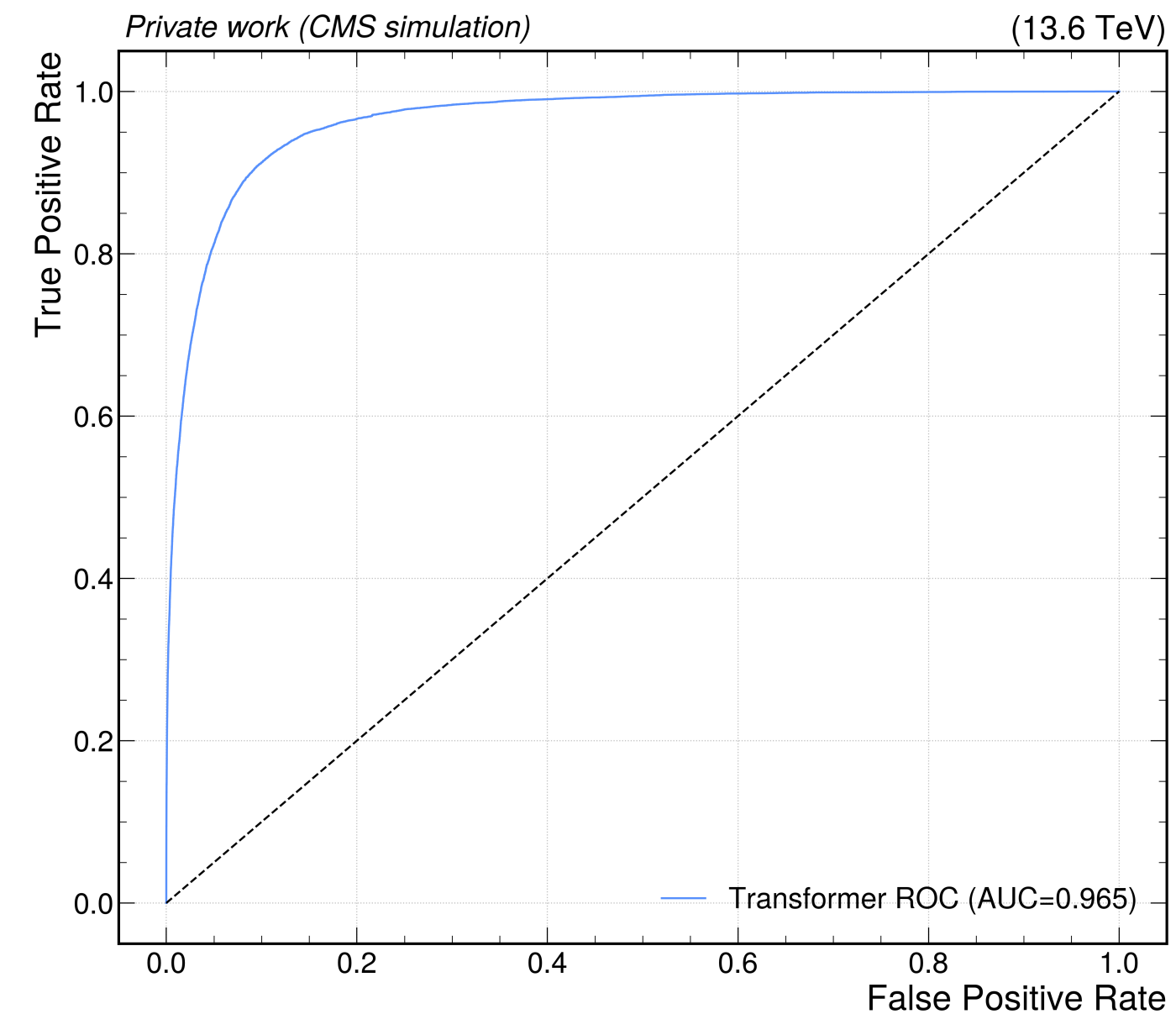
DNN : 1
Transformer : 1
BDT : 3

Comparison of models - Performance (ROC curves and training time)

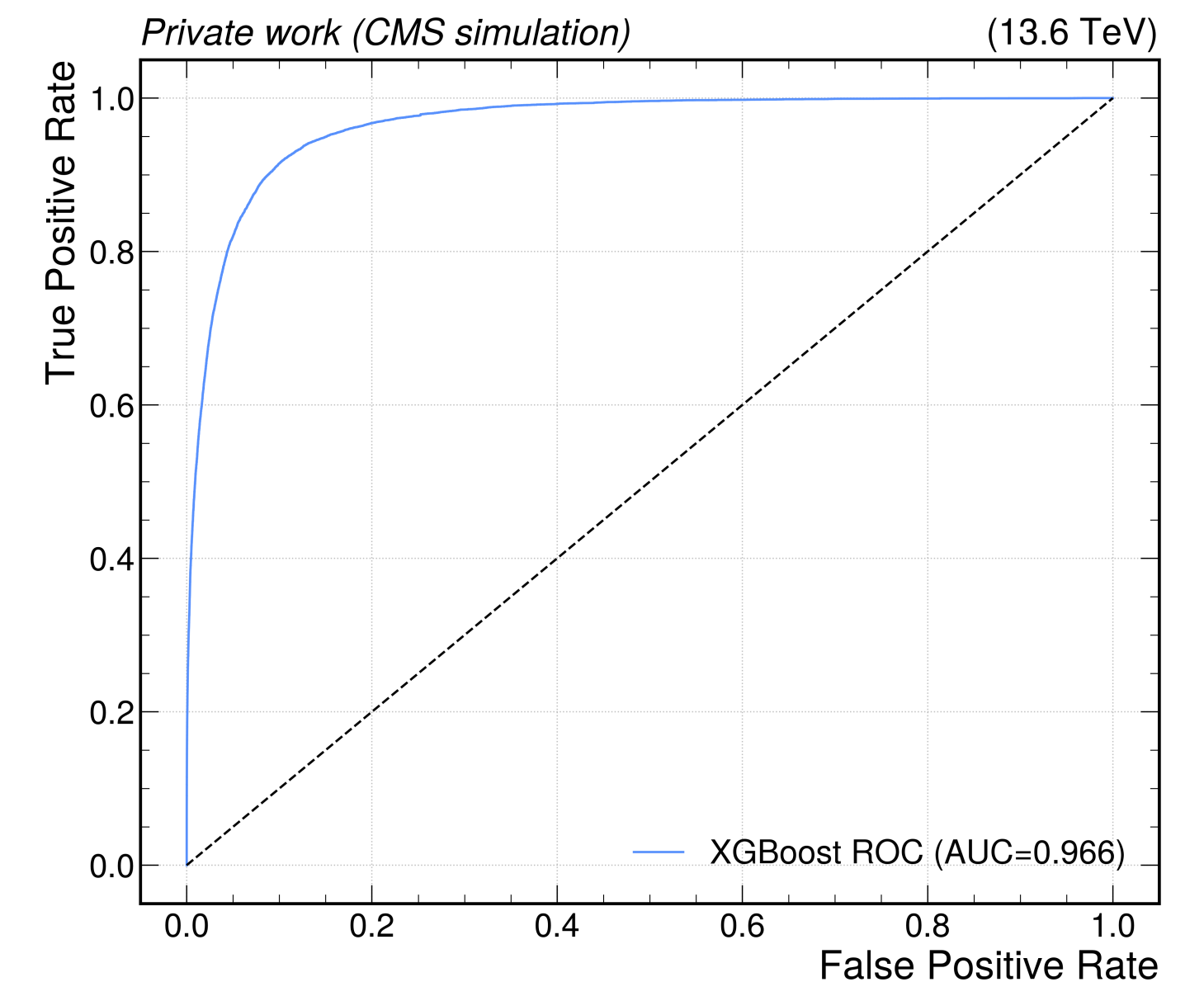
DNN



Transformer



BDT



Model	DNN	Transformer	BDT
AUC	0.965	0.969	0.966

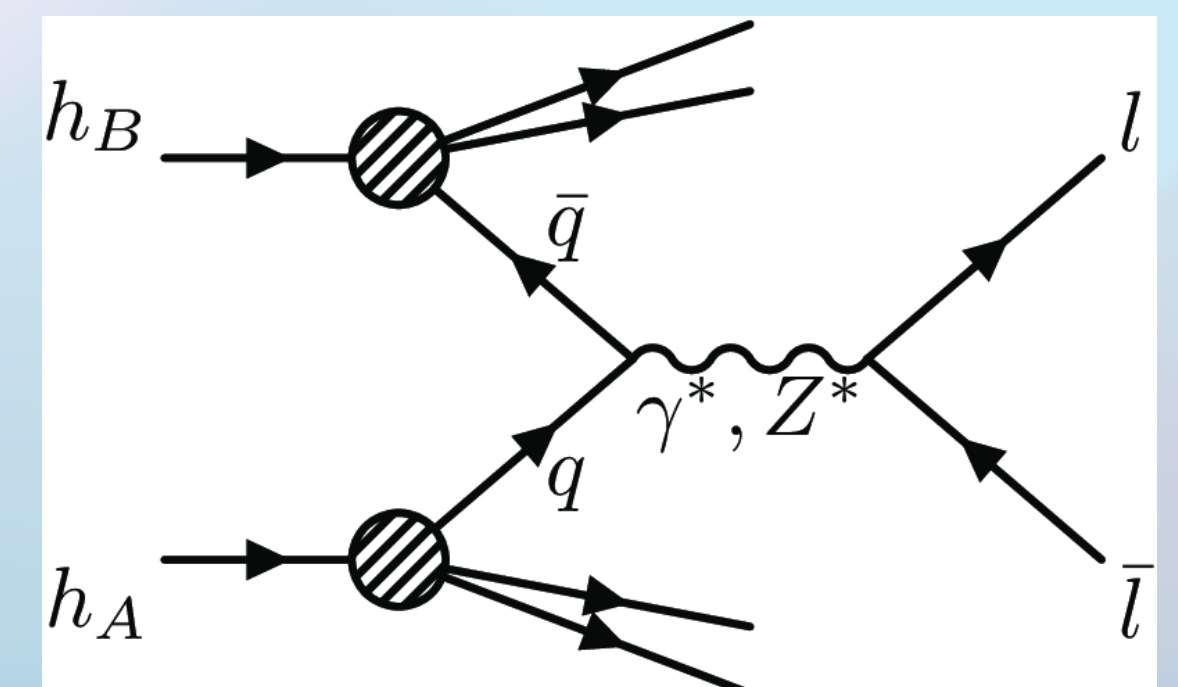
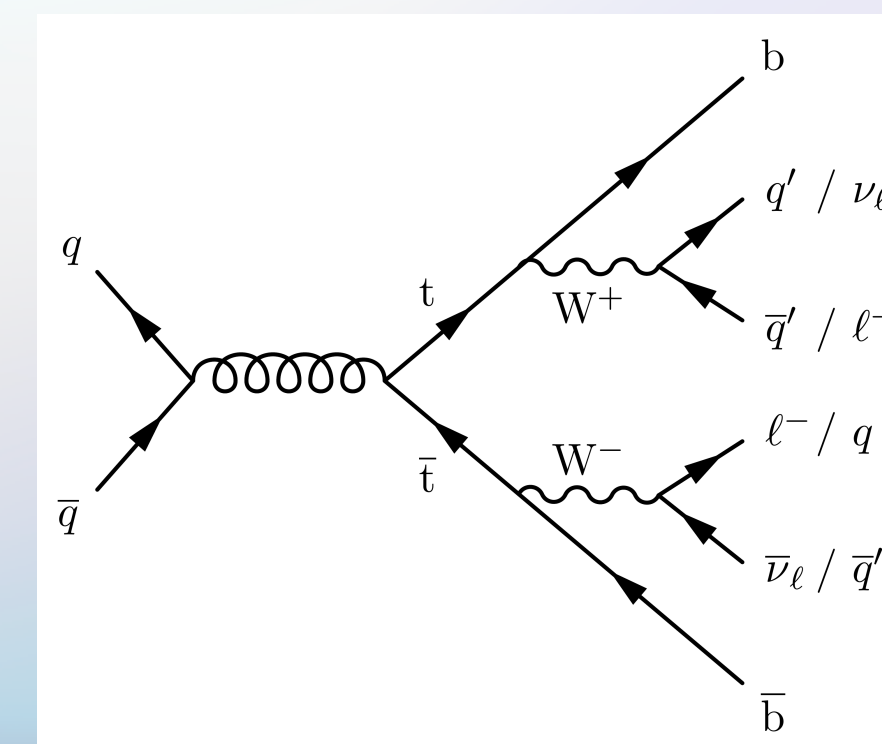
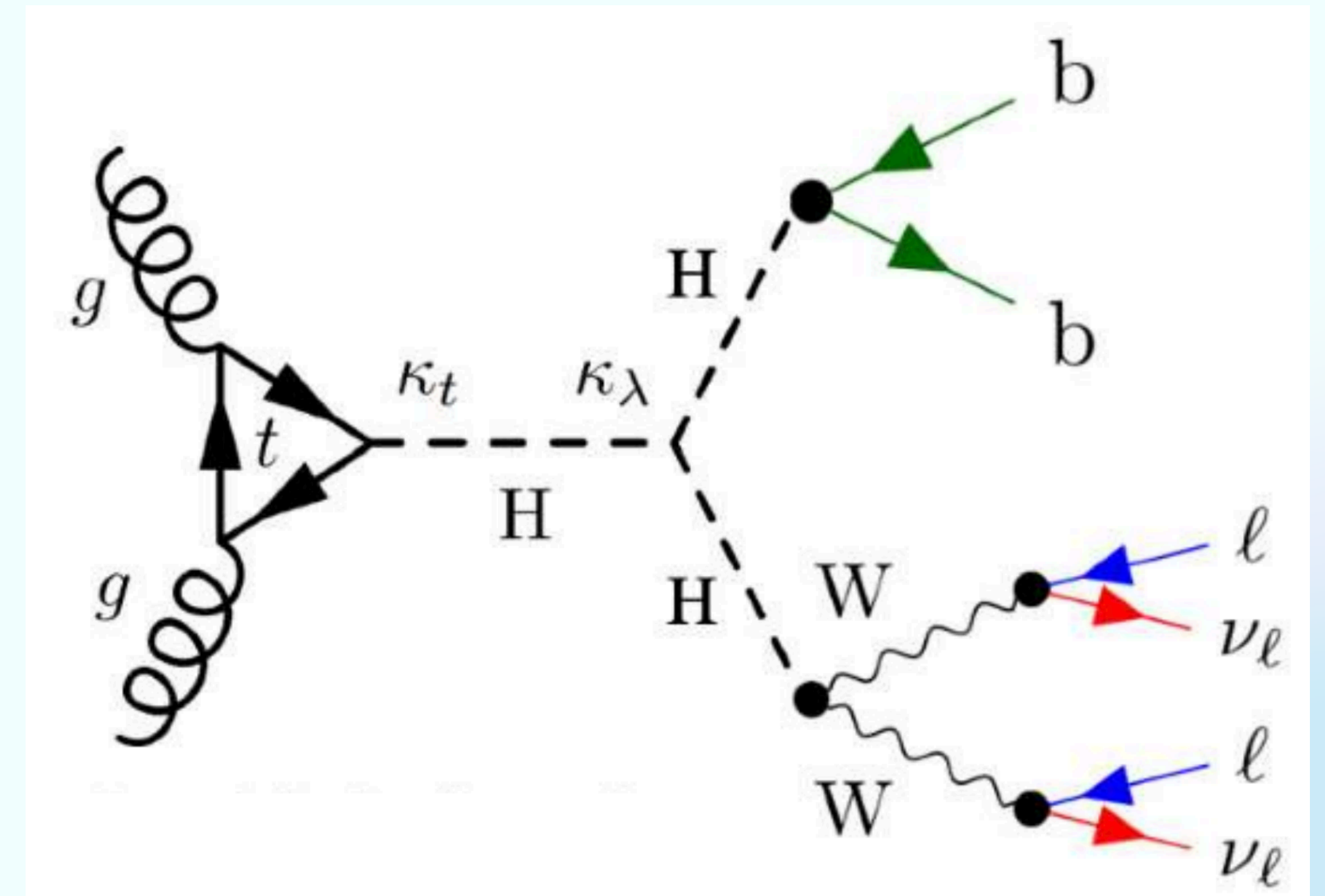
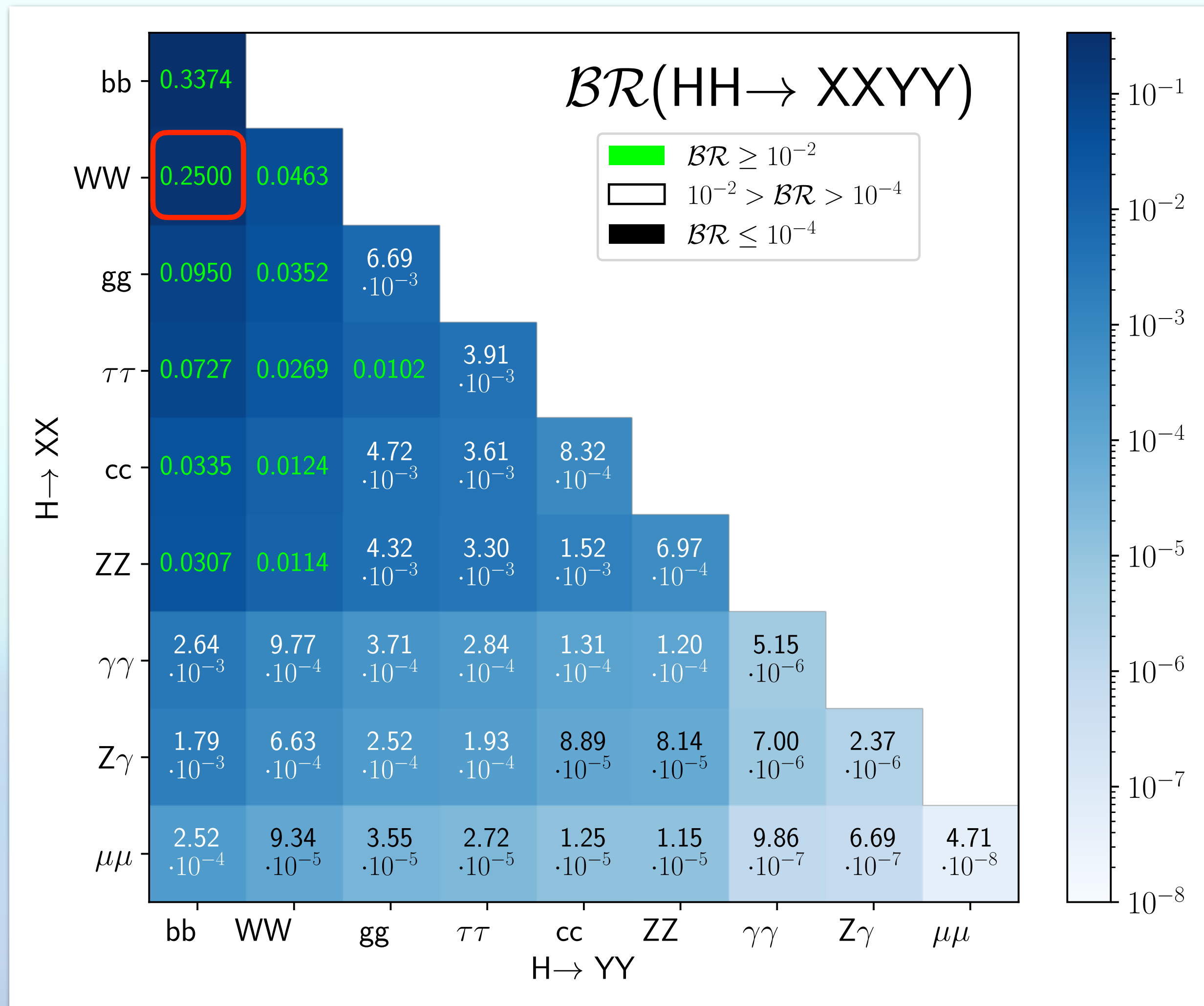
Model	Total Time	Epochs*	Time / Epoch	Early Stopping
Transformer	90 min	4	22.5 min	No
DNN	28 min	16	106 s	Yes
BDT	45 s	200	0.2 s	No

Data Analysis

Data Analysis

$HH \rightarrow bbWW$

$$BR(HH \rightarrow bbWW \rightarrow bbl\nu l\nu) = 2.64 \%$$



Data Analysis

Data, MC samples and trigger

- 62.4 fb^{-1} collected data in 2022 and 2023
- Muon, MuonEG and EGamma datasets
- Single + di-lepton triggers
- MC samples for major bkg ($t\bar{t}$, DY+jets, VV, ST, V+jets)
- 4 ggF signal sample with different κ_λ
- 9 VBF signal sample with different κ_V , κ_{2V} , κ_λ for coupling scans

Data Analysis

Object definitions

Muons

Loose

- $p_T > 5 \text{ GeV}$
- $|\eta| < 2.4$
- $|d_{xy}| < 0.05 \text{ cm}$
- $|d_z| < 0.1 \text{ cm}$
- $d/\sigma_d \text{ (SIP}_{3D}) < 8$
- $I_\mu < 0.4 \times p_T^\mu$
- loose POG WP

Fakeable

- $p_T > 10 \text{ GeV}$
- pNet of nearby jet < medium WP
- jet relative ISO* < 0.8 (none)[†] (corrected in manuscript)

Tight

- medium WP
- pNet of nearby jet < medium

Electrons

Loose

- $p_T > 5 \text{ GeV}$
- $|\eta| < 2.5$
- $|d_{xy}| < 0.05 \text{ cm}$
- $|d_z| < 0.1 \text{ cm}$
- $d/\sigma_d \text{ (SIP}_{3D}) < 8$
- $I_e < 0.4 \times p_T^e$
- missing hits ≤ 1
- loose POG WP
- cleaned against loose muons by $\Delta R \geq 0.3$

Fakeable

- $p_T > 10 \text{ GeV}$
- $\sigma_{i\eta i\eta} < 0.011/0.030^1$
- H/E < 0.1
- $1/E - 1/p > -0.04$
- conversion rejection
- missing hits = 0
- tight WP no ISO
- pNet of nearby jet < tight/medium³
- jet relative ISO < 0.7 (none)[†]

