



中國科學院為能物招加完所 Institute of High Energy Physics Chinese Academy of Sciences

## SM Higgs search with H→yy at CMS

#### Junquan TAO (IHEP/CAS)

#### **Guoming CHEN (IHEP/CAS)**

2010年高能物理年会,南昌

April 16-21, 2010

### Outline

Introduction & motivation

Sensitivity of Higgs searching with  $H \rightarrow \gamma \gamma \otimes CMS$ 

> Overview of the H $\rightarrow\gamma\gamma$  analysis in CMS PTDR

 $\geq$  Improved analysis @14TeV with MC samples @NLO ( $\sigma$ )

- Improved Cut-based analysis with new selections
- Improved event-optimized analysis related to TMVA.

Ongoing work and near futrue plan with LHC collisions data @ 7TeV

Conclusion

### Introduction

LHC @ CERN: CMS (Compact Muon Solenoid), ATLAS (A Toroidal LHC Apparatus), ALICE (A Large Ion Collider Experiment) & LHCb (LHCbeauty).

#### Main Physics goal: "God particle"-Higgs Boson

#### LHC for Higgs: —Direct searching: SM Higgs & non-SM Higgs

-Indirect searching: Precision measurements of  $m_t$  and  $m_w$ 

What is the source of mass? What breaks SU(2)<sub>L</sub> × U(1)<sub>y</sub> ?





#### Motivation: The SM $H \rightarrow \gamma \gamma$ search

#### **Search for the Higgs Particle**

#### **Experiments**











### MC samples for signal & backgronds

Higgs Mass used in analysis: mH=120GeV,130GeV,140GeV and 150GeV
 The cross section and BR for signal were used: (NLO)

mH	120 GeV	130 GeV	140 GeV	150 GeV
σ (gg fusion)( <b>pb</b> )	36.4	31.6	27.7	24.5
σ (IVB fusion) ( <b>pb</b> )	4.5	4.1	3.8	3.6
σ (HW, HZ, ttH) ( <b>pb</b> )	3.3	2.6	2.1	1.7
Total ( <b>pb</b> )	44.2	38.3	33.6	29.7
BR (H→ ♀ ♀ )	2.21x10 <sup>-3</sup>	2.24x10 <sup>-3</sup>	1.95x10 <sup>-3</sup>	1.40x10 <sup>-3</sup>
Inclusive o x BR (fb)	97.5	86.0	65.5	41.5

#### Backgouds Cross section with PYTHIA (LO) and "K factor" (LO→NLO)

Process	P <sub>t hat</sub> (GeV)	σ after preselection (pb)	K-factor
pp→үү (born)	>25	<b>45</b>	1.5
pp→ ɣ ɣ (box)	>25	36	1.2
pp→ γ +jets	>25	600	1.72 (2 prompt) 1.0 (1 prompt+1 fake)
pp→jets	>50	4800	1.0

#### **Overview of H** $\rightarrow \gamma\gamma$ analysis process in CMS PTDR

(In CMS Physics Technical Design Report 2) Before 2006, older CMS softerware: OSCAR+ORCA ...; now CMSSW.

#### "cut-based" analysis:

- 1). Passing the High Level Trigger of photons
- 2). Kinematic cuts on 2 Rec. photons: ET( $\gamma$ 1)>40GeV  $\Sigma$  ET( $\gamma$ 2)>35GeV  $\langle |\eta|<2.5$
- 3). Photon isolation selection: Track ISO; ECAL ISO; HCAL ISO
- "event-optimized" analysis: Photon ISO ANN analysis & event optimization ANN analysis (preselections ET(γ)>40GeV、 |η|<2.5 and loose ISO)
  </p>



#### **Event optim. ANN 6 inputs**

- ✓ Photons ISO: NNiso1, NNiso2
- ✓ Et<sub>γ1</sub>/M<sub>γγ</sub>
- ✓ Et<sub>γ2</sub> /M<sub>γγ</sub>
- ✓ | η 1- η 2| of γγ
- $\checkmark$  P<sub>L</sub> of  $\gamma\gamma$  (Higgs candidates)

#### Drawbacks of the analysis in CMS PTDR

- > No  $\gamma/\pi^0$  discrimination (to suppress the reducible backgrounds with fake photon).
- When event optimization analysis, the other information of one event was not considered, such as jets and MET etc., to separate the signal and backgrouds.

