

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

— **Johanna Matthiesen**¹, Johannes Haller¹, Roman Kogler², Daniel Savoie¹ —

¹Universität Hamburg - ²DESY, Hamburg

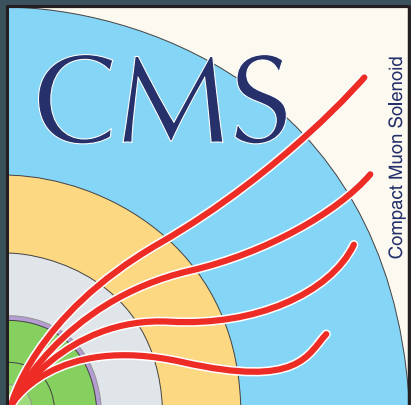


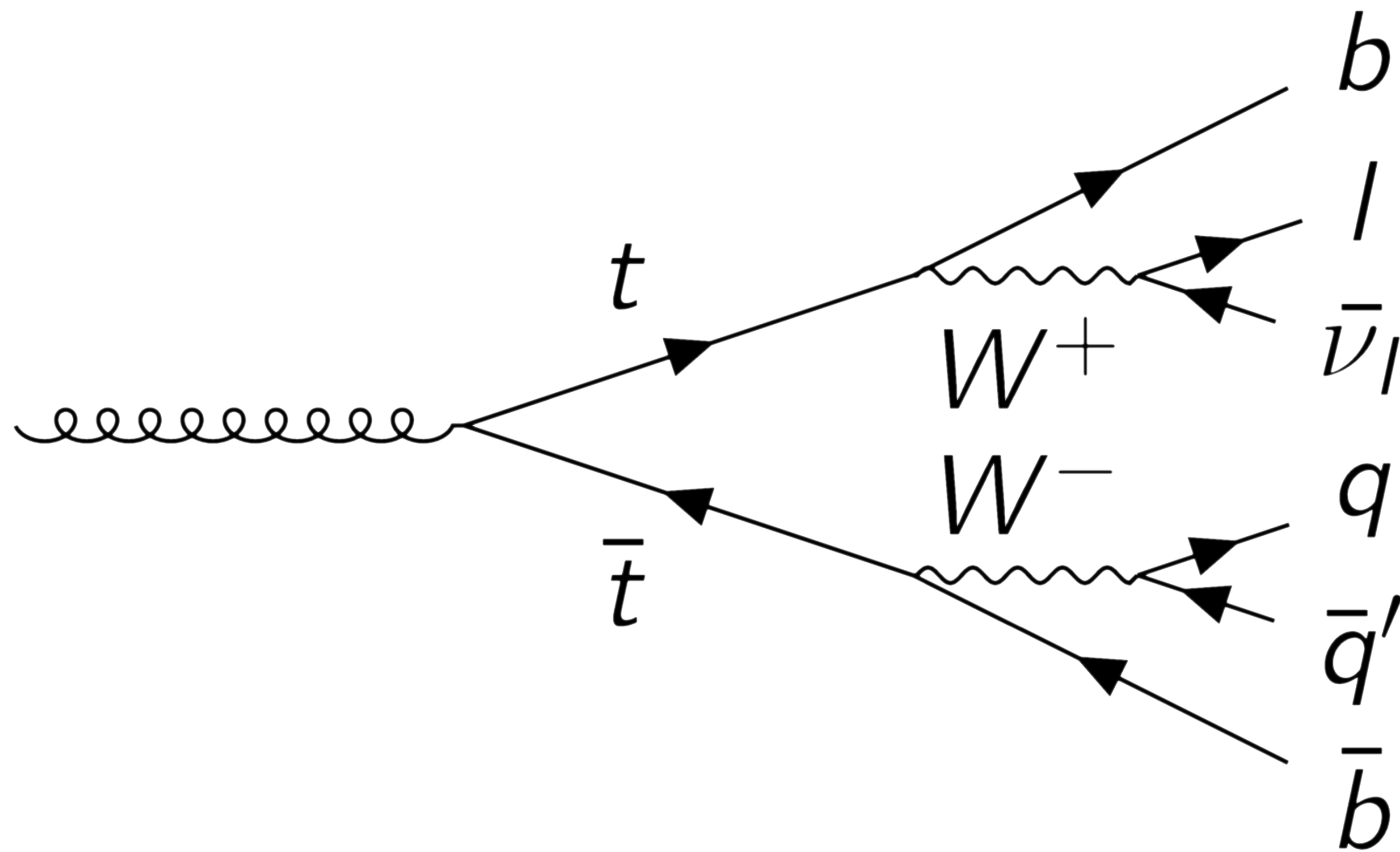
Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

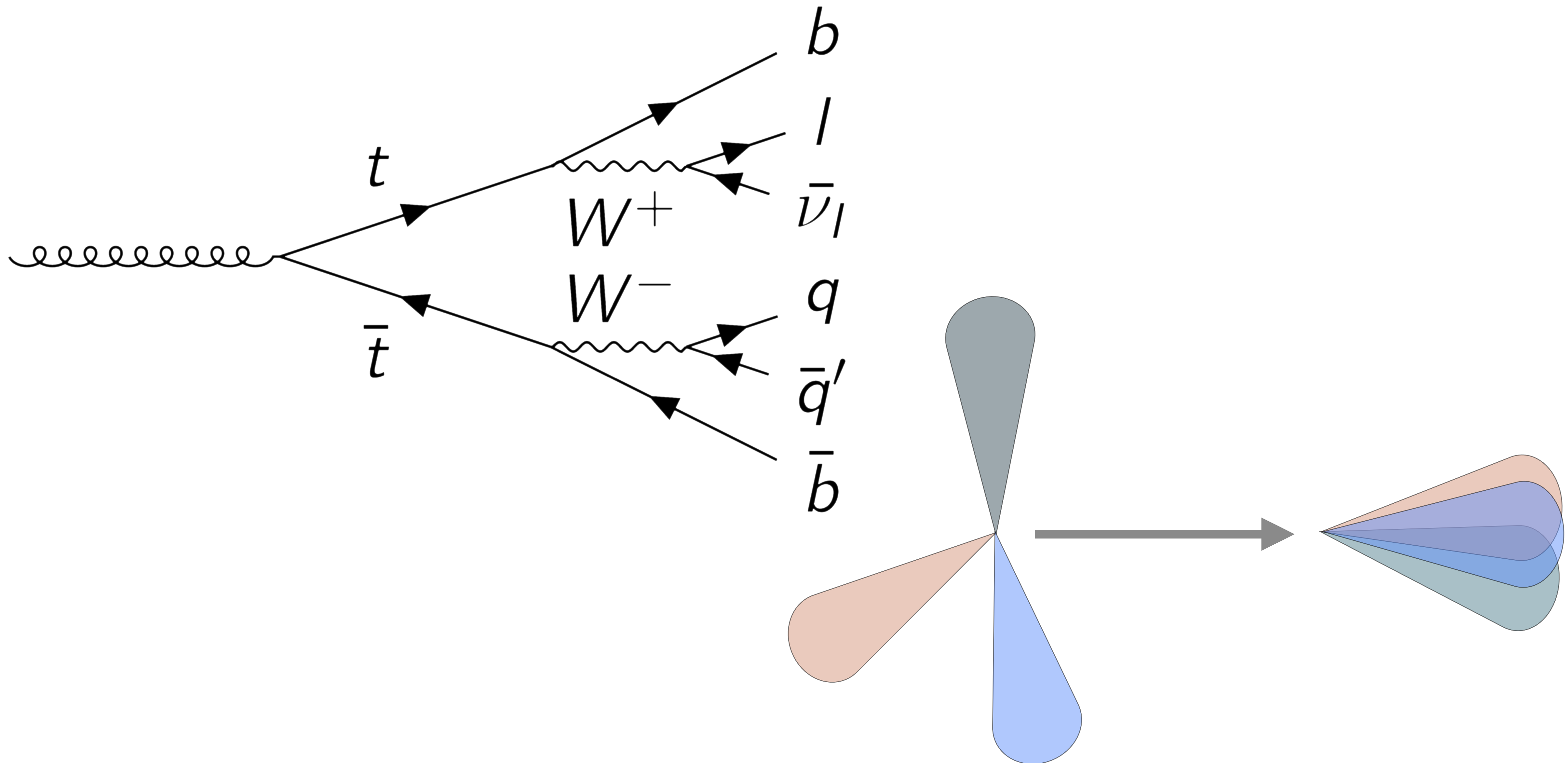
CLUSTER OF EXCELLENCE
QUANTUM UNIVERSE

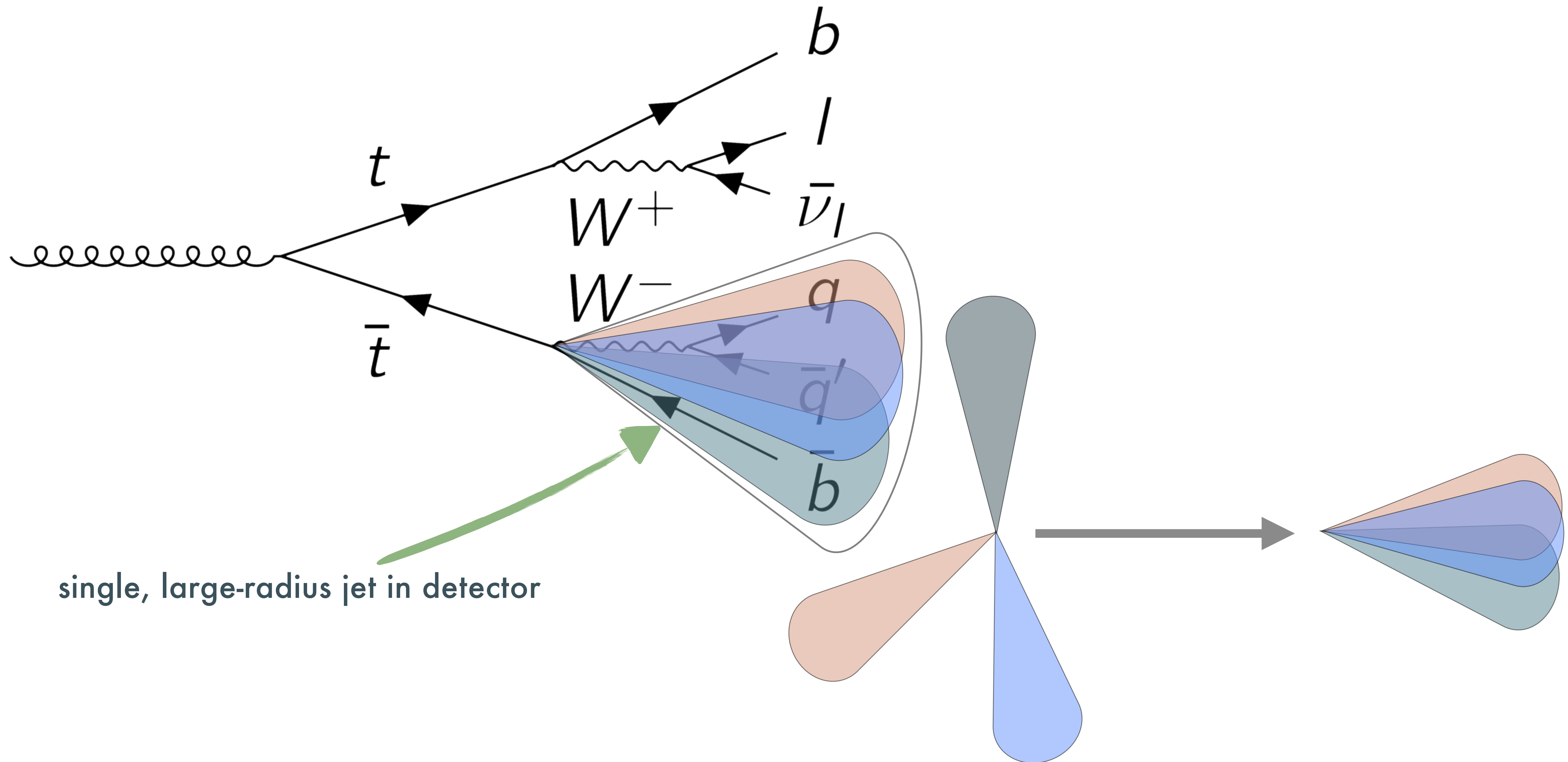


FSP CMS
Erforschung von
Universum und Materie



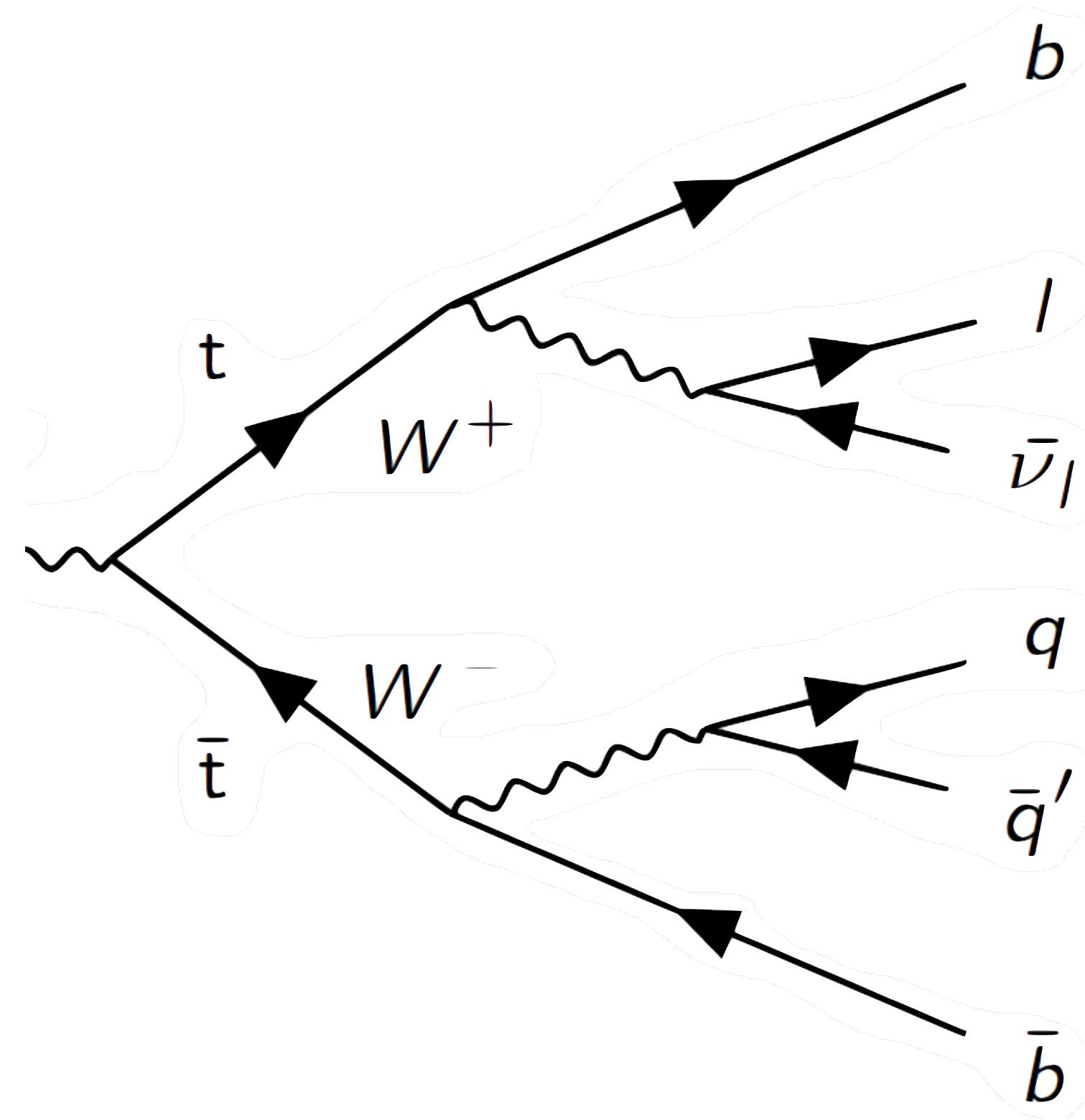






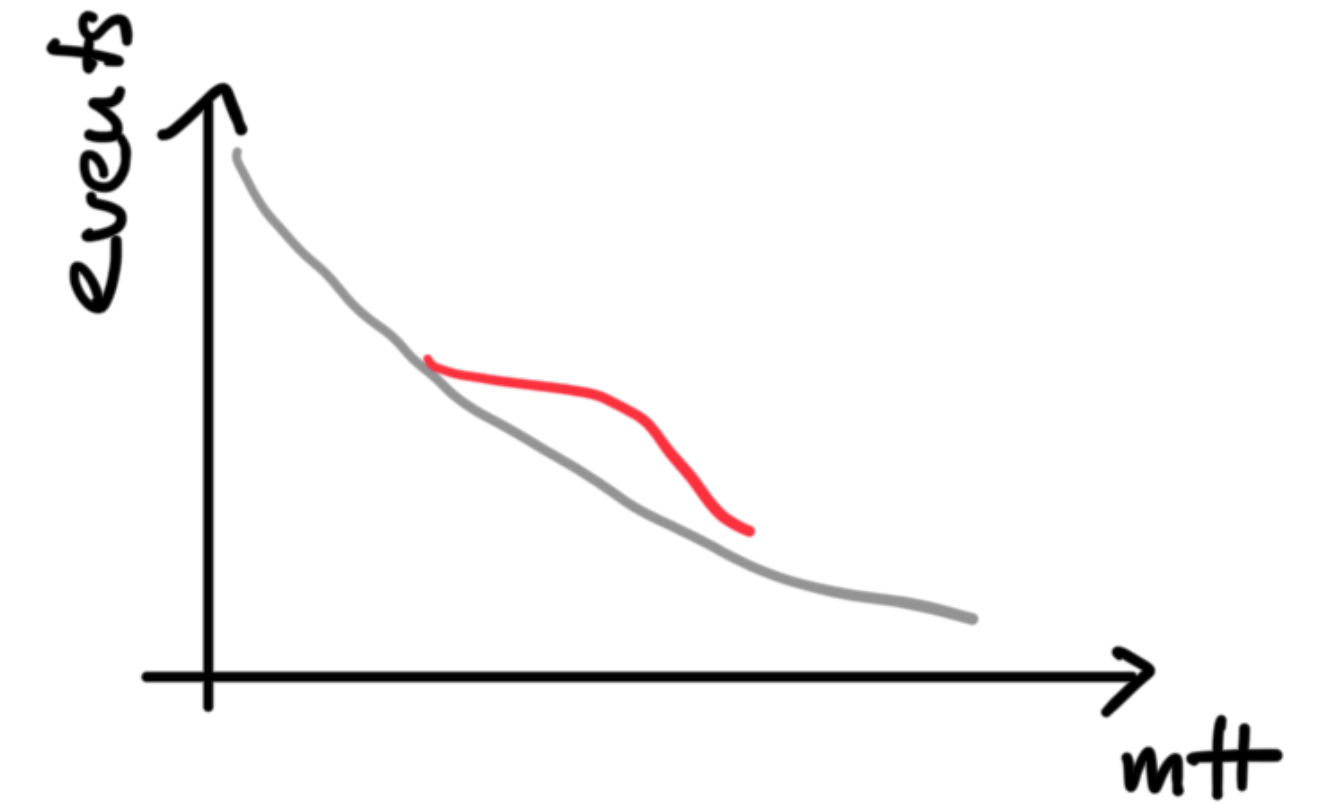
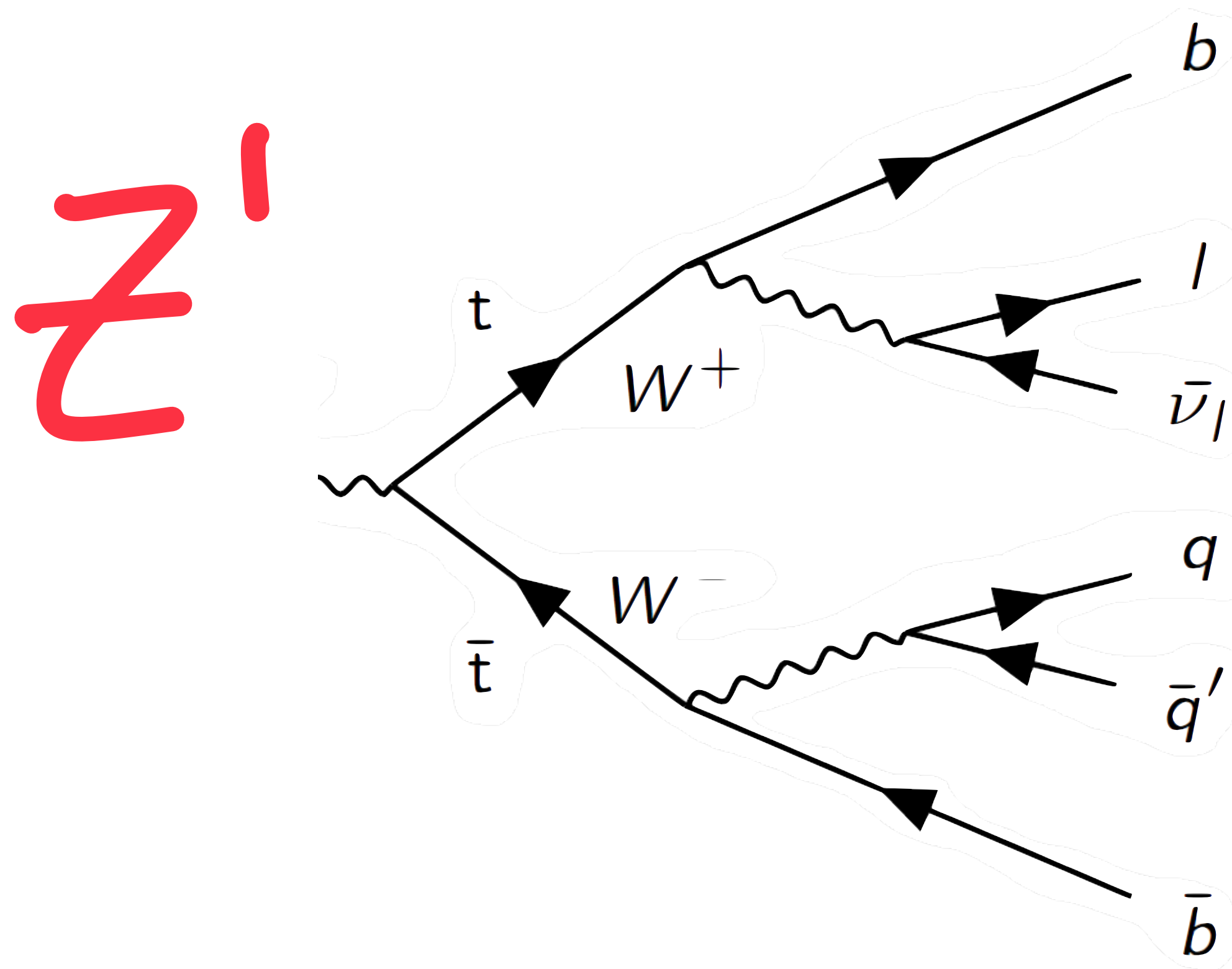
Boosted Top Quarks

— in BSM signals —



Boosted Top Quarks

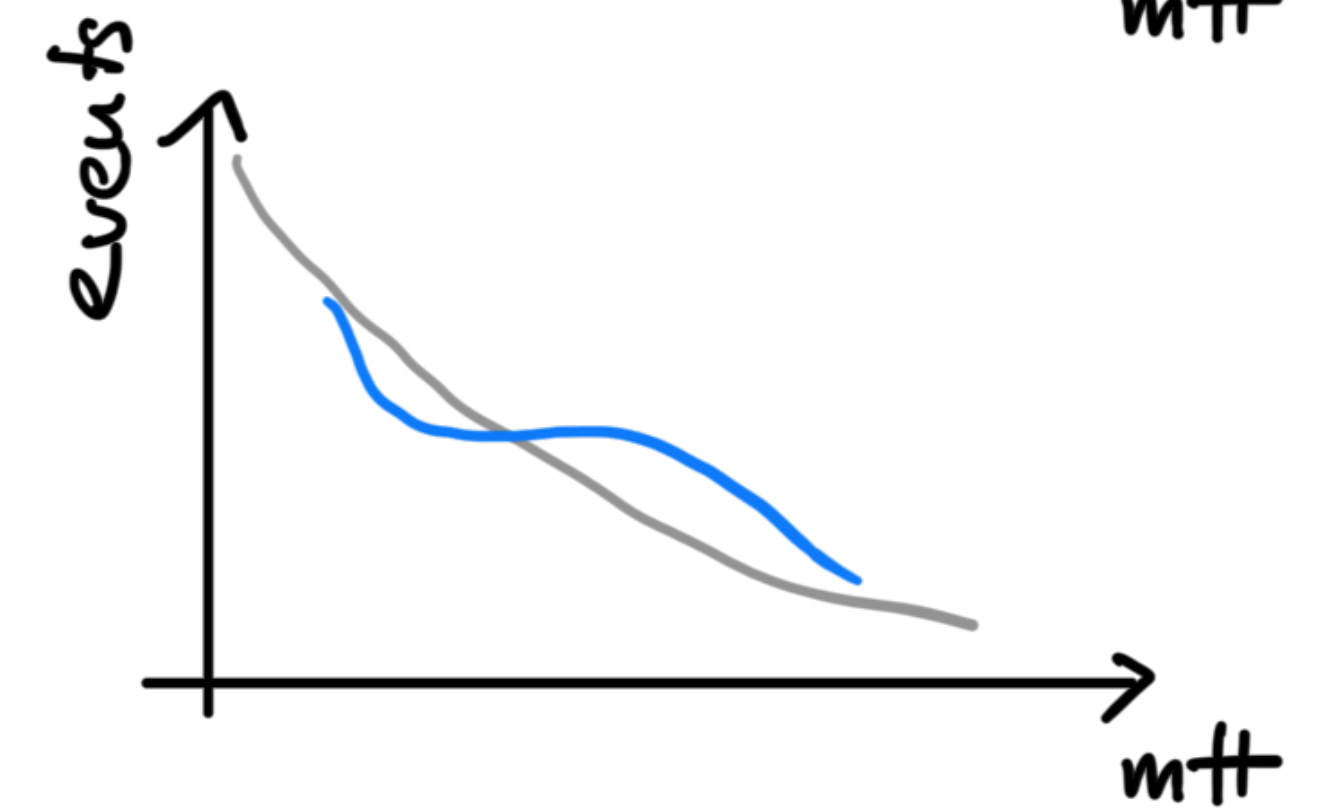
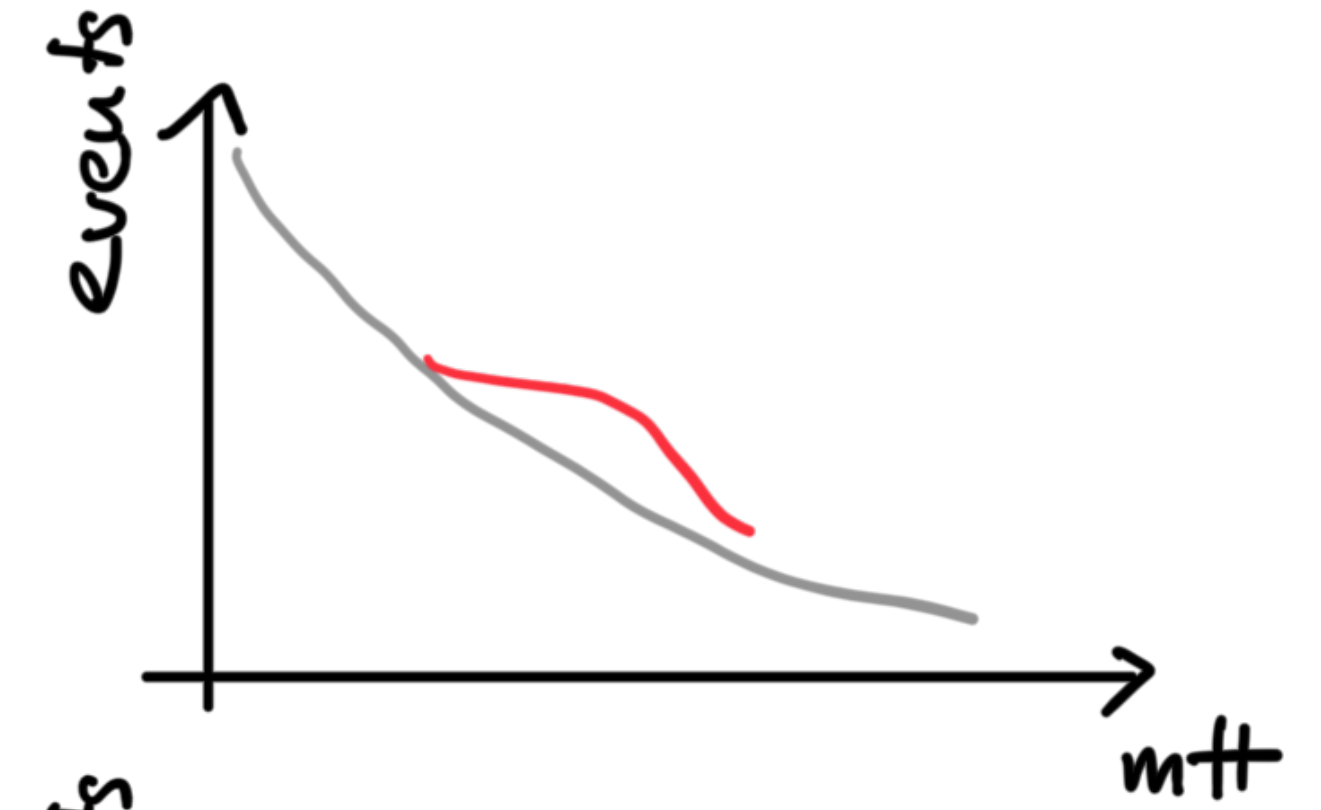
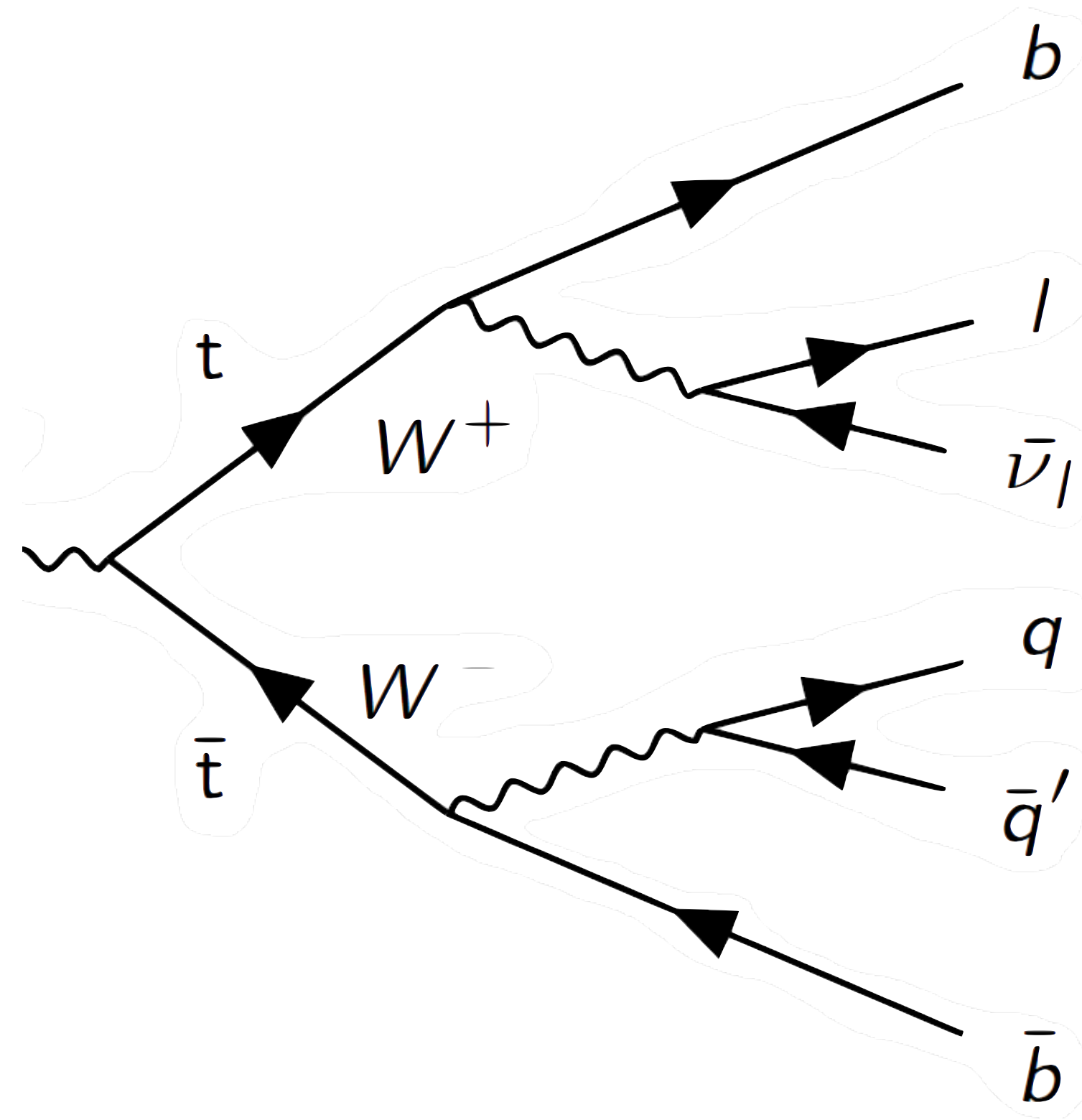
— in BSM signals —



Boosted Top Quarks

— in BSM signals —

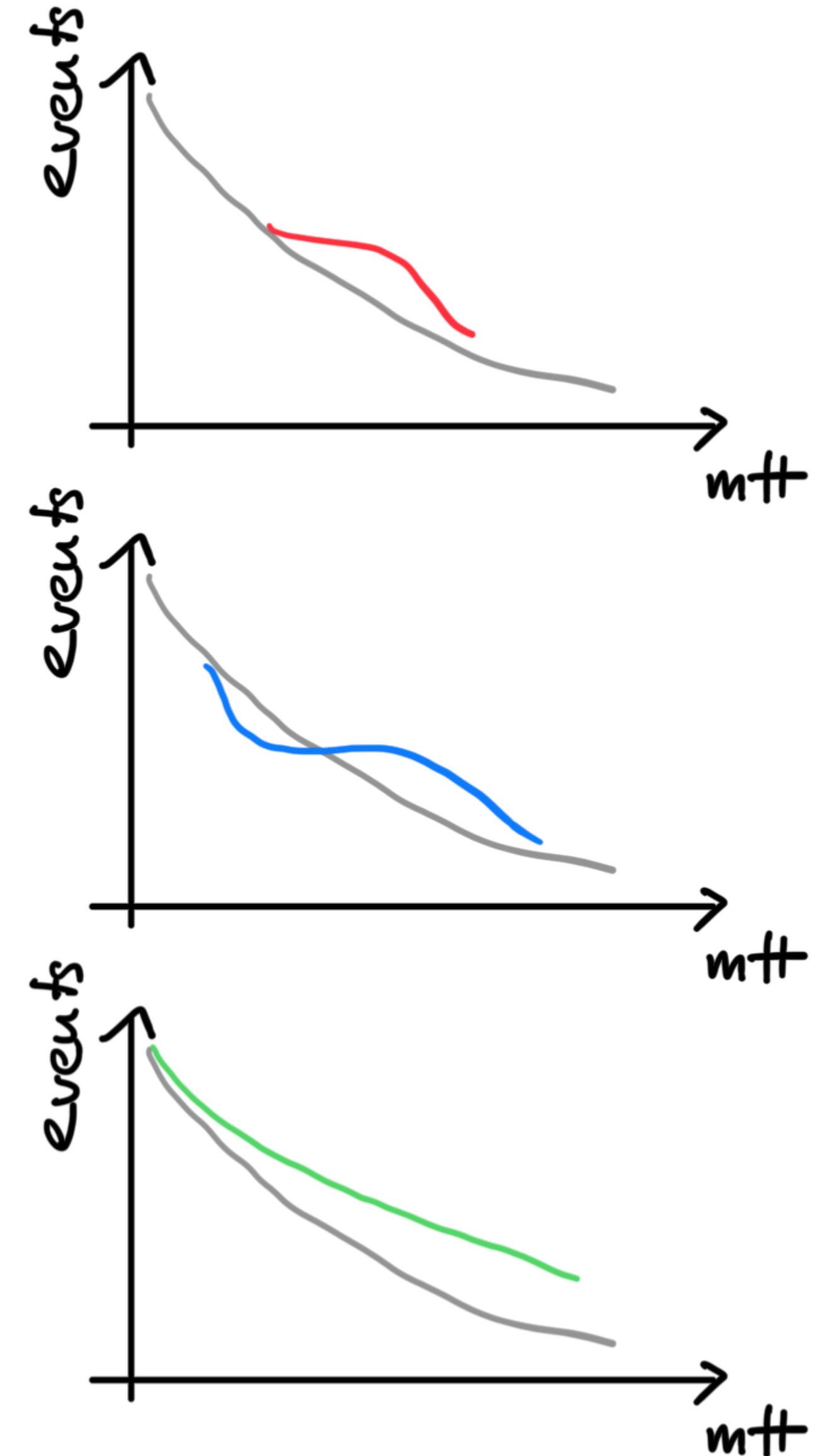
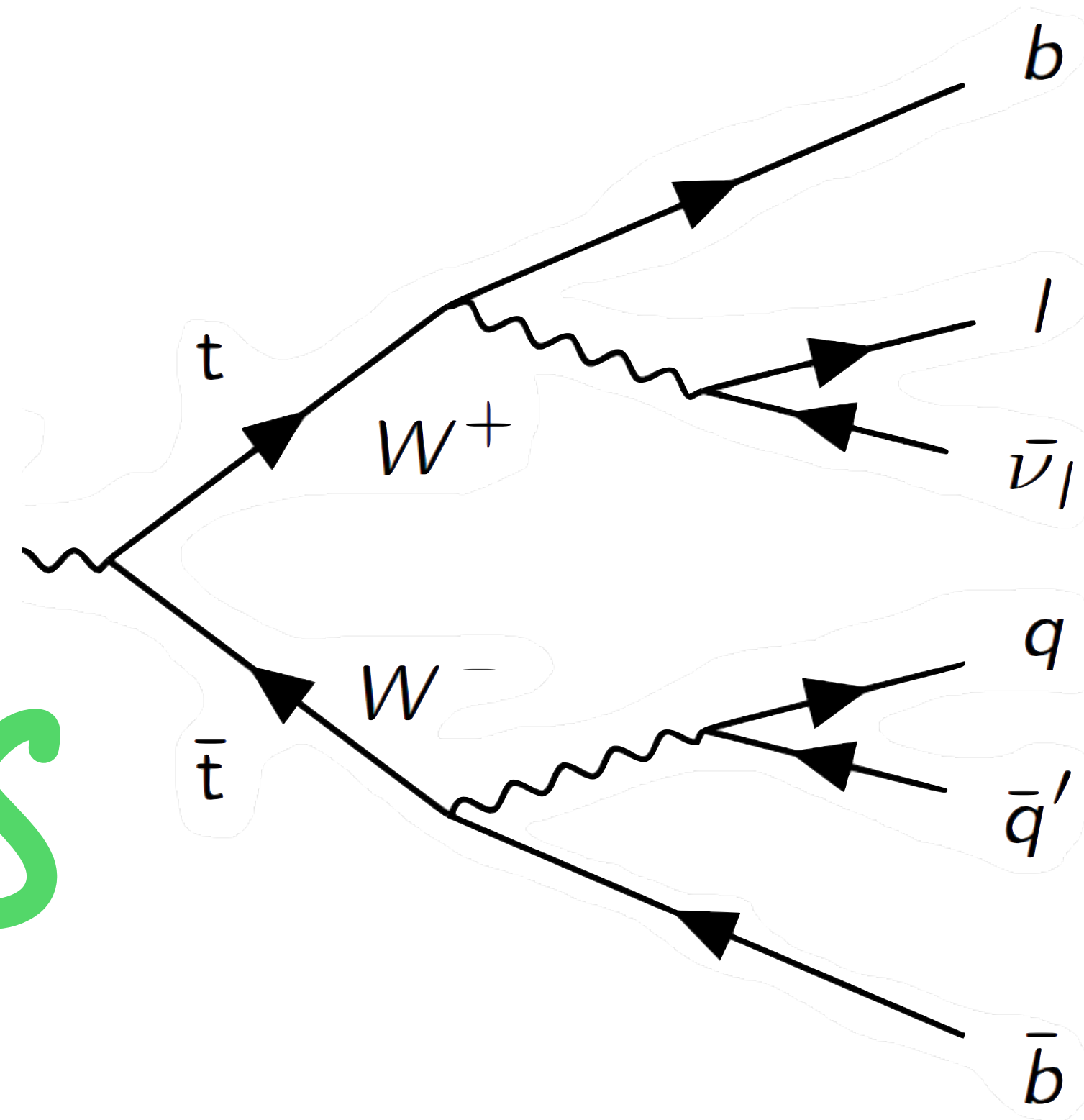
A/H Z'



