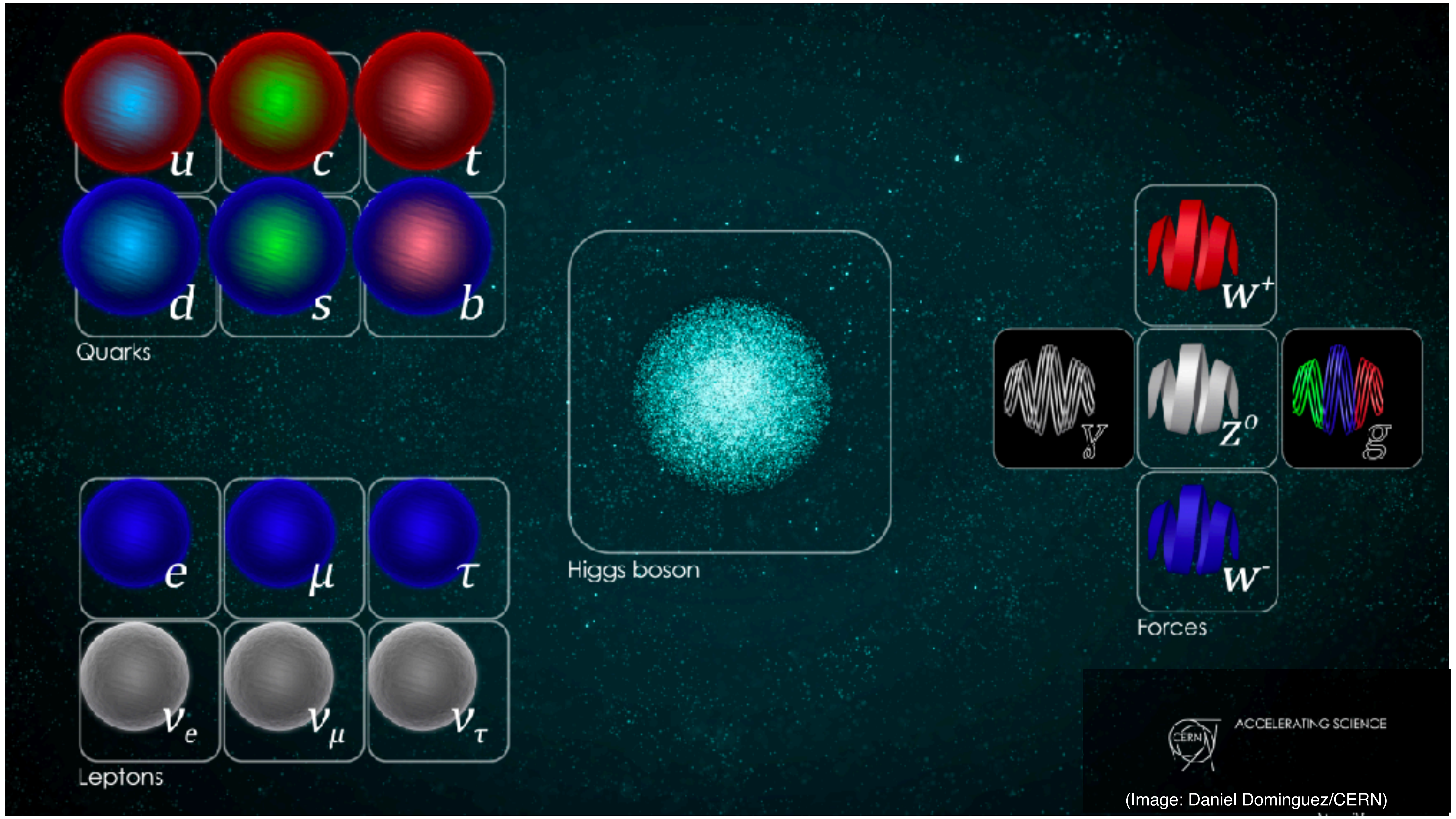
Standard model four-top production at 13 TeV in the all-hadronic final state with CMS Run II data

Melissa Quinnan

Tuesday June 21, 2022

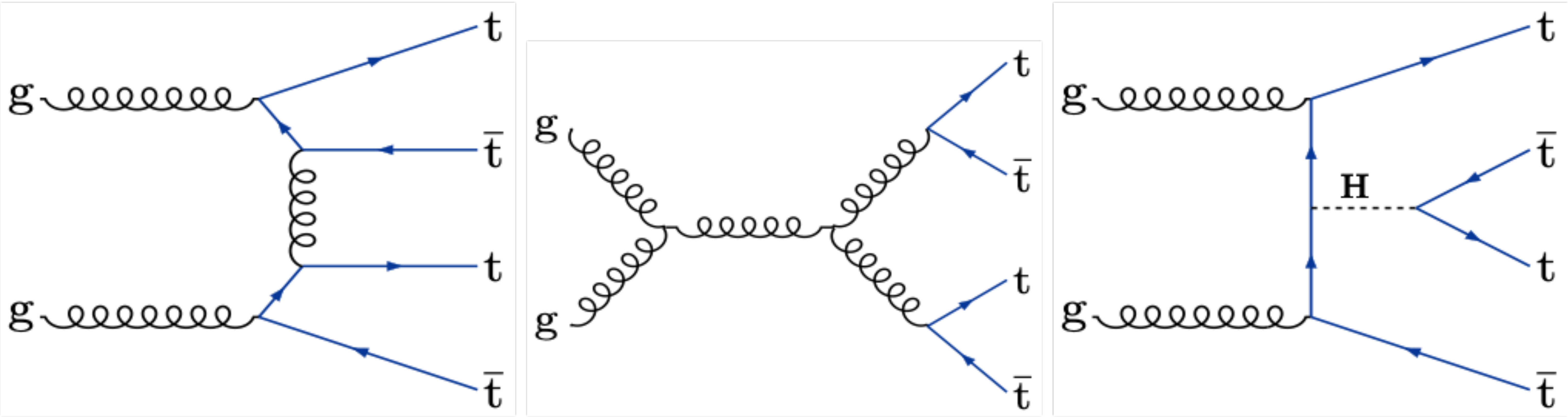
Goals

- Objective of this analysis:
  - Probe an important SM process (four-top production)
- Goals of this presentation:
  - Provide example of how particle physics analyses are conducted at detectors like CMS
  - Showcase potential of machine-learning techniques to improve such analyses



Big picture goal: test the standard model to better understand its properties & potentially reveal new physics!

Four-Top Production

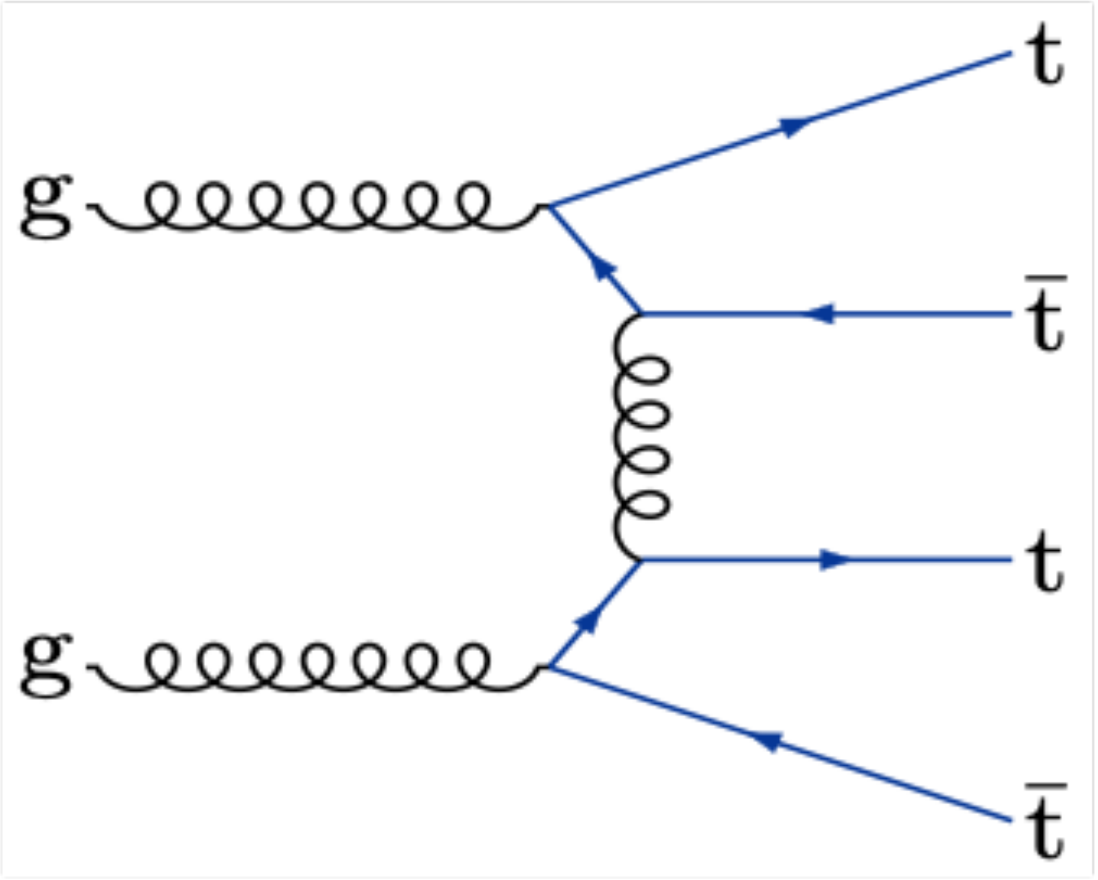


- Rare Standard Model (SM) process with ~12 fb (13 TeV NLO QCD + EW)*
- Important test of SM
- Enhancement of cross section predicted by some beyond-the-Standard-Model (BSM) physics
  - ex: extra BSM Higgs-like bosons decaying to tops

*Rikkert Frederix, Davide Pagani, Marco Zaro. arXiv:1711.02116

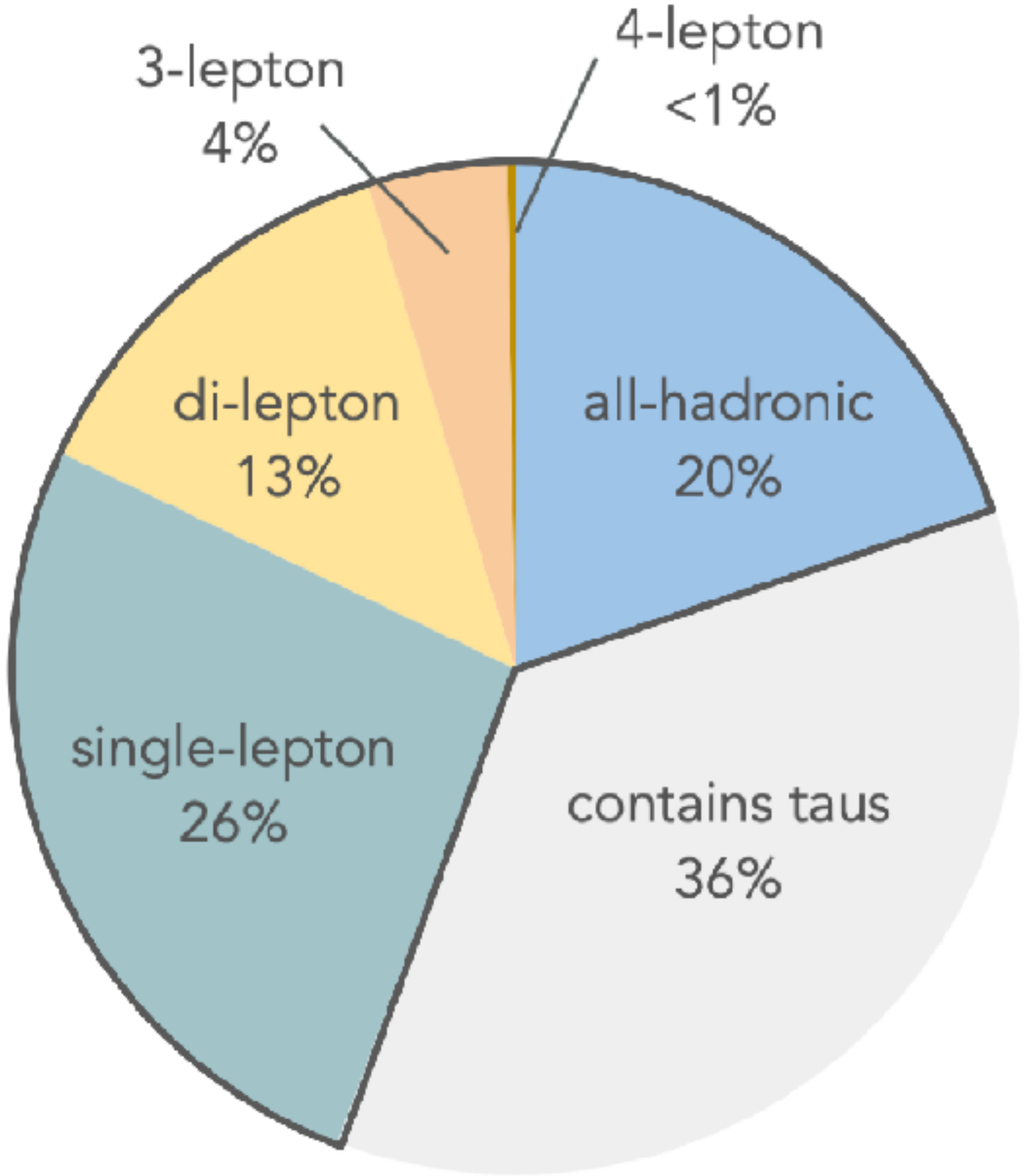
Four -Top Final States

- Top quarks decay into b-quark and a W boson, which in turn decays either leptonically or hadronically
- Multiple possible final states (“channels”) depending on how each top decays



“Leptonic” top decay
- decays to lepton and a neutrino

“Hadronic” top decay
- decays to jets



Published Results

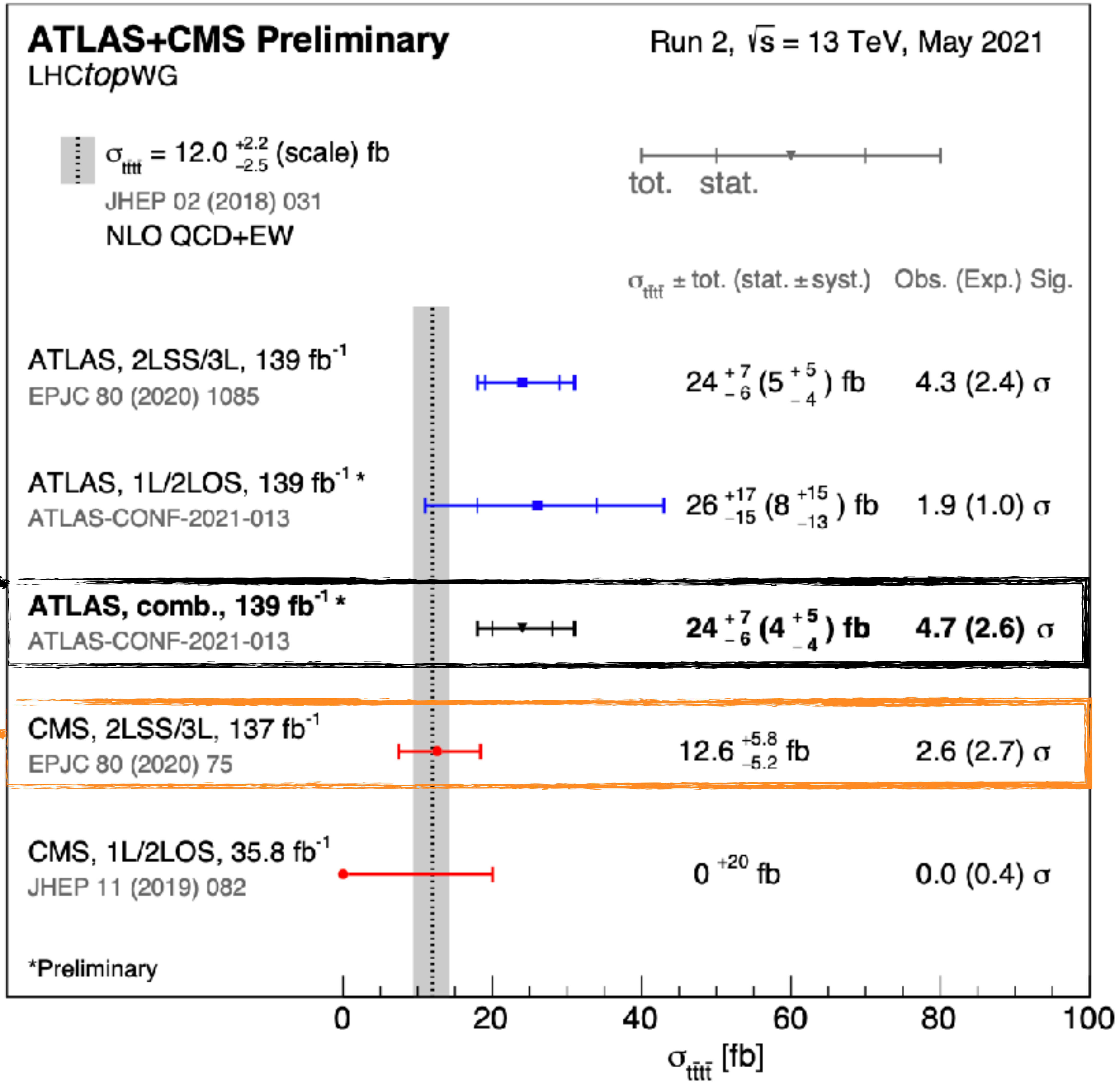
σ = standard deviations
3σ = "evidence"
5σ = "discovery"

ATLAS:

Recent Run II ATLAS Results*
Combined multiple final states:
Same-Sign Di-Lepton + Multi-Lepton (2LSS/3L) + Single Lepton (1L) + Opposite-Sign Di-Lepton (2LOS)

CMS:

Same-Sign 2-Lepton + Multi-Lepton (2LSS/3L) published Run II analysis+

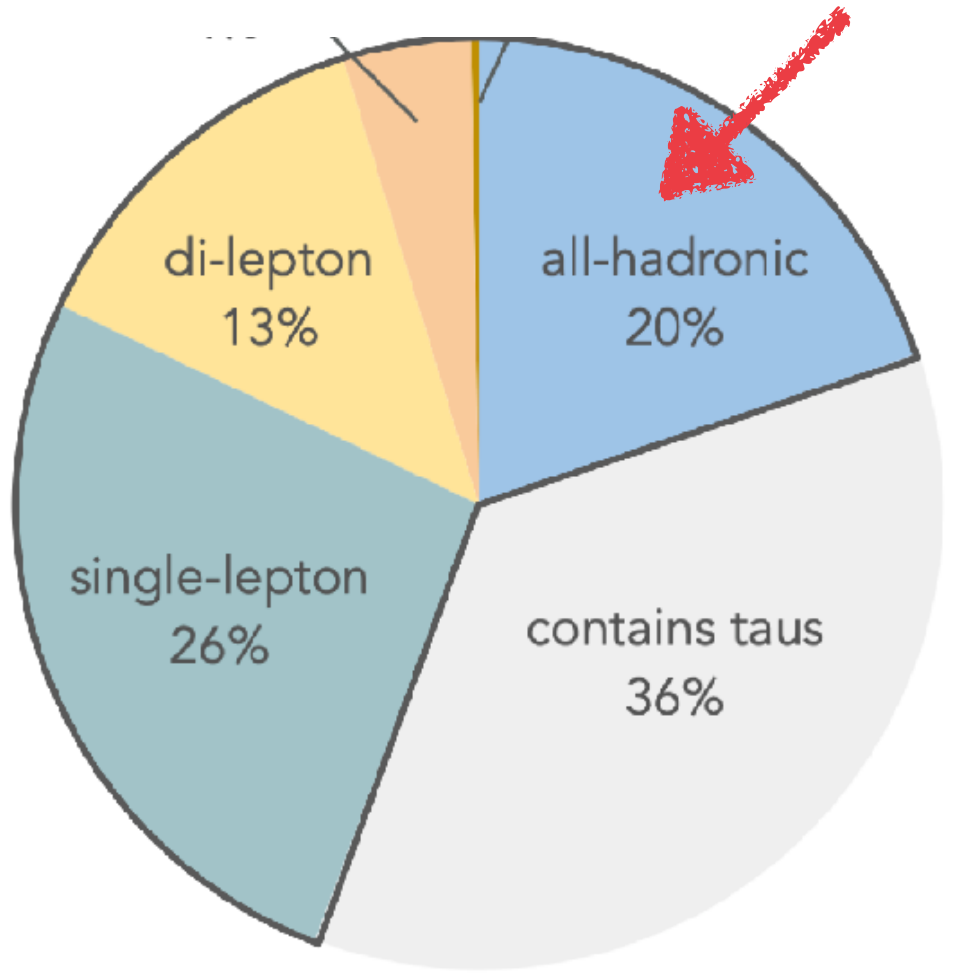


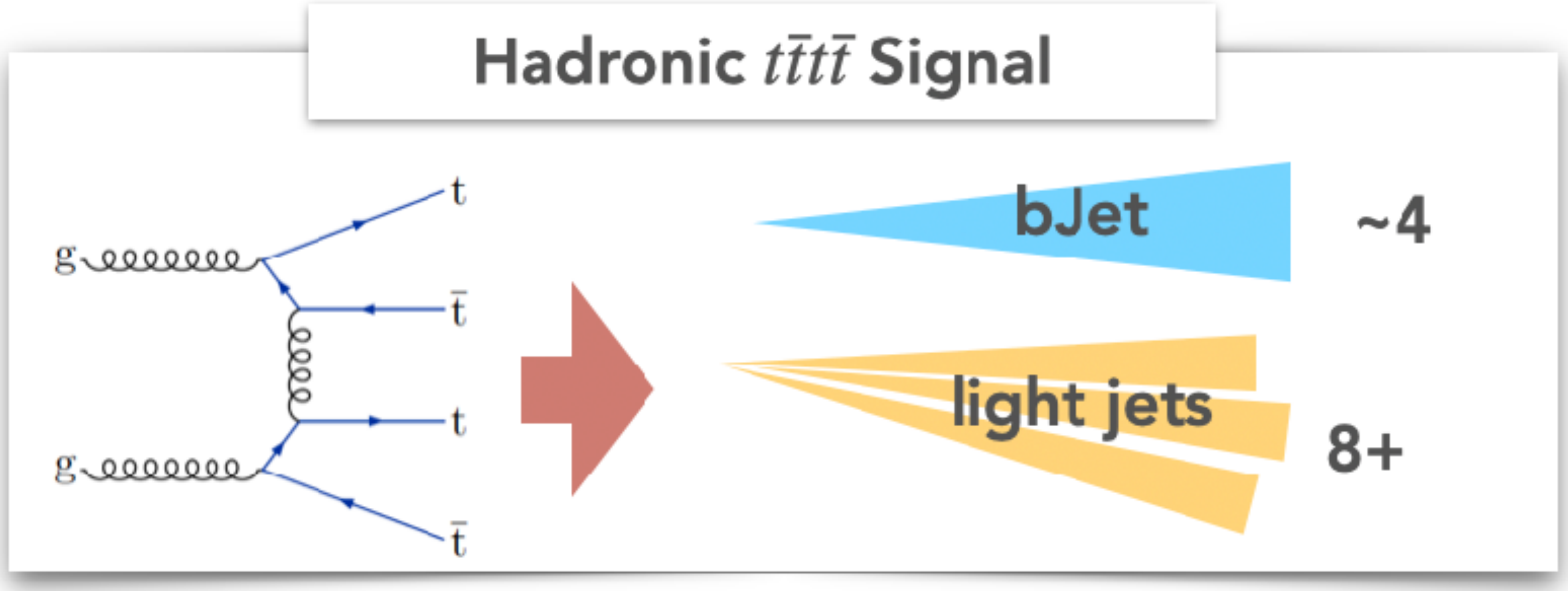
Next: CMS combined Run II analysis in multiple final states underway

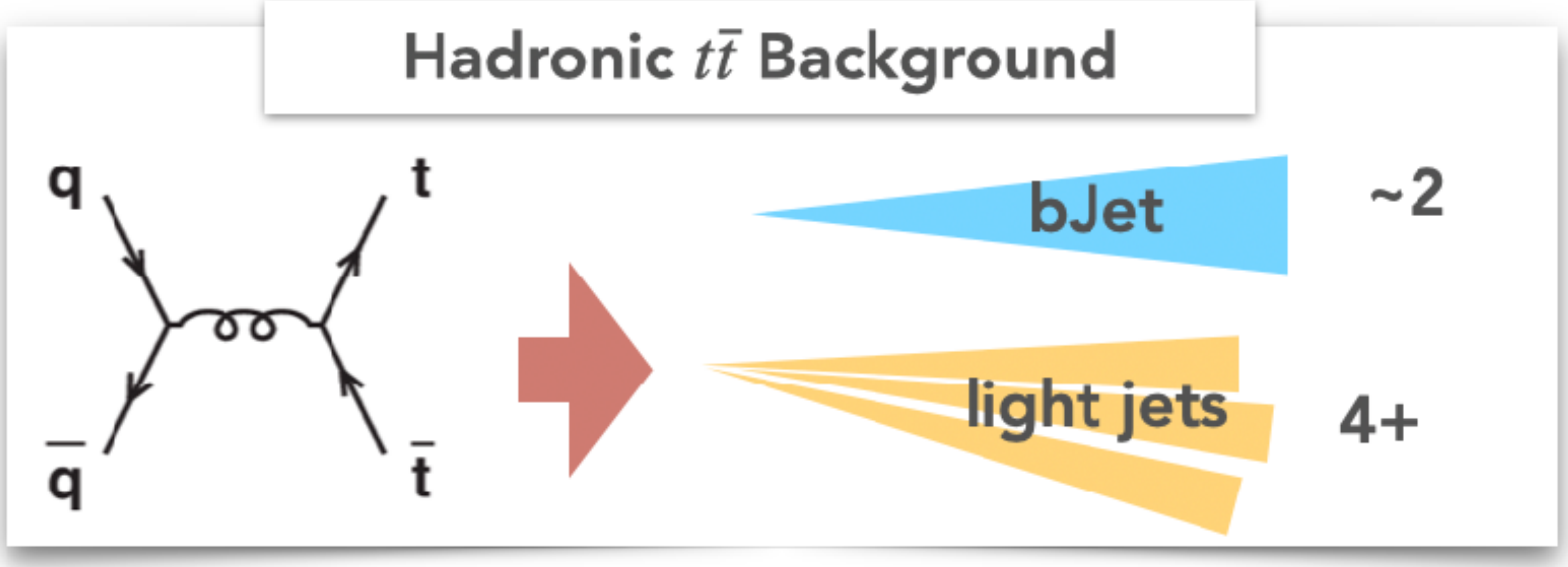
*ATLAS Collaboration. arXiv:2007.14858
+CMS Collaboration. CMS-TOP-18-003. arXiv:1908.06463

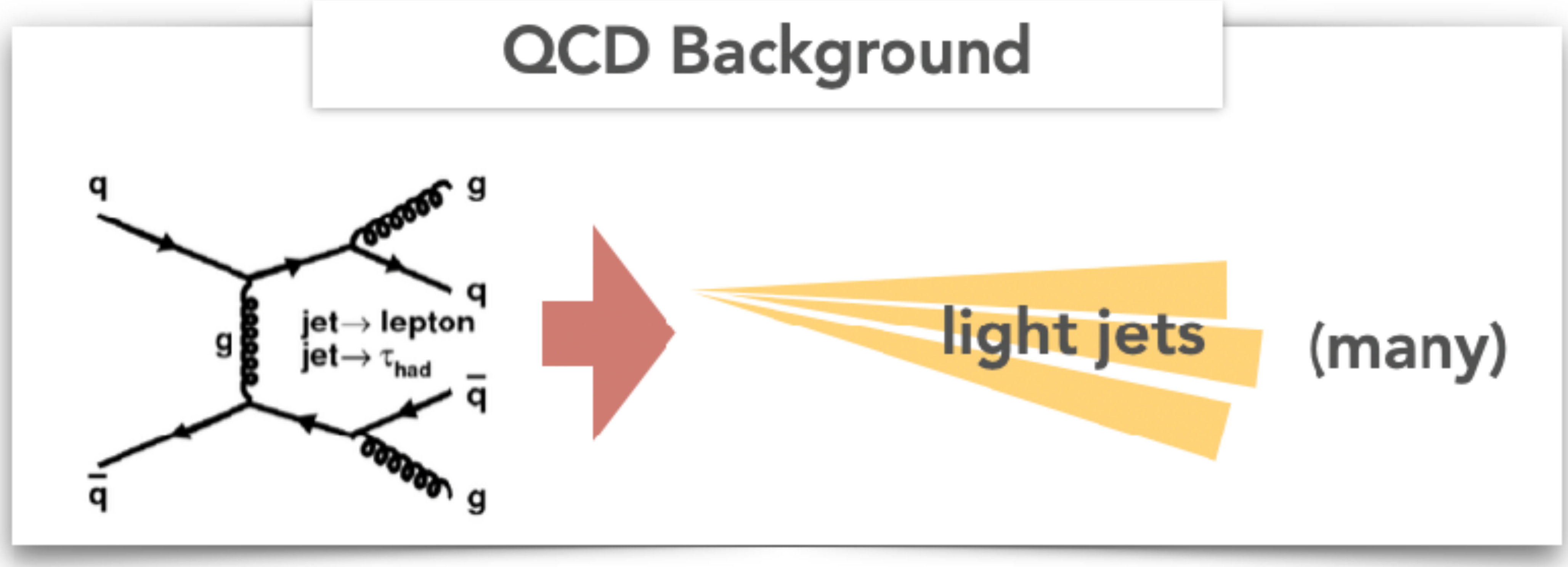
The All-Hadronic Channel

- First all-hadronic $t\bar{t}t\bar{t}$ search
- Challenges:
  - Very significant QCD+$t\bar{t}$ backgrounds with similar signatures compared to $t\bar{t}t\bar{t}$
  - QCD difficult to model in simulation
- Opportunity to showcase some novel machine learning (ML) tools for data-driven background estimation

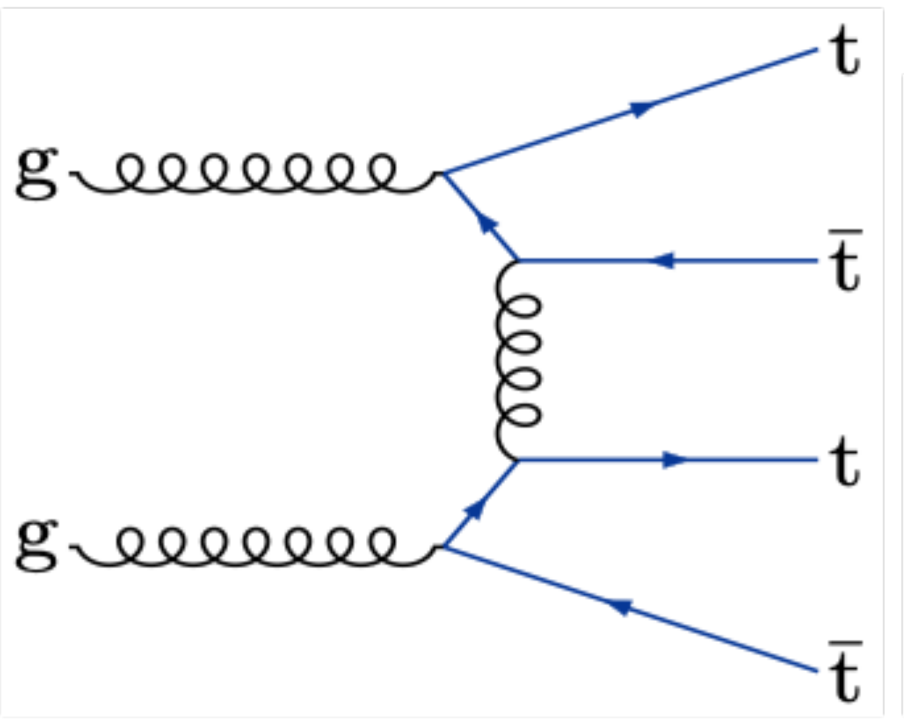




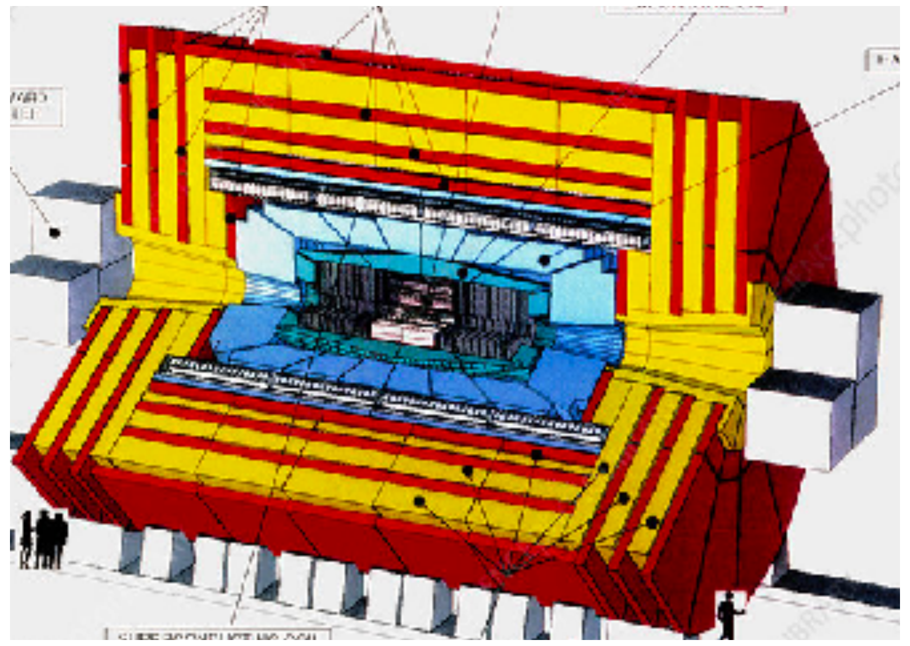




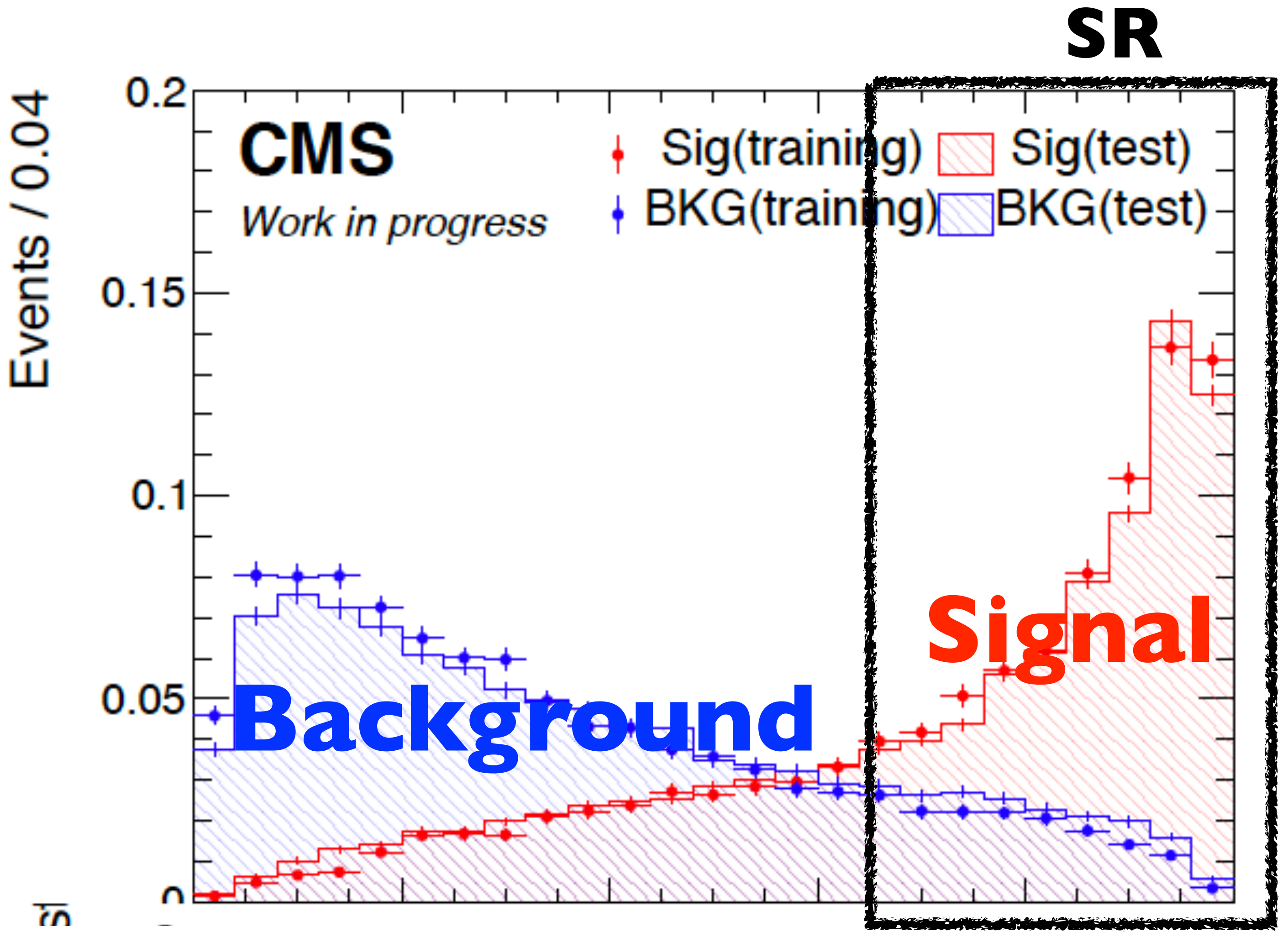
Analysis Roadmap



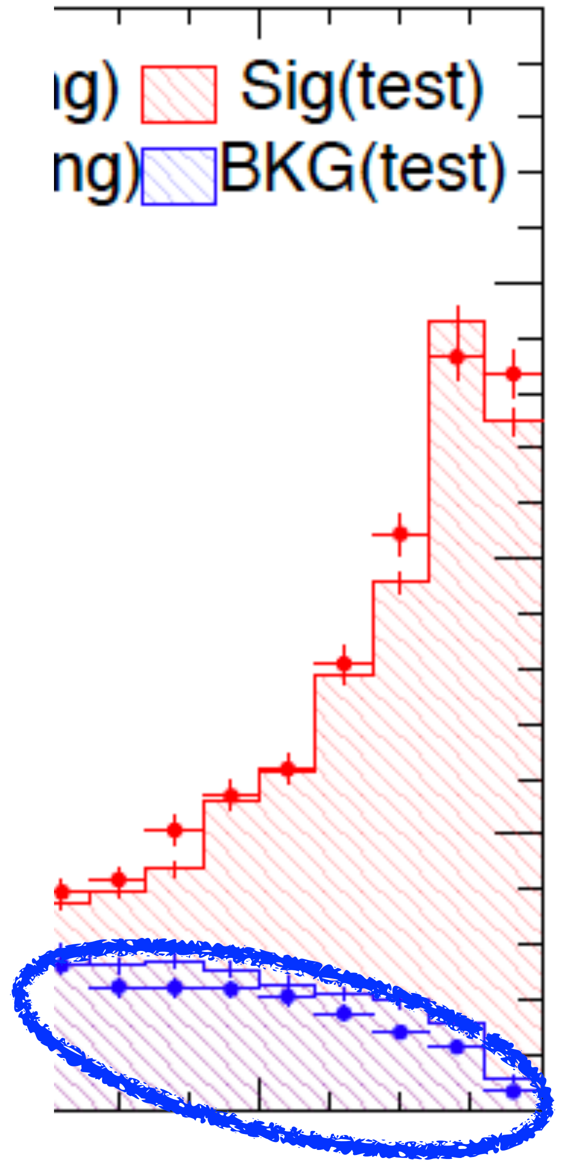
1) Select an interesting physics hypothesis to study (signal)



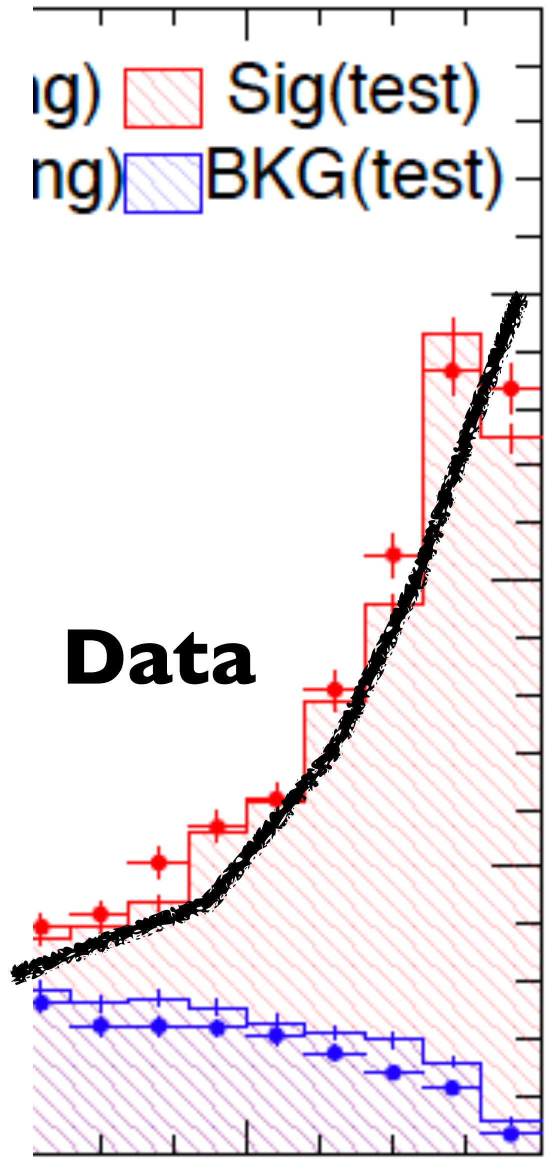
2) Collect data



3) Optimize a signal-enriched region



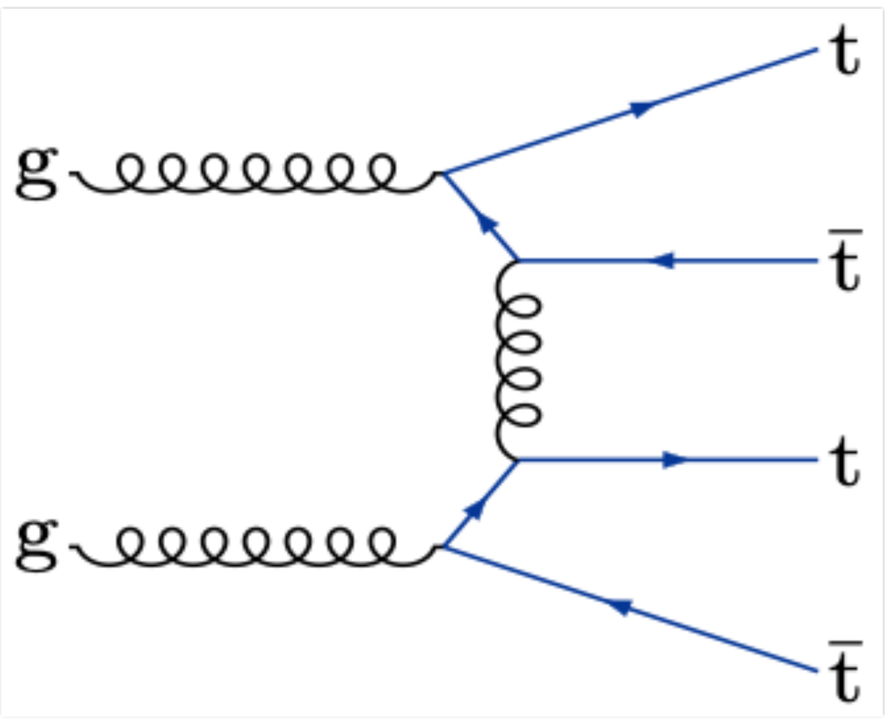
4) Estimate background



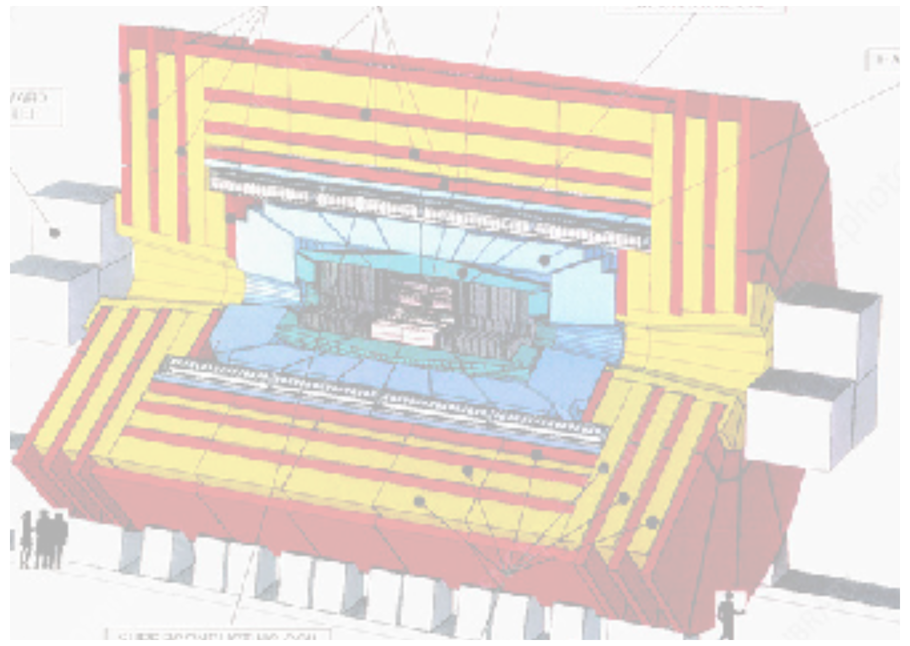
5) Compare expectations with data

Do results agree with SM predictions?

