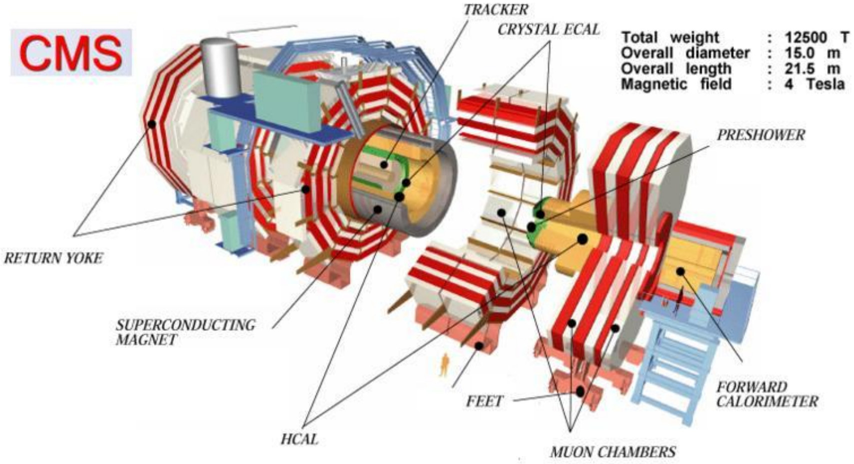
Search for the decay of a heavy Higgs boson into two lighter Higgs bosons of different mass in final states with b quarks and tau leptons at the LHC

23.05.22

In this talk...

A beyond-Standard-Model search for additional Higgs bosons:

- Inspired by the next-to-minimal supersymmetric Standard Model (NMSSM)
- Majority of background estimation from data
- Multiclassification using a neural net
- Statistical inference to set upper exclusion limits on the signal cross section x BR



Why search for BSM physics?

The Standard Model (SM) provides our current best description of fundamental particles

However: SM has experimental and theoretical shortcomings (e.g. Gravitation, Dark Matter) → search for SM extension or new fundamental theory

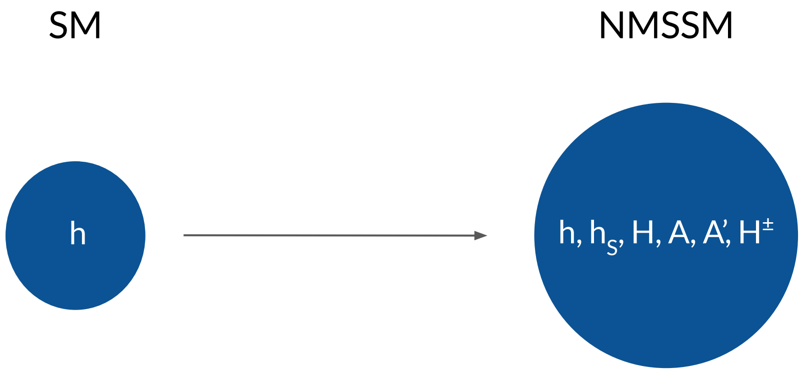
Supersymmetric theories:

- Class of SM extensions
- Imposes additional symmetry on SM → new particles
- Can be parametrised in many ways e.g. the next-to-minimal supersymmetric Standard Model (**NMSSM**)

The NMSSM Higgs Sector (1)

Next-to-minimal realisation of supersymmetry. Extends the SM Higgs sector with an additional SU(2) doublet + singlet

→ **NMSSM Higgs Sector contains 7 Higgs boson with rich phenomenology**



The NMSSM Higgs Sector (2)

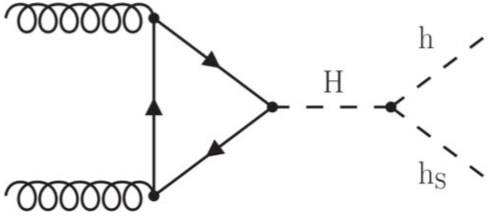
Next-to-minimal realisation of supersymmetry. Extends the SM Higgs sector with an additional SU(2) doublet + singlet

→ **NMSSM Higgs Sector contains 7 Higgs boson with rich phenomenology**

Scalars: h, h_s, H
Pseudoscalars: A, A'
Charged: $H^{\pm}$

We consider the 3 scalars in this analysis:

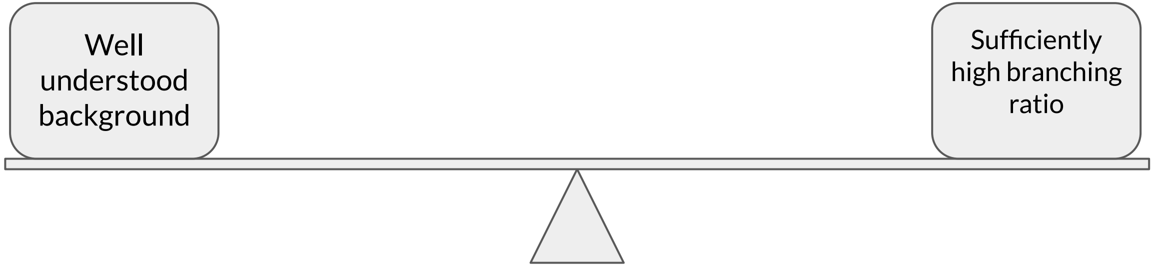
- Consider h to have properties of previously discovered Higgs boson
- h_s is considered 'light', H is considered 'heavy'
- $m(h_s)$ and m(H) are free parameters



$H \rightarrow hh_s$ final states with b quarks and tau leptons (1)

We target final states with a b quark and tau lepton pair (i.e. $\mathbf{H \rightarrow hh_s \rightarrow bb\tau\tau}$).

