Machine Learning with Augmentation for Boosting di-Higgs Searches at the LHC

W. S. Cho
(Seoul National University)

in collaboration with
K. Ban, D. Lee, C.B.Park, J.H.Park, S.C.Park

Deep Learning for the Natural Science
1st March 2019 Hamburg

About our study (DL wise...)

1) Improvement of the performance, by feeding augmented features to NN discriminative model :

⟶ Missing info. in data blocks, can exist everywhere.

⟶ Augmentation of missing features (in entry-by-entry basis), NOT the 'data augmentation' for the amount of data.

⟶ <u>Prescription for the augmentation</u>

: Given visible/available information (in entry-by-entry basis), Augment the missing blocks defined by ALL possible/relevant hypothesis (for both signals & backgrounds) which can contribute to data.

About our study (DL wise...)

2) Role of physics, for discriminative models embodied by Neural Network :

→ comparing by the quality of input features

→ HL (high level, from physical invariances) vs RL(raw level)

→ RL may be enough [P. Baldi, P. Sadowski, D. Whiteson 2014].

→ Physics matters for DL ?

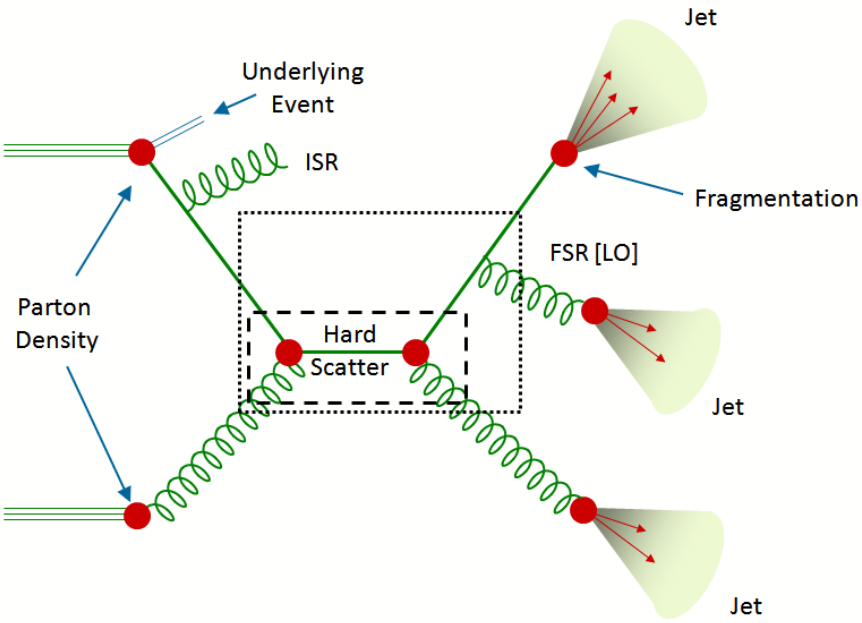
→ We believe 'Yes' in many ways & at least for the augmentation for auged HL vs auged RL. here :)

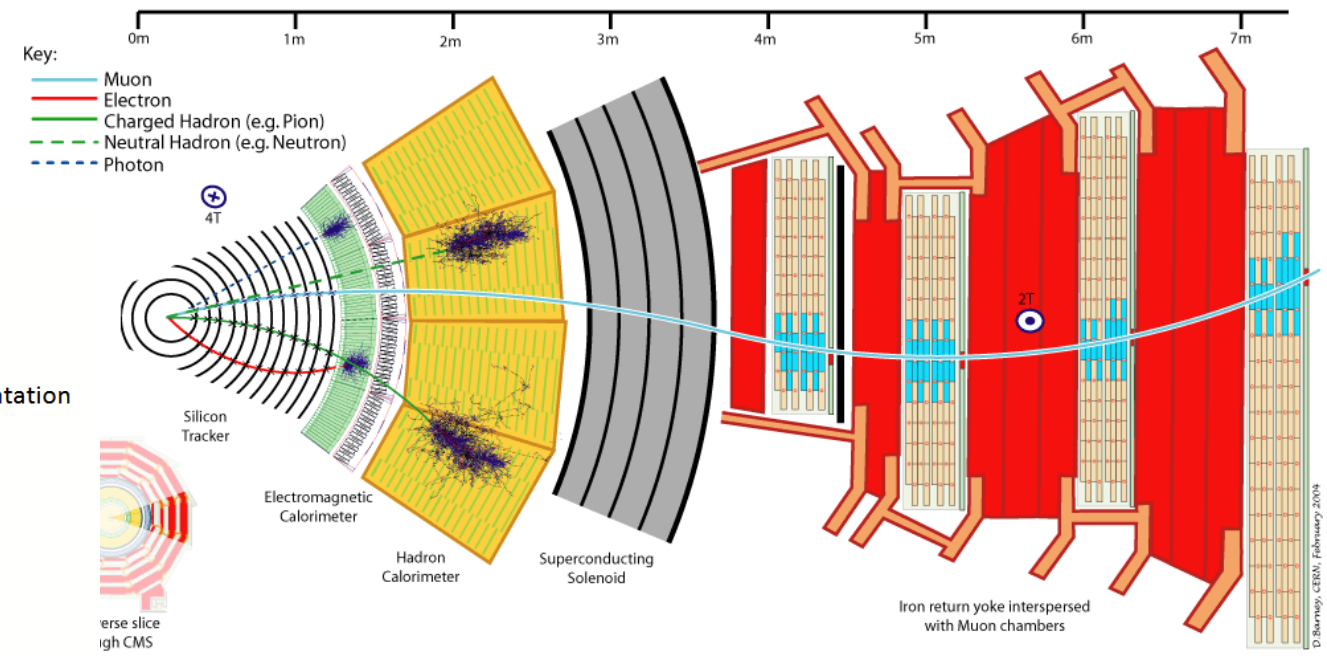
3) Our NN models are all simple (?) MLP (multi-layer perceptrons) type :

→ implemented using Keras @ tensorflow-backend (~10 Dense layers (~10L x 500 nodes))
→ GPU resources (Titan-XP) used for training.
→ Regularisation (drop-out, batch-normalization), optimiser (Adam)

→ no innovations in NN architecture yet.

Where our domain problem is
= discrimination of hard scattering processes at the LHC
(di-Higgs production vs other BG)





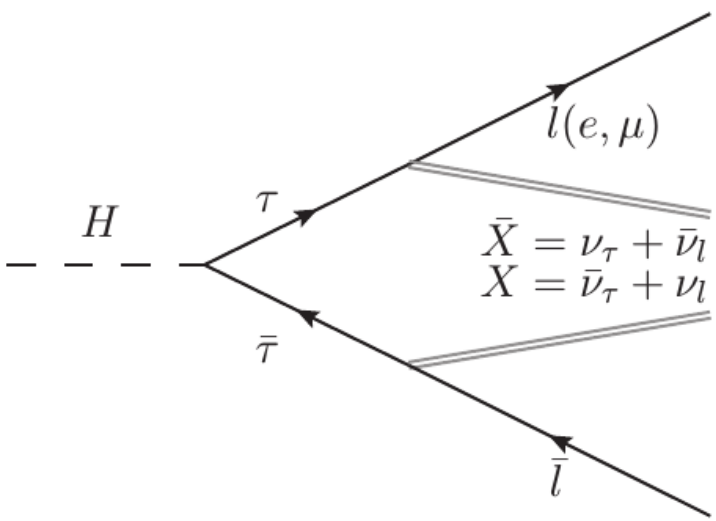
from arXiv:1002.1708 [hep-ex]

: LHC data analysis, based on the long chain/set of effective models in various different scales and systems, from high energy vacuum to the electrical digits from detector facilities

Elusive di-Higgs production and decays
(also with multiple invisible particles)

- **1/1000 times smaller production rate** than single Higgs production (discovered at the LHC in 2012)

