EXPLORING BOOSTED TOP QUARK DECAYS USING RUN 3 DATA COLLECTED BY THE CMS EXPERIMENT

— **Johanna Matthiesen**[1], Johannes Haller[1], Roman Kogler[2], Daniel Savoiu[1] —

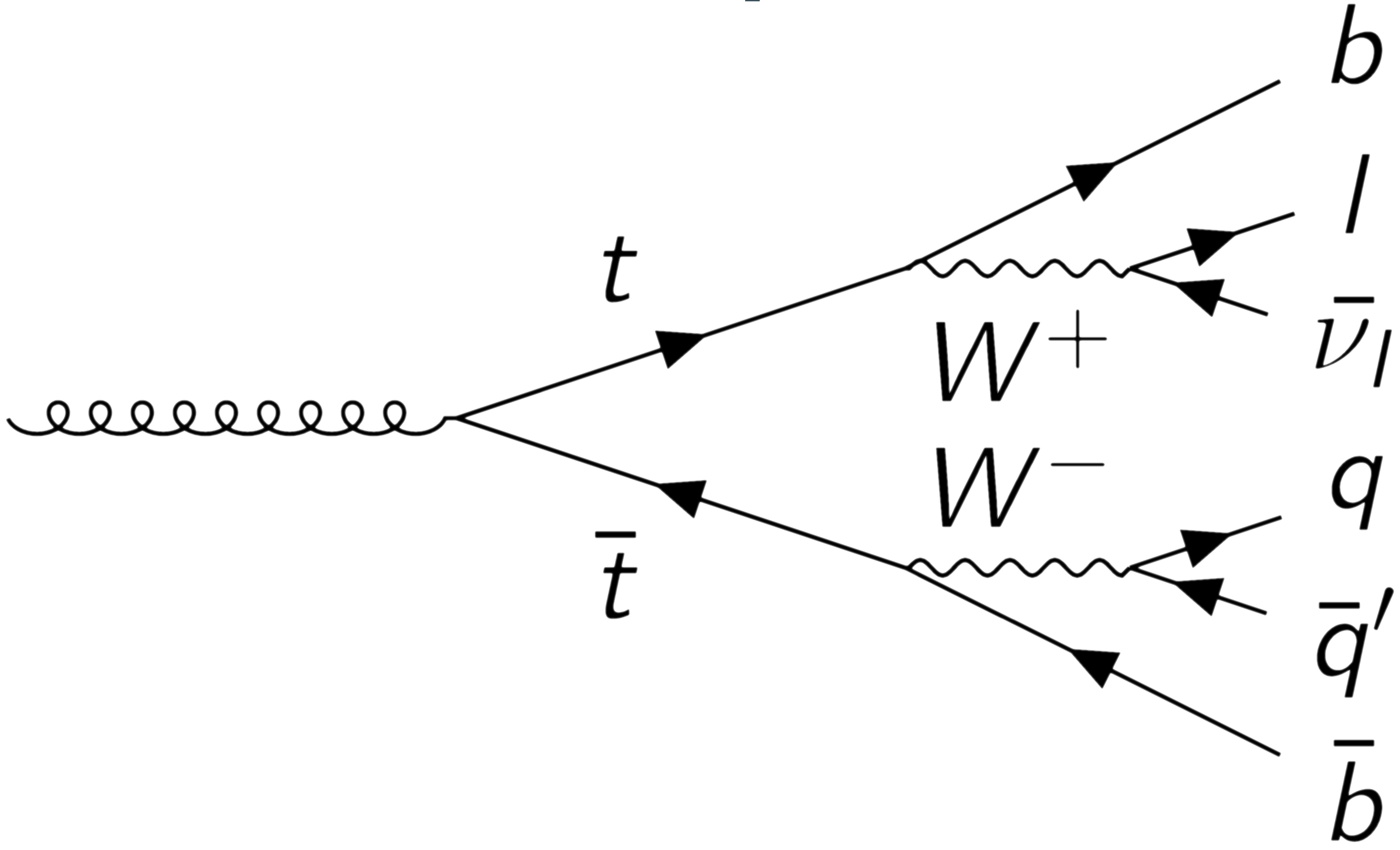
[1]Universität Hamburg - [2]DESY, Hamburg

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

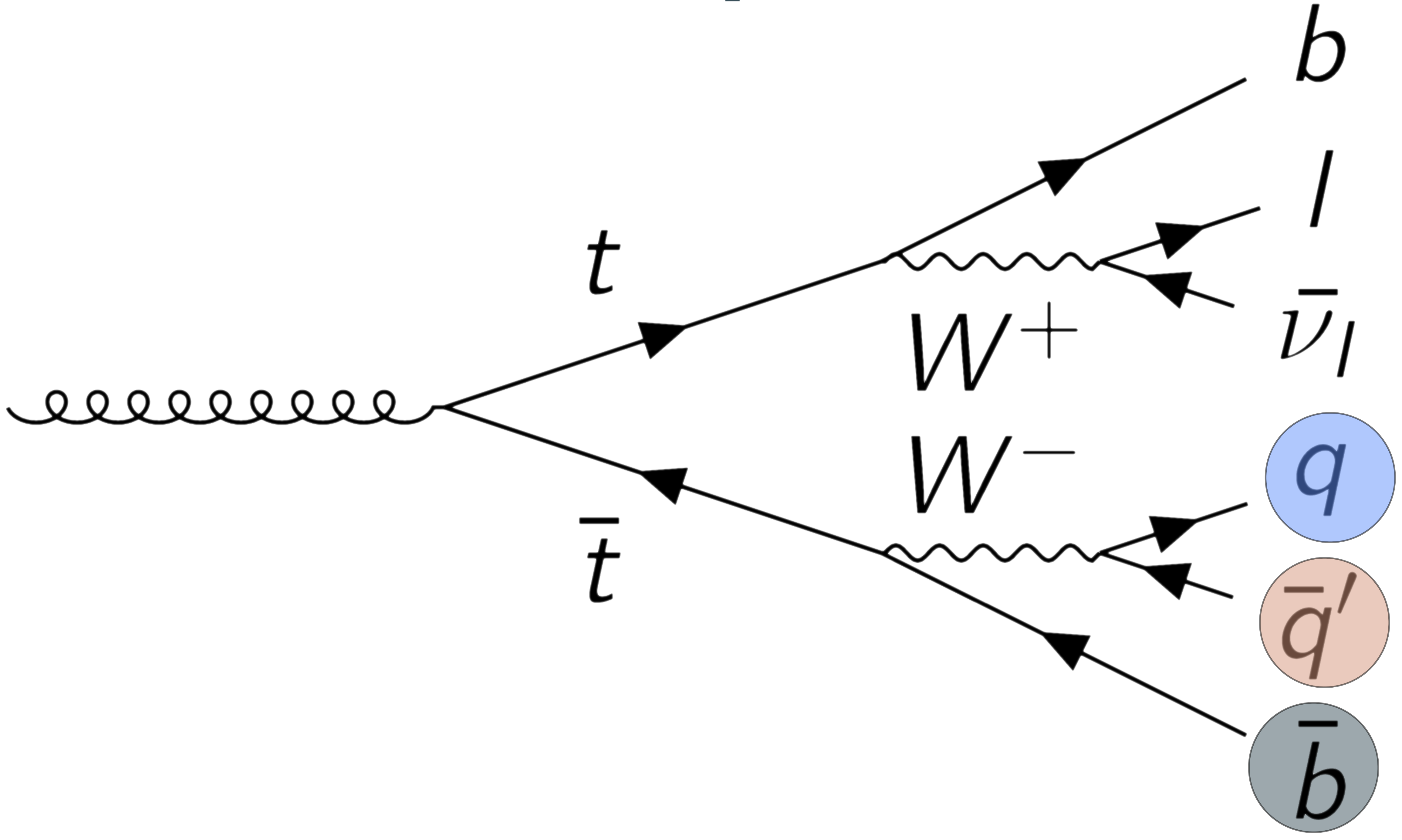
CLUSTER OF EXCELLENCE
QUANTUM UNIVERSE

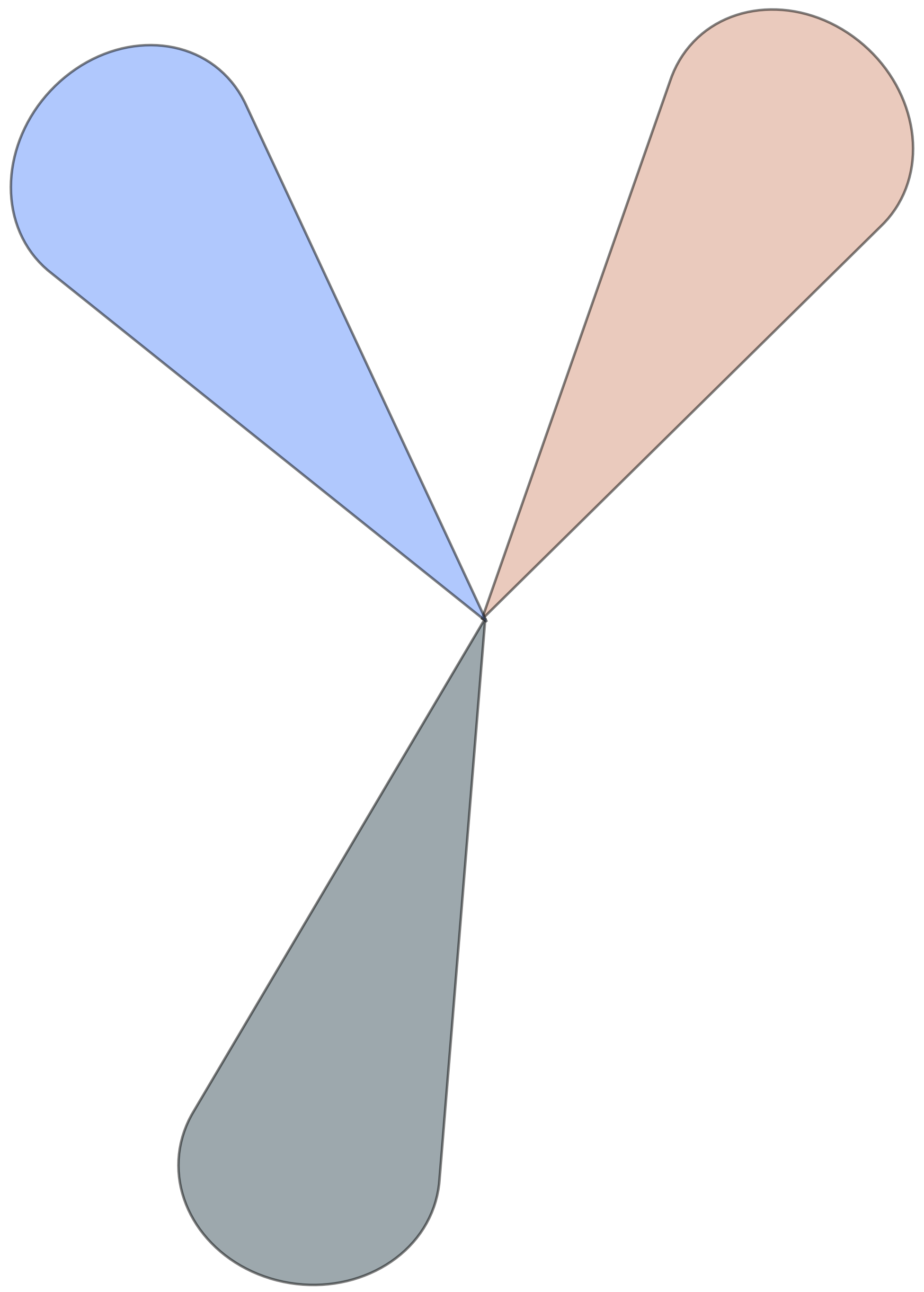
FSP CMS
Erforschung von
Universum und Materie

Boosted Top Quarks

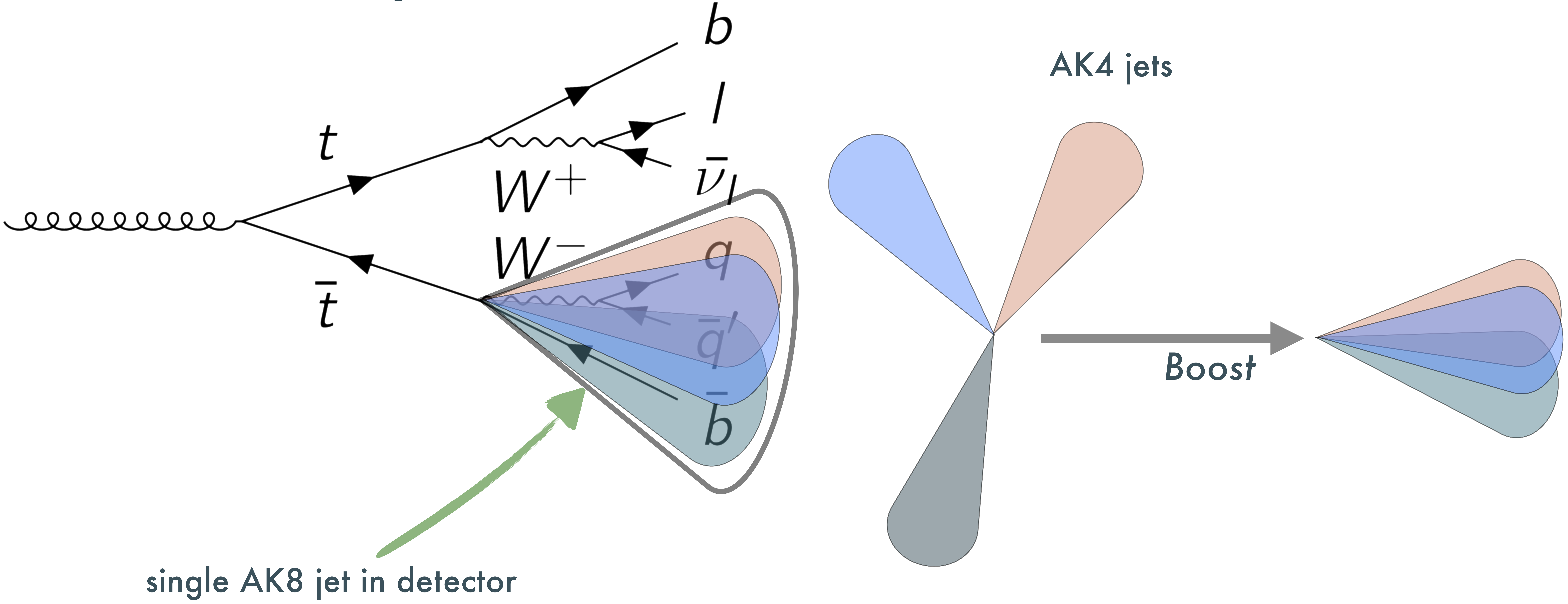


Boosted Top Quarks



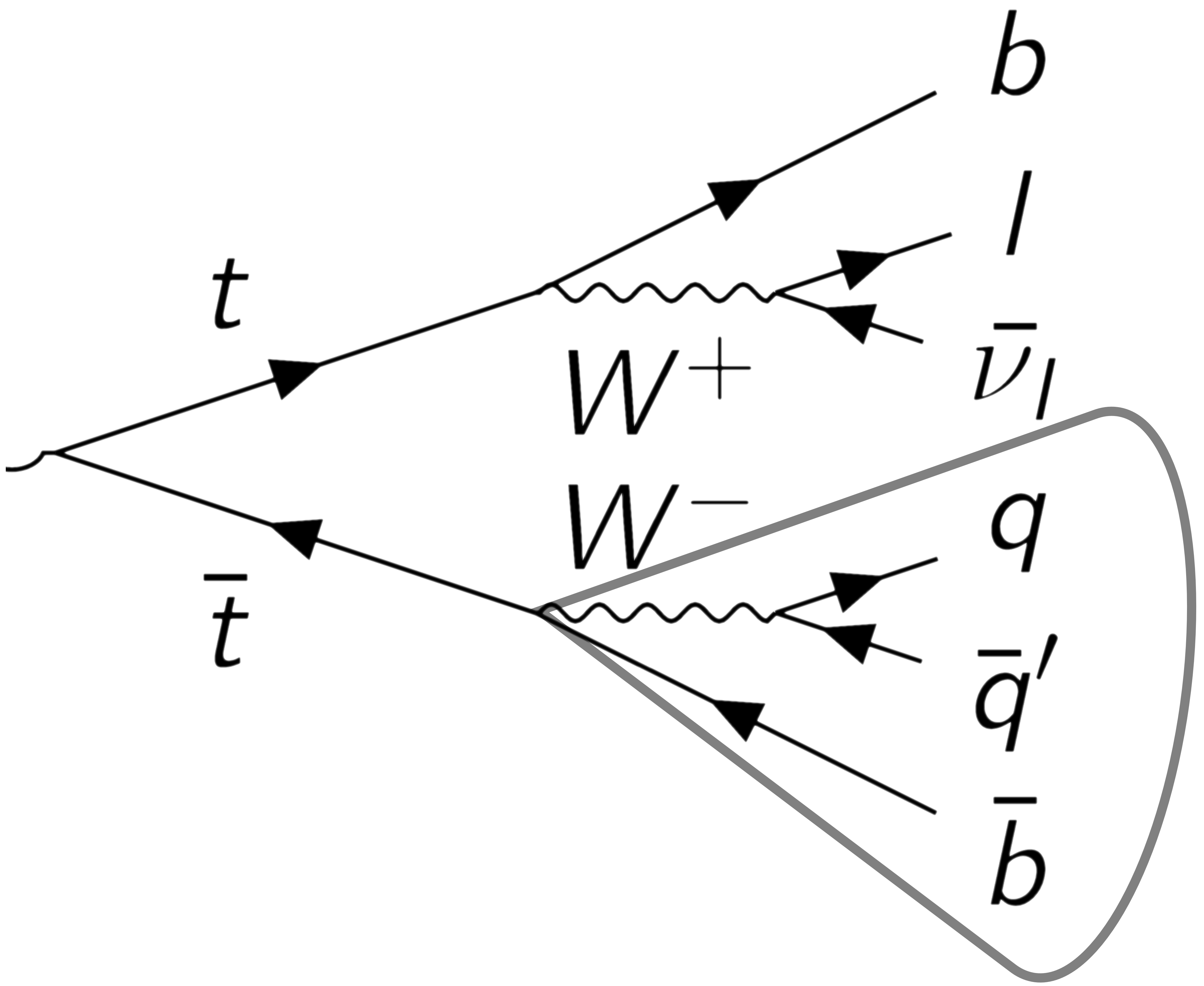


Boosted Top Quarks



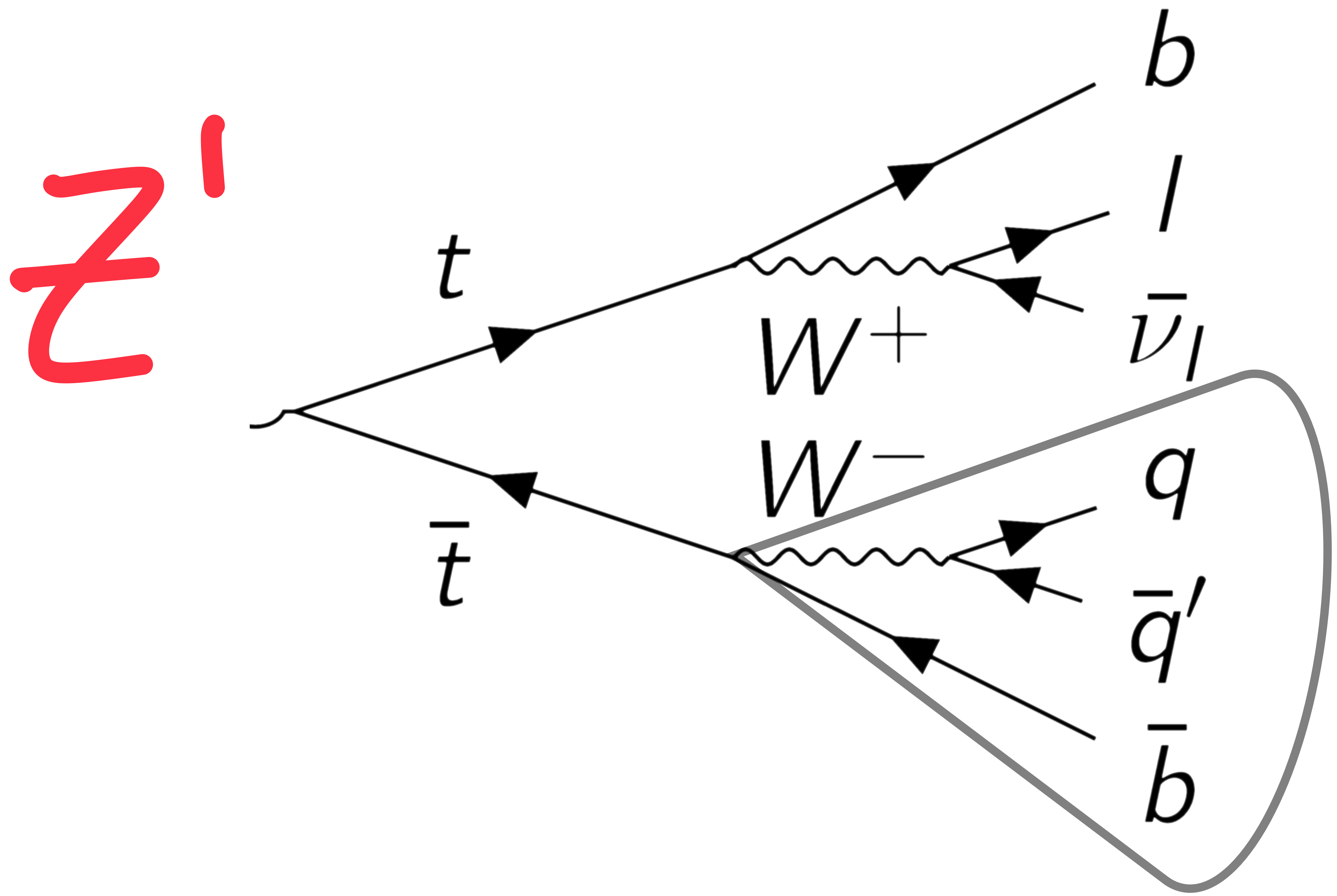
Boosted Top Quarks

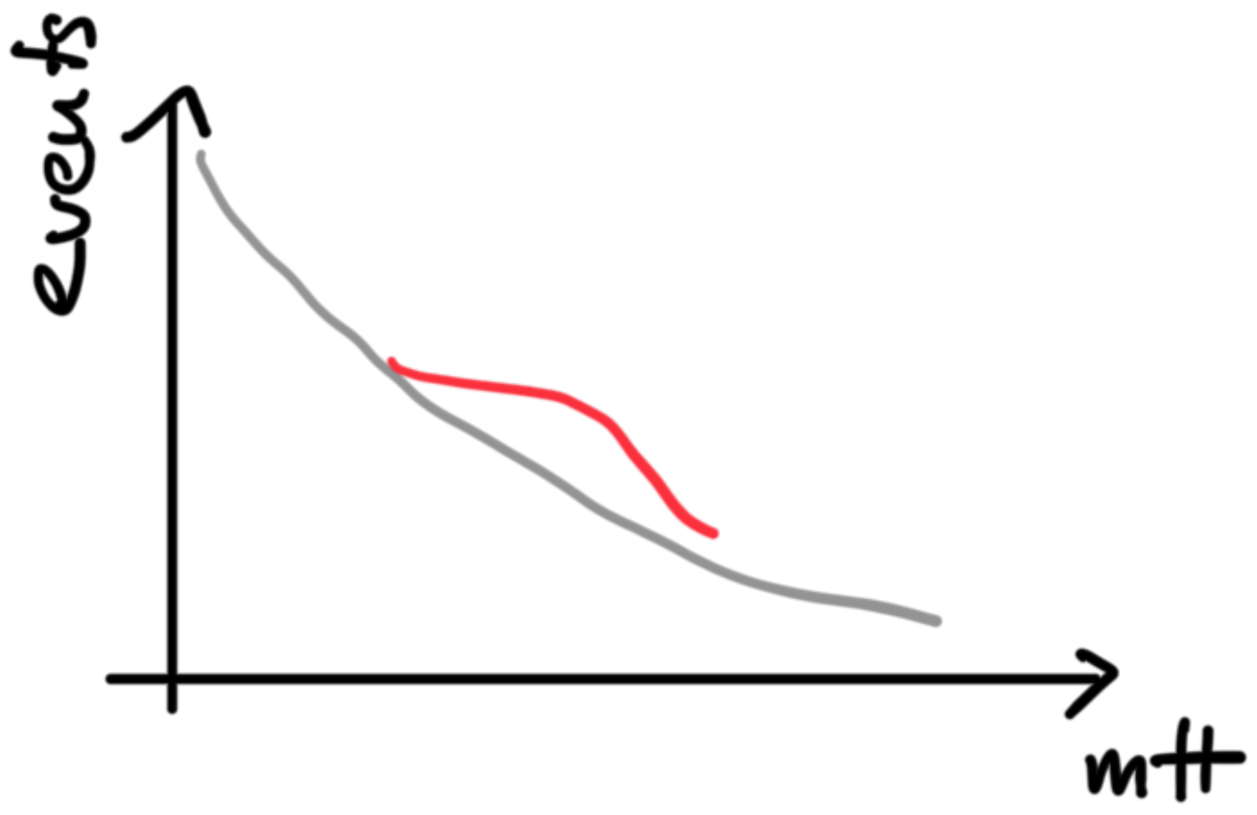
— in BSM signals —



Boosted Top Quarks

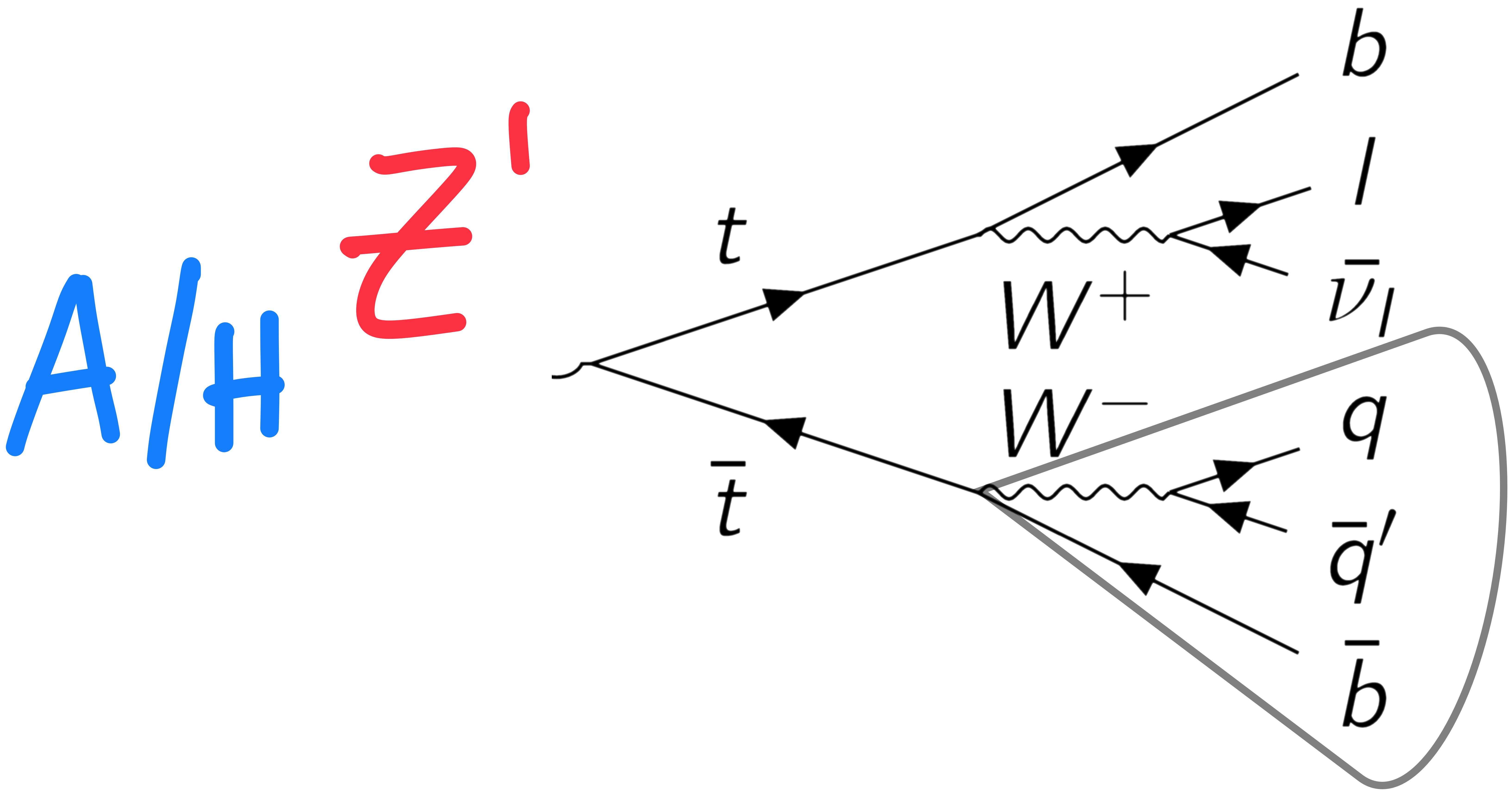
— in BSM signals —

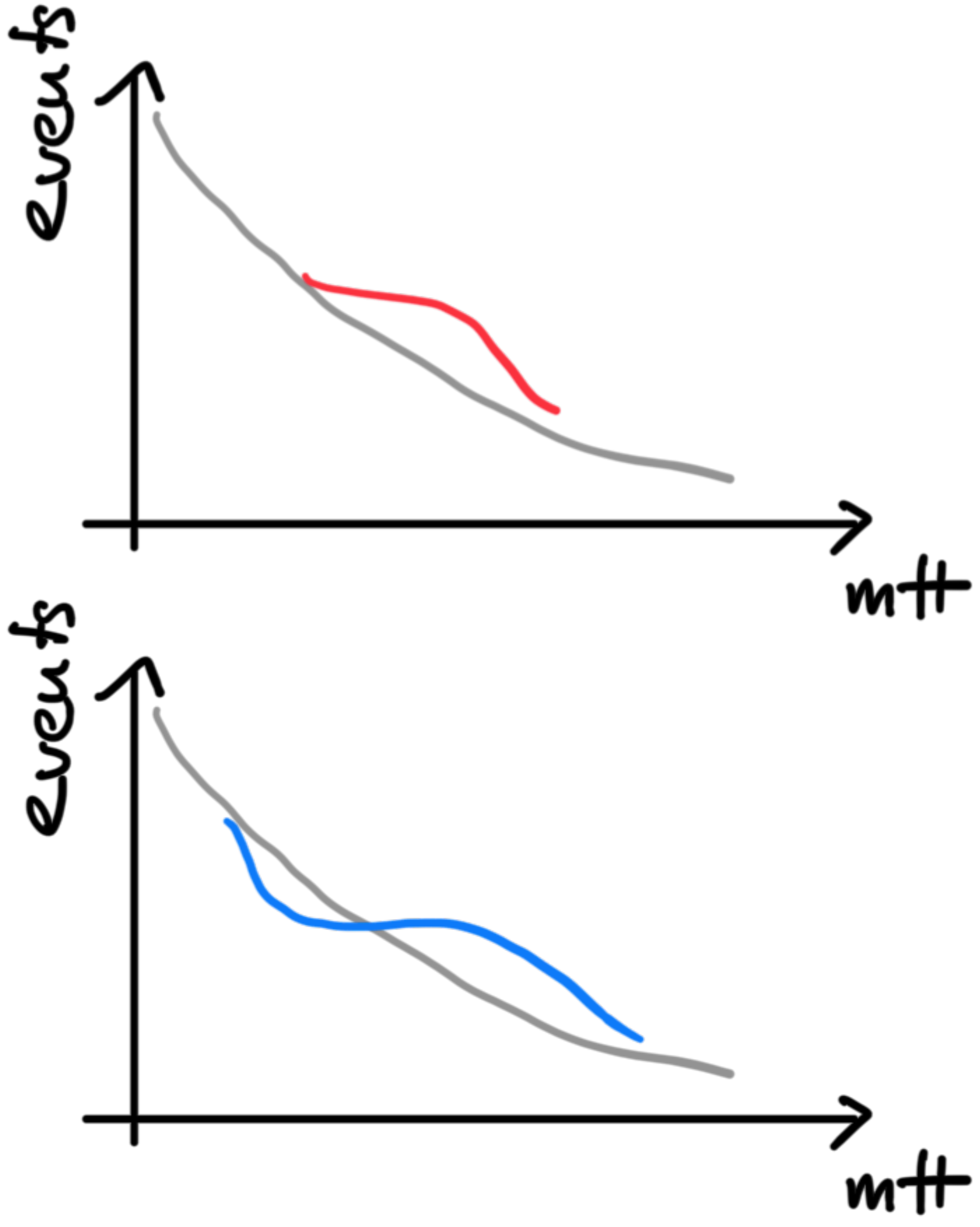




Boosted Top Quarks

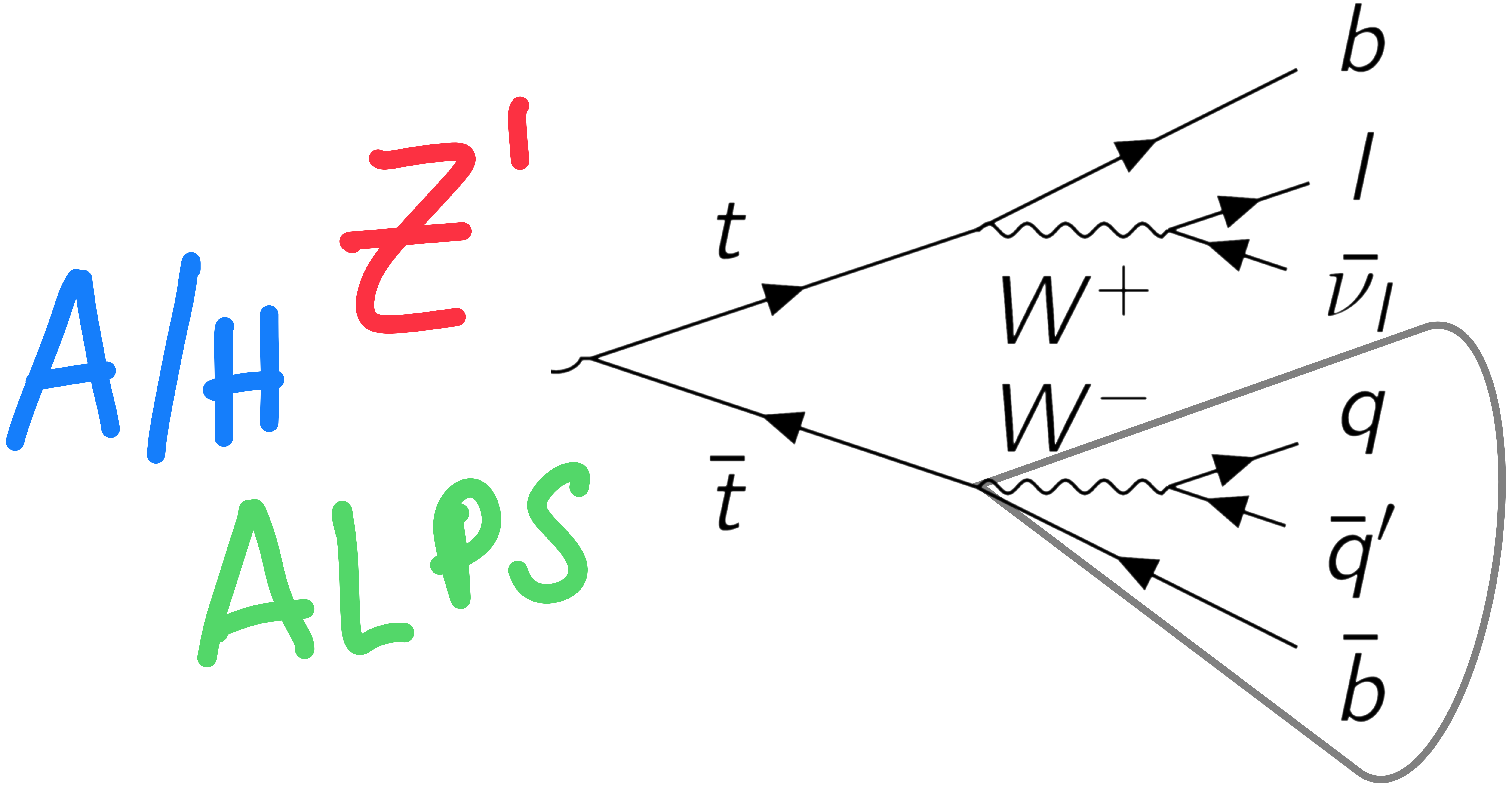
— in BSM signals —

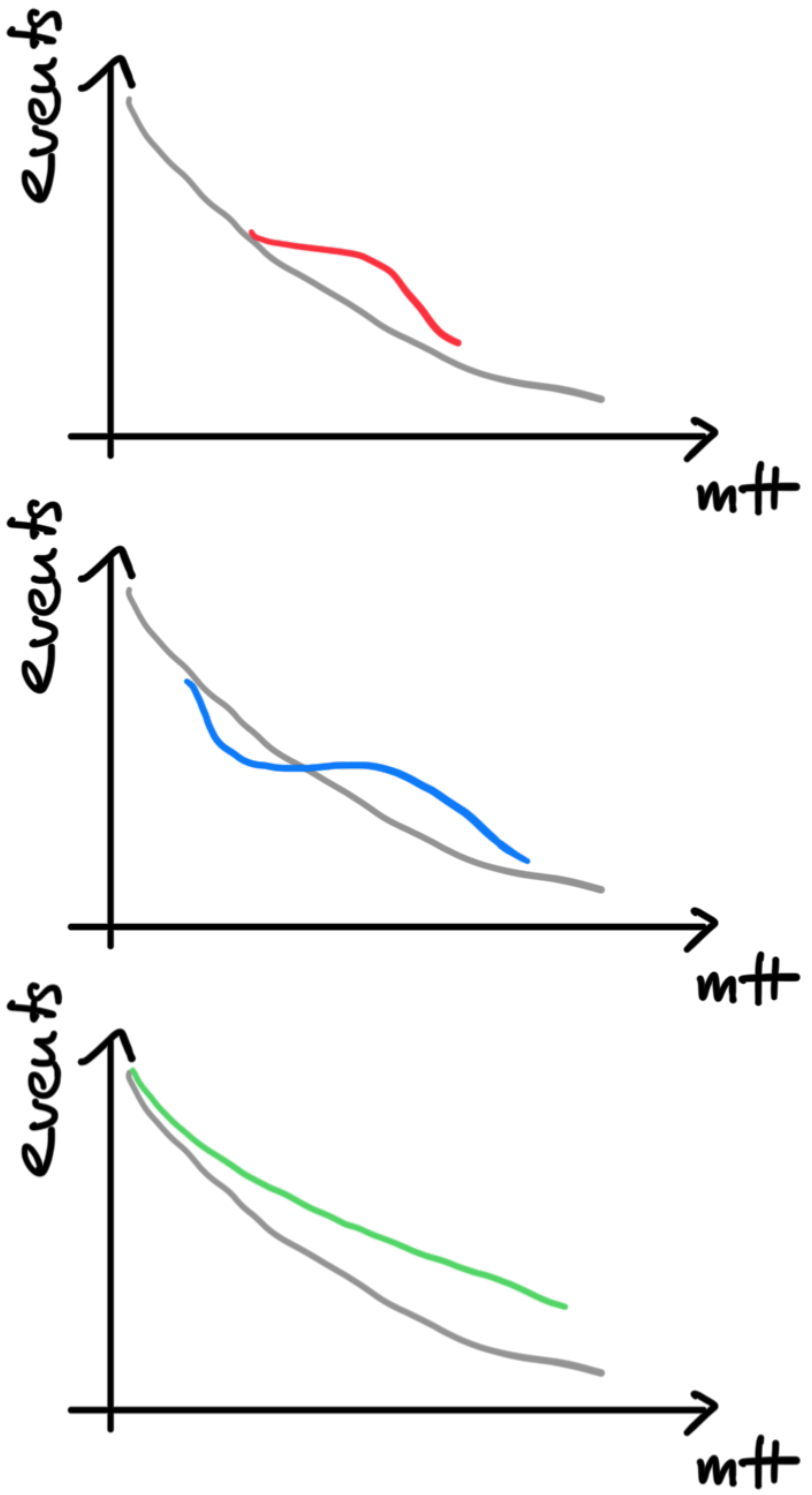




Boosted Top Quarks

— in BSM signals —





Identifying Boosted Top Quarks

— in CMS —

- different approaches to top tagging:
  - **machine learning** algorithms
    DeepAK8, ParticleNet, ...
    JINST 15 (2020) P06005 CMS-DP-2020-002
  - **cut based** using jet substructure techniques
    N-subjettiness ratio (τ_{32}) & soft drop mass (m_{SD})
    CMS-DP-2025-010

