Studies for the measurement of the ZZ -> 2l2v cross section in ATLAS and for the search for aQGCs

S. Tzamarias, K. Kordas, ***A. Tsiamis***, D. Fakoudis: (AUTH)

D. Sampsonidou: (SJTU,TDLI),

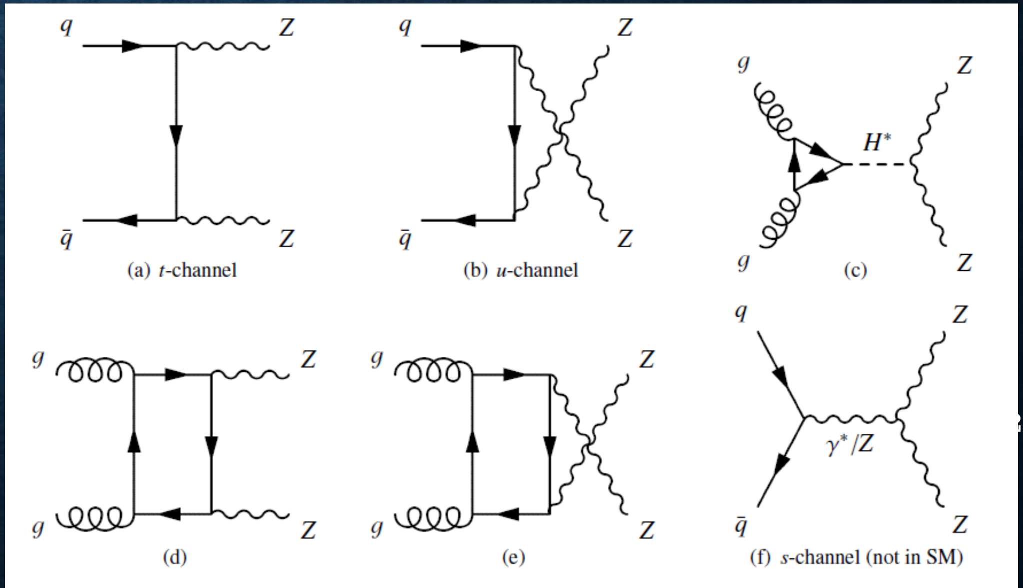
ATLAS EXPERIMENT

ARISTOTLE UNIVERSITY OF THESSALONIKI

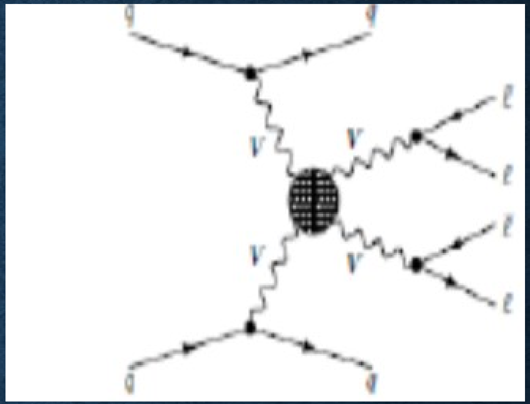
INTRODUCTION

- Production of boson pairs provide an opportunity to study the electroweak sector by looking for deviations from predicted total and differential cross-sections, that could be an indication of new physics.
- ZZ produced via :
  - quark-antiquark ($q\bar{q}$) annihilation
  - gluon-gluon fusion (gg) via a quark loop.
  - Higgs boson decay.
- 2 dominant decay modes
  - 4 leptons
    - Cleaner Signal
    - Small Branching Ratio
  - 2 leptons 2 neutrinos
    - Missing energy due neutrinos -> Not so clean Signal
    - Bigger Branching Ratio.

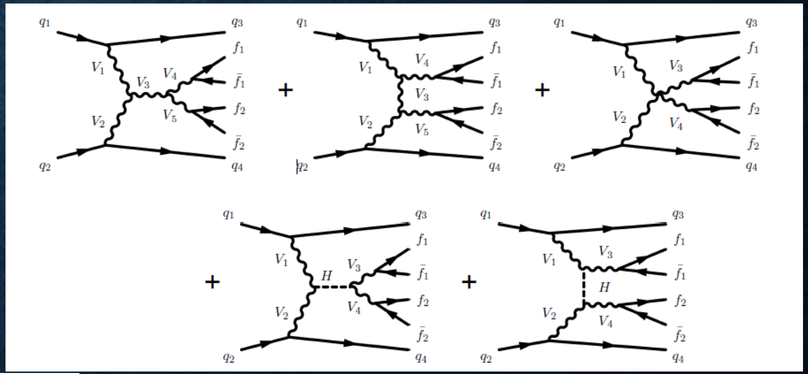
- In high $\sqrt{\hat{s}}$, results from ZZ->2l2v could improve the sensitivity to new physics, with respect to ZZ->4l alone.
- Our analysis focus on 2 leptons and 2 neutrinos decay channel.
- ZZ production also useful for:
  - High order calculations of perturbative Quantum chromodynamics
  - pp->ZZ production is a background to the SM Higgs boson process and to many searches for physics beyond the SM.



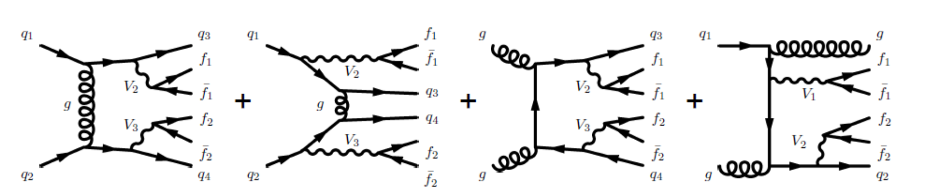
VECTOR BOSON SCATTERING (VBS)



- Self interactions between electroweak vector bosons:
  - Only EWK vertices involved.
  - 3 or 4 gauge bosons at a single vertex (triple and quartic gauge couplings, respectively).
- VBS is initiated by quarks from the colliding protons; both quarks radiate vector bosons (V=W, Z) which then interact.
  - The vector bosons decay in fermions (2l2v,llll)
  - two high energy jets, in a forward-backward topology, with two vector bosons in between (“VBS topology”).
  - Events with QCD vertices can also result in the VBS topology.
- Pure EWK production :
  - Probes the mechanism of electroweak symmetry breaking (EWSB) in the Standard Model (SM).



