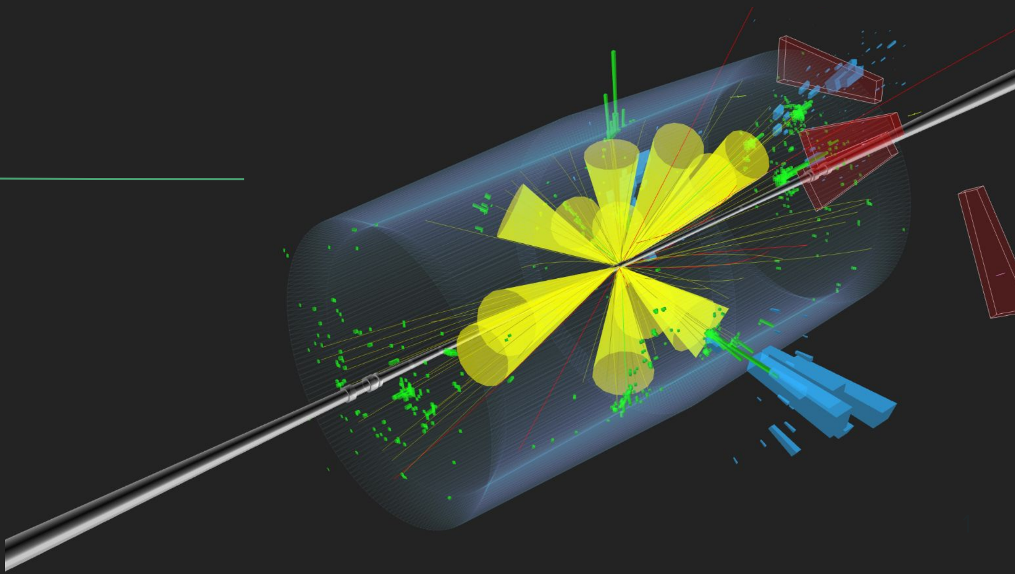
Searching for Extreme Events in Multi-lepton Data from the LHC

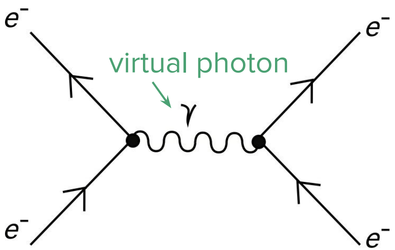
Xinyue Wu (University of Rochester)

Mentors:
Ben Tovar, Kelci Mohrman
(University of Notre Dame)

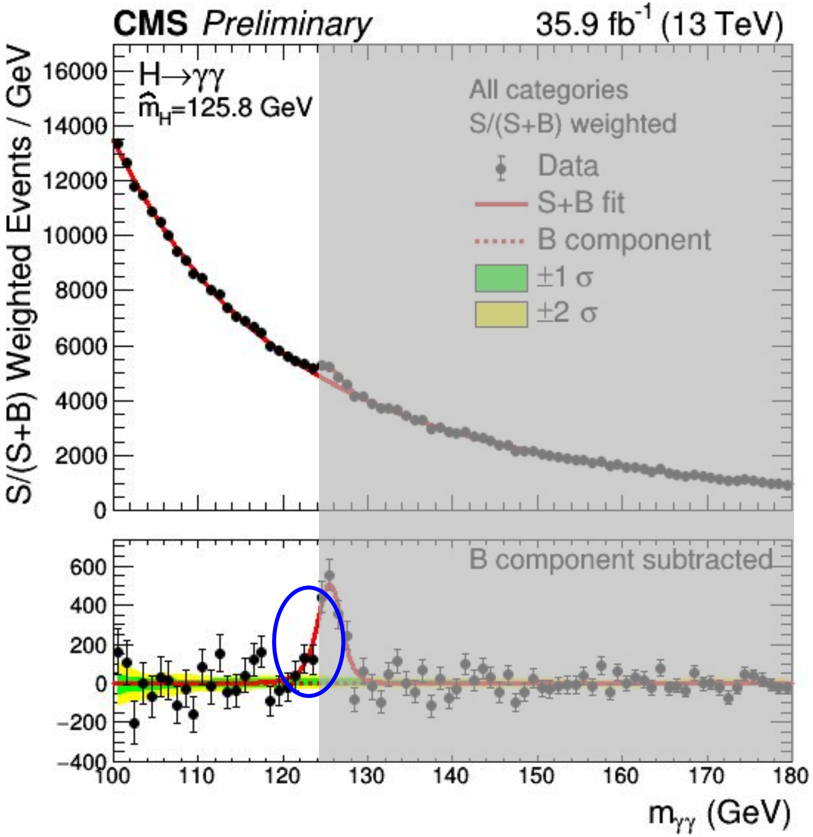


Motivation

- Despite its many successes, the Standard Model (SM) is still incomplete
- Since there is no new particles discovered so far, particles beyond SM might be **off-shell** at the LHC
  - e.g.



