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

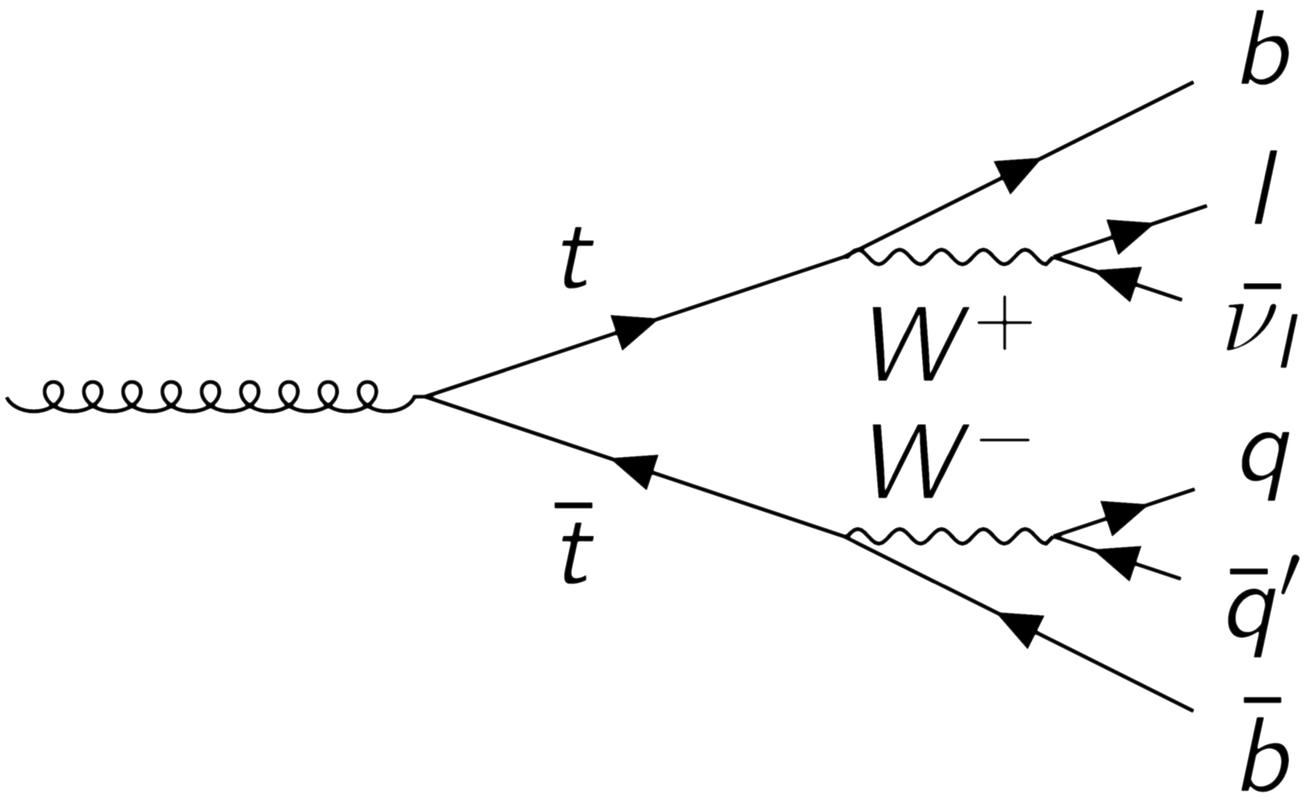
— **Johanna Matthiesen**[1], Johannes Haller[1], Roman Kogler[2], Daniel Savoiu[1] —

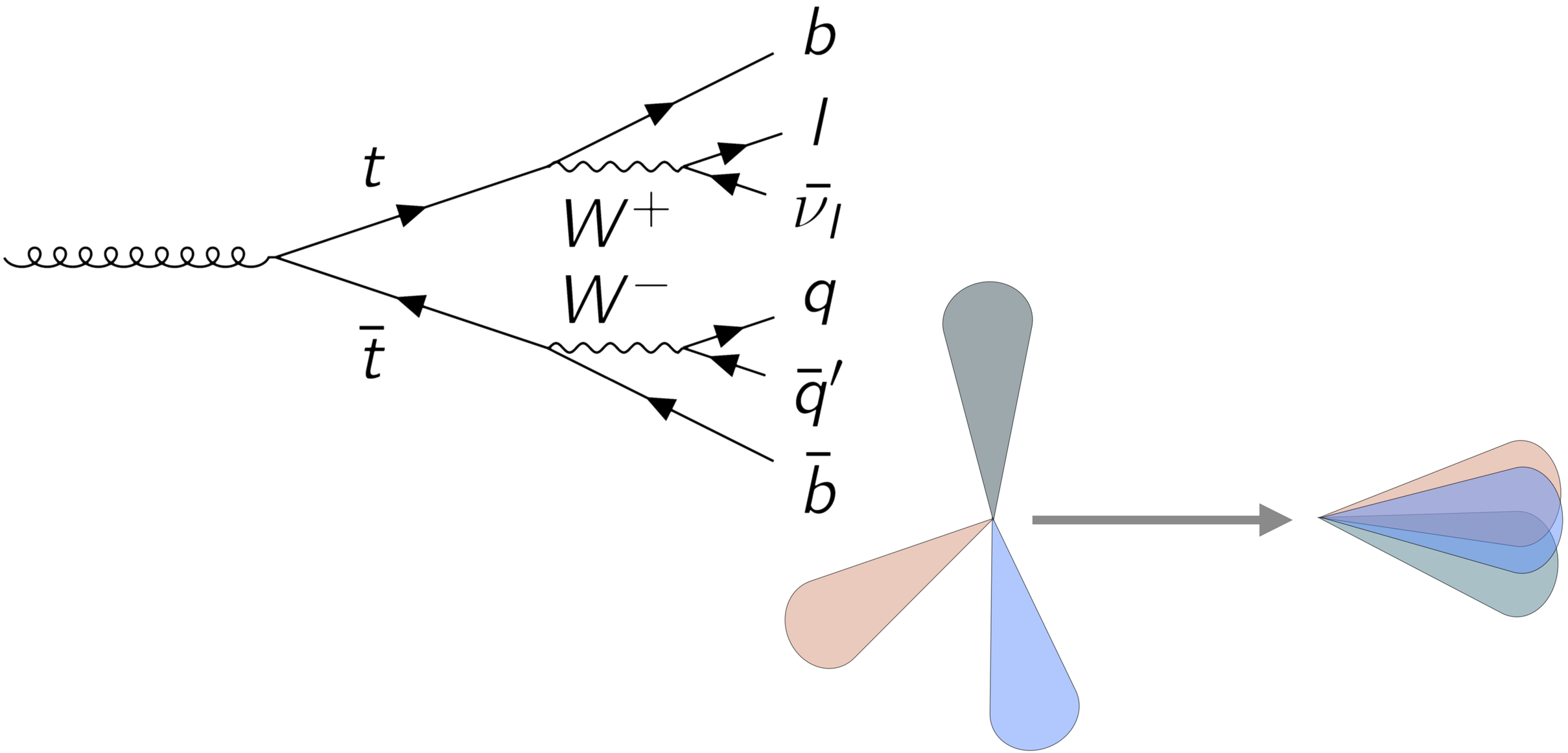
[1]Universität Hamburg - [2]DESY, Hamburg

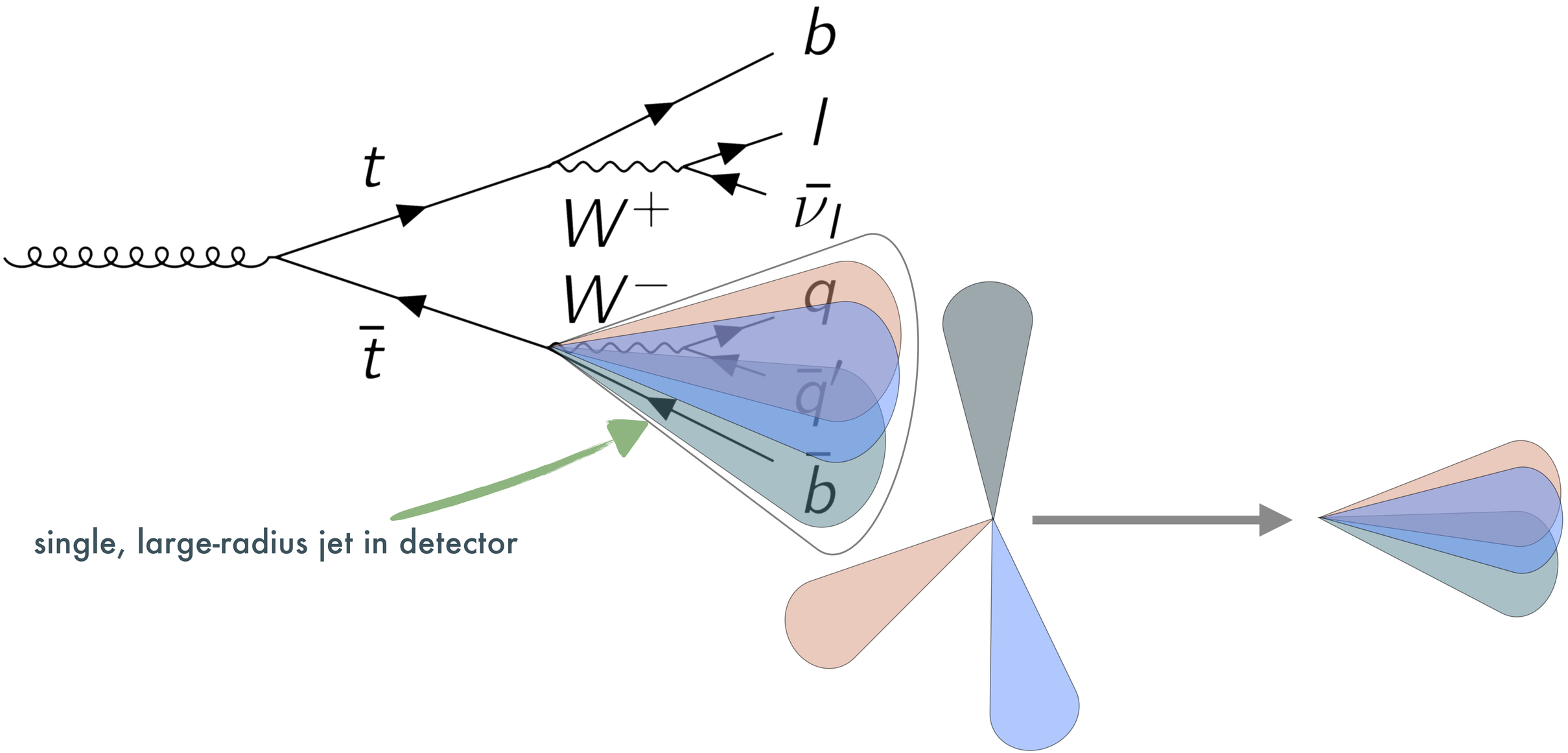
Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

CLUSTER OF EXCELLENCE
QUANTUM UNIVERSE

FSP CMS
Erforschung von Universum und Materie


b
l
t
$\bar{\nu}_l$
W^+
W^-
q
$\bar{t}$
$\bar{q}'$
$\bar{b}$


b
l
t
W^+
$\bar{\nu}_l$
W^-
q
$\bar{t}$
$\bar{q}'$
$\bar{b}$


b
l
t
$\bar{\nu}_l$
W^+
W^-
q
$\bar{q}'$
$\bar{t}$
$\bar{b}$
single, large-radius jet in detector

Identifying Boosted Top Quarks

— in CMS—

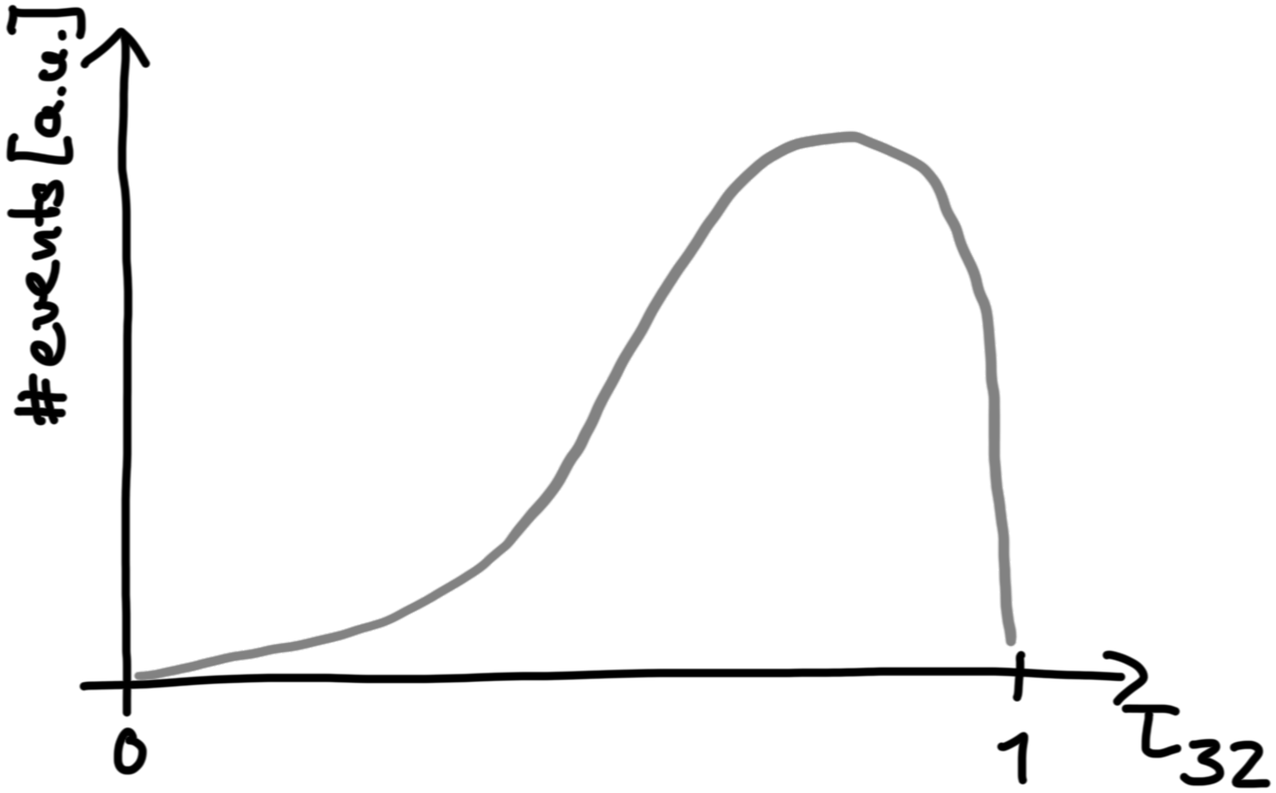
- different approaches to top tagging:
  - **machine learning** algorithms
    DeepAK8, ParticleNet, ...
  - **cut based** using jet substructure techniques
    N-subjettiness, soft drop mass

Identifying Boosted Top Quarks

— in CMS—

- different approaches to top tagging:
  - **machine learning** algorithms
    DeepAK8, ParticleNet, ...
  - **cut based** using jet substructure techniques
    N-subjettiness, soft drop mass

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

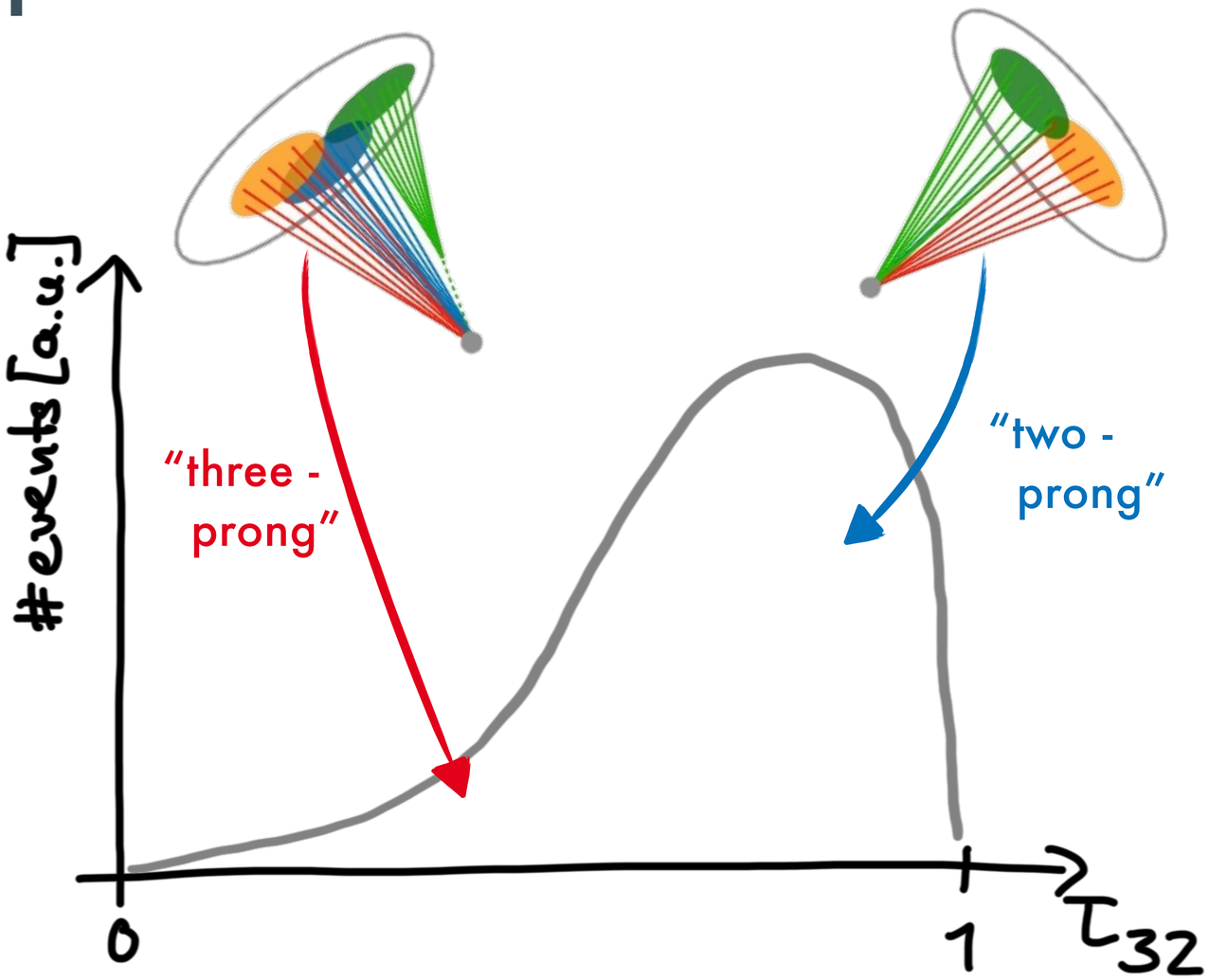


Identifying Boosted Top Quarks

— in CMS—

- different approaches to top tagging:
  - **machine learning** algorithms
    DeepAK8, ParticleNet, ...
  - **cut based** using jet substructure techniques
    N-subjettiness, soft drop mass