EWK vertices



QCD vertices

DOMINANT BACKGROUNDS IN *ZZ->2l2v* CHANNEL

- WZ - > 3l
    - Consists of events where one of the 3 leptons is not detected. Genuine E_T^{miss} from W neutrino
    - Large contribution in the Signal Region.
- WW/$t\bar{t}$/Wt/Ztt (non resonant)
    - contribute with 2 leptons from different mother particles. Genuine E_T^{miss} from neutrinos.
    - relative probability at production is ee:μμ:eμ = 1:1:2 .
- Z->l+l- +jets
    - Contribute to SR due to fake E_T^{miss}.
- Other contributions (4l, VVV, W+jets, ttV)
    - Small contributions

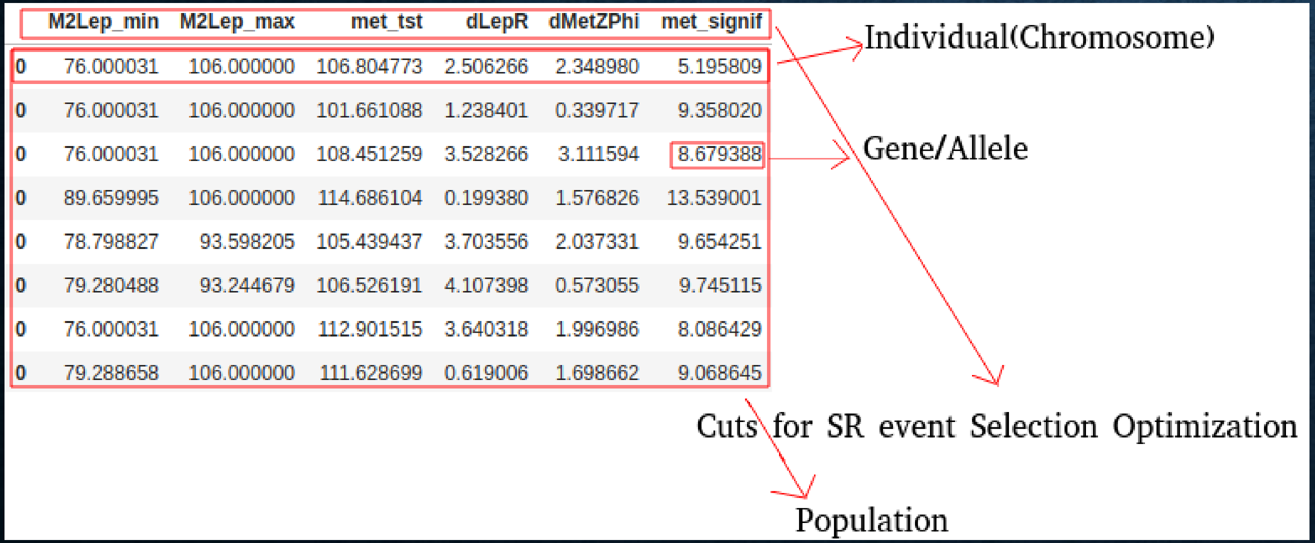
SIGNAL REGION SELECTION

- Goal: Find an optimal set of kinematic requirements (cuts) that maximizes significance.
  - $S = \sqrt{2\left((S+B)\ln\left[1+\frac{S}{B}\right] - S\right)}$
- *Cut and Count method*:
  - Use one variable at a time to define the cut at a certain point where significance starts to drop.
    - This way we ignore any correlations between the variables.
- Grid search : Computationally demanding .
  - Need for an optimization technique.
- Several Classification algorithms, such as Boosted Decision Trees (BDTs) and Artificial Neural Networks have been deployed for separating signal from background.
  - Both methods result in a classifier score. Signal region is defined by a cut on the score(not the original kinematic variables)
- In our current analysis we employ a Genetic Algorithm , for searching the optimal set of square cuts.

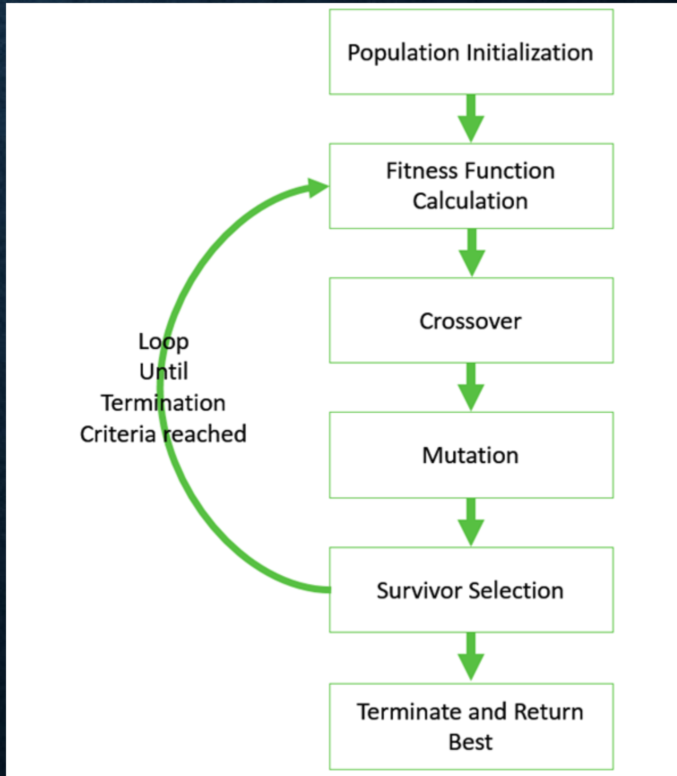
OPTIMIZATION USING GENETIC ALGORITHM

- GA is search heuristic(rather than algorithm) inspired by Darwin's natural selection, based on the concept of *survival of the fittest*.
- Each solution to the problem (set of cut values) are represented as a vector of numbers, which is referred as *individual*.
- Each individual has a score (significance for our case) , indicating how fit an individual is.

- Initialization: Initialize several individuals with random cut values

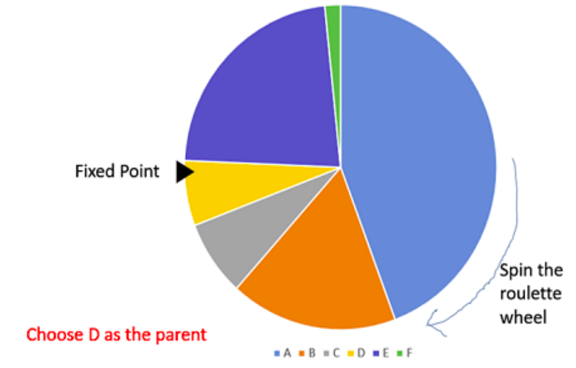


	M2Lep_min	M2Lep_max	met_tst	dLepR	dMetZPhi	met_signif
0	76.000031	106.000000	106.804773	2.506266	2.348980	5.195809
0	76.000031	106.000000	101.661088	1.238401	0.339717	9.358020
0	76.000031	106.000000	108.451259	3.528266	3.111594	8.679388
0	89.659995	106.000000	114.686104	0.199380	1.576826	13.539001
0	78.798827	93.598205	105.439437	3.703556	2.037331	9.654251
0	79.280488	93.244679	106.526191	4.107398	0.573055	9.745115
0	76.000031	106.000000	112.901515	3.640318	1.996986	8.086429
0	79.288658	106.000000	111.628699	0.619006	1.698662	9.068645



- At each generation(loop):
  - We create new individuals (offsprings) by combining 2 individuals in the population. Weather the new individual is produced or not is controlled by a certain probability (*crossover rate*).
  - The selection of the individuals to be combined is proportional to its fitness (significance).
  - *Crossover*-> Combination of the genes of the 2 individuals, to produce a new individual(offspring)

