



MINISTÈRE  
DE L'ENSEIGNEMENT SUPÉRIEUR,  
DE LA RECHERCHE  
ET DE L'INNOVATION  
REPUBLIQUE FRANÇAISE



PHAST  
PHYSIQUE  
ET ASTROPHYSIQUE  
UNIVERSITÉ DE LYON



# Search for additional heavy Higgs bosons decaying to tau lepton pair in the CMS experiment at LHC with machine learning techniques

Ph.D. thesis defense

Lucas TORTEROTOT, under the direction of Colin BERNET

Institut de Physique des deux Infinis – Lyon

July 8, 2021



## Why do we search for...?

### Current standard model status

- Robust and predictive (top quark,  $W$ ,  $Z$  and one Higgs boson...)
- Still not good enough, unable to explain some observations such as:
  - ▶ dark matter
  - ▶ matter vs antimatter asymmetry
  - ▶ naturalness problem
  - ▶ ...
- Go beyond with a new model!
- Consequences of this new model? **Test it!**

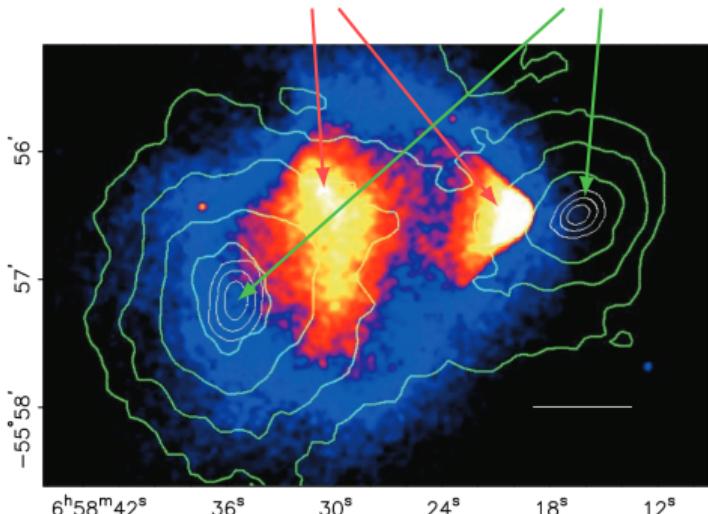
▷ D. Clowe et al. "A Direct Empirical Proof of the Existence of Dark Matter". *Astrophysical Journal* **648**.2 (Aug. 2006). DOI: 10.1086/508162.

## Why do we search for...?

### Current standard model status

- Robust and predictive (top quark,  $W$ ,  $Z$  and one Higgs boson...)
- Still not good enough, unable to explain some observations such as:
  - ▶ dark matter →
  - ▶ matter vs antimatter asymmetry
  - ▶ naturalness problem
  - ▶ ...
- Go beyond with a new model!
- Consequences of this new model? **Test it!**

Galaxies from: **visible matter** **gravitational lensing**



Difference due to **dark matter**!

▷ D. Clowe et al. "A Direct Empirical Proof of the Existence of Dark Matter". *Astrophysical Journal* **648**.2 (Aug. 2006). DOI: 10.1086/508162.

# Keywords in title

Search for **additional heavy Higgs bosons decaying to tau lepton pair** in the **CMS experiment at LHC**

# Keywords in title

Search for **additional heavy Higgs bosons decaying to tau lepton pair** in the **CMS experiment at LHC**

**Part I**  
*Phenomenology*

# Keywords in title

Search for **additional heavy Higgs bosons decaying to tau lepton pair** in the **CMS experiment at LHC**

**Part I**  
*Phenomenology*

**Part II**  
*Experimental device*

# Keywords in title

Search for **additional heavy Higgs bosons decaying to tau lepton pair** in the **CMS experiment at LHC**

Part I  
Phenomenology

Part II  
Experimental device

Part III  
 $H/A \rightarrow \tau\tau$  analysis

# Keywords in title

Search for **additional heavy Higgs bosons decaying to tau lepton pair** in the **CMS experiment at LHC**

Part I  
Phenomenology

Part II  
Experimental device

Part III  
 $H/A \rightarrow \tau\tau$  analysis

+ Part IV: **with machine learning techniques**

## 1 Phenomenology

## 2 Experimental device

## 3 $H/A \rightarrow \tau\tau$ analysis

## 4 Machine learning

# Higgs bosons in the MSSM

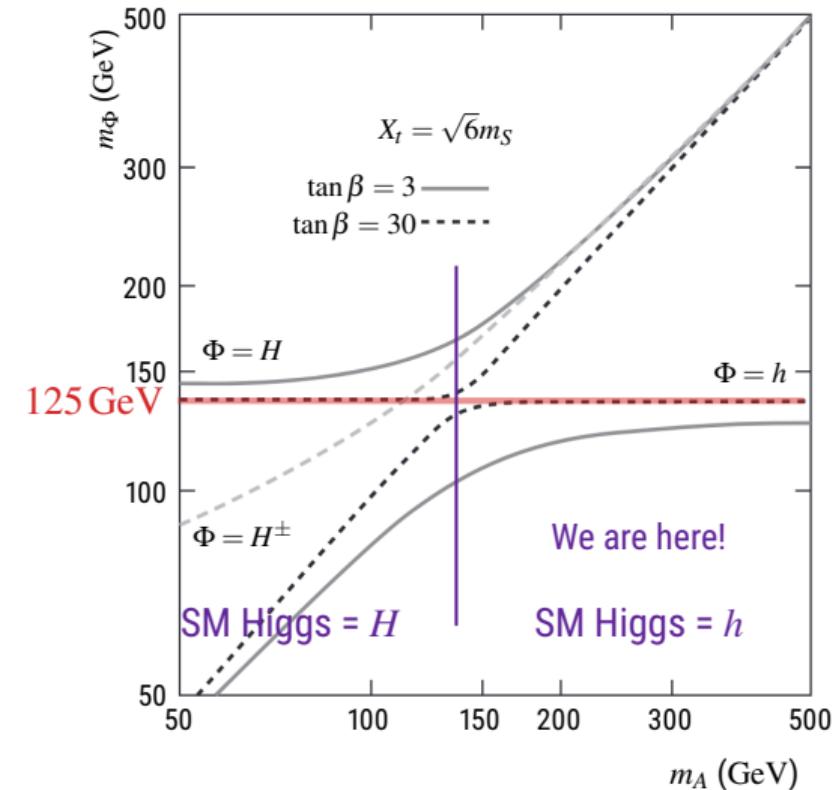
**Minimal Supersymmetric extension of Standard Model**

## 5 Higgs bosons

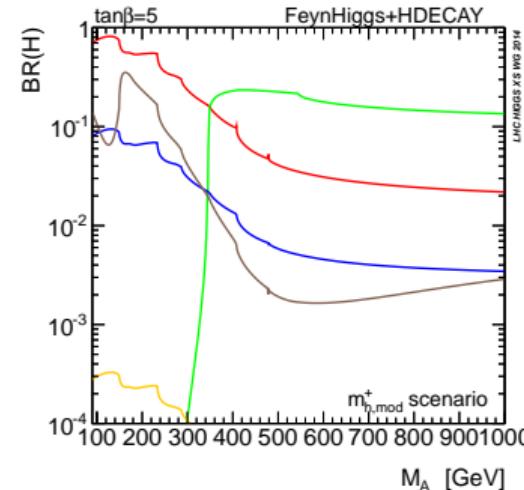
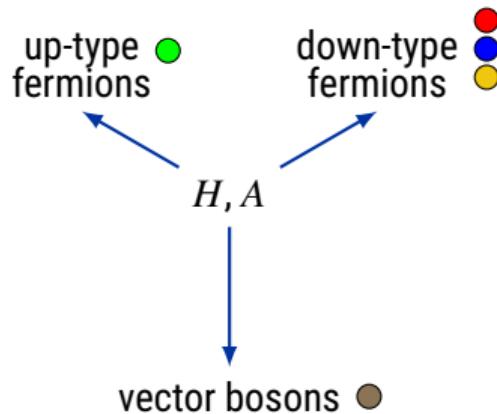
light scalar	$h$	SM or MSSM
heavy scalar	$H$	MSSM or SM
pseudo-scalar	$A$	MSSM
+ charged	$H^+$	MSSM
- charged	$H^-$	MSSM

Main parameters:  $m_A$  and  $\tan \beta$ .

- ▷ The CMS Collaboration. "Search for additional neutral MSSM Higgs bosons in the  $\tau\tau$  final state in  $pp$  collisions at  $\sqrt{s} = 13$  TeV". *Journal of High Energy Physics* **09.007** (Sept. 2018). doi: [10.1007/JHEP09\(2018\)007](https://doi.org/10.1007/JHEP09(2018)007).
- ▷ Y. Nagashima. *Beyond the Standard Model of Elementary Particle Physics*. Weinheim: Wiley-VCH, June 2014. URL: <http://cds.cern.ch/record/1620277>.

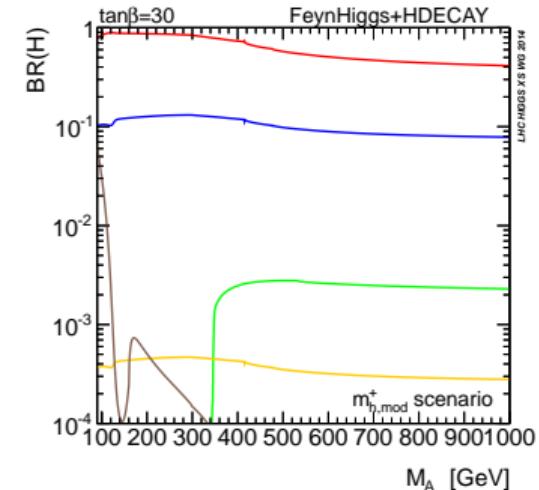
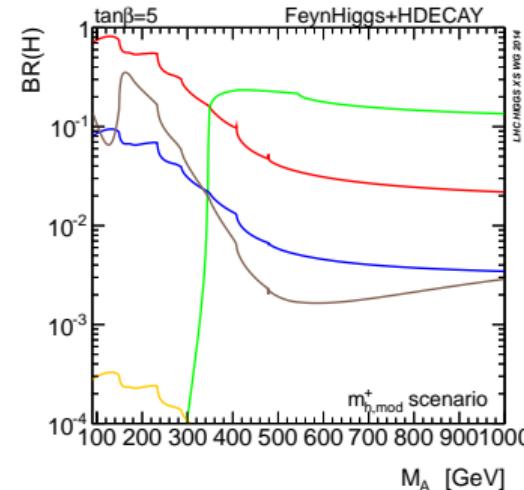
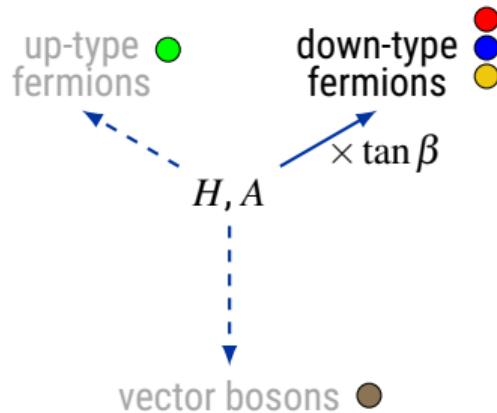


# $H/A \rightarrow \tau\tau?$



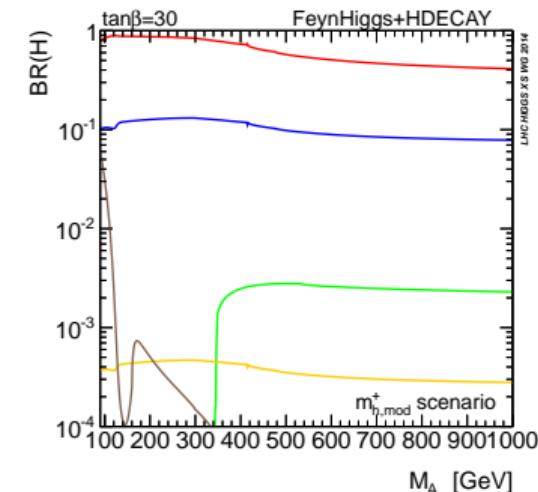
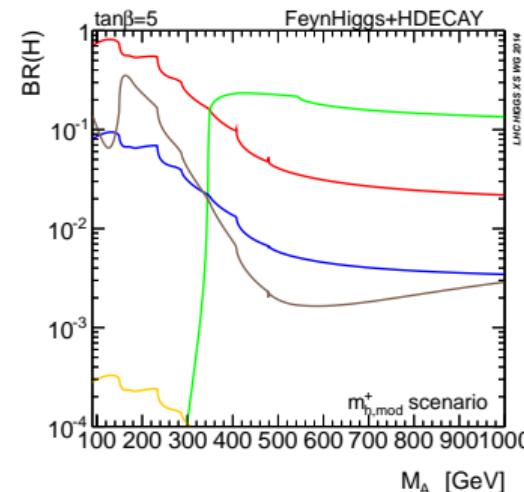
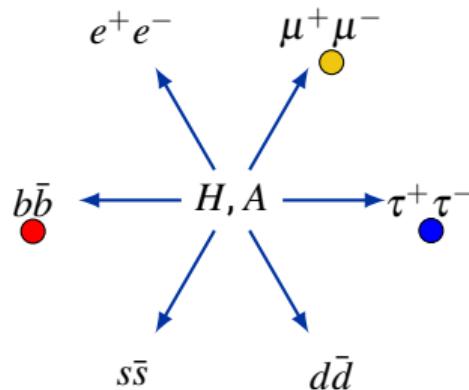
- ▷ The CMS Collaboration. "Search for additional neutral MSSM Higgs bosons in the  $\tau\tau$  final state in  $pp$  collisions at  $\sqrt{s} = 13\text{TeV}$ ". *Journal of High Energy Physics* **09**.007 (Sept. 2018). DOI: 10.1007/JHEP09(2018)007.
- ▷ LHC Higgs Cross Section Working Group. "Higgs Properties". *Handbook of LHC Higgs Cross Sections*. 3. CERN Yellow Reports: Monographs. Geneva: CERN, 2013. URL: <https://cds.cern.ch/record/1559921>.

# $H/A \rightarrow \tau\tau?$ – enhanced and suppressed couplings



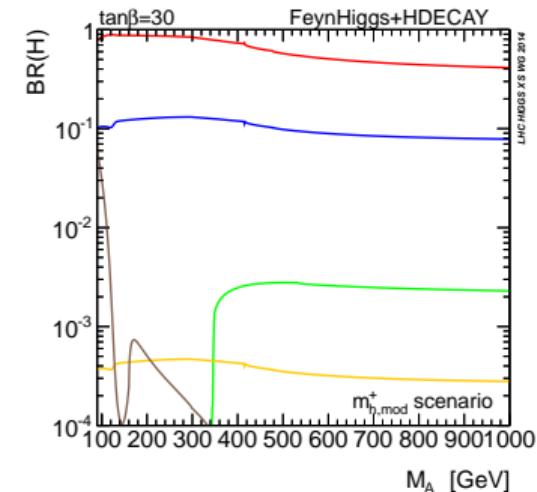
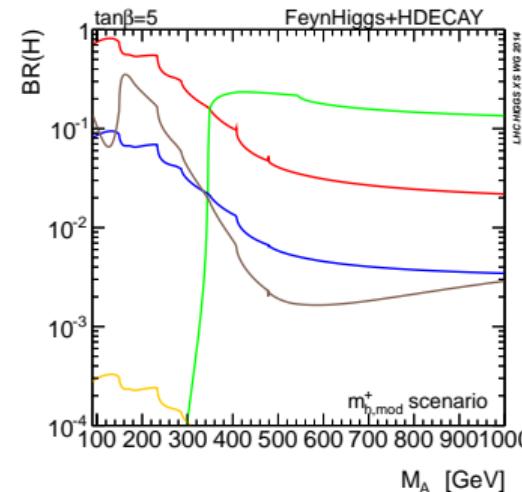
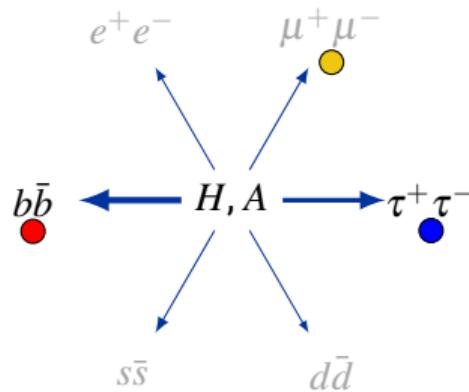
- ▷ The CMS Collaboration. "Search for additional neutral MSSM Higgs bosons in the  $\tau\tau$  final state in  $pp$  collisions at  $\sqrt{s} = 13\text{TeV}$ ". *Journal of High Energy Physics* **09**.007 (Sept. 2018). DOI: 10.1007/JHEP09(2018)007.
- ▷ LHC Higgs Cross Section Working Group. "Higgs Properties". *Handbook of LHC Higgs Cross Sections. 3. CERN Yellow Reports: Monographs*. Geneva: CERN, 2013. URL: <https://cds.cern.ch/record/1559921>.

# $H/A \rightarrow \tau\tau?$



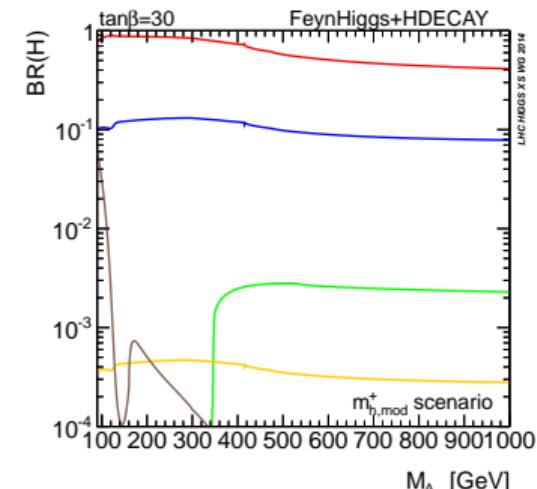
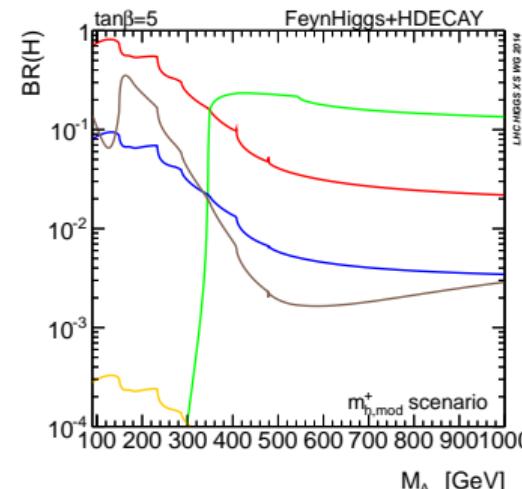
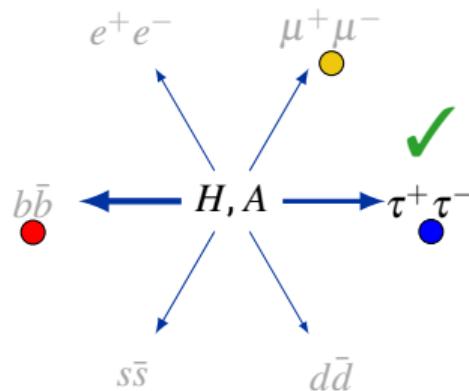
- ▷ The CMS Collaboration. "Search for additional neutral MSSM Higgs bosons in the  $\tau\tau$  final state in  $pp$  collisions at  $\sqrt{s} = 13\text{TeV}$ ". *Journal of High Energy Physics* **09**.007 (Sept. 2018). DOI: 10.1007/JHEP09(2018)007.
- ▷ LHC Higgs Cross Section Working Group. "Higgs Properties". *Handbook of LHC Higgs Cross Sections. 3. CERN Yellow Reports: Monographs*. Geneva: CERN, 2013. URL: <https://cds.cern.ch/record/1559921>.

# $H/A \rightarrow \tau\tau?$ – Higgs couplings and particle masses



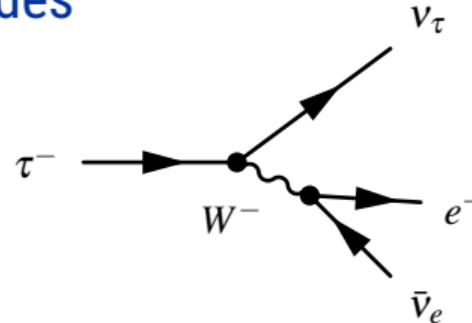
- ▷ The CMS Collaboration. "Search for additional neutral MSSM Higgs bosons in the  $\tau\tau$  final state in  $pp$  collisions at  $\sqrt{s} = 13\text{TeV}$ ". *Journal of High Energy Physics* **09**.007 (Sept. 2018). DOI: [10.1007/JHEP09\(2018\)007](https://doi.org/10.1007/JHEP09(2018)007).
- ▷ LHC Higgs Cross Section Working Group. "Higgs Properties". *Handbook of LHC Higgs Cross Sections. 3.* CERN Yellow Reports: Monographs. Geneva: CERN, 2013. URL: <https://cds.cern.ch/record/1559921>.

# $H/A \rightarrow \tau\tau?$ – avoid hadronic background

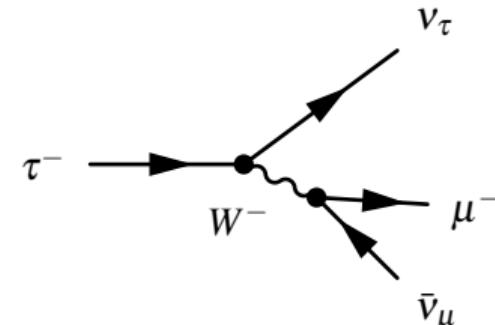
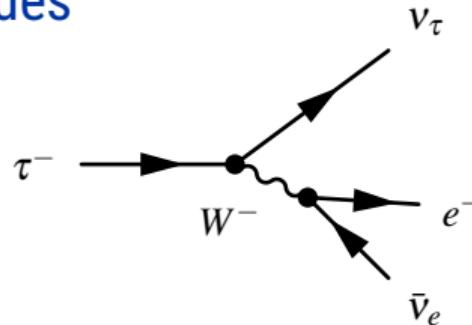


- ▷ The CMS Collaboration. "Search for additional neutral MSSM Higgs bosons in the  $\tau\tau$  final state in  $pp$  collisions at  $\sqrt{s} = 13\text{TeV}$ ". *Journal of High Energy Physics* **09**.007 (Sept. 2018). DOI: [10.1007/JHEP09\(2018\)007](https://doi.org/10.1007/JHEP09(2018)007).
- ▷ LHC Higgs Cross Section Working Group. "Higgs Properties". *Handbook of LHC Higgs Cross Sections. 3. CERN Yellow Reports: Monographs*. Geneva: CERN, 2013. URL: <https://cds.cern.ch/record/1559921>.

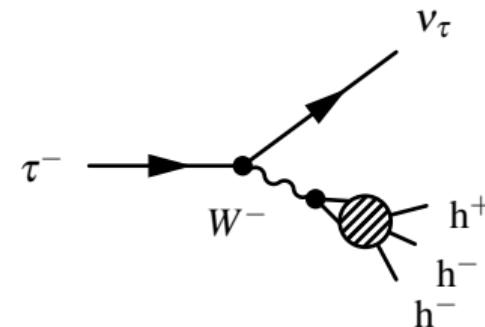
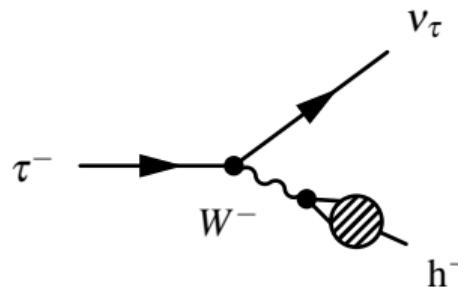
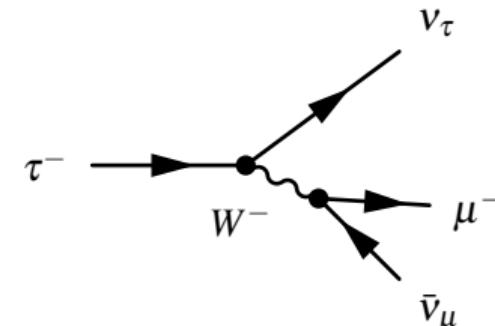
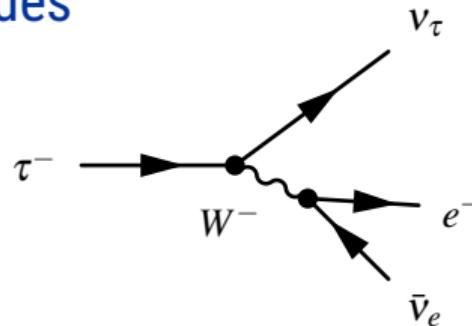
# $\tau$ decay modes

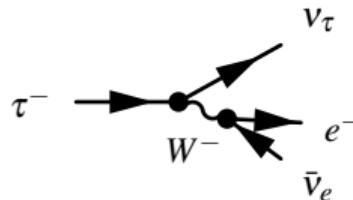
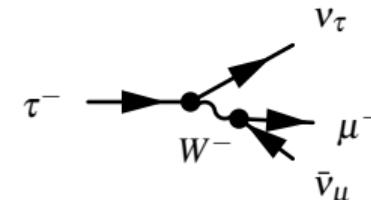
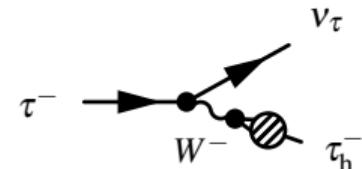


# $\tau$ decay modes



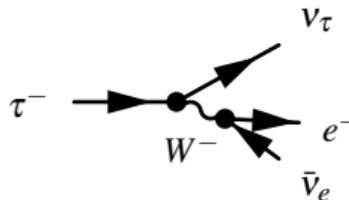
# $\tau$ decay modes



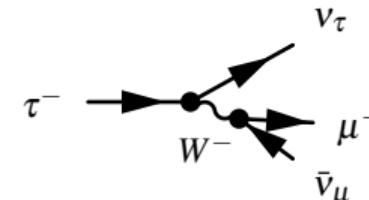
$H/A \rightarrow \tau\tau \rightarrow L_1 L_2$  $\tau \rightarrow e + v_e + \bar{v}_\tau \Rightarrow e$   
17.8 % $\tau \rightarrow \mu + v_\mu + \bar{v}_\tau \Rightarrow \mu$   
17.4 % $\tau \rightarrow \text{hadrons} + v_\tau \Rightarrow \tau_h$   
64.8 %

$$H/A \rightarrow \tau\tau \rightarrow L_1 L_2$$

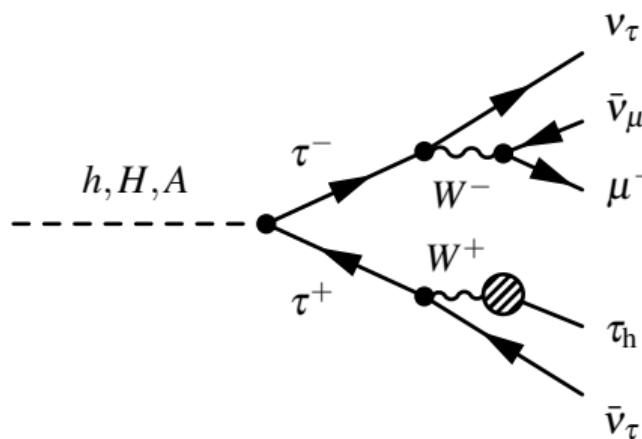
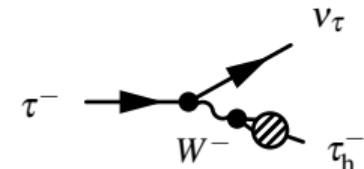
$$\tau \rightarrow e + v_e + \bar{v}_\tau \Rightarrow e \\ 17.8\%$$



$$\tau \rightarrow \mu + v_\mu + \bar{v}_\tau \Rightarrow \mu \\ 17.4\%$$



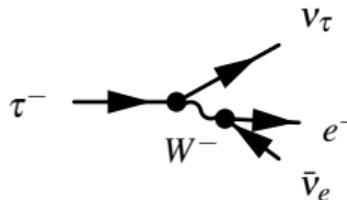
$$\tau \rightarrow \text{hadrons} + v_\tau \Rightarrow \tau_h \\ 64.8\%$$



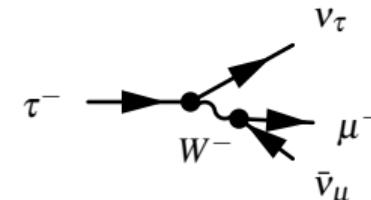
▷ Particle Data Group. "Review of Particle Physics". *Progress of Theoretical and Experimental Physics* 8 (Aug. 2020). DOI: 10.1093/ptep/ptaa104.

$$H/A \rightarrow \tau\tau \rightarrow L_1 L_2$$

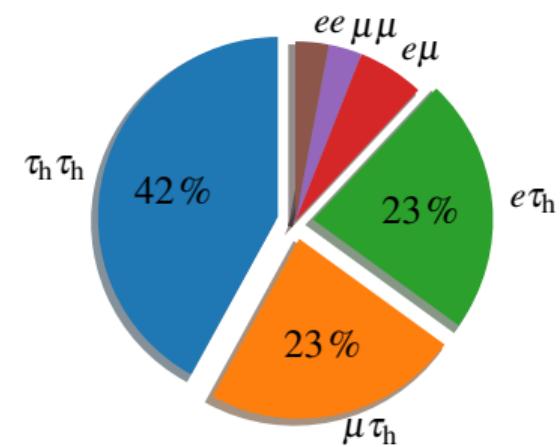
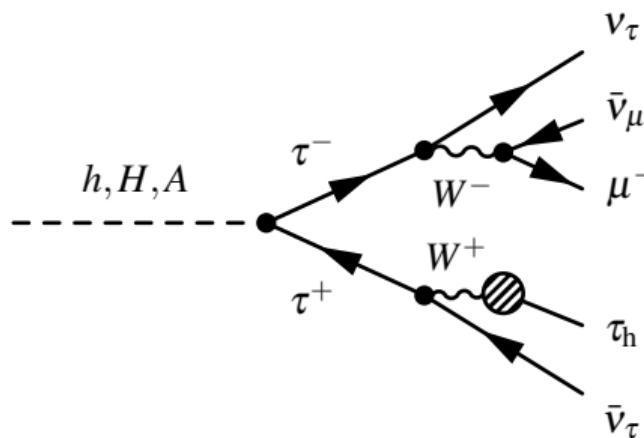
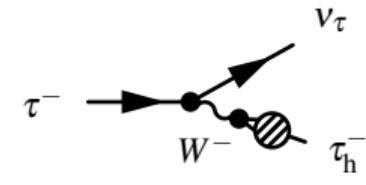
$$\tau \rightarrow e + v_e + \bar{v}_\tau \Rightarrow e \\ 17.8\%$$



$$\tau \rightarrow \mu + v_\mu + \bar{v}_\tau \Rightarrow \mu \\ 17.4\%$$



$$\tau \rightarrow \text{hadrons} + v_\tau \Rightarrow \tau_h \\ 64.8\%$$



▷ Particle Data Group. "Review of Particle Physics". *Progress of Theoretical and Experimental Physics* 8 (Aug. 2020). DOI: 10.1093/ptep/ptaa104.

## 1 Phenomenology

## 2 Experimental device

## 3 $H/A \rightarrow \tau\tau$ analysis

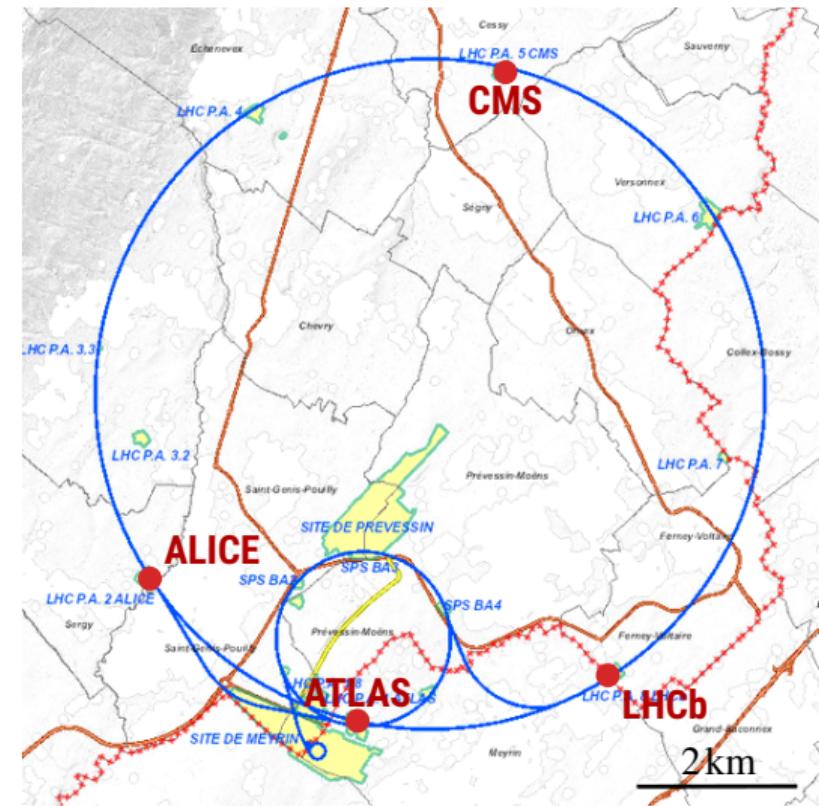
## 4 Machine learning

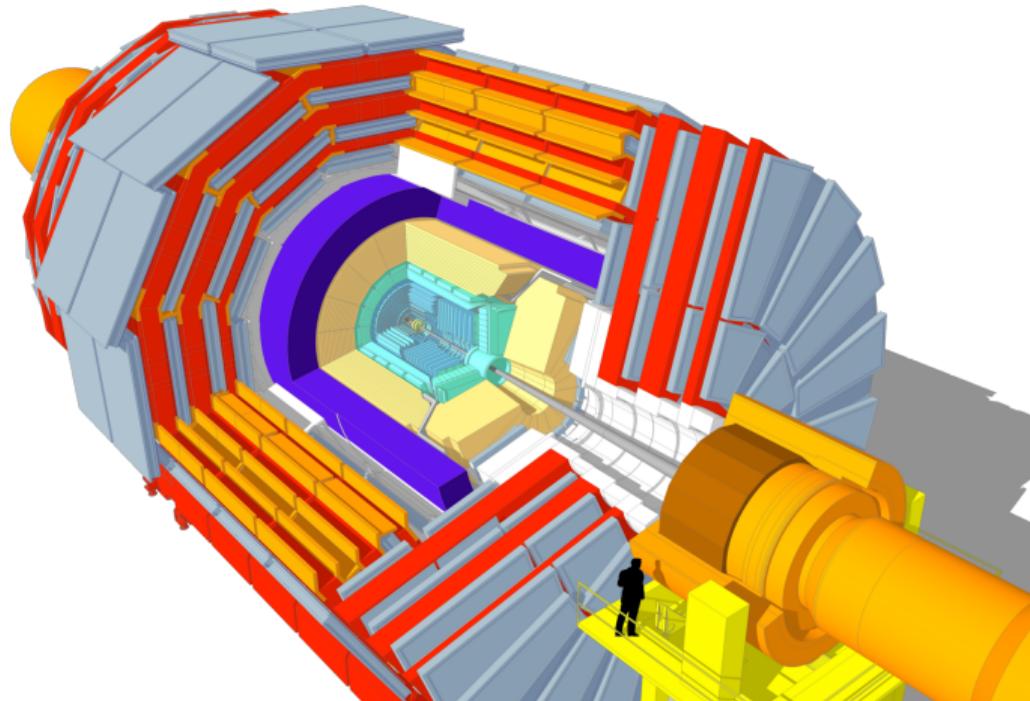
# Principle

$$E = mc^2$$

mass (new particles) from the collision energy

# CERN LHC

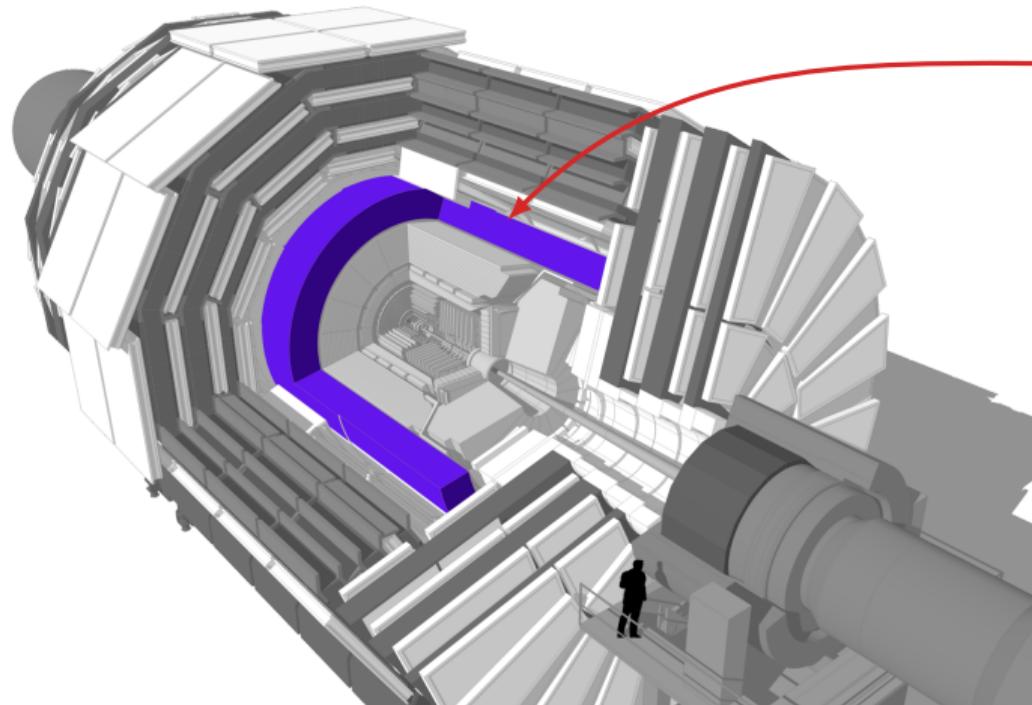




## CMS detector

- Mass:  $\sim 14000\text{t}$
- Diameter: 15 m
- Length: 28.7 m

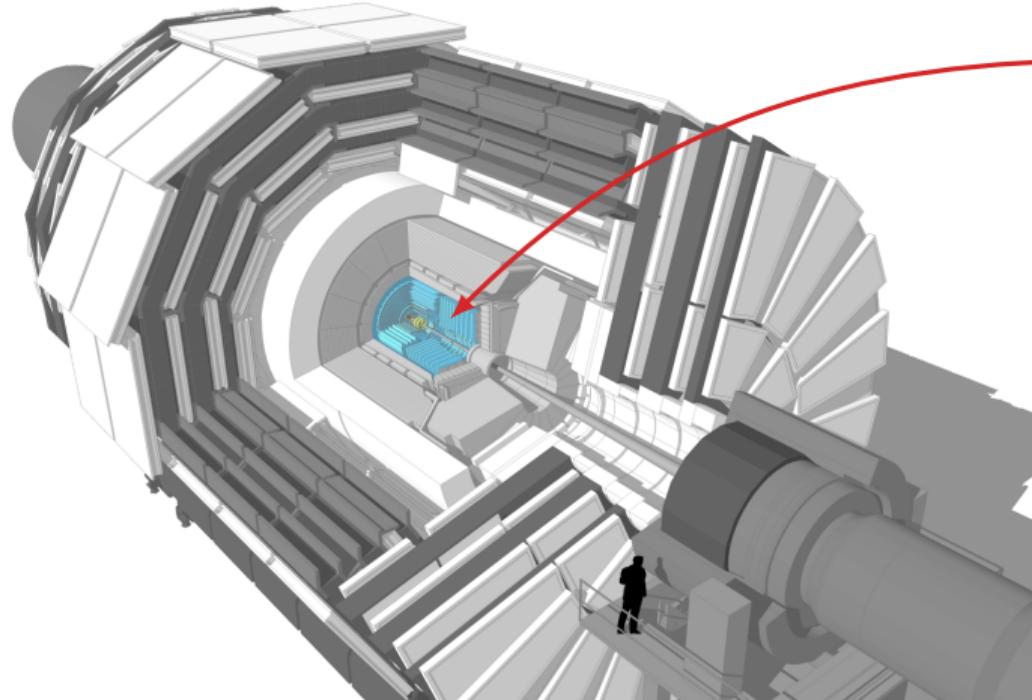
⇒ How to see the particles?



