UNIVERSITY OF NOTRE DAME

Advisor: Prof. Kevin Lannon

Committee: Prof. Graham Peaslee,
Prof. Mike Hildreth, Prof. Adam Martin

Search for new physics impacting associated top production in multilepton final states using the framework of effective field theory

Thesis Defense: February 21, 2023

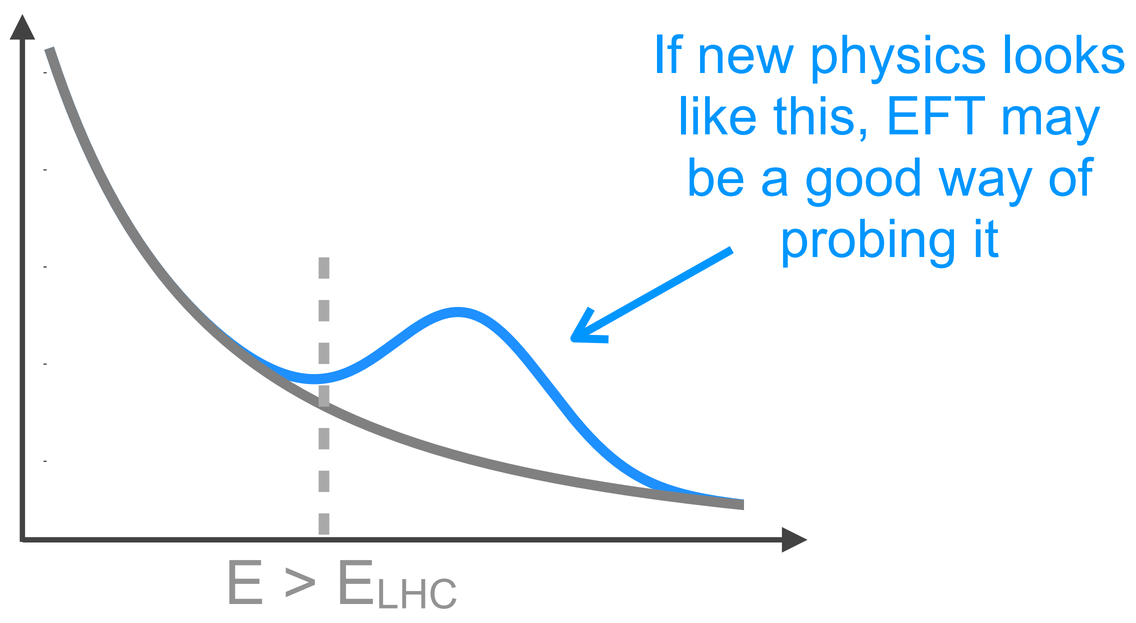
Kelci Mohrman

Motivation for indirect searches for new physics

- There are strong indications that the SM is not the complete description of nature, but there's no guarantee that the new particles would be light enough to be produced on shell at the LHC

- Indirect methods of probing higher mass scales are thus becoming increasingly interesting in the search for new physics at the energy frontier

- Effective field theory (EFT) is an example of such an indirect probe, and offers a model independent method of extending the discovery reach of the LHC



Brief introduction to EFT

- In general an effective theory is a low energy approximation for a more fundamental underlying theory
- SM EFT treats the SM as the lowest order term in an expansion of higher-dimensional operators, that describe physics at a scale Λ, interacting with a strength determined by a dimensionless parameter called a Wilson coefficient *c*
- If all Wilson coefficients (WCs) are 0, the SM Lagrangian is recovered thus –> a non-zero WC indicates new physics

Wilson Coefficient (strength of interaction)

Energy scale of the new physics

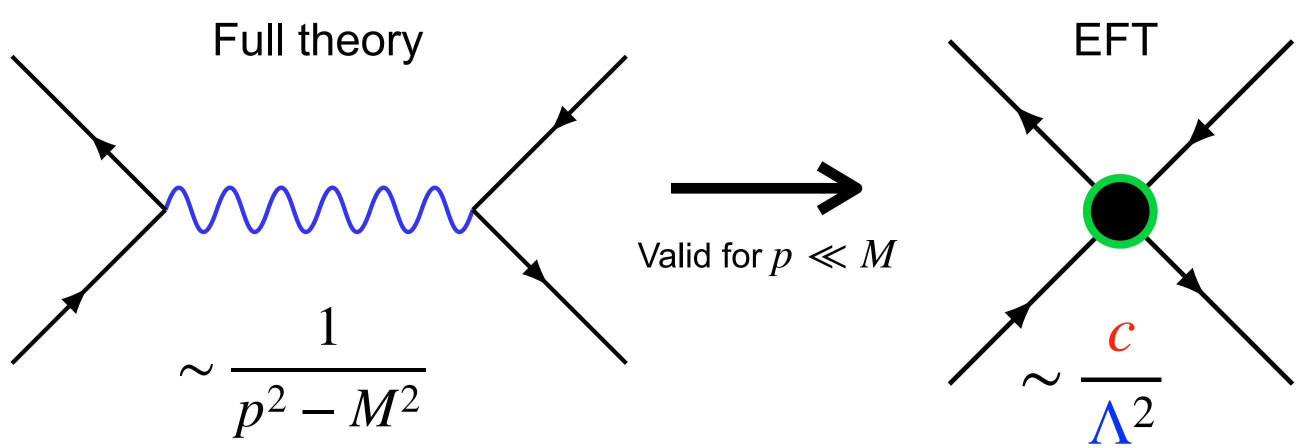
Operators are built of products of SM fields and their derivatives

$$\mathscr{L}_{EFT} = \mathscr{L}_{SM} + \sum_i \frac{c_i}{\Lambda}\mathcal{O}_i^{(5)} + \sum_i \frac{c_i}{\Lambda^2}\mathcal{O}_i^{(6)} + \dots$$

We focus on the **dim 6** terms, as they are the lowest order terms that contribute

Using $\mathscr{L}_{EFT} = \mathscr{L}_{SM} + \sum_i \frac{c_i}{\Lambda} \mathcal{O}_i^{(5)} + \sum_i \frac{c_i}{\Lambda^2} \mathcal{O}_i^{(6)} + \ldots$ to describe interactions

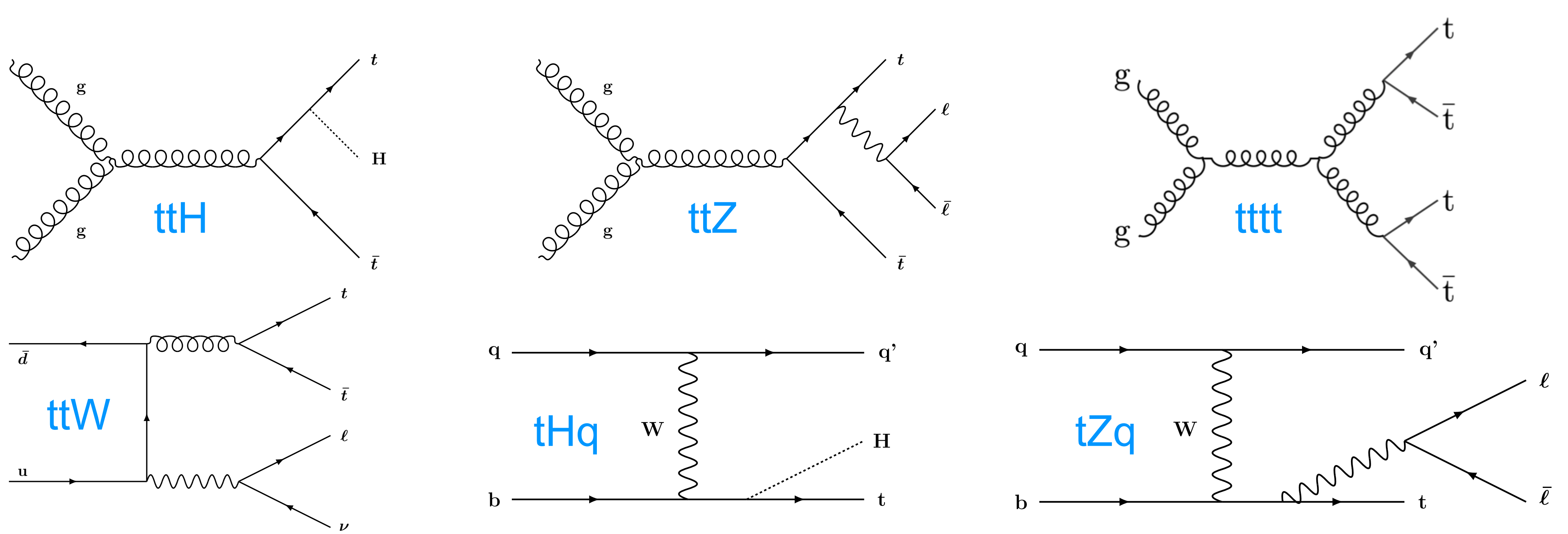
- Example: If a heavy particle can't be produced on-shell at the LHC, may be hard to find via a direct search, but EFT can describe the interaction with an EFT operator, where the strength of the interaction is determined by the WC



- "Model independent", do not need to know details about the new physics in order to describe the effects
- Provides a systematic way of describing heavy new physics

Overview of analysis goals

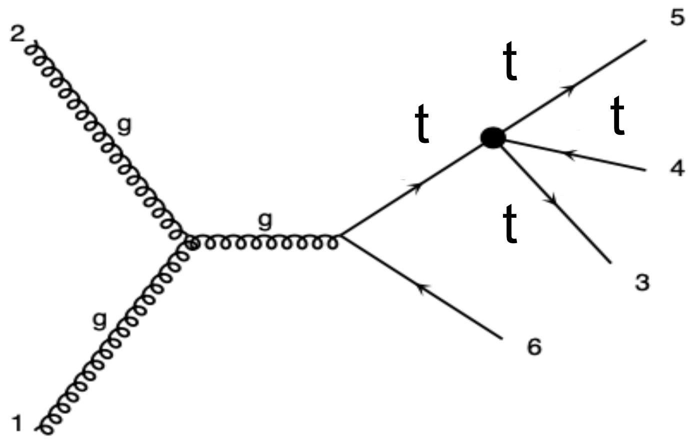
- EFT can be used to study many different sectors, but we aim to use EFT to probe new physics impacting associated top production
  - Signal processes: ttH, ttlnu, ttll, tllq, tHq, tttt
  - These processes are relatively rare, involve heavy particles, and may be an interesting region for new physics to be hiding
  - Global approach, aiming to probe all effects of dimension-6 EFT operators (involving top quarks) that can impact these processes



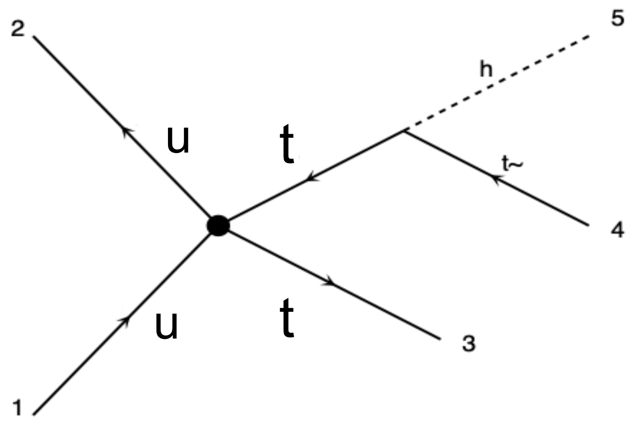
A few example associated top production diagrams

EFT operators impacting associated top processes

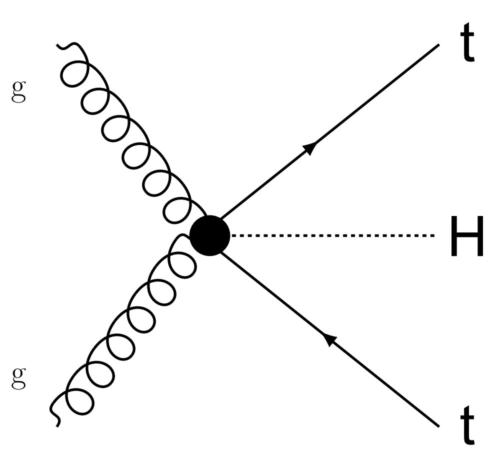
- We study 26 WCs that significantly impact associated top processes, the operators fall into 4 main categories:



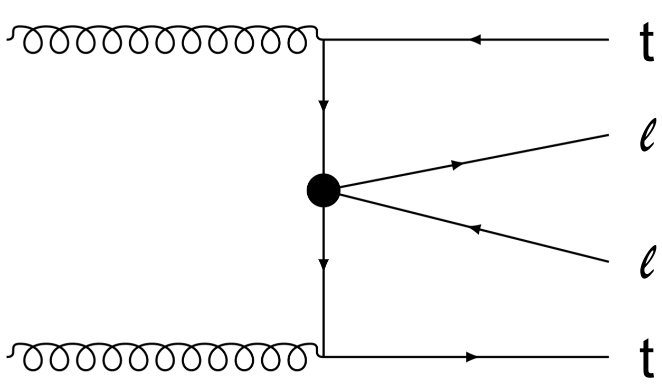
Operators involving 4 heavy quarks



Operators involving two heavy quarks and 2 light quarks



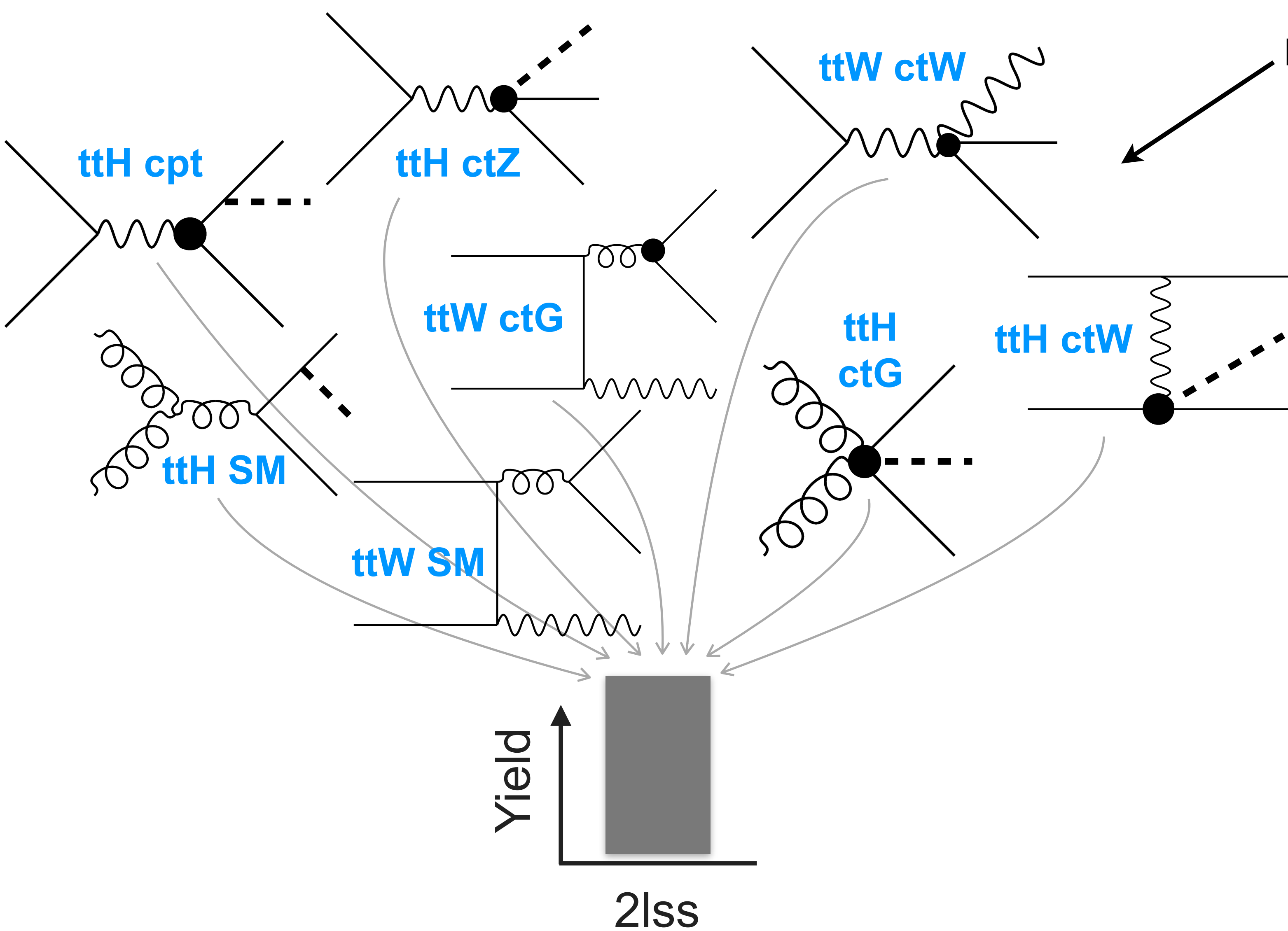
Operators involving heavy quarks and bosons



Operators involving 2 heavy quarks and 2 leptons

Challenges of a multilepton EFT analysis

- We focus on multilepton signatures, many advantages but also leads to analysis-level challenges
  - Multiple signal processes contribute to the same final state signatures
  - Many WCs can affect the processes, interfere with each other and the SM

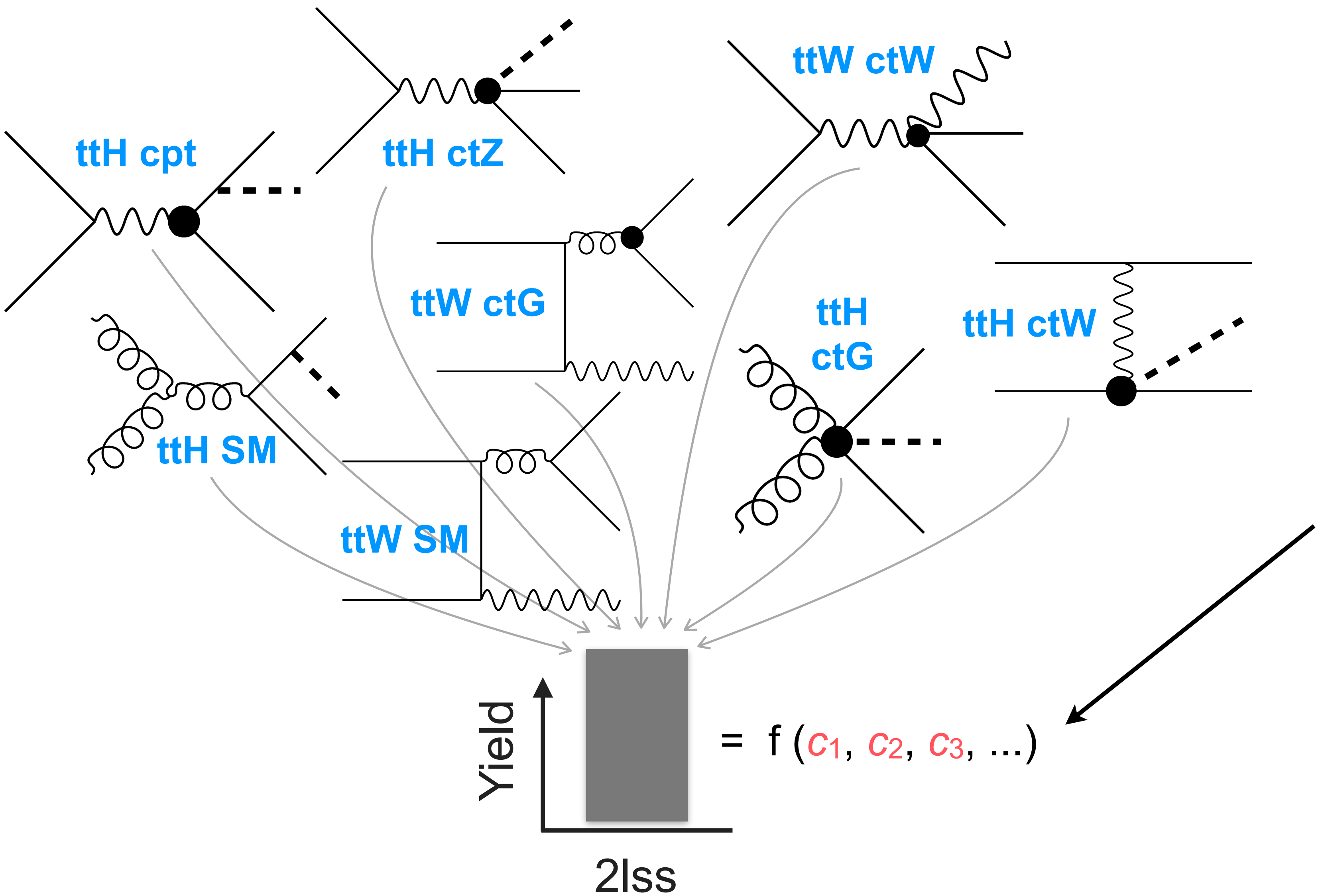


E.g. both ttH and ttW contribute to the 2lss final state, and both are impacted by many WCs (just a few examples of WCs are shown here)

It is therefore crucial to analyze the effects of all relevant operators across all channels simultaneously

Challenges of a multilepton EFT analysis

- We focus on multilepton signatures, many advantages but also leads to analysis-level challenges
  - Multiple signal processes contribute to the same final state signatures
  - Many WCs can affect the processes, interfere with each other and the SM



Want to know number of predicted events in a given detector-level observable bin, as a function all of the WCs, i.e. Yield = Yield(c_1, c_2, c_3, ...)

With this info, we can compare prediction to the data and find the best fit value for the WCs

Quadratic dependence of the weights on the WCs

- Matrix element can be written as sum of SM and new physics:

$$\mathscr{M} = \mathscr{M}_{SM} + \sum_i c_i \mathscr{M}_i$$

c_i are the Wilson coefficients

- Since, $\sigma \propto \mathscr{M}^2$, cross sections depend quadratically on the Wilson coefficients c_i

- E.g. for just one c_1:

SM

Interference with SM

Pure NP

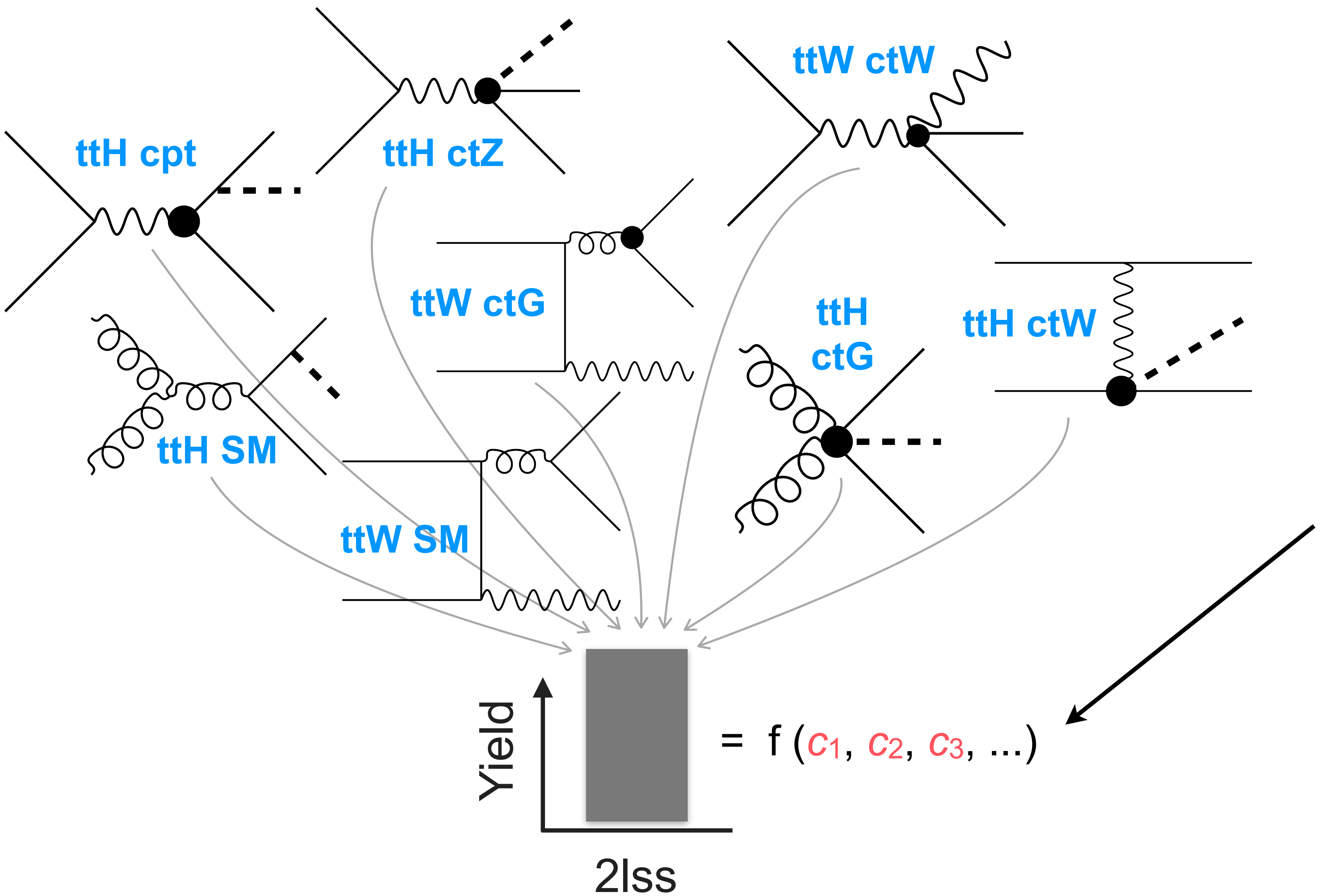
$$d\sigma(c_1) \propto |\mathscr{M}_{SM} + c_1 \mathscr{M}_1|^2 \propto s_0 + s_1 c_1 + s_2 c_1^2$$

In general this is an *n*-dimensional quadratic (*n* = number of WCs)

- Each event's weight will also depend quadratically on the WCs, which we can find via a reweighting procedure

Challenges of a multilepton EFT analysis

- We focus on multilepton signatures, many advantages but also leads to analysis-level challenges
  - Multiple signal processes contribute to the same final state signatures
  - Many WCs can affect the processes, interfere with each other and the SM

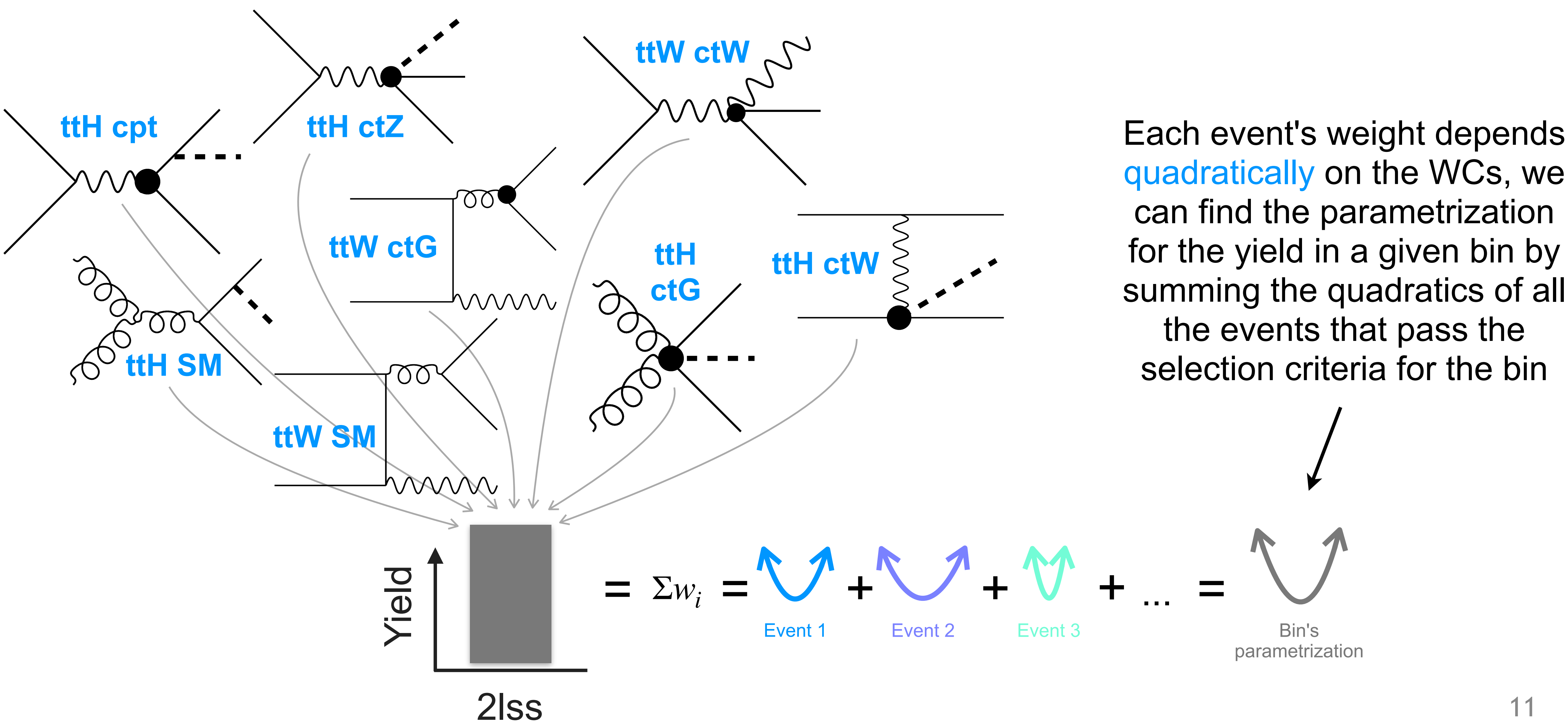


Want to know number of predicted events in a given detector-level observable bin, as a function all of the WCs, i.e. Yield = Yield(c_1, c_2, c_3, ...)

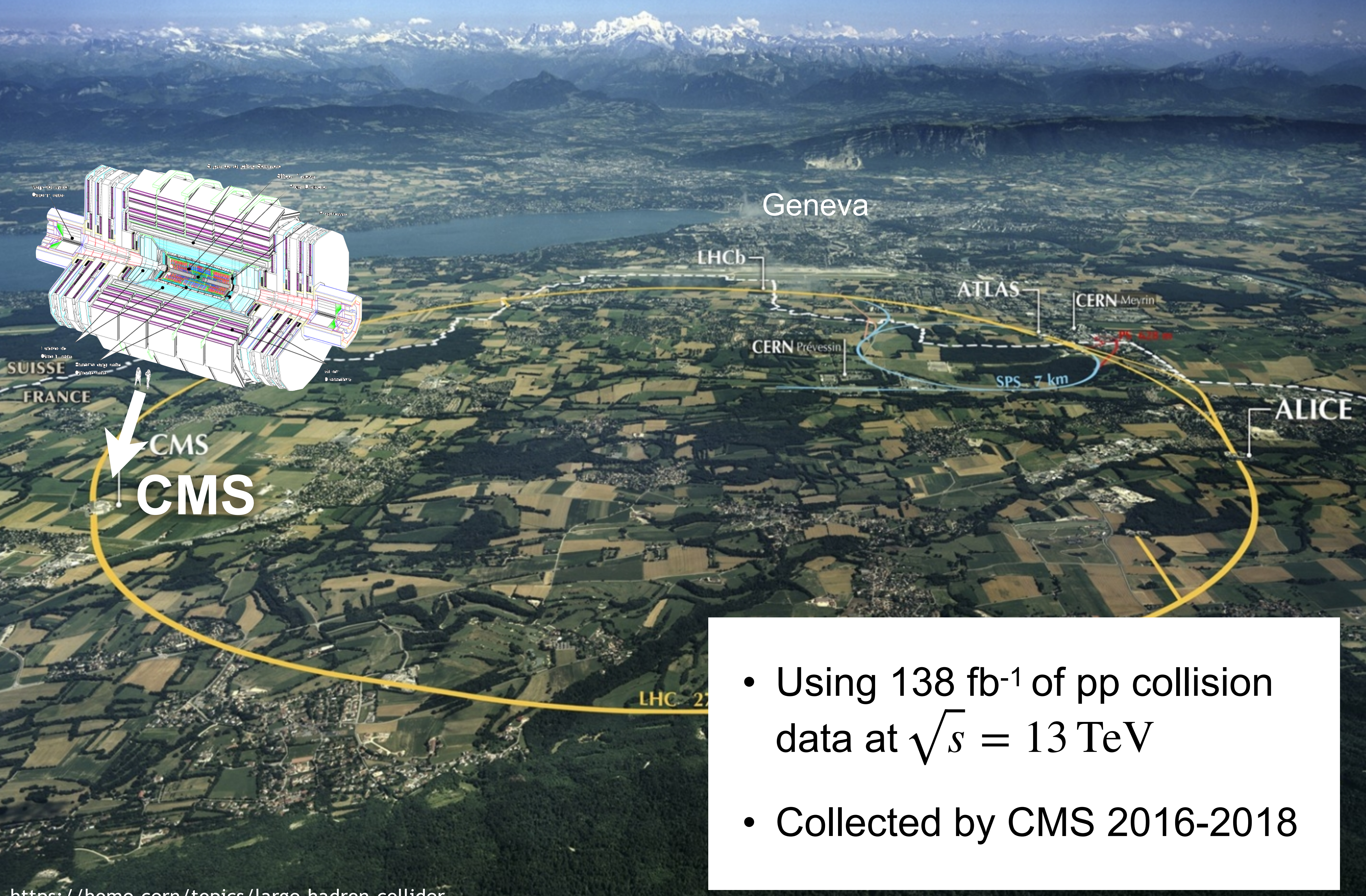
With this info, we can compare prediction to the data and find the best fit value for the WCs

Challenges of a multilepton EFT analysis

- We focus on multilepton signatures, many advantages but also leads to analysis-level challenges
  - Multiple signal processes contribute to the same final state signatures
  - Many WCs can affect the processes, interfere with each other and the SM



