Higgs Coupling Measurement with papas

Colin Bernet (IPNL)

FCC-ee Higgs Working Group

Dec 11, 2017

Goal

- Measure the Higgs couplings for various detector hypotheses and scenarios
- Need:
  - fast simulation (papas)
  - detector models
  - fitting infrastructure
- First step (today): reproduce existing results
  - LEP3 note (validate papas vs CMS full simulation, analysis code)
  - TLEP note (validate global fit method and model)
- Later:
  - update running scenario ($\sqrt{s}$, lumi)
  - try other detector models and show improvement w/r CMS

Outline

- Tools:
  - papas
  - heppy analysis framework
- Analyses (vs LEP3 note)
  - ZH→llX, ZH→llbb, ZH→ννbb
  - yield extraction
- Global fit for the couplings (vs TLEP note)
- Summary and plans

Papas: Tracker

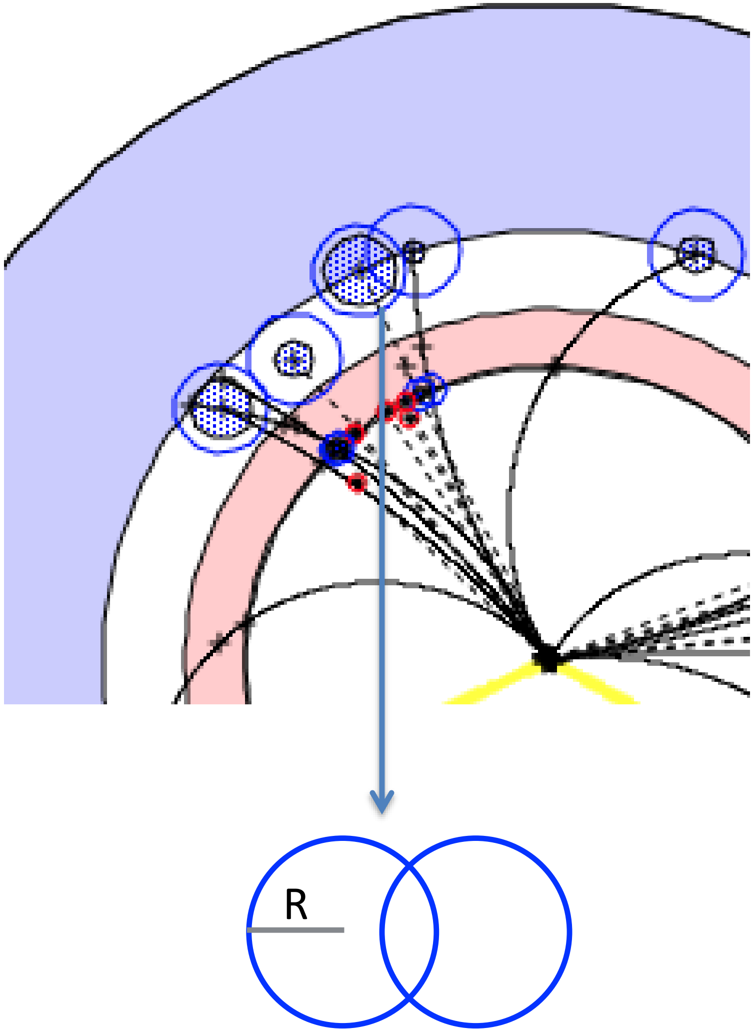
```
class Tracker(DetectorElement):

    def __init__(self):
        volume = VolumeCylinder('tracker', 1.29, 1.99)
        mat = material.void
        super(Tracker, self).__init__('tracker', volume,  mat)

    def acceptance(self, track):
        pt = track.p3() .Pt()
        eta = abs(track.p3() .Eta())
        if eta < 1.35 and pt>0.5:
            return random.uniform(0,1)<0.95
        elif eta < 2.5 and pt>0.5:
            return random.uniform(0,1)<0.9
        else:
            return False

    def resolution(self, track):
        return 1.1e-2
```

- Define:
  - simple geometry (cylinder)
  - acceptance model
  - resolution model
  - (+ B field)

Easy to create / easy to change
Python script → extreme flexibility w/r to a « card » system

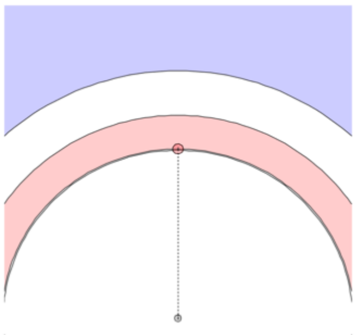
Papas: Calorimeters

- Simple geometry (2 cylinders)
- Material
  - hadron shower in ECAL
- Energy resolution and response
- Acceptance
  - thresholds
- Cluster size R
  - models calorimeter granularity

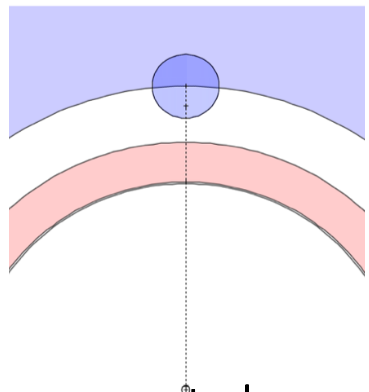


Sum energy and create a merged cluster

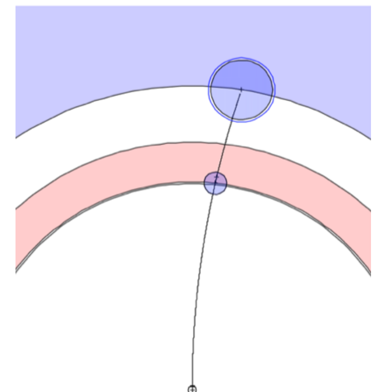
Papas: Particle Flow



photon

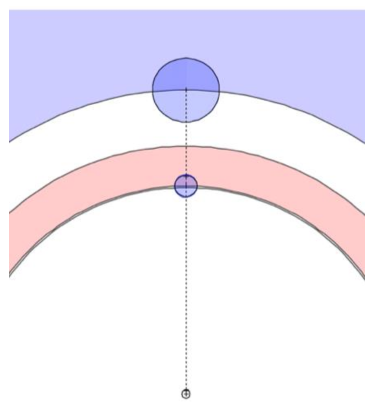


neutral
hadron

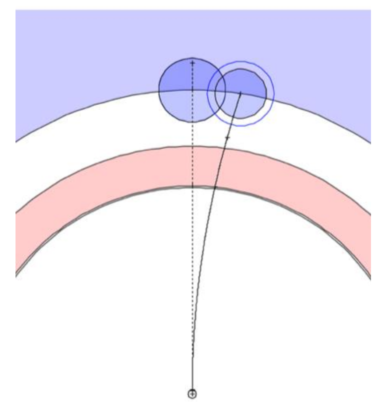


charged
hadron

Full PF algorithm
similar to CMS



neutral hadron
+ photon



charged and
neutral hadrons

Papas: e and mu

- Integration in PF would be very difficult
- Treated separately
  - no lepton / hadron fakes
- Using the Delphes CMS parametrization
- Isolation:
  - particle-based isolation w/r to the surrounding particles from PF

```
def electron_acceptance(self, ptc):
    """Delphes parametrization
    https://github.com/delphes/delphes/blob/master/cards/delphes_card_CMS.tcl
    96d6bcf
    """
    rnd = random.uniform(0, 1)
    if ptc.pt() < 10.:
        return False
    else:
        eta = abs(ptc.eta())
        if eta < 1.5:
            return rnd < 0.95
        elif eta < 2.5:
            return rnd < 0.85
        else:
            return False
```

