

Examining the CP properties of the top-Yukawa and the ggH coupling

Marco Menen

Higgs 2024, Uppsala, 06.11.2024

Based on [\[2309.03146\]](#) and [\[2406.03950\]](#)

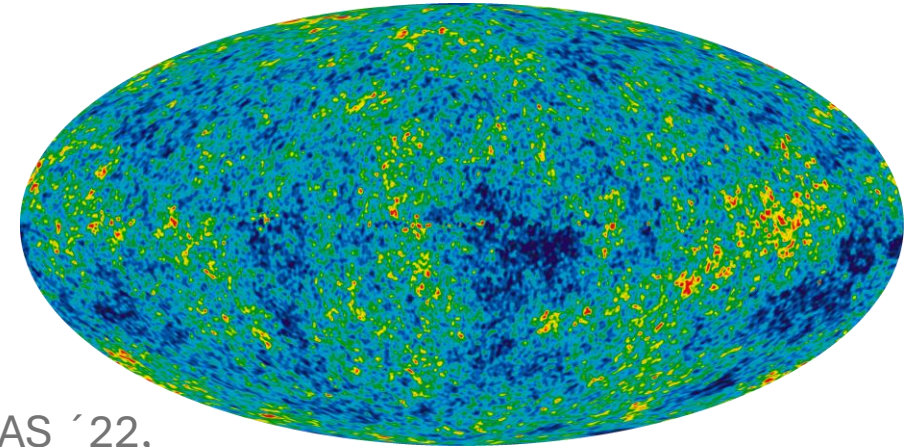
In collaboration with H. Bahl, A. Carnelli, F. Deliot, E. Fuchs, M. Hannig, A. Kotsokechagia, M. Saimpert, L. Olivier Schoeffel



Overview & Motivation

The need for CP violation:

- The SM fails to explain the observed baryon asymmetry of the universe
- Higgs CP violation could help with this issue
- Current experimental limits leave room for CP violation in the Higgs sector [CMS '21](#), [ATLAS '21](#), [ATLAS '22](#), [CMS '22](#), [CMS '23](#), [ATLAS '24](#) ...



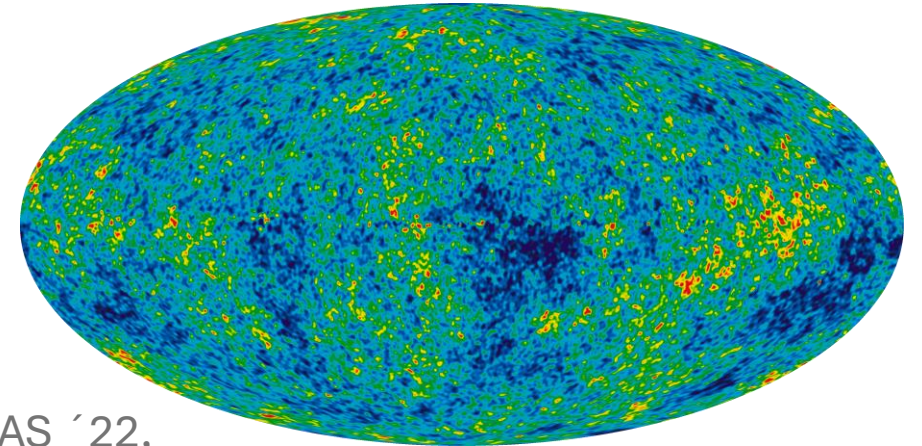
The Higgs-top quark coupling:

- Well accessible at the LHC, only $\mathcal{O}(1)$ Yukawa coupling
- Complementary constraints from indirect probes ($ggF2j$) and direct probes (ttH)

Overview & Motivation

The need for CP violation:

- The SM fails to explain the observed baryon asymmetry of the universe
- Higgs CP violation could help with this issue
- Current experimental limits leave room for CP violation in the Higgs sector [CMS '21](#), [ATLAS '21](#), [ATLAS '22](#), [CMS '22](#), [CMS '23](#), [ATLAS '24](#) ...



The Higgs-top quark coupling:

- Well accessible at the LHC, only $\mathcal{O}(1)$ Yukawa coupling
- Complementary constraints from indirect probes (**ggF2j**) and direct probes (**ttH**)
- **Two good choices**
(Unlike US presidential election)
(Personal bias included)

Parameterization and framework

- Higgs characterization model: Higgs H assumed to be mixed CP state

Artoisenet et al. '13

Mapping to
SMEFT: backup

- Allows for (effective) couplings with modified CP-character via CP-even (CP-odd) coupling modifiers c_i (\tilde{c}_i)

Also: $\alpha_i = \tan^{-1} \frac{\tilde{c}_i}{c_i}$

$$\mathcal{L}_{\text{Yuk}} \supset -\frac{y_t^{\text{SM}}}{\sqrt{2}} \bar{t}(c_t + i\gamma_5 \tilde{c}_t)tH$$

Direct test of y_t
few events (ttH)

$$\mathcal{L}_{ggH} = -\frac{1}{4v} \left(-\frac{\alpha_s}{3\pi} c_g G_{\mu\nu}^a G^{\mu\nu,a} + \frac{\alpha_s}{2\pi} \tilde{c}_g G_{\mu\nu}^a \tilde{G}^{\mu\nu,a} \right) H$$

Indirect test of y_t
many events (ggF)

with $c_g, \tilde{c}_g \rightarrow c_t, \tilde{c}_t$ in the heavy top limit and without low-mass colored BSM particles in the ggF loop

Structure

Part I: ggF2j

- Main production mode at LHC
- 2 jets in the final state needed to construct CP-odd observables
- Combine all CP information in a classifier in the $H \rightarrow \gamma\gamma$ channel
- Many measurements conducted in VBF-like signal region due to more $q\bar{q}$ initial state events, what about ggF-like signal region

Structure

Part I: ggF2j

- Main production mode at LHC
- 2 jets in the final state needed to construct CP-odd observables
- Combine all CP information in a classifier in the $H \rightarrow \gamma\gamma$ channel
- Many measurements conducted in VBF-like signal region due to more $q\bar{q}$ initial state events, what about ggF-like signal region

