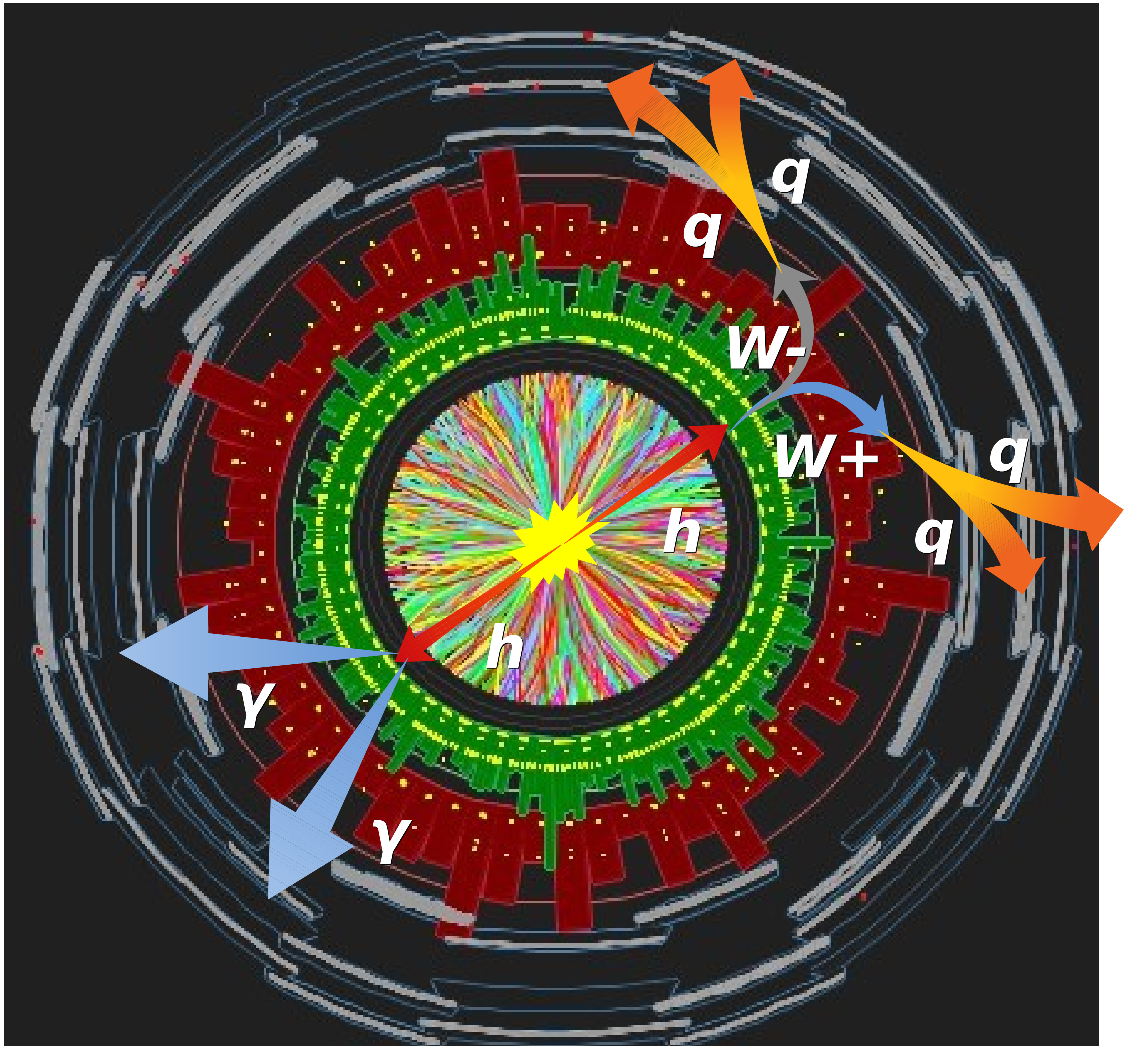
BSM gg → H → hh → WWγγ Search with jjjjγγ final state

Yaquan Fang, Xiaohu Sun
28-05-2014
IHEP, Beijing

Introduction

- Final states jjjjγγ for searching
  - SM hh production
  - BSM gg → X → hh production

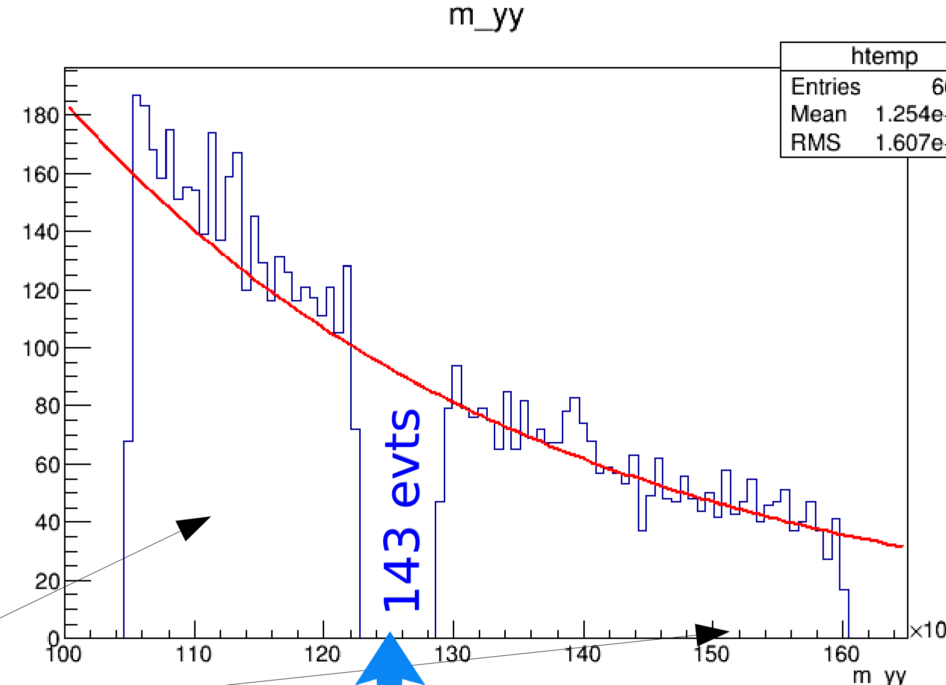


Signal region event yields [EXP]