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

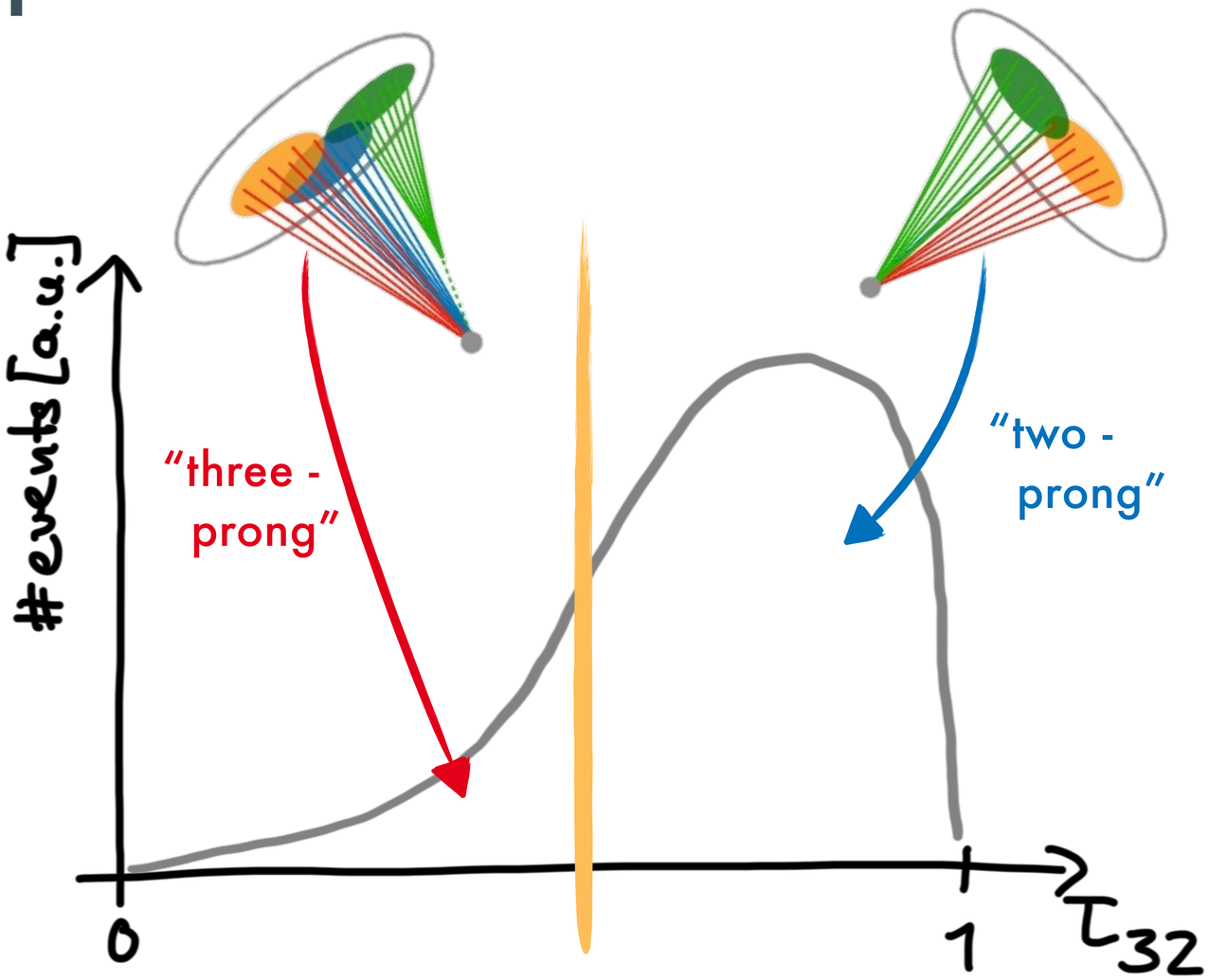


Identifying Boosted Top Quarks

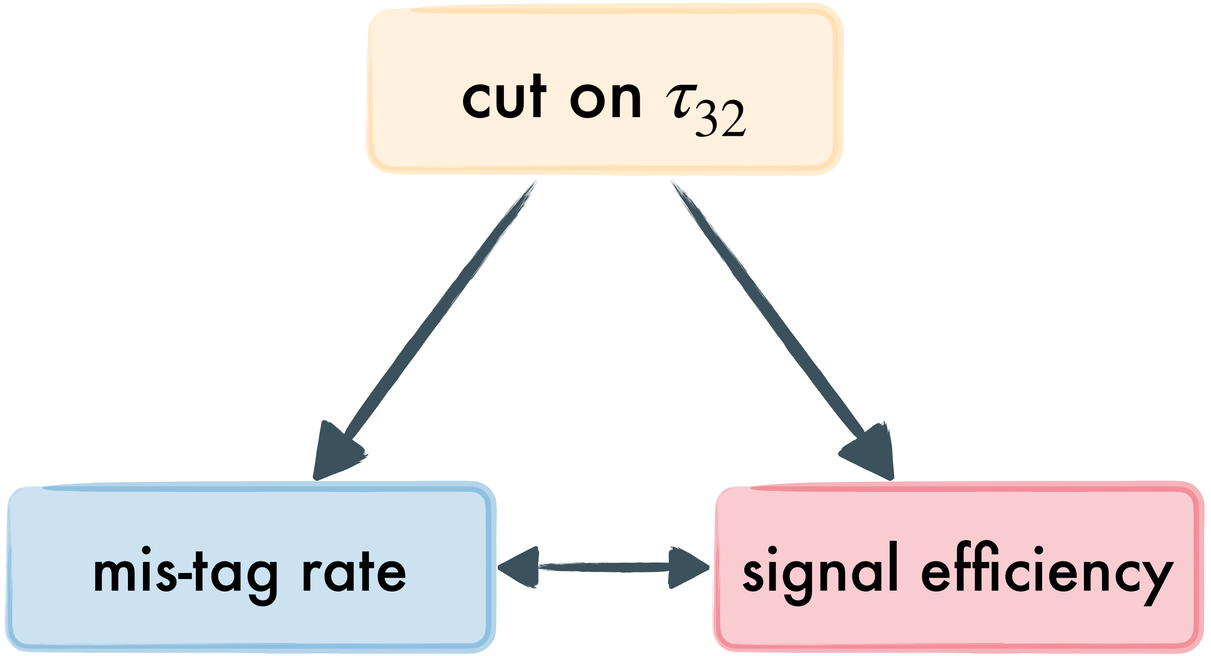
— in CMS—

- different approaches to top tagging:
  - **machine learning** algorithms
    DeepAK8, ParticleNet, ...
  - **cut based** using jet substructure techniques
    N-subjettiness, soft drop mass