- Decays to b quarks (which hadronise to form b-jets):
  - (+) **high branching ratio**
  - (-) less well understood background
- The τ leptons are identified via their decay products (e, μ, hadrons) :
  - (+) **well understood background**
  - (-) lower branching ratio

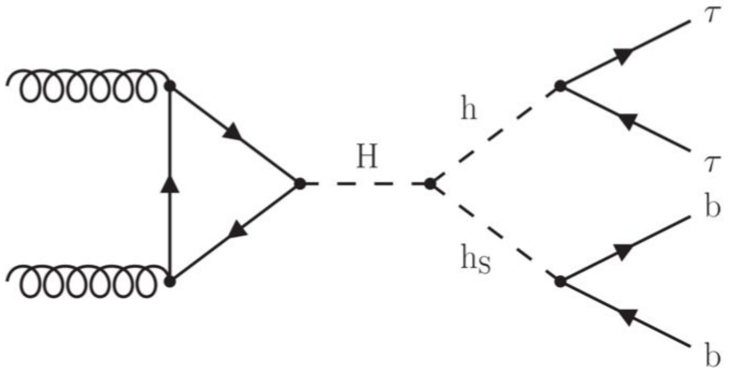


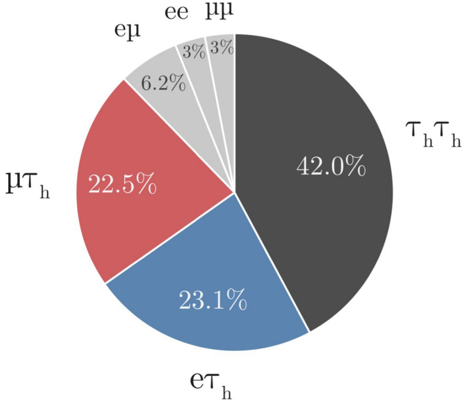
Felix Heyen - felix.heyen@vub.be Inter-University Institute for High Energies, VUB, Brussels

$H \rightarrow hh_s$ final states with b quarks and tau leptons (2)

An analysis targeting $\mathbf{H \rightarrow h(\tau\tau)h_s(bb)}$ was published in 2021 by the CMS collaboration [1]:

- Considered to full CMS Run-2 data (2016 + 2017 + 2018 data taking periods)
- Targets 3 most common τ-pair decay modes





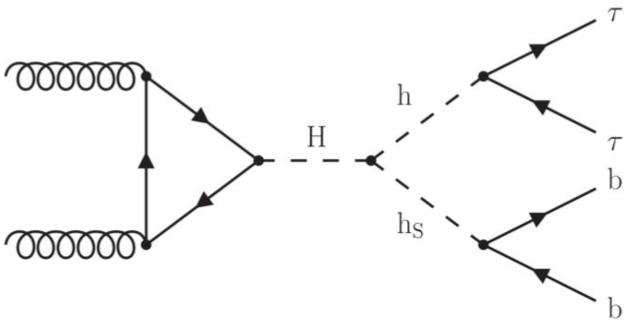
[1] https://doi.org/10.1007/JHEP11(2021)057

$H \rightarrow hh_s$ final states with b quarks and tau leptons (3)

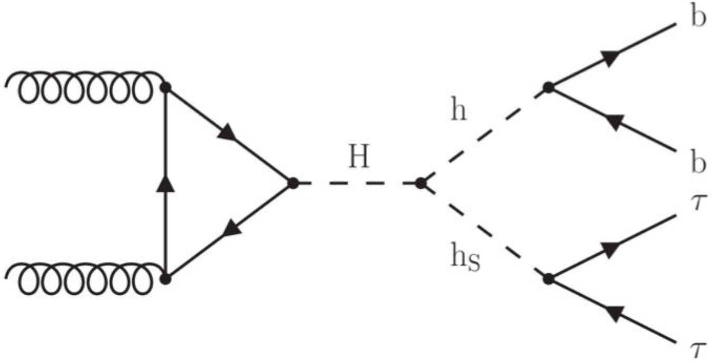
This analysis is a proof of concept to extend [1] with the **$H \rightarrow h(bb)h_s(\tau\tau)$** final state.

A number of simplifications w.r.t original analysis are made:

- Only consider 2018 period
- Only consider $\tau_h\tau_h$ decay mode



+

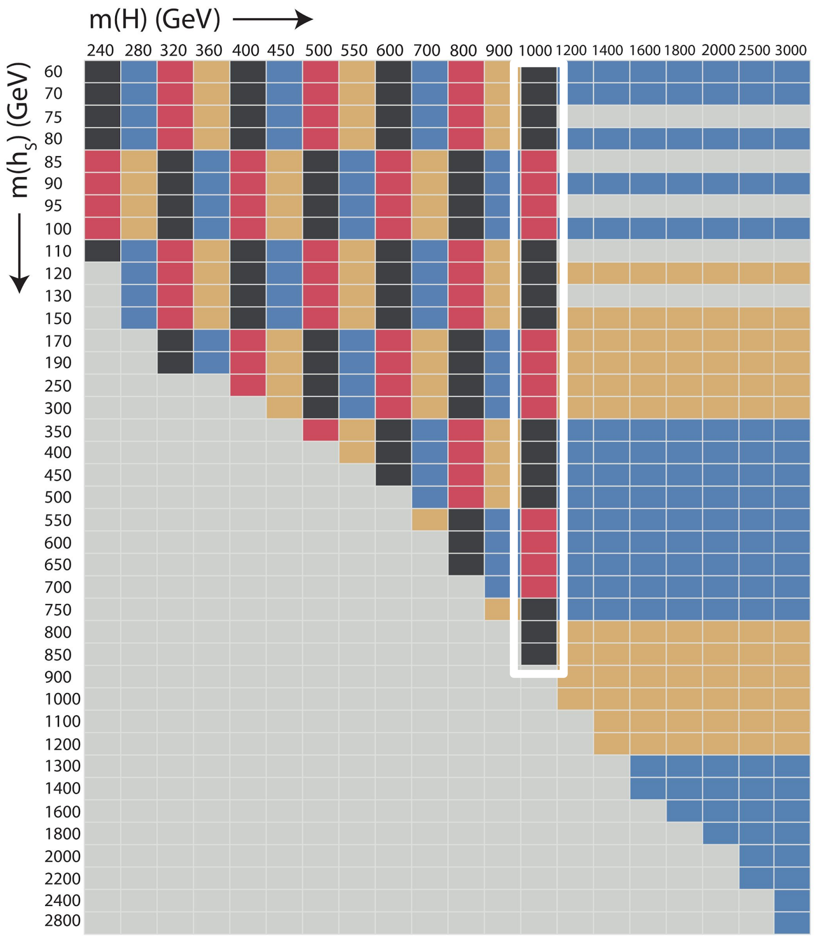


[1] https://doi.org/10.1007/JHEP11(2021)057

Signal Simulation

$H \rightarrow h(bb)h_s(\tau\tau)$ / $h(\tau\tau)h_s(bb)$ samples have been privately produced:

- MadGraph5_aMC@NLO
- 27 mass combinations with **$m(H) = 1000$ GeV** and **$m(h_s) = [60, 850]$ GeV** were produced (on-shell)
- 200,000 events for each mass point
- Production required ~ 100,000 CPU hours (local resources, NEMO)



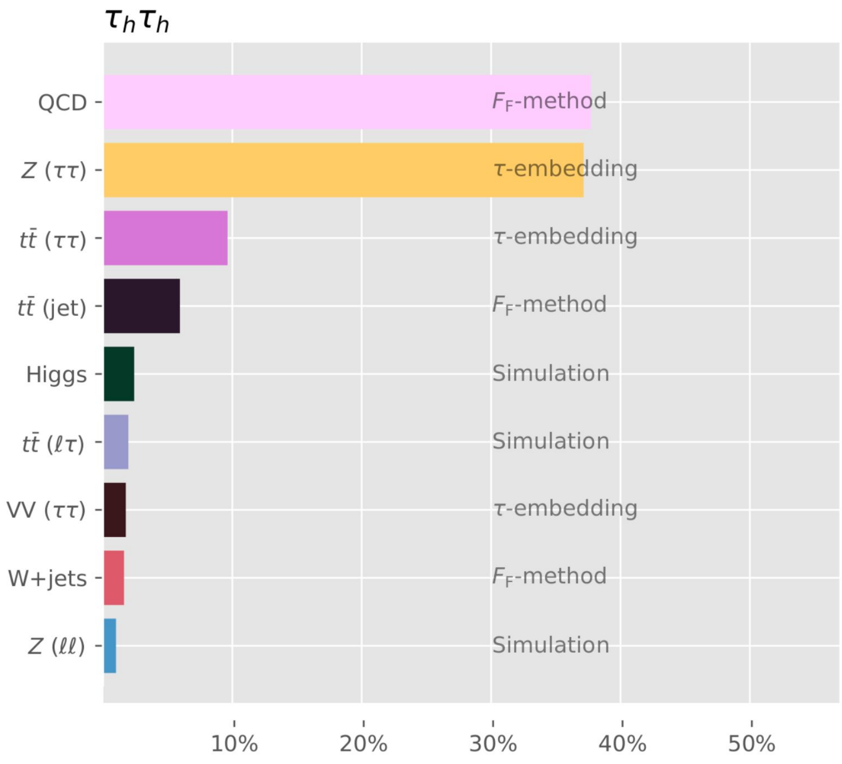
Background Estimation (1)

Processes that produce

- A genuine or ‘fake’ tau lepton pair
- A genuine or ‘fake’ b-Jet pair

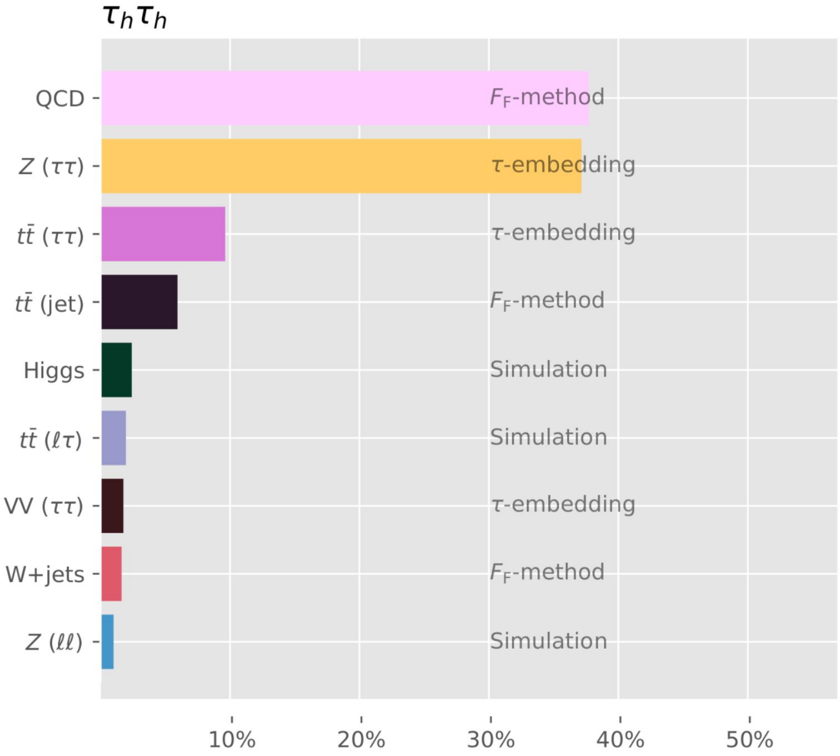
Most prominent in $\tau_h\tau_h$ channel:

- QCD multijet production
- $Z \rightarrow \tau\tau$



Background Estimation (2)

1. **τ-embedding** for all events with two genuine tau leptons
2. **F_F method** for events entering the analysis due to misidentified quark/gluon induced jets
3. **Monte Carlo (MC) simulation** for SM h → ττ events + misc.