Tight

- tight WP with ISO
- pNet of nearby jet < medium
- electron MVA > 0.3

AK4 jets

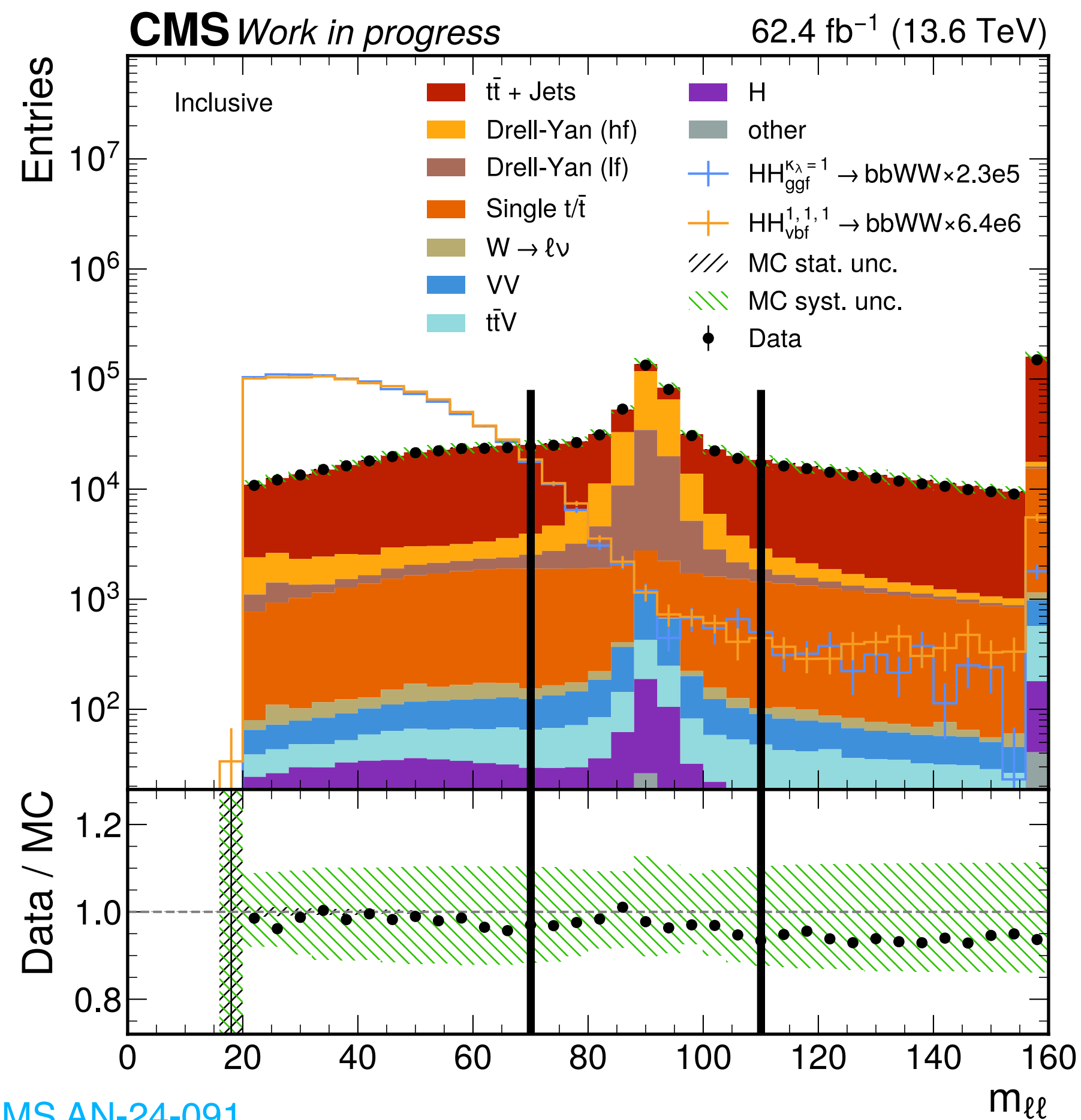
- $p_T > 25 \text{ GeV}$
- $|\eta| < 2.4$
- tight ID
- cleaned against fakeable leptons by $\Delta R \geq 0.4$
- AK4 b-jets with medium PNet WP (80%/1%)

AK8 jets

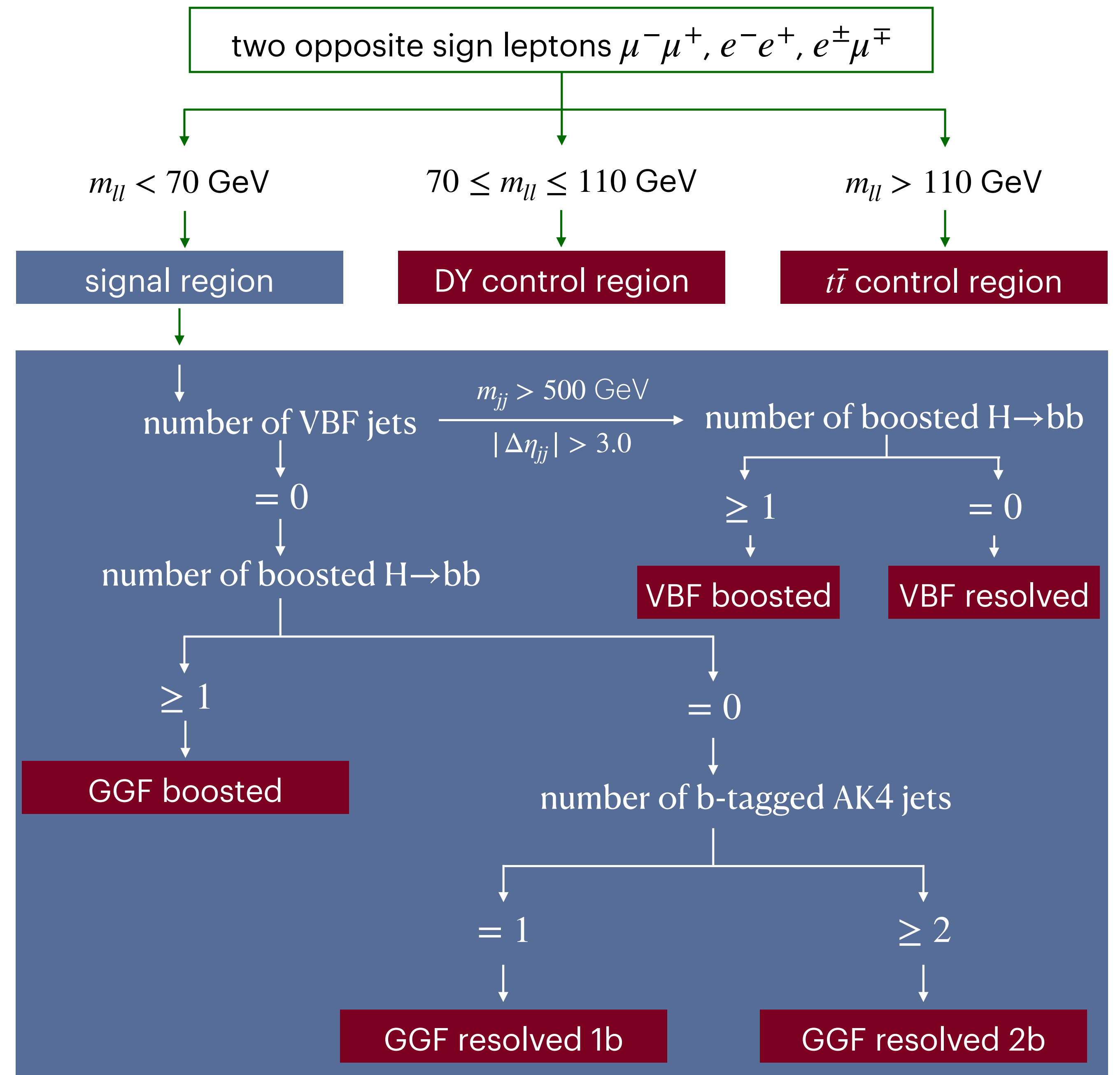
- $p_T > 200 \text{ GeV}$
- $|\eta| < 2.4$
- tight ID
- at least 2 subjets with $p_T \geq 20 \text{ GeV}$, $|\eta| \leq 2.4$
- $30 \leq m_{SD} \leq 210 \text{ GeV}$, $\tau_2/\tau_1 \leq 0.75$
- cleaned against fakeable leptons by $\Delta R \geq 0.4$
- A8 b-jets with
- $particleNetWithMass_HbbvsQCD > 0.92$, 26%(22%)/2% ggF(VBF)

Data Analysis

Event selection

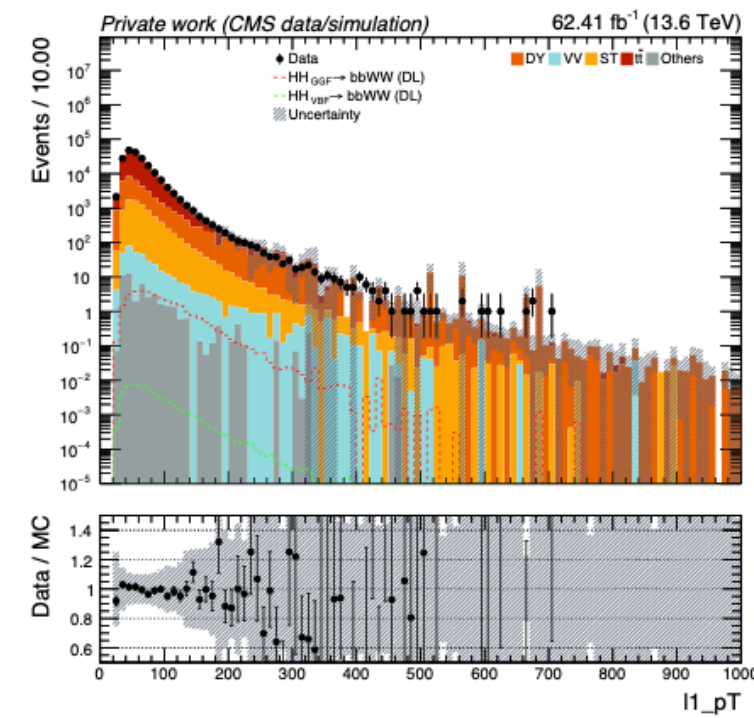


[CMS AN-24-091](#)

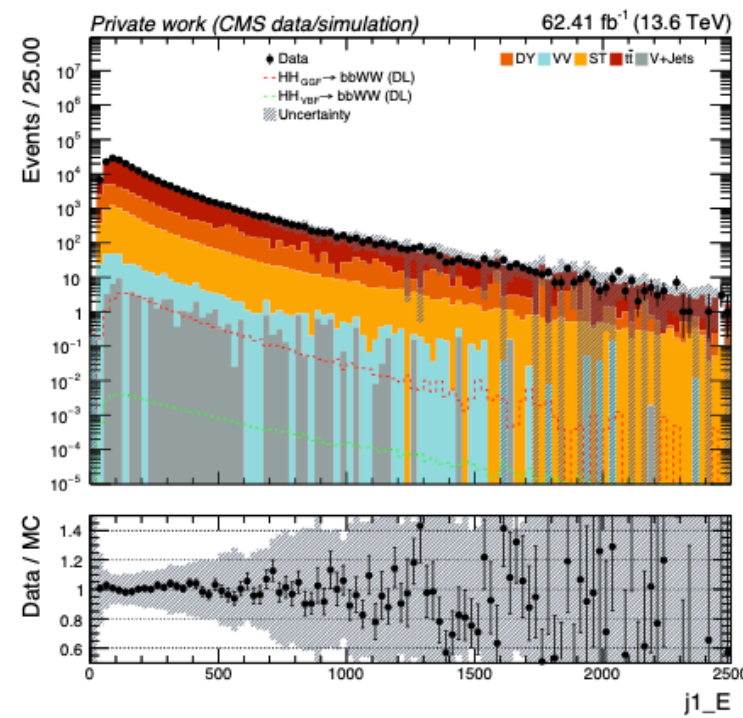


Data Analysis

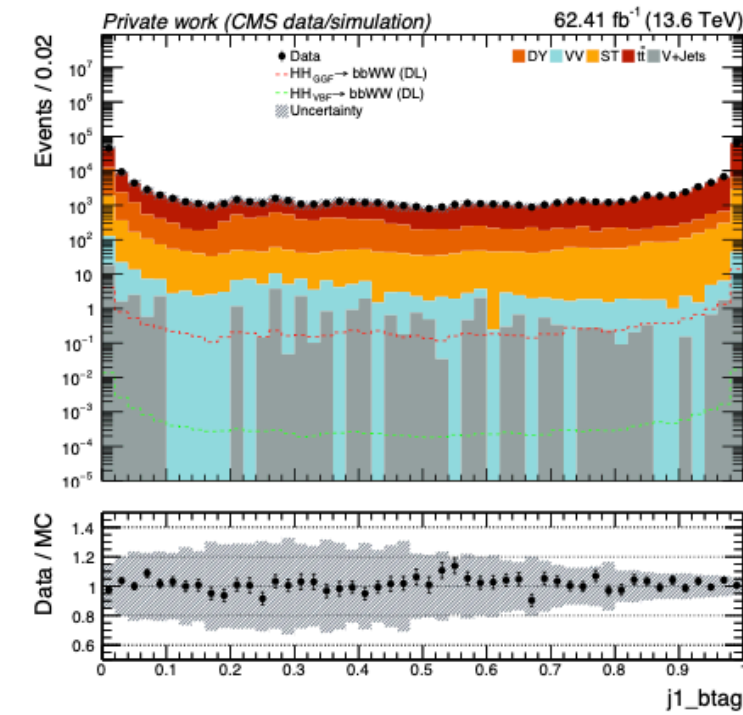
Data/MC match & BDT inputs



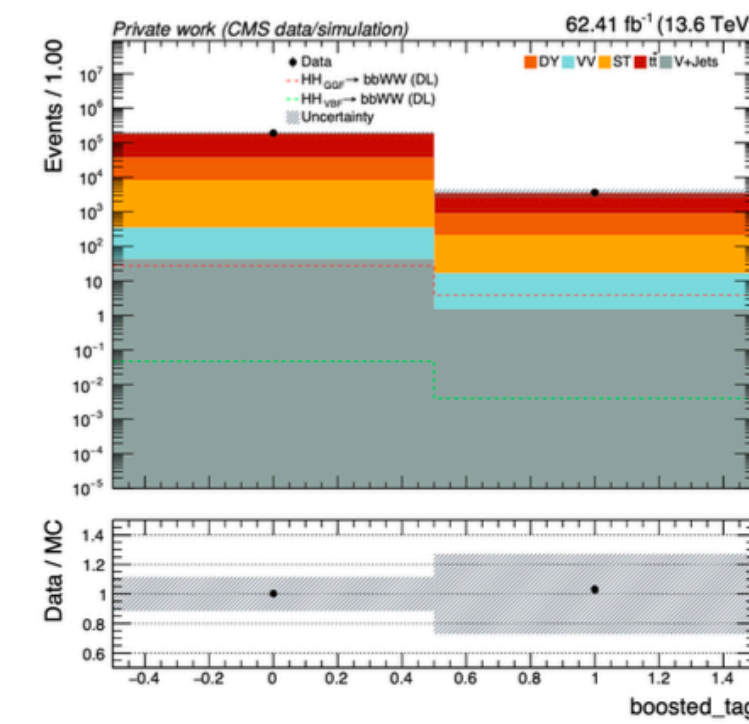
(A) p_T of the leading lepton



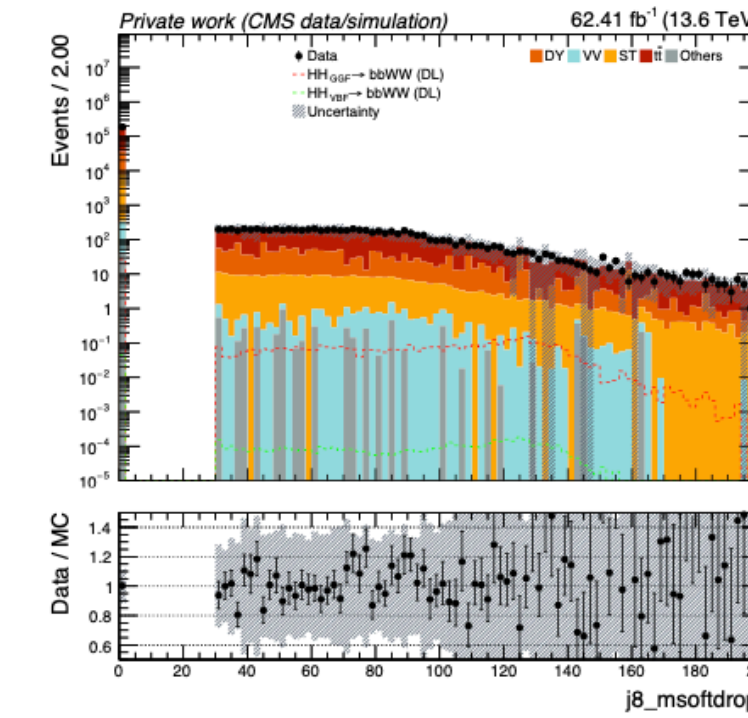
(B) Energy of the leading jet



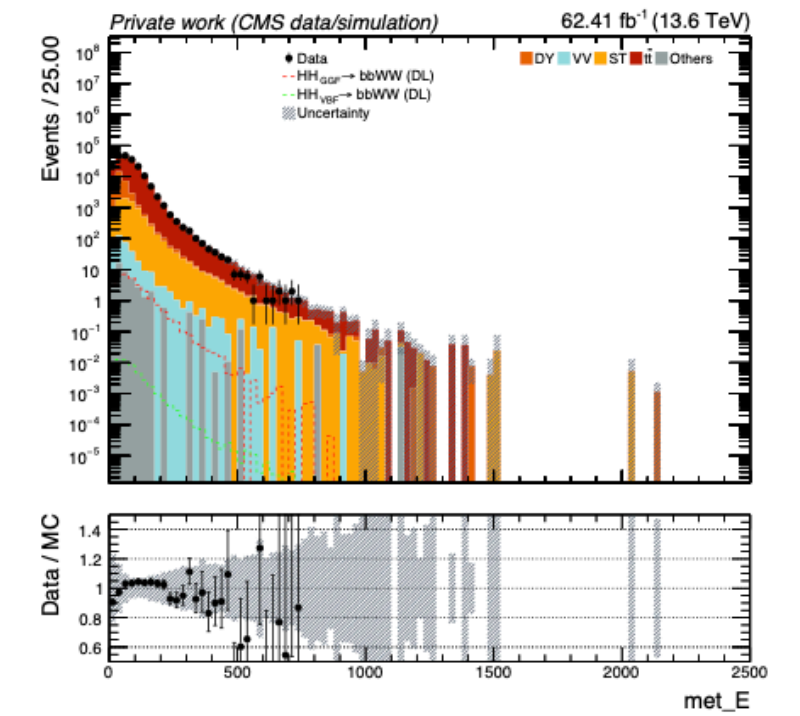
(C) b-tagging score of the leading jet



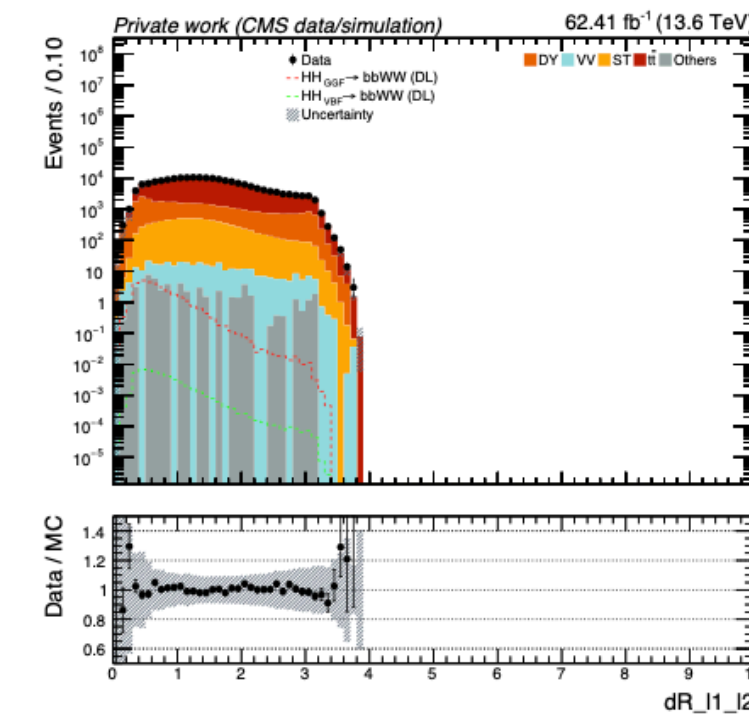
(D) Boosted category flag



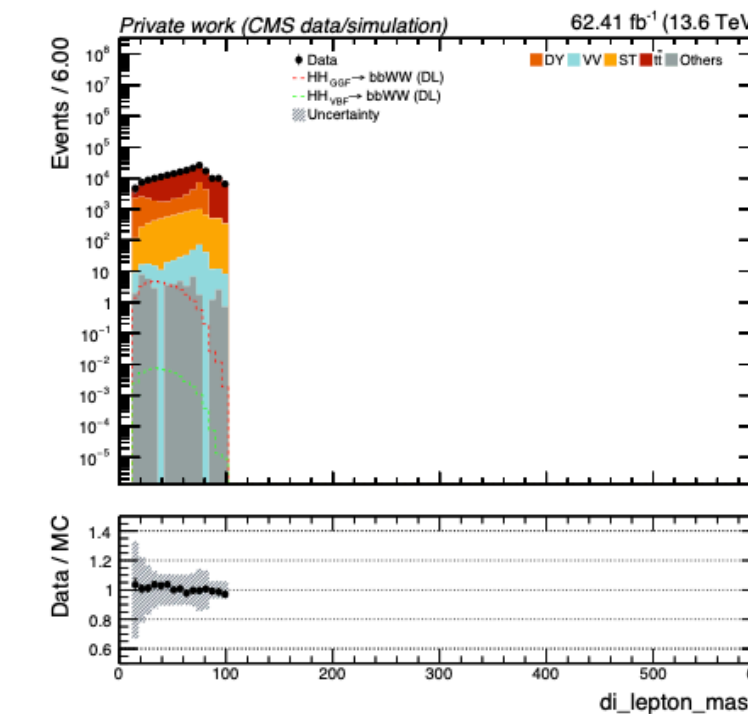
(E) Soft drop mass of the leading AK8 jet



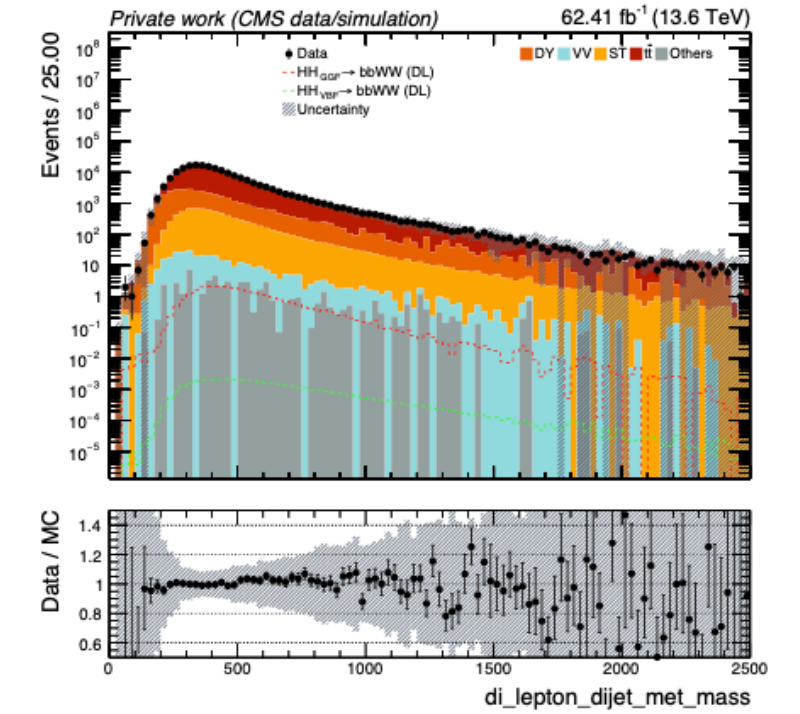
(F) E_T^{miss}



(G) ΔR between the two leptons



(H) Mass of the dilepton system



(I) Mass of the dilepton + di-jet + E_T^{miss} system

Data Analysis

Corrections to the data and the simulation

Scalefactors

- despite the high precision in generators/detector modeling, MC can still not excel
- various sources
- typically derived by comparing the performance of physics objects

$$SF = \frac{\epsilon_{data}}{\epsilon_{MC}}$$

- data taking issues
 - EE in 2022, BPix in 2023
 - separate MC campaigns
 - jet veto maps
- pileup reweighting
- trigger efficiency
- noise filters to reject anomalous events
- top quark p_T reweighting
- btagging SFs and renormalization of weights
- SF for AK8 b-tagged jet pair tagging
- lepton corrections for ID/ISO/resolution
- Z p_T reweighting
- JEC/JER

