Machine Learning with Topological Augmentation for boosting di-Higgs searches at the LHC

W. S. Cho

(Seoul National University)

in collaboration with

Ka Young Ban, Dongsub Lee, Chan Beom Park, Jae-hyeon Park & Seong Chan Park

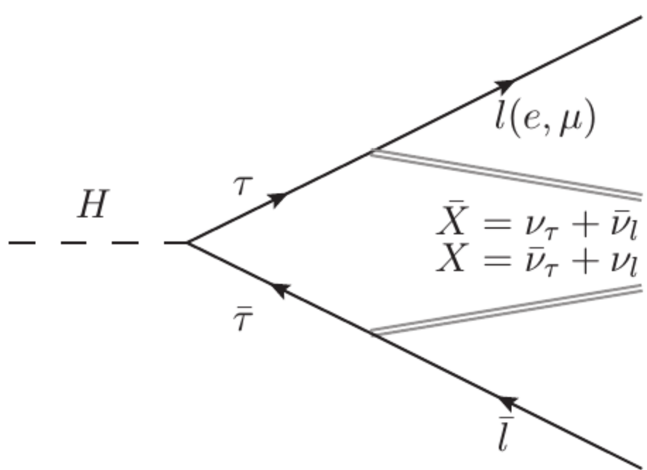
Higgs Couplings 2018 in Tokyo

Elusive di-Higgs production and decays
with multiple invisible particles

- ✓ $\sigma(gg \longrightarrow hh) = 39.64\ {}^{+4.4}_{-6.0(\mathrm{Scale})}\ {}^{+-2.1}_{(\mathrm{PDF})}\ {}^{+-2.2}_{(\alpha s)}$ fb @ [14TeV, m_h=125 GeV]

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

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



ex) H⟶ ττ ⟶ 2ℓ + MET with 4 neutrinos

HH ⟶ bbWW & bbττ on ttbar BG

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

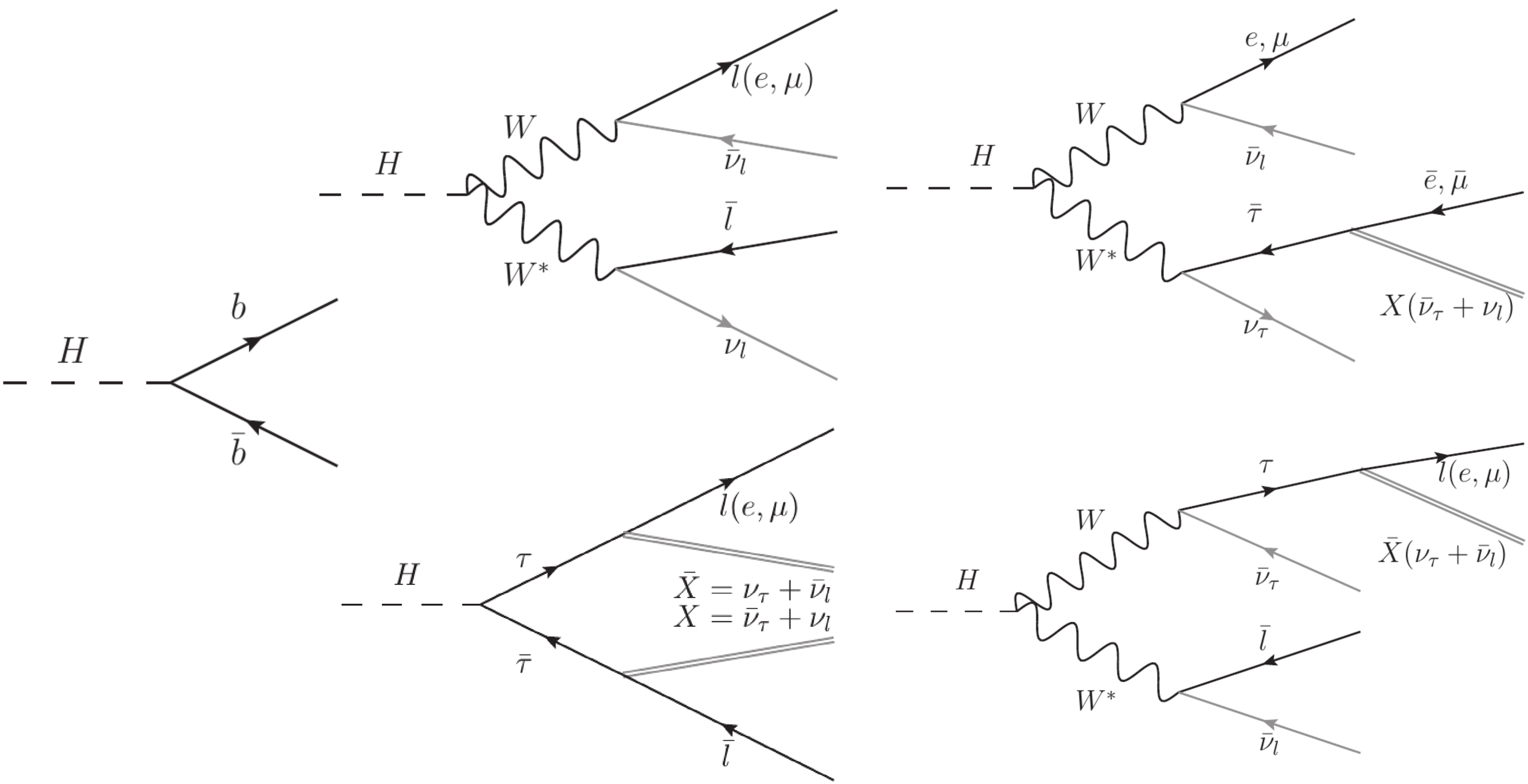
categorized further by tau decay kinematics