Boosted Top Quarks

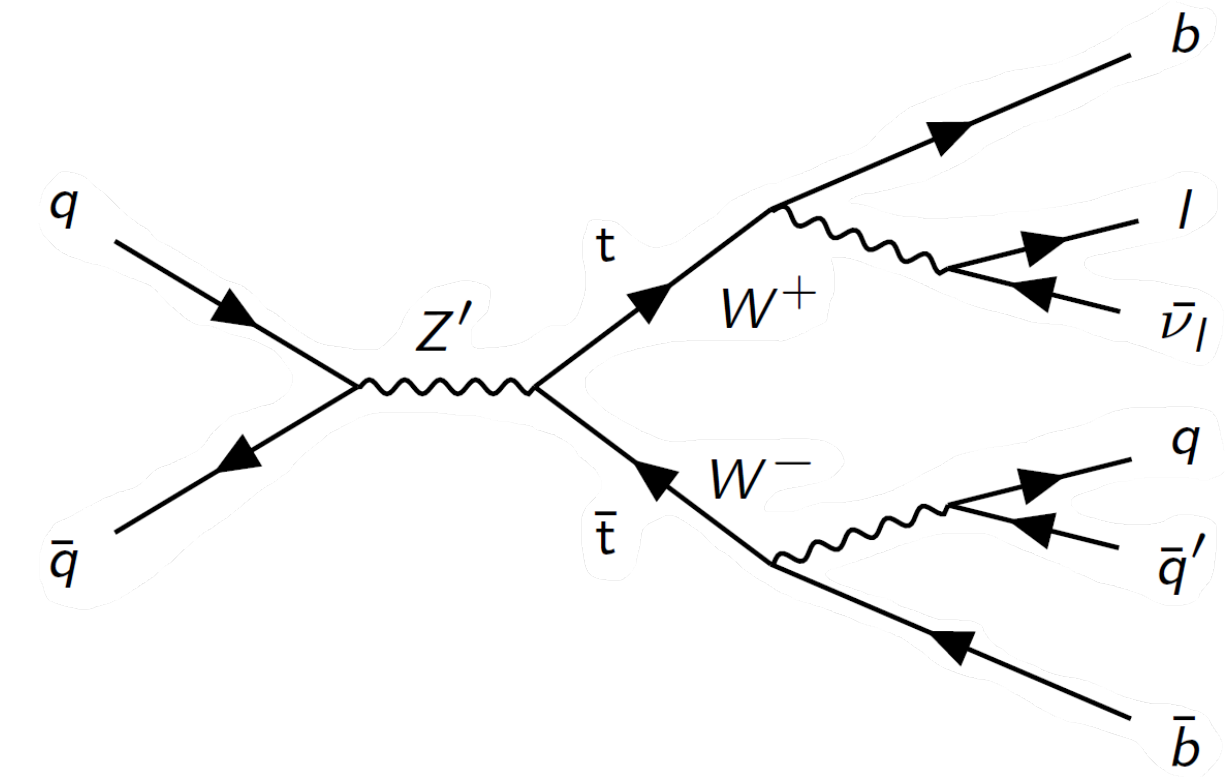
— in BSM signals —

A/H Z'
ALPS



Search for Heavy Z Like Boson

— analysis overview —



Search for Heavy Z Like Boson

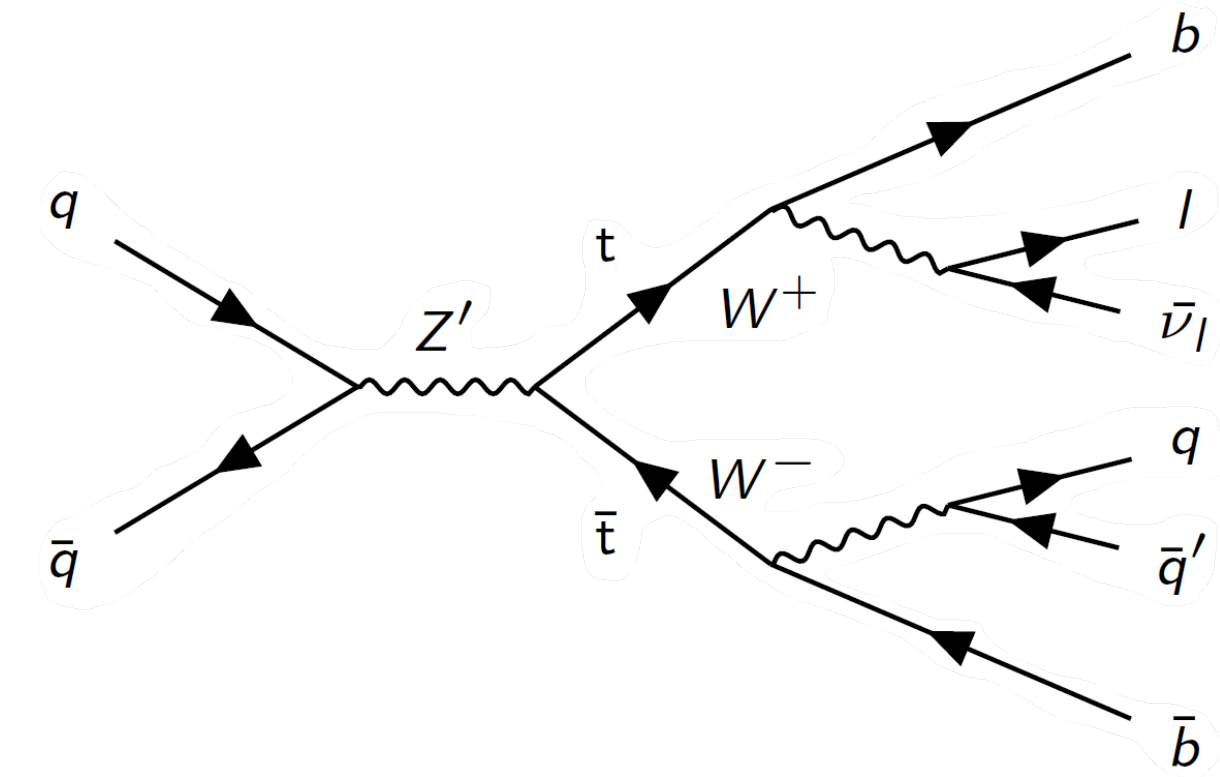
— analysis overview —

I. Selection

II. Reconstruction

III. Event Classification

IV. Statistical Analysis



Search for Heavy Z Like Boson

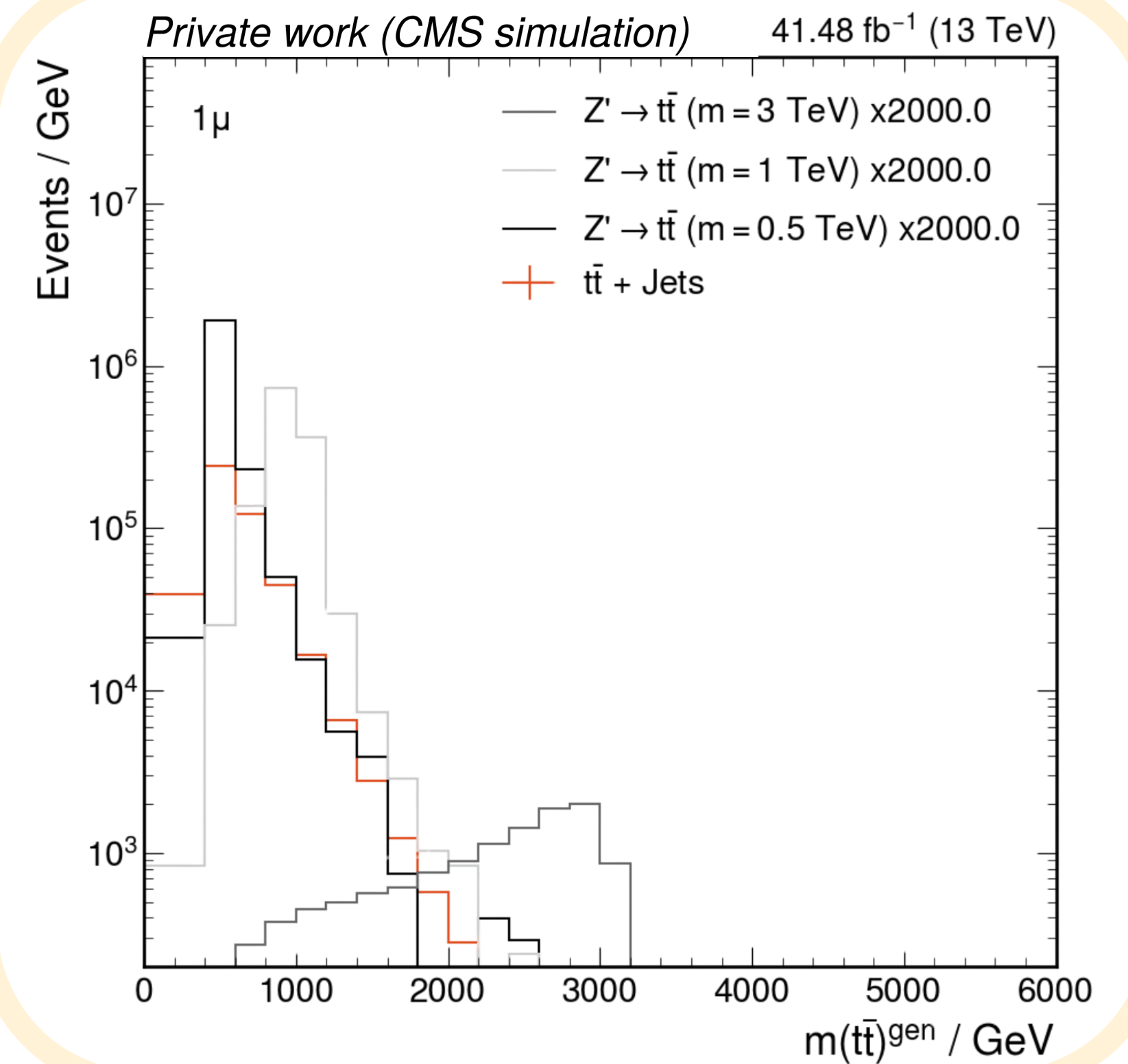
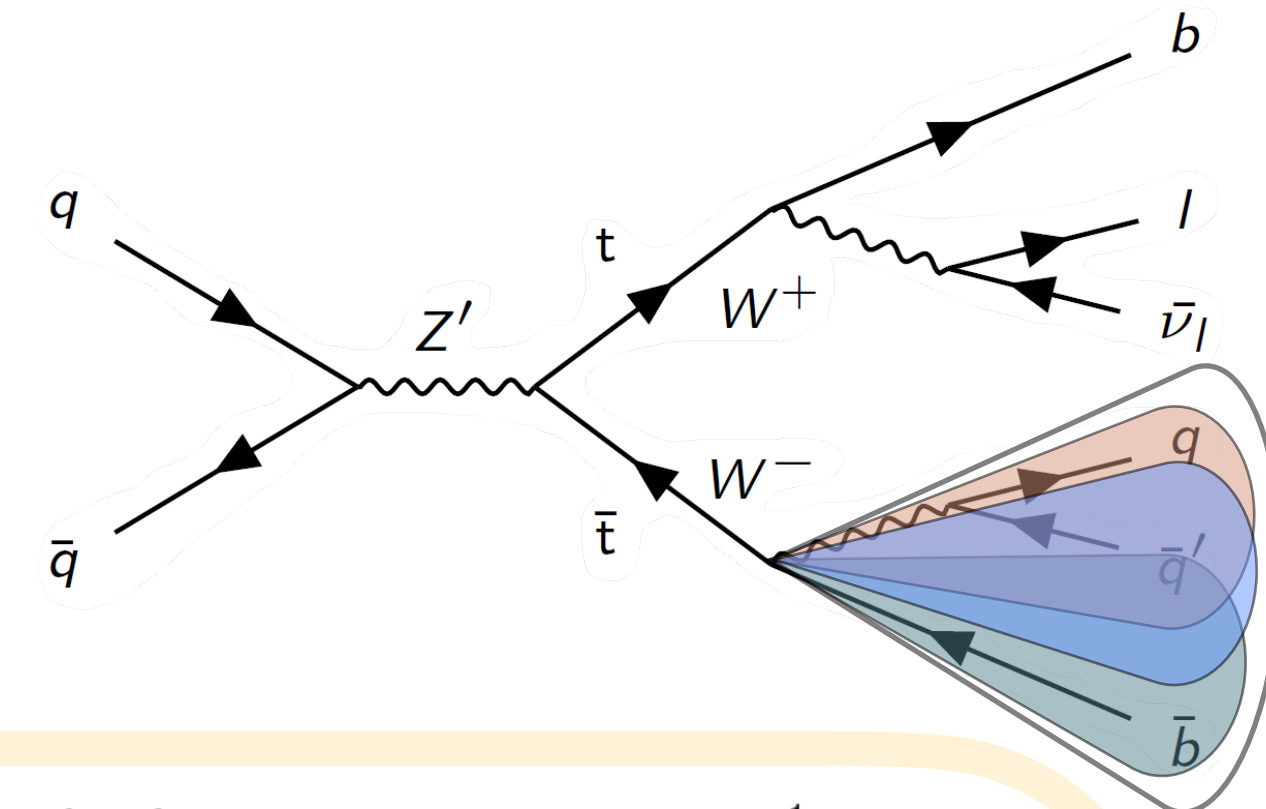
— analysis overview —

I. Selection

II. Reconstruction

III. Event Classification

IV. Statistical Analysis



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-2020-025

N-subjettiness

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

JHEP 1103:015,2011 - [arXiv:1011.2268](https://arxiv.org/abs/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})

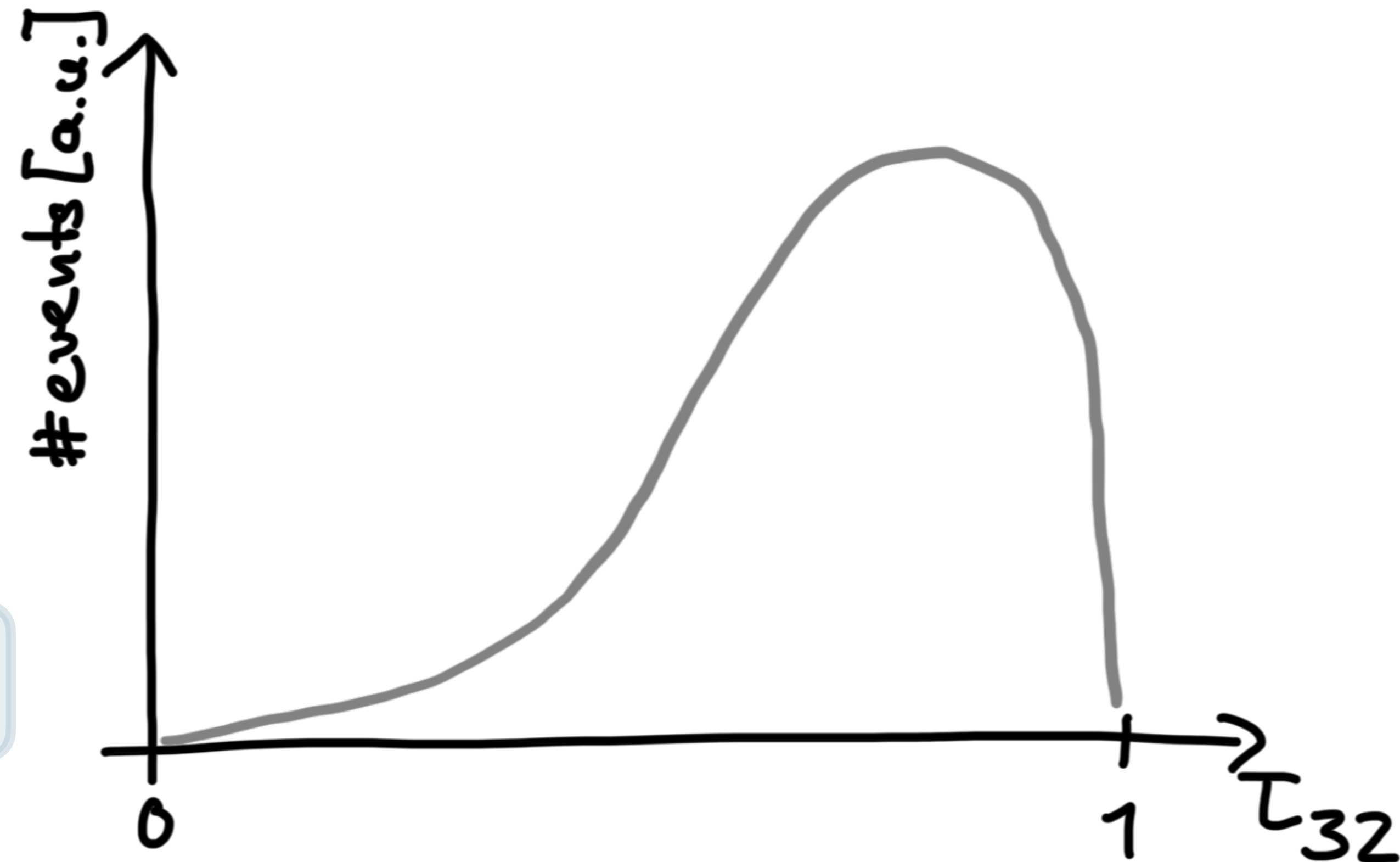
CMS-DP-2020-025

N-subjettiness

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

JHEP 1103:015,2011 - [arXiv:1011.2268](https://arxiv.org/abs/1011.2268)

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



Identifying Boosted Top Quarks

— in CMS —

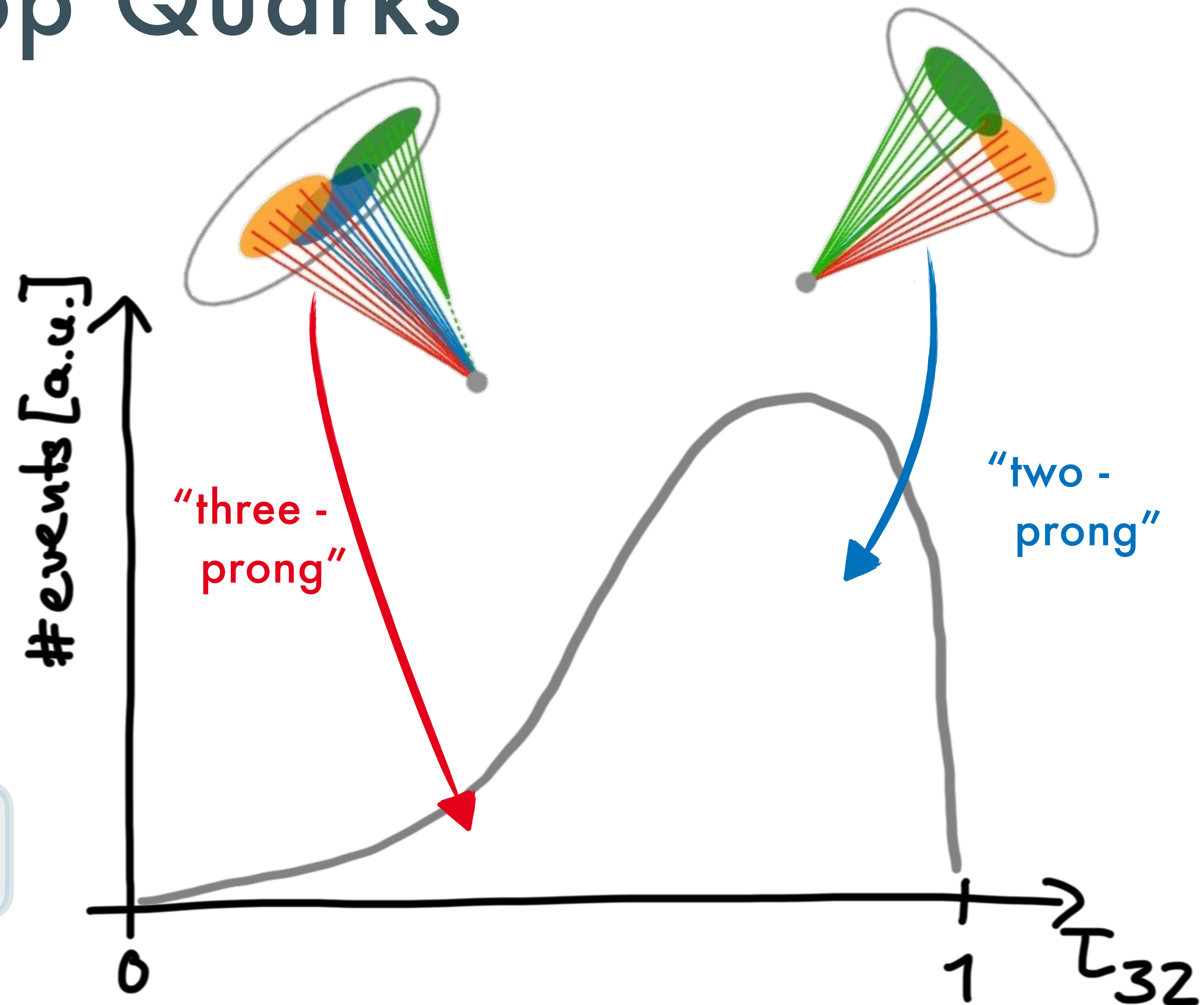
- different approaches to top tagging:
 - ▶ **machine learning** algorithms
DeepAK8, ParticleNet, ...
JINST 15 (2020) P06005 CMS-DP-2020-002
 - ▶ **cut based** using jet substructure techniques
N-subjettiness ratio (τ_{32}) & soft drop mass (m_{SD})
CMS-DP-2020-025

N-subjettiness

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

JHEP 1103:015,2011 - [arXiv:1011.2268](https://arxiv.org/abs/1011.2268)

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



Identifying Boosted Top Quarks

— in CMS —

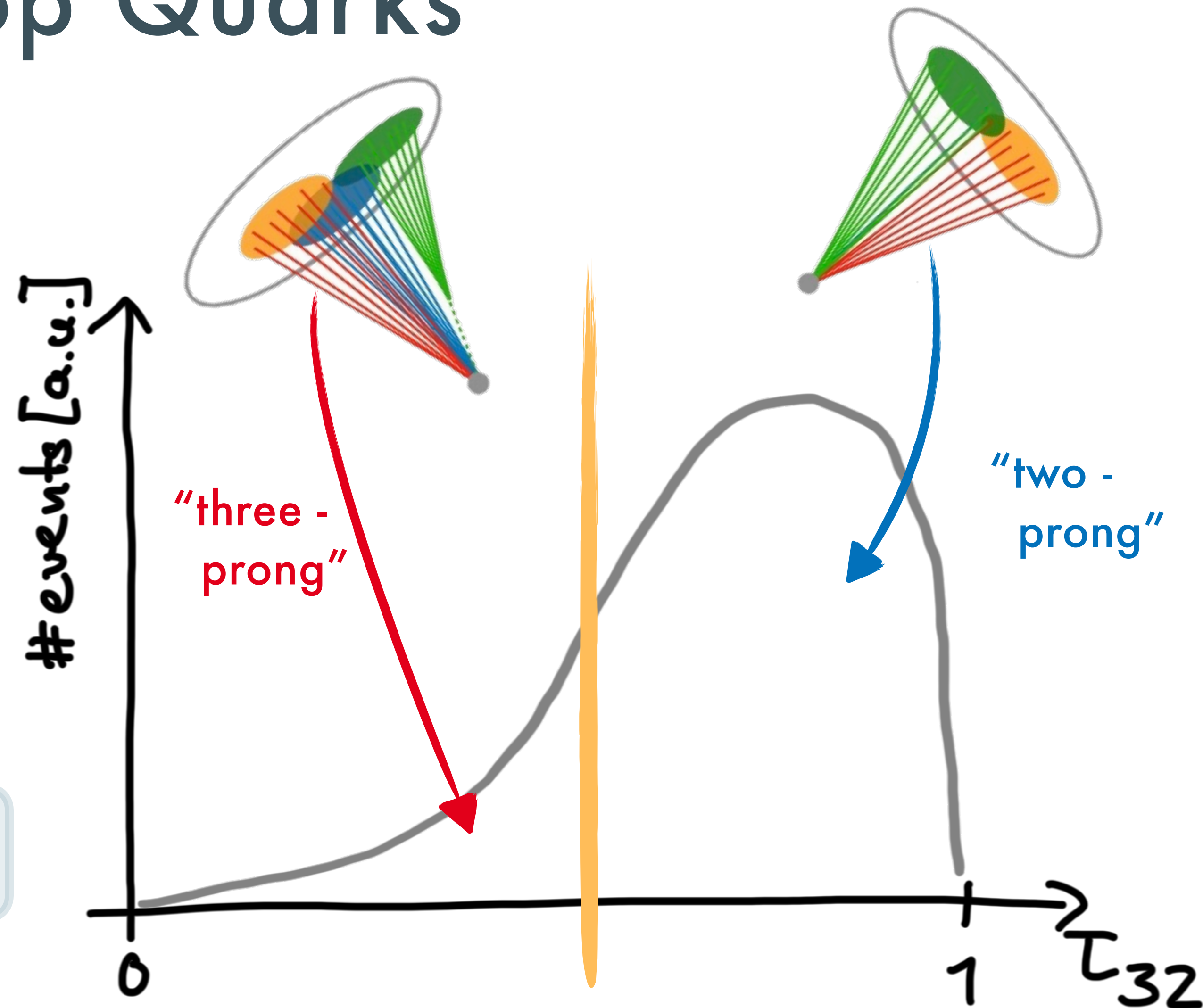
- different approaches to top tagging:
 - ▶ **machine learning** algorithms
DeepAK8, ParticleNet, ...
JINST 15 (2020) P06005 CMS-DP-2020-002
 - ▶ **cut based** using jet substructure techniques
N-subjettiness ratio (τ_{32}) & soft drop mass (m_{SD})
CMS-DP-2020-025

N-subjettiness

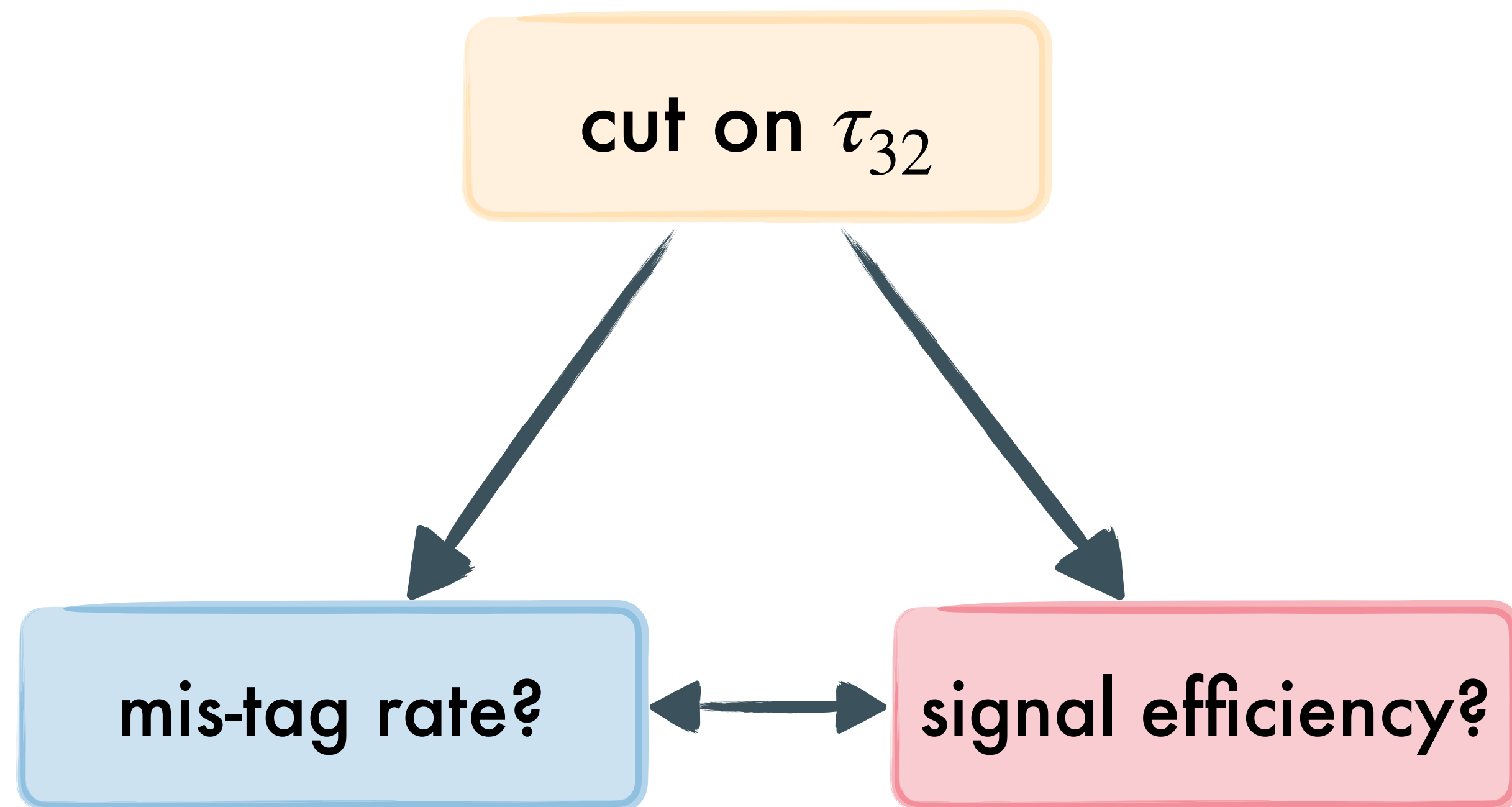
$$\tau_N = \frac{1}{d_0} \sum_k p_{T,k} \min(\Delta R_{1,k}, \Delta R_{2,k}, \dots, \Delta R_{N,k})$$

JHEP 1103:015,2011 - [arXiv:1011.2268](https://arxiv.org/abs/1011.2268)

