A CASE STUDY ON $b\bar{b}h$ ASSOCIATED PRODUCTION

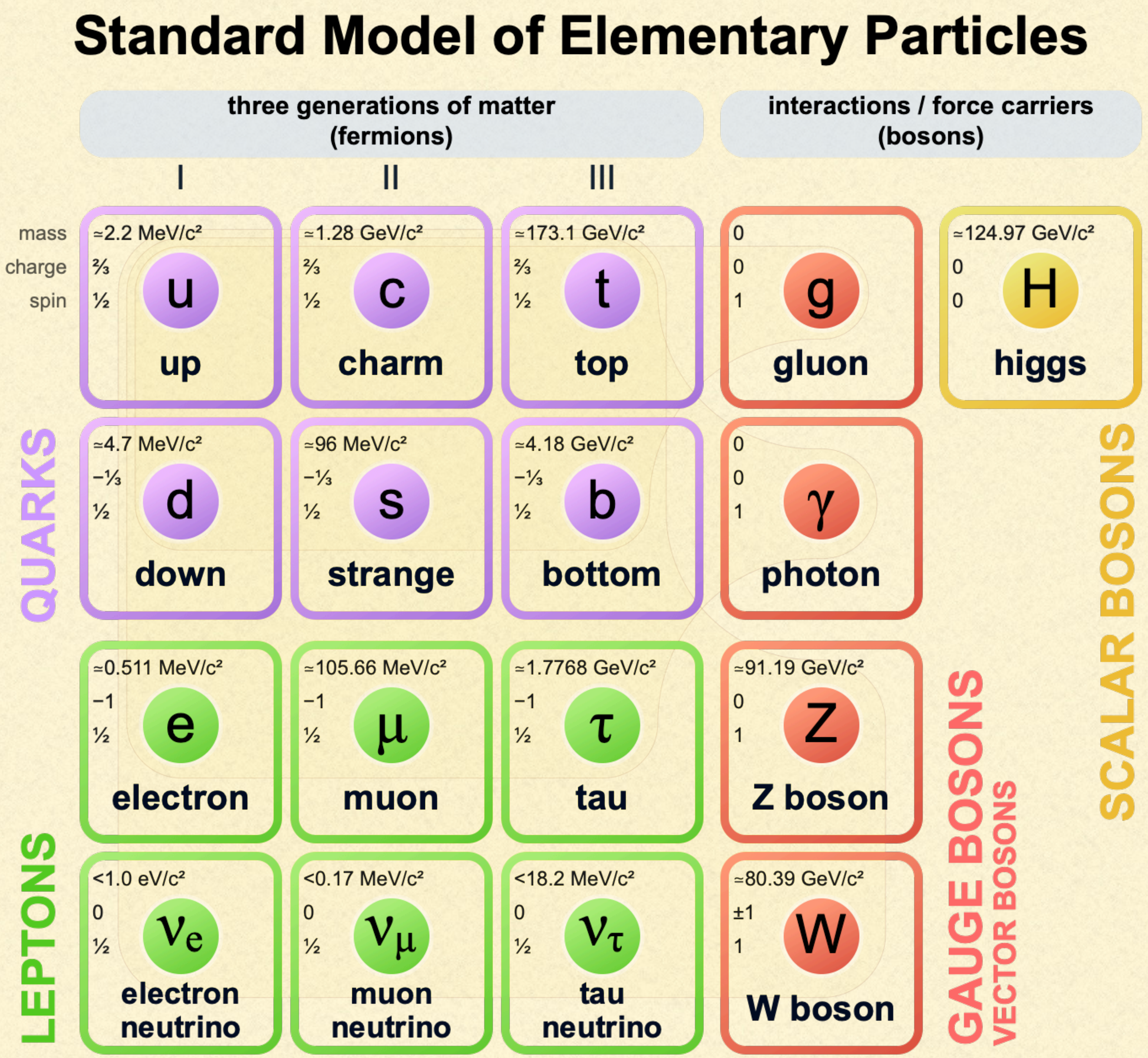
its physics implication, and machine interpretation

HPNP, March 25th 2021

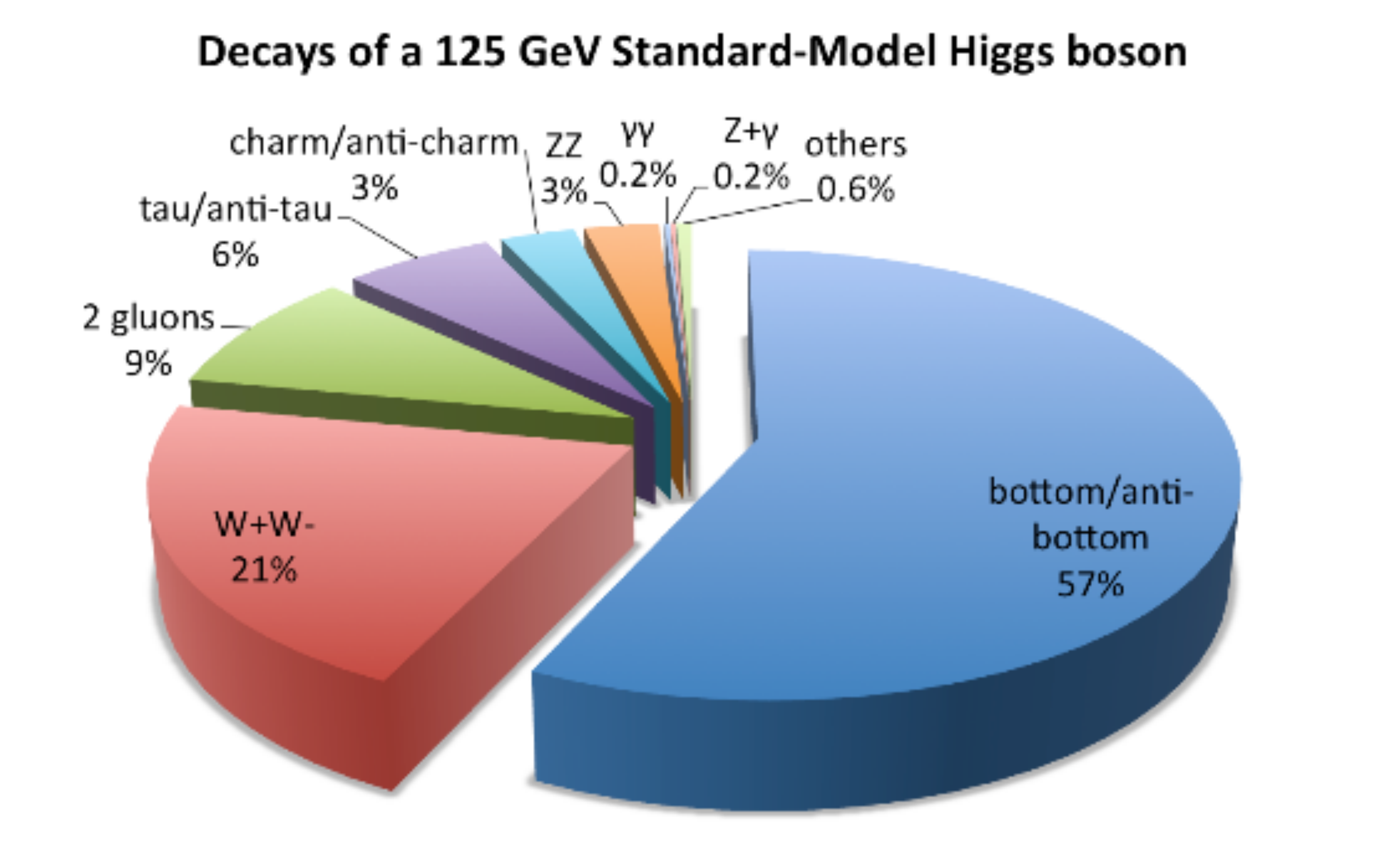
Zhuoni Qian

Based on work with Christophe Grojean, Ayan Paul (arXiv: 2011.13945)