Papas: Jets

- Fastjet
  - input:
    reconstructed particles
  - ee_kt algorithm
  - exclusive reconstruction
    (e.g. njets=2)
    - different from our previous analyses where an a posteriori resummation was done
- Jet energy correction:
  - flat 1.1 factor
    (~ as in CMS)
- b tagging:
  - input = ROC curve
  - set desired efficiency
    (fixes fake rate)
  - gen b matching: fraction of rec jet energy arising from B hadron > 0.01
    - extremely efficient and pure
  - matched: apply efficiency
  - unmatched: apply fake rate

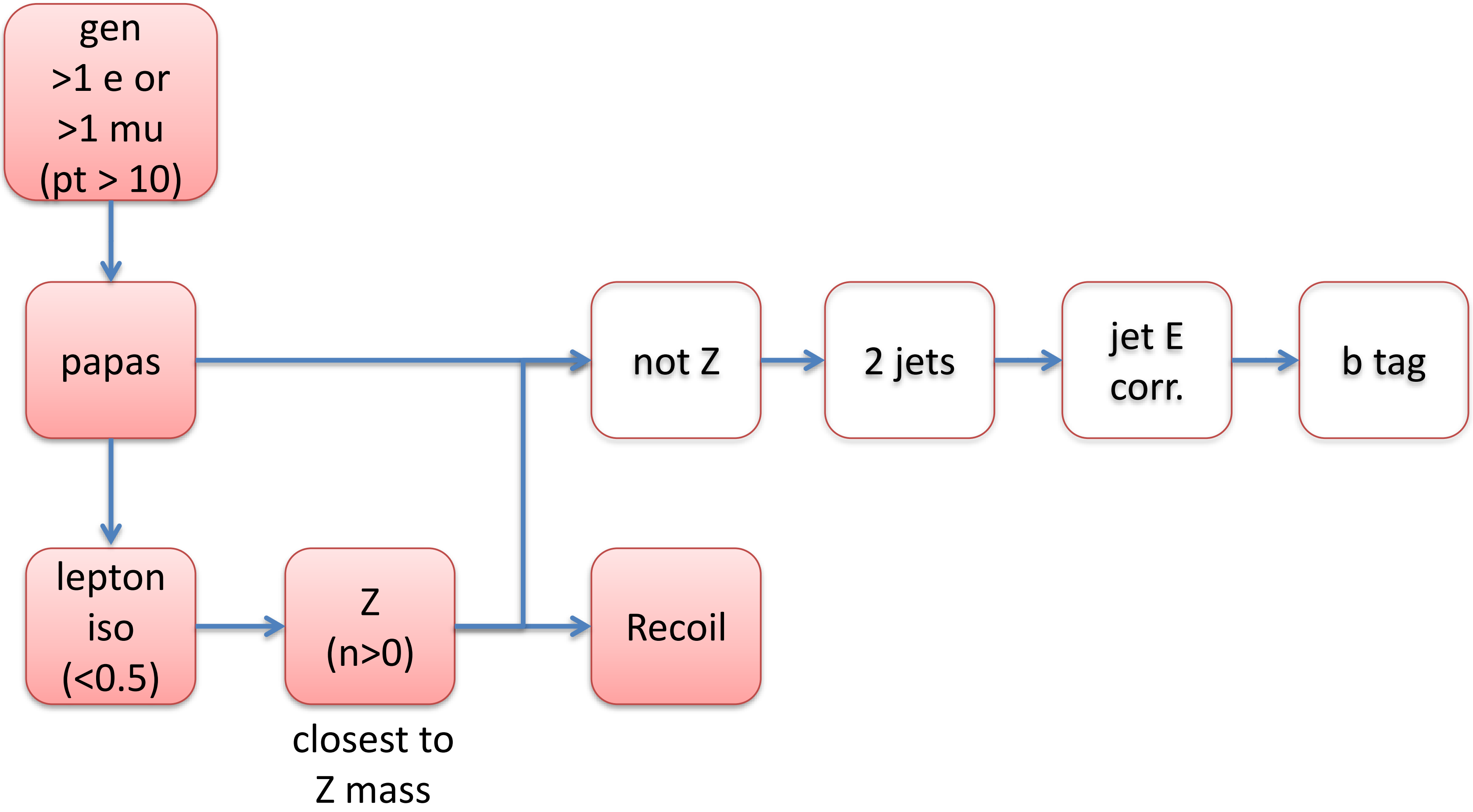
heppy

- Event processing framework
- Modular
  - ~ CMSSW, Gaudi, Athena, Marlin, but much lighter
- Written in python
- Can read from root trees, CMS, FCC, LCIO
- Widely used in CMS
- Many tools:
  - batch processing, physics tools, …

Outline

- Tools:
  - papas
  - heppy analysis framework
- Analyses (vs LEP3 note)
  - ZH→llX, ZH→llbb, ZH→nunubb
  - yield extraction
- Global fit for the couplings (vs TLEP note)
- Summary and plans

ZH→llX: Analysis Sequence



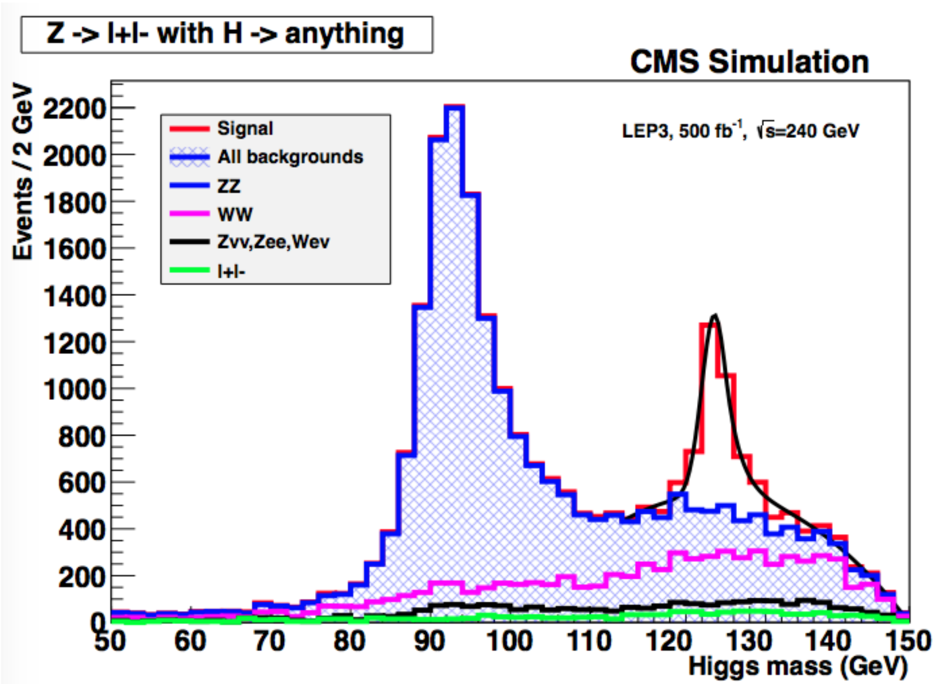
https://github.com/cbernet/fcc-ee-higgs/blob/master/analysis_ee_ZH_llbb.py