  - Discovery of the Higgs boson by bump hunting (right)
- **Effective Field Theory (EFT)** is an indirect approach of probing higher mass scales, which extends the discovery reach of LHC



Introduction

- EFT: - an expansion of SM with higher dimensional operators

  - a framework that allows us to model a wide variety of interactions

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 study 26WCs in the dim 6 terms

- Associated top production
  - Relatively rare, heavy particles (top quarks) involved
- Multi-lepton data
  - Top production events with multiple charged leptons in the final state
  - Clean signatures, efficient triggers, few backgrounds

CMS-TOP-19-001, arXiv:2012.04120

Methodology

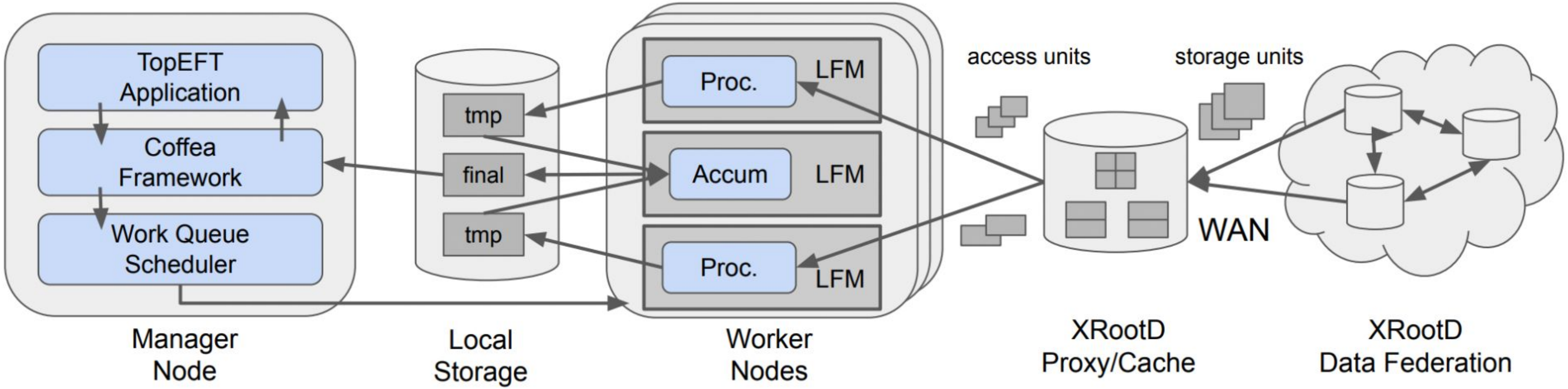
Overview

- Take data from root files as input
- Write a processor (extreme_events.py) and an accumulator to store selected data into dataframes as output
- Run the processor at scale starting from one dataset to the full dataset
- Compare prediction with observation
- Visualize interesting events

System Architecture

- Components:
    - **Coffea**: a framework designed for HEP experiment analysis using the scientific python ecosystem
    - **TopCoffea**: an application designed for top quark analysis using coffea
    - **Work Queue**: an execution system for creating and managing scalable manager-worker style programs
- The **extreme events processor** is built based on a TopCoffea processor (topEFT) to generate multiple pandas dataframes of different groups of selected events
- A **dataframe accumulator** is built for this new type of output

System Architecture



B. Tovar, et al., topEFT-IPDPS-2022

Event Searching

- The full Run 2 data is given as input to the extreme events processor
- 3927 events in total pass the selection
- Events are further filtered by interesting characteristics:
  - Lepton, jet multiplicity
  - S_T, H_T
  - Invariant mass
  - Leading leptons, jets p_T
- Multiple dataframes with top event information (event, run, luminosity block, etc.), event quantities (number of leptons and jets, S_T, etc.), and object level information (p_T, η of leptons or jets) get accumulated to the output

EFT Study

- Data samples
  - Number of extreme events is read from dataframe directly
- Monte Carlo samples
  - Another topcoffea processor (topEFT) generates coffea histograms as output
  - Number of corresponding extreme events is obtained from the bins
- Data vs. MC
  - yield = (cross section) * (luminosity) * acceptance * efficiency
  - acceptance * efficiency = sum(weights of events that pass) / sum(weights of all events)
  - Plug in 26 WCs at approximately expected $\pm 2\sigma$ limits one at a time
  - **Compare the modified yield (prediction) with the observation from data to qualitatively examine the effect of EFT**

Visualization

- Write a processor (find_file.py) to match the **skimmed NDCMS file** with the **NanoAOD file**
- Use Dasgoclient to find the parent **MiniAOD file** of the NanoAOD files
- Use Dasgoclient again to find the **AOD file** from the MiniAOD file
- Extract the **event** of interest by edmCopyPickMerge
- Generate an **IG file** with iSpy
- **Display** the event in the browser

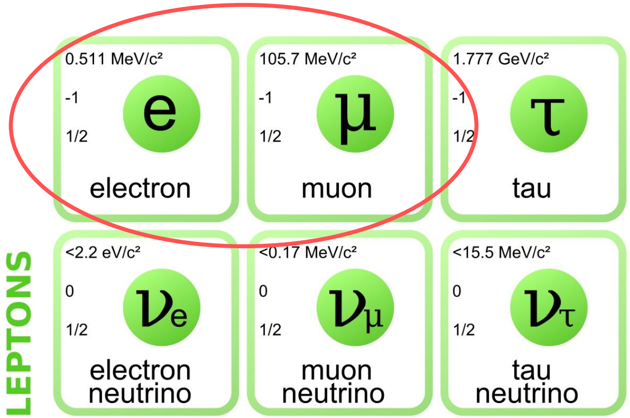
Run into a problem with **proxy** here. A general solution to the tedious process could be future work.

