Higgs at CMS with 1, 10, 30 fb^{-1}

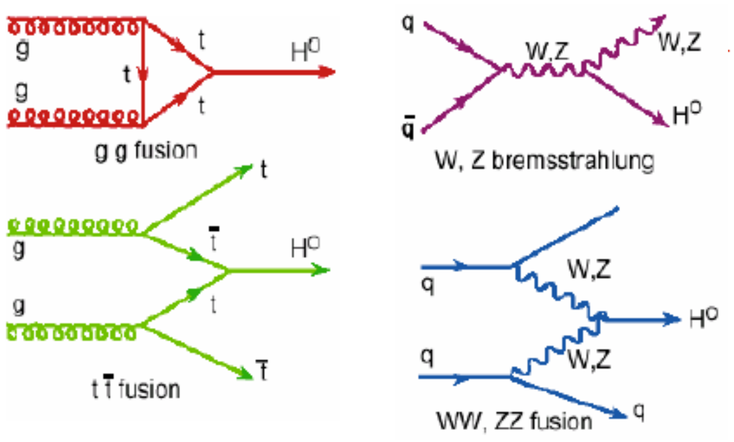
LCWS – ILC

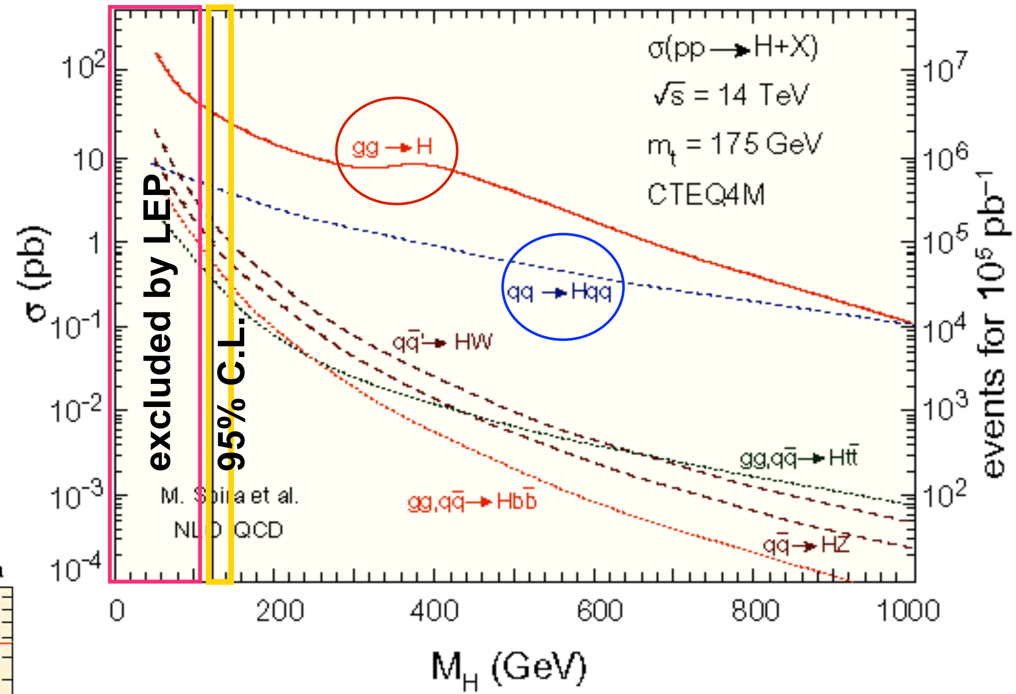
2007 INTERNATIONAL LINEAR COLLIDER WORKSHOP

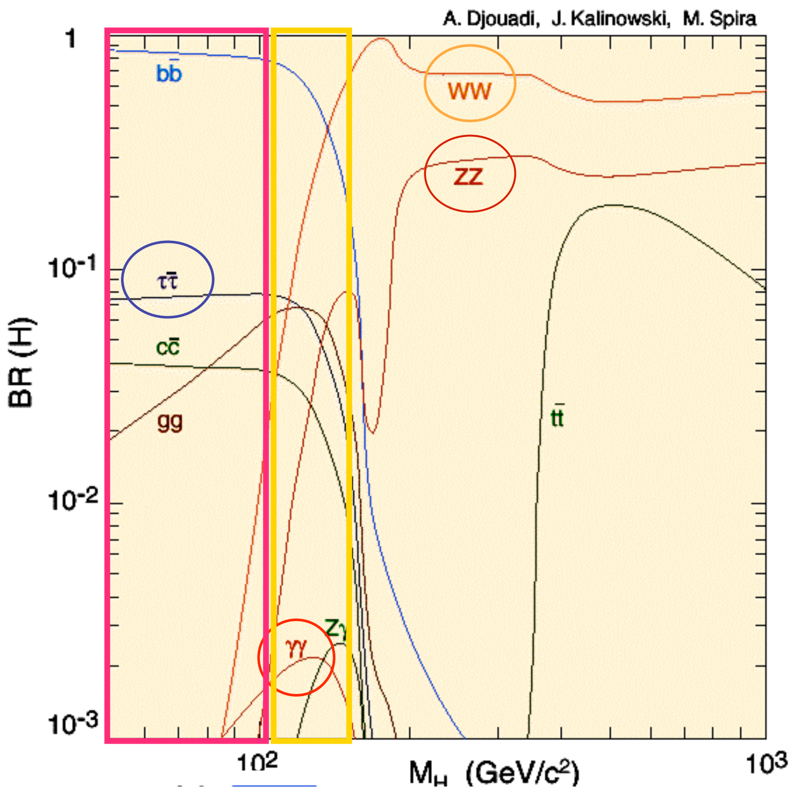
30 May – 3 June 2007 @ DESY

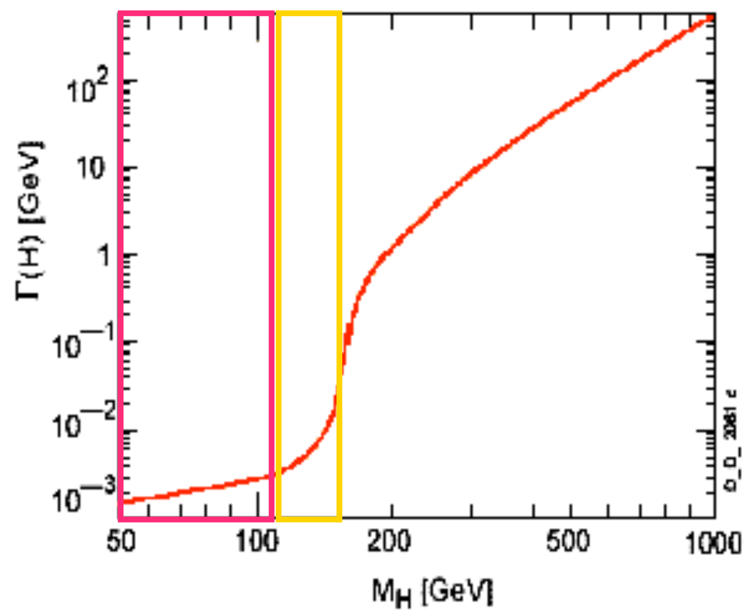
Worldwide Study of the Physics and Detectors for Future Linear e^+e^- Colliders

S. Bolognesi – INFN Torino


g
g
t
t
t
H0
g g fusion
q
q̄
W,Z
W,Z
H0
W, Z bremsstrahlung
g
g
t
t̄
t
t̄
H0
t t̄ fusion
q
q
W,Z
W,Z
H0
q
WW, ZZ fusion


σ(pp→H+X)
√s = 14 TeV
mt = 175 GeV
CTEQ4M
gg→H
qq→Hqq
qq̄'→HW
gg,qq̄→Htt̄
gg,qq̄→Hbb̄
qq̄→HZ
excluded by LEP
95% C.L.
M. Spira et al.
NLO QCD
σ (pb)
events for 10^5 pb^-1
M_H (GeV)


A. Djouadi, J. Kalinowski, M. Spira
bb̄
ττ
cc̄
gg
WW
ZZ
tt̄
γγ
Zγ
BR (H)
M_H (GeV/c²)


Γ(H) [GeV]
M_H [GeV]

- Higgs production xsec (mainly **gg fusion and VBF**) decreases with the increasing of Higgs mass
- Higgs width increases with its mass

Roughly speaking, the difficulty of Higgs detection increases with its mass so at low lumi (1,10,30 fb^{-1}) you will see low masses…

… except for very low masses (M_H<130 GeV) where the Higgs decay channels are a big experimental challenge !!

- Higgs couples to heaviest available fermion **(b,τ)** ...

  … until **WW, ZZ** thresholds open.

  H → γγ the favourite (one loop) decay in a very unlucky region
- Due to large QCD backgrounds
  - (quite) no hope to trigger/extract fully hadronic final state

    e.g. $\sigma(H\rightarrow bb) \approx 20$ pb

    $\sigma(bb) \approx 500\ \mu b$
  - look for final state with lepton (or γ) or associated production (very low xsec)

Main Higgs channels

- **$H\to ZZ^*\to 4l$**
- **$H\to WW^*\to l\nu l\nu$**

} early discovery channels
measure Higgs properties (mass, width, xsec) already with 30 fb^{-1} !!