Part II: ttH

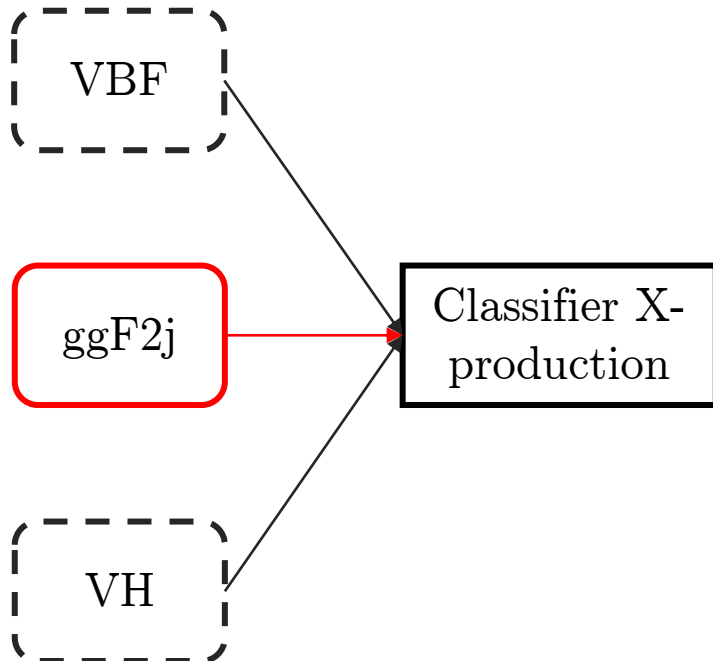
- Fewer events but coupling measured directly
- Measuring CP-odd variables not feasible currently, concentrate on CP-sensitive observables
- Combine $H \rightarrow \gamma\gamma$, $H \rightarrow b\bar{b}$ and multilep. final states
- Extension of the current STXS $p_{T,H}$ binning by a second observable

Part I: ggF2j

Analysis strategy

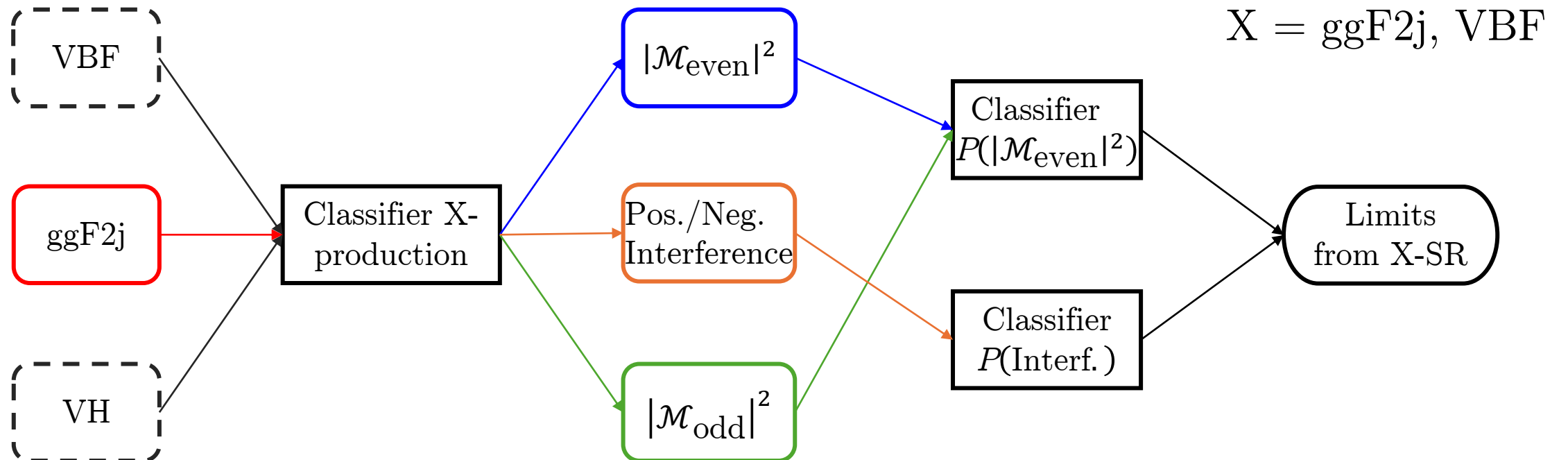
- First distinguish ggF2j from relevant Higgs backgrounds (backup)
- Current most sensitive study in VBF-like signal region [CMS '22](#)

$$X = \text{ggF2j}, \text{VBF}$$



Analysis strategy

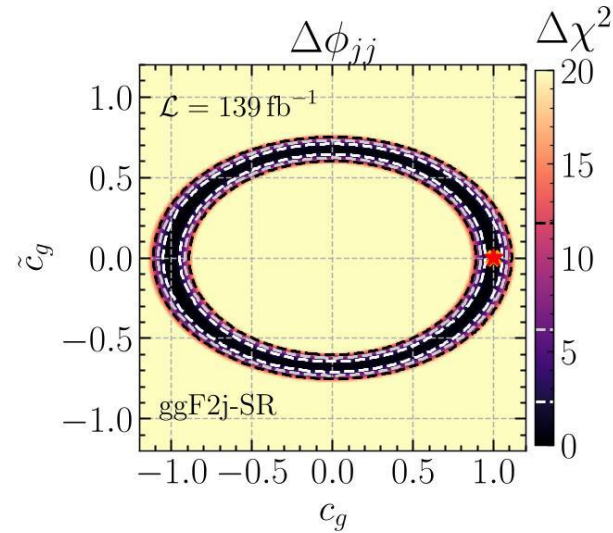
$$|\mathcal{M}_{\text{ggF2j}}|^2 = c_g^2 |\mathcal{M}_{\text{even}}|^2 + \underbrace{2c_g \tilde{c}_g \text{Re}[\mathcal{M}_{\text{even}} \mathcal{M}_{\text{odd}}^*]}_{\text{Interference}} + \tilde{c}_g^2 |\mathcal{M}_{\text{odd}}|^2$$