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 next week

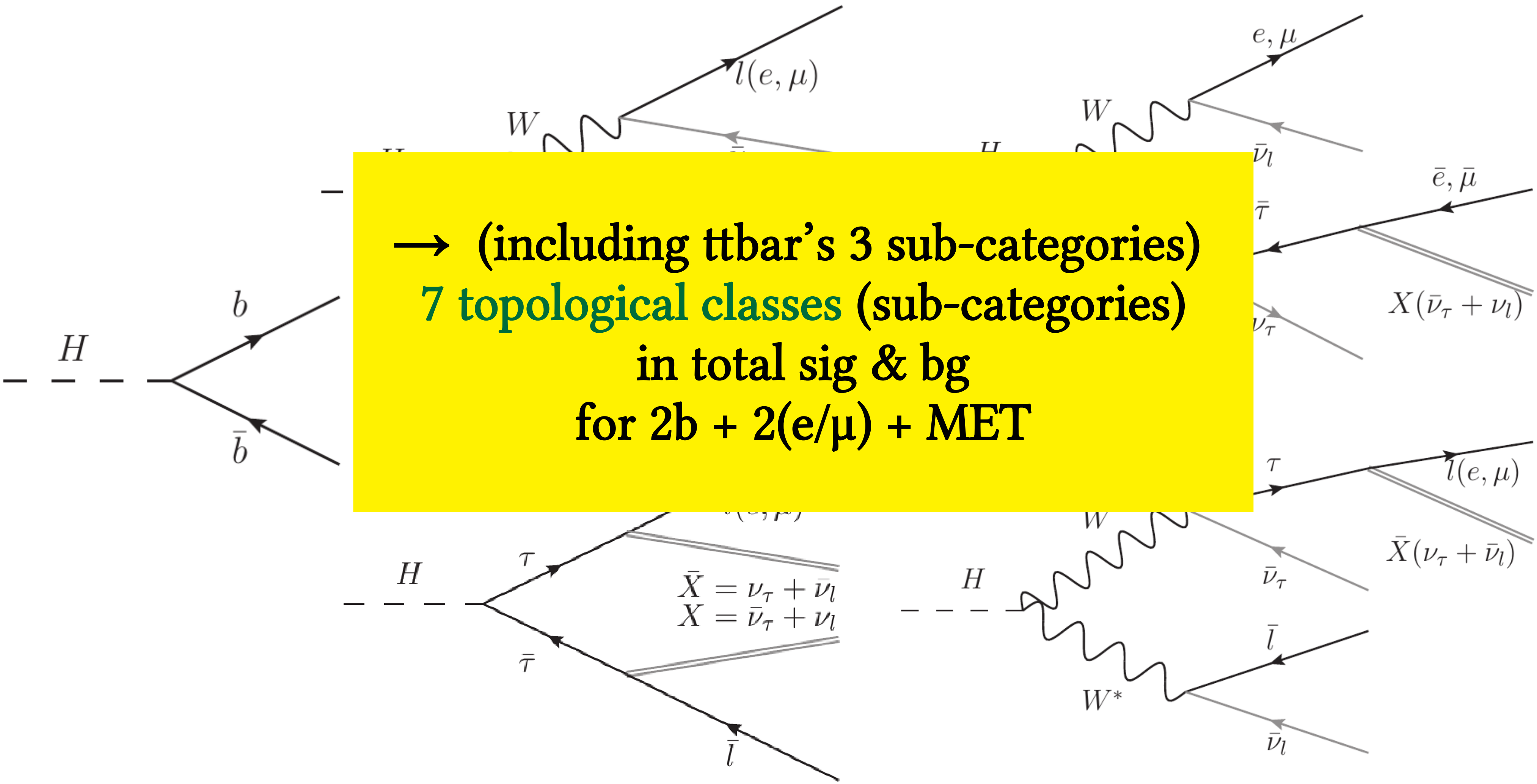
ex) 3rd cartegory : HH ⟶ 2b + 2(e/μ) + MET (4 sub-categories)

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

Building Machine learning HEP event classifiers
for the processes with large missing information

- **Machines, greedy on (HEP) data, eating and digesting data @ any level …**

$\longrightarrow$ 'DNN(deep neural networks) show similar performances even when trained only with raw-level features' [P. Baldi, P. Sadowski, D. Whiteson]

- Not even with DNN, ML have accomplished great success in HEP, including the Higgs discovery and etc, using (conventional) visible feature variables.

….

- “Okay, then do we lose our jobs?”

- “No. Not yet.”

- “Then, what should we do?”

- “Just do your best to feed more delicious data to the machines.”

- “How can we feed better data to machines? (especially for better HEP event classifiers?)”

- “Look, machines must STILL suffer from serious feature deficiency for the process data with large missing information… so, just find out and feed the entire missing part.”

- “It might not be possible..in many cases.”

- “Then, just guess the most plausible missing data (on MET) under various hypothetical models & feed all of them to machines, they will be able to eat all of them and digest, though! ”

- “It sounds not so smart, but let’s just try the `**machine learning with augmentation**’ , and let’s see. ”

….

(Topological) Augmentation of invisible missing momenta in under-constrained systems using OptiMass[ref]

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)

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

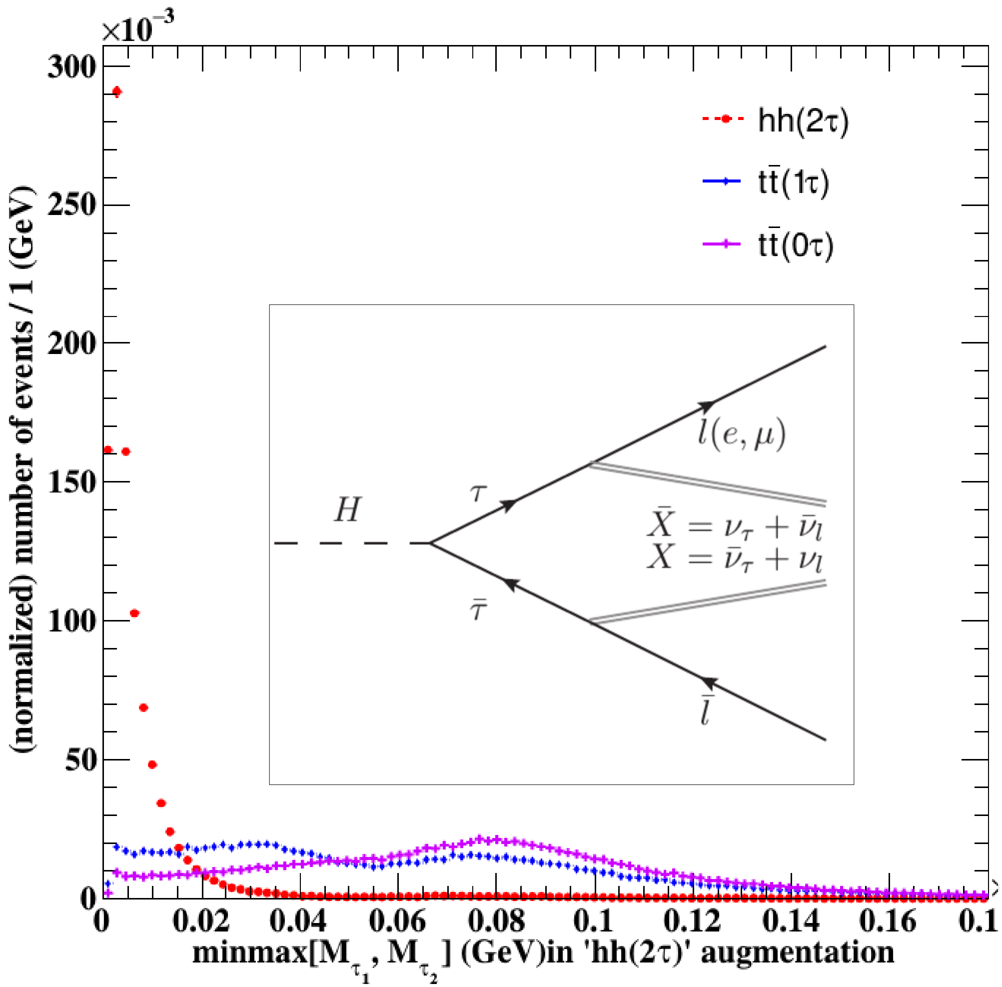
via the optimization of Augmented Lagrangian :