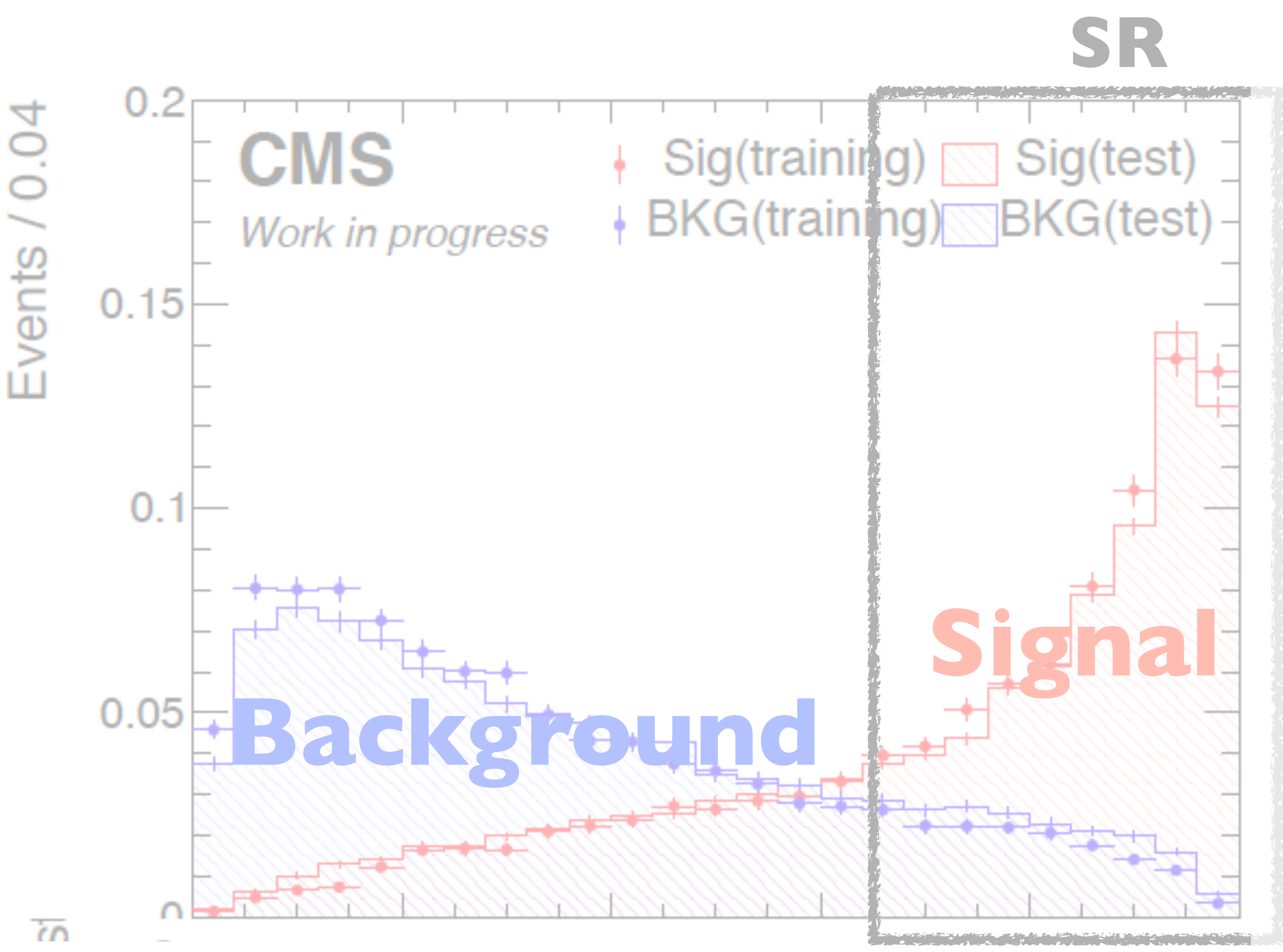
Analysis Roadmap



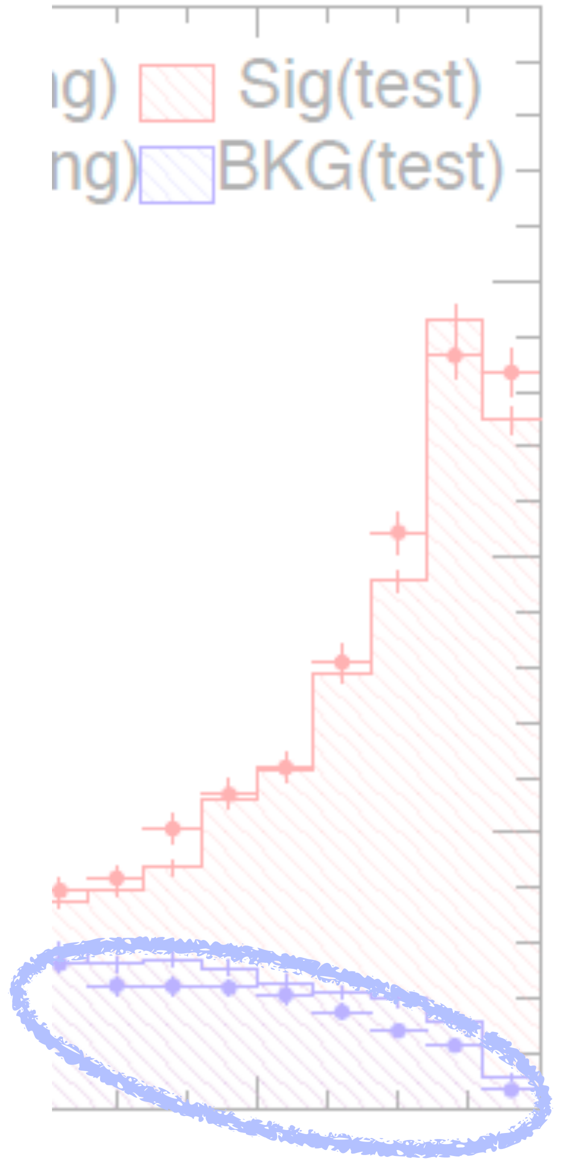
1) Select an interesting physics hypothesis to study (signal)



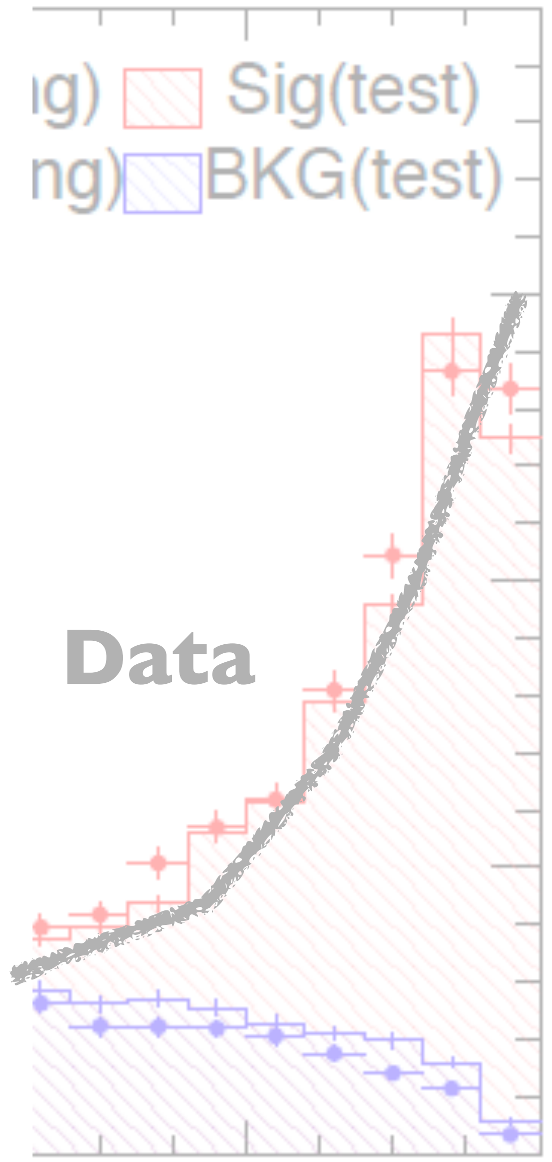
2) Collect data



3) Optimize a signal-enriched region



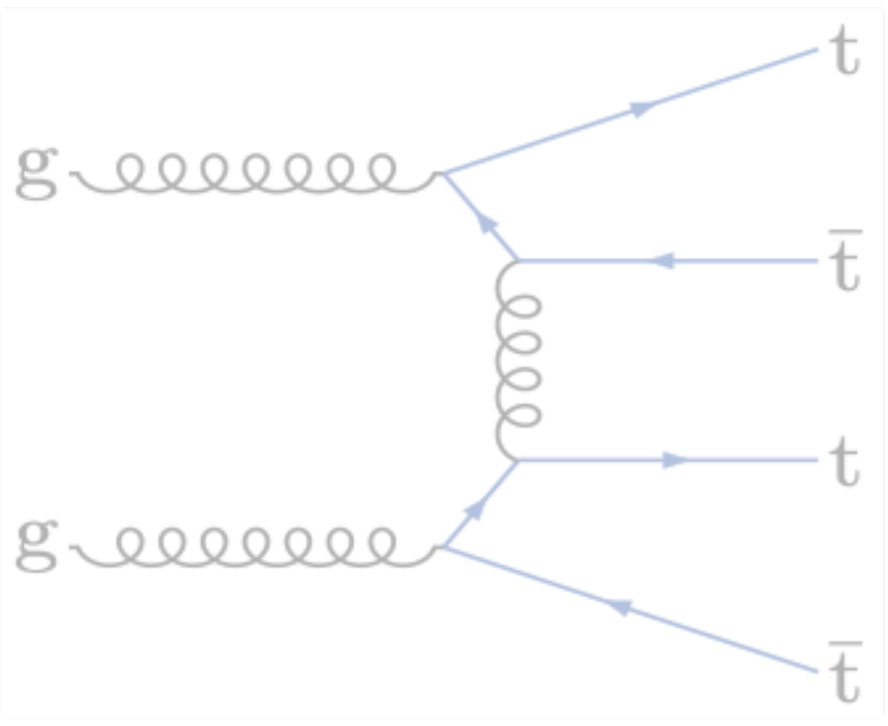
4) Estimate background



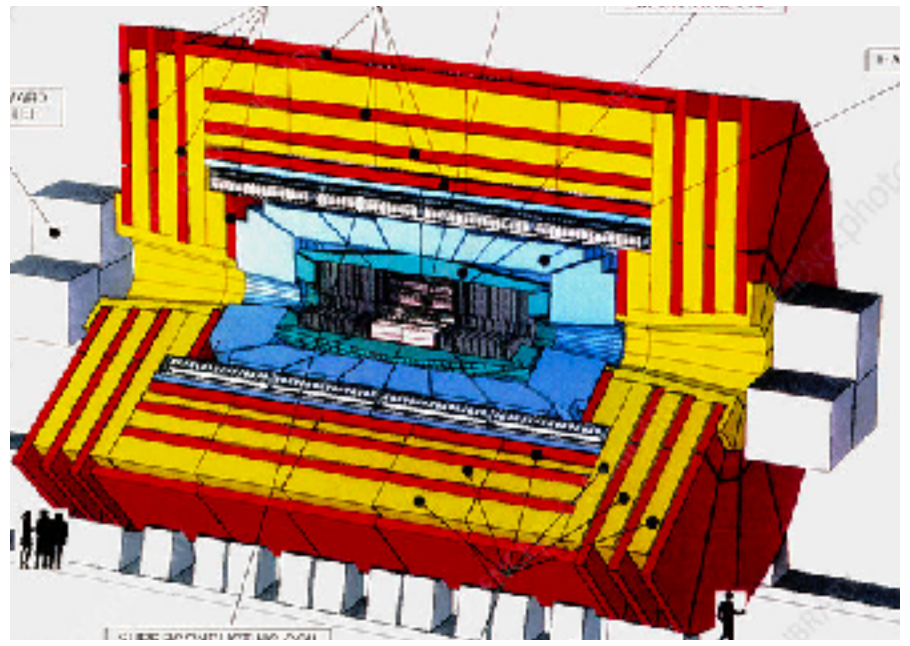
5) Compare expectations with data

Do results agree with SM predictions?

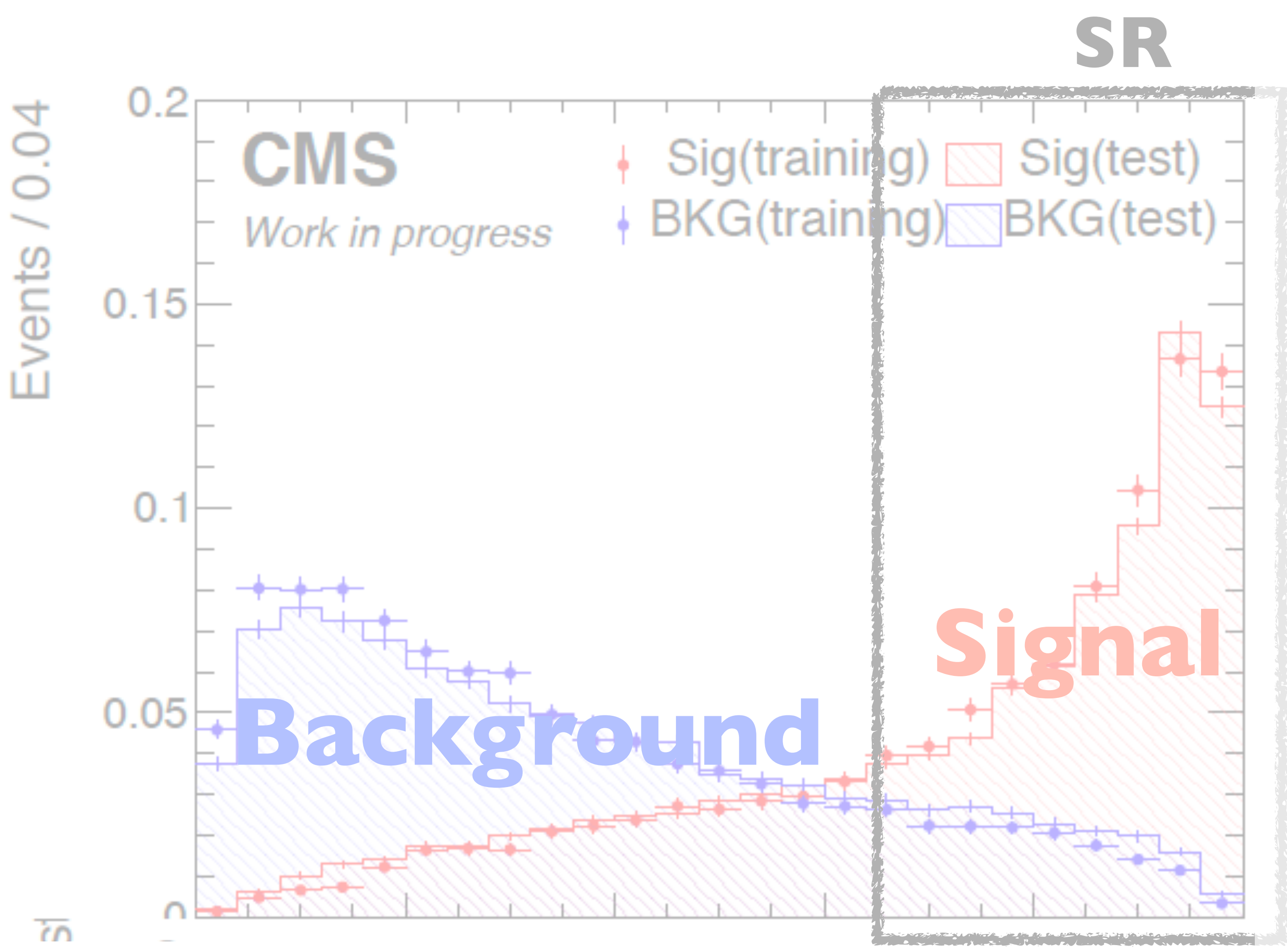
Analysis Roadmap



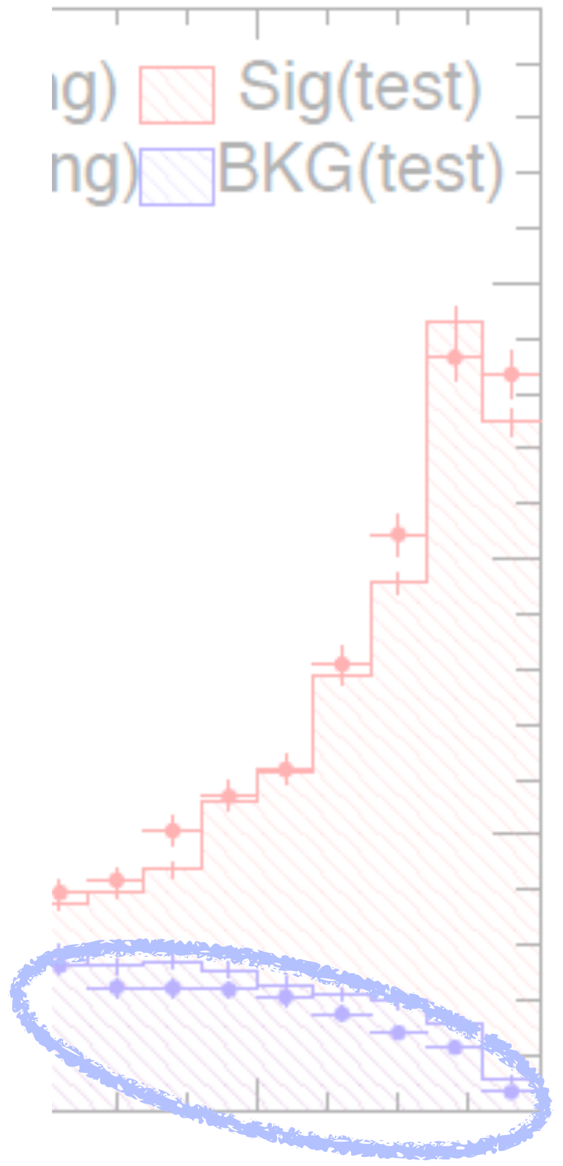
1) Select an interesting physics hypothesis to study (signal)



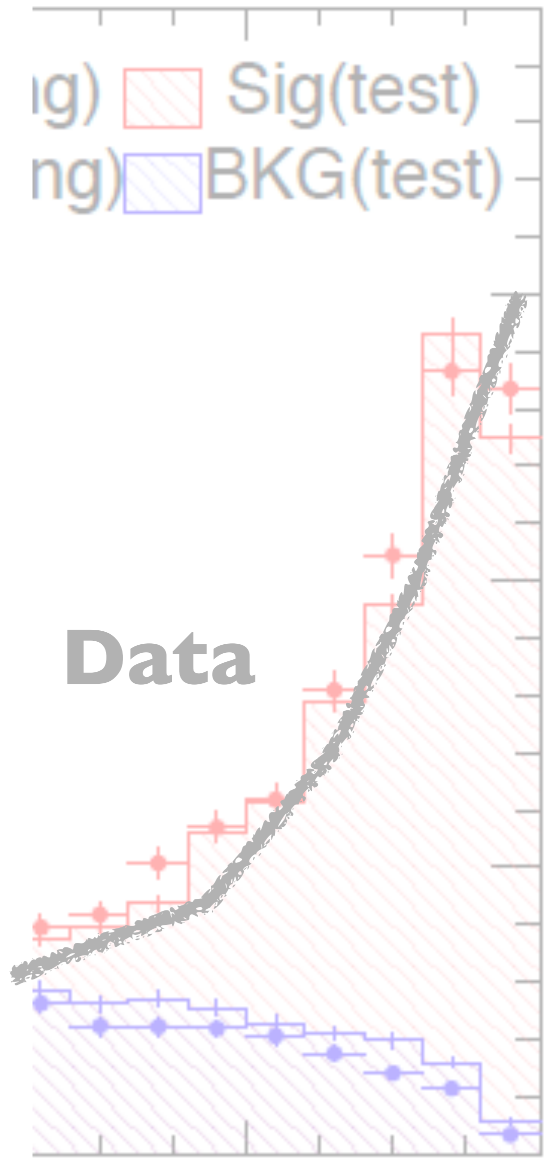
2) Collect data



3) Optimize a signal-enriched region



4) Estimate background

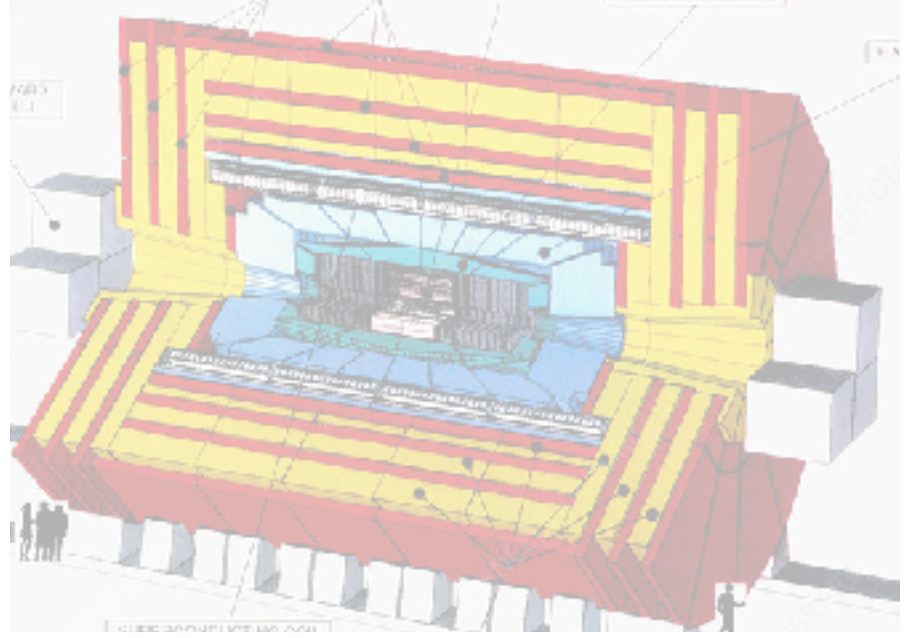


5) Compare expectations with data

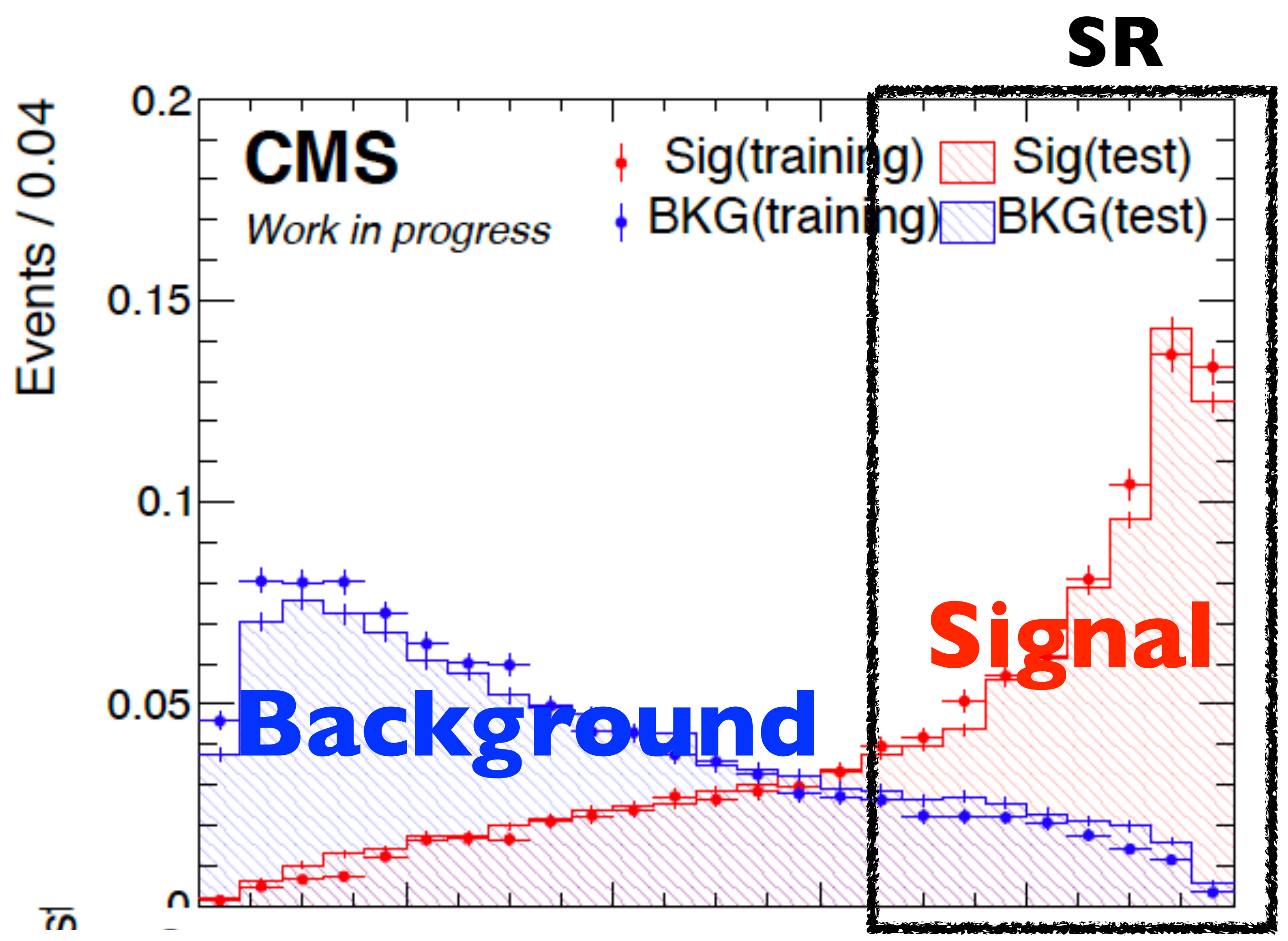
Do results agree with SM predictions?

Analysis Roadmap

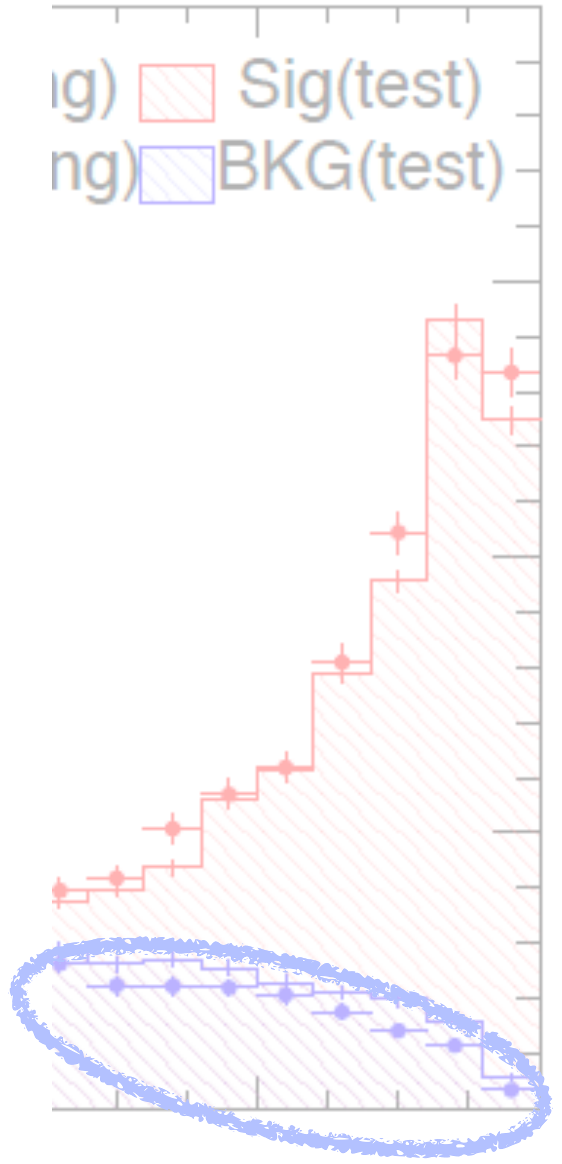
1) Select an interesting physics hypothesis to study (signal)



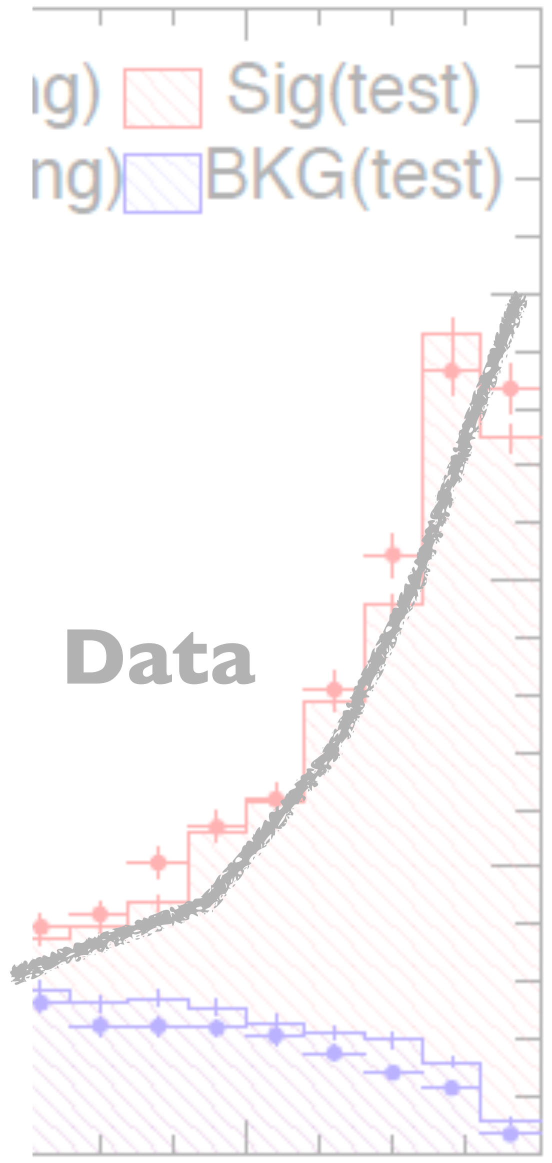
2) Collect data



3) Optimize a signal-enriched region



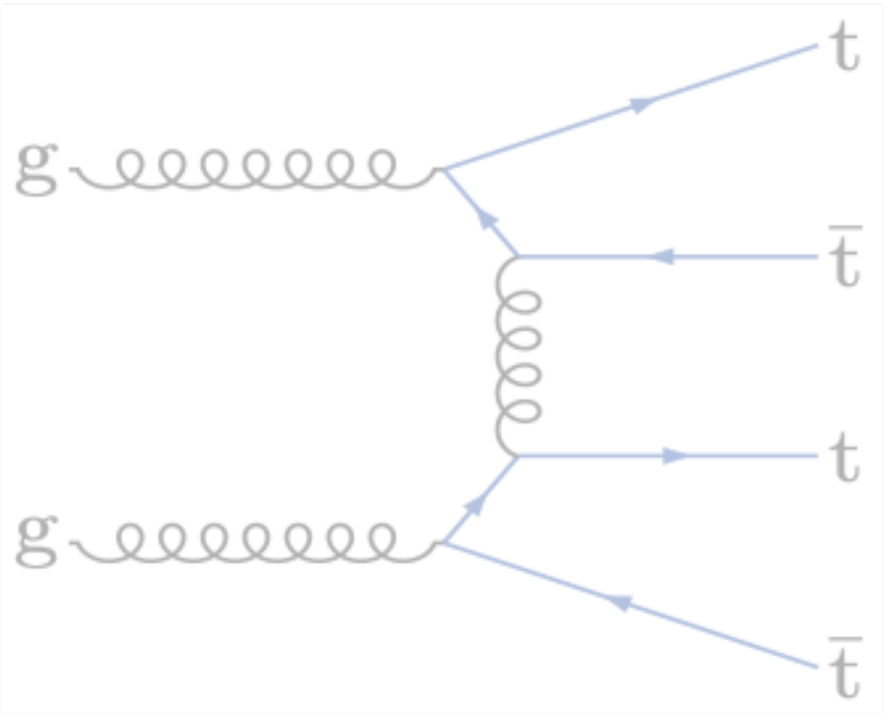
4) Estimate background



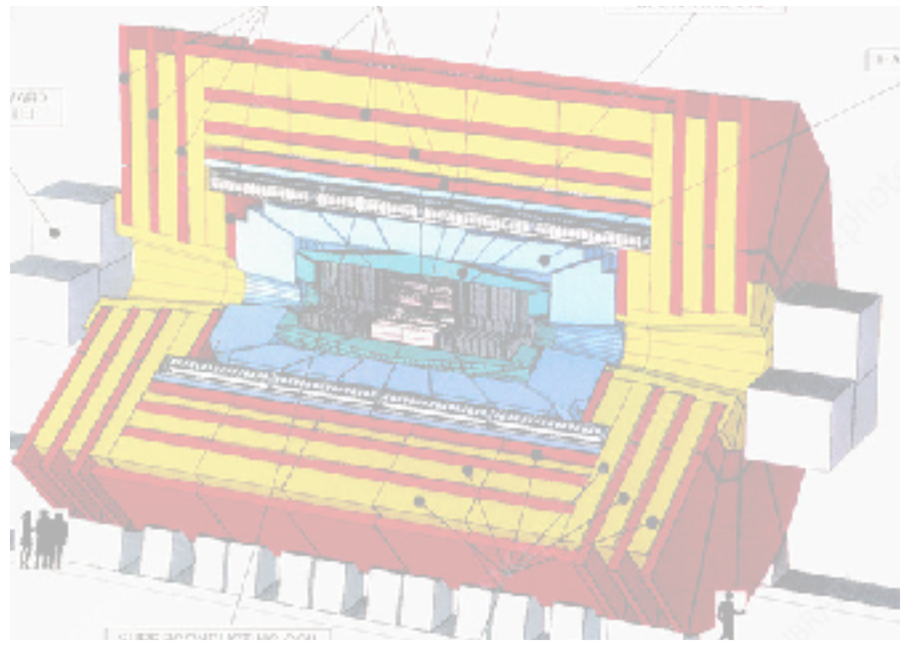
5) Compare expectations with data

Do results agree with SM predictions?

