Search for a heavy pseudoscalar boson A decaying into a Z and h boson in the semileptonic final state

HEP 2018
"Recent Developments in High Energy Physics and Cosmology"

Emilios Ioannou[1], Alberto Zucchetta[2], Jehad Mousa[1], Panos Razis[1]

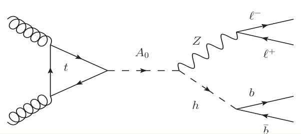
[1]University of Cyprus, [2]University of Zurich

Athens, Greece, 28/3 -1/4/2018

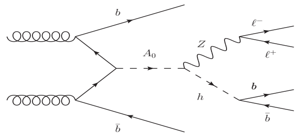
Physics Motivations

- A pseudoscalar boson A is predicted in many BSM theories (2HDM, MSSM).
- In 2HDM (MSSM included), if $m_h + m_Z \lesssim m_A \lesssim 2m_t$, the branching ratio $A \to Zh$ might be important $\to$ strong dependence on model parameters.
- Clean final state thanks to Z decays, large $\mathrm{BR}(h \to b\bar{b})$.
- The SM-like Higgs boson is interpreted as the light scalar h with $m_h = 125\ GeV$.
- A pseudoscalar A could be generated by two Higgs mechanisms.

gluon-gluon fusion



associated b-quarks production



In this presentation

- ✓ Signal generation.
- ✓ Analysis strategy.
- ✓ Background estimation.
- ✓ Discriminators (angular, event shape).
- ✓ Systematic uncertainties.
- ✓ Results.

⋆ Results of blind analysis are being presented.

Signal generation

University of Cyprus

- **Next-to-leading-order generators** are available **aMC@NLO**.
  - 2HDMtII_NLO UFO model with QCD corrections for the type II 2HDM, both the top and bottom are massive and their yukawa are renormalized like their masses, i.e. on-shell.
- Possibility to fully simulate loop-induced processes (full top, bottom loop).
- Use **MadSpin** to have the correct $A \to Zh$, $Z \to \ell\ell$, $h \to b\bar{b}$ spin correlations.
- Mass points (500K events for each point): $m_A = 225, 250, 275, 300, 325, 350, 400, 500, 600, 700, 800, 900, 1000 GeV$.
- Samples were generated for the processes:
  - $A \to Zh \to \ell\ell b\bar{b}$
  - $A \to Zh \to \nu\bar{\nu}b\bar{b}$
  - $b\bar{b}A \to Zh \to \ell\ell 2b2\bar{b}$
  - $b\bar{b}A \to Zh \to \nu\bar{\nu}2b2\bar{b}$

Background and signal samples

Data and Monte Carlo samples

- Backgrounds are represented by all the processes with large MET or 2 leptons, and b-jets in the final state.

Data
SingleMu/Run2016(B-H)-23Sep2016-v*
SingleElectron/Run2016(B-H)-23Sep2016-v*
MET/Run2016(B-H)-23Sep2016-v*
Monte Carlo backgrounds
DYJetsToLL_M-50_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8
DYJetsToLL_M-50_HT-*_TuneCUETP8M1_13TeV-madgraphMLM-pythia8
DY*JetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8
DYJetsToNuNu_M-50_HT-*_TuneCUETP8M1_13TeV-madgraphMLM-pythia8
WJetsToLNu_M-50_HT-*_TuneCUETP8M1_13TeV-madgraphMLM-pythia8
TT_TuneCUETP8M1_13TeV-powheg-pythia8
TTToSemilepton_TuneCUETP8M1_13TeV-powheg-pythia8
TTTo2L2Nu_TuneCUETP8M1_13TeV-powheg-pythia8
ST_*_13TeV-amcatnlo-pythia8_TuneCUETP8M1
WWTo*_13TeV_amcatnloFXFX_madspin_pythia8
WZTo*_13TeV_amcatnloFXFX_madspin_pythia8
ZZTo*_13TeV_amcatnloFXFX_madspin_pythia8
ZH_HToBB_ZTo*_M125_13TeV_amcatnloFXFX_madspin_pythia8
WH_HToBB_WTo*_M125_13TeV_amcatnloFXFX_madspin_pythia8
Monte Carlo signal
GluGluToAToZhToLLBB_M*_13TeV-madgraphMLM-pythia8
GluGluToAToZhToNuNuBB_M*_13TeV-madgraphMLM-pythia8
BBAToZhToLLBB_M*_13TeV-madgraphMLM-pythia8
BBAToZhToNuNuBB_M*_13TeV-madgraphMLM-pythia8

- Golden JSON (35.9 fb^{-1}): `Cert_271036-284044_13TeV_23Sep2016ReReco_Collisions16_JSON.txt`
- All samples are from the Summer16 official CMS production
- Pile-up reweighting using the official recipe ($\sigma_{in} = 69200 \mu b$)
- LO samples corrected for NLO QCD and NLO EWK contributions

The analysis in a nutshell

- Identify and select the decay products in the detector:
  - Two opposite sign, same flavour, isolated leptons or large $\not{E}_T$.
  - Two jets tagged as originated from b-quarks.
- Build intermediate candidates Z ($\ell^+\ell^-$ *or* $\nu\bar{\nu}$) and h ($b\bar{b}$) up to the A candidate.
- The signal would manifest itself as a narrow peak in the mass distribution over the SM background continuum.
- Define appropriate control regions (CR) to check background shape and normalization with data by inverting selections.