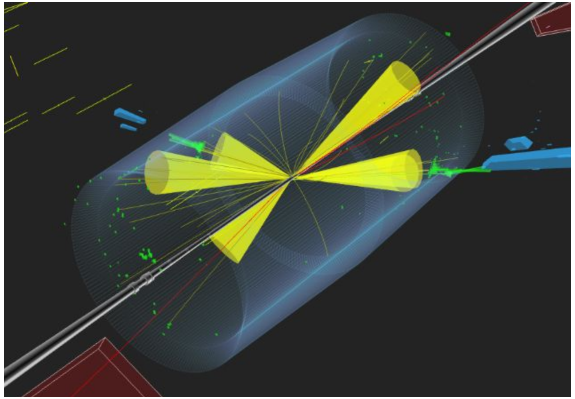
T. McCauley, 2017 J. Phys.: Conf. Ser. 898 072030

Event Analysis

Background

- Charged leptons
  - Only electrons and muons are included in the data
- Jet
  - A collimated spray of hadrons and other particles produced by a quark or a gluon
  - Fly in roughly the same direction (i.e. distributed in a cone)
- Invariant mass
  - Rest mass, independent of the reference frame
  - Calculated from energy and momentum



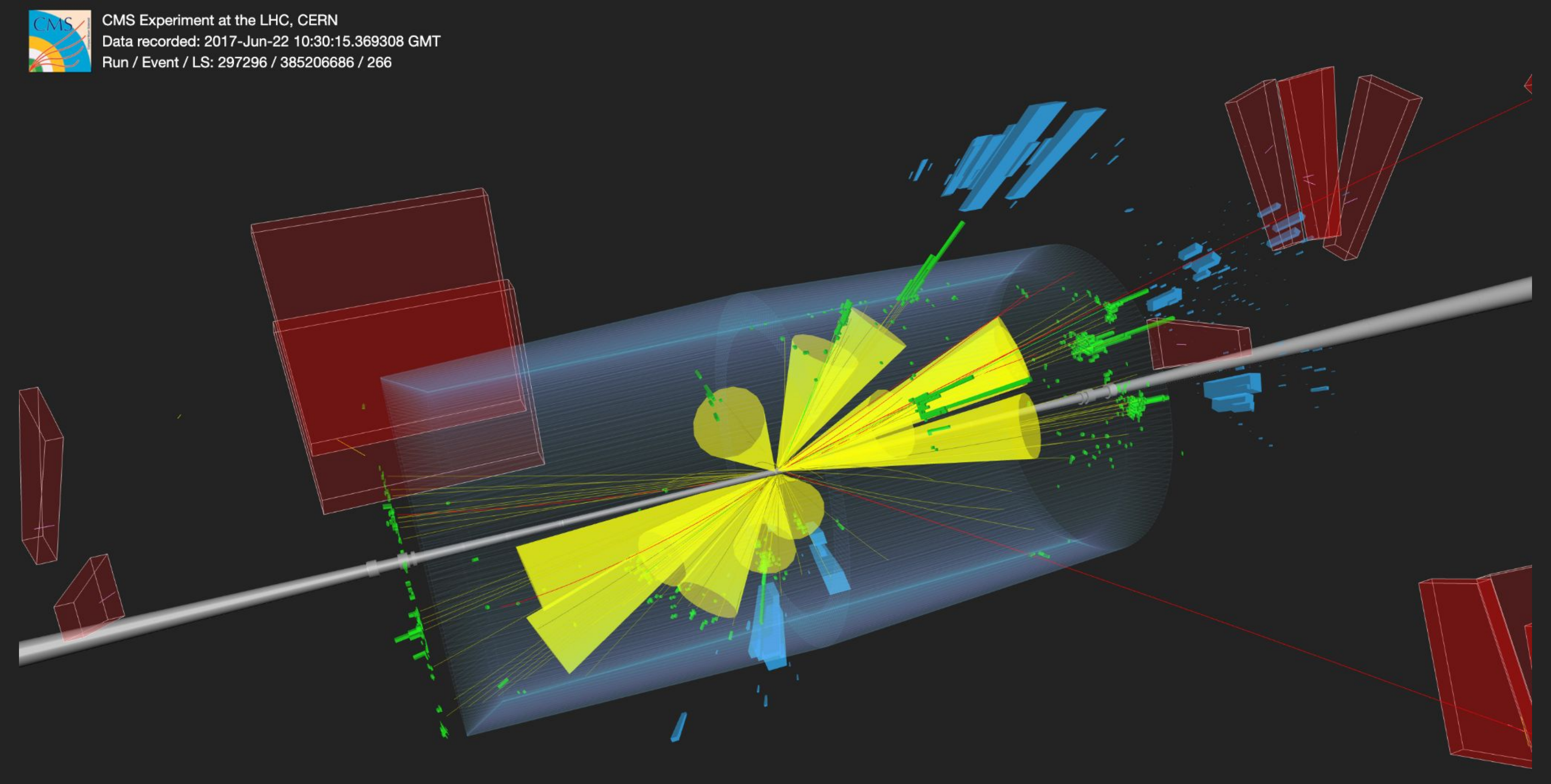


Example of Event Visualization: 1TeV Jet p_T

Jet Multiplicity

	dataset	year	event	njets	nleps
0	DoubleMuon	2017	385206686	12	2
