Discovering Double Higgs Production with Machine Learning Tools

Alexandre Alves

UNIFESP-Diadema

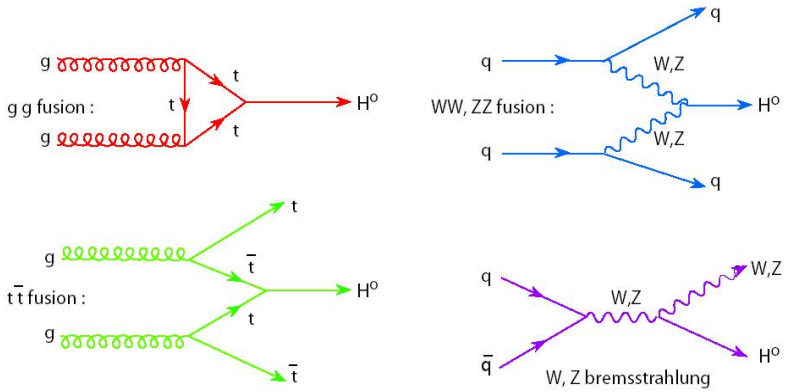
June 22th, 2017

Outline

- Why double Higgs production is important?
- Production and decays of Higgs bosons pairs
- Optimizing cut-and-count
- ML tools to discover $pp \rightarrow HH$
- Further research

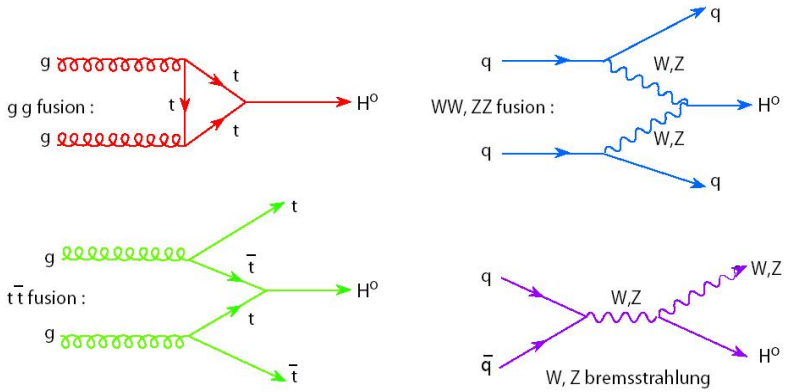
Scalar potential

After Higgs boson discovery, it remains, for example:



Scalar potential

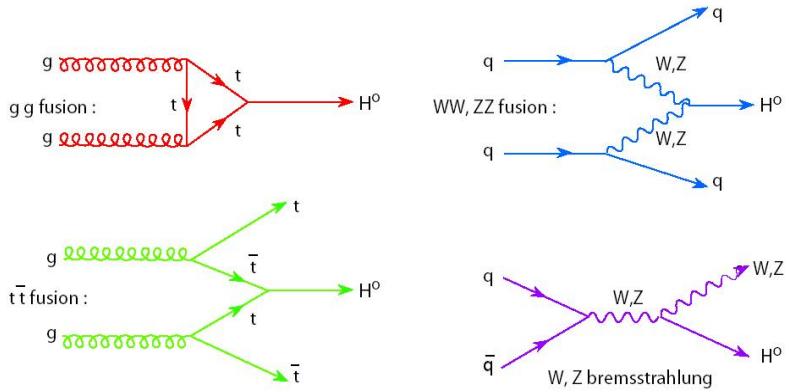
After Higgs boson discovery, it remains, for example:



- Measuring its properties like SPIN and COUPLINGS

Scalar potential

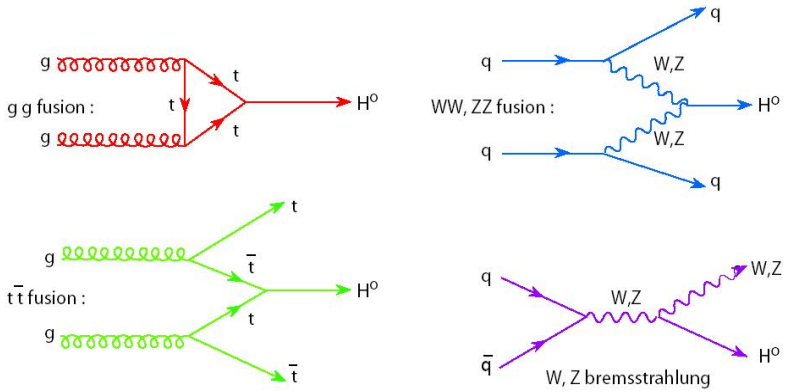
After Higgs boson discovery, it remains, for example:



- Measuring its properties like SPIN and COUPLINGS
- Determining its Total Width and observe all the decay channels

Scalar potential

After Higgs boson discovery, it remains, for example:

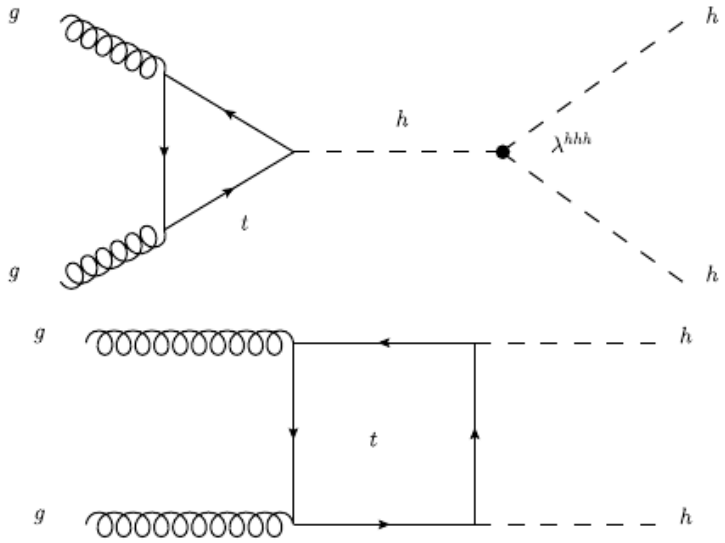


- Measuring its properties like SPIN and COUPLINGS
- Determining its Total Width and observe all the decay channels
- Studying its Self-Interactions ➡ Stability, Inflation, BSM

$$
\begin{aligned}
V(|H|^2) &= \mu^2|H|^2 + \frac{1}{2}\lambda|H|^4\ ,\ H \to h(x) + v \\
&= m_h^2 h^2 + \lambda_{hhh} hhh + \lambda_{hhhh} hhhh
\end{aligned}
$$