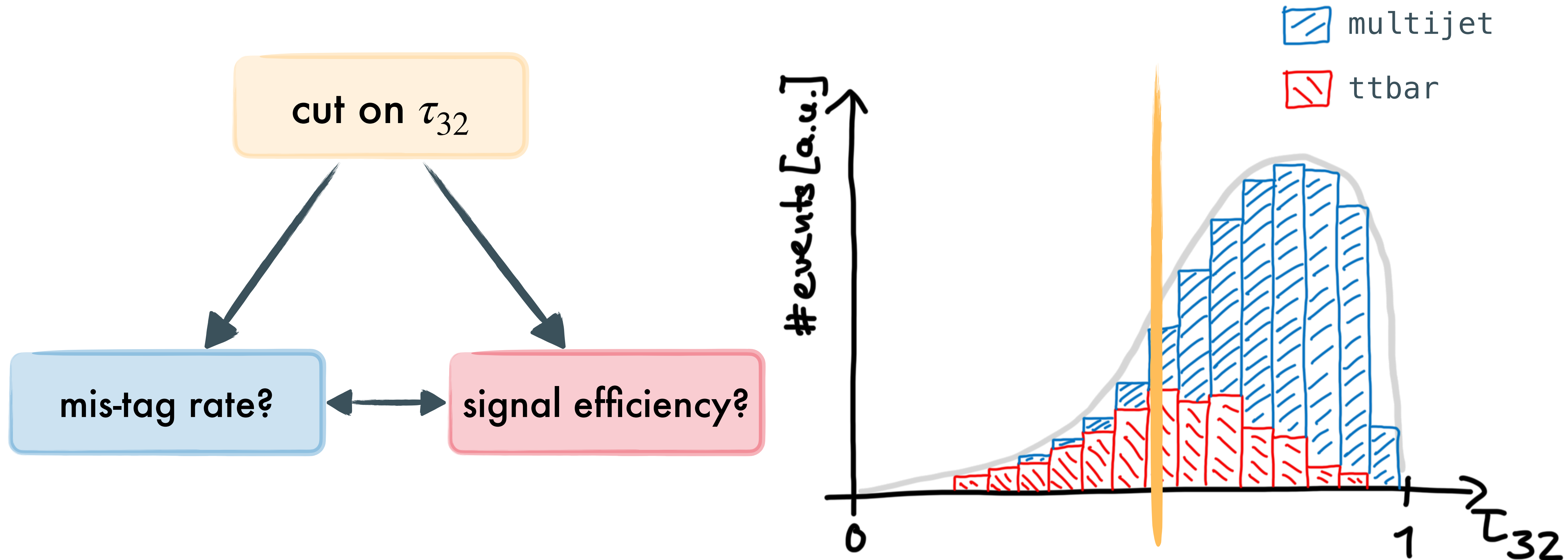
$$\tau_{32} := \tau_3 / \tau_2$$



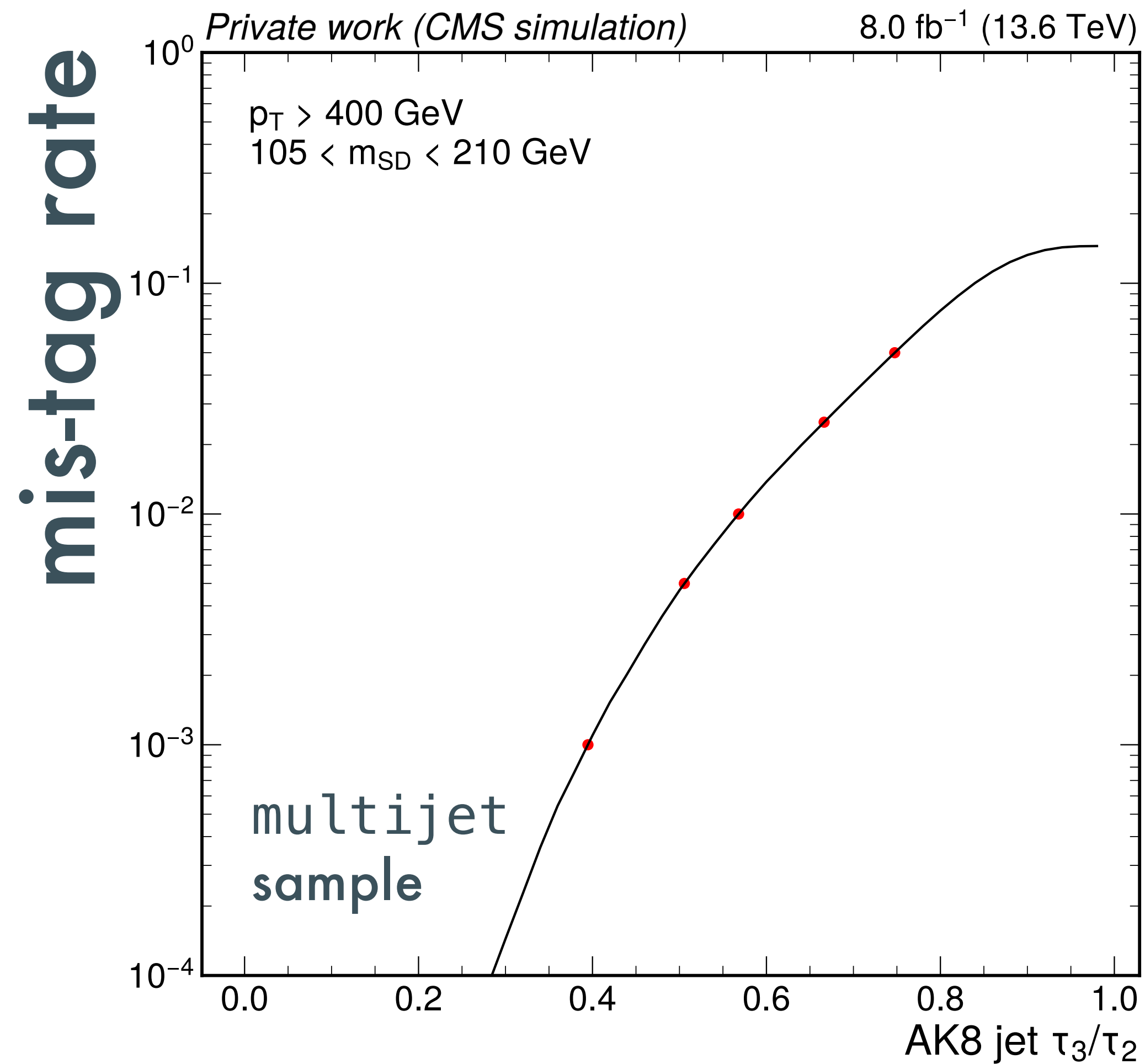
Classifying the Cut on τ_{32}



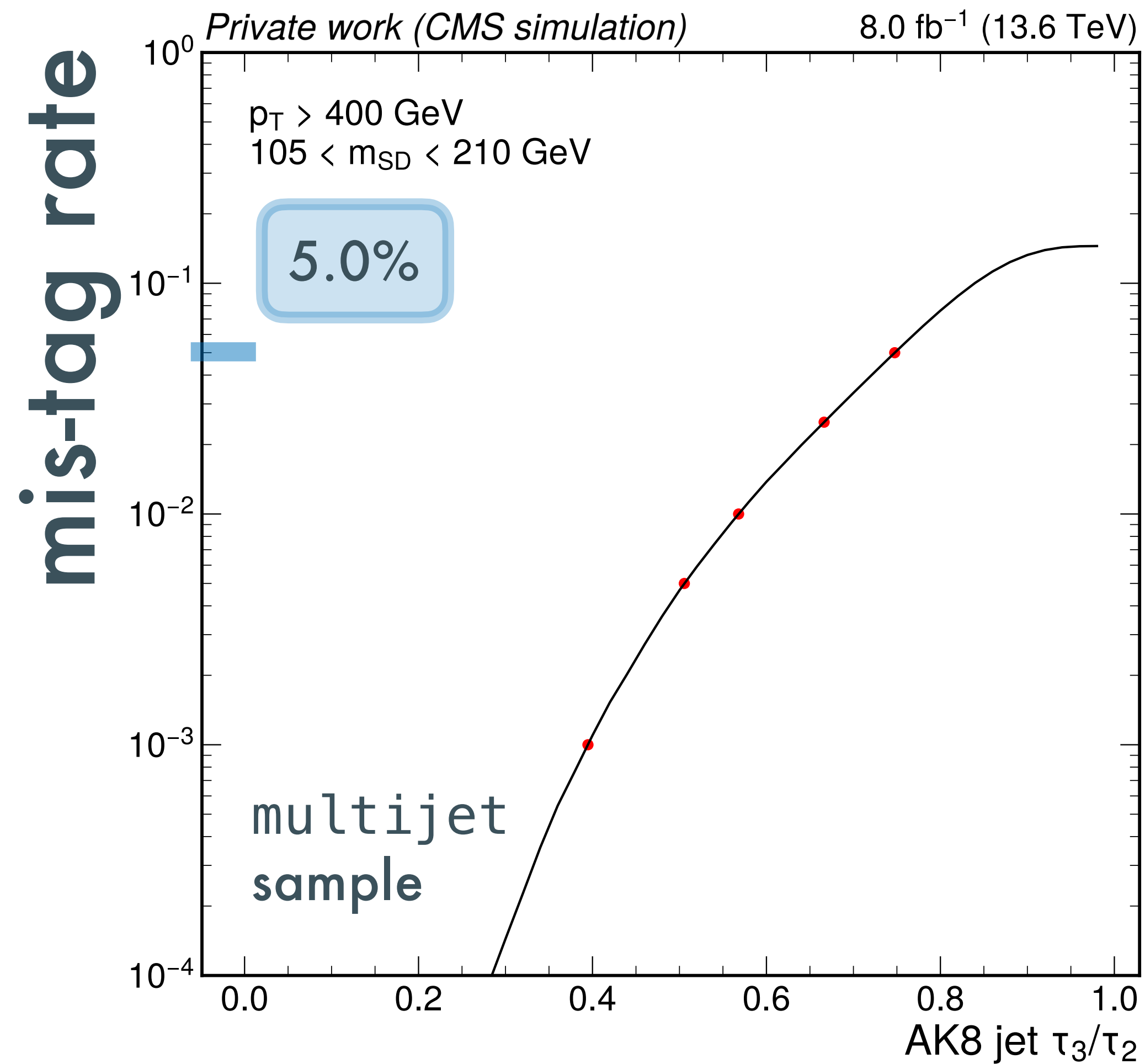
Classifying the Cut on τ_{32}



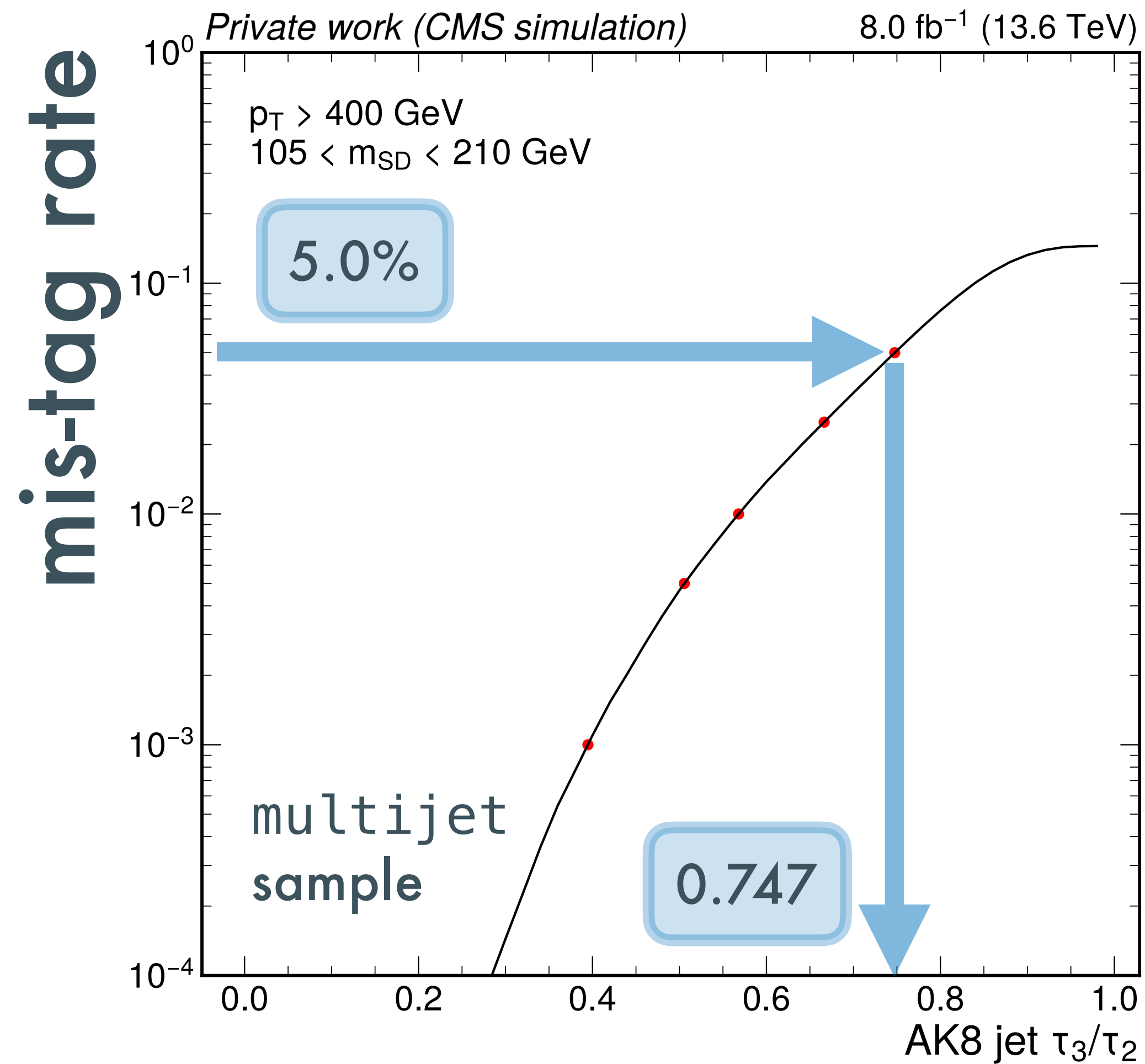
Working Points in One Slide



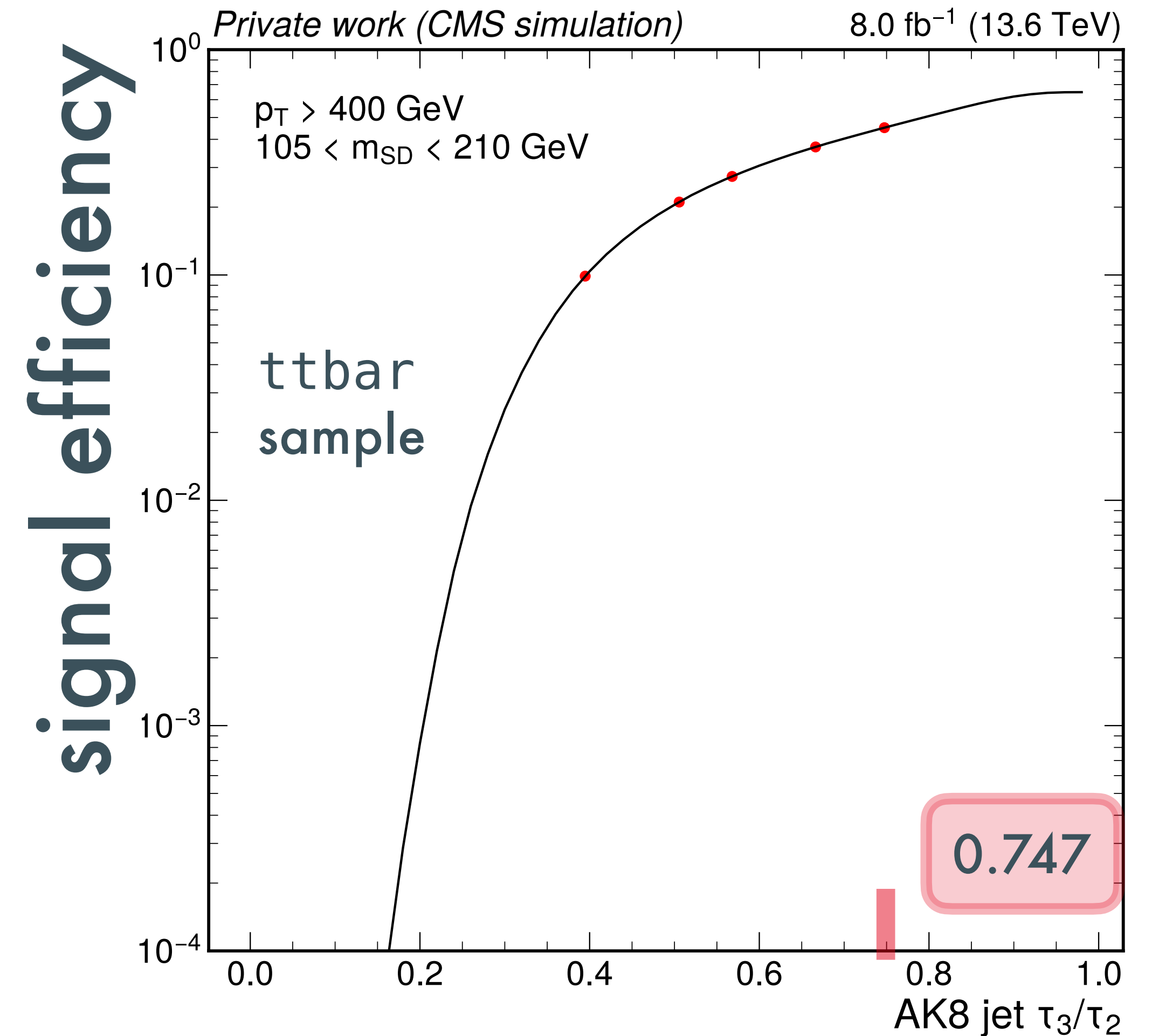
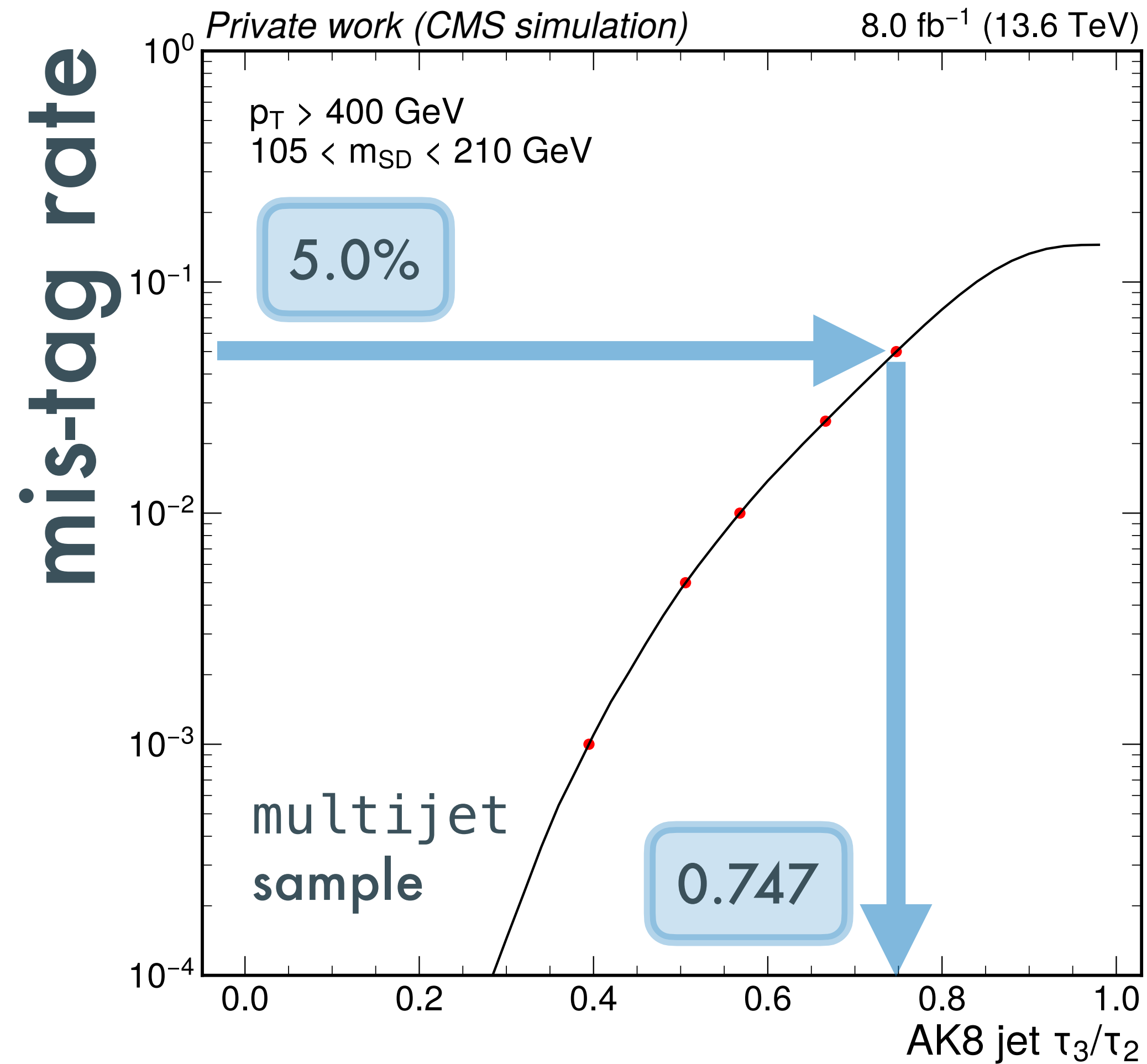
Working Points in One Slide



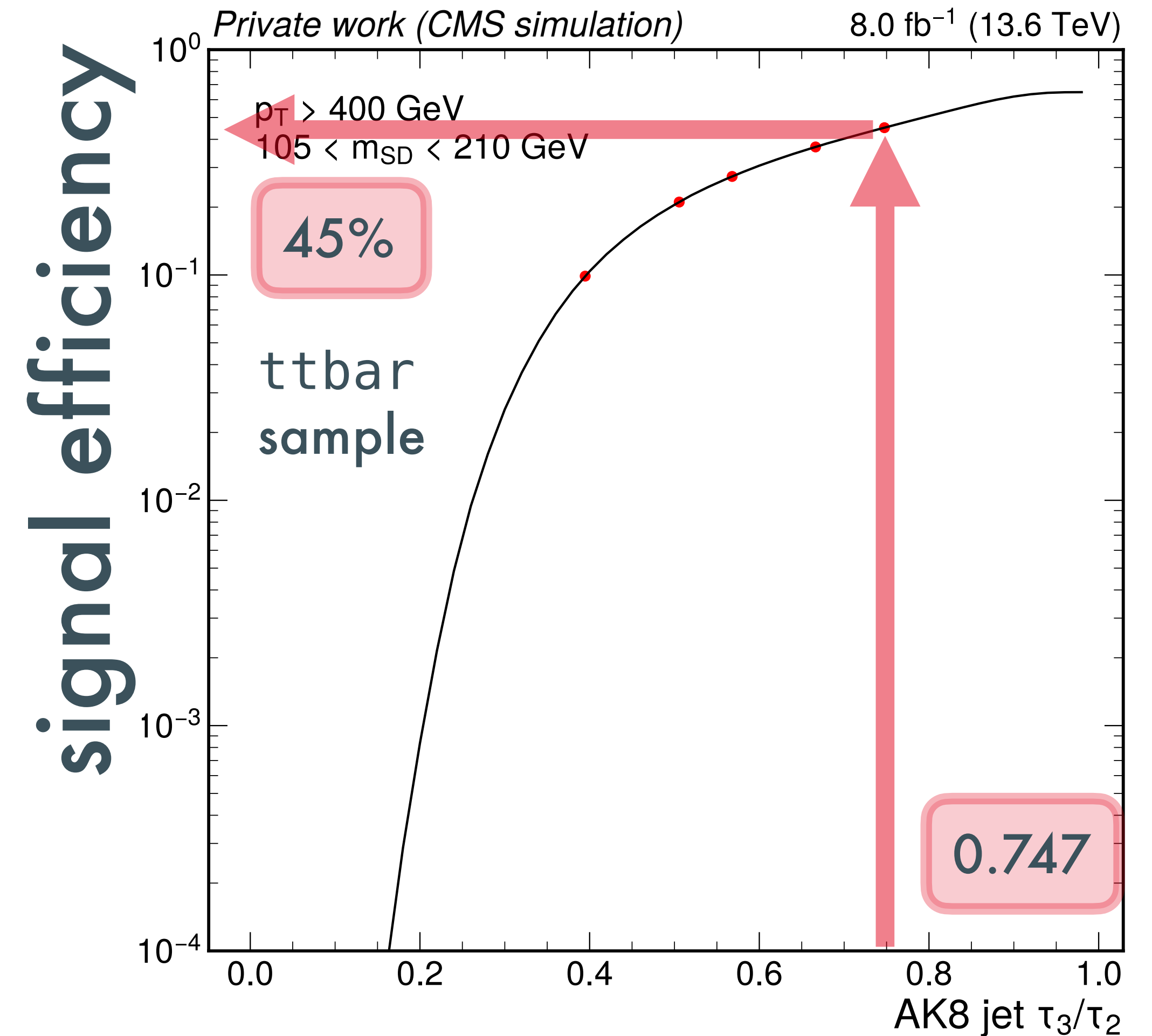
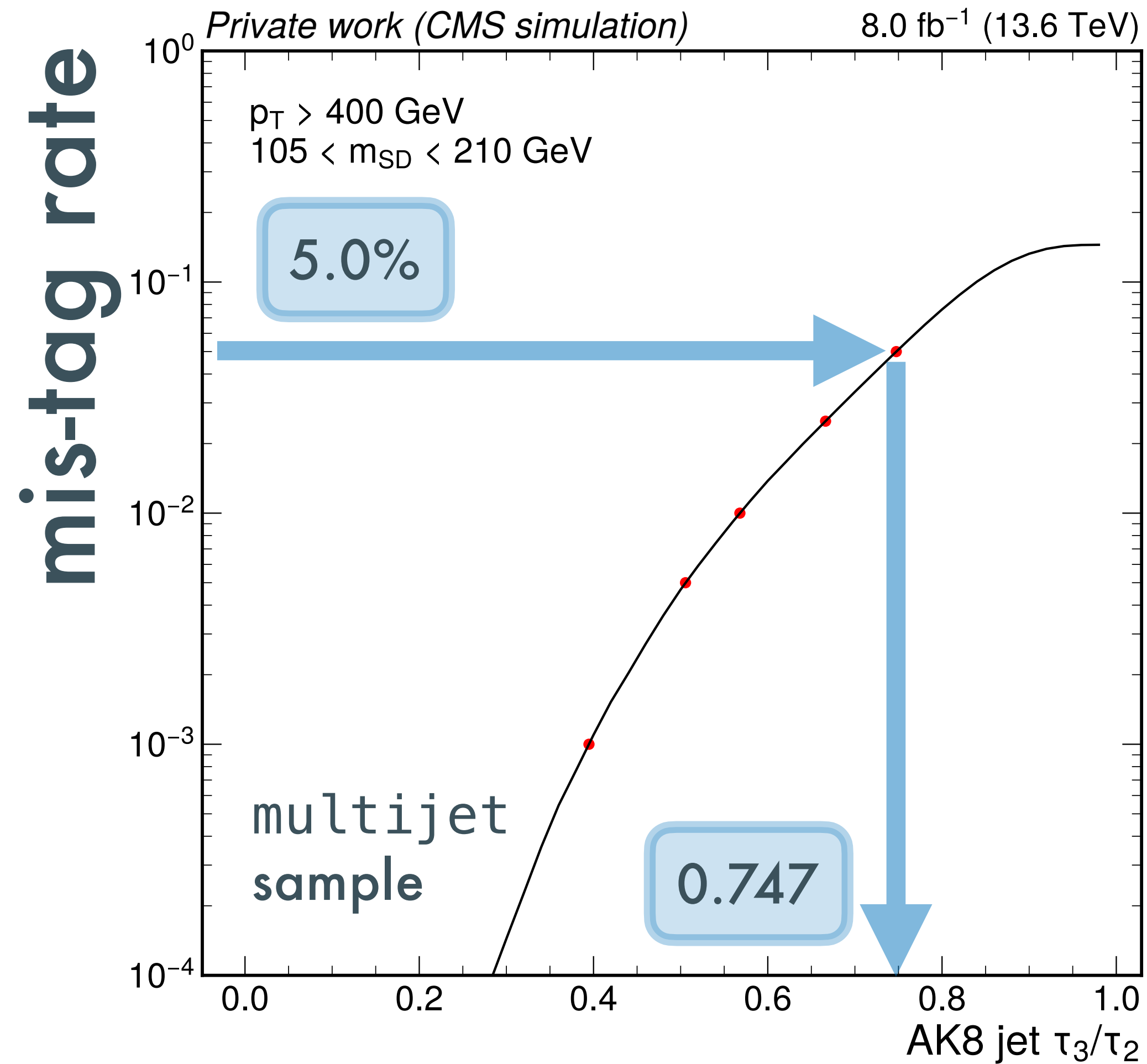
Working Points in One Slide



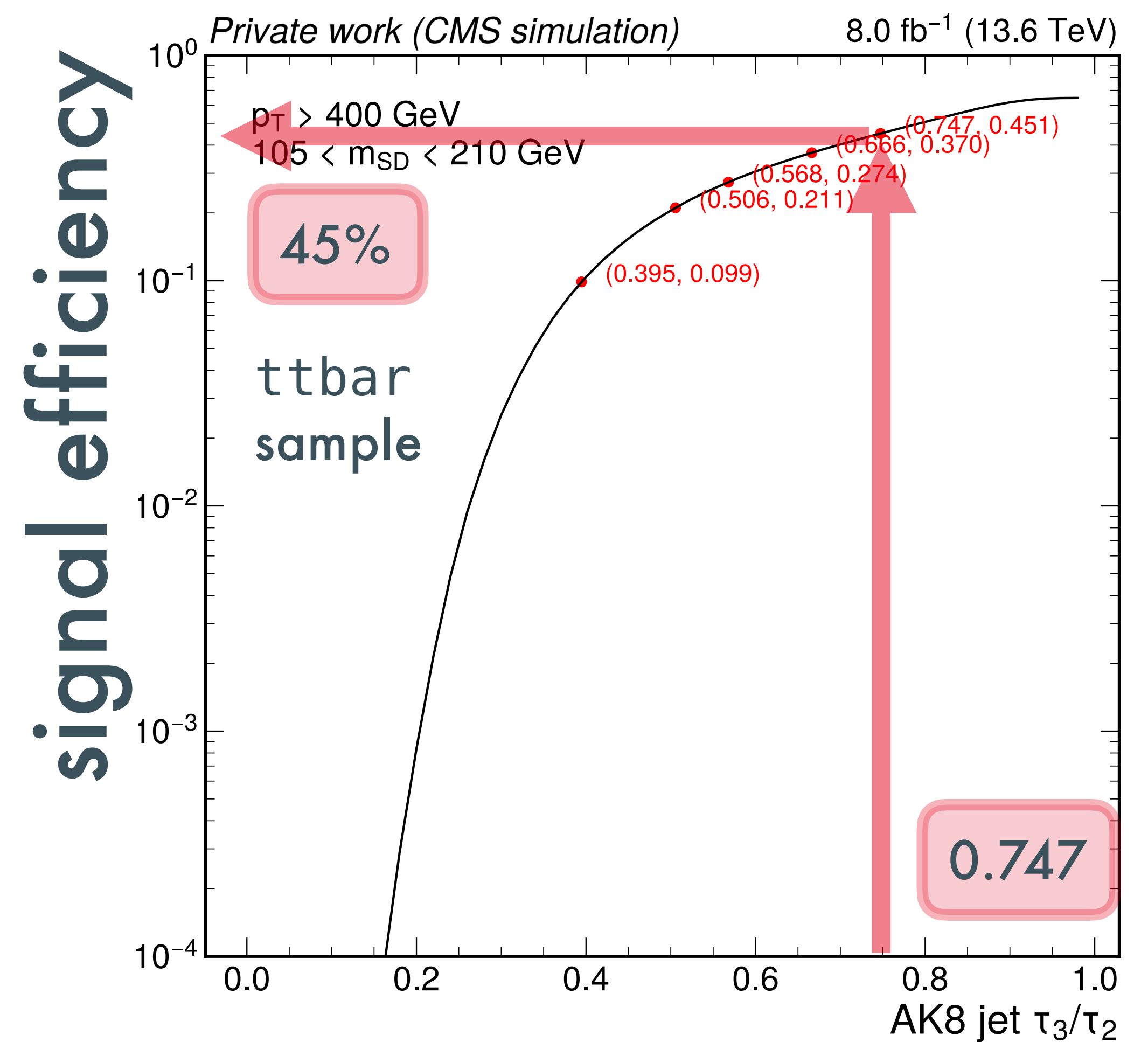
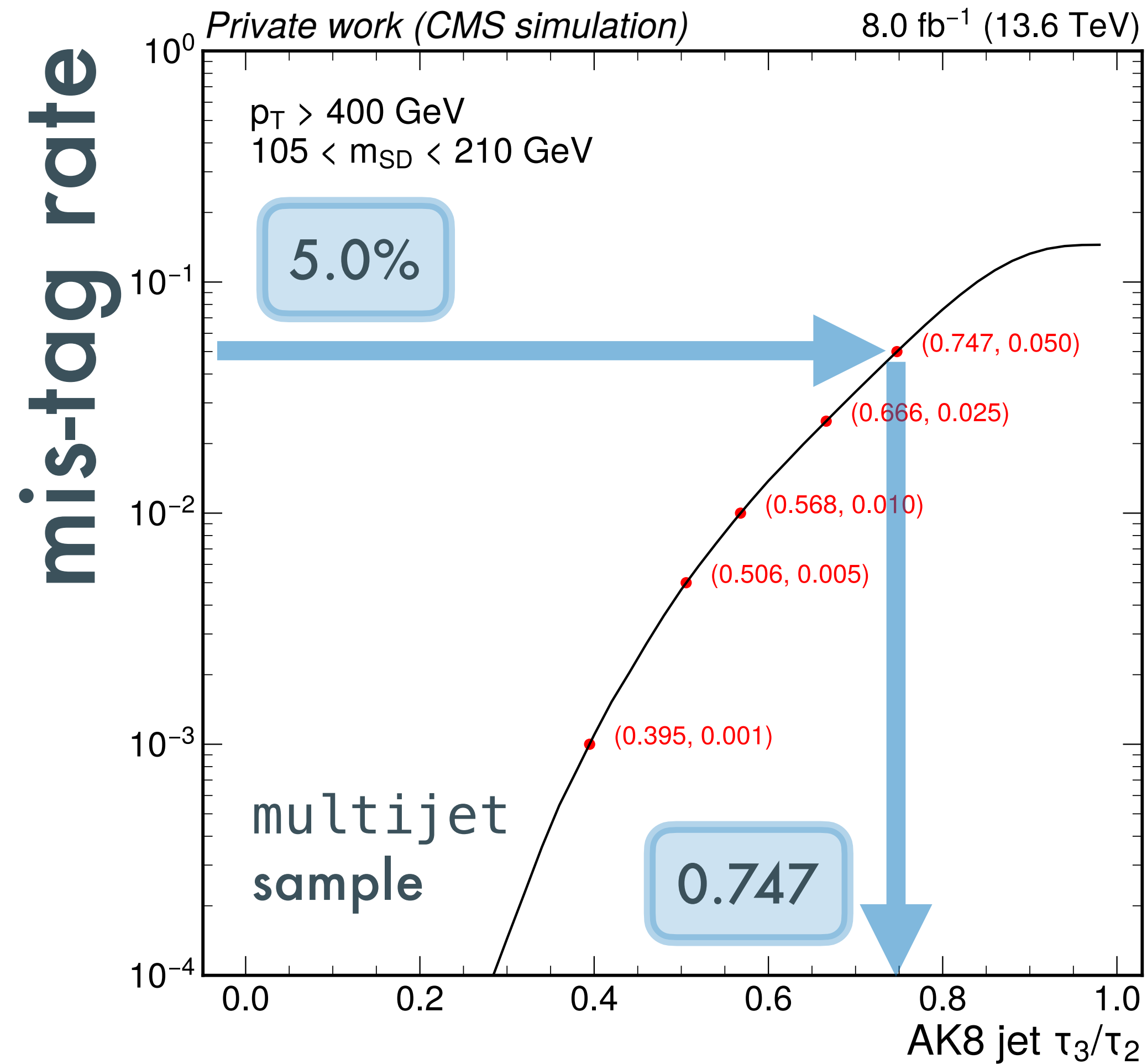
Working Points in One Slide



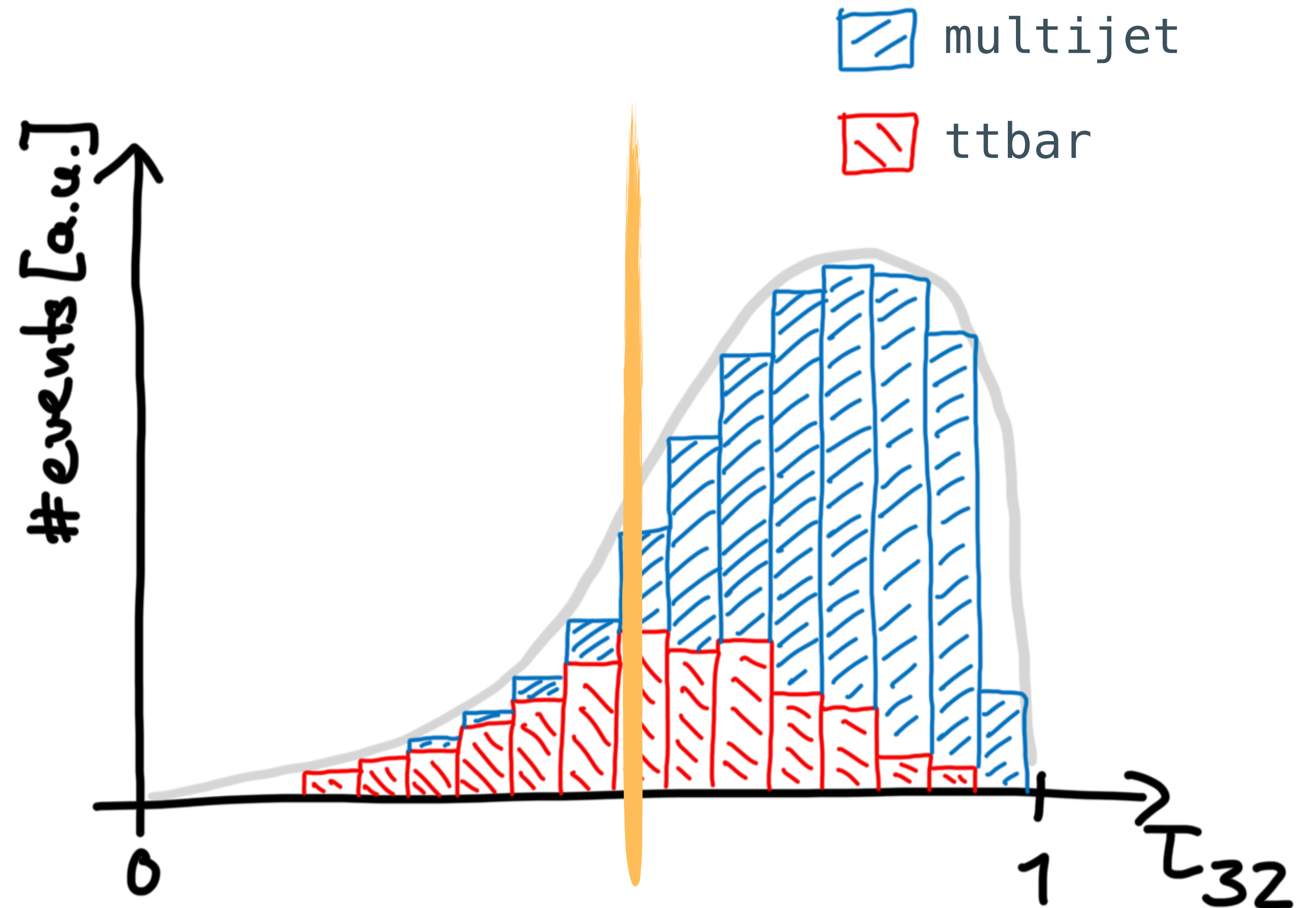
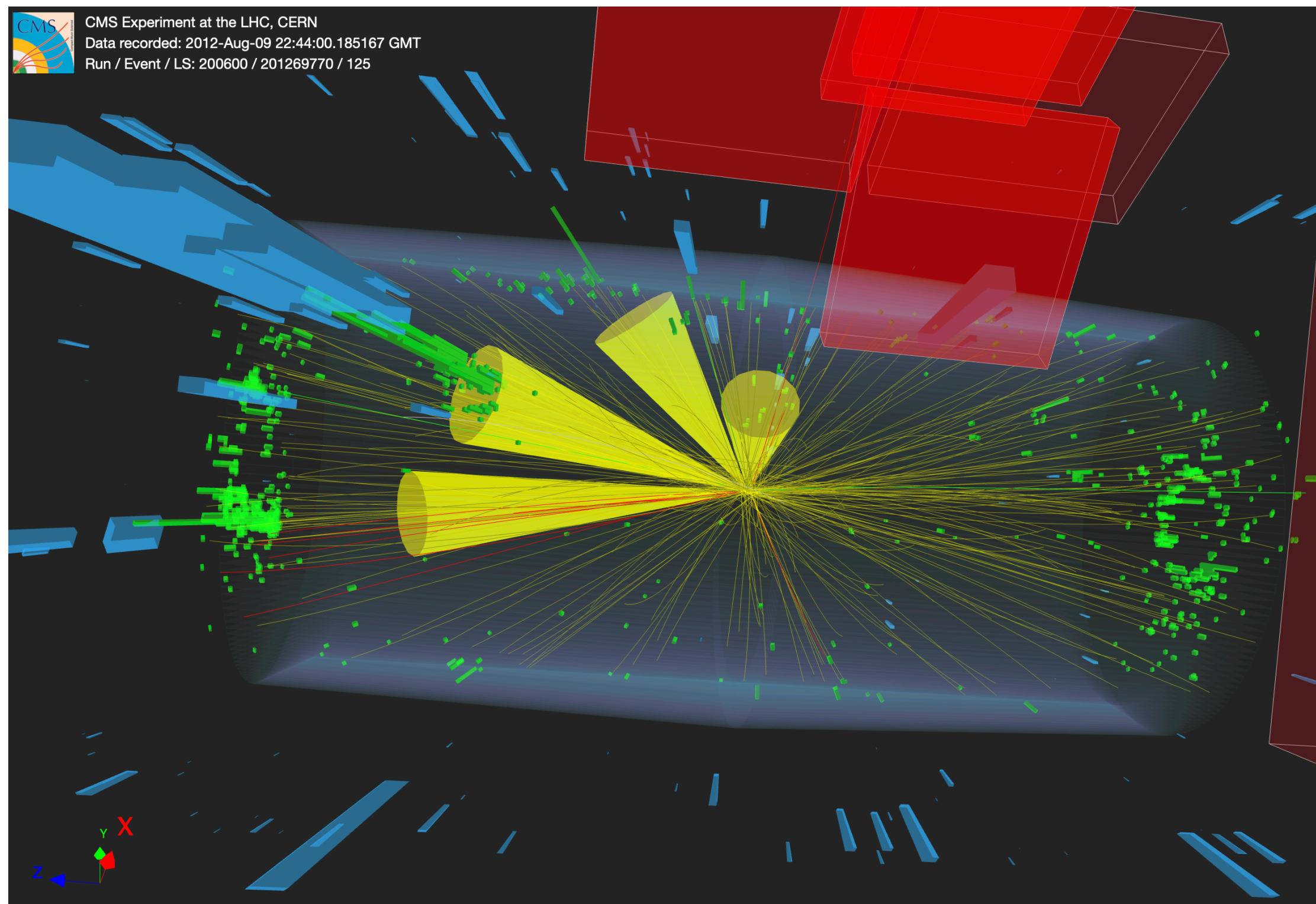
Working Points in One Slide



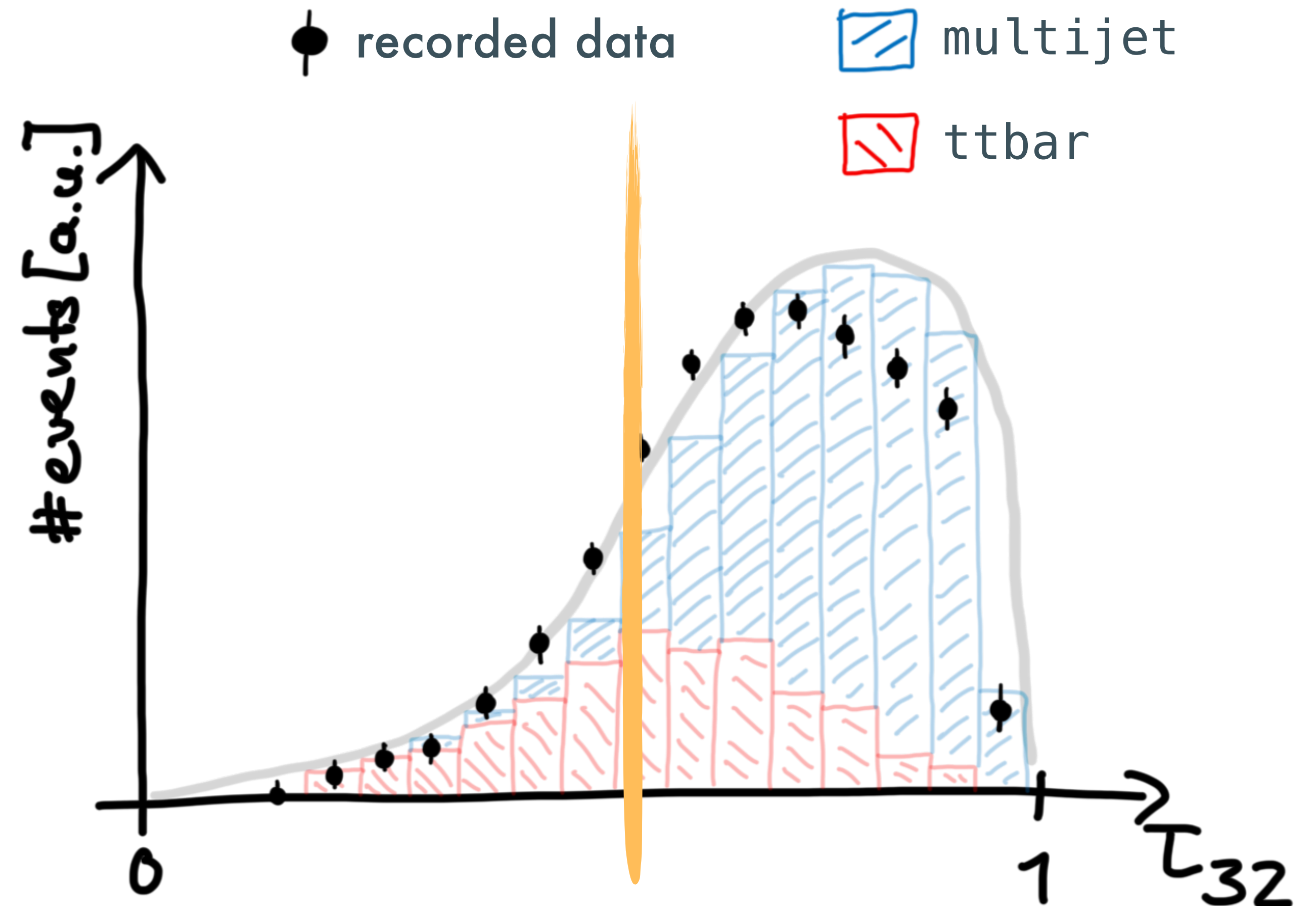
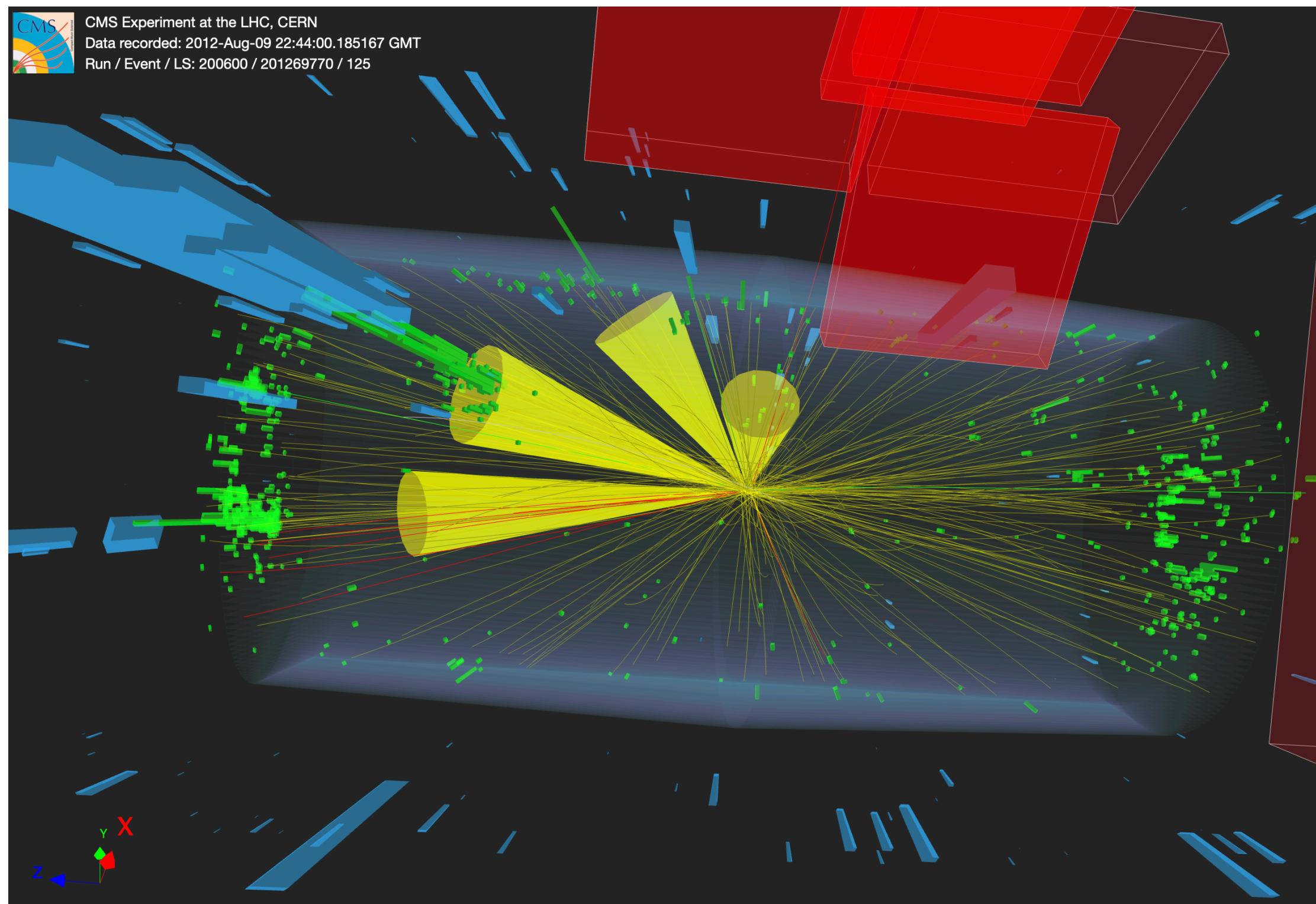
Working Points in One Slide