1	MuonEG	2017	975066055	11	2
2	EGamma	2018	210307202	11	2
3	EGamma	2018	9852065	11	2
4	DoubleMuon	2016	2954912008	10	2
5	DoubleMuon	2017	541866020	10	2
6	DoubleEG	2017	17367514	10	2
7	DoubleMuon	2018	1299057415	10	2

- Observation from real data:
  - 1 event with 12 jets (p_T of all jets > 30GeV)
  - 8 events with ≥ 10 jets
- Yield of ≥ 10 jets events from MC samples:
  - SM predicts ~2 but observe 8
  - Some WCs have a fairly large impact on the yield



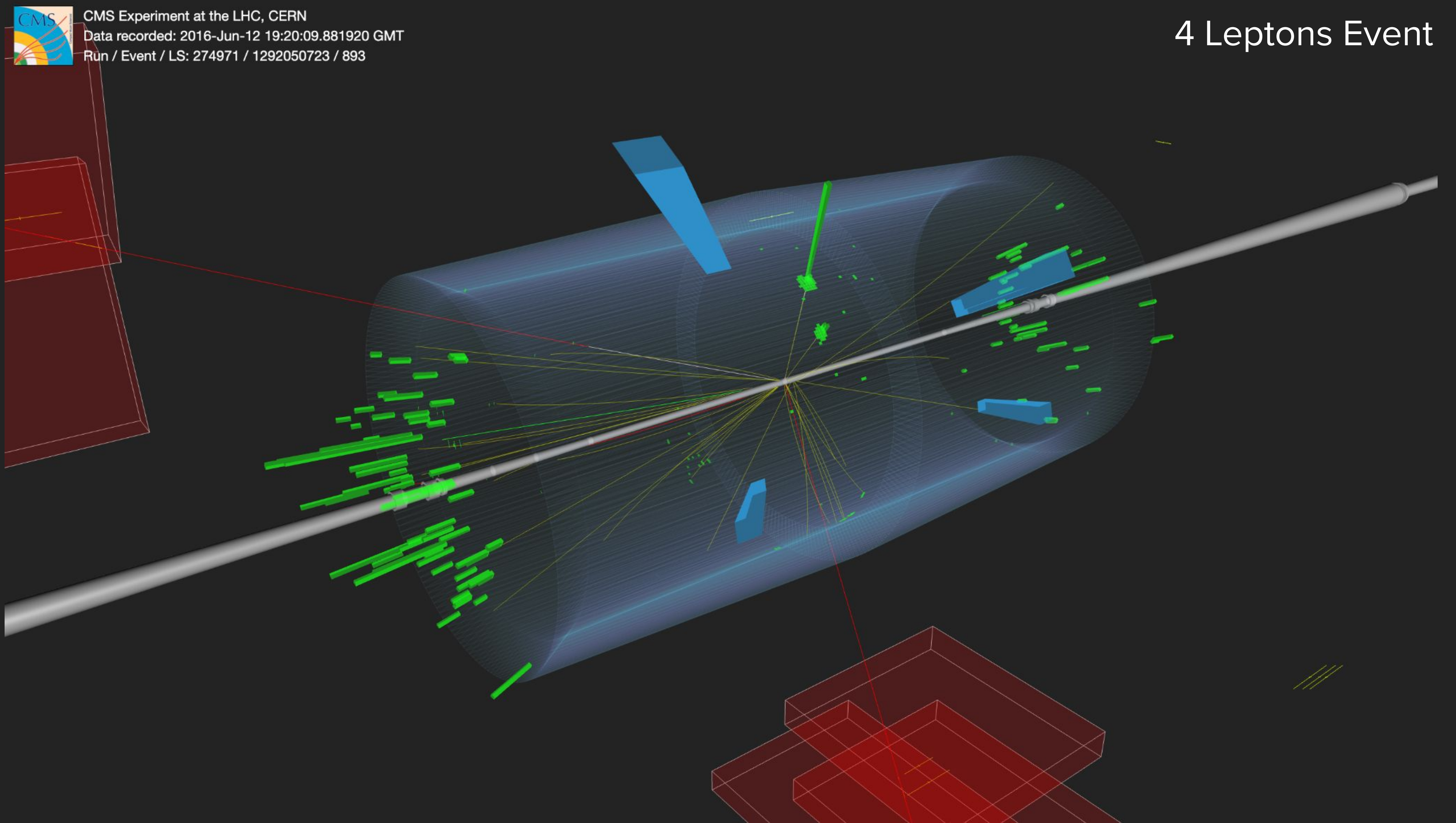
12 Jets Event

Lepton Multiplicity

SM	0.067
cQei	0.068
cQlMi	0.068
cQq13	0.090
cQq83	0.076
cpt	0.116
ctG	0.248
ctZ	0.104
ctlSi	0.068
allWCs	0.515

