Probing electroweak symmetry breaking with Higgs pair production at the LHC

Juan Rojo
STFC Rutherford Fellow
Rudolf Peierls Center for Theoretical Physics
University of Oxford

Higgs as Probe and Portal (HPP) meeting
Nikhef, 30/09/2016

Probing electroweak symmetry breaking with Higgs pair production at the LHC

Juan Rojo
VU Amsterdam & Theory group, Nikhef

from tomorrow!

Higgs as Probe and Portal (HPP) meeting
Nikhef, 30/09/2016

Fingerprinting the Higgs sector

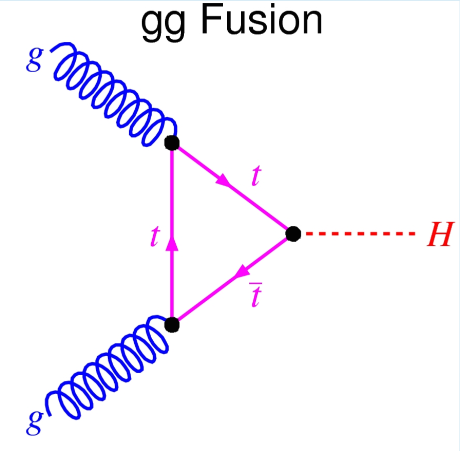
In the **Standard Model** the properties of the Higgs sector are **uniquely determined**

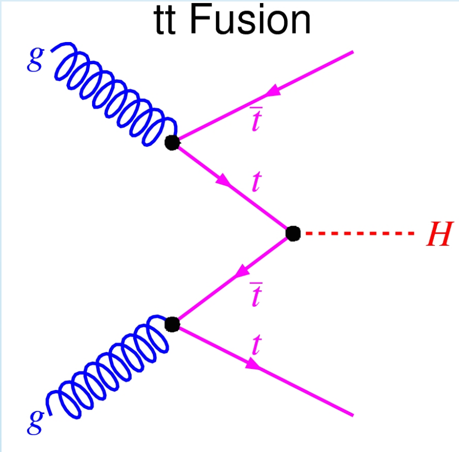
Any deviation from the tight SM predictions would be a **smoking gun for Physics beyond the SM**

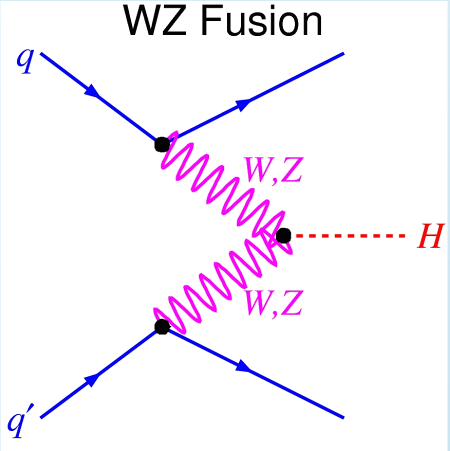
BSM model	Deviations in Higgs coupling to		
	W, Z weak bosons	bottom quarks	photons
New heavy Higgs boson	6%	6%	6%
Two-Higgs Doublet model	1%	10%	1%
Composite Higgs	-3%	-9%	-9%
New heavy top-like quark	-2%	-2%	+2%

A ***precision of a few percent*** *in* ***Higgs couplings measurements*** *is the goal!*

This precision required both in **experimental data** and in **theory calculations** of **Higgs production**

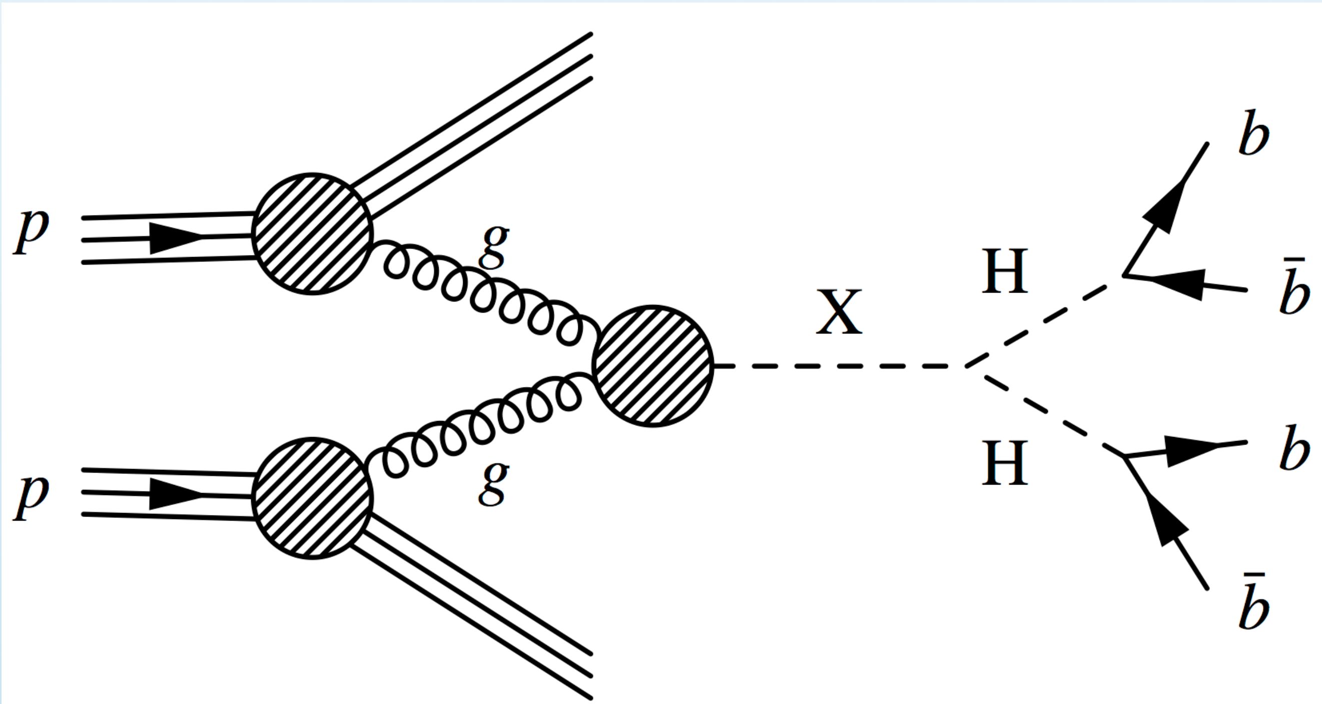






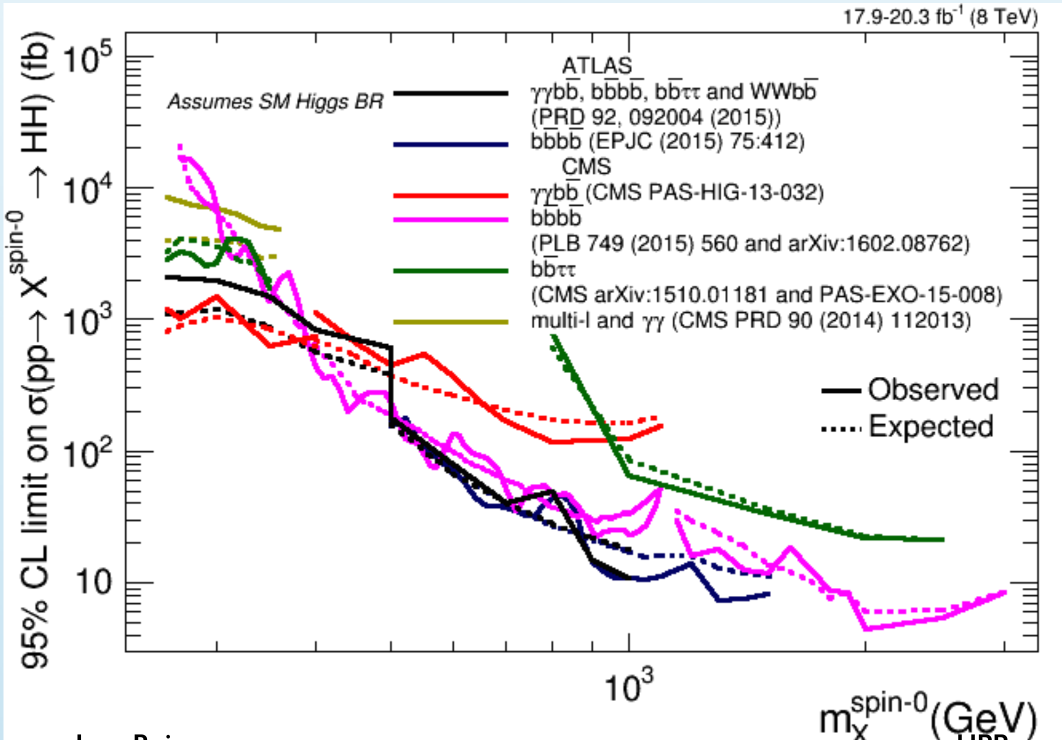
Higgs Pair Production at the LHC

- **Double Higgs production** allows accessing **crucial components of the Higgs sector**:
  - ☑ Reconstruct the **full electroweak symmetry breaking potential**
  - ☑ Probe the **Higgs self-interaction**
  - ☑ Probe the **doublet nature** of the Higgs by means of the **hhVV coupling**
- In the SM, **hh rates are small:** in the leading gluon-fusion production mode, the **cross-section at 14 TeV** is only **40 fb,** further suppressed by branching fractions
- Rates for double Higgs production **generically enhanced in BSM scenarios**, and **LHC searches** in various final states have already started at **Run I**

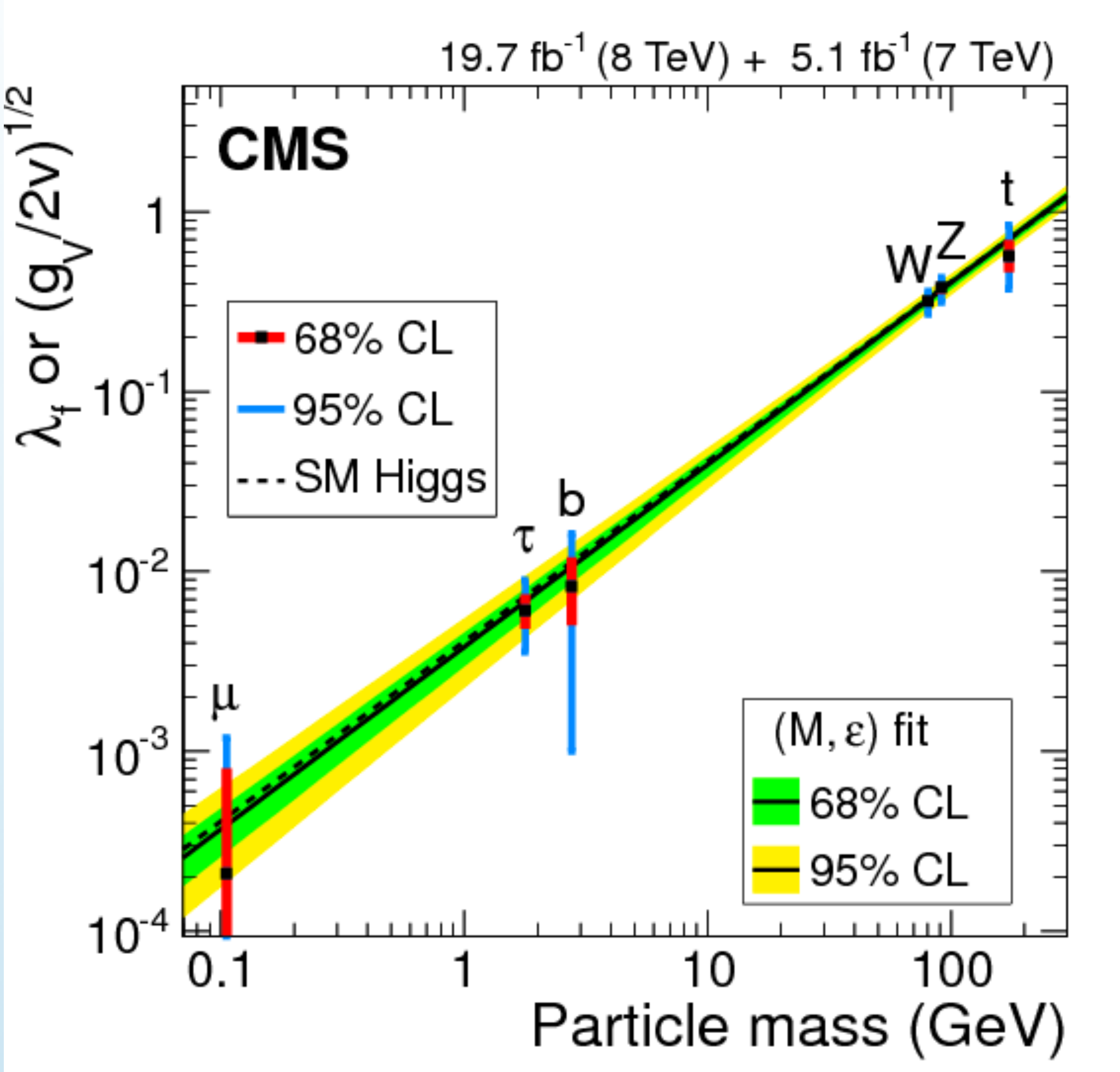


Higgs Pair Production at the LHC

- **Double Higgs production** allows accessing **crucial components of the Higgs sector**:
  - ☑ Reconstruct the **full electroweak symmetry breaking potential**
  - ☑ Probe the **Higgs self-interaction**
  - ☑ Probe the **doublet nature** of the Higgs by means of the **hhVV coupling**
- In the SM, **hh rates are small:** in the leading gluon-fusion production mode, the **cross-section at 14 TeV** is only **40 fb,** further suppressed by branching fractions
- Rates for double Higgs production **generically enhanced in BSM scenarios**, and **LHC searches** in various final states have already started at **Run I**