Pre-selections of $Z \to \ell^+\ell^-$

- At least two opposite sign, same flavour e or μ.
- At least two jets with $p_T > 30\ GeV$.
- The Higgs candidate (h) is reconstructed from the highest-CSV jets.
- No boost requirements.

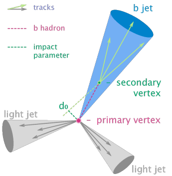
Pre-selections of $Z \to \nu\bar{\nu}$

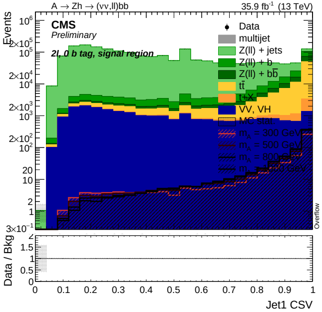
- No isolated leptons ($e,\ \mu,\ \tau$).
- $\not{E}_T > 200\ GeV$.
- At least two jets with $p_T > 30\ GeV$.
- The Higgs candidate (h) is reconstructed from highest-CSV jets.
- $p_T^h > 200\ GeV$ (back-to-back).
- No jet closer in $\Delta\phi < 0.4$ to the $\not{E}_T$.

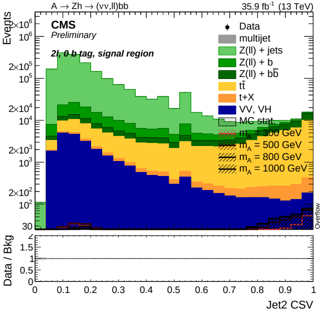
Combined-Secondary Vertex (CSV)

University of Cyprus

- The SM Higgs candidate will be reconstructed from the highest-CSV jets.
  - CSV: algorithm which identifies jets as b-jets, taking into account:
    - ✓ Kinematical and topological secondary vertex variables.
    - ✓ Information from track impact parameters.







CSV working point selections

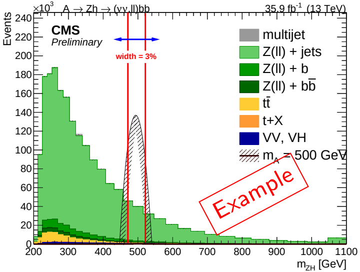
Working point	Cut
Loose	0.5426
Medium	0.8484
Tight	0.9535

- A proper CSV working point selection should be defined.
- Which should be the most suitable criterion?
  - ✓ CSV working point selection → larger significance.

University of Cyprus

- How do we compute the Significance?

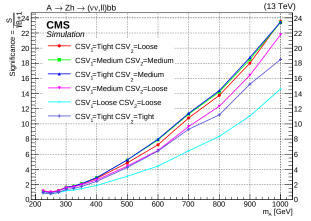
(If one selection can reject background events from the signal region, while signal events remain unaffected)

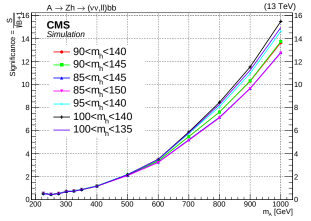


✓ Choosing a window with 3% width for each A candidate.

✓ Integrating the bins (signal and background) which are included in the window, separately.

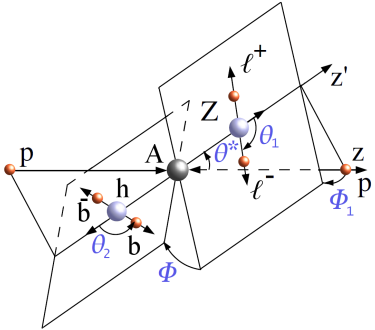
✓ Using the formula $FOM = \frac{Signal}{\sqrt{Background+1}}$ to extract the Significance for each mass point.





- CSV working point selection: (Medium, Medium) has been selected because of the larger significance.
- Higgs mass window: (**black line**) $100 < m_h < 140$ GeV→ improved significance by up to 10-15%.

- In order to distinguish the signal from background, two discriminators are introduced:
  - Angular discriminator ($Z + b\bar{b}$).
  - Event shape discriminator ($t\bar{t}$).
- Angular Discriminator:
  - A distinctive signature of signal is that the Z boson is transversally polarized w.r.t. its flight direction for CP conservation
  - Additional discrimination comes from the $A \to Zh \to \ell\ell bb$ production ($\cos\theta^*$, Φ_1) and decay angles ($\cos\theta_1$, $\cos\theta_2$, Φ)



$\theta*$: angle between the A flight direction and the beam in the A rest frame.