- HH ⟶ bbWW & HH ⟶ bbττ channels
  - sizable branching ratios
  - huge ttbar backgrounds
  - large MET from multiple neutrinos



ex) H⟶ ττ ⟶ 2ℓ + MET with 4 neutrinos

channel	BR (%)
bbbb	~33
bbWW	**~25**
bbττ	**~7.3**
WWWW	~4.3
bbrr	~0.27

HH → bbWW & bbττ VS top-quark pair production (BG)

in 3 categorical signatures for 2b + 2L ($n_{\ell=e,\mu} + n_{\tau(h)}=2$) + MET

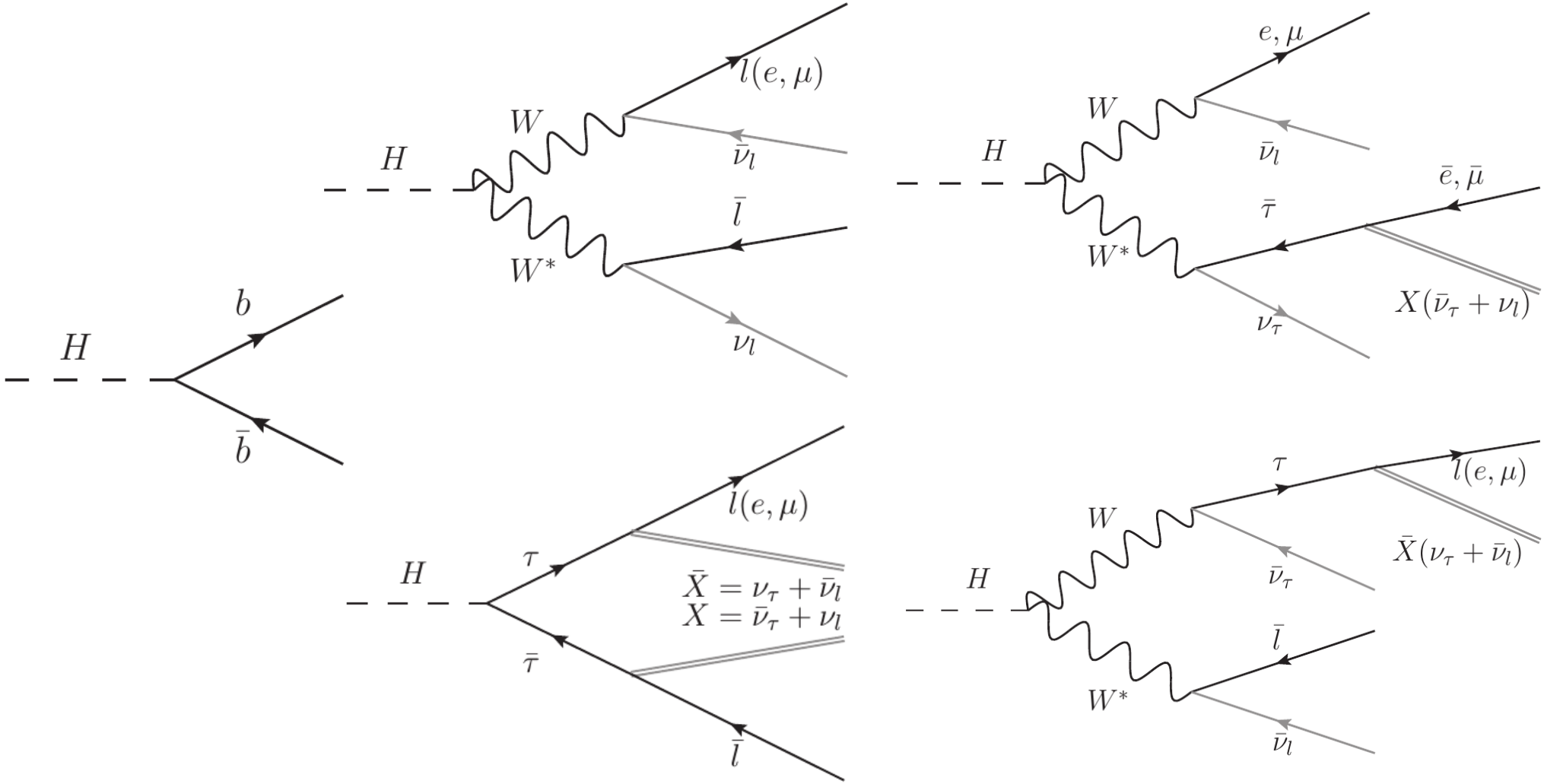
categorized further by tau decay kinematics

Channel	Leptons	X section	Topology $* \, l = e \, or \, \mu$
HH2Tau	0	~1.2	$hh \to b\,b\,\tau\,\tau \to b\,b\,\tau_h\,\tau_h + met$
TT2Tau	0	~5097.2	$\bar{t}t \to b\,w\,b\,w \to b\,b\,\tau\,\tau + met \to b\,b\,\tau_h\,\tau_h + met$
HH2Tau	1	~1.3	$hh \to b\,b\,\tau\,\tau \to b\,b\,\tau_h\,l + met$
HH2W*W	1	~0.15	$hh \to b\,b\,w\,w^* \to b\,b\,\tau\,l + met \to b\,b\,\tau_h\,l + met$
HH2WW*	1	~0.15	$hh \to b\,b\,w\,w^* \to b\,b\,l\,\tau + met \to b\,b\,l\,\tau_h + met$
TT2Tau	1	~5546.3	$\bar{t}t \to b\,w\,b\,w \to b\,b\,\tau\,\tau + met \to b\,b\,\tau_h\,l + met$
TT1Tau	1	~29700.2	$\bar{t}t \to b\,w\,b\,w \to b\,b\,\tau\,l + met \to b\,b\,\tau_h\,l + met$
HH2Tau	2	~0.36	$hh \to b\,b\,\tau\,\tau \to b\,b\,l\,l + met$
HH2W*W1Tau	2	~0.08	$hh \to b\,b\,w\,w^* \to b\,b\,\tau\,l + met \to b\,b\,l\,l + met$
HH2WW*1Tau	2	~0.08	$hh \to b\,b\,w\,w^* \to b\,b\,l\,\tau + met \to b\,b\,l\,l + met$
HH2WW0Tau	2	~0.47	$hh \to b\,b\,w\,w^* \to b\,b\,l\,l + met$
TT2Tau	2	~1508.7	$\bar{t}t \to b\,w\,b\,w \to b\,b\,\tau\,\tau + met \to b\,b\,l\,l + met$
TT1Tau	2	~16158.3	$\bar{t}t \to b\,w\,b\,w \to b\,b\,\tau\,l + met \to b\,b\,l\,l + met$
TT0Tau	2	~43263.9	$\bar{t}t \to b\,w\,b\,w \to b\,b\,l\,l + met$