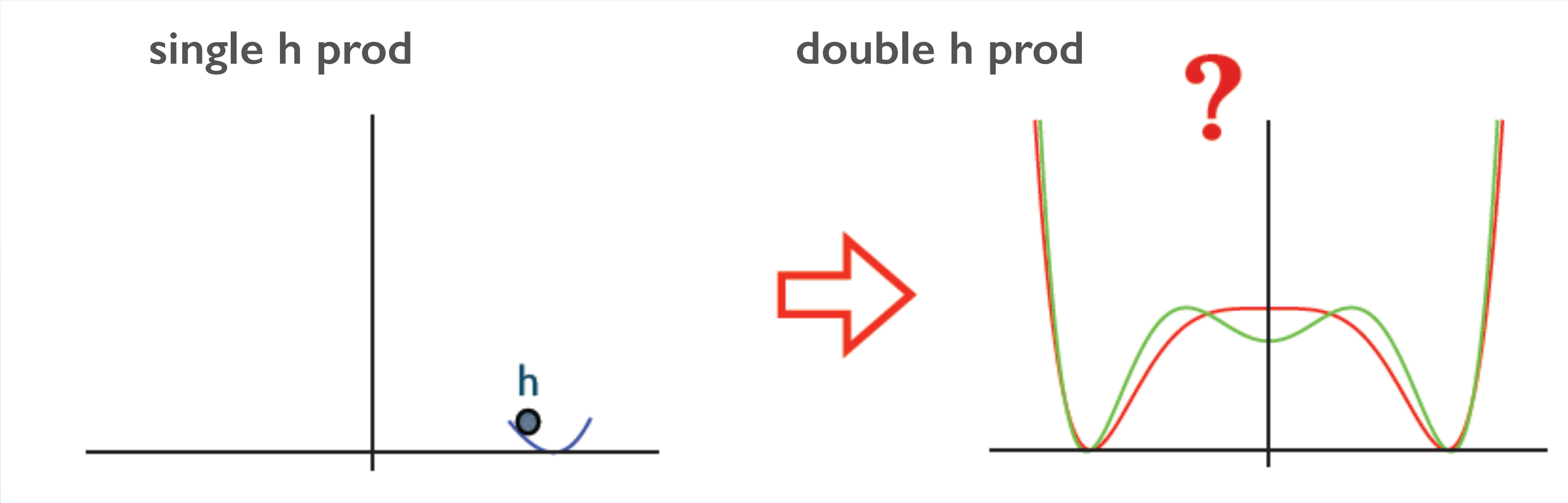
EW symmetry breaking: what do we know



- **Yukawa/Couplings** between Higgs and SM particles **proportional to mass**
- The Higgs boson is **responsible to break EW symmetry and give particles mass**
- However, we still lack understanding of ***why and how*** EWS is broken
- What are the **dynamics** underlying EW symmetry breaking?

EW symmetry breaking: what we don't know

- **Current measurements** (couplings in single Higgs production) probe **Higgs potential close to minimum**
- **Double Higgs production** essential to **reconstruct the full Higgs potential** and clarify EWSB mechanism
- The Higgs potential is *ad-hoc*: **many other EWSB mechanisms conceivable**



Higgs mechanism

$$V(h) = m_h^2 h^\dagger h + \frac{1}{2}\lambda(h^\dagger h)^2$$

Coleman-Weinberg mechanism

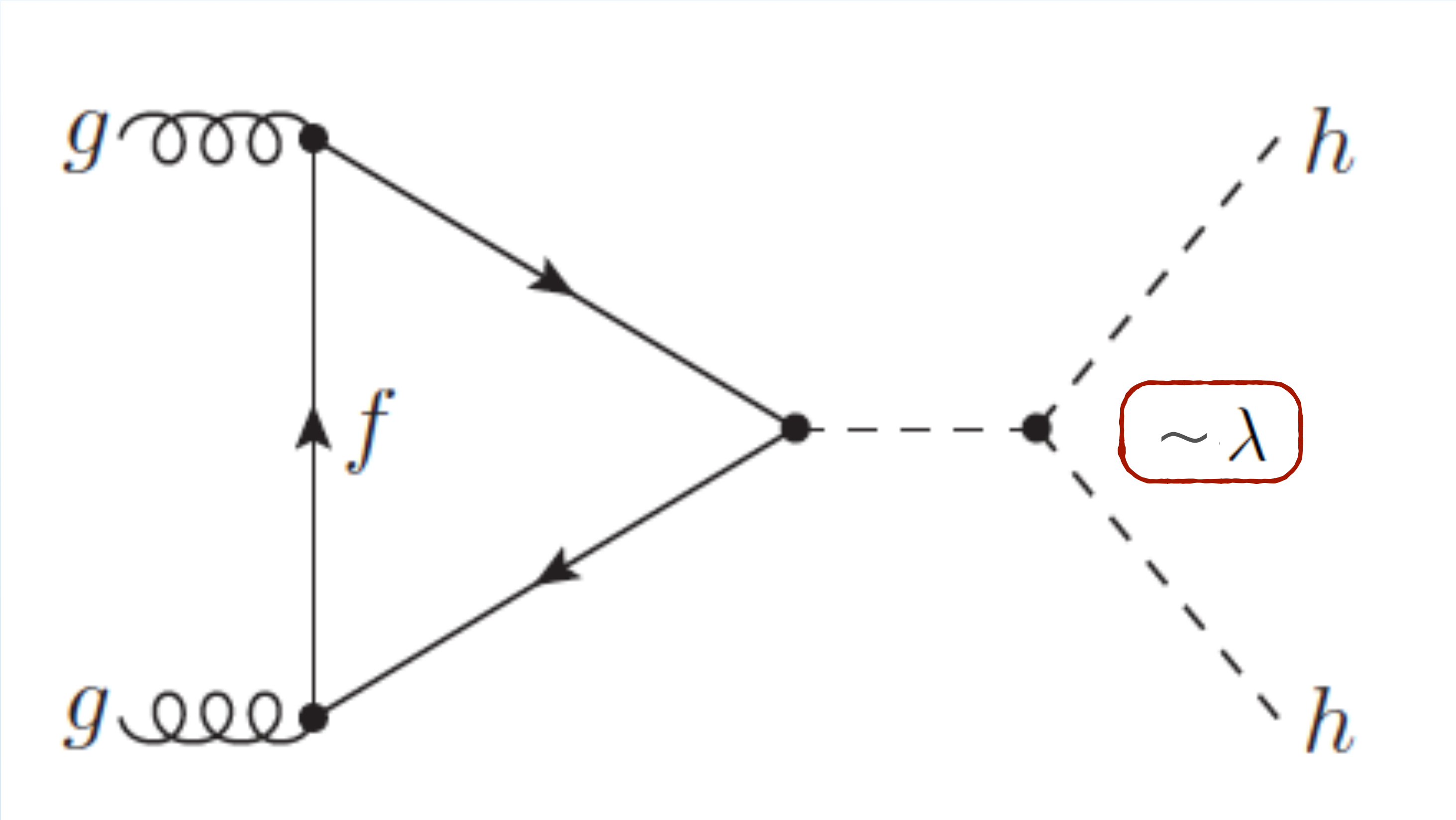
$$V(h) \to \frac{1}{2}\lambda(h^\dagger h)^2 \log\left[\frac{(h^\dagger h)}{m^2}\right]$$

Each possibility associated to **completely different EWSB mechanism**, with crucial implications for the **hierarchy problem**, the structure of quantum field theory, and **New Physics at the EW scale**

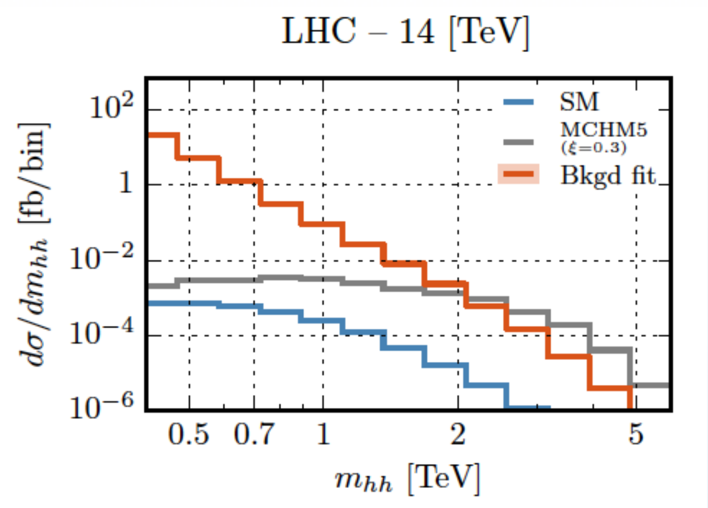
Arkani-Hamed, Han, Mangano, Wang, arxiv:1511.06495

EW symmetry breaking: what we don't know

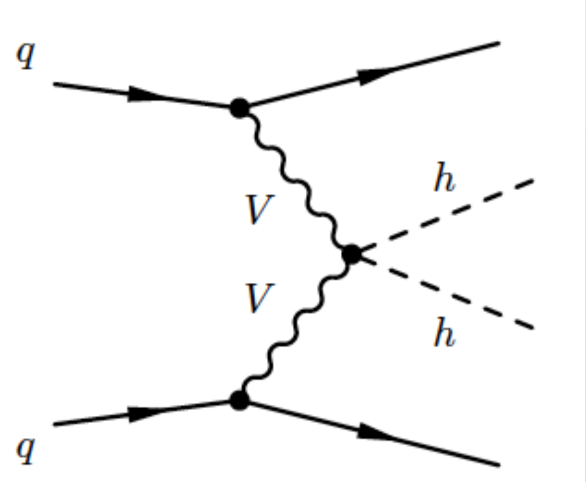
- **Current measurements** (couplings in single Higgs production) probe **Higgs potential close to minimum**
- **Double Higgs production** essential to **reconstruct the full Higgs potential** and clarify EWSB mechanism
- The Higgs potential is *ad-hoc*: **many other EWSB mechanisms conceivable**



$$V(h) = m_h^2 h^\dagger h + \frac{1}{2}\lambda(h^\dagger h)^2$$



Higgs pair production by vector-boson fusion in the 4b final state



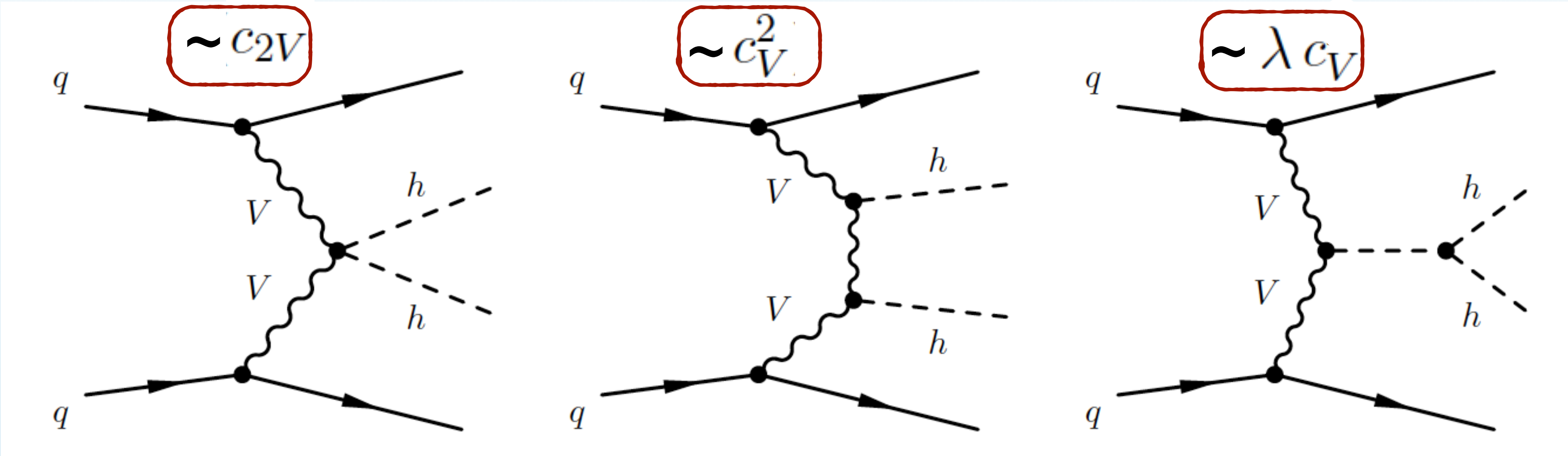
- *F. Bishara, R. Contino and J. Rojo, in preparation*

see also:

- *J. K. Behr, D. Bortoletto, J. A. Frost, N. P. Hartland, C. Issever and J. Rojo, EPJC76, no. 7, 386 (2016), arXiv:1512.08928*
- *M. Gouzevitch, A. Oliveira, J. Rojo, R. Rosenfeld, G. P. Salam and V. Sanz, JHEP 1307, 148 (2013), arXiv:1303.6636*

EW symmetry breaking: what we don't know

- In the absence of the Higgs boson, the amplitude for **vector-boson scattering** (VBS) **grows** with the partonic center-of-mass energy, until eventually **unitarity is violated**
- In the SM, the Higgs boson **unitarizes the high-energy behaviour** of VBS amplitudes



at high energies

$$\mathcal{A}(V_L V_L \to hh) \simeq \frac{\hat{s}}{v^2}(c_{2V} - c_V^2)\,,$$

Is this **cancellation exact** (as in SM, $c_{2V} = c_V^2$) or only **approximate (BSM, $c_{2V} \neq c_V^2$)?**

No model-independent information on c_{2V} available so far at the LHC

Even for small deviation of the SM couplings, **striking signals within the reach of Run II!**

Exploiting the VBF channel for di-Higgs

- Signal generation with **MadGraph5** using customized UFO model
- **Event rates** can increase by **up to a factor 30** as compared to SM if new physics is present
- At a 100 TeV collider, **10^5 events before cuts** even for SM couplings

Signal: VBF $hh \to b\bar{b}b\bar{b}$

$\{c_V, c_{2V}, c_3\}$		LHC 14 TeV σ (fb)	LHC 14 TeV $N_{ev}(\mathcal{L} = 3\,\text{ab}^{-1})$	FCC 100 TeV σ (fb)	FCC 100 TeV $N_{ev}(\mathcal{L} = 10\,\text{ab}^{-1})$
{1,1,1}	SM	0.26	780	14.8	$1.5 \cdot 10^5$
{1,0,1}		4.4	$1.3 \cdot 10^4$	593	$5.9 \cdot 10^6$
{1,2,1}		2.5	$7.5 \cdot 10^3$	471	$4.7 \cdot 10^6$
{1,0,0}		5.8	$1.7 \cdot 10^4$	656	$6.6 \cdot 10^6$
{1,0,-1}		7.5	$2.3 \cdot 10^4$	731	$7.3 \cdot 10^6$
{1,1,0}		0.64	$1.9 \cdot 10^3$	29.8	$3.0 \cdot 10^5$
{0.84,0.40,0.48}	MCHM5 $\xi = 0.3$	0.78	$2.3 \cdot 10^3$	75.7	$7.6 \cdot 10^5$

Exploiting the VBF channel for di-Higgs

- Generation of **QCD multijet backgrounds** highly CPU time-intensive
- Generated at LO with **Sherpa** (weighted and unweighted events), **cross-checked with ALGEN**
- **Gluon-fusion di-Higgs** production now background to VBF production
- The **irreducible 4b multijet background** is **seven orders of magnitude larger than the SM signal** at the generation level. How to overcome this huge difference?