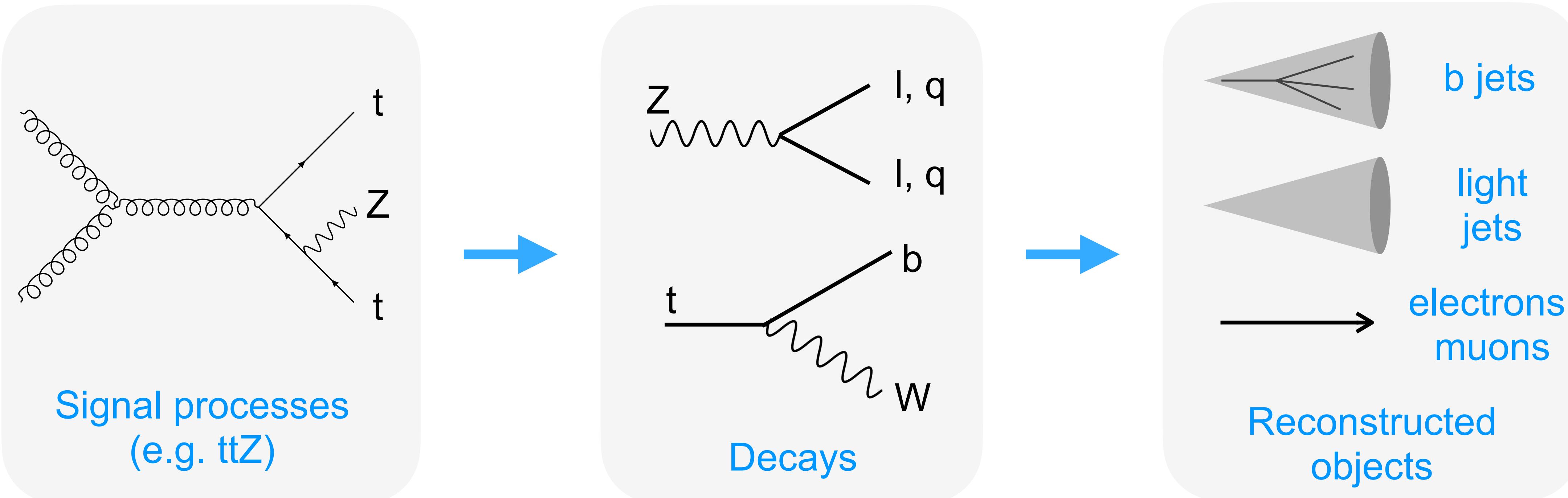
The LHC and CMS



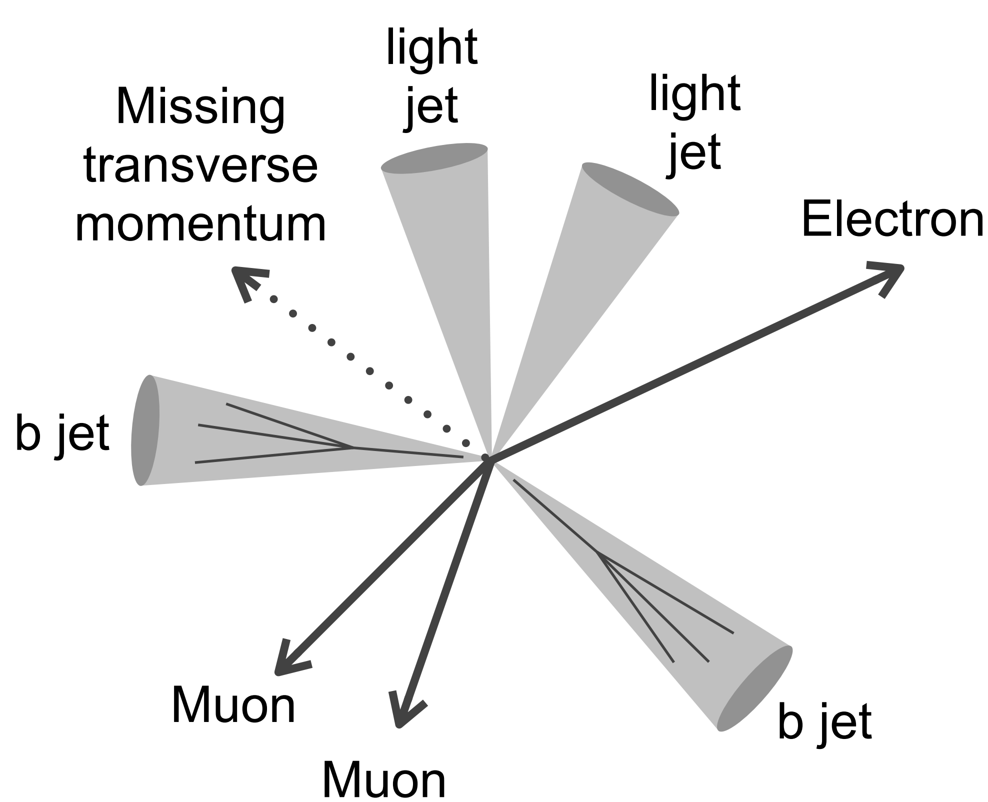
- Using 138 fb^{-1} of pp collision data at $\sqrt{s} = 13\,\mathrm{TeV}$
- Collected by CMS 2016-2018

https://home.cern/topics/large-hadron-collider

Experimental signatures

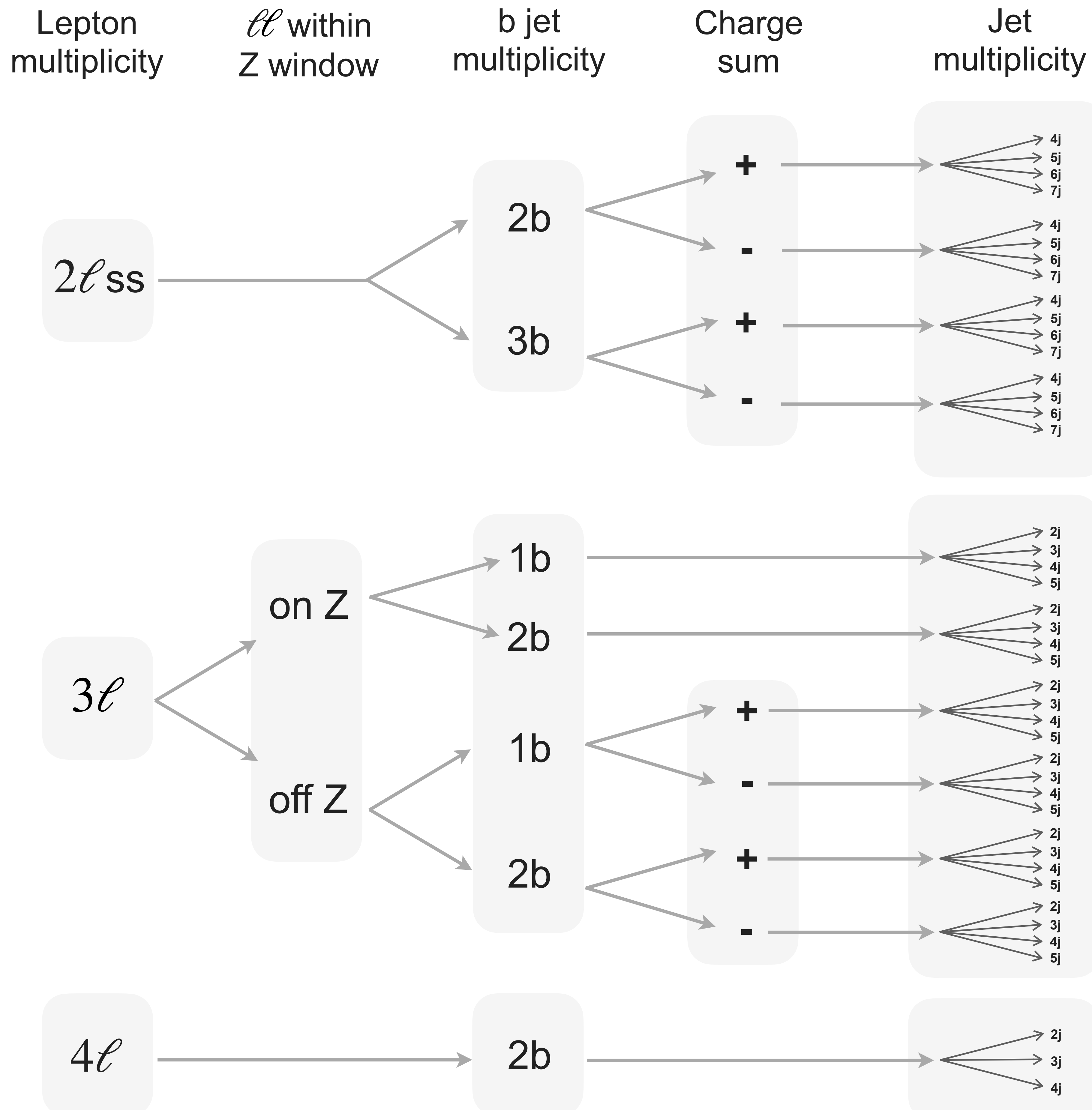


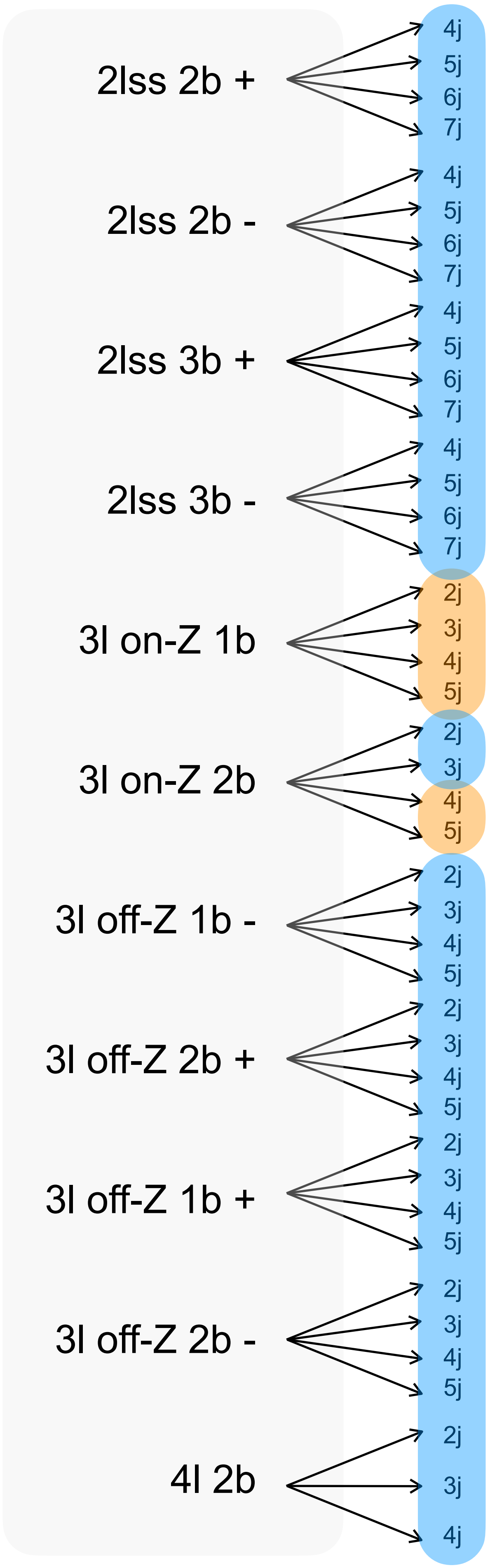
- We're interested in leptonic decays of associated top processes
- These lead to signatures of leptons, jets, and b jets



Event selection summary

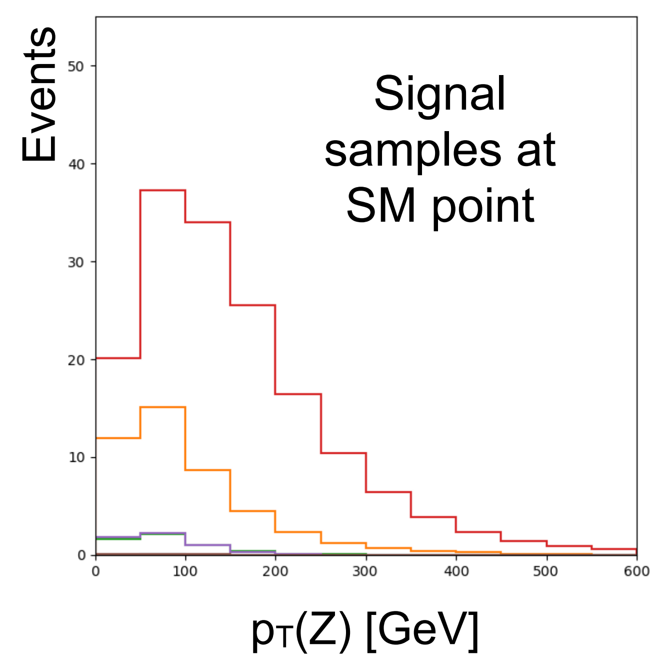
Since EFT impacts each process differently and the goal is to gain sensitivity to the EFT effects: The purpose of our event selection categorization is to differentiate between the admixture of processes

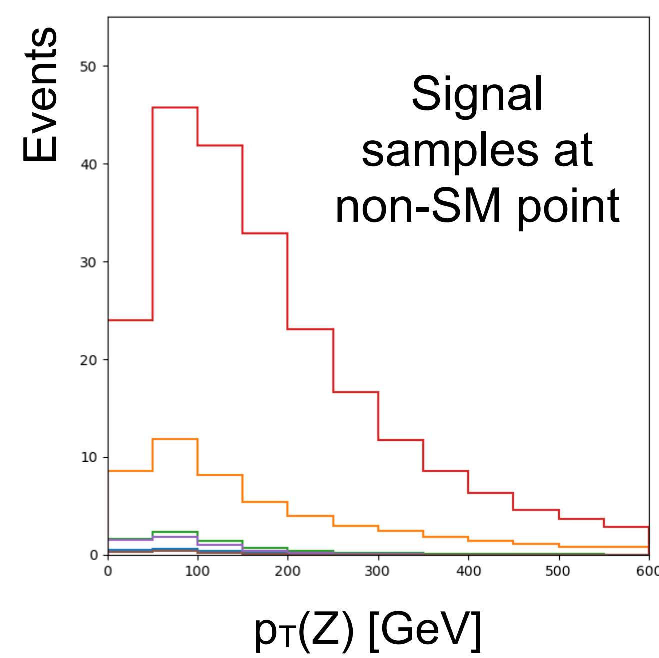




Improving sensitivity with differential distributions

- In order to improve sensitivity, we fit a differential kinematic distribution for each of the 43 categories
- Use different variables ($p_T(lj)_0$, $p_T(Z)$) in different regions to optimize sensitivity to EFT effects

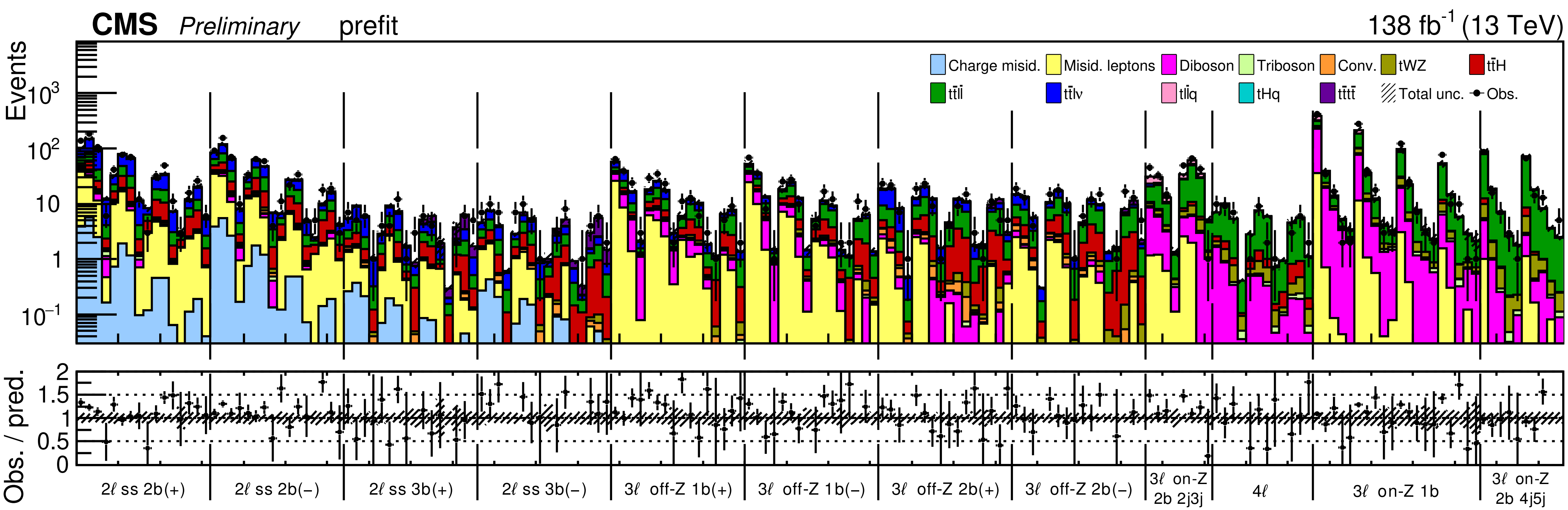




When we reweight to a non-SM point, we can see the shape and normalization of the distribution changes

Summary of event selection and categorization

- Binning the 43 categories according to these kinematical distributions results in 178 total bins
- The predicted yield in each bin depends quadratically on the 26 WCs
- The goal is to turn the knobs on the 26 WCs to figure out the ranges of WC values values that lead to yields that are consistent with the observed data



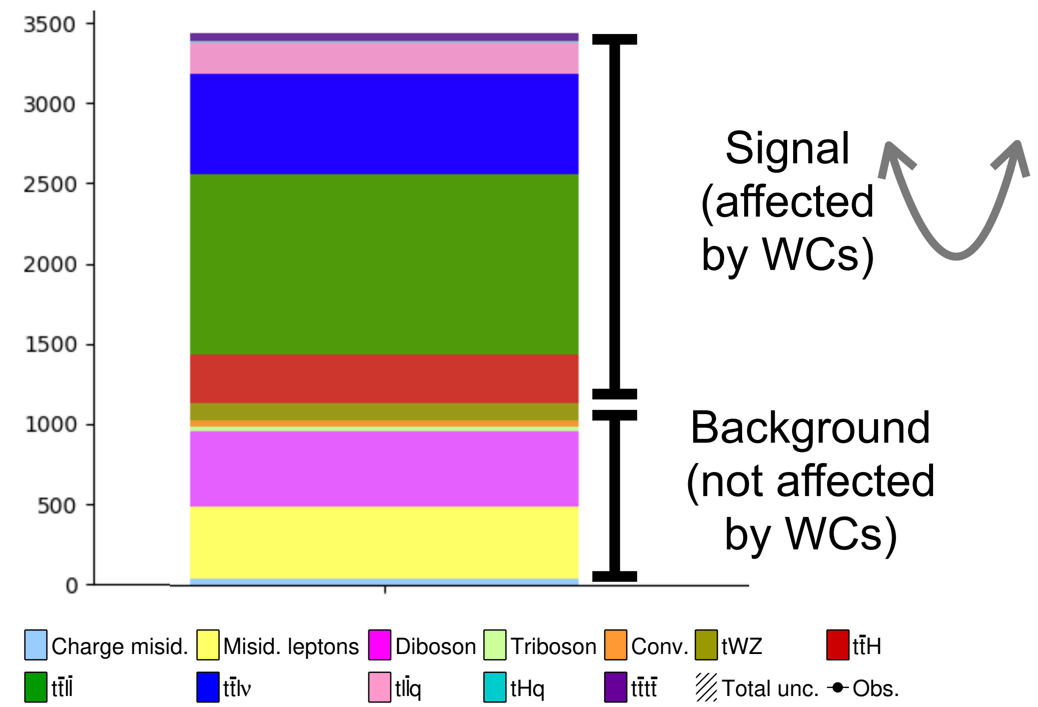
Backgrounds and systematics

- Signal processes (impacted by the EFT) are not the only contributions to our signal regions → about 1/3 of yield is background

- Main backgrounds: From processes that lead to the same final states as our signal processes, and from misidentified leptons

- Model backgrounds with combination of MC and data-driven approaches

Yield integrated over all signal regions



- Various systematic uncertainties (impacting signal and background) also must be accounted for in the fit

The likelihood

- The likelihood characterizes the probability of measuring the observed number of events, given the theory i.e. $L = P(\text{data}|\text{theory})$
- Write the likelihood as a product over the 178 bins in the analysis, each treated as an independent Poisson measurement, with a mean corresponding to the predicted yield (which is a quadratic function of the WCs)
- We want to find the WC values that best agree with the data (i.e. that maximize the likelihood)

Product over the N=178 bins in the analysis

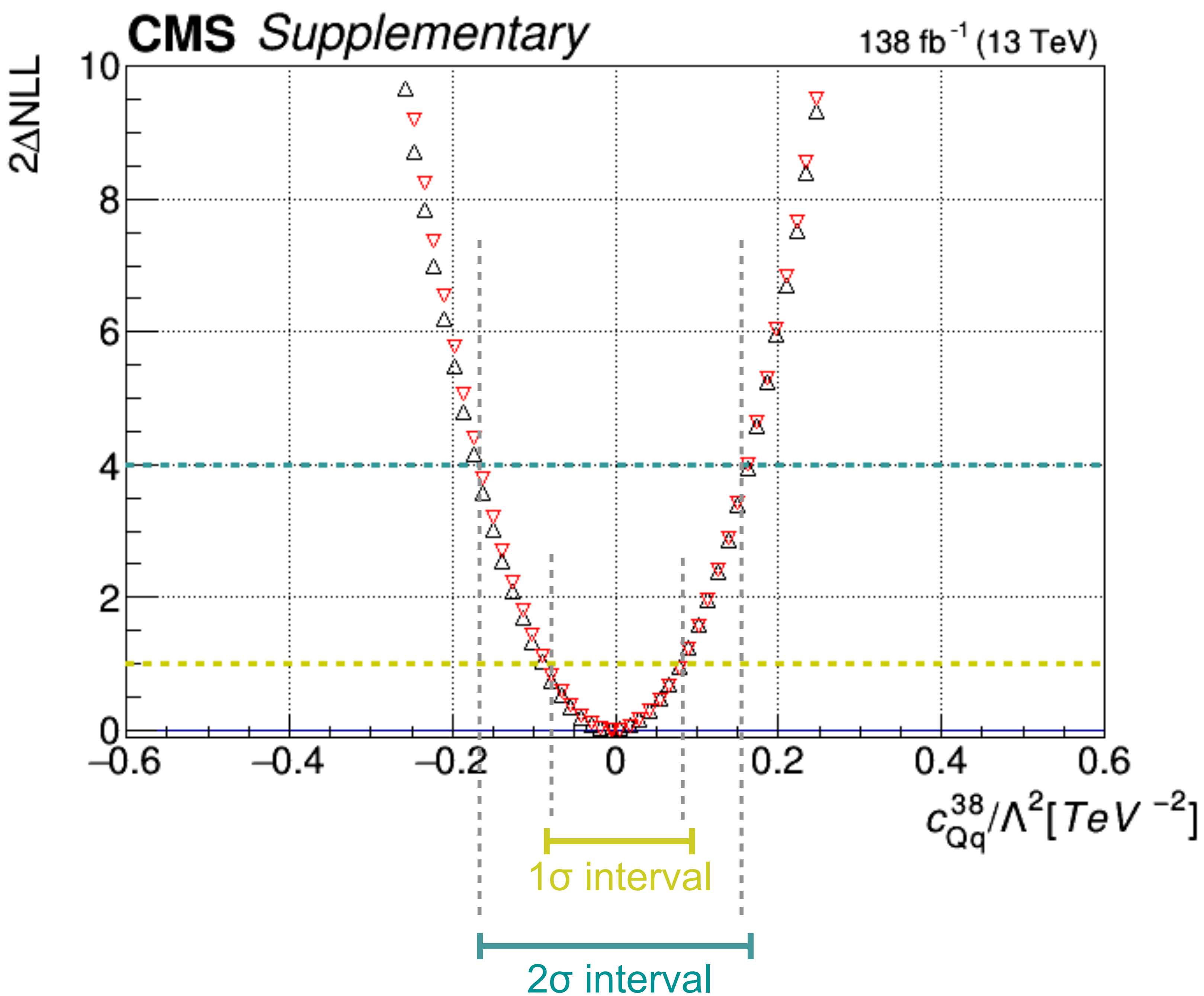
n is the observed yield in the bin

m is the prediction in each bin, depends quadratically on the WCs θ, and the dependence on θ is different in every bin

$$L = \prod_{i=1}^{N} \frac{m(\theta)_i^{n_i}}{n_i!} e^{-m(\theta)_i}$$

Extracting the confidence intervals

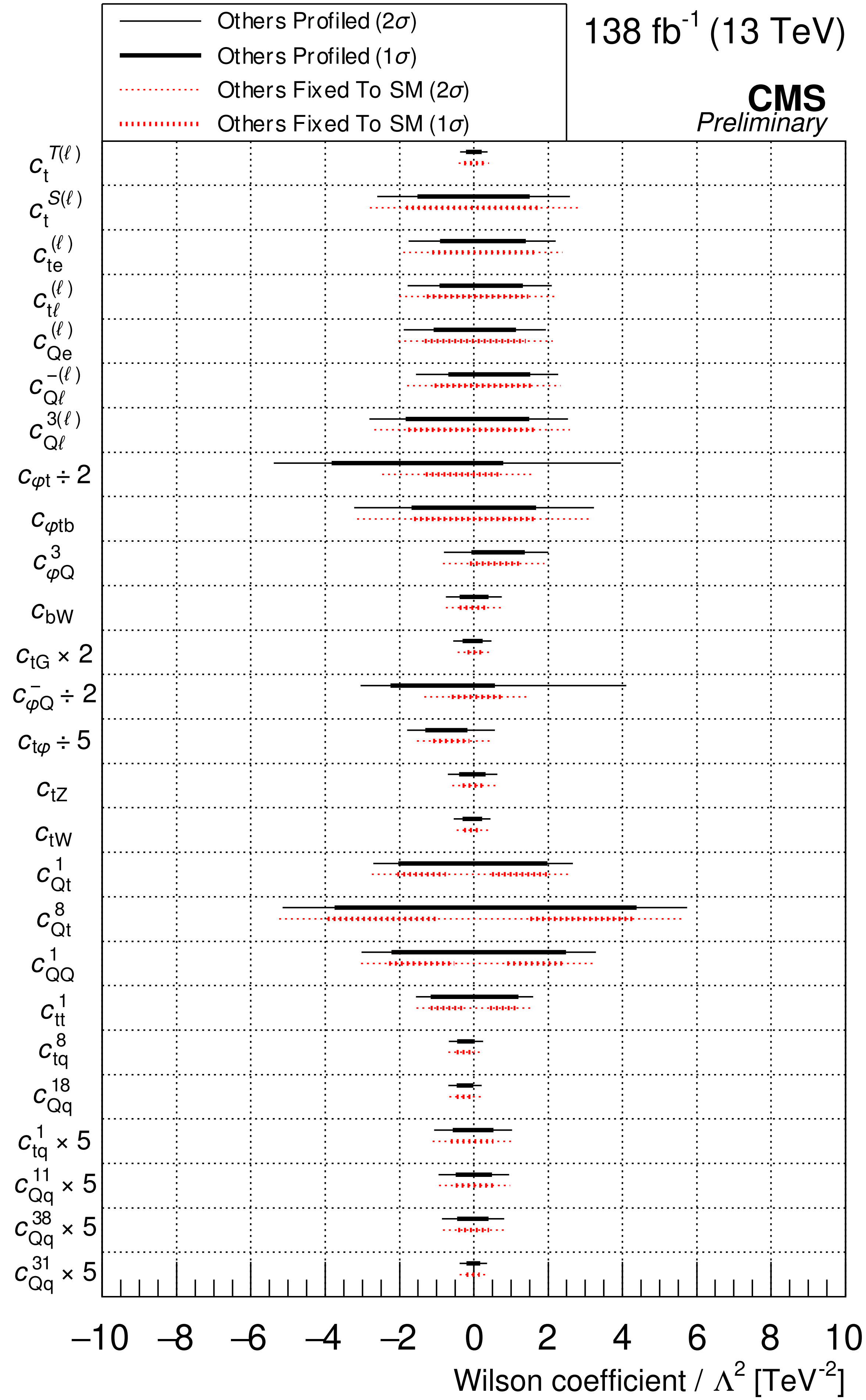
- For each WC, we scan across a range of values, profiling the other 25 WCs
- We can then read off the best fit point and the one and two standard deviation confidence intervals from the scans



Since WC=0 is within the interval, this result is consistent with the SM

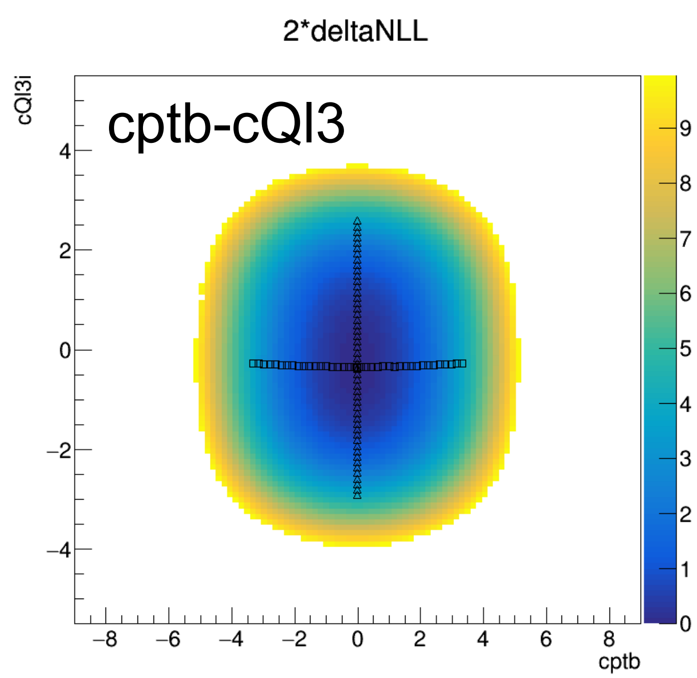
Results

- Results are consistent with the SM
- While these results do not indicate signs of new physics, we can still use them to gain insights about the implications
- For most of the WCs, sensitivity is limited by statistics, though a subset of the WCs are limited by systematics

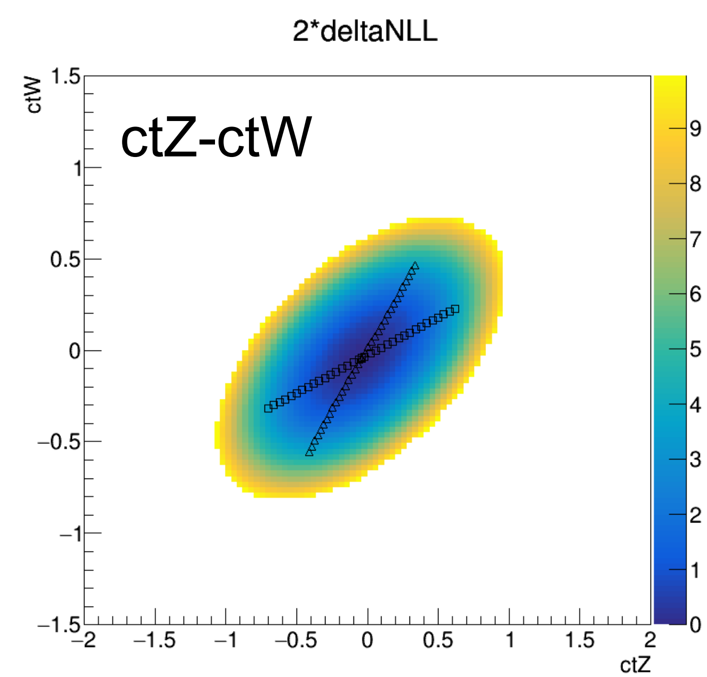


Discussion of results: Correlated WCs

- We've looked at the results for the 1d scans, but in principle would be interesting to explore the full 26d likelihood "surface"
- Very difficult to explore high-dimensional spaces, but we can at least look at pairs of WCs in 2d
- 325 pairs in total, so we identify the potentially interesting pairs and make 2d scans for these, for example:



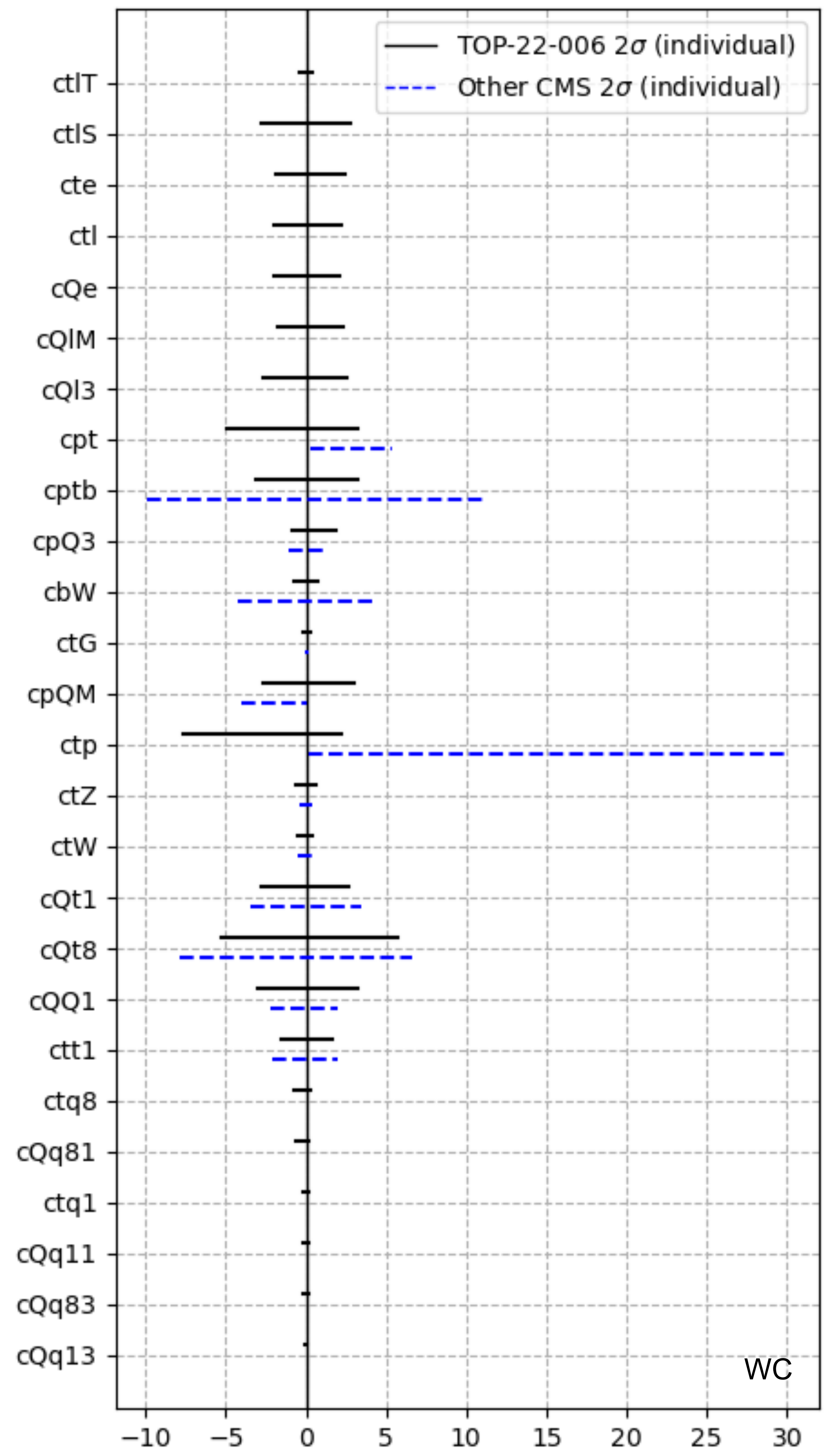
Example of two WCs that do not have significant interference



Strong linear correlation, due to large interference between these two WCs

Discussion of results: Comparisons to other results

- For most of the WCs, the results of this analysis are tighter than or competitive to other CMS results (for 6 of the WCs we set the only limits)
- Could be useful to perform a combination with other analyses that study orthogonal signal regions
- Can also learn from other analysis to apply techniques and ideas to next iteration of this analysis



Discussion of results in terms of energy scale

- The limits set by the analysis are on the ratio c/Λ^2, recall EFT Lagrangian:

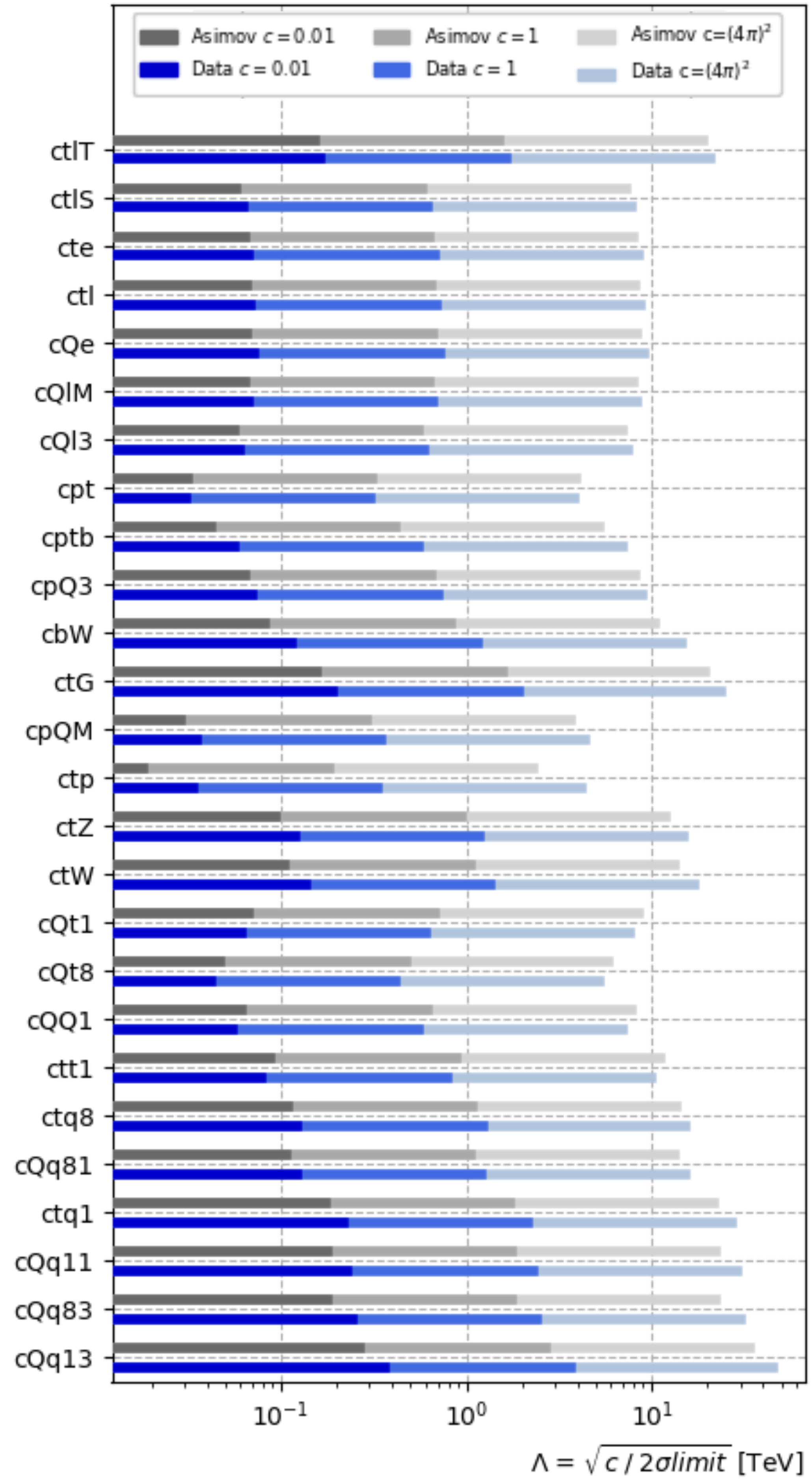
$$\mathscr{L}_{EFT} = \mathscr{L}_{SM} + \sum_i \frac{c_i}{\Lambda^2}\mathcal{O}_i^{(6)}$$

- We can assume a value for c, and explore the implications on the energy scale Λ:

$$\text{Observed limit} = \frac{c}{\Lambda^2} \quad \Longrightarrow$$

$$\Lambda = \sqrt{\frac{c}{\text{Observed limit}}}$$

- E.g. if c=1, the energy range explored by this analysis generally extends to $\Lambda = \sim 1$ TeV



Summary and future directions

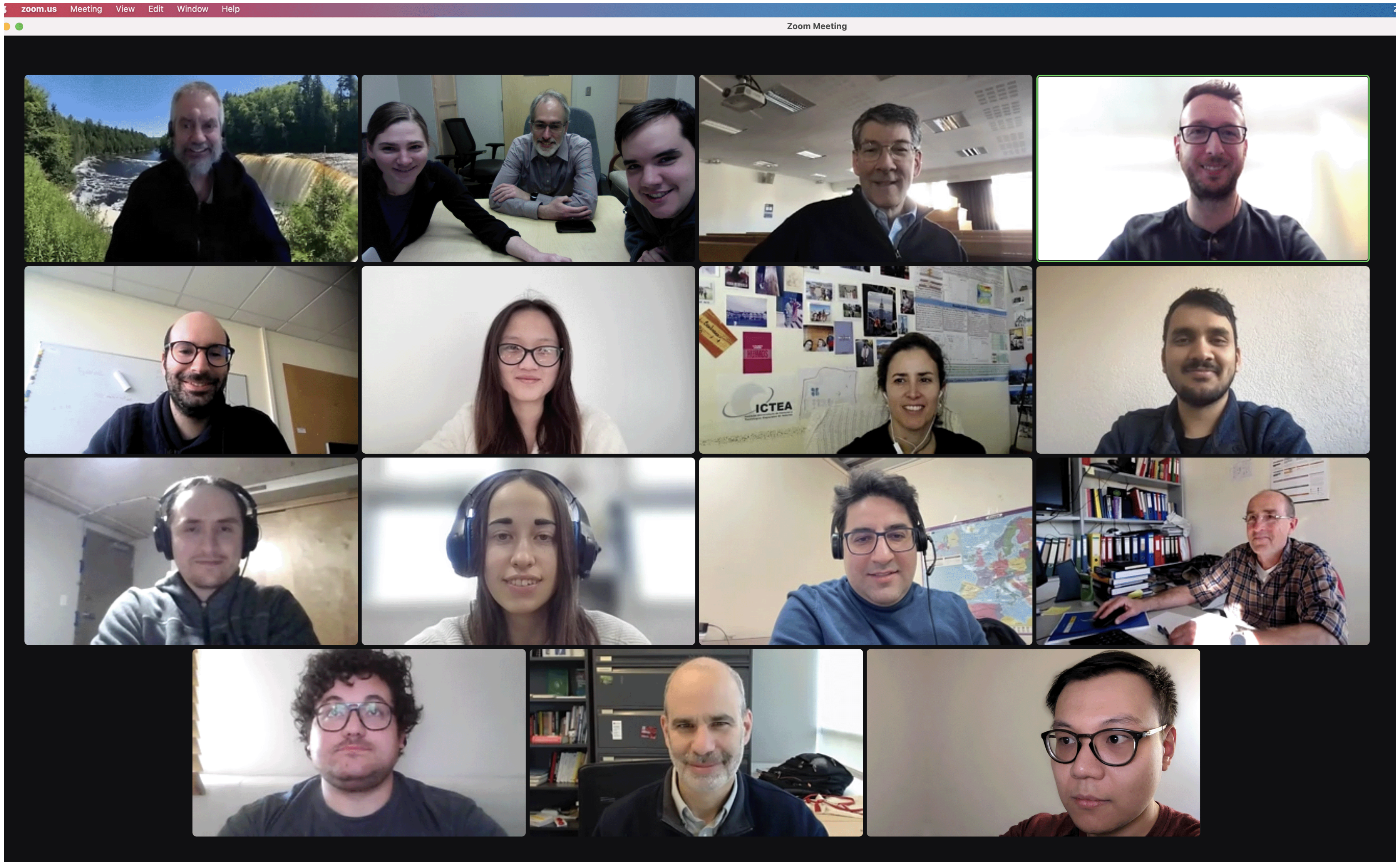
- This analysis has searched for new physics impacting associated top production in multilepton final states within the context of EFT
    - The results are consistent with the SM

- There are many directions in which the analysis could be improved and expanded, e.g.:
    - Collect more data
    - Improvements in EFT modeling
    - Optimizations of categorizations and kinematic variables
    - Targeting more signal processes and other final states

- This analysis represents one small step towards the eventual goal of a global EFT combination across all sectors at the LHC
    - Such a global combination will probe energy ranges beyond what is accessible on-shell, and could result in the discovery of new physics
    - Even if the results remain consistent with the SM, the comprehensive limits on EFT parameters will constrain theoretical models and point to the energy frontier beyond which new physics yet lie

Thank you!

- I would like to thank the committee
- Thanks to the TOP-22-006 analysis team →
- Thanks to the CCL team and all the folks at the CRC
- And thanks to the theorists with whom we've had the pleasure of collaborating



Aashwin Basnet 1, Ken Bloom 3, Florencia Canelli 5, Enrique Palencia Cortezon 2, Sergio Sánchez Cruz 5, Andrea Trapote Fernández 2, Reza Goldouzian 4, Bárbara Álvarez González 2, Xuan González 2, Mike Hildreth 4, Kevin Lannon 4, John Lawrence 4, Sascha Liechti 5, C. Eddie Mcgrady 4, Kelci Mohrman 4, Hannah Nelson 4, Wynona Wan 4, Andrew Wightman 3, Brian Winer 1, Furong Yan 3, Brent Yates 1, Henry Yockey 4, Mateusz Zarucki 4

1: The Ohio State University 2: University of Oviedo 3: University of Nebraska 4: University of Notre Dame 5: University of Zurich

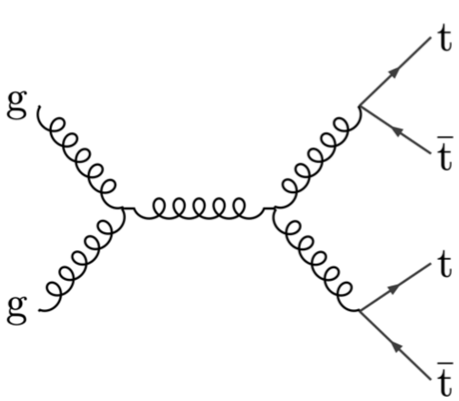
Updates to analysis during CMS review

- During the internal CMS review of this analysis, we were asked to update the 4t cross section (to which we normalize our LO samples) to a new calculation of 13.37 fb (arxiv 2212.03259)
- This request was made Feb 15, 2023, so the analysis results with the updated cross section are not included in the version of my thesis shared with the committee
- However, the impacts on the results are fairly minimal (about a 10% improvement in the 2sigma CIs for the 4-heavy WCs)
- The results in the thesis will be updated (to match our CMS publication) before officially submitting to the grad school

Backup

Comparison to TOP-19-001

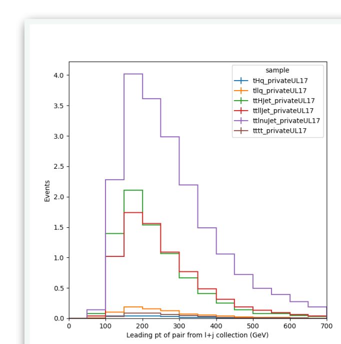
- TOP-22-006 builds on the techniques and tools developed in the TOP-19-001 analysis
- Some of the important improvements over TOP-19-001 include:
    - Including an additional signal process (tttt) in addition to the 5 already included in TOP-19-001
    - Studying 10 additional Wilson coefficients in addition to the 16 probed in TOP-19-001, for a total of 26
    - Using the full Run 2 data set (TOP-19-001 only used 2017)
    - Fitting differential distributions, which allows us to gain additional sensitivity (TOP-19-001 followed a more inclusive approach, fitting 35 categories based primarily on object multiplicities)



More process

Category	WCs
Two-heavy	$c_{t\varphi}, c^{-}_{\varphi Q}, c^{3}_{\varphi Q}, c_{\varphi t}, c_{\varphi tb}, c_{tW}, c_{tZ}, c_{bW}, c_{tG}$
Two-heavy-two-lepton	$c^{3(\ell)}_{Q\ell}, c^{-(\ell)}_{Q\ell}, c^{(\ell)}_{Qe}, c^{(\ell)}_{t\ell}, c^{(\ell)}_{te}, c^{S(\ell)}_{t}, c^{T(\ell)}_{t}$
Two-light-two-heavy	$c^{31}_{Qq}, c^{38}_{Qq}, c^{11}_{Qq}, c^{18}_{Qq}, c^{1}_{tq}, c^{8}_{tq}$
Four-heavy	$c^{1}_{QQ}, c^{1}_{Qt}, c^{8}_{Qt}, c^{1}_{tt}$

More WCs



More differential

Detector and data

Data samples

- We use the full Run 2 UL samples, corresponding to 137.64 fb^{-1}

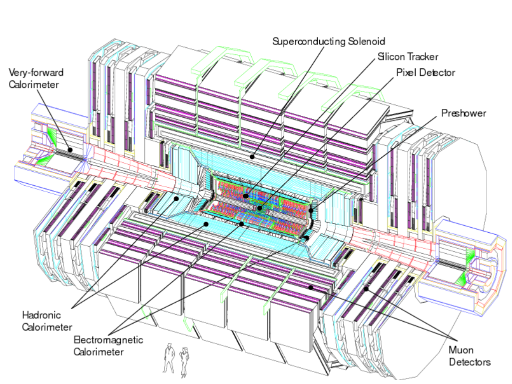
Year	Golden JSON	Int. Lumi (fb^{-1})
2016	Cert_271036-284044_13TeV_Legacy2016_Collisions16_JSON	36.33
2017	Cert_294927-306462_13TeV_UL2017_Collisions17_GoldenJSON	41.48
2018	Cert_314472-325175_13TeV_Legacy2018_Collisions18_JSON	59.83

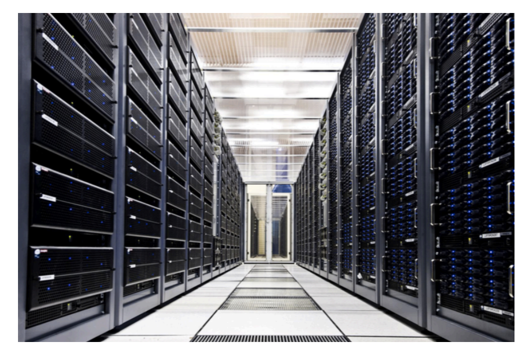
Table 2: JSON files with certified luminosity blocks used for each data-taking year.

- We use the following datasets for each year (full list of triggers can be found in the AN, and in the backup slides):
    - 2016: SingleMuon, SingleElectron, DoubleMuon, DoubleEG, MuonEG
    - 2017: SingleMuon, SingleElectron, DoubleMuon, DoubleEG, MuonEG
    - 2018: SingleMuon, EGamma, DoubleMuon, MuonEG

CMS trigger rate monitoring

- Bunch crossings ~40MHz, → too much data to record/store
- Purpose of trigger: Reduce event rate to manageable ~1kHz while keeping as many potentially interesting events as possible

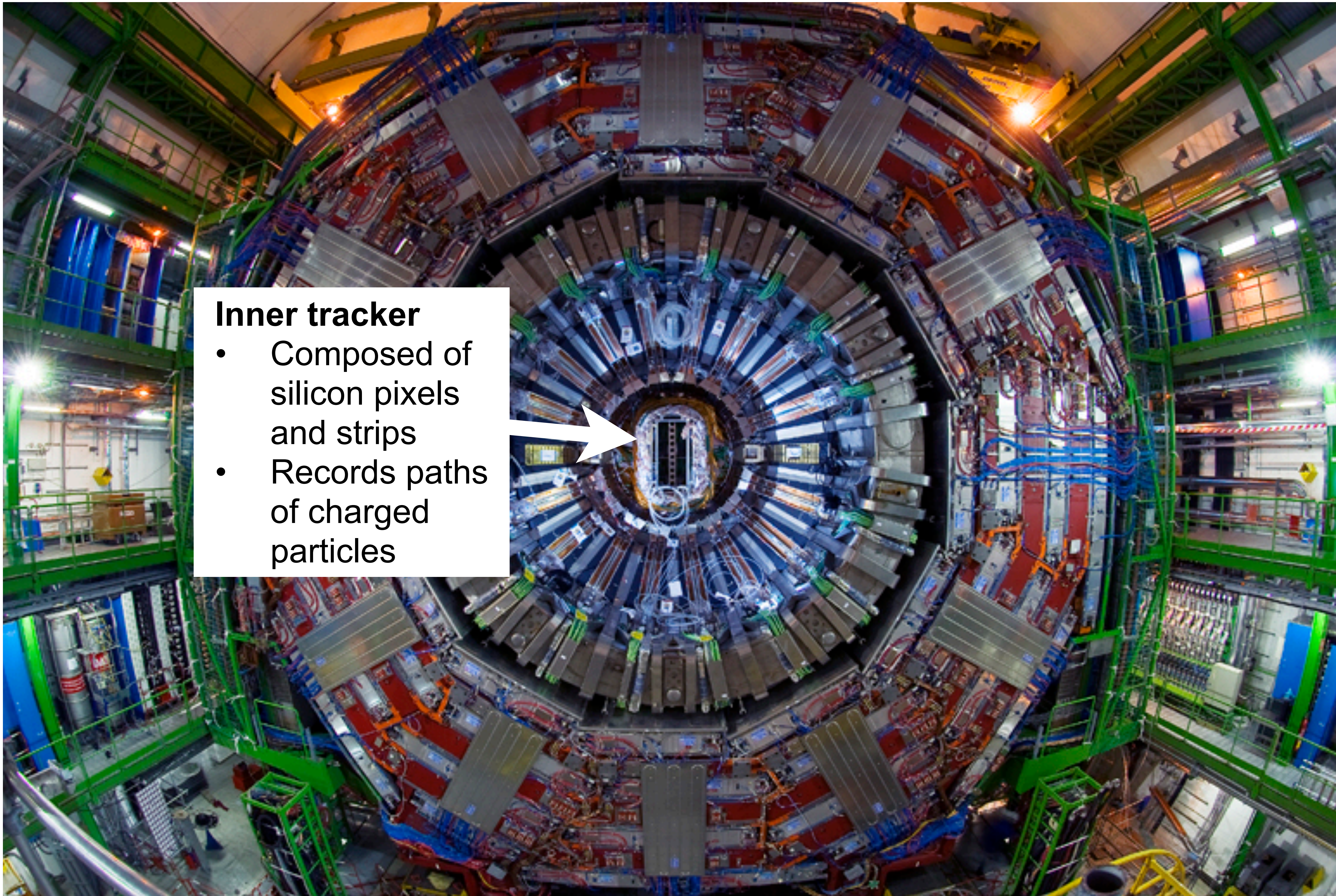




Bunch crossings in CMS → 40 MHz → L1 trigger → 100 kHz → High level trigger → 1 kHz → Offline reconstruction and analysis

Important to monitor the rates of the L1 and HLT paths, unexpected rates can be a useful early warning sign of issues in various parts of the detector

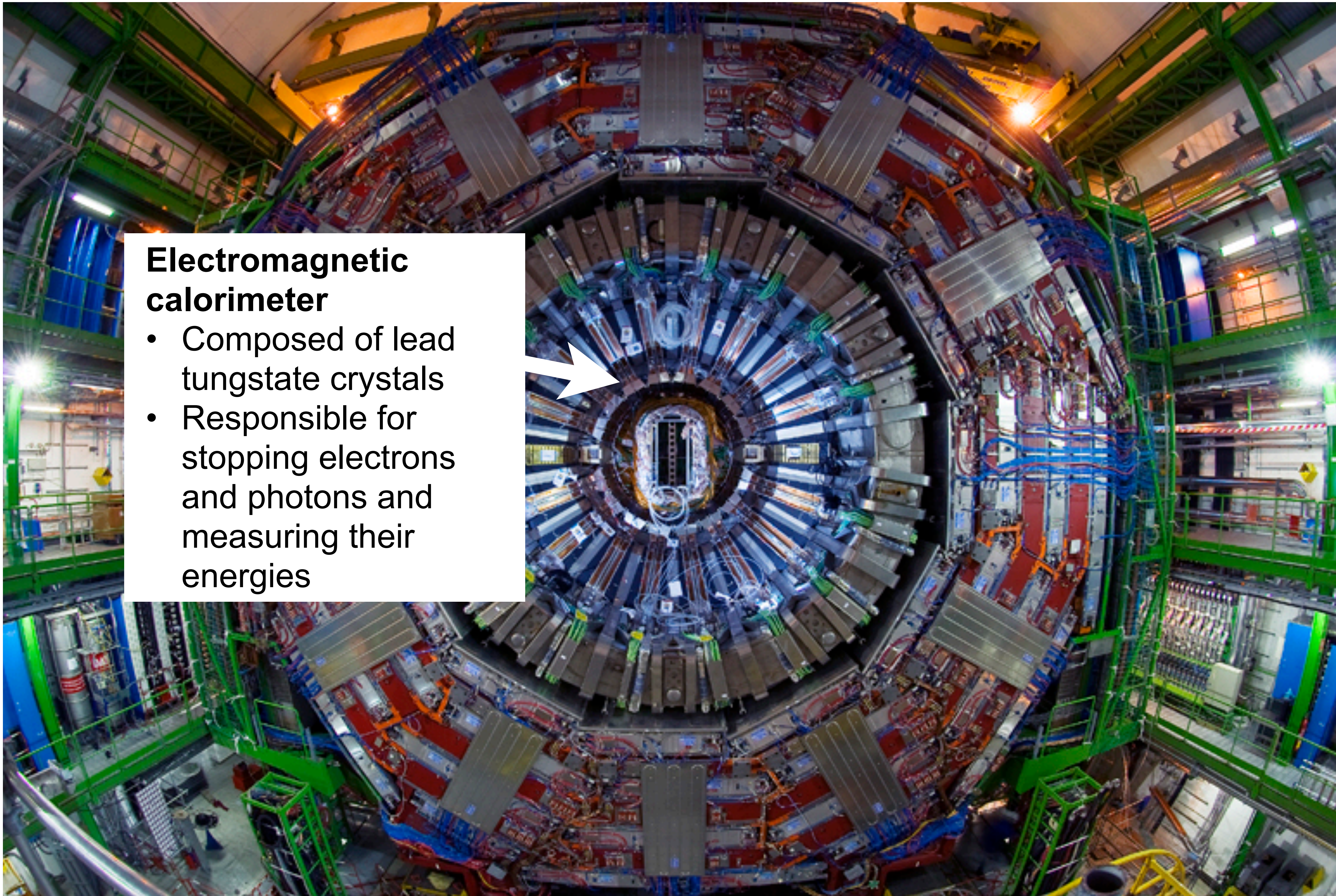
The subdetectors of CMS



Inner tracker

- Composed of silicon pixels and strips
- Records paths of charged particles

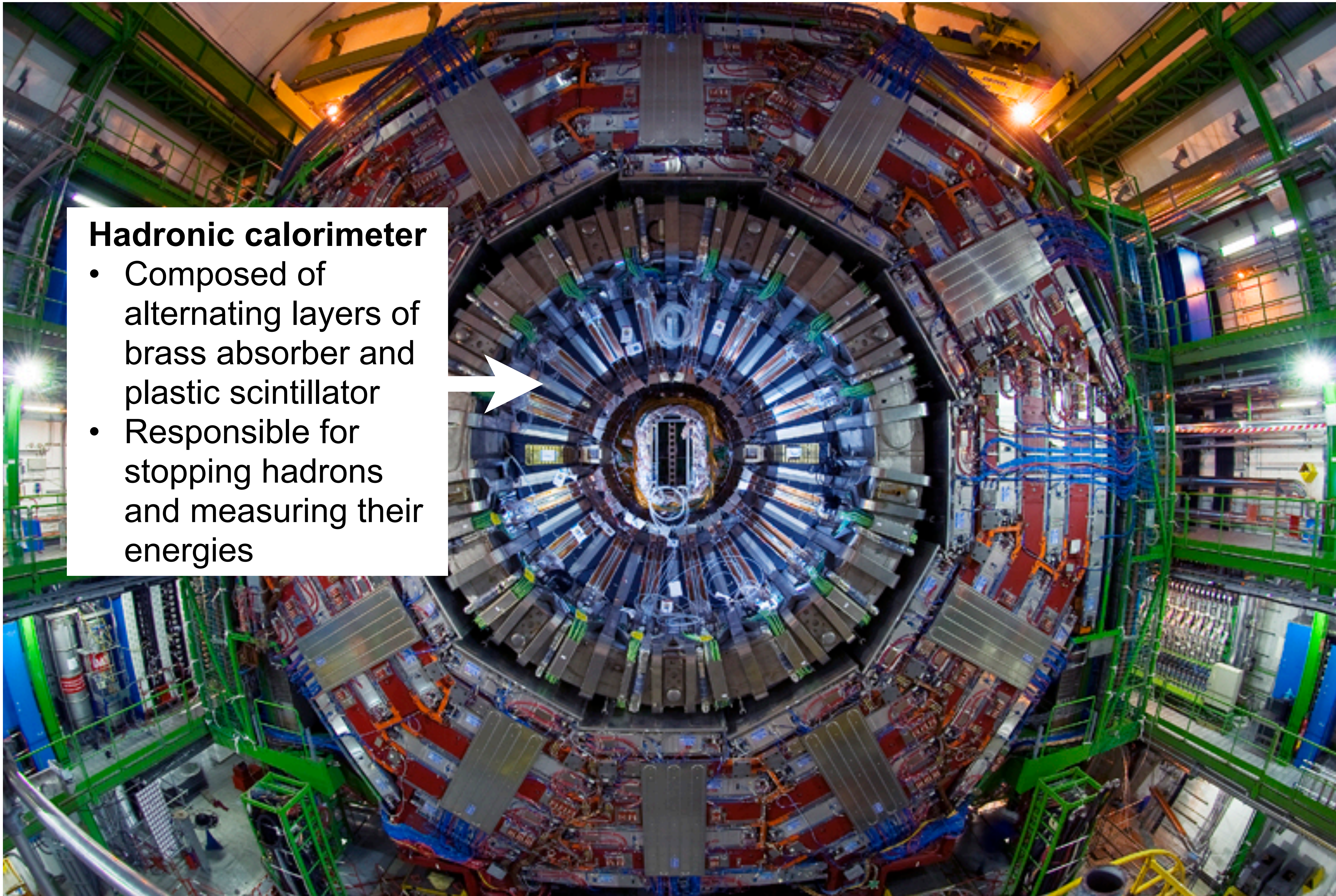
The subdetectors of CMS



Electromagnetic calorimeter

- Composed of lead tungstate crystals
- Responsible for stopping electrons and photons and measuring their energies

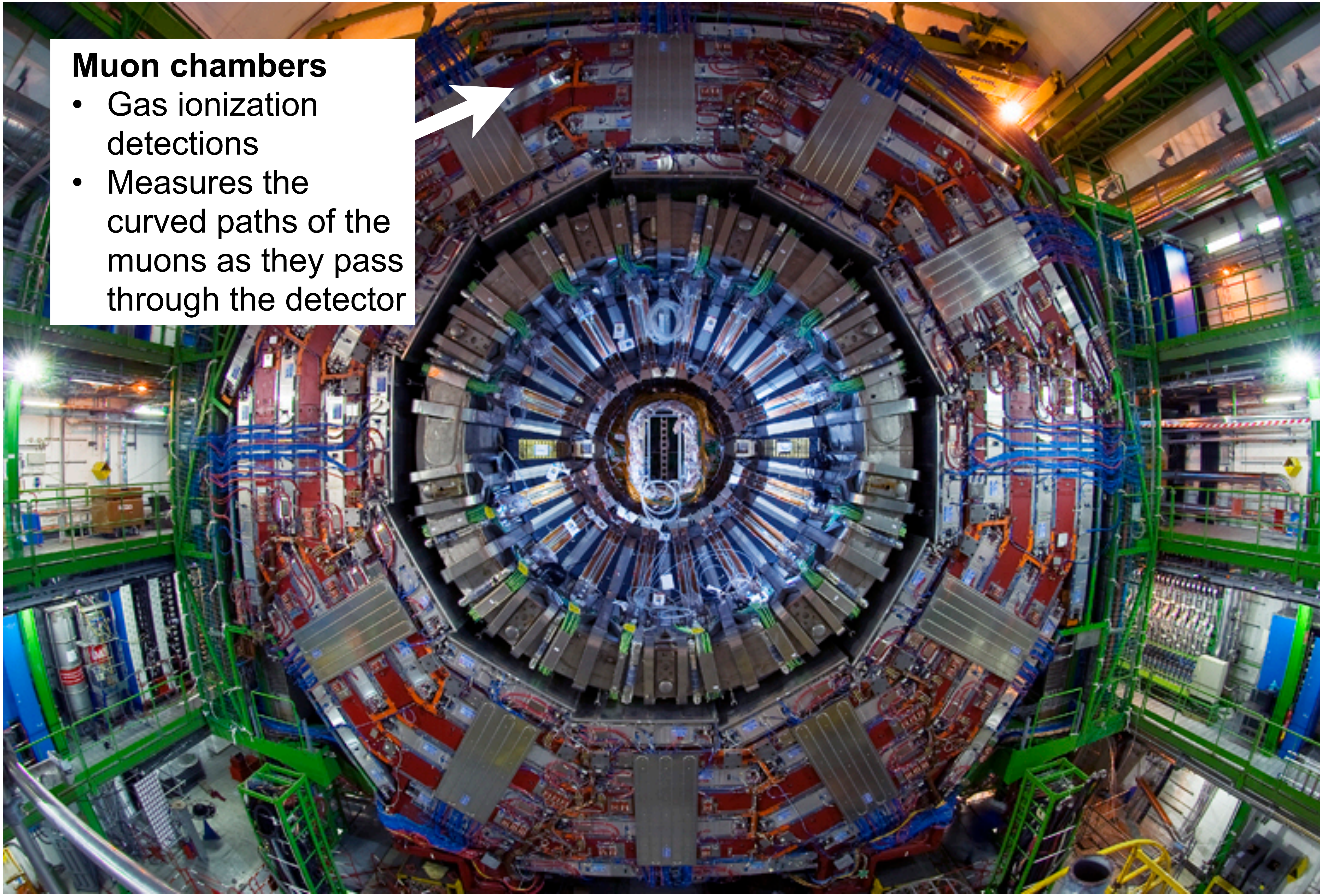
The subdetectors of CMS



Hadronic calorimeter

- Composed of alternating layers of brass absorber and plastic scintillator
- Responsible for stopping hadrons and measuring their energies

The subdetectors of CMS

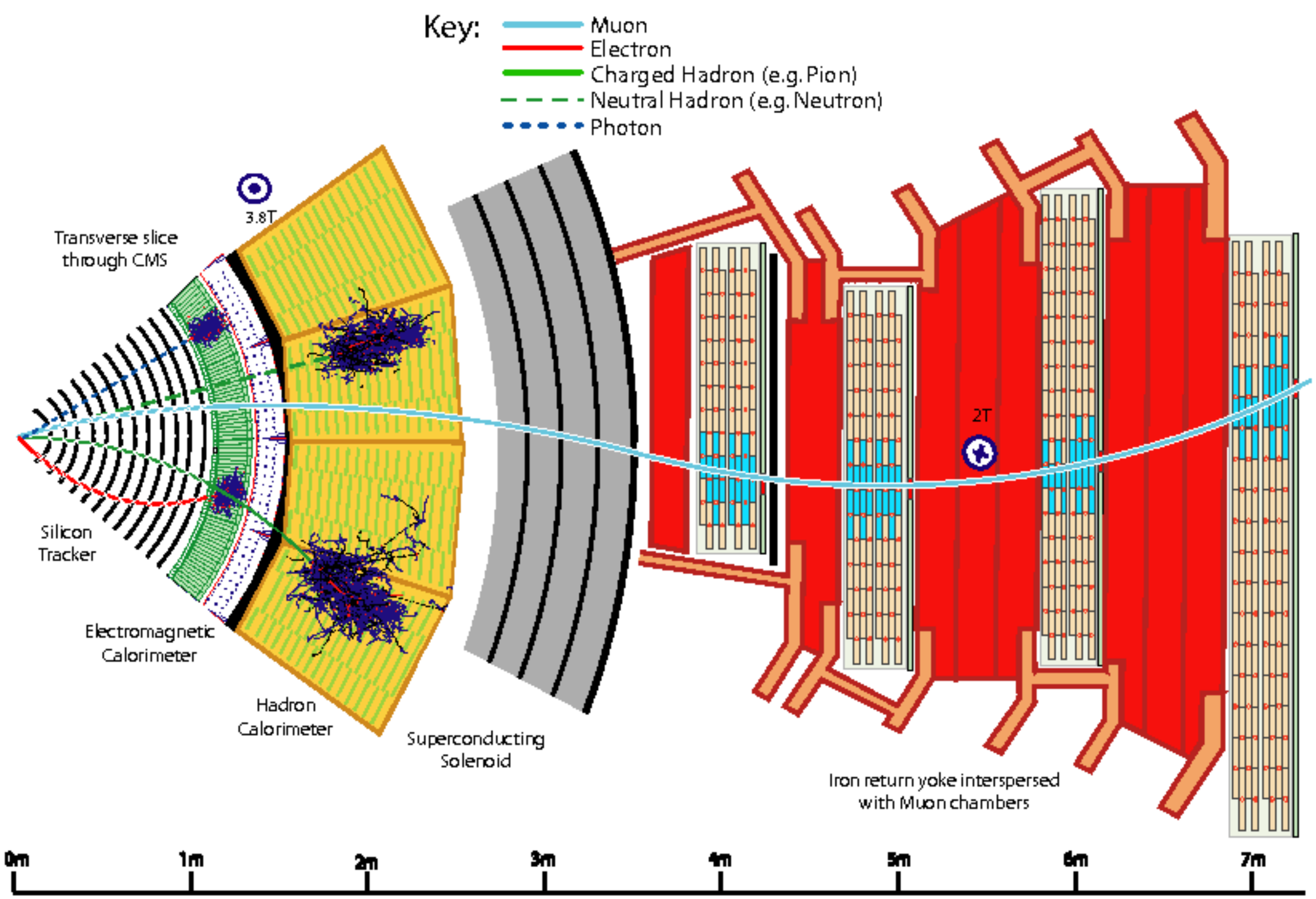


Muon chambers

- Gas ionization detections
- Measures the curved paths of the muons as they pass through the detector

Object reconstruction

- CMS uses a holistic reconstruction technique called **particle flow** to correlate the elements from each subdetector and construct a global picture of each event
- First identifies **muons**, then **electrons** and **isolated photons**, finally **charged hadrons**, **neutral hadrons**, **non-isolated photons**



CMS solenoid magnet and solenoid magnetic field

- Superconducting solenoid, magnetic field of ~4T
- Requires about 18,000 A of current
- Steel return yoke to guide lines of magnetic field around outside of the solenoid back into other end (and becomes magnetized by the field and thus passively adds to field)
- Solenoid:
  - Use Ampere's law to find the magnetic field, depends on the current in the wire (I) and the number of turns per length in the wire (N/L):

$$\oint \vec{B} \cdot d\vec{l} = \mu_0 I_{inc}$$

$$BL = \mu_0 NI$$

$$B = \mu_0 \frac{N}{L} I$$

Assume top leg of loop is far enough away from the coil that the field is negligible

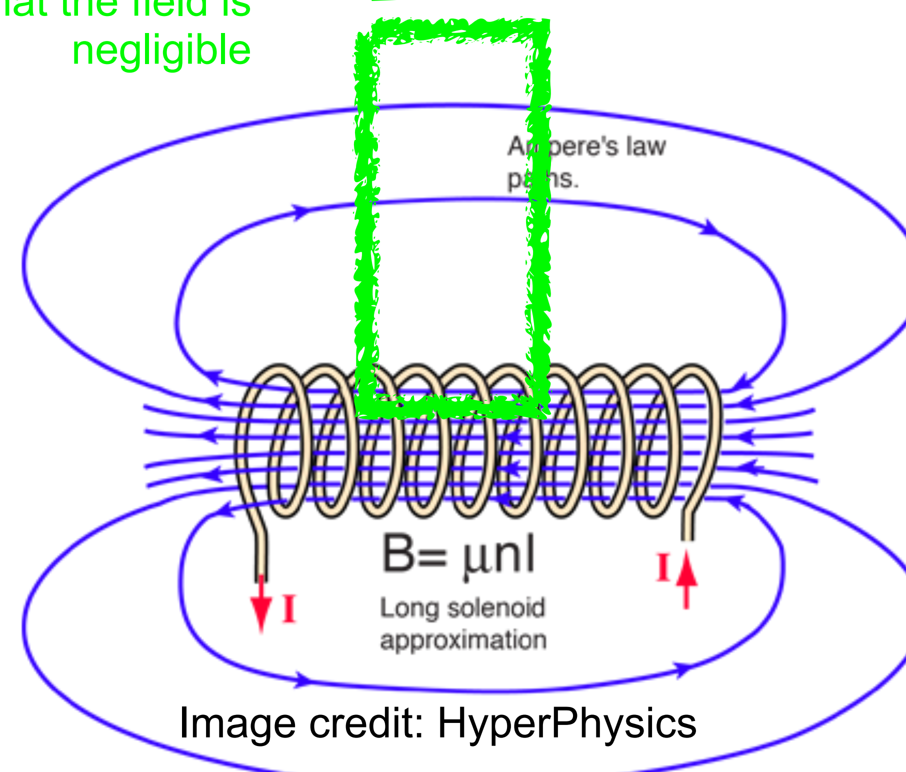


Image credit: HyperPhysics

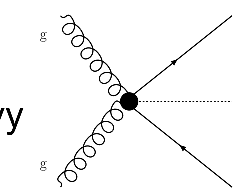
Modeling the EFT effects with MC

Modeling the signal contribution

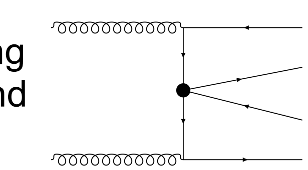
- We generate MC samples for our six signal processes (ttH, ttlnu, ttll, tllq, tHq, tttt) using MG with the dim6top model (1802.07237) to incorporate the relevant EFT effects in the MC
  - The dim6top model uses Warsaw basis of dimension-6 operators
  - The dim6top model is LO, so we include an extra jet in the matrix element (when possible) to improve the modeling at high jet-multiplicities and to capture relevant EFT dependence that enters with an extra parton
  - We include 26 WCs (all WCs from dim6top that significantly impact the processes contributing to our data samples)

Category	WCs
Two-heavy	$c_{t\varphi}, c^{-}_{\varphi Q}, c^{3}_{\varphi Q}, c_{\varphi t}, c_{\varphi tb}, c_{tW}, c_{tZ}, c_{bW}, c_{tG}$
Two-heavy-two-lepton	$c^{3(\ell)}_{Q\ell}, c^{-(\ell)}_{Q\ell}, c^{(\ell)}_{Qe}, c^{(\ell)}_{t\ell}, c^{(\ell)}_{te}, c^{S(\ell)}_{t}, c^{T(\ell)}_{t}$
Two-light-two-heavy	$c^{31}_{Qq}, c^{38}_{Qq}, c^{11}_{Qq}, c^{18}_{Qq}, c^{1}_{tq}, c^{8}_{tq}$
Four-heavy	$c^{1}_{QQ}, c^{1}_{Qt}, c^{8}_{Qt}, c^{1}_{tt}$

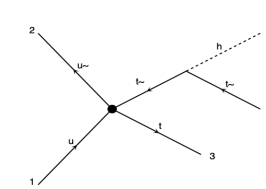
Operators involving heavy quarks and bosons



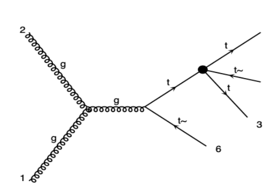
Operators involving 2 heavy quarks and 2 leptons



Operators involving two light quarks and 2 heavy quarks



Operators involving 4 heavy quarks



- We generated ~300M events in total (using dedicated and opportunistic computational resources at Notre Dame) using the same configurations as UL central samples, so our private samples are equivalent to corresponding central samples in terms of detector conditions etc.