N-subjettiness

$$\tau_N = \frac{1}{d_0} \sum_k p_{T,k} \min(\Delta R_{1,k}, \Delta R_{2,k}, \ldots, \Delta R_{N,k})$$

JHEP 1103:015,2011 - arXiv:1011.2268

Identifying Boosted Top Quarks

— in CMS —

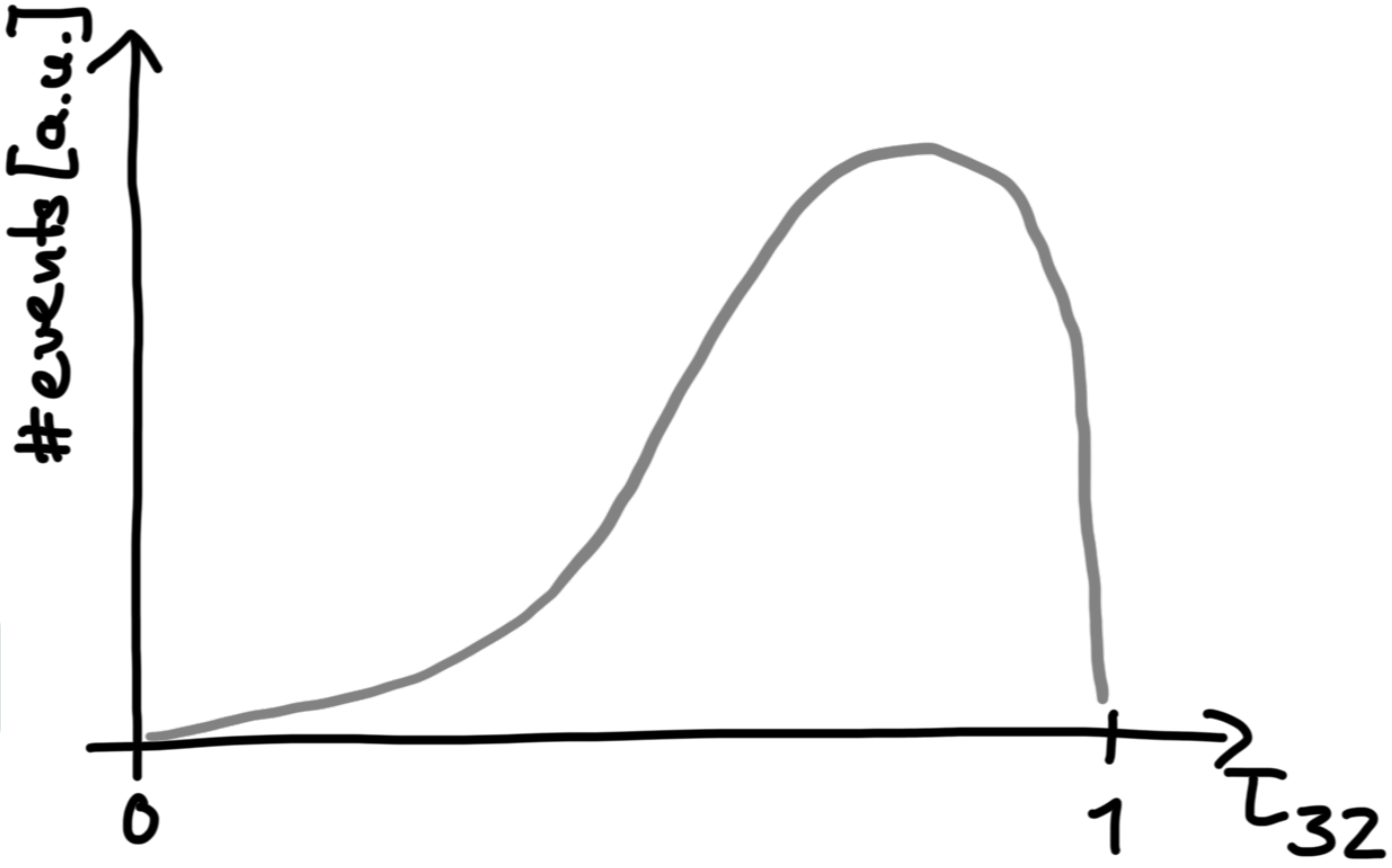
- different approaches to top tagging:
  - **machine learning** algorithms
    DeepAK8, ParticleNet, ...
    JINST 15 (2020) P06005 CMS-DP-2020-002
  - **cut based** using jet substructure techniques
    N-subjettiness ratio (τ_{32}) & soft drop mass (m_{SD})
    CMS-DP-2025-010

N-subjettiness

$$\tau_N = \frac{1}{d_0} \sum_k p_{\mathrm{T},k} \min(\Delta R_{1,k}, \Delta R_{2,k}, \ldots, \Delta R_{N,k})$$

JHEP 1103:015,2011 - arXiv:1011.2268

$$\tau_{32} := \tau_3/\tau_2$$



Identifying Boosted Top Quarks

— in CMS —

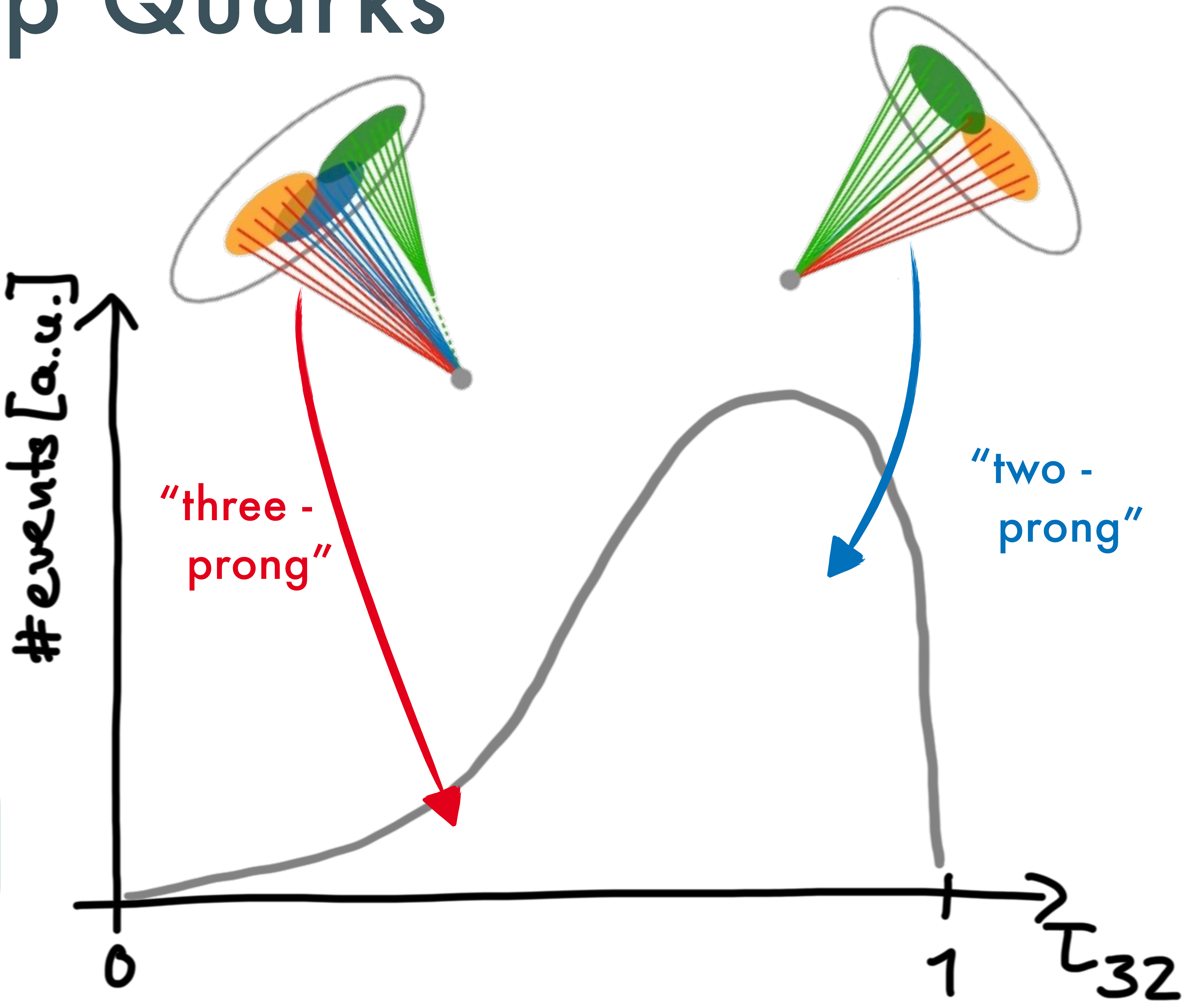
- different approaches to top tagging:
  - **machine learning** algorithms
    DeepAK8, ParticleNet, …
    JINST 15 (2020) P06005 CMS-DP-2020-002
  - **cut based** using jet substructure techniques
    N-subjettiness ratio (τ_{32}) & soft drop mass (m_{SD})
    CMS-DP-2025-010

N-subjettiness

$$\tau_N = \frac{1}{d_0} \sum_k p_{T,k} \min(\Delta R_{1,k}, \Delta R_{2,k}, \ldots, \Delta R_{N,k})$$

JHEP 1103:015,2011 - arXiv:1011.2268

$$\tau_{32} := \tau_3/\tau_2$$



Identifying Boosted Top Quarks

— in CMS —

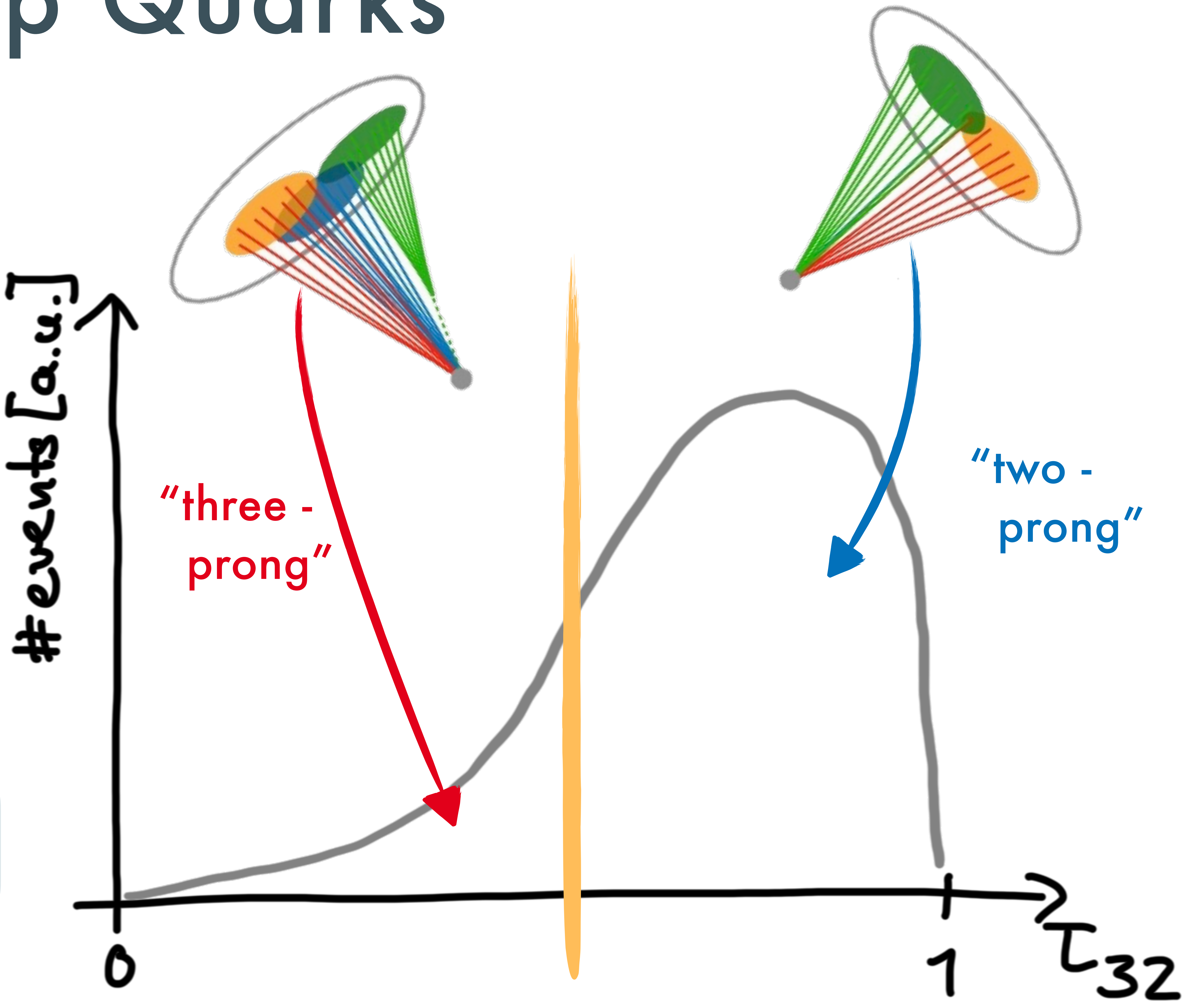
- different approaches to top tagging:
  - **machine learning** algorithms
    DeepAK8, ParticleNet, ...
    JINST 15 (2020) P06005 CMS-DP-2020-002
  - **cut based** using jet substructure techniques
    N-subjettiness ratio (τ_{32}) & soft drop mass (m_{SD})
    CMS-DP-2025-010

N-subjettiness

$$\tau_N = \frac{1}{d_0} \sum_k p_{\mathrm{T},k} \min(\Delta R_{1,k}, \Delta R_{2,k}, \dots, \Delta R_{N,k})$$

JHEP 1103:015,2011 - arXiv:1011.2268

$$\tau_{32} := \tau_3/\tau_2$$

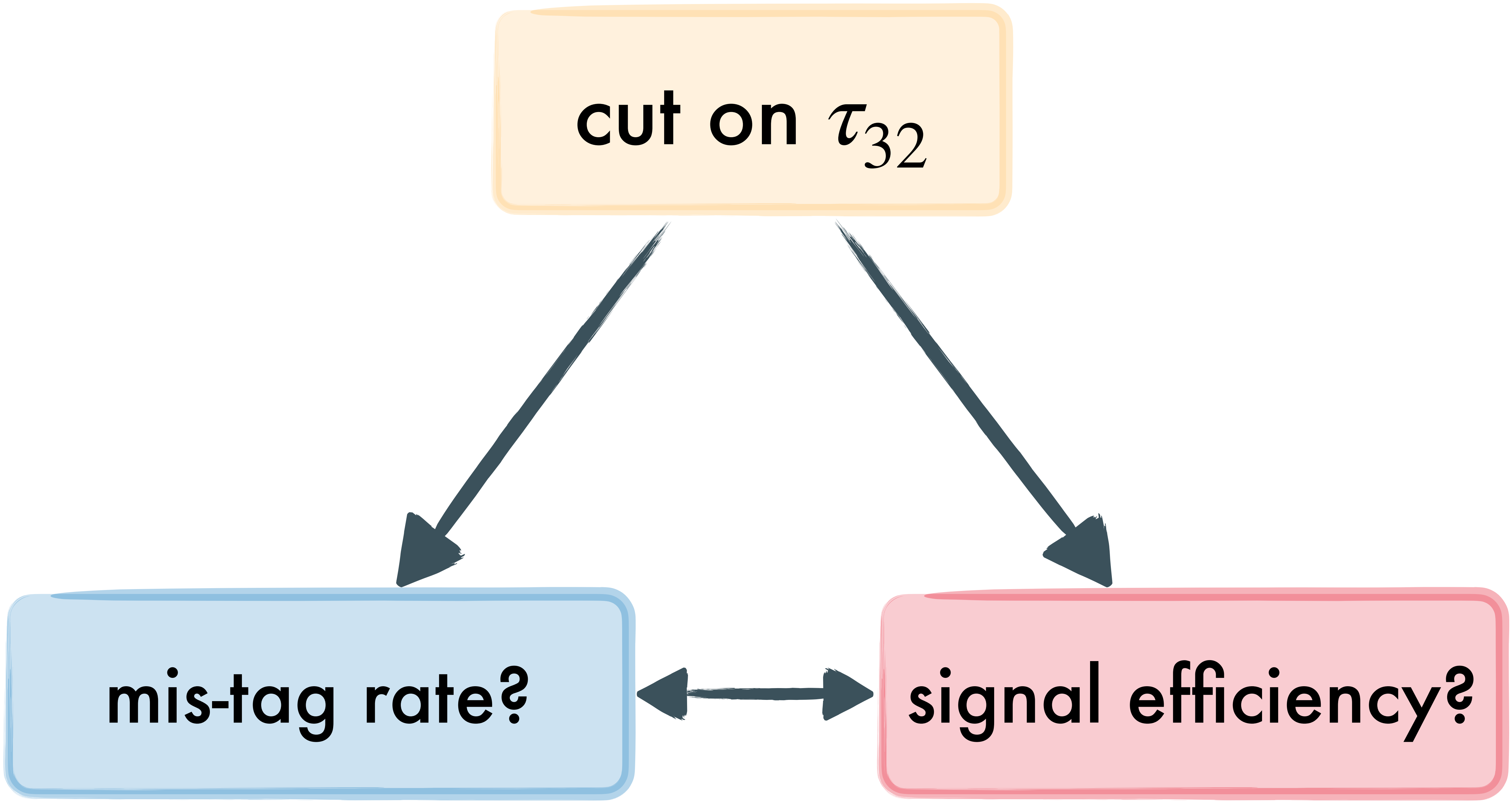


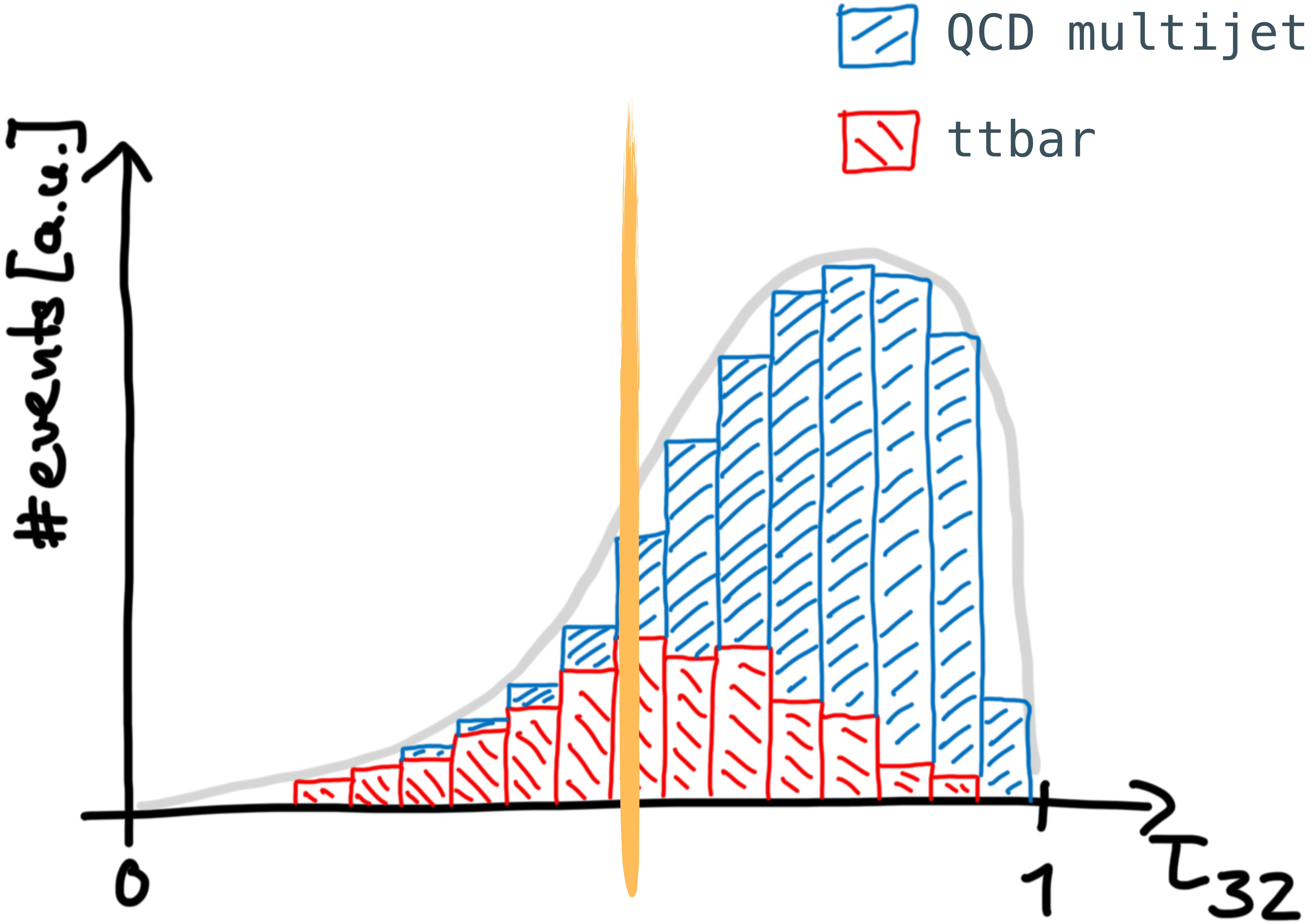
Separating Signal From Background

cut on τ_{32}

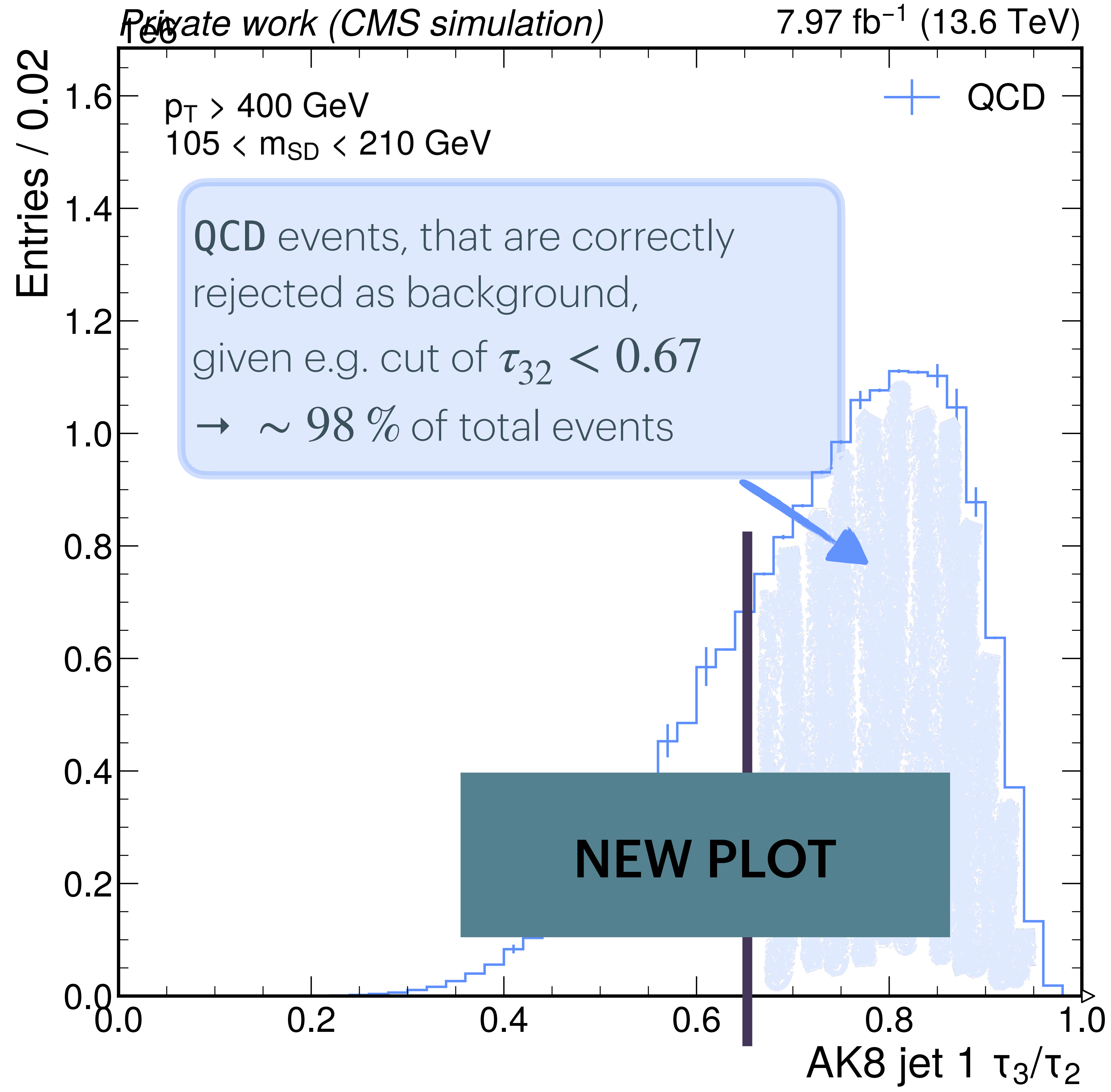
mis-tag rate? ⟷ signal efficiency?

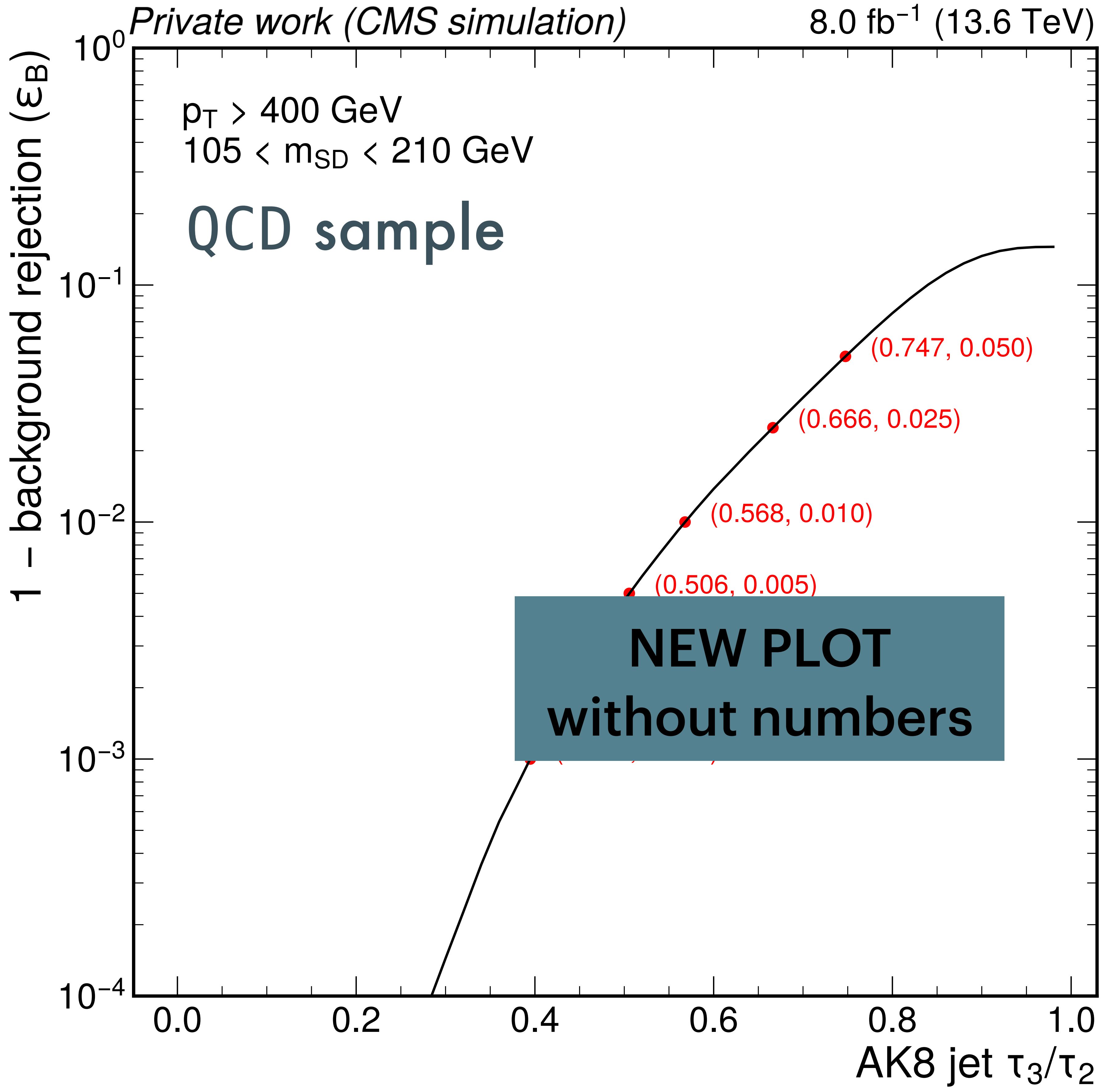
Separating Signal From Background





Mis-Tag Rate

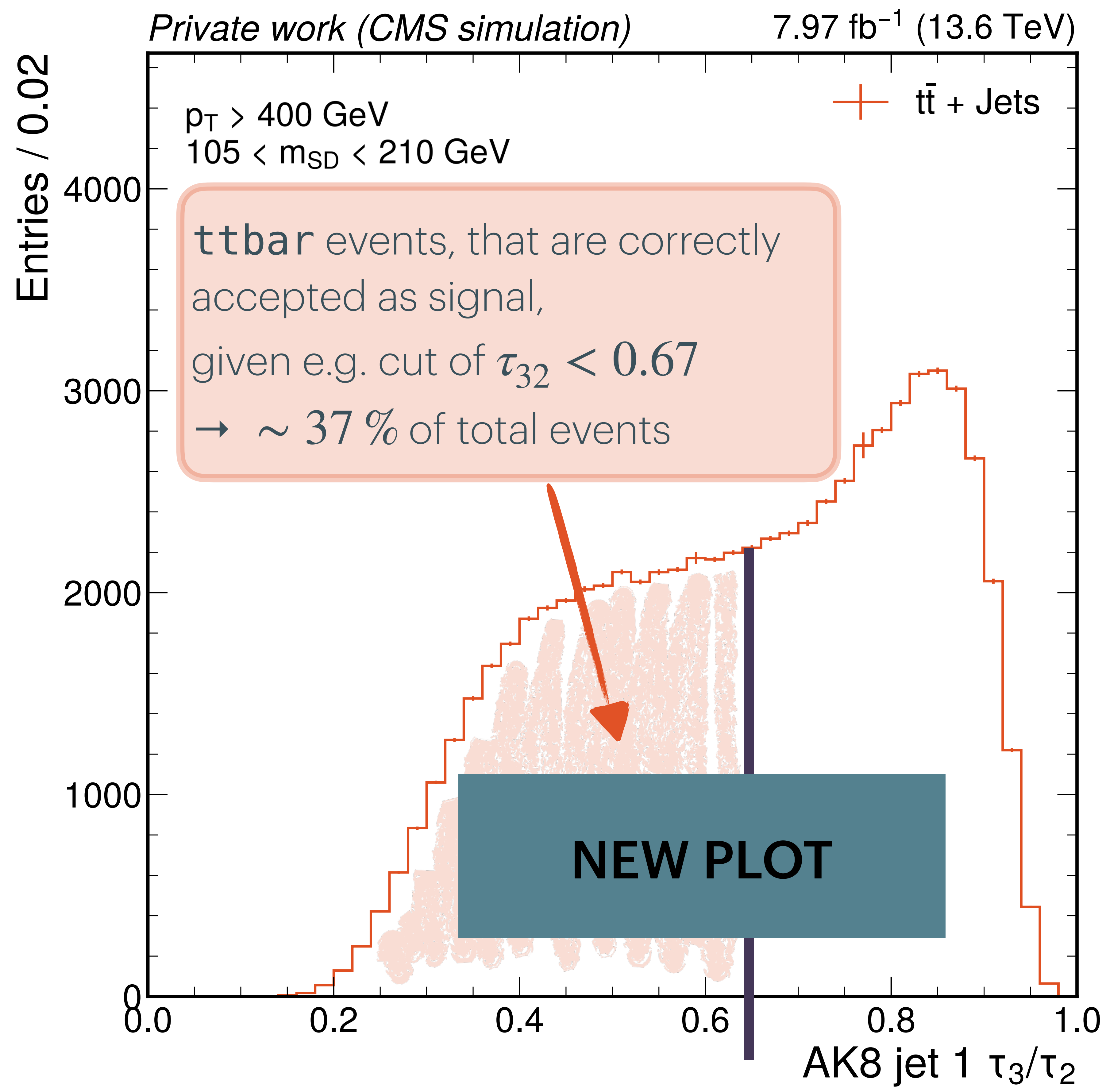


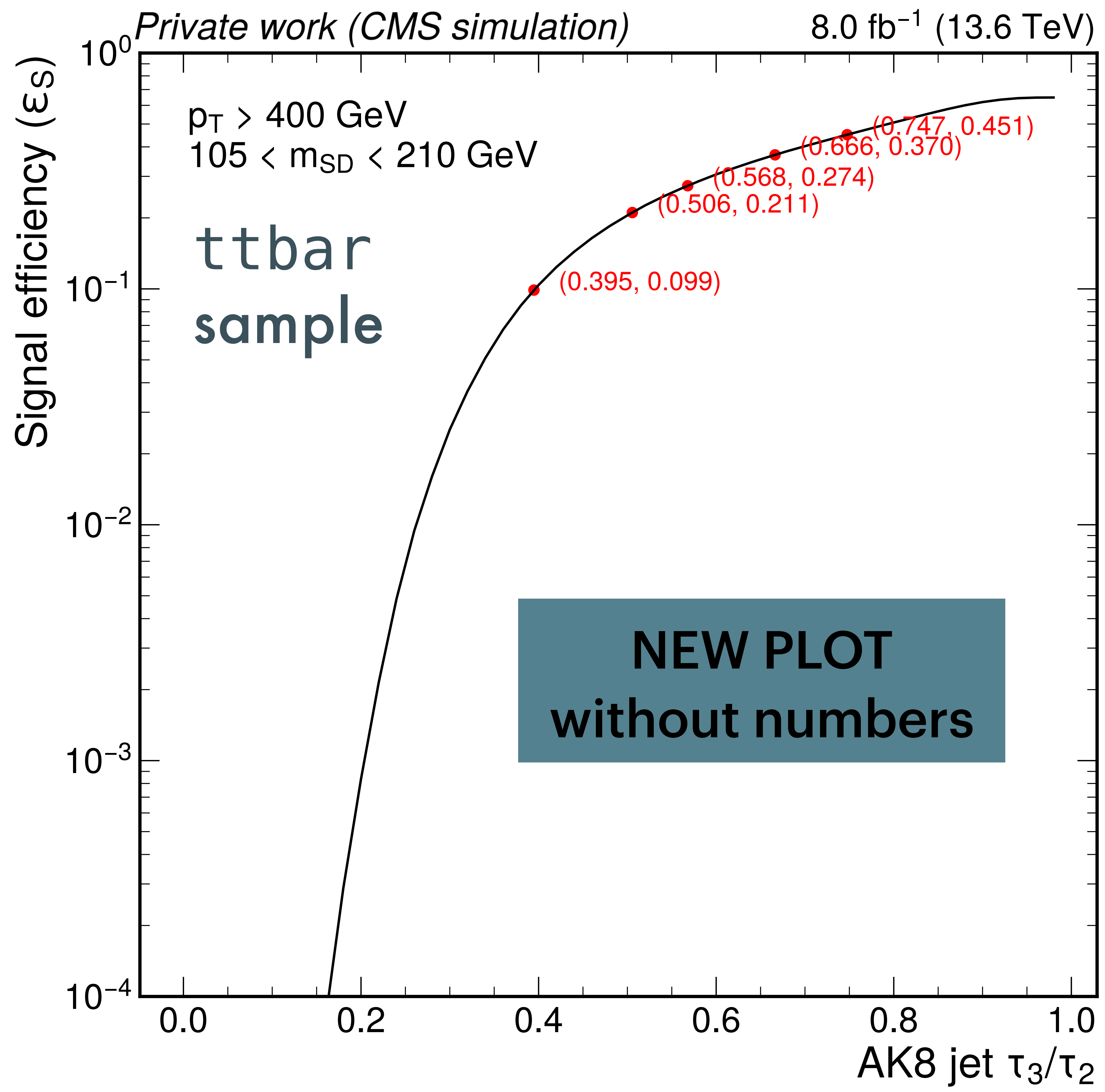


$$\frac{\text{number of QCD events with a FatJet passing selection AND (incorrectly) tagged as a top jet given the } \tau_{32} \text{ cut}}{\text{total number of events in QCD dataset}}$$