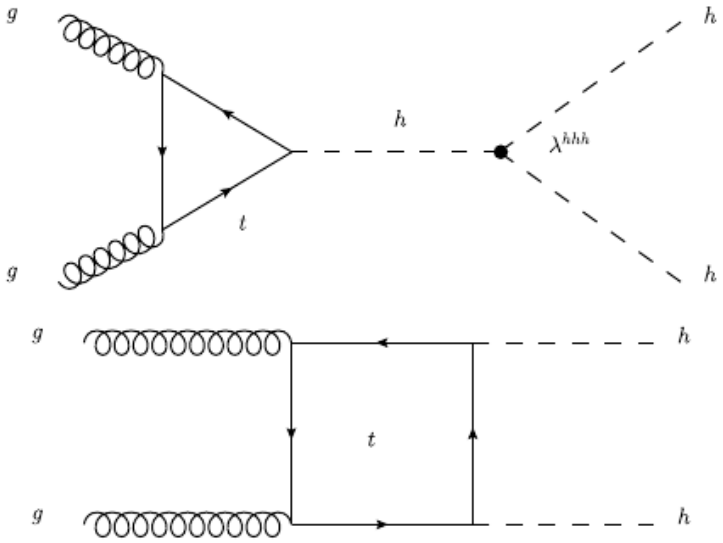
Measuring λ_{hhh}

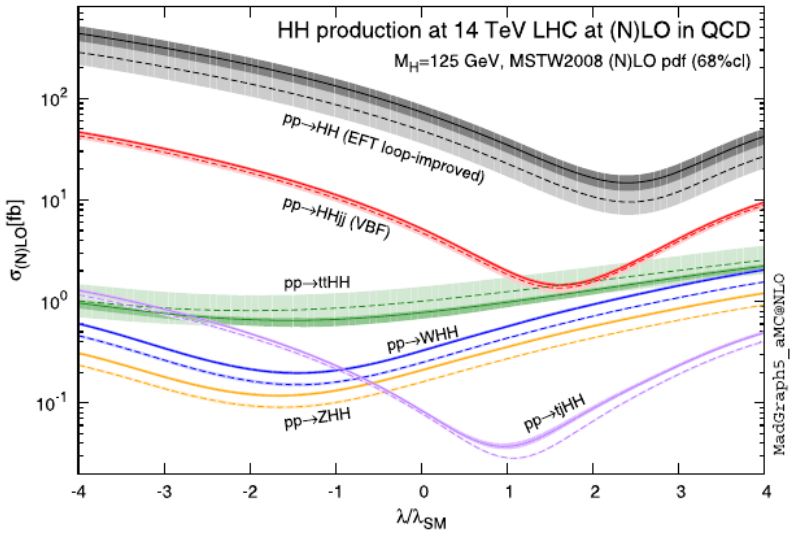
SM Double Higgs Production @ LHC



Measuring λ_{hhh}

SM Double Higgs Production @ LHC

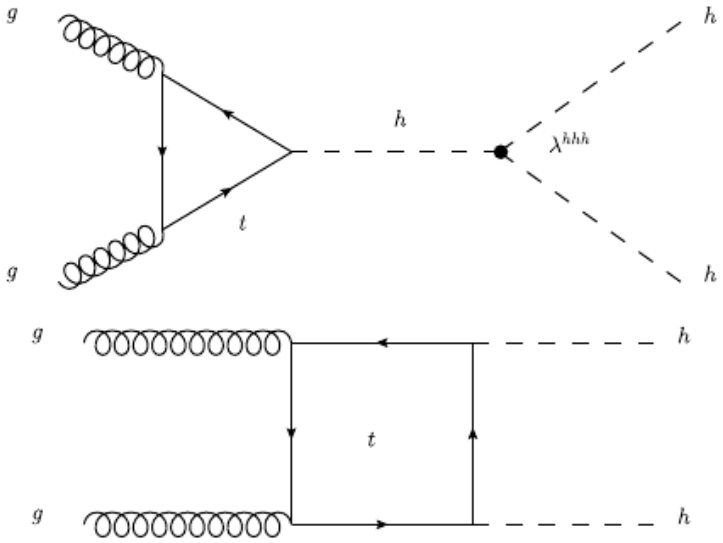


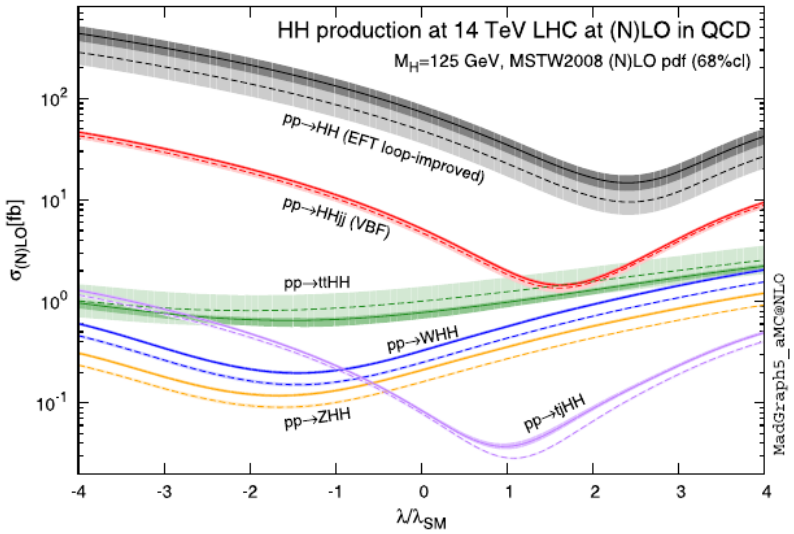


- Destructive interference between triangle and box in the SM

Measuring λ_{hhh}

SM Double Higgs Production @ LHC

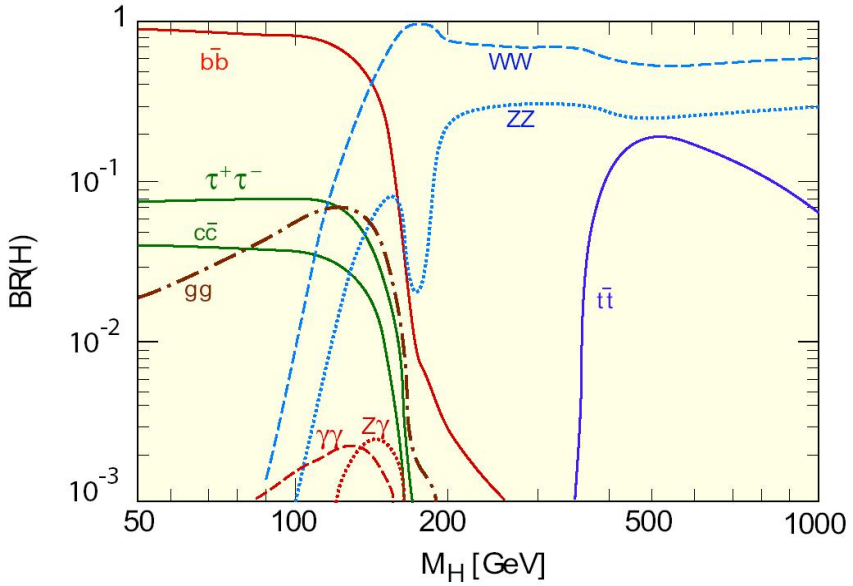




- Destructive interference between triangle and box in the SM
- current status: $|\lambda - \lambda_{SM}|/\lambda_{SM} \sim 30\%$ @95% CL

Measuring λ_{hhh}

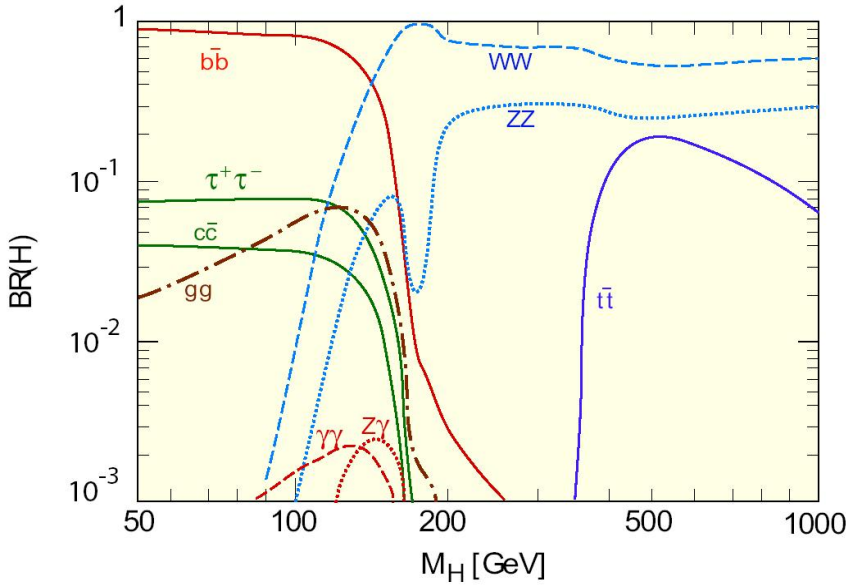
Decay channels



- $hh \rightarrow b\bar{b}b\bar{b}$: largest BR, $\sim 32\%$, large QCD backgrounds