- Observation:
  - 50 events with 4 leptons
  - No event with more than 4
- Yield of > 4 leptons events: (left)
  - SM prediction agrees with observation
  - WCs have no large impact on yield

4 Leptons Event



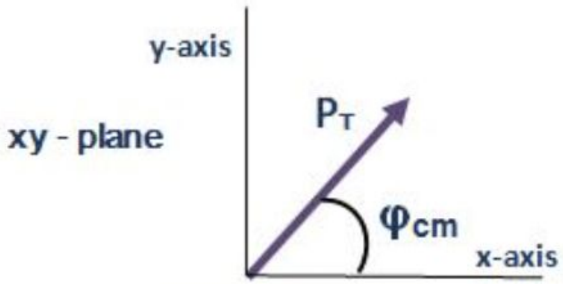
Invariant Mass

SM	49
cQei	77
cQlMi	76
cQq13	127
cQq83	93
cpt	109
ctG	118
ctZ	99
ctlSi	75
allWCs	974

- Observation:
  - 84 with > 2TeV
  - 7 with > 3TeV
  - 4 with > 4TeV
  - 1 with 5.1TeV
- Yield of > 2TeV: (left)
  - MC underpredicting
  - Invariant mass distribution is fairly sensitive to WCs

Background

- Transverse momentum p_T
  - Momentum in the transverse plane (perpendicular to the beamline)
  - Good indication of how energetic the collision was
- S_T
  - Scalar sum of the p_T of all jets and leptons
- H_T
  - Scalar sum of the p_T of all jets



S_T and H_T

- S_T: observe 2 events with > 3TeV, 9 events with > 2TeV
- H_T: observe 1 event with > 3TeV, 1 event with > 2TeV
- Highest S_T and H_T event:

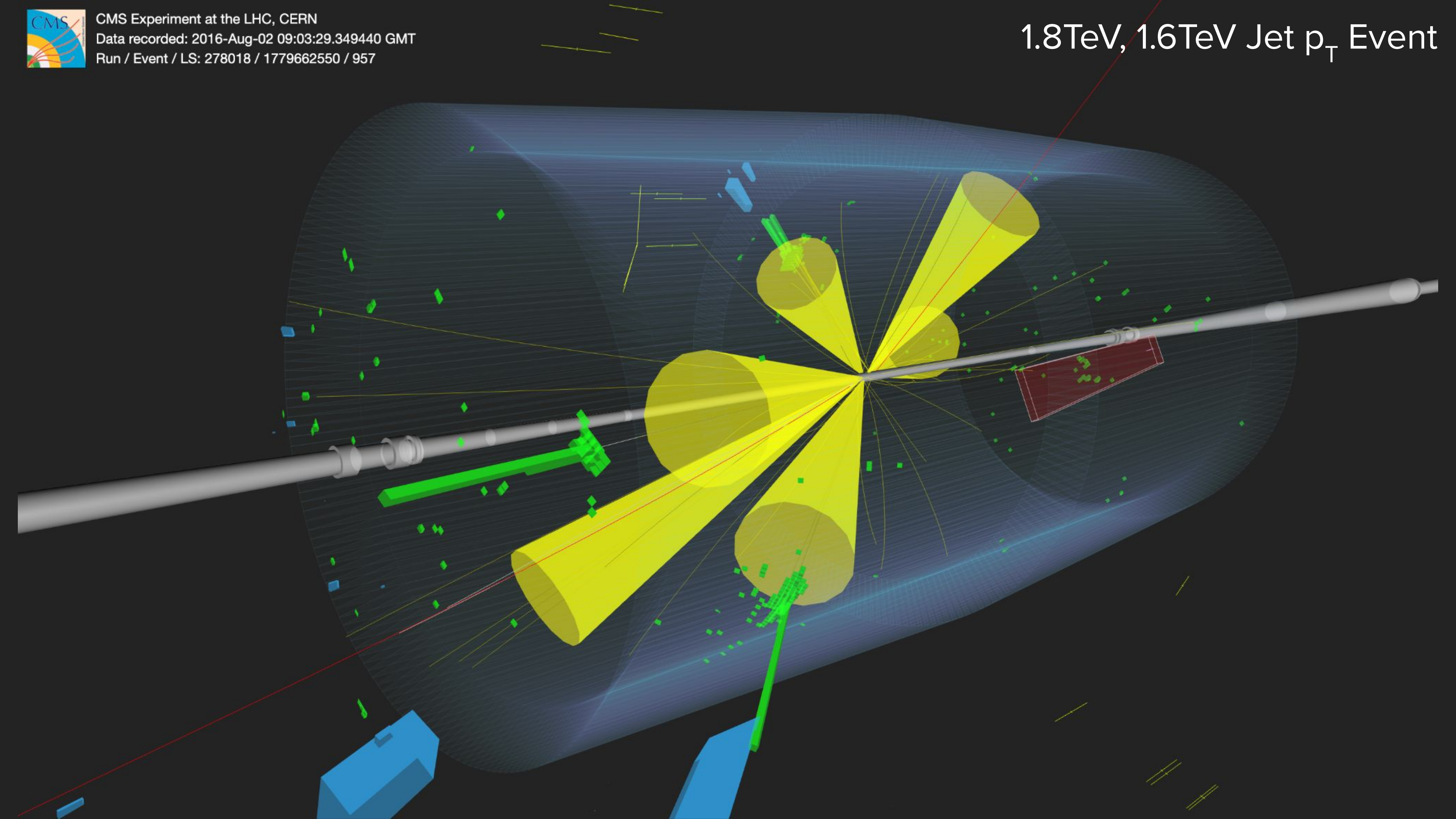
	dataset	year	event	nleps	njets	S_T	H_T
0	SingleElectron	2016APV	1779662550	2	4	3730.113898	3476.75000