Jet Substructure in Recorded Data



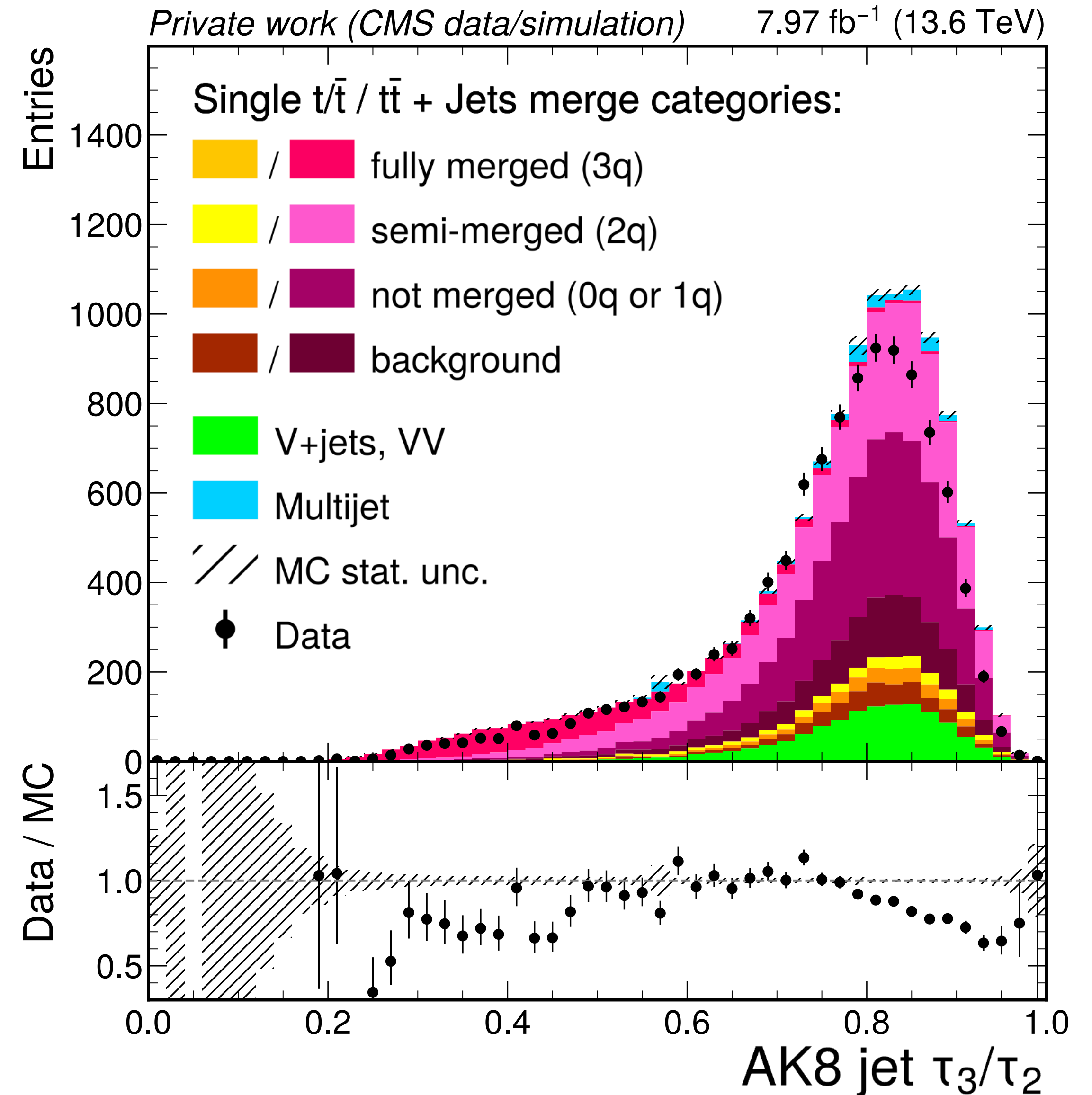
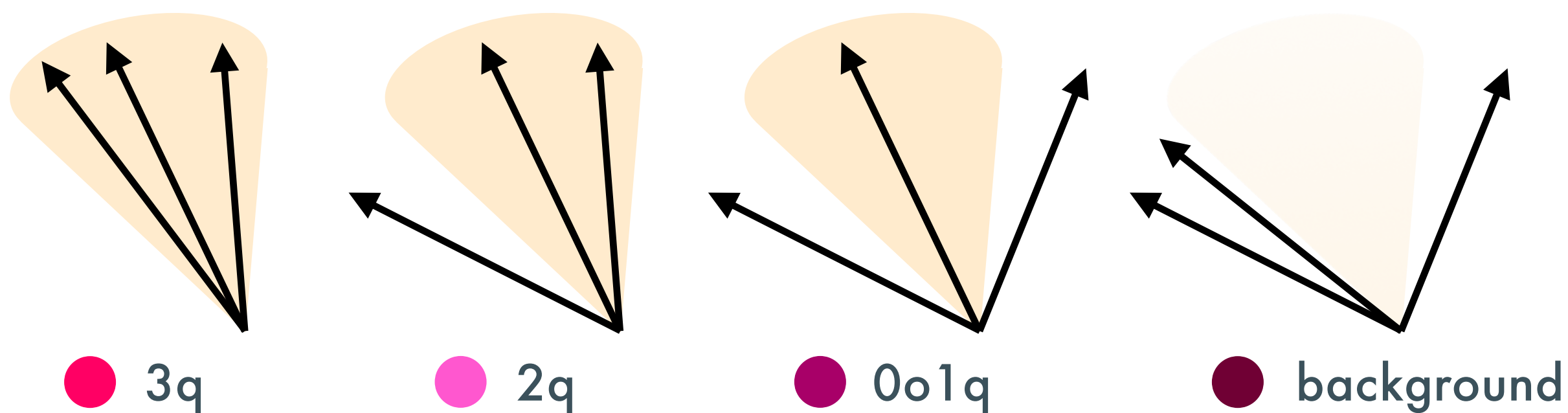
Jet Substructure in Recorded Data



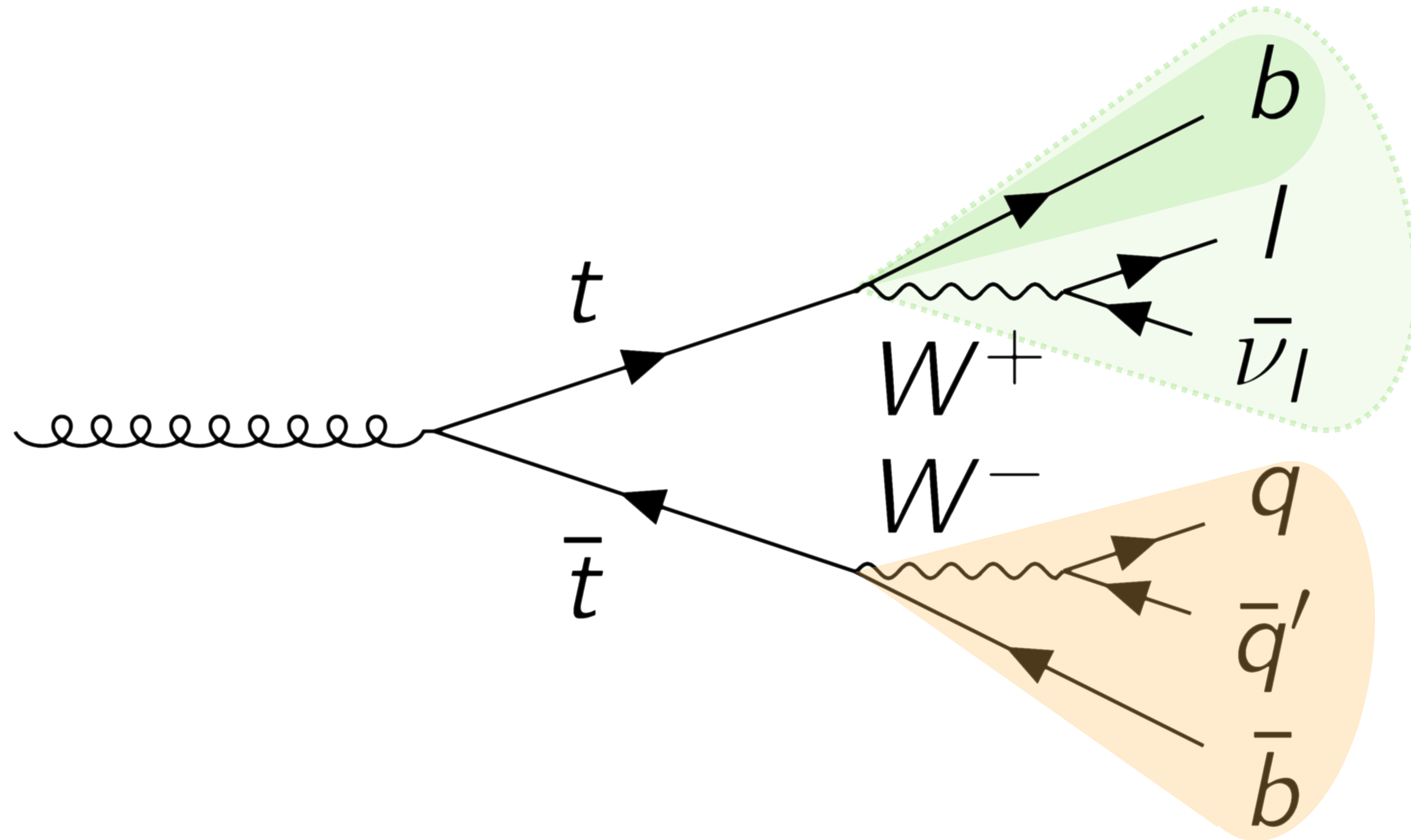
Scale Factors

— to correct MC simulation —

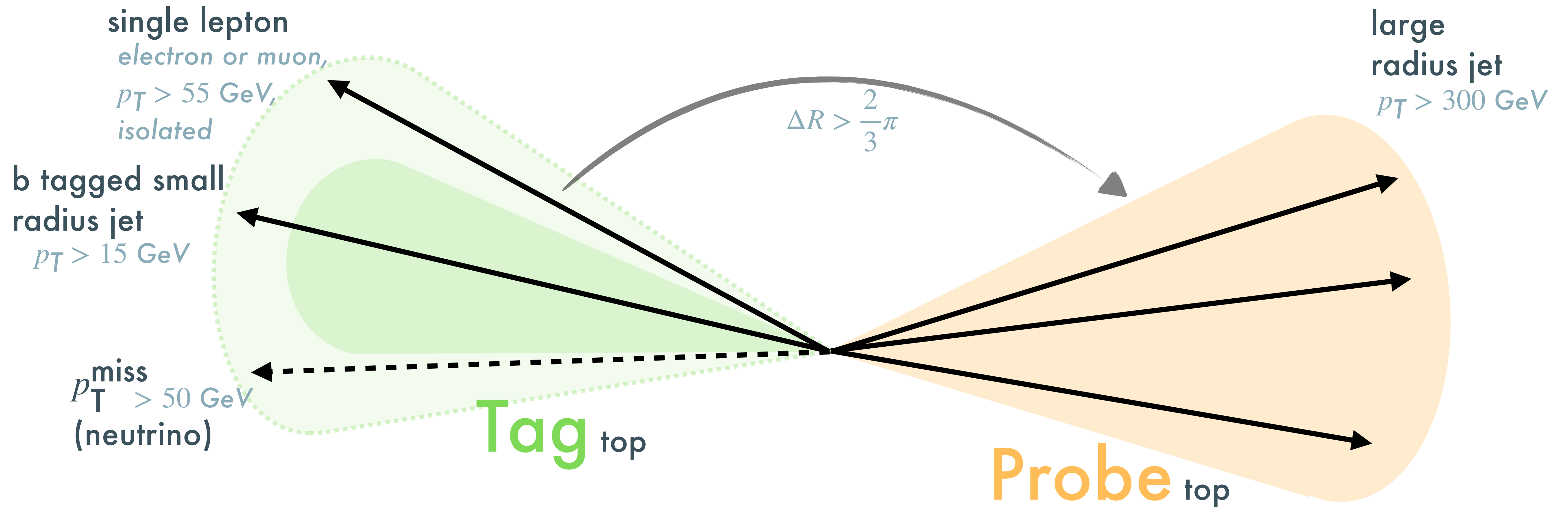
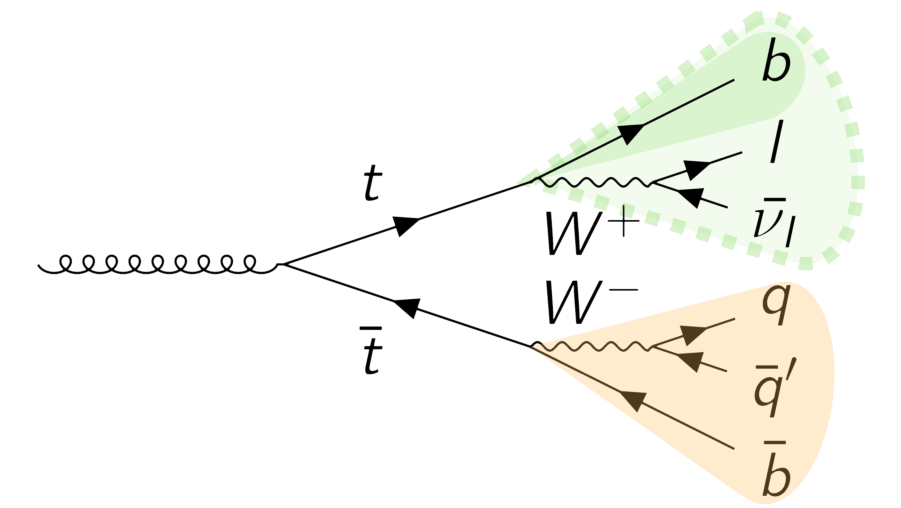
- derive correction factors for each
 - ▶ data taking era
 - ▶ working point
 - ▶ p_T bin of large-radius jet (AK8 jet)
 - ▶ merge scenario of top quark decay products



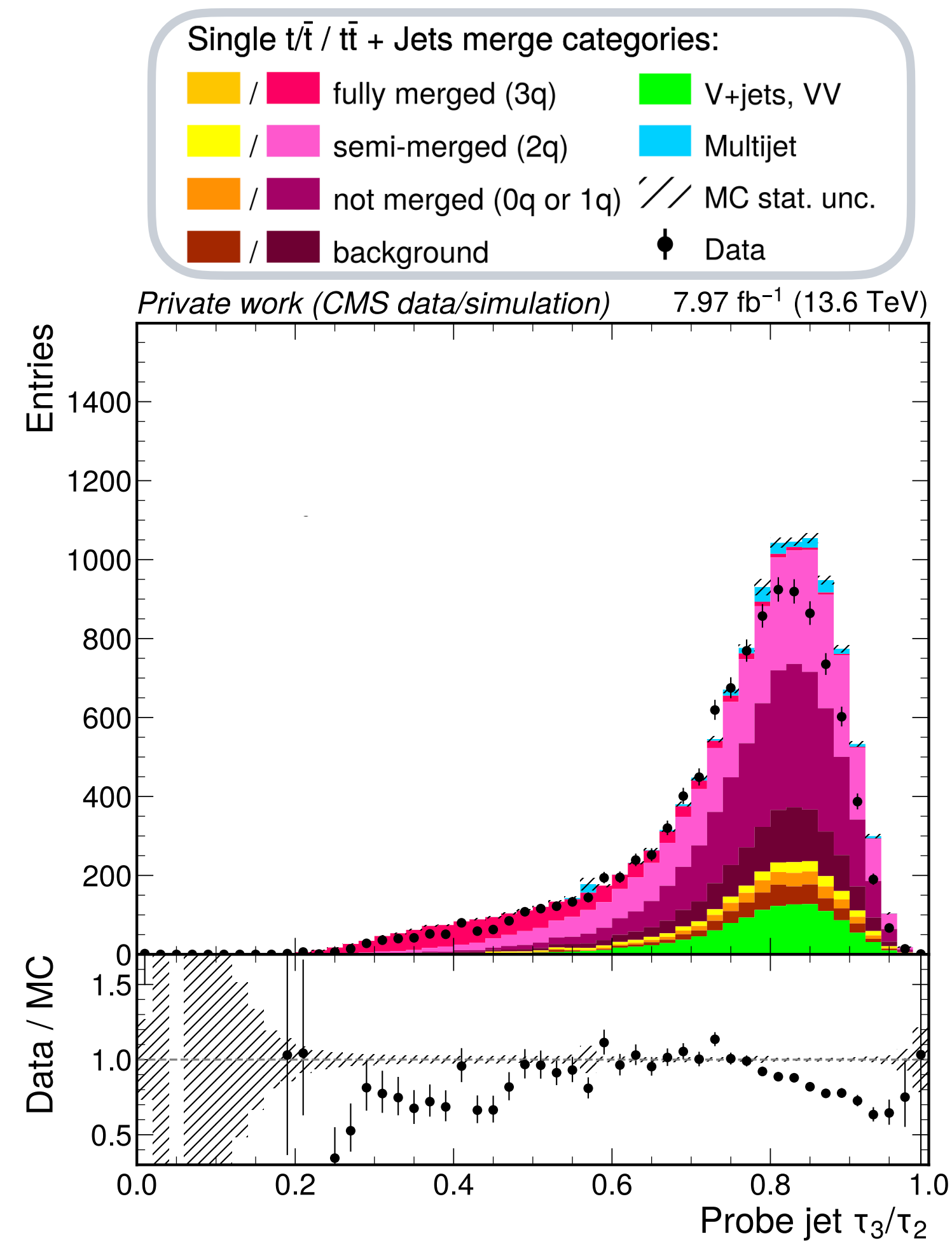
Tag and Probe Method With $t\bar{t}$ Pairs



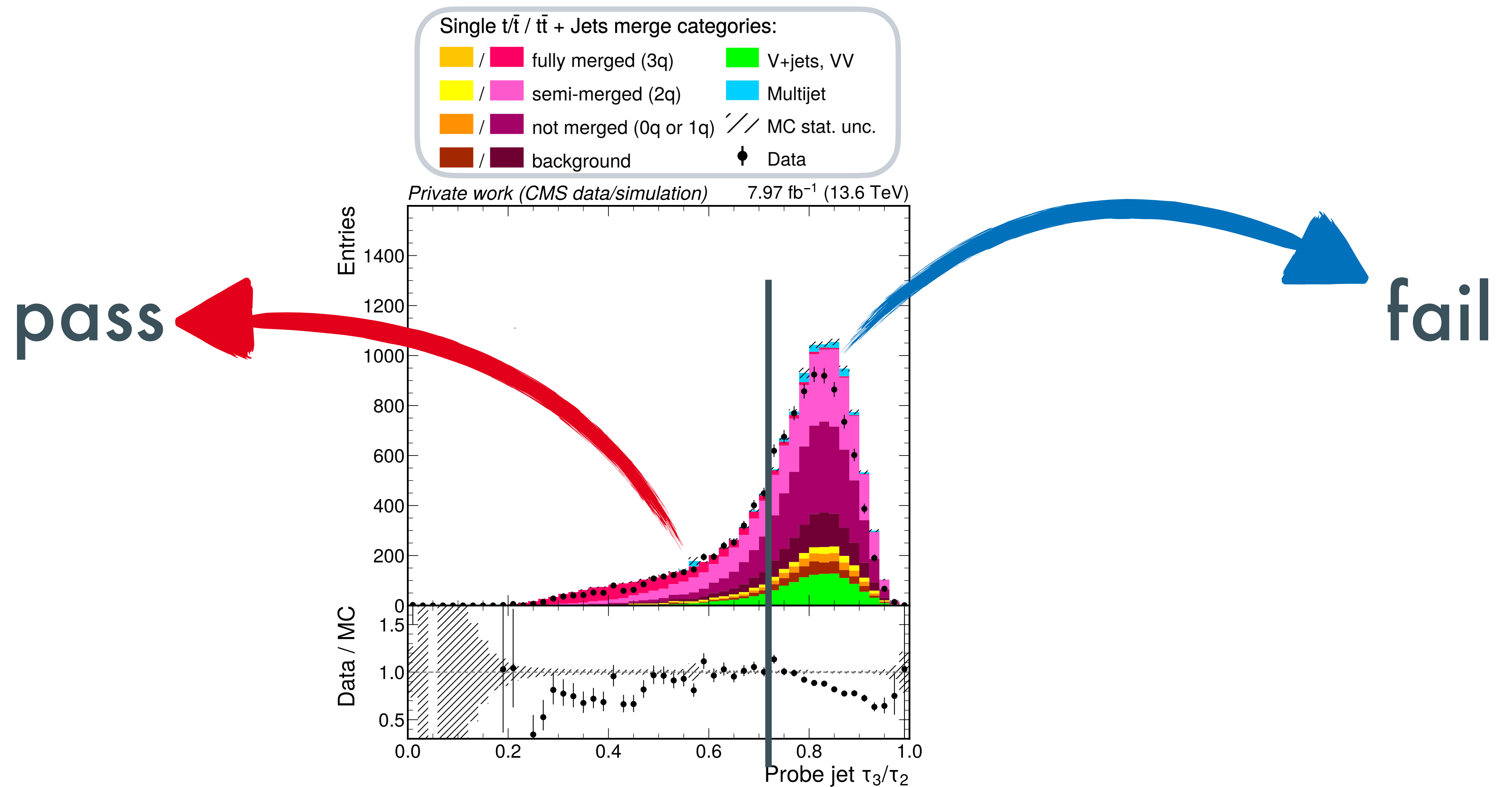
Tag and Probe Method With $t\bar{t}$ Pairs



Probe Jet: τ_{32}



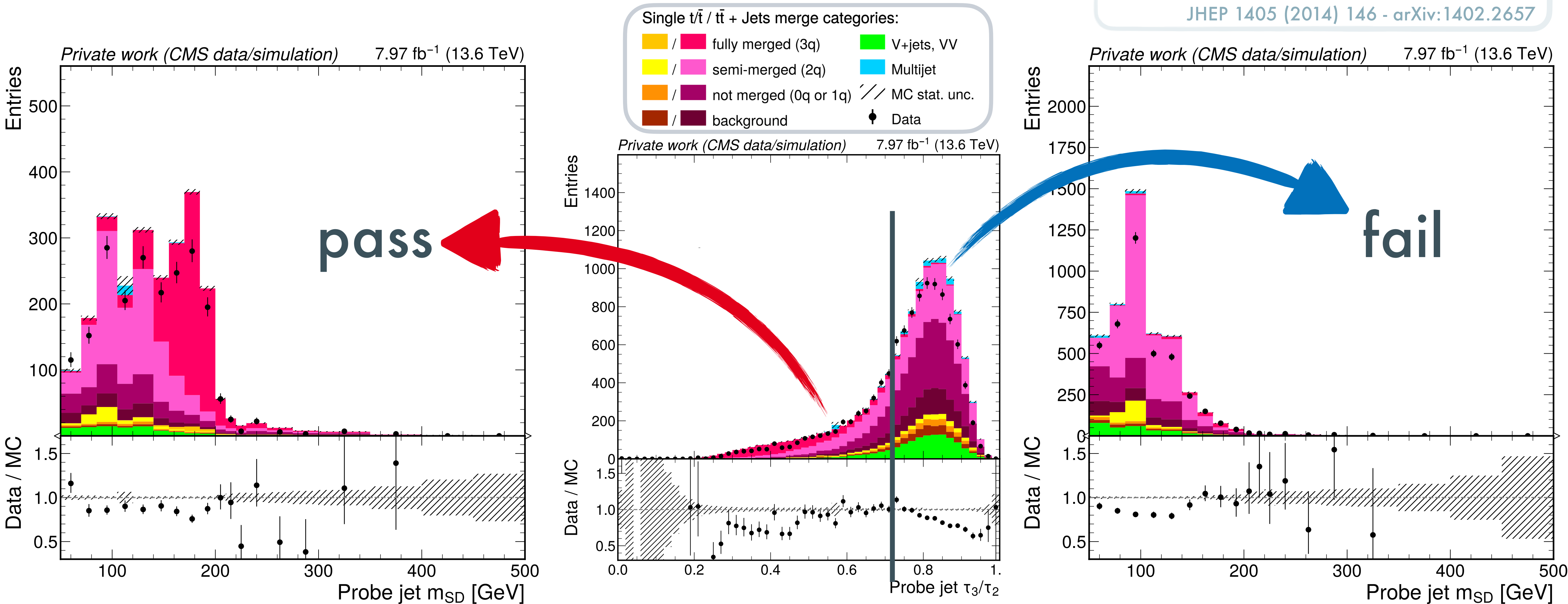
Probe Jet: τ_{32}



Probe Jet: τ_{32} and m_{SD}

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

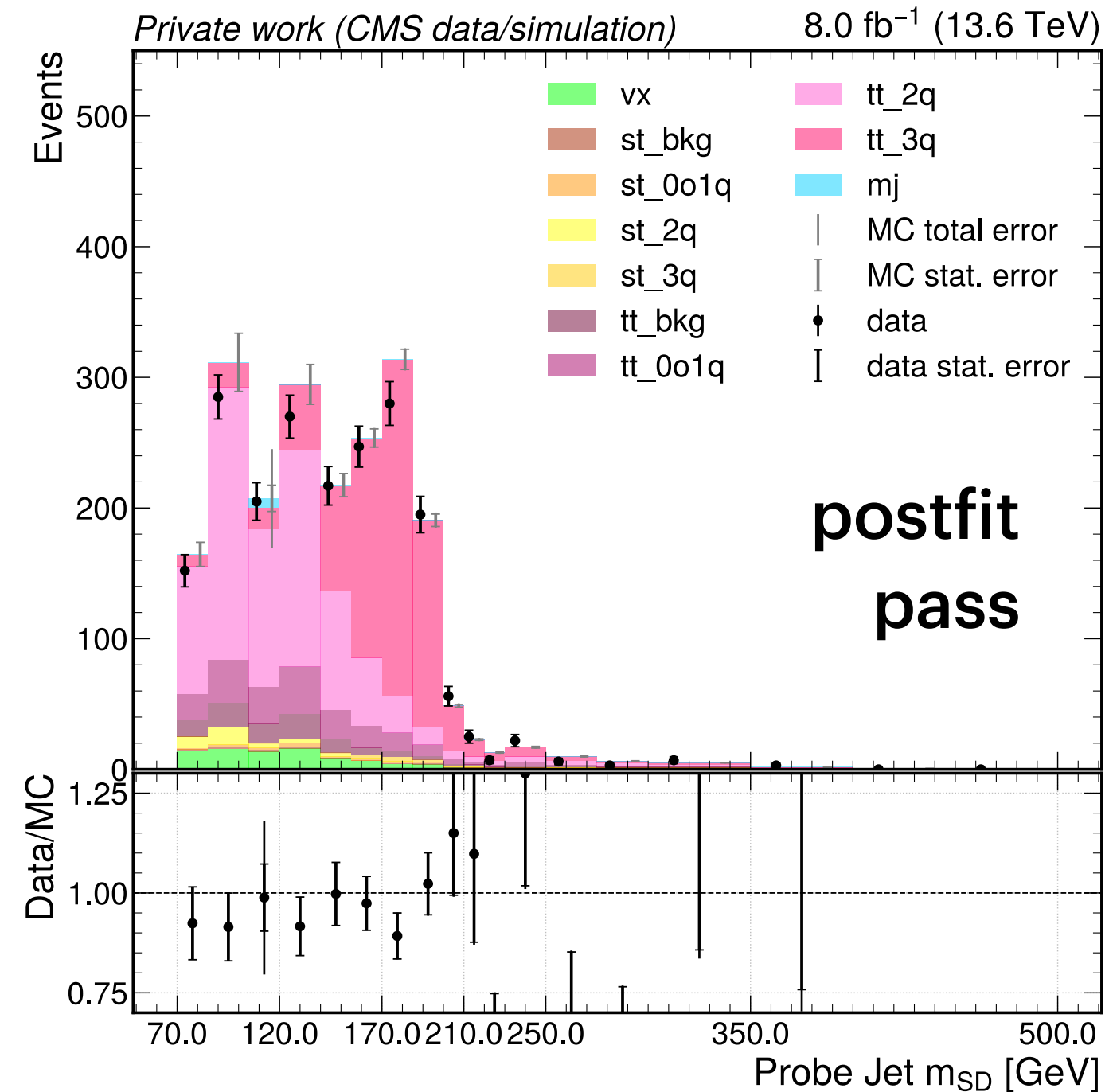
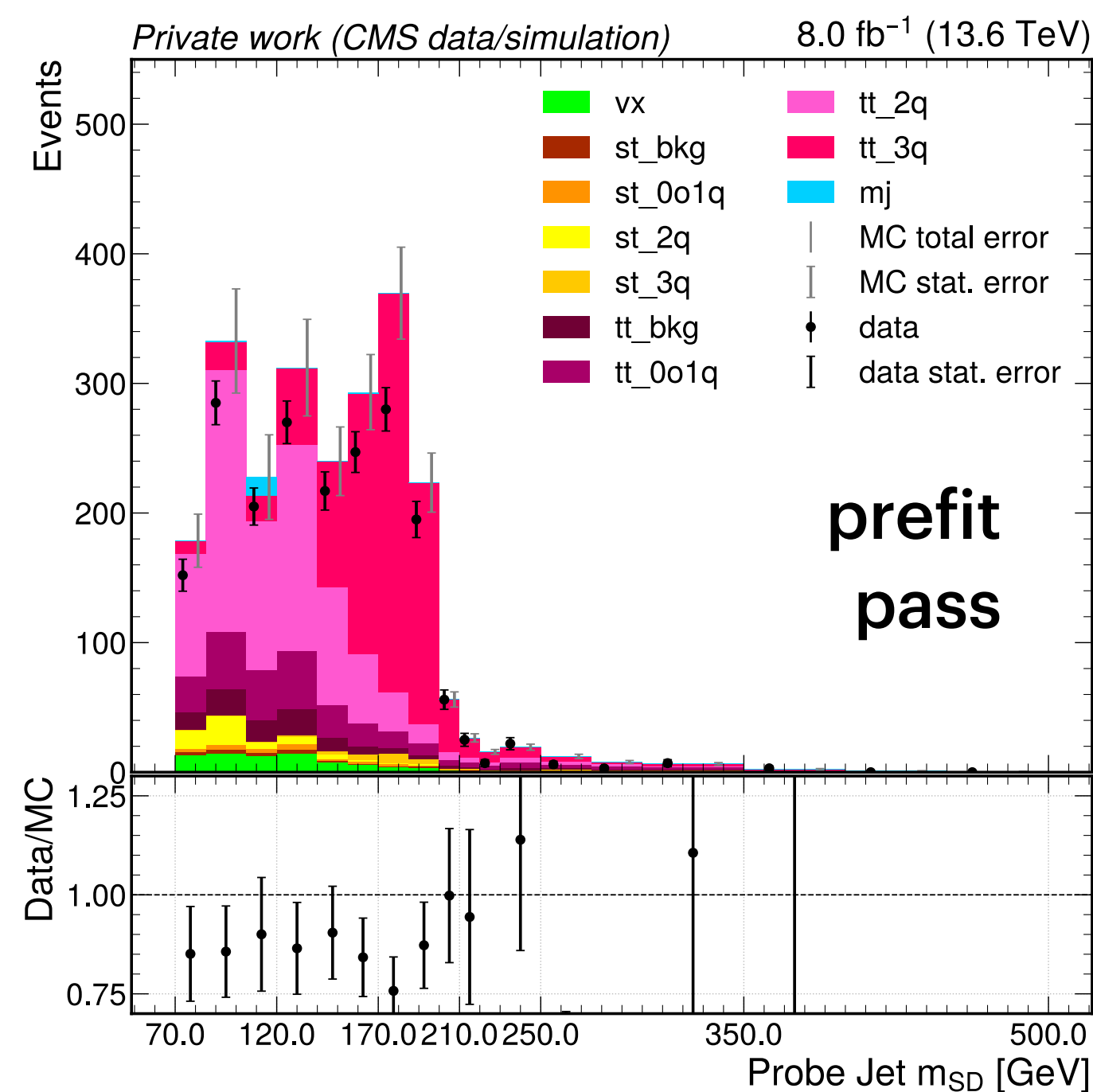
JHEP 1405 (2014) 146 - arXiv:1402.2657



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, jet energy corrections largest

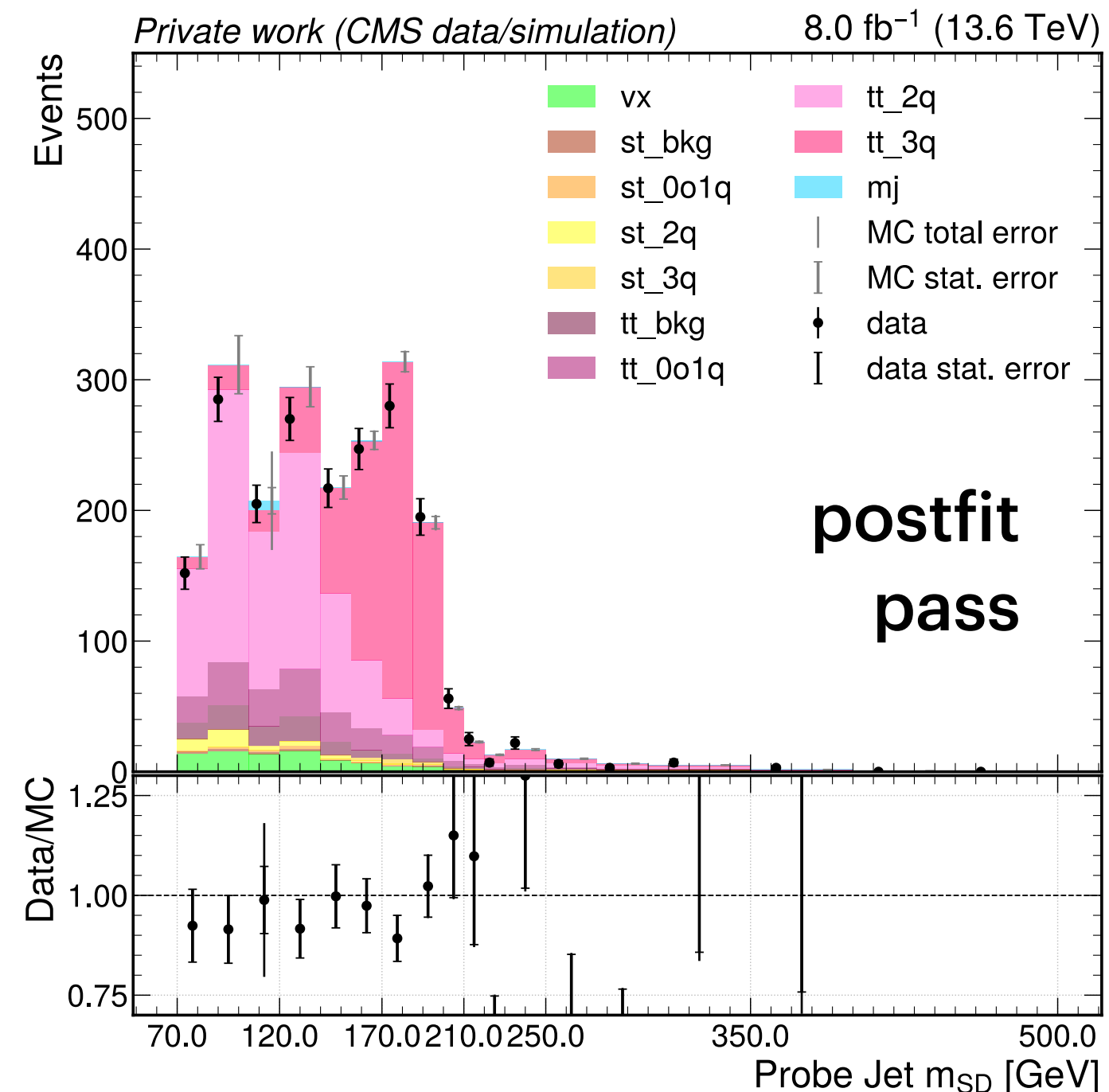
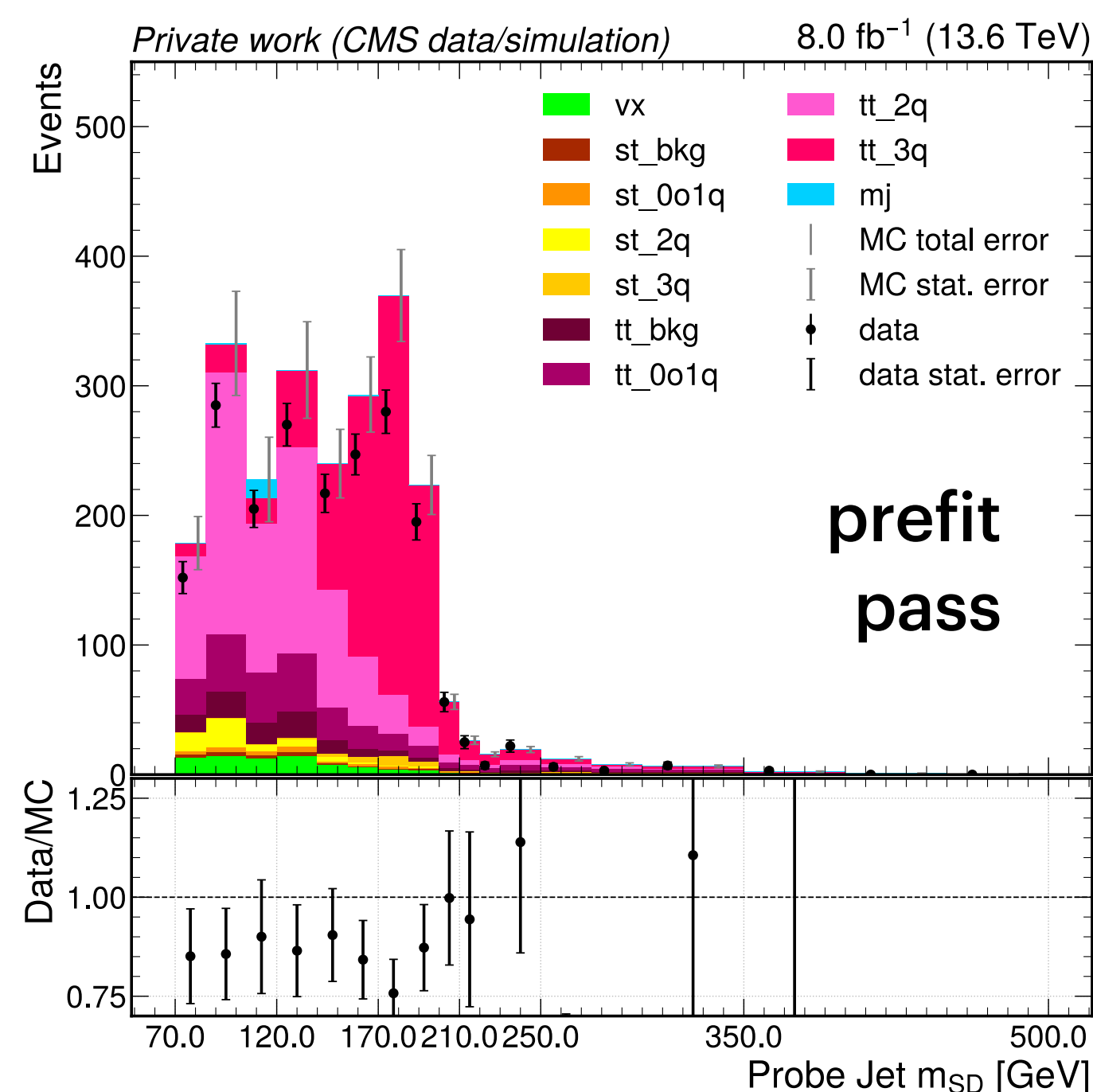


SF Derivation

— 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, jet energy corrections largest

**Quick derivation of SF
essential to be used in
analyses.**

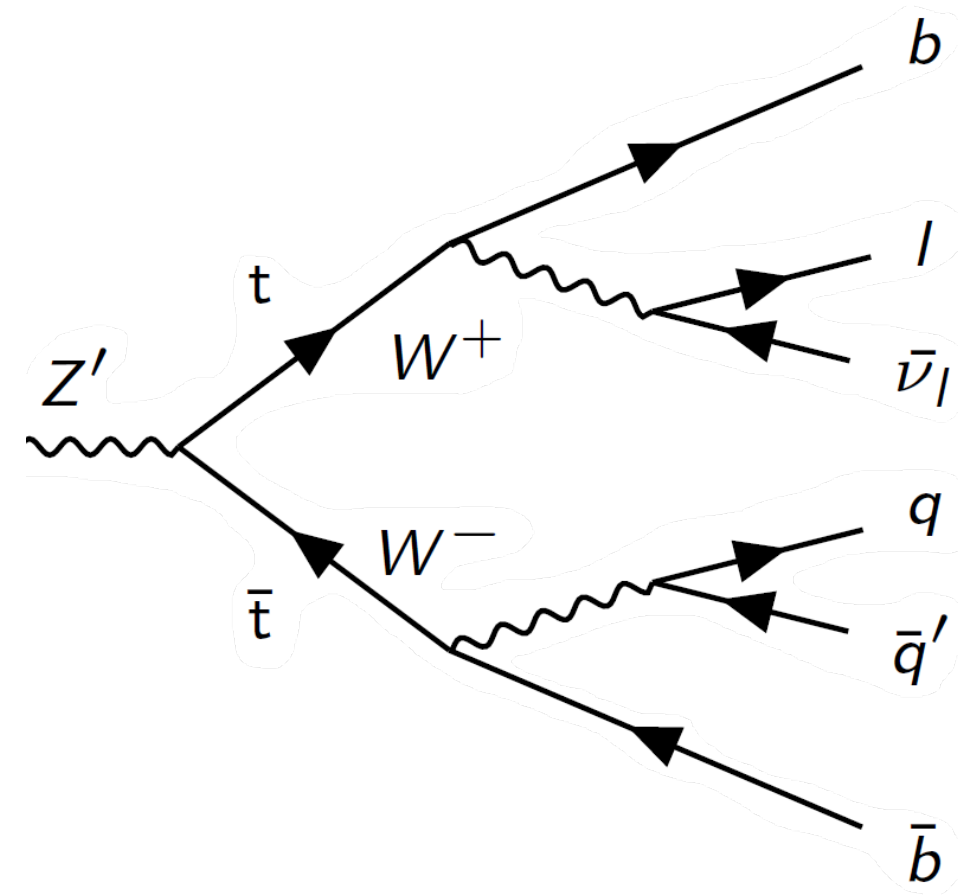


**column
flow**
github.com/columnflow/columnflow

Search for New Physics Effects ...

calibrate events/objects

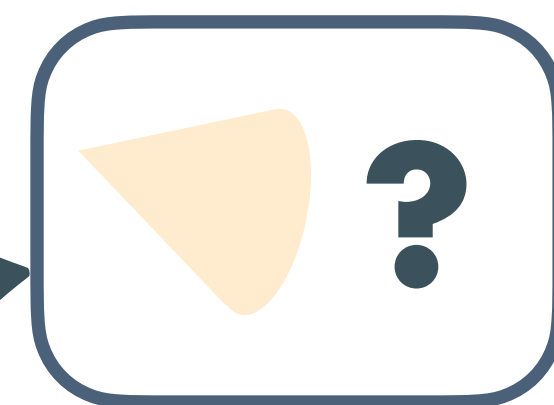
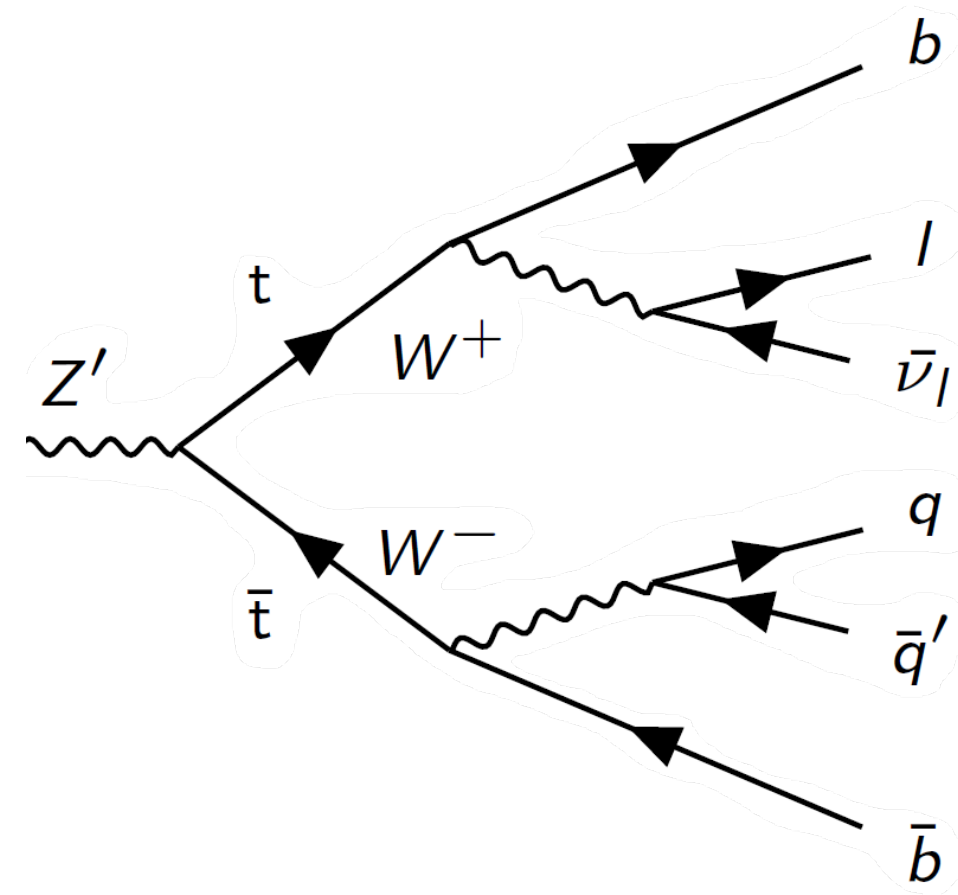
select events



Search for New Physics Effects ...