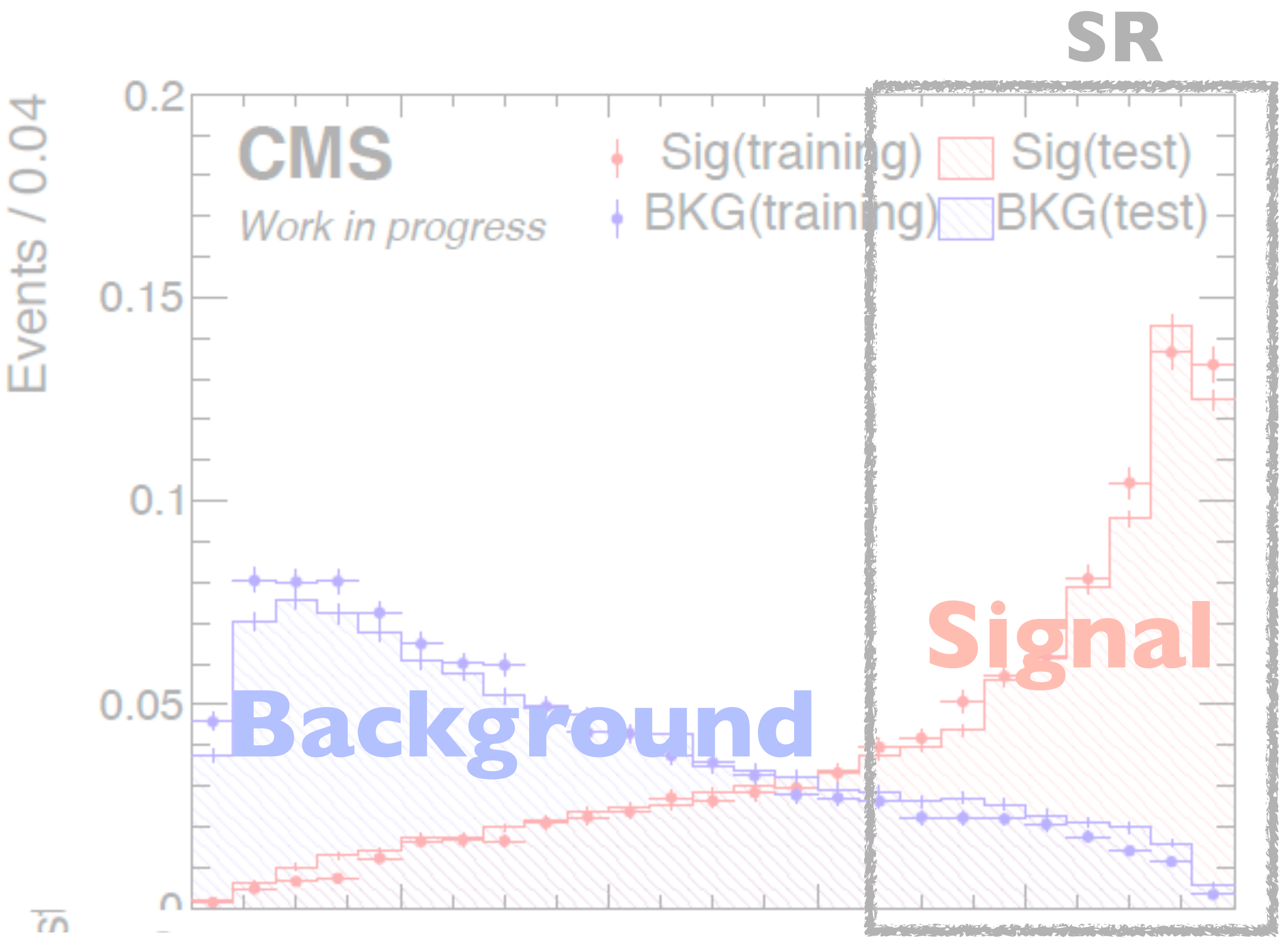
Analysis Roadmap



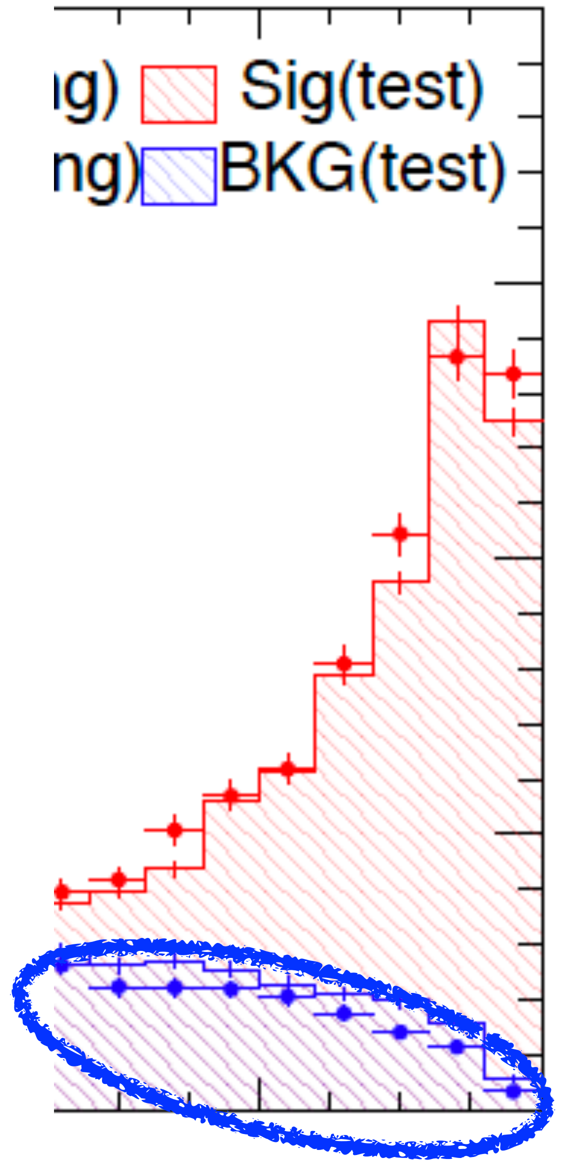
1) Select an interesting physics hypothesis to study (signal)



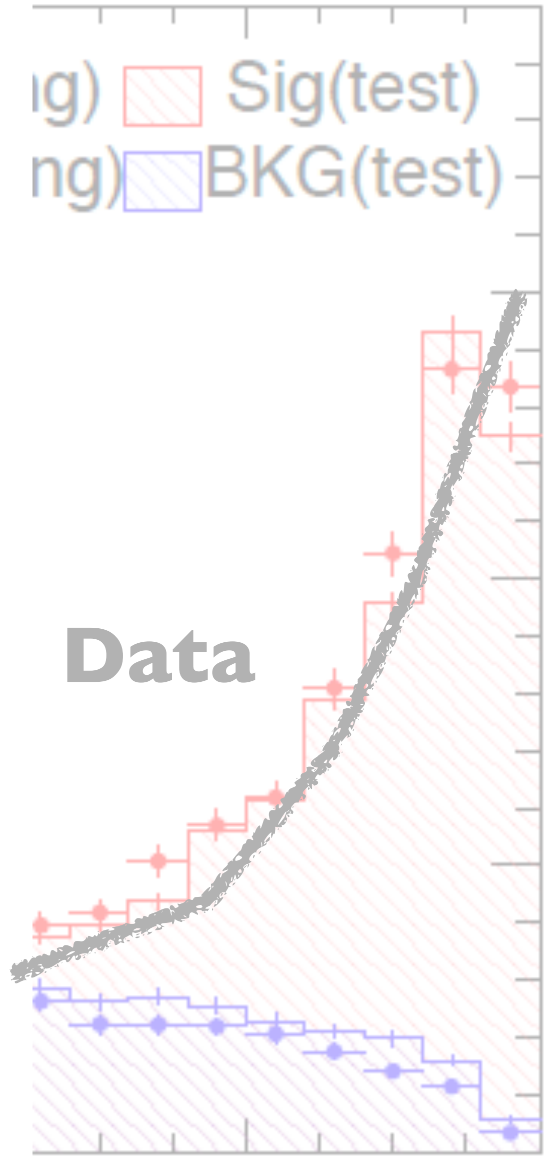
2) Collect data



3) Optimize a signal-enriched region



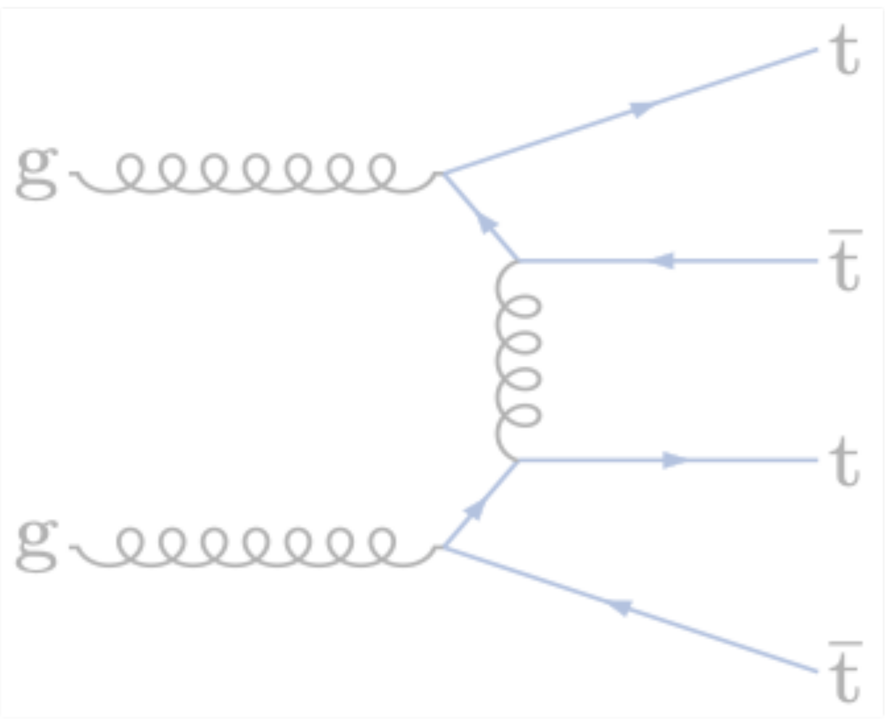
4) Estimate background



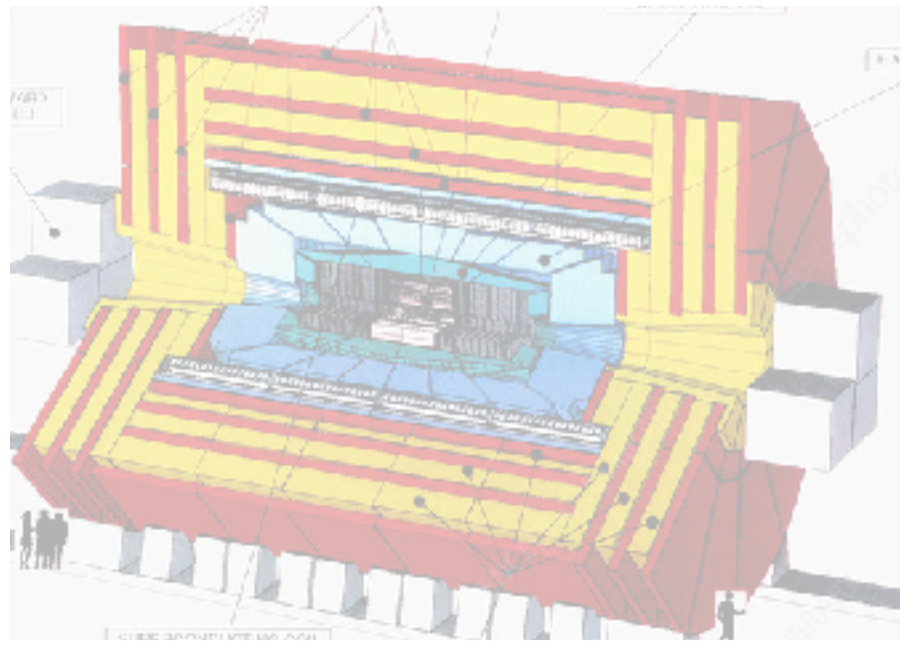
5) Compare expectations with data

Do results agree with SM predictions?

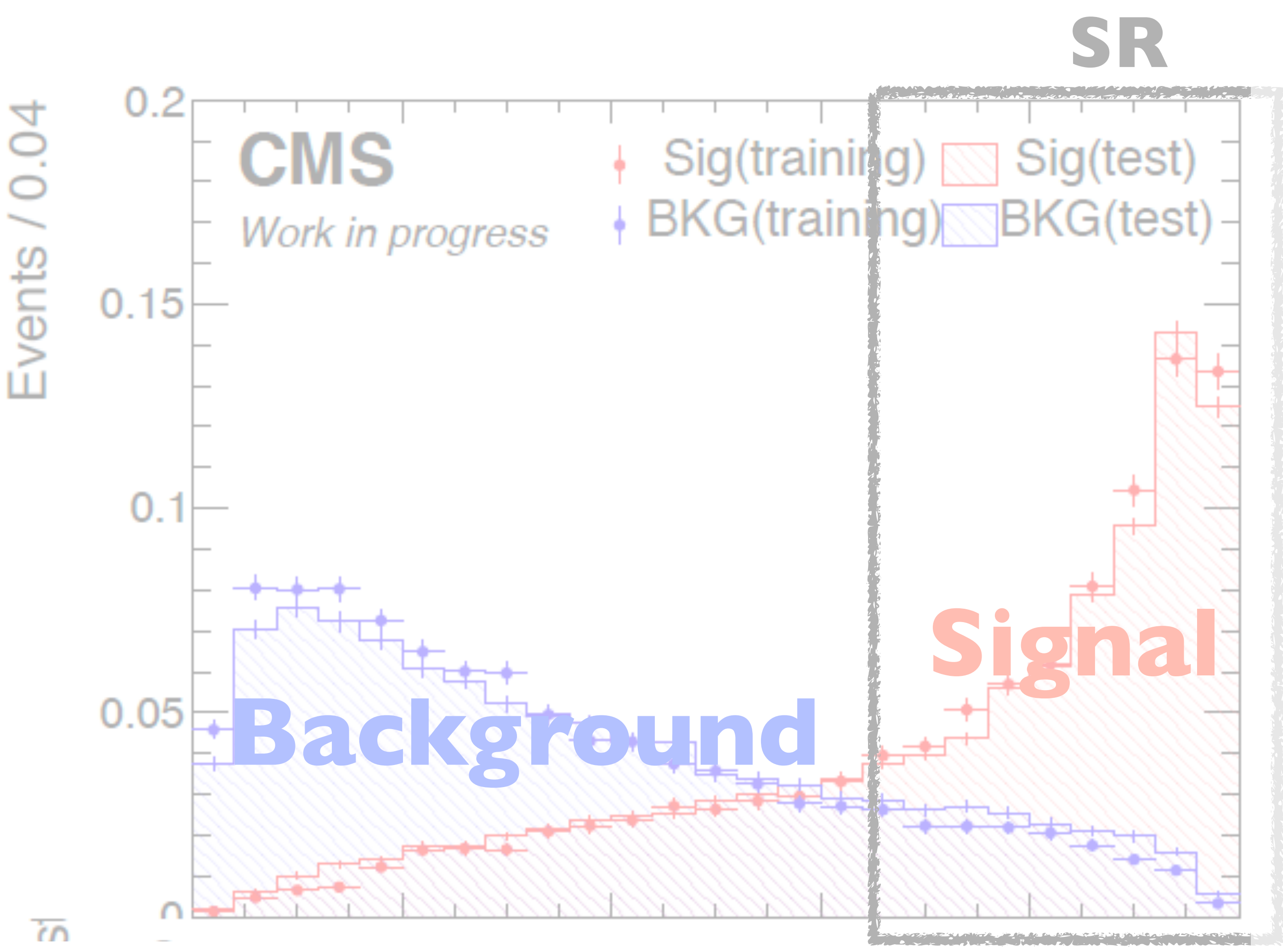
Analysis Roadmap



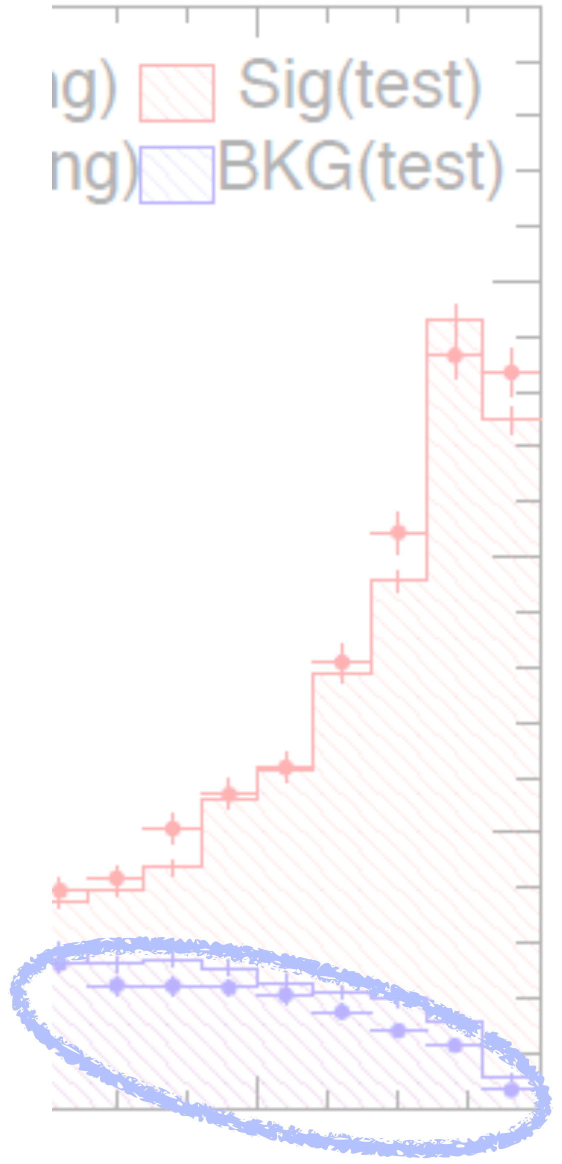
1) Select an interesting physics hypothesis to study (signal)



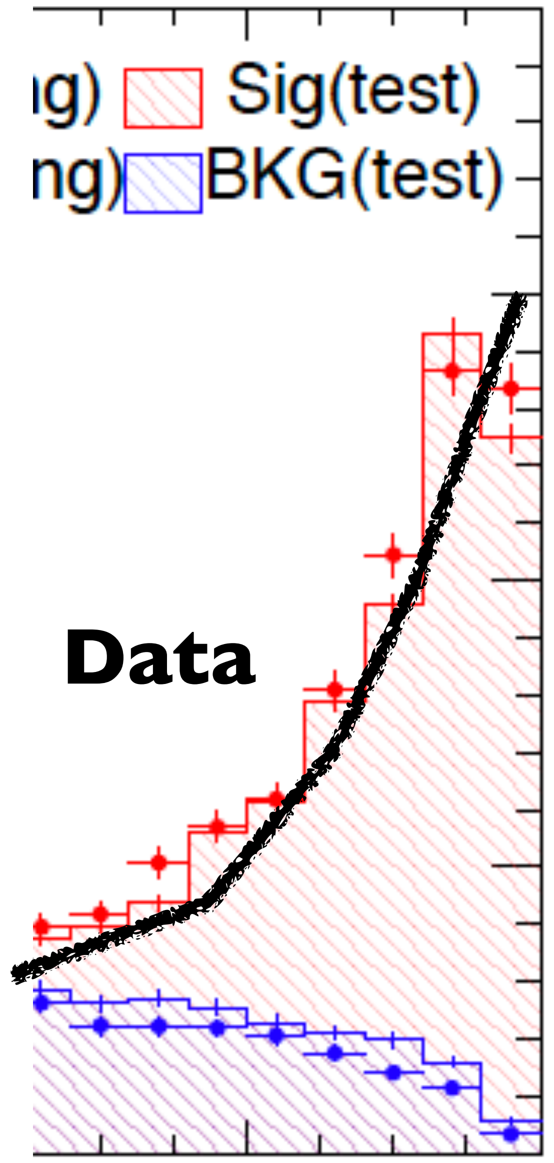
2) Collect data



3) Optimize a signal-enriched region



4) Estimate background

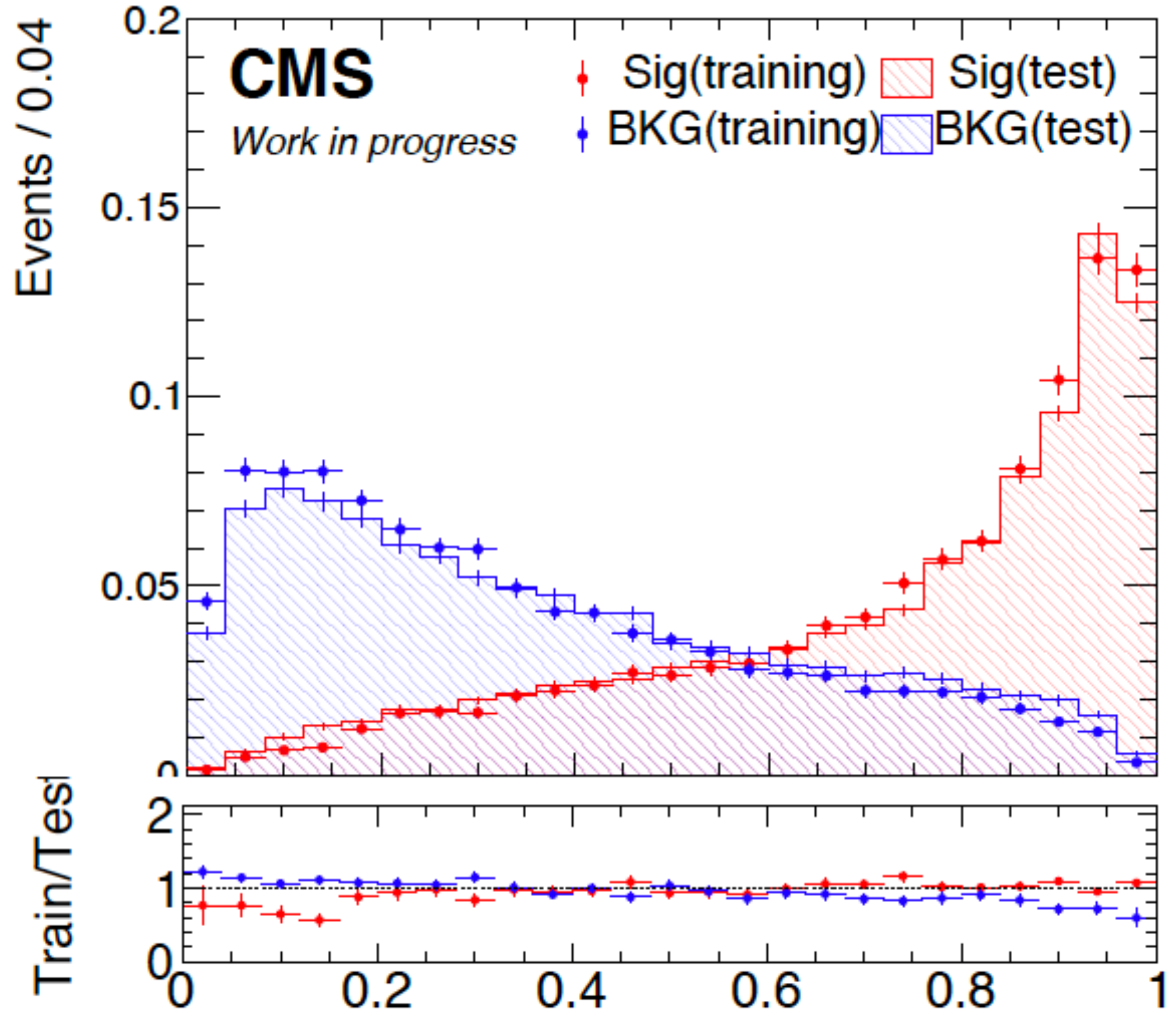


5) Compare expectations with data

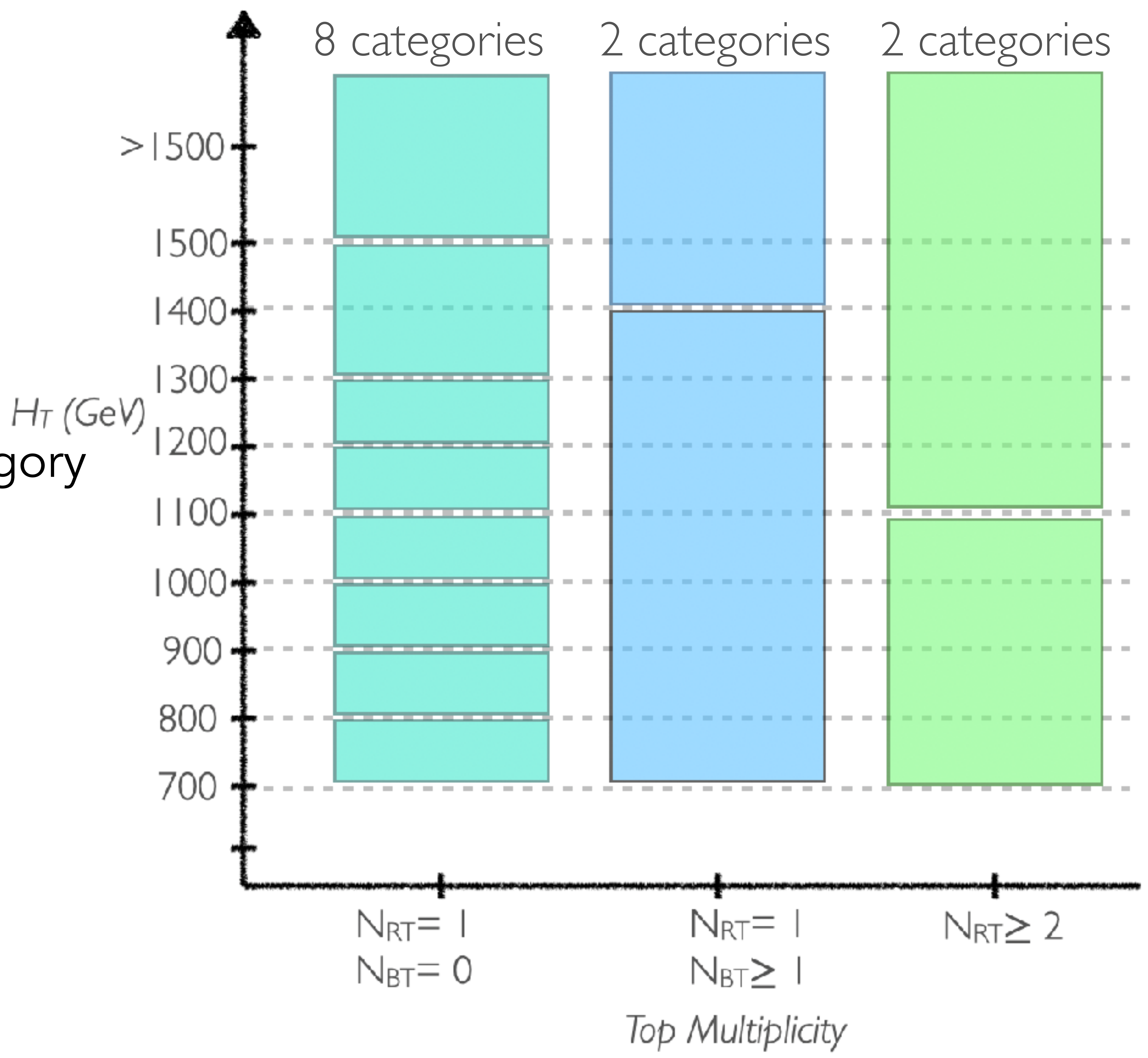
Do results agree with SM predictions?

Signal Region, Simplified

- Categorize signal region (SR) by sensitive variables
  - Number of tops (high p_T "boosted" tops (N_{BT}) and moderate p_T "resolved" tops (N_{RT}))
  - H_T, or scalar sum of transverse jet momentum
- Perform binned maximum likelihood fit to a discriminating variable distribution in each SR category
  - Use BDT discriminant distribution



Summary of SR Categories



Estimating Backgrounds

- Signal+ minor backgrounds - use MC simulation
- QCD+$t\bar{t}$ background - MC unreliable!
- Need method to predict the BDT shape of these backgrounds in SR categories!
- Solution: Use data in CRs to estimate backgrounds with help from ML techniques!

Define control regions (CRs) orthogonal to SR

↓

in CRs: QCD+$t\bar{t}$ = (data - other processes)

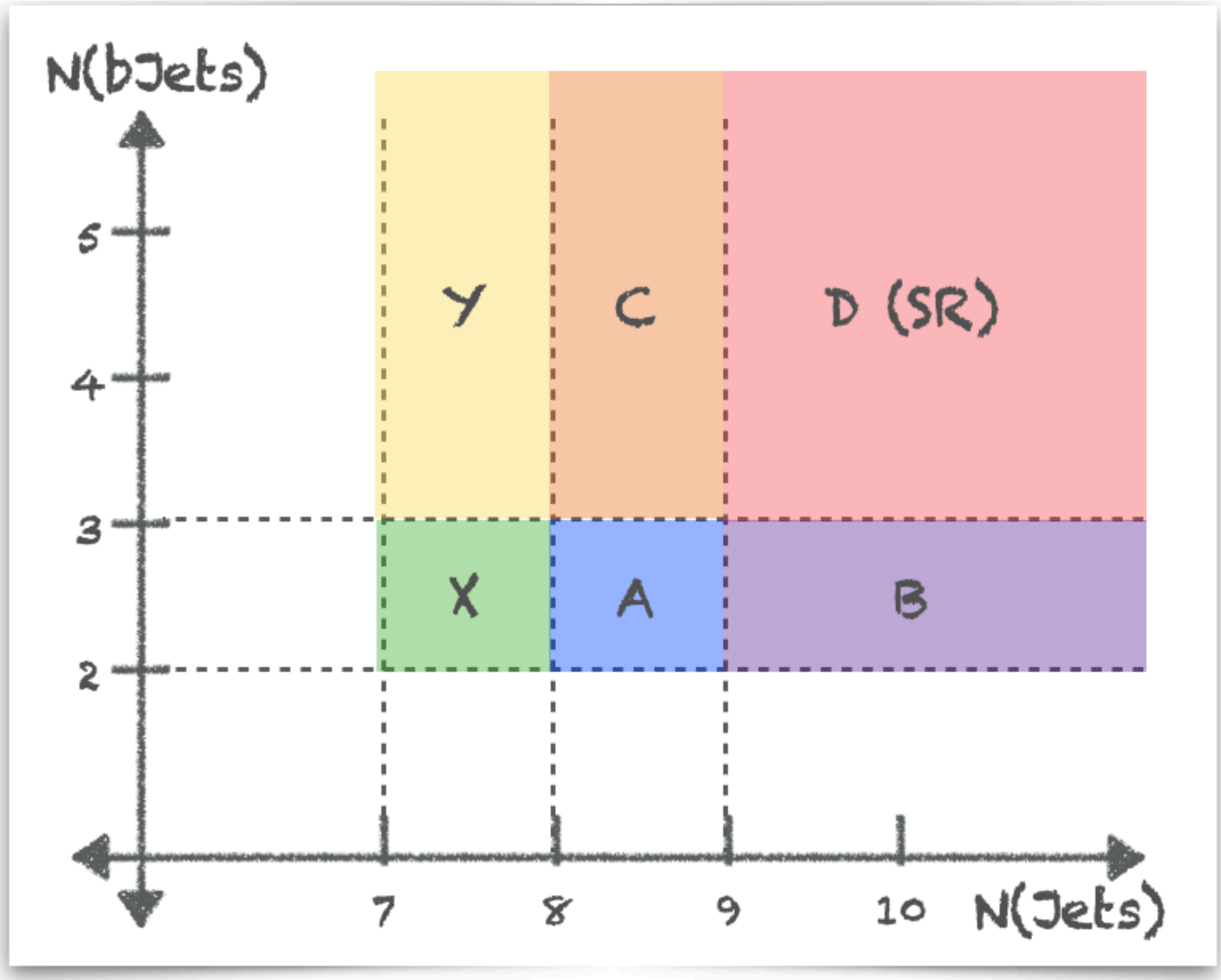
↙ ↘

Predict QCD+$t\bar{t}$ normalization (number of events) using extended ABCD formula

Predict QCD+$t\bar{t}$ BDT shape using a neural network (NN)

QCD+$t\bar{t}$ Normalization

- Goal: predict number of QCD+$t\bar{t}$ events in SR from data
- Split phase space by N(Jets) and N(bJets)
- Extrapolate from data in CRs to predict yields in SR (region "D")



$$D\,(SR) \cong \underbrace{\left(\frac{B\;C}{A}\right)}_{\text{ABCD term}} \underbrace{\left(\frac{C\;X}{A\;Y}\right)}_{\text{higher order term}}$$

QCD+$t\bar{t}$ yield =
data - ($t\bar{t}t\bar{t}$+$t\bar{t}$X+other processes)

*S. Choi. arXiv:1906.1083

QCD+$t\bar{t}$ Shape

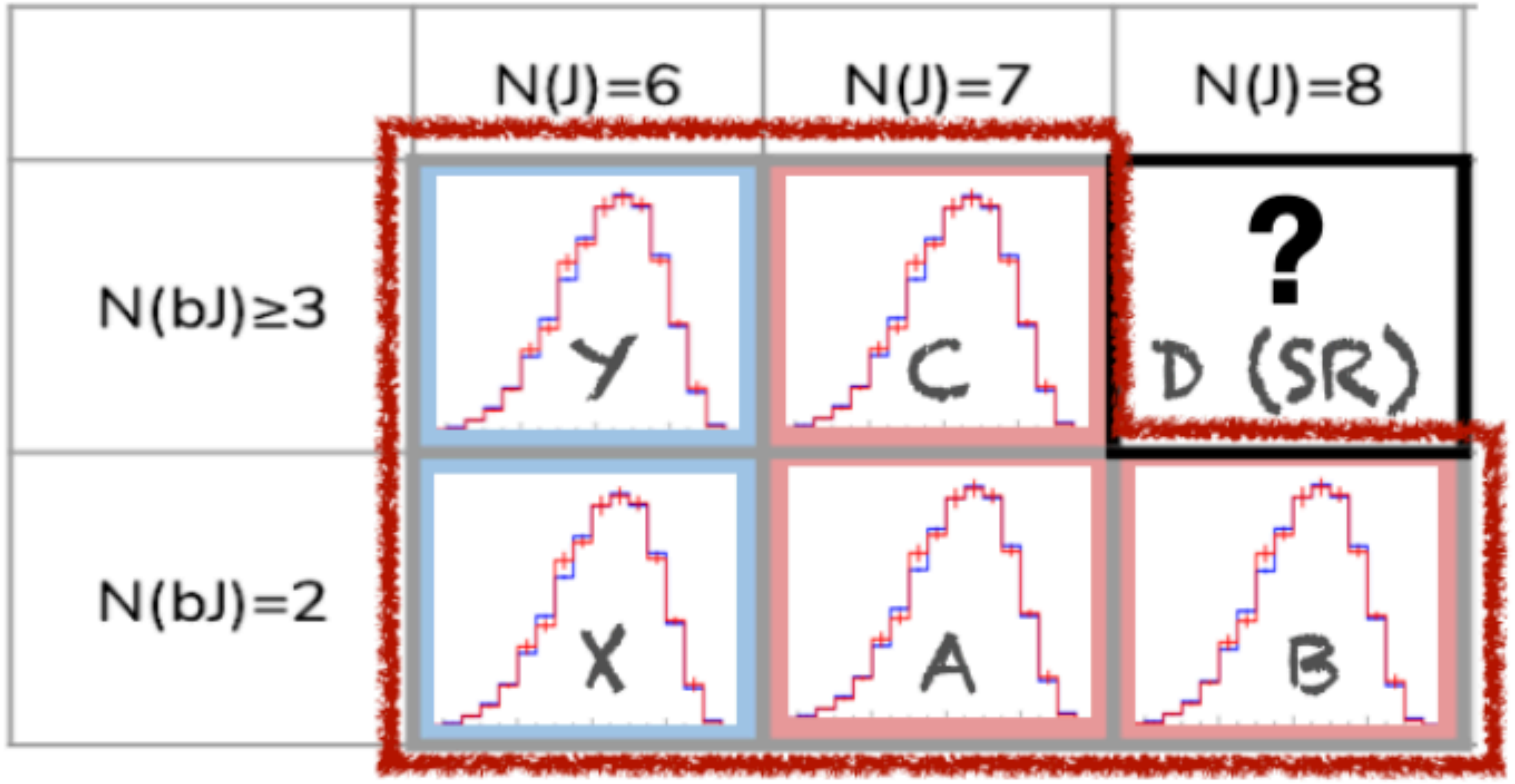
QCD+$t\bar{t}$ prediction = data - ($t\bar{t}t\bar{t}$+$t\bar{t}$X+other processes)

- Goal: predict shape of QCD+$t\bar{t}$ BDT in SR from data
- Neural net (NN) that finds transformation from input distribution → target distribution
  - "Neural autoregressive flow"*
  - Learns to transform distributions

input distribution: $t\bar{t}$ simulation

target distribution: QCD+$t\bar{t}$ prediction

- Trained on five CRs
  - Simultaneously predict BDT and H_T distributions
- After predicting distributions, they are normalized to predicted yields



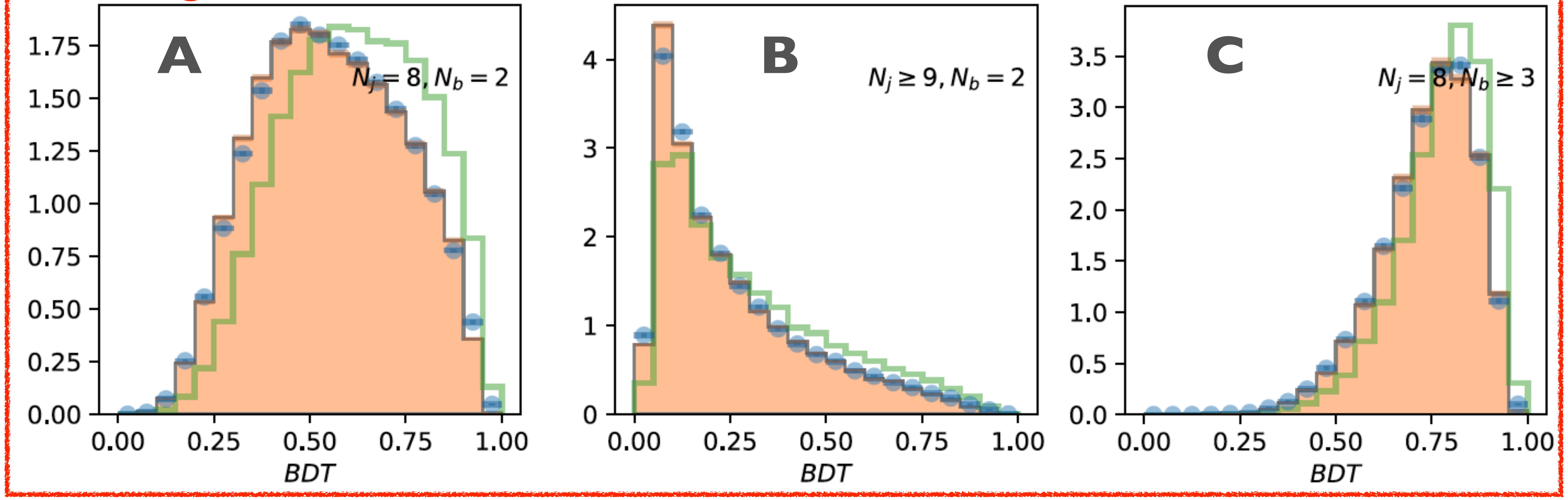
CR distributions used for training

*Huang, Krueger, Lacoste, Courville. *Neural Autoregressive Flows.* arXiv:1804.00779
*S. Choi. arXiv:2008.0363

QCD+$t\bar{t}$ Shape

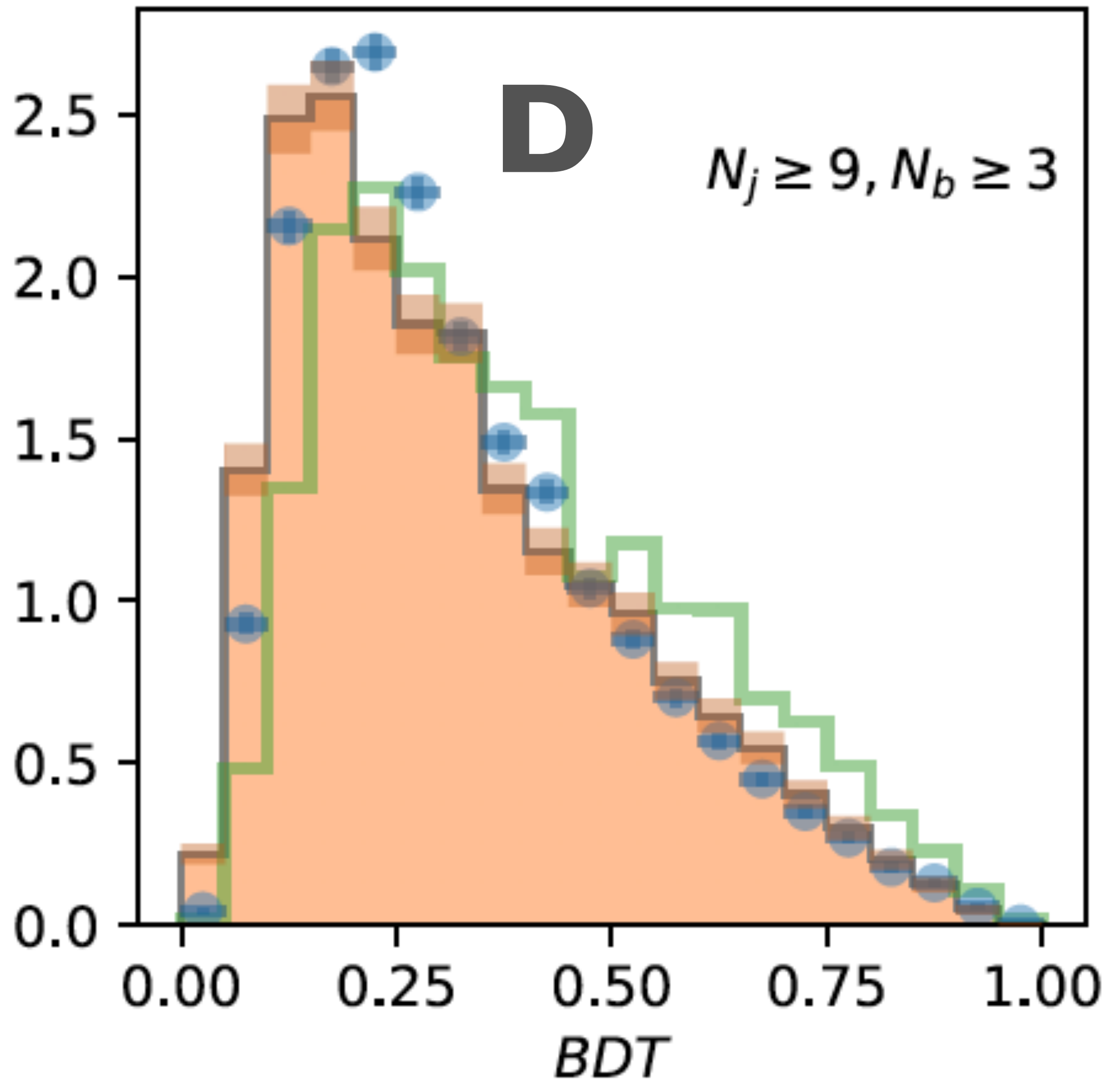
data
original $t\bar{t}$ distribution
predicted distribution

Training: start with $t\bar{t}$ and learn transformation to match data



- Example: morphing $t\bar{t}$ simulated samples to match data in a N(resolved tops)=0 region
- Normalized to $t\bar{t}$ yields

