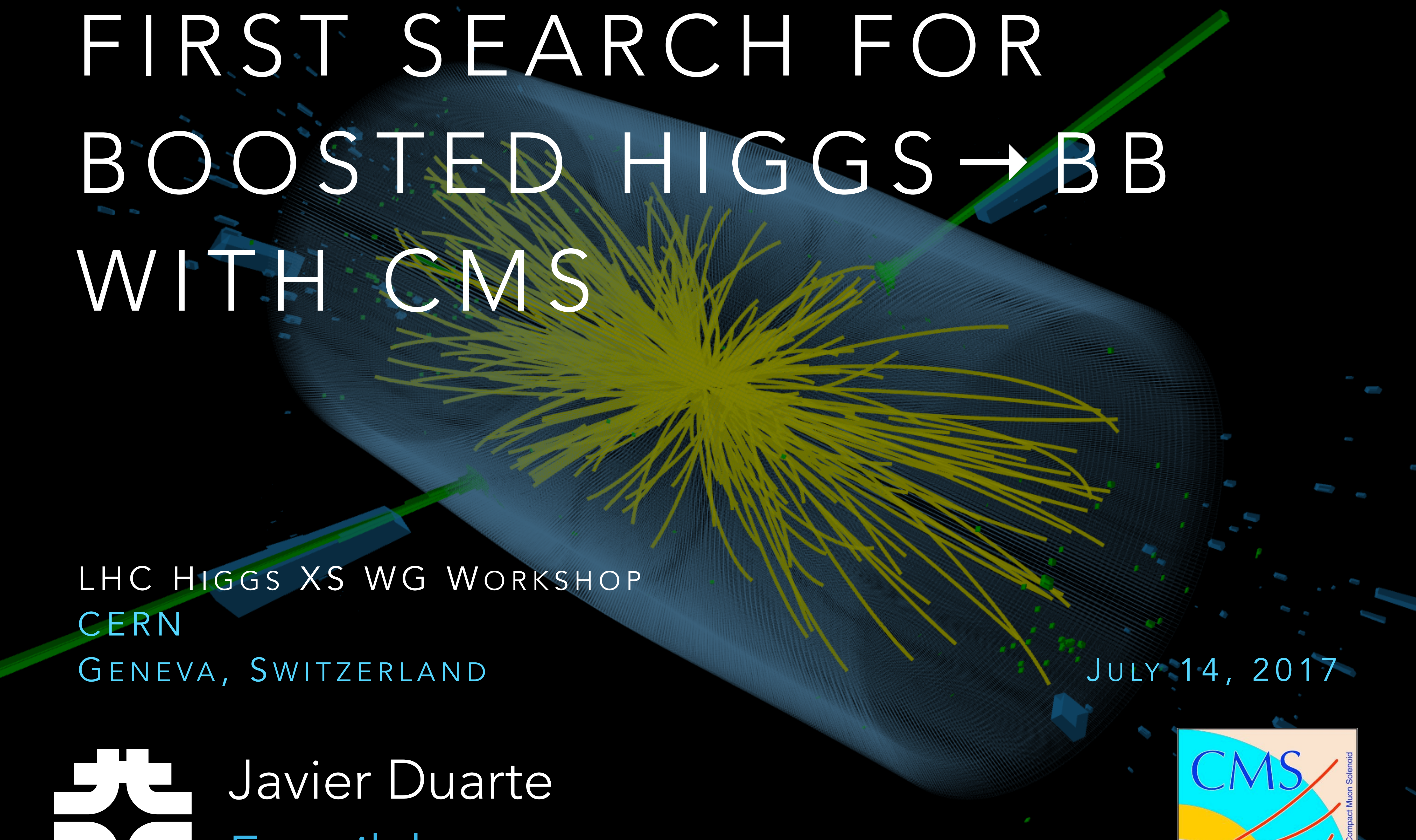
BOOSTED GGF H(BB)

FIRST SEARCH FOR BOOSTED HIGGS→BB WITH CMS



LHC Higgs XS WG Workshop

CERN

Geneva, Switzerland

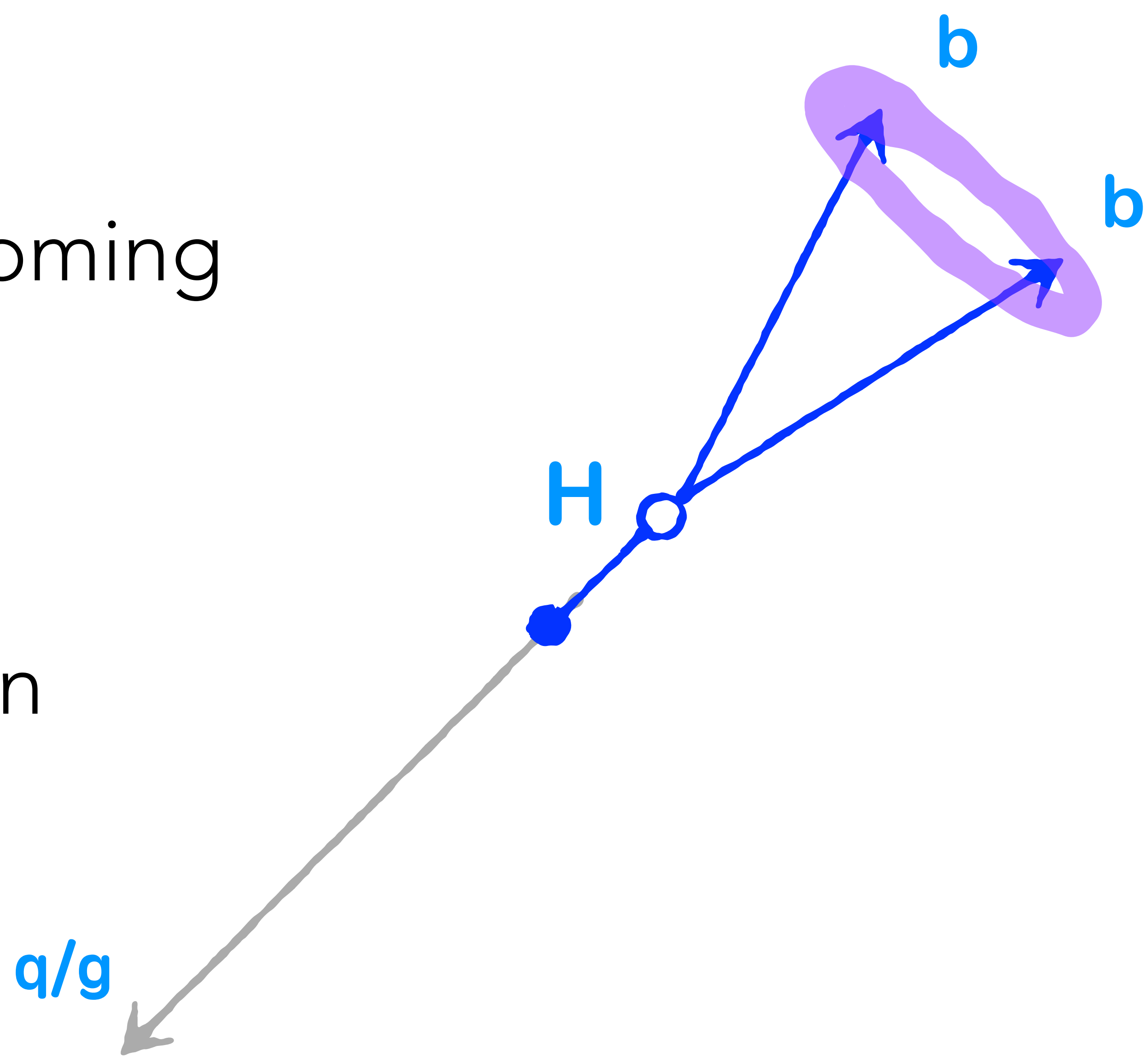
July 14, 2017

Javier Duarte

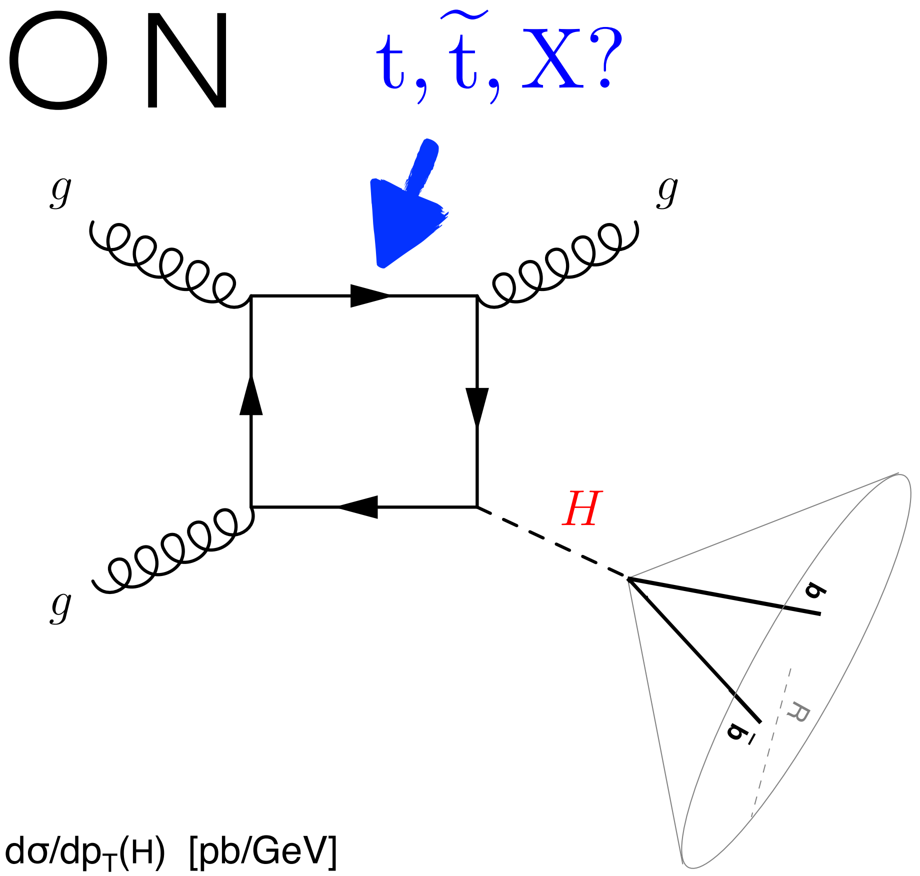
Fermilab

OUTLINE

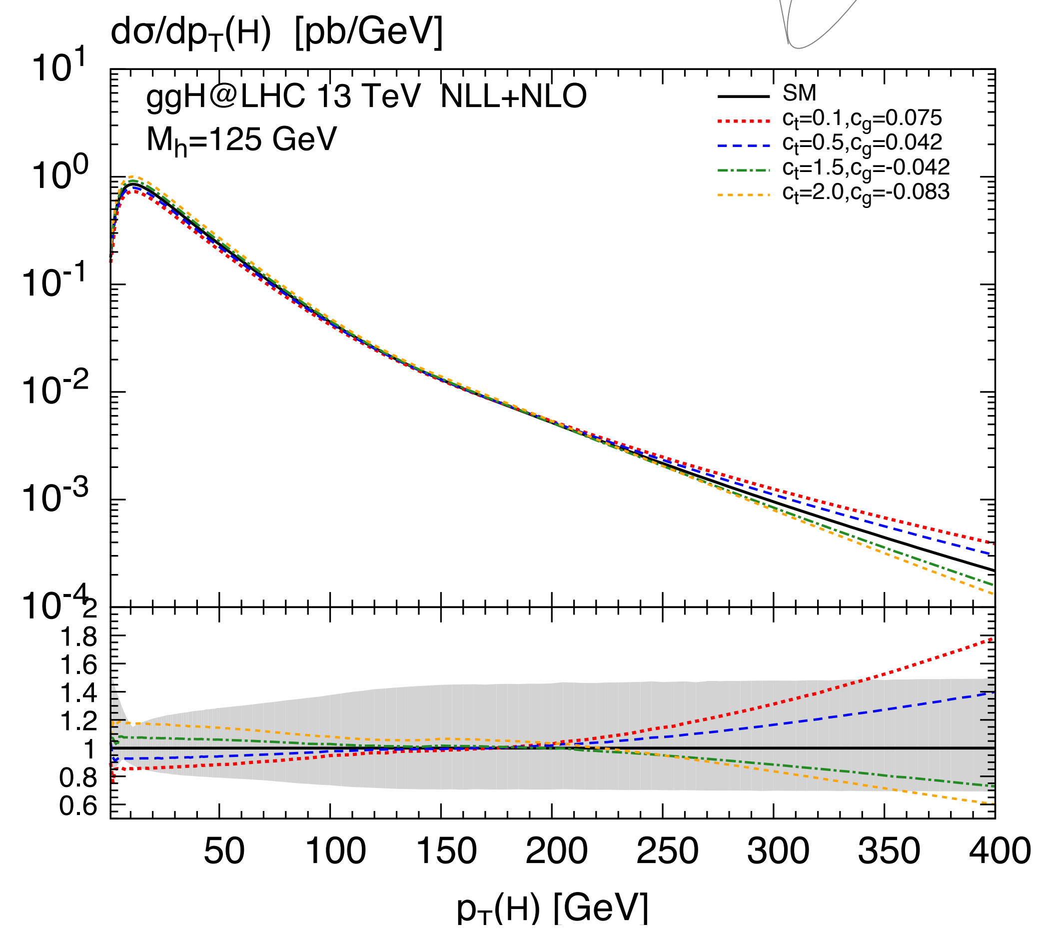
- Motivation
- Experimental techniques
  - Jet substructure and grooming
  - Double-b-tagging
- Event selection
- Data-driven QCD estimation
- Higgs p_T modeling
- Results
- Summary and outlook



MOTIVATION

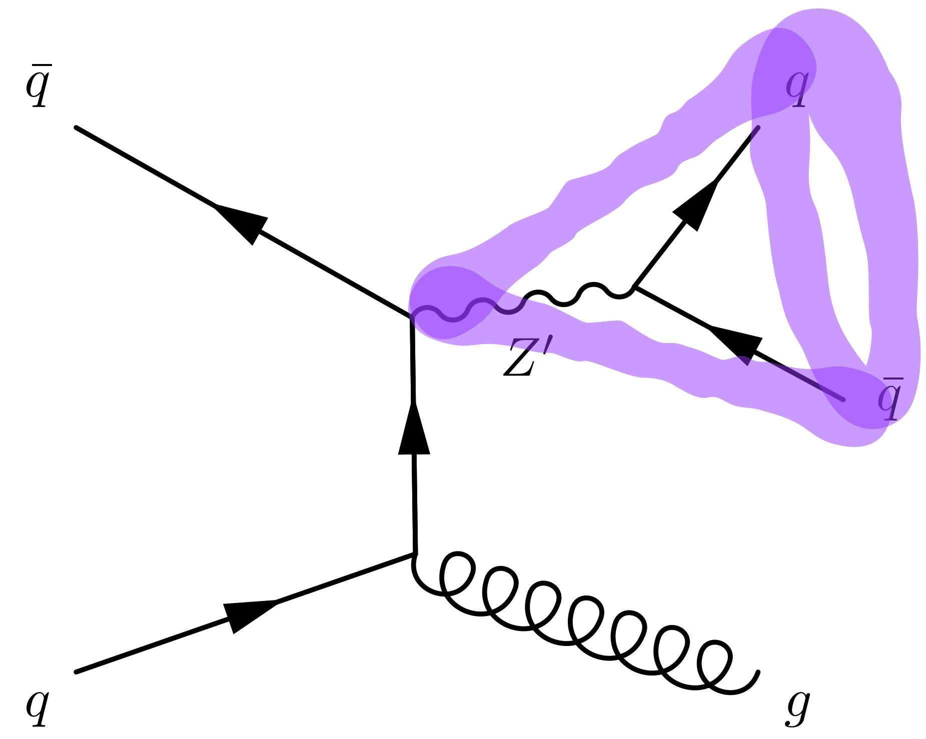


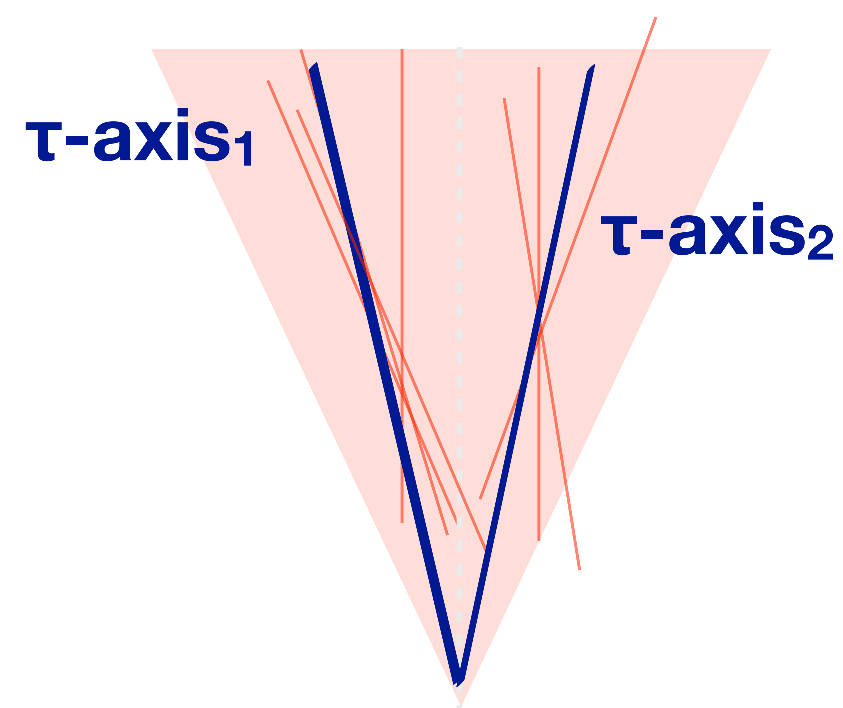
- Search for gg → H → bb historically thought impossible due to overwhelming and difficult to predict QCD background
- We can access this process in the boosted dijet topology
- Probing Higgs couplings at high momentum transfer (Q) accesses large new physics energy scale (Λ)



SO HOW CAN WE DO IT?

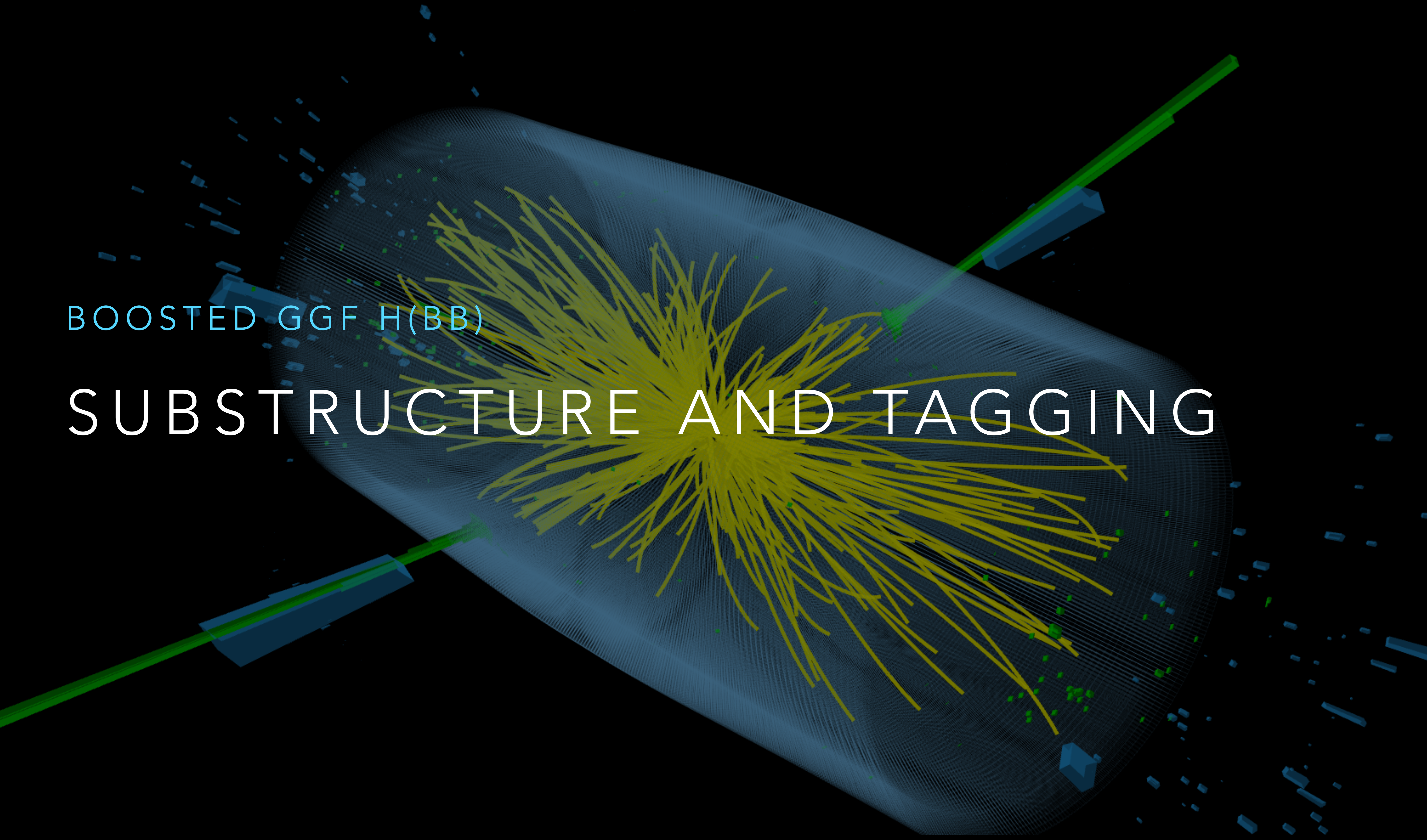
- Inspiration from boosted Z′+jet search?
  - ✔ Use ISR jet to get you above the trigger threshold
  - ✔ Requires one boosted fat jet
  - ✔ Substructure and jet grooming to enhance S/B
  - ✔ Data-driven background estimate
- Inspiration from machine learning and b-tagging?
  - ✔ Double b-tagger selects fat jets containing two b-quarks





BOOSTED GGF H(BB)

SUBSTRUCTURE AND TAGGING

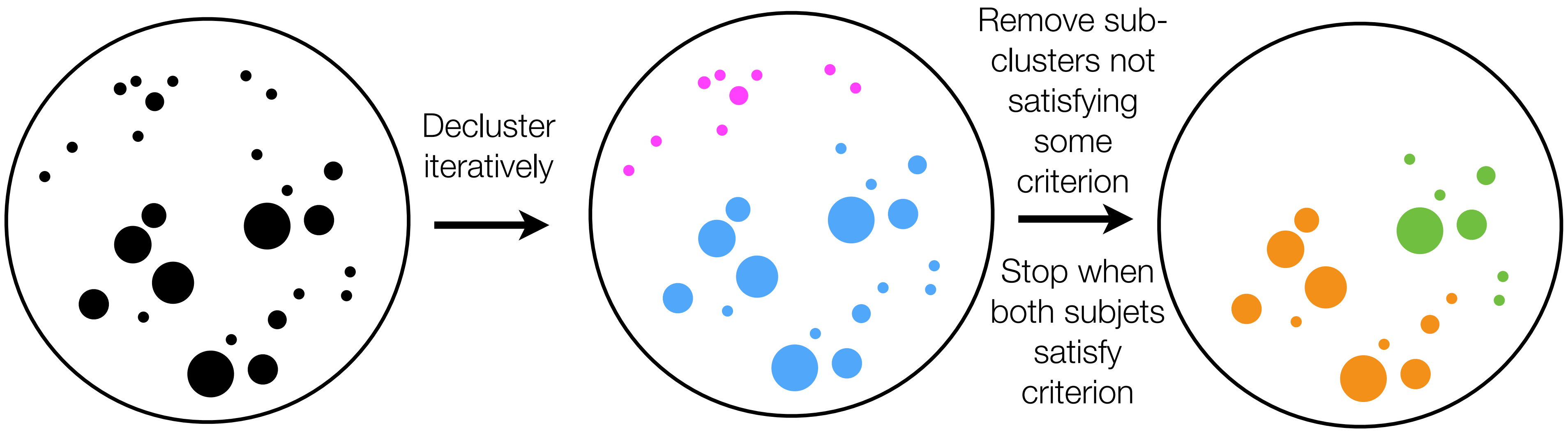


Javier Duarte
Fermilab

arXiv:1307.0007
arXiv:1402.2657

JET MASS

- Provides good separation between W/Z/H-jets from q/g jets
- Grooming removes soft and wide-angle radiation (soft drop / modified mass soft drop)



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

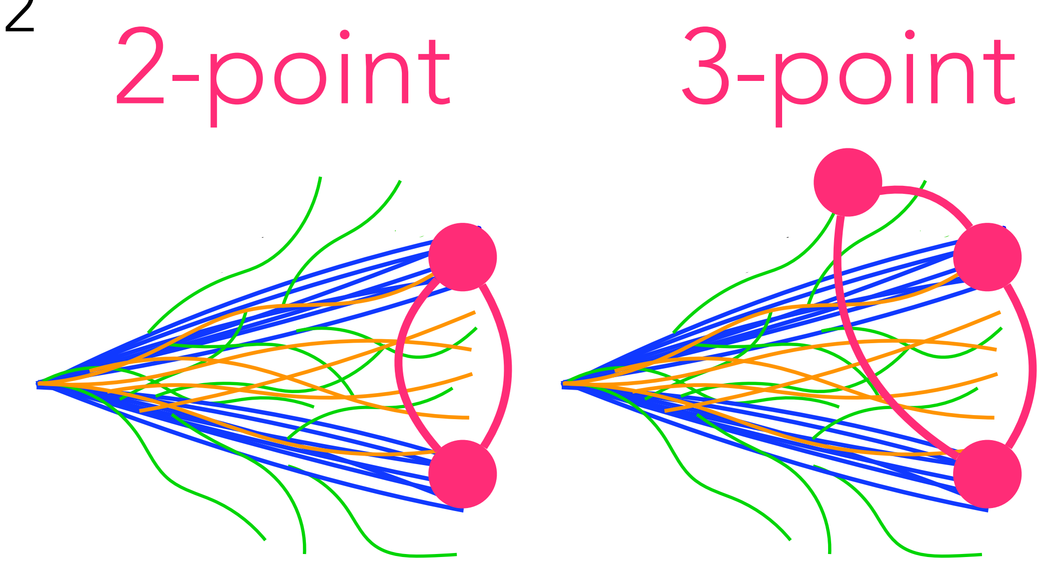
CMS: $z_{cut} = 0.1$, $\beta = 0$

arXiv:1609.07483

JET SUBSTRUCTURE

- How many "prongs" are in the jet?
- Generalized energy correlation functions are sensitive to N-point correlations within a jet
  - A two-pronged jet has ${}_2e_3 < ({}_1e_2)^2$
  - Taking a ratio gives N^1_2

$$N_2^\beta = \frac{{}_2e_3^\beta}{({}_1e_2^\beta)^2}$$



$${}_1e_2^\beta = \sum_{1\le i<j\le n_J} z_i z_j \Delta R_{ij}^\beta$$

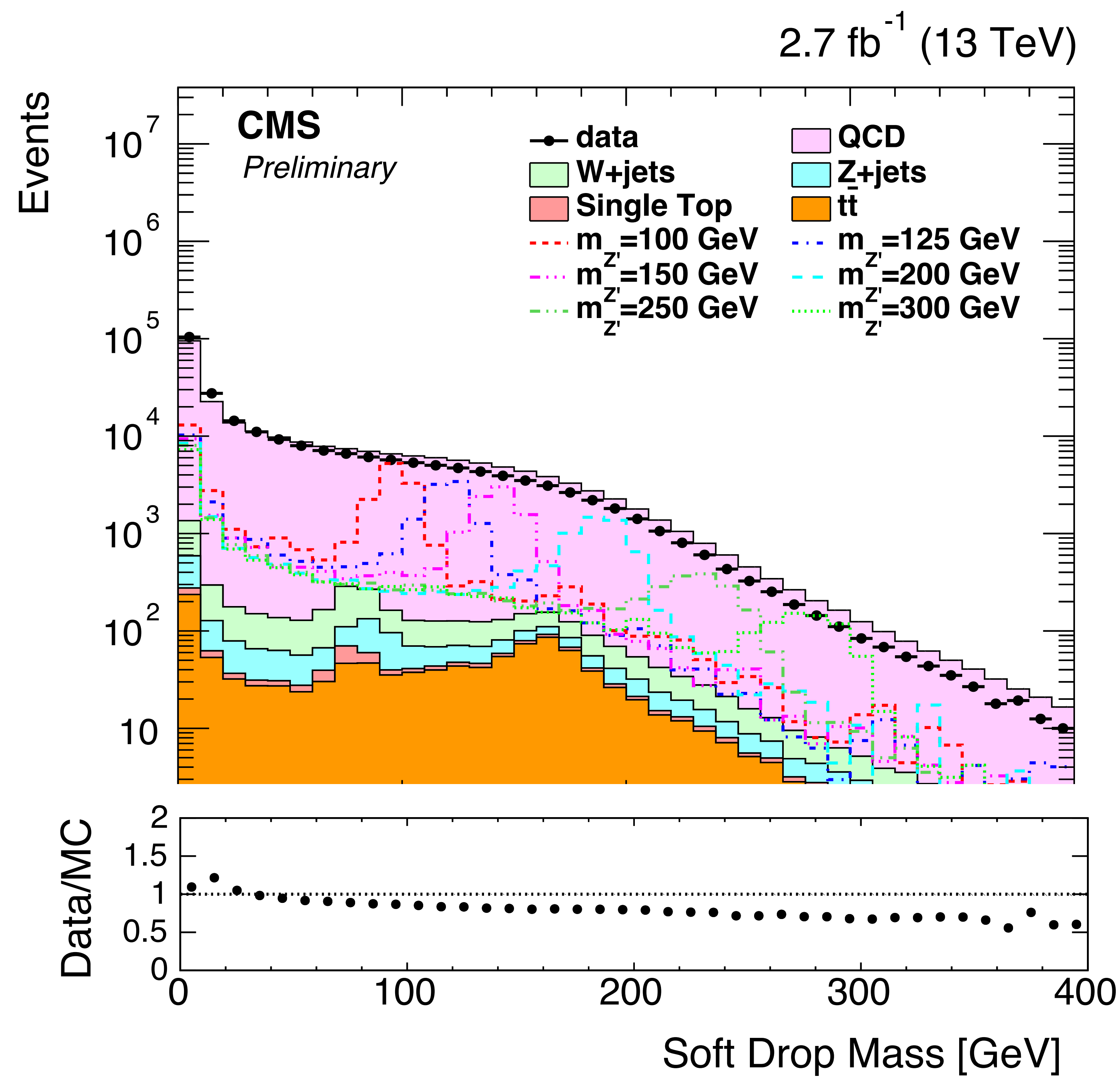
$${}_2e_3^\beta = \sum_{1\le i<j<k\le n_J} z_i z_j z_k \min\{\Delta R_{ij}^\beta \Delta R_{ik}^\beta, \Delta R_{ij}^\beta \Delta R_{jk}^\beta, \Delta R_{ik}^\beta \Delta R_{jk}^\beta\}$$