- Signal region (ask yy cuts && njets>=4):
  - mass(yy) is required by $|m_h - \Delta m_h - m_{\gamma\gamma}| < 2\sigma_{\gamma\gamma}$
  - where mh=125.6, deltamh=0.15, sigma=1.6

SIGNAL REGION

bkg components	# of evt
SM H (ggH,VBF,VH,ttH)	~8
Continuum	~143
	~151



Sideband data

* bkg in signal region estimated by fitting to exponential with sideband data

A very rough est on upper limits

- With the estimated bkg ~151 events, one can fluctuate from the background-only hypothesis with an excess of 2sigma (95%) ~24.6
  - Use this number to roughly estimate the exp upper limits
  - For this est, NO systematic uncertainties are considered
  - Working point: all yy-cuts && Njets>=4

	Lumi (pb-1)	Branching ratio	Cut eff	Upper limit
Non-resonant SM HH	20,000	4.48e-4	15%	**18 pb**
Resonant 300GeV	20,000	4.48e-4	9%	**30 pb**

↑ hh→WWyy→jjjjyy

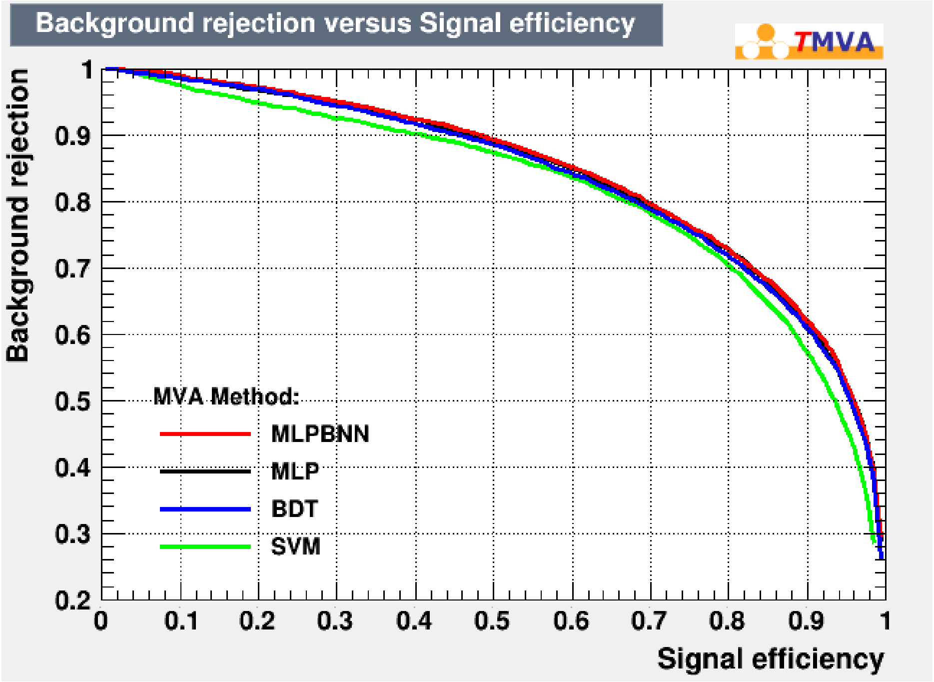
↑ gg → H → hh

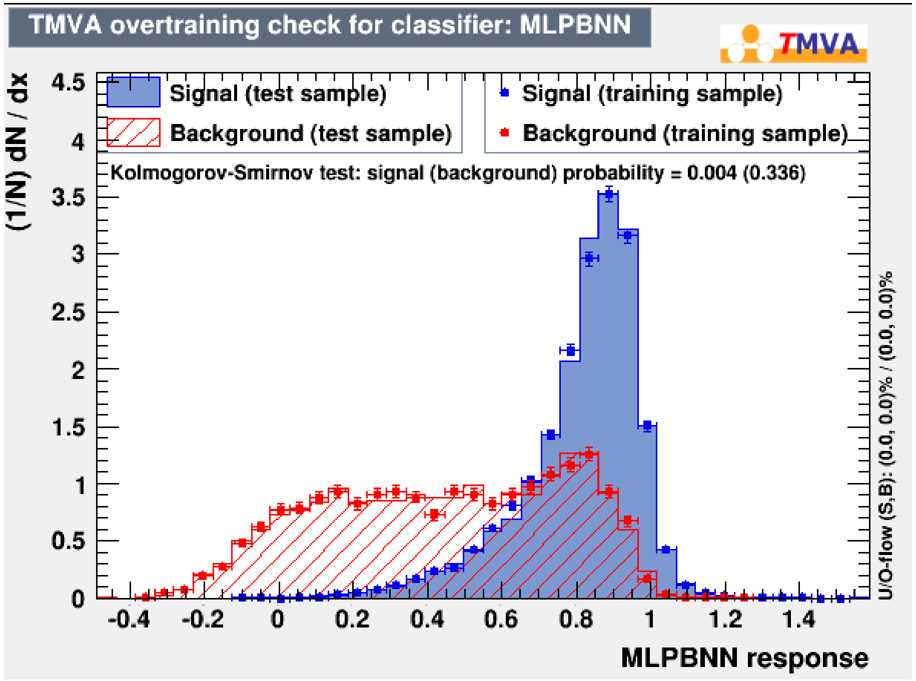
Compared to the exp upper limits from bbyy analysis:
* Non-resonant: 1.0 pb
* Resonant: 1.5 pb