Higgs couplings and $b\bar{b}h$

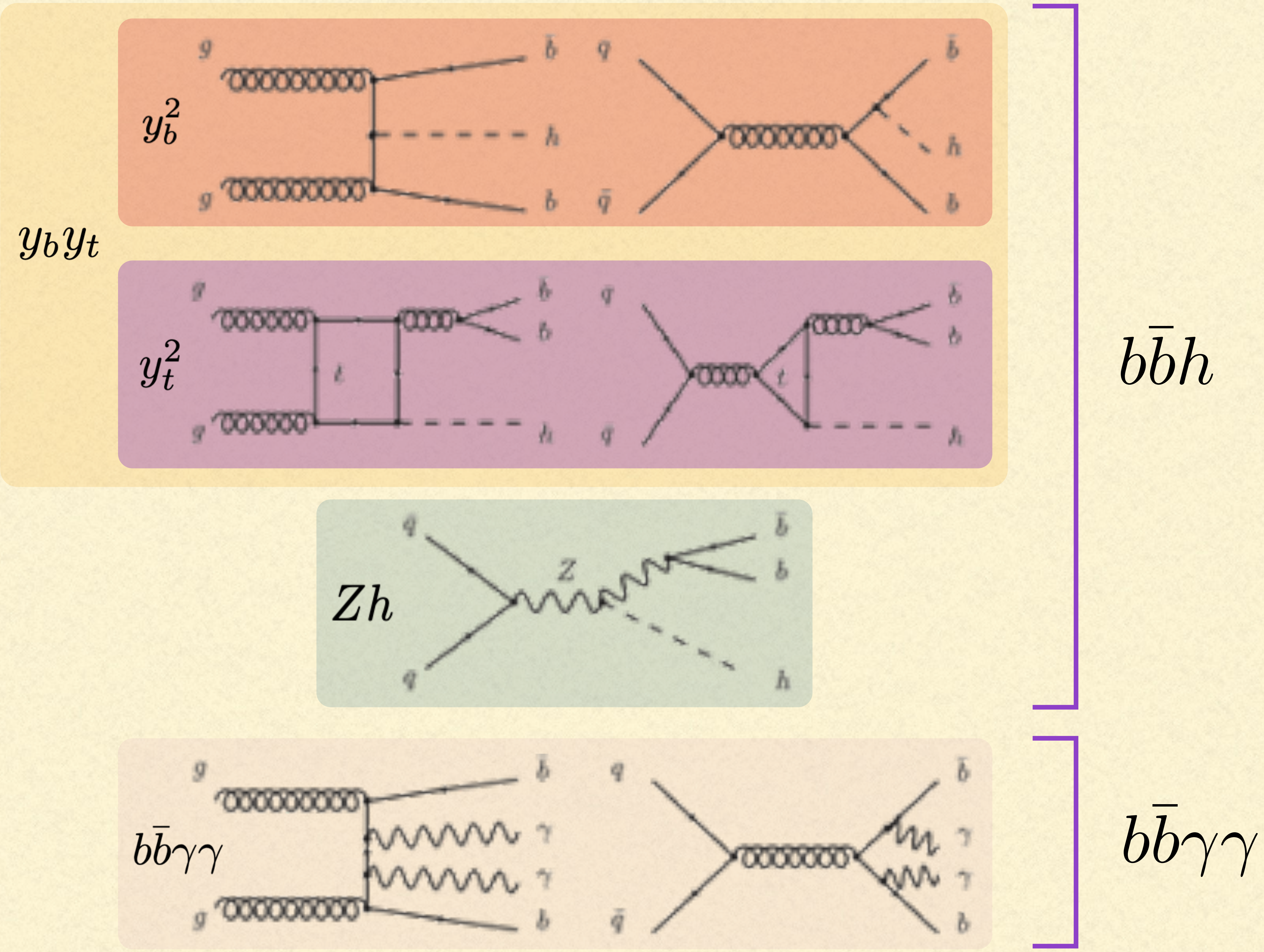


- Bottom Yukawa measurement is a recent achievement: $Vh, h \to b\bar{b}$
- The sign (or phase) of the Yukawa couplings have not been well measured
- There are possible interplays between Yukawa phases of various couplings in EDM measurements and collider physics



(Multi-channel) Signal embedded in large backgrounds

Goal: measure bottom-Yukawa couplings



Traditional cut-based analysis cannot separate the different $b\bar{b}h$ contributions – no y_b sensitivity at HL-LHC

Basic selection (14 TeV HL-LHC):

signals

Channel	LO σ (fb)	NLO-k-fact	6 ab^{-1} [#evt]	$2b$-jets[%]
y_b^2	0.0648	1.5	583	7.7%
$y_b y_t$	-0.00829	1.9	-95	4.0%
y_t^2	0.123	2.5	1,840	12%
Zh	0.0827	1.3	645	21%
$\sum b\bar{b}h$	0.262	-	2,970	-
$b\bar{b}\gamma\gamma$	12.9	1.5	116,000	14%

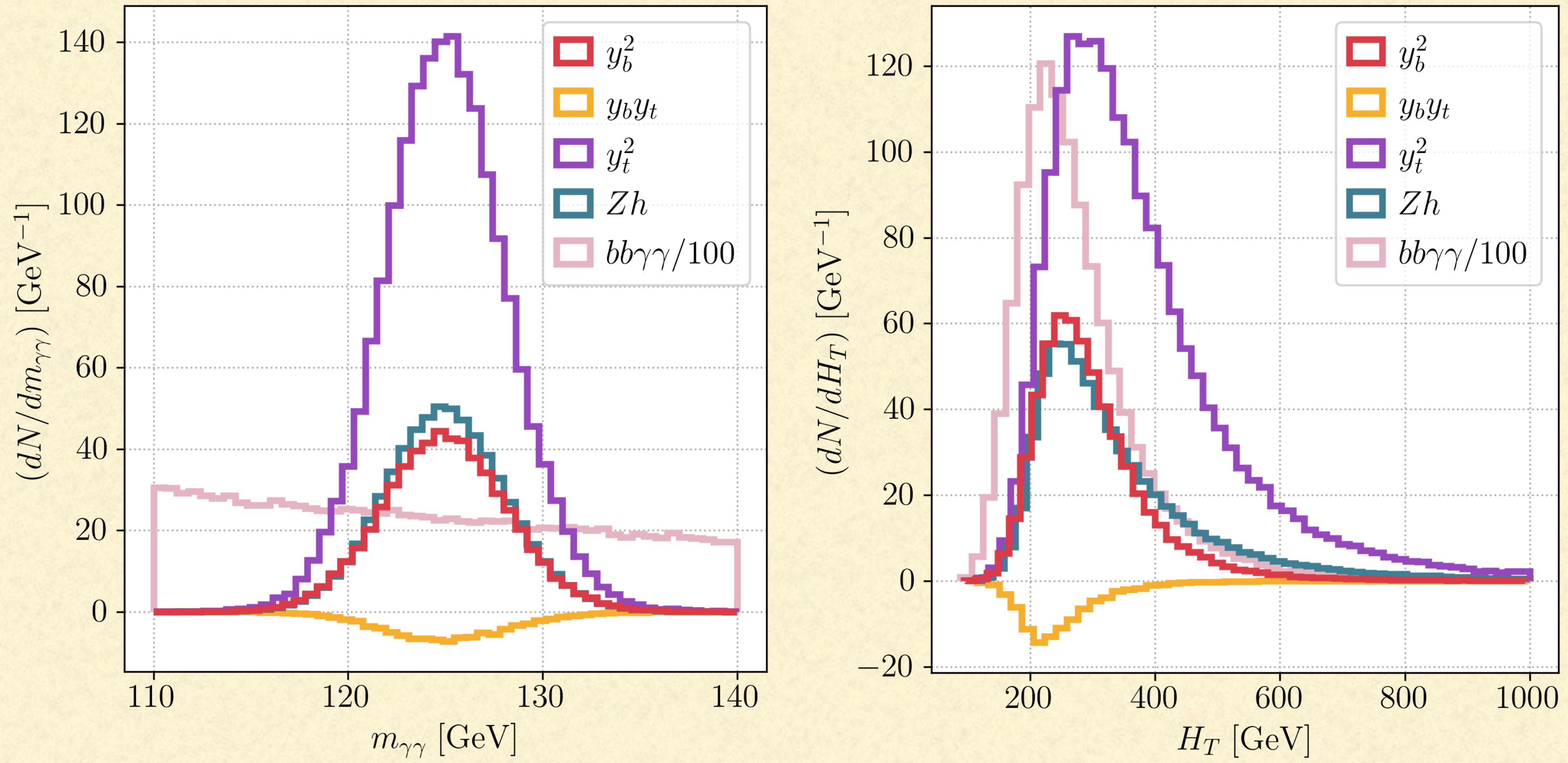
$b\bar{b}h$ background QCD-QED background

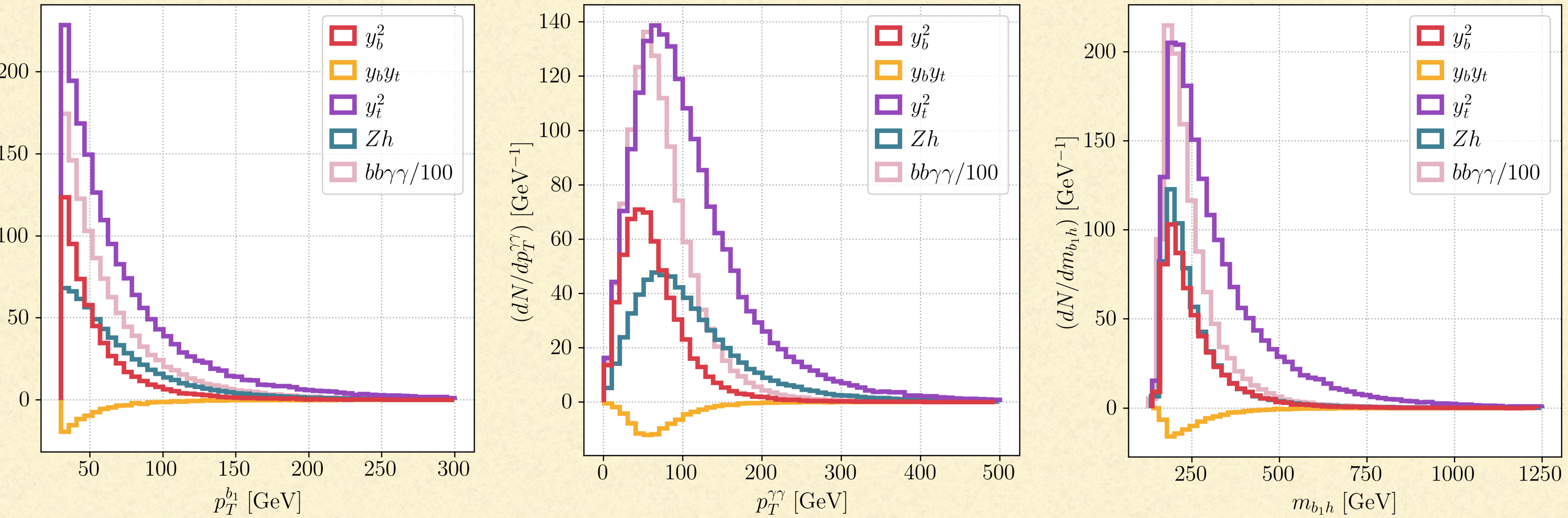
$$p_T^{bjet} > 30\,\mathrm{GeV},\ p_T^{\gamma jet} > 20\,\mathrm{GeV},$$
$$\eta_{bjet,\gamma jet} < 2.5,\ 110 < m_{\gamma\gamma}\,(\mathrm{GeV}) < 140.$$