Overview of EFT parametrization

- Matrix element can be written as the sum of SM and new physics components:

$$\mathscr{M} = \mathscr{M}_{SM} + \sum_i c_i \mathscr{M}_i$$

c_i are the Wilson coefficients

- Since, $\sigma \propto \mathscr{M}^2$, cross sections depend quadratically on the Wilson coefficients c_i

- E.g. for just one c_1:

$$d\sigma(c_1) \propto |\mathscr{M}_{SM} + c_1 \mathscr{M}_1|^2 \propto \underbrace{s_0}_{\text{SM}} + \underbrace{s_1}_{\text{Interference with SM}} c_1 + \underbrace{s_2}_{\text{Pure NP}} c_1^2$$

- In principle, could solve for the s_0, s_1, s_2 by generating 3 samples at distinct points in the WC space, but this would be very impractical if considering multiple WCs

Using MG reweighting to parametrize event weights

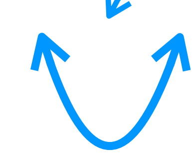
- Instead of parametrizing an inclusive or differential cross section, we can use the MadGraph generator's reweighting procedure to parametrize the weight of each generated event
    - Given an event generated under a specific theoretical scenario (e.g. an event at a particular point in WC space), MadGraph can determine the weight of the event under alternative theoretical scenarios (e.g. at other distinct points in WC space)
    - Since the event weight essentially corresponds to a small piece of the cross section, each event's weight depends quadratically on the WCs
    - For each event, we ask MadGraph to find the weight at a set of distinct points in WC space (which we call reweight points)
    - With enough reweight points (378 are required for 26 WCs), we can fully determine all of the coefficients of the 26-dimensional quadratic function for each event's weight

$$w_i\left(\frac{\vec{c}}{\Lambda^2}\right) = \underbrace{s_{0i}}_{\text{SM}} + \sum_j \underbrace{s_{1ij}}_{\text{Interference between EFT and SM}} \frac{c_j}{\Lambda^2} + \sum_j \underbrace{s_{2ij}}_{\text{Pure EFT}} \frac{c_j^2}{\Lambda^4} + \sum_{j \neq k} \underbrace{s_{3ijk}}_{\text{Interference among EFT contributions}} \frac{c_j}{\Lambda^2}\frac{c_k}{\Lambda^2}$$

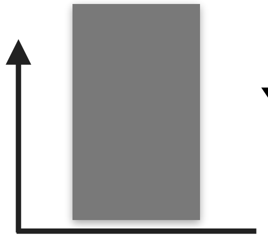
Quadratic parametrization of each event's weight

- MC events each carry a weight, which indicates how much of the cross section the event accounts for
- Since the event weight corresponds to a small piece of the cross section, it should also depend quadratically on the WCs
- So for every event, we can write the weight as a quadratic in terms of the WCs:

$$w_i = \overset{\text{SM}}{s_{0i}} + \sum_j \overset{\text{Interference between EFT and SM}}{s_{1ij}}\, c_j + \sum_j \overset{\text{Pure EFT}}{s_{2ij}}\, c_j^2 + \sum_{j \neq k} \overset{\text{Interference among EFT contributions}}{s_{3ijk}}\, c_j c_k =$$

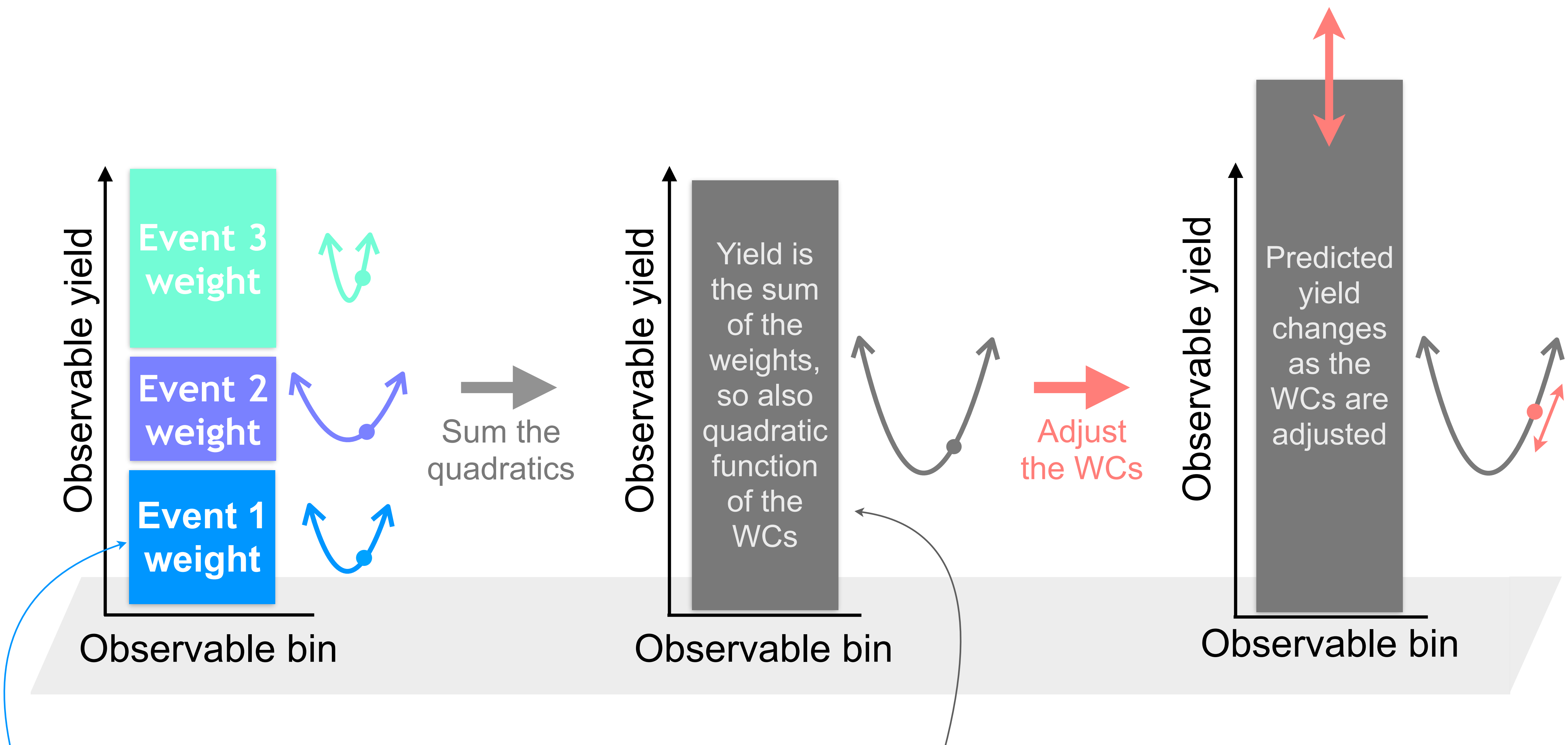


- To find the yield in a given bin, sum the weight functions for all events passing event selection criteria for the bin:



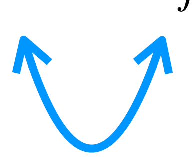
$$\text{Yield} = \Sigma w_i =$$

Summary of how the parameterized weights are used to find the yield in a given bin



$$w_i = s_{0i} + \sum_j s_{1ij}\, c_j + \sum_j s_{2ij}\, c_j^2 + \sum_{j \neq k} s_{3ijk}\, c_j c_k$$

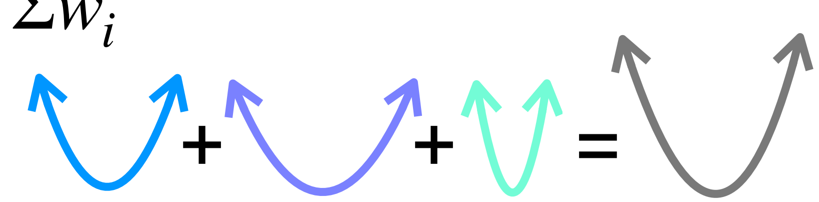
=



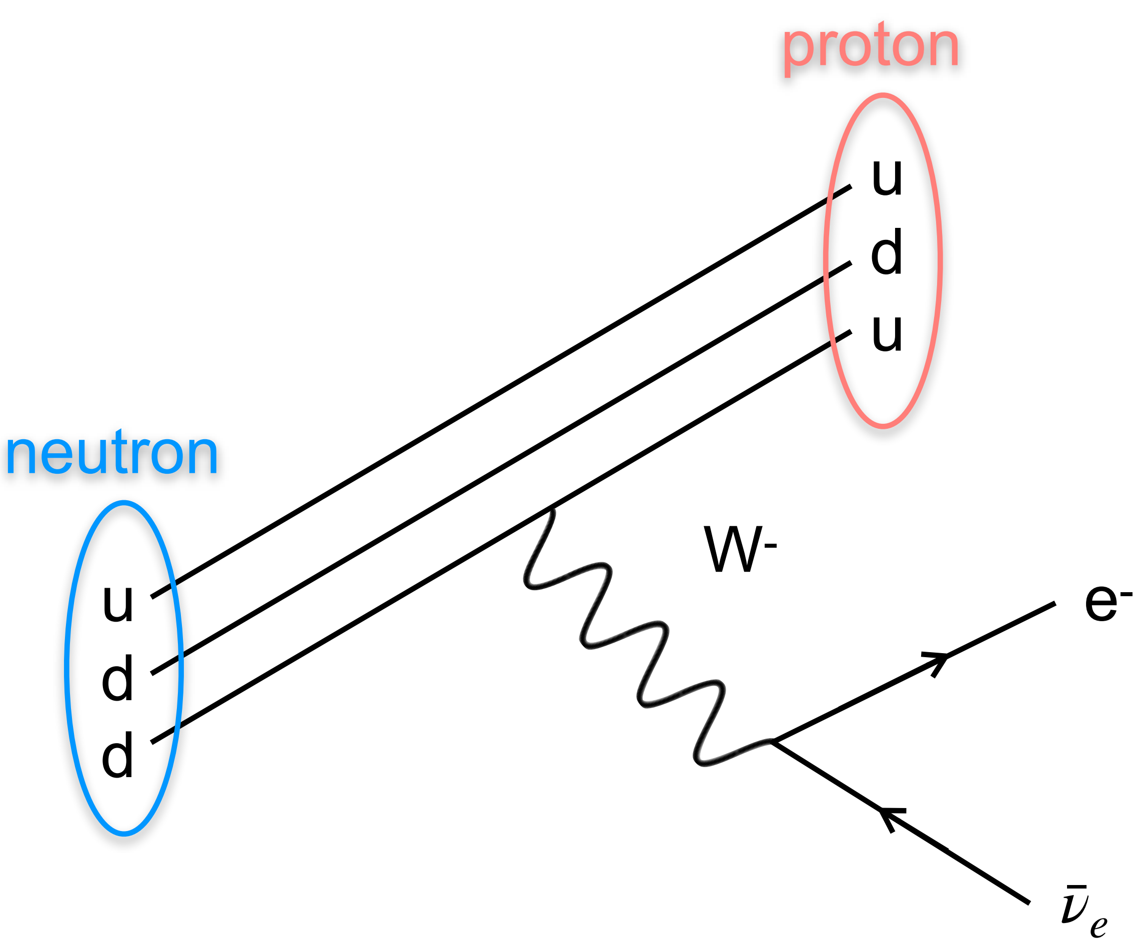
The structure constants of the n-dimensional quadric are determined with MG reweighting

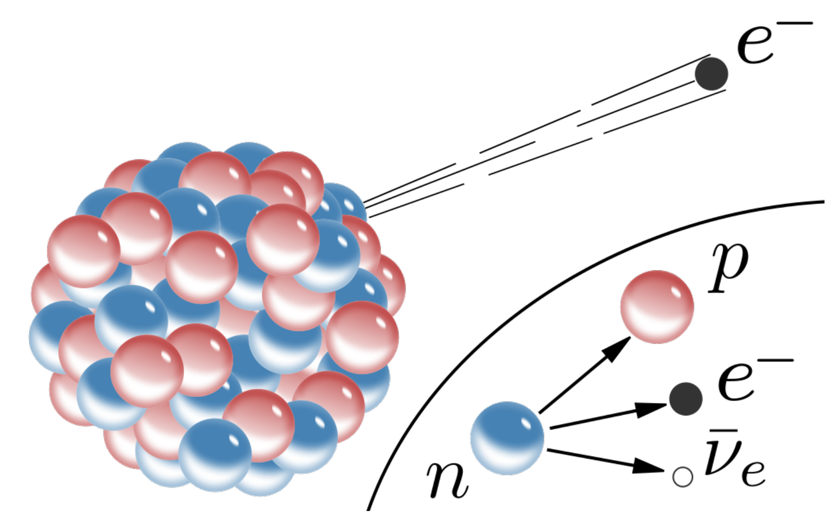
$$\text{Yield} = \Sigma w_i$$

=



Fermi theory for beta decay





When q/m_W is small, interaction can be written as dimension-6 four fermion interaction between d, u, e, and $\bar{\nu}_e$

$$i\mathcal{M} = \left(-i\frac{g}{\sqrt{2}}\right)^2 \left[\bar{u}(\nu_u)\gamma^u \frac{1-\gamma^5}{2} u(\mu)\right] \frac{-i}{q^2 - m_W^2 + i\epsilon}\left(\eta_{\mu\nu} - \frac{q_\mu q_\nu}{m_W^2}\right)\left[\bar{u}(\nu_u)\gamma^u \frac{1-\gamma^5}{2} u(\mu)\right]$$

← Matrix element

$(q \ll m_W)$

$$\frac{-i}{q^2 - m_W^2 + i\epsilon}\left(\eta_{\mu\nu} - \frac{q_\mu q_\nu}{m_W^2}\right) = i\frac{\eta_{\mu\nu}}{m_W^2} + \mathcal{O}\left(\frac{1}{m_W^4}\right)$$

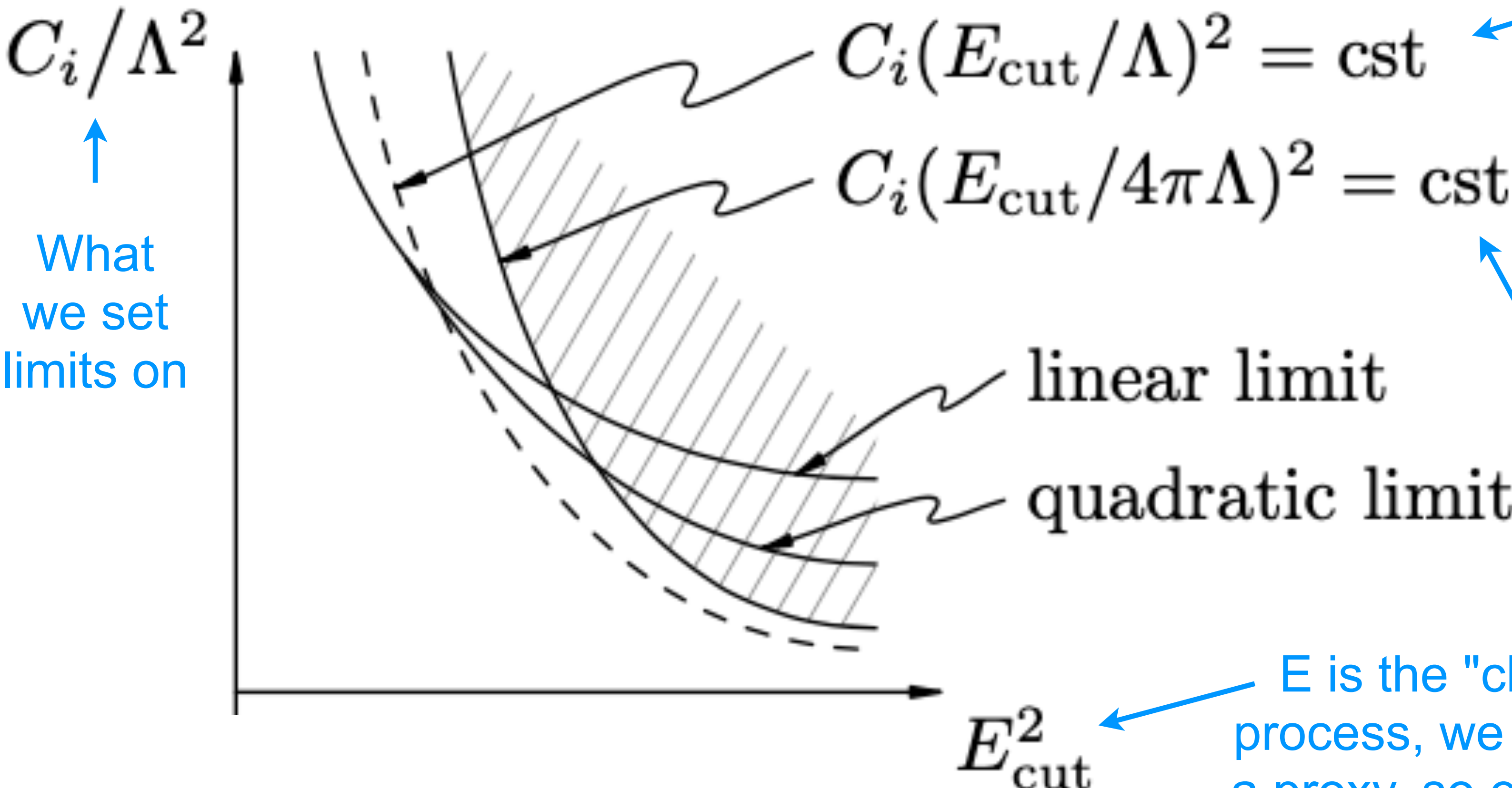
Operators involving two quarks and one or more bosons			
Operator	Definition	WC	Lead processes affected
‡$O_{\mathrm{u}\varphi}^{(ij)}$	$\bar{\mathrm{q}}_i \mathrm{u}_j \tilde{\varphi}\,(\varphi^\dagger\varphi)$	$c_{\mathrm{t}\varphi} + ic^I_{\mathrm{t}\varphi}$	$\mathrm{t\bar{t}H}$, tHq
$O_{\varphi\mathrm{q}}^{1(ij)}$	$(\varphi^\dagger i\overleftrightarrow{D}_\mu\varphi)(\bar{\mathrm{q}}_i\gamma^\mu \mathrm{q}_j)$	$c^-_{\varphi Q} + c^3_{\varphi Q}$	$\mathrm{t\bar{t}H}$, $\mathrm{t\bar{t}l\nu}$, $\mathrm{t\bar{t}l\bar{l}}$, tHq, $\mathrm{tl\bar{l}q}$
$O_{\varphi\mathrm{q}}^{3(ij)}$	$(\varphi^\dagger i\overleftrightarrow{D}^I_\mu\varphi)(\bar{\mathrm{q}}_i\gamma^\mu\tau^I \mathrm{q}_j)$	$c^3_{\varphi Q}$	$\mathrm{t\bar{t}H}$, $\mathrm{t\bar{t}l\nu}$, $\mathrm{t\bar{t}l\bar{l}}$, tHq, $\mathrm{tl\bar{l}q}$
$O_{\varphi\mathrm{u}}^{(ij)}$	$(\varphi^\dagger i\overleftrightarrow{D}_\mu\varphi)(\bar{\mathrm{u}}_i\gamma^\mu \mathrm{u}_j)$	$c_{\varphi\mathrm{t}}$	$\mathrm{t\bar{t}H}$, $\mathrm{t\bar{t}l\nu}$, $\mathrm{t\bar{t}l\bar{l}}$, $\mathrm{tl\bar{l}q}$
‡$O_{\varphi\mathrm{ud}}^{(ij)}$	$(\tilde{\varphi}^\dagger iD_\mu\varphi)(\bar{\mathrm{u}}_i\gamma^\mu \mathrm{d}_j)$	$c_{\varphi\mathrm{tb}} + ic^I_{\varphi\mathrm{tb}}$	$\mathrm{t\bar{t}H}$, $\mathrm{tl\bar{l}q}$, tHq
‡$O_{\mathrm{uW}}^{(ij)}$	$(\bar{\mathrm{q}}_i\sigma^{\mu\nu}\tau^I \mathrm{u}_j)\ \tilde{\varphi}\mathrm{W}^I_{\mu\nu}$	$c_{\mathrm{tW}} + ic^I_{\mathrm{tW}}$	$\mathrm{t\bar{t}H}$, $\mathrm{t\bar{t}l\nu}$, $\mathrm{t\bar{t}l\bar{l}}$, tHq, $\mathrm{tl\bar{l}q}$
‡$O_{\mathrm{dW}}^{(ij)}$	$(\bar{\mathrm{q}}_i\sigma^{\mu\nu}\tau^I \mathrm{d}_j)\ \varphi\mathrm{W}^I_{\mu\nu}$	$c_{\mathrm{bW}} + ic^I_{\mathrm{bW}}$	$\mathrm{t\bar{t}H}$, $\mathrm{t\bar{t}l\bar{l}}$, tHq, $\mathrm{tl\bar{l}q}$
‡$O_{\mathrm{uB}}^{(ij)}$	$(\bar{\mathrm{q}}_i\sigma^{\mu\nu}\mathrm{u}_j)\ \tilde{\varphi}\mathrm{B}_{\mu\nu}$	$(c_{\mathrm{W}}c_{\mathrm{tW}} - c_{\mathrm{tZ}})/s_{\mathrm{W}} + i(c_{\mathrm{W}}c^I_{\mathrm{tW}} - c^I_{\mathrm{tZ}})/s_{\mathrm{W}}$	$\mathrm{t\bar{t}H}$, $\mathrm{t\bar{t}l\nu}$, $\mathrm{t\bar{t}l\bar{l}}$, tHq, $\mathrm{tl\bar{l}q}$
‡$O_{\mathrm{uG}}^{(ij)}$	$(\bar{\mathrm{q}}_i\sigma^{\mu\nu}T^A \mathrm{u}_j)\ \tilde{\varphi}G^A_{\mu\nu}$	$\mathrm{g_s}(c_{\mathrm{tG}} + ic^I_{\mathrm{tG}})$	$\mathrm{t\bar{t}H}$, $\mathrm{t\bar{t}l\nu}$, $\mathrm{t\bar{t}l\bar{l}}$, tHq, $\mathrm{tl\bar{l}q}$

Operators involving two quarks and two leptons			
Operator	Definition	WC	Lead processes affected
$O_{\ell\mathrm{q}}^{1(ijkl)}$	$(\bar{\ell}_i\gamma^\mu\ell_j)(\bar{\mathrm{q}}_k\gamma^\mu \mathrm{q}_\ell)$	$c^{-(\ell)}_{Q\ell} + c^{3(\ell)}_{Q\ell}$	$\mathrm{t\bar{t}l\nu}$, $\mathrm{t\bar{t}l\bar{l}}$, $\mathrm{tl\bar{l}q}$
$O_{\ell\mathrm{q}}^{3(ijkl)}$	$(\bar{\ell}_i\gamma^\mu\tau^I\ell_j)(\bar{\mathrm{q}}_k\gamma^\mu\tau^I \mathrm{q}_\ell)$	$c^{3(\ell)}_{Q\ell}$	$\mathrm{t\bar{t}l\nu}$, $\mathrm{t\bar{t}l\bar{l}}$, $\mathrm{tl\bar{l}q}$
$O_{\ell\mathrm{u}}^{(ijkl)}$	$(\bar{\ell}_i\gamma^\mu\ell_j)(\bar{\mathrm{u}}_k\gamma^\mu \mathrm{u}_\ell)$	$c^{(\ell)}_{\mathrm{t}\ell}$	$\mathrm{t\bar{t}l\bar{l}}$
$O_{\mathrm{e\bar{q}}}^{(ijkl)}$	$(\bar{\mathrm{e}}_i\gamma^\mu \mathrm{e}_j)(\bar{\mathrm{q}}_k\gamma^\mu \mathrm{q}_\ell)$	$c^{(\ell)}_{Q\mathrm{e}}$	$\mathrm{t\bar{t}l\bar{l}}$, $\mathrm{tl\bar{l}q}$
$O_{\mathrm{eu}}^{(ijkl)}$	$(\bar{\mathrm{e}}_i\gamma^\mu \mathrm{e}_j)(\bar{\mathrm{u}}_k\gamma^\mu \mathrm{u}_\ell)$	$c^{(\ell)}_{\mathrm{te}}$	$\mathrm{t\bar{t}l\bar{l}}$
‡$O_{\ell\mathrm{equ}}^{1(ijkl)}$	$(\bar{\ell}_i \mathrm{e}_j)\ \varepsilon\ (\bar{\mathrm{q}}_k \mathrm{u}_\ell)$	$c^{S(\ell)}_{\mathrm{t}} + ic^{SI(\ell)}_{\mathrm{t}}$	$\mathrm{t\bar{t}l\bar{l}}$, $\mathrm{tl\bar{l}q}$
‡$O_{\ell\mathrm{equ}}^{3(ijkl)}$	$(\bar{\ell}_i\sigma^{\mu\nu}\mathrm{e}_j)\ \varepsilon\ (\bar{\mathrm{q}}_k\sigma_{\mu\nu}\mathrm{u}_\ell)$	$c^{T(\ell)}_{\mathrm{t}} + ic^{TI(\ell)}_{\mathrm{t}}$	$\mathrm{t\bar{t}l\nu}$, $\mathrm{t\bar{t}l\bar{l}}$, $\mathrm{tl\bar{l}q}$

Where the covariant derivative is: $D_\mu = \partial_\mu - ig_s T^A G^A_\mu - igS^I W^I_\mu - ig' Y_\psi B_\mu$

Looking at Fig. 1 from the dim6top paper

- The dim6top paper (1802.07237) section 5 points 6 and 7 discuss interpretations of limits in terms of validity, summarized in Fig 1
- We can try to run our analysis with different Ecut values, and see where we stand on the plot



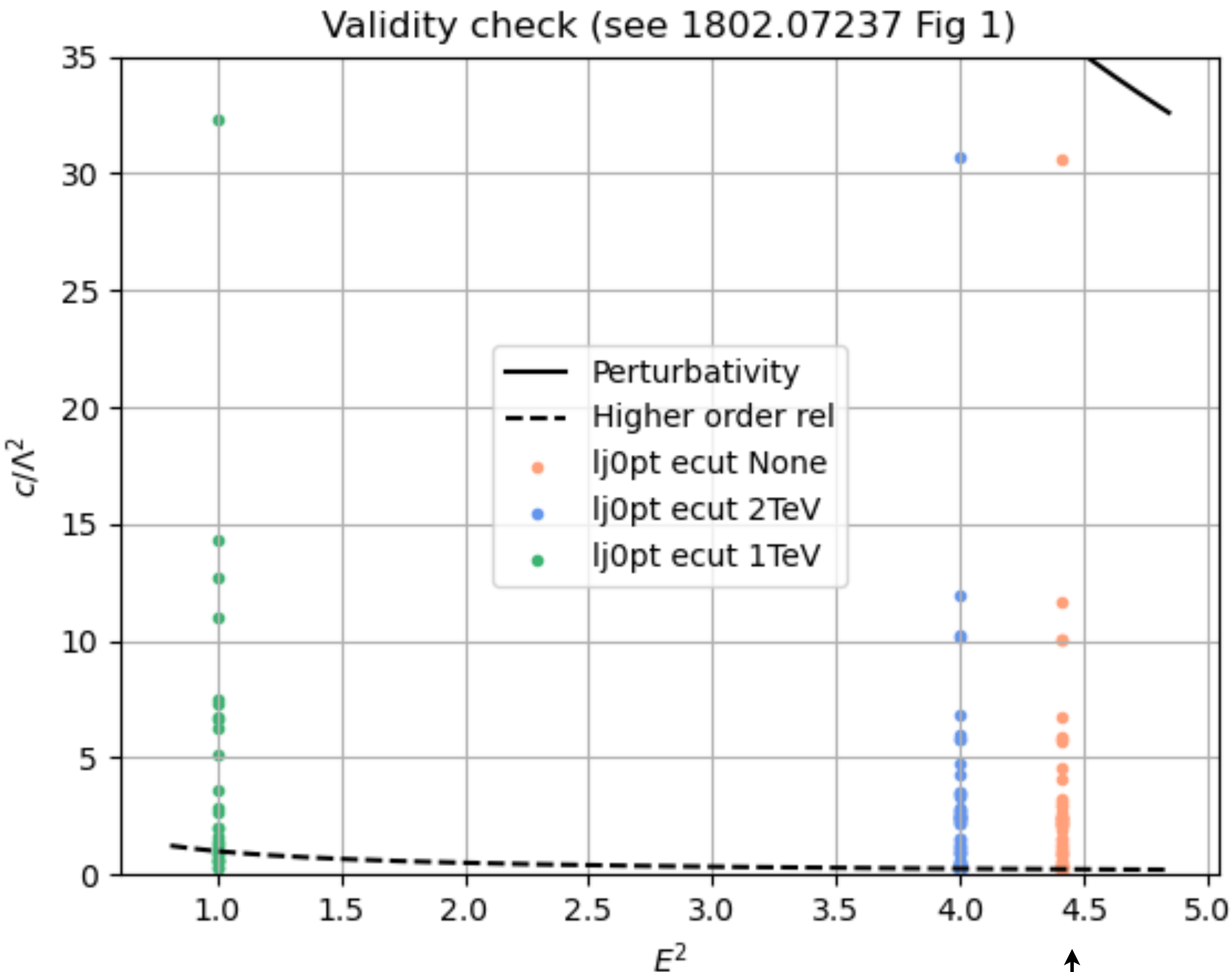
What we set limits on

Higher order terms become important

EFT breaks down here, we definitely do not want to be in this region

dim6top doesn't commit to a cst value, but we can plug in 1 for our tests because why not

E is the "characteristic energy scale" of the process, we will try to use something like S_T as a proxy, so e.g. Ecut = 50GeV means we throw out all events in our analysis with $S_T > 50$



This plot shows abs of each + and - 2 sigma limit, so there should be 52 numbers for each E_cut, from this we see our limits do not blow up when we place an E_cut

Note the red points are without an E_cut, so their placement on the x axis is a bit arbitrary

Object selection and event selection details

Jet and b-jet objects selection details for reference

- Jet are reconstructed by clustering PF candidates using the anti-kt algorithm with distance parameter $\Delta R = 4$ as implemented in the FastJet package
- Charged hadrons not coming from the primary vertices are subtracted from the PF candidates considered in the clustering
- Fake jets (mainly arising from calorimeter noise) are rejected by requiring reconstructed jets to pass a set of tight jet identification criteria, the selected jets are referred to as AK4 PF CHS jets
- The jet energy corrections derived by the JetMET FOG are applied as a function of jet E_T and eta
- The recommended jet energy resolution smearing is also applied
- Within the tracker acceptance the jet tracks are required to be compatible with the primary vertex
- Jets are required to have a pt>30GeV and abs(eta)<2.4
- Jets are also required to be separated from leptons candidates passing the loose selection (by requiring $\Delta\mathrm{R} = \sqrt{(\eta^{\ell} - \eta^{jet})^2 + (\phi^{\ell} - \phi^{jet})^2} > 0.4$)
- Identifying b-jets:
  - The DeepJet b-tagging algorithm is used
  - This distinguishes between b-jets and jets originating from light quarks, gluons, and charm quarks
  - We use the UL medium and loose working points defined by the BTV POG

Jet energy corrections

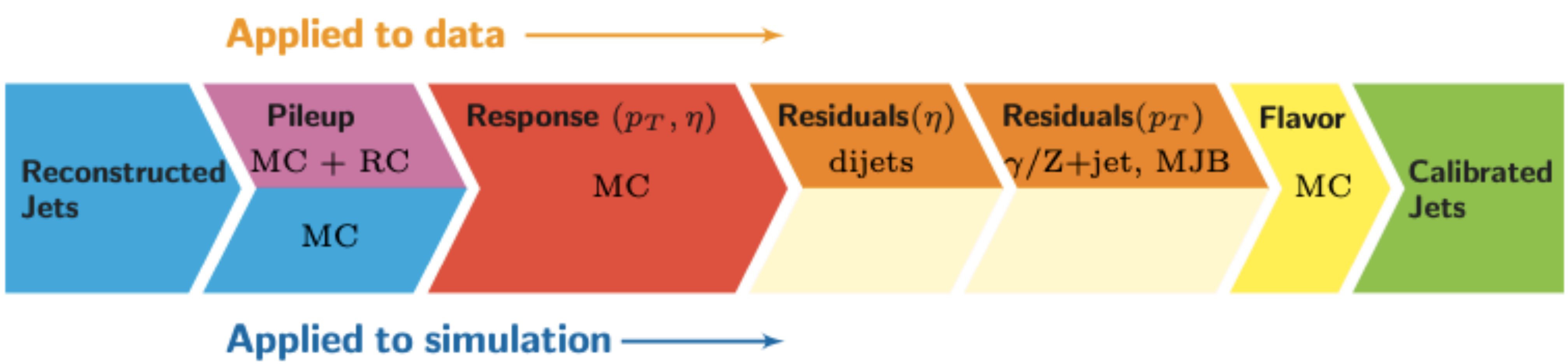


Figure 2: Consecutive stages of JEC, for data and MC simulation. All corrections marked with MC are derived from simulation studies, RC stands for random cone, and MJB refers to the analysis of multijet events.

https://arxiv.org/pdf/1607.03663.pdf

Object selection for electrons, muons, jets, and b jets

- Jets and b tagging:
  - Generally follows the standard recommendations by the POGs
  - We require jets to have p_T>30 GeV, $|\eta|$< 2.4, identify b-jets with the DeepJet b-tagging algorithm
- Leptons: Using custom MVA selection, standard usage in multilepton analyses (and reviewed by the POGs)
- Efficiencies for electrons and muons to pass the selections defined in the tables to the right have been calculated, and the SFs are being approved by the EGamma and muon groups (presentations [here] and [here])

Electrons			
Observable	Loose	Fakeable	Tight
Cone-p_T	> 7 GeV	> 10 GeV	> 10 GeV
$\|\eta\|$	< 2.5	< 2.5	< 2.5
$\|d_{xy}\|$	< 0.05 cm	< 0.05 cm	< 0.05 cm
$\|d_z\|$	< 0.1 cm	< 0.1 cm	< 0.1 cm
d/σ_d	< 8	< 8	< 8
I_e	$< 0.4 \times p_T$	$< 0.4 \times p_T$	$< 0.4 \times p_T$
$\sigma_{i\eta i\eta}$	–	$<$ { 0.011 / 0.030 }[1]	$<$ { 0.011 / 0.030 }[1]
H/E	–	< 0.10	< 0.10
1/E - 1/p	–	> -0.04	> -0.04
Conversion rejection	–	✓	✓
Missing hits	≤ 1	$= 0$	$= 0$
EGamma POG MVA	>WP-loose[2]	>WP-90 (>WP-loose)[2]†	>WP-loose[2]
Deep Jet of nearby jet	–	<WP-interp. (<WP-medium)[2]	<WP-medium[2]
Jet relative isolation[4]	–	< 1.0 (–) †	–
Prompt-e MVA	–	< 0.90 (> 0.90)	> 0.90

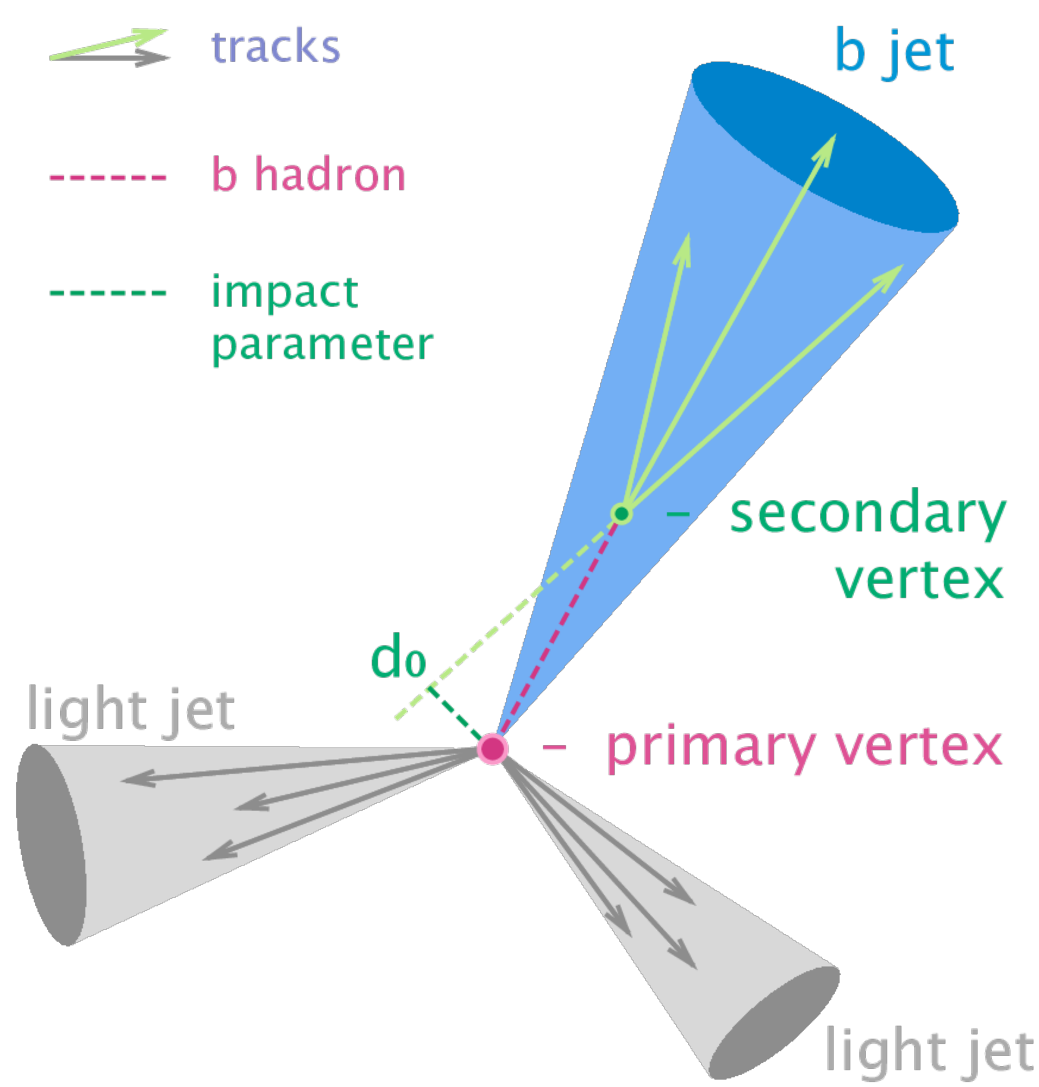
[1] Barrel / endcaps.
[2] WPs as defined by EGamma POG (see Section 4.1.1).
[3] WPs as defined by BTV POG (see Section 4.2).
[4] Defined as $1/p_T^{ratio}$-1 if the electron is matched to a jet within $\Delta R < 0.4$ or as the PF relative isolation with ΔR=0.4 otherwise.
† Fails (passes) the requirement prompt-e MVA > 0.80.