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

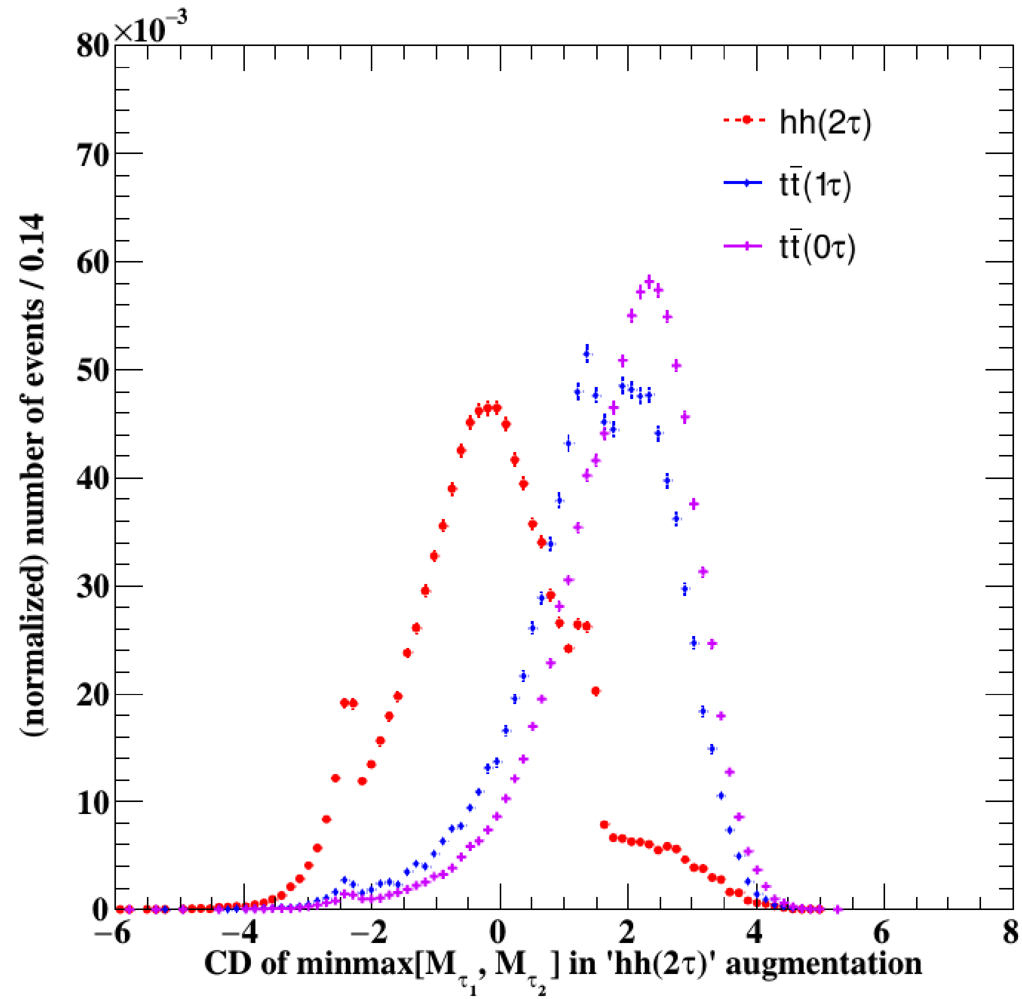
Augmentation with hh(sig) & ttbar(bg) processes