-

Signal Efficiency

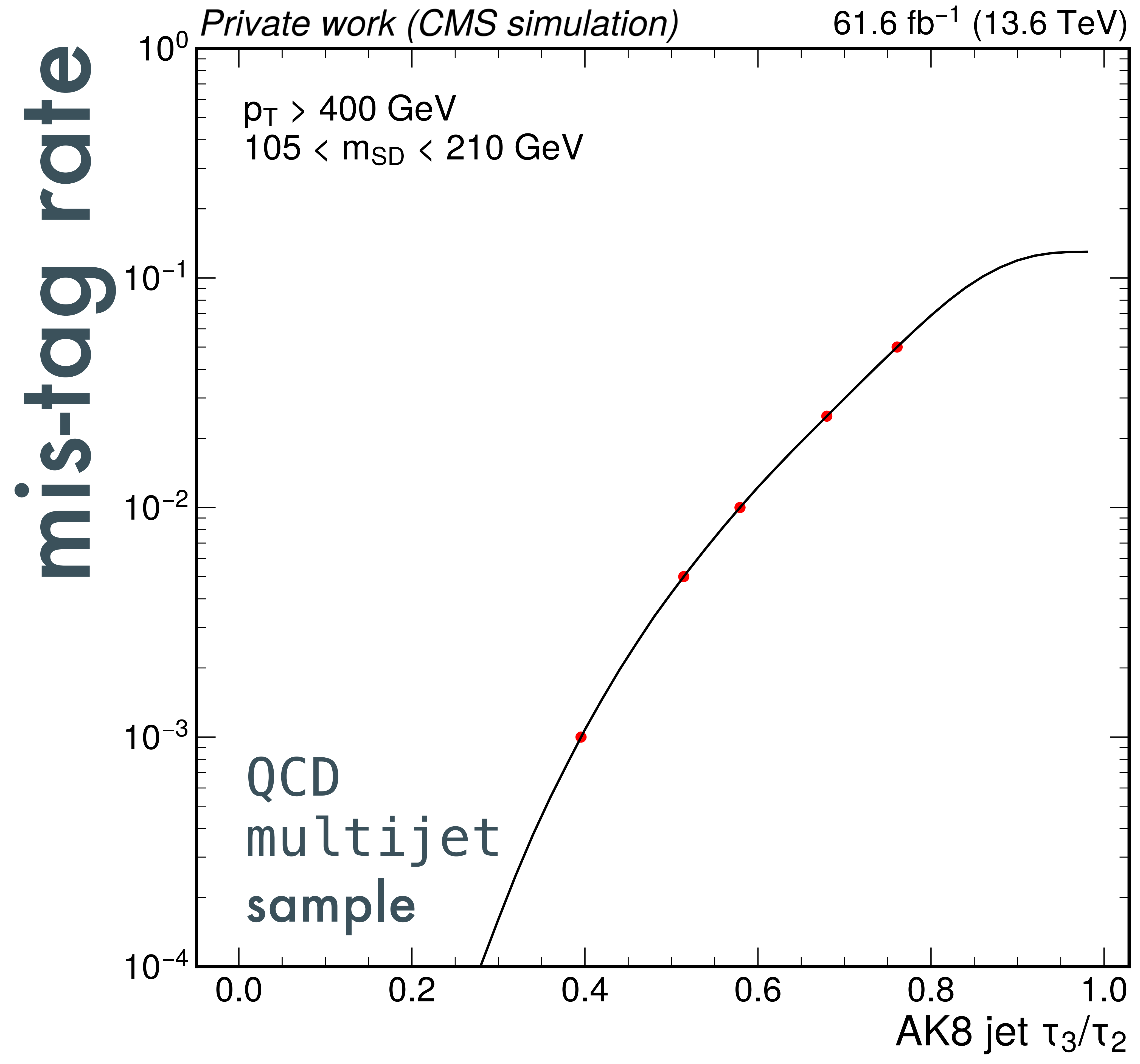




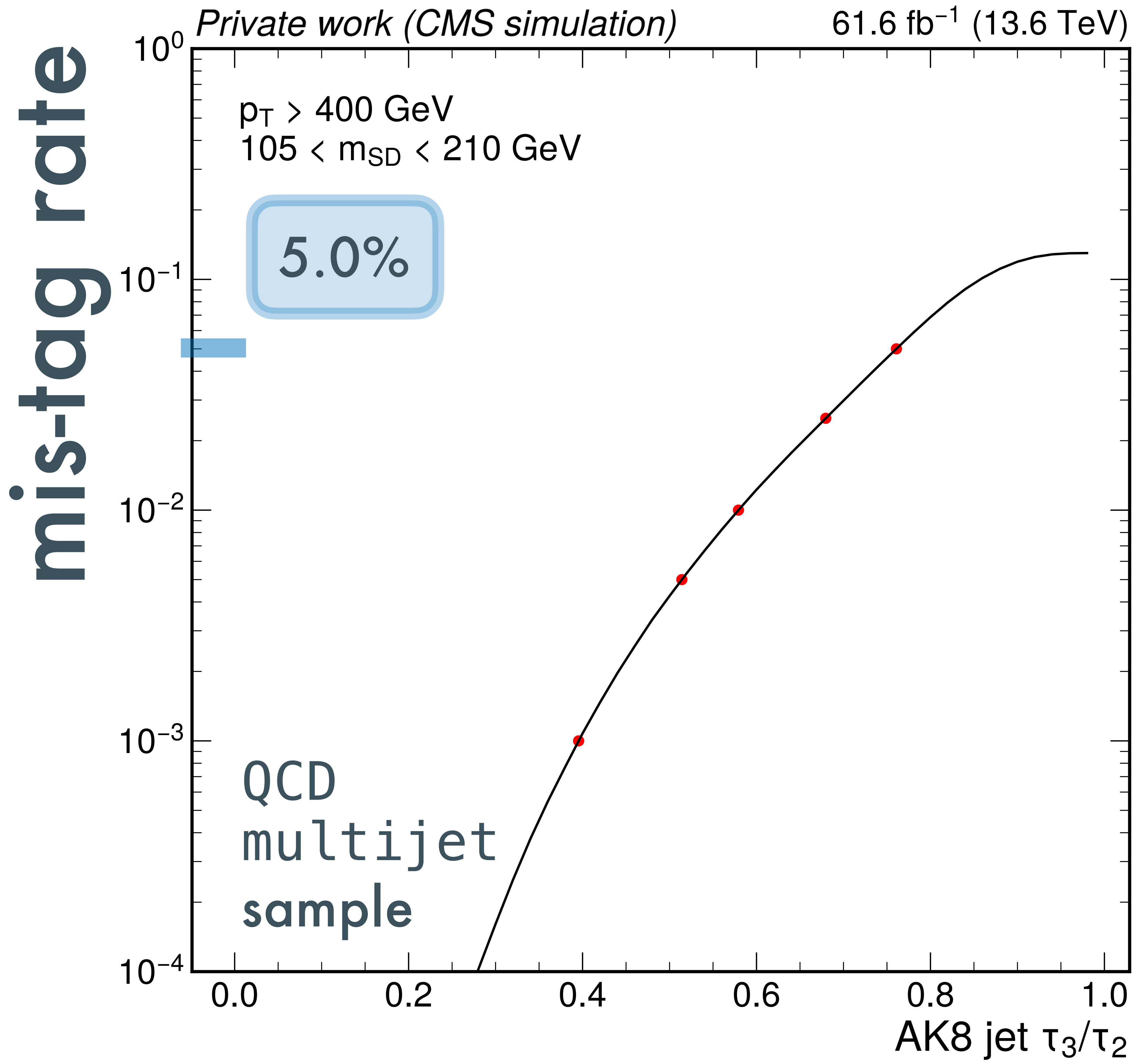
$$\frac{\text{number of ttbar events with a FatJet passing selection AND (correctly) tagged as a top jet given the } \tau_{32} \text{ cut}}{\text{total number of events in ttbar dataset}}$$

-

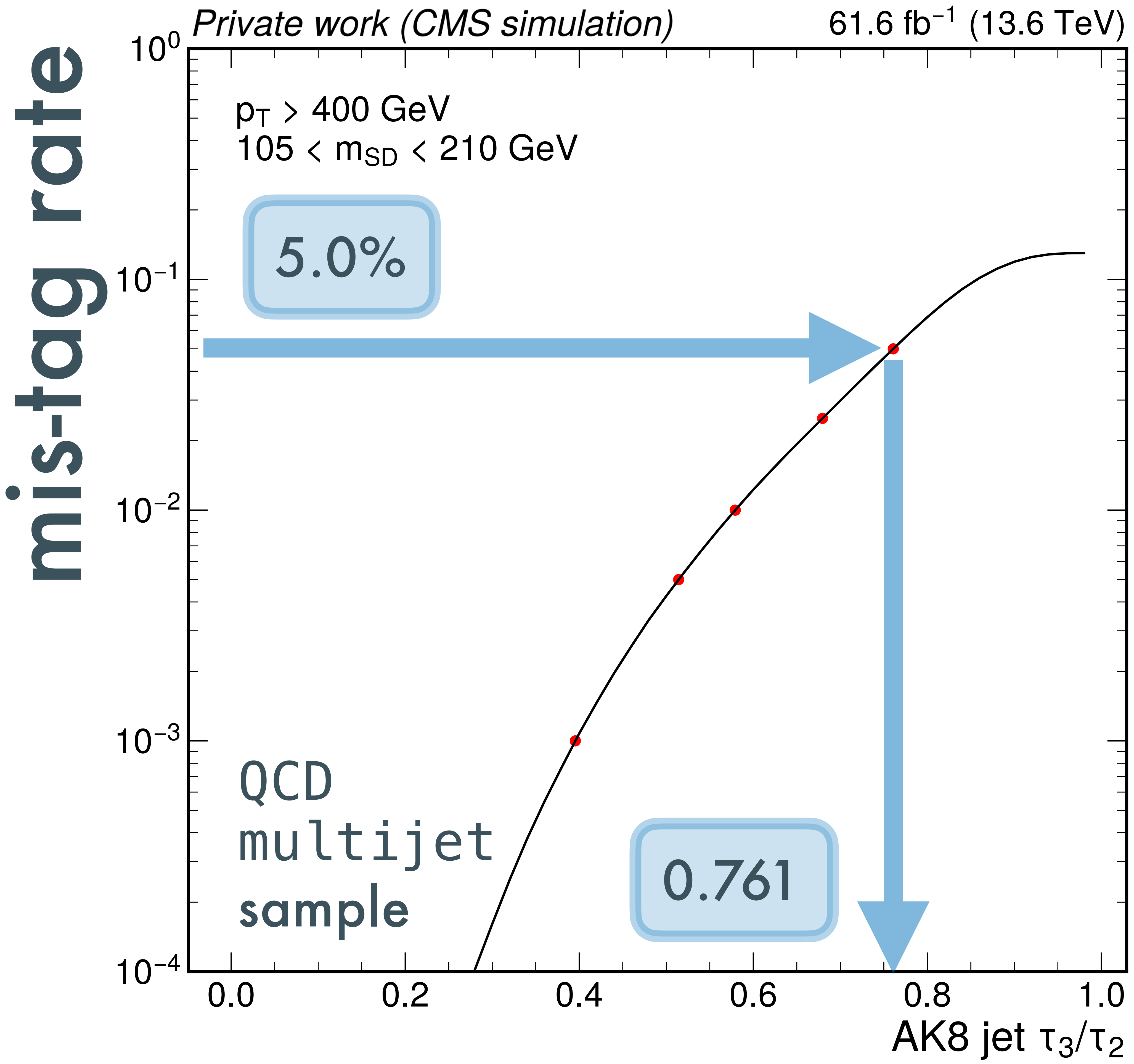
Working Points in One Slide



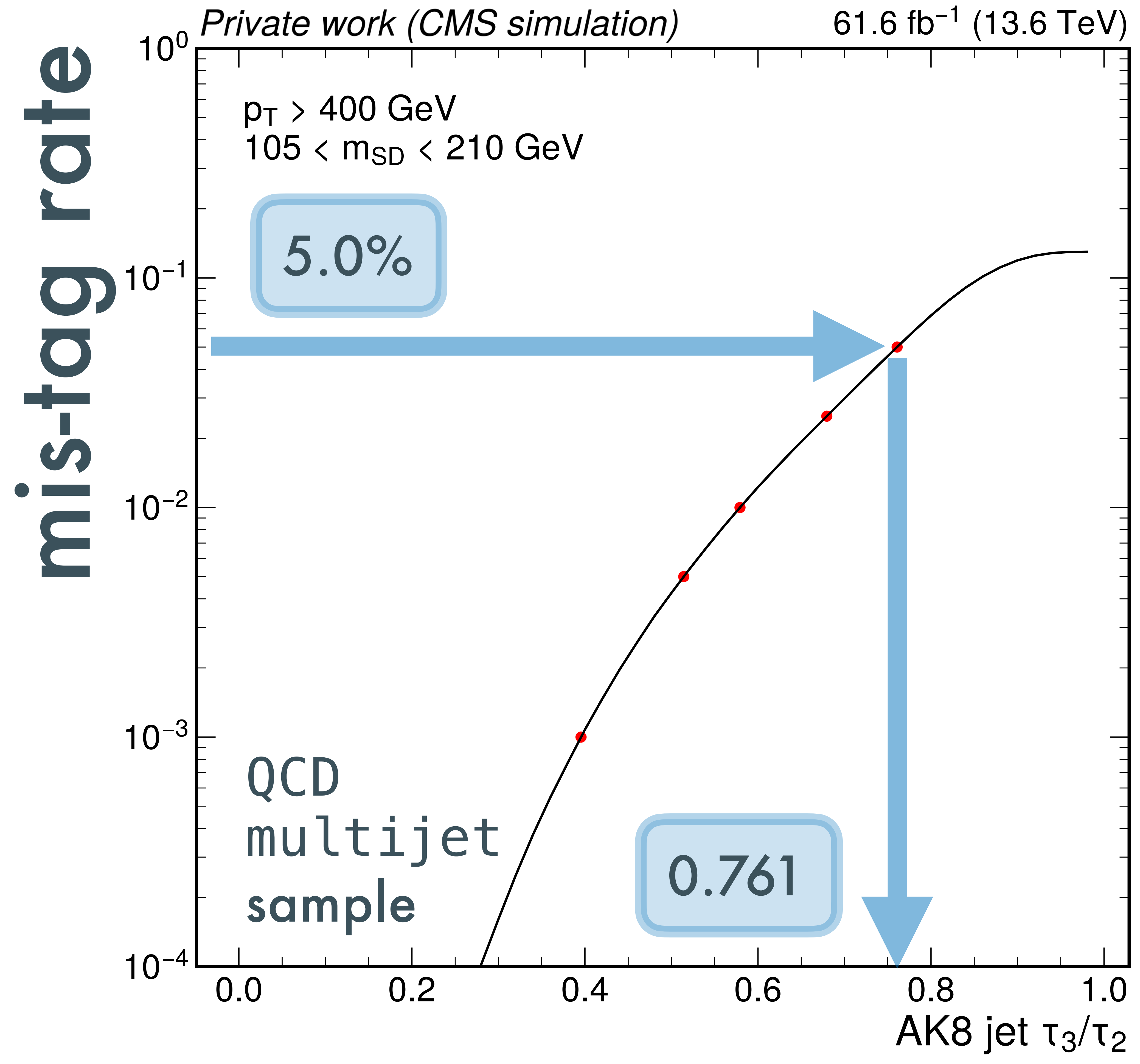
Working Points in One Slide

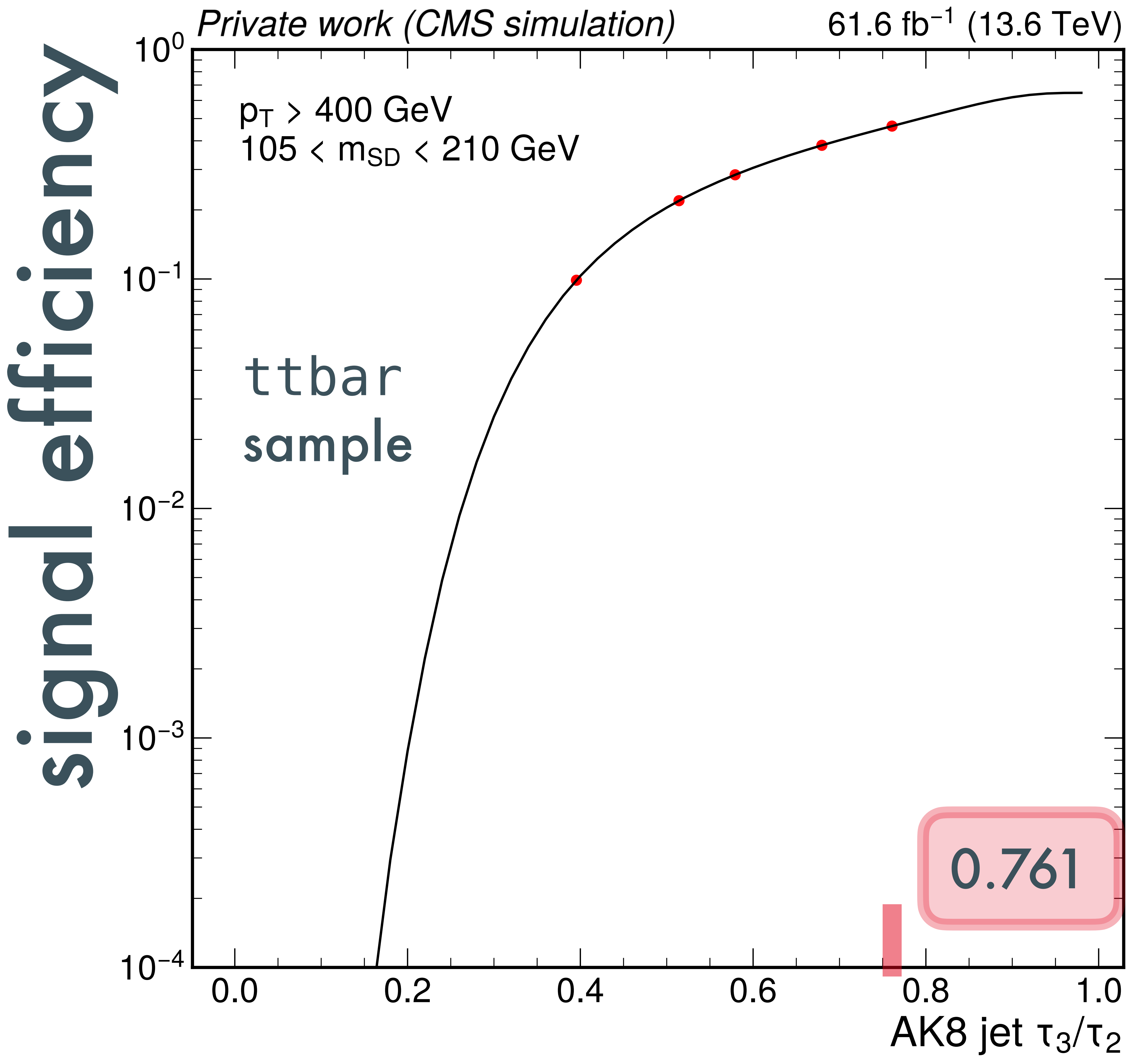


Working Points in One Slide

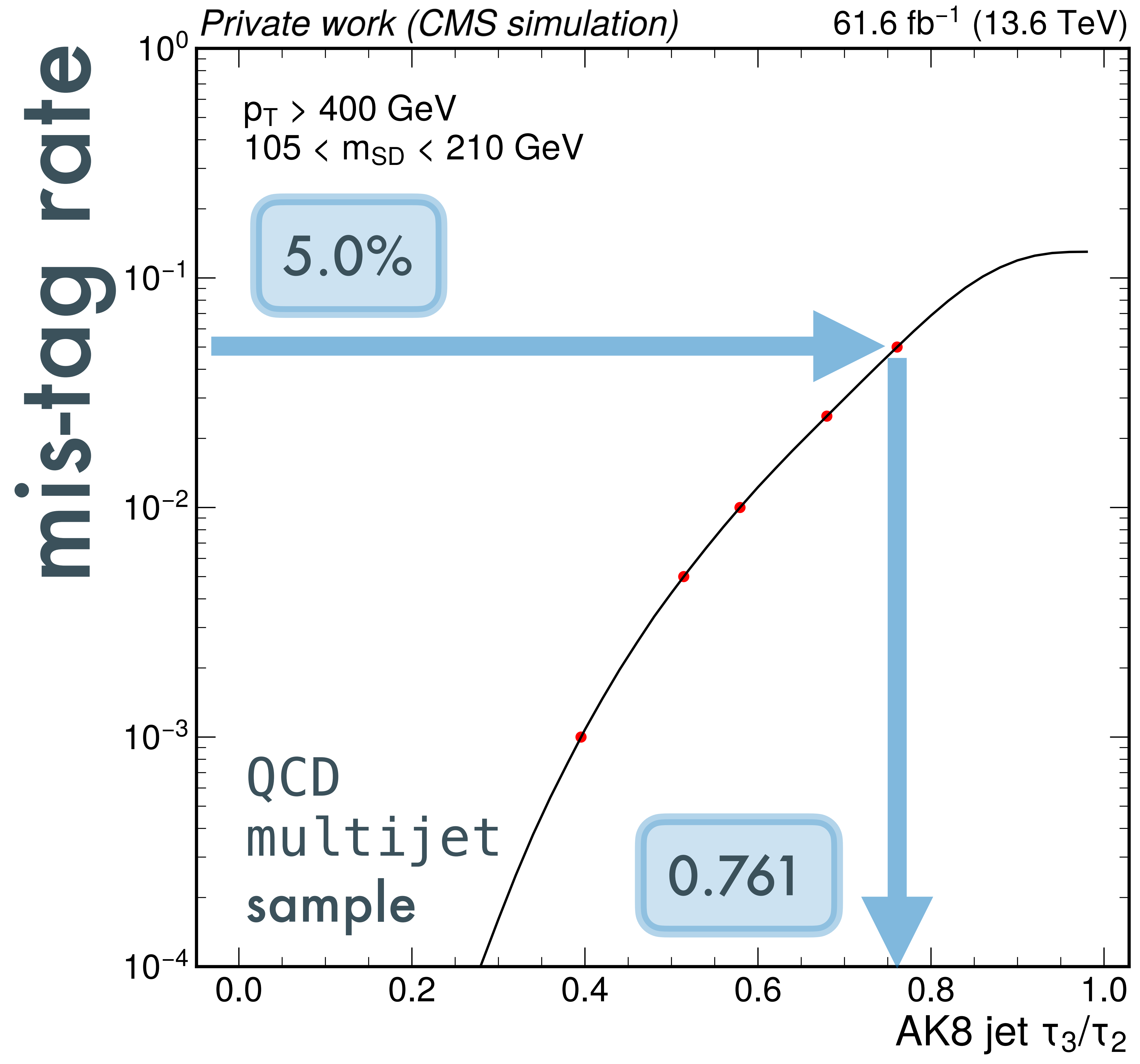


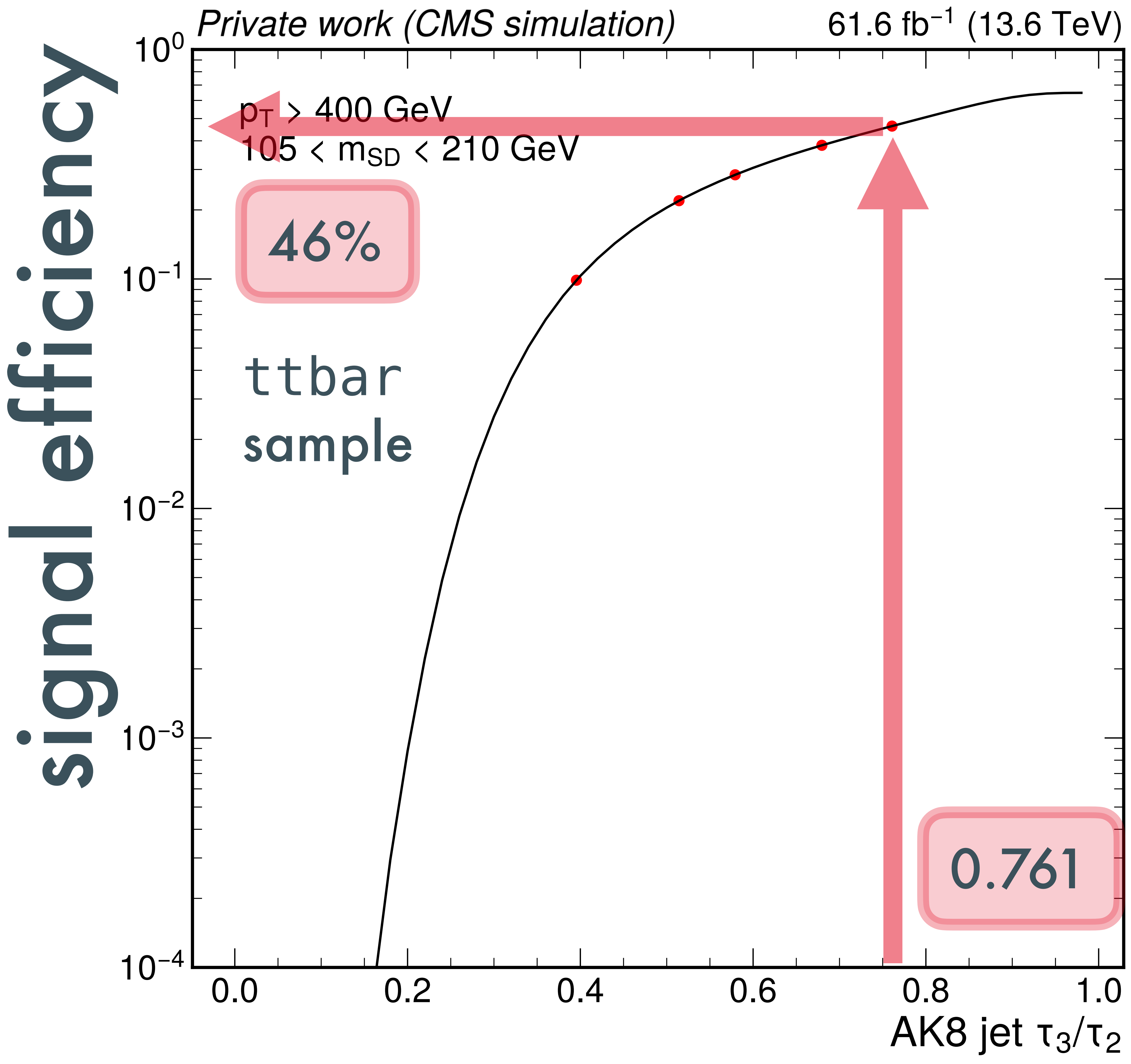
Working Points in One Slide



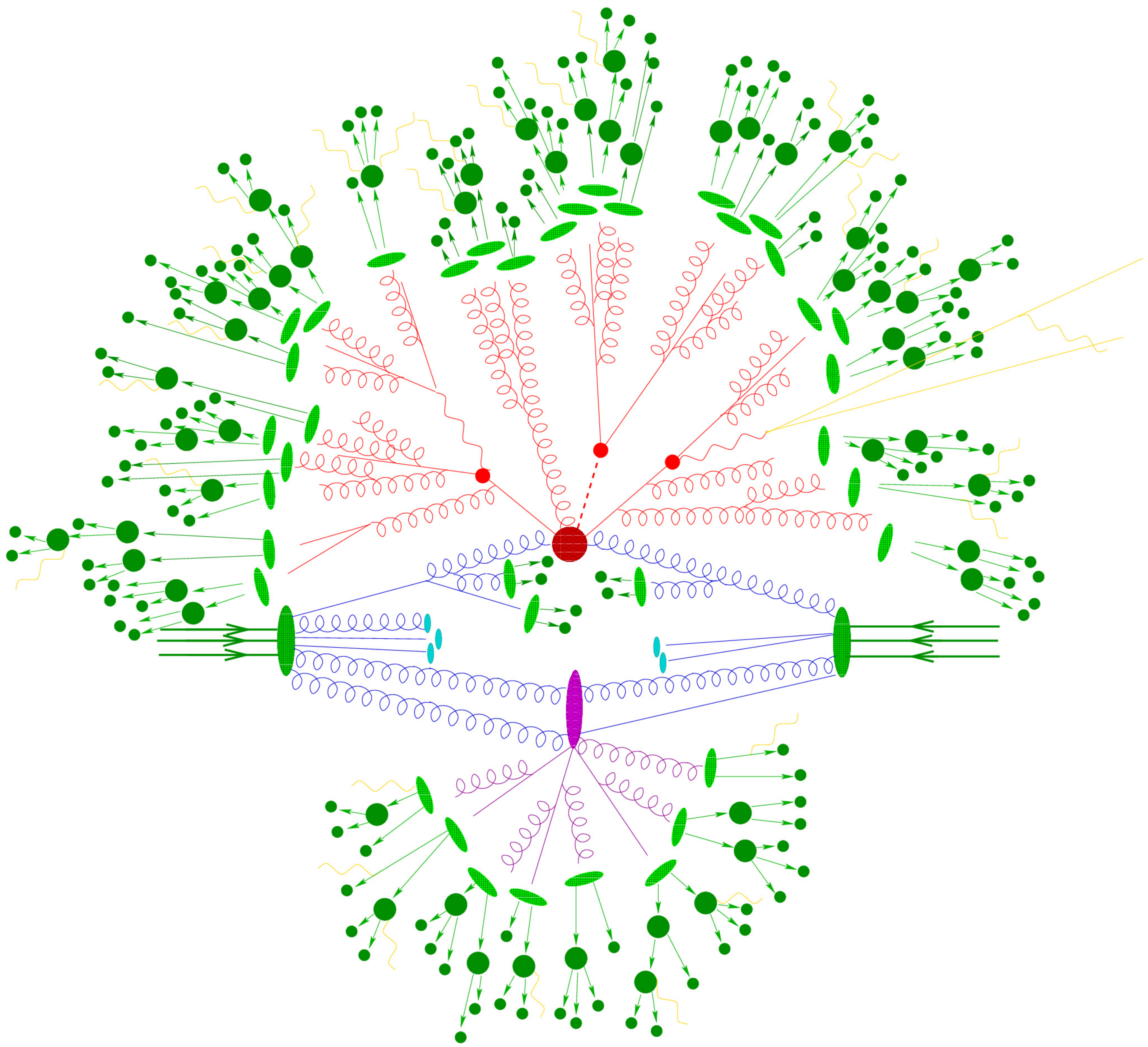


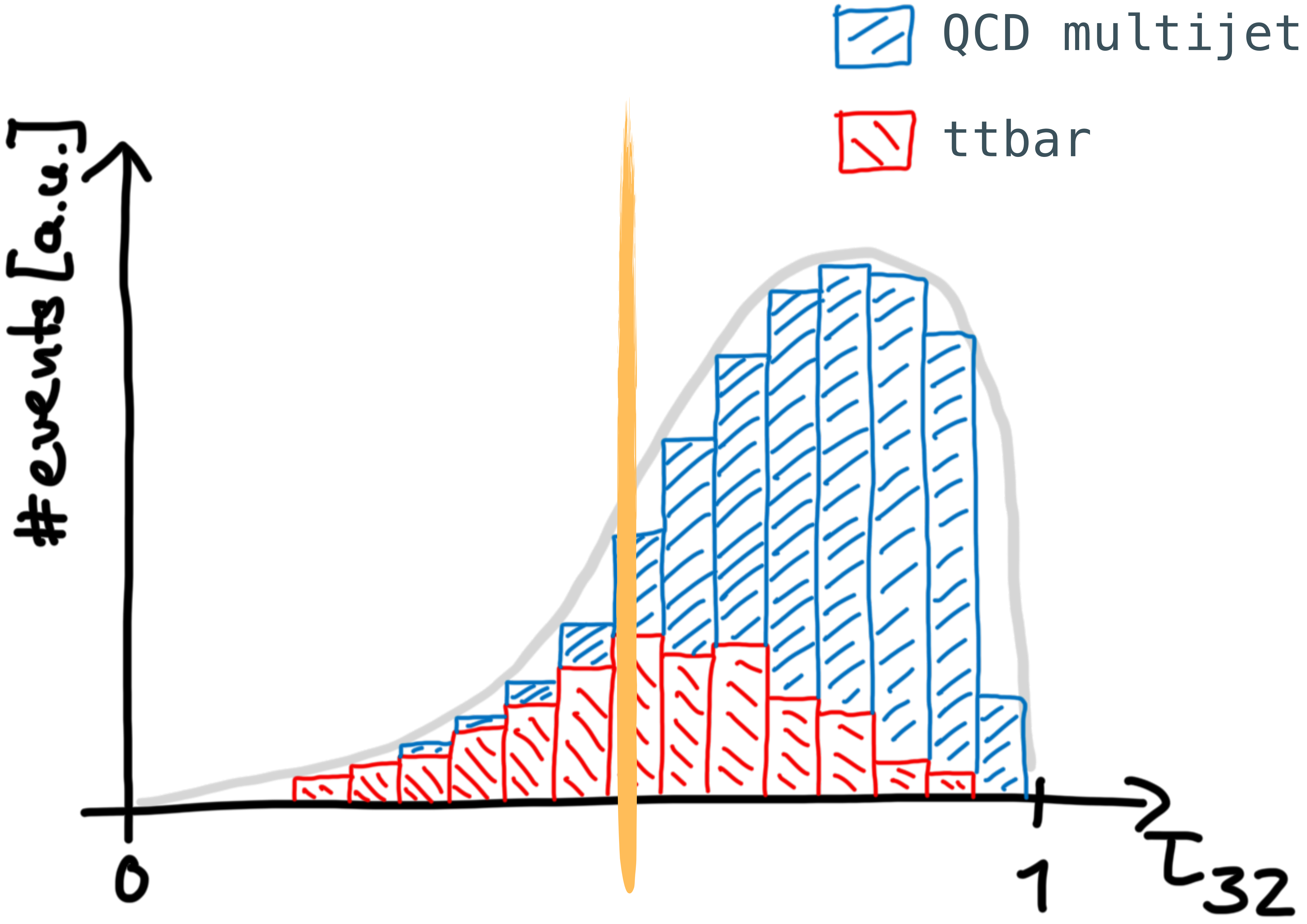
Working Points in One Slide



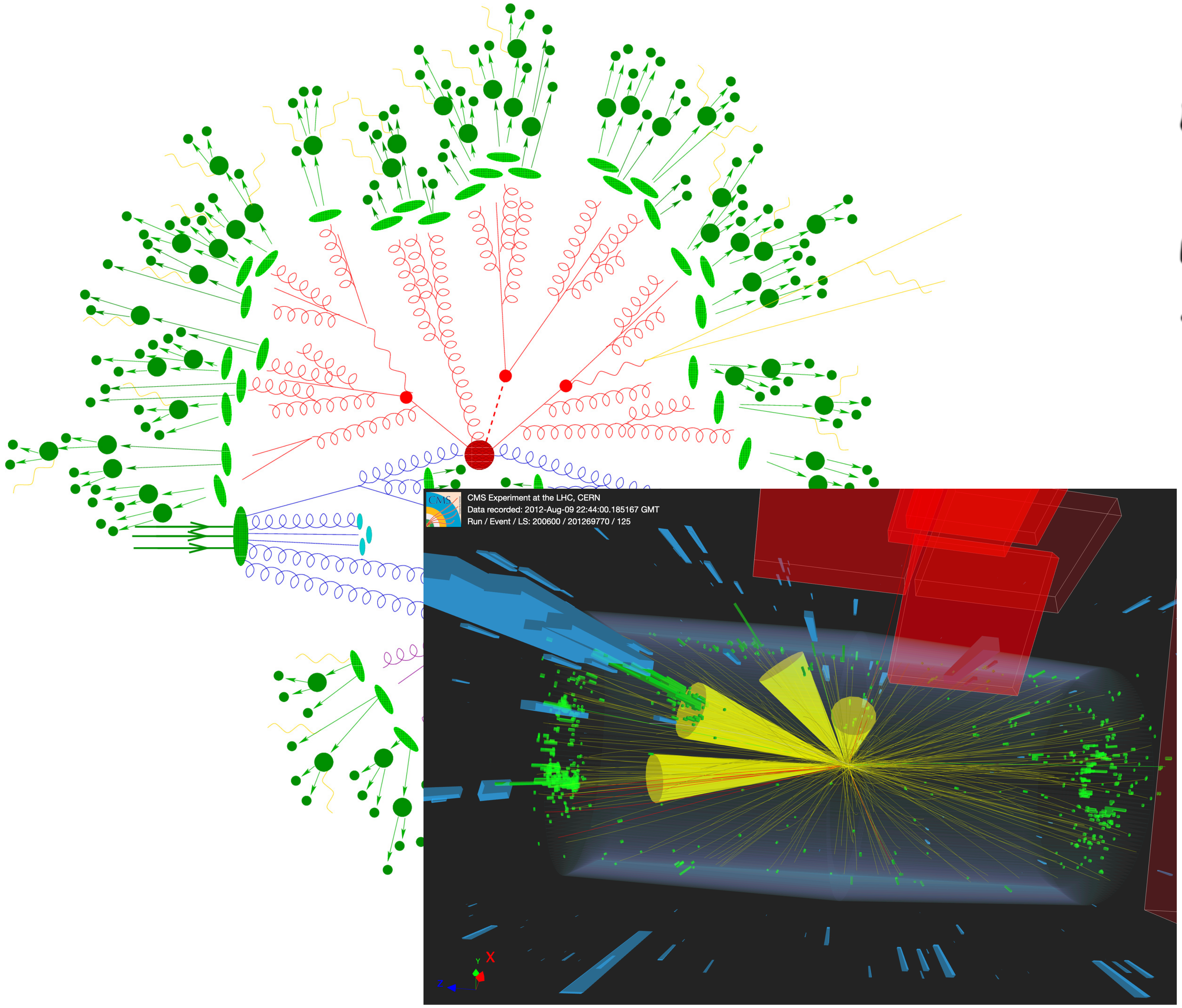


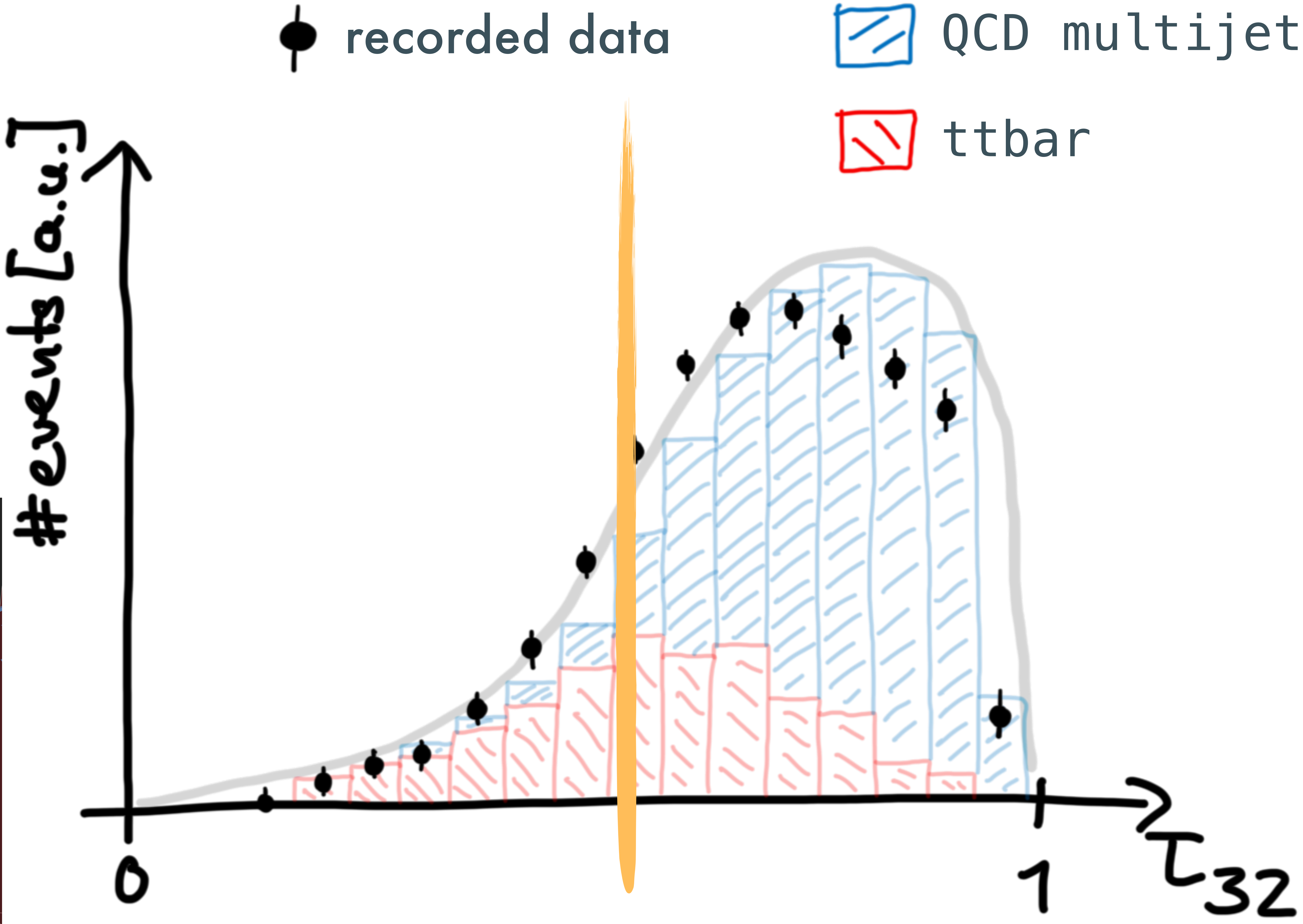
Jet Substructure in Recorded Data





Jet Substructure in Recorded Data

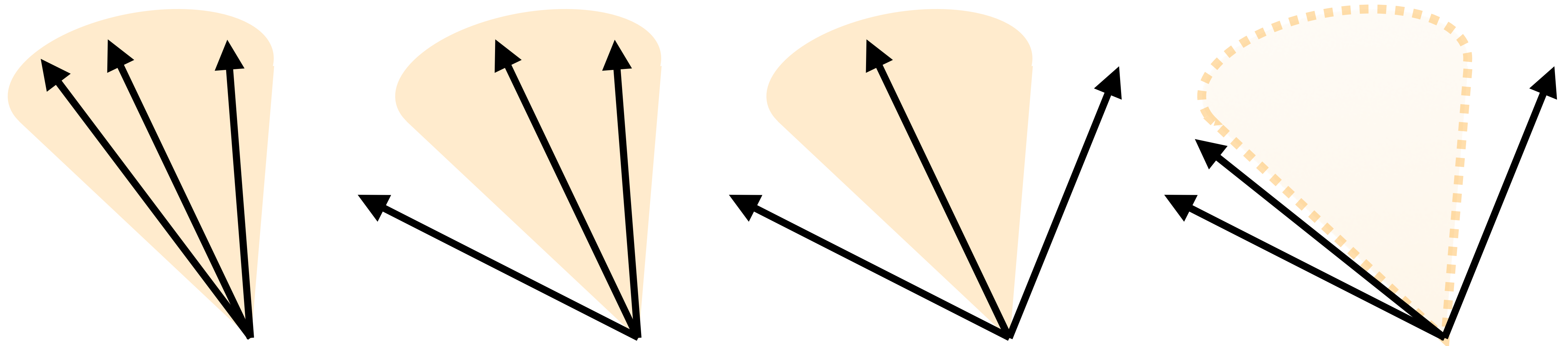




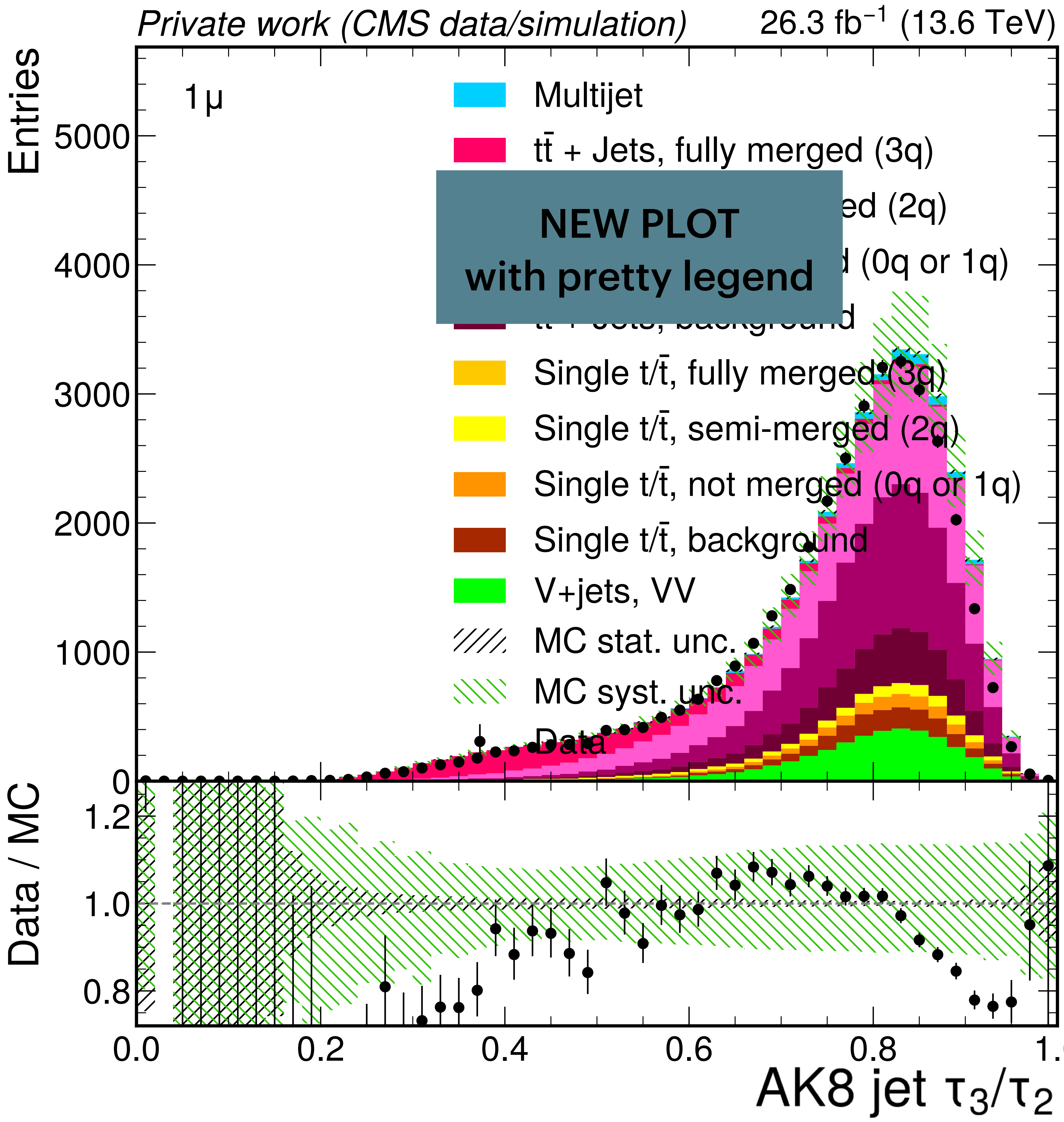
Scale Factors

— to correct MC simulation —

- derive correction factors for each
  - data taking era
  - working point
  - p_T bin of large-radius jet (AK8 jet)
  - merge scenario of top quark decay products