- Train a CP-even and a CP-odd classifier in a ggF2j-SR and a VBF-SR

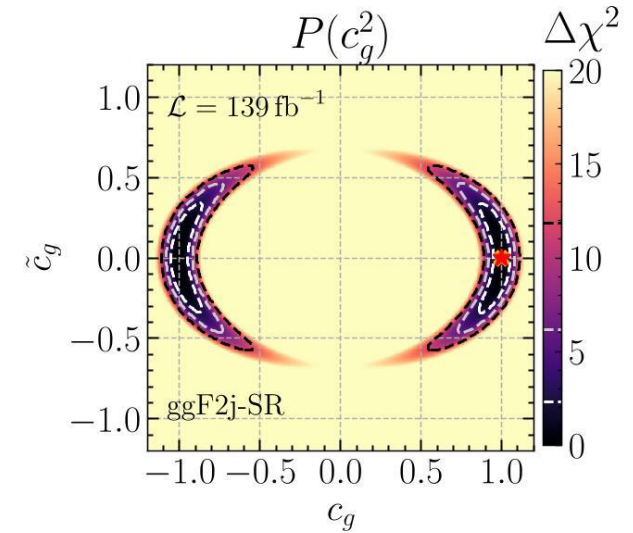
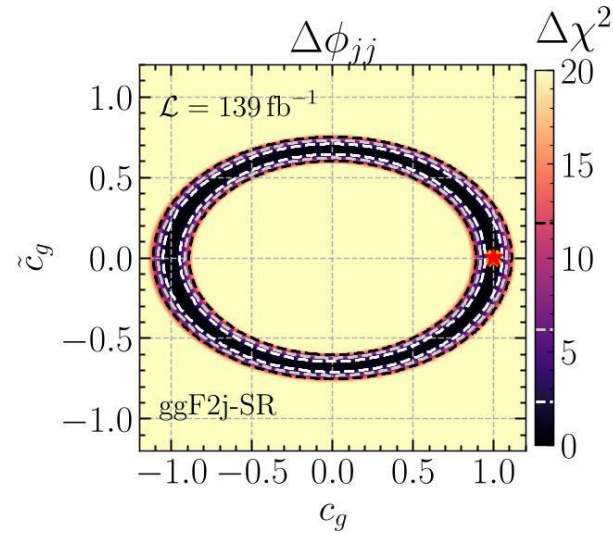
ggF2j signal region

- Ellipse from total rate
- $\Delta\phi_{jj}$ alone is not able to resolve the ellipse



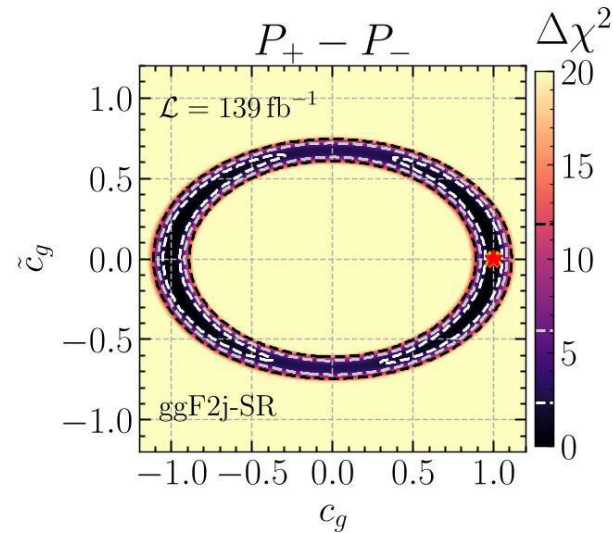
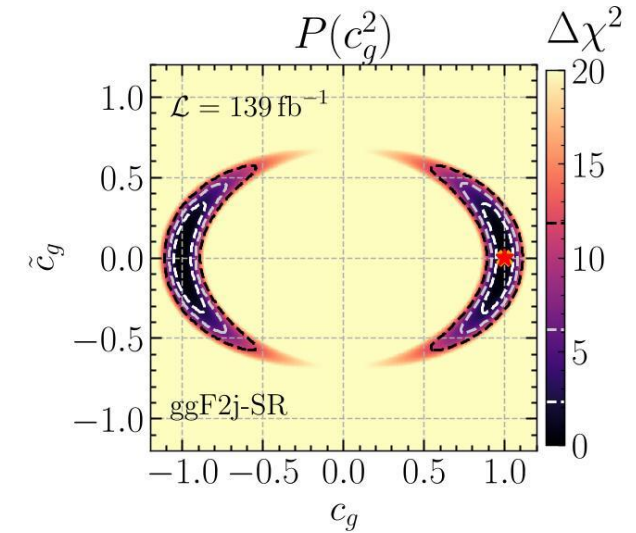
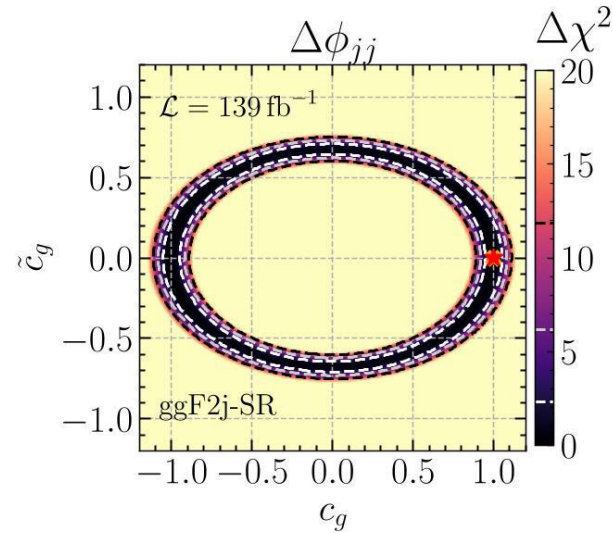
ggF2j signal region

- Ellipse from total rate
- $\Delta\phi_{jj}$ alone is not able to resolve the ellipse
- CP-even classifier can resolve the ellipse well