from K.Y.Ban's talk at KEK

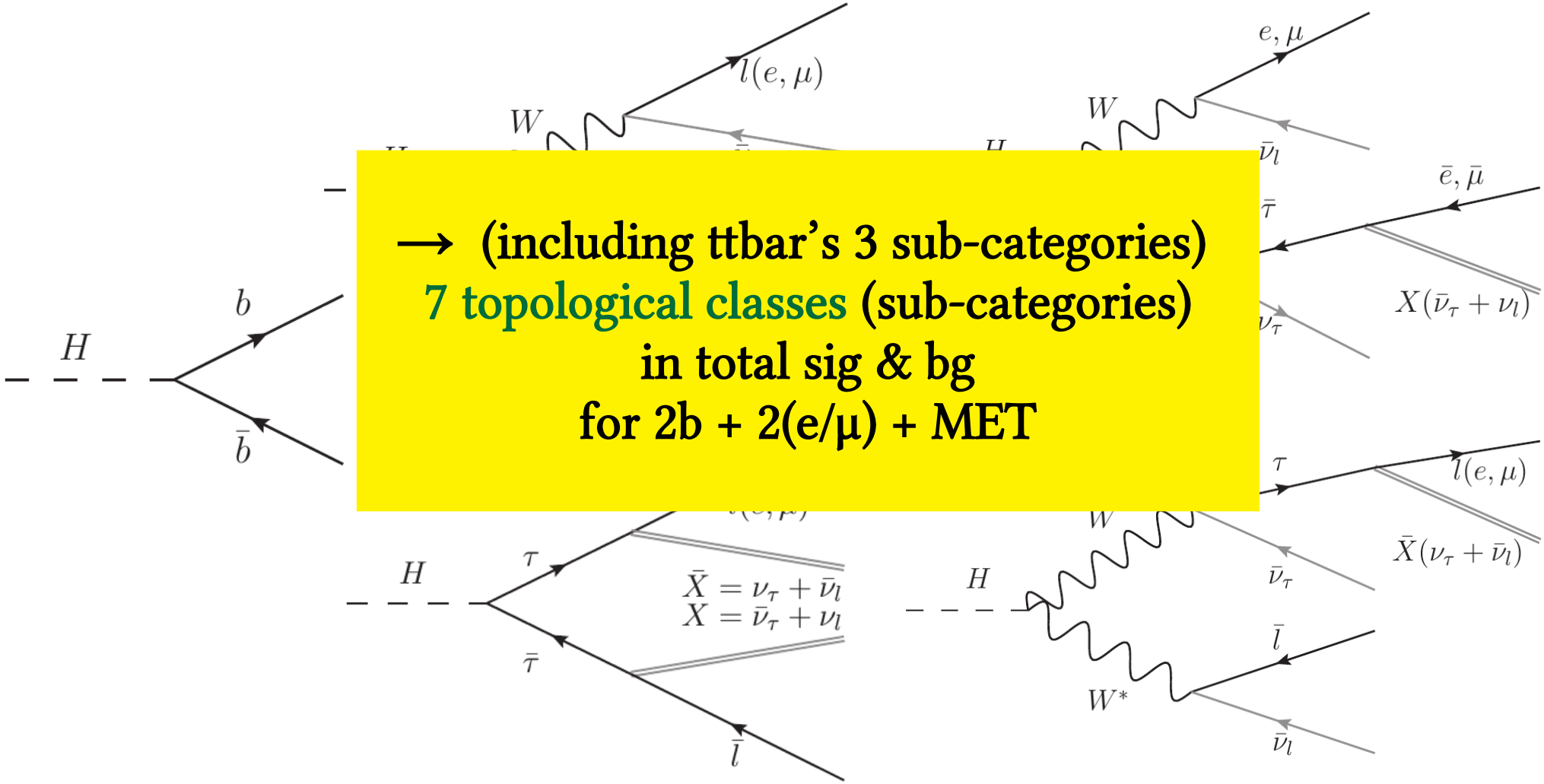
ex) 3rd cartegory : HH ⟶ 2b + 2(e/μ) + MET (4 sub-categories)

from bbWW & bbtautau production



ex) 3rd cartegory : HH ⟶ 2b + 2(e/μ) + MET (4 sub-categories)

from bbWW & bbtautau production



⟶ (including ttbar's 3 sub-categories)

7 topological classes (sub-categories)

in total sig & bg

for 2b + 2(e/μ) + MET

(Topological) Augmentation of invisible missing momenta in under-constrained systems using OptiMass package

OptiMASS provides…

1) **Augmented** invisible 4 momentum vectors (by given physical hypothesis, **h**)

$$\tilde{q}^*$$

2) Minimized mass variable $\bar{M}(p, \tilde{q}^*)$

3) Compatibility Distance (D) of the event with respect to the hypothesis (h)

$$D^l(p) = \left(\sum_{i=1}^{n_c} \left| \vec{C}_i(p, \tilde{q}^*) \right|^2 \right)^{1/2}$$

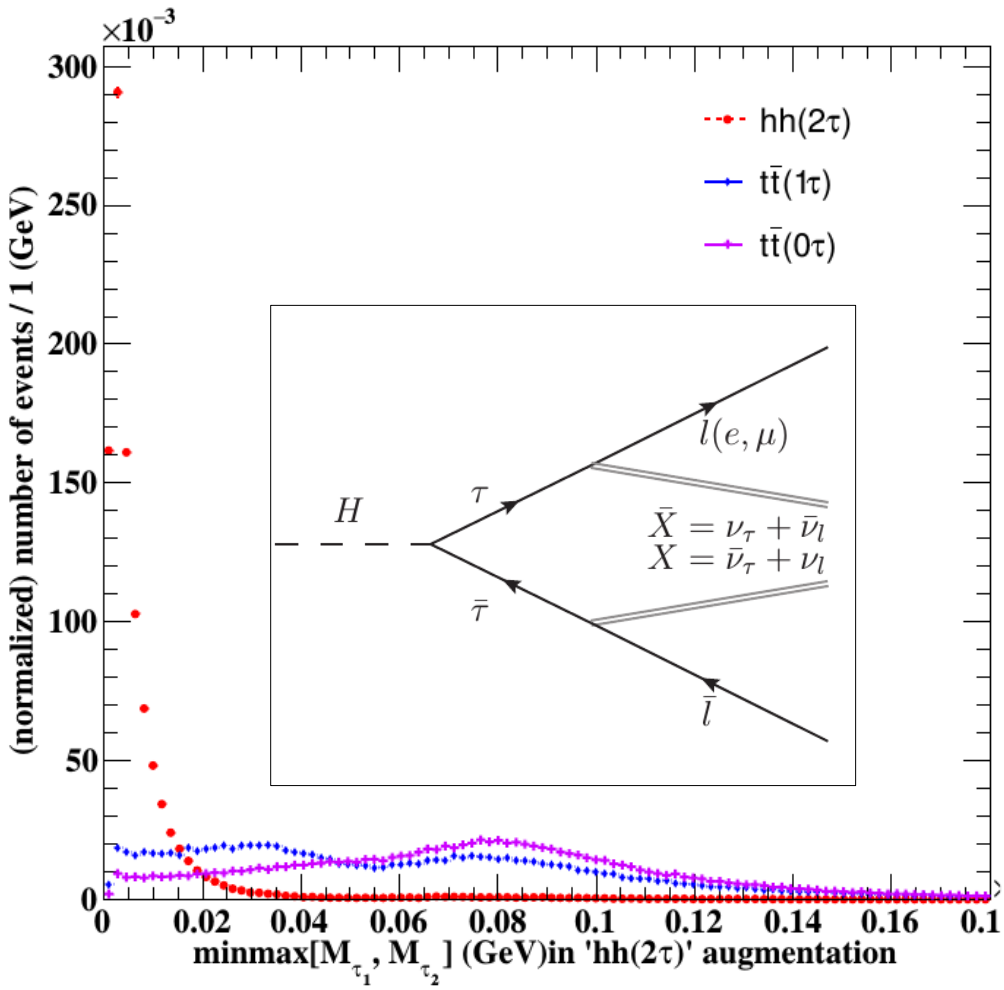
via the optimization of Augmented Lagrangian :

$$\tilde{\mathcal{L}}^l(p, \tilde{q}; \vec{\lambda}, \mu) = \tilde{M}^l(p, \tilde{q}) - \sum_{i=1}^{n_c} \vec{\lambda}_i \vec{C}_i^l + \frac{1}{\mu} \sum_{i=1}^{n_c} \left| \vec{C}_i^l \right|^2$$