Leading Lepton and Jet p_T

- Leading lepton p_T: observe 1 event with > 800GeV, 1 event with > 600GeV, 12 events with >500GeV
- Leading jet p_T: observe 6 events with > 1TeV
  - The two highest p_T jets (1.8TeV, 1.6TeV) are from the same event (also the highest S_T&H_T event)

	pt_j_max	event	S_T	category
0	1798.0	1779662550	3730.113898	2lss_p
1	1161.0	2168723274	2572.719021	3l_onZ_1b
2	1132.0	526956127	1788.570972	2lss_m
3	1109.0	1082289551	2152.337328	3l_onZ_1b
4	1107.0	294824687	2506.260368	2lss_p
5	1037.0	806177876	2088.711299	2lss_p

1.8TeV, 1.6TeV Jet p_T Event



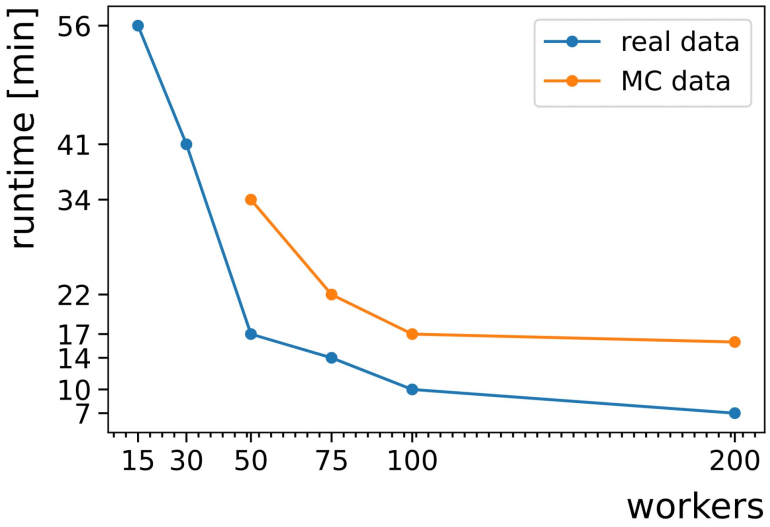
Summary

- EFT parameters have fairly large effects on the prediction
- The impacts are larger at higher energies but less on distributions like number of leptons
- Distributions sensitive to EFT effects:
    - Jet multiplicity
    - Invariant mass
    - Leading leptons, jets p_T
- Several events with high energies could be studied further
- Statistical analysis will quantitatively narrow down the intervals of WCs

Computation Performance

Overview

- Run at scale with different numbers of workers using condor
- One task per core, at peak 800 tasks running at the same time



- Resources usage:

Cores	4
Memory (GB)	18
Disk (GB)	36

- Approximate data size:

	Actual Size (GB)	WQ Transfer (GB)
real data	560	150
MC data	550	266

Bottleneck

- For MC data, the average processing wall time with 200 workers (∼104s) is about twice of which with 100 (and other numbers of) workers (∼57s)
    - i.e. If we double the number of concurrent tasks, tasks run much slower
- For real data, this feature is not obvious due to relatively small data size
- Each task should run independently of each other and the only common resource is getting files from **xrootd**
- Thus **IO** is the bottleneck which should be fixed in the future work of CMS