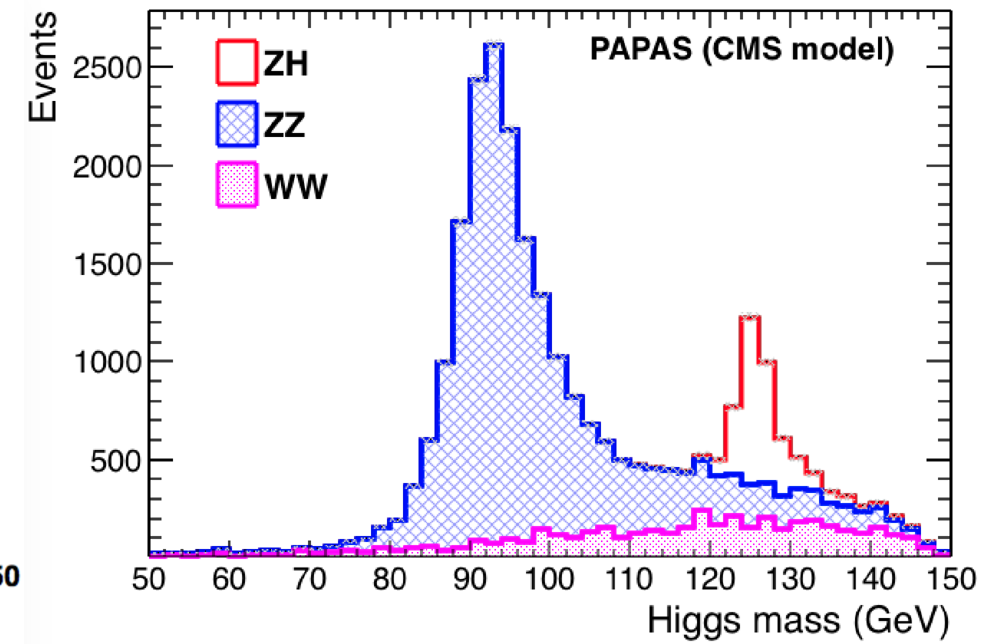
ZH→llX: Final Selection

- Leptons:
  - iso < 0.2, same flavour, opposite charge
- Z candidate
  - |m - 91| < 4 (not 5)
  - $p_T > 10$
  - $p_z < 50$
  - acollinearity > 100
  - cross > 10 (not acoplanarity)
- if jets present, photon fraction < 0.8

The cuts in red are the ones that were used for the LEP3 note. But they differ from what is written therein.
(Thanks P. Janot)

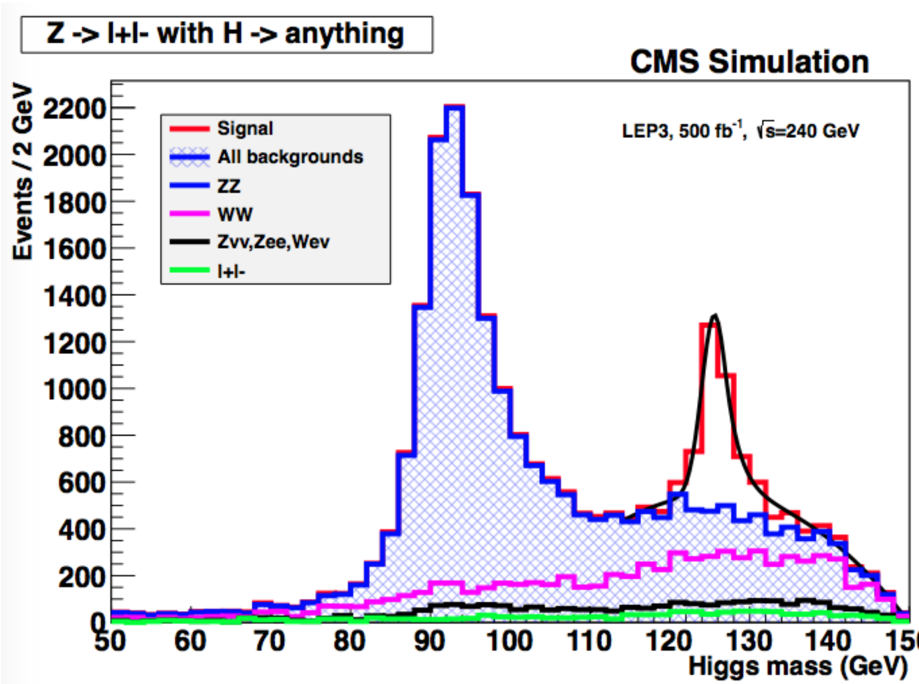
Recoil mass

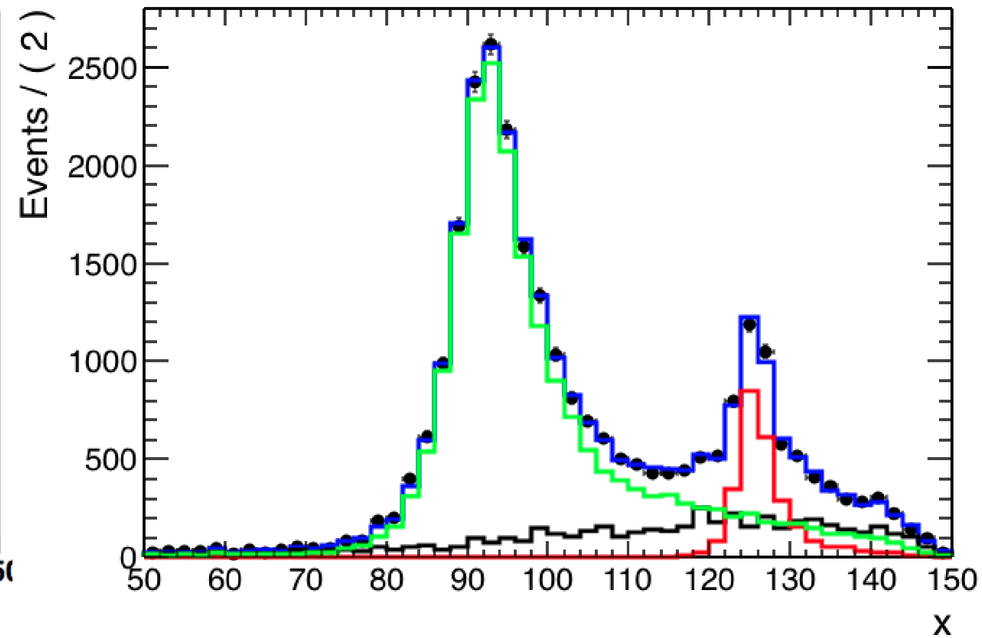




Still missing minor backgrounds
Lepton (electron?) resolution slightly too bad?
10% too many events. lepton isolation too efficient?
Will not matter much anyway