Observable and distributions:

Understanding differences in shapes

- $p_T^{b_1}, \, p_T^{b_2}, \, p_T^{\gamma_1}, \, p_T^{\gamma\gamma},$
- $\eta_{b_{j1}}, \, \eta_{b_{j2}}, \, \eta_{\gamma_1}, \, \eta_{\gamma\gamma},$
- $n_{bjet}, \, n_{jet}, \, \Delta R^{b\gamma}_{\min}, \, \Delta\phi^{bb}_{\min},$
- $m_{\gamma\gamma}, \, m_{bb}, \, m_{b_1 h}, \, m_{b\bar{b}h}, \, H_T.$

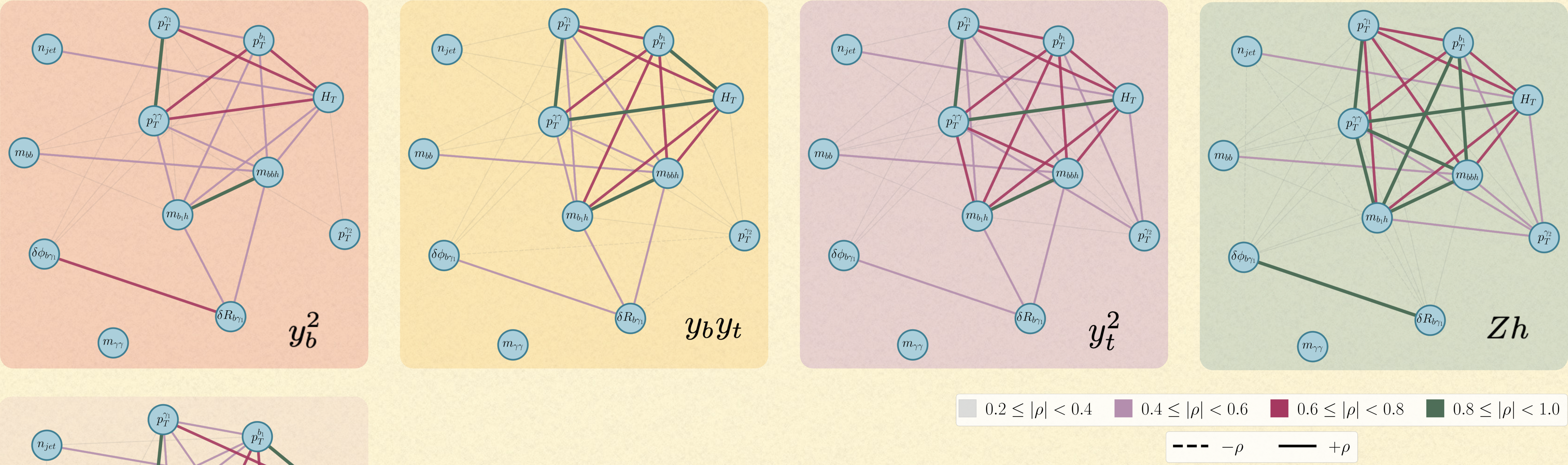


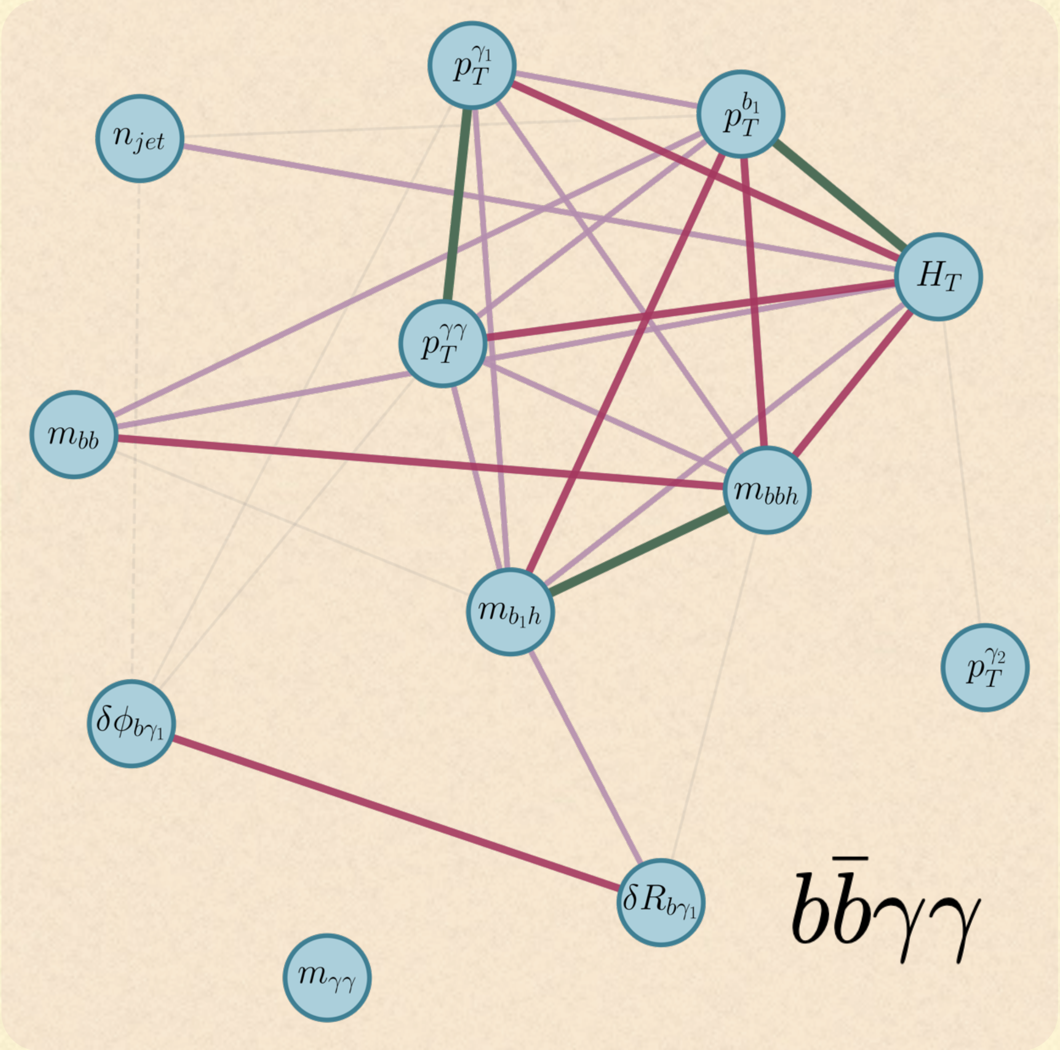


The choice of variables is important:

- Momenta four vectors are not easily interpretable
- Kinematic variables are interpretable but there is no clear “complete set”

Into higher dimensions (the correlation):





- The multi-channel multivariable correlation pattern.
- MVAs (BDT, NN) > 1D cuts <= higher order correlation
- Importance of observables measure

Analysis optimised with BDT classification:

Predicted no. of events at HL-LHC

Actual no. of events

Channel	y_b^2	$y_b y_t$	y_t^2	Zh	$bb\gamma\gamma$	total
y_b^2	170	54	51	122	189	586
$y_b y_t$	-7	-24	-4	-20	-40	-95
y_t^2	238	112	452	546	487	1,835
Zh	22	28	21	416	161	648
$bb\gamma\gamma$	2,183	2,450	151	8,045	101,591	115,779
$\mathcal{Z}_j$	3.33	0.47	10.	4.36	317	

$$\mathcal{Z}_j = \frac{|N_{jj}|}{\sqrt{\sum_i N_{ij}}}$$

About ~60% gain in significance over traditional cut-based analyses (2σ).