Data Analysis

Uncertainties

Experimental

- Luminosity
- Pileup
- Trigger
- Lepton ID/ISO
- Top quark p_T reweighting
- JES/JER
- b-jet tagging
 - JEC/JER
 - flavor purity
 - statistical and c-quark jet uncertainties
 - correlations
- boosted bb-jet tagging

Theory

- Inclusive cross sections, PDF
- Matrix-element scale variations
- Unconstrained normalizations
- PDF-induced shape variations
- Parton shower scales (ISR/FSR)
- MC statistics

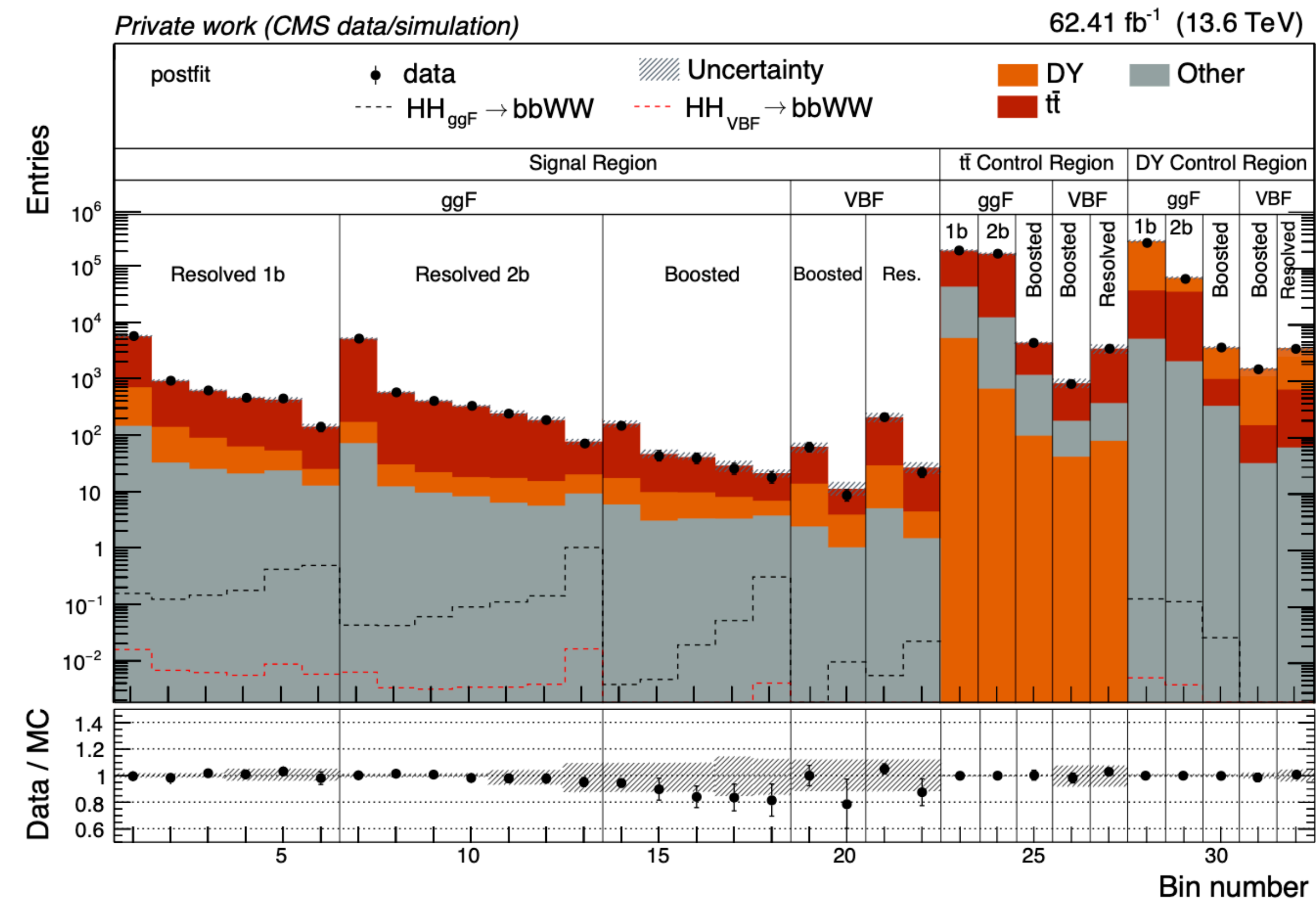
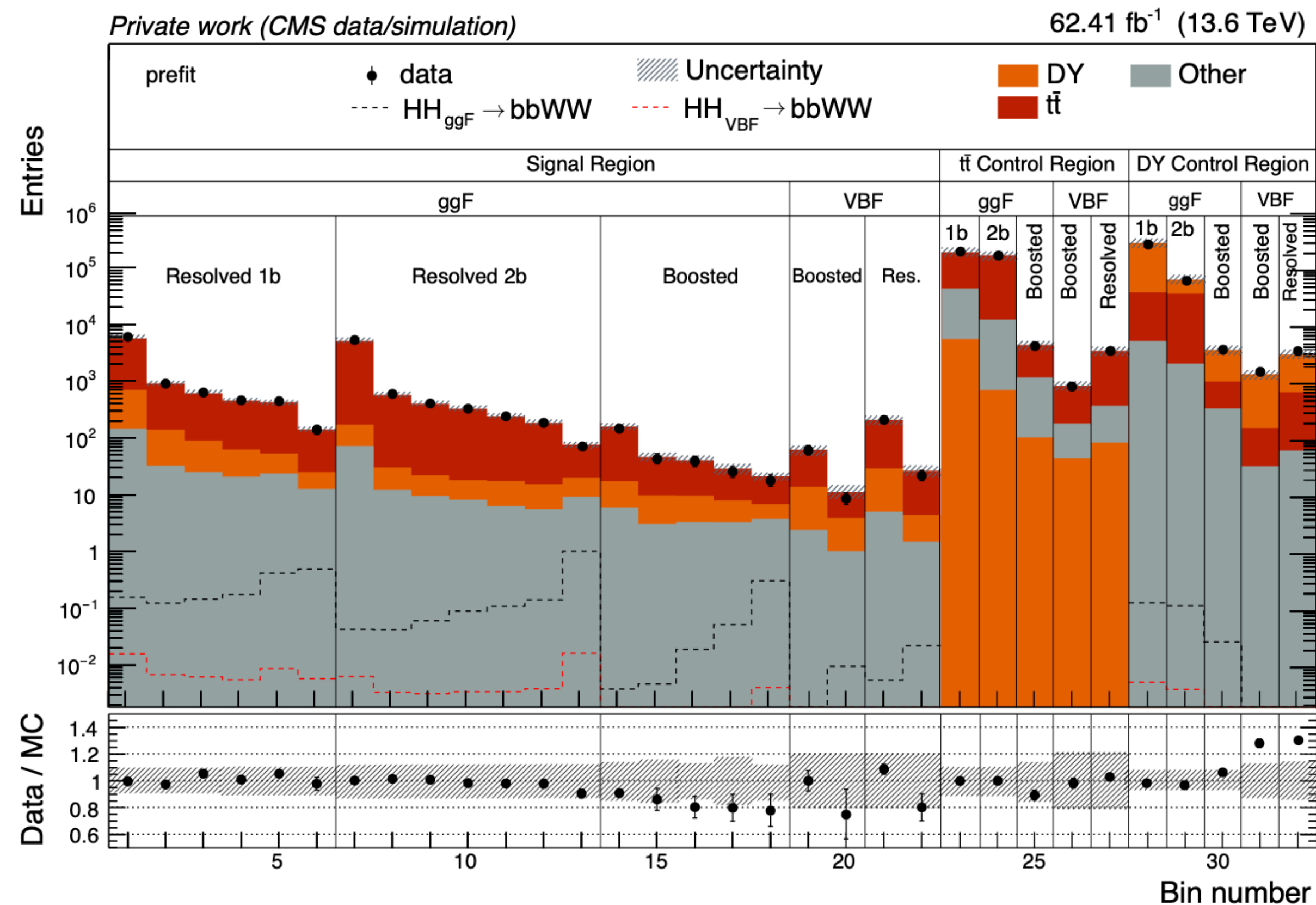
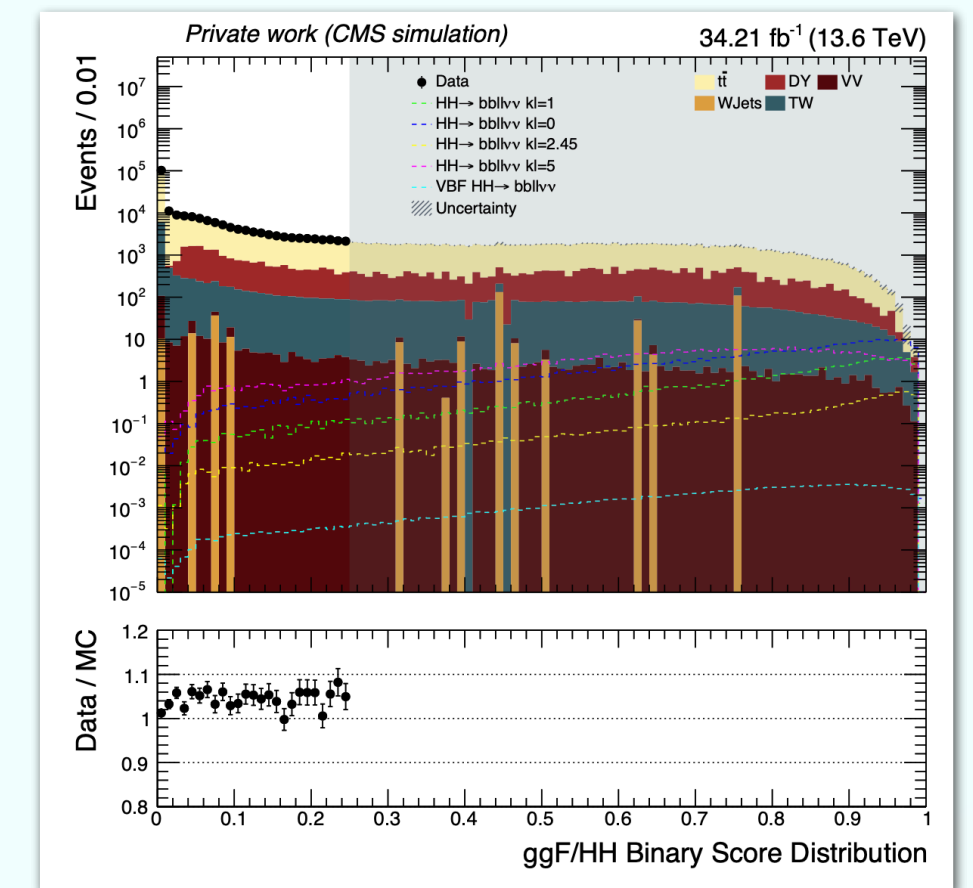
NPs in the fit

Data Analysis

Binning strategy

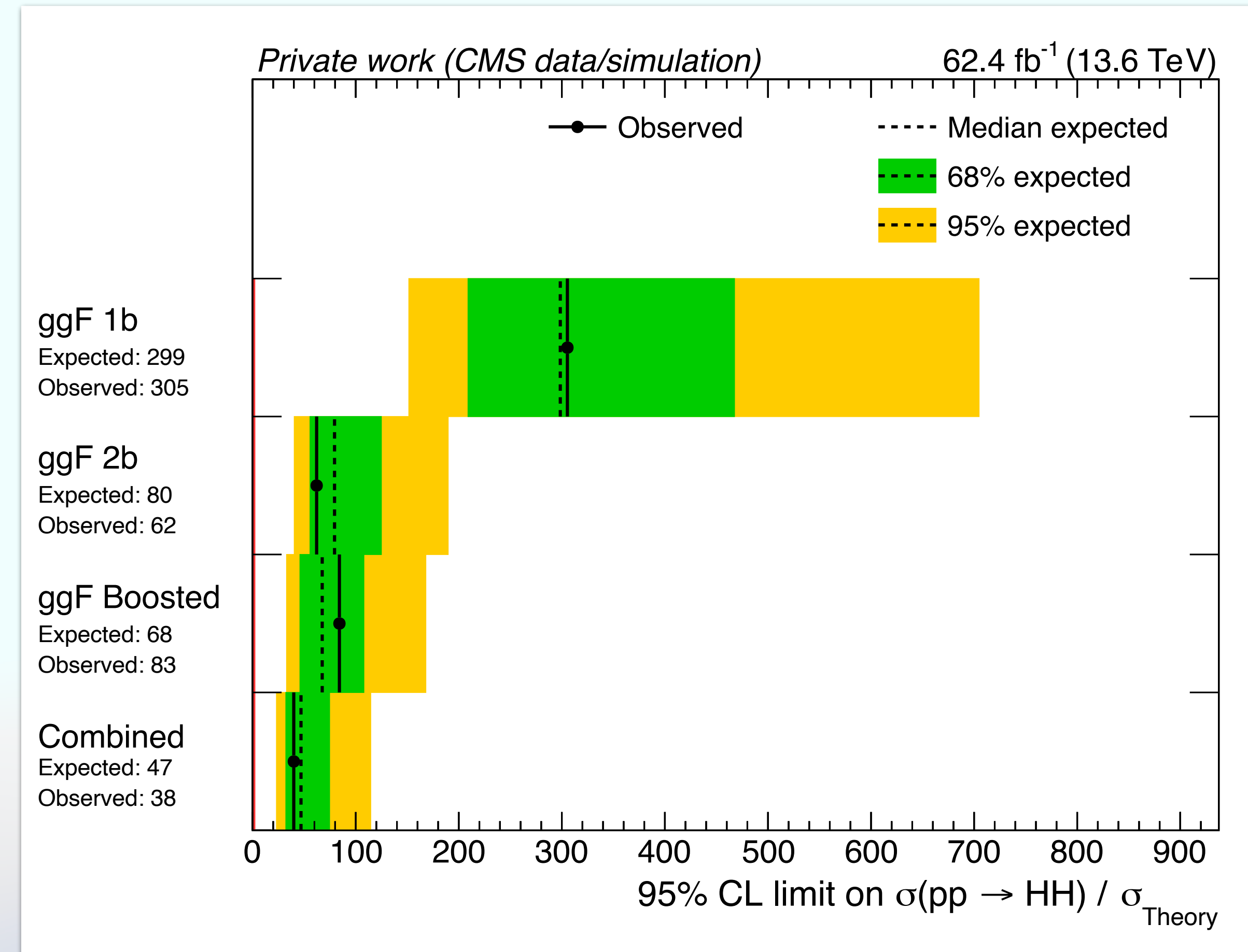
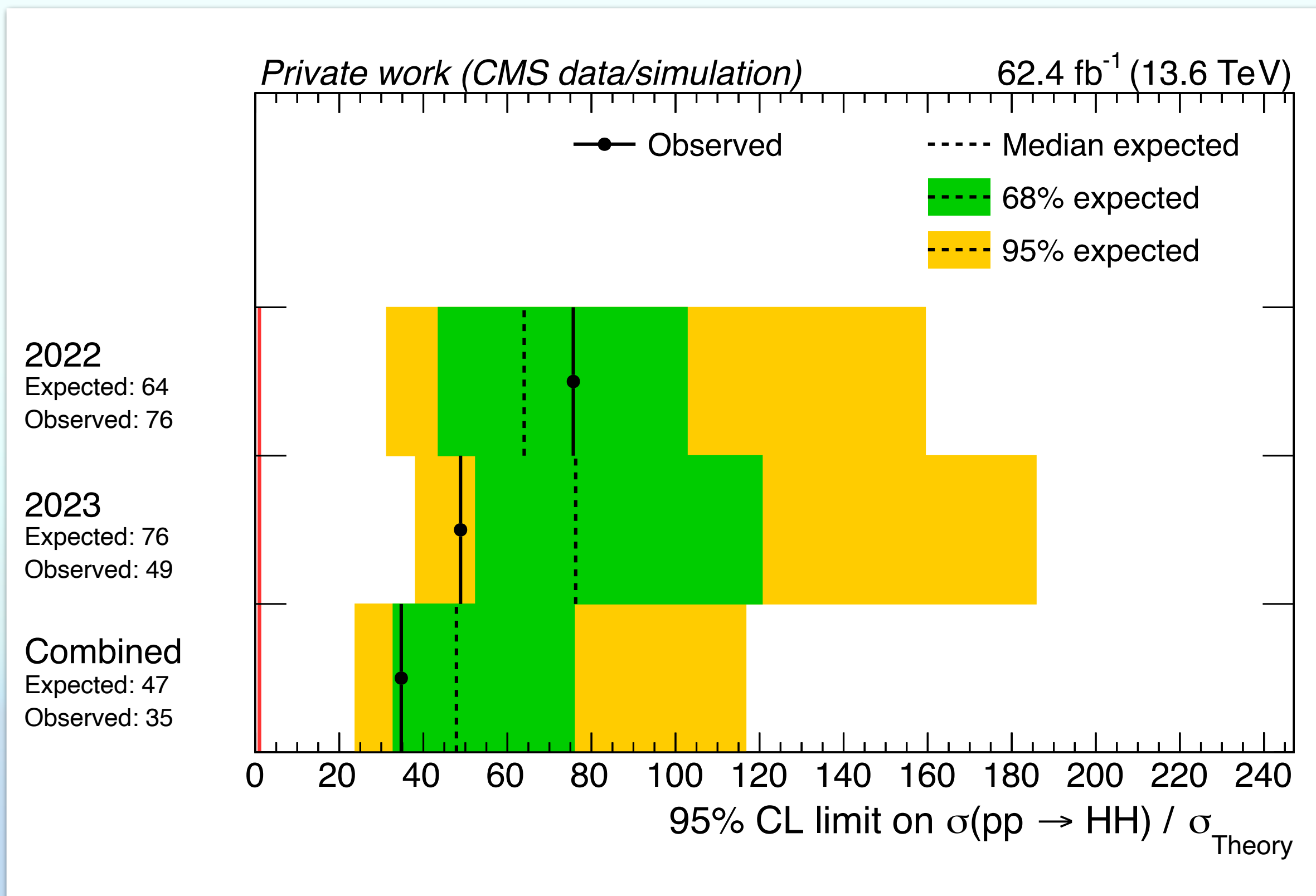
for signal, from right to left

- at least 10 MC events
 - max 20-30% uncertainty
 - coarser bins toward left
- single bin for CRs



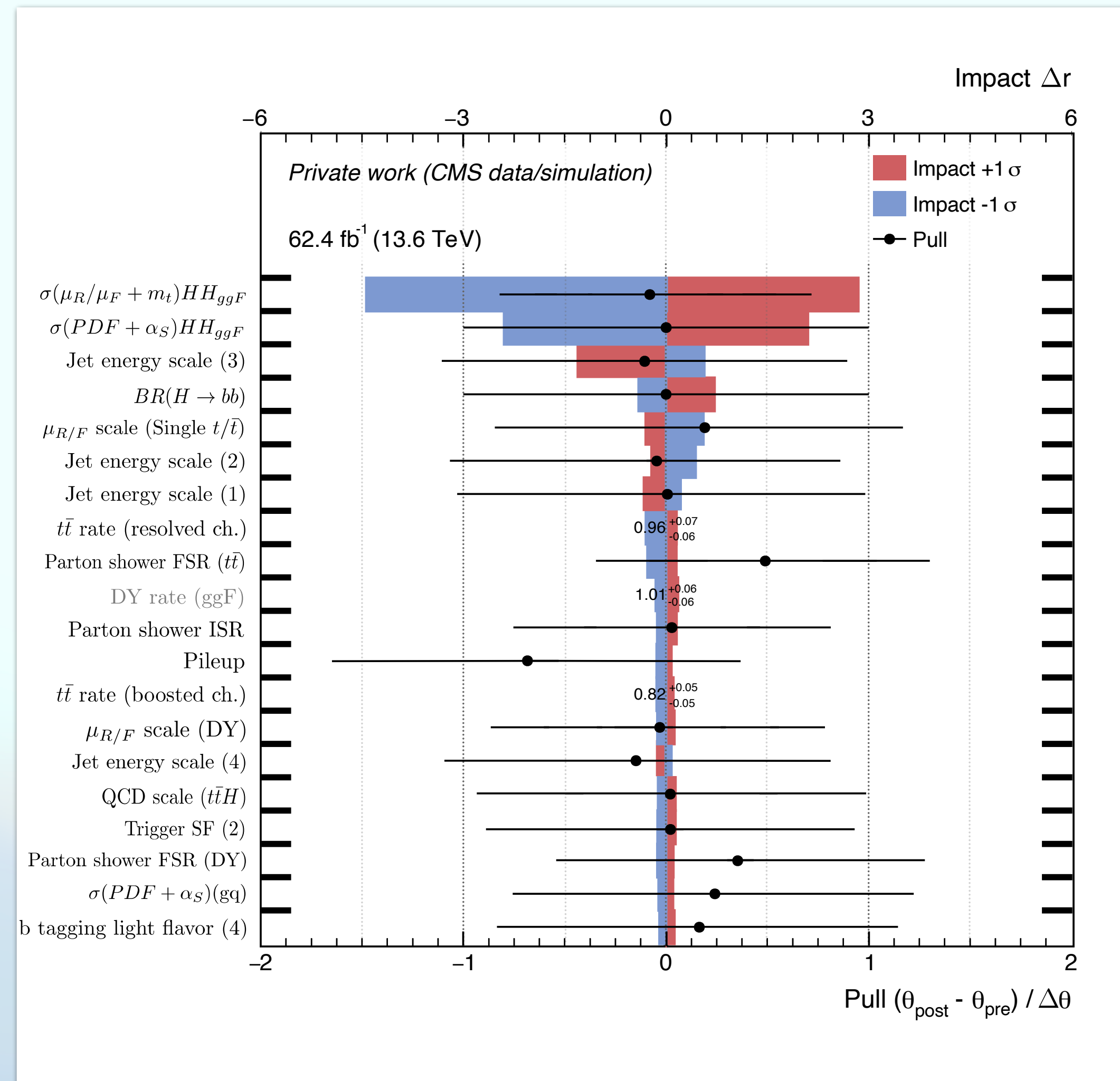
Results

Upper limits



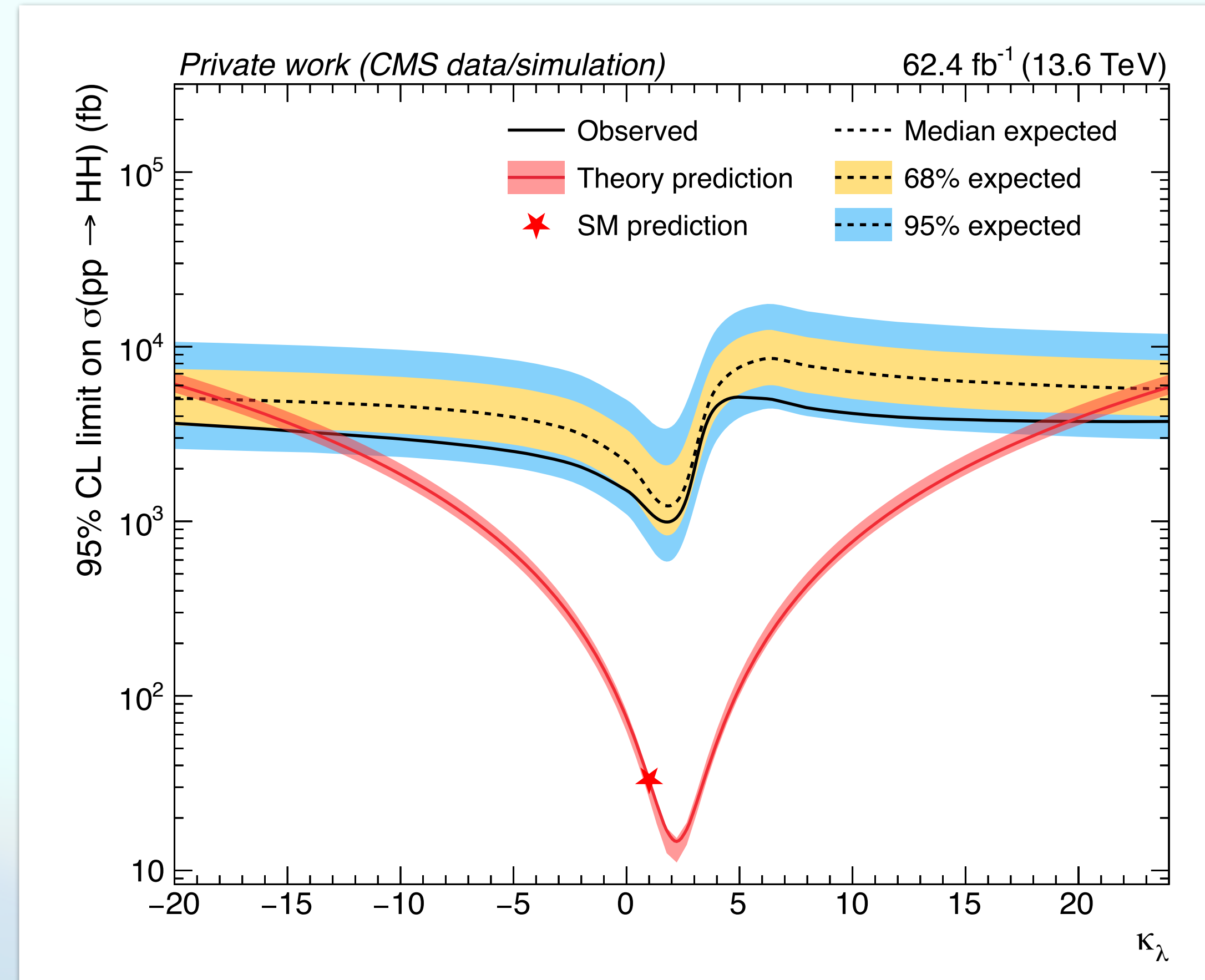
Results

Impacts of systematic uncertainties



Results

Coupling scan



[-13.3, 19.0] observed, [-20.0, 23.7] expected

Conclusion and Outlook

Compared to Run 2 analysis (27):

$$\bullet \sqrt{\frac{62.4}{138}} \times 35 \approx 23.5$$

Run 3 analysis:

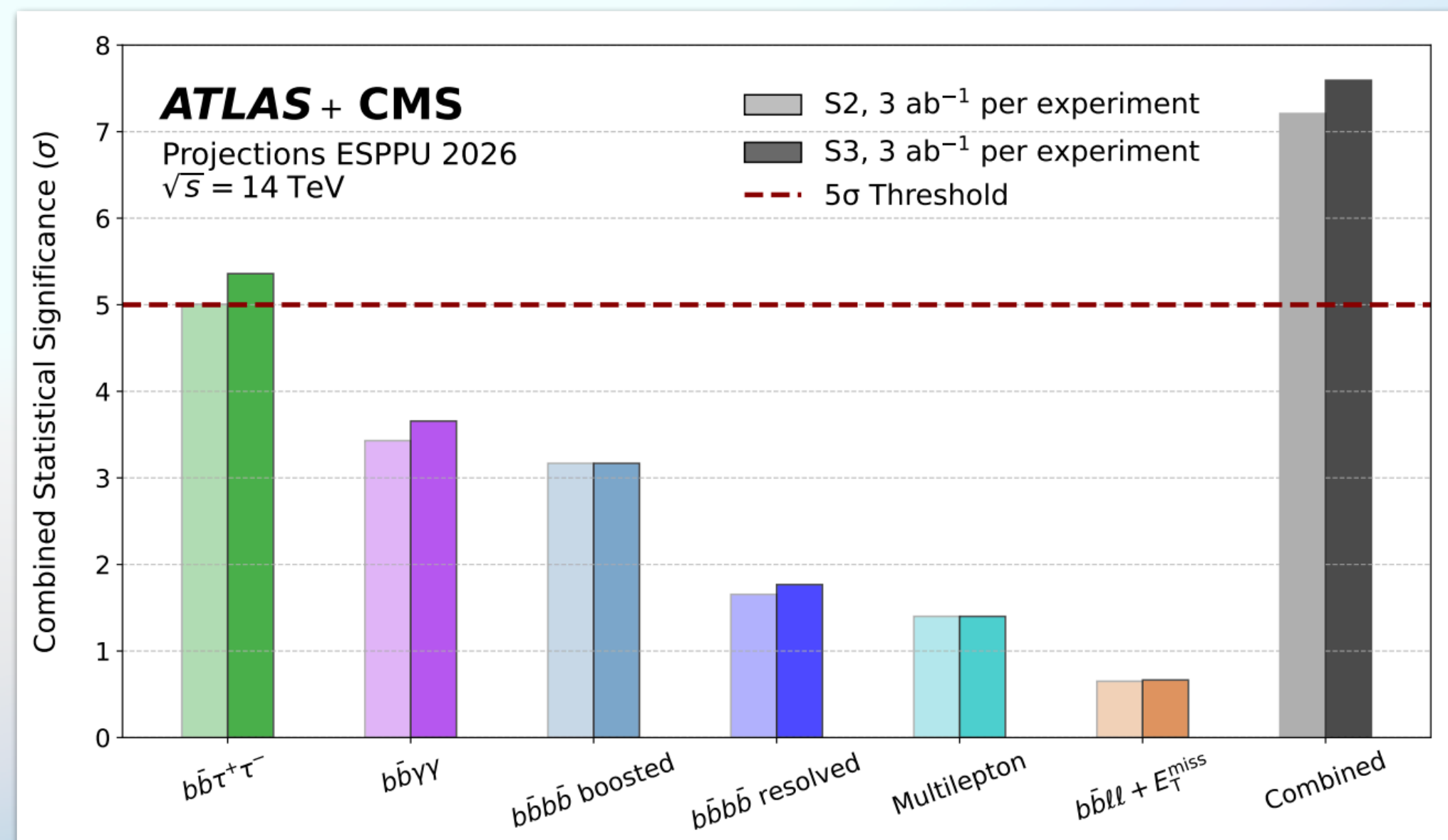
- 12.7 (18.6) observed (expected)
- Two-stage ML architecture
- Multi-category simultaneous fit

Run 3 delivered more than 300 fb^{-1}

- some issues
- better reconstruction & calibration
- better analysis techniques involving ML

HL-LHC 3000 fb^{-1}

- global combination



Thank you !

Backup

Backup

Scalar sector

$$\mathcal{L}_{\text{scalar}} = D_\mu \phi^\dagger D^\mu \phi - V(\phi^\dagger \phi) \text{ with } \phi \text{ doublet under } SU(2)$$

$$V(\phi^\dagger \phi) = -\mu^2 (\phi^\dagger \phi) + \lambda (\phi^\dagger \phi)^2$$

$$\text{non-zero VEV } v^2 = \mu^2 / \lambda = 246 \text{ GeV}$$

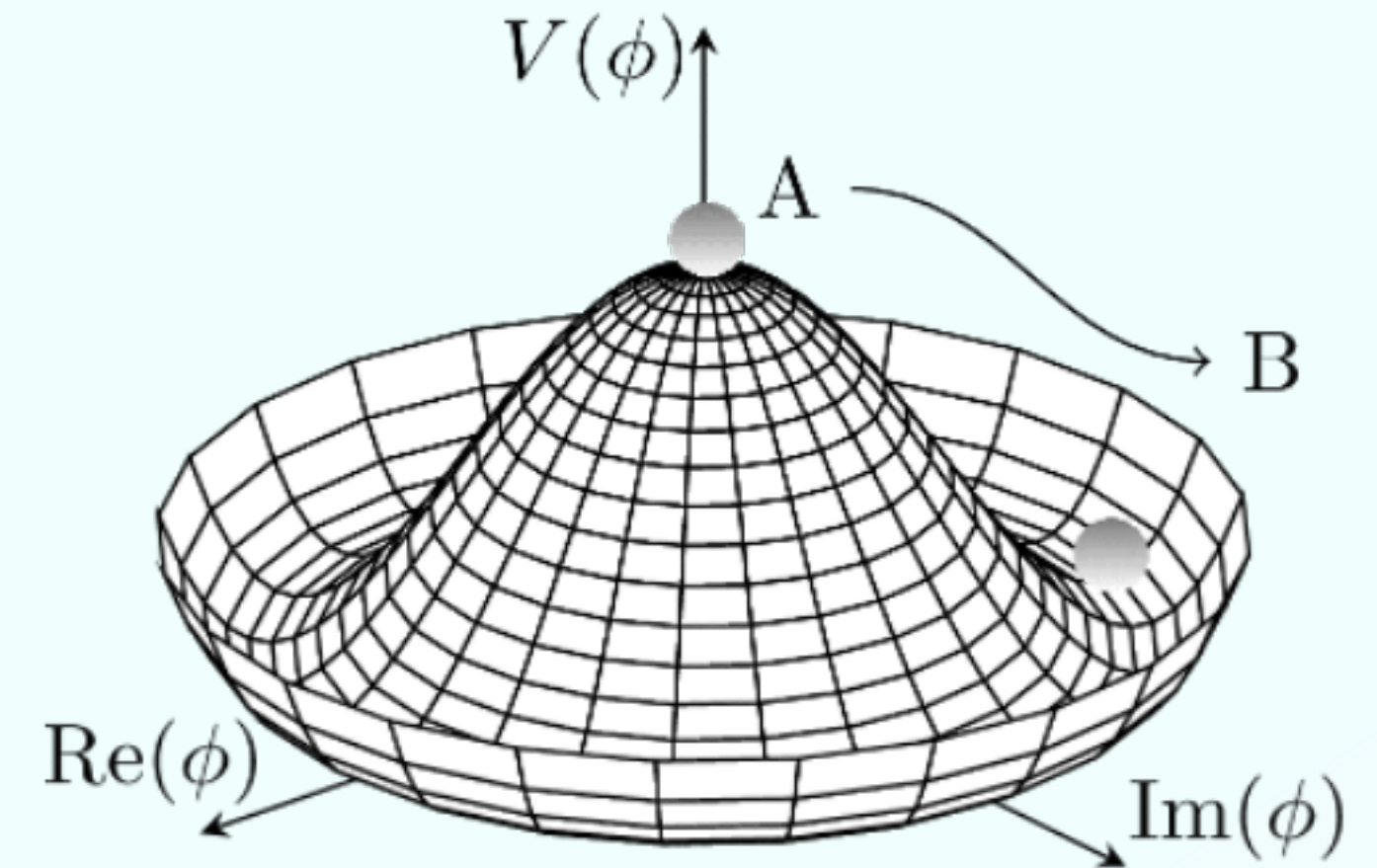
$$\phi(x) = \frac{1}{\sqrt{2}} \exp(i\sigma^i \xi_i(x)) \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} \xrightarrow{\text{EWSB}} \phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}$$

$$\mathcal{L}_{\text{scalar}} = D_\mu \phi^\dagger D^\mu \phi + \mu^2 (\phi^\dagger \phi) - \lambda (\phi^\dagger \phi)^2$$

$$= \frac{v^2}{8} \left(2W_\mu^i W^{i\mu} + g'^2 B_\mu B^\mu - 2g'g B_\mu W^{3\mu} \right) \left(1 + \frac{h}{v} \right)^2$$

$$+ \underbrace{\frac{1}{2} \left(\partial_\mu h \partial^\mu h \right)}_{\text{kinetic term}} - \underbrace{\frac{\lambda v^2 h^2}{2}}_{\text{mass term}} - \underbrace{\frac{\lambda v h^3}{3}}_{\text{trilinear coupling}} - \underbrace{\frac{\lambda}{4} h^4}_{\text{quartic coupling}} - \frac{\lambda v^4}{4}$$

$$m_H = \sqrt{2\lambda} v = 125 \text{ GeV}$$



$$W_\mu^\pm = \frac{1}{\sqrt{2}} \left(W_\mu^1 \pm iW_\mu^2 \right)$$

$$\begin{pmatrix} Z_\mu \\ A_\mu \end{pmatrix} = \begin{pmatrix} \cos \theta_W & -\sin \theta_W \\ \sin \theta_W & \cos \theta_W \end{pmatrix} \begin{pmatrix} W_\mu^3 \\ B_\mu \end{pmatrix}$$

$$\left(m_W^2 W_\mu^+ W^{-\mu} + \frac{1}{2} m_Z^2 Z_\mu Z^\mu \right) \left(1 + \frac{h}{v} \right)^2$$

$$m_W = \frac{gv}{2} = 80 \text{ GeV}$$

$$m_Z = \sqrt{g^2 + g'^2} \frac{v}{2} = 91 \text{ GeV}$$

$$\mathcal{L}_{\text{Yukawa}} = - \sum_i Y_l^{ij} (\bar{L}^i \cdot \phi) l^j + h.c.$$

$$- \frac{Y_l v}{\sqrt{2}} (\bar{l}_L l_R + \bar{l}_R l_L) \left(1 + \frac{h}{v} \right)$$

+ quarks (CKM)

Backup

Higgs self-coupling

$$\mathcal{L}_{\text{scalar}} = D_\mu \phi^\dagger D^\mu \phi - V(\phi^\dagger \phi) \text{ with } \phi \text{ doublet under } SU(2)$$

$$V(\phi^\dagger \phi) = -\mu^2 (\phi^\dagger \phi) + \lambda (\phi^\dagger \phi)^2$$

$$\phi(x) = \frac{1}{\sqrt{2}} \exp(i\sigma^i \xi_i(x)) \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} \xrightarrow{EWSB} \phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}$$

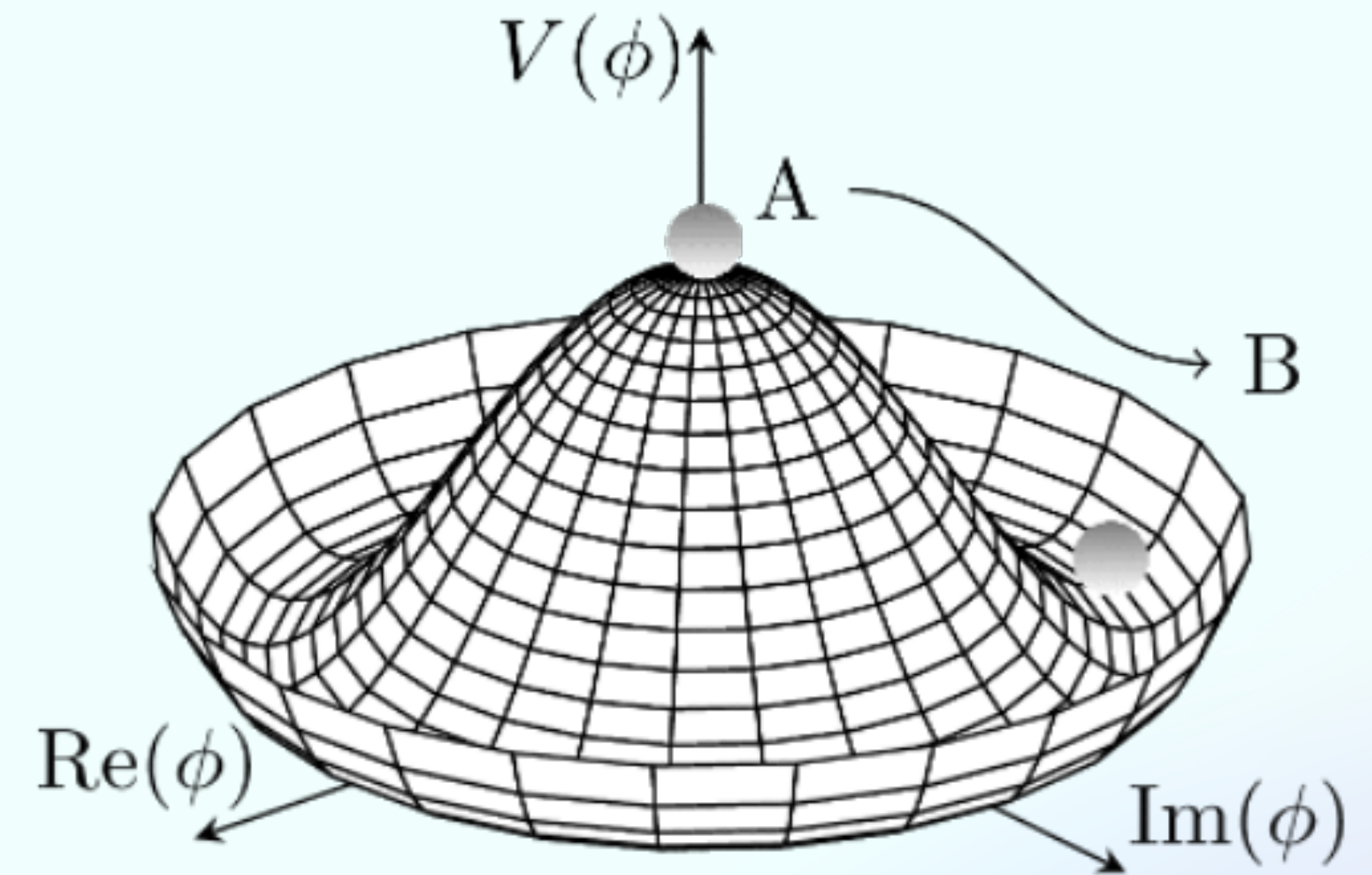
$$\text{non-zero VEV } v^2 = \mu^2 / \lambda = 246 \text{ GeV}$$

$$\mathcal{L}_{\text{scalar}} = D_\mu \phi^\dagger D^\mu \phi + \mu^2 (\phi^\dagger \phi) - \lambda (\phi^\dagger \phi)^2$$

$$= \frac{v^2}{8} \left(2W_\mu^i W^{i\mu} + g'^2 B_\mu B^\mu - 2g'g B_\mu W^{3\mu} \right) \left(1 + \frac{h}{v} \right)^2$$

$$\underbrace{+\frac{1}{2} \left(\partial_\mu h \partial^\mu h \right)}_{\text{kinetic term}} - \underbrace{\frac{\lambda v^2 h^2}{2}}_{\text{mass term}} - \underbrace{\frac{\lambda v h^3}{3}}_{\text{trilinear coupling}} - \underbrace{\frac{\lambda}{4} h^4}_{\text{quartic coupling}} - \frac{\lambda v^4}{4}$$

$$m_H = \sqrt{2\lambda}v = 125 \text{ GeV}$$



Core objective: Reconstruct shape of the Higgs potential

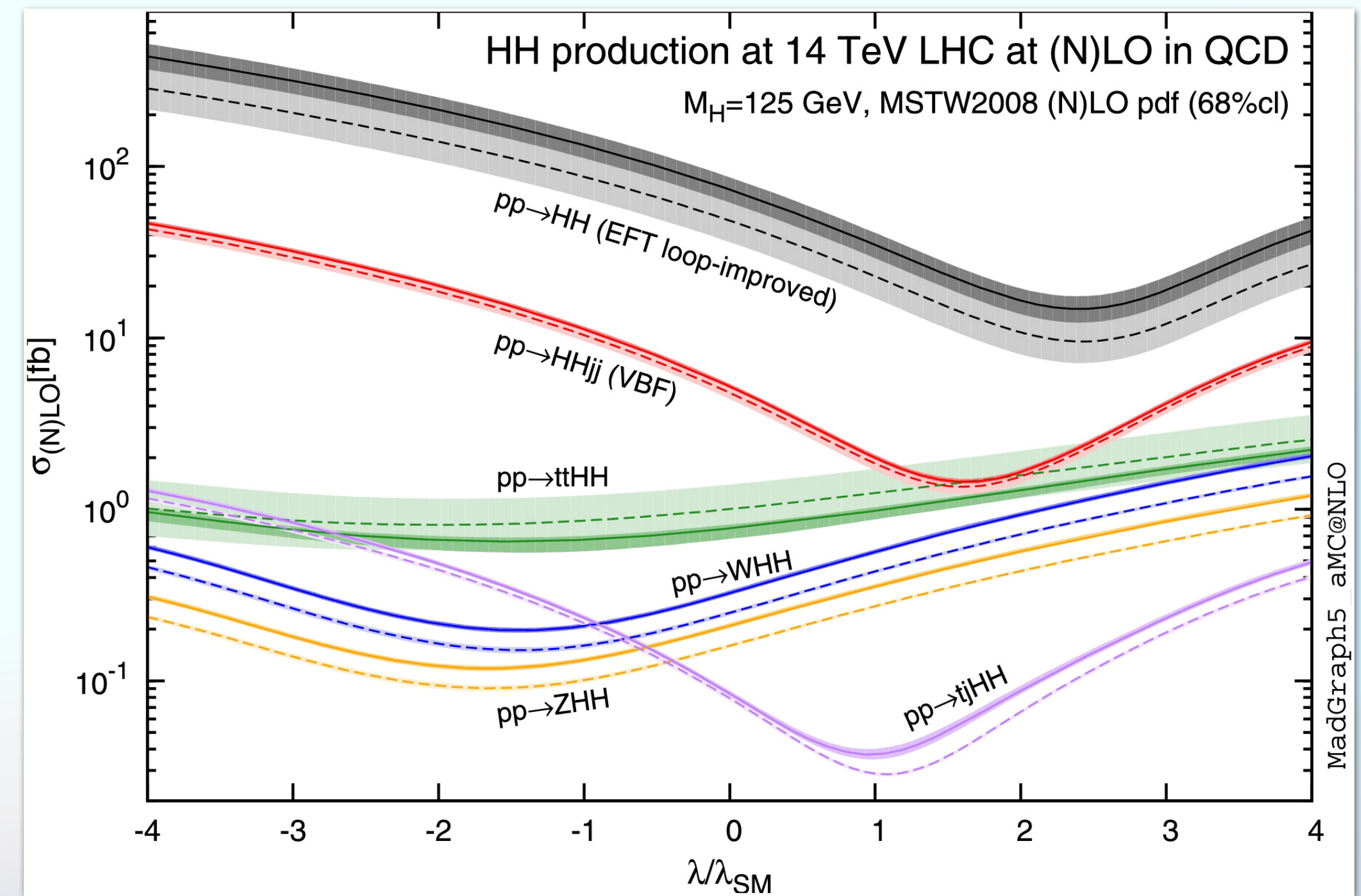
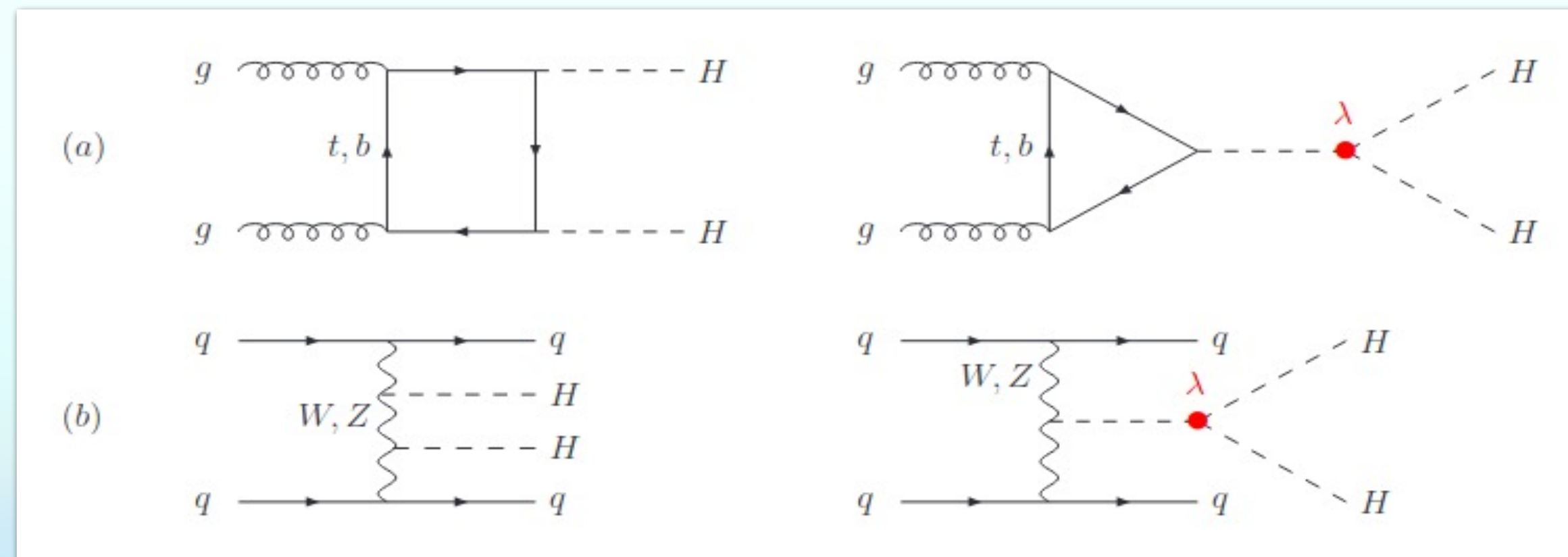
Measuring di-Higgs production