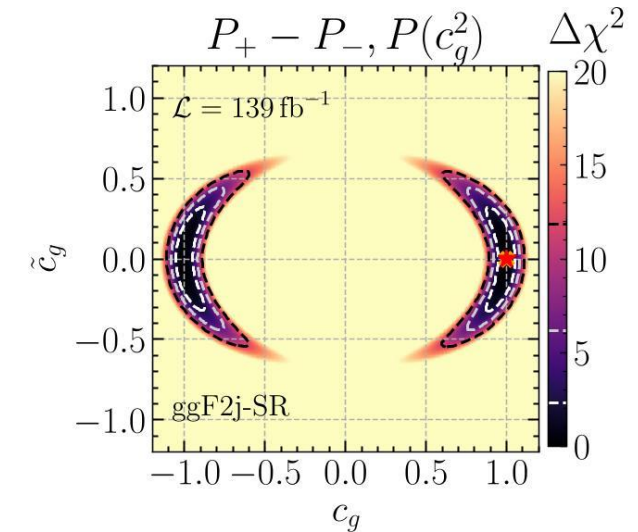
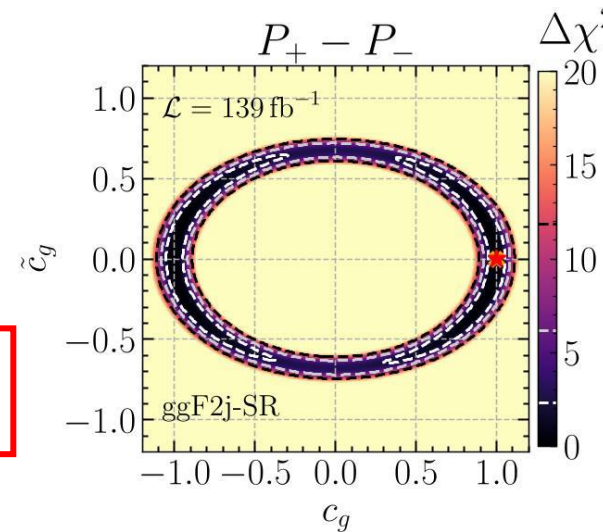
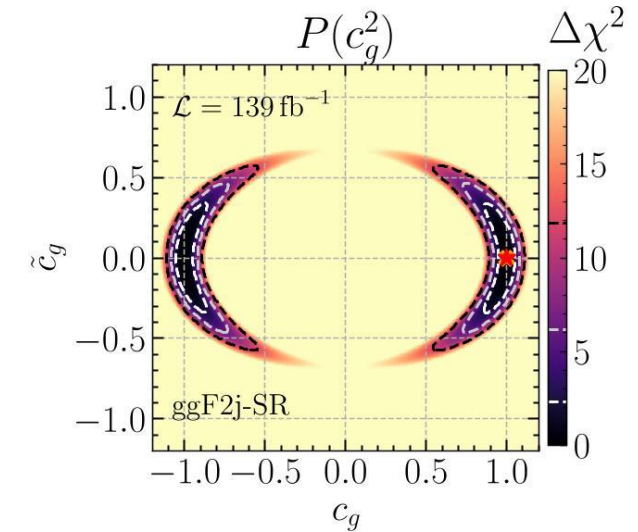
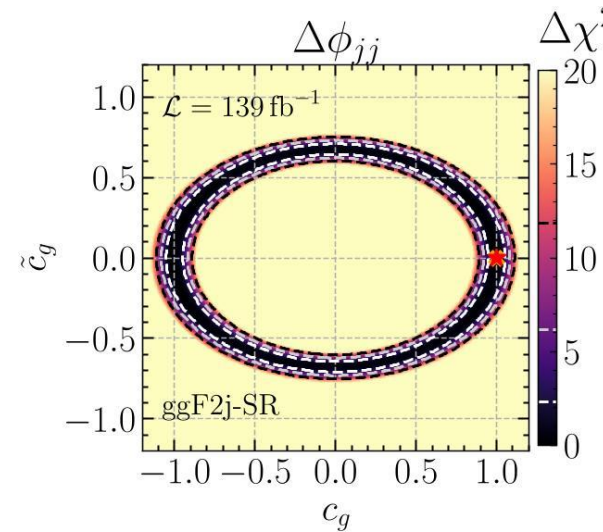
ggF2j signal region

- Ellipse from total rate
- $\Delta\phi_{jj}$ alone is not able to resolve the ellipse
- CP-even classifier can resolve the ellipse well
- CP-odd classifier performs much worse

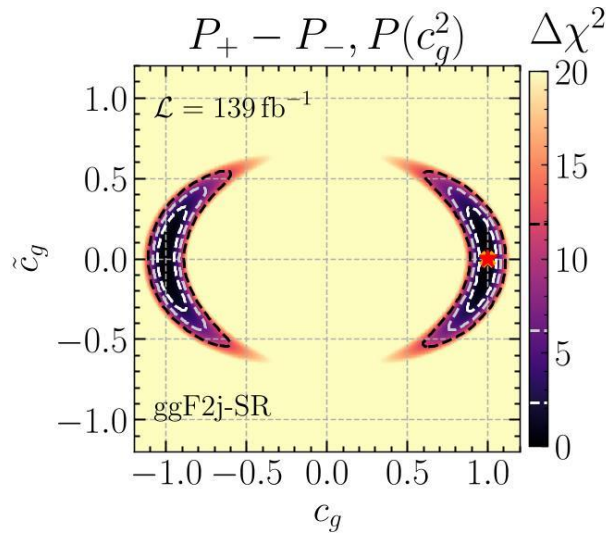


ggF2j signal region

- Ellipse from total rate
- $\Delta\phi_{jj}$ alone is not able to resolve the ellipse
- CP-even classifier can resolve the ellipse well
- CP-odd classifier performs much worse
- Combined: $|\alpha_g| \leq 15^\circ @ 1\sigma$

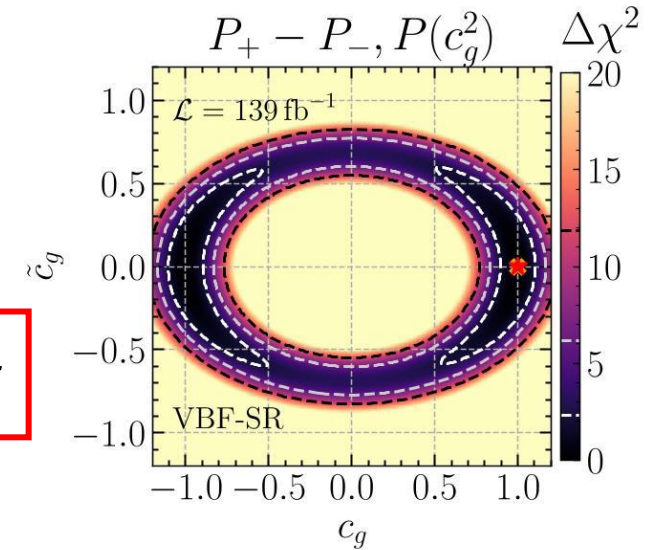


Comparison between signal regions



$$|\tilde{c}_g| \leq 15^\circ @ 1\sigma$$

$$|\tilde{c}_g| \leq 25^\circ @ 1\sigma$$



- Ellipse in VBF-SR much wider due to lower number of events
- Most sensitivity in the VBF-SR comes from interference events (Backup)
- Caveat: No non-Higgs backgrounds taken into account

