RESURRECTING y_b FROM $b\bar{b}h$ STUDY WITH KINEMATIC SHAPES

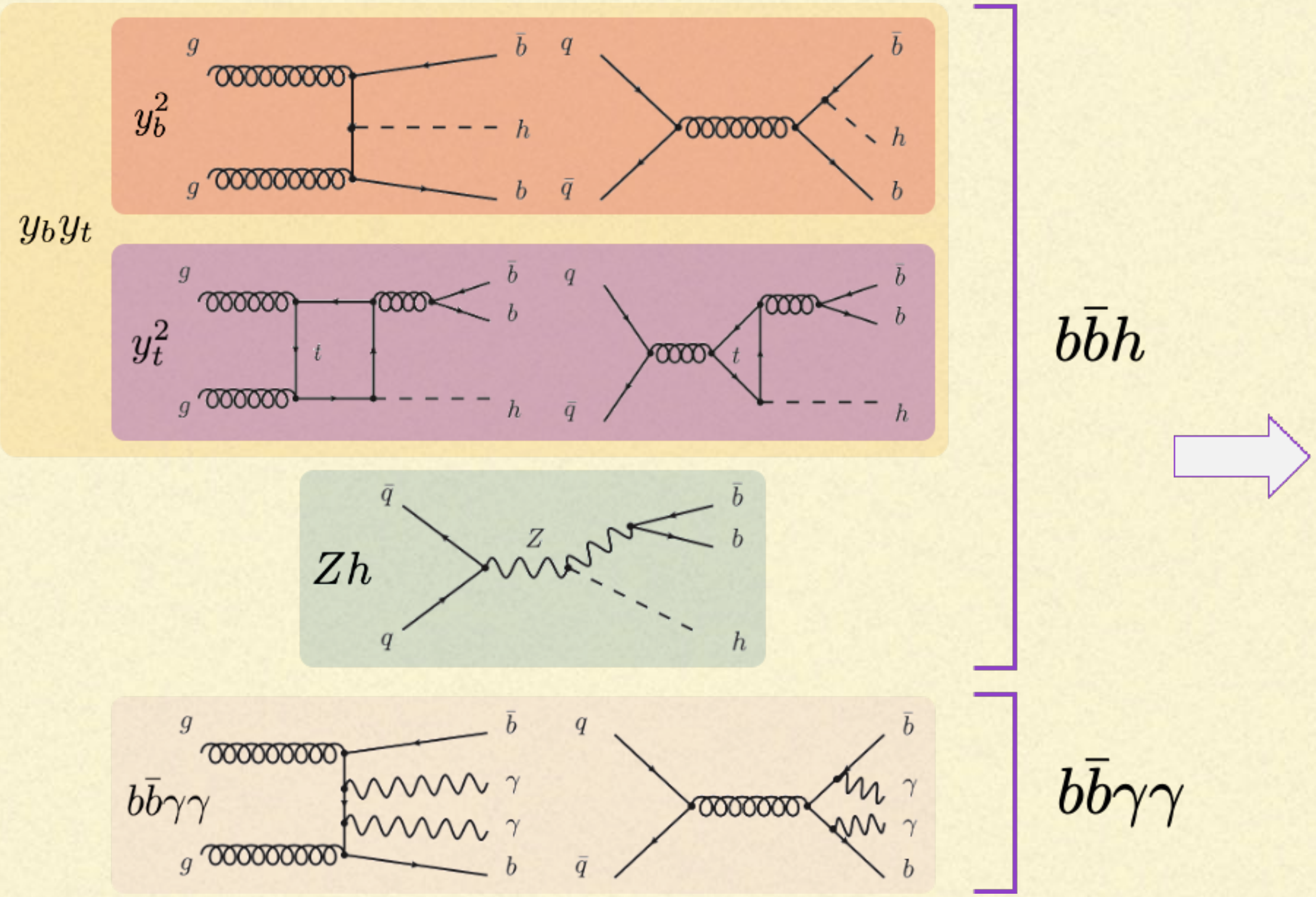
Zhuoni Qian @ Pheno, May 2021
On behalf of C. Grojean, A. Paul
arXiv: 2011.13945

$b\bar{b}h$ signal at the (HL-)LHC, y_b sensitivity

Motivation: Bottom Yukawa measure is a recent achievement:

o Phase of the Yukawa not well measured

o Interplay between Yukawa phases in EDM and collider



C. Grojean, A. Paul, ZQ(2020)

$b\bar{b}h$ signal at the (HL-)LHC, y_b sensitivity

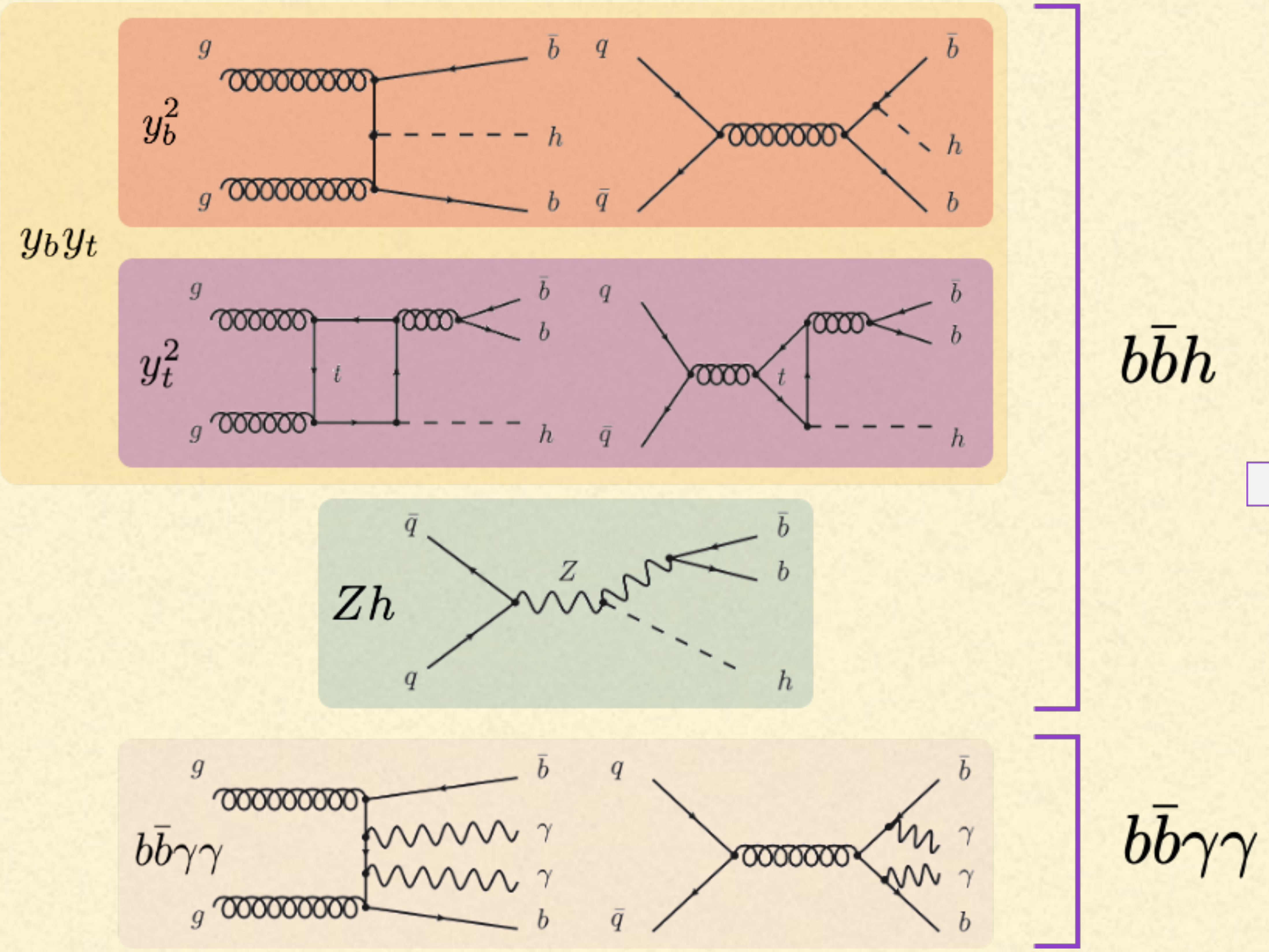
Motivation: Bottom Yukawa measure is a recent achievement:

o Phase of the Yukawa not well measured

o Interplay between Yukawa phases in EDM and collider

Previous consensus: hopeless to separate from all the $b\bar{b}h$ contributions at HL-LHC:

→ no y_b sensitivity



C. Grojean, A. Paul, ZQ(2020)

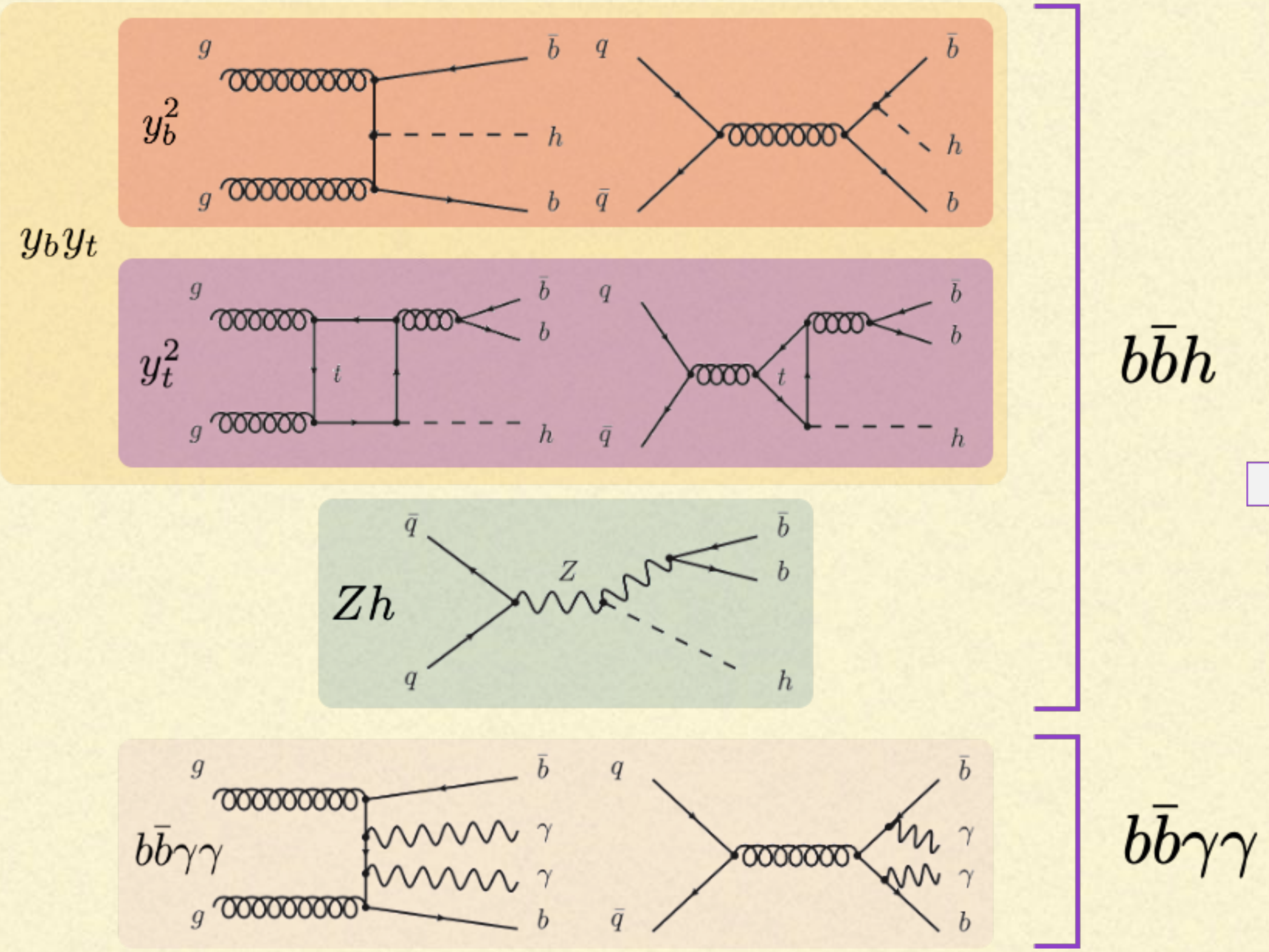
$b\bar{b}h$ signal at the (HL-)LHC, y_b sensitivity

Motivation: Bottom Yukawa measure is a recent achievement:

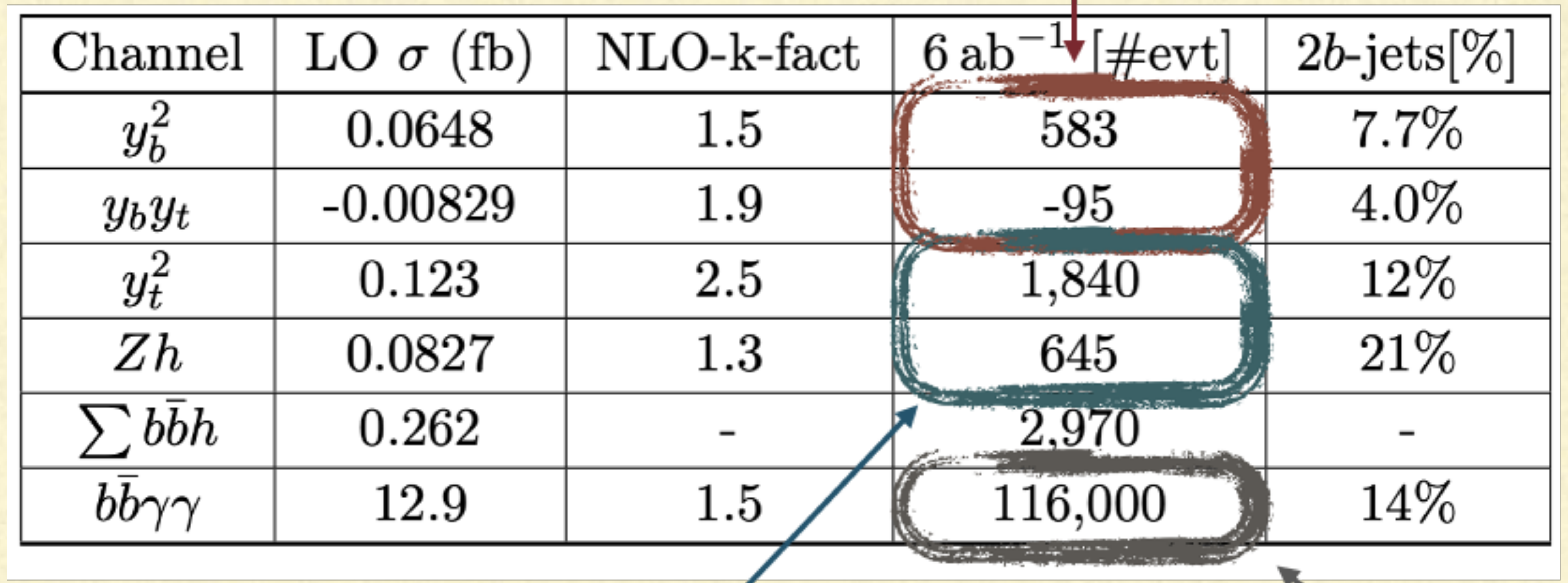
- Phase of the Yukawa not well measured
- Interplay between Yukawa phases in EDM and collider

Previous consensus: hopeless to separate from all the $b\bar{b}h$ contributions at HL-LHC:

→ no y_b sensitivity



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

$b\bar{b}\gamma\gamma$ **background**

C. Grojean, A. Paul, ZQ(2020)

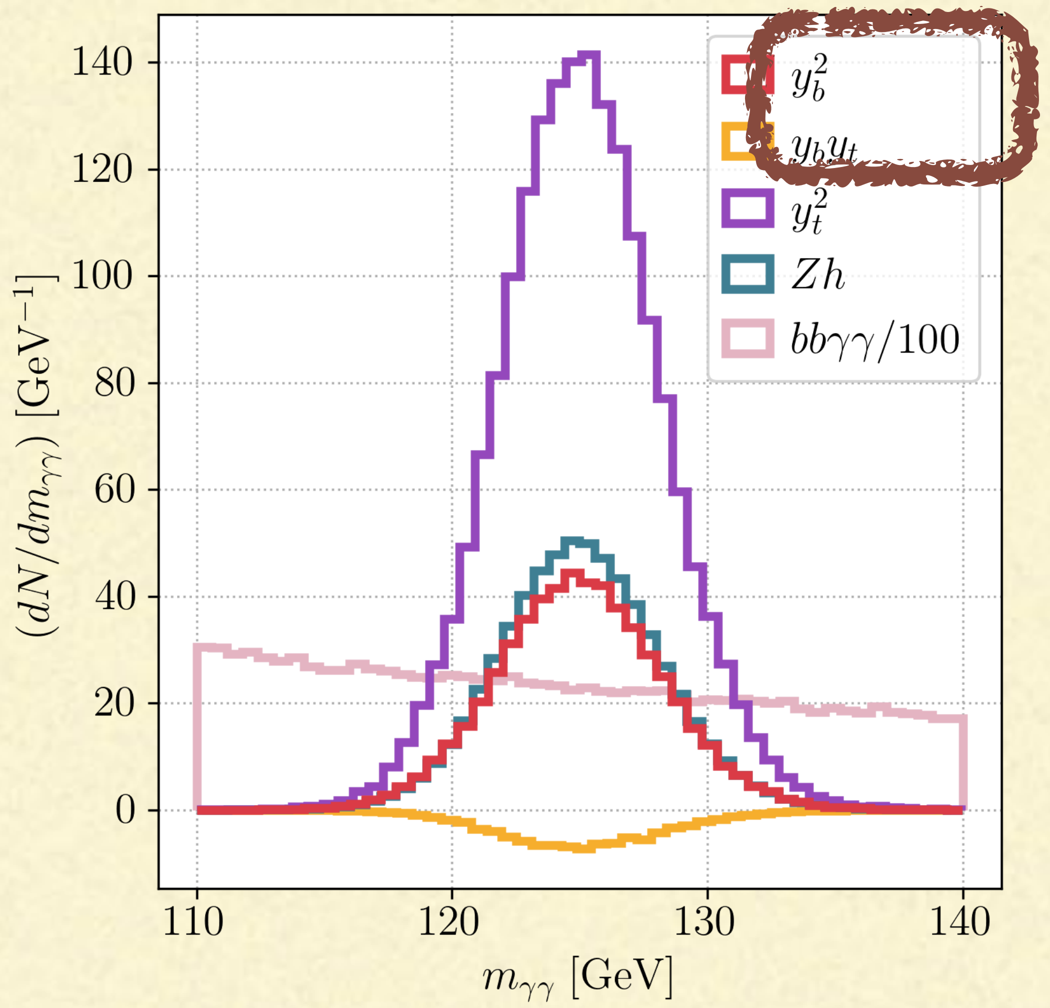
1D Differential Distribution

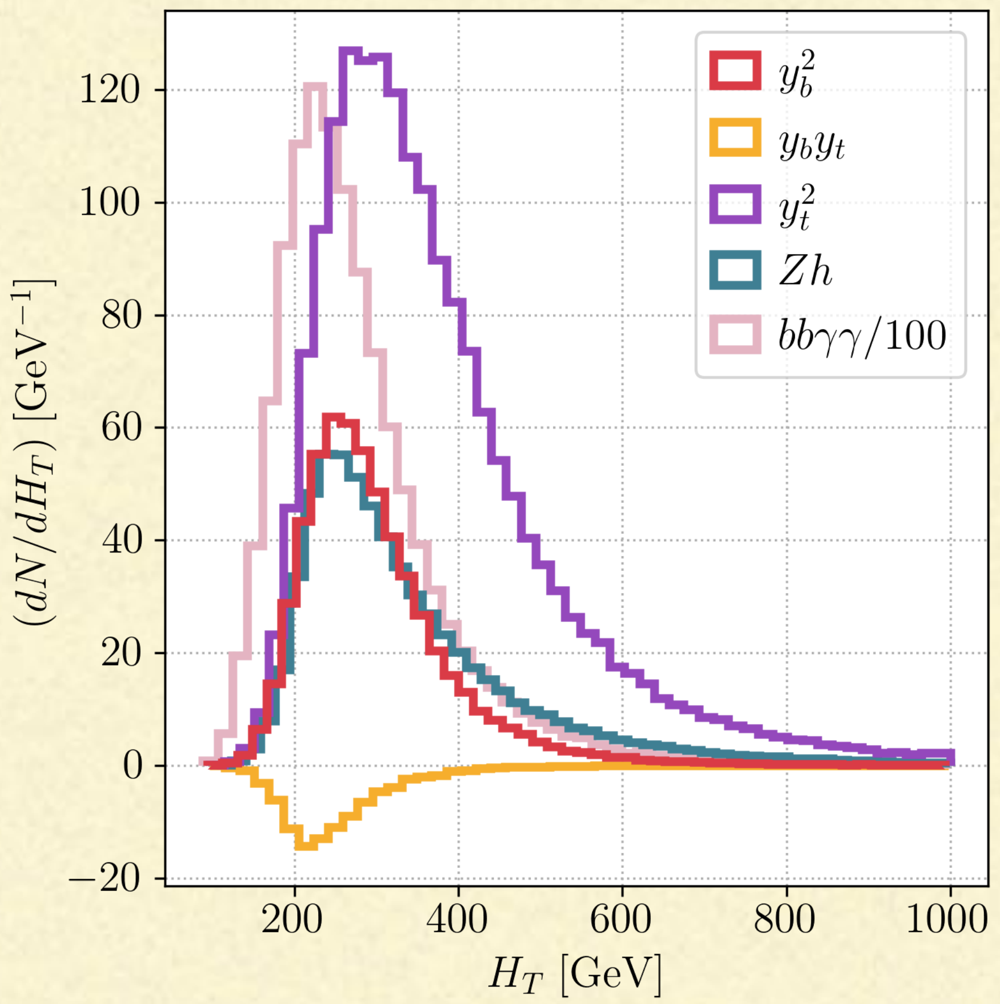
Observable and distributions:

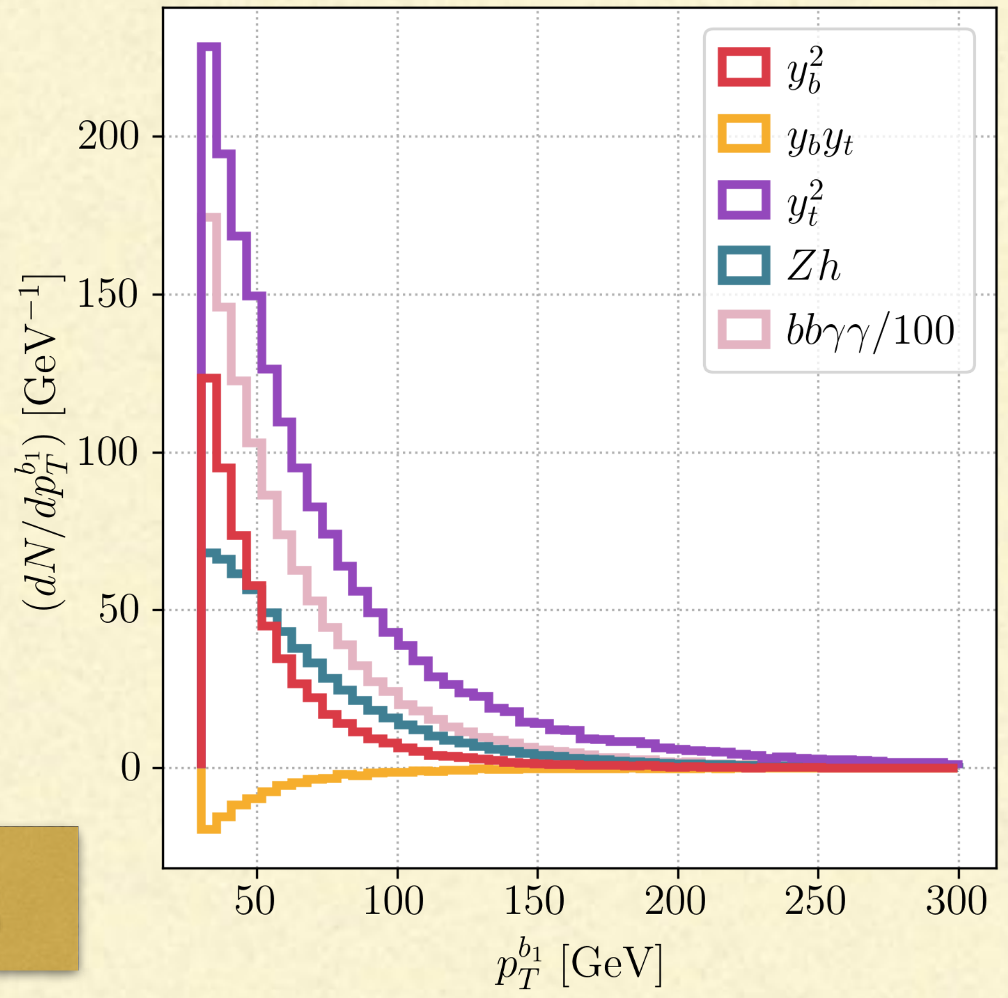
- $p_T^{b_1}, p_T^{b_2}, p_T^{\gamma_1}, p_T^{\gamma\gamma},$
- $\eta_{b_{j1}}, \eta_{b_{j2}}, \eta_{\gamma_1}, \eta_{\gamma\gamma},$
- $n_{bjet}, n_{jet}, \Delta R_{\text{min}}^{b\gamma}, \Delta\phi_{\text{min}}^{bb},$
- $m_{\gamma\gamma}, m_{bb}, m_{b_1 h}, m_{b\bar{b}h}, H_T.$

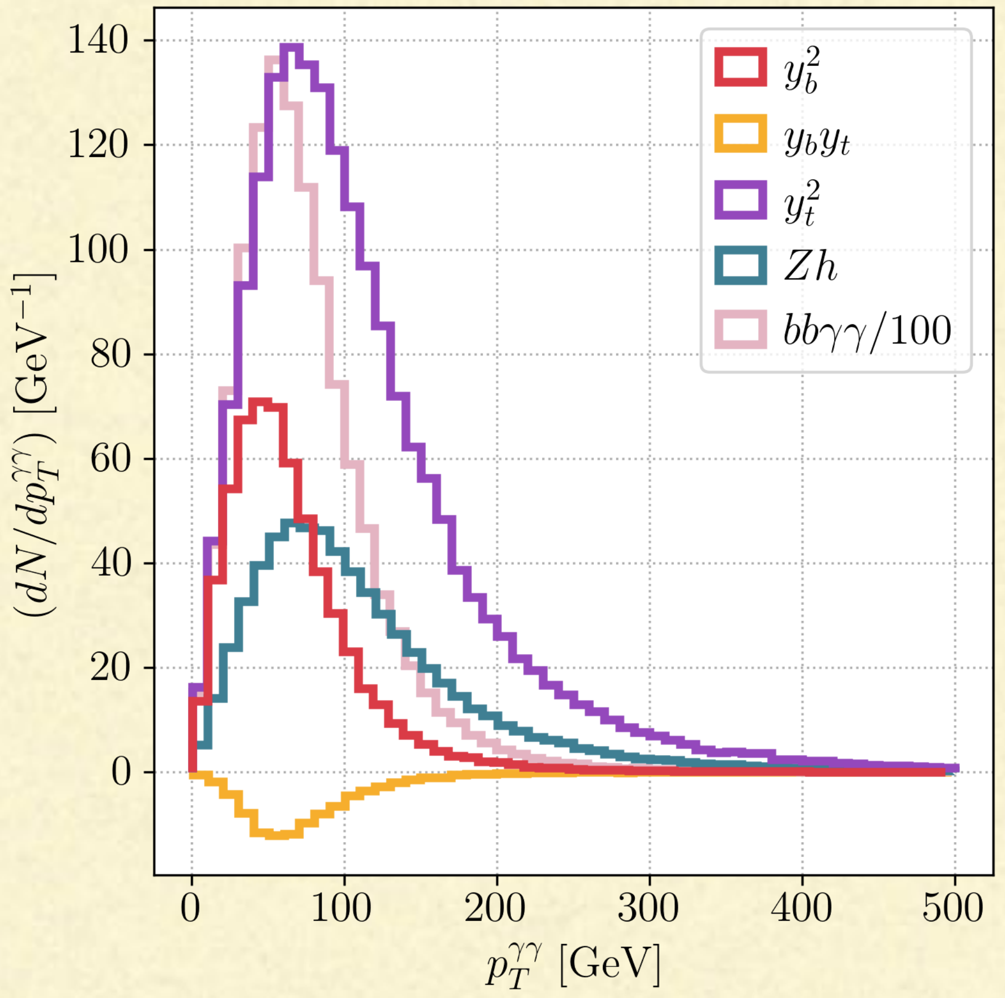
Begin with an over-complete set of median/high level observables