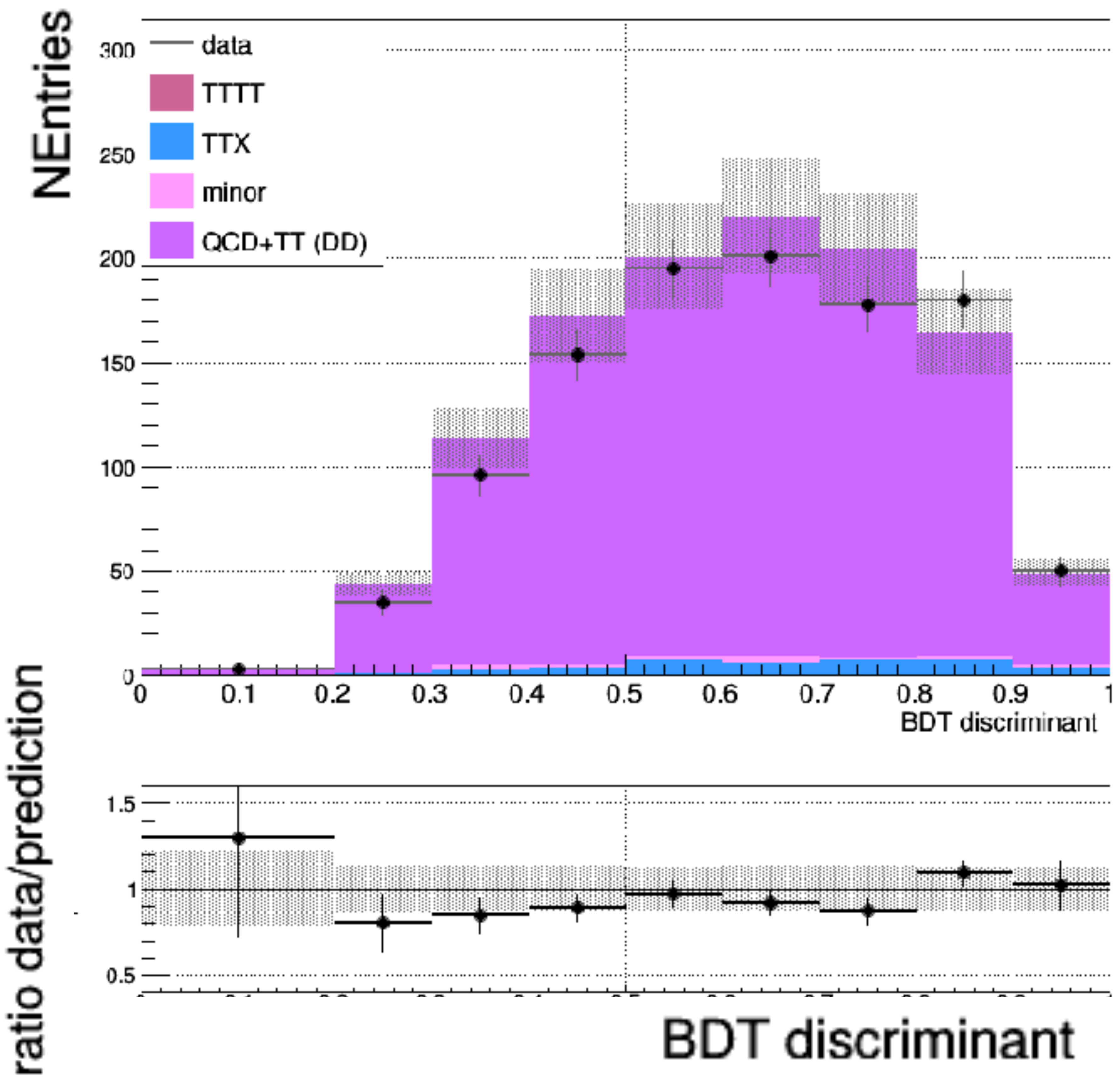
Result: Apply learned transformation to $t\bar{t}$ in target region to predict data



Plots by Prof. Suyong Choi

Validating the Background Estimation

- VR identical to SR except with 8 instead of 9+ jets
- Residual differences are assigned as a systematic uncertainty in corresponding SR categories



$N_{RT} \geq 2$, $700 \leq H_T < 1400$ GeV

SR and corresponding CRs

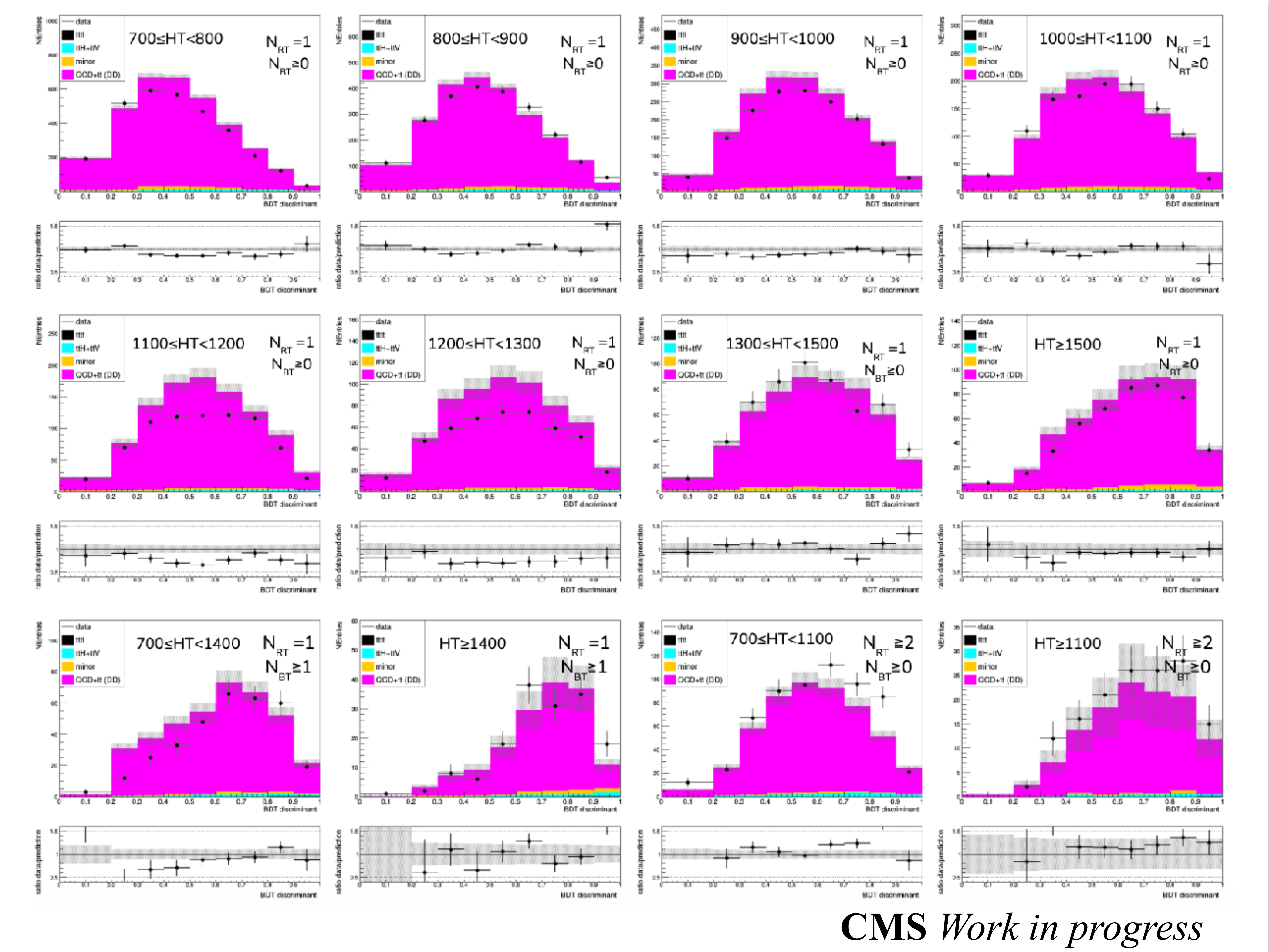
	N(J)=6	N(J)=7	N(J)=8	N(J)≥9
N(bJ)≥3	N(J)=6 N(bJ)≥3	Y N(J)=7 N(bJ)≥3	C N(J)=8 N(bJ)≥3	D N(J)≥9 N(bJ)≥3
N(bJ)=2	N(J)=6 N(bJ)=2	X N(J)=7 N(bJ)=2	A N(J)=8 N(bJ)=2	B N(J)≥9 N(bJ)=2

VR and corresponding CRs

	N(J)=6	N(J)=7	N(J)=8	N(J)≥9
N(bJ)≥3	Y N(J)=6 N(bJ)≥3	C N(J)=7 N(bJ)≥3	D N(J)=8 N(bJ)≥3	N(J)≥9 N(bJ)≥3
N(bJ)=2	X N(J)=6 N(bJ)=2	A N(J)=7 N(bJ)=2	B N(J)=8 N(bJ)=2	N(J)≥9 N(bJ)=2

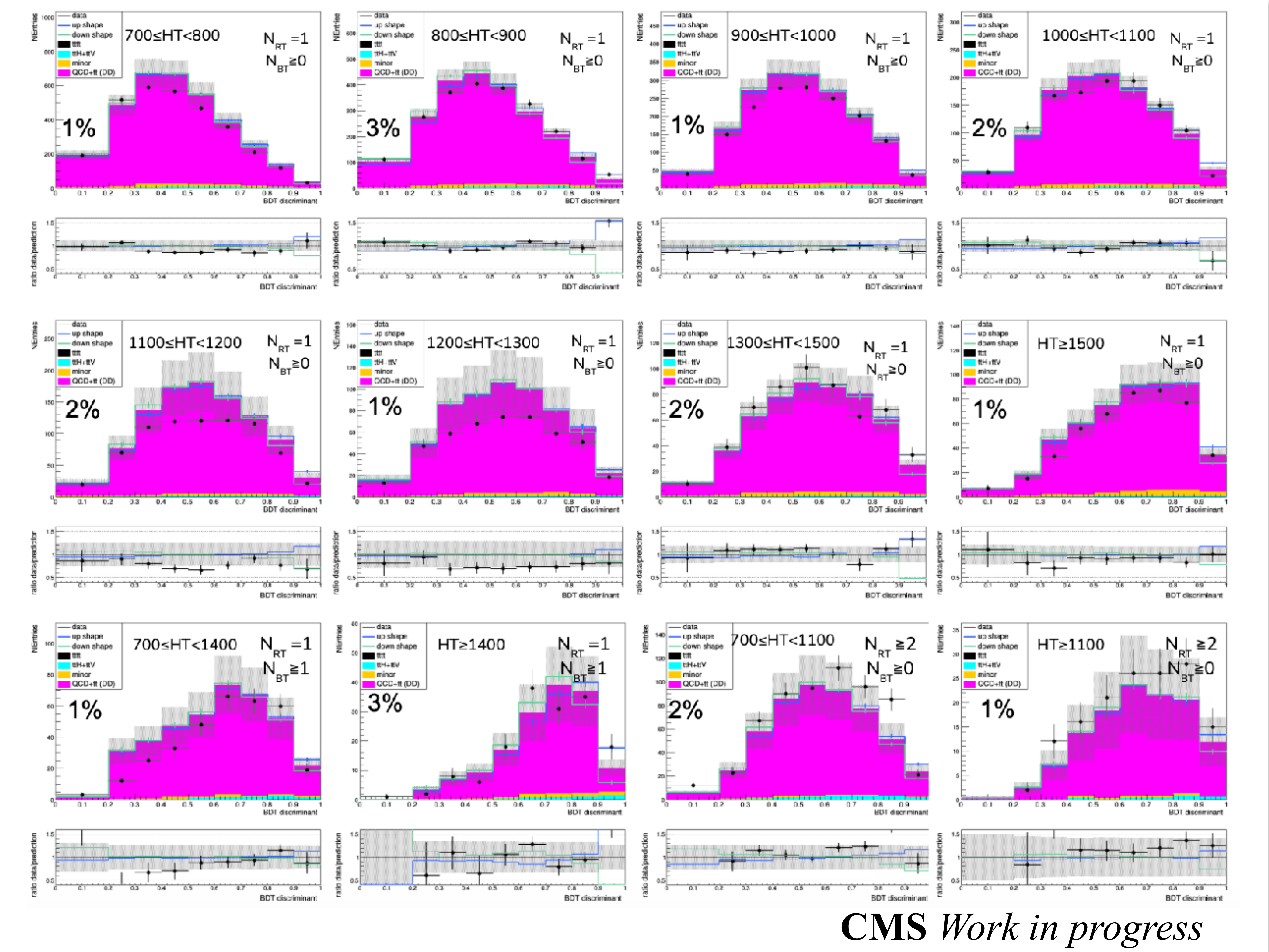
VR Distributions (2016)

- Statistical uncertainties only
- 12 VR categories



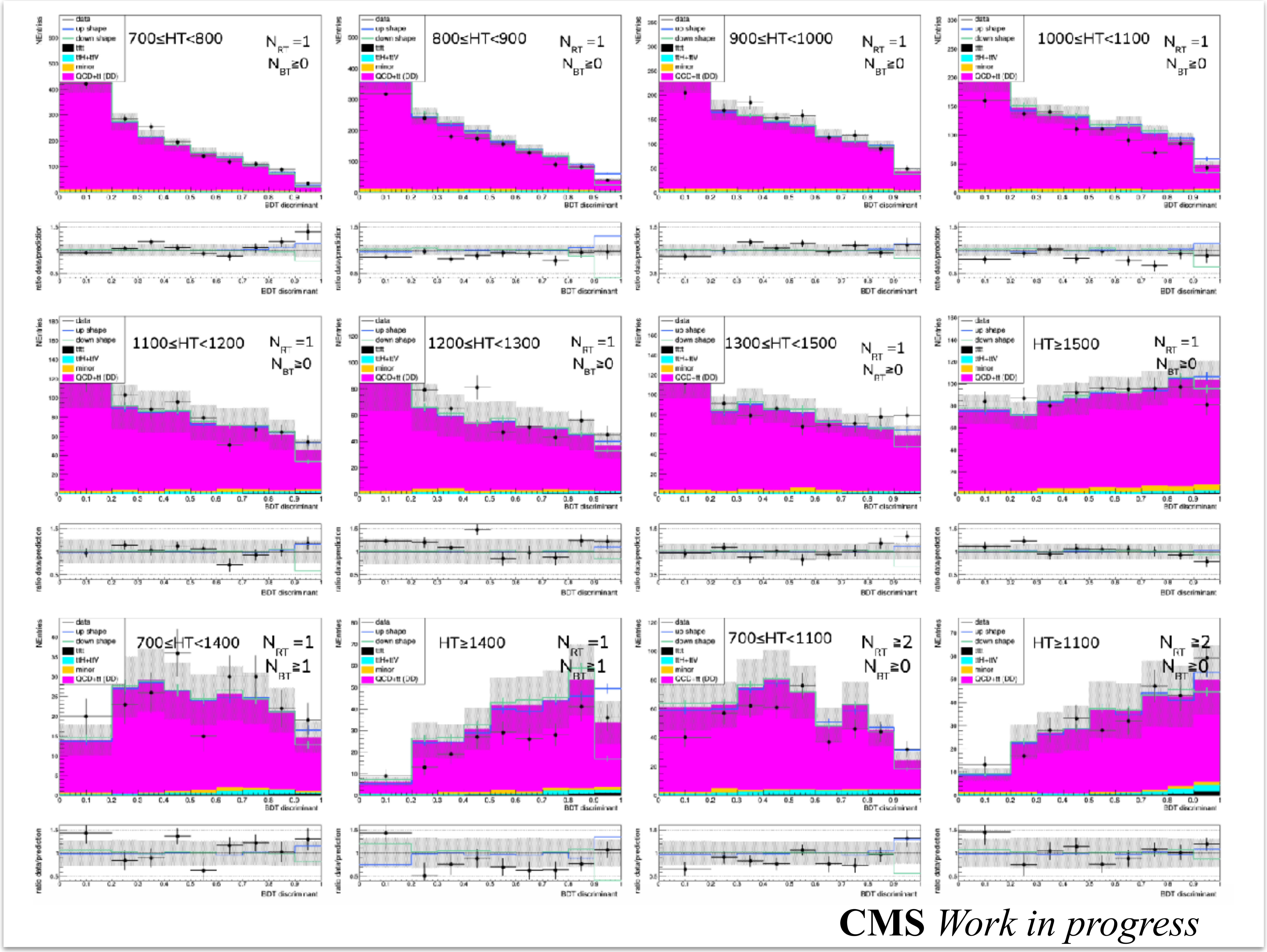
VR Distributions (2016)

- Statistical and VR residual disagreement shown
- 12 VR categories
- Residual VR uncertainties propagated to corresponding SR category



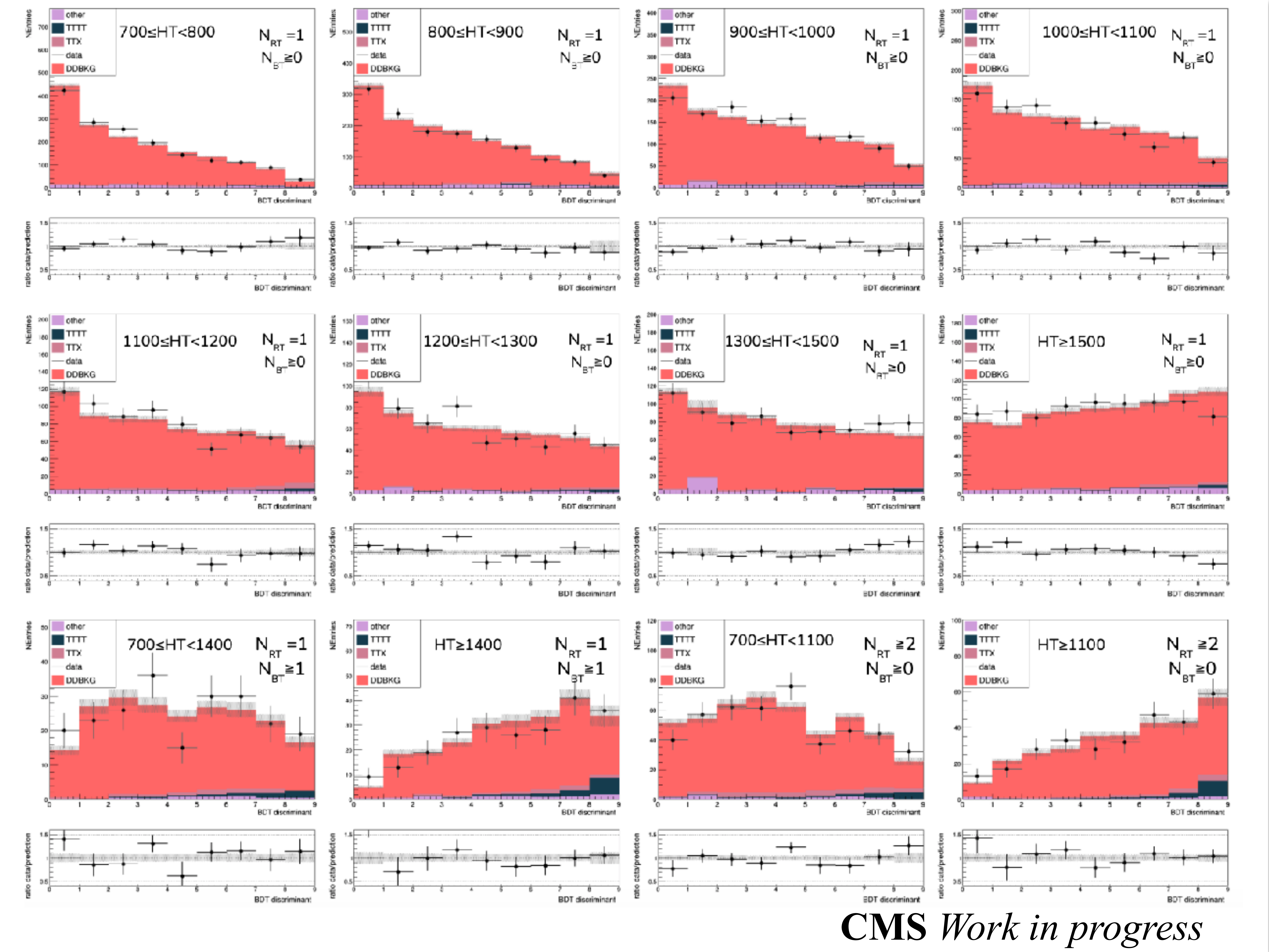
SR Distributions (2016)

- Statistical, VRnorm, and VRshape uncertainties are shown
- 12 SR categories



Postfit Distributions (Signal+Background) (2016)

- post-fit uncertainties are shown
- 12 SR categories



Results

CMS *Work in progress*

No fit to data

	2016 (36 fb^{-1})	2017 (42 fb^{-1})	2018 (60 fb^{-1})	Run II (137 fb^{-1})
Median expected limit: Expected significance (s.d.)	8.91 0.25	8.97 0.23	8.31 0.26	4.88 0.43

With fit to data

	2016 (36 fb^{-1})	2017 (42 fb^{-1})	2018 (60 fb^{-1})	Run II (137 fb^{-1})
Median expected limit: +1σ expected limit +2σ expected limit Expected significance (s.d.)	8.72 13.2 19.7 0.25	7.56 11.1 15.8 0.28	7.06 10.4 14.8 0.29	4.20 6.05 8.44 0.49
Observed limit: Observed significance (s.d.)	12.9 1.10	14.0 1.76	10.2 1.15	8.39 2.25

"Limits" are on $\mu = \sigma/\sigma_{SM}$

- Expected and observed significances and limits x SM cross section in all-hadronic channel alone.
- Per year and for full Run II
- Best fit signal strength: $5.10^{+2.31}_{-2.05}$ (68% CL)

Conclusions & Next Steps

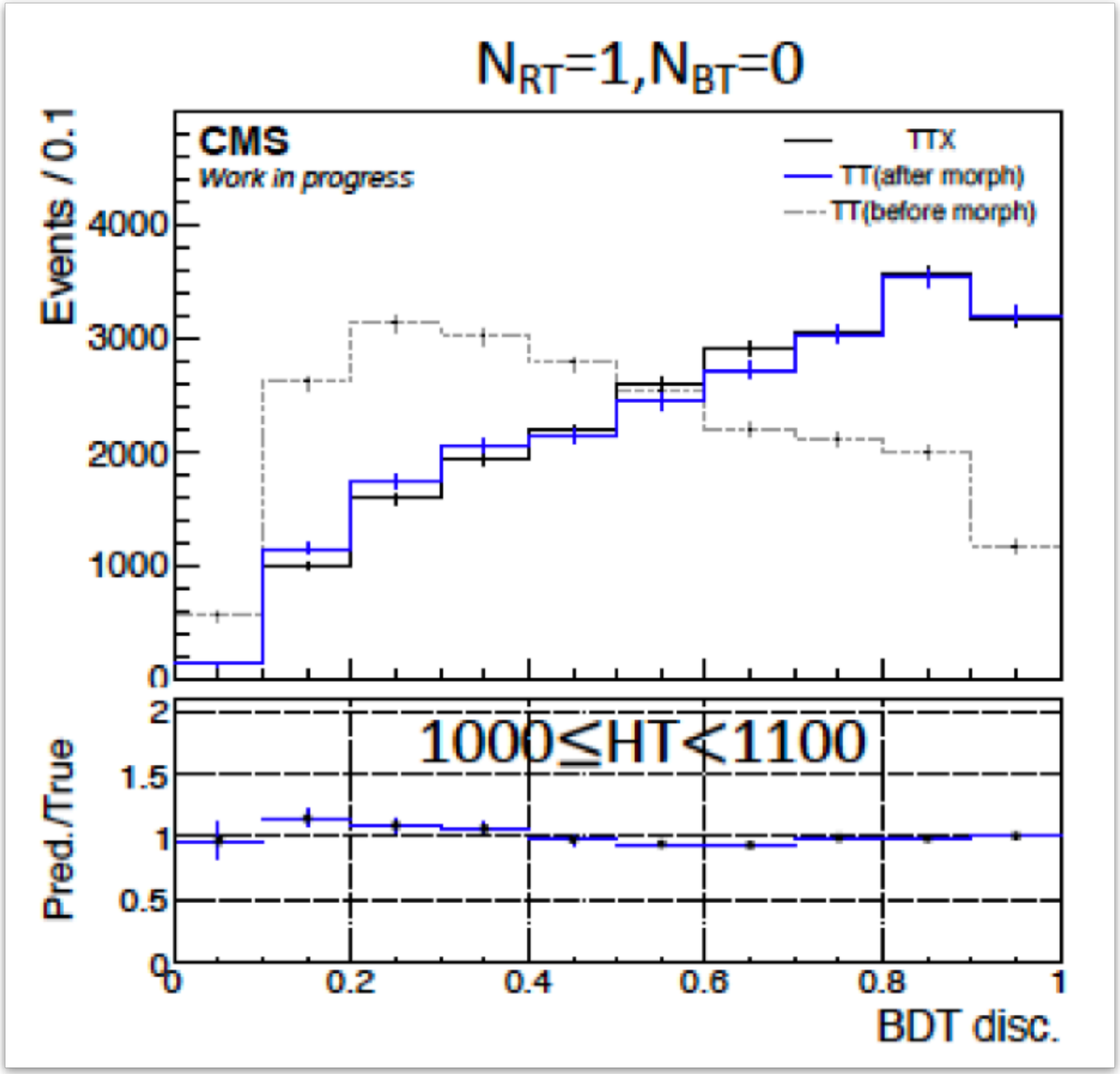
- Analysis status: currently under review
  - Combined CMS Run II tttt result in multiple final states expected this summer
  - Defended Ph.D. March 2022
- Participated in a complex and challenging analysis built from scratch
  - Exciting results, with signal excess to be explored in future BSM interpretations
  - First all-hadronic four-top analysis
  - Showcases novel ML tools potentially useful for future analyses
  - Motivates possible future BSM interpretations

Backup

QCD+$t\bar{t}$ Shape

- Example: Closure test morphing $t\bar{t}$ simulated samples to predict $t\bar{t}$X rather than data-driven QCD+$t\bar{t}$
  - Can predict final shape even when input and target distributions are very different

— $t\bar{t}$X
— predicted $t\bar{t}$ distribution
- - - original $t\bar{t}$ distribution



Plots by Prof. Suyong Choi

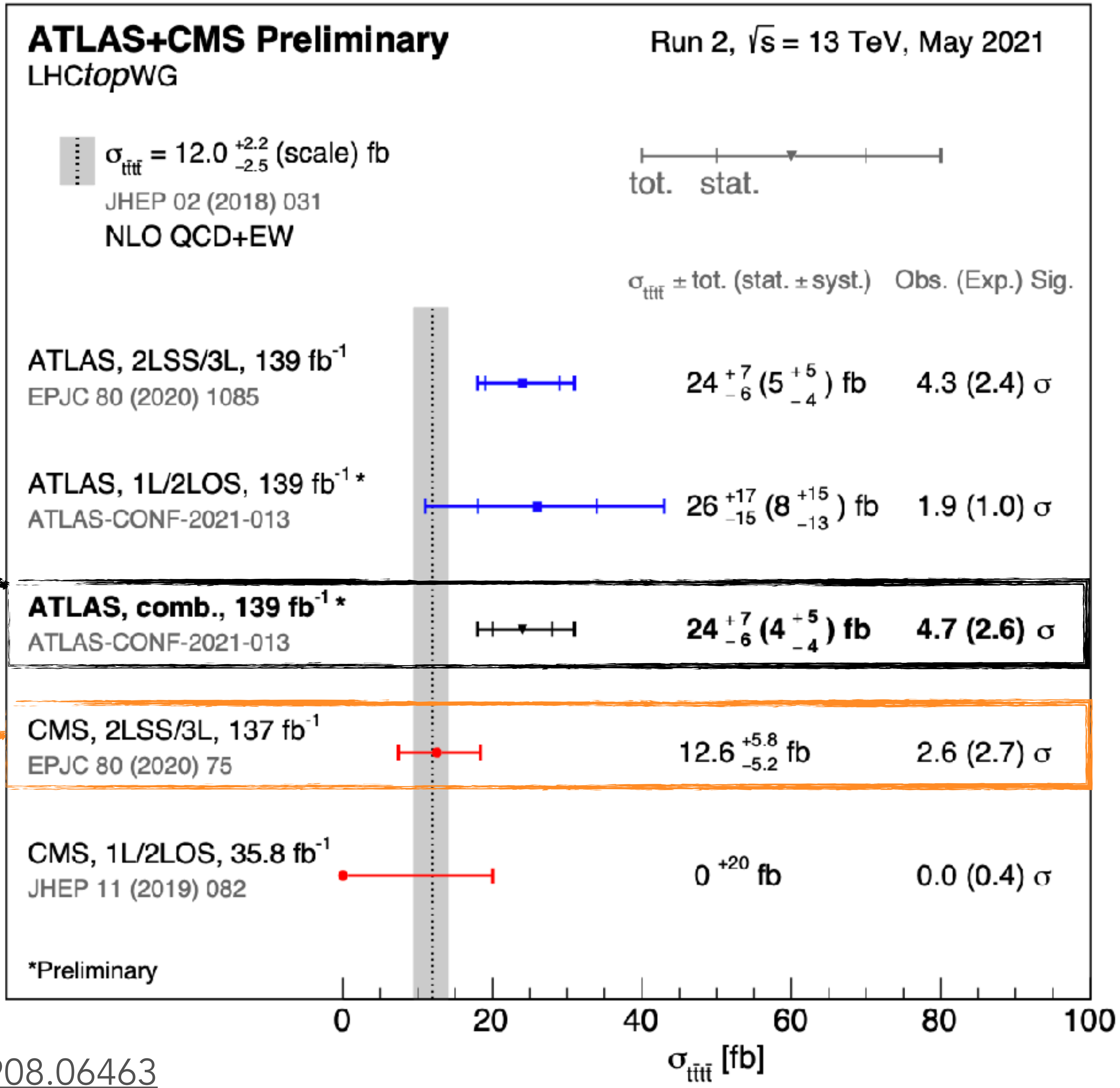
Published Results

ATLAS:

Recent Run II ATLAS Results*
Combined multiple final states:
Same-Sign Di-Lepton + Multi-Lepton (2LSS/3L) + Single Lepton (1L) + Opposite-Sign Di-Lepton (2LOS)

CMS:

Same-Sign 2-Lepton + Multi-Lepton (2LSS/3L) published Run II analysis+



σ = standard deviations
3σ = "evidence"
5σ = "discovery"

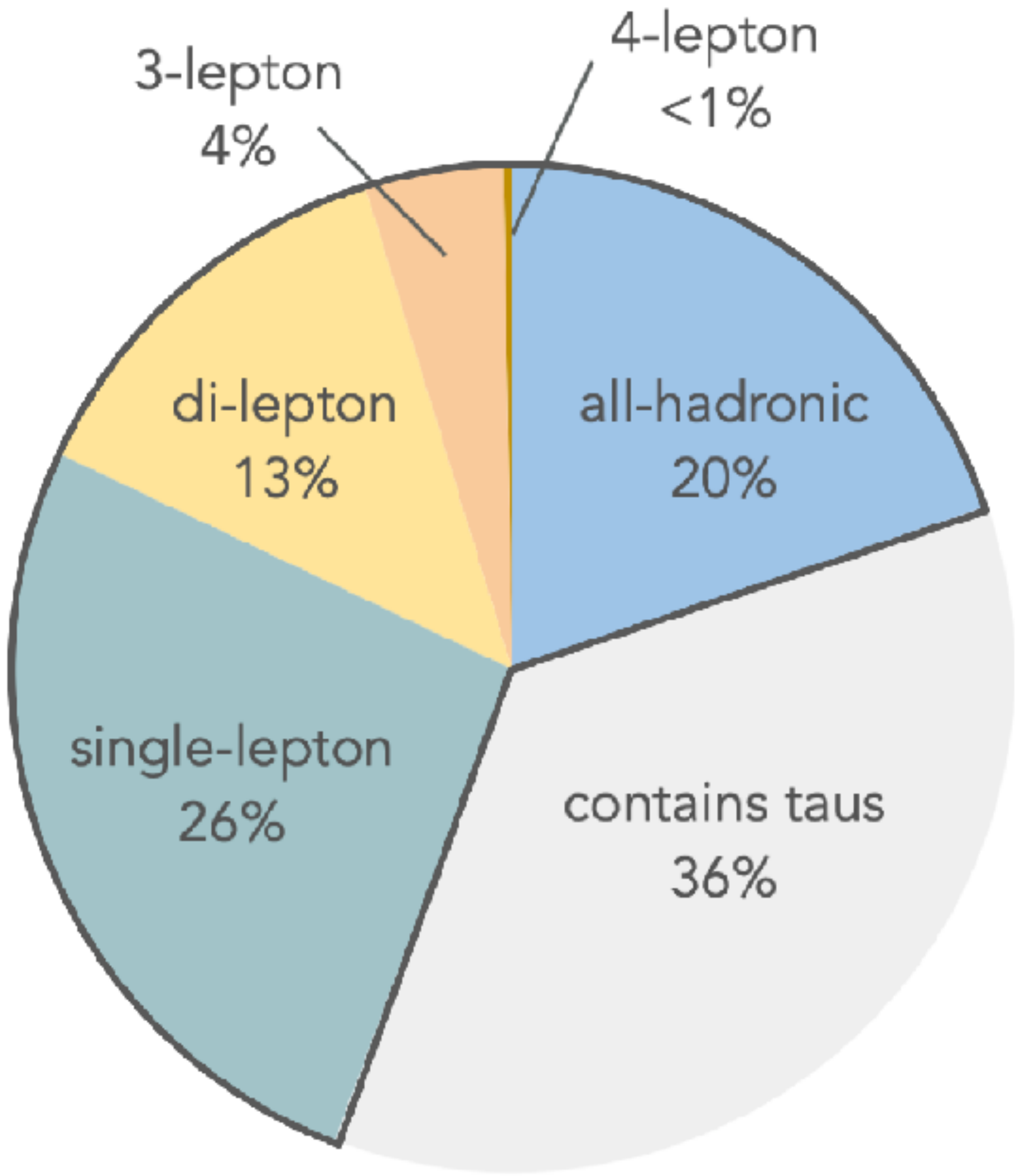
*ATLAS Collaboration. arXiv:2007.14858
+CMS Collaboration. CMS-TOP-18-003. arXiv:1908.06463

Four-Tops - CMS Multi-Channel Collaboration

CMS-TOP-21-005

- CMS Run II combination of multiple final states:

= existing Same-Sign Di-Lepton/Multi-Lepton (2LSS/3L) + updated 2LOS + 1L + new AllHad



"Opposite-Sign Di-Lepton" "Single-Lepton" "All-Hadronic"

published 2016 analysis*

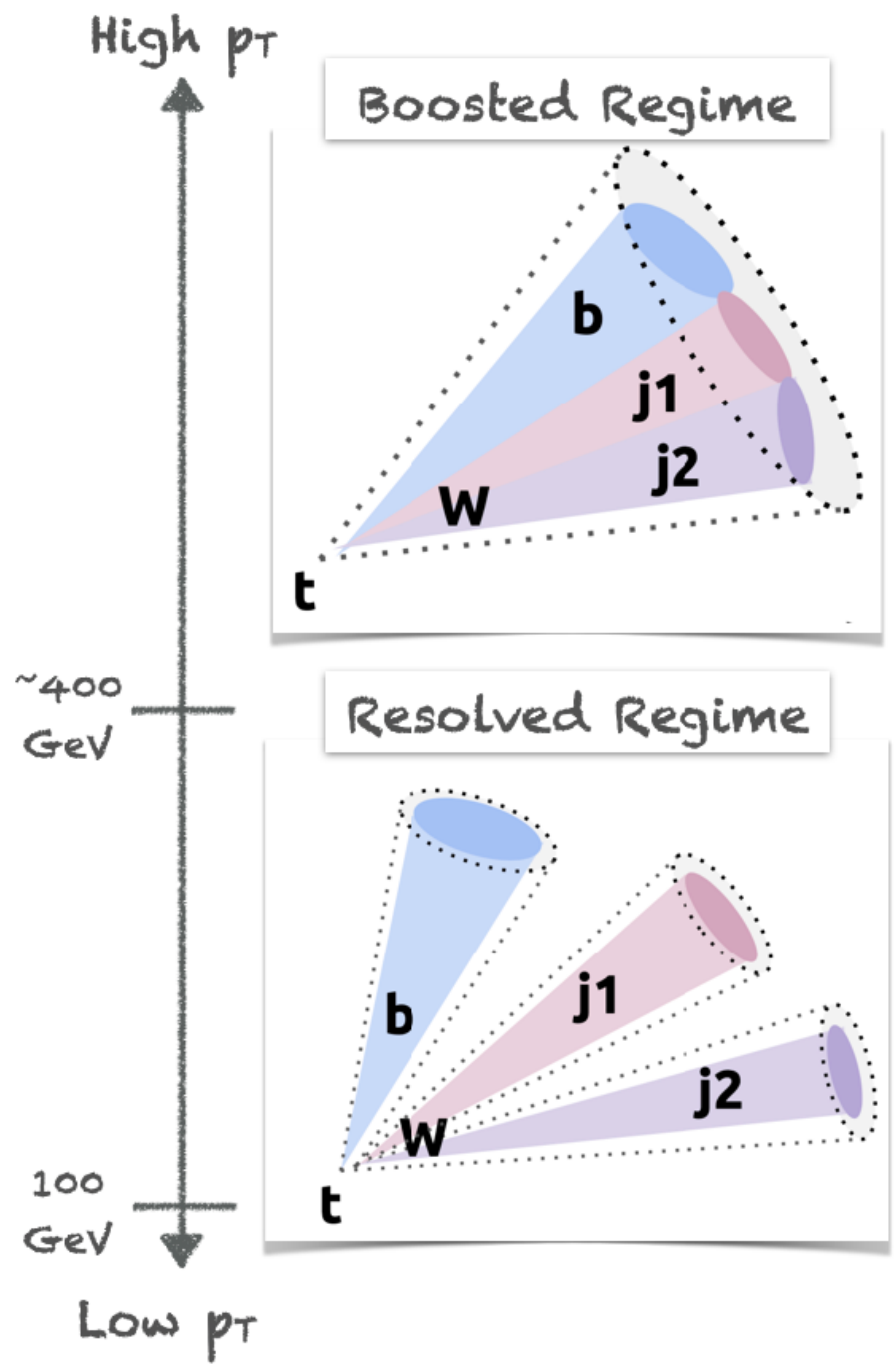
first zero-lepton $t\bar{t}t\bar{t}$ analysis