Part II: ttH

Analysis strategy

- Look for extension of STXS-framework to combine decay channels with focus on CP
- Test most promising CP-even observables, no need to distinguish t & \bar{t}
- Variables are generally not Lorentz-invariant

Observable	Definition	Frame
$p_{T,H}$	--	lab, $t\bar{t}$, $t\bar{t}H$
$\Delta\eta_{t\bar{t}}$	$ \eta_t - \eta_{\bar{t}} $	lab, H , $t\bar{t}H$
$\Delta\phi_{t\bar{t}}$	$ \phi_t - \phi_{\bar{t}} $	lab, H , $t\bar{t}H$
$m_{t\bar{t}}$	$(p_t + p_{\bar{t}})^2$	frame-invariant
$m_{t\bar{t}H}$	$(p_t + p_{\bar{t}} + p_H)^2$	frame-invariant

Observables

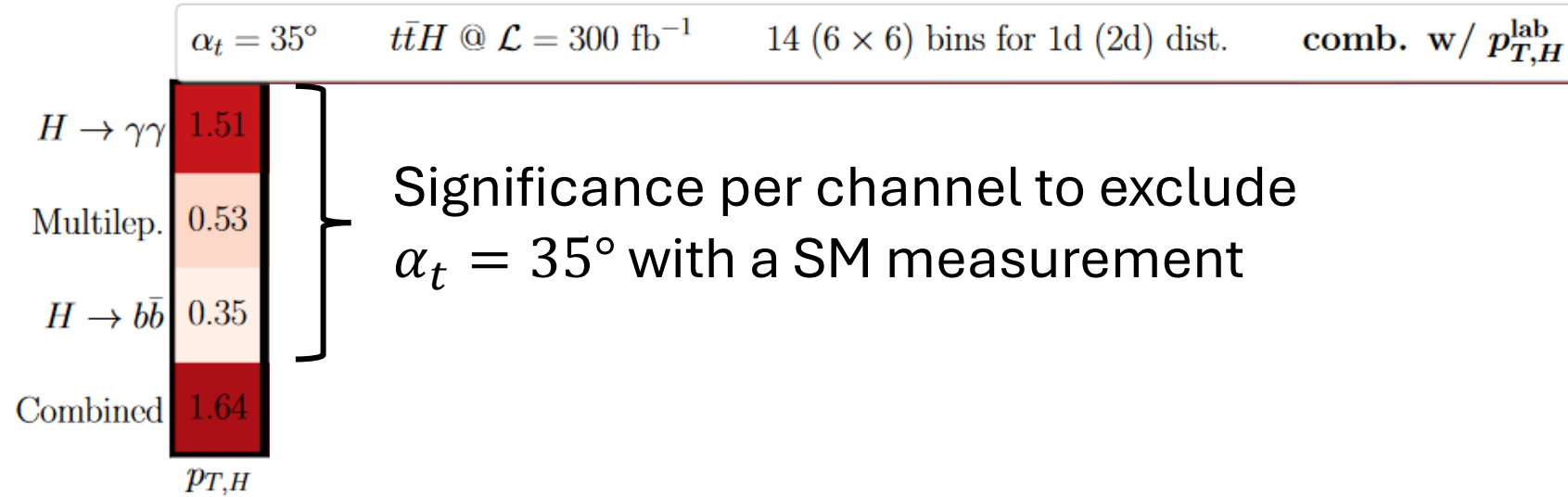
Observable	Definition	Frame
$p_{T,H}$	--	lab, $t\bar{t}$, $t\bar{t}H$
$\Delta\eta_{t\bar{t}}$	$ \eta_t - \eta_{\bar{t}} $	lab, H , $t\bar{t}H$
$\Delta\phi_{t\bar{t}}$	$ \phi_t - \phi_{\bar{t}} $	lab, H , $t\bar{t}H$
$m_{t\bar{t}}$	$(p_t + p_{\bar{t}})^2$	frame-invariant
$m_{t\bar{t}H}$	$(p_t + p_{\bar{t}} + p_H)^2$	frame-invariant
ϕ_C	$\cos^{-1} \left(\frac{ (p_{p,1} \times p_{p,2})(p_t \times p_{\bar{t}}) }{ p_{p,1} \times p_{p,2} p_t \times p_{\bar{t}} } \right)$	H
$\cos \theta^*$	$\frac{\vec{p}_t \vec{n}}{ \vec{p}_t \vec{n} }$	$t\bar{t}$
b_1	$\frac{(\vec{p}_t \times \vec{n})(\vec{p}_{\bar{t}} \times \vec{n})}{p_{T,t} p_{T,\bar{t}}}$	all
b_2	$\frac{(\vec{p}_t \times \vec{n})(\vec{p}_{\bar{t}} \times \vec{n})}{ \vec{p}_t \vec{p}_{\bar{t}} }$	all
b_3	$\frac{p_t^x p_{\bar{t}}^x}{p_{T,t} p_{T,\bar{t}}}$	all
b_4	$\frac{p_t^z p_{\bar{t}}^z}{ \vec{p}_t \vec{p}_{\bar{t}} }$	all