Background processes						
Process	Program	Generation	$\sigma_{\rm LO}$ (fb)		K–factor	
			LHC14	FCC100	LHC14	FCC100
$4b$	`Sherpa2.2`	$N_{\rm ev}$ = 50M weighted	$1.1 \cdot 10^6$	$1.6 \cdot 10^7$	1.7	1.7
$2b2j$	`Sherpa2.2`	$N_{\rm ev}$ = 50M weighted	$2.6 \cdot 10^8$	$3.8 \cdot 10^9$	1.3	1.3
$t\bar{t}jj$	`Sherpa2.2`	$N_{\rm ev}$ = 10M weighted	$1.9 \cdot 10^4$	$1.6 \cdot 10^6$	1.6	1.6
$4b2j$	`ALPGEN`	$N_{\rm ev}$ = 6M(2M) unweighted	$5.4 \cdot 10^4$	$2.4 \cdot 10^6$	1.7	1.7
$2b4j$	`ALPGEN`	$N_{\rm ev}$ = 260k unweighted	10^7	$5.2 \cdot 10^8$	1.3	1.3
$gg \to hh \to b\bar{b}b\bar{b}$	`aMC@NLO`	$N_{\rm ev}$ = 1M unweighted	6.2	272	2.4	2.2

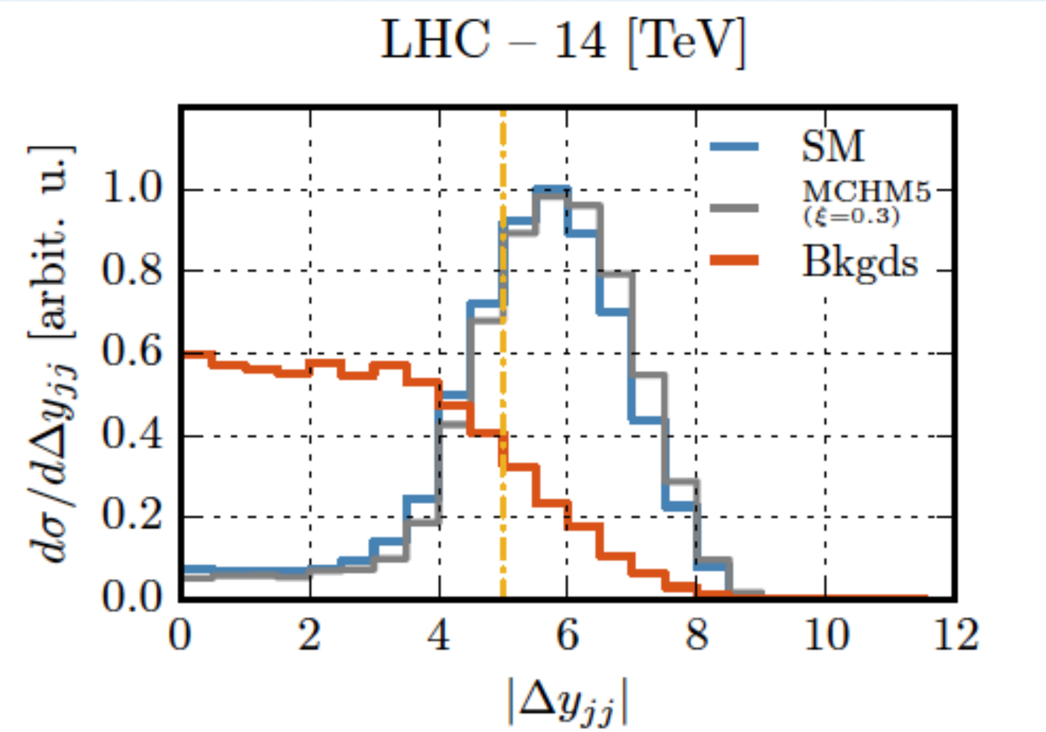
Killing backgrounds with VBF topology

The huge QCD jet backgrounds can be **reduced by exploiting the VBF topology**

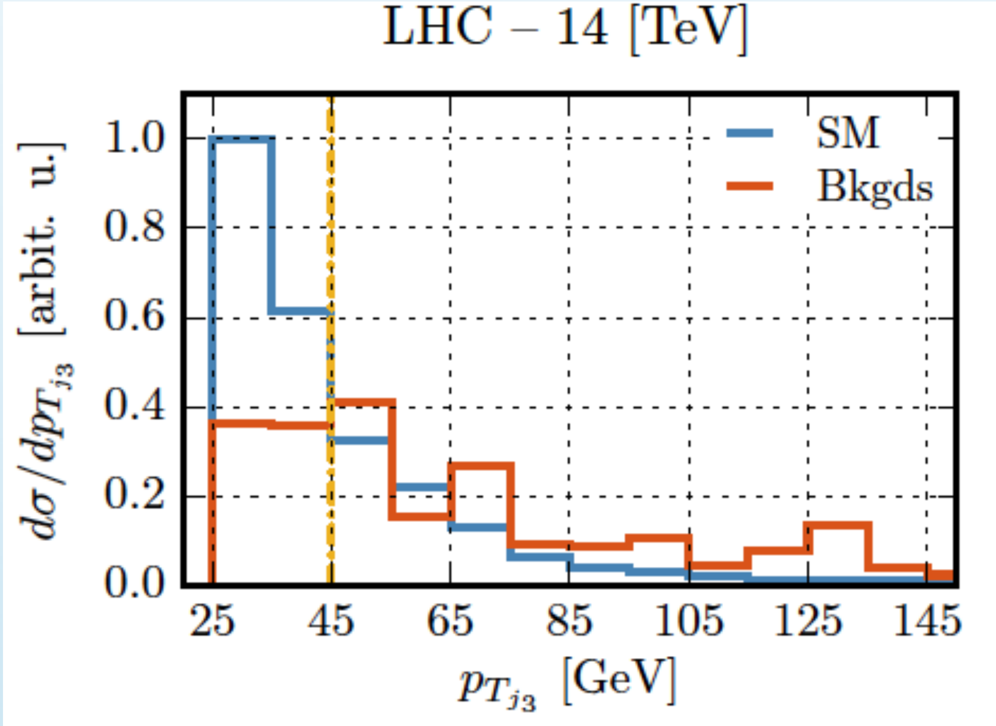
Require **two forward jets**, **separated in rapidity**, plus a **veto in hadronic activity** in the central region

Additional cuts in the **di-Higgs invariant mass m_{hh}** further reduce the QCD multijet cross-sections

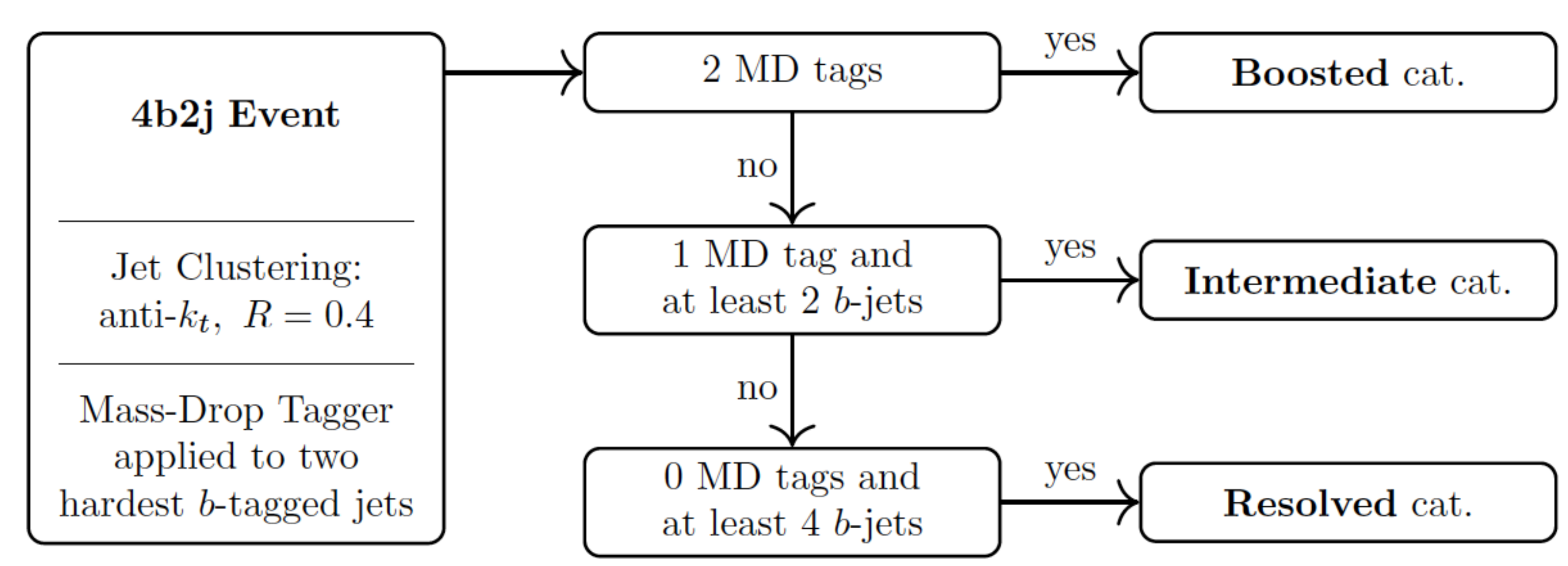
Dijet rapidity separation cut



Central jet veto cut

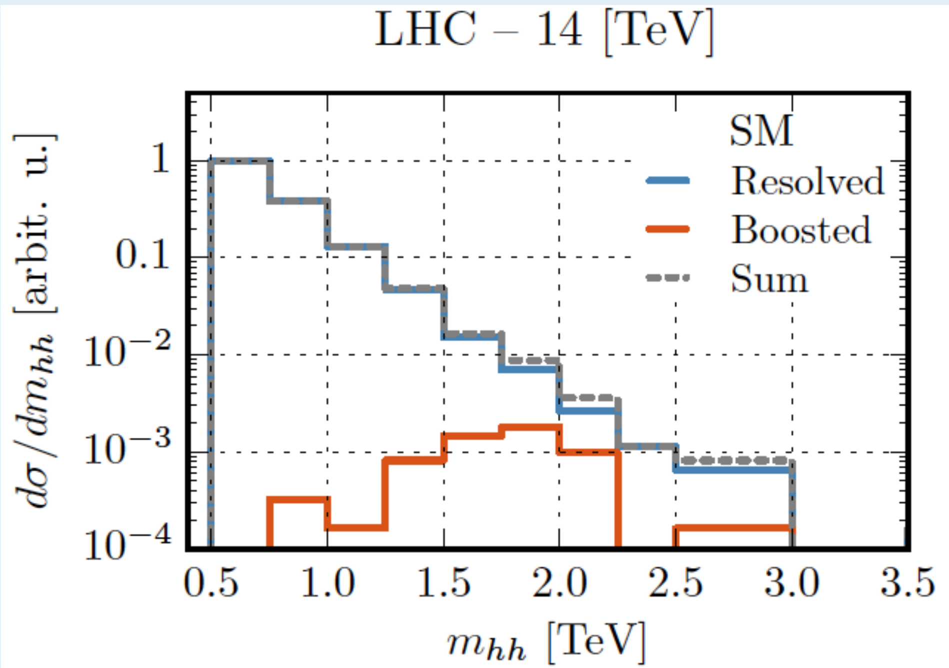


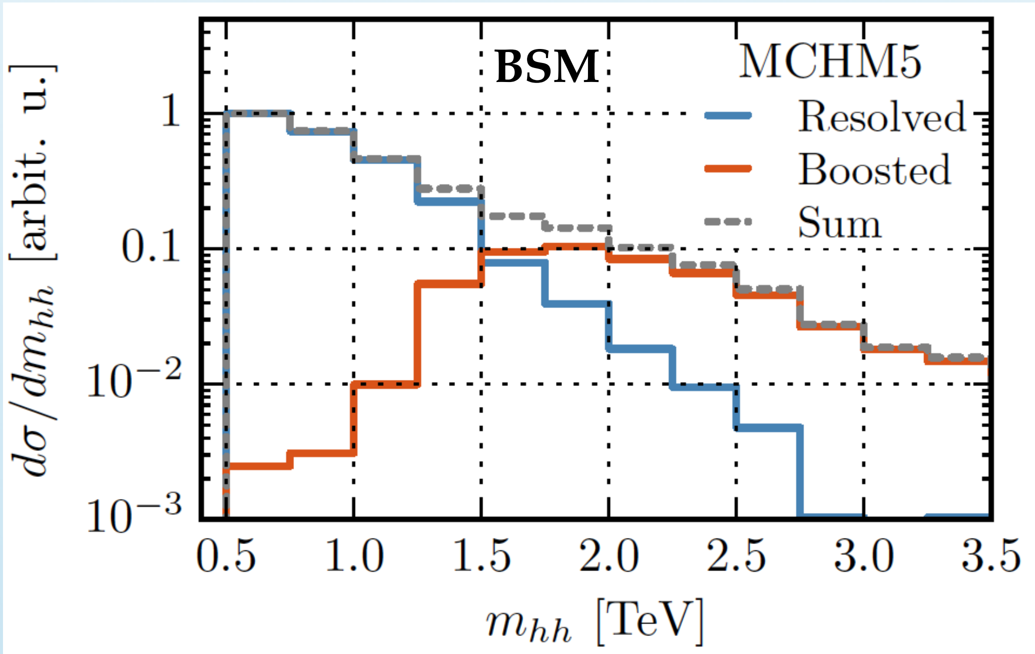
Scale-invariant Tagging



Determine, **event-by-event**, degree of **boost of di-Higgs system** and optimise selection accordingly

In the **SM, resolved selection** dominates, but for **BSM couplings**, **boosted topology** important