Individual 1		Individual 2		Offspring	
M2Lep_min	82.319510	M2Lep_min	76.000031	M2Lep_min	76.000031
M2Lep_max	89.861942	M2Lep_max	106.000000	M2Lep_max	106.000000
met_tst	189.573846	met_tst	100.125271	met_tst	189.573846
dLepR	2.496918	dLepR	0.539875	dLepR	2.496918
dMetZPhi	0.173729	dMetZPhi	2.269436	dMetZPhi	2.269436
met_signif	7.839628	met_signif	9.117347	met_signif	7.839628



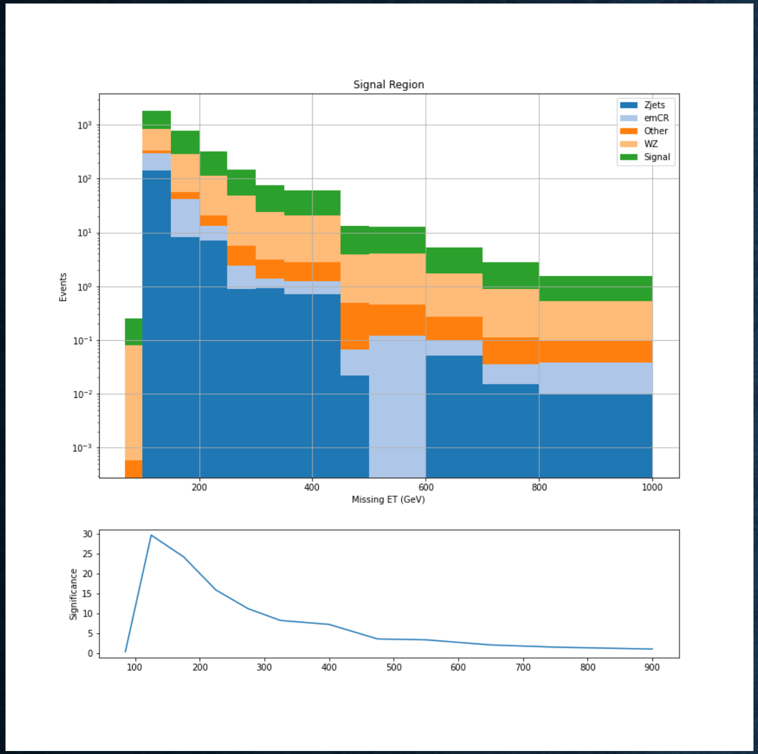
Chromosome	Fitness Value
A	8.2
B	3.2
C	1.4
D	1.2
E	4.2
F	0.3

Roulette Selection

  - Replace least fitted individuals with the offsprings
  - Apply *mutation* . Each individual in the population mutates with a certain *mutation rate*. Each gene mutates with a certain *mutation probability*.

Individual		Mutated Individual	
M2Lep_min	82.319510	M2Lep_min	76.000031
M2Lep_max	89.861942	M2Lep_max	89.861942
met_tst	189.573846	met_tst	189.573846
dLepR	2.496918	dLepR	2.496918
dMetZPhi	0.173729	dMetZPhi	0.173729
met_signif	7.839628	met_signif	9.922509

- Repeat for several generations
- The individual with the best score among all generations is returned as the solution.

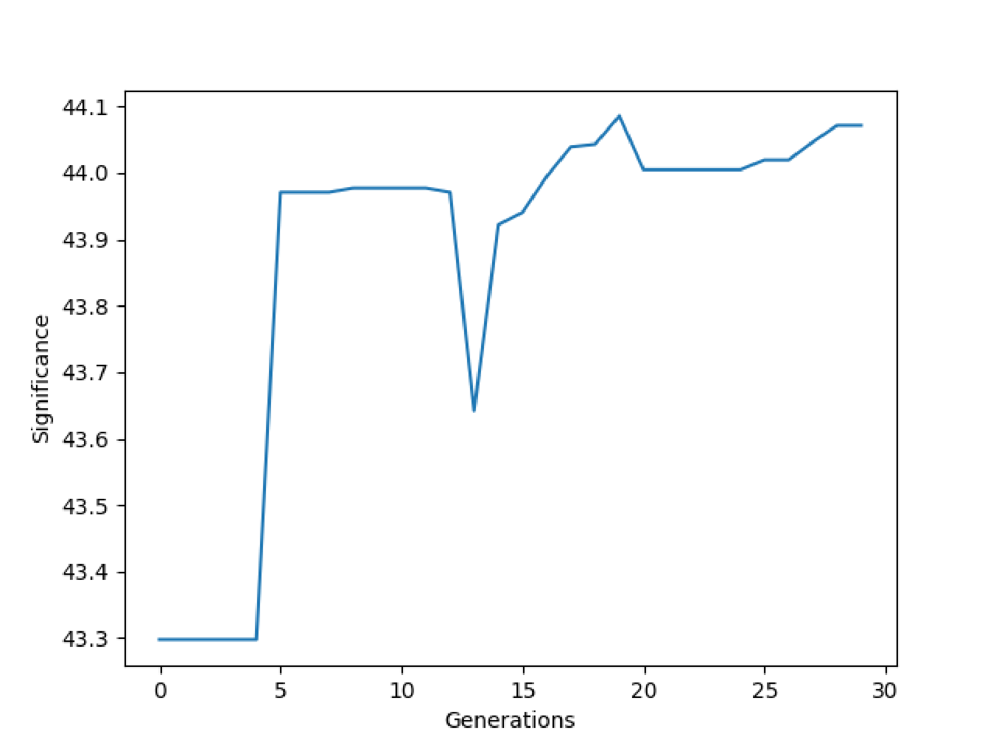


SIGNAL REGION OPTIMIZATION RESULTS (INCLUSIVE)

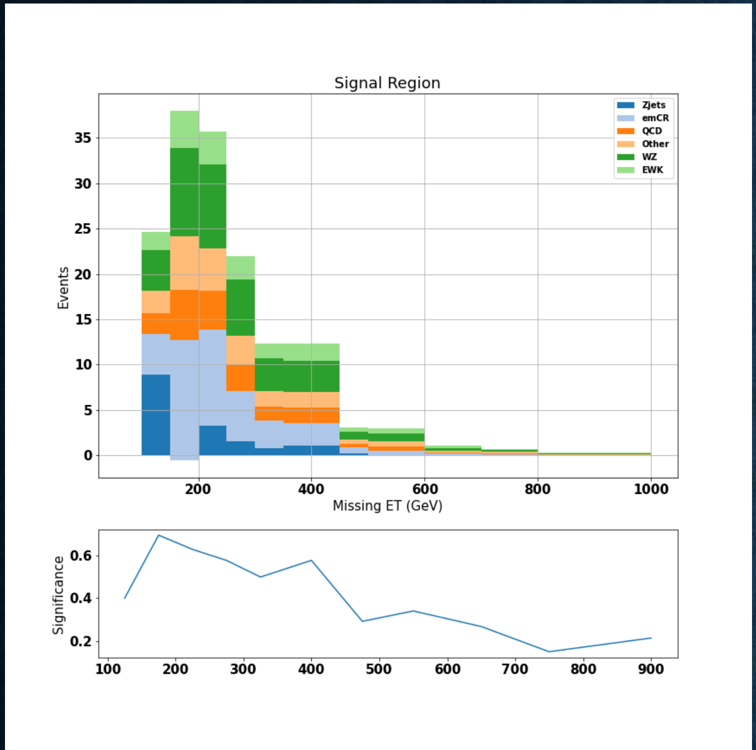
GA hyperparameters				
Generations	population size	crossover Rate	mutation rate	mutation probability
30	100	0.6	0.5	0.4