5%

What if more cuts

- The previous estimation is based on yy-cuts and njets>=4
- Apply more cuts on jjjjyy to see how much the upper limits will be improved in order to determine if it is worthy to continue
- Trying to train a BDT/NN/SVM in a quick way to see the performance

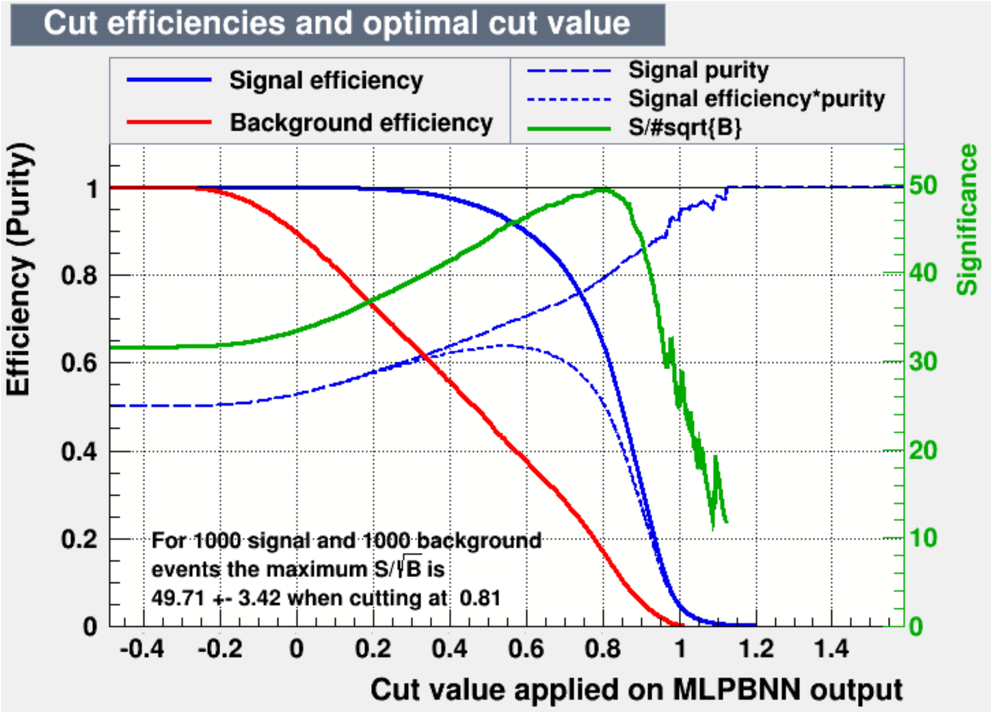




Training samples: signal [MC, ~15K]; bkg [sideband data njets>=2, ~6K]
Test samples are independent of training samples with the same statistics

MVA cuts

- Consider S/sqrt(B), in order to find the MVA cut
  - In this way, S or B are not required to be normalized
  - and S/sqrt(B) is only a relative value



--- ==

--- Classifier (#signal, #backgr.) Optimal-cut S/sqrt(B) NSig NBkg **EffSig** EffBkg

--- ---

--- MLPBNN: (1000,1000) 0.8082 (49.7+-3.42) 627.3654 159.2894 **0.6274** 0.1593

--- ---

A very rough est on upper limits (rev.)

- These est are based on yy cuts && njets>=4 && MVA cut
  - Additional multiply the efficiencies for S (**63**%) and B (**16**%)
  - So bkg evt is now 151 * 16% = 24, leading to a 2sigma fluctuation of ~10 evts

	Lumi (pb-1)	Branching ratio	Cut eff	Upper limit
Non-resonant SM HH	20,000	4.48e-4	15% * 63%	**12 pb**
Resonant 300GeV	20,000	4.48e-4	9% * 63%	**20 pb**

hh→WWyy→jjjjyy

gg → H → hh

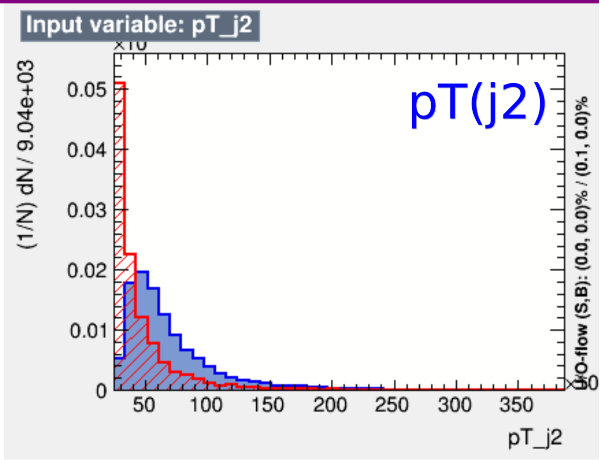
Compared to the exp upper limits from bbyy analysis:
* Non-resonant: 1.0 pb
* Resonant: 1.5 pb