#### Backgroud decision/estmation with sideband was not included.

Didn't give out the selection efficiency in event optimization analysis. The number of events is unknown for signal and backgrounds respectively.

Etc.

### Chance for us to participate in the H $\rightarrow \gamma\gamma$ analysis at CMS

#### $\gamma/\pi^0$ discrimination: unconverted photons in Barrel

#### Unconverted case in Barrel

PT bins (GeV)	<b>π 0 rejection efficiency</b> for keeping 90% photon efficincy (%) (BDT in Barrel)				
	N12	N12 M3 F6			
20-25	69.6±0.6	51.2±0.6	63.8±0.6	72.5±0.6	
25-35	61.6±0.4	39.4±0.4	57.0±0.4	66.9±0.4	
35-45	50.7±0.5	30.9±0.4	48.5±0.5	60.0±0.5	
45-55	40.0±0.5	26.0±0.4	41.7±0.5	50.9±0.5	
55-65	34.5±0.5	23.9±0.4	34.1±0.5	42.8±0.5	
65-75	29.6±0.5	21.7±0.4	29.8±0.5	37.0±0.5	



Unconverted case in Endcap:

$P_T (\text{GeV})$	保持90%光子效率的 $\pi^0$ 排斥率(%)
20-25	$64.1 {\pm} 2.0$
25-35	$62.1 \pm 1.3$
35-45	$57.1 \pm 1.3$
45-55	$54.4 \pm 1.2$
55-65	$51.8 \pm 1.0$

#### $\gamma/\pi^0$ discrimination: Converted photons

Shower shape viriables and converted tracks etc.:

ET range	$\pi$ 0 rejection for keeping 90% γ efficiency (%)
20-25	31.5
25-35	48.0
35-45	67.3
45-55	74.0
55-65	75.1
65-75	71.5



### Improved Cut-based analysis with M<sub>H</sub>=120GeV



### Event-optimized analysis: Photon ISO

- Information with different cone size around the photon used for inputs.
  - Training signal: Reco. ¥ from gg fusion H (Isolated).
  - Training background: Reco. ¥ from jets sample (Nonisolated)

Results: signal efficiency vs bkg rejection





For keeping 90% signal (ISO photons), ~8% higher bkg (Non-ISO) rejection

### Event-optimized analysis: Event Opt.



NNiso for both photons as inputs.

Add 6 new inputs: Jets & MET etc.

Considering preselection efficiency for signal: ~65%, if keeping 50% signal efficiency here, the combined signal eff. will be ~33%.

#### Much higher bkg rejection eff.: ~16.4%

### **Event-optimized analysis**

Direct cut on the TMVA-BDT output here.
Required NNevt>0.05

> Kinematic cuts on photons: PT(Leading  $\gamma$ )>40.0, PT(Trailing  $\gamma$ )>35.0,  $|\eta|<2.5$ 

>NN( $\gamma/\pi 0$ ) cuts: (NN>NNmin), NNmin is the same as cut-based analysis

mH (GeV)	S <sub>L</sub> @ 1fb-1	5 σ Discovery	<b>3</b> σ evidence	95% CL exclusion
120	1.76	8.0 /fb	2.9 /fb	1.3 /fb
130	1.55	10.4 /fb	3.7 /fb	1.6 /fb
140	1.23	16.4 /fb	5.9 /fb	2.6 /fb
150	0.69	52.0 /fb	18.7 /fb	8.2 /fb



Junquan Tao

#### Normolized to 1/fb, signal $\times$ 10

### Analysis with 6 categories

Different s/b in different categories based on R9 and pseudo-rapidity of photons.

Discovery sensitivity for different Higgs mass



mH (GeV)	5 σ Discovery	3 σ evidence	95% CL exclusion
120	7.8 /fb	2.8 /fb	1.3 /fb
130	10.0 /fb	3.6 /fb	1.6 /fb
140	15.9 /fb	5.7 /fb	2.5 /fb
150	36.8 /fb	13.2 /fb	6.0 /fb

### Estimation of background

Error on bkg estimation with "sideband".