Predicted no. of events at FCC-hh

Actual no. of events

	y_b^2	$y_b y_t$	y_t^2	Zh	$bb\gamma\gamma$	total
y_b^2	32,074	15,112	10,966	6,579	8,959	73,690
$y_b y_t$	-964	-6,815	-907	-583	-1,820	-11,089
y_t^2	48,772	45,751	148,669	39,598	26,484	309,274
Zh	1,860	4,498	2,280	12,661	2,282	23,581
$bb\gamma\gamma$	172,088	373,436	106,335	126,429	7,952,834	8,731,122
$\mathcal{Z}_j$	63.7	10.4	288	29.4	2,813	

$$\mathcal{Z}_j = \frac{|N_{jj}|}{\sqrt{\sum_i N_{ij}}}$$

About ~60% gain in significance over traditional cut-based analyses.

Physics Interpretation: real bottom Yukawa κ-scheme $\mathcal{L} \supset -\kappa_b \frac{m_b}{v} \bar{b} b h$

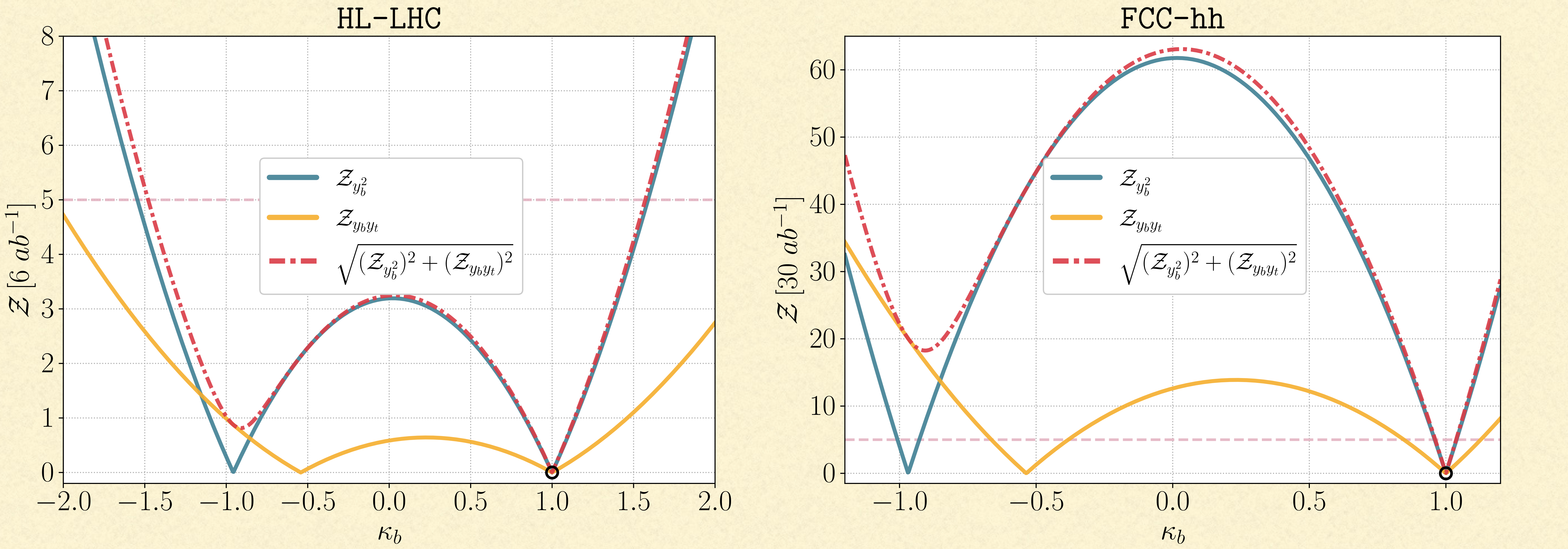
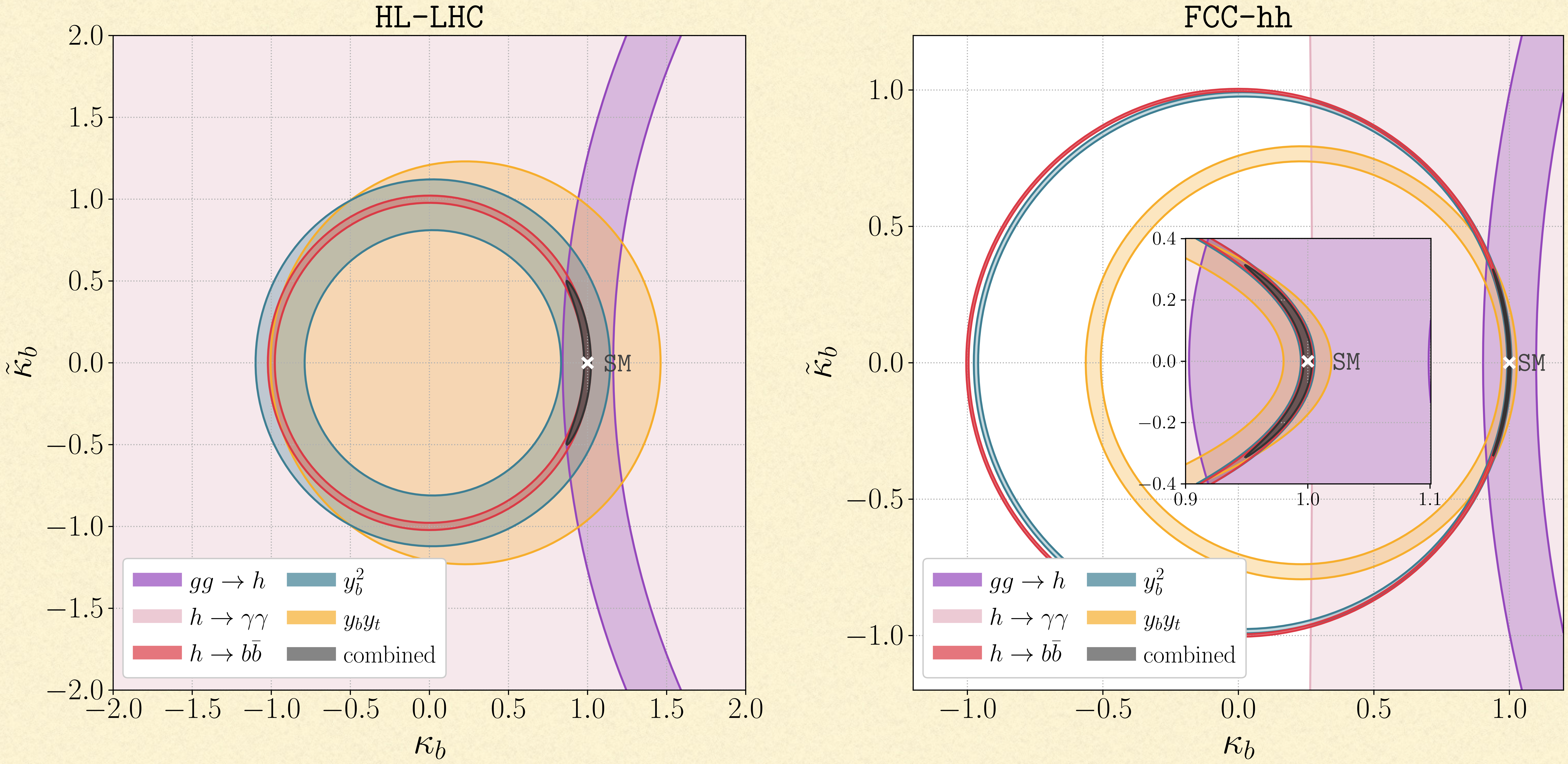


Figure 7. *Significance, $\mathcal{Z}$, as a function of κ_b at HL-LHC (ATLAS+CMS combined, 6 ab^{-1}) and FCC-hh (30 ab^{-1}). A SM signal is injected.*