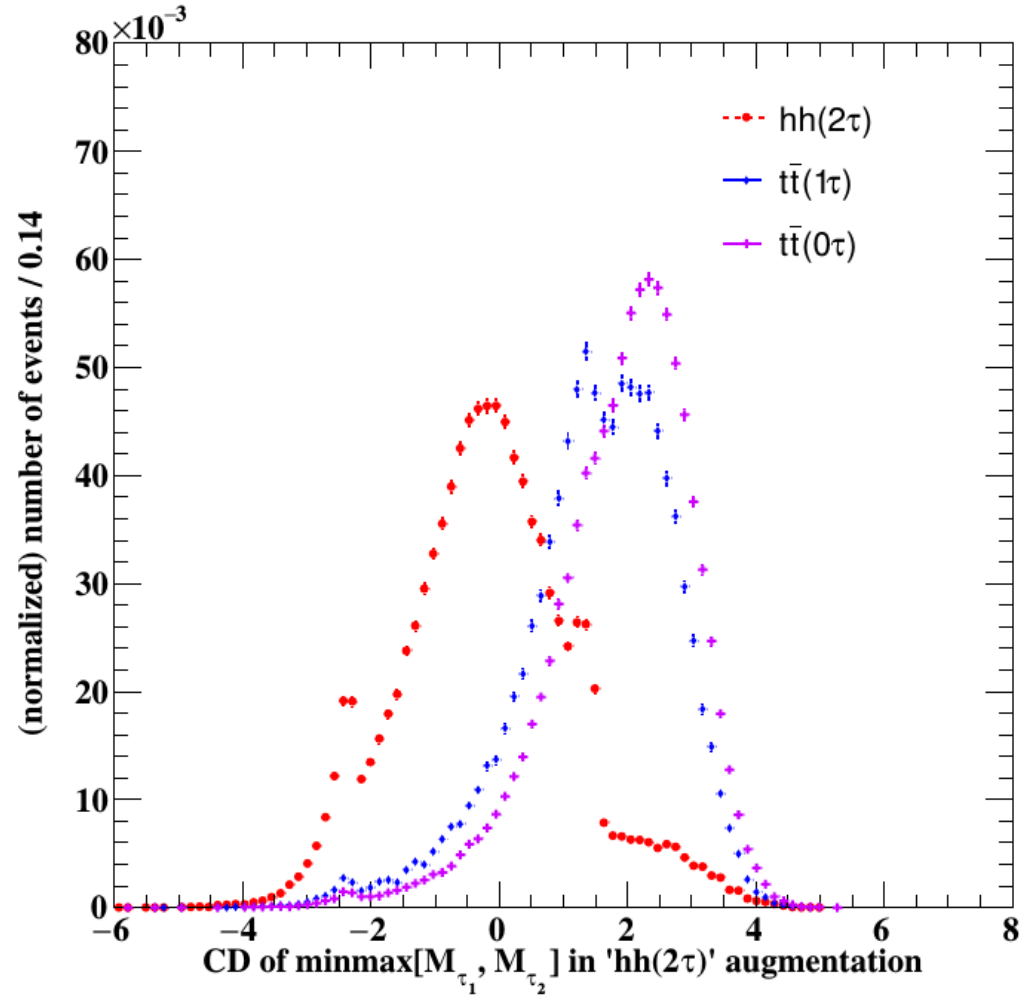
Augmented missing features of data by HH(sig) & ttbar(bg) scenarios

Augmented features 1 : HL (optimass + compatibility distance) [ex1]