Angular observables

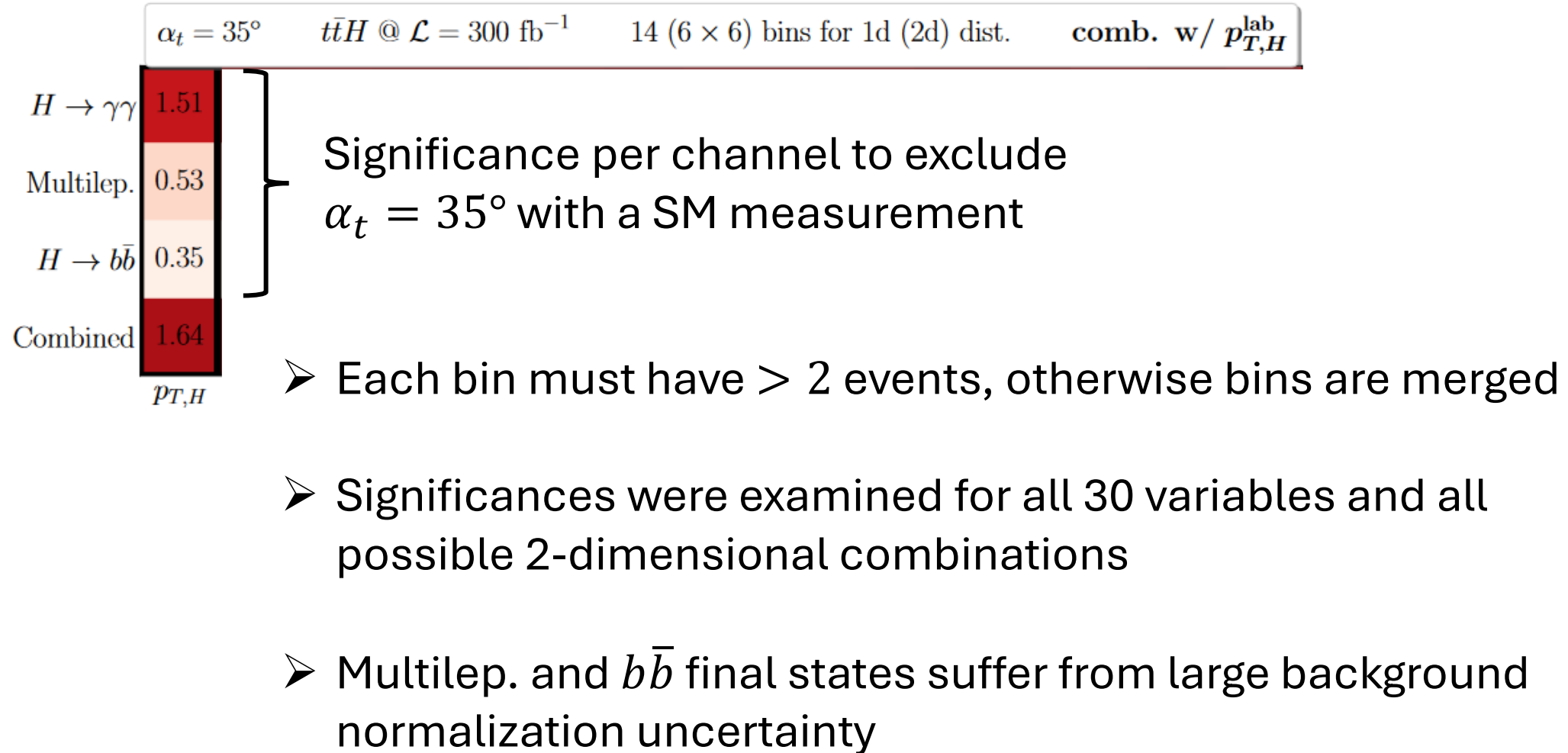
Variables based on top-quark momenta

Collins, Soper '77, Gunion, He '96, Q. Cao et al. '20

Significances



Significances



Significances

	$\alpha_t = 35^\circ$	$t\bar{t}H @ \mathcal{L} = 300 \text{ fb}^{-1}$						14 (6×6) bins for 1d (2d) dist.				comb. w/ $p_{T,H}^{\text{lab}}$			
$H \rightarrow \gamma\gamma$	1.51	1.56	1.54	1.56	1.55	1.52	1.55	1.48	1.5	1.51	1.58	1.59	1.5	1.58	1.51
Multilep.	0.53	0.69	0.9	0.89	0.87	0.73	0.69	0.54	0.52	0.45	0.77	0.8	0.48	0.82	0.56
$H \rightarrow b\bar{b}$	0.35	0.43	0.52	0.52	0.51	0.45	0.44	0.38	0.36	0.29	0.47	0.49	0.3	0.5	0.38
Combined	1.64	1.76	1.86	1.87	1.85	1.75	1.75	1.62	1.63	1.61	1.82	1.84	1.6	1.85	1.65
	$p_{T,H}$	$\Delta\eta_{t\bar{t}}$	$\Delta\phi_{t\bar{t}}$	b_1	b_2	b_3	b_4	$m_{t\bar{t}}$	$m_{t\bar{t}H}$	$p_{T,H}$	$\Delta\eta_{t\bar{t}}$	$ \cos\theta^* $	b_1	b_2	b_3
	lab frame							indep.		$t\bar{t}$ frame					

- Lab and $t\bar{t}$ frame shown due to overall higher significances
- Combinations of variables with $p_{T,H}$ can yield close-to-optimal significances
Comparison with BDT: backup
- Best combined 2d significance: 1.91σ ($\Delta\phi_{t\bar{t}}^{\text{lab}} + b_4^{\text{lab}}$, not shown here)