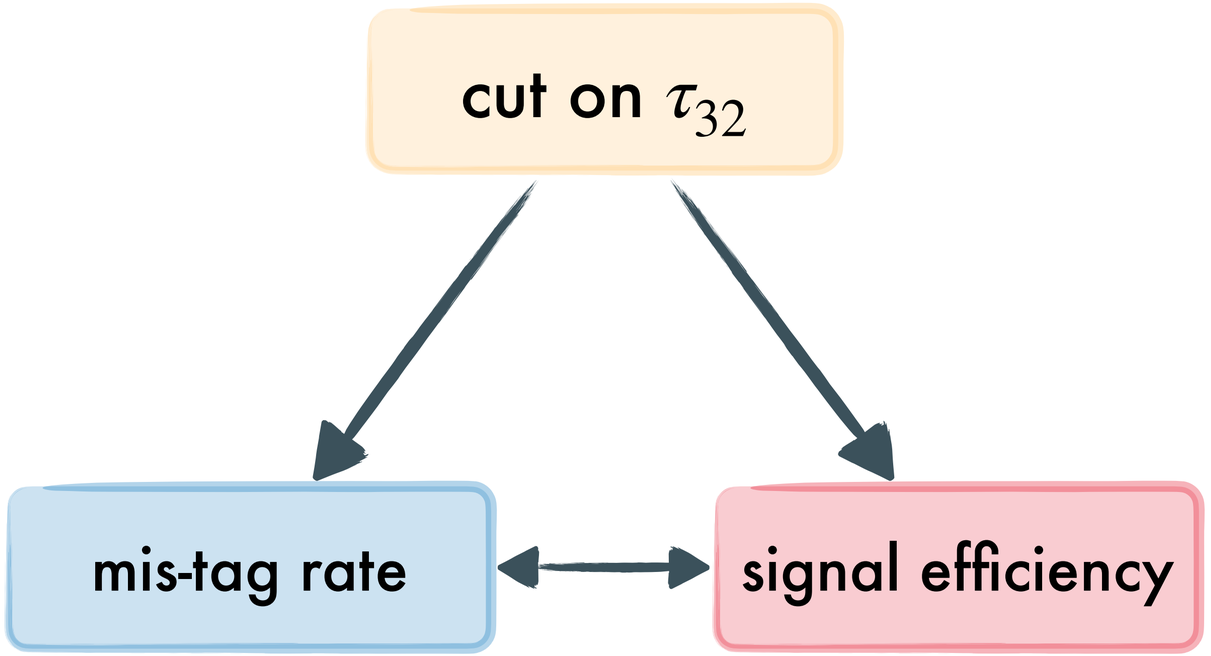
$$\tau_{32} := \tau_3/\tau_2$$

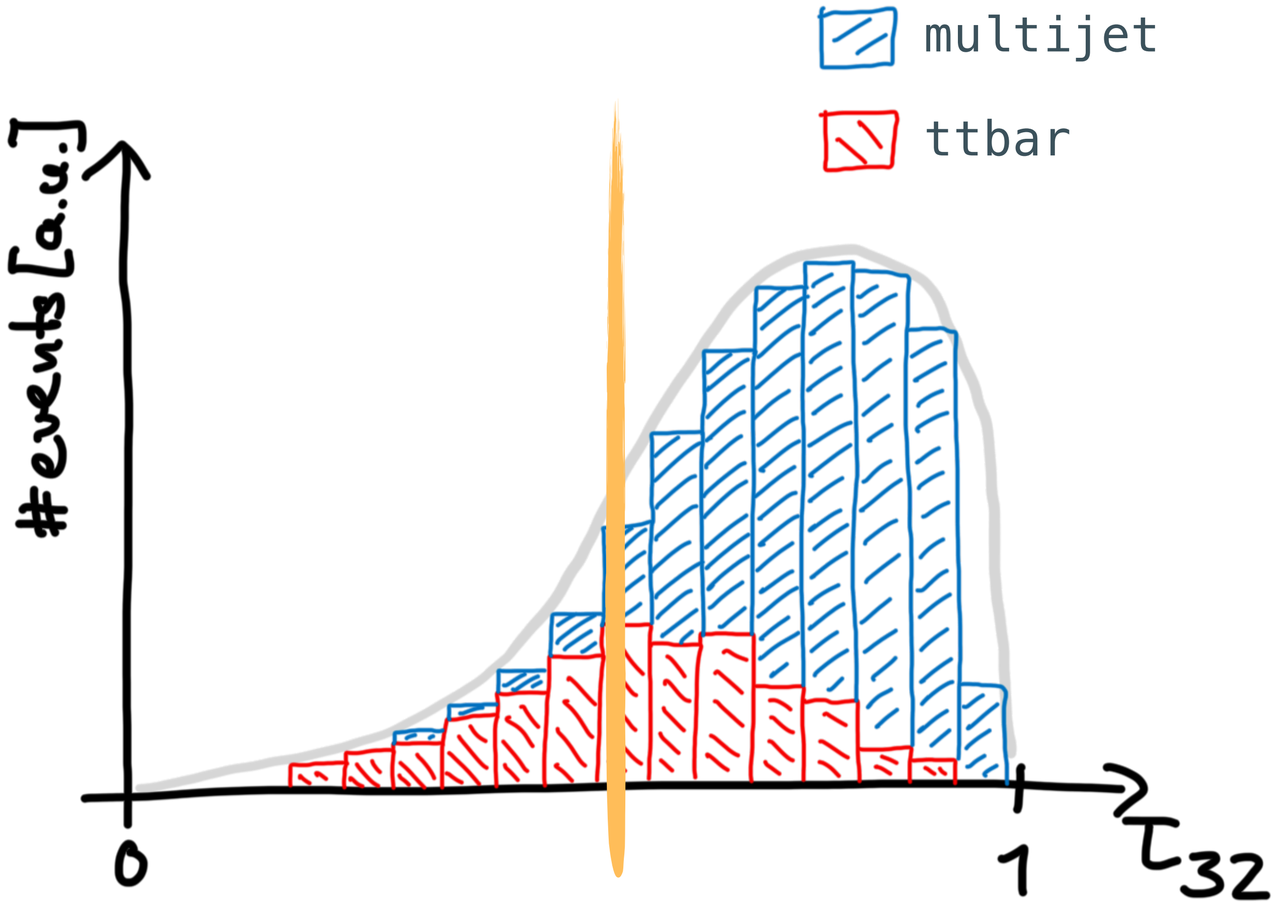


Classifying the Cut on τ_{32}

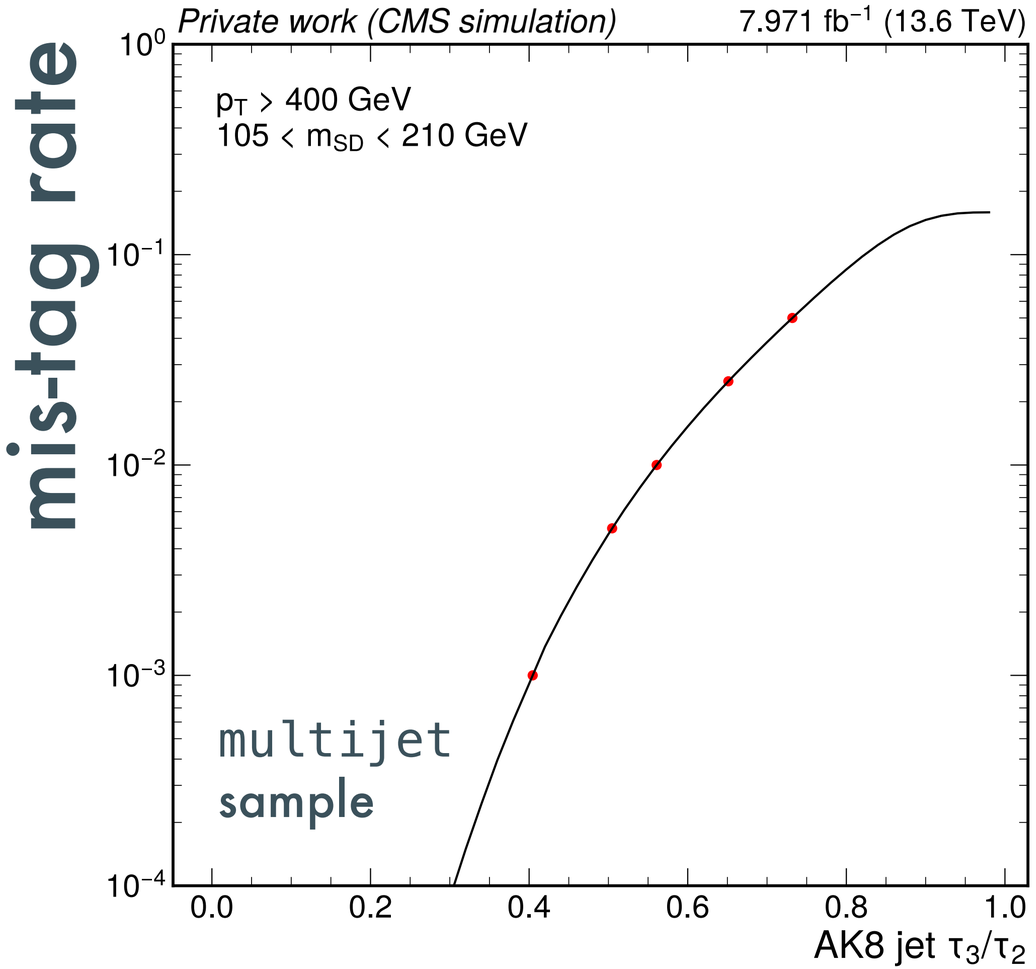


Classifying the Cut on τ_{32}

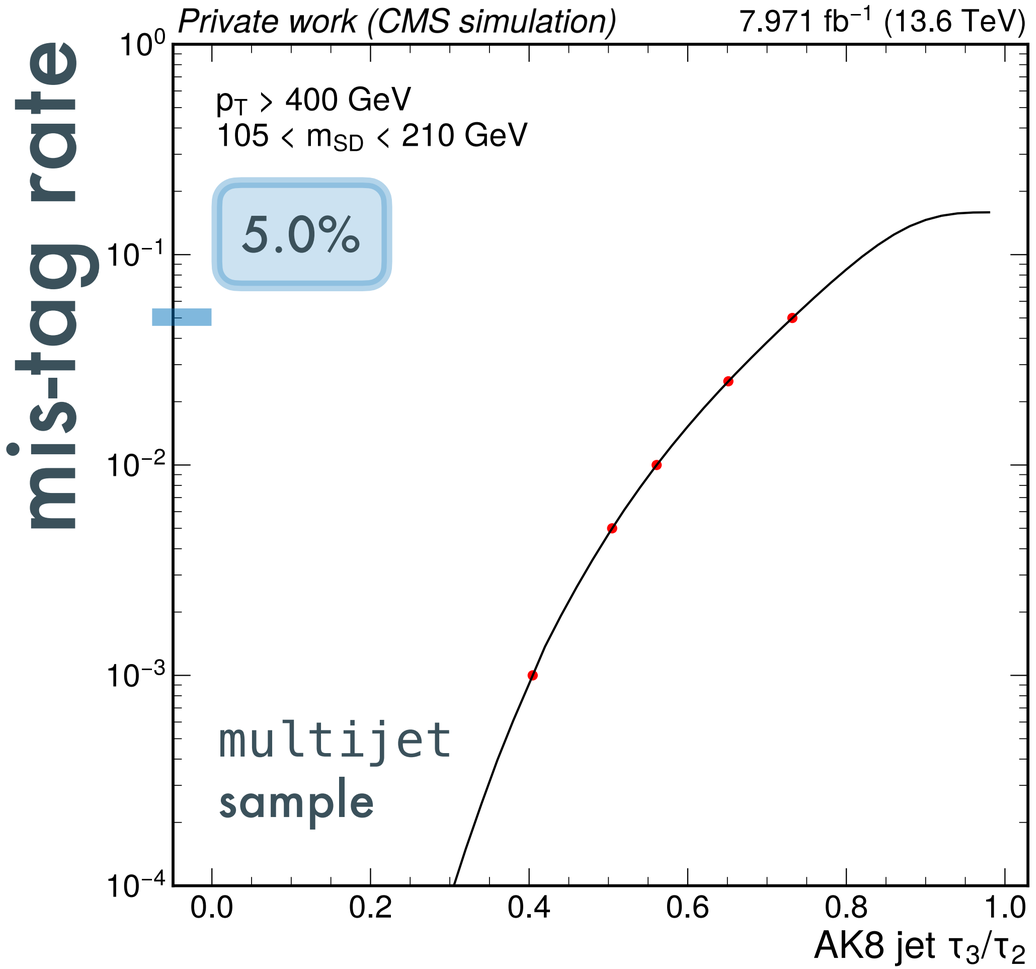




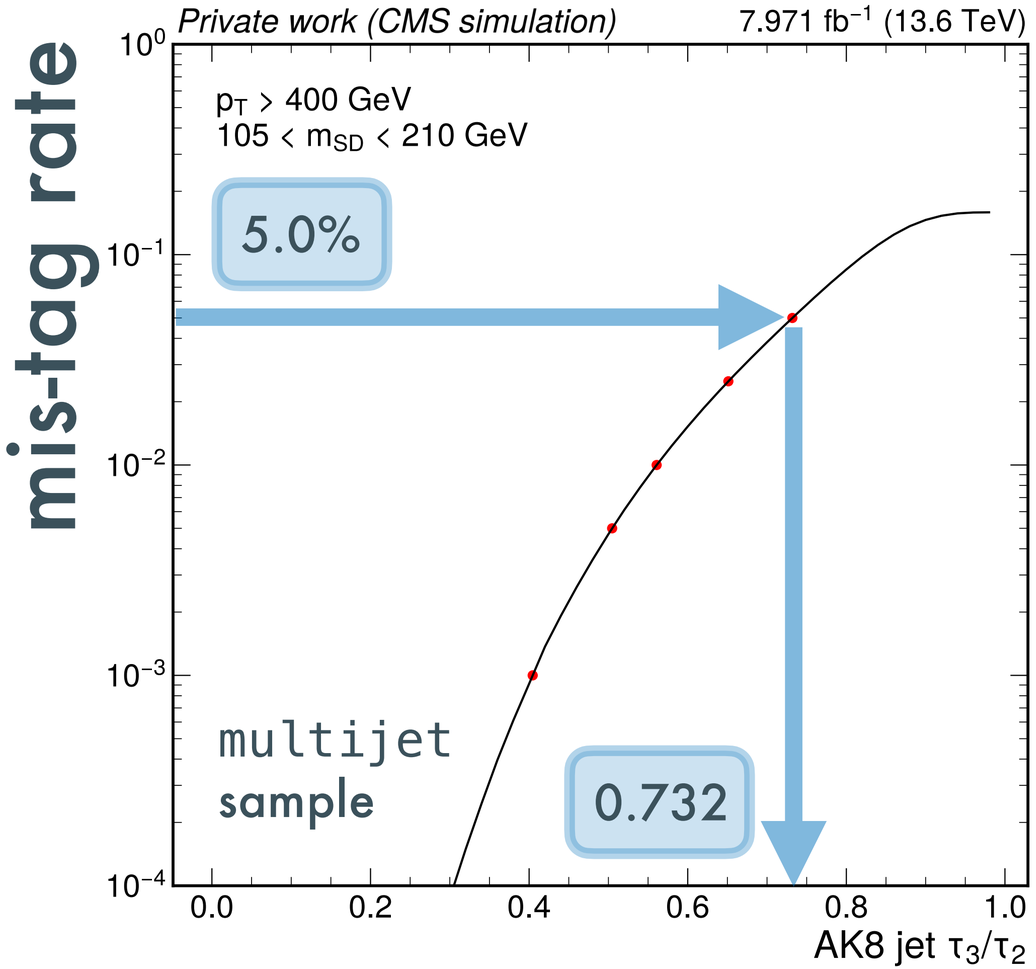
Working Points in One Slide



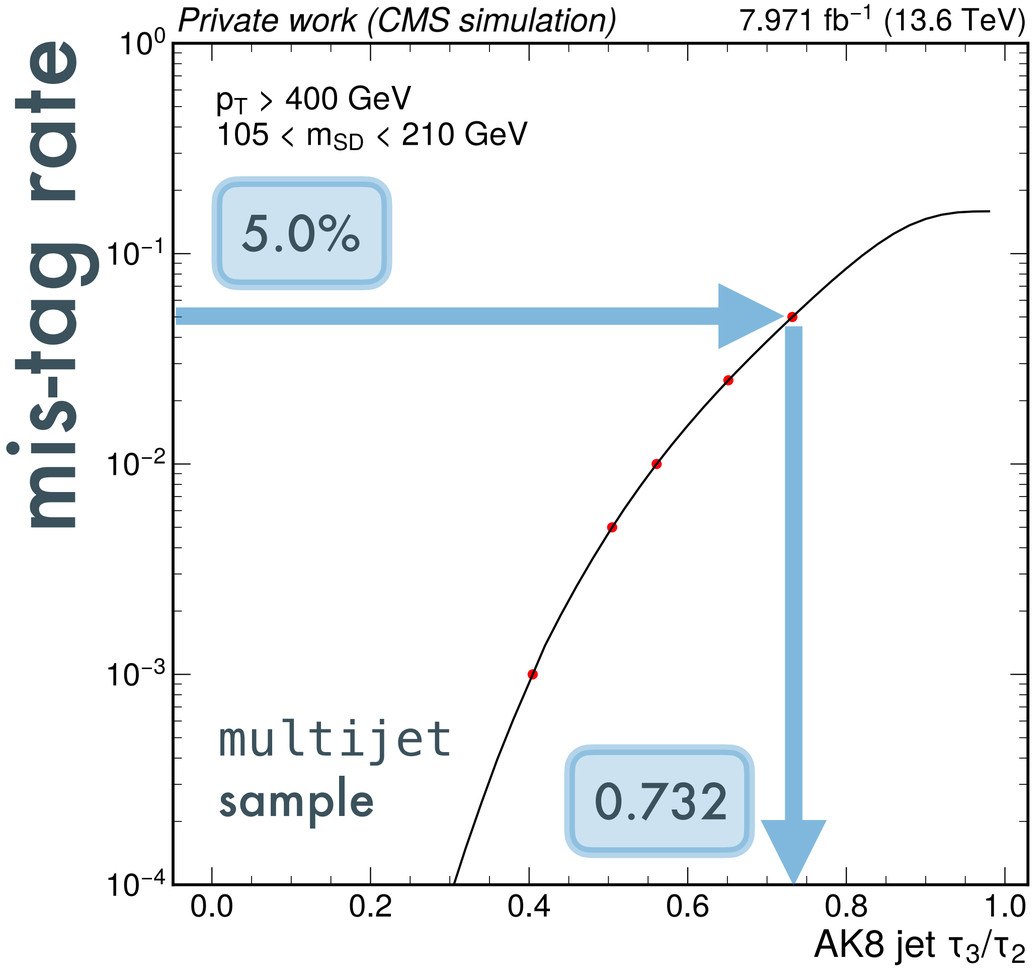
Working Points in One Slide

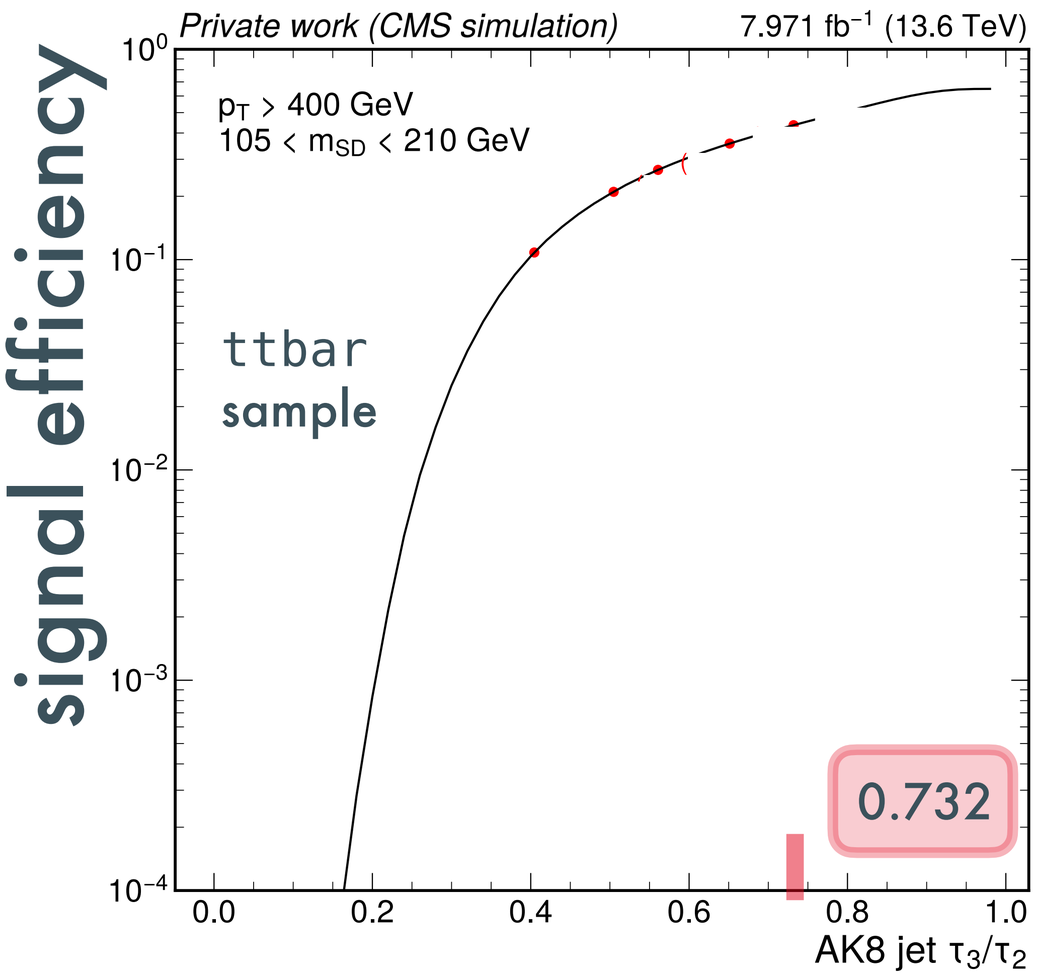


Working Points in One Slide

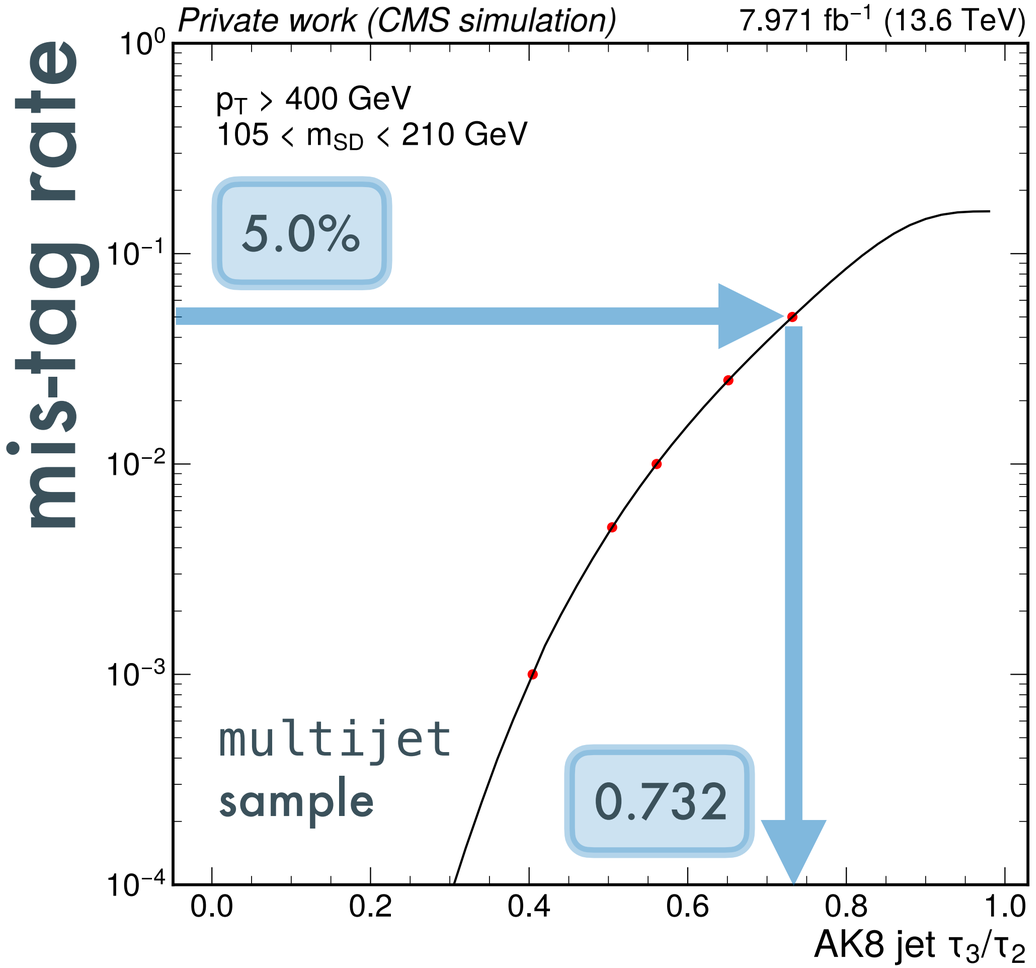


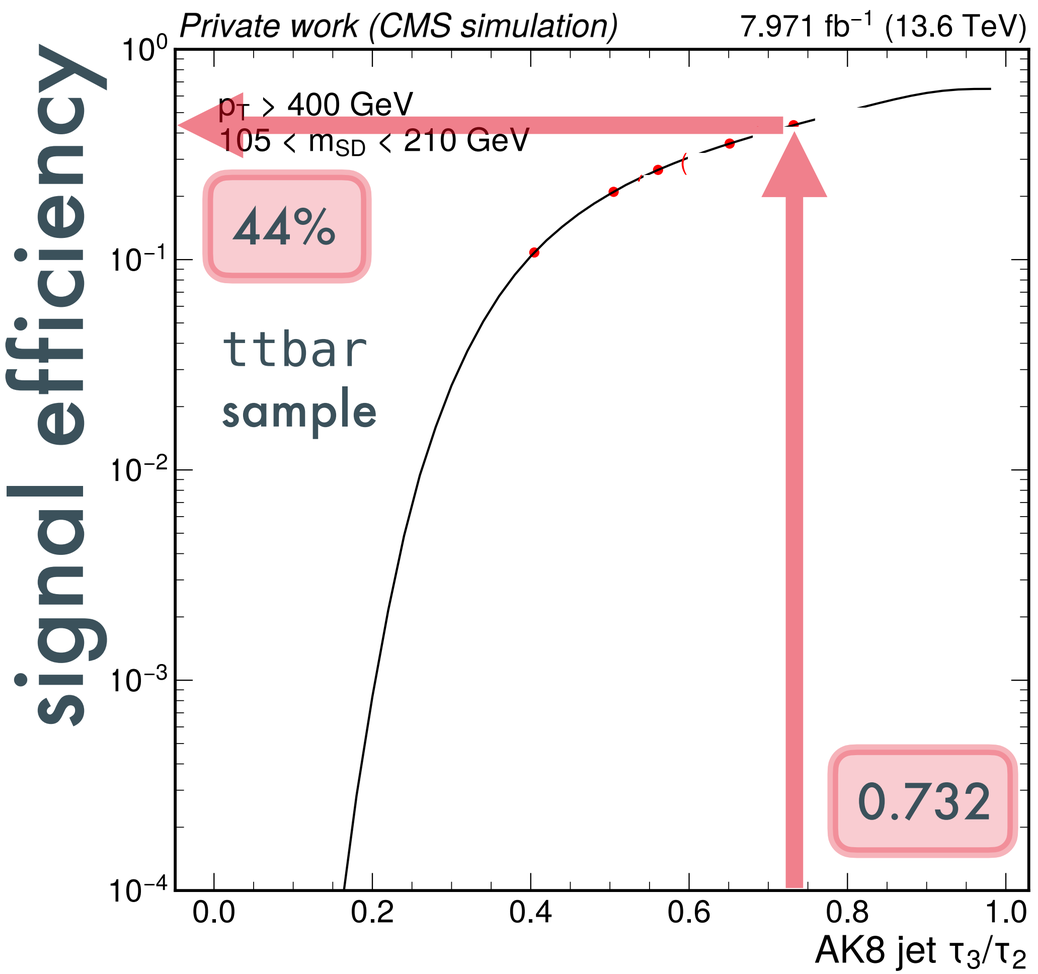
Working Points in One Slide



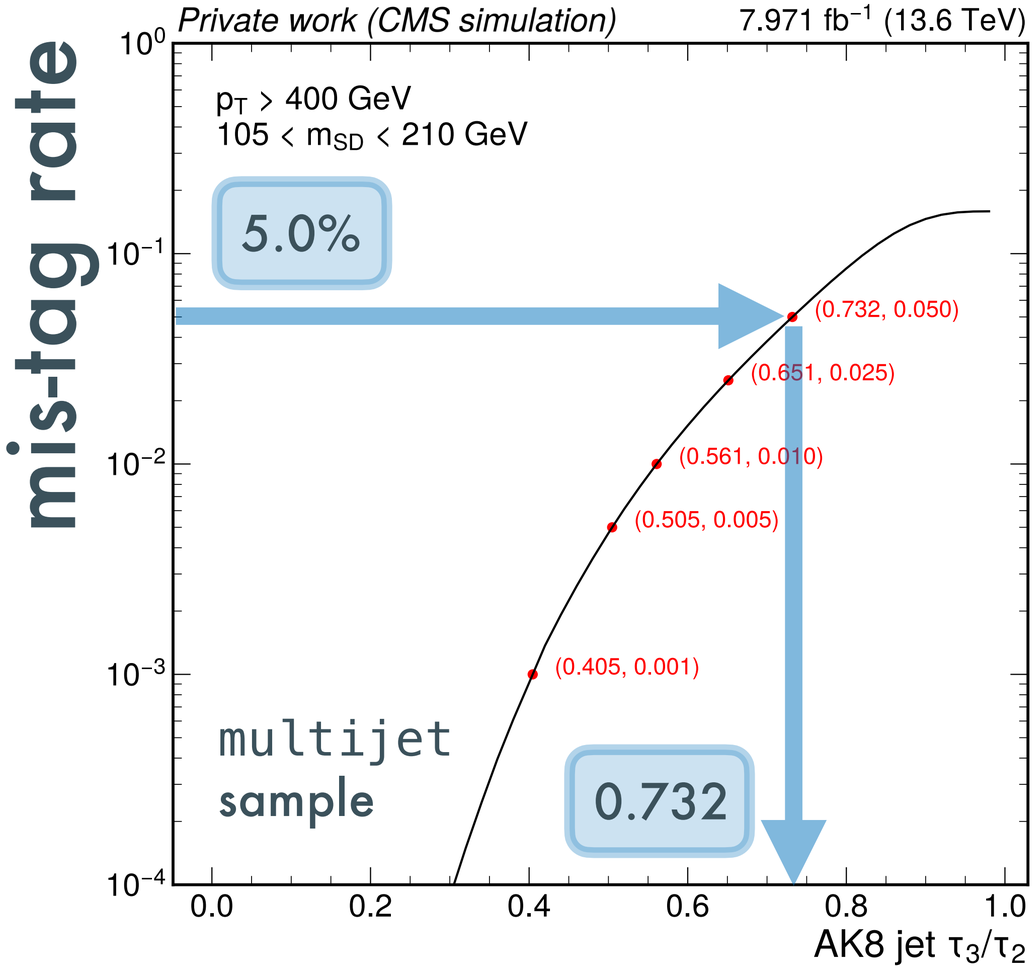


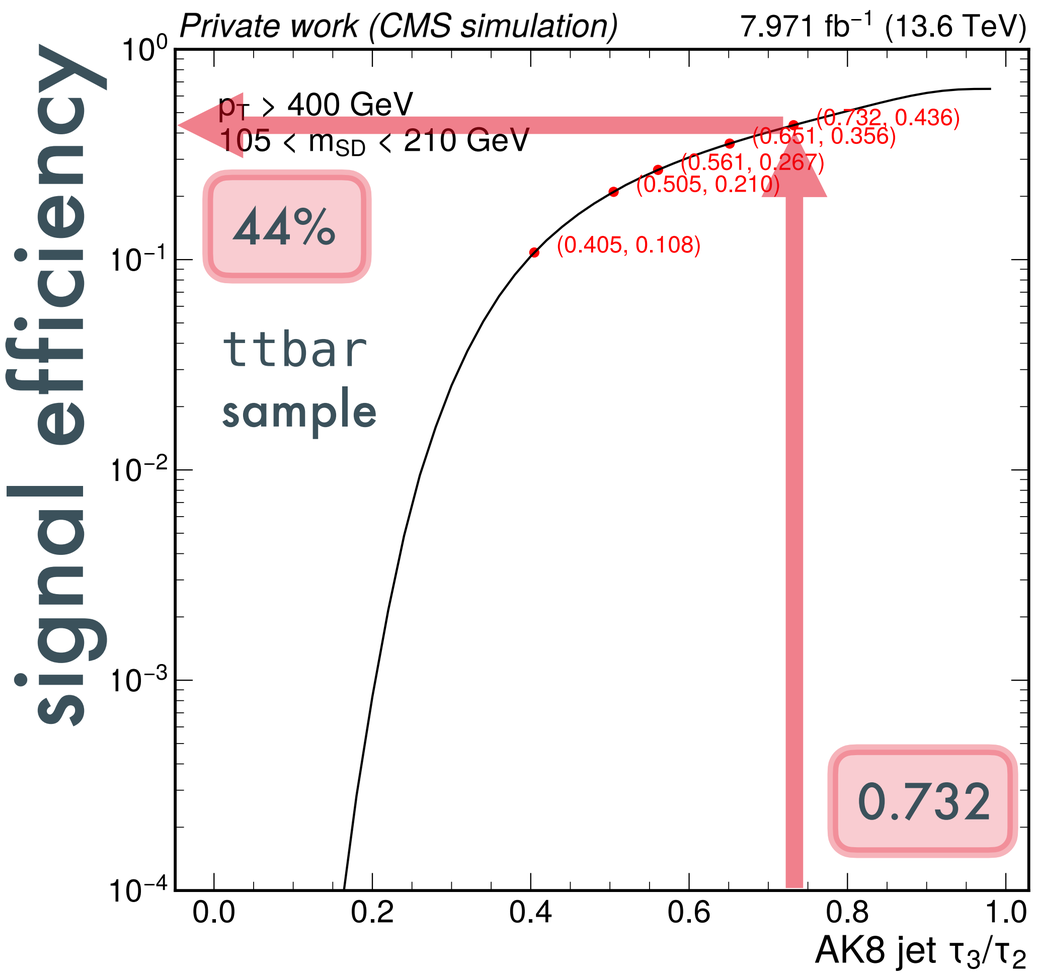
Working Points in One Slide



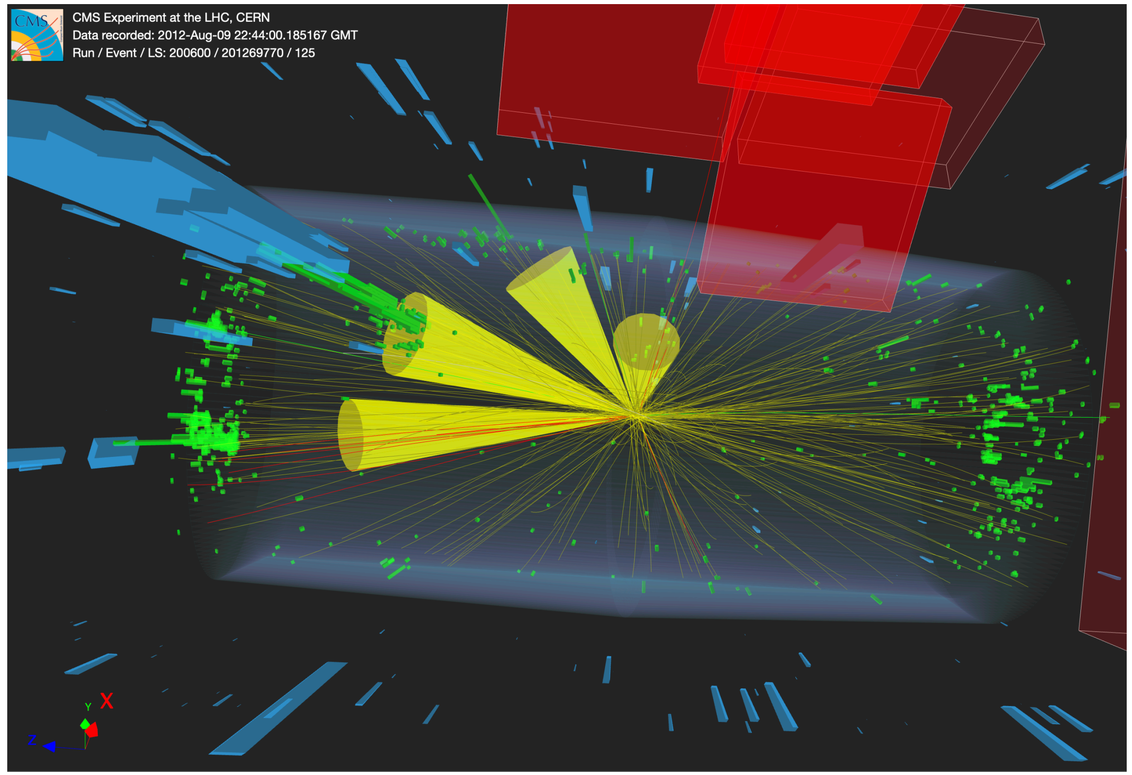


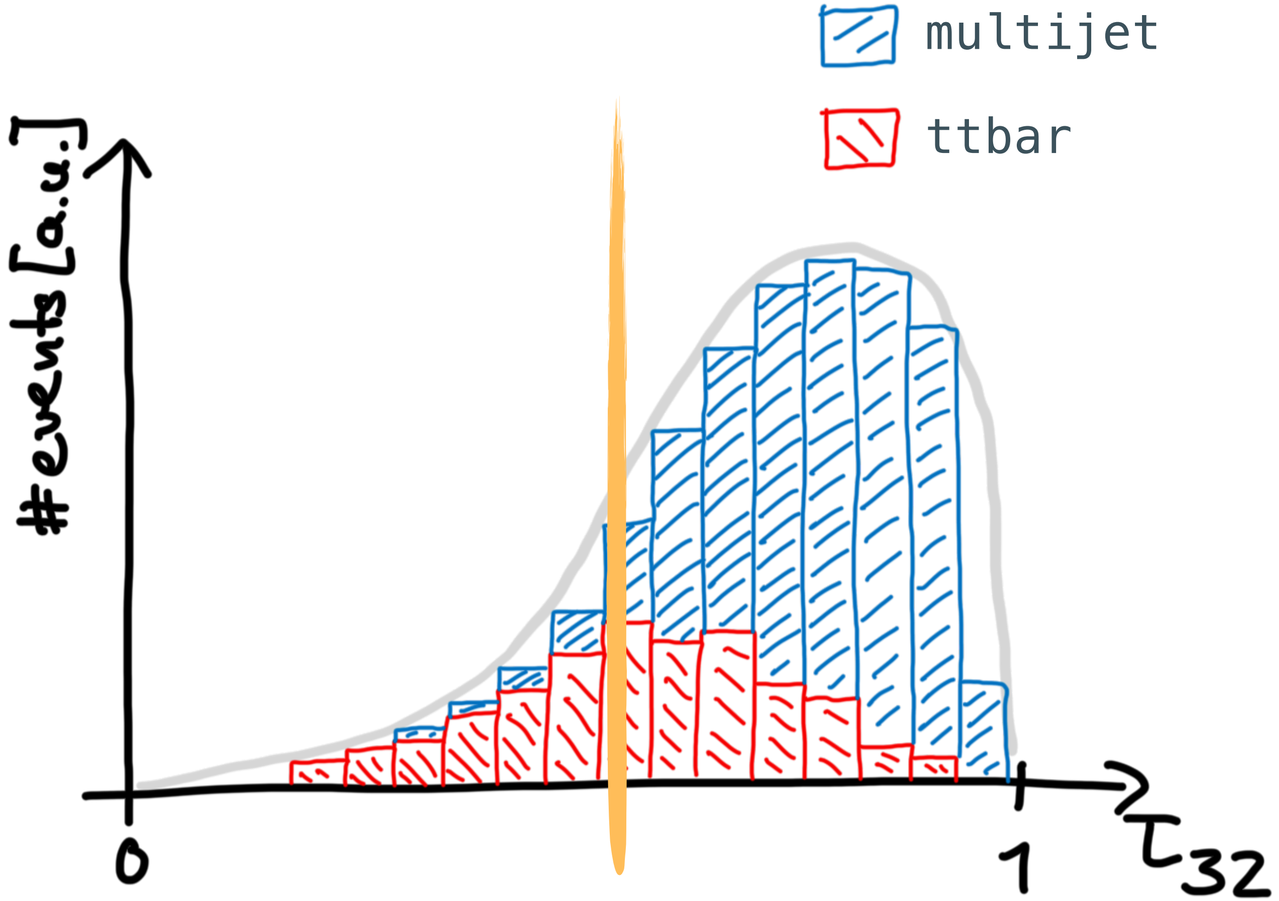