OptiMass of 3 procs (Sig & Bg)
in HH(bbττ) augmentation

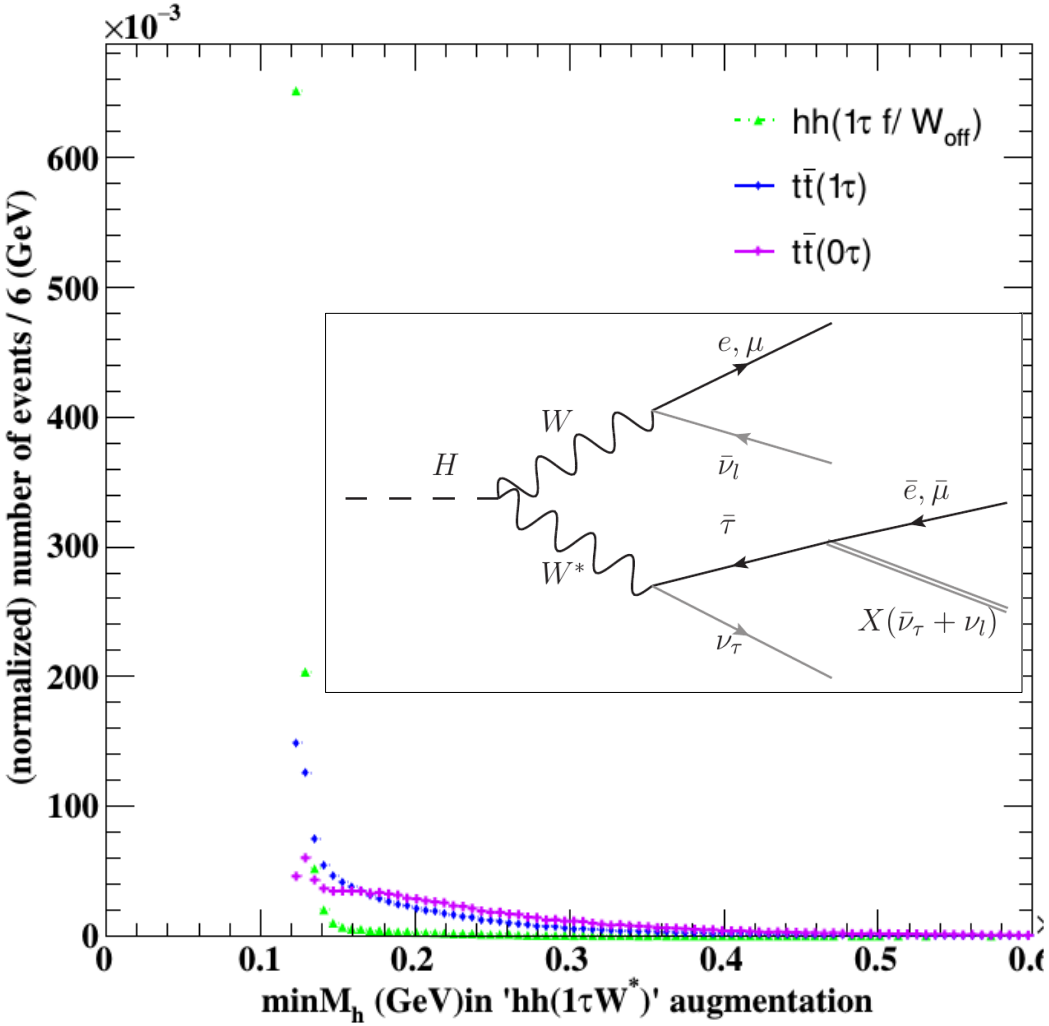


Com. Dist. of 3 procs (Sig & Bg)
in HH(bbττ) augmentation

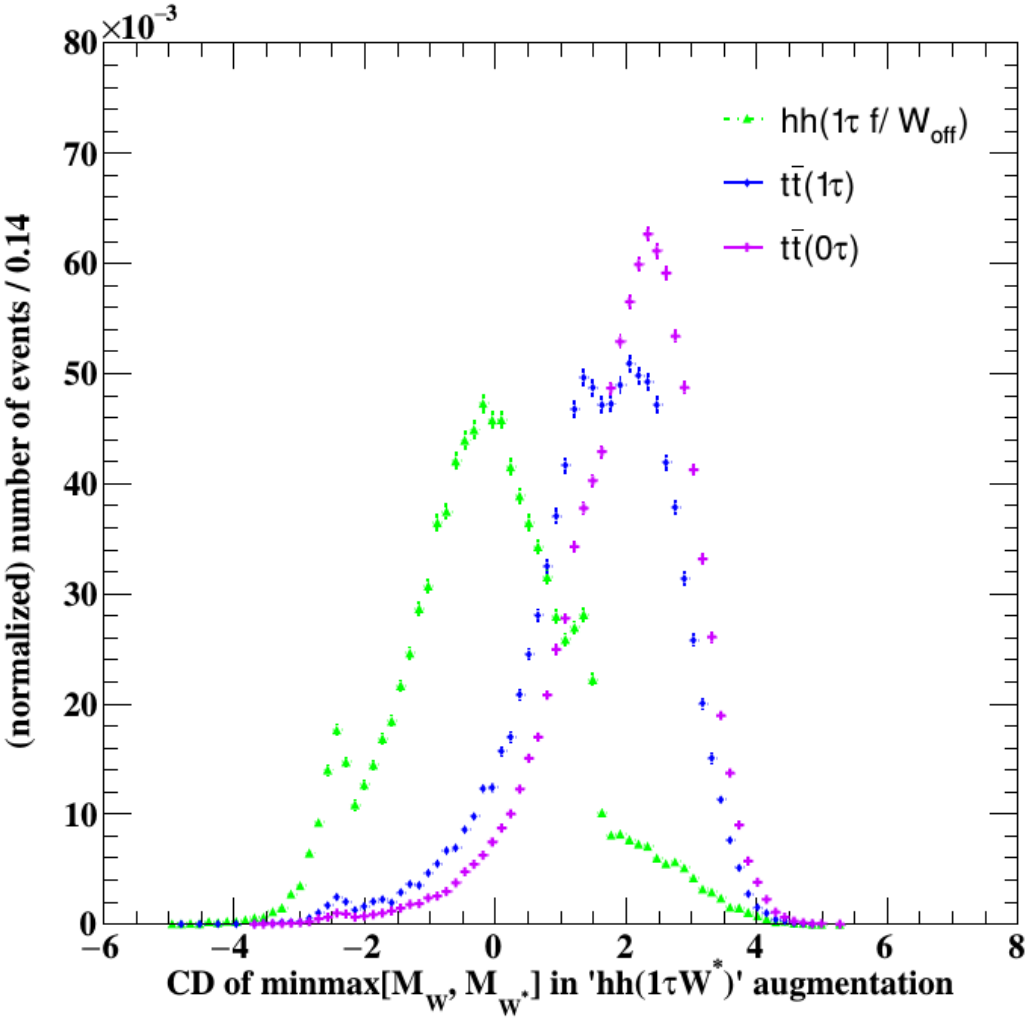


Augmented features 1 : HL (optimass + compatibility distance) [ex2]

OptiMass of 3 procs (Sig & Bg)
in HH(bbWW*) augmentation

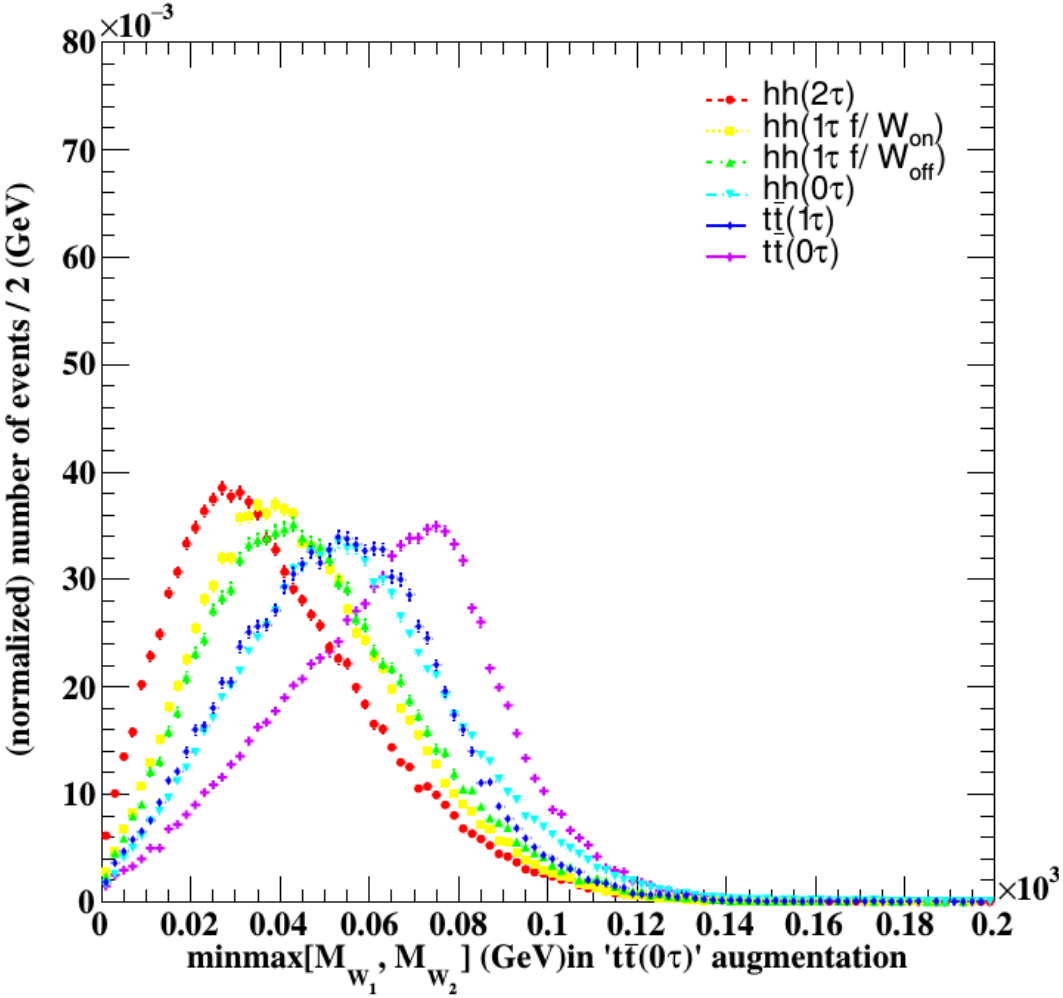


Com. Dist. of 3 procs (Sig & Bg)
in HH(bbWW*) augmentation

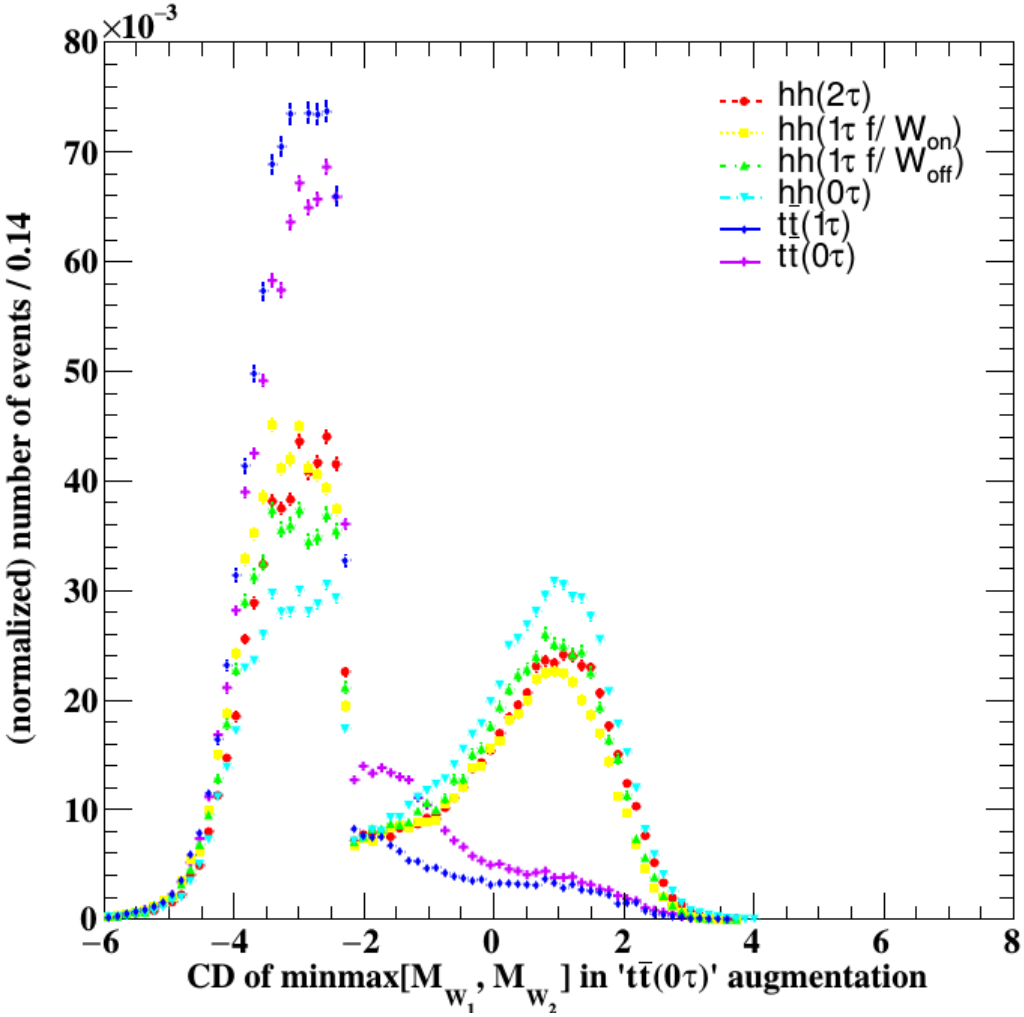


Augmented features 1 : HL (optimass + compatibility distance) [ex3]