## Solenoid

- Niobium titanium coil
- Superconducting
- $\sim 18\,000\text{ A}$
- 4 T in the inner volume

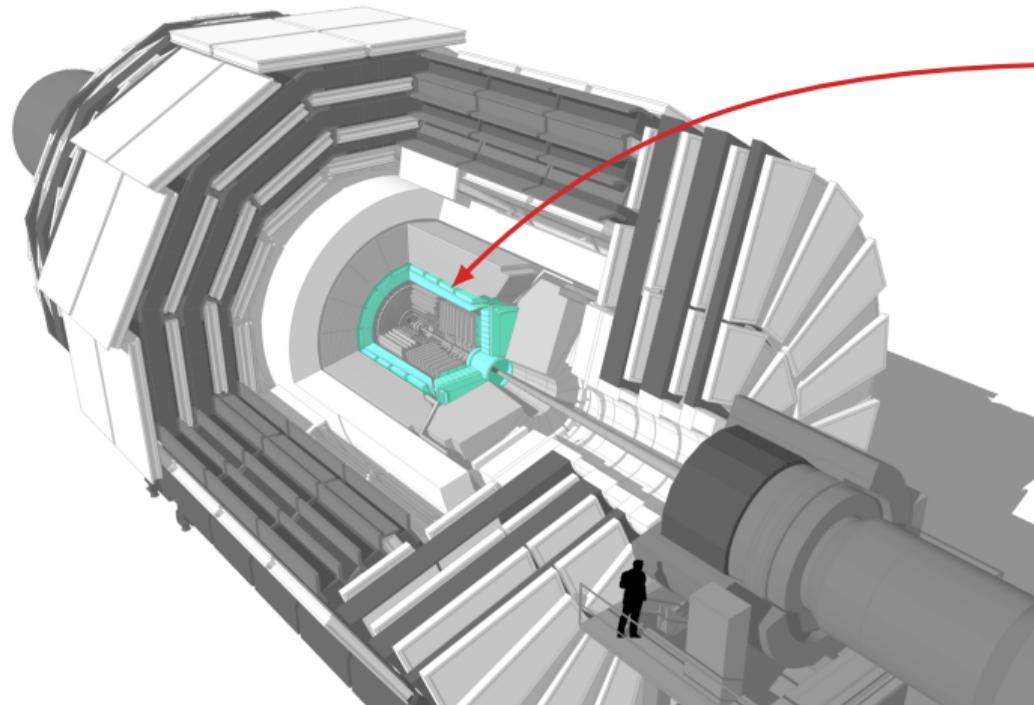
⇒ Bends charged particles trajectories in the transverse plane



## Tracker

- Made of Silicon
- Inner: pixels ( $100 \times 150 \mu\text{m}^2$ ,  
 $\sim 1.9 \text{ m}^2$ ,  $\sim 124 \text{ M}$  channels)
- Outer: microstrips ( $80 - 180 \mu\text{m}$ )  
 $\sim 200 \text{ m}^2 \sim 9.6 \text{ M}$  channels

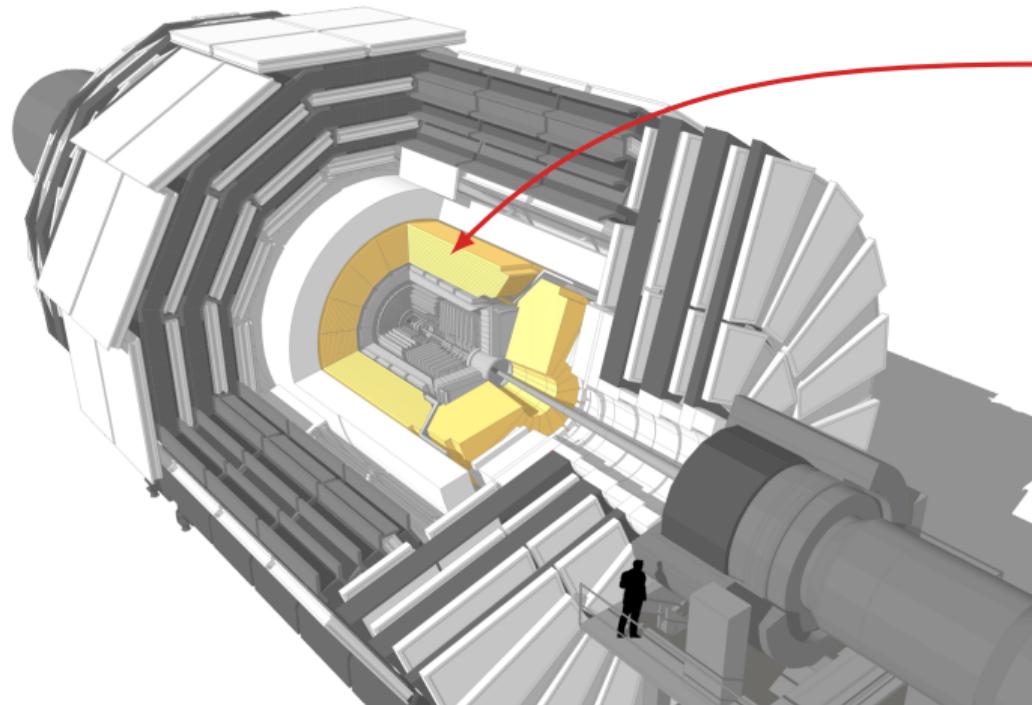
⇒ Charged particles leave hits when going through



## Electromagnetic CALorimeter

- $\sim 76\,000$  scintillating PbWO<sub>4</sub> crystals

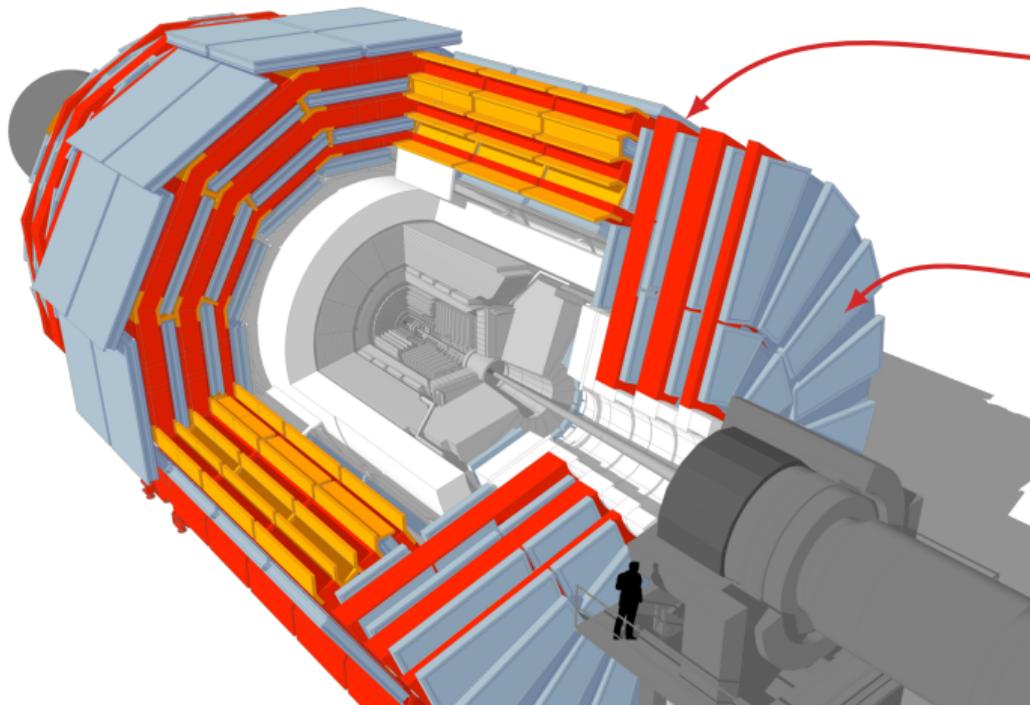
⇒ Electrons and photons are stopped,  
energy deposits



## Hadronic CALorimeter

- Brass + plastic scintillator,  
~ 7000 channels

⇒ Hadrons are stopped, energy deposits



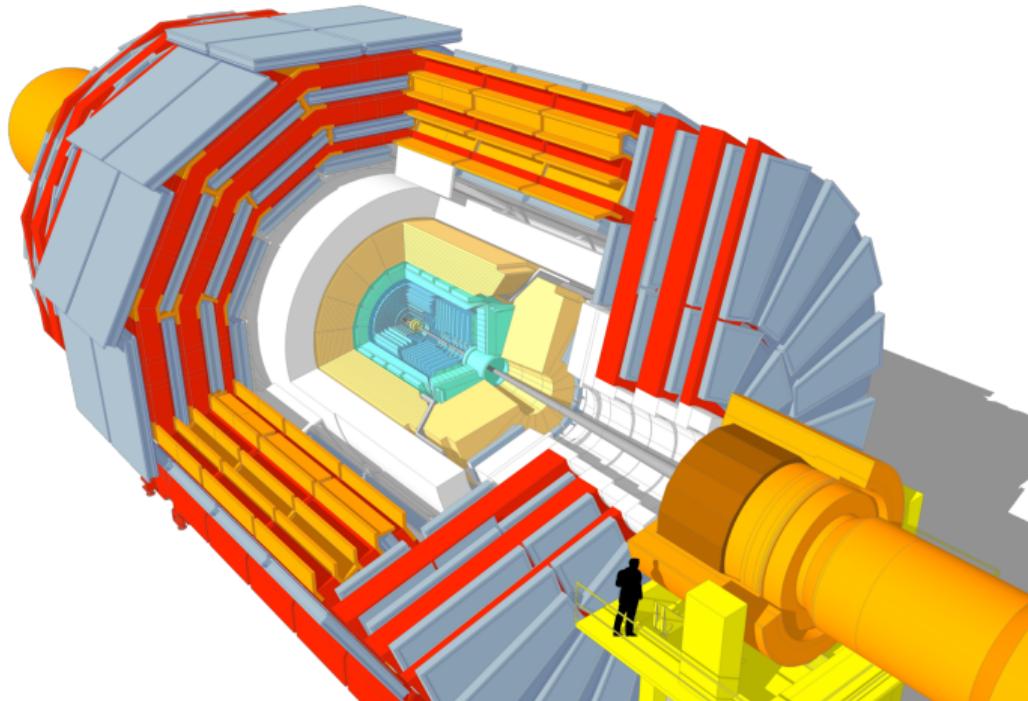
### Steel return yoke (red)

- Allows for 2 T magnetic field around the solenoid

### Muon chambers (blue-gray)

- Barrel: 250 drift tubes, 480 resistive plate chambers
- Endcaps: 540 cathode strips, 576 resistive plate chambers

⇒ Charged particles leave hits when going through (only muons do)

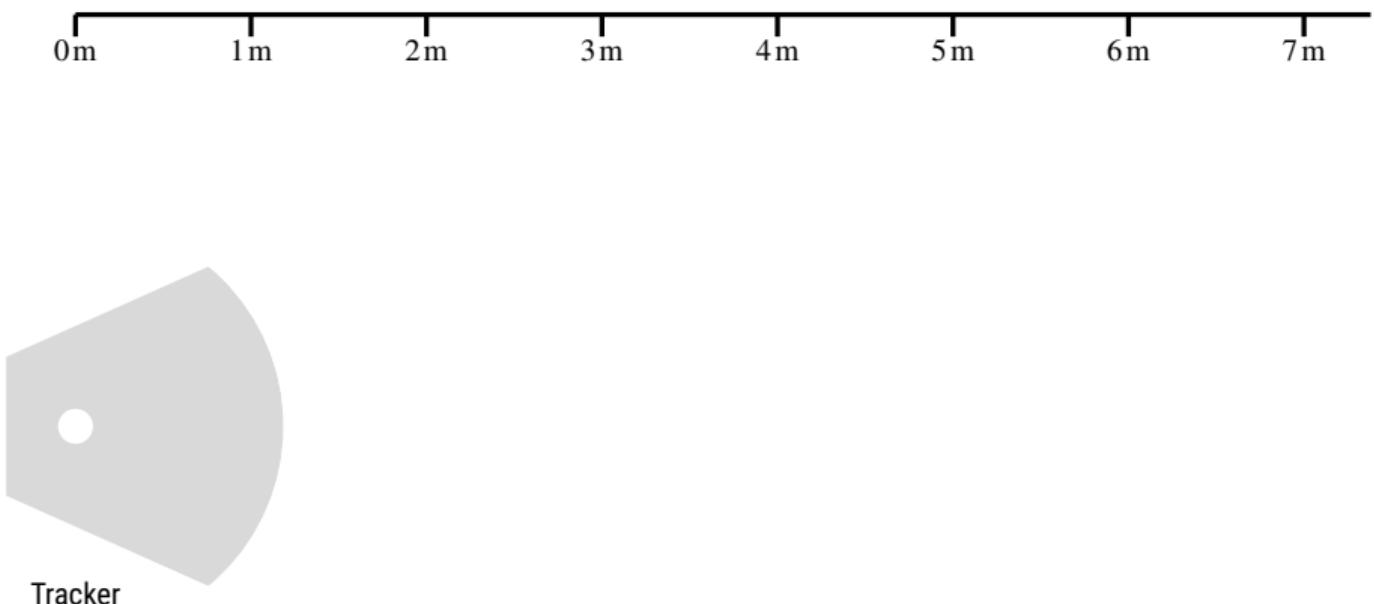


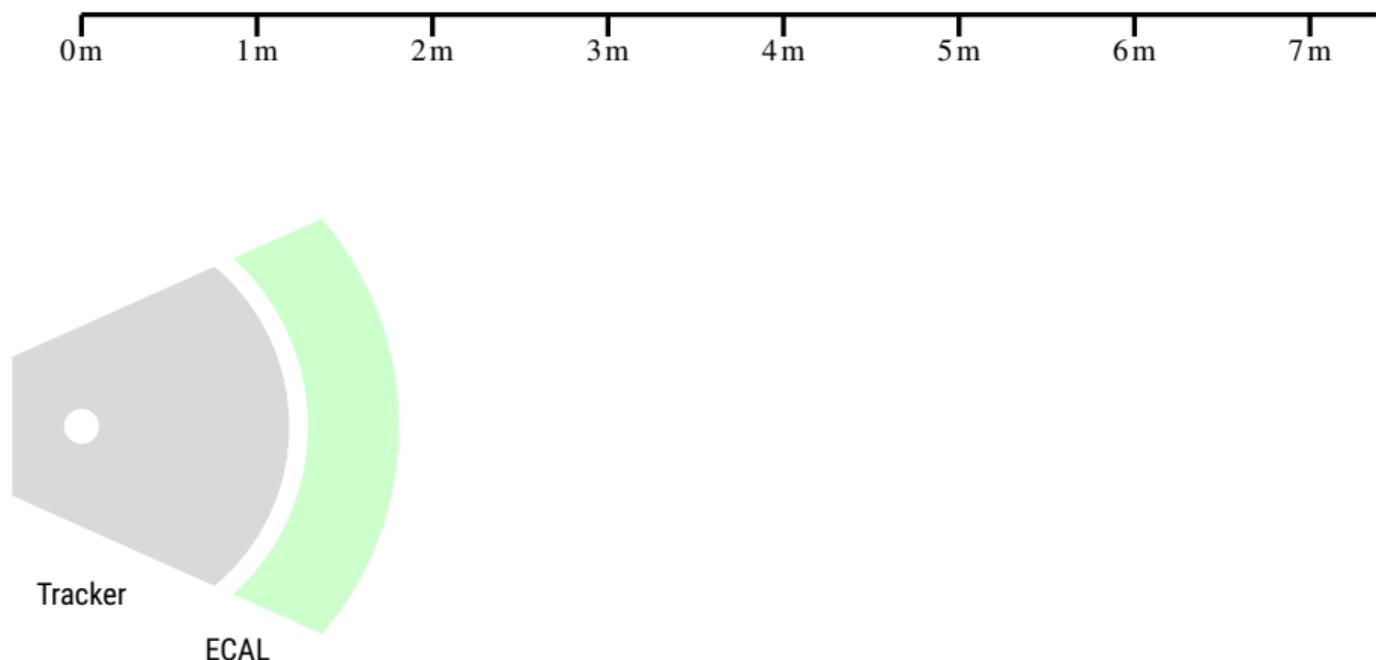
## Sensitive parts of CMS

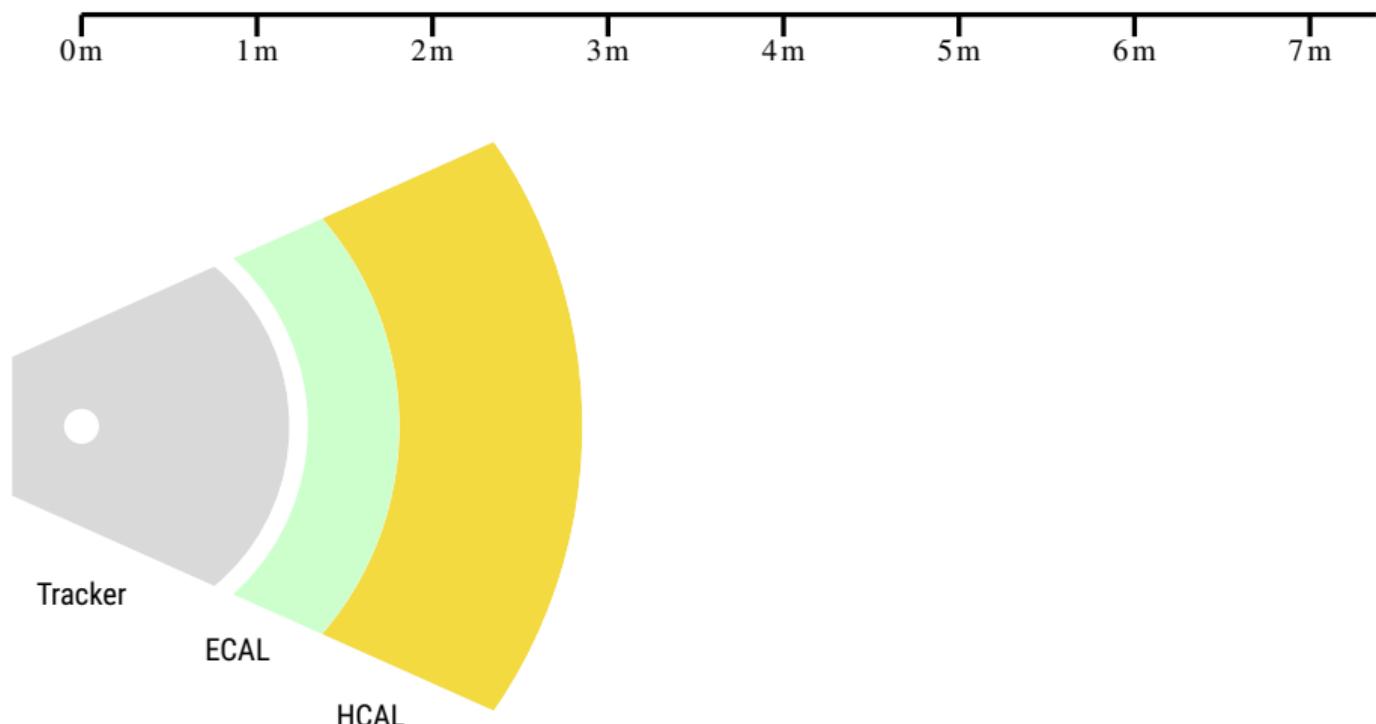
Combine sub-detectors signals to determine which particles were there!

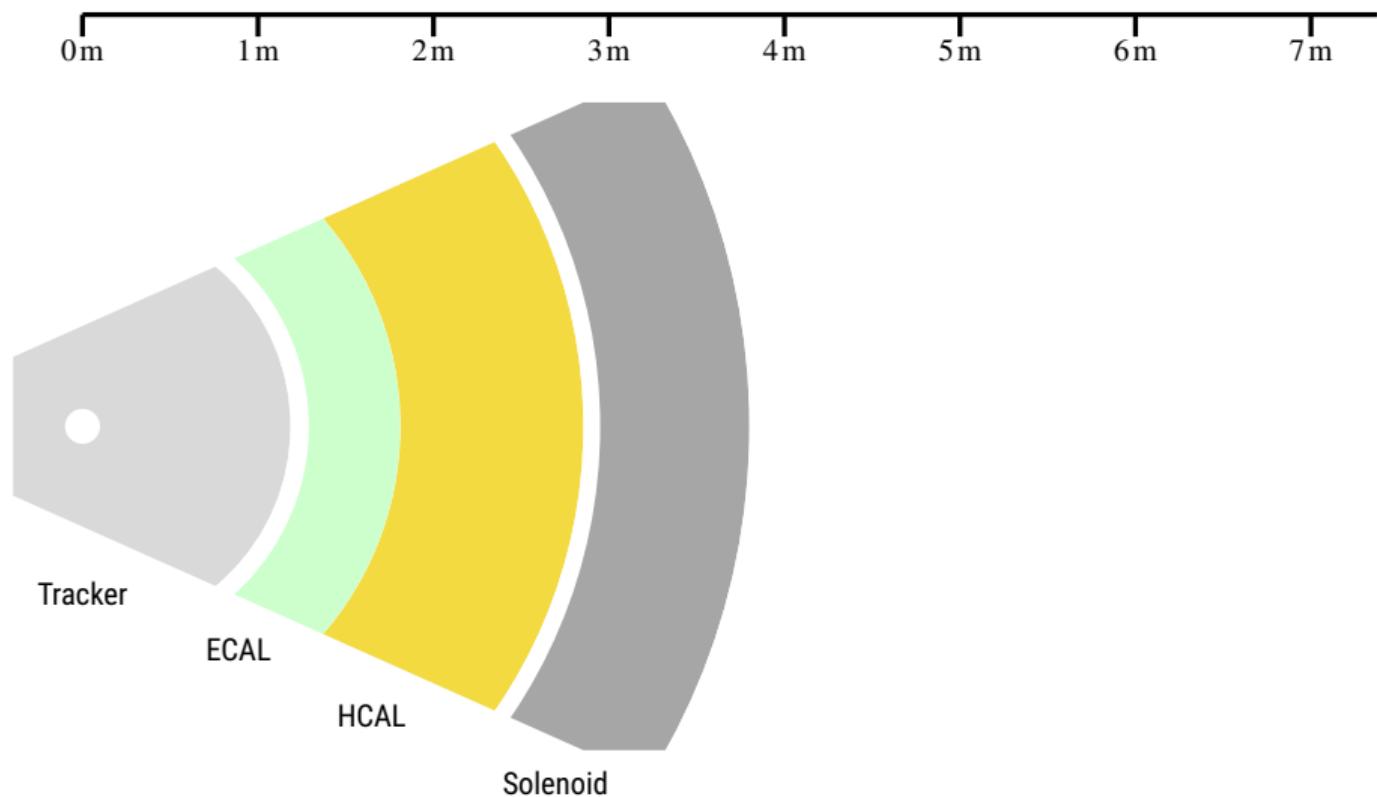


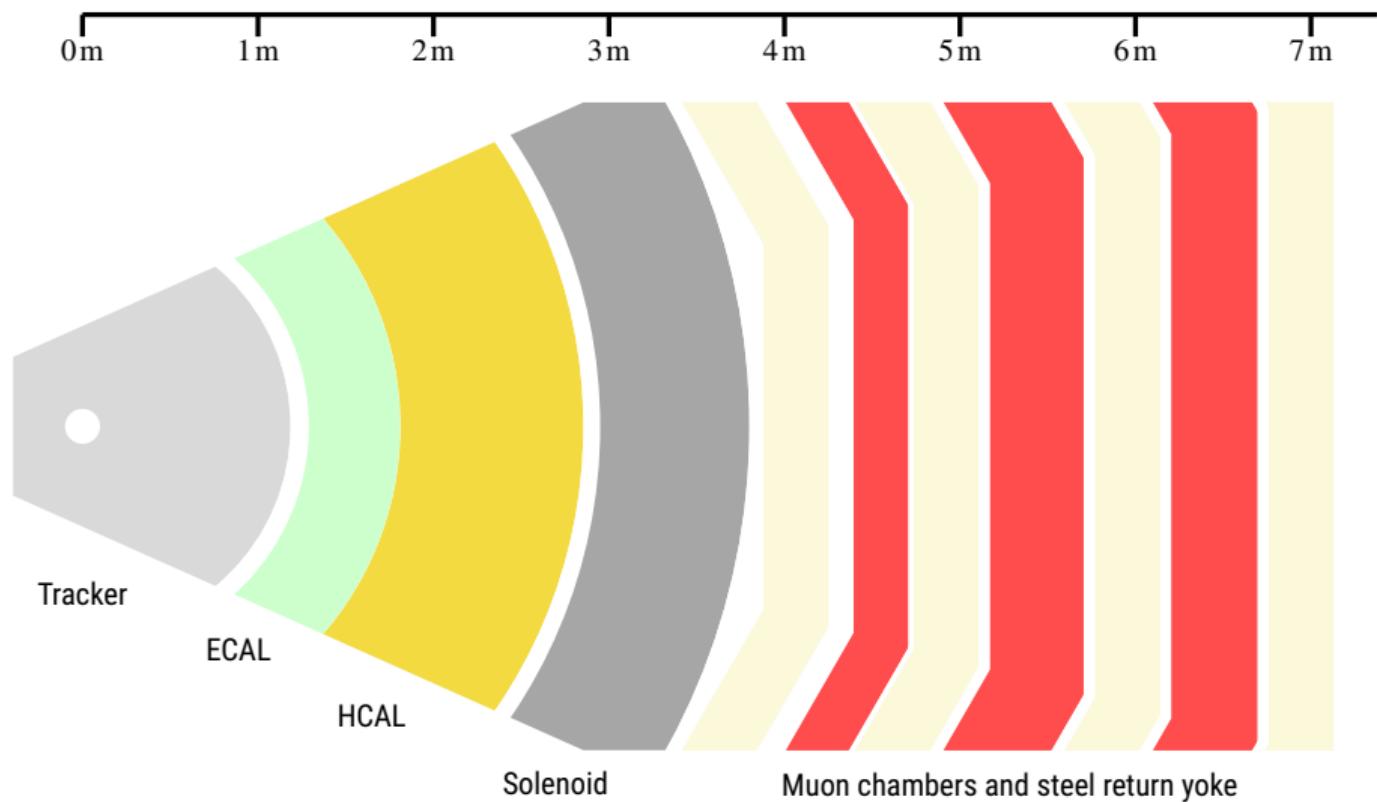


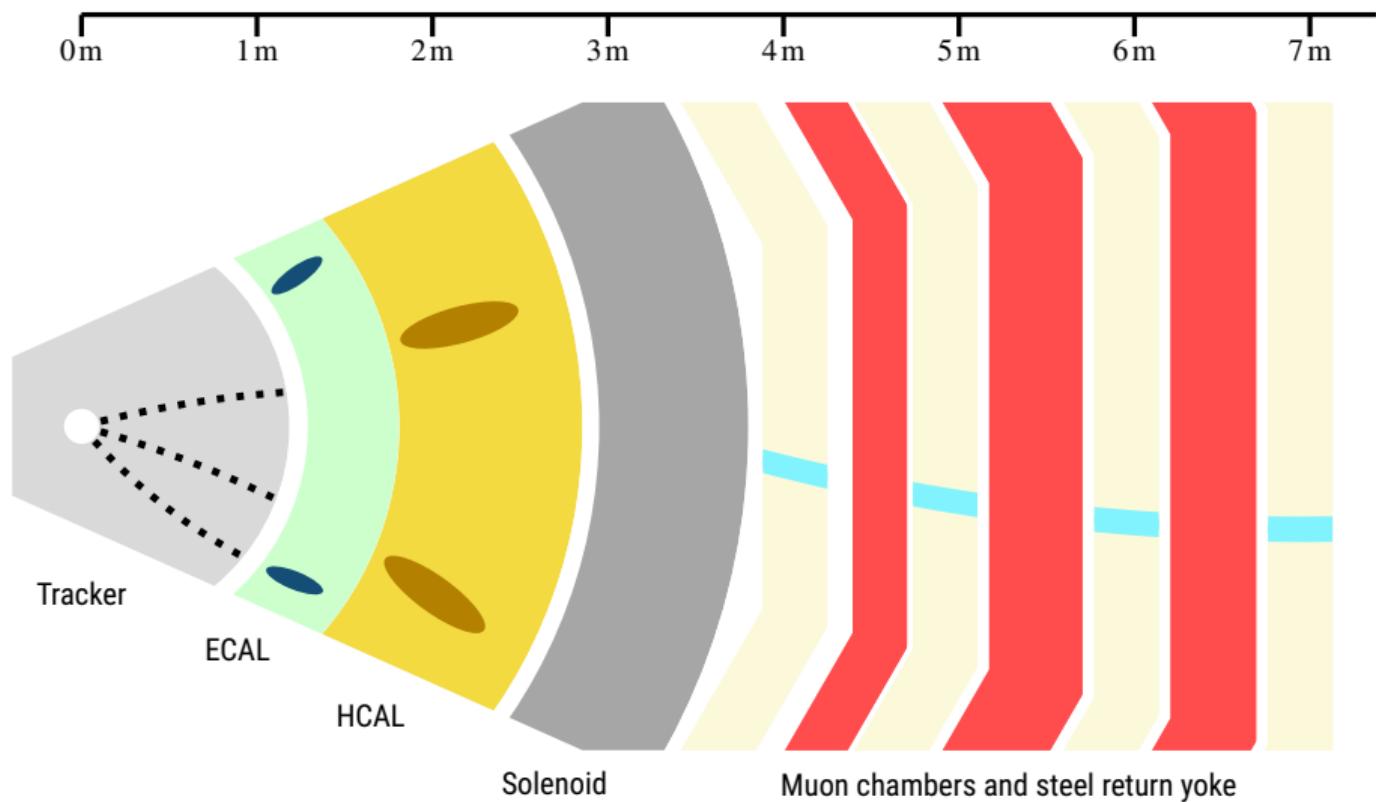


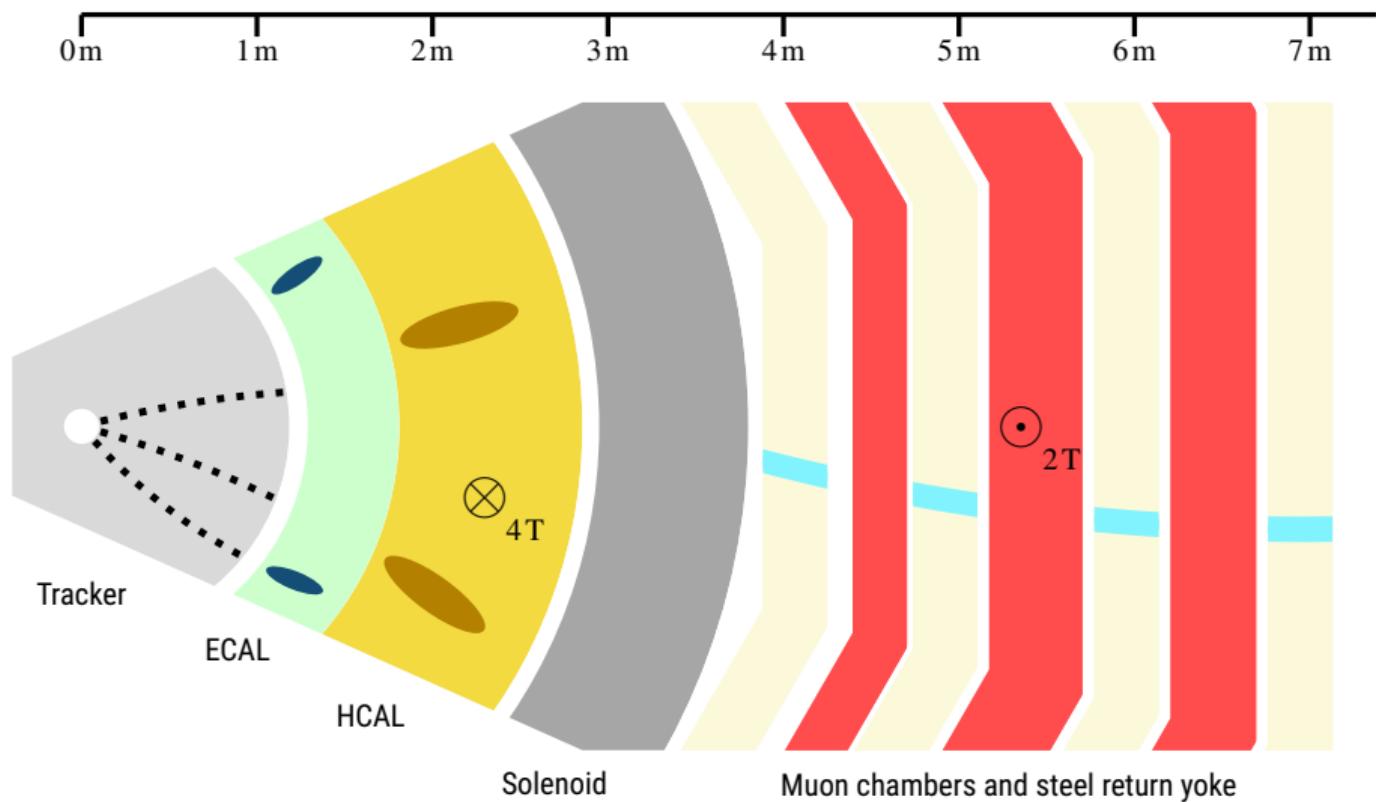


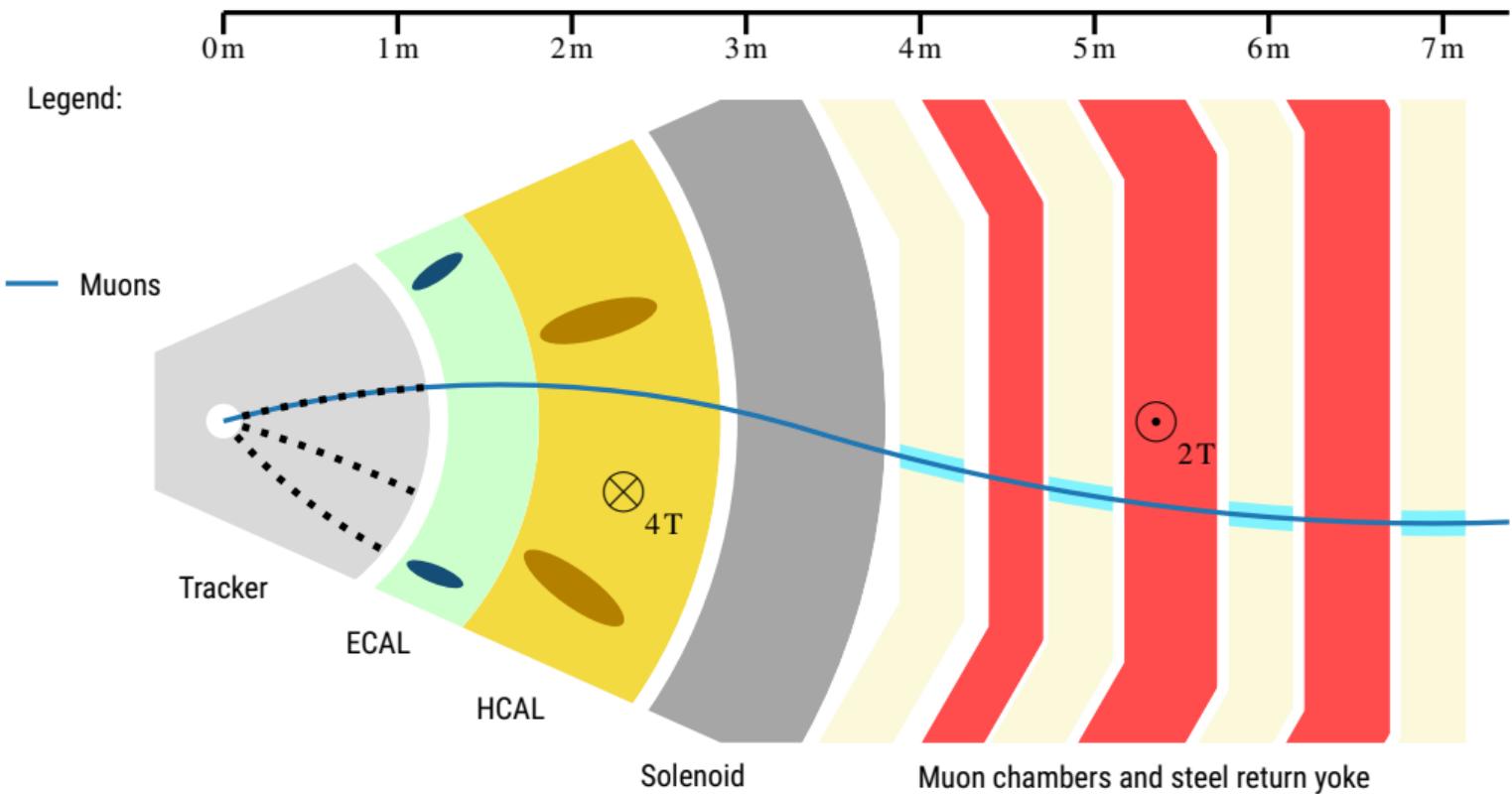






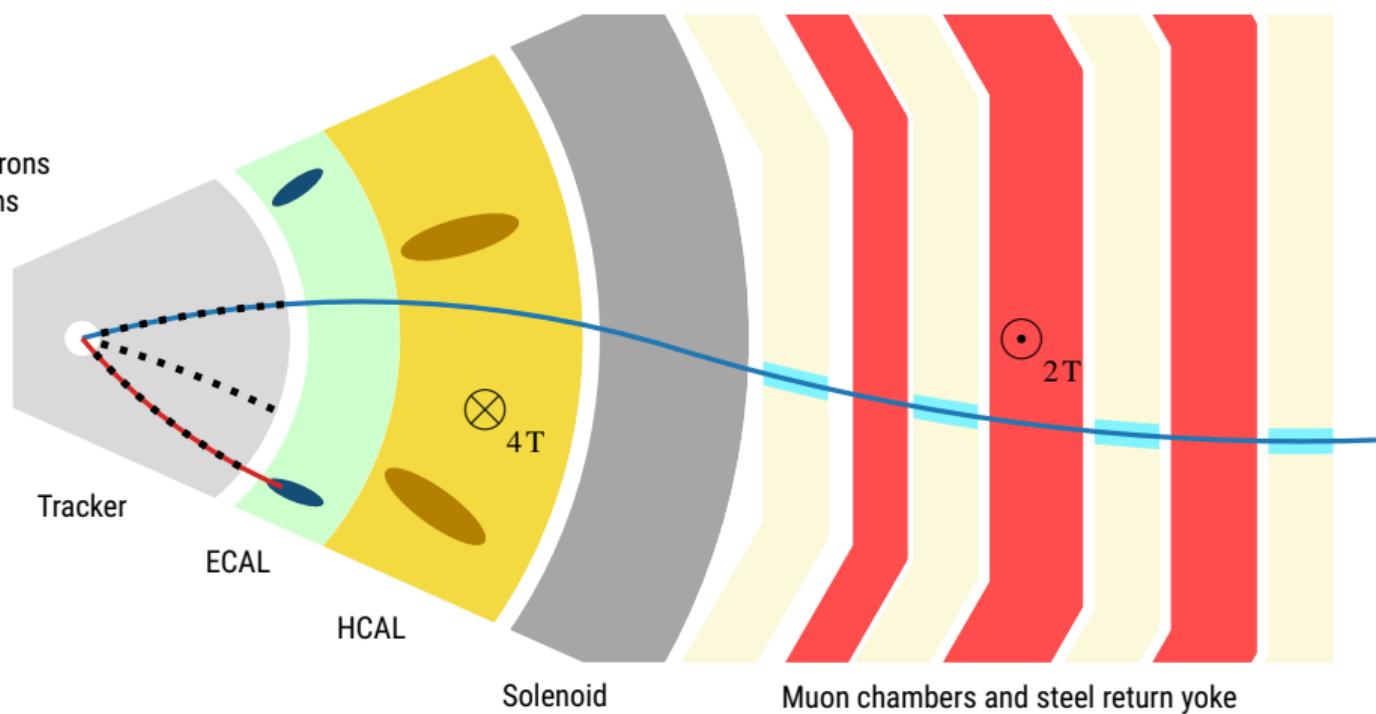


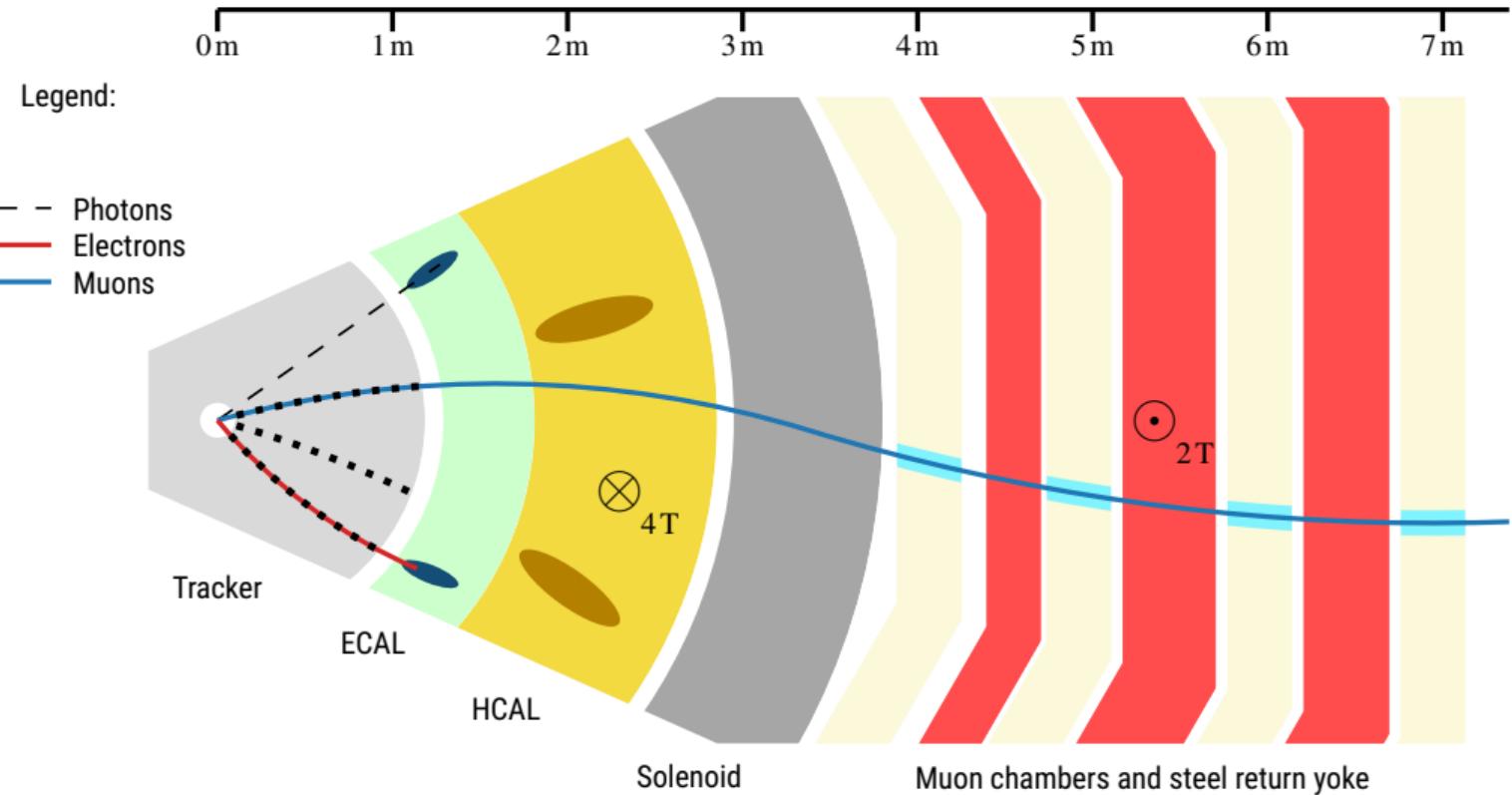


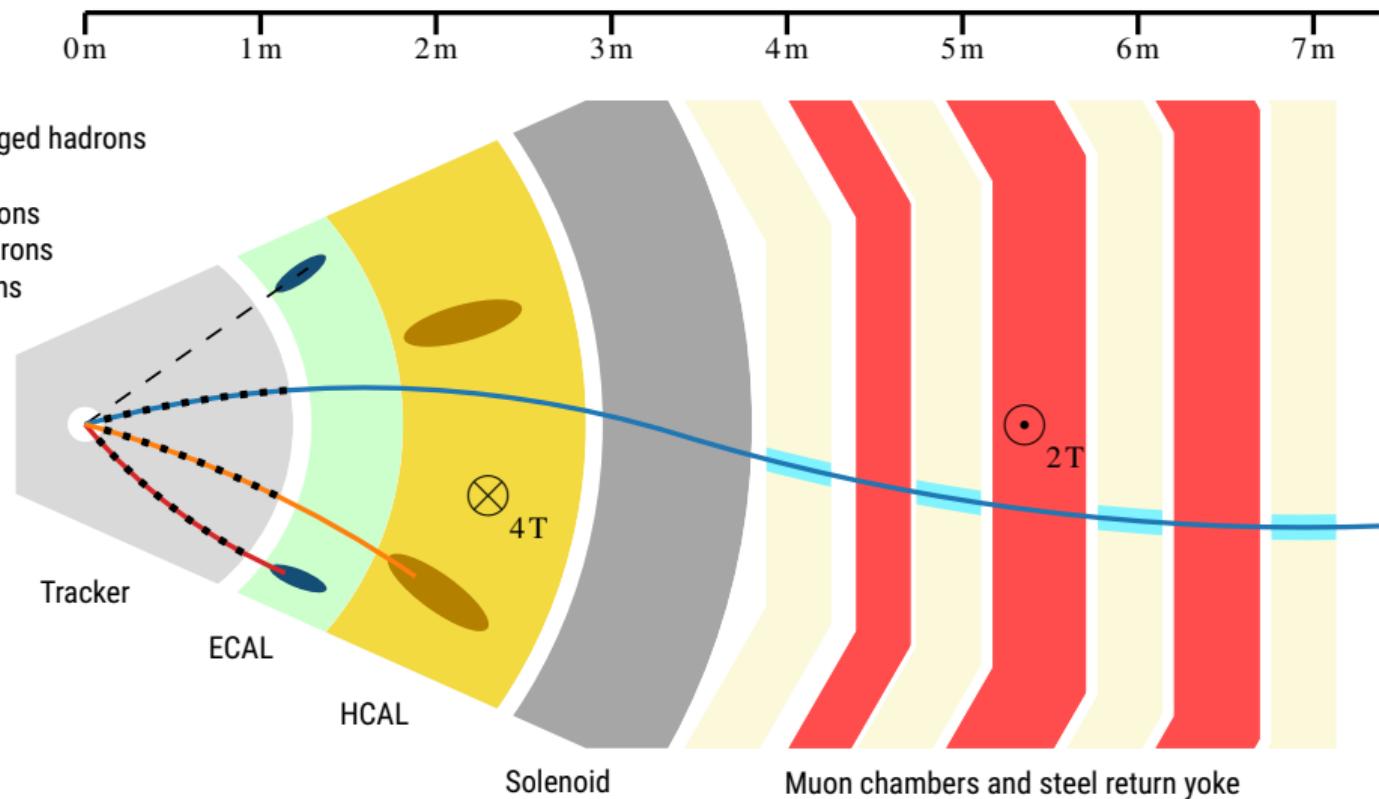


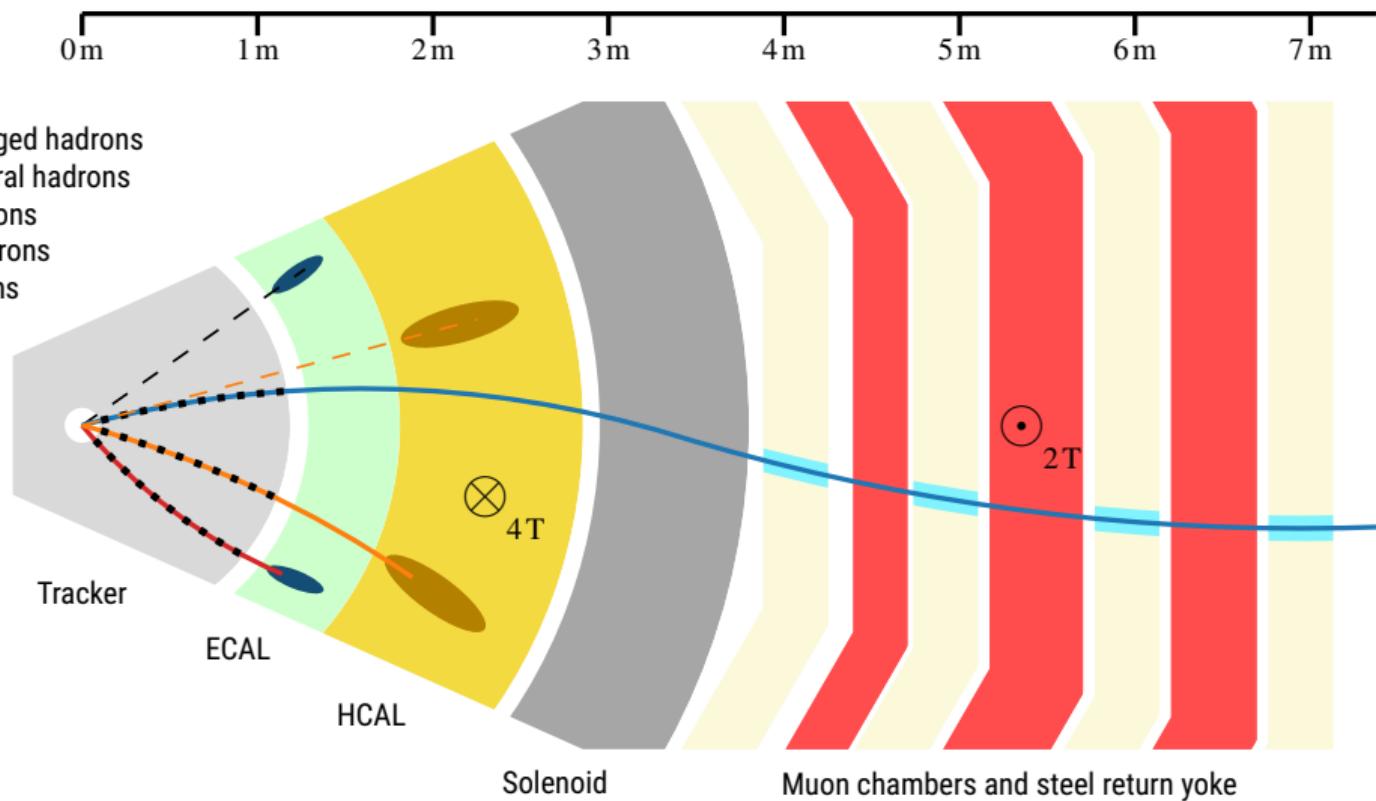
**Legend:**

- Electrons
- Muons



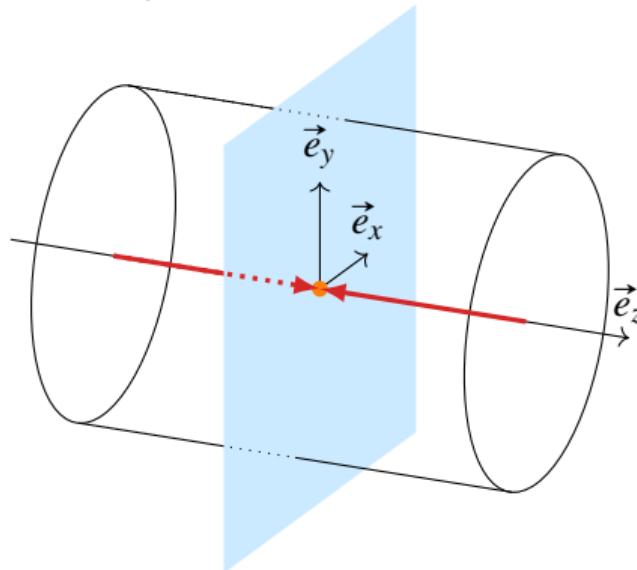




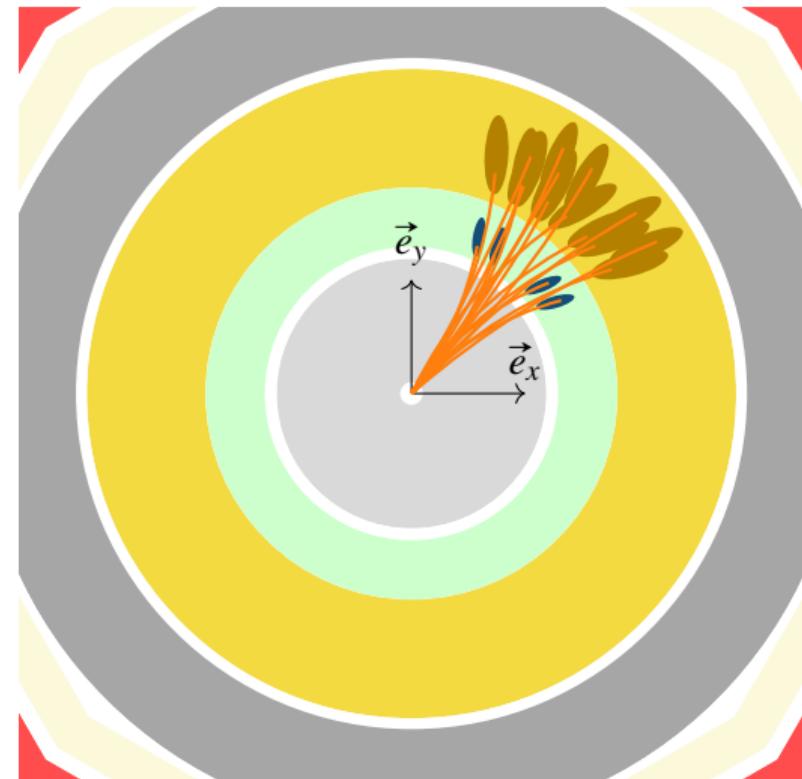


# Conserved momentum and neutrinos: missing transverse energy (MET)

$(\vec{e}_x, \vec{e}_y)$  = transverse plane ( $\eta = 0$ )