θ_1 : angle between the ℓ^- and the Z flight direction in the Z rest frame.

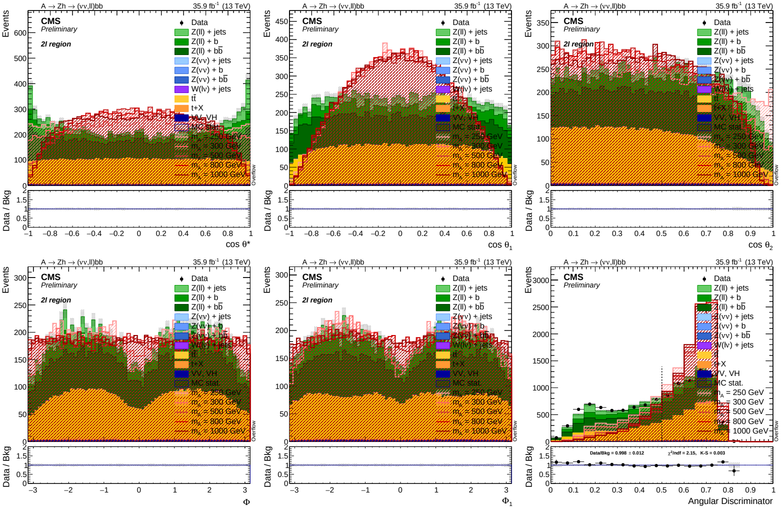
θ_2 : same as θ_1, but for H and jets. The jet is chosen by convention.

Φ : angle between the Z and h decay planes.

Φ_1 : angle between the Z decay plane and the plane where the Z and the beam directions lie.

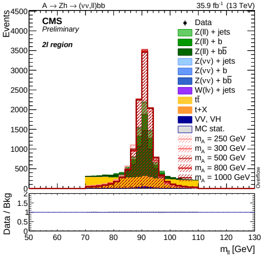
Angular variables

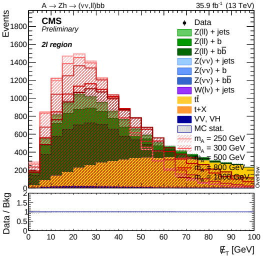
- $\cos\theta*$ and $\cos\theta_1$ most discriminating, take advantage of the transversal polarization of the Z
- Likelihood discriminator trained against signal and Z+b(b) background.

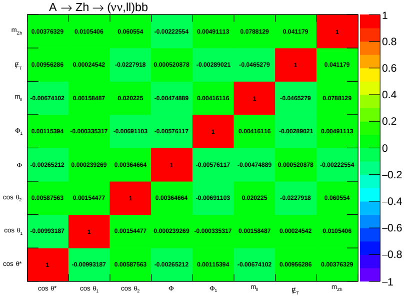


Event shape Discriminator

- Also $t\bar{t}$ is large ($1/3 \sim 1/2$) of the total background
- Dedicated discriminator trained against $t\bar{t}$
- Discriminating variables: Z mass, MET



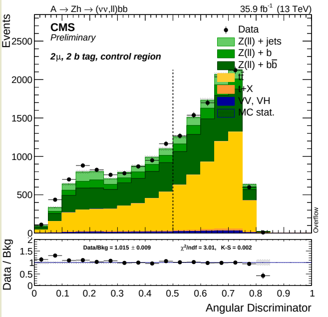


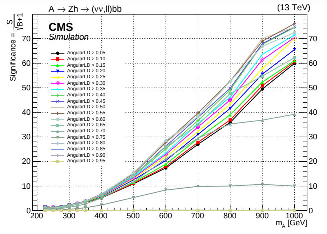


- Check the correlation matrix.
- Likelihood discriminator ensures optimal performance in absence of correlations between the variables.
- LD inputs also uncorrelated w.r.t. A boson mass.
- This is necessary to avoid any background sculpting when cutting on the discriminator.
- Correlation is very small $\rightarrow$ allowing implementation of cut on the discriminators.