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

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

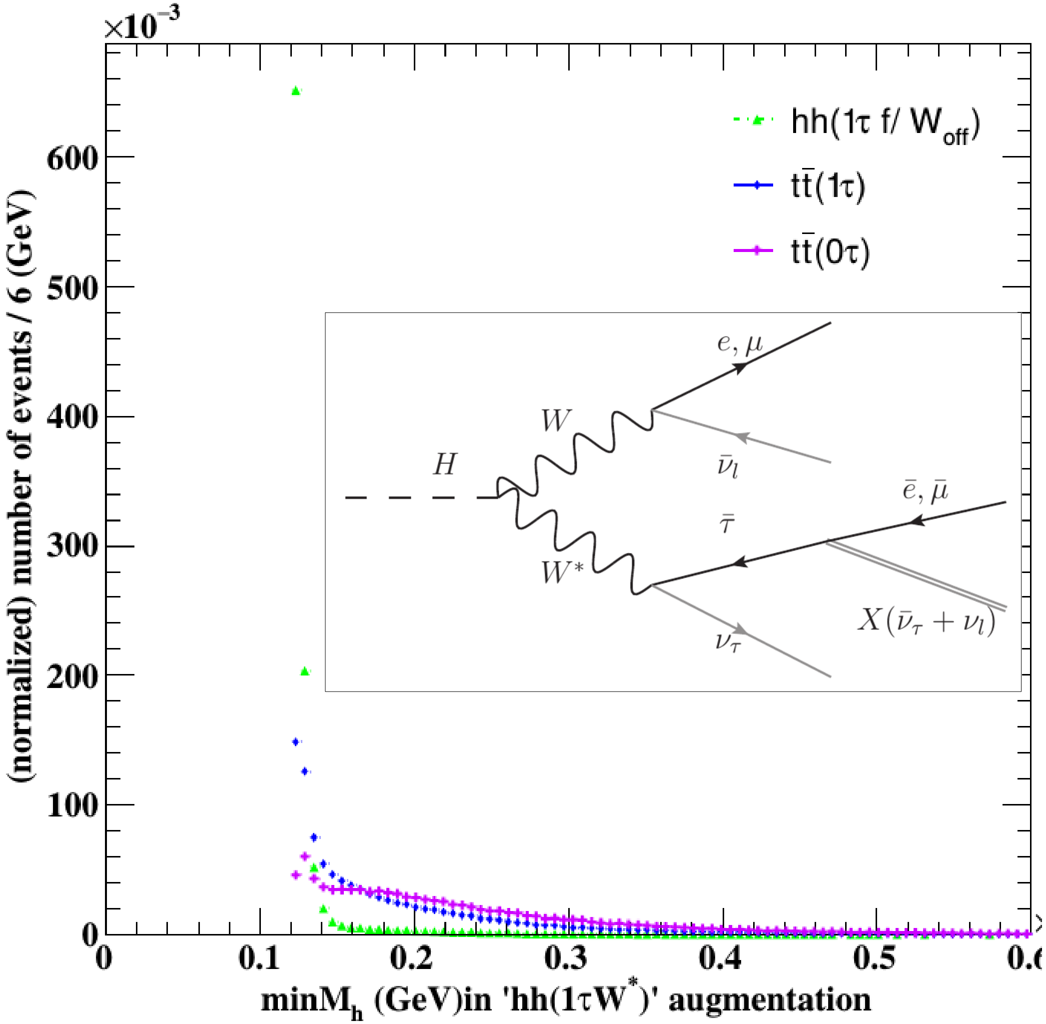


Com. Dist. of 3 procs (Sig & Bg) in HH(bbττ) augmentation

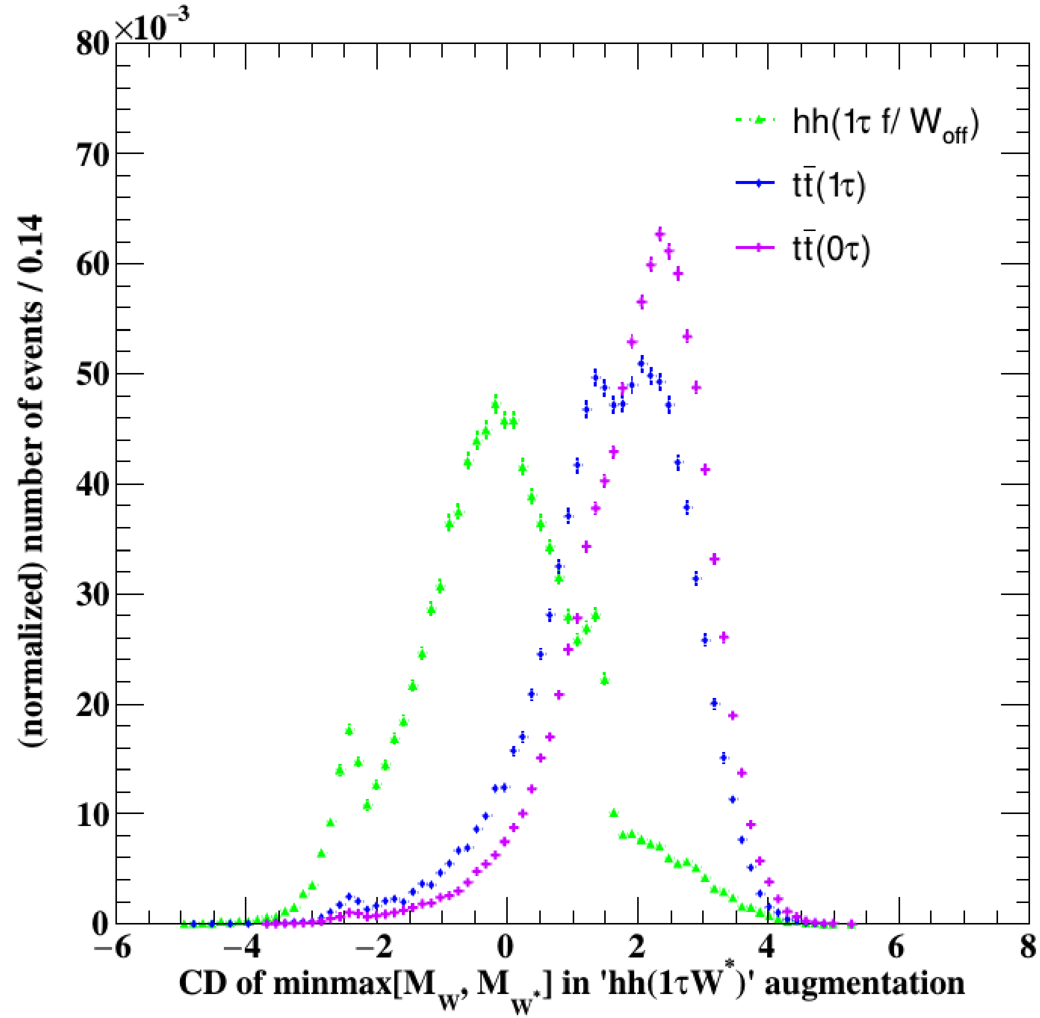


Augmented features 1 : HL (optimass + compatibility distance) [ex2]