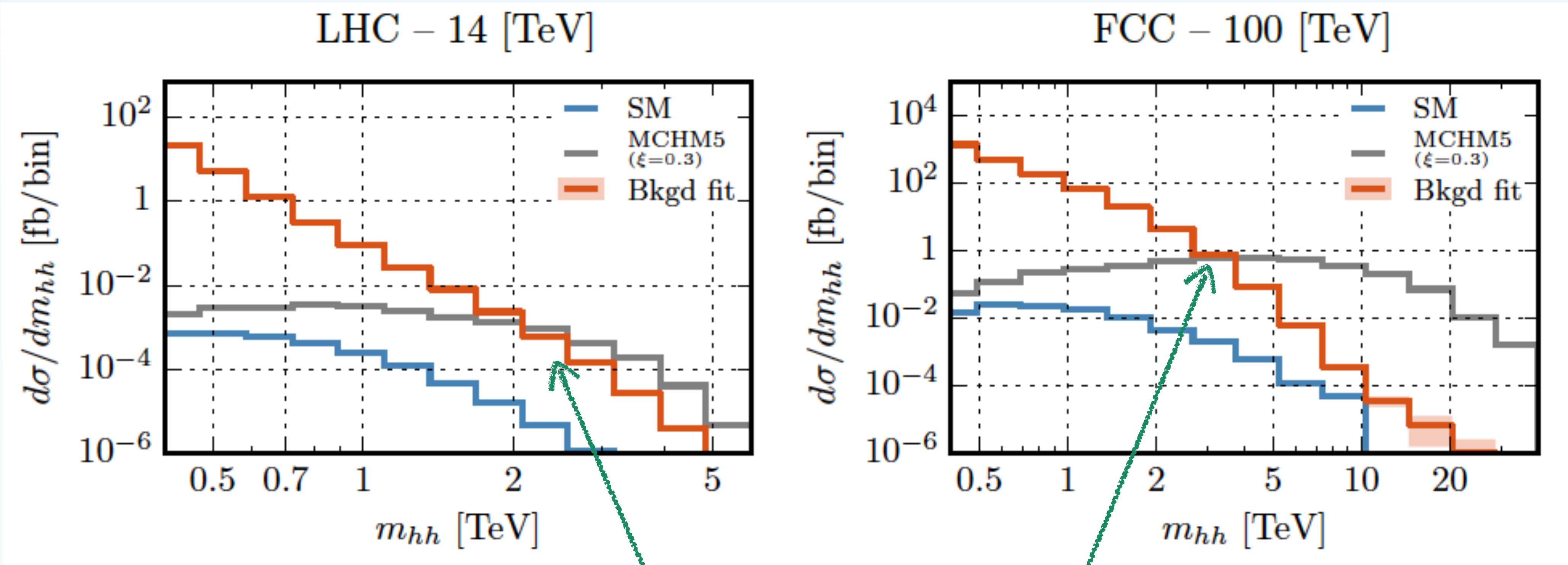


Probing the high-energy regime

After selection and analysis cuts, backgrounds are still **overwhelmingly large for m_{hh} close to threshold**

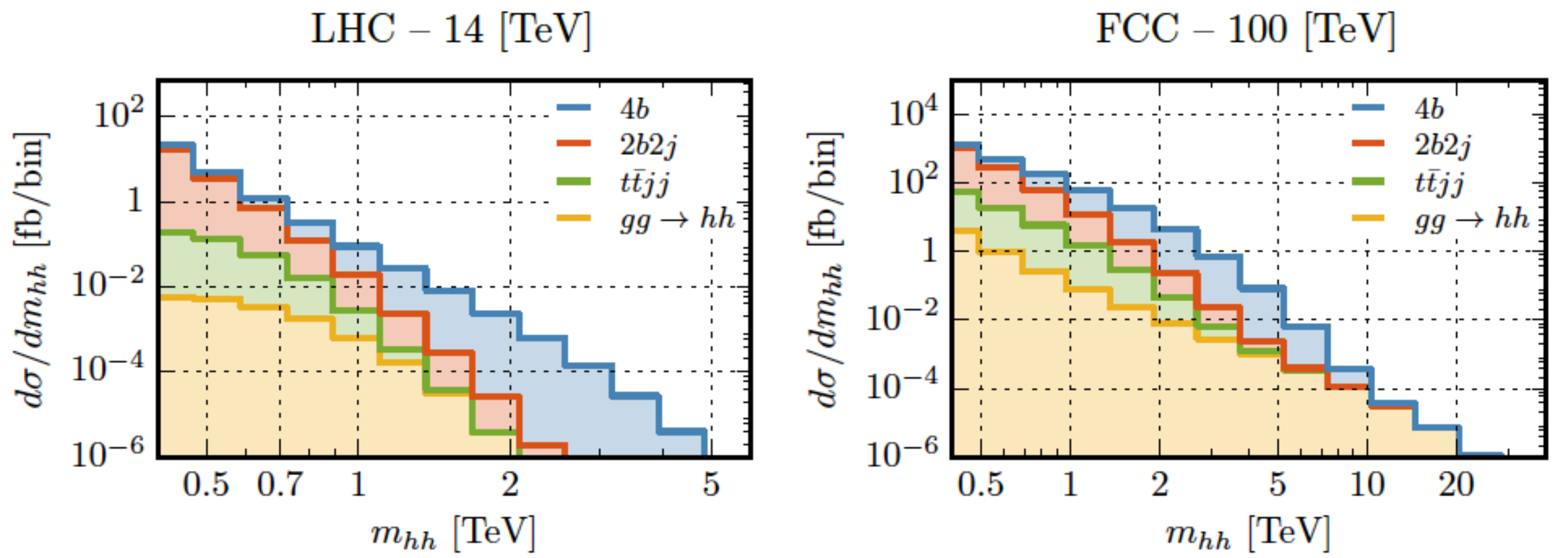
For BSM couplings, the ratio of cross-sections **between BSM and SM increases dramatically** as we probe the regions of **large m_{hh}**, eventually dominating over backgrounds

This is the key of **the sensitivity to c_{2V}** despite the tiny SM cross-sections



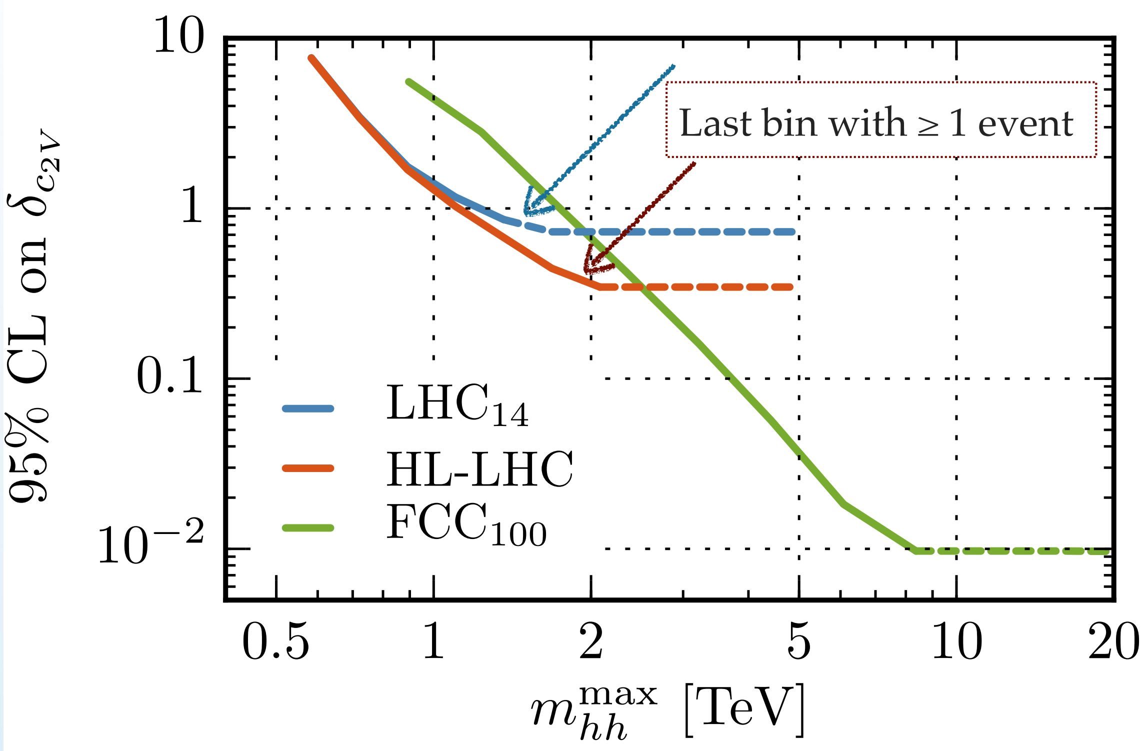
BSM signal (MCHM5 benchmark point) dominates over background

Background decomposition



- At 14 TeV, contribution from **reducible 2b2j multijet** backgrounds (mistagged light jets) dominate up to **m_{hh} ~800 GeV**, when the **irreducible 4b background** becomes the most important
- Same trend at 100 TeV, with now **gg>hh becoming dominant** at highest invariant masses, **m_{hh} ~8 TeV**
- Reliability of QCD multijet background estimation has been carefully validated with **two independent MC generators, Sherpa and ALPGEN**

Sensitivity to the hhVV coupling



95% CL on $\delta_{c_{2V}}$

$$\mathrm{LHC}_{14} : [-0.59, 0.66]$$
$$\text{HL-LHC} : [-0.21, 0.24]$$
$$\mathrm{FCC}_{100} : [-0.01, 0.01].$$

The **sensitivity to c_{2V}** improves significantly the **higher the values of m_{hh} that can be probed**

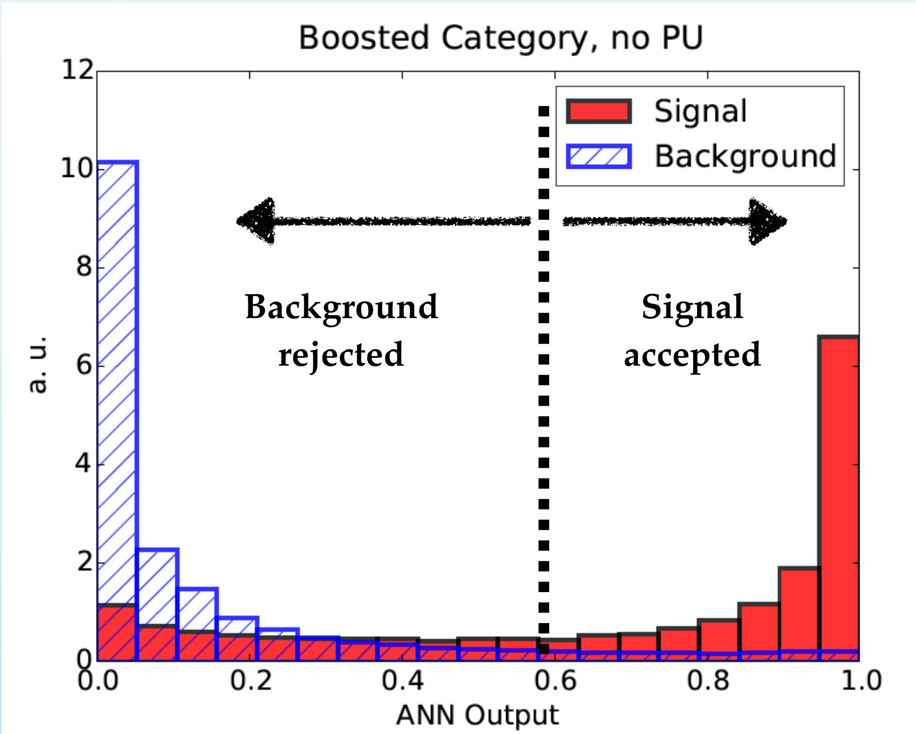
In the absence of new resonances, **c_{2V}** can be constrained **down to 60% (20%) of the SM value at the 95%CL** at the LHC (HL-LHC), assuming SM couplings

Take-away message

Searches for **di-Higgs production in the vector-boson-fusion channel** should start already during **Run II**, without waiting for the HL-LHC!

The di-Higgs frontier at the LHC

- Higgs pair production is a **cornerstone of the LHC program** for the coming years, allowing to reconstruct the **EWSB potential** and to test the nature of the **EWSB mechanism**
- The **4b final state** offers the highest yields, but requires taming the overwhelming QCD background
- In the VBF channel, the **steep rise of the cross-section with the di-Higgs invariant mass in the case of deviations from the SM couplings** is the key for the high sensitivity of this process
- Strong motivation to start to **search for VBF di-Higgs now**, no need to wait until the HL-LHC!
- The ultimate optimization of the extraction of the Higgs couplings in VBF might require the use of **multivariate techniques**, as we those used in the di-Higgs gluon-fusion analysis