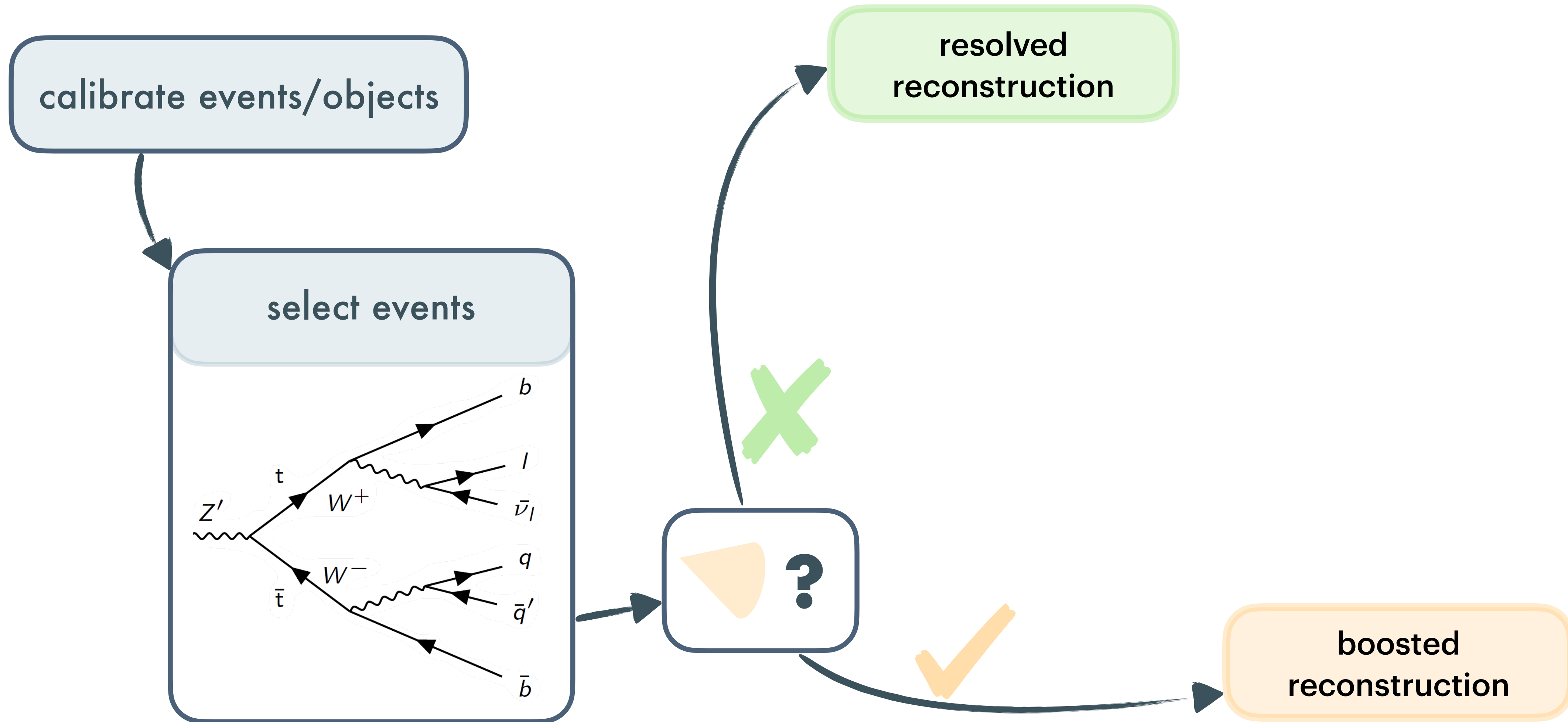
calibrate events/objects

select events

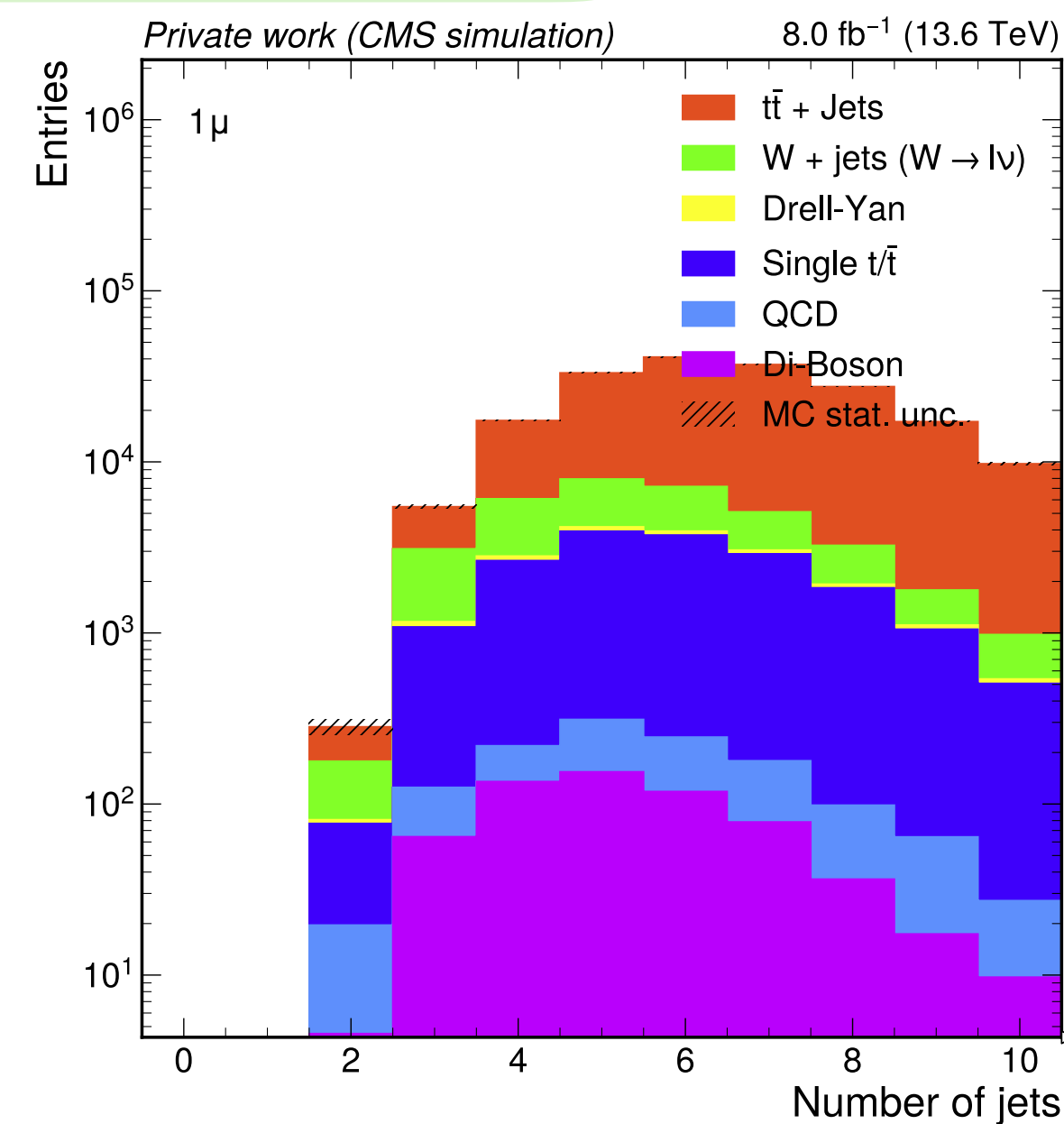
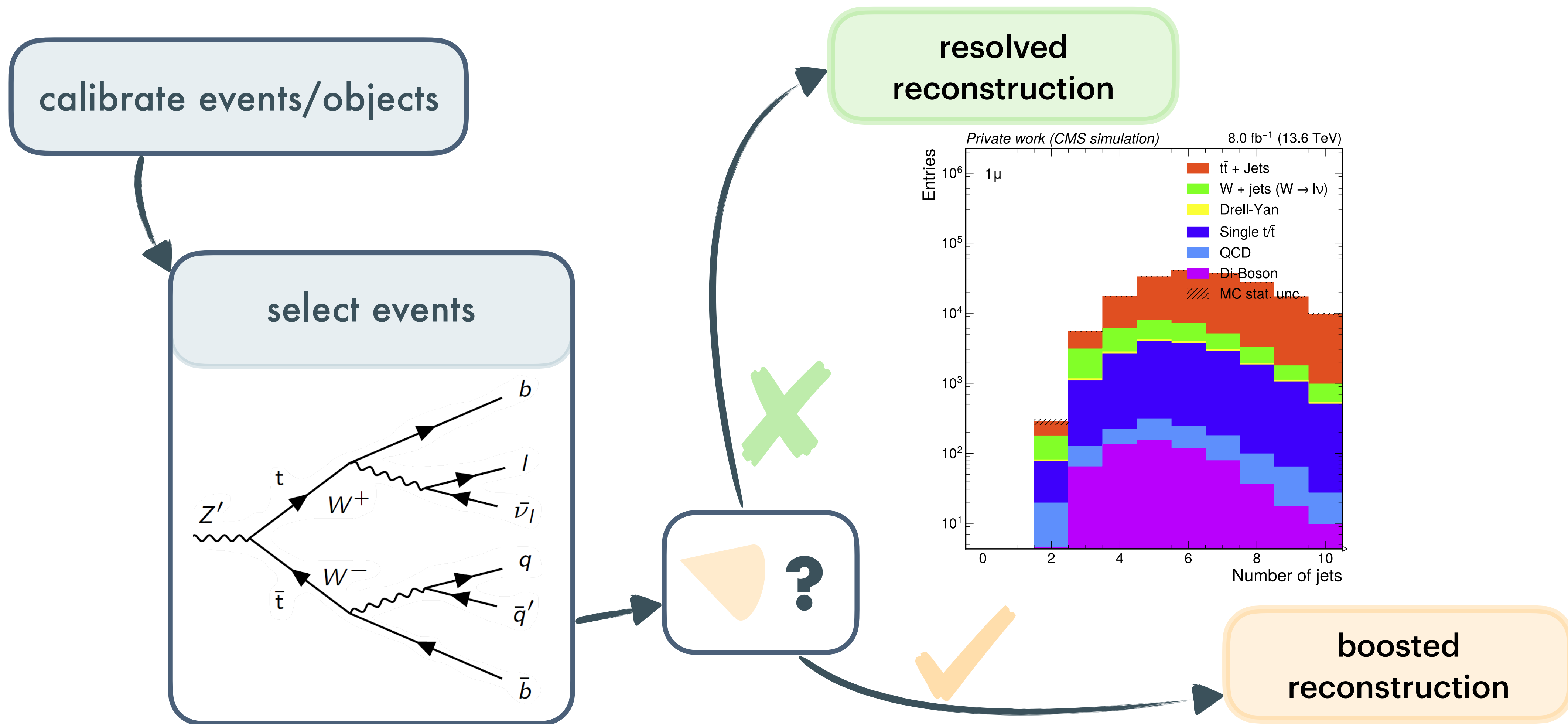


boosted
reconstruction

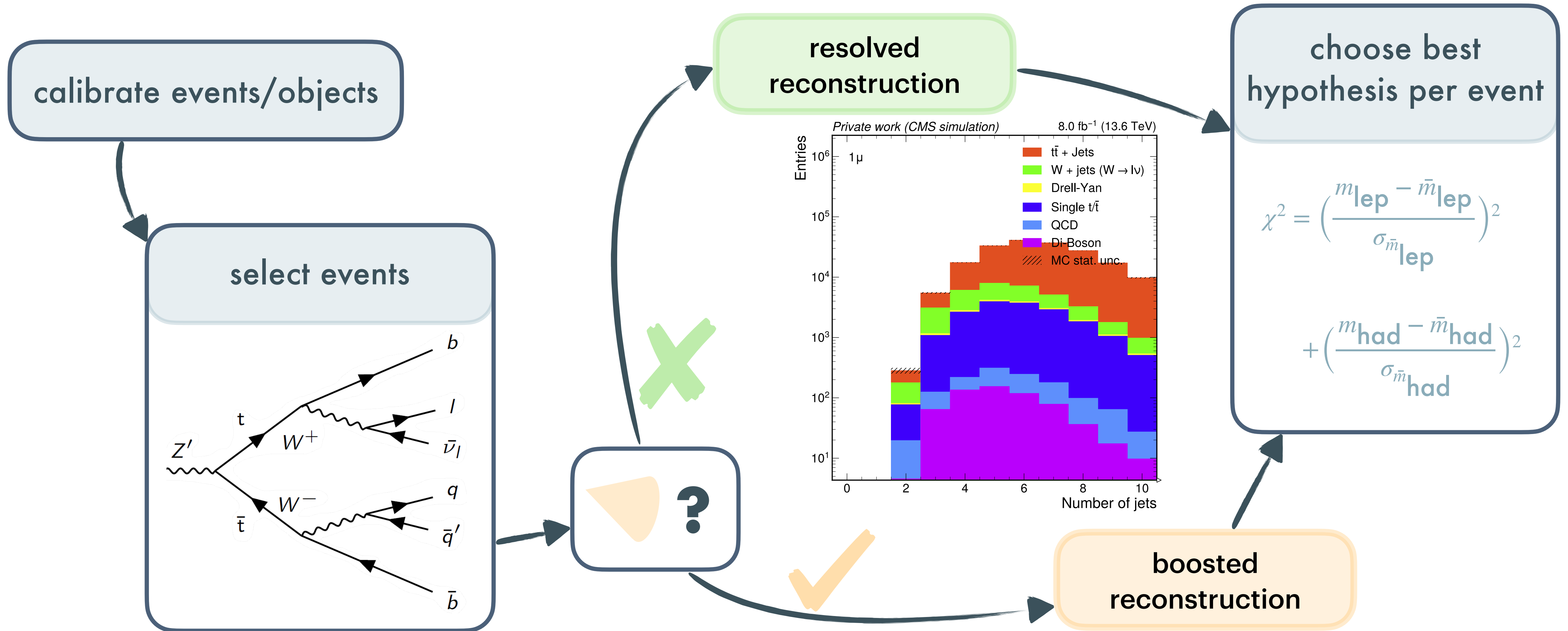
Search for New Physics Effects ...



Search for New Physics Effects ...



Search for New Physics Effects ...

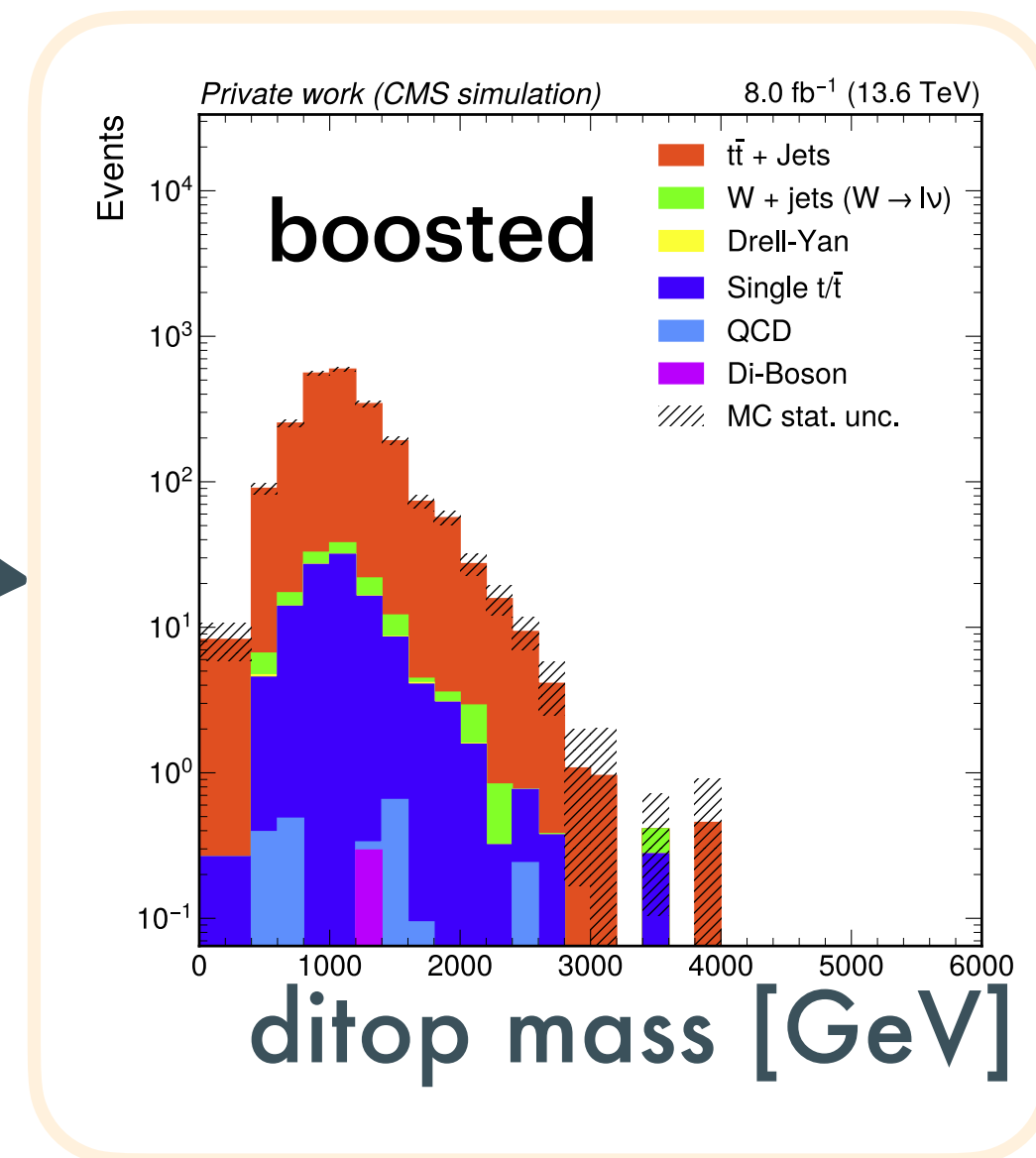
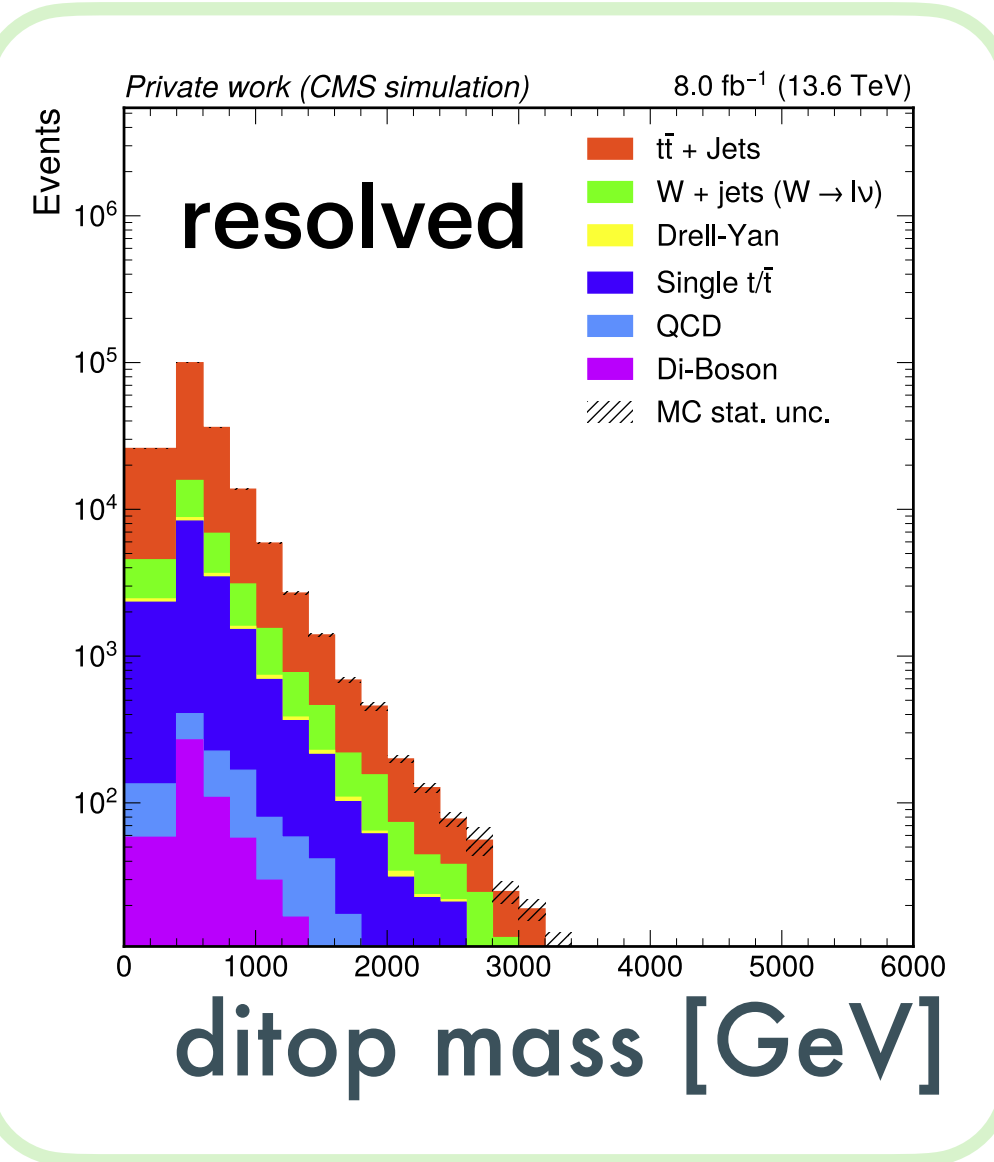


... In the Ditop Mass Spectrum

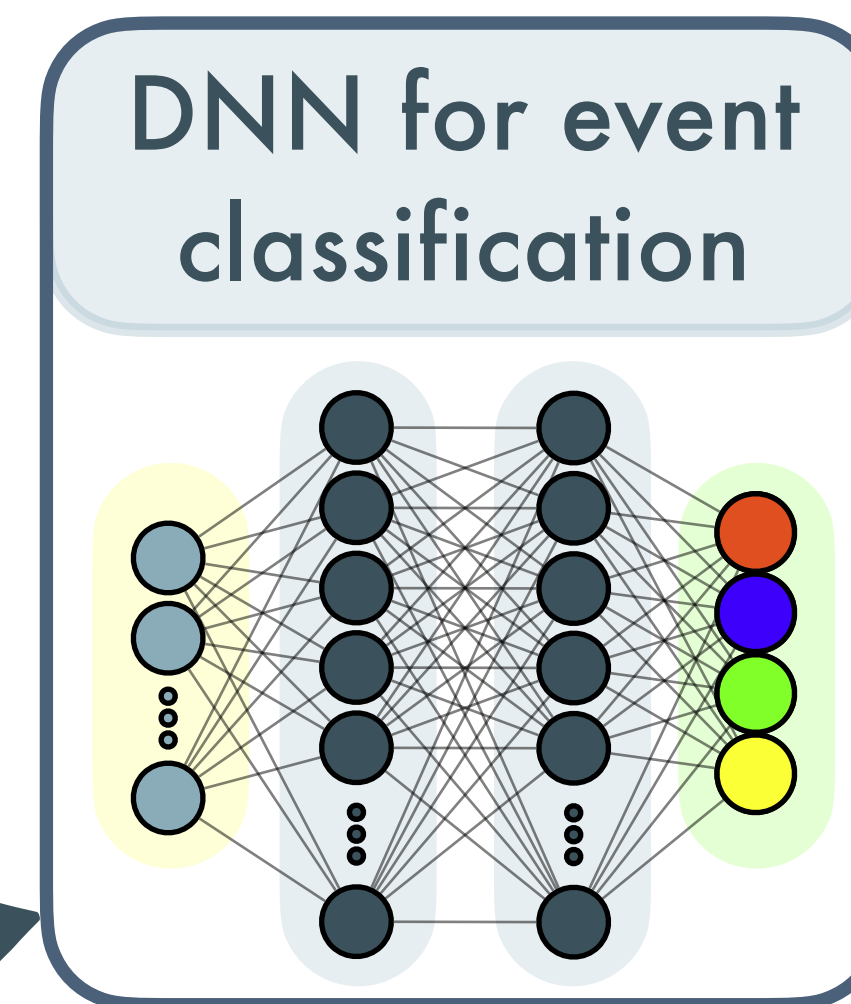
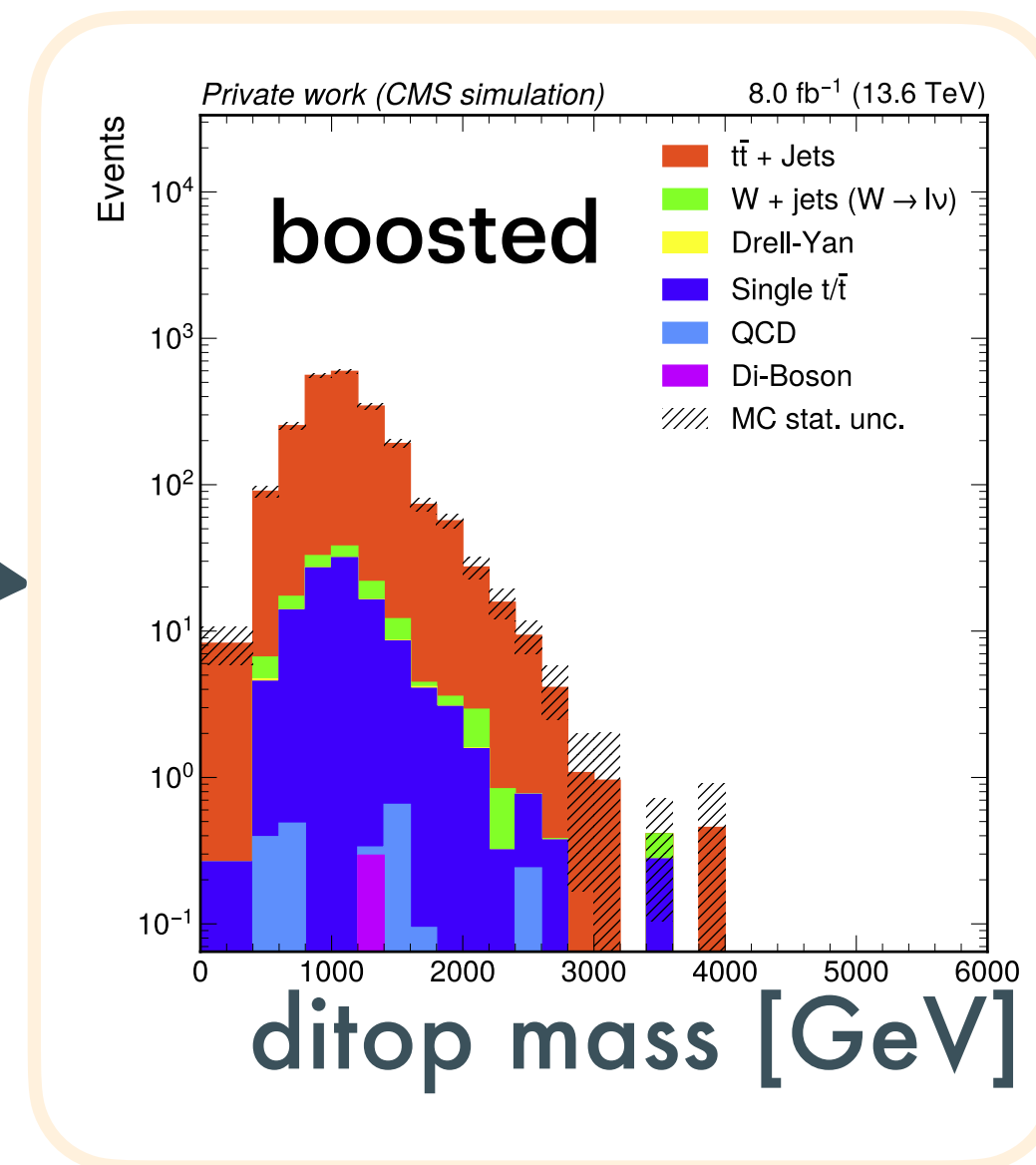
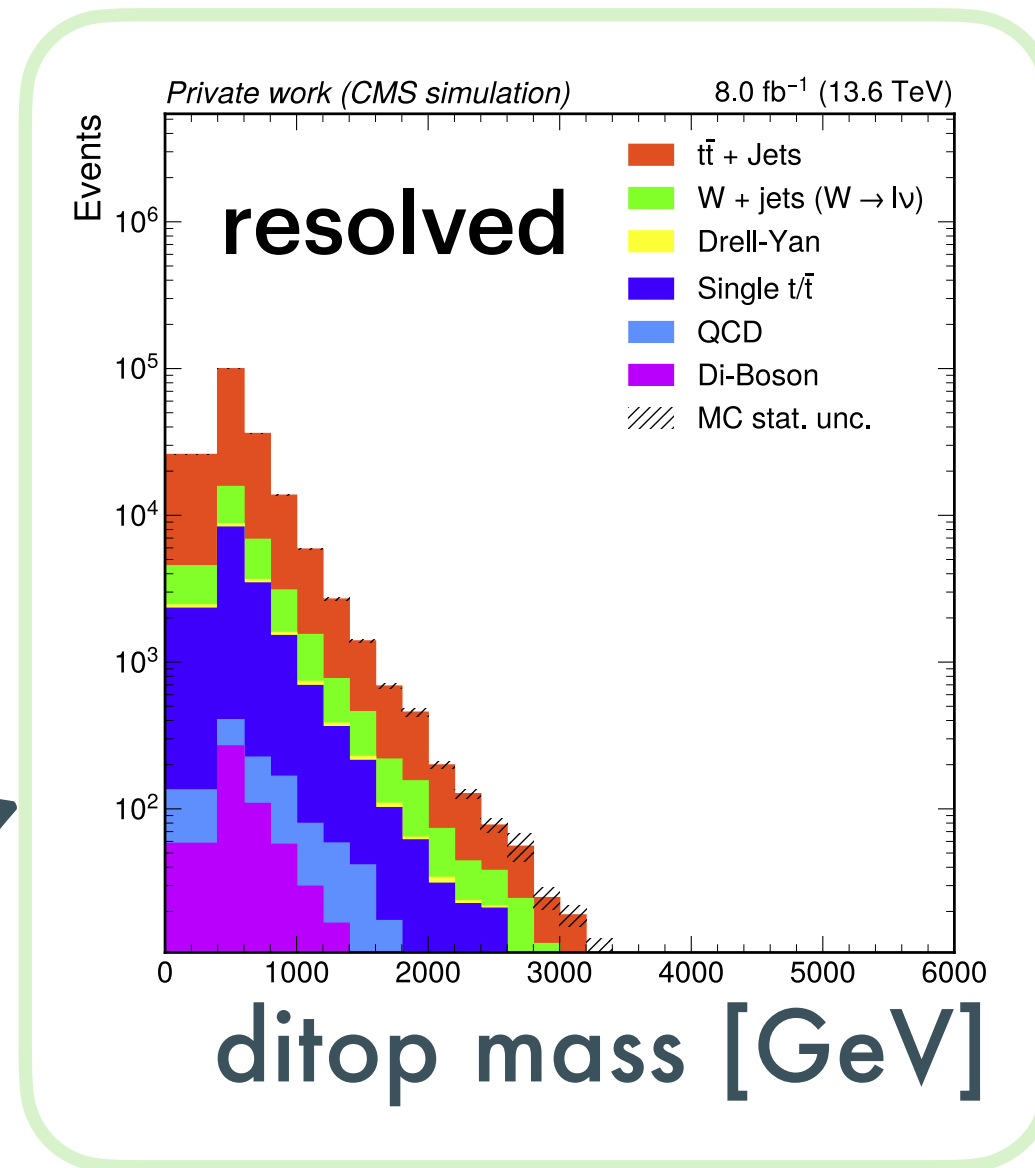
best
hypo-
thesis