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

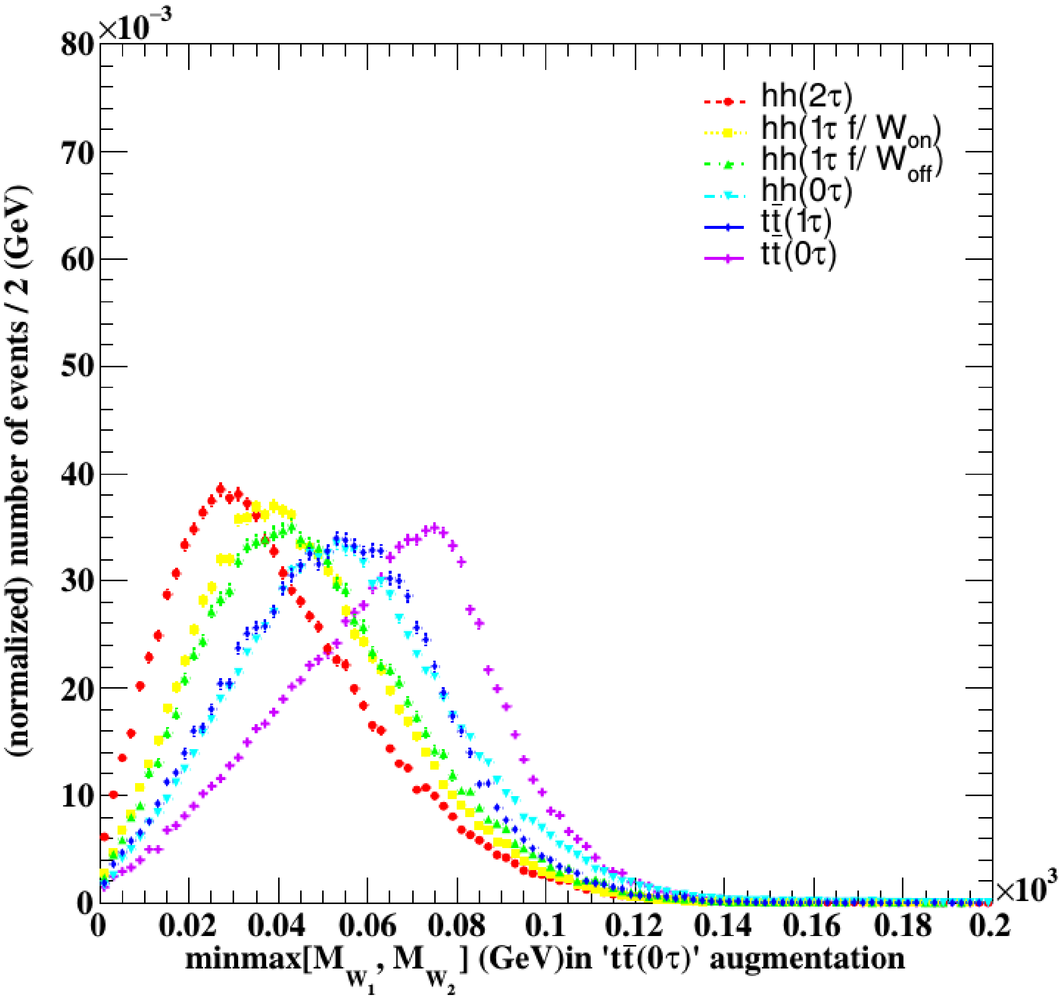


Com. Dist. of 3 procs (Sig & Bg) in HH(bbWW*) augmentation

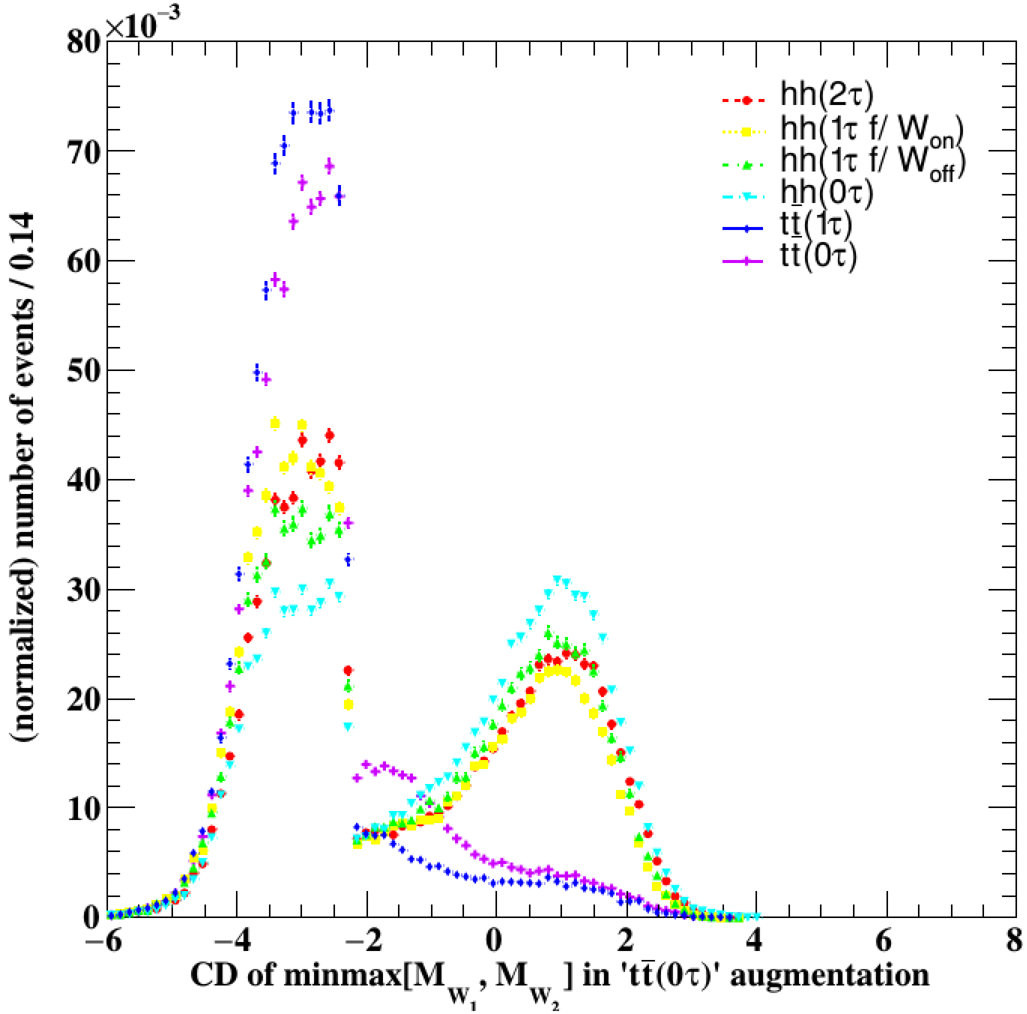


Augmented features 1 : HL (optimass + compatibility distance) [ex3]