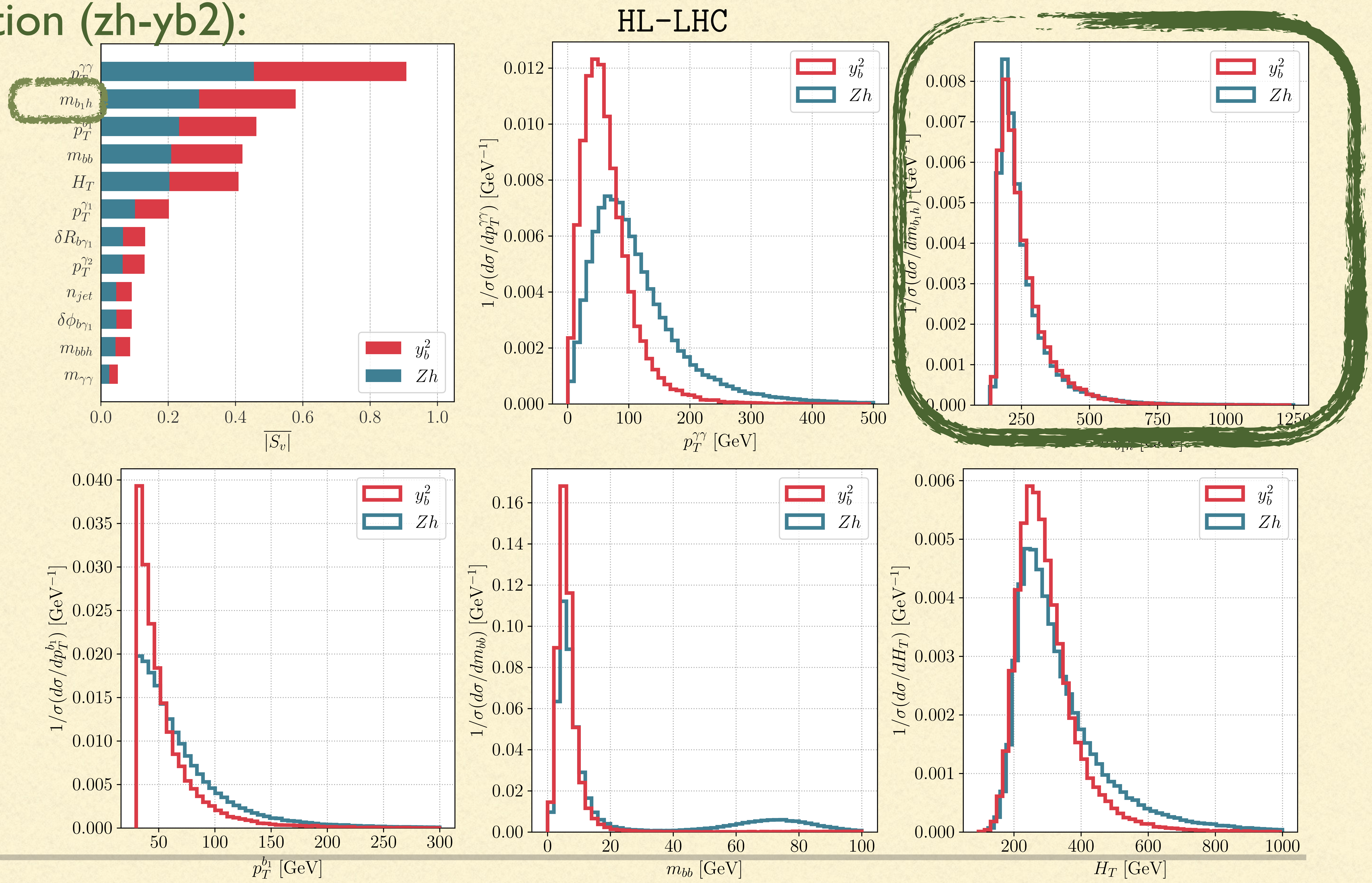
=> Unambiguous sign determination at FCC-hh.

Physics Interpretation: complex bottom Yukawa (CP-phase)

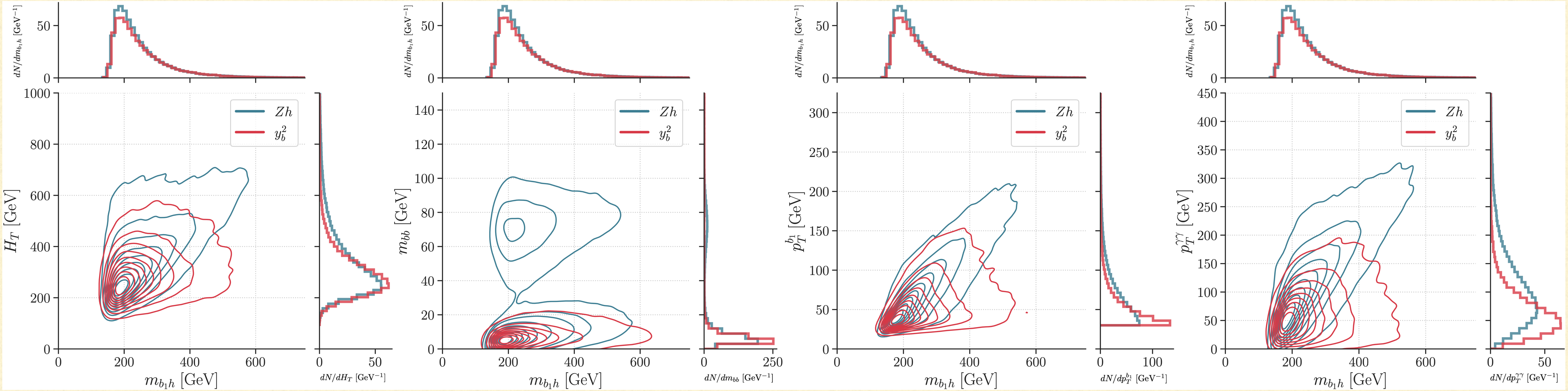
$$\mathcal{L} \supset -\frac{m_b}{v}(\kappa_b \bar{b} b + i\tilde{\kappa}_b \bar{b}\gamma_5 b)h$$



Machine Interpretation (zh-yb2):



Machine Interpretation (zh-yb2):



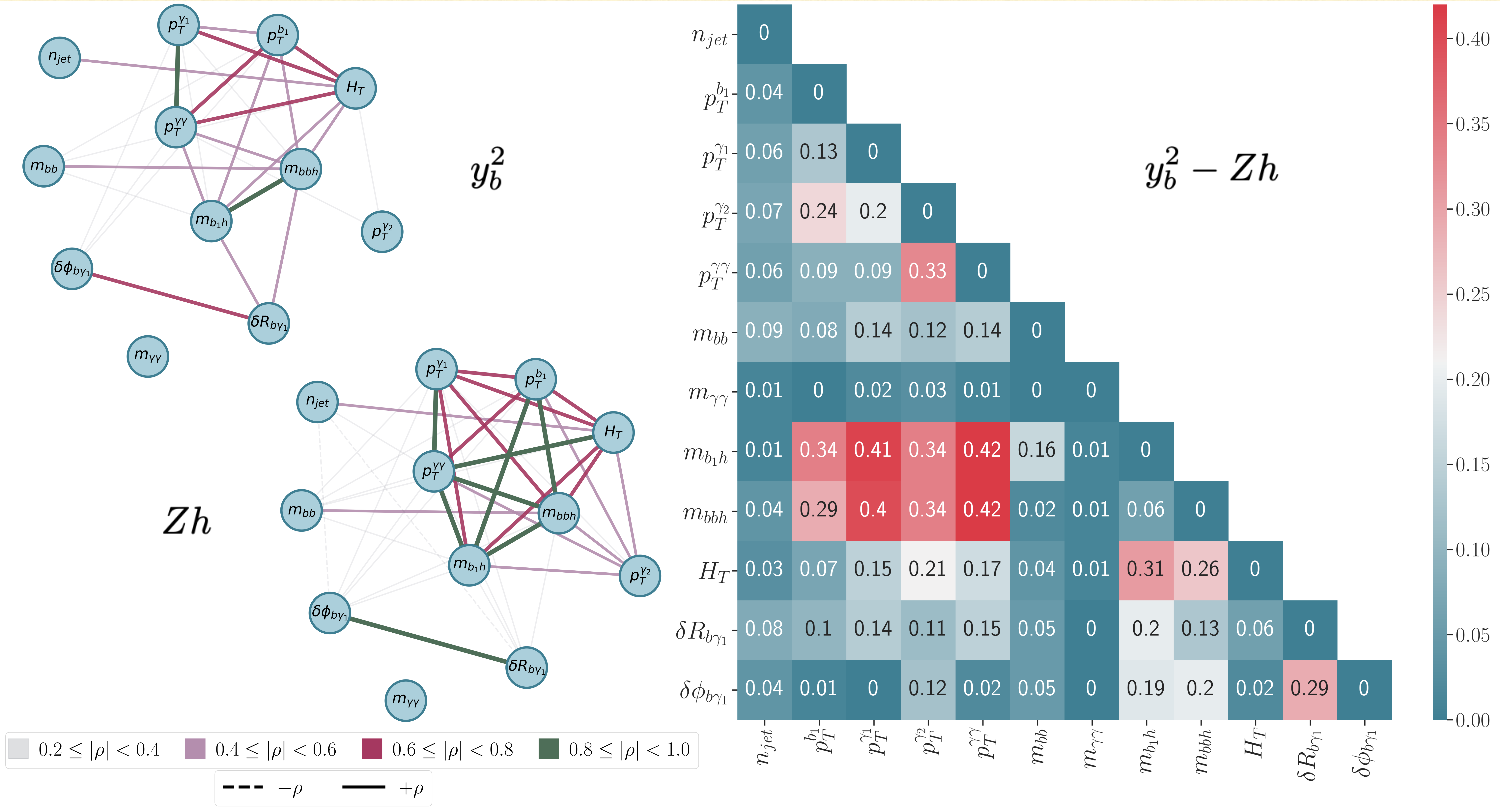
- Cut-based analyses start to falter with multivariate correlations – difficult to visualize and interpret
- Machine learning algorithms excel at multivariate analyses

- Machine learning algorithms are essentially black-boxes – not good for understanding the underlying dynamics

Machine Interpretation (zh-yb2):

"importance of m_{b1h} variable visualised through correlation"