$\beta = 1$

EXO-17-001

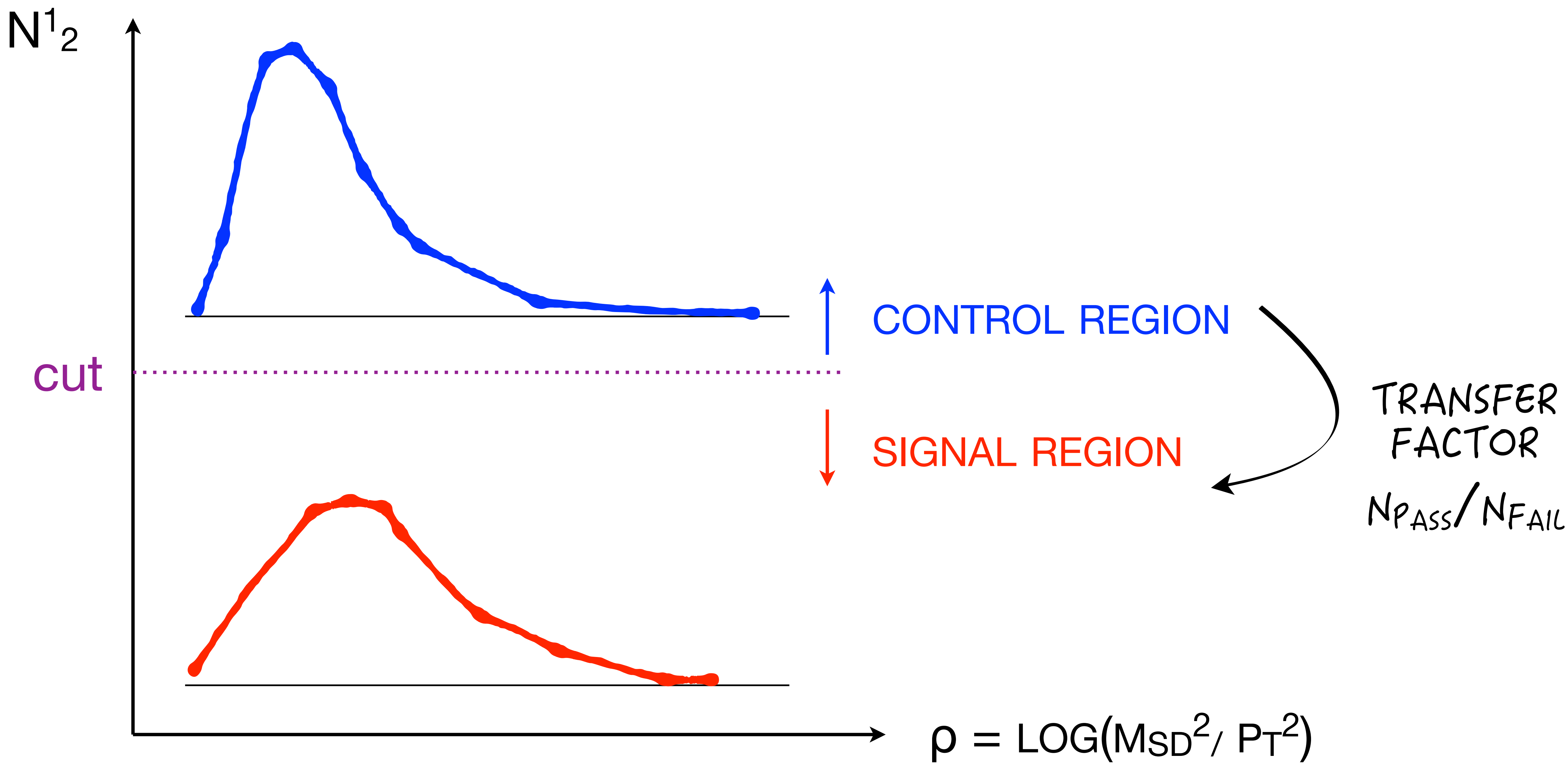
N^1_2 IN PRACTICE

- Here's what the boosted Z′+jet analysis looks like after kinematic selection
- Difficult to use the QCD Monte Carlo to predict the background in this phase space
- Fitting this mass distribution directly requires high order polynomial → large background uncertainties
- Can we try a data-driven sideband prediction?



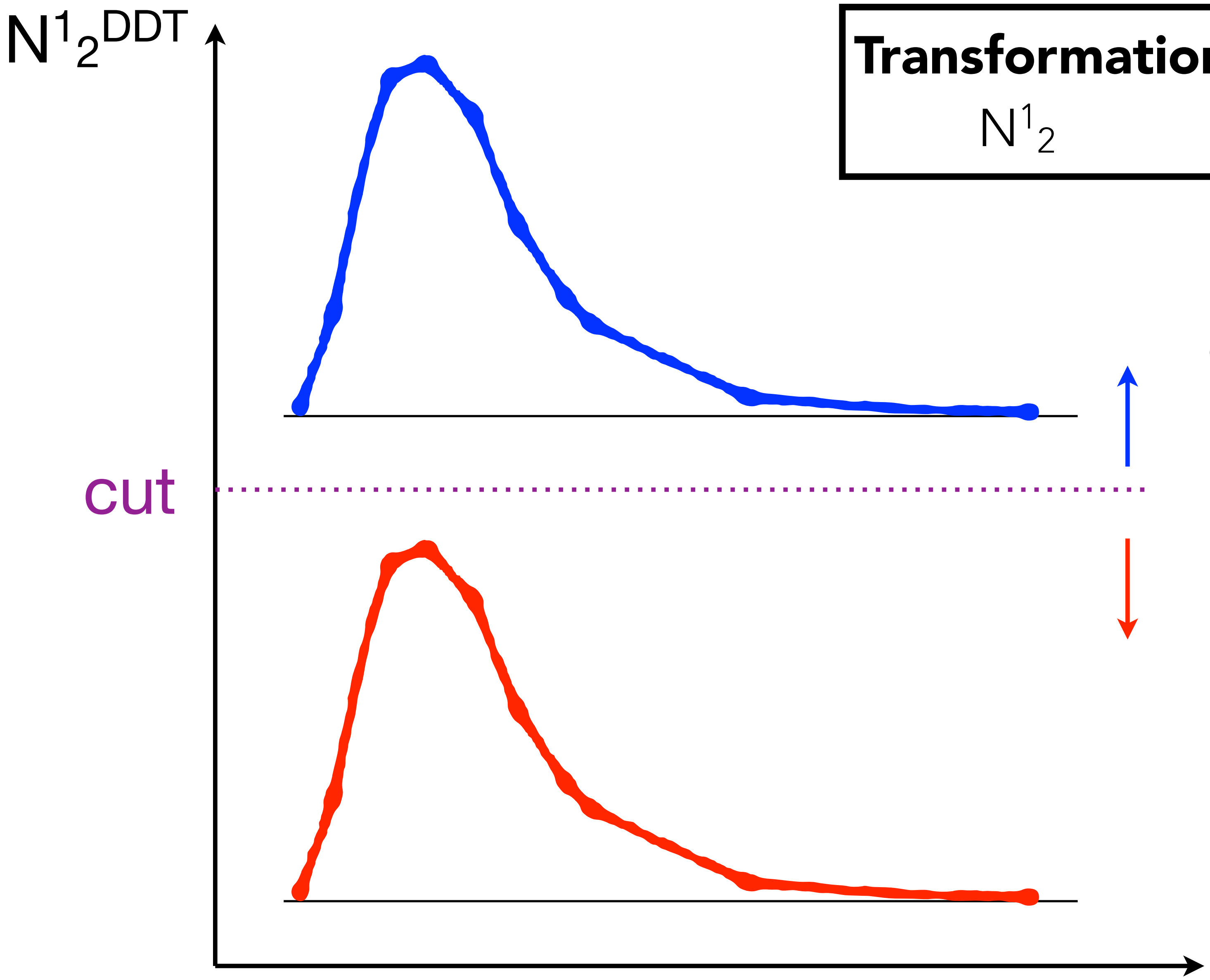
SIDEBAND QCD PREDICTION

- Core idea: predict QCD jet mass distribution from region failing the tagger
- Possible problem: does tagger sculpt jet mass distribution?



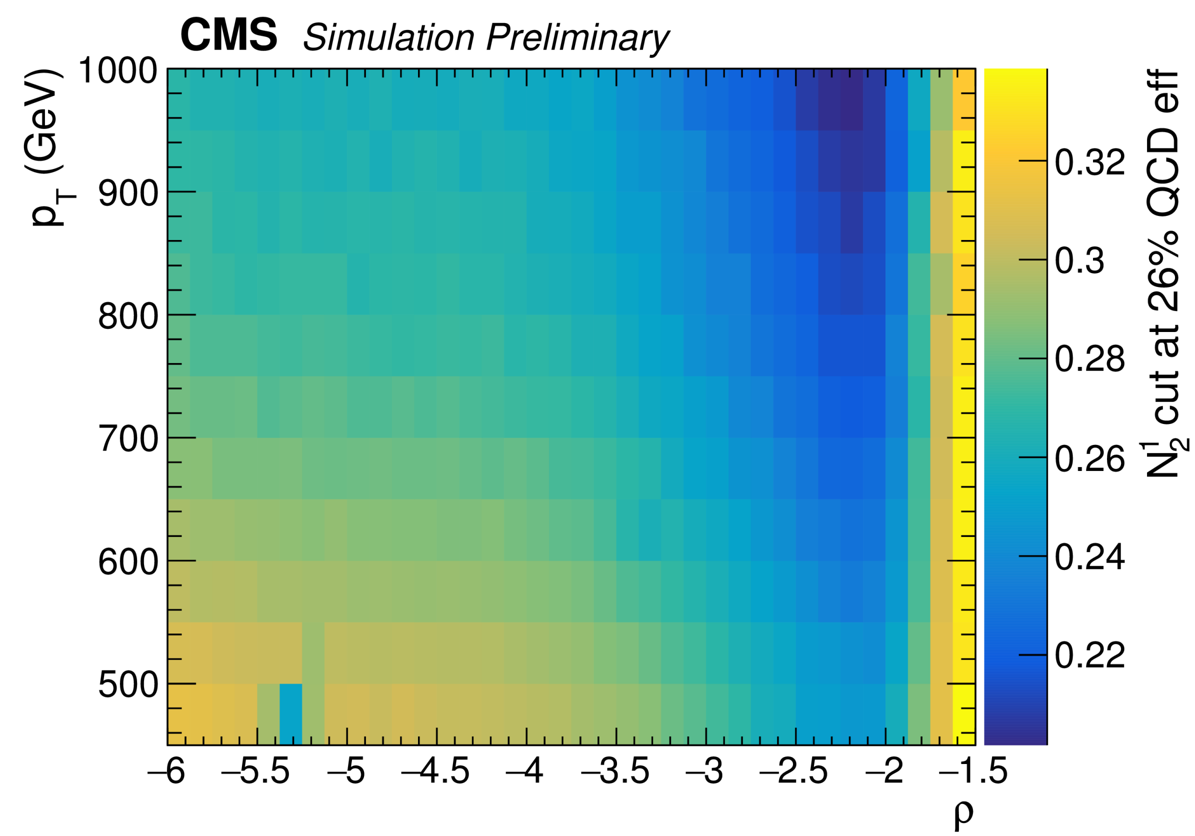
SIDEBAND QCD PREDICTION

- Solution: define new substructure variable intended to be decorrelated from jet mass



Transformation:

$N^1_2 \rightarrow N^1_2{}^{DDT} = N^1_2 - N^1_{2\,(26\%\ \text{quantile})}$

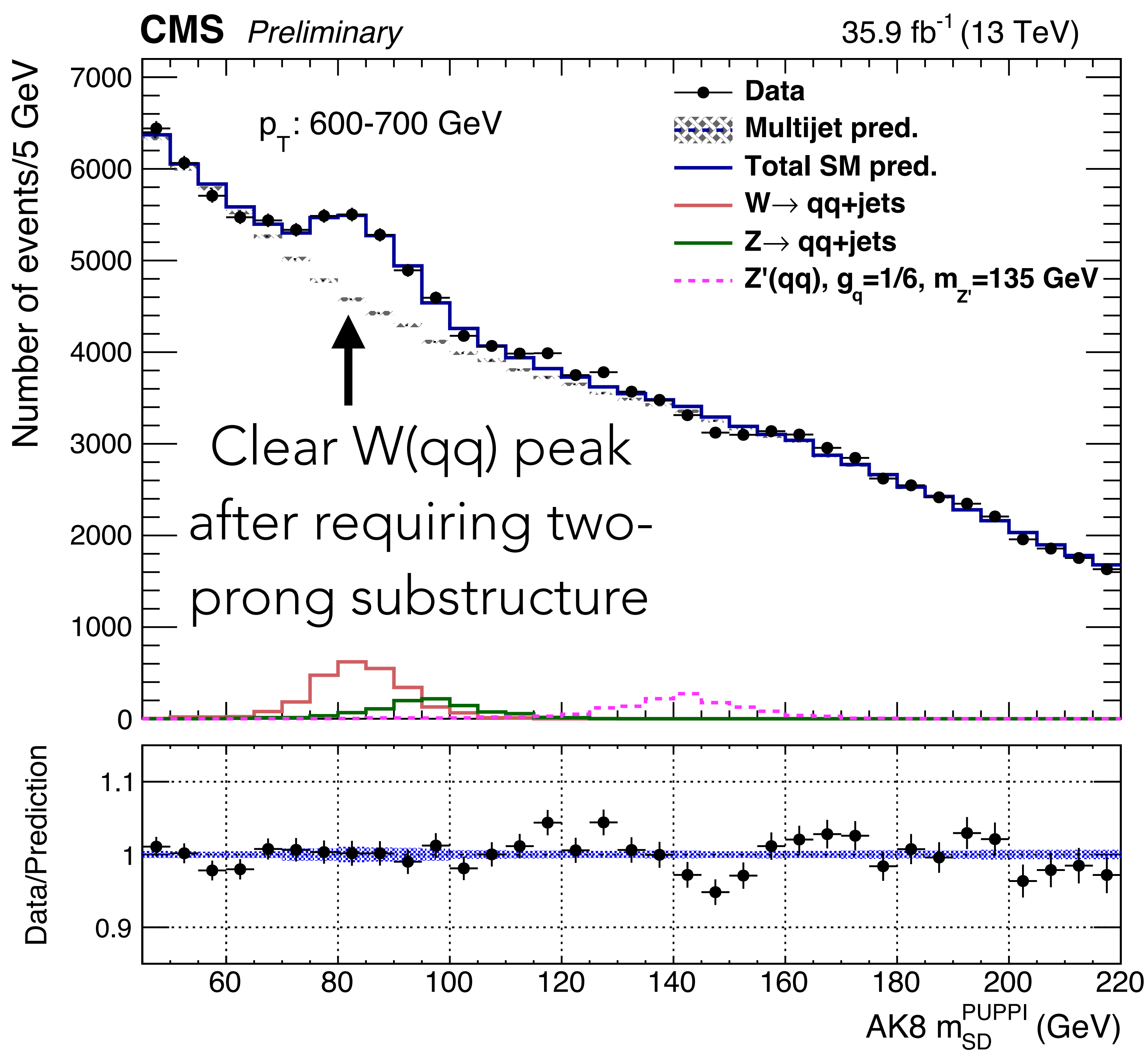


ρ (MASS-LIKE VARIABLE)

EXO-17-001

Z′ RESULTS

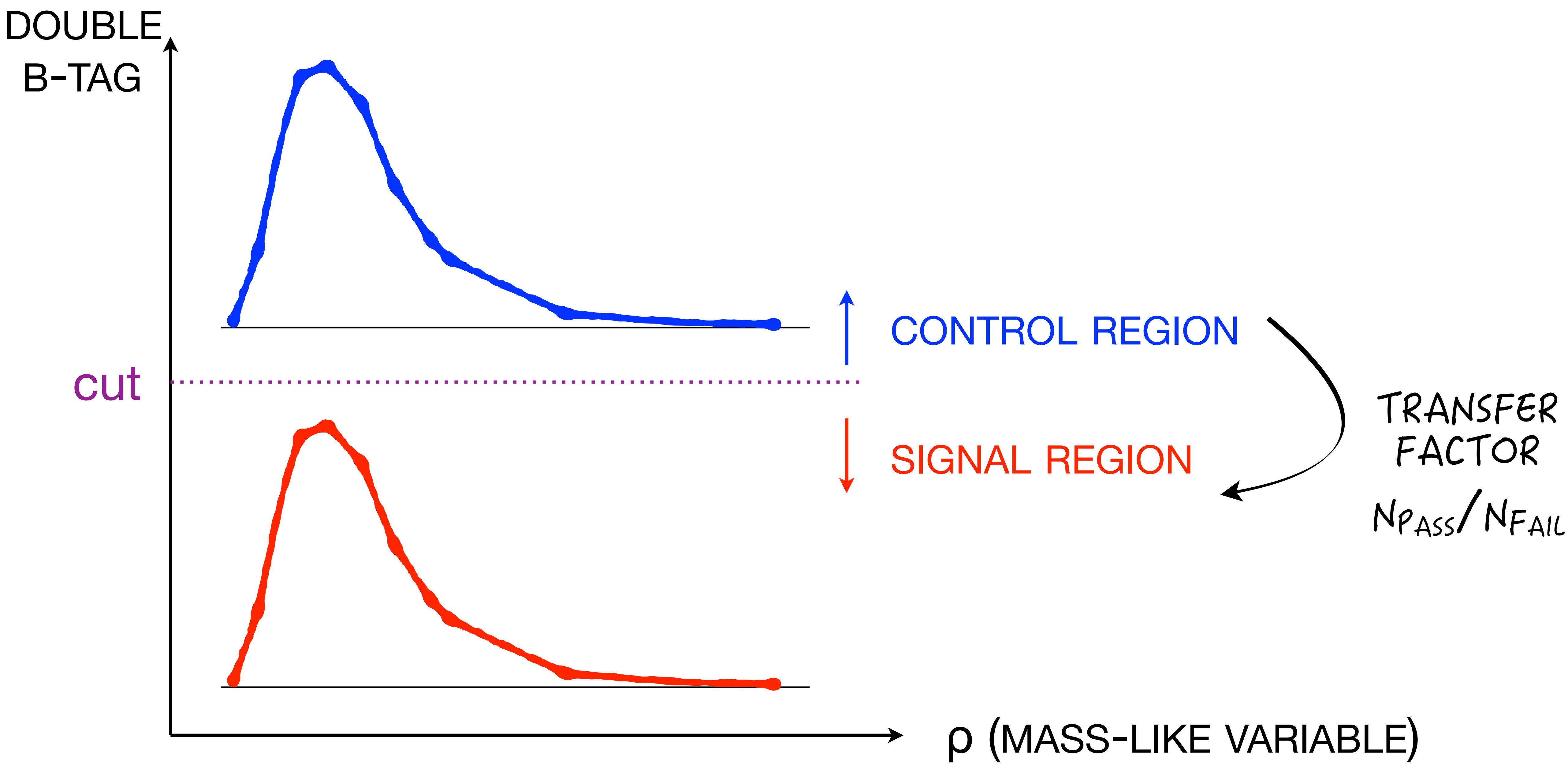
- Jet mass distribution is fit down to 40 GeV



background prediction from failing region allows good precision

SIDEBAND QCD PREDICTION (REDUX)

- Can we use the same QCD prediction when using a double-b tagger?
  - Yes — if it's sufficiently decorrelated from jet mass and p_T



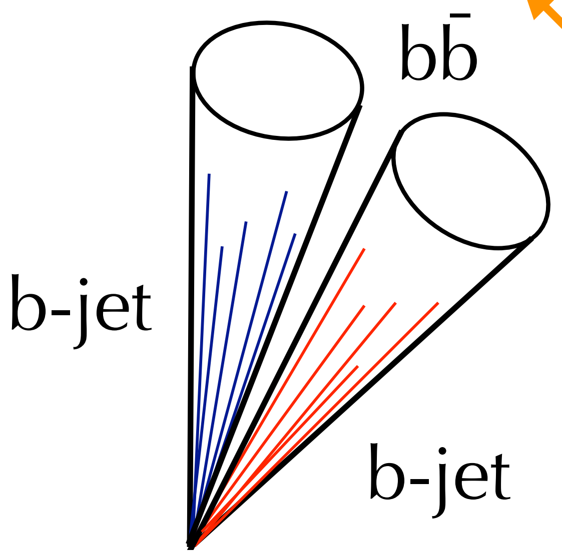
BOOSTED H(BB)

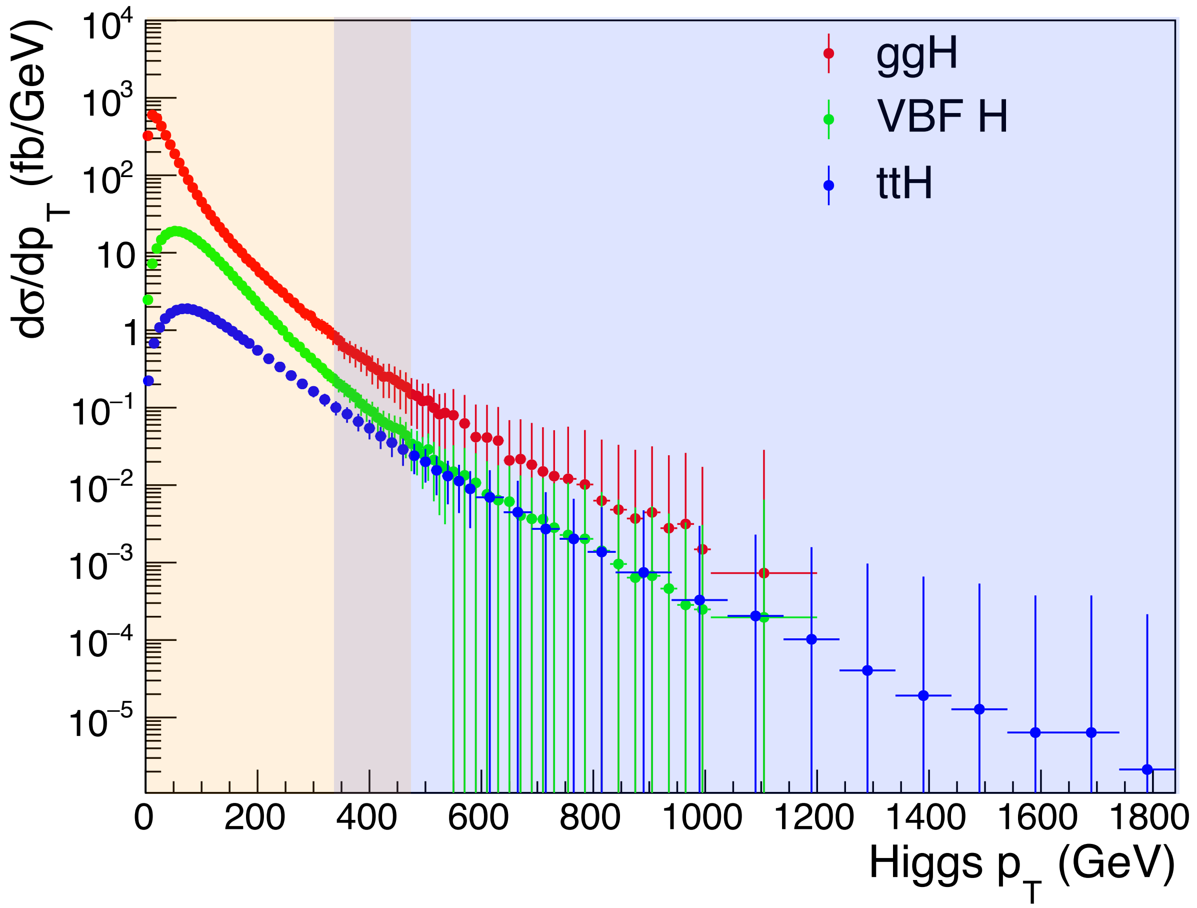
- With large boost, both b quarks merge into a single large radius jets
- How can we best exploit the presence of the b-quarks in the jet in a tagger?