Yield extraction

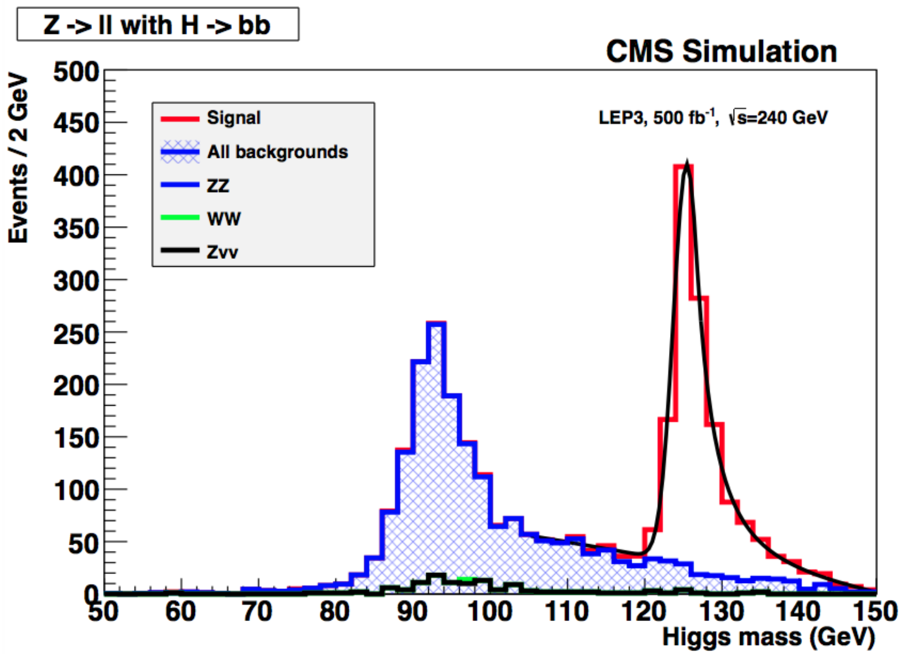


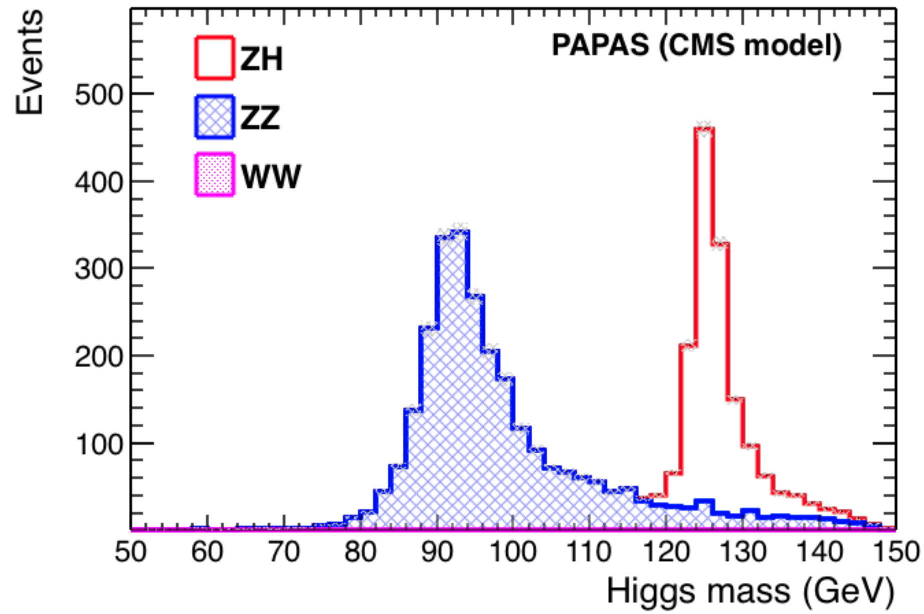


- fit had hoc function
  - yield uncertainty overestimated
    - function might not be adequate
    - limited MC yield

- template fit
  - **use the same templates to generate pseudodata and to fit**
    - both effects eliminated
    - we will have an excellent knowledge of our templates when we take data

ZH→llbb

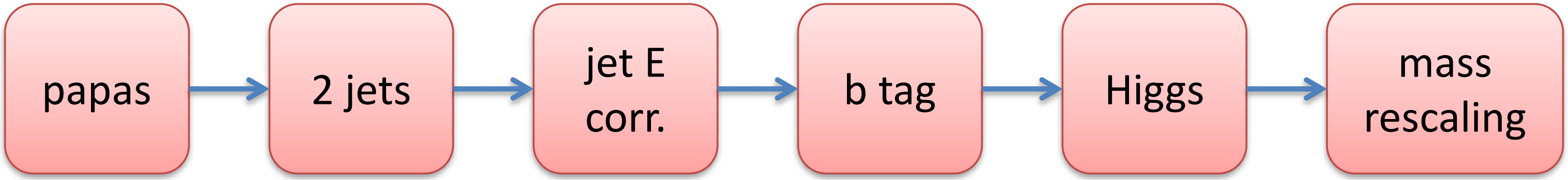




Tuned b tag:
eff = 60%, fake = 0.3%
b1 or b2

Still missing minor backgrounds
Lepton (electron?) resolution slightly too bad?
10% too many events. lepton isolation too efficient?
Will not matter much anyway

ZH→vvbb: Analysis Sequence



- Higgs mass rescaled:
  - jet energy scaled by a common factor by to bring missing mass to mZ
  - 2^{nd} degree equation
  - only in final plot, not for the cuts

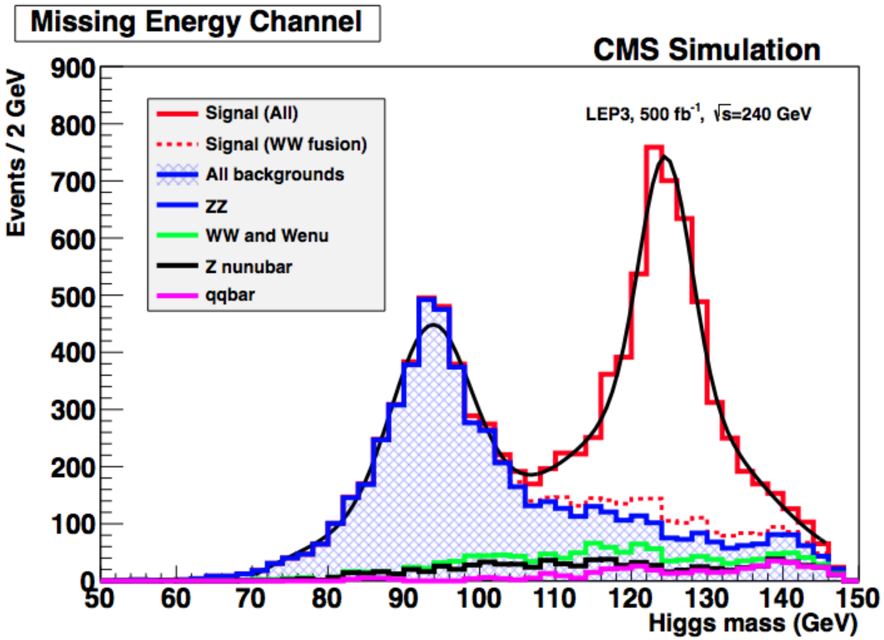
https://github.com/cbernet/fcc-ee-higgs/blob/master/analysis_ee_ZH_llbb.py