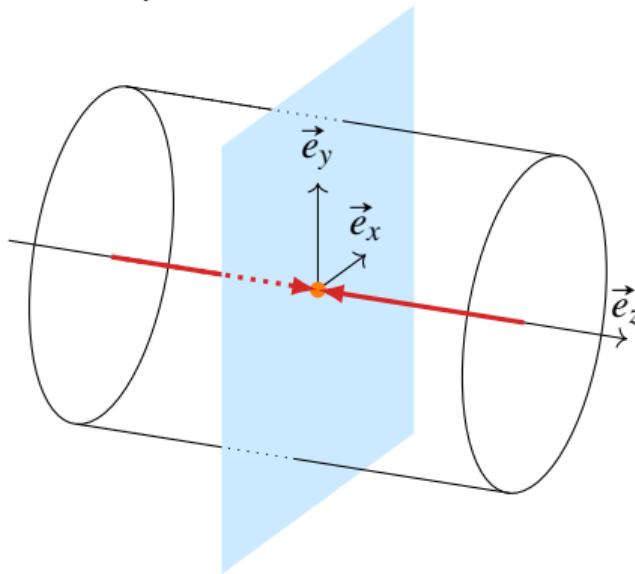


$$\sum_{\text{final state}} \vec{p}_{\text{T}} = \sum_{\text{initial state}} \vec{p}_{\text{T}} = \vec{0}$$

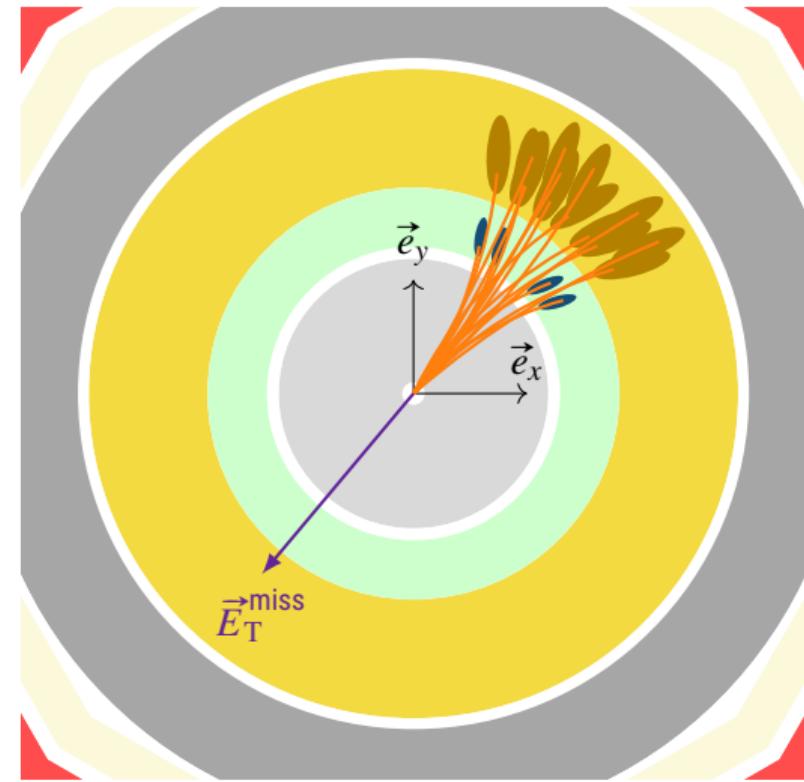


# Conserved momentum and neutrinos: missing transverse energy (MET)

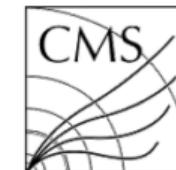
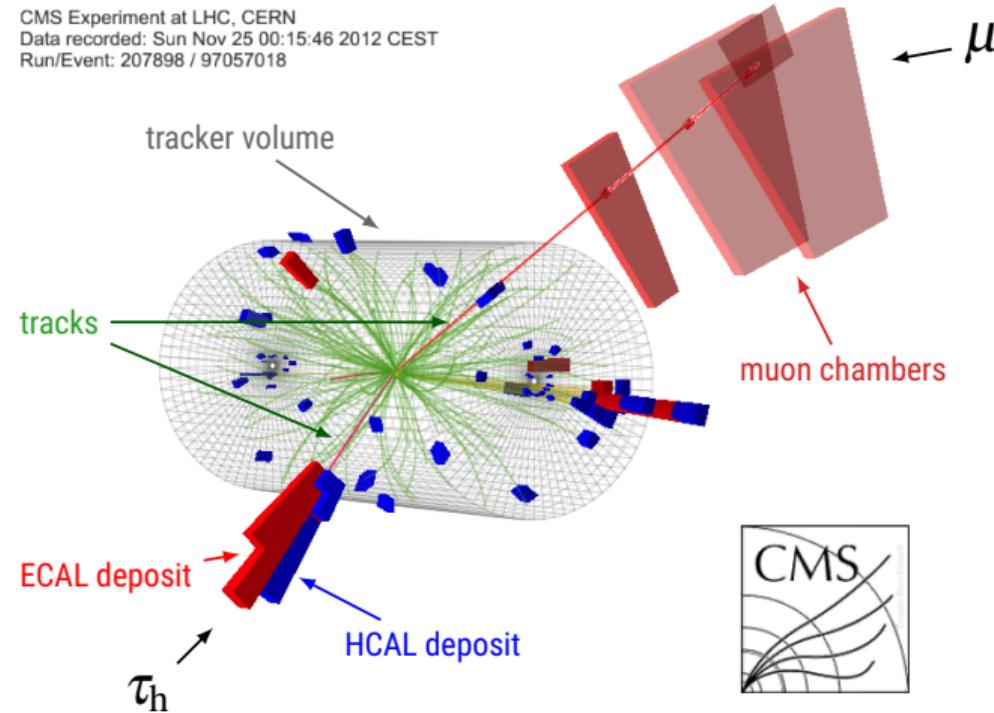
$(\vec{e}_x, \vec{e}_y)$  = transverse plane ( $\eta = 0$ )



$$\sum_{\text{final state}} \vec{p}_{\text{T}} = \sum_{\text{initial state}} \vec{p}_{\text{T}} = \vec{0} \Rightarrow \vec{E}_{\text{T}}^{\text{miss}} = - \sum_{\text{visible particles}} \vec{p}_{\text{T}}$$



# Event display: $h \rightarrow \tau\tau \rightarrow \mu\tau_h$ candidate from real data



## 1 Phenomenology

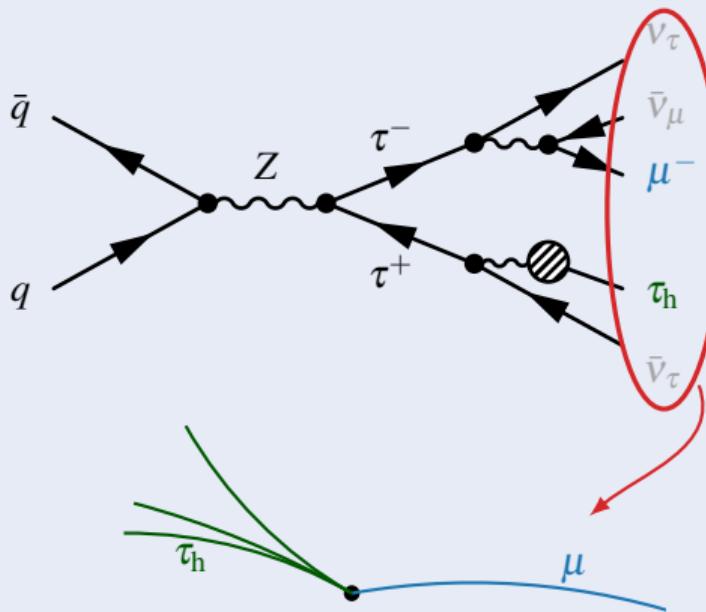
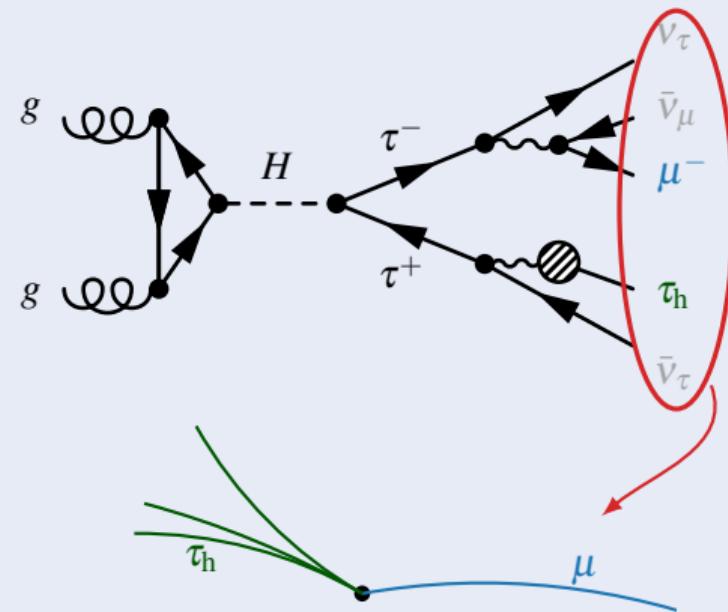
## 2 Experimental device

3  $H/A \rightarrow \tau\tau$  analysis

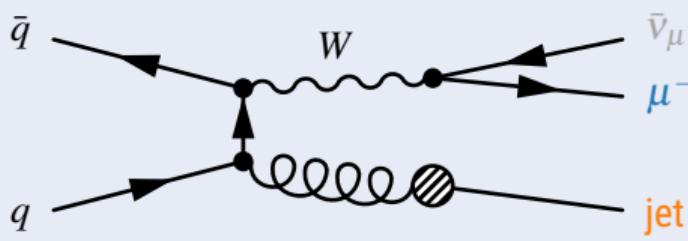
## 4 Machine learning

# Backgrounds?

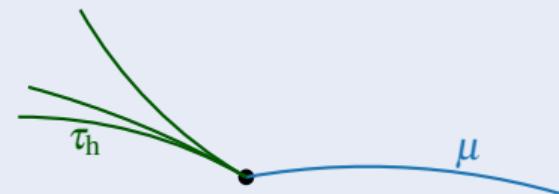
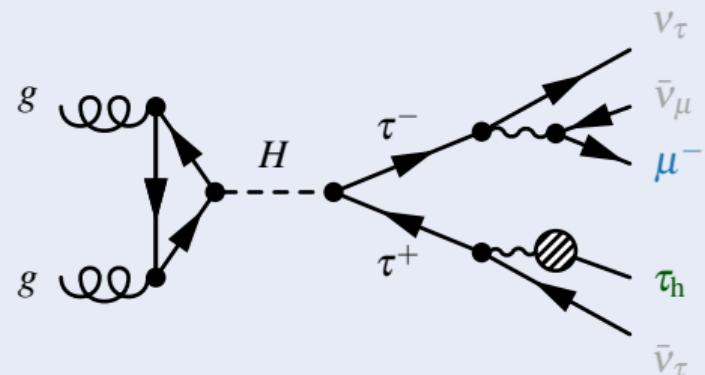
## Drell-Yan background

 $H \rightarrow \tau\tau \rightarrow \mu\tau_h$  signal

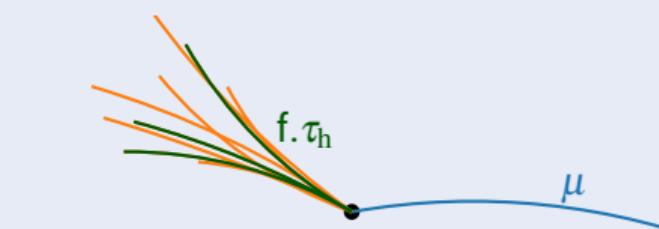
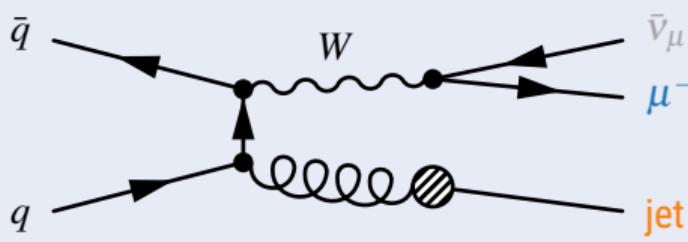
## $W + \text{jets}$ background



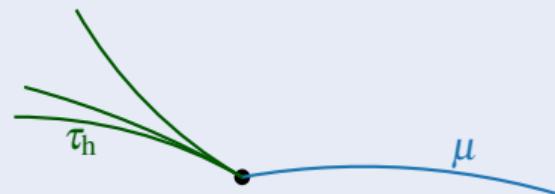
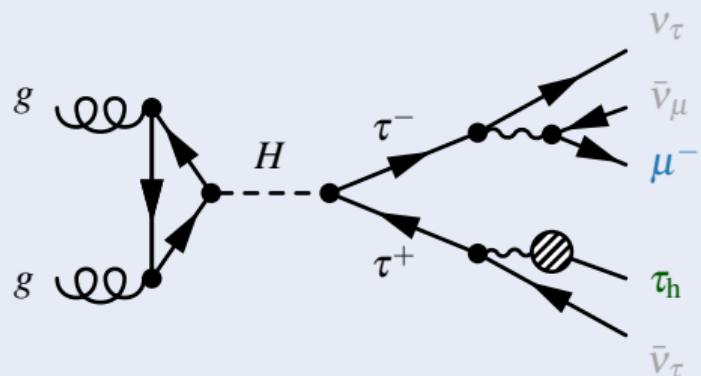
## $H \rightarrow \tau\tau \rightarrow \mu\tau_h$ signal



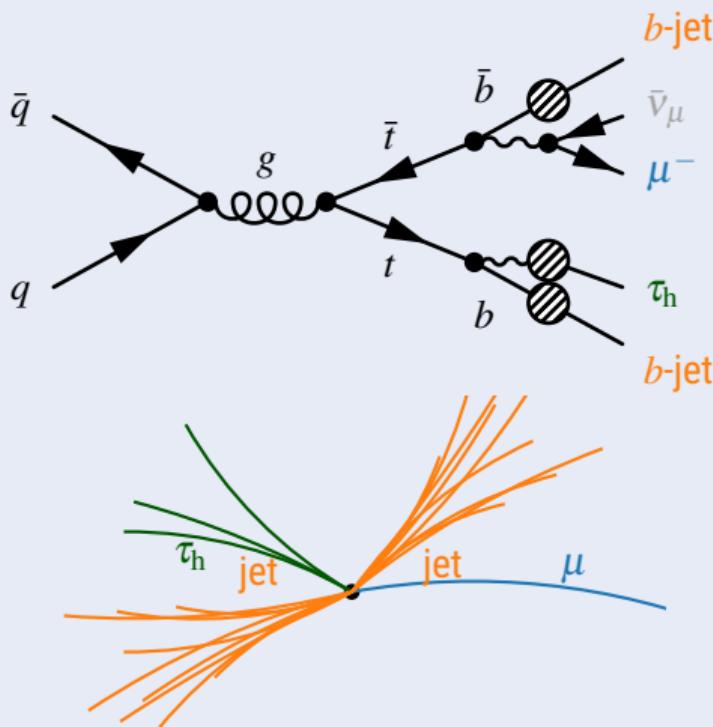
## $W + \text{jets}$ background, jet $\rightarrow$ fake $\tau_h$



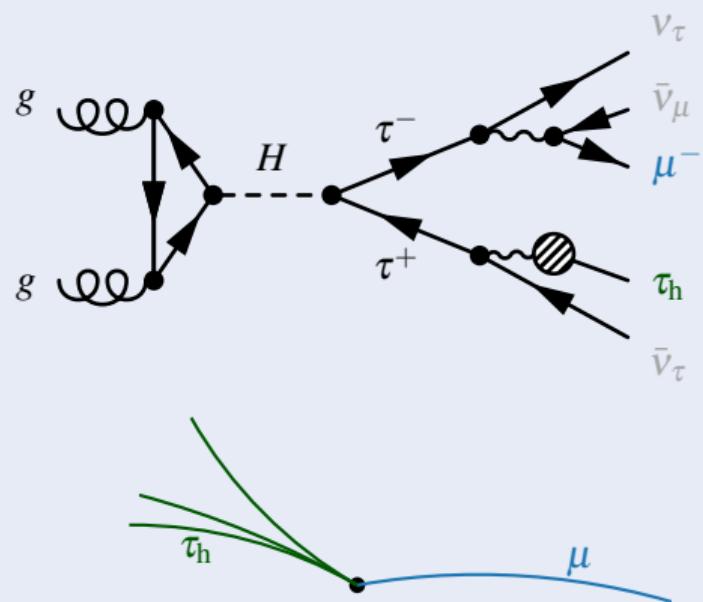
## $H \rightarrow \tau\tau \rightarrow \mu\tau_h$ signal



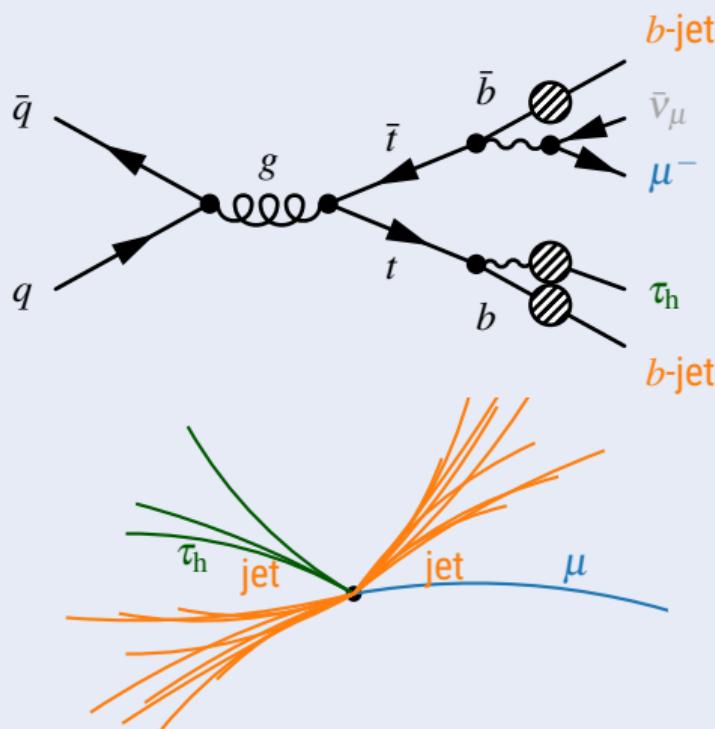
## $t\bar{t}$ background



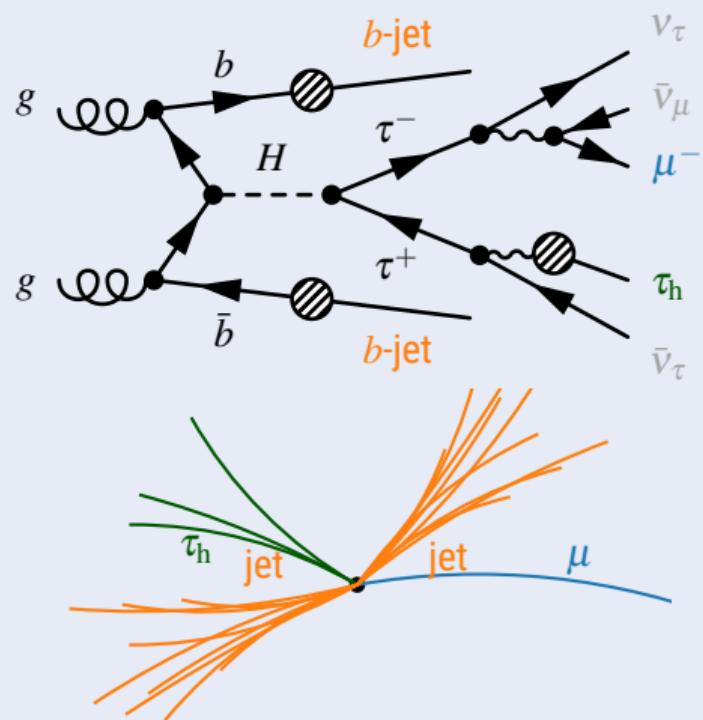
## $H \rightarrow \tau\tau \rightarrow \mu\tau_h$ signal



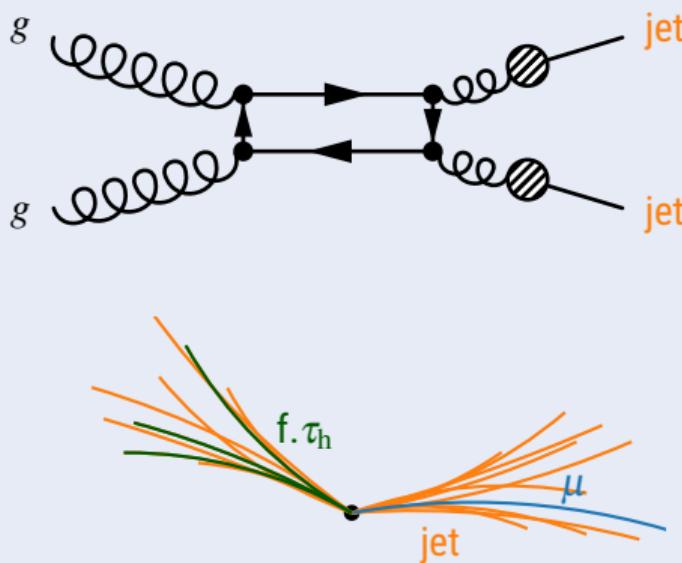
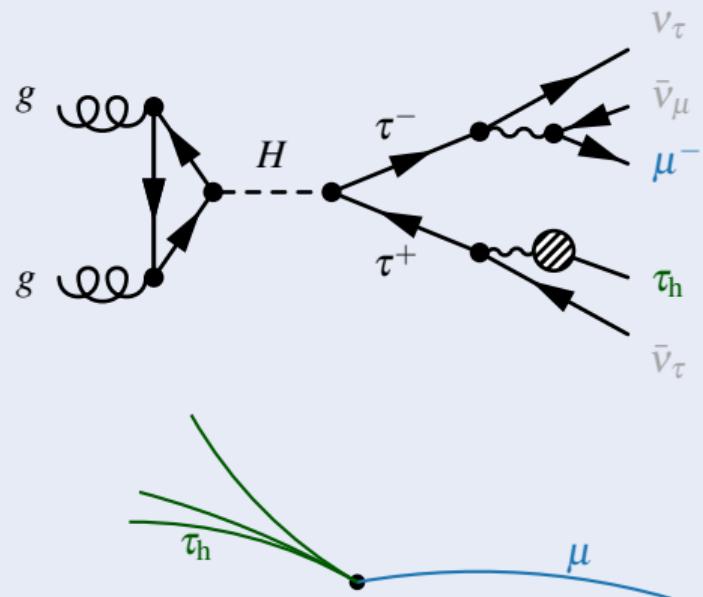
## $t\bar{t}$ background



## $H$ production with $b$ -jets



## QCD background

 $H \rightarrow \tau\tau \rightarrow \mu\tau_h$  signal

# Backgrounds modelling?

## Pure simulation (Monte-Carlo)

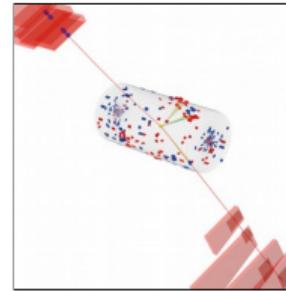
- Generate and simulate processes.
- Available for all except QCD.

## Data-driven techniques

- Better observation - background agreement.
- In some case, lower uncertainties.

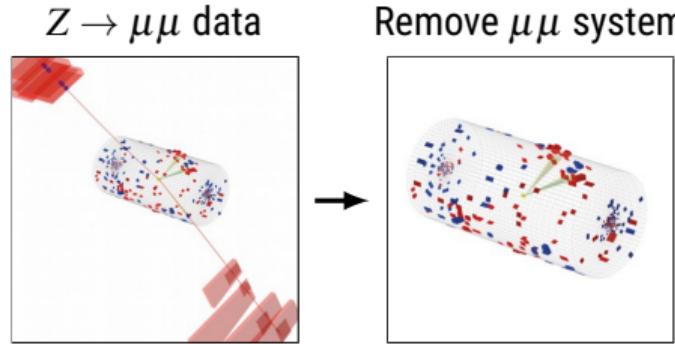
# Embedded events & genuine $\tau$ leptons

$Z \rightarrow \mu\mu$  data



- ▷ The CMS Collaboration. "An embedding technique to determine  $\tau\tau$  backgrounds in proton-proton collision data". *Journal of Instrumentation* **14**.06 (June 2019). DOI: [10.1088/1748-0221/14/06/p06032](https://doi.org/10.1088/1748-0221/14/06/p06032).

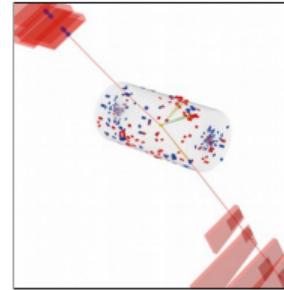
# Embedded events & genuine $\tau$ leptons



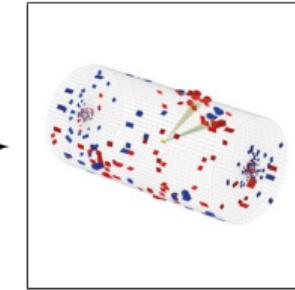
- ▷ The CMS Collaboration. "An embedding technique to determine  $\tau\tau$  backgrounds in proton-proton collision data". *Journal of Instrumentation* **14**.06 (June 2019). DOI: [10.1088/1748-0221/14/06/p06032](https://doi.org/10.1088/1748-0221/14/06/p06032).

# Embedded events & genuine $\tau$ leptons

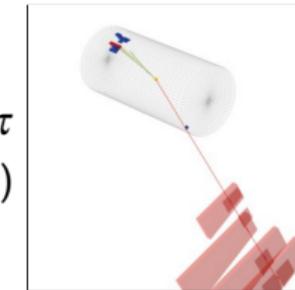
$Z \rightarrow \mu\mu$  data



Remove  $\mu\mu$  system



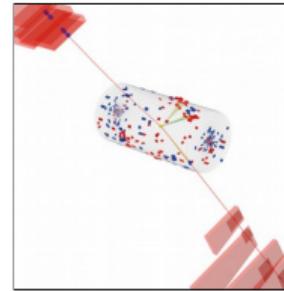
Simulate  $Z \rightarrow \tau\tau$   
 $(\vec{p}_\tau \simeq \vec{p}_\mu)$



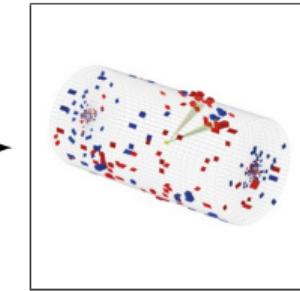
- ▷ The CMS Collaboration. "An embedding technique to determine  $\tau\tau$  backgrounds in proton-proton collision data". *Journal of Instrumentation* **14**.06 (June 2019). DOI: [10.1088/1748-0221/14/06/p06032](https://doi.org/10.1088/1748-0221/14/06/p06032).

# Embedded events & genuine $\tau$ leptons

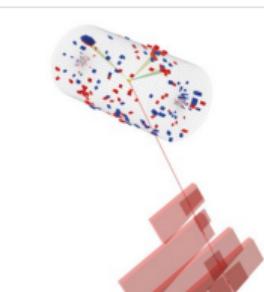
$Z \rightarrow \mu\mu$  data



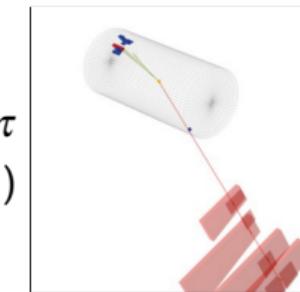
Remove  $\mu\mu$  system



$Z \rightarrow \tau\tau$  embedded event



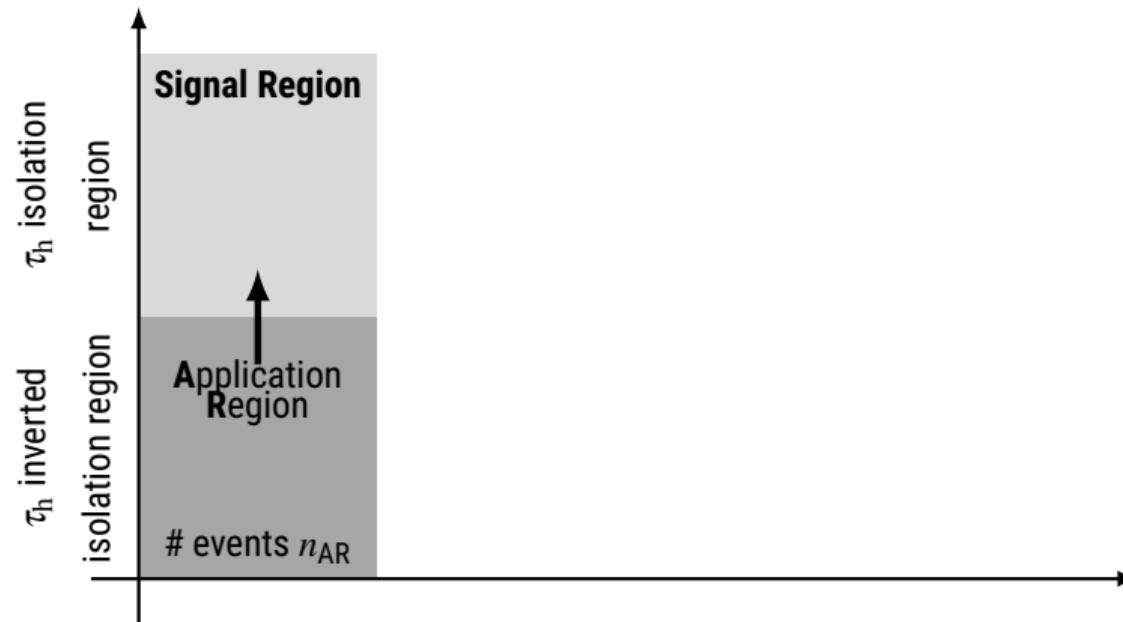
Simulate  $Z \rightarrow \tau\tau$   
 $(\vec{p}_\tau \simeq \vec{p}_\mu)$



- ▷ The CMS Collaboration. "An embedding technique to determine  $\tau\tau$  backgrounds in proton-proton collision data". *Journal of Instrumentation* 14.06 (June 2019). DOI: 10.1088/1748-0221/14/06/p06032.

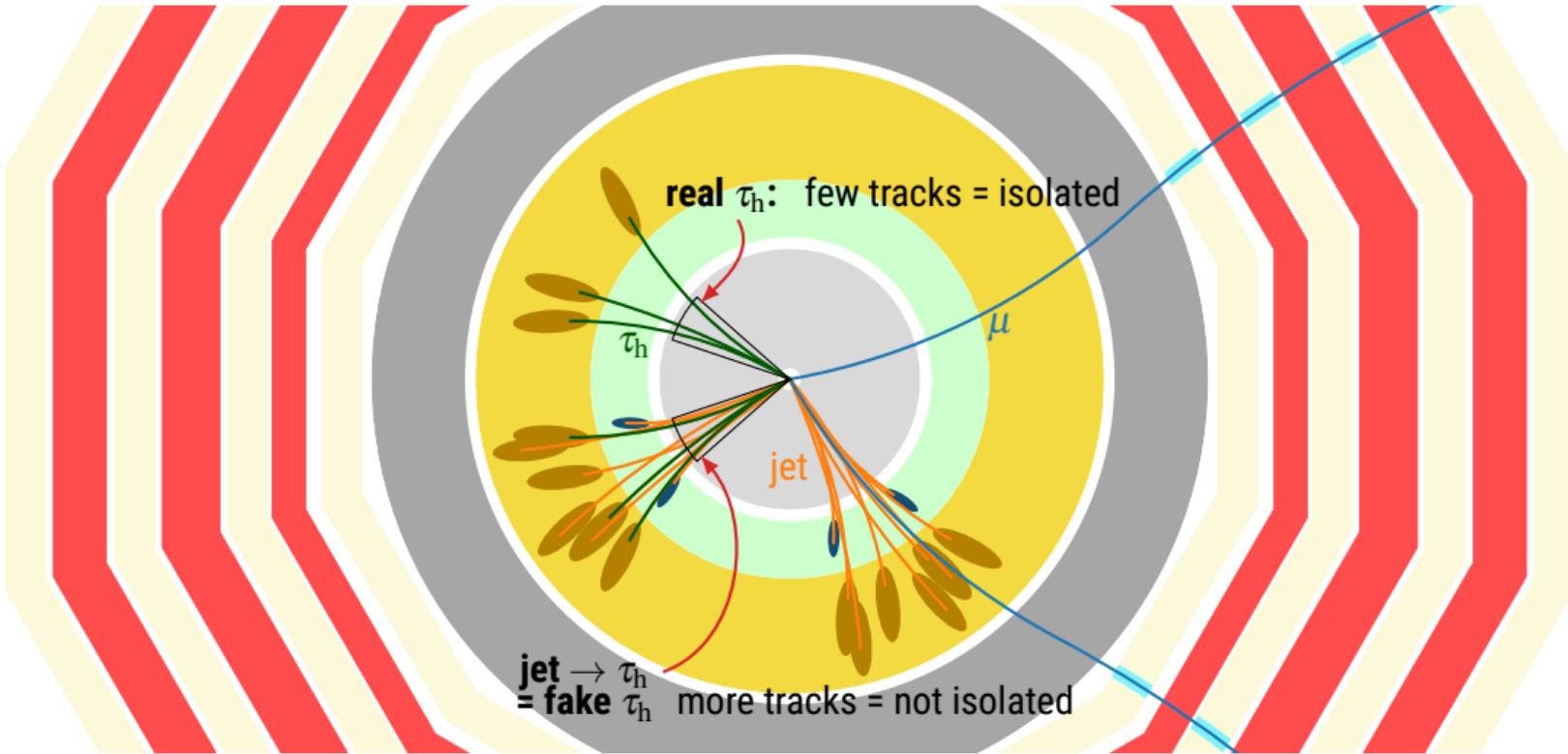
# The Fake Factor method & jets faking $\tau_h$

- ▶ How many events contain misidentified  $\tau_h$  (fake taus) in the Signal Region (SR)?



- ▷ J. Andrejkovic & J. Bechtel. "Data-driven background estimation of fake-tau backgrounds in di-tau final states with the full Run-II dataset". *CMS analysis Note* (June 2020). URL: [https://cms.cern.ch/iCMS/jsp/db\\_notes/noteInfo.jsp?cmsnoteid=CMS%20AN-2019/170](https://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%20AN-2019/170).

## Particles isolation – qualitatively



# The FF method: determination regions definitions

## QCD multijet ( $\tau_h\tau_h$ , $\mu\tau_h$ and $e\tau_h$ channels)

Same as SR, except:

- same signs for  $L_1$  and  $L_2$  electric charges (opposite signs in the SR).

▷ J. Andrejkovic & J. Bechtel. "Data-driven background estimation of fake-tau backgrounds in di-tau final states with the full Run-II dataset". *CMS analysis Note* (June 2020). URL: [https://cms.cern.ch/iCMS/jsp/db\\_notes/noteInfo.jsp?cmsnoteid=CMS%20AN-2019/170](https://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%20AN-2019/170).

# The FF method: determination regions definitions

## QCD multijet ( $\tau_h\tau_h$ , $\mu\tau_h$ and $e\tau_h$ channels)

Same as SR, except:

- same signs for  $L_1$  and  $L_2$  electric charges (opposite signs in the SR).

## $W + \text{jets}$ ( $\mu\tau_h$ and $e\tau_h$ channels)

Same as SR, except:

- transverse mass  $m_T^{(\ell)} > 70 \text{ GeV}$  ( $m_T^{(\ell)} < 70 \text{ GeV}$  in the SR);
- no  $b$ -jet (allowed in the SR).

▷ J. Andrejkovic & J. Bechtel. "Data-driven background estimation of fake-tau backgrounds in di-tau final states with the full Run-II dataset". *CMS analysis Note* (June 2020). URL: [https://cms.cern.ch/iCMS/jsp/db\\_notes/noteInfo.jsp?cmsnoteid=CMS%20AN-2019/170](https://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%20AN-2019/170).

# The FF method: determination regions definitions

## QCD multijet ( $\tau_h\tau_h$ , $\mu\tau_h$ and $e\tau_h$ channels)

Same as SR, except:

- same signs for  $L_1$  and  $L_2$  electric charges (opposite signs in the SR).

## $W + \text{jets}$ ( $\mu\tau_h$ and $e\tau_h$ channels)

Same as SR, except:

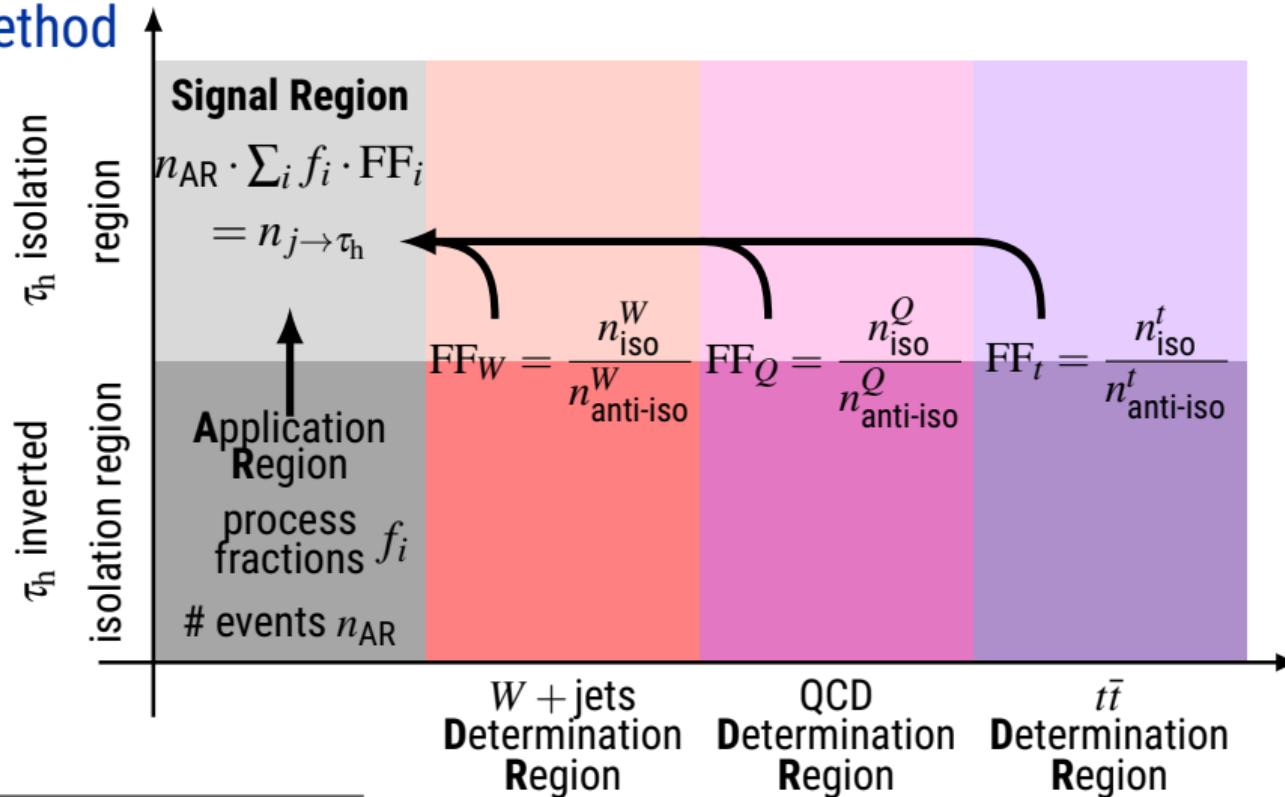
- transverse mass  $m_T^{(\ell)} > 70 \text{ GeV}$  ( $m_T^{(\ell)} < 70 \text{ GeV}$  in the SR);
- no  $b$ -jet (allowed in the SR).

## $t\bar{t}$ ( $\mu\tau_h$ and $e\tau_h$ channels)

Estimation from simulated samples, same selection as in SR.

▷ J. Andrejkovic & J. Bechtel. "Data-driven background estimation of fake-tau backgrounds in di-tau final states with the full Run-II dataset". *CMS analysis Note* (June 2020). URL: [https://cms.cern.ch/iCMS/jsp/db\\_notes/noteInfo.jsp?cmsnoteid=CMS%20AN-2019/170](https://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%20AN-2019/170).

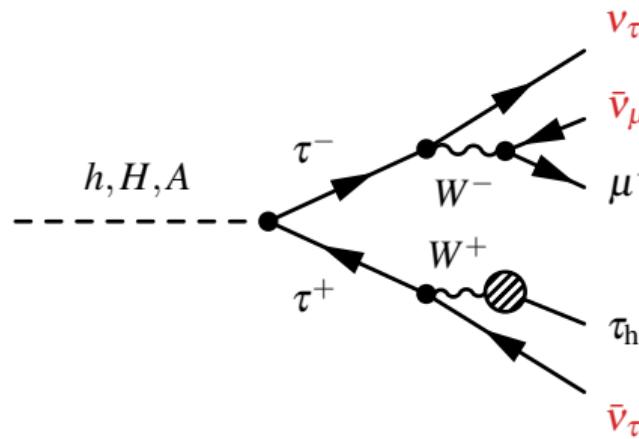
# The FF method



▷ J. Andrejkovic & J. Bechtel. "Data-driven background estimation of fake-tau backgrounds in di-tau final states with the full Run-II dataset". *CMS analysis Note* (June 2020). URL: [https://cms.cern.ch/iCMS/jsp/db\\_notes/noteInfo.jsp?cmsnoteid=CMS%20AN-2019/170](https://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%20AN-2019/170).

