Search for exotic Higgs boson decays with CMS and fast machine learning solutions for the LHC

Ho Fung Tsoi [U. of Wisconsin-Madison]

Special EPE seminar at U. of Washington

March 5th, 2024

Short bio

I joined **CMS@CERN** in 2020 as a PhD candidate in the group of Sridhara Dasu at Wisconsin-Madison, with the following activities:

Physics analysis

- Run 2 search h(125) $\rightarrow$ aa $\rightarrow$ $\tau\tau$bb, and combination with $\mu\mu$bb [2402.13358]

Machine learning

- CICADA: Level-1 anomaly detection trigger for model-agnostic new physics searches [CMS-DP-2023-086]
- Symbolic regression for nanosecond inference on FPGAs [2305.04099] [2401.09949]

Other technical [calorimeter layer-1 trigger subsystem]

- Software for data quality monitoring
- On-call operations

Collaboration services

- L3 co-convener of SUS Monte-Carlo & Interpretation (09/23-)
- Contact person of HIG Monte-Carlo: subgroup contact (01/21-08/22) and overall contact (09/22-08/23)

Agenda

I. Search for exotic Higgs boson decays h(125) → aa → $\tau\tau$bb / $\mu\mu$bb

II. Anomaly detection trigger at the Level-1 for model-agnostic new physics searches

III. Symbolic regression for hardware-efficient ML inference with nanosecond-latency

IV. A slide on the technical: calorimeter layer-1 trigger subsystem

Agenda

I. Search for exotic Higgs boson decays h(125) → aa → $\tau\tau$bb / $\mu\mu$bb

II: a lower-level improvement of I

II. Anomaly detection trigger at the Level-1 for model-agnostic new physics searches

III: a lower-level improvement of II

III. Symbolic regression for hardware-efficient ML inference with nanosecond-latency

IV. A slide on the technical: calorimeter layer-1 trigger subsystem

Analysis

I. Search for exotic Higgs boson decays $h(125) \rightarrow aa \rightarrow \tau\tau bb$ / $\mu\mu bb$

II. Ano … new physics searches

III. Sym … th nanosecond-latency

IV. A sli … em

CMS-HIG-22-007

CERN-EP-2023-284
2024/02/22

arXiv:2402.13358v1 [hep-ex] 20 Feb 2024

Search for exotic decays of the Higgs boson to a pair of pseudoscalars in the $\mu\mu$bb and $\tau\tau$bb final states

The CMS Collaboration*

Abstract

A search for exotic decays of the Higgs boson (H) with a mass of 125 GeV to a pair of light pseudoscalars a_1 is performed in final states where one pseudoscalar decays to two b quarks and the other to a pair of muons or τ leptons. A data sample of proton-proton collisions at $\sqrt{s} = 13$ TeV corresponding to an integrated luminosity of 138 fb^{-1} recorded with the CMS detector is analyzed. No statistically significant excess is observed over the standard model backgrounds. Upper limits are set at 95% confidence level (CL) on the Higgs boson branching fraction to $\mu\mu$bb and to $\tau\tau$bb, via a pair of a_1s. The limits depend on the pseudoscalar mass m_{a_1} and are observed to be in the range $(0.17–3.3)\times10^{-4}$ and $(1.7–7.7)\times10^{-2}$ in the $\mu\mu$bb and $\tau\tau$bb final states, respectively. In the framework of models with two Higgs doublets and a complex scalar singlet (2HDM+S), the results of the two final states are combined to determine model-independent upper limits on the branching fraction $\mathcal{B}(\mathrm{H} \rightarrow a_1 a_1 \rightarrow \ell\ell bb)$ at 95% CL, with ℓ being a muon or a τ lepton. For different types of 2HDM+S, upper bounds on the branching fraction $\mathcal{B}(\mathrm{H} \rightarrow a_1 a_1)$ are extracted from the combination of the two channels. In most of the Type II 2HDM+S parameter space, $\mathcal{B}(\mathrm{H} \rightarrow a_1 a_1)$ values above 0.23 are excluded at 95% CL for m_{a_1} values between 15 and 60 GeV.

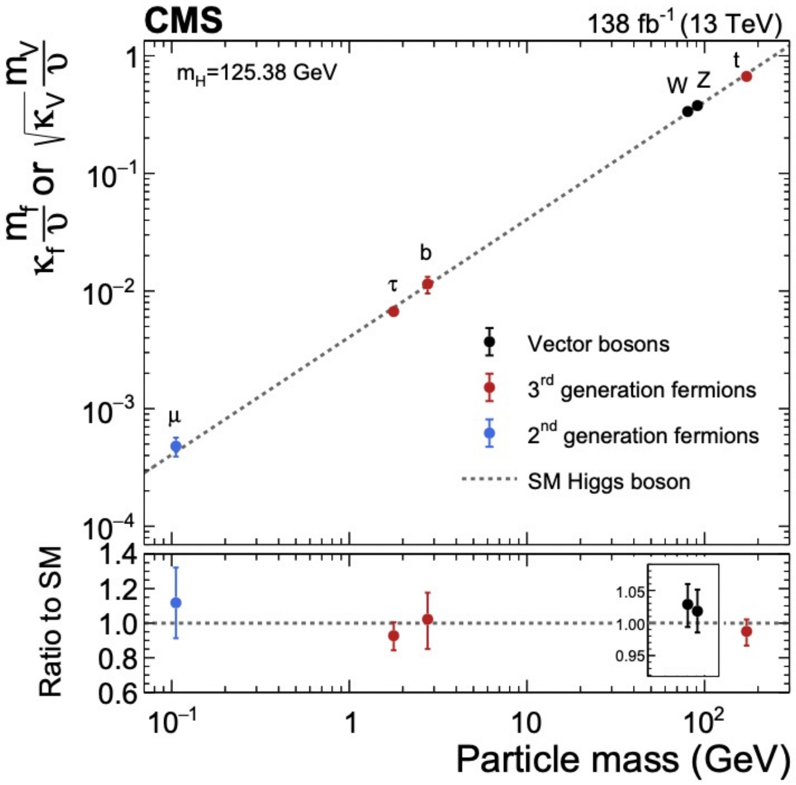
Submitted to the European Physical Journal C

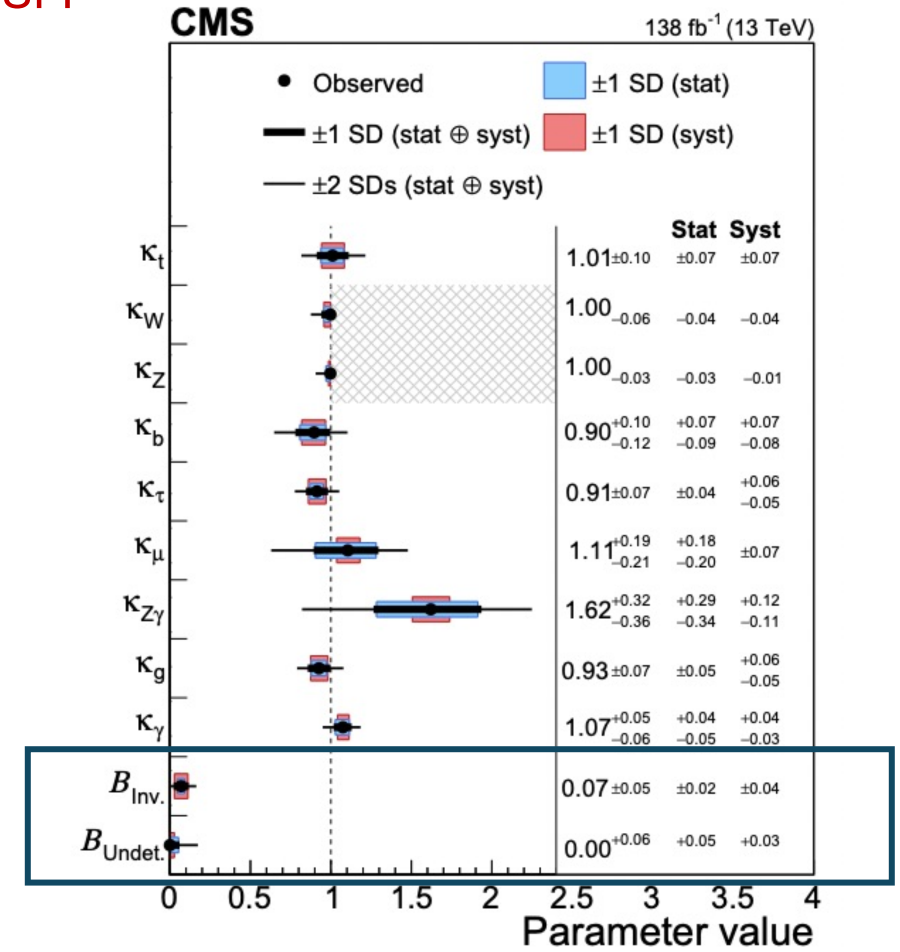
- **My contribution**
  - Main analyzer of the $\tau\tau$bb channel
  - Combination with $\mu\mu$bb

- [arXiv:2402.13358], submitted to EPJC

Analysis. Where we stand now in the scalar sector?

- A SM-like Higgs boson h(125) discovered in 2012
- 10 years since then, collected data corresponding to 30x more Higgs production allowed more precise measurements of the particle
- Despite good compatibility with the SM prediction so far, there is still room for exotic Higgs decays to new particles beyond the SM





Nature 607, 60-68 (2022)

Analysis. BSM extended scalar sector: 2HDM+S

2HDM+S: Two Higgs doublet (Φ_1, Φ_2) model + a singlet S

Not ruled out by all the LHC measurements published so far!

- 7 physical states: (h, H, A, $H^{\pm}$) + (s, a)
- One of the scalars could be SM-like $\mathrm{h} = \mathrm{h}(125)$
- The pseudoscalar, a, could be light enough to allow $\mathrm{h}(125) \rightarrow aa \rightarrow \mathrm{SM}$
- Four coupling types of 2HDM+S can forbid tree-level flavor-changing-neutral-current

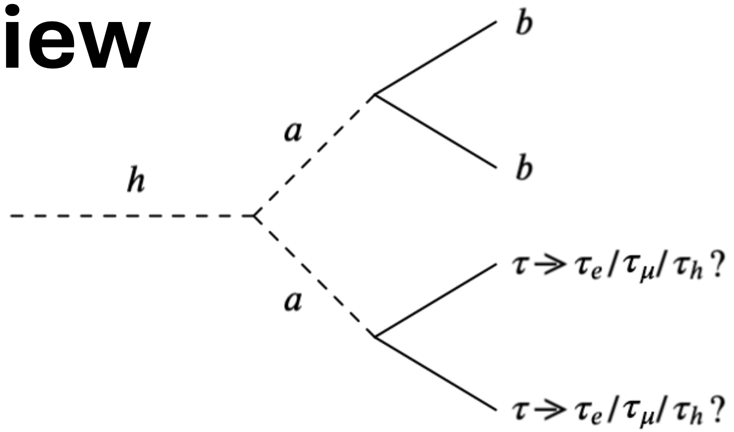
	Type-I	Type-II	Type-III	Type-IV
Charged lepton	Φ_1	Φ_2	Φ_2	Φ_1
Up-type quark	Φ_1	Φ_1	Φ_1	Φ_1
Down-type quark	Φ_1	Φ_2	Φ_1	Φ_2

Why exotic Higgs decays and 2HDM+S?

- Tiny natural width of the SM Higgs boson → small coupling to a light state can lead to sizeable signature
- Additional portal to the hidden sector
- Contains NMSSM as a particular case (type-II)
- Provides possible dark matter candidate

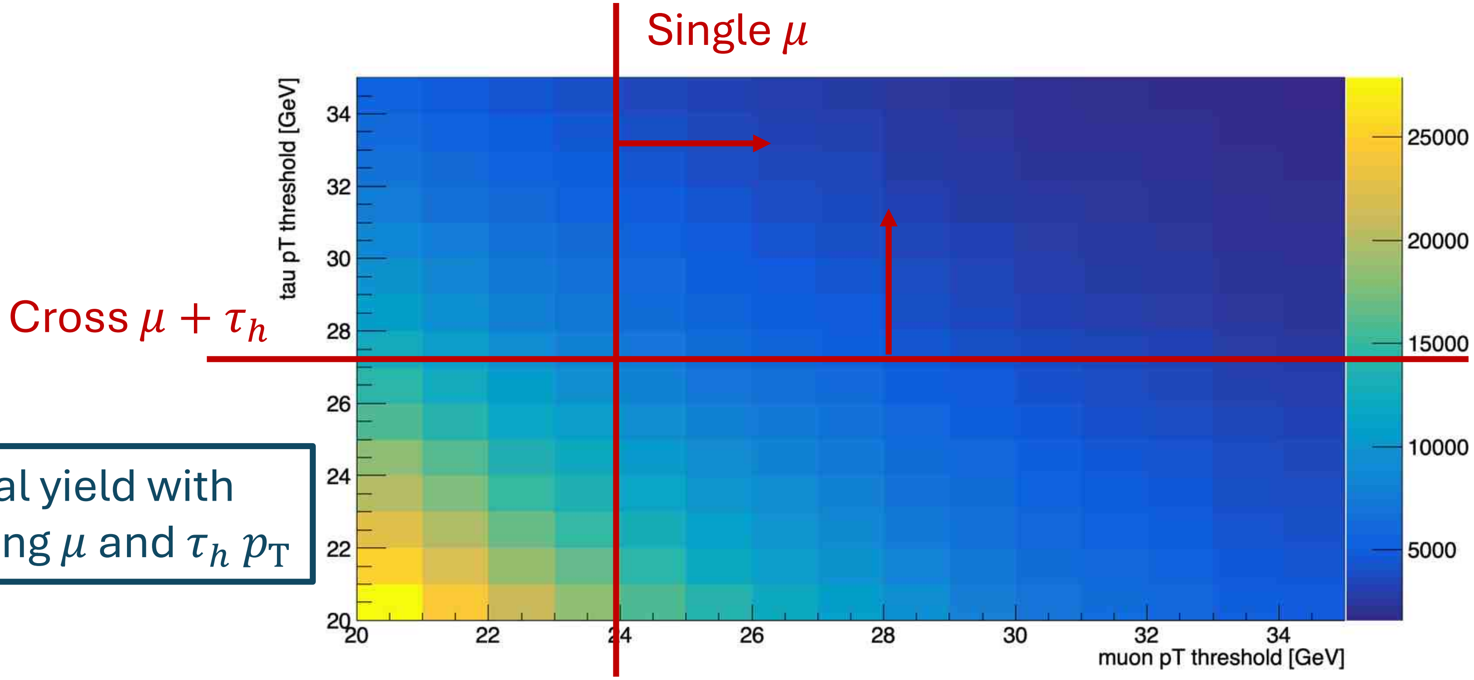
Analysis. Analysis overview

In 2HDM+S, the decay $\mathrm{h}(125) \to aa$ can be possible if $m_a < m_h/2$
- Probe the final state with 2 b quarks and 2 tau leptons: large $\mathrm{B}(a \to \mathrm{ff})$
- Target 3 $\tau\tau$ final states: $\boldsymbol{e\mu/e\tau_h/\mu\tau_h}$

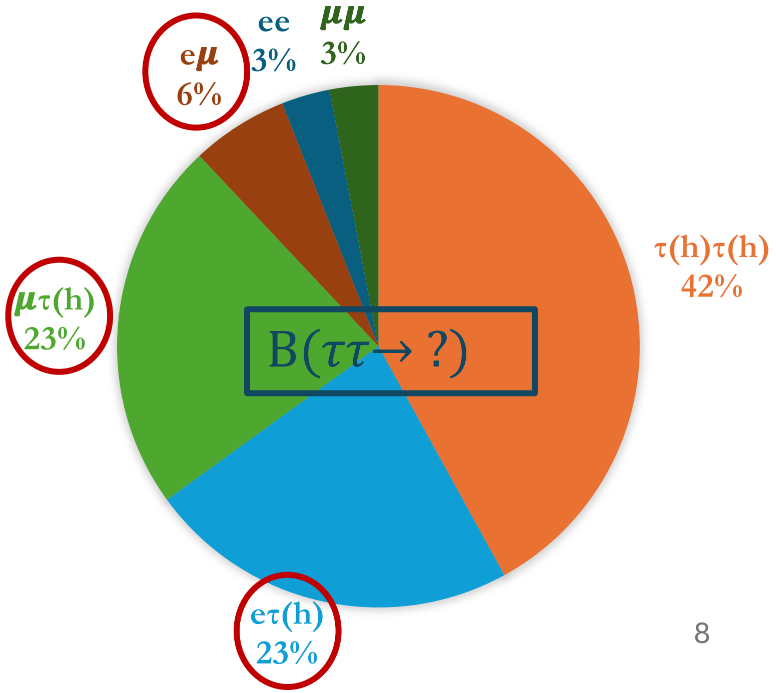


arXiv:2402.13358

Challenges
- 4 soft objects in the final state
- Difficult signature to trigger on
- Online triggers already removed > 90% of the signal!

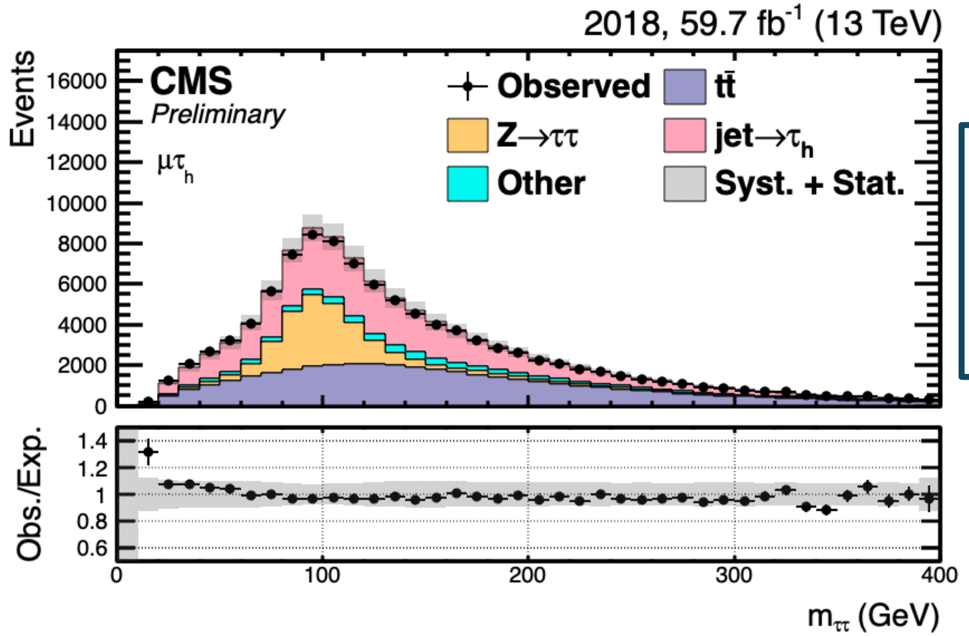


Signal yield with varying μ and τ_h p_T

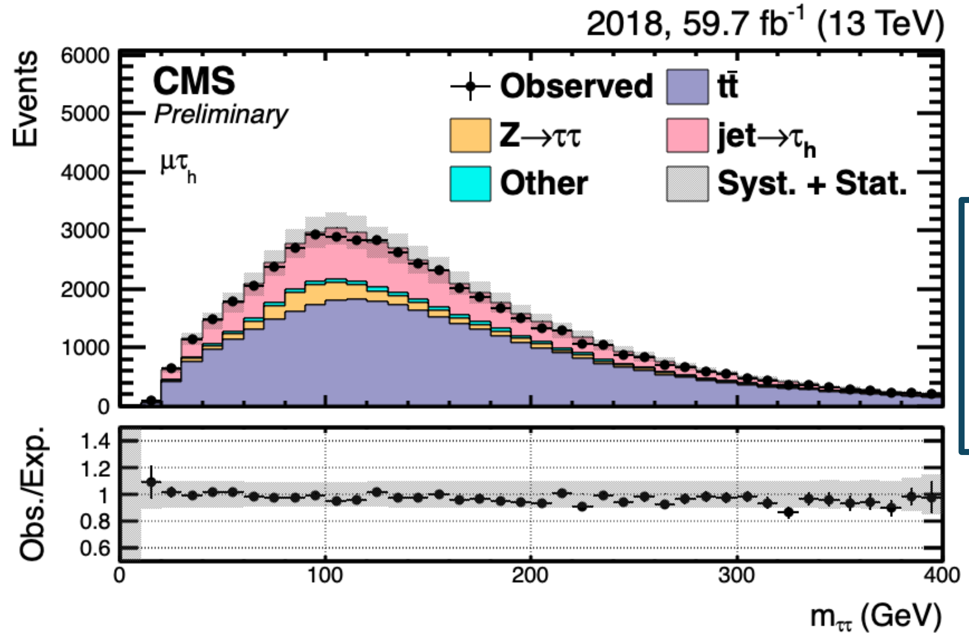


Analysis. Background estimation

- **Top quark pair $t\bar{t}$**
  - ➢ Simulation
- **Drell-Yan Z → $\tau\tau$**
  - ➢ Embedding technique, select $\mu\mu$ events from data and replace with simulated τ leptons arXiv:1903.01216
- **Jet → τ_h fakes**
  - ➢ Data-driven method, measure fake rate from $Z(\mu\mu)$ + jets data events and apply to anti-isolated τ_h
- **Other minor background (single top, diboson, W+jets, $Z(ee/\mu\mu)$+jets, SM Higgs)**
  - ➢ Simulation



Background composition in $m_{\tau\tau}$ distribution (= 1 b tagged jet)



Background composition in $m_{\tau\tau}$ distribution (> 1 b tagged jets)

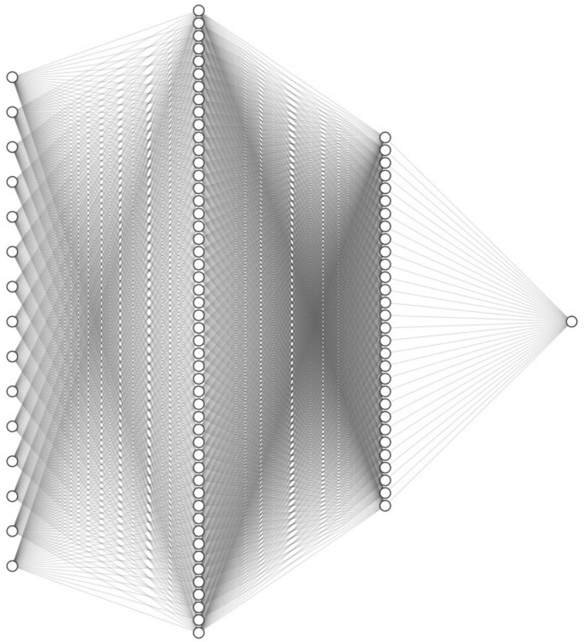
arXiv:2402.13358

Analysis. Event categorization using DNN

arXiv:2402.13358

- Two basic categories based on number of b tagged jets (= 1 or > 1)
- Each category is further optimized by a neural network (NN) classifier trained to separate signal from background

1bNN/2bNN



1bNN/2bNN trained with ~15/30 kinematic features

$m_{\tau\tau}$ distribution

Baseline selection

$m_{\tau\tau}$ from events with 1 b-jet

$m_{\tau\tau}$ from events with 2 b-jets

1bNN

2bNN

$m_{\tau\tau}$ 1 b-jet SR1

$m_{\tau\tau}$ 1 b-jet SR2

$m_{\tau\tau}$ 1 b-jet SR3

$m_{\tau\tau}$ 1 b-jet CR

$m_{\tau\tau}$ 2 b-jet SR1

$m_{\tau\tau}$ 2 b-jet SR2

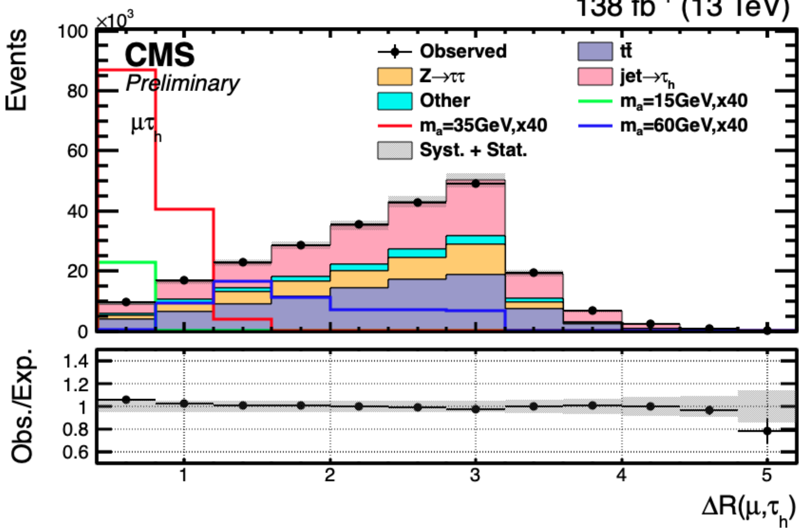
$m_{\tau\tau}$ 2 b-jet CR

Binned likelihood fit to $m_{\tau\tau}$ for signal extraction

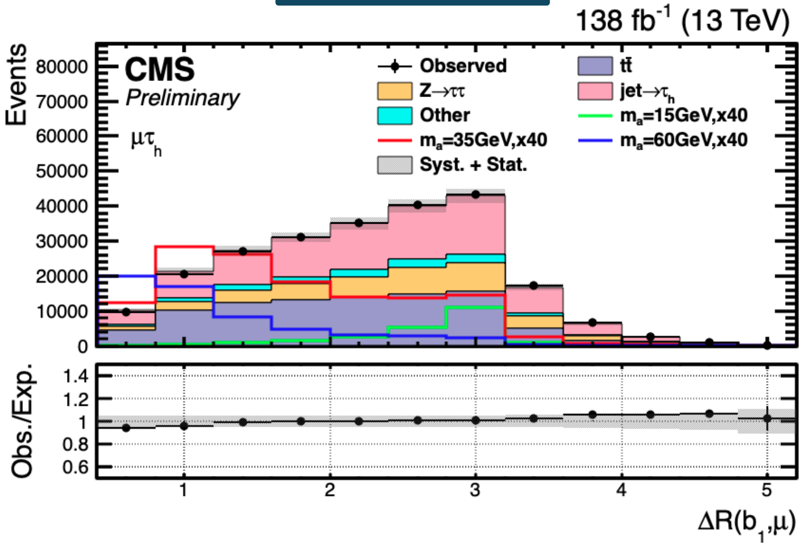
Analysis. Input features to DNN

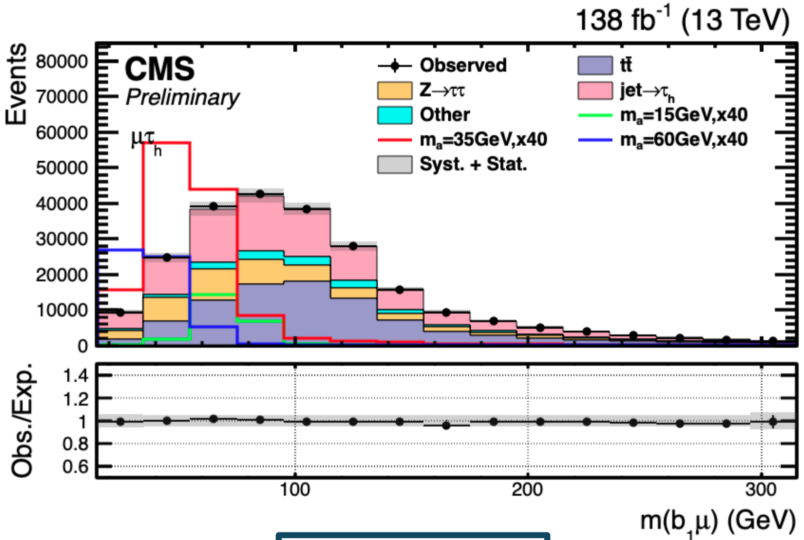
arXiv:2402.13358

$\Delta R(\mu, \tau_h)$

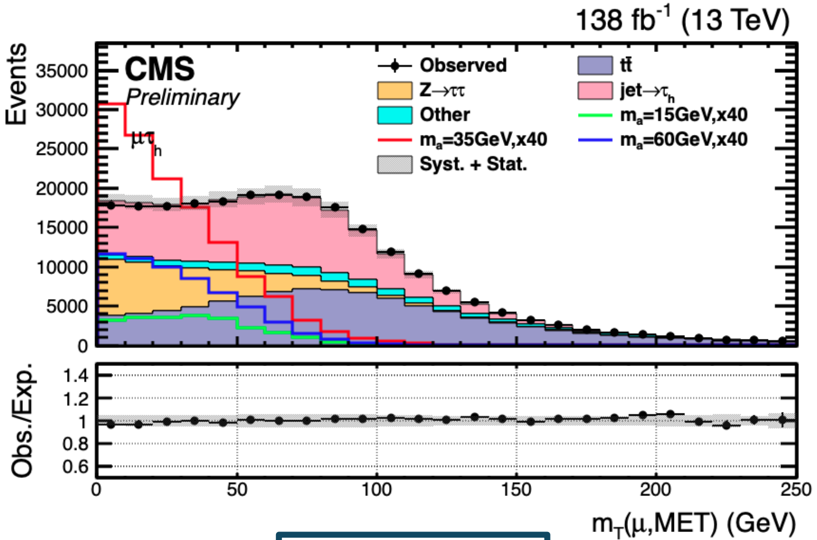


$\Delta R(b_1, \mu)$

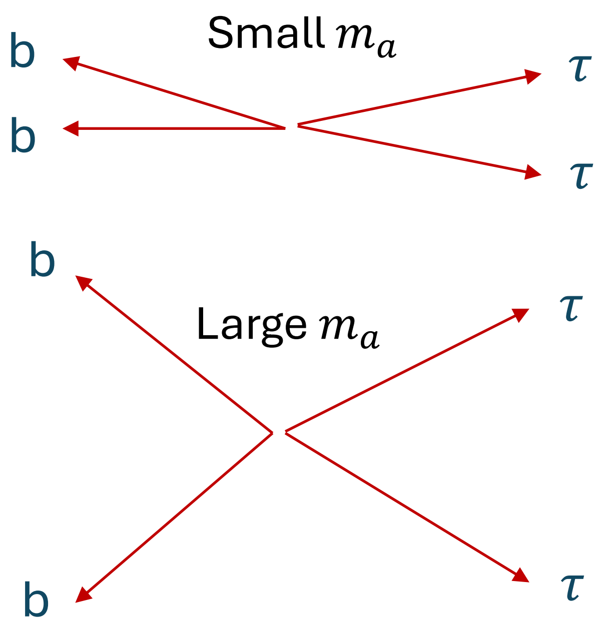




$m(b_1\mu)$



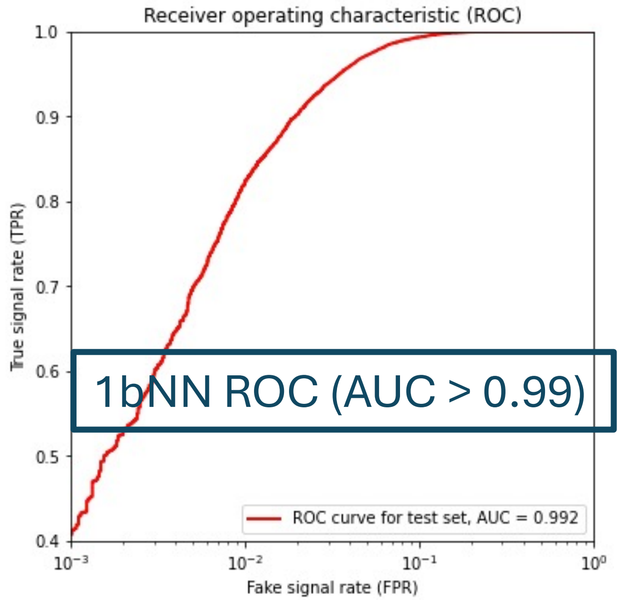
$m_{\mathrm{T}}(\mu, \mathrm{MET})$

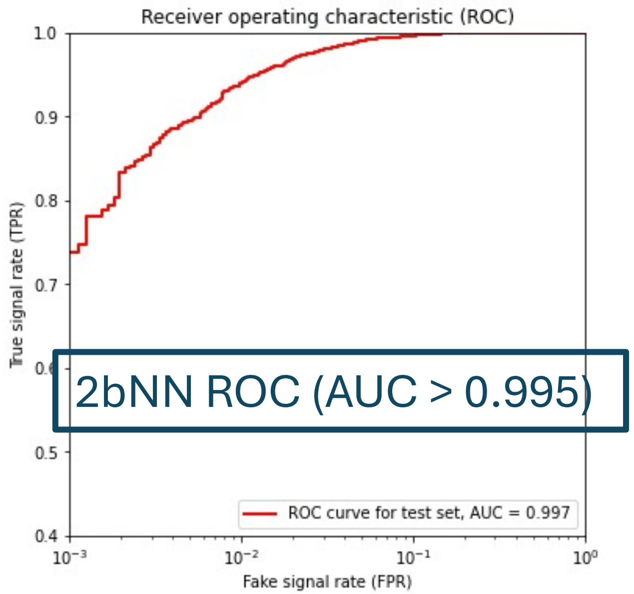


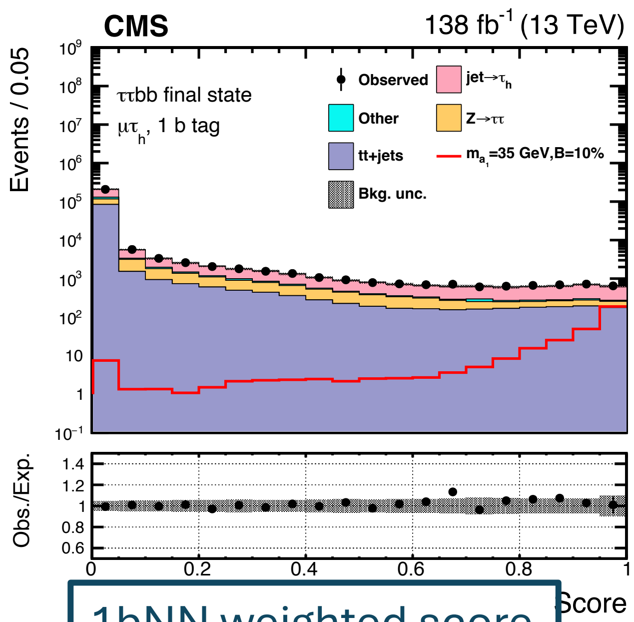
Other discriminating variables include

- $m(\mathrm{b}_1\tau_h)$, $m(\mathrm{b}_1\tau\tau)$, … except $m(\tau\tau)$
- $\Delta R(\mathrm{b}_1, \tau_h)$, $\Delta R(\mathrm{bb}, \tau\tau)$, …
- p_{T} of μ, τ_h, $\tau\tau$, b_1, …