HL-LHC

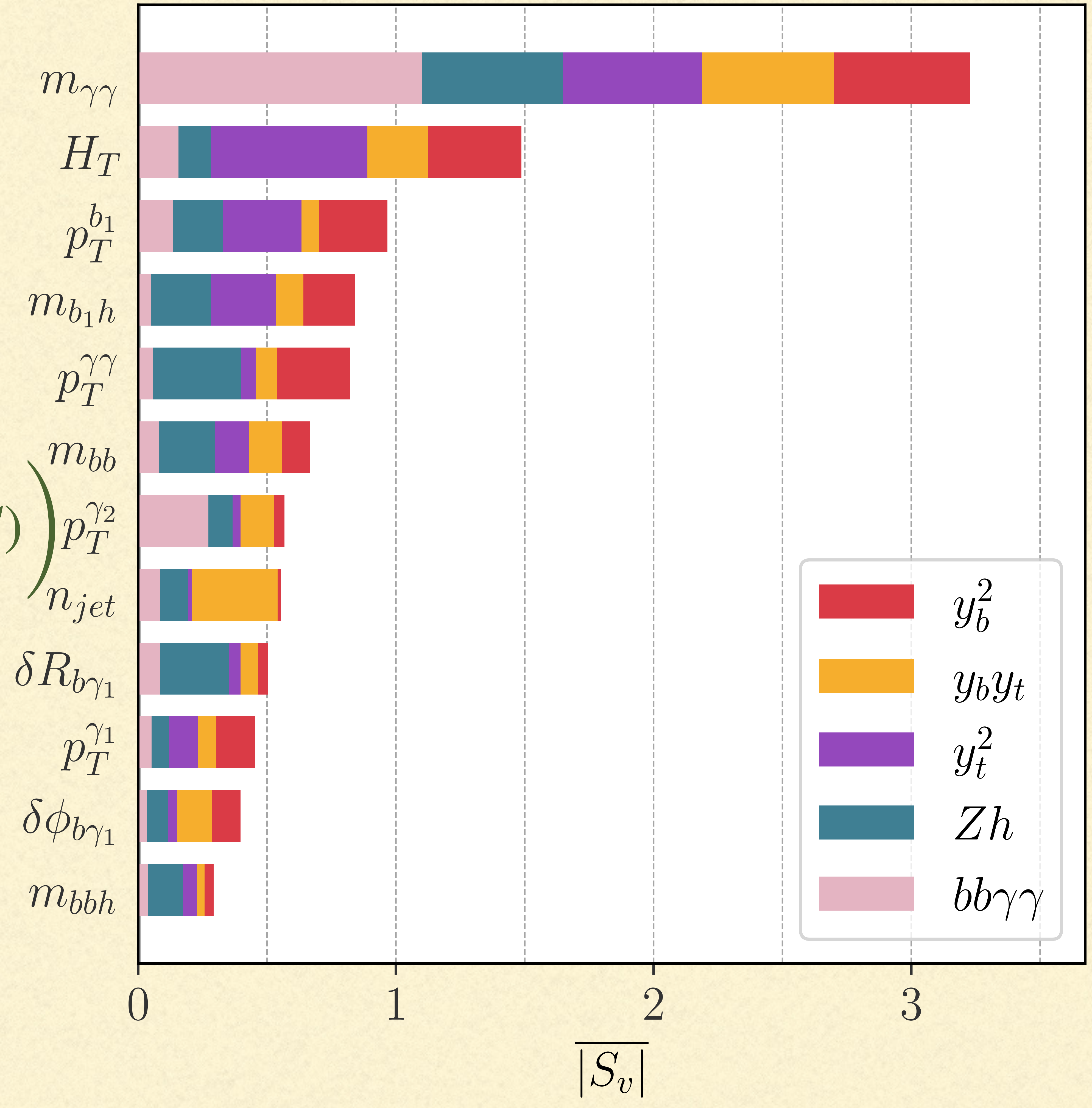


Machine Interpretation (full): Using shapley value:

- Calculated by SHAP on BDT model
- Shapley value: an importance "measures" of given observable, through marginalising its contribution over the observable set:

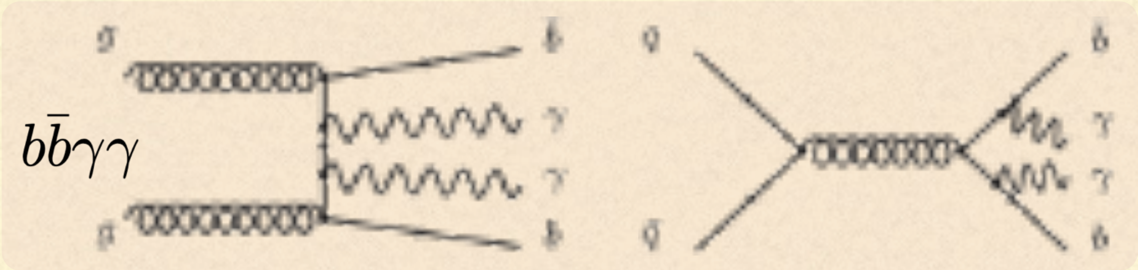
$$\phi_j(val) = \sum_{S \subseteq \{x_1,\ldots,x_p\} \setminus \{x_j\}} \frac{|S|!(p-|S|-1)!}{p!} \left(val\left(S \cup \{x_j\}\right) - val(S) \right)$$

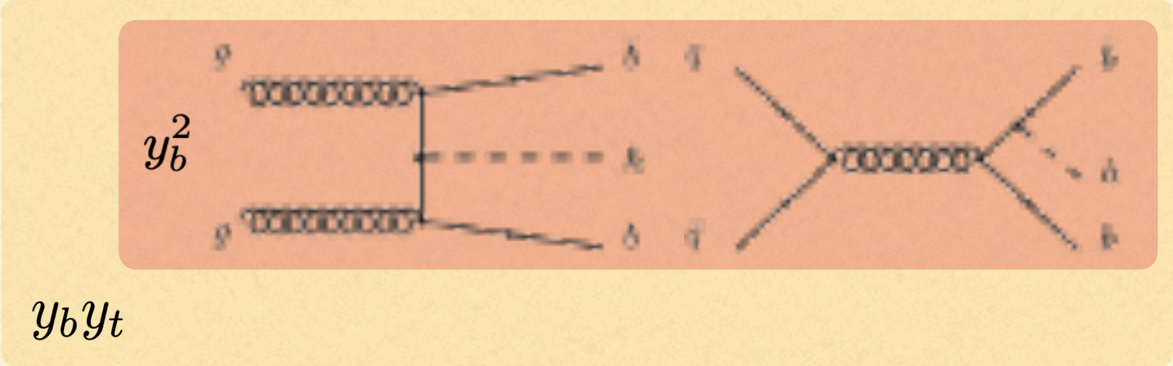
- Feature importance: the averaged absolute value of shapley value for a given observable: $I_j = \sum_{i=1}^{n} |\phi_j^{(i)}|$
- Reduction of observable d.o.f ; local attribution of importance with additivity

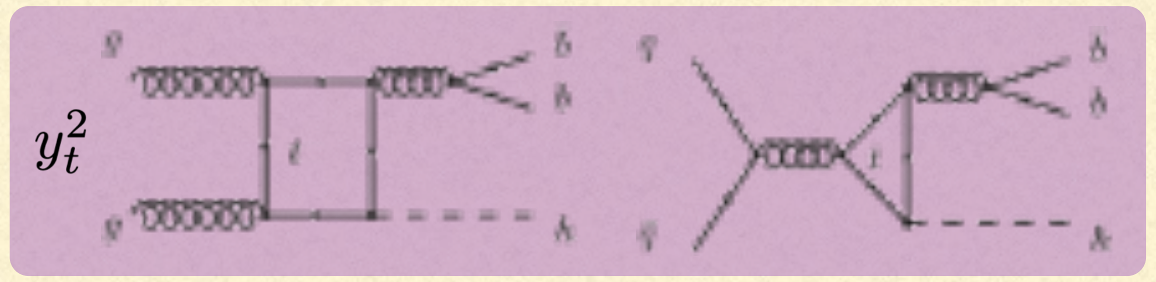


An interpretable framework:

Interpretable variables







QCD-QED Backgrounds

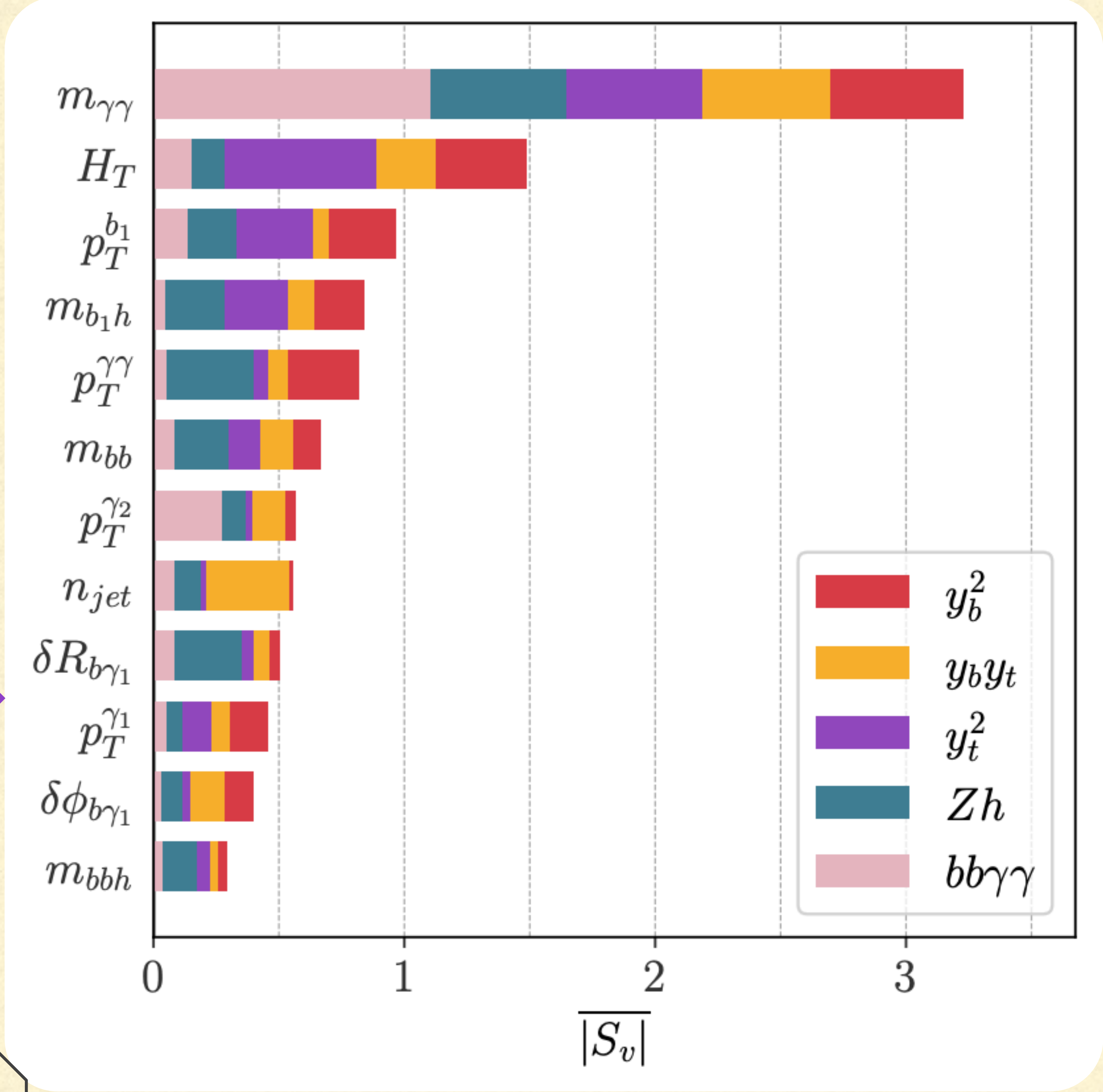
Signal

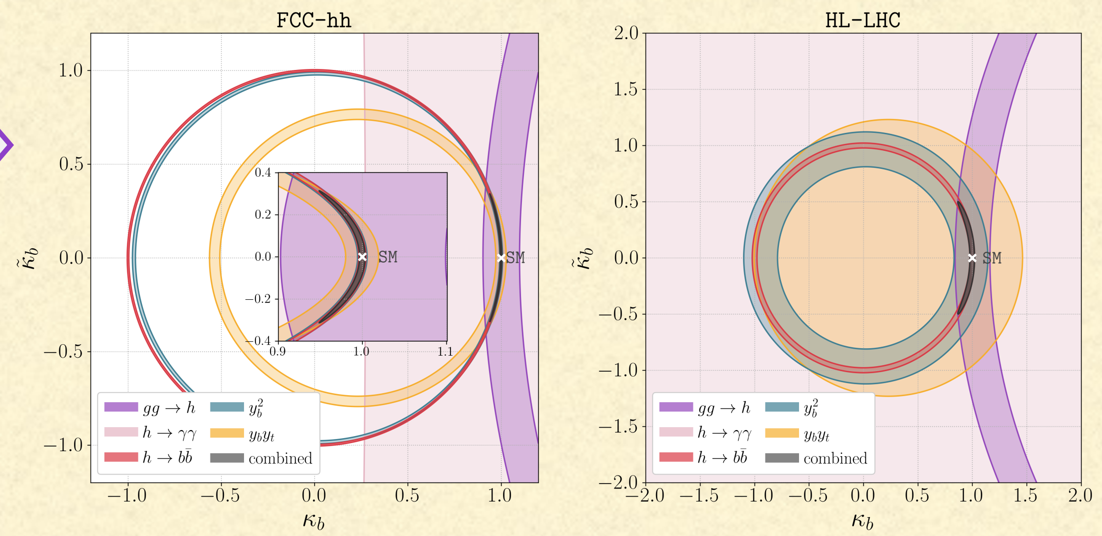
$b\bar{b}h$ backgrounds

Shapley values + Physics Insights

Boosted Decision Trees + Signal Classification

Interpretable model





Some Conclusions:

- Multivariate analysis are the key to distinguishing signals that cut-based analyses using only 1D or low-dimension info miserably fail at.
- Multivariate methods can be better understood and supplemented with interpretable machine learning tools such as the Shapley values.
- Associated production of $b\bar{b}h$ stands to gain at FCC as its production rate grows faster than backgrounds with rising energies. => sensitivity on a complex phase of y_b can be comparable to that from $h \to b\bar{b}$.

References

Measurement of $H \to b\bar{b}$

- J. M. Butterworth, A. R. Davison, M. Rubin and G. P. Salam, *Jet substructure as a new Higgs search channel at the LHC*, Phys. Rev. Lett. 100 (2008) 242001, [arXiv:0802.2470].
- ATLAS collaboration, M. Aaboud et al., *Observation of $H \to b\bar{b}$ decays and VH production with the ATLAS detector*, Phys. Lett. B 786 (2018) 59–86, [arXiv:1808.08238].
- CMS collaboration, A. M. Sirunyan et al., *Observation of Higgs boson decay to bottom quarks*, Phys. Rev. Lett. 121 (2018) 121801, [arXiv:1808.08242].

Recent theory papers on $b\bar{b}h$

- N. Deutschmann, F. Maltoni, M. Wiesemann and M. Zaro, *Top-Yukawa contributions to $b\bar{b}h$ production at the LHC*, JHEP 07 (2019) 054, [arXiv:1808.01660].
- D. Pagani, H.-S. Shao and M. Zaro, *RIP $Hb\bar{b}$: How other Higgs production modes conspire to kill a rare signal at the LHC*, arXiv:2005.10277.

Papers on Higgs couplings fits

- M. Cepeda et al., *Report from Working Group 2: Higgs Physics at the HL-LHC and HE-LHC*, vol. 7, pp. 221–584. 12, 2019. arXiv:1902.00134.
- J. de Blas et al., *Higgs Boson Studies at Future Particle Colliders*, JHEP 01 (2020) 139, [arXiv:1905.03764].

Shapley values and interpretable machine learning

- L. S. Shapley, *Notes on the n-Person Game-II: The Value of an n-Person Game*, Rand Corporation (1951).
- L. S. Shapley, *A Value for n-person Games*. Contributions to the Theory of Games 2.28 (1953): 307-317.
- C. Molnar, *Interpretable Machine Learning*. Lulu, 2020. [Link]
- S. M. Lundberg and S.-I. Lee, *A unified approach to interpreting model predictions*, in Advances in Neural Information Processing Systems (I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan et al., eds.), vol. 30, pp. 4765–4774, Curran Associates, Inc., 2017. arXiv:1705.07874.
- S. M. Lundberg, G. G. Erion and S.-I. Lee, Consistent Individualized Feature Attribution for Tree Ensembles, arXiv e-prints (Feb. 2018) , [arXiv:1802.03888].
- S. M. Lundberg, G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair et al., *From local explanations to global understanding with explainable AI for trees*, Nature Machine Intelligence 2 (2020) 56–67.