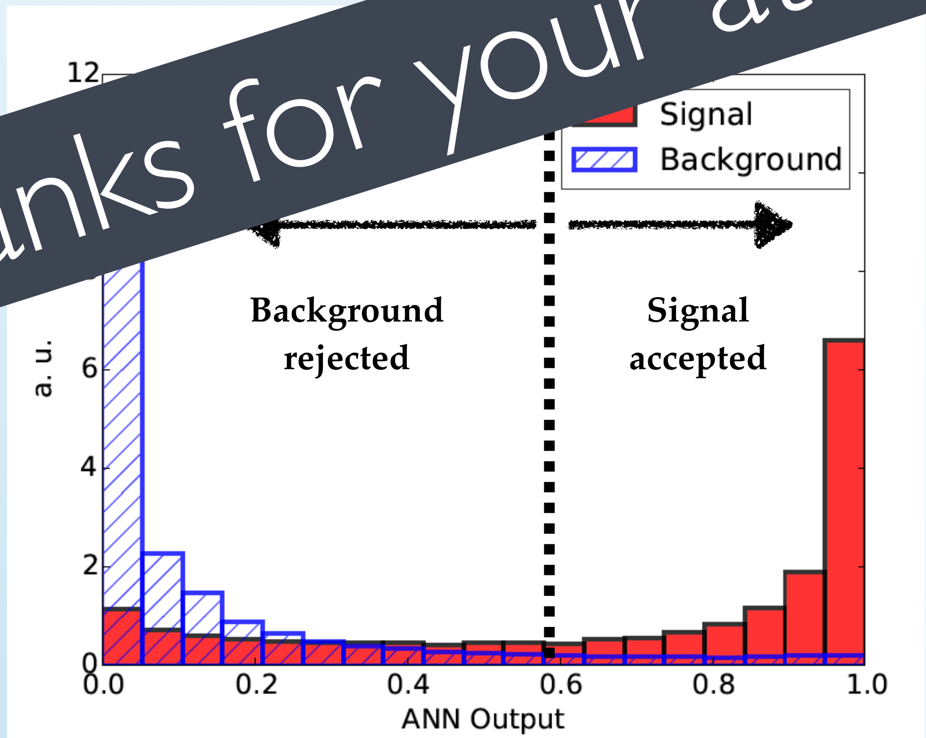


J. K. Behr, D. Bortoletto, J. A. Frost, N. P. Hartland, C. Issever and J. Rojo, arXiv:1512.08928

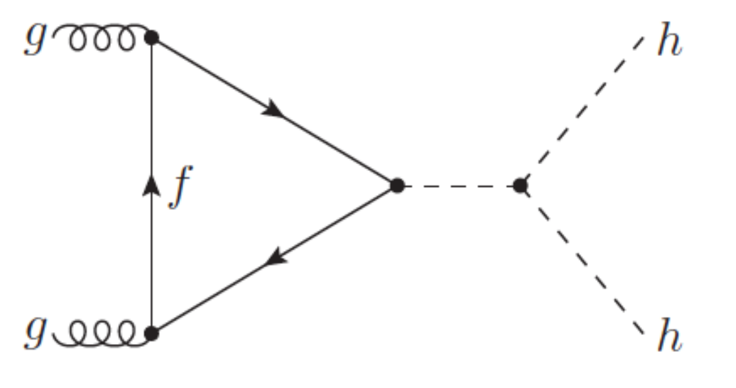
The di-Higgs frontier at the LHC

- Higgs pair production is a **cornerstone of the LHC program** for the coming years, allowing to reconstruct the **EWSB potential** and to test the nature of the **EWSB mechanism**
- The **4b final state** offers the highest yields, but requires taming the overwhelming QCD background
- In the VBF channel, the **steep rise of the cross-section with the di-Higgs invariant mass in the case of deviations from the SM couplings** is the key for the high sensitivity of this process
- Strong motivation to start to **search for VBF di-Higgs now**, no need to wait u… LHC!
- The ultimate optimization of the extraction of the Higgs co… the use of **multivariate techniques**, as we those used in the di…

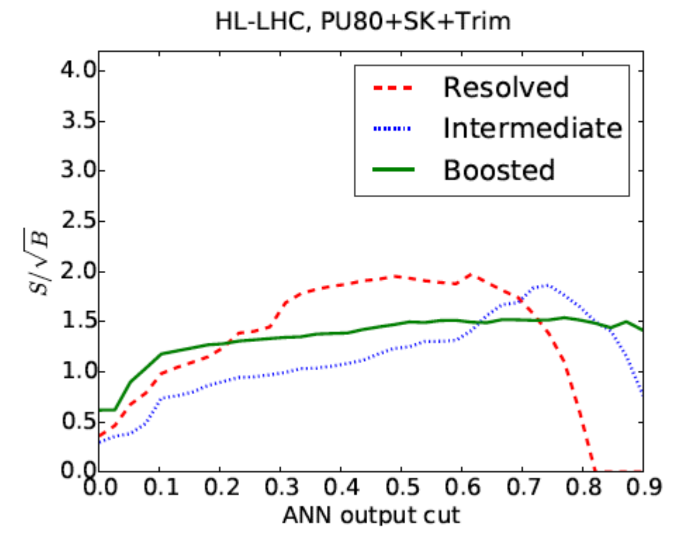
Thanks for your attention!



J. K. Behr, D. Bortoletto, J. A. Frost, N. P. Hartland, C. Issever and J. Rojo, arXiv:1512.08928



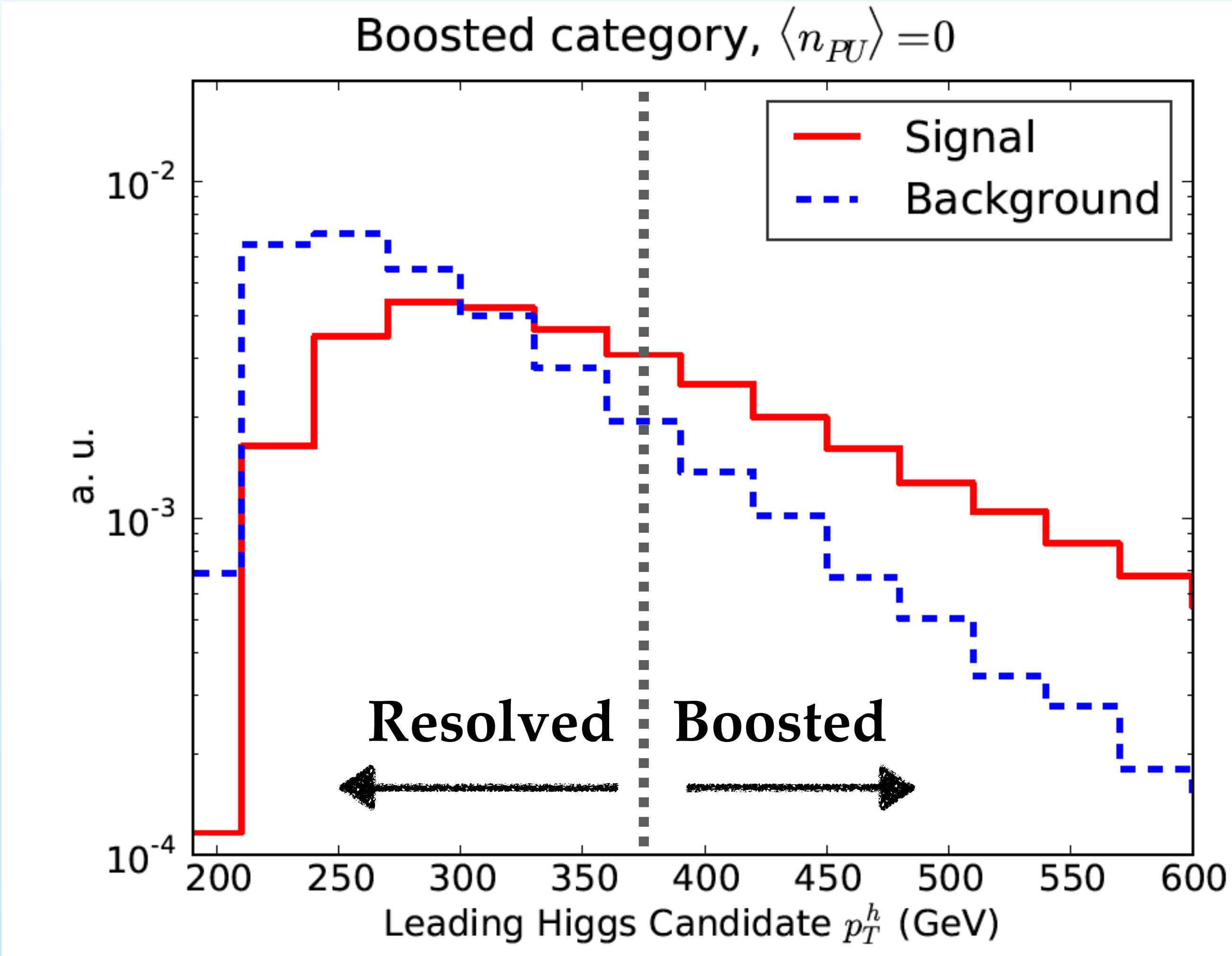
Higgs pair production by gluon-fusion in the 4b final state



- *J. K. Behr, D. Bortoletto, J. A. Frost, N. P. Hartland, C. Issever and J. Rojo, EPJC76, no. 7, 386 (2016), arXiv:1512.08928*
- *M. Gouzevitch, A. Oliveira, J. Rojo, R. Rosenfeld, G. P. Salam and V. Sanz, JHEP 1307, 148 (2013), arXiv:1303.6636*

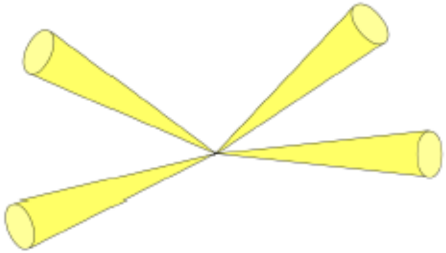
Higgs pair production at the LHC

- Focus on the **hh->4b final state:** largest rates, but **overwhelming QCD multijet background**
- Made competitive requiring the **di-Higgs system to be boosted** and exploiting kinematic differences between signal and QCD background with **jet substructure**
- Mandatory to optimize the **boosted b-tagging techniques,** impressive recent progress by ATLAS and CMS



Selection Strategy

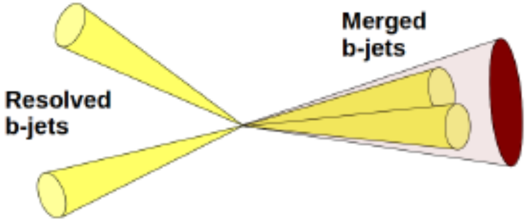
Resolved



- ≥ 4 b-tagged small-R jets
- Higgs reconstruction from leading 4 jets
- Choice that minimises mass difference between dijet systems

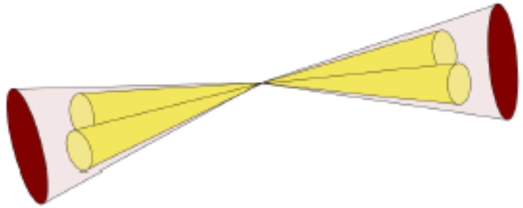
Scale-invariant resonance tagging
Gouzevitch, Oliveira, JR, Rosenfeld, Salam and Sanz, arxiv:1303.6636

Intermediate



- $= 1$ large-R jet (Higgs-tagged $+$ b-tagged) (leading Higgs)
- ≥ 2 b-tagged small-R jets
- $\Delta R > 1.2$ w.r.t. large-R jet
- Higgs reconstruction from leading 2 small-R jets
- Choice that minimises mass difference of dijet system and large-R jet

Boosted

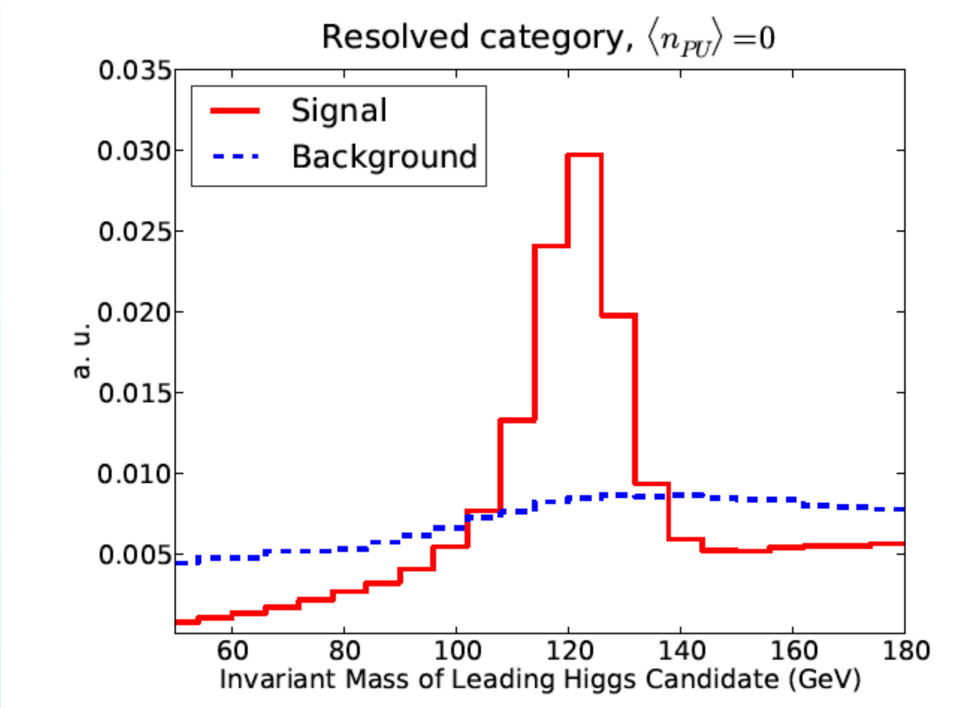


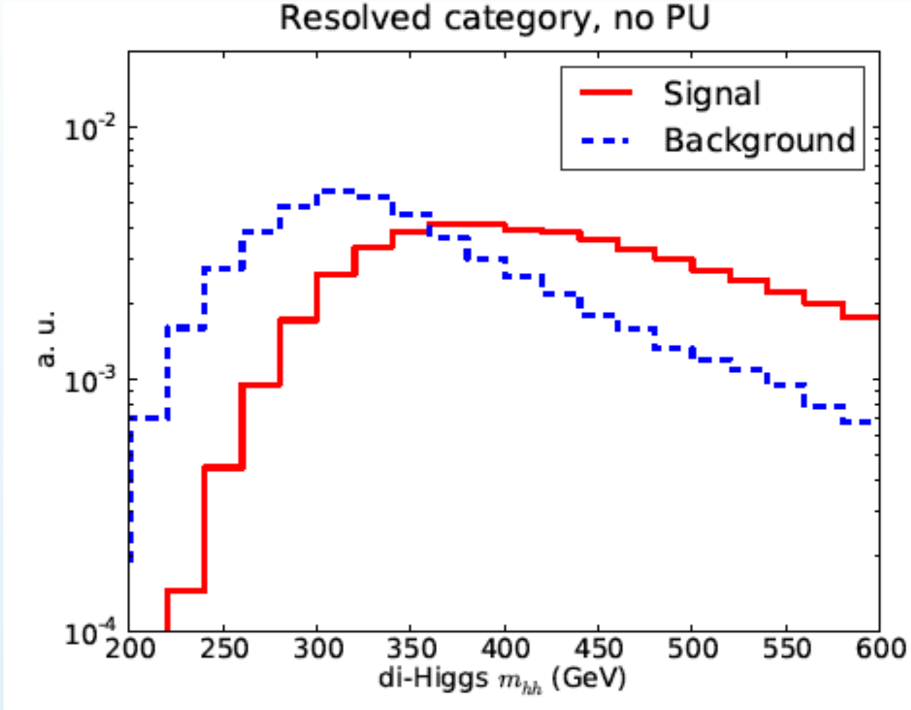
- ≥ 2 large-R jets (Higgs-tagged $+$ b-tagged)
- Leading two jets taken as Higgs candidates