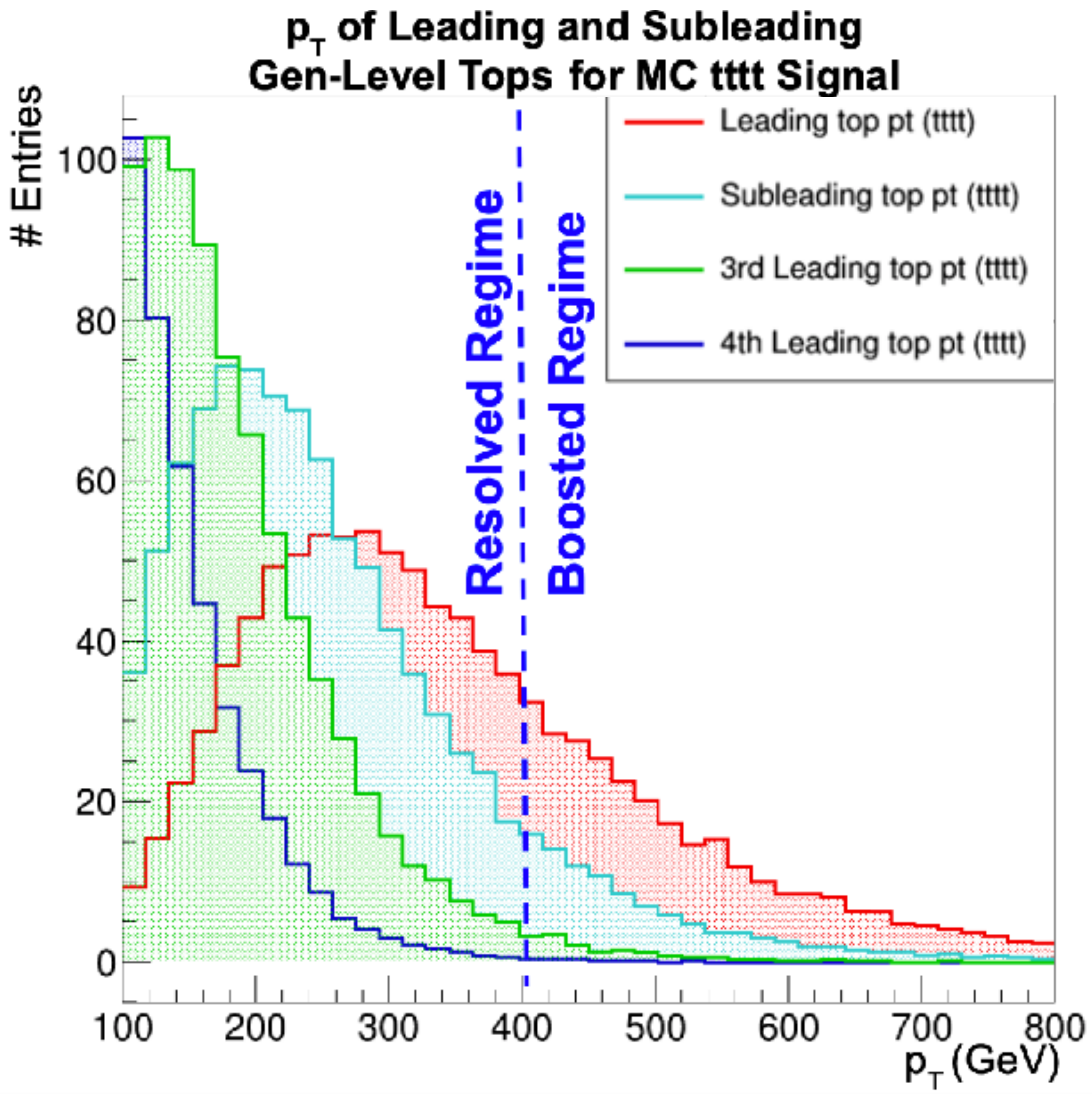
NEW Run II results

*CMS Collaboration. CMS-TOP-17-019. arXiv:1906.02805

Hadronic Top Tagging

*CMS Collaboration. CMS-PAS-JME-18-002





- All-Hadronic analysis uses **boosted** and **resolved** top tagging

"DeepAK8" boosted top tagger*

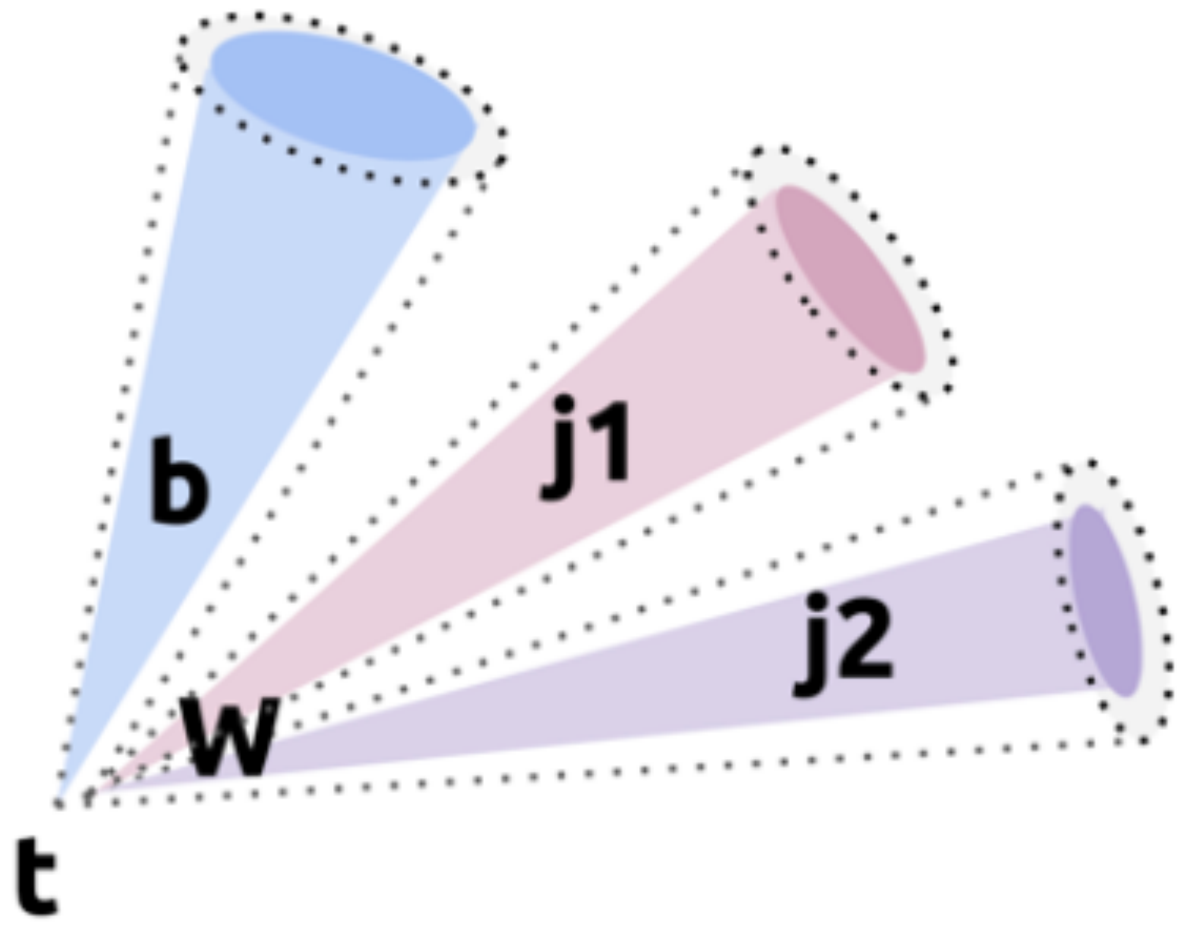
Custom resolved top tagger compatible with centrally-produced data samples:

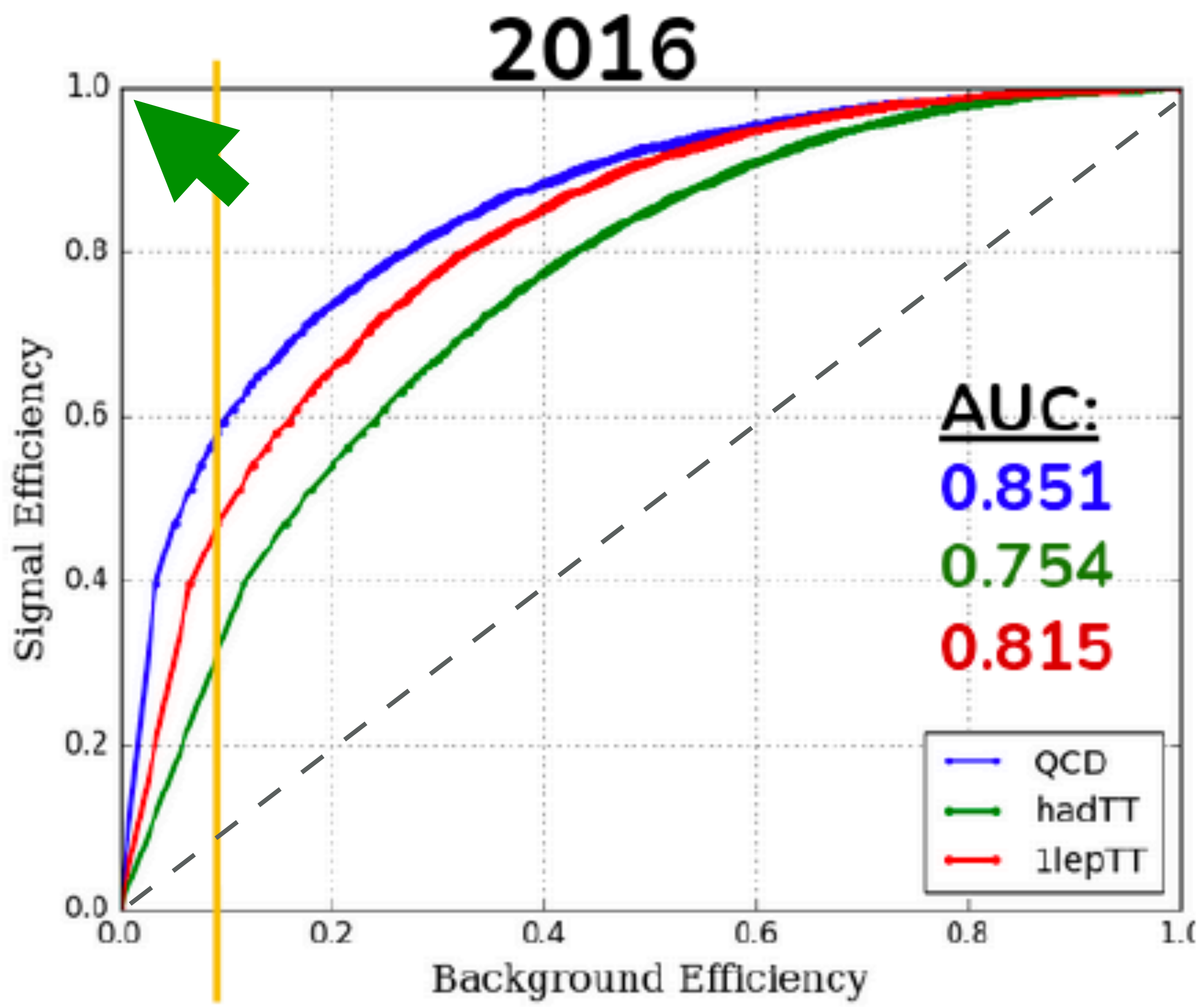
- Significantly reduces time and space needed for computing

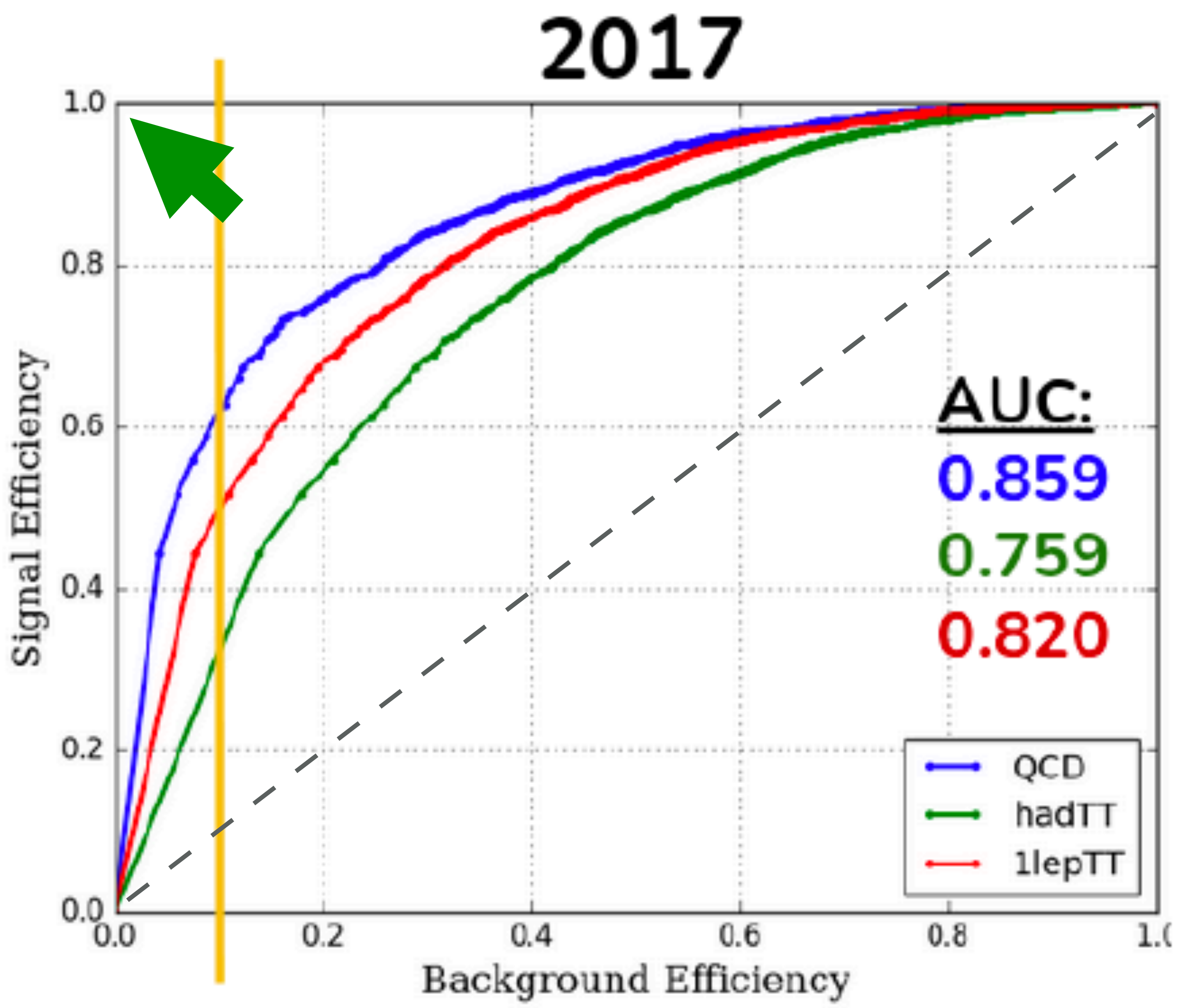
Building a Resolved-Top Tagging Algorithm

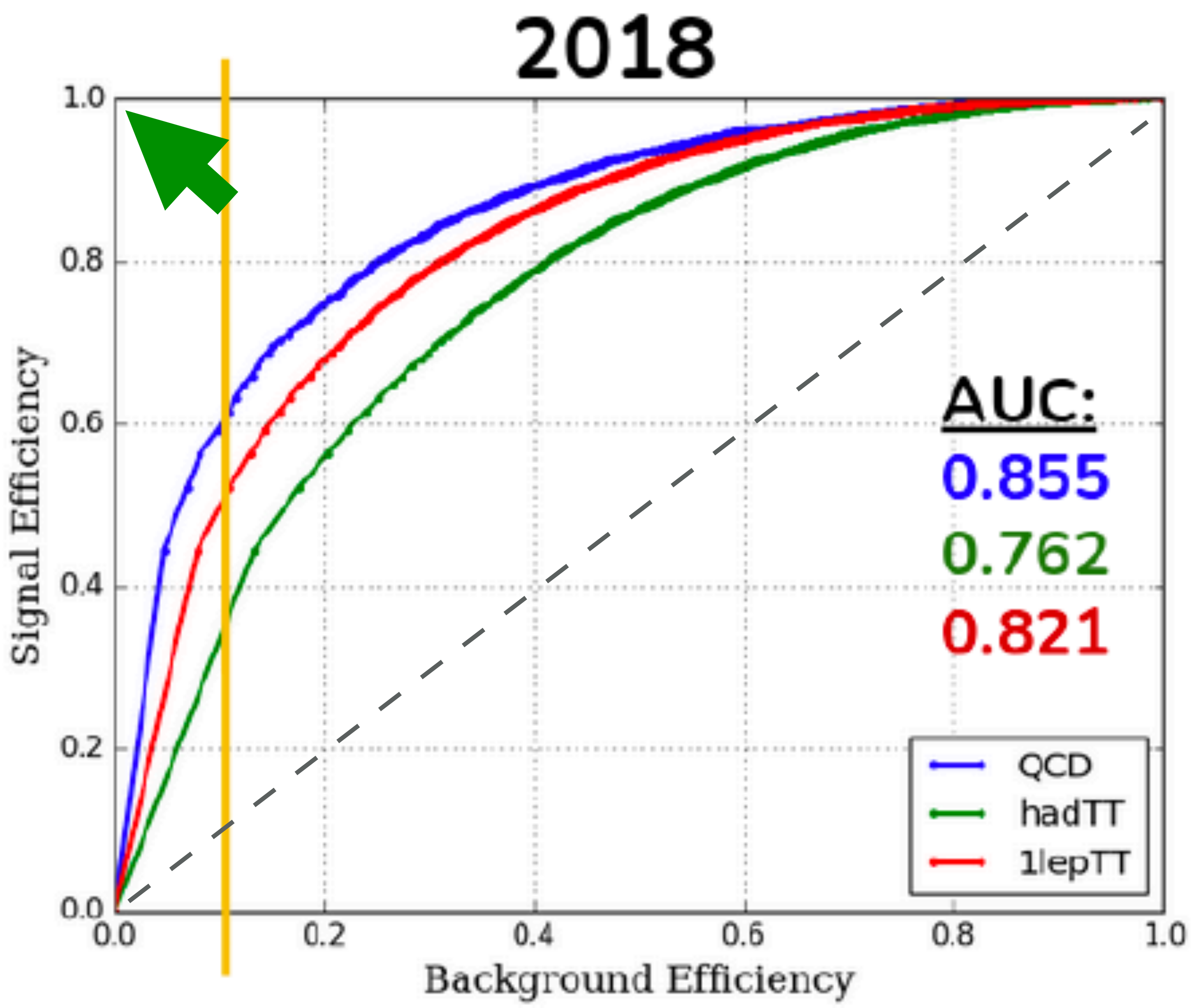
*CMS Collaboration. arXiv:1707.03316

- Custom resolved top tagger based on tagger from 2016 SUSY stop analysis*
  - Assembles 3-jet top candidates (bJet +2 jet candidates from W)
  - BDT trained on single-lepton+dilepton $t\bar{t}$ simulated samples
  - Inputs b and c tagging information & jet kinematics









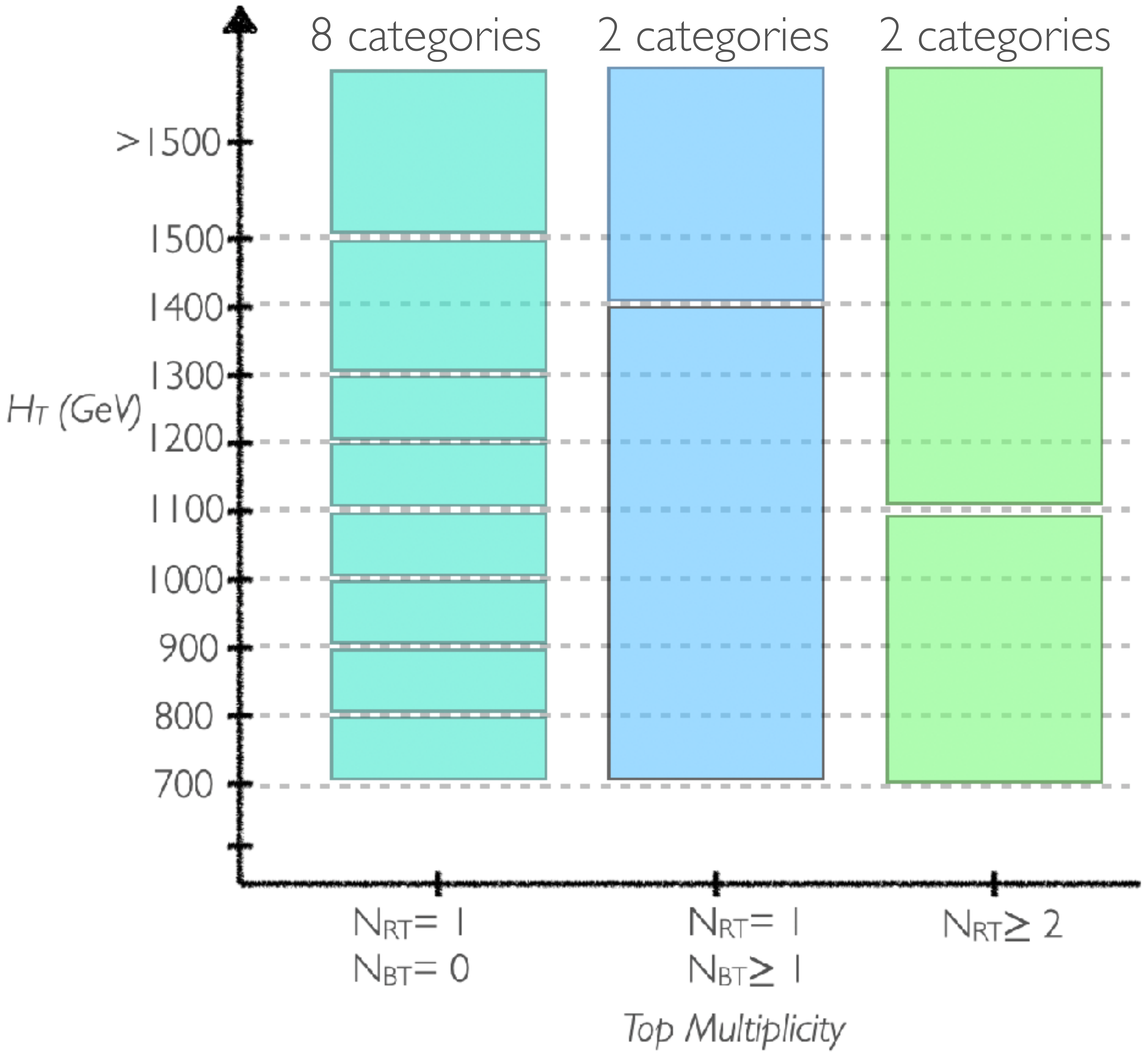
Selecting Events

- Categorize signal region (SR) by:
  - Resolved top multiplicity (1, 2+)
  - Boosted top multiplicity (0, 1+)
  - H_T

Baseline Selection:

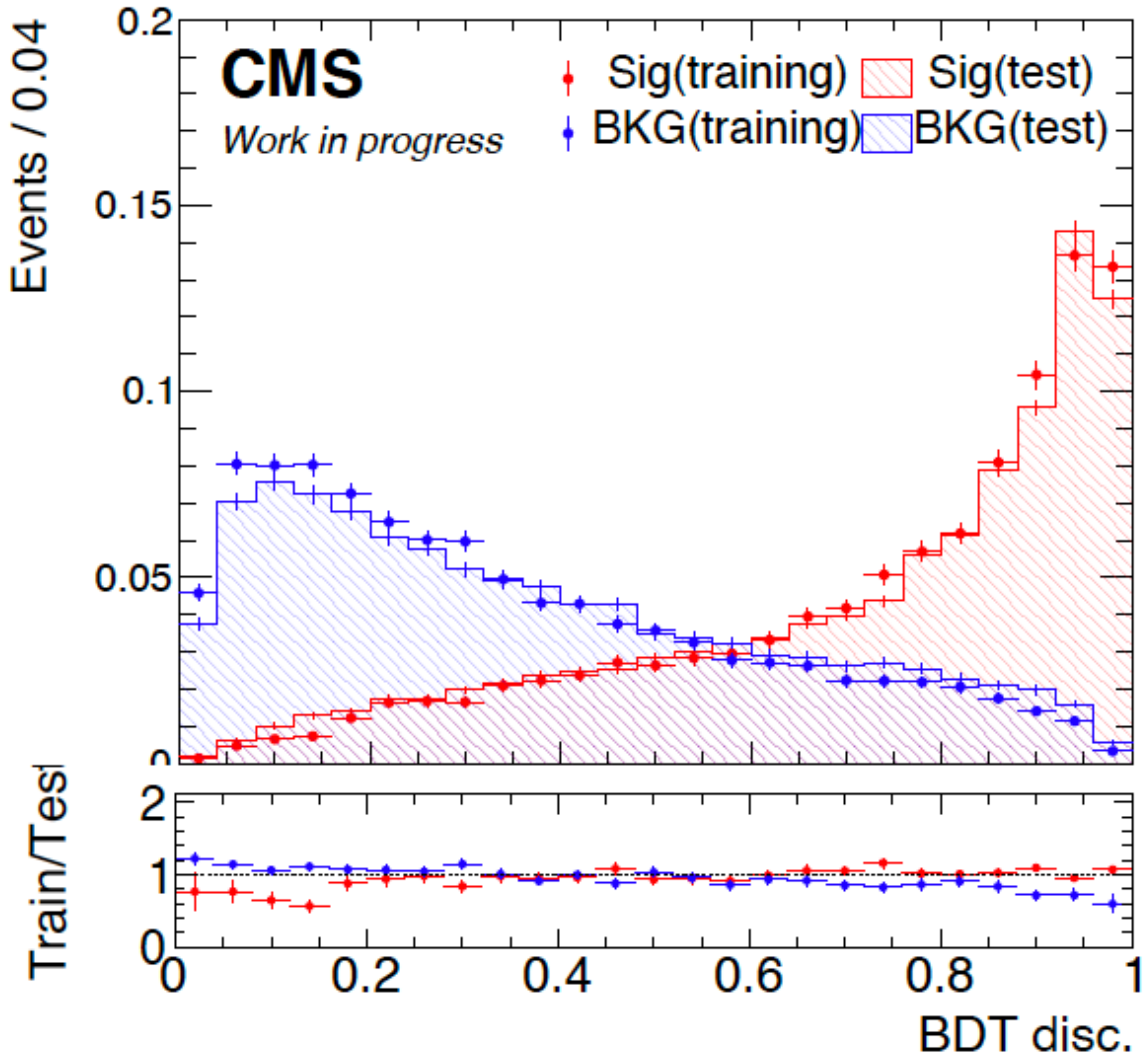
N(leptons)=0, N(Jets)≥9, N(bJets)≥3, H_T≥700 GeV

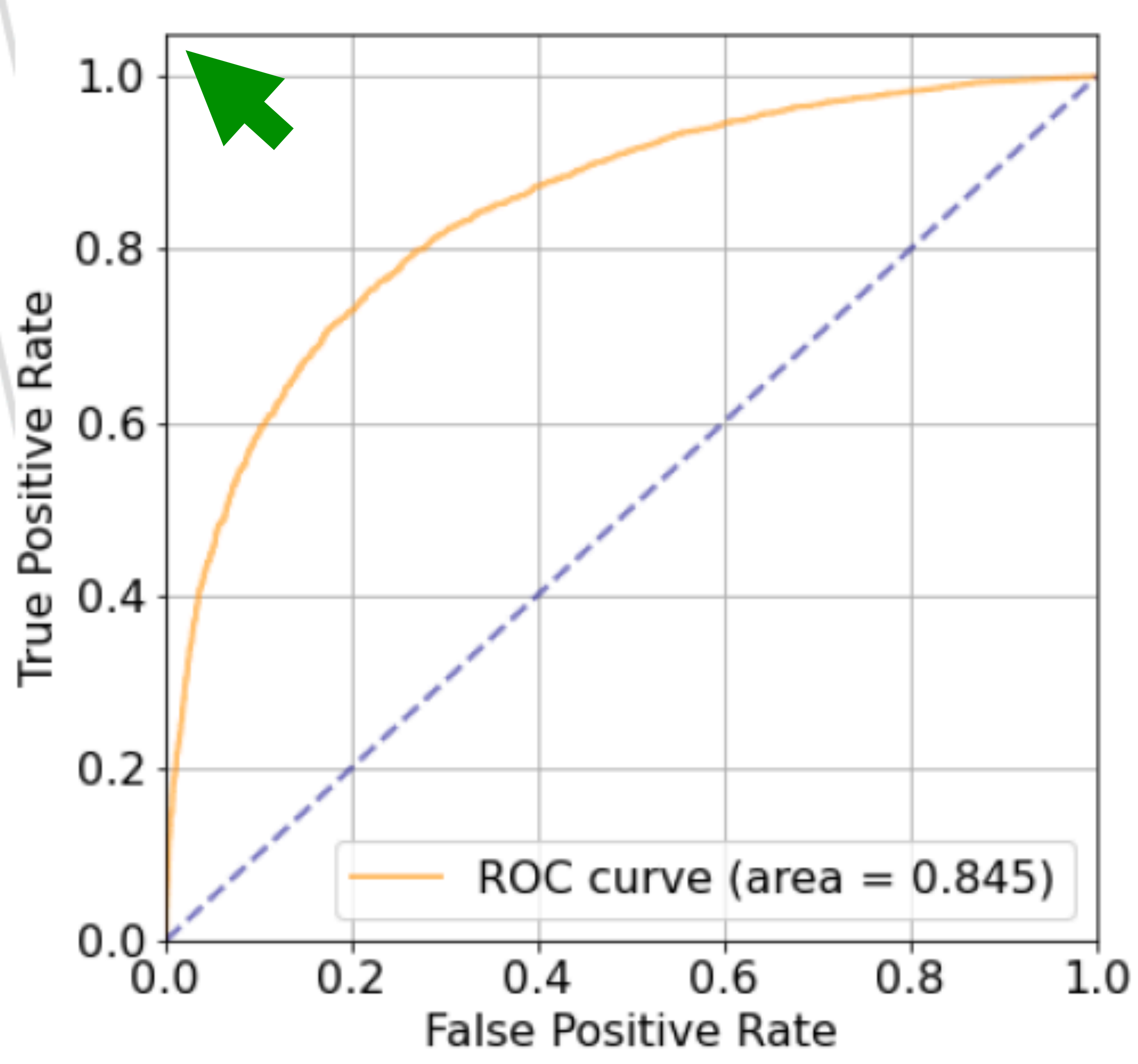
Summary of SR Categories



Discriminating Signal from Background - BDT

- Trained a boosted decision tree (BDT) to discriminate between $t\bar{t}t\bar{t}$ and QCD+$t\bar{t}$ events
- Input optimized set of 20 kinematic variables related to jets, bJets, resolved tops and angular variables





Plots by Hayoung Oh

Documentation

Important links:

- CADI line
- Most recent AN
- Thesis draft (under review)

VR-Derived Uncertainties

- Derive two uncertainties in the VR in order to account for discrepancies between BDT predictions and data
- Uncertainties derived in VR and applied to corresponding SR category

1. VR normalization uncertainty

$$VR_{norm} = \sqrt{(1 - \underbrace{< mean >}_{})^2 + \underbrace{< RMS >^2}_{}}$$

Deviation from 1 of weighted mean of data/prediction ratios across BDT bins
~**offset** in normalization

Weighted RMS of data/prediction ratios across BDT bins
~**spread** of disagreement

Statistics committee recommendation derived using mean and RMS of data/prediction ratios weighted across BDT histogram bins.

2. VR shape uncertainty

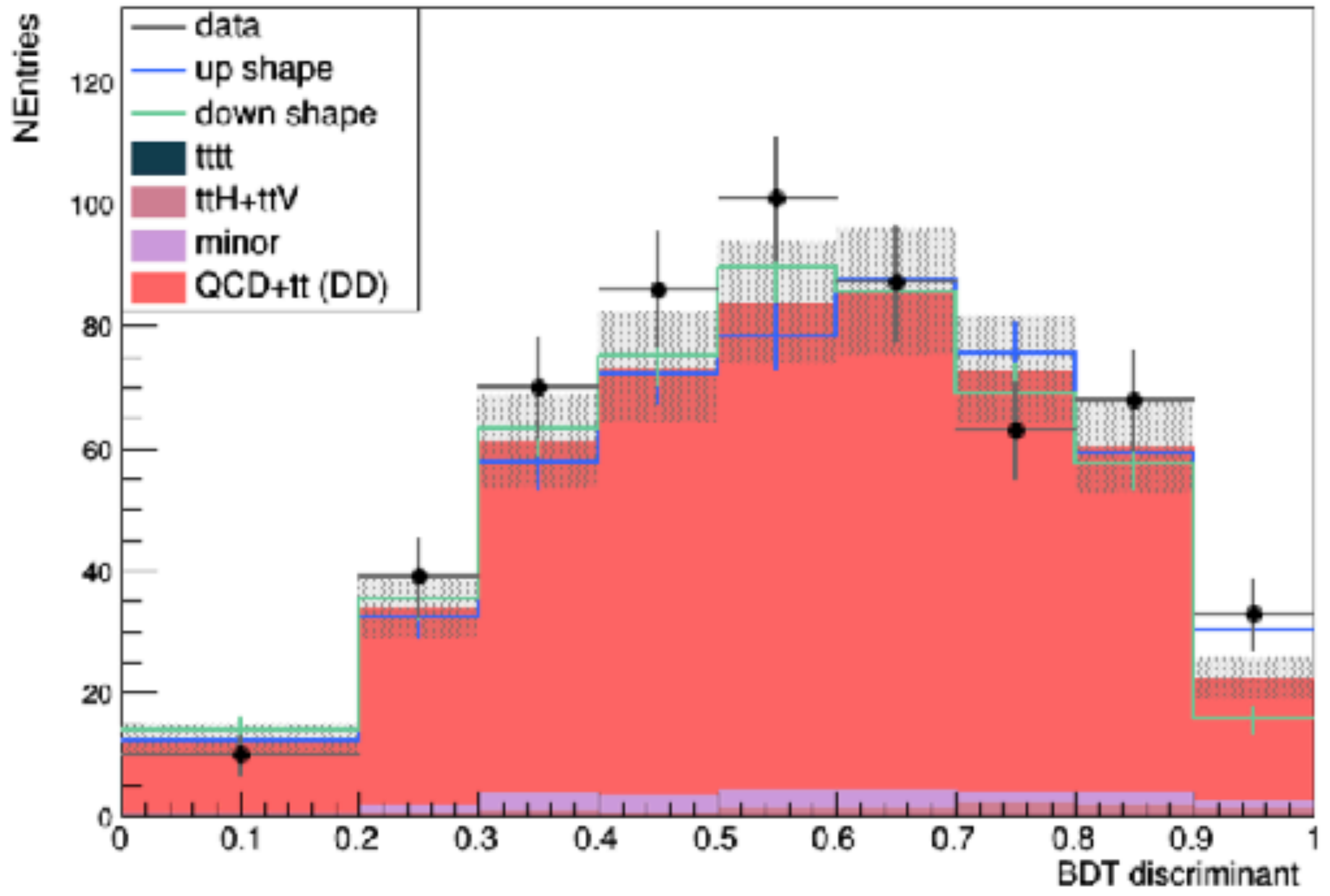
Example: 2% shift in BDT value of each event

Up:

$$BDT_{up} = BDT + (BDT \cdot 0.02)$$

Down:

$$BDT_{down} = BDT - (BDT \cdot 0.02)$$



Accounts for remaining disagreement in high-BDT bins after applying VR normalization uncertainty by shifting BDT values by % up and down. Minimum shift is 1%. Up and down templates renormalized.

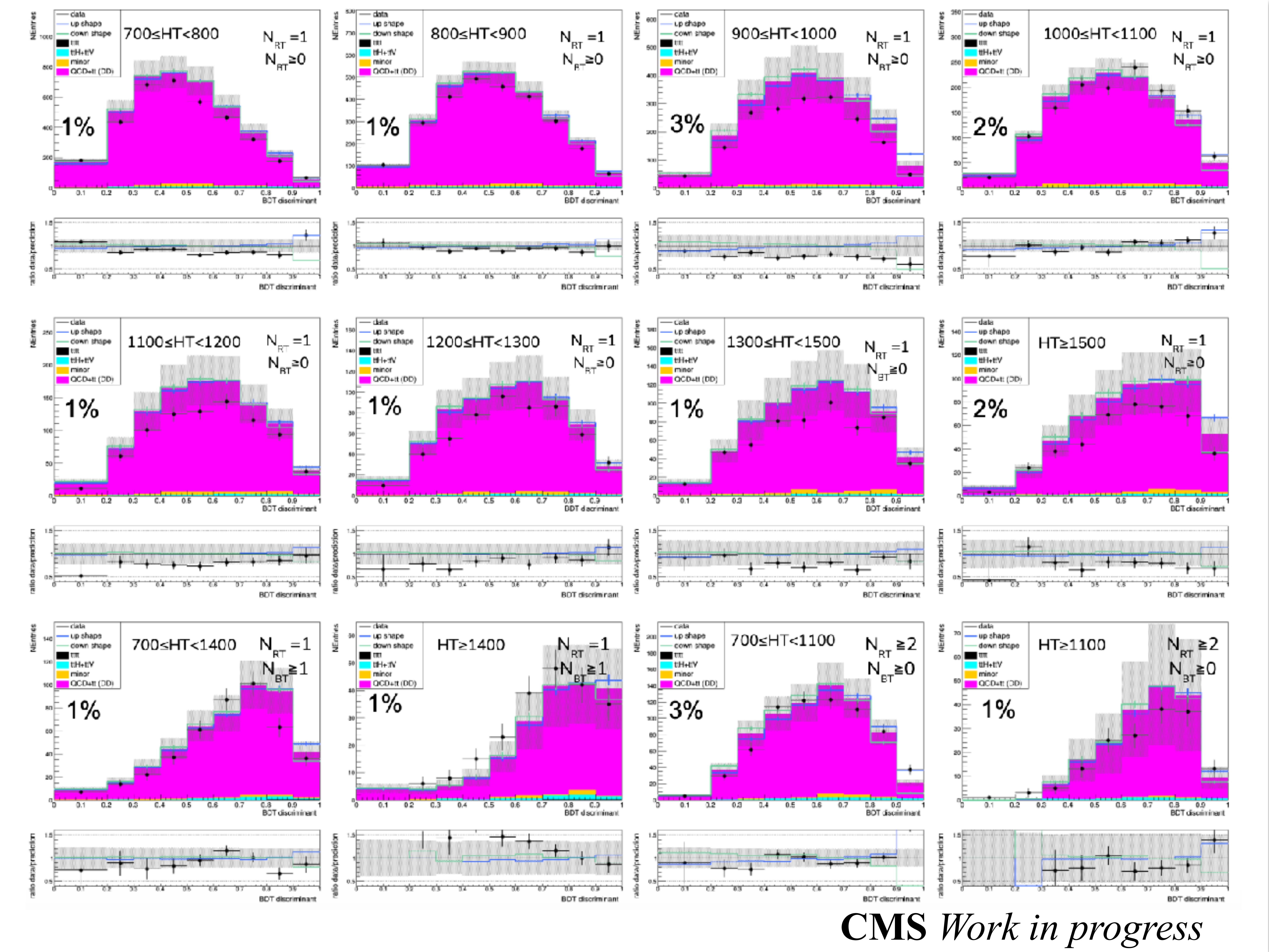
Systematic Uncertainties

- Largest uncertainties come from the discrepancy between data-driven background prediction and data (evaluated in VR) and statistics of QCD+$t\bar{t}$ background
- Not impacted by theoretical uncertainties on modeling of $t\bar{t}$

Systematic Uncertainty	Signal Uncertainty (%)	Background Uncertainty (%)
Discrepancy in VR between data and QCD+tt prediction	-	5-37
Statistics of data-driven backgrounds	-	5-30
ttX (ttH, ttW, ttZ) theoretical cross section normalization	-	26
Jet energy scale (JES)	5-20	5-20
Jet energy resolution (JER)	Up to 20	Up to 20
Statistics of simulated samples	Up to 20	Up to 20
Final state radiation (FSR)	Up to 20	Up to 20
Renormalization and factorization scales (μ_R/μ_F)	Up to 20	Up to 20

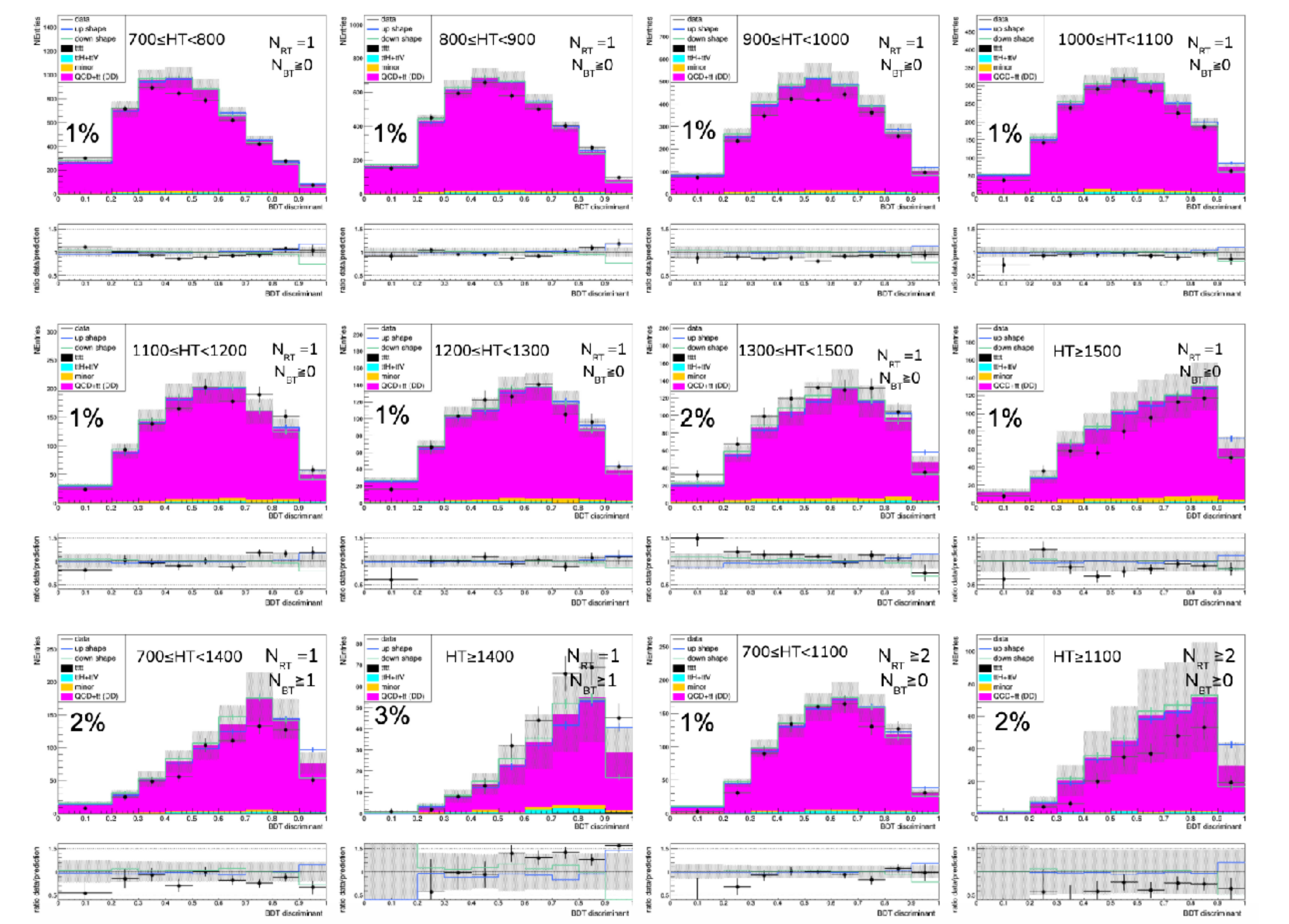
VR Distributions (2017)

- Statistical, VRnorm, and VRshape uncertainties are shown
  - Shape shift % labeled
- 12 VR categories
- Uncertainties propagated to corresponding SR category



VR Distributions (2018)