Measuring λ_{hhh}

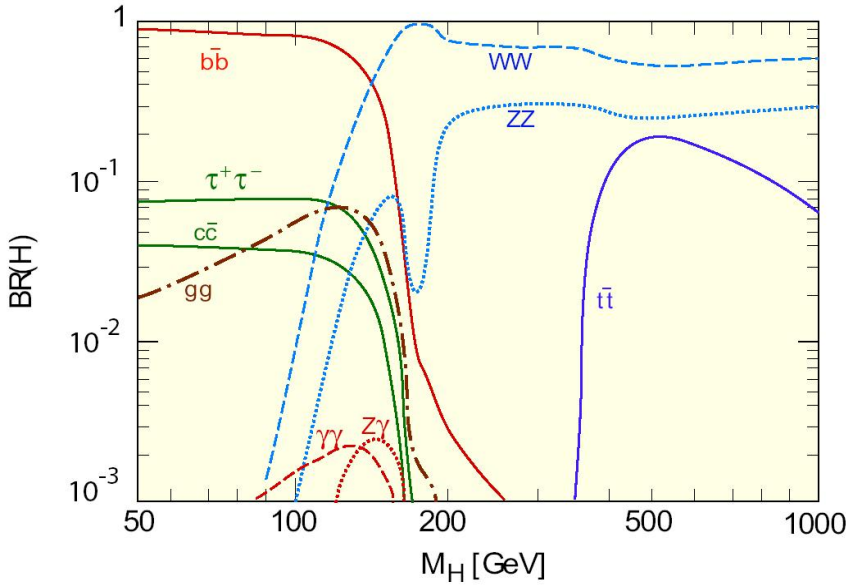
Decay channels



- $hh \to b\bar{b}b\bar{b}$: largest BR, $\sim 32\%$, large QCD backgrounds
- $hh \to b\bar{b}W^+W^-$: decent BR, but $t\bar{t}$ is background

Measuring λ_{hhh}

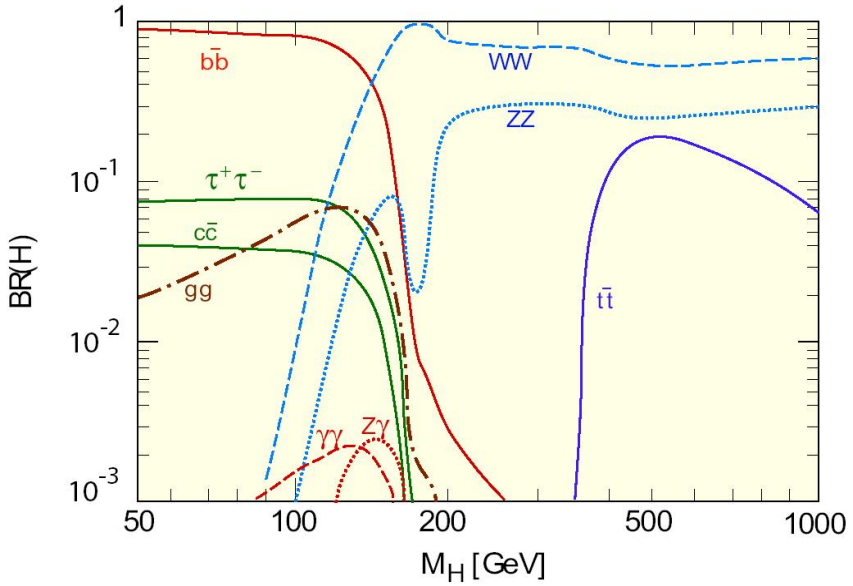
Decay channels



- $hh \to b\bar{b}b\bar{b}$: largest BR, $\sim 32\%$, large QCD backgrounds
- $hh \to b\bar{b}W^+W^-$: decent BR, but $t\bar{t}$ is background
- $hh \to b\bar{b}\tau^+\tau^-$: promising with efficient τ, b-tagging

Measuring λ_{hhh}

Decay channels



- $hh \to b\bar{b}b\bar{b}$: largest BR, $\sim 32\%$, large QCD backgrounds
- $hh \to b\bar{b}W^+W^-$: decent BR, but $t\bar{t}$ is background
- $hh \to b\bar{b}\tau^+\tau^-$: promising with efficient τ, b-tagging
- $hh \to b\bar{b}\gamma\gamma$: BR $\sim 0.26\%$, but it's very clean!
- $b\bar{b}\gamma\gamma$ has the best prospects, around 2σ @ 14TeV HL-LHC

Measuring λ_{hhh}

Our simulations for the $b\bar{b}\gamma\gamma$ channel (1704.07395)

- Signal+Backgrounds with `Mad5aMC@NLO+Pythia6+Delphes3`
- NLO and NNLO K-factors, but without top mass effects
- All dominant and subdominant backgrounds taken into account

Measuring λ_{hhh}

Our simulations for the $b\bar{b}\gamma\gamma$ channel (1704.07395)

- Signal+Backgrounds with `Mad5aMC@NLO+Pythia6+Delphes3`
- NLO and NNLO K-factors, but without top mass effects
- All dominant and subdominant backgrounds taken into account
- $b\bar{b}\gamma\gamma$, $c\bar{c}\gamma\gamma$, $jj\gamma\gamma$, $bj\gamma\gamma$

Measuring λ_{hhh}

Our simulations for the $b\bar{b}\gamma\gamma$ channel (1704.07395)

- Signal+Backgrounds with `Mad5aMC@NLO+Pythia6+Delphes3`
- NLO and NNLO K-factors, but without top mass effects
- All dominant and subdominant backgrounds taken into account
- $b\bar{b}\gamma\gamma$, $c\bar{c}\gamma\gamma$, $jj\gamma\gamma$, $bj\gamma\gamma$ $\Longrightarrow$ c,j mistagged as b
- $t\bar{t}h$, $c\bar{c}\gamma j$, $b\bar{b}h$, Zh, $b\bar{b}jj$

Measuring λ_{hhh}

Our simulations for the $b\bar{b}\gamma\gamma$ channel (1704.07395)

- Signal+Backgrounds with `Mad5aMC@NLO+Pythia6+Delphes3`
- NLO and NNLO K-factors, but without top mass effects
- All dominant and subdominant backgrounds taken into account
- $b\bar{b}\gamma\gamma$, $c\bar{c}\gamma\gamma$, $jj\gamma\gamma$, $bj\gamma\gamma$ $\Longrightarrow$ c,j mistagged as b
- $t\bar{t}h$, $c\bar{c}\gamma j$, $b\bar{b}h$, Zh, $b\bar{b}jj$ $\Longrightarrow$ small xsec, $P(j \to \gamma) \sim 10^{-4}$
- Simulations checked against 2 theoretical and 2 exp works