Working Points in One Slide



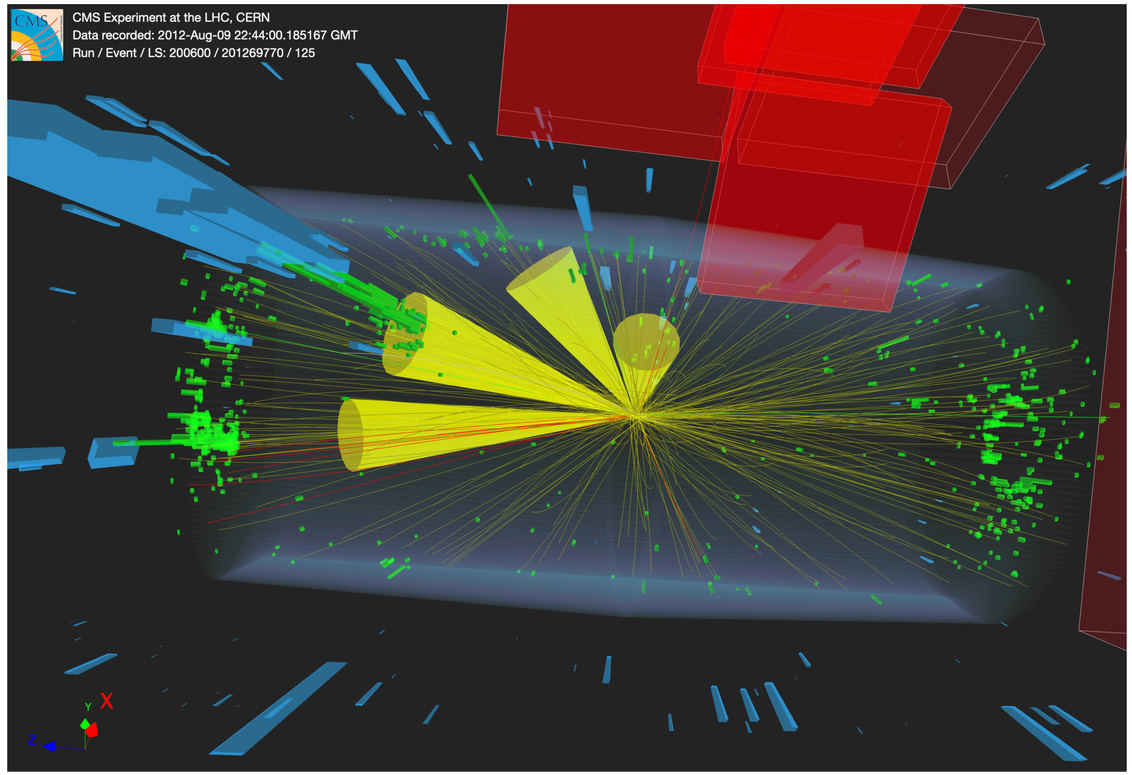


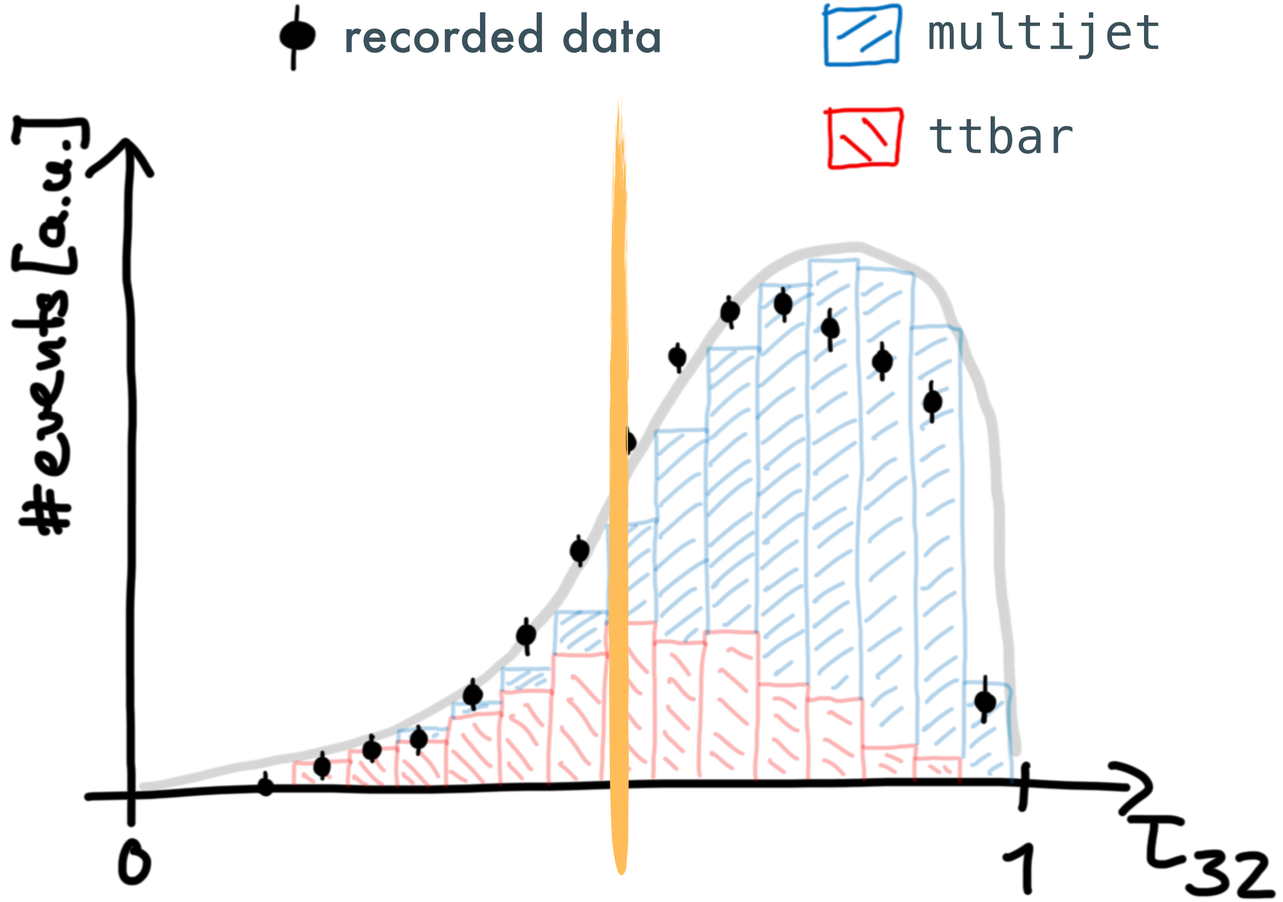
Jet Substructure in Recorded Data





Jet Substructure in Recorded Data

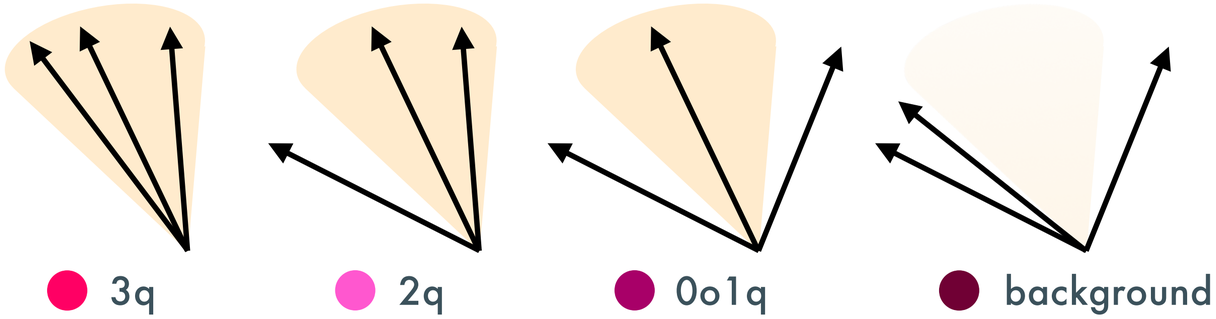


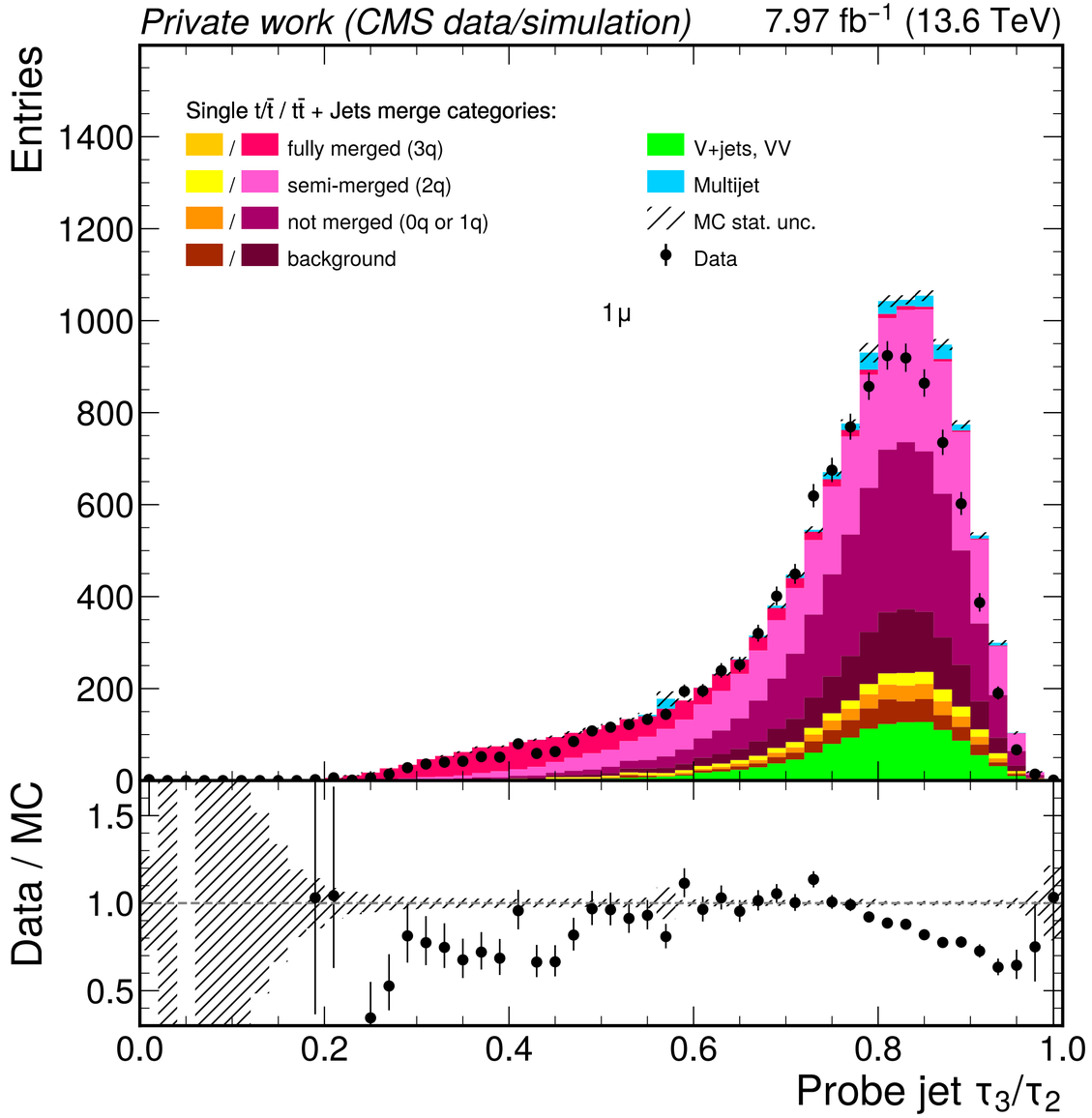


Scale Factors

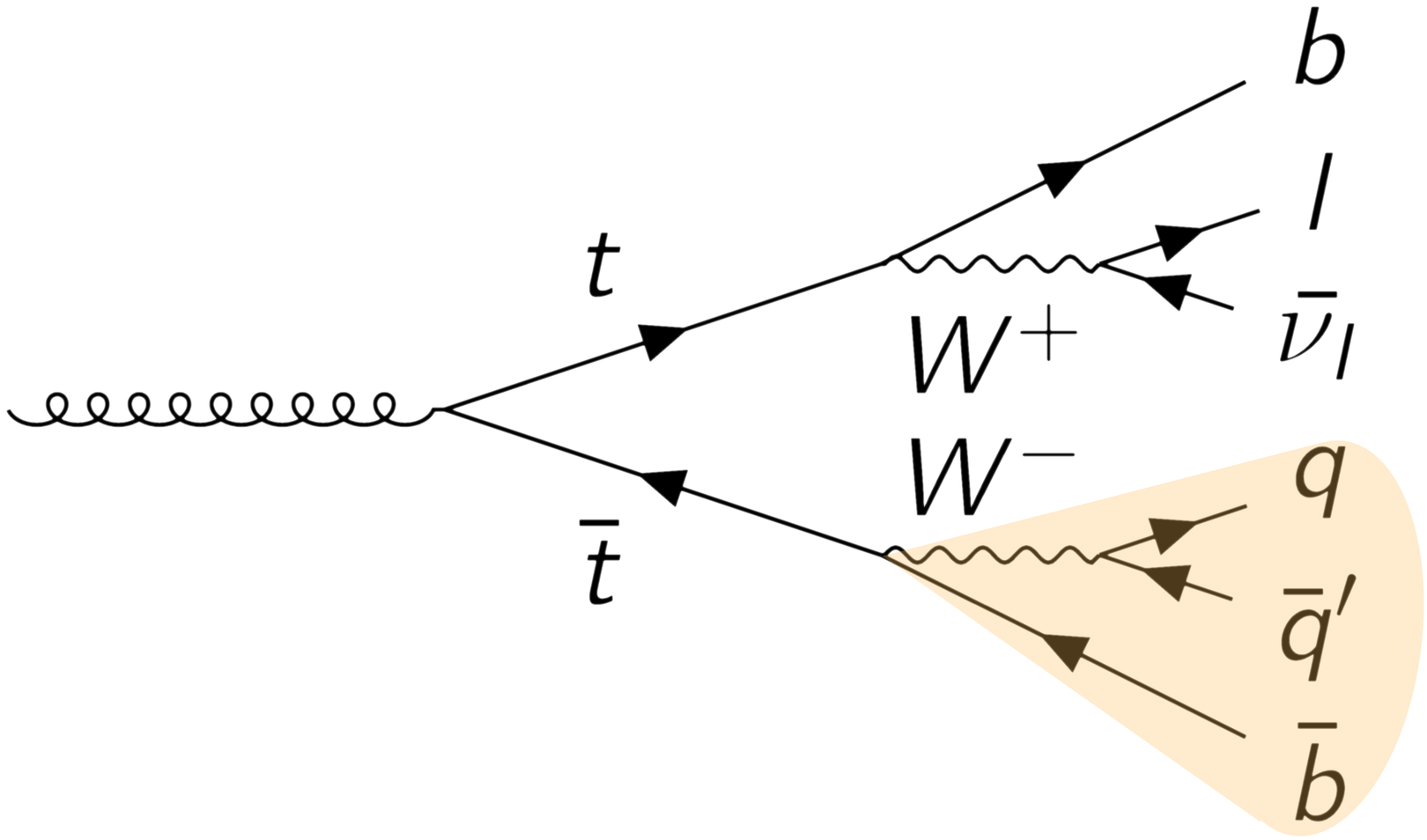
— to correct MC simulation —

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

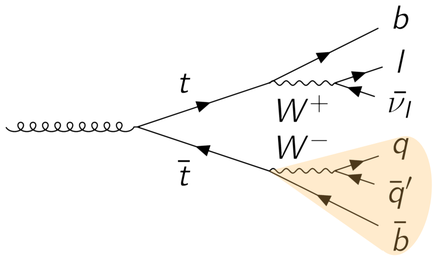




Tag and Probe Method With $t\bar{t}$ Pairs



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



single lepton
electron or muon,
$p_T > 55$ GeV,
isolated

large
radius jet
$p_T > 300$ GeV

b tagged small
radius jet
$p_T > 15$ GeV

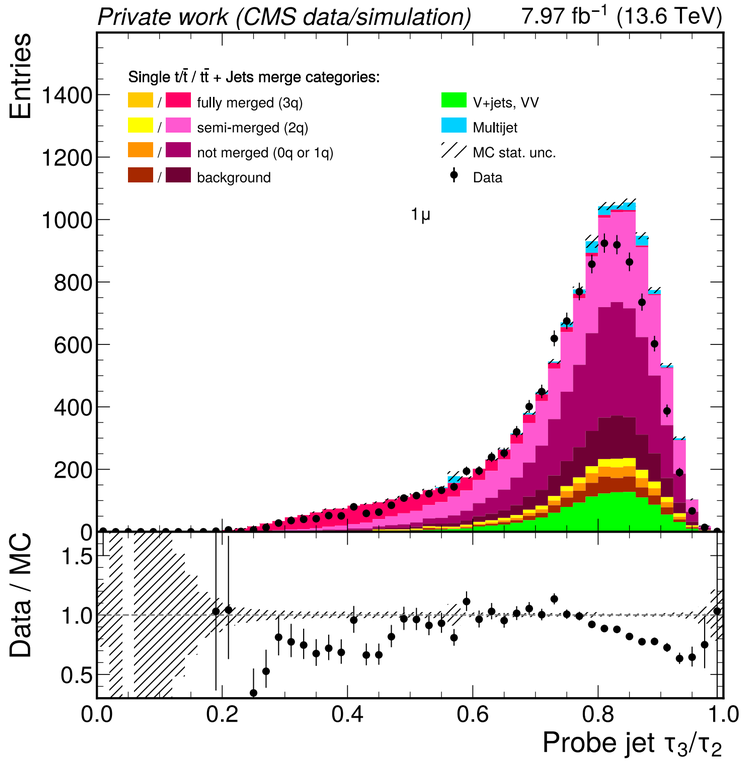
$\Delta R > \frac{2}{3}\pi$

$p_T^{miss} > 50$ GeV
(neutrino)