... In the Ditop Mass Spectrum

best
hypo-
thesis

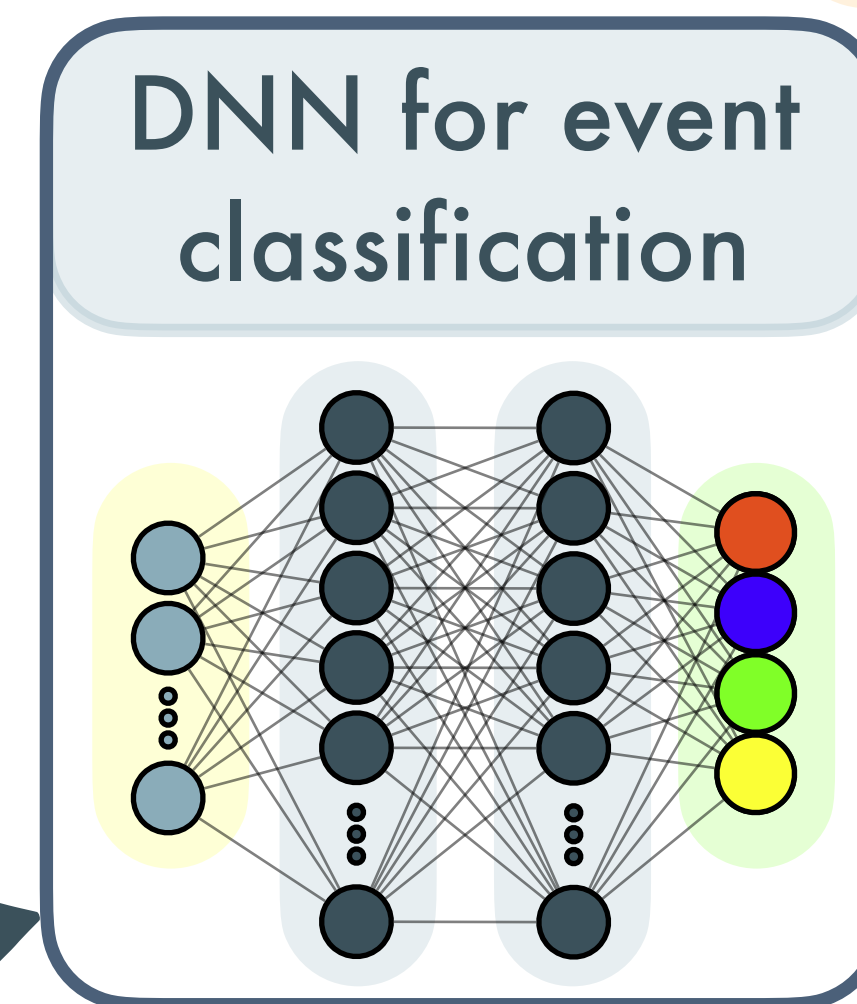
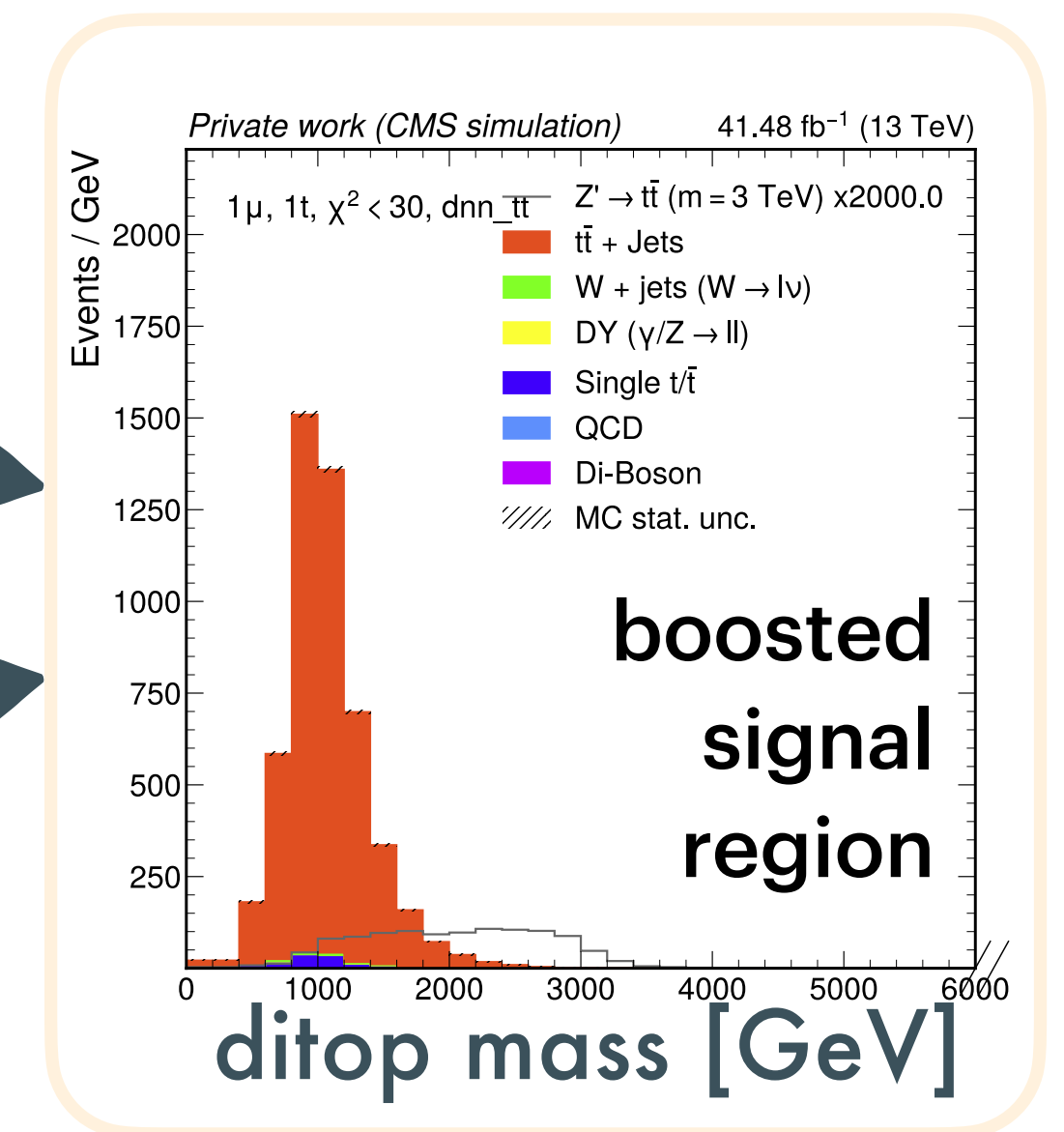
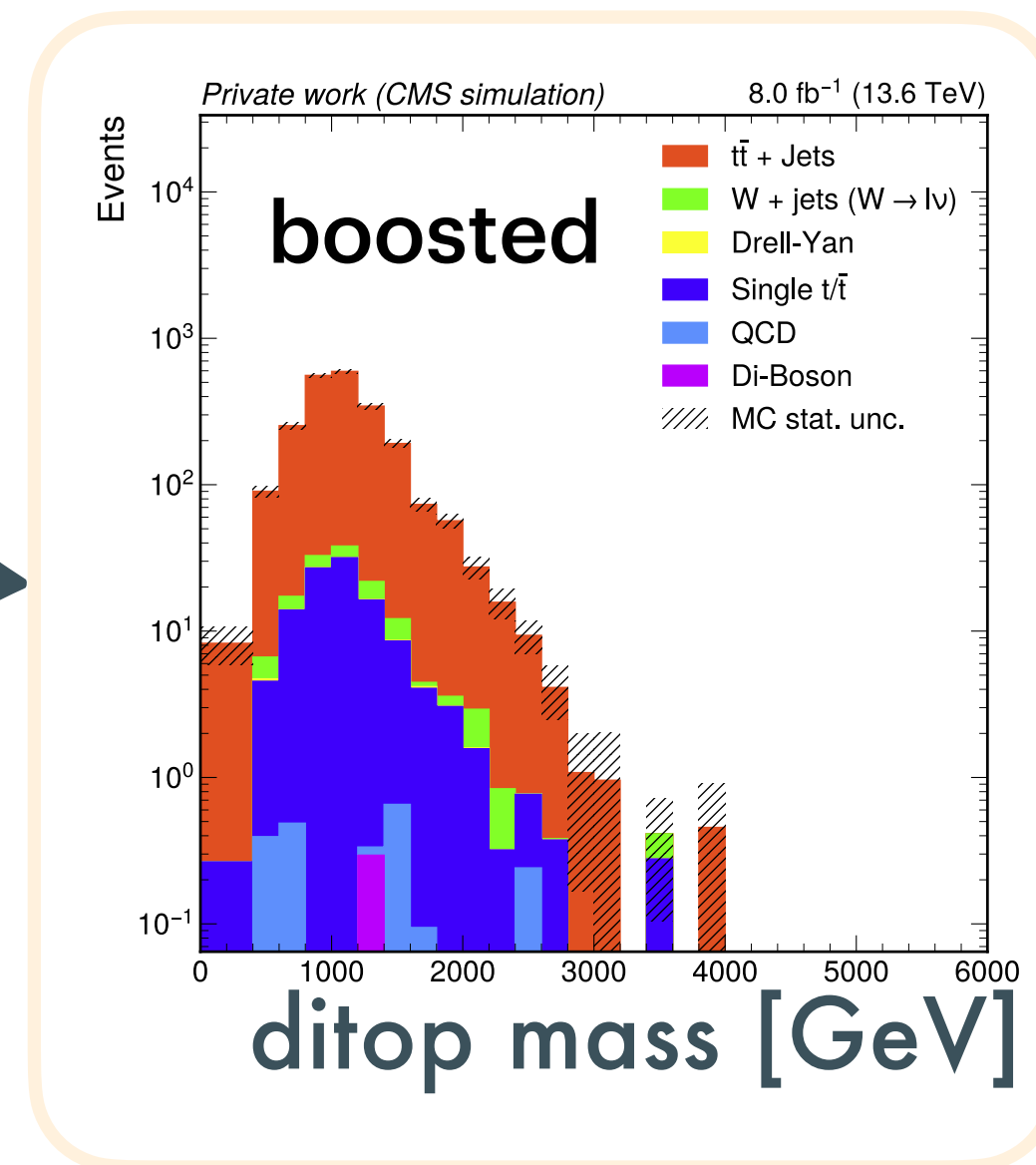
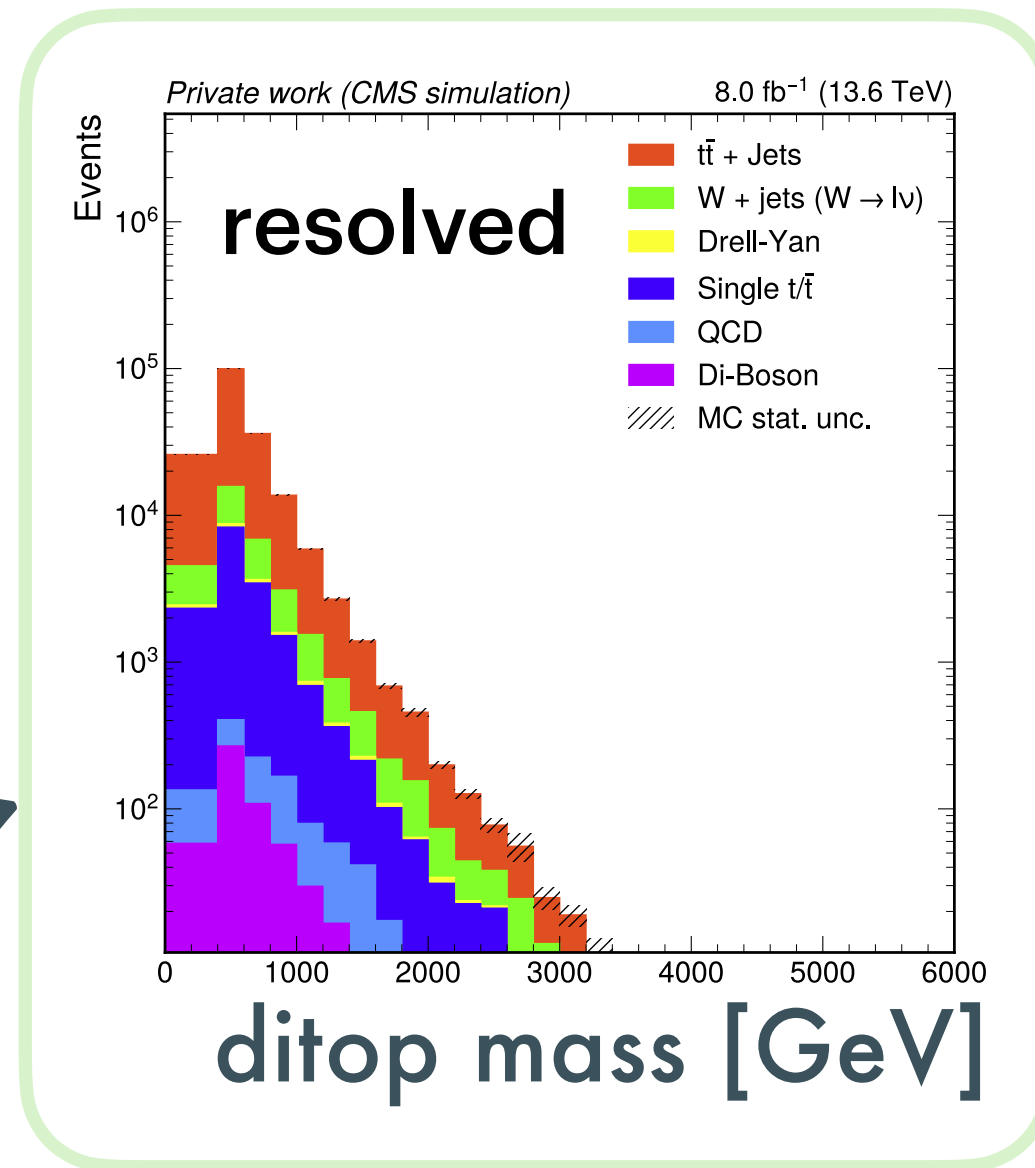


... In the Ditop Mass Spectrum



selected events

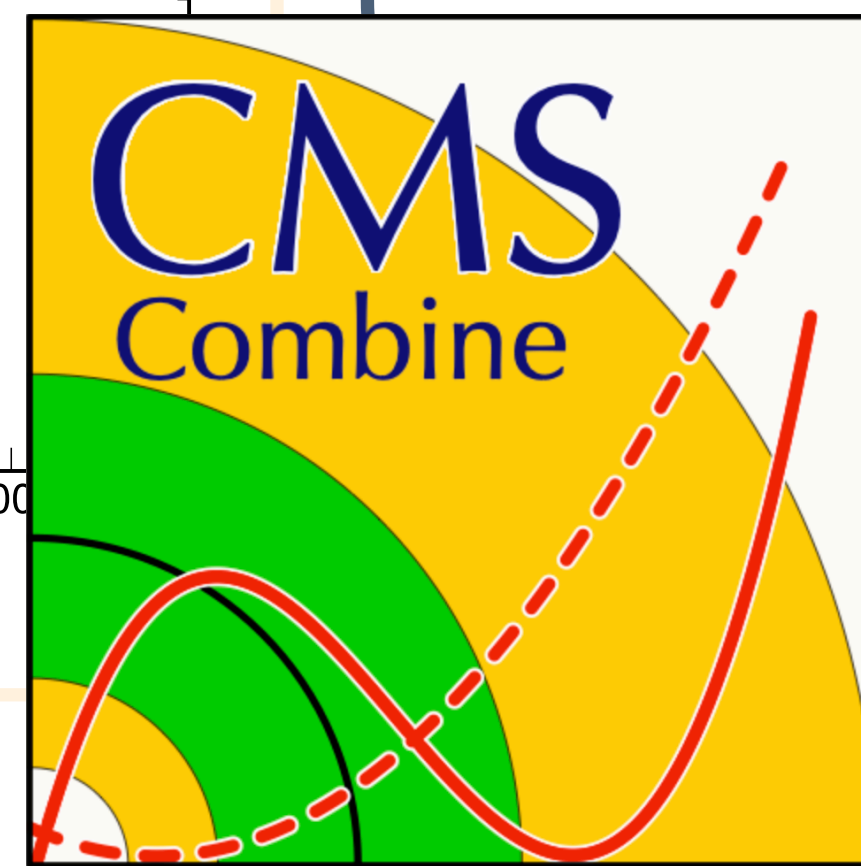
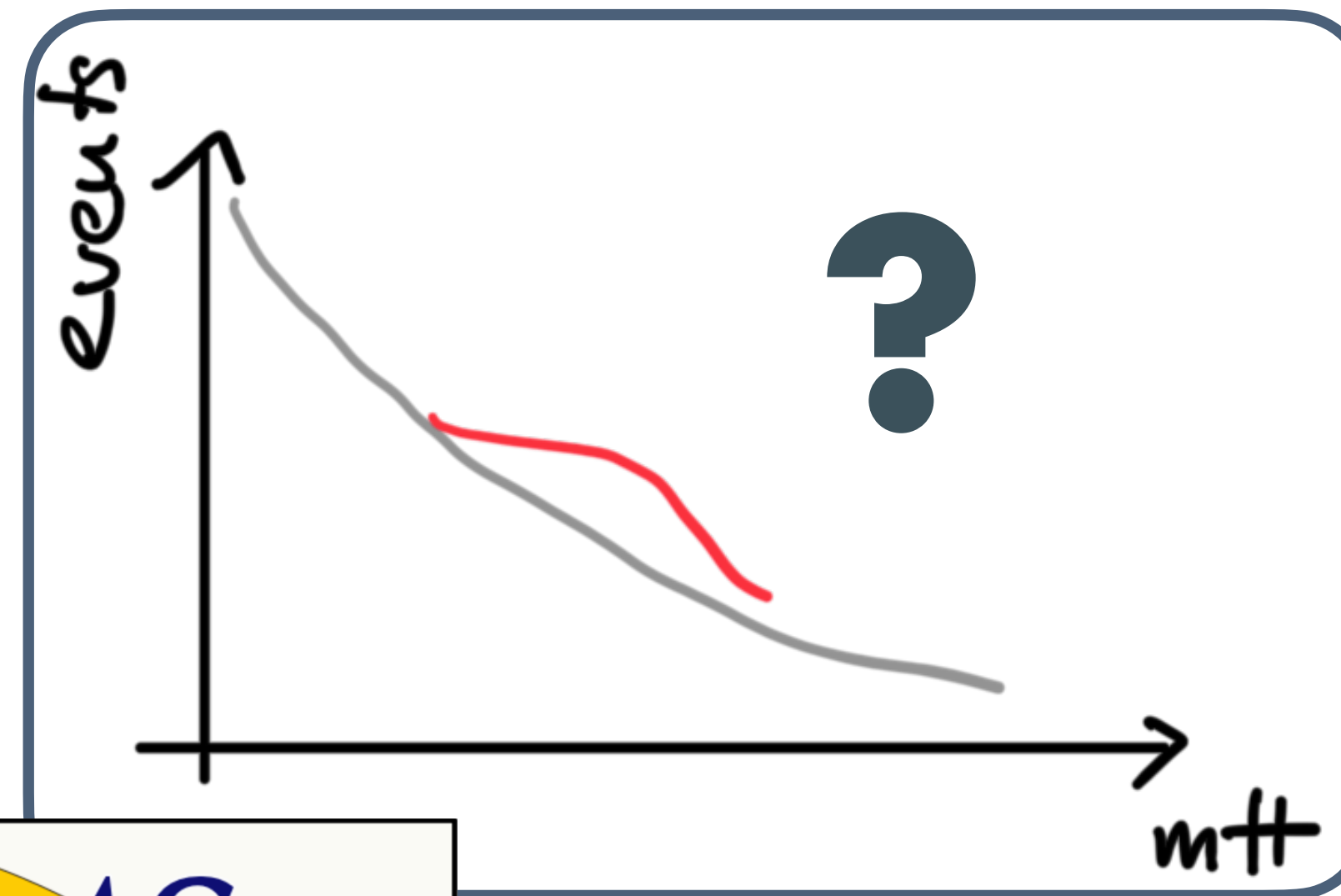
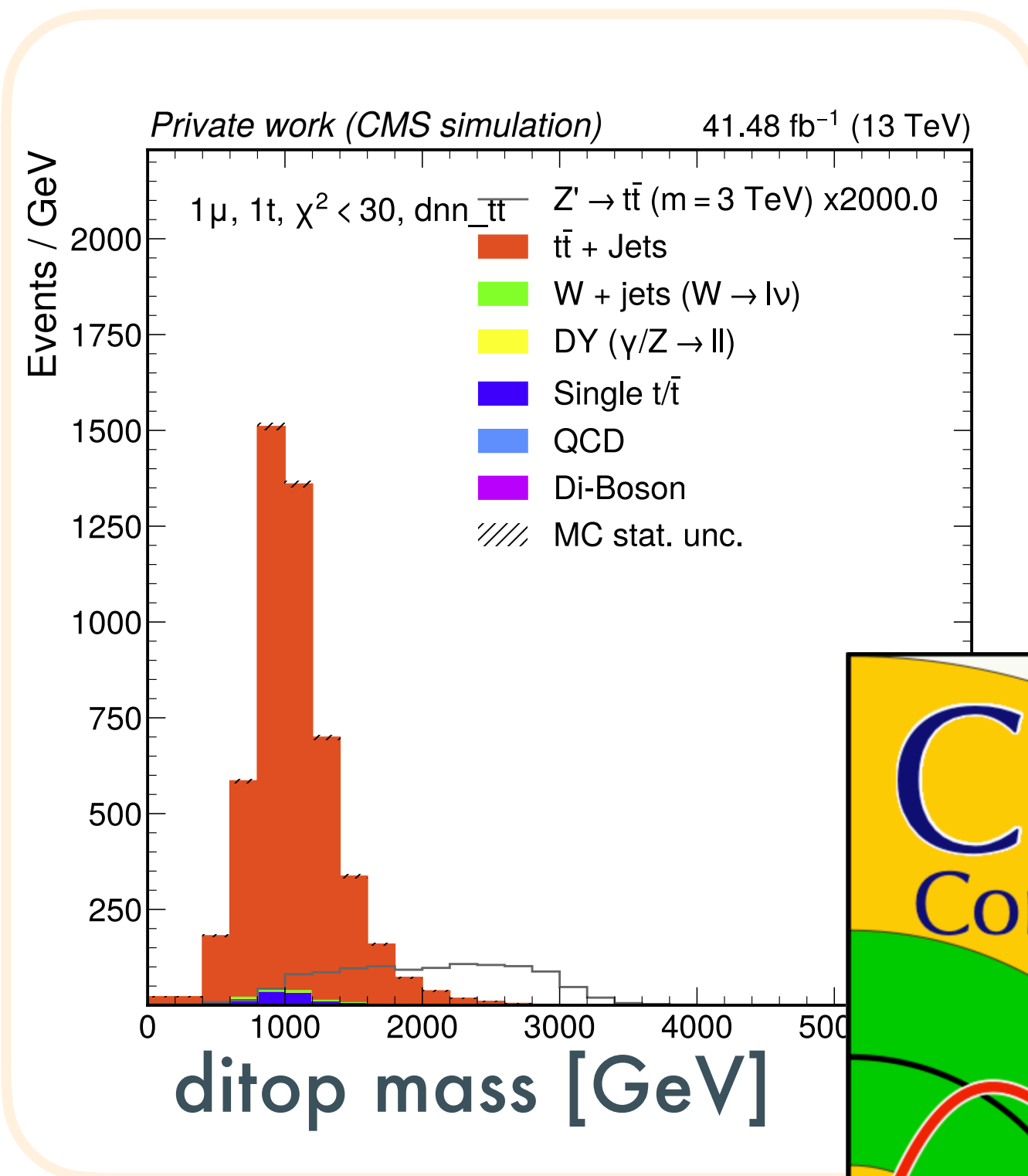
... In the Ditop Mass Spectrum



best hypothesis

selected events

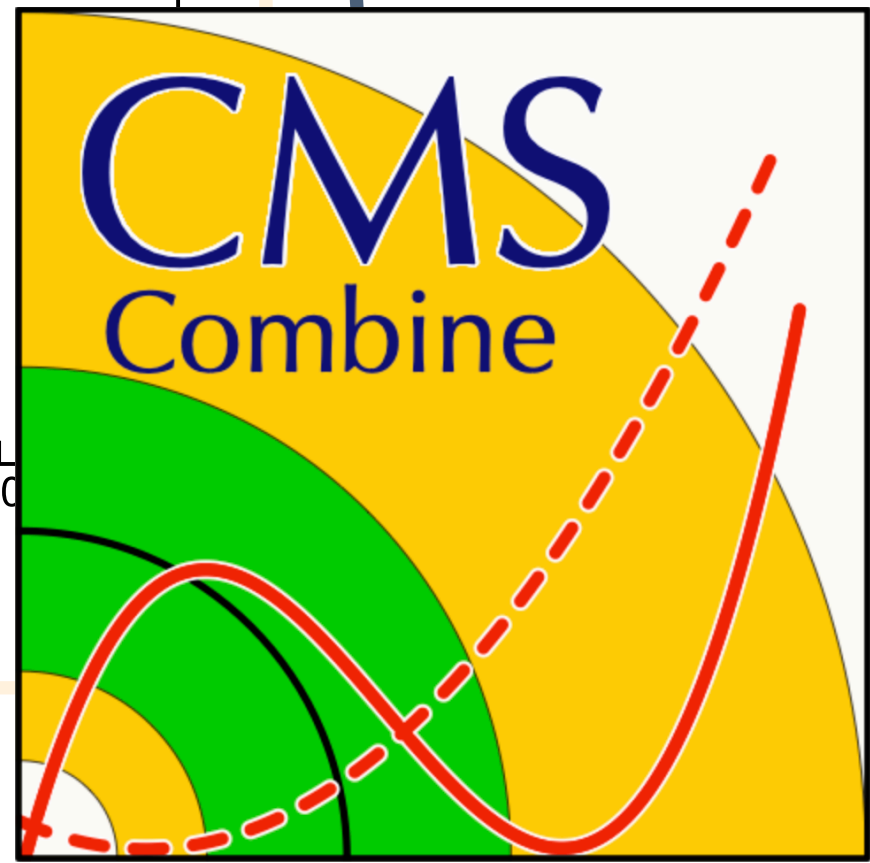
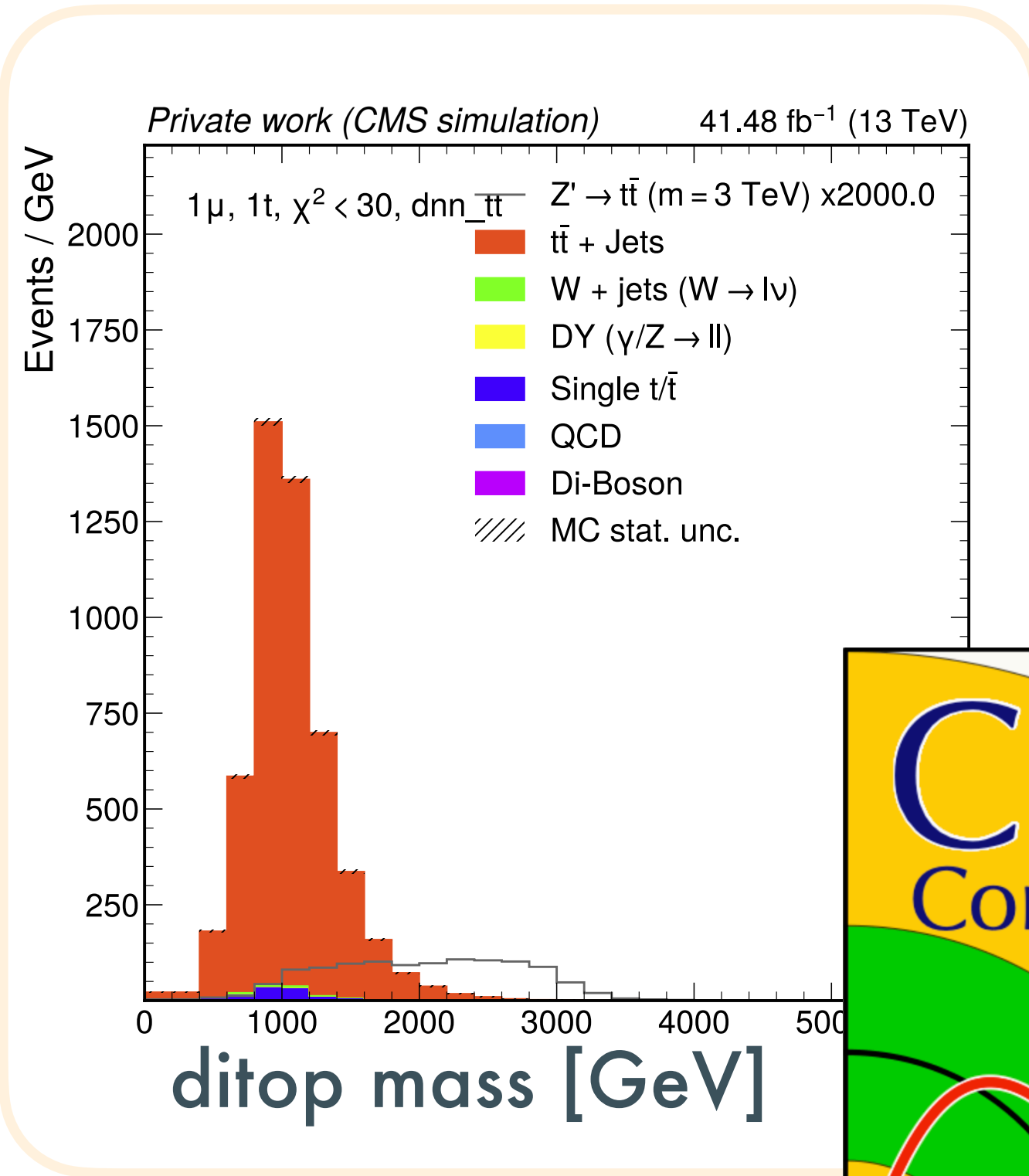
Statistical Analysis



CMS Combine

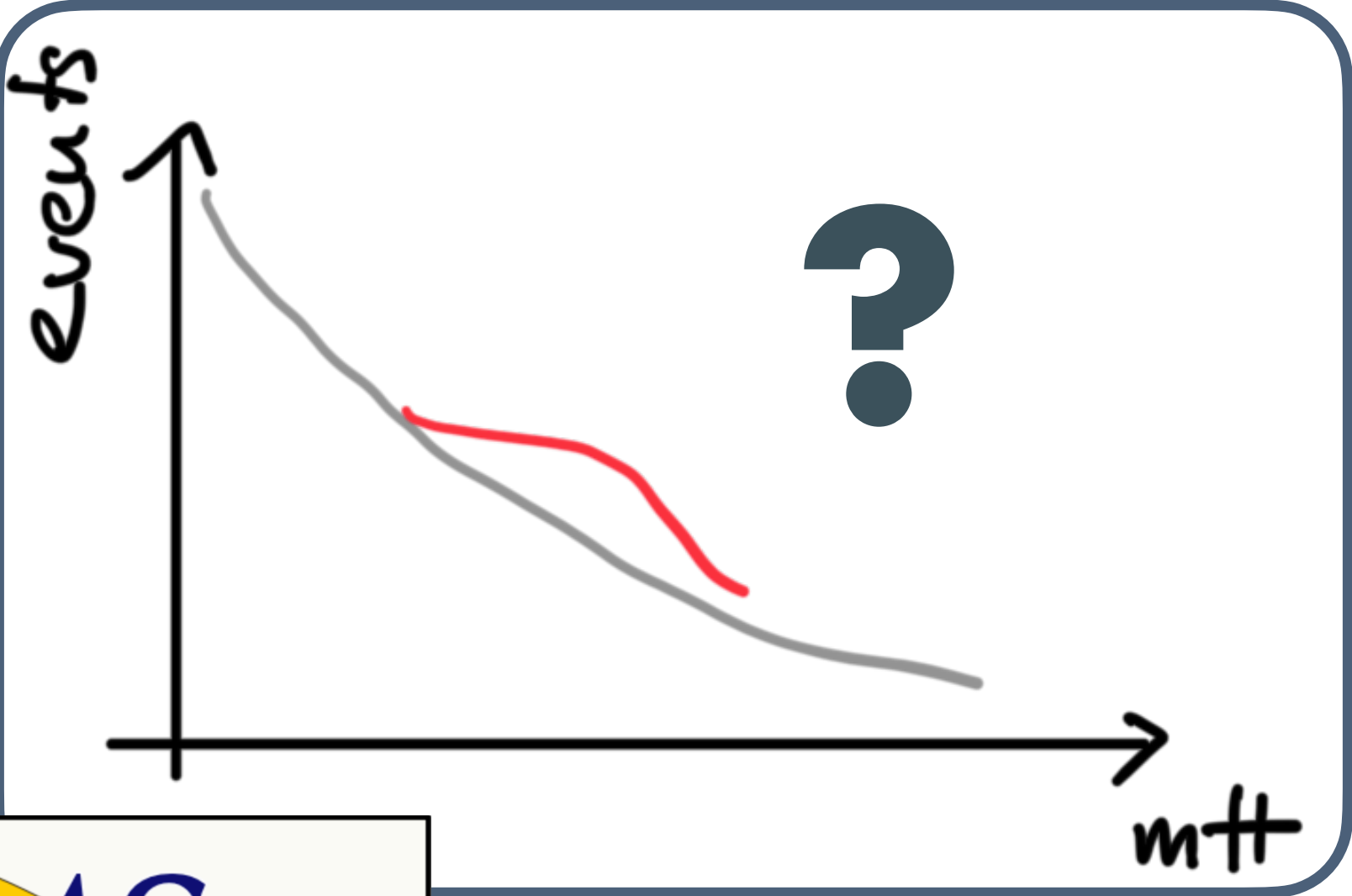
Comput.Softw.Big Sci. 8 (2024) 1, 19 - arXiv:2404.06614

Statistical Analysis

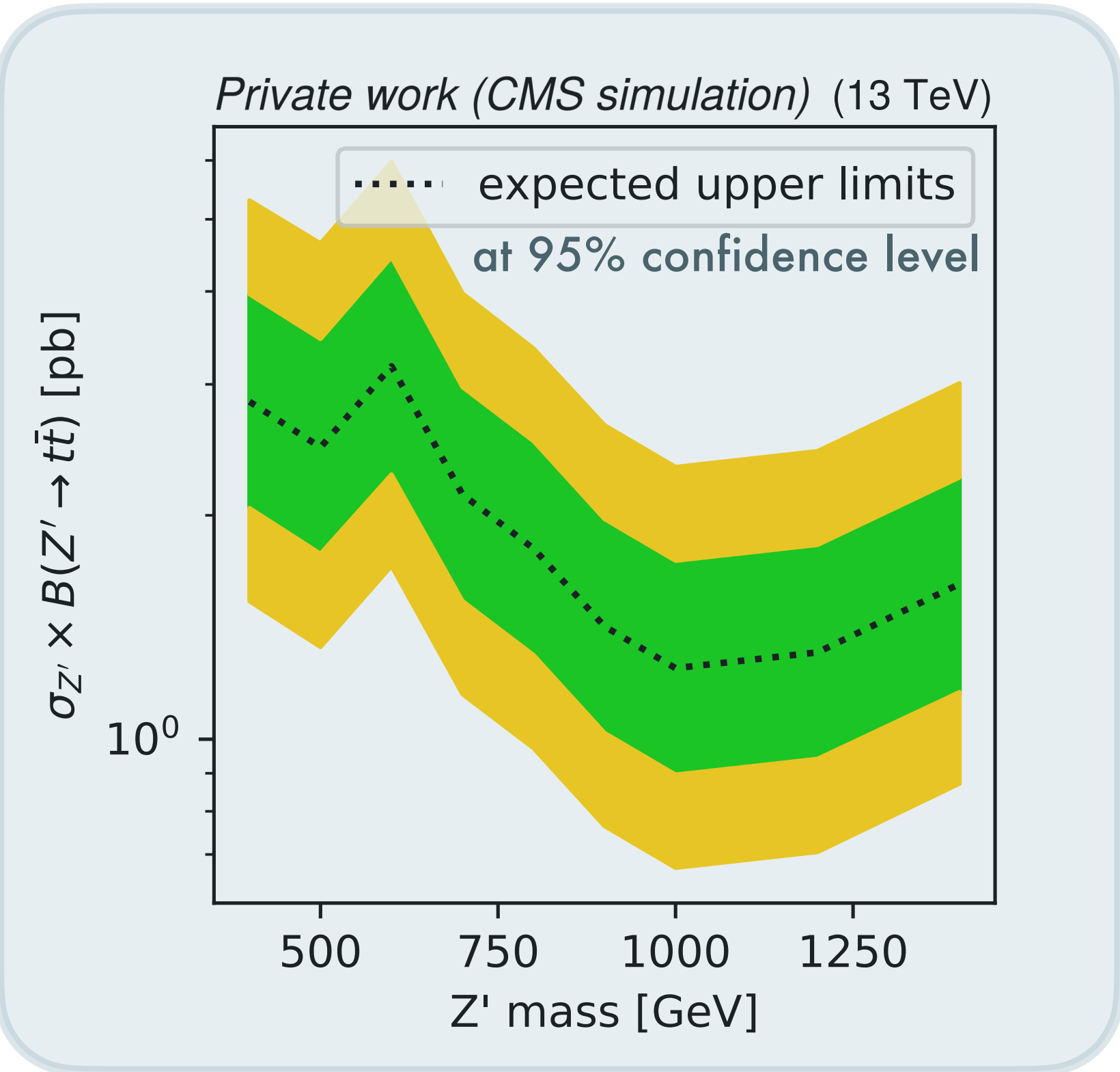


CMS Combine

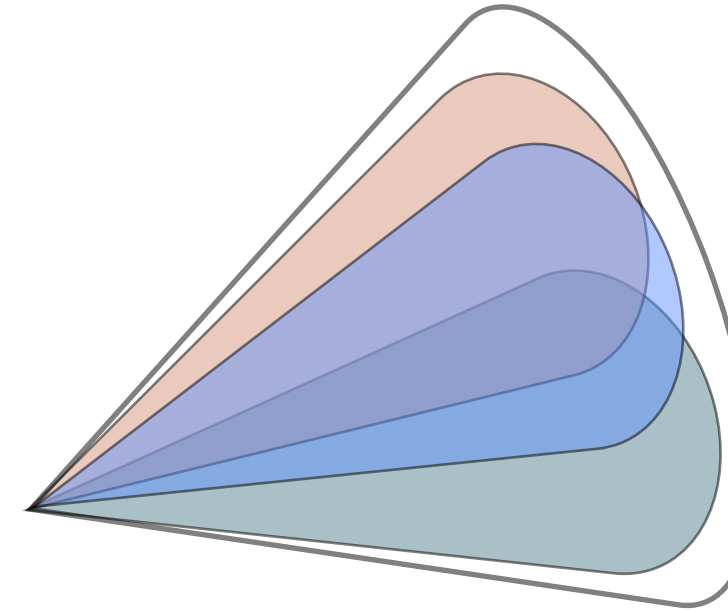
Comput.Softw.Big Sci. 8 (2024) 1, 19 - arXiv:2404.06614



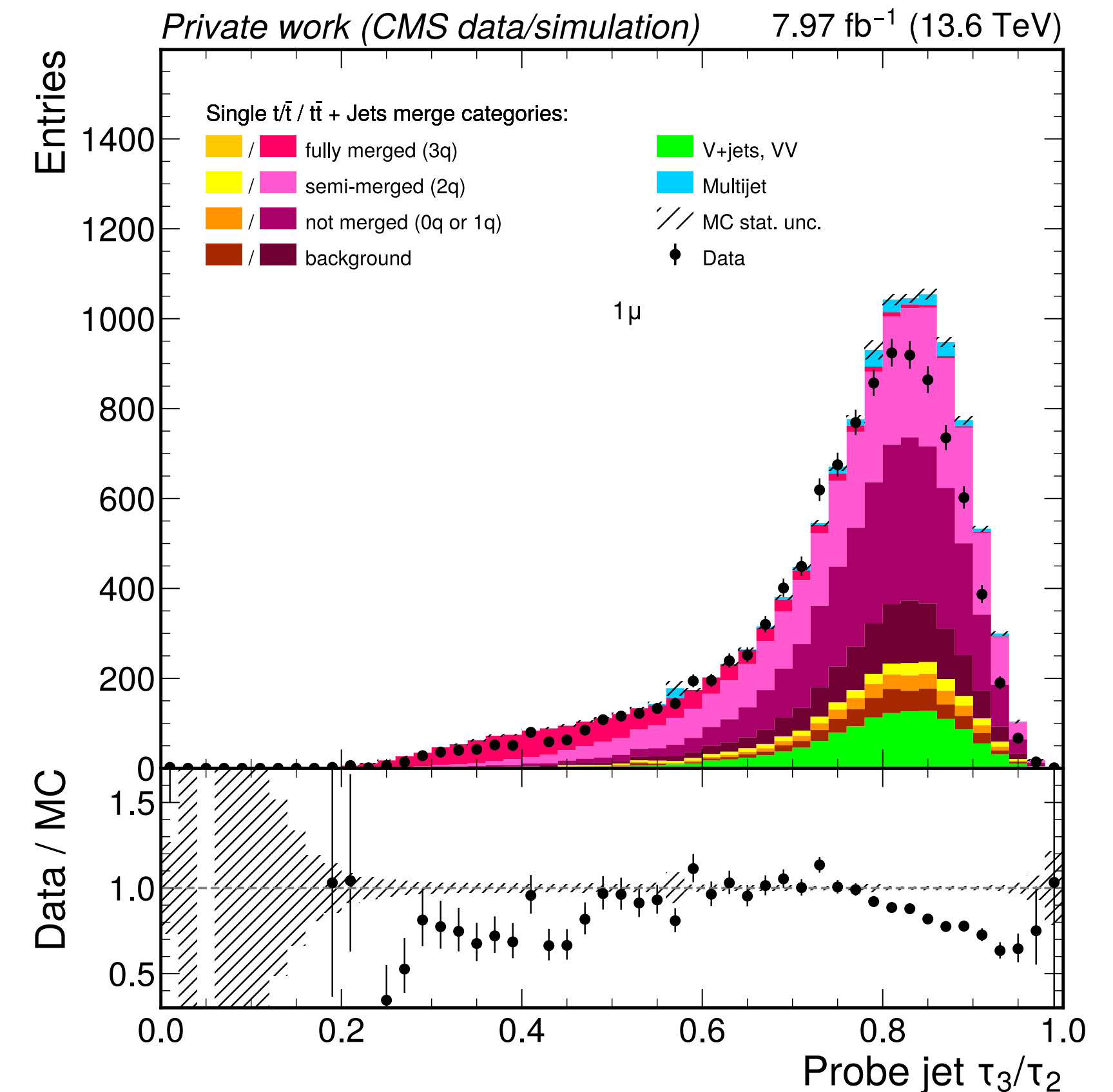
for illustration:



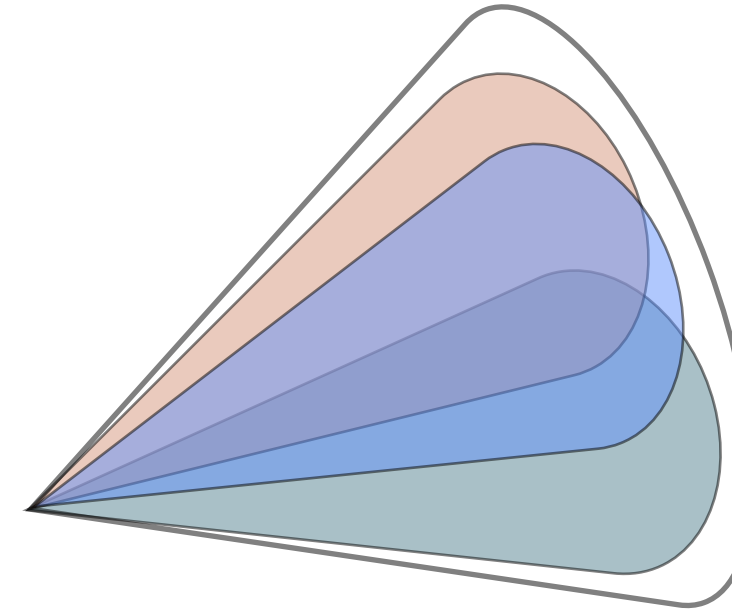
Summary



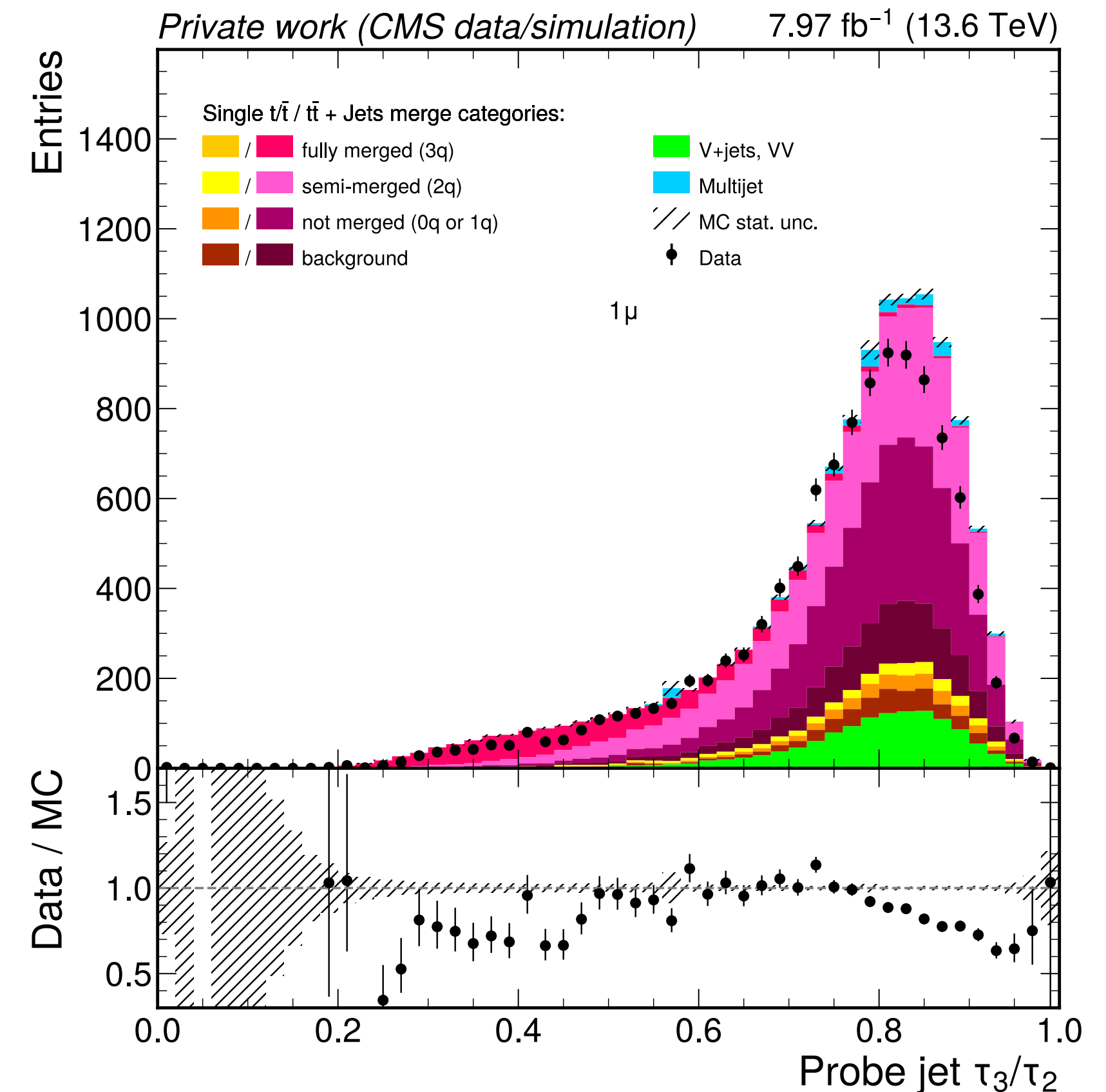
- identification of boosted objects essential step in many analyses
- **boosted top quarks** as potential window to BSM physics
here: heavy Z/Higgs bosons, non resonant $t\bar{t}$ production
 - Run 2 search expected to be published very soon
 - Run 3 search setup to look into new data
- understand **cut based top tagging algorithm**
using jet substructure and soft drop mass
 - derived working points and scale factors



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 expected to be published very soon
 - Run 3 search setup to look into new data
- understand **cut based top tagging algorithm**
using jet substructure and soft drop mass
 - derived working points and scale factors



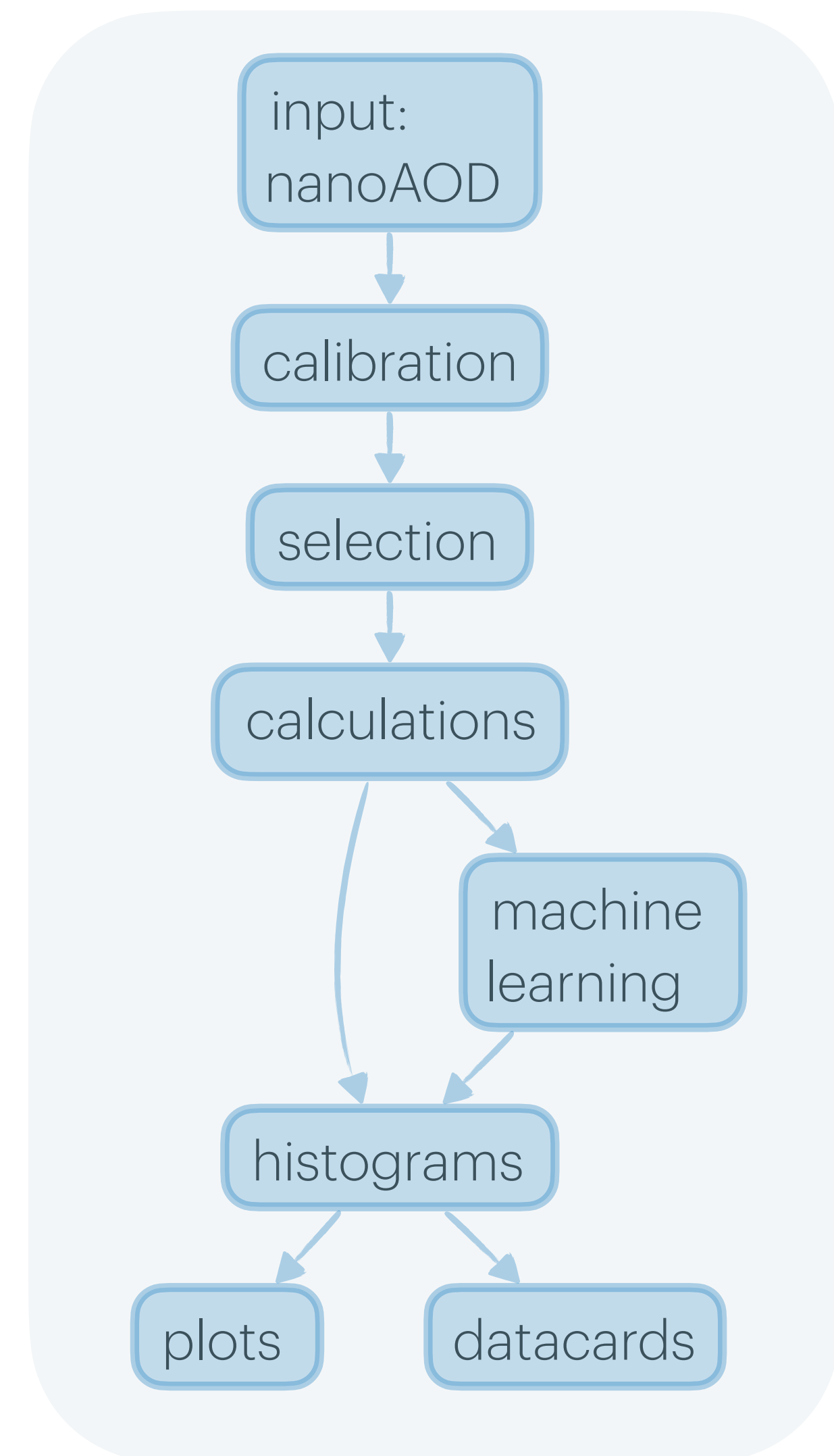
Stay tuned!

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



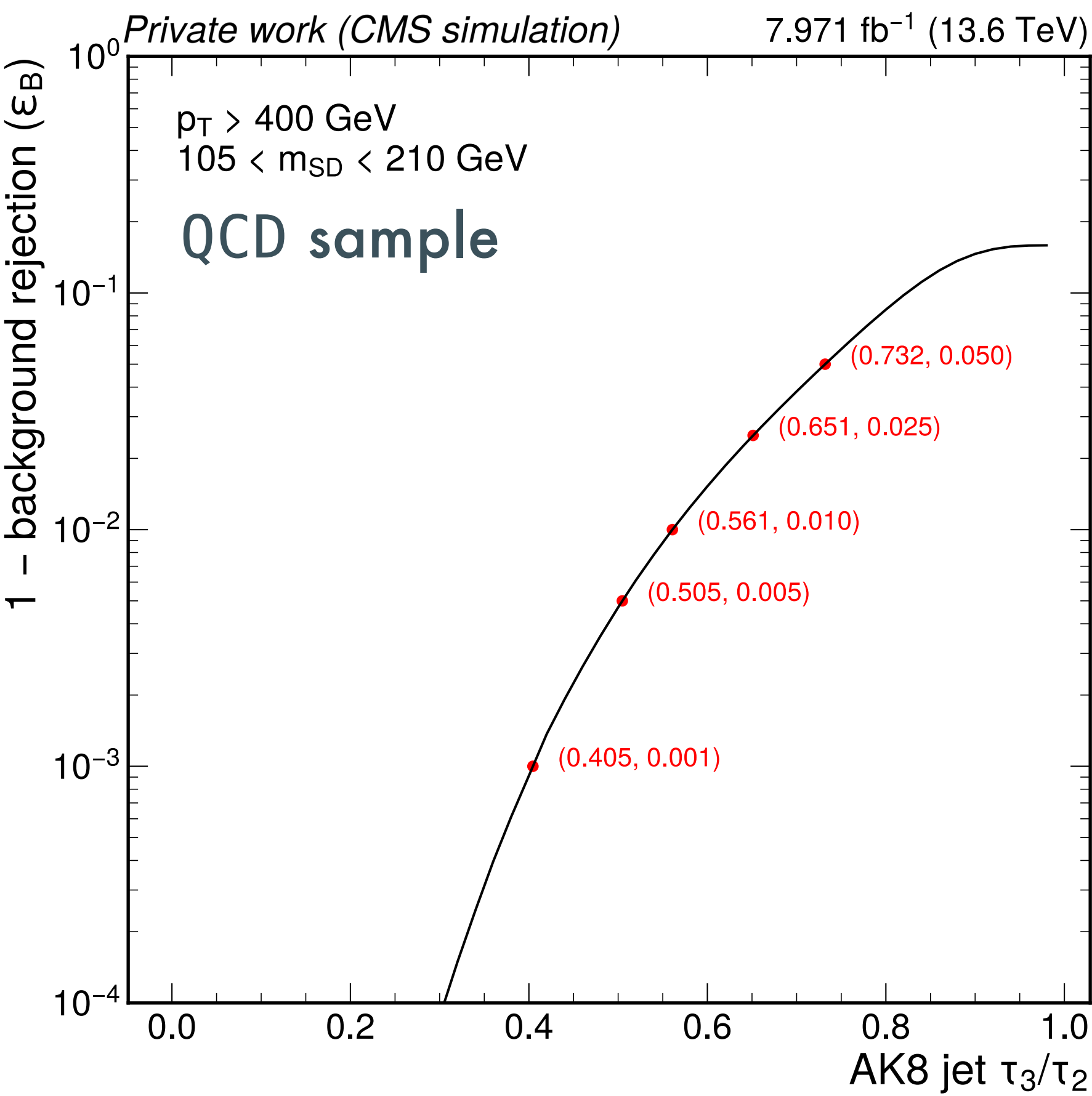
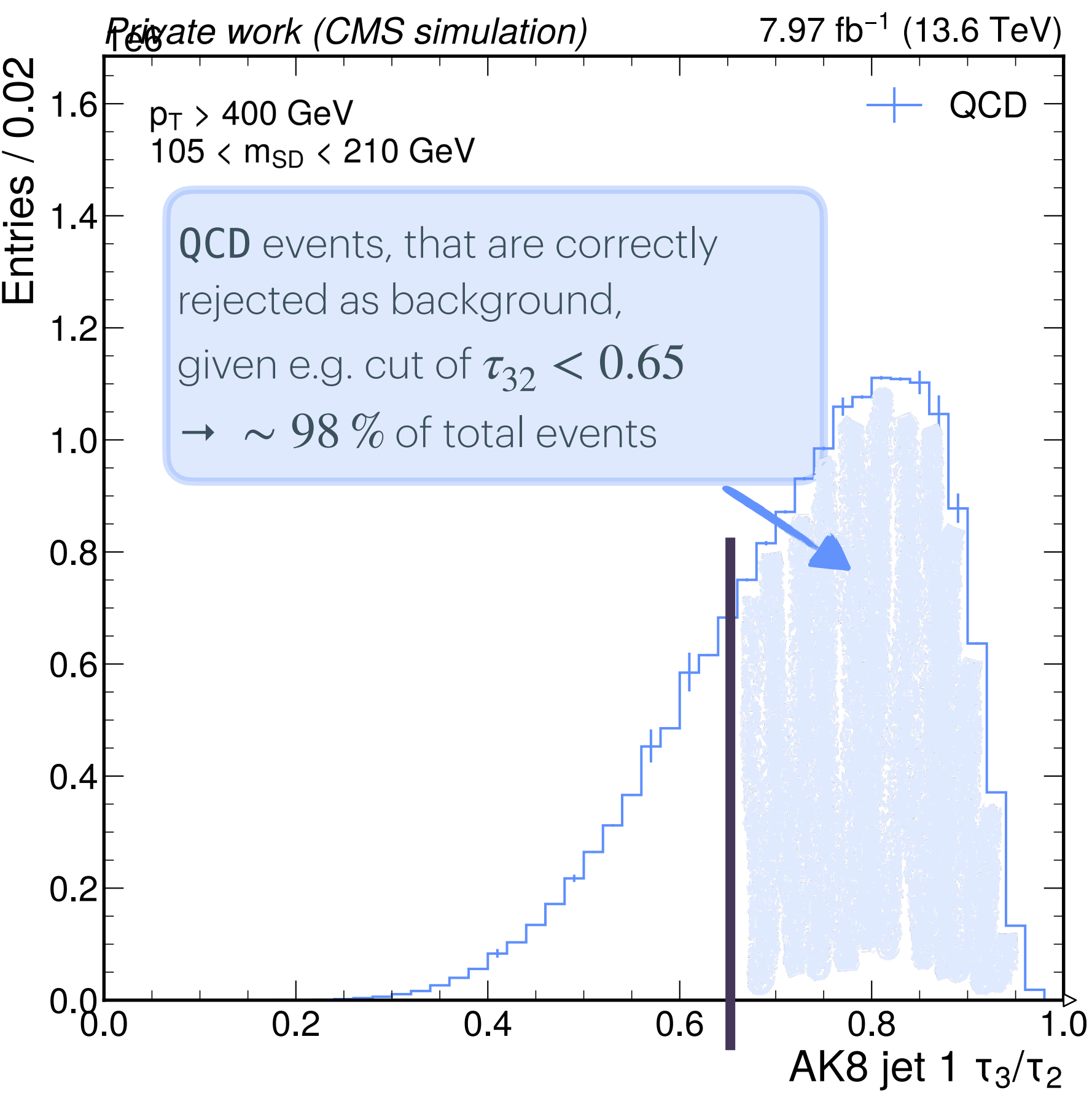
CF [repo](#)
column
flow

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
 - $|\eta| < 2.4$
 - $p_T > 400 \text{ GeV}$
→ boosted regime
 - in $t\bar{t}$ bar sample to determine signal efficiency:
uniquely matched to parton-level top quark
 - additional softdrop mass constraint:
 $105 < m_{SD} < 210 \text{ GeV}$
→ window around top mass

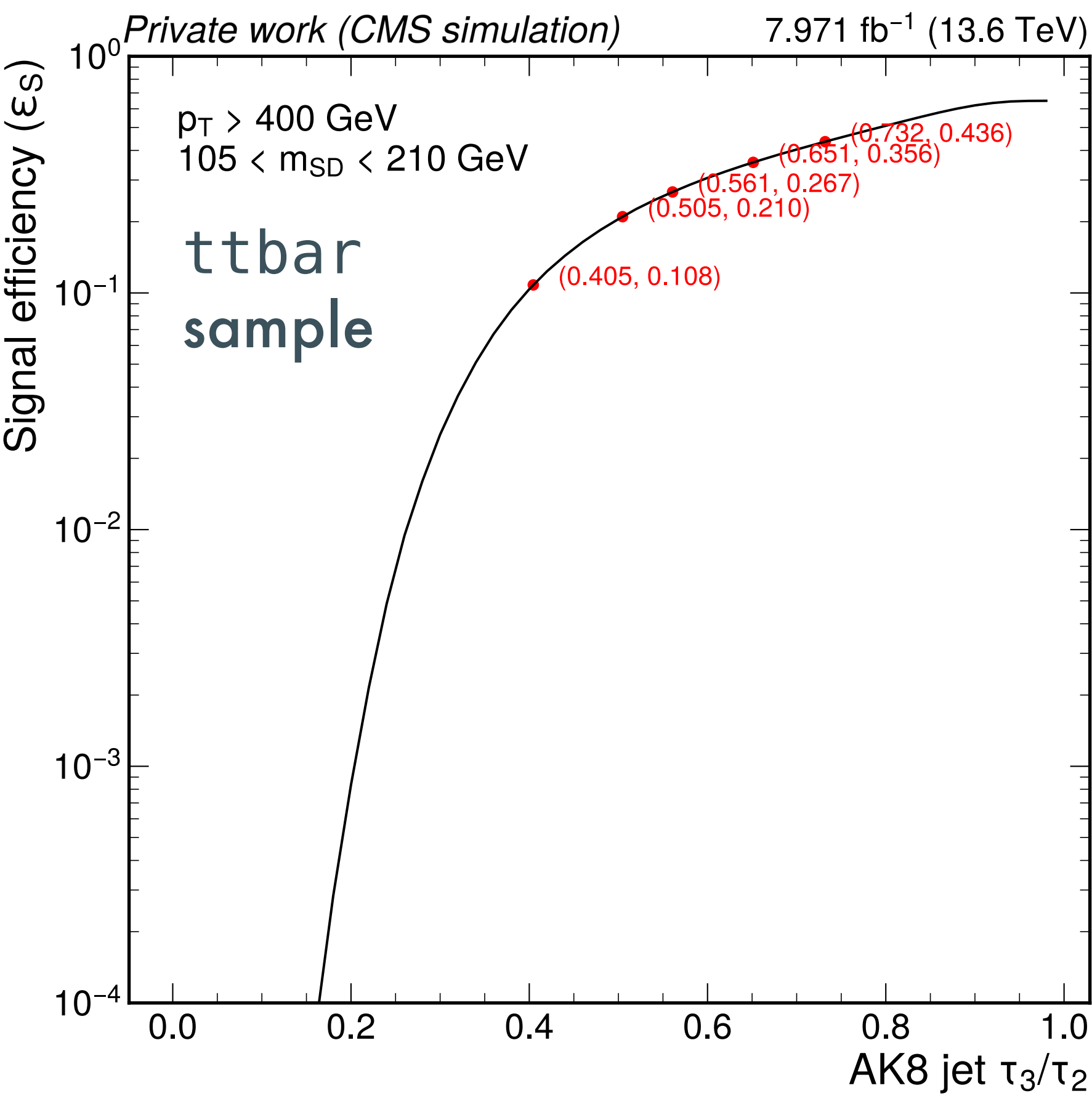
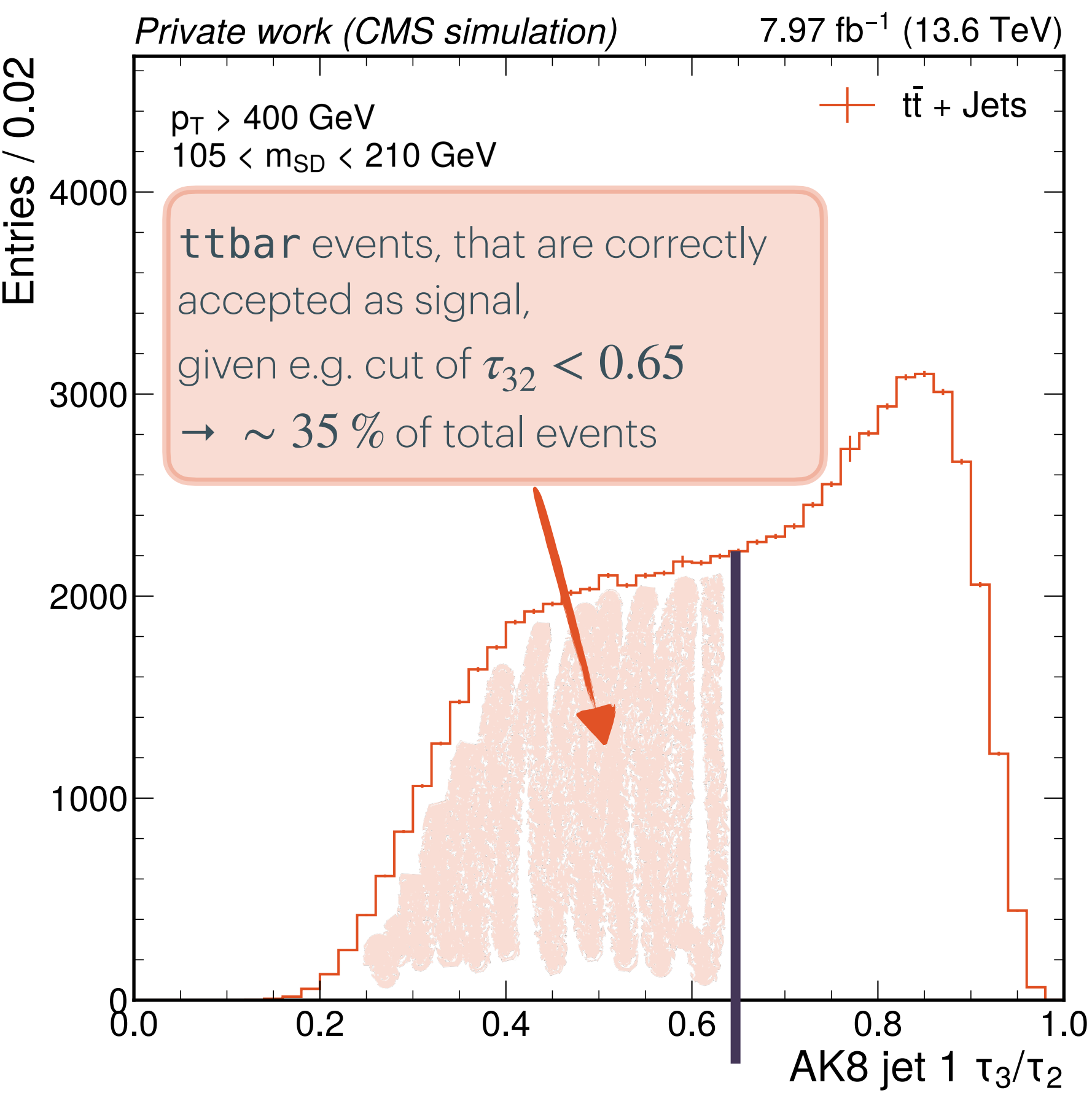
Mis-Tag Rate



number of QCD events with a FatJet passing selection AND (incorrectly) tagged as a top jet given the τ_{32} cut

total number of events in QCD dataset

Signal Efficiency



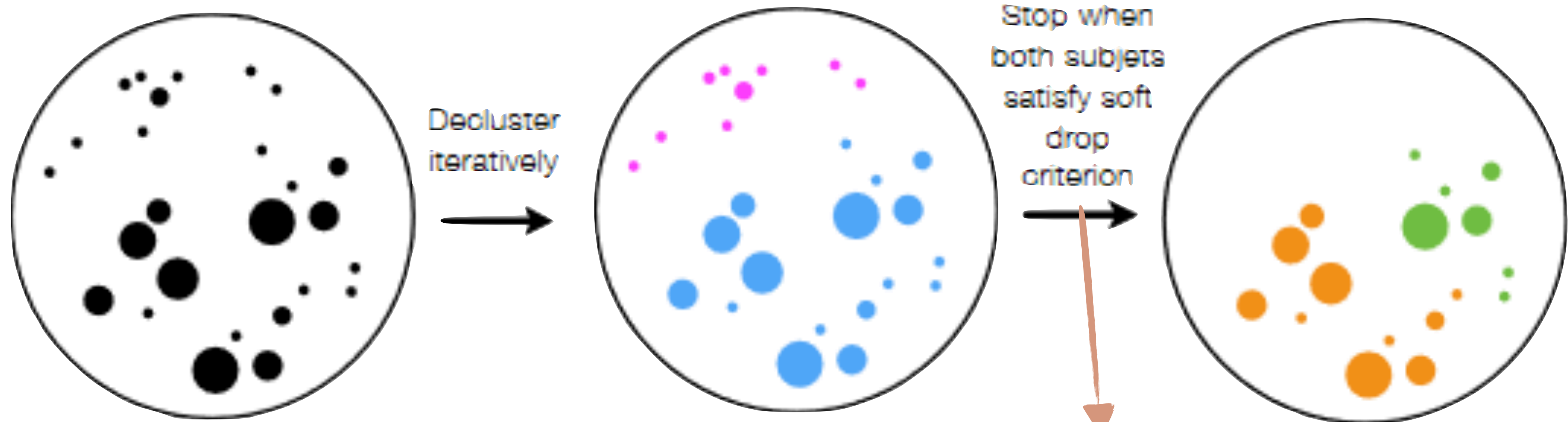
number of tt̄bar events with a FatJet passing selection AND (correctly) tagged as a top jet given the τ_{32} cut

total number of events in tt̄bar dataset

m_{SD} Of FatJets

Soft Drop

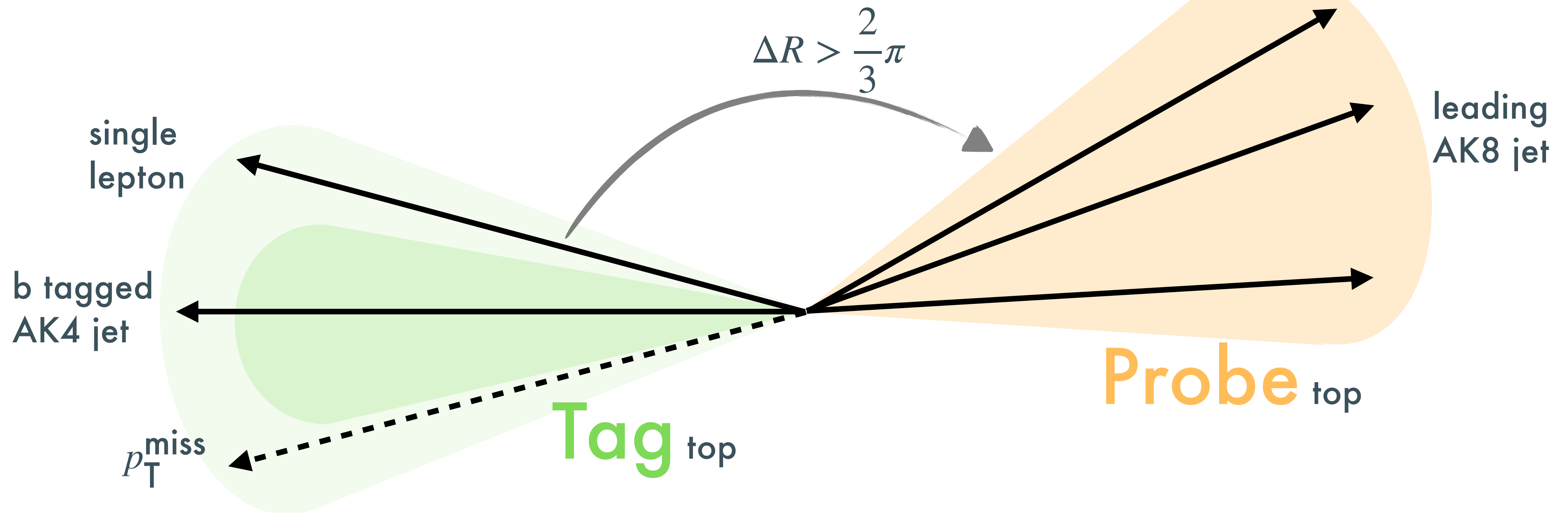
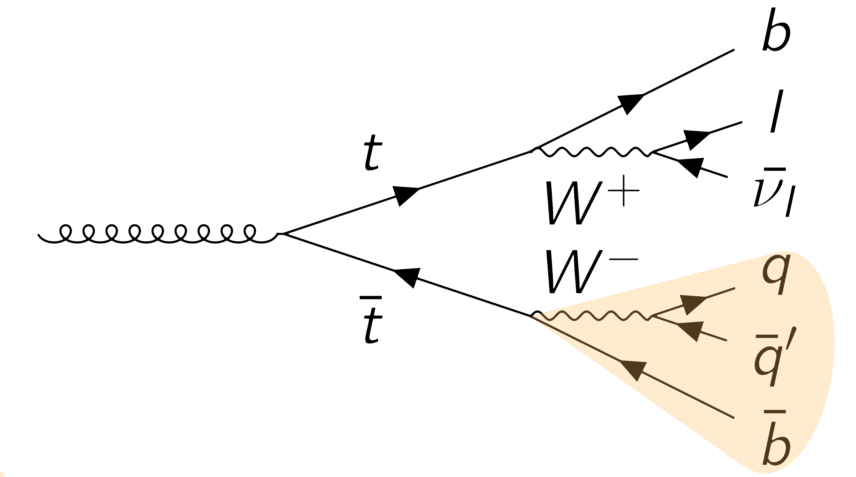
image credit: J. Dolen
from [cms-opendata-workshop](https://cms-opendata-workshop.github.io/)



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

→ fails for wide-angle soft radiation

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

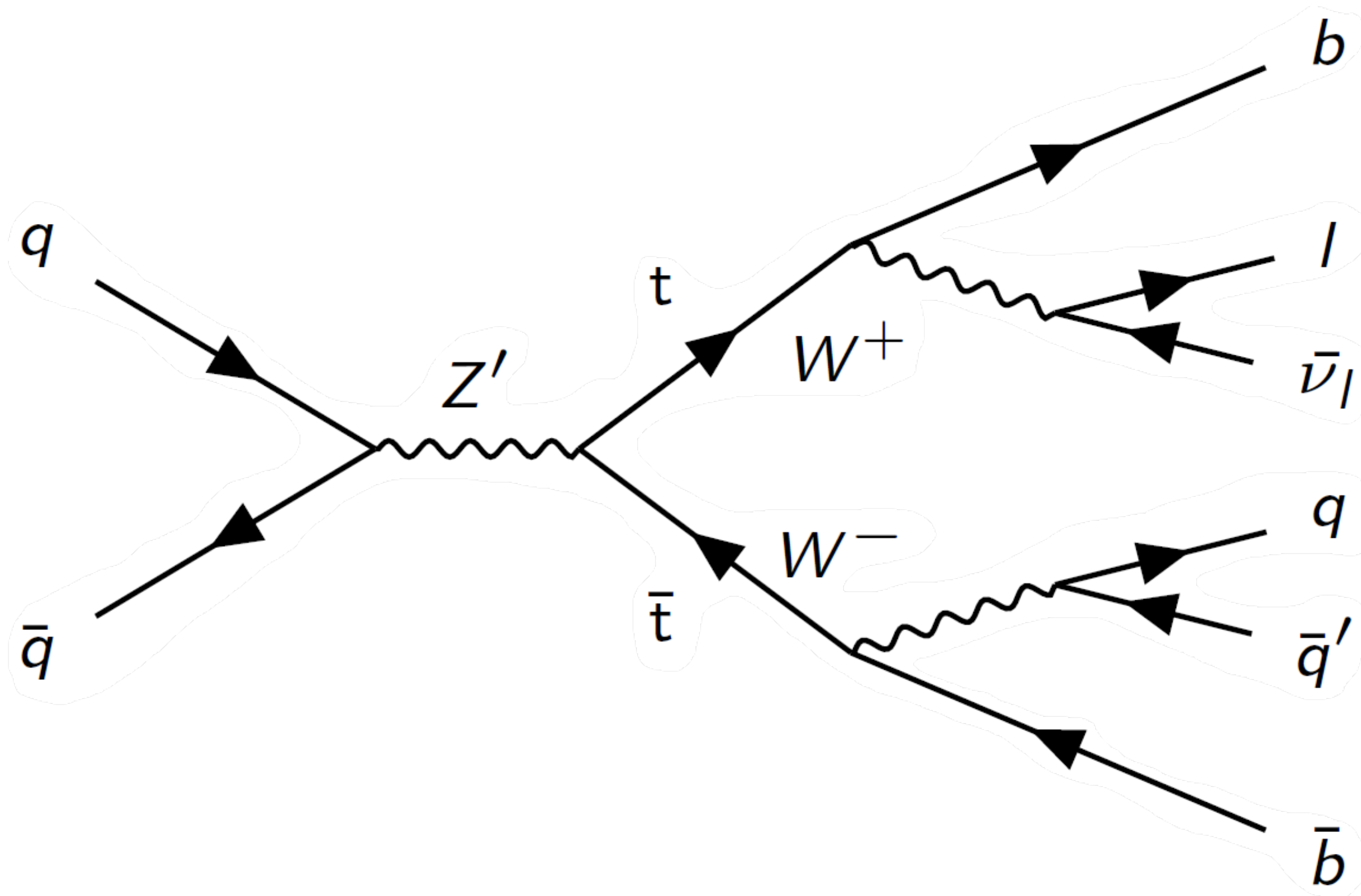


- single lepton
- $p_{T, \text{lep}} > 55 \text{ GeV}$
- b-tagged AK4 jet close to lepton
- $p_{T, \text{AK4}} > 15 \text{ GeV}$
- $p_T^{\text{miss}} > 50 \text{ GeV}$
- $p_{T(W_{\text{lep}})} > 150 \text{ GeV}$
- $p_{T, \text{AK8}} > 300 \text{ GeV}$
- jet-lepton 2D isolation

Baseline Selection

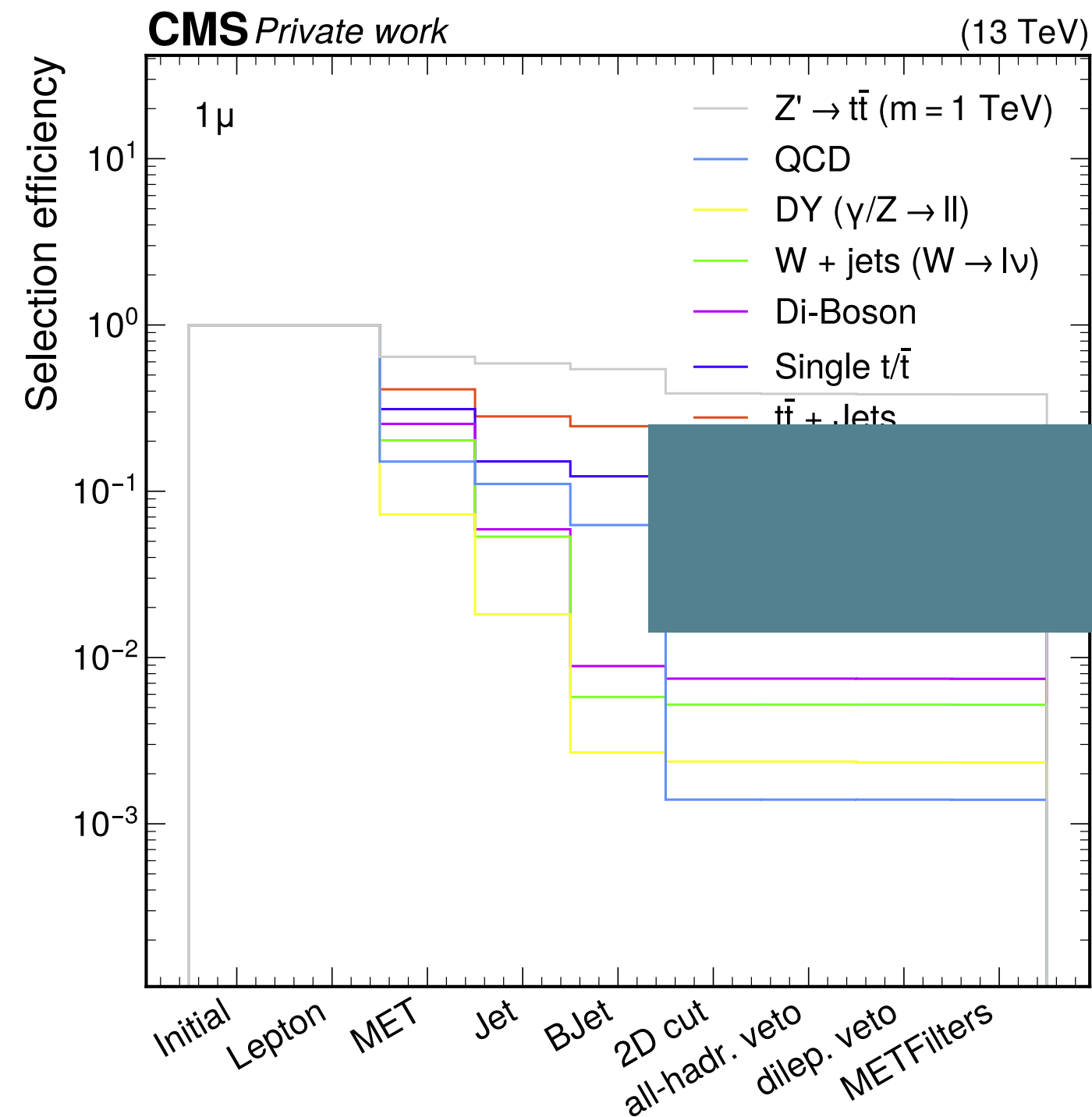
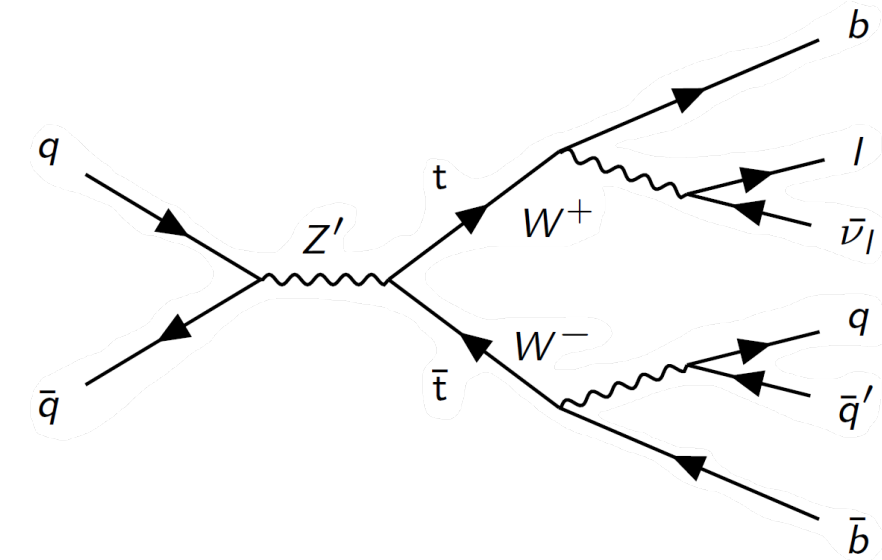
— in numbers —

muon/electron channel

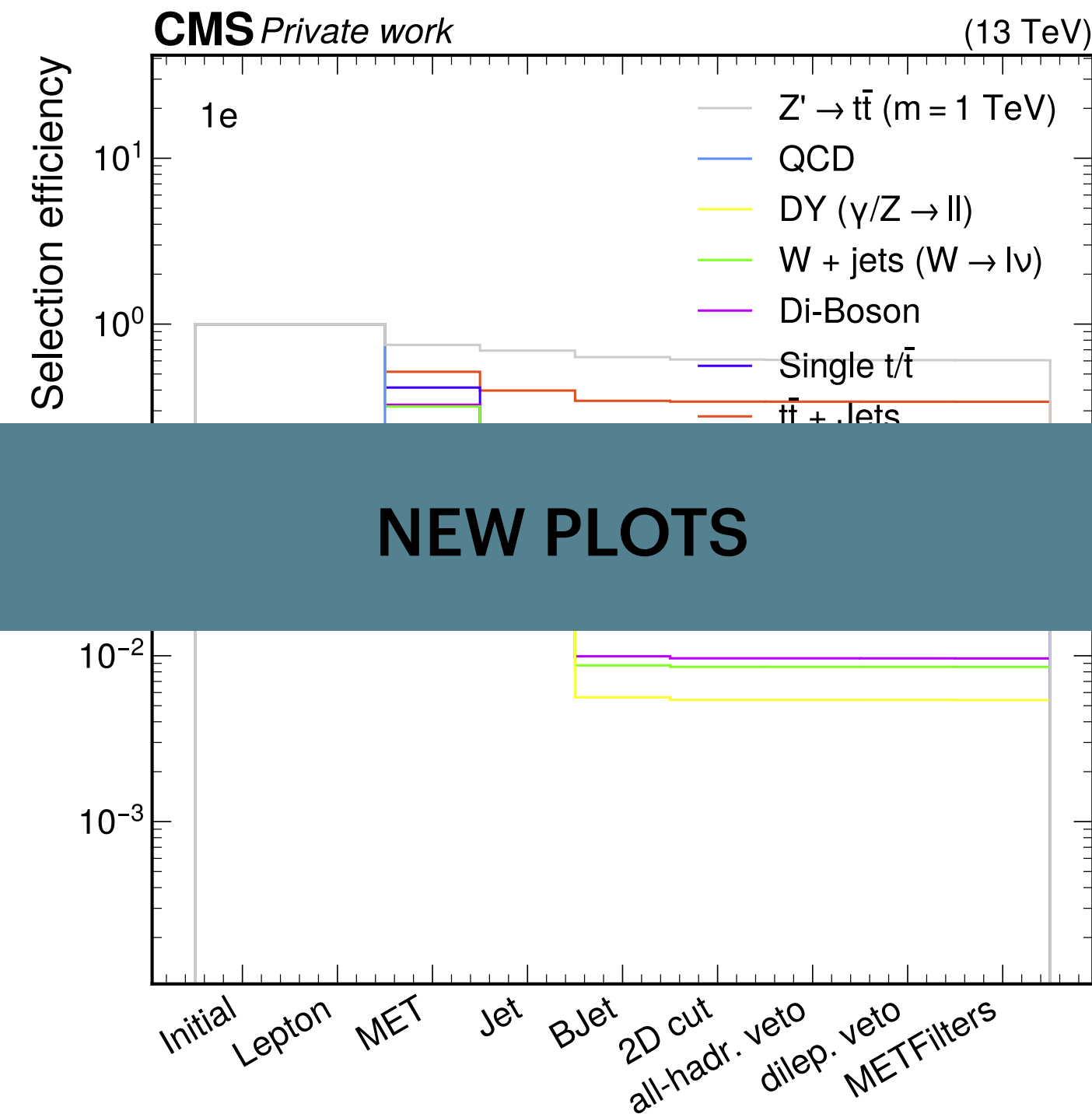


- lepton selection:
 - single muon/electron trigger
 - low p_T : $30/35 < p_T < 55/120$ GeV and isolation
 - high p_T : $p_T > 55/120$ GeV and 2D isolation
 - $|\eta^{\mu/e}| < 2.4$
- $p_T^{\text{miss}} > 70/60$ GeV
- at least two jets with $|\eta| < 2.5$ and:
 - $p_T^{\text{jet1}} > 50/50$ GeV
 - $p_T^{\text{jet2}}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