Selection Optimization Discriminators

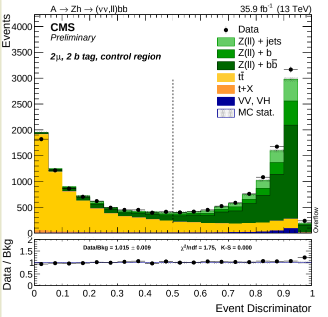
Angular discriminator

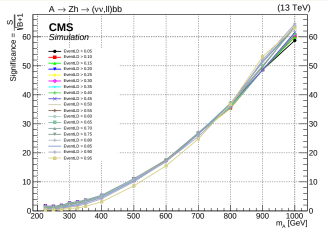




- Optimized selection: $AngularLD > 0.5$

Event shape discriminator



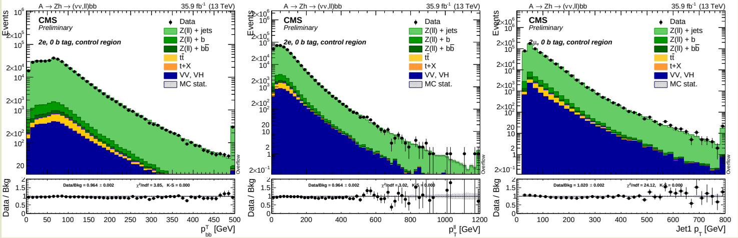


- Optimized selection: $EventLD > 0.5$

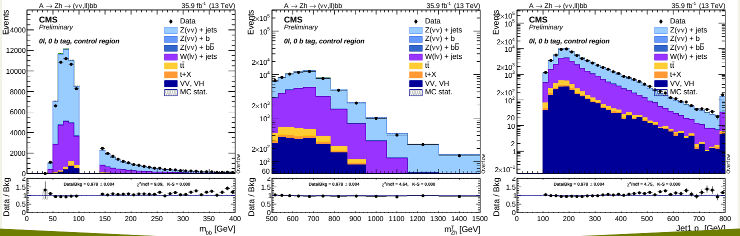
Data-driven background normalization (Z+jets)

Region	Z mass [GeV]	h mass [GeV]	CSV_1	CSV_2	$\not{E}_T$ [GeV]
Z+jets	$70 < m_{\ell\ell} < 100$	$m_{bb} < 100, m_{bb} > 140$	!Medium	!Medium	< 100
	$e,\ \mu,\ \tau$ veto	$m_{bb} < 100, m_{bb} > 140$	!Medium	!Medium	> 200

$Z \to \ell^+\ell^-$



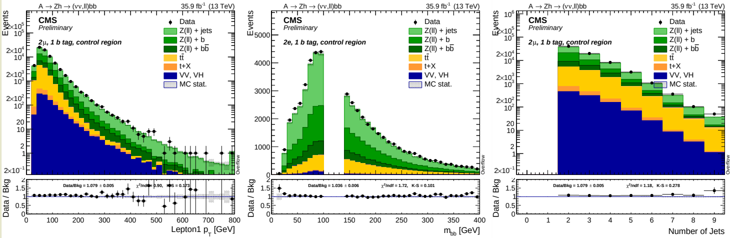
$Z \to \nu\bar{\nu}$



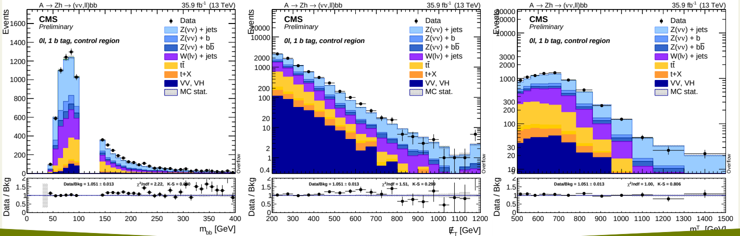
Data-driven background normalization (Z+b)

Region	Z mass [GeV]	h mass [GeV]	CSV_1	CSV_2	$\not{E}_T$ [GeV]
Z+b	$70 < m_{\ell\ell} < 100$	$m_{bb} < 100, m_{bb} > 140$	Medium	!Medium	< 100
	$e,\ \mu,\ \tau$ veto	$m_{bb} < 100, m_{bb} > 140$	Medium	!Medium	> 200

$Z \to \ell^+\ell^-$



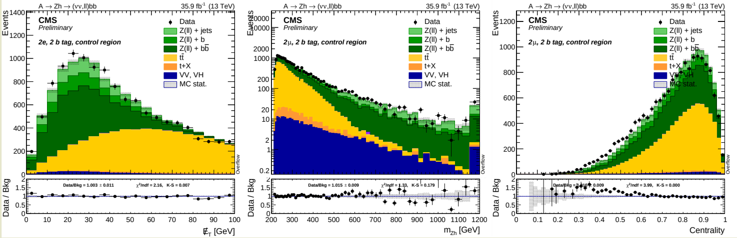
$Z \to \nu\bar{\nu}$



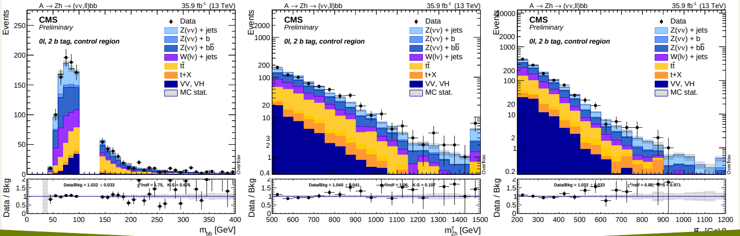
Data-driven background normalization ($Z + b\bar{b}$)

Region	Z mass [GeV]	h mass [GeV]	CSV_1	CSV_2	$\not{E}_T$[GeV]
Z+bb	$70 < m_{\ell\ell} < 100$	$m_{bb} < 100, m_{bb} > 140$	Medium	Medium	< 100
	0ℓ	$m_{bb} < 100, m_{bb} > 140$	Medium	Medium	> 200