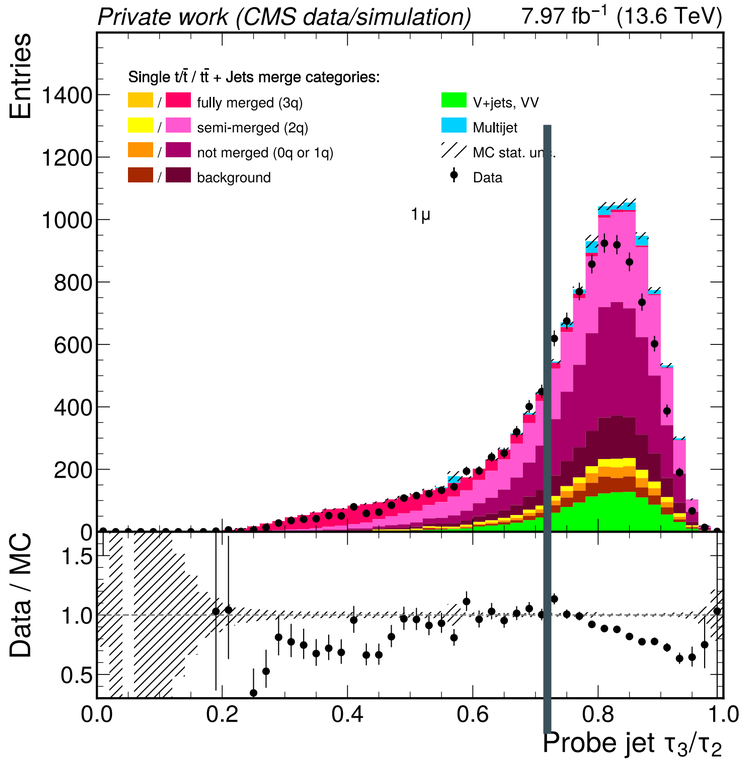
Tag top

Probe top

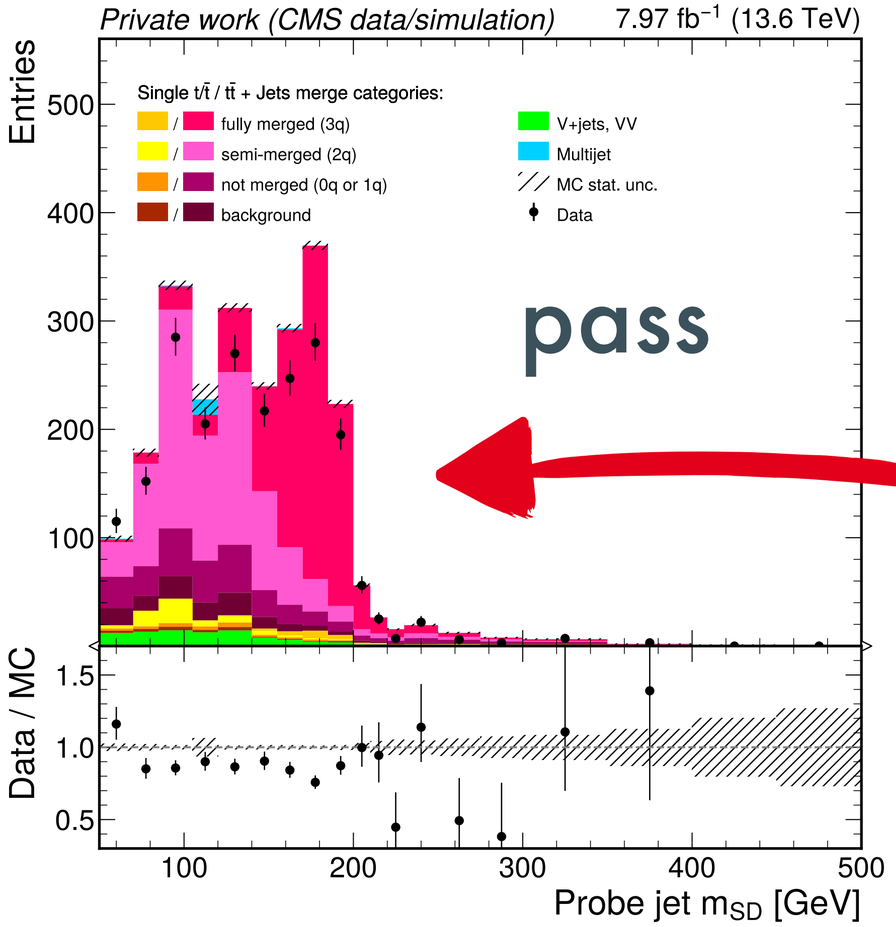
τ_{32} and m_{SD} of Probe Jet

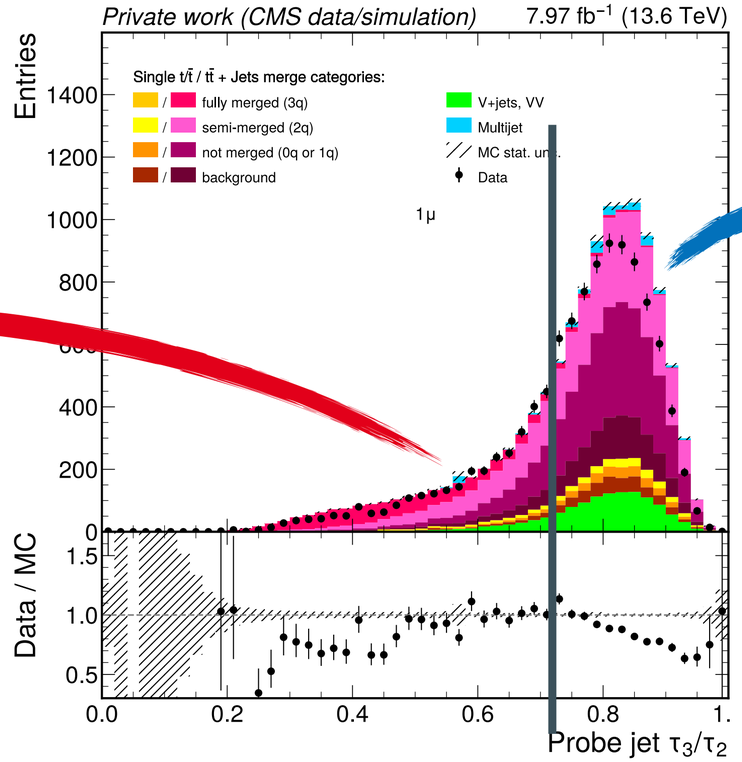


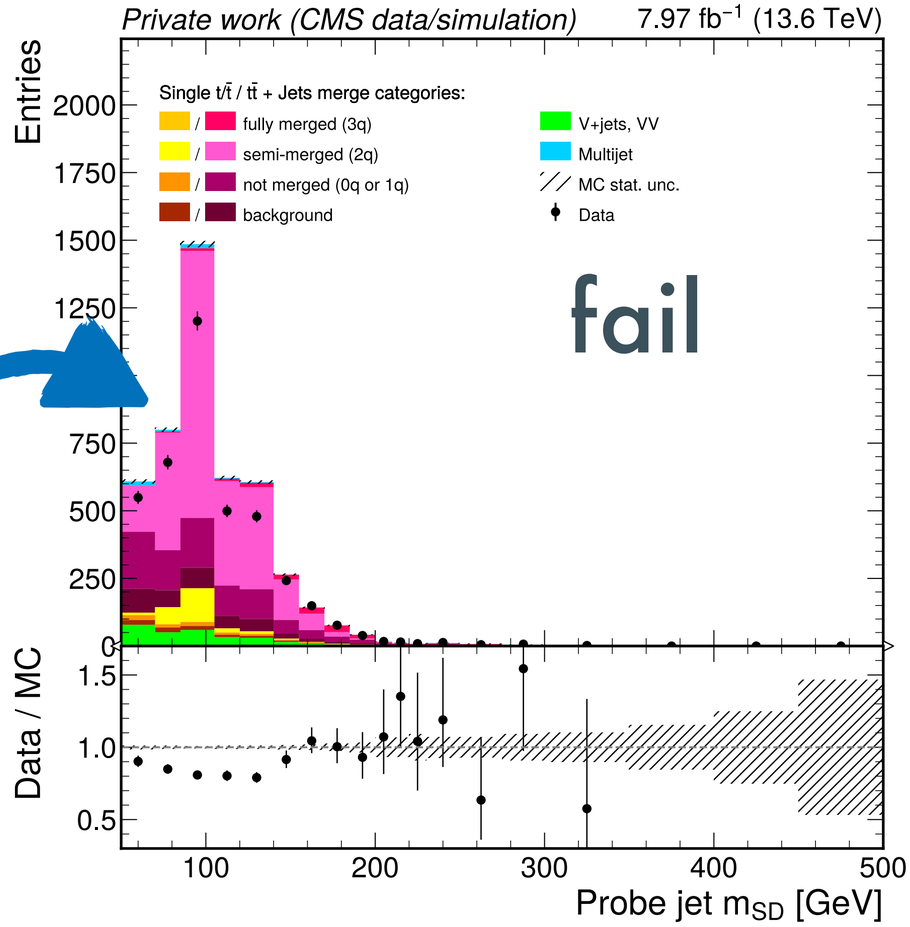
τ_{32} and m_{SD} of Probe Jet



τ_{32} and m_{SD} of Probe Jet



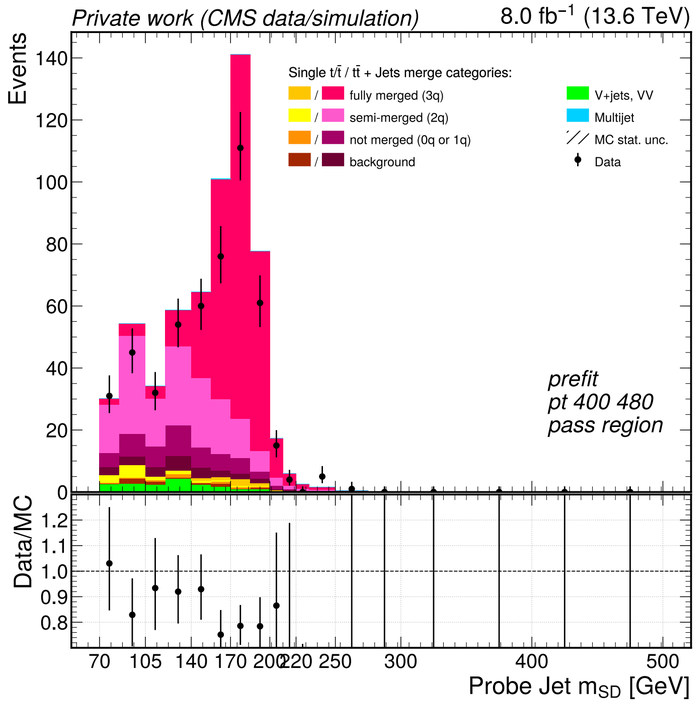


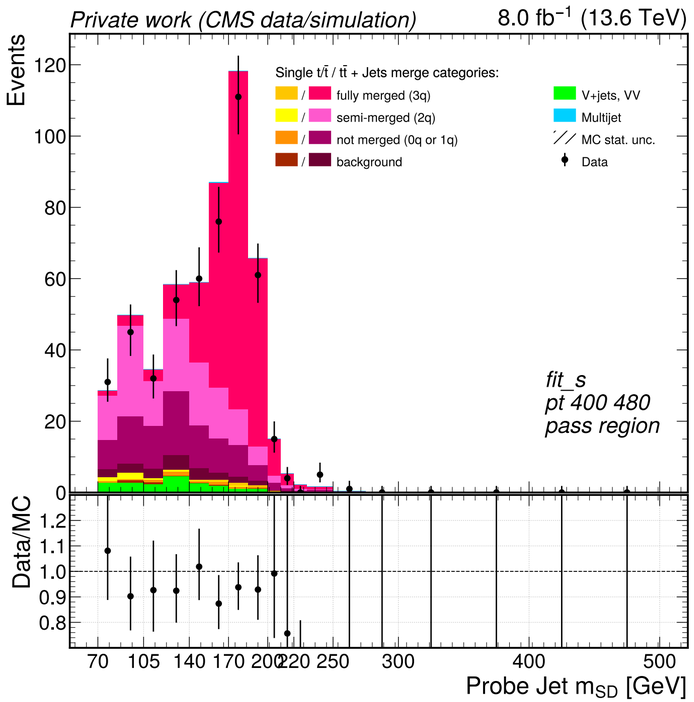


SF Derivation

— 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

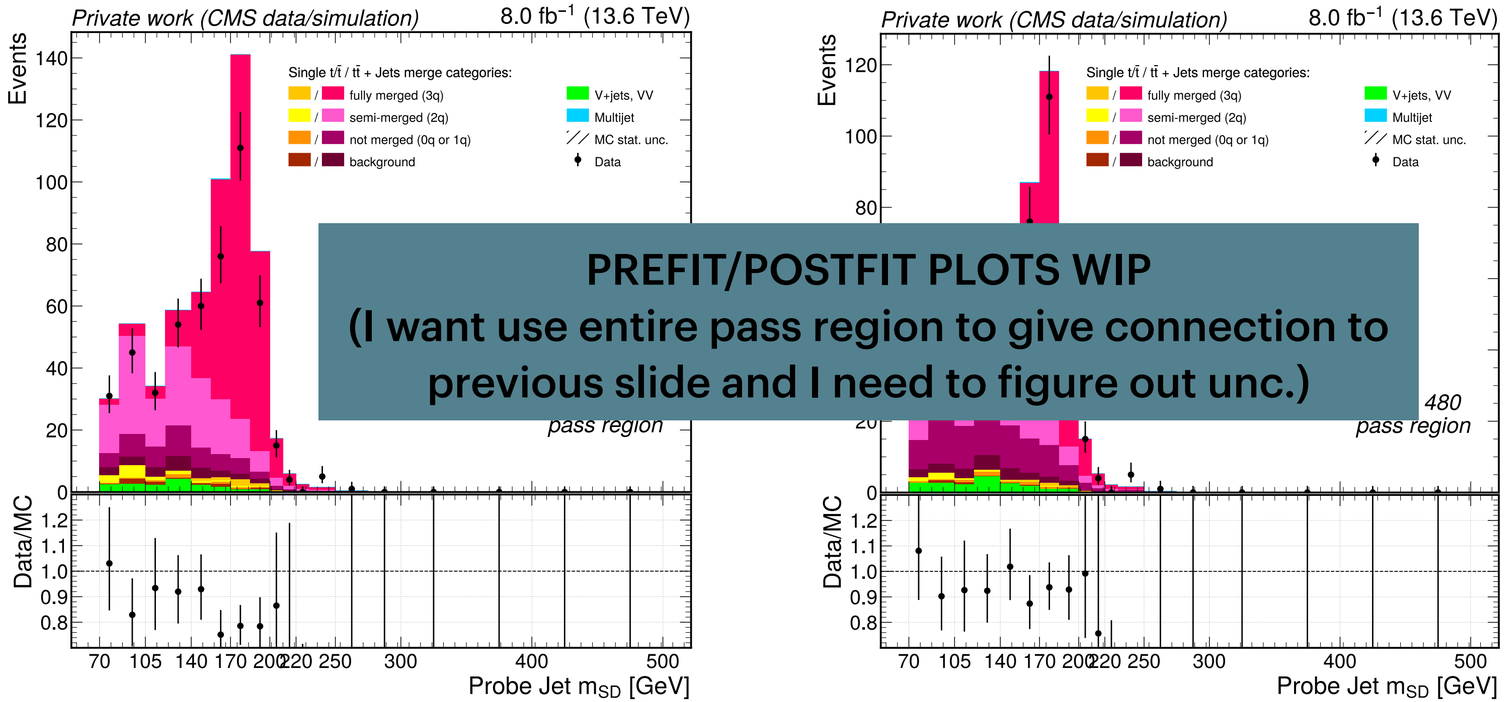




SF Derivation

— 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



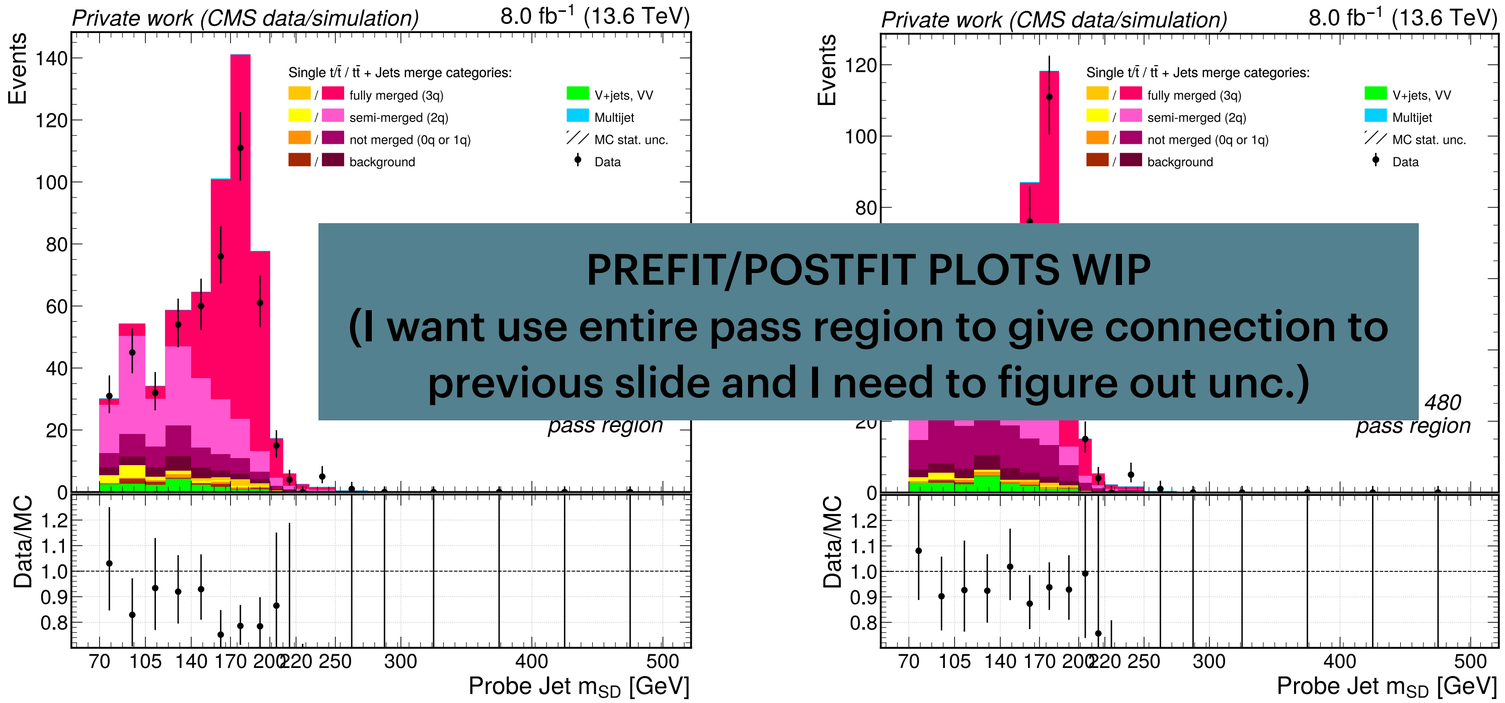
PREFIT/POSTFIT PLOTS WIP
(I want use entire pass region to give connection to previous slide and I need to figure out unc.)

SF Derivation

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

github.com/columnflow/columnflow
combine paper: arXiv:2404.06614

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

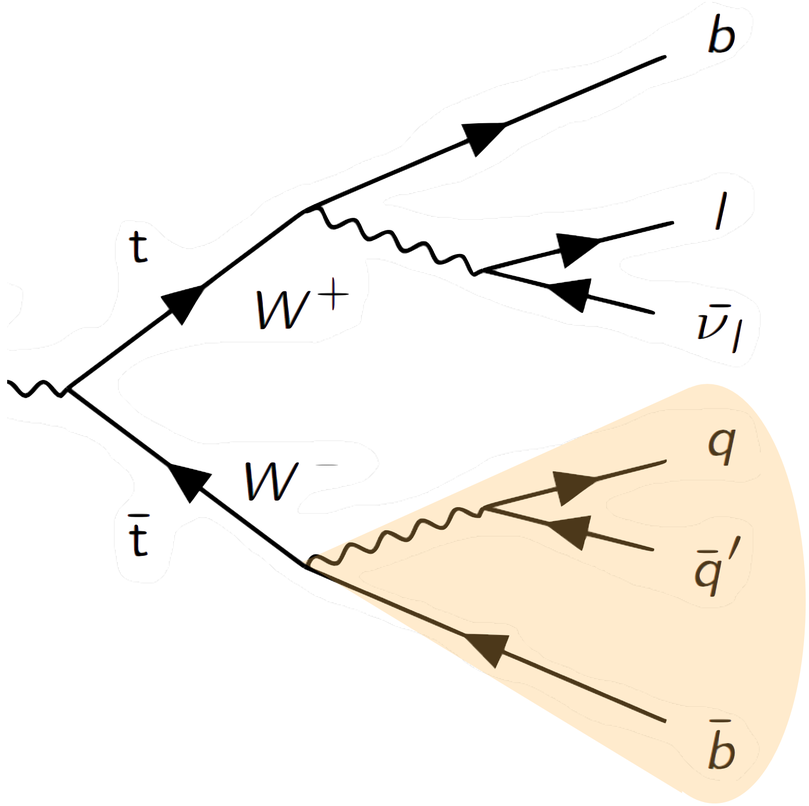


PREFIT/POSTFIT PLOTS WIP
(I want use entire pass region to give connection to previous slide and I need to figure out unc.)

Quick derivation of SF essential to be used in analyses.