- **$H\to WW^*\to jjl\nu$ / $l\nu l\nu$ in VBF**
- **$H\to\tau\tau$ in VBF**

} significance > 5(3) with 30 fb^{-1}
but good comprehension of detector needed (jet, MET, τ in lept. and hadr. decay)

- **$H\to\gamma\gamma$** very difficult analysis with still quite unpredictable background
- **$ttH\to ttbb$** at least 60 fb^{-1} (many jets also with low p_T (<30 GeV) → bad reso/eff)
- other channels (mainly **associated production**) can help EXCLUDING Higgs (e.g. $WH\to WWW^*\to Wl\nu l\nu$)

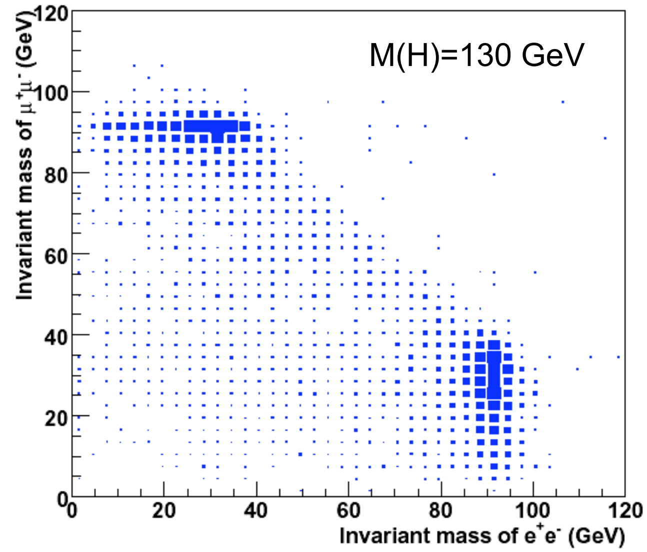
	channel	$O(\sigma \times BR)$	studied M_H
	$H\to ZZ^*\to 4l$	5-100 fb	130-500 GeV
	$H\to WW^*\to l\nu l\nu$	0.5-2.5 pb	120-200 GeV
VBF	$H\to WW^*\to jjl\nu$	200-900 fb	120-250 GeV
VBF	$H\to WW^*\to l\nu l\nu$	50-250 fb	120-200 GeV
VBF	$H\to\tau\tau$	50-150 fb	115-145 GeV
	$H\to\gamma\gamma$	50-100 fb	115-150 GeV

❑ **Analysis focusing on**

- improuvement of the reconstruction
- backgr. and syst. from data
- correct statistical treatment of results

$H \rightarrow ZZ^{(*)} \rightarrow 4l$

- very sensible for **M(H) = 130 to 500** (except 150-190 where WW open)
  - early discovery: statistical observation involving a small number of events
  - compatibility with SM expectation:
    preserving the phase space for more involved characterization
    measuring xsec, M_H, width (spin, CP …)



- usual cuts
  - **isolated lepton from primary vertex with high p_T** (trigger)
  - **one on-shell Z** greater than 50% for M(H)>115
    greater than 85% for M(H)>150

- three channels
  - **4μ**: golden channel
  - **2e 2μ**: highest BR but lower reso/effic on electrons
  - **4e**: most difficult (important to recover low p_T electrons)

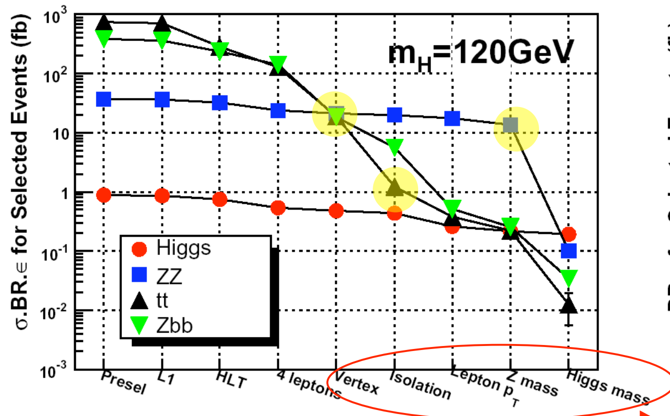
2e2μ analysis

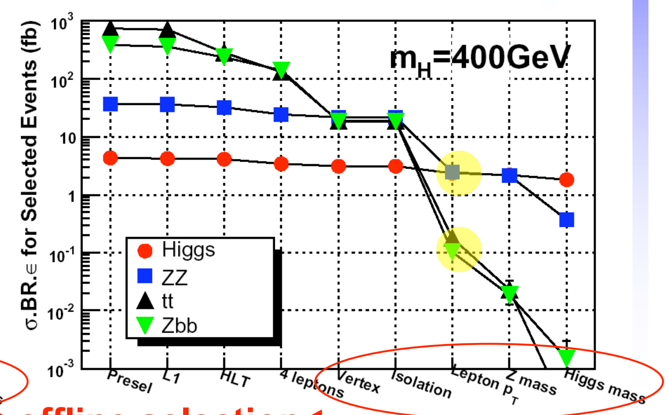
- backgrounds:

  ZZ(*)/γ*, tt, Zbb

  (Zcc found to be negligible)

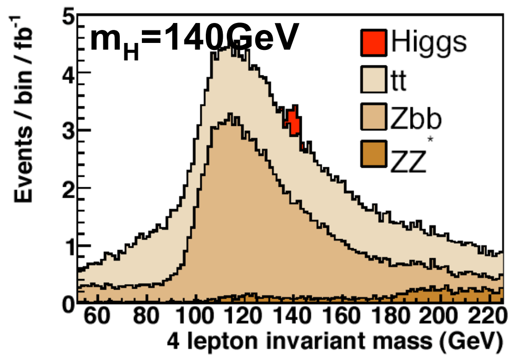
- reconstruction
  - likelihood approach to discriminate real / fake $e^{+/-}$
    - ECAL-Tracker matching, shower shape
    - e^+e^- with highest likelihood selected
  - internal bremsstrahlung recovery:
    - **40%-10%** events with γ (p_T>5 GeV) radiation from lepton (1/3 from μ)
    - recovered γ with **ΔR(γ,l) < 0.3**

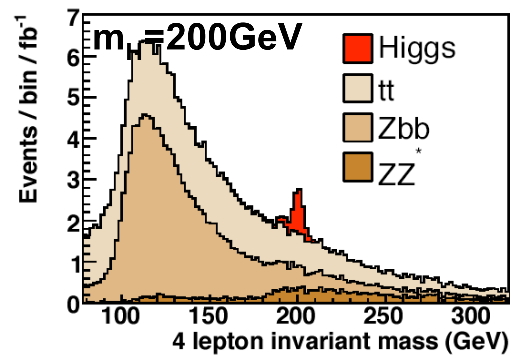




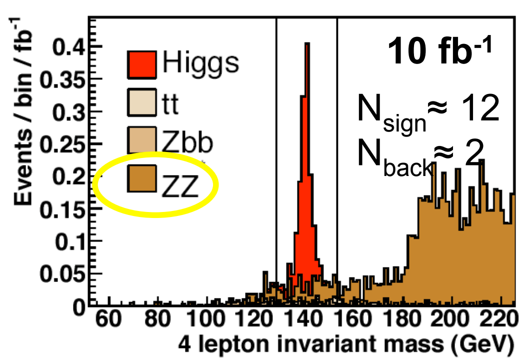
offline selection

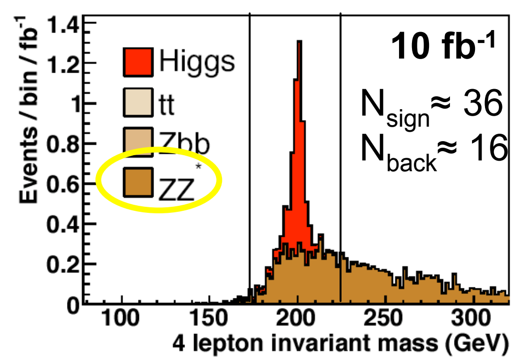
before





after



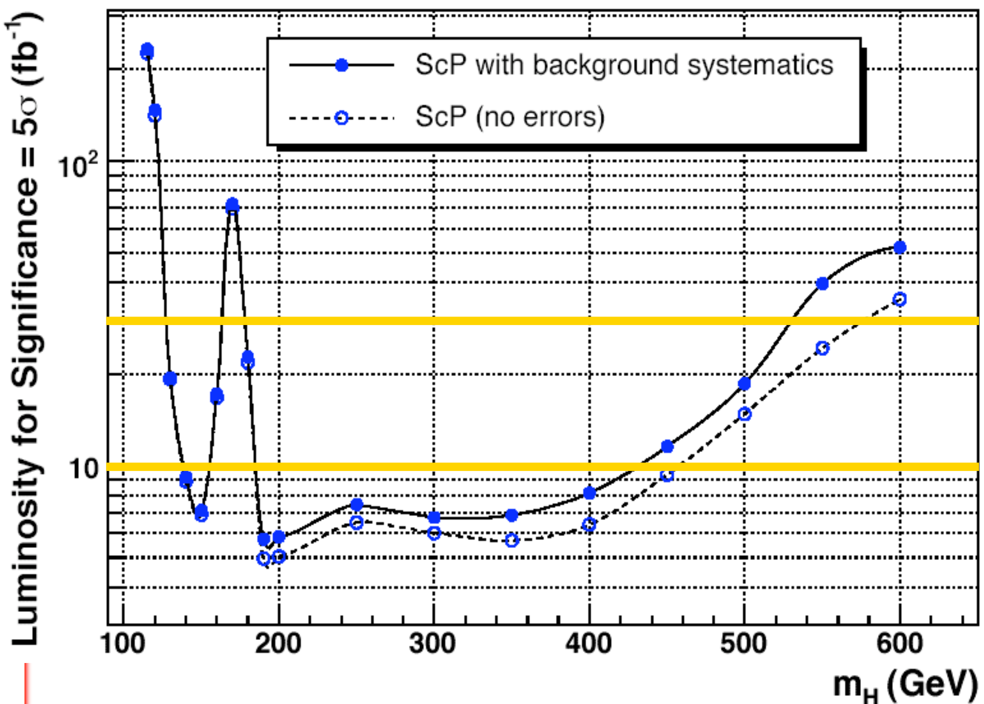


$2e2\mu$ results

- Background normalized from sidebands **$\Delta B = \Delta B$ stat $\oplus$ ΔB theory**

  ΔB stat increases with m_H **from 2% (m_H 120) to 30% (m_H 600)** because of events decreasing in sidebands w.r.t. signal window