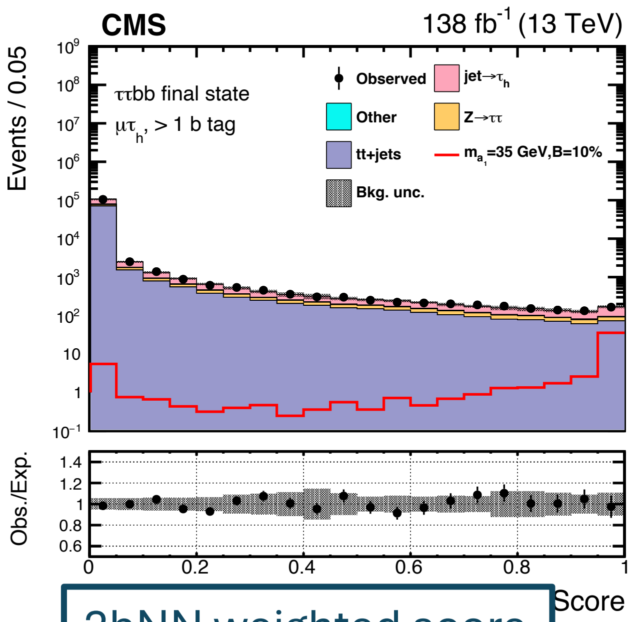
Analysis. DNN classifier performance







1bNN weighted score



2bNN weighted score

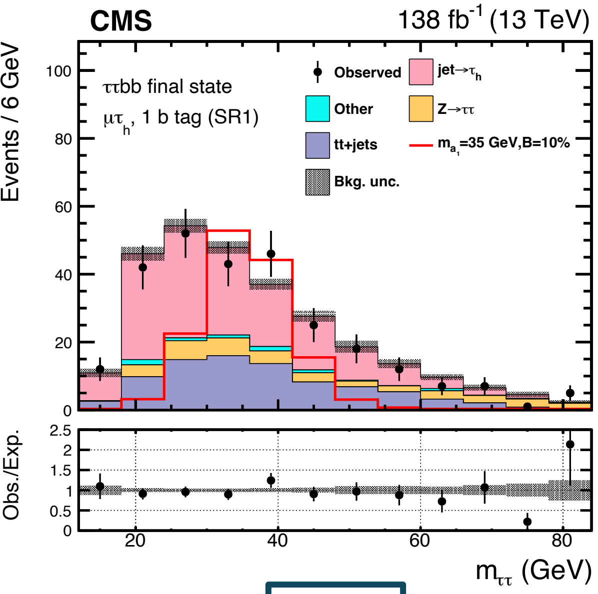
arXiv:2402.13358

$\mu\tau_h$ final state

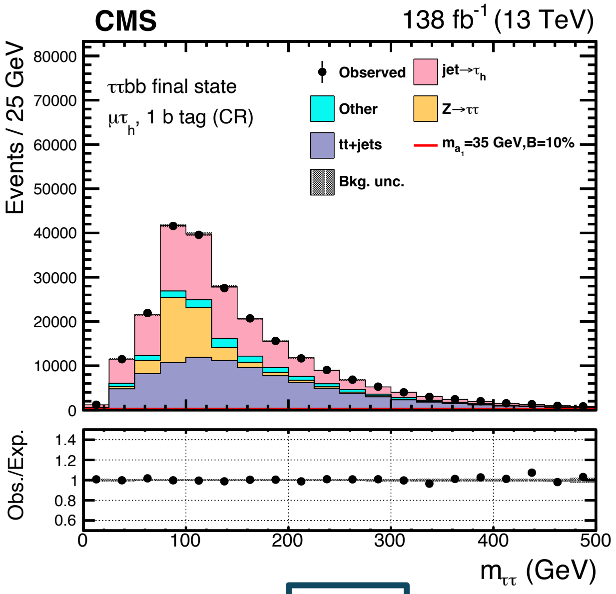
- Excellent separation power
- In addition to the "= 1 b-tag" and "> 1 b-tag" bins, further categorization is based on the classifier score
- Category thresholds on the score determined by a grid scan in the signal sensitivity
- Other final states in backup

Analysis. Results

$\mu\tau_h$ "1 b-tag" $m_{\tau\tau}$

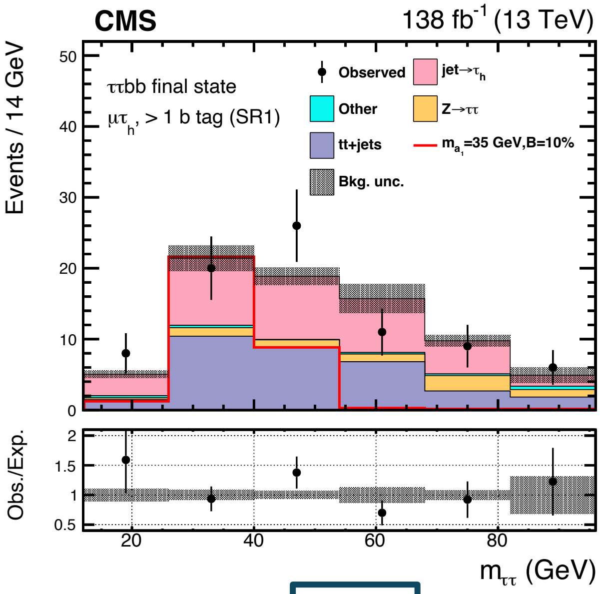


SR1

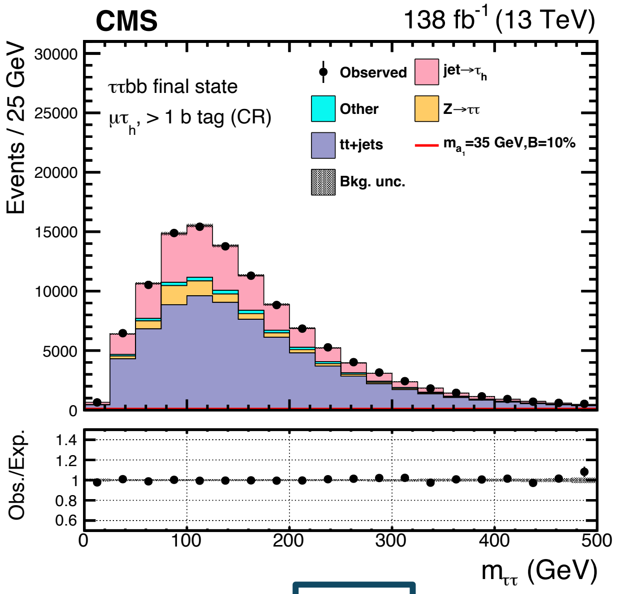


CR

$\mu\tau_h$ "> 1 b-tag" $m_{\tau\tau}$



SR1

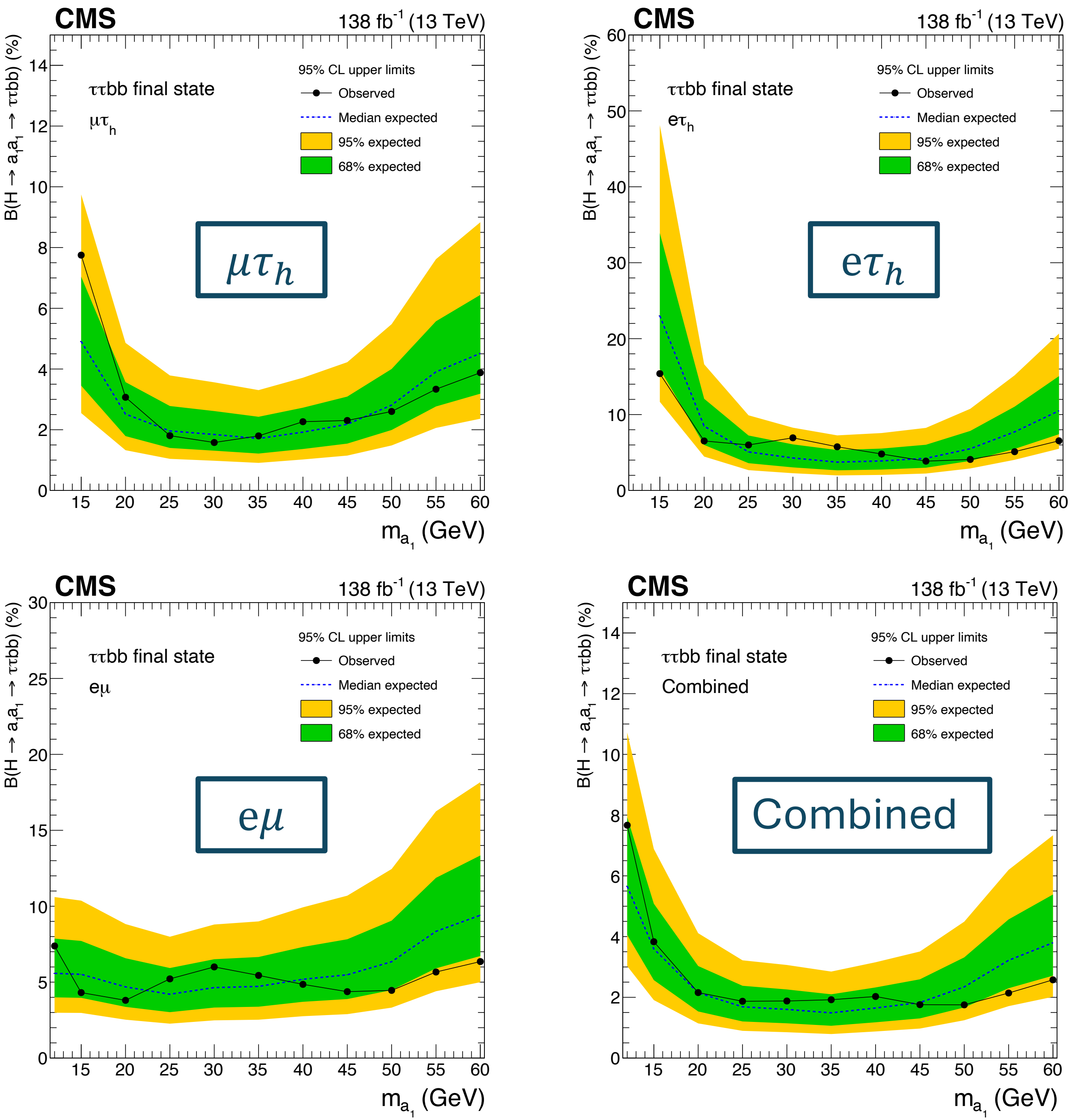


CR

- 7 optimized $m_{\tau\tau}$ distributions in the $\mu\tau_h$ final state
  - "1 b-tag" category: 3 SR + 1 CR
  - "> 1 b-tag" category: 2 SR + 1 CR
- Data agreed with SM prediction across categories, no significant excess in data observed
- All other categories and final states in backup

Analysis. Exclusion limits

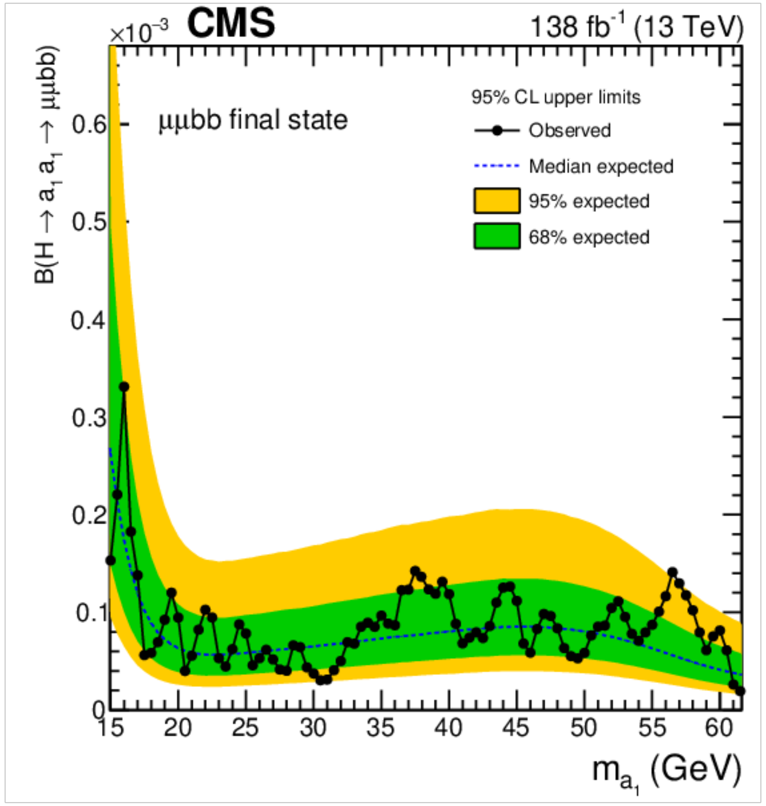
arXiv:2402.13358



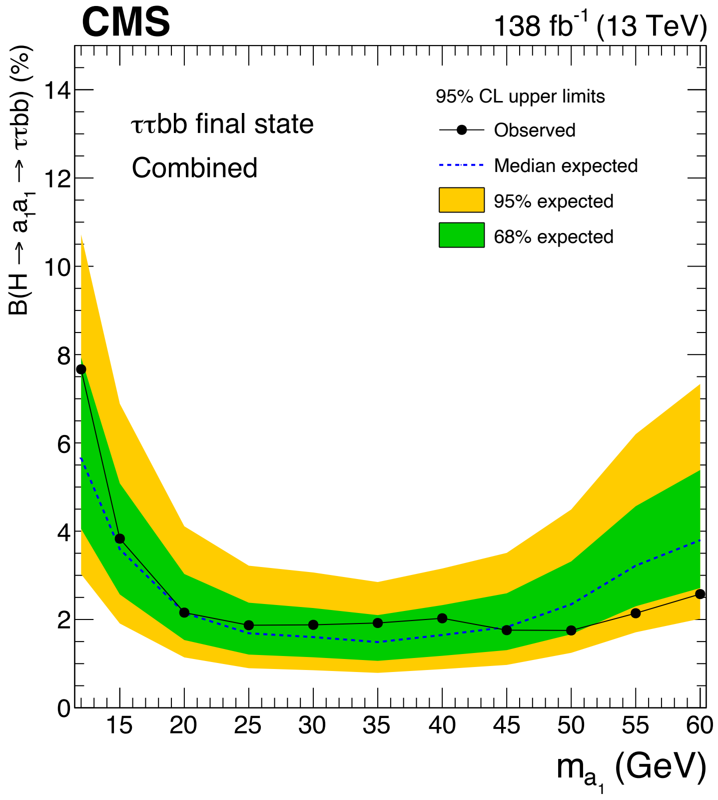
- Data agreed with SM prediction, no significant excess observed
- Model-independent exclusion limits set at 95% CL on the signal strength $\mathrm{B}(\mathrm{h} \rightarrow aa \rightarrow \mathrm{bb}\tau\tau)$
- Sensitivity improved beyond the increased luminosity, compared to the 2016 results [arXiv:1805.10191]
  - Largely due to the ML-based optimization (~50% more sensitive than the cut-based)

Analysis. Combination of $\tau\tau$bb and $\mu\mu$bb

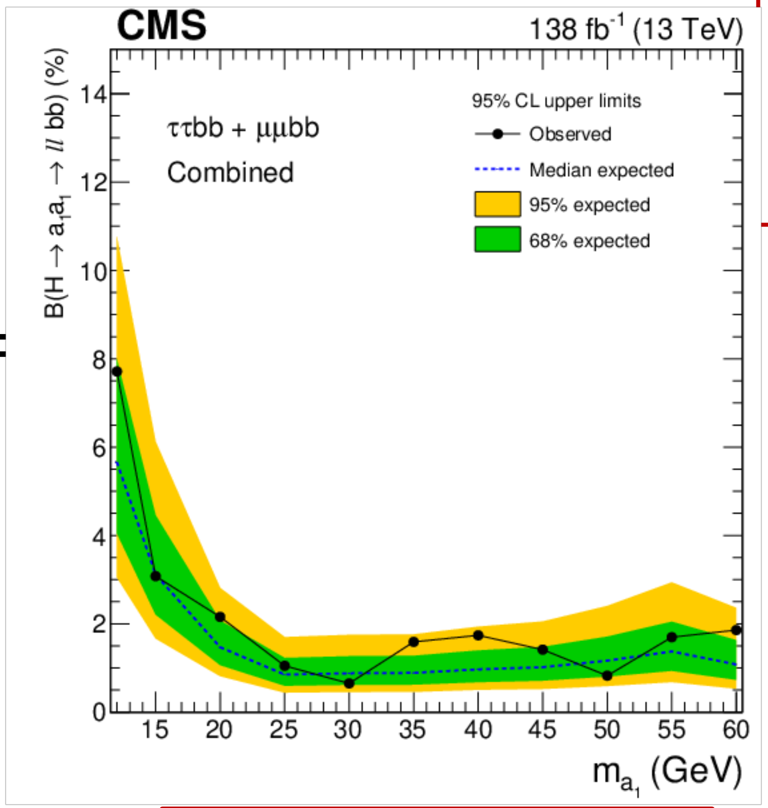
arXiv:2402.13358



+



=



95% CL limits on $B(\mathrm{h} \to aa \to \mu\mu\mathrm{bb})$

95% CL limits on $B(\mathrm{h} \to aa \to \tau\tau\mathrm{bb})$

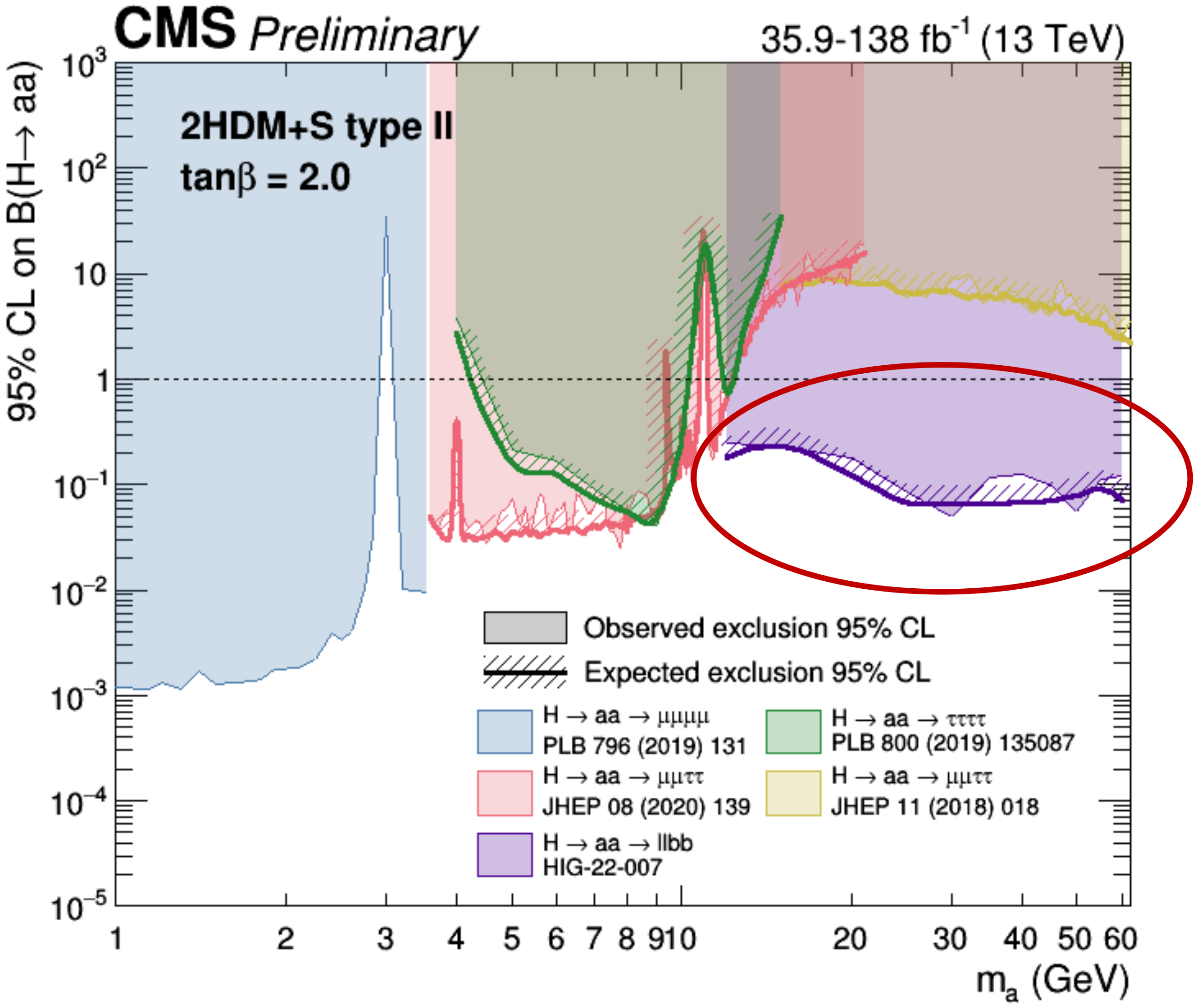
95% CL limits on $B(\mathrm{h} \to aa \to ll\mathrm{bb})$ in the 2HDM+S scenario

- Sensitivity is mainly driven by the $\tau\tau$bb channel

Analysis. Combination of $\tau\tau$bb and $\mu\mu$bb

arXiv:2402.13358

Summary2HDMSRun2

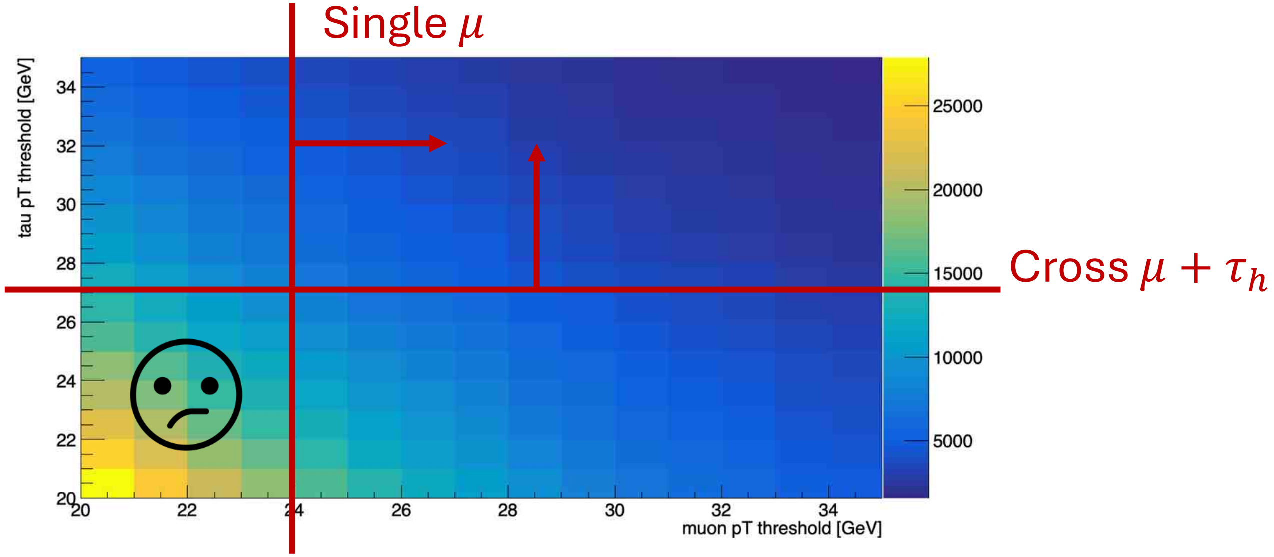


Summary of 95% CL limits on $B(\mathrm{h} \to aa)$ in the 2HDM+S type-II

Analysis. Takeaway

- More than 90% signal rejected already by the online triggers
- Many analyses are in a similar situation

Signal yield with varying μ and τ_h p_T



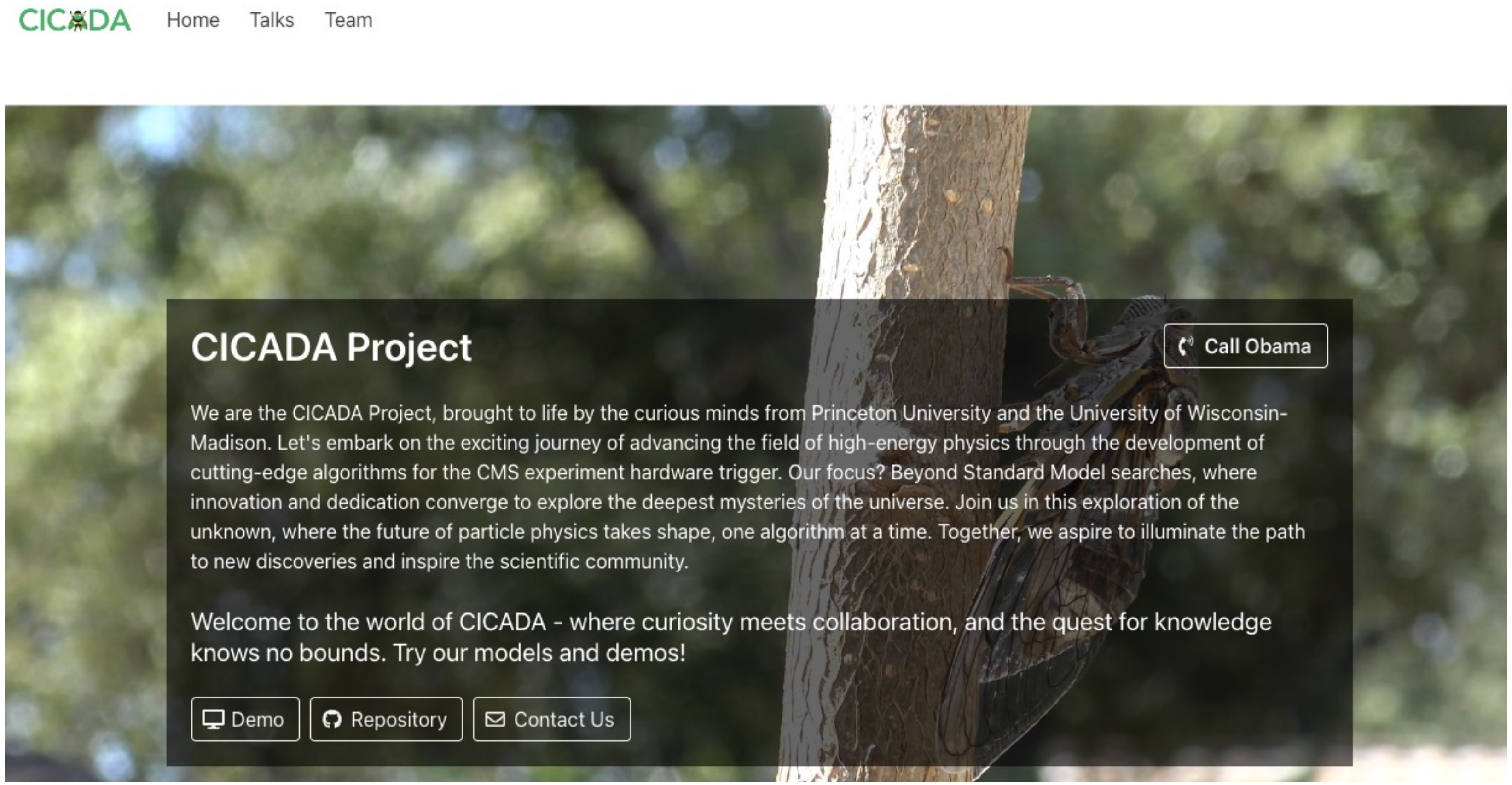
Can the situation be improved at a lower level?

➢ Next: anomaly detection trigger at the Level-1 for model-agnostic new physics searches

Anomaly trigger

I. Search for exotic Higgs boson decays h(125) → aa → $bb\tau\tau$ / $bb\mu\mu$

II. Anomaly detection trigger at the Level-1 for model-agnostic new physics searches



nd-latency

CICADA

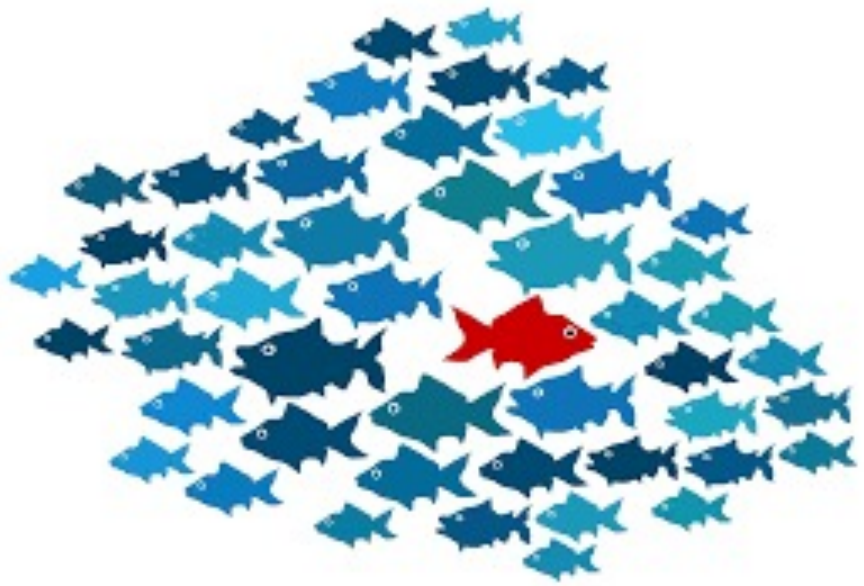
Calorimeter Image Convolutional Anomaly Detection Algorithm

- **My contribution**
  - Main developer of the ML model: architecture design and model training
- Public DP note [CMS-DP-2023-086]
- Deployment expected in the CMS L1 trigger menu during Run 3 in the second half of 2024

Anomaly trigger. Anomaly detection

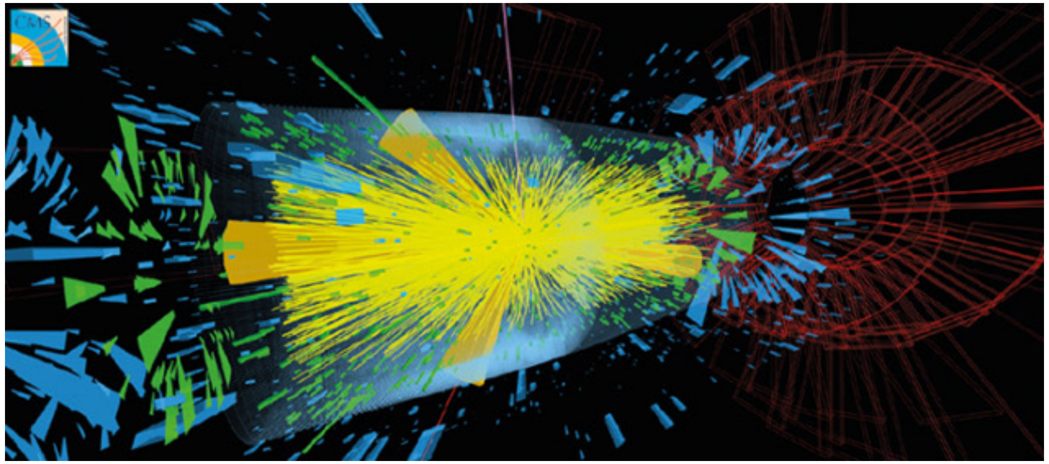
Anomaly detection using unsupervised learning

- Learning
  - Does not require labeled data
  - Assumes majority of data are normal
  - Learns the normal behavior from the data
- Inference
  - Identifies deviations from the learned normal pattern as anomalies
- No model assumptions on the anomalies



Anomaly trigger. In the context of LHC

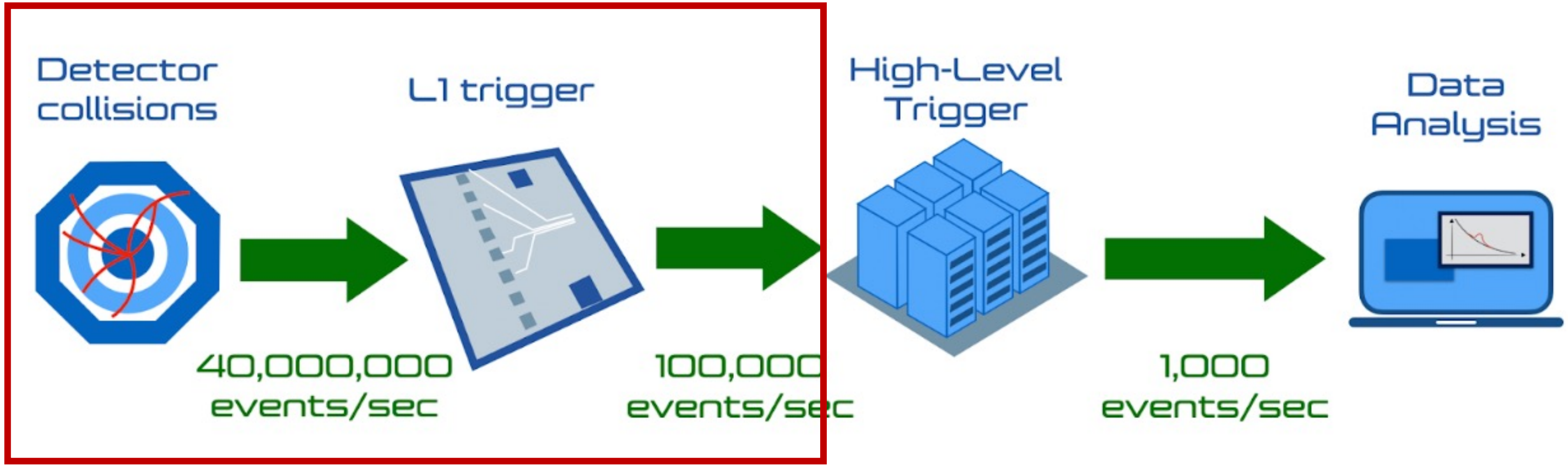
Model-agnostic searches for most rare/new physics at once!