OptiMass of 3 procs (Sig & Bg)
in ttbar(0tau) augmentation

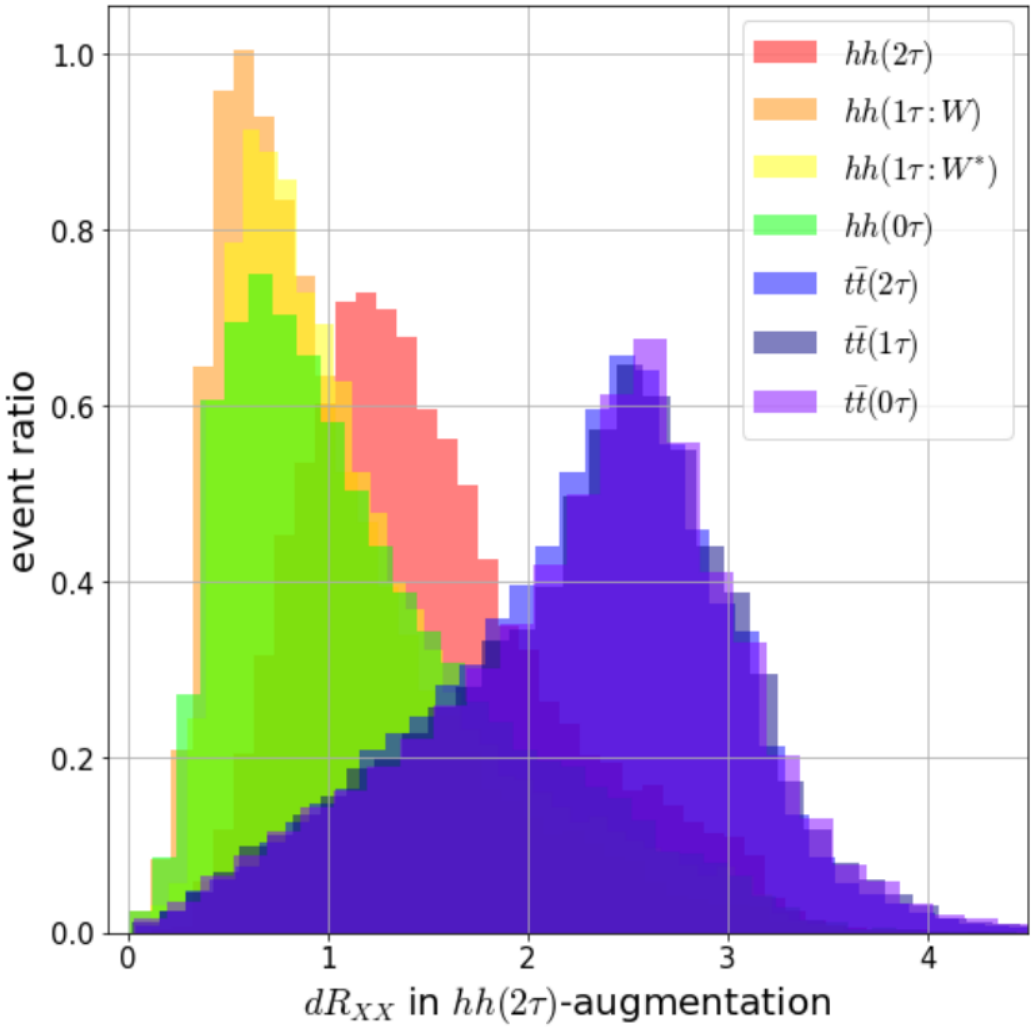


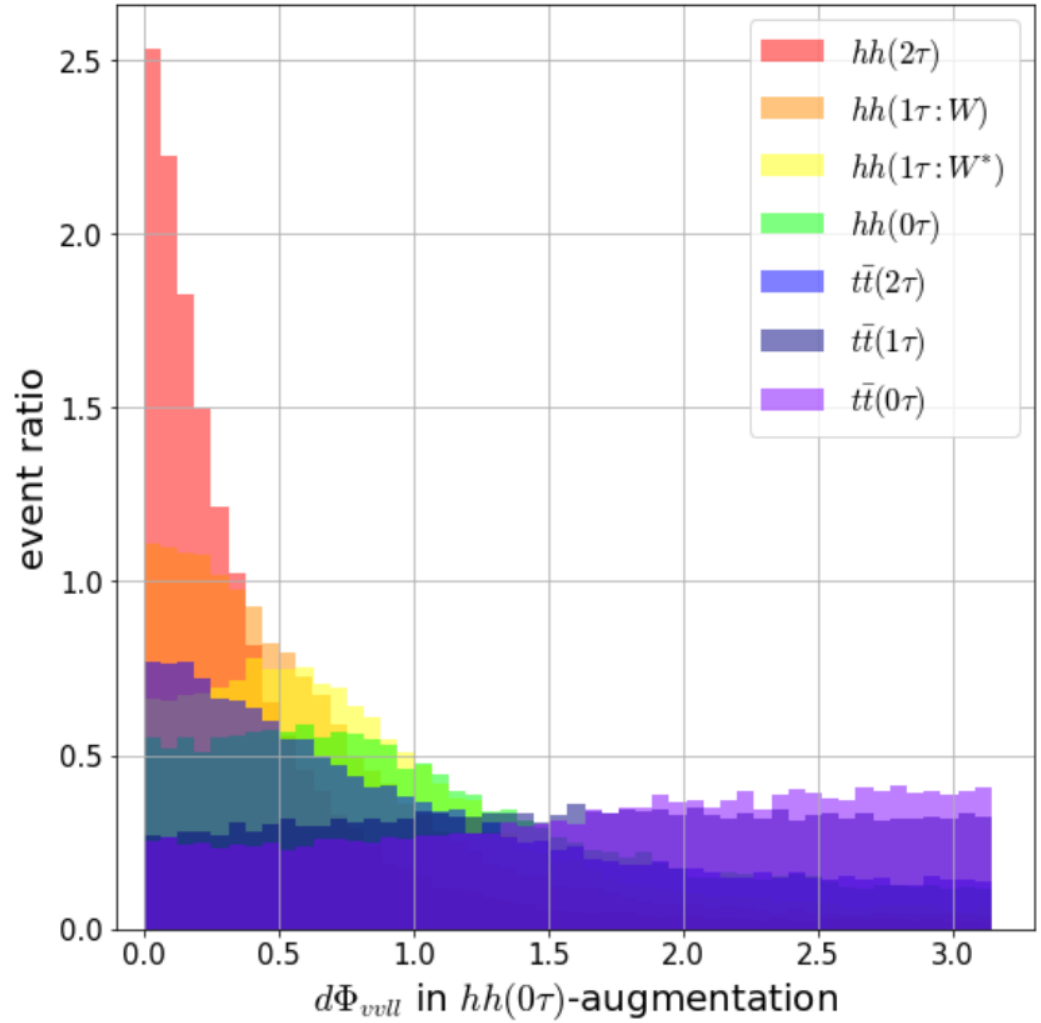
Com. Dist. of 3 procs (Sig & Bg)
in ttbar(0tau) augmentation



Augmented Features 2 : RL feature variables

(raw momenta, angular variables) ⟶ Ndim ~ (100-200)





Many augmented features (@HL, @RL) for a given event entry

from

(subsystem optimass) x (kinematic constraint profile)

---

for (a given augmentation model)