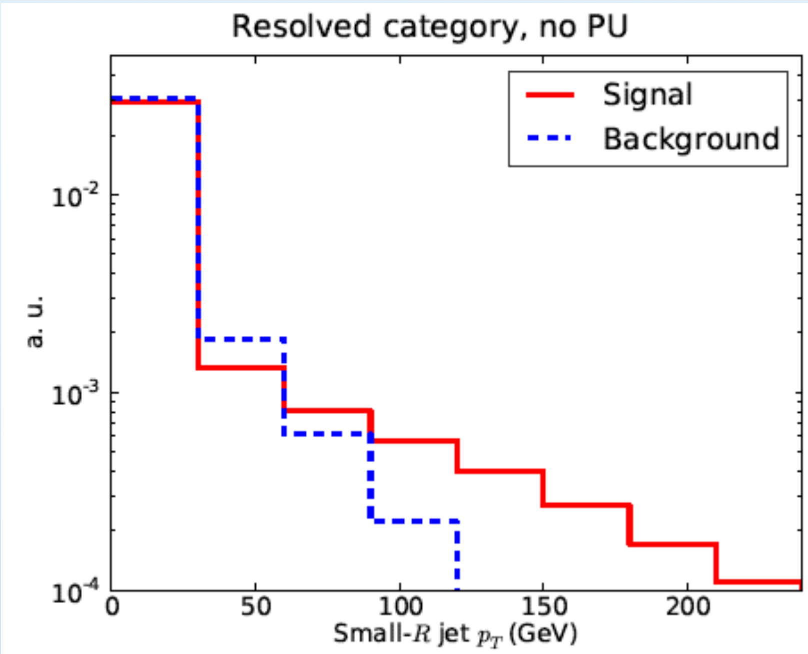
$+$ **Loose Higgs mass window cut**: $|m_{h,j} - 125 \text{ GeV}| < 40 \text{ GeV}$, $j = 1, 2$

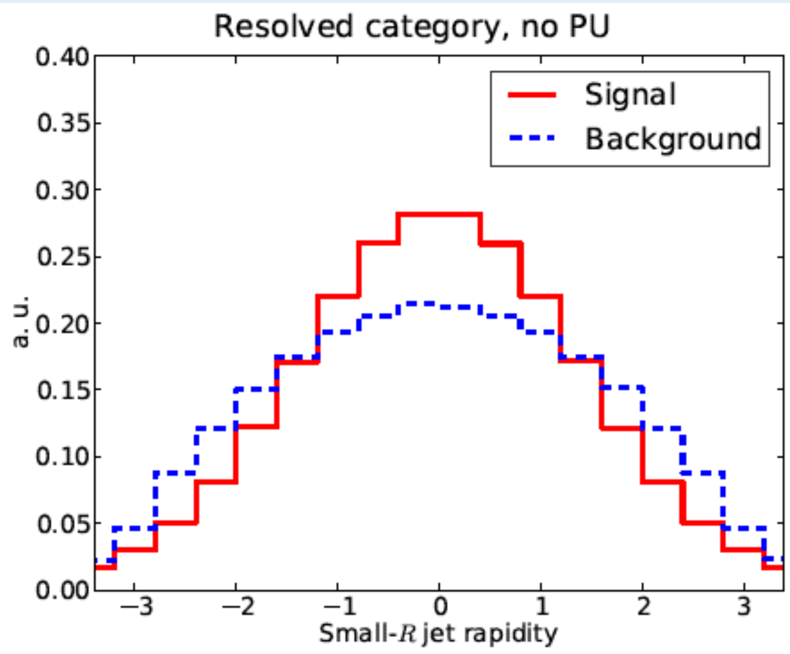
$+$ **Rank categories** by $S/\sqrt{B}$ to make them **exclusive**: boosted $>$ intermediate $>$ resolved

Cut-based analysis - resolved category

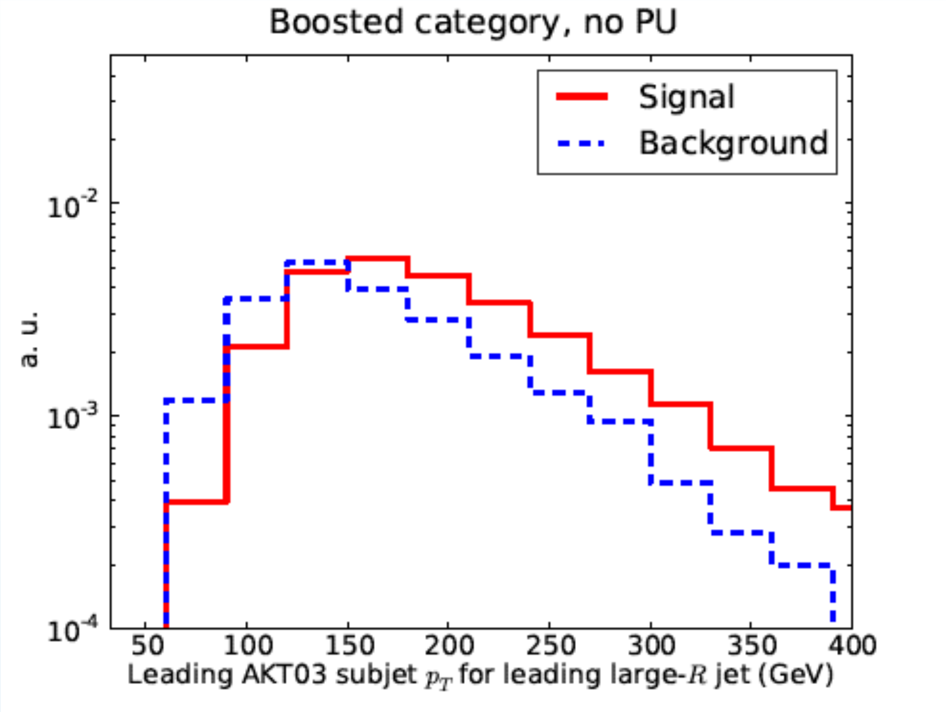


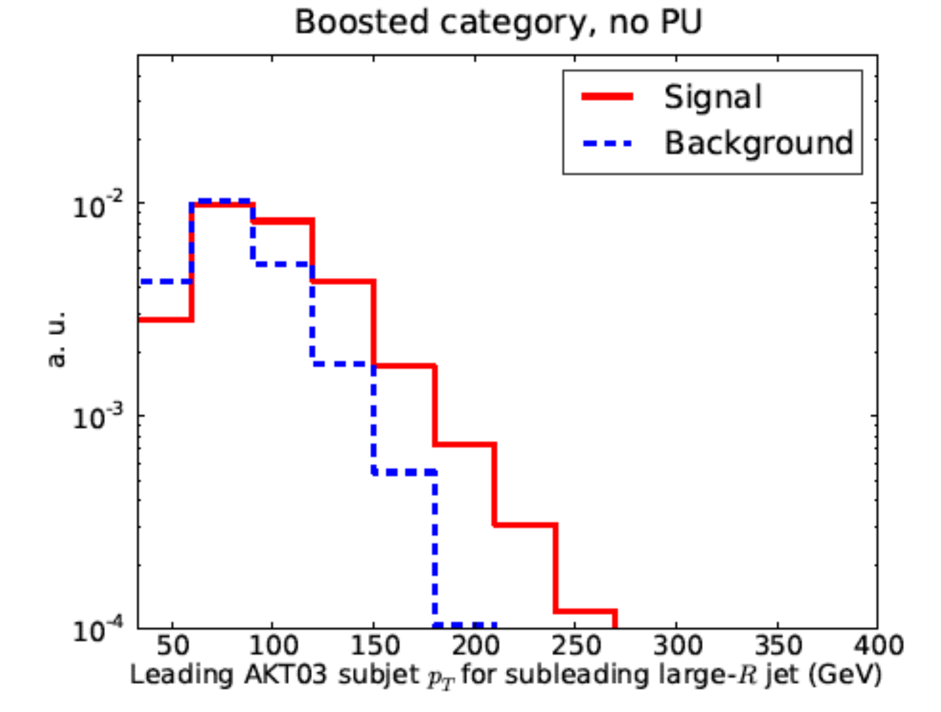


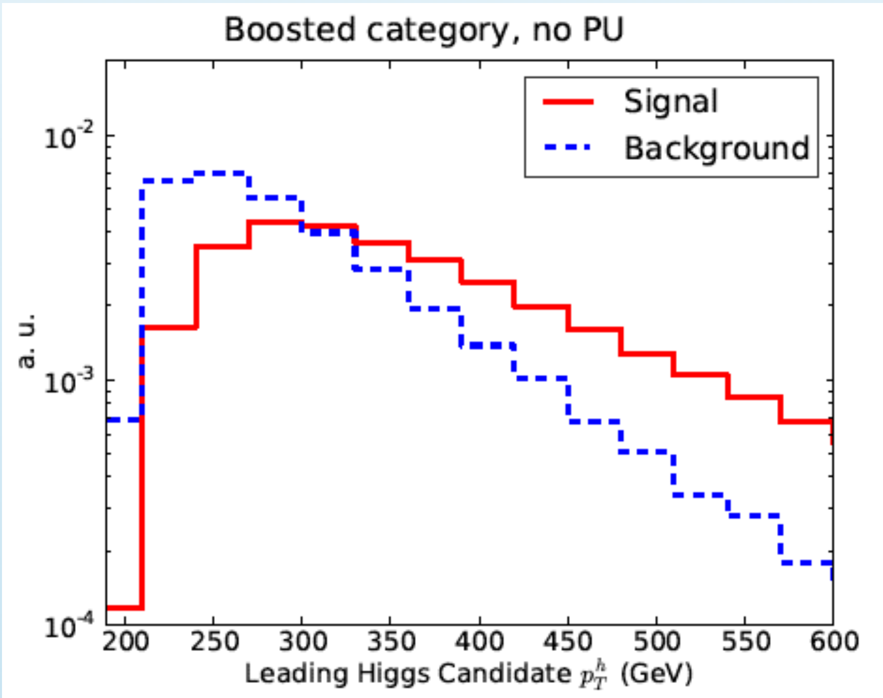


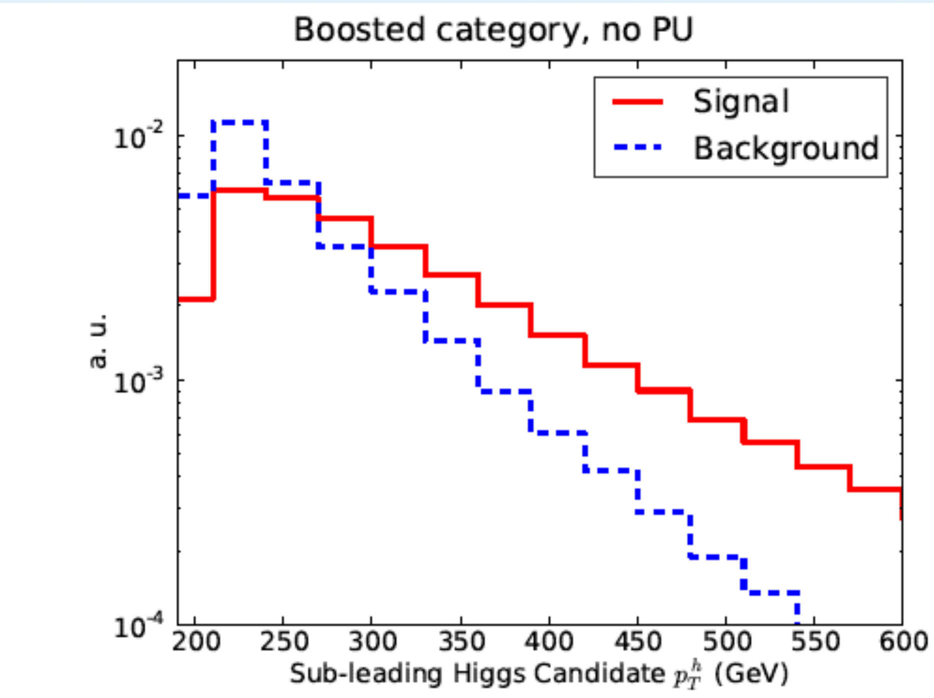


Cut-based analysis - boosted category









Jet substructure variables

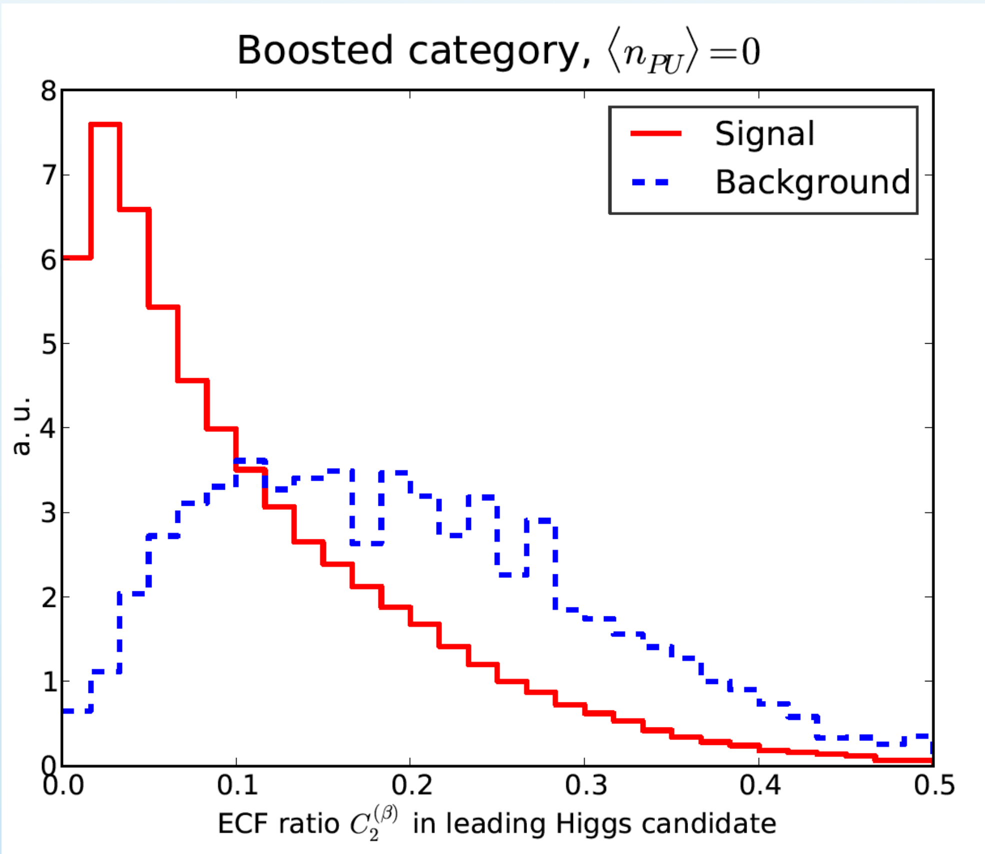
- **Substructure variables** quantify differences in internal structure between **QCD jets** and **jets from the decay of heavy resonances**
- QCD radiation tends to be **soft** and **collinear,** while decay products of resonances **share momentum evenly**