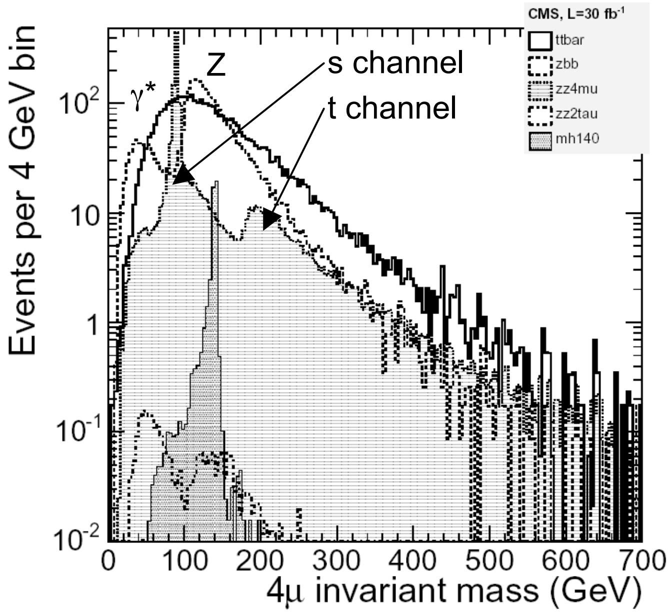
  ΔB theory from PDF, QCD scale, NLO ZZ xsec → **0.5% - 4.5%**



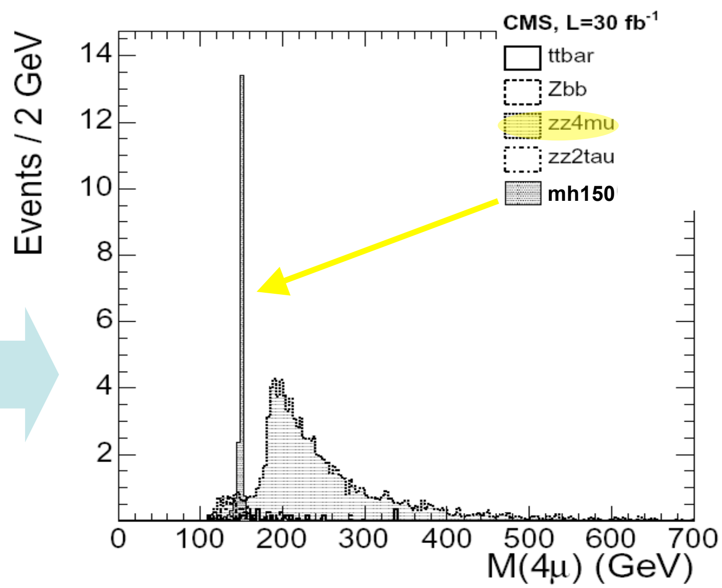
- Luminosity VS m_H **(same shape of 4μ and 4e)**
  - **m_H 150** high BR and low backgrounds
  - **m_H 170** low BR at the H→WW turn on
  - **m_H 200** strong enhancement of BR for $m_H > 2m_Z$
  - **m_H 250** decreasing of signal while ZZ background remains high
  - **m_H 250-350** decreasing of ZZ background
  - **m_H > 350** decreasing of signal xsec and BR (due to H→tt)

4μ analysis

MC generated



Reconstructed M(4μ) after selection



Half of the events used to **optimize cuts with GARCON*** which allows to obtain smooth M(4μ) dependent cuts:

three main critical cuts uncorrelated:

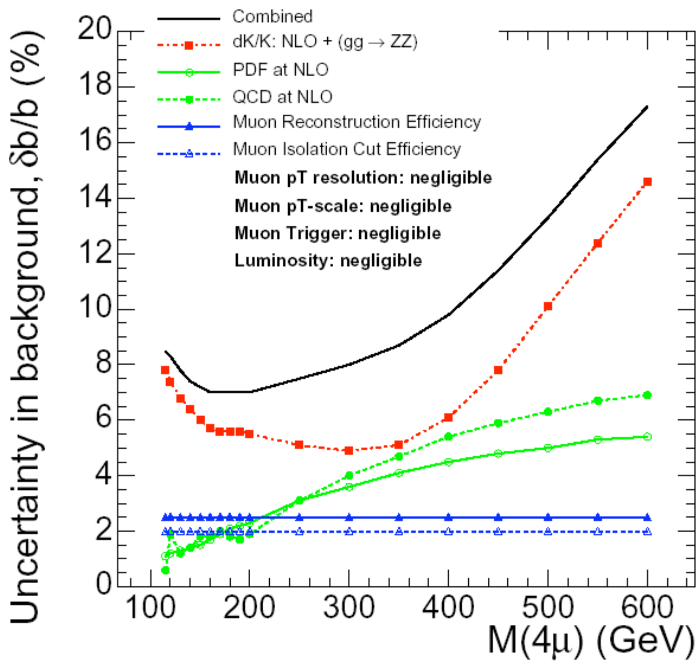
- muon isolation
- p_T of the second lowest p_T muon
- M(4μ) window (≈ 2σ where σ ≈ Γ_H + reso)

other half of the events used to compute significance

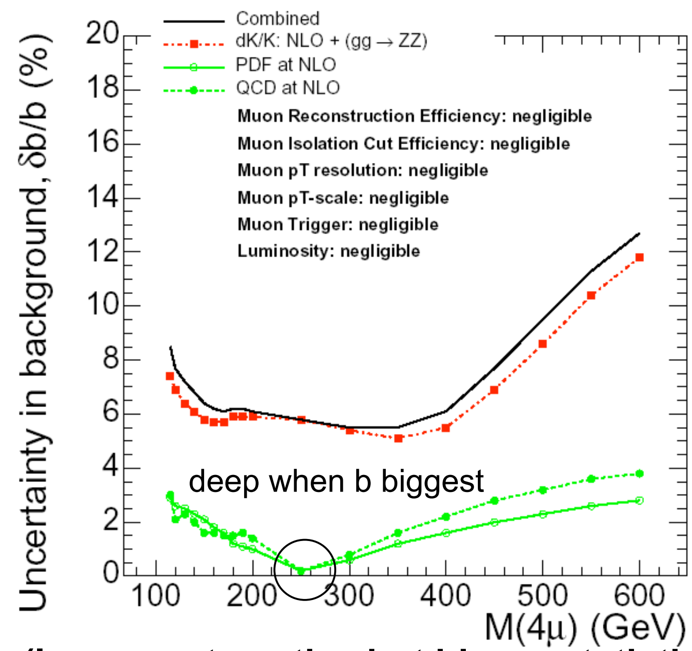
* **G**enetic **A**lgorithm for **R**ectangular **C**uts **O**ptimizatio**N** allows to check effectively a large set of cuts which, in a straightforward approach, would take an astronomical amount of time

4μ background systematics

Ratio H→ 4μ to Z→2μ (≈ 1 fb[-1])



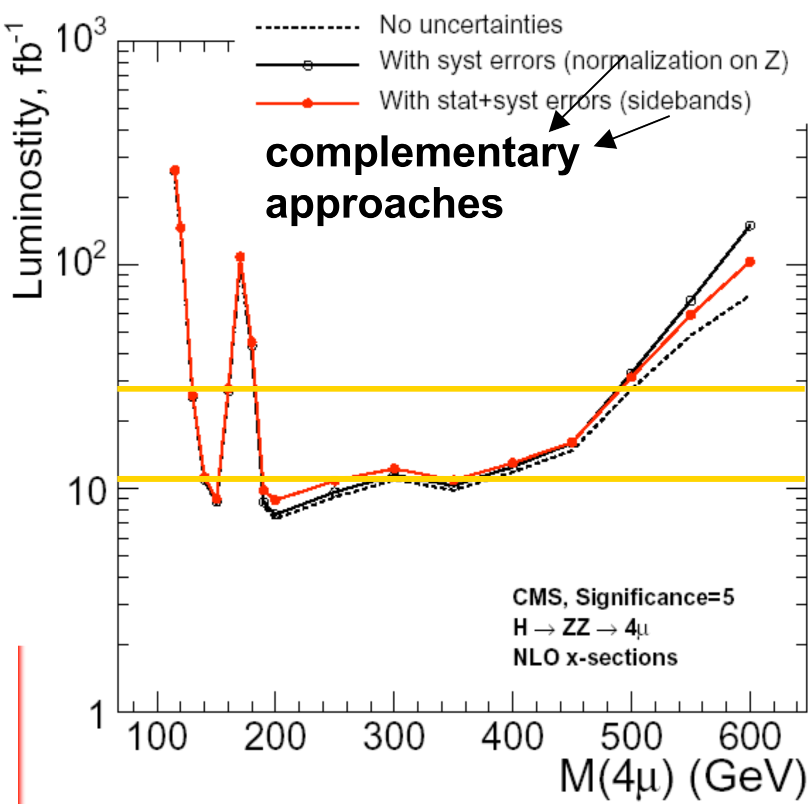
Normalization from sidebands



(lower systematics but bigger statistical error)

• new process NLO gg→ZZ ≈ (20±8)% LO xsec (different initial state so variations of QCD scale do not necessary give a feel for its relative importance)

4μ results

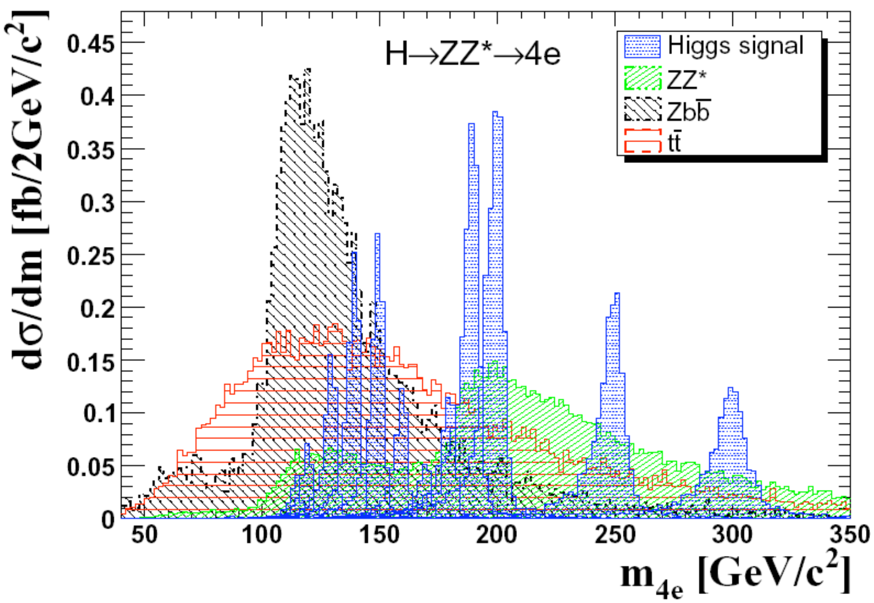


❑ problem of **significance overestimatimation** of a local discovery in **searching for a localized new phenomenon in a wide phase space**