Conclusion

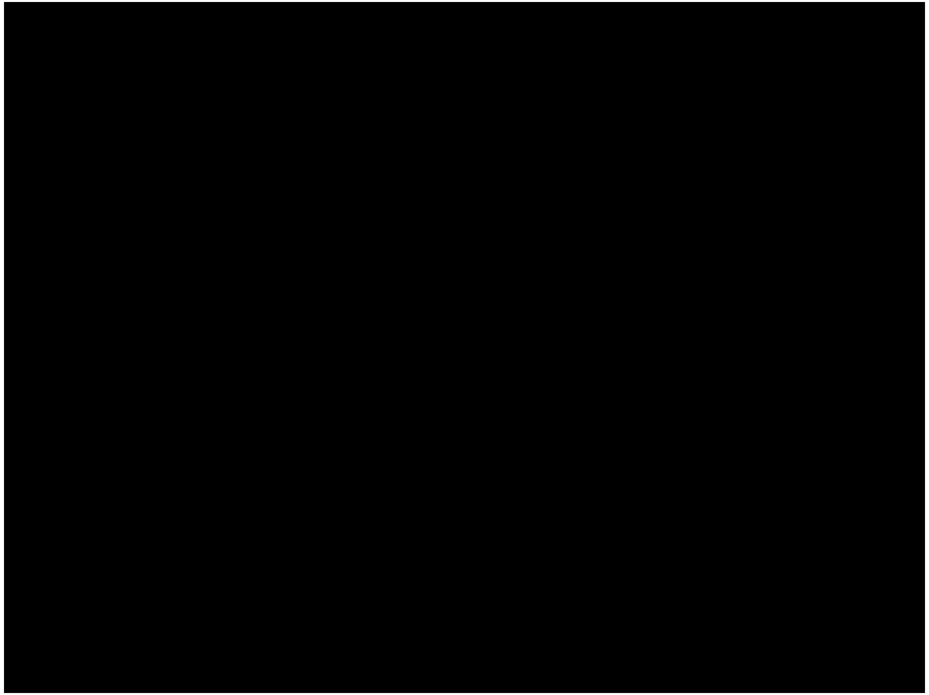
- ☑ Two processors and one accumulator are written to search and visualize extreme events from the multi-lepton data
- ☑ Several events with interesting characteristics have been found
- ☑ EFT parameters have fairly large effects on the prediction of high energy events
- ☑ While the processor can be ran at scale, there is still a bottleneck to be fixed in the future work

What I learned this summer…

- ➢ Use of Pandas dataframe, Awkward array, Coffea processor, Root files
- ➢ Use HTCondor to launch jobs for applications that needs hundreds of cores
- ➢ Use dasgoclient and iSpy to track files and visualize events
- ➢ Study logs from Work Queue to identify bottlenecks of applications that run at scale
- ➢ EFT, virtual particles, HEP experiment quantities (p_T, η, invariant mass…)
- ➢ How an experimental HEP research is planned and conducted
- ➢ Grad school experience in experimental HEP
- ➢ A variety of research in HEP (CMS related research, phenomenology, neutrino physics…)
- ➢ The list goes on…

Backup

1.8TeV Jet pT Event Animation



MC tr_log (mean)

workers	category	memory	disk	bytes_received	bytes_sent	wall_time
50	accumulating	170.319231	318.065385	0.000000	0.0	4.522875
	preprocessing	117.028176	317.000000	0.000000	0.0	3.313709
	processing	606.008680	317.000000	42.762257	0.0	58.471657
75	accumulating	165.590038	317.712644	0.000000	0.0	4.047876
	preprocessing	117.024480	317.000000	0.000000	0.0	3.846523
	processing	605.210095	317.000000	42.769169	0.0	57.045569
100	accumulating	167.505747	317.762452	0.000000	0.0	4.232760
	preprocessing	117.029099	317.000000	0.000000	0.0	4.085918
	processing	605.488185	317.000000	42.765954	0.0	56.255694
200	accumulating	167.019157	317.842912	0.000000	0.0	4.814565
	preprocessing	117.021709	317.000000	0.000000	0.0	5.090015
	processing	607.398810	317.000000	42.765311	0.0	104.775493

WCs

```
self.wc_ranges_differential = {
           'cQQ1' : (-4.0,4.0),   'cQei' : (-4.0,4.0),
           'cQl3i': (-5.5,5.5),   'cQlMi': (-4.0,4.0),
           'cQq11': (-0.7,0.7),   'cQq13': (-0.35,0.35),
           'cQq81': (-1.7,1.5),   'cQq83': (-0.6,0.6),
           'cQt1' : (-4.0,4.0),   'cQt8' : (-8.0,8.0),
           'cbW'  : (-3.0,3.0),   'cpQ3' : (-4.0,4.0),
           'cpQM' : (-10.0,17.0), 'cpt'  : (-15.0,15.0),
           'cptb' : (-9.0,9.0),   'ctG'  : (-0.8,0.8),
           'ctW'  : (-1.5,1.5),   'ctZ'  : (-2.0,2.0),
           'ctei' : (-4.0,4.0),   'ctlSi': (-5.0,5.0),
           'ctlTi': (-0.9,0.9),   'ctli' : (-4.0,4.0),
           'ctp'  : (-11.0,35.0), 'ctq1' : (-0.6,0.6),
           'ctq8' : (-1.4,1.4),   'ctt1' : (-2.1,2.1),
       }
```