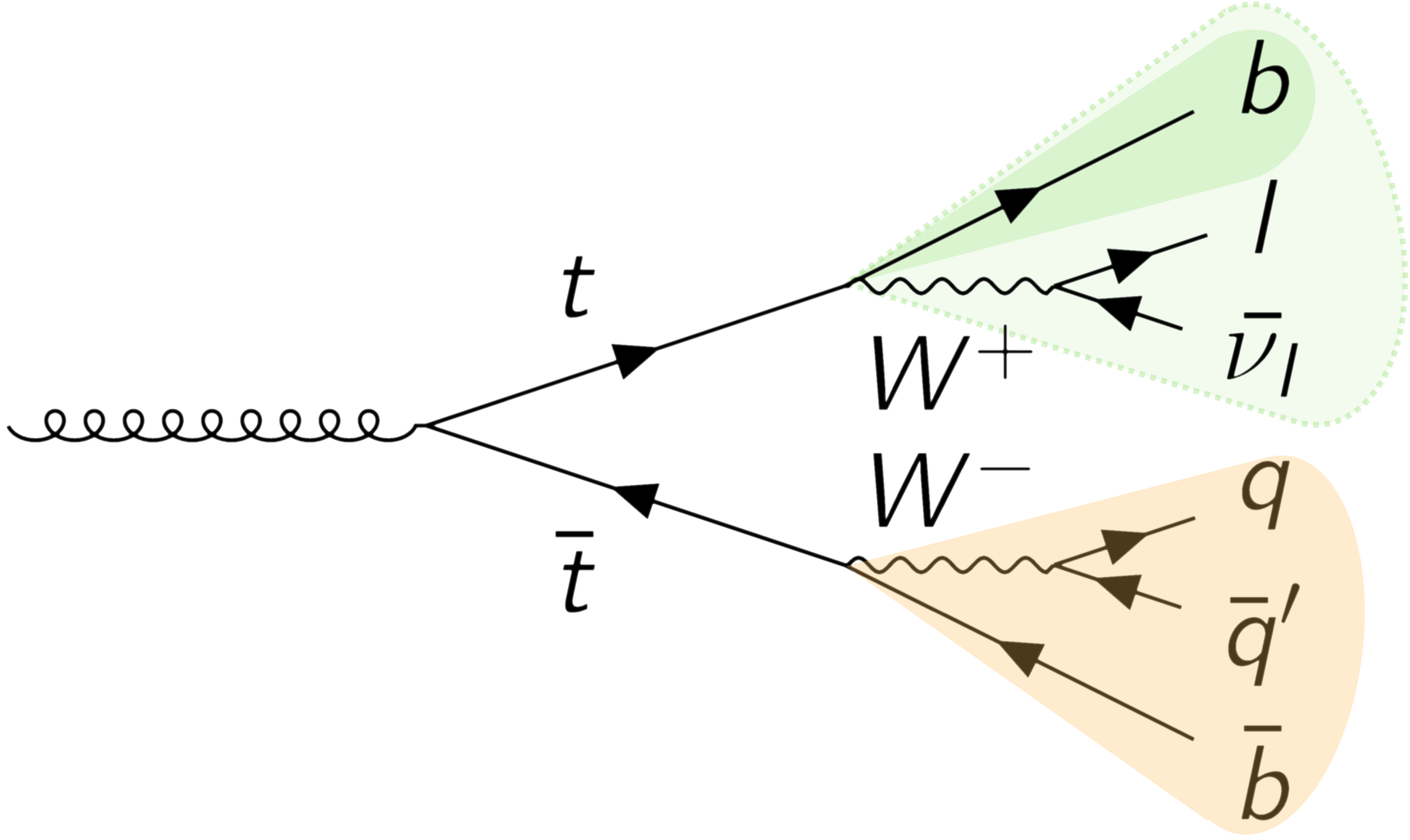


3q   2q   0o1q   background

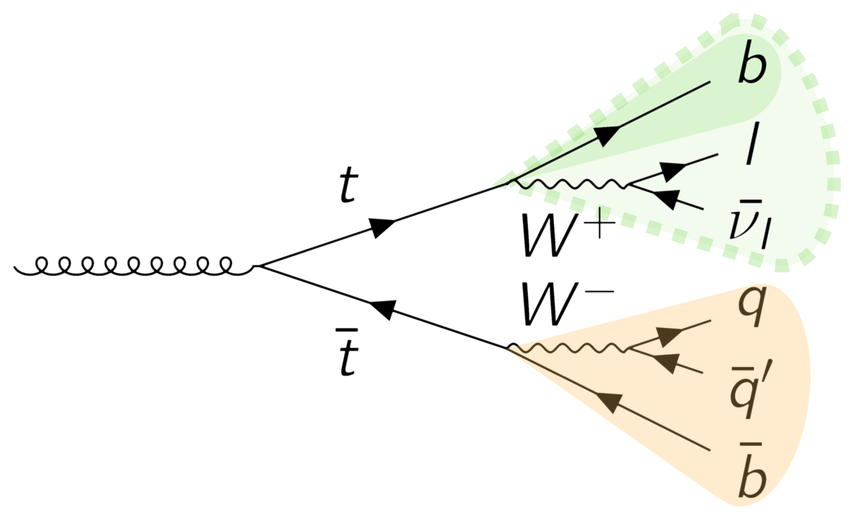


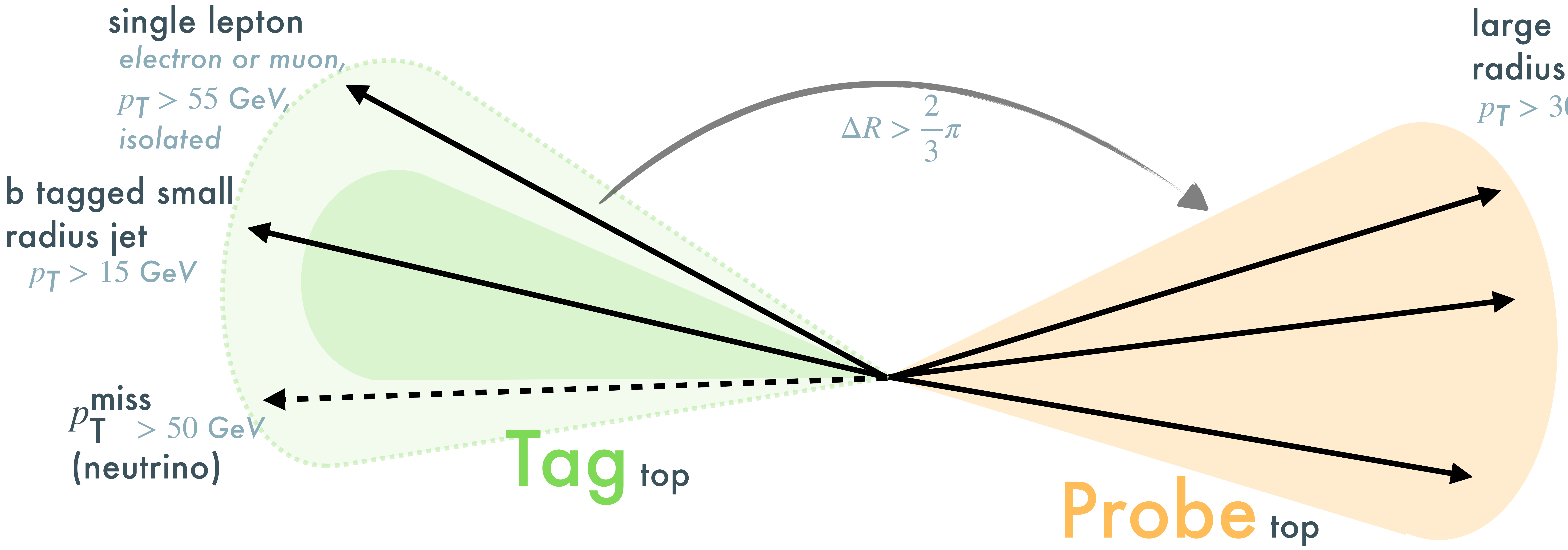
NEW PLOT with pretty legend

Tag and Probe Method With $t\bar{t}$ Pairs

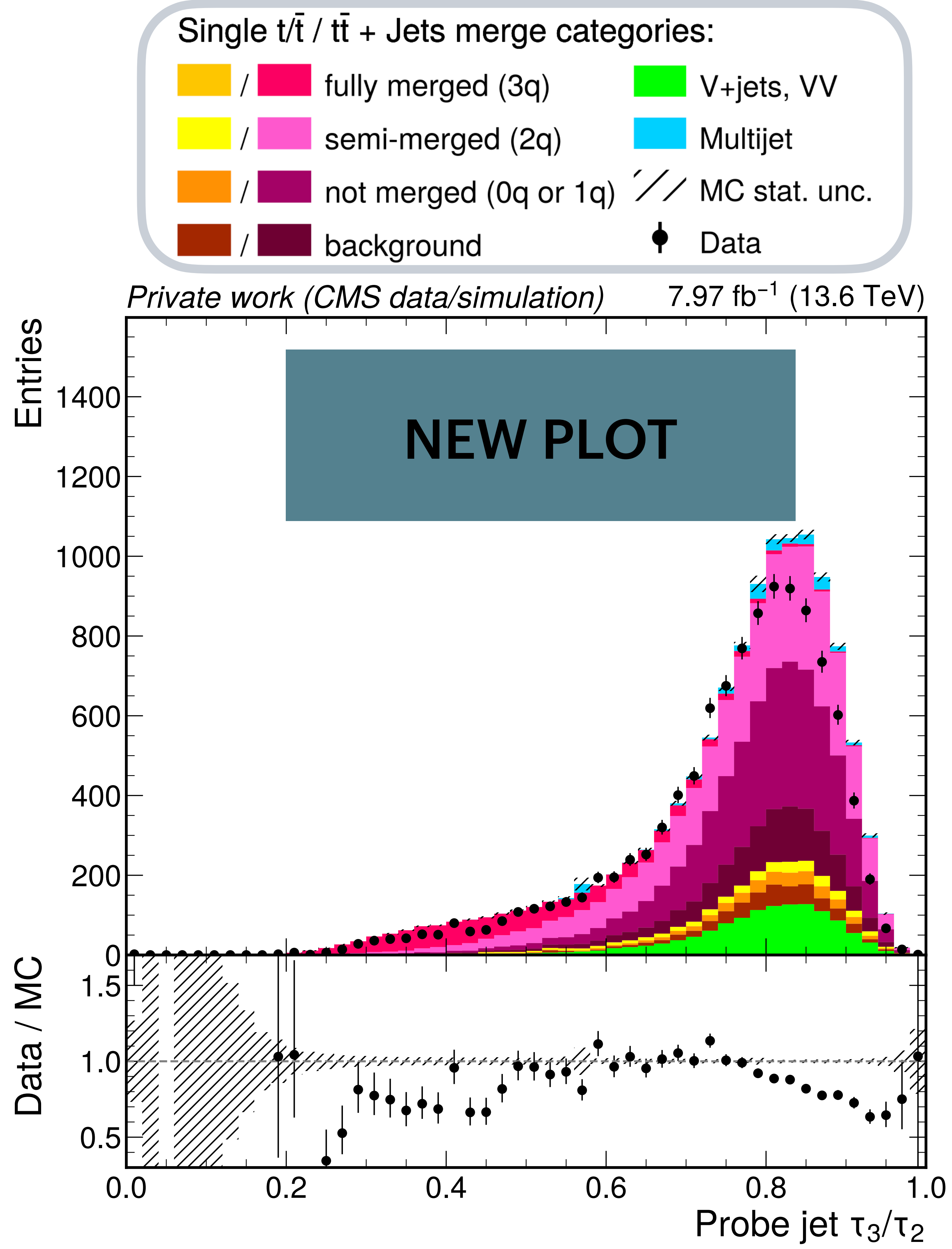


Tag and Probe Method With $t\bar{t}$ Pairs



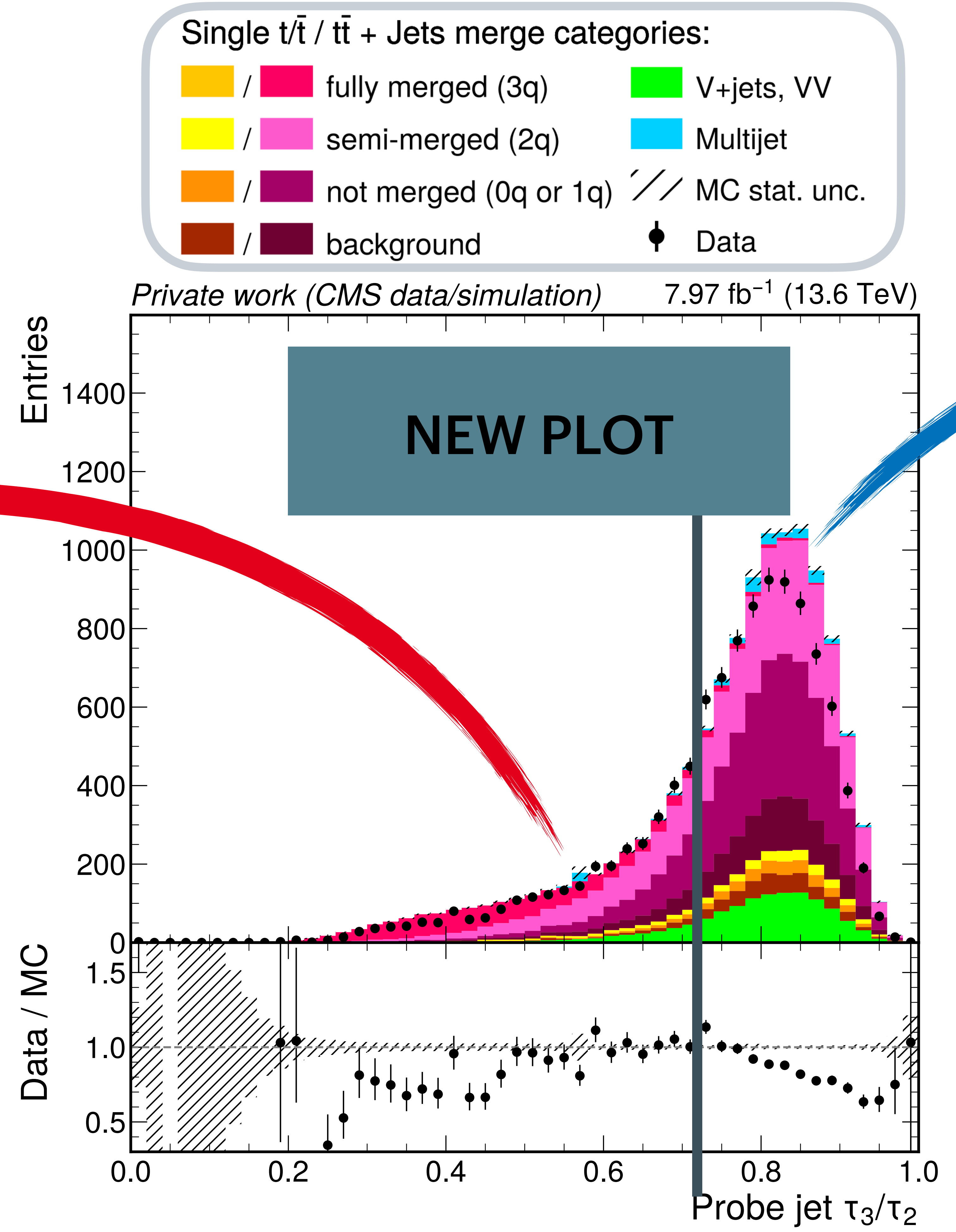


Probe Jet: τ_{32}



Probe Jet: τ_{32}

passing
τ_{32} cut



failing
τ_{32} cut

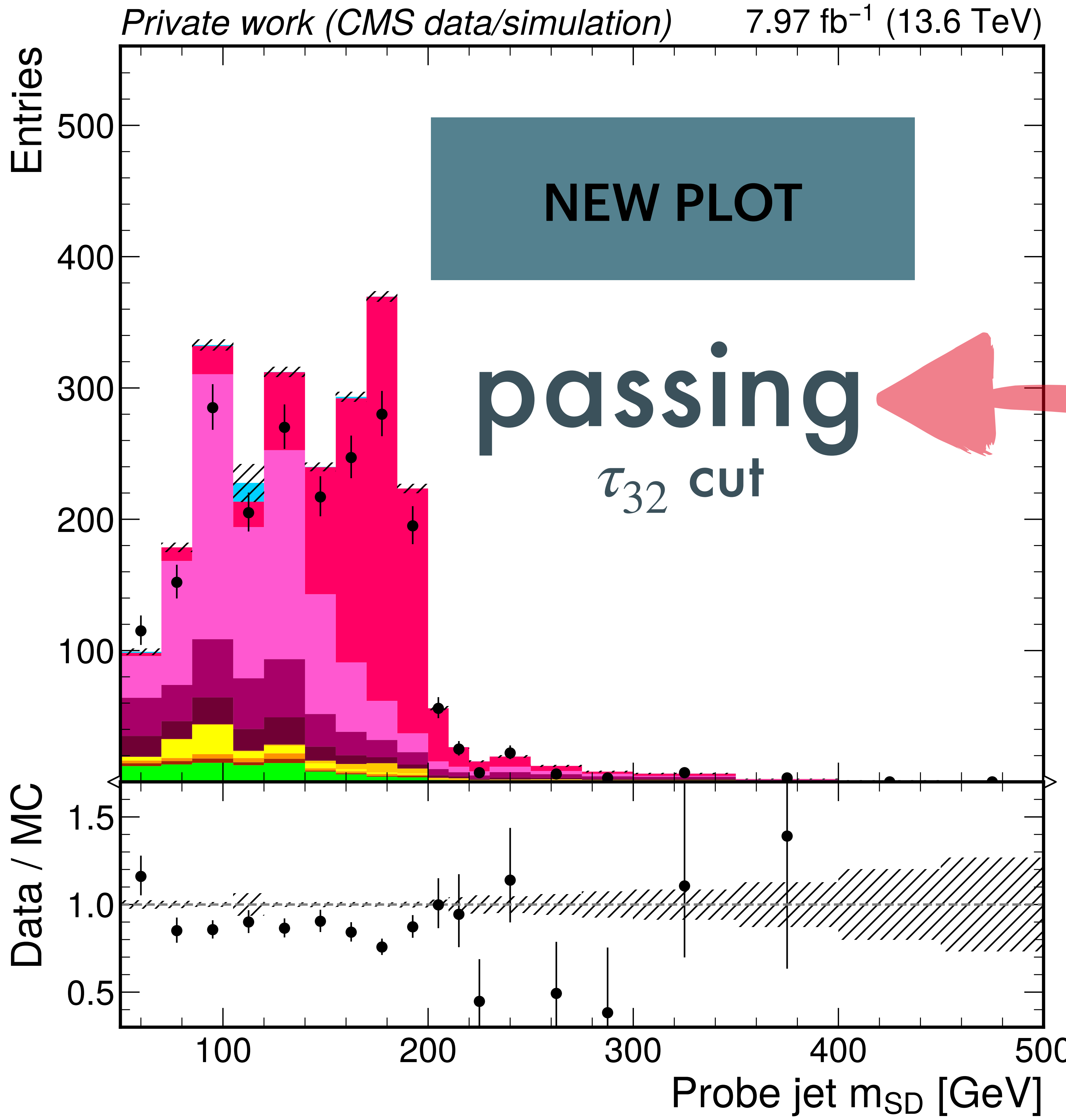
Probe Jet: τ_{32} and m_{SD}

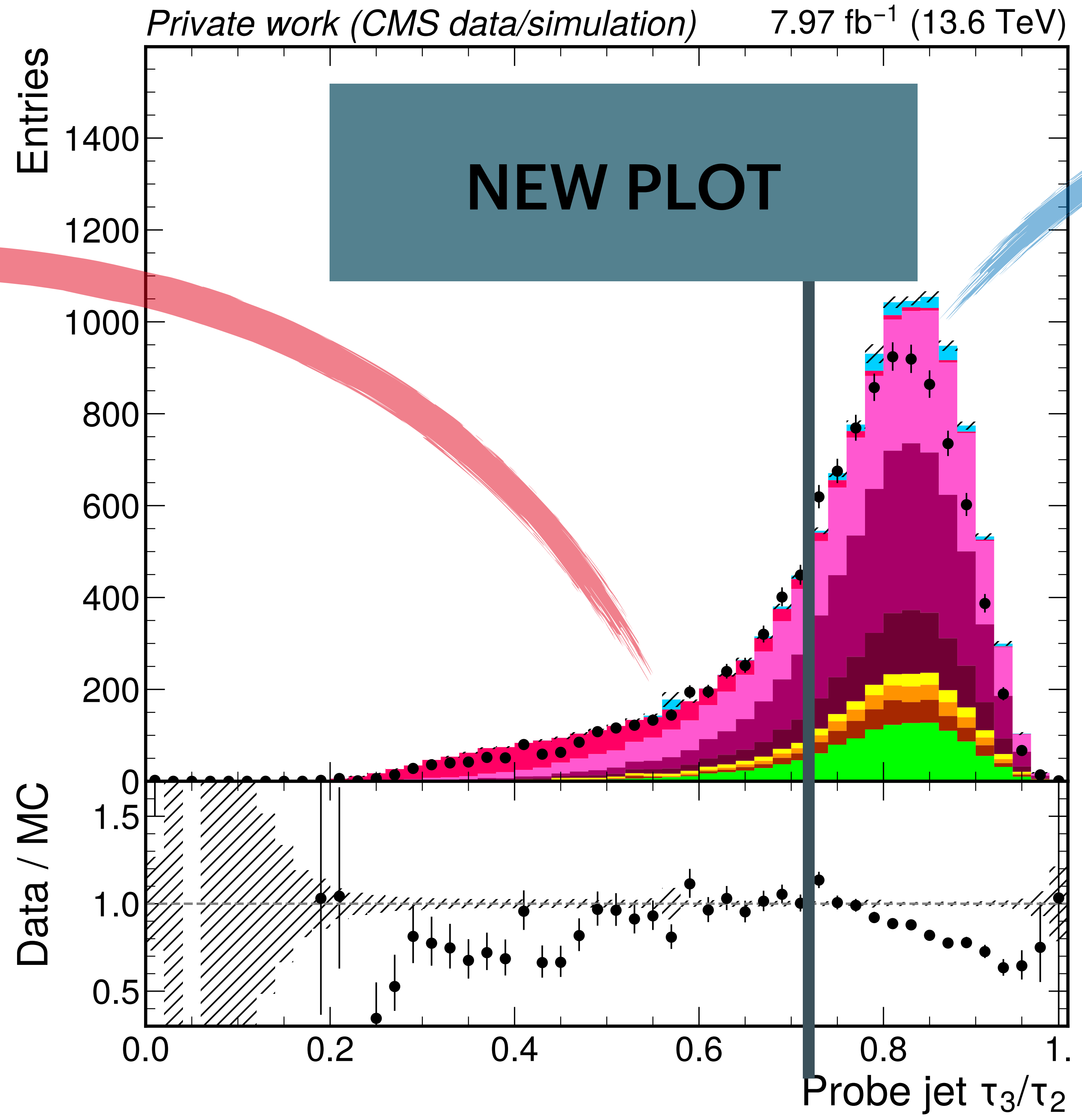
soft drop criterion $\frac{\min(p_{T1}, p_{T,2})}{p_{T1} + p_{T2}} > z_{\mathrm{cut}} \cdot \left(\frac{\Delta R_{12}}{R_0}\right)^{\beta}$

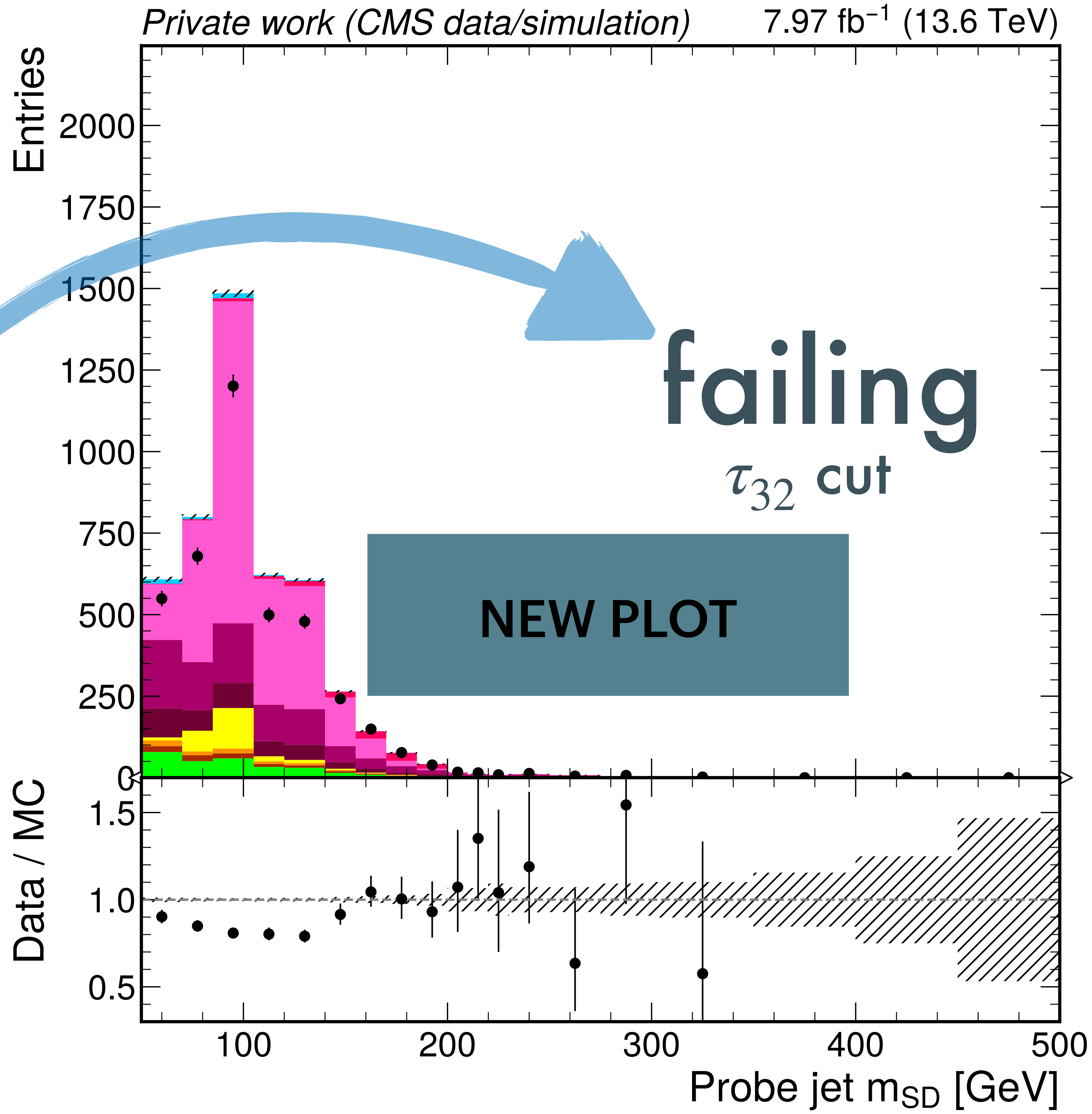
JHEP 1405 (2014) 146 - arXiv:1402.2657

Single t/$\bar{t}$ / t$\bar{t}$ + Jets merge categories:

- / fully merged (3q)
- / semi-merged (2q)
- / not merged (0q or 1q)
- / background
- V+jets, VV
- Multijet
- MC stat. unc.
- Data



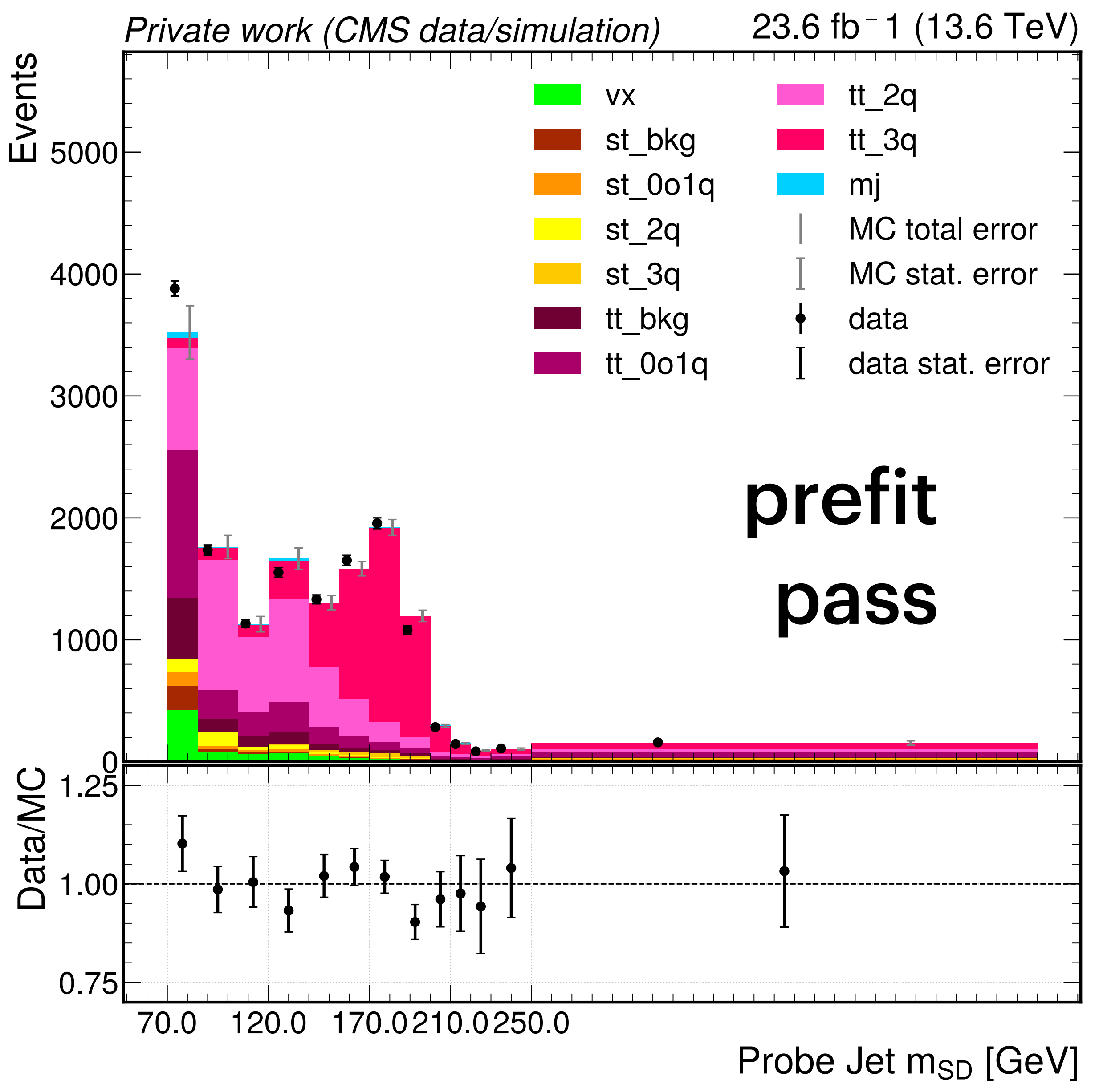


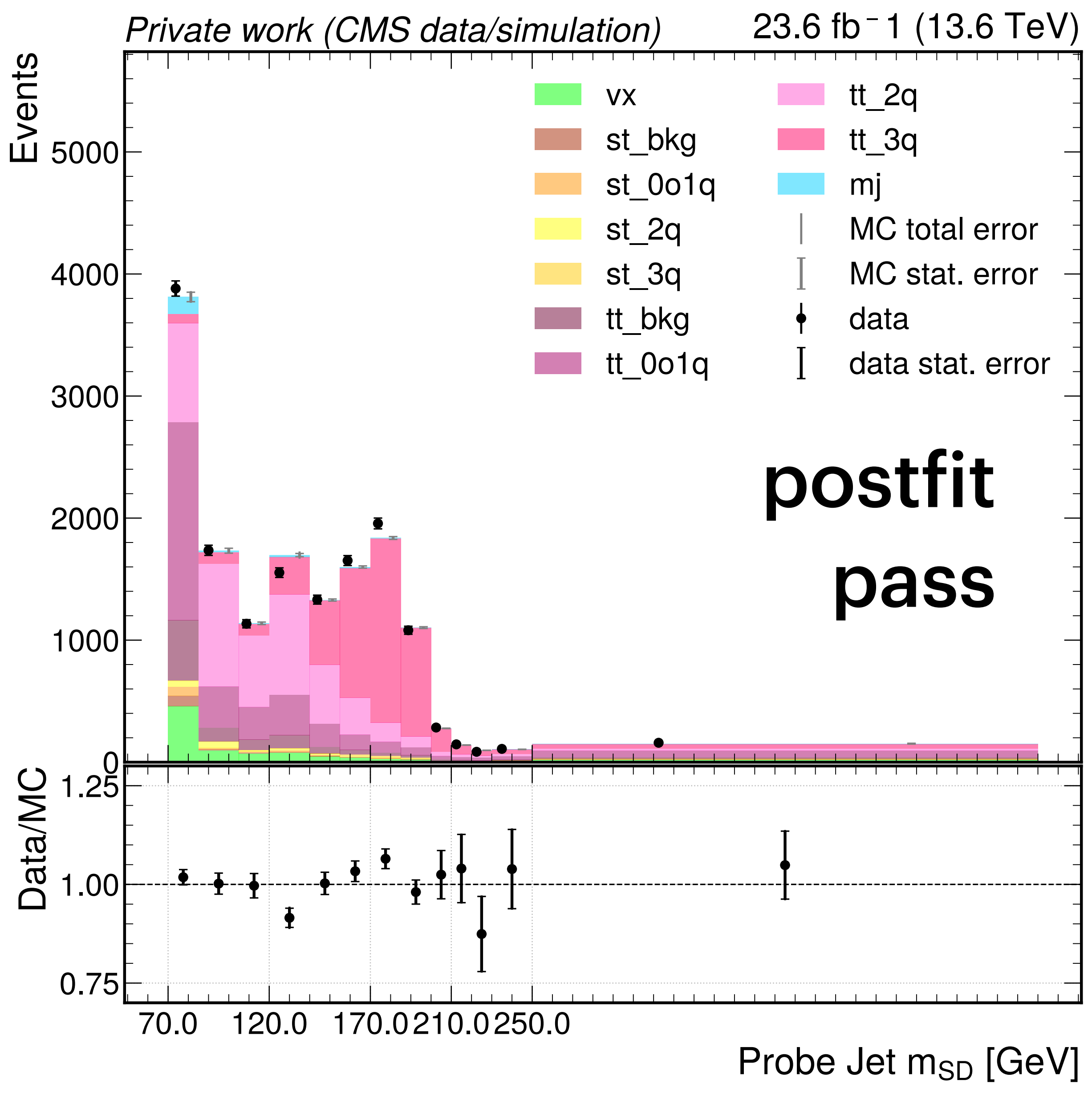


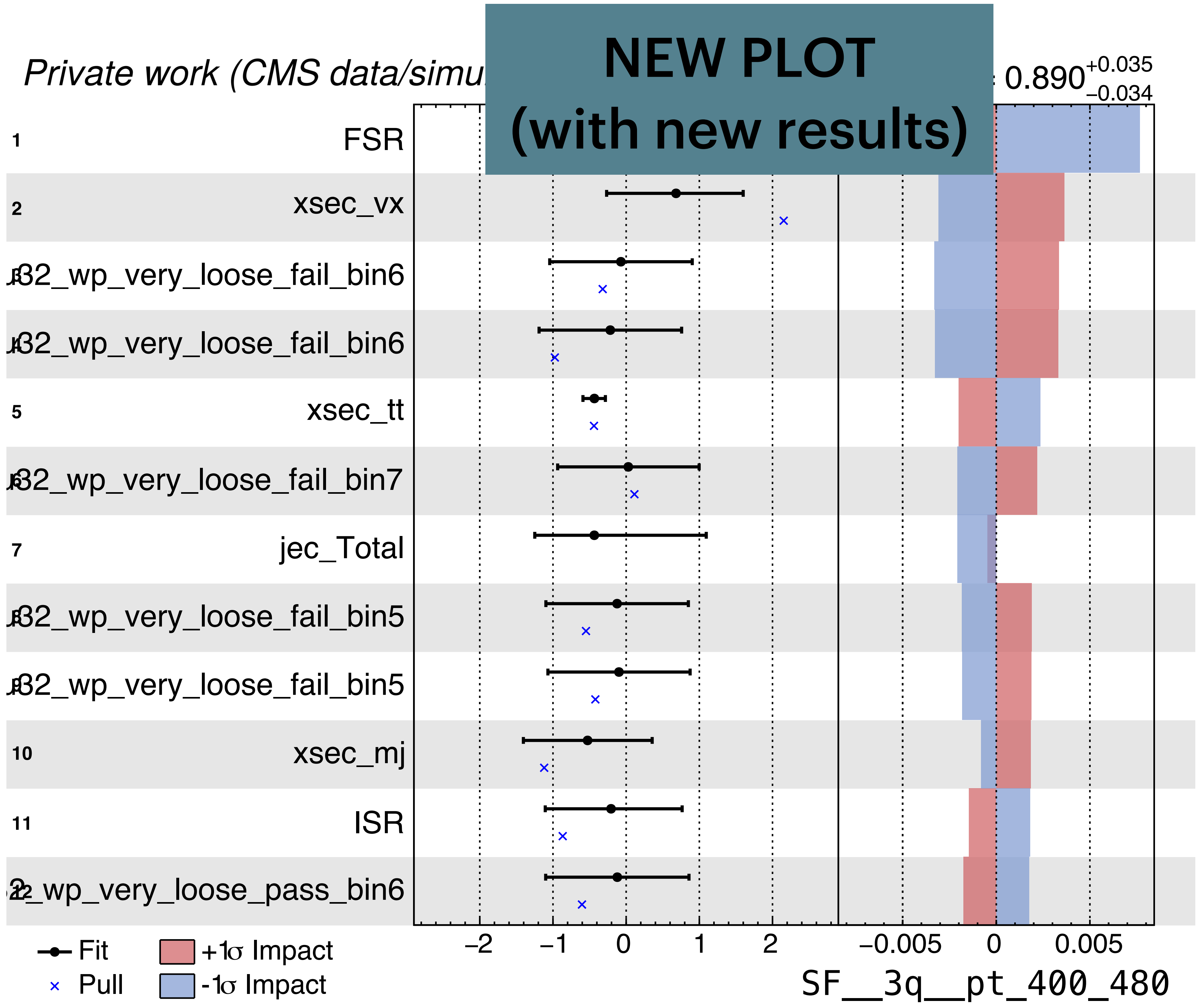
SF Derivation

— simultaneous fit for merge categories and p_T of probe jet —

- constraint: $N_{\text{total}} = N_{\text{pass}} + N_{\text{fail}} = \text{const}$
- syst. uncertainties considered, FSR largest impact