Measuring λ_{hhh}

Our simulations for the $b\bar{b}\gamma\gamma$ channel (1704.07395)

- Signal+Backgrounds with `Mad5aMC@NLO+Pythia6+Delphes3`
- NLO and NNLO K-factors, but without top mass effects
- All dominant and subdominant backgrounds taken into account
- $b\bar{b}\gamma\gamma$, $c\bar{c}\gamma\gamma$, $jj\gamma\gamma$, $bj\gamma\gamma$ $\Longrightarrow$ c,j mistagged as b
- $t\bar{t}h$, $c\bar{c}\gamma j$, $b\bar{b}h$, Zh, $b\bar{b}jj$ $\Longrightarrow$ small xsec, $P(j \to \gamma) \sim 10^{-4}$
- Simulations checked against 2 theoretical and 2 exp works
- After basic selections and 3 ab^{-1}

$$p_T(j) > 20 \text{ GeV},\ p_T(\gamma) > 20 \text{ GeV},\ |\eta(j,\gamma)| < 2.5$$
$$100 \text{ GeV} < M_{bb(\gamma\gamma)} < 150 \text{ GeV}$$

signal	$b\bar{b}\gamma\gamma$	$c\bar{c}\gamma\gamma$	$jj\gamma\gamma$	$b\bar{b}\gamma j$	$t\bar{t}h$	$c\bar{c}\gamma j$	$b\bar{b}h$	Zh	total backgrounds
42.6	1594.5	447.7	160.3	137	101.1	38.2	2.4	1.8	2483

Measuring λ_{hhh}

Updating previous results

- Our simulations with cut analysis of previous works

Reference	Kinematic cuts	AMS(σ) (S/B)
(A) [1]	$p_{T_{\gamma(b)}} > 20(45)$ GeV, $\lvert\eta_{b,\gamma}\rvert < 2.5$ $\lvert M_{bb} - m_h\rvert < 20$ GeV, $\lvert M_{\gamma\gamma} - m_h\rvert < 2.3$ GeV $\Delta R_{b\gamma} > 1.0$, $\Delta R(\gamma\gamma) < 2.0$	1.54(0.30)
(B) [2]	$p_{T_{b,\gamma}} > 50$ GeV, $\lvert\eta_{b,\gamma}\rvert < 2.5$, $\Delta R_{b\gamma} > 0.4$, $\Delta R(bb) < 2.5$ $110 < M_{bb} < 135$ GeV, $\lvert M_{\gamma\gamma} - m_h\rvert < 5$ GeV, $M_{bb\gamma\gamma} > 350$ GeV $\lvert\eta_H\rvert < 2$, $P_{T_H} > 100$ GeV	1.33(0.39)
(C) [3]	$p_{T_{b,\gamma}} > 30$ GeV, $\lvert\eta_{b,\gamma}\rvert < 2.5$ $\lvert M_{bb} - m_h\rvert < 12.5$ GeV, $\lvert M_{\gamma\gamma} - m_h\rvert < 5$ GeV $M_{bb\gamma\gamma} > 350$ GeV	1.51(0.17)
(D) [4]	$p_{T_{1(2)}} > 30(50)$ GeV, $\lvert\eta_{b,\gamma}\rvert < 2.4$ $\Delta R_{b\gamma} > 1.5$, $\Delta R(bb,\gamma\gamma) < 2$ $\lvert M_{bb} - m_h\rvert < 20$ GeV, $\lvert M_{\gamma\gamma} - m_h\rvert < 5$ GeV	1.76(0.27)
ATLAS [5]	$p_{T_\gamma} > 30(30)$ GeV, $p_{T_\gamma} > 40(25)$ GeV, $\lvert\eta_{b,\gamma}\rvert < 2.4$ $\Delta R_{b\gamma} > 0.4$, $\Delta R(bb,\gamma\gamma) < 2$, $p_{T_{bb,\gamma\gamma}} > 110$ GeV $\lvert M_{bb} - m_h\rvert < 25$ GeV, $123 < M_{\gamma\gamma} < 128$ GeV	1.73(0.28)

Best current result for this channel $\sim 1.8\sigma$

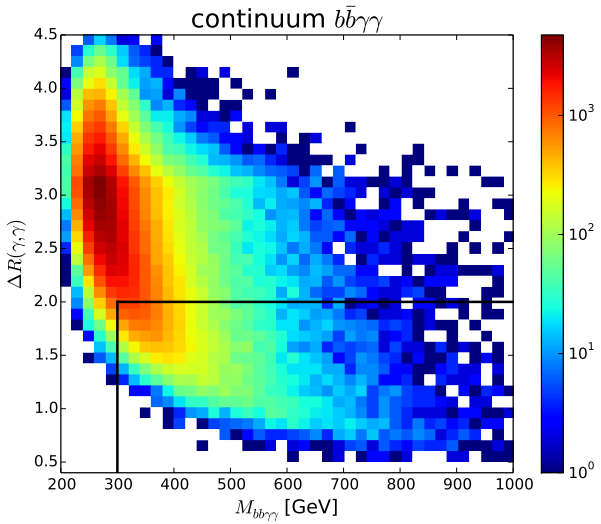
Cut-and-count, an art?