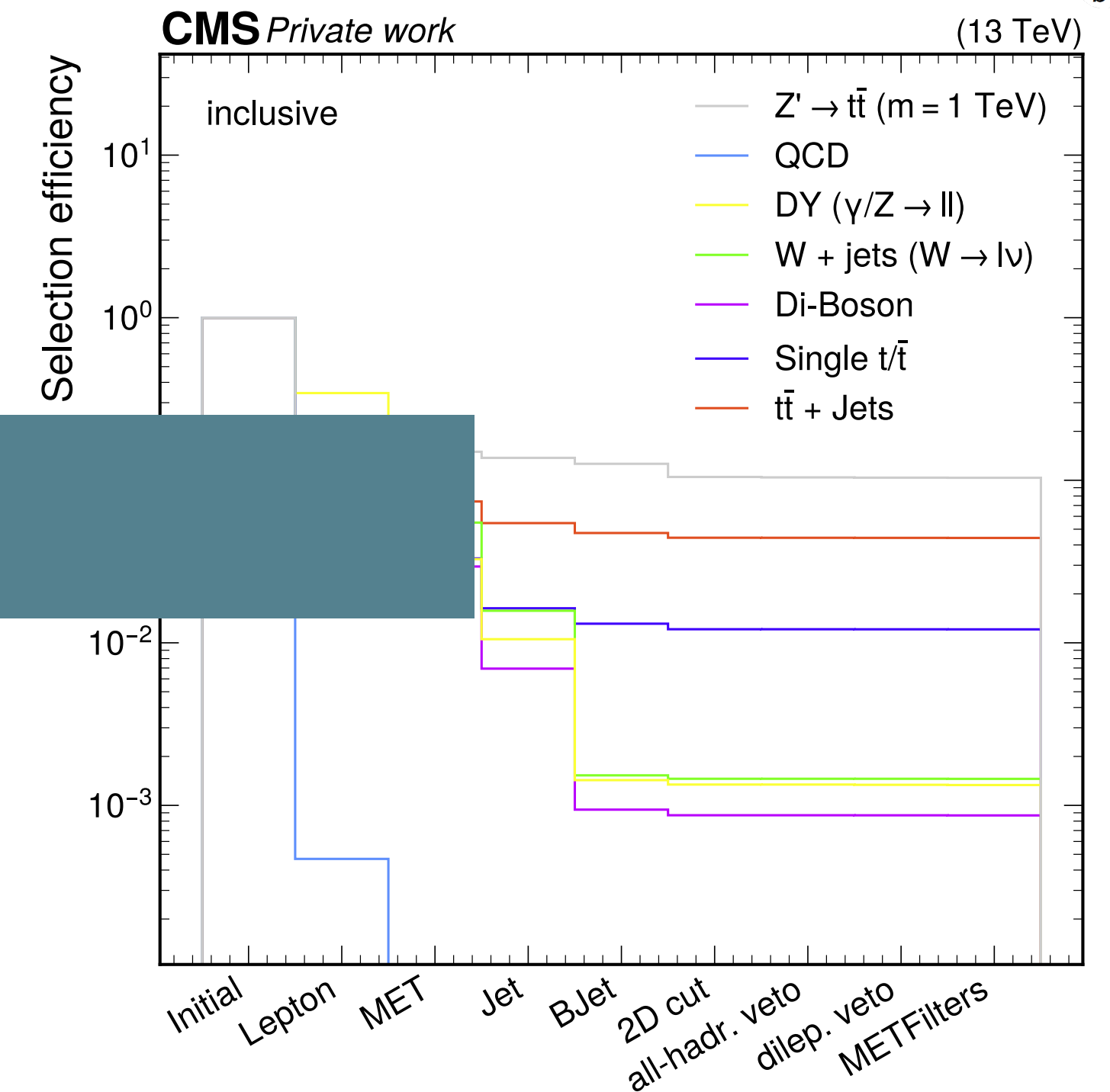
Selection Efficiencies



muon channel



electron channel

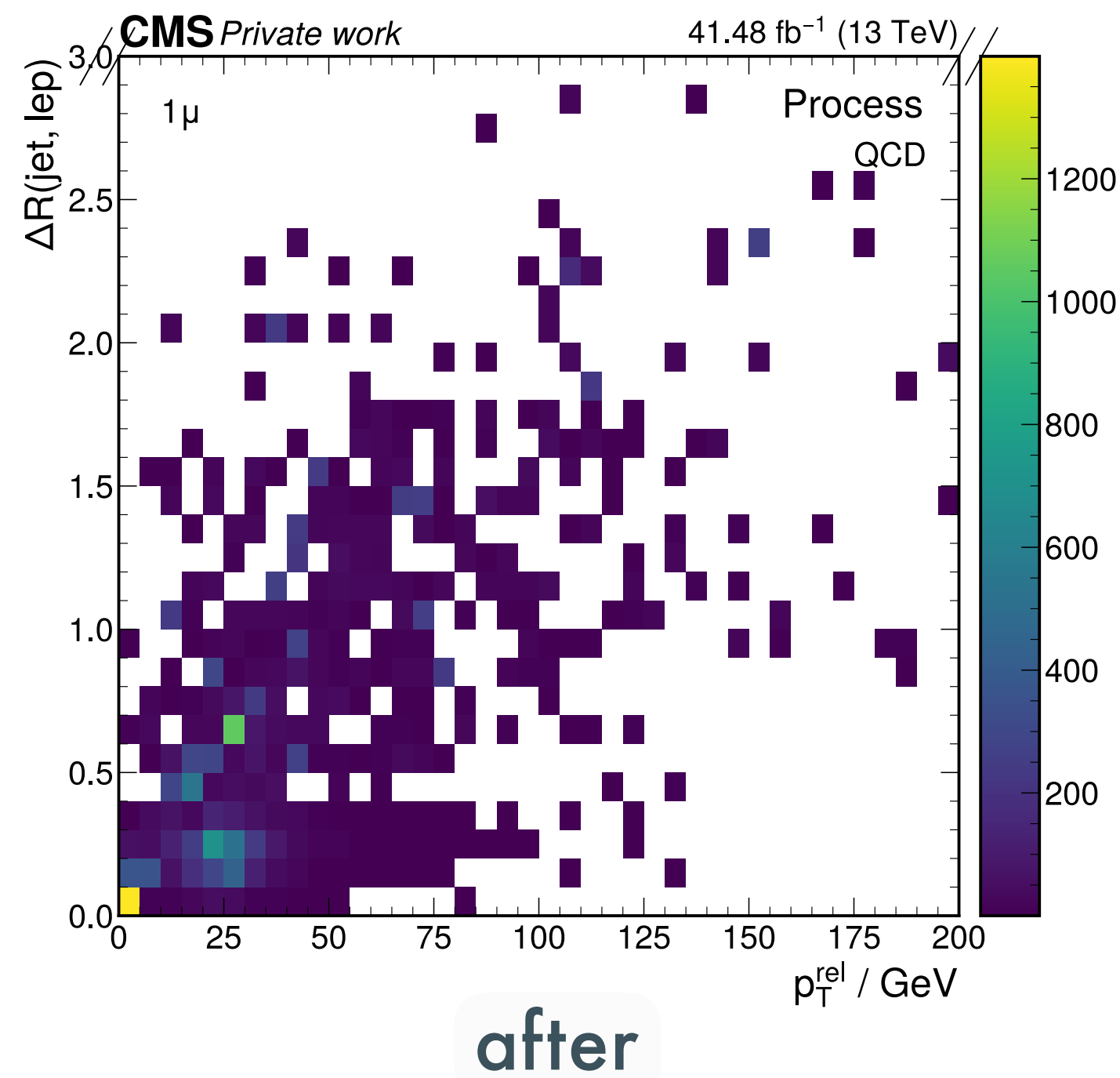
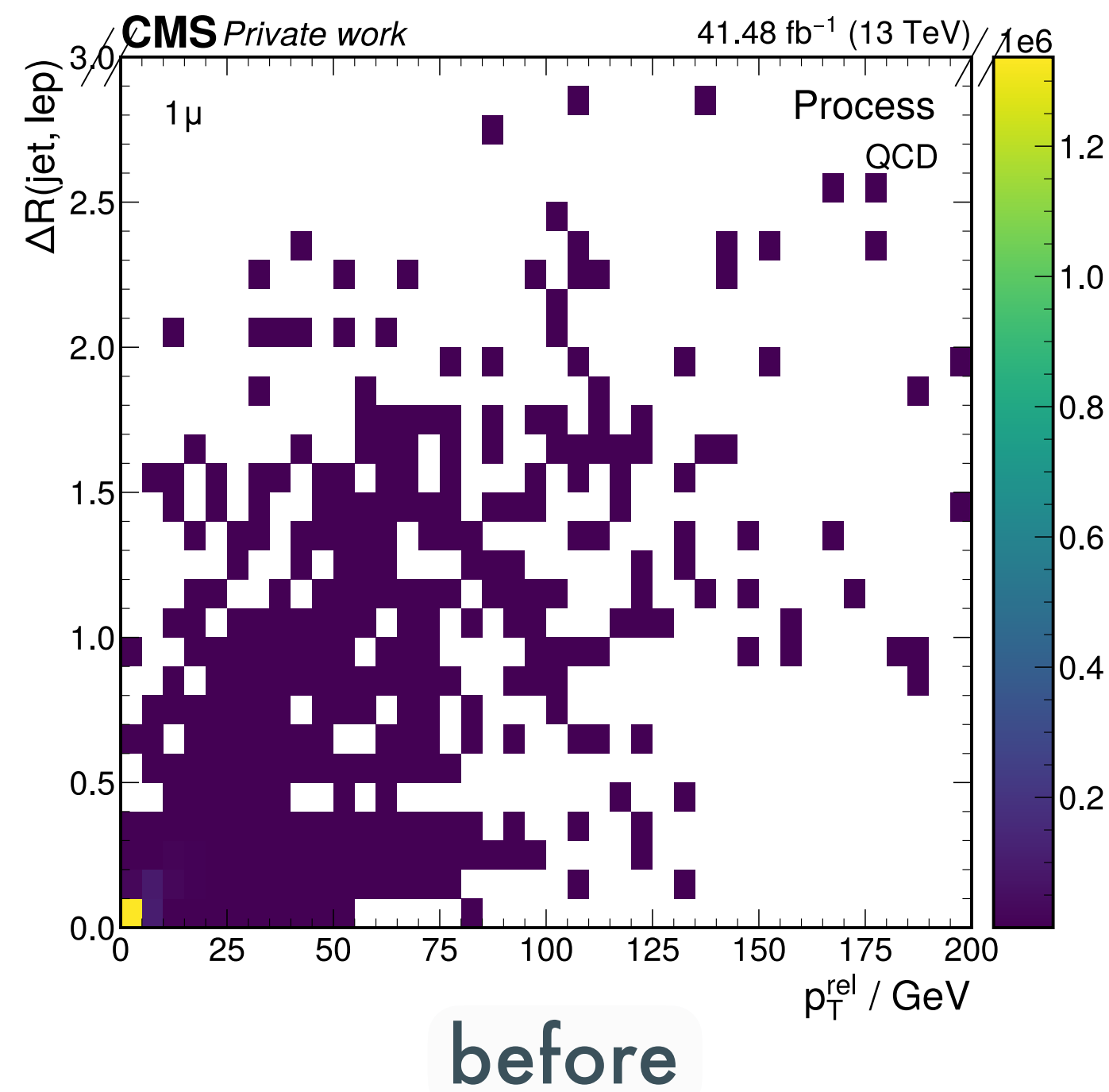
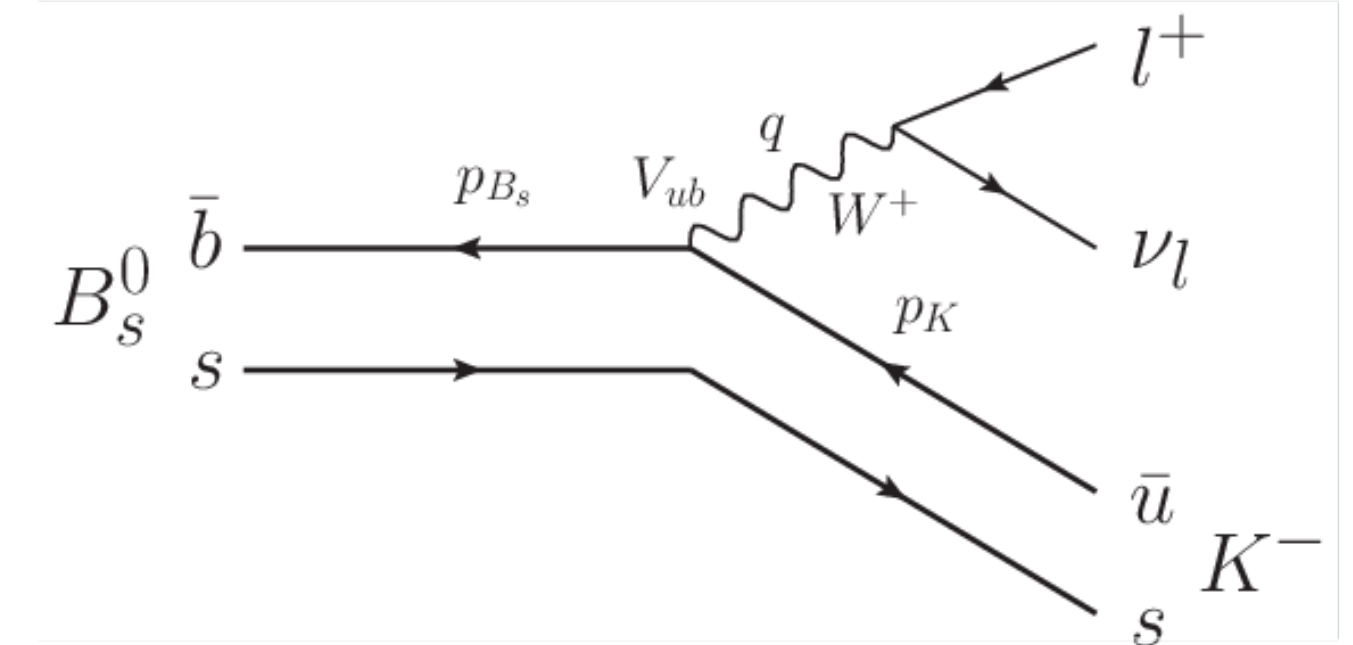


inclusive channel

Jet-Lepton 2D Cut

— filtering leptonic B meson decays —

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



- 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

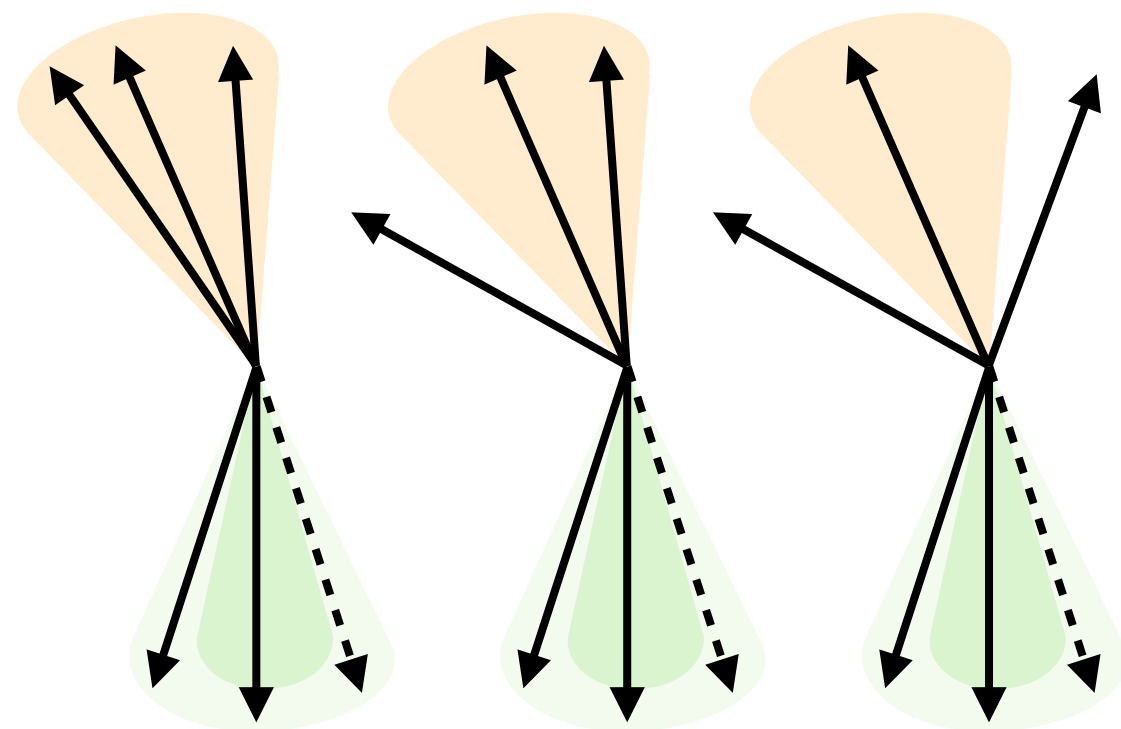
Combine Physics Model

- 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 Uncertainties

— 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

Impacts

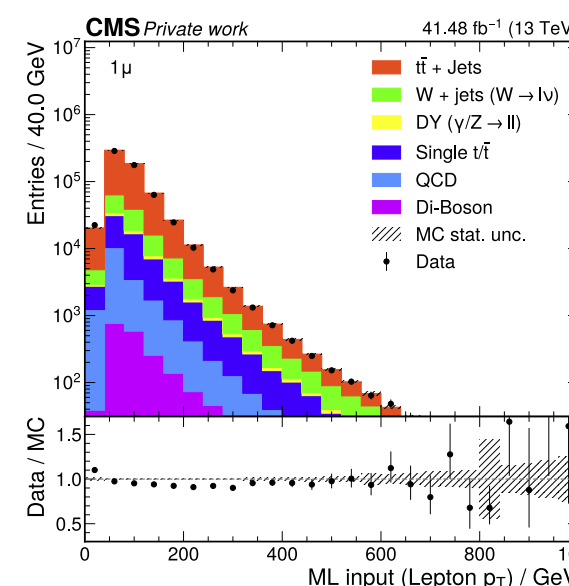
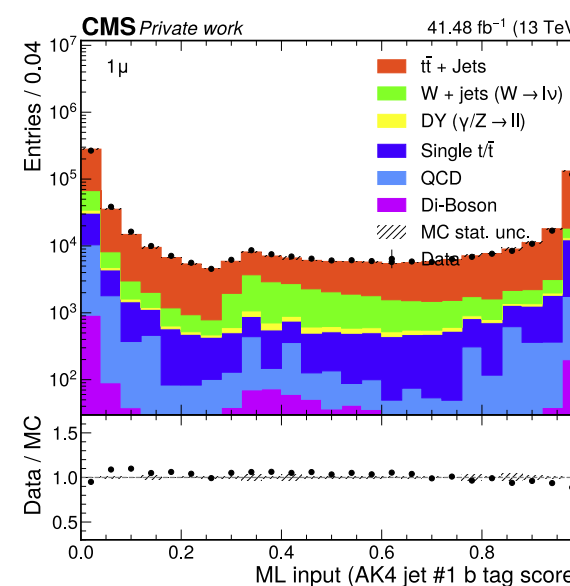
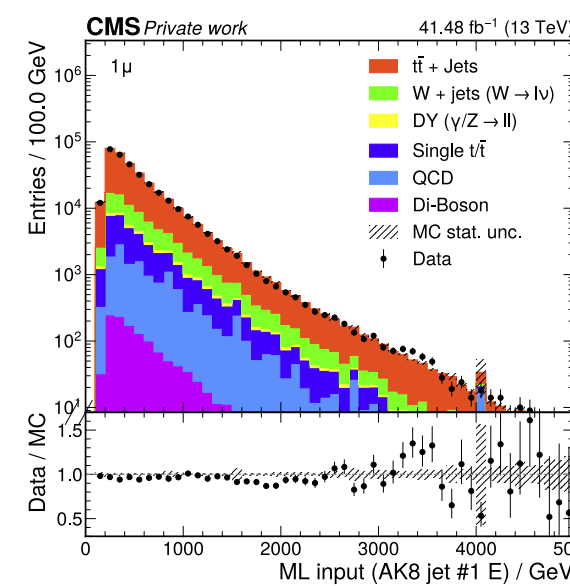
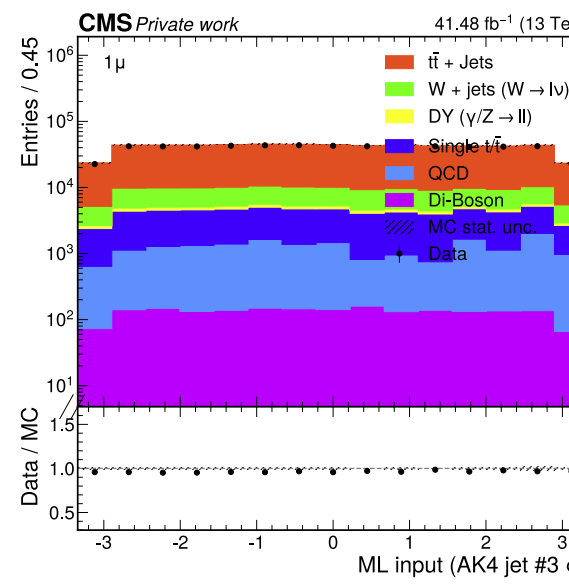
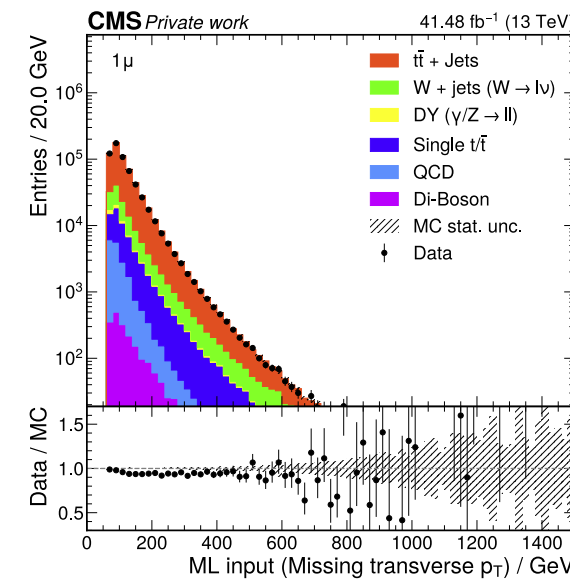
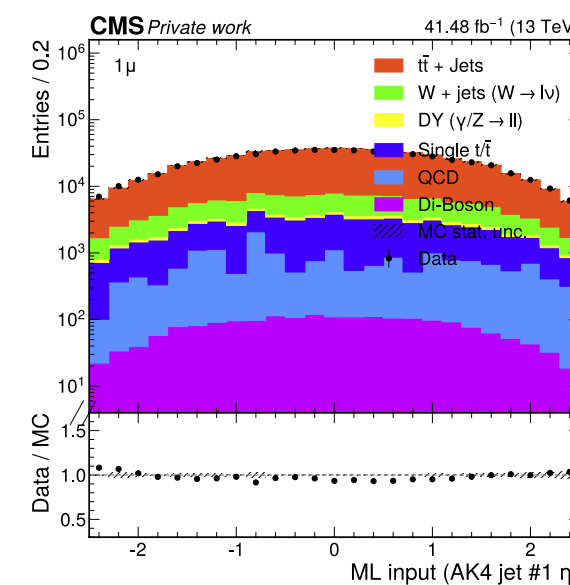
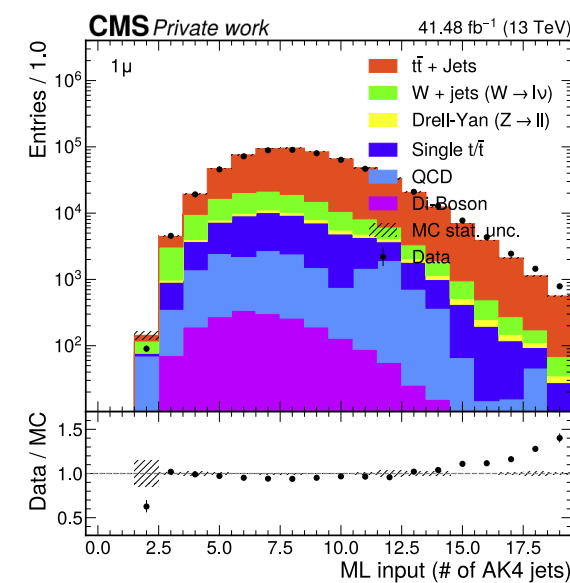
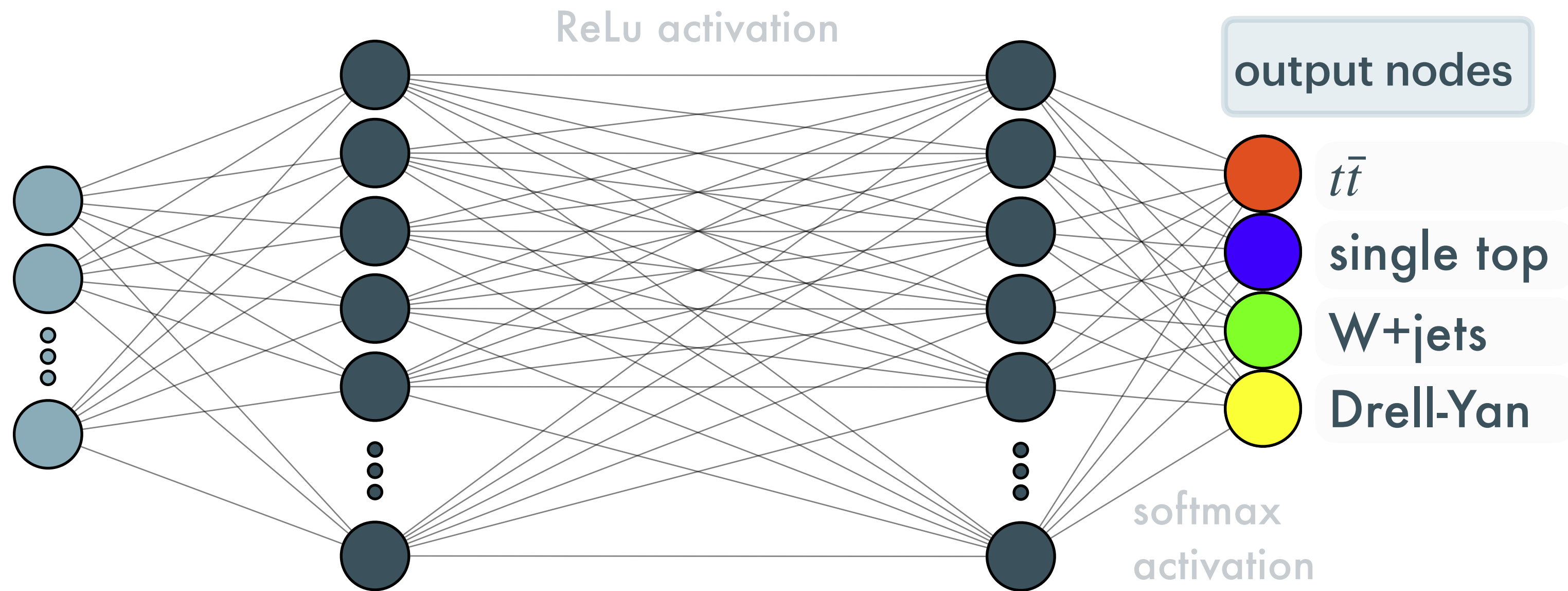
— Text —

DNN for Event Classification

— network architecture —

59 input features:

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

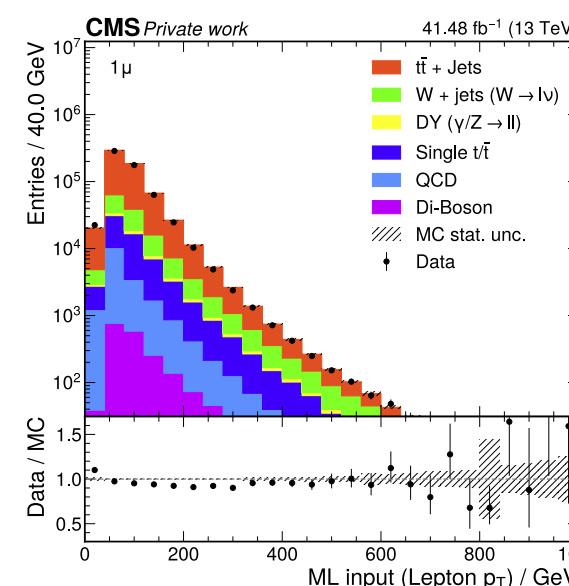
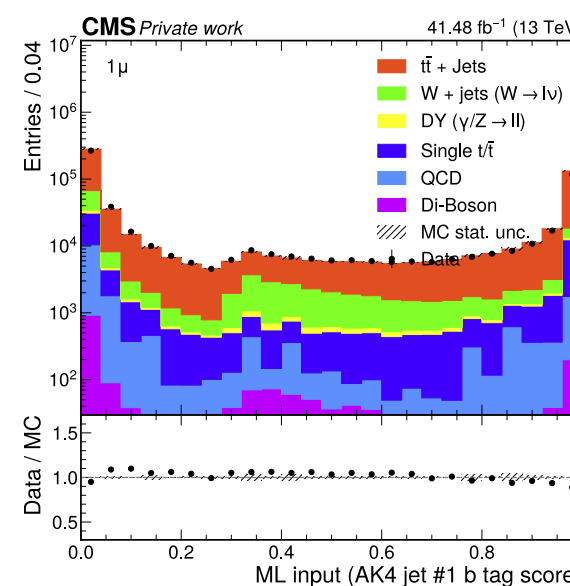
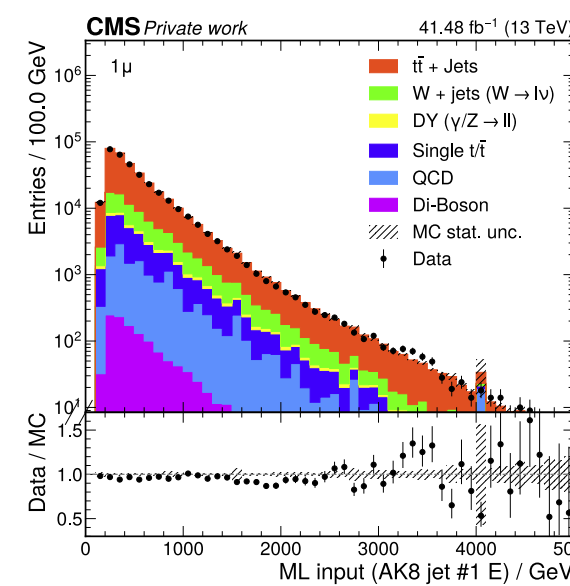
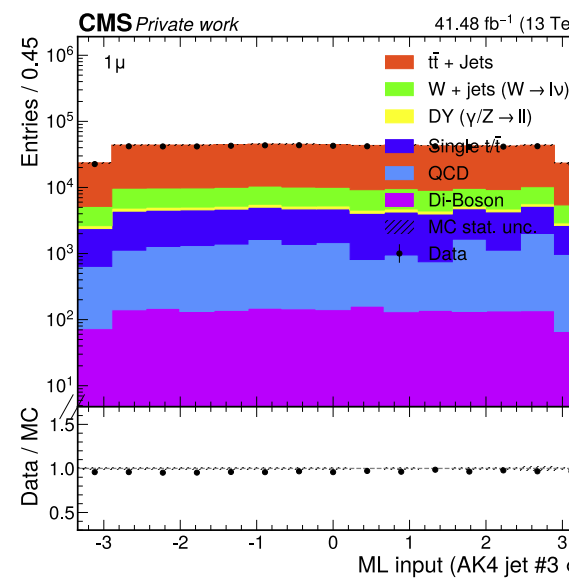
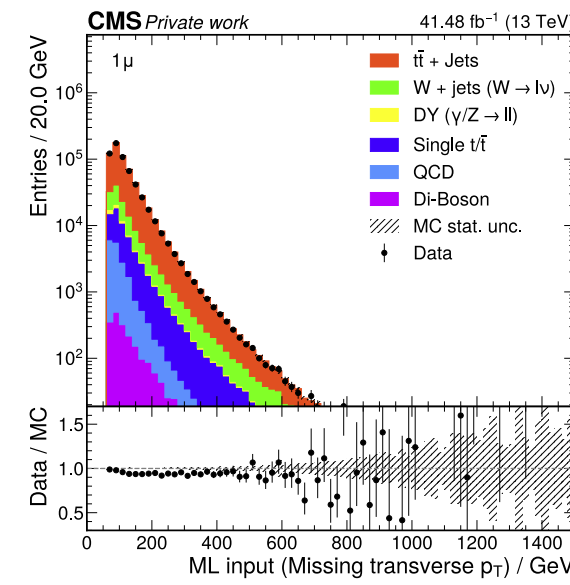
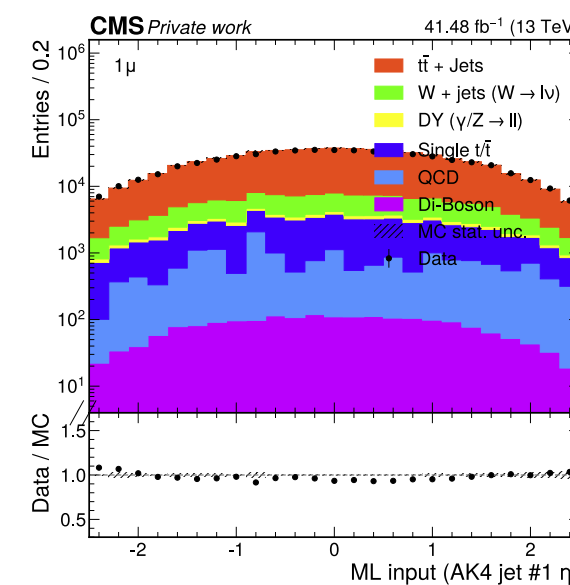
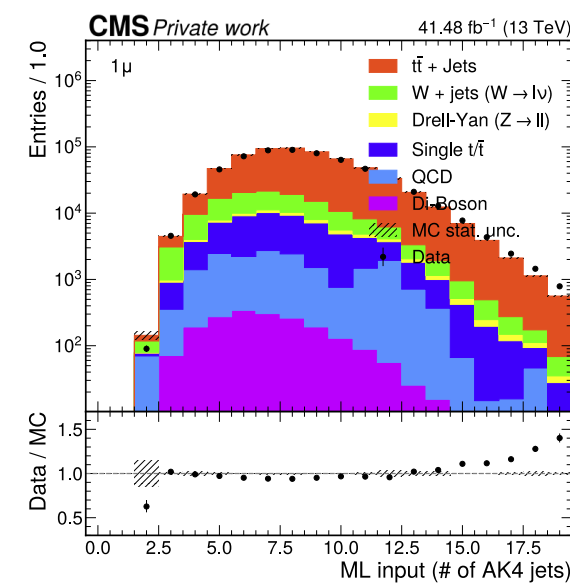
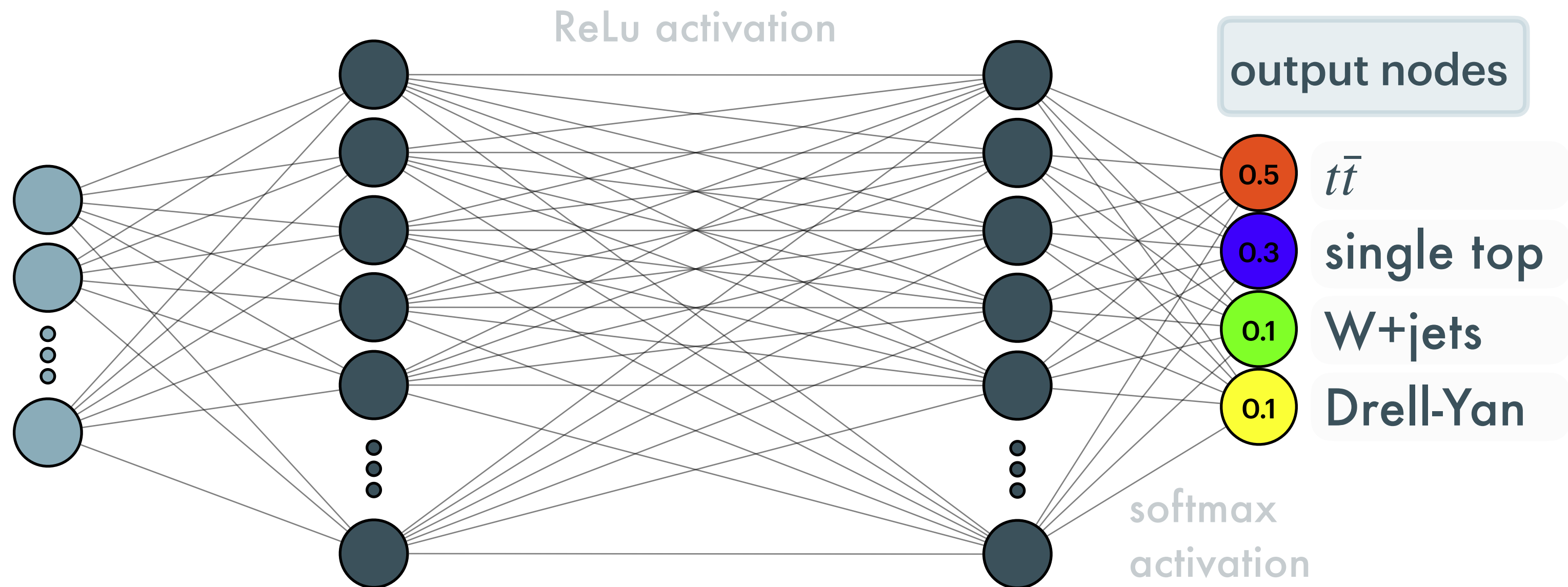


DNN for Event Classification

— 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 ≤ 500
- folds = 5
- learning rate = 0.0005
- validation fraction = 0.25