- Normal events
  - Processes dominating the collisions
  - E.g. soft QCD
- Anomalous events
  - Processes with small xs or new physics
  - E.g. Higgs, SUSY, ...

Anomaly trigger. Why at L1 trigger?

First round of event selection ⟶ It is lost if it is lost!

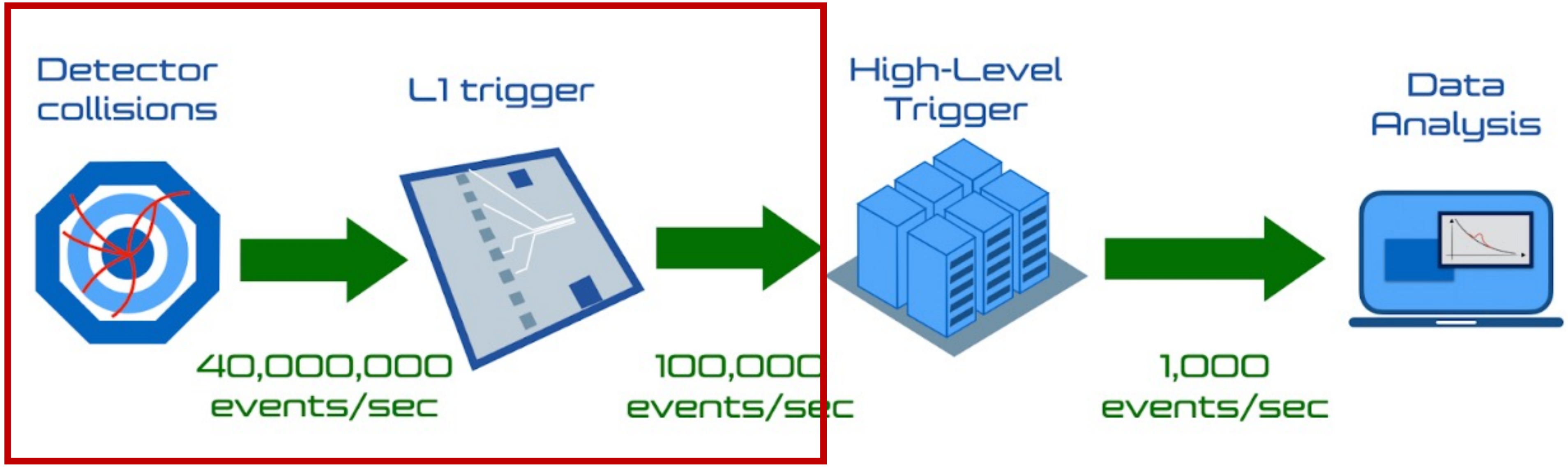


No BSM discovery at the LHC (yet!). Three possibilities:

- New physics not possible at the current LHC energy
- Not enough data collected
- Already hiding there, but we have been looking at the wrong places or using the wrong event selections

Anomaly trigger. Why anomaly trigger?

First round of event selection ⟶ It is lost if it is lost!

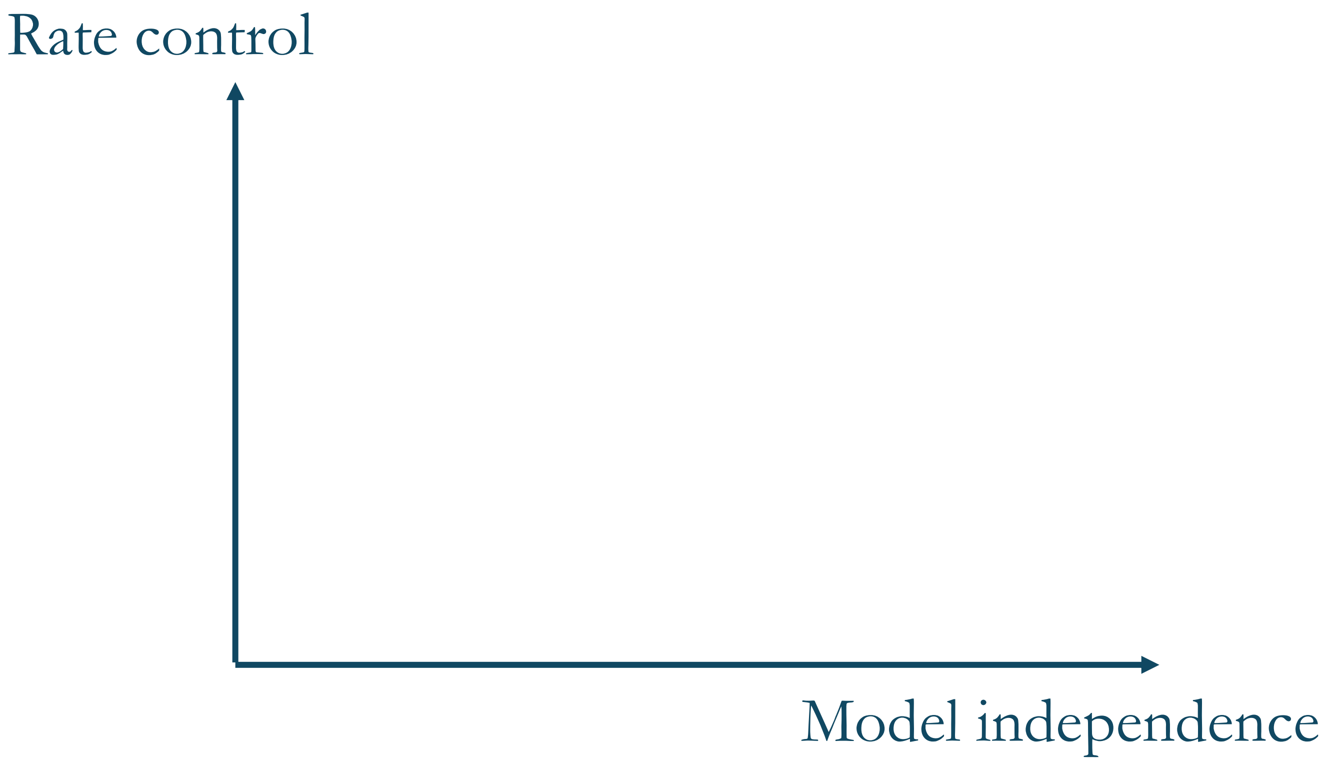


ML-based anomaly detection triggers

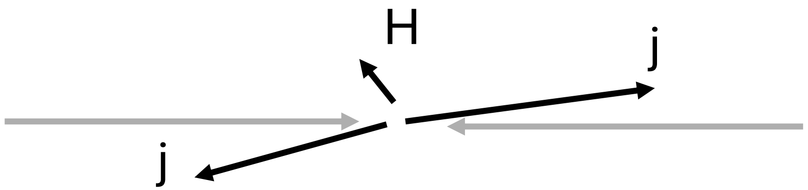
- Minimize human bias, completely data-driven
- ML can unearth unknown and complex correlation
- New physics searches in a model-agnostic way

Sometimes we don't know exactly what we are searching for...

Anomaly trigger. Anomaly vs. traditional triggers

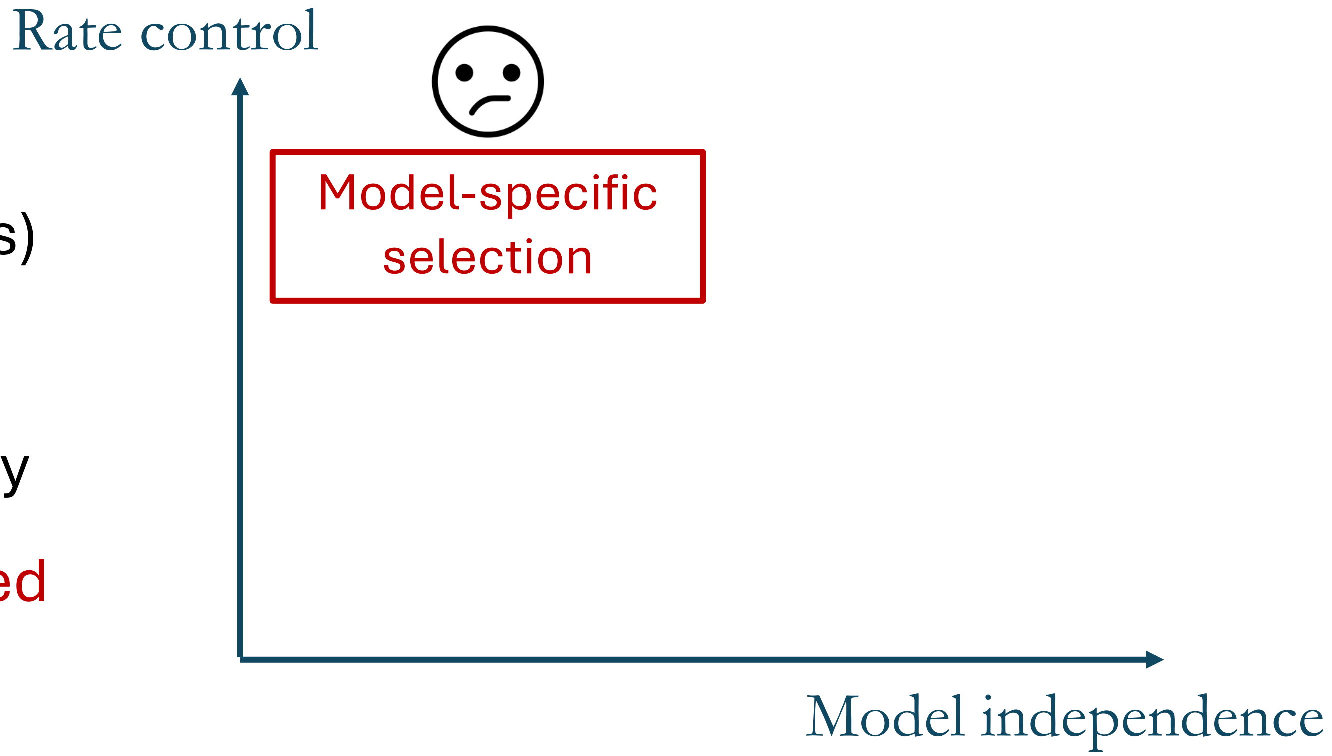


Anomaly trigger. Anomaly vs. traditional triggers



Model-specific triggers (e.g. VBF Higgs)

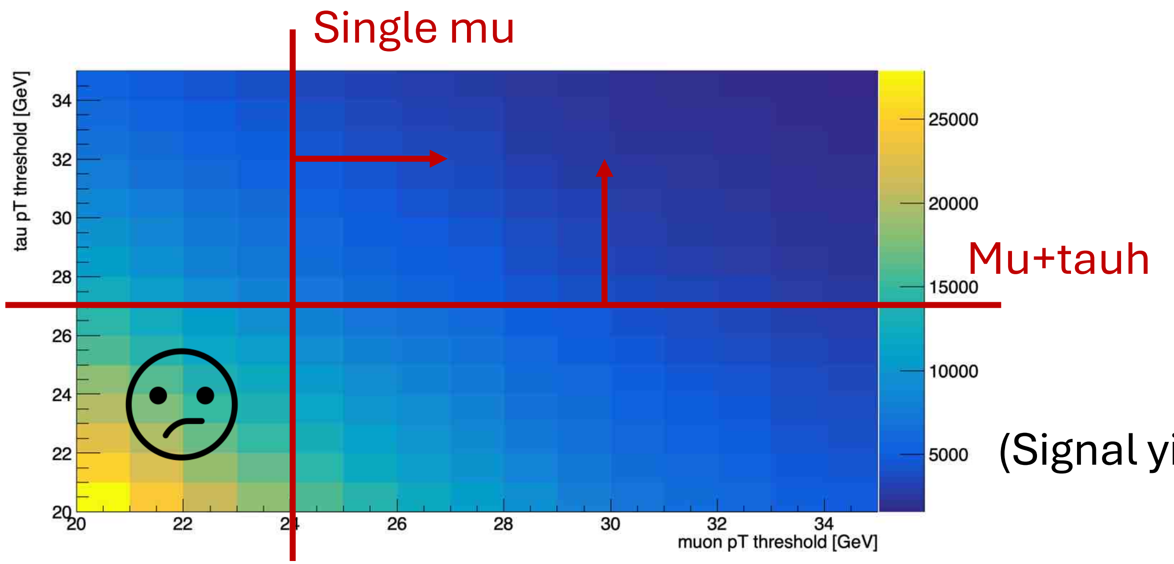
- ✓ Low rate
- ✓ Very sensitive to a particular topology
- Model-specific, cannot be generalized for other searches

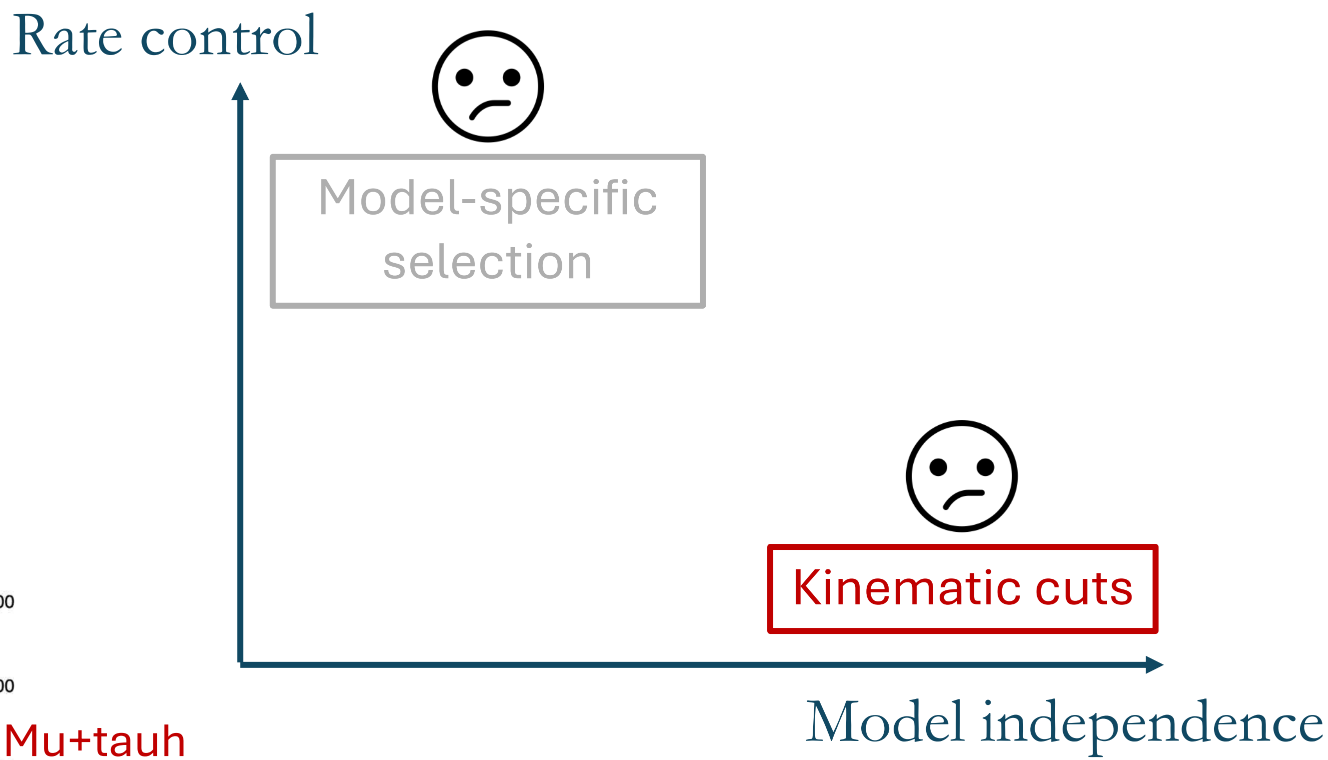


Anomaly trigger. Anomaly vs. traditional triggers

Object-based triggers (e.g. single muon)

- ✓ Fairly model-independent
- ✓ Can trigger on a wide range of signals
- Usually impose high threshold cuts in order to suppress rate, thus largely rejecting soft signals



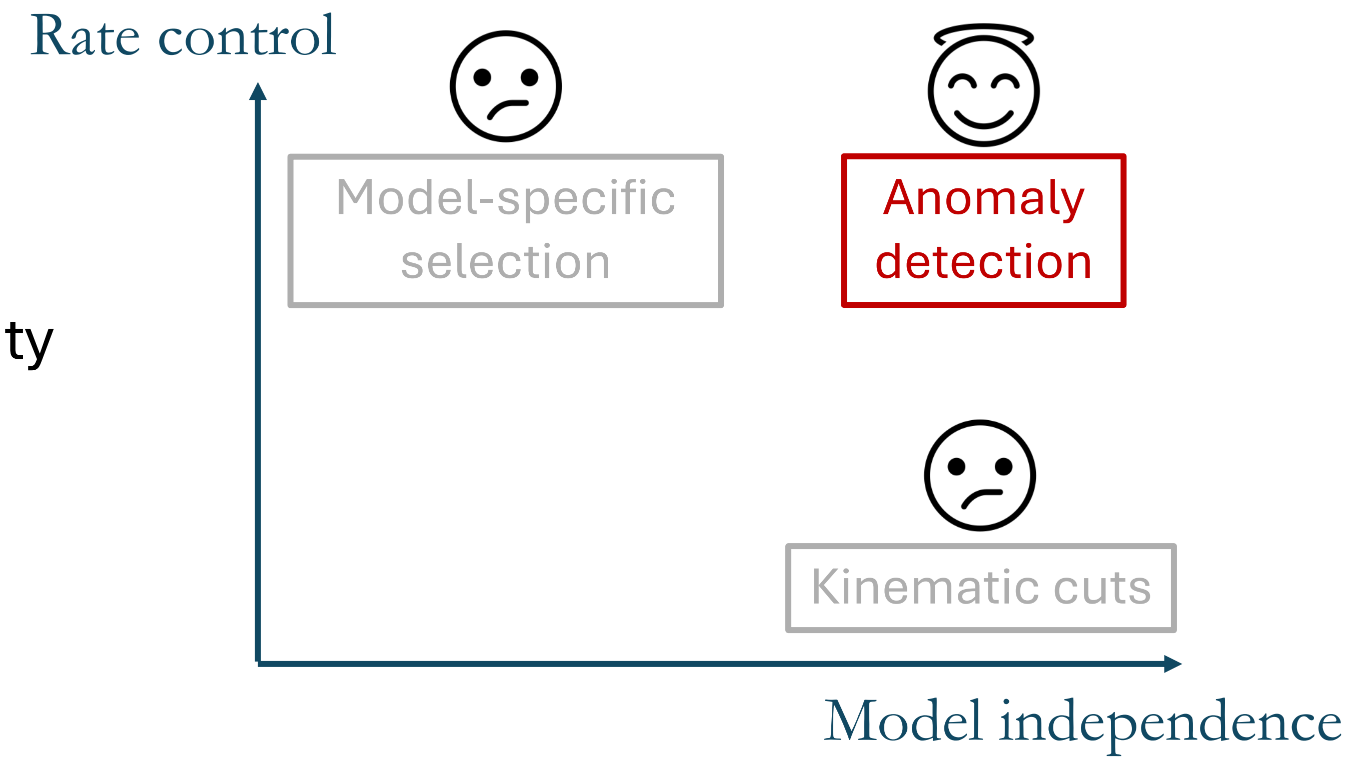


(Signal yield of h125 → aa → $\tau\tau$bb, before/after L1 cuts)

Anomaly trigger. Anomaly vs. traditional triggers

Anomaly triggers

- ✓ No model input needed
- ✓ Good rate control since high sensitivity



Anomaly trigger. CMS efforts

CICADA [CMS-DP-2023-086]

- Inputs from calorimeters
- Low-level information
  - Calorimeter energy deposits in trigger towers

AXOL1TL [CMS-DP-2023-079]

- Inputs from GT
- High-level information
  - pT/eta/phi of jet/muon/EG/MET

I will talk about CICADA in the following slides (similar story from AXOL1TL)

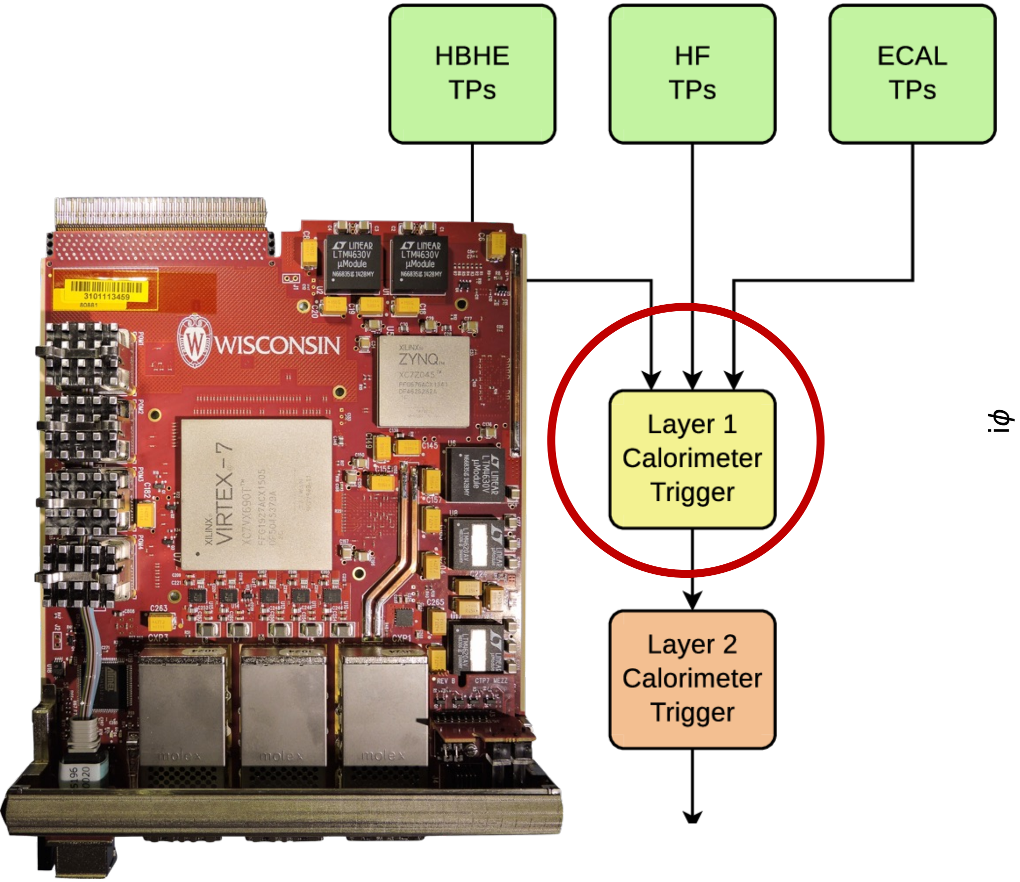
Anomaly trigger. CICADA

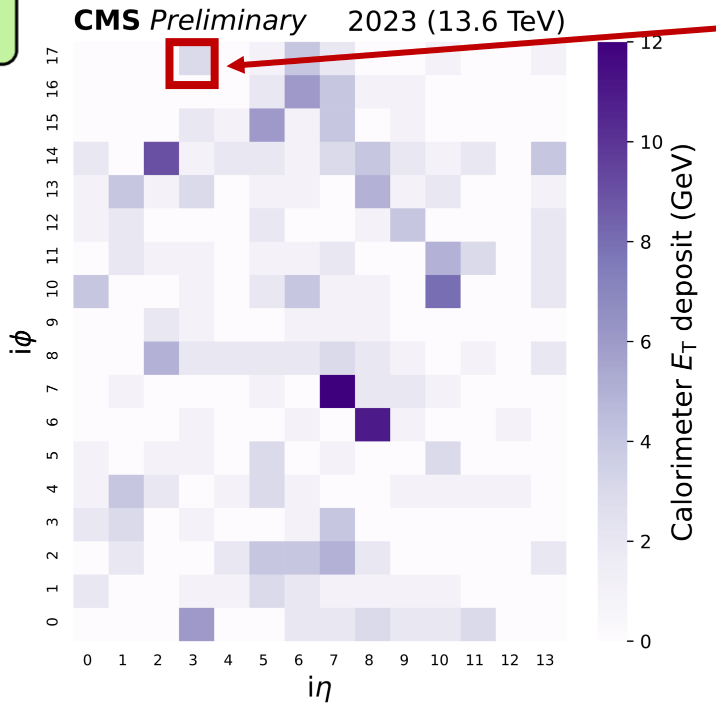
Calorimeter Image Convolutional Anomaly Detection Algorithm

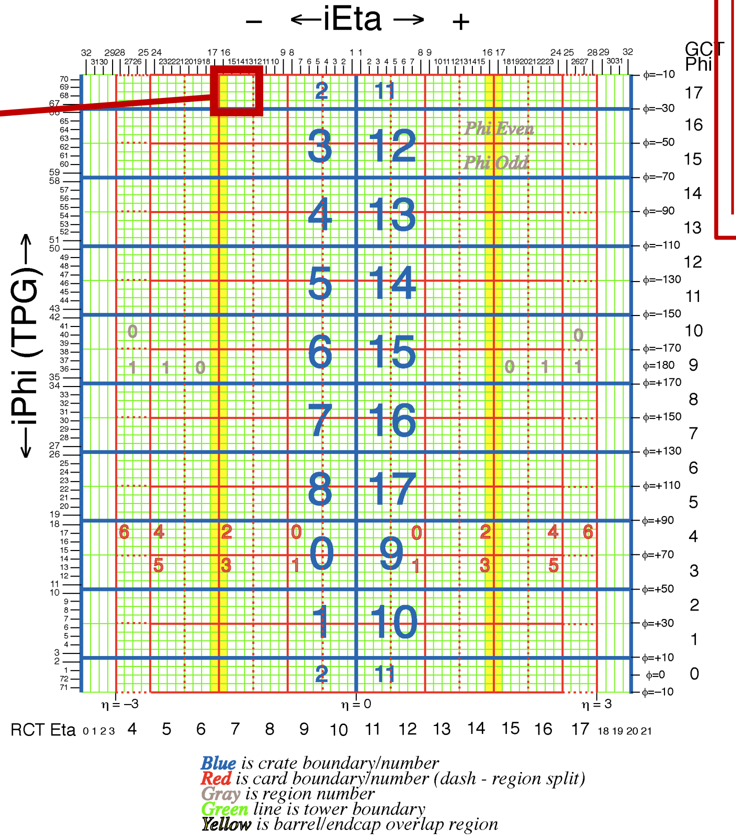
- Image-like inputs from calorimeter energy deposits at trigger towers
- CNN autoencoders
- Model compression with knowledge distillation and quantization
- HLS conversion using hls4ml

- CMS-DP-2023-086
- https://cicada.web.cern.ch/

Anomaly trigger. Model inputs







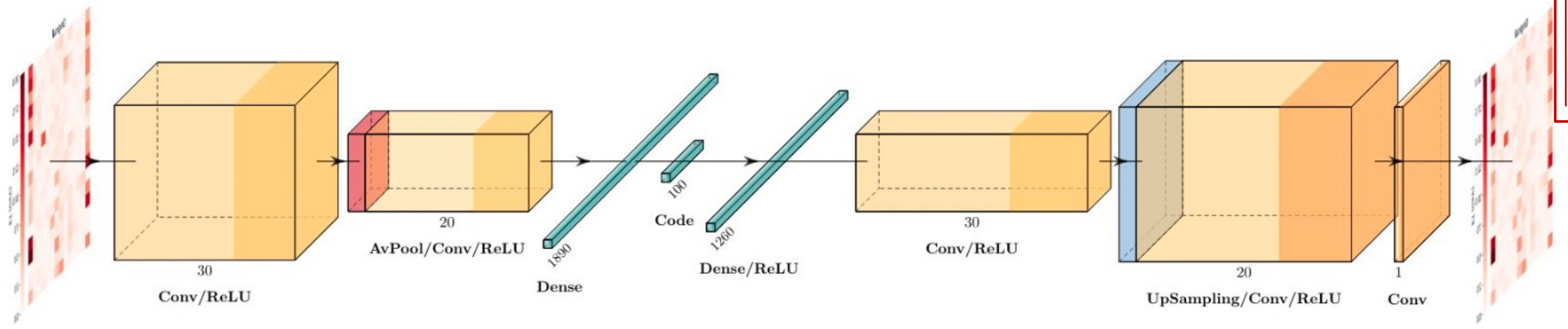
CMS-DP-2023-086

Inputs from CaloLayer-1

- ECAL+HCAL energy deposits in 18Phi x 14Eta = 252 trigger regions
- Low-level information, independent of jet reconstruction etc.