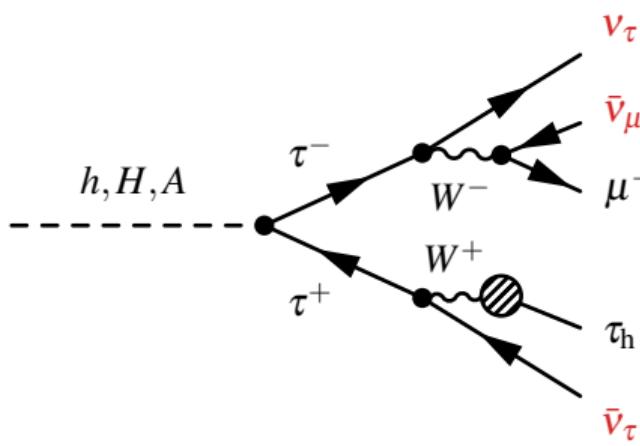
# Discriminant variable?

- ▶  $E_T^{\text{miss}}$  due to neutrinos.
- ▷ No invariant mass!

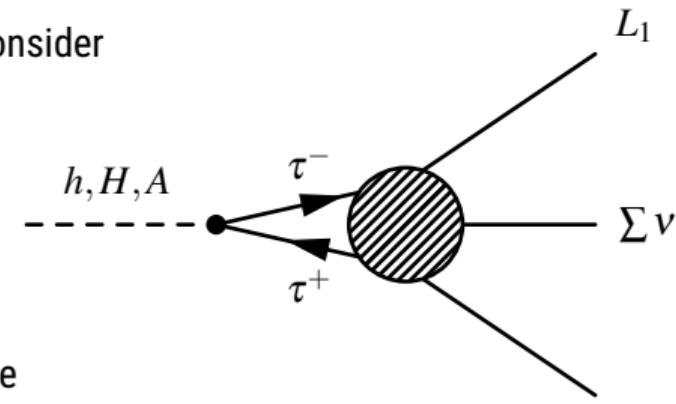


# Discriminant variable?

- ▶  $E_T^{\text{miss}}$  due to neutrinos.
- ▷ No invariant mass!



- ▶ Consider



where

- $L_1 = \mu$ ;
- $L_2 = \tau_h$ ;
- $\Sigma v \simeq E_T^{\text{miss}}$ ;

with respect to the left side.

# Discriminant variable: $m_T^{\text{tot}}$

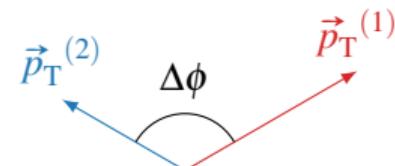
- ▶ For  $L_1, L_2$  and  $E_T^{\text{miss}}$  system,
  - ▷ in the transverse plane (use  $E_T^{\text{miss}}$ ),
  - ▷ for  $E_i \gg m_i$  (highly relativistic case),

deriving the "invariant" mass would then lead to

the **total transverse mass**,  $m_T^{\text{tot}}$

$$m_T^{\text{tot}} = \sqrt{m_T^2(L_1, E_T^{\text{miss}}) + m_T^2(L_2, E_T^{\text{miss}}) + m_T^2(L_1, L_2)}$$

$$m_T(1,2) = \sqrt{2p_T^{(1)} p_T^{(2)} (1 - \cos \Delta\phi)}$$

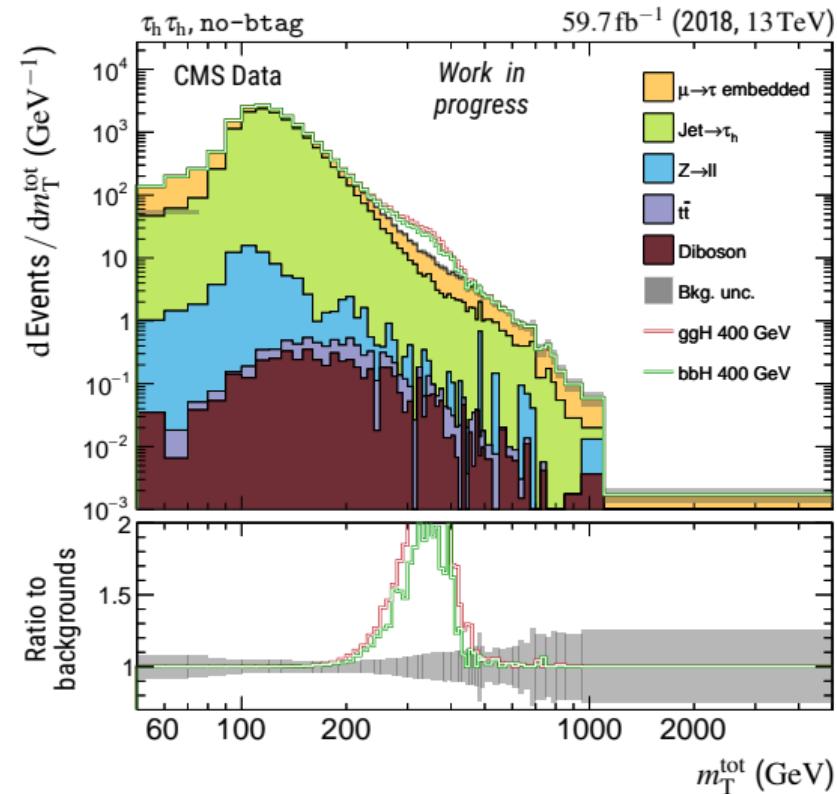


# Results obtained in this thesis?

# $m_T^{\text{tot}}$ distributions

## ► Backgrounds = SM expectations:

- ▷ DY  $Z \rightarrow \tau\tau$  and some  $t\bar{t}$  in  $\mu \rightarrow \tau$  embedded
- ▷ QCD,  $W + \text{jets}$  and some  $t\bar{t}$  in Jet  $\rightarrow \tau_h$
- ▷  $Z \rightarrow ee + Z \rightarrow \mu\mu$  in  $Z \rightarrow ll$
- ▷ Remaining  $t\bar{t}$  in  $t\bar{t}$
- ▷ Other small backgrounds in Diboson

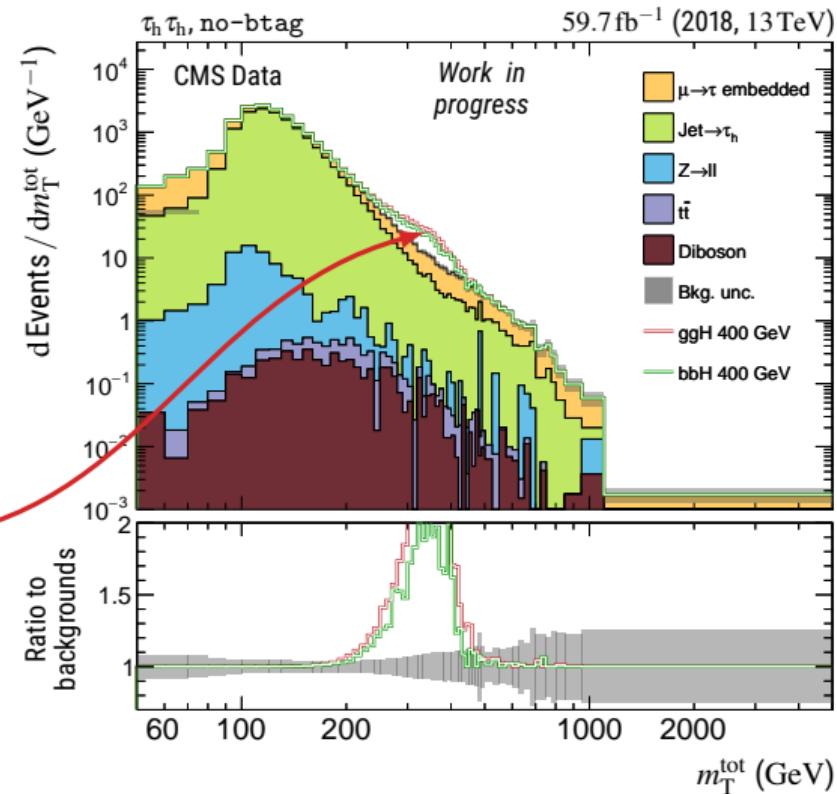


# $m_T^{\text{tot}}$ distributions

## ► Backgrounds = SM expectations:

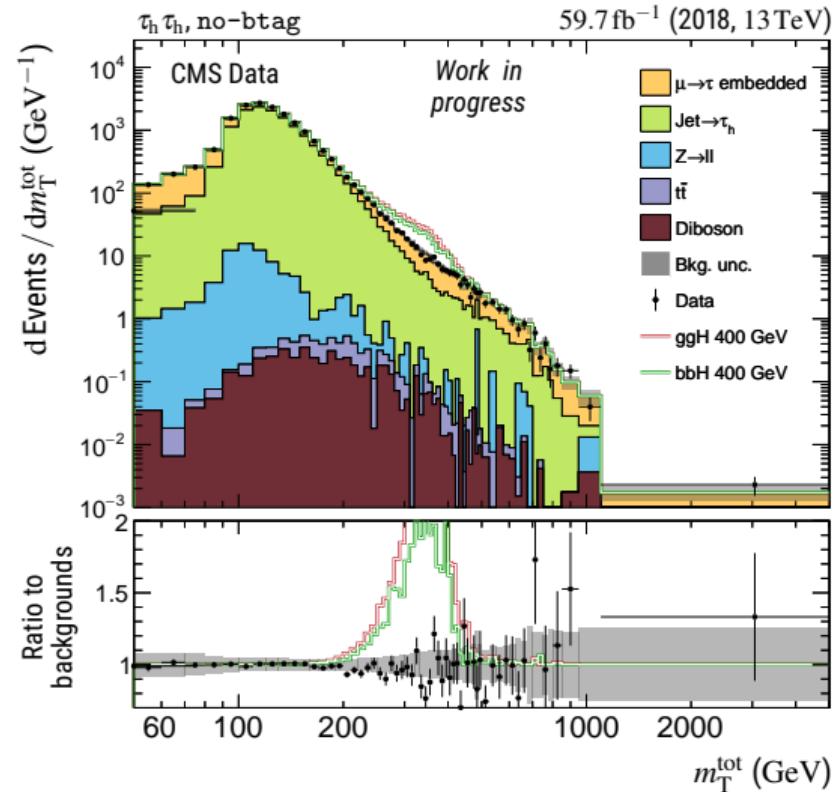
- ▷ DY  $Z \rightarrow \tau\tau$  and some  $t\bar{t}$  in  $\mu \rightarrow \tau$  embedded
- ▷ QCD,  $W + \text{jets}$  and some  $t\bar{t}$  in Jet  $\rightarrow \tau_h$
- ▷  $Z \rightarrow ee + Z \rightarrow \mu\mu$  in  $Z \rightarrow ll$
- ▷ Remaining  $t\bar{t}$  in  $t\bar{t}$
- ▷ Other small backgrounds in Diboson

►  $H$  at 400 GeV expected  $\sigma \times \mathcal{BR} = 1 \text{ pb}$  signal.



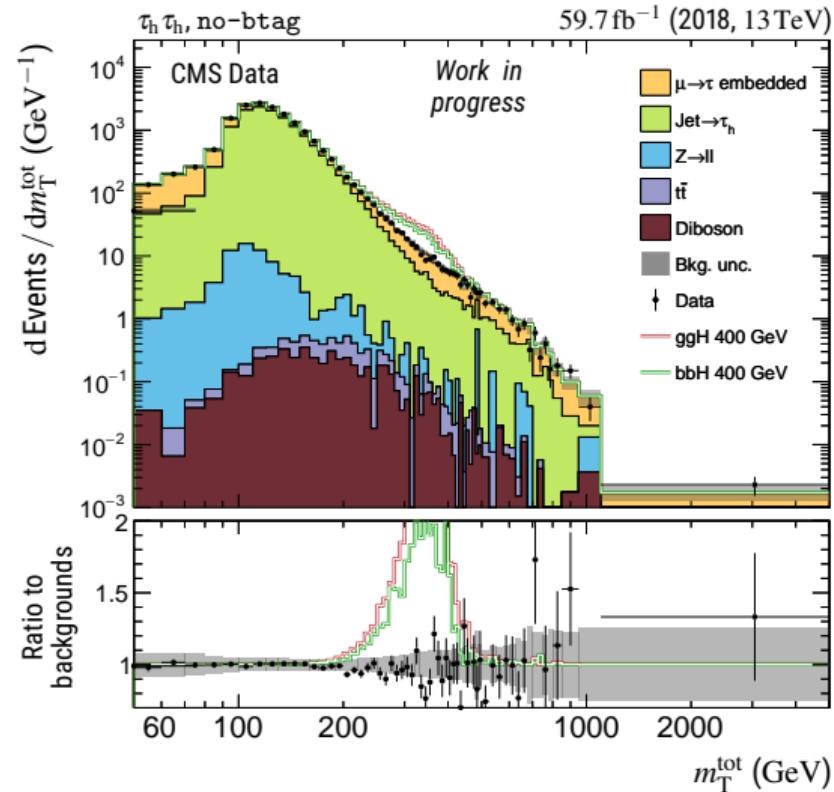
# $m_T^{\text{tot}}$ distributions

- ▶ Backgrounds = SM expectations:
  - ▷ DY  $Z \rightarrow \tau\tau$  and some  $t\bar{t}$  in  $\mu \rightarrow \tau$  embedded
  - ▷ QCD,  $W + \text{jets}$  and some  $t\bar{t}$  in Jet  $\rightarrow \tau_h$
  - ▷  $Z \rightarrow ee + Z \rightarrow \mu\mu$  in  $Z \rightarrow ll$
  - ▷ Remaining  $t\bar{t}$  in  $t\bar{t}$
  - ▷ Other small backgrounds in Diboson
- ▶  $H$  at 400 GeV expected  $\sigma \times \mathcal{BR} = 1 \text{ pb}$  signal.
- ▶ Compare to observed events (black dots).



# $m_T^{\text{tot}}$ distributions

- ▶ Backgrounds = SM expectations:
  - ▷ DY  $Z \rightarrow \tau\tau$  and some  $t\bar{t}$  in  $\mu \rightarrow \tau$  embedded
  - ▷ QCD,  $W + \text{jets}$  and some  $t\bar{t}$  in Jet  $\rightarrow \tau_h$
  - ▷  $Z \rightarrow ee + Z \rightarrow \mu\mu$  in  $Z \rightarrow ll$
  - ▷ Remaining  $t\bar{t}$  in  $t\bar{t}$
  - ▷ Other small backgrounds in Diboson
- ▶  $H$  at 400 GeV expected  $\sigma \times \mathcal{BR} = 1 \text{ pb}$  signal.
- ▶ Compare to observed events (black dots).
- ▶ Data/Bkg agreement  $\rightarrow$  **exclusion limits** on  $\sigma \times \mathcal{BR}$



# $m_T^{\text{tot}}$ distributions

## ► Background contributions:

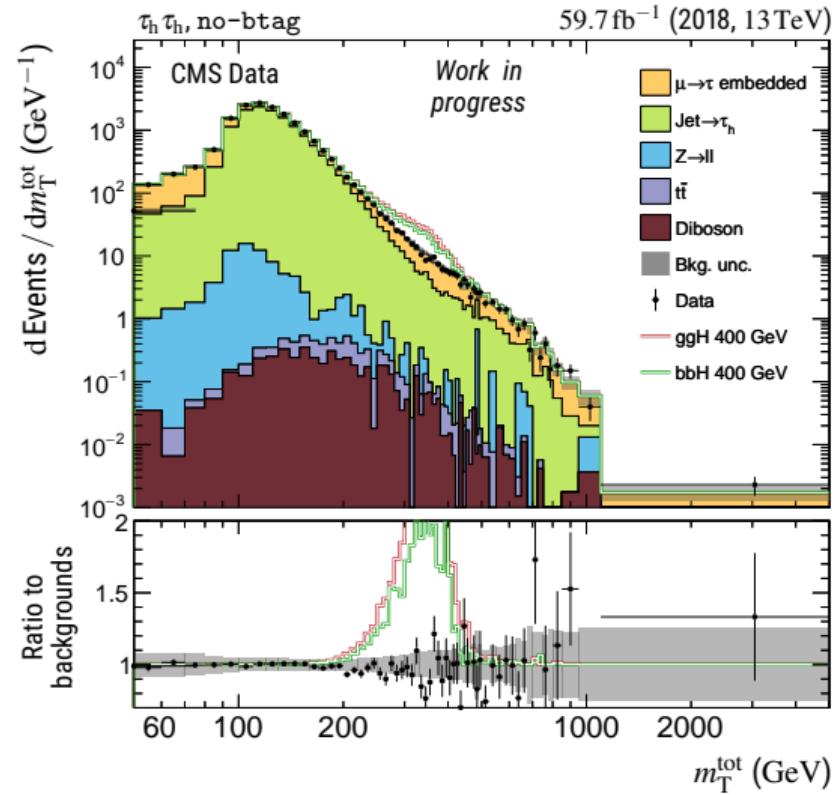
**Not just a plot!**

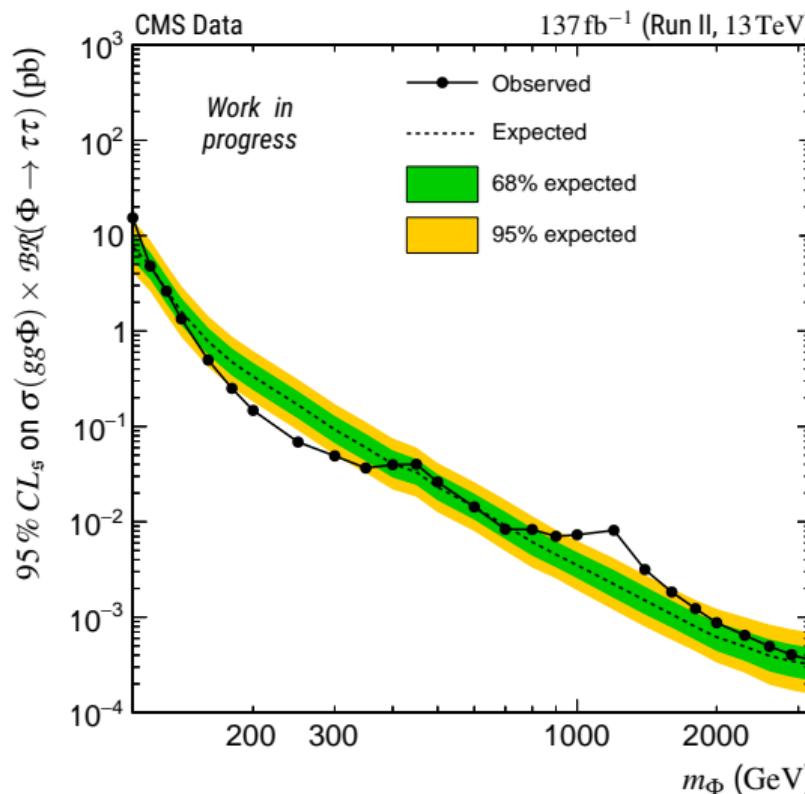
### ► Lot of hard work to obtain this: $\tau$ embedded

- ▷ simulated events
- ▷ QCD,  $W + \text{jets}$  and some  $t\bar{t}$  in Jet  $\rightarrow \tau_h$
- ▷ detector issues
- ▷  $Z \rightarrow ee + Z \rightarrow ll$
- ▷ uncertainties measured

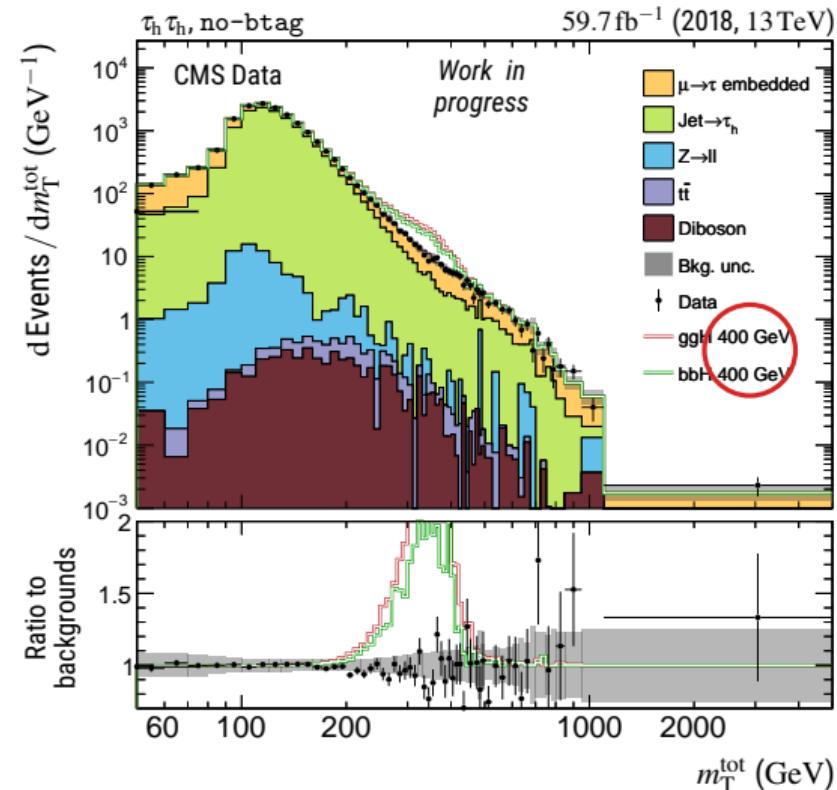
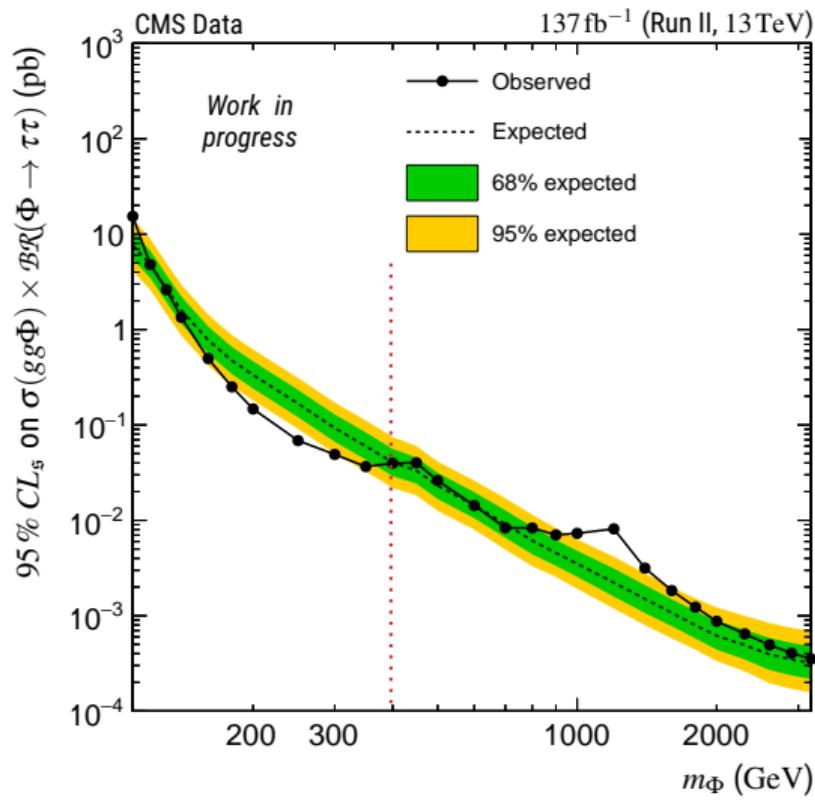
### ► Collaborative work:

- ▷ Other collaboration
- ▷ Karlsruhe Institute of Technology (DE)
- ▷ Imperial College (UK)  $\mathcal{BR} = 1 \text{ pb}$  signal.
- ▷ DESY (DE) observed events (black dots).
- ▷ HEPHY (AT)  $\rightarrow$  exclusion limits on  $\sigma \times \mathcal{BR}$
- ▷ IP2I (FR)

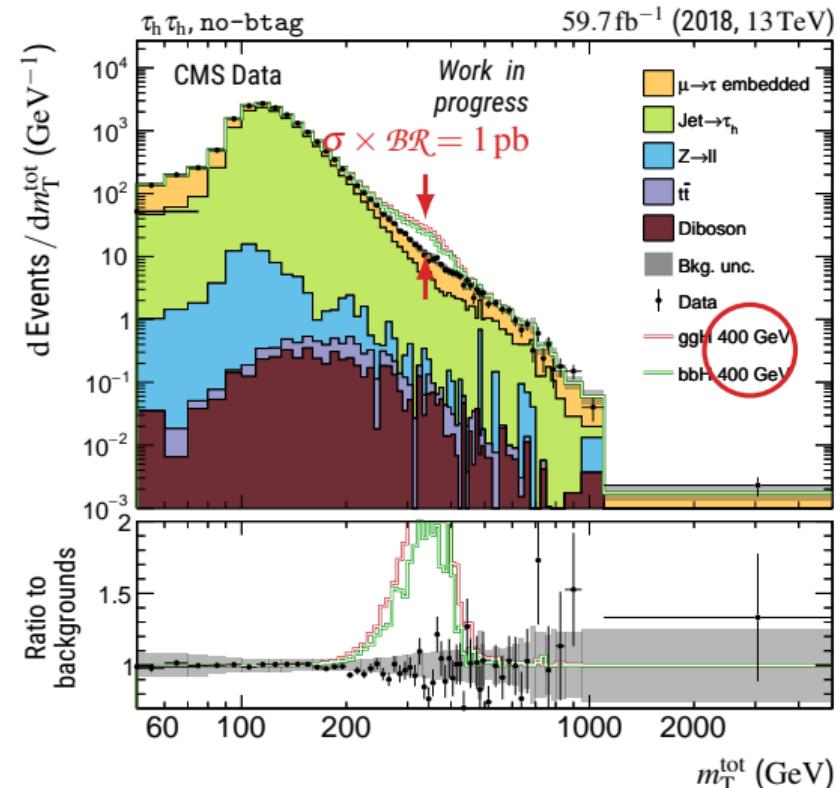
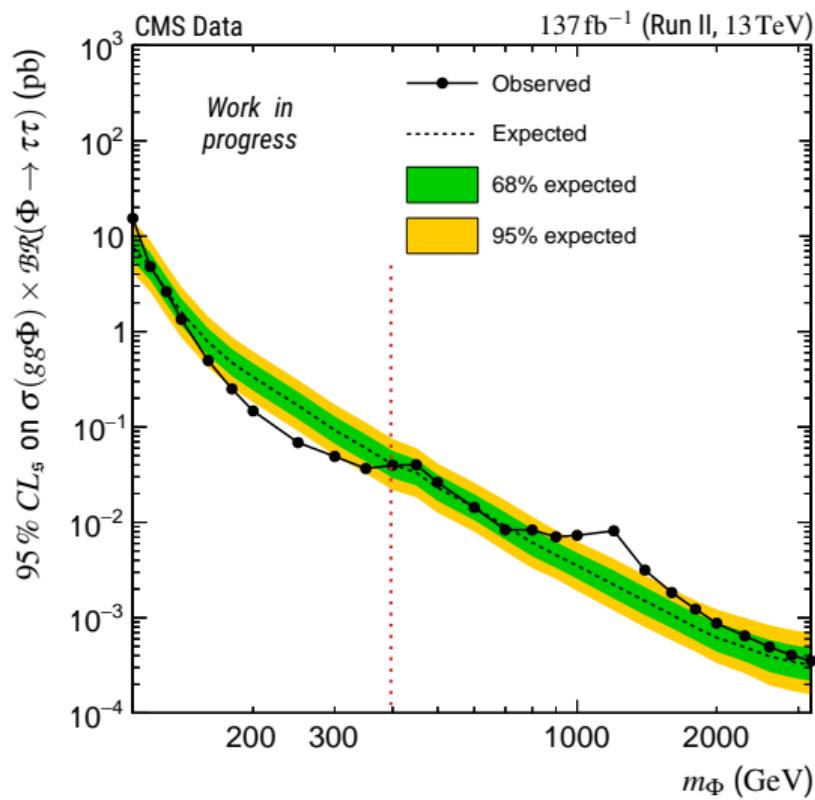




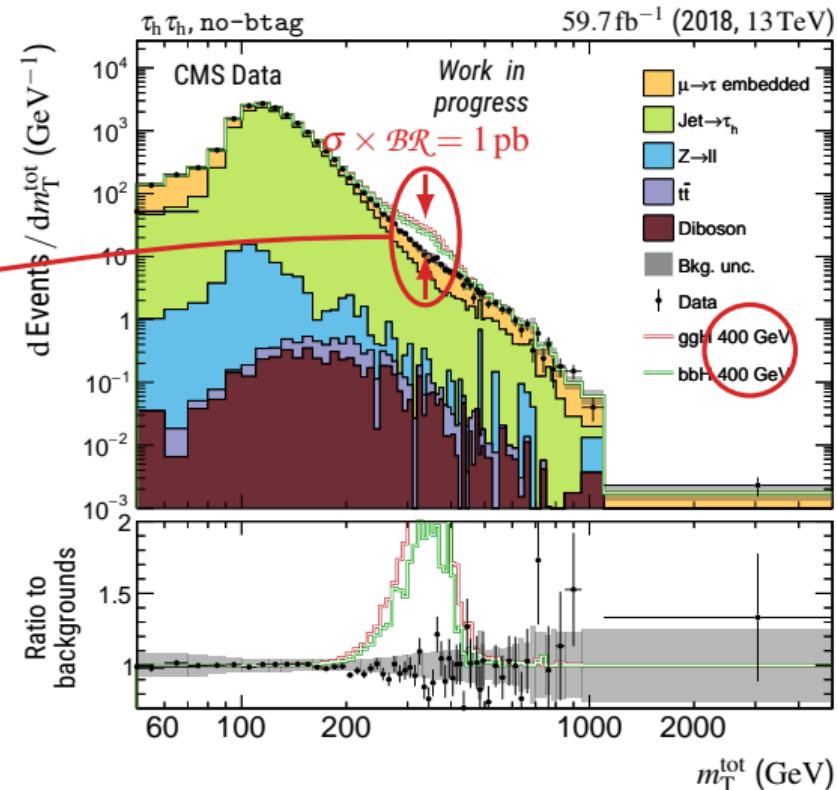
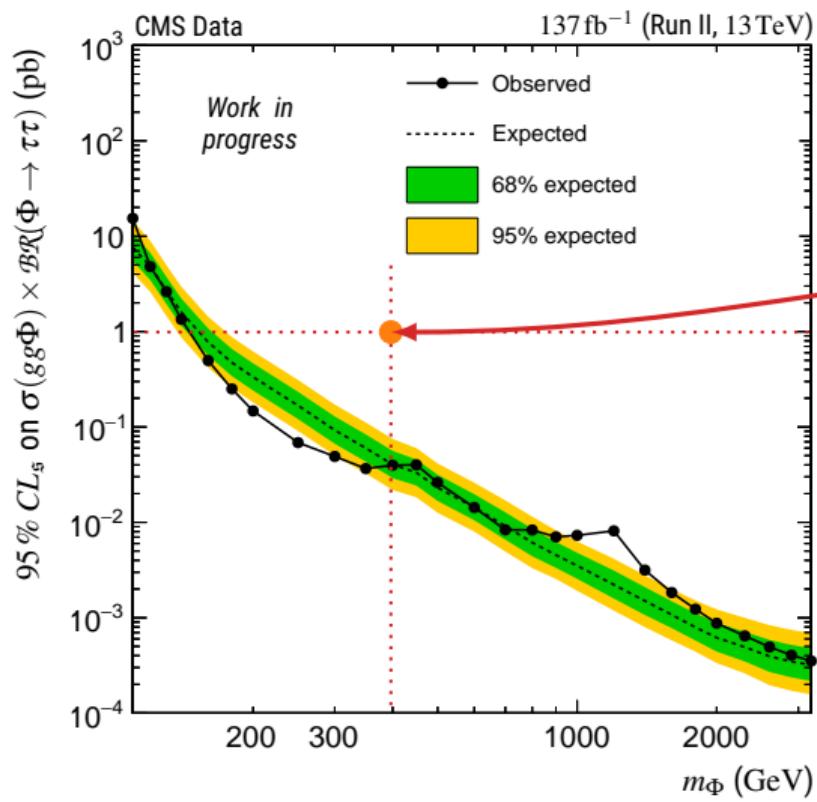
- ▷ A. L. Read. "Modified frequentist analysis of search results (the  $CL_s$  method)". *Workshop on confidence limits, CERN, Geneva, Switzerland, 17-18 Jan 2000: Proceedings*. CERN-OPEN-2000-205. May 2000. URL: <http://cds.cern.ch/record/451614>.



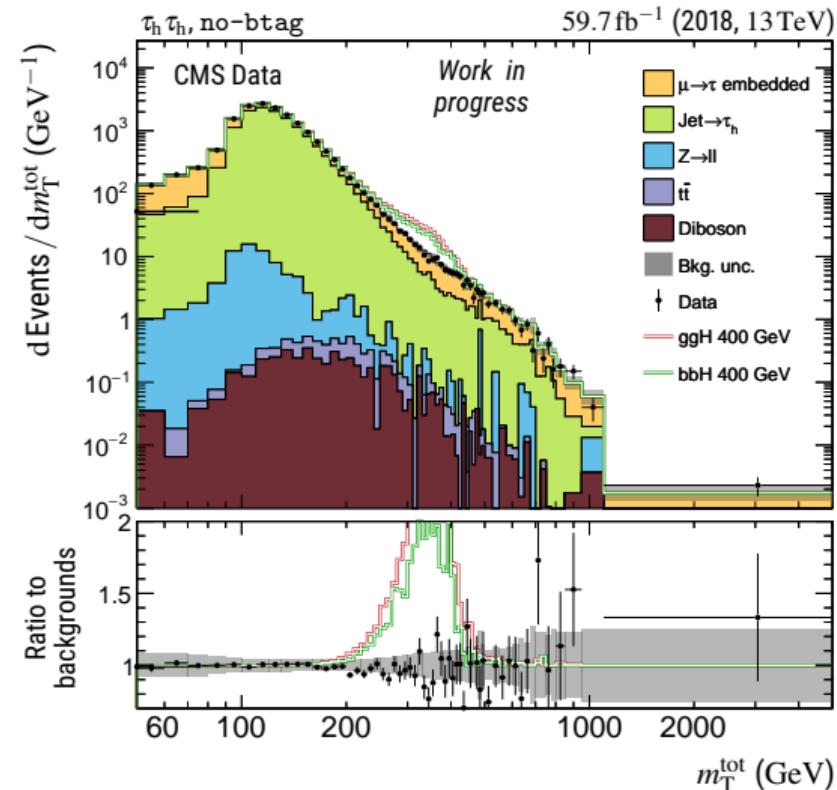
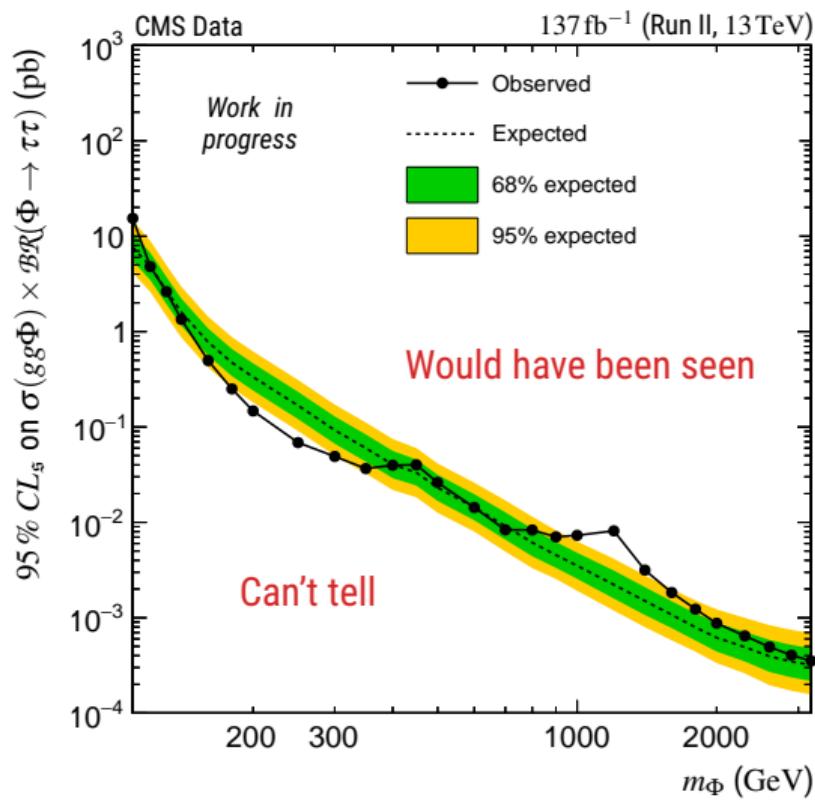
▷ A. L. Read. "Modified frequentist analysis of search results (the  $CL_s$  method)". *Workshop on confidence limits, CERN, Geneva, Switzerland, 17-18 Jan 2000: Proceedings*. CERN-OPEN-2000-205. May 2000. URL: <http://cds.cern.ch/record/451614>.



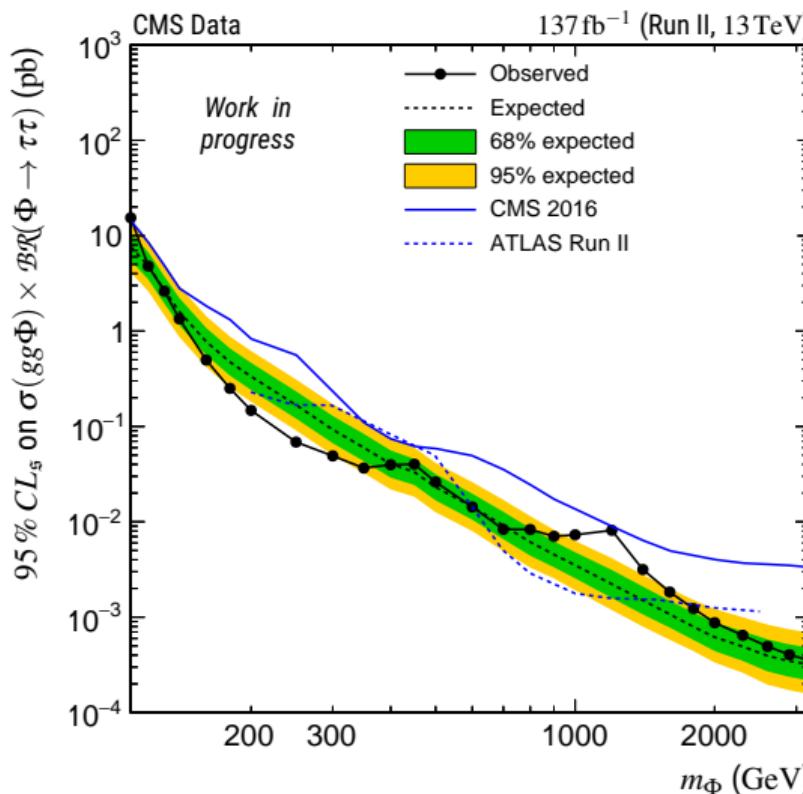
▷ A. L. Read. "Modified frequentist analysis of search results (the  $CL_s$  method)". *Workshop on confidence limits, CERN, Geneva, Switzerland, 17-18 Jan 2000: Proceedings*. CERN-OPEN-2000-205. May 2000. URL: <http://cds.cern.ch/record/451614>.



▷ A. L. Read. "Modified frequentist analysis of search results (the  $CL_s$  method)". *Workshop on confidence limits, CERN, Geneva, Switzerland, 17-18 Jan 2000: Proceedings*. CERN-OPEN-2000-205. May 2000. URL: <http://cds.cern.ch/record/451614>.



▷ A. L. Read. "Modified frequentist analysis of search results (the  $CL_s$  method)". *Workshop on confidence limits, CERN, Geneva, Switzerland, 17-18 Jan 2000: Proceedings*. CERN-OPEN-2000-205. May 2000. URL: <http://cds.cern.ch/record/451614>.



### ► CMS 2016:

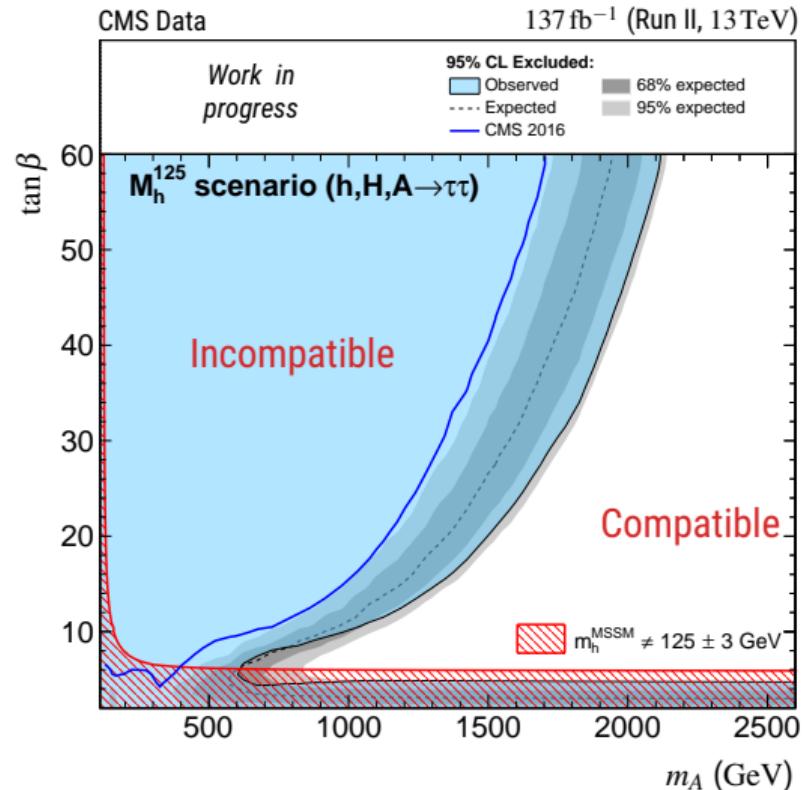
- ▷ [The CMS Collaboration](#). "Search for additional neutral MSSM Higgs bosons in the  $\tau\tau$  final state in  $pp$  collisions at  $\sqrt{s} = 13 \text{ TeV}$ ". *Journal of High Energy Physics* **09**.007 (Sept. 2018). DOI: [10.1007/JHEP09\(2018\)007](https://doi.org/10.1007/JHEP09(2018)007).

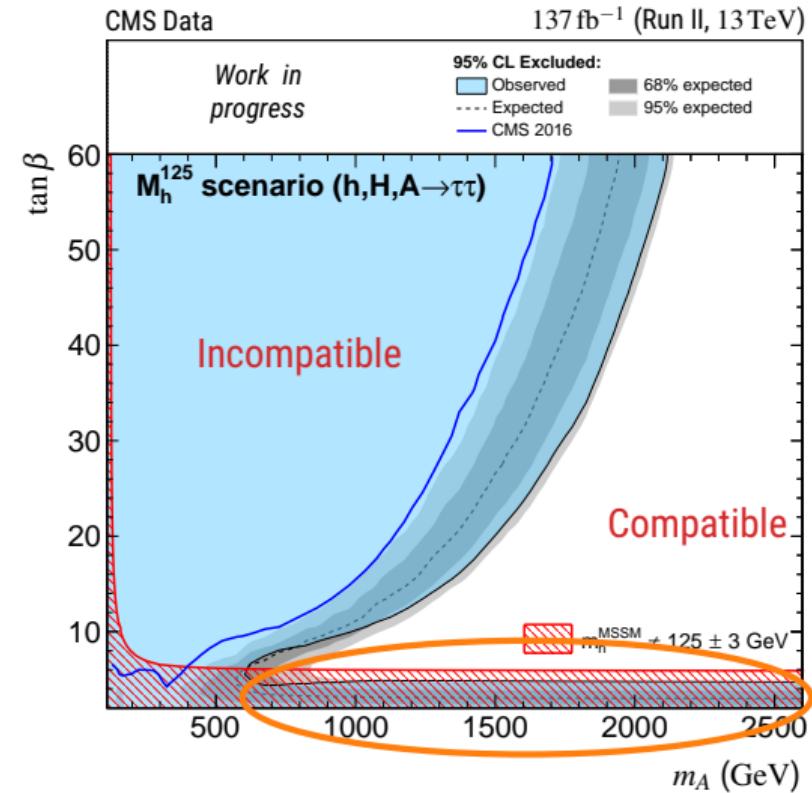
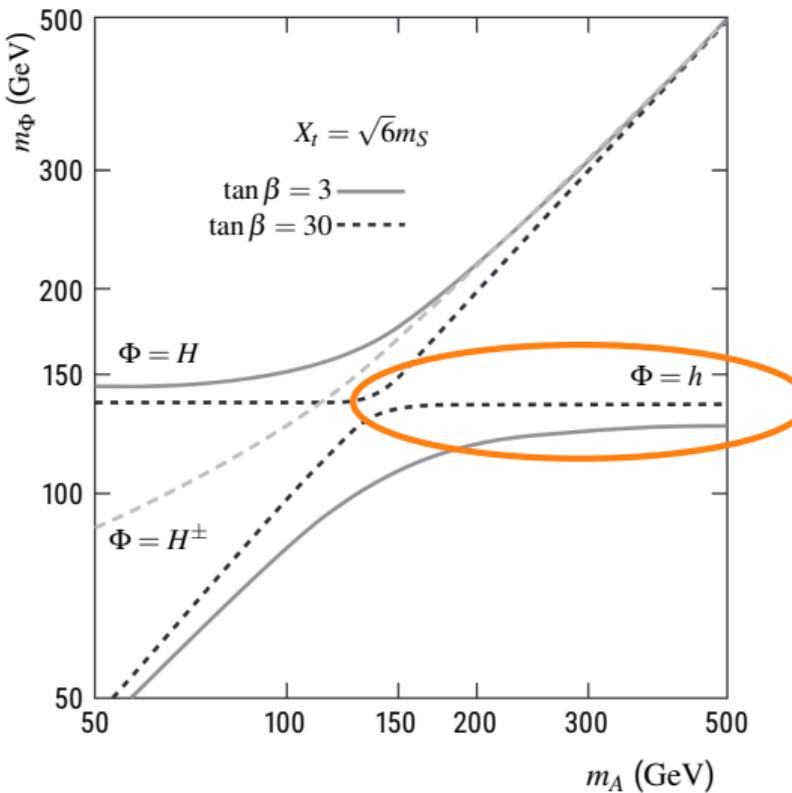
### ► ATLAS Run II:

- ▷ [The ATLAS Collaboration](#). "Search for Heavy Higgs Bosons Decaying into Two Tau Leptons with the ATLAS Detector Using  $pp$  Collisions at  $\sqrt{s} = 13 \text{ TeV}$ ". *Physical Review Letters* **125** (5 July 2020), p. 051801. DOI: [10.1103/PhysRevLett.125.051801](https://doi.org/10.1103/PhysRevLett.125.051801).