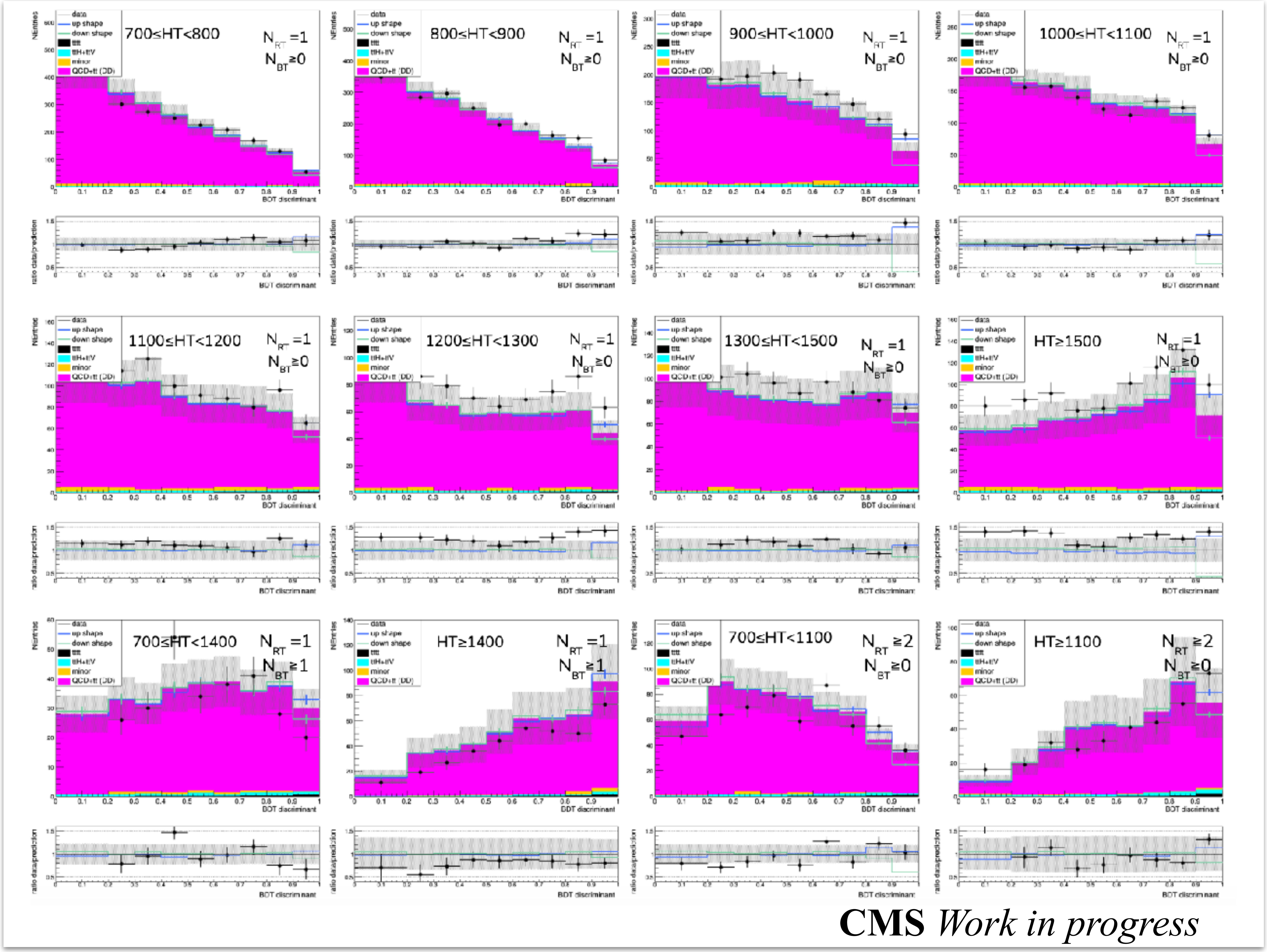
- Statistical, VRnorm, and VRshape uncertainties are shown
  - Shape shift % labeled
- 12 VR categories
- Uncertainties propagated to corresponding SR category



CMS *Work in progress*

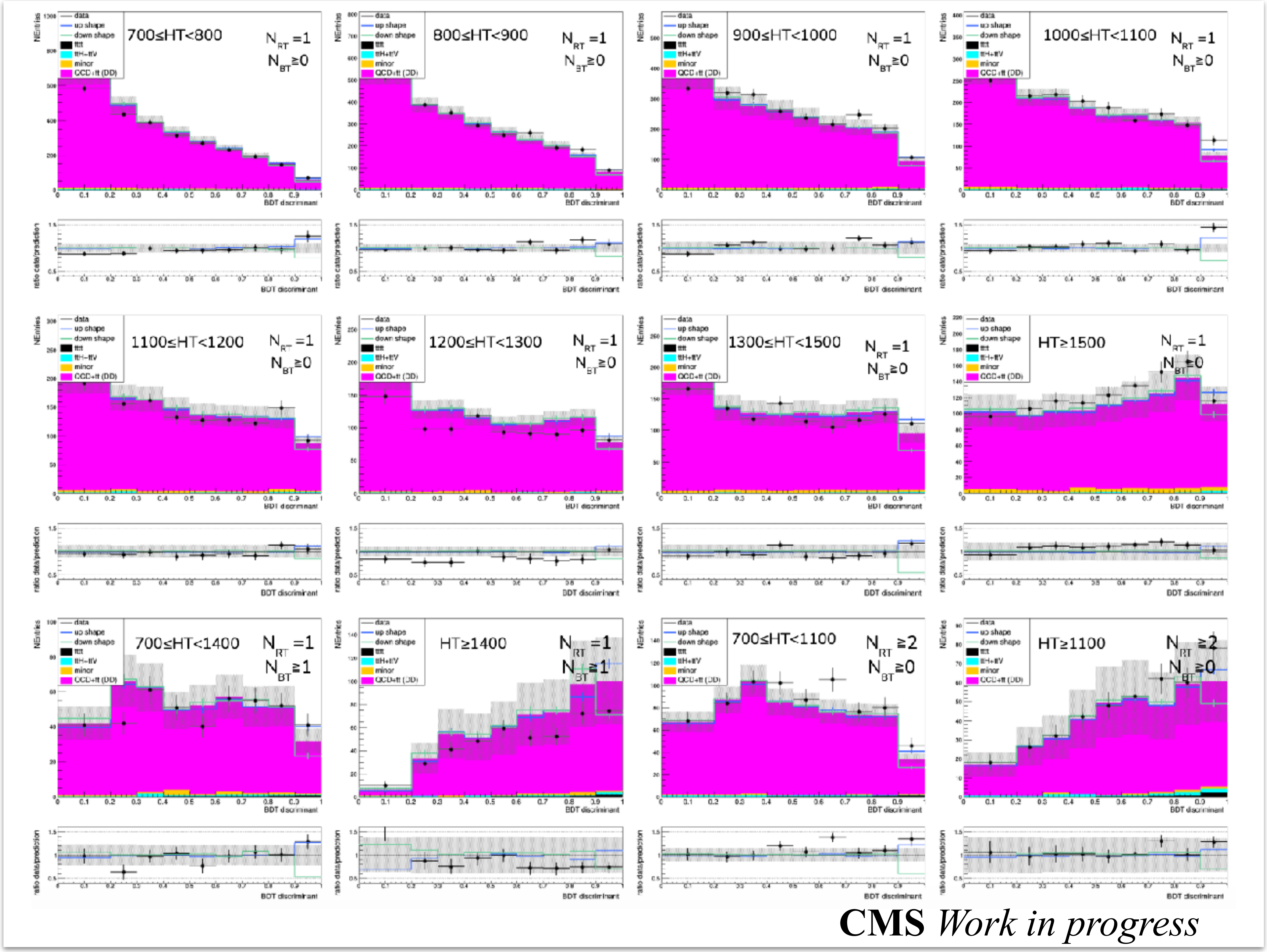
SR Distributions (2017)

- Statistical, VRnorm, and VRshape uncertainties are shown
- 12 SR categories



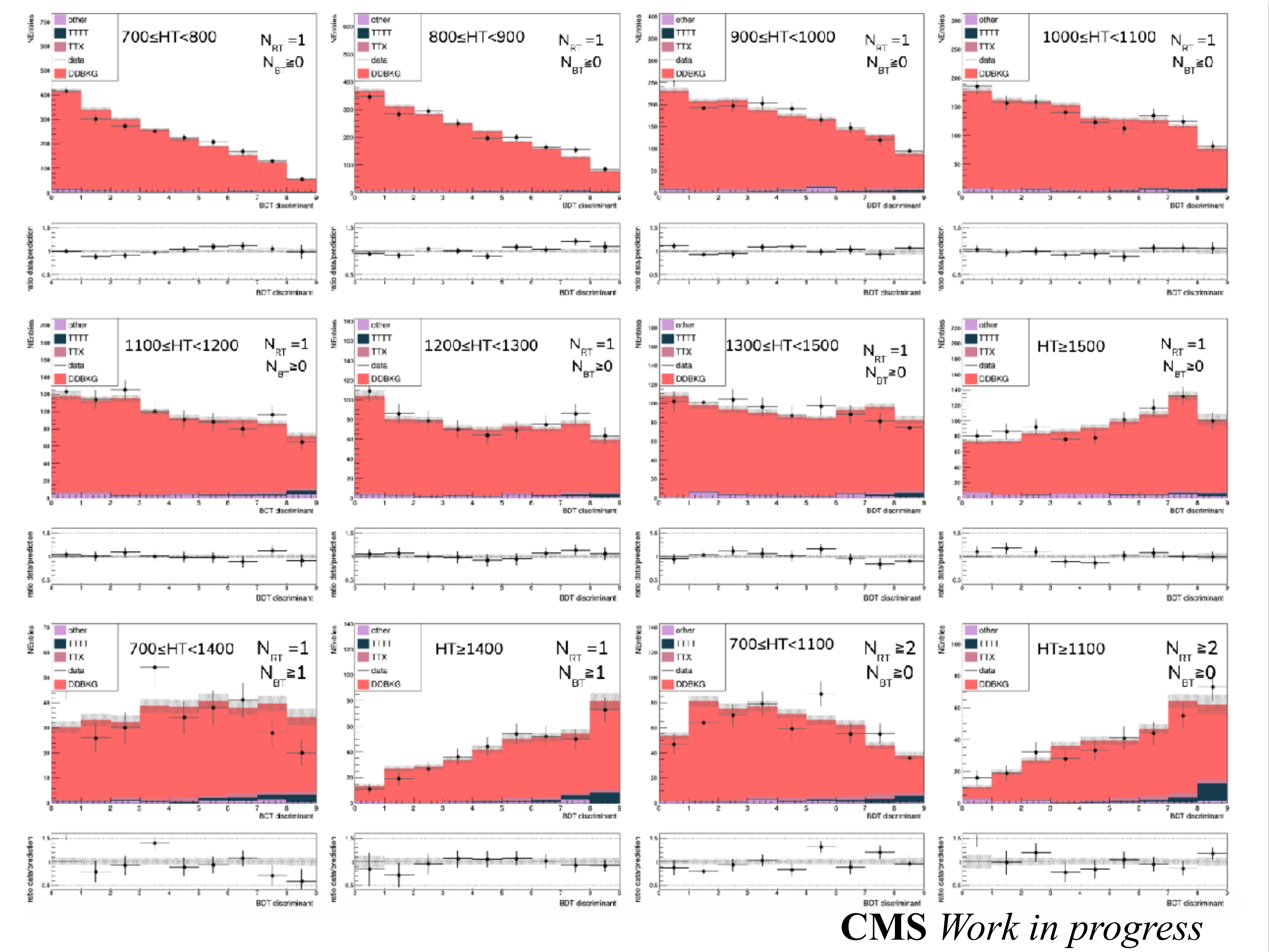
SR Distributions (2018)

- Statistical, VRnorm, and VRshape uncertainties are shown
- 12 SR categories



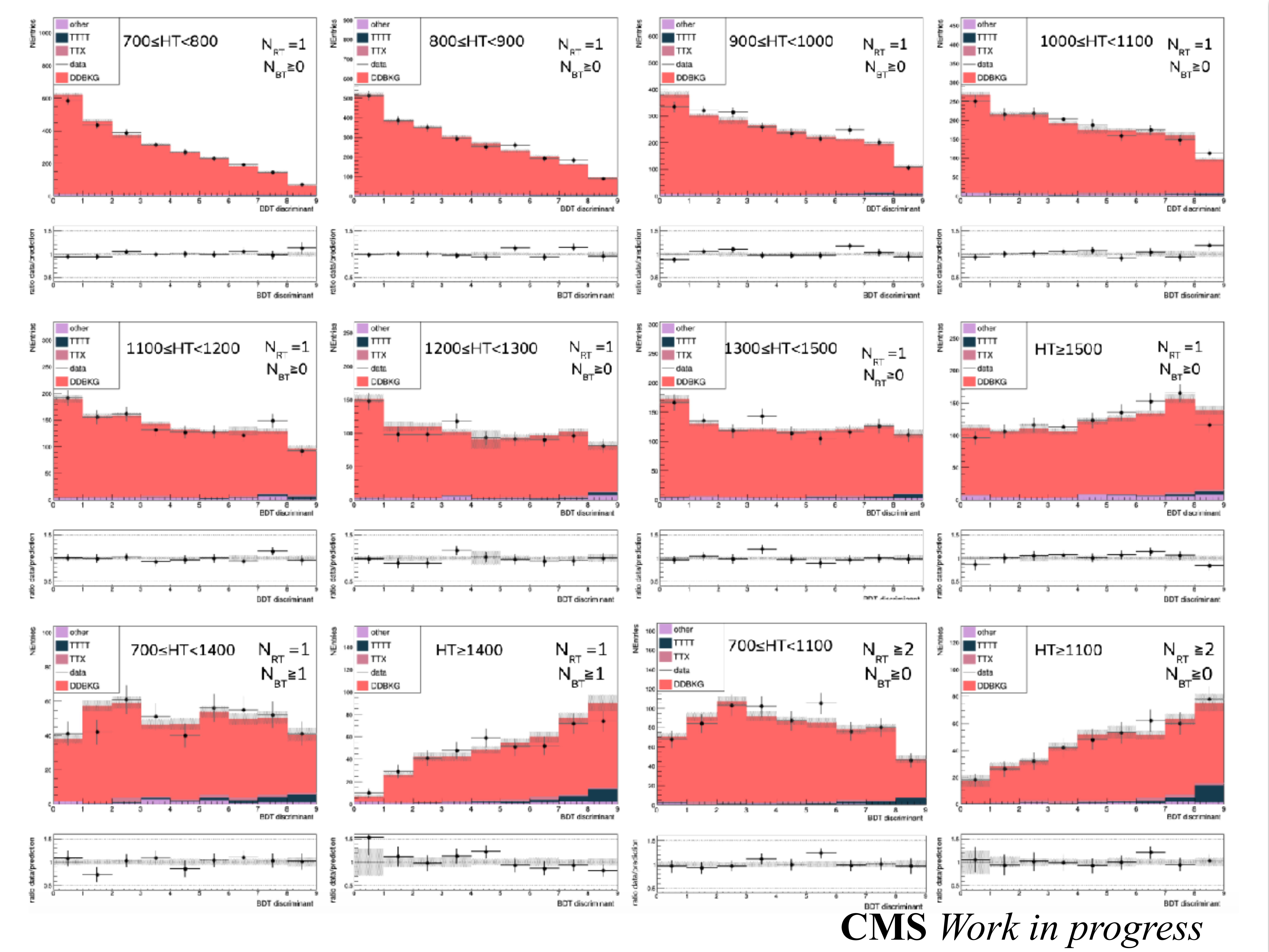
Postfit Distributions (Signal+Background) (2017)

- post-fit uncertainties are shown
- 12 SR categories



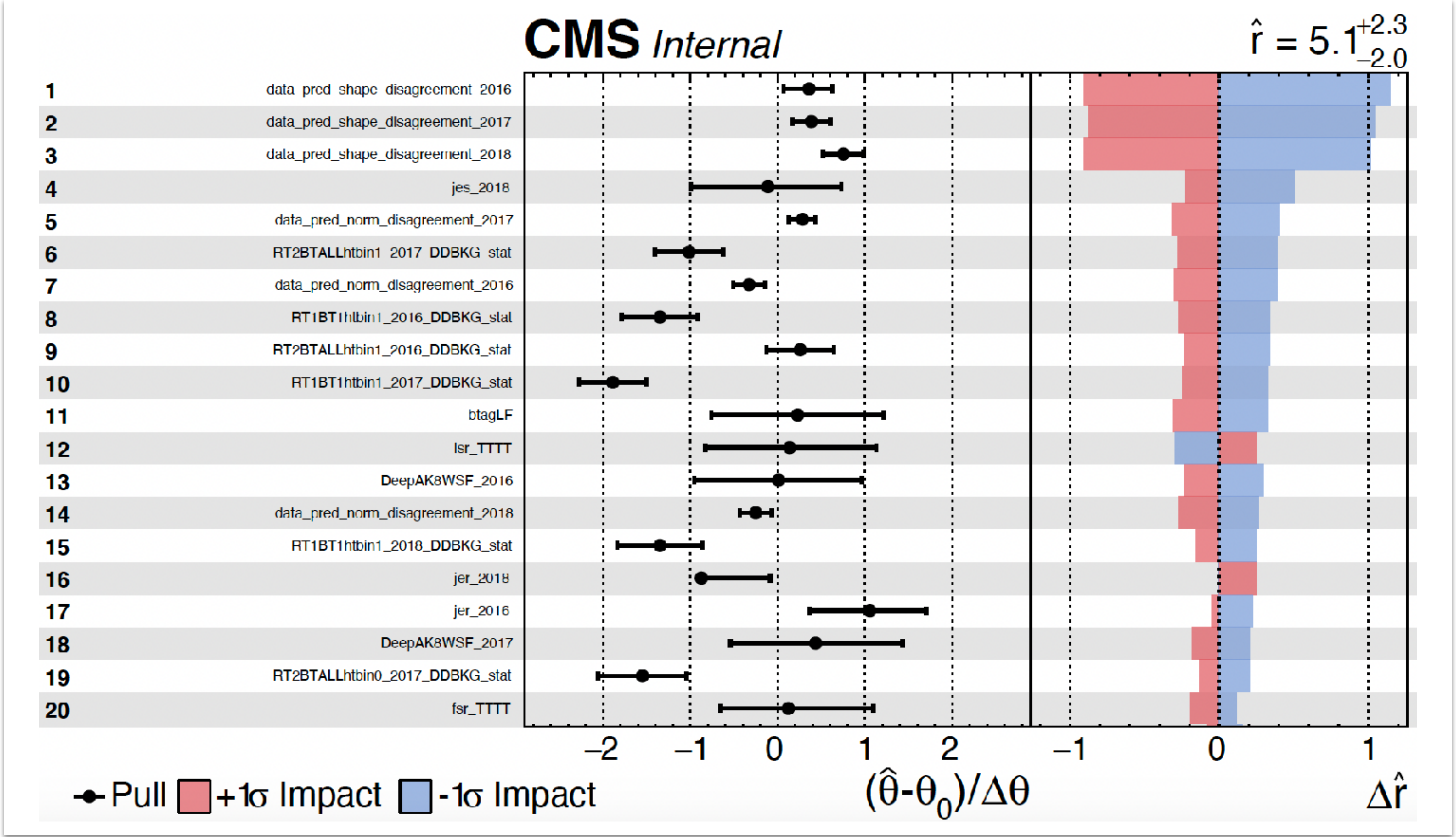
Postfit Distributions (Signal+Background) (2018)

- post-fit uncertainties are shown
- 12 SR categories



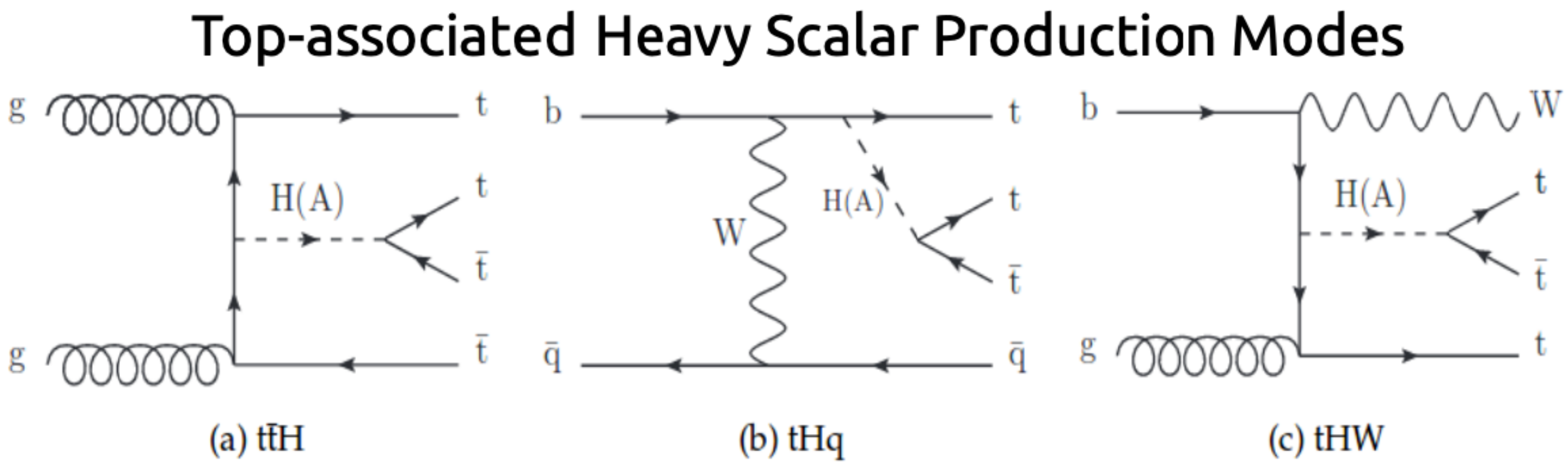
Impacts

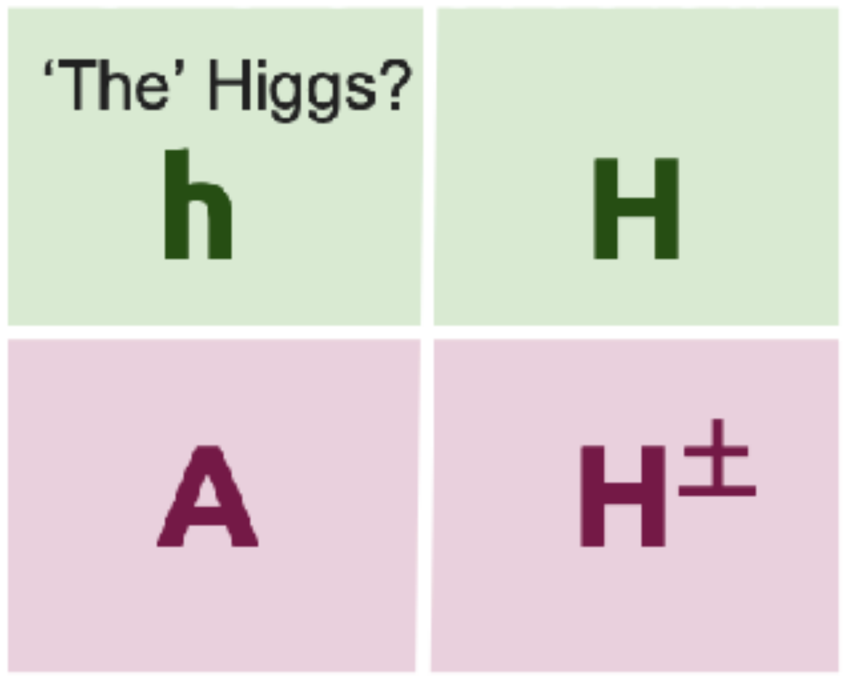
- first 20, ranked
- VR-derived shape uncertainties most important



Four Tops Beyond the Standard Model (BSM)

- Extra BSM Higgs-like bosons decaying to tops would produce an enhanced $t\bar{t}t\bar{t}$ cross section





5 Higgs: One doublet of neutral bosons (including 125 GeV Higgs) and one doublet of a psuedoscalar and heavy charged Higgs

- This is predicted for example by 2HDMs (2 Higgs Doublet Models), where the SM is extended to include 2 doublets of scalar or psuedoscalar bosons

Analysis Strategy

- Final state: all jets (zero leptons)
- Most dominant backgrounds: **ttbar & QCD**
- Categorize signal region (SR) by **top multiplicity**:
  - Resolved top multiplicity (1, 2+)
  - Boosted top multiplicity (0, 1+)
  - H_T
- Variable used to discriminate signal vs. background: **BDT discriminant**
- 2016+2017+2018

Use **data-driven** methods to predict **1) normalization** and **2) shape** of BDT discriminant for QCD+ttbar background

(MC used for minor backgrounds)

Use hadronic top taggers, including a custom NanoAOD-compatible **resolved top tagger**

Trained event-level **BDT** on tttt signal vs. **ttbar + QCD** backgrounds using 20 kinematic variables as inputs

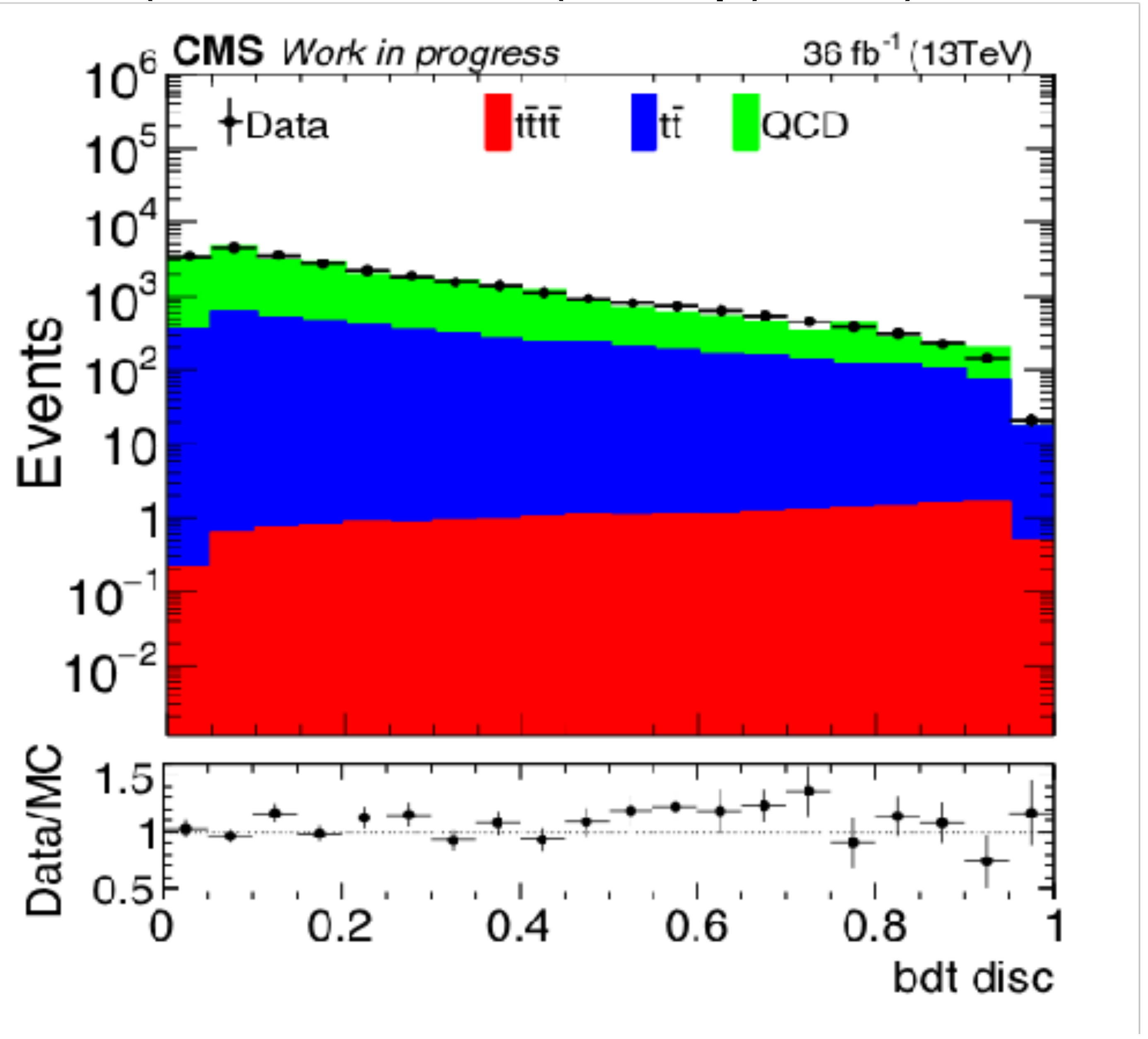
Event Level BDT - Details

- Training: 10k tttt signal, 5k QCD + 5k ttbar background (MC samples, 2016, passing baseline)
- BDT: CatBoost
  - Learning rate: 0.005
  - Loss function: Cross Entropy
  - 1 million iterations
- Tests performed:
  - NN vs. BDT (BDT slightly better performance)
  - CatBoost vs. XGBoost (CatBoost better performance)
  - Training: QCD+ttbar vs. ttbar alone (QCD+ttbar better)
  - Input optimization

Inputs

- N(Jets)
- N(bJets)
- met
- met/sqrt(HT)
- HT
- N(boosted Ws)
- HT(bJets)
- pT(leading resolved top)
- pT(leading bJet)
- Sum (ak8 puppi jet mass)
- | phi (Jet 1)- phi (Jet2)|
- | phi (bJet 1)- phi (bJet2)|
- | eta (Jet 1)- eta (Jet2)|
- | eta (bJet 1)- eta (bJet2)|
- pT(7th jet)
- sphericity
- aplanarity
- centrality
- Sum(pT(leading 6 jets))/HT
- mean bJet discriminant value

BDT shape comparison for 2016 data vs MC
(baseline + N(restop) = 0)



Resolved Top Tagger - Details

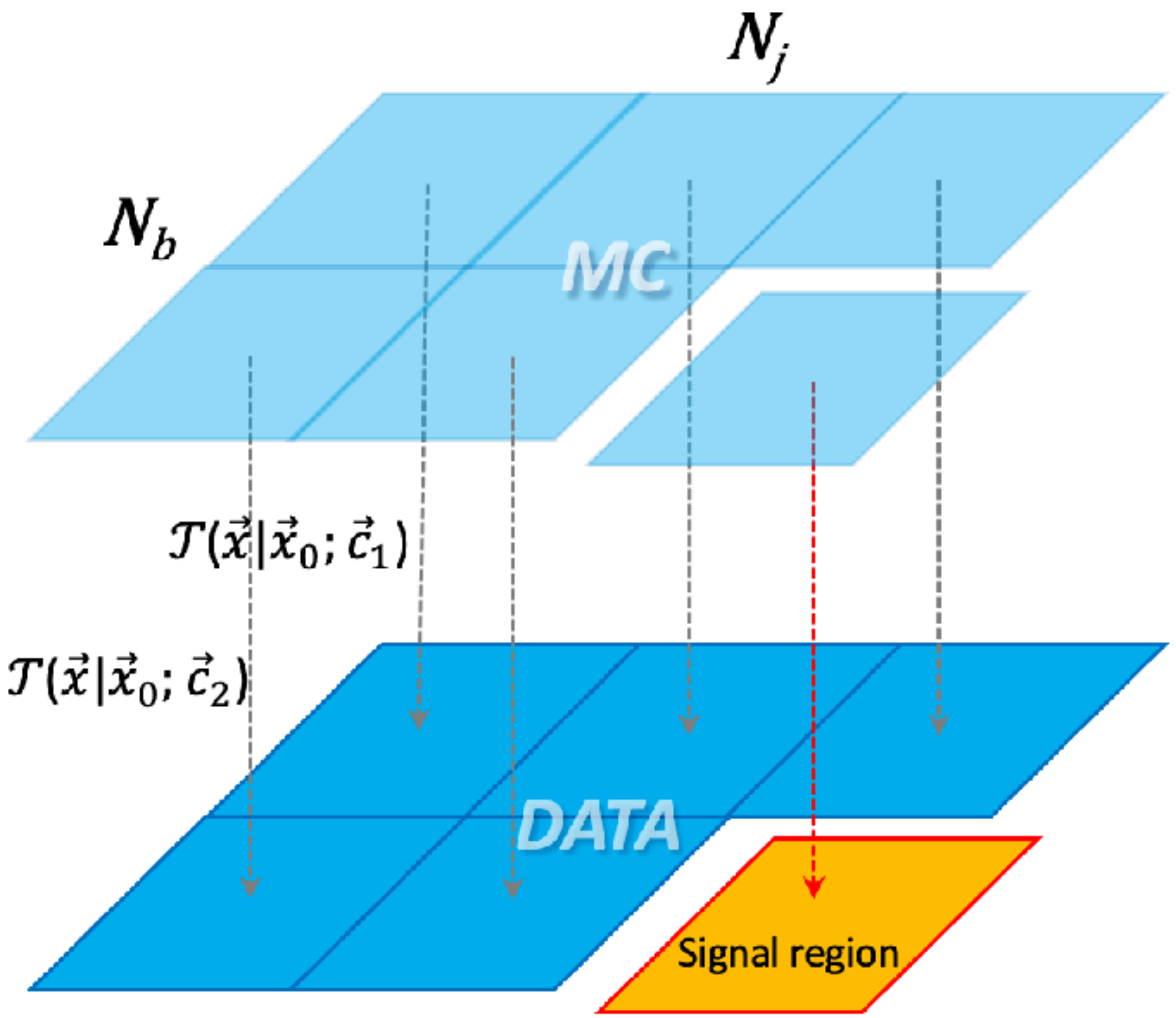
- Training: 100k single-lepton TT signal, 100k dilepton TT background (MC samples, 2016)
  - Training baseline: MET>100, N(jets)>=5, N(deepflavour bjets)>=1
- Candidate requirements:
  - Jets pass jet preselection
  - No jets overlapping within $\Delta R<0.4$ or sharing jets between candidates
  - 2 non b-candidate jets mass within 40 GeV of 80GeV (W candidate), top candidate mass within 80GeV of 175 GeV
  - BDT Score > WP corresponding to 10% mistag rate in QCD
  - Mistag rate and efficiency SFs applied
- BDT: XGBoost
  - Learning rate: 0.05
  - Flat pT reweighting
  - 10% of training used for testing
  - Loss function: logistic regression for binary classification
  - Max depth = 6
  - 2000 iterations

Inputs

- Individual jet masses
- Total top candidate mass
- Total W candidate mass
- Jet DeepJet b-tag scores
- Jet DeepJet c-tag scores
- Jet number of constituents
- Jet quark gluon likelihood scores
- Top cand pT x ΔR(bjet and W cand)
- W cand pT x ΔR(jet1 and jet2)
- $pT_{jet2} x\Delta R(\text{jet1 and jet2})^2 /(pT_{jet1} + pT_{jet2})$

ABCDnn - Details

- Learns the transformation between two distributions (for example data and simulation) in the same condition c
  - For example, each control region and the signal region are under some condition
- These transformations are found using a Neural autoregressive flow
- The training seeks to minimize the maximum-mean-discrepancy (MMD) between the source and target distributions
  - MMD is a measure of the difference between two probability distributions. As a loss function it computes the differential between a predicted distribution (generated by a network) and an observed distribution (of data within a mini-batch). Useful reference.

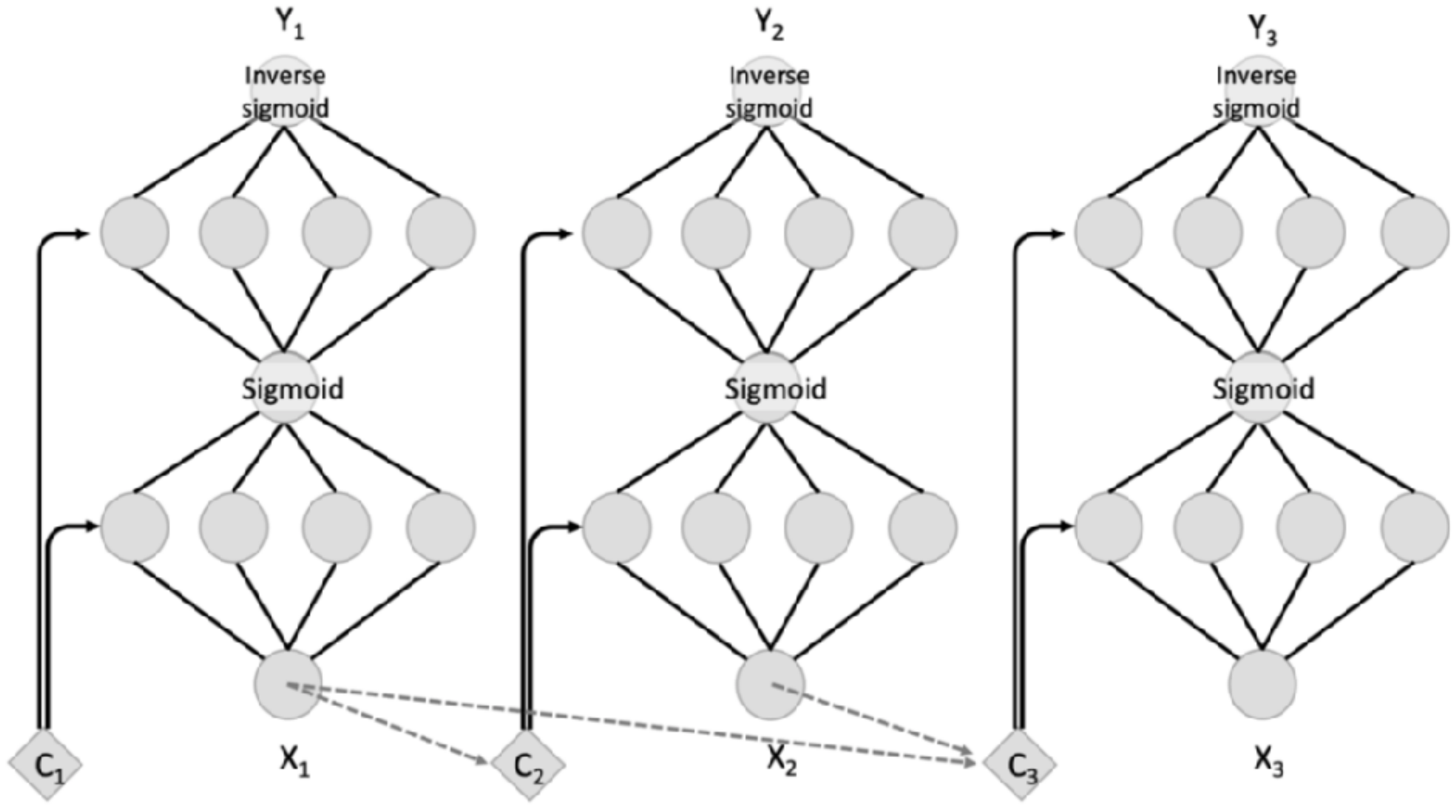


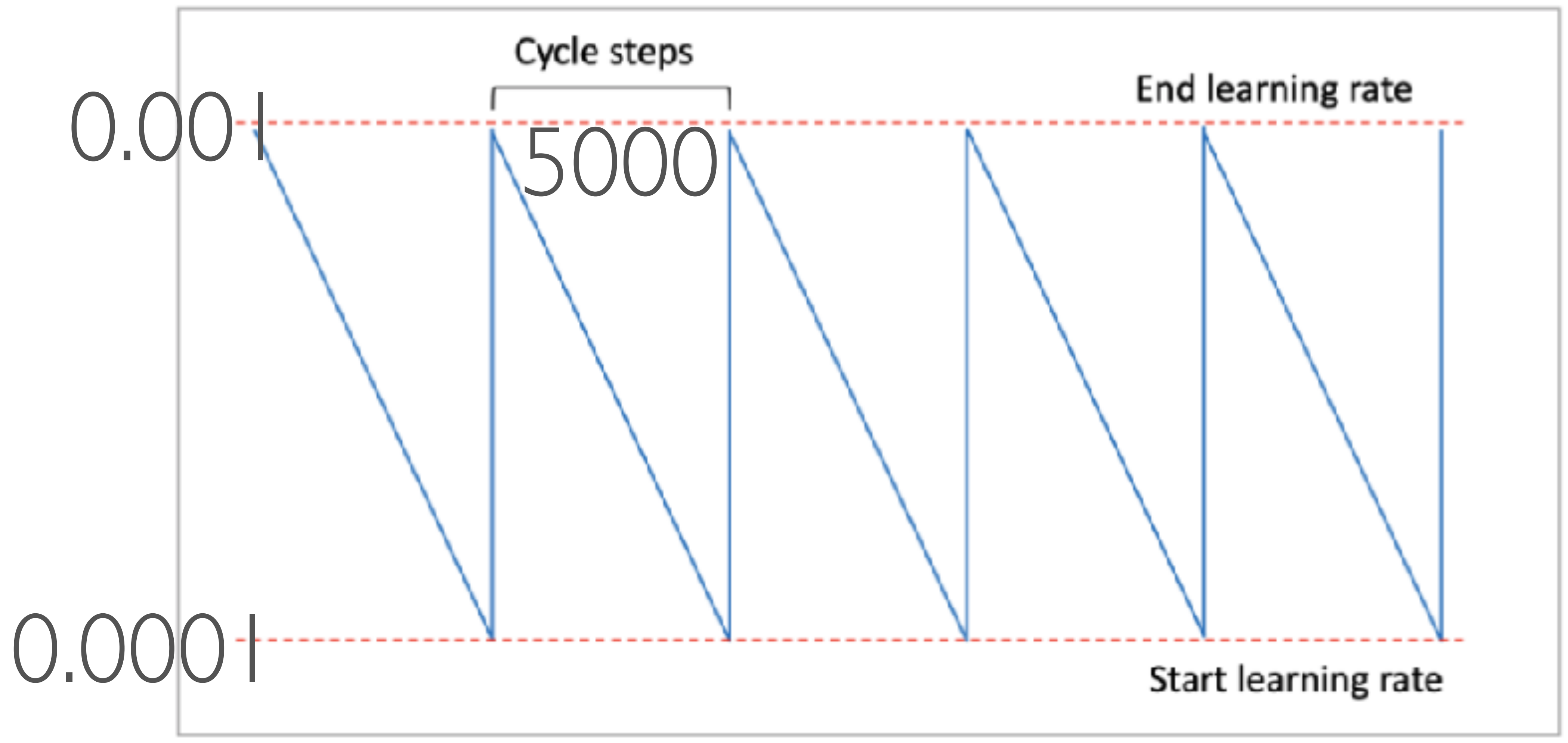
$$\hat{p}_{data}(\vec{x}|\vec{c}_i) = \mathcal{T}(\vec{x},\vec{x}_0|\,\vec{c}_i) \otimes p_{MC}(\vec{x}_0|\vec{c}_i)$$