Anomaly trigger. Model architecture

CMS-DP-2023-086



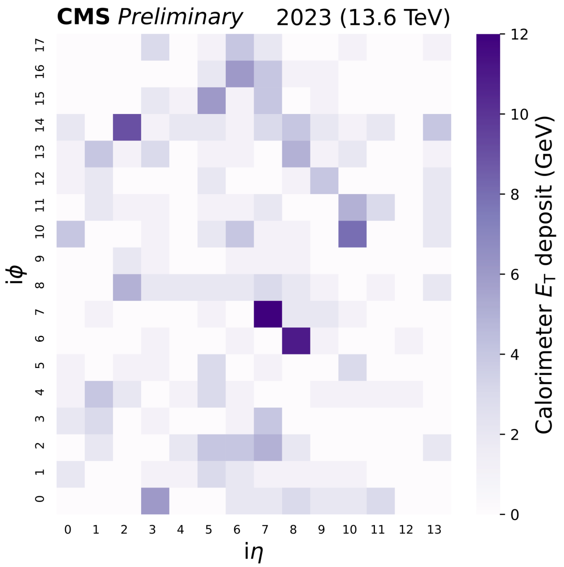
Autoencoder-based anomaly detection

- Input is 2D tensor from the calorimeter region energy
- Encoder and decoder are CNN
- Unsupervised learning
  - Train only on ZeroBias data to learn input reconstruction

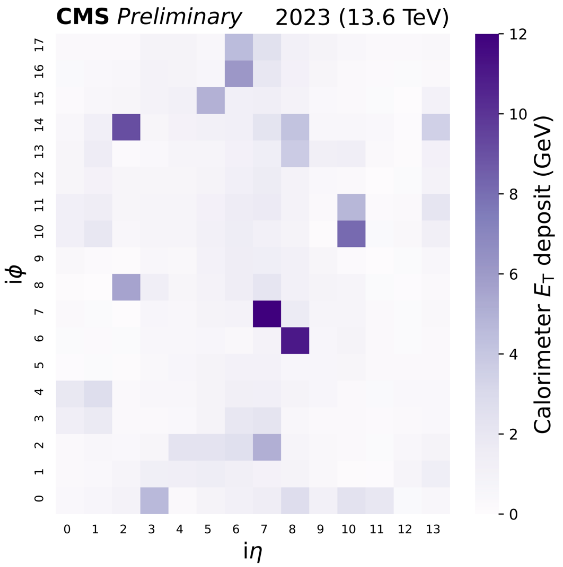
Anomaly trigger. Model output expectation

CMS-DP-2023-086

Single ZB data event

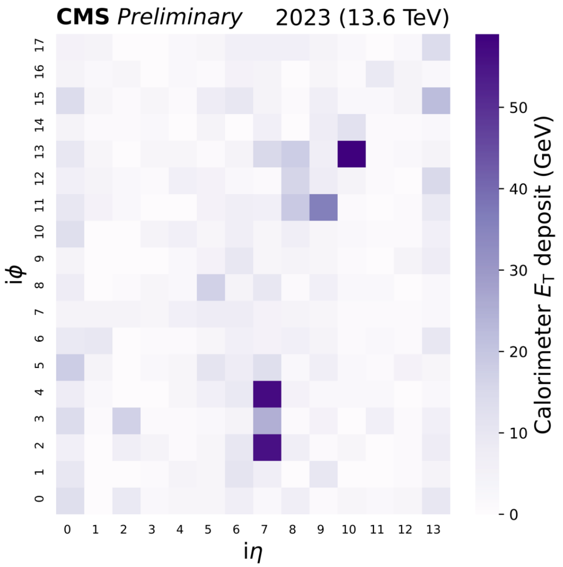


Input

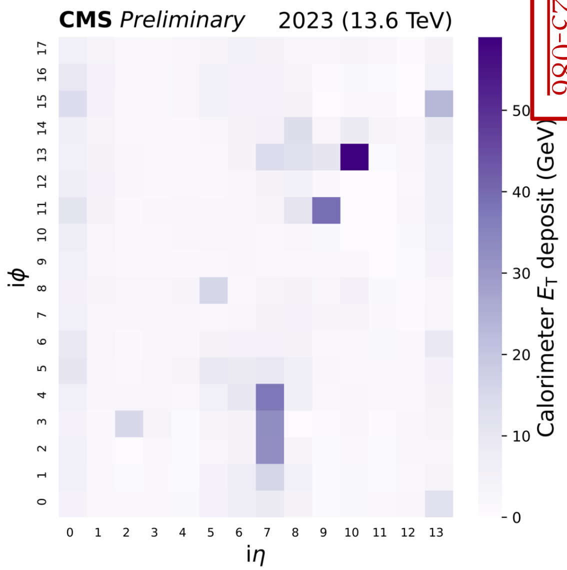


Output

Single BSM MC event



Input



Output

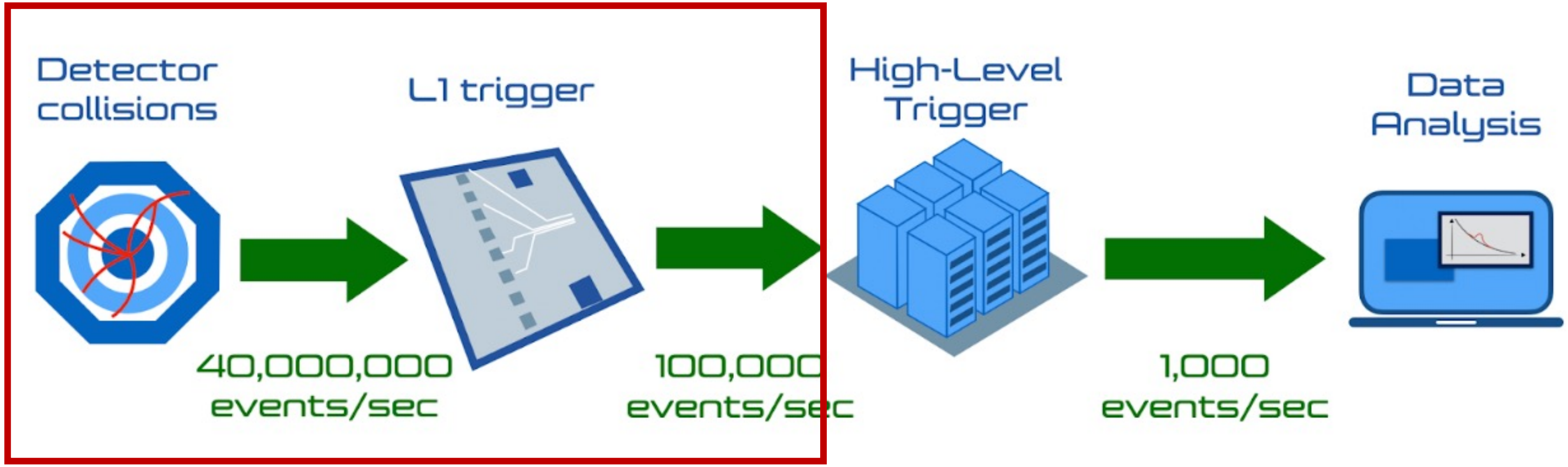
Expectation

- Good reconstruction on normal events (ZB used for training)
- Bad reconstruction on anything else (rare SM or BSM, never seen in training)

Goal

- Construct metric like mean-squared error MSE(input, output) as anomaly score for triggering

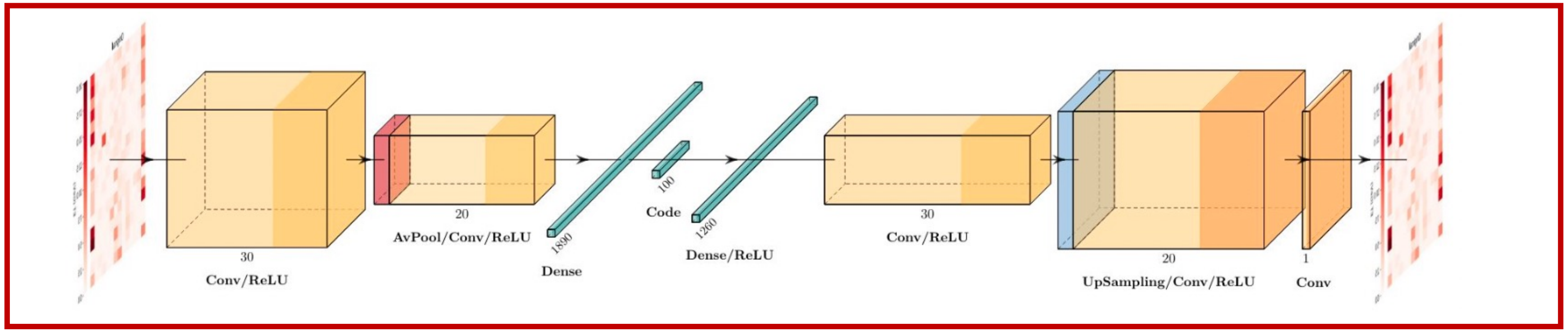
Anomaly trigger. Challenges from L1 constraints



Require algorithms to be ultra fast and lightweight but still physics sensitive

- Extremely low trigger latency < O(1) μs
- Limited computational resources from a single FPGA board
- Trigger region around background rate < O(10^{-4})

Anomaly trigger. Model compression for L1 constraints



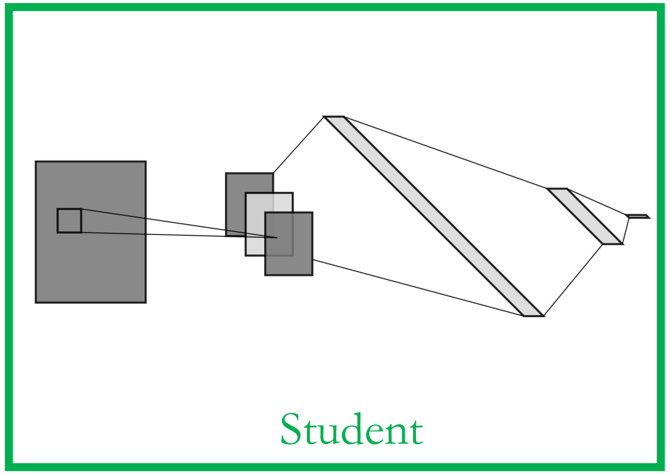
Compute
MSE(input, output)

Anomaly score

CMS-DP-2023-086

Knowledge Distillation

- Train a smaller model (student) under the guidance of a bigger model (teacher)
- The student can directly learn to regress MSE from teacher's outputs



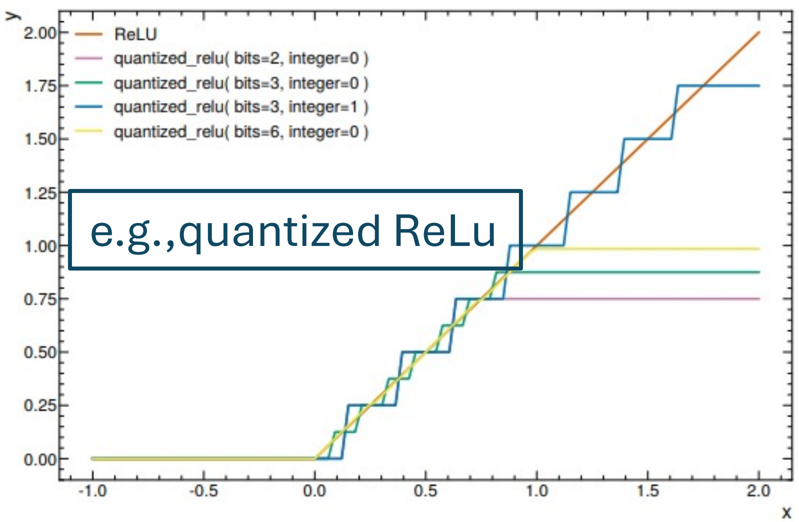
Compare with
teacher's score

Score regression

hls4ml

QKeras github

hls4ml github



Quantization-aware training (QKeras)

- Model weights quantized to fixed bit widths
- Train a quantized model rather than quantize a trained model

→ x10 reduction
in resources/latency

Anomaly trigger. Model compression for L1 constraints

CMS-DP-2023-086

Layer (type)	Output Shape	Param #
input (InputLayer)	[(None, 18, 14, 1)]	0
conv2d_1 (Conv2D)	(None, 18, 14, 20)	200
relu_1 (Activation)	(None, 18, 14, 20)	0
pool_1 (AveragePooling2D)	(None, 9, 7, 20)	0
conv2d_2 (Conv2D)	(None, 9, 7, 30)	5430
relu_2 (Activation)	(None, 9, 7, 30)	0
flatten (Flatten)	(None, 1890)	0
latent (Dense)	(None, 80)	151280
dense (Dense)	(None, 1890)	153090
reshape2 (Reshape)	(None, 9, 7, 30)	0
relu_3 (Activation)	(None, 9, 7, 30)	0
conv2d_3 (Conv2D)	(None, 9, 7, 30)	8130
relu_4 (Activation)	(None, 9, 7, 30)	0
upsampling (UpSampling2D)	(None, 18, 14, 30)	0
conv2d_4 (Conv2D)	(None, 18, 14, 20)	5420
relu_5 (Activation)	(None, 18, 14, 20)	0
output (Conv2D)	(None, 18, 14, 1)	181

Total params: 323,731
Trainable params: 323,731
Non-trainable params: 0

Teacher

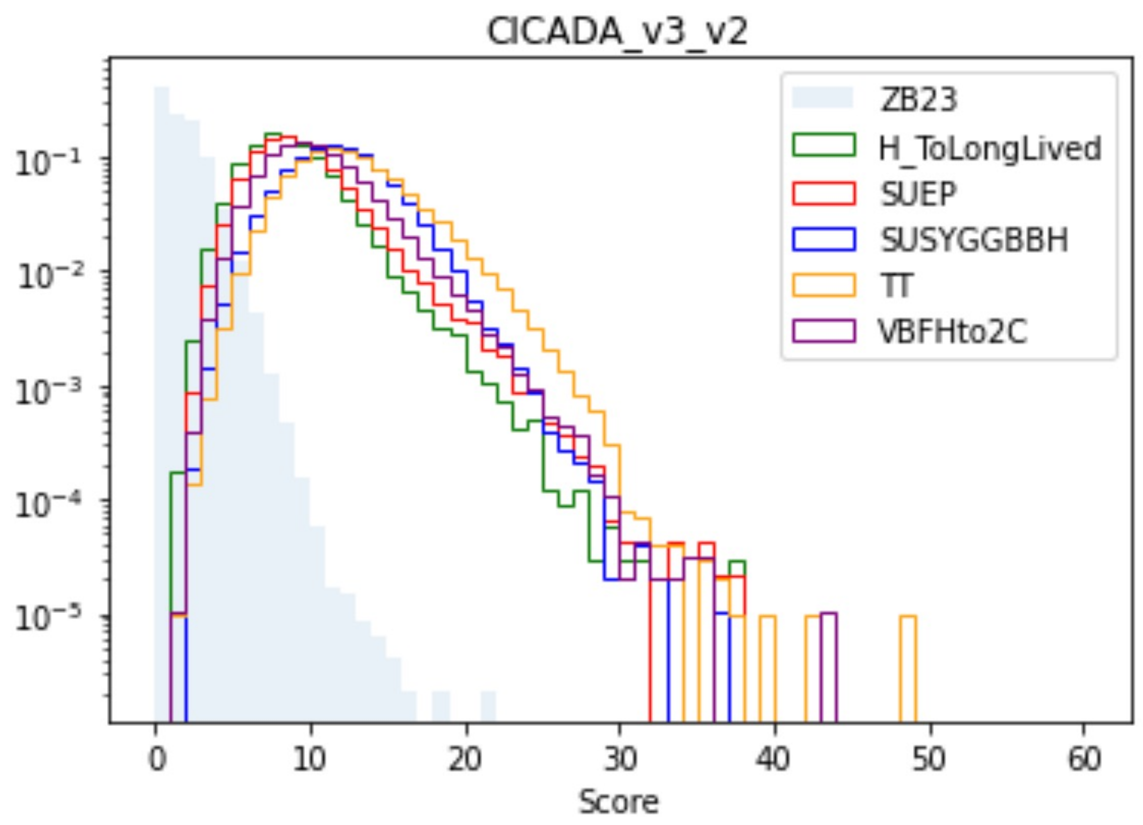
Layer (type)	Output Shape	Param #
In (InputLayer)	[(None, 252)]	0
reshape (Reshape)	(None, 18, 14, 1)	0
conv (QConv2D)	(None, 8, 6, 3)	27
relu1 (QActivation)	(None, 8, 6, 3)	0
flatten (Flatten)	(None, 144)	0
dense1 (QDense)	(None, 20)	2880
relu2 (QActivation)	(None, 20)	0
output (QDense)	(None, 1)	20

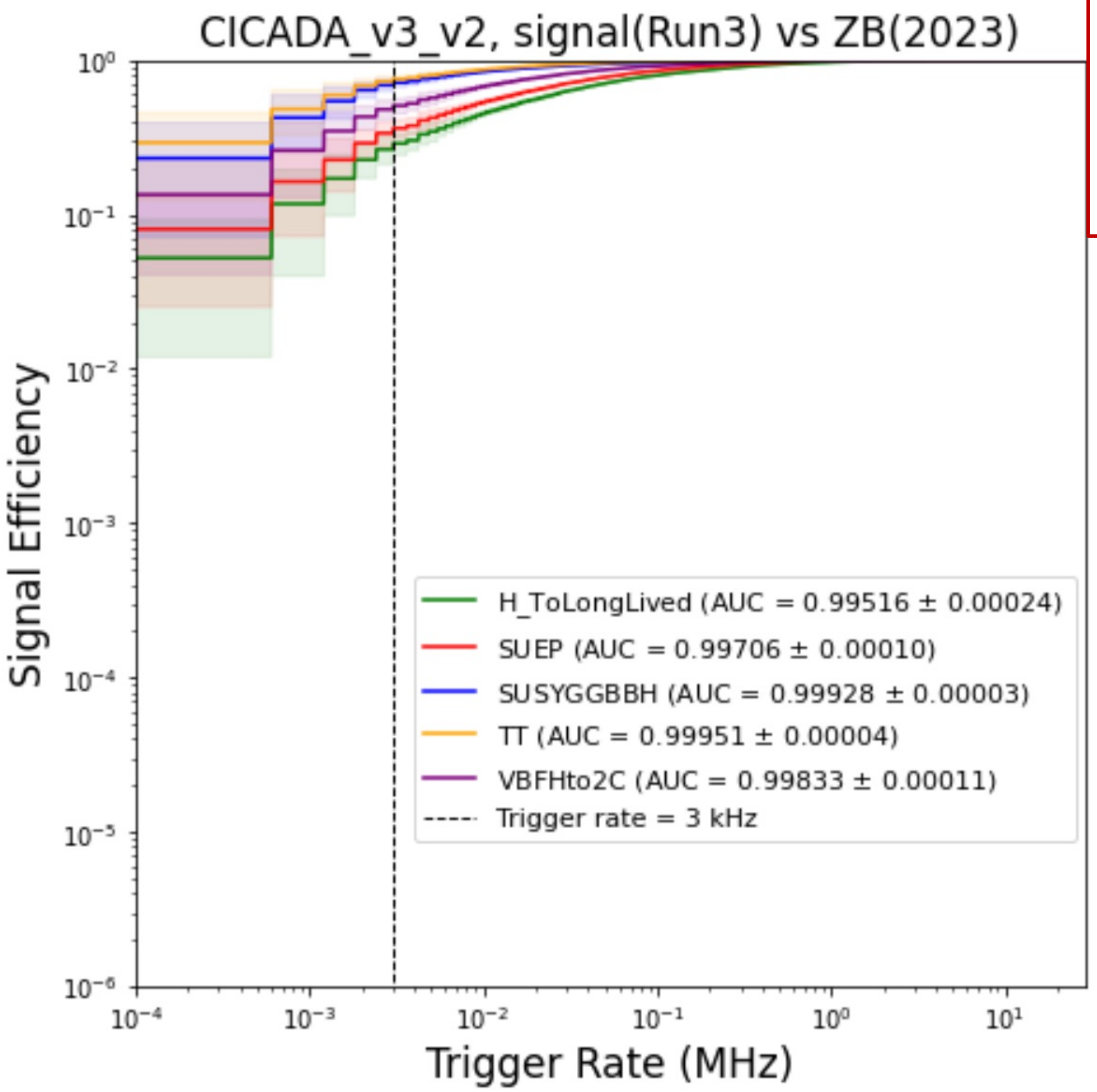
Total params: 2,927
Trainable params: 2,927
Non-trainable params: 0

Student

- 300k parameters go down to 3k parameters, while preserving most of the performance
- Inference latency ~ 100 nanoseconds
- Computational resources fit to a single FPGA board by large margins

Anomaly trigger. Anomaly score distributions

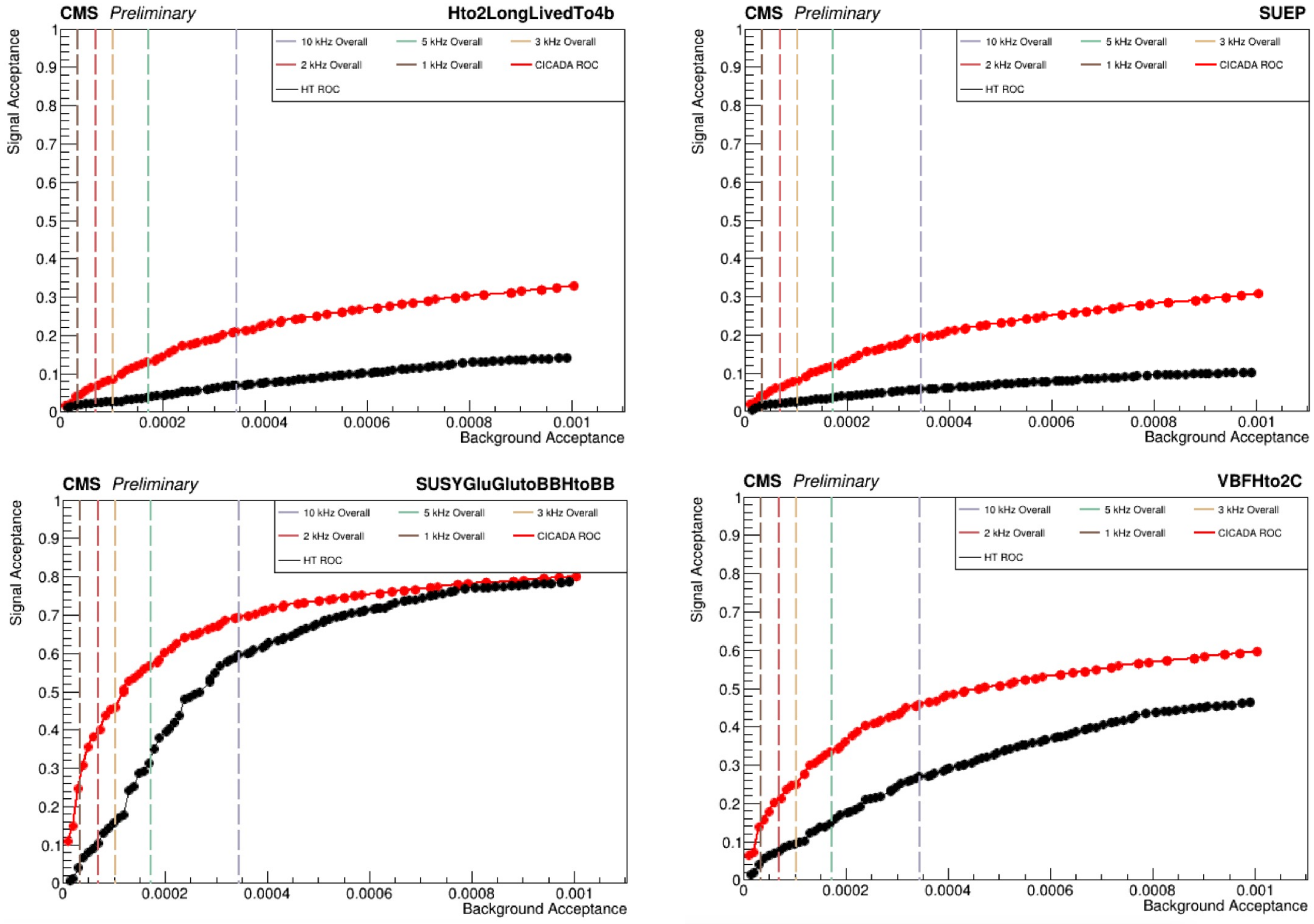




CMS-DP-2023-086

- Low anomaly score for ZB events
- High anomaly score for rare SM / BSM

Anomaly trigger. Compare with cut-based HT trigger



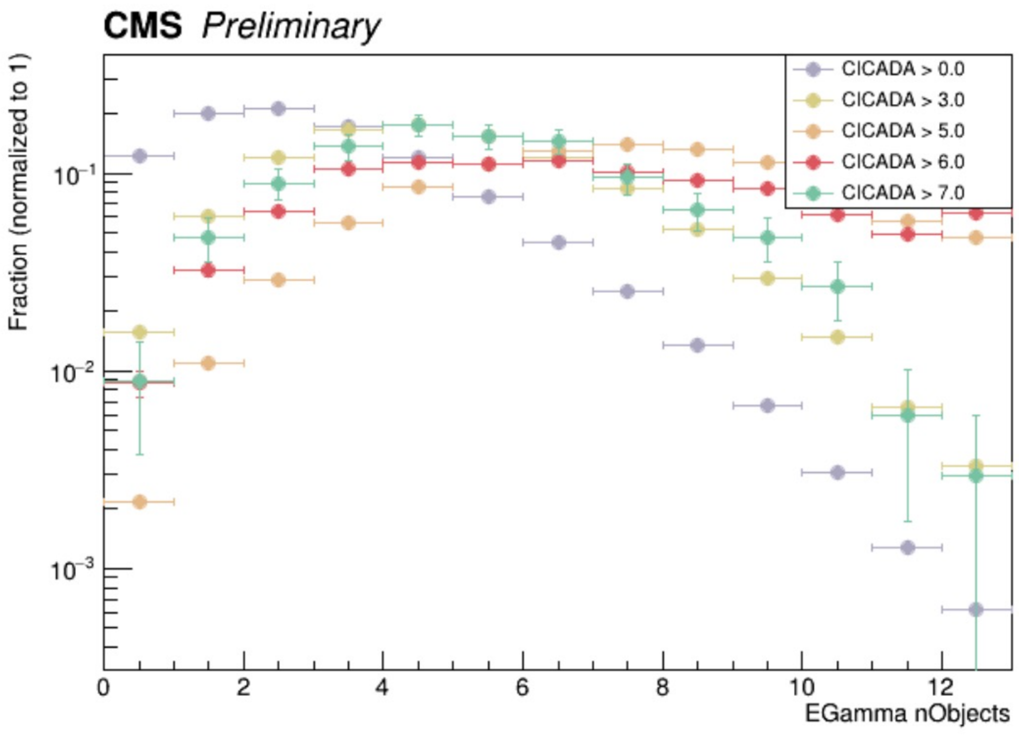
- HT trigger vs. CICADA
- Region of interest
  - $O(10^{-4})$ bkg rate, or
  - $O(1)$ kHz trigger rate

CMS-DP-2023-086

Anomaly trigger. Some trigger objects (EGamma)

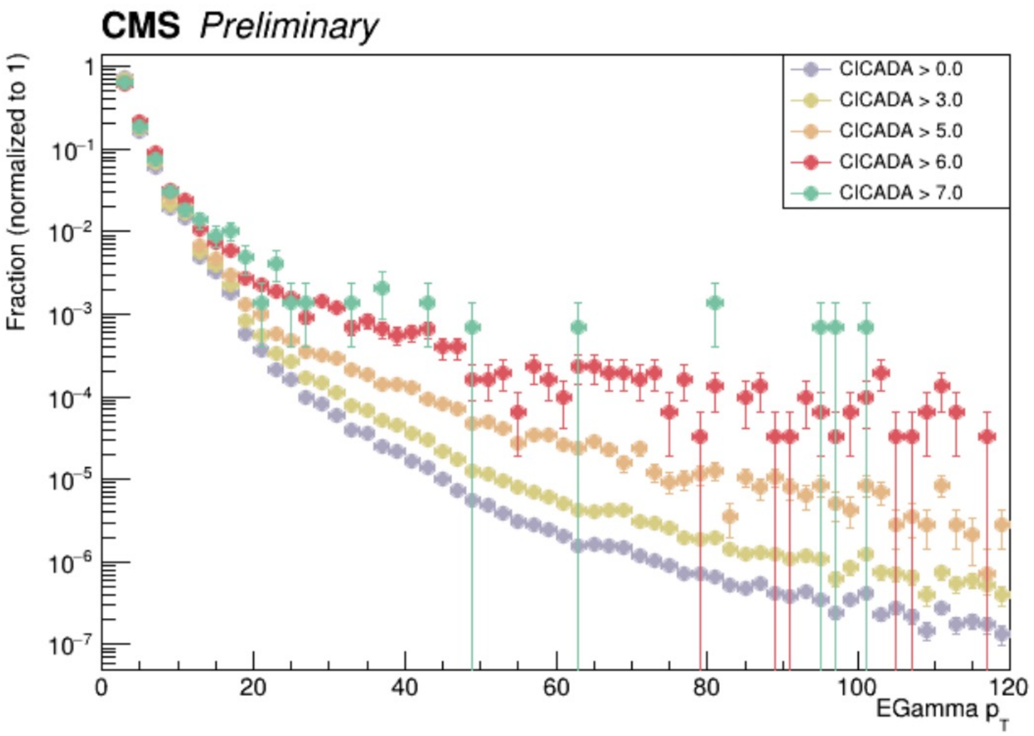
CMS-DP-2023-086

nObjects



- CICADA has preference for more EG objects
  - 3 or more

Pt



- Slight preference for higher Pt

- Still sensitive to low Pt objects as well
- Jet and Tau in backup

Anomaly trigger. Takeaway

Deployment expected in the second half of 2024 during Run 3!

- A major bottleneck in the development was the tight resource constraints at L1
- Tricks of model compression used here may not work in every other case
- We need more modeling options for higher flexibility

➢ Next: symbolic regression for hardware-efficient ML inference with nanosecond-latency

Symbolic regression

I. Search for exotic Higgs boson decays $h(125) \rightarrow aa \rightarrow bb\tau\tau$ / $bb\mu\mu$

II. Anomaly detection trigger at the Level-1 for model-agnostic new physics searches

III. Symbolic regression for hardware-efficient ML inference with nanosecond-latency

Computer Science > Machine Learning

[Submitted on 6 May 2023 (v1), last revised 17 Jan 2024 (this version, v2)]

Symbolic Regression on FPGAs for Fast Machine Learning Inference

Ho Fung Tsoi, Adrian Alan Pol, Vladimir Loncar, Ekaterina Govorkova, Miles Cranmer, Sridhara Dasu, Peter Elmer, Philip Harris, Isobel Ojalvo, Maurizio Pierini

SR on FPGAs with hls4ml [2305.04099]
→ 5 nanoseconds latency on the LHC jet tagging!

Computer Science > Machine Learning

[Submitted on 18 Jan 2024]

SymbolNet: Neural Symbolic Regression with Adaptive Dynamic Pruning

Ho Fung Tsoi, Vladimir Loncar, Sridhara Dasu, Philip Harris

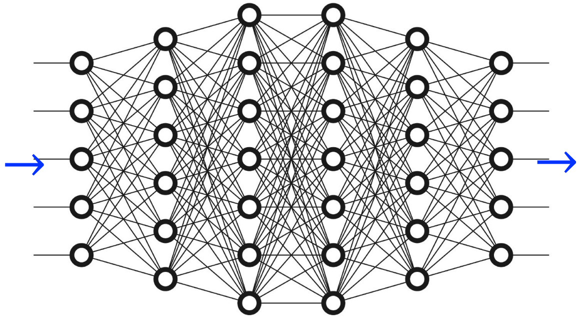
New architecture to scale SR to higher input dimensions [2401.09949]
→ SR on MNIST and SVHN

Symbolic regression. Introduction

- A ML technique that seeks to discover analytic functions that approximate a dataset

Neural network

Dataset input →



→ Output

Large black box

vs.

Symbolic regression

Dataset input → $f(x) = x_0^3 - \sin(x_1 x_3) + x_2 \log(5 + x_4) + \cdots$ → Output

Closed-form analytic equation

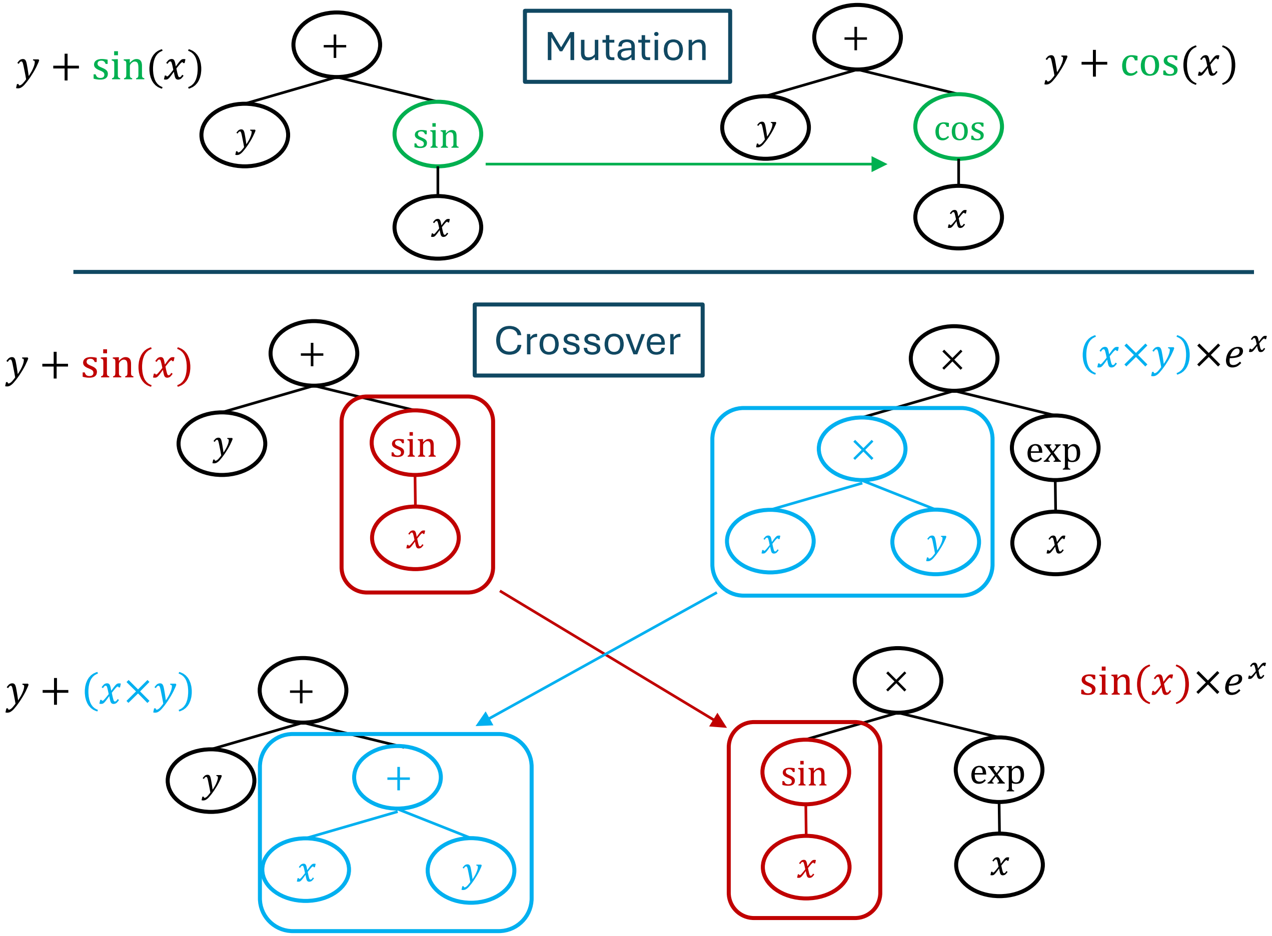
	Traditional regression (linear, polynomial,...)	**Symbolic regression**
Fit inputs	Pre-specified functional form required	Only primitive operators needed e.g. $+, \times, ^\wedge, \sin, \exp, \ldots$
Functional form	Fixed throughout the fit	Dynamically evolving throughout the fit
Expressivity	Low due to only one functional form available	High due to vast equation search space

- Furthermore, symbolic models allow human-interpretability and compact representation of the data
  - Great potential for low latency ML!