Muons			
Observable	Loose	Fakeable	Tight
p_T	> 5 GeV	> 10 GeV	> 10 GeV
$\|\eta\|$	< 2.4	< 2.4	< 2.4
$\|d_{xy}\|$	< 0.05 cm	< 0.05 cm	< 0.05 cm
$\|d_z\|$	< 0.1 cm	< 0.1 cm	< 0.1 cm
d/σ_d	< 8	< 8	< 8
I_μ	$< 0.4 \times p_T$	$< 0.4 \times p_T$	$< 0.4 \times p_T$
PF muon	>WP-loose[1]	>WP-loose[1]	>WP-medium[1]
Deep Jet of nearby jet	–	<WP-interp. (<WP-medium)[2]	<WP-medium[2]
Jet relative isolation[3]	–	<0.5 (–) †	–
Prompt-μ MVA	–	< 0.85 (> 0.85)	> 0.85

[1] WPs as defined by Muon POG (see Section 4.1.2).
[2] Upper cut on the Deep Jet score defined with a linear interpolation from Deep Jet WP-medium at cone-p_T 20 GeV to Deep Jet WP-loose at cone-p_T 45 GeV, taking the Deep Jet WPs as defined by JetMET POG (see Section ??).
[3] Defined as 1/`jetPtRatio`-1 if the muon is matched to a jet within $\Delta R < 0.4$ or as the PF relative isolation with ΔR=0.4 otherwise.
† Fails (passes) the requirement prompt-μ MVA > 0.85.

b-tagging

- Top quarks almost always decay into a W and a b
- Identifying jets that arising from b quarks is called b-tagging
- To distinguish from jets arising from hadronization of lighter quarks or gluons: take advantage of b hadrons'
  - **large mass** (decay products with relatively large transverse momentum relative to jet axis)
  - **long lifetimes** (displaced tracks, secondary vertices)
  - Sizable **semileptonic branching ratio**



"b Tagging in ATLAS and CMS:" arxiv 1709.01290

Event selection details

- All jets required to have $|\eta| < 2.4$ and $p_T > 30$ GeV, all electrons require $|\eta| < 2.5$, all μ require $|\eta| < 2.4$, with lepton p_T cuts (in GeV):
  - 2lss 1st and 2nd: $p_T > 25$, $p_T > 15$
  - 3l 1st, 2nd, and 3rd: $p_T > 25$, $p_T > 15$, $p_T > 15$ (10) for e (μ)
  - 4l 1st, 2nd, 3rd, and 4th: $p_T > 25$, $p_T > 15$, $p_T > 15$ (10) for e (μ), $p_T > 15$ (10) for e (μ)

Category	Leptons	$m_{\ell\ell}$	b-tags	Lepton charge sum	Jets	Differential variable
2ℓss 2b	2	No requirement	2	> 0, <0	4,5,6,≥7	$p_T(lj)_0$
2ℓss 3b	2	No requirement	≥ 3	> 0, <0	4,5,6,≥7	$p_T(lj)_0$
3ℓ off-Z 1b	3	$\|m_Z - m_{\ell\ell}\| > 10\,\mathrm{GeV}$	1	> 0, <0	2,3,4,≥5	$p_T(lj)_0$
3ℓ off-Z 2b	3	$\|m_Z - m_{\ell\ell}\| > 10\,\mathrm{GeV}$	≥ 2	> 0, <0	2,3,4,≥5	$p_T(lj)_0$
3ℓ on-Z 1b	3	$\|m_Z - m_{\ell\ell}\| \leq 10\,\mathrm{GeV}$	1	No requirement	2,3,4,≥5	$p_T(Z)$
3ℓ on-Z 2b	3	$\|m_Z - m_{\ell\ell}\| \leq 10\,\mathrm{GeV}$	≥ 2	No requirement	2,3	$p_T(lj)_0$
3ℓ on-Z 2b	3	$\|m_Z - m_{\ell\ell}\| \leq 10\,\mathrm{GeV}$	≥ 2	No requirement	4,≥5	$p_T(Z)$
4ℓ	≥4	No requirement	≥ 2	No requirement	2,3,≥4	$p_T(lj)_0$

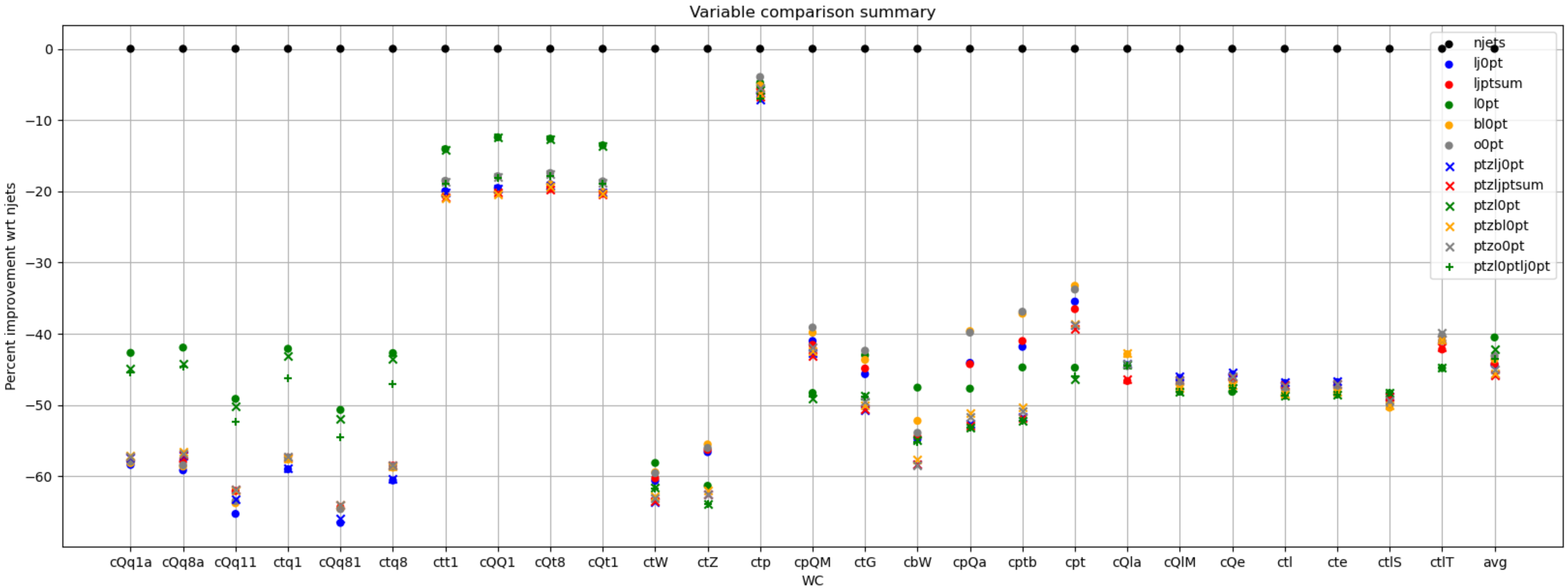
Signal selection yields

- For the full run 2 dataset, we have the following expected SM yields (~3400 total expected):

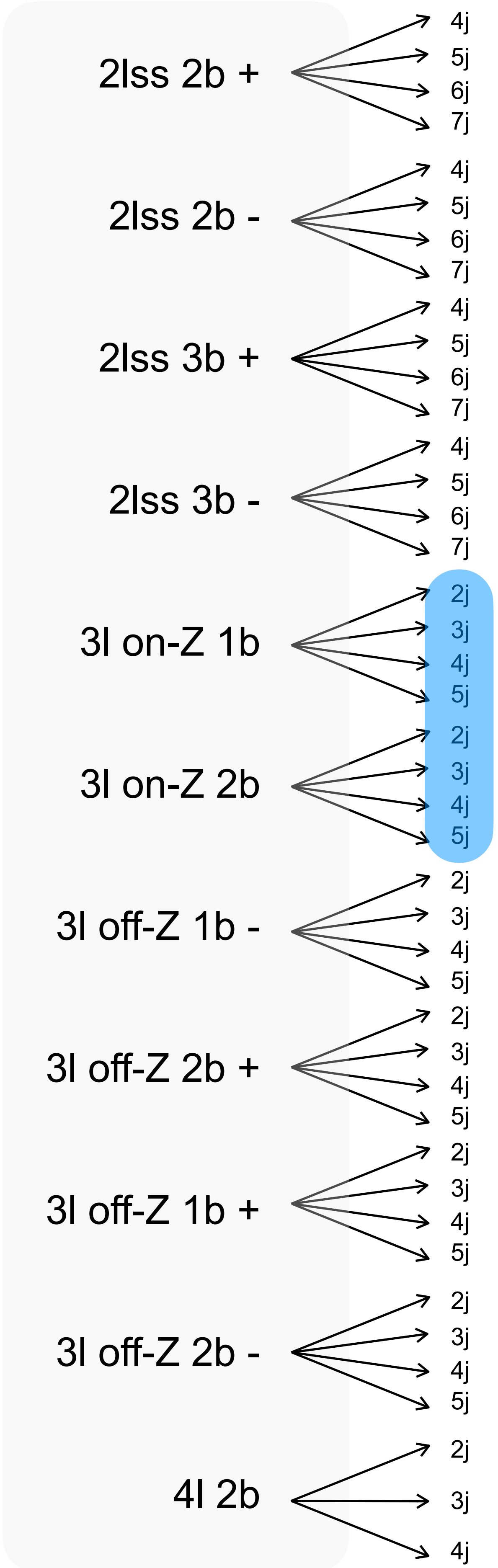
	2ℓss 3b -	2ℓss 3b +	2ℓss 2b -	2ℓss 2b +	3ℓ 1b -	3ℓ 1b +	3ℓ 2b -	3ℓ 2b +	3ℓ on-Z 1b	3ℓ on-Z 2b	4ℓ 2b
tWZ	0.46	0.47	6.66	6.7	4.78	4.77	1.69	1.69	63.34	20.3	2.72
Diboson	0.1	0.25	12.13	15.37	30.17	30.51	2.02	3.19	339.31	33.79	4.81
Triboson	0.04	0.07	2.16	3.14	0.95	1.33	0.1	0.17	16.02	2.6	0.45
Charge flips	1.62	1.57	17.5	17.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Nonprompt	6.72	8.94	112.63	120.55	57.0	56.52	11.86	10.47	55.38	10.05	0.0
Conversions	0.96	0.86	11.47	9.55	3.37	2.94	3.24	3.07	0.78	0.68	0.01
Sum bkg	9.9	12.16	162.55	172.72	96.28	96.06	18.91	18.6	474.83	67.42	7.99
$t\bar{t}l\nu$	12.53	23.69	144.17	272.75	25.7	47.37	27.25	50.67	10.18	11.42	0.03
$t\bar{t}l\bar{l}$	12.28	12.28	119.13	119.66	51.58	51.08	48.44	49.77	320.7	295.8	40.22
$t\bar{t}H$	9.51	9.47	83.29	83.48	23.51	23.24	22.92	22.71	9.61	9.7	3.5
$tl\bar{l}q$	0.47	0.87	6.51	11.75	5.46	9.53	2.4	4.23	111.51	48.11	0.01
tHq	0.12	0.23	1.42	2.6	0.47	0.84	0.35	0.61	0.35	0.23	0.03
$t\bar{t}t\bar{t}$	9.61	9.53	7.58	7.46	0.87	0.85	4.88	4.92	0.21	1.3	0.55
Sum sig	44.52	56.08	362.11	497.72	107.59	132.9	106.24	132.91	452.56	366.56	44.34
Sum expected	54 ± 5	68 ± 6	525 ± 48	670 ± 60	204 ± 22	229 ± 24	125 ± 10	152 ± 12	927 ± 129	434 ± 44	52 ± 6
Observation	71.0	68.0	608.0	781.0	233.0	270.0	148.0	158.0	1074.0	466.0	50.0

Optimizing sensitivity

- We have 43 analysis bins, and in principle we could choose different kinematic distribution for each of the 43 categories
- Because of the global nature of our analysis, we can't choose a different variable per WC, so for each category we wanted to identify a variable that targets WCs that contribute strongly to the given category, or identify variables that provide generally good sensitivity to a broad range of EFT effects

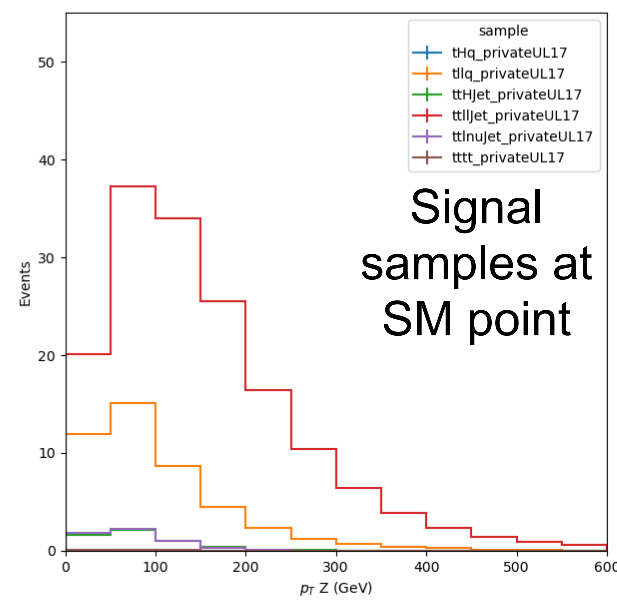


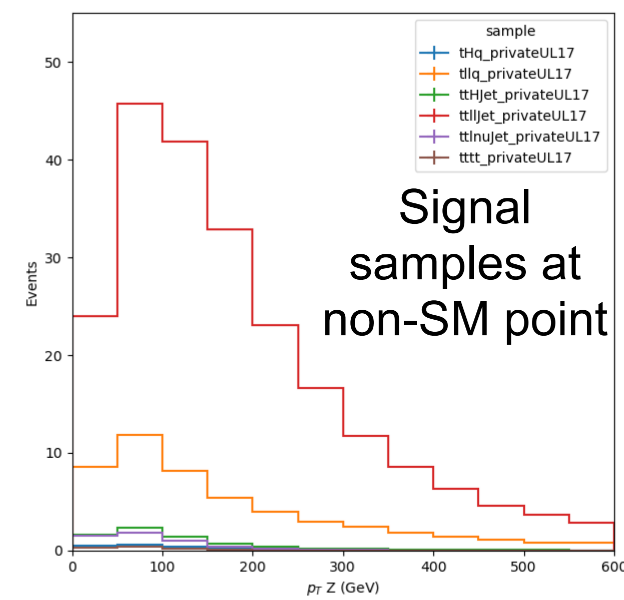
We studied a wide variety of differential distributions and compared the sensitivity they provided in order to identify variables and combinations of variables that optimized our sensitivity



The p_{TZ} variable

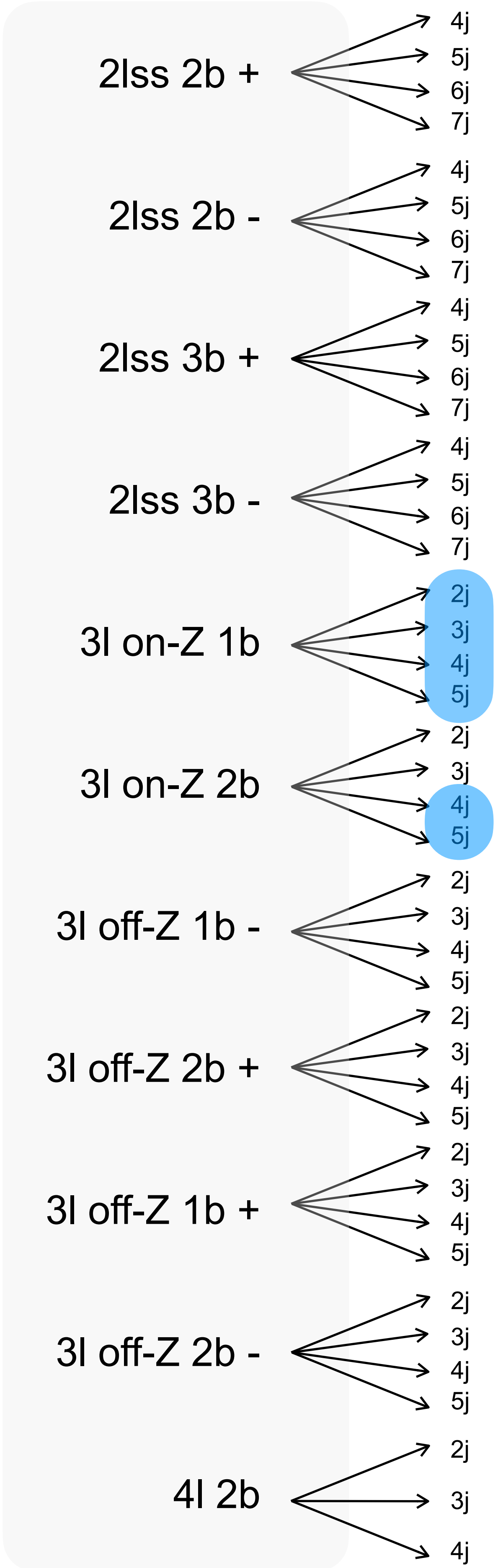
- For most of the on-Z categories, we use p_{TZ}
- Defined as the p_T of the ll pair from the Z window, helps to gain sensitivity to WCs involving vertices with Z bosons



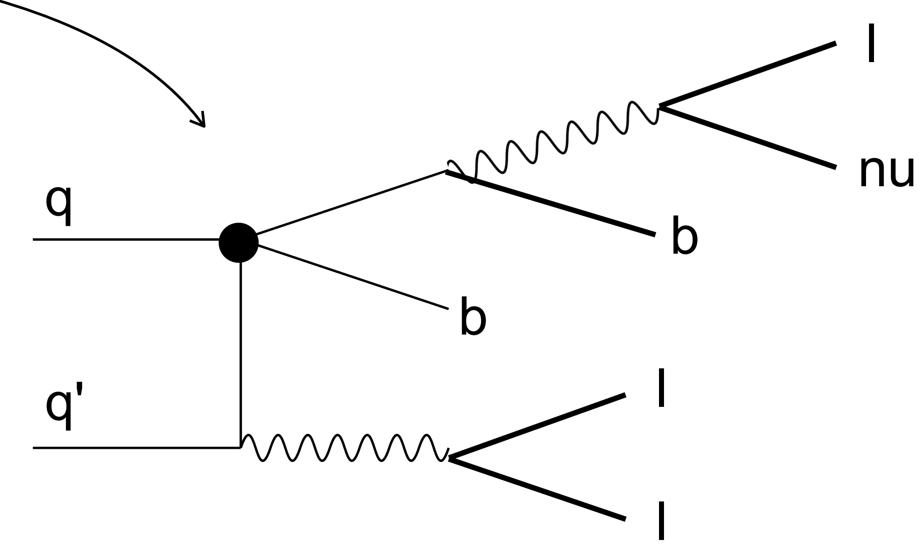


When we reweight to a non-SM point, we can see the shape and normalization of the ptz distribution changes

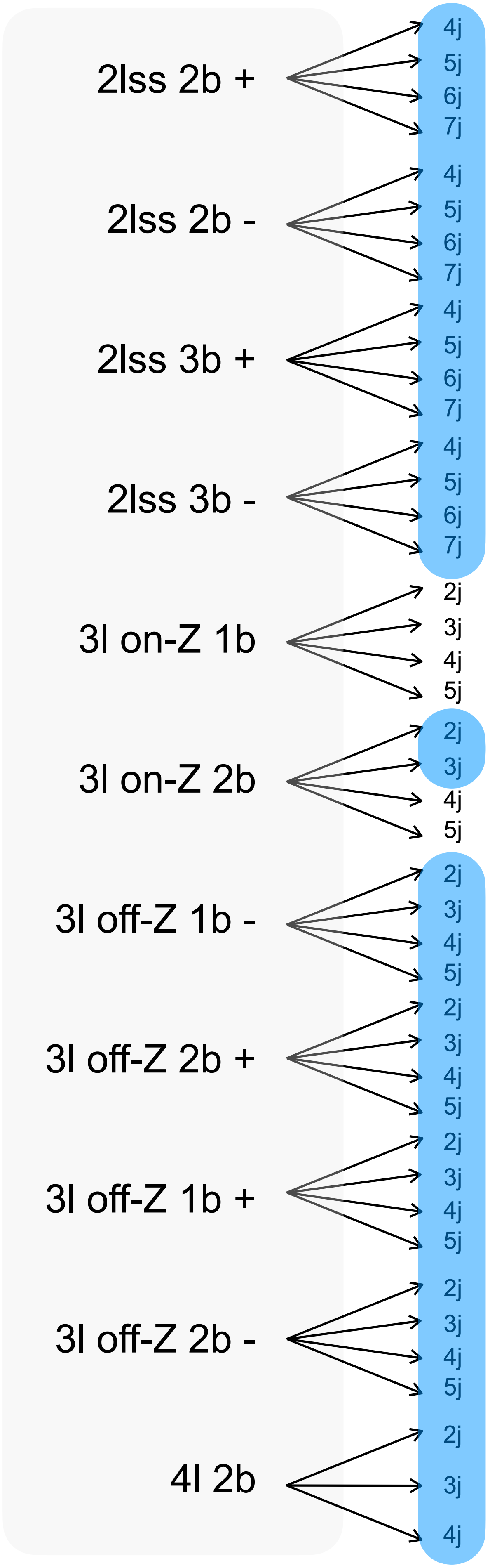
The p_{TZ} variable



- However, for two of the on-Z categories, there are also important contributions from WCs that do *not* involve a Z
- So p_{TZ} is not not a good variable to use in these bins since the Z has nothing to do with the EFT

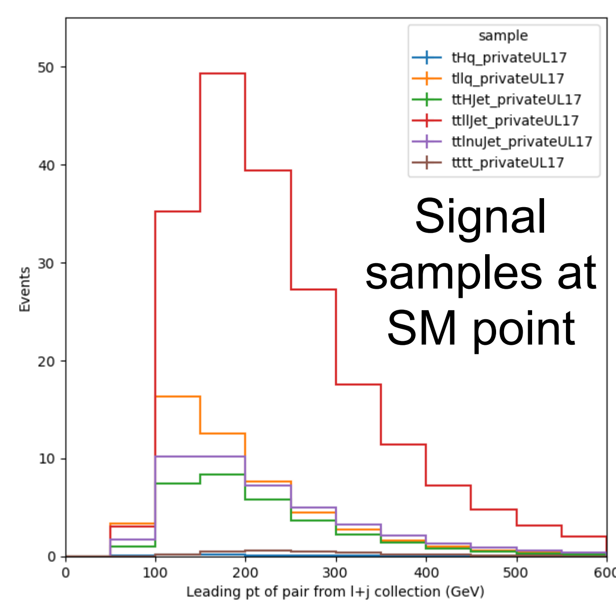


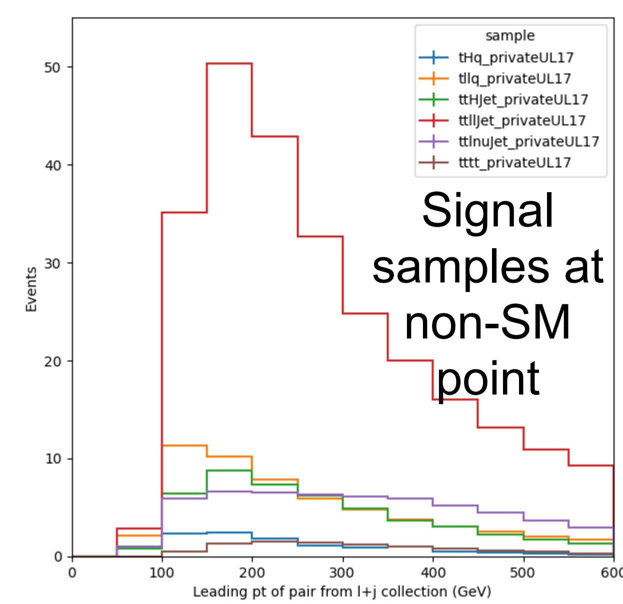
cQq13 and cQq83 contribute via diagrams like these, and in these cases the Z is not at all related to the EFT vertex



The p_{Tlj0} variable

- For all other categories, we use p_{Tlj0} (pt of the leading pair from lep+jet collection)
- Contributions from many EFT vertices grow with energy, so would make sense to use a variable related to the highest energy objects in the event
- Since the variable may contain leptons and/or jets, it provides broadly good sensitivity to various EFT effects



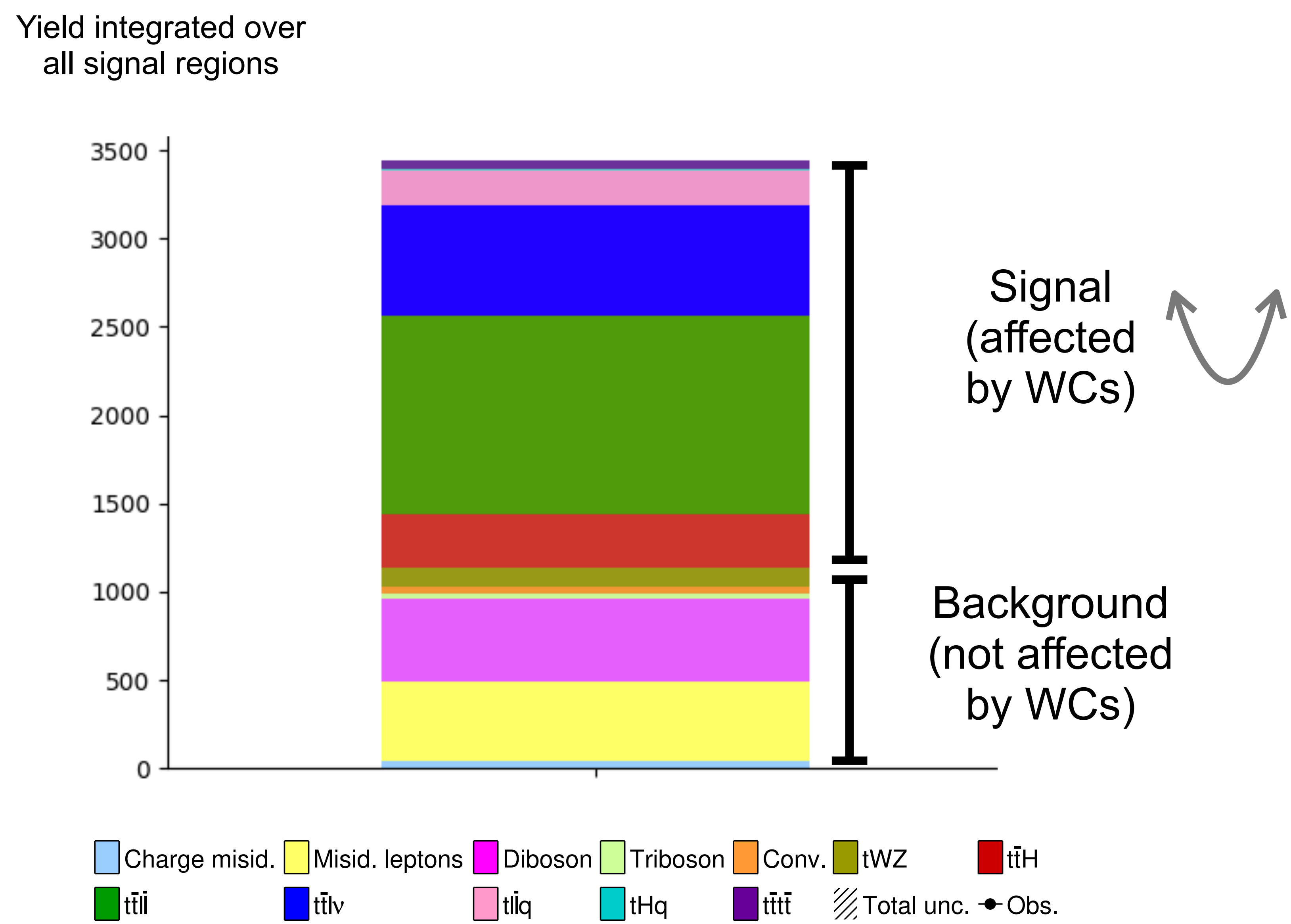


In these p_{Tlj0} distribution we can see the sensitivity to EFT effects generally comes from the high pt tail

Backgrounds and CRs

Backgrounds

- Any process that contributes to the signal region categories but is not significantly impacted by any of our 26 WCs is considered to be background
- The backgrounds make up about a third of the yield in our signal regions



Backgrounds

- Any process that contributes to the signal region categories but is *not* significantly impacted by any of our 26 WCs is considered to be background
- Backgrounds are categorized as:
    - Irreducible: All final state leptons are prompt
    - Reducible: Arise from misreconstruction or misidentification of objects

Background contributions

Reducible

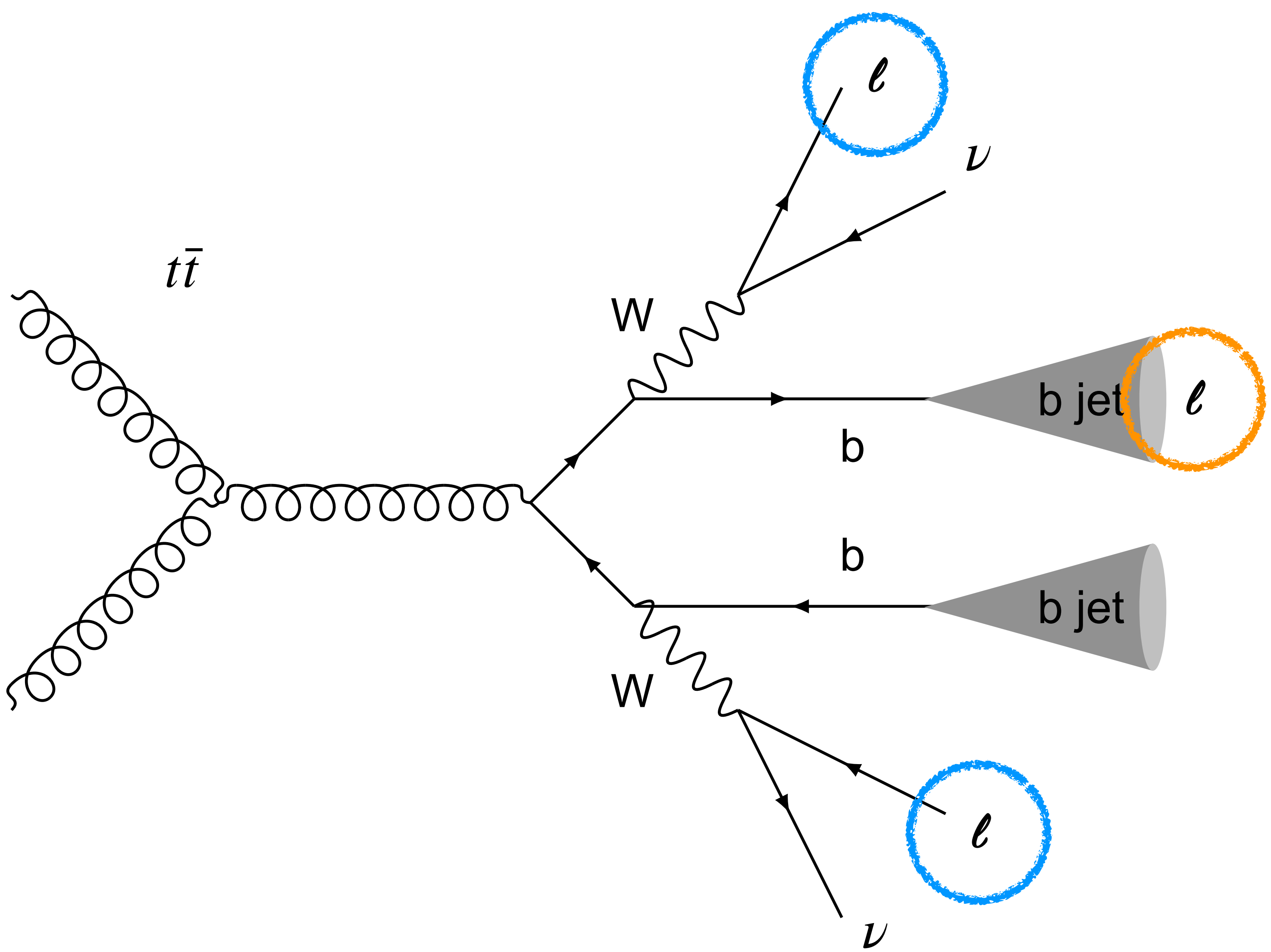
Mainly $t\bar{t}$ with a nonprompt lepton, smaller contributions from charge flips and $\gamma \rightarrow e^+e^-$ conversions