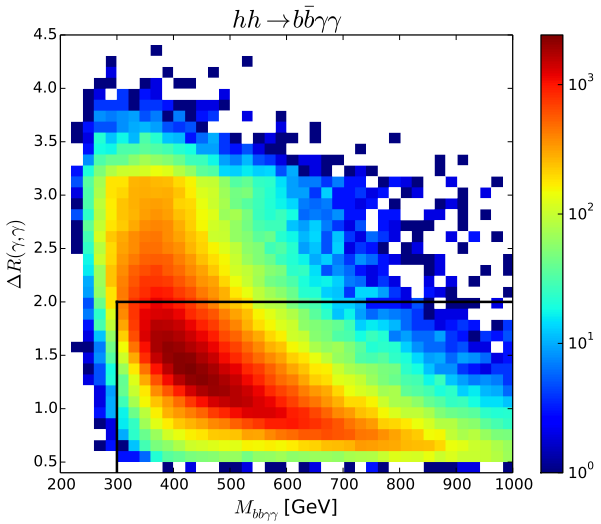
Deciding the cut selection to optimize signal significance

Cut-and-count, an art?

Deciding the cut selection to optimize signal significance

- Most widely employed technique in particle pheno
- How is it exactly done? **By eye!** Huge waste of data ($$)

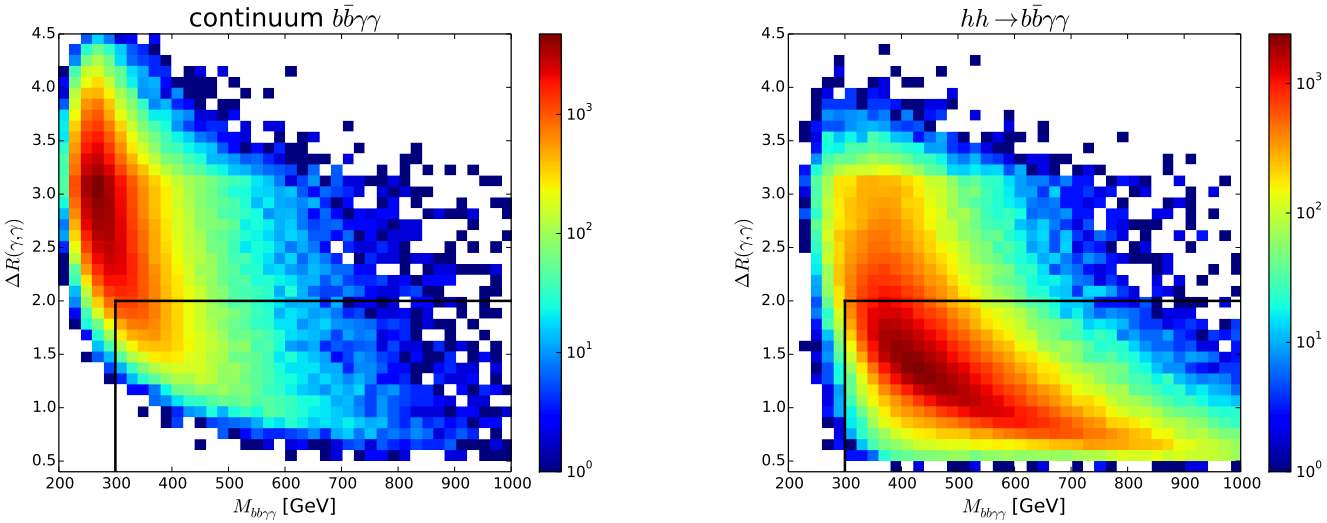


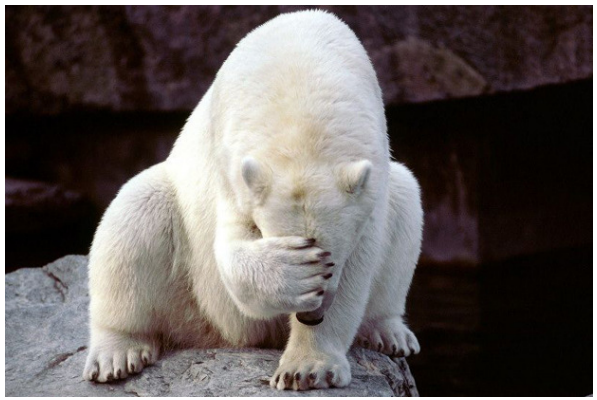


Cut-and-count, an art?

Deciding the cut selection to optimize signal significance

- Most widely employed technique in particle pheno
- How is it exactly done? **By eye!** Huge waste of data ($$)





Cut-and-count should be a science!

Tuning cut thresholds

Cut-and-count should be a science!

Tuning cut thresholds

- Simple idea: let's tune the cut thresholds!

Cut-and-count should be a science!

Tuning cut thresholds

- Simple idea: let's tune the cut thresholds!
- Problem :(prohibitively large grids, $\mathcal{O}(10^{14})$ points

Cut-and-count should be a science!

Tuning cut thresholds

- Simple idea: let's tune the cut thresholds!
- Problem :(prohibitively large grids, $\mathcal{O}(10^{14})$ points
- Solution ;) let's ask data scientists

Cut-and-count should be a science!

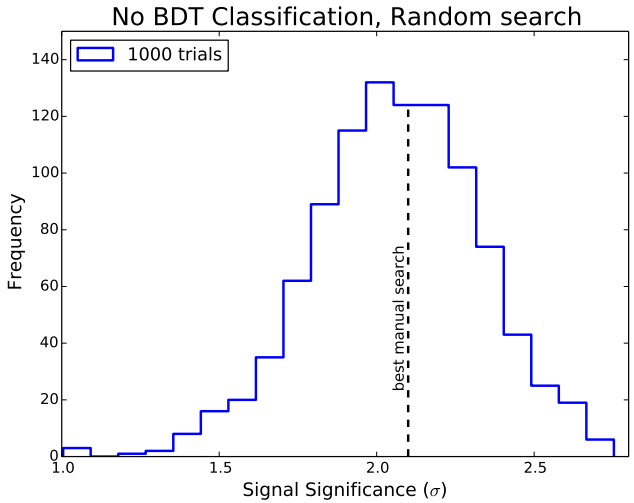
Tuning cut thresholds

- Simple idea: let's tune the cut thresholds!
- Problem :(prohibitively large grids, $\mathcal{O}(10^{14})$ points
- Solution ;) let's ask data scientists
- **Bayesian optimization algorithms do the job**  `HyperOpt`

Cut-and-count should be a science!