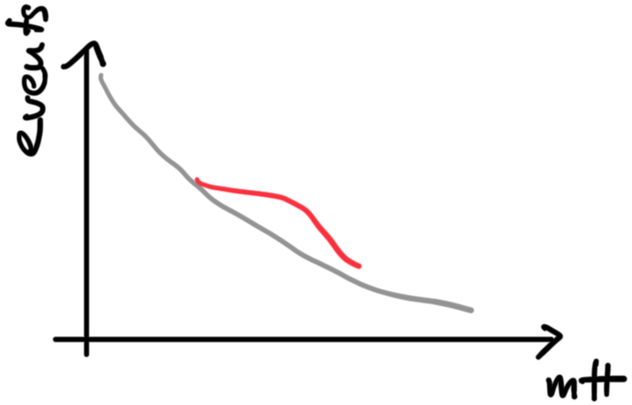
Boosted Top Quarks

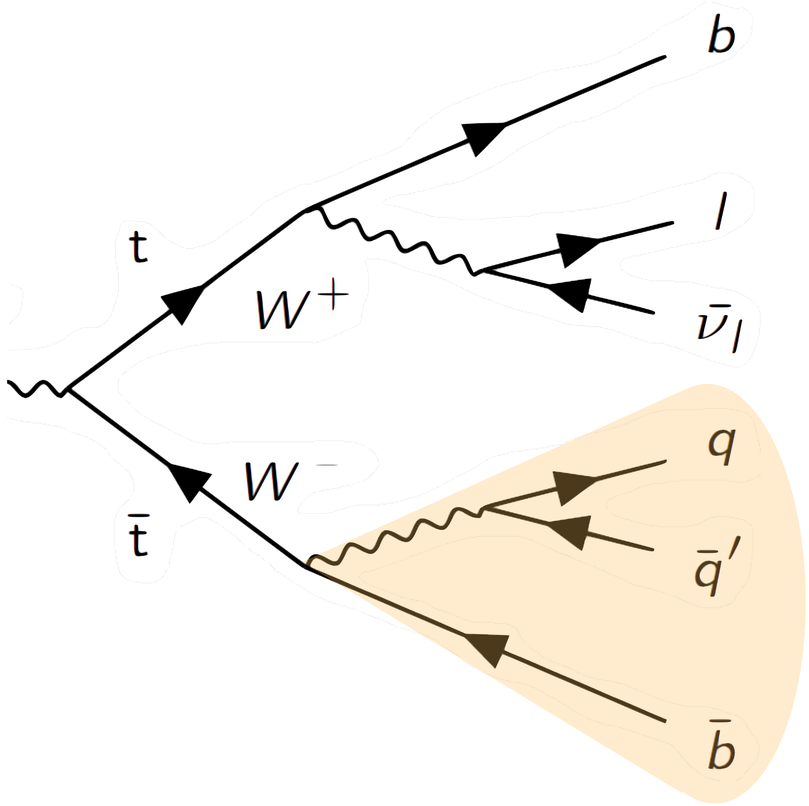
— in BSM signals —



Boosted Top Quarks

— in BSM signals —

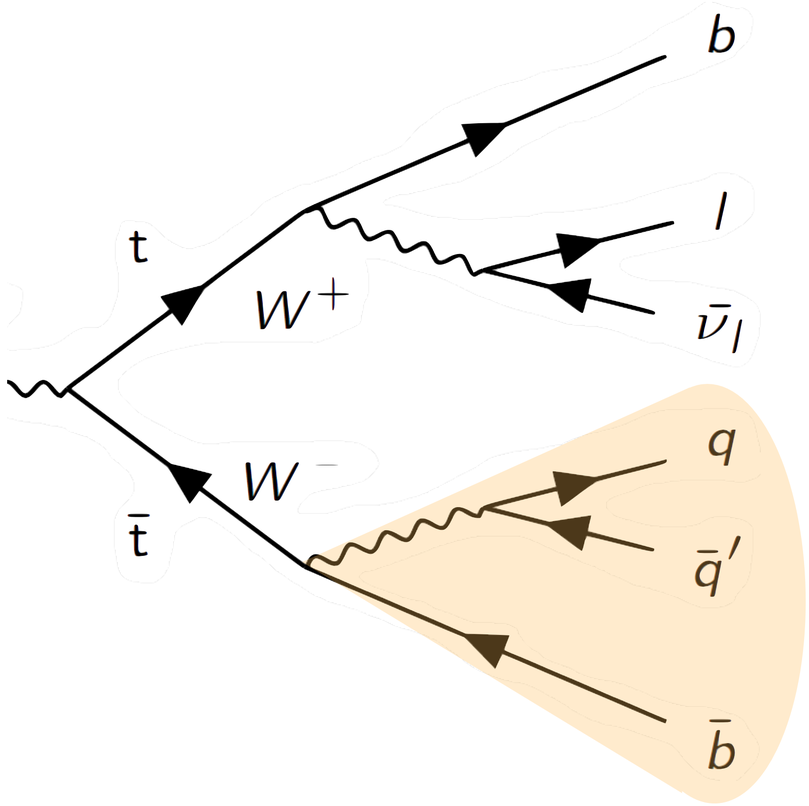


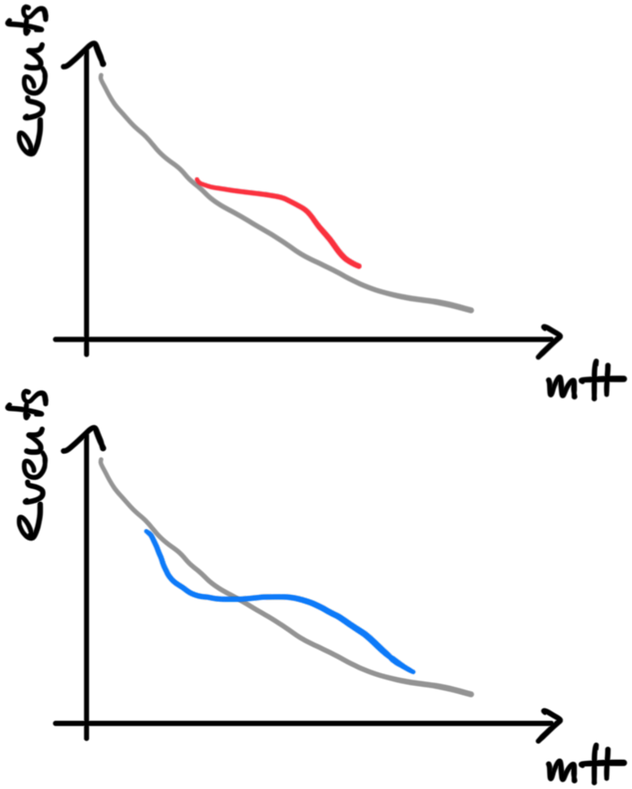


Boosted Top Quarks

— in BSM signals —

A/H Z'





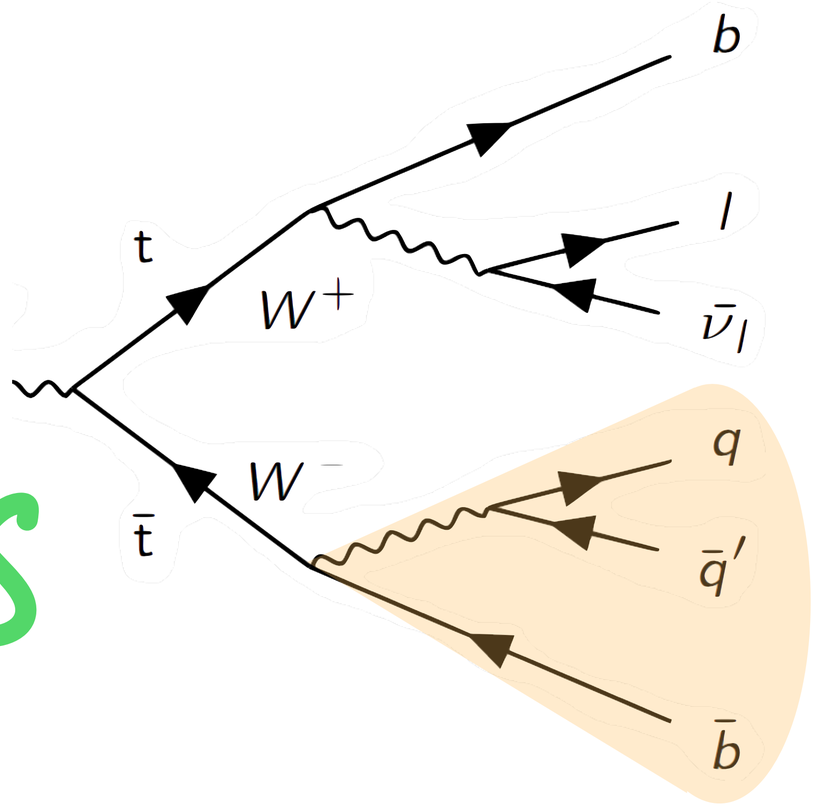
Boosted Top Quarks

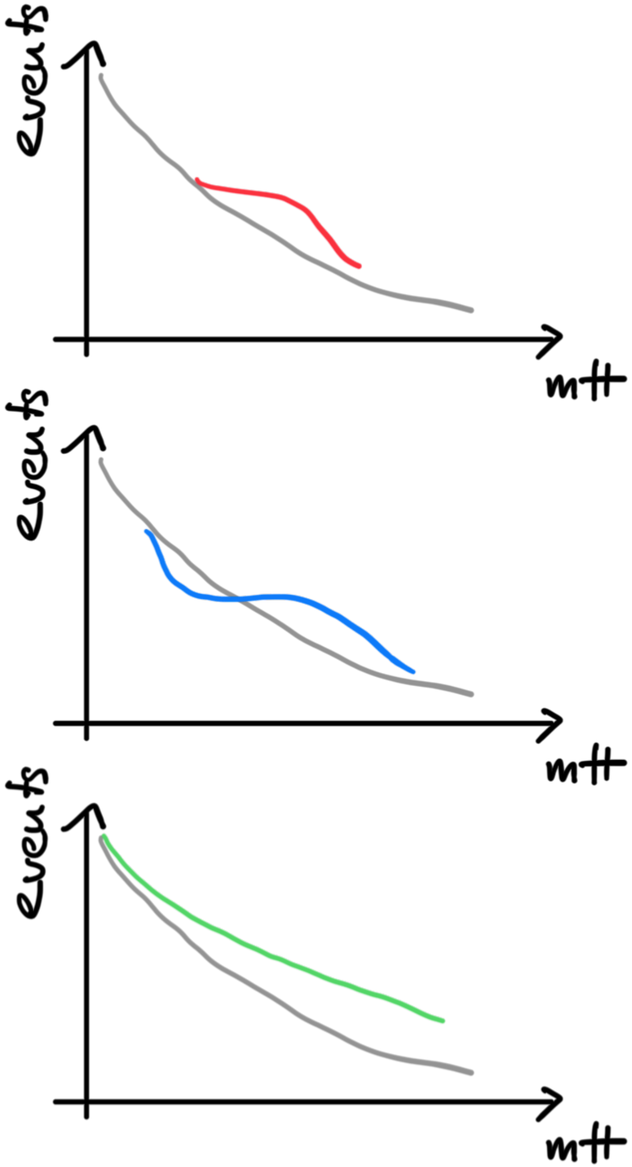
— in BSM signals —

A/H

Z'

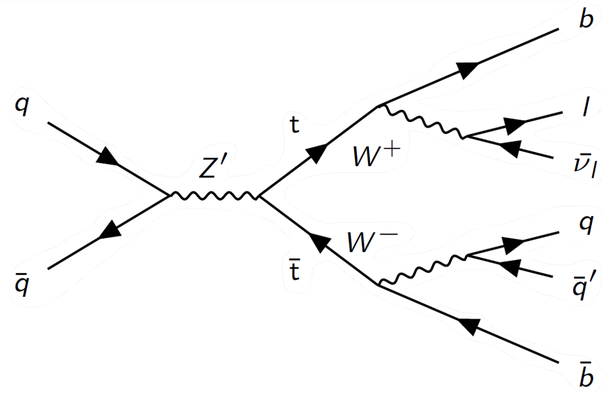
ALPS



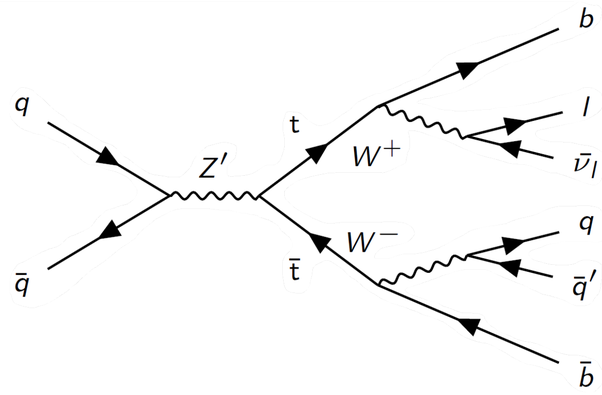


Analysis Strategy

- search for heavy Z like boson
- mass range from about $0.4 - 9$ TeV
- samples to be produced for Run 3 analysis



Analysis Strategy



- search for heavy Z like boson
- mass range from about $0.4 - 9$ TeV
- samples to be produced for Run 3 analysis

I. Selection

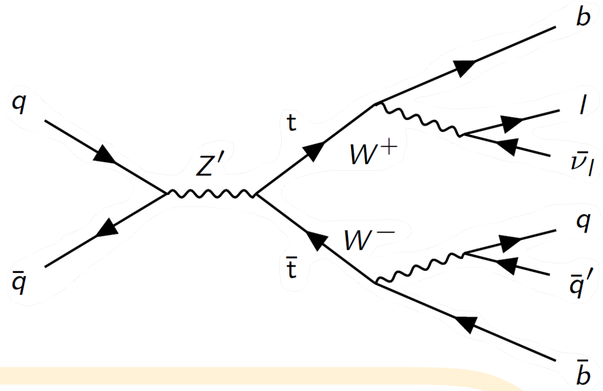
II. Reconstruction

III. Event Classification

IV. Statistical Analysis

Analysis Strategy

- search for heavy Z like boson
- mass range from about $0.4 - 9$ TeV
- samples to be produced for Run 3 analysis

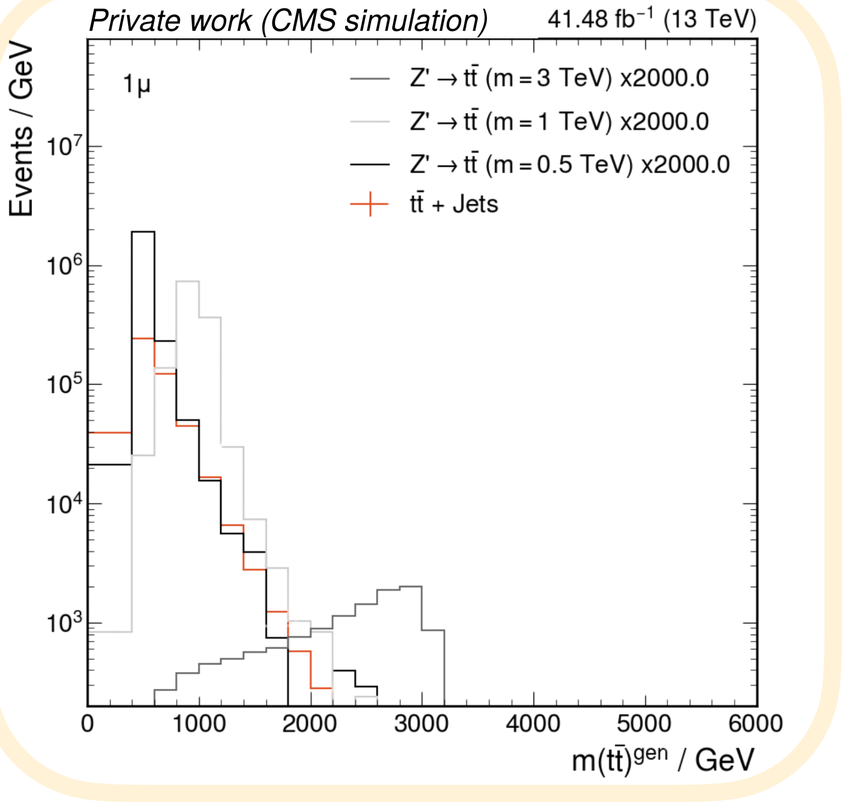


I. Selection

II. Reconstruction

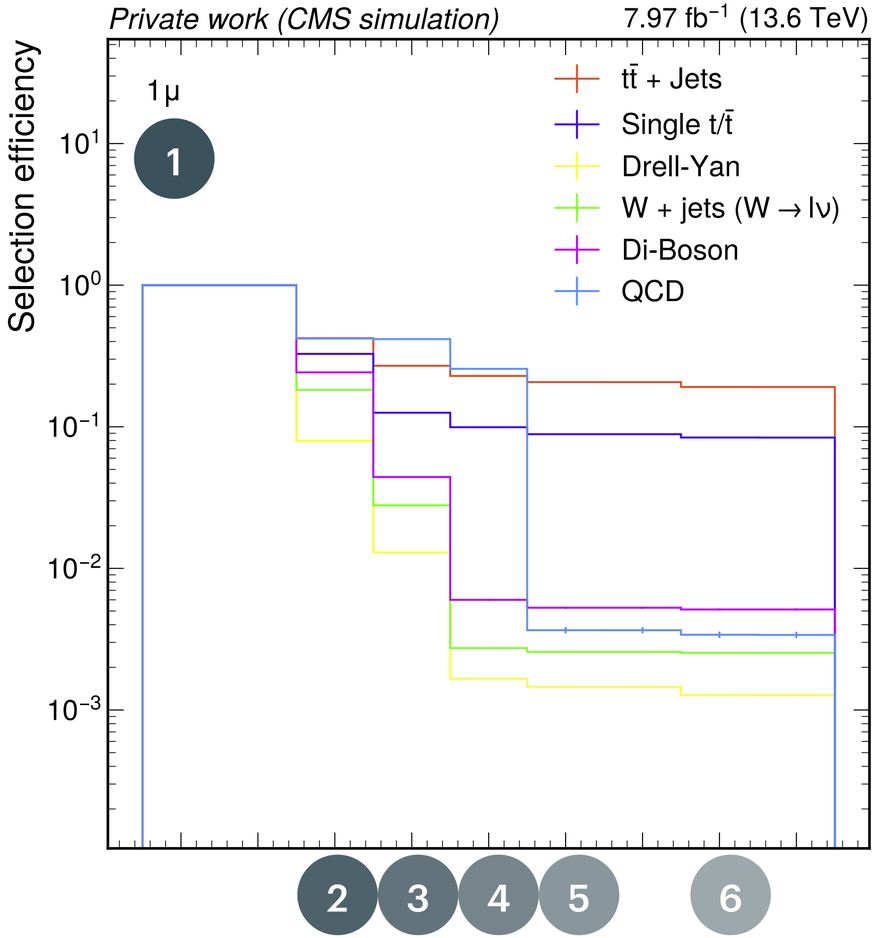
III. Event Classification

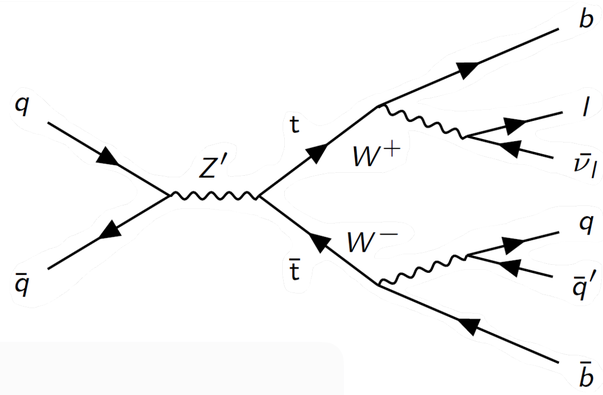
IV. Statistical Analysis



Event Selection

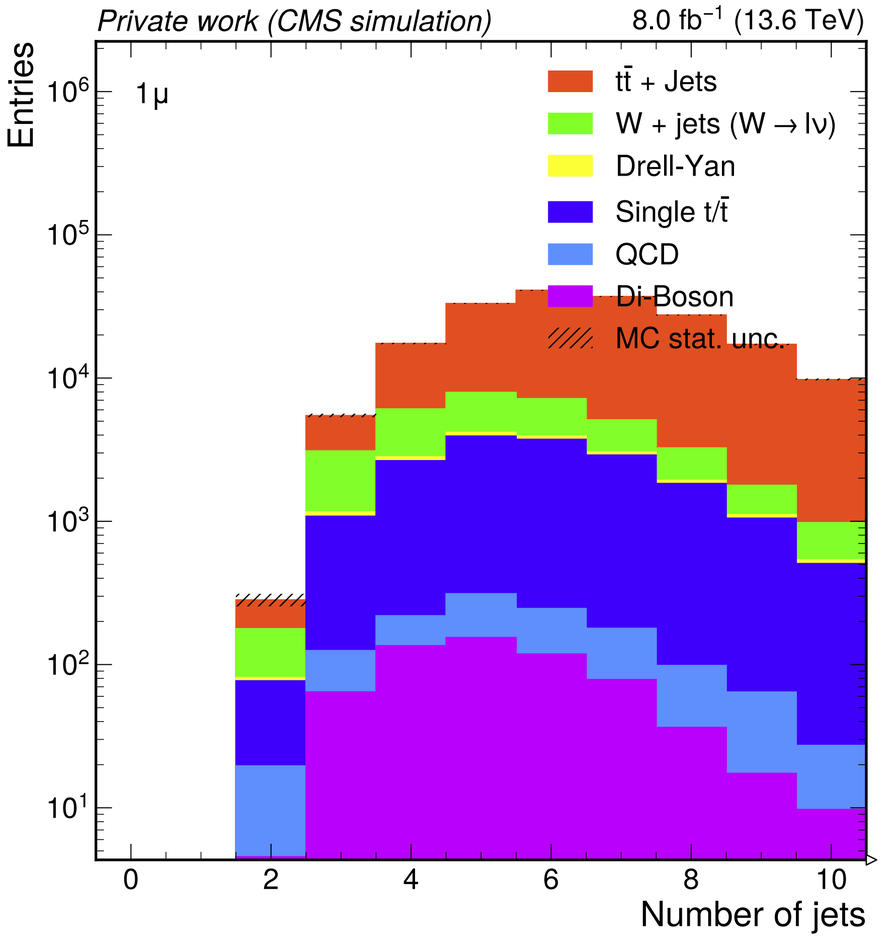
— efficiency per process —

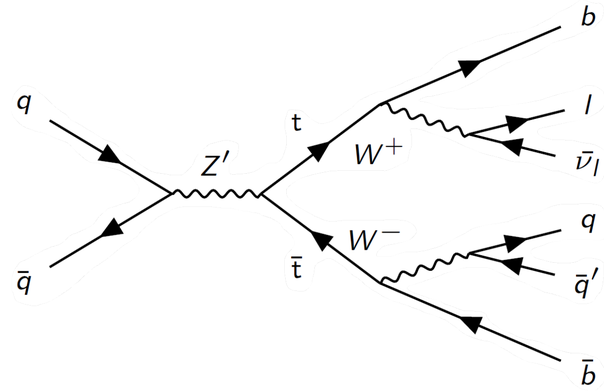




1. exactly one electron/muon
2. missing transverse momentum
3. at least two jets
4. at least one b-tagged jet
5. at high p_T of lepton:
   jet-lepton 2D isolation
   (relative momentum and spatial distance)
6. orthogonality to other channels