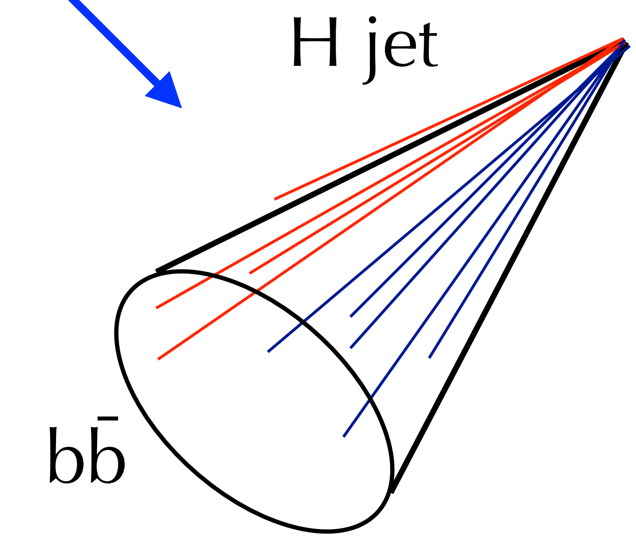
$\Delta R(b\bar{b}) \sim 2m_H/p_T$

two separated b-jets (R=0.4)

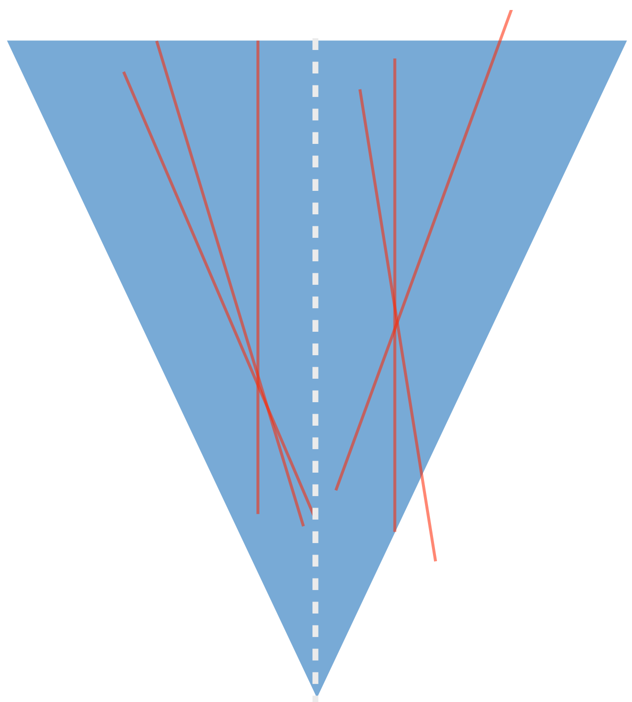
one merged double b-jet (R=0.8)





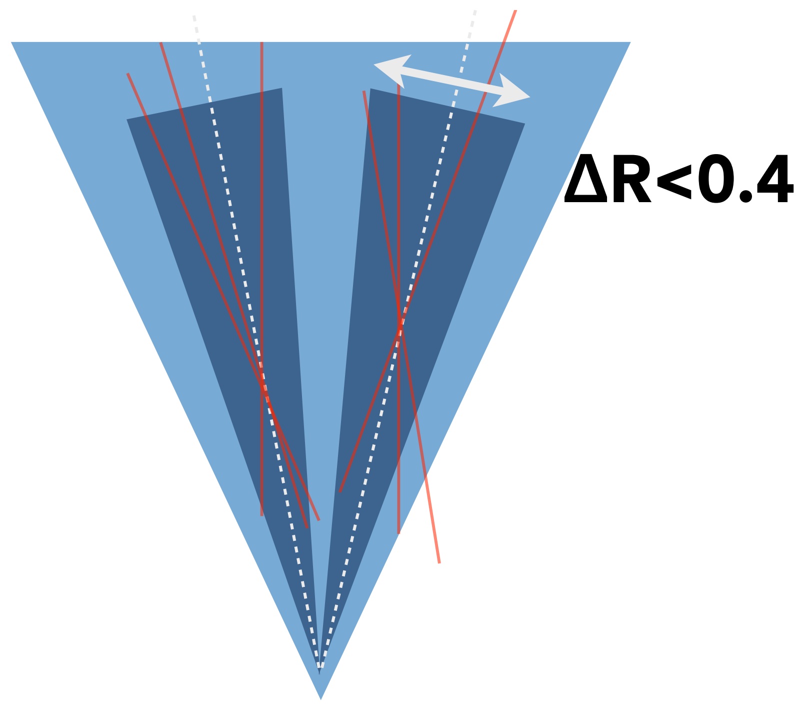


MULTIPLE APPROACHES



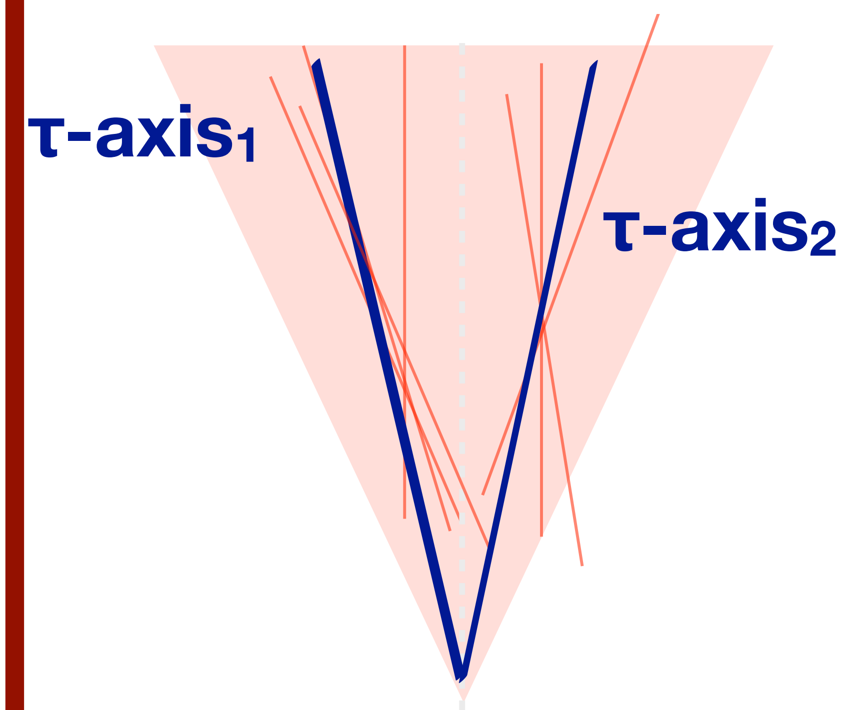
fatjet

- Based on standard b-tagging algorithm
- Not designed for two b's in the same jet



subjets

- Defines sub-jets
- Standard b-tagging applied to each subject

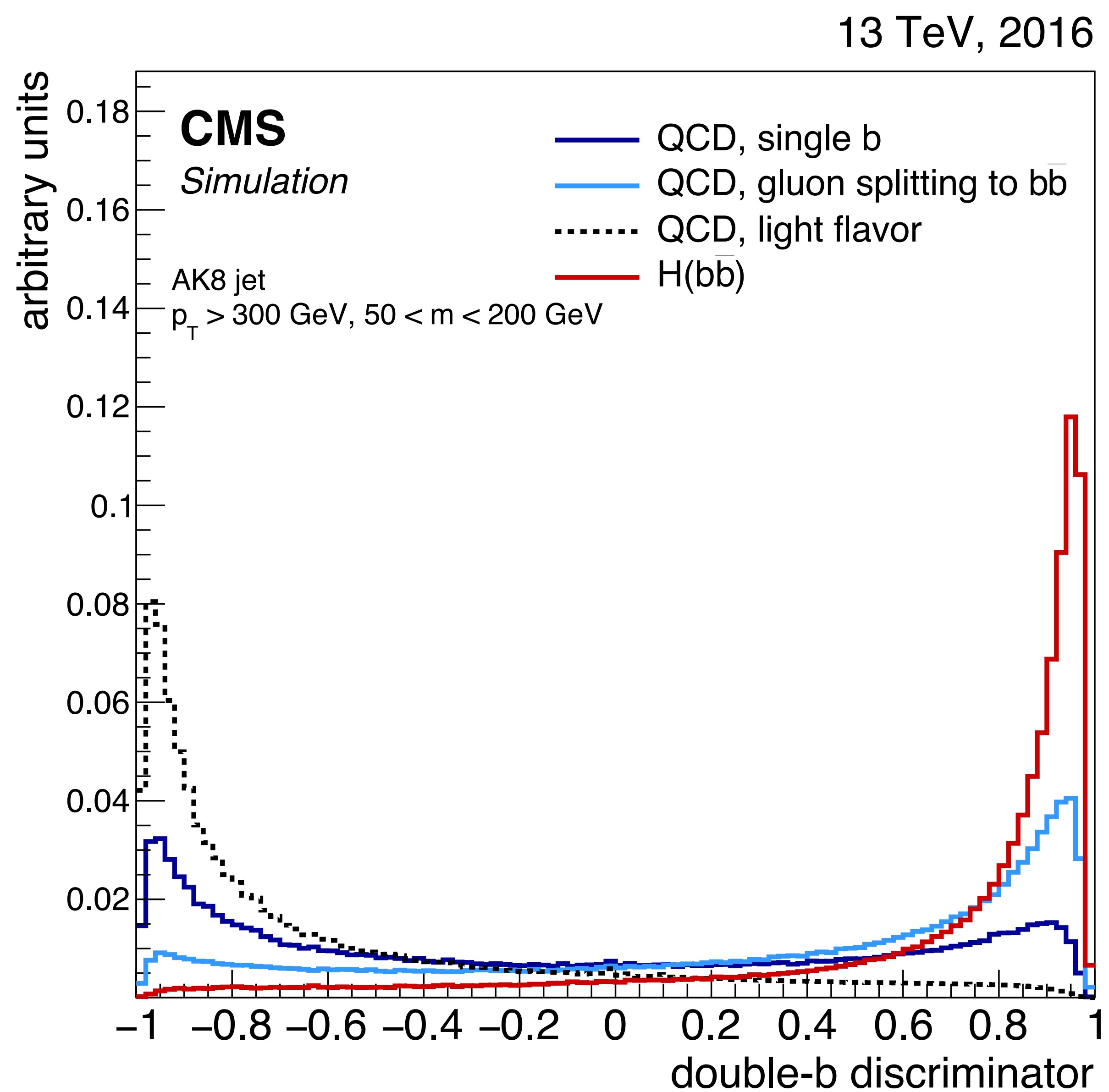


double-b

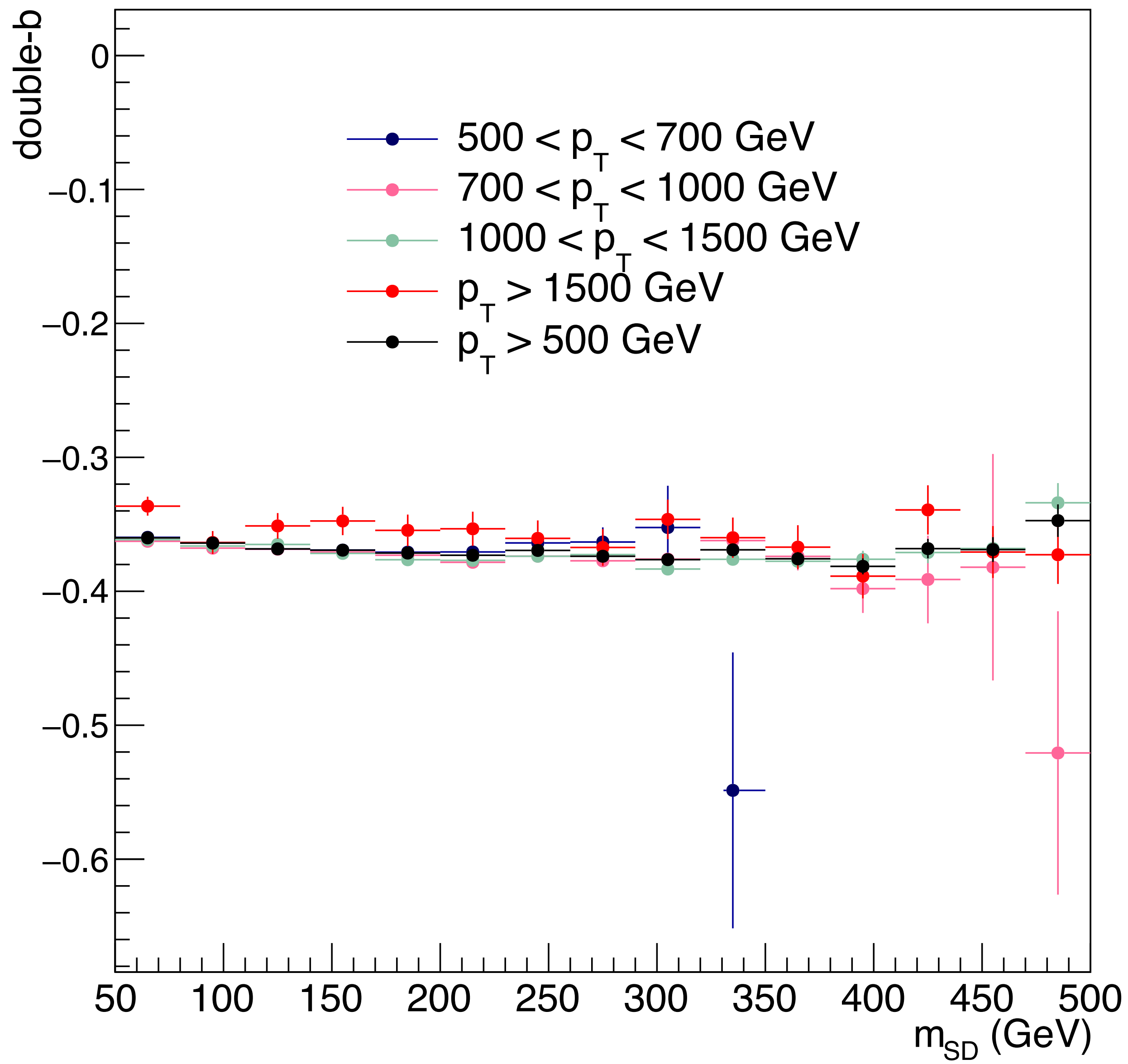
- Identifies two b hadron decay chains in the same fat jet
- Does not define subjects, but uses N-subjettiness axes

DOUBLE B-TAGGER

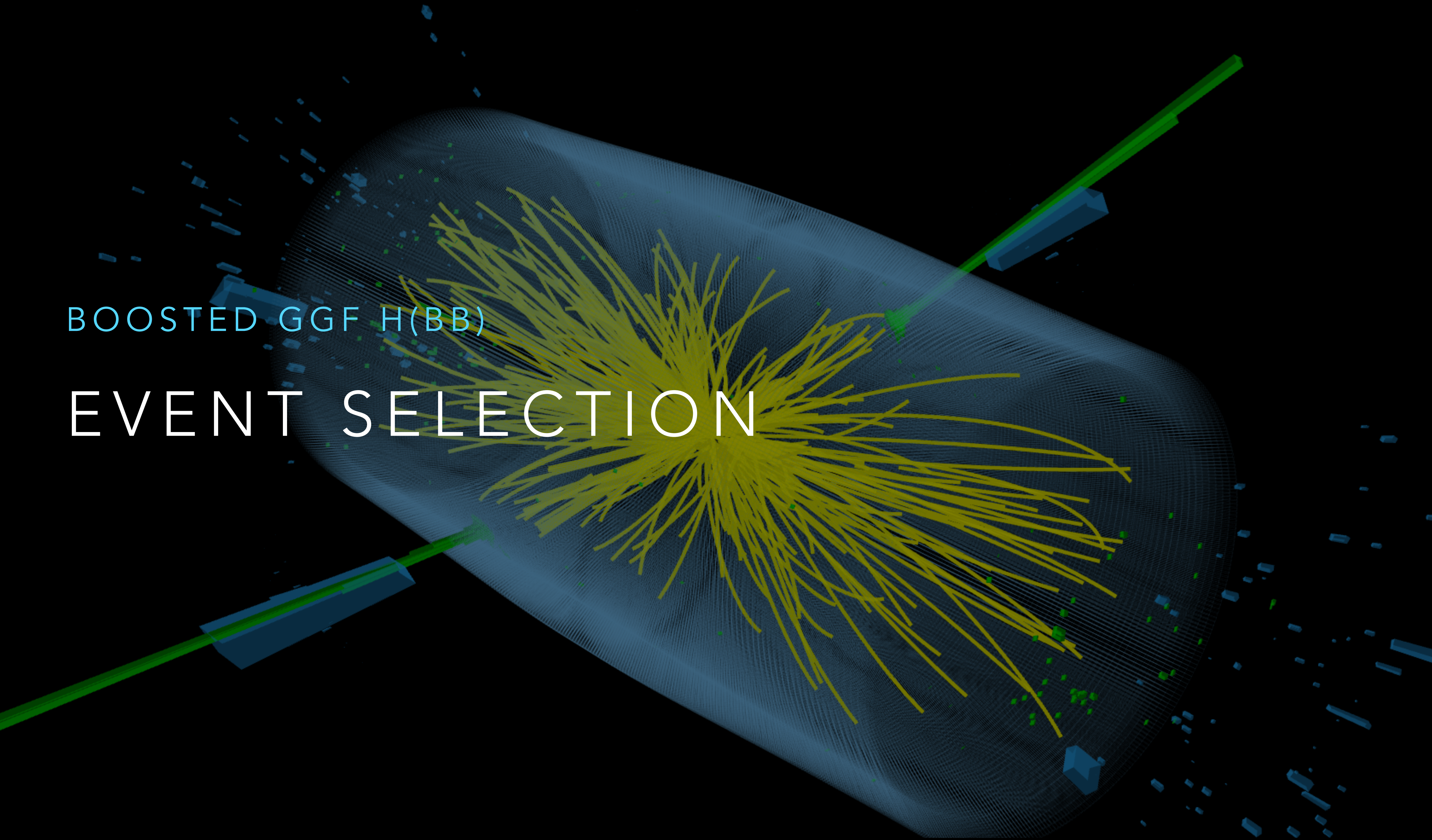
- Combines tracking and vertexing information in a multivariate classifier with 27 observables
- Targets the bb signal with additional aims:
  - jet mass and p_T independent
  - cover a very wide p_T range
  - inputs are chosen to avoid p_T correlation
    - e.g. no ΔR-like variables, no substructure info



CORRELATIONS?



- No strong correlations in double-b tagger versus m_{SD} or p_T in QCD background



BOOSTED GGF H(BB)

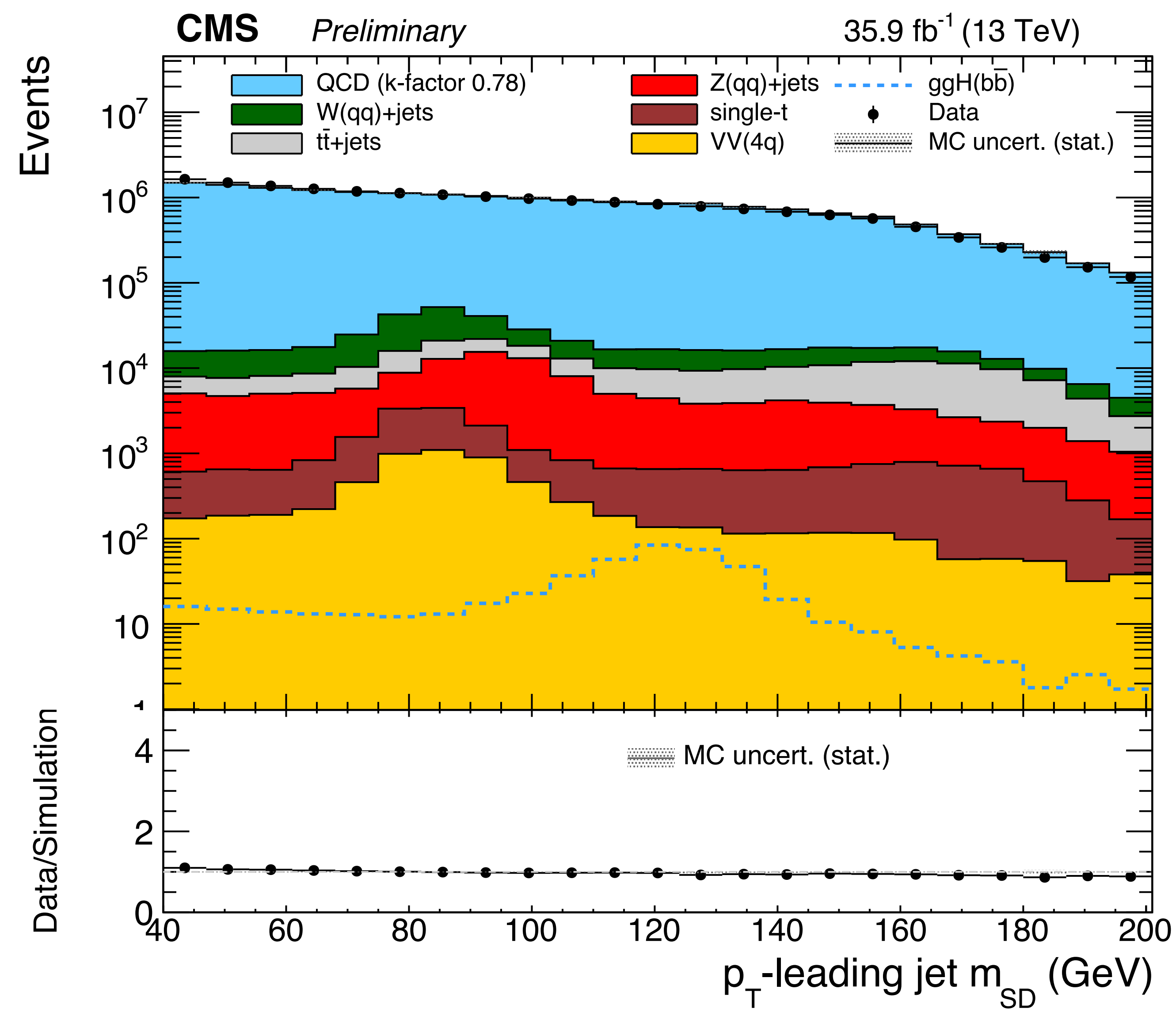
EVENT SELECTION

Javier Duarte
Fermilab

HIG-17-010

EVENT SELECTION

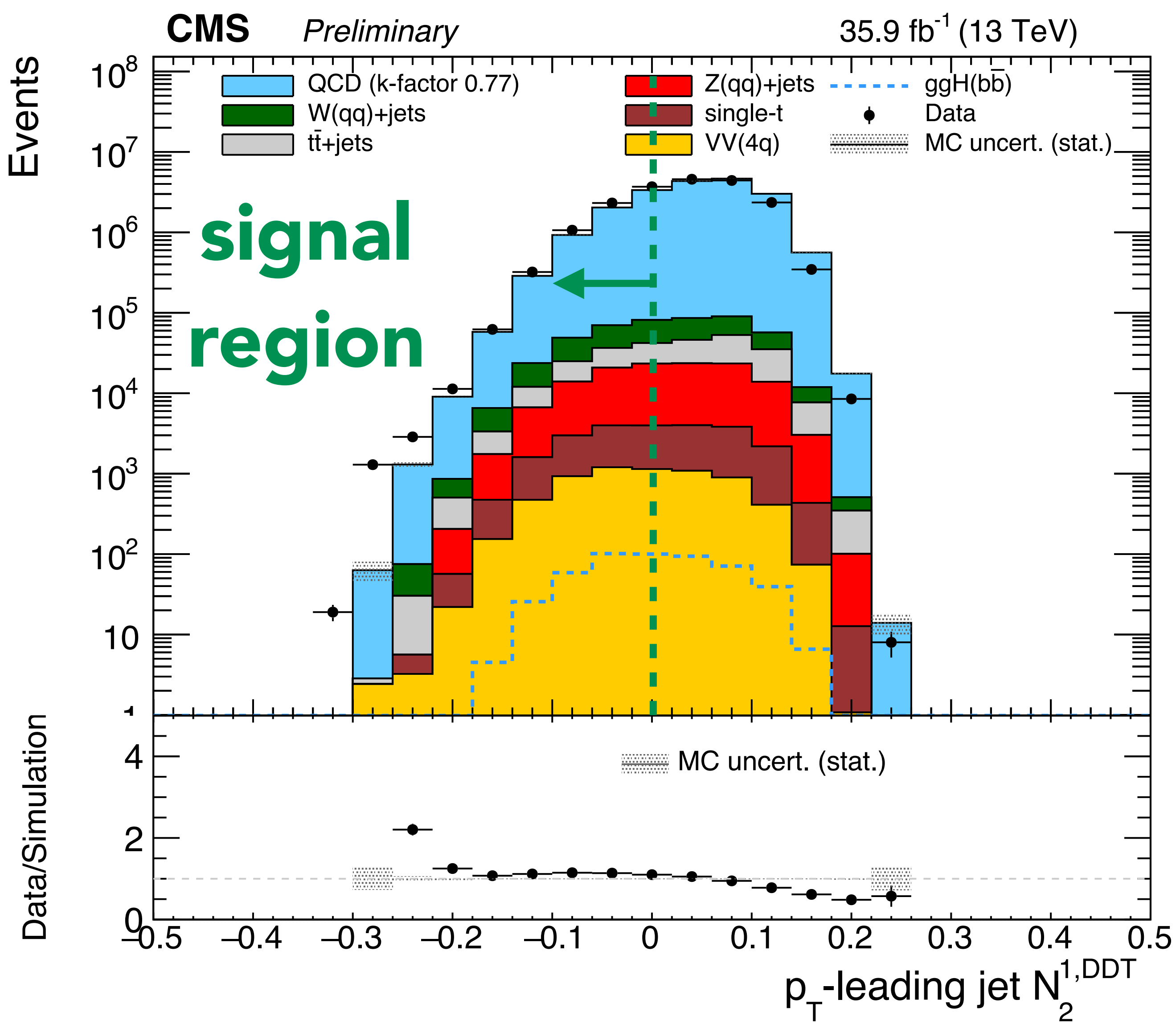
- Online selection asks for a high p_T single jet or large hadronic activities
  - $p_T > 360$ GeV ($m > 30$) or $\Sigma\, p_T > 900$ GeV
- Offline: Highest p_T jet
  - $p_T > 450$ GeV, $|\eta| < 2.5$
  - jet mass $m_{SD} > 40$ GeV
- lepton veto, p_T^{miss} veto
- $-6.0 < \rho = \log(m_{SD}^2 / p_T^2) < -2.1$

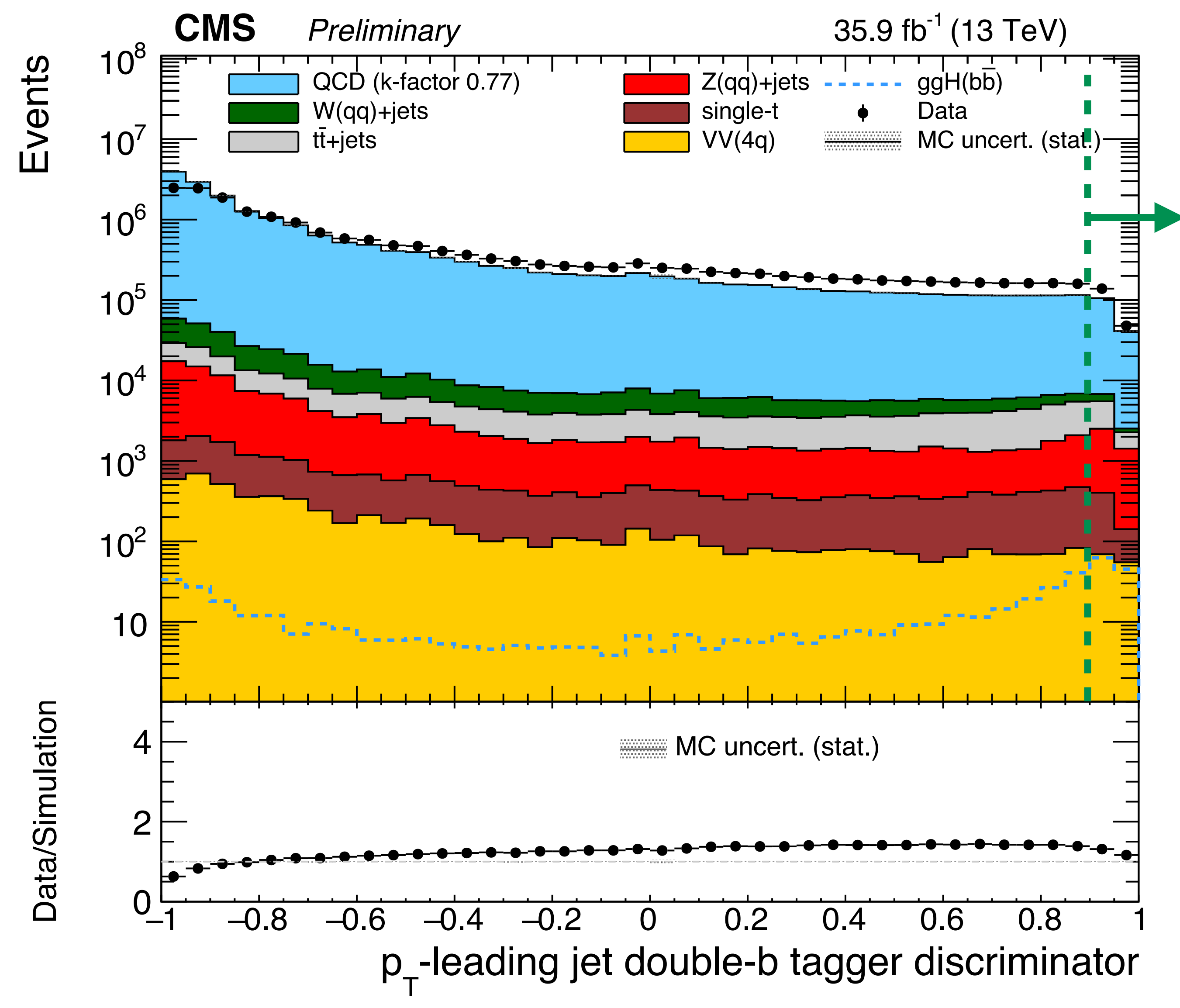


EVENT SELECTION

Substructure: two prong discrimination, ~50% sig. efficiency, 26% bkg. efficiency

Double-b tagger: ~30% sig. efficiency, 1% bkg. efficiency (tight working point)