ABCDnn - Details

- Uses tensorflow/keras
- Loss function: maximum mean discrepancy (MMD)
- Optimizer: Adam
  - Learning rate: sawtooth scheduler
  - Exponential decay term 1 (β_1) = 0.9
  - Exponential decay term 2 (β_2) = 0.999
  - Epsilon = 1e-5
- Size of minibatches: = 1024
- # of nodes of hidden layers = 64
- Depth of neural network = 3
- Dimension of output given to conditioner = 30

- C = autoregressive "conditioner" that outputs weights and biases, incorporating information from previous variables
- x = input variables
- y = output variables
- Sigmoid = activation function between layers





Sawtooth learning rate scheduling

Data-Prediction Discrepancy Uncertainty Definition

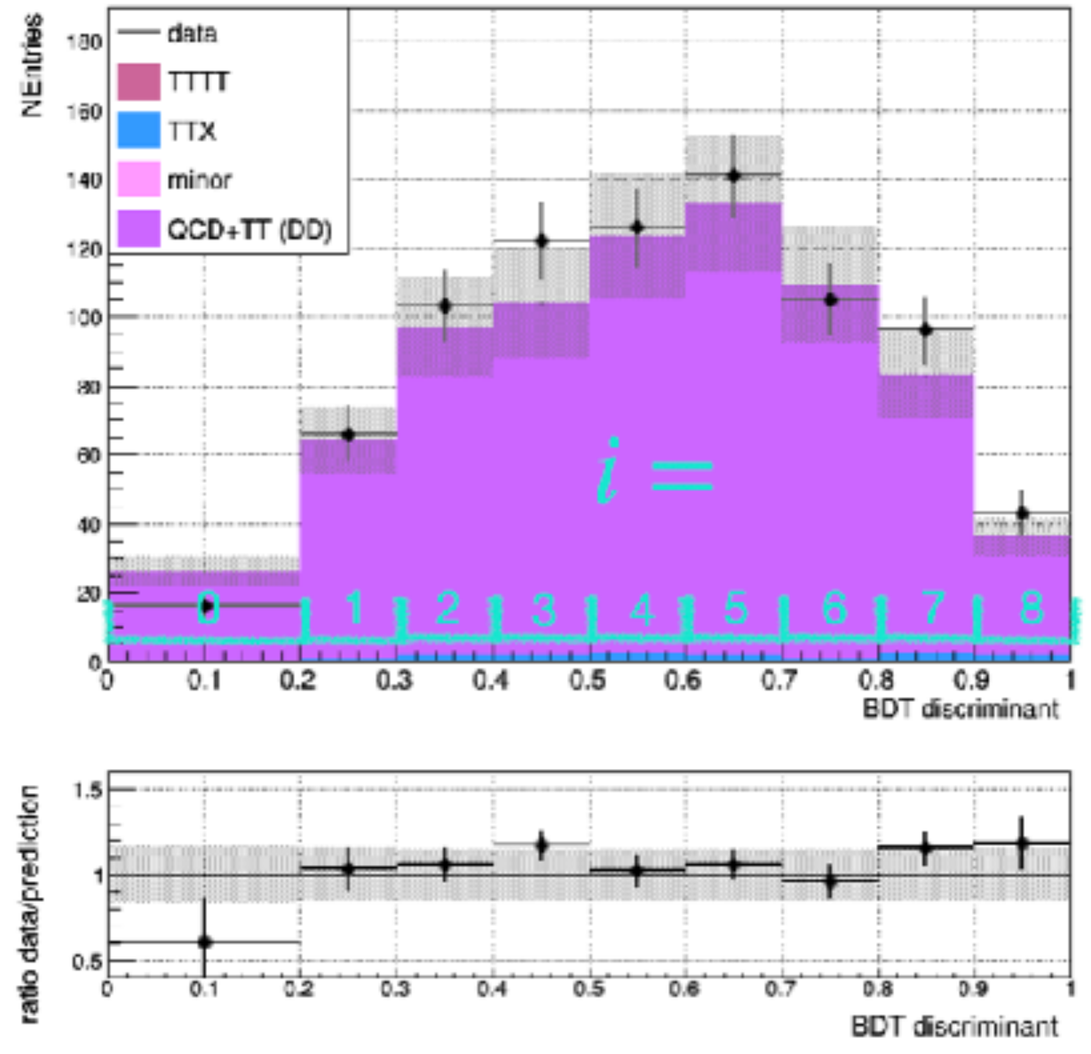
$$N_{DP} = \sqrt{\underbrace{(1 - \langle f \rangle)^2}_{} + \underbrace{\langle RMS \rangle^2}_{}}$$

Deviation from 1 of weighted mean of data/prediction ratios across BDT bins
~**offset** in normalization

Weighted RMS of data/prediction ratios across BDT bins
~**spread** of disagreement

$$\langle f \rangle = \frac{\sum_i f_i w_i}{\sum_i w_i} \quad f_i = \frac{(N_{data})_i}{(N_{pred})_i} \quad w_i = (N_{pred})_i \quad \langle RMS \rangle = \sqrt{\langle f^2 \rangle - \langle f \rangle^2}$$

- Statistics Committee Recommendation
- Uncertainty to account for potential non-closure of the method, as observed in VR bins
- Derived in VR bins and applied to SR bins
- Weighted per BDT discriminant histogram bin "i" - reflects distribution of histogram events



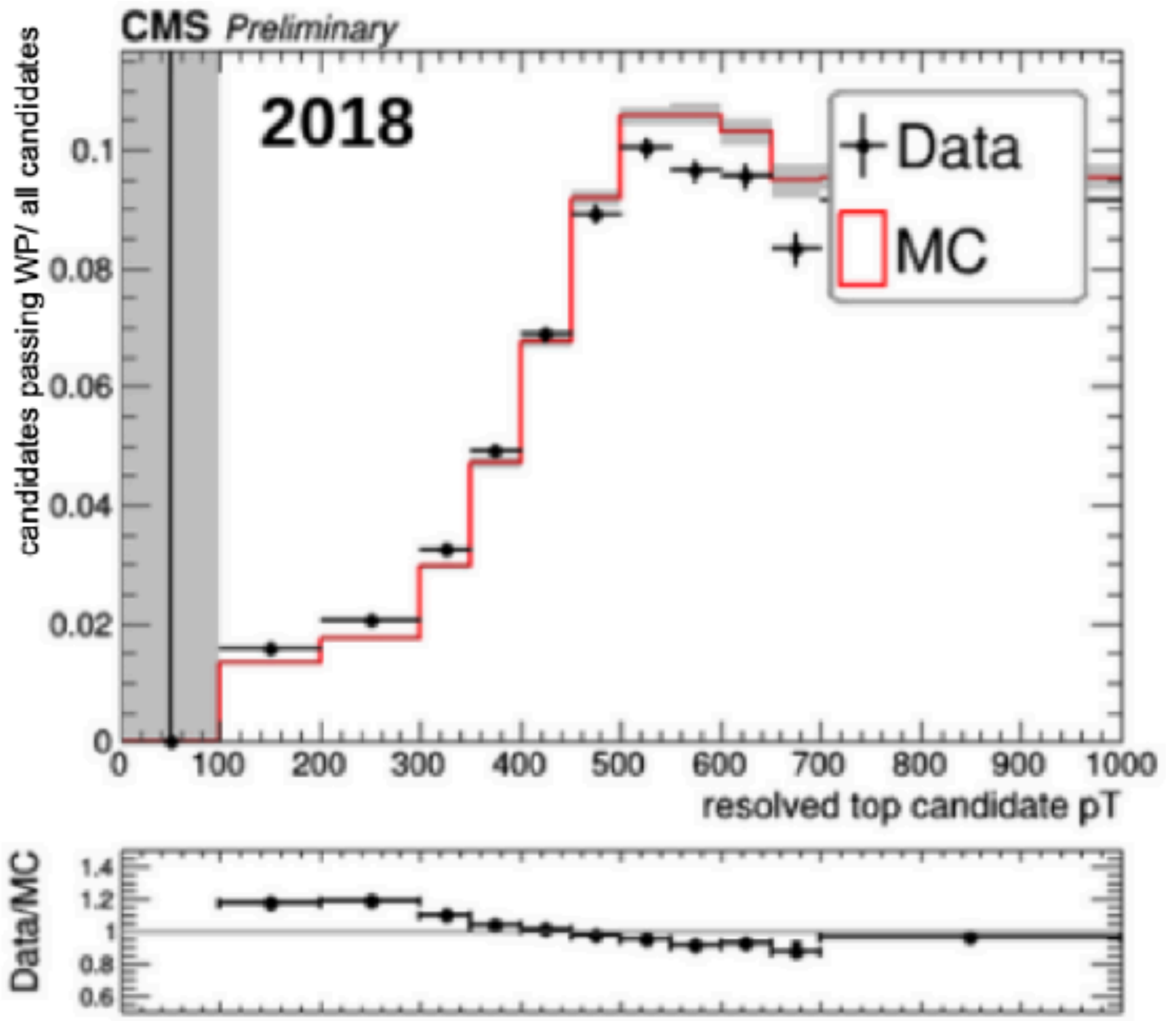
Resolved Top Tagger Scale Factors

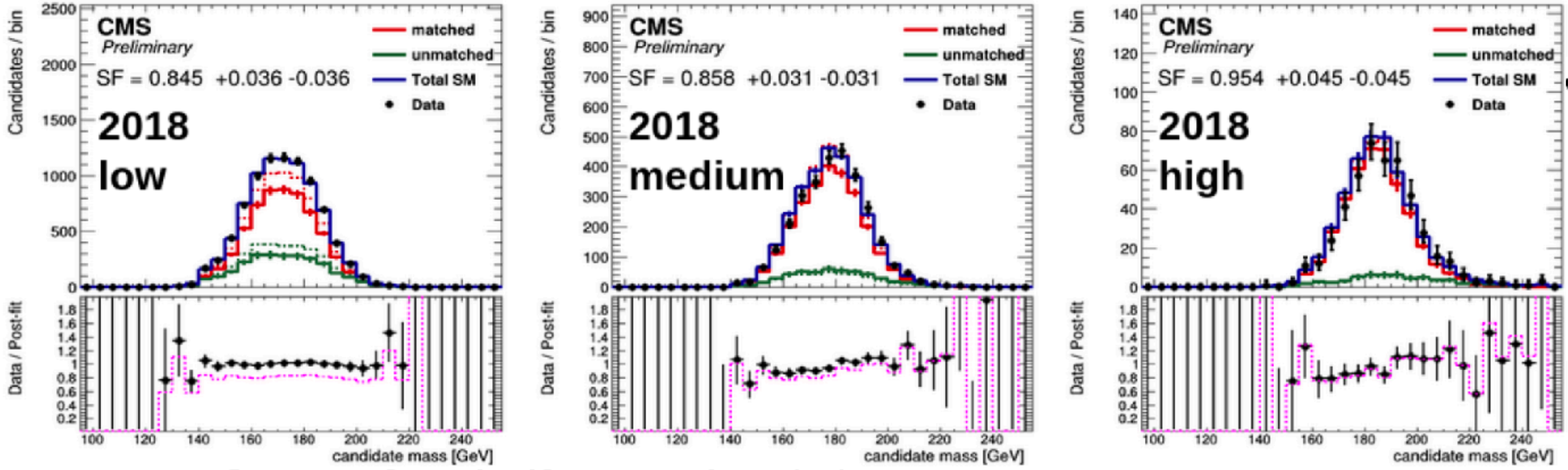
- Mistag Rate SFs:
  - Top candidates in 0-lepton, 1 bjet control region targeting QCD
  - Calculate SFs as function of pT
  - Gen-tagged tops removed
- Efficiency SFs:
  - Top candidates in single-muon control region targeting 1-lepton ttbar
  - Calculate SFs in low (100-300 GeV), medium(300-500 GeV), & high (> 500GeV) pT ranges
  - Mistag rate applied to non-gen tops
  - SFs extracted from template fit of top mass distribution to data

$$\varepsilon = \frac{cands\ passing\ WP}{all\ cands}$$

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

Mistag rates measured in data and simulation





Data vs pre- and post-fit distributions in top candidate mass for top candidates passing the WP

Analysis Objects

Jets
PF anti-kT ΔR=0.4 jets
pT ≥ 35 GeV
|η| ≤ 2.4
loose (tight) PF jetId 2016(2017/2018)
standard JECs + JER smearing

bJets
DeepJet medium WP
pass jet pre-selection

p_T^{miss}
type-1 corrections

Veto Electrons
pT ≥ 15 GeV
|η| < 2.5
mvaFall17V2noIso “loose”
miniPFRelIso_all ≤ 0.4

Veto Muons
pT ≥ 15 GeV
|η| < 2.5
miniIsoId “loose”
cutBasedId “loose”

Boosted Tops
PF anti-kT ΔR=0.8 jets
pT ≥ 400 GeV
|η| ≤ 2.4
separated from resolved tops by ΔR>0.8
DeepAK8 medium WP

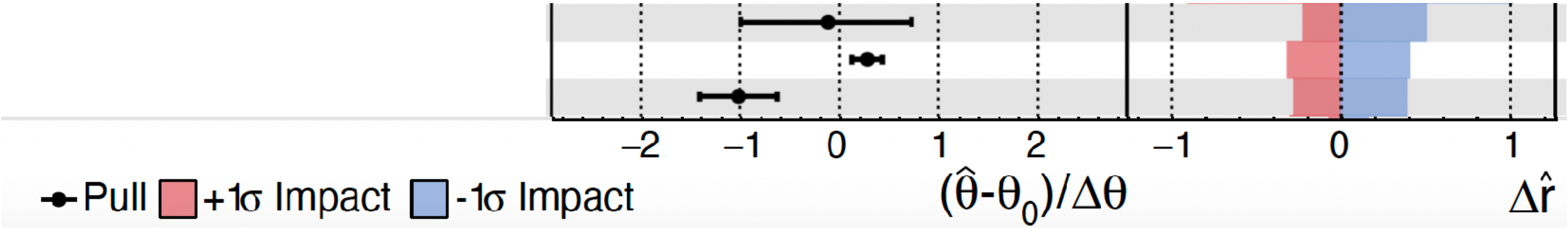
Boosted Ws
PF anti-kT ΔR=0.8 jets
pT ≥ 200 GeV
|η| ≤ 2.4
separated from resolved tops by ΔR>0.8
DeepAK8 medium WP

Resolved Tops
Custom NanoAOD tagger
pT ≥ 100 GeV
|η| ≤ 2.4
assembled from 3-jet candidates (passing jet pre-selection)
WP ~ 10% mistag rate in QCD

(Standard vertex selections also applied)

Statistics

- To summarize, a binned maximum likelihood fit is performed simultaneously in signal region (SR) categories and using systematic uncertainties as nuisance parameters in order to extract a signal strength based on a profile likelihood method.
- **Profile likelihood:** differentiates between the signal vs. background-only hypotheses in a way independent of nuisances given a sufficiently large sample size.
- **Limits** express the compatibility of a statistical model with the signal strength
- **Significance:** Qualitatively, this is the statistical fluctuation in a background-only probability distribution required to explain some observation
- **Pulls and impacts:** Indicate if uncertainties are moved from their pre-fit values or constrained by the data. large pulls or constrains can sometimes indicate a problem with the fit



NN vs. BDT

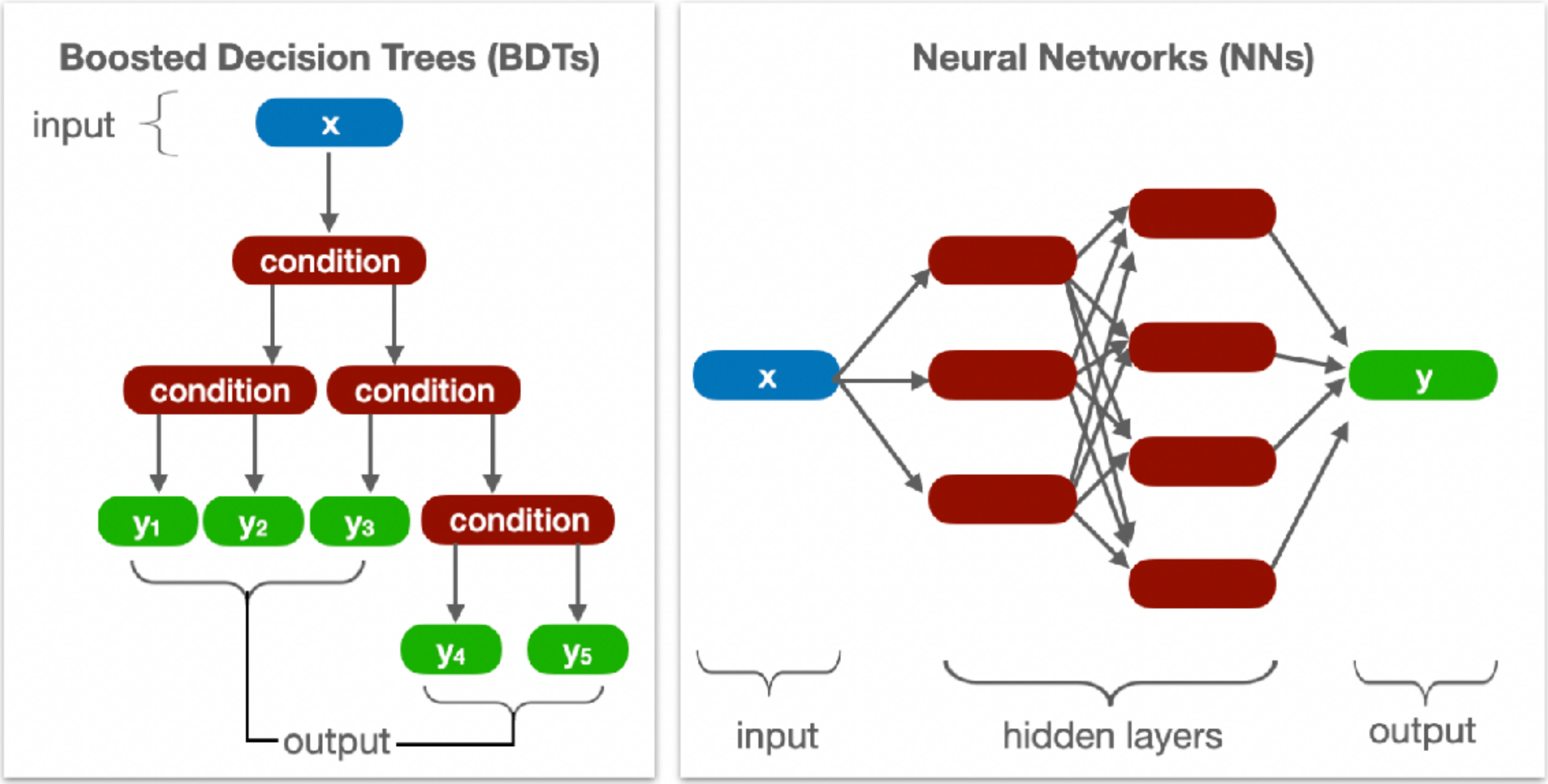


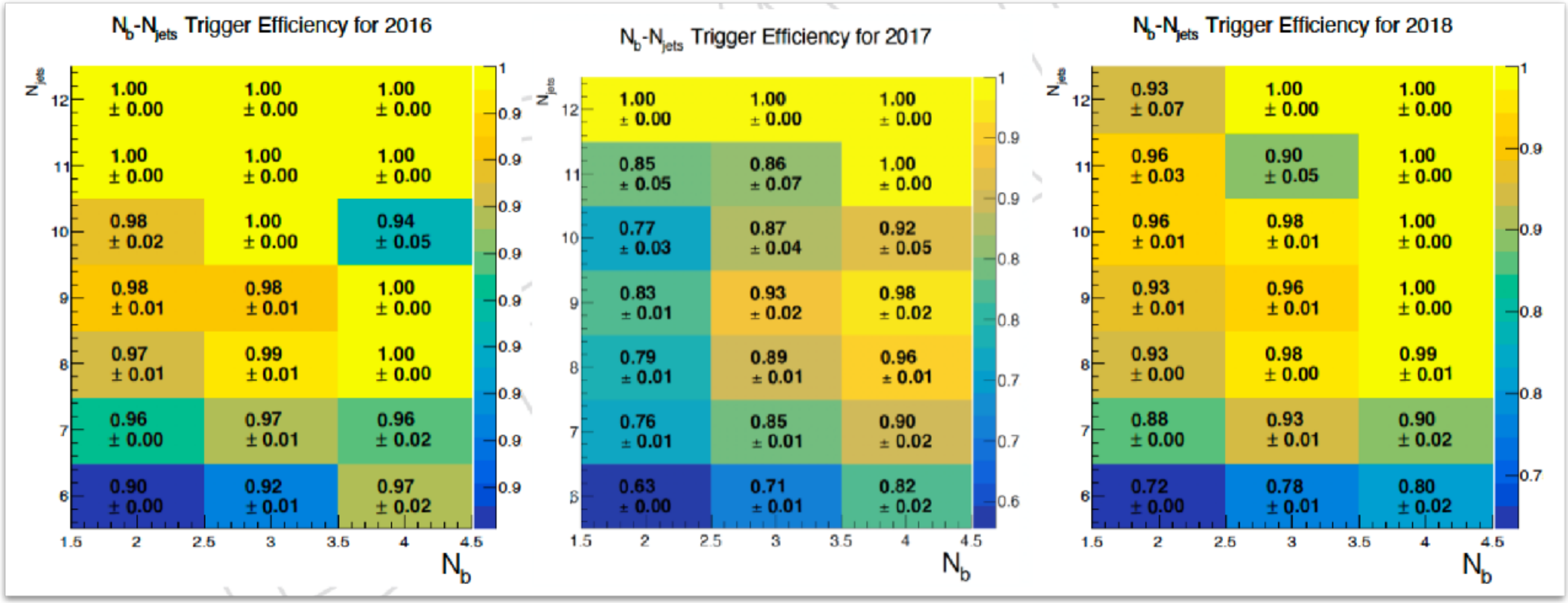
Figure 5.1: Schematic comparing the structure of a BDT (left) vs. a NN (right) given inputs x and outputs y. Note that these are just example structures and that x and y can be vectors of multiple inputs and outputs.

Triggers

$$\text{Trigger Efficiency} = \frac{\text{\# events passing OR of triggers and muon triggers}}{\text{\# events passing muon triggers}}$$

Muon triggers: HLT_IsoMu24 or HLT_IsoMu27

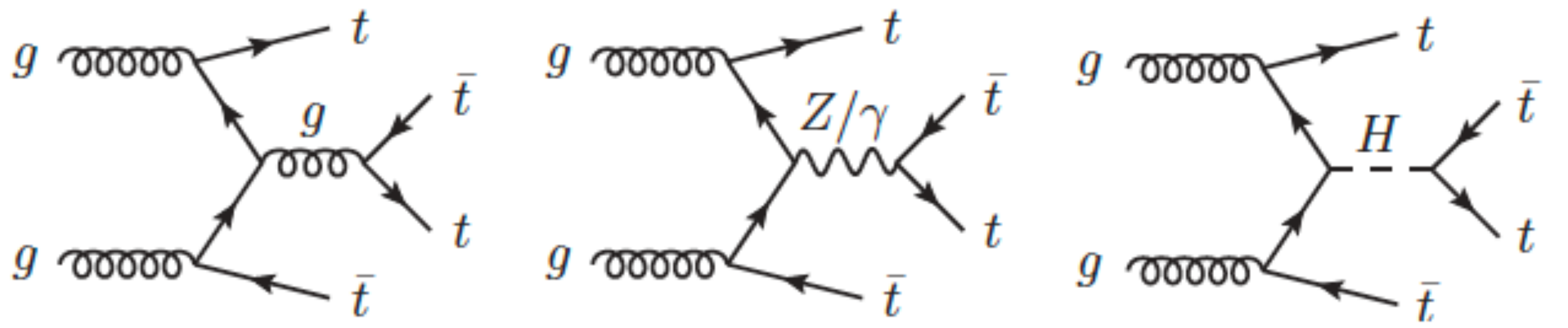
Data-taking period	HLT path
2016	HLT_PFHT400_SixJet30_DoubleBTagCSV_p056
	HLT_PFHT450_SixJet40_BTagCSV_p056
2017	HLT_PFHT380_SixJet32_DoubleBTagCSV_p075
	HLT_PFHT430_SixJet40_BTagCSV_p080
	HLT_PFHT380_SixPFJet32_DoublePFBTagCSV_2p2
	HLT_PFHT430_SixPFJet40_PFBTagCSV_1p5
2018	HLT_PFHT380_SixPFJet32_DoublePFBTagDeepCSV_2p2
	HLT_PFHT430_SixPFJet40_PFBTagDeepCSV_1p5
	HLT_PFHT400_SixPFJet32_DoublePFBTagDeepCSV_2p94
	HLT_PFHT450_SixPFJet36_PFBTagDeepCSV_1p59



What's Next?

*Cao, Chen, Liu. arXiv:1602.01934v1
+CMS Collaboration. CMS-TOP-18-003. arXiv:1908.06463

- Further refine and apply the machine-learning based analysis tools introduced here:
  - NN-based data-driven background estimation
  - Accessible resolved top tagger
- Four-tops will be interesting to us in the future!
  - Good probe of the top-Higgs Yukawa coupling*
    - Independent of the Higgs decay width*



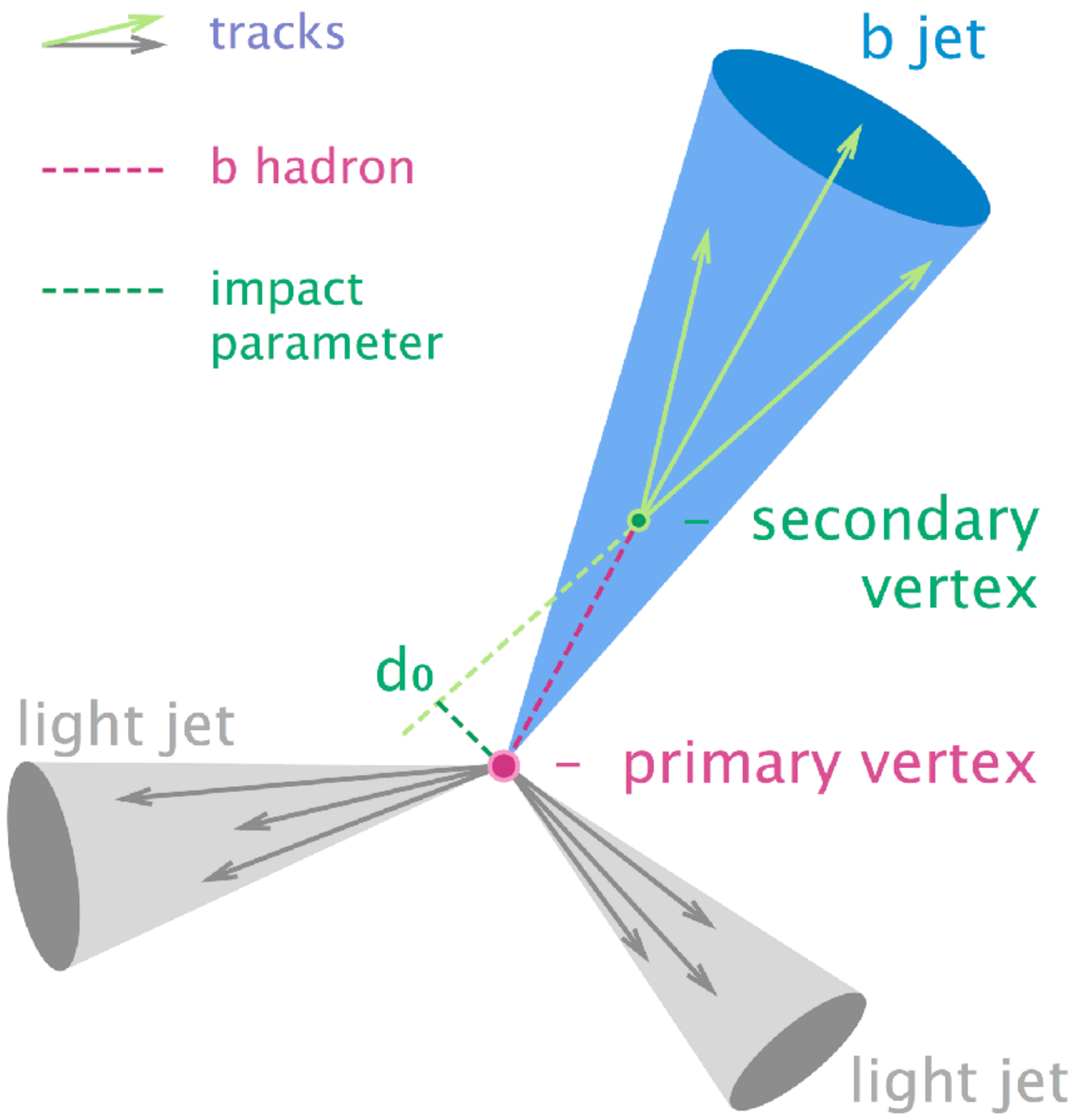
$$\sigma(t\bar{t}t\bar{t}) = \sigma^{\rm SM}(t\bar{t}t\bar{t})_{g+Z/\gamma} + \kappa_t^2\sigma^{\rm SM}_{\rm int} + \boxed{\kappa_t^4\sigma^{\rm SM}(t\bar{t}t\bar{t})_H}$$

We expect κ_t=1 $\kappa_t = y_{t(obs)}/y_{t(SM)}$

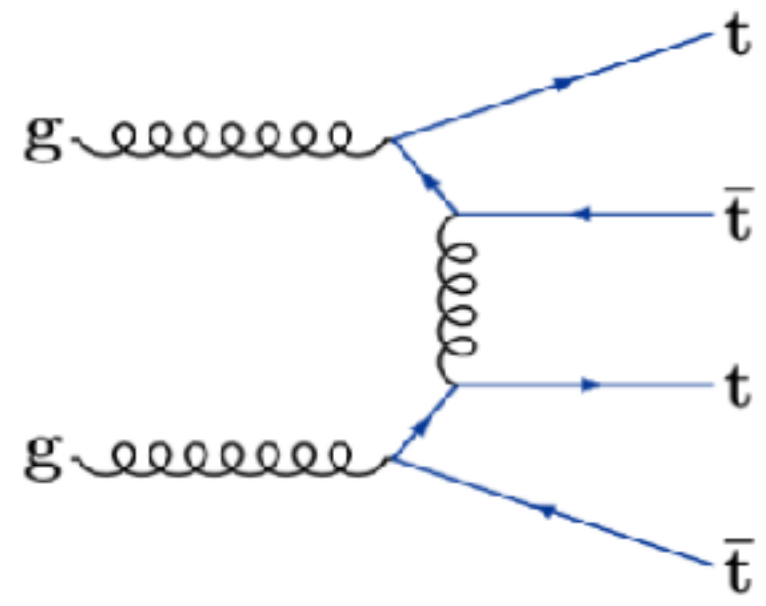
Results from CMS Same-Sign 2-Lepton + Multi-Lepton (2LSS/3L) published Run II analysis:+
$\kappa_t < 1.7 \pm 0.3$ at 95% confidence level

(y_t is top-Higgs Yukawa coupling)

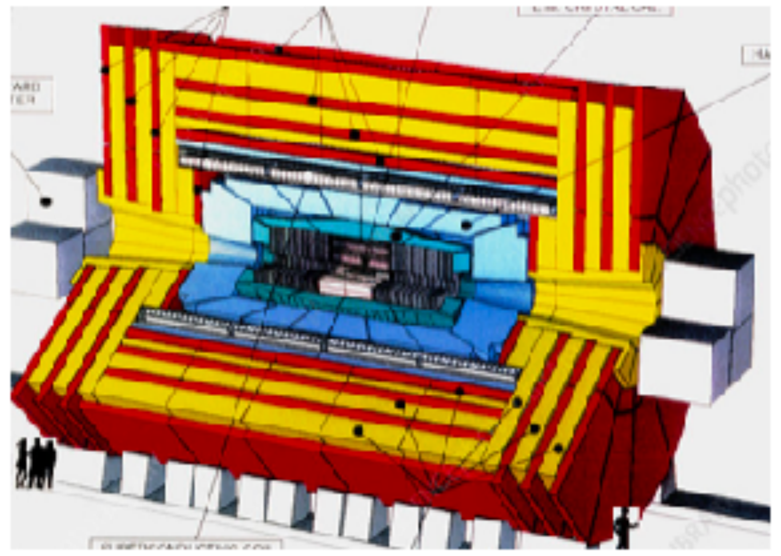
B-Tagging


tracks
b hadron
impact parameter
b jet
secondary vertex
d_0
light jet
primary vertex
light jet

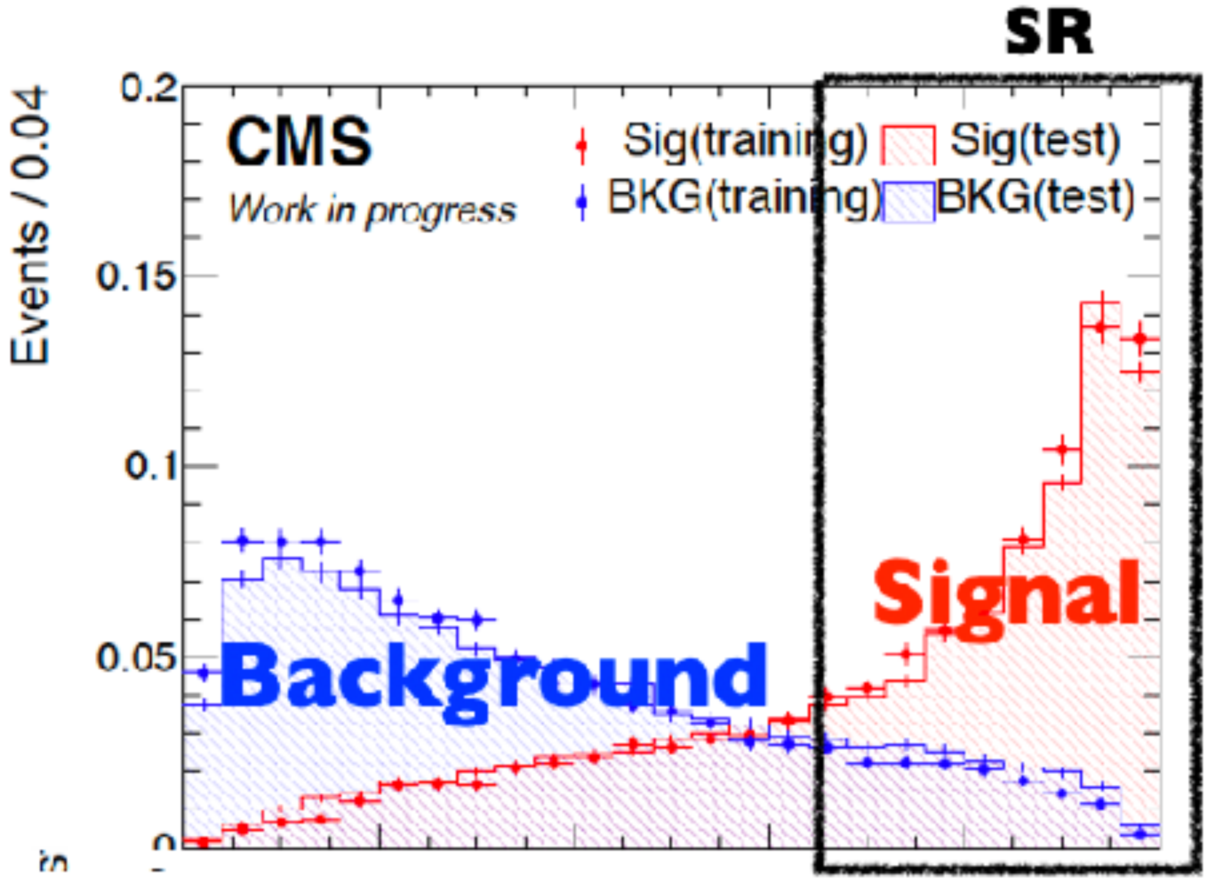
Analysis Strategy



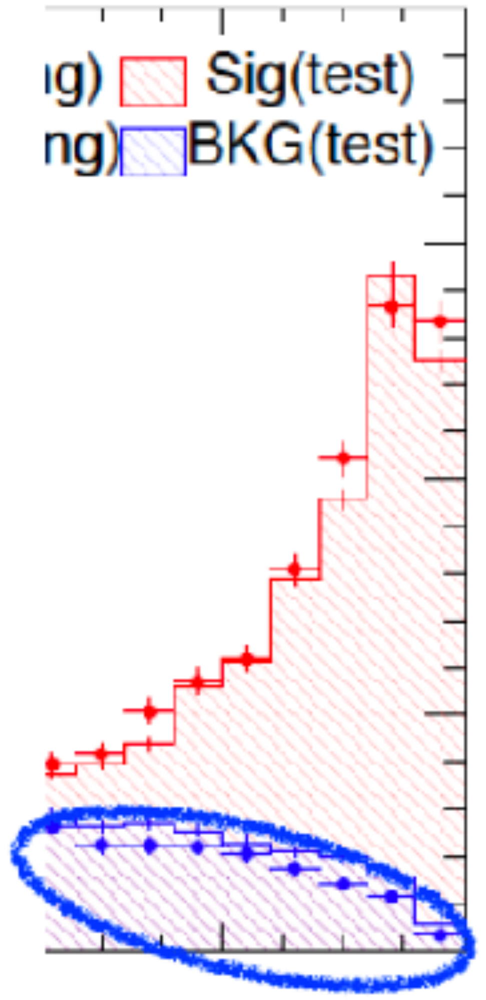
1) Select an interesting physics hypothesis to study (signal)