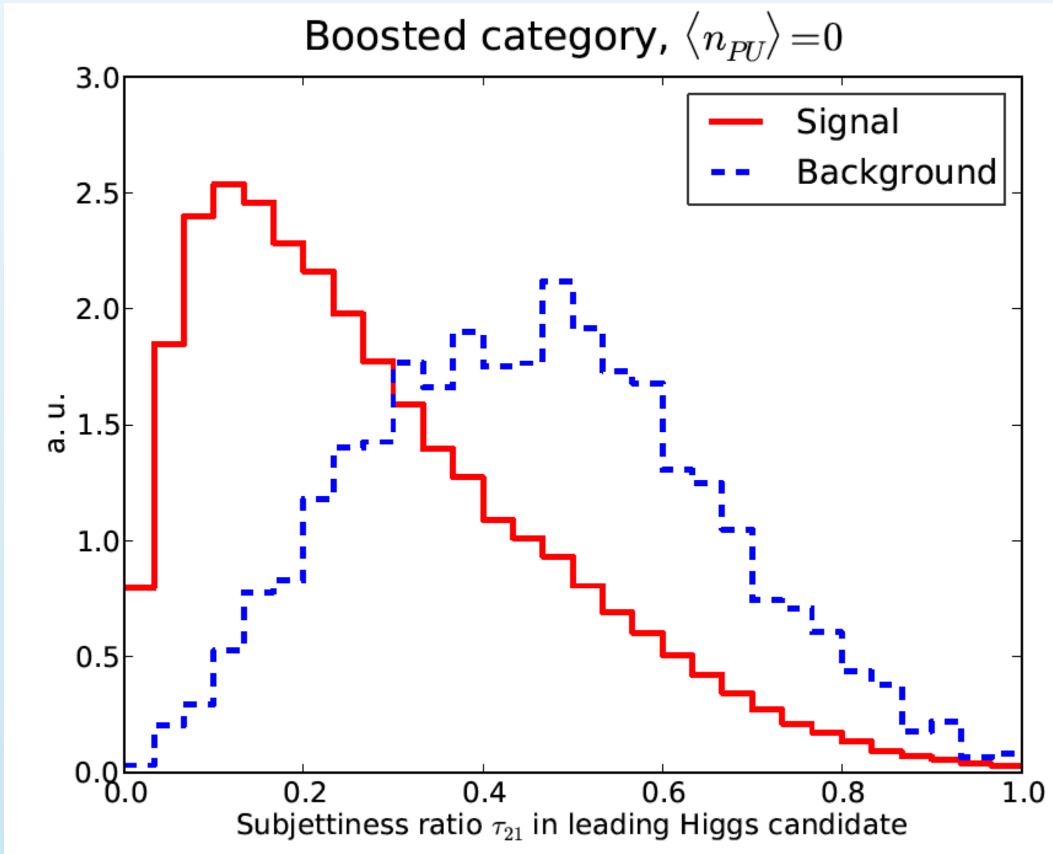
$$C_2^{(\beta)} \equiv \frac{\mathrm{ECF}(3,\beta)\mathrm{ECF}(1,\beta)}{[\mathrm{ECF}(2,\beta)]^2}$$

$$\tau_N \equiv \frac{1}{d_0}\sum_k p_{T,k} \cdot \min\left(\delta R_{1k}, \ldots, \delta R_{Nk}\right)$$

Energy Correlation Functions ratio

τ_{21}: 2-to-1 Subjettiness ratio





Cut-based analysis results

- Assume **HL-LHC integrated luminosity of L=3000 fb^{-1}**
- At the end of the cut-based analysis (level **C2**) **backgrounds still still overwhelmingly larger than the signal**, and signal significances rather low
- Use **MVA to fully exploit the kinematic discrimination power** between signal and background events

HL-LHC, Resolved category, no PU										
	Cross-section [fb]						S/B		$S/\sqrt{B}$	
	$hh4b$	total bkg	$4b$	$2b2j$	$4j$	$t\bar{t}$	tot	$4b$	tot	$4b$
C1a	9	$2.2 \cdot 10^8$	$6.9 \cdot 10^4$	$1.5 \cdot 10^7$	$2.0 \cdot 10^8$	$2.1 \cdot 10^5$	$4.0 \cdot 10^{-8}$	$1.3 \cdot 10^{-4}$	0.03	1.9
C1b	9	$2.2 \cdot 10^8$	$6.9 \cdot 10^4$	$1.5 \cdot 10^7$	$2.0 \cdot 10^8$	$2.1 \cdot 10^5$	$4.0 \cdot 10^{-8}$	$1.3 \cdot 10^{-4}$	0.03	1.9
C1c	2.6	$4.4 \cdot 10^7$	$1.6 \cdot 10^4$	$3.2 \cdot 10^6$	$4.1 \cdot 10^7$	$8.8 \cdot 10^4$	$6.1 \cdot 10^{-8}$	$1.6 \cdot 10^{-4}$	0.02	1.1
C2	0.5	$4.9 \cdot 10^3$	$1.7 \cdot 10^3$	$2.9 \cdot 10^3$	$2.1 \cdot 10^2$	47	$1.1 \cdot 10^{-4}$	$2.9 \cdot 10^{-4}$	0.4	0.6

HL-LHC, Boosted category, no PU										
	Cross-section [fb]						S/B		$S/\sqrt{B}$	
	$hh4b$	total bkg	$4b$	$2b2j$	$4j$	$t\bar{t}$	tot	$4b$	tot	$4b$
C1a	3.9	$4.6 \cdot 10^7$	$1.1 \cdot 10^4$	$2.9 \cdot 10^6$	$4.3 \cdot 10^7$	$2.4 \cdot 10^4$	$8.2 \cdot 10^{-8}$	$3.4 \cdot 10^{-4}$	0.03	2.0
C1b	2.7	$3.7 \cdot 10^7$	$7.5 \cdot 10^3$	$2.1 \cdot 10^6$	$3.5 \cdot 10^7$	$2.2 \cdot 10^4$	$7.4 \cdot 10^{-8}$	$3.7 \cdot 10^{-4}$	0.03	1.7
C1c	1.0	$3.9 \cdot 10^6$	$8.0 \cdot 10^2$	$2.3 \cdot 10^5$	$3.7 \cdot 10^6$	$7.1 \cdot 10^3$	$2.6 \cdot 10^{-7}$	$1.3 \cdot 10^{-3}$	0.03	2.0
C2	0.16	$2.5 \cdot 10^2$	53	$1.9 \cdot 10^2$	13	1.6	$5.7 \cdot 10^{-4}$	$2.7 \cdot 10^{-3}$	0.5	1.1

- Accurate treatment of **b-jet fakes from light and charm jets** crucial for this final state

Multivariate techniques

Large number of kinematic variables to disentangle signal and background, **how to combine them?**

Multivariate techniques: Identify automatically kinematical variables with most discrimination power

Higgs pT

Higgs m

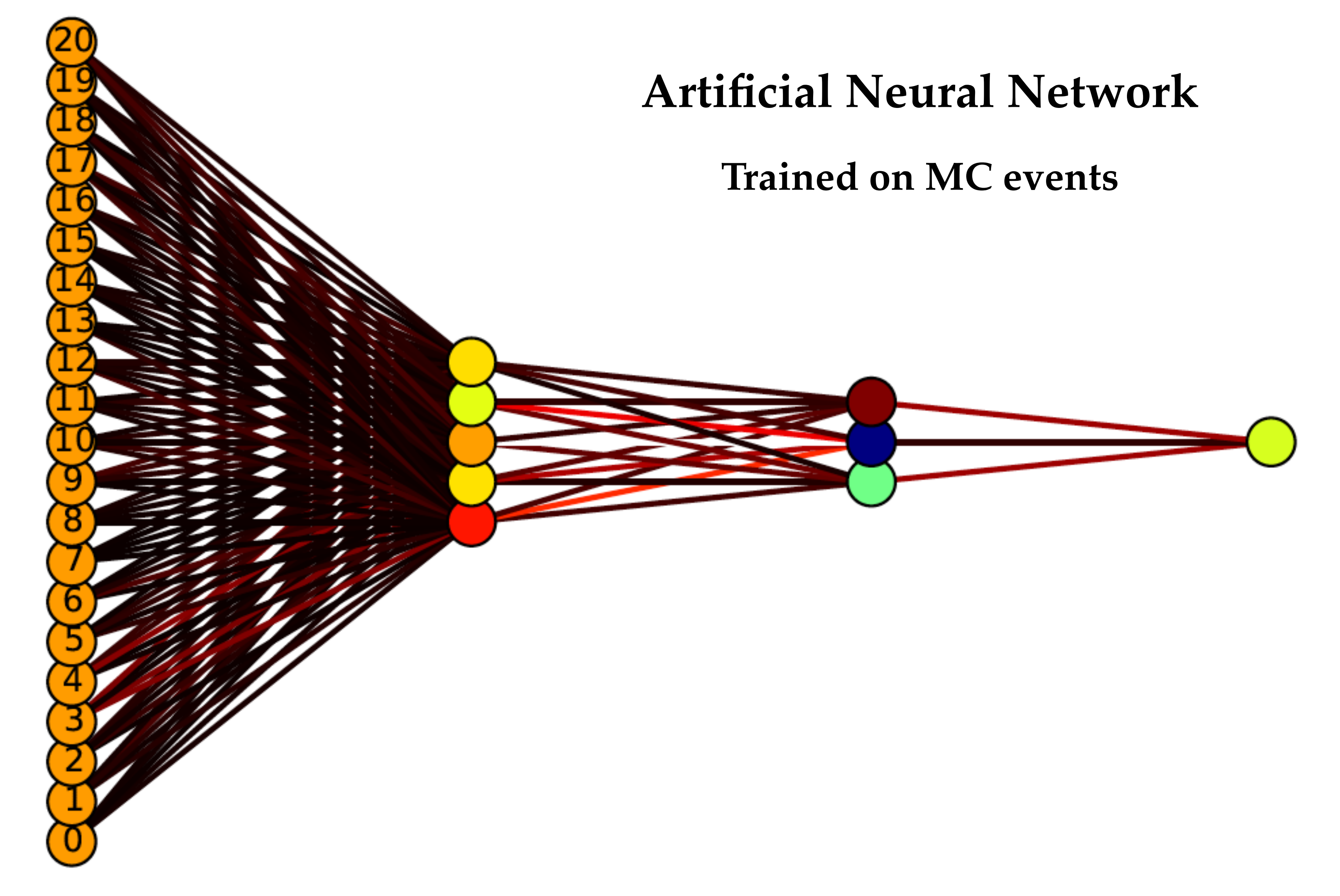
di-Higgs m

ECF

τ_{12}

Subjet pT

.....



Multivariate techniques

Given a set of N_{var} kinematic variables $\{k_i\}$ associated to MC event i, and a set of ANN weight parameters $\{\omega\}$, the ANN output y_i interpreted as **probability that this event originates from signal process**

$$y_i = P(y'_i = 1|\{k\}_i, \{\omega\}) \,,$$

With y'_i the true MC classification: $y'_i=1$ for signal, $y'_i=0$ for background

The **general classification probability** including background events is

$$P(y'_i|\{k\}_i, \{\omega\}) = y_i^{y'_i}(1-y_i)^{1-y'_i}$$

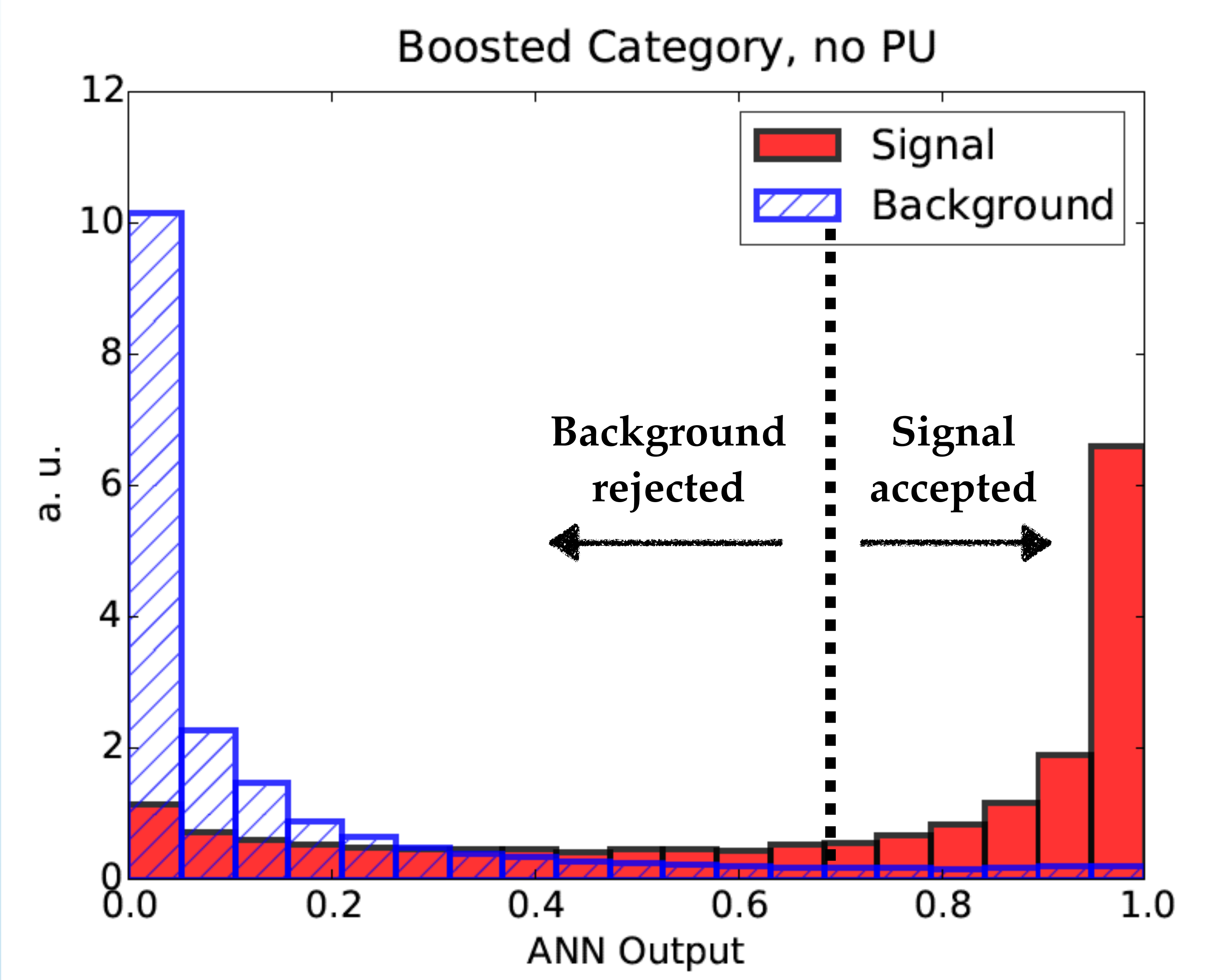
Thus the **error function to be minimised during the training** is the **cross-entropy:**

$$E(\{\omega\}) \equiv -\log\left(\prod_i^{N_{\rm ev}} P(y'_i|\{k\}_i, \{\omega\})\right)$$
$$= \sum_i^{N_{\rm ev}} \left[y'_i \log y_i + (1-y'_i)\log(1-y_i)\right]$$