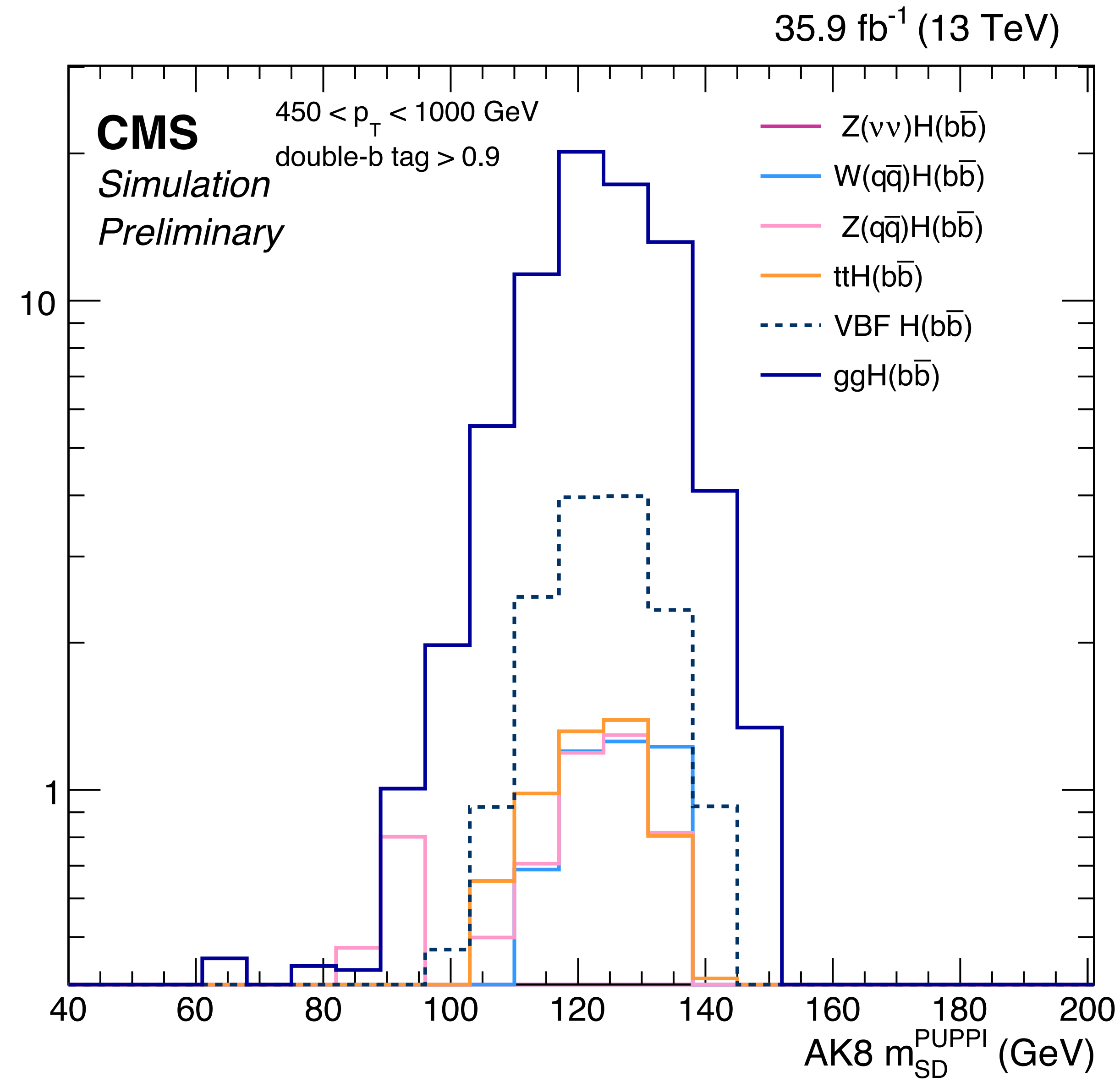




Javier Duarte
Fermilab

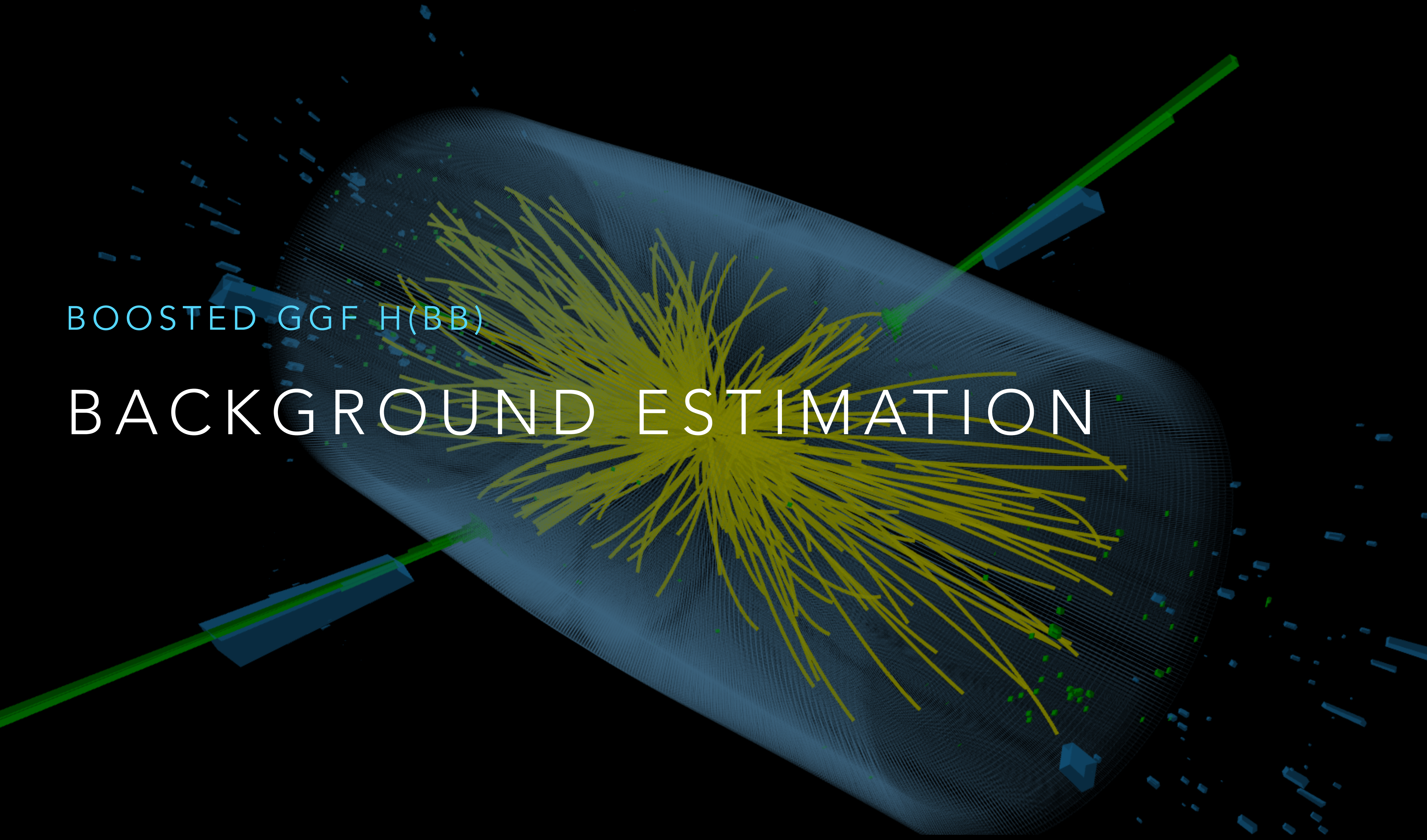
SIGNAL COMPOSITION

- Analysis is inclusive in Higgs production mode
- Dominant contribution is ggF (74%)
    - 12% VBF
    - 8% VH
    - 6% ttH



BOOSTED GGF H(BB)

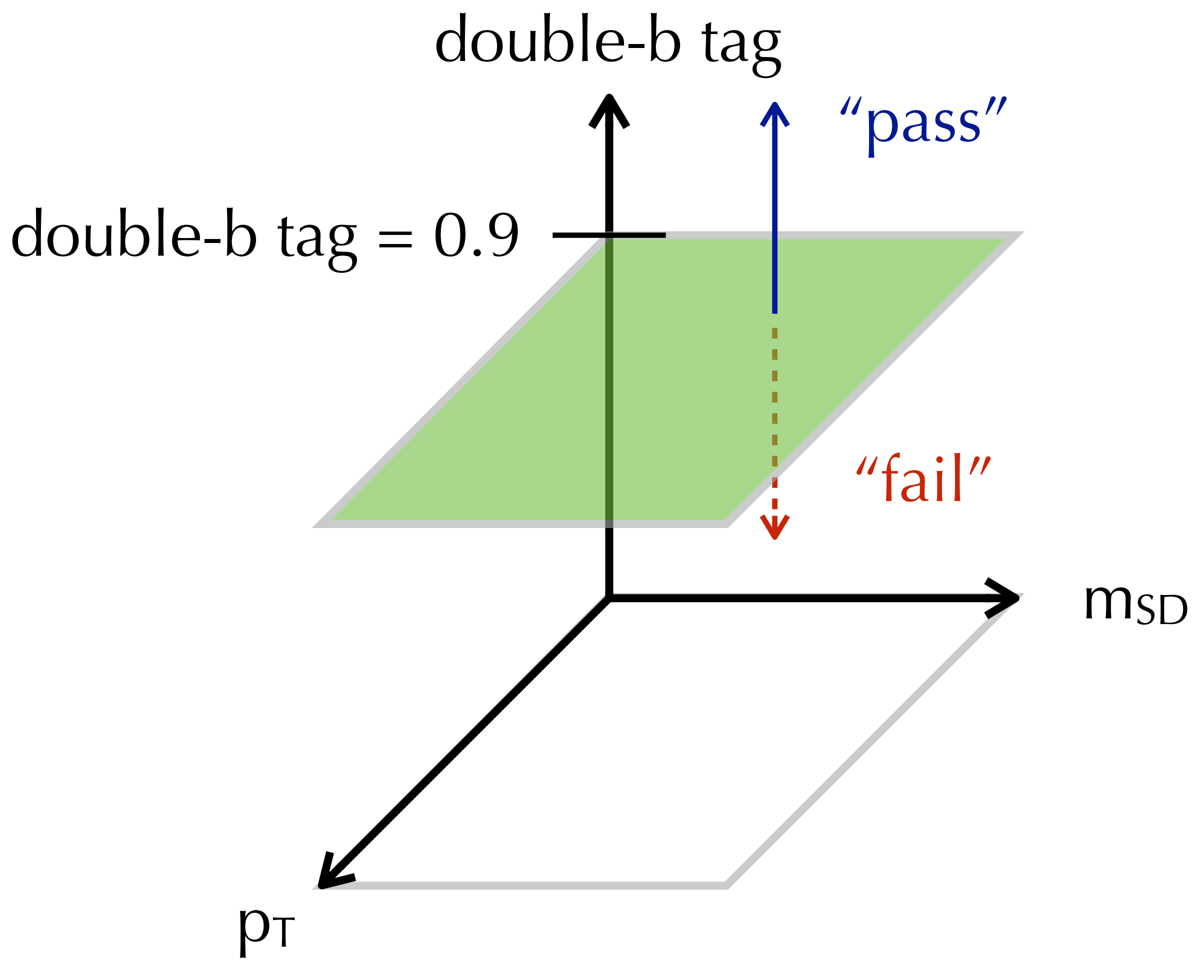
BACKGROUND ESTIMATION



Javier Duarte
Fermilab

BACKGROUND STRATEGY

- Backgrounds estimated from data
  - QCD (90%): from failing double b-tag x transfer factor
  - tt+jets (3%): from 1μ control region
- Backgrounds estimated from MC including NLO QCD + EWK corrections and jet mass, resolution, and substructure tagging scale factors
  - W/Z+jets (5%)
  - single-t, VV (<1%)



QCD TRANSFER FACTOR

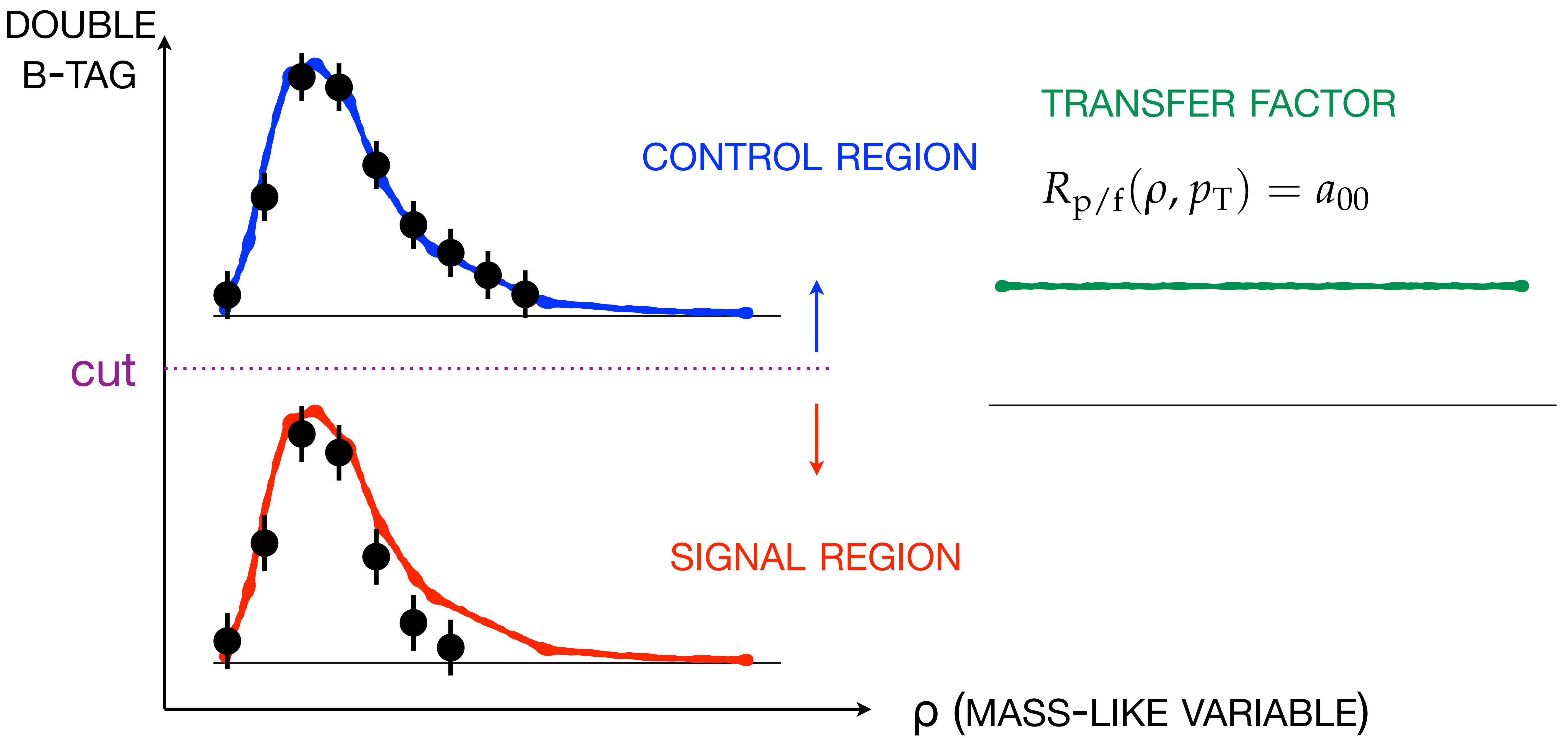
- If the double b-tagger were completely uncorrelated from jet mass and p_T, the transfer factor would be flat
- Taylor expand as a polynomial in ρ and p_T to parameterize any small correlations
- F-test determined 2nd order in ρ and 1st order in p_T is sufficient to fit the ratio

$$N_{\mathrm{pass}}^{\mathrm{QCD}}(m_{\mathrm{SD}}, p_{\mathrm{T}}) = R_{\mathrm{p/f}}(\rho, p_{\mathrm{T}}) \cdot N_{\mathrm{fail}}^{\mathrm{QCD}}(m_{\mathrm{SD}}, p_{\mathrm{T}})$$

$$N_{\mathrm{pass}}^{\mathrm{QCD}}(m_{\mathrm{SD}i}, p_{\mathrm{T}j}) = \boxed{\left(\sum_{k,\ell} a_{k\ell}\rho_{ij}^{k} {p_{\mathrm{T}}}_j^{\ell}\right)} \cdot N_{\mathrm{fail}}^{\mathrm{QCD}}(m_{\mathrm{SD}i}, p_{\mathrm{T}j})$$

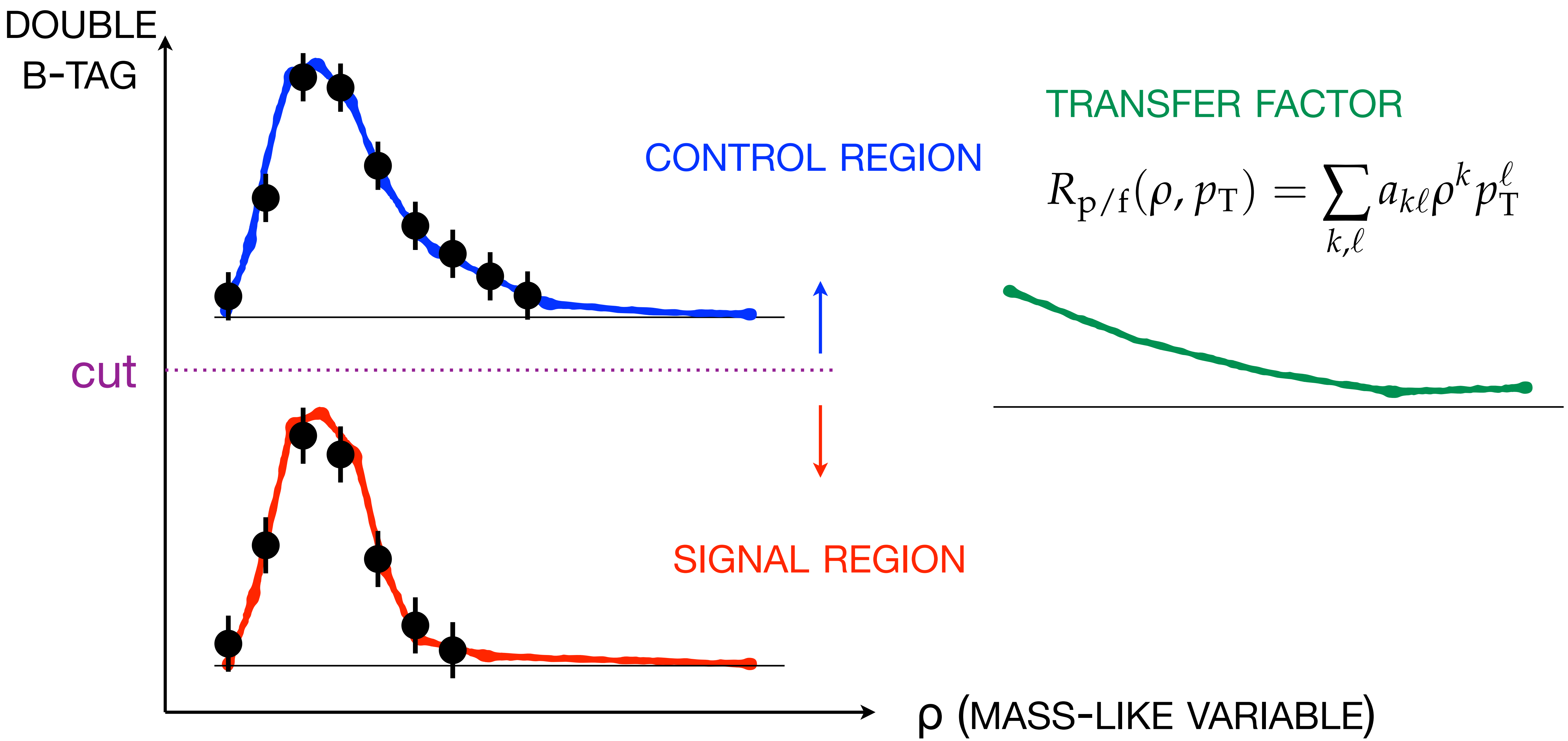
FITTING TRANSFER FACTOR

- Pre-fit both regions have the same predicted shape



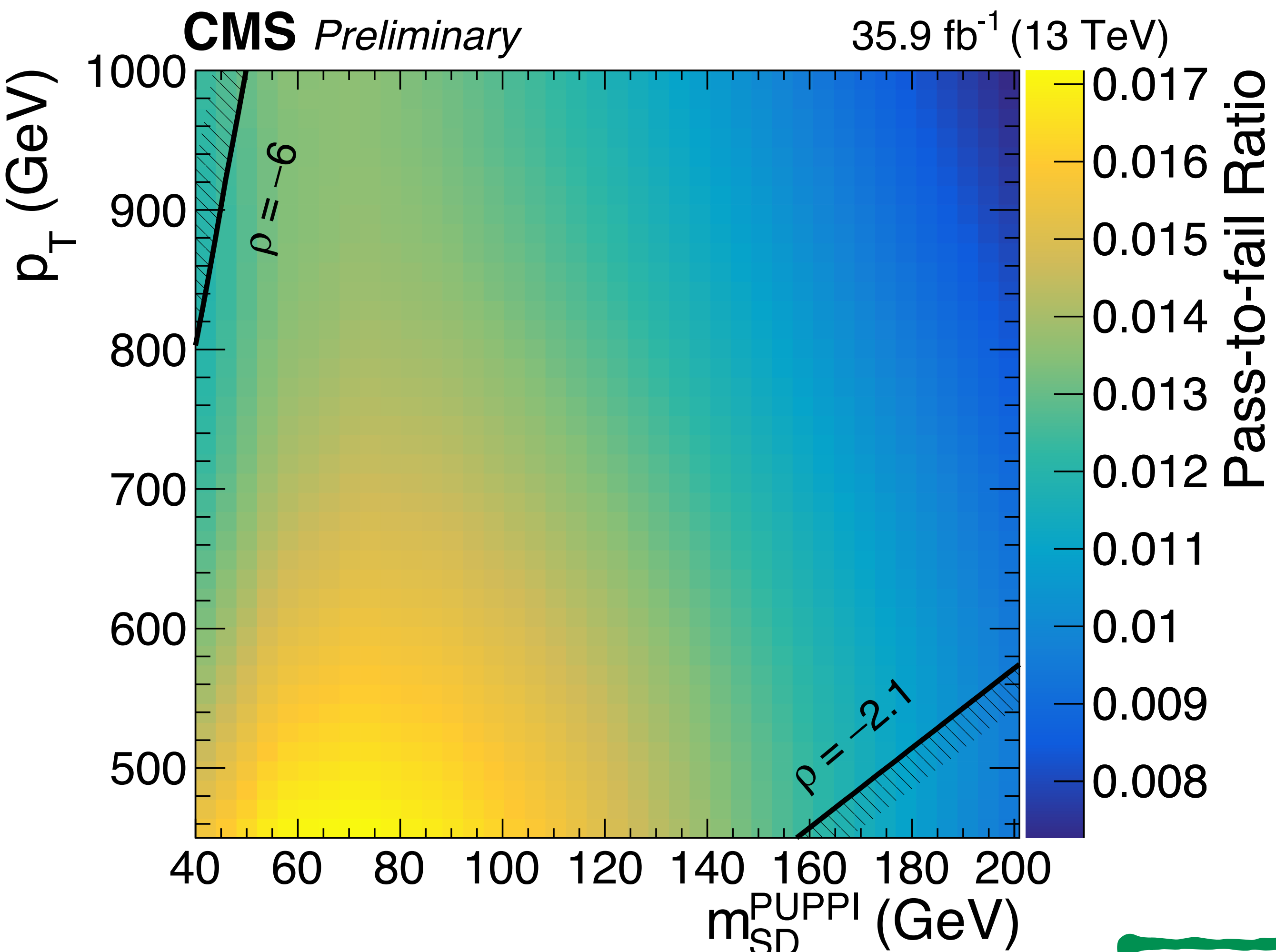
FITTING TRANSFER FACTOR

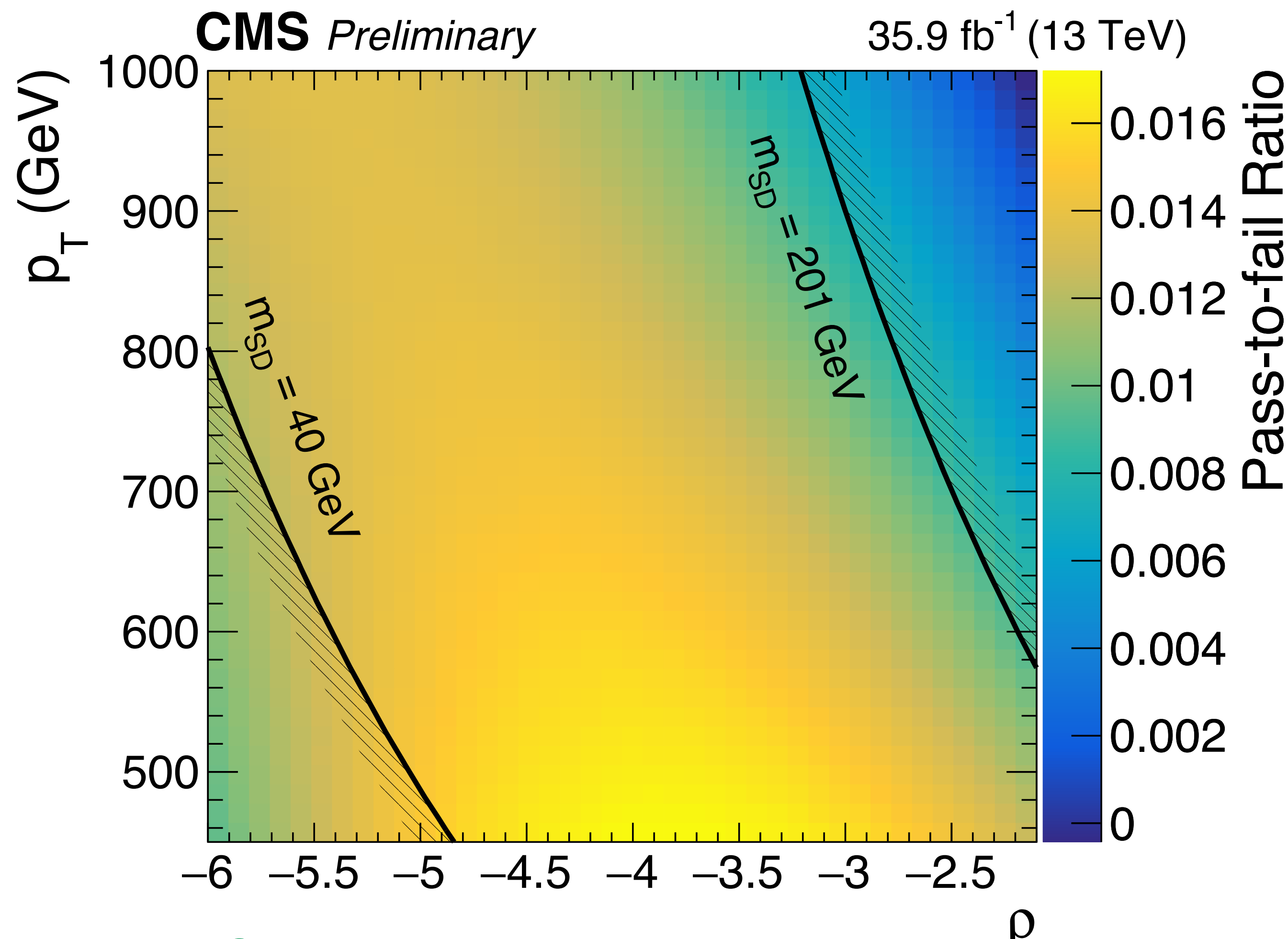
- Post-fit signal region has slightly different shape with the ratio given by the polynomial transfer factor