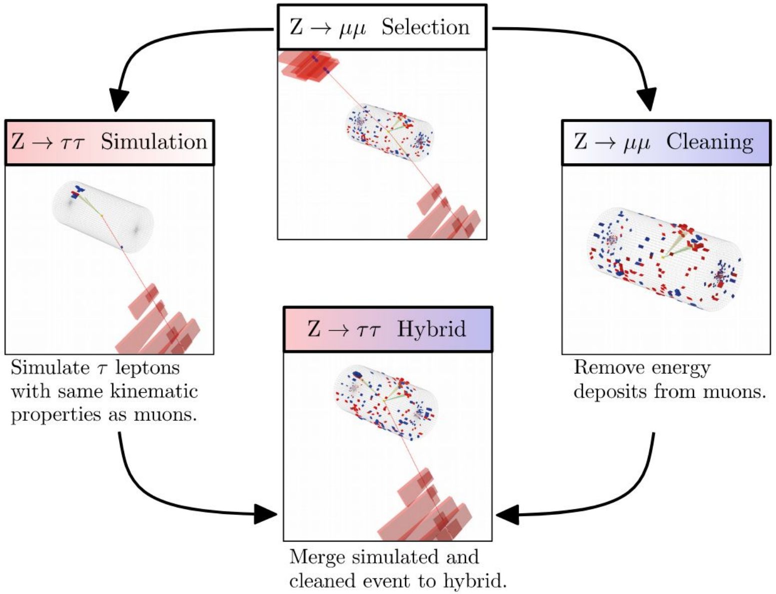


Background Estimation: Tau-Embedding

Used to model background processes with genuine τ-pairs:

1. Selecting $\mu\mu$ events from data
2. Remove muon energy deposits
3. Simulate τ-leptons in place of muons
4. Merge simulation with original event

2019 JINST 14 P06032



Reduced associated uncertainties vs full MC estimation

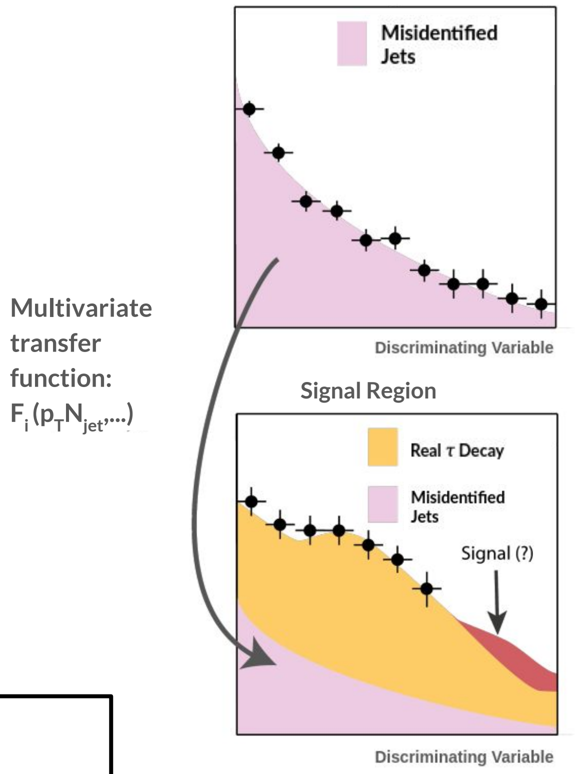
Background Estimation: F_F Method

Used to model jets incorrectly identified as τ_h:

1. Bckgd. processes are selected by process specific properties e.g.:
   - QCD: same sign
   - tt: additional leptons
2. Define sideband region enriched in misidentified jets
3. Calculate multivariate transfer function F_i $\rightarrow$ apply to sideband region

No MC corrections needed in estimating signal region

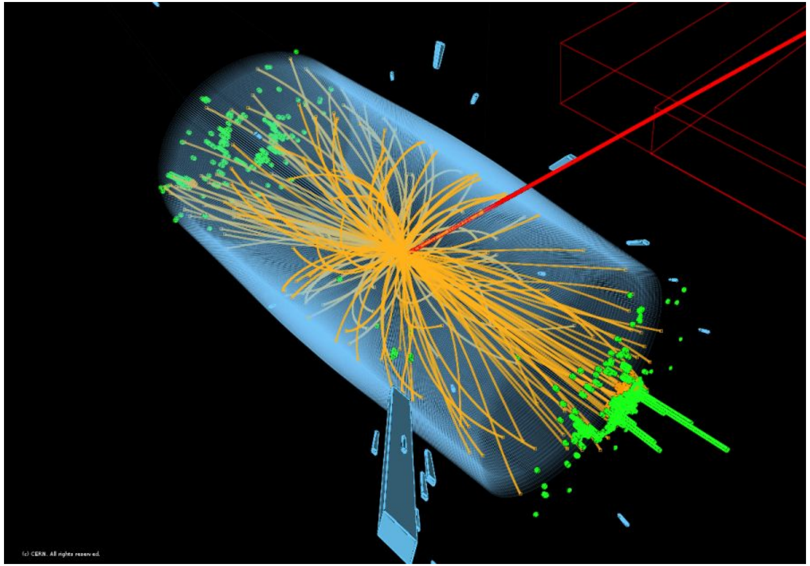
Sideband region: e.g. same sign QCD taus



Background Estimation: Simulation

Used to estimate a number of minor backgrounds e.g.:

- SM h → ττ
- Z → ee, Z → μμ
- tt(*l*τ)