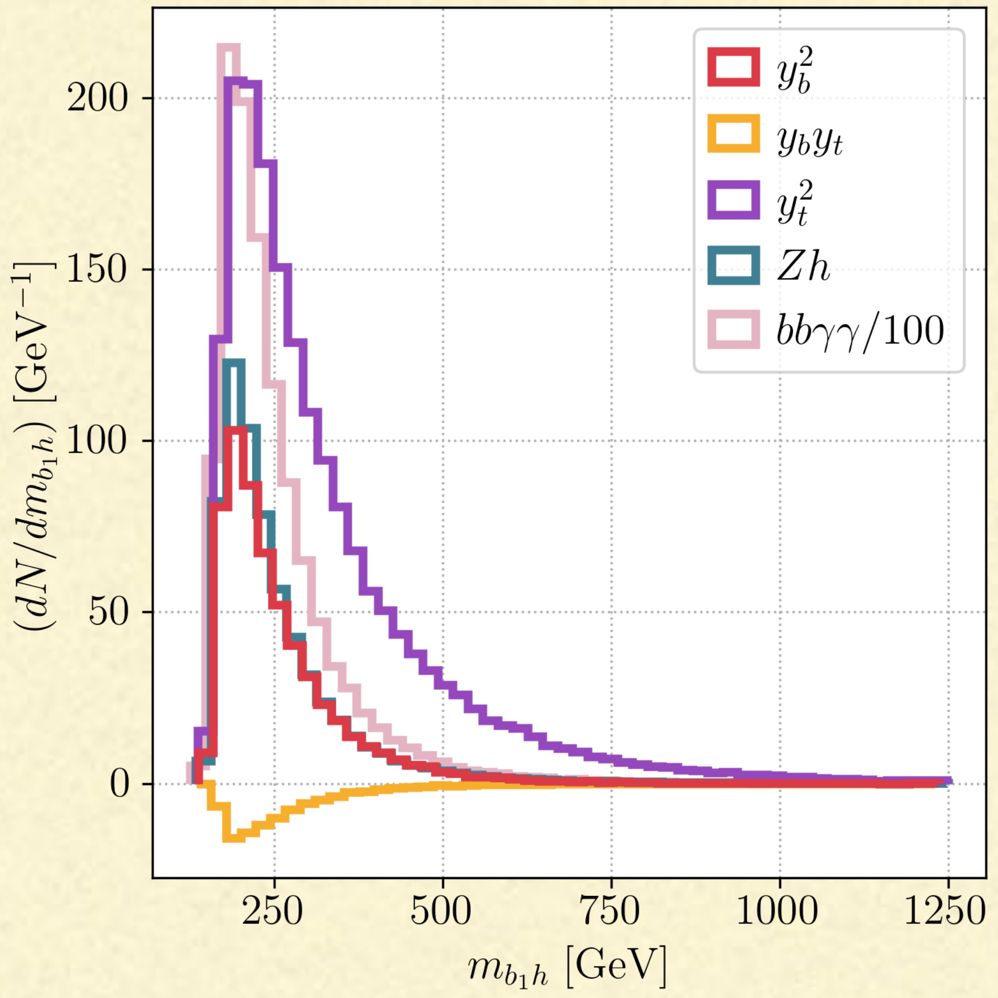
Challenging to distinguish channels





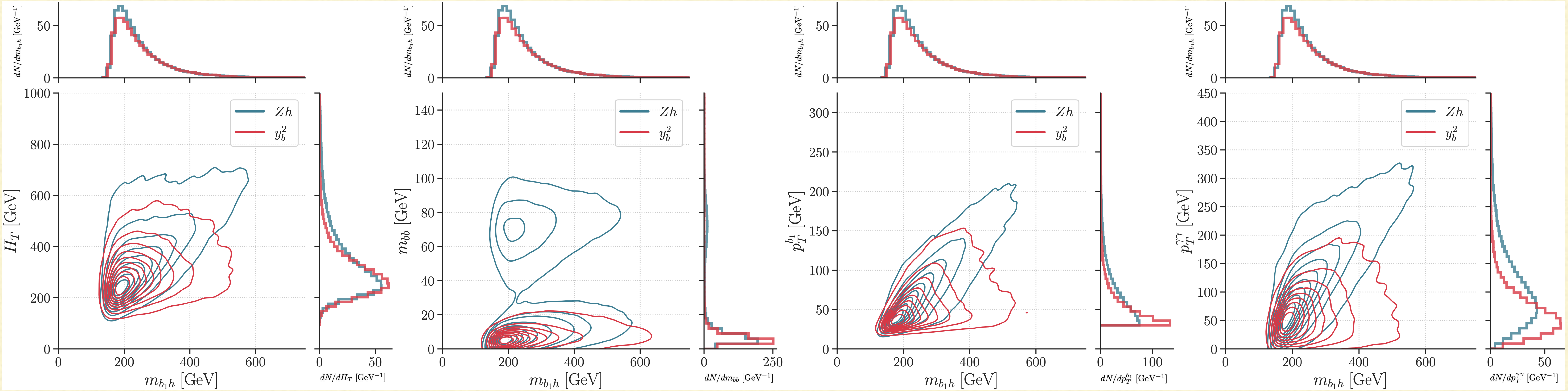






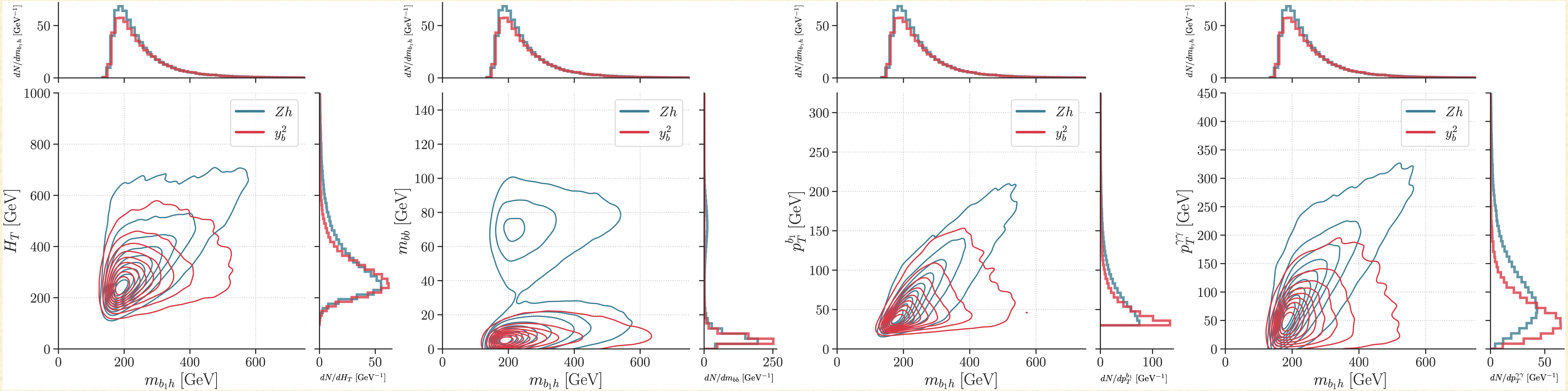
C. Grojean, A. Paul, ZQ(2020)

2D Differential Distribution (y_b^2 and Zh)



C. Grojean, A. Paul, ZQ(2020)

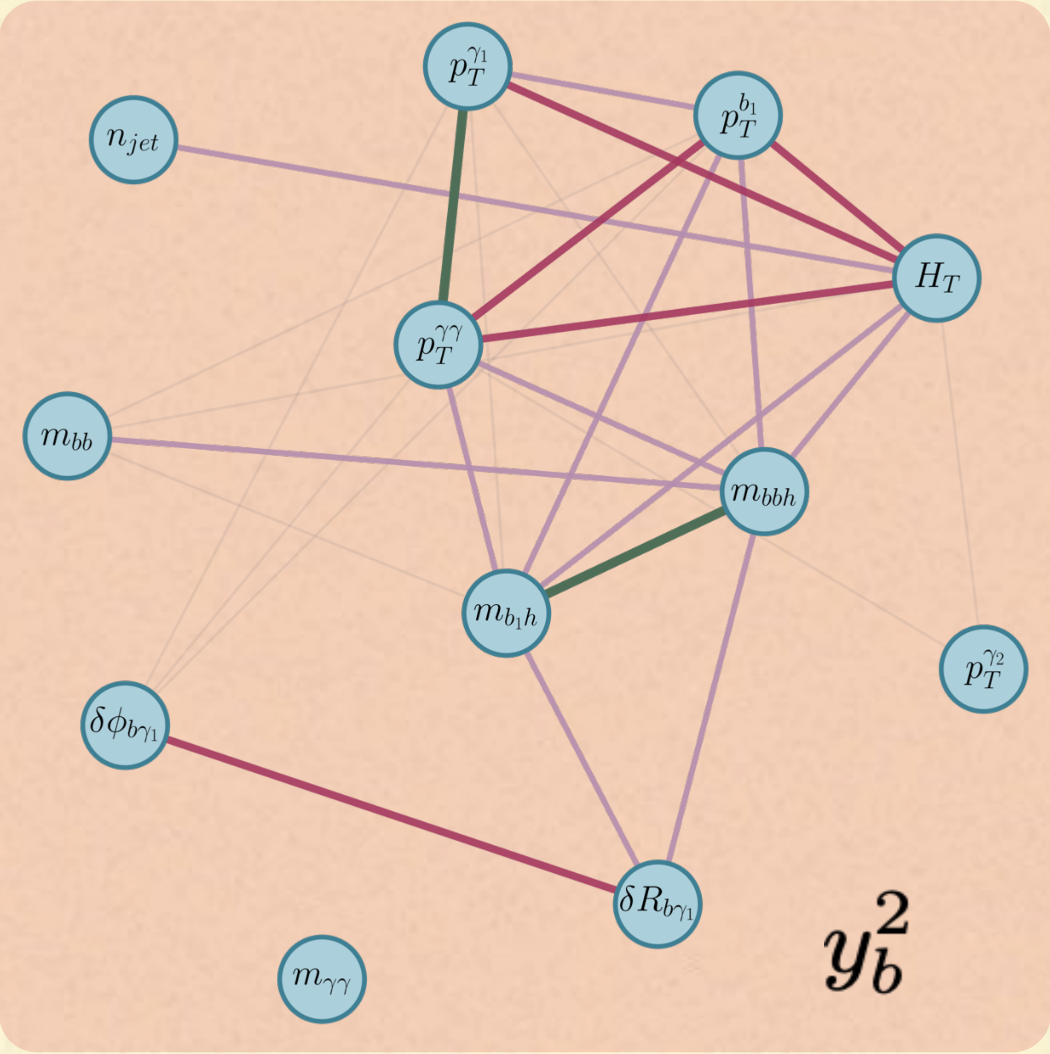
2D Differential Distribution (y_b^2 and Zh)

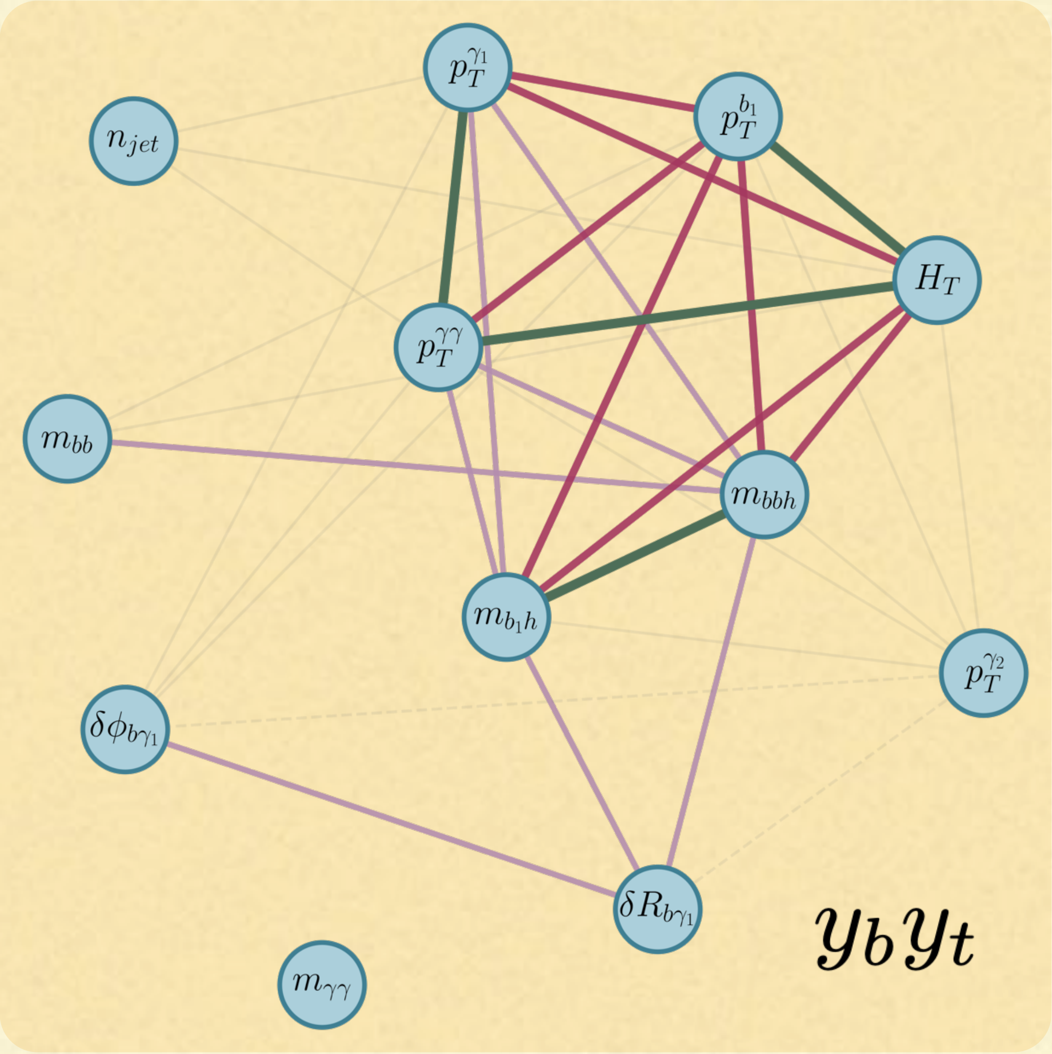


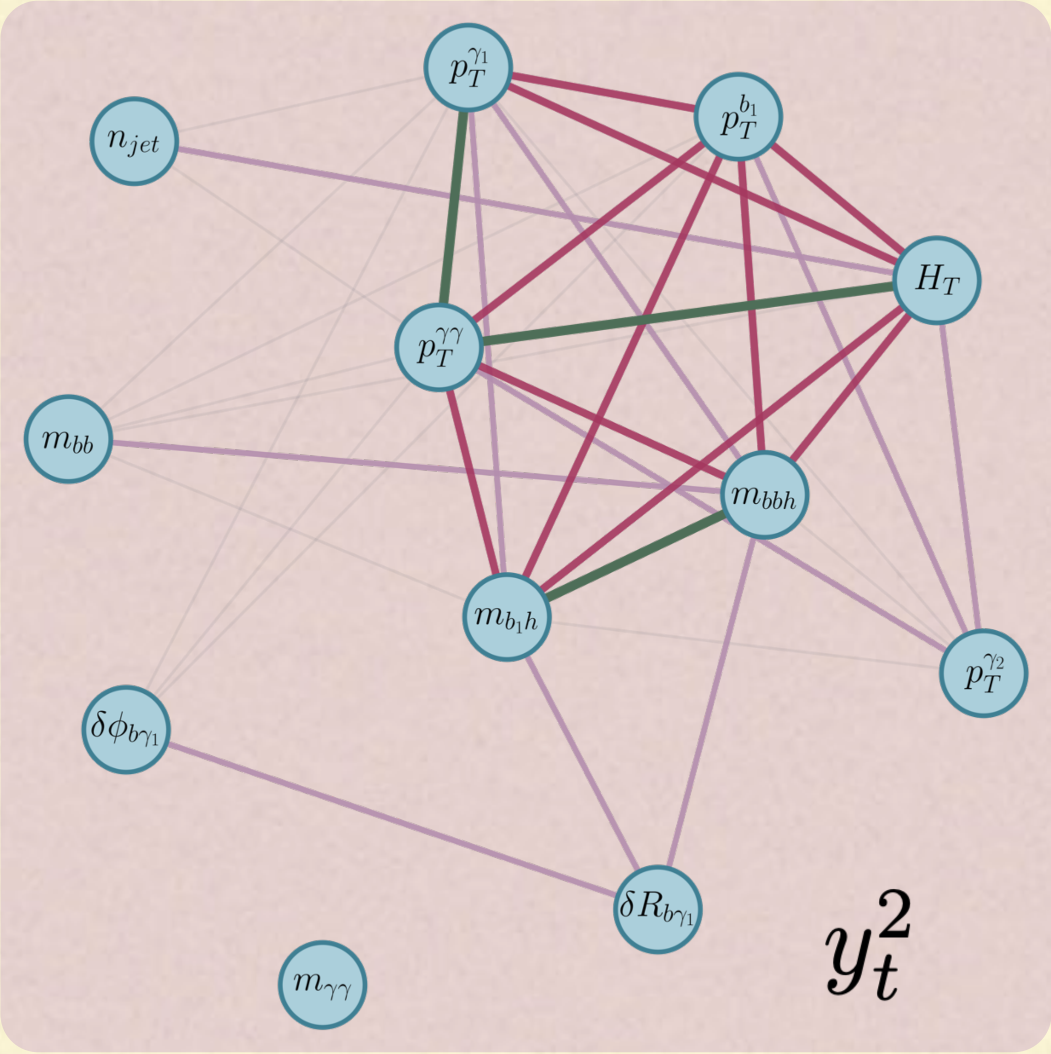
- High dimensional features reveal further difference
- Designed cut, smarter/optimal observable, matrix element method

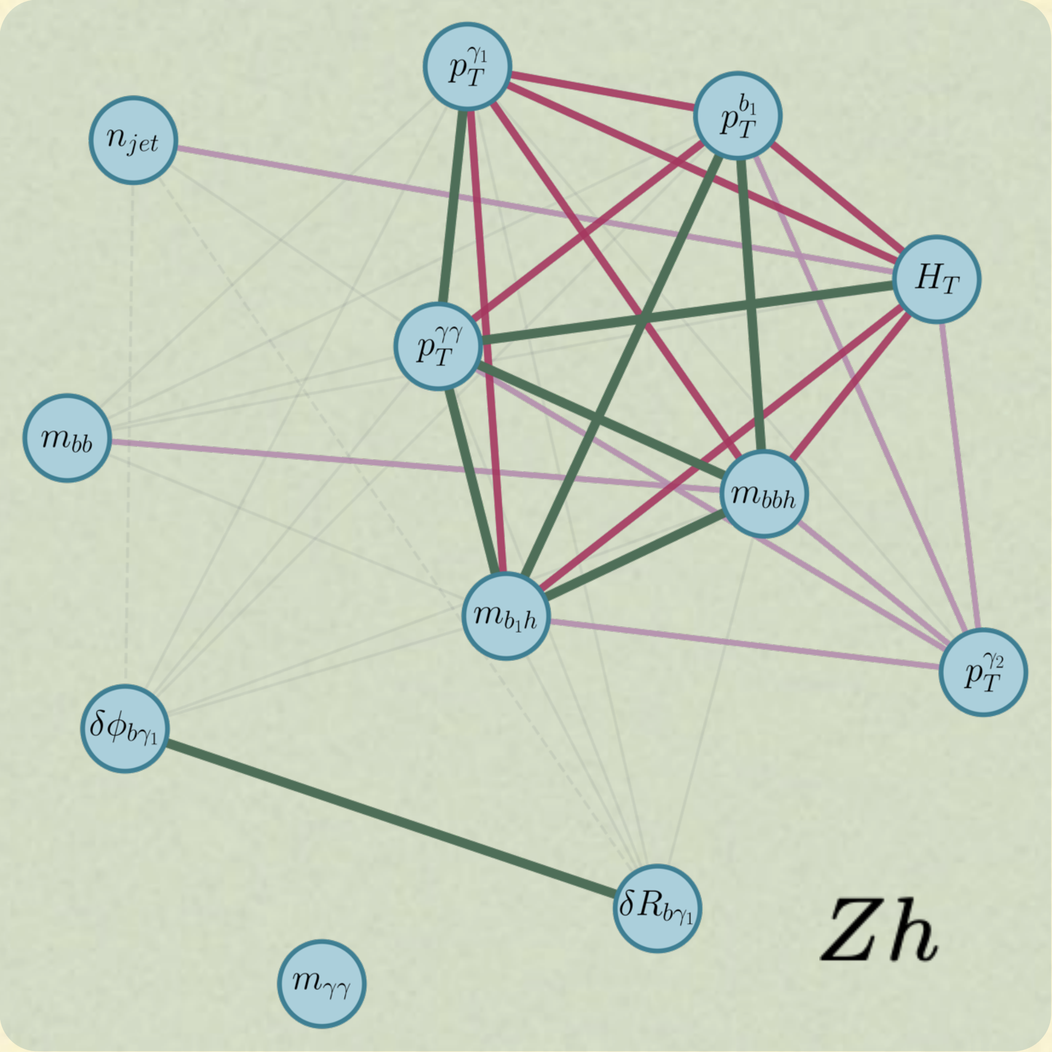
C. Grojean, A. Paul, ZQ(2020)