FINAL TRANSFER FACTOR

- Two views of the same transfer factor function

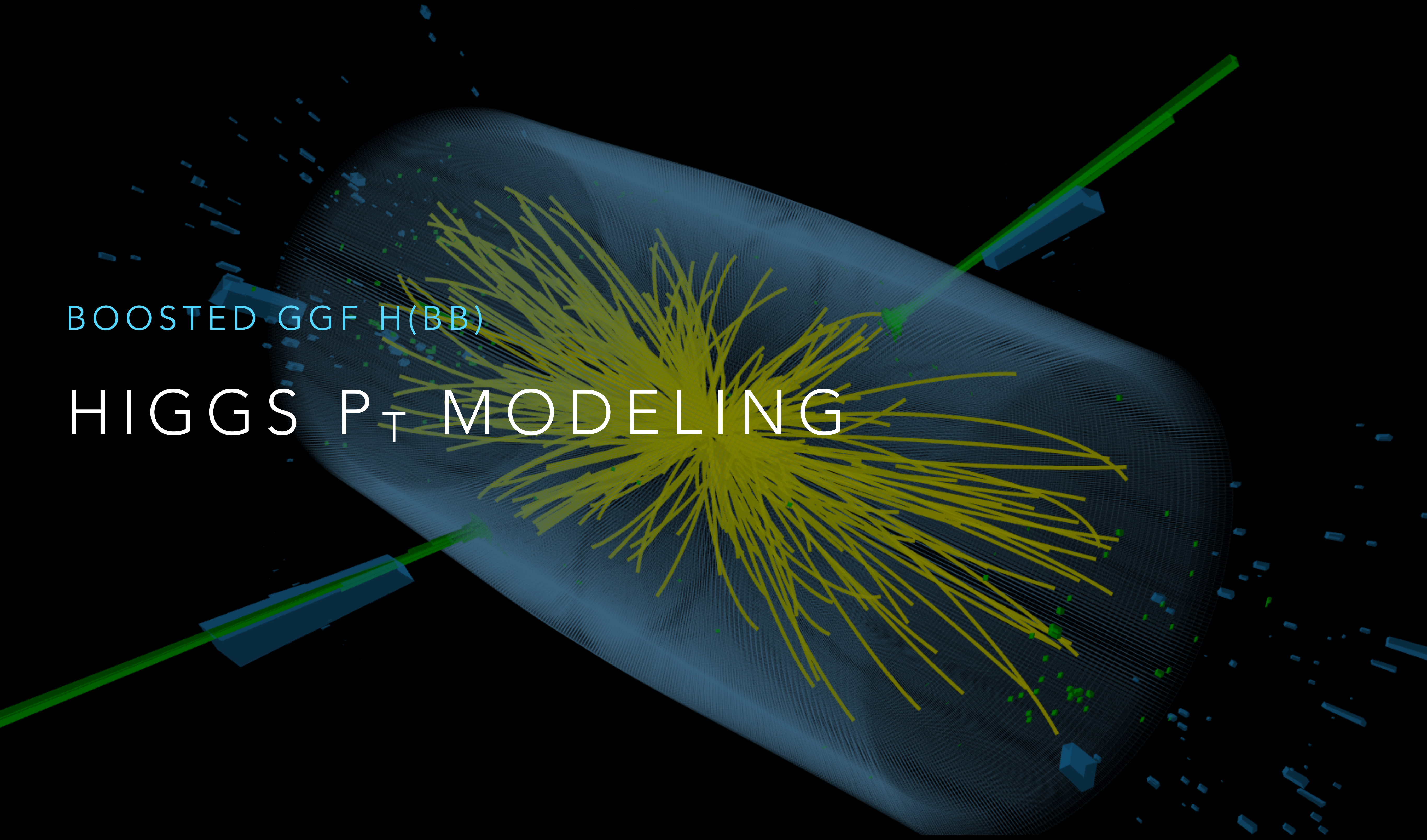




$$N_{\mathrm{pass}}^{\mathrm{QCD}}(m_{\mathrm{SD}i}, p_{\mathrm{T}j}) = \left(\sum_{k,\ell} a_{k\ell}\rho_{ij}^{k} {p_{\mathrm{T}}}_{j}^{\ell}\right) \cdot N_{\mathrm{fail}}^{\mathrm{QCD}}(m_{\mathrm{SD}i}, p_{\mathrm{T}j})$$

BOOSTED GGF H(BB)

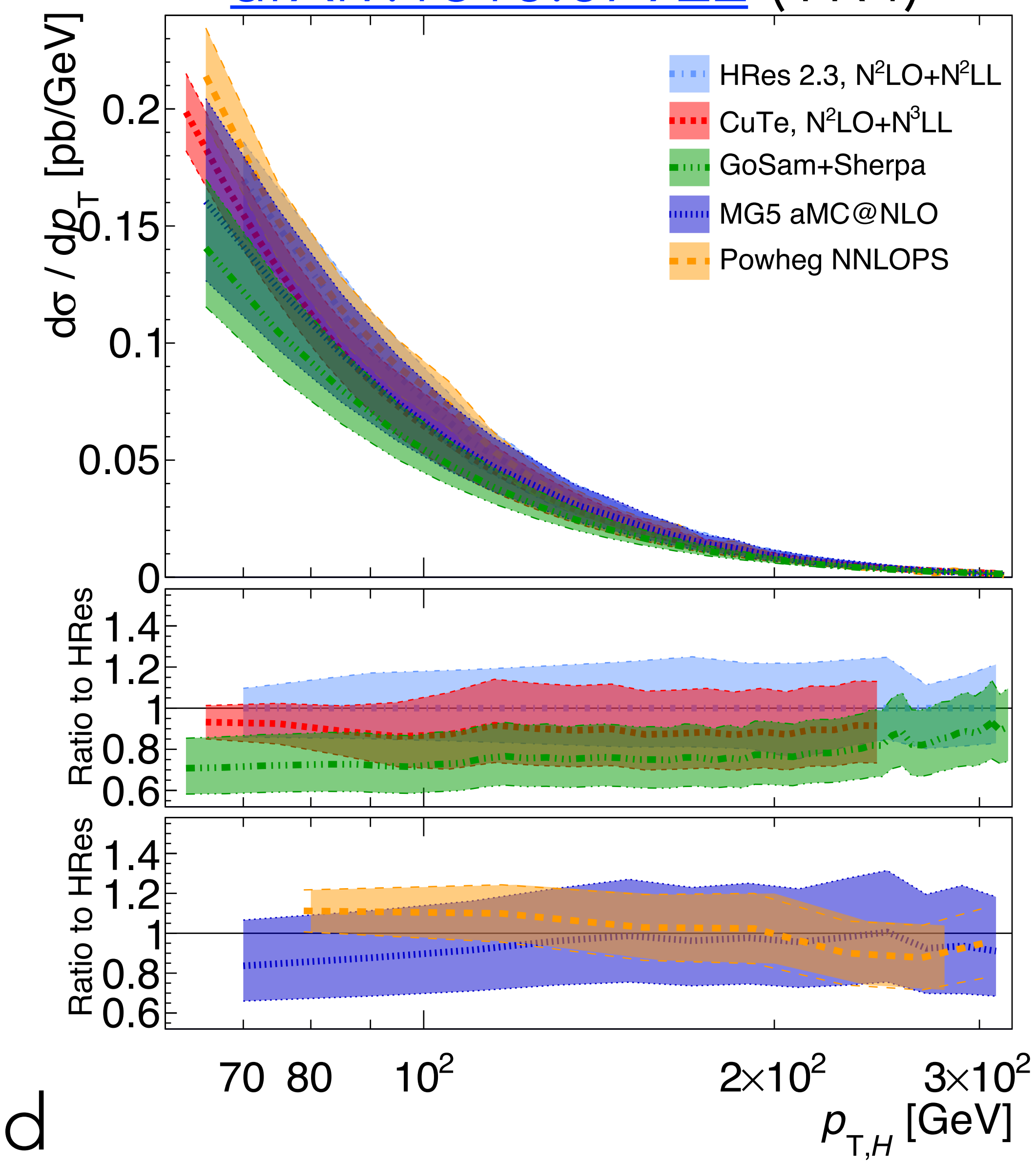
HIGGS P_T MODELING



HIGGS PT SPECTRUM

- Other CMS Higgs results use Powheg: 1 jet + $m_t = \infty$, arXiv:1111.2854
- We want to account for both effects of **higher order corrections** and **finite top mass**
  - No real NLO + finite top mass calculation available in the literature
-
-
-

arXiv:1610.07922 (YR4)

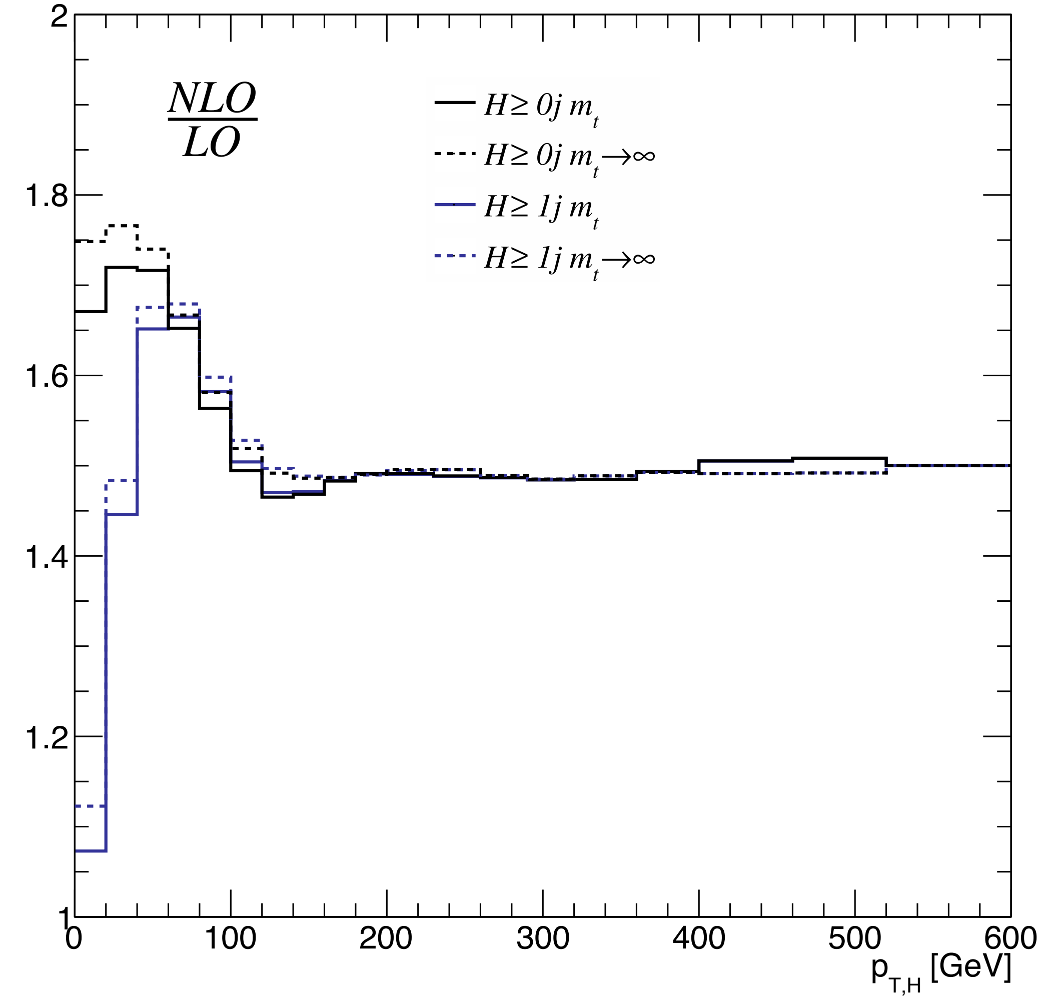


≪ 1 TeV

Disclaimer: we made an educated choice and assigned a reasonable uncertainty, but it's not the only possible choice. We also provide results with CMS standard Powheg sample

HIGGS PT SPECTRUM

- Other CMS Higgs results use Powheg: 1 jet + $m_t = \infty$, arXiv:1111.2854
- We want to account for effects of higher order corrections and finite top mass
  - No real NLO + finite top mass calculation available in the literature
- Adopt a factorized approach:
- LO H+0-2jet, finite m_t, p_t^H up to 600 GeV, including WW acceptance cuts arXiv:1410.5806 → We build on this
-
-

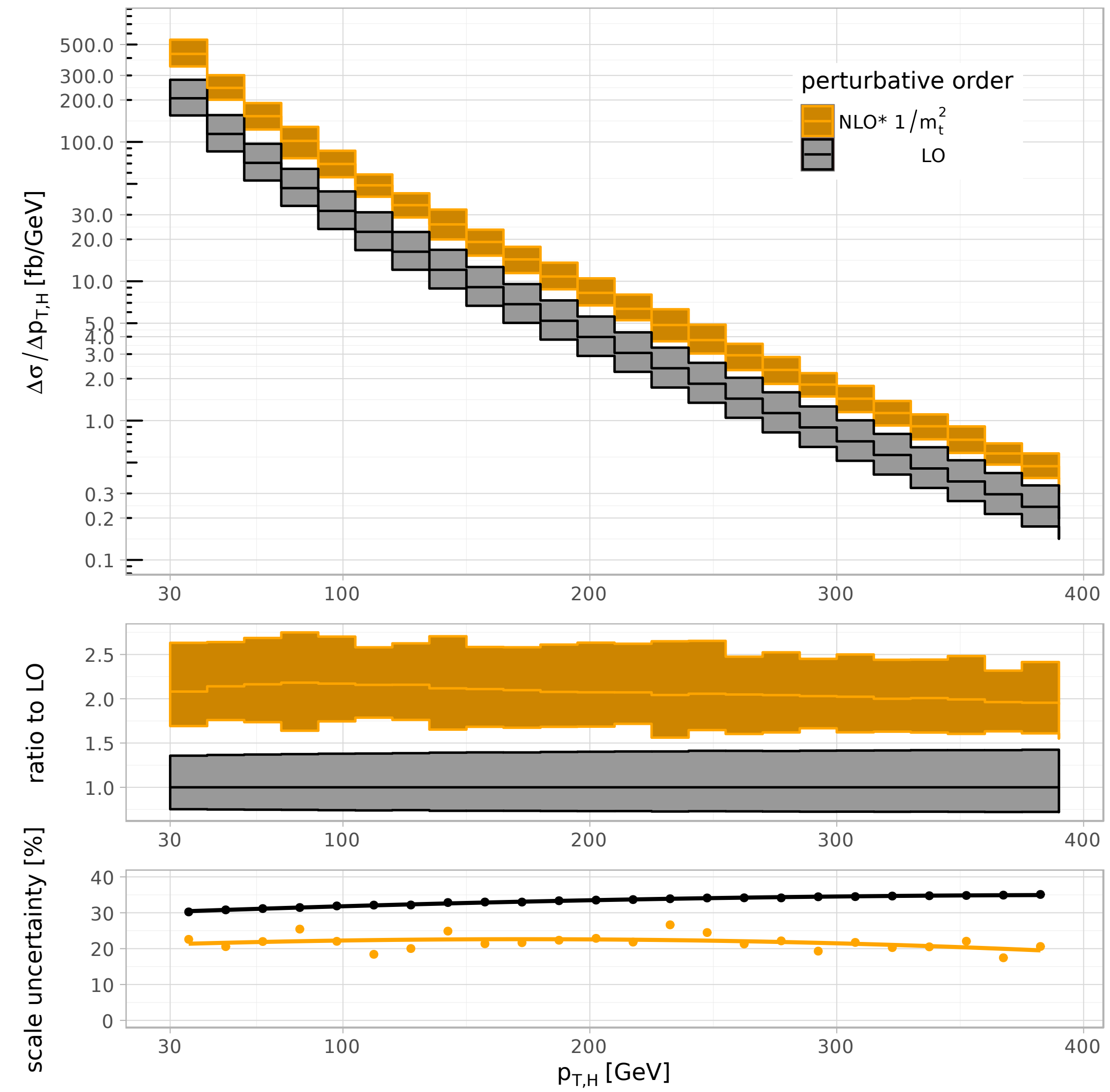


$$\mathrm{GF\ H(NNLO}+m_\mathrm{t}) = (1\ \mathrm{jet}\ m_\mathrm{t}\to\infty)\times\frac{\mathrm{MG\ LO}\ 0-2\ \mathrm{jet}\ m_\mathrm{t}}{(1\ \mathrm{jet}\ m_\mathrm{t}\to\infty)}$$

CKKW merged

HIGGS PT SPECTRUM

- Other CMS Higgs results use Powheg: 1 jet + $m_t = \infty$, arXiv:1111.2854
- We want to account for both effects of higher order corrections and finite top mass
  - No real NLO + finite top mass calculation available in the literature
- Adopt a factorized approach:
- LO H+0-2jet, finite m_t, p_t^H up to 600 GeV, including WW acceptance cuts arXiv:1410.5806 → We build on this
- NLO H+1jet finite m_t up to $1/m_t^4$ expansion: arXiv:1609.00367
-

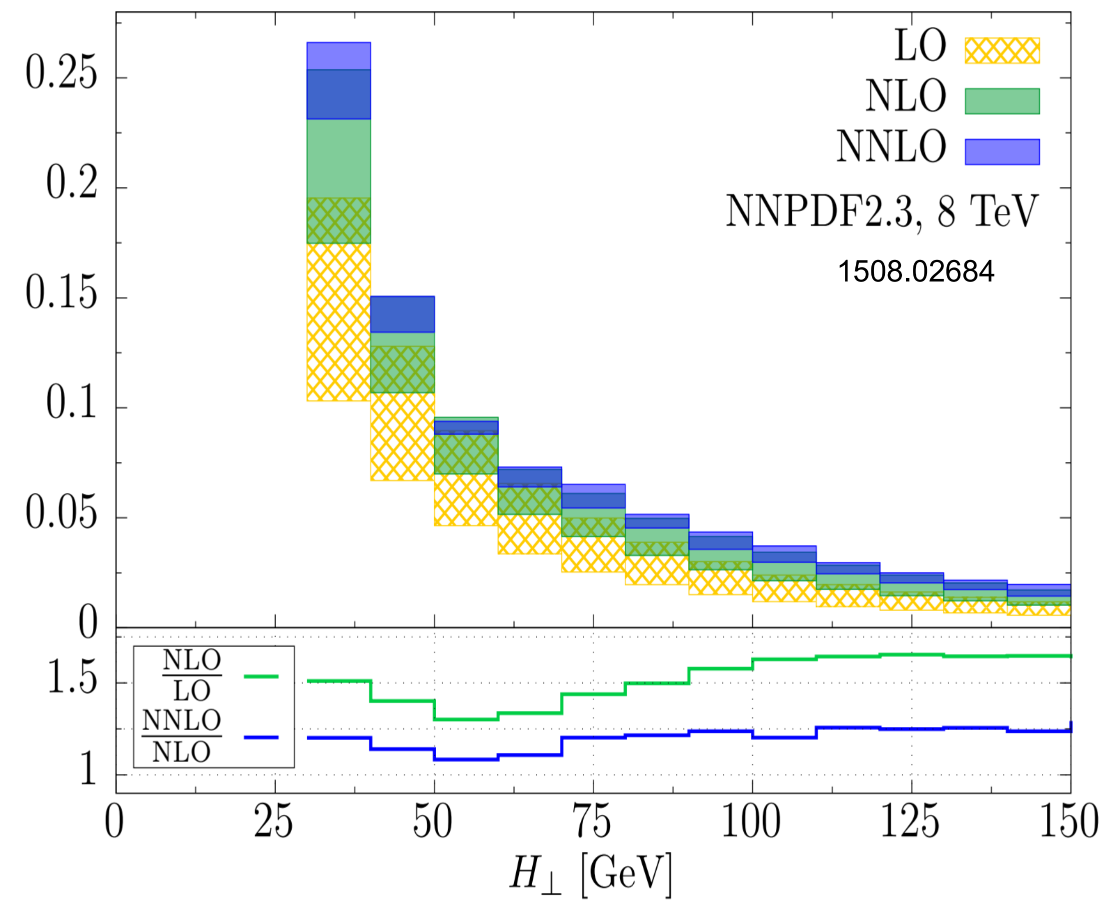


$$\mathrm{GF\ H(NNLO} + m_t) = (1\ \mathrm{jet}\ m_t \to \infty) \times \frac{\mathrm{MG\ LO\ 0-2\ jet}\ m_t}{(1\ \mathrm{jet}\ m_t \to \infty)} \times \frac{\mathrm{NLO\ 1\ jet}\ m_t}{\mathrm{LO\ 1\ jet}\ m_t}$$

CKKW merged factor of 2

HIGGS PT SPECTRUM

- Other CMS Higgs results use Powheg: 1 jet + $m_t = \infty$, arXiv:1111.2854
- We want to account for both effects of higher order corrections and finite top mass
  - No real NLO + finite top mass calculation available in the literature
- Adopt a factorized approach:
- LO H+0-2jet, finite m_t, p_t^H up to 600 GeV, including WW acceptance cuts arXiv:1410.5806 → We build on this
- NLO H+1jet finite m_t up to $1/m_t^4$ expansion: arXiv:1609.00367
- NNLO H+1jet, $m_t = \infty$, p_T^H up to ~200 GeV, arXiv:1408.5325, arXiv:1302.6216, arXiv:1504.07922, arXiv:1505.03893, arXiv:1508.02684

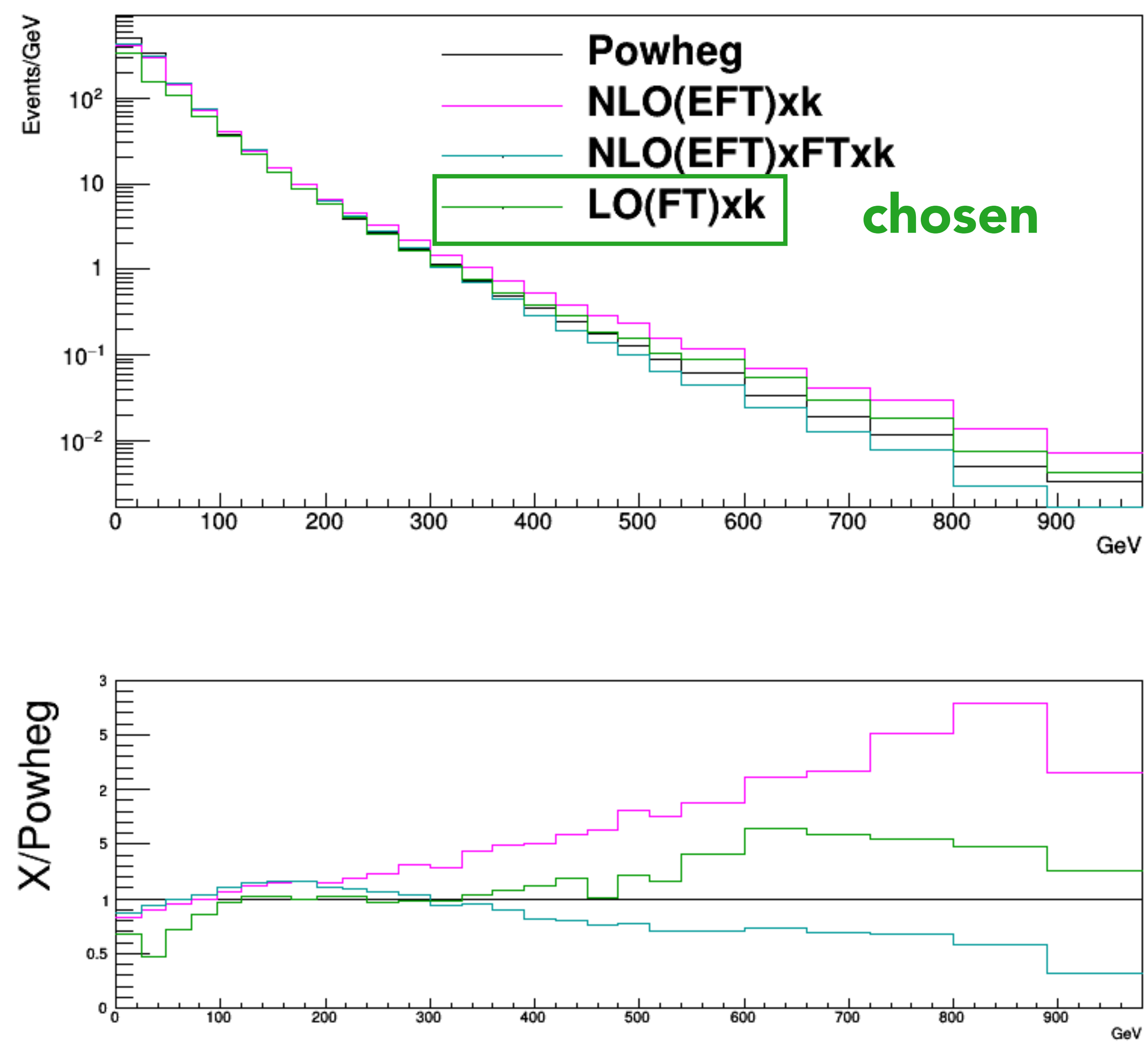


$$\text{GF H}(\text{NNLO} + m_t) = (1\ \text{jet}\ m_t \to \infty) \times \frac{\text{MG LO } 0-2\ \text{jet}\ m_t}{(1\ \text{jet}\ m_t \to \infty)} \times \frac{\text{NLO 1 jet}\ m_t}{\text{LO 1 jet}\ m_t} \times \frac{\text{NNLO 1 jet}\ m_t \to \infty}{\text{NLO 1 jet}\ m_t \to \infty}$$

CKKW merged | factor of 2 | factor of 1.25

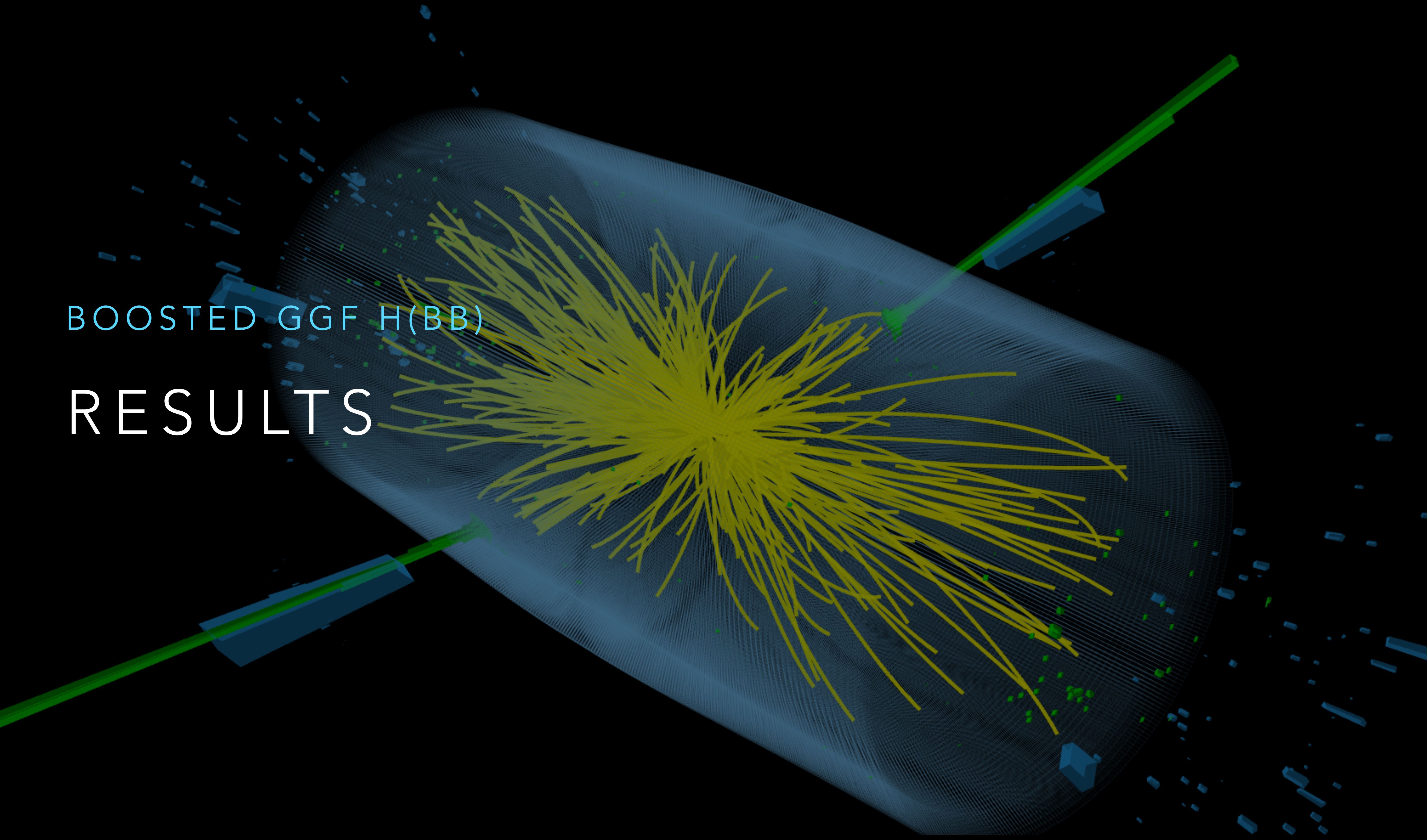
HIGGS PT SPECTRUM

- Pythia version of CKKW-L merged 0,1,2jet LO finite top mass
- ME generation in aMC@NLO (ptj > 20) with xqcut = 30 GeV
- CKKW shower is extended down to a merging scale of TMS = 20 GeV
- Two factorized systematic uncertainties:
  - 30% overall normalization
  - 30% linear change in slope (no effect on overall norm.)