Events are weighted with a factor

$$\beta = L_{\text{int}} \cdot \sigma \cdot \frac{1}{N_{\text{sim}}}$$

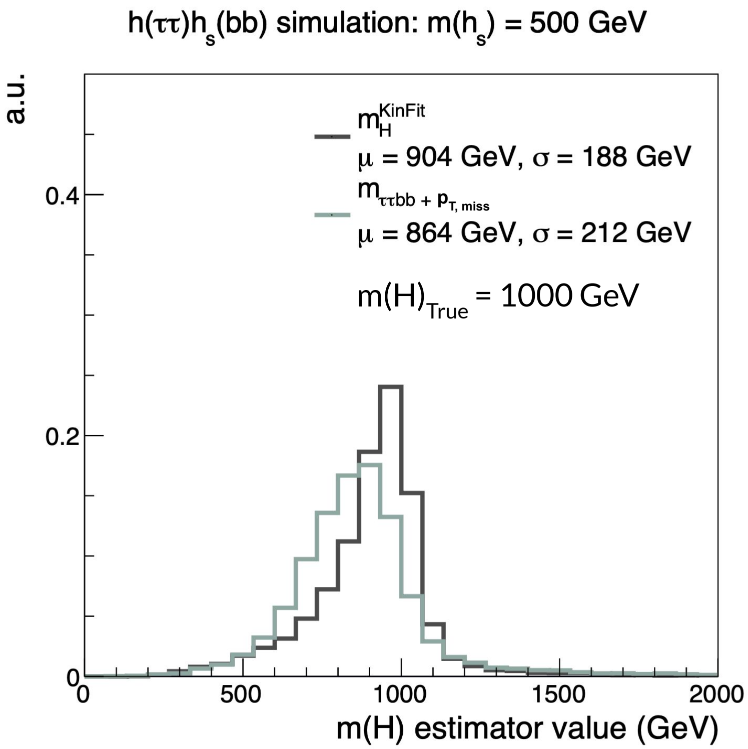
- L_{int}: Integrated luminosity
- σ: process cross section
- N_{sim}: # of sim. events

Kinematic Fit of bb + $\tau\tau$ Mass (1)

χ^2 estimate of the mass of the bb$\tau\tau$-system:

- Minimise χ^2 for range of $m(h_s)$ hypotheses ([30 GeV, 3000 GeV])
- Fit with lowest minimal χ^2 value is chosen
- Output: $m(H)$ estimator, $m(h_s)$, χ^2 value of fit
- Improves resolution of the reconstructed mass $m_{bb\tau\tau}$

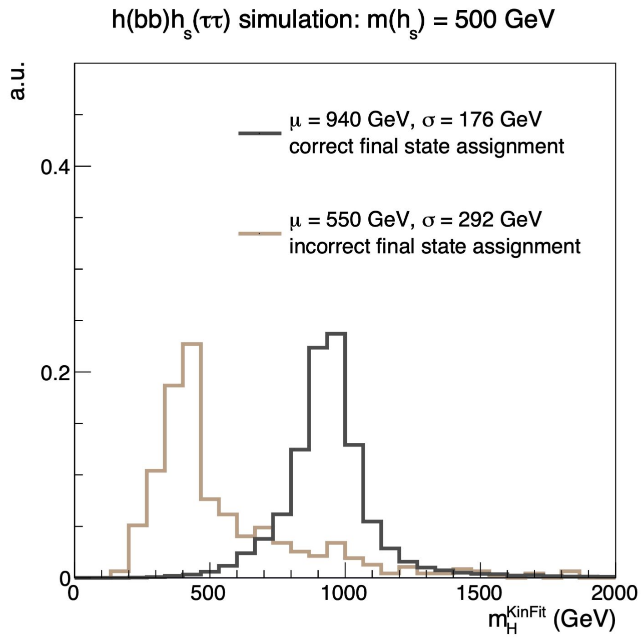
$$\chi^2 = \chi^2_{b_1} + \chi^2_{b_2} + \chi^2_{\text{recoil}}$$

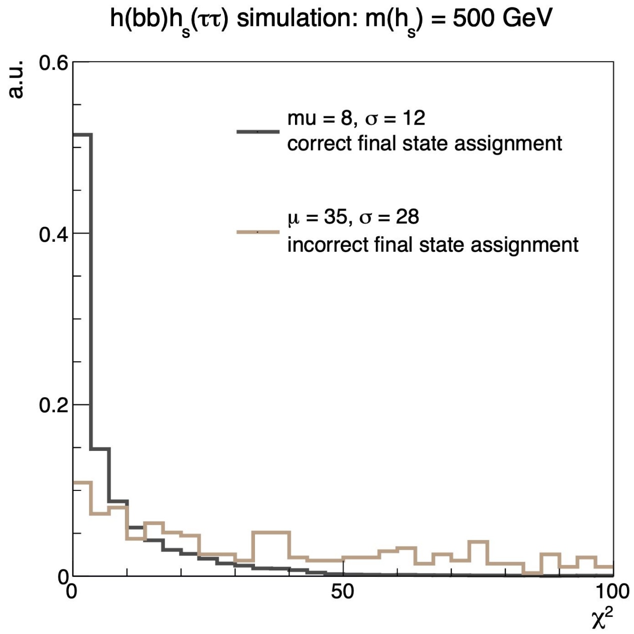


CMS Simulation: Work in Progress

Kinematic Fit of bb + $\tau\tau$ Mass (2)

Fit performed once for each signal final state, outputs of both are saved



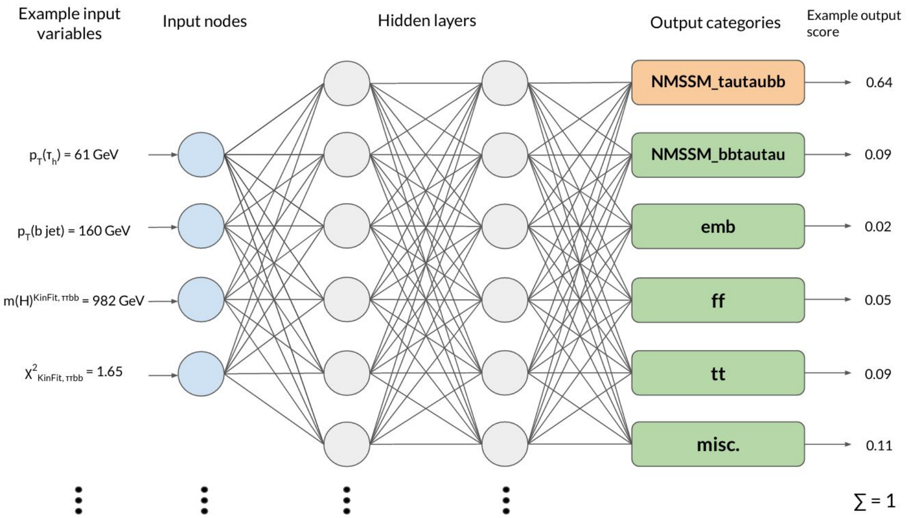


CMS Simulation: **Work in Progress**

Event Classification (1)