- test the SM by probing the shape
- hints to the BSM (require EFT)

Backup

Production modes and cross section

$$\frac{d\sigma}{dm_{HH}} \propto |\mathcal{A}_{\square} - \kappa_{\lambda} \mathcal{A}_{\triangle}|^2$$



Backup

ggF HH box&triangular diagram interference

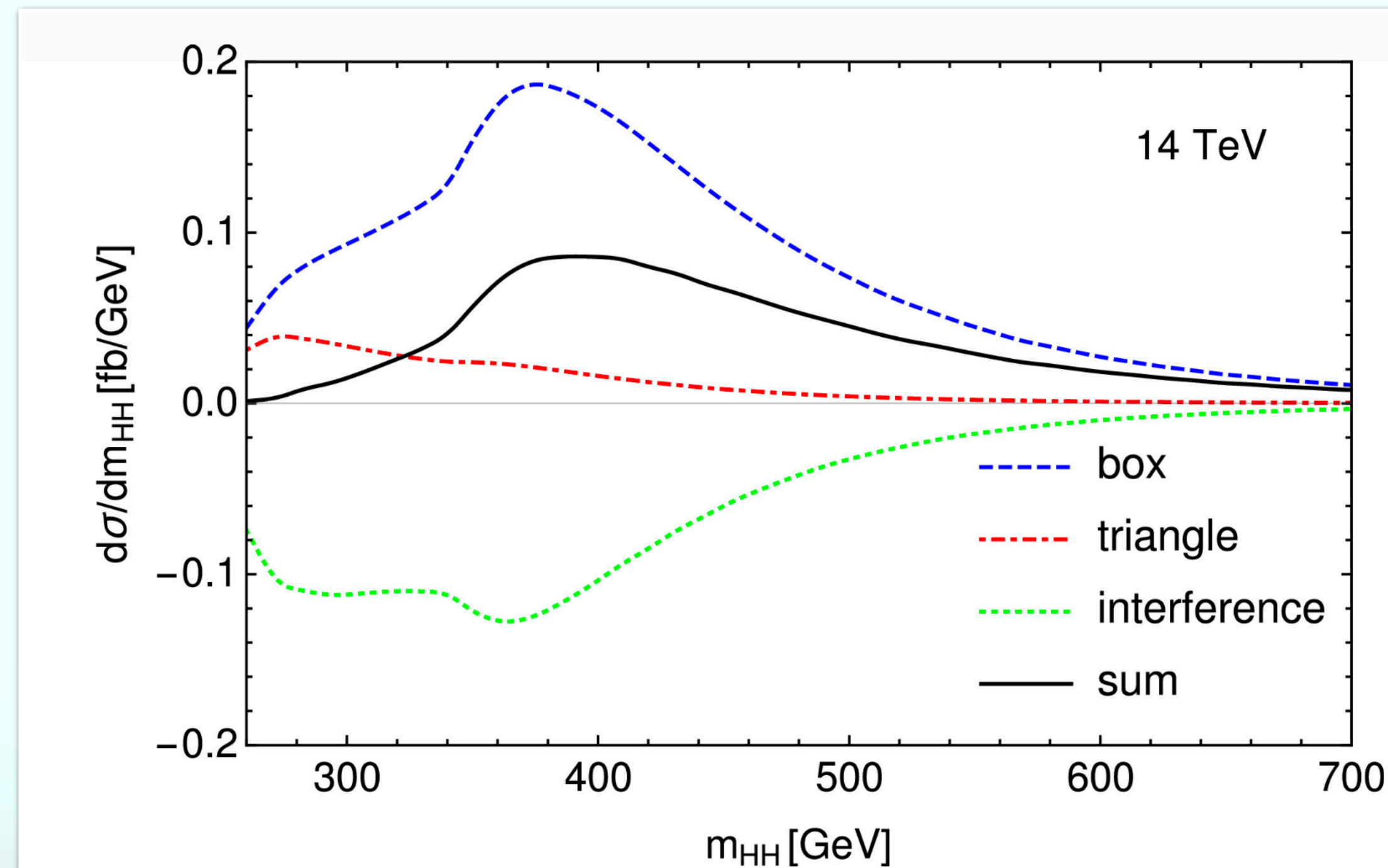
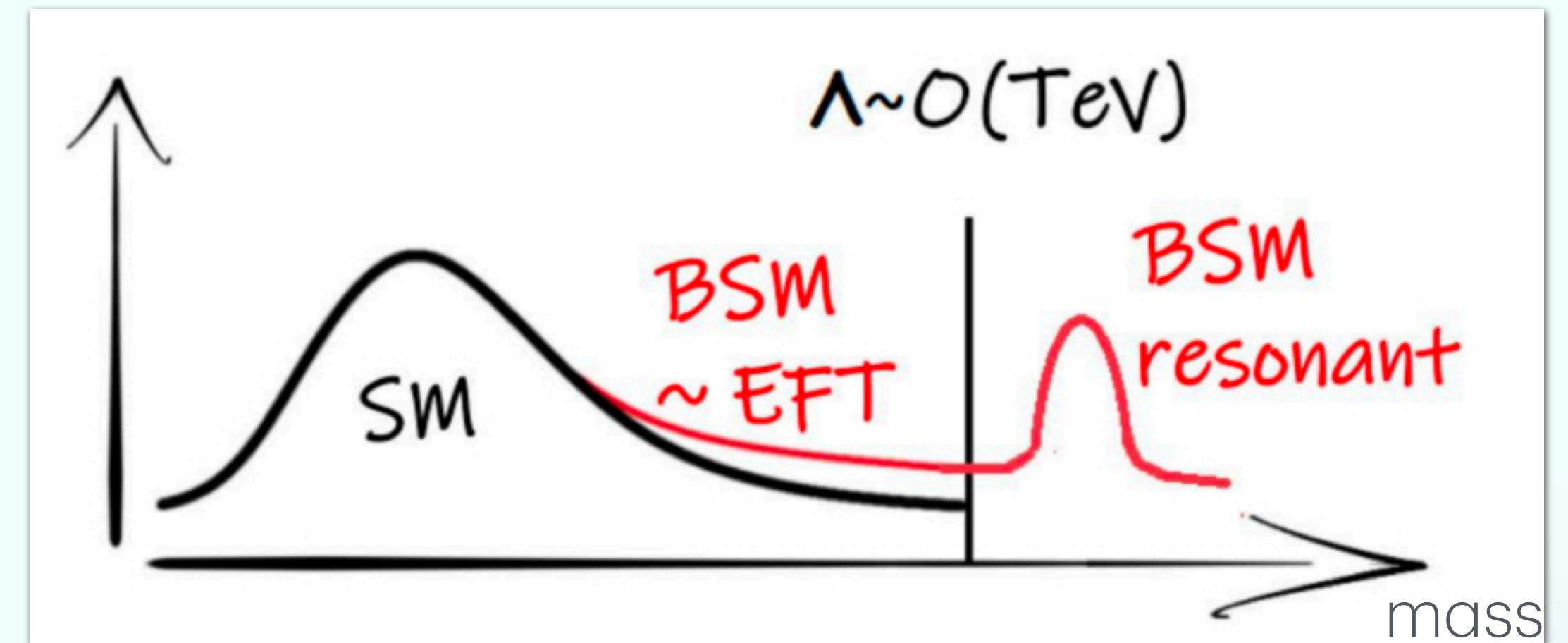


FIGURE 1.4 | Differential cross-section of producing a pair of Higgs bosons at the LO as function of the invariant mass of the Higgs pair. The figure breaks down the total probability into contributions from different underlying Feynman diagrams involved in gluon-gluon fusion, including how these contributions interfere with each other. The final, observable quantity is the sum of all these individual contributions and their interferences [25].

Backup

Probing new physics via EFTs

- λ_{HHH} hints BSM
- to parametrize these effects



SMEFT

- Expanded in canonical mass dimensions
- $\mathcal{L}_{SMEFT} = \mathcal{L}_{SM} + \frac{c_6}{\Lambda^2}(\phi^\dagger \phi)^3 + \dots$
- Higgs is SU(2) doublet, EWSB is linear
- Motivates κ_λ search

HEFT

- Expanded by chiral loops & topology
- EWSB is non-linear, Higgs is a scalar singlet
- single H / di-Higgs
- larger deviations
- κ_{2V}

Backup

SMEFT & HEFT Lagrangian

$$\mathcal{L}_{\text{SMEFT}} = \mathcal{L}_{SM} + \sum_{d>4}^{\infty} \sum_i \frac{c_i^{(d)}}{\Lambda^{d-4}} \mathcal{O}_i^{(d)}$$

- Odd values of d cause lepton number violations forcing us to consider only the even values (usually only dim-6)

$$\mathcal{L}_{\text{HEFT}} = \mathcal{L}_{LO} + \sum_{L=1}^{\infty} \left(\frac{1}{16\pi^2} \right)^L c_i^{(L)} \mathcal{O}_i^{(L)}$$

HEFT well suited for models with two EWSB (Higgs vev + higher energies) → composite-Higgs, etc

Backup

MC signal samples - Parameter points

TABLE 4.1 | The production cross-section of the double Higgs GGF signal samples at NLO precision.

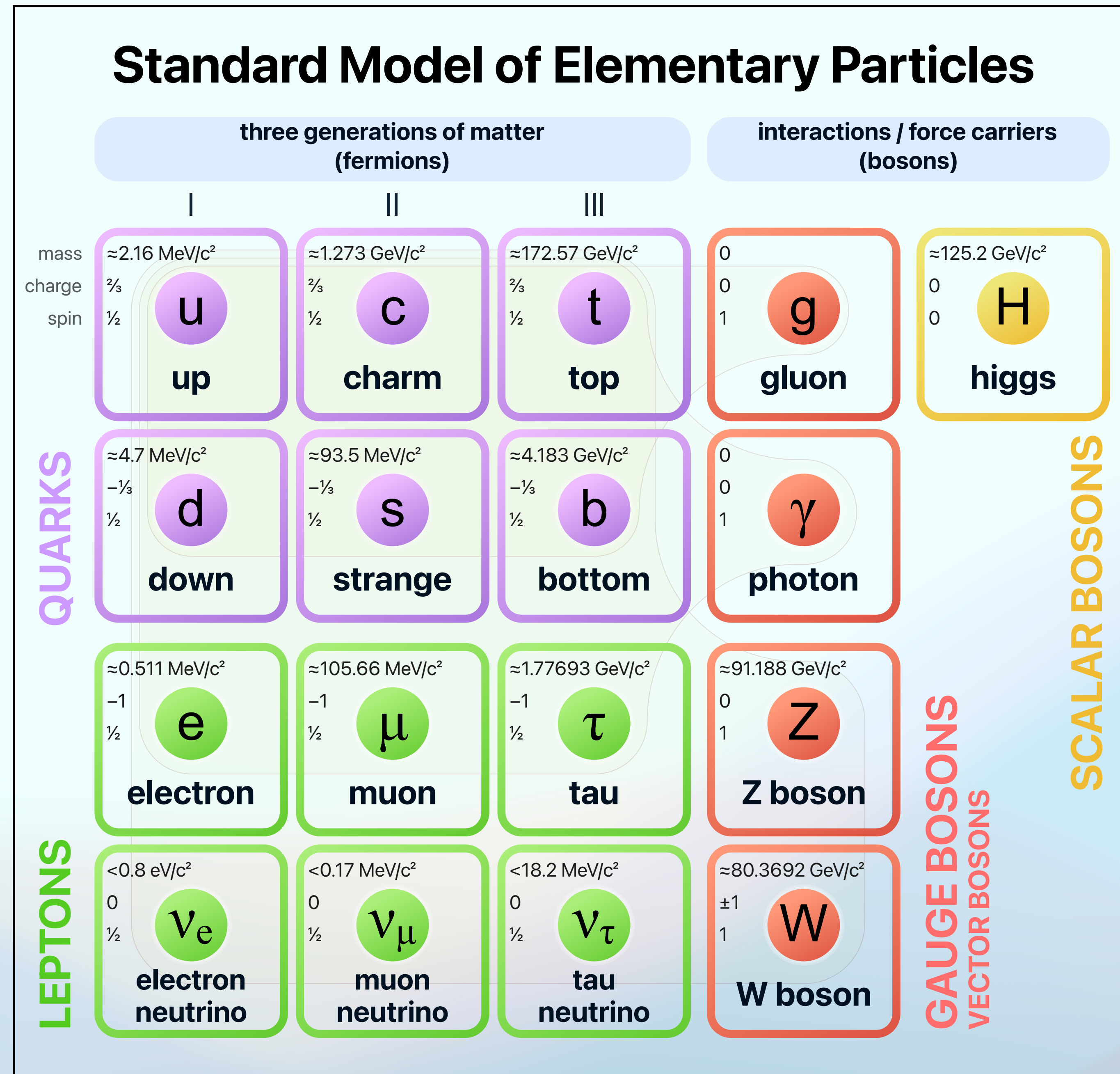
Sample	κ_λ	Cross-section [fb]
SM	1	34.13
ggF 2	0	65.30
ggF 3	2.45	12.85
ggF 4	5	83.49

TABLE 4.2 | List of VBF HH signal samples with their corresponding coupling modifiers used in the analysis.

Sample	κ_V	κ_{2V}	κ_λ	Cross-section [fb]
SM	1.00	1.00	1.00	1.9
VBF 2	1.74	1.37	14.40	394.6
VBF 3	-0.96	0.96	-1.43	1.11
VBF 4	-0.76	1.44	-19.30	354.8
VBF 5	-1.60	2.72	-1.36	11.58
VBF 6	-0.01	0.03	10.20	0.01
VBF 7	-1.21	1.94	-0.94	3.76
VBF 8	-1.83	3.57	-3.39	16.63
VBF 9	-2.12	3.87	-5.96	672.1