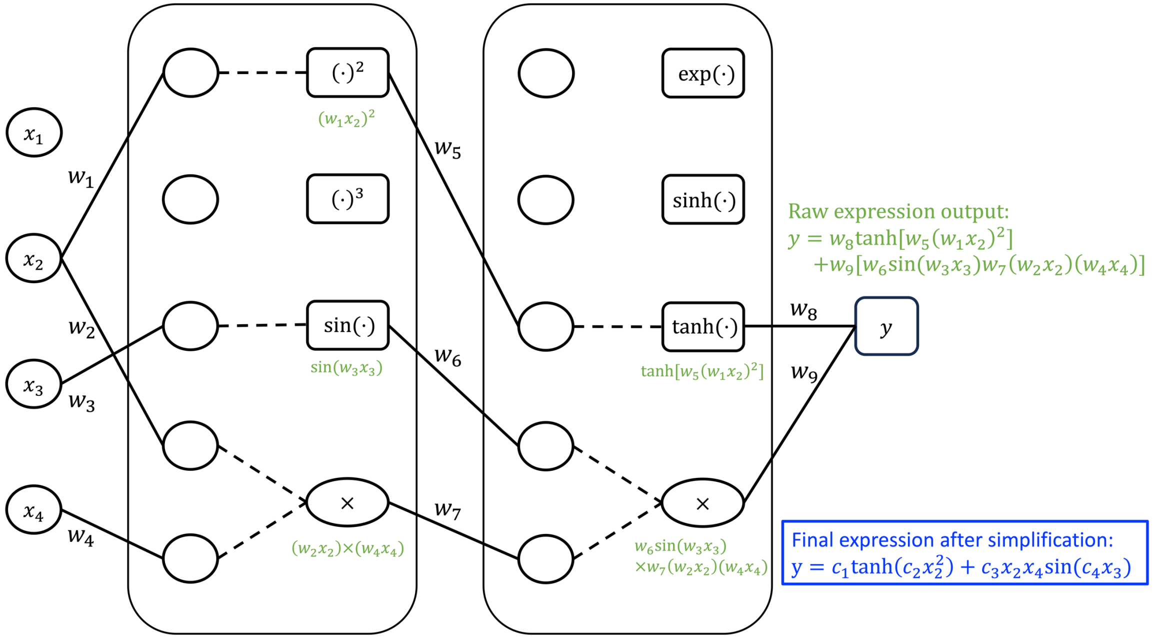
Symbolic regression. How to do SR?

Two popular approaches

- **Genetic programming (discrete nature)**
  - Construct equation in a tree representation, then mutate and crossover substructures by mimicking biological evolution

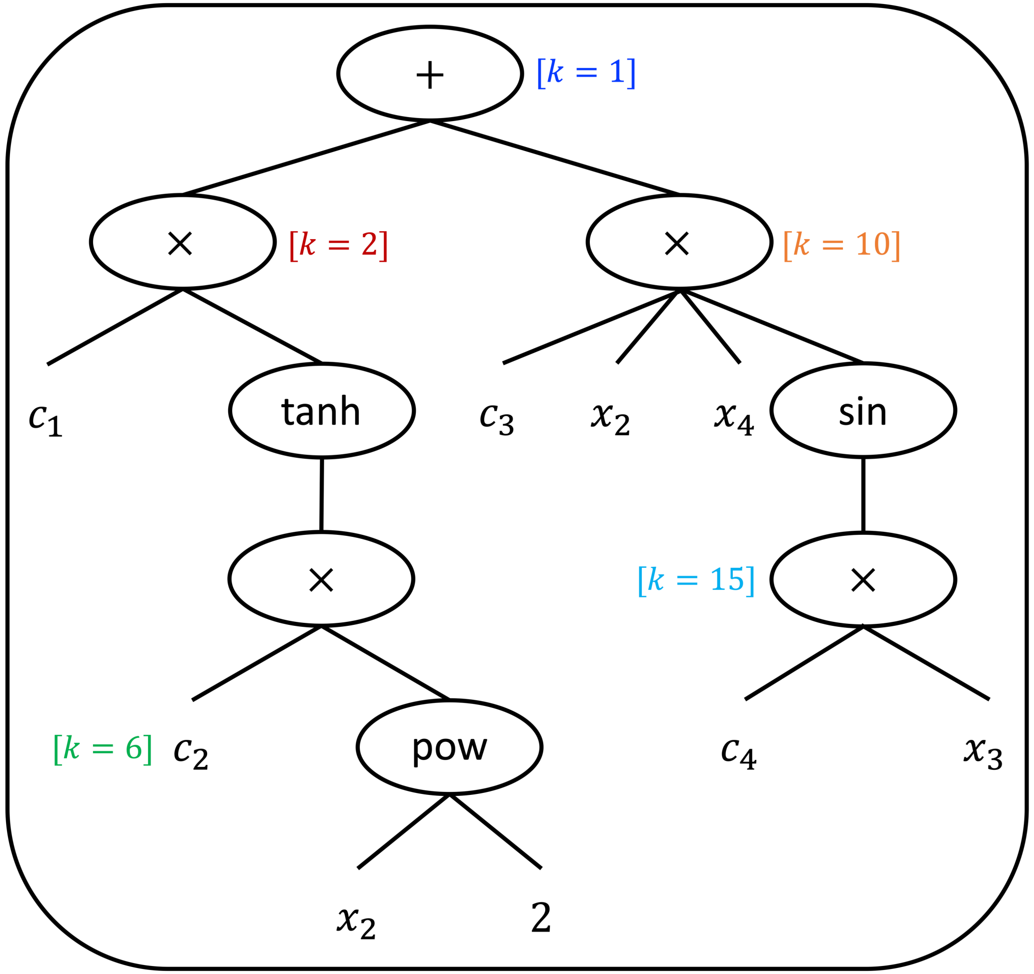


- **Gradient method (continuous nature)**
  - Construct neural network with custom math operators as activation, train in sparsity and then unroll to obtain closed-form equation



arXiv:2401.09949

Symbolic regression. Equation complexity as model size



Each step, k, of the tree traversal:
$[k = 1]$: $c_1 \tanh(c_2 x_2^2) + c_3 x_2 x_4 \sin(c_4 x_3)$
$[k = 2]$: $c_1 \tanh(c_2 x_2^2)$
$[k = 3]$: c_1
$[k = 4]$: $\tanh(c_2 x_2^2)$
$[k = 5]$: $c_2 x_2^2$
$[k = 6]$: c_2
$[k = 7]$: x_2^2
$[k = 8]$: x_2
$[k = 9]$: 2
$[k = 10]$: $c_3 x_2 x_4 \sin(c_4 x_3)$
$[k = 11]$: c_3
$[k = 12]$: x_2
$[k = 13]$: x_4
$[k = 14]$: $\sin(c_4 x_3)$
$[k = 15]$: $c_4 x_3$
$[k = 16]$: c_4
$[k = 17]$: x_3

Expression complexity = total traversal steps
= 17

- Counting the number of tree nodes (can be done using sympy preorder_traversal)
- Assuming every math operation is equally weighted (naively)

Symbolic regression. Adapting Symbolic models on FPGA

arXiv:2305.04099

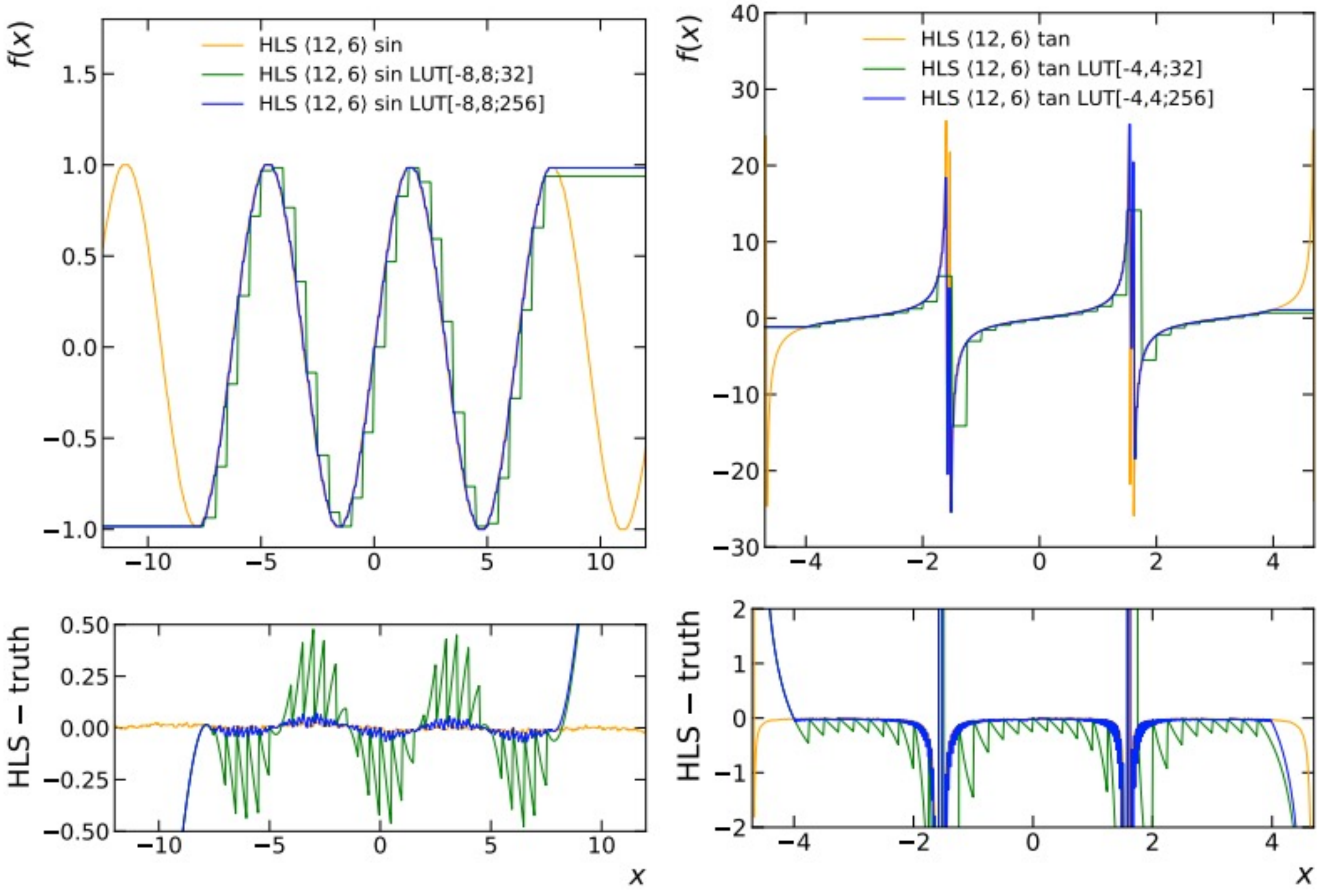
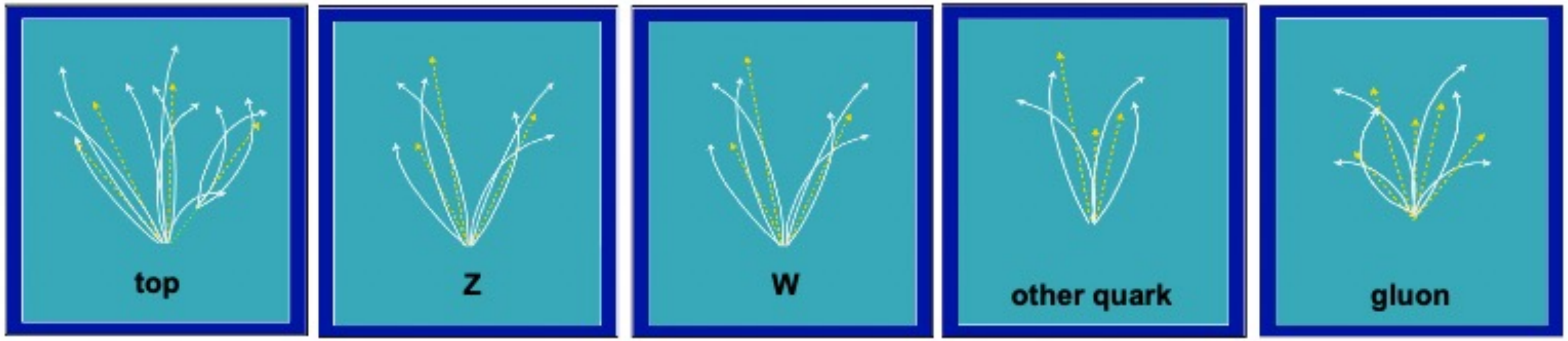


Figure 1. The sine (left) and tangent (right) functions evaluated with and without the use of LUTs, implemented in HLS with precision ⟨12, 6⟩, i.e. 12 bits variable with 6 integer bits. The LUT notation reads: [range start, range end; table size] for table definition. The lower panel shows the function deviation from the truth.

Math operations such as $\tan(\cdot)$ can be computationally expensive on an FPGA

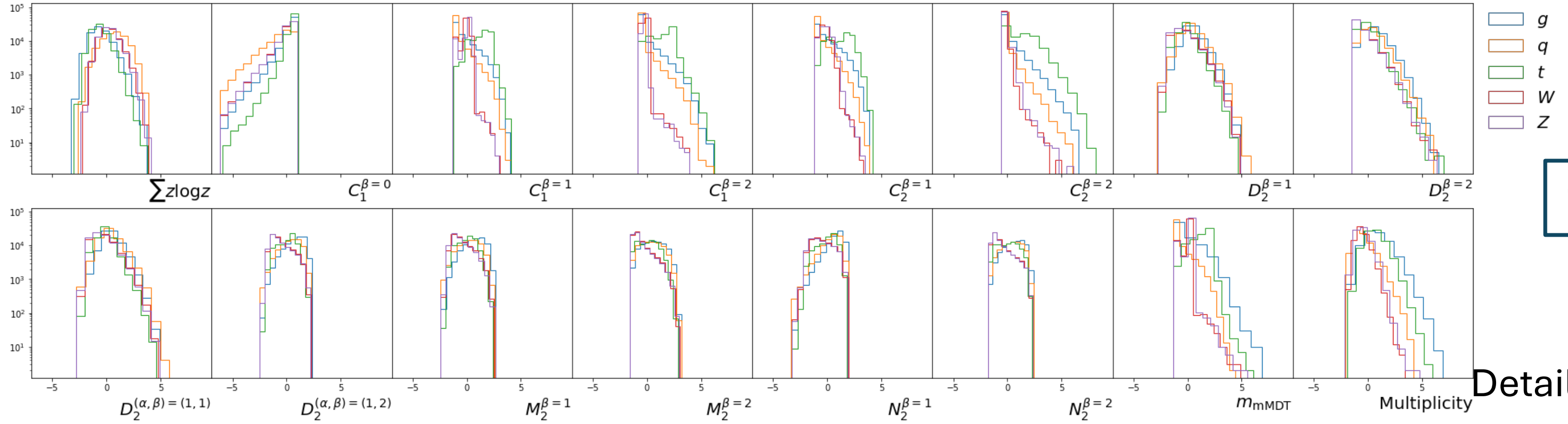
- Approximation of math function with look-up table
    - Look-up table: use an array that maps input to output
    - Replace runtime computation by much simpler array indexing
- An alternative approach for resource saving (latency-aware training) in backup

Symbolic regression. Benchmark on LHC jet tagging



A physics benchmark: LHC jet tagging dataset for fast ML studies [Zenodo]

- Dataset of ~1M simulated boosted jets produced from LHC proton-proton collisions
- Each jet is represented by 16 high-level features
- 5-class classification → {gluon, light-quark, W, Z, top}



16 input features

Observables

m_{mMDT}
$N_2^{\beta=1,2}$
$M_2^{\beta=1,2}$
$C_1^{\beta=0,1,2}$
$C_2^{\beta=1,2}$
$D_2^{\beta=1,2}$
$D_2^{(\alpha,\beta)=(1,1),(1,2)}$
$\sum z \log z$
Multiplicity

Detailed description of the dataset in [1709.08705]

Symbolic regression. Symbolic jet taggers

arXiv:2305.04099

- Use genetic programming-based library PySR to perform symbolic regression
- HLS converter developed for symbolic expressions in hls4ml

Single-math-class models {primitive math operations allowed in the fit}

- **Polynomial** $\{+, -, \times\}$
- **Trigonometric** $\{+, -, \times, \sin(\cdot)\}$
- **Exponential** $\{+, -, \times, \text{Gauss}(\cdot) = \exp(-(\cdot)^2)\}$
- **Logarithmic** $\{+, -, \times, \log(\text{abs}(\cdot))\}$

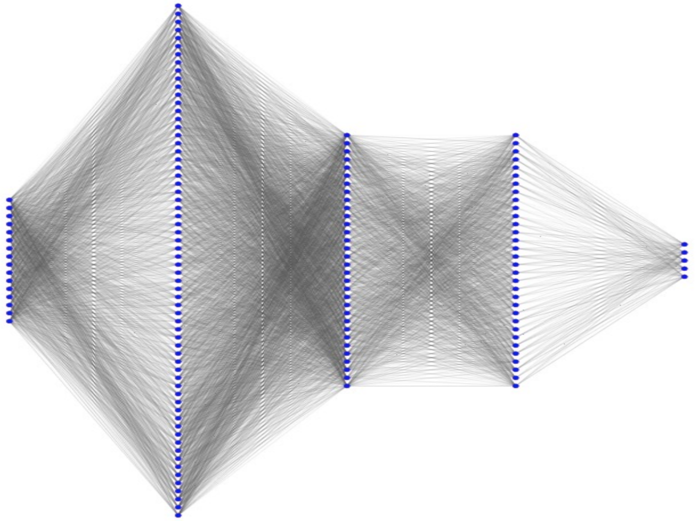
Model	Expression for the t tagger with $c_{\max} = 40$	AUC
Polynomial	$C_1^{\beta=2} + 0.09m_{\text{mMDT}}(2C_1^{\beta=1} + M_2^{\beta=2} - m_{\text{mMDT}} - \text{Multiplicity} - (1.82C_1^{\beta=1} - M_2^{\beta=2})(C_1^{\beta=2} - 0.49m_{\text{mMDT}}) - 3.22) - 0.53$	0.914
Trigonometric	$\sin(0.06(\sum z \log z)M_2^{\beta=2} - 0.25C_1^{\beta=2}(-C_1^{\beta=1} + 2C_1^{\beta=2} - M_2^{\beta=2} + \text{Multiplicity} - 8.86) - m_{\text{mMDT}} + 0.06\text{Multiplicity} - 0.4)$	0.925
Exponential	$0.23C_1^{\beta=1}(-m_{\text{mMDT}} + \text{Gauss}(0.63\text{Multiplicity}) + 1) - \text{Gauss}(C_1^{\beta=1}) + 0.45C_1^{\beta=2} - 0.23m_{\text{mMDT}}$ $+0.23\text{Gauss}((4.24 - 1.19C_2^{\beta=1})(C_1^{\beta=2} - m_{\text{mMDT}})) + 0.15$	0.920
Logarithmic	$C_1^{\beta=2} - 0.1m_{\text{mMDT}}(\text{Multiplicity} \times \log(\text{abs}(\text{Multiplicity})) + 2.2) - 0.02\log(\text{abs}(\text{Multiplicity}))$ $-0.1(C_1^{\beta=2}(C_1^{\beta=1} - 1.6M_2^{\beta=2} + m_{\text{mMDT}} + 1.28) - m_{\text{mMDT}} - 0.48)\log(\text{abs}(C_1^{\beta=2})) - 0.42$	0.923

Table 2. Expressions generated by PySR for the t tagger in different models with $c_{\max} = 40$. Operator complexity is set to 1 by default. Constants are rounded to two decimal places here.

Symbolic regression. Compare SR with baseline NN

arXiv:2305.04099

Baseline NN taggers



MLP [16-64-32-32-5], $O(10^3)$ parameters

vs.

Symbolic taggers (analytic expressions)

Tagger	Expression for the trigonometric model with $c_{\max} = 20$	AUC
g	$\sin(-2C_1^{\beta=1} + 0.31C_1^{\beta=2} + m_{\mathrm{mMDT}} + \mathrm{Multiplicity} - 0.09\mathrm{Multiplicity}^2 - 0.79)$	0.897
q	$-0.33(\sin(m_{\mathrm{mMDT}}) - 1.54)(\sin(-C_1^{\beta=1} + C_1^{\beta=2} + \mathrm{Multiplicity}) - 0.81)\sin(m_{\mathrm{mMDT}}) - 0.81$	0.853
t	$\sin(C_1^{\beta=1} + C_1^{\beta=2} - m_{\mathrm{mMDT}} + 0.22(C_1^{\beta=2} - 0.29)(-C_1^{\beta=2} + C_2^{\beta=1} - \mathrm{Multiplicity}) - 0.68)$	0.920
W	$-0.31(\mathrm{Multiplicity} + (2.09 - \mathrm{Multiplicity})\sin(8.02C_1^{\beta=2} + 0.98)) - 0.5$	0.877
Z	$(\sin(4.84m_{\mathrm{mMDT}}) + 0.59)\sin(m_{\mathrm{mMDT}} + 1.14)\sin(C_1^{\beta=2} + 4.84m_{\mathrm{mMDT}}) - 0.94$	0.866

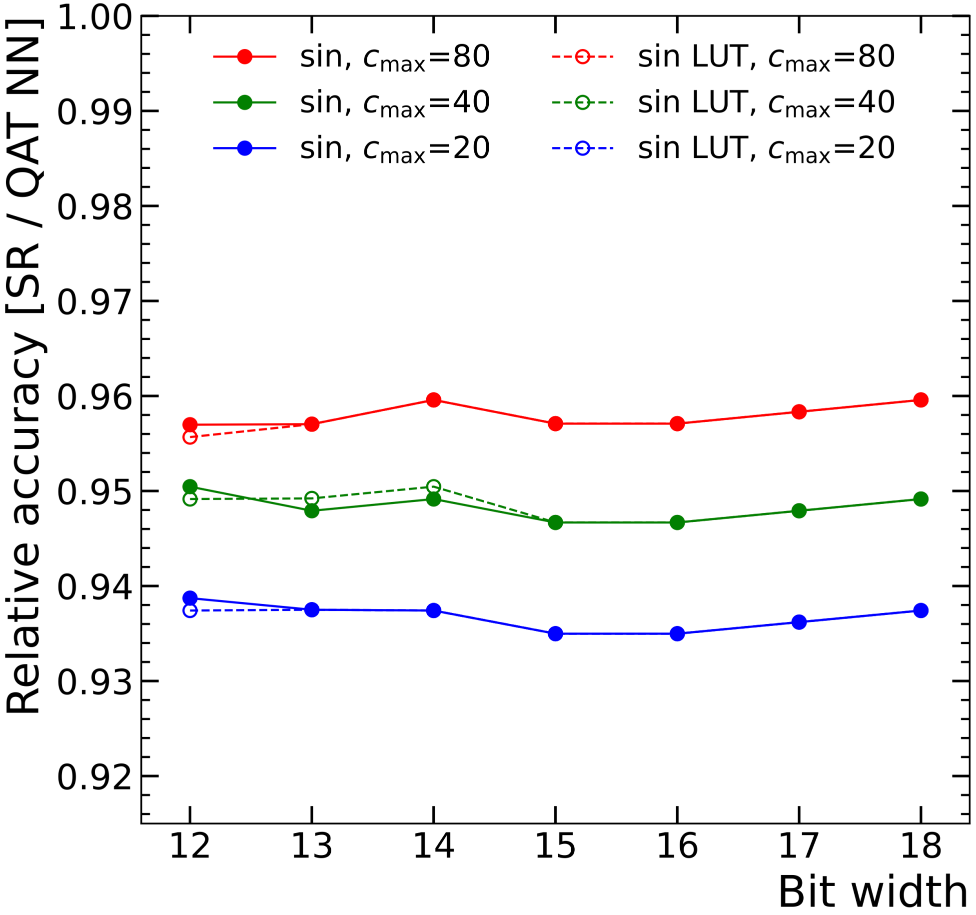
Table 1. Expressions generated by PySR for the trigonometric model with $c_{\max} = 20$. Operator complexity is set to 1 by default. Constants are rounded to two decimal places for readability. Area under the receiver operating characteristic (ROC) curve, or AUC, is reported.

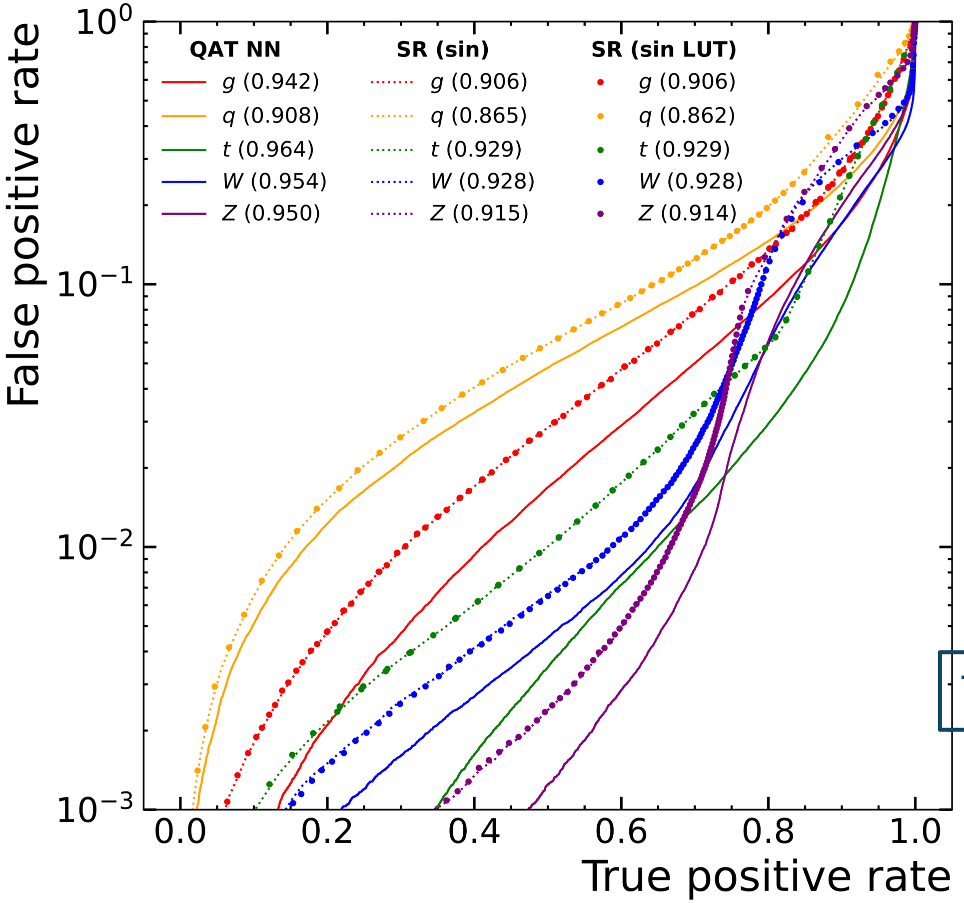
Evaluation metrics

- Physics performance (jet tagging accuracy)
- Inference latency and resource utilization on an FPGA (DSPs and LUTs)

Symbolic regression. Physics performance

arXiv:2305.04099

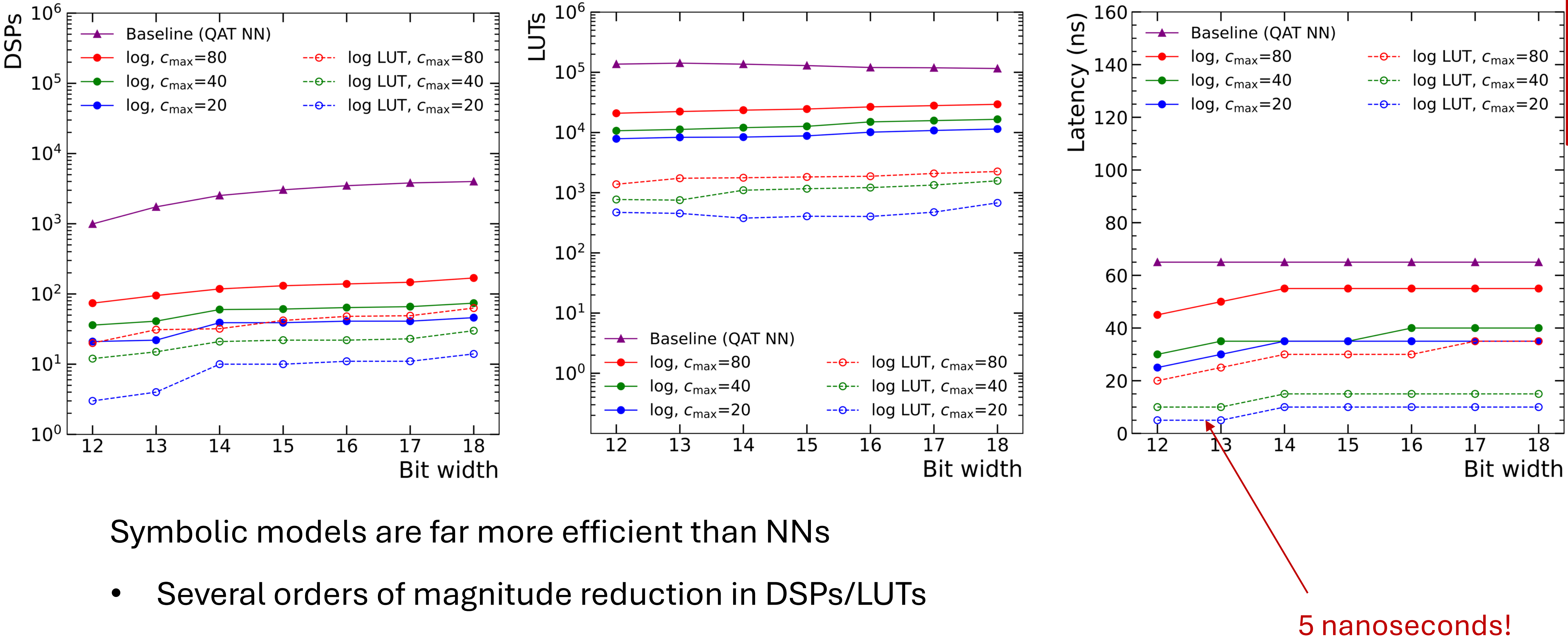




Trigonometric model

- $c_{\max}$: maximum equation complexity allowed in the fit
  - Very simple symbolic models are already competitive to large neural network with O(10^3) parameters

Symbolic regression. Latency and resource utilization



arXiv:2305.04099

Symbolic models are far more efficient than NNs

- Several orders of magnitude reduction in DSPs/LUTs
- Several times faster inference speed

while physics performance is comparable

Symbolic regression. SR vs. traditional compression

arXiv:2401.09949

TABLE IV

RESOURCE UTILIZATION AND LATENCY ON AN FPGA FOR QUANTIZED AND PRUNED (QP) NNS AND SYMBOLIC EXPRESSIONS LEARNED BY SYMBOLNET. THE MODEL SIZE IS QUOTED IN TERMS OF THE NUMBER OF NEURONS PER HIDDEN LAYER FOR DNN, AND THE NUMBER OF FILTERS FOR CNN, WHERE, FOR EXAMPLE, $(16)_3$ MEANS 16 FILTERS WITH A KERNEL SIZE OF 3×3. THE INITIATION INTERVAL (II) IS QUOTED IN CLOCK CYCLES. THE NUMBERS IN PARENTHESES INDICATE THE PERCENTAGE OF TOTAL AVAILABLE RESOURCE UTILIZATION. THE RELATIVE ACCURACY AND THE ROC AUC ARE EVALUATED WITH RESPECT TO THE SAME DNN/CNN IMPLEMENTED IN FLOATING POINT PRECISION AND WITHOUT PRUNING.