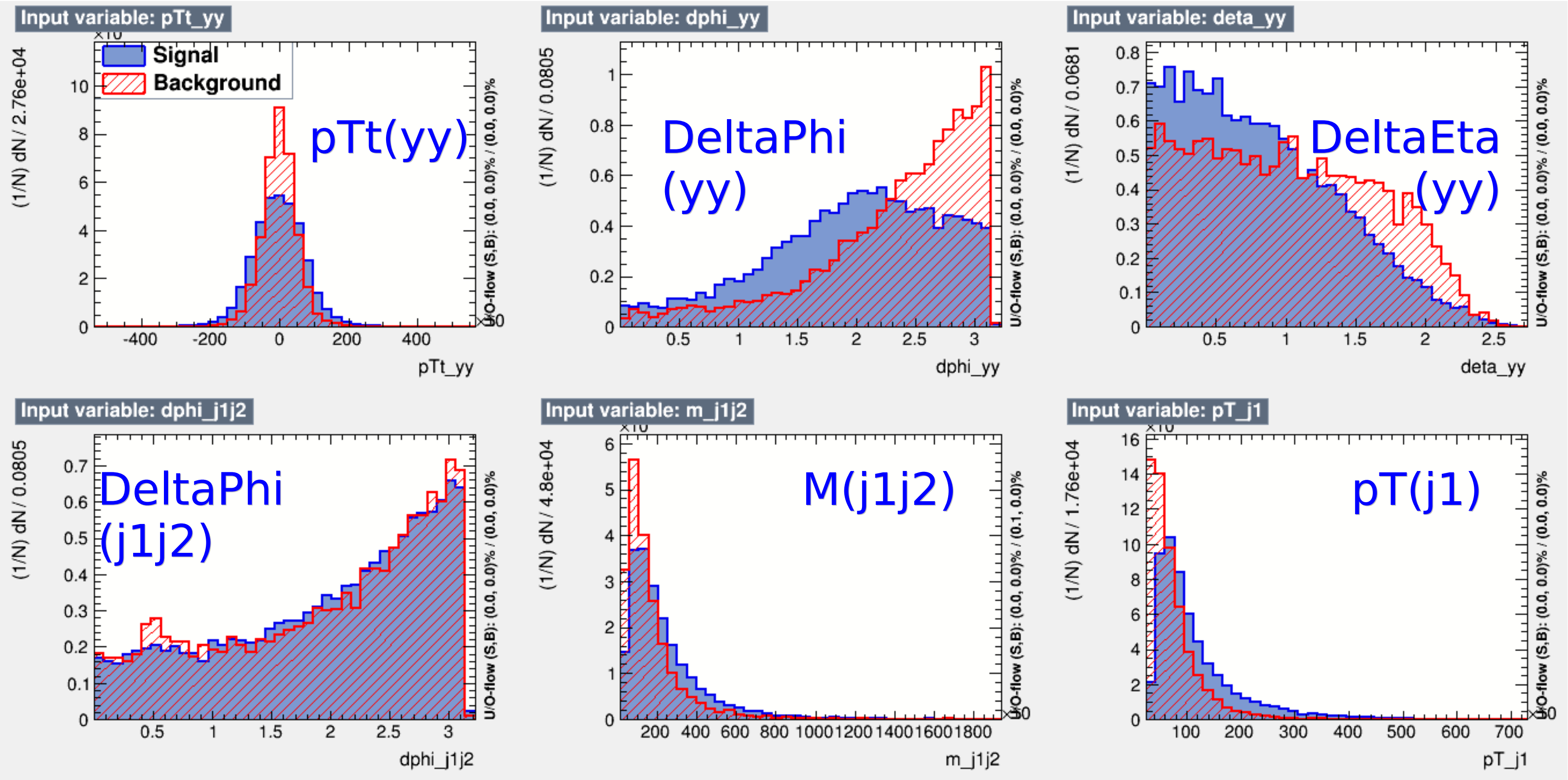
8%

Factor of 2

Input variables

- Input variables are quite preliminary and simple
- More careful choices are needed in future
- For time-being, y-variables and jet-variables
- In future, also need to explore y-jet-variables





JxJy not has to be j1j2

- Till now I only assume j1 j2 are from a real W boson and use them to reconstruct it for simplicity
- An adaptive method is used to improve the correctness of finding the two jets from a same W boson here
    - By asking the invariant mass of jx jy, and choose the pair with the mass closest to W boson mass from PDG