Backup

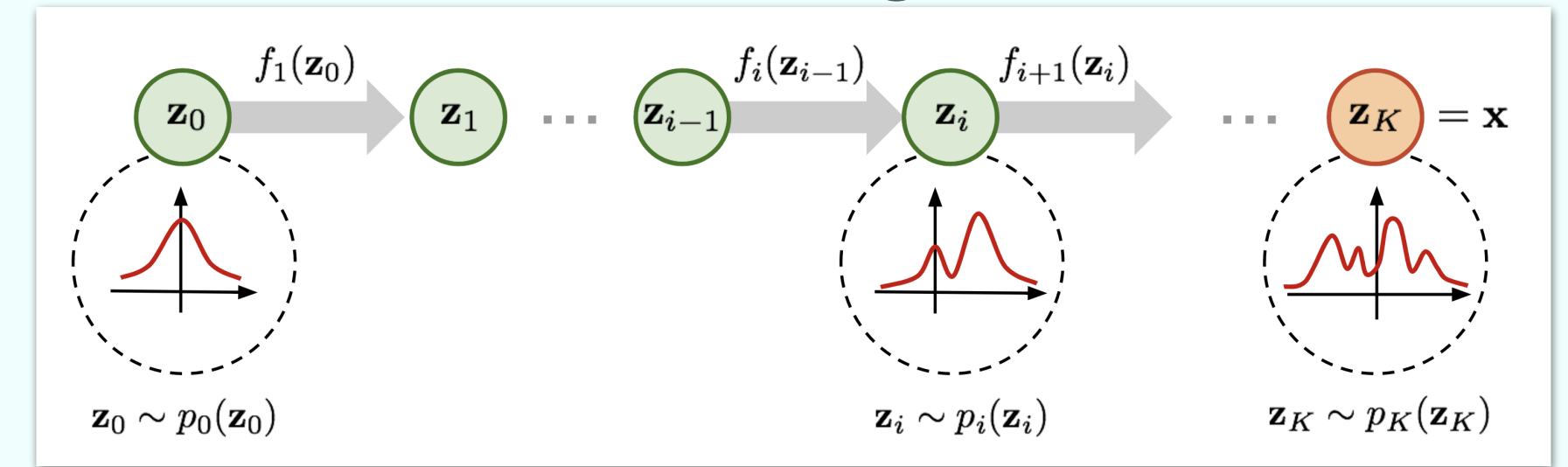
SM particles



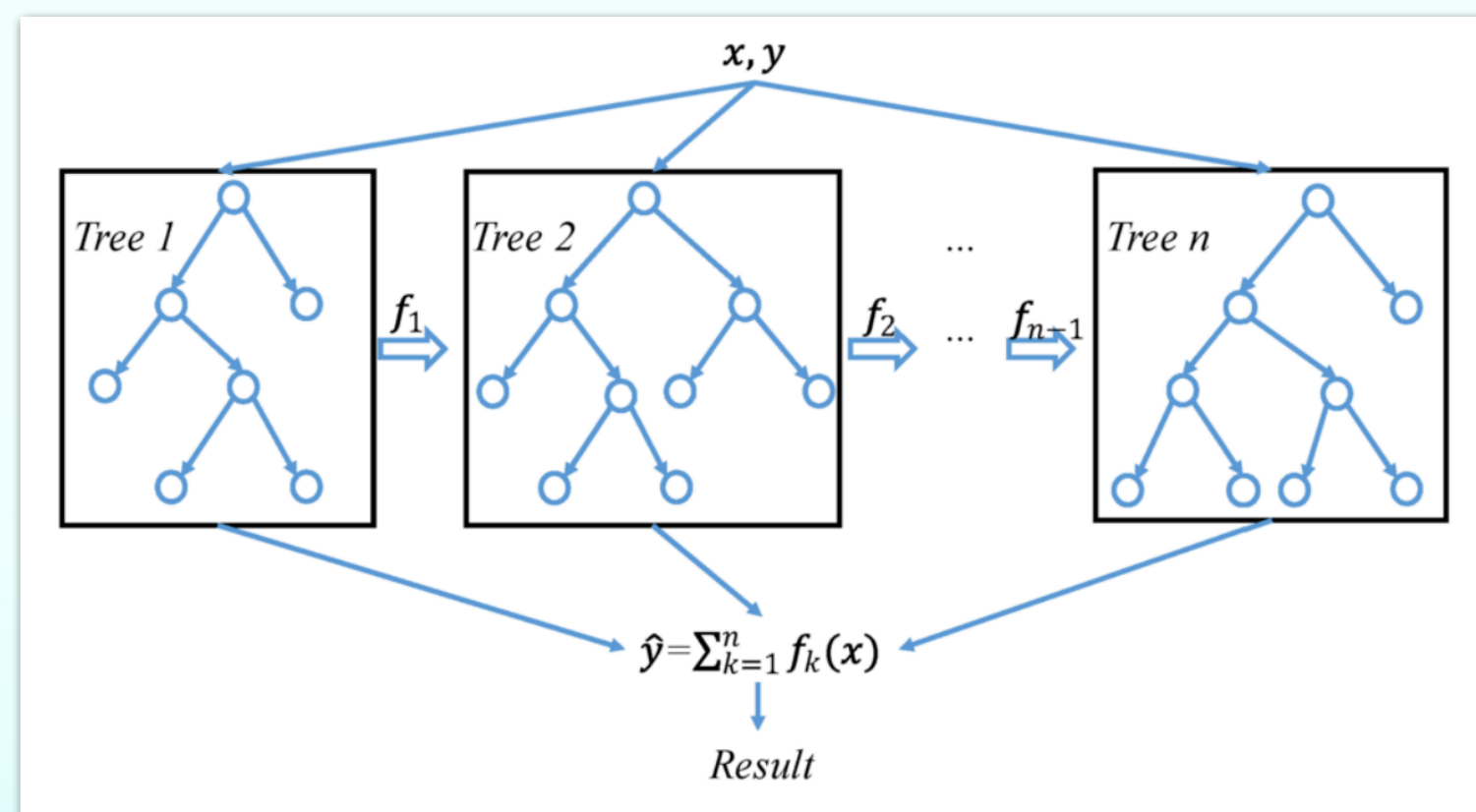
Backup

Common supervised ML models in HEP

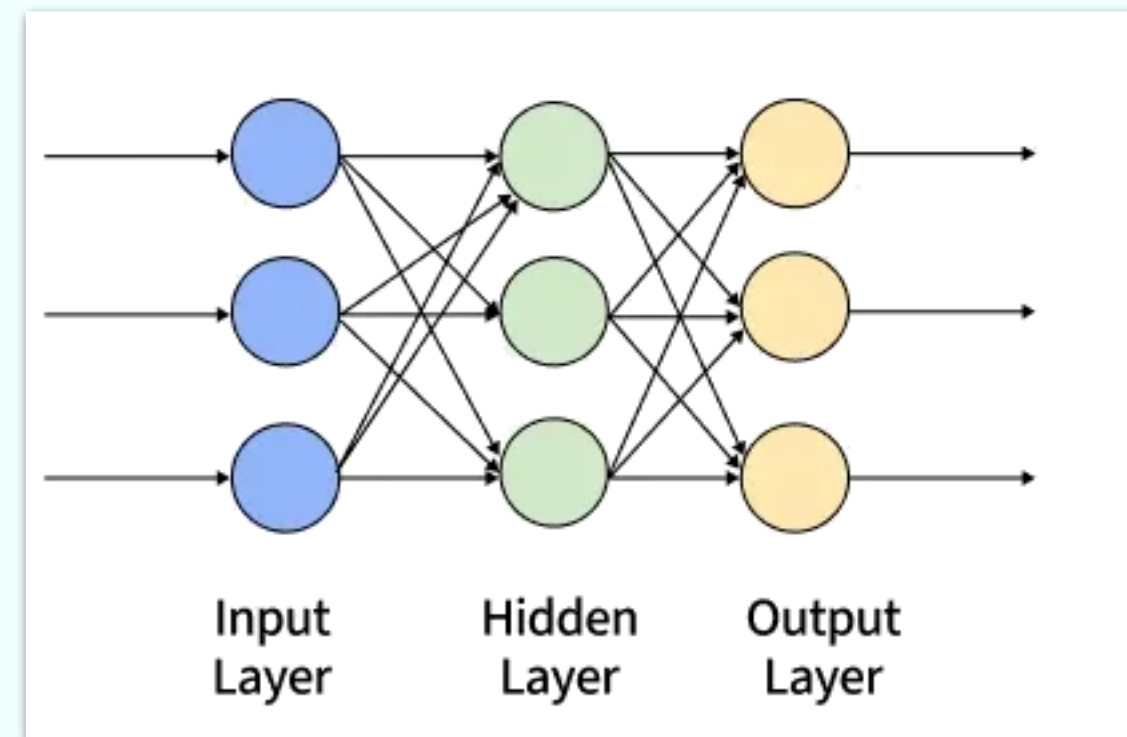
Normalizing flows



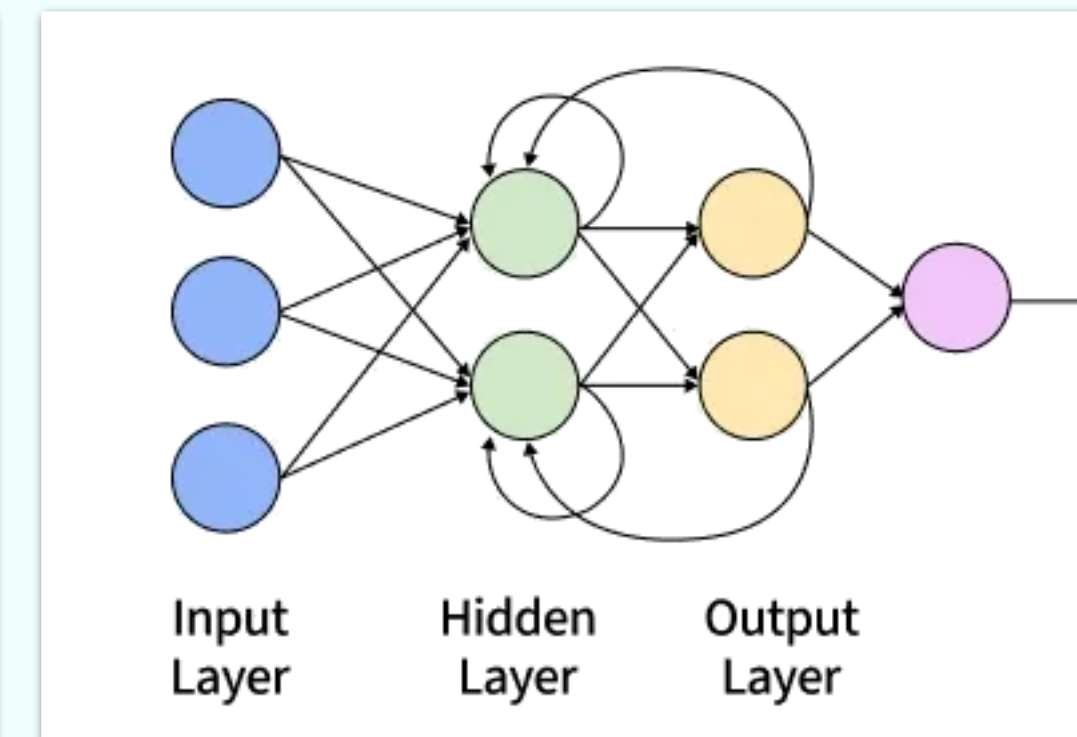
BDT



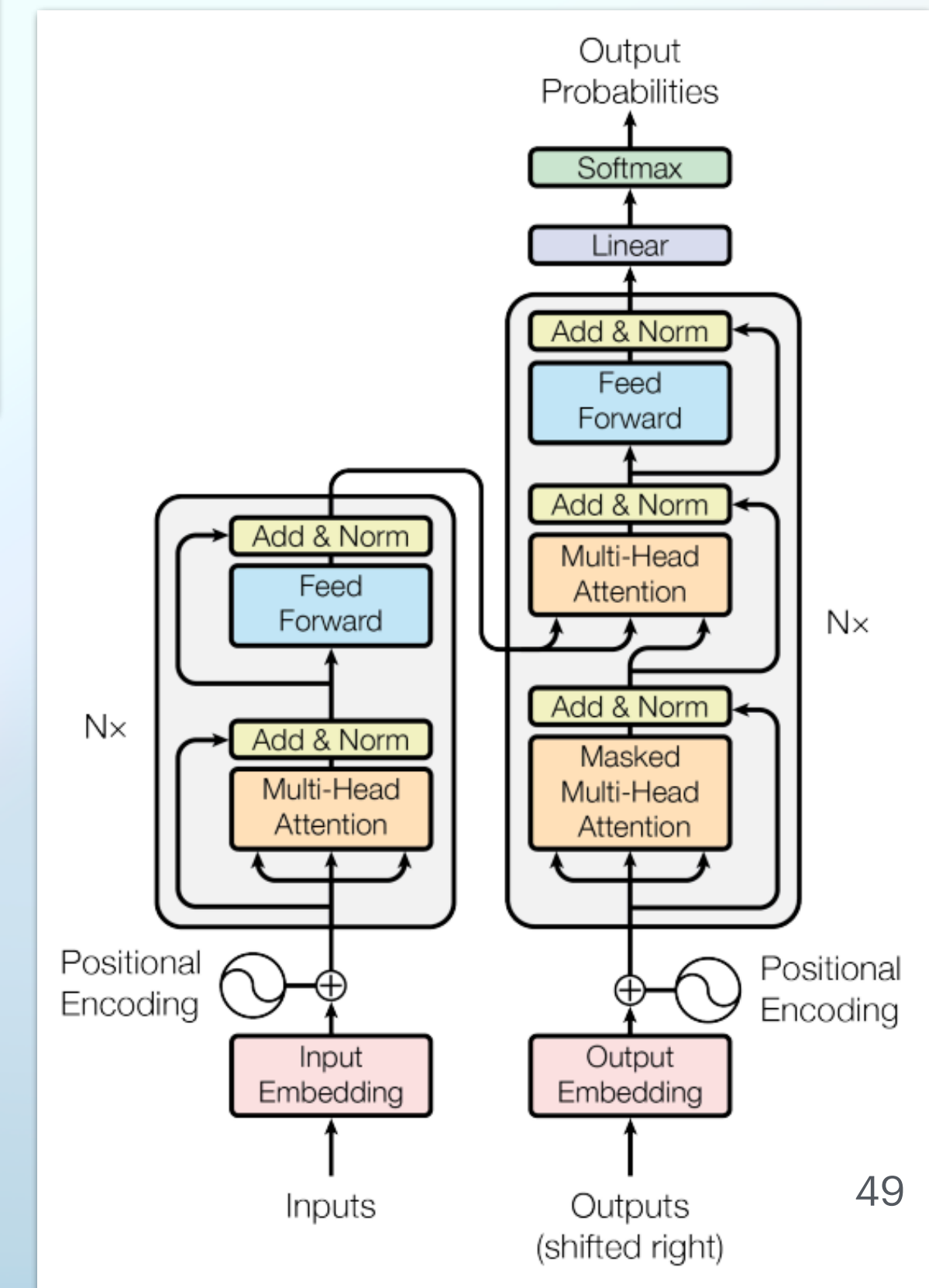
DNN



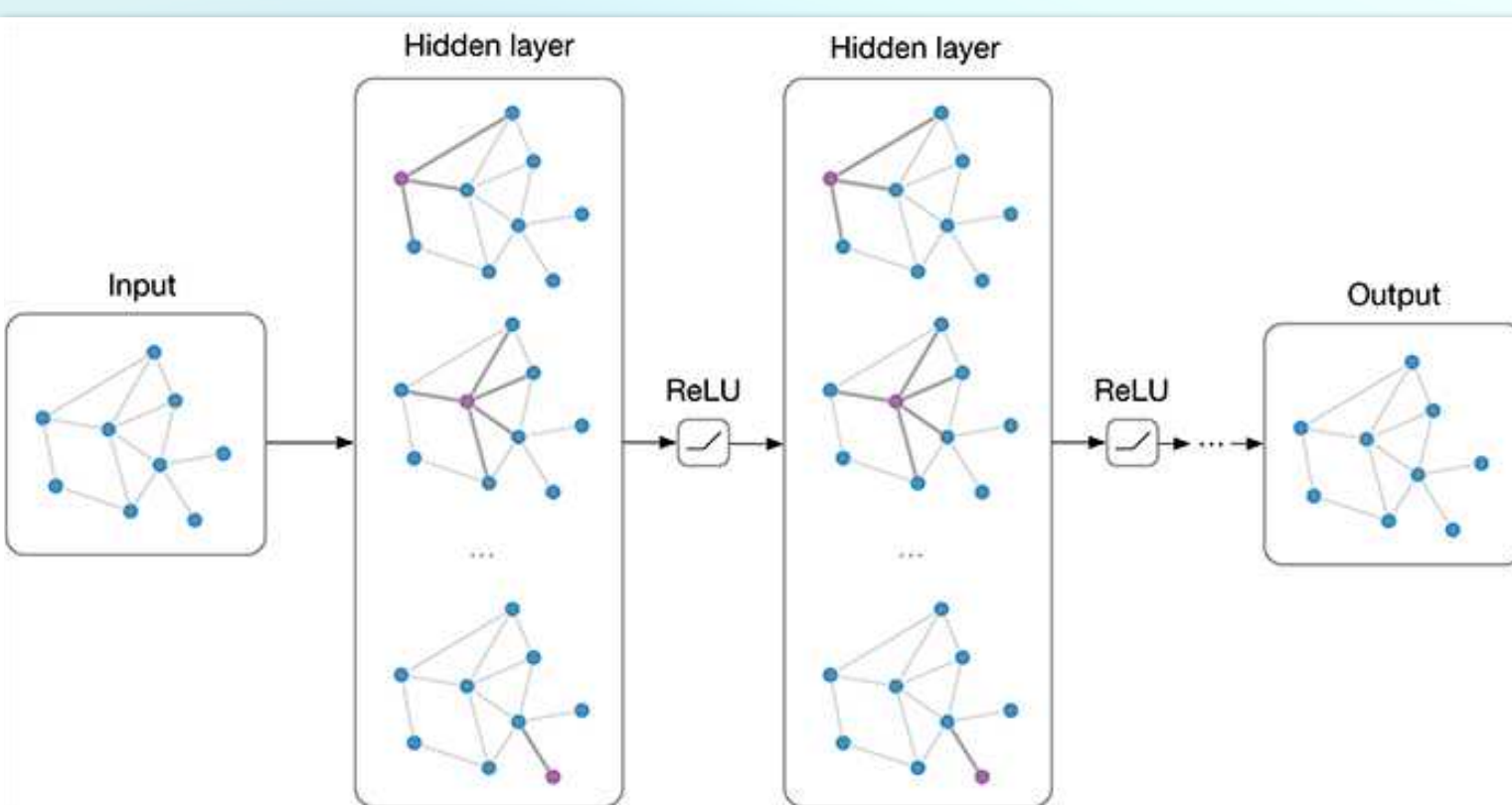
RNN



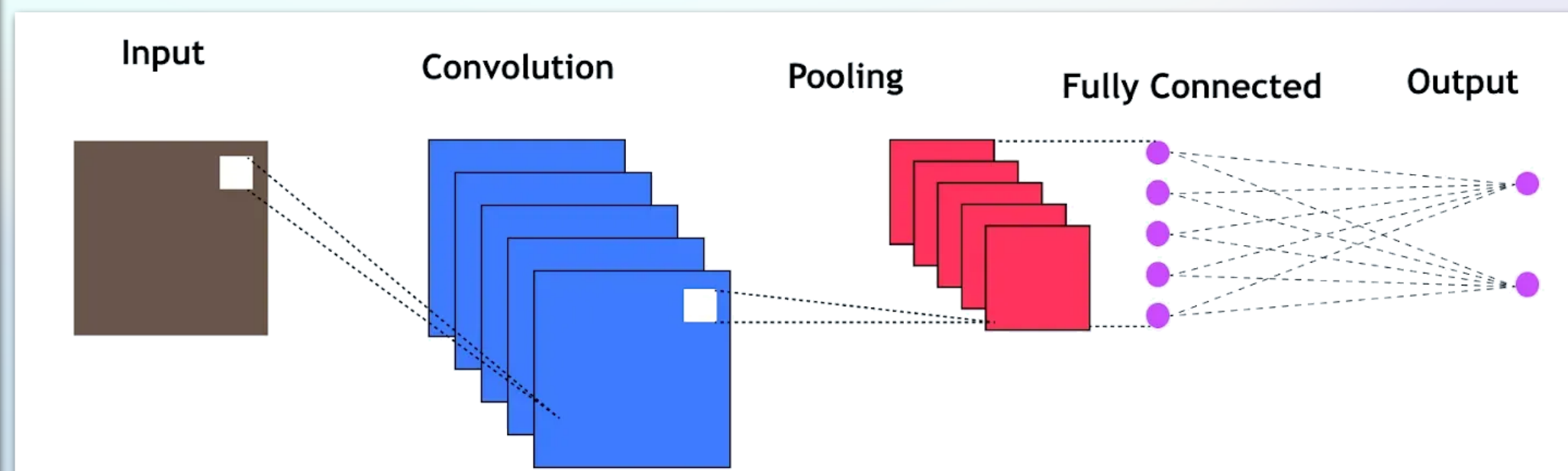
Transformers



GNN



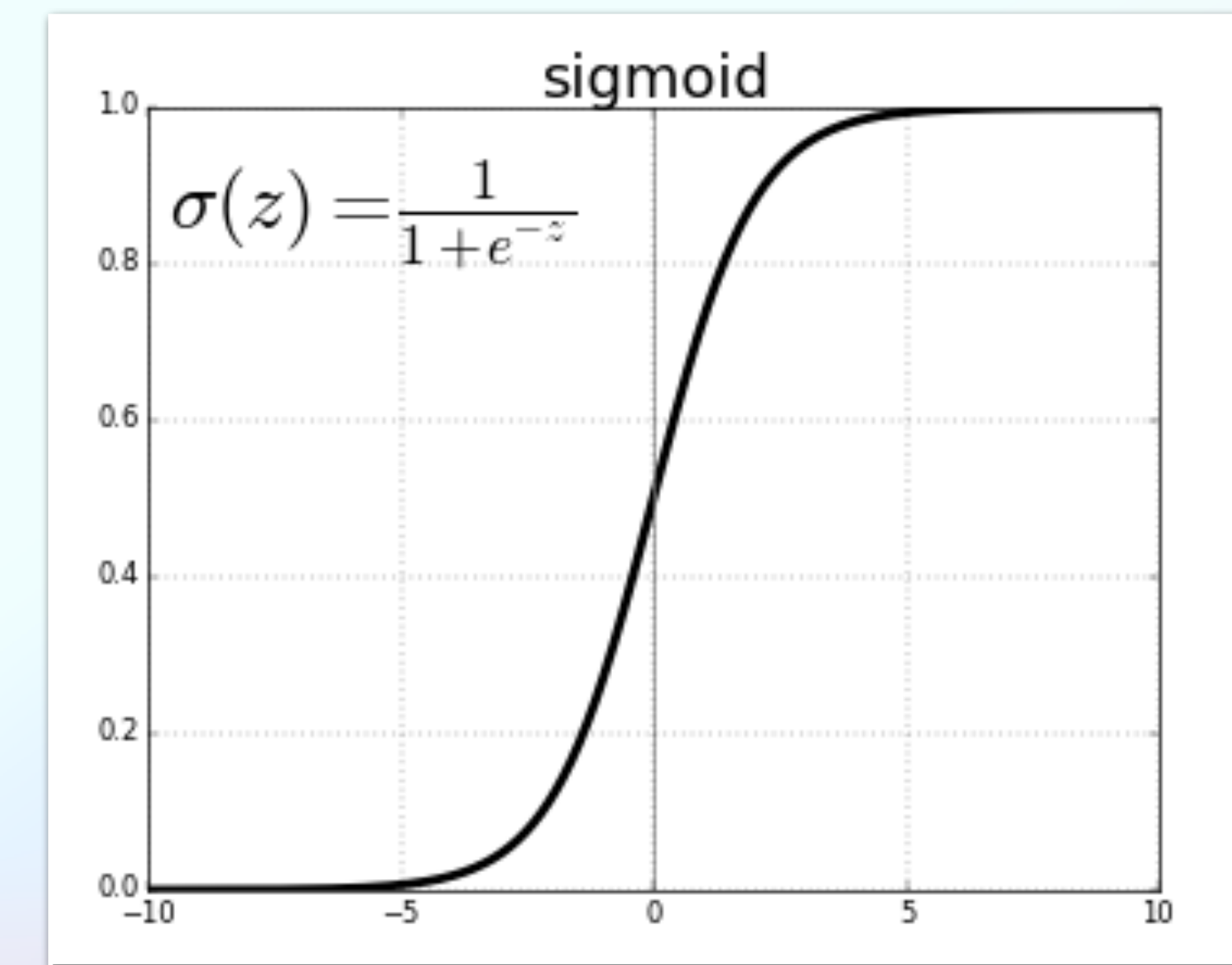
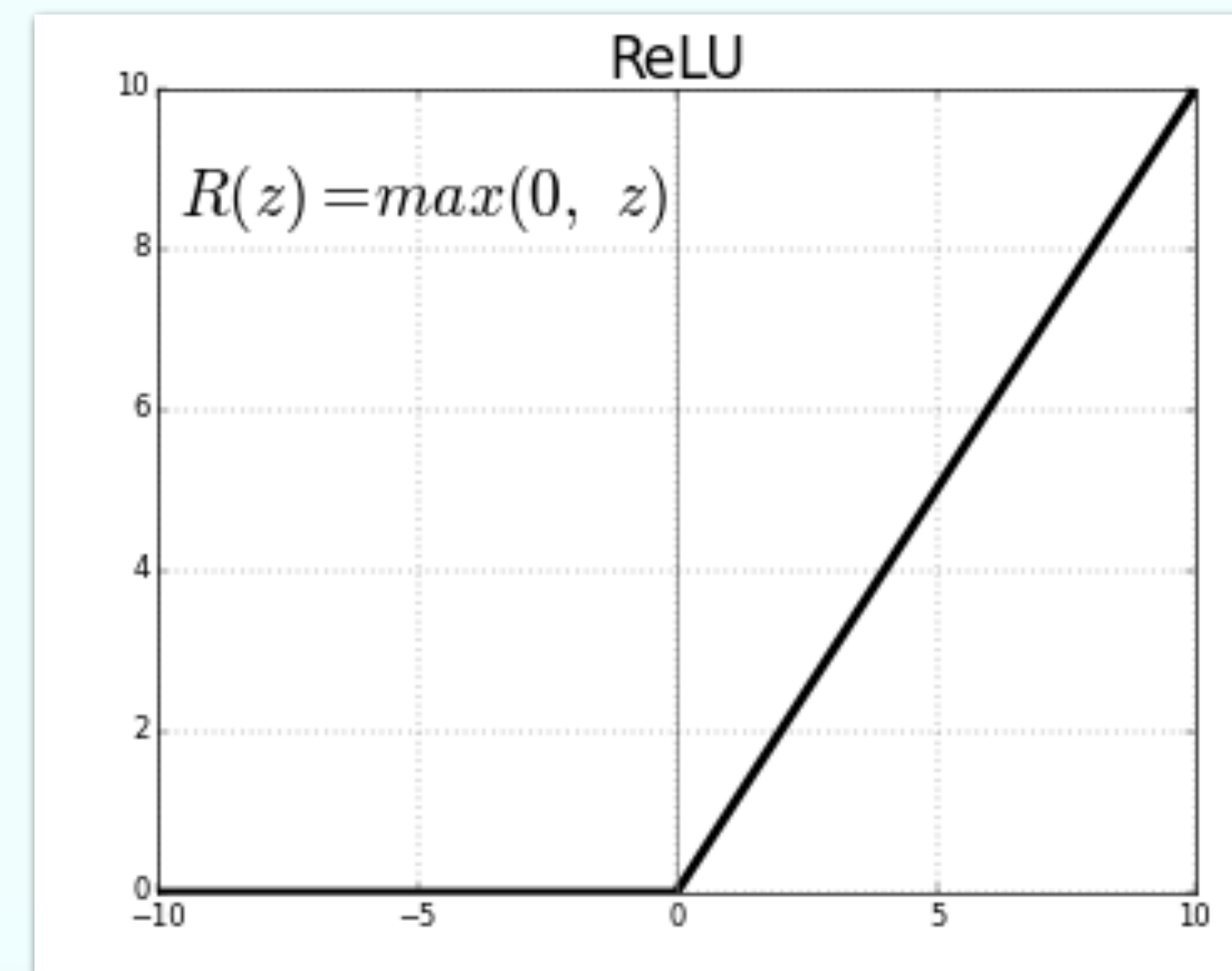
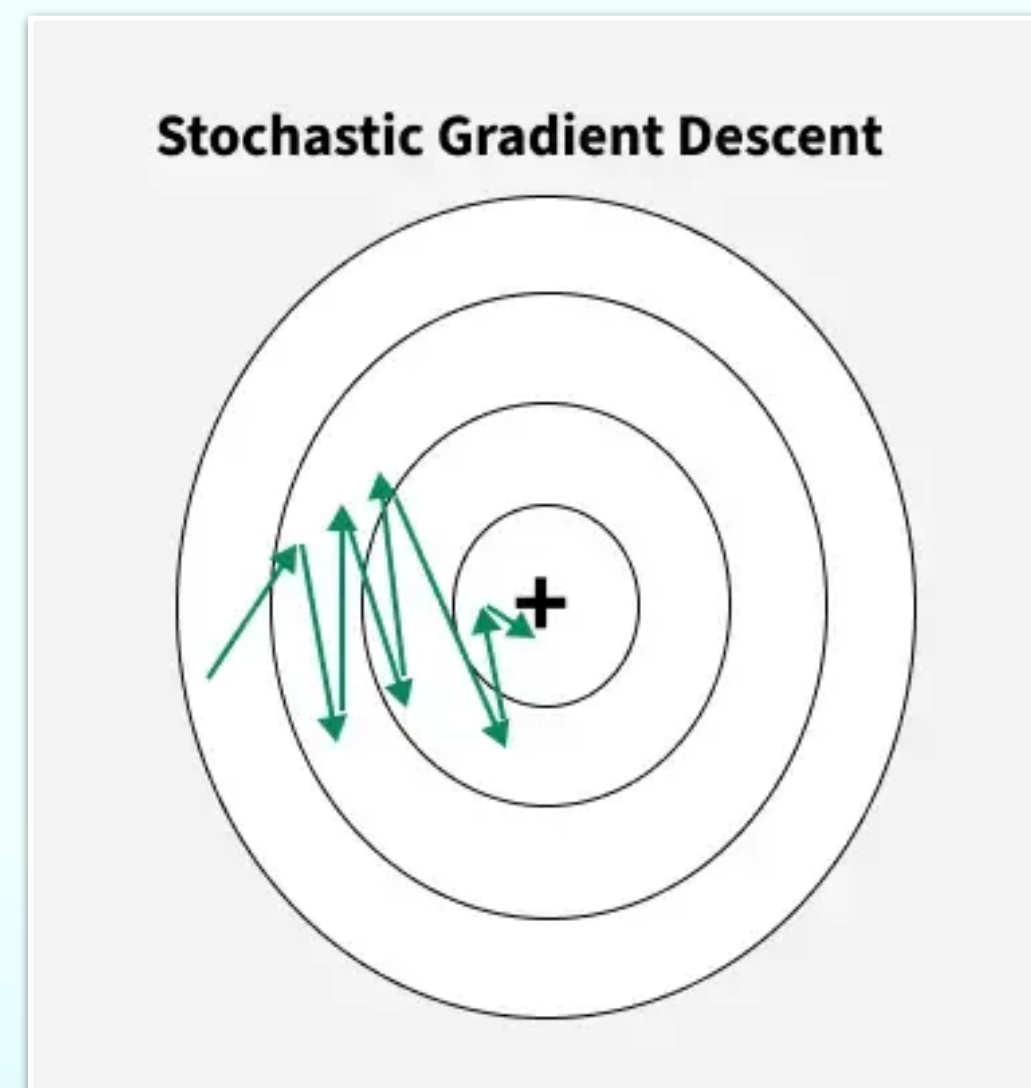
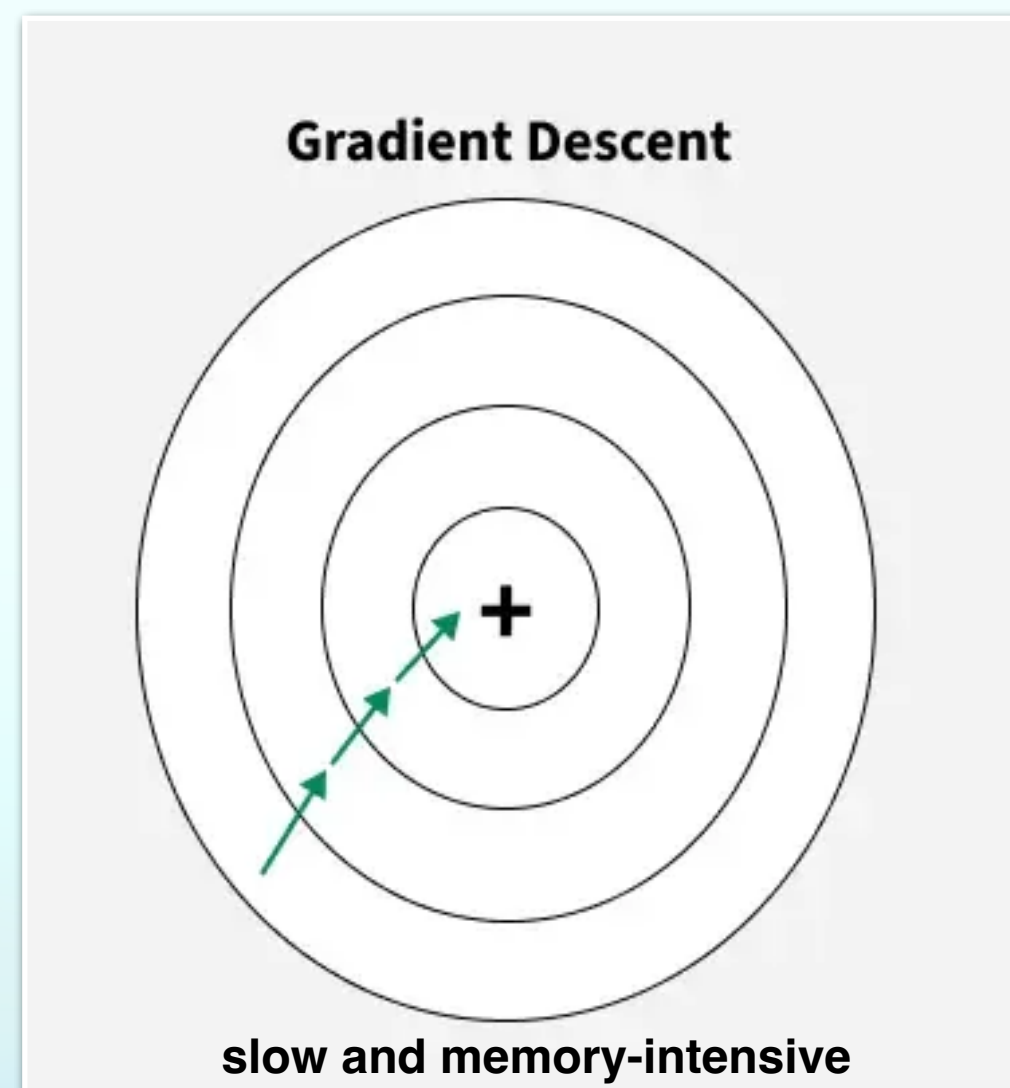
CNN