$Z \to \ell^+\ell^-$



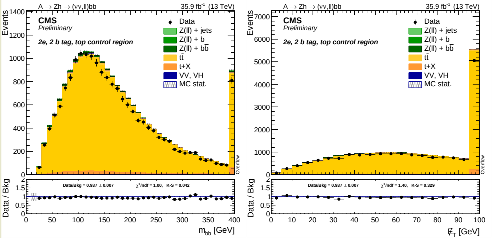
$Z \to \nu\bar{\nu}$



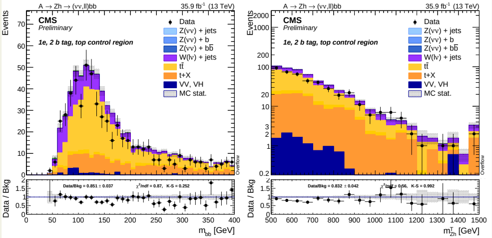
Data-driven background normalization (Top quark)

Region	Z mass [GeV]	h mass [GeV]	CSV_1	CSV_2	$\not{E}_T$ [GeV]
Top and * W+jets	$m_{\ell\ell} < 70$, $m_{\ell\ell} > 110$	$m_{bb} < 100$, $m_{bb} > 140$	Medium	Medium	-
	$1e$ or 1μ	$m_{bb} < 100$, $m_{bb} > 140$	Medium	Medium	> 200

$Z \rightarrow \ell^+\ell^-$



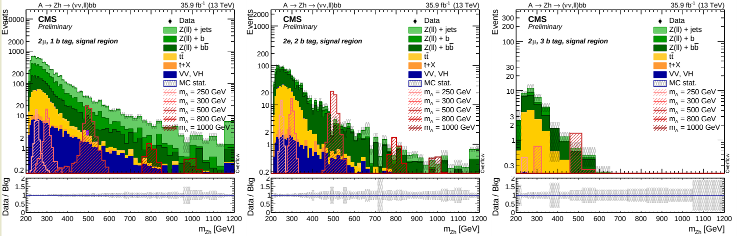
$Z \rightarrow \nu\bar{\nu}$



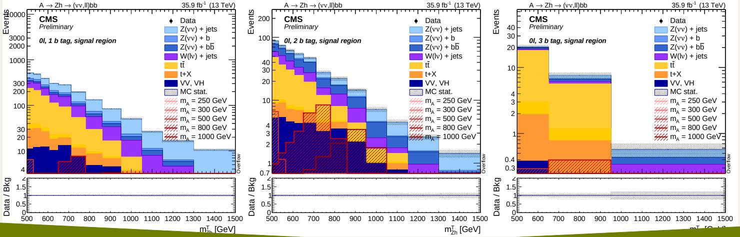
Signal Regions

Region	Z mass [GeV]	h mass [GeV]	CSV_1	CSV_2	$\not{E}_T$[GeV]
SR	$70 < m_{\ell\ell} < 100$	$100 < m_{bb} < 140$	Medium	Medium	< 100
	0ℓ	$100 < m_{bb} < 140$	Medium	Medium	> 200

$Z \to \ell^+\ell^-$ (Three different signal regions)



$Z \to \nu\bar{\nu}$ (Three different signal regions)



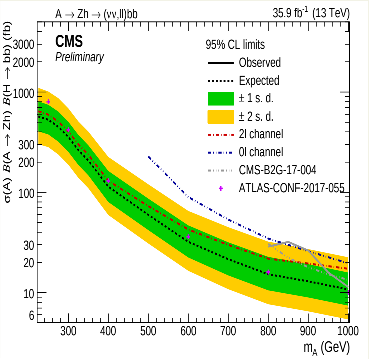
University of Cyprus

- Summary of systematic shape uncertainties for backgrounds and signals

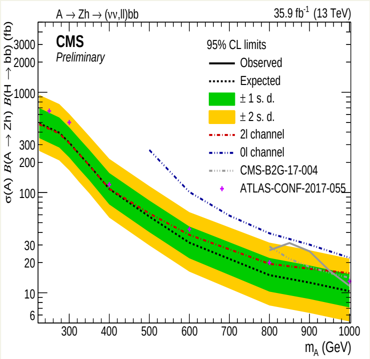
	Shape	Main backgrounds (Drell-Yan, $t\bar{t}$)	Other electroweak (single-top, VV, Vh)	Signal
Lepton and trigger efficiency	✓	-	2-3%	2-3%
Jet Energy Scale	✓	-	5%	2-6%
Jet Energy Resolution	✓	-	2%	1-2%
b-tagging	✓	-	4%	5-8%
Unclustered MET	✓	-	1%	1%
Pile-up	✓	-	1%	1%
PDF	✓	-	3-5%	4-8%
top quark p_T (only $t\bar{t}$)	✓	8–15%	-	-
Factorization and renormalization scale	✓	2-6%		6-14%
Monte Carlo modeling	✓	1-15 %		-
Monte Carlo statistics	✓	1–20%		-
Interpolation to SR		1–12%		-
Extrapolation to 3 b-tag SR		20–26% (3 b-tag only)		-
Cross section		-	2-10%	-
Luminosity		-	2.5%	2.5%

Results

gluon-gluon fusion



b-quark associated



- Expected limits with all systematics, included MC statistics.
- Gained $\sim$20% from discriminator optimization.
- Lost $5-15\%$ due to MC stat. uncertainty (depends on mass point).