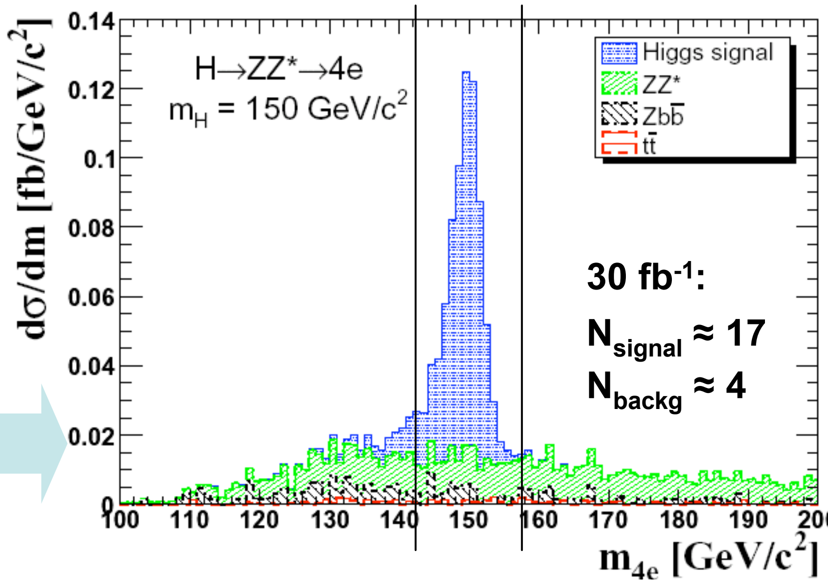
- check the **consistency with expected properties**:
  - xsec and variables not used in the analysis
  - M(4μ) shape consistency with sign+back hypothesis
- **decrease a priori the open phase space:**
  - M_H prior probability could be forced to be consistent with the fit to precision EW measurements
  - use the early data for a first hint and then discard them from analysis

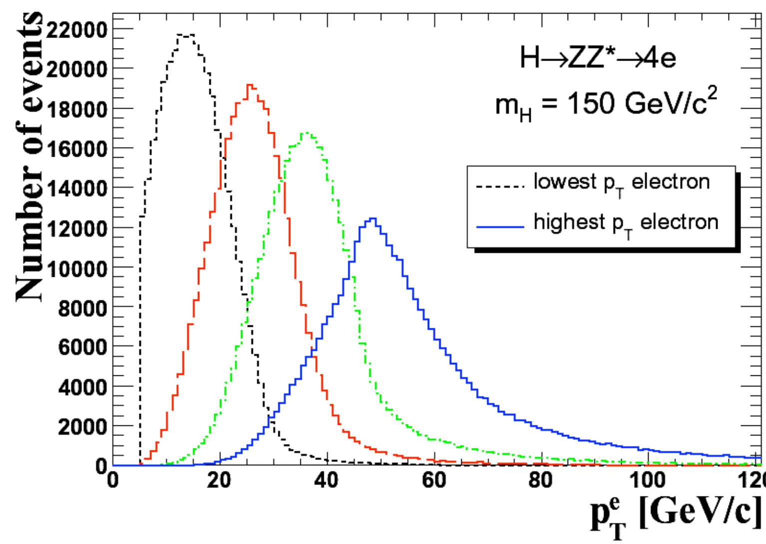
4e analysis

After trigger and preselection



After full analysis selection



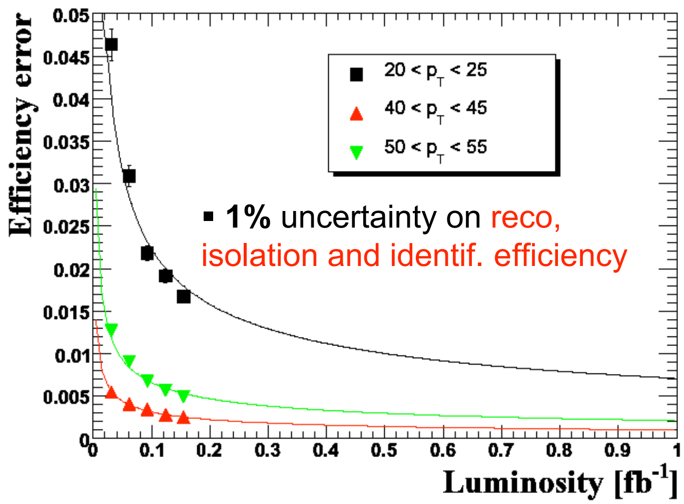


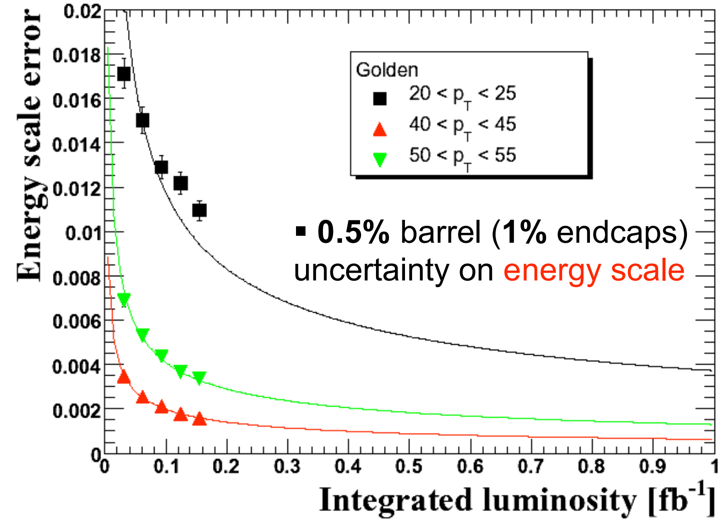
- **Optimization of low p_T $e^{+/-}$ reco**
- cuts to reject fakes are separately optimized for **different Bremsstr. $e^{+/-}$ classes**

4e: systematics & reults

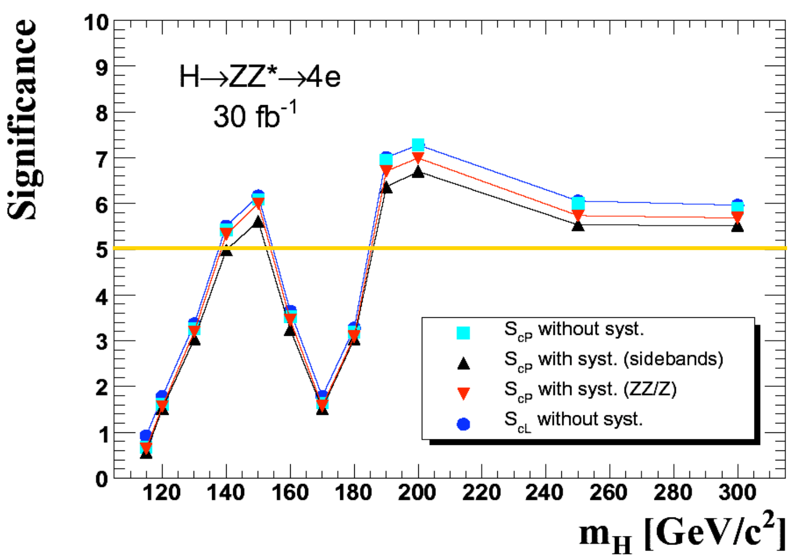
- Use $Z \rightarrow e^+e^-$ with one golden $e^{+/-}$, second $e^{+/-}$ used to estimate uncertainties

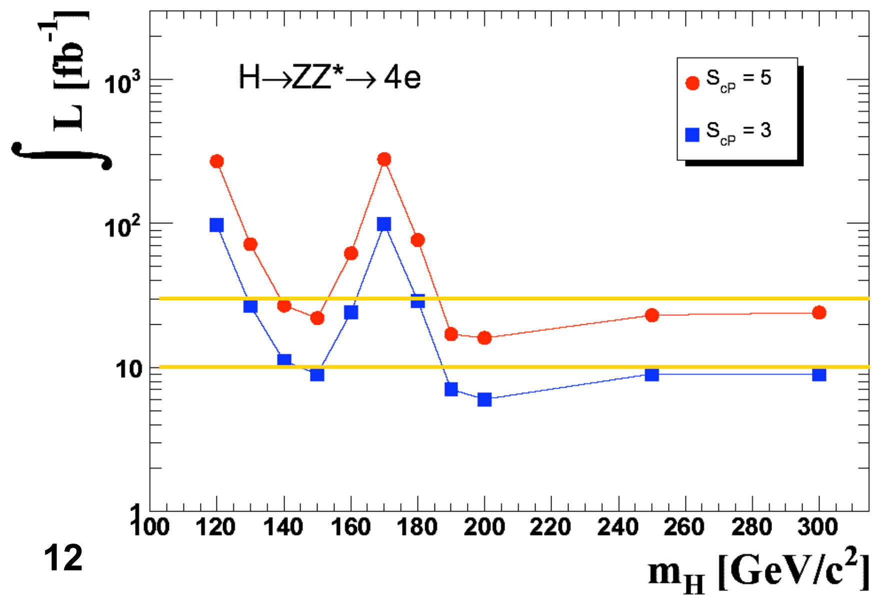
(best resolution on the Jacobian peak: $p_T \approx m_Z/2$, low $|\eta|$)





- **Tracker "radiography"** measuring the amount of $e^{+/-}$ Bremsstralhung
  - (**2%** material budget with 10 fb^{-1})





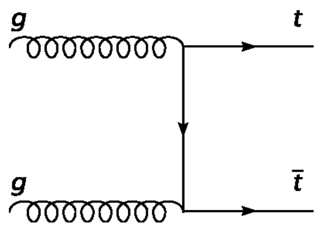
$H \rightarrow WW^{(*)} \rightarrow l\nu l\nu$ [M(H)=150-180]