Backup

General training procedure - continued

activation functions



learning rate = length of the arrows

Backup

Practical considerations in HEP analyses

Class imbalance and weighting

Class reweighting

Oversampling

Undersampling

Negative event weights

Feature engineering

Low-level variables

High-level variables

Interpretability and validation

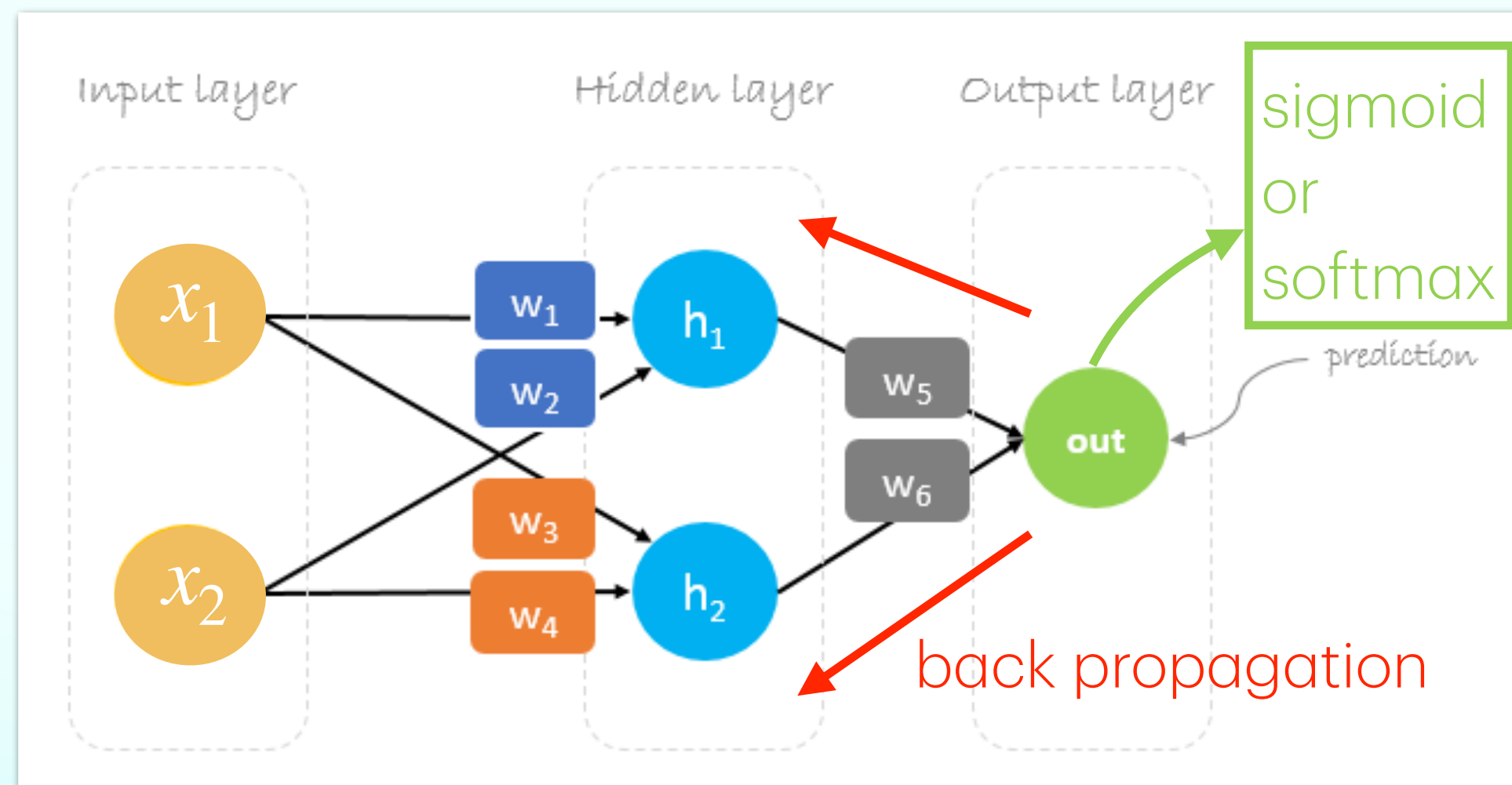
Feature importance

Partial dependence plots

SHAP values (SHapley Additive
exPlanations)

Backup

(Deep) Neural Networks



a : output of a neuron

$$a = \sigma \left(\sum_{i=1}^n w_i x_i + b \right)$$

w : weight

x : input

b : bias

σ : activation function

DNNs in HEP

Approximation for any continuous function

Automated feature engineering

Scalability with data

Flexibility of architecture

Considerations:

- Input pre-processing (scale of the inputs)
- Negative event weights
- Computational resources (architecture)
- Hyperparameter optimization

Current research:

- Generative models for fast simulation
- Real-time inference on FPGAs
- Anomaly detection

Backup

DNN implementation for bbWW

- Fully connected network with 3 hidden layers (32, 16, 8 neurons)
- ReLU activation
- output layer with a sigmoid activation
- event weights are used from cutflow study + weight balance between signal & bkg is assured
- negative weight events are removed, around 0.5% of total events
- binary cross-entropy
- Adam optimizer
- batch size of 1048, early stopping of 5 epochs on the validation loss, heuristic hyper-parameter search, 5 fold CV
- Platt scaling
 - sigmoid squashes any number into 0-to-1 range but it doesn't mean it's truly probability
 - Platt scaling uses a validation dataset to train and adjust the sigmoid output to match the actual odds.

Backup

Transformer implementation for bbWW

- Binary classification
- Transformer encoder-only model
- The model receives a sequence of 8 embeddings representing the event: 2 leptons, 4 AK4 jets, 1 AK8 jet, and the MET
- Embeddings include $(P_x P_y P_z E)$ and object-specific parameters
- each embedding is a MLP
- 6 encoder layers stacked
- 4 attention heads
- positional encoding is not used
- mean of the encoder output is used (b,s,d)
- k-fold CV is not used due to computational expense
- Platt scaling

Backup

Transformers in HEP

Advantages

- global correlations
- variable-length input
- scalability

Considerations

- computationally expensive, to mitigate
 - sparse/hierarchical/linear attention
 - input pruning (reduce the sequence length)
- choice of embedding also affect comp. time
 - dense/Lorentz-invariant/learnable

Backup

BDTs in HEP

- XGBoost (Extreme Gradient Boosting) - L1/L2
- LightGBM (Light Gradient Boosting Machine)
 - asymmetrical trees

Advantages

- Performance on tabular data
- Training speed and efficiency
- Robustness to small datasets

Considerations

- Feature engineering dependency
- Correlated variables (selecting one)
- Discrete nature and gradient descent

Current trend

- Differentiable Decision Trees
- Fast inference on FPGAs for trigger
- Gradient boosting for reweighting

Backup

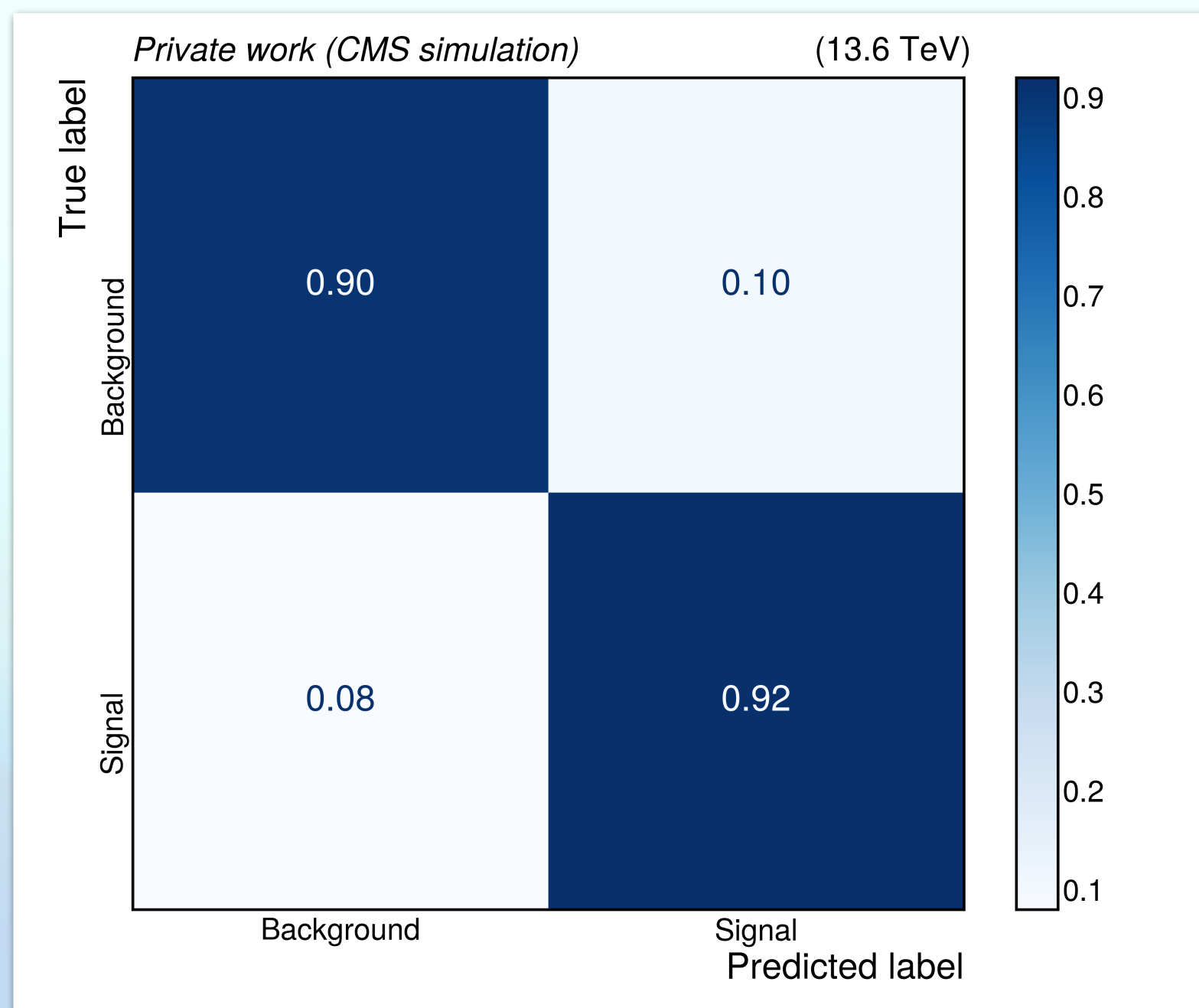
BDT for bbWW

- 200 boosting rounds with a maximum depth of 3
- learning rate 0.1, decaying by a factor of 0.99 every round
- early stopping if no improvement over 30 rounds
- L1 and L2 regularizations
- Taylor expansion to calculate both the first and second-order derivatives (Hessian) of the loss function
- XGBoost

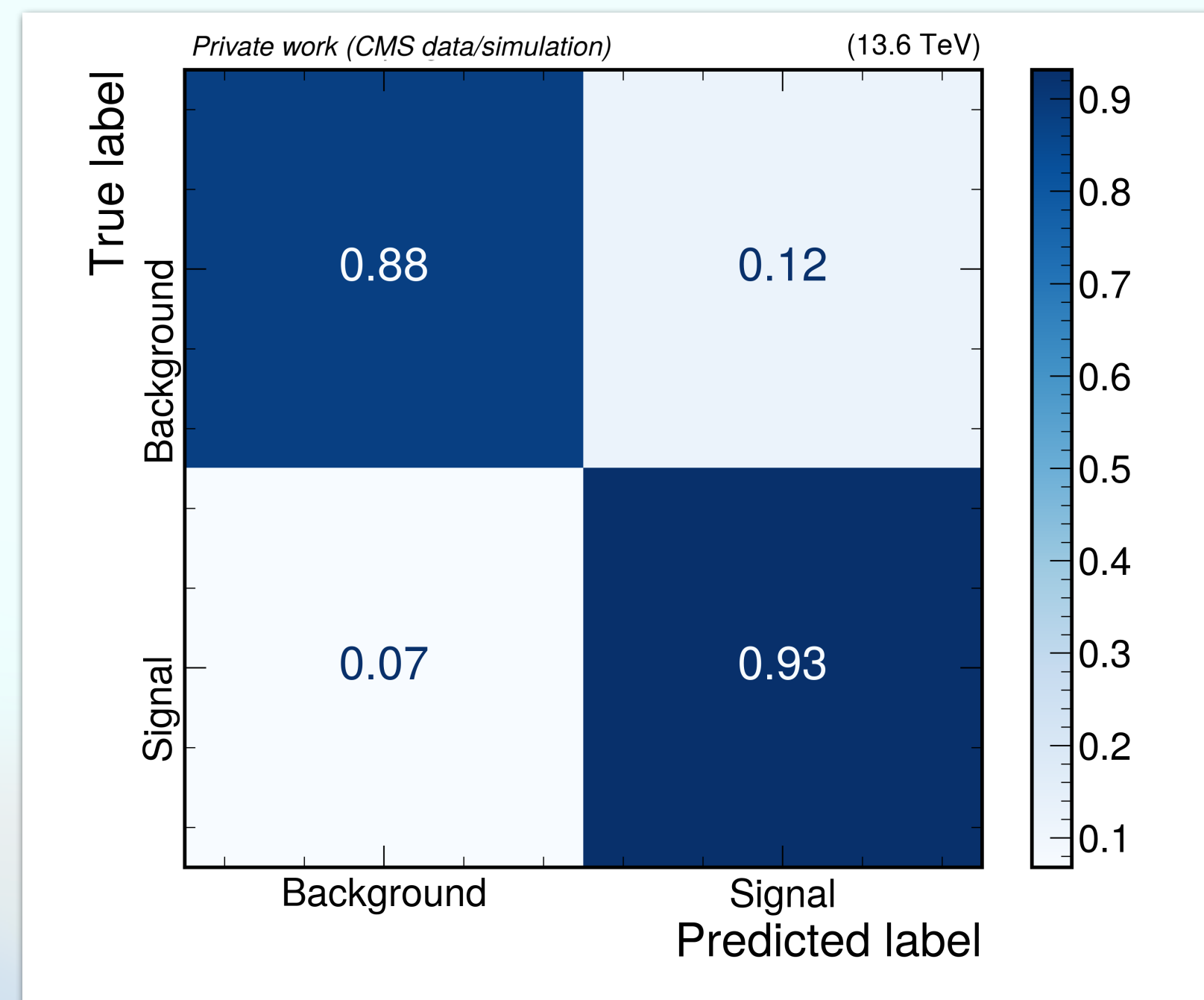
Backup

Comparison of models - Performance (confusion matrices)