	LHC jet tagging (five classes)								
	Model size (input dim. = 16)	Precision	BRAMs	DSPs	FFs	LUTs	II	Latency	Rel. acc.
QP DNN	[64, 32, 32, 5], **90% pruned**	⟨6, 0⟩	4 (0.1%)	28 (0.4%)	2739 (0.1%)	7691 (0.7%)	1	55 ns	94.7%
SR	**Mean complexity of the five expr. = 18**	⟨12, 8⟩	0 (0%)	3 (0%)	109 (0%)	177 (0%)	1	**10 ns**	93.3%
	MNIST (ten classes)								
	Model size (input dim. = 28×28×1)	Precision	BRAMs	DSPs	FFs	LUTs	II	Latency	Rel. acc.
QP CNN	[$(16, 16, 24)_3$, 42, 64, 10], **92% pruned**	⟨6, 0⟩	66 (1.5%)	216 (3.2%)	18379 (0.8%)	29417 (2.5%)	788	4.0 μs	86.8%
SR	**Mean complexity of the ten expr. = 133**	⟨18, 10⟩	0 (0%)	160 (2.3%)	6424 (0.3%)	7592 (0.6%)	1	**125 ns**	85.3%
	SVHN (binary "1" vs. "7")								
	Model size (input dim. = 32×32×3)	Precision	BRAMs	DSPs	FFs	LUTs	II	Latency	Rel. AUC
QP CNN	[$(16, 16, 24)_3$, 42, 64, 1], **92% pruned**	⟨6, 0⟩	62 (1.4%)	77 (1.1%)	16286 (0.7%)	27407 (2.3%)	1029	5.2 μs	94.0%
SR	**Complexity = 311**	⟨10, 4⟩	0 (0%)	38 (0.6%)	1945 (0.1%)	3029 (0.3%)	1	**195 ns**	94.5%

- Compare symbolic models with traditional compression methods i.e. quantized and pruned NNs
- SR still uses much lower resources and runs much faster on FPGAs

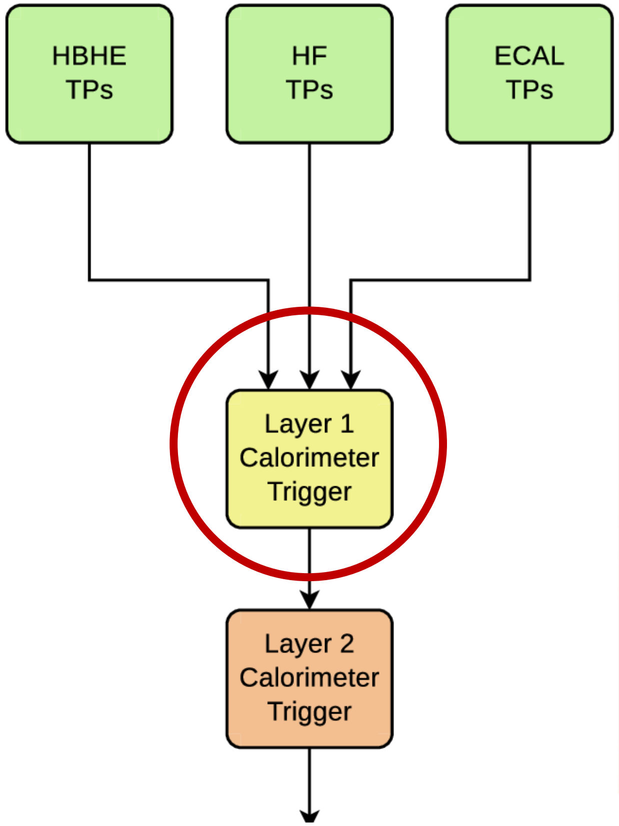
Symbolic regression. Takeaway

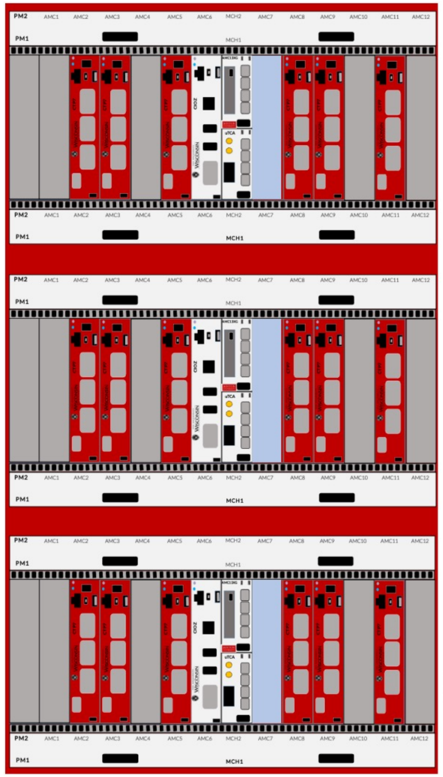
- SR can be a promising alternative to NN-based models for solving critical tasks in resource-constrained systems such as the LHC experiments

- Potential applications at the LHC
  - Standalone classifier for object identification etc.
  - Standalone regressor for particle energy reconstruction etc.
  - New model compression technique for overcoming the L1 constraints

Technical: CaloLayer-1

I. Search for exotic Higgs boson decays $h(125) \rightarrow aa \rightarrow bb\tau\tau$ / $bb\mu\mu$
II. Anomaly detection trigger at the Level-1 for model-agnostic new physics searches
III. Symbolic regression for hardware-efficient ML inference with nanosecond-latency
IV. A slide on the technical: calorimeter layer-1 trigger subsystem

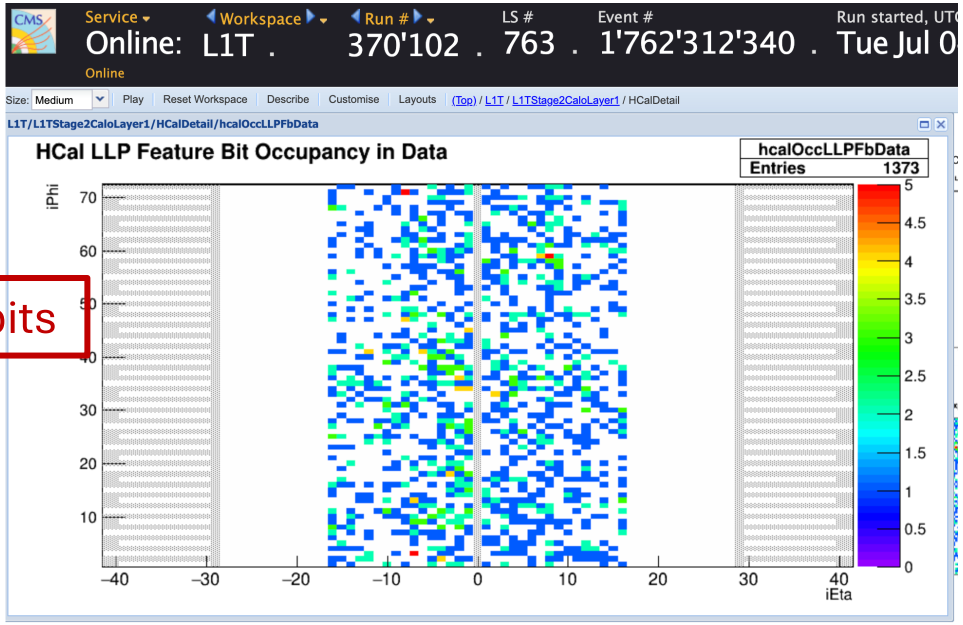




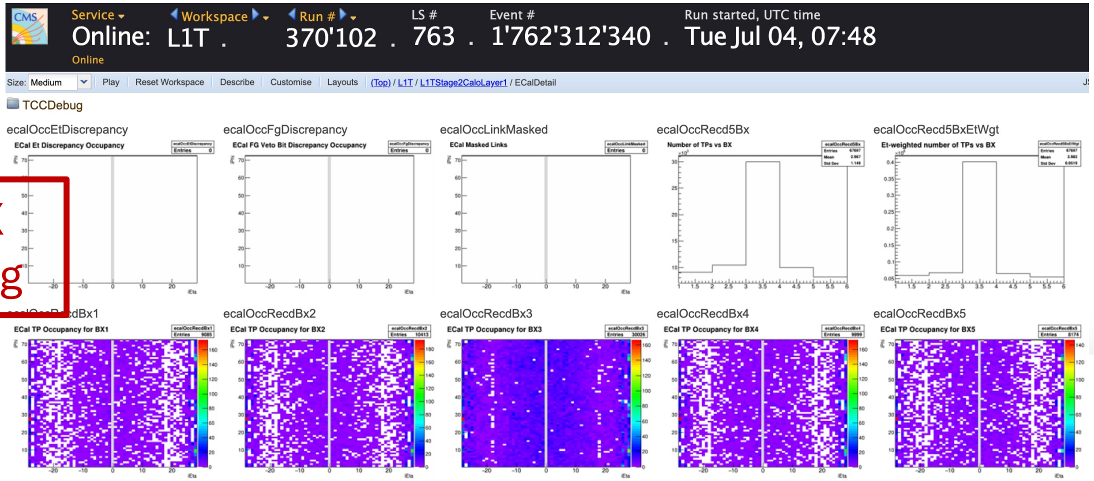
- The Calorimeter layer-1 trigger subsystem (CaloLayer-1) was developed and managed by the Wisconsin group
- My contribution
    - Developed and maintained the software for the Data Quality Monitoring (DQM)

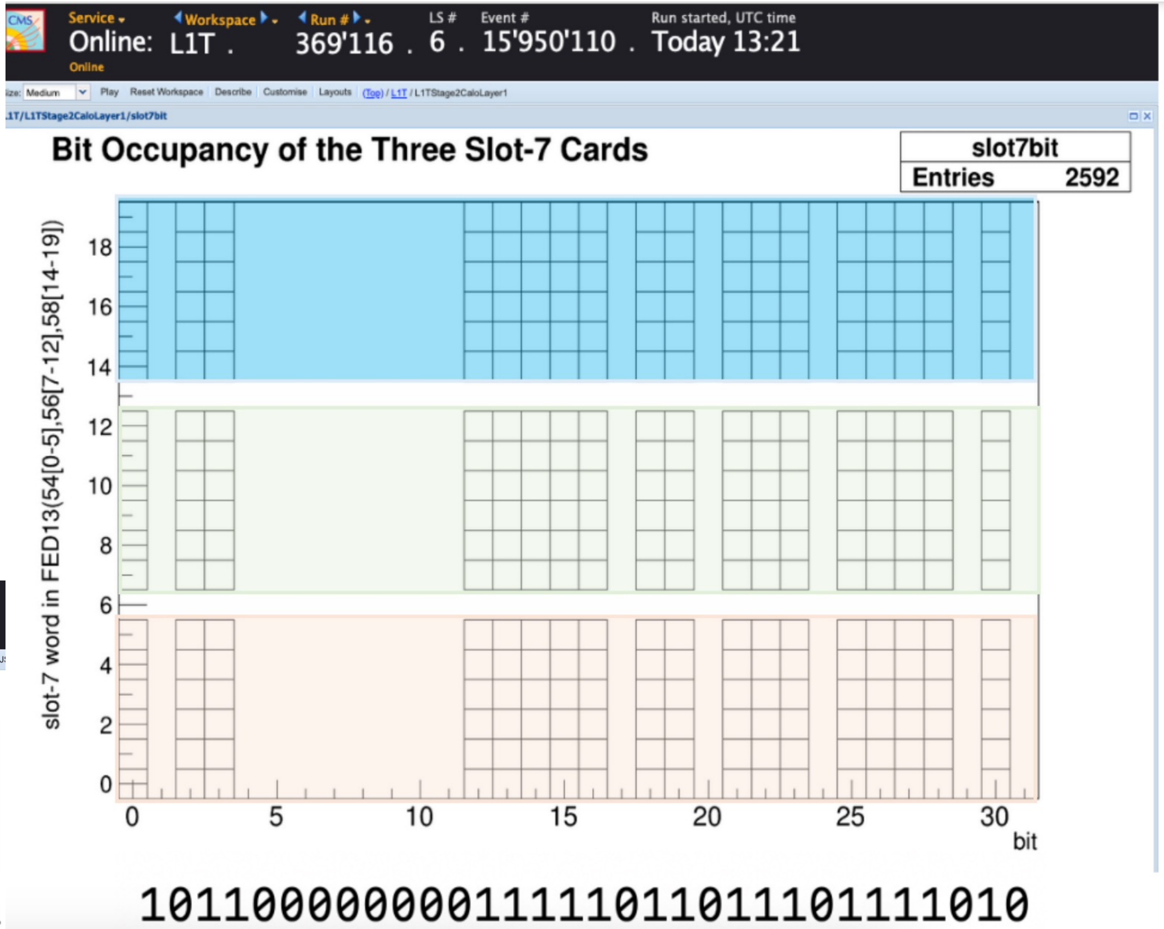
CaloLayer-1. Data quality monitoring

HCAL LLP bits



ECAL 5-BX for prefiring





New cards for CICADA anomaly detection trigger

- Commissioned the monitoring for the three projects shown here

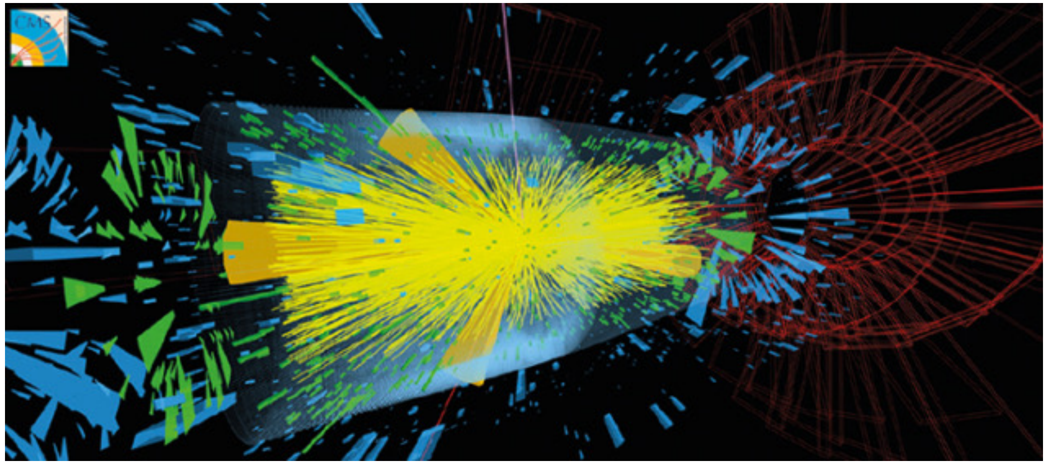
Summary and outlook

- No BSM physics discovered at the LHC (yet!)

- Seen improvements using ML over traditional methods in many areas
  - Analyses, reconstruction, low latency L1T,...

- Increasing needs of ML almost everywhere

- Postdoc outlook
  - Continue new physics searches
  - Novel ML methods for HEP applications
  - Low latency ML based online algorithms
  - Phase 2 upgrade

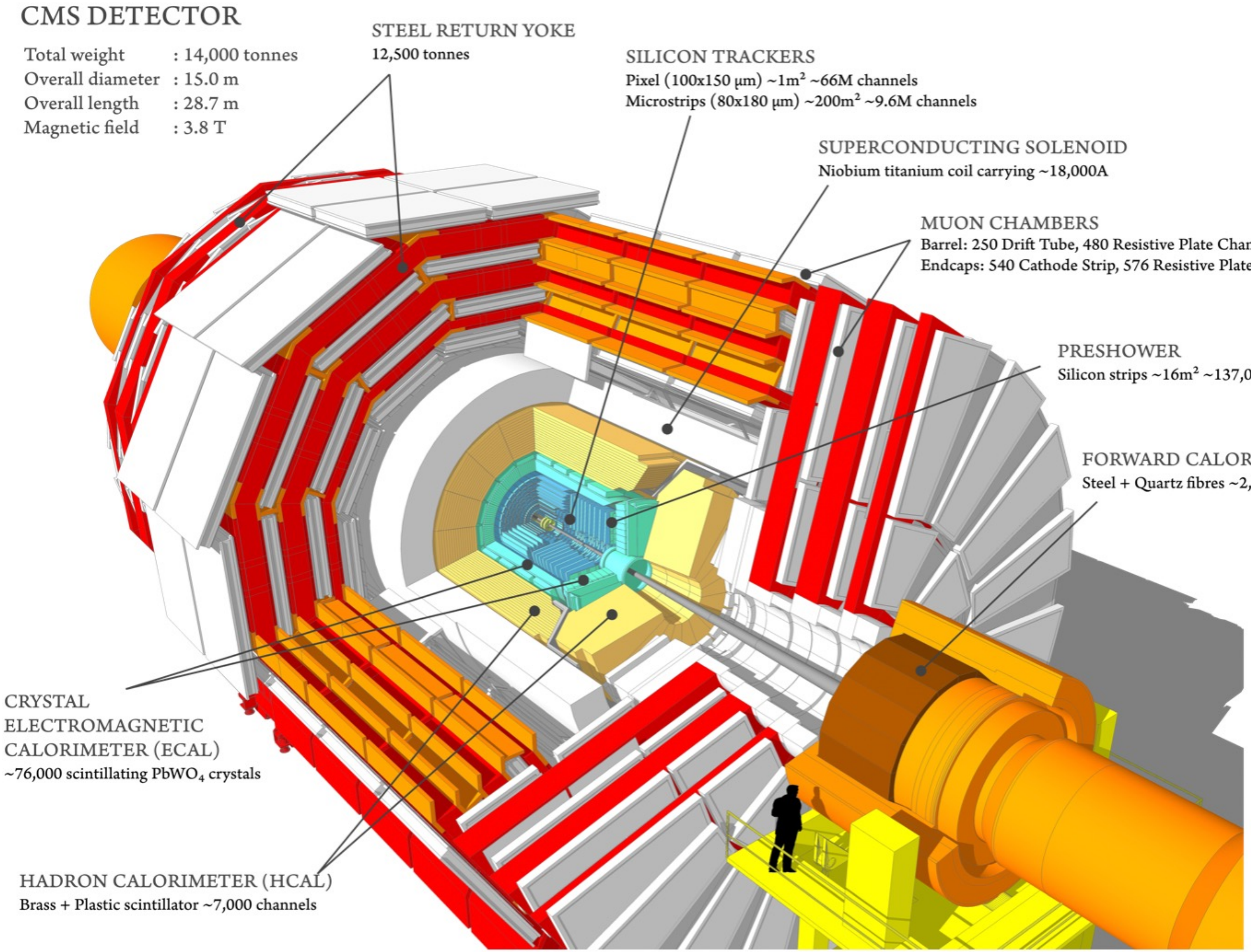


Thank you!

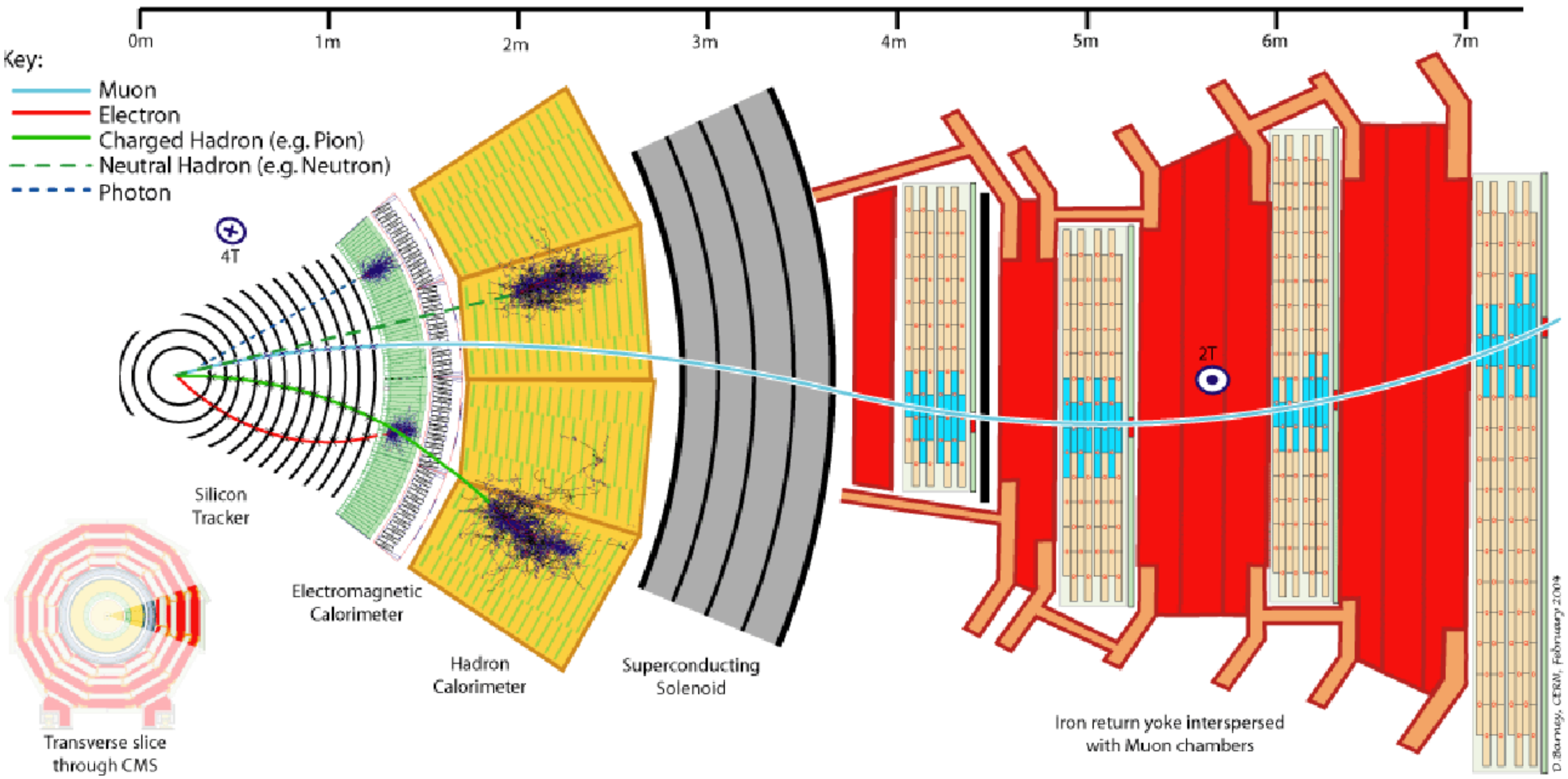
Backup

Backup - Analysis

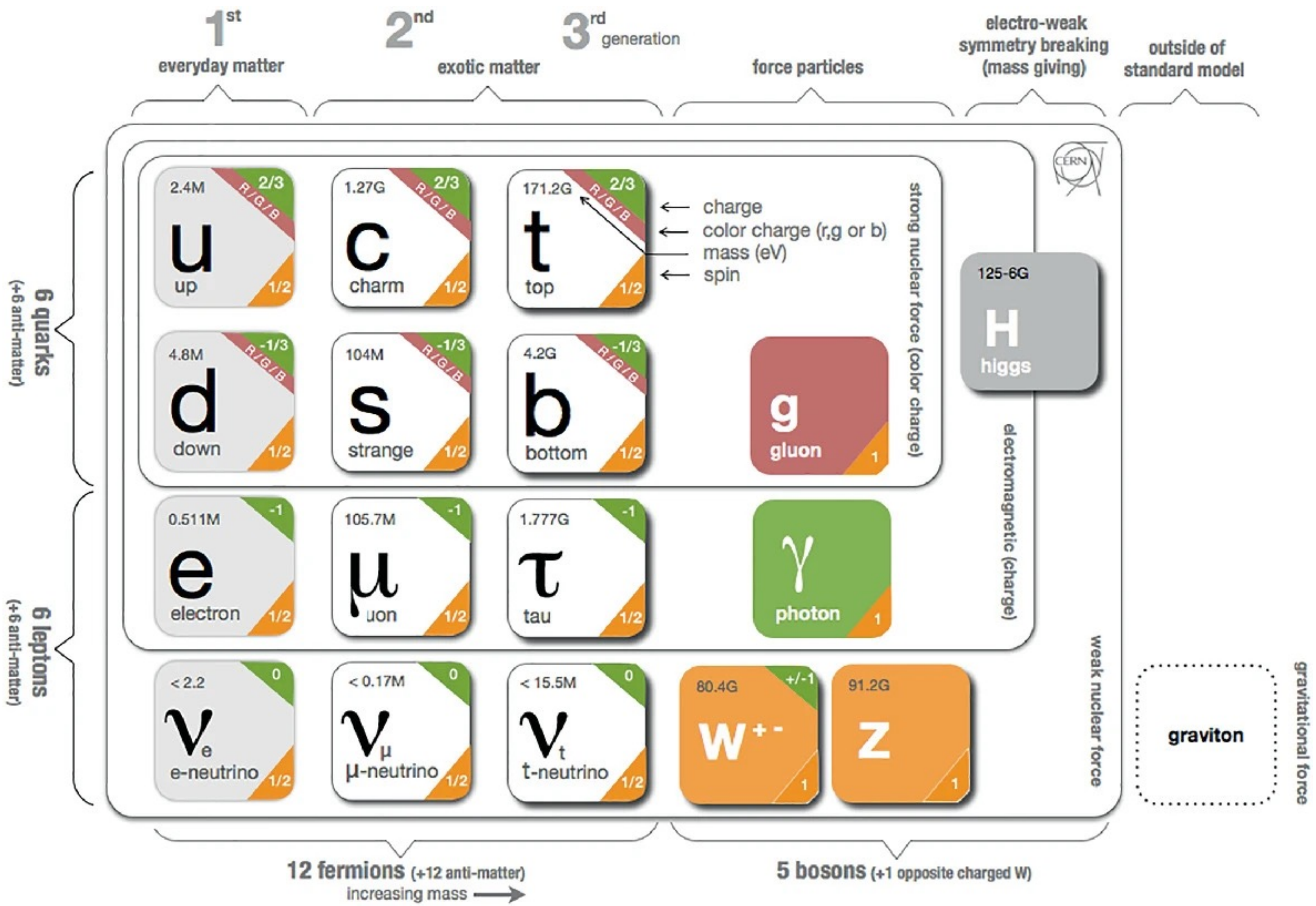
Analysis. CMS detector



Analysis. Particle-flow reconstruction



Analysis. The standard model of particle physics

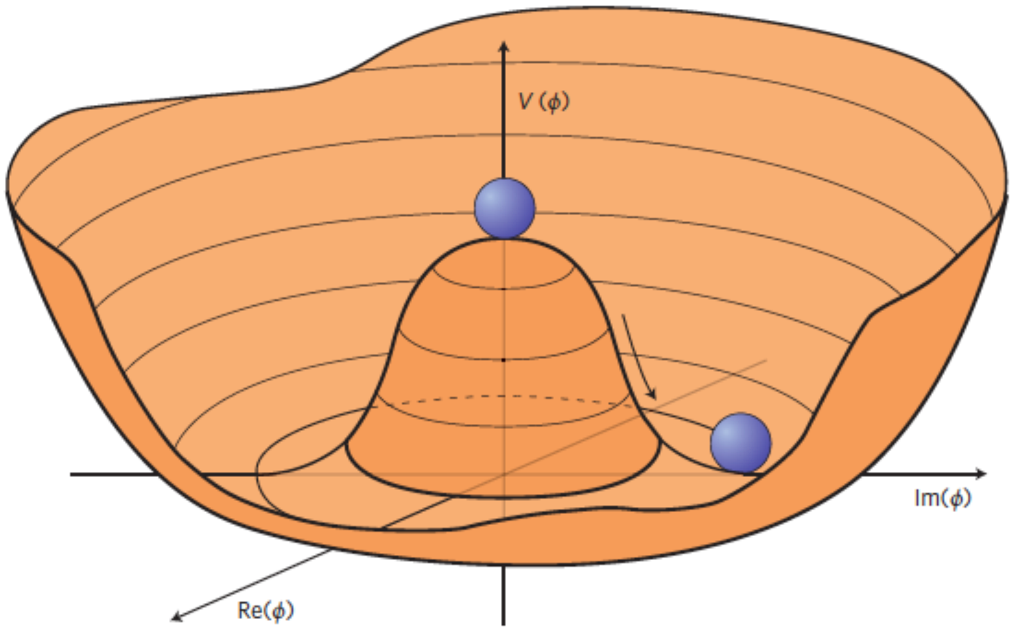


SM

- Fermions are matter particles
- Gauge bosons are force carriers (strong, weak, and EM forces)
- Higgs mechanism gives rise mass

But how about
dark matter, hierarchy, gravity,... ?