✓ Uncertainty of the fit function
 ✓ Statistics error with the events for fitting

> Fitted with *a 3rd order polynomial function*.

➢At 10 fb-1, for 1 whole category, the error from the fit function is ~0.6%; the statistical error is ~0.9%. The total error is estimated to be 1.1%.

For 6 categories, total error: ~2.3%.





### Systematic error on the signal

#### > 20% uncertainties on the signal in total

Sources	Uncertainties
Theory	~15%
Int. Luminosity	~3%
Trigger	~1%
Others	~1%

### Event-optimized results with sys. error

With 1.1% systematic error of bkg and 20% uncertainty on signal, the significance:

#### whole samples as 1 category

mH (GeV)	5 σ Discovery	<b>3</b> σ evidence	95% CL exclusion
120	12.6 /fb	4.5 /fb	2.0 /fb
130	16.1 /fb	5.8 /fb	2.6 /fb
140	25.9 /fb	9.3 /fb	4.1 /fb
150	80.6 /fb	29.0 /fb	12.8 /fb

With 2.3% systematic error of bkg and 20% uncertainty on signal, the significance:

#### Analysis with 6 categories

mH (GeV)	5 σ Discovery	<b>3</b> σ evidence	95% CL exclusion
120	12.3 /fb	4.4 /fb	2.1 /fb
130	15.8 /fb	5.7 /fb	2.6 /fb
140	25.3 /fb	9.1 /fb	4.0 /fb
150	55.6 /fb	20.0 /fb	9.4 /fb

### Results when assuming 30% uncertainty on Bkg

Backgrouds generated with PYTHIA (LO)

For the "K-factor" uncertainty of ~20-30% (LO→NLO), we assumed another 30% uncertainty on Bkg here

>The **discovery sensitivities**:

mH (GeV)	5 o Discovery	<b>3</b> σ evidence	95% CL exclusion
120	15.7 /fb	5.7 /fb	2.5 /fb
130	20.3 /fb	7.3 /fb	3.3 /fb
140	32.1 /fb	11.5 /fb	5.1 /fb

For the insensitivity of the Higgs research with mH=150GeV by the channel  $H \rightarrow \gamma \gamma$ , the result with 30% uncertainty on Bkg is not shown here.

# Significance and sensitivity of Higgs research with the channel $H \rightarrow \gamma \gamma @ CMS$



### Ongoing work with collisions @ 7TeV: Photon Calibration

> With LHC 7TeV collisions, 2010 is THE YEAR for calibration activities.

**ECAL calibration with**  $\pi \mathbf{0} \rightarrow \gamma \gamma$  and  $\eta \rightarrow \gamma \gamma$  at the startup.

> Cluster of  $\gamma$  candidates based on 3x3 crystals array, if the energy is higher, there will be overlapping between 2  $\gamma$  candidates

>Try to solve the shower overlapping problem in the higher energy region. The parametric shower shape profile method is being used for such purpose.

In a higher energy region, to keep the precision of photon energy measurement.



### **Near futrue plan with collisions @ 7TeV: Di-photons in CMS**

pb/GeV/c

- σ (Y Y + X) CDF published with 200pb<sup>-1</sup> (hepex/0412050)
- Due to higher cross section, CMS will have equivalent statistics with ~10pb<sup>-1</sup>.
- How to solve the photon purity problem?
- The parametric shower shape profile method as in CDF analysis (unconverted photon case).

**"Template** method" trying with the outputs of **Neural Network with**  $\gamma/\pi 0$  discrmination analysis.





profile method. Also shown are the total systematic uncertainties on these efficiencies, and the measured efficiency of the data as a function of photon  $P_t$ .

FIG. 3. Simulated  $\tilde{\chi}^2$  distributions for 15 GeV/c photons (solid) and  $\pi^0$ 's (dashed).

### Conclusion

> We finished the SM  $H \rightarrow \gamma \gamma$  analysis with 14TeV @CMS based on the MC samples @NLO.