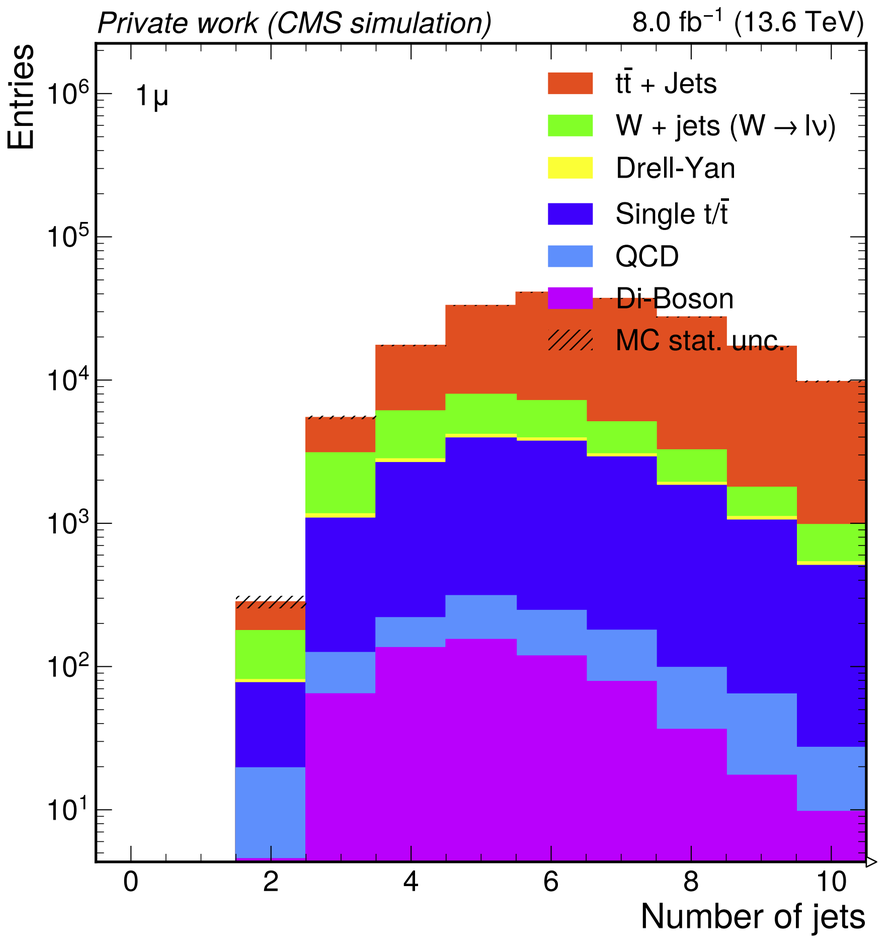
Reconstruct $t\bar{t}$ System

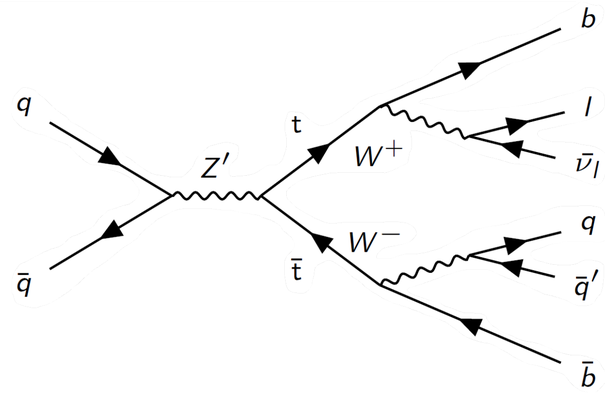




- <u>boosted regime</u>:
  assign top tagged large-radius jet to hadronic leg
- <u>resolved regime</u>:
  full jet combinatorics for both decay legs
  - choose best hypothesis based on χ^2 criterion

Reconstruct $t\bar{t}$ System

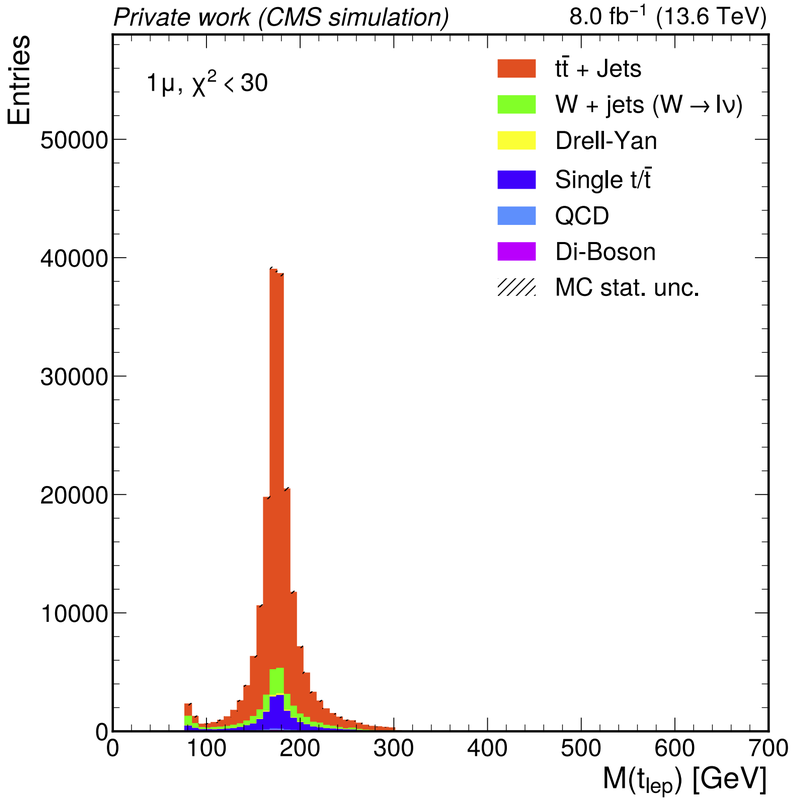




- <u>boosted regime</u>:
  assign top tagged large-radius jet to hadronic leg
- <u>resolved regime</u>:
  full jet combinatorics for both decay legs
  - choose best hypothesis based on χ^2 criterion

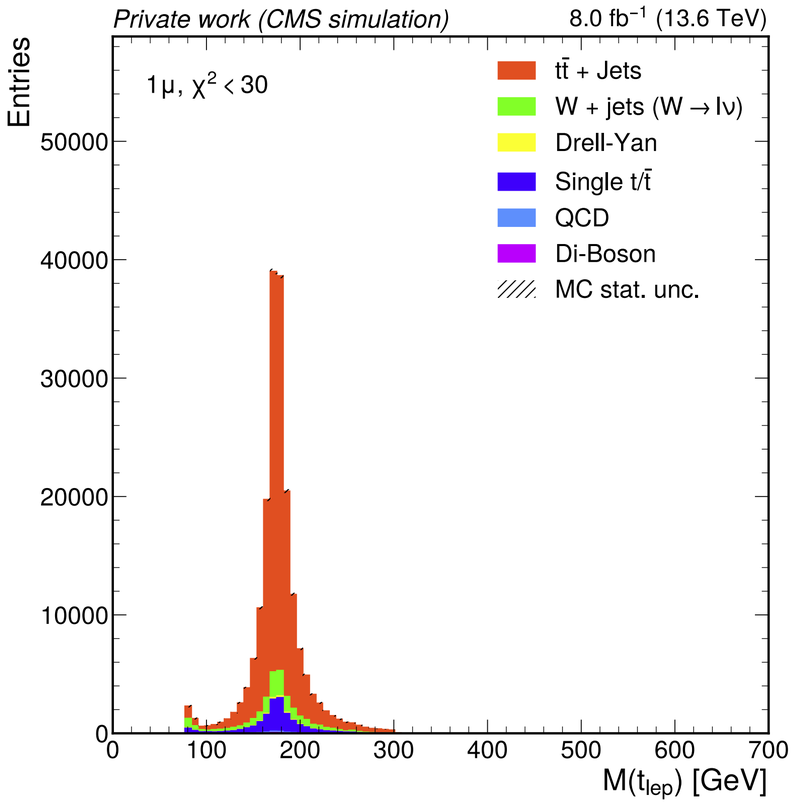
- generally, large combinatorics due to many jets in events
- memory intensive, also due to columnar based analysis
  - limit number of jets used for reconstruction without impacting efficiency of reconstruction

Reconstructed $t\bar{t}$ System

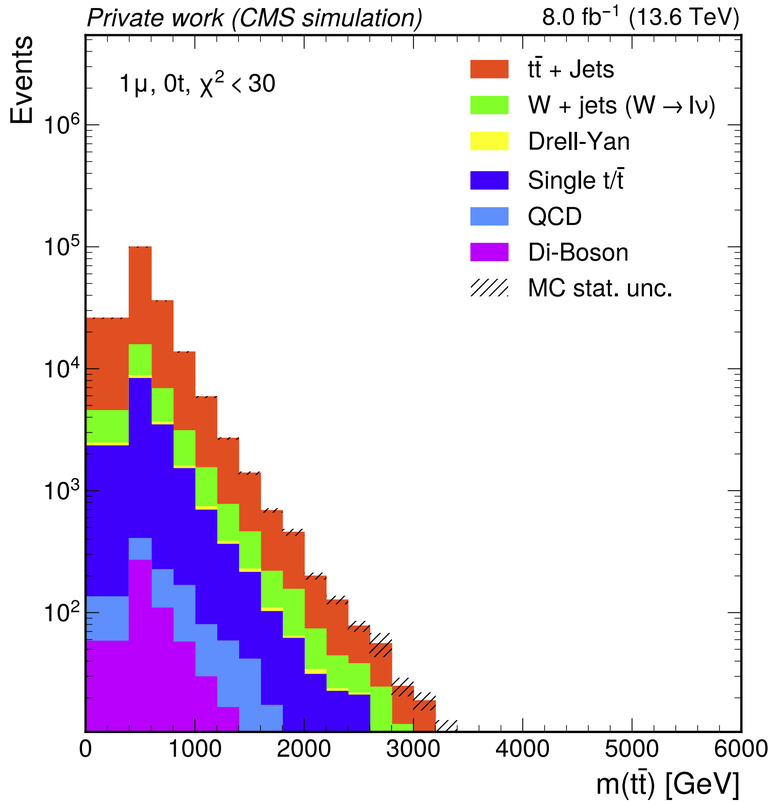


reconstructed mass of leptonic decay leg

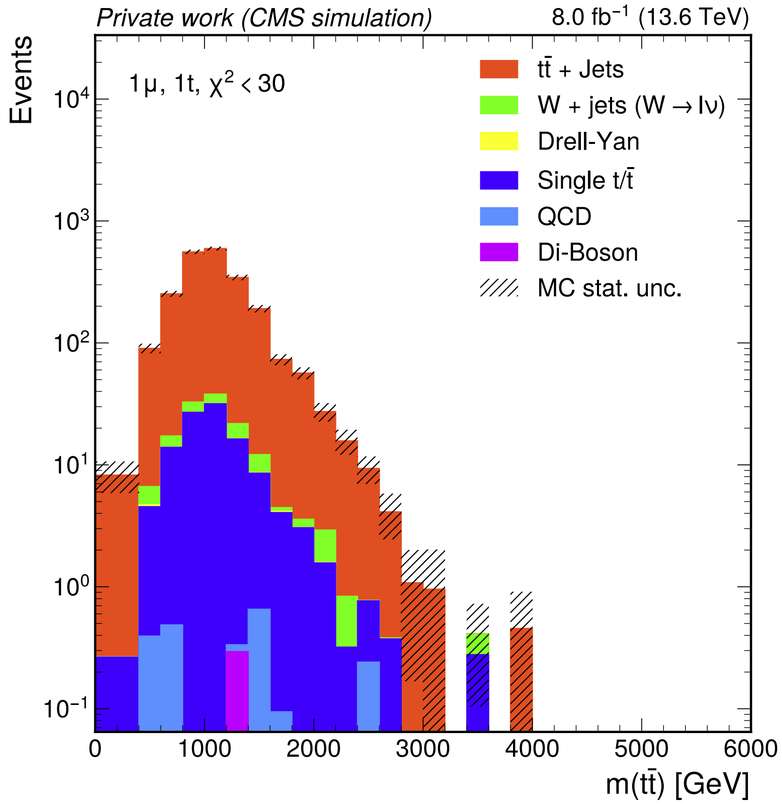
Reconstructed $t\bar{t}$ System



reconstructed mass of leptonic decay leg



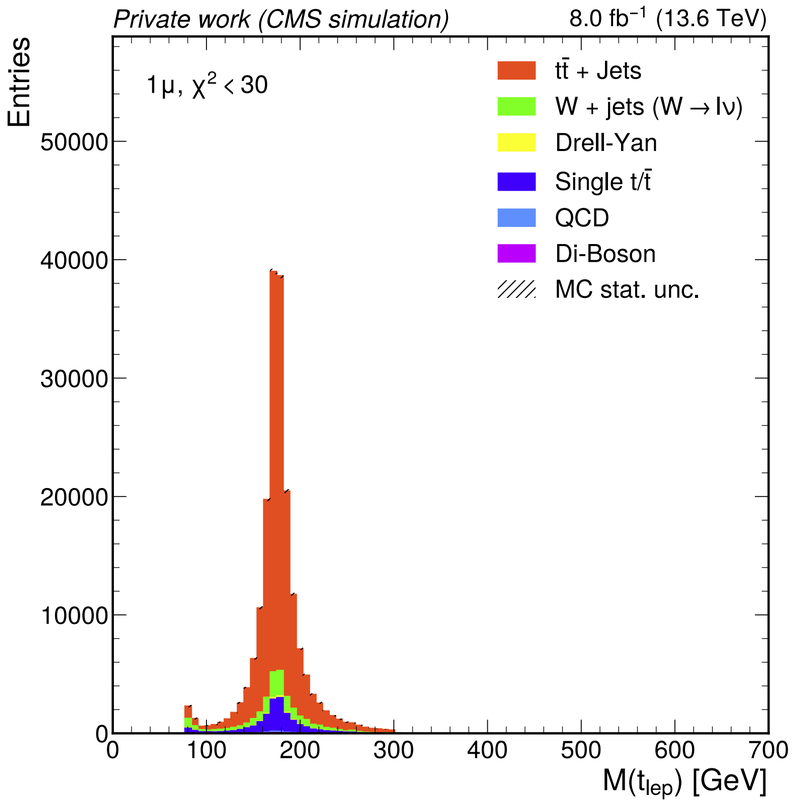
resolved regime



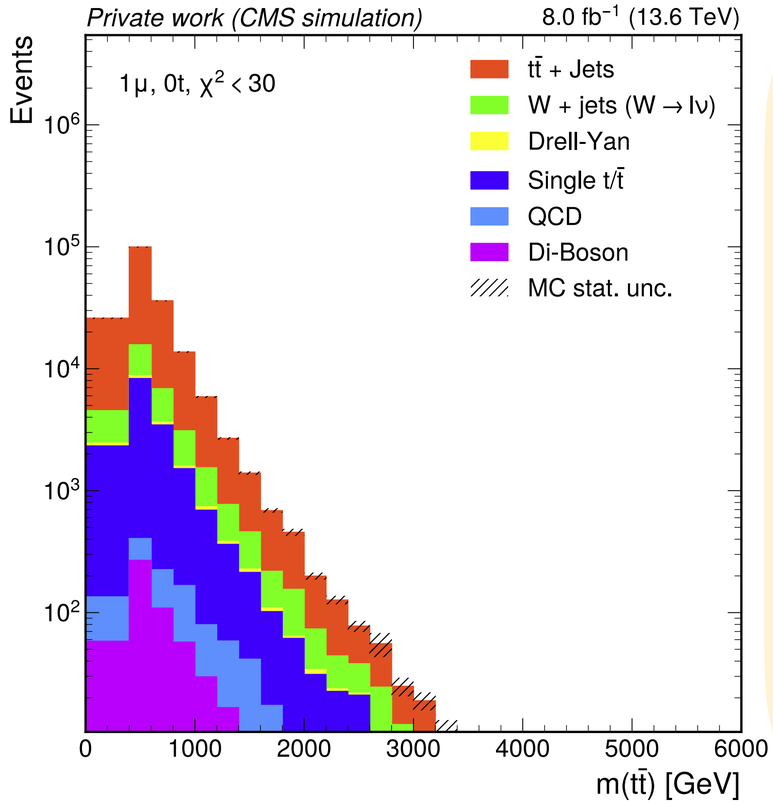
boosted regime

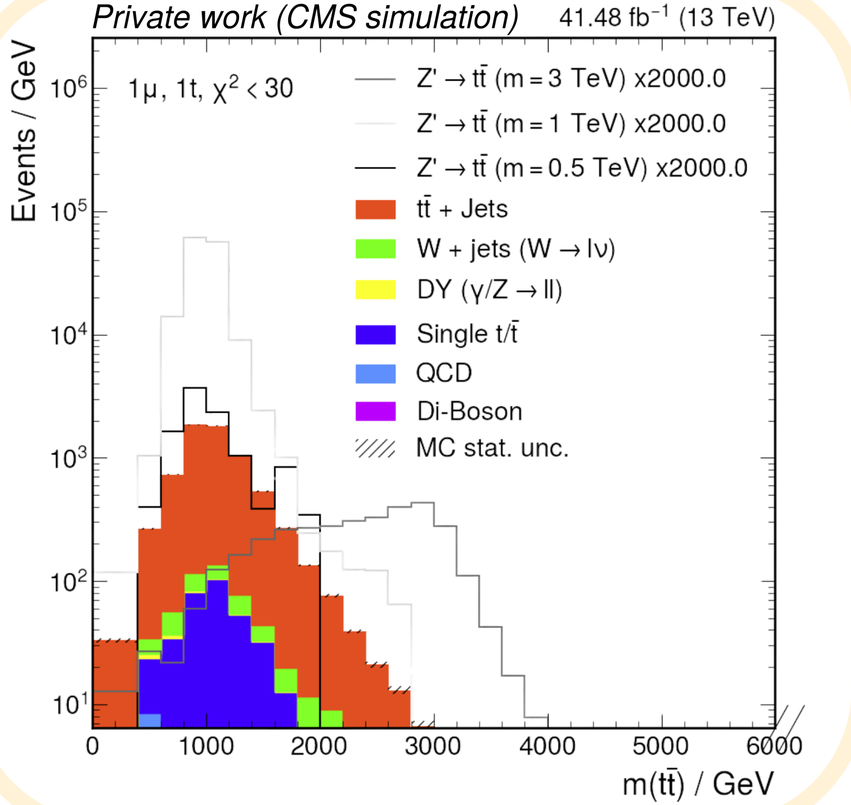
ditop mass

Reconstructed $t\bar{t}$ System



reconstructed mass of leptonic decay leg





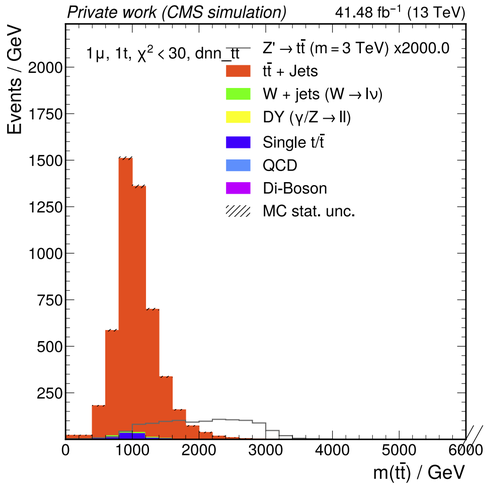
resolved regime — ditop mass — boosted regime

Outlook

— Run 2: DNN approach and statistical analysis —

- simple neural network setup for event classification:
  - 59 low level input features
  - 2 hidden layers
  - 4 output nodes: ttbar, single top, W+jets, DY

- enrich signal region with ttbar events to enhance sensitivity
- use control regions to constrain backgrounds

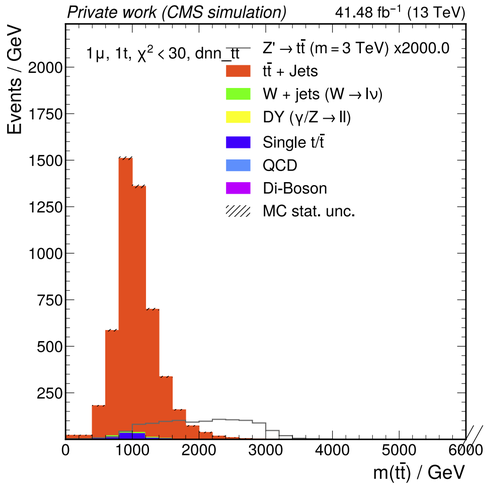


Outlook

— Run 2: DNN approach and statistical analysis —

- simple neural network setup for event classification:
  - 59 low level input features
  - 2 hidden layers
  - 4 output nodes: ttbar, single top, W+jets, DY

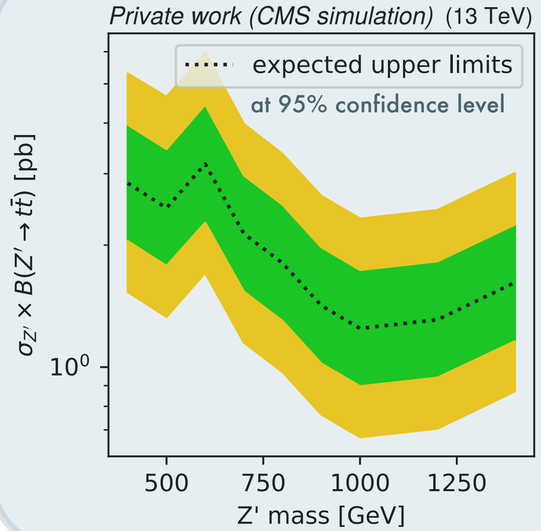
- enrich signal region with ttbar events to enhance sensitivity
- use control regions to constrain backgrounds



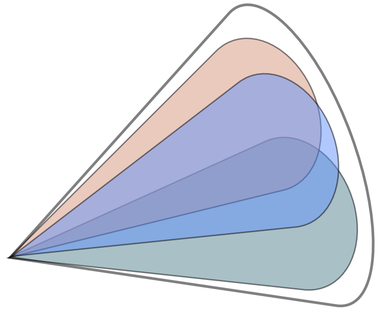
- template-based binned likelihood fit with ditop mass as sensitive variable using CMS combine tool