Tuning cut thresholds

- Simple idea: let's tune the cut thresholds!
- Problem :(prohibitively large grids, $\mathcal{O}(10^{14})$ points
- Solution ;) let's ask data scientists
- **Bayesian optimization algorithms do the job** ➡ `HyperOpt`
- **Just a week ago! Paper on Random Search (1706.09907,Bhat et al)**



Turning cut-and-count into a science

SMBO+TPE, aka Bayesian Optimization

- Sequential Model-based Global Optimization of $f(x)$
- Propose a cheap surrogate function to propose x^*, $y^* = f(x^*)$

Turning cut-and-count into a science

SMBO+TPE, aka Bayesian Optimization

- Sequential Model-based Global Optimization of $f(x)$
- Propose a cheap surrogate function to propose x^*, $y^* = f(x^*)$

SMBO(f, M_0, T, S)

1 $\mathcal{H} \leftarrow \emptyset$,
2 For $t \leftarrow 1$ **to** T,
3 $x^* \leftarrow \text{argmin}_x \; S(x, M_{t-1})$,
4 Evaluate $f(x^*)$, ▷ *Expensive step*
5 $\mathcal{H} \leftarrow \mathcal{H} \cup (x^*, f(x^*))$,
6 Fit a new model M_t to $\mathcal{H}$.
7 **return** $\mathcal{H}$

- $S(x) = EI_{y^*}(x) = \int_{-\infty}^{\infty} max(y^* - y, 0) p_M(y|x) dy$
- TPE algorithm to estimate $p_M(x|y) \Longrightarrow$ `HyperOpt`

Turning cut-and-count into a science.

Tuning cut thresholds

- Alves, Ghosh, Sinha, arXiv:1704.07395, submitted to PRD
- Generative ML model `HyperOpt` to maximize an AMS function

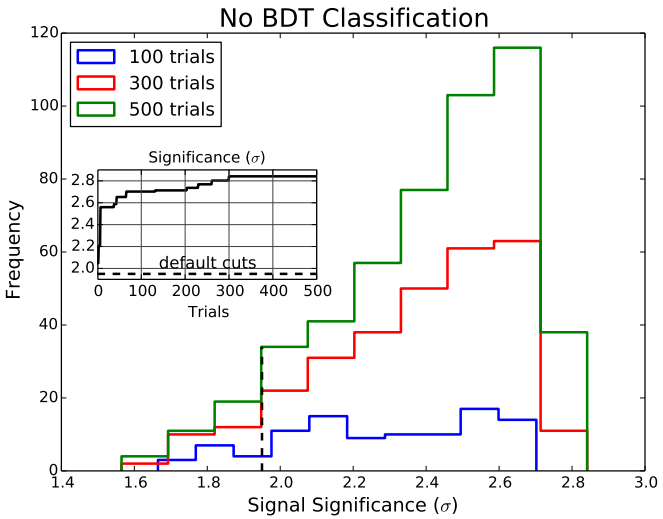
$$\frac{S}{\sqrt{B + (\varepsilon B)^2}}$$

Turning cut-and-count into a science.

Tuning cut thresholds

- Alves, Ghosh, Sinha, arXiv:1704.07395, submitted to PRD
- Generative ML model `HyperOpt` to maximize an AMS function

$$\frac{S}{\sqrt{B + (\varepsilon B)^2}}$$



$\sim$50% improvement compared to the best previous result!

Using BDTs

Cut+ML

- AI algorithms like BDT and DNN improve event classification

Using BDTs

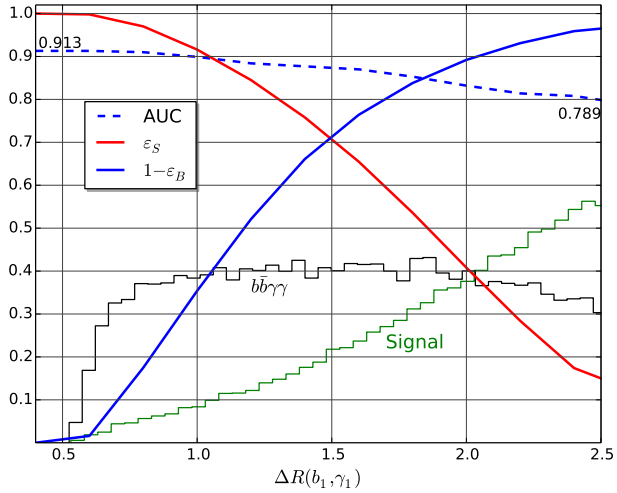
Cut+ML

- AI algorithms like BDT and DNN improve event classification
- BDTs are faster and easier to train, used by LHC people
- Many tool in the market, `TMVA`, `Sklearn`, `XGBoost`

Using BDTs

Cut+ML

- AI algorithms like BDT and DNN improve event classification
- BDTs are faster and easier to train, used by LHC people
- Many tool in the market, `TMVA`, `Sklearn`, `XGBoost`
- Overlooked issue: hard cuts decrease ML performance