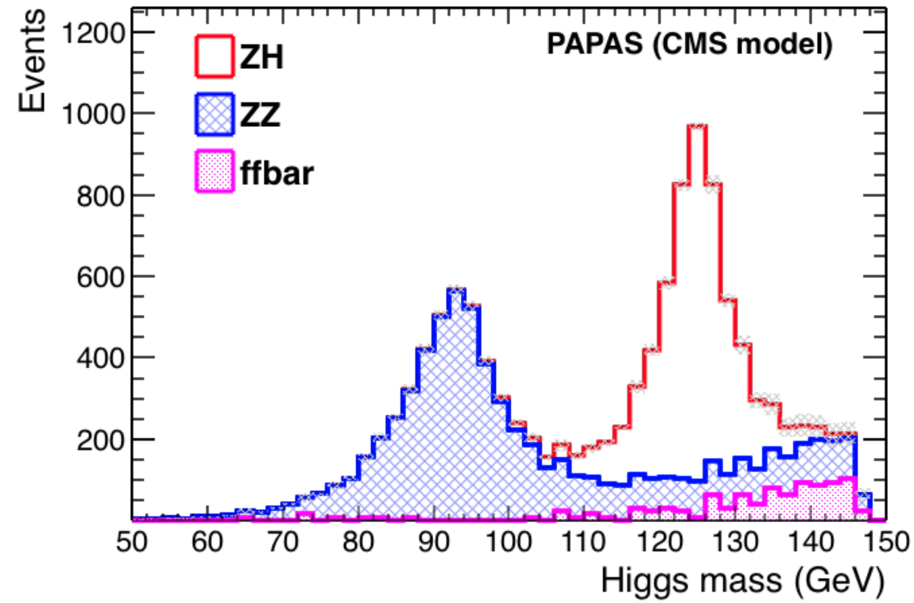
ZH→ννbb: Final Selection

- Missing Z:
  - $65 < m_{miss} < 125$
- Higgs candidate
  - b tag as in ZH→llbb:
    - eff = 60%, fake = 0.3%
    - b1 or b2
  - $p_T > 10$
  - $p_z < 50$
  - acollinearity > 100
  - cross > 10 (not acoplanarity)

The cut in red is the one that was used for the LEP3 note. But it differs from what is written therein.

ZH→ννbb

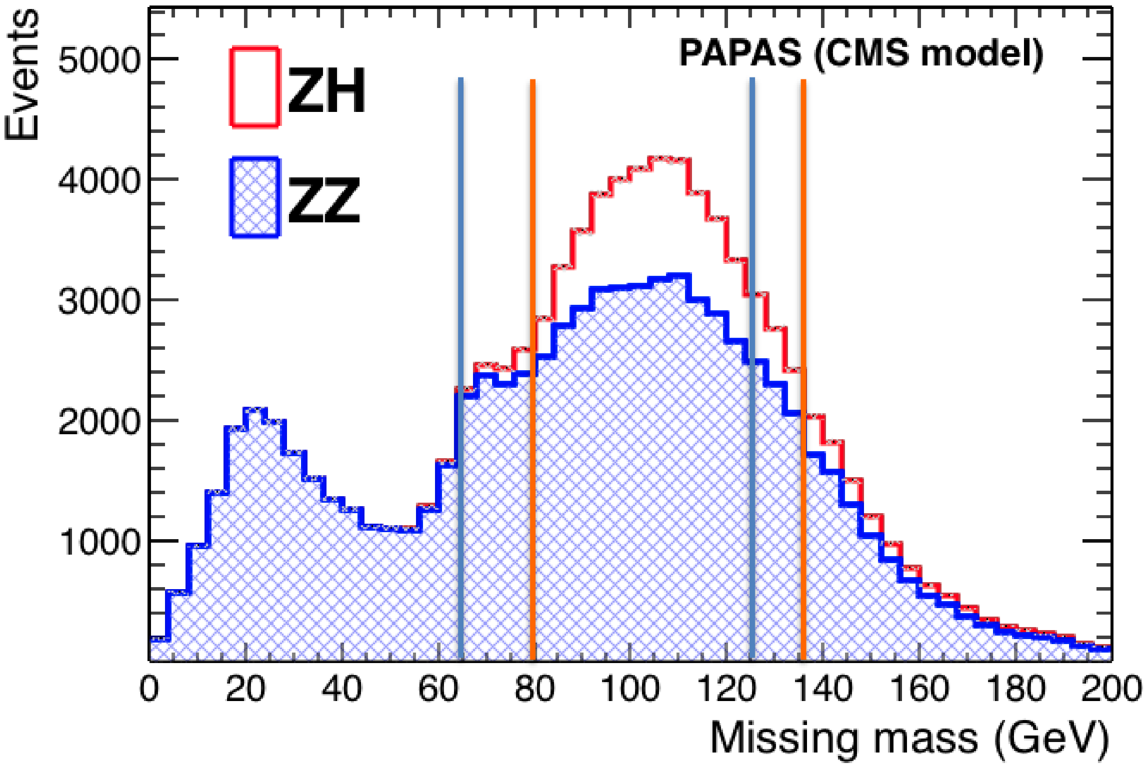




Too many events at high mass (see next slide)

Resolution matches, indicates that the calorimeter model and the particle flow in papas are fine.

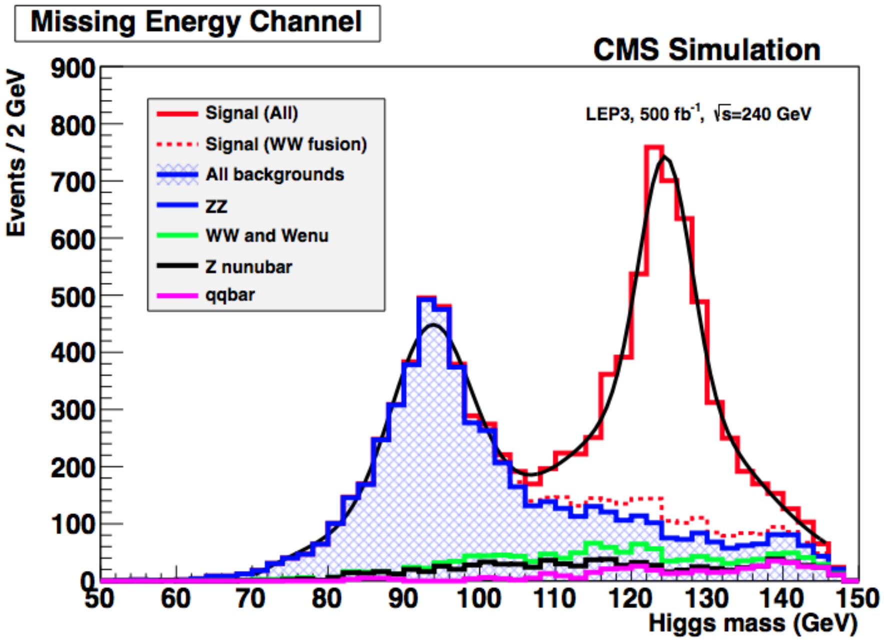
ZH→ννbb: missing mass

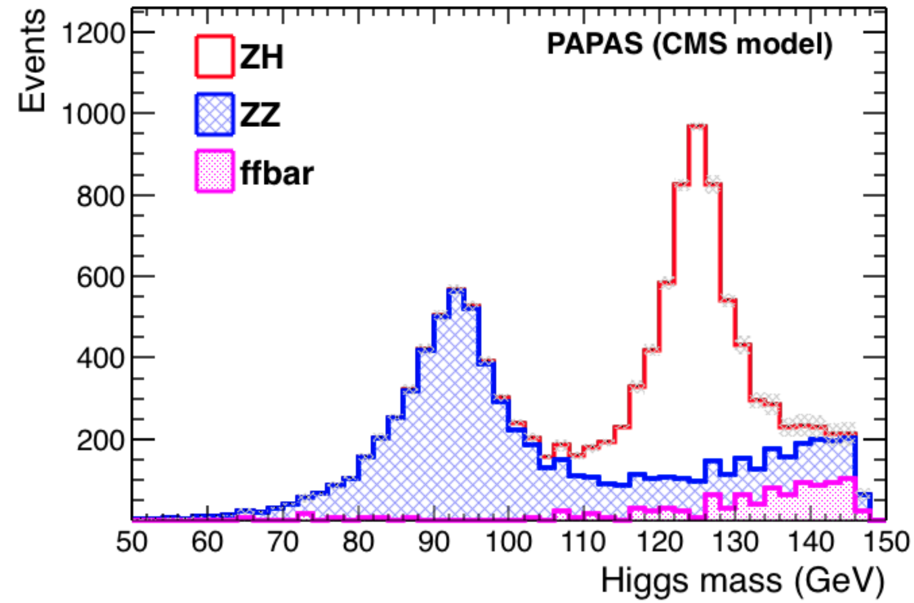


Low missing mass
→ high higgs mass

- missing Z mass too large
  - jet energy too low
  - jet energy calibration to be reviewed.
- Poor man reoptimization:
  - 80 < m < 135