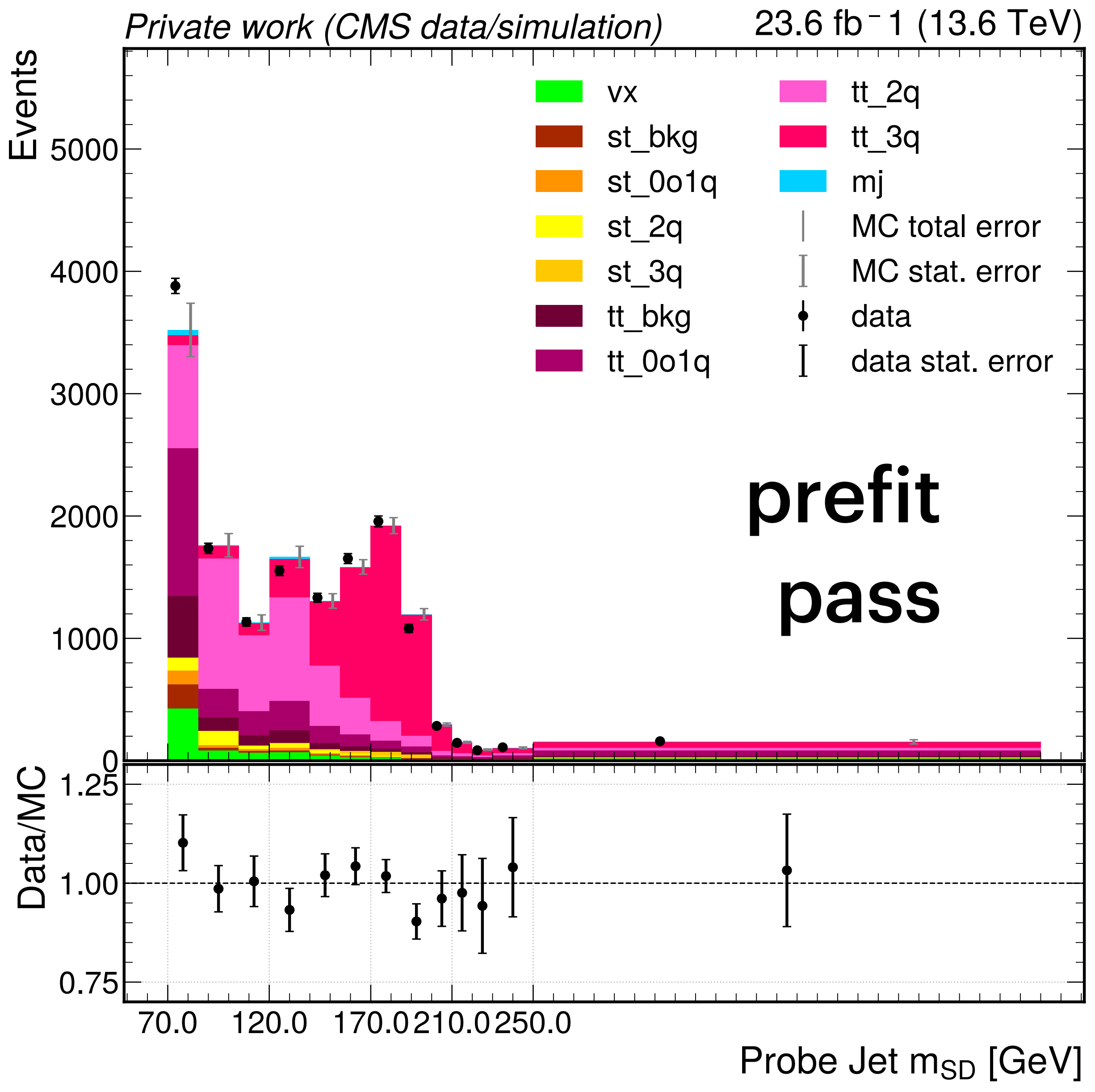


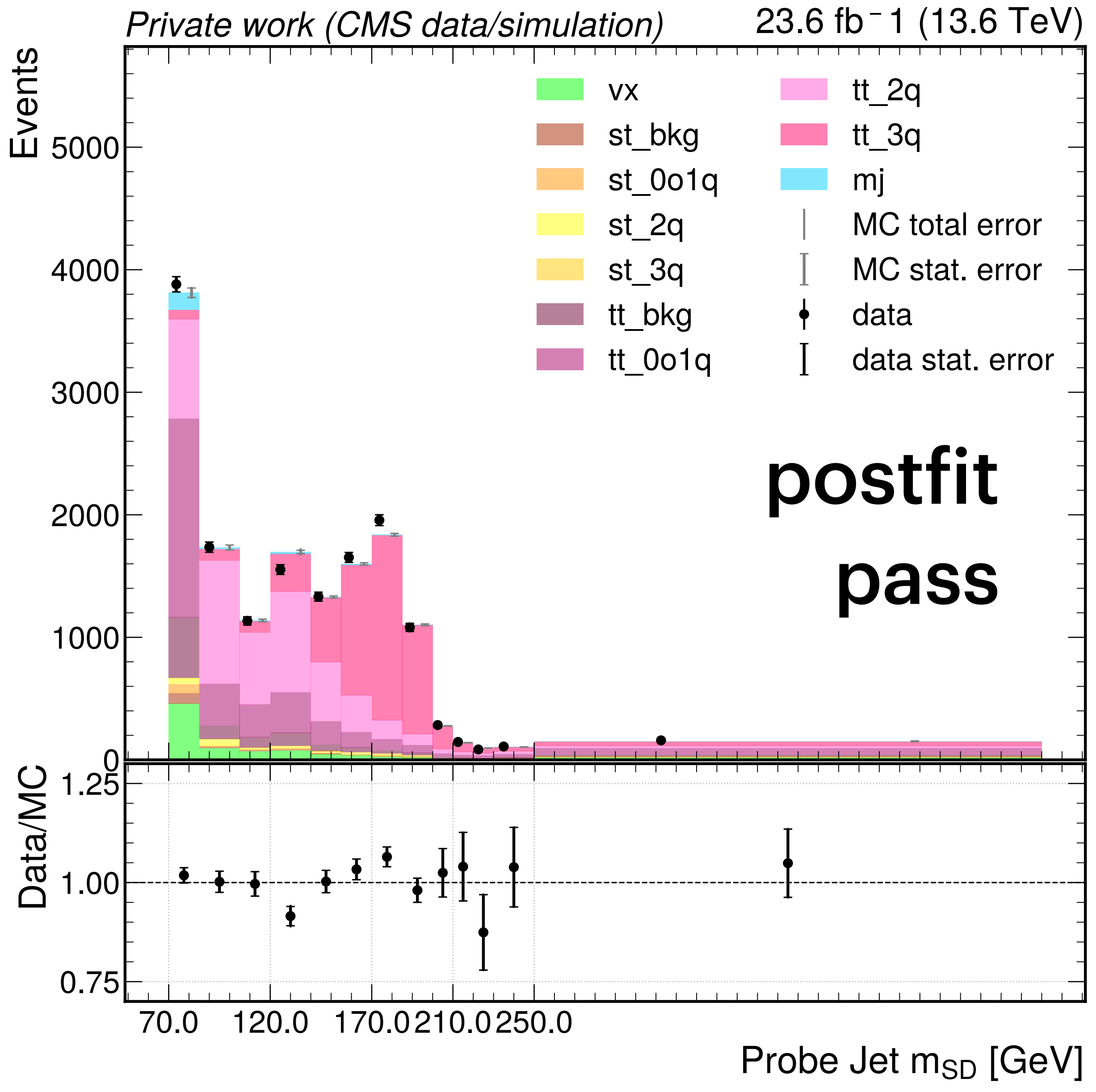
SF Derivation

— simultaneous fit for merge categories and p_T of probe jet —

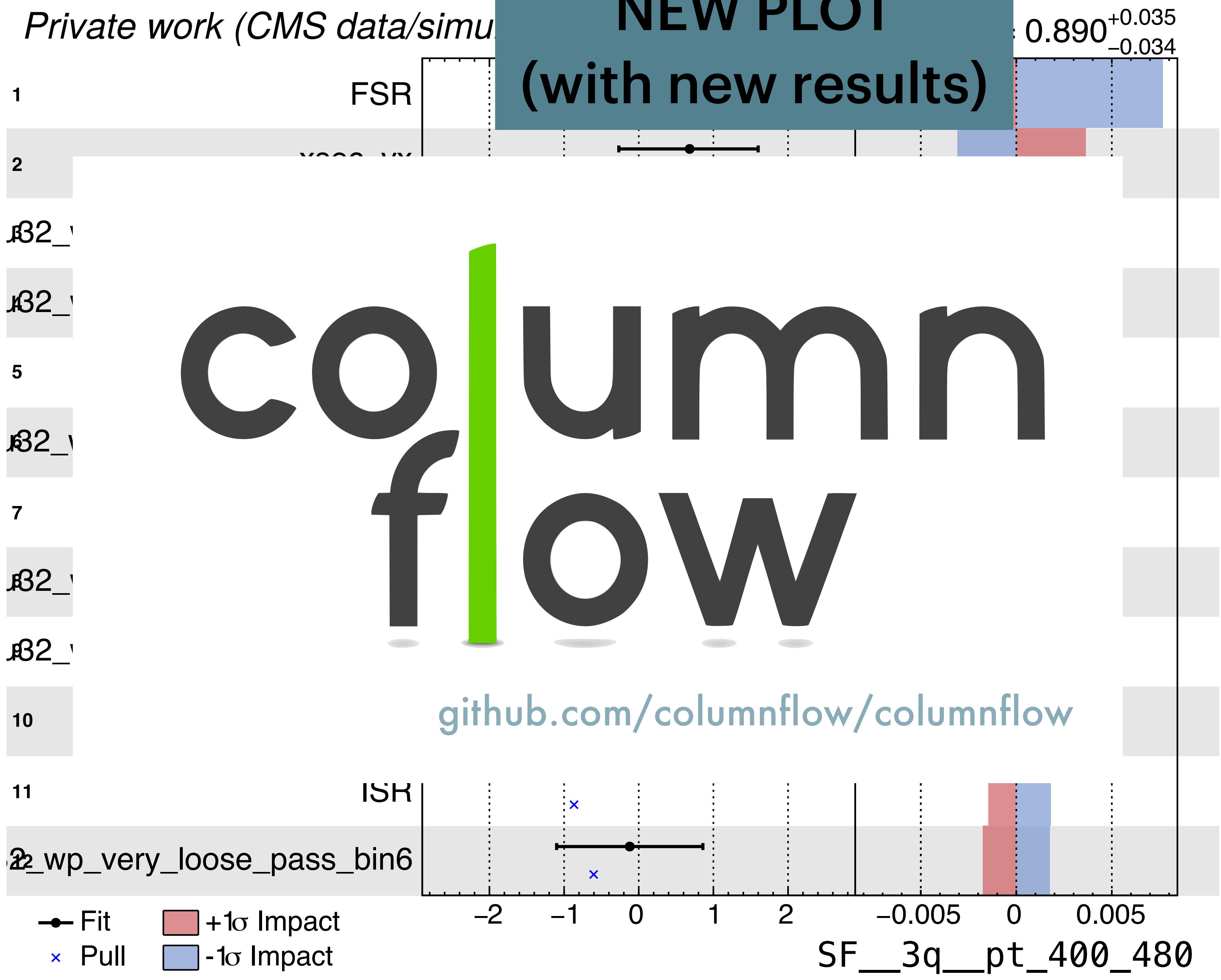
- constraint: $N_{\text{total}} = N_{\text{pass}} + N_{\text{fail}} = \text{const}$
- syst. uncertainties considered, FSR largest impact

Quick and automated derivation of SF essential to be used in analyses.



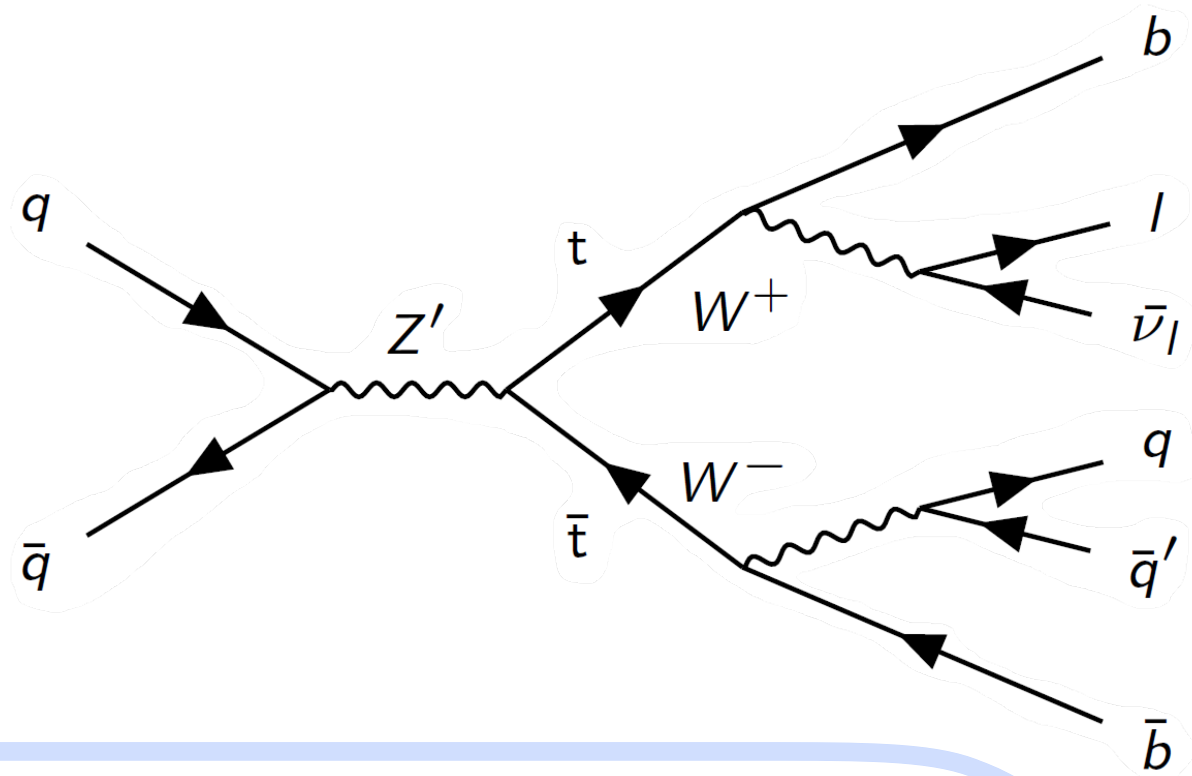


NEW PLOT (with new results)



Analysis Strategy

— searching for Z' bosons —



I. Selection

$t\bar{t}$ events with lepton+jets final state objects

II. Reconstruction

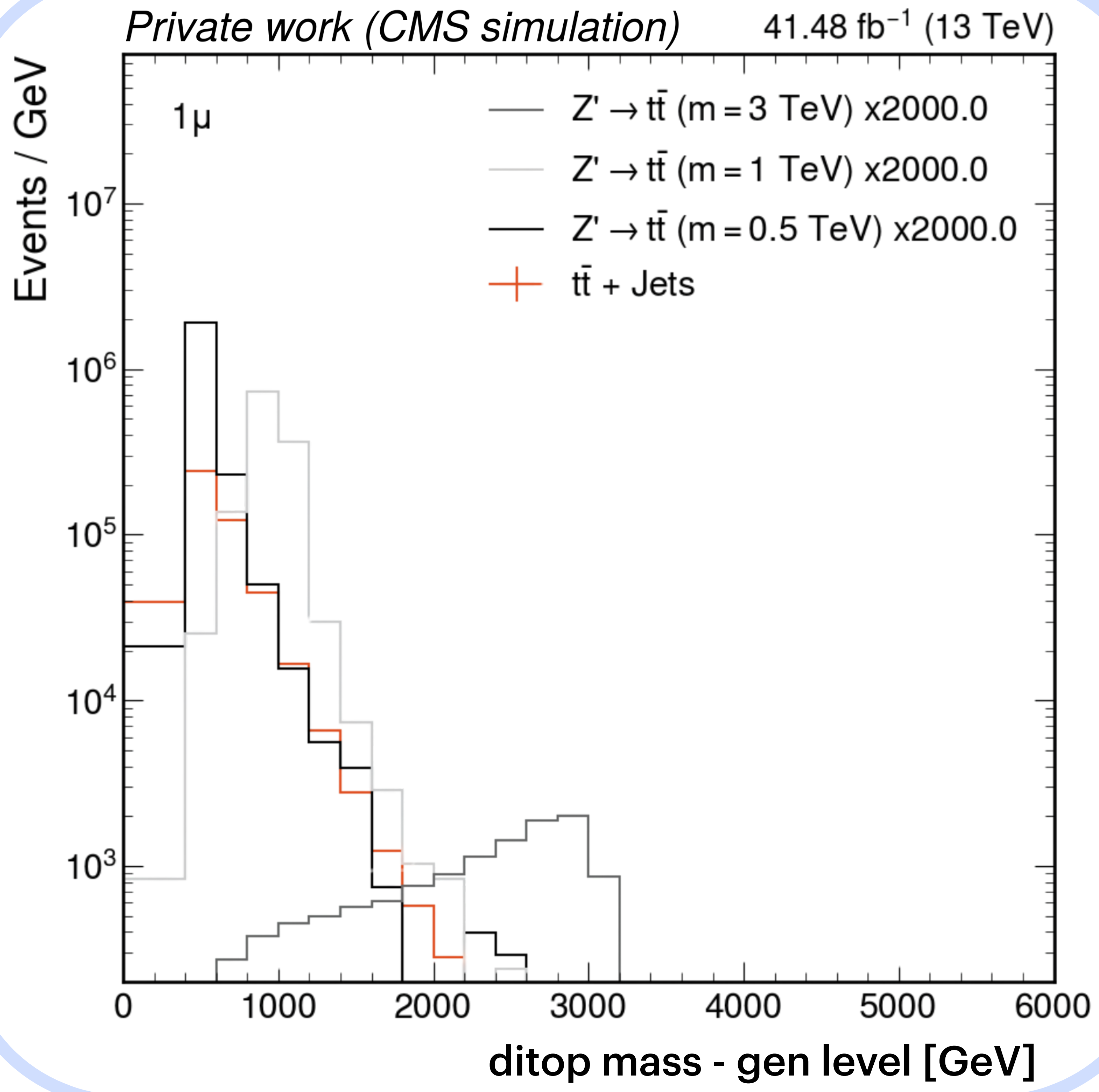
kinematics of the $t\bar{t}$ system

III. Event Classification

using machine learning

IV. Statistical Analysis

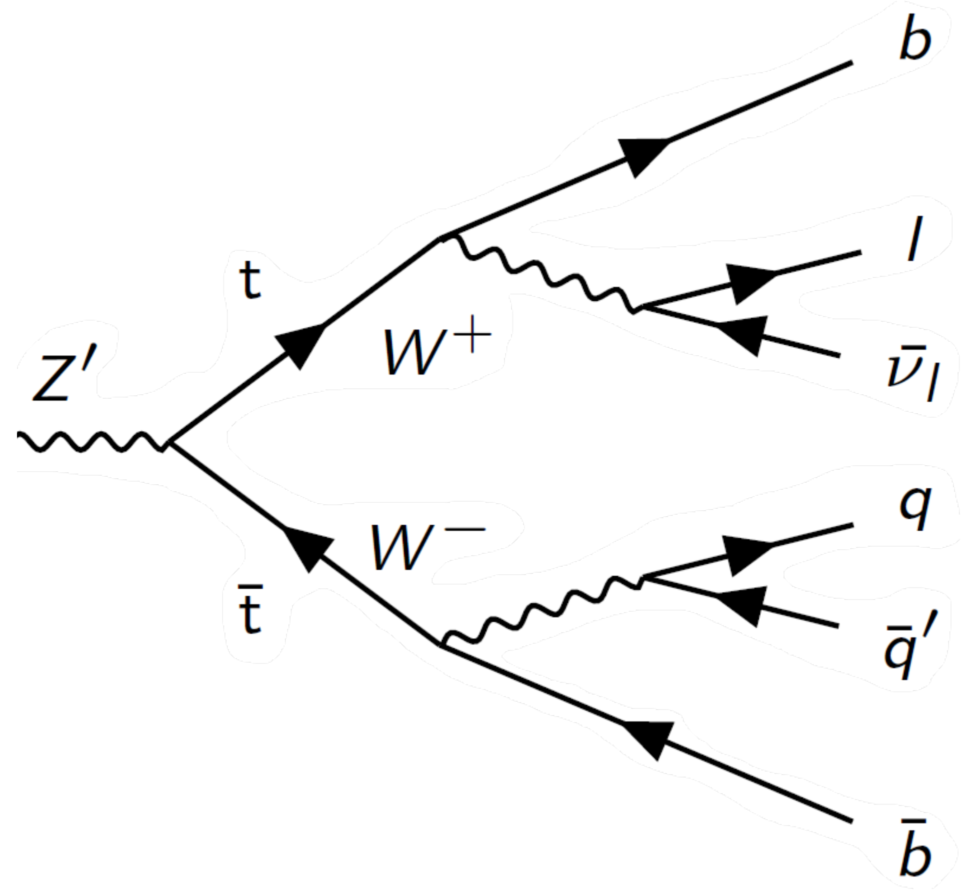
bump hunt or set limits



Search for New Physics Effects ...

calibrate events/objects

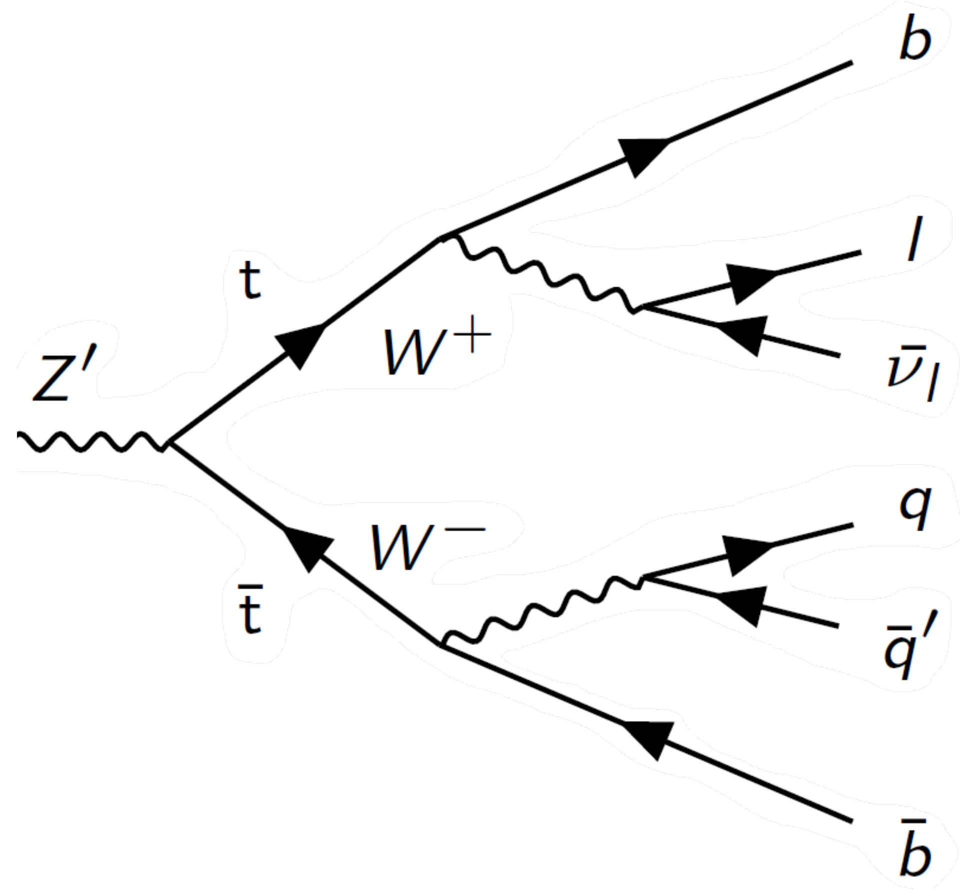
select events



Search for New Physics Effects ...

calibrate events/objects

select events



AK8 jet present?

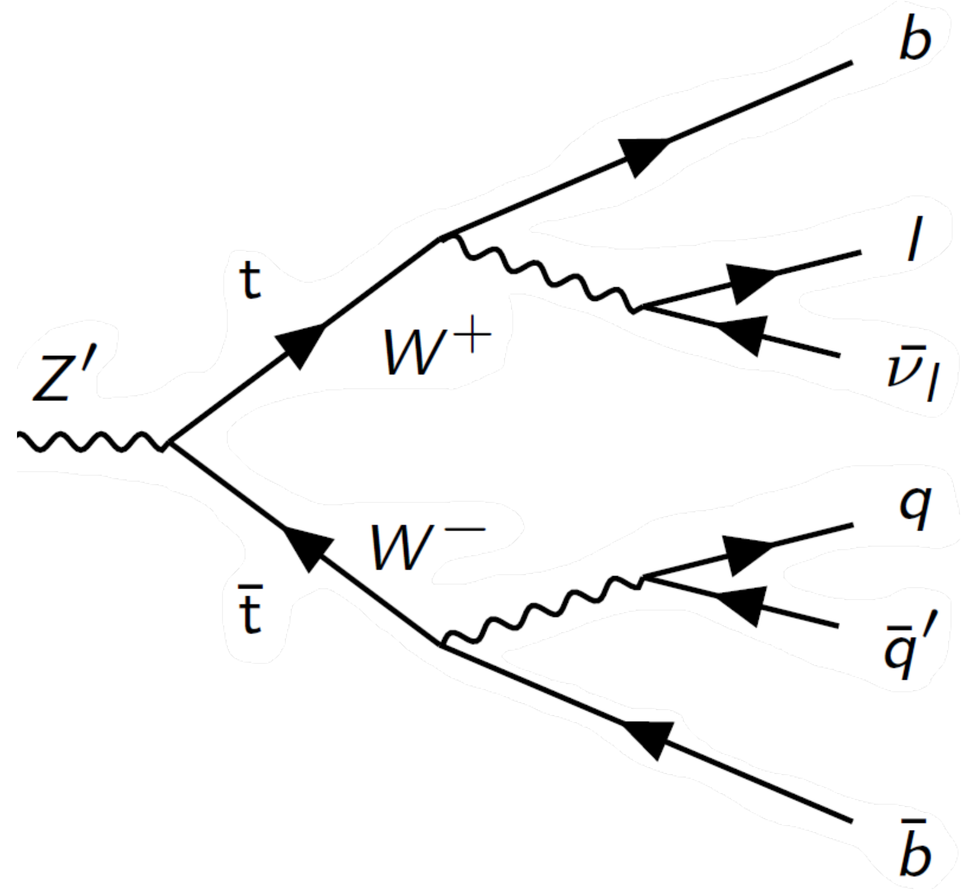
✓

boosted reconstruction

Search for New Physics Effects …

calibrate events/objects

select events



AK8 jet present?

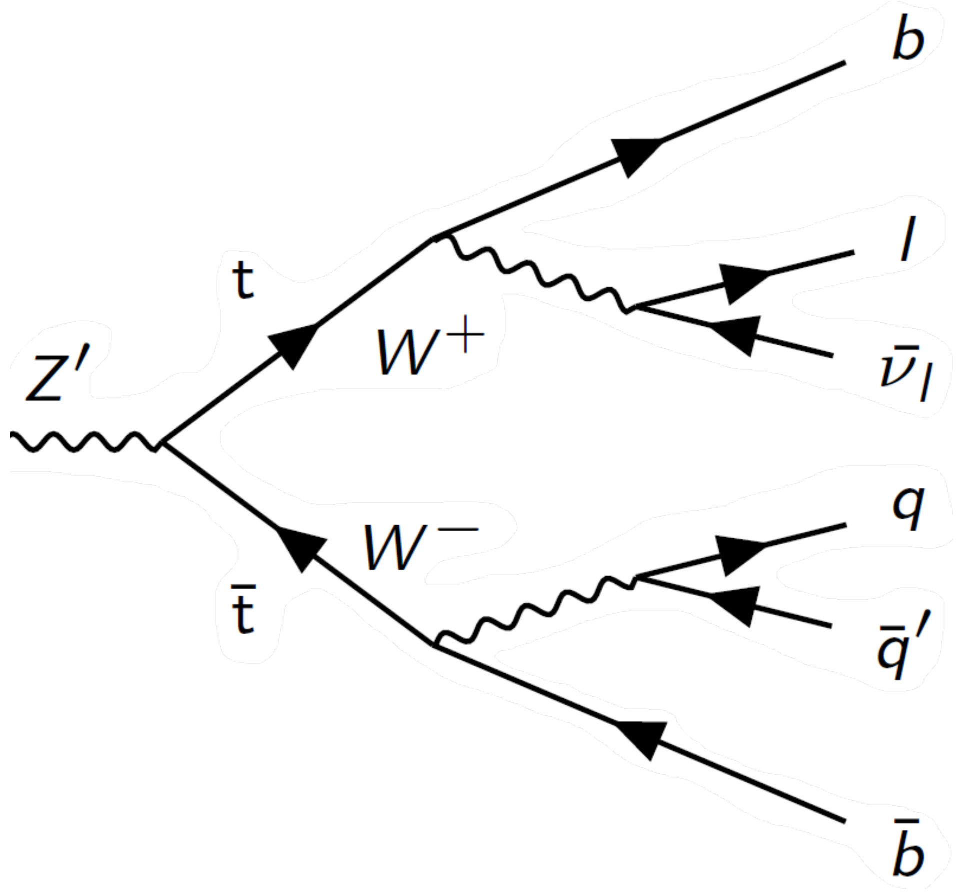
resolved reconstruction

boosted reconstruction

Search for New Physics Effects ...

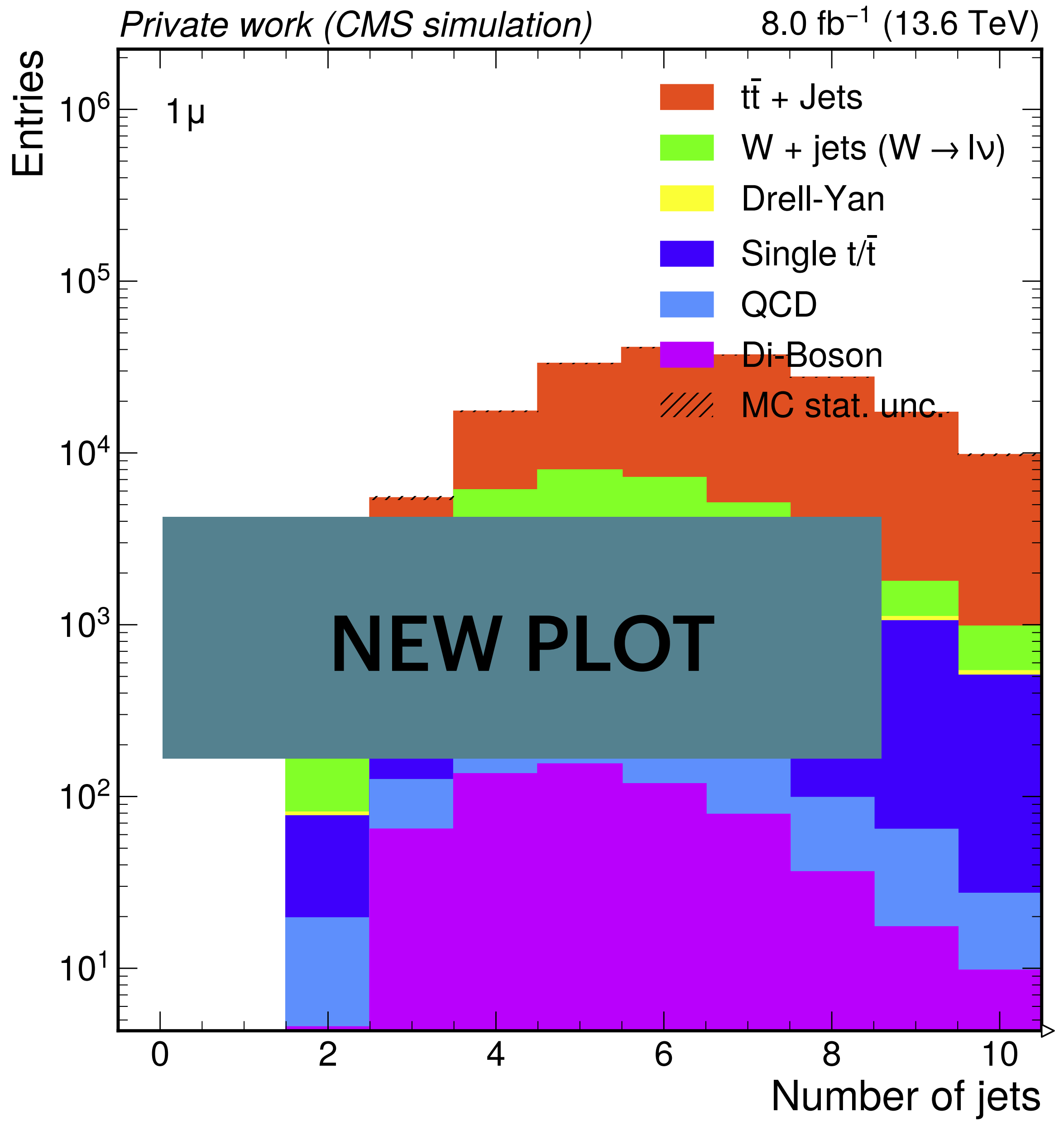
calibrate events/objects

select events



AK8 jet present?

resolved reconstruction

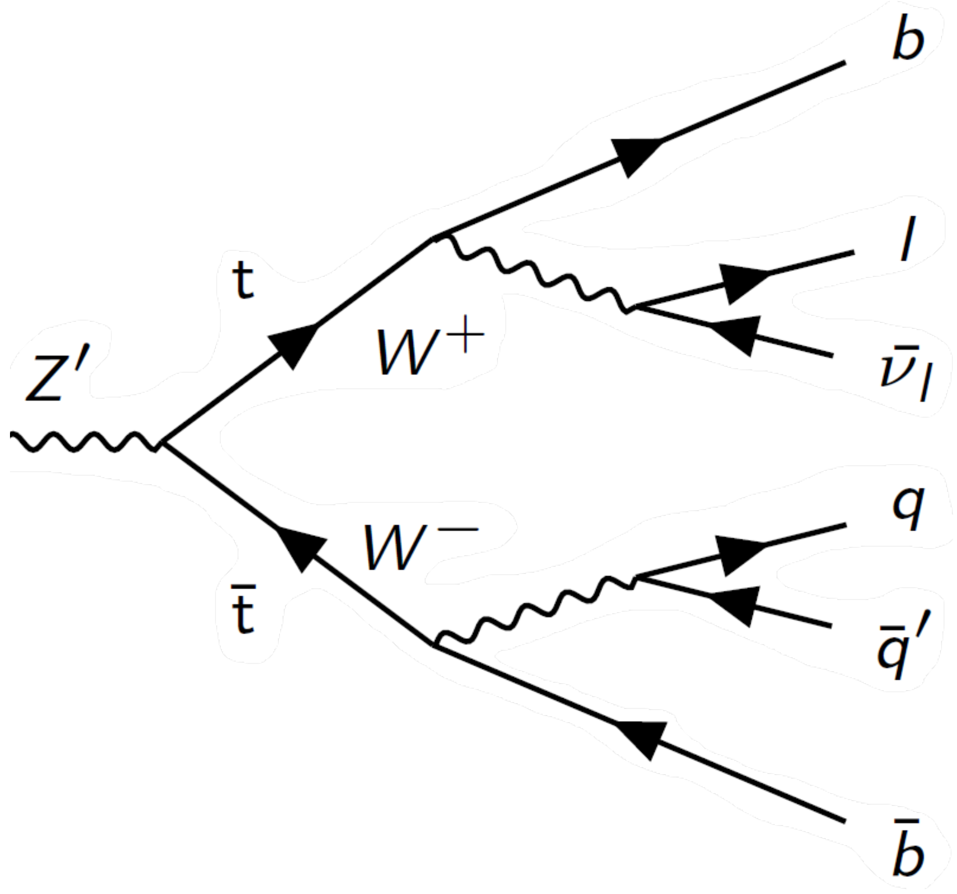


boosted reconstruction

Search for New Physics Effects …

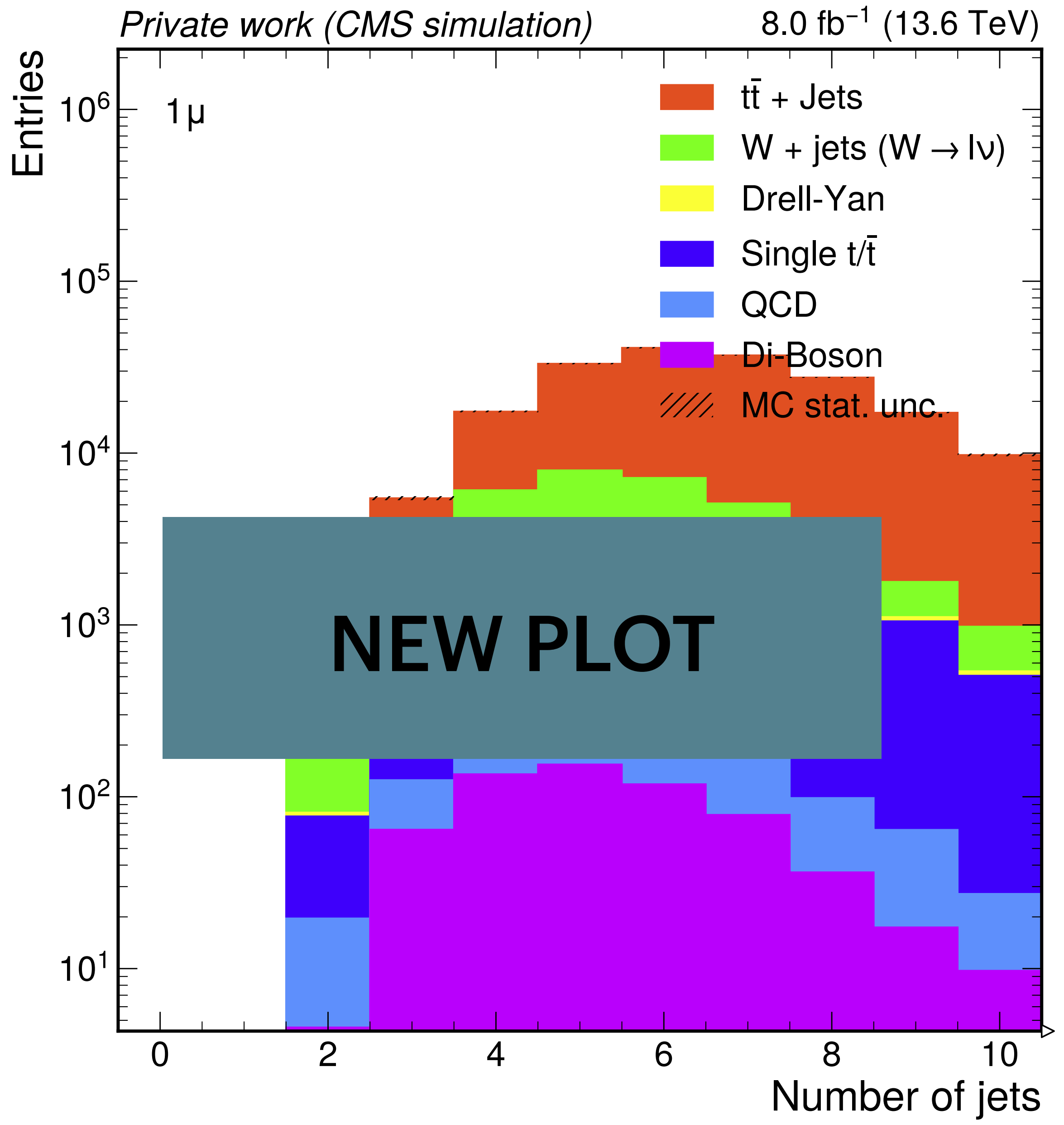
calibrate events/objects

select events



AK8 jet present?