For both cut-based analysis and event optimized analysis, with the backgound can be fixed by the sidebands fit, better results can be obtained.

For Higgs mass less than 140 GeV, 5 or discovery result or at least a strong exist evidence of Higgs can be obtained with the data of 20 /fb.

> Expect to **contribute** a lot to the task force of SM  $H \rightarrow \gamma \gamma$  analysis @ CMS with the LHC colision data.



### Backup



Di-photon invariant mass spectrum after the selection for the cut-based analysis. Events are normalized to an integrated luminosity of 1 fb–1 and the Higgs signal, shown for different masses, is scaled by a factor 10.

### Higher order $\gamma\gamma$ Generator Studies



- Interesting from a theoretical point of view because prompt diphoton production is between pure QCD and QED - It's a background to light Higgs searches (and new physics in some BSM models)

- CDF/D0 show that Pythia (LO) can not describe accurately the shapes of  $\gamma\gamma$ +X events => need for NLO calculations

27

M<sub>yy</sub> (GeV)

### Applying to other Higgs mass samples

Events/GeV @ 1fb<sup>-</sup>  $H \rightarrow \gamma \gamma M$  = 130GeV (×10) 70 > Applying the TMVA result and the final H→γγ M. =140GeV (×10)  $H \rightarrow \gamma \gamma M$  =150GeV (×10) selections to other signal samples with jets p\_>50GeV 60  $\gamma$ +jets (2 prompt  $\gamma$ ) mH=130GeV,140GeV and 150GeV 50 γγ box 🚫 γγ born 40 Analysis with CSA07 samples, 30 normolized to 1/fb, signal  $\times 10$ 20 10 Selection efficiencies for signal samples 100 110 120 140 160 130 150 170 180 M...(GeV) Selection **Selectio** mH (GeV) events mΗ events Eff. @1fb-1 (GeV) n Eff. @1fb-1 120 gg fusion 22.26% 17.9 130 21.08% 14.9 IVB 14.34% 1.4 16.07% 1.5 ZH,WH,ttH 11.87% 0.9 12.63% 0.7 140 20.79% 11.2 150 18.59% 6.4 16.36% 1.2 15.84% 0.8 12.60% 0.5 12.42% 0.3

 $H \rightarrow \gamma \gamma M_{..} = 120 \text{GeV} (\times 10)$ 





Sub-process	PTDR	now	tools used now
gg fusion	NLO	NNLO	HggTotal
VBF	NLO	NLO	VV2H
VH	NLO	NLO	V2HV
ttH	LO	LO	HQQ

### **Pre-analysis**

Filter: at least 2 Rec. correctedPhotons with pT Min.=20GeV.

L1 & HLT: Previous studies show ~99% efficiency for L1 and HLT after analysis selection. The inefficiency due to trigger is negligible.

- > Also **pseudo-rapidity cuts** on Photons:  $|\eta|$ <2.5
- Primary vertex selection:
- ✓ Using the default object "offlinePrimaryVerticesFromCTFTracks".
- $\checkmark$  For gg fusion with mH=120GeV, the PV rec. efficiency is ~98%.
- ✓ If no PV, using the nominal vertex (0.0, 0.0, 0.0) instead.

Only z position correction as PTDR has the same result as the default P.V. in the invariant mass calculation of 2 photons



### Variables as TMVA inputs

#### NN 12 variables (N-12)

- $V_0 = |S_{2 \times 5Right} S_{2 \times 5Left}|/S_{25}$
- $V_1 = \sigma_{\eta\eta}/0.0004$
- $V_2 = \sigma_{\phi\phi}/0.001$  (假如 $\sigma_{\phi\phi} < 0.001$ , 否则 $V_2 = 0.0$ )
- $V_3 = S_1/S_9$
- $V_4 = (S_9 S_1)/(S_{25} S_1)$
- $V_5 = S_4/S_{25}$
- $V_6 = |S_{2 \times 5Bottom} S_{2 \times 5Top}|/S_{25}$
- $V_7 = E_{3 \times 2Ration}$
- $V_8 = S_6/S_9$