Analysis. Higgs mechanism



- The scalar sector of the SM is described by a scalar doublet called the Higgs field ϕ

- The Higgs potential takes the form

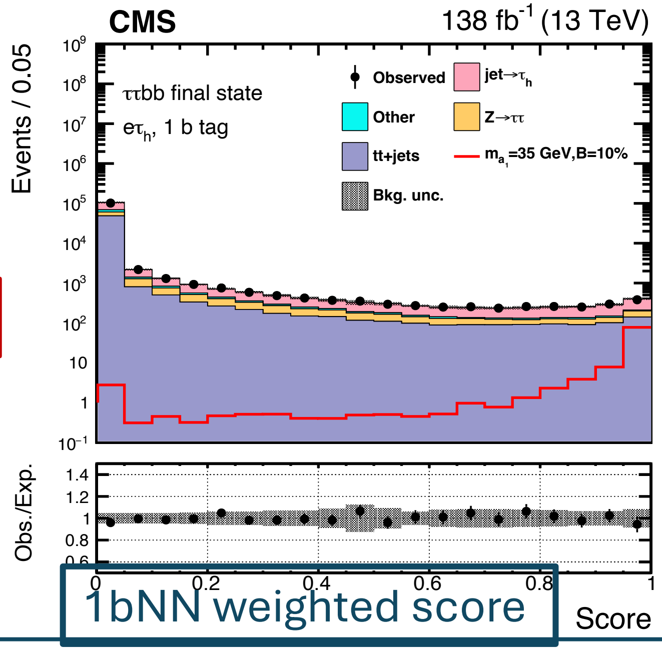
$$V(\phi) = -\mu^2\phi^2 + \lambda\phi^4$$

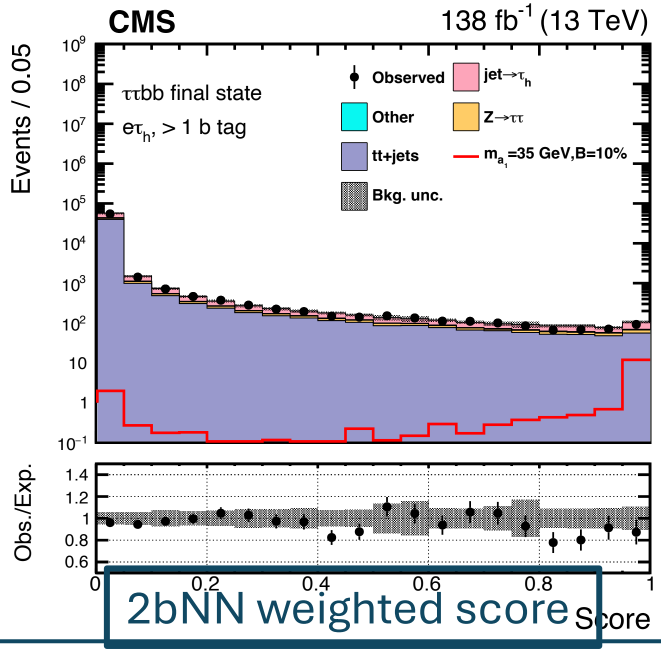
- The Higgs field prefers a nonzero value at the vacuum state $<\phi> \neq 0$
  - W, Z bosons acquire mass from this spontaneous electroweak symmetry breaking
  - Fermions acquire mass via Yukawa coupling

Analysis. DNN classifier performance

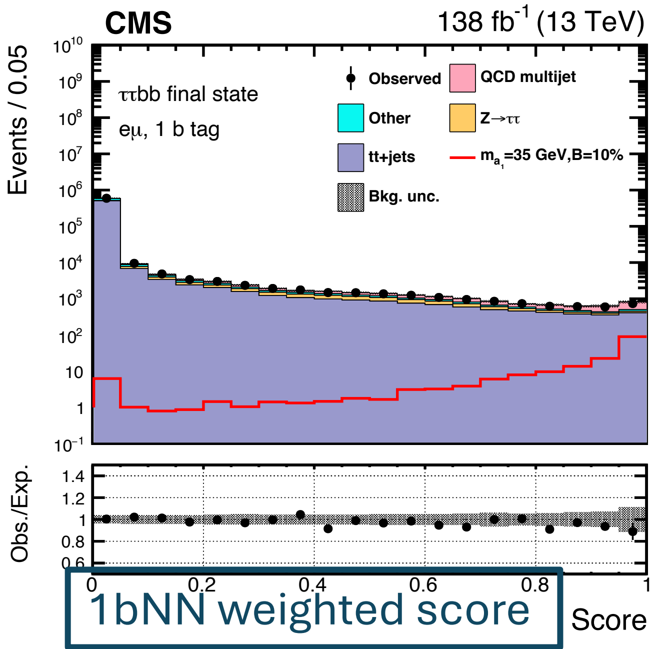
arXiv:2402.13358

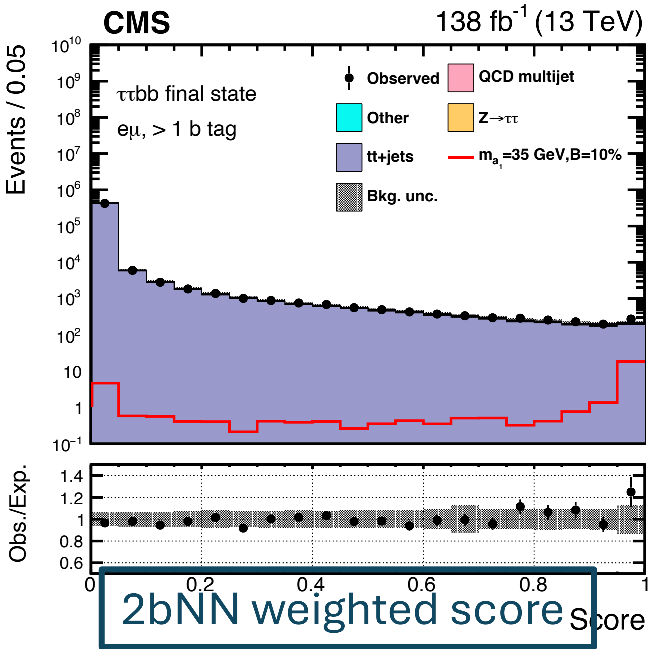
$e\tau_h$





$e\mu$





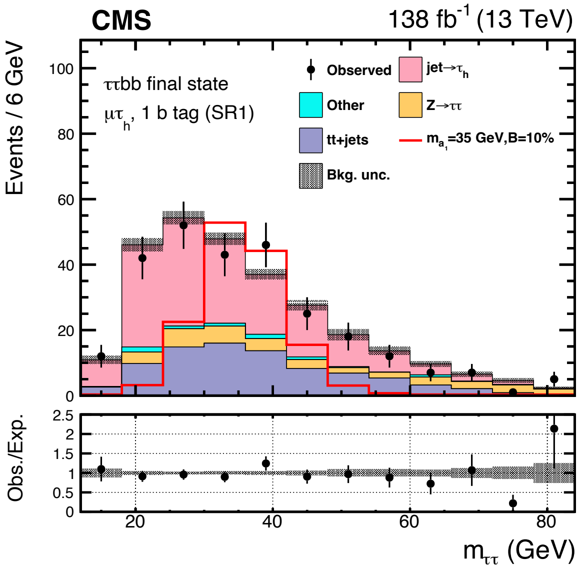
$e\tau_h$ and $e\mu$ final states

- Excellent separation power
- In addition to the "= 1 b-tag" and "> 1 b-tag" bins, further categorization is based on the classifier score
- Category thresholds on the score determined by a grid scan in the signal sensitivity

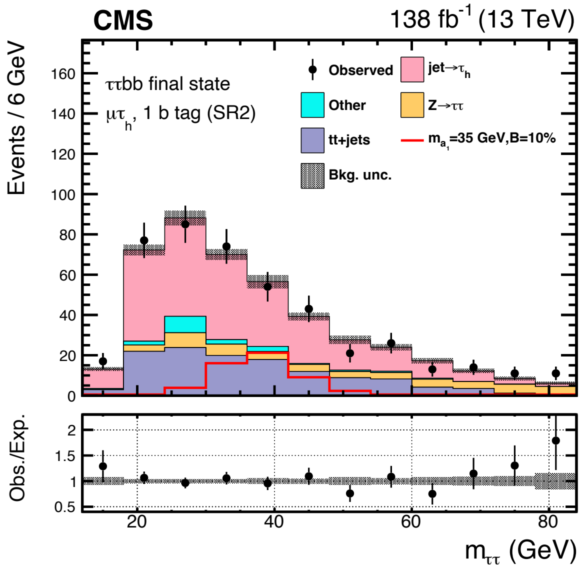
Analysis. Results

$\mu\tau_h$ "1 b-tag" $m_{\tau\tau}$

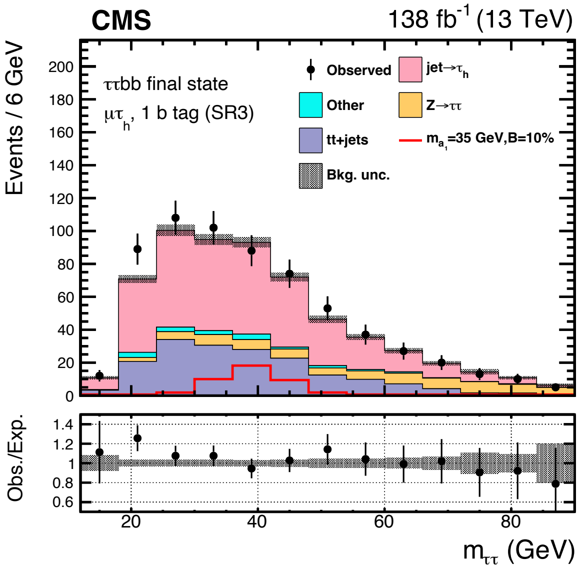
arXiv:2402.13358



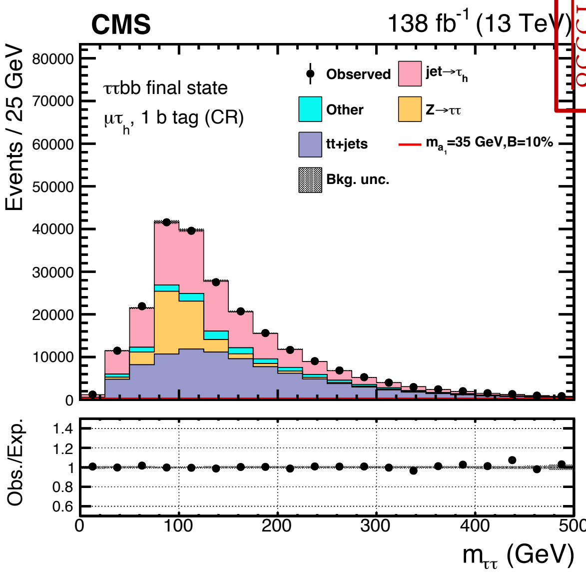
SR1



SR2



SR3



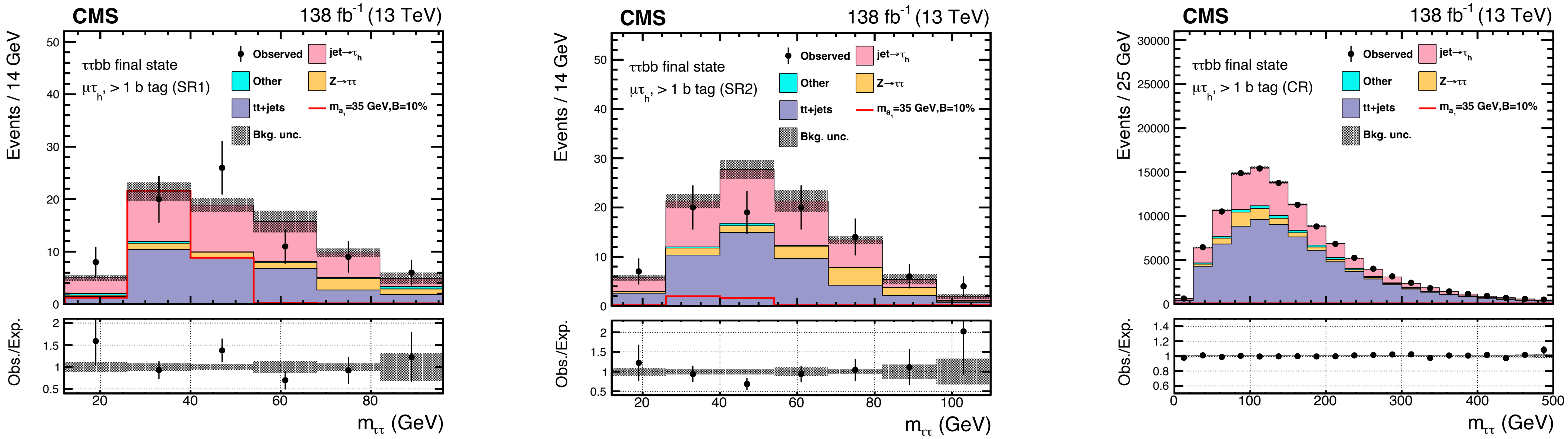
CR

- 7 optimized $m_{\tau\tau}$ distributions in the $\mu\tau_h$ final state
  - "1 b-tag" category: 3 SR + 1 CR
  - "> 1 b-tag" category: 2 SR + 1 CR

- Data agreed with SM prediction across categories, no significant excess in data observed

Analysis. Results

$\mu\tau_h$ "> 1 b-tag" $m_{\tau\tau}$



SR1 SR2 CR

- 7 optimized $m_{\tau\tau}$ distributions in the $\mu\tau_h$ final state
  - "1 b-tag" category: 3 SR + 1 CR
  - "> 1 b-tag" category: 2 SR + 1 CR

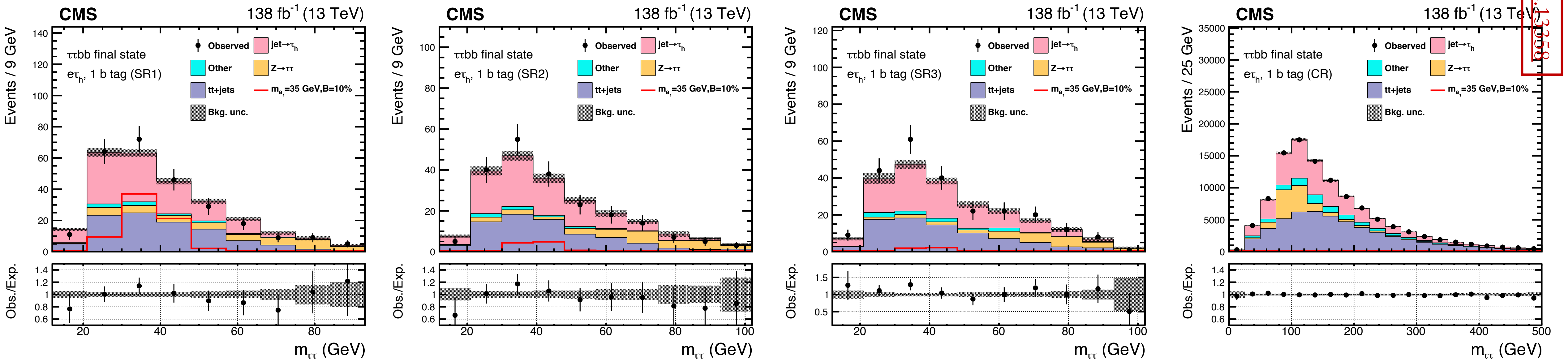
- Data agreed with SM prediction across categories, no significant excess in data observed

arXiv:2402.13358

Analysis. Results

$e\tau_h$ "1 b-tag" $m_{\tau\tau}$

arXiv:2402.13358



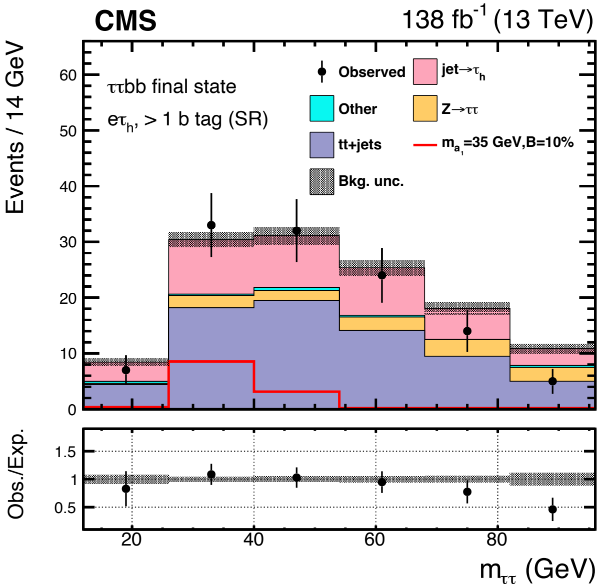
SR1 SR2 SR3 CR

- 6 optimized $m_{\tau\tau}$ distributions in the $e\tau_h$ final state
  - "1 b-tag" category: 3 SR + 1 CR
  - "> 1 b-tag" category: 1 SR + 1 CR

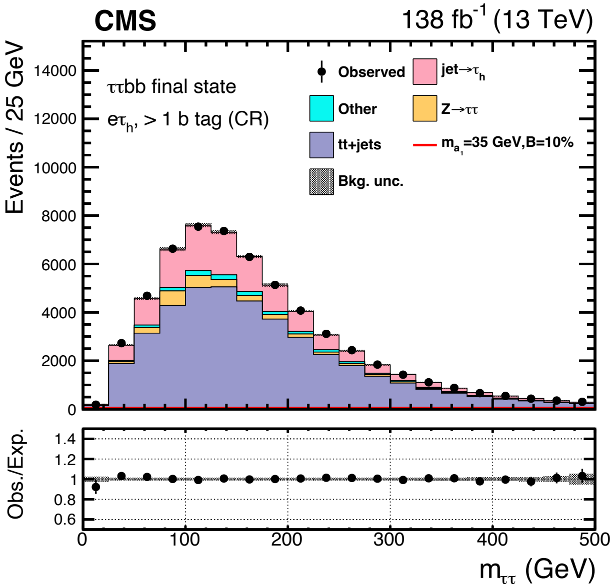
- Data agreed with SM prediction across categories, no significant excess in data observed

Analysis. Results

$e\tau_h$ "> 1 b-tag" $m_{\tau\tau}$



SR



CR

- 6 optimized $m_{\tau\tau}$ distributions in the $e\tau_h$ final state
  - "1 b-tag" category: 3 SR + 1 CR
  - "> 1 b-tag" category: 1 SR + 1 CR

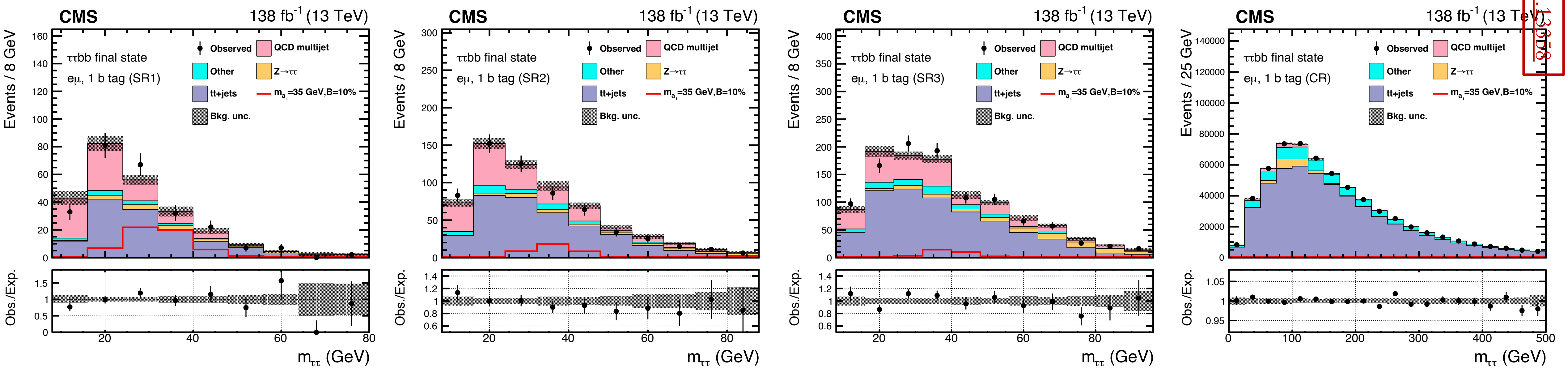
- Data agreed with SM prediction across categories, no significant excess in data observed

arXiv:2402.13358

Analysis. Results

arXiv:2402.13358

$e\mu$ "1 b-tag" $m_{\tau\tau}$



SR1 SR2 SR3 CR

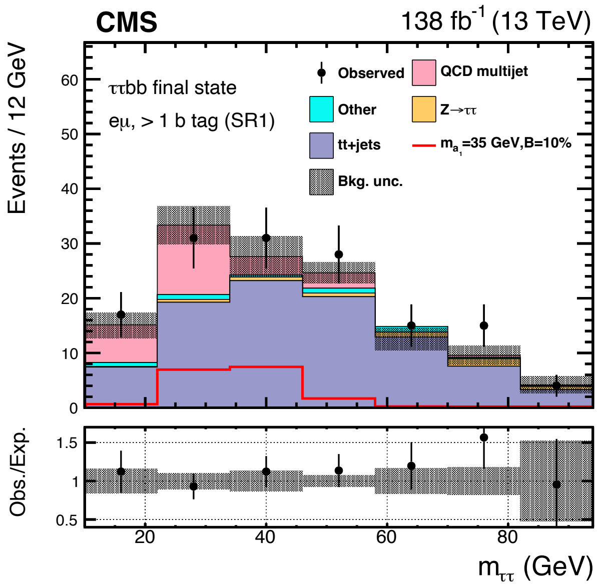
- 7 optimized $m_{\tau\tau}$ distributions in the $e\mu$ final state
  - "1 b-tag" category: 3 SR + 1 CR
  - "> 1 b-tag" category: 2 SR + 1 CR

- Data agreed with SM prediction across categories, no significant excess in data observed

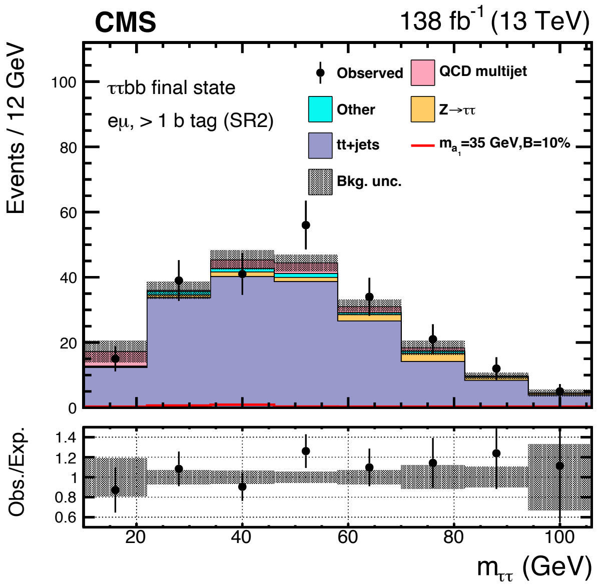
Analysis. Results

arXiv:2402.13358

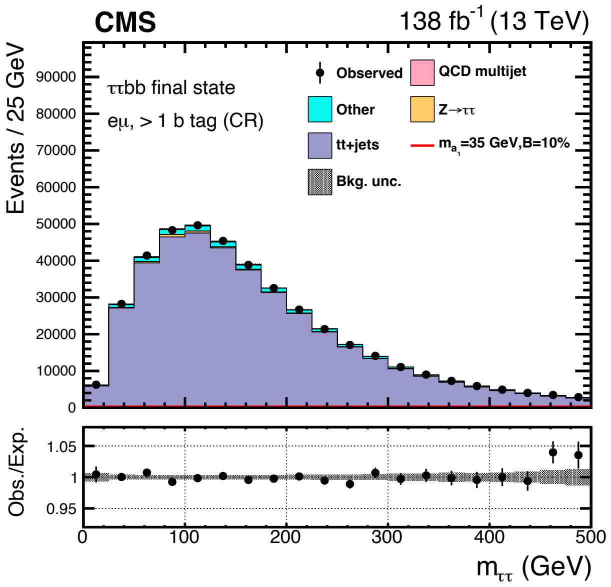
$e\mu$ "> 1 b-tag" $m_{\tau\tau}$



SR1



SR2



CR

- 7 optimized $m_{\tau\tau}$ distributions in the $e\mu$ final state
  - "1 b-tag" category: 3 SR + 1 CR
  - "> 1 b-tag" category: 2 SR + 1 CR

- Data agreed with SM prediction across categories, no significant excess in data observed

Backup – Anomaly trigger

Anomaly trigger. Model compression for L1 constraints


```
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 input (InputLayer)          [(None, 18, 14, 1)]       0

 conv2d_1 (Conv2D)           (None, 18, 14, 20)        200

 relu_1 (Activation)         (None, 18, 14, 20)        0

 pool_1 (AveragePooling2D)   (None, 9, 7, 20)          0

 conv2d_2 (Conv2D)           (None, 9, 7, 30)          5430

 relu_2 (Activation)         (None, 9, 7, 30)          0

 flatten (Flatten)           (None, 1890)              0

 latent (Dense)              (None, 80)                151280

 dense (Dense)               (None, 1890)              153090

 reshape2 (Reshape)          (None, 9, 7, 30)          0

 relu_3 (Activation)         (None, 9, 7, 30)          0

 conv2d_3 (Conv2D)           (None, 9, 7, 30)          8130

 relu_4 (Activation)         (None, 9, 7, 30)          0

 upsampling (UpSampling2D)   (None, 18, 14, 30)        0

 conv2d_4 (Conv2D)           (None, 18, 14, 20)        5420

 relu_5 (Activation)         (None, 18, 14, 20)        0

 output (Conv2D)             (None, 18, 14, 1)         181

=================================================================
Total params: 323,731
Trainable params: 323,731
Non-trainable params: 0
```

Naïve autoencoder model

Encoder (compressor)

Latent space (compressed input)

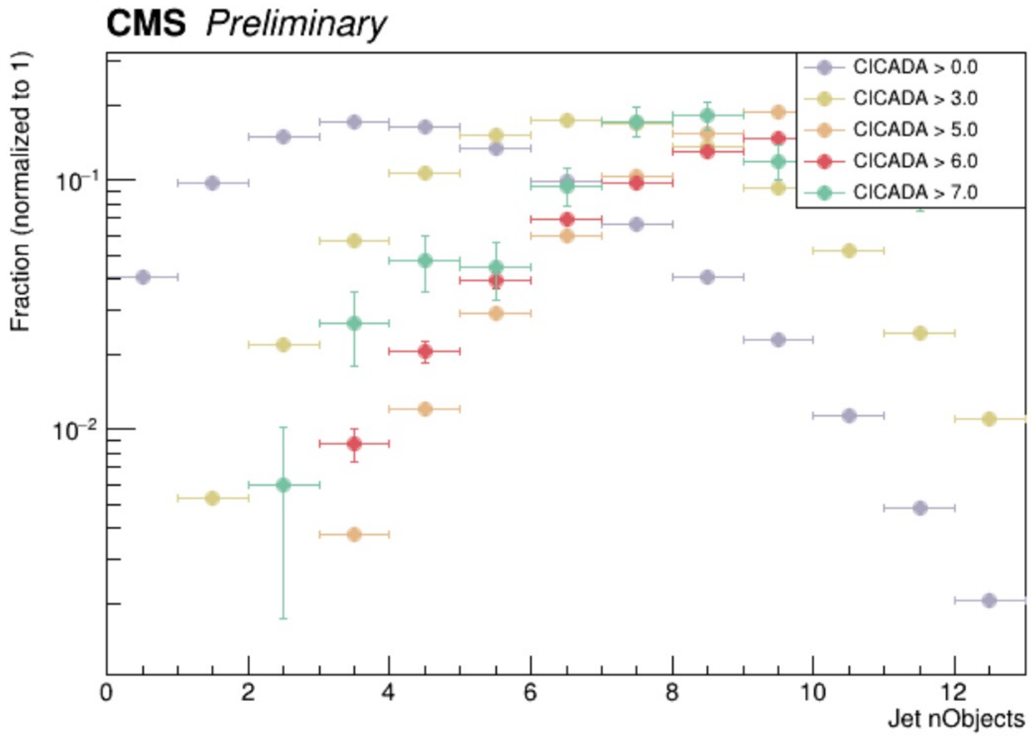
Decoder (decompressor)

Despite good performance, model size of 300k parameters will certainly fail the L1 constraints...

Anomaly trigger. Some trigger objects (Jet)

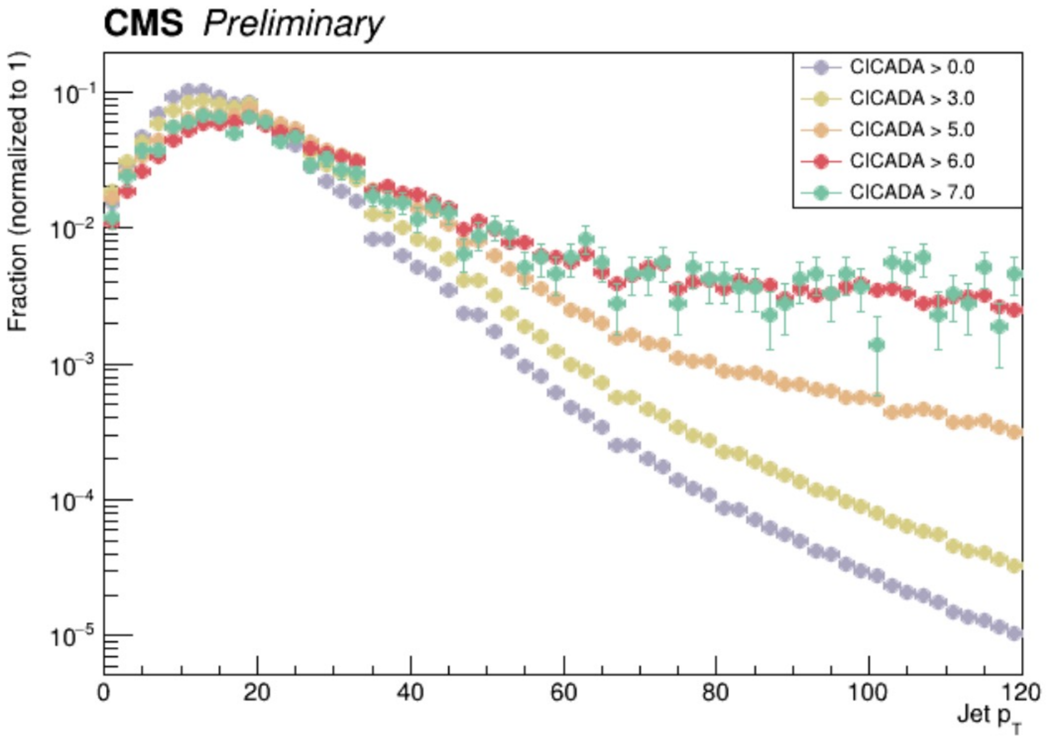
CMS-DP-2023-086

nObjects



- CICADA's most consistent preference is for high jet multiplicity

Pt



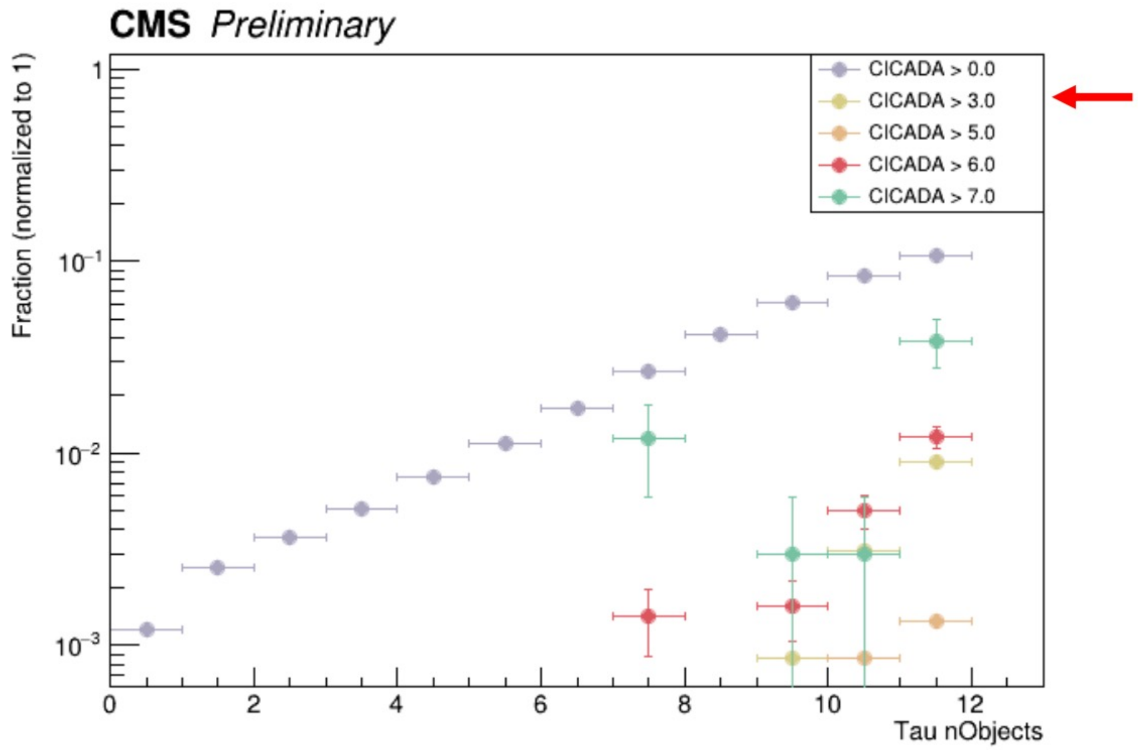
- Slight preference for higher Pt

- Still sensitive to low Pt objects as well

Anomaly trigger. Some trigger objects (Tau)

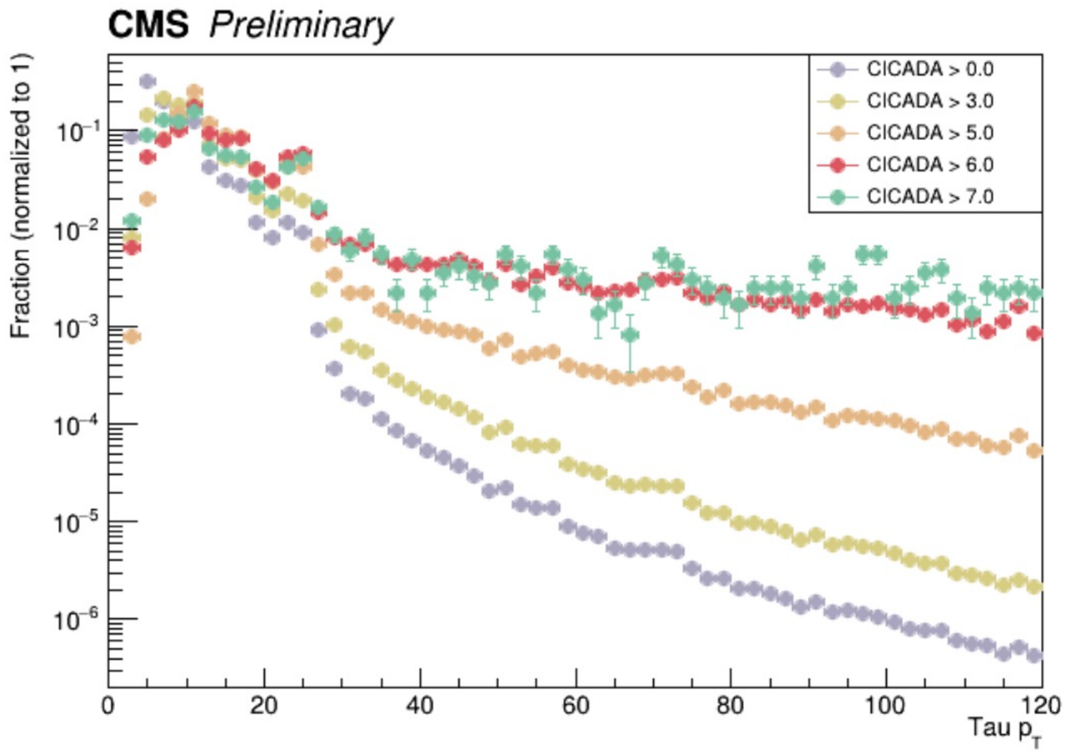
CMS-DP-2023-086

nObjects



- If this looks a bit confusing, it's bad legend placement
- Nearly 97% of high score CICADA events have 12 trigger taus

Pt

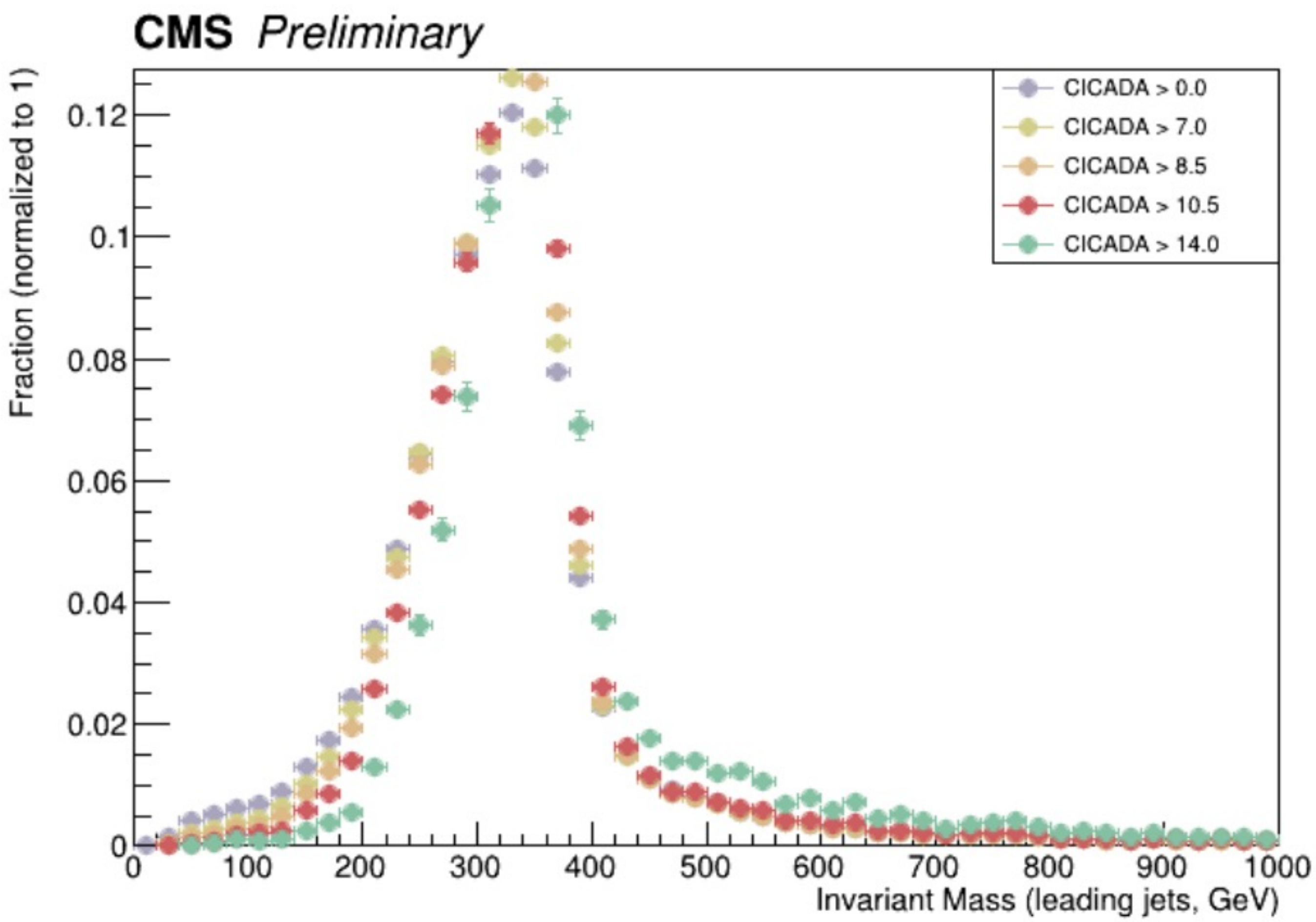


- Similar to other objects, there are slight preferences for higher Pt

- Still sensitive to low Pt objects as well

Anomaly trigger. Signal shaping?