▷ A. L. Read. "Modified frequentist analysis of search results (the  $CL_s$  method)". *Workshop on confidence limits, CERN, Geneva, Switzerland, 17-18 Jan 2000: Proceedings*. CERN-OPEN-2000-205. May 2000. URL: <http://cds.cern.ch/record/451614>.

- ▶ Model dependent limits:
  - ▷ Fix high-order MSSM parameter,
  - ▷ Explore  $(m_A, \tan \beta)$  plane,
  - ▷ Is MSSM compatible with data? Look for:
    - ▷ additional signal from  $H$  and  $A$
    - ▷ compatibility with the already discovered  $h$



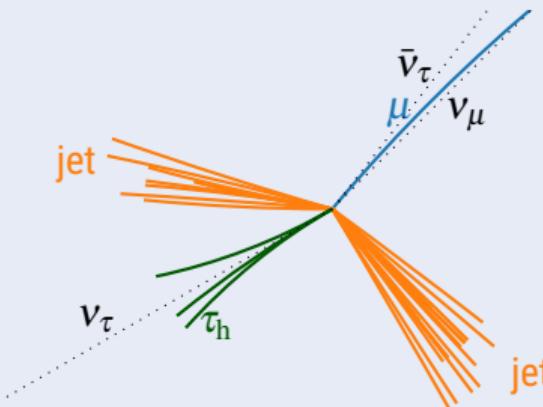


- ▶ Remember: invariant mass not fully available:
  - ▷ neutrinos in di- $\tau$  events.

- ▶ Remember: invariant mass not fully available:
  - ▷ neutrinos in di- $\tau$  events.

## What's here

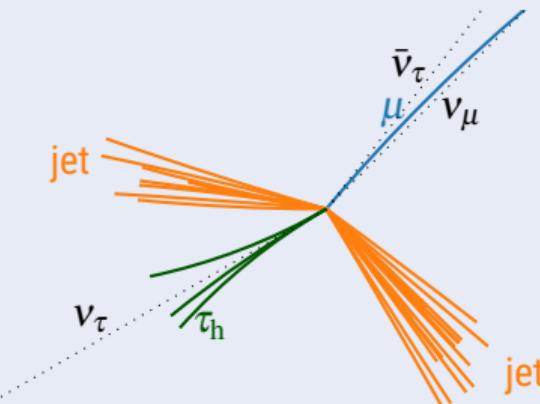
(e.g. VBF Higgs production + decay to  $\tau\tau, \mu\tau_h$  channel)



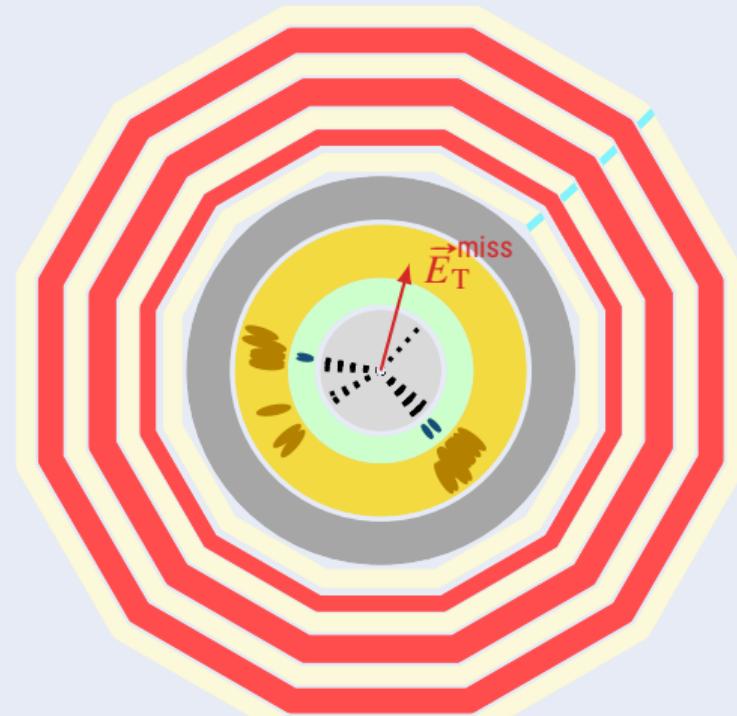
- ▶ Remember: invariant mass not fully available:
  - ▷ neutrinos in di- $\tau$  events.

## What's here

(e.g. VBF Higgs production + decay to  $\tau\tau, \mu\tau_h$  channel)



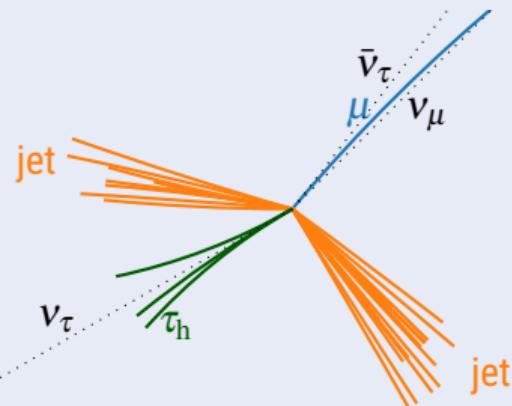
What CMS sees: no neutrinos but  $E_T^{\text{miss}}$



- ▶ Remember: invariant mass not fully available:
  - ▷ neutrinos in di- $\tau$  events.

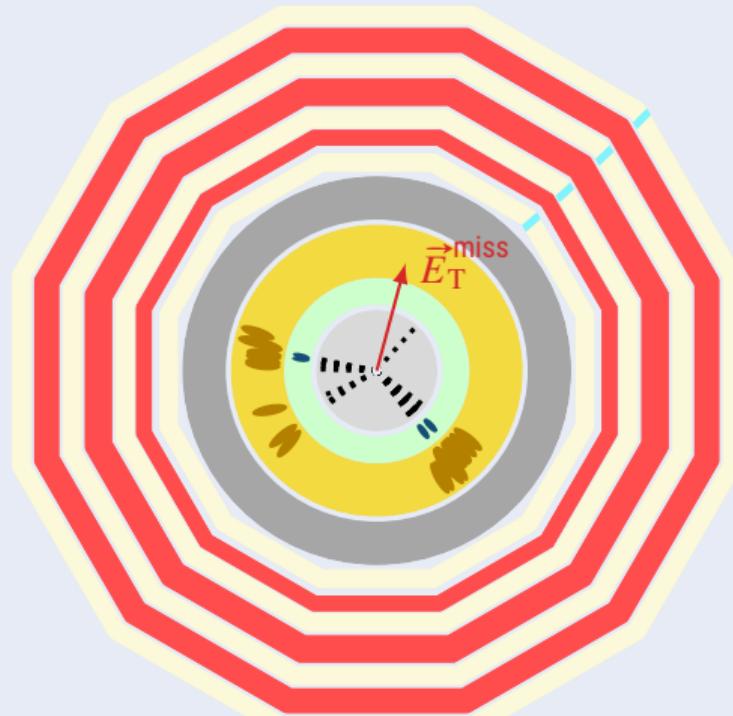
## What's here

(e.g. VBF Higgs production + decay to  $\tau\tau, \mu\tau_h$  channel)



- ▶ It would be great to have a di- $\tau$  mass estimator!
  - ▷ What about **machine learning?**

What CMS sees: no neutrinos but  $E_T^{\text{miss}}$



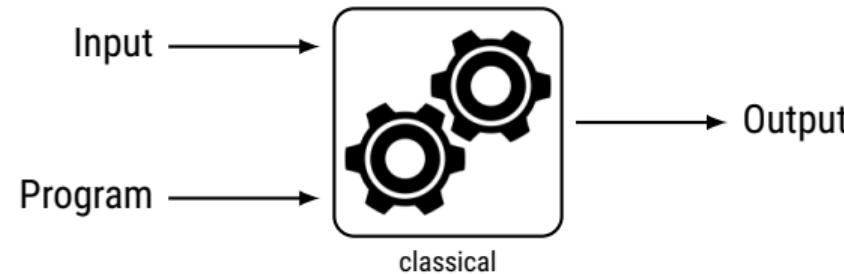
## 1 Phenomenology

## 2 Experimental device

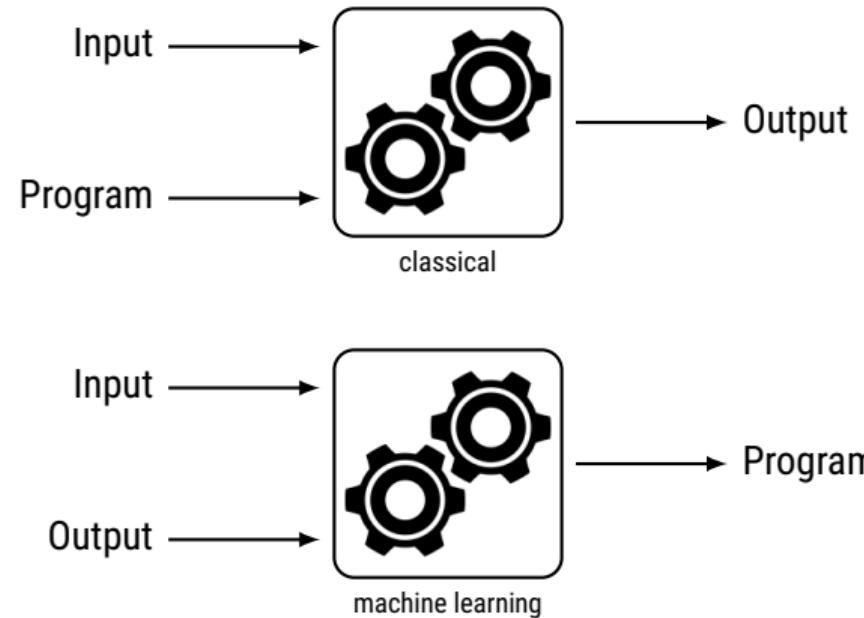
## 3 $H/A \rightarrow \tau\tau$ analysis

## 4 Machine learning

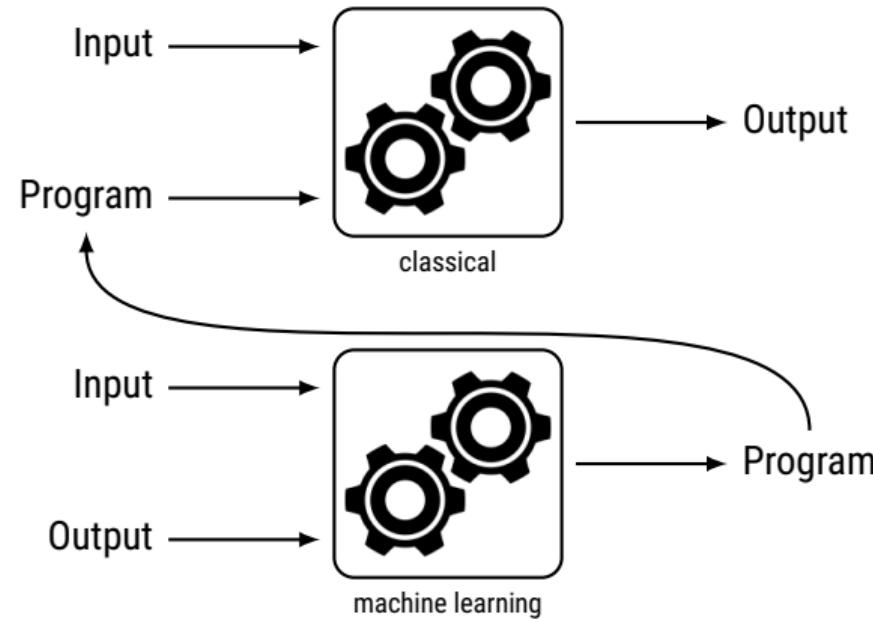
# What is *machine learning*? – A brief introduction



# What is *machine learning*? – A brief introduction



# What is *machine learning*? – A brief introduction



# What is *machine learning*? – A brief introduction

**Aim:** find a function (program) mapping features (input) to a target (output)

# What is *machine learning*? – A brief introduction

**Aim:** find a function (program) mapping features (input) to a target (output)

- ▶ Categorical target ⇒ Classification

e.g. cat or dog on the image



- ▶ C. Bernet. *The Data Frog – Image Recognition: Dogs vs Cats!* URL:  
<https://thedatafrog.com/en/articles/dogs-vs-cats/>.

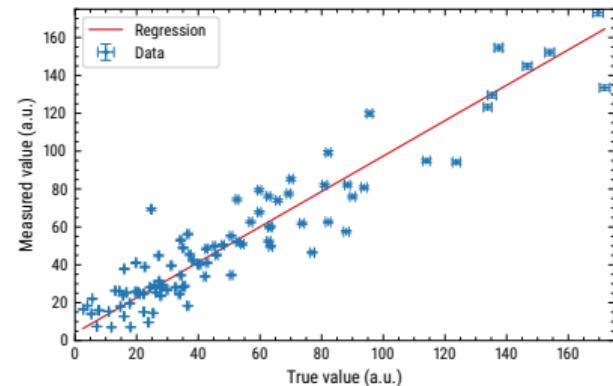
# What is *machine learning*? – A brief introduction

**Aim:** find a function (program) mapping features (input) to a target (output)

- ▶ Categorical target  $\Rightarrow$  Classification  
e.g. cat or dog on the image



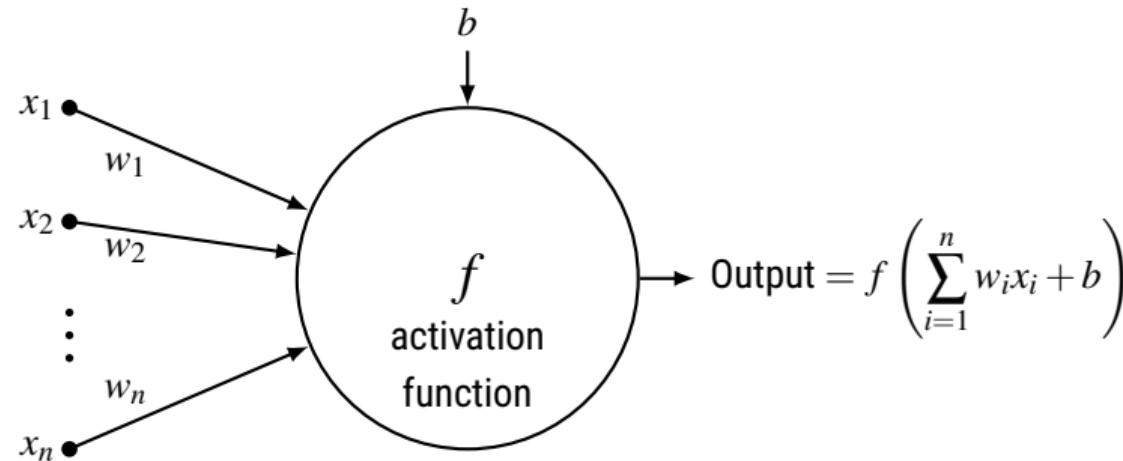
- ▶ Continuous target  $\Rightarrow$  Regression  
e.g. discriminating variable!  
Linear case:



- ▶ C. Bernet. *The Data Frog – Image Recognition: Dogs vs Cats!* URL:  
<https://thedatafrog.com/en/articles/dogs-vs-cats/>.

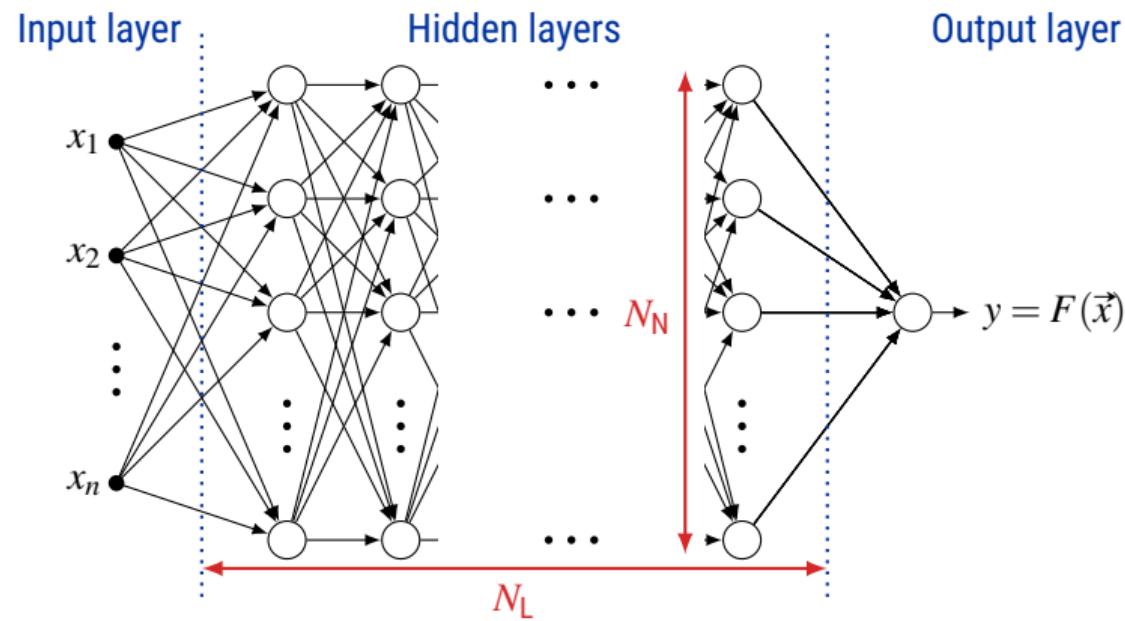
What if the target is not linear wrt. input?

# Neurons in ML



- ▶ Parameters:  $w_1, w_2, \dots, w_n, b$
- ▶ Equivalent to linear regression for  $f = \mathbb{1} : x \mapsto x$

# (Deep) Neural Networks

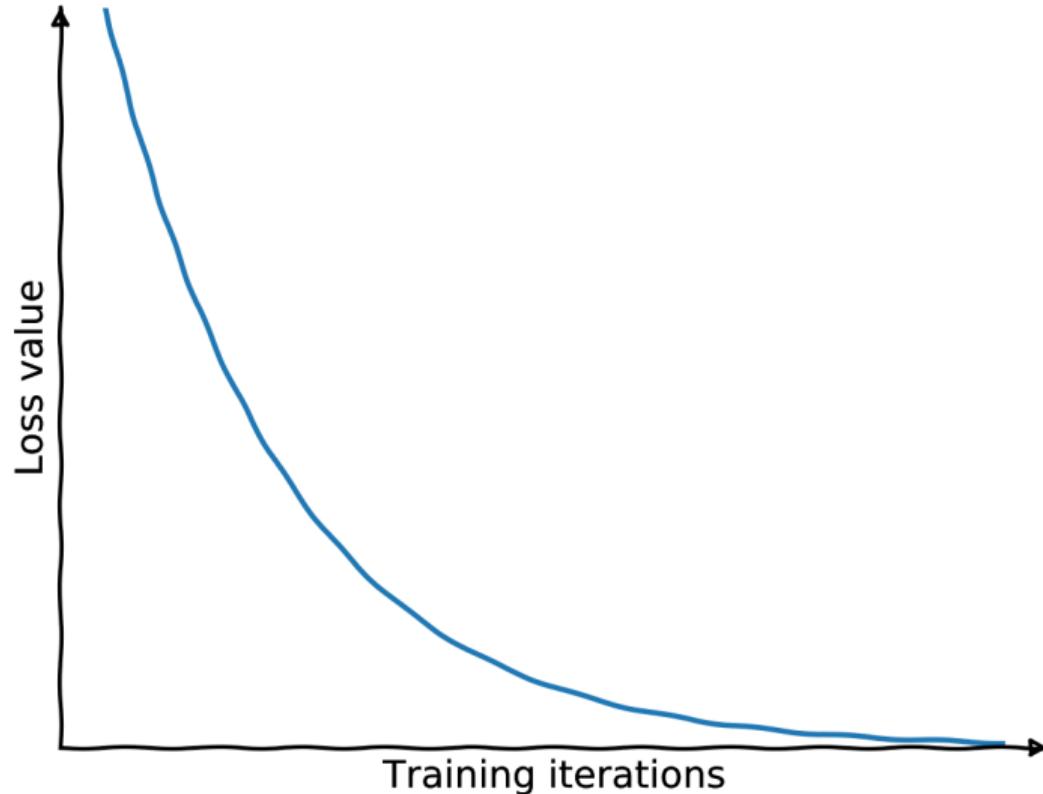


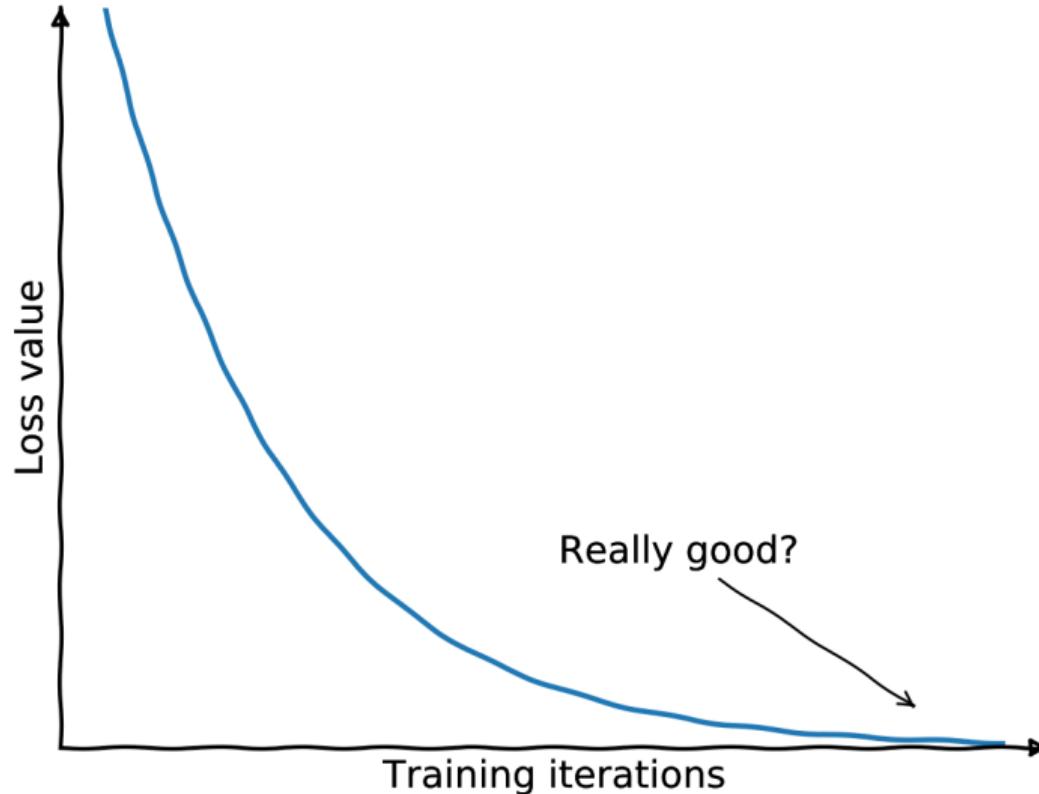
# How to train a neural network?

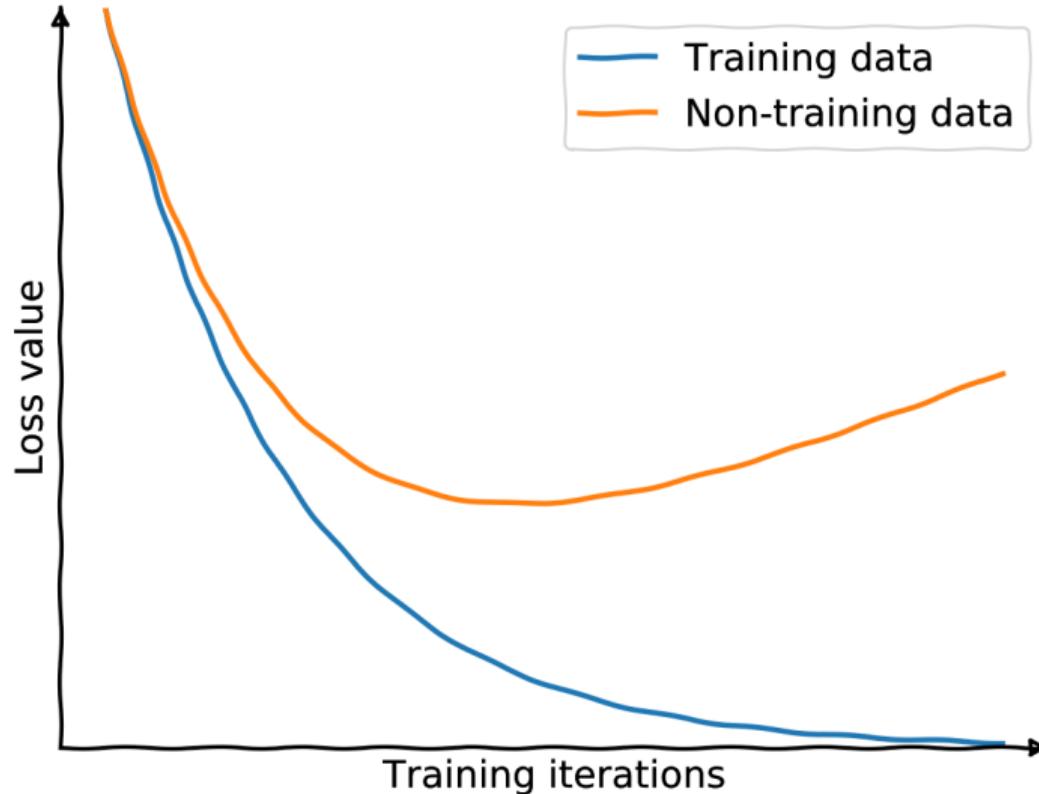
Train = optimize parameters  $(w_1, w_2, \dots, w_n, b)$  for each neuron.

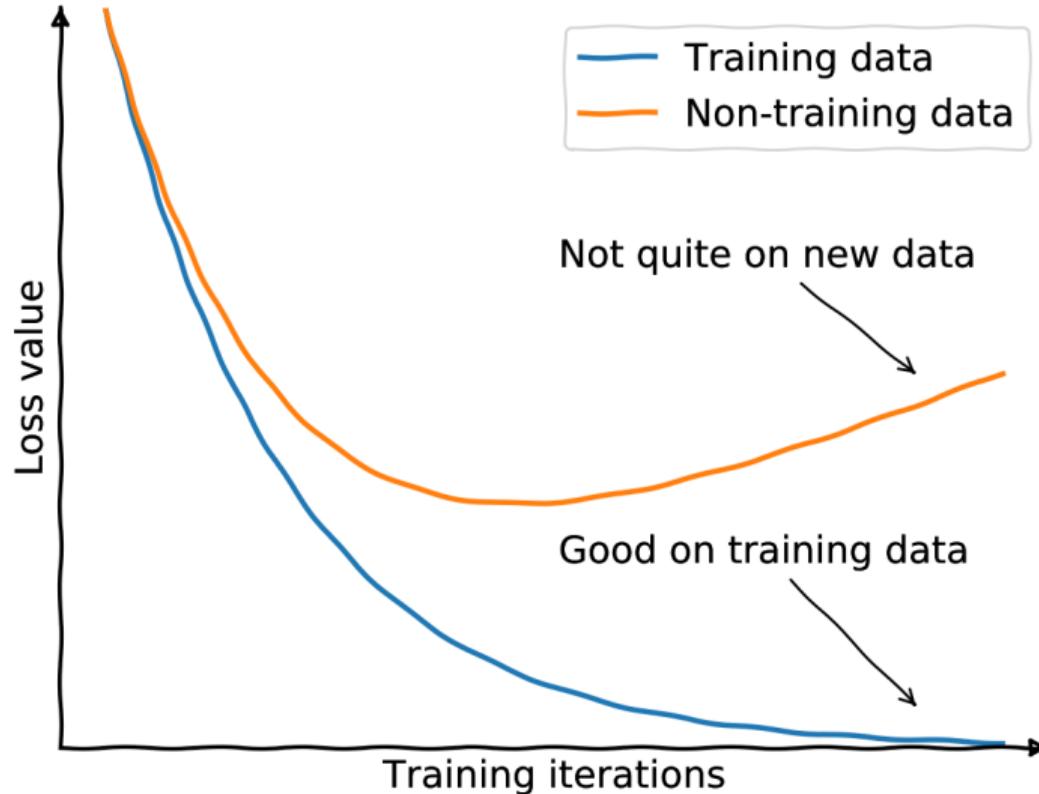
- ▶ Get a **training dataset** = examples of inputs  $\vec{x}_i$  with corresponding outputs  $y_i$
- ▶ Compare the model predictions  $F(\vec{x}_i)$  to the true values  $y_i$ 
  - ▷ Define a **loss function**  $\mathcal{L}$  such that its minimum is reached when  $F(\vec{x}_i) = y_i$
  - ▷ Change the parameters a bit, aiming at minimizing  $\mathcal{L}(F(\vec{x}_i), y_i)$
  - ▷ Repeat

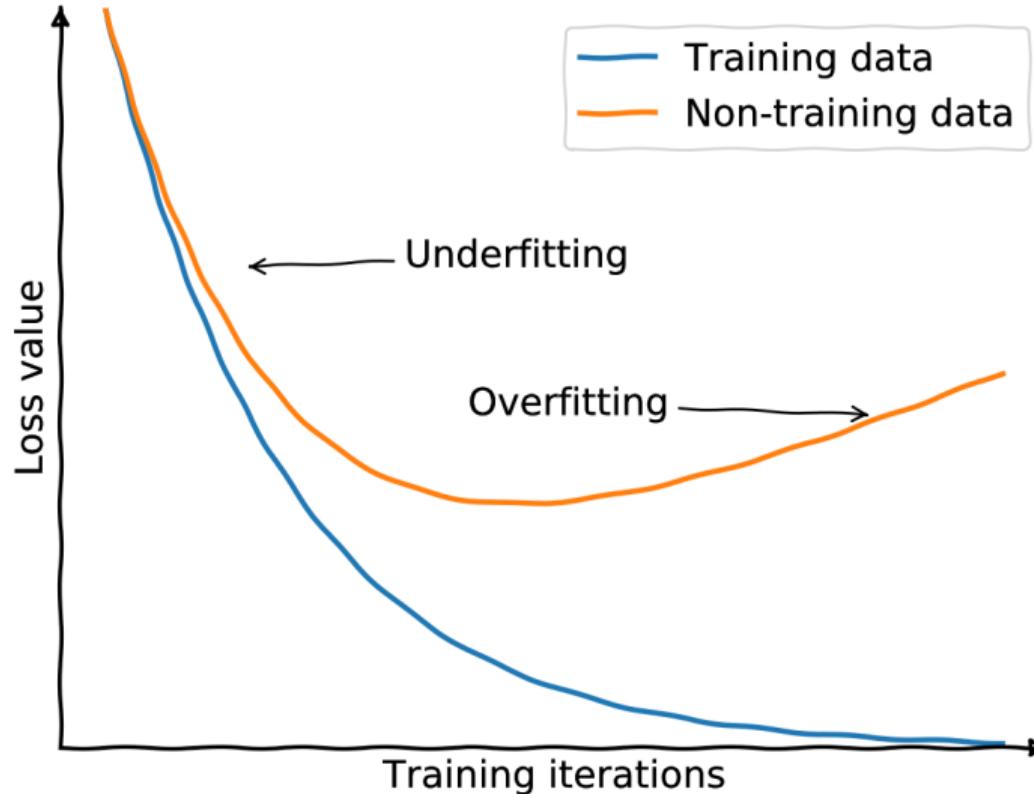
When to stop training?

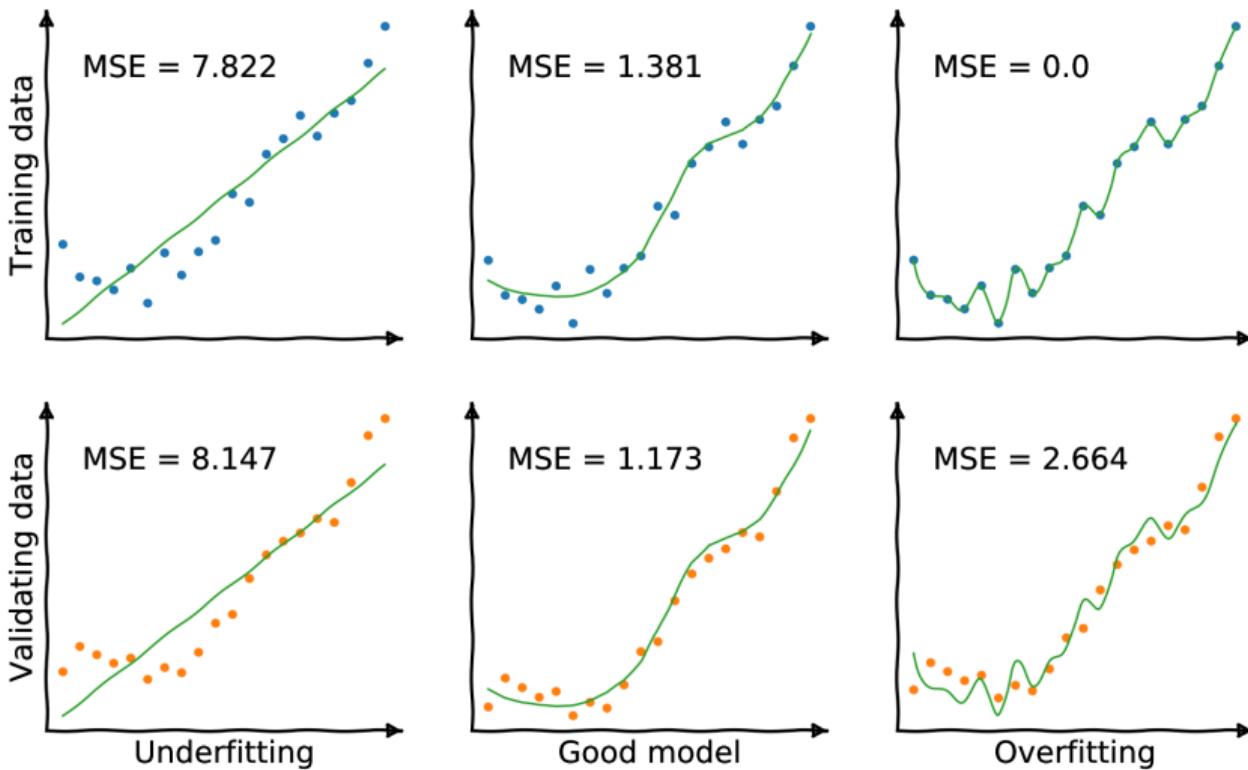


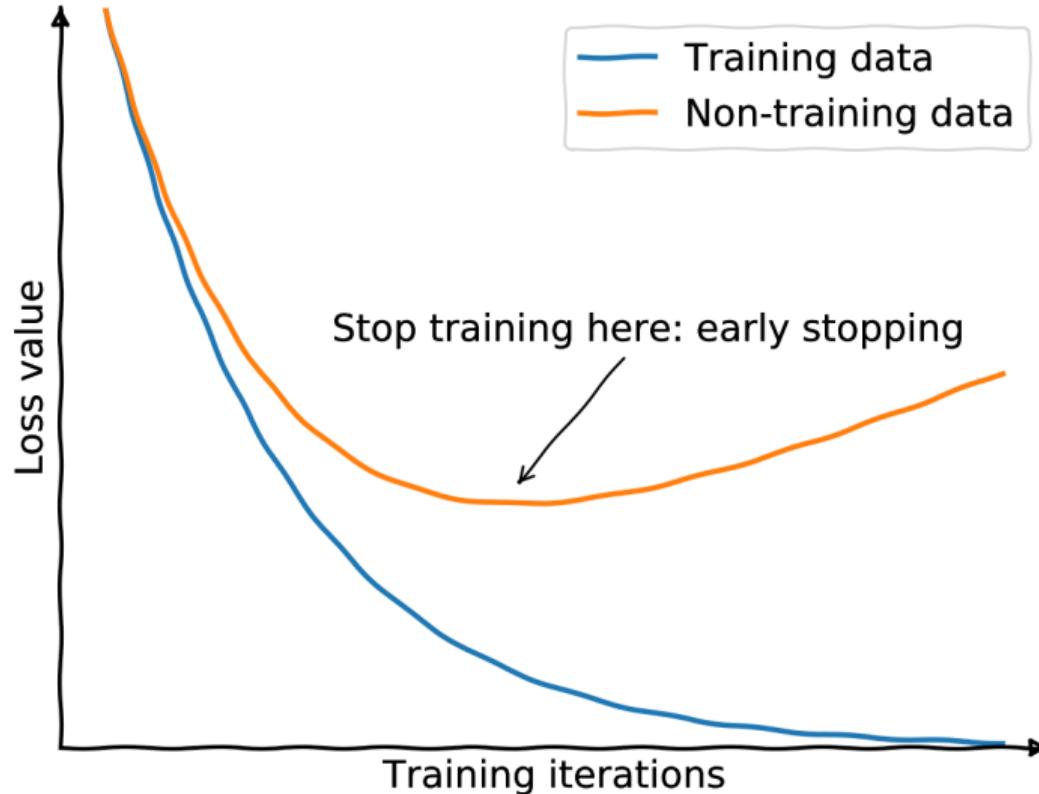


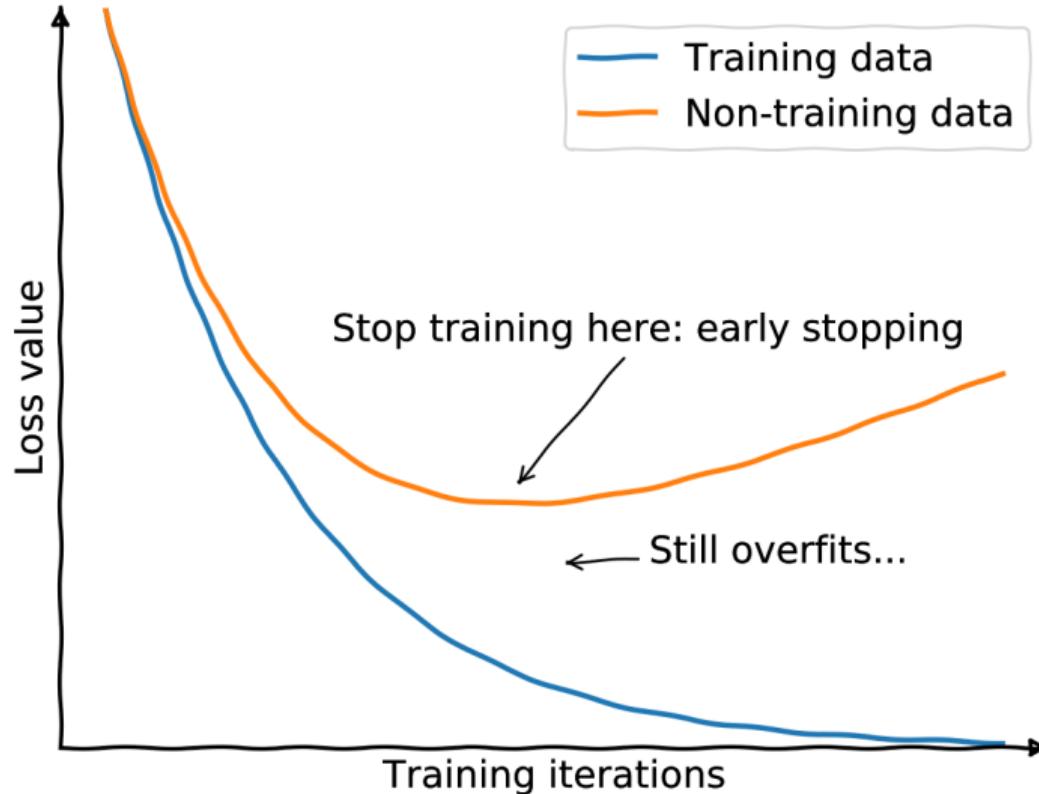


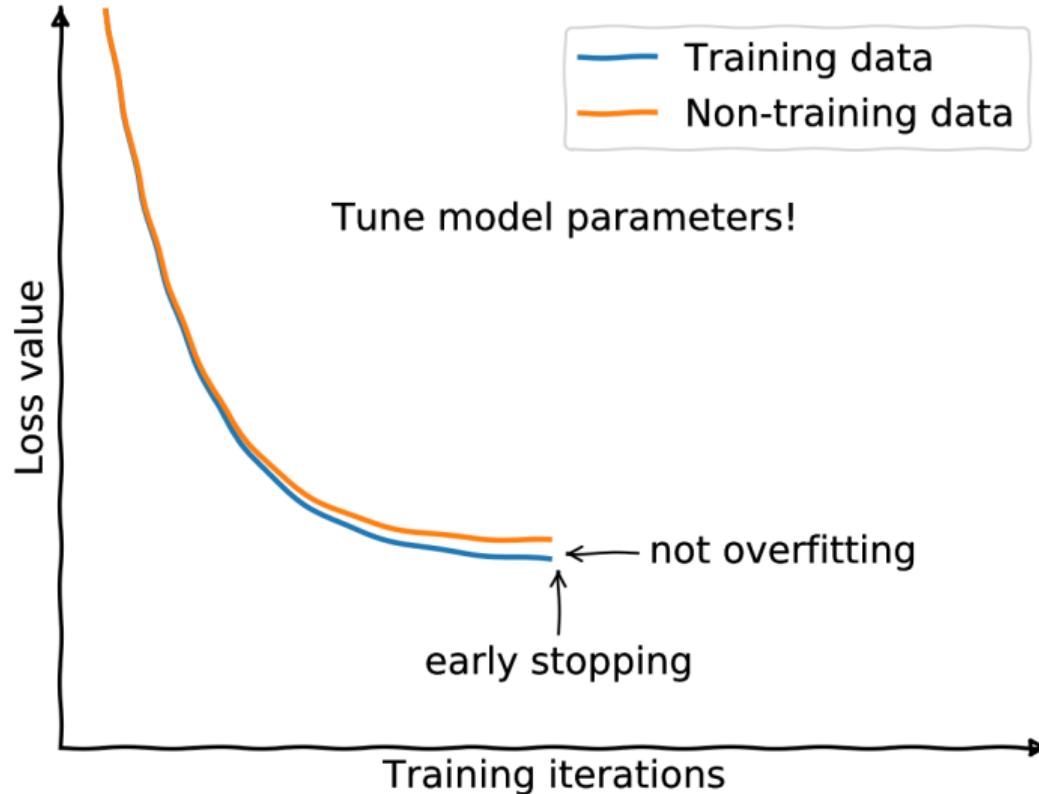


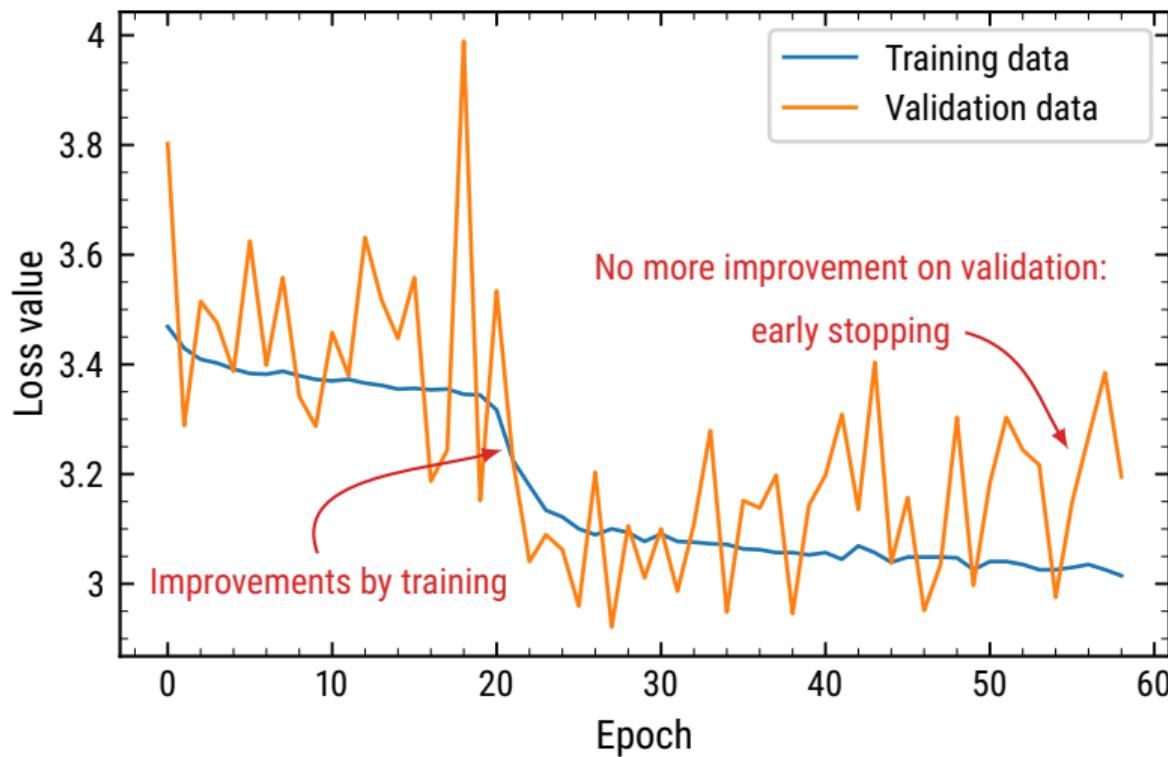












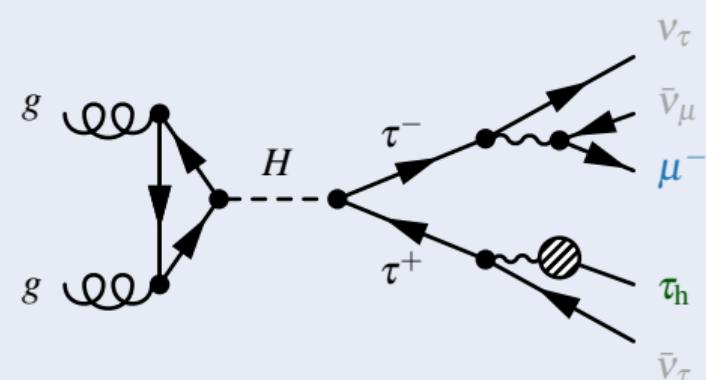
# Build a neural network: target and inputs?