6%

Using j1j2

The correctness of finding jet pair from the same real W boson

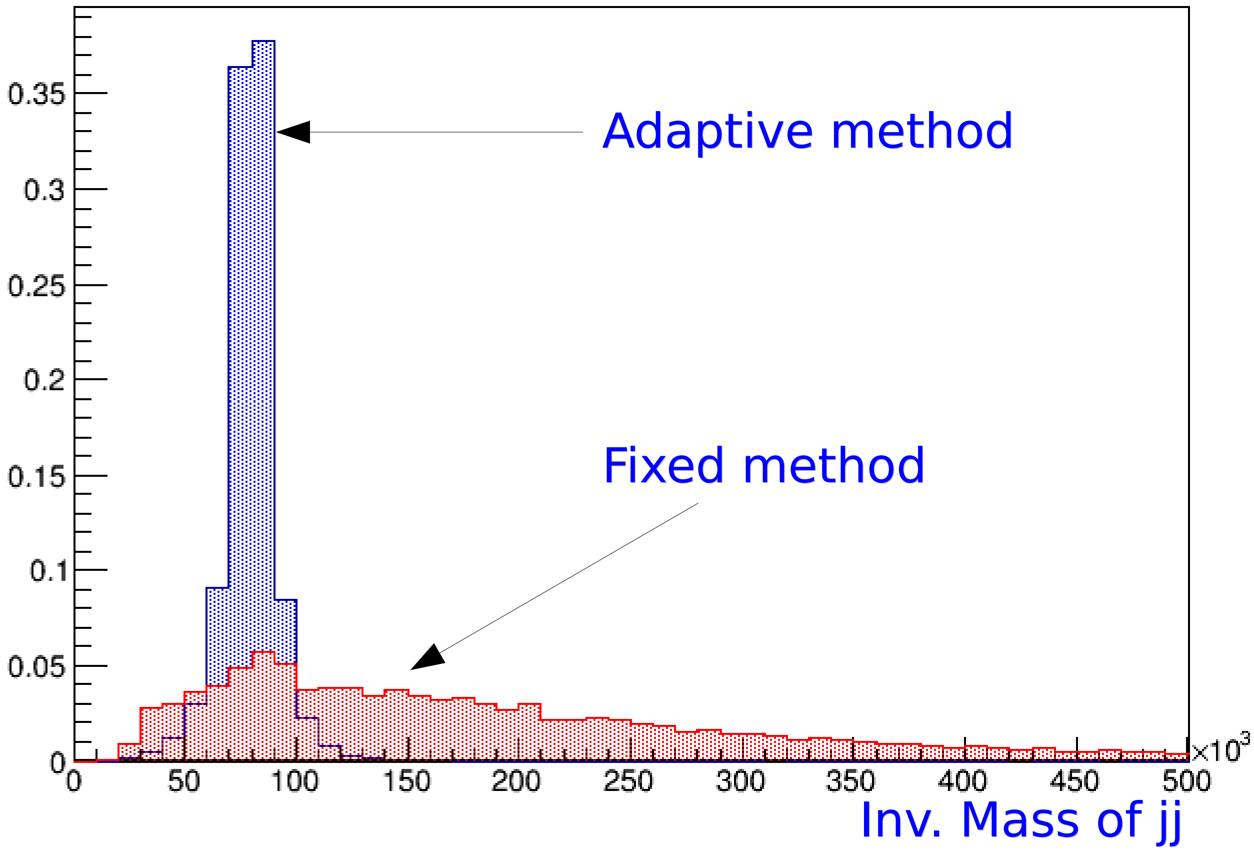
24%

Adaptive jxjy

*these correctness are calculated by using signal MC only @ mH=300GeV

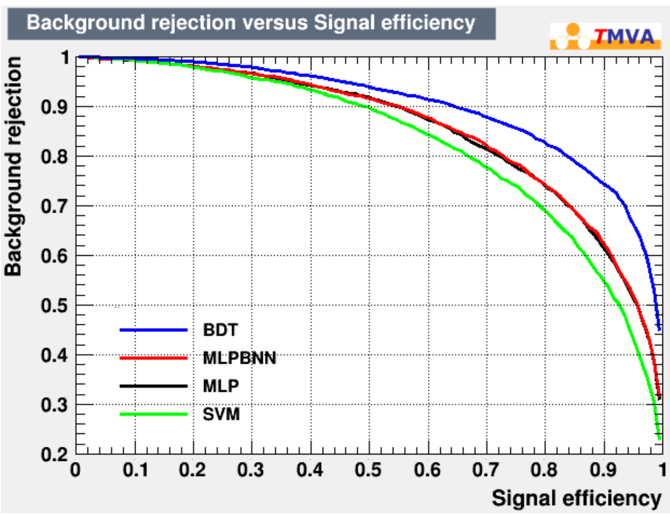
JxJy not has to be j1j2

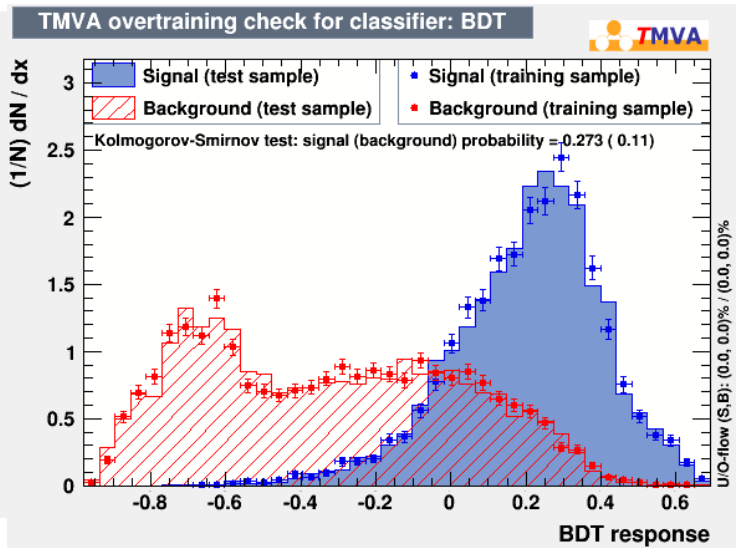
- Compare the invariant mass of jxjy in adaptive method and the one of j1j2 in the fixed method

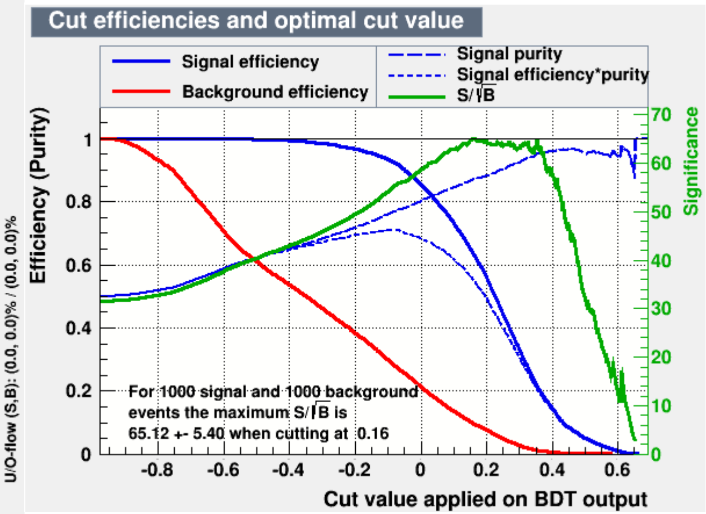


MVA with JxJy

- Train with the variables from JxJy instead of J1J2
- Much more better performance is obtained