Inclusive ZZ production: Signal Region
81.71< M_{ll} <105.62 GeV
ΔR_{ll} <1.85
E_T^{miss}>96.2
$\Delta\varphi(E_T^{miss}$,ll)>2.14
$\frac{E_T^{miss}}{H_T}$ >0.54
$E_T^{miss}{}_{significance}$>9.92
b jet veto

	all	ee	mm
signal	1923.2±15.87	911.37±10.52	1011.83±11.89
WZ	917.01±8.33	439.78±6.27	477.24±5.48
Zjets	159.41±19.47	83.2±15.07	76.21±12.32
Non-Resonant	201.63±3.42	101.93±2.46	99.7±2.38
Other	59.06±2.27	27.97±1.79	31.09±1.39
Signal	1923.2±15.87	911.37±10.52	1011.83±10.52
Background	1337.12±21.57	652.88±16.61	684.24±13.77
Significance	44.33	30.18	32.49



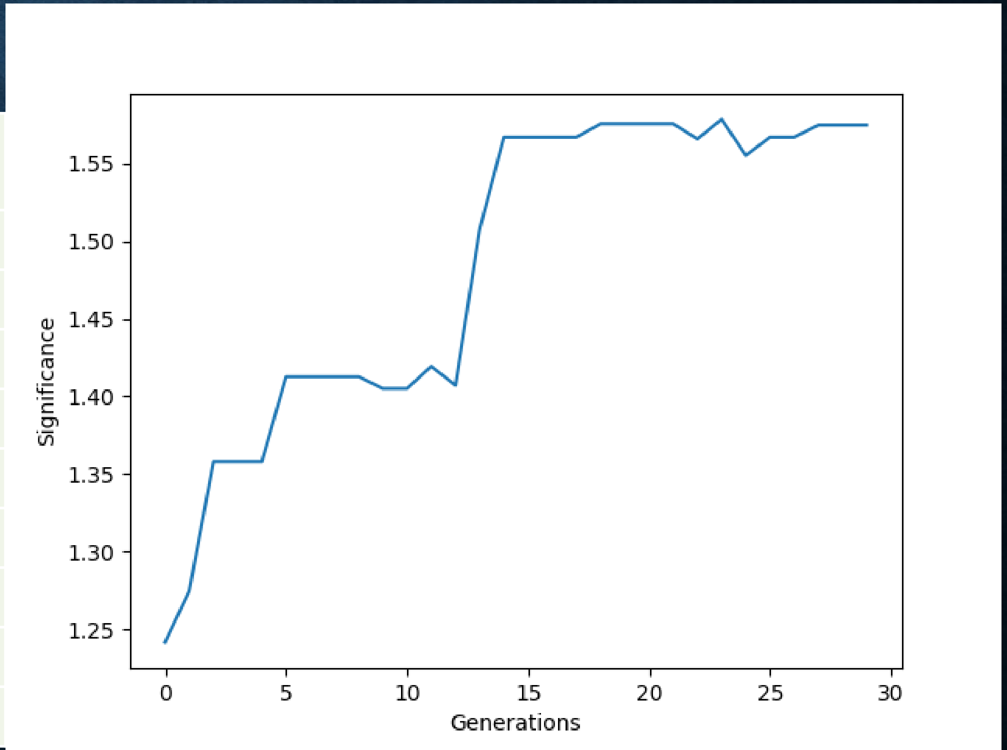
SIGNAL REGION OPTIMIZATION RESULTS (VBS)



GA hyperparameters				
Generations	population size	crossover Rate	mutation rate	mutation probability
30	100	0.6	0.5	0.4

ZZ+2jets production: VBS EWK Region
84.72< M_{ll} <96.07 GeV
M_{jj} >583.77
$\Delta\eta_{jj}$>2.05
$p_T(leading\ jet)$>47.00
$p_T(sub\_leading\ jet)$>45.08
$E_T^{miss}{}_{significance}$ > 8.82
E_T^{miss}>106.34
b jet veto

	all	ee	mm
EWK	17.39±0.12	8.62±0.09	8.77±0.09
QCD	19.57±0.59	9.74±0.44	9.83±0.4
WZ	39.06±0.56	19.03±0.45	20.03±0.34
Zjets	15.3±7.16	2.49±5.07	12.81±5.06
Non-resonant	40.82±1.35	19.6±0.95	21.21±0.96
Other	21.12±0.61	10.5±0.45	10.62±0.41
Signal	17.39±0.12	8.62±0.09	8.77±0.09
Background	116.3±7.36	51.63±5.22	64.67±5.19
Significance	1.57	1.17	1.07



BACKGROUND ESTIMATE BY A SIMULTANEOUS FIT

- To calculate background contributions from different sources :
  - We define Control Regions, which in principle are mostly populated by one source of background, but each region is a different mixture of the various event types(ZZ->2l2v,WW,WZ,Zjets etc.).
    - 3lCR : Consists of 3 leptons. Mostly populated by WZ events
    - emCR : Consists of different flavor leptons. Mostly populated by non-resonant events (WW,tt~, etc.).
    - Zjets CR: Inverse ${E_T^{miss}}_{significance}$ cut. Mostly populated by Zjets and non-resonant events.
  - Define scaling factors (for signal and Background) , to correct the MC predictions to match the observations at the various regions.
  - Simultaneous fit in all regions (Signal and Control).
  - We do not use just the event counts, but distributions of kinematic variables (histograms).
- For demonstration purposes, we perform a fit, using the expectations from the MC contributions as Data ("Asimov dataset")
  - This way we can study the expected uncertainty on the estimation of the scaling factors.
- Consider 2 sources of theory uncertainties:
  - PDF uncertainties
  - Renormalization and Factorization Scale uncertainties

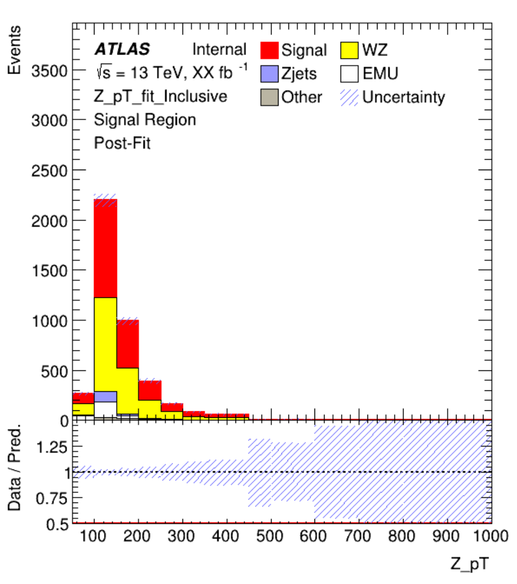
DEFINITION OF SIGNAL AND CONTROL REGIONS