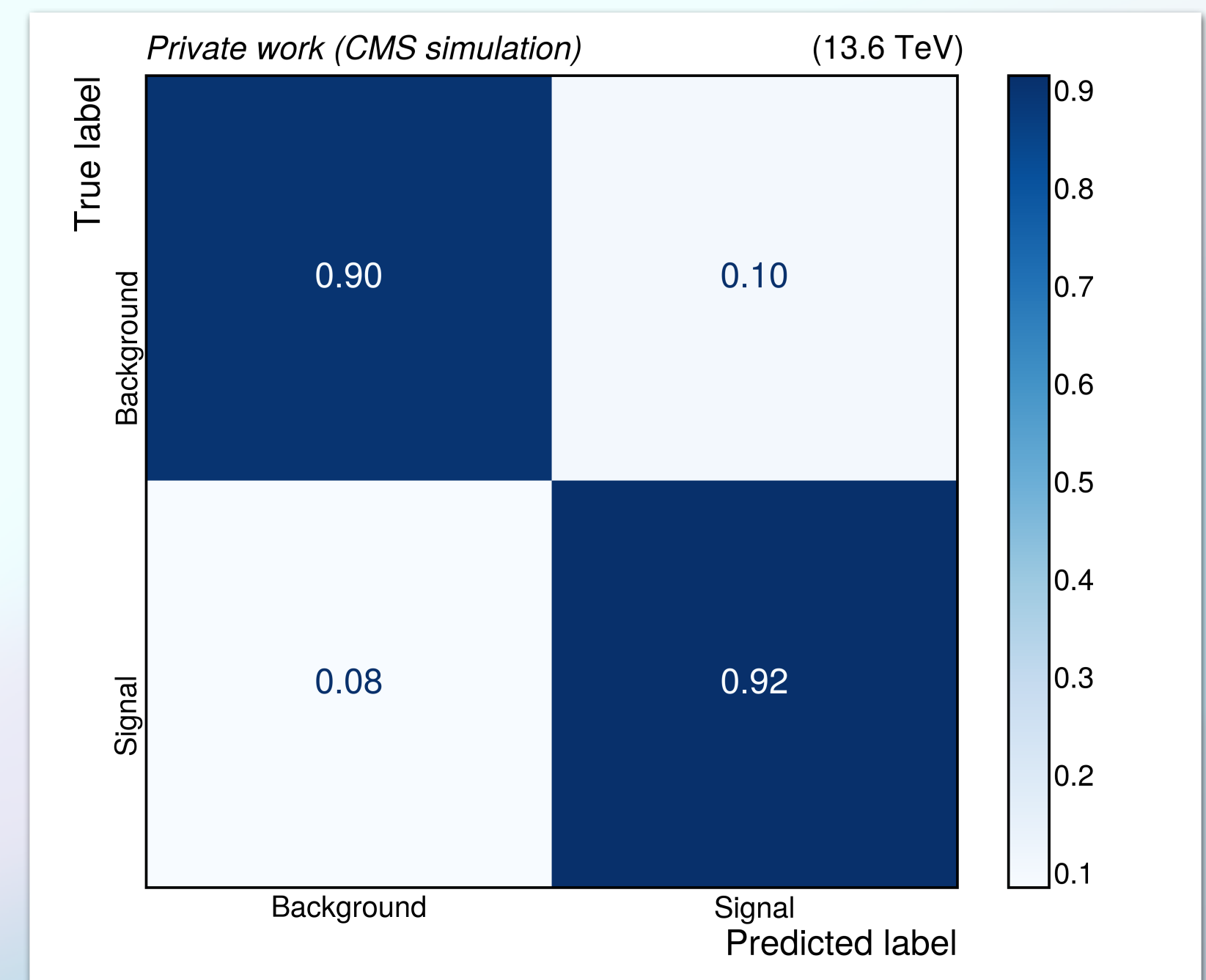
DNN



Transformer

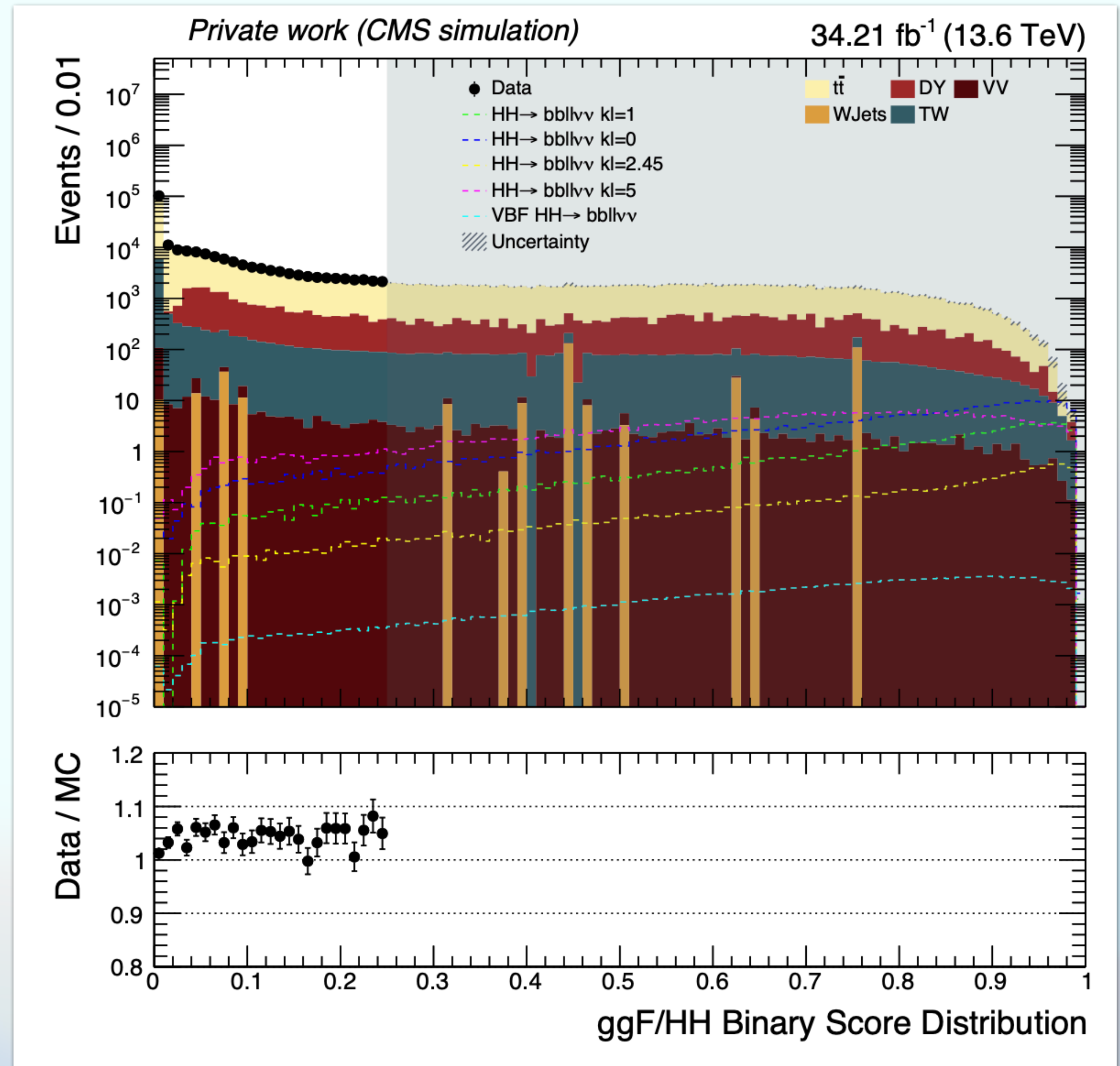


BDT



Backup

ML score distribution



Backup

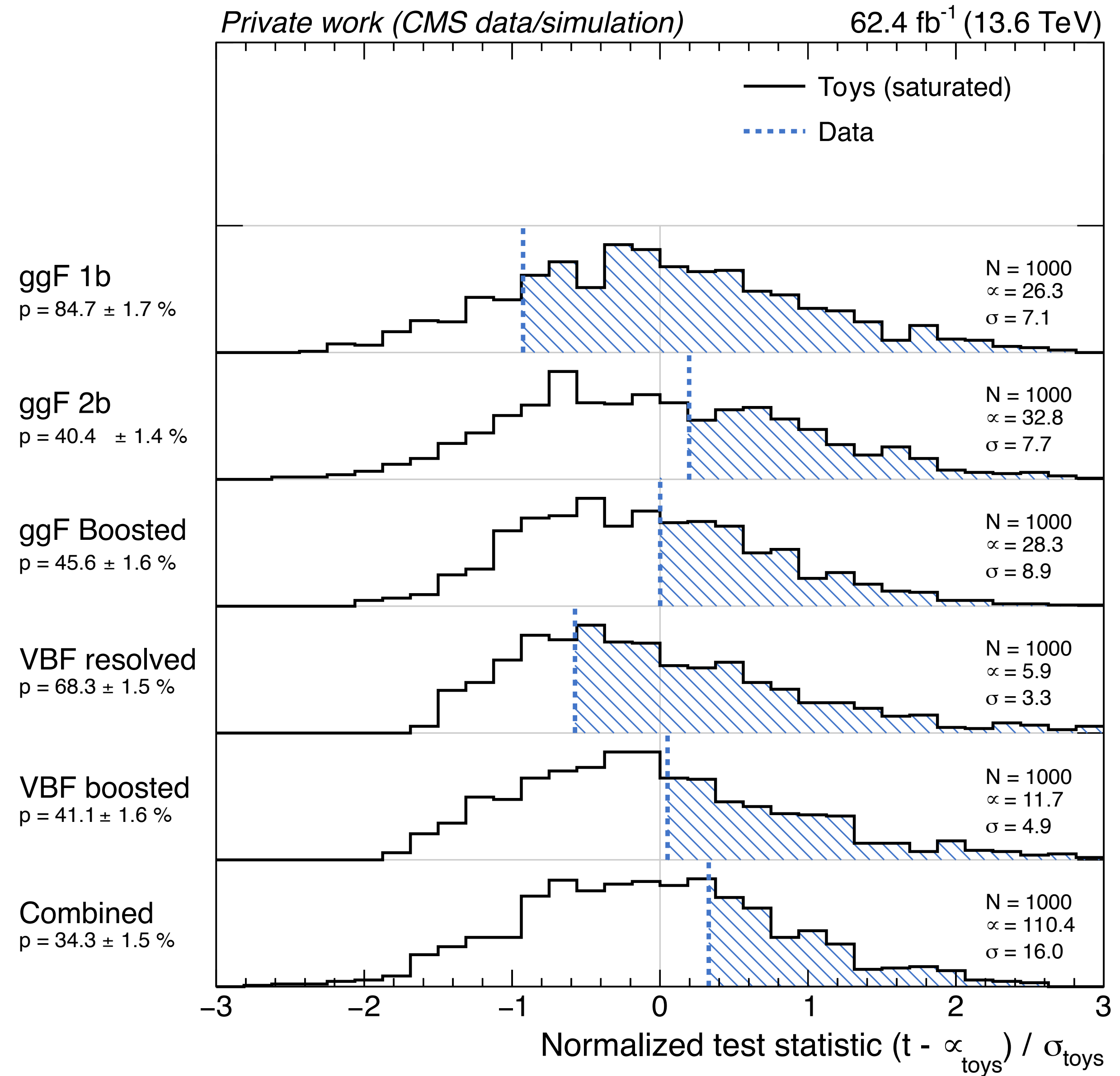
Event yields in SR categories

Table 1: Event yields for various SR categories.

Category	Tot. Signal	TT	DY	ST	Others	Tot. MC	Data	Data/MC
ggF boos. ee	0.7 ± 0.0	33.7 ± 1.2	27.6 ± 4.8	3.3 ± 0.2	5.0 ± 0.5	69.7 ± 5.1	63	0.90 ± 0.13
ggF boos. $\mu\mu$	2.5 ± 0.0	122.9 ± 2.7	95.4 ± 21.6	11.0 ± 0.5	15.3 ± 0.9	244.5 ± 22.0	271	1.11 ± 0.12
ggF boos. $e\mu$	2.0 ± 0.0	189.1 ± 3.7	3.5 ± 0.6	17.8 ± 0.6	12.4 ± 1.0	222.8 ± 4.2	241	1.08 ± 0.07
ggF res. 1b ee	3.9 ± 0.1	3903.9 ± 56.0	2499.8 ± 105.3	301.3 ± 4.8	105.5 ± 3.2	6810.6 ± 138.3	6232	0.92 ± 0.02
ggF res. 2b ee	3.1 ± 0.1	3945.4 ± 56.5	321.4 ± 30.6	124.4 ± 2.3	24.8 ± 1.0	4415.9 ± 69.9	3980	0.90 ± 0.02
ggF res. 1b $\mu\mu$	12.5 ± 0.2	14220.7 ± 200.4	10180.2 ± 251.1	1036.1 ± 15.1	330.8 ± 6.2	25768.2 ± 416.5	26479	1.03 ± 0.02
ggF res. 2b $\mu\mu$	10.0 ± 0.1	14889.0 ± 209.7	1114.1 ± 67.1	415.5 ± 6.4	81.7 ± 2.1	16501.2 ± 241.1	17014	1.03 ± 0.02
ggF res. 1b $e\mu$	12.9 ± 0.2	22592.4 ± 317.6	153.7 ± 6.3	1680.4 ± 24.1	301.0 ± 6.4	24737.9 ± 347.6	25022	1.01 ± 0.02
ggF res. 2b $e\mu$	10.3 ± 0.2	23257.8 ± 326.8	20.0 ± 2.0	686.8 ± 12.0	85.5 ± 2.3	24050.2 ± 338.0	24096	1.00 ± 0.02
VBF res. ee	0.2 ± 0.0	74.0 ± 1.9	40.2 ± 7.4	1.6 ± 0.2	2.8 ± 0.4	118.6 ± 7.8	124	1.05 ± 0.12
VBF boos. ee	0.0 ± 0.0	14.4 ± 0.7	28.9 ± 10.2	0.6 ± 0.1	1.4 ± 0.3	45.3 ± 10.3	42	0.93 ± 0.25
VBF res. $\mu\mu$	0.6 ± 0.0	244.6 ± 4.5	165.8 ± 25.5	4.7 ± 0.3	8.9 ± 0.8	423.9 ± 56.2	491	1.16 ± 0.19
VBF boos. $\mu\mu$	0.1 ± 0.0	46.6 ± 1.4	82.4 ± 18.7	2.1 ± 0.2	3.4 ± 0.5	134.6 ± 18.8	153	1.14 ± 0.18
VBF res. $e\mu$	0.6 ± 0.0	393.5 ± 6.6	2.5 ± 0.6	8.1 ± 0.6	11.7 ± 1.0	415.5 ± 7.0	434	1.04 ± 0.05
VBF boos. $e\mu$	0.1 ± 0.0	78.6 ± 1.9	2.6 ± 0.4	2.9 ± 0.2	4.4 ± 0.6	88.5 ± 2.2	82	0.93 ± 0.10

Backup

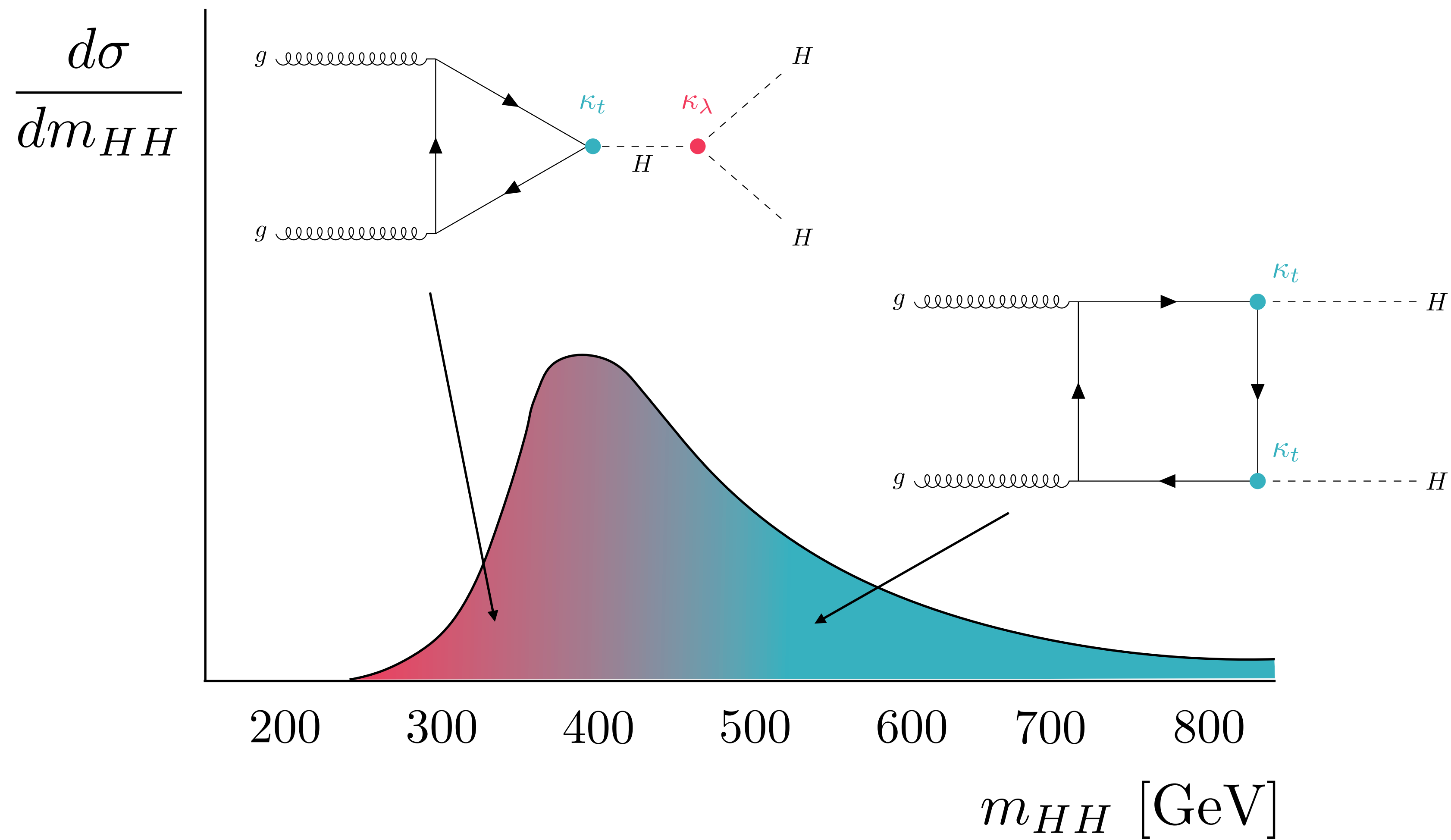
Goodness of fit



Backup

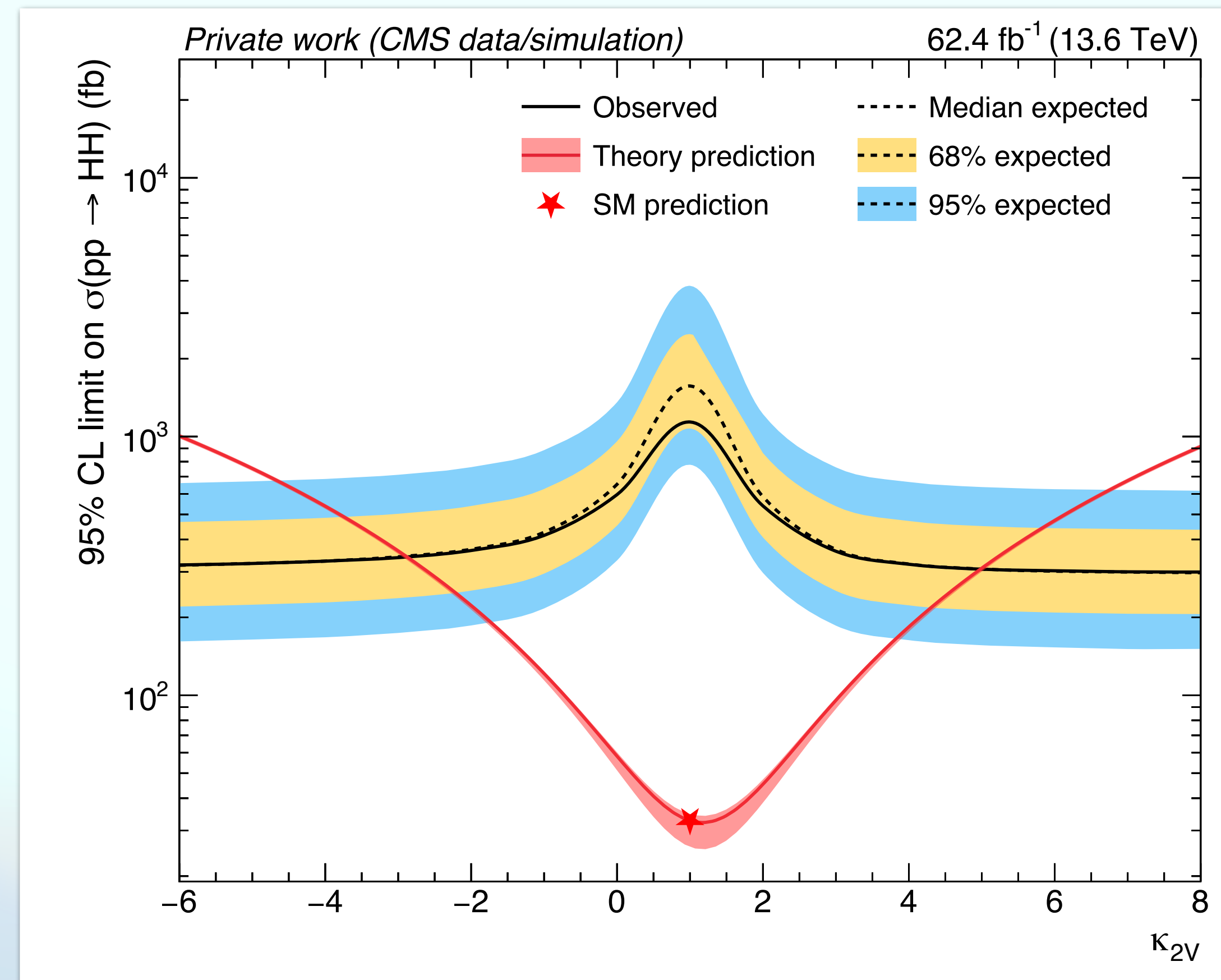
Run 3 scan comparison

	κ_λ	κ_{2V}
This thesis	[-13.3, 19.0] observed, [-20.0, 23.7] expected	[-2.9, 5.0] both observed and expected
Run 3	[-9.1, 15.7] observed, [-13.3, 19.8] expected	[-0.20, 2.25] observed, [-0.48, 2.54] expected



Backup

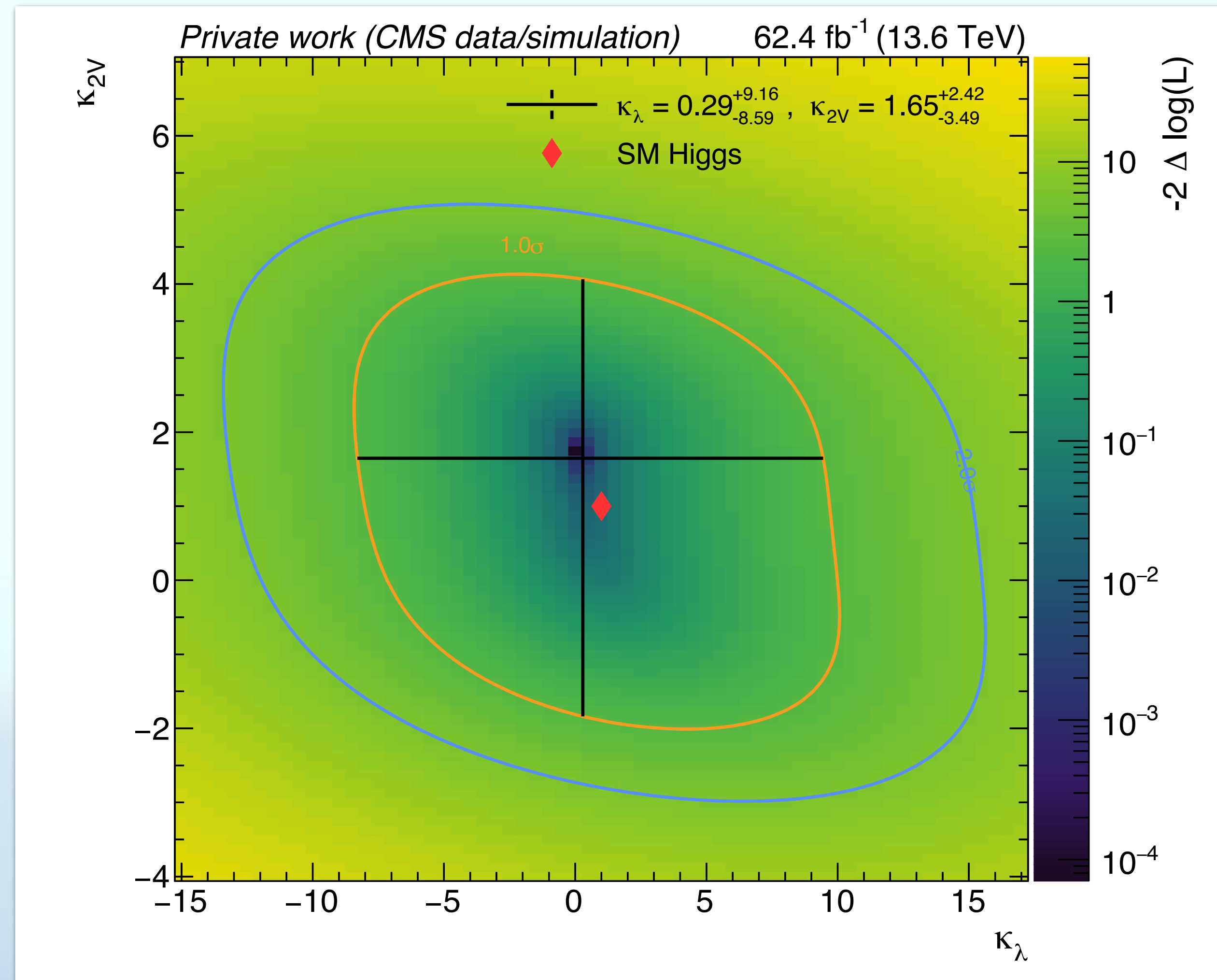
Coupling scans



[-2.9, 5.0] both observed and expected

Results

Coupling scans



Results

Coupling scans