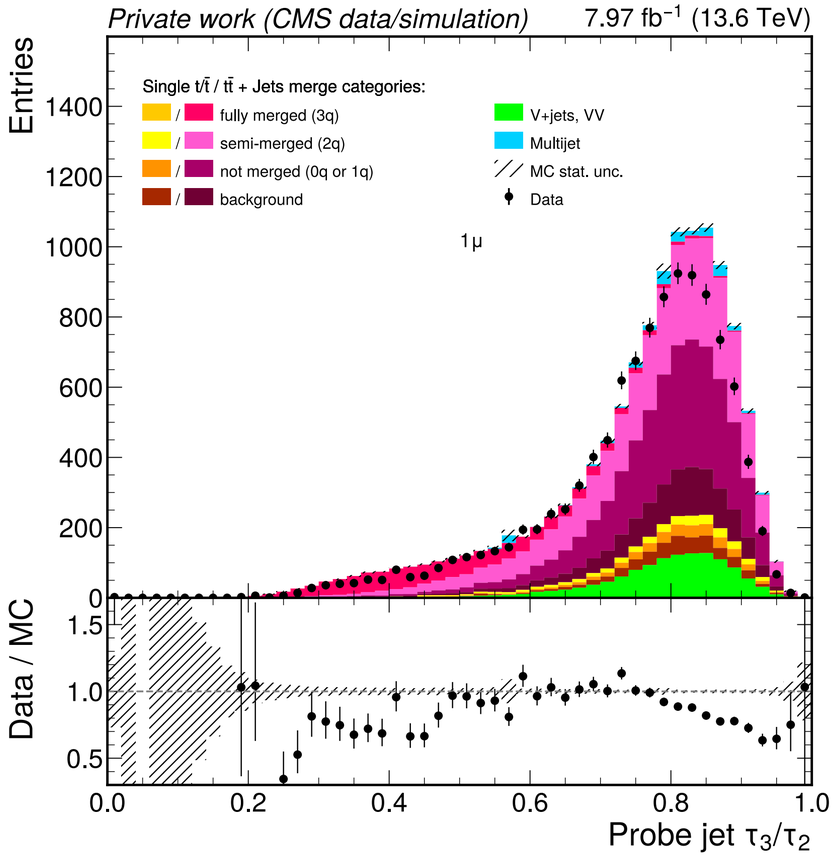
- if no significant excess observed: set upper limits on $\sigma_{Z'} \times BR(Z' \to t\bar{t})$ at 95 % confidence level

for illustration:



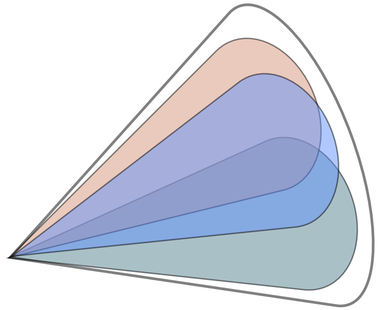
Summary





- identification of boosted objects essential step in many analyses
- **boosted top quarks** as potential window to BSM physics
  - Run 2 analysis expected to be published very soon
  - Run 3 analysis setup to look into new data
- understand **cut based top tagging algorithm** using jet substructure and soft drop mass
  - derive working points and scale factors

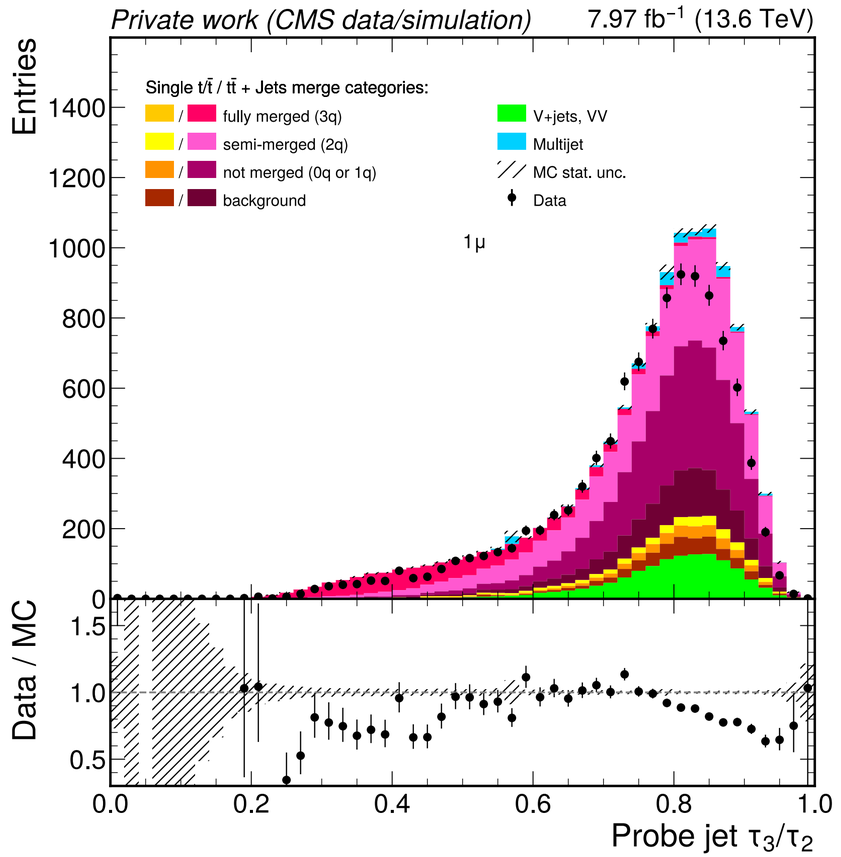
Summary



- identification of boosted objects essential step in many analyses

- **boosted top quarks** as potential window to BSM physics
  - Run 2 analysis expected to be published very soon
  - Run 3 analysis setup to look into new data

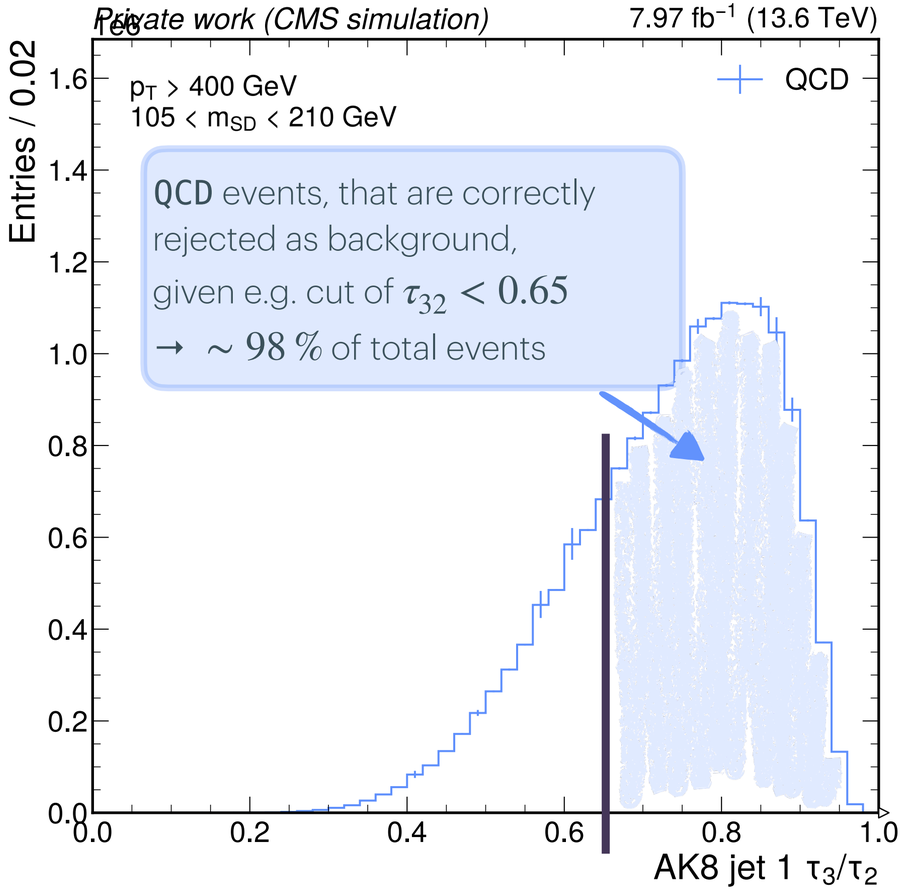
- understand **cut based top tagging algorithm** using jet substructure and soft drop mass
  - derive working points and scale factors

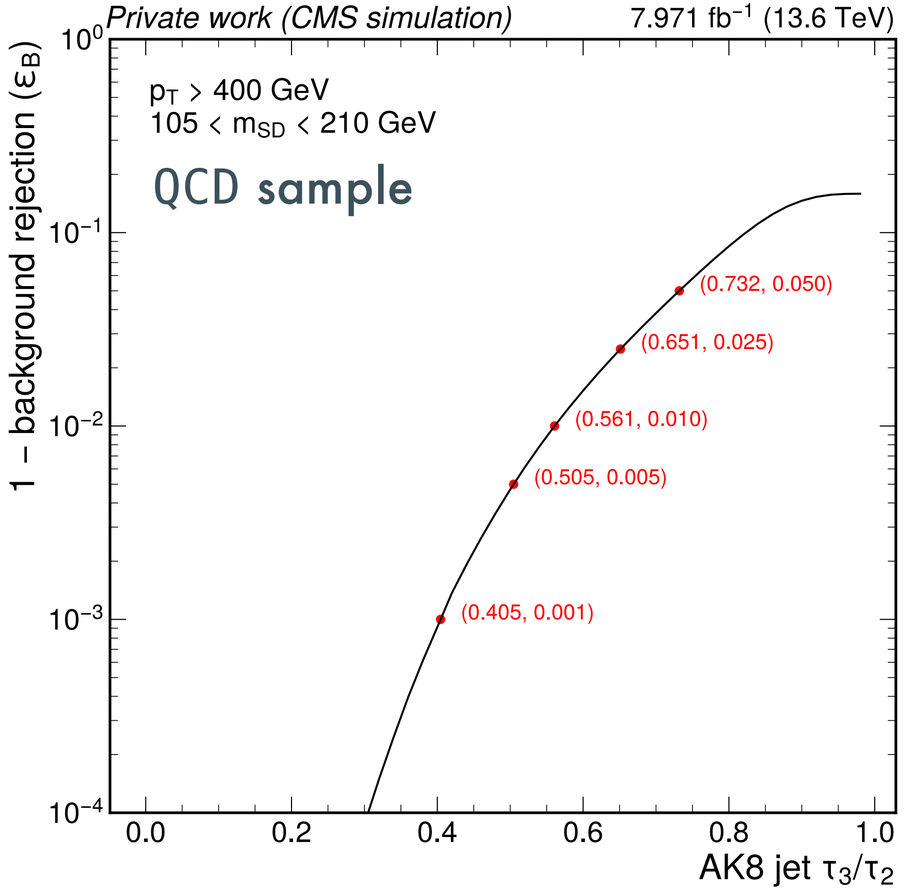


Stay tuned!

BACKUP

Mis-Tag Rate

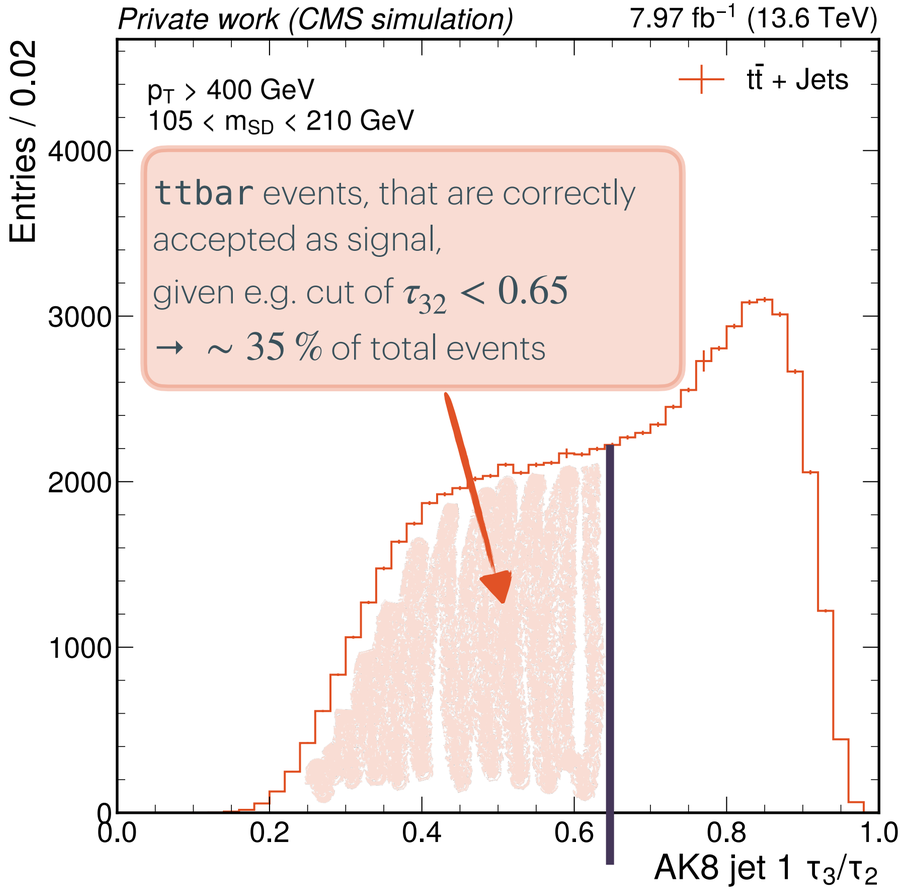


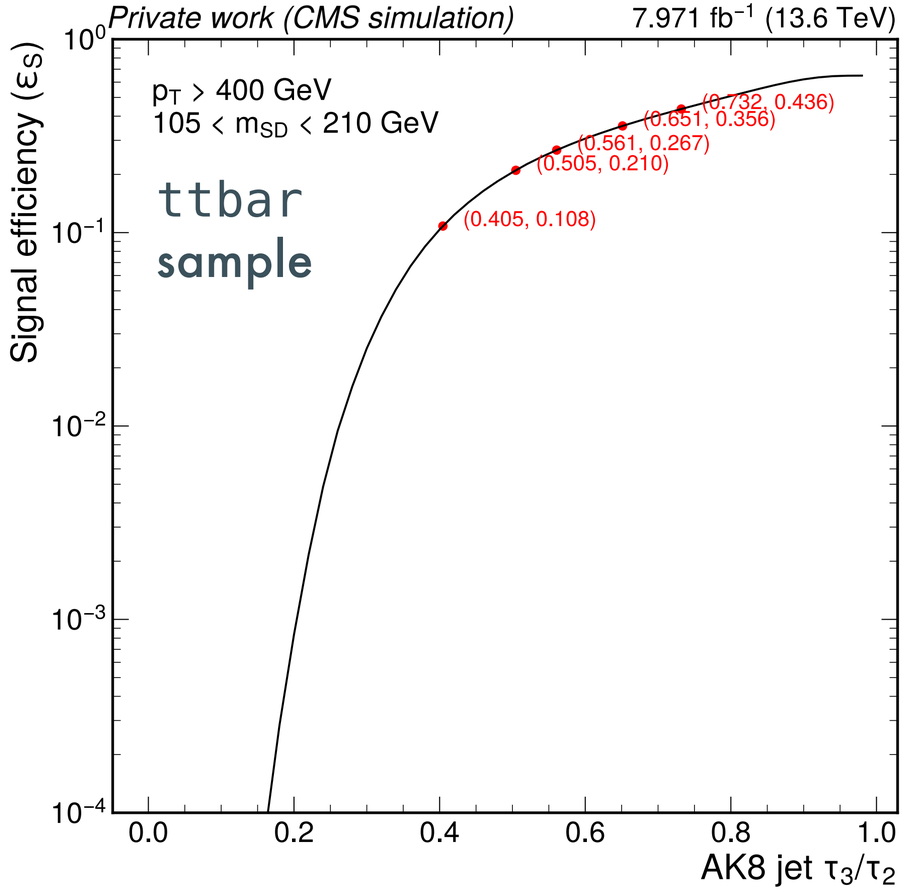


number of QCD events with a FatJet passing selection AND (incorrectly) tagged as a top jet given the τ_{32} cut

total number of events in QCD dataset

Signal Efficiency





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

total number of events in `ttbar` dataset

m_{SD} Of FatJets

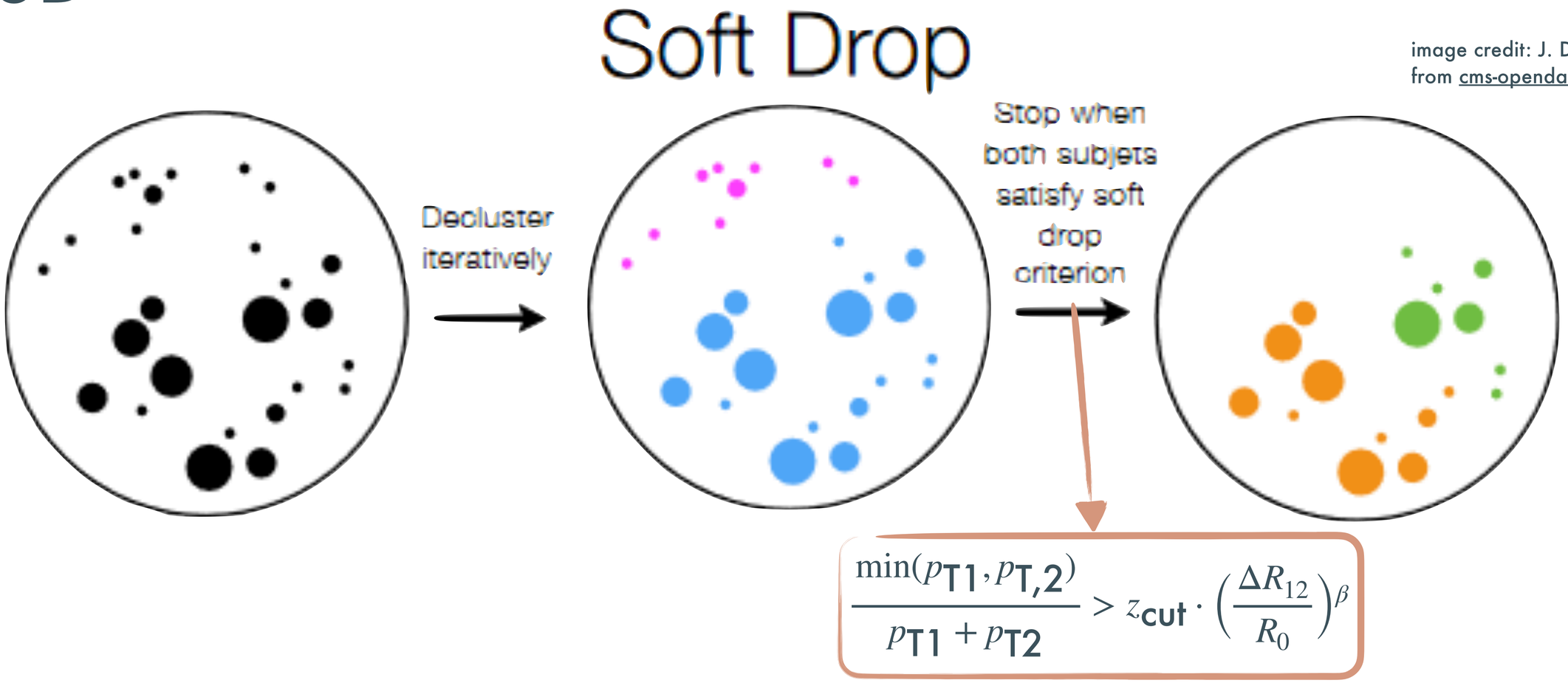
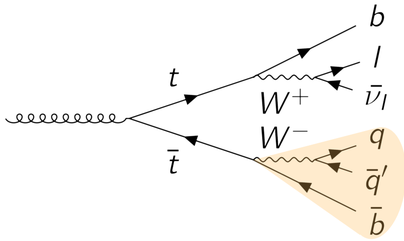
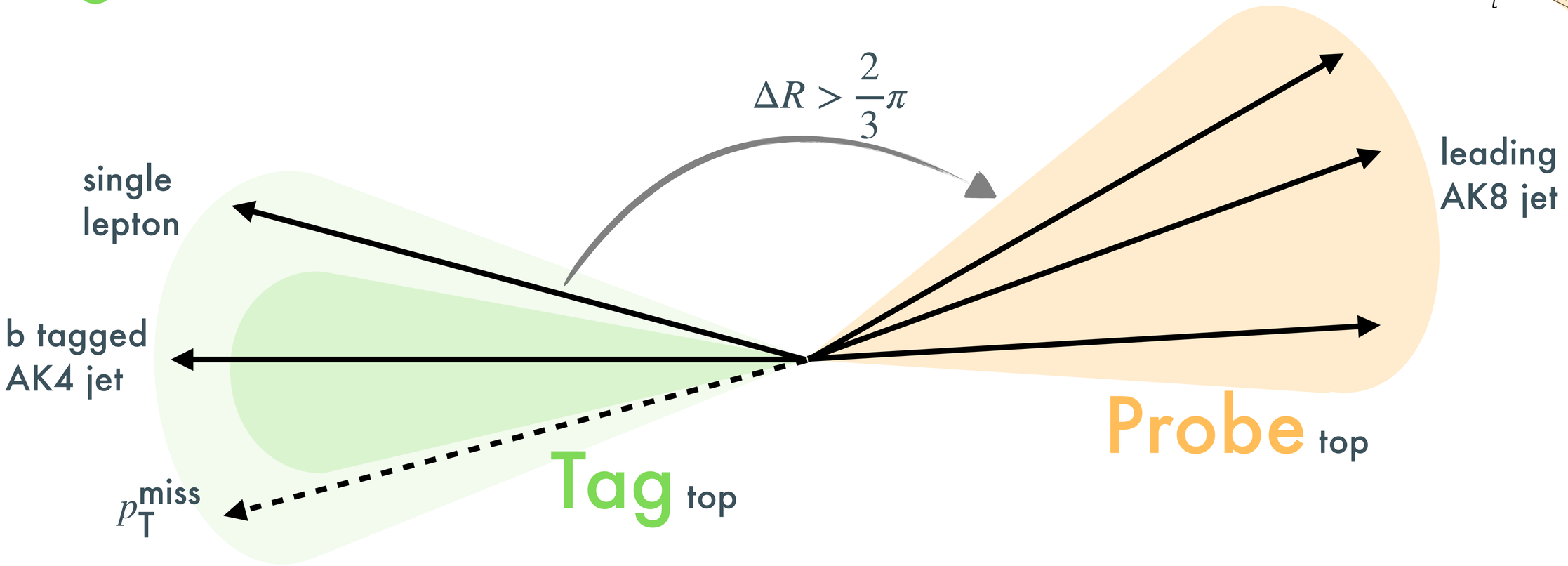


image credit: J. Dolen
from cms-opendata-workshop

→ 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

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

Impacts

— Text —