- ▶ Model target: generated Higgs mass.

- ▶ Model inputs:

- ▶  $\tau_1$  (here =  $\mu^-$ ) and  $\tau_2$  (here =  $\tau_h$ )  $p_T, \eta, \phi$ ;
- ▶ PuppiMET  $p_T, \phi$ ;
- ▶ METcov xx, xy and yy;
- ▶ Number of neutrinos from reco tau decays;
- ▶  $m_T^{(1,MET)}, m_T^{(2,MET)}, m_T^{(1,2)}, m_T^{\text{tot}}$  (Puppi);
- ▶ jet 1, jet 2  $p_T, \eta, \phi$ ;
- ▶ Additionnal Hadronic Activity  $p_T, \eta, \phi, N_{\text{jets}}^{\text{AHA}}$ ;
- ▶ npvsGood  $\rightarrow$  how much PU.

$$m_T^{\text{tot}} = \sqrt{m_T^2(\tau_1, E_T^{\text{miss}}) + m_T^2(\tau_2, E_T^{\text{miss}}) + m_T^2(\tau_1, \tau_2)} , \quad m_T(1,2) = \sqrt{2 p_T^{(1)} p_T^{(2)} (1 - \cos \Delta\phi)}$$



# Build a neural network: hyperparameters?

- ▶ NN hyperparameters (and other tested values):

- ▷ **Adam** optimizer (Adadelta, SGD),
- ▷ Weight initialized with **Glorot uniform** (Glorot normal, normal, uniform),
- ▷ Custom  $\mathcal{L}_{\text{MA}\sqrt{\text{PE}} \times b}$  loss ( $\mathcal{L}_{\text{MAPE}}$ ,  $\mathcal{L}_{\text{MAE}}$ ,  $\mathcal{L}_{\text{MSE}}$ ),
- ▷ **Softplus** activation function (ReLU, ELU, SELU, Exponential),

$$\text{softplus}(x) = \ln(1 + e^x)$$

- ▷ **3** hidden layers (2 to 5),
- ▷ **1000** neurons per hidden layer (200 to 2000 per steps of 100).

# Datasets?

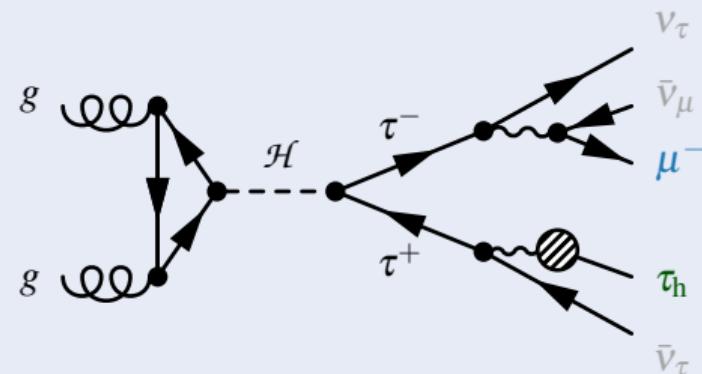
► Generate  $\mathcal{H} \rightarrow \tau\tau$  events:

- ▷  $\mathcal{H}$  is SM Higgs with a different mass,
- ▷  $\mathcal{H}$  produced by gluon fusion,
- ▷ set  $\mathcal{BR}(\mathcal{H} \rightarrow \tau\tau) = 1$  to avoid non di- $\tau$  events.

► All final states used simultaneously for training:

- ▷  $\tau_h\tau_h, \mu\tau_h, e\tau_h, \mu\mu, e\mu, ee$ .

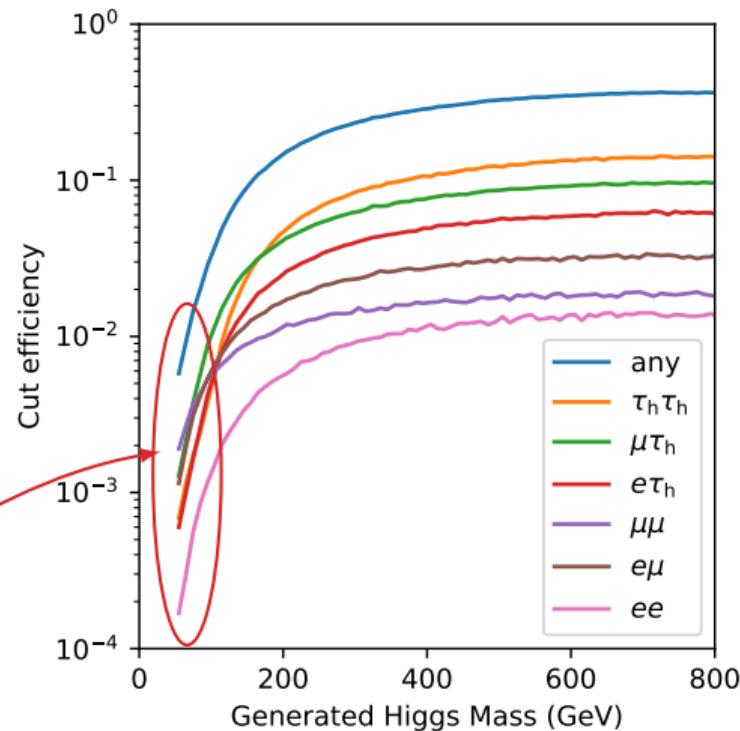
$$gg \rightarrow \mathcal{H} \rightarrow \tau\tau \rightarrow \mu\tau_h$$



# Datasets: low mass boundary

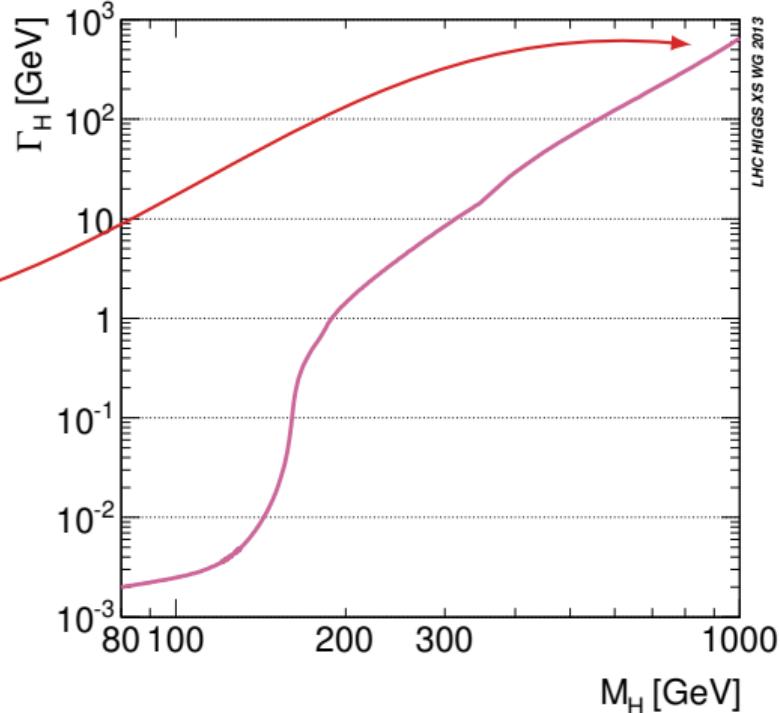
- ▶ Event selection:
  - ▷ Same as in the MSSM  $H/A \rightarrow \tau\tau$  analysis,
  - ▷ Add  $\mu\mu$  and  $ee$  channels.
- ▶ Events amount drops at low mass:
  - ▷ due to  $p_T$  cuts,
  - ▷ go down to 50 GeV for  $m_H$ .

Less than 1 % pass the selection



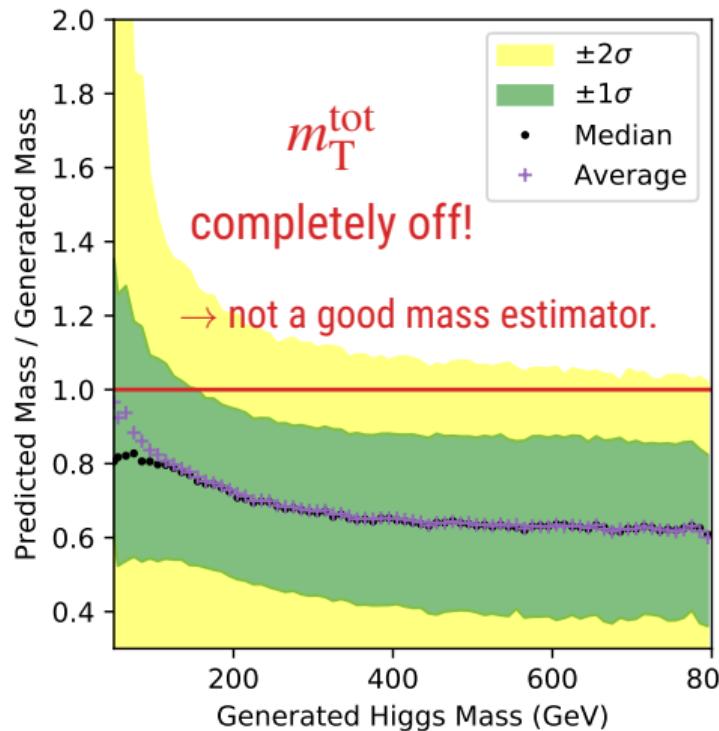
# Datasets: high mass boundary

- ▶ Higgs width  $\simeq$  Higgs mass around 1 TeV:
  - ▷ can't have coherent mass points,
  - ▷ go up to 800 GeV for  $m_H$ .



- ▷ LHC Higgs Cross Section Working Group. "Higgs Properties". *Handbook of LHC Higgs Cross Sections*. 3. CERN Yellow Reports: Monographs. Geneva: CERN, 2013. URL: <https://cds.cern.ch/record/1559921>.

# DNN's $m_{\text{ML}}$ predictions vs $m_{\text{T}}^{\text{tot}}$

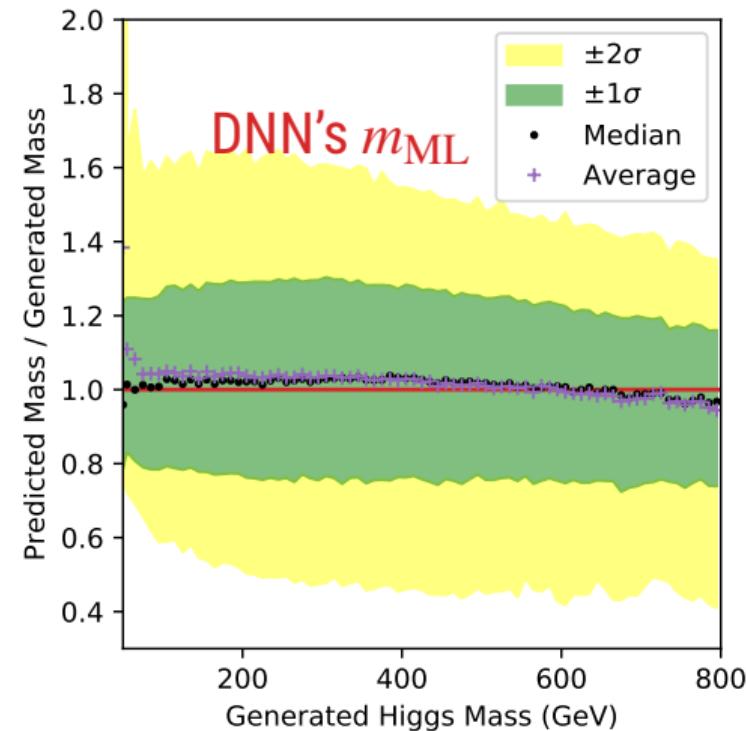
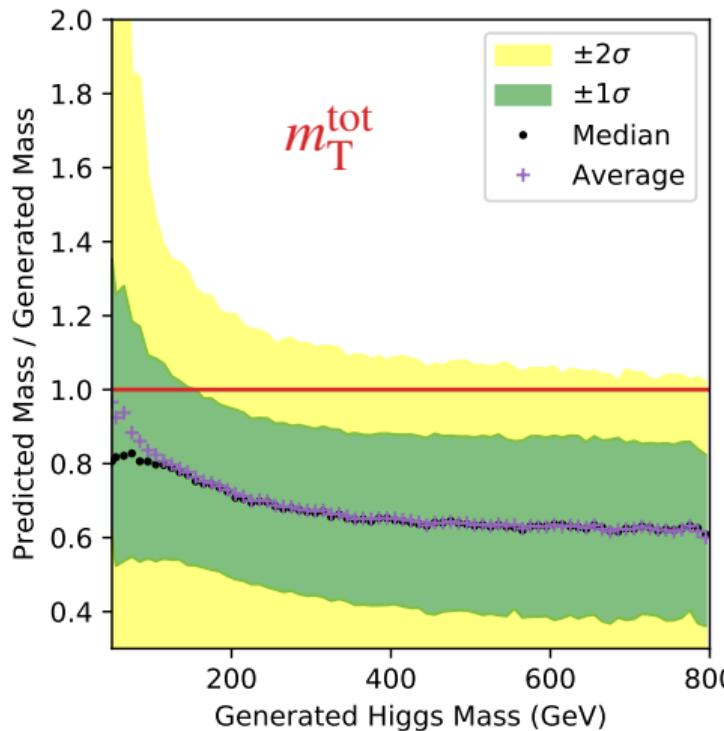


► Model's response:

$$r = \frac{\text{prediction}}{\text{true value}} = \frac{m_{\text{ML}}}{m_{\mathcal{H}}} \text{ or } \frac{m_{\text{T}}^{\text{tot}}}{m_{\mathcal{H}}}$$

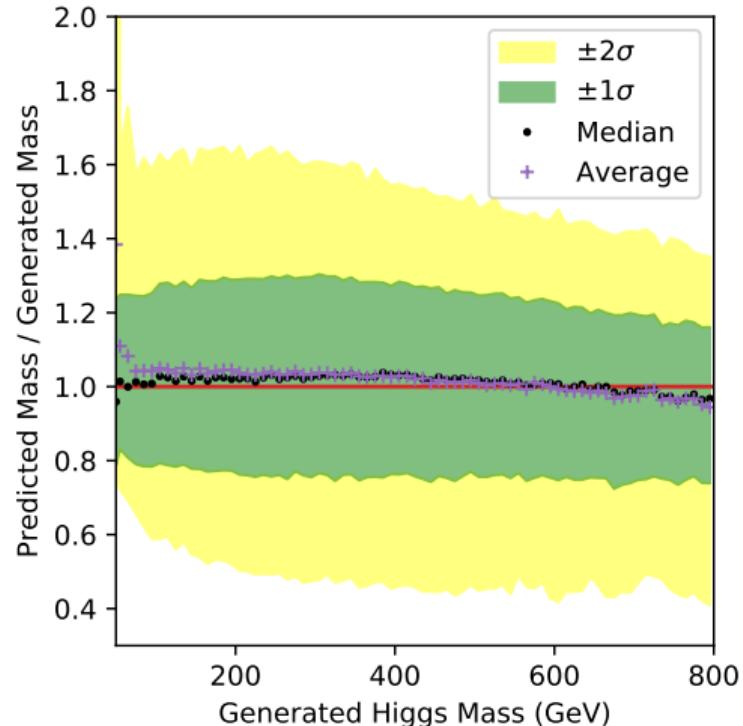
► Closer to 1 is better for black dots.

# DNN's $m_{\text{ML}}$ predictions vs $m_{\text{T}}^{\text{tot}}$



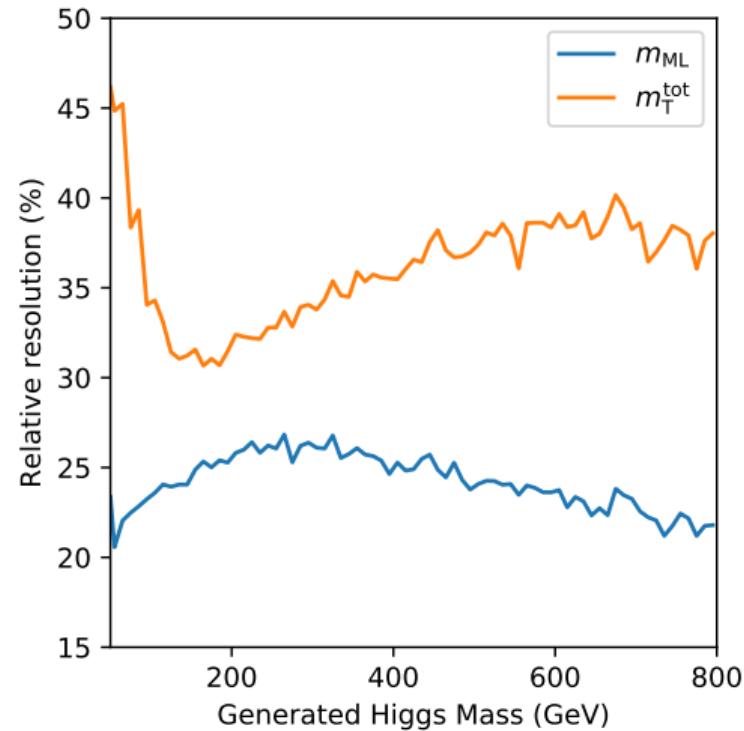
# DNN's $m_{\text{ML}}$ predictions vs $m_{\text{T}}^{\text{tot}}$

- $r = 1.00 \pm 0.05$  from 80 to 800 GeV
- $\mathcal{H}$  mass reconstruction achieved ✓

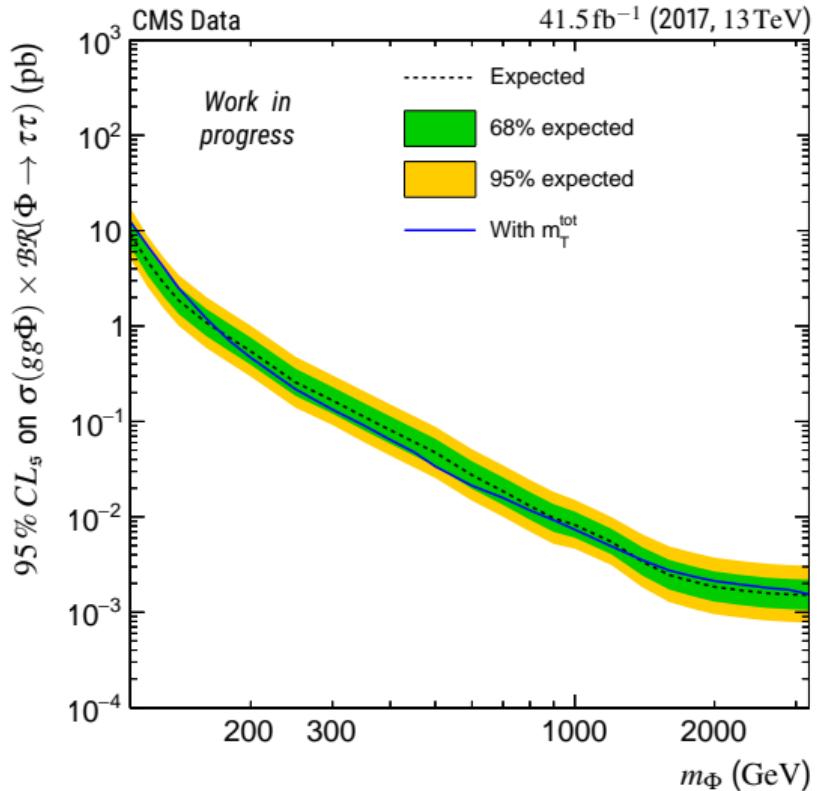


# Using the model to get a discriminating variable

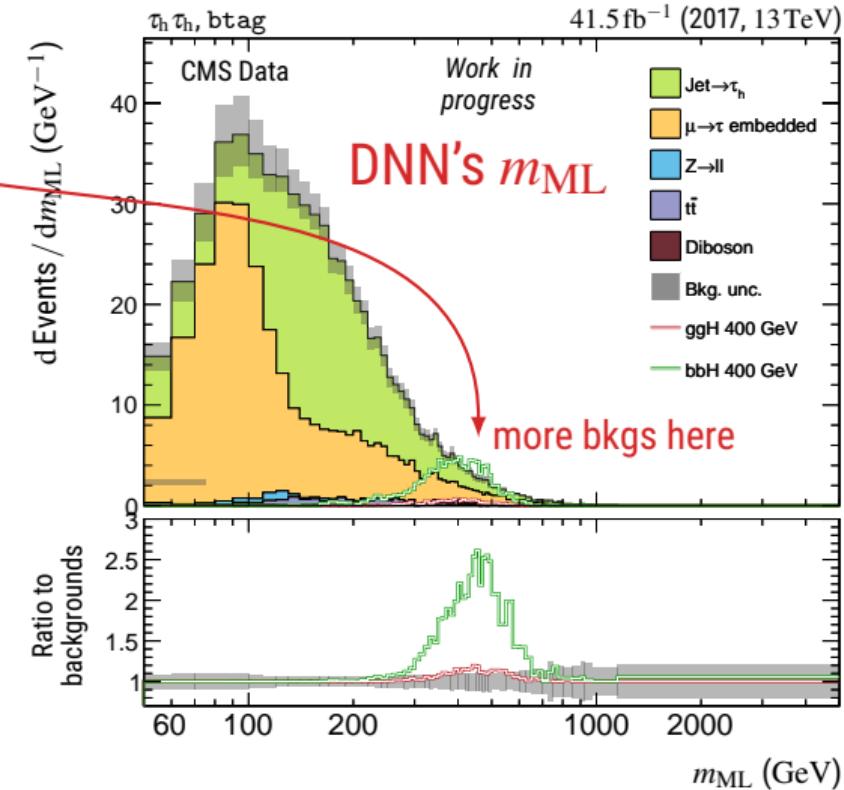
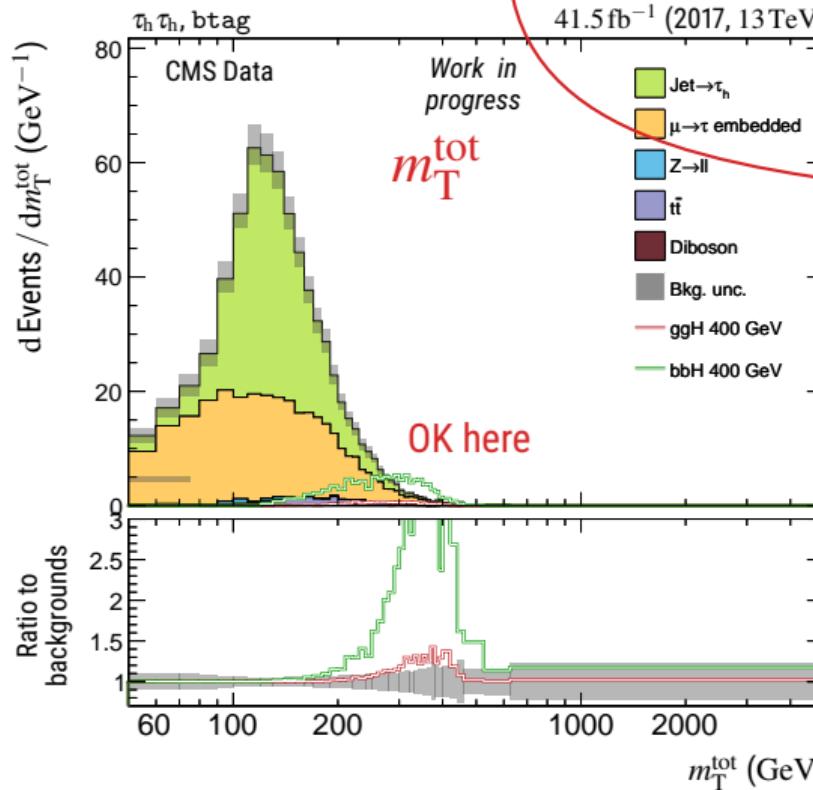
- ▶ In the  $H/A \rightarrow \tau\tau$  analysis, discriminating variable =  $m_T^{\text{tot}}$ .
- ▶  $m_T^{\text{tot}}$  is equal to the invariant mass assuming:
  - ▷ all neutrinos are a single particle with  $\vec{p}_T = \vec{E}_T^{\text{miss}}$ ,
  - ▷ all is going on in the transverse plane (any  $p_z = 0$ ).
- ▶ Our model has a better resolution on  $m_H$  than  $m_T^{\text{tot}}$ .



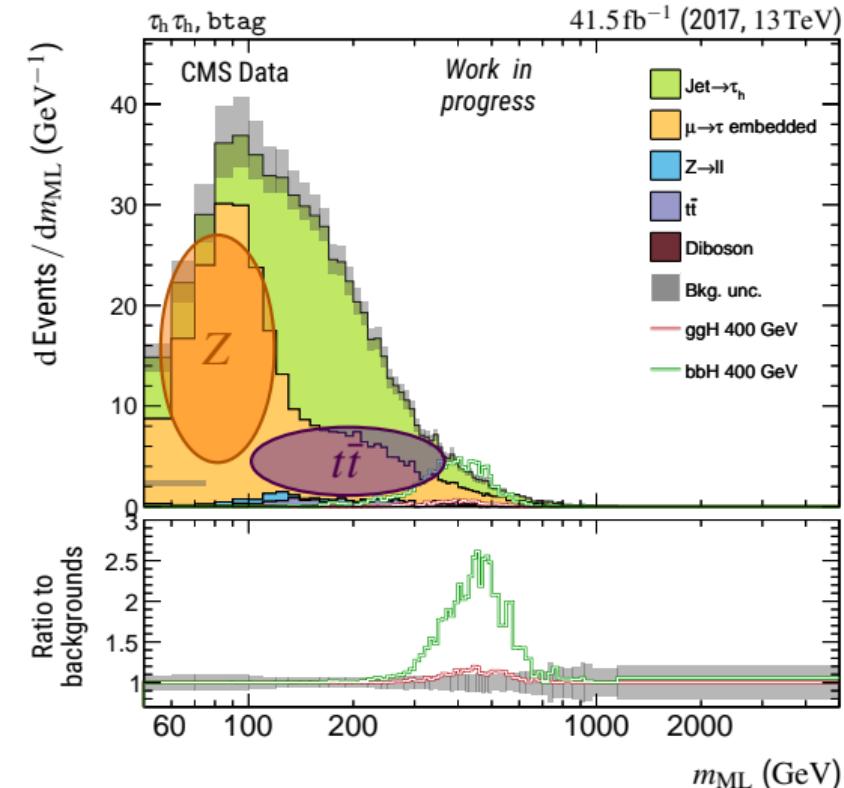
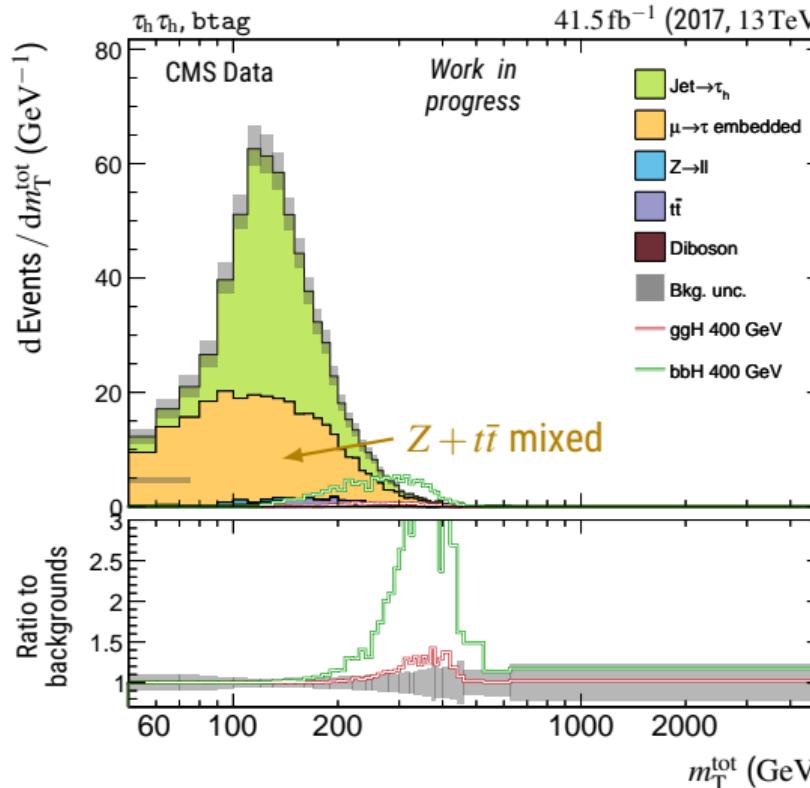
- ▶ Proceed to the search for massive Higgs boson  $\Phi$  with di- $\tau$  events on the 2017 era.
- ▶ Use  $m_{\text{ML}}$  as discriminating variable.
- ▶ Not really better than with  $m_T^{\text{tot}}$ ... Why?



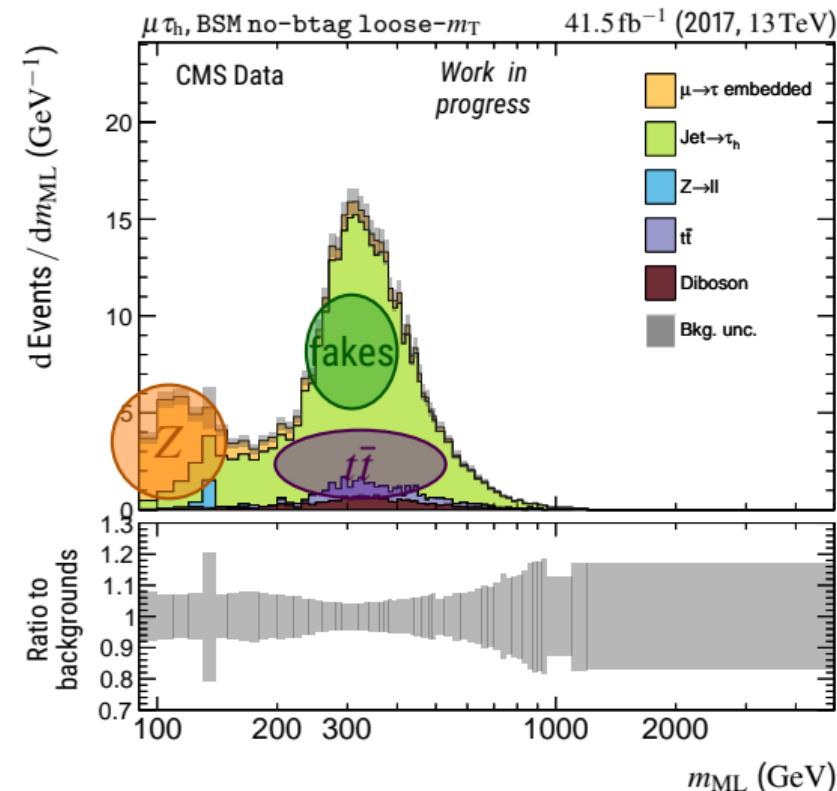
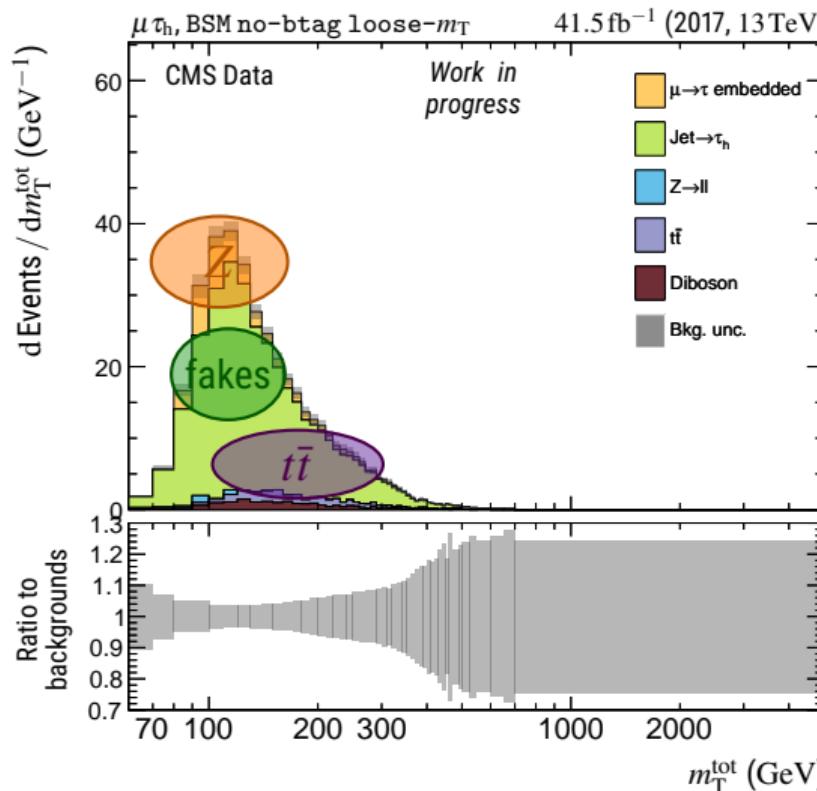
► Large fakes  $\tau_h$  high mass tails falling into the signal region  $\Rightarrow$  lowered signal to background ratio.



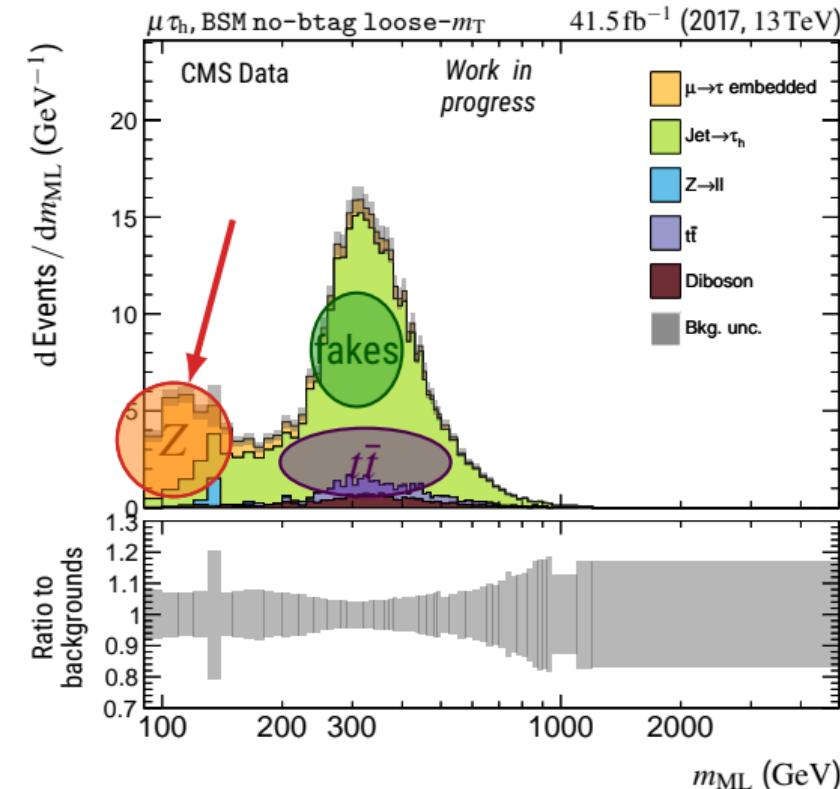
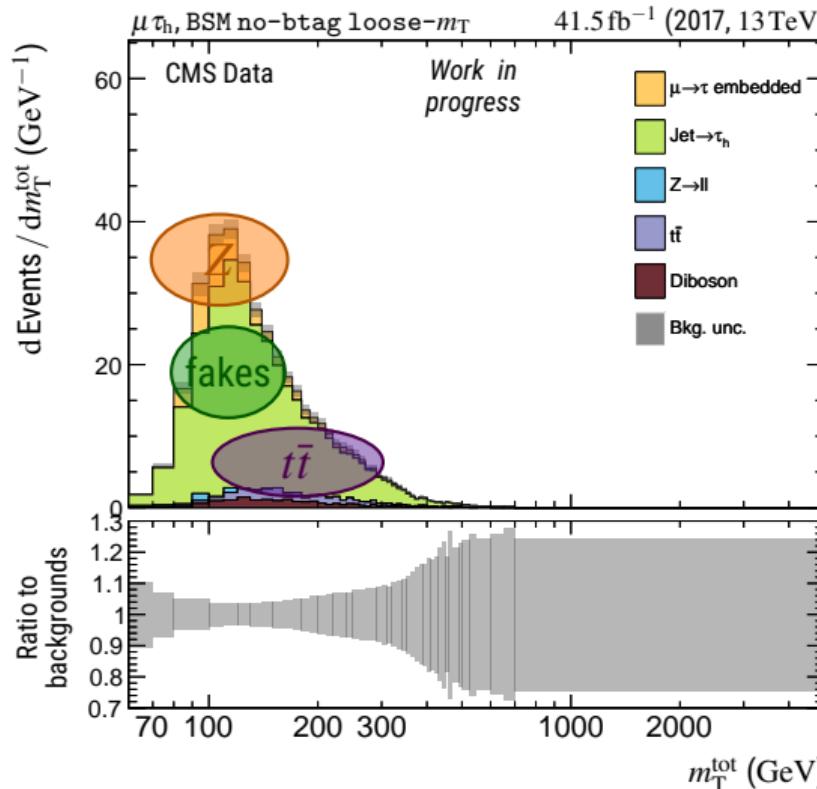
► BUT  $Z / t\bar{t}$  separation! See the two  $m_{ML}$  embedded components, not present with  $m_T^{\text{tot}}$ .



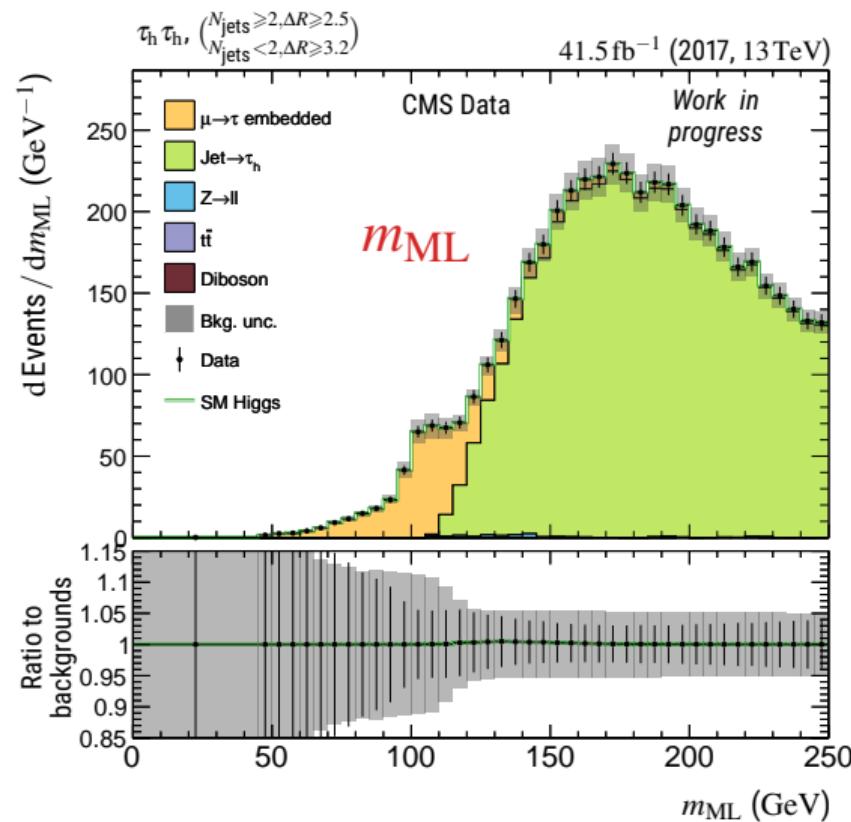
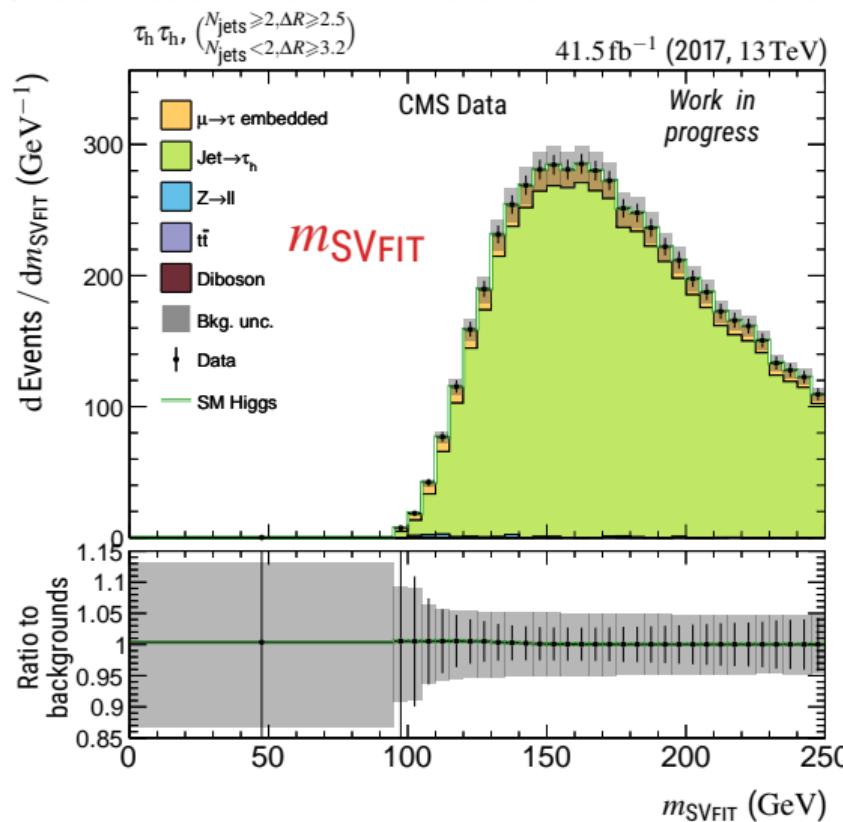
► Visible in other categories too. But here, the SVFIT mass is above 250 GeV and our model sees a  $Z$  signal!



► Visible in other categories too. But here, the SVFit mass is above 250 GeV and our model sees a  $Z$  signal!