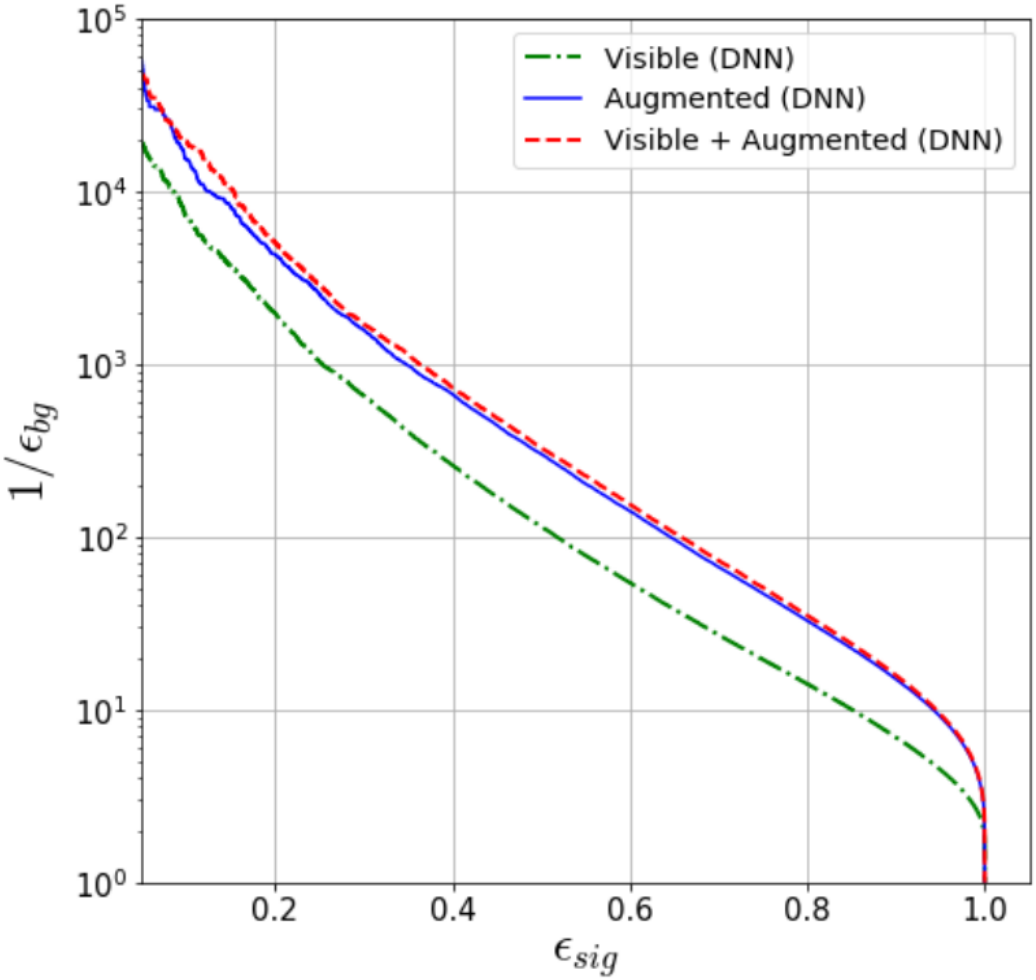
Results

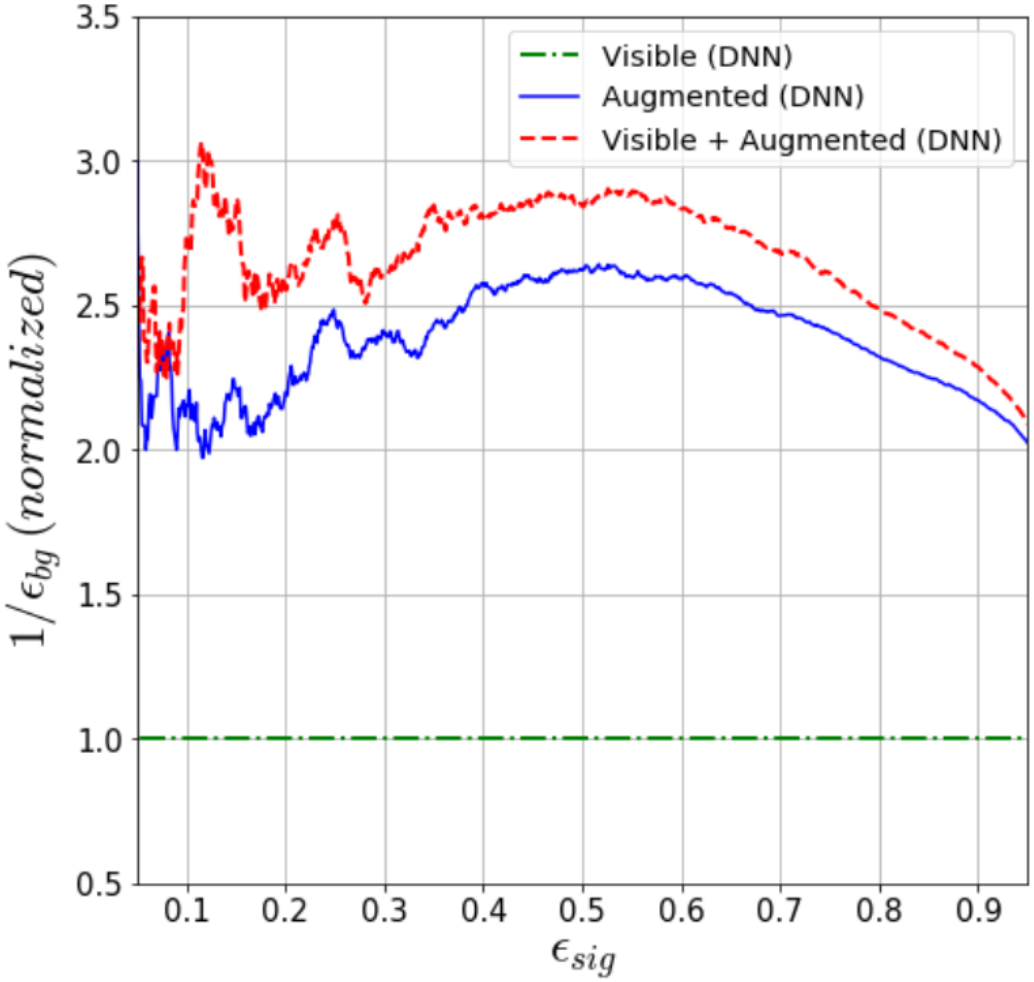
Improvement with the augmented features (2L=2l case)

[Feature variable set definition]