ZH→ννbb

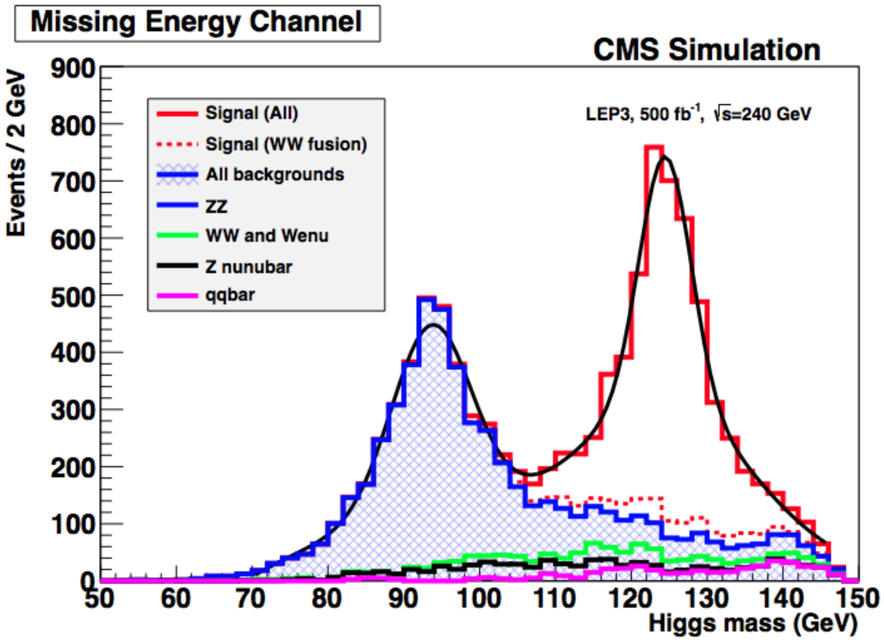


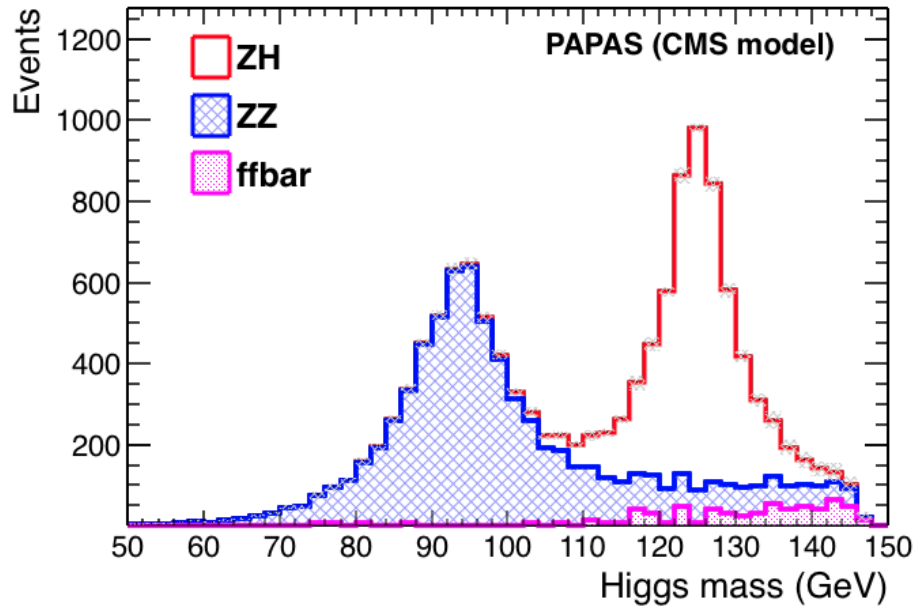


Too many events at high mass (see next slide)

Resolution matches, indicates that the calorimeter model and the particle flow in papas are fine.

ZH→ννbb: missing mass: new cuts





Excess at high mass mostly disappears.

But 20% too many events overall.
Do proper jet energy correction and review…

Outline

- Tools:
  - papas
  - heppy analysis framework
- Analyses (vs LEP3 note)
  - ZH→llX, ZH→llbb, ZH→nunubb
  - yield extraction
- Global fit for the couplings (vs TLEP note)
- Summary and plans

Global Fit

- Concept from M. Peskin
  - https://arxiv.org/abs/1312.4974
- Our implementation (M. Bachtis)
  - https://github.com/bachtis/tlep-couplings
  - used for the TLEP paper
  - fitting code checked, looks correct to me
- Rewritten to add goodness of fit tests:
  - https://github.com/cbernet/tlep-couplings
  - exact same results

Global Fit: kappa framework

$$\kappa_j \equiv g_j / g_j^{SM}$$

coupling deviation

Deviation of the observed yield w/r SM

$$\frac{\sigma_i \times \mathrm{BR}_f}{(\sigma_i \times \mathrm{BR}_f)_{SM}} = \frac{\kappa_i^2 \kappa_f^2}{\kappa_H^2}$$

$$\kappa_H^2 \equiv \Gamma / \Gamma_{SM} = \sum_j \kappa_j^2 \mathrm{BR}_{j,SM}$$

Full width deviation assuming no other mode

- Measure a set of yields in various channels:
  - i : ZH, WWH
  - f : bb, ττ, …
- Fit to find the best values of $\kappa_j, \mathrm{BR}_{\mathrm{inv}}$

$$\kappa_H^2 \equiv \Gamma / \Gamma_{SM} = \frac{\sum_j \kappa_j^2 \mathrm{BR}_{j,SM}}{1 - \mathrm{BR}_{\mathrm{inv}}}$$

Full width deviation allowing for invisible decay modes

More precisely

Fitting additive parameters instead of κ_j

yield ratio w/r SM set to 1

yield uncertainty (here at 240 GeV)

accounts for the fact that the 240 GeV measurement will also be done at 350 GeV

```
f350 = 0.95
f.addConstraint('Zh','(1+Z)*(1+Z)','Z',1,0.004*f350)
# f.addConstraint('Wh','(1+W)*(1+W)','W',1,0.004)
# f.addConstraint('Wh250','(1+W)*(1+W)','W',1,0.02)
f.addConstraint('Whbb240','(1+W)*(1+W)*(1+b)*(1+b)/width','W,b,width',1,0.022)
f.addConstraint('Whbb350','(1+W)*(1+W)*(1+b)*(1+b)/width','W,b,width',1,0.006)
f.addConstraint('Zhbb','(1+Z)*(1+Z)*(1+b)*(1+b)/width','Z,b,width',1,0.002*f350)
```

κ_H^2

Gaussian pdf for all **yields** (σ = uncertainty)
Likelihood = product of the pdfs
The fit varies the **parameters of interests** $\kappa_j, \mathrm{BR}_{\mathrm{inv}}$
Uncertainty on the POIs finally taken from the fit