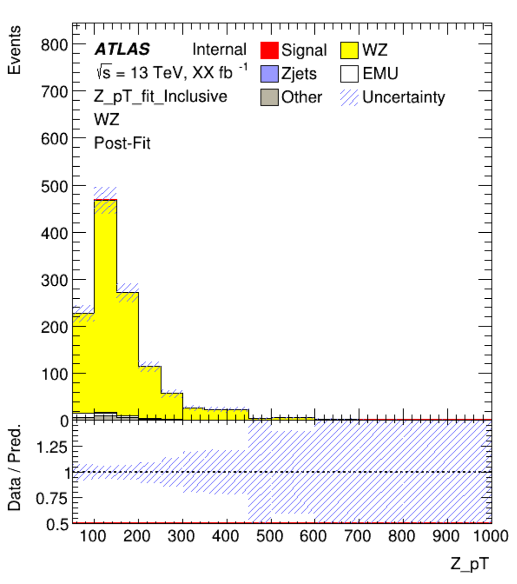
Inclusive ZZ production: Signal Region
76< M_{ll} < 106 GeV
E_T^{miss}> 70 GeV
ΔR_{ll}<1.8
$\Delta\varphi(E_T^{miss}$,ll)>2.3
$\frac{E_T^{miss}}{H_T}$ >0.5
$E_T^{miss}{}_{significance}$ > 10
bjet veto

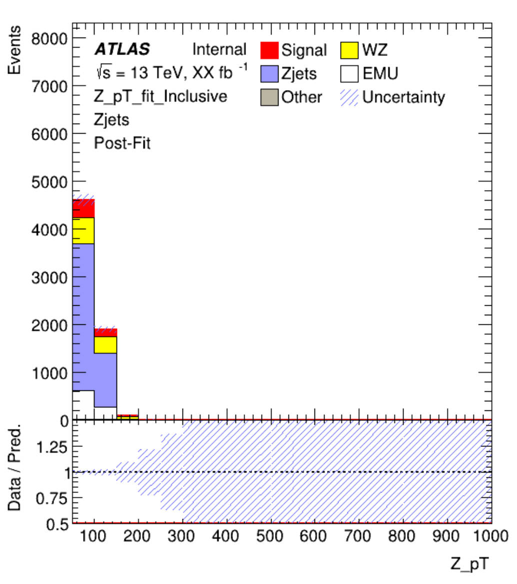
ZZ+2jets production: VBS EWK Region
76< M_{ll} < 106 GeV
$E_T^{miss}{}_{significance}$ > 10
$p_T(leading\ jet)$> 45 GeV
$p_T(leading\ jet)$>40 GeV
$\Delta\eta_{jj}$ > 2
E_T^{miss} > 70 GeV
m_{jj} > 550
bjet veto

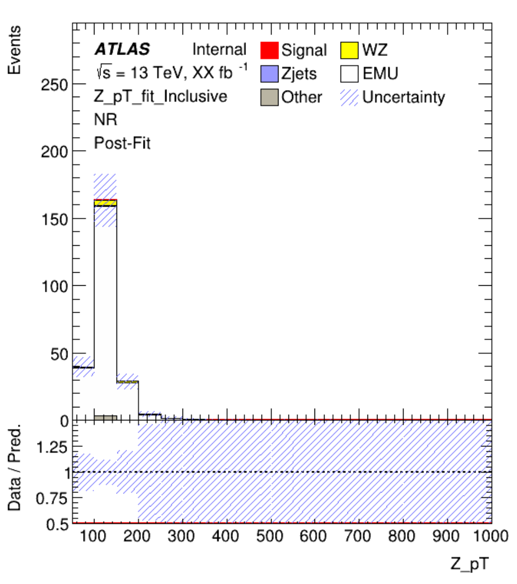
- Control Regions defined with the same phase space as the Signal Region,
  - *3lCR*: 3 leptons in the final state
  - *emCR* : 2 opposite flavor leptons in the final state
  - *Zjets* : Invert $E_T^{miss}{}_{significance}$ < 9 for Inclusive phase space and $E_T^{miss}{}_{significance}$ < 10 for VBS phase space.
- For VBS phase space we define an extra control region, a *QCD enhanced region,* by requiring 100<m_{jj} <550 .

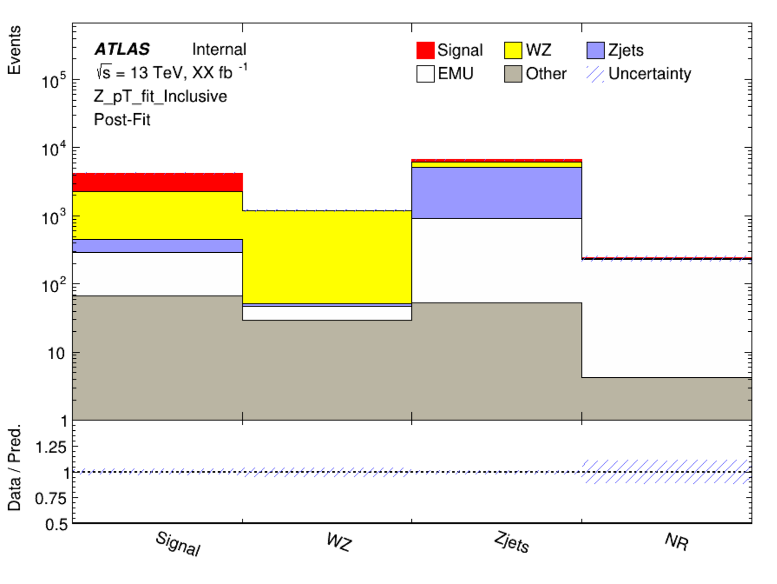
SIMULTANEOUS FIT RESULTS (INCLUSIVE)