Multiclass classification based on **Neural Net** (NN) with 4 background categories and 2 signal categories:

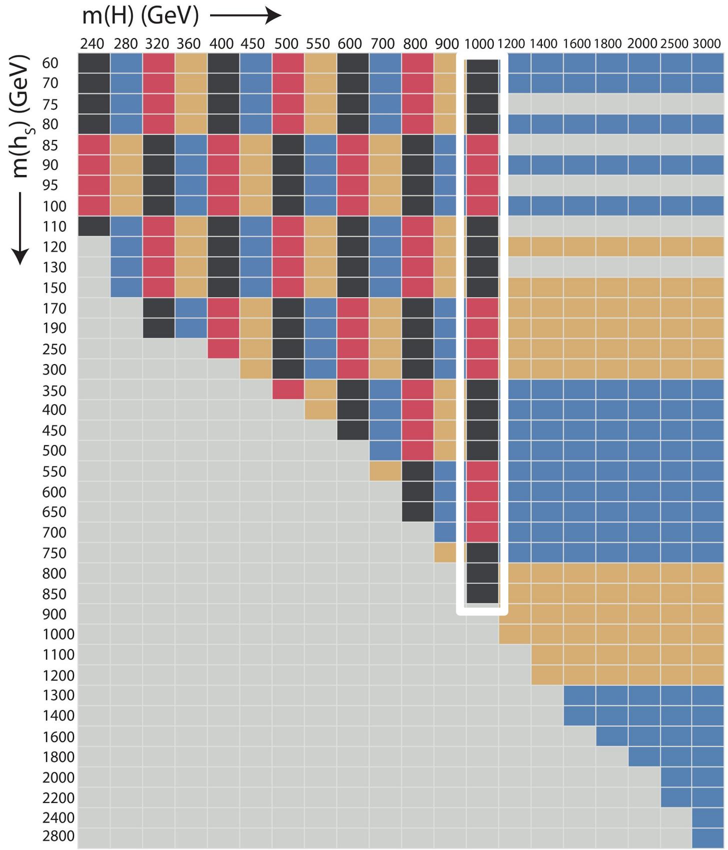
- 28 Input features
- NN returns a probability-like score for each category. The event is assigned to category with highest score with value NN_{max}



Event Classification (1)

Multiclass classification based on **Neural Net** (NN) with 4 background categories and 2 signal categories:

- 28 Input features
- NN returns a probability-like score for each category. The event is assigned to category with highest score with value NN_{max}
- 7 NNs are trained for 7 different mass groupings

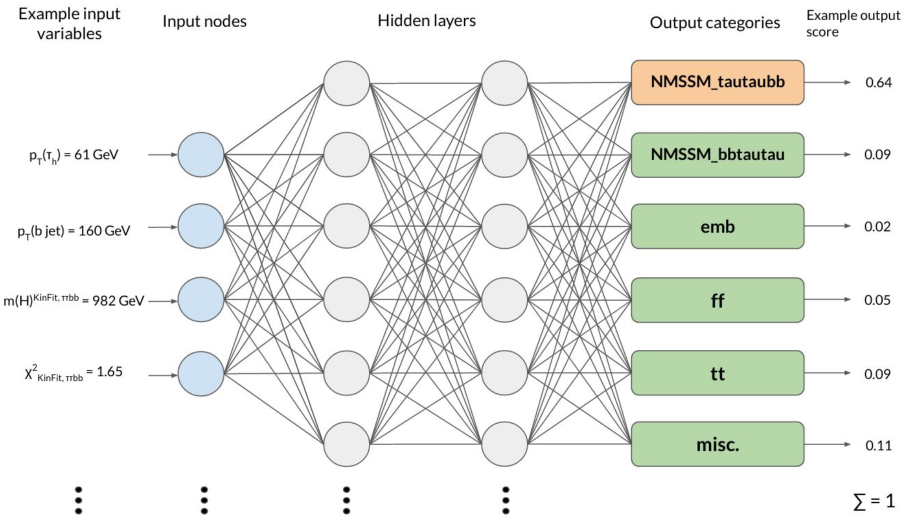


Event Classification (2)

NN features:

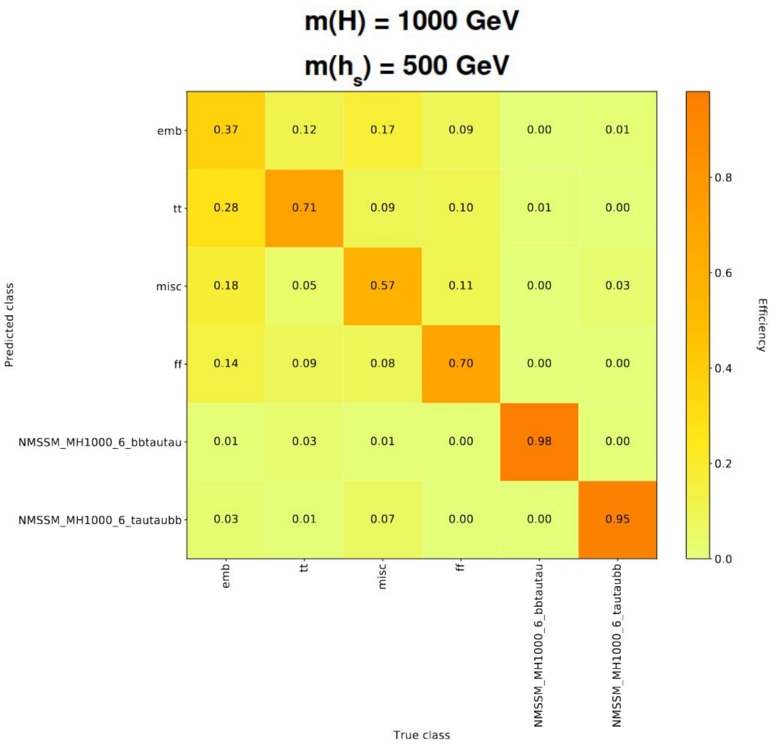
- Fully connected feed-forward network
- 2 hidden layers with 200 nodes
- Minimise the empirical risk function (*balanced batches, N = 600*)