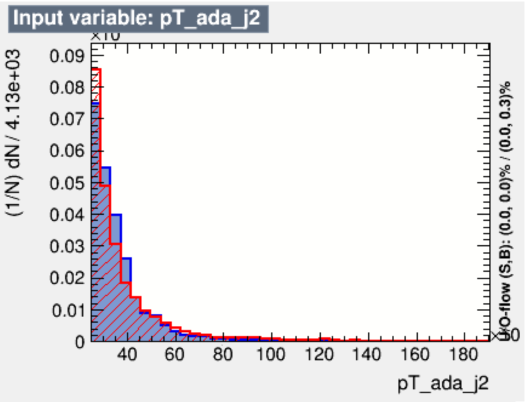


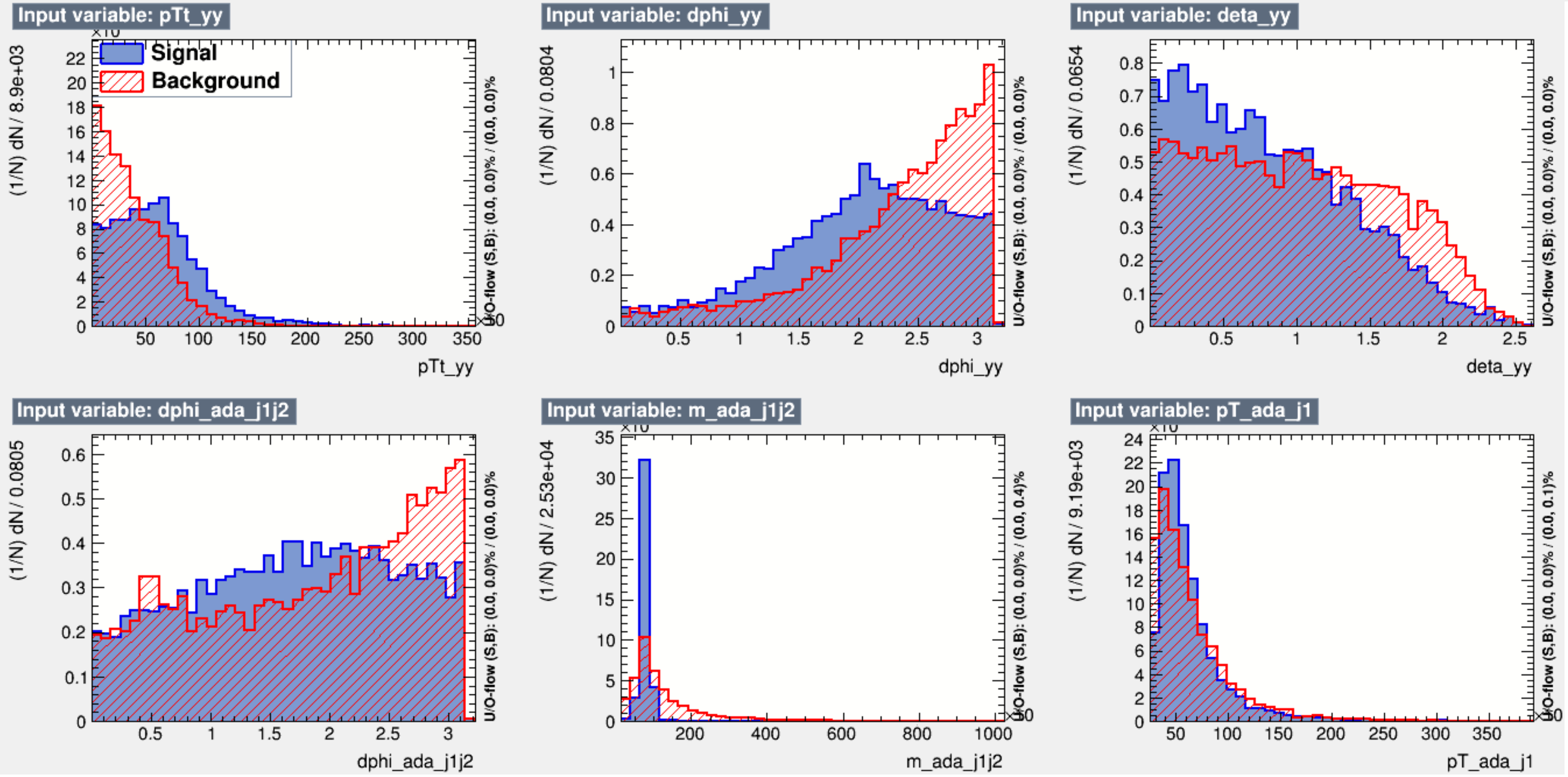

```
- ==========================================================
- Classifier   (  #signal, #backgr.)  Optimal-cut  S/sqrt(B) NSig NBkg   EffSig   EffBkg
- ----------------------------------------------------------------------------------------
-       MLP:  (1000,1000) 0.6192 ( 55 +- 4.8) 515.8762  87.92532   0.5159  0.08793
-    MLPBNN:  (1000,1000) 0.6068 ( 54.9 +- 4.8)  545.217  98.76543   0.5452  0.09877
-       BDT:  (1000,1000) 0.1597 ( 65.1 +- 5.4) 633.2395  94.54983   0.6332  0.09455
- ----------------------------------------------------------------------------------------
```

The signal eff is kept 63% while the bkg eff is lower 9%
The relative S/sqrt(B) is leveled up to 65.1

MVA inputs with adaptive method

- J1J2 variables are replaced by the Jx Jy variables obtained by adaptive method





To-do list (in the order of priority)

- Instead of using j1 j2, find jx jy to be used to reconstruct only one W boson as precisely as possible (preliminarily explored)
- Using these new variables from adaptive method in MVA (trying)
- Introduce more variables into MVA training: y-jet variables, event shape variables (only yy side, only jets side)
- Explain the background components by using Du Chun's samples (jjjj,jjjy,jjyy): reading them now, mimicking the cuts from ATLAS
- Measure eff(m_yy) & uncertainties
- Freeze MVA cuts at some point, and measure eff(MVA) & uncertainties
- Learn and use Hfitter to build up the statistical model
- Start documentation
- Try deep learning if possible