'**Visible**' : variable sets used in CMS PAS 17-006

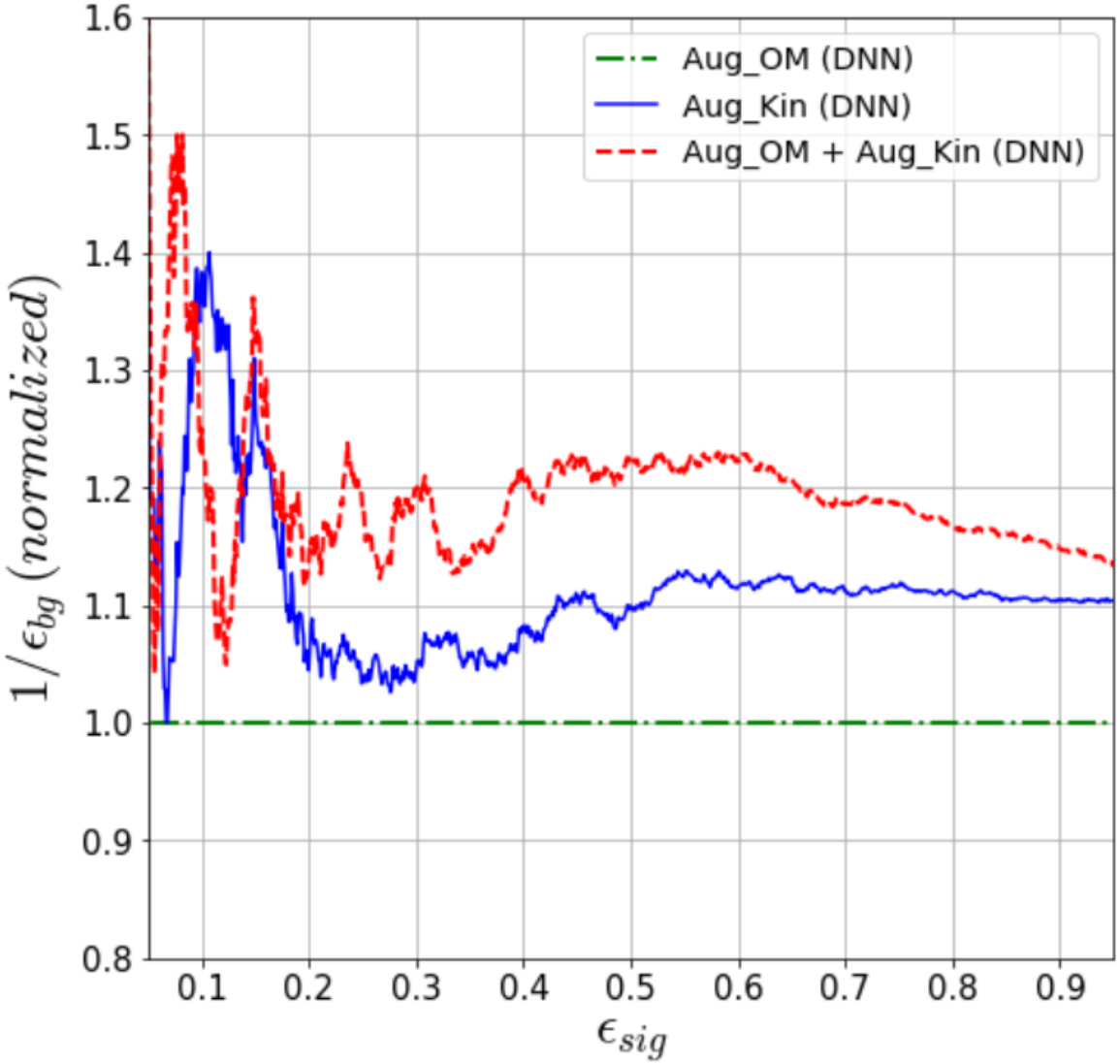
'**Augmented**' : augmented HL [Ndim~50] + RL [Ndim ~ 200]

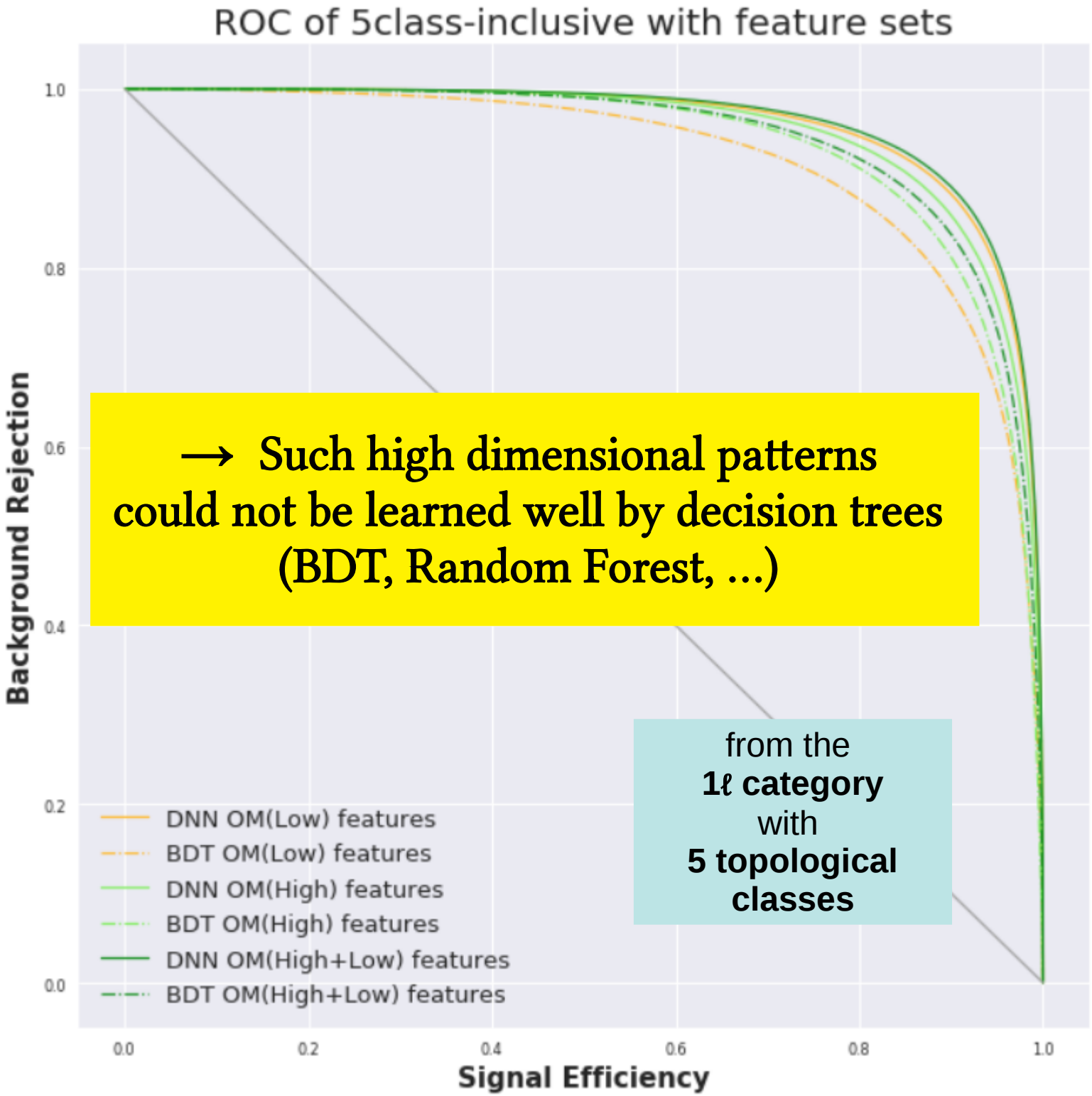




auged HL vs auged RL

ML with augmented raw-level(RL) features (blue) in n_{dim}(~200)
(shows better performance than)
ML with augmented high-level(HL) (green) in n_{dim}(~50)




ROC of 5class-inclusive with feature sets
Background Rejection
Signal Efficiency
0.0
0.2
0.4
0.6
0.8
1.0
DNN OM(Low) features
BDT OM(Low) features
DNN OM(High) features
BDT OM(High) features
DNN OM(High+Low) features
BDT OM(High+Low) features
→ Such high dimensional patterns could not be learned well by decision trees (BDT, Random Forest, ...)
from the
1ℓ category
with
5 topological classes

Conclusion

- **Di-Higgs searches in 2b+ 2L($n_{\ell=e,\mu}$ + $n_{\tau(h)}$=2)+MET channels**

  hh $\longrightarrow$ bbWW & bb$\tau_l\tau_l$ VS ttbar

- **Kinematically distinctive processes in S & B**

  7/5/2 topological classes (for 2ℓ / 1ℓ+$1\tau_h$ / $2\tau_h$ categories) by the tau decay kinematics

- **Augmented HL & RL features for each topological class**

  HL: OptiMasses & Compatibility Distances from it [Ndim~50]

  RL: dR, dPhi, Mt, Pt, … of missing d.o.f [Ndim~200]

Conclusion

- **NN discriminative model**

  can be trained well, even with raw-level augmented features in large dim. space (~200).

  1) (augmented only) > (visible only)

  2) (augmented RL) > (augmented HL)

  3) (augmented + visible) : the best

  $\longrightarrow$ **The amount of data for relevant analysis can be reduced to ~ 1/3 for a wide range of signal efficiency (2L = 2ℓ case).**

Our Challenges & Etc

- **Design of a smarter neural network structure**

  which can learn the hidden symmetries / patterns in the data,

  1) **with less amount of NN capacity**

  2) **with less amount of data**

- **Natural scientists must be good candidates for doing it**

  1) with their domain data, (tend to) possessing rather robust mathematical symmetries (than for daily life models)

  2) from diverse fields in the nature.

ML for the NS ↔ NS for the ML !

Danke schön !