Conclusions

University of Cyprus

In this presentation

- An overview of the $A \to Zh \to \ell^+\ell^-(\nu\bar{\nu})b\bar{b}$ analysis has been presented.
- All results which have been presented, are from blind analysis.
- After optimization of selections, the improvement is significantly better.
- The analysis is stable and optimized.

Thanks for your attention!

BackUp

Physics objects

Electrons

- Trigger: `HLT_Ele27_WP*_Gsf_v*` OR `HLT_Ele115_CaloIdVT_GsfTrkIdT_v*`
  - SF with T&P package
- $p_T^{1,2} > 30,\ 10$ GeV, $|\eta| < 2.5$
- Cut-based `Loose` Id
- PFIso included in Id
  - SF from EGamma POG

Muons

- Trigger: `HLT_IsoMu24_v` OR `HLT_IsoTkMu24_v`
  - SF for trigger from Muon POG
- $p_T^{1,2} > 30,\ 10$ GeV, $|\eta| < 2.4$
- At least one `Tight`, the other `Loose`
- PFIsolation $\Delta\beta$ corrected $< 0.15, 0.25$
  - SF for Id+Iso from Muon POG

Taus

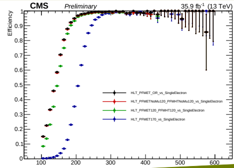
- $p_T > 18$ GeV, $|\eta| < 2.5$
- `Loose HPSPFTauId`

Jets

- AK4 PFJets with CHS
- default JEC L1L2L3(L2L3)
- $p_T > 30$ GeV, $|\eta| < 2.5$
- `PFLooseId`
- $\Delta R_{\ell,j} > 0.4$
- CSVIVFV2 b-tagging
  - SF from BTV POG

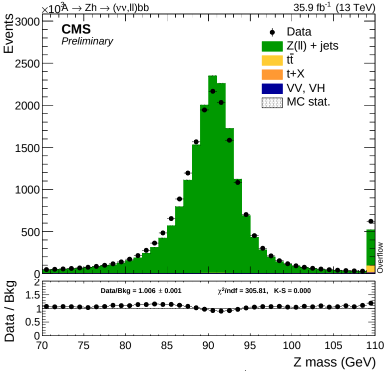
MET

- Trigger: `HLT_PFMET110_PFMHT110_*`
  - 95% efficient for $MET > 200$ GeV
- Type-I corrected MET
- Recommended filters applied



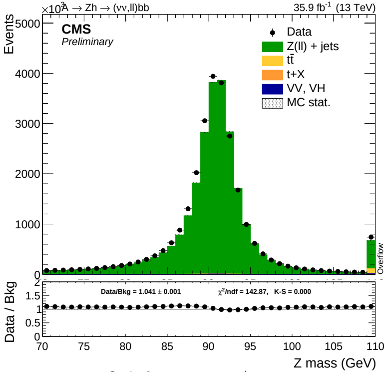
- Reconstruction, identification, isolation applied in both data and MC (with SF).
- Good out-of-the-box agreement: $\sim 4\%$ in $Z \to \mu^+\mu^-$