Diving into higher dimension



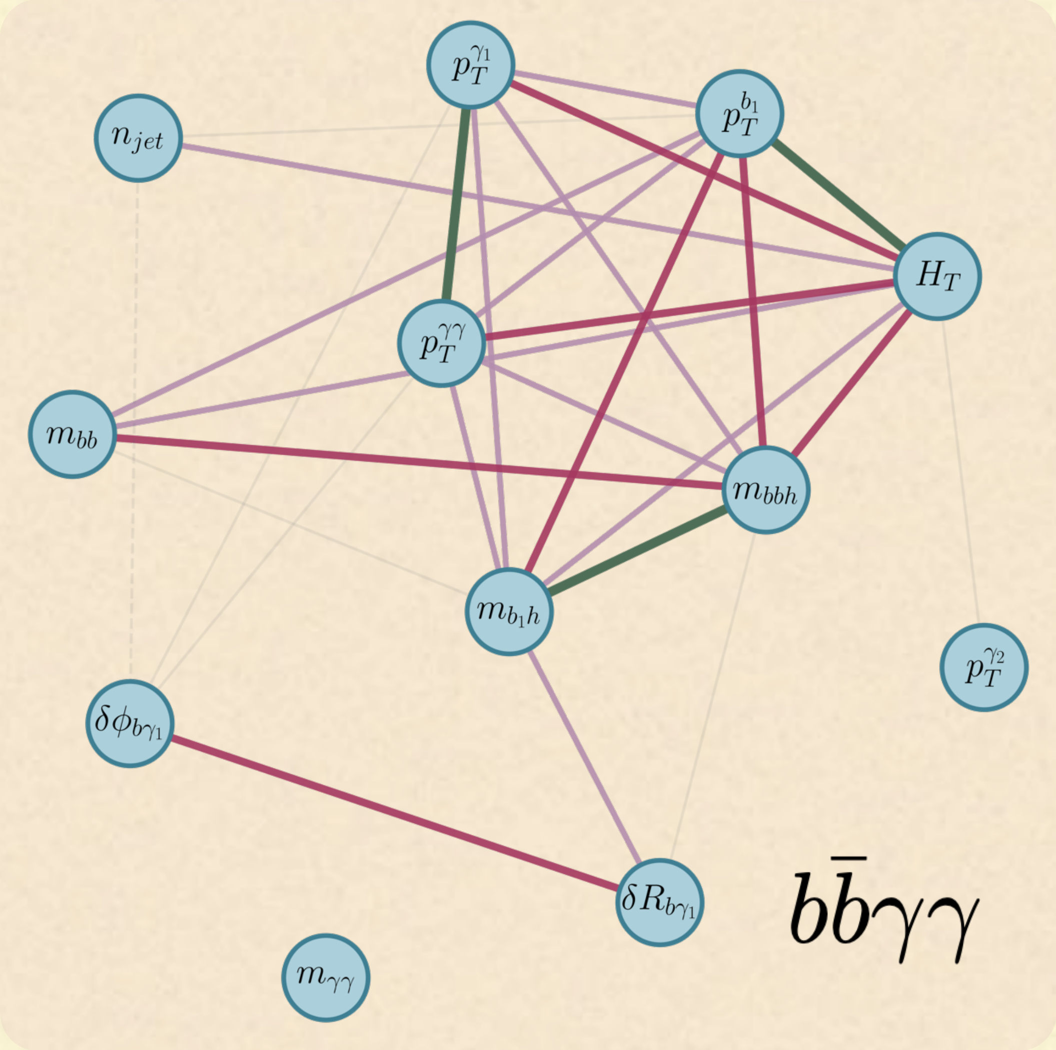






⬜ $0.2 \le |\rho| < 0.4$ 🟪 $0.4 \le |\rho| < 0.6$ 🟥 $0.6 \le |\rho| < 0.8$ 🟩 $0.8 \le |\rho| < 1.0$

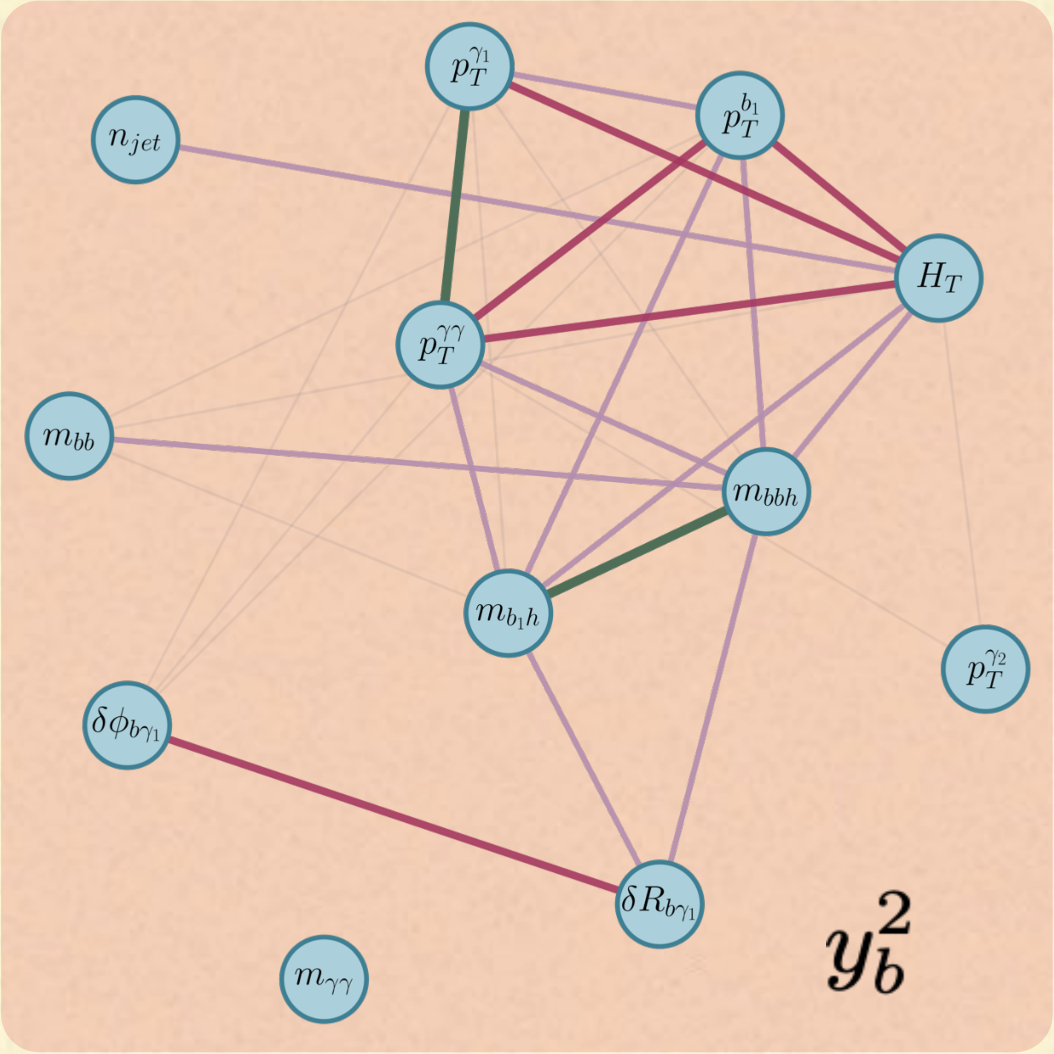
--- $-\rho$ — $+\rho$

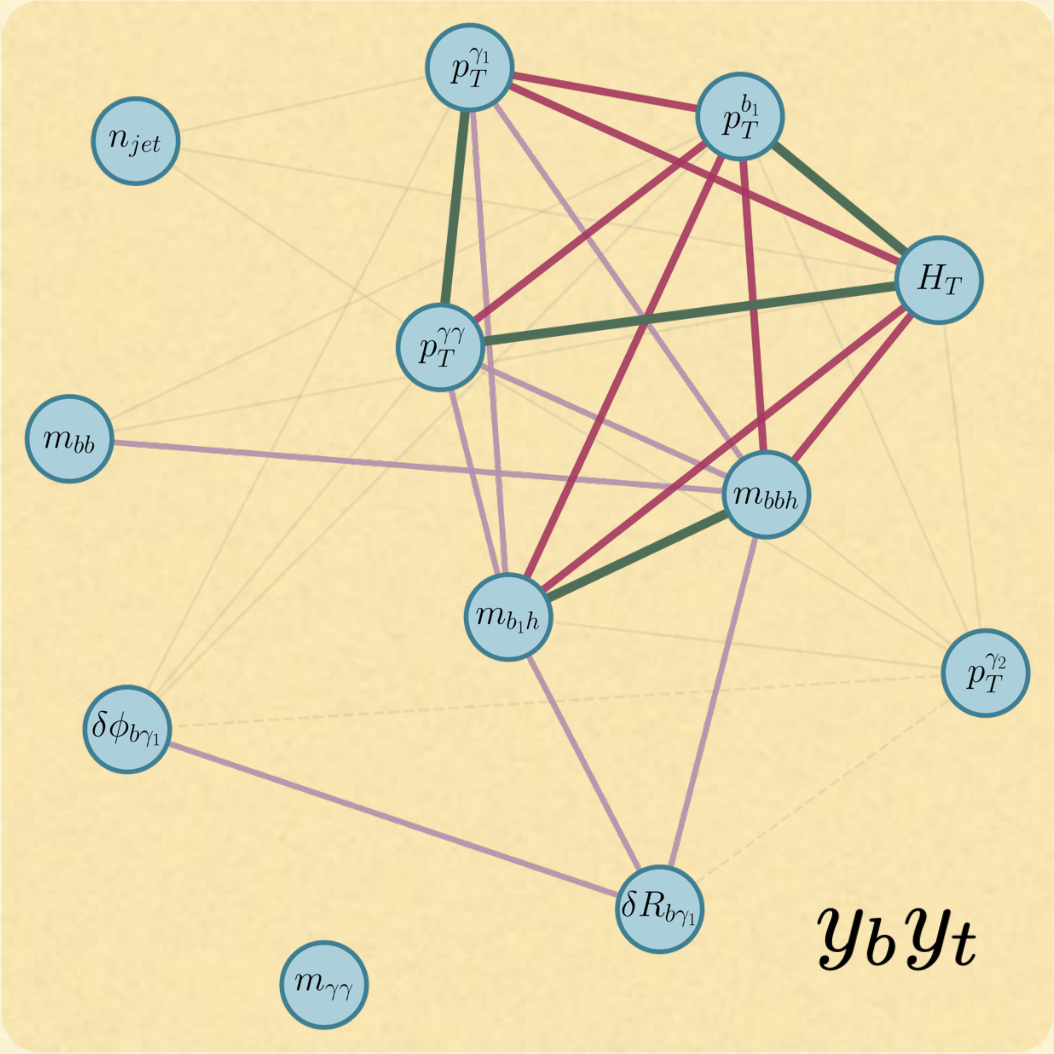


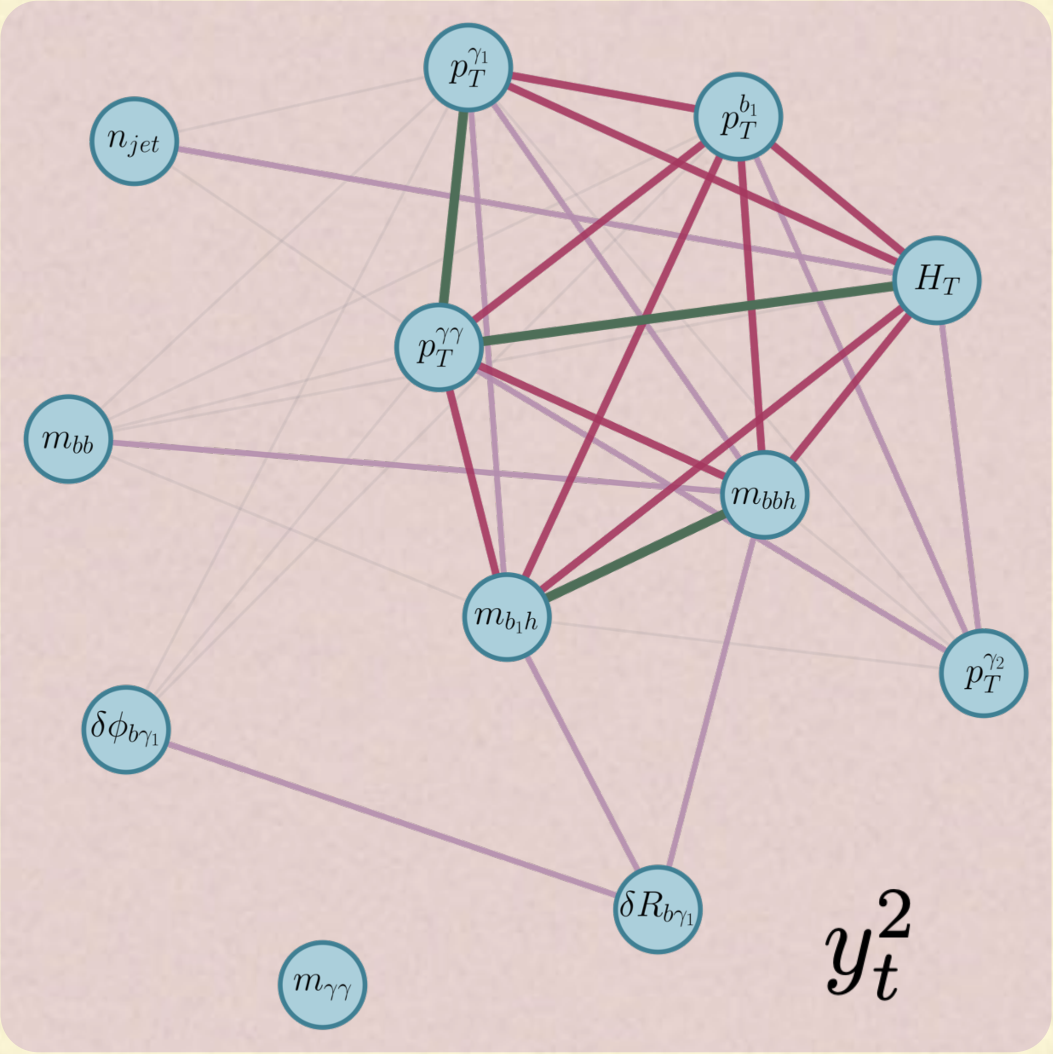
Into higher dimensions:

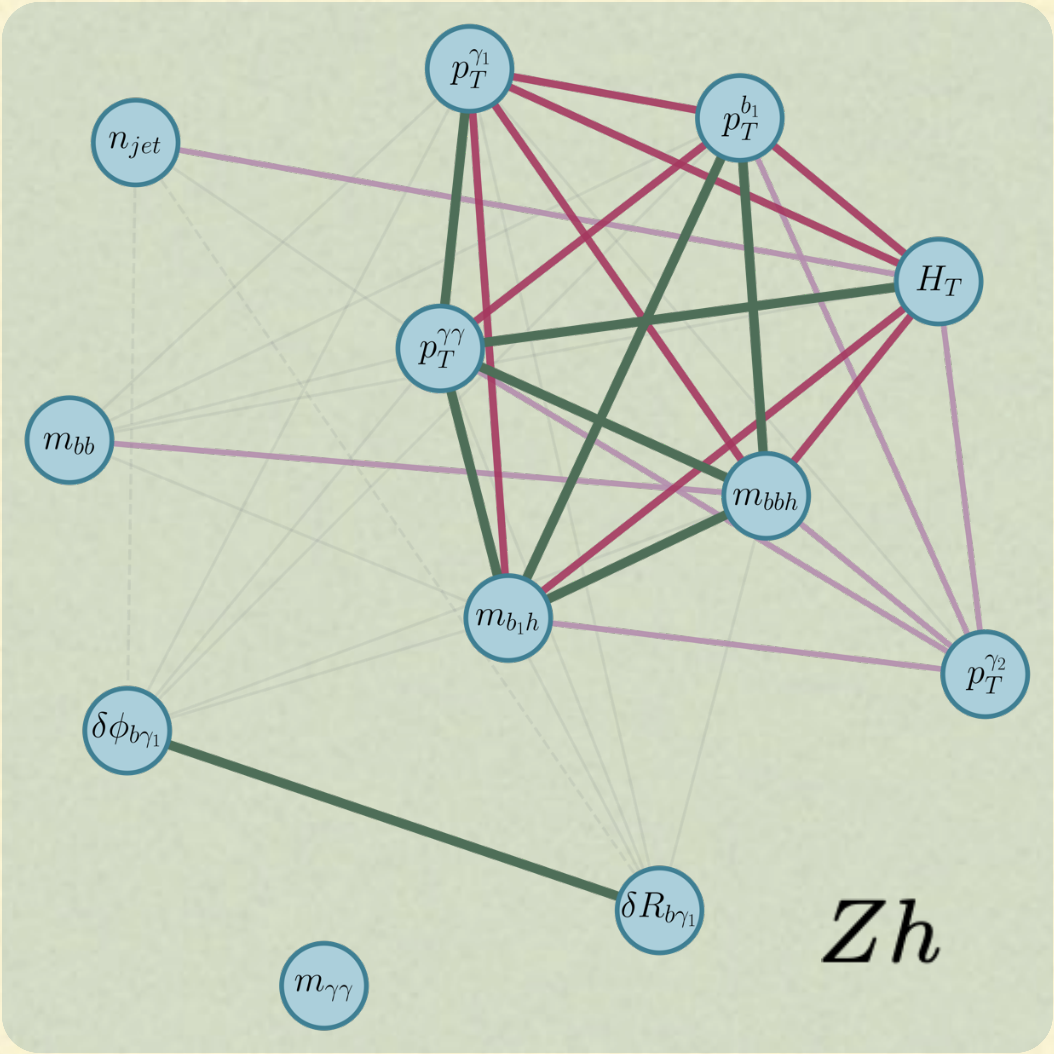
C. Grojean, A. Paul, ZQ(2020)

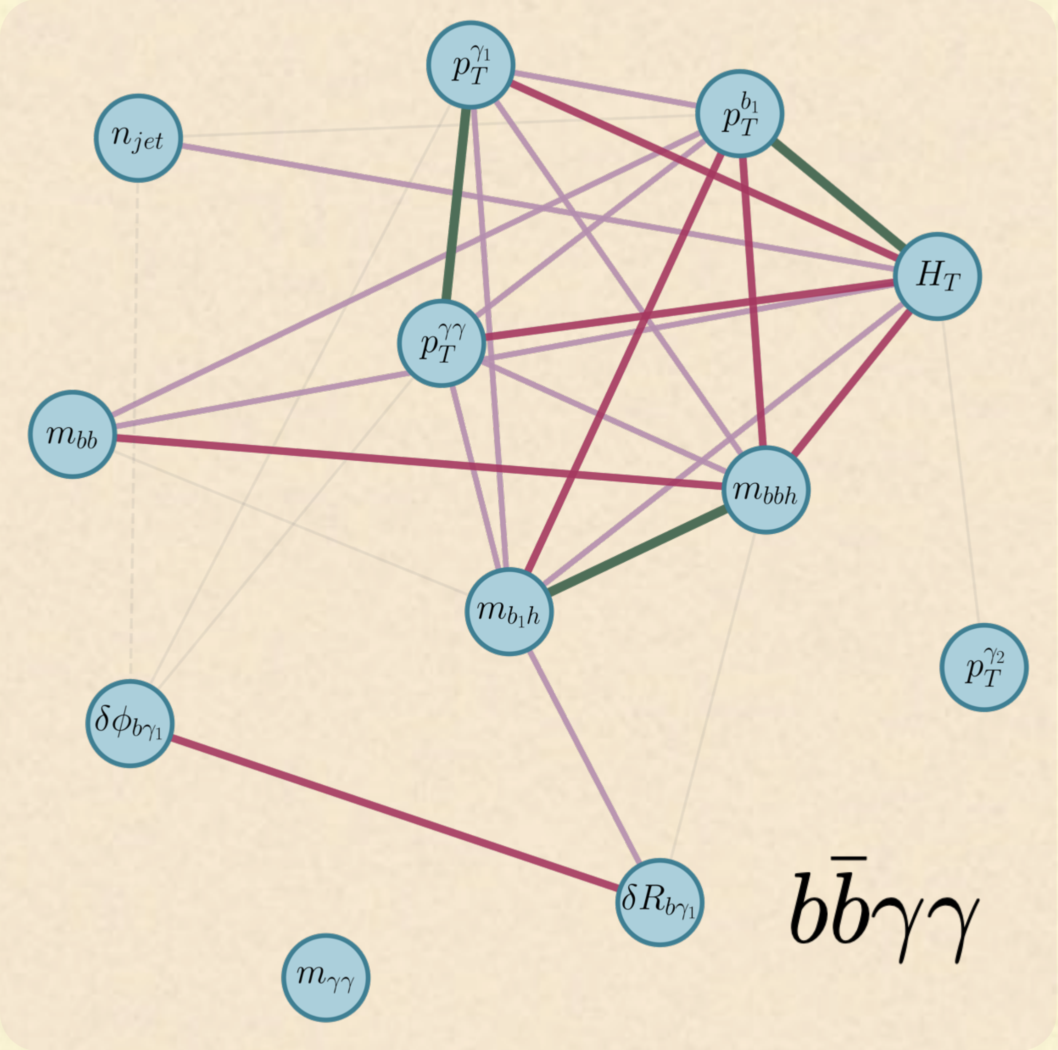
Diving into higher dimension











⬜ $0.2 \leq |\rho| < 0.4$ 🟪 $0.4 \leq |\rho| < 0.6$ 🟥 $0.6 \leq |\rho| < 0.8$ 🟩 $0.8 \leq |\rho| < 1.0$

--- $-\rho$ — $+\rho$

Into higher dimensions:

- The channel-specific multivariable correlation pattern (detector level)
- NLO (colour) effects, Parton shower, detector effects, etc.
- By Multivariable Analysis (MVA, e.g. BDT, NN) on simulated events

C. Grojean, A. Paul, ZQ(2020)

An Importance Measure/Distribution:

Machine Interpretation: Shapley value (2012):

- Shapley value: an importance "measure" of "group member", through marginalising contribution over the set:

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

- Shapley value approach log-likelihood ratio in binary-class:

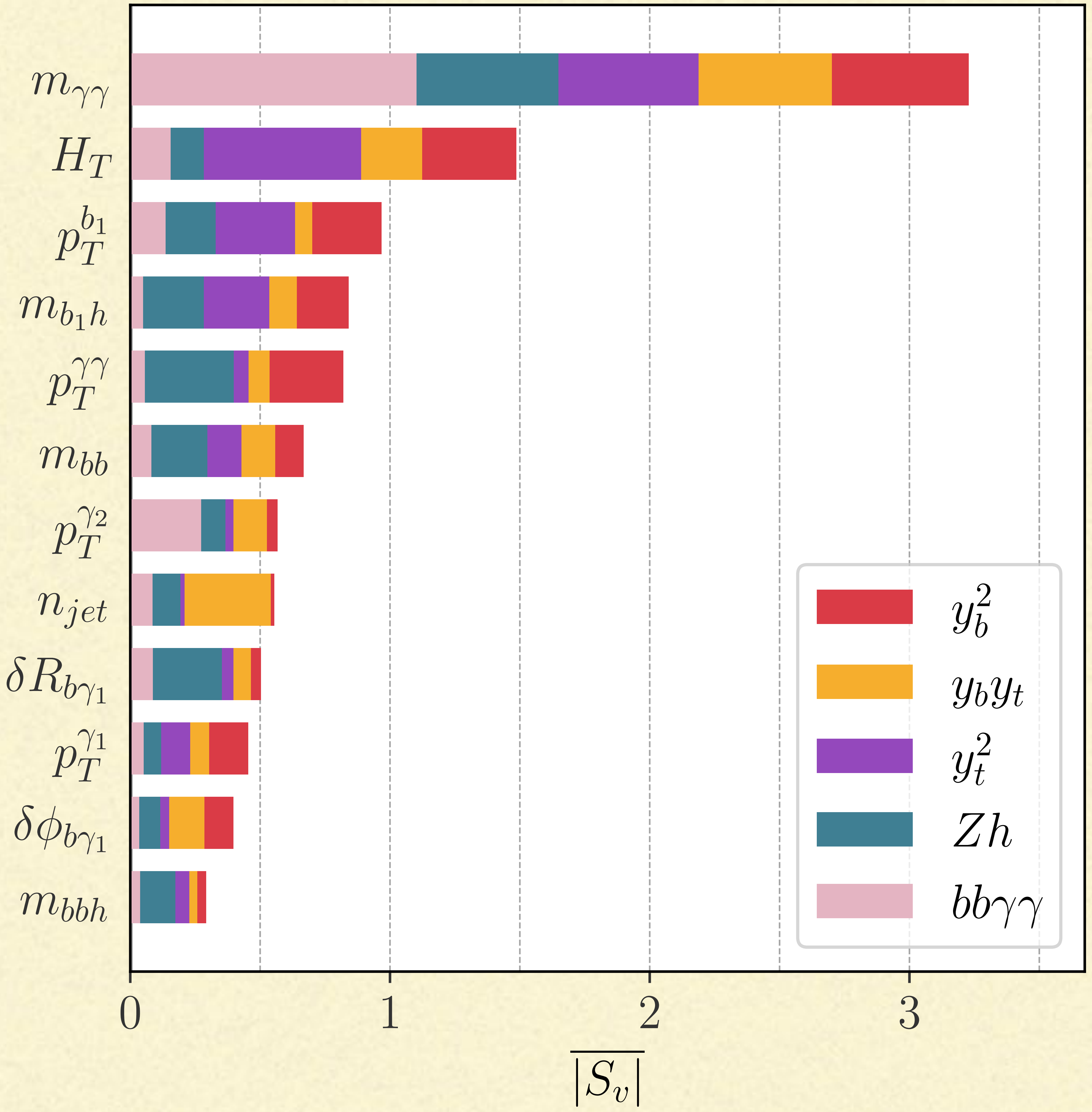
$$LL = \log(\frac{\sum_{i,j}\int f_i f_j|\mathcal{M}_1^{ij\to\vec{f}}|^2}{\sum_{I,J}\int f_I f_J|\mathcal{M}_2^{IJ\to\vec{f}}|^2}) \approx -S^{(n)}(v_1,..v_k)$$

- Feature importance: the averaged abs value of local Shapley:

$$I_j = \sum_{i=1}^{n} |\phi_j^{(i)}|$$

- Reduction of d.o.f. , Additivity over phase space, Distribution Correlation and more.. (Ongoing)

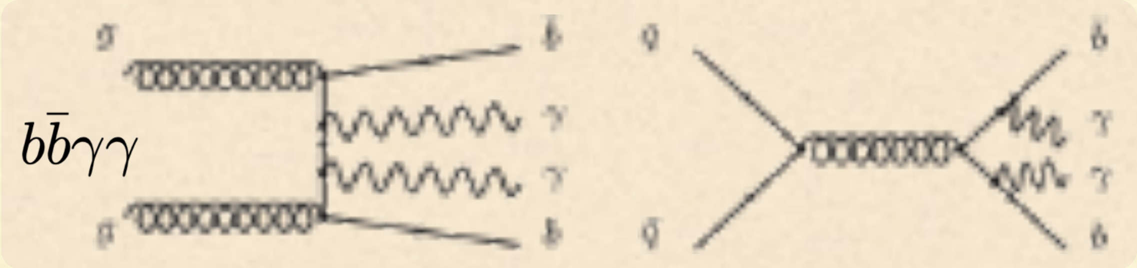
Shapley Feature Importance:

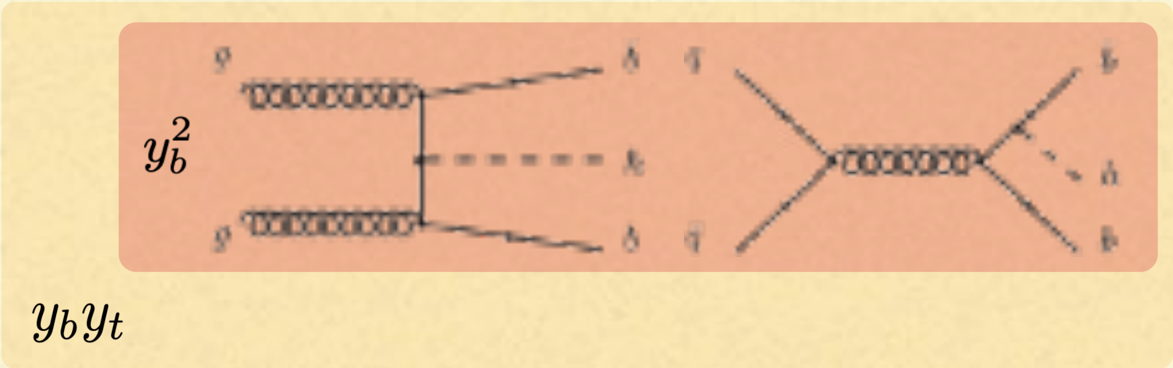


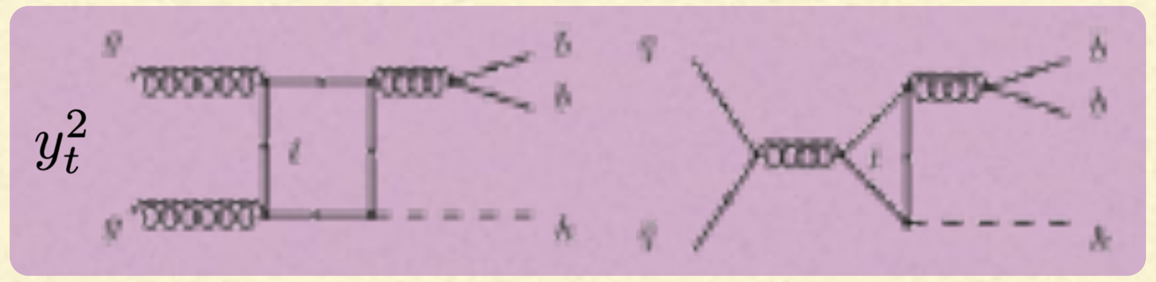
C. Grojean, A. Paul, ZQ(2020)

An Interpretable Framework:

Interpretable variables







QCD-QED Backgrounds

Signal

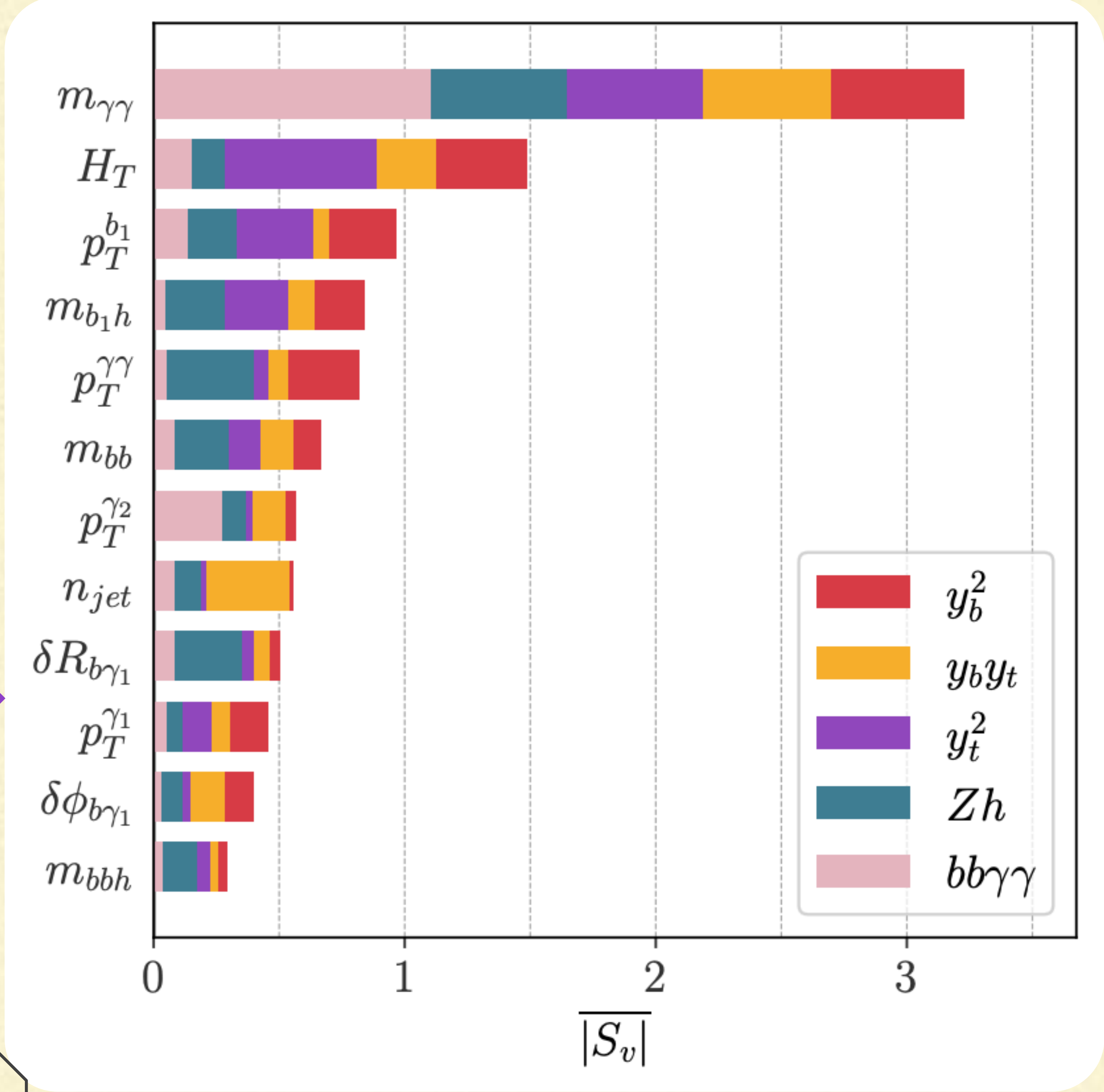
$b\bar{b}h$ backgrounds

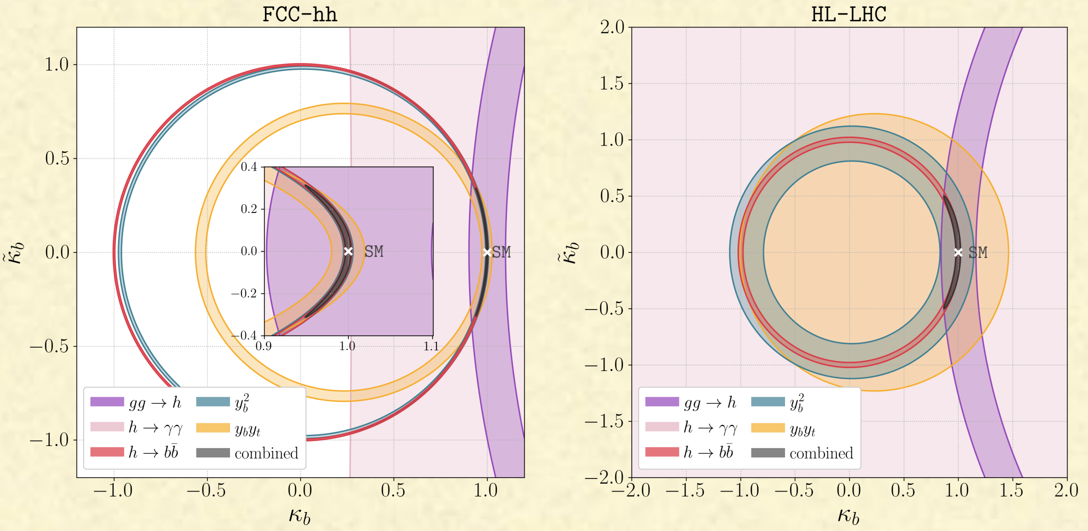
Shapley values + Physics Insights

Boosted Decision Trees + Signal Classification

Interpretable model

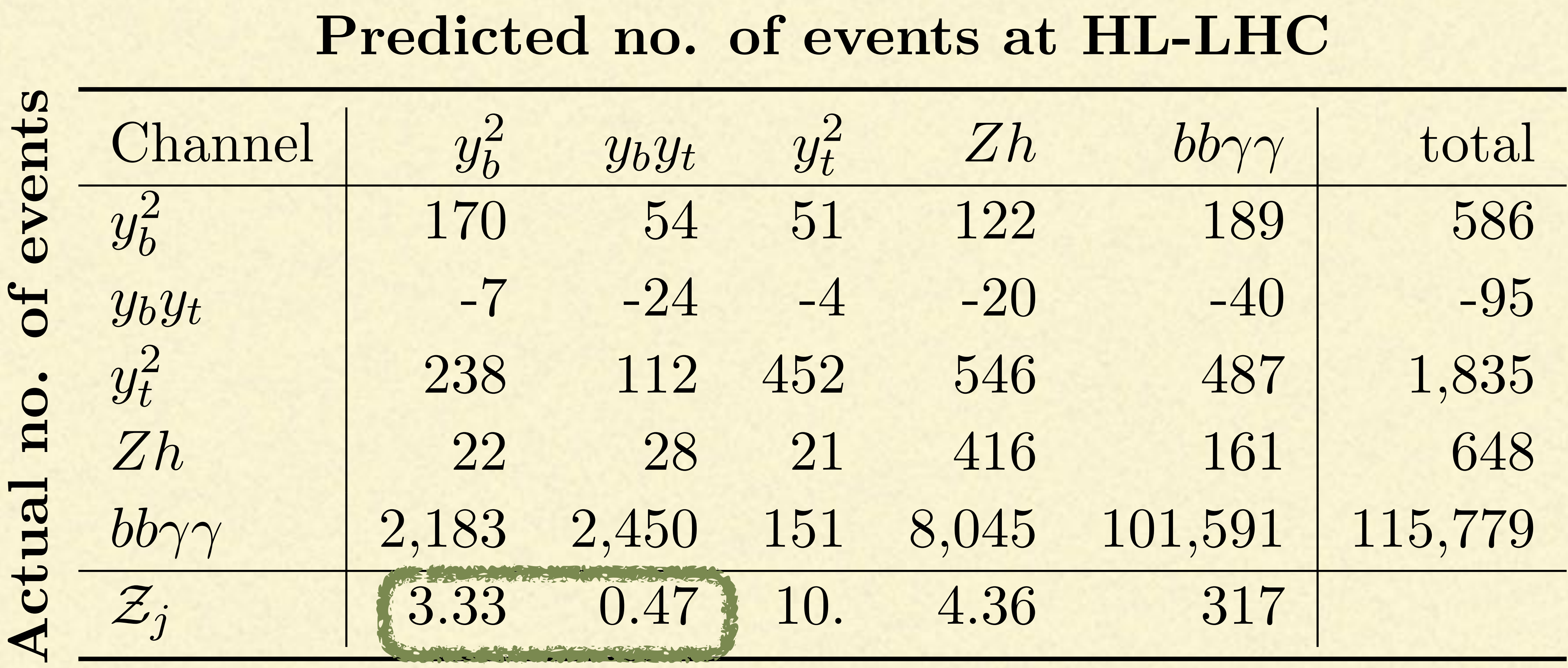
Interpretable results





C. Grojean, A. Paul, ZQ(2020)

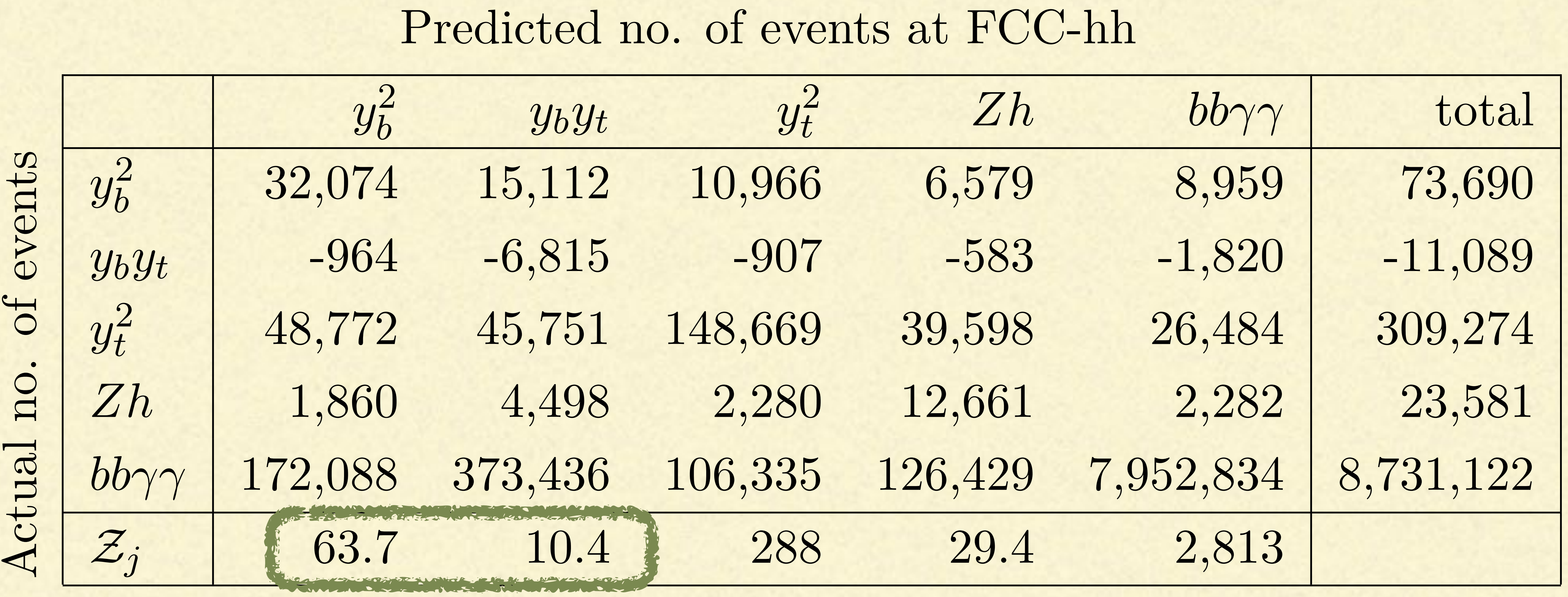
Improved Channel Sensitivity:



Predicted no. of events at HL-LHC

Actual no. of events

Channel	y_b^2	y_by_t	y_t^2	Zh	$bb\gamma\gamma$	total
y_b^2	170	54	51	122	189	586
y_by_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	



Predicted no. of events at FCC-hh

Actual no. of events

	y_b^2	y_by_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_by_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	

Optimised BDT/NN classification

$$Z_j = \frac{|N_{jj}|}{\sqrt{\sum_i N_{ij}}}$$

About $\sim 60\,\%$ gain in **significance** over traditional cut analysis.

C. Grojean, A. Paul, ZQ(2020)

Physics Interpretation:

A Real Bottom Yukawa: κ-scheme

$$\mathcal{L} \supset -\kappa_b \frac{m_b}{v} \bar{b} b h$$

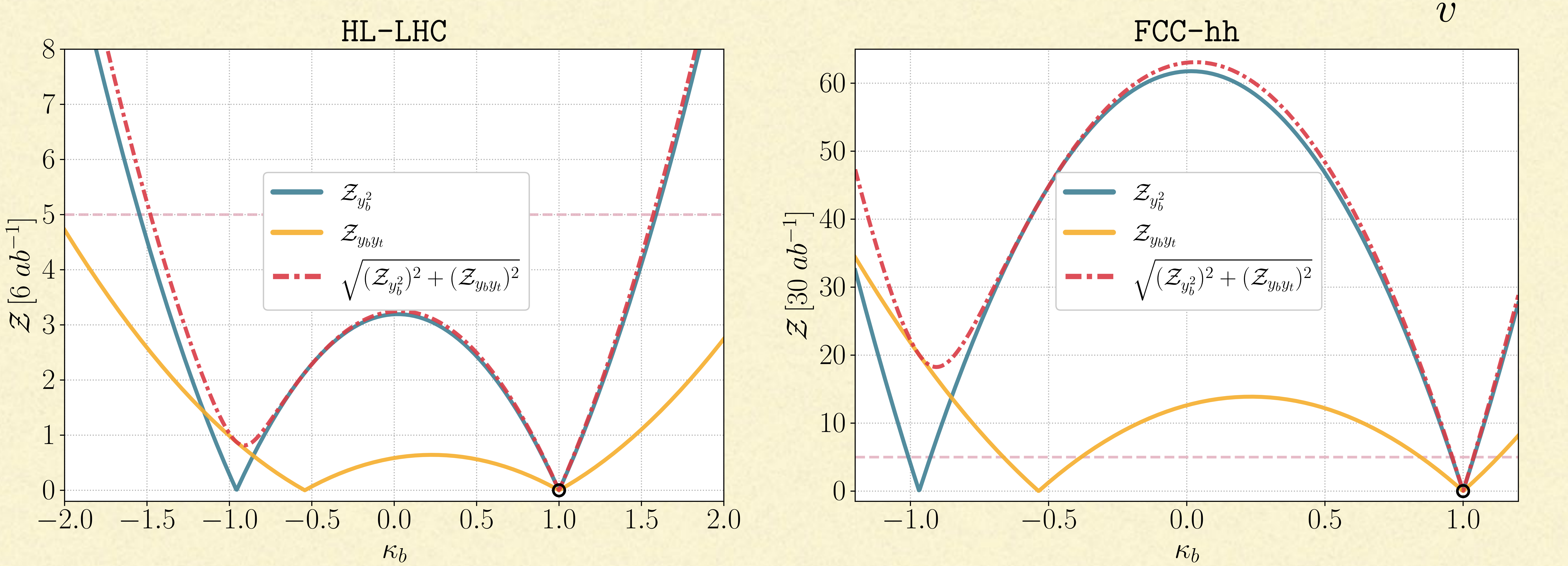


Figure 7. *Significance, $\mathcal{Z}$, as a function of κ_b at HL-LHC (ATLAS+CMS combined, 6 ab^{-1}) and FCC-hh (30 ab^{-1}). A SM signal is injected.*

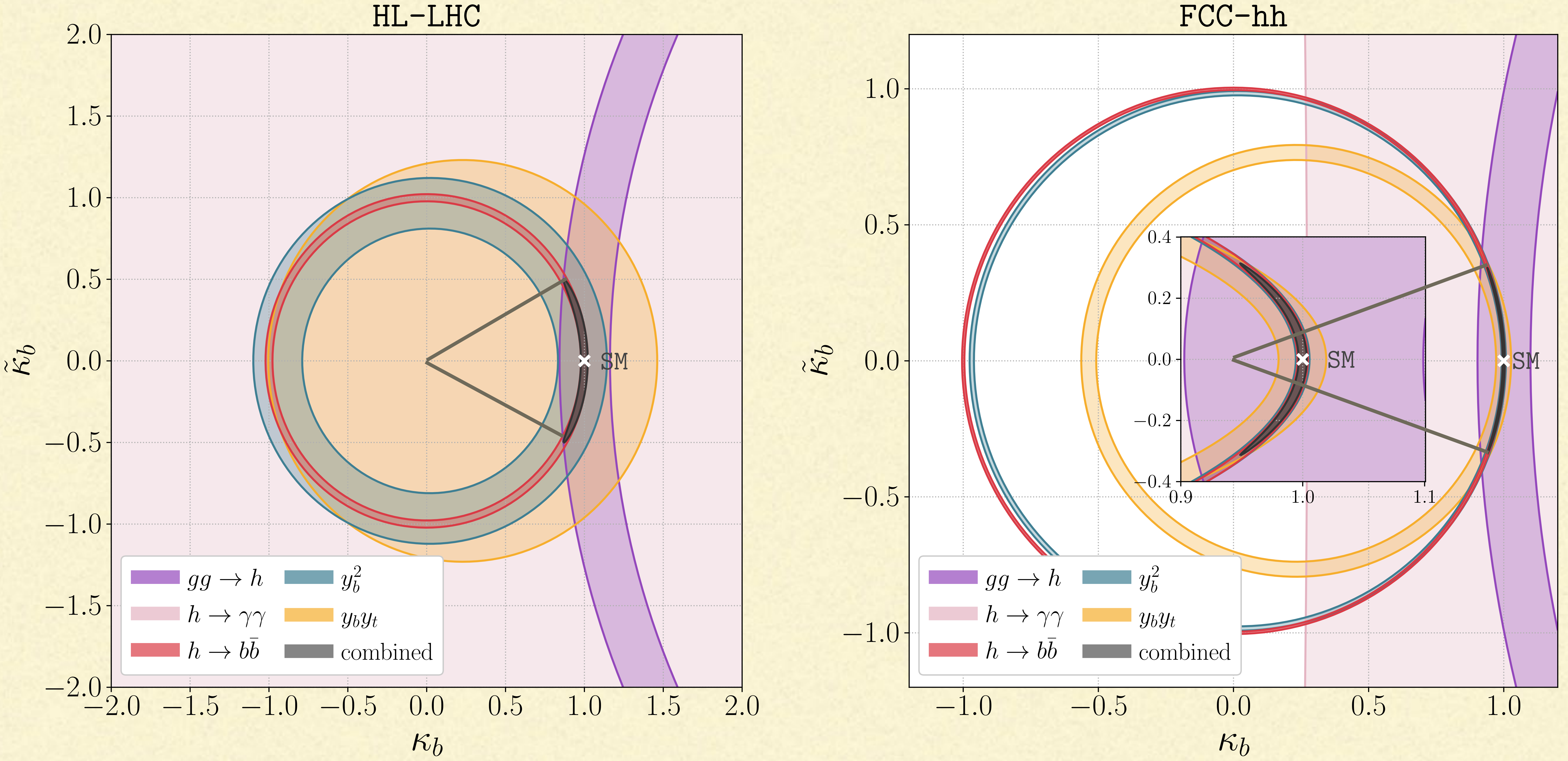
=> Unambiguous sign determination at FCC-hh.