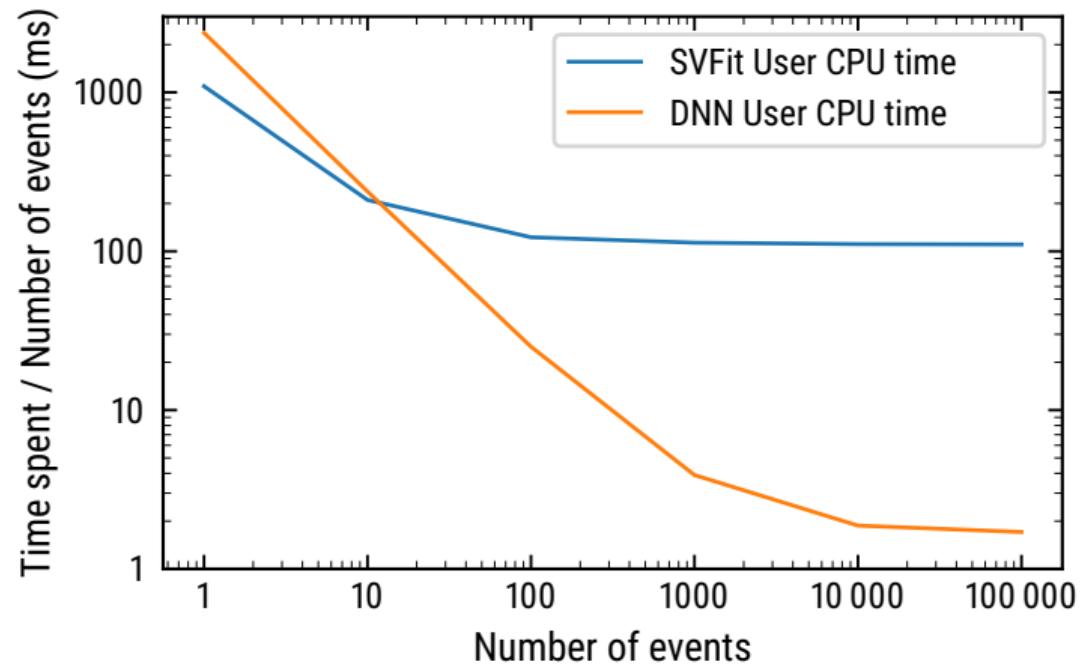


► Our model finds  $Z \rightarrow \tau\tau$  events when SVFit does not!



- ▶ Computing time: DNN (Python) is  $\sim 60\times$  faster than SVFIT (C++)!

- ▶ SVFIT:
  - ▷ fit to find the best mass
  - ▷ for each event
- ▶ DNN:
  - ▷ fit done once (training)
  - ▷ apply the DNN formula



# What to conclude from this thesis?

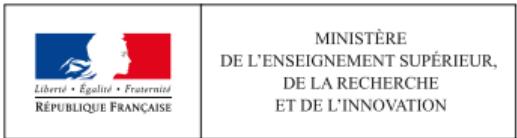
# Conclusion & prospects: $H/A \rightarrow \tau\tau$

- ▶ MSSM  $H/A \rightarrow \tau\tau$  analysis on full Run II:
  - ▷ 4 final states:  $\tau_h\tau_h$ ,  $\mu\tau_h$ ,  $e\tau_h$  and  $e\mu$ ,
  - ▷ Model independent exclusion limits on  $\sigma \times \mathcal{BR}$ ,
  - ▷ Model dependent exclusion contours in the  $(m_A, \tan \beta)$  plane.
- ▶ CMS paper HIG-21-001 on its way for publication:
  - ▷ Leading-edge until Run III corresponding results!
- ▶ No evidence for MSSM.

# Conclusion & prospects: ML project

- ▶ Successful  $m_{\mathcal{H}}$  reconstruction in di- $\tau$  events.
  - ▷ Not only MSSM  $H/A \rightarrow \tau\tau$  but any  $X \rightarrow \tau\tau$  analysis could benefit.
- ▶  $m_{\text{ML}}$  vs  $m_{\text{T}}^{\text{tot}}$ :
  - ▷ A good mass estimator is not equivalent to a good discriminating variable.
  - ▷ Still, we already have the same performances at this point.
- ▶  $m_{\text{ML}}$  vs  $m_{\text{SVFIT}}$ :
  - ▷ Similar Higgs sensitivity for some event topologies.
  - ▷ Better  $Z$  estimation observed (the model has been trained on  $\mathcal{H} \rightarrow \tau\tau$  with various masses only).
  - ▷ Could be improved by updating the training datasets (other kinds of events).
  - ▷ Faster (about 60 times!).
  - ▷ Very promising as a SVFIT successor.

# Merci !



- ▷ Colin
- ▷ Gaël, Ece
- ▷ Hugues
- ▷ Aurélien, Antoine L.
- ▷ Jean-François, Grégoire
- ▷ Corentin, Martin
- ▷ autres doctorants
- ▷ groupe CMS
- ▷ Antoine C.
- ▷ personnels



- ▷ Günter, Roger
- ▷ Artur
- ▷ Sebastian B., Maximilian
- ▷ Sebastian W.
- ▷ Janek
- ▷ Felix



- ▷ Daniel
- ▷ David
- ▷ Georges



- ▷ équipe des guides
- ▷ Jacob
- ▷ Jean
- ▷ Giuseppe
- ▷ Juska
- ▷ Yi
- ▷ Mikko



- ▷ Aleksei
- ▷ Mareike



- ▷ Janik
- ▷ Suman

► Phenomenology:

- ▷ SM, SUSY, 2HDM,  $\tan \beta$  [► slide 53](#)
- ▷ How histograms unveil particles [► slide 58](#)

► JERC with  $\gamma + \text{jets}$  events:

- ▷ Jet energy calibration [► slide 64](#)
- ▷ Jet energy resolution [► slide 72](#)

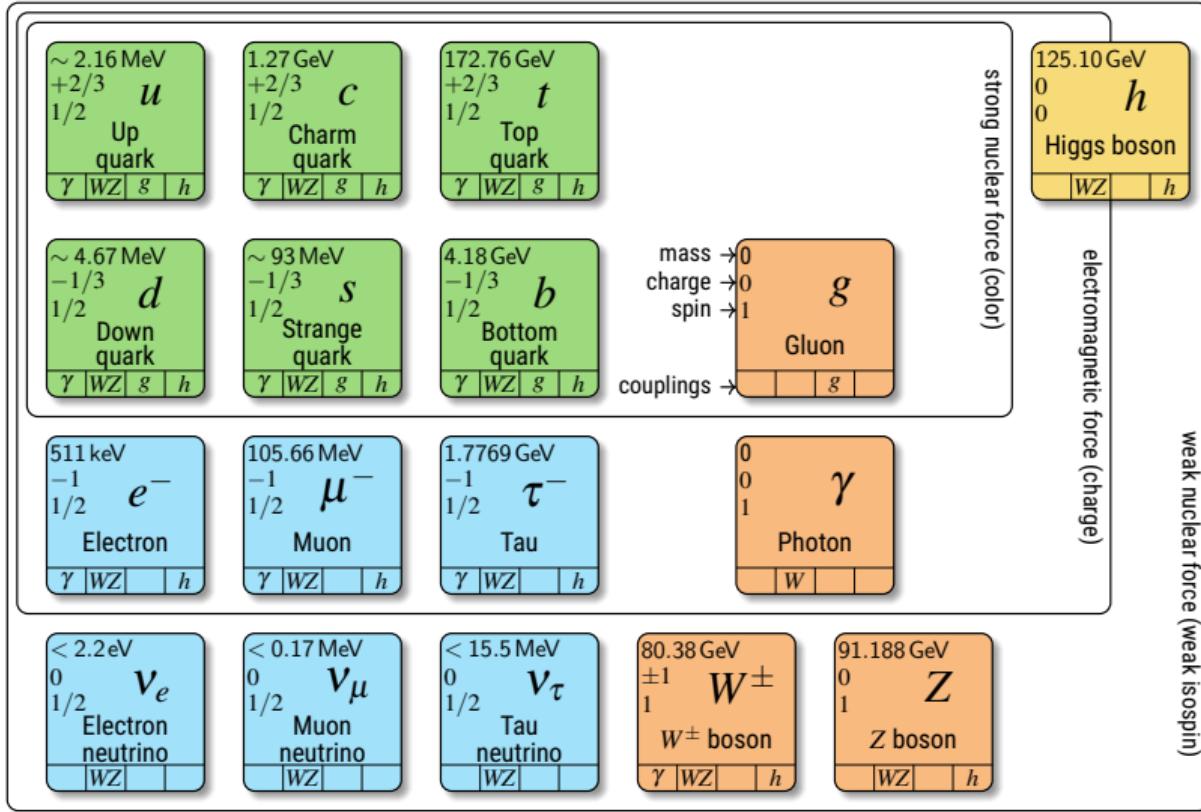
► MSSM  $H/A \rightarrow \tau\tau$ :

- ▷ Triggers in the  $\tau_h \tau_h$  channel [► slide 73](#)
- ▷ Fake factors for subleading  $\tau_h$  [► slide 74](#)

► Machine Learning:

- ▷ Custom loss function for the DNN [► slide 75](#)

# The Standard Model



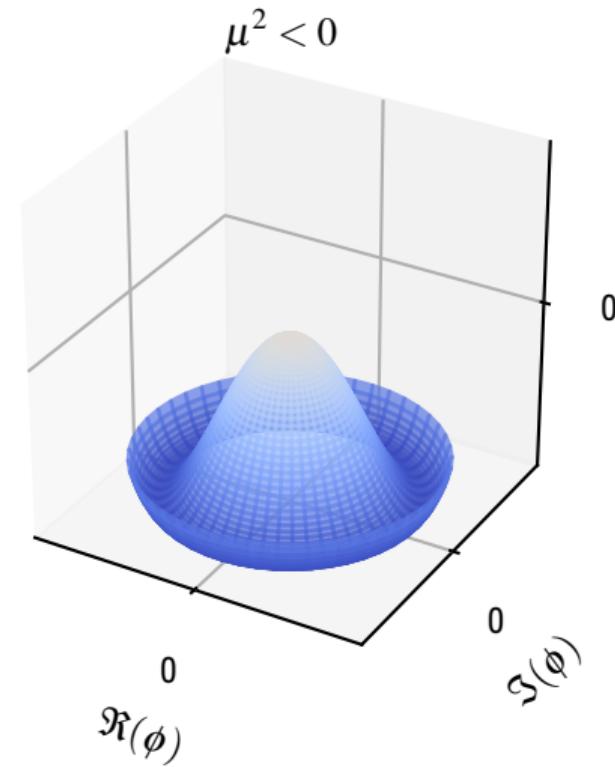
# Higgs boson in the Standard Model

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_3 + i\phi_4 \\ \phi_1 + i\phi_2 \end{pmatrix}$$

$$V(\phi) = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2, \quad \lambda > 0$$

$$\langle \phi \rangle_0 = \frac{v}{\sqrt{2}} = \sqrt{\frac{-\mu^2}{2\lambda}} \neq 0$$

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}$$

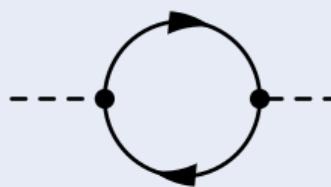


# The Standard Model and naturalness problem

► Higgs mass measured:  $m_h = 125.10 \pm 0.14 \text{ GeV}$

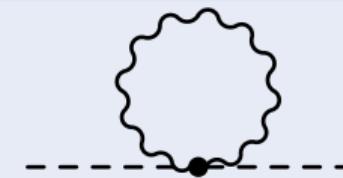
► Higgs mass derivation:  $m_h^2 = m_{h0}^2 - \frac{3}{8\pi^2} y_t^2 \Lambda^2 + \frac{1}{16\pi^2} g^2 \Lambda^2 + \frac{1}{16\pi^2} \lambda^2 \Lambda^2 + \dots$

top quark



$$-\frac{3}{8\pi^2} y_t^2 \Lambda^2 \sim -(2 \text{ TeV})^2$$

vector bosons



$$+\frac{1}{16\pi^2} g^2 \Lambda^2 \sim +(0.7 \text{ TeV})^2$$

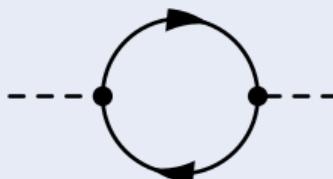
Higgs itself



$$+\frac{1}{16\pi^2} \lambda^2 \Lambda^2 \sim +(0.5 \text{ TeV})^2$$

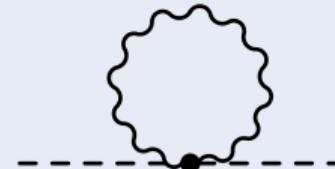
# Supersymmetry

top quark



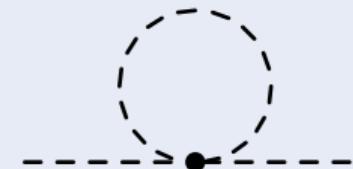
$$\sim -(2 \text{ TeV})^2$$

vector bosons



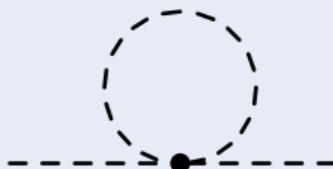
$$\sim +(0.7 \text{ TeV})^2$$

Higgs itself



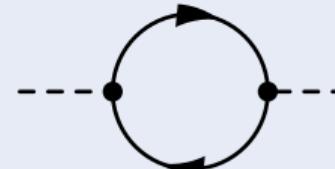
$$\sim +(0.5 \text{ TeV})^2$$

stop quark



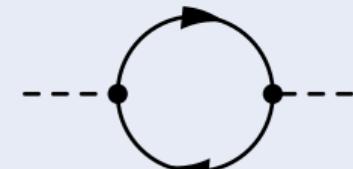
$$\sim +(2 \text{ TeV})^2$$

bosinos



$$\sim -(0.7 \text{ TeV})^2$$

Higgsinos



$$\sim -(0.5 \text{ TeV})^2$$

## 2 Higgs doublets models for supersymmetry

$$\langle \phi_1 \rangle_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v_1 \end{pmatrix}, \quad \langle \phi_2 \rangle_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v_2 e^{i\xi} \end{pmatrix}$$

$$\tan \beta = \frac{\langle \phi_2 \rangle_0}{\langle \phi_1 \rangle_0} = \frac{v_2}{v_1}$$

- ▷ J. F. Gunion et al. *The Higgs hunter's guide*. T. 80. Upton, NY: Brookhaven Nat. Lab., 1989. URL: <https://cds.cern.ch/record/425736>.

# Using histograms

- ▶ Find a discriminating variable:
  - ▷ for uncorrelated  $\tau$  pairs, it's random
  - ▷ for  $\tau$  pairs coming from a particle (Higgs?), not random.
- ▶ For one  $\tau$  pair only, impossible to say!
- ▶ With many events, a difference may show up.

# The rabbit analogy

- ▶ What the theorists say:
  - ▷ There is a white rabbit that once lived in a casino.
  - ▷ The rabbit loved watching people playing dices.
  - ▷ He was happy when the result of dice was 4.
  - ▷ So when he sees a dice, he turns it so that the result is 4.
  - ▷ But this rabbit is very shy and nobody has seen him since the casino closure.
- ▶ The only way to know if he's here is to throw a dice and come back to see the result.
  - ▷ If the rabbit has been here, the dice will show a 4!

# The rabbit analogy

- ▶ Dice results: 4

# The rabbit analogy

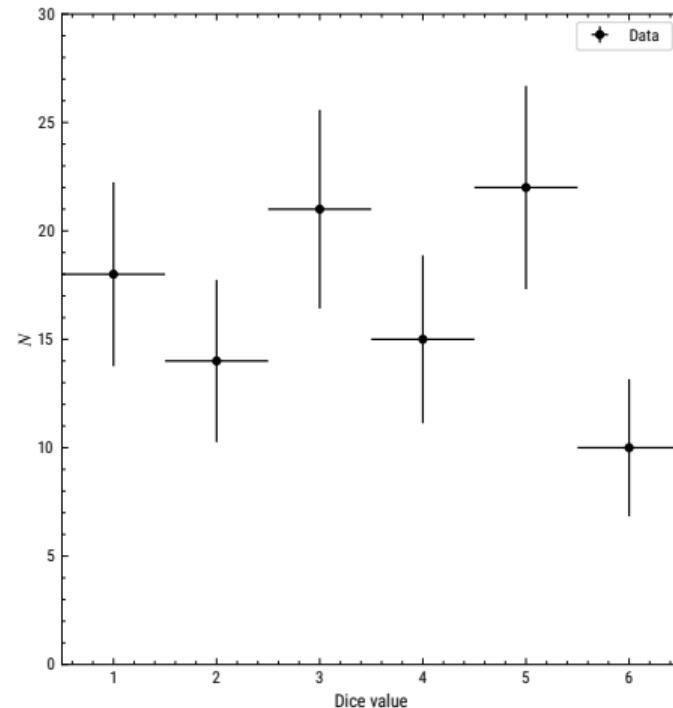
- ▶ Dice results: 4, 2

# The rabbit analogy

- ▶ Dice results: 4, 2, 4, 1, 3, 2, 5, 1, 1, 6...

# The rabbit analogy

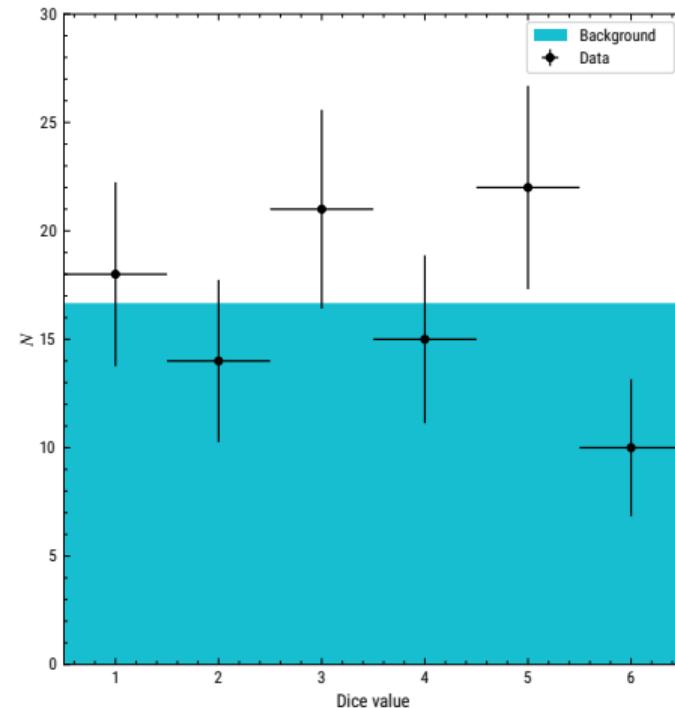
On 100 days →



Not really conclusive...

# The rabbit analogy

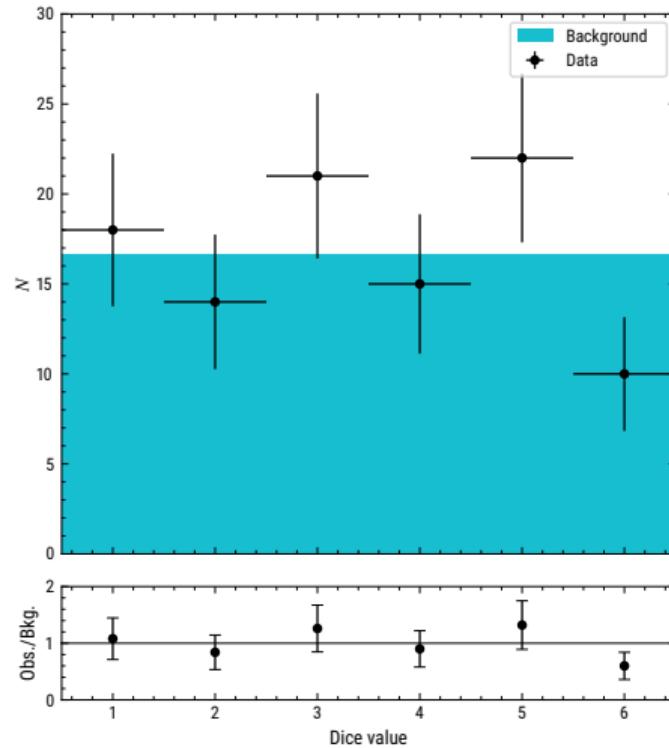
On 100 days →



Comparing with predictions!

# The rabbit analogy

On 100 days →

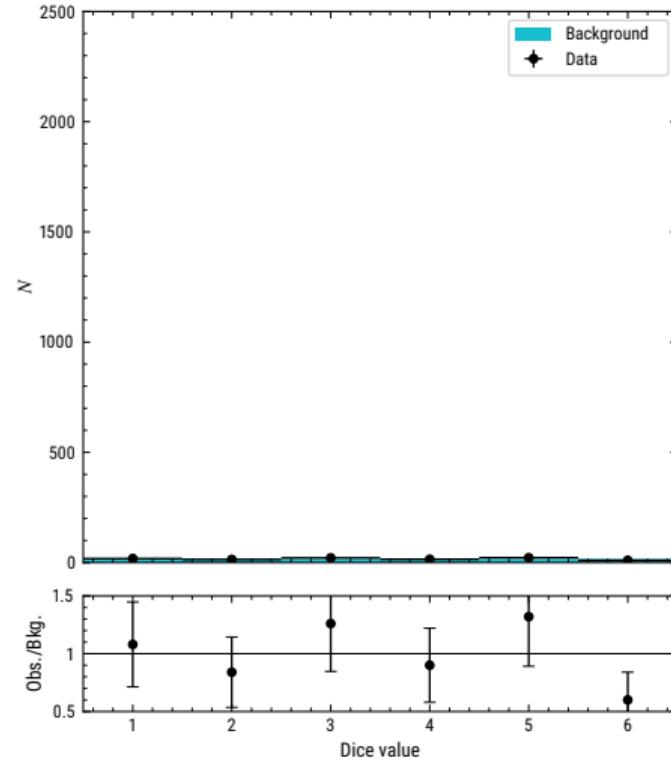


Also add ratio plot:  
observed / predictions

# The rabbit analogy

On 100 days →

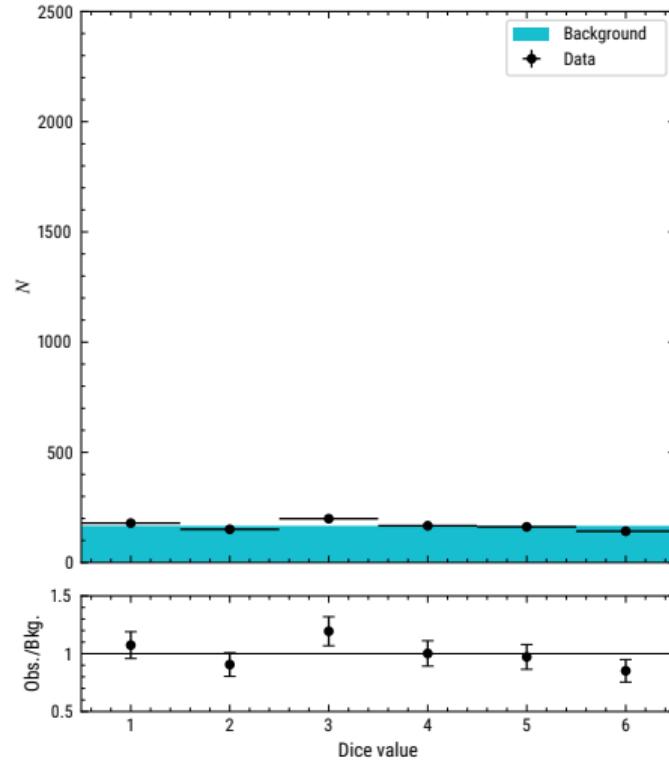
Fill up with more data!



# The rabbit analogy

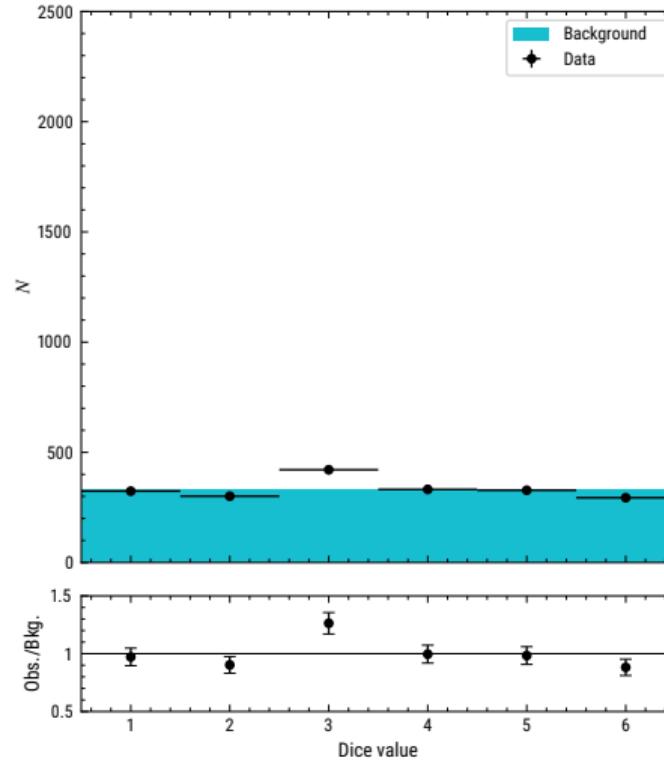
On 1000 days →

Fill up with more data!



# The rabbit analogy

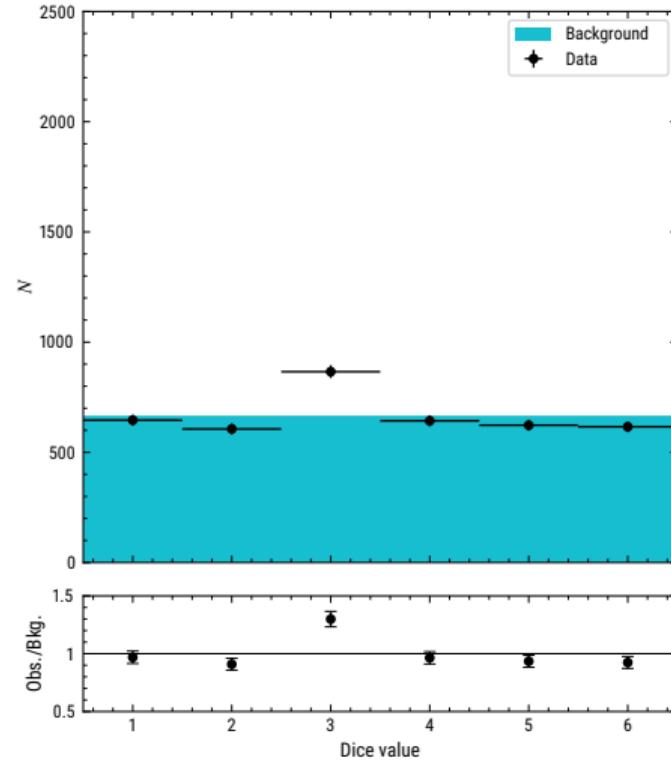
On 2000 days →



Fill up with more data!

# The rabbit analogy

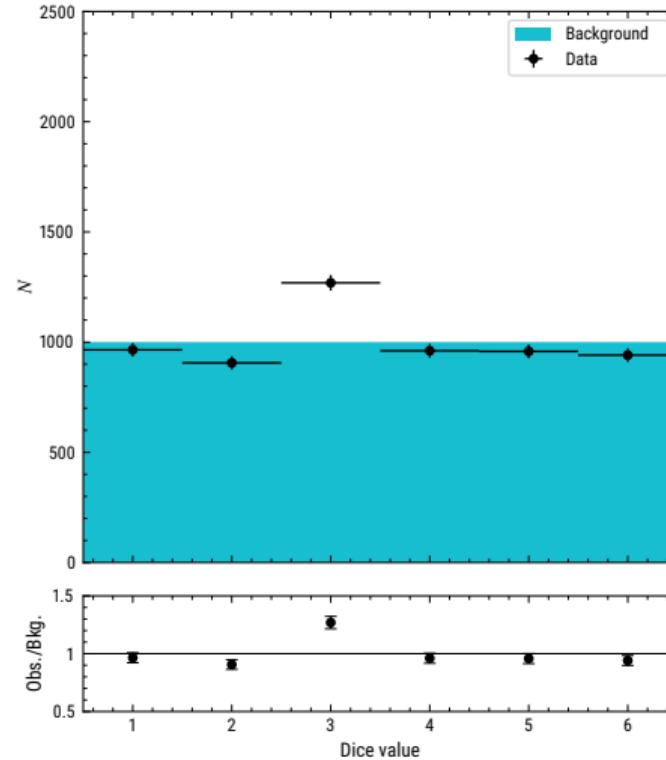
On 4000 days →



Fill up with more data!

# The rabbit analogy

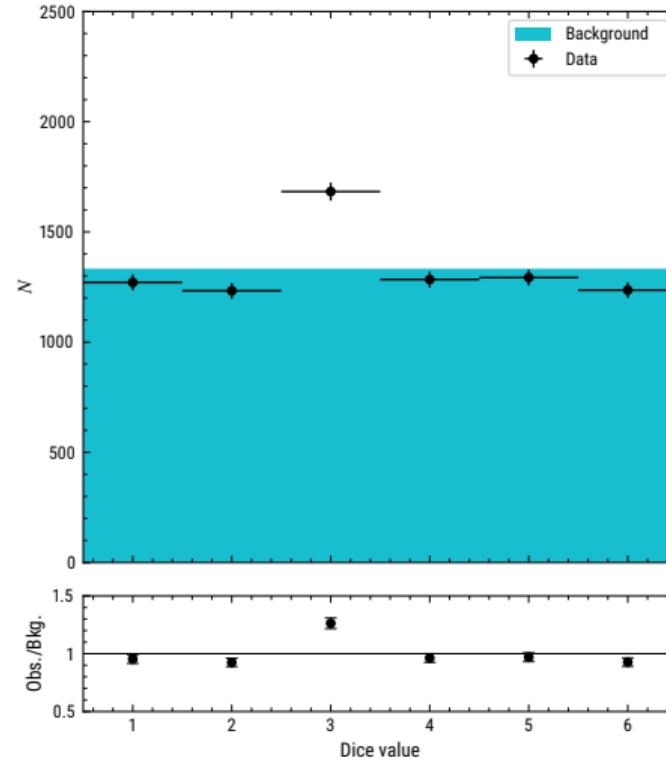
On 6000 days →



Fill up with more data!

# The rabbit analogy

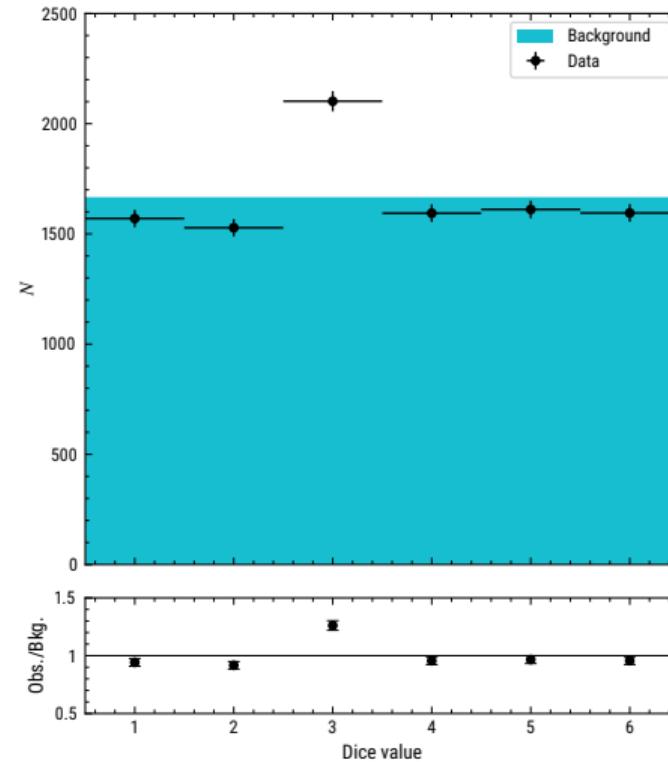
On 8000 days →



Fill up with more data!

# The rabbit analogy

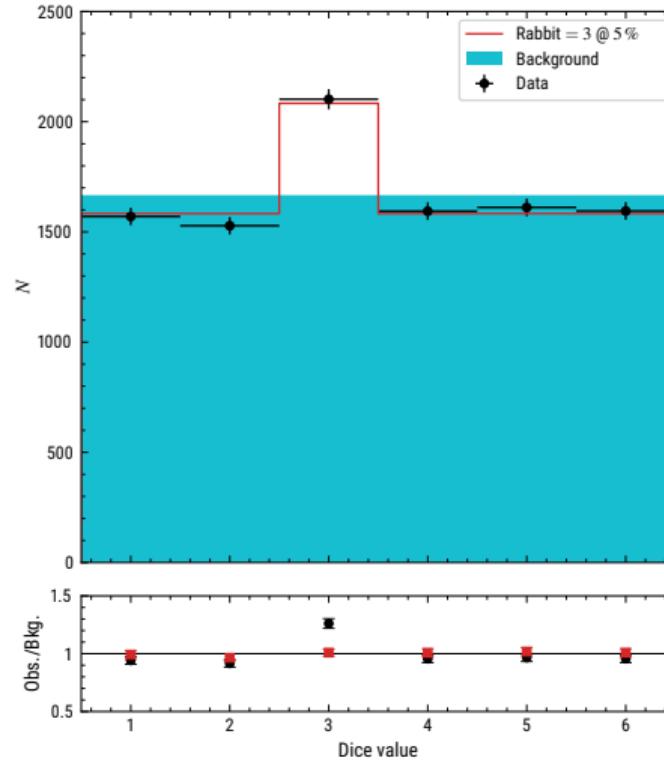
On 10 000 days →



Fill up with more data!

# The rabbit analogy

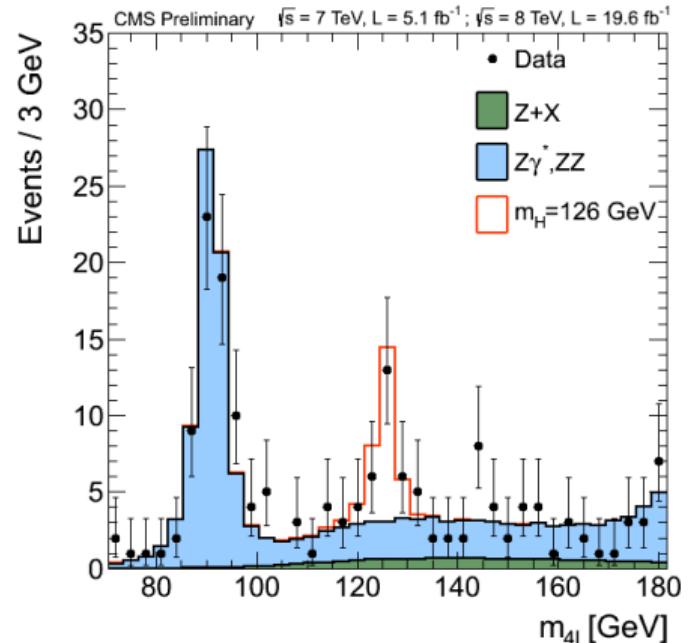
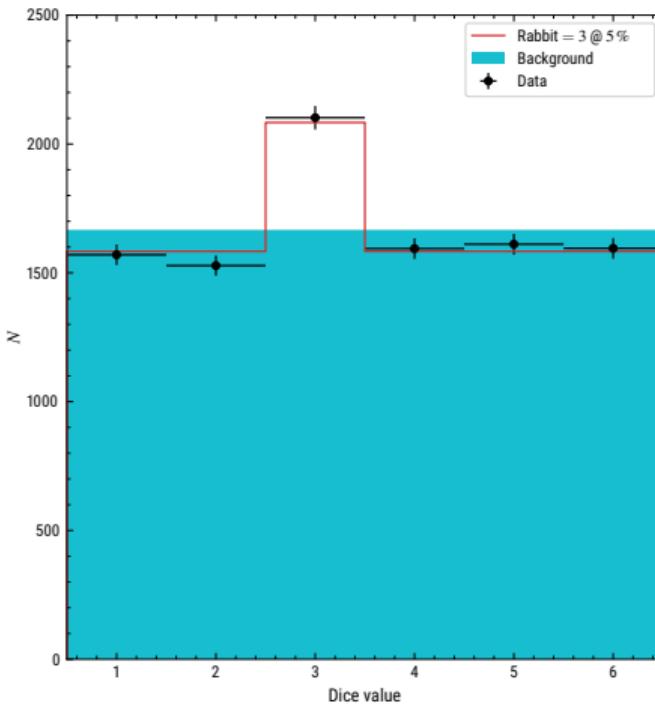
On 10 000 days →



In red, hypothesis of the rabbit with 3 as preferred result (instead of 4!), with a probability to show up of 5 %.

# The rabbit analogy

Search for	the rabbit	the Higgs
Observed data	dice values	$pp$ collisions outgoing particles
Discriminating variable	dice value	invariant or transverse mass
Backgrounds predictions	random dice	Standard Model processes
Amount of data	number of days	luminosity
Signal probability	rabbit's shyness	process cross-section



- ▷ The CMS Collaboration. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC". *Physics Letters B* **716.1** (2012), pp. 30–61. DOI: [10.1016/j.physletb.2012.08.021](https://doi.org/10.1016/j.physletb.2012.08.021).
- ▷ The CMS Collaboration. *Properties of the Higgs-like boson in the decay  $H \rightarrow ZZ \rightarrow 4\ell$  in  $pp$  collisions at  $\sqrt{s} = 7$  and 8 TeV*. URL: <https://twiki.cern.ch/twiki/bin/view/CMSPublic/Hig13002TWiki>.

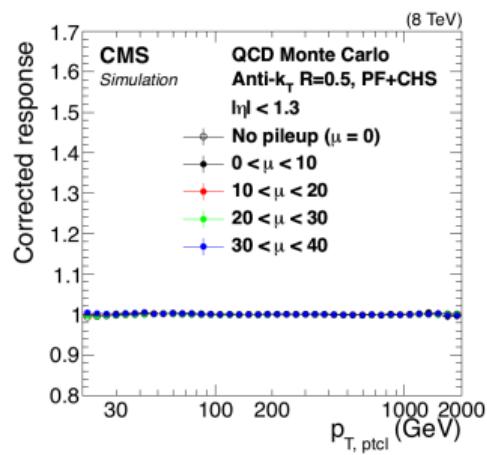
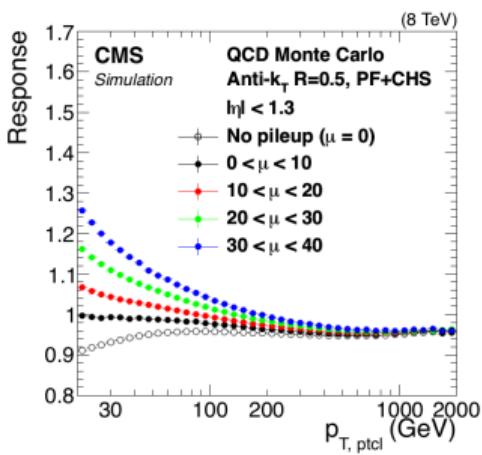
► Levels of knowledge

particle	(ptcl)
reconstructed	(reco)
corrected	(corr)

► Jet response

$$R = \frac{p_T}{p_{T\text{ptcl}}}$$

## Reconstructed Jets

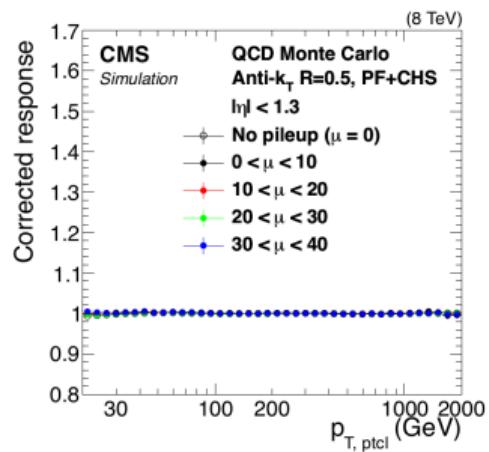
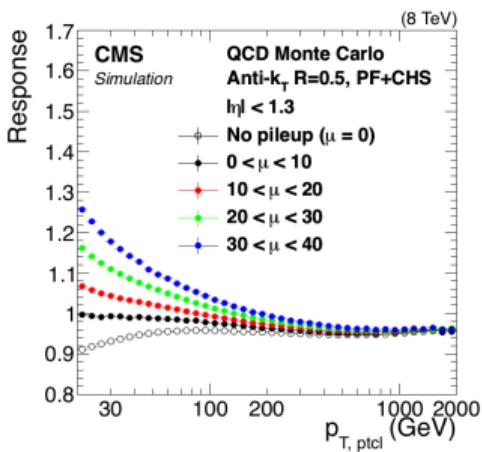


Applied to data

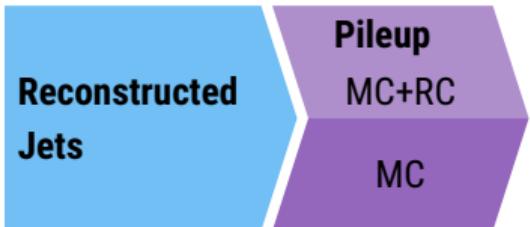
Reconstructed  
Jets

Calibrated  
Jets

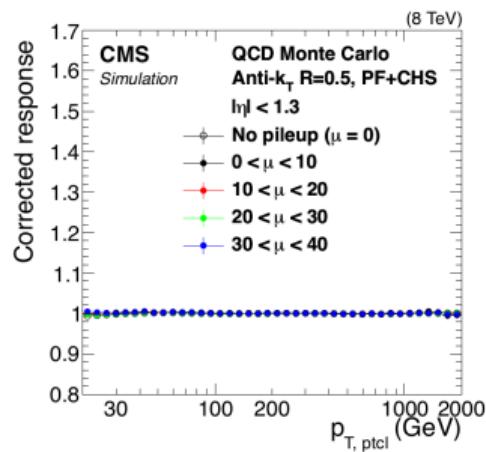
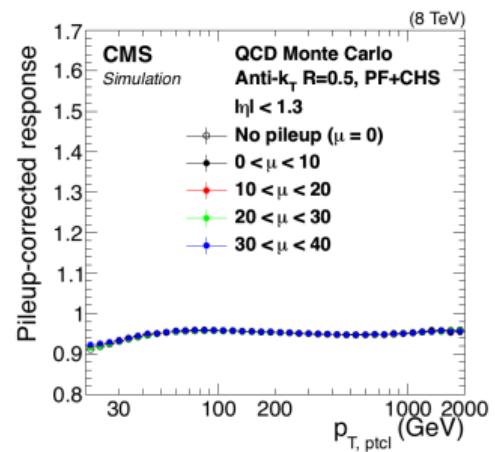
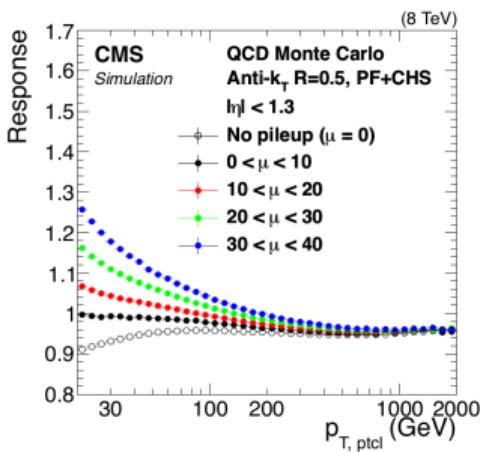
Applied to simulation



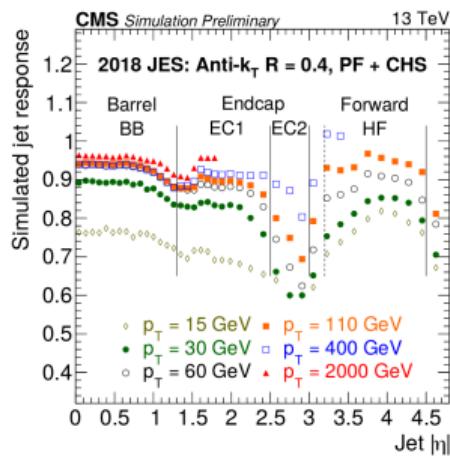
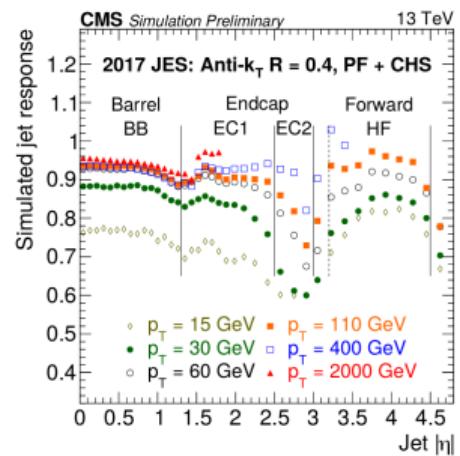
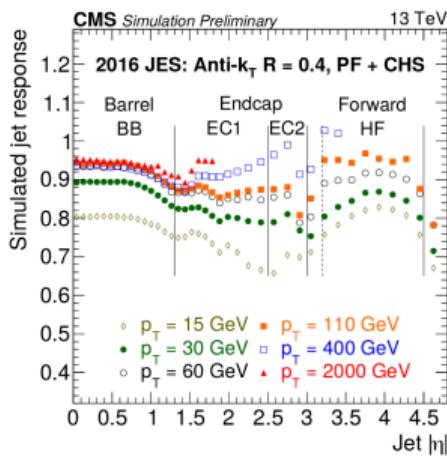
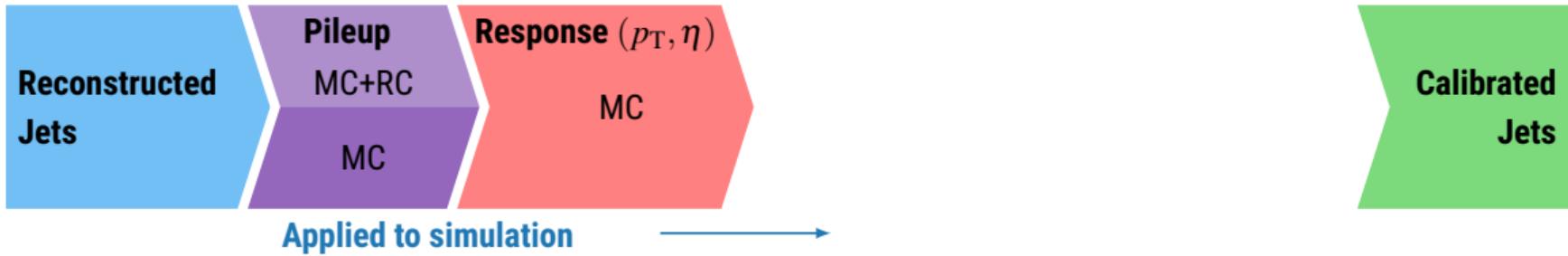
Applied to data