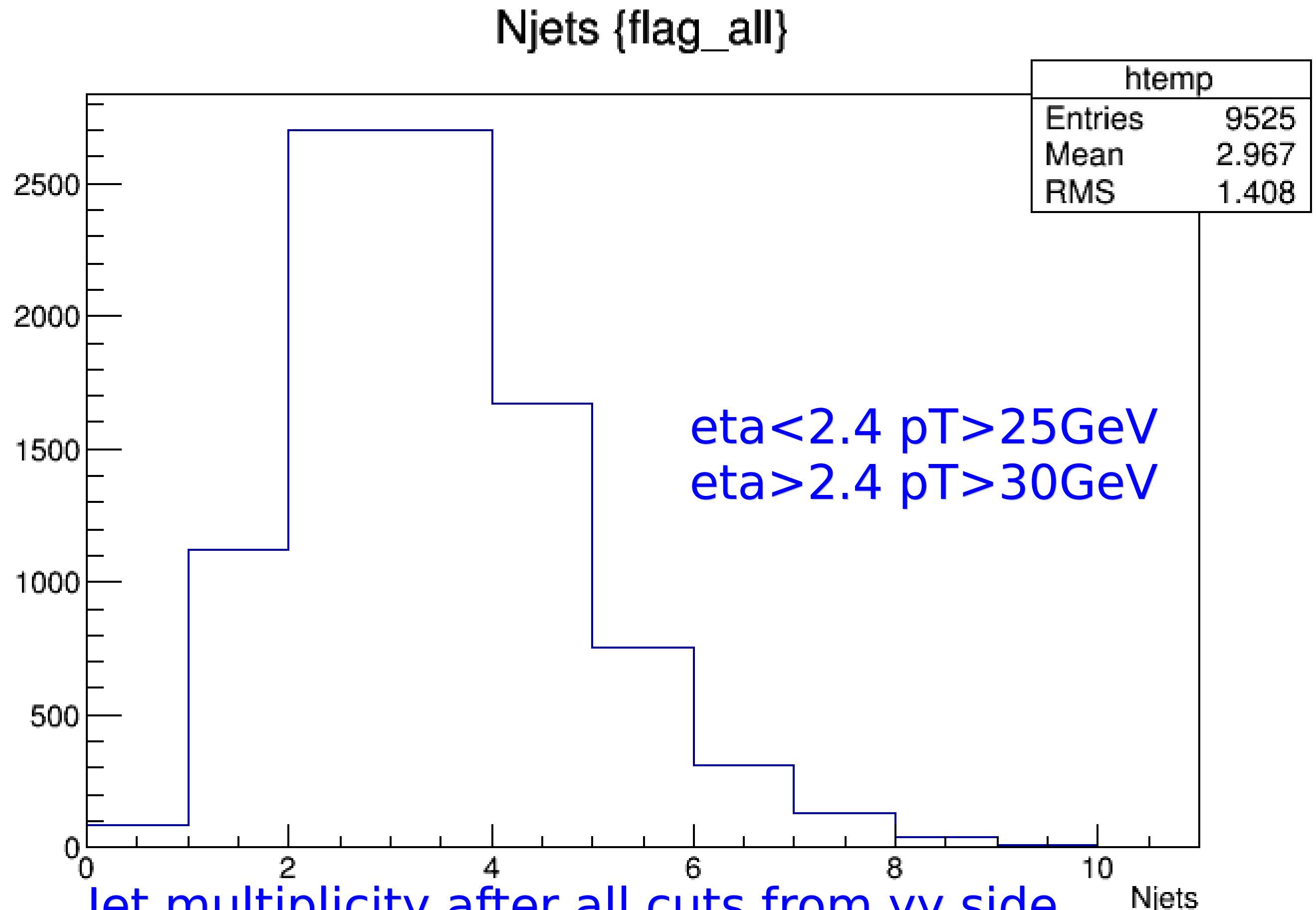
Backup

Introduction

- Sample page

Additional cuts

- Additionally, we will cut on kinematics of the children from the other Higgs boson
  - To do some studies on cuts: deltaPhi(j,j) deltaEta(j,j)
  - To find the more correct combinations of jets originating from the same W boson



Jet multiplicity after all cuts from yy side

Cutflow (mainly yy-cuts)

- Use h → yy trigger and get the gamma pair
- Follow the same preselection in HSG1, check on cutflow (private MC production 28k):

	Unweighted evt	Cut eff (%)
generated	28000	100%
trigger	19953	71%
GRL	19953	71%
detector errors	19953	71%
vertex tracks	19953	71%
pre-selection	15547	56%
photon pT	13996	50%
photon ID	12041	43%
photon isolation	9525	34%
diphoton mass	9516	34%

Additionally	Njets>=4	2913	10%
And	XXX	-	5%

Descriptions of cuts

	descriptions
generated	Generated events
trigger	EF g35 loose g25 loose (8TeV)
GRL	Good Run List
detector errors	LAr error, TileError, BadJet
vertex tracks	Primary vertex track requirements
pre-selection	At least two loose photons
photon pT	1^{st} > 40 GeV, 2^{nd} > 30 GeV
photon ID	IsEM & 0x45fc01 == 0
photon isolation	Calo (etcone40) < 6GeV && track < 2.6GeV
diphoton mass	[100GeV, 160GeV]