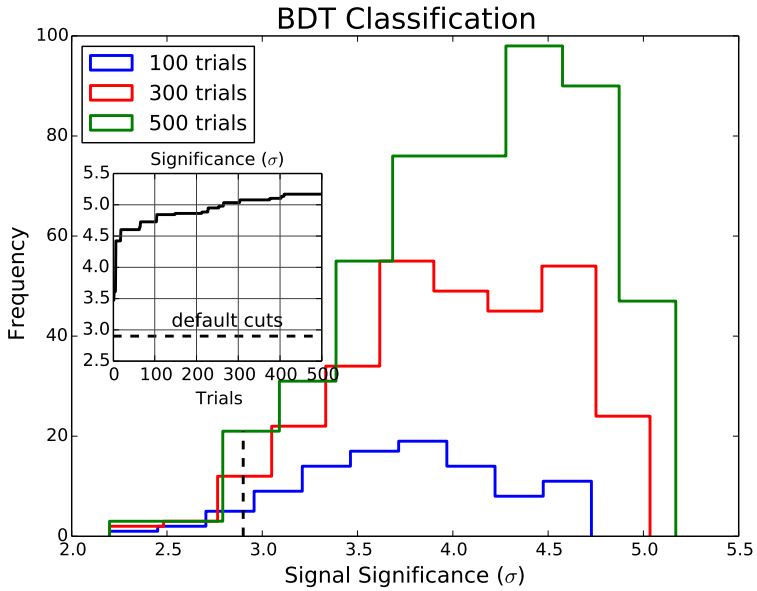


- Why not optimizing cuts and ML tool at the same time?

Using BDTs

Cut+ML

- Tuning cut thresholds and ML hyperparameters
- Another huge improvement in significance!



We're getting closer to the dreamed 5σ! But wait...systematics use to screw our lives

Using BDTs

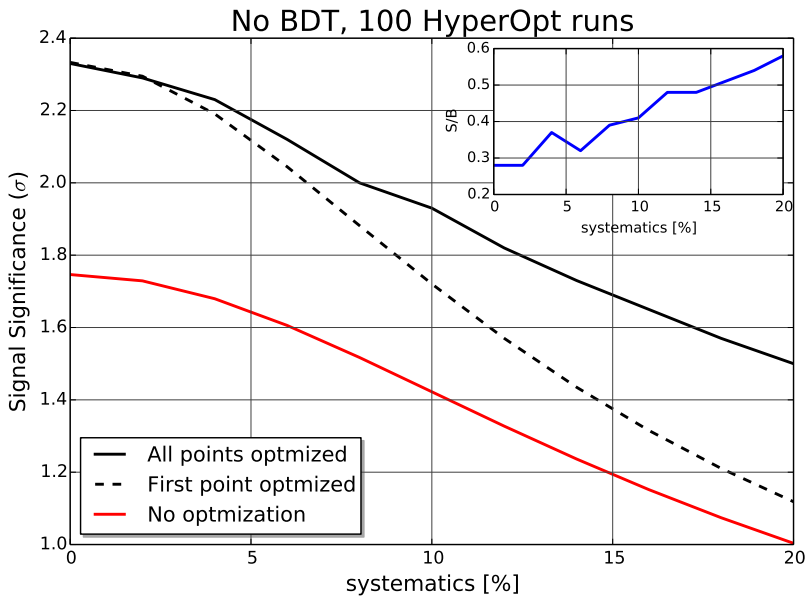
Taking systematic uncertainties into account

- If S/B is small, systematics become important

Using BDTs

Taking systematic uncertainties into account

- If S/B is small, systematics become important
- How does `HyperOpt` behave when ε comes into play?



The algorithm learns how to tame the systematics!

Using BDTs

Taking systematic uncertainties into account

- Automatized cuts + ML hyperparameters

$$p_T(1) > 72 \text{ GeV}, \; p_T(2) > 20 \text{ GeV}$$
$$\Delta R_{ij} > 0.15, \; \Delta R_{ii} < 3.6$$
$$M_{b\bar{b}\gamma\gamma} > 370 \text{ GeV}, \; p_{T_{ii}} > 145 \text{ GeV}, \; M_{b_1\gamma_1} > 100 \text{ GeV}$$
$$|M_{bb} - m_h| < 27 \text{ GeV}, \; |M_{\gamma\gamma} - m_h| < 11 \text{ GeV}$$
$$\texttt{number of trees} = 157$$
$$\texttt{learning rate} = 0.101$$
$$\texttt{maximum tree depth} = 14$$
$$\texttt{min\_child\_weight} = 5$$

Prevents overfitting, tune the cuts, tame systematics at the same time with $\mathcal{O}(10^2)$ trials!

Using BDTs+MVA

Final results

- With cross-checked realistic background simulations
- Taking systematics into account
- Using better AMS metrics
- Jointly tuning cuts+BDT+MVA

systematics (%)	Cut-and-count	BDT	MVA
0	2.34[1.76]	3.88	5.05
10	**1.93**[1.43]	**3.57**	**4.64**
20	1.51[1.0]	3.10	3.60

Very close to discovery @ 14TeV LHC with 3 ab^{-1}!

Further research

- From BDT to DNN
- Measure SM and non-SM λ_{hhh} in the HL-LHC
- Measure SM and non-SM λ_{hhhh} in 100TeV colliders
- Test the technique with classic signals: DM, SUSY, other Higgs channels
- Getting the tool implemented in a general purpose package like MadAnalysis

Further research

- From BDT to DNN
- Measure SM and non-SM λ_{hhh} in the HL-LHC
- Measure SM and non-SM λ_{hhhh} in 100TeV colliders
- Test the technique with classic signals: DM, SUSY, other Higgs channels
- Getting the tool implemented in a general purpose package like MadAnalysis

THANK YOU

THANK YOU VERY MUCH!!!