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

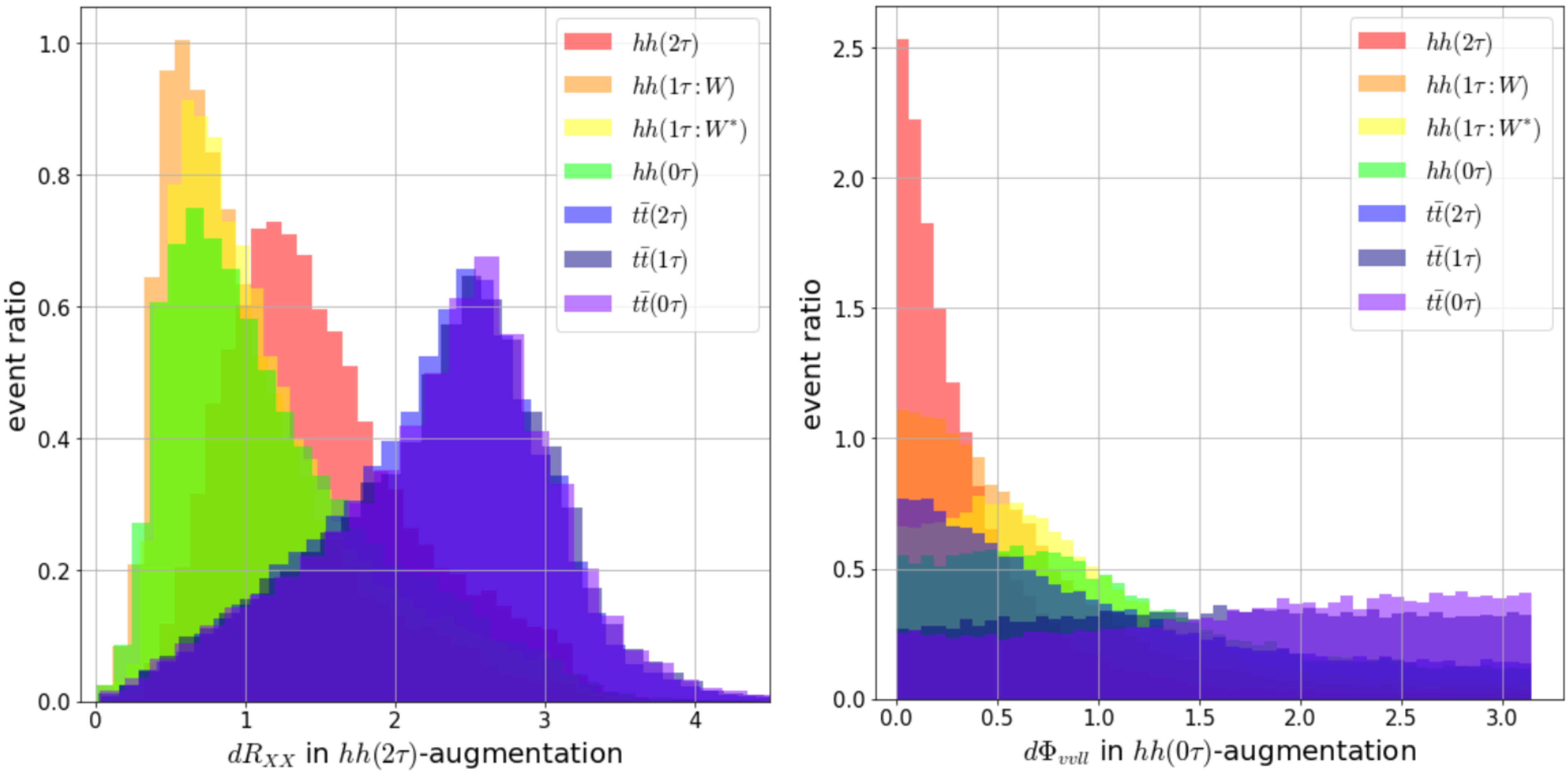


Com. Dist. of 3 procs (Sig & Bg) in ttbar(0tau) augmentation



Augmented Features 2 : raw level variables

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



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

from

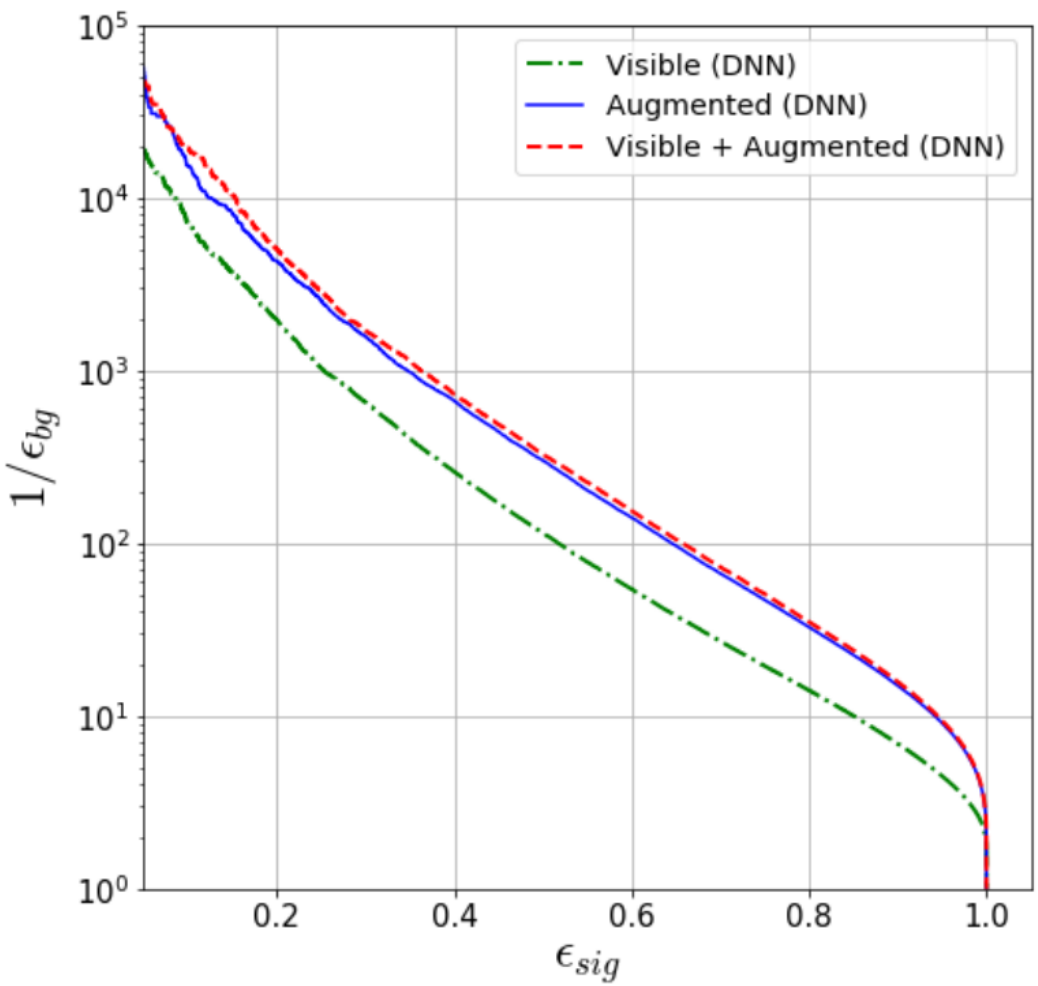
(subsystem optimass) x (constraint profile)
----
for (a given augmentation model)

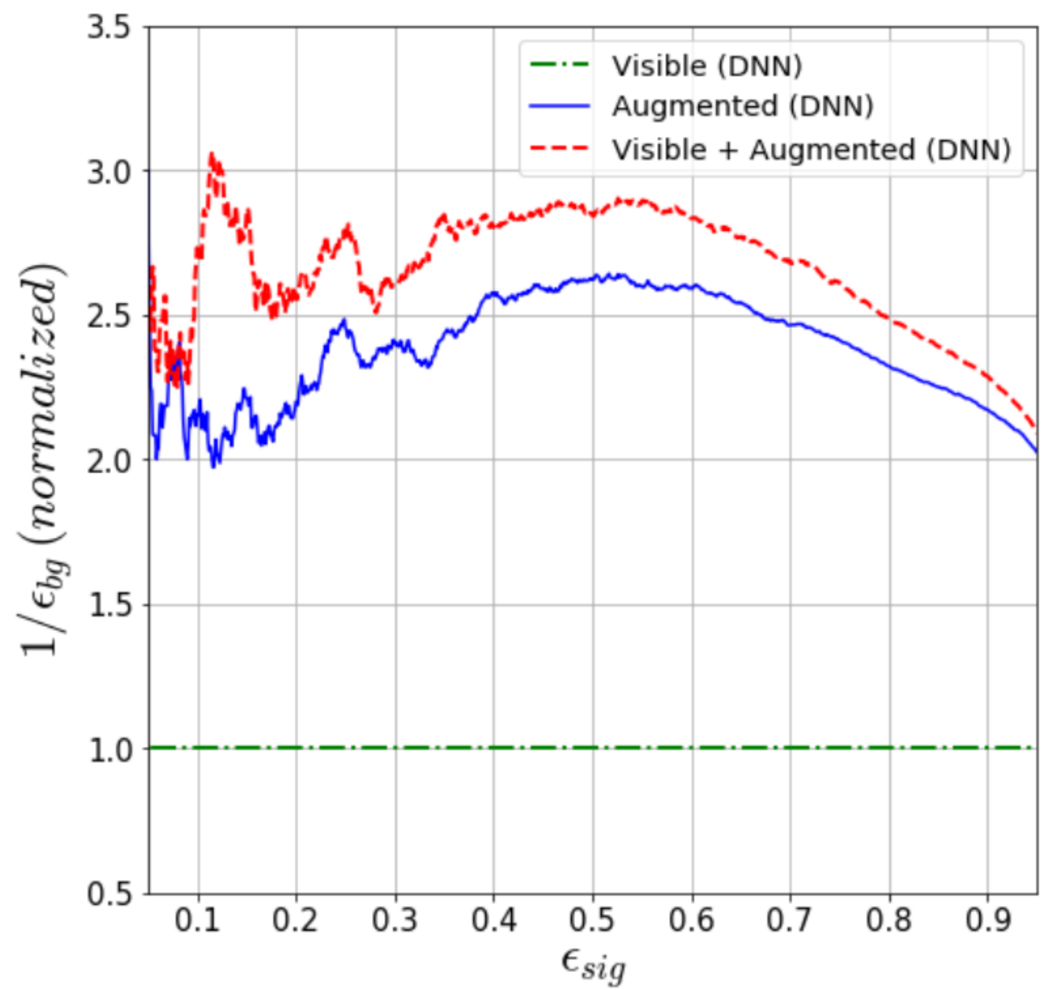
Supervising DNN classifier with the augmented features (2L=2l case)

[Feature variable set definition]