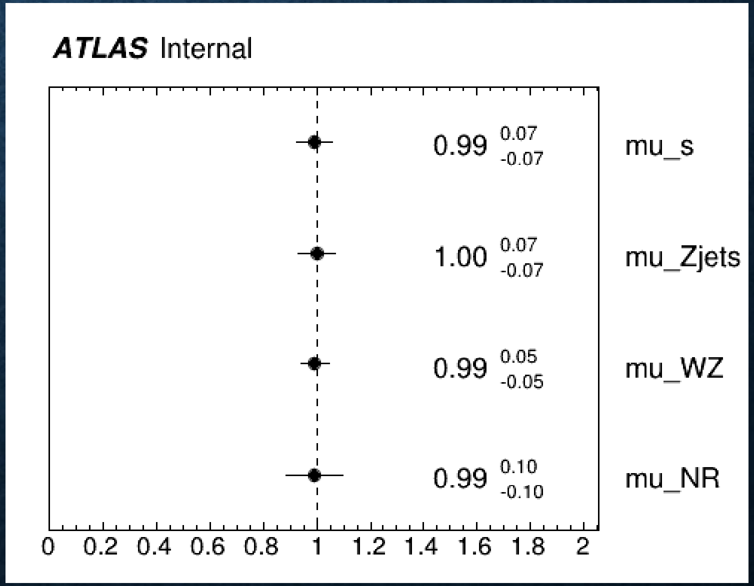






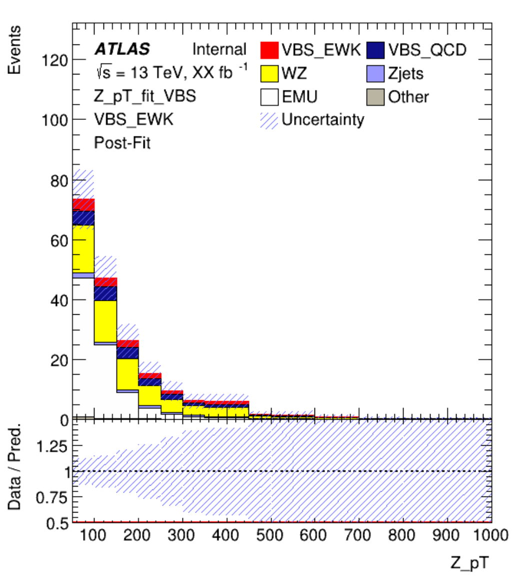


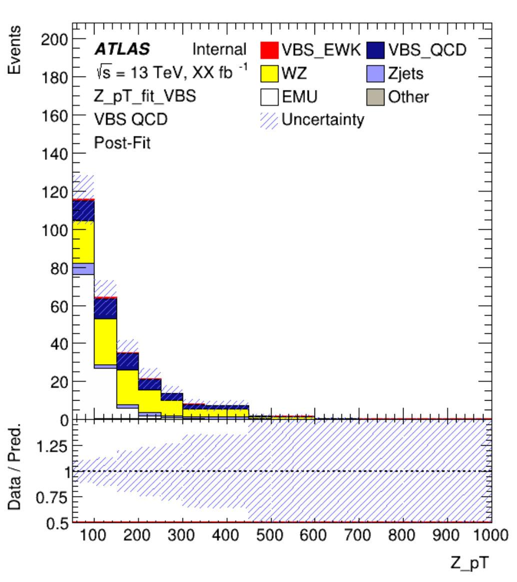


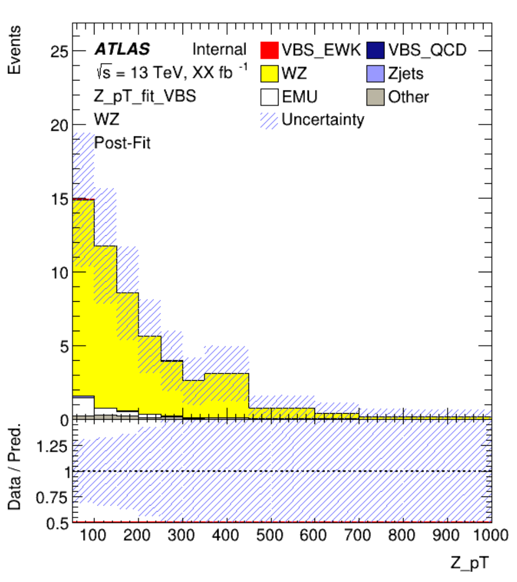


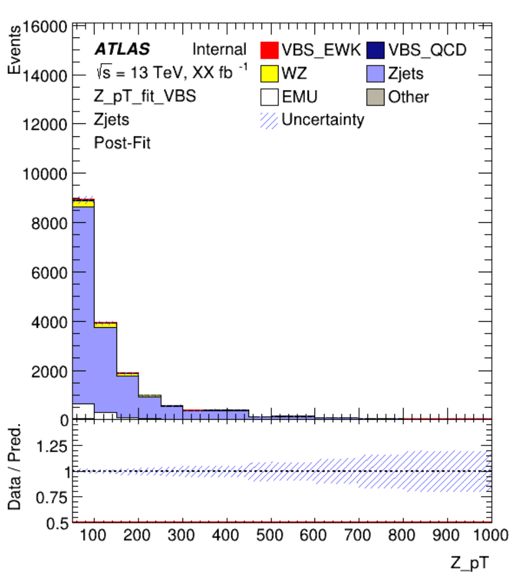
MC expectation as Data

SIMULTANEOUS FIT RESULTS (VBS)