resolved reconstruction



boosted reconstruction

choose best jet combination per event

$$\chi^2 = \left(\frac{m_{\text{lep}} - \bar{m}_{\text{lep}}}{\sigma_{\bar{m}_{\text{lep}}}}\right)^2 + \left(\frac{m_{\text{had}} - \bar{m}_{\text{had}}}{\sigma_{\bar{m}_{\text{had}}}}\right)^2$$

… In the Ditop Mass Spectrum

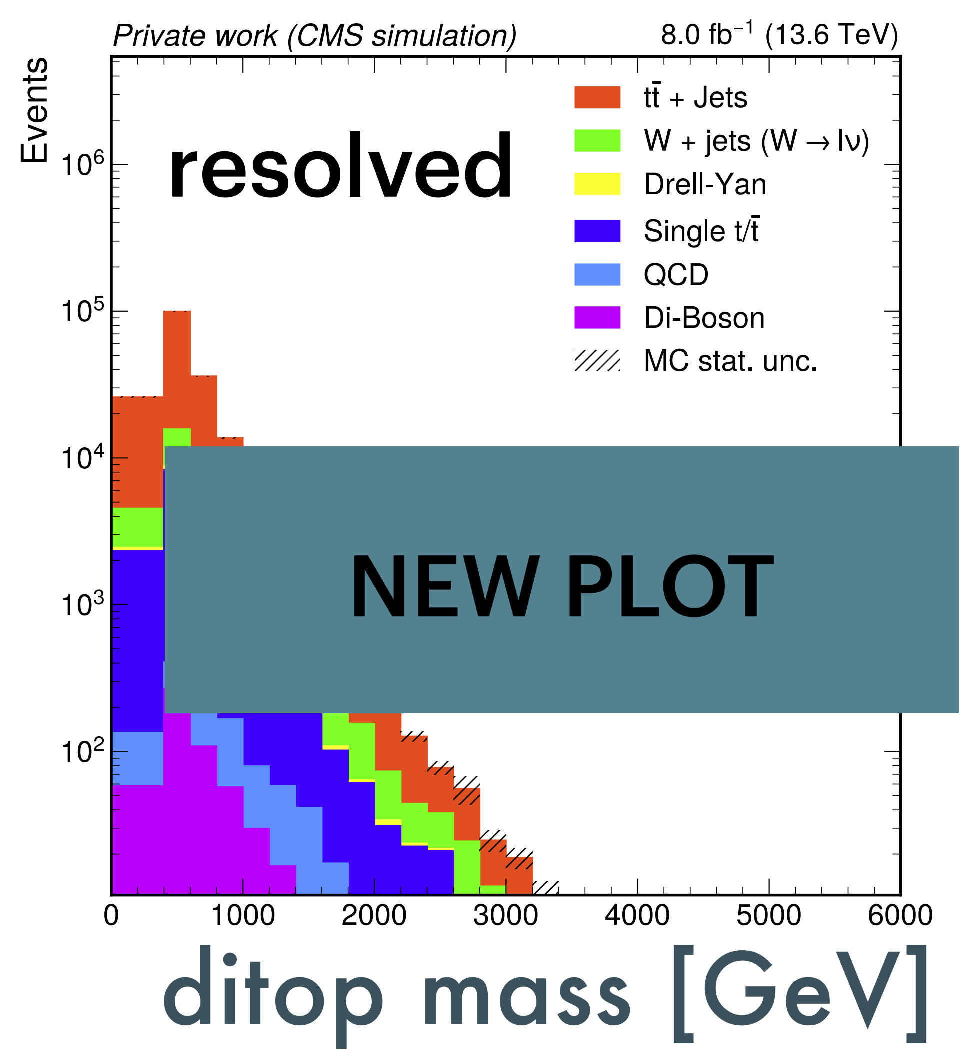
NEW PLOT

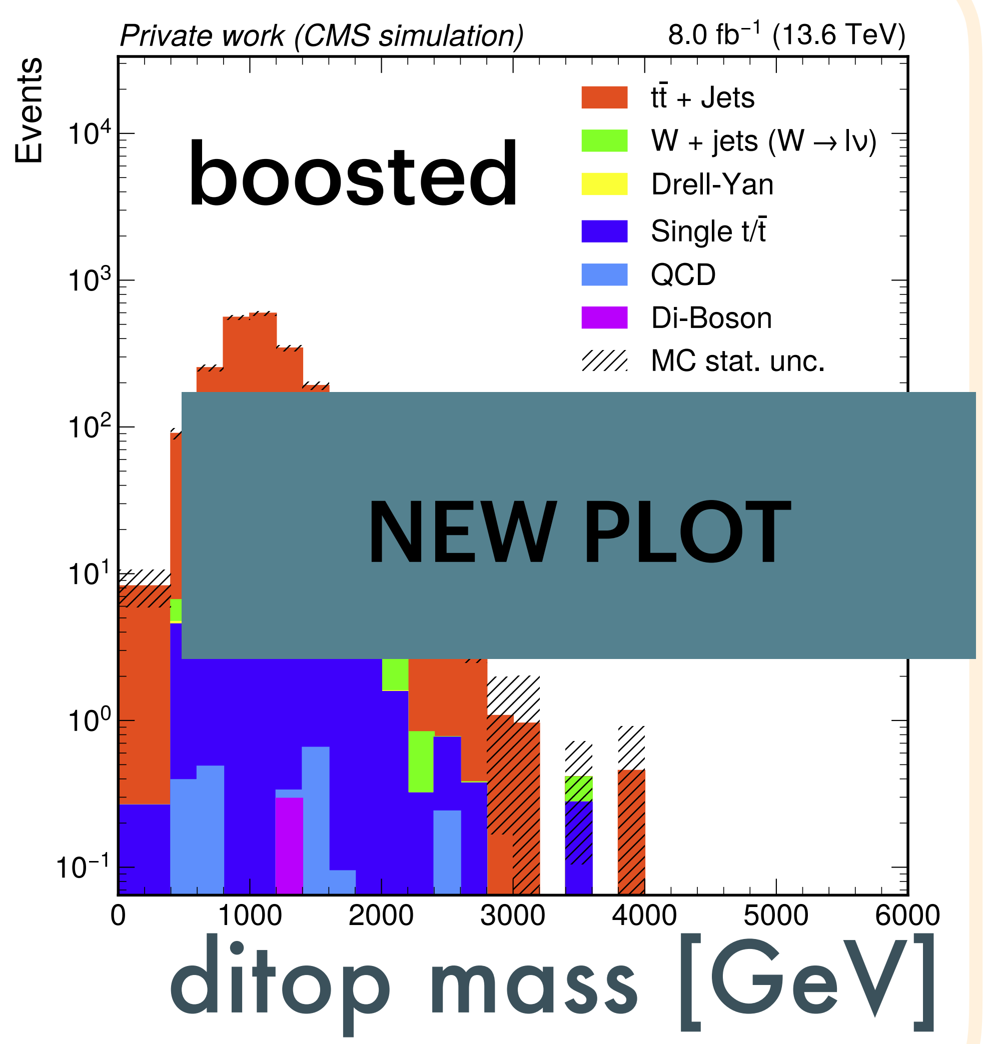
$\chi^2 < 30$

NEW PLOT

... In the Ditop Mass Spectrum

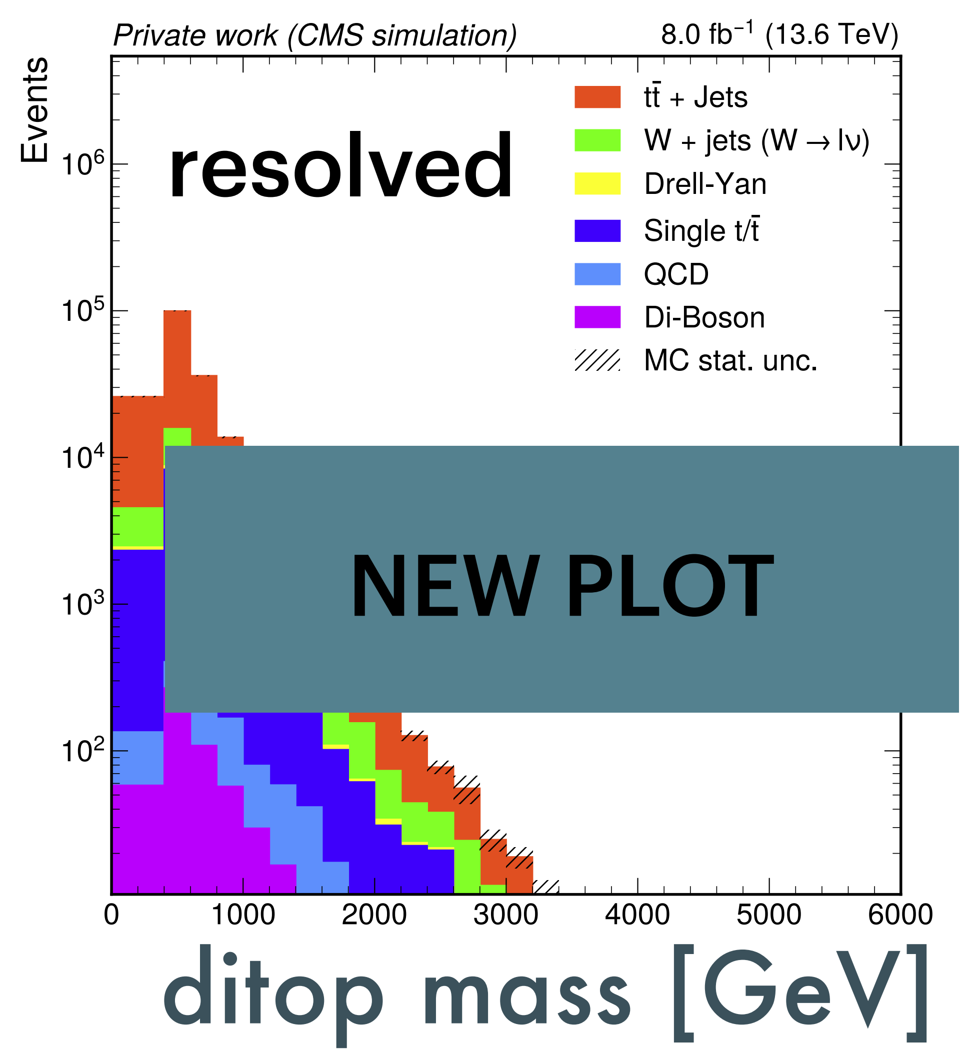
$\chi^2 < 30$

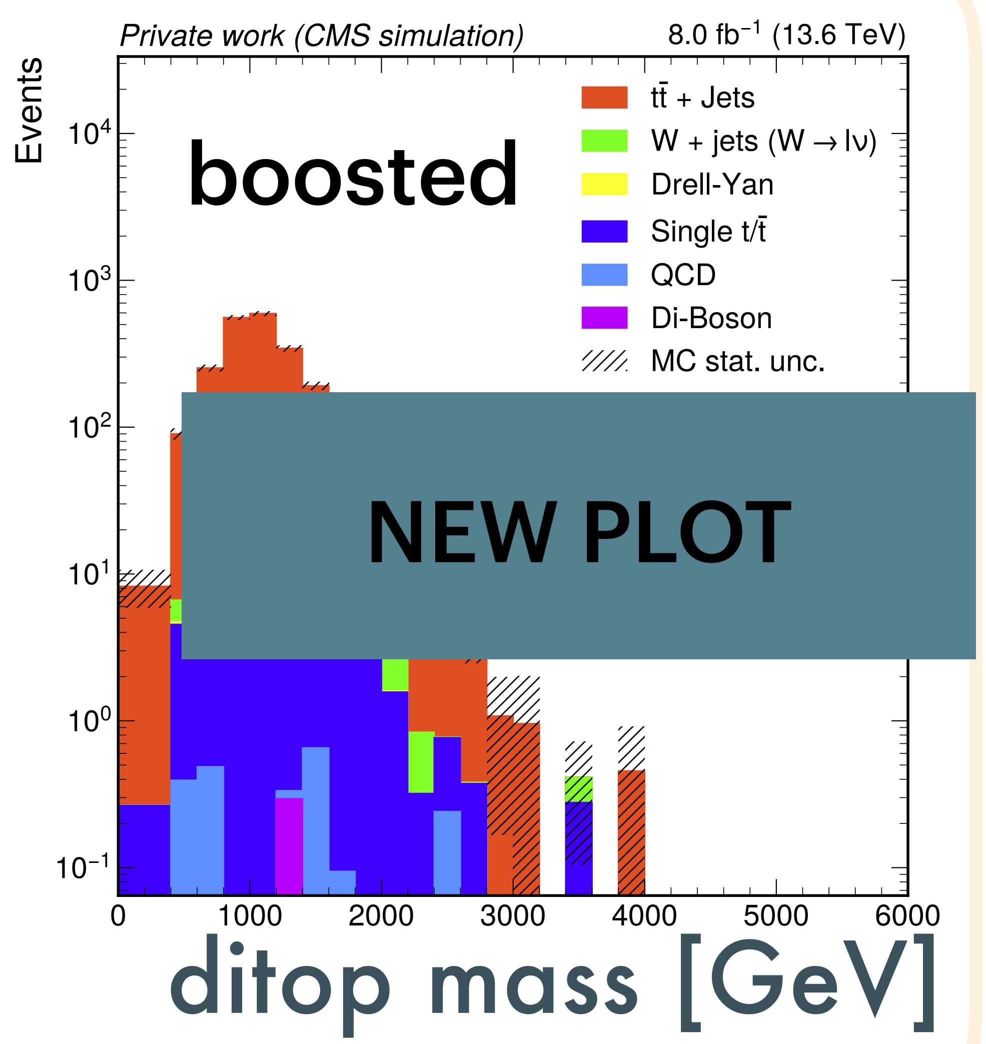




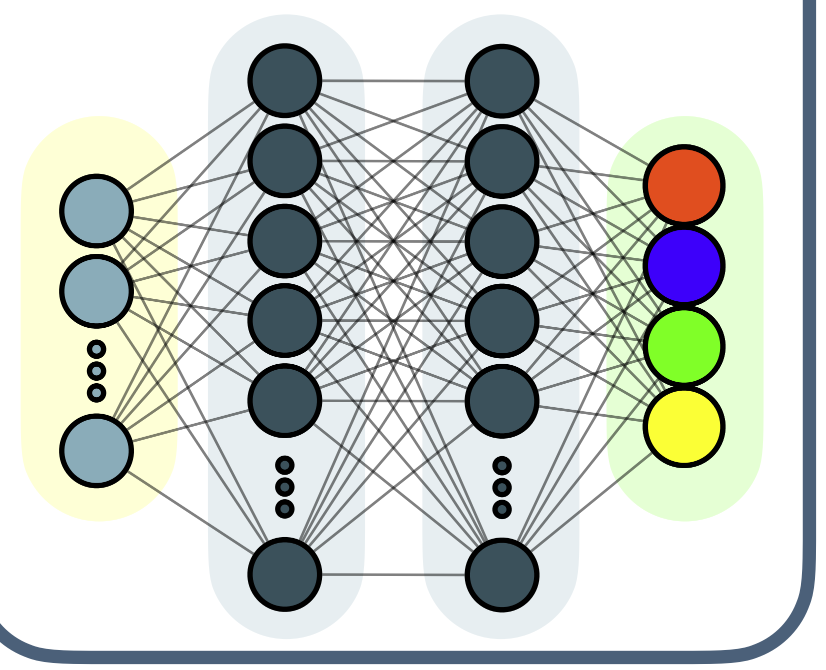
… In the Ditop Mass Spectrum

$\chi^2 < 30$





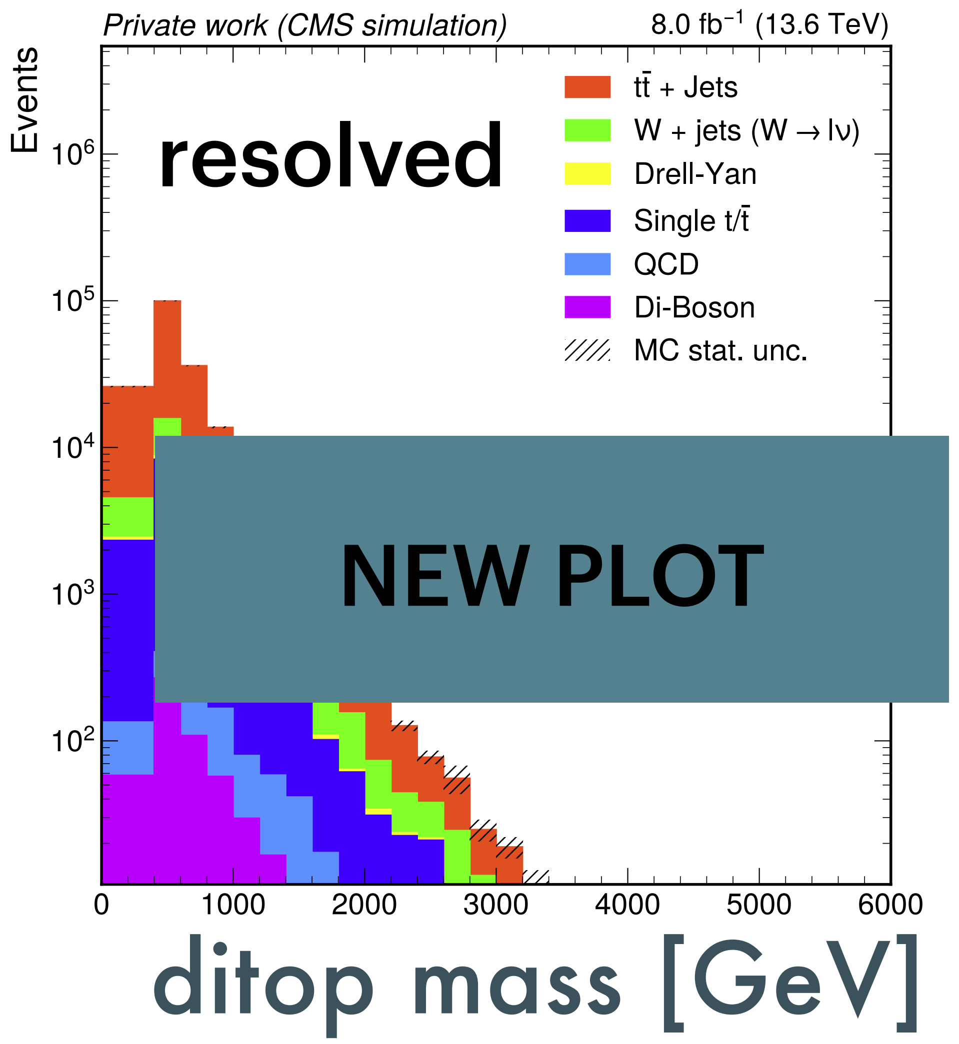
DNN for event classification

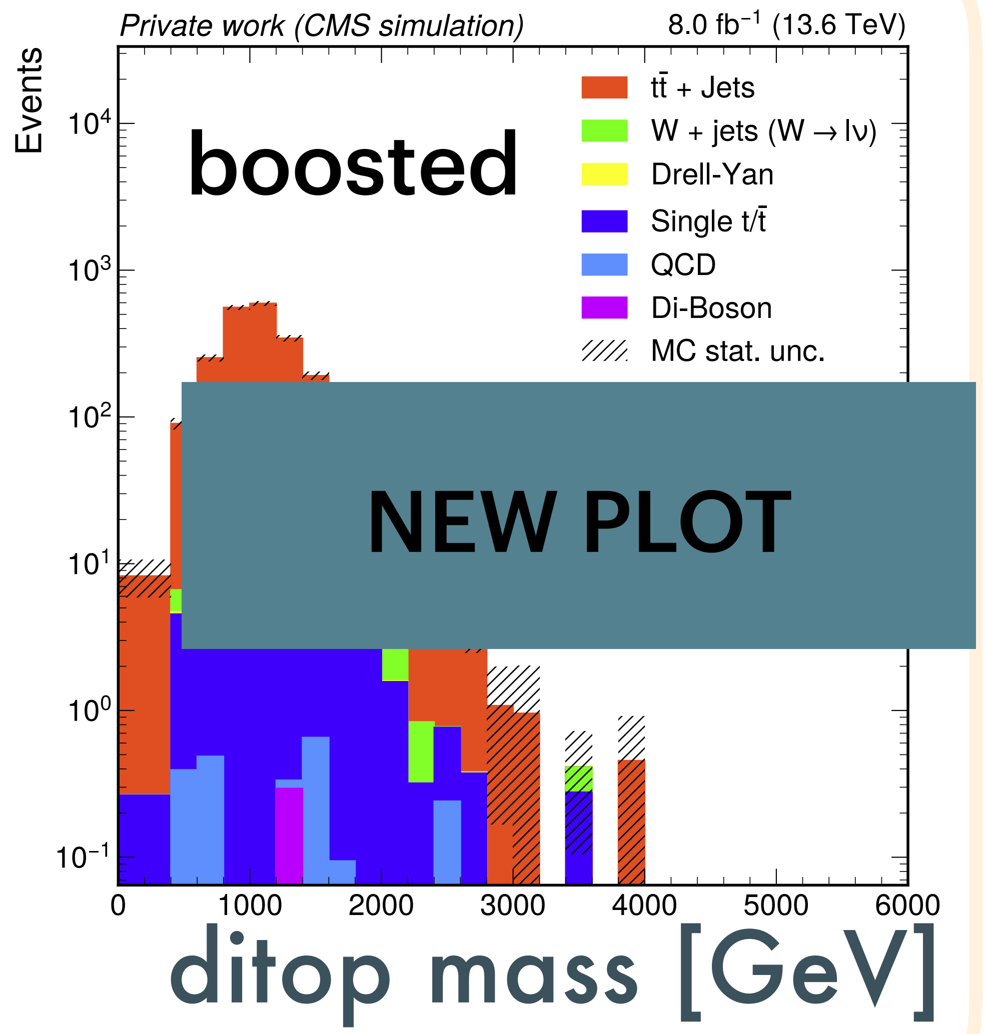


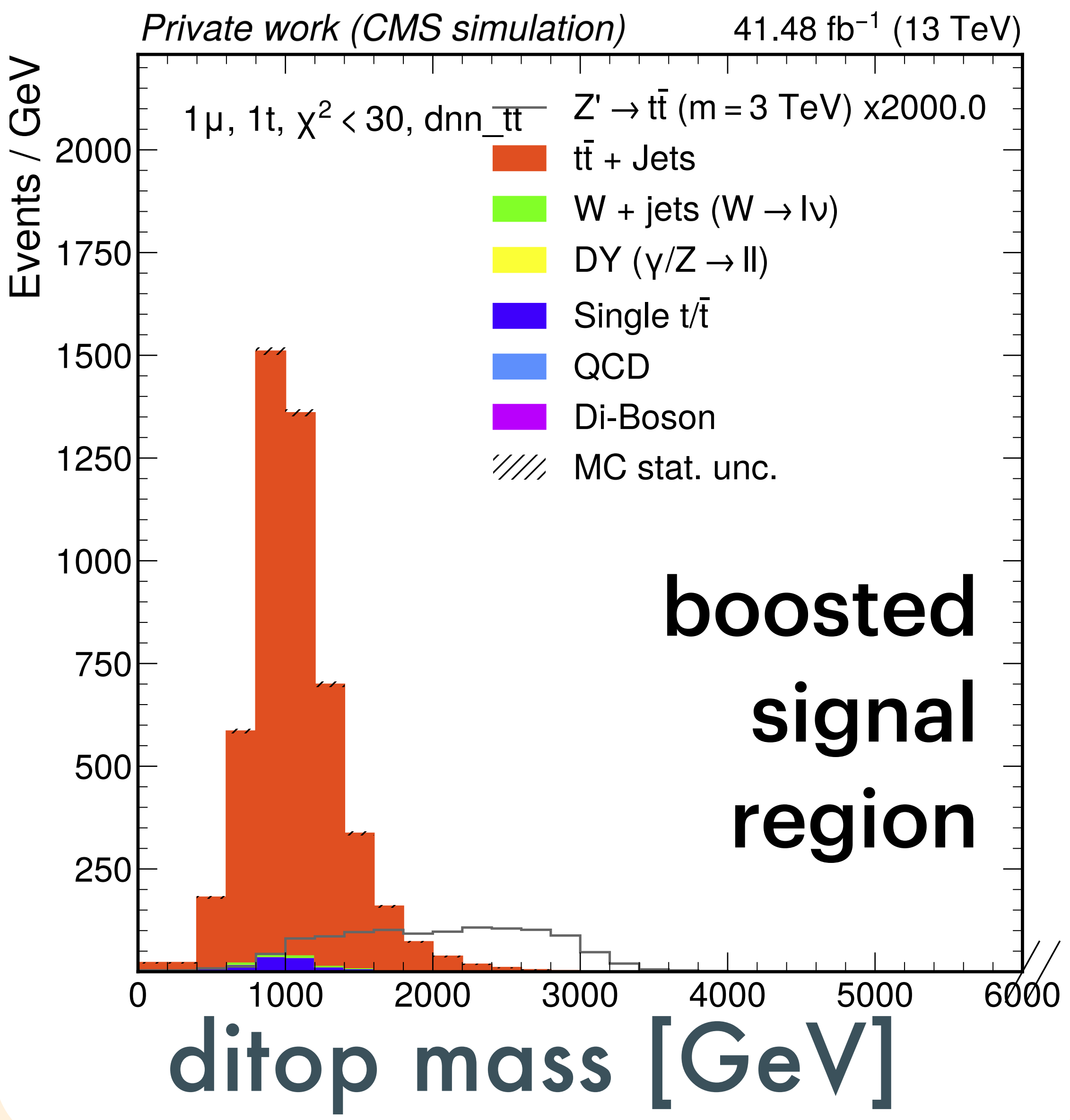
selected events

... In the Ditop Mass Spectrum

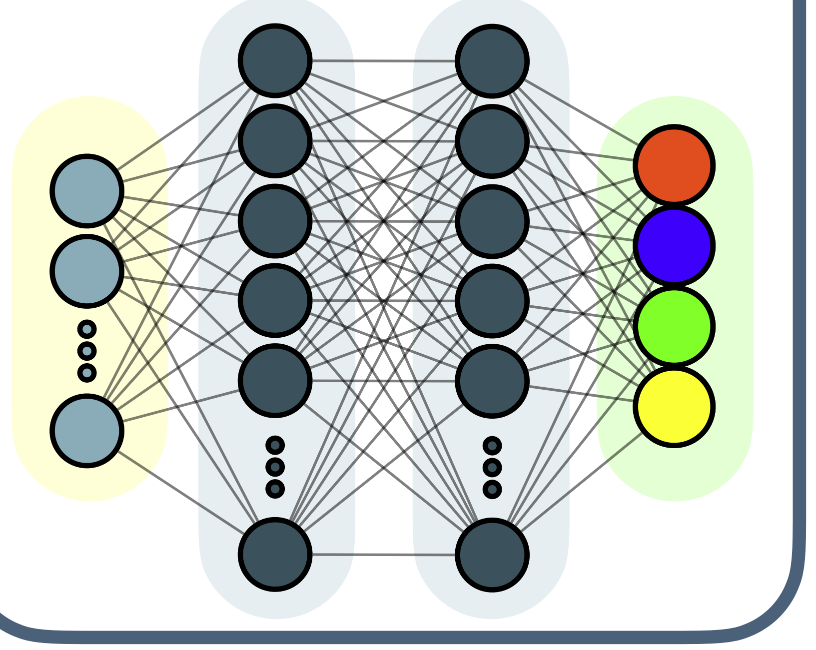
$\chi^2 < 30$





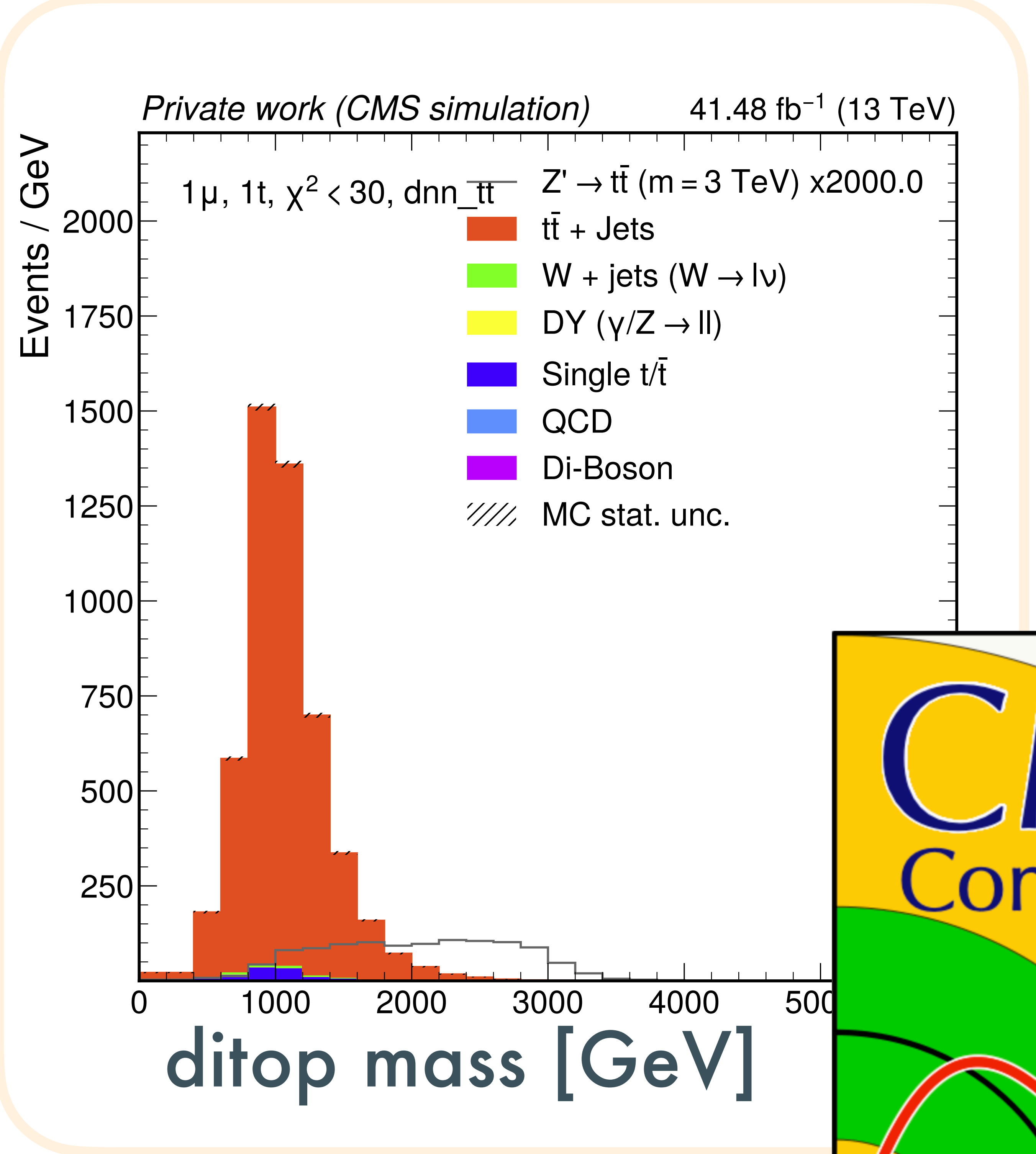


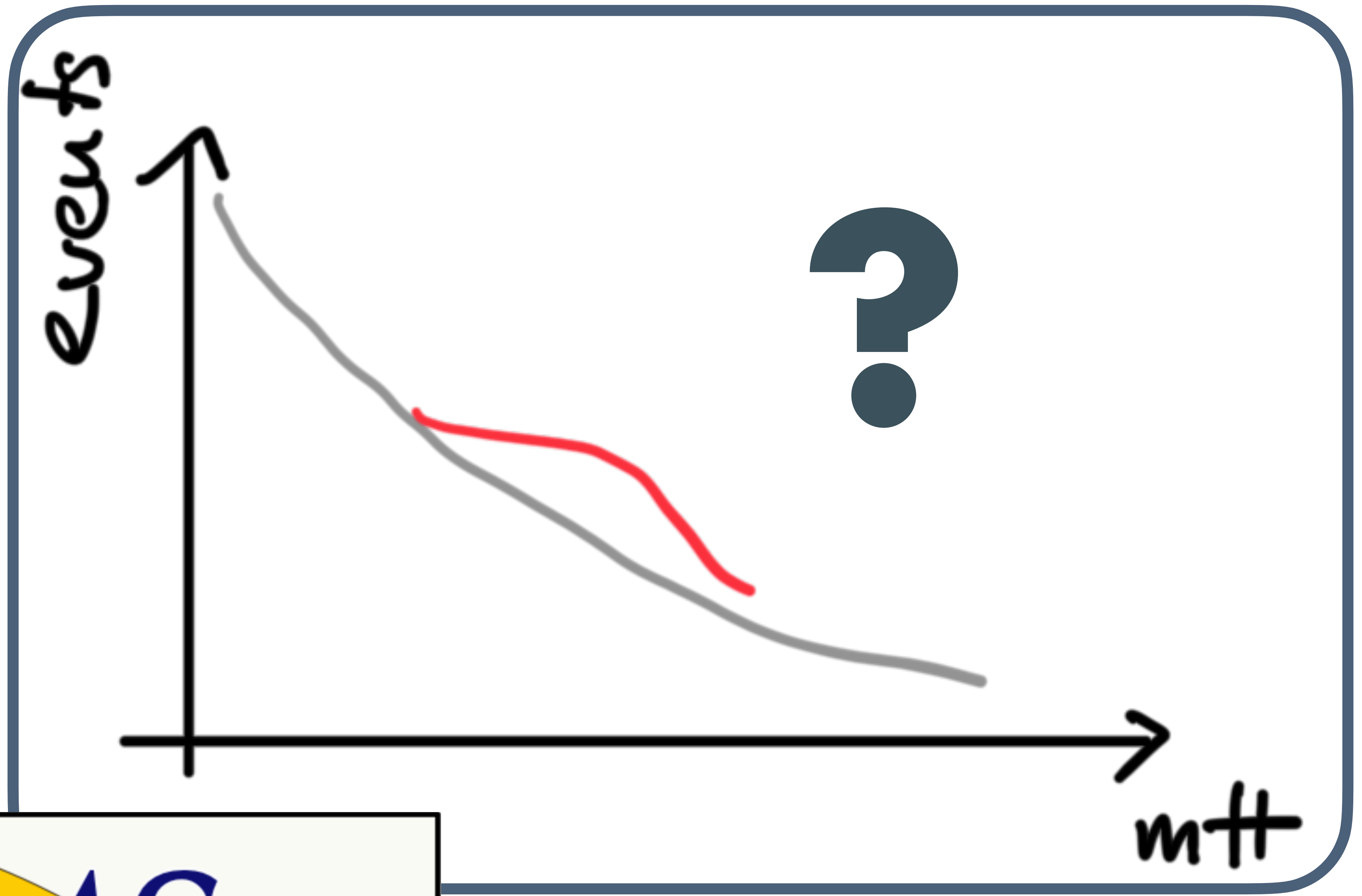
DNN for event classification



selected events

Statistical Analysis

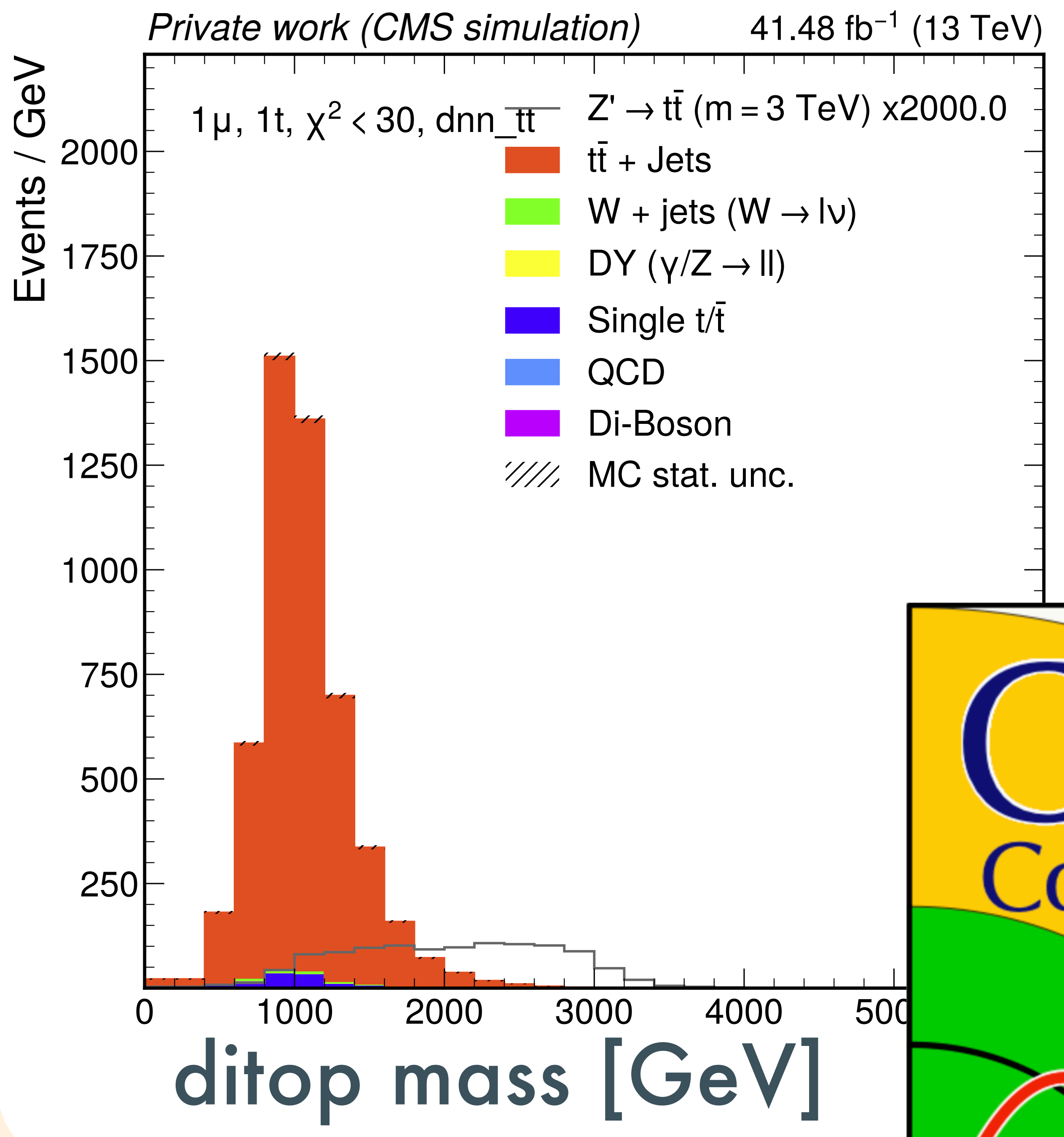


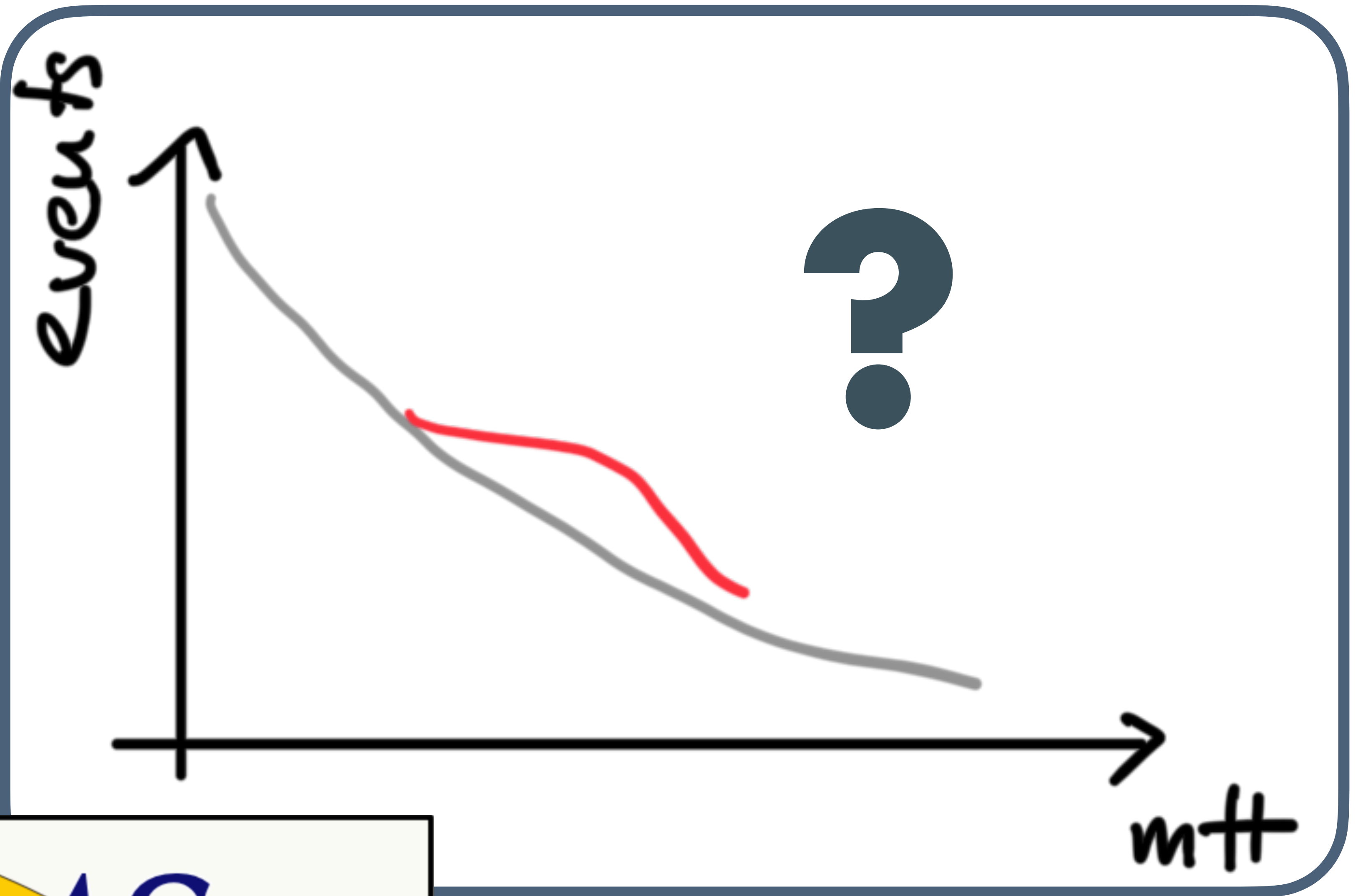


CMS Combine

Comput.Softw.Big Sci. 8 (2024) 1, 19 - arXiv:2404.06614

Statistical Analysis

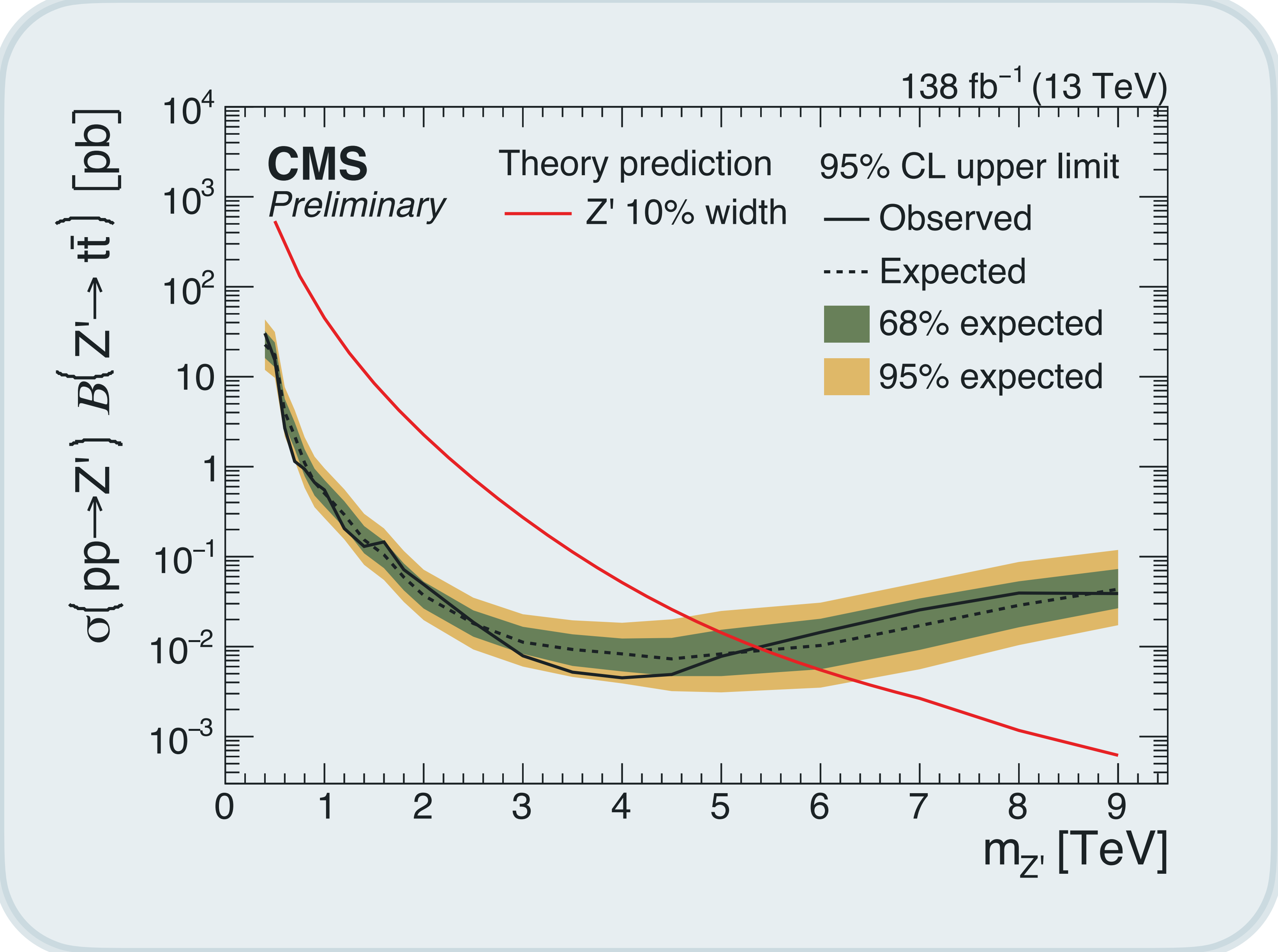




CMS Combine

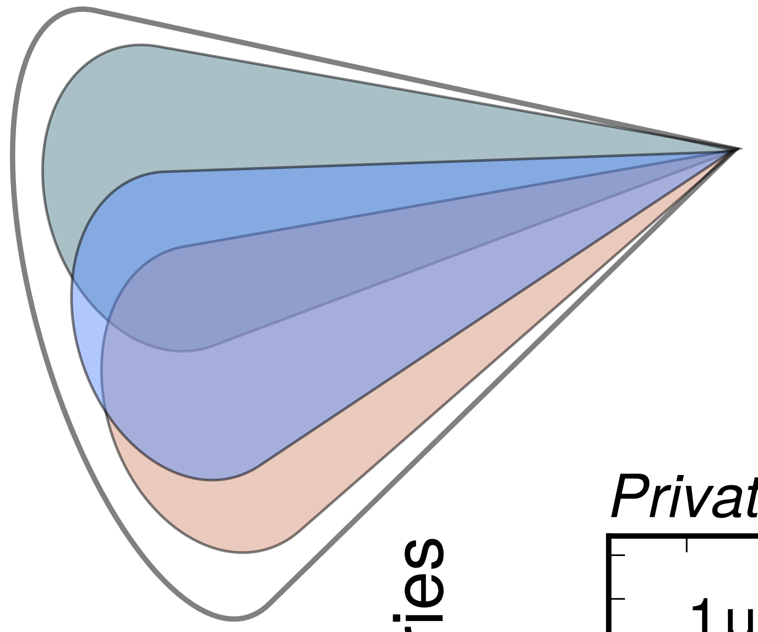
Comput.Softw.Big Sci. 8 (2024) 1, 19 - arXiv:2404.06614

CMS-PAS-B2G-22-006

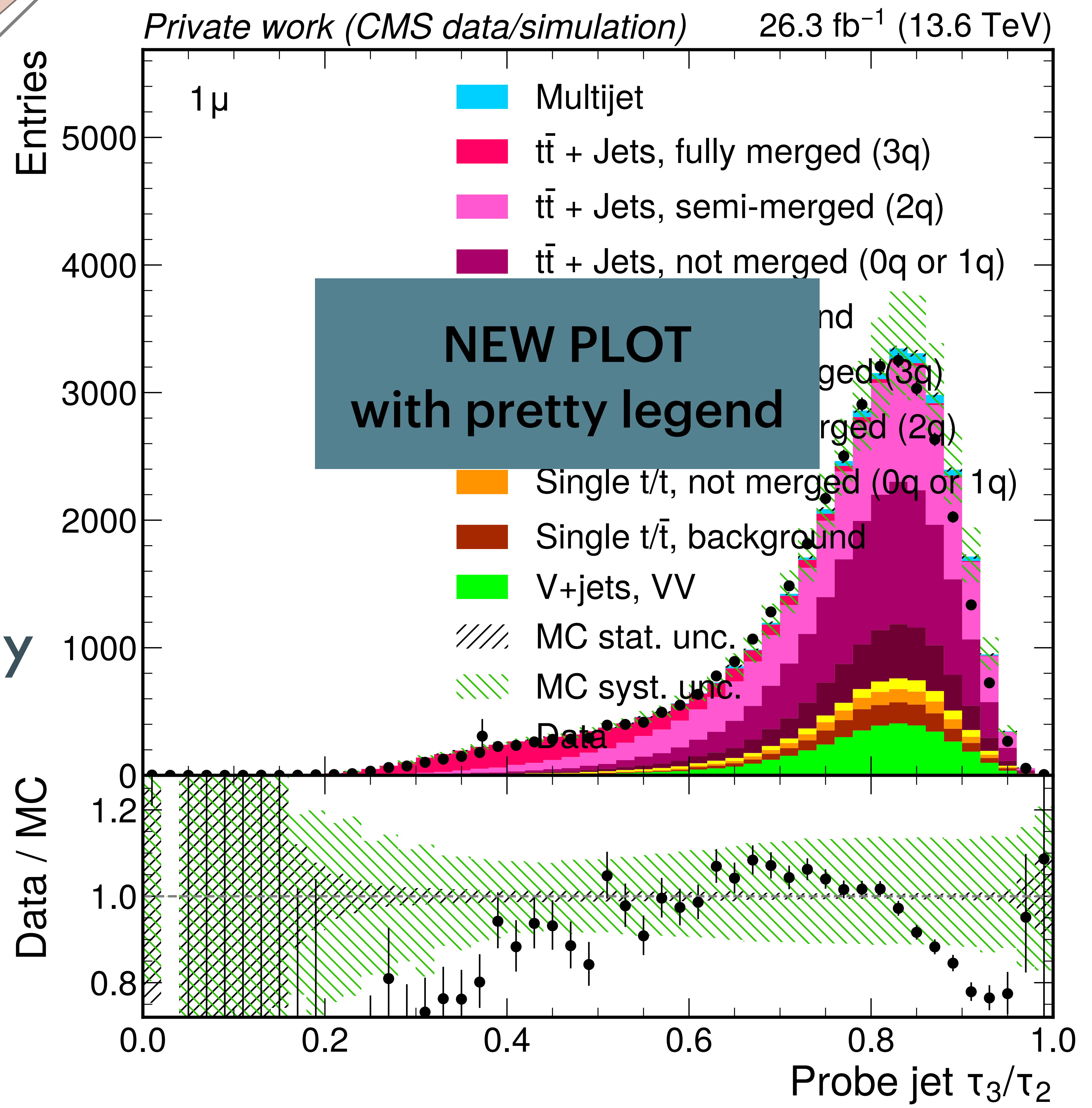


improve limits found in Run 2 analysis

Summary



- identification of boosted objects essential step in many analyses
- **boosted top quarks** as potential window to BSM physics
  here: heavy $Z/$Higgs bosons, non resonant $t\bar{t}$ production
  - Run 2 search public under CMS-PAS-B2G-22-006
  - Run 3 search to use more data and improved analysis strategy
- **cut based top tagging algorithm** using jet substructure and soft drop mass as baseline for top tagging
  - derived working points and scale factors

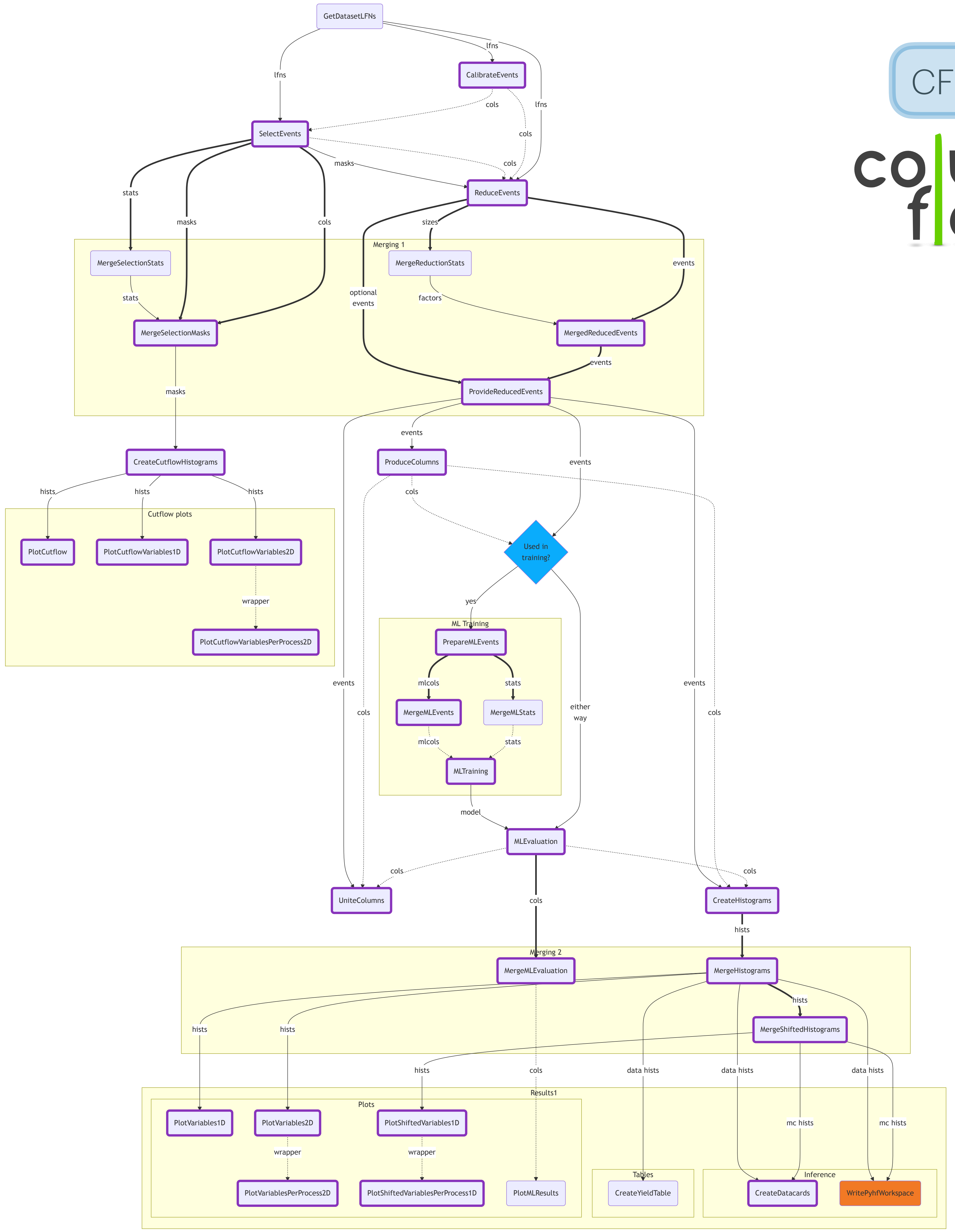


BACKUP

New Framework for Run 3

— columnar-based analysis with columnflow (CF) —

- new framework in development in CMS
- advantages:
  - columnar analysis
  - fully orchestrated with intermediate outputs
  - Python-based