BOOSTED GGF H(BB)

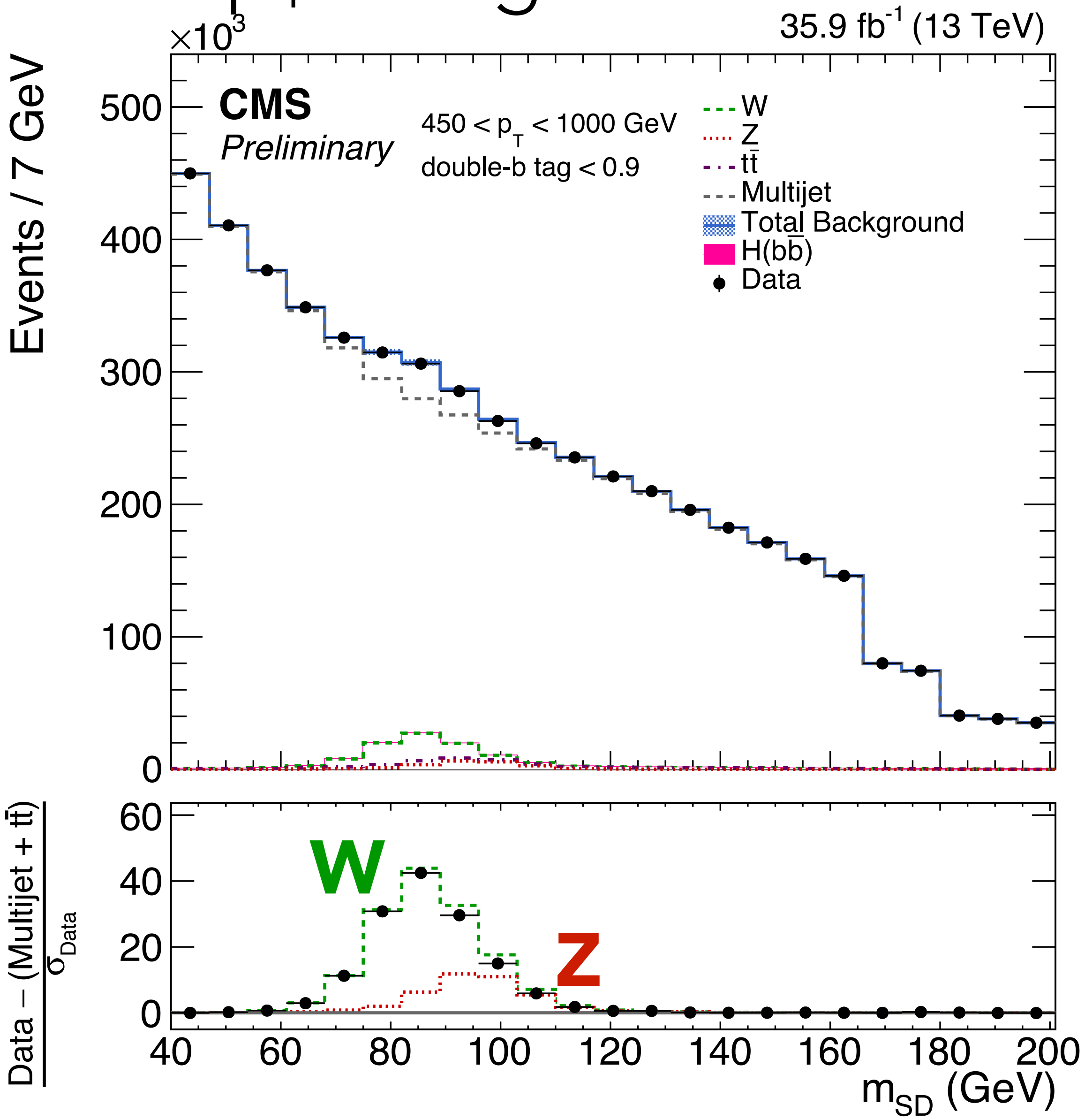
RESULTS

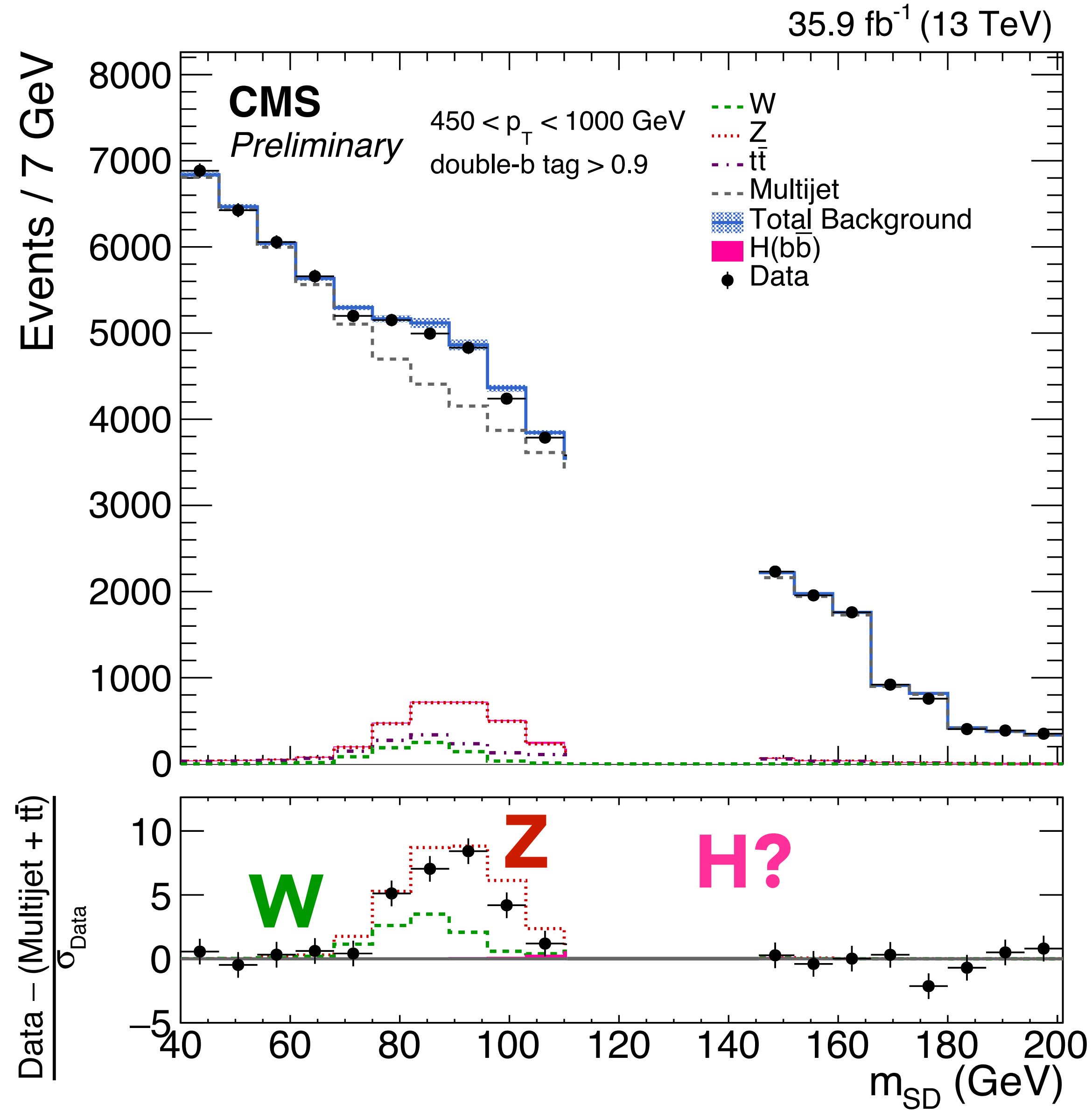


HIG-17-010

FIT RESULTS

- Simultaneous fit for Z(bb) and H(bb)
- All p_T categories



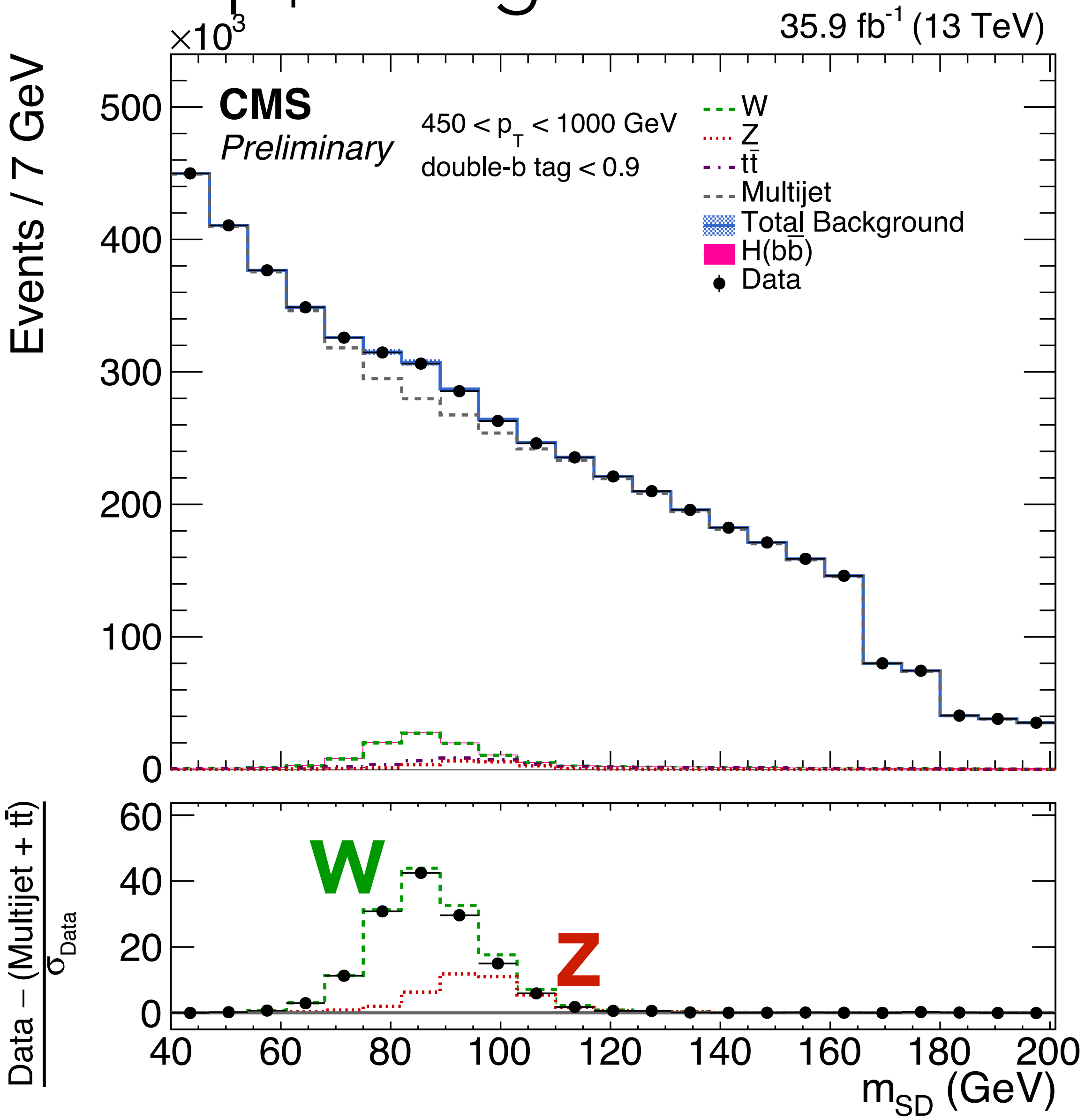


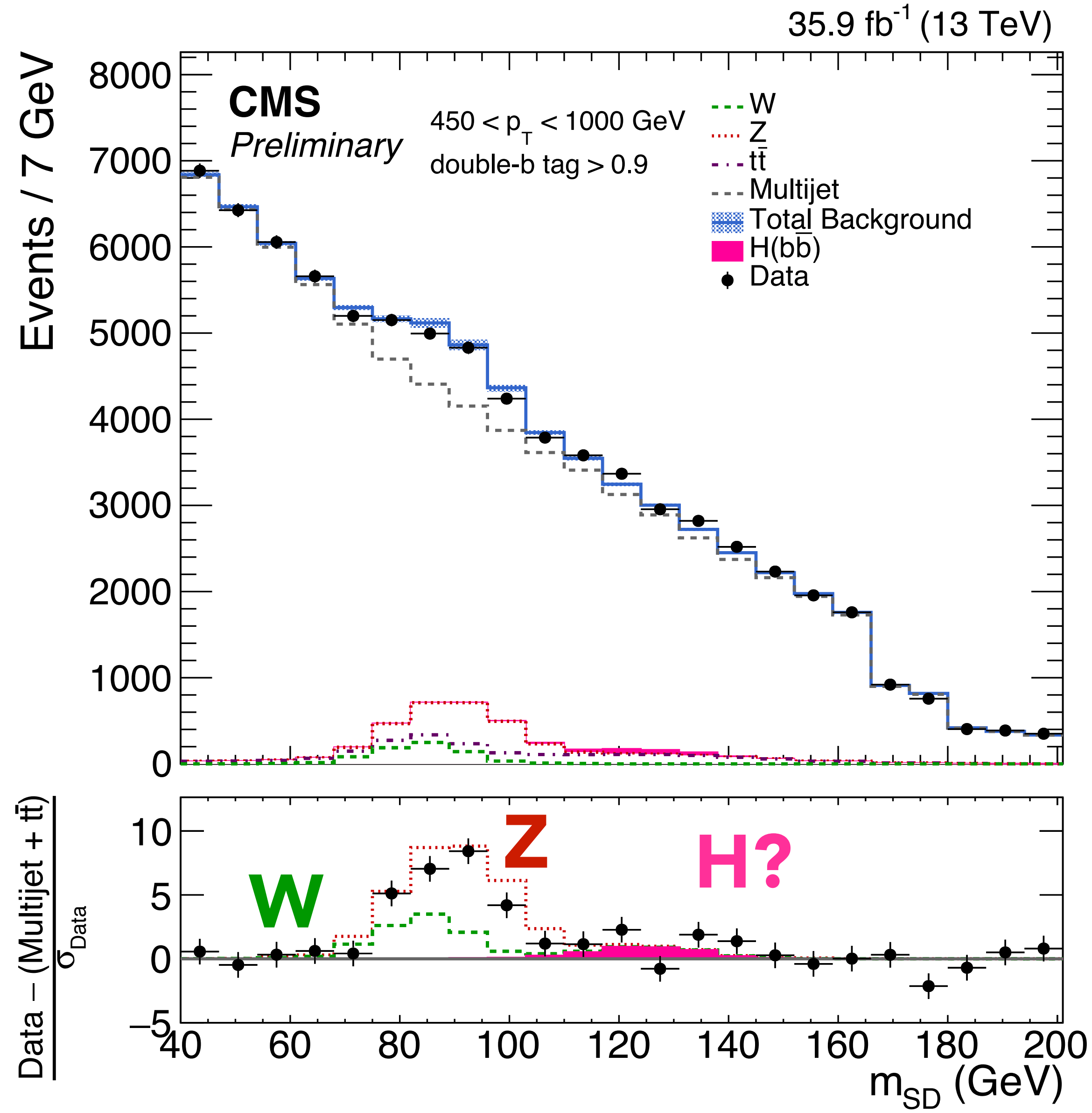
SM candles: Z(bb) peak provides in-situ constraint of H(bb) signal systematics

observed Z(bb) significance: 5.1σ, $\mu_Z = 0.78^{+0.23}_{-0.19}$

FIT RESULTS

- Simultaneous fit for Z(bb) and H(bb)
- All p_T categories





observed H(bb) significance:

1.5σ, $\mu_H = 2.3^{+1.8}_{-1.6}$

FIT RESULTS

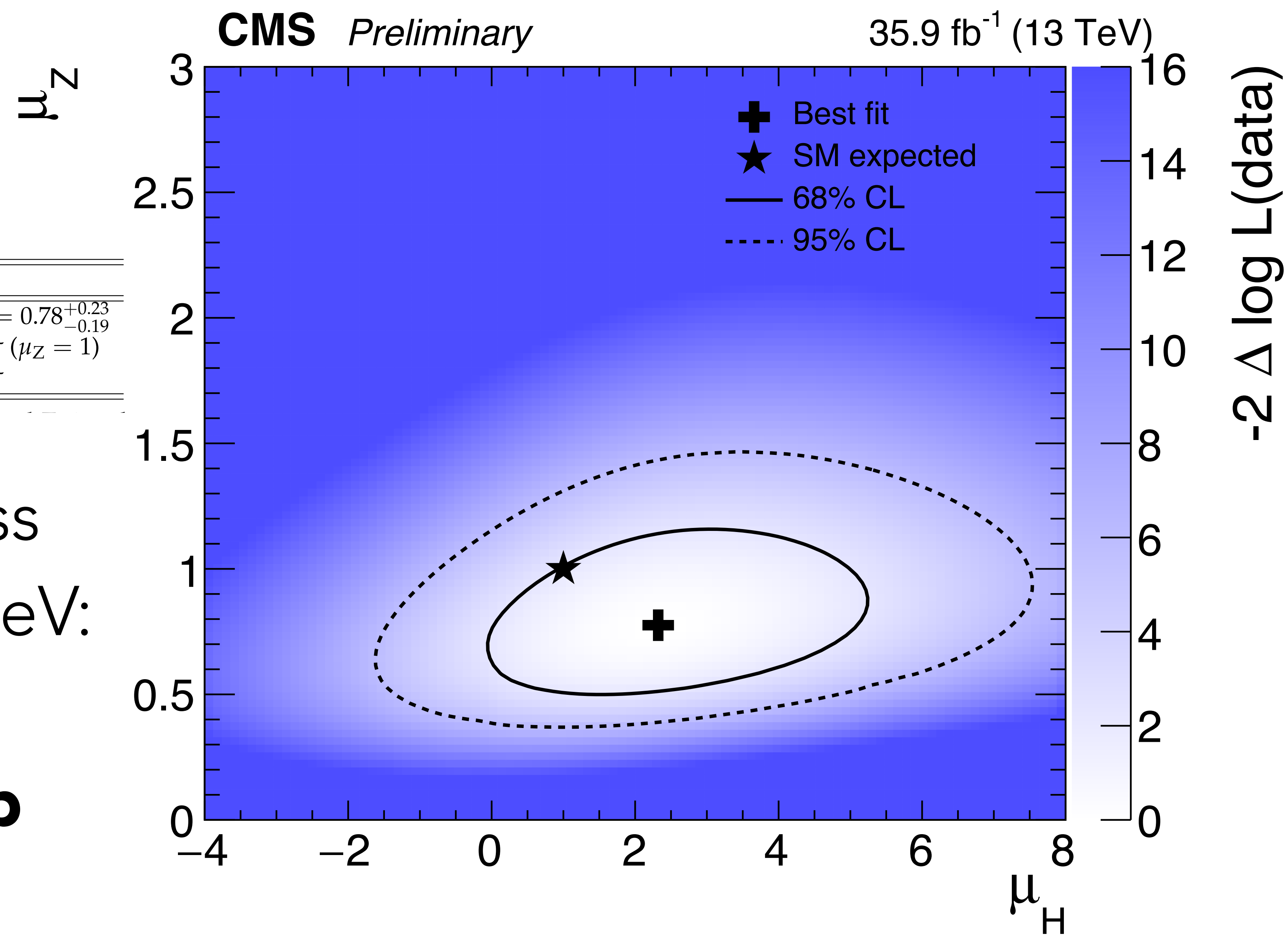
- Two dimensional likelihood scan

	H	H no p_T corrections	Z
Observed best fit	$\mu_H = 2.3^{+1.8}_{-1.6}$	$\mu'_H = 3.2^{+2.2}_{-2.0}$	$\mu_Z = 0.78^{+0.23}_{-0.19}$
Expected significance	0.7σ ($\mu_H = 1$)	0.5σ ($\mu'_H = 1$)	5.8σ ($\mu_Z = 1$)
Observed significance	1.5σ	1.6σ	5.1σ

measured visible cross sections for $p_T > 450$ GeV:

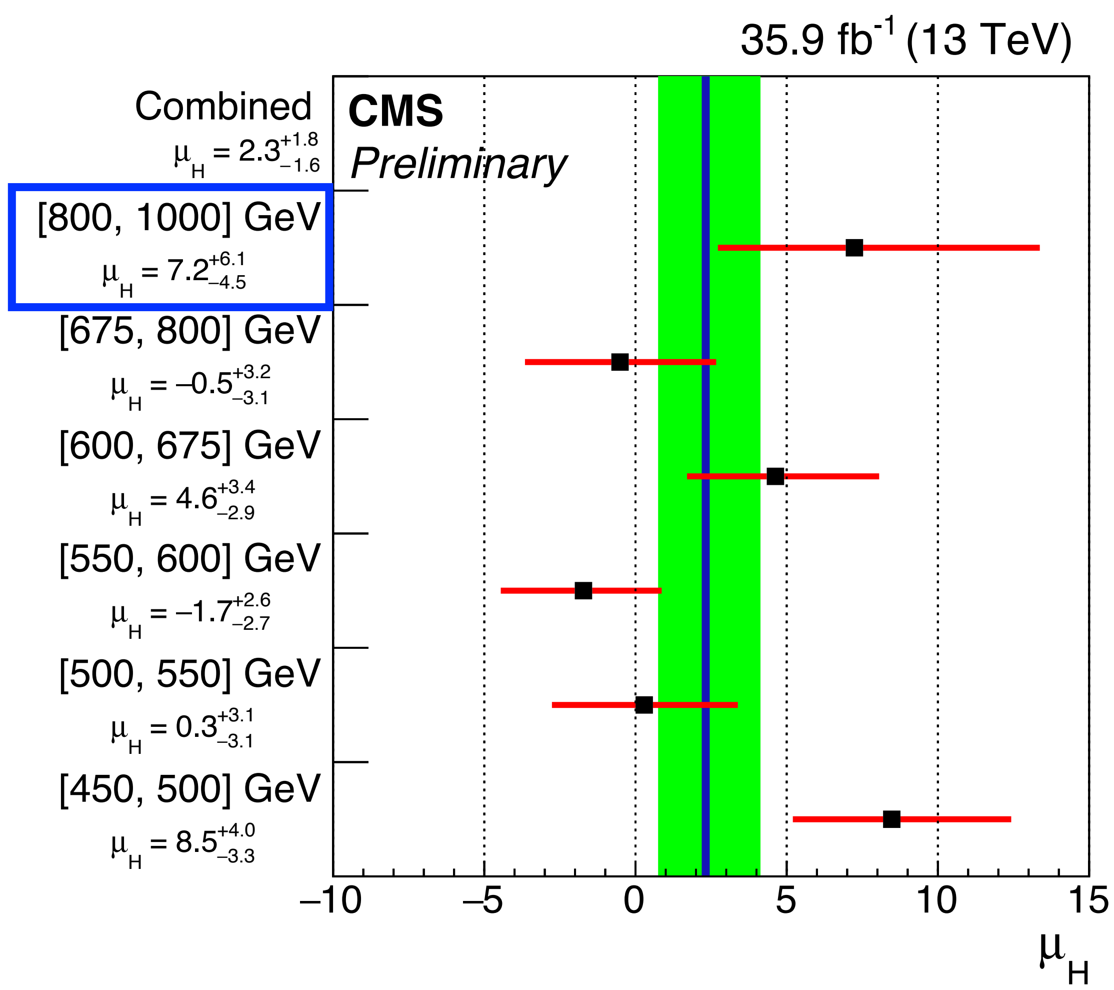
$\sigma_H = 74^{+51}_{-49}$ fb

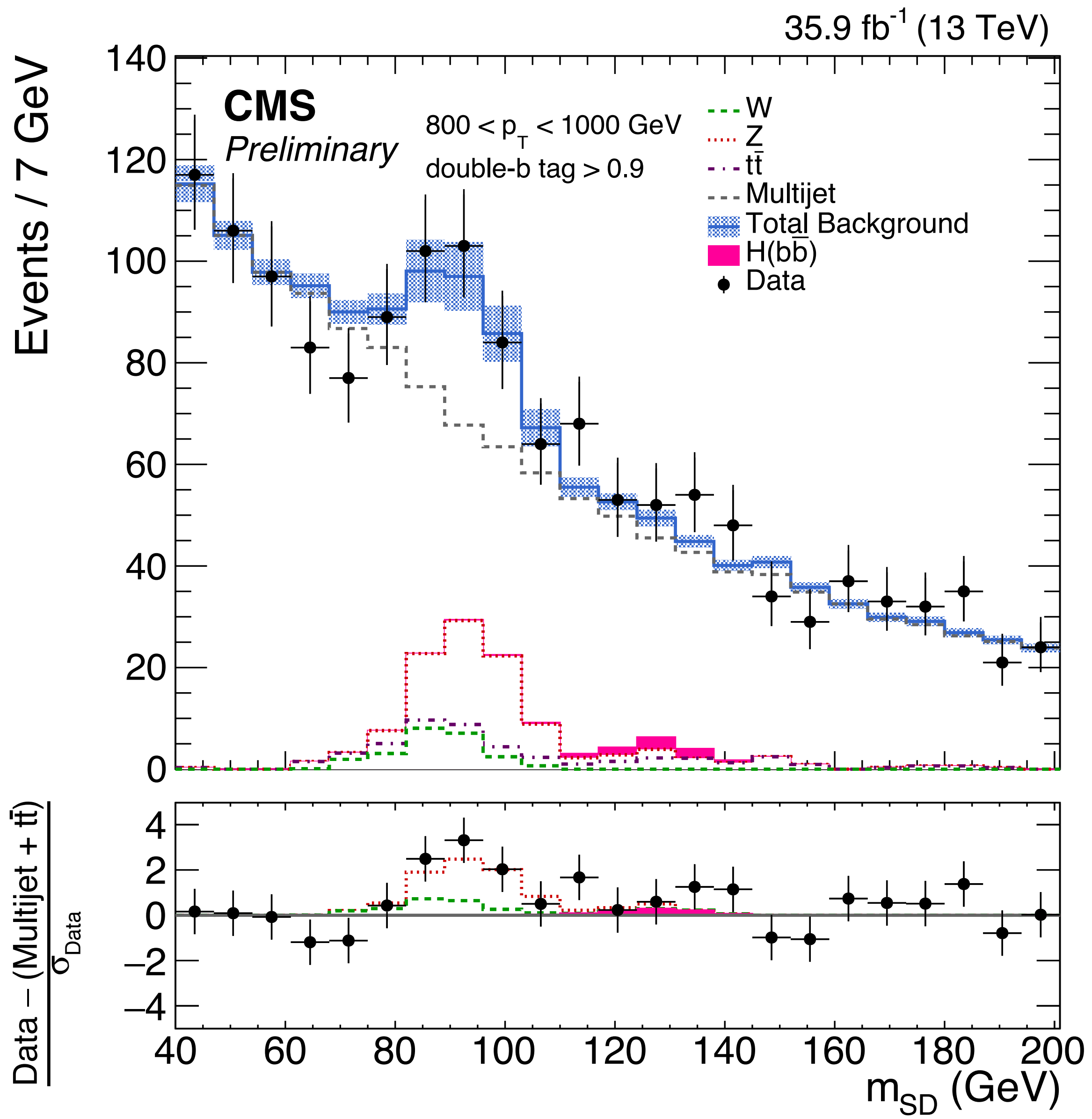
$\sigma_Z = 0.85^{+0.26}_{-0.21}$ pb



HIG-17-010

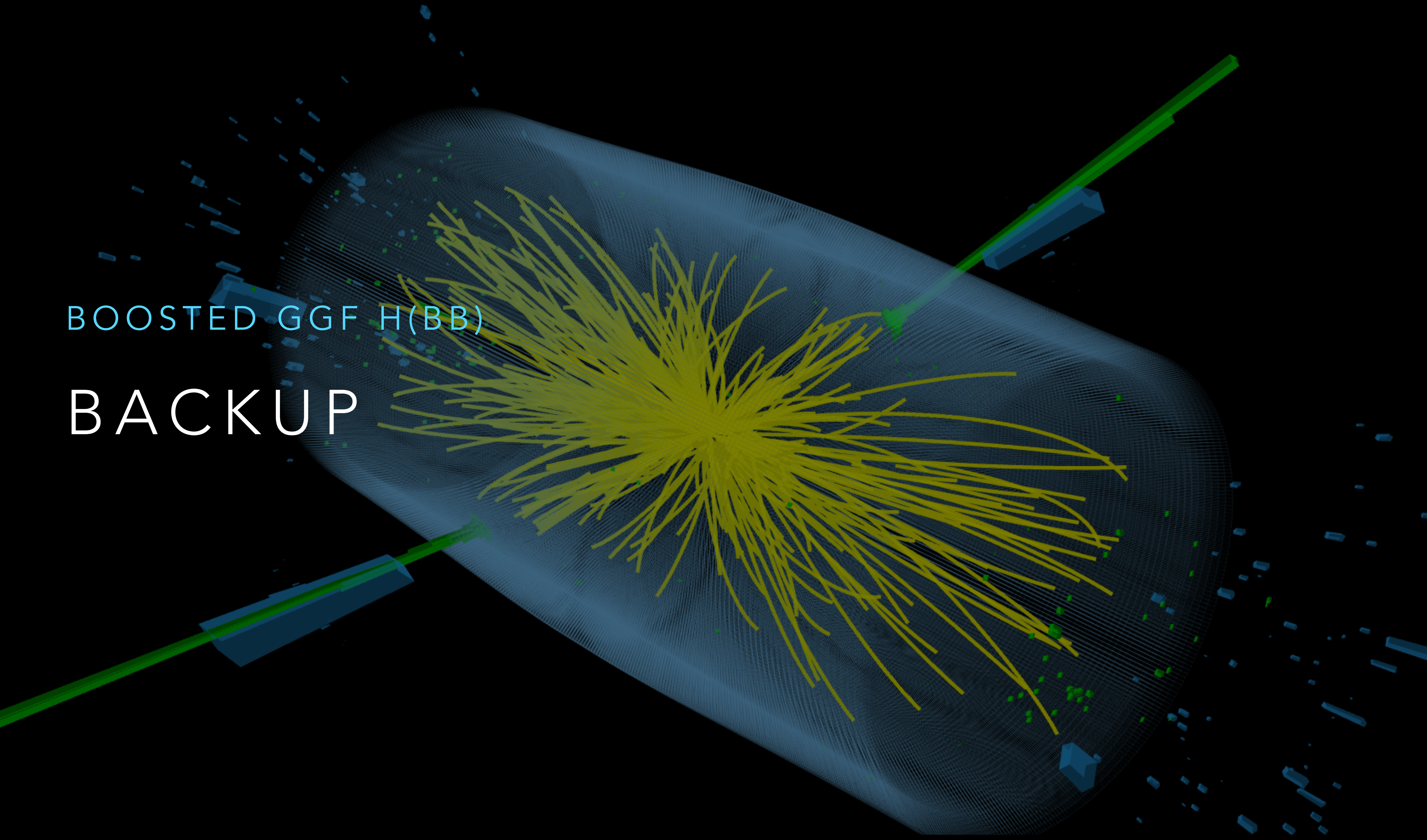
P_T CATEGORIES





SUMMARY AND OUTLOOK

- First LHC search for gg → H → bb in boosted topology
  - First observation of Z(bb) in single-jet topology, 5.1σ observed (5.8σ expected)
  - Observed significance of H(bb) is 1.5σ
- Measured cross sections agree with SM
- Search probes **previously unexplored** regions of Higgs phase space
- **New** and **generic** strategy to search for boosted hadronic Higgs decays
  - Future prospects are bright
  - Means we need help from LHC H XS WG for best possible theory prediction in boosted Higgs regime — $\mathbf{p_T^H}$ up to **1 TeV** and beyond…