#### 3 Moments variables (M-3): M<sub>MAJ</sub><sup>2</sup>, LAT& PZM

- $V_9 = \lambda_2/\lambda_1$
- $V_{10} = (E_2 + S_1)/S_9$
- $V_{11} = (E_2 + S_1)/S_4$





图 5.5: (a)端盖使用神经网络方法时的示意图。通过连接ECAL端盖簇射中心与对撞点 (0.0,0.0,0.0)的连线,找出与preshower的交点。(b)神经网络中使用到的preshower每 一层11条硅微条(相交点左右±5条硅微条),其大约为2cm宽

- Preshower X-层上11个变量: V<sub>0</sub>、V<sub>1</sub>、...、V<sub>10</sub> = <sup>E<sub>X</sub></sup>/<sub>C<sub>X</sub></sub>。其中i = 0,±1,±2,...,±5,即中间硅微条(X<sub>imp</sub>所在硅微条),左右±1一直到±5的硅微条。E<sub>i</sub><sup>X</sup>为该层上第i根硅微条的能量读出。对于中心及±1的硅微条,C<sub>X</sub> = 0.02,对其余硅微条C<sub>X</sub> = 0.01。
- Preshower Y-层上11个变量: V<sub>11</sub>、V<sub>12</sub>、...、V<sub>21</sub> = <sup>E<sub>i</sub>Y</sup>/<sub>C<sub>Y</sub></sub>。其中i的定义与X-层一样。E<sup>Y</sup><sub>i</sub>为 该层上第i根硅微条的能量读出。对于中心及±1的硅微条, C<sub>Y</sub> = 0.04, 对其余硅微 条C<sub>Y</sub> = 0.02。
- ECAL端盖中有关能量沉积的3个变量: V<sub>22</sub> = S<sub>1</sub>/C<sub>EE</sub>、V<sub>23</sub> = S<sub>9</sub>/C<sub>EE</sub>、V<sub>24</sub> = S<sub>25</sub>/C<sub>EE</sub>。 其中S<sub>1</sub>、S<sub>9</sub>、S<sub>25</sub>的定义与桶部的一样,分别为最大能量沉积的晶体的能量、中 心晶体周围3 × 3晶体矩阵中的总能量、中心晶体周围5 × 5晶体矩阵中的总能 量。假如S<sub>25</sub> < 500GeV,系数C<sub>EE</sub> = 500GeV;假如500GeV < S<sub>25</sub> < 1000GeV, 系数C<sub>EE</sub> = 1000GeV;假如1000GeV < S<sub>25</sub> < 7000GeV,系数C<sub>EE</sub> = 7000GeV。

#### Application of $\gamma/\pi 0$ discrimination: Unconverted case

Selection of unconverted case: "N<sub>trk</sub><sup>ConvID</sup>=0" method, "track finding for identification of converted photons"



NN cut value vs SC Et: (ANN results from CMS AN-2008/063)

Junquan Tao

#### Resuls of $\gamma/\pi 0$ discrimination: Converted case



### 其他研究组的结果

UCSD (University of California, San Diego) Marco Pieri (convenor)

 可确定本底, CMS Note-2006/112 "cut-based"的结果 (老数据): mH=120GeV,不考虑系统误差

Analysis	5 $\sigma$ discovery no systematic error
Counting experiment	27.4 /fb
1 category	24.5 /fb
4 categories	21.3 /fb
12 categories	19.3 /fb

 CSA07优化分析结果(2008.12.04的报告, 不能确定 本底): mH=120GeV,不考虑系统误差, L(5 σ)~11fb<sup>-1</sup>

#### >Università di Roma "La Sapienza" and INFN Sez.

**Roma**, Francesco Pandolfi (2008.11.21的报告) CSA07分析结果,**有本底谱**,**不考虑系统误差**时5σ发现所

需要的积分亮度

mH=120GeV	19.3 /fb
mH=130GeV	20.5 /fb
mH=140GeV	28.0 /fb
mH=150GeV	53.1 /fb
	Junquan Tao





### **Event-optimized analysis: inputs**

#### ightarrow Et<sub>y1</sub> /M<sub>yy</sub>, Et<sub>y2</sub> /M<sub>yy</sub>, | n 1- n 2| and PL are almost independent of the Higgs mass



36