Trying the TLEP model

	Model-independent fit				Constrained fit			
Coupling	TLEP-240	TLEP		ILC	TLEP		ILC	
g_{HZZ}	0.16%	**0.15%**	(0.18%)	0.9%	0.05%	(0.06%)	0.31%	0.148
g_{HWW}	0.85%	**0.19%**	(0.23%)	0.5%	0.09%	(0.11%)	0.25%	0.195
g_{Hbb}	0.88%	**0.42%**	(0.52%)	2.4%	0.19%	(0.23%)	0.85%	0.442
g_{Hcc}	1.0%	**0.71%**	(0.87%)	3.8%	0.68%	(0.84%)	3.5%	0.717
g_{Hgg}	1.1%	**0.80%**	(0.98%)	4.4%	0.79%	(0.97%)	4.4%	0.802
$g_{H\tau\tau}$	0.94%	**0.54%**	(0.66%)	2.9%	0.49%	(0.60%)	2.6%	0.546
$g_{H\mu\mu}$	6.4%	**6.2%**	(7.6%)	45%	6.2%	(7.6%)	45%	6.202
$g_{H\gamma\gamma}$	1.7%	**1.5%**	(1.8%)	14.5%	1.4%	(1.7%)	14.5%	1.49
BR_{exo}	0.48%	**0.45%**	(0.55%)	2.9%	0.16%	(0.20%)	0.9%	**6.74**

This should read BR_{inv}

Can reproduce all values except BRexo (factor 10 difference)

Check fit model

```
###Here add the constraints 'name','formula','dependents',mean value ,error
################################################
f.addConstraint('Zh','(1+Z)*(1+Z)','Z',1,0.004*0.95)
f.addConstraint('Wh','(1+W)*(1+W)','W',1,0.004)
f.addConstraint('Wh250','(1+W)*(1+W)','W',1,0.02)
f.addConstraint('Zhbb','(1+Z)*(1+Z)*(1+b)*(1+b)/width','Z,b,width',1,0.002*0.95)
f.addConstraint('Zhcc','(1+Z)*(1+Z)*(1+c)*(1+c)/width','Z,c,width',1,0.012*0.95)
f.addConstraint('Zhgg','(1+Z)*(1+Z)*(1+g)*(1+g)/width','Z,g,width',1,0.014*0.95)
f.addConstraint('ZhWW','(1+Z)*(1+Z)*(1+W)*(1+W)/width','Z,W,width',1,0.009*0.95)
f.addConstraint('ZhZZ','(1+Z)*(1+Z)*(1+Z)*(1+Z)/width','Z,width',1,0.031*0.95)
f.addConstraint('Zhtautau','(1+Z)*(1+Z)*(1+tau)*(1+tau)/width','Z,tau,width',1,0.007*0.95)
f.addConstraint('Zhgammagamma','(1+Z)*(1+Z)*(1+gamma)*(1+gamma)/width','Z,gamma,width',1,0.03*0.95)
f.addConstraint('Zhmumu','(1+Z)*(1+Z)*(1+mu)*(1+mu)/width','Z,mu,width',1,0.13*0.95)
f.addUniformConstraint('Zhinv','inv','inv') ####->Means free floating
```

There is no inclusive WW→H measurement in the LEP3 / TLEP notes

Otherwise correct

https://github.com/bachtis/tlep-couplings/blob/master/runTLEP_250_350_Standalone_Floating.py#L21

Fixing fit model

```
f350 = 0.95
f.addConstraint('Zh','(1+Z)*(1+Z)','Z',1,0.004*f350)
# f.addConstraint('Wh','(1+W)*(1+W)','W',1,0.004)
# f.addConstraint('Wh250','(1+W)*(1+W)','W',1,0.02)
f.addConstraint('Whbb240','(1+W)*(1+W)*(1+b)*(1+b)/width','W,b,width',1,0.022)
f.addConstraint('Whbb350','(1+W)*(1+W)*(1+b)*(1+b)/width','W,b,width',1,0.006)
f.addConstraint('Zhbb','(1+Z)*(1+Z)*(1+b)*(1+b)/width','Z,b,width',1,0.002*f350)
f.addConstraint('Zhcc','(1+Z)*(1+Z)*(1+c)*(1+c)/width','Z,c,width',1,0.012*f350)
f.addConstraint('Zhgg','(1+Z)*(1+Z)*(1+g)*(1+g)/width','Z,g,width',1,0.014*f350)
f.addConstraint('ZhWW','(1+Z)*(1+Z)*(1+W)*(1+W)/width','Z,W,width',1,0.009*f350)
f.addConstraint('Zhtautau','(1+Z)*(1+Z)*(1+tau)*(1+tau)/width','Z,tau,width',1,0.007*f350)
f.addConstraint('ZhZZ','(1+Z)*(1+Z)*(1+Z)*(1+Z)/width','Z,width',1,0.031*f350)
f.addConstraint('Zhgammagamma','(1+Z)*(1+Z)*(1+gamma)*(1+gamma)/width','Z,gamma,width',1,0.03*f350)
f.addConstraint('Zhmumu','(1+Z)*(1+Z)*(1+mu)*(1+mu)/width','Z,mu,width',1,0.13*f350)
f.addUniformConstraint('Zhinv','inv') ####->Means free floating
```

Table 6: Statistical precision of the TLEP measurement of $\sigma_{WW \to H} \times BR(H \to b\bar{b})$. For illustration, the ILC potential at the same centre-of-mass energies is also indicated.

$\sqrt{s}$ (GeV)	TLEP	ILC
240 - 250	**2.2%**	10.5%
350	**0.6%**	1.0%

After the fix

	Model-independent fit				Constrained fit				
Coupling	TLEP-240	TLEP		ILC	TLEP		ILC		
g_{HZZ}	0.16%	**0.15%**	(0.18%)	0.9%	0.05%	(0.06%)	0.31%	0.148	0.149
g_{HWW}	0.85%	**0.19%**	(0.23%)	0.5%	0.09%	(0.11%)	0.25%	0.195	**0.329**
g_{Hbb}	0.88%	**0.42%**	(0.52%)	2.4%	0.19%	(0.23%)	0.85%	0.442	0.463
g_{Hcc}	1.0%	**0.71%**	(0.87%)	3.8%	0.68%	(0.84%)	3.5%	0.717	0.745
g_{Hgg}	1.1%	**0.80%**	(0.98%)	4.4%	0.79%	(0.97%)	4.4%	0.802	0.840
$g_{H\tau\tau}$	0.94%	**0.54%**	(0.66%)	2.9%	0.49%	(0.60%)	2.6%	0.546	0.580
$g_{H\mu\mu}$	6.4%	**6.2%**	(7.6%)	45%	6.2%	(7.6%)	45%	6.202	6.206
$g_{H\gamma\gamma}$	1.7%	**1.5%**	(1.8%)	14.5%	1.4%	(1.7%)	14.5%	1.49	1.504
BR_{exo}	0.48%	**0.45%**	(0.55%)	2.9%	0.16%	(0.20%)	0.9%	6.74	6.77

Main change: uncertainty on W coupling almost x2

Including direct constraint on $\mathrm{BR}_{\mathrm{inv}}$

Bug found: the fit was defining: $$\kappa_H^2 \equiv \Gamma/\Gamma_{SM} = \frac{\sum_j \kappa_j^2 \mathrm{BR}_{j,SM}}{1 - \mathrm{BR}_{\mathrm{inv}}^2}$$