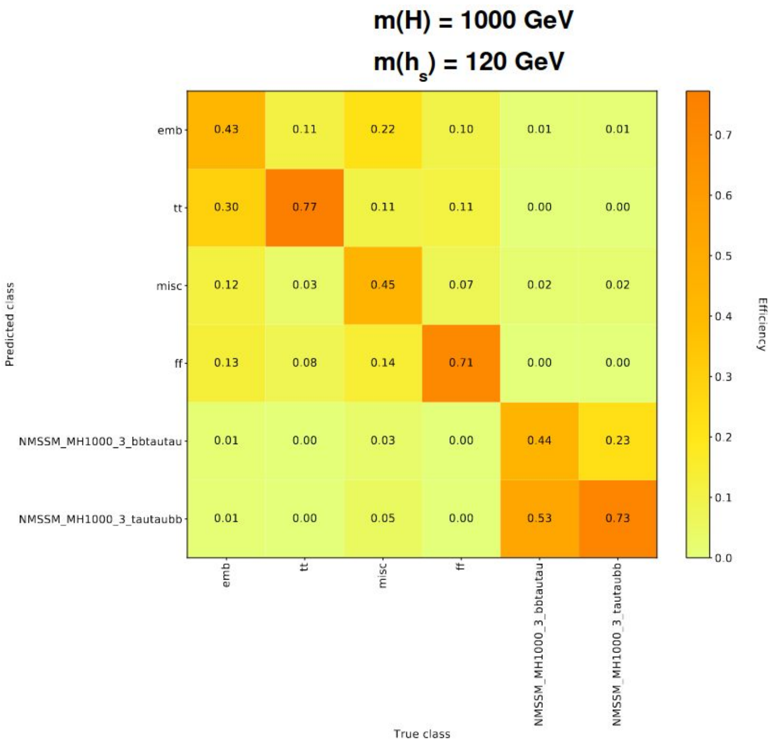
$$R_{\mathrm{emp}} = -\sum_{j=1}^{N} w^j \sum_{i=1}^{C} t_i^j \cdot \log(y_i^j)$$



Event Classification (3)

Comparison of different $m(h_s)$ mass points:





Statistical Inference

Calculate 95% CL $\mathrm{CL_s}$ limits on $\mathbf{H \rightarrow hh_s}$ assuming absence of signal:

- Calculate likelihood function based on a poisson ansatz for individual bin content

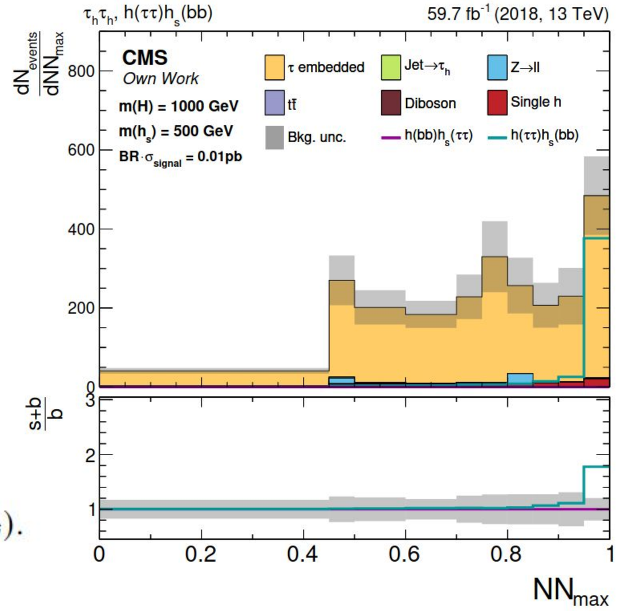
$$\mathcal{P}(k\,;\,\lambda) = \frac{\lambda^k e^{-\lambda}}{k!} \qquad \mathcal{L}(d\,|\,\mu \cdot s(\boldsymbol{\theta}) + b(\boldsymbol{\theta})) = \prod_{i \in \text{bins}} \mathcal{P}(d_i\,|\,\mu \cdot s(\boldsymbol{\theta}) + b(\boldsymbol{\theta})) \text{ x } \prod_{j \in \text{nuis.}} C(\hat{\theta}_j\,|\,\theta_j).$$

- Profile likelihood test statistic

$$q_\mu = -2\ln\left(\frac{\mathcal{L}(d\,|\,\mu \cdot s(\hat{\boldsymbol{\theta}}_\mu) + b(\hat{\boldsymbol{\theta}}_\mu))}{\mathcal{L}(d\,|\,\hat{\mu} \cdot s(\hat{\boldsymbol{\theta}}) + b(\hat{\boldsymbol{\theta}})}\right)$$