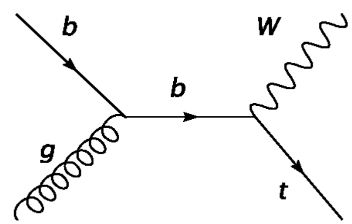
❑ **No narrow peak** →
- **high S/B needed**
- **good background shape control necessary** (normalization from data)
- **mass independent cuts**

▪ **signal:** all leptonic W decays (**0.5 - 2.3 pb** with a peak at $M_H \approx 160$ GeV)

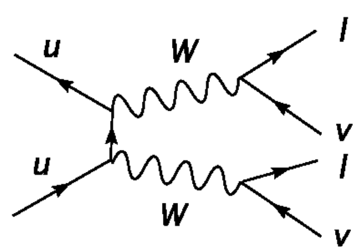
▪ **backgrounds:**

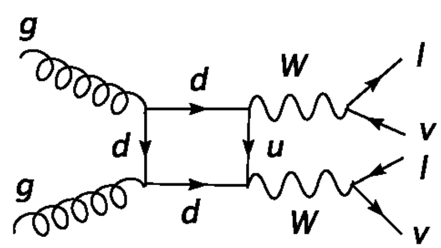
tt, tWb (≈ **90 pb**)





WW, WZ, ZZ (≈ **15 pb**)





(ggWW)

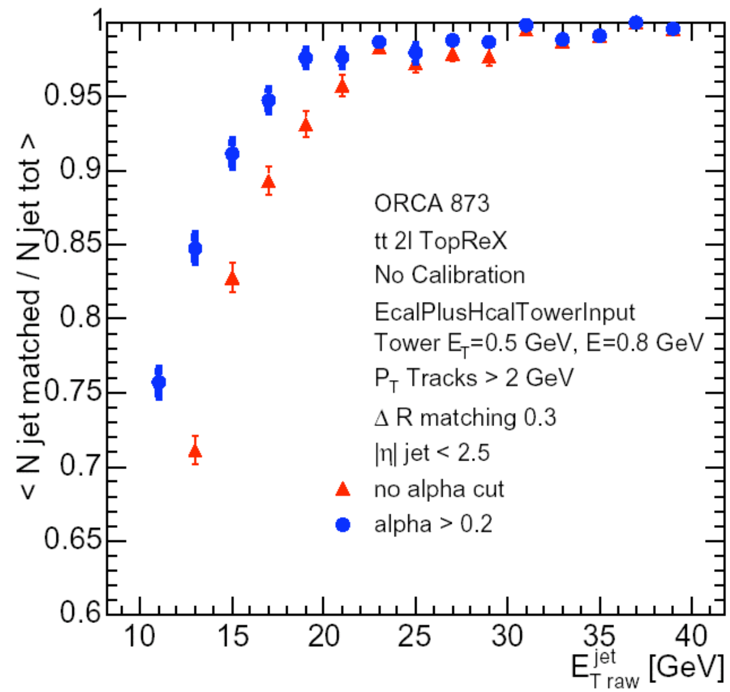
Z Drell-Yan not considered but checked that after selection should be < 2% of the total background

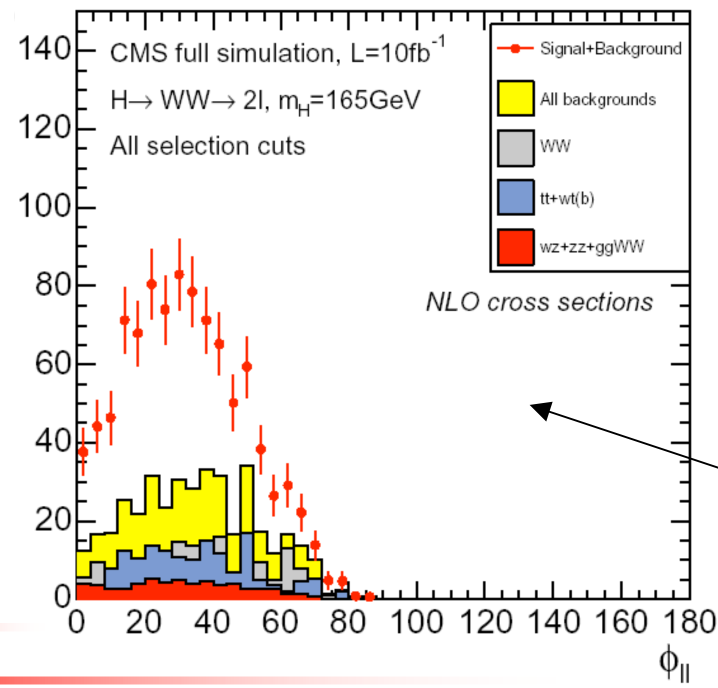
$l\nu l\nu$ analysis

- **central jet veto** (|η|<2.5, E_T>20 GeV)
  - no calibration (energy is not needed)
  - discrimination between *real* and *fake* jets (PU, UE, FSR, ISR, detector noise)

$$\alpha = \frac{\sum p_T(tracks)}{E_T(jet)}$$

α > 0.2 for jets with $15 < E_T < 20$ GeV





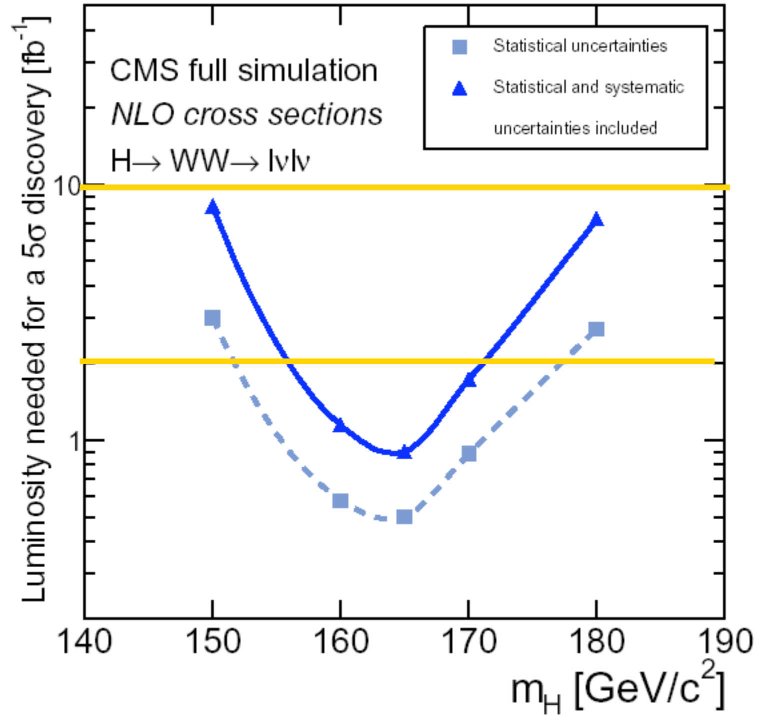
- **high MET** (> 50 GeV)
- **ee, eμ, μμ** reconstruction and selection
  - intermediate m(ll)
  - little φ(ll) in the transverse plane

lνlν results

❑ ΔB from data defining free signal region varying the analysis cuts

- ΔB (tt) ≈ 16% dominated by jet energy scale
- ΔB (WW) ≈ 17% dominated by statistic
- ΔB (WZ) ≈ 20% dominated by the presence of tt also

(values for 5 fb^{-1})

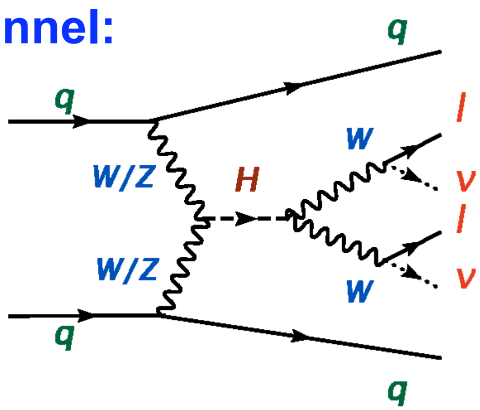


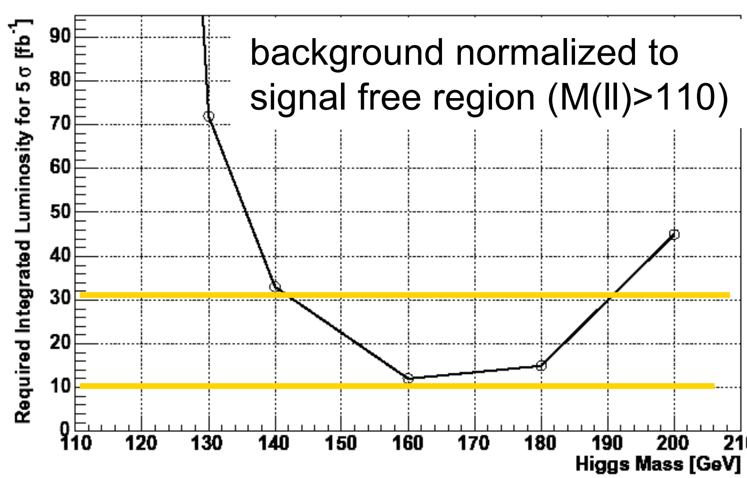
❑ tWb, ggWW small fraction of B:

- normalization region difficult to find
- syst uncertainties from MC

theoretical error dominates (20%, 30%)

Similar promising analysis specifically in **VBF channel:**





qqH with H→WW→lνjj [M(H) = 120-250]