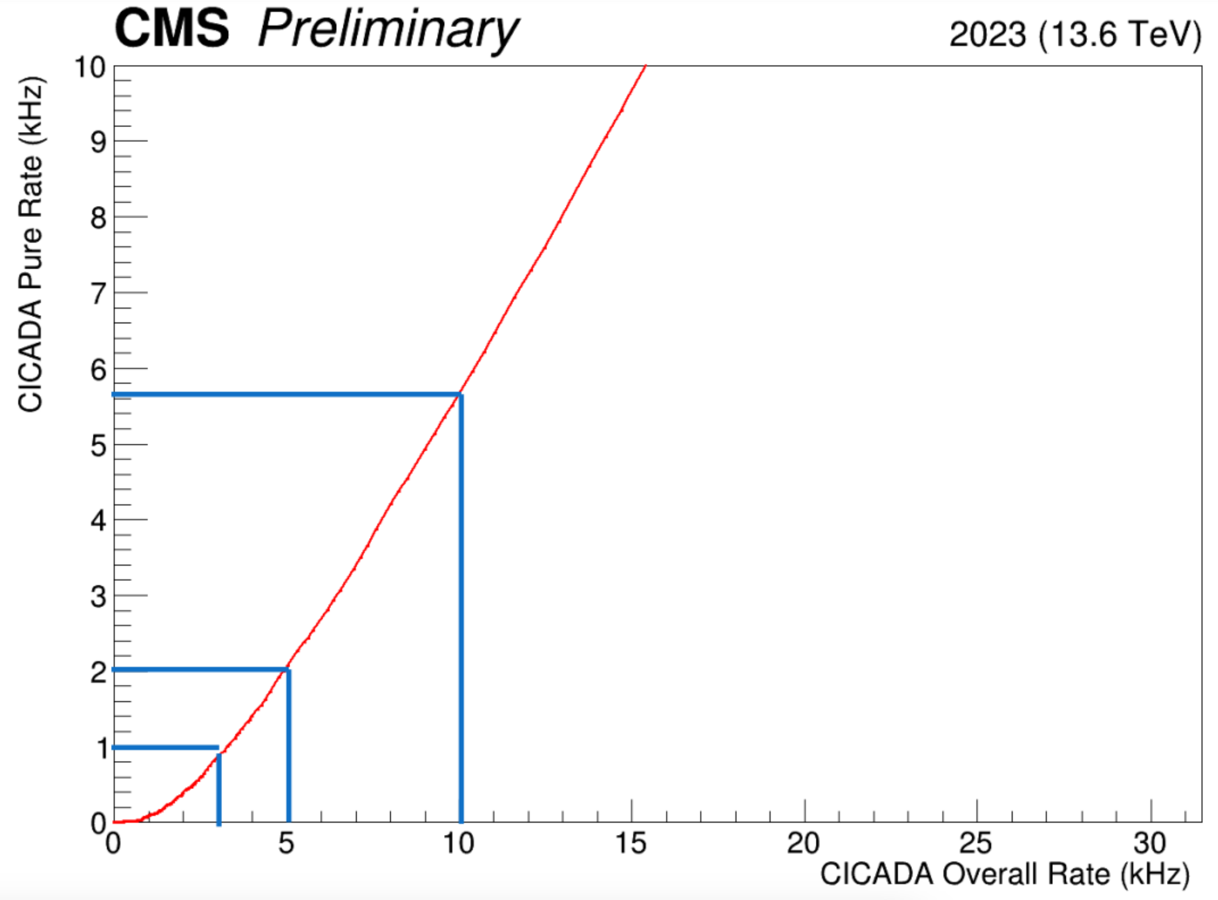
CMS-DP-2023-086

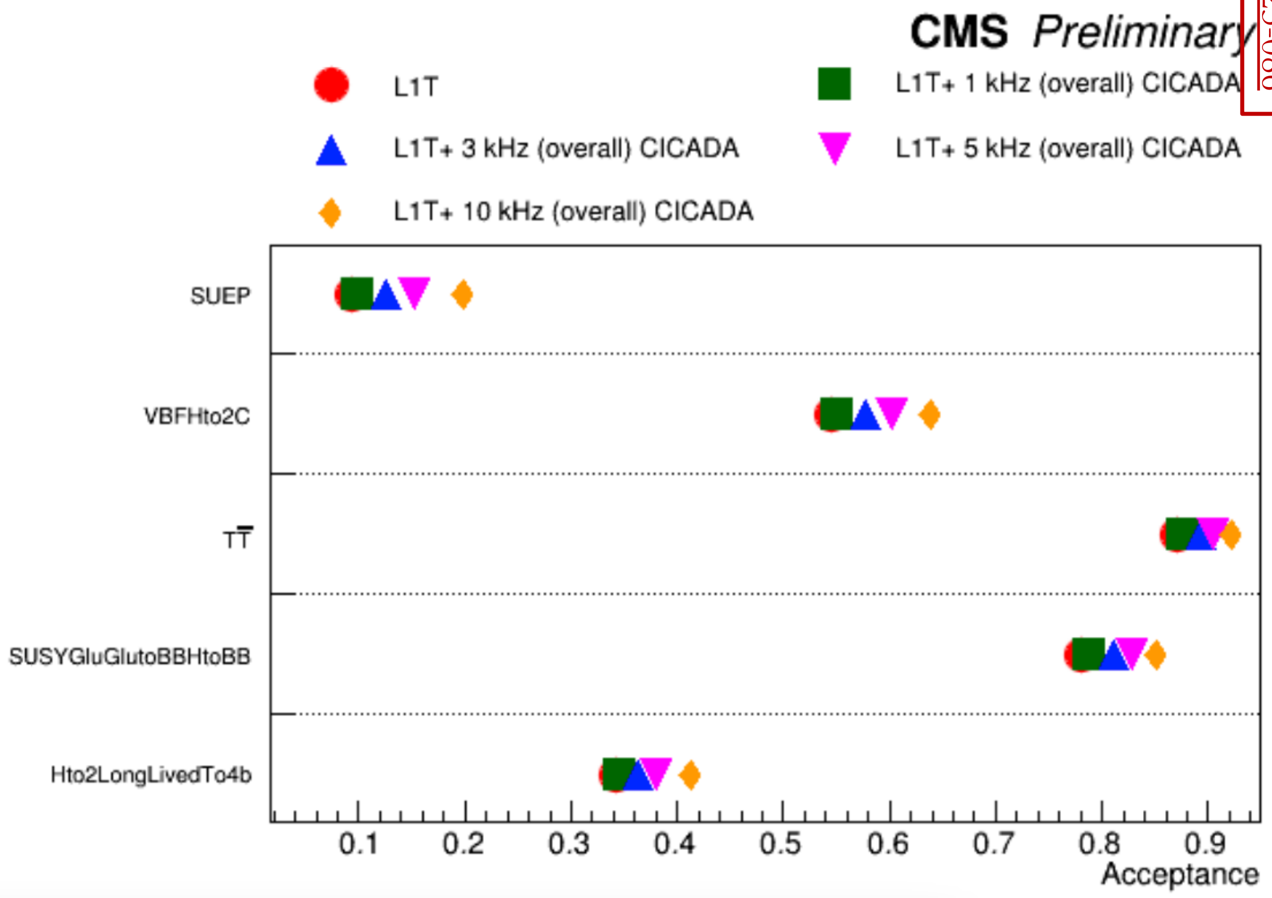


Example: /SUSYGluGlutoBBHtoBB_NarrowWidth_M-350_TuneCP5_13p6TeV_pythia8/Run3Winter23MiniAOD126X_mcRun3_2023_forPU65_v1-v2/MINIAODSIM

- Invariant mass of 2 leading jets

Anomaly trigger. Adding to the current triggers

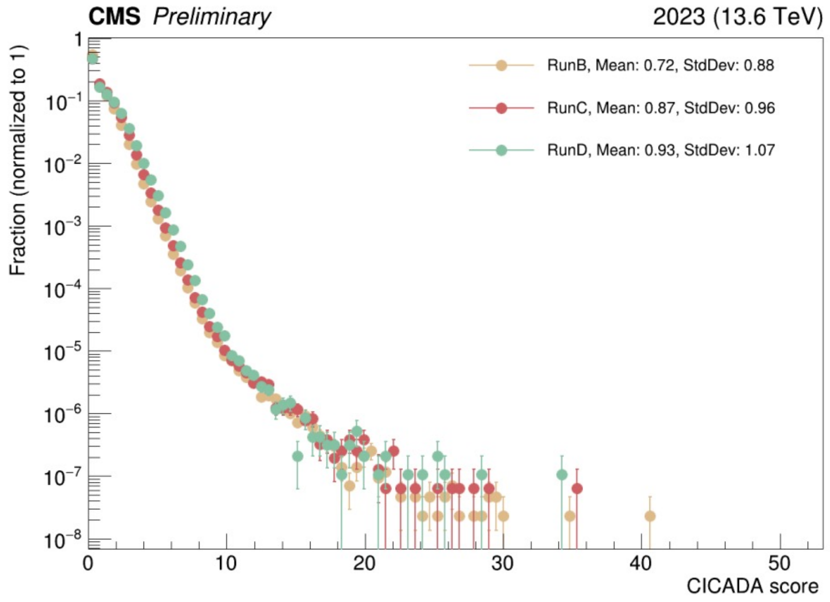


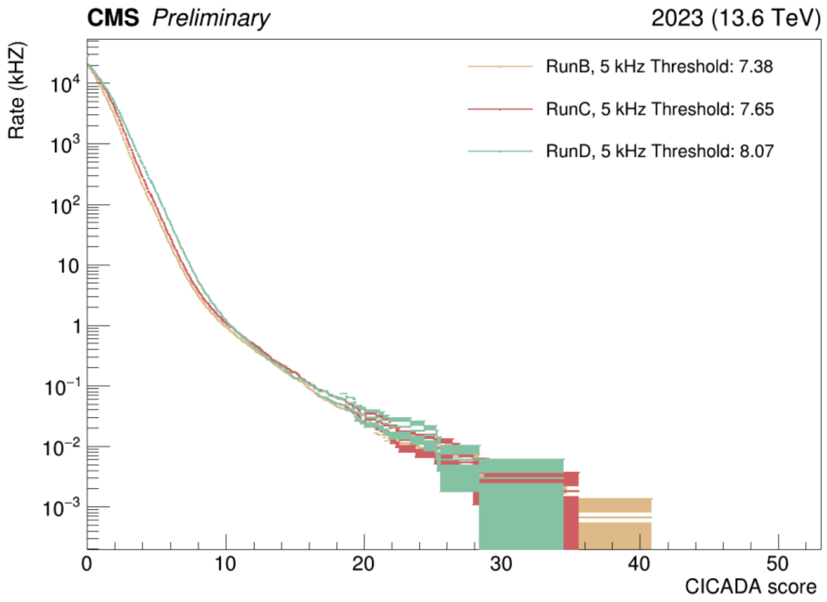


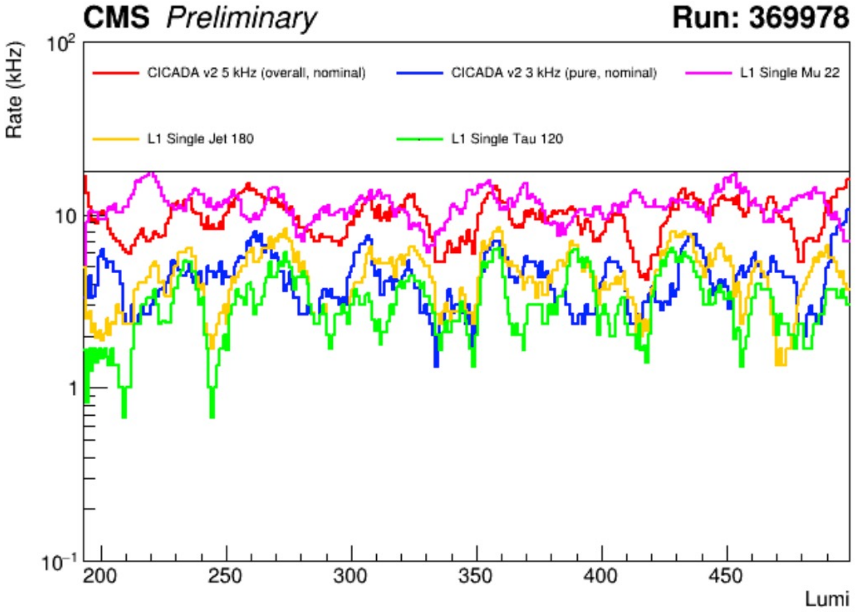
CMS-DP-2023-086

Anomaly trigger. Stability as a trigger

CMS-DP-2023-086



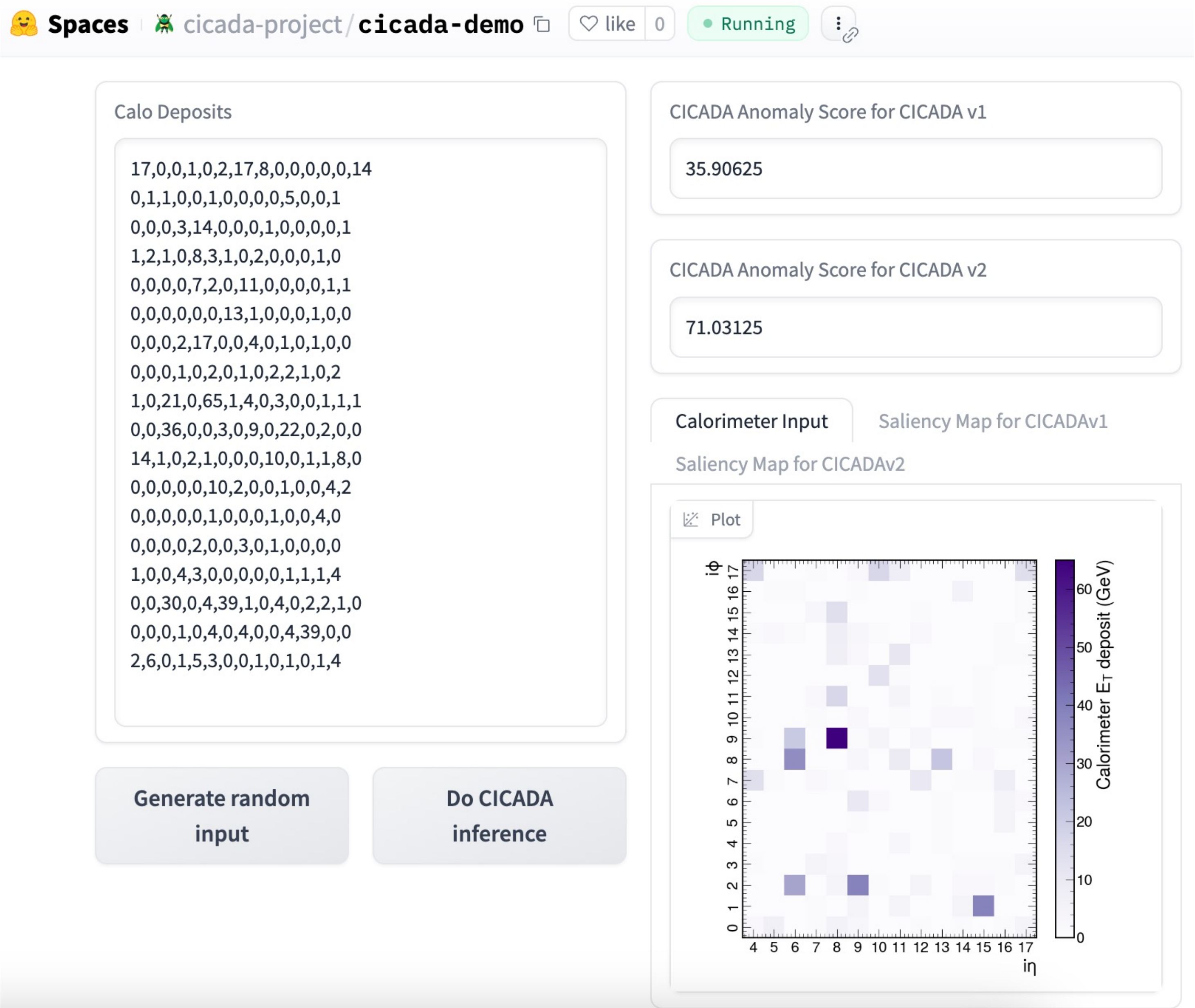




- Flexible trigger: tunable threshold for rate control
- Stable in a run and across runs

Anomaly trigger. CICADA public demo

CMS-DP-2023-086



https://cicada.web.cern.ch/

Backup – Symbolic regression

Symbolic regression. Latency-aware training

Latency = # of clock cycles (cc)

- An alternative resource-saving approach: latency-aware training
  - Define operator complexity by its # of cc required on FPGA, and incorporate this time cost into the model training
  - E.g., $\tan(\cdot)$ will be penalized more than $\sin(\cdot)$
  - Models are trained balancing between accuracy and latency

Operator	# of cc
$+$	1
$-$	1
$\times$	1
log(abs(·))	4
sin(·)	8
tan(·)	48
cosh(·)	8
sinh(·)	9
exp(·)	3

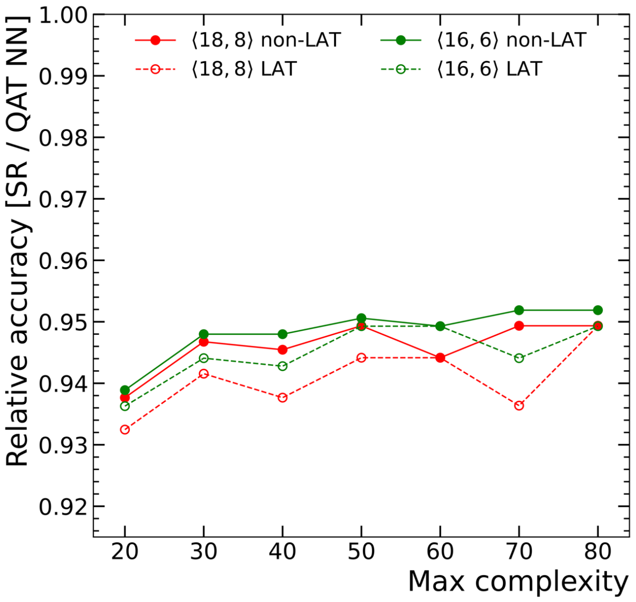
Evaluated on a Xilinx VU9P FPGA

arXiv:2305.04099

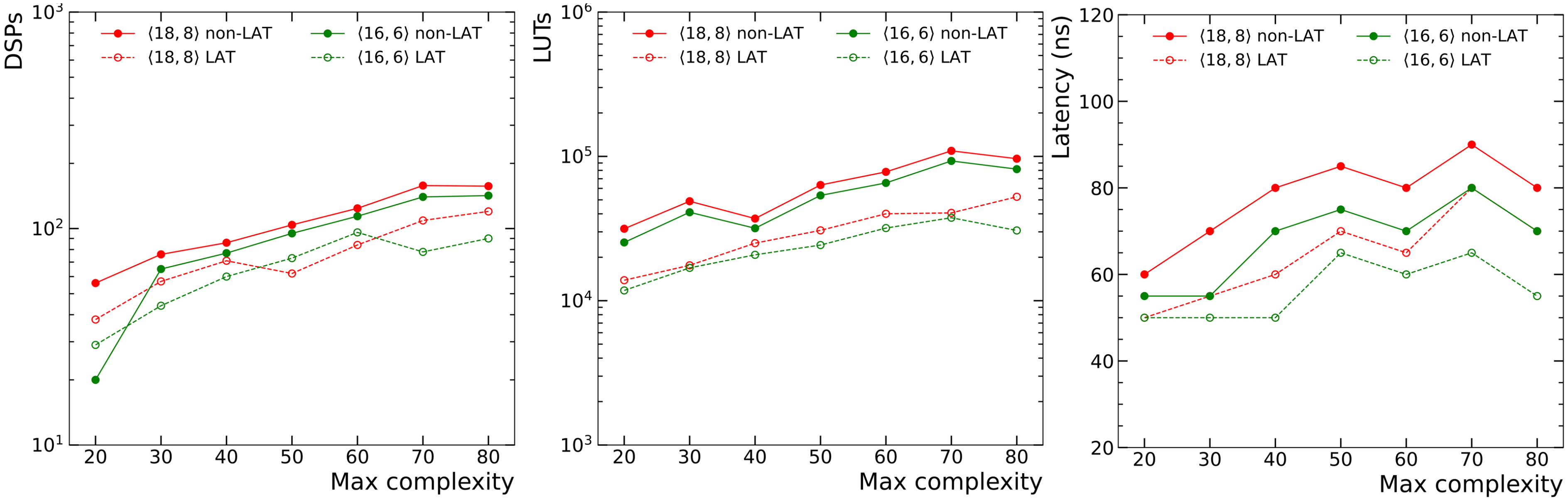
Symbolic regression. Latency-aware training

Operator complexity	Expression for the t tagger with $c_{\max} = 40$	AUC
All 1's (PySR default)	$0.11(C_1^{\beta=1} + C_1^{\beta=2} + \log(\mathrm{abs}(C_1^{\beta=2}))) - 0.48 m_{\mathrm{mMDT}} - 0.05\,\mathrm{Multiplicity}(\mathrm{Multiplicity} + \log(\mathrm{abs}(m_{\mathrm{mMDT}})))$ $-\sin(-C_1^{\beta=2} + 0.14 C_2^{\beta=1} m_{\mathrm{mMDT}}) + 0.11\sinh(C_1^{\beta=1}) - 0.24$	0.930
No. of clock cycles at $\langle 16, 6\rangle$	$0.04((\sum z\log z) + C_1^{\beta=1} + C_2^{\beta=1} - m_{\mathrm{mMDT}} - (\mathrm{Multiplicity} - 0.2)(\mathrm{Multiplicity} + \log(\mathrm{abs}(C_1^{\beta=2}))))$ $-\sin(-C_1^{\beta=1} - C_1^{\beta=2} + 1.23 m_{\mathrm{mMDT}} + 0.58)$	0.924
No. of clock cycles at $\langle 18, 8\rangle$	$0.04\,\mathrm{Multiplicity}(C_1^{\beta=2}(C_1^{\beta=2} - m_{\mathrm{mMDT}}) - \mathrm{Multiplicity} - \log(\mathrm{abs}(C_1^{\beta=2}((\sum z\log z) + 0.23))))$ $-\sin(-C_1^{\beta=1} - C_1^{\beta=2} + 1.19 m_{\mathrm{mMDT}} + 0.61)$	0.926

Table 3. Expressions generated by PySR for the t tagger with $c_{\max} = 40$, implemented with and without LAT. Constants are rounded to two decimal places for readability.



Accuracy



Resources and latency

Symbolic regression. A compact symbolic model for MNIST

TABLE II

AN EXAMPLE OF A COMPACT SYMBOLIC MODEL WITH **a mean complexity of 90 and an overall accuracy of 80%**, LEARNED BY SYMBOLNET IN THE MNIST DATASET. CONSTANTS ARE ROUNDED TO 2 SIGNIFICANT FIGURES FOR THE PURPOSE OF DISPLAY.

Class	Expression (symbolic model classifying MNIST digits)	Complexity	AUC
0	$0.094\sin(0.41x_{374}-0.53x_{378}+0.66x_{484})+0.15(-0.3x_{184}-0.17x_{239}-0.12x_{269}-0.27x_{271}-0.14x_{318}+0.72x_{352}-0.6x_{358}-0.19x_{374}+0.55x_{377}-0.32x_{415}-0.23x_{456}-0.26x_{485}-0.4x_{510}-0.53x_{627}+0.25x_{637}-0.19x_{658}+0.55x_{711})(0.44x_{102}-0.29x_{156}-0.41x_{212}-0.29x_{271}-0.22x_{302}-0.11x_{371}-0.5x_{398}-0.41x_{428}-0.24x_{430}+0.84x_{433}+0.6x_{436}+0.11x_{462}+0.62x_{490}-0.45x_{509}-0.066x_{539}-0.4x_{541}-0.13x_{568}+0.22x_{580}-0.58x_{627}-0.25x_{658})$	129	0.993
1	$\exp(-26.0(0.11x_{102}+0.056x_{158}+0.21x_{176}+0.08x_{178}+0.093x_{182}+0.93x_{205}+0.11x_{212}+0.15x_{235}+0.27x_{248}-0.033x_{267}+0.067x_{271}+0.24x_{302}-0.063x_{323}+0.095x_{327}-0.067x_{350}-0.12x_{378}+0.18x_{430}+x_{438}-0.067x_{462}-0.092x_{489}+0.18x_{510}-0.024x_{568}+0.12x_{580}+0.23x_{637}+0.24x_{711}+0.13x_{713}+0.24x_{715}+0.27x_{96}+0.28)^2)$	91	0.995
2	$0.54\sin(0.59x_{124}+0.35x_{156}-0.39x_{318}-0.41x_{350}-0.46x_{371}-0.41x_{374}-0.6x_{415}+0.18x_{431}+0.14x_{465}+1.1x_{473}+0.7x_{509}+0.38x_{515}+0.88x_{528}+0.38x_{554}+0.77x_{611}+0.39x_{637}+0.1x_{99}-0.8)+0.53$	58	0.966
3	$-0.042x_{158}+0.062x_{178}-0.039x_{235}-0.12x_{291}-0.063x_{316}+0.045x_{318}+0.061x_{404}-0.066x_{458}+0.032x_{485}-0.1x_{487}-0.074x_{489}-0.12x_{490}+0.038x_{515}+0.036x_{517}-0.06x_{541}+0.36x_{563}-0.043x_{572}+0.048x_{611}+0.28$	56	0.907
4	$0.76\exp(-4.7(0.47x_{124}+0.42x_{126}+0.49x_{128}+0.14x_{176}+0.28x_{182}+0.44x_{184}+0.17x_{212}+x_{239}+0.88x_{267}+0.81x_{322}+0.43x_{323}+0.33x_{350}+0.4x_{543}+0.3x_{554}+0.5x_{568}+0.35x_{623})^2)-0.082\times(-0.2x_{124}-0.34x_{182}+0.39x_{429}-0.69x_{568}-0.66x_{713}+0.68)(1.4x_{102}+0.58x_{182}+0.75x_{208}+0.51x_{215}+0.29x_{235}+0.47x_{322}-0.53x_{323}-0.7x_{325}+0.23x_{355}+0.53x_{358}-1.4x_{374}-1.5x_{398}-0.63x_{431}-1.5x_{456}-0.68x_{462}-1.1x_{465}+0.48x_{541}+0.83x_{568}+4.9x_{66}+1.3x_{71}+1.3x_{713}+1.4x_{96})$	141	0.980
5	$\exp(-2.4(-0.15x_{124}+0.13x_{158}+0.59x_{190}+0.98x_{248}-0.13x_{267}-0.35x_{323}-0.68x_{325}-x_{327}-0.78x_{355}+0.17x_{404}-0.5x_{456}-0.19x_{490}-0.41x_{510}-0.6x_{515}+0.15x_{568}-0.63)^2)-0.012x_{128}-0.12x_{358}+0.03x_{371}+0.069x_{374}-0.031x_{436}-0.019x_{485}+0.042x_{580}+0.026x_{623}$	79	0.928
6	$0.21x_{102}+0.3x_{103}+0.42x_{107}-0.054x_{215}-0.057x_{217}-0.093x_{269}-0.065x_{271}-0.068x_{302}-0.08x_{322}+0.068x_{358}+0.04x_{374}+0.12x_{414}+0.021x_{431}+0.069x_{485}-0.063x_{489}-0.078x_{510}+0.081x_{515}+0.047x_{543}-0.056x_{568}+0.065x_{572}-0.05x_{580}+0.35x_{64}+0.43x_{66}+0.22x_{68}+0.34x_{69}+0.29x_{71}+0.35x_{73}+0.56x_{78}+0.18x_{99}+0.1$	89	0.975
7	$0.98\exp(-3.1(-x_{124}-0.61x_{126}-0.81x_{128}-0.97x_{156}-0.24x_{184}-0.23x_{350}+0.073x_{355}-0.28x_{376}-0.13x_{377}-0.62x_{378}-0.72x_{404}-0.6x_{415}-0.62x_{431}-0.092x_{433}-0.43x_{458}-0.87x_{485}-0.94x_{539}-0.27x_{541}-0.84x_{581}-0.37x_{623})^2)$	68	0.968
8	$-0.68\sin(0.16x_{156}-0.35x_{176}+0.43x_{302}+0.19x_{318}+0.23x_{327}+0.41x_{376}-0.2x_{414}-0.4x_{428}+0.46x_{433}-0.33x_{467}+0.27x_{487}+0.3x_{515}-0.34x_{528}+0.25x_{541}+0.58x_{658}+0.43x_{689}+1.1)+0.64$	55	0.924
9	$-0.051\sin(0.59x_{124}+0.35x_{156}-0.39x_{318}-0.41x_{350}-0.46x_{371}-0.41x_{374}-0.6x_{415}+0.18x_{431}+0.14x_{465}+1.1x_{473}+0.7x_{509}+0.38x_{515}+0.88x_{528}+0.38x_{554}+0.77x_{611}+0.39x_{637}+0.1x_{99}-0.8)-0.054x_{126}-0.066x_{158}-0.082x_{190}-0.11x_{205}+0.059x_{208}+0.016x_{215}-0.039x_{217}-0.0092x_{235}-0.11x_{248}-0.047x_{271}+0.093x_{316}-0.04x_{322}+0.069x_{327}+0.07x_{352}-0.059x_{414}+0.069x_{429}+0.038x_{431}+0.048x_{436}-0.057x_{467}-0.044x_{517}-0.067x_{541}+0.065x_{637}-0.06x_{658}+0.11x_{711}+0.1x_{713}+0.16x_{715}+0.029$	136	0.926

- Symbolic model can be ultra compact in size while having competitive performance

- Can visualize the entire model in a single table, as compared to black box NN with thousands or more parameters

arXiv:2401.09949