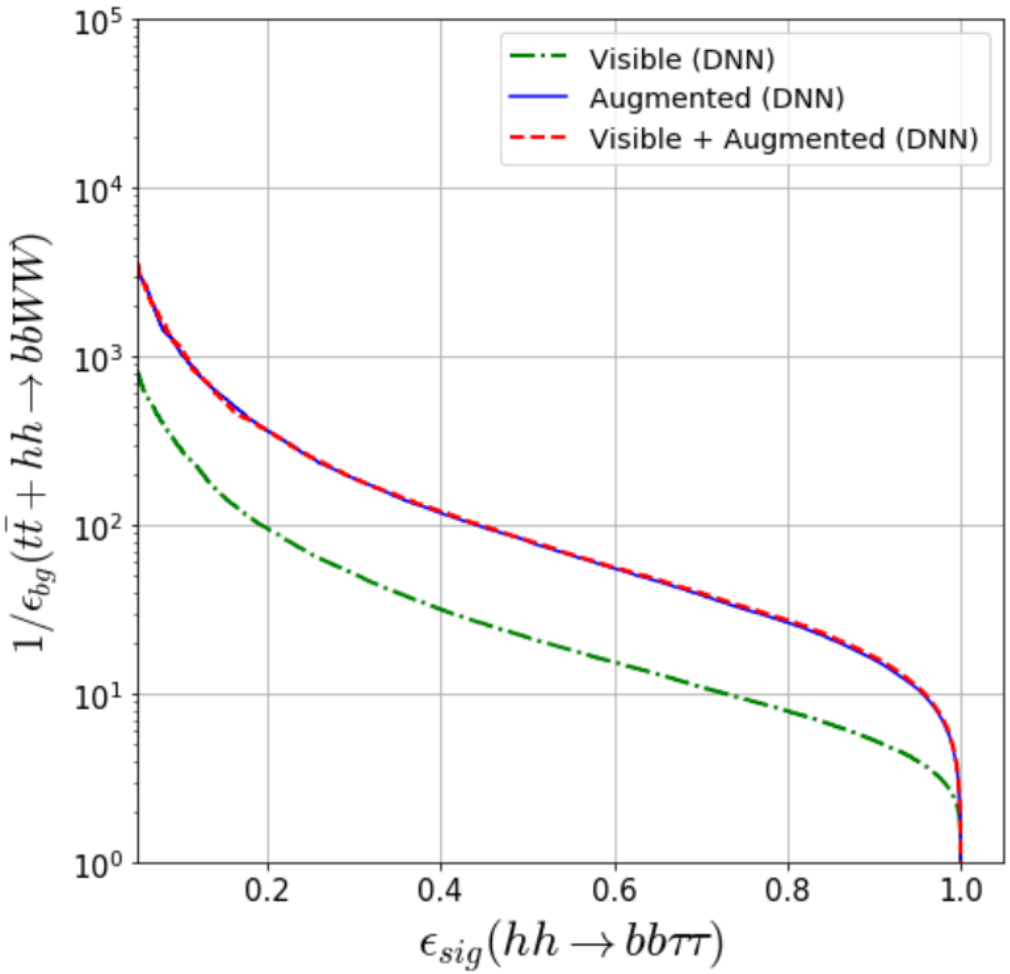
'**Visible**' : variable sets used in CMS PAS 17-006

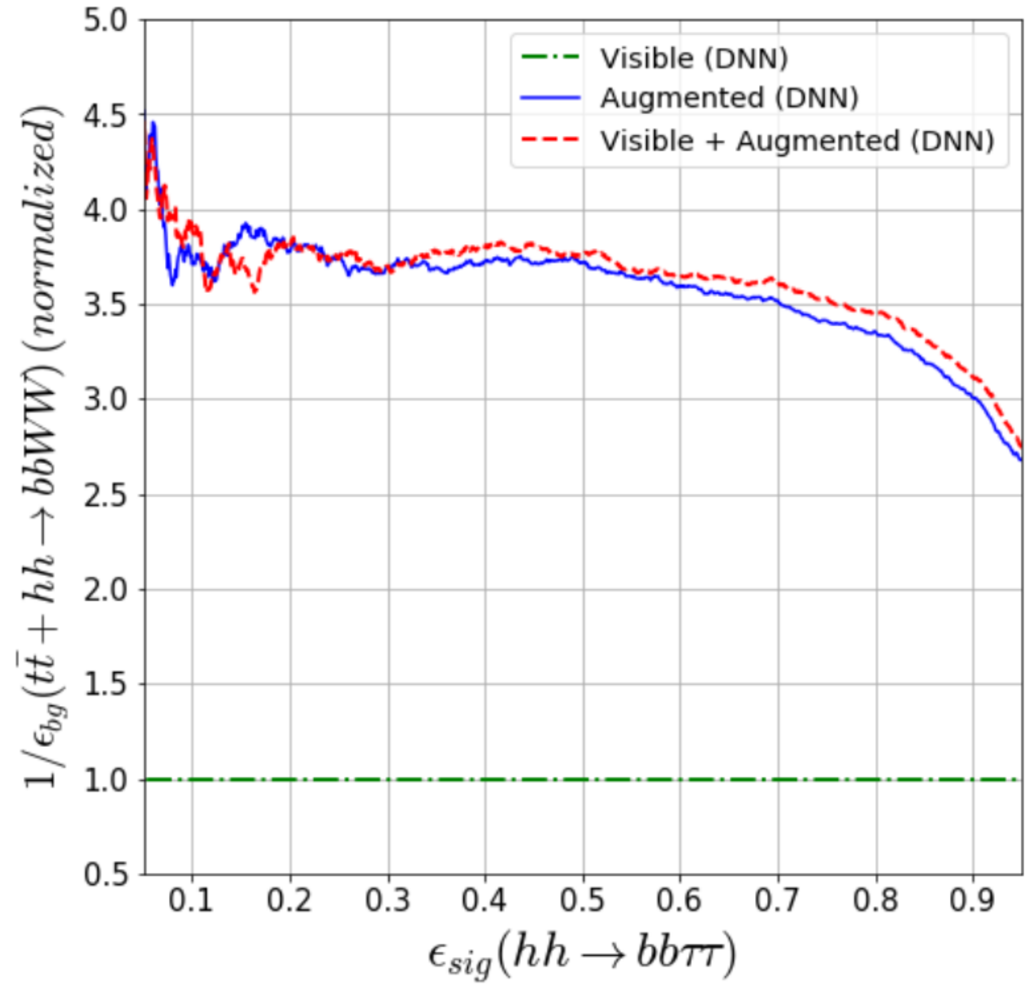
'**Augmented**' : augmented HL [Ndim~50] + RL [Ndim ~ 200]