C. Grojean, A. Paul, ZQ(2020)

Physics Interpretation:

A complex Bottom Yukawa (CP-phase)

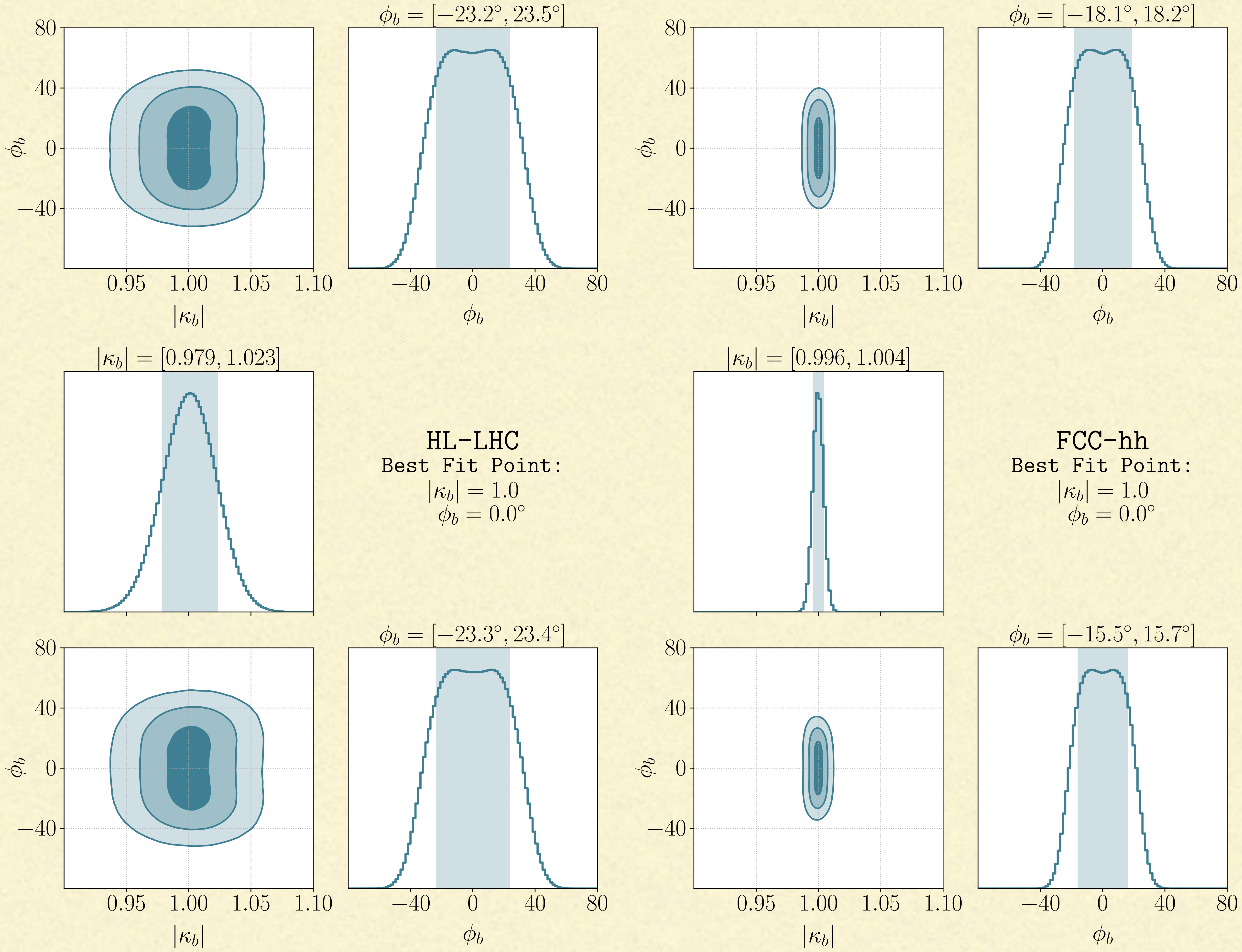
$$\mathscr{L} \sim -\frac{m_b}{v}(\kappa_b \bar{b} b + i\tilde{\kappa}_b \bar{b}\gamma_5 b)h$$



C. Grojean, A. Paul, ZQ(2020)

Physics Interpretation:

A complex Bottom Yukawa (CP-phase)



Comparison to LHC:

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

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

+ 15% to indirect bounds

Comparison to EDM:

Hadronic EDM (free of y_e 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: $\sum A\kappa_e\tilde{\kappa}_{\mathbf{q}} + B\tilde{\kappa}_e\kappa_{\mathbf{q}}$ => $\tilde{\kappa}_b \lesssim 0.5$

C. Grojean, A. Paul, ZQ(2020)

Conclusions:

- Associated production of $b\bar{b}h$ stands to gain at HL-LHC, FCC
- Direct sensitivity on a complex phase of y_b from interference term, compared to $gg \to h, h \to \gamma\gamma$, or e-EDM, n-EDM
- Multi-channel multi-dimensional final states benefit from or rely on MVA
- MVA (BDT, NN ML etc) can be better understood with importance measure such as Shapley values, retaining interpretability.

Backup

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.

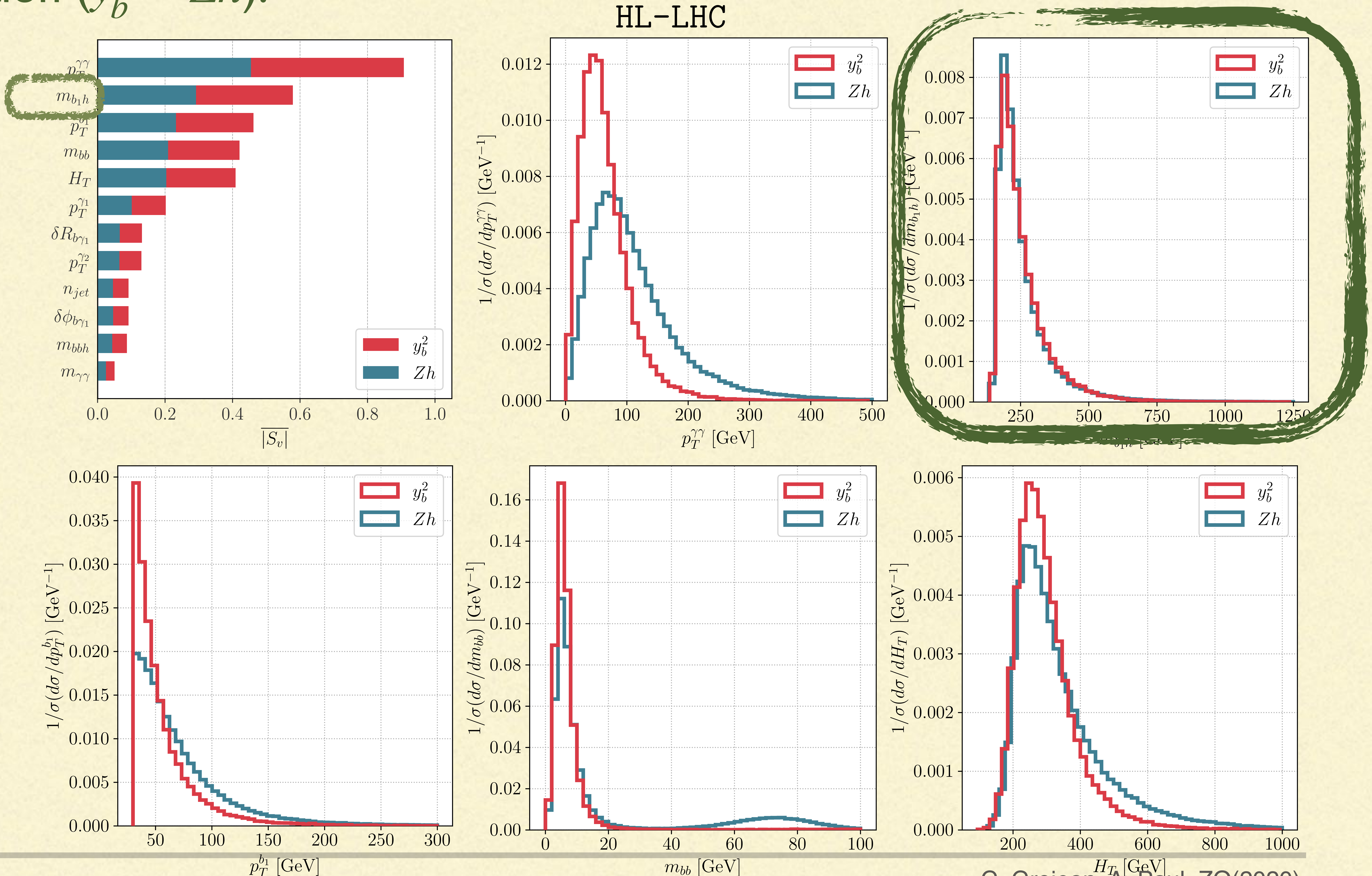
Papers on Higgs couplings fits

- 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.

Machine Interpretation ($y_b^2 - Zh$):

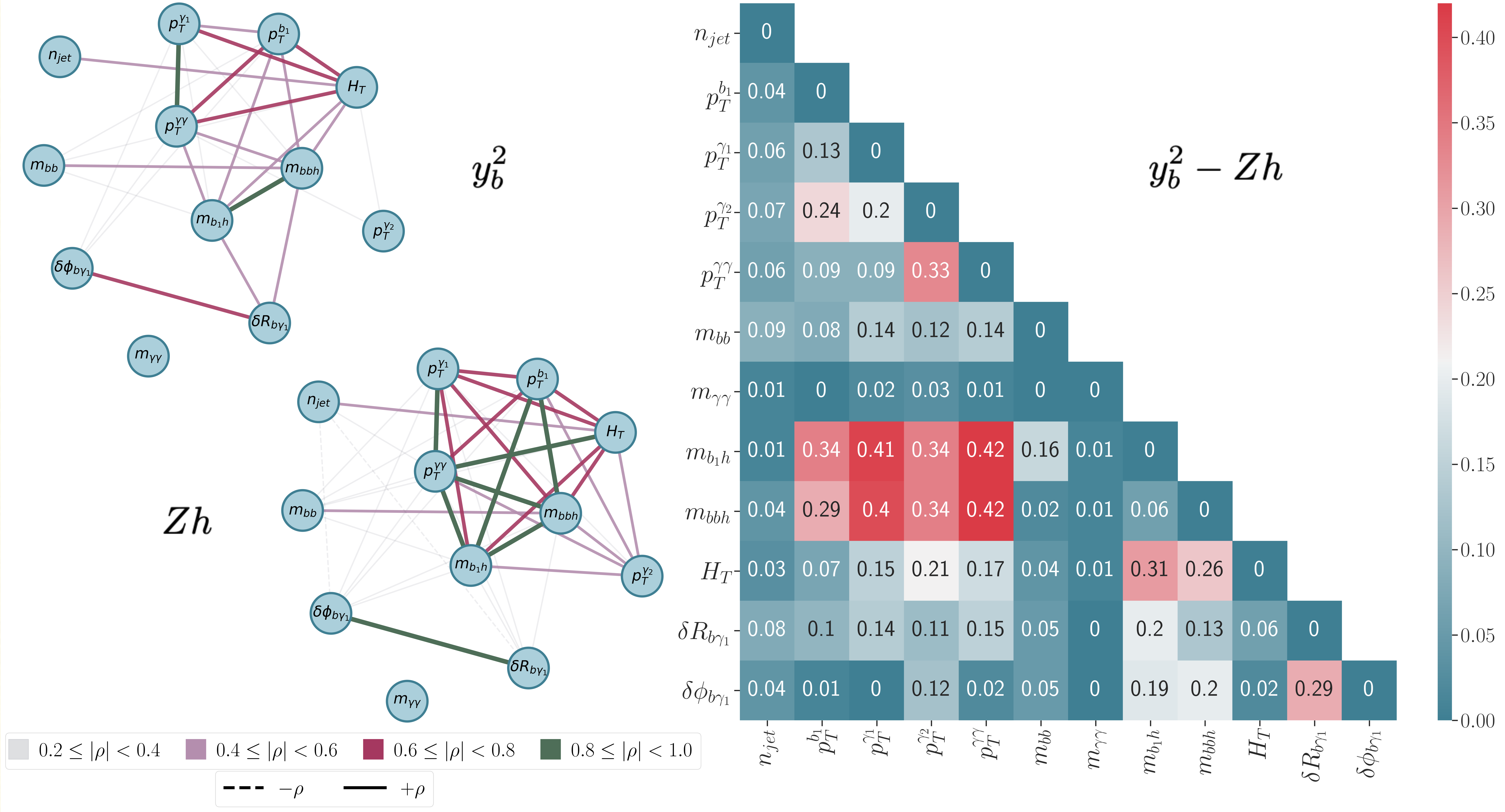


C. Grojean, A. Paul, ZQ(2020)

Machine Interpretation ($y_b^2 - Zh$):

“importance of m_{b1h} variable visualised through correlation”