+ BR ≈ 5.5 BR(lνlν) → xsec ≈ 0.02 - 0.8 pb

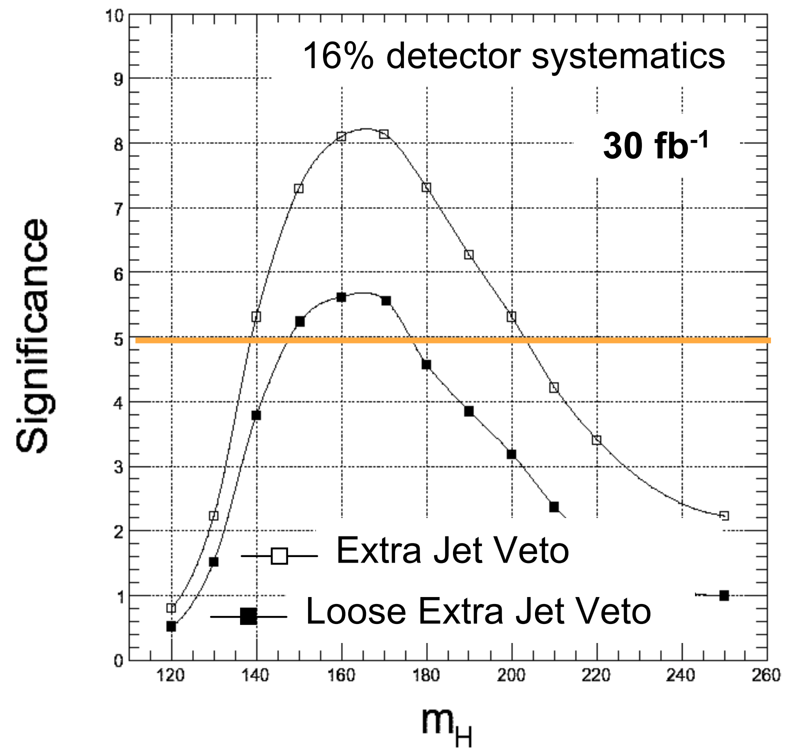
+ you can reconstruct the **Higgs mass**

- big amount of background → strong cuts → good knowledge of physics needed (measure backgrounds from data) :

- **tt + jets** (≈ 840 pb)
- **Wtb** (≈ 100 pb)
- **VV + jets** (≈ 100 pb)
- **V + jets** (≈ 700 pb)

✓ multiple jets xsec will be precisely measured from data

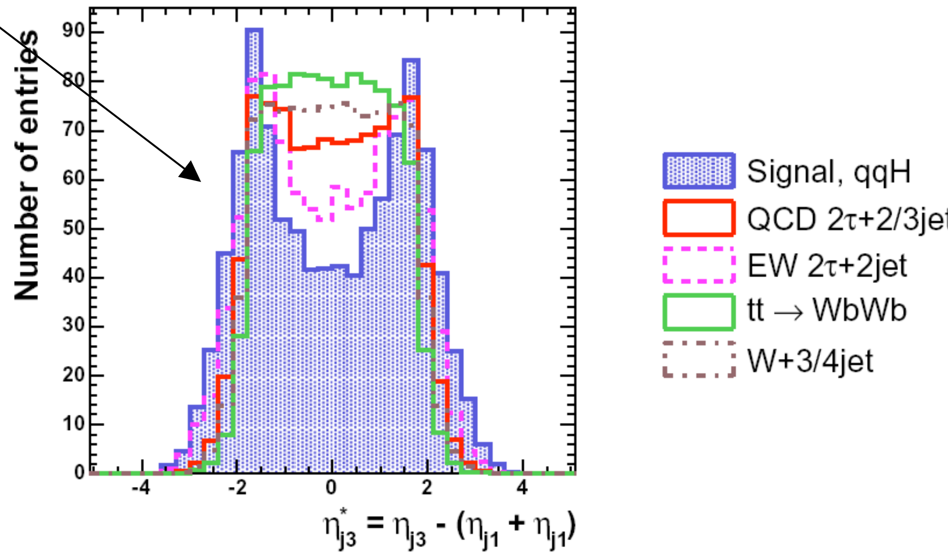
✓ many systematics about jets will be understood and resolved from data



qqH with $H \to \tau\tau \to$ lep + jet [M(H)< 150]

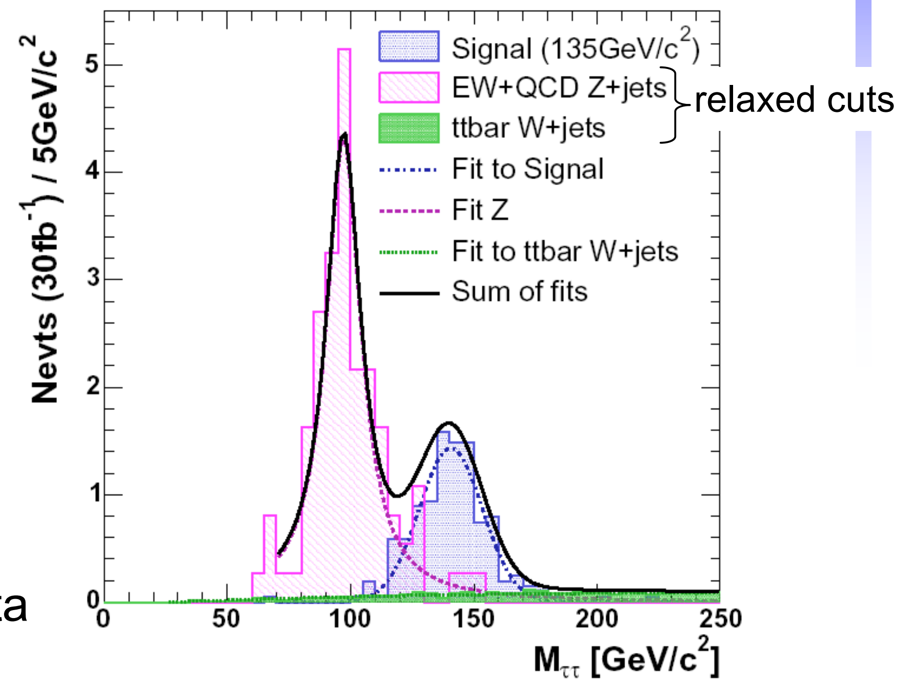
- backgrounds:
  - Z/γ^* + jets (irreducible),
  - $W \to l\nu$ + jets
  - $tt \to bl\nu bl\nu$

  (W→lν + jets and tt→blνblν:) with one jet misidentified as τ-jet
- complex signal kinematics:
  - forward jets with high rapidity gap (no color exchange)
    - MC calibration
    - central jet veto applied (with cut on α parameter)
  - high p_T lepton (e or μ)
  - MET: resolution 20% after correction
  - τ-jet identification
    - trigger on little (ΔR) isolated jet
    - offline impurity 2.7%
      efficiency 30% (mainly due to p_T, η cuts and request of isolation)
    - energy resolution 11.3%



$H \rightarrow \tau\tau$ results

- M($\tau\tau$) computed using collinear approximation of visible part of τ decay products and neutrinos
  - M($\tau\tau$) overestimated 5 GeV because of over-corrected MET
  - M($\tau\tau$) resolution of 9.1%



- **Significance exceeds 3σ at 30 fb^{-1}**
  - number of events computed from data using the M($\tau\tau$) fit (envisaged to do it in a region unaffected from signal)
  - error (σ_B) only from the fit:
    - 10k **toy MC data distributions** following the fit (with the number of events equiv. to 30 fb^{-1})
    - each sample refitted with free scale factors for the three independent fit
    - uncertainty = spread of the number of background events in the 10k samples

M_H [GeV]	115	125	135	145
S_{cP} at 30fb^{-1} ($\sigma_B = 7.8\%$)	3.97	3.67	3.94	2.18

Inclusive H→γγ [M(H)=115-150]

- inclusive signal production but with very low BR≈0.002
- pp→ γγ (irreducible)

  pp→ jets / γ + jets (reducible) with one jet misidentified as γ

  } very big background and very detector dependent + not well known QCD physics (big k factor in γ+jets events)