$Z \to e^+e^-$



$Scale\ factor = 1.006 \pm 0.001$

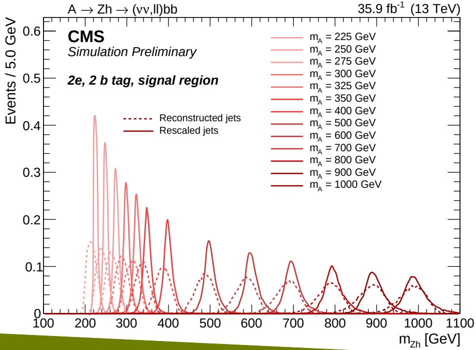
$Z \to \mu^+\mu^-$

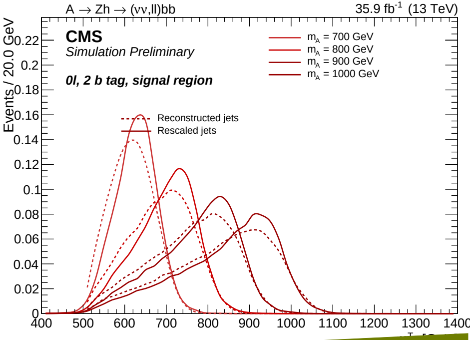


$Scale\ factor = 1.041 \pm 0.001$

Kinematic constraints

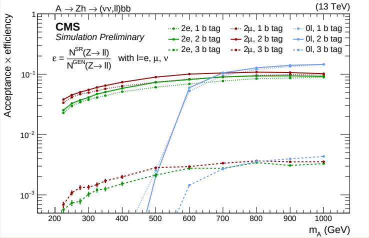
- Exploit the measurement of the Higgs mass m_{bb} to rescale the jets' p_T to match $m_{jj} = 125$ GeV
- The rescale factor is proportional to the JEC uncertainty of each jet.
- New widths are about $1/4$ for low masses, $1/2$ for higher masses
  - at high m_A, the resolution is dominated by p_T^{bb}
- No so effective in the $Z \to \nu\nu$ channel, where the MET resolution is dominating
- Similar constrain is applied to the leptos' p_T to match m_Z, but the impact on the m_A resolution is much smaller.



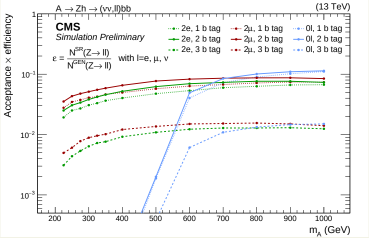


Signal Efficiency

gluon - gluon fusion



b-quarks associated



Comments

- Overall signal efficiency is 5-15%.
- Half of the events recovered in the 1 b-tag category.
- 3 b-tag category is useless for gluon-gluon fusion signal but is not negligible in the b-quark associated production ($\approx$ 20% fraction of events with the b-quarks in acceptance).

Systematic uncertainties

Jet Energy Scale:

JES is varied by its uncertainty and propagated to composite objects ($\not{E}_T$). Effect on the A peak is negligible after the kinematic fit.

Jet Energy Resolution:

JER is evaluated by smearing the jet four-momentum according to the measured uncertainty.

Missing Transverse Energy:

We consider standard recipes for Unclustered MET only. Variations for JES, JER and leptons already included in the corresponding systematics.

Pile-up and PDF:

Followed standard prescriptions as reported in TWikis.

Systematic uncertainties (continue)

University of Cyprus

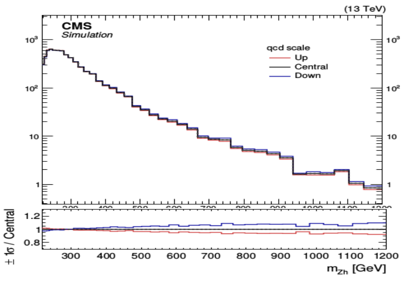
Lepton uncertainties:

propagated effect of Trigger, Reco, Id and Isolation scale factors.

	Trigger	Id + Iso (1ℓ)	Id + Iso (2ℓ)
Electrons	0.9% (2%)	1.9%	3.6%
Muons	3% (0.5%)	1.0%(2%+1.5%)	1.9% (4%)

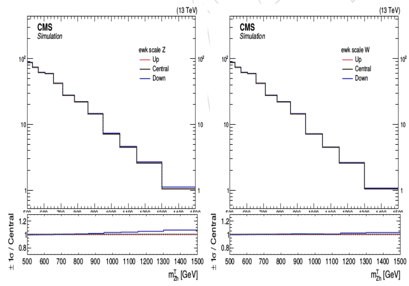
QCD renormalization and factorization scale:

MC reweighted charging Q by factor 2.



Electroweak NLO corrections:

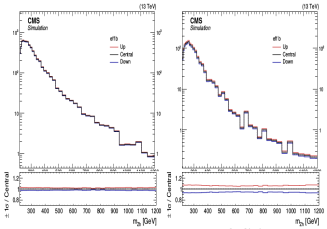
Take the full effect of the correction as systematic uncertainty.



Systematic uncertainties (continue)

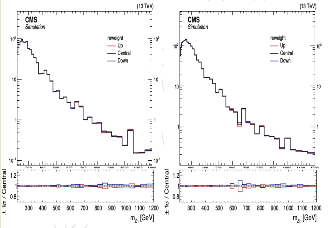
B-tagging:

Scale factors are varied by their error, weights propagated to the final distributions.



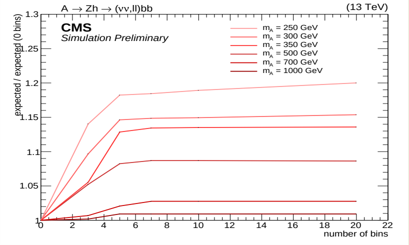
MC modeling:

reweight in the Centrality variable to give the background a degree of freedom and cover minor mismodelings.



MC statistics:

Bins varied by their statistic uncertainty for each background independently. Only the 5 bins with the highest S/B are considered.