BSM gg → H → hh → bbyy

	SM	Resonant NWA				
	Benchmark	260 GeV	300 GeV	350 GeV	500 GeV	1000 GeV
Generated	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
Trigger	73.1%	72.5%	71.6%	71.8%	73.6%	81.0%
Preselection	57.3%	56.7%	56.1%	56.2%	57.7%	65.1%
Photon p_T	51.6%	51.6%	49.8%	49.2%	52.5%	62.4%
Photon Identification	45.3%	44.2%	42.8%	42.6%	46.4%	56.2%
Isolation	39.1%	33.1%	33.8%	35.9%	40.6%	47.4%
$105 < m_{\gamma\gamma} < 160$ GeV	39.0%	33.0%	33.8%	35.9%	40.5%	47.4%
2 Central Jets	33.9%	25.5%	26.9%	29.8%	36.2%	45.1%
Tagging	12.5%	8.4%	8.9%	10.0%	14.1%	19.1%
b p_T Cuts	10.1%	4.8%	5.6%	7.2%	12.0%	18.1%
$95 < m_{b\bar{b}} < 135$ GeV	7.4%	4.0%	4.3%	5.3%	8.6%	14.2%

Reconstructed jets >= 4

- By matching with truth information of jets, one find calculate the correctness of the well-reconstructed jet pairs

specific	#evts	%
none (all cuts form yy side)	2914	-
any pairs from W	1887	65%
1 pair from W	1704	**58%**
2 pairs from W	183	**6%**
the pair from 1st W	1243	43%
the pair from 2nd W	827	28%

*all numbers except the case of 'none' are calculated by asking njets>=4
pT > 25 GeV when |eta| < 2.4
pT > 30 GeV when |eta| > 2.4

So, instead of asking for >=4 jets and trying to reconstruct both Ws, one may keep higher statistics and obtain better sensitivities by
Looking also 2,3,>=4 jets and reconstruct only one of the Ws from 2 jets

Reconstructed jets **>=2**

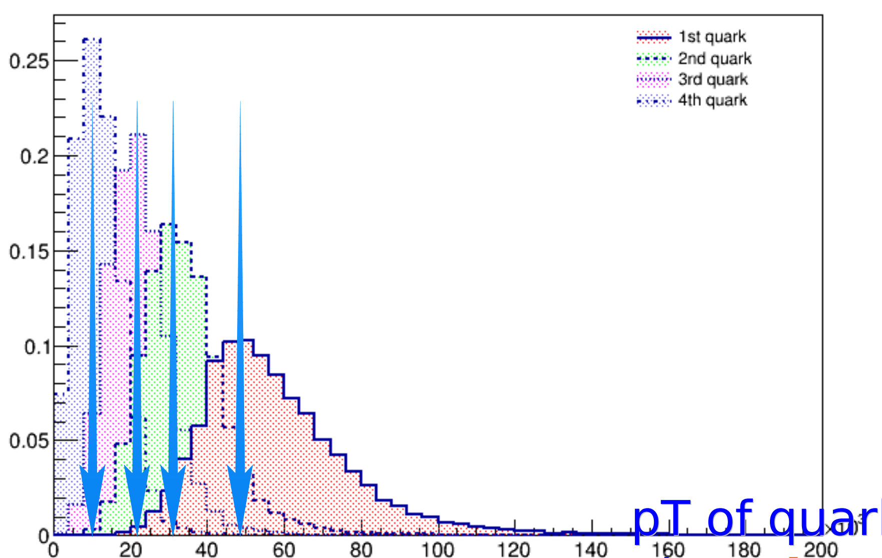
- Redo the cutflow and the correctness tables

	Unweighted evt	Cut eff (%)
generated	28000	100%
trigger	19953	71%
GRL	19953	71%
detector errors	19953	71%
vertex tracks	19953	71%
pre-selection	15547	56%
photon pT	13996	50%
photon ID	12041	43%
photon iso	9525	34%
diphoton mass	9516	34%
Njets>=2	**8312**	**30%**

specific	**#evts**	**%**
none (all cuts form yy side)	8312	-
any pairs from W	3909	47%
1 pair from W	3726	**45%**
2 pairs from W	183	**2%**
the pair from 1st W	2631	32%
the pair from 2nd W	1461	18%

BSM gg → H → hh → WWyy → jjjjyy

- Triggered by the recent results on H → hh → bbyy search
  - There is a significance of 3.0 standard deviations at mH=300GeV
- We look at second largest decay: H → hh → WWyy → jjjjyy
- Look at the pT of jets in parton level, find it quite difficult to reconstruct all four jets



pT of quarks at **parton** lvl

pT threshold	evts(parton lvl)	efficiencies
non	19430	100%
5 GeV	17168	88%
10 GeV	11389	59%
15 GeV	5437	28%
20 GeV	1963	10%
25 GeV	**594**	**3%**

A first look at sideband region

Njets>=4

- Sideband region:
  - mass(yy) within [100,160] GeV
  - mass(yy) is excluded from $|m_h - \Delta m_h - m_{\gamma\gamma}| < 2\sigma_{\gamma\gamma}$
  - where mh=125.6, deltamh=0.15, sigma=1.6

sideband	# of evt
ggH	0.467175
VBF	0.123474
WH	0.0638113
ZH	0.0405459
ttH	0.138622
Continuum	?
In data	1170

Continuum?

There are large components in backgrounds not yet clear
Need to at least introduce pp→jjjjyy and pp→jjyy

Bkg samples, use bbyy continuum samples?

A first look at signal region

Njets>=4

- Signal region:
  - mass(yy) is required by $|m_h - \Delta m_h - m_{\gamma\gamma}| < 2\sigma_{\gamma\gamma}$
  - where mh=125.6, deltamh=0.15, sigma=1.6

sideband	# of evt
ggH	4.91724
VBF	1.0963
WH	0.570564
ZH	0.374228
ttH	1.34295
Continuum	?
est bkg*	143

* bkg in signal region estimated by fitting to exponential with sideband data