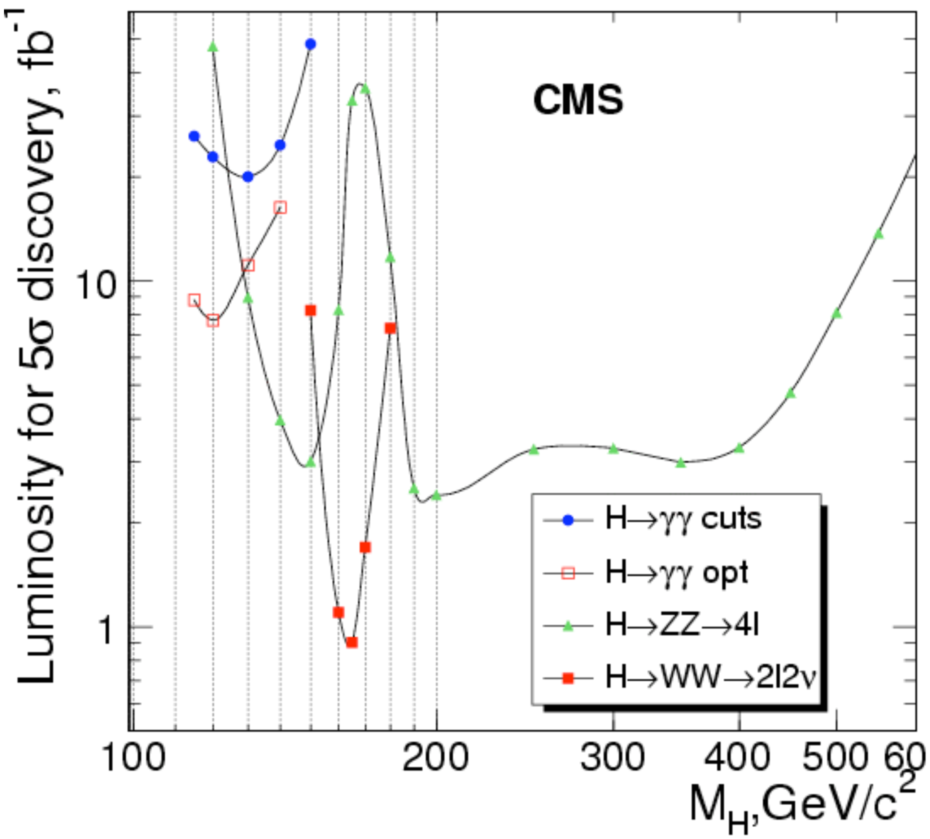
  Drell-Yan e^+e^-

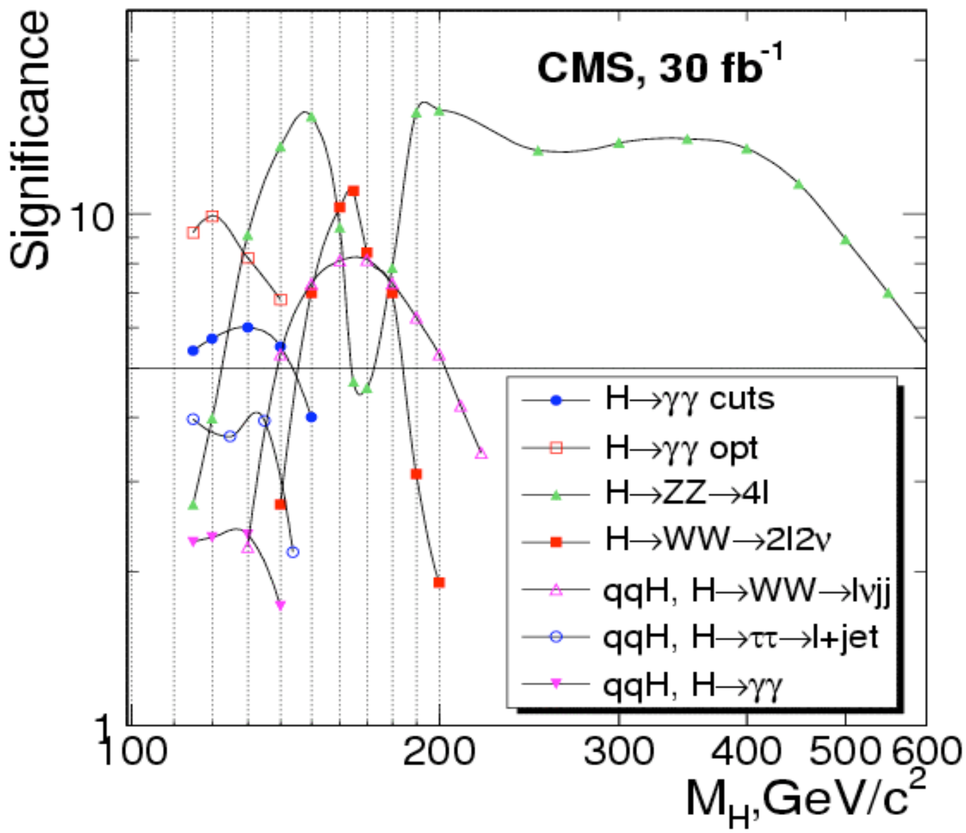
Great deal of uncertainty in the benchmark estimate of luminosity …

… this will **not be a systematic error on real data** since the background will be measured from data (thanks to the big sidebands signal free)

- **Analysis based on NN trained**
  - **on sidebands for backgr.** (1% systematic error on the background interpolation under the Higgs peak)
  - **on MC for signal**

Conclusions



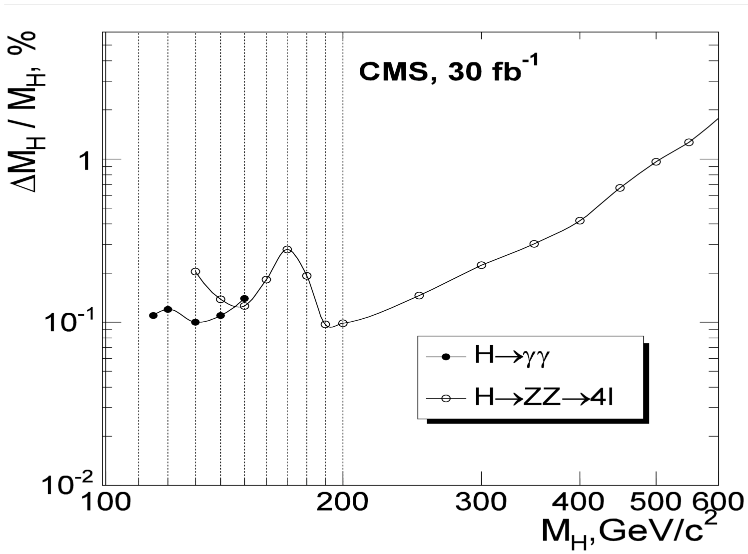


- These are inverse fbs of **(w.)u.d.** !!
  - detector systematics: jets, γ, MET (e and μ from Z→ll)
  - multiple jets background xsec: V+jets, VV+jets, tt

Back-up slides

S. Bolognesi – INFN Torino

Higgs properties measurement



- The mass can be measured with a precision between 0.1 % and 3.7 %
- The intrinsic width can only be measured when the Higgs boson is heavier than 190 GeV, with precisions around 25%, the experimental resolution dominating for lower masses
- The production cross-section can be determined with a precision around 20% for masses in the range 130 GeV-150 GeV and above 190 GeV

μ experimental systematics

- μ reco efficiency by counting # of Z in single μ HLT sample with p_T>20

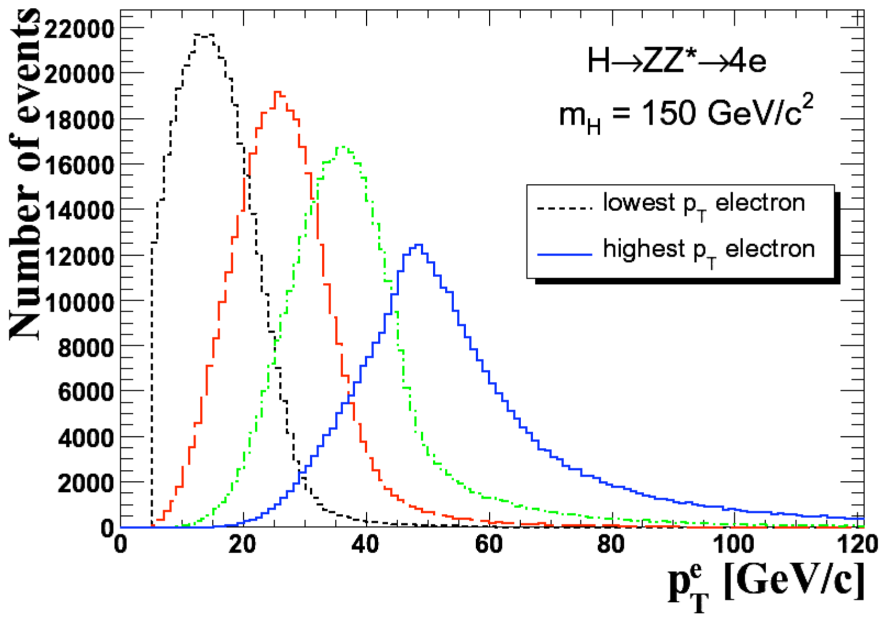
$$N_{Z(TRK)} = \epsilon_{HLT} \cdot \epsilon_{TRK} \cdot N_Z,$$
$$N_{Z(SAM)} = \epsilon_{HLT} \cdot \epsilon_{SAM} \cdot N_Z,$$
$$N_{Z(GMR)} = \epsilon_{HLT} \cdot \epsilon_{GMR} \cdot N_Z$$

$$\epsilon_{GMR} = (N_{Z(GMR)})^2/(N_{Z(TRK)}N_{Z(SAM)}),$$
$$\epsilon_{TRK} = N_{Z(GMR)}/N_{Z(SAM)},$$
$$\epsilon_{SAM} = N_{Z(GMR)}/N_{Z(TRK)}.$$

- μ p_T scale and resolution from J/ψ and Z peak

- trigger on single μ → efficiency ≈ 100% without sizeable uncertainty

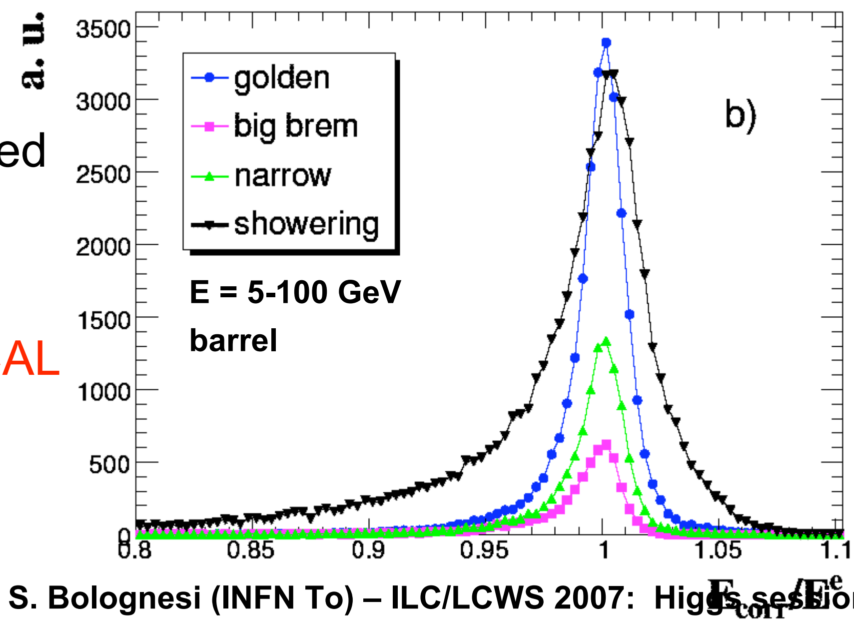
4e: electron reco



- **Optimization of low p_T $e^{+/-}$ reco**:
  - supercluster (cluster of cluster)
  - dedicated tracking with GSF using energy loss modeling

  to recover
  - Bremsstr. and initial showering in Tracker
  - energy ϕ spread due to magnetic field

- cuts to reject fakes are separately optimized for **different Bremsstr. $e^{+/-}$ classes**
  - supercluster size
  - ϕ and E matching between tracker and ECAL
  - $f_{brem} = (p_{in} - p_{out}) / p_{in}$



R_9 shower shape variable

Fraction of the super-cluster energy found inside the **3 by 3 array of crystals** centred around the highest energy crystal.