- single command triggers complete workflow
  - very quick turnaround time

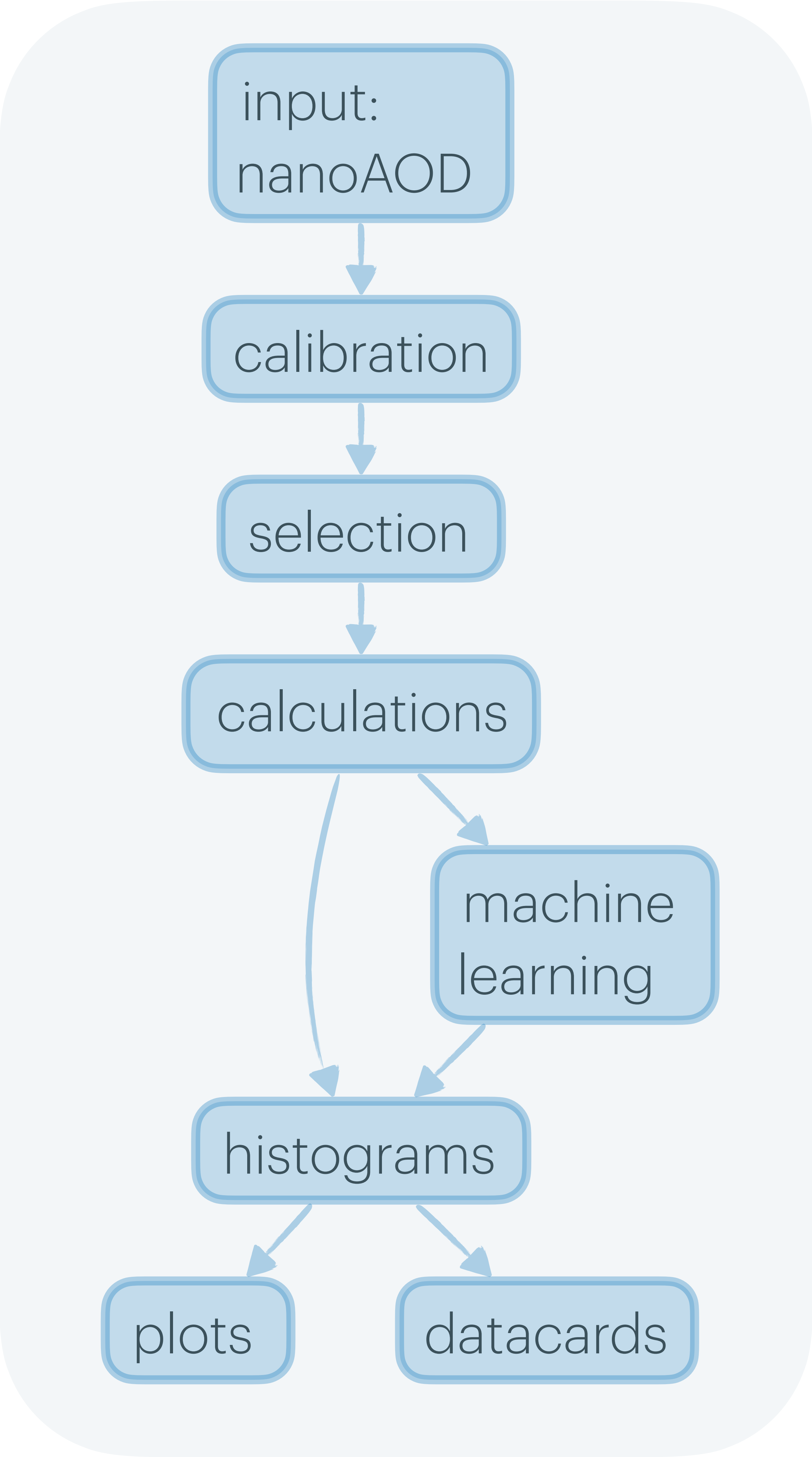


CF repo

New Framework for Run 3

— columnar-based analysis with columnflow (CF) —

- new framework in development in CMS
- advantages:
  - columnar analysis
  - fully orchestrated with intermediate outputs
  - Python-based
- single command triggers complete workflow
  - very quick turnaround time



CF repo

columnflow

Procedure and Selection

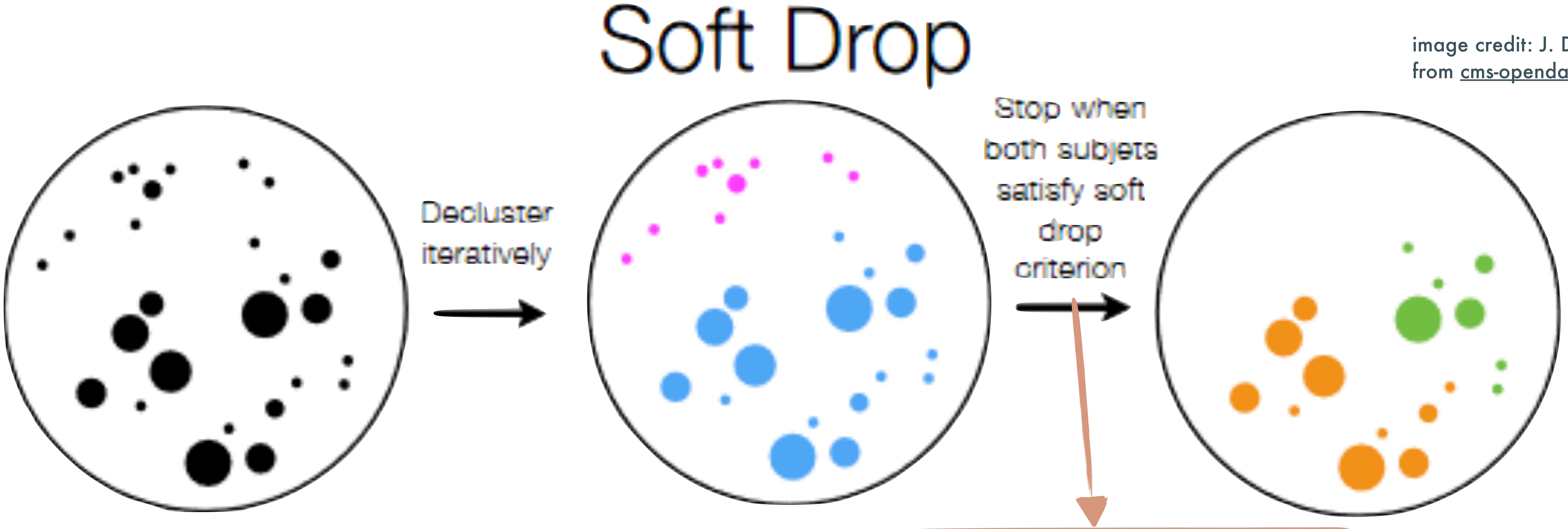
— for WP derivation —

1. Set mis-tag rates defining the WP:
   `0.1%` → `very tight`
   `0.5%` → `tight`
   `1.0%` → `medium`
   `2.5%` → `loose`
   `5.0%` → `very loose`
2. Plot mis-tag rate and determine cut values on τ_{32}
3. Plot signal efficiency and determine signal efficiency for these cut values/WP

- select FatJet (AK8 jet) with
  - $|\eta| < 2.4$
  - $p_\mathrm{T} > 400$ GeV
    → boosted regime
  - in `ttbar` sample to determine signal efficiency:
    uniquely matched to parton-level top quark
  - additional softdrop mass constraint:
    $105 < m_\mathrm{SD} < 210$ GeV
    → window around top mass

m_{SD} Of FatJets

image credit: J. Dolen
from cms-opendata-workshop



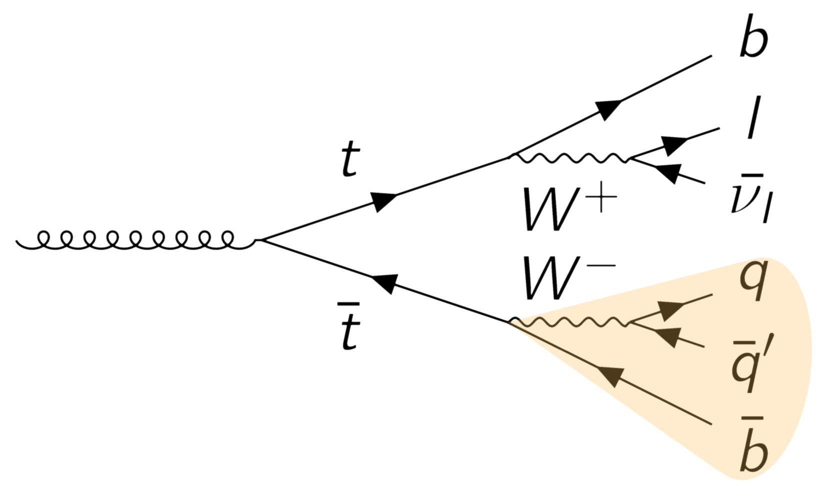
$$\frac{\min(p_{T1}, p_{T,2})}{p_{T1} + p_{T2}} > z_{cut} \cdot \left(\frac{\Delta R_{12}}{R_0}\right)^{\beta}$$

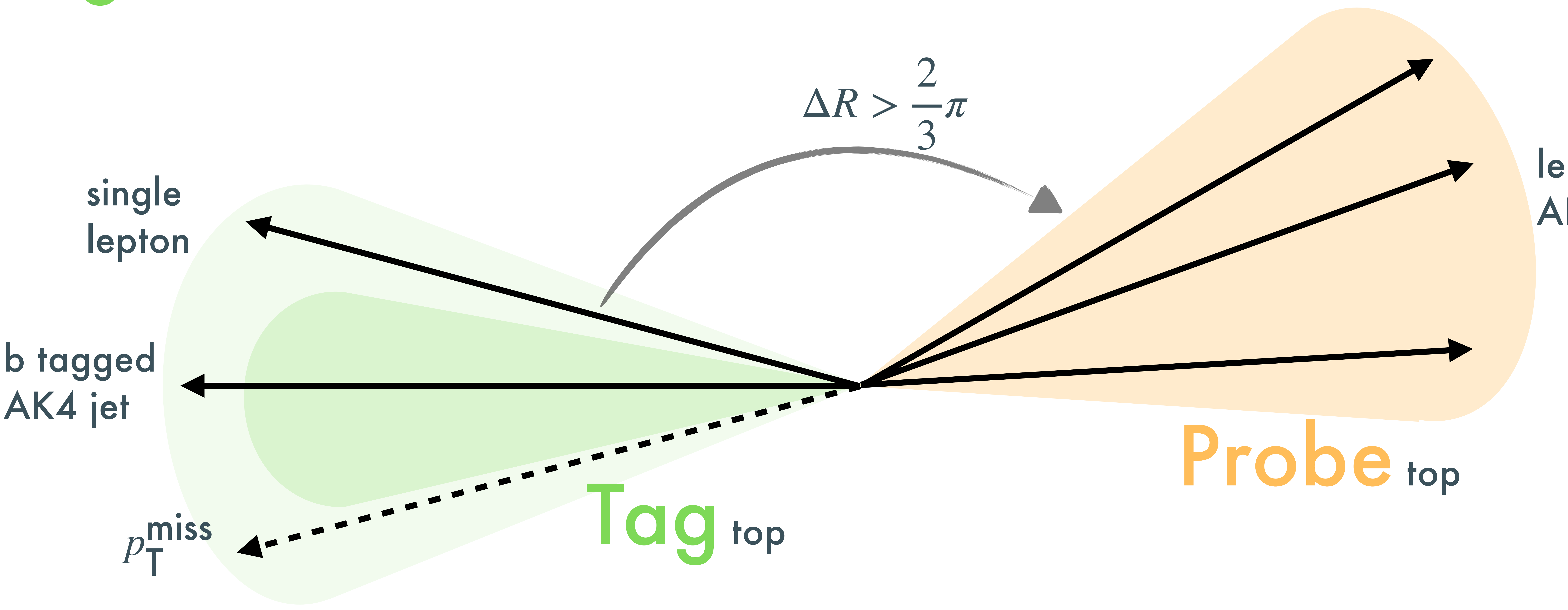
$\beta = 0$

$z_{cut} = 0.1$

→ fails for wide-angle soft radiation

Tag and Probe Method With $t\bar{t}$ Pairs





- single lepton
- $p_{T,\,lep} > 55$ GeV

- b-tagged AK4 jet close to lepton
- $p_{T,\,AK4} > 15$ GeV

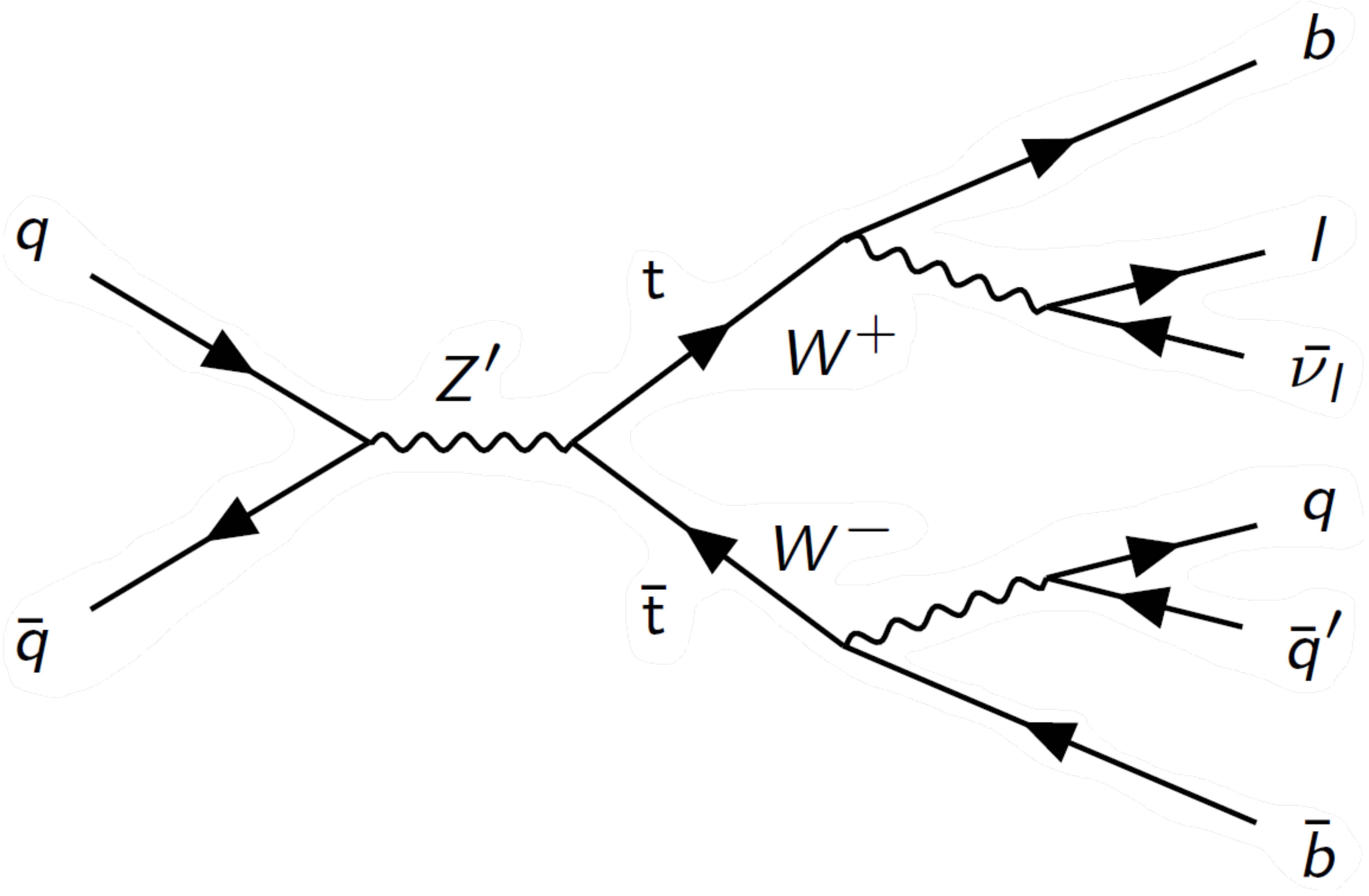
- $p_T^{miss} > 50$ GeV
- $p_T(W_{lep}) > 150$ GeV

- $p_{T,\,AK8} > 300$ GeV
- jet-lepton 2D isolation

Baseline Selection

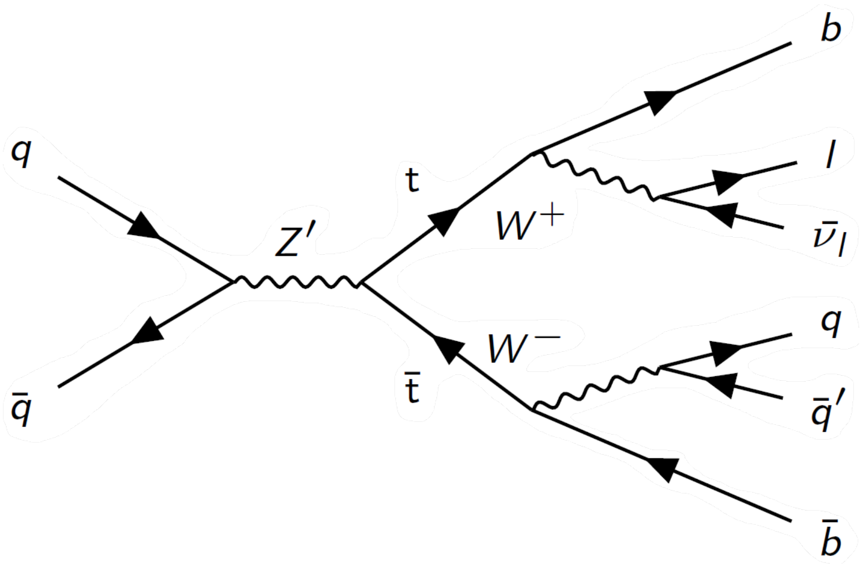
— in numbers —

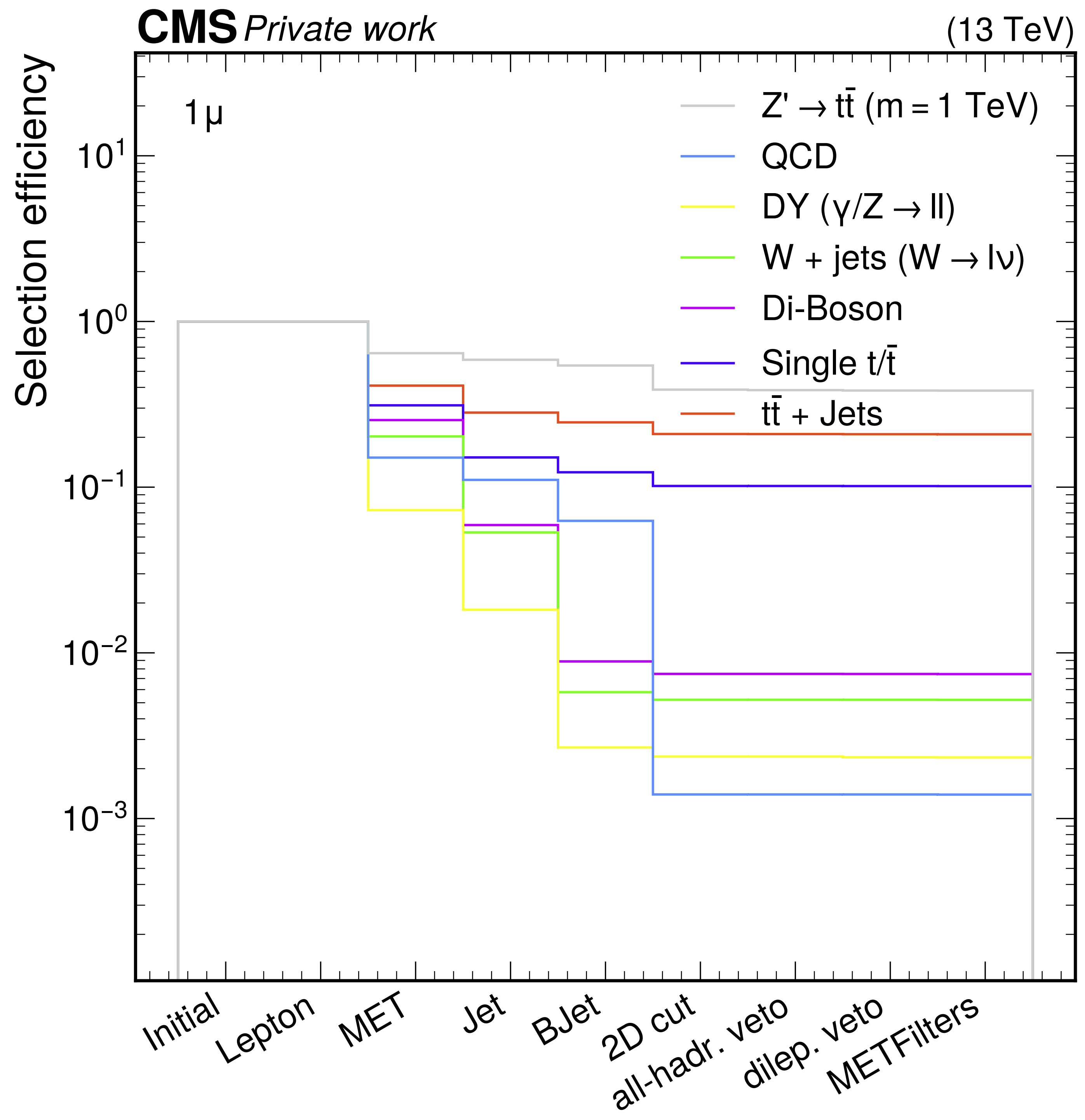
muon/electron channel



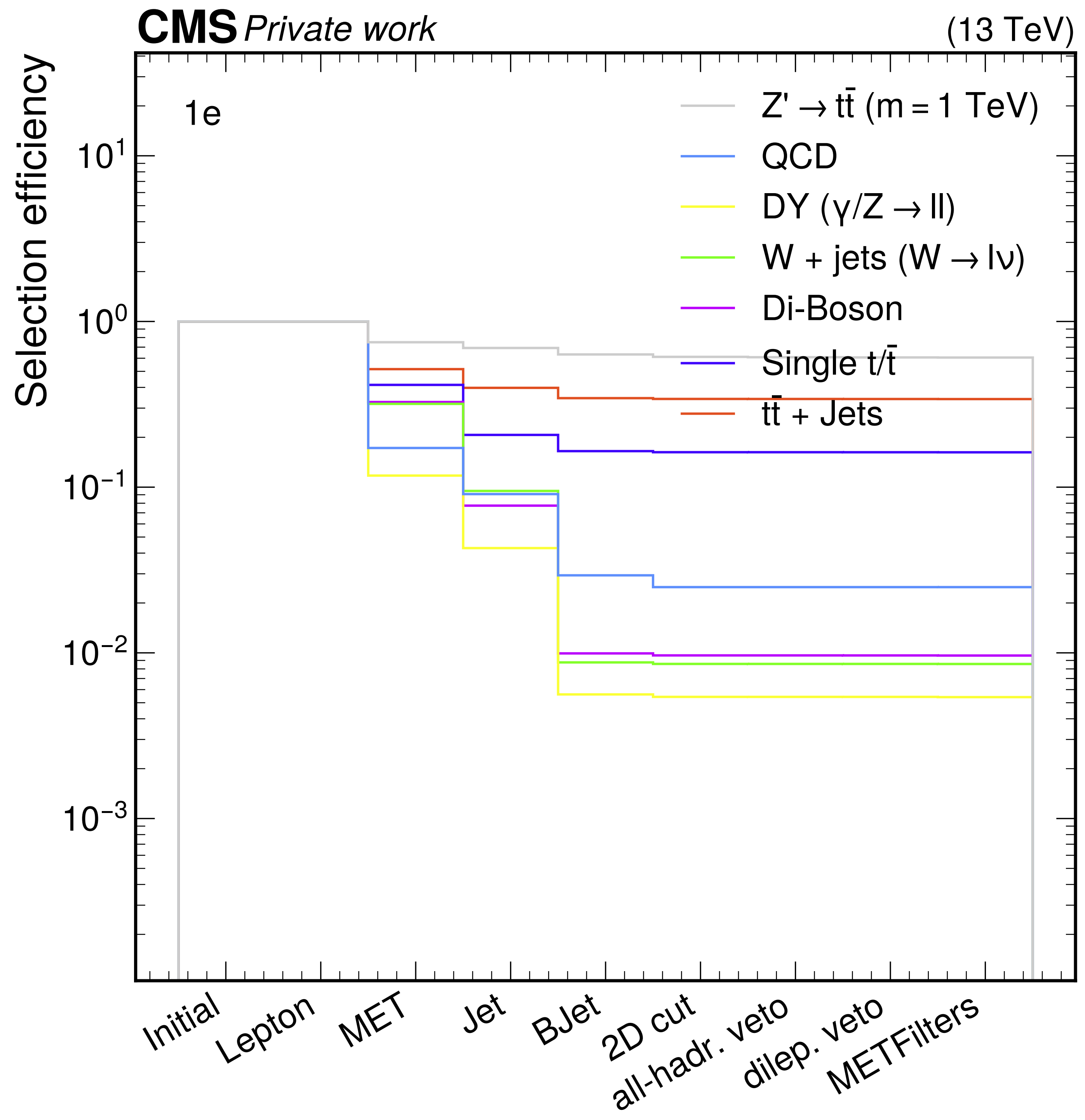
- lepton selection:
  - single muon/electron trigger
  - low p_T: $30/35 < p_T < 55/120$ GeV and isolation
  - high p_T: $p_T > 55/120$ GeV and 2D isolation
  - $|\eta^{\mu/e}| < 2.4$
- $p_T^{miss} > 70/60$ GeV
- at least two jets with $|\eta| < 2.5$ and:
  - $p_T^{jet1} > 50/50$ GeV
  - $p_T^{jet2} > 50/40$ GeV
  - at least one b-tagged jet