Irreducible

Mainly diboson (mostly WZ), smaller contributions from rare processes like tribosons and tWZ

The misidentified lepton background

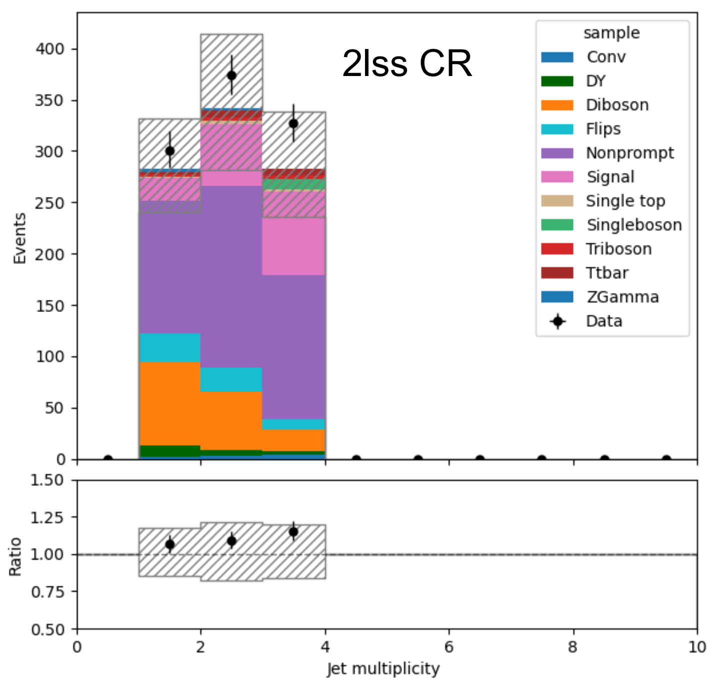
- If a lepton (e.g. from a b jet) is incorrectly identified as coming promptly from the primary vertex, then ttbar can populate our signal regions



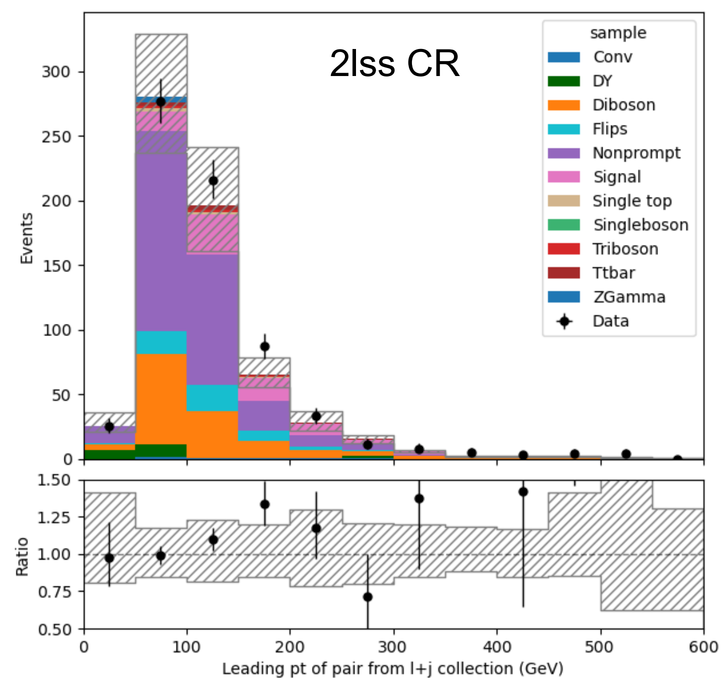
E.g. this $t\bar{t}$ event might be incorrectly categorized as a 3l off-Z signal event, when it should really be classified as a 2l opposite-sign event (which is not part of our signal selection)

Estimating the reducible background contribution

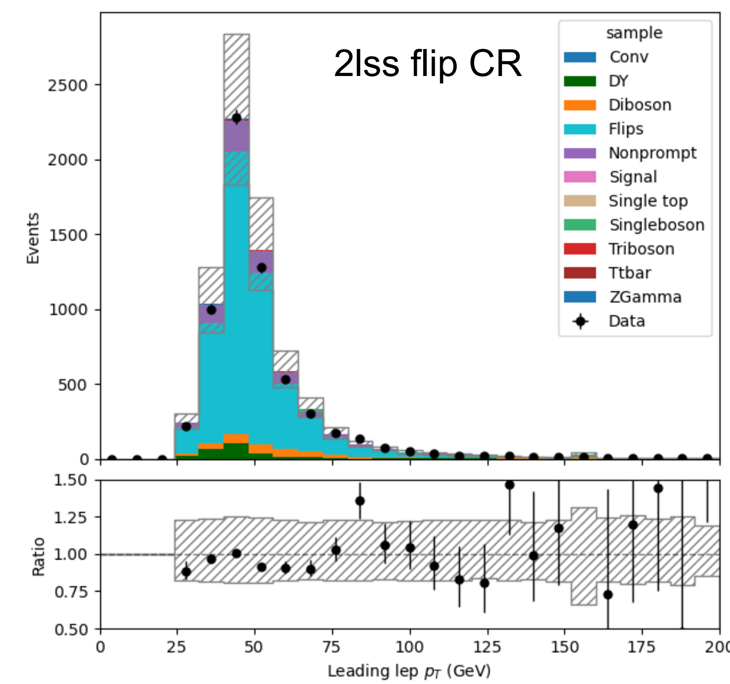
- Reducible backgrounds contribute to the signal region via misreconstructed or misidentified objects, and are modeled primarily with data-driven approaches
  - Nonprompt: Arises when objects that are not true prompt leptons pass the tight lepton selection criteria, rates measured in a region dominated by non-prompt leptons, and the contribution to the signal region (SR) is estimated by applying the rates to leptons in an application region (AR), which is identical to the SR expect requiring at least one of the leptons to be not tight
  - Charge flips: Arises from the mismeasurement of the sign of an electron in an opposite sign pair, rates are measured from MC truth information, and the contribution to the SR is estimated by applying the probabilities to events in an AR (which is identical to the SR except requiring opposite-sign pairs)
  - Conversions: Arises from $\gamma \rightarrow e^+e^-$ conversions, modeled with MC using a $t\bar{t}\gamma$ sample



jet multiplicity, UL17



pt of leading pair from l+j collection, UL17



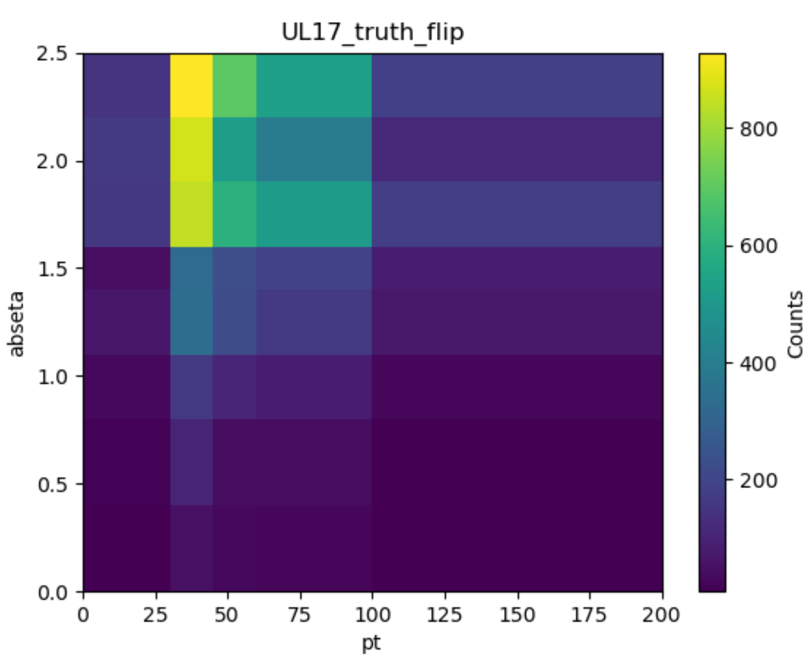
Leading lep pt, UL17

Nonprompt background estimation

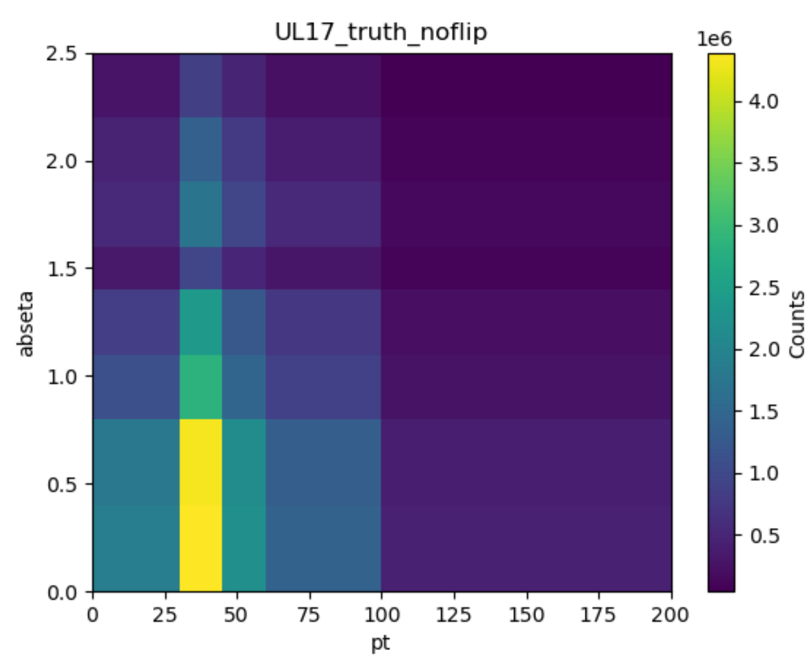
- Non prompt leptons are one of the leading sources of background, estimated with the fake-factor (FF) method, calculated by the ttH multilepton analysis
- The fake rates are measured in a multi-jet control region, these rates are then applied to leptons in the application region (which has all of the same requirements as the signal region, except the requirement that all leptons are tight is inverted, i.e. at leas one lepton must not be tight) in order to estimate the number of events in the signal region
- Uncertainties affecting the measurement of the FF weights:
    - Statistics in the MR
    - Subtraction of prompt lepton contamination in the MR
    - Differences in the background composition between the MR (dominated by multi-jet background) and the AR (dominated by ttbar+jets background)
- The effect on the fake rate due to overall uncertainty on the FF measurement is taken into account by varying the entire map of fake factor weights up and down by one standard deviation, with three separate nuisance parameters used to parameterize this uncertainty
- Additionally, the residual difference between the FF estimation with multijet and tt simulated samples is taken as an additional source of systematic uncertainties
- In addition, limited statistics in the AR of the FF method have a significant effect on the estimate of the fake rate and must be considered a separate source of uncertainty, using the Barlow-Beeston method

Charge flips (1): The measurement

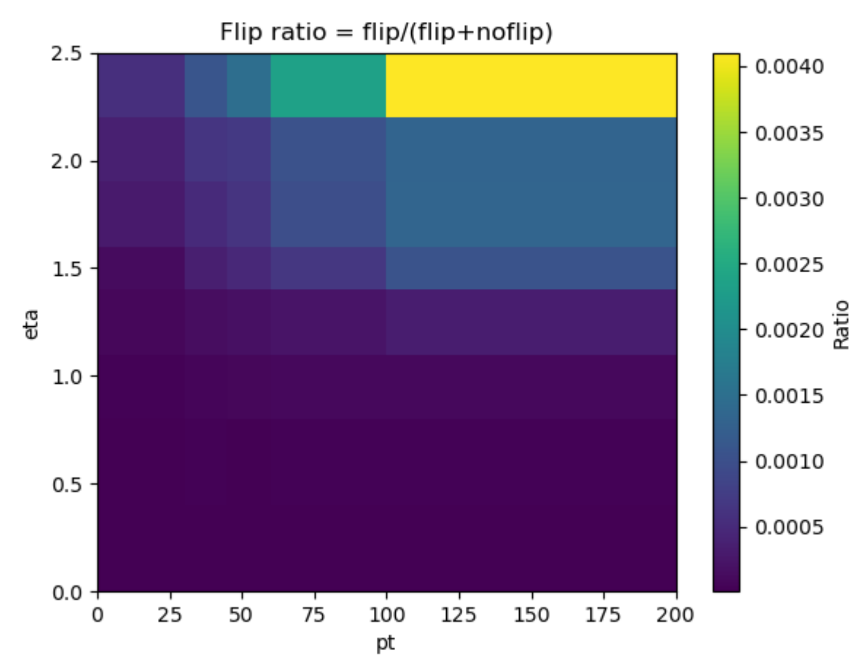
- Count the number of charge flips (according to MC truth) to find the flip probability in pt and eta bins, using DY and ttbar samples
- For this measurement, use tight electrons (with the additional requirement of tightCharge>=2, since we apply this to both electrons in the 2lss event selection)
- Using the same binning as AN-19-127
- For example, here are the measurements for 2017, seems to make sense conceptually (larger flip probability at higher pt and larger abs(eta))



Charge flips



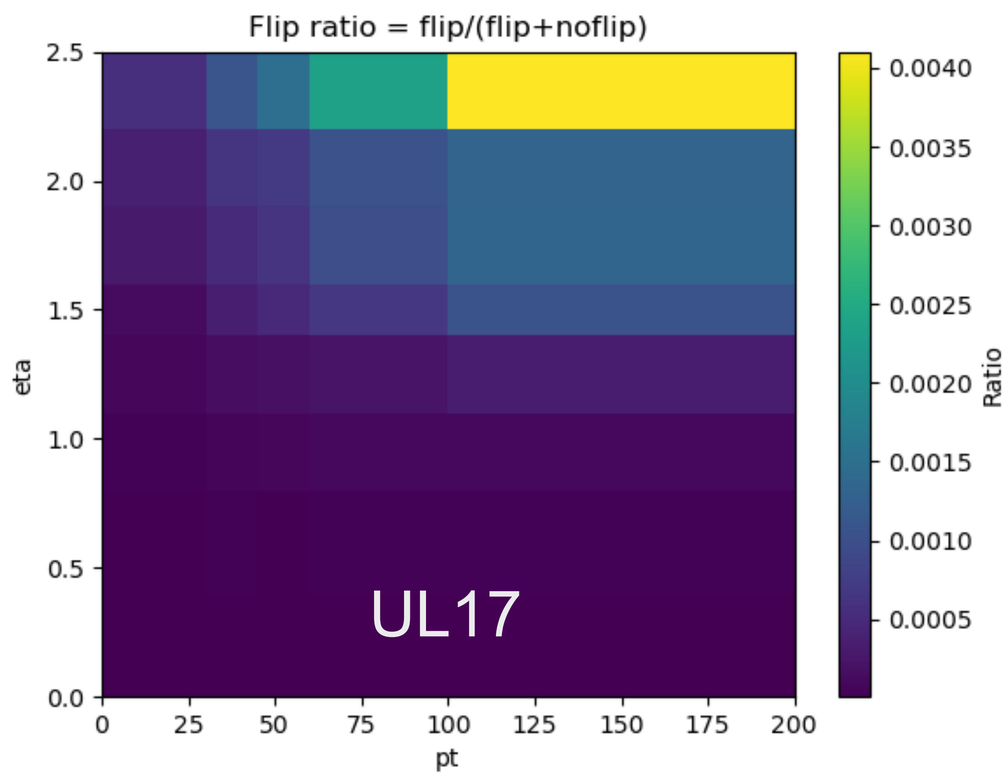
Not charge flips

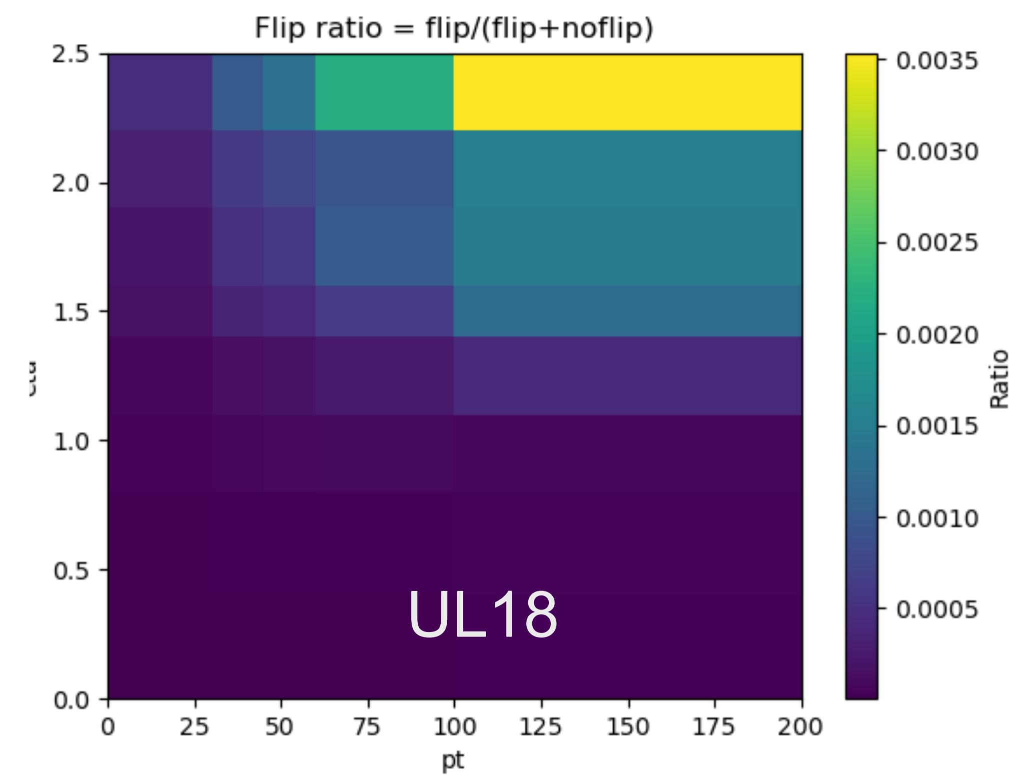


Flip probability = charge flip / (charge lips + not charge flips)

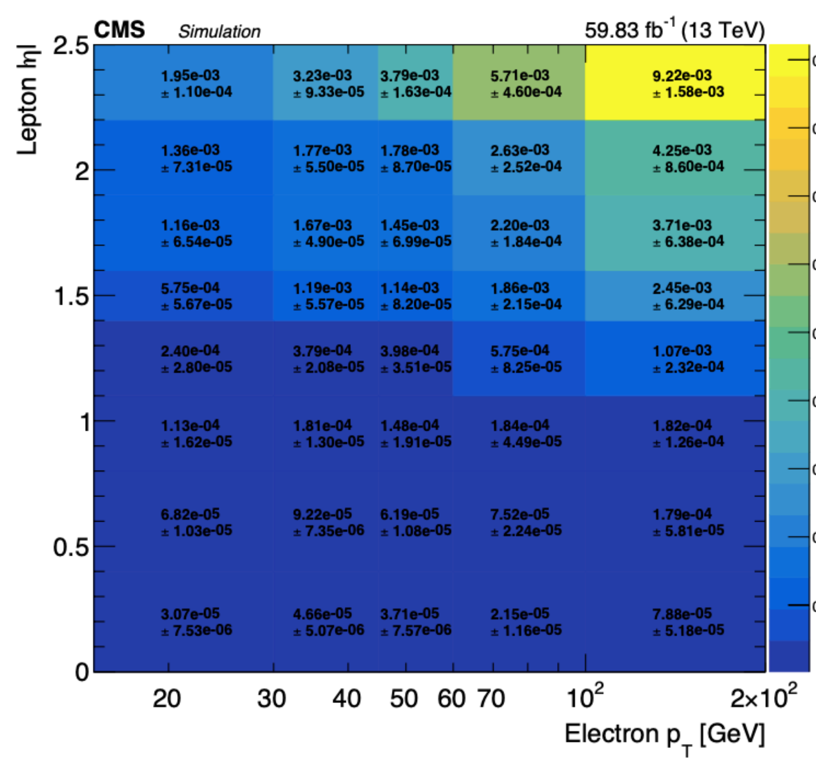
Charge flips (2): Comparing against AN-19-127 for 2017 and 2018

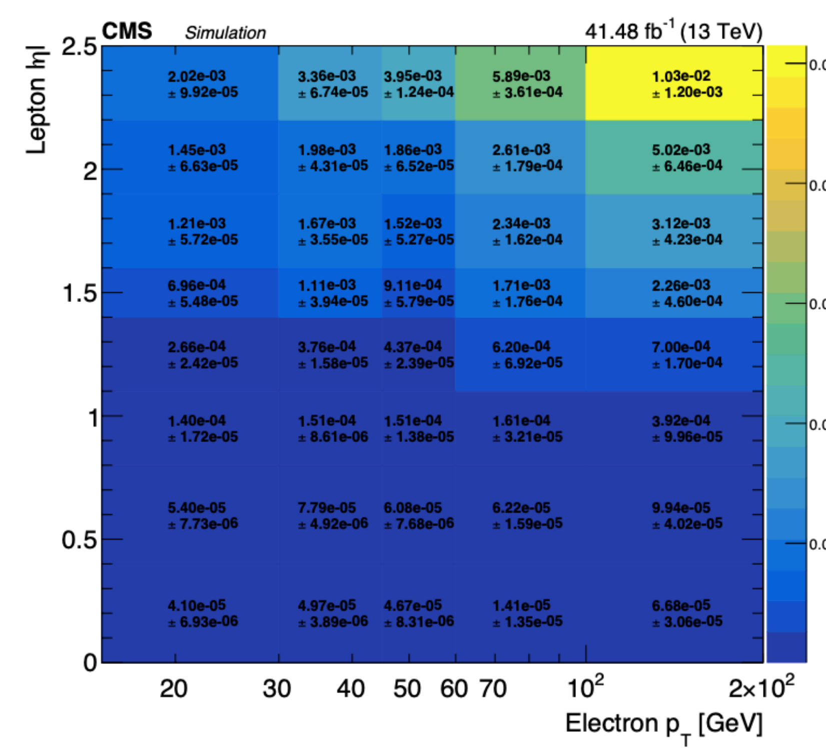
Our UL17 and UL18 are qualitatively similar, and also match AN-19-127 fairly well →





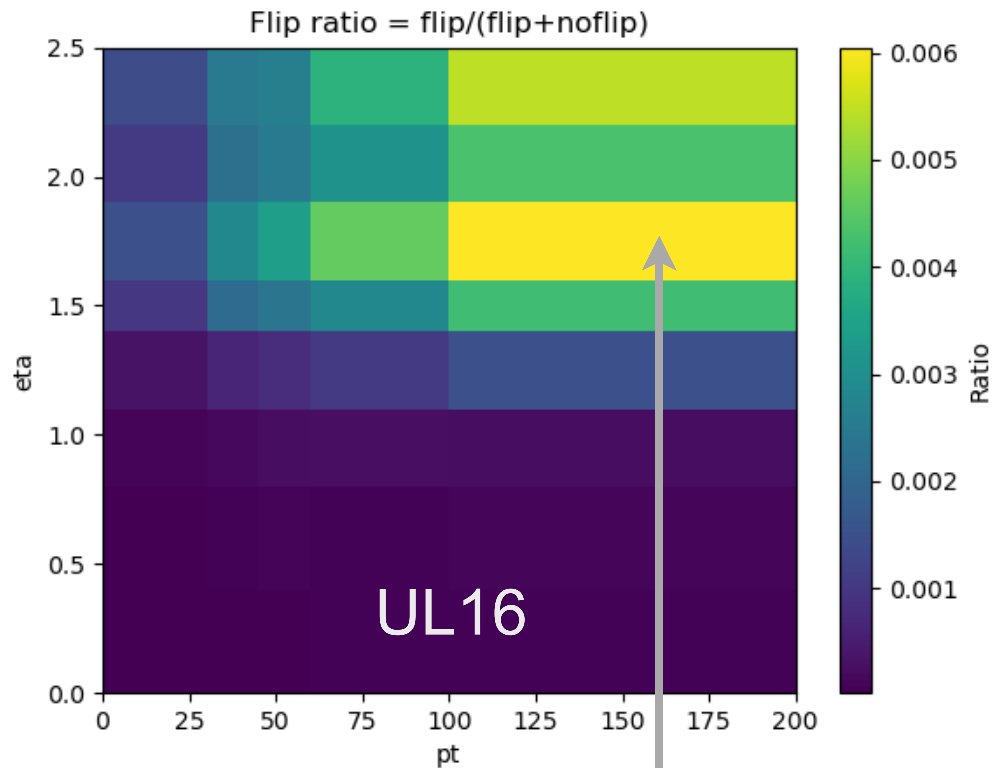
AN-19-127's numbers →

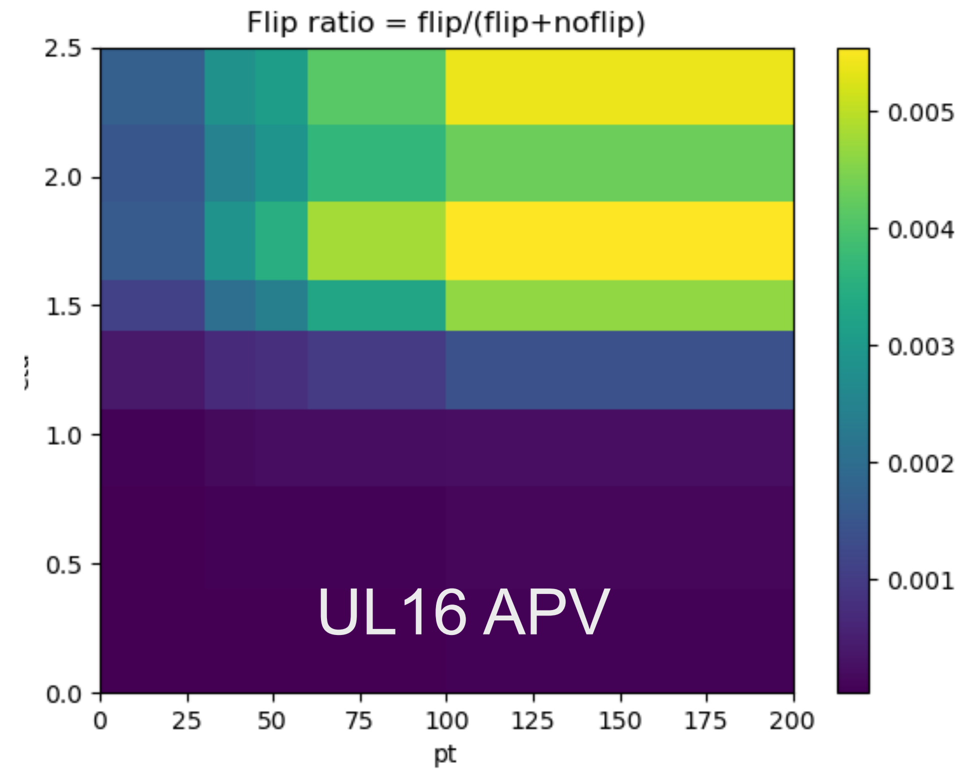




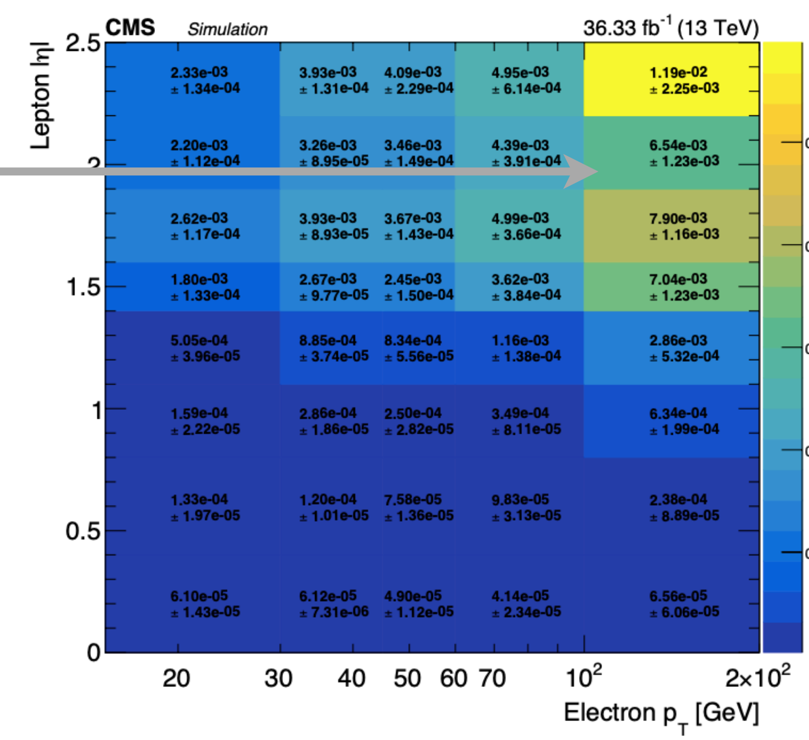
Charge flips (3): Comparing against AN-19-127 for 2016

Our UL16 and UL16 APV are similar to each other, but not to UL17/18



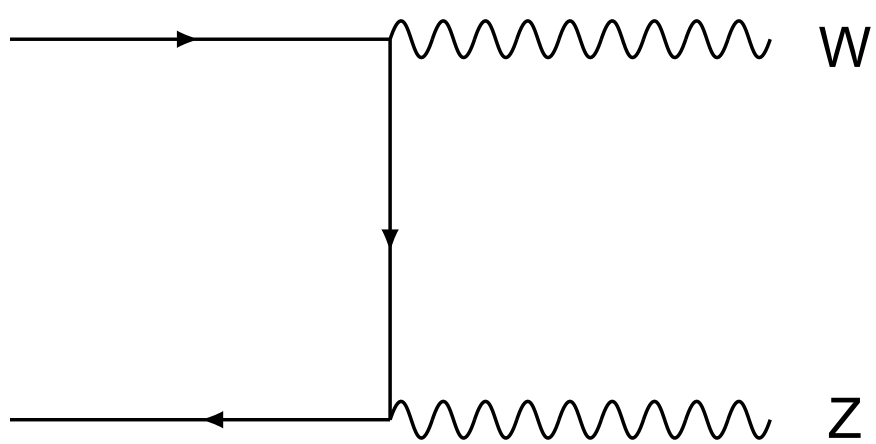


Note that our 2016 probabilities reproduce the feature at eta ~2 that is also apparent in AN-19-127's 2016 probabilities

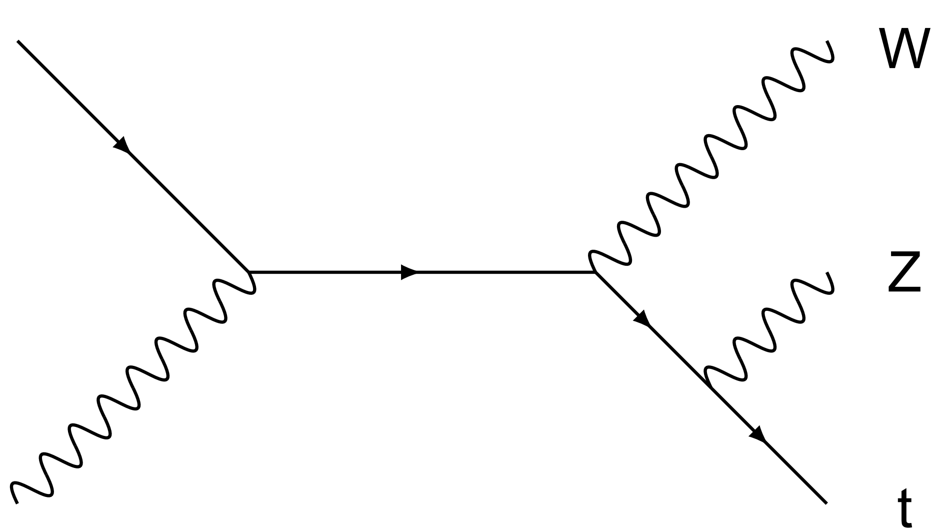


Irreducible backgrounds

- Any process that contributes to the signal region categories but is not significantly impacted by any of our 26 WCs is considered to be background
- Irreducible backgrounds come from processes that lead to the same final state lepton signatures as our signal processes



Main contributions are from dibosons (mostly WZ)

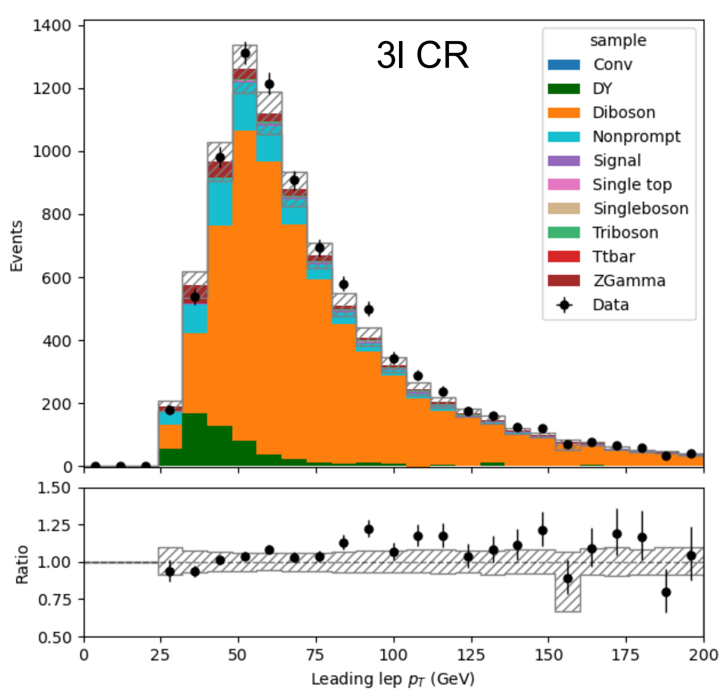


Smaller contributions from rare processes (tWZ and tribosons)

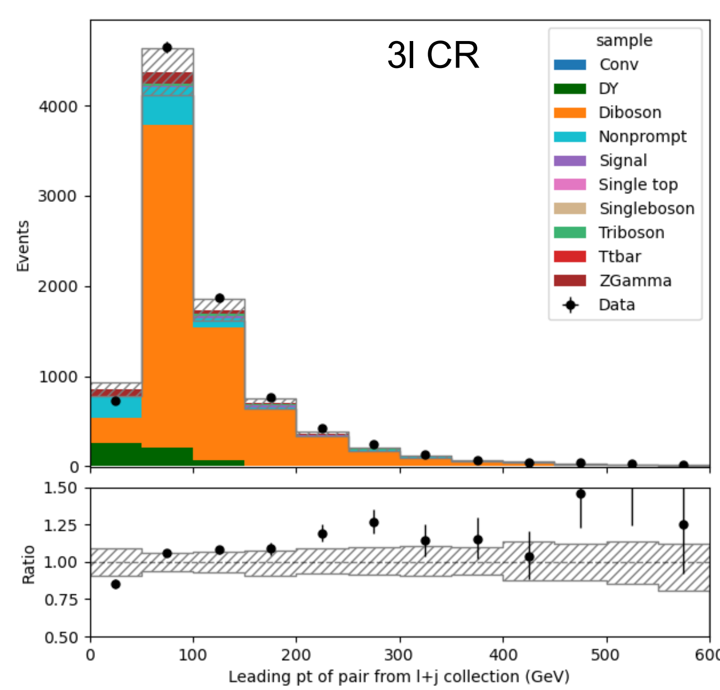
Modeling the irreducible background contribution