HL-LHC

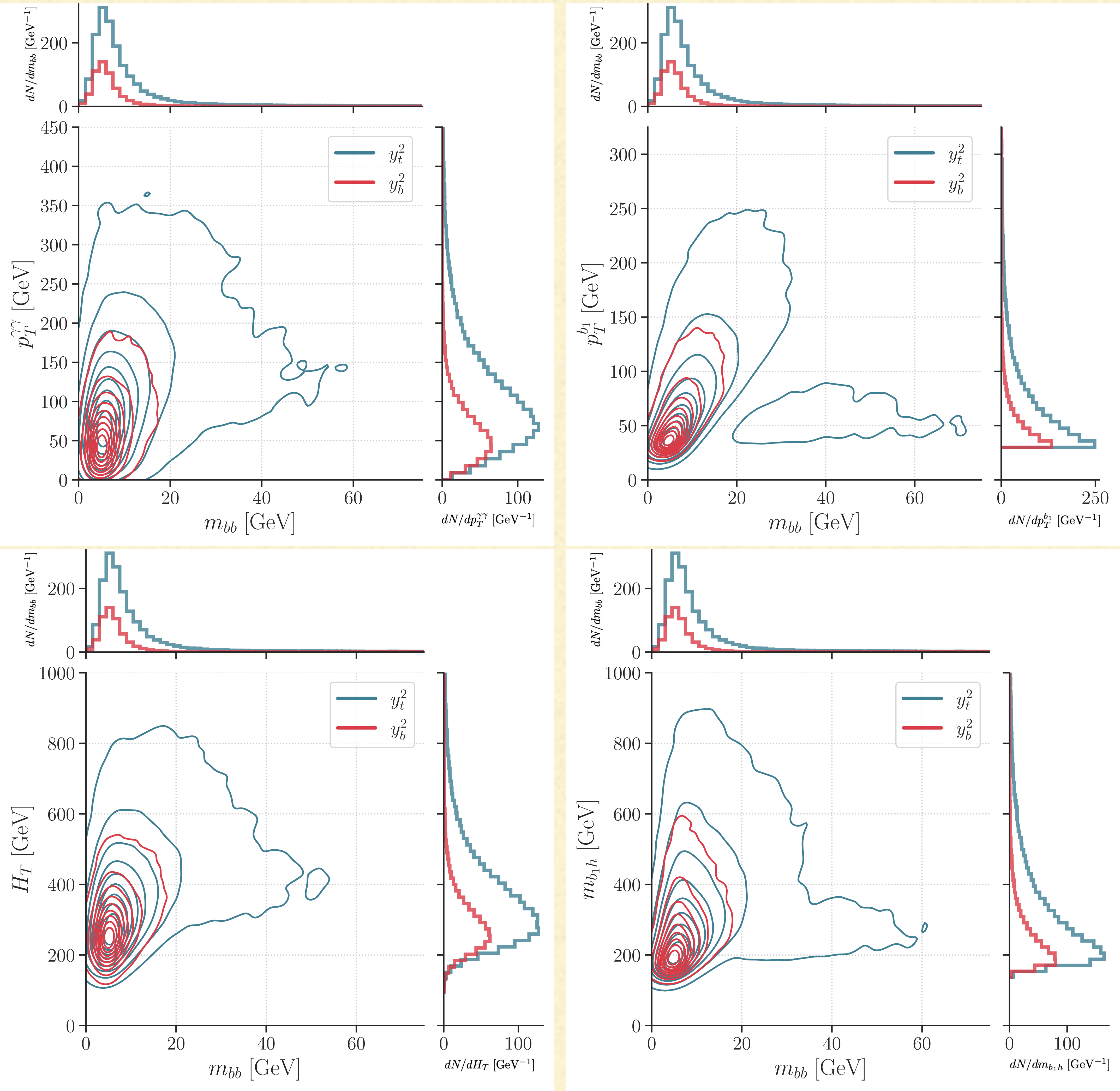


	n_{jet}	$p_T^{b_1}$	$p_T^{\gamma_1}$	$p_T^{\gamma_2}$	$p_T^{\gamma\gamma}$	m_{bb}	$m_{\gamma\gamma}$	m_{b_1h}	m_{bbh}	H_T	$\delta R_{b\gamma_1}$	$\delta\phi_{b\gamma_1}$
n_{jet}	0											
$p_T^{b_1}$	0.04	0										
$p_T^{\gamma_1}$	0.06	0.13	0									
$p_T^{\gamma_2}$	0.07	0.24	0.2	0								
$p_T^{\gamma\gamma}$	0.06	0.09	0.09	0.33	0							
m_{bb}	0.09	0.08	0.14	0.12	0.14	0						
$m_{\gamma\gamma}$	0.01	0	0.02	0.03	0.01	0	0					
m_{b_1h}	0.01	0.34	0.41	0.34	0.42	0.16	0.01	0				
m_{bbh}	0.04	0.29	0.4	0.34	0.42	0.02	0.01	0.06	0			
H_T	0.03	0.07	0.15	0.21	0.17	0.04	0.01	0.31	0.26	0		
$\delta R_{b\gamma_1}$	0.08	0.1	0.14	0.11	0.15	0.05	0	0.2	0.13	0.06	0	
$\delta\phi_{b\gamma_1}$	0.04	0.01	0	0.12	0.02	0.05	0	0.19	0.2	0.02	0.29	0

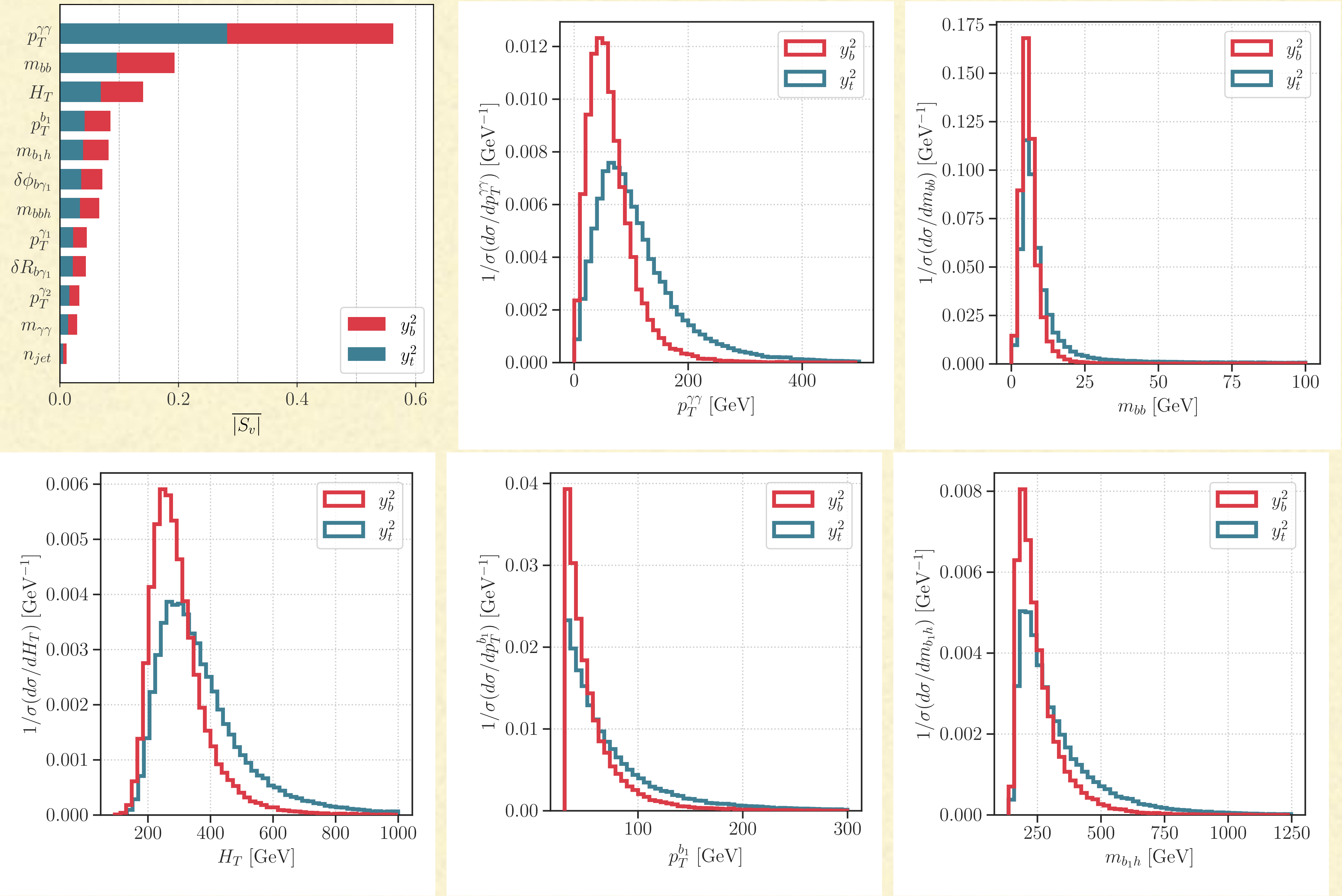
C. Grojean, A. Paul, ZQ(2020)

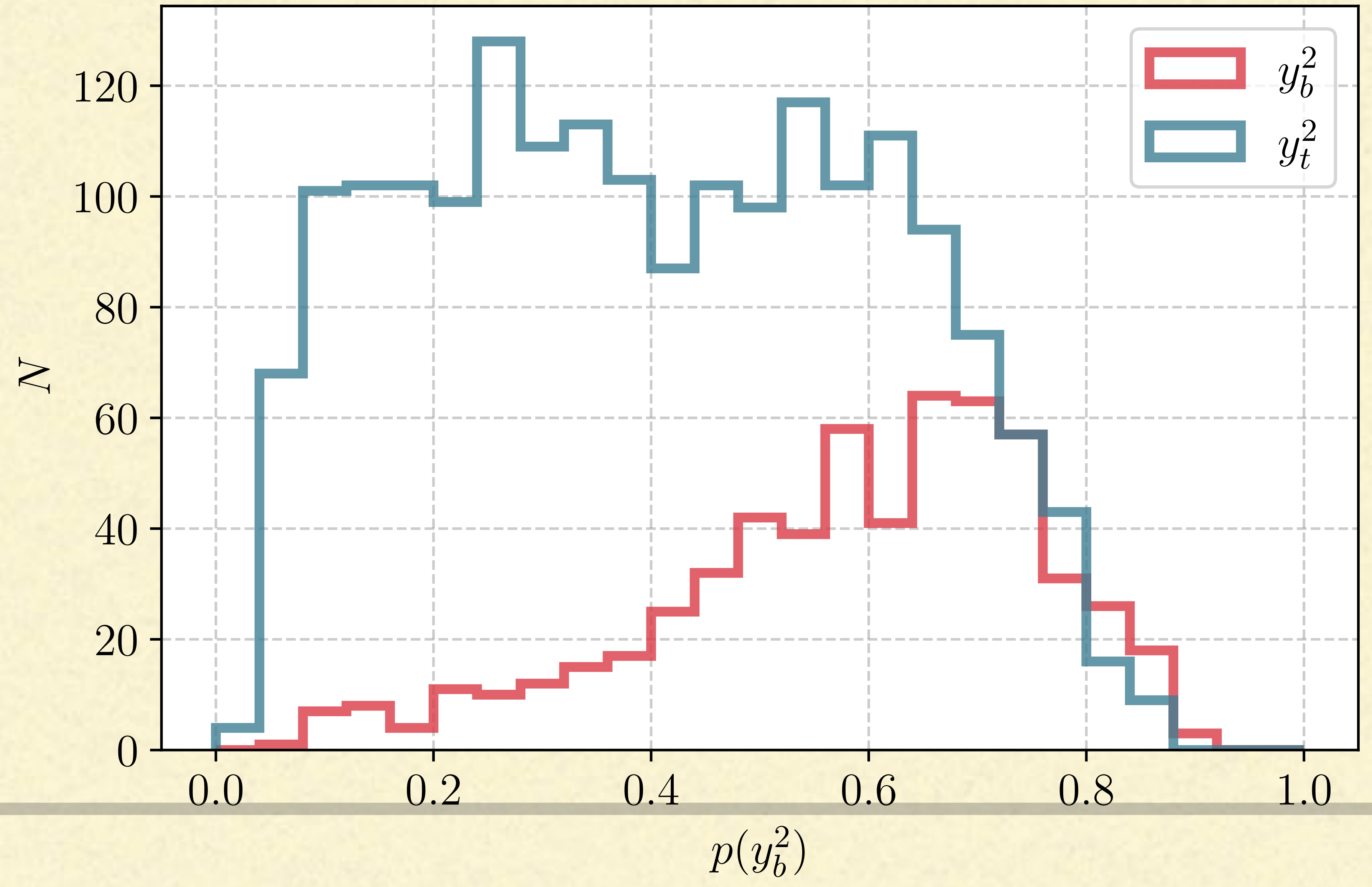
Machine Interpretation ($y_b^2 - y_t^2$):

HL-LHC



HL-LHC

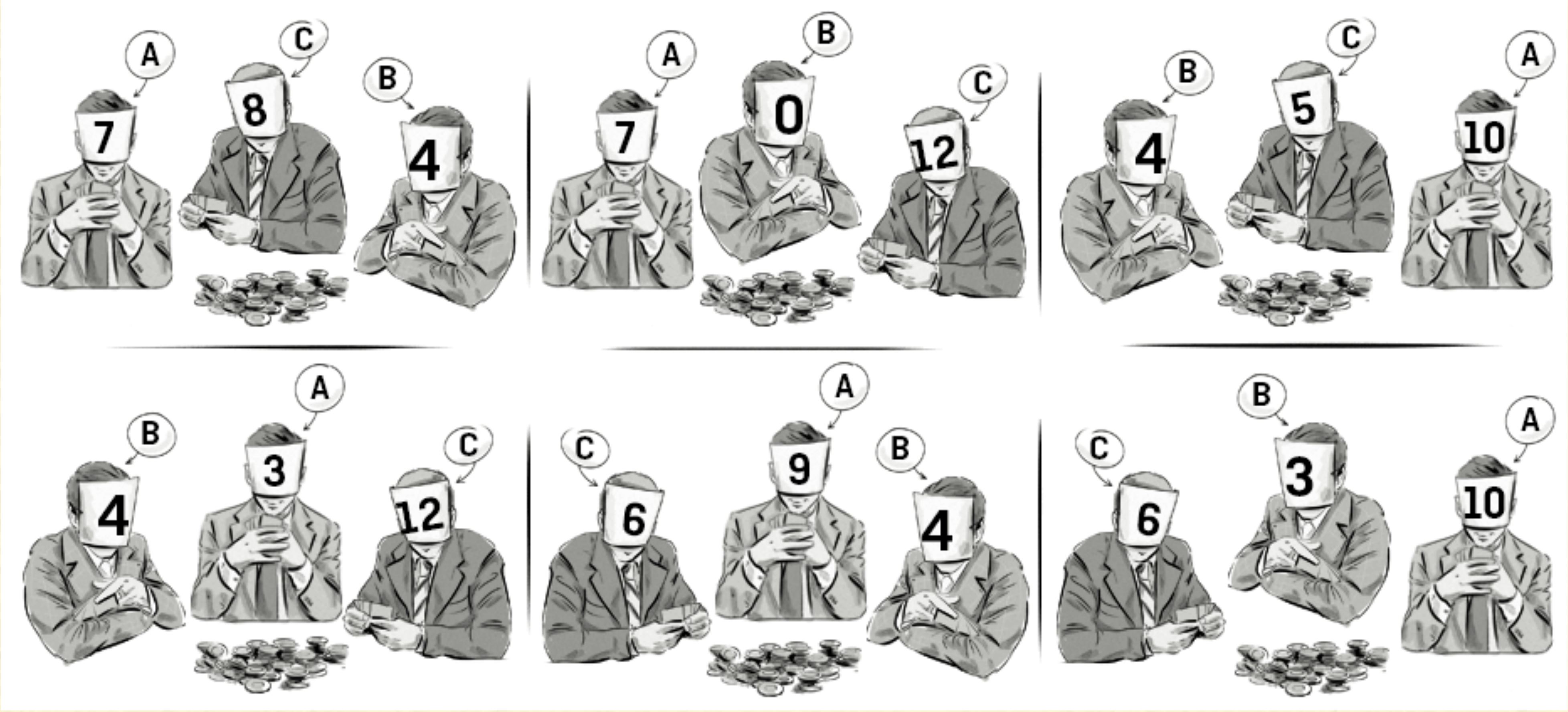




Shapley value from Cooperative game theory :

The value of each player and each combination of players

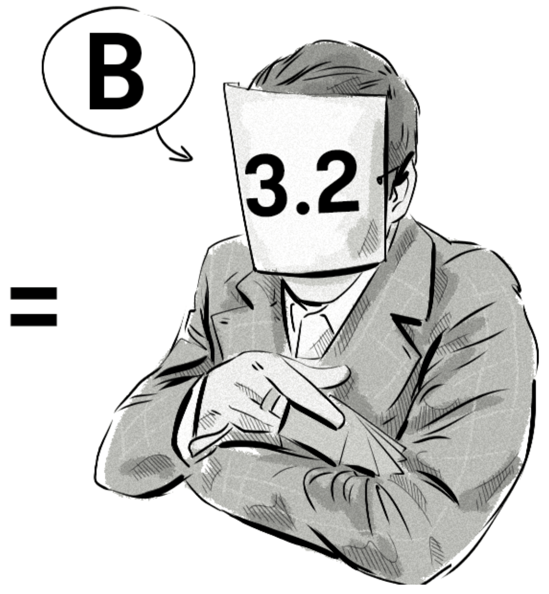
The value of the player in each game



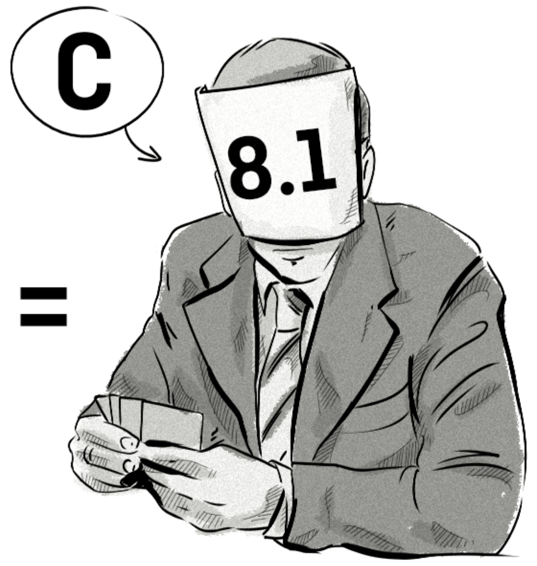
Marginalized values

(7+7+10+3+9+10) / 6 =

(4+0+4+4+4+3) / 6 =



(8+12+5+12+6+6) / 6 =



The most important player

L. S. Shapley, Notes on the n-Person Game-II: The Value of an n-Person Game (1951).

$b\bar{b}h$: Additional background 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

$b\bar{b}h$: Additional background discussion

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