Selection Efficiencies

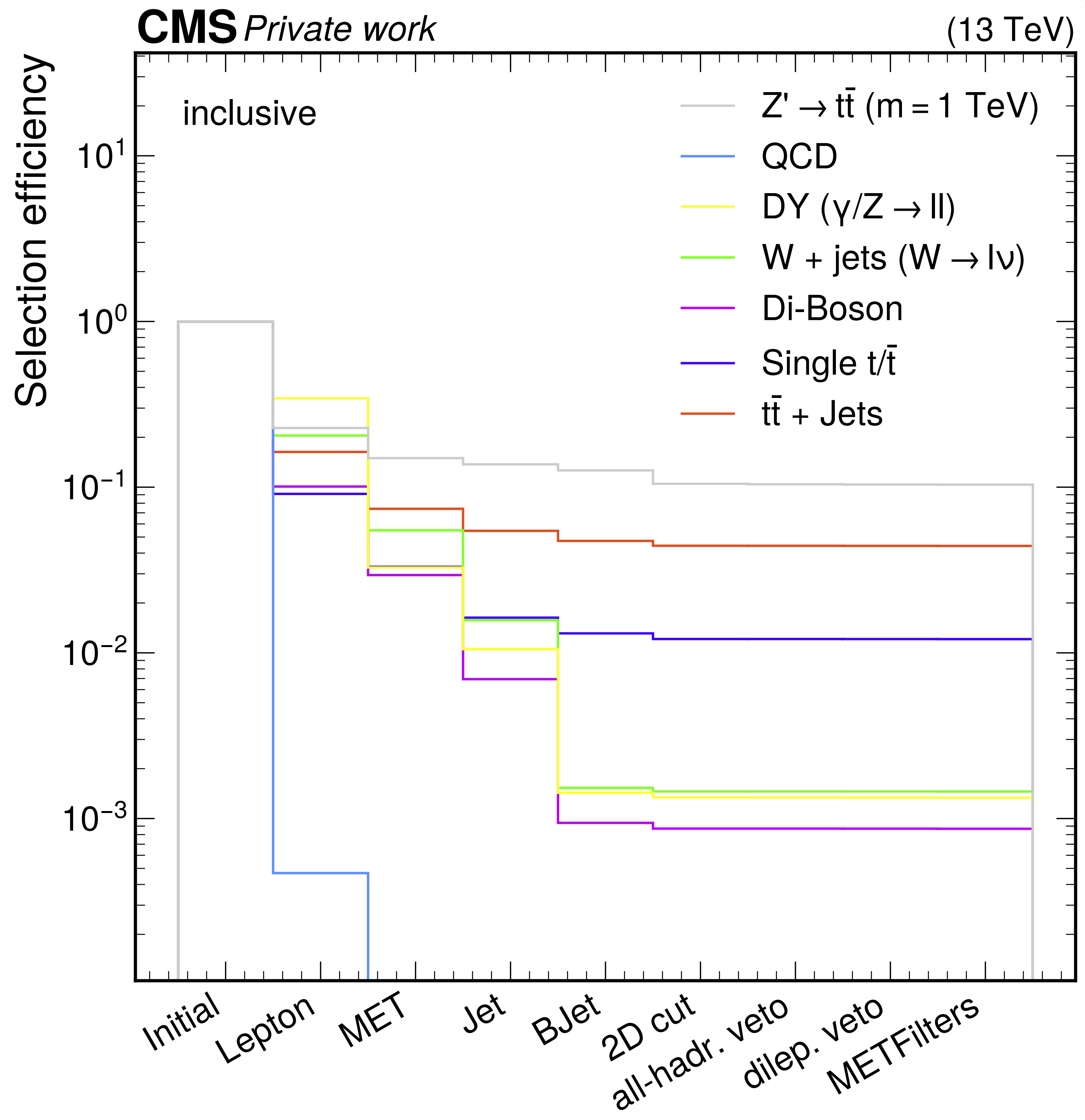




muon channel



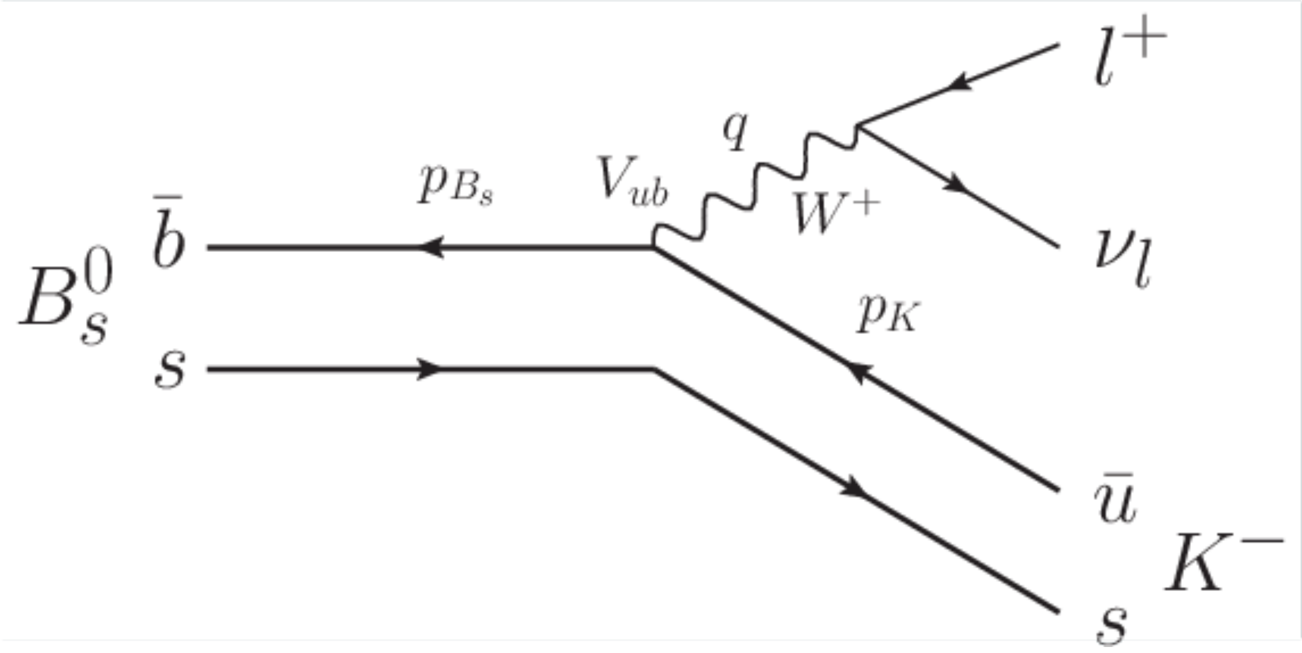
electron channel



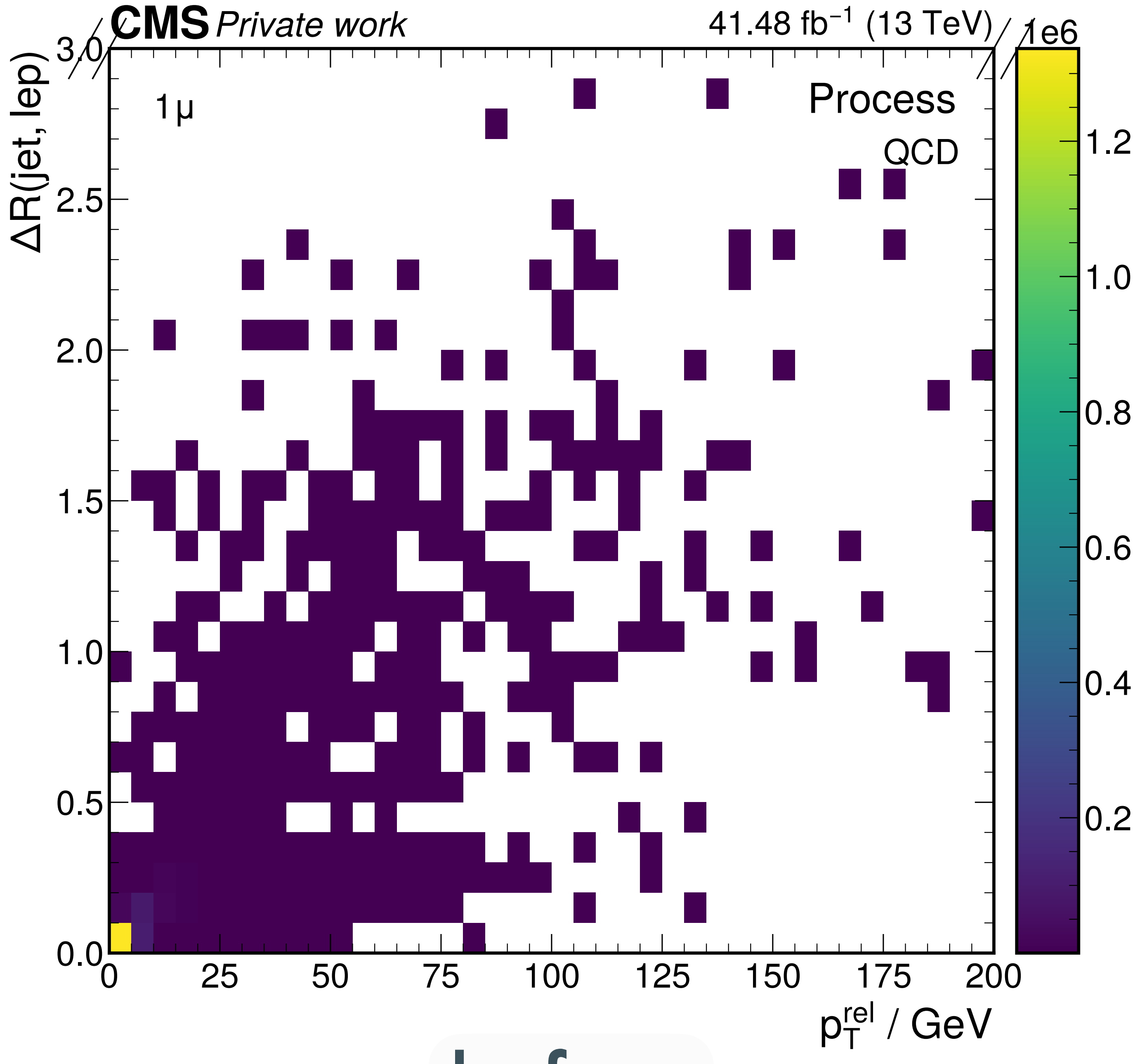
inclusive channel

Jet-Lepton 2D Cut

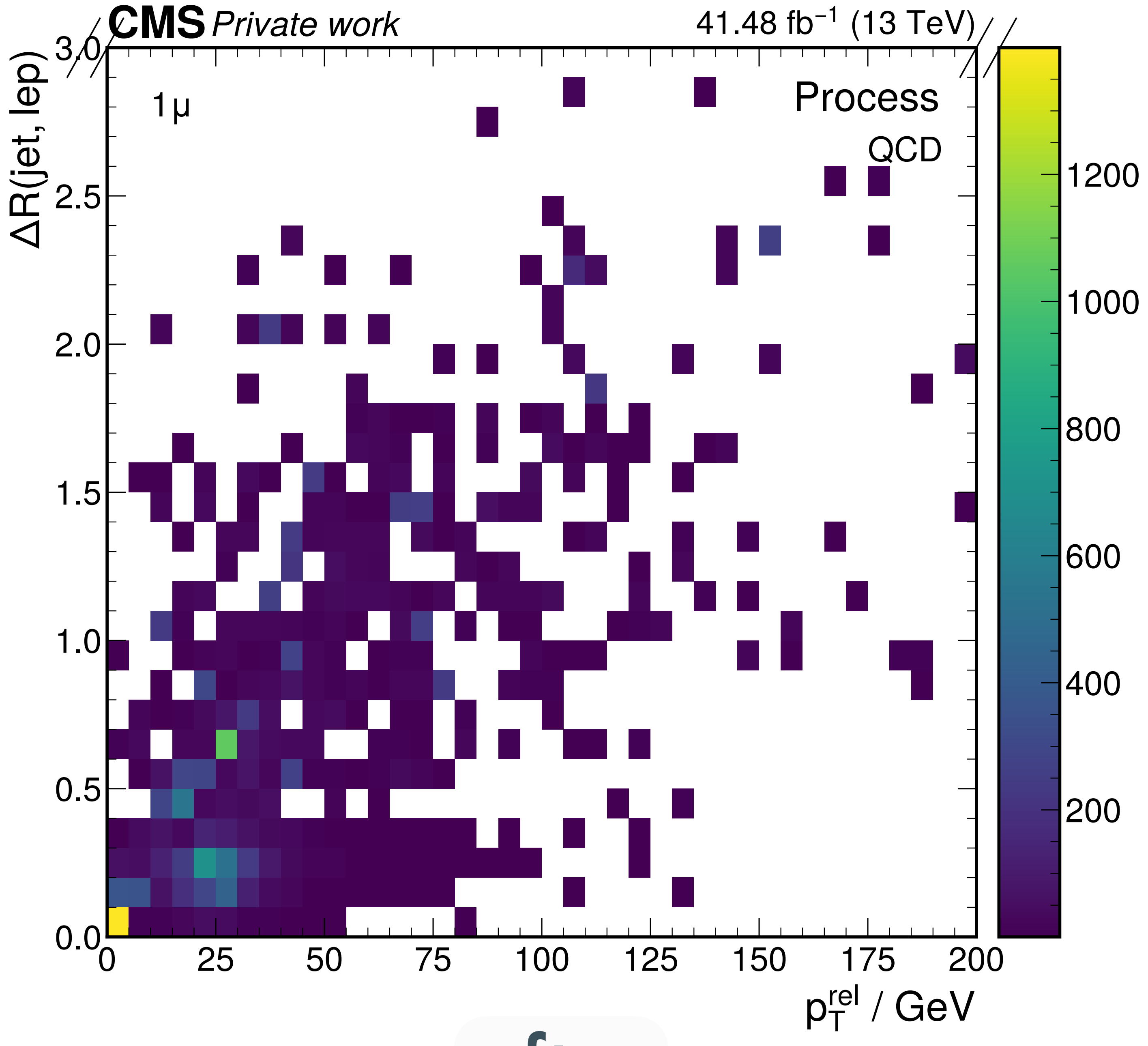
— filtering leptonic B meson decays —



$\Delta R_{\text{min}}(\text{lep, jet}) > 0.4$ or $p_{\text{T, rel}}(\text{lep, jet}) > 25$ GeV



before



after

- note different color scale!
- 2D cut only in high p_T regime of lepton
    - contribution of low p_T contributions in first bin
- reduces QCD background

Merge Category Algorithm

— process_ids producer—

Process	Decay mode	# of merged quarks	Subprocess
tt	semi-leptonic	3	tt_3q
		2	tt_2q
		1 or 0	tt_1o0q
	dileptonic or fully hadronic	any	tt_bkg
st	hadronic	3	st_3q
		2	st_2q
		1 or 0	st_1o0q

Process	Top decay	Assoc. W decay	# of merged quarks	Subprocess
st_tW	leptonic	hadronic	3	st_bkg*
			2	st_2q
			1 or 0	st_1o0q
st_tW	leptonic	leptonic	any	st_bkg
st_other	leptonic	--	any	st_bkg

process ID set per event from gen level info:

1. Check if st or tt.
2. Check if bkg:
   no probejet
   OR top decay merged into probejet not hadronic
   OR associated b quark and both associated W decay products merge into probejet
3. Check if 3q or 2q:
   not bkg
   AND exactly 3/2 quarks merged into probejet
4. Check if 0o1q:
   not bkg
   AND 0 or 1 quarks are merged into probejet

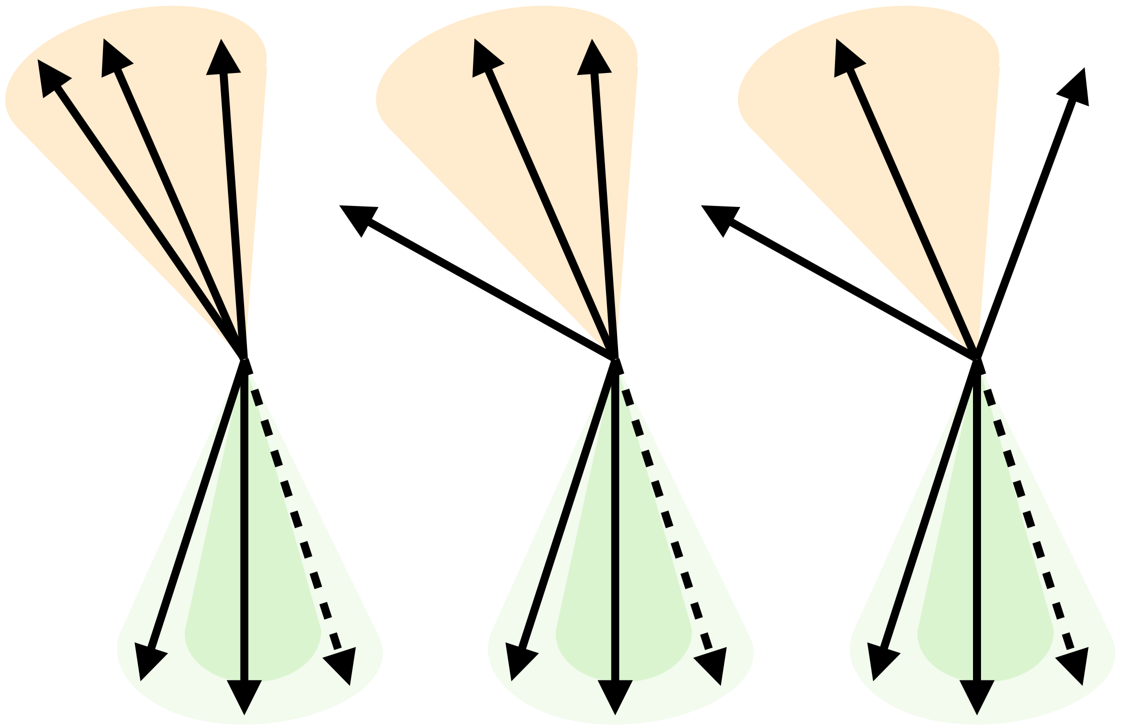
Combine Physics Model

- with seesaw condition, keeping N_{total} constant
- SF_{pass} as parameter-of-interest, determining SF_{fail}
- one fit per combination of year/era and WP
  - simultaneous fit for merge categories and p_{T} bins of probe jet:

Seesaw Condition

$$N_{\text{total}} = N_{\text{pass}} + N_{\text{fail}} = \text{const}$$

$$SF_{\text{fail}} = (1 - SF_{\text{pass}}) \cdot \frac{N_{\text{pass}}^{\text{prefit}}}{N_{\text{fail}}^{\text{prefit}}} + 1$$



```
SF__3q/2q/0o1q__22__pt_300_400
SF__3q/2q/0o1q__22__pt_400_480
SF__3q/2q/0o1q__22__pt_480_600
SF__3q/2q/0o1q__22__pt_600_inf
```

→ 12 scale factors

Main Uncertainities

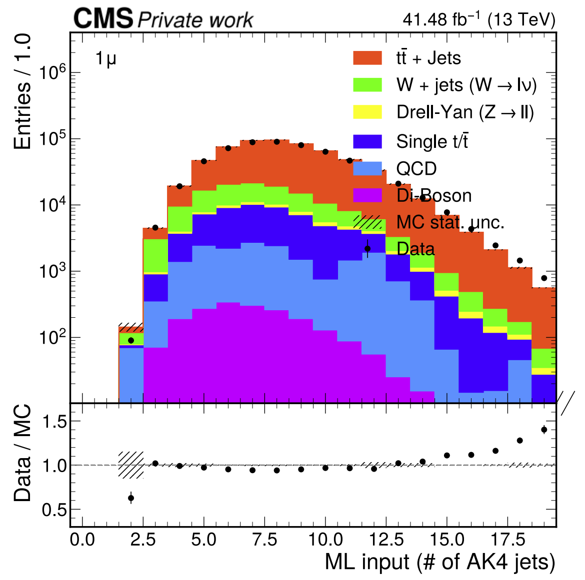
— in Run 2 —

- checked Run 2 impacts (JMAR Meeting 06.06.23) and wrote down five highest for each SF in each pT bin:

`FSR q2qgMuR/g2ggMuR/x2xgMuR`	available in Nano as `PSWeight`, not split (single NP)
`xsec tt/st/mj/vx`	already used in v8 fit
`scale tt/st`	available in Nano as `LHEScaleWeight`
`toppt B`	available
`JES`	all unc. sources available from central json
`PU`	available (`minbias_xs`)
`btagging`	to be investigated

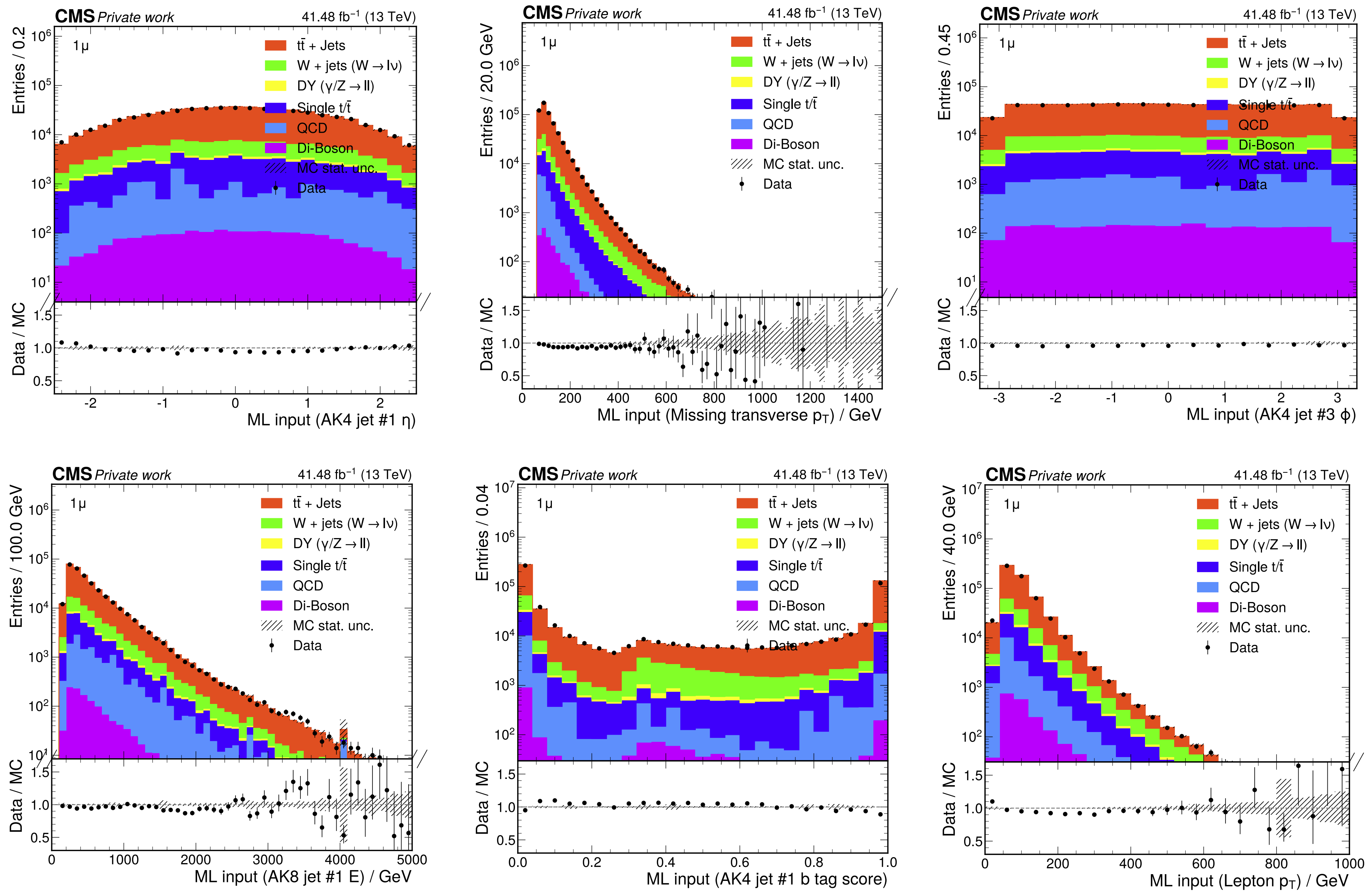
DNN for Event Classification

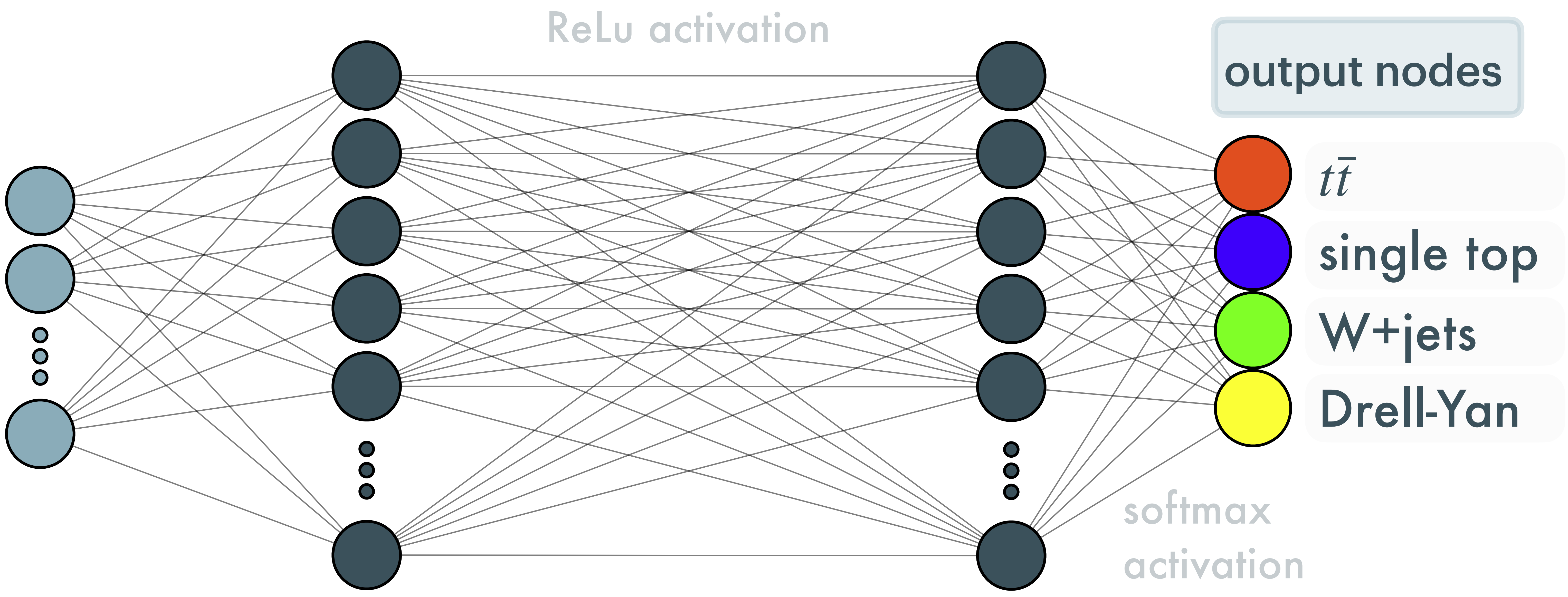
— network architecture —



59 input features:

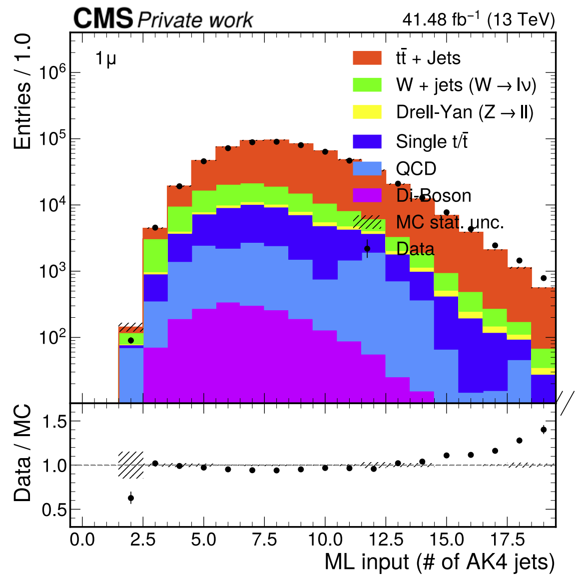
- jet multiplicities
- kinematic properties of objects (MET, jets, lepton)





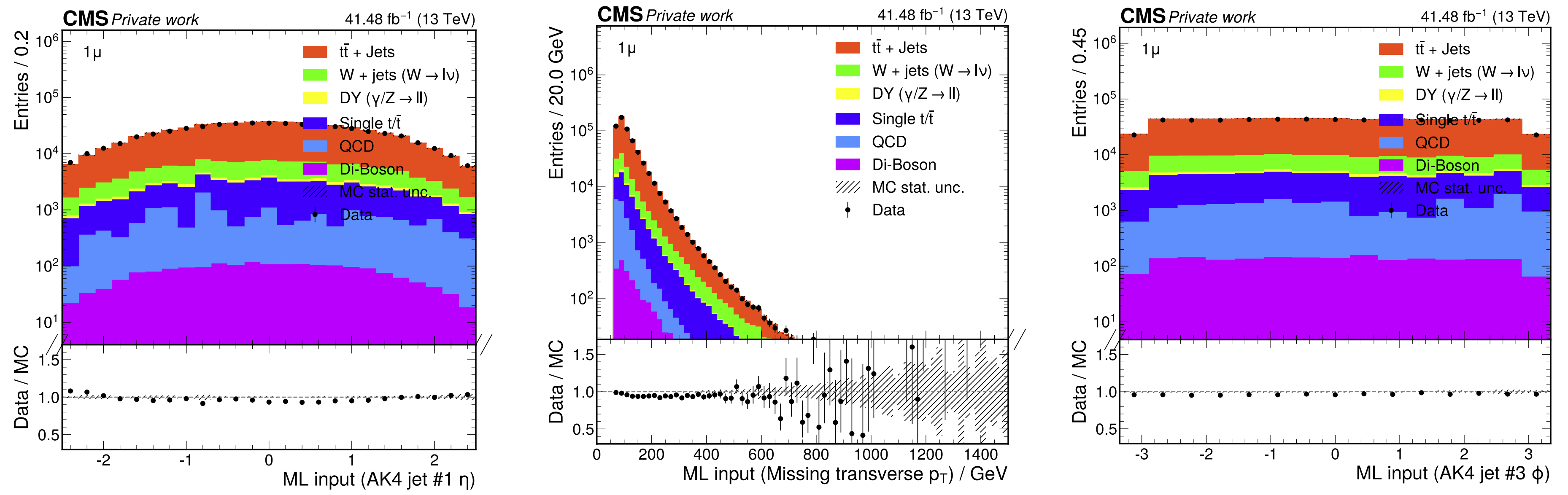
DNN for Event Classification

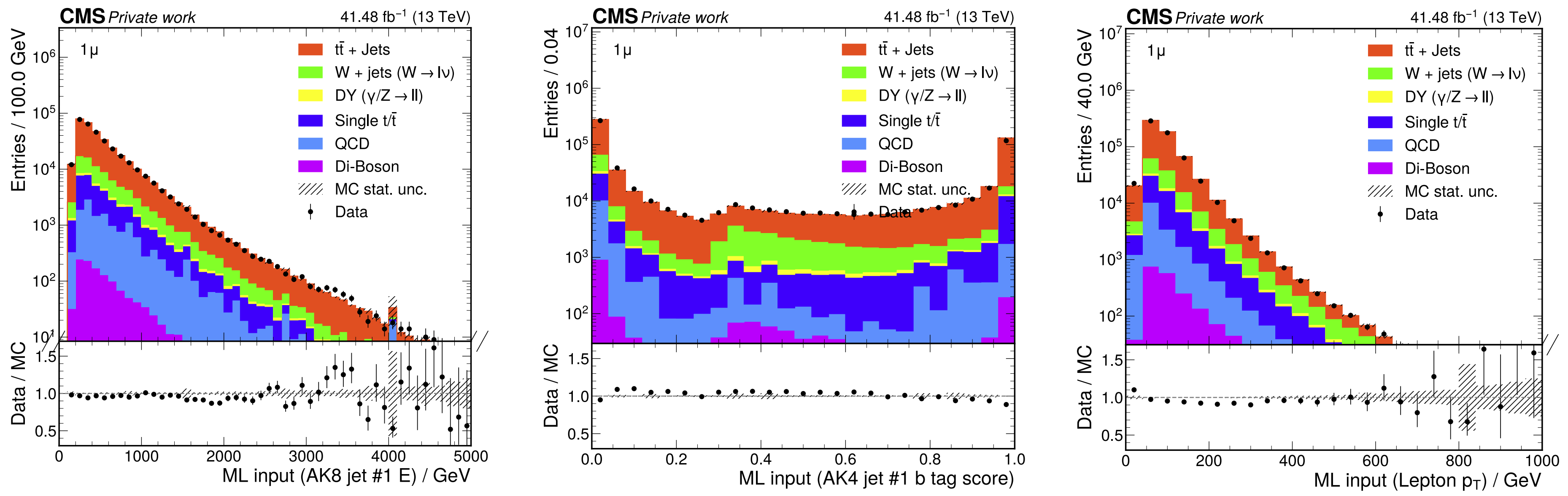
— network architecture —

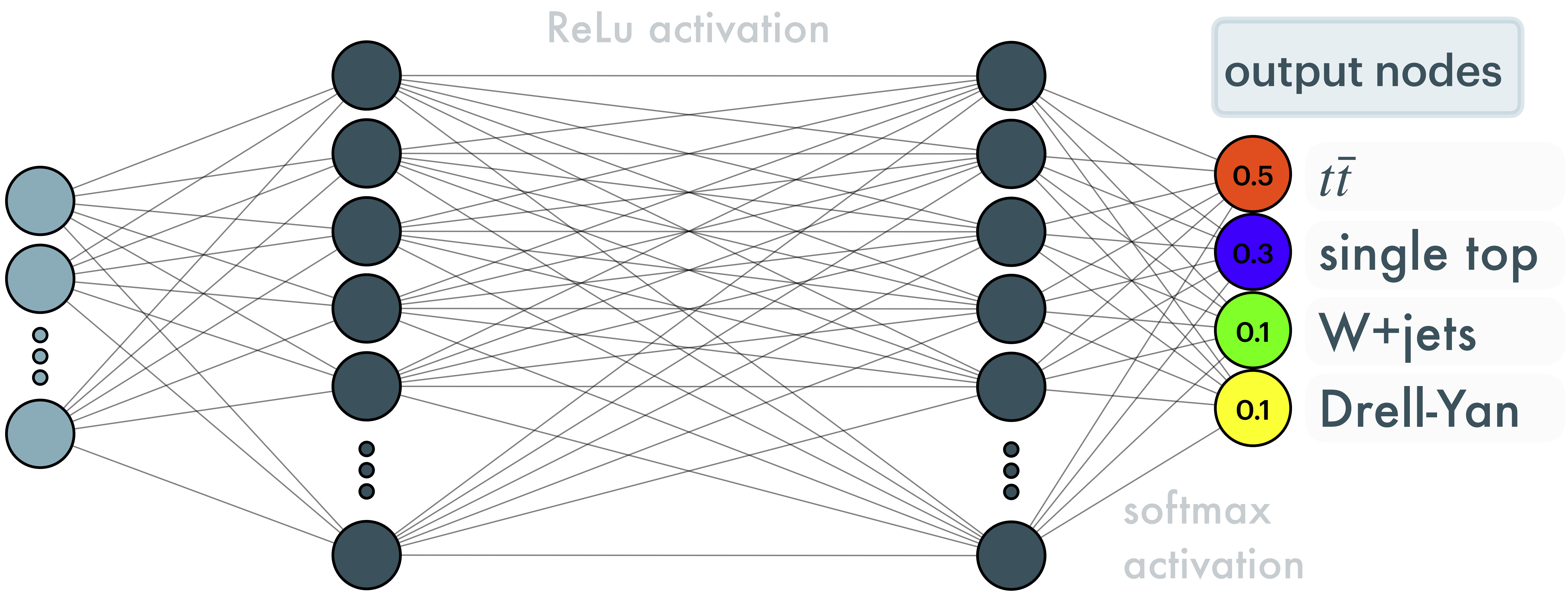


59 input features:

- jet multiplicities
- kinematic properties of objects (MET, jets, lepton)







DNN Parameter

59 input features

jet multiplicity	5 AK4 jets	3 AK8 jets	lepton	MET
N_{AK4}	E	E	E	p_{T}
N_{AK8}	p_{T}	p_{T}	p_{T}	ϕ
	η	η	η	
	ϕ	ϕ	ϕ	
	mass	softdrop mass		
	b-tag score	τ_{21}		
		τ_{32}		

model hyperparameters

- AdamOptimizer
- ReLu activation for hidden layers
- softmax activation for output layers
- batchsize = 32768
- drop out rate = 0.5
- epochs $\leq$ 500
- folds = 5
- learning rate = 0.0005
- validation fraction = 0.25