- Irreducible backgrounds arise from processes that lead to the same final state signatures as our signal processes
- Irreducible background are primarily diboson processes, with a very small contribution from triboson processes
- These backgrounds are modeled with MC
- In order to check the modeling we study a 3l CR dominated by dibosons (require three tight leptons, exactly 0 medium b tagged jets, no N_{jet} requirement), and as shown in the plots below, we observe reasonable agreement in this CR

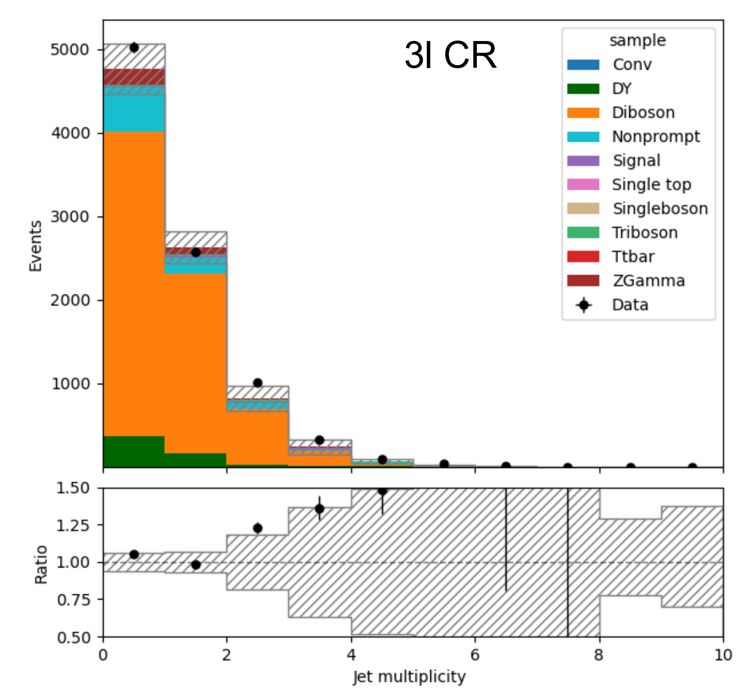
Note: In all CR plots, error bands represent the systematic uncertainties



Leading lep pt, UL17



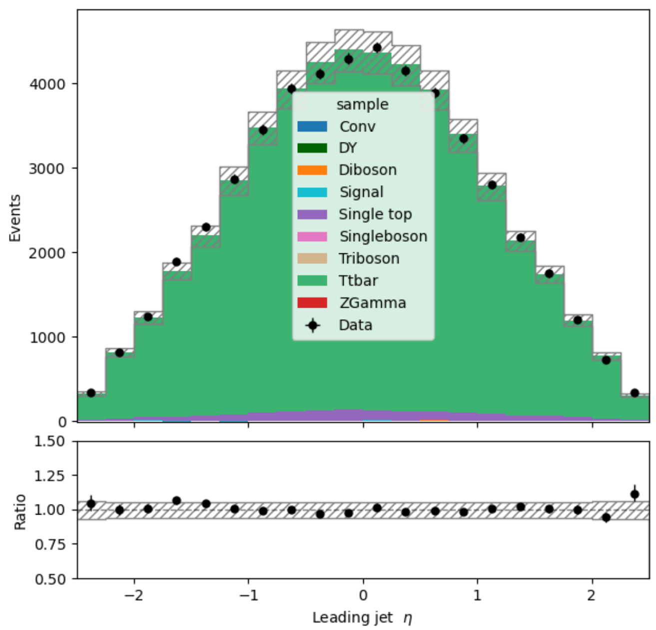
pt of leading pair from l+j collection, UL17



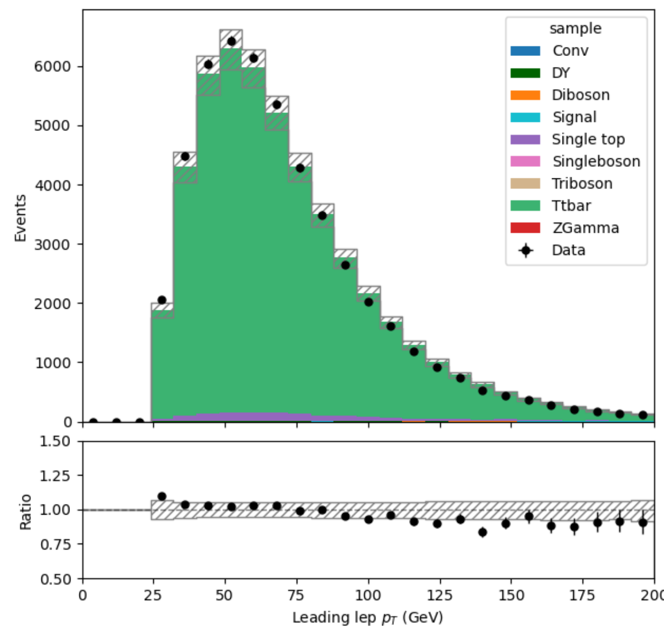
jet multiplicity, UL17

Data/MC agreement

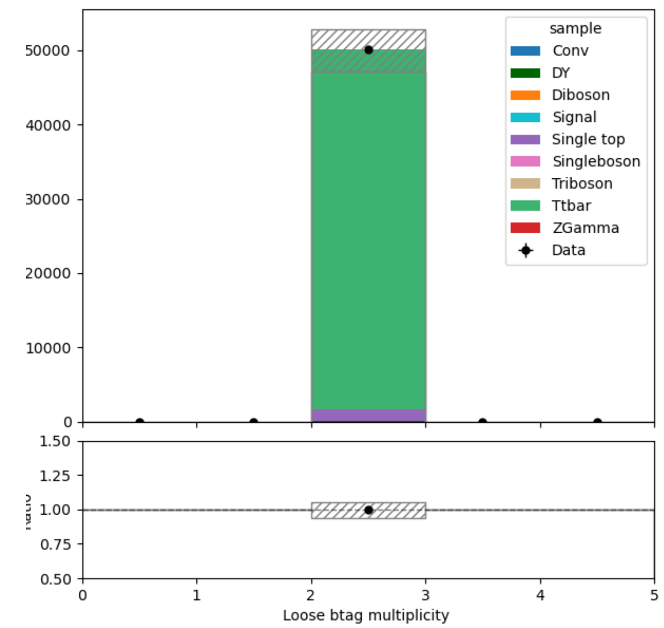
- We define CRs to check the data/MC agreement and ensure the modeling of the objects looks reasonable, from these CRs we see reasonable agreement between data/MC, the shaded bands correspond to the systematic uncertainties (full set of CR plots available in AN)



Leading jet eta, UL17

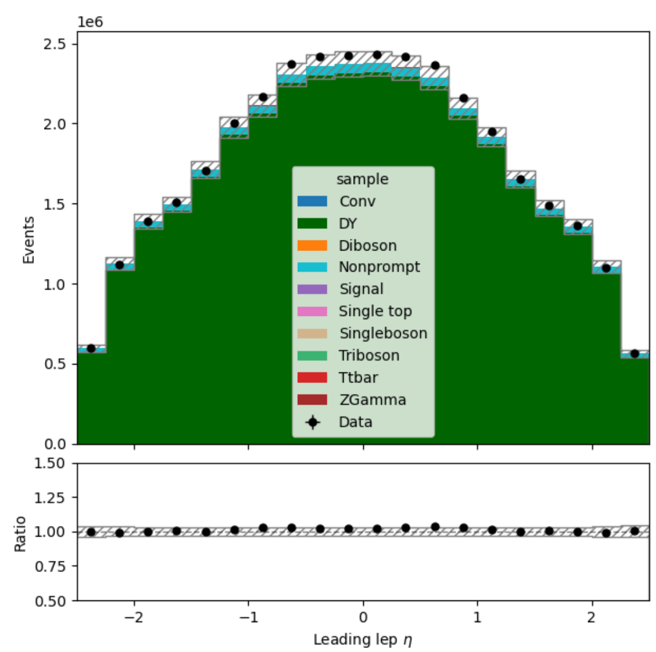


Leading lep p_T, UL17

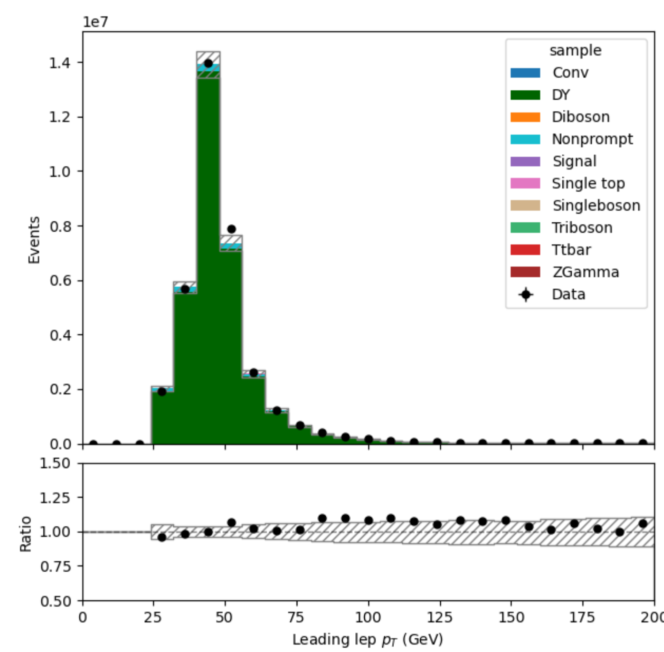


Loose btag multiplicity, UL17

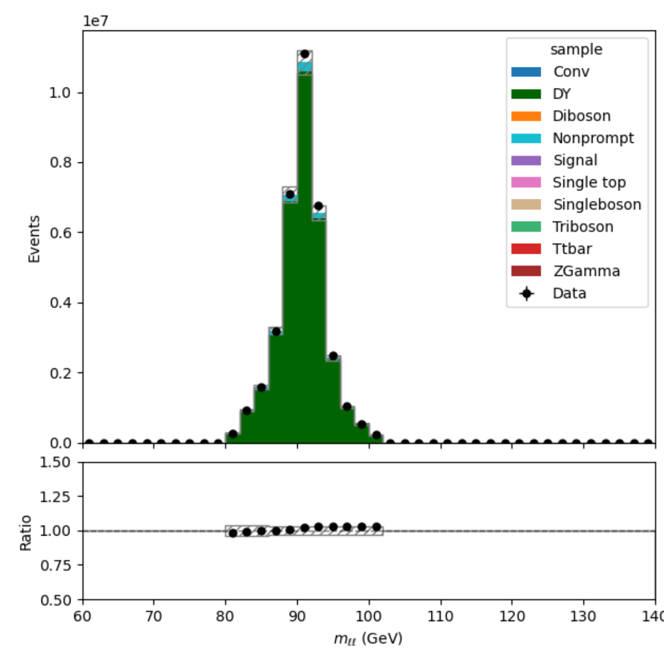
← <u>2los $t\bar{t}$ CR</u>: Require 2 opposite-sign tight leptons, two jets, 2 medium b-tags



Leading lepton eta, UL17



Leading lep p_T, UL17



Invariant mass, UL17

← <u>2los Z CR</u>: Require two opposite-sign tight leptons, m_{ll} within 10GeV of the Z, 0 medium b tagged jets (no requirement on N_{jets})

Summary of CR requirements

- 2lss: Mainly fakes (but diboson charge flips, and signal processes also contribute)
  - 2lss
  - Exactly 1 medium b
  - 1j <= njets <=2j
- 2lss flip: Mainly flips (but dibosons and fakes also contribute)
  - 2lss (required to be ee)
  - Invmass within 30 GeV of the Z
  - No requirement on b jets
  - njets <= 3j
- 2los ttbar: Mainly ttbar
  - 2los (require em)
  - Exactly 2 medium b
  - njets == 2
- 2los Z: Mainly DY
  - 2los (require ee or mm)
  - Invmass within 10GeV of the Z
  - Exactly 0 medium b
  - njets >= 0
- 3l: Mainly dibosons (but fakes and conversions also contribute)
  - 3l
  - Exactly 0 medium b tags
  - njets >= 0

Uncertainties

Summary of uncertainties

- For most WCs, the statistical uncertainties dominate over systematics
- For 6 WCs, systematics are more important
  - From 2light2heavy: cQq18, ctq8
  - From 2heavyWithBosons: ctp, ctG, cpQM, cpt
- For the systematics-dominated WCs:
  - The leading sources are the theory uncertainties on NLO xsec
  - Other important sources are various modeling-related uncertainties (e.g. ISR, FSR, renorm., diboson njet)
  - Leading experimental source is the uncertainty on non prompt background

Source	Type	Notes
Luminosity	rate	Applied to signal and all MC backgrounds
Muon ID/Iso	rate+shape	Applied to signal and all MC backgrounds
Electron ID/Iso/reco	rate+shape	Applied to signal and all MC backgrounds
Muon scale	rate+shape	Applied to signal and all MC backgrounds
Trigger efficiency	rate+shape	Applied to signal and all MC backgrounds
Pileup	rate+shape	Applied to signal and all MC backgrounds
L1 prefiring	rate+shape	Applied to signal and all MC backgrounds
JES	rate+shape	Applied to signal and all MC backgrounds
JER	rate+shape	Applied to signal and all MC backgrounds
b-Tag HF correlated	rate+shape	Applied to signal and all MC backgrounds
b-Tag HF uncorrelated	rate+shape	Applied to signal and all MC backgrounds
b-Tag LF correlated	rate+shape	Applied to signal and all MC backgrounds
b-Tag LF uncorrelated	rate+shape	Applied to signal and all MC backgrounds
Fakes FF measurement	rate+shape	Related to FF measurement for the fakes background.
Fakes AR stats	rate+shape	Due to statistics in the AR for the fakes background. Uncorrelated across event categories.
Charge Flip	rate	Uncertainty on the charge flip background
QCD Scale ($t\bar{t}H$)	rate	Scale uncertainty for theoretical cross section NLO $t\bar{t}H$ prediction
QCD Scale ($t\bar{t}\gamma$)	rate	Scale uncertainty for theoretical cross section NLO $t\bar{t}\gamma$ +jets prediction
QCD Scale ($t\bar{t}$ V)	rate	Scale uncertainty for theoretical cross section NLO $t\bar{t}l\nu$ and $t\bar{t}l\bar{l}$ prediction
QCD Scale (tHq)	rate	Scale uncertainty for theoretical cross section NLO tHq prediction
QCD Scale ($t\bar{t}t\bar{t}$)	rate	Scale uncertainty for theoretical cross section NLO $t\bar{t}t\bar{t}$ prediction
QCD Scale (V)	rate	Scale uncertainty for theoretical cross section NNLO W and Z prediction
QCD Scale (VV)	rate	Scale uncertainty for theoretical cross section NLO diboson prediction
QCD Scale (VVV)	rate	Scale uncertainty for theoretical cross section NLO triboson prediction
pdf (gg)	rate	Pdf uncertainty for theoretical cross section gg-initiated processes except ($t\bar{t}H$ $t\bar{t}l\bar{l}$, $t\bar{t}\gamma$ +jets)
pdf ($q\bar{q}$)	rate	Pdf uncertainty for theoretical cross section $q\bar{q}$-initiated processes ($tl\bar{l}q$, $t\bar{t}l\nu$, Diboson, Triboson)
pdf (qg)	rate	Pdf uncertainty for theoretical cross section qg-initiated processes (tHq)
Renorm. and fact. scales	rate+shape	Uncertainties in the matrix-element generators due to renormalization and factorization scales. Applied to signal and all MC backgrounds. Uncorrelated among processes
Parton Shower	rate+shape	Parton-shower scale uncertainties: Applied to signal and all MC backgrounds. FSR is correlated among processes. ISR is decorrelated among processes with different initial states (similar to PDF uncertainties described above).
Missing parton uncertainty	rate+shape	Introduced to account for the fact that an extra parton could not be included in the privately produced single top samples: Applied only to $tl\bar{l}q$ and tHq.
Diboson jet-dependent	rate + shape	Accounts for jet multiplicity mismodelling in diboson samples

Likelihood fitting

Statistical framework

- After event selection and histogramming -> 43 categories, 178 total bins
- The prediction in each bin is a 26d quadratic in terms of the WCs, which needs to be taken into account by the statistical framework (combine)
  - In principle, could encode the quadratic dependence by treating each term in the quadratic as a separate process in combine
  - However, interference terms can be negative (and combine does not handle negative normalizations)
  - So we use the model developed by Andrea Massironi et al. in AN-20-204 to rearrange into terms that are positive by construction

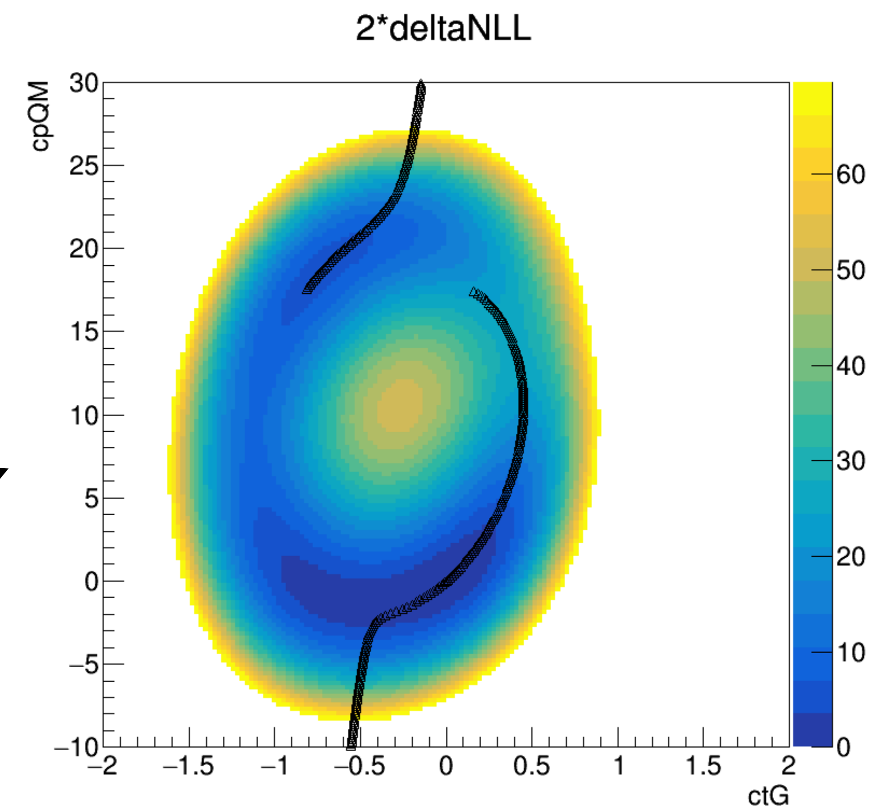
$$
\begin{aligned}
N = {} & \underbrace{S}_{\text{"sm"}} \cdot \Big(1 - \sum_i k_i + \sum_{i,i<j} \sum_j k_i \cdot k_j\Big) \\
& + \Big[\sum_i k_i - \sum_{i \neq j} k_i \cdot k_j\Big] \cdot \underbrace{(S + L_i + Q_i)}_{\text{"lin"}} \\
& + \sum_i (k_i^2 - k_i) \cdot \underbrace{Q_i}_{\text{"quad"}} \\
& + \sum_{i,i<j} \sum_j k_i \cdot k_j \cdot \underbrace{[S + L_i + L_j + Q_i + Q_j + 2 \cdot M_{ij}]}_{\text{"mixed"}}
\end{aligned}
$$

The terms "sm", "lin", "quad", and "mixed" are positive by construction, and we have a template histogram for each

- The combine PhysicsModel provides the connection between templates and WCs
- In the fit, WCs are the POIs: Scan 1 and profile the other 25

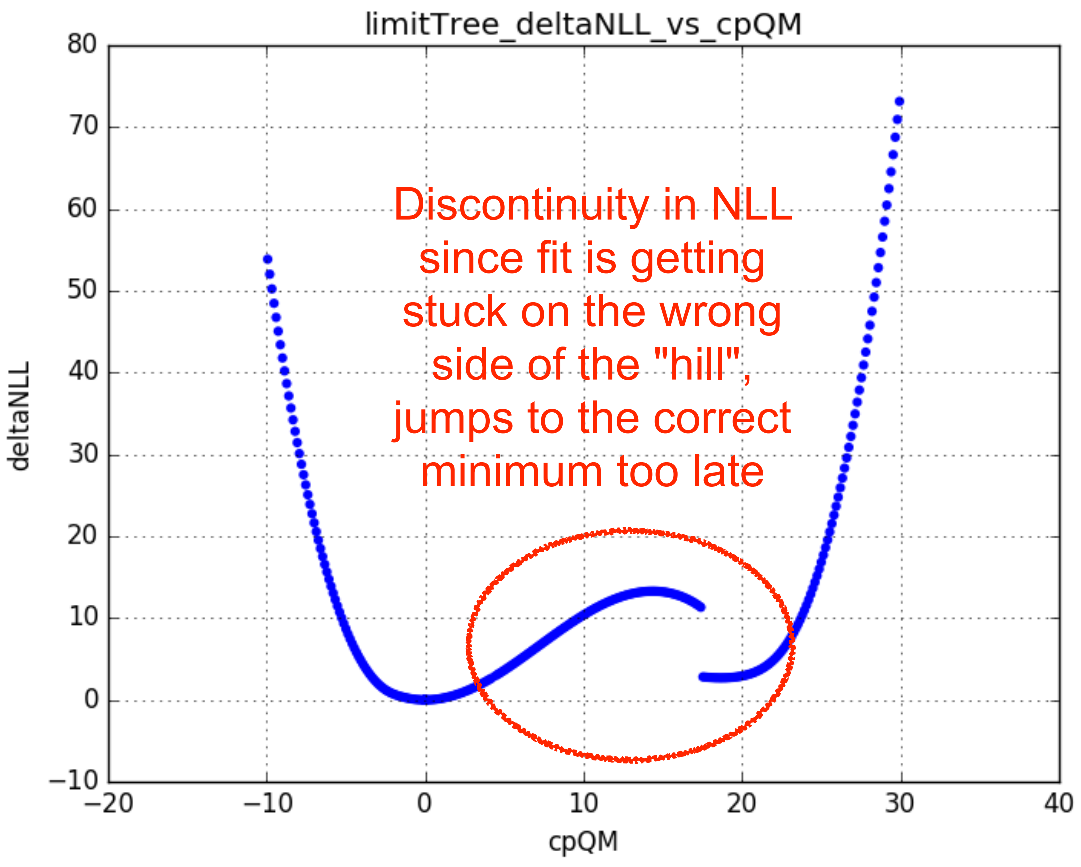
Fitting framework

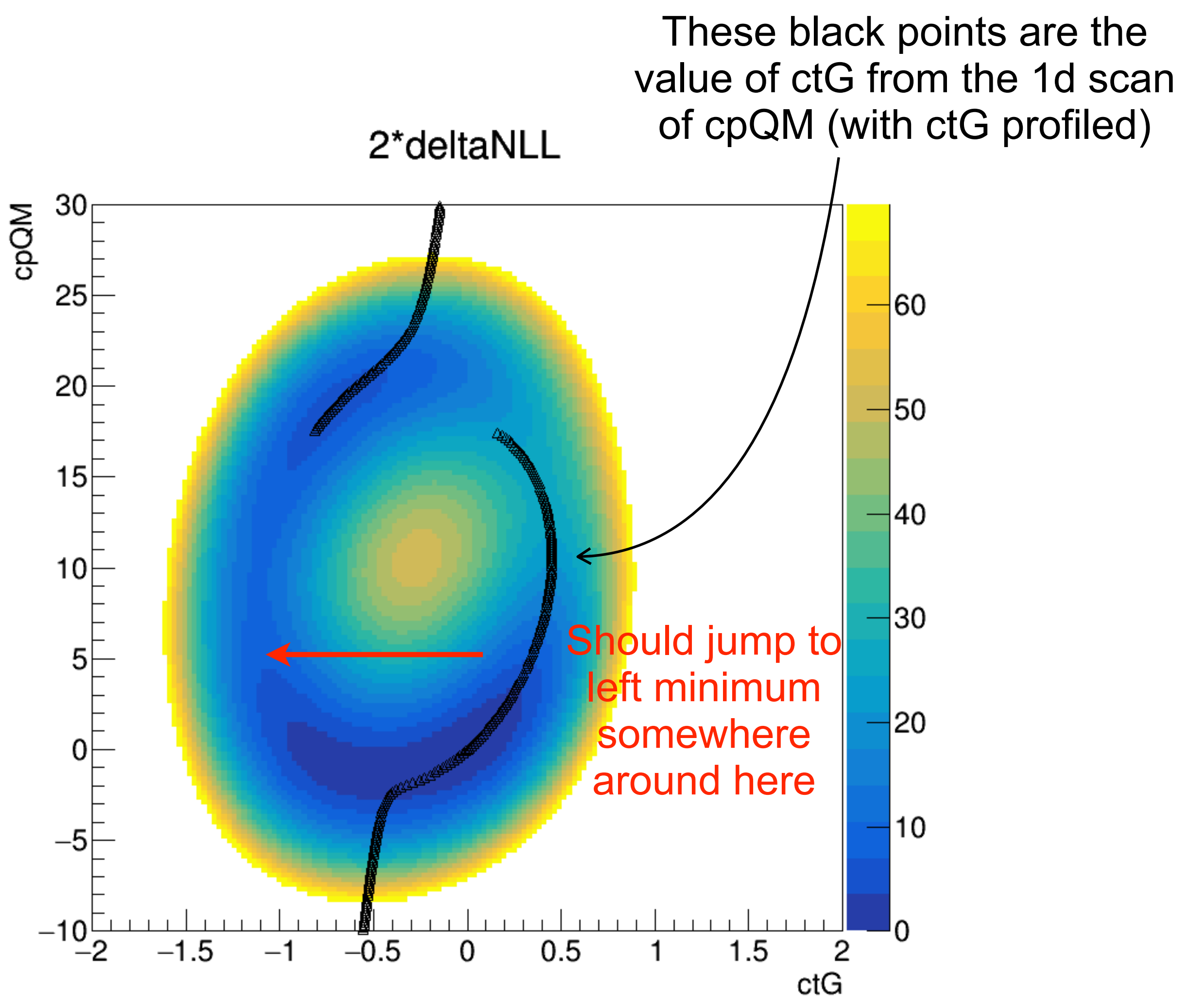
- We use Combine to perform the likelihood fit, POIs are the 26 WCs
- The yields in the bins are quadratically parameterized by the 26 WCs included in this analysis, encoded by breaking down the full 26-dimensional quadratic shape into individual templates whose normalizations are set by the PhysicsModel
- To extract the 2σ confidence intervals for each WC, we perform 1d scans for each of the 26 WCs (profiling the other parameters of interest)
- The profiled scans lead to challenges with false minima, we have developed an approach for navigating these false minima



Investigating discontinuities in likelihood scans

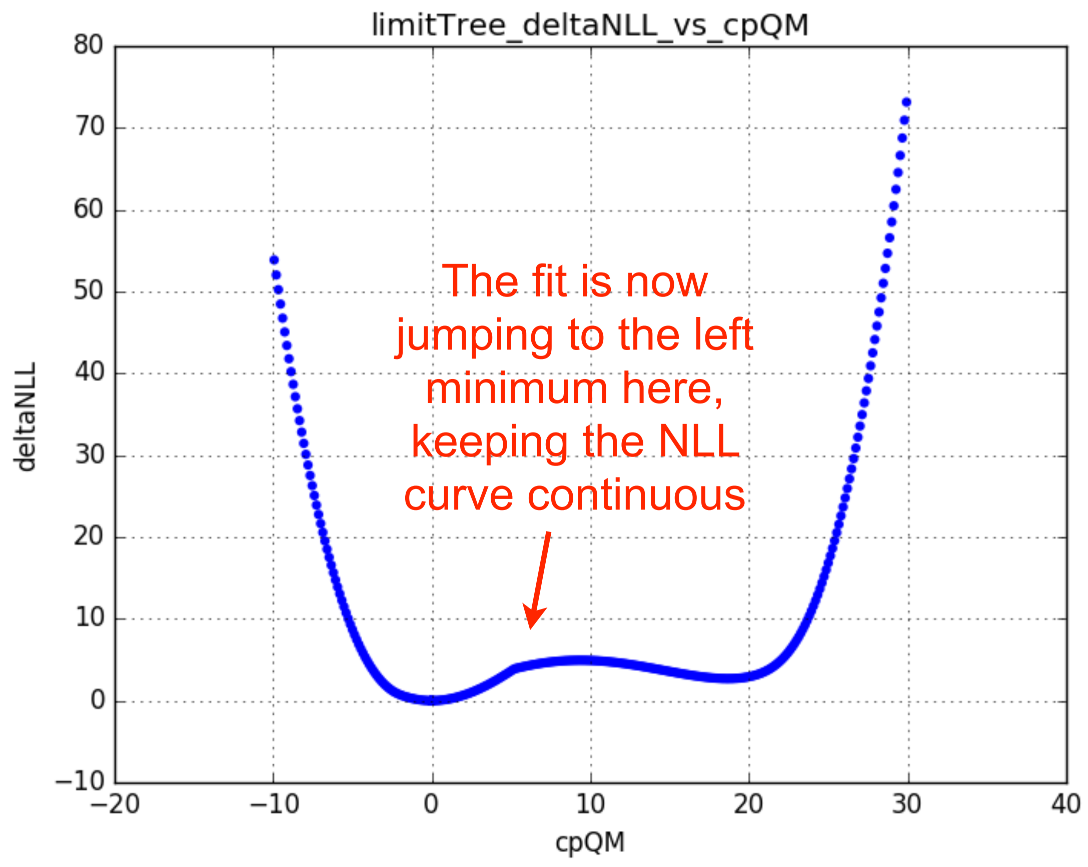
- To understand the cause of discontinuities, we looked at a 2d scan (scanning both cpQM and ctG, still ignoring all other WCs), tracked the ctG parameter, and overlaid the fitted value of ctG at each scan point
- From this we can see that the fit is getting stuck on the wrong side of a "hill" in the NLL, and cannot "see" the lower valley on the other side till it gets most of the way around the hill

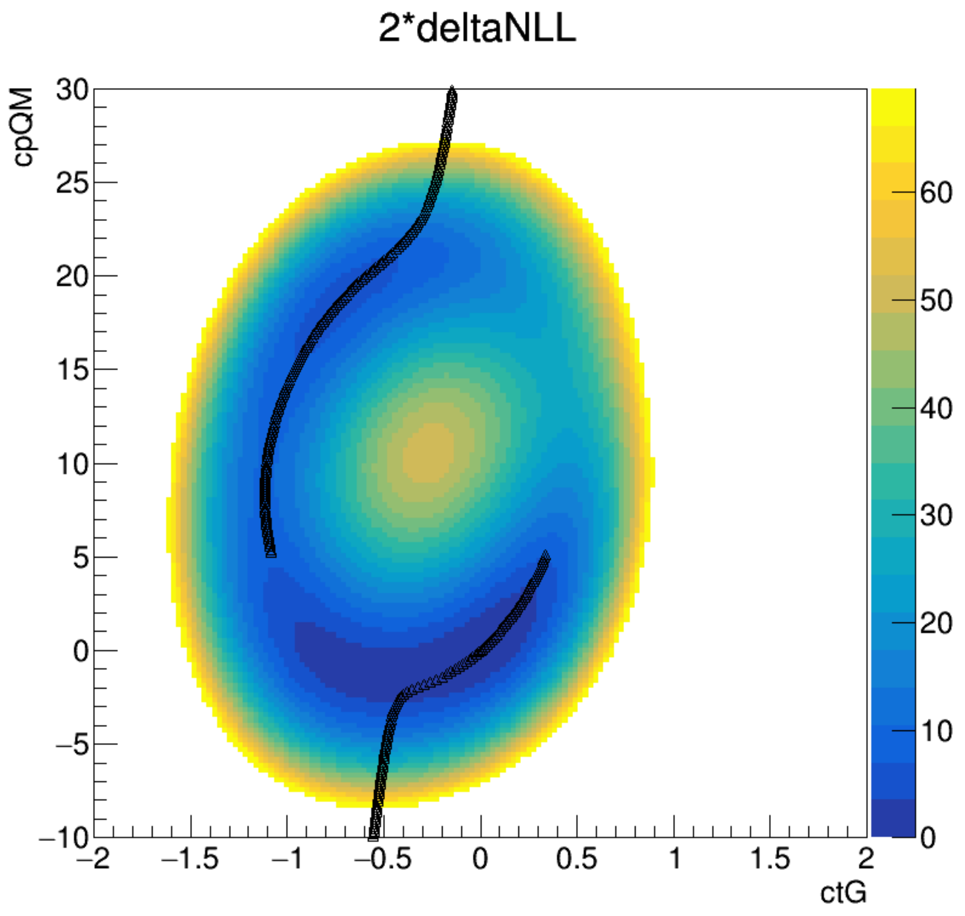




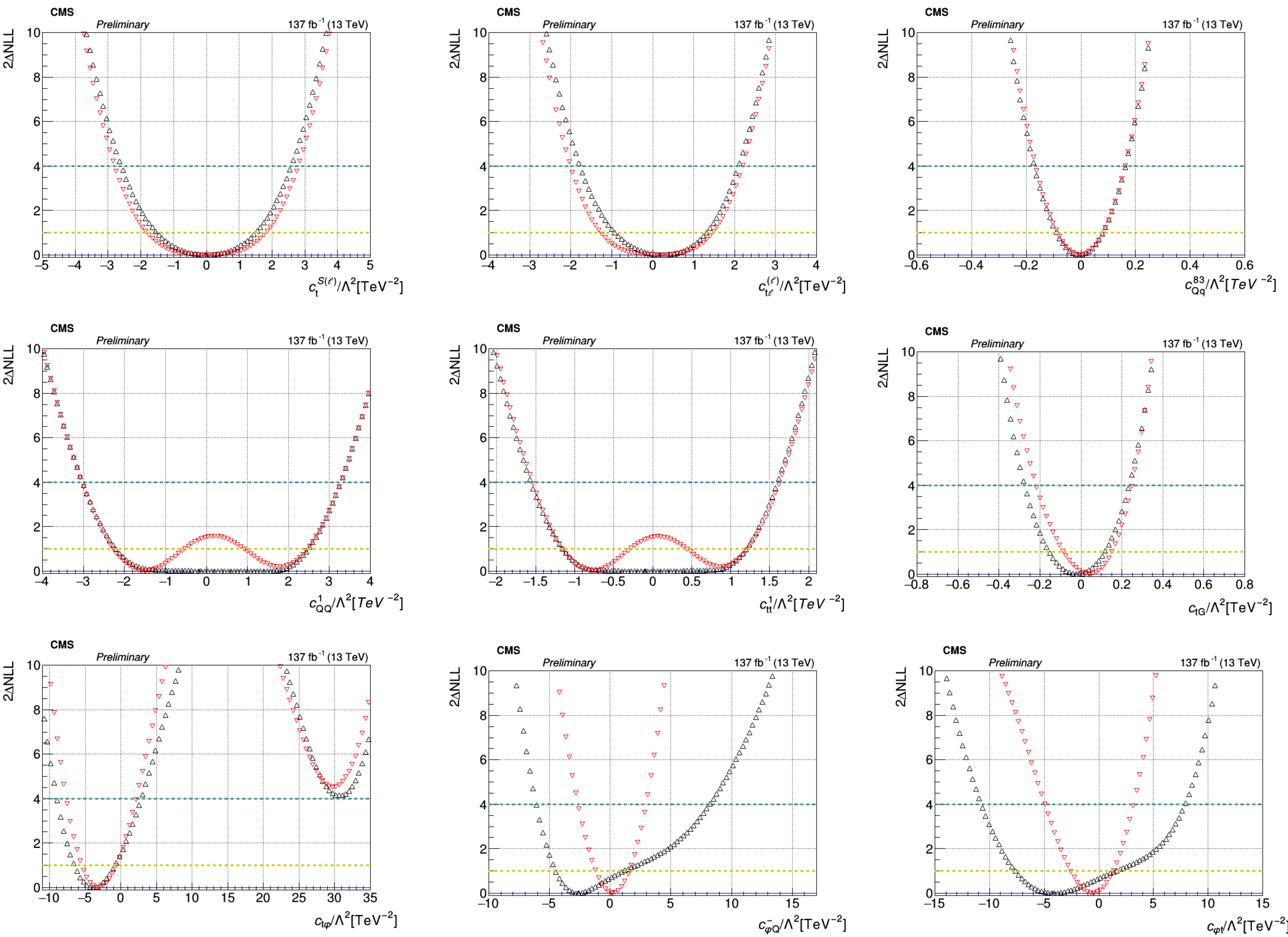
Investigating discontinuities in likelihood scans