- $\mathrm{CL_s}$ method

$$\mathrm{CL_s} = \frac{p_\mu}{1 - p_0}$$



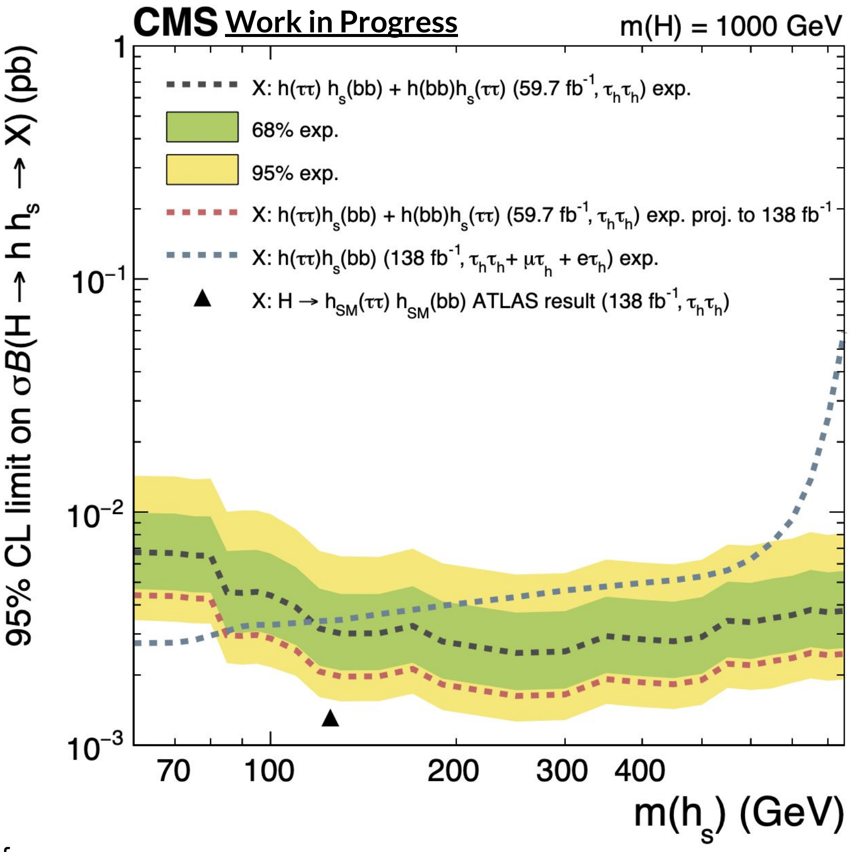
- Analysis is model-independent until this point
- Can be interpreted in terms of any X→xh model by scaling expected yields of the signal final states (assume equal yields here)

Results (1)

Analysis sensitivity is gauged by the calculated expected exclusion limits on

$$\sigma\,B(H \to hh_s \to bb + \tau\tau)$$

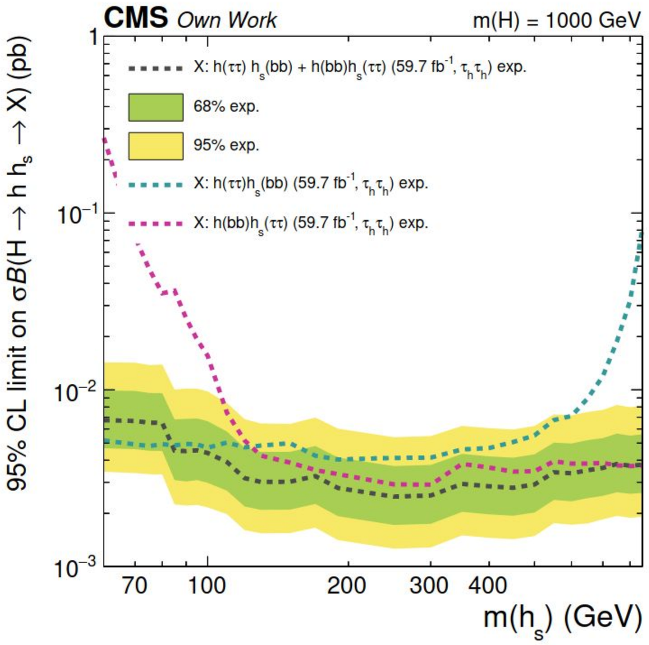
Combination of **$h(bb)h_s(\tau\tau)$ + $h(\tau\tau)h_s(bb)$** final states (**red**) improves upon sensitivity of $h(\tau\tau)h_s(bb)$ analysis (**blue**) significantly for $m(h_s)$ > 80 GeV

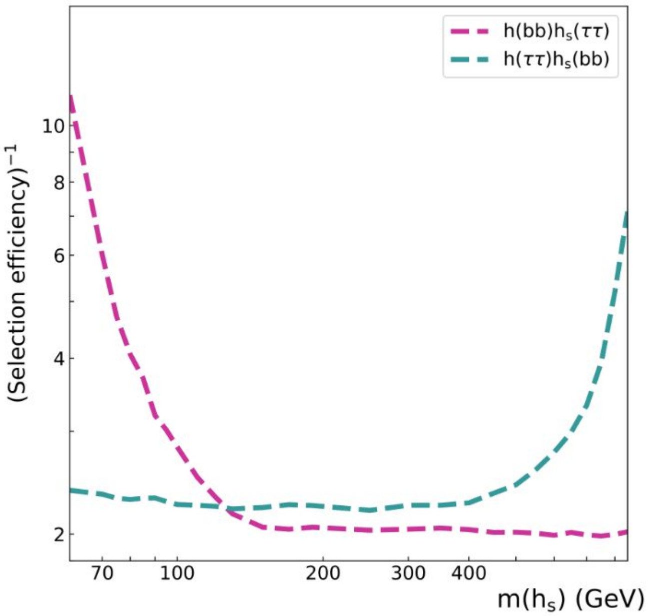


[2] http://cdsweb.cern.ch/record/ 2777236/files/ATLAS-CONF-2021-030.pdf

Results (2)

Taking a look at the results split by signal final states:





→ Analysis sensitivity is driven by selection efficiency

Conclusion

A novel search for **H** → $\mathbf{hh_s}$ decays was presented:

- NMSSM inspired but largely model independent
- Employs background estimation methods such as τ-embedding and the F_F method
- Multiclassification is performed using multivariate NN with two signal classes
- Significant improvement in exclusion sensitivity is achieved when $\mathbf{h(bb)h_s(\tau\tau) + h(\tau\tau)h_s(bb)}$ final states are considered simultaneously

Thank you!

Backup - F_F Method