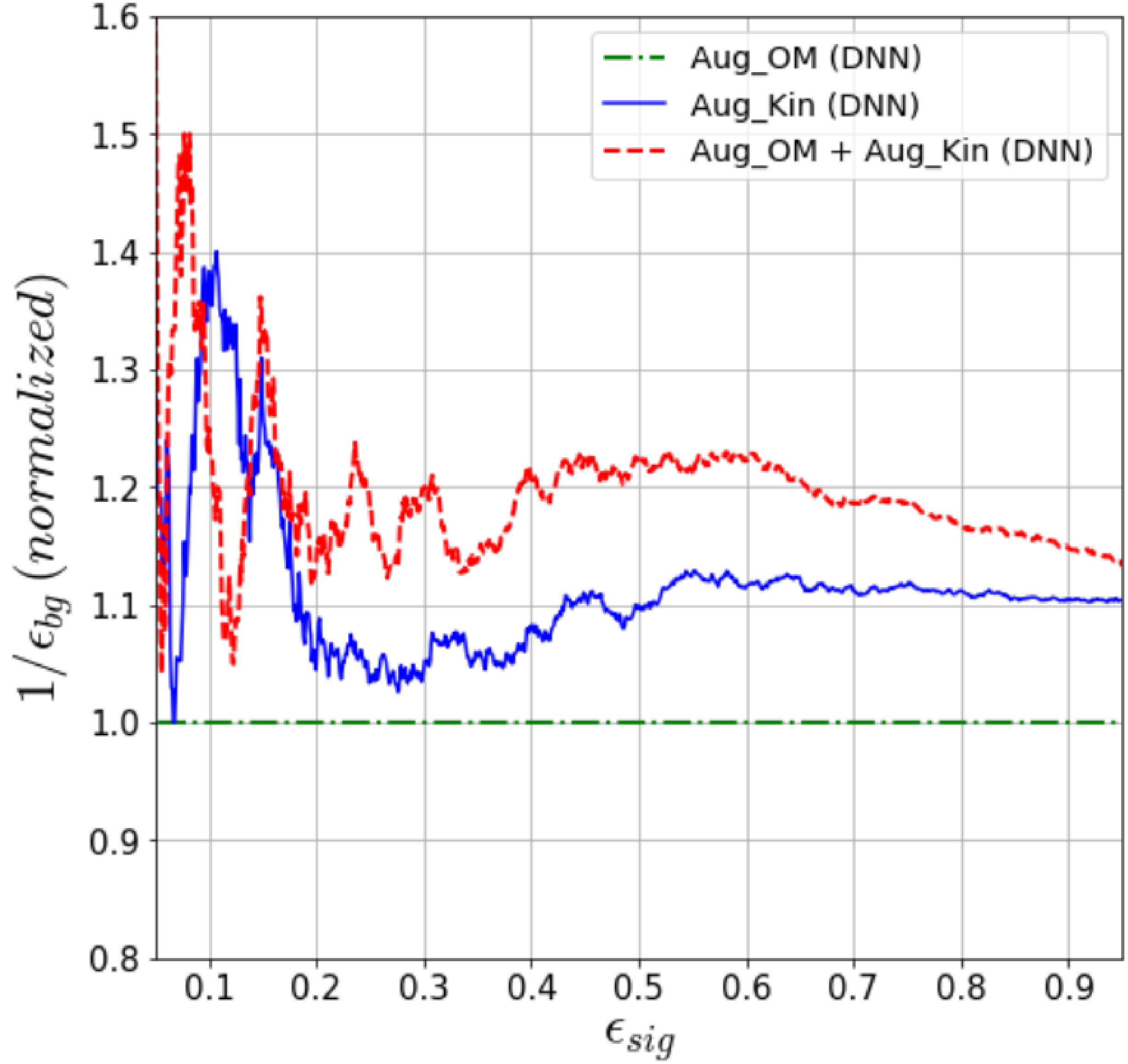


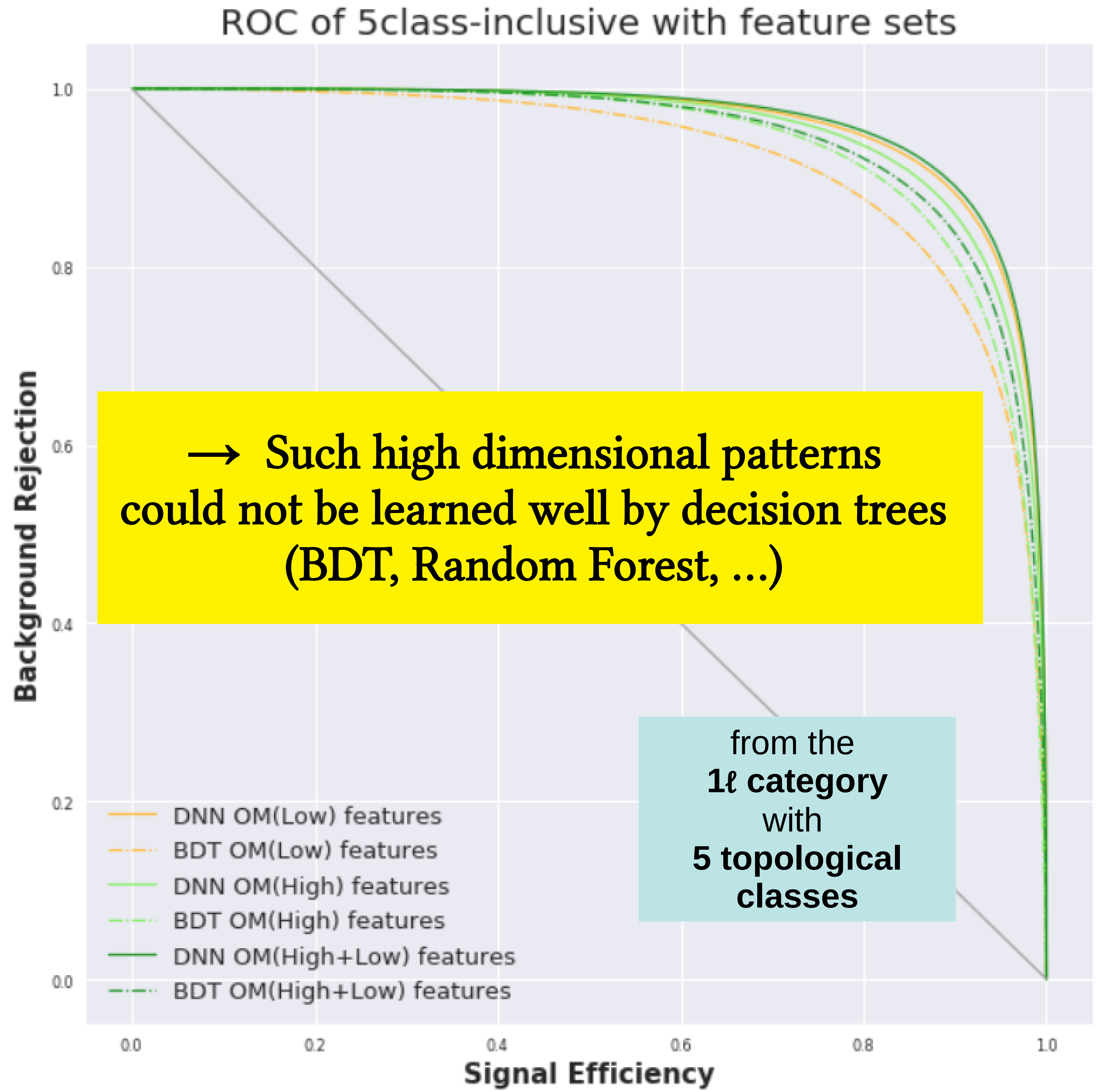
DNN classifier for hh(bbττ) vs others(ttbar + hh(bbWW)) shows more enhanced significance





DNN classifiers can even be trained better using augmented raw-level(RL) features in n_{dim}(~200) >> n_{dim}(HL: ~50)





→ Such high dimensional patterns could not be learned well by decision trees (BDT, Random Forest, ...)

from the **1ℓ category** with **5 topological classes**

+ lots of new results
in the talk
by K. Y. Ban [Pheno@KEK] next week
including more on
0ℓ & 1ℓ categorical signatures !

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 / $1\ell+1\tau_h$ / $2\tau_h$ categories) by the tau decay kinematics

- **Augmented high-level & raw-level features for each topological classes**

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

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

- **Supervising DNN classifiers**

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

  (augmented only) >> (visible only).

  (augmented + visible) : the best.

  $\longrightarrow$ **Bg efficiency can be reduced to ~ 1/3 for a wide range of sig efficiency (2L = 2ℓ case).**

Thank you