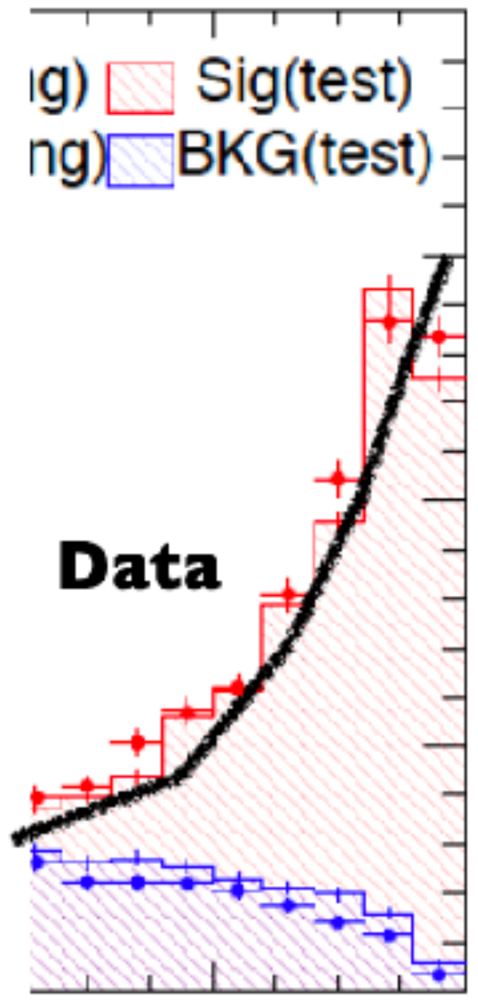
2) Collect data



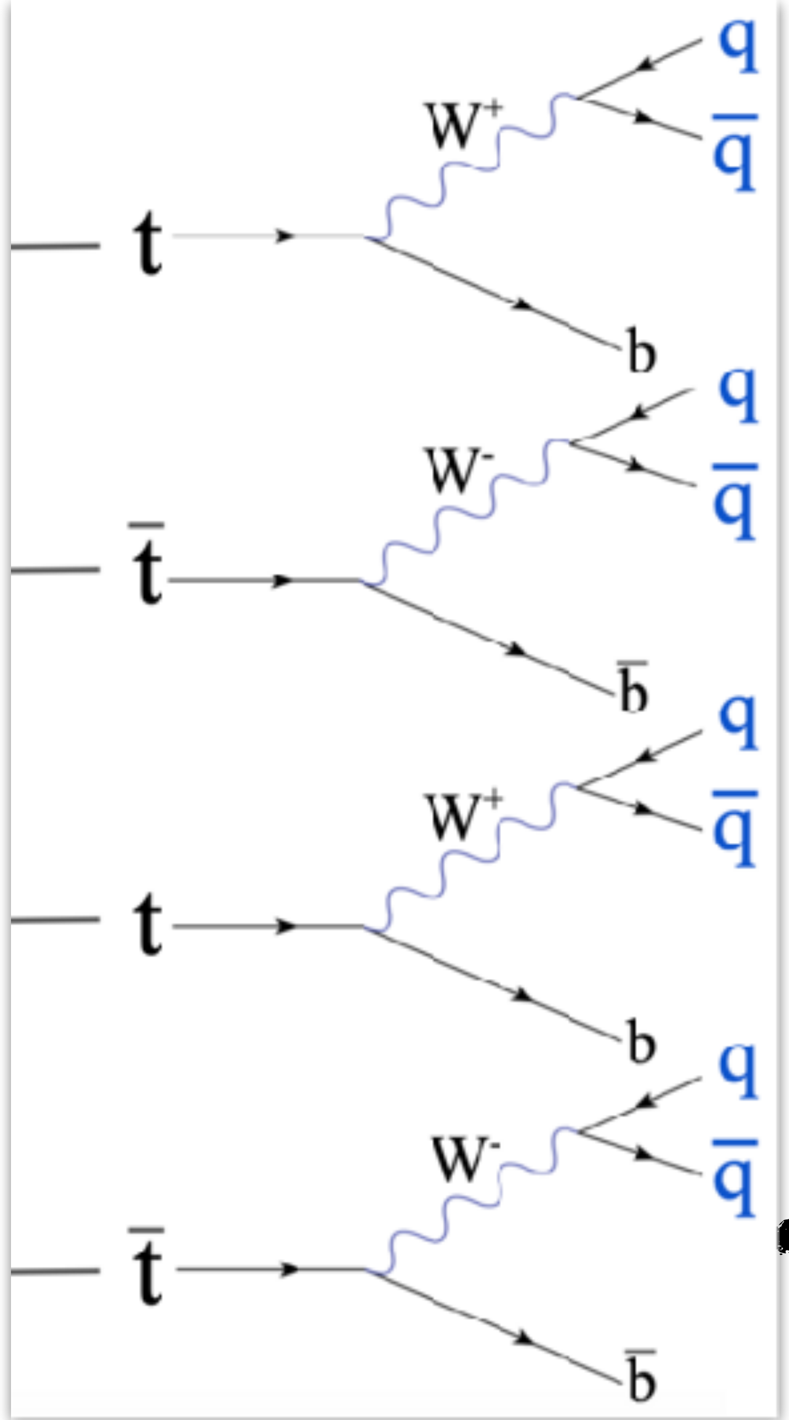
3) Optimize a signal-enriched region



4) Estimate background

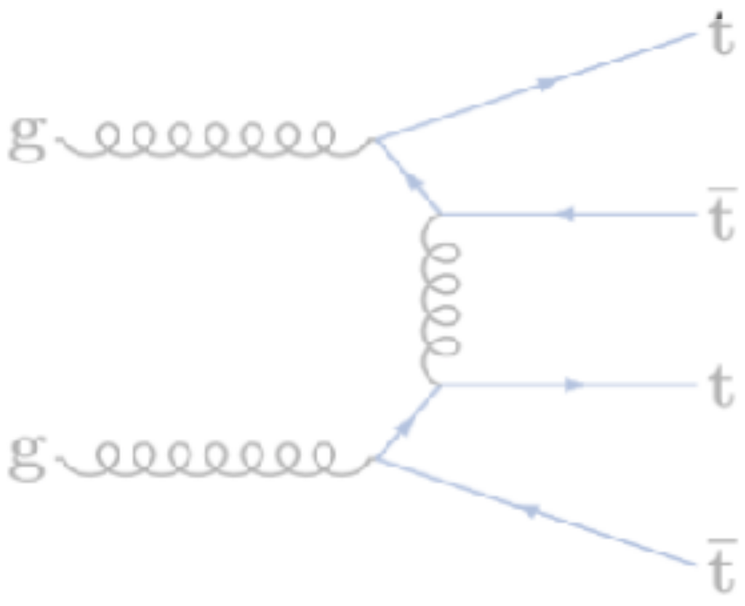


5) Compare expectations with data

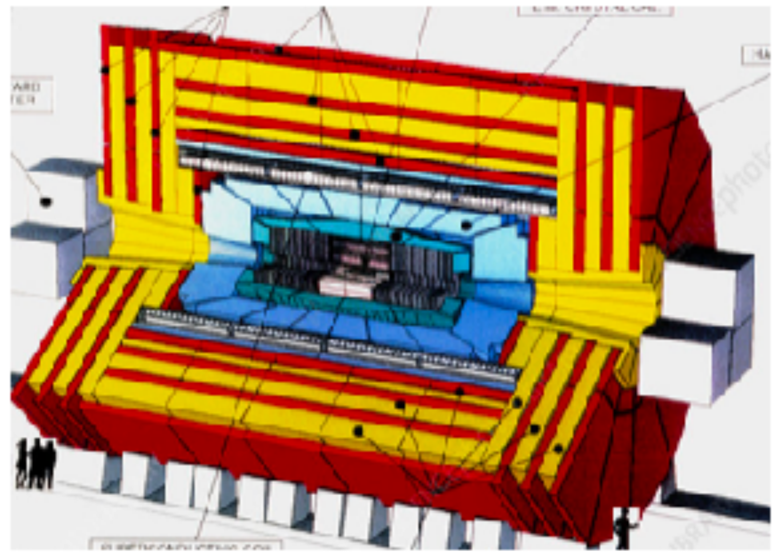


Signal:
All-hadronic $t\bar{t}t\bar{t}$

Analysis Strategy

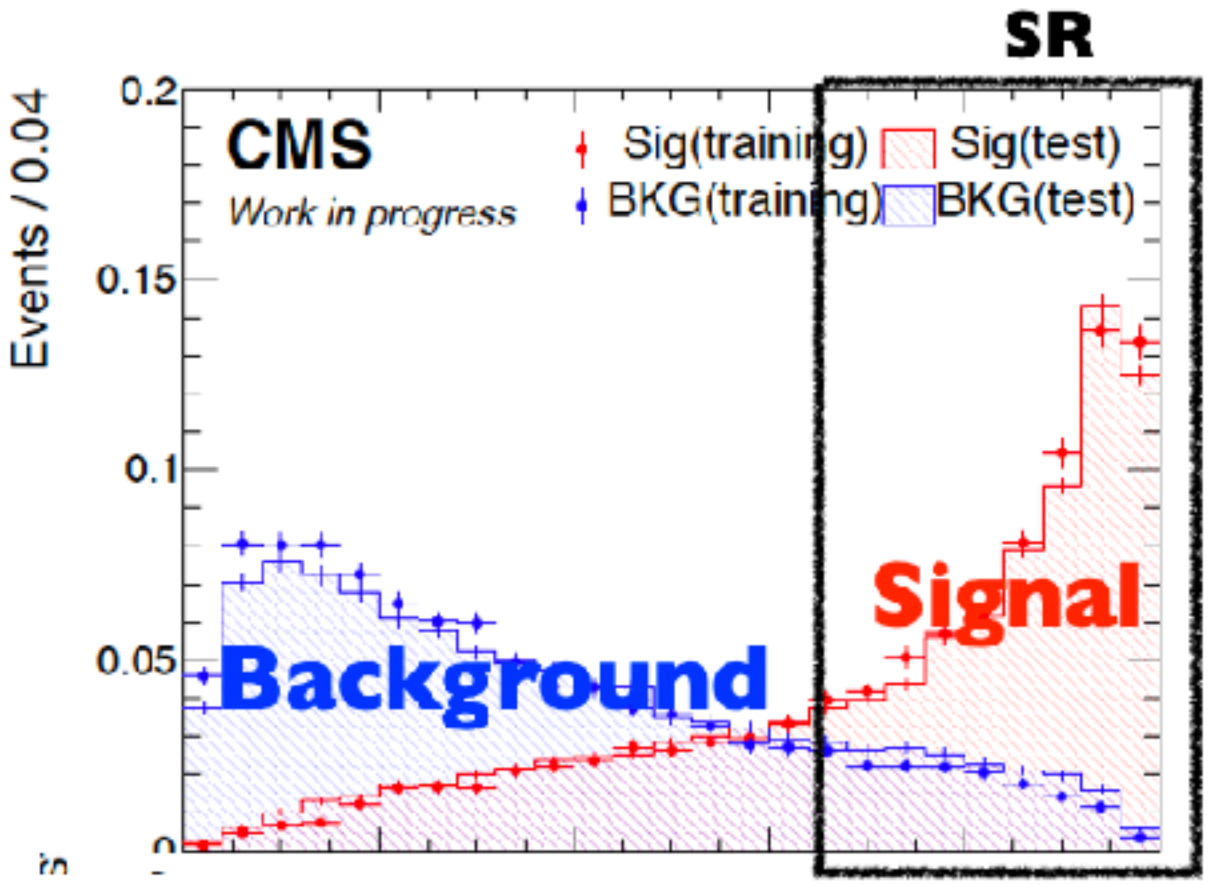


1) Select an interesting physics hypothesis to study (signal)

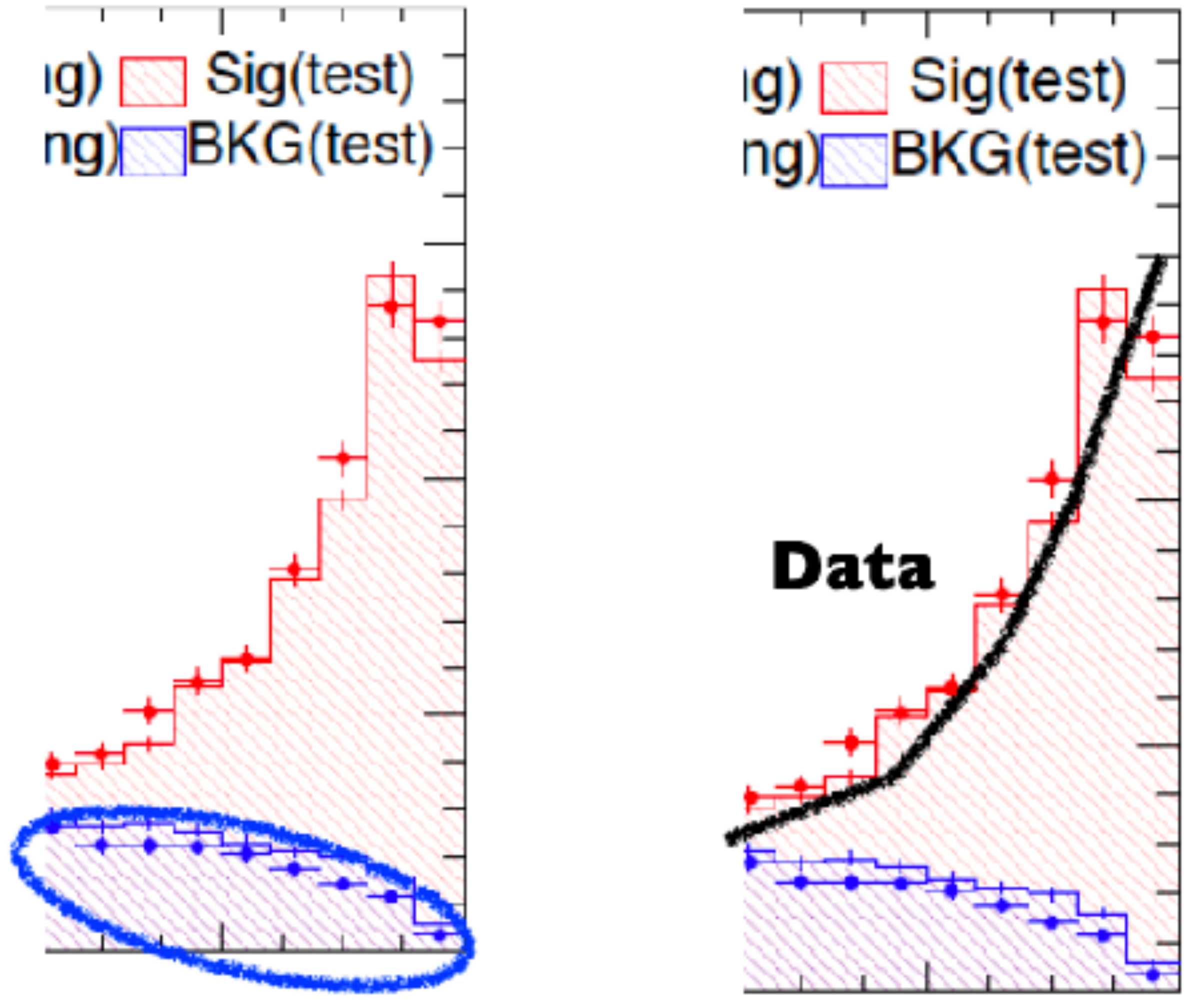


2) Collect data

Run II data from CMS



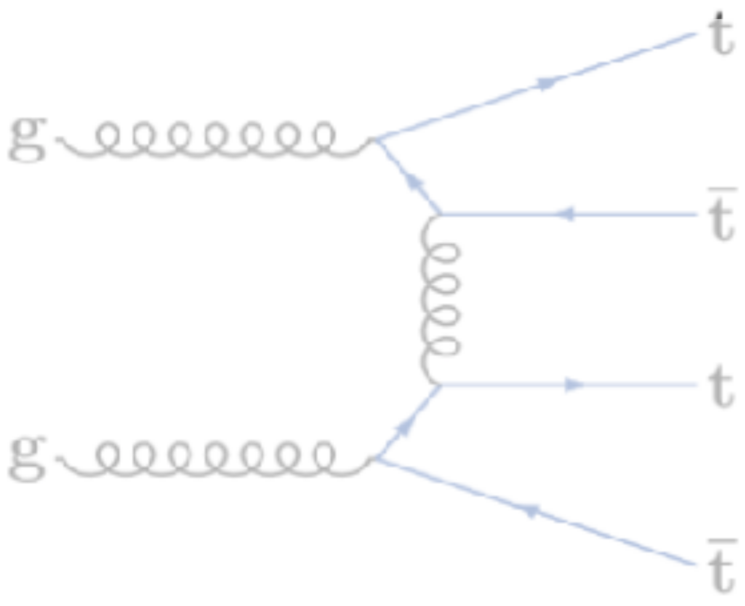
3) Optimize a signal-enriched region



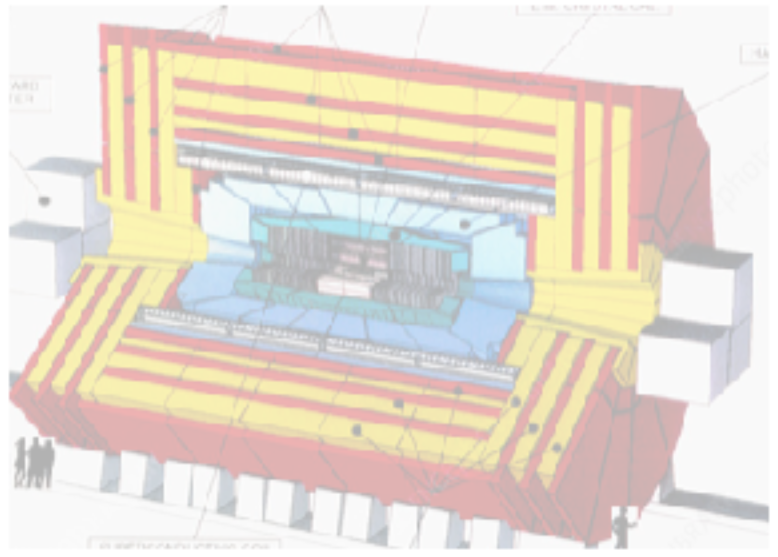
4) Estimate background

5) Compare expectations with data

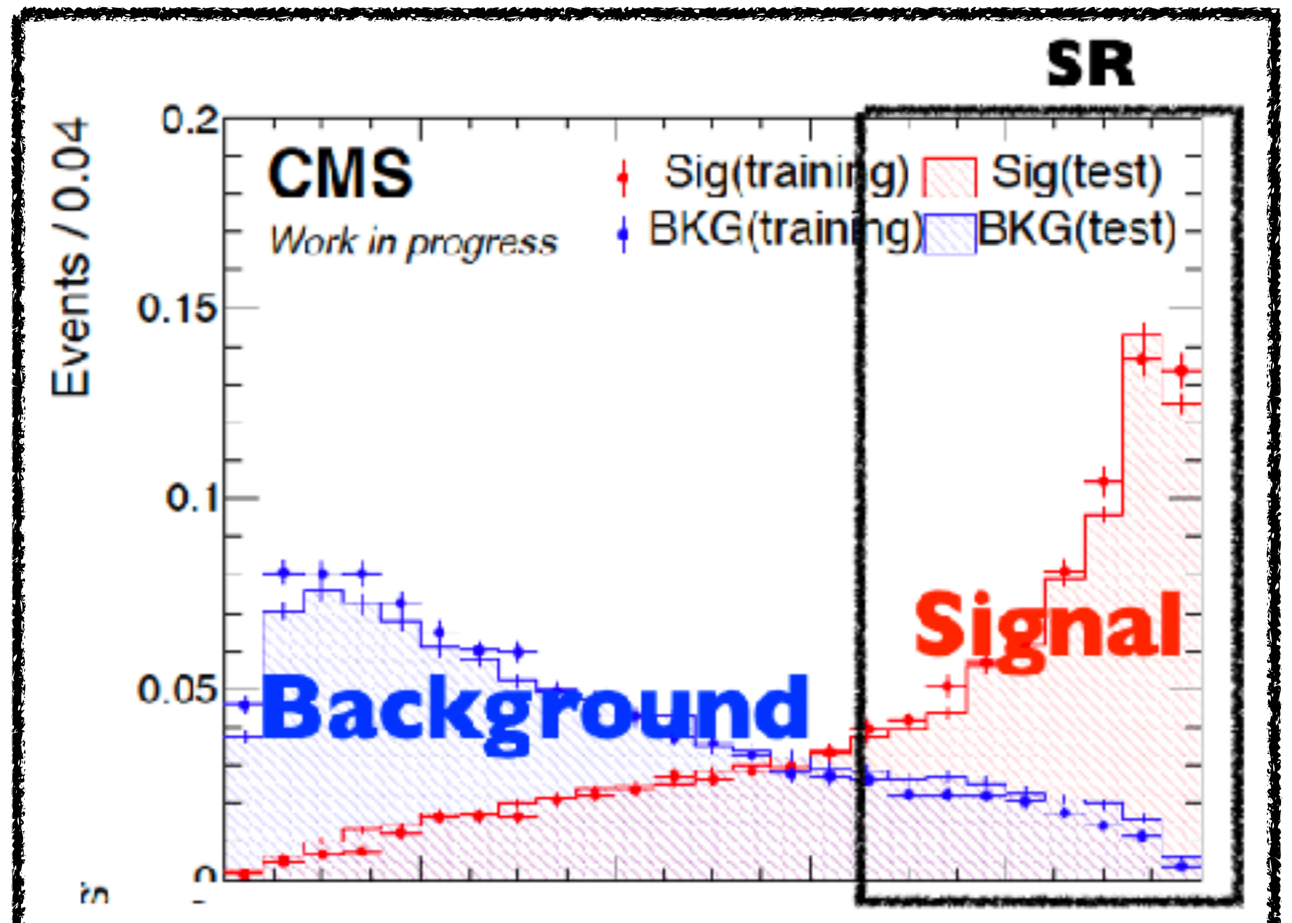
Analysis Strategy



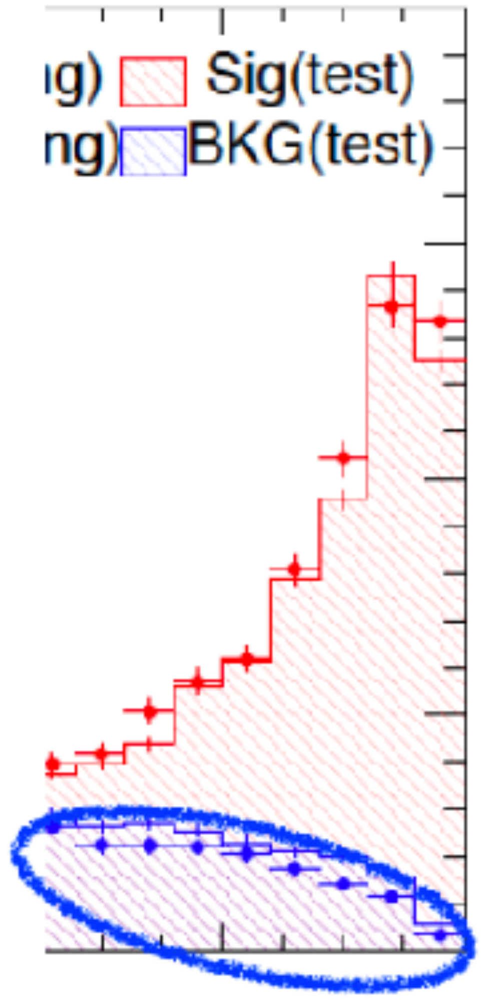
1) Select an interesting physics hypothesis to study (signal)



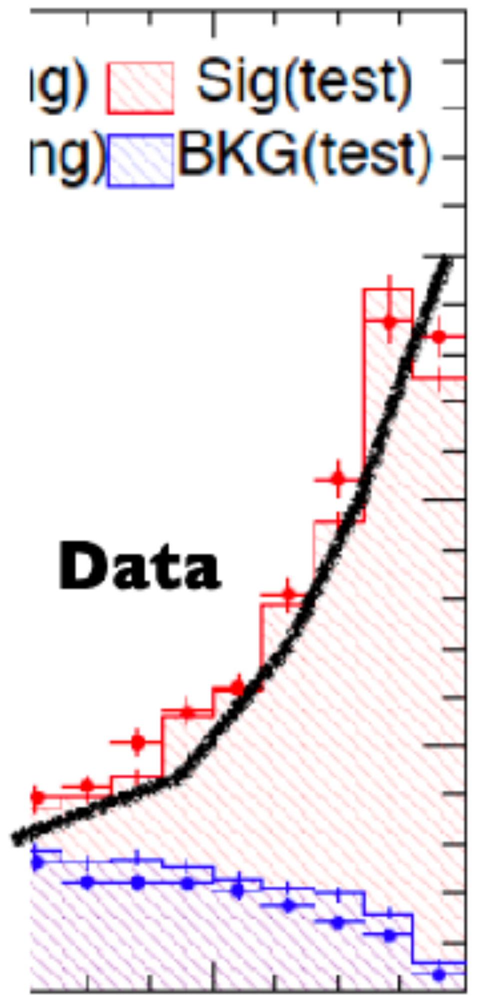
2) Collect data



3) Optimize a signal-enriched region



4) Estimate background

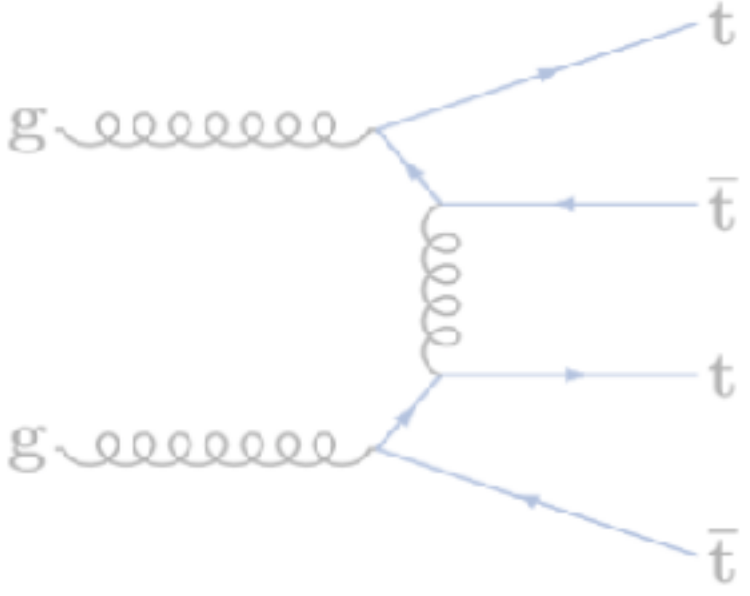


5) Compare expectations with data

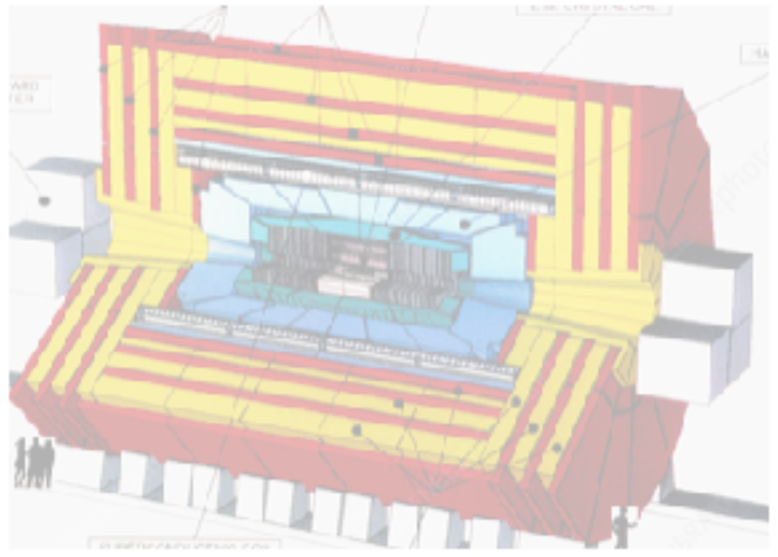
Identify signal region and discriminate vs. background

- Define baseline selection and signal region (SR) categories
- Identify hadronically decaying top quarks
- Optimize a variable that separates signal from background
  - Use a boosted decision tree (BDT)

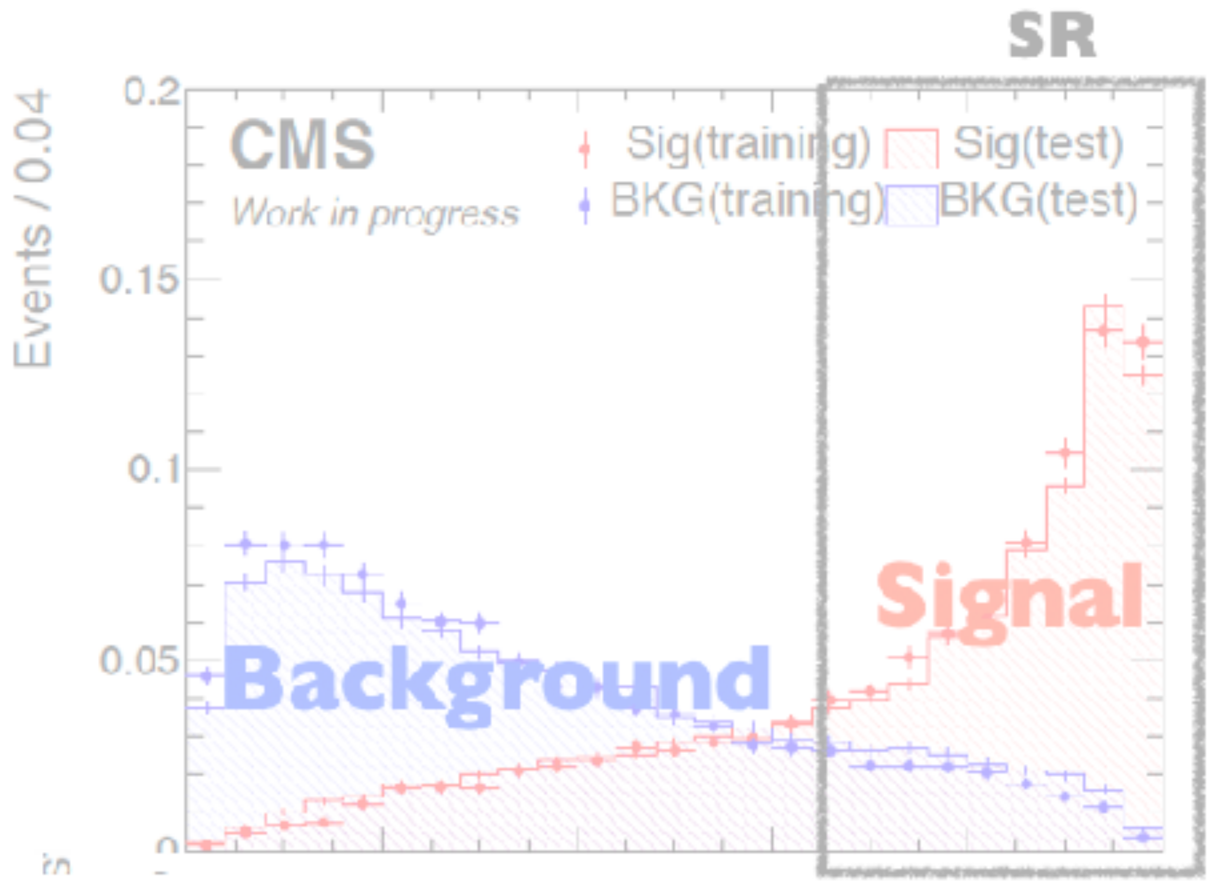
Analysis Strategy



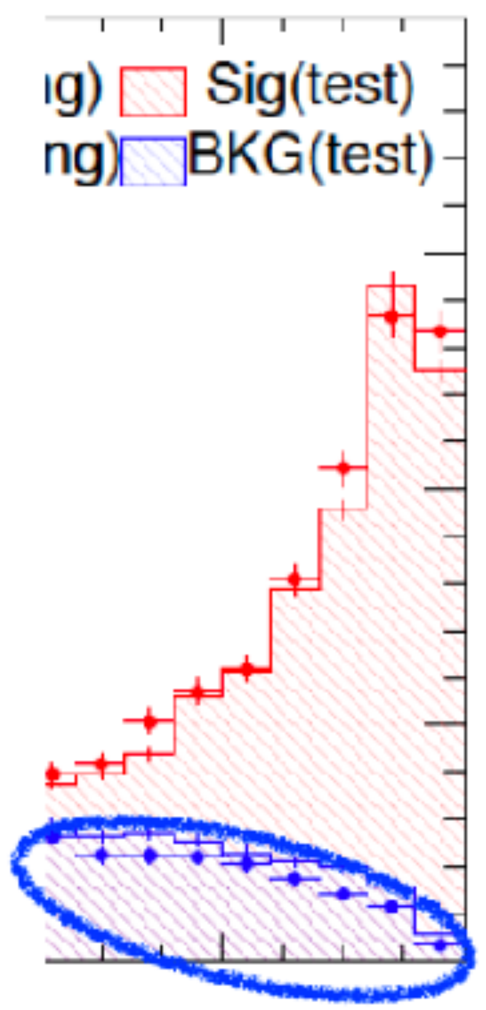
1) Select an interesting physics hypothesis to study (signal)



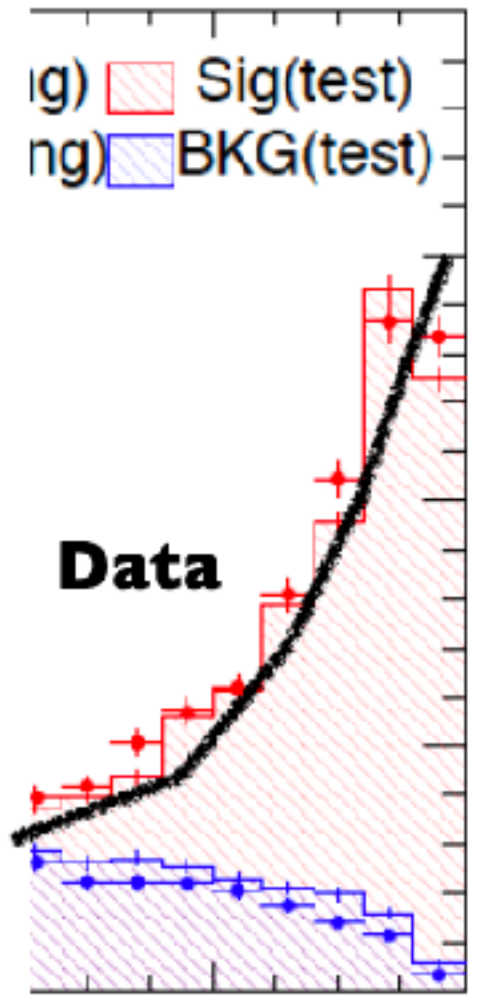
2) Collect data



3) Optimize a signal-enriched region



4) Estimate background



5) Compare expectations with data

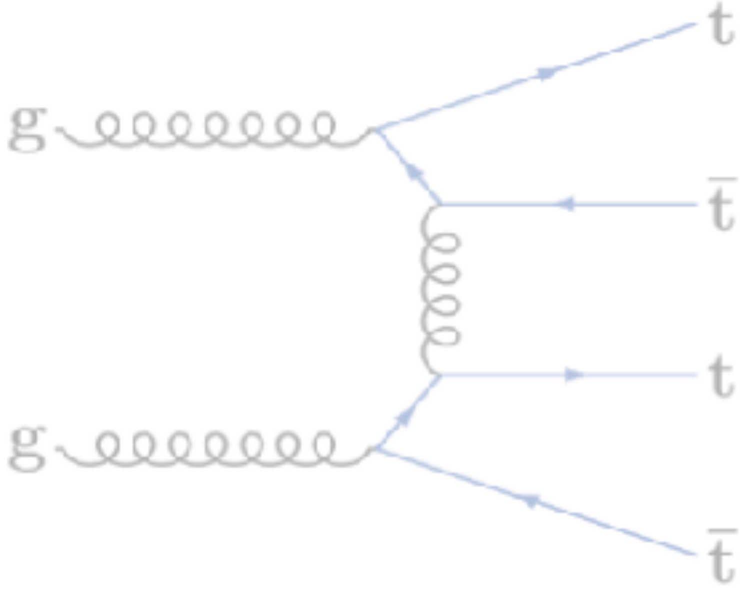
Estimate Backgrounds

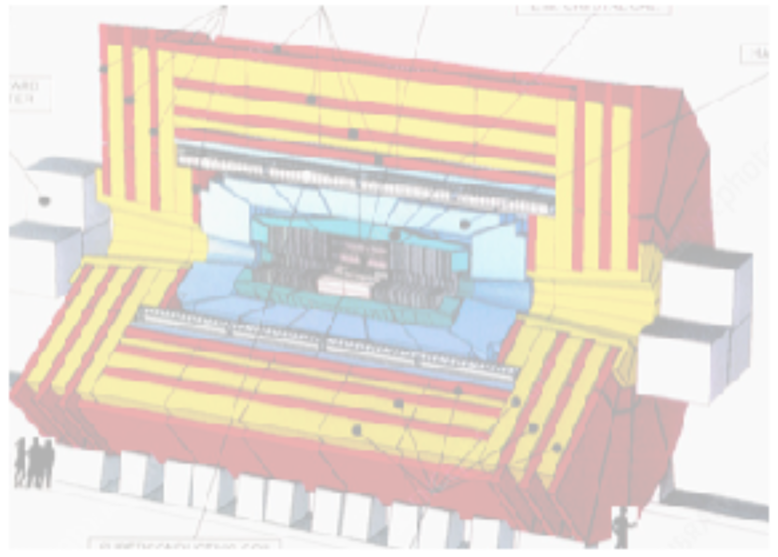
Use simulation for minor backgrounds ($t\bar{t}$H, $t\bar{t}$W/Z, single top…)

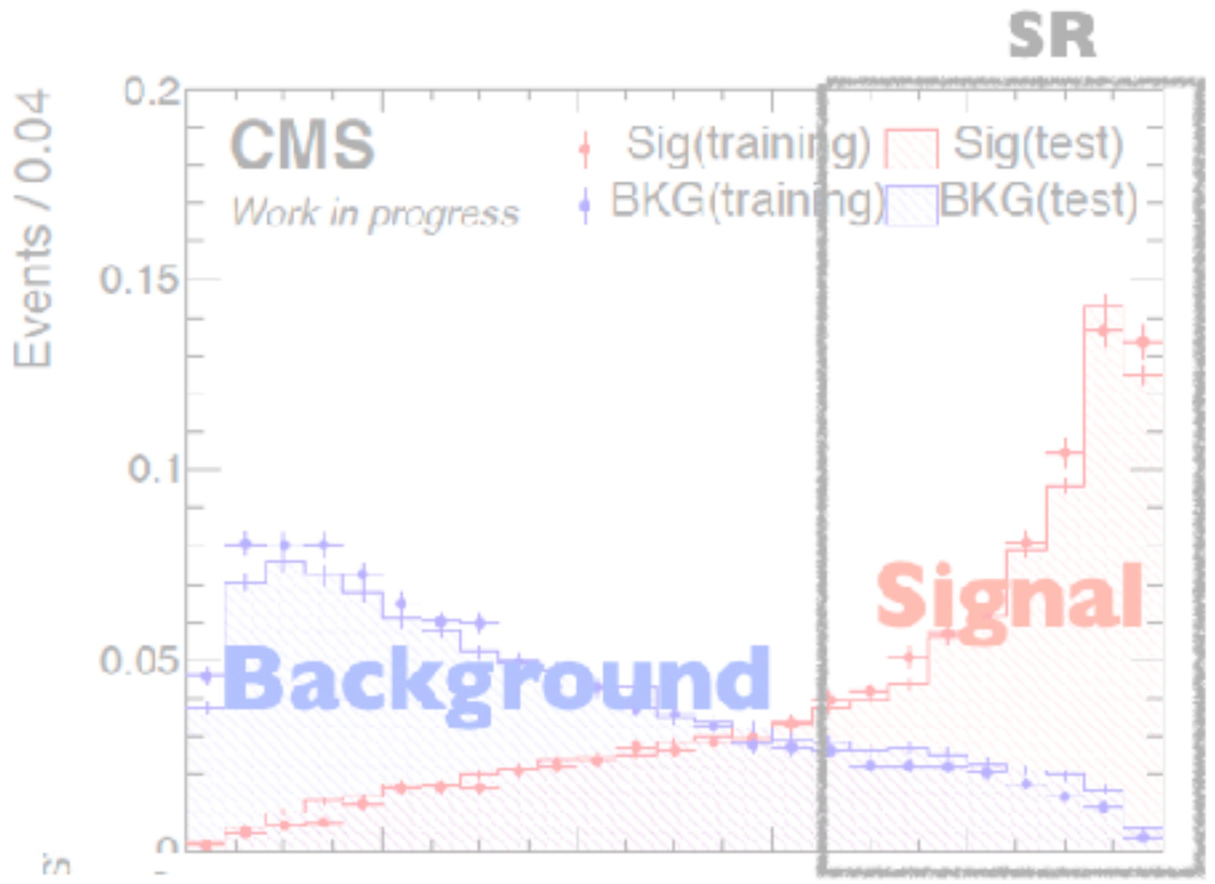
Major backgrounds (QCD + $t\bar{t}$):

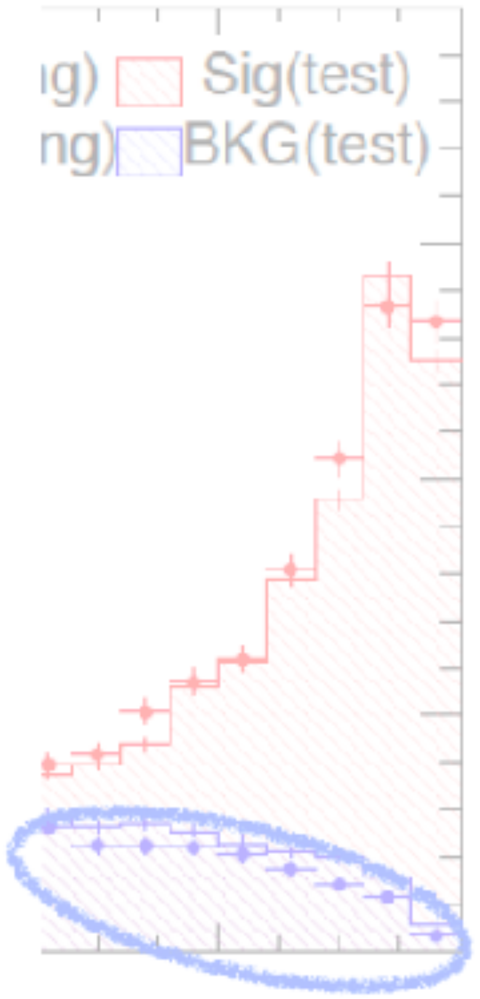
- Extrapolate from data in control regions (CRs) to predict backgrounds in SR
- Predict background yields using variation of "ABCD" method and background shapes using a neural net (NN)

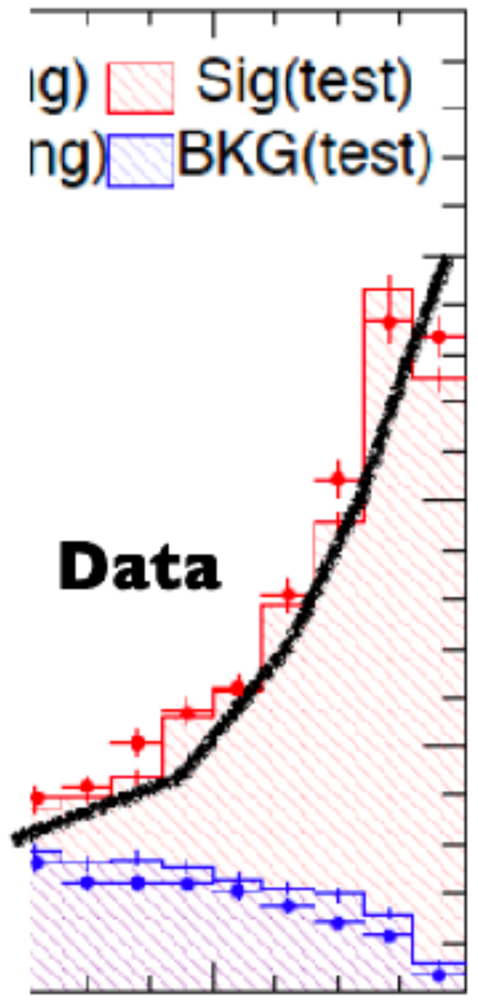
Analysis Strategy











1) Select an interesting physics hypothesis to study (signal)

2) Collect data

3) Optimize a signal-enriched region

4) Estimate background

5) Compare expectations with data

Compare results with standard model expectations!