- For the "scan-cpQM-profile-ctG" case, we tried 10 different starting points for ctG at each of the 300 scan points for cpQM
- This allowed the fit to find and jump to the right minimum at the problematic cpQM ~ 5 region
- This makes the NLL plot continuous (though its derivative does not look to be continuous, but this is probably expected)





Some example 1d scans



Discussion of results: Interpretation of sensitivity

- The sensitivity to most of the WCs comes from a wide range of bins across all selection categories

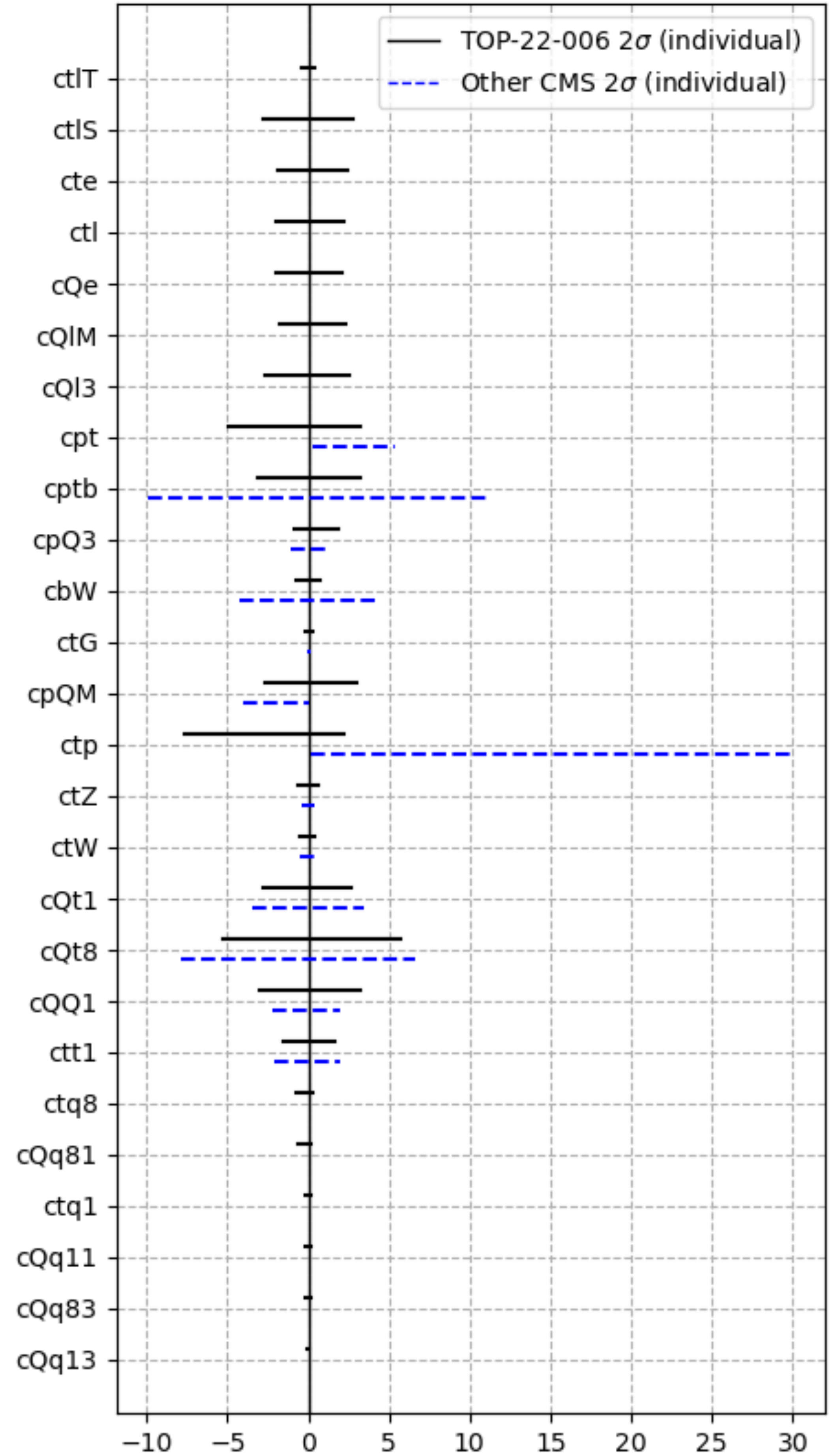
Grouping of WCs	WCs	Lead categories
Two heavy two leptons	$c_{Q\ell}^{3(\ell)}$, $c_{Q\ell}^{-(\ell)}$, $c_{Qe}^{(\ell)}$, $c_{t\ell}^{(\ell)}$, $c_{te}^{(\ell)}$, $c_{t}^{S(\ell)}$, $c_{t}^{T(\ell)}$	3ℓ off-Z
Four heavy	c_{QQ}^{1}, c_{Qt}^{1}, c_{Qt}^{8}, c_{tt}^{1}	2ℓss
Two heavy two light "$\mathrm{t\bar{t}l\nu}$-like"	c_{Qq}^{11}, c_{Qq}^{18}, c_{tq}^{1}, c_{tq}^{8}	2ℓss
Two heavy two light "$\mathrm{tl\bar{l}q}$-like"	c_{Qq}^{31}, c_{Qq}^{38}	3ℓ on-Z
Two heavy with bosons "$\mathrm{t\bar{t}l\bar{l}}$-like"	c_{tZ}, $c_{\varphi t}$, $c_{\varphi Q}^{-}$	3ℓ on-Z and 2ℓss
Two heavy with bosons "tXq-like"	$c_{\varphi Q}^{3}$, $c_{\varphi tb}$, c_{bW}	3ℓ on-Z
Two heavy with bosons with significant impacts on many processes	c_{tG}, $c_{t\varphi}$, c_{tW}	3ℓ and 2ℓss

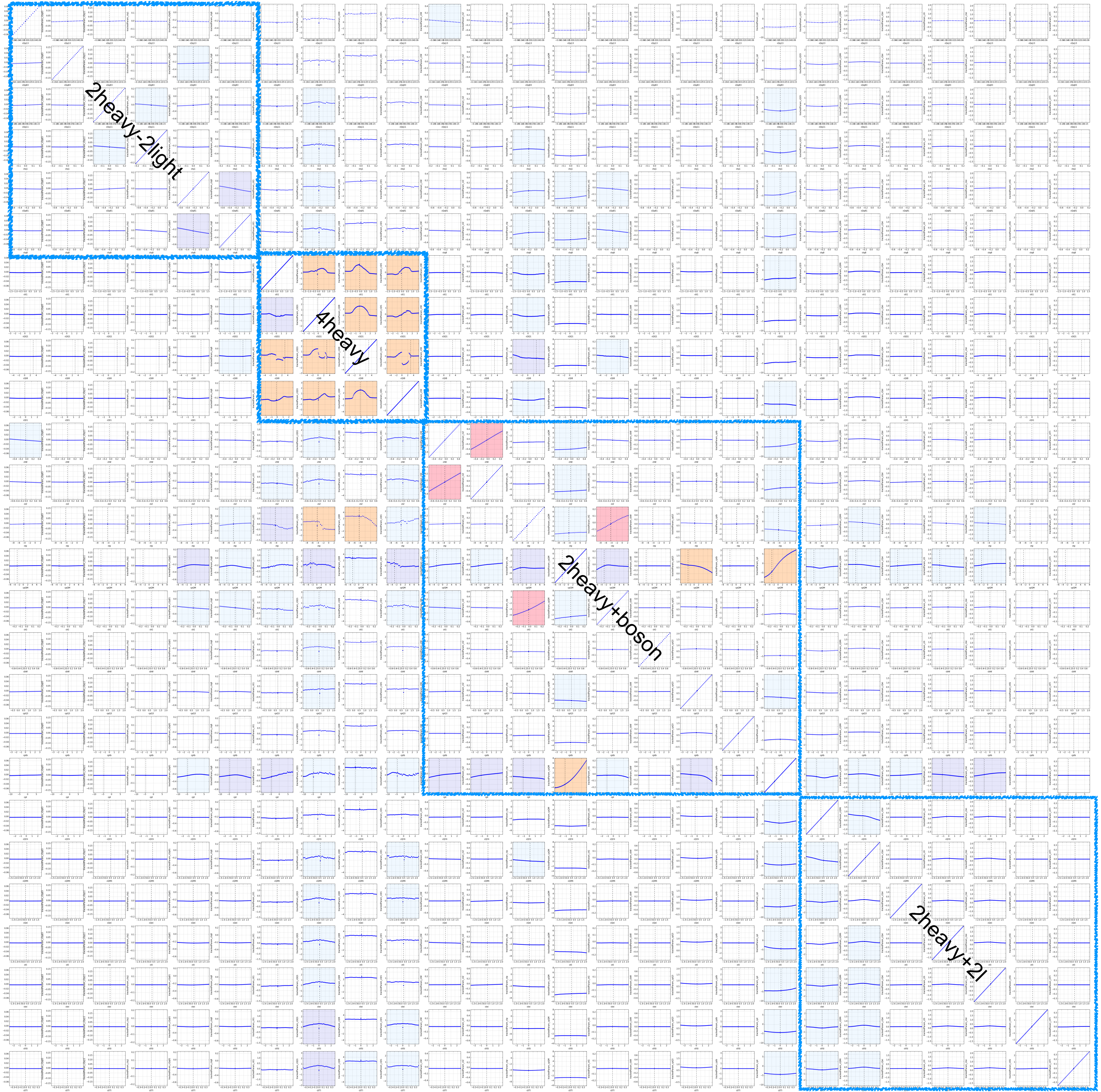
For these WCs, the sensitivity comes from a relatively clear-cut subset of the bins

For these WCs, the sensitivity comes from a more complex admixture of categories

Discussion of results: Comparisons to other results

- For most of the WCs, the results of this analysis are tighter than or competitive to other CMS results
- Important to keep in mind the many differences between the analyses
- Could be useful to perform a combination with other analyses that study orthogonal signal regions
- Can also apply techniques and ideas from other analyses to next iteration of this analysis
    - Other final states or processes (e.g. 2los, $t\bar{t}\gamma$)
    - More sensitive differential distributions (e.g. angular distributions)




2heavy-2light
4heavy
2heavy+boson
2heavy+2l

Comparing against the correlation matrices

- The plots on the previous page (organized by the shape of the tracked parameter paths) share a lot of features with the correlation matrices (top: regular, bottom: with robust hess)
- Though we should expect to see some differences, as the correlation plots are related to how things look very close to the minimum, while the tracked param plots show how things look over the whole 2sig intervals

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1	1	0.03	0.031	-0.00	-0.017	-0.00	-0.000	-0.00	0.00	-0.00	-0.06	-0.03	0.017	0.010	0.008	-0.0	0.015	0.001	0.02	0.00	0.004	0.001	0.001	0.003	2.52	-8.38
2	0.0307	1	-0.00	0.00	0.022	0.004	-0.000	0.00	-0.00	-0.00	-0.00	-0.00	0.002	-0.01	0.000	0.00	0.006	-0.000	-0.01	-0.00	0.003	0.000	0.00	0.000	8.25	8.84E
3	0.031	-0.00	1	-0.07	0.049	0.000	-0.000	0.00	-0.00	-0.00	-0.01	0.02	-0.004	-0.00	-9.70	-3.9	0.001	-1.69E	0.000	0.000	0.001	0.001	0.000	0.000	-0.00	-0.00
4	-0.002	0.00	-0.07	1	-0.000	-0.05	-0.001	-0.00	0.00	-0.00	-0.01	0.02	0.005	-0.00	0.002	-3.9	0.001	1.48E	-0.00	-4.60	0.000	0.001	0.001	0.001	8.88	-7.97
5	-0.017	0.02	0.049	-0.00	1	-0.17	-0.00	-0.00	0.00	-0.00	-0.00	0.00	0.056	0.134	-0.07	5.73	-0.00	0.000	0.11	-0.01	0.016	0.00	-0.00	0.014	2.65	-0.00
6	-0.009	0.00	0.000	-0.05	-0.171	1	-0.00	-0.00	0.01	0.00	-0.01	-0.00	0.077	0.068	-0.09	-0.0	0.001	0.000	0.07	-0.00	0.006	0.001	0.001	0.016	0.00	-0.00
7	-0.000	-0.0	-0.000	-0.00	-0.003	-0.00	1	0.26	-0.38	0.614	0.00	-0.01	-0.04	0.033	0.000	0.00	-0.00	-0.000	0.05	-0.00	0.003	0.000	-0.00	0.005	0.00	0.000
8	-0.001	0.00	0.001	-0.00	-0.001	-0.00	0.264	1	-0.93	0.619	0.01	-0.00	-0.03	0.017	0.015	0.00	-0.000	-0.000	0.03	0.001	-0.01	-0.00	-0.00	-0.00	0.00	0.000
9	0.004	-0.0	-0.000	0.00	0.009	0.013	-0.385	-0.93	1	-0.82	-0.00	0.00	-0.003	-0.01	-0.03	-0.0	0.006	0.000	-0.00	-0.00	0.001	0.000	0.000	-0.000	-0.00	-0.00
10	-0.002	-0.0	-0.000	-0.00	-0.000	0.00	0.614	0.61	-0.82	1	0.00	-0.01	0.050	-0.01	0.017	0.00	-0.00	0.000	-0.02	0.00	-0.00	-0.00	-0.00	-0.00	0.00	0.000
11	-0.065	-0.0	-0.01	-0.01	-0.005	-0.01	0.004	0.01	-0.00	0.00	1	0.55	0.020	0.093	-0.04	0.00	0.017	-0.002	0.112	-0.00	-0.01	0.000	-0.01	0.004	-0.00	1.09E
12	-0.033	-0.0	0.024	0.02	0.003	-0.00	-0.01	-0.00	0.00	-0.01	0.556	1	0.008	0.080	-0.01	0.00	0.004	-0.001	0.102	-2.79	0.001	0.023	-0.00	0.026	-0.00	0.000
13	0.017	0.00	-0.00	0.00	0.056	0.077	-0.04	-0.0	-0.00	0.050	0.02	0.00	1	0.088	0.478	1.4	-0.01	0.012	-0.09	0.018	-0.06	-0.00	-0.02	-0.03	0.00	0.000
14	0.010	-0.0	-0.00	-0.00	0.134	0.068	0.033	0.01	-0.01	-0.01	0.09	0.080	0.088	1	0.175	0.00	-0.07	0.004	0.76	-0.04	0.002	0.032	-0.03	0.065	-0.00	0.000
15	0.008	0.00	-9.70	0.00	-0.074	-0.09	0.000	0.01	-0.0	0.017	-0.04	-0.01	0.478	0.175	1	-0.0	-0.01	0.005	0.00	-0.00	-0.00	0.00	-0.00	0.004	0.00	0.000
16	-0.000	0.00	-3.92	-3.90	5.73E	-0.00	0.000	0.00	-0.00	0.000	0.00	0.00	1.48E	0.002	-0.00	1	0.004	-0.014	0.00	0.000	-0.000	-0.00	-0.00	-0.000	-4.81	2.73E
17	0.015	0.00	0.001	0.00	-0.008	0.001	-0.00	-0.00	0.00	-0.00	0.01	0.00	-0.017	-0.07	-0.01	0.00	1	-0.003	-0.08	-0.01	-0.00	-0.00	0.001	-0.00	9.56	-0.00
18	0.001	-0.00	-1.69	1.48	0.000	0.000	-0.000	-0.00	0.00	0.000	-0.00	-0.00	0.012	0.004	0.005	-0.0	-0.00	1	0.00	0.000	-0.000	5.77E	-0.00	-3.05	-1.71	3.55E
19	0.021	-0.0	0.000	-0.00	0.118	0.076	0.054	0.03	-0.00	-0.02	0.11	0.10	-0.09	0.769	0.008	0.00	-0.08	0.002	1	-0.07	0.026	0.02	-0.04	0.090	-0.00	0.000
20	0.004	-0.00	0.000	-4.60	-0.01	-0.00	-0.004	0.00	-0.00	0.00	-0.00	-2.79	0.018	-0.04	-0.00	0.00	-0.01	0.000	-0.07	1	-0.09	0.001	0.05	0.035	0.00	0.000
21	0.004	0.00	0.001	0.00	0.016	0.006	0.003	-0.01	0.00	-0.00	-0.01	0.00	-0.060	0.002	-0.00	-0.0	-0.00	-0.000	0.02	-0.09	1	-0.00	-0.03	-0.06	-0.00	-0.00
22	0.001	0.00	0.001	0.00	0.006	0.001	0.000	-0.00	0.00	-0.00	0.000	0.02	-0.004	0.032	0.005	-0.0	-0.00	5.77E	0.02	0.001	-0.00	1	-0.00	0.009	0.00	0.000
23	0.001	0.00	0.000	0.00	-0.003	0.001	-0.000	-0.00	0.00	-0.00	-0.01	-0.00	-0.02	-0.03	-0.00	-0.0	0.001	-0.000	-0.04	0.05	-0.03	-0.00	1	-0.03	0.00	6.51E
24	0.003	0.00	0.000	0.00	0.014	0.016	0.005	-0.00	-0.00	-0.00	0.00	0.02	-0.03	0.065	0.004	-0.0	-0.00	-3.05E	0.09	0.035	-0.06	0.00	-0.03	1	-0.00	-6.90
25	2.52E	8.25	-0.000	8.88	2.65E	0.000	0.000	0.00	-0.00	0.000	-0.00	-0.00	0.001	-0.000	0.000	-4.8	9.56E	-1.71E	-0.00	0.001	-0.00	0.000	0.000	-0.000	1	0.03
26	-8.38E	8.84	-0.000	-7.97	-0.000	-0.00	0.000	0.00	-0.00	0.000	1.09	0.000	0.000	0.000	0.000	2.73	-0.000	3.55E	0.000	0.000	-0.000	0.000	6.51	-6.90	0.03	1

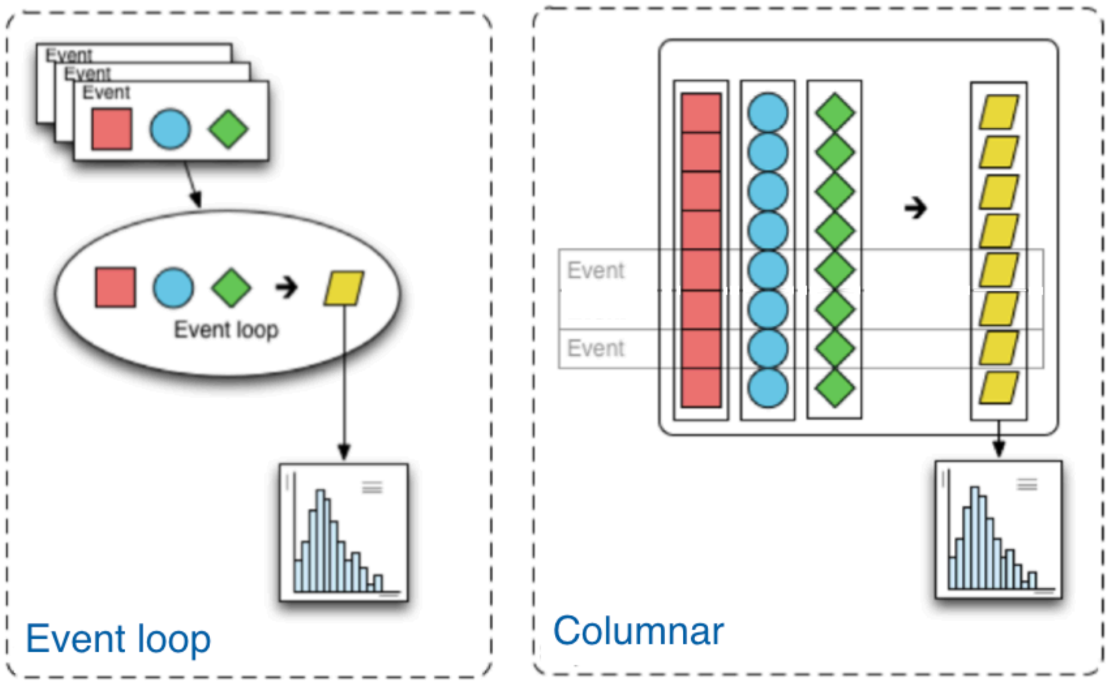
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1	1	0.03	0.033	-0.00	-0.016	-0.00	-7.96E	-0.00	0.01	-0.00	-0.07	-0.03	0.018	0.011	0.009	-0.0	0.015	0.001	0.021	0.003	0.003	0.001	0.001	0.003	3.35	-6.76
2	0.0314	1	-0.00	0.00	0.023	0.004	-0.000	-5.53	-0.00	0.000	-0.00	-0.00	0.003	-0.013	0.000	0.00	0.007	-9.13E	-0.01	-0.00	0.002	0.000	0.002	-0.000	9.05	6.25E
3	0.033	-0.00	1	-0.07	0.053	0.000	-0.001	0.00	0.00	-0.00	-0.01	0.02	-0.004	-0.003	-0.00	-2.4	0.001	-1.11E	0.000	0.000	0.001	0.001	0.000	0.000	-0.00	-0.00
4	-0.002	0.00	-0.07	1	-0.000	-0.06	-0.001	-0.00	0.00	-0.00	-0.01	0.02	0.004	-0.004	0.002	-4.7	0.001	1.01E	-0.00	0.000	0.000	0.001	0.000	0.001	4.18	-6.52
5	-0.016	0.02	0.053	-0.00	1	-0.18	0.000	-0.00	0.03	-0.00	-0.00	0.00	0.049	0.150	-0.07	-3.0	-0.01	0.000	0.13	-0.00	0.015	0.007	-0.00	0.015	-0.00	-0.00
6	-0.009	0.00	0.000	-0.06	-0.180	1	-0.001	-0.00	0.04	-0.00	-0.01	-0.01	0.066	0.089	-0.09	-0.0	-0.00	0.000	0.09	-0.00	0.005	0.001	-0.00	0.018	0.00	-0.00
7	-7.96E	-0.00	-0.001	-0.00	0.000	-0.00	1	-0.10	0.21	0.354	0.00	-0.01	-0.03	0.045	-0.00	0.00	-0.00	-0.000	0.07	-0.00	0.010	0.001	-9.71	0.010	-2.40	0.000
8	-0.001	-5.53	0.000	-0.00	-0.004	-0.00	-0.102	1	0.24	0.389	0.02	0.01	-0.065	0.012	-0.01	0.00	-0.00	-0.000	0.044	-0.00	0.007	0.000	0.000	0.007	-7.14	9.23E
9	0.011	-0.00	0.000	0.00	0.037	0.044	0.210	0.24	1	0.403	0.00	-0.01	-0.02	-0.04	-0.08	-0.0	0.013	-4.05E	0.02	0.010	-0.04	-0.00	-0.02	-0.02	-0.00	-0.00
10	-0.003	0.00	-0.001	-0.00	-0.008	-0.00	0.354	0.38	0.40	1	0.00	-0.01	-0.02	-0.02	-0.02	3.13	-0.000	-0.000	-0.03	0.000	0.004	-0.00	0.004	-0.000	0.00	5.79E
11	-0.072	-0.00	-0.01	-0.01	-0.005	-0.01	0.007	0.02	0.00	0.00	1	0.59	0.016	0.111	-0.04	0.00	0.012	-0.002	0.13	0.001	-0.01	0.002	-0.02	0.007	-0.00	6.28E
12	-0.039	-0.00	0.024	0.02	0.003	-0.01	-0.016	0.01	-0.0	-0.01	0.59	1	0.007	0.095	-0.01	0.00	0.002	-0.001	0.11	0.00	0.000	0.026	-0.00	0.029	-0.00	0.000
13	0.018	0.00	-0.00	0.00	0.049	0.066	-0.03	-0.06	-0.02	-0.02	0.016	0.00	1	0.108	0.472	-2.5	-0.02	0.011	-0.07	0.023	-0.06	-0.00	-0.02	-0.03	0.00	0.000
14	0.011	-0.0	-0.00	-0.00	0.150	0.089	0.045	0.01	-0.04	-0.02	0.111	0.09	0.108	1	0.180	0.00	-0.06	0.004	0.76	-0.03	-0.00	0.035	-0.04	0.068	-0.00	0.000
15	0.009	0.00	-0.00	0.00	-0.079	-0.09	-0.00	-0.01	-0.08	-0.02	-0.04	-0.01	0.472	0.180	1	-0.0	-0.01	0.005	0.01	-0.00	-0.00	0.00	-0.00	0.006	5.80	0.000
16	-0.000	0.00	-2.45	-4.70	-3.04E	-0.00	0.000	0.00	-0.00	3.13	0.00	0.00	-2.55E	0.002	-0.00	1	0.004	-0.012	0.00	0.000	-0.000	-0.00	-0.00	-9.16	-3.7	1.75E
17	0.015	0.00	0.001	0.00	-0.016	-0.00	-0.006	-0.00	0.01	-0.00	0.012	0.002	-0.026	-0.067	-0.01	0.00	1	-0.003	-0.07	-0.01	0.000	-0.00	0.001	-0.006	3.98	-0.00
18	0.001	-9.13	-1.11E	1.01	0.000	0.000	-0.000	-0.00	-4.05	-0.00	-0.00	-0.00	0.011	0.004	0.005	-0.0	-0.003	1	0.00	0.000	-0.000	7.49E	-0.00	-4.18	-6.29	3.90E
19	0.021	-0.0	0.000	-0.00	0.136	0.099	0.074	0.04	0.02	-0.03	0.13	0.11	-0.072	0.768	0.016	0.00	-0.070	0.002	1	-0.06	0.027	0.028	-0.05	0.095	-0.00	0.000
20	0.003	-0.00	0.000	0.00	-0.009	-0.00	-0.008	-0.00	0.01	0.000	0.00	0.002	0.023	-0.038	-0.00	0.00	-0.01	0.000	-0.06	1	-0.17	0.001	0.058	0.030	0.00	0.00
21	0.003	0.00	0.001	0.00	0.015	0.005	0.010	0.00	-0.04	0.004	-0.01	0.000	-0.062	-0.001	-0.00	-0.0	0.000	-0.000	0.02	-0.17	1	-0.00	-0.03	-0.05	-0.00	-0.00
22	0.001	0.00	0.001	0.00	0.007	0.001	0.001	0.00	-0.00	-0.00	0.00	0.026	-0.003	0.035	0.005	-0.0	-0.00	7.49E	0.02	0.001	-0.00	1	-0.00	0.012	0.00	0.00
23	0.001	0.00	0.000	0.00	-0.005	-0.00	-9.71E	0.00	-0.02	0.004	-0.02	-0.00	-0.028	-0.042	-0.00	-0.0	0.001	-0.000	-0.05	0.058	-0.03	-0.00	1	-0.02	0.00	7.19E
24	0.003	-0.00	0.000	0.00	0.015	0.018	0.010	0.00	-0.02	-0.00	0.00	0.02	-0.03	0.068	0.006	-9.1	-0.00	-4.18E	0.09	0.030	-0.05	0.012	-0.02	1	-0.00	-2.24
25	3.35E	9.05	-0.000	4.18	-0.000	0.000	-2.40E	-7.14	-0.00	0.000	-0.00	-0.00	0.001	-0.000	5.80E	-3.7	3.98E	-6.29E	-0.00	0.000	-0.00	0.000	0.000	-0.000	1	0.03
26	-6.76E	6.25	-0.000	-6.52	-0.000	-0.00	0.000	9.23	-0.00	5.79E	6.28	0.000	0.000	0.000	0.000	1.75	-0.00	3.90E	0.00	0.000	-0.000	0.000	7.19E	-2.24	0.03	1

Computational aspects of the analysis

Minimizing analysis turnaround time with coffea

- Since the nanoAOD → histogram step in the analysis workflow is usually run many times as the analysis is developed and improved, our goal was to minimize turnaround time of this step

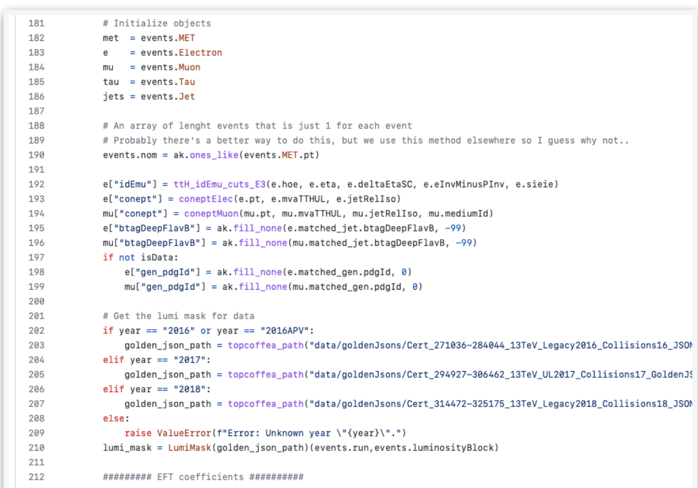
- We chose to use the Coffea framework for this step
  - Uses a columnar approach
  - Relies on the scientific python ecosystem and specialized packages for HEP



https://coffeateam.github.io/coffea/_images/columnar.png

- Scale out using the Work Queue framework to build manager-worker applications and manage the remote tasks

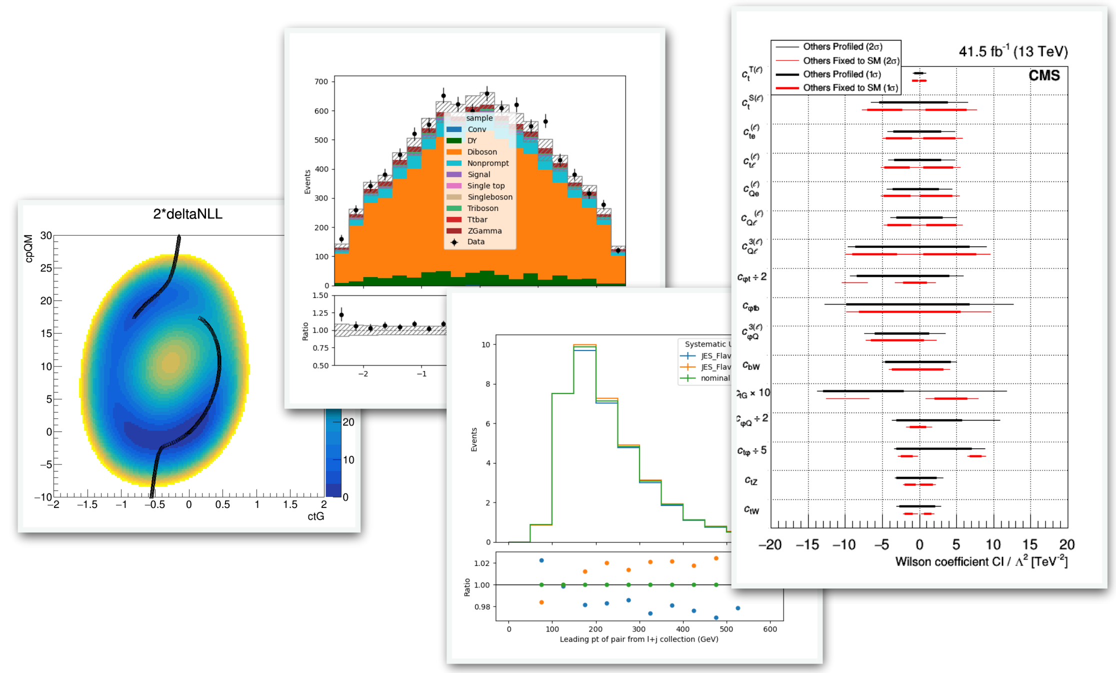
Mutually beneficial collaboration with CS colleagues



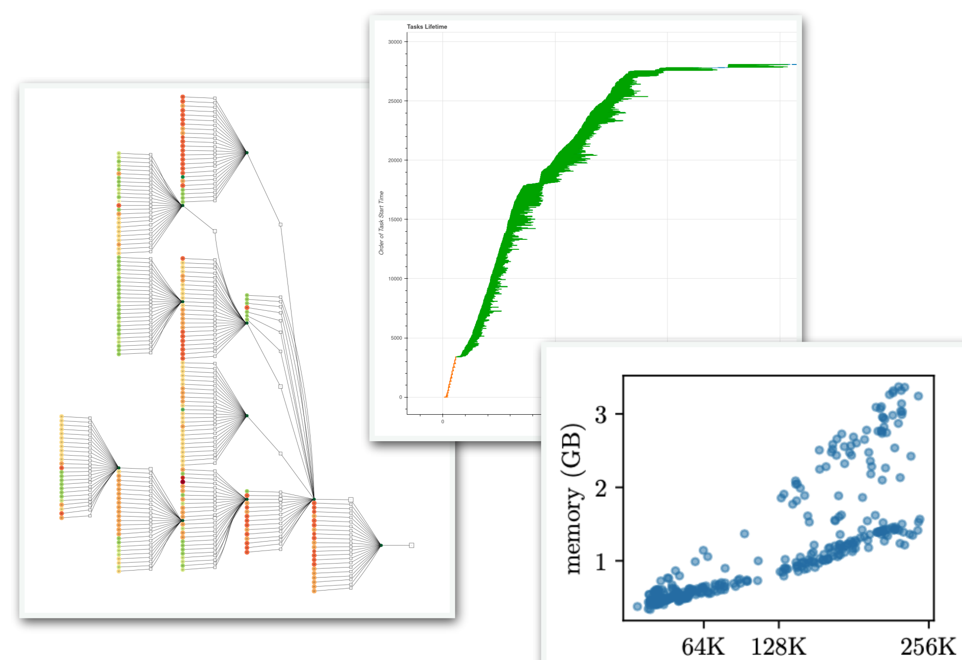
```python
        # Initialize objects
        met  = events.MET
        e    = events.Electron
        mu   = events.Muon
        tau  = events.Tau
        jets = events.Jet

        # An array of lenght events that is just 1 for each event
        # Probably there's a better way to do this, but we use this method elsewhere so I guess why not..
        events.nom = ak.ones_like(events.MET.pt)

        e["idEmu"] = ttH_idEmu_cuts_E3(e.hoe, e.eta, e.deltaEtaSC, e.eInvMinusPInv, e.sieie)
        e["conept"] = coneptElec(e.pt, e.mvaTTHUL, e.jetRelIso)
        mu["conept"] = coneptMuon(mu.pt, mu.mvaTTHUL, mu.jetRelIso, mu.mediumId)
        e["btagDeepFlavB"] = ak.fill_none(e.matched_jet.btagDeepFlavB, -99)
        mu["btagDeepFlavB"] = ak.fill_none(mu.matched_jet.btagDeepFlavB, -99)
        if not isData:
            e["gen_pdgId"] = ak.fill_none(e.matched_gen.pdgId, 0)
            mu["gen_pdgId"] = ak.fill_none(mu.matched_gen.pdgId, 0)

        # Get the lumi mask for data
        if year == "2016" or year == "2016APV":
            golden_json_path = topcoffea_path("data/goldenJsons/Cert_271036-284044_13TeV_Legacy2016_Collisions16_JSON
        elif year == "2017":
            golden_json_path = topcoffea_path("data/goldenJsons/Cert_294927-306462_13TeV_UL2017_Collisions17_GoldenJS
        elif year == "2018":
            golden_json_path = topcoffea_path("data/goldenJsons/Cert_314472-325175_13TeV_Legacy2018_Collisions18_JSON
        else:
            raise ValueError(f"Error: Unknown year \"{year}\".")
        lumi_mask = LumiMask(golden_json_path)(events.run,events.luminosityBlock)

        ######### EFT coefficients ##########
```

Analysis code based on coffea

Physics results



CS results



CS results lead to improvements in analysis code and the underlying packages we depend on