nJets ≥ 10 yield

SM	2.176771572665743	cQQ1	4.881141971491163
cQei	2.316530515824015	cQl3i	2.203684563212246
cQlMi	2.3064875089632952	cQq11	2.9495139578430014
cQq13	2.5831742429276017	cQq81	2.889410500247003
cQq83	2.3597289071715624	cQt1	6.673692389925052
cQt8	5.910526963958538	cbW	2.351264532604
cpQ3	2.2982535645396944	cpQM	2.5755497590796366
cpt	4.744387606439089	cptb	2.2439000020882514
ctG	4.506612404414624	ctW	2.9006076895865887
ctZ	3.062872238878099	ctei	2.293294058528366
ctlSi	2.2783078729477495	ctlTi	2.3456824136857266
ctli	2.317186762439446	ctp	3.9970527590678064
ctq1	2.8488545681685427	ctq8	3.056552010615471
ctt1	5.192903940205572	allWCs	27.84587700789577

nleps ≥ 4 yield

SM	0.0666218998762968	cQQ1	0.0666414566756765
cQei	0.06843203140234011	cQl3i	0.06662189578701458
cQlMi	0.06831440223154565	cQq11	0.07906971051456381
cQq13	0.08986723059012523	cQq81	0.08772927580623205
cQq83	0.07631948294506452	cQt1	0.06727199327433016
cQt8	0.06740732203787383	cbW	0.1723208456229724
cpQ3	0.07704494123266814	cpQM	0.10814914671370131
cpt	0.116339294362893	cptb	0.10462039945794104
ctG	0.2484872020165991	ctW	0.11285359394730379
ctZ	0.10361924099054434	ctei	0.06838301402783485
ctlSi	0.06760184185412475	ctlTi	0.06830768654078799
ctli	0.07070707303451959	ctp	0.09853535445419992
ctq1	0.08091781814181444	ctq8	0.09470733387068539
ctt1	0.06659280318863631	allWCs	0.515225482312428

invmass

0	5119.891508
1	4661.199477
2	4178.783497
3	4097.395059
4	3540.074779
5	3255.766280
6	3194.158088
7	2951.814710
8	2943.704305
9	2847.873707
10	2827.410250

invmass ≥ 2TeV yield

SM	49.29994475853063	cQQ1	66.57442357155549
cQei	77.16230254165818	cQl3i	70.73855823718007
cQlMi	75.80395674243763	cQq11	125.69366737299683
cQq13	127.06322525410205	cQq81	114.82337129557996
cQq83	93.21596287149254	cQt1	75.50215178177055
cQt8	72.28046888454307	cbW	102.0082326767067
cpQ3	68.31866547295127	cpQM	59.41795277221684
cpt	108.98492919853041	cptb	62.642181580278255
ctG	118.03341848737091	ctW	114.7304577984265
ctZ	99.32774593099475	ctei	74.54814481878303
ctlSi	75.28985325815539	ctlTi	84.50868007284119
ctli	76.18813125445114	ctp	57.1834405730811
ctq1	102.5319677405466	ctq8	113.93560088073654
ctt1	68.42523352052429	allWCs	974.5608822599422

S_T (≥ 2TeV)

	dataset	year	event	nleps	njets	S_T	H_T
0	SingleElectron	2016APV	1779662550	2	4	3730.113898	3476.75000
1	DoubleEG	2016APV	1201516424	3	5	3170.825983	1867.68750
2	DoubleMuon	2018	2168723274	3	2	2572.719021	1615.00000
3	MuonEG	2017	294824687	2	5	2506.260368	2424.75000
4	MuonEG	2016APV	3141982364	3	5	2155.942602	1892.06250
5	DoubleEG	2017	1082289551	3	3	2152.337328	1825.03125
6	DoubleMuon	2018	916826155	3	4	2106.409133	1862.31250
7	DoubleMuon	2016	806177876	2	6	2088.711299	1966.40625
8	DoubleMuon	2017	2022679122	3	4	2015.399828	1706.65625

H_T (≥ 1.6TeV)

	dataset	year	event	nleps	njets	S_T	H_T
0	SingleElectron	2016APV	1779662550	2	4	3730.113898	3476.750000
1	MuonEG	2017	294824687	2	5	2506.260368	2424.750000
2	DoubleMuon	2016	806177876	2	6	2088.711299	1966.406250
3	MuonEG	2016APV	3141982364	3	5	2155.942602	1892.062500
4	DoubleEG	2016APV	1201516424	3	5	3170.825983	1867.687500
5	DoubleMuon	2018	916826155	3	4	2106.409133	1862.312500
6	DoubleEG	2017	1082289551	3	3	2152.337328	1825.031250
7	DoubleMuon	2017	2022679122	3	4	2015.399828	1706.656250
8	MuonEG	2018	710990208	2	6	1856.277738	1629.937500
9	DoubleMuon	2018	2168723274	3	2	2572.719021	1615.000000
10	DoubleMuon	2018	526956127	2	7	1788.570972	1604.937500

pt_l_max (≥ 500GeV)

0	880.983582
1	622.755850
2	592.126648
3	583.641733
4	583.272522
5	552.726990
6	539.088025
7	522.723145
8	522.189606
9	515.303772
10	513.344109
11	501.016697

pt_j_max ($\geq$ 800GeV)

0	1798.0
1	1161.0
2	1132.0
3	1109.0
4	1107.0
5	1037.0
6	963.0
7	893.5
8	871.0
9	853.0