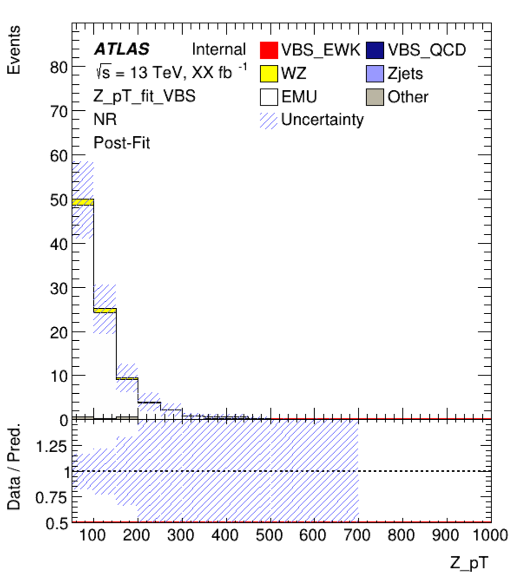


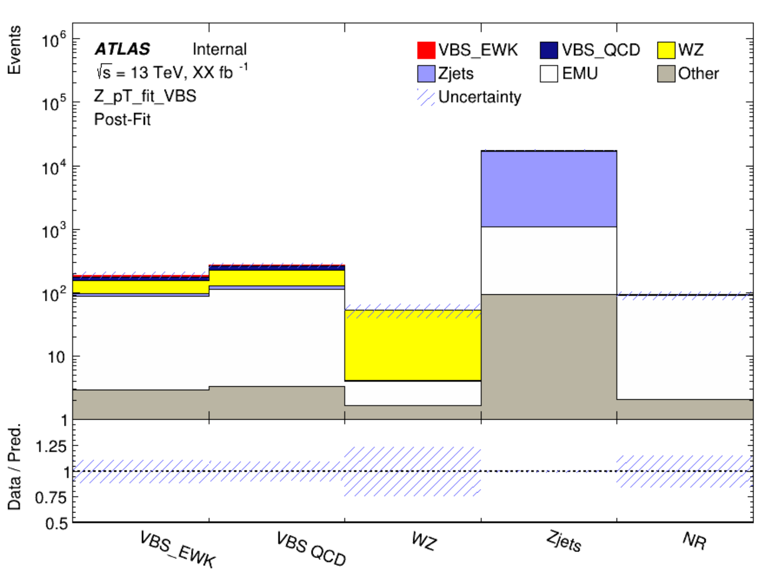


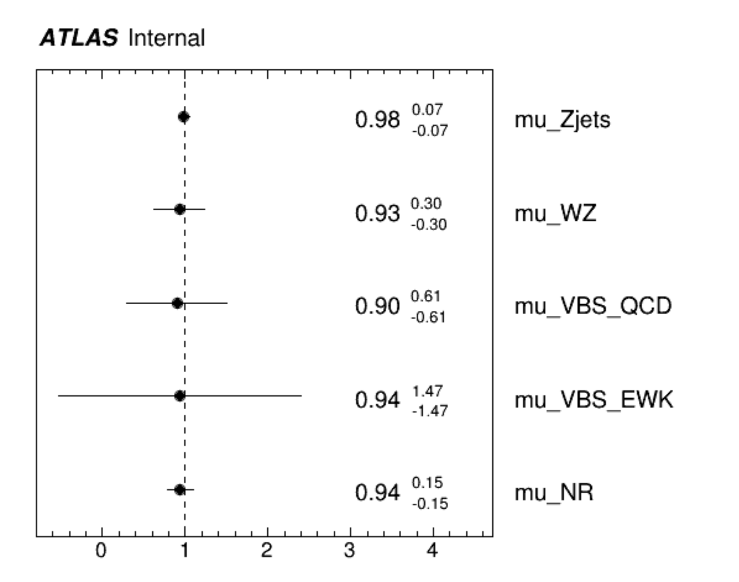




MC expectation as Data





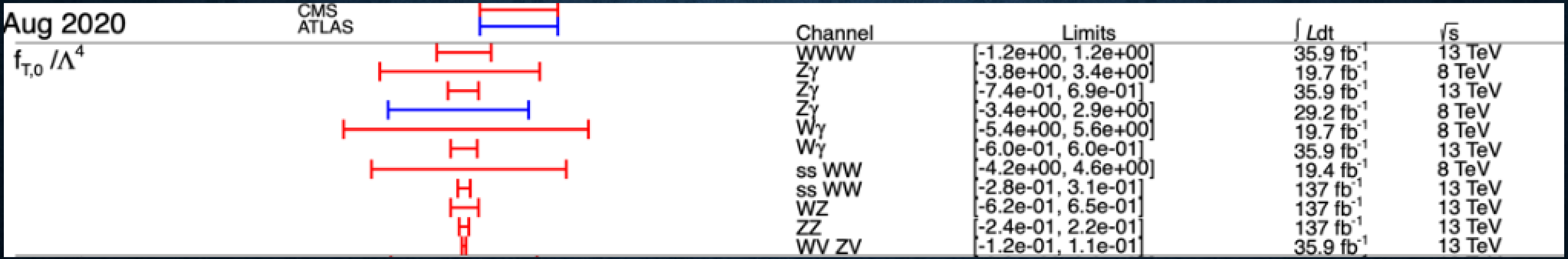


NO EXPECTATION OF VBS EWK MEASUREMENT. EFT FOR NEW PHYSICS.

- Due to low statistics, we do not expect an observation EWK-only production in the VBS phase space.
- This is already established by ATLAS (>5σ) by combining ZZ->4l and ZZ->2l2v : https://arxiv.org/pdf/2004.10612.pdf.
- EWK production cross section is in agreement with the SM.
- We want to check if the experimental data are indicating the presence of new physics.
- The Effective Field Theory (EFT) is the natural way to expand the SM such that the gauge symmetries are respected
  - Provides guidance for searching new physics.
- Λ : energy scale of the new physics.
- O_i *: dim 6 and 8 operators*
- c_i f_i : dimensionless coefficients *"Wilson coefficients"*

$$\mathcal{L} = \mathcal{L}_{SM} + \sum_i \frac{c_i}{\Lambda^2}\mathcal{O}_i + \sum \frac{f_i O_i}{\Lambda^4}$$

- Dimension-8 operators are dominant in anomalous QGC couplings.
- The dimension-8 operators that describe the pure anomalous QGC effects can be divided into three categories:
  - Longitudinal (FS) : FS0,FS1,FS2
  - Transverse(FT) : FT0, FT1, FT2,FT5,FT6,FT7,FT8,FT9
  - Mixed (FM) : FM1,FM2,FM3,FM4,FM5,FM7



limits on FT0.

Picture obtained by : https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsSMPaTGC

EFT SAMPLE PRODUCTION AND STUDIES

- We use the Madgraph5 generator and the Eboli Gonzales models for the generation of EWK EFT ZZjj -> 2l2vjj samples.
  - MadGraph + Pythia
  - Process : p p -> ZZ , QCD=0
    - ZZ -> $e^+ e^-$ $v_\tau v\sim_\tau$ j j (Decay using madspin module)
  - N_events = 20k

- We study :
  - The Decomposition technique
  - The effect of EFT operators at various distributions

DECOMPOSITION

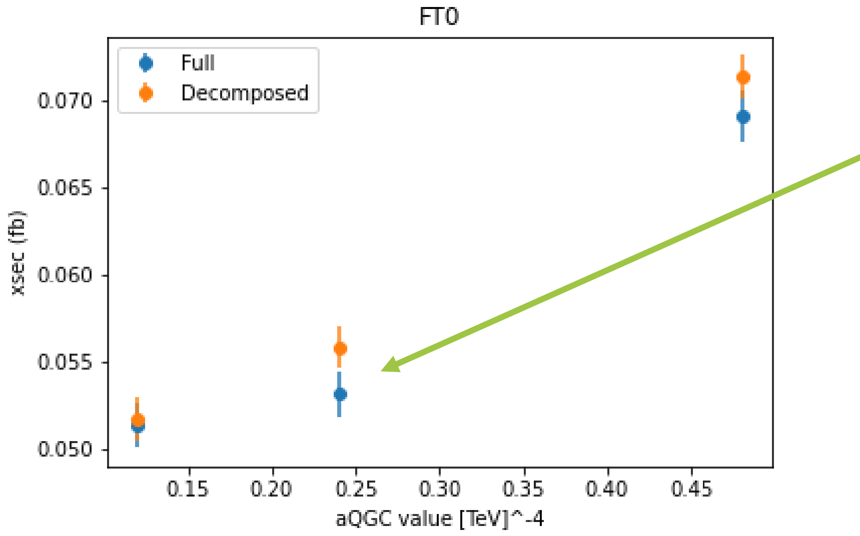
- The EFT cross section can be split in 4 terms:

$$\left|\mathcal{A}_{SM} + \sum_i \frac{f_i}{\Lambda^4}\mathcal{A}_i\right|^2 = |\mathcal{A}_{SM}|^2 + \underbrace{\sum_i 2\frac{f_i}{\Lambda^4}\Re(\mathcal{A}_i^*\mathcal{A}_{SM})}_{\text{Interference SM-EFT}} + \underbrace{\sum_i \frac{f_i^2}{\Lambda^8}|\mathcal{A}_i|^2}_{\text{Pure EFT (quadratic term)}} + \underbrace{\sum_{i,j,i\neq j} 2\frac{f_i f_j}{\Lambda^8}\Re(\mathcal{A}_i^*\mathcal{A}_j)}_{\text{EFT cross term}}$$

- Instead of producing the total process, we can produce the SM, interference, quadratic and cross terms (for 2 or more active EFT operators)
- Useful for rescaling these samples, for different values of coefficients.
    - This dramatically reduces computational time