The shower shape variable R_9 very useful in discriminating between photons and jets. Because it looks in a small 3 × 3 crystal area inside the super-cluster it can provide information about narrow jets

Signal photons sometimes have low values of R_9 due to conversions, but usually R_9 provides additional isolation information about the super-cluster.

4e additional backgrounds

NOT explicitly considered but taken into account to choose cuts:

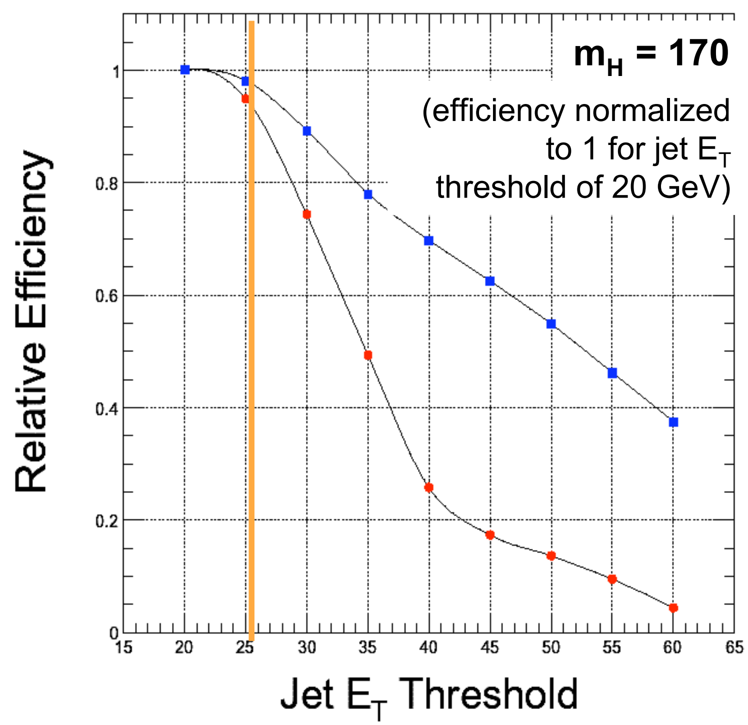
- electrons from D/B decay in QCD jets
- fake primary electrons due to early γ conversions
- π^o $\pi^\pm$ overlap

(e.g. Z+jets)

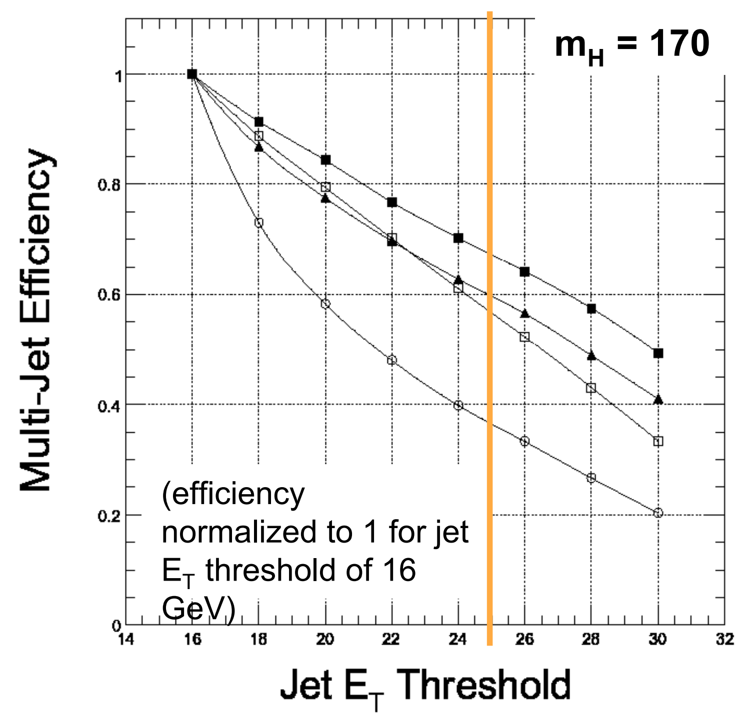
qq + H→lνjj : jets (1)

- Strong E_T cuts needed
  - for keeping an acceptable resolution (jets with E_T<30 GeV very difficult to calibrate)
  - for eliminating fake jets (most of PU jets with E_T<30 GeV)
- Strong E_T cuts affect efficiency:
  - Parton-jet matching efficiency
  - Efficiency of requiring at least 4 jets

signal forward quarks
signal quarks from W decay



signal
tt + jets
W + 4 jets
W + 3 jets



qq + H→lνjj : **jets** (2)

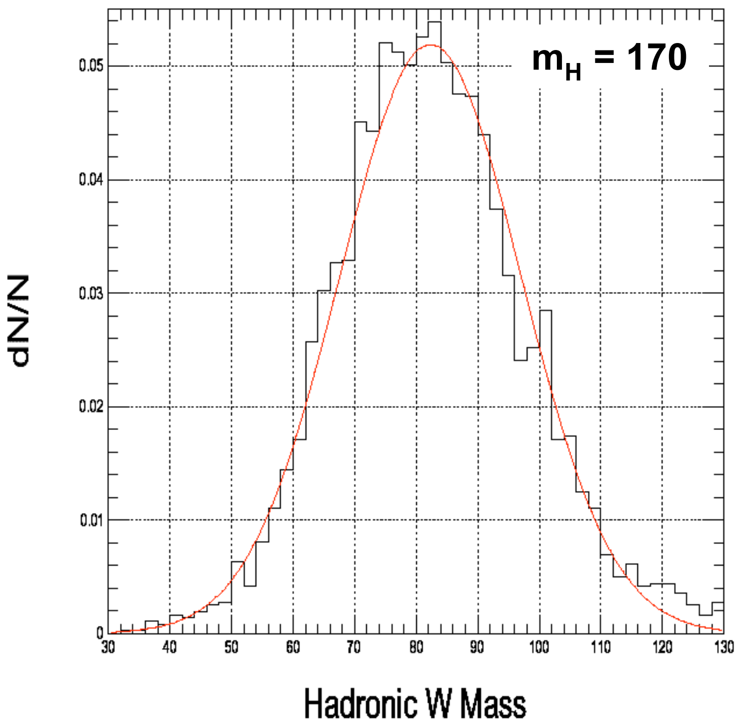
❑ **tag jets** misidentified with jets from FRS, ISR, PU, UE, detector noise …

In the signal this increases the chance of misidentification central jets from W

❑ **jets from W**:

- best possible resolution of **15 GeV** !!
- other central jets (E_T>20 GeV in 60% of events) often (20%) with higher E_T than jets from W

M(W→jj) using parton-jet matching



- MC calibration from QCD jet samples
- Iterative cone algorithm (ΔR=0.6)
- Fast Simulation for some backgrounds

$\gamma\gamma$ analysis

- γ reconstruction and preselection
  - **ECAL crystal resolution** from $W \rightarrow e\nu$ calibration after 10fb^{-1}: 0.3% barrel, 1.0% endcaps

    γ **reconstruction efficiency** ≈ 100% in the ECAL acceptance
  - **vertex refitted** from high p_T tracks → 5 mm resolution in 81% of the cases (needed to have right γ direction → precise m_H)
  - NN to combine the **isolation** variables (Tracker, ECAL, HCAL)
- Analysis performed with NN:
  - NN input: $E_T/M(\gamma\gamma)$ signal has higher E_T

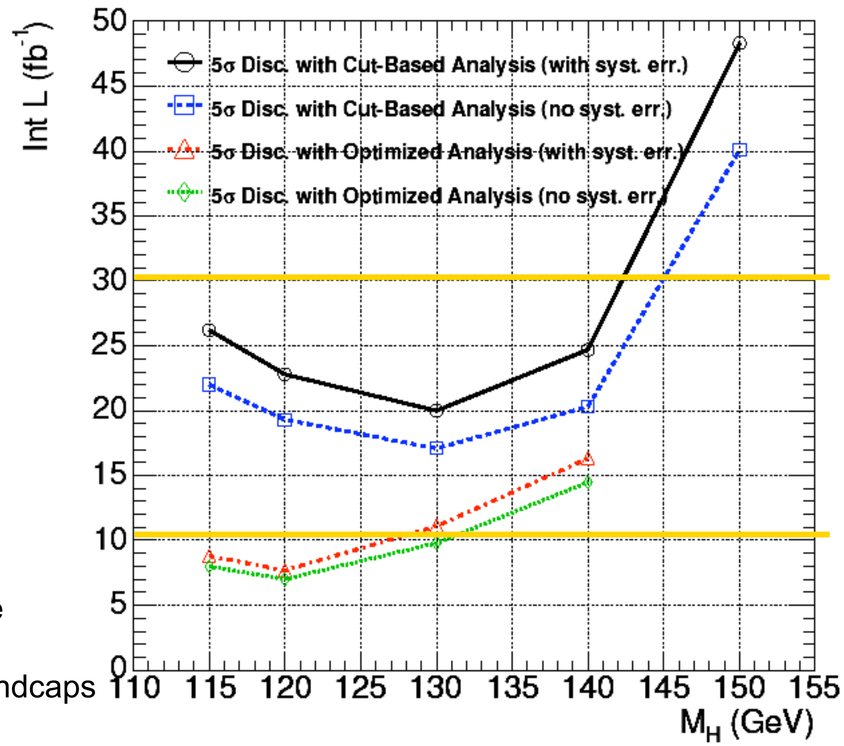
    $\Delta\eta(\gamma\gamma)$ backgr. have high mass only if high $\Delta\eta$

    NN_{isol} output against jets

    longitudinal momentum
  - trained separately on 6 categories:

    3 steps of R_9
    - signal events with better mass resolution have higher R_9
    - jets and π^0 have lower R_9

    barrel / endcaps
    - signal events in barrel have better resolution
    - higher background in the endcaps



Significance computation

- Counting experiment approach ($\mathbf{S_{cP}}$)
  probability from a **Poisson distribution** with mean N_B to observe $N \geq N_B + N_S$ converted in equivalent number of Gaussian standard deviations
- Log-likelihood ratio significance ($\mathbf{S_{cL}}$)
  likelihood ratio of probability of observing data in the signal+background hypothesis to the probability of observing the data in background only hypothesis