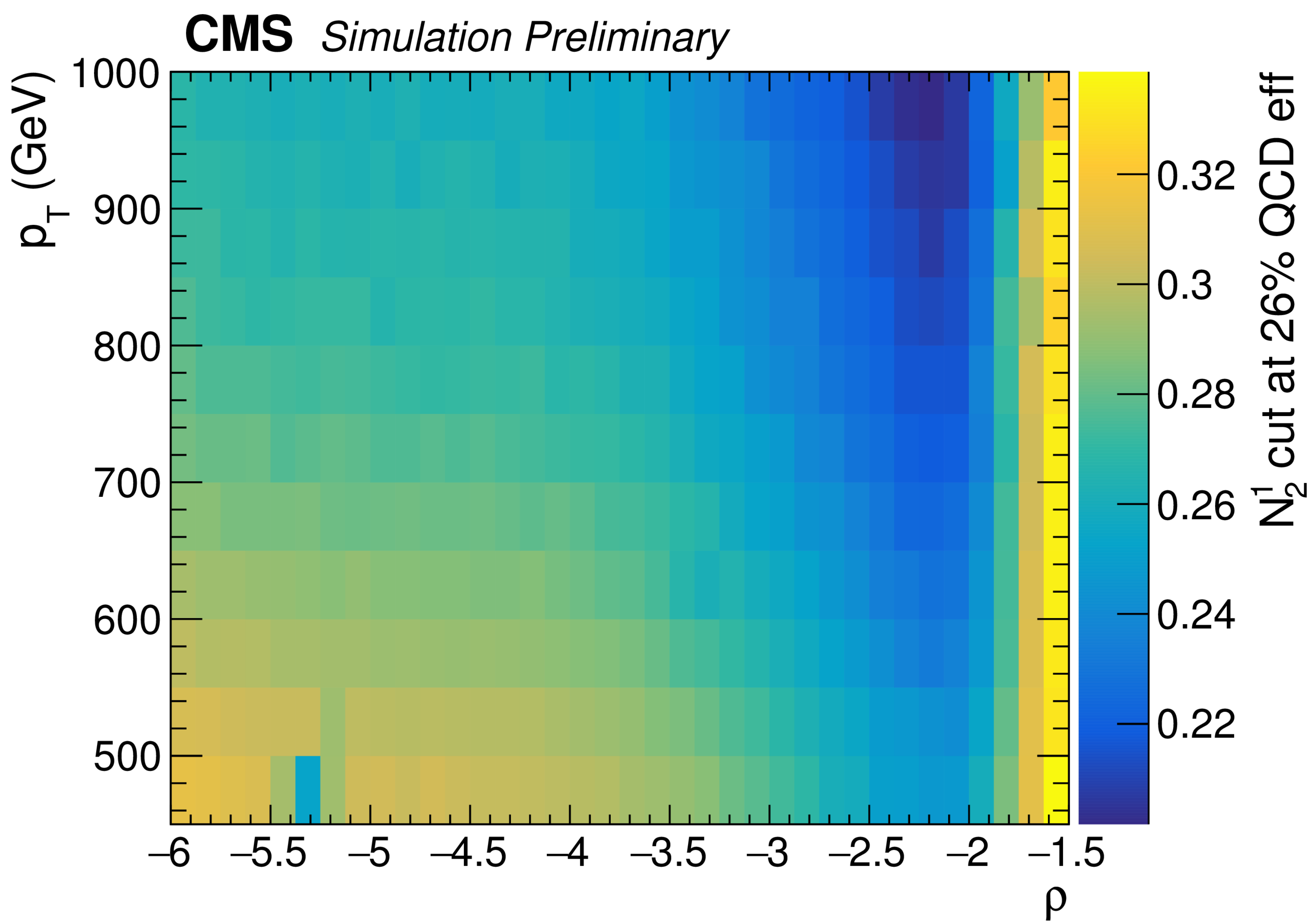
BOOSTED GGF H(BB)

BACKUP

Javier Duarte
Fermilab

$N^1{}_2{}^{DDT}$

- Cut value map used to transform $N^1{}_2$

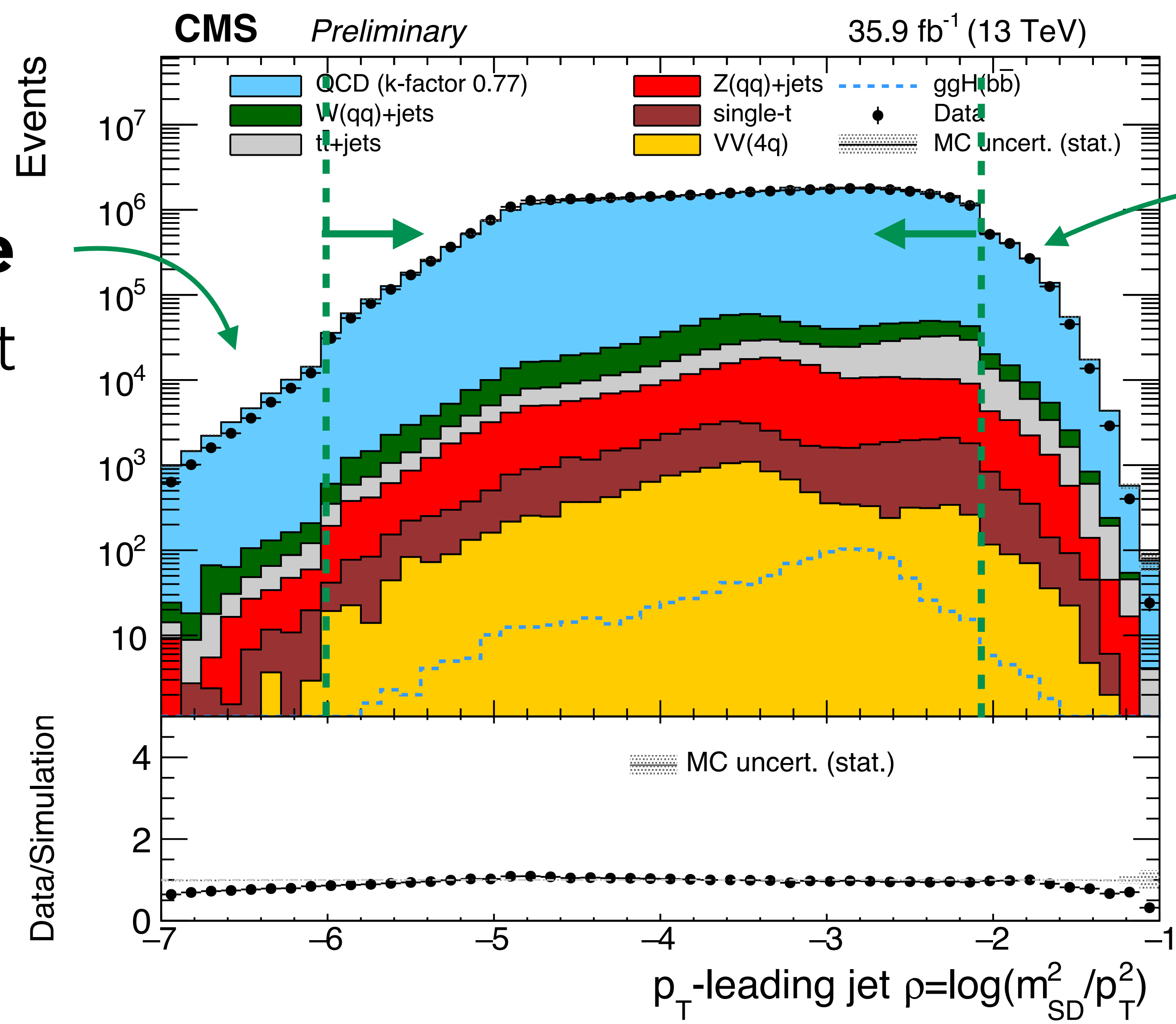


HIG-17-010

EVENT SELECTION

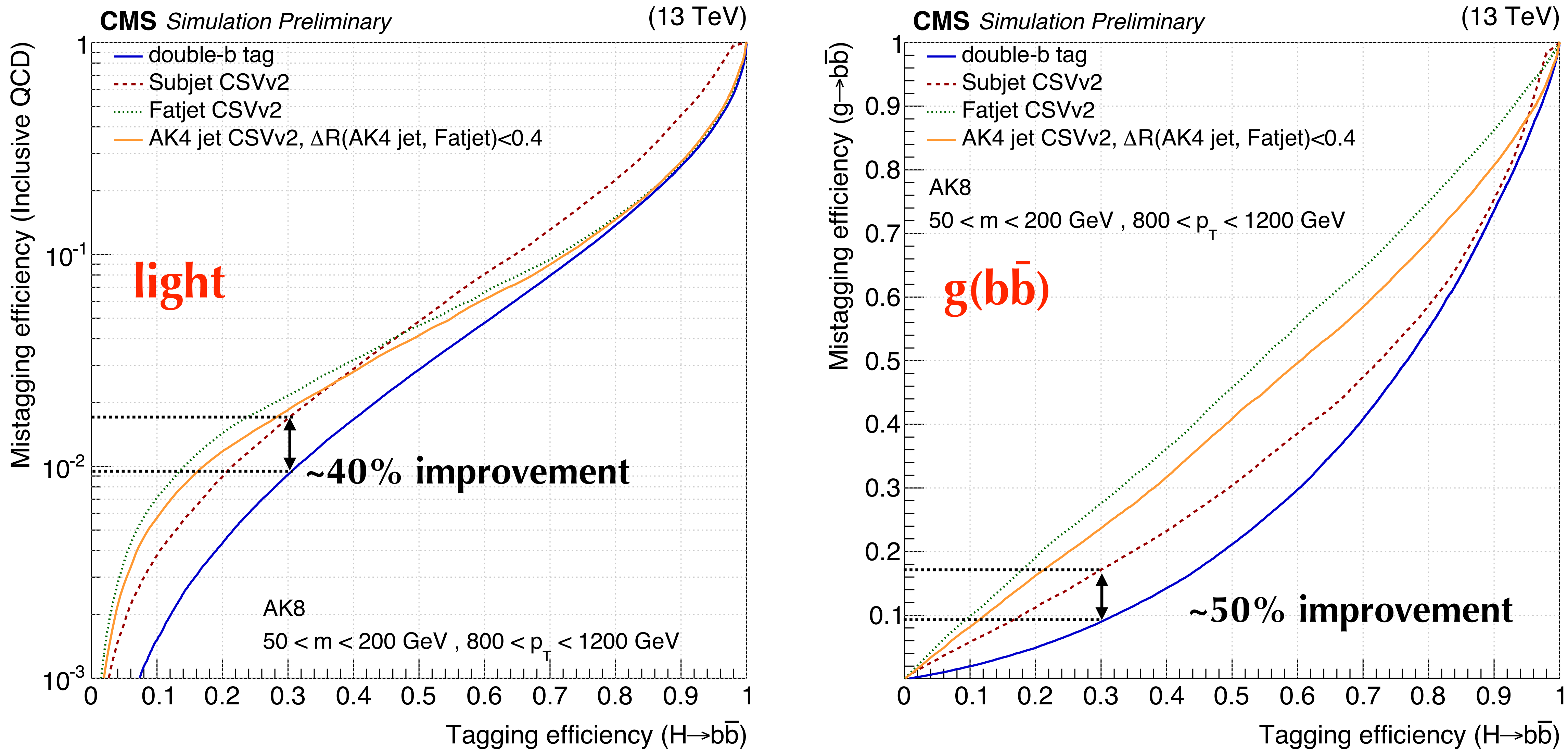
$-6.0 < \rho < -2.1$

non-perturbative regime of the soft drop mass calculation



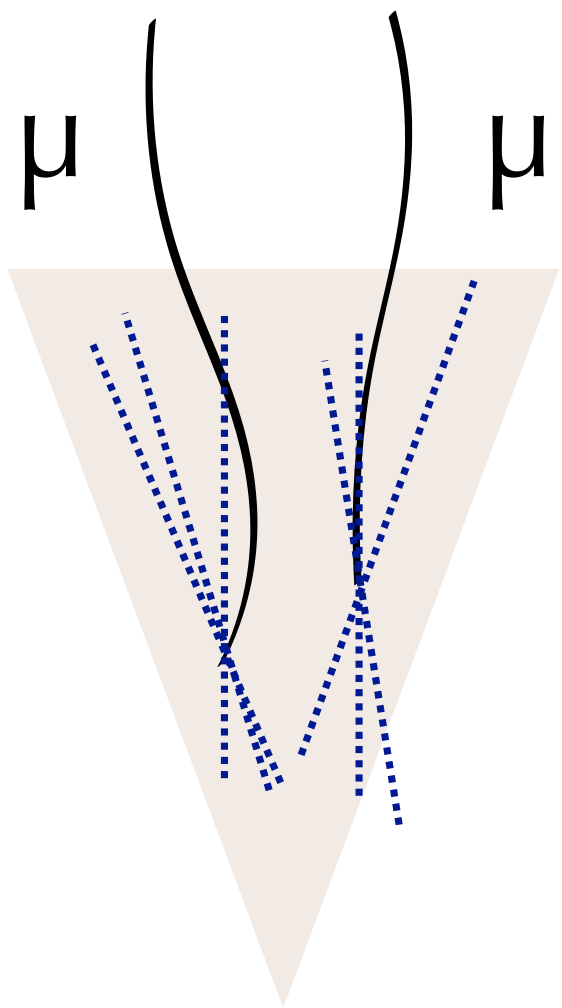
finite cone effects from the AK8 jet clustering

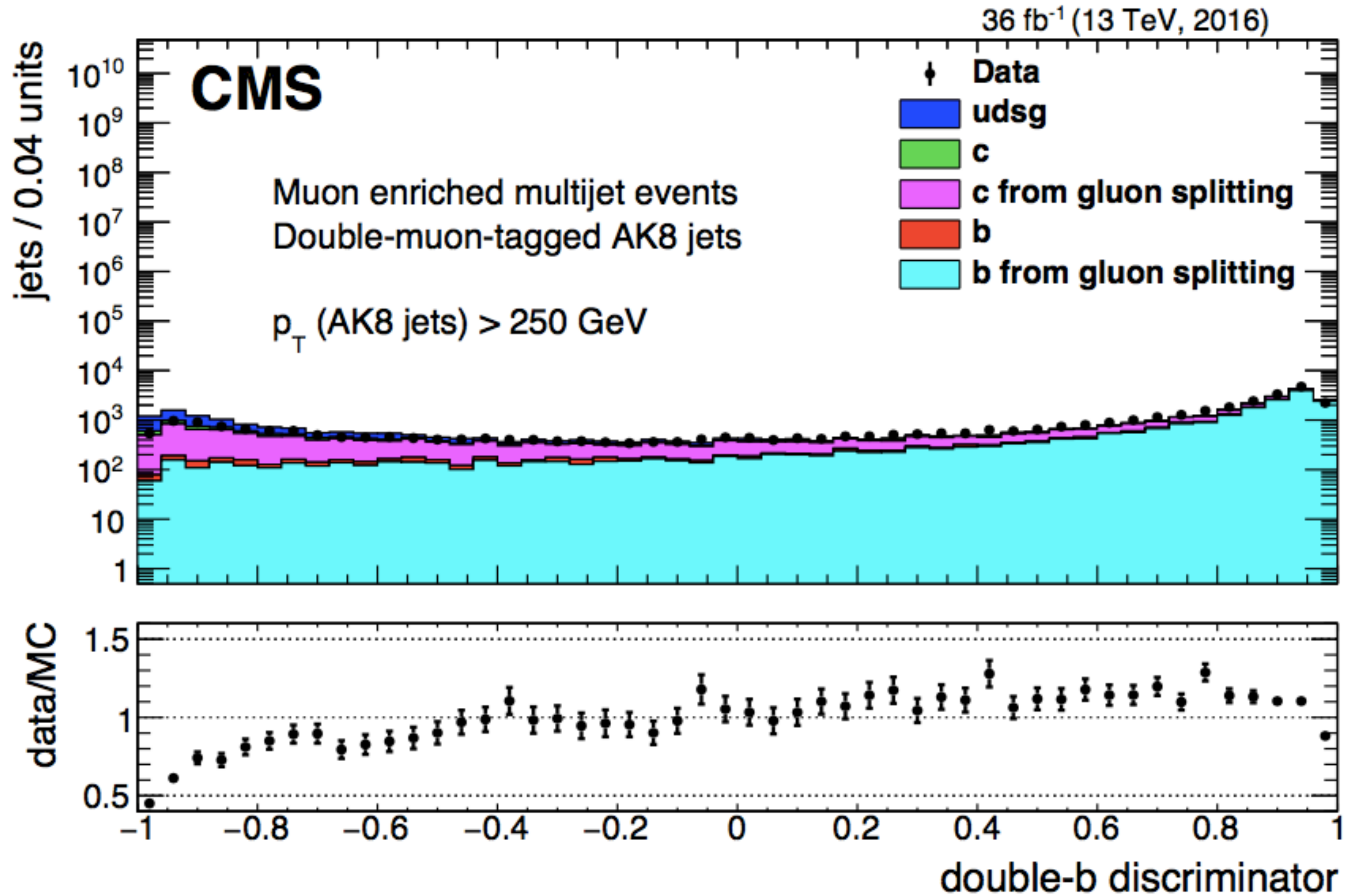
EFFICIENCY AND MIS-TAG



- Mis-tag is reduced by more than 40% at 30% signal efficiency for a tight working point

EFFICIENCY IN DATA

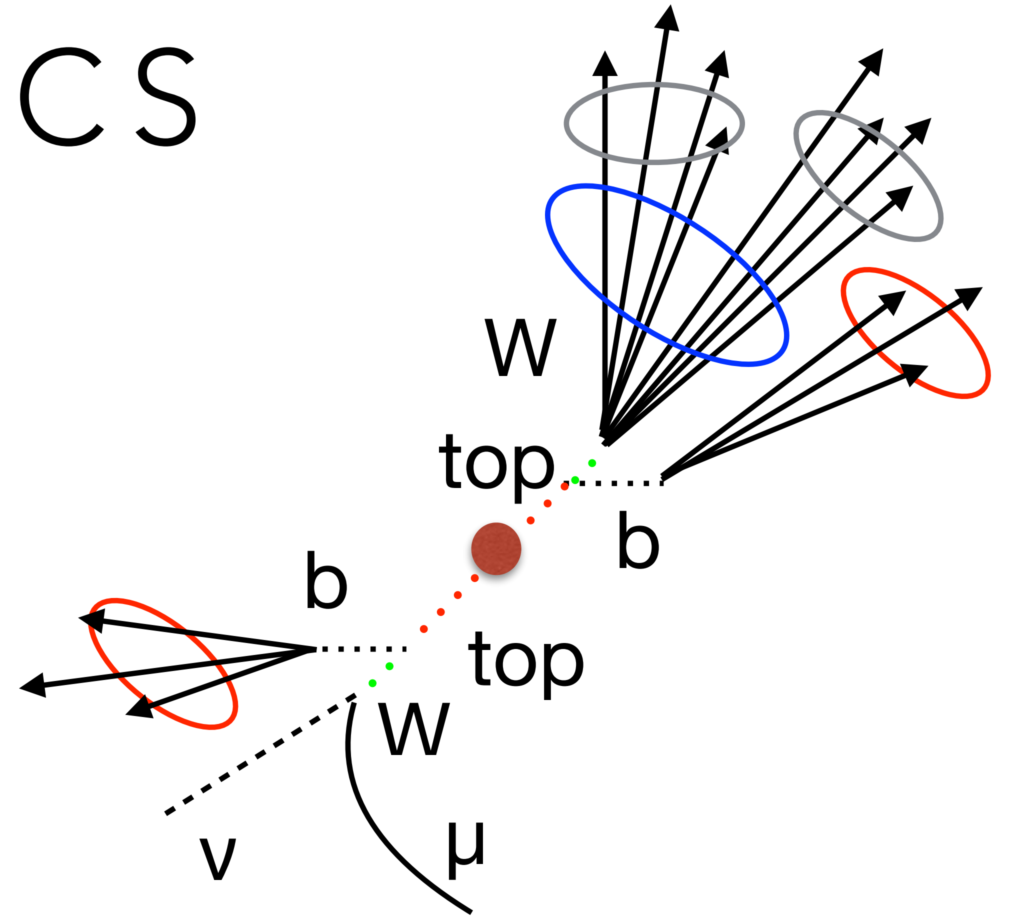




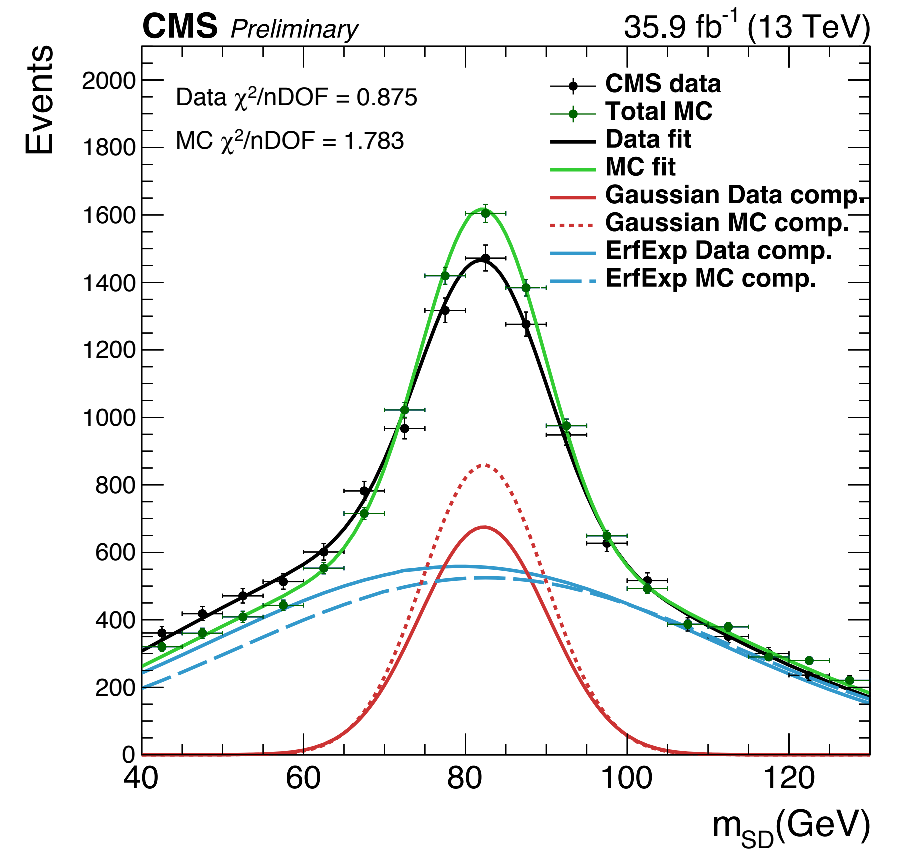
- Using g(bb) jet as proxy in double muon tagged jet sample
- Associated data/MC uncertainty 3-5%

SYSTEMATICS

Systematic uncertainty source	Type (shape or normalization)	Relative size (or description)
QCD transfer factor	both	profile $a_{k\ell}$ and QCD normalization
Luminosity	normalization	2.5%
V-tag ($N_2^{1,DDT}$) efficiency	normalization	4.3%
Muon veto efficiency	normalization	0.5%
Electron veto efficiency	normalization	0.5%
Trigger efficiency	normalization	4%
Muon ID efficiency	shape	up to 0.2%
Muon isolation efficiency	shape	up to 0.1%
Muon trigger efficiency	shape	up to 8%
$t\bar{t}$ normalization SF	normalization	from 1μ CR: 8%
$t\bar{t}$ double-b mis-tag SF	normalization	from 1μ CR: 15%
W/Z NLO QCD corrections	normalization	10%
W/Z NLO EWK corrections	normalization	15% – 35%
W/Z NLO EWK ratio decorrelation	normalization	5% – 15%
double-b tagging efficiency	normalization	4%
Jet energy scale	normalization	up to 10%
Jet energy resolution	normalization	up to 15%
Jet mass scale	shape	shift m_{SD} peak by $\pm$0.4%
Jet mass resolution	shape	smear m_{SD} distribution by $\pm$9%
Jet mass scale p_T	normalization	0.4%/100 GeV (p_T)
Monte Carlo statistics	normalization	-
H p_T correction (gluon fusion)	both	30%