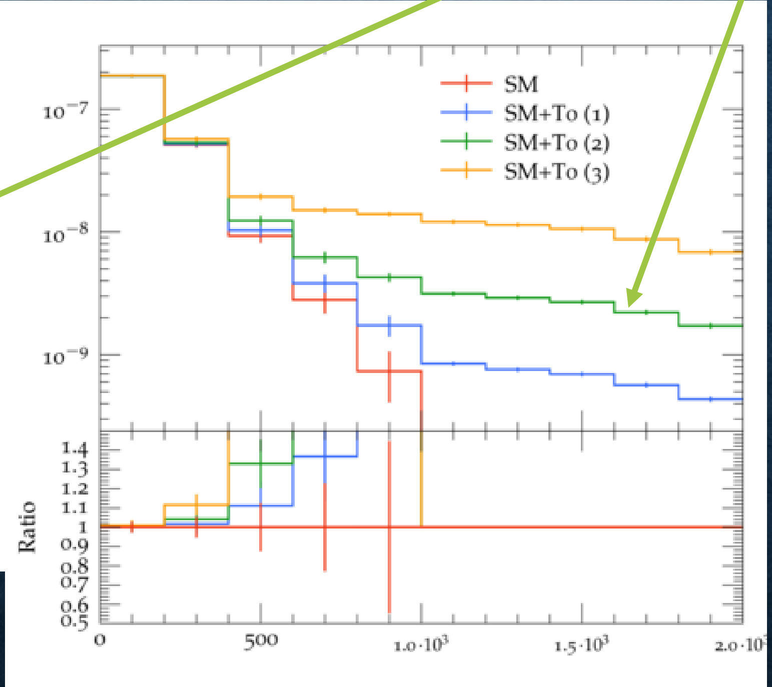
RESULTS FOR FT0 OPERATOR

	xsec (fb)		
	FT0 =0.12 $[TeV]^{-4}$	FT0 =0.24 $[TeV]^{-4}$	FT0 =0.48 $[TeV]^{-4}$
SM	0.05002 ± 0.00122	0.05002 ± 0.00122	0.05002 ± 0.00122
FULL	0.05133 ± 0.00124	0.05312 ± 0.00127	0.06913 ± 0.00152
Decomposition	0.05171 ± 0.00122	0.05582 ± 0.00122	0.07135 ± 0.00125
Δ(FULL,DEC) (%)	0.74	5.08	3.21
QUAD	0.00122 ± 0.00002	0.00486 ± 0.00007	0.01944 ± 0.00028
INT	0.00047 ± 0.00001	0.00094 ± 0.00002	0.00188 ± 0.00004



Comparison of FULL and Decomposed for different FT0 values

Current limit



pT_Z (GeV)

- Full and decomposed are within statistical agreement,
  - FULL having consistently lower yields than Decomposed.
  - Madgraph experts trust the decomposition procedure, so we used these samples to examine the differential cross section.
- In high pTZ values, SM+FT0 is enhanced compared to SM.
  - effect increasing with higher values of the operator.
- Assuming that only one operator contributes at a time
  - we can set limits of the contribution of this operator , by comparing it with the SM predictions for various sensitive variables (pTZ, mjj, MET etc.)

SUMMARY

- Employed a genetic algorithm for optimizing signal region in both inclusive and VBS phase space
- Performed a demonstration of a simultaneous fit for estimating the expected uncertainty of the scaling factors for Signal and Background, using MC expectation as Data (Asimov Dataset)
- Validated the decomposition technique for FT0 EFT operator, and the effect of the FT0 EFT operator at high ptZ values.

THANK YOU

REFERENCES

- *Observation of electroweak production of two jets and a Z-boson pair with the ATLAS detector at the LHC*
- *Measurement of ZZ production in the ℓℓνν final state with the ATLAS detector in pp collisions at s√=13 TeV*
- *Practical Statistics for the LHC*
- *Genetic Algorithms – Introduction :https://www.tutorialspoint.com/genetic_algorithms/genetic_algorithms_introduction.htm*

BACK UP

SCALE UNCERTAINTIES

- Missing Higher orders in perturbation QCD calculations are estimated by performing scale variations, i.e. varying the renormalization and factorization scales by a factor of 2.
  - 9 point variation: $\{\mu_R, \mu_F\} \otimes$\{0.5,0.5\},\{0.5,1\},\{1,0.5\},\{1,1\},\{2,1\},\{1,2\},\{2,2\},\{2,0.5\},\{0.5,2\}
  - 7 point variation: $\{\mu_R, \mu_F\} \otimes$\{0.5,0.5\},\{0.5,1\},\{1,0.5\},\{1,1\},\{2,1\},\{1,2\},\{2,2\}
  - 3 point variation: $\{\mu_R, \mu_F\} \otimes$\{0.5,0.5\},\{1,1\},\{2,2\}

- Calculate estimated yields for each variation by replacing weight_gen with the respective weight variation ($\{\mu_R, \mu_F\}$ =\{1,1\} corresponds to the nominal weight_gen)
- Calculate the differences between the nominal value and each variation
  - Max_error=max[Yields($\mu_{R,i}, \mu_{F,i}$)- Yields($\mu_{R,0}, \mu_{F,0}$)]
  - Min_error=min[Yields($\mu_{R,i}, \mu_{F,i}$)- Yields($\mu_{R,0}, \mu_{F,0}$)]
  - Where Yields($\mu_{R,i}, \mu_{F,i}$):Yields for ith scale variation , Yields($\mu_{R,0}, \mu_{F,0}$) :Yields for the central scale

PDF UNCERTAINTIES

- PDF uncertainties: There are several sources of uncertainty that affect the determination of PDFs:
  - Experimental uncertainties entering the datasets used in the PDF fits => encapsulated in the PDF error eigensets
  - Uncertainty on functional form used in the PDF fits -> encapsulated in the PDF error eigensets
  - Missing Higher Order Uncertainties -> the scale variations are supposed to give an estimate of this uncertainty
  - (other theory uncertainties: flavour scheme, nuclear effects, … are usually not taken into account, but some of these effects (e.g. flavour scheme) are probed when comparing different PDF sets)
- For NNPDF sets:
  - ensemble of PDFs is provided.
  - the used value is the mean of all the ensembles
  - the uncertainty is the standard deviation .
- Calculate estimated yields for each variation again by replacing weight_gen with the respective weight variation
  - X0 is the mean of all X1..X100
  - The uncertainty is calculated by:

$$\sqrt{\frac{\sum(X_\iota - X_0)^2}{N}}$$

SAMPLE PRODUCTION

- Samples produced on lxplus, using 21.6.57 AthGeneration version
- Job option obtained by ZZ->4l group (Alexandros Marantis, Ioannis Maznas)
  - MadGraph + pythia
  - Process : p p -> ZZ , QCD=0
    - ZZ -> $e^+ e^-$ $v_\tau v\sim_\tau$ j j (Decay using madspin)
- Rivet for analysis
  - (Script obtained by previous VBS analysis, and modified to run for Rivet version 3.1.2)
  - 21.6.67 AthGeneration version was used to compile and run Rivet analysis.
- N_events = 20k.
  - Due to some errors in production, multiple INT samples were produced for FT8 and FT9.
    - FT8 : 3 samples of 5k events (15k total)
    - FT9 : 40 samples of 500 events (20k total)
- drjj cut applied at generation level to avoid singularities.
  - For FT0,1,2,5,6,7 drjj > >0.05
  - For FT8 and FT9 drjj > 0.1