Instead of: $$\kappa_H^2 \equiv \Gamma/\Gamma_{SM} = \frac{\sum_j \kappa_j^2 \mathrm{BR}_{j,SM}}{1 - \mathrm{BR}_{\mathrm{inv}}}$$

Constraint model for BRinv:

- Gaussian centred on 0, σ = 0.3% (P. Janot)
- Allow values between 0 and 1%

Including direct constraint on BR_{inv}

	Model-independent fit				Constrained fit		
Coupling	TLEP-240	TLEP		ILC	TLEP		ILC
g_{HZZ}	0.16%	**0.15%**	(0.18%)	0.9%	0.05%	(0.06%)	0.31%
g_{HWW}	0.85%	**0.19%**	(0.23%)	0.5%	0.09%	(0.11%)	0.25%
g_{Hbb}	0.88%	**0.42%**	(0.52%)	2.4%	0.19%	(0.23%)	0.85%
g_{Hcc}	1.0%	**0.71%**	(0.87%)	3.8%	0.68%	(0.84%)	3.5%
g_{Hgg}	1.1%	**0.80%**	(0.98%)	4.4%	0.79%	(0.97%)	4.4%
$g_{H\tau\tau}$	0.94%	**0.54%**	(0.66%)	2.9%	0.49%	(0.60%)	2.6%
$g_{H\mu\mu}$	6.4%	**6.2%**	(7.6%)	45%	6.2%	(7.6%)	45%
$g_{H\gamma\gamma}$	1.7%	**1.5%**	(1.8%)	14.5%	1.4%	(1.7%)	14.5%
BR_{exo}	0.48%	**0.45%**	(0.55%)	2.9%	0.16%	(0.20%)	0.9%

0.149
0.329
0.463
0.745
0.840
0.580
6.206
1.504
0.46

BRinv now correct!
Actually, the direct constraints are not needed.
The bug fix is enough to bring BRinv to the correct value.
Mike had certainly found the bug but had not committed.

Summary and plans

- Preparatory work:
  - 3 analyses done:
    - ZH→llX, ZH→llbb, ZH→ννbb
  - papas and global fit validated
    - a few details to be checked (jet E correction)
  - many analyses still uncovered
    - ZH→ττ, ZH→WW, ZH→γγ, WWH→bb, …
      - with many subchannels
  - tools ready, let's team up, happy to help

- Do the study
  - update running scenarios
  - use other detector models
    - CLIC-ILD ready in papas, parameter scan
  - reoptimize analyses for each detector model

Thanks to Michalis Bachtis, Patrick Janot, and Janik von Ahnen

Backup

« 0p6 » model, also available in tlep-couplings

```
###Here add the constraints 'name','formula','dependents',mean value ,error
################################################
f.addConstraint('Zh','(1+Z)*(1+Z)','Z',1,0.004*0.95)
f.addConstraint('Whbb350','(1+W)*(1+W)*(1+b)*(1+b)/width','W,b,width',1,0.006)
f.addConstraint('Whbb','(1+W)*(1+W)*(1+b)*(1+b)/width','W,b,width',1,0.022*0.95)
f.addConstraint('WhWW350','(1+W)*(1+W)*(1+W)*(1+W)/width','W,width',1,0.024)
f.addConstraint('Zhbb','(1+Z)*(1+Z)*(1+b)*(1+b)/width','Z,b,width',1,0.002*0.95)
f.addConstraint('Zhcc','(1+Z)*(1+Z)*(1+c)*(1+c)/width','Z,c,width',1,0.012*0.95)
f.addConstraint('Zhgg','(1+Z)*(1+Z)*(1+g)*(1+g)/width','Z,g,width',1,0.014*0.95)
f.addConstraint('ZhWW','(1+Z)*(1+Z)*(1+W)*(1+W)/width','Z,W,width',1,0.009*0.95)
f.addConstraint('ZhZZ','(1+Z)*(1+Z)*(1+Z)*(1+Z)/width','Z,width',1,0.031*0.95)
f.addConstraint('Zhtautau','(1+Z)*(1+Z)*(1+tau)*(1+tau)/width','Z,tau,width',1,0.007*0.95)
f.addConstraint('Zhgammagamma','(1+Z)*(1+Z)*(1+gamma)*(1+gamma)/width','Z,gamma,width',1,0.03*0.95)
f.addConstraint('Zhmumu','(1+Z)*(1+Z)*(1+mu)*(1+mu)/width','Z,mu,width',1,0.13*0.95)
f.addUniformConstraint('Zhinv','inv','inv') ####->Means free floating
```

should not be there since Whbb350 is accounted for

analysis not mentioned anywhere?

https://github.com/bachtis/tlep-couplings/blob/master/runTLEP_250_350_Standalone_Floating_with_0p6.py

« 0p6 » model

	Model-independent fit			
Coupling	TLEP-240	TLEP		ILC
g_{HZZ}	0.16%	**0.15%**	(0.18%)	0.9%
g_{HWW}	0.85%	**0.19%**	(0.23%)	0.5%
g_{Hbb}	0.88%	**0.42%**	(0.52%)	2.4%
g_{Hcc}	1.0%	**0.71%**	(0.87%)	3.8%
g_{Hgg}	1.1%	**0.80%**	(0.98%)	4.4%
$g_{H\tau\tau}$	0.94%	**0.54%**	(0.66%)	2.9%
$g_{H\mu\mu}$	6.4%	**6.2%**	(7.6%)	45%
$g_{H\gamma\gamma}$	1.7%	**1.5%**	(1.8%)	14.5%
BR_{exo}	0.48%	**0.45%**	(0.55%)	2.9%

TLEP model	**my fix**	
0.148	0.149	0.148
0.195	**0.329**	**0.317**
0.442	0.463	0.460
0.717	0.745	0.744
0.802	0.840	0.838
0.546	0.580	0.578
6.202	6.206	6.206
1.49	1.504	1.485
3.369	**3.387**	**3.376**

The other model is close to my fix (but is not understood)
It is not the one used for the TLEP paper either
BRexo still 10 times larger
Uncertainty on W coupling almost x2

ILC (250 fb-1, 250 GeV)

	Model-independent fit			
Coupling	TLEP-240	TLEP		ILC
g_{HZZ}	0.16%	**0.15%**	(0.18%)	0.9%
g_{HWW}	0.85%	**0.19%**	(0.23%)	0.5%
g_{Hbb}	0.88%	**0.42%**	(0.52%)	2.4%
g_{Hcc}	1.0%	**0.71%**	(0.87%)	3.8%
g_{Hgg}	1.1%	**0.80%**	(0.98%)	4.4%
$g_{H\tau\tau}$	0.94%	**0.54%**	(0.66%)	2.9%
$g_{H\mu\mu}$	6.4%	**6.2%**	(7.6%)	45%
$g_{H\gamma\gamma}$	1.7%	**1.5%**	(1.8%)	14.5%
BR_{exo}	0.48%	**0.45%**	(0.55%)	2.9%

TLEP model	**my fix**	**0p6 TLEP model**	**ILC (Peskin)**
0.148	0.149	0.148	0.78
0.195	**0.329**	**0.317**	4.6
0.442	0.463	0.460	4.7
0.717	0.745	0.744	6.4
0.802	0.840	0.838	6.1
0.546	0.580	0.578	5.2
6.202	6.206	6.206	-
1.49	1.504	1.485	18.8
3.369	**3.387**	**3.376**	**0.54**

https://arxiv.org/abs/1312.4974 table 5
not directly comparable: twice smaller lumi, 250 instead of 240 GeV
BR inv < 1 % !