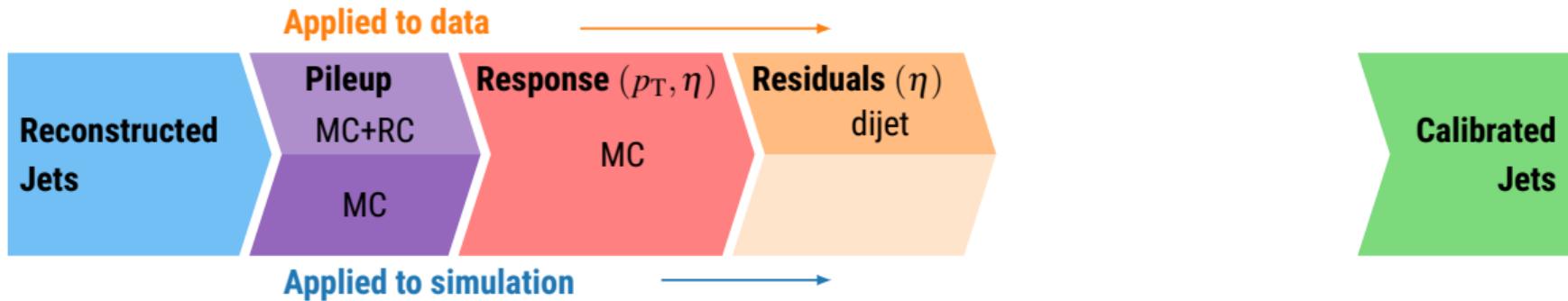


Applied to simulation

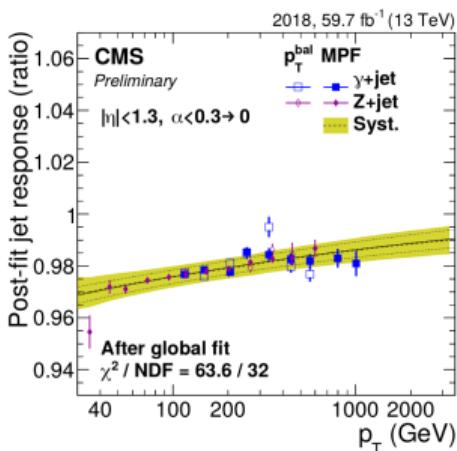
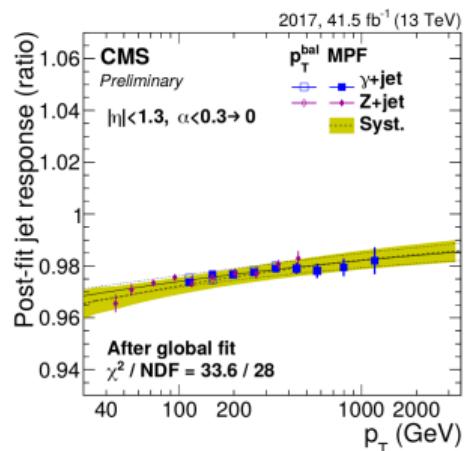
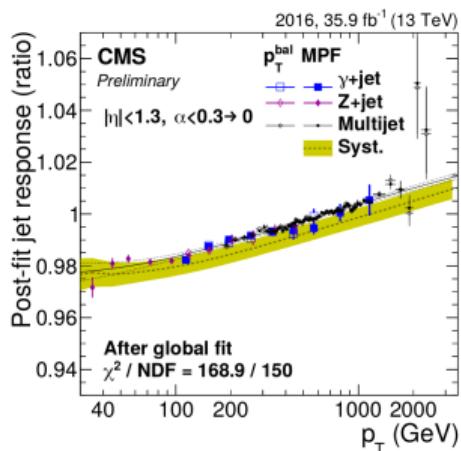
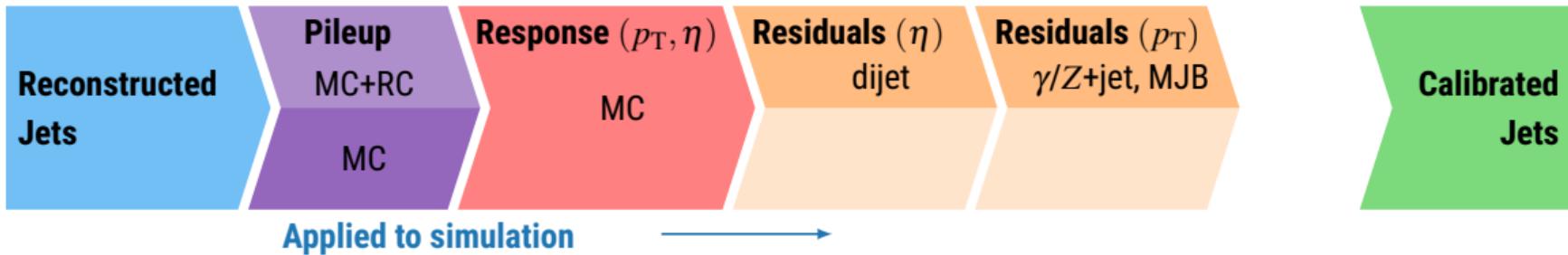


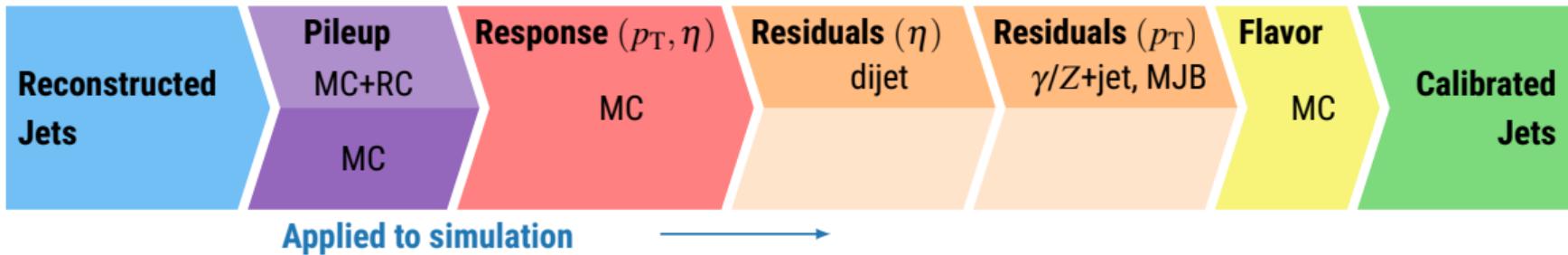
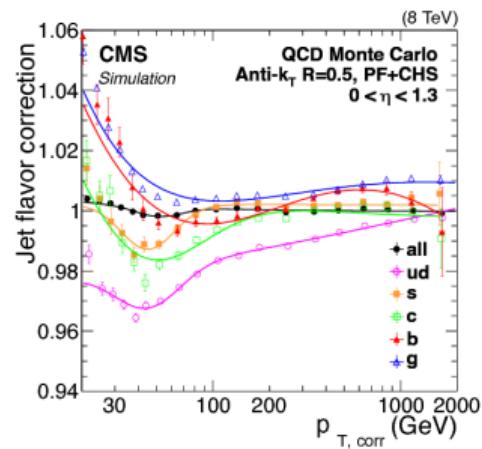
## Applied to data

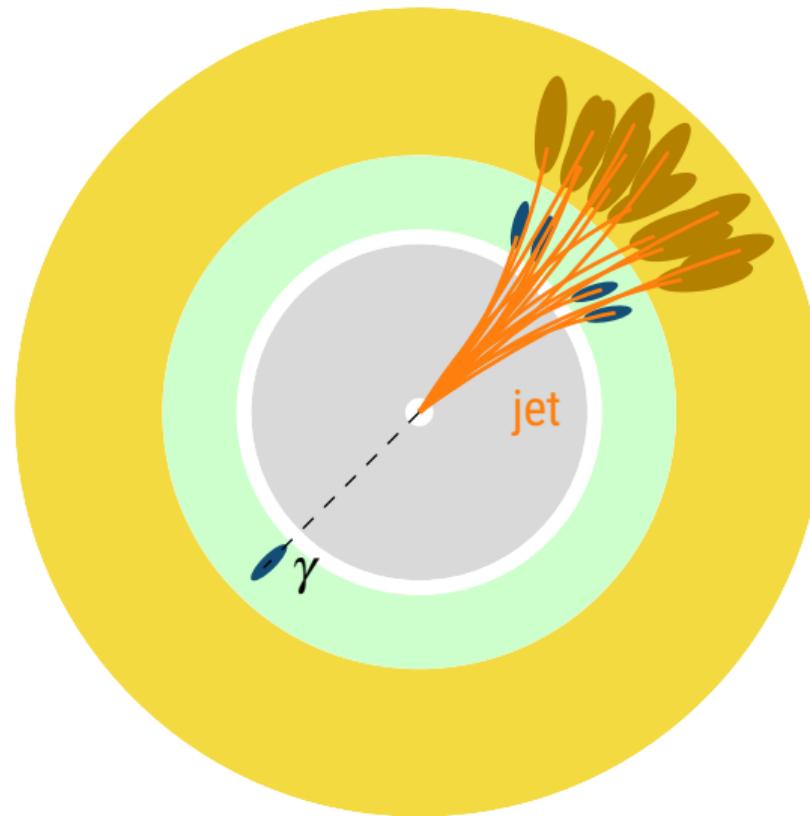


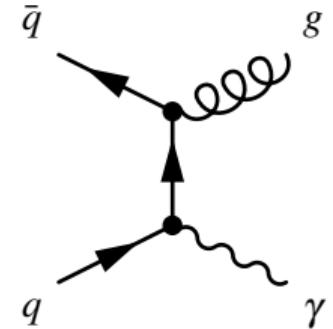
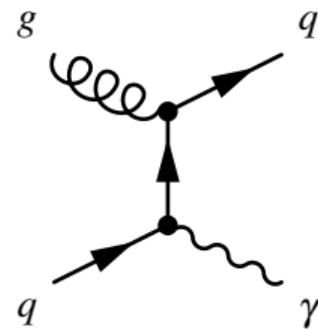
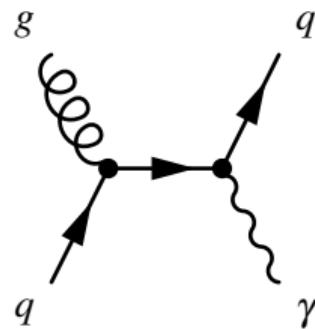


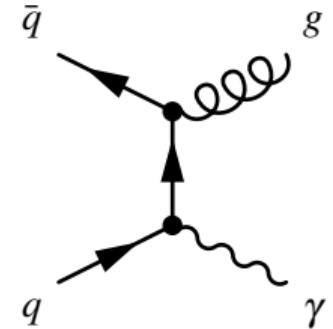
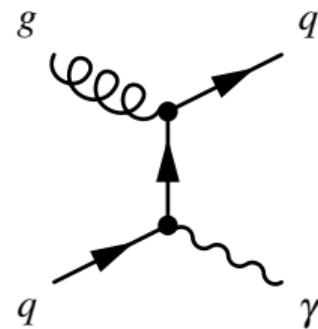
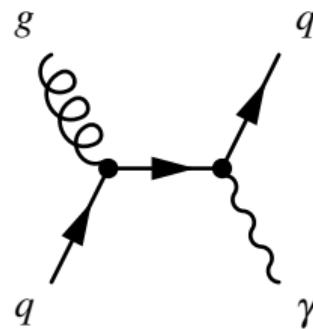
## Applied to data



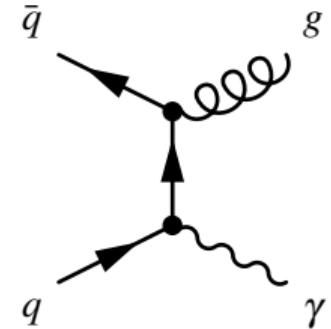
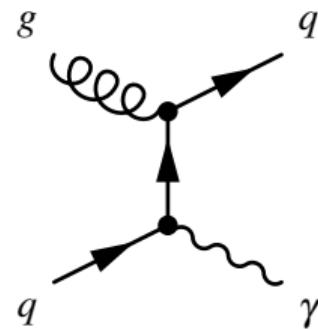
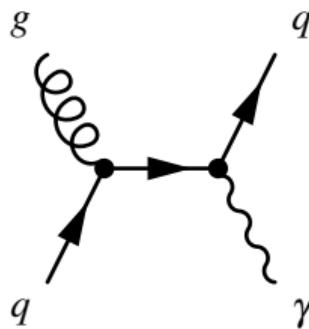
**Applied to data****Applied to simulation**





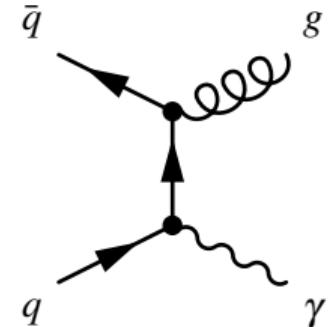
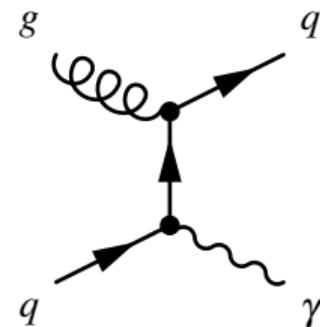
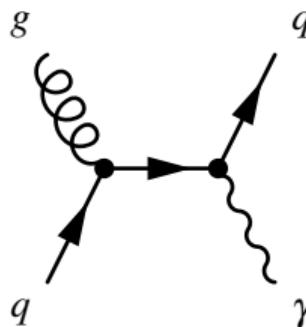


$$\vec{p}_{T \text{ ptcl}}^\gamma + \vec{p}_{T \text{ ptcl}}^{\text{jet}} = \vec{0} \Rightarrow p_{T \text{ ptcl}}^\gamma = p_{T \text{ ptcl}}^{\text{jet}}$$



$$\vec{p}_{\text{T ptcl}}^\gamma + \vec{p}_{\text{T ptcl}}^{\text{jet}} = \vec{0} \Rightarrow p_{\text{T ptcl}}^\gamma = p_{\text{T ptcl}}^{\text{jet}}$$

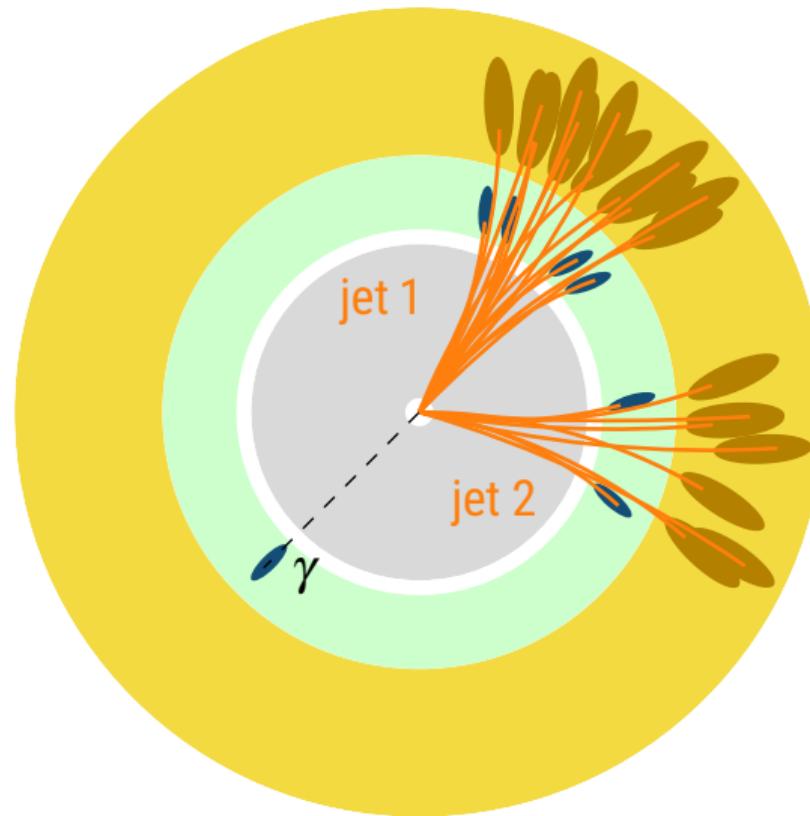
$$R = \frac{p_{\text{T ptcl}}^{\text{jet}}}{p_{\text{T ptcl}}^{\gamma}} = \frac{p_{\text{T ptcl}}^{\text{jet}}}{p_{\text{T ptcl}}^\gamma} \simeq \frac{p_{\text{T ptcl}}^{\text{jet}}}{p_{\text{T ptcl}}^{\gamma}}$$

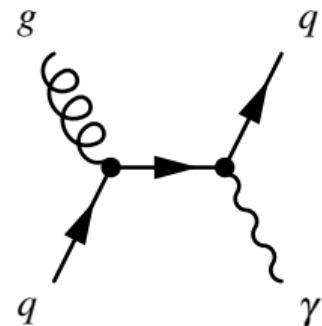


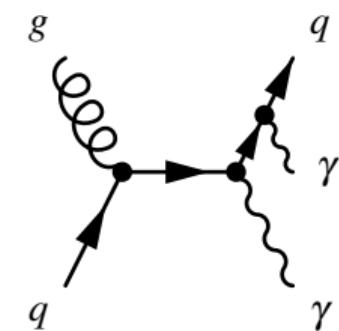
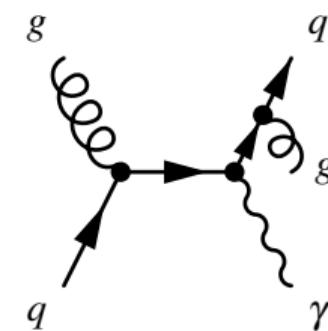
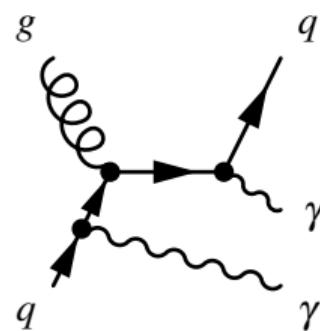
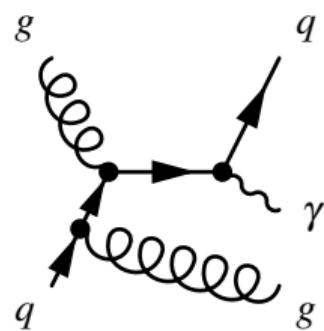
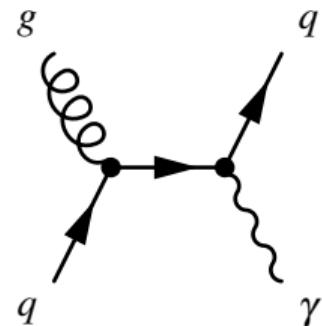
$$\vec{p}_{\text{T ptcl}}^{\gamma} + \vec{p}_{\text{T ptcl}}^{\text{jet}} = \vec{0} \Rightarrow p_{\text{T ptcl}}^{\gamma} = p_{\text{T ptcl}}^{\text{jet}}$$

$$R = \frac{p_{\text{T ptcl}}^{\text{jet}}}{p_{\text{T ptcl}}^{\gamma}} = \frac{p_{\text{T reco}}^{\text{jet}}}{p_{\text{T ptcl}}^{\gamma}} \simeq \frac{p_{\text{T reco}}^{\text{jet}}}{p_{\text{T reco}}^{\gamma}}$$

$$R_{bal} = \frac{p_{\text{T reco}}^{\text{jet}}}{p_{\text{T}}^{\gamma}}$$







$$R_{bal} = \frac{p_T^{\text{jet 1}}}{p_T^\gamma}$$

$$\alpha = \frac{p_T^{\text{jet 2}}}{p_T^\gamma}$$

$$\vec{p}_{T\text{ptcl}}^{\gamma} + \vec{p}_{T\text{ptcl}}^{\text{recoil}} = \vec{0}$$

$$\vec{p}_{\text{Tptcl}}^{\gamma} + \vec{p}_{\text{Tptcl}}^{\text{recoil}} = \vec{0}$$

$$\underbrace{\vec{p}_{\text{Treco}}^{\gamma} + R_{MPF} \vec{p}_{\text{Tptcl}}^{\text{recoil}}}_{\vec{p}_{\text{Treco}}^{\text{recoil}}} = -\vec{E}_{\text{T}}^{\text{miss}} \Rightarrow \boxed{R_{MPF} = 1 + \frac{\vec{p}_{\text{T}}^{\gamma} \cdot \vec{E}_{\text{T}}^{\text{miss}}}{|\vec{p}_{\text{T}}^{\gamma}|^2}}$$

# Jet Energy Resolution

- ▶ Remember  $R_{bal}$  definition,

$$R_{bal} = \frac{p_{T\text{reco}}^{\text{1st jet}}}{p_{T\text{reco}}^\gamma}$$

# Jet Energy Resolution

- ▶ Remember  $R_{bal}$  definition,

$$R_{bal} = \frac{p_{T\text{reco}}^{\text{1st jet}}}{p_{T\text{reco}}^\gamma}$$

Then

$$R_{bal} = \underbrace{\frac{p_{T\text{reco}}^{\text{1st jet}}}{p_{T\text{ptcl}}^{\text{1st jet}}}}_{\sigma_{\text{jet}} = \text{JER}} \times \underbrace{\frac{p_{T\text{ptcl}}^{\text{1st jet}}}{p_{T\text{ptcl}}^\gamma}}_{\text{PLI}} \times \underbrace{\frac{p_{T\text{ptcl}}^\gamma}{p_{T\text{reco}}^\gamma}}_{\sigma_\gamma \equiv 1}$$

- ▶ PLI: Particle Level Imbalance (pile-up, radiations, neutrinos...),  $\rightarrow 0$  when  $\alpha \rightarrow 0$ .

# Jet Energy Resolution

- ▶ Remember  $R_{bal}$  definition,

$$R_{bal} = \frac{p_{T\text{reco}}^{\text{1st jet}}}{p_{T\text{reco}}^\gamma}$$

Then

$$R_{bal} = \underbrace{\frac{p_{T\text{reco}}^{\text{1st jet}}}{p_{T\text{ptcl}}^{\text{1st jet}}}}_{\sigma_{\text{jet}} = \text{JER}} \times \underbrace{\frac{p_{T\text{ptcl}}^{\text{1st jet}}}{p_{T\text{ptcl}}^\gamma}}_{\text{PLI}} \times \underbrace{\frac{p_{T\text{ptcl}}^\gamma}{p_{T\text{reco}}^\gamma}}_{\sigma_\gamma \equiv 1}$$

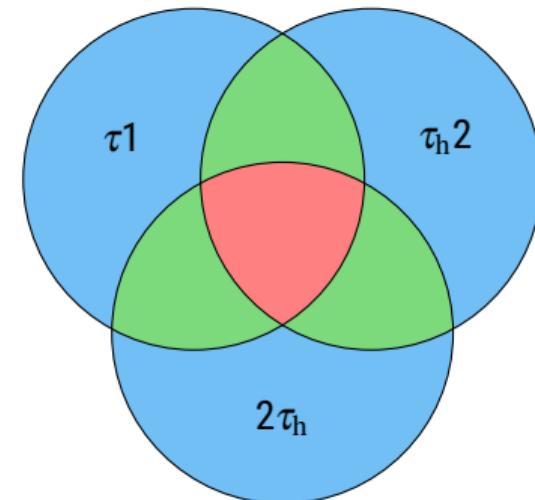
- ▶ PLI: Particle Level Imbalance (pile-up, radiations, neutrinos...),  $\rightarrow 0$  when  $\alpha \rightarrow 0$ .

$$\text{JER} = \sigma_{\text{jet}} = \sqrt{\sigma_{R_{bal}}^2 - \sigma_{\text{PLI}}^2}$$

# Triggers in the $\tau_h \tau_h$ channel

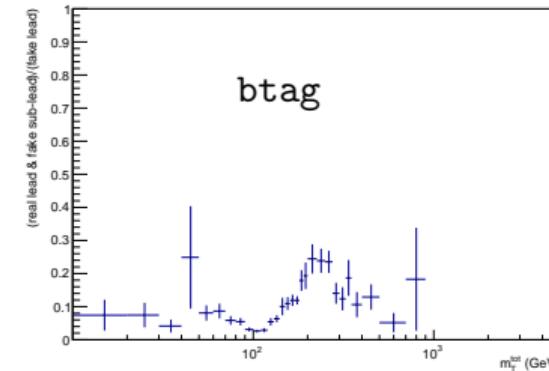
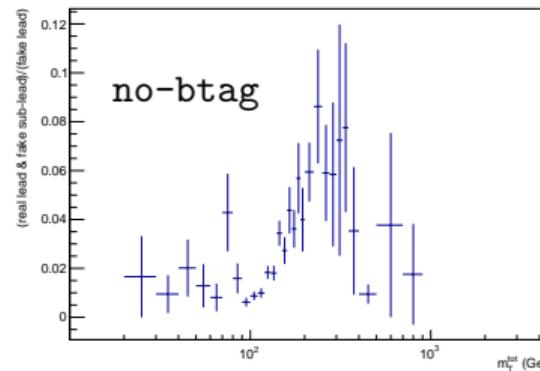
- ▶ In the manuscript: page 118, formula (4.5)

$$\begin{aligned}\varepsilon = & \varepsilon(2\tau_h) + \varepsilon(\tau_h 1) + \varepsilon(\tau_h 2) \\ & - \varepsilon(2\tau_h + \tau_h 1) - \varepsilon(2\tau_h + \tau_h 1) - \varepsilon(\tau_h 1 + \tau_h 2) \\ & + \varepsilon(2\tau_h + \tau_h 1 + \tau_h 2)\end{aligned}$$

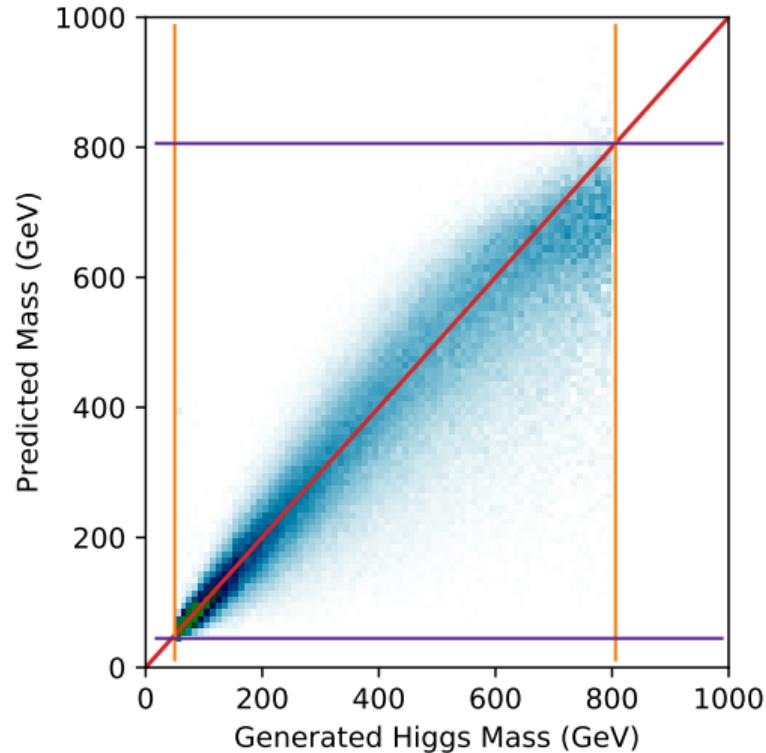
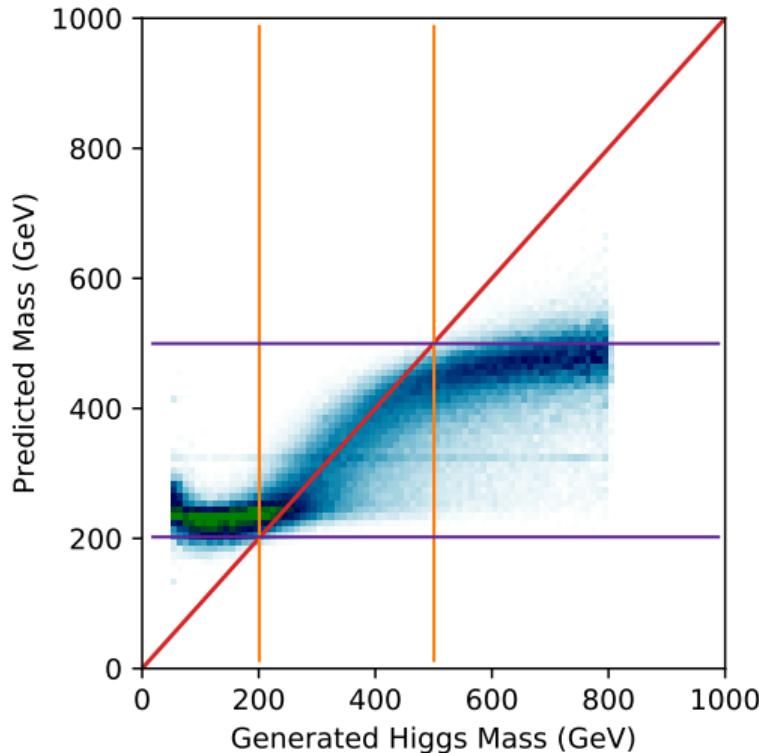


# Fake factors for subleading $\tau_h$

- ▶ The  $\tau_h \tau_h$  fake factors are measured for the leading  $\tau_h$  candidate only.
  - ▷ The subleading one can be either a genuine or fake  $\tau_h$ .
- ▶ At this point, underestimation of events in which only the subleading  $\tau_h$  is a fake.
  - ▷ Adding these back using MC.
- ▶ Small fraction of fakes < 10% in no-btag, < 30% in btag (due to  $t\bar{t}$ ):



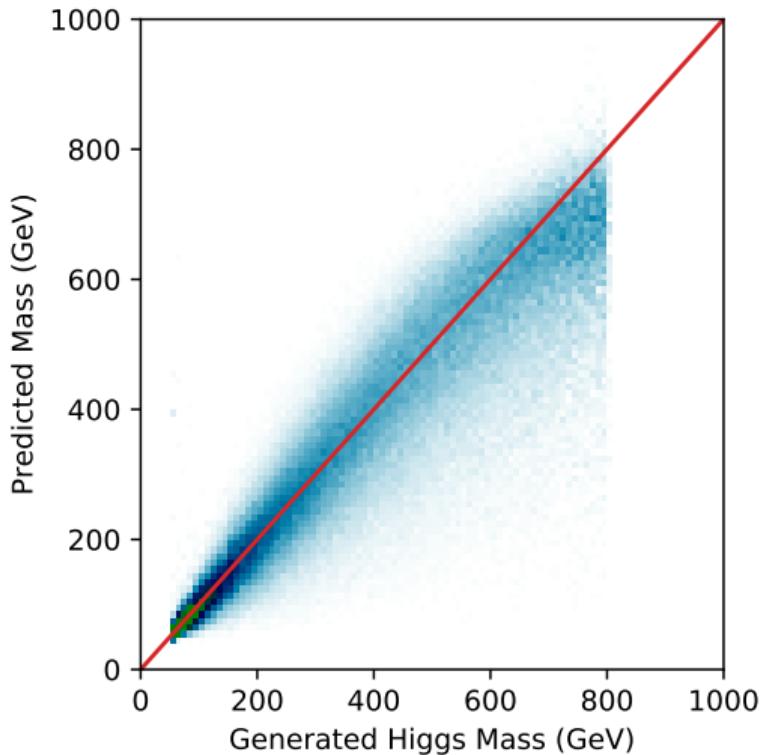
- ▷ J. Andrejkovic et al. "BSM  $H \rightarrow \tau\tau$  analysis on full Run 2 CMS data at  $\sqrt{s} = 13$  TeV". *CMS analysis Note* (2021). URL:  
[https://cms.cern.ch/iCMS/jsp/db\\_notes/noteInfo.jsp?cmsnoteid=CMS%5C20AN-2020/218](https://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%5C20AN-2020/218).

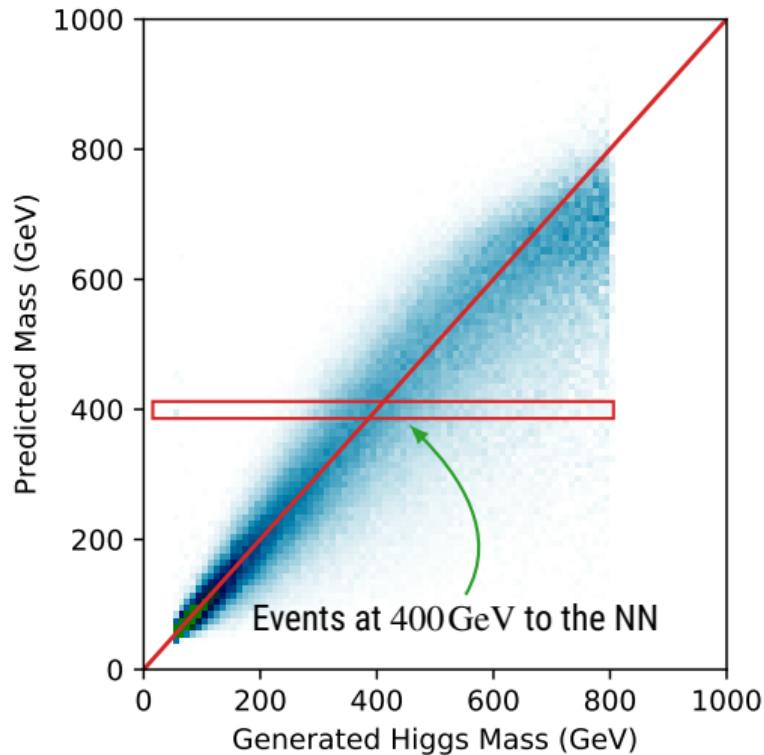
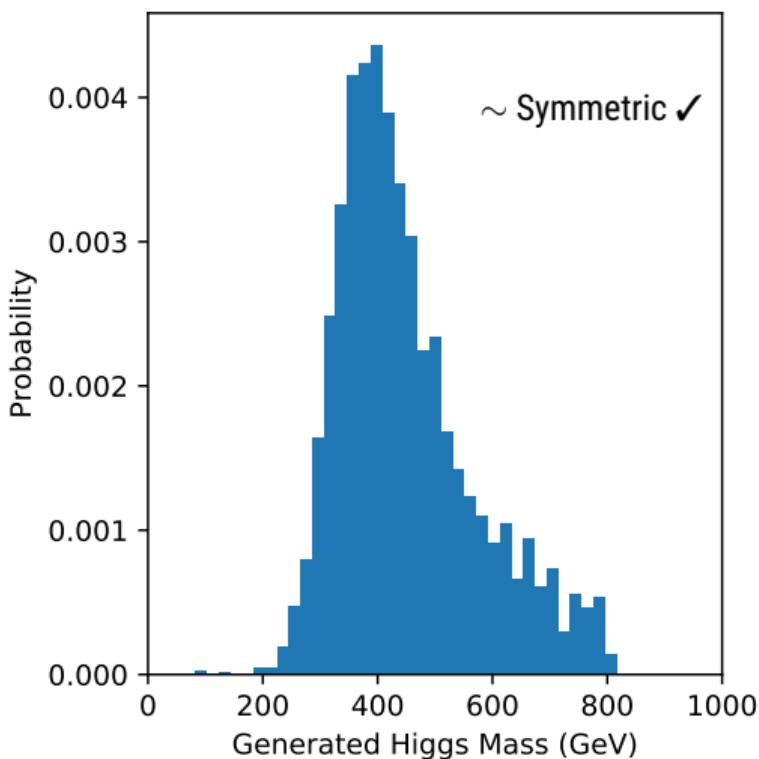


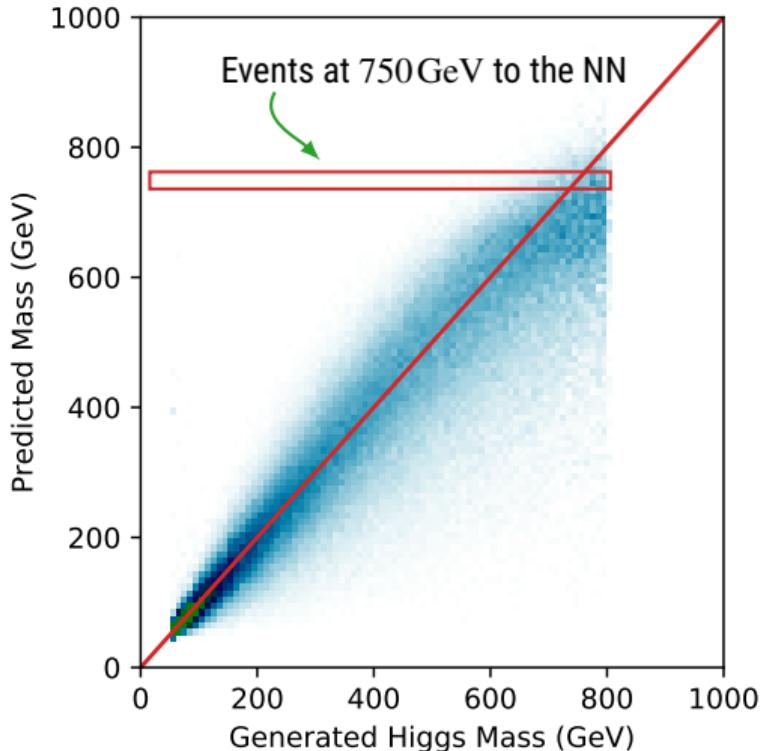
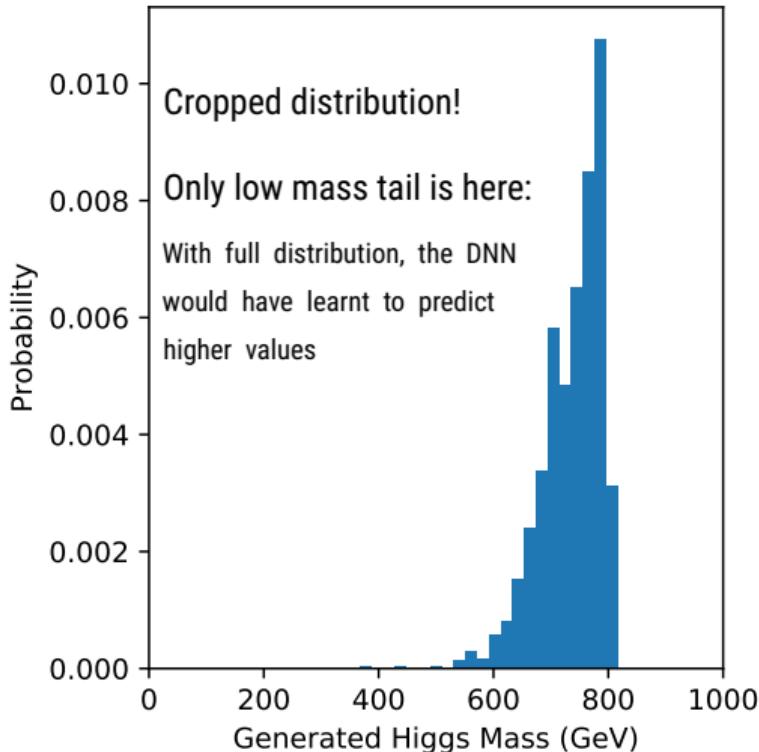
► How to cope with the boundaries?

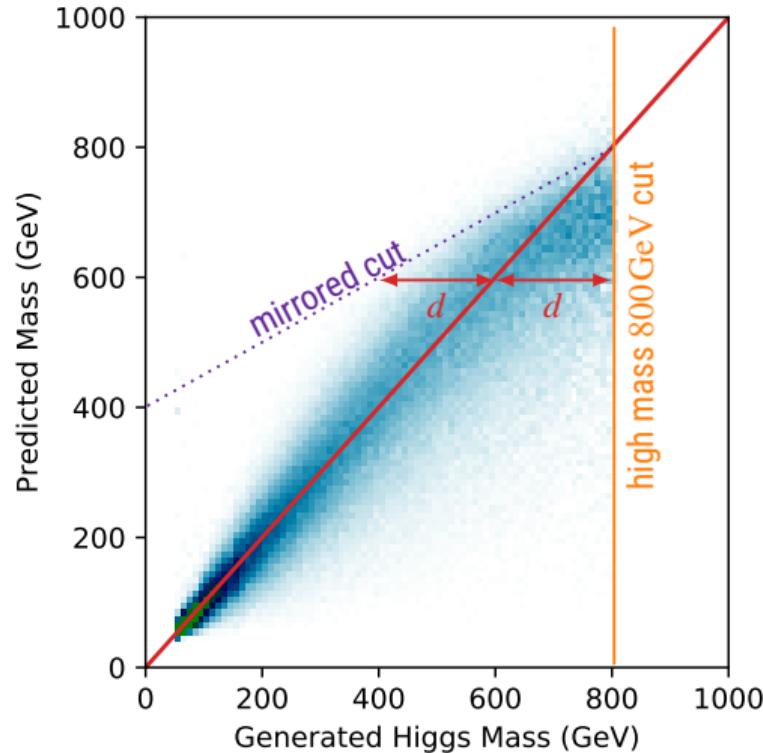
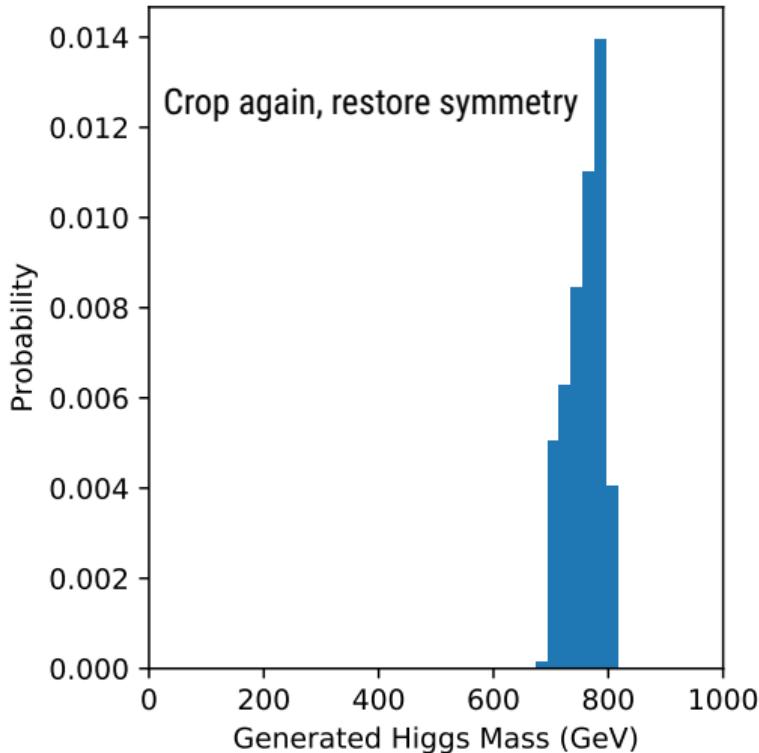
- ▷ Bias to be balanced,
- ▷ Extend the mass range?
  - ▷ Would be nice!
  - ▷ Not always feasible...

► Every horizontal slice is one predicted value:  
▷ "Same family" of events to the NN.









$$\mathfrak{L}_{\text{MA}\sqrt{\text{PE}} \times b}(y_{\text{true}}, y_{\text{pred}}) = \mathfrak{L}_{\text{MA}\sqrt{\text{PE}}}(y_{\text{true}}, y_{\text{pred}}) \\ \times \begin{cases} 0 & \text{if } (y_{\text{true}}, y_{\text{pred}}) \in \text{area 3} \\ 0.1 & \text{if } (y_{\text{true}}, y_{\text{pred}}) \in \text{area 4} \\ 1 & \text{else} \end{cases}$$

$$\mathfrak{L}_{\text{MA}\sqrt{\text{PE}}}(y_{\text{true}}, y_{\text{pred}}) = \mathfrak{L}_{\text{MAPE}}(y_{\text{true}}, y_{\text{pred}}) \times \sqrt{y_{\text{true}}} \\ = \left| \frac{y_{\text{pred}} - y_{\text{true}}}{y_{\text{true}}} \right| \times \sqrt{y_{\text{true}}} \\ \Leftrightarrow \mathfrak{L}_{\text{MA}\sqrt{\text{PE}}}(y_{\text{true}}, y_{\text{pred}}) = \left| \frac{y_{\text{pred}} - y_{\text{true}}}{\sqrt{y_{\text{true}}}} \right|.$$