Systematic uncertainties (continue)

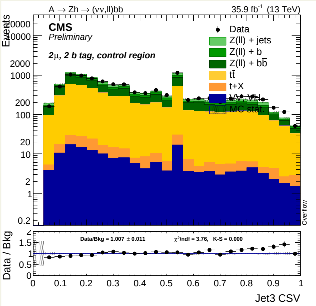
Interpolation to the SR

- In some CR, there is a non-negligible mismodeling in m_{bb}
- This may lead to a bias in the estimation of the background normalization
- Take the difference between the lower and higher sideband as systematic uncertainty

	1 b-tag	2 b-tag	3 b-tag
0ℓ	12.1%	1.2%	3.6%
$2e$	1.3%	5.4%	0.7%
2μ	1.2%	4.7%	7.1%

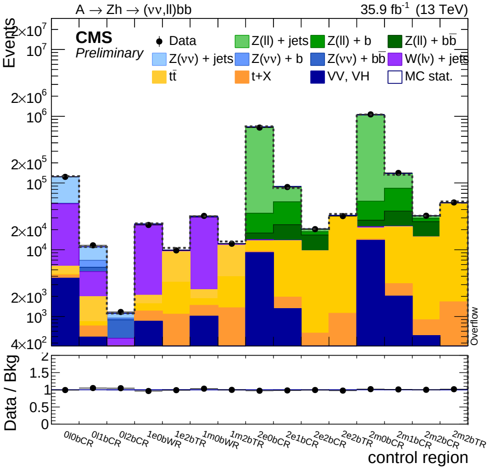
Interpolation to the SR

- Mismodeling in 3rd jet CSV observed consistently in all control regions
- Likely that this effect also yields an excess of data in the 3 b-tag SR
- Take the difference from 1 between the data and bkg simulation in 3 b-tag control regions (exaclty as the 2 b-tag CR, but with 3 or more b-tags) as systematic uncertainty
- The difference is 20-25% in all final states



- A background-only fit is performed deriving the background scale factors.

parameter	scale factor
DYJets-0b	0.9880 ± 0.0006
DYJets-1b	1.154 ± 0.006
DYJets-2b	1.021 ± 0.011
TTbar-1l	0.925 ± 0.007
TTbar-2l	0.965 ± 0.003
WJetsToLNu	0.988 ± 0.003



Interpretation in 2HDM

University of Cyprus

2HDM in nutshell

- Two type of 2HDM are taken into account:

Type I: all quarks coupled to the same Higgs doublet (Φ_2). Discrete symmetry ($\Phi_1 = -\Phi_1$).

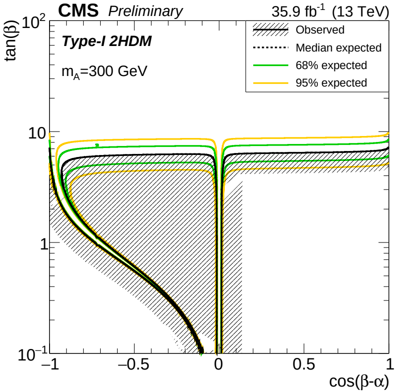
Type II: Right-handed quarks ($Q = 2/3$) coupled to Φ_2. Right-handed quarks ($Q = -1/3$) and leptons coupled to Φ_1.

- Large number of free parameters: m_h, m_H, m_A, $m_{H^\pm}$, m_{12}, β, α, $\lambda_{6,7}$
- Recommendations from the Higgs Working Group:
  - $m_H = m_A = m_{H^\pm}$: A searches not really dependent on m_H, $m_{H^\pm}$ anyway
  - $m_{12}^2 = m_A \frac{\tan\beta}{1+\tan^2\beta}$: discrete Z_2 symmetry broken as in MSSM
  - $\lambda_{6,7} = 0$ to avoid CP-violation at tree level
- Two parameters left other than m_A:

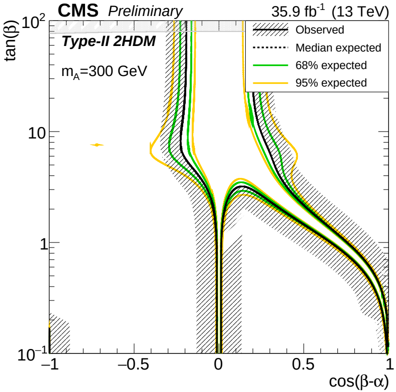
α : mixing angle of the two doublets

$\tan\beta$: the ratio of the *vev* of the two doublets

- For each mass point, perform a scan in $tan\beta$ and $\cos(\beta-\alpha)$
  - $0.1 < \tan\beta < 100$,
  - $-1 \leq \cos(\beta-\alpha) \leq 1$ $\qquad 0 \leq \beta - \alpha \leq \pi$
- Cross sections and Br calculated at NNLO with SuSHi 1.6.1 and 2HDMC 1.7.0



Expected limit on $\sigma \times Br$ in Type I 2HDM



Expected limit on $\sigma \times Br$ in Type II 2HDM

- Expected exclusion up to $\tan\beta \approx 6 \sim 7$ in Type I.
- Large parameter space excluded in Type II, also at high $\tan\beta$.