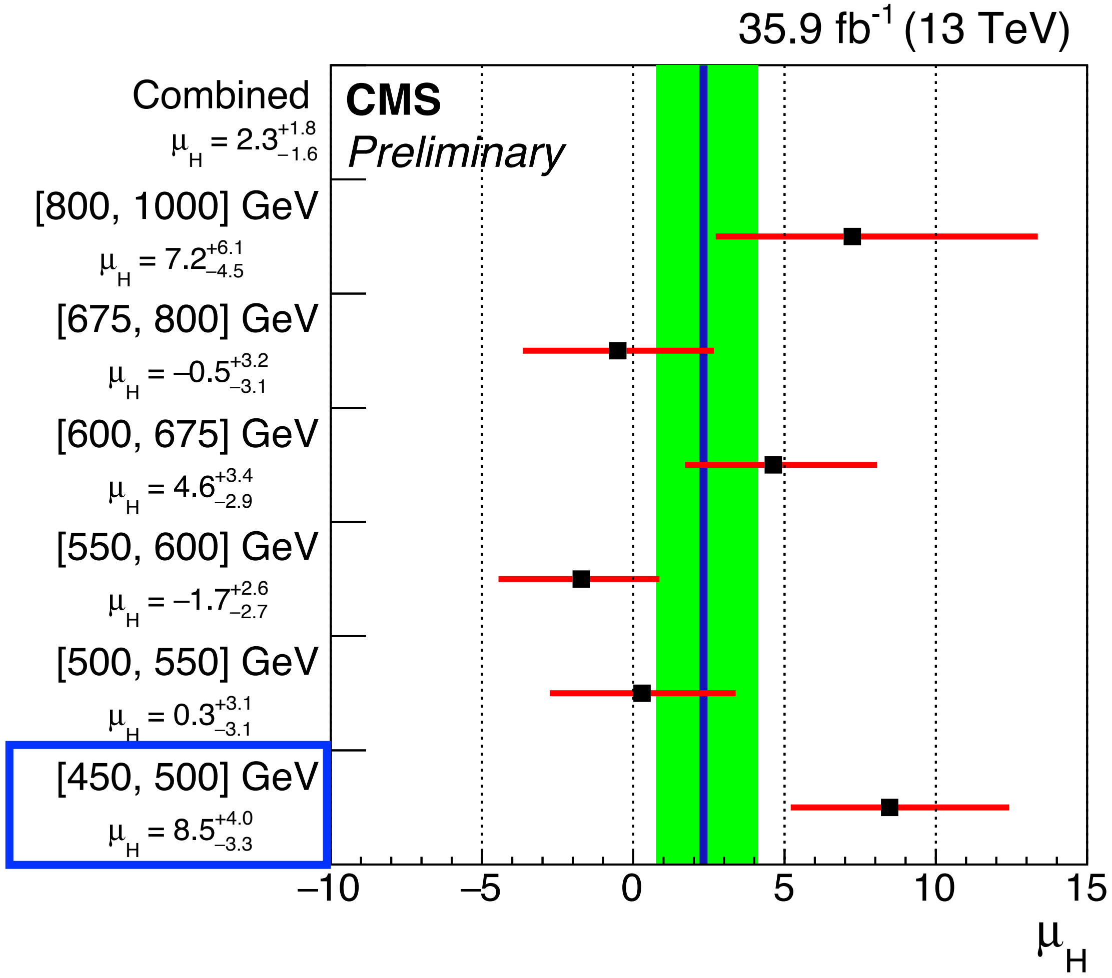


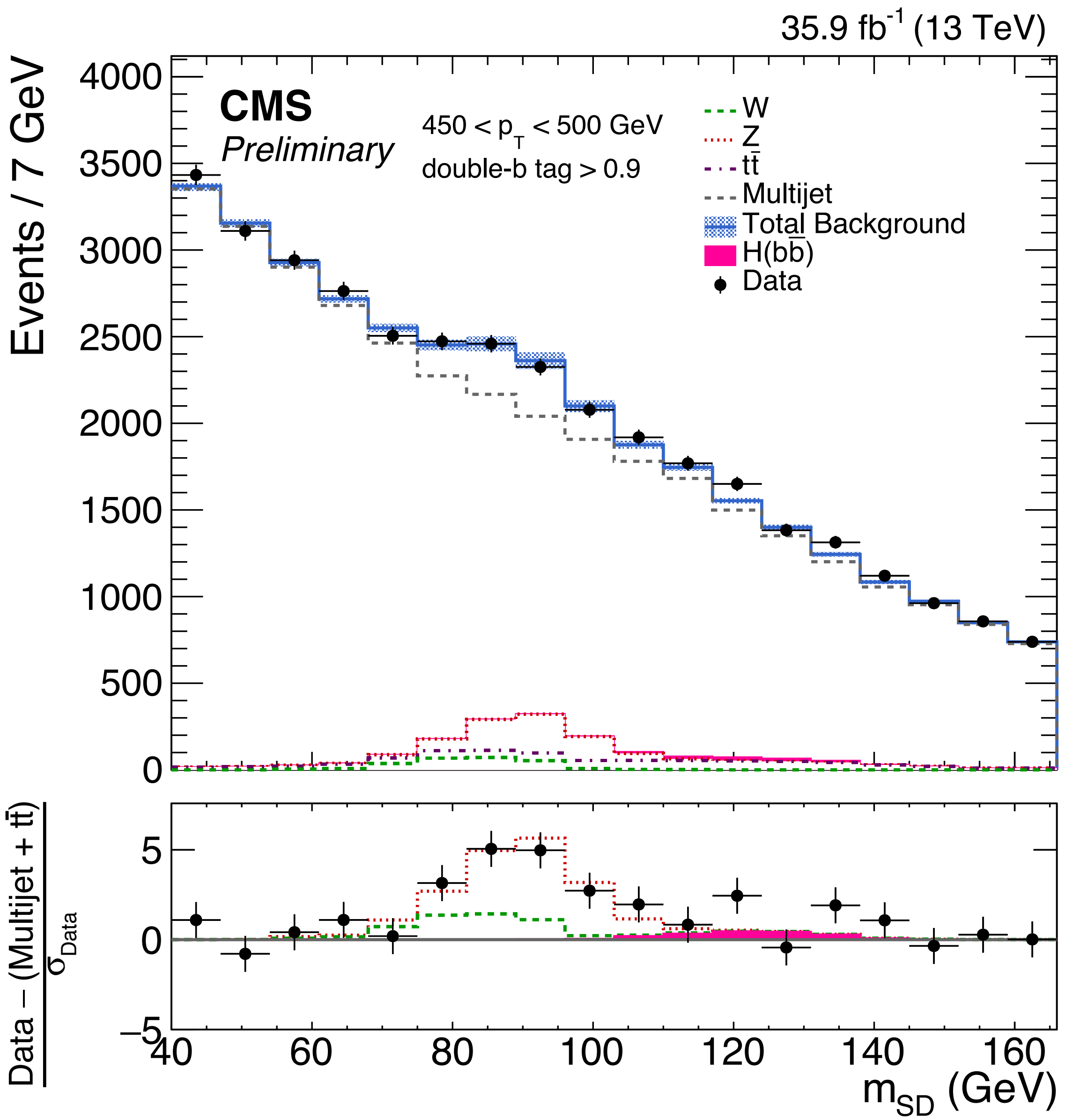
- Signal systematic uncertainties from merged W sample in semi-leptonic ttbar events (external constraint)
- SM candles: presence of W/Z in final jet mass distribution provides additional in-situ constraint



HIG-17-010

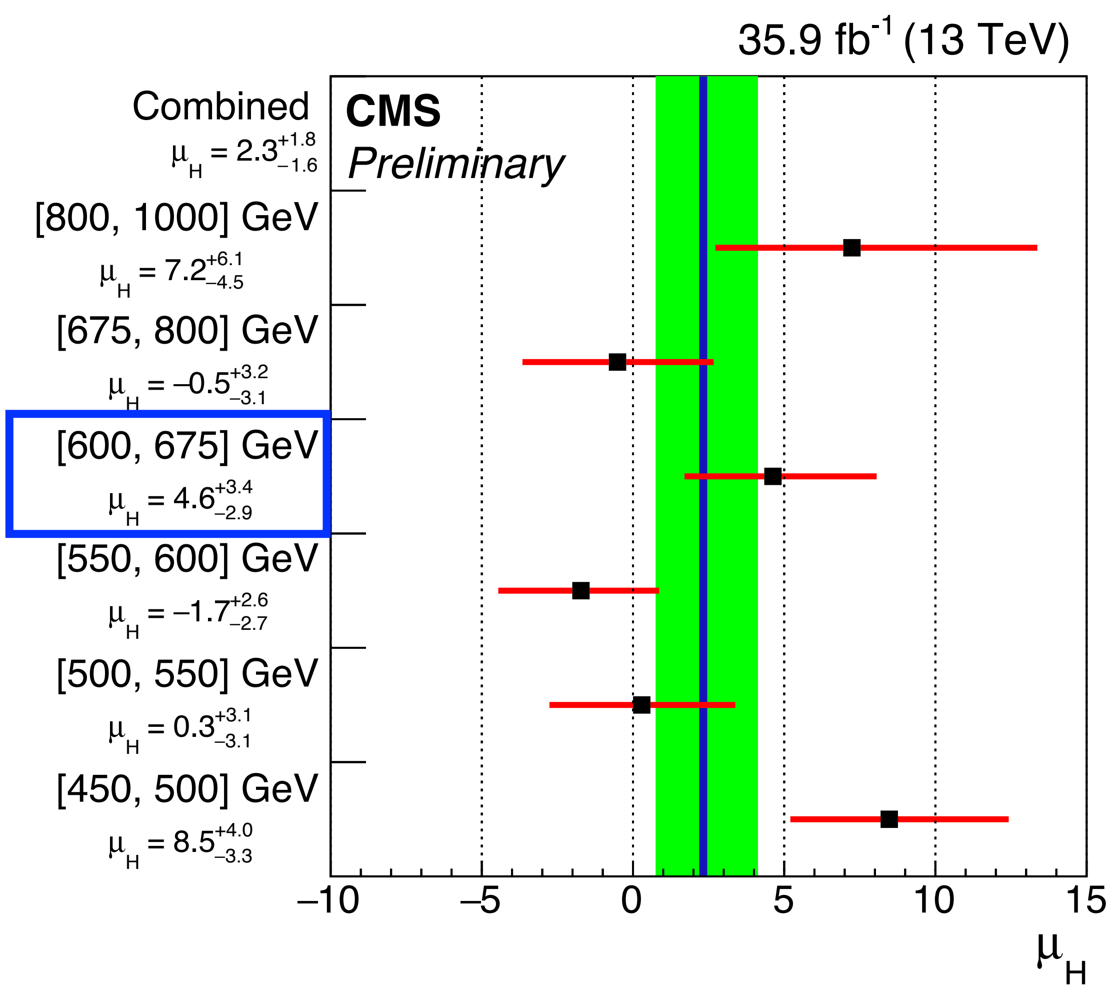
P_T CATEGORIES

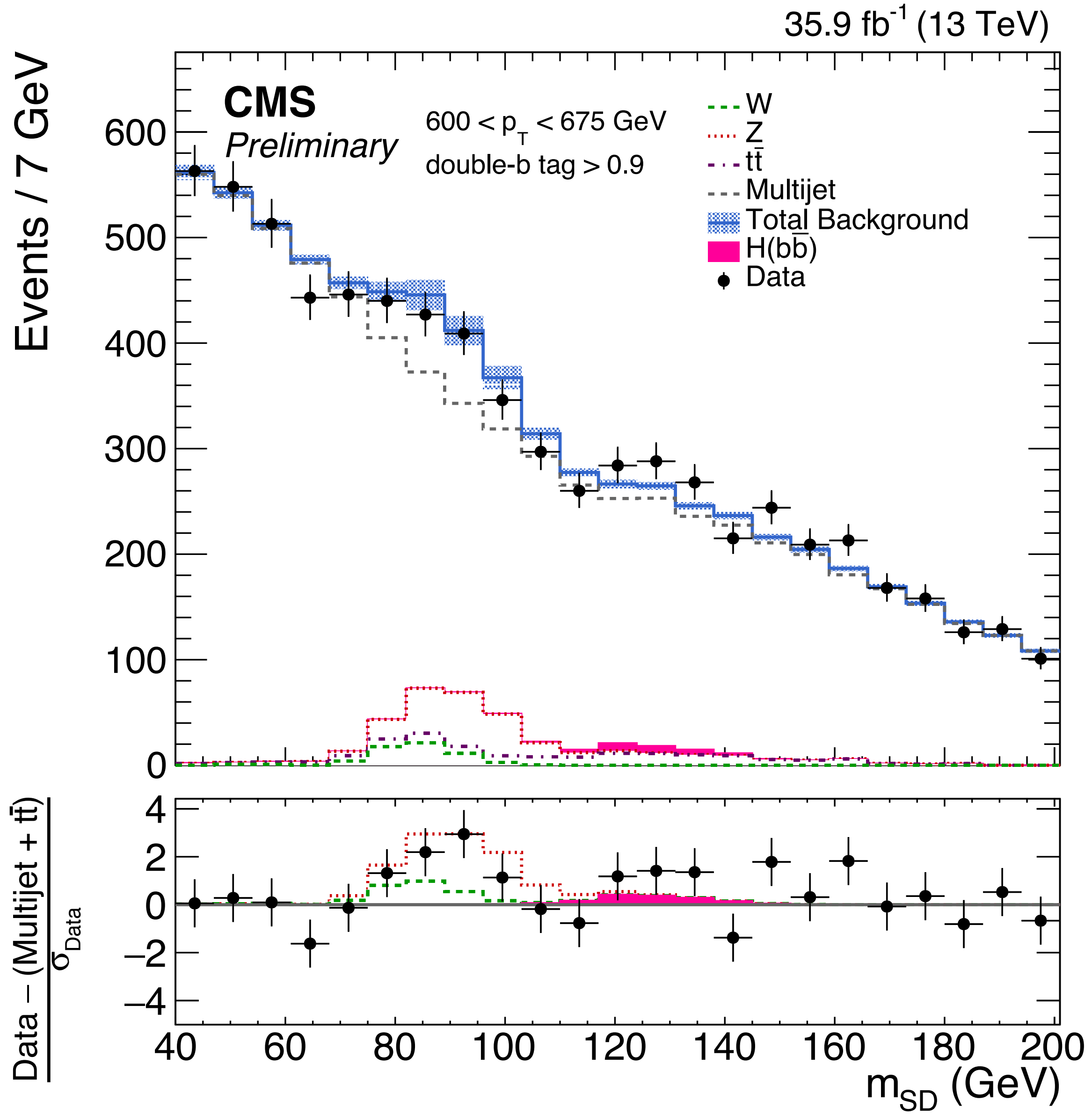


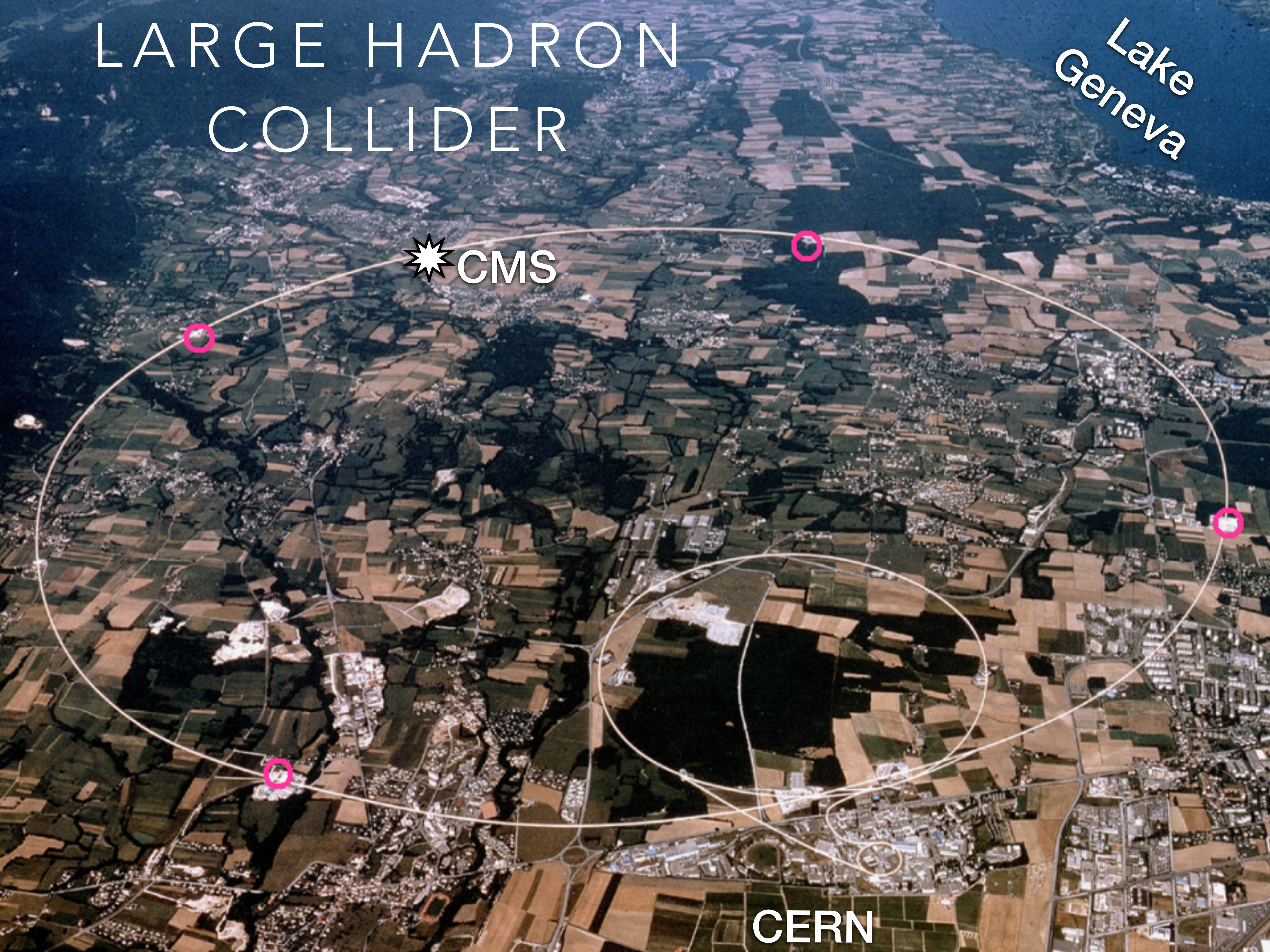


[HIG-17-010](HIG-17-010)

P_T CATEGORIES






LARGE HADRON COLLIDER
Lake Geneva
CMS
CERN

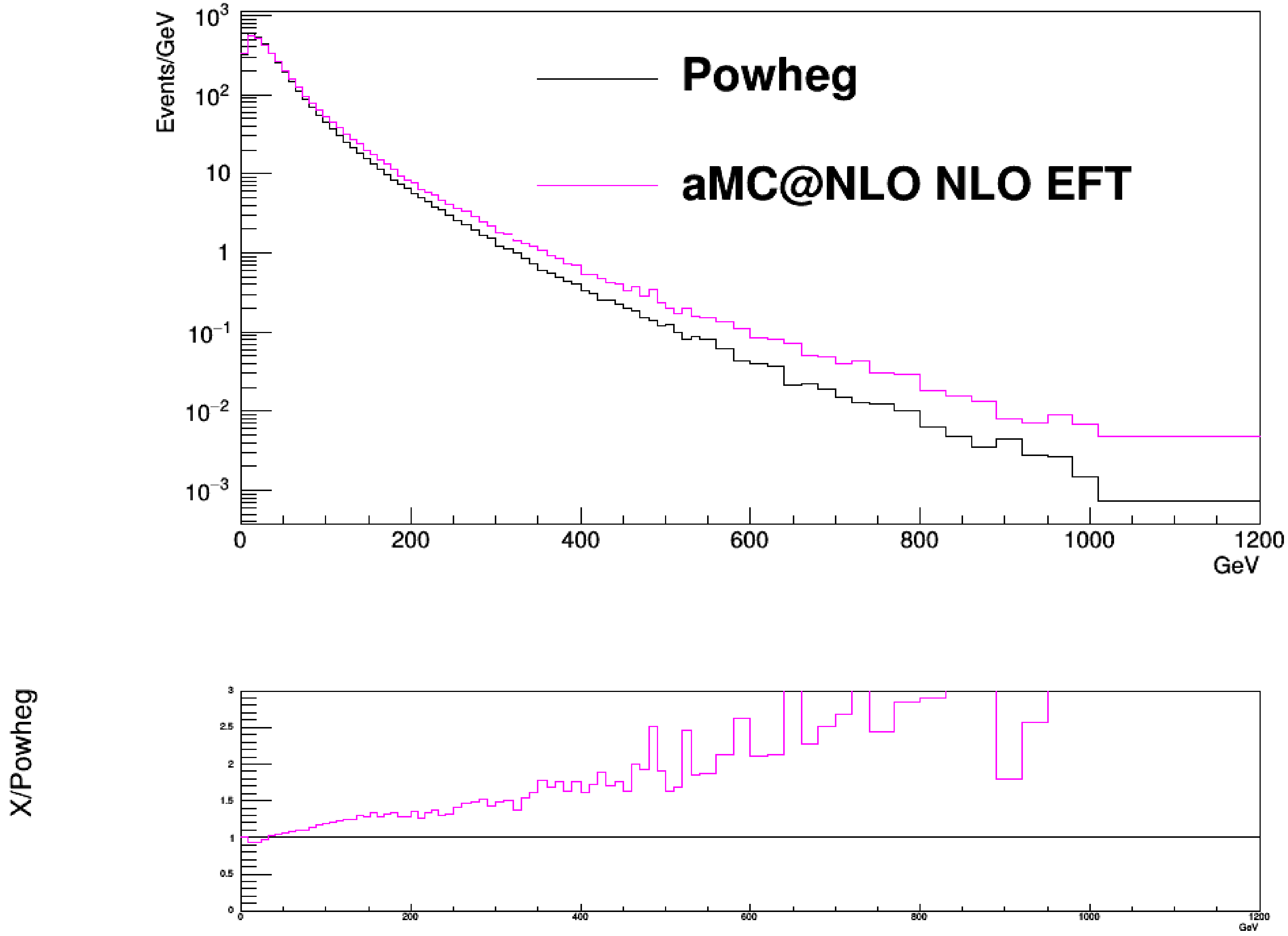
What is the best Higgs p_T :Options

- The key is to identify two different effects
  - Finite top mass effect
  - NNLO differential corrections

- What are the orders known:
  - Differential EFT : NNLO H+1jet production
  - Finite top mass : almost NLO
  - At MC level EFT : NLO H+0/1/2jet
  - At MC level finite top mass : LO 0/1/2

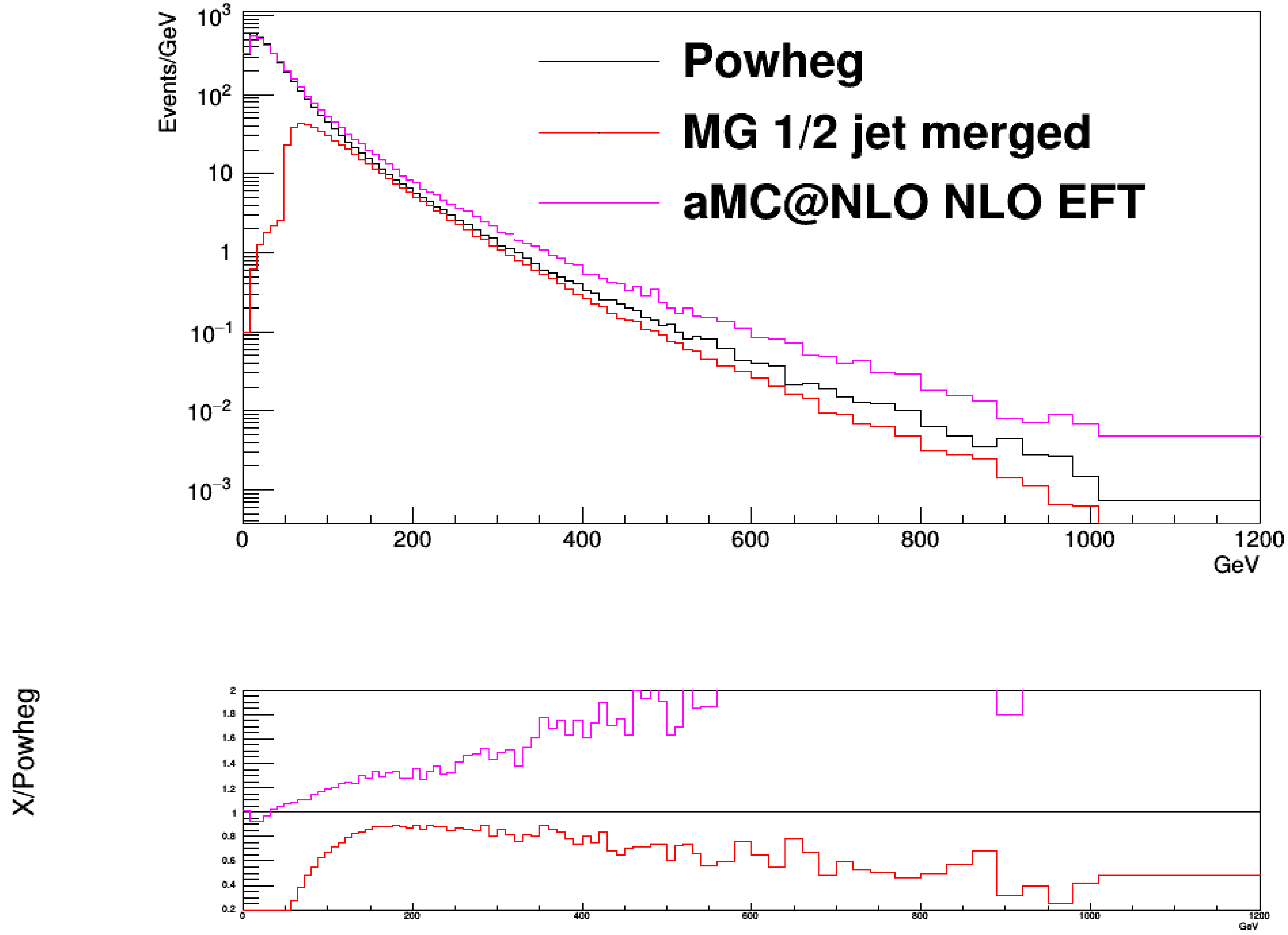
Going to EFT

- When going to EFT large gain



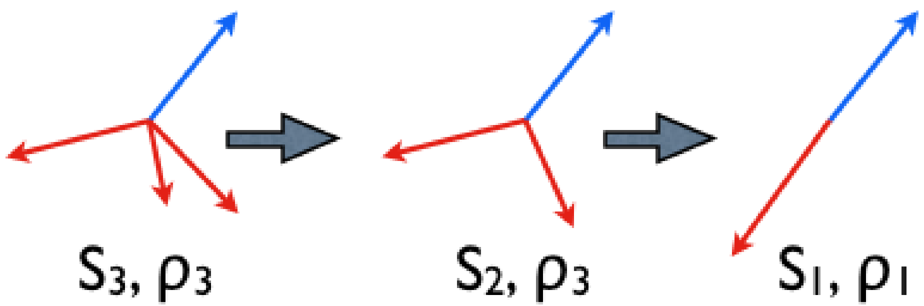
Going to EFT

- Adding the finite top mass merged LO its lower



Idea: Reduce the dependence to the merging scale MS.

- Start by generate events with $N_1..N_2$ ME partons, hard and well separated
- Assume an event with n ME partons, reconstruct the possible shower histories, pick one according to the occurence probabilities



- Each clustering step i is characterized by the emission scale ρ_i, reweight by the product of $\alpha_s(\rho_i)/\alpha_s(ME)$
- For i=2..n (!=N_2):
  - Generate one emission ρ with ρ_i as starting scale.
  - If $\rho > \rho_{i+1} \Rightarrow$ reject the event.
    - This is equivalent to the product of Sudakovs $\Pi(\rho_i, \rho_{i+1})$, i=2..n-1.
- if not HME: generate an emission at $\rho < \rho_n$, if $\rho >$MS $\Rightarrow$ reject the event.
- if HME, accept the event and start shower with ρ_n.

Practical use: Madgraph 5

Essential parameters for matching/merging:

- ickkw
  - Applies the α_s reweighting at each QCD vertex in the ME calculation. K_T-MLM, Shower-K_T:1 CKKW-L, UMEPS:0
- xqcut
  - Defines the minimal K_T between the partons (+beam) at ME level.
- auto_ptj_mjj
  - Set to False: leaves the xqcut be the only cut applied to ME partons$\Rightarrow$ ptj, mmjj=0
- maxjetflavor
  - QCD partons with pdgId$\leq$maxjetflavor are affected by xqcut ptj,etc... Otherwise, affected by ptb, mmbb, etc... That means that for a n-Flavour prediction, maxjetflavor = n

Practical use: main89.cc

- main89ckkwl.cmnd: CKKWL. Essential parameters are
  - `Merging:TMS = XXX.`
    - The merging scale
  - `Merging:Process = UUU`
    - Type of process, e.g. pp>LEPTONS,NEUTRINOS
  - `Merging:nJetMax = WWW`
    - Maximal number of additional jets in the matrix element
  - `Merging:doPTLundMerging = on`
    - Set the merging scale definition to $P_{T,evol}$ (cfr definition in the manual)
- main89umeps.cmnd: UMEPS. Essential parameters are
  - `Merging:TMS = XXX.`
  - `Merging:Process = (e.g.) pp>LEPTONS,NEUTRINOS`
  - `Merging:nJetMax = WWW`
  - `Merging:doUMEPSTree = on`
    - Reweight events according to the UMEPS prescription for tree-level configurations)