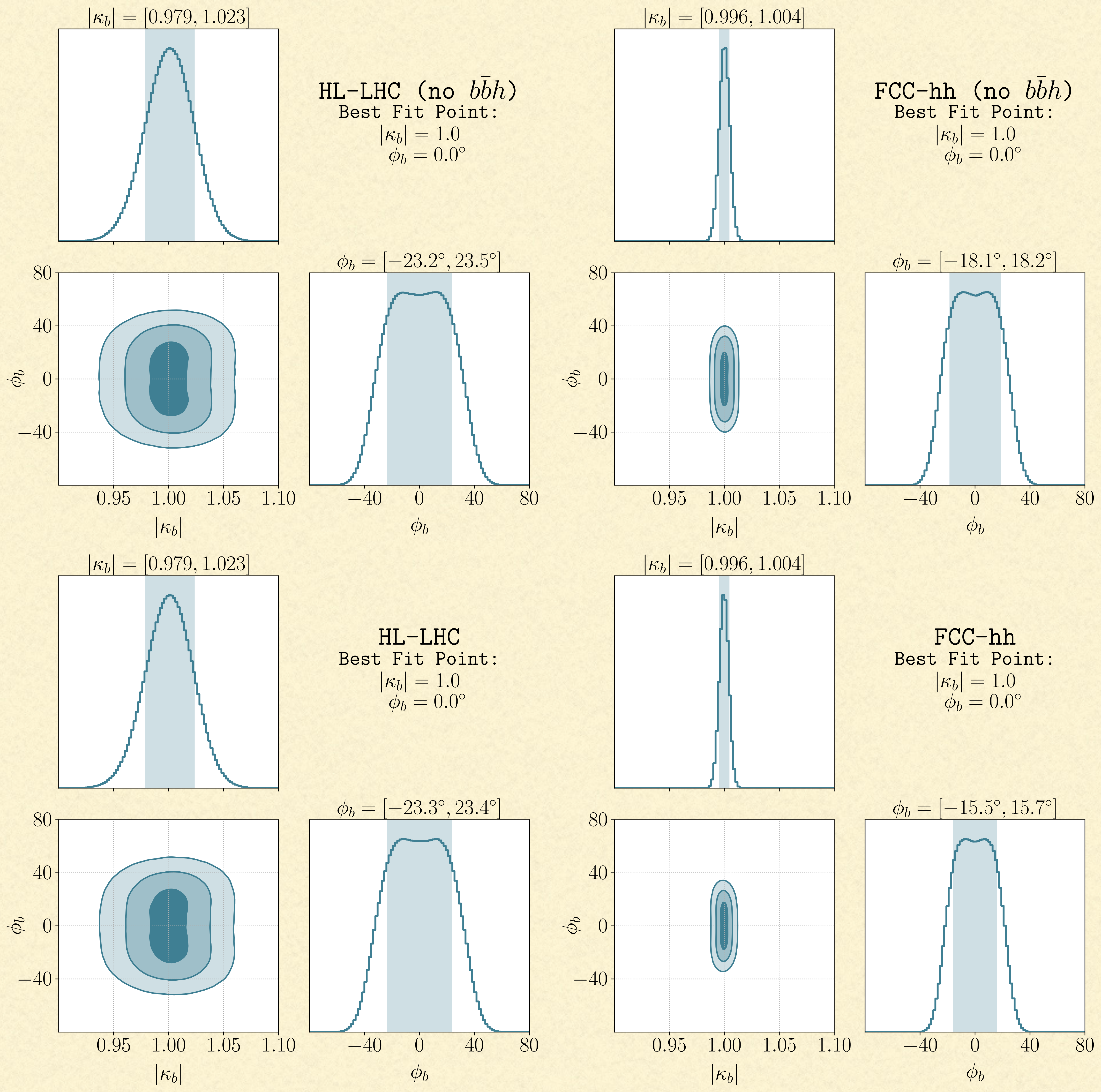
Backup

Additional backgrounds and discussion :

- VBF: light-jet veto kills the VBF while careful simulation is further needed.
- di-Higgs: both mbb and mγγ clustered around the Higgs-mass peak, distinct final state shape to be separate
- $gg \to Zh$: small at HL-LHC, but grows rapidly with s, and comparable but subdominant to the yb-sensitive channels at FCC-hh. Can be further distinguished as the case of $q\bar{q} \to Zh$.
- Fakes: ccxaa, jjxaa, caa, jjja, etc.: subdominant yet comparable to bbxaa. Needs attention and study in future for better control

Systematics estimates:

systematics	HL-LHC (6 ab^{-1})		FCC-hh (30 ab^{-1})	
	y_b^2	$y_b y_t$	y_b^2	$y_b y_t$
0%	3.33	0.47	63.7	10.4
0.5%	3.26	0.46	32.2	3.44
1%	3.06	0.42	17.9	1.80
5%	1.41	0.18	3.72	0.36



HL-LHC: $\phi_b = [-23.2°, 23.5°]$ => $\tilde{\kappa}_b \lesssim$ 0.4

Negligible improvement from bbxh study

FCC-hh: $\phi_b = [-15.5°, 15.7°]$ => $\tilde{\kappa}_b \lesssim$ 0.3

~15% improvement from bbxh study

Comparison to:

Hadronic EDM (free of electron EDM assumption):

nEDM: $\sum A\kappa_q\tilde{\kappa}_{\mathbf{q}} + B\tilde{\kappa}_{\mathbf{q}}\kappa_{\mathbf{q}}$ => $\tilde{\kappa}_b \lesssim 5.$

Electron EDM:

eEDM: $\sum A\kappa_e\tilde{\kappa}_{\mathbf{q}} + B\tilde{\kappa}_e\kappa_{\mathbf{q}}$ => $\tilde{\kappa}_b \lesssim 0.5$

One slide on Shapley value:

The Shapley value is defined via a value function val of players in S.

The Shapley value of a feature value is its contribution to the payout, weighted and summed over all possible feature value combinations:

$$\phi_j(val) = \sum_{S \subseteq \{x_1,\ldots,x_p\} \setminus \{x_j\}} \frac{|S|!\,(p-|S|-1)!}{p!} \left(val\left(S \cup \{x_j\}\right) - val(S)\right)$$

where S is a subset of the features used in the model, x is the vector of feature values of the instance to be explained and p the number of features. $val_x(S)$ is the prediction for feature values in set S that are marginalized over features that are not included in set S:

$$val_x(S) = \int \hat{f}\,(x_1,\ldots,x_p)d\mathbb{P}_{x \notin S} - E_X(\hat{f}\,(X))$$

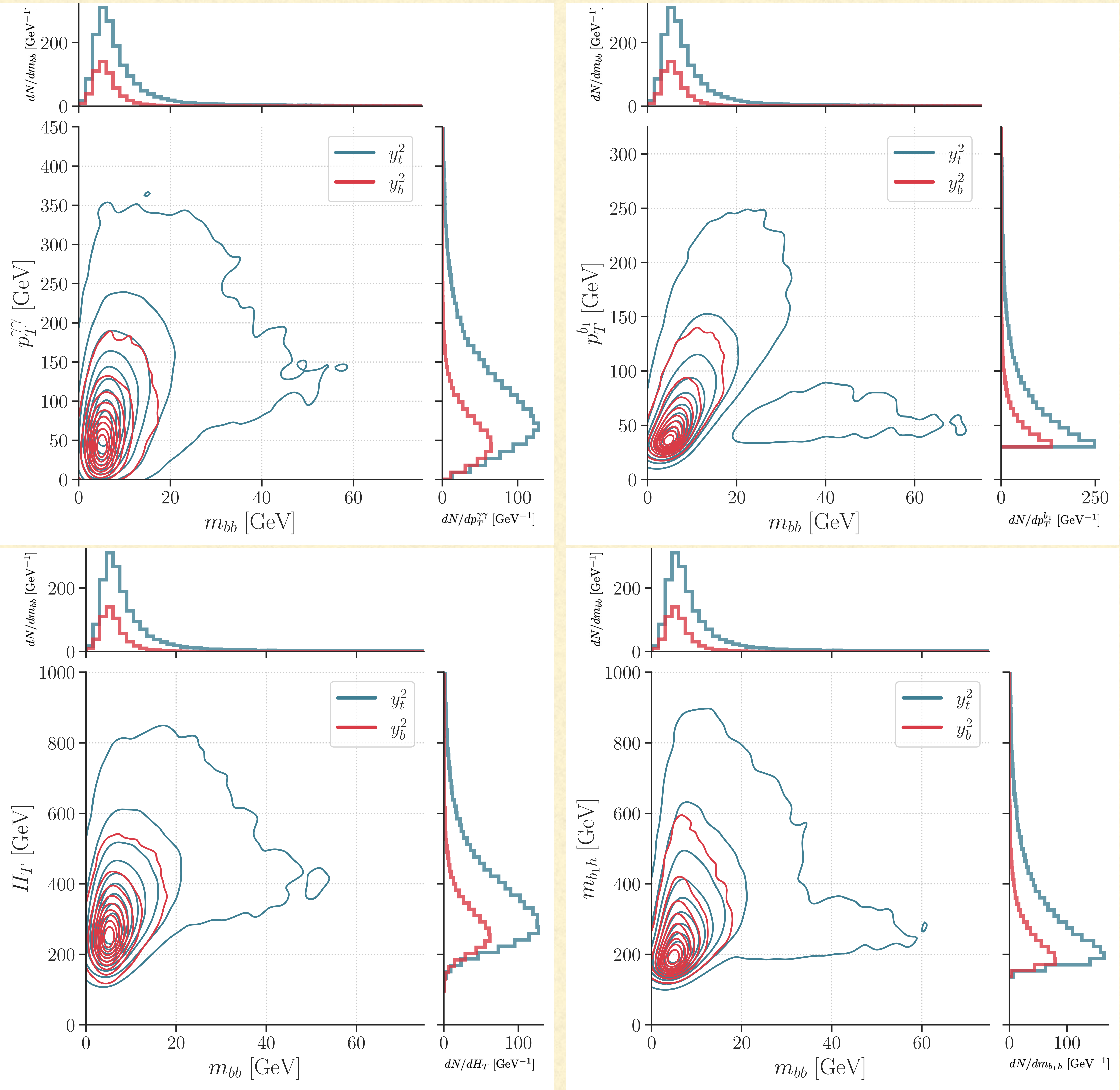
You actually perform multiple integrations for each feature that is not contained S. A concrete example: The machine learning model works with 4 features x1, x2, x3 and x4 and we evaluate the prediction for the coalition S consisting of feature values x1 and x3:

$$val_x(S) = val_x(\{x_1, x_3\}) = \int_{\mathbb{R}} \int_{\mathbb{R}} \hat{f}\,(x_1, X_2, x_3, X_4)d\mathbb{P}_{X_2 X_4} - E_X(\hat{f}\,(X))$$

https://christophm.github.io/interpretable-ml-book/shapley.html

yt2-yb2

HL-LHC



HL-LHC