ANN training performed with **Genetic Algorithms** using **cross-validation stopping**

Multivariate techniques

Combining information from all kinematic variables in MVA: **excellent signal/background discrimination**

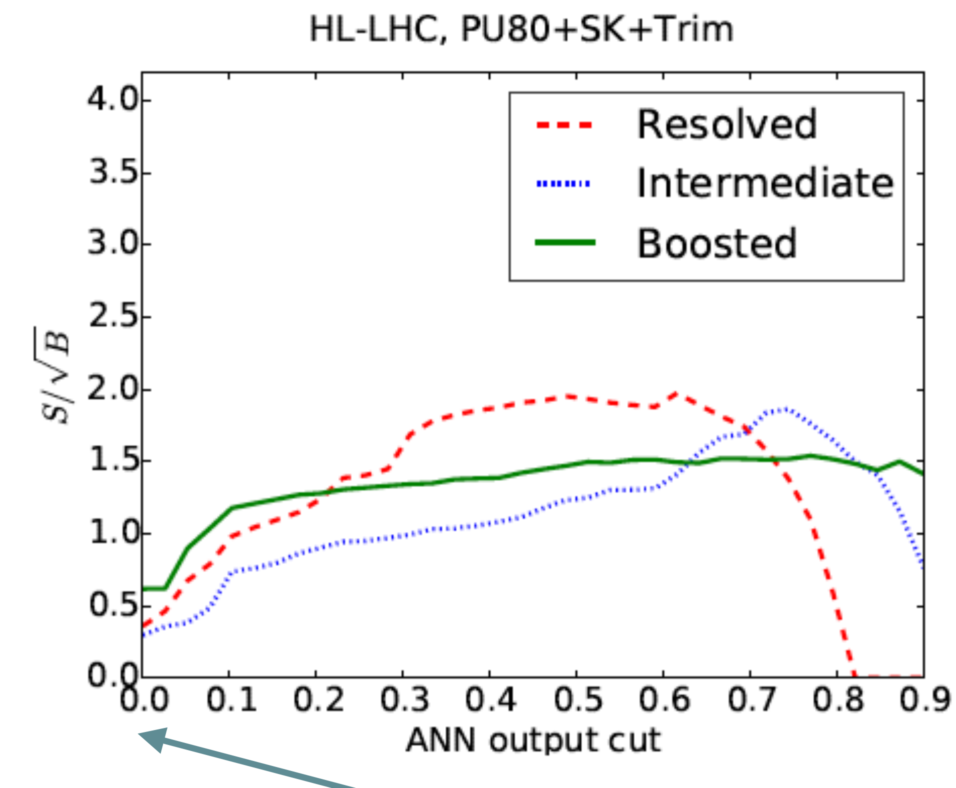


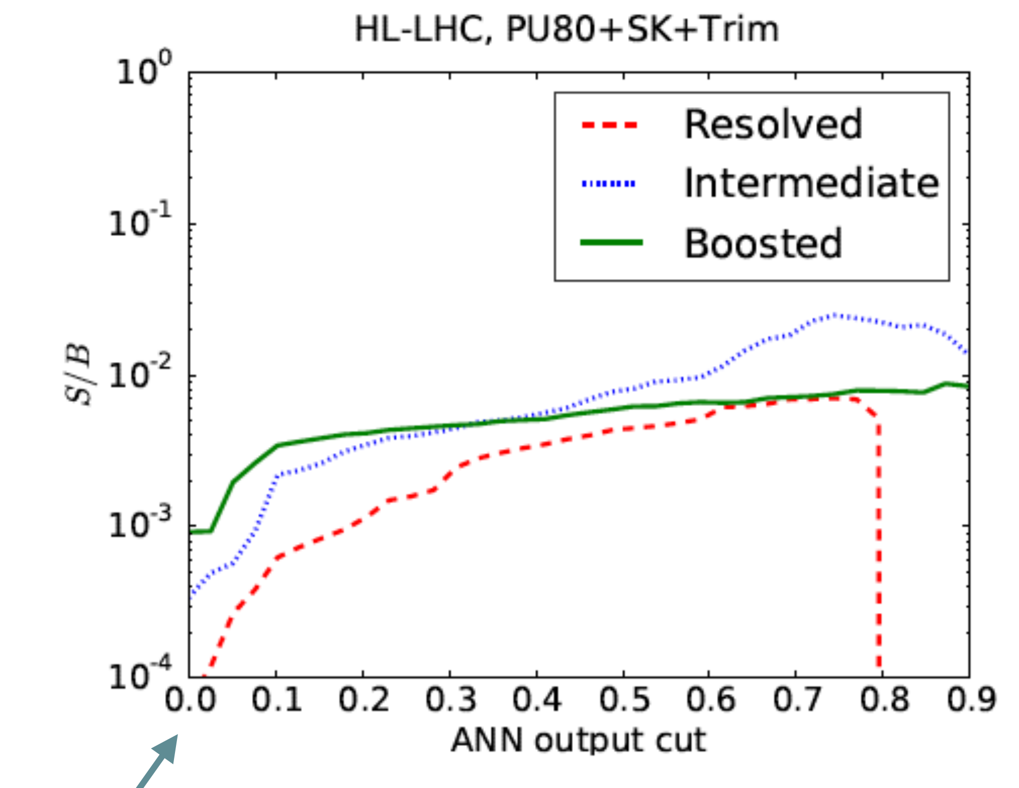
Signal significance

- Use of **multivariate techniques** allows to **substantially improve the signal significance** for this process as compared to a **traditional cut -based analysis**
- The total combined significance is enough to **observe Higgs pair production in the 4b final state** at the HL-LHC. Substantial improvement if reducible backgrounds (fakes) can be eliminated

$$\left(\frac{S}{\sqrt{B}}\right)_{\rm tot} \simeq 3.1\ (1.0)\,, \quad \mathcal{L} = 3000\ (300)\ {\rm fb}^{-1}$$

$$\left(\frac{S}{\sqrt{B_{4b}}}\right)_{\rm tot} \simeq 4.7\ (1.5)\,, \quad \mathcal{L} = 3000\ (300)\ {\rm fb}^{-1}$$





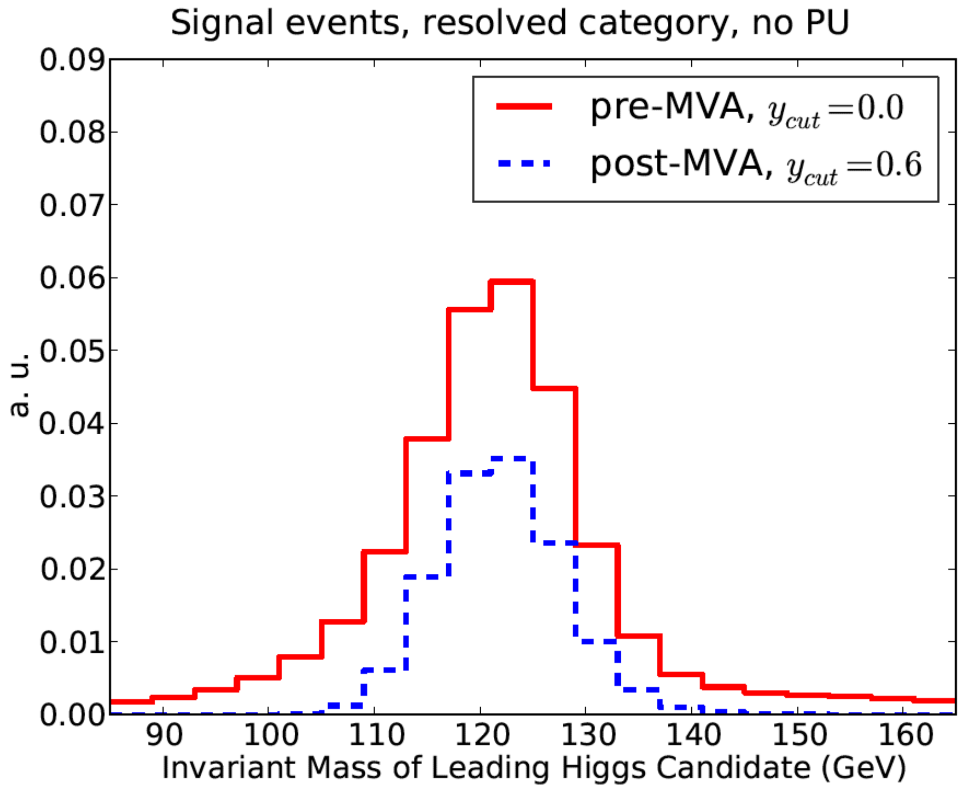
Pre-MVA result

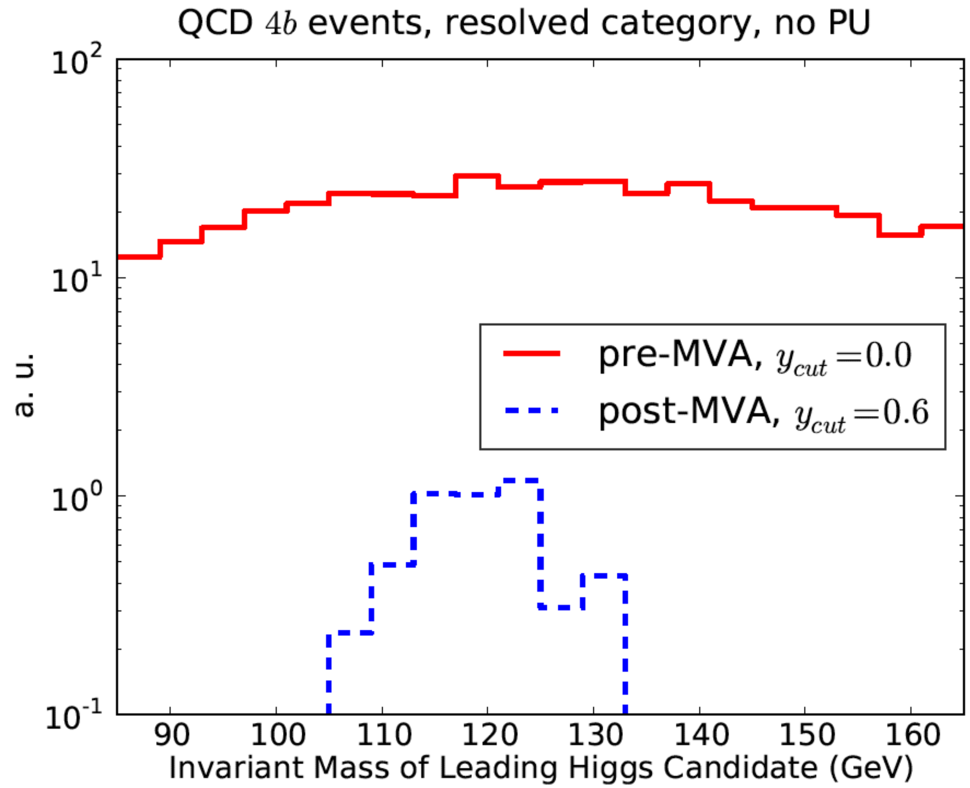
Opening the Black Box

- ANNs are sometimes criticised by acting as **black boxes**, with little control/understanding of what is happening inside them
- But ANNs are simply a **set of combined kinematical cuts**, nothing mysterious in them
- To verify this, plot kin distributions **after and before the ANN cut**: we can then determine the **effective kinematic cuts** are being optimised by the MVA
- This info should be enough to **perform a cut-based analysis** and achieve similar signal significance

Opening the Black Box

- ANNs are sometimes criticised by acting as **black boxes**, with little control/understanding of what is happening inside them
- But ANNs are simply a **set of combined kinematical cuts**, nothing mysterious in them
- To verify this, plot kin distributions **after and before the ANN cut**: we can then determine the **effective kinematic cuts** are being optimised by the MVA
- This info should be enough to **perform a cut-based analysis** and achieve similar signal significance





The background "Higgs mass peak" now mimics the signal one!

A measurement of the self-coupling

- Higgs pair production is a **cornerstone of the LHC program** for the next years, providing evidence of **Higgs self-interactions and** allowing to reconstruct the **EWSB potential**

- The **4b final state** offers the highest yields, but requires taming an overwhelming QCD background. At the **HL-LHC**, **"observation" signal significances** will be achieved in this final state. A first measurement might even be possible **by the end of Run II**, but extremely challenging (limited by systematics).

- Now working on estimating the accuracy on the **extraction of the Higgs self-coupling** that can be achieved at the **LHC Run II**, the **HL-LHC** and at a **100 TeV FCC**

- **The optimization of the extraction of the Higgs self-coupling might require the use of an additional MVA** (different learning task than signal/background classification)