Significances

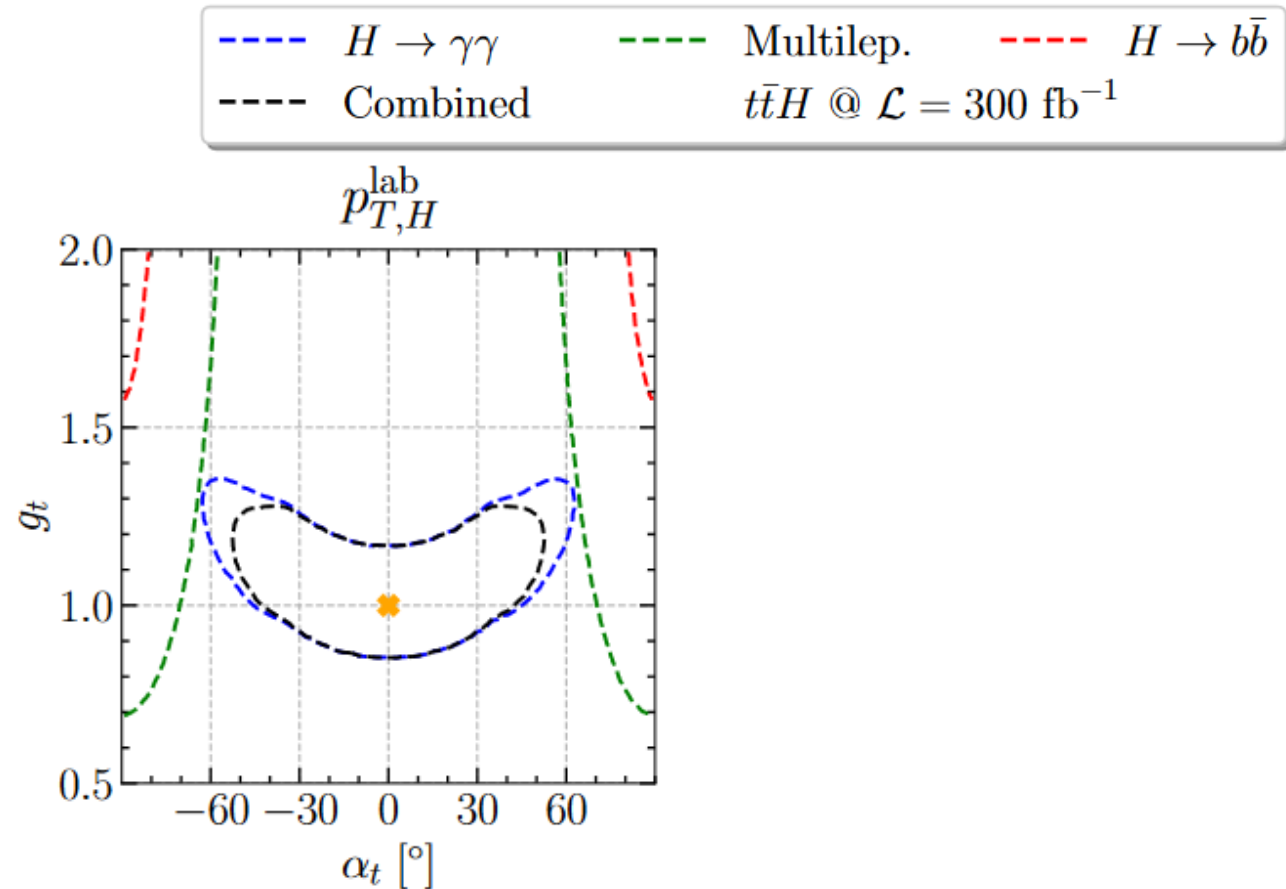
	$\alpha_t = 35^\circ$	$t\bar{t}H @ \mathcal{L} = 300 \text{ fb}^{-1}$													
		14 (6 × 6) bins for 1d (2d) dist.										comb. w/ $p_{T,H}^{\text{lab}}$			
$H \rightarrow \gamma\gamma$	1.51	1.56	1.54	1.56	1.55	1.52	1.55	1.48	1.5	1.51	1.58	1.59	1.5	1.58	1.51
Multilep.	0.53	0.69	0.9	0.89	0.87	0.73	0.69	0.54	0.52	0.45	0.77	0.8	0.48	0.82	0.56
$H \rightarrow b\bar{b}$	0.35	0.43	0.52	0.52	0.51	0.45	0.44	0.38	0.36	0.29	0.47	0.49	0.3	0.5	0.38
Combined	1.64	1.76	1.86	1.87	1.85	1.75	1.75	1.62	1.63	1.61	1.82	1.84	1.6	1.85	1.65
	$p_{T,H}$	$\Delta\eta_{t\bar{t}}$	$\Delta\phi_{t\bar{t}}$	b_1	b_2	b_3	b_4	$m_{t\bar{t}}$	$m_{t\bar{t}H}$	$p_{T,H}$	$\Delta\eta_{t\bar{t}}$	$ \cos\theta^* $	b_1	b_2	b_3
	lab frame							indep.			$t\bar{t}$ frame				

- Lab and $t\bar{t}$ frame shown due to overall higher significances
- Combinations of variables with $p_{T,H}$ can yield close-to-optimal significances
Comparison with BDT: backup
- Best combined 2d significance: 1.91σ ($\Delta\phi_{t\bar{t}}^{\text{lab}} + b_4^{\text{lab}}$, not shown here)

Exclusion limits

➤ Expected exclusion limits taking $p_{T,H}^{\text{lab}}$ only:

➤ $H \rightarrow \gamma\gamma$ dominates the constraints



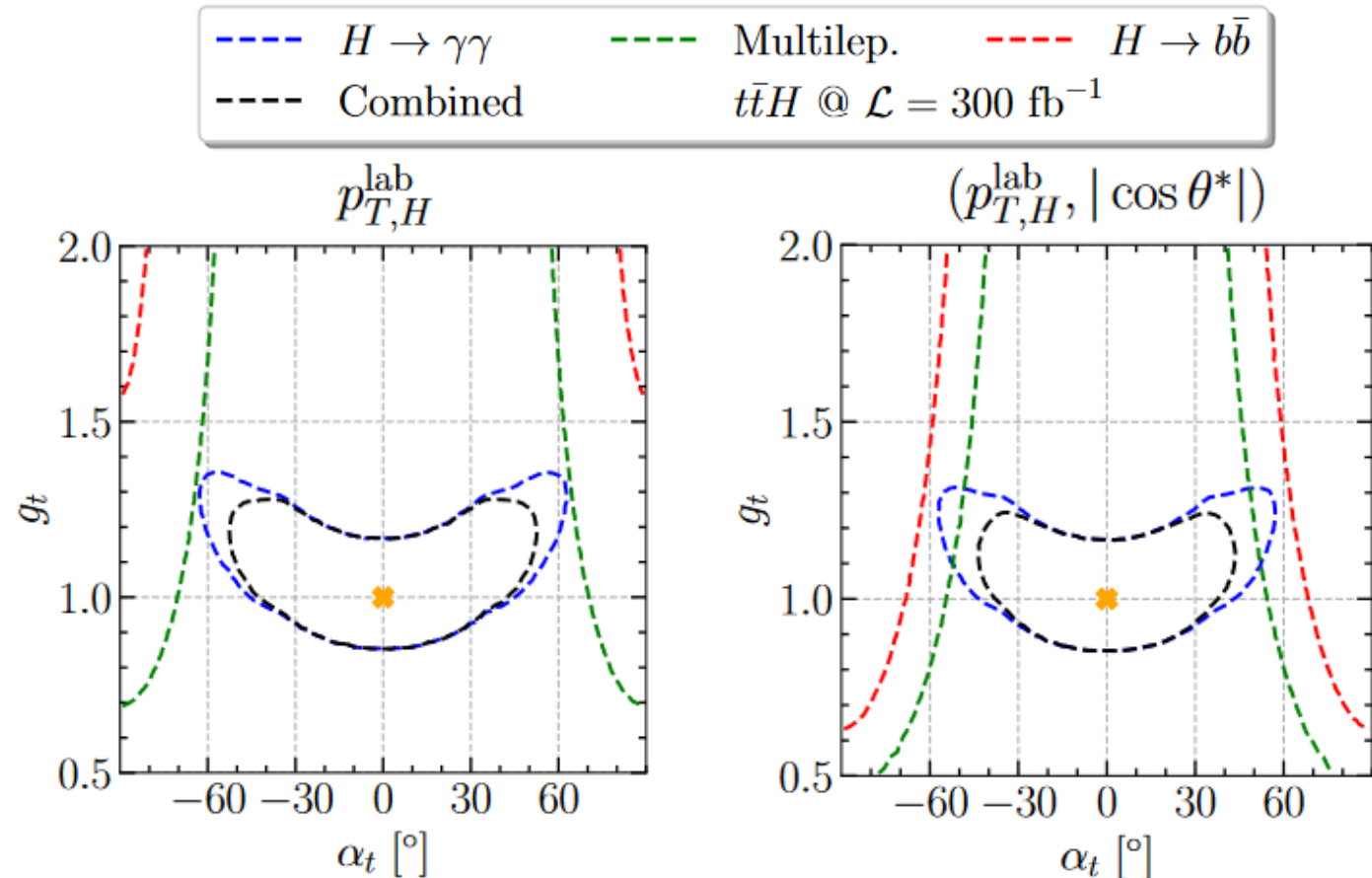
Exclusion limits

- Expected exclusion limits taking $p_{T,H}^{\text{lab}}$ only and taking $p_{T,H}^{\text{lab}} + |\cos \theta^*|$:

- $H \rightarrow \gamma\gamma$ dominates the constraints

- Strong potential improvement from a second variable, here: $\alpha_t \lesssim 52^\circ \rightarrow \alpha_t \lesssim 43^\circ$

- Very similar results for $p_{T,H}^{\text{lab}} + b_2^{\text{lab}} / \Delta\eta_{t\bar{t}}$



Conclusion & Takeaways

[2309.03146]

Part I: ggF2j

- Performed a multivariate CP analysis in a ggF2j-like SR
- Expect much stronger constraints than in VBF-like SR (driven by CP-even observables)
- Improvement at 139fb^{-1} (1σ):
 $\alpha_{g/t} \lesssim 25^\circ \rightarrow \alpha_{g/t} \lesssim 15^\circ$
- Outlook: Investigate behaviour of the irreducible background

Conclusion & Takeaways

[2309.03146]

Part I: ggF2j

- Performed a multivariate CP analysis in a ggF2j-like SR
- Expect much stronger constraints than in VBF-like SR (driven by CP-even observables)
- Improvement at 139fb^{-1} (1σ):
 $\alpha_{g/t} \lesssim 25^\circ \rightarrow \alpha_{g/t} \lesssim 15^\circ$
300fb⁻¹ (2σ) similar
- Outlook: Investigate behaviour of the irreducible background

Part II: ttH

[2406.03950]

- Studied CP violation combining three Higgs decay channels
- b_2^{lab} , $\Delta\eta_{t\bar{t}}$ and $|\cos\theta^*|$ as a second STXS dimension have close to optimal sensitivity towards CP
- Improvement at 300fb^{-1} (2σ):
 $\alpha_t \lesssim 52^\circ \rightarrow \alpha_t \lesssim 43^\circ$
- Outlook: Agree on common top quark definition

Backup

Backup: Other notations

$$\text{Here: } \mathcal{L}_{ggH} = -\frac{1}{4v} \left(-\frac{\alpha_s}{3\pi} c_g G_{\mu\nu}^a G^{\mu\nu,\alpha} + \frac{\alpha_s}{2\pi} \tilde{c}_g G_{\mu\nu}^a \tilde{G}^{\mu\nu,a} \right) H$$

$$\text{Separate top loop: } \mathcal{L}_{ggH} = -\frac{\alpha_s \pi}{v} \left(c_{gg} G_{\mu\nu}^a G^{\mu\nu,\alpha} + \tilde{c}_{gg} G_{\mu\nu}^a \tilde{G}^{\mu\nu,a} \right) H$$

Used in CMS `22

$$\Rightarrow c_g = 1 + 12\pi^2 c_{gg}, \quad \tilde{c}_g = -8\pi^2 \tilde{c}_{gg}$$

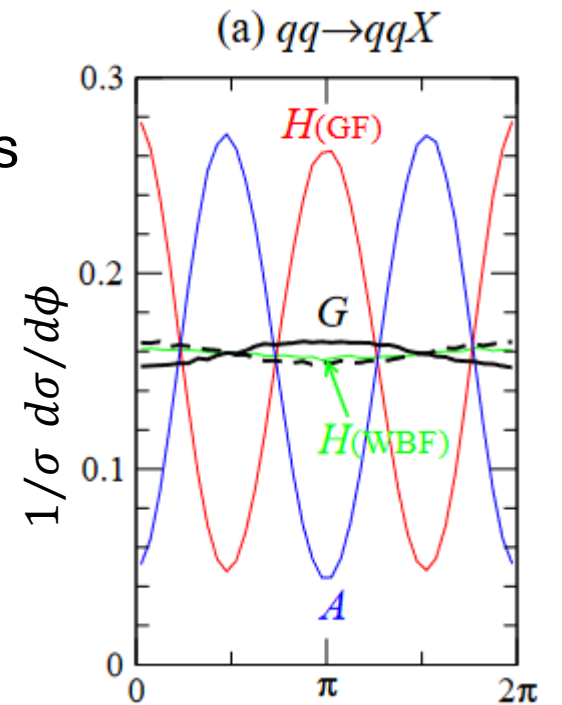
$$\text{SMEFT: } c_{g/t} \sim 1 + \sum_i c_i / \Lambda^2 + \dots, \quad \tilde{c}_{g/t} \sim \sum_i \tilde{c}_i / \Lambda^2 + \dots$$

c_i, \tilde{c}_i : Wilson coefficients

Backup: CP information in ggF2j

$$|\mathcal{M}_{\text{ggF2j}}|^2 = c_g^2 |\mathcal{M}_{\text{even}}|^2 + \underbrace{2c_g \tilde{c}_g \text{Re}[\mathcal{M}_{\text{even}} \mathcal{M}_{\text{odd}}^*]}_{\text{Interference}} + \tilde{c}_g^2 |\mathcal{M}_{\text{odd}}|^2$$

- Current most sensitive study from CMS in VBF-like signal region CMS '22
 - VBF-SR is enriched in ggF2j $q\bar{q}$ initial state and interference events
 - Does the $q\bar{q}$ initial state have most CP-sensitivity?
 - Problems: Initial study conducted with heavy Higgs, in VBF-like signal region, only variable used is $\Delta\phi_{jj}$
- ⇒ ggF2j-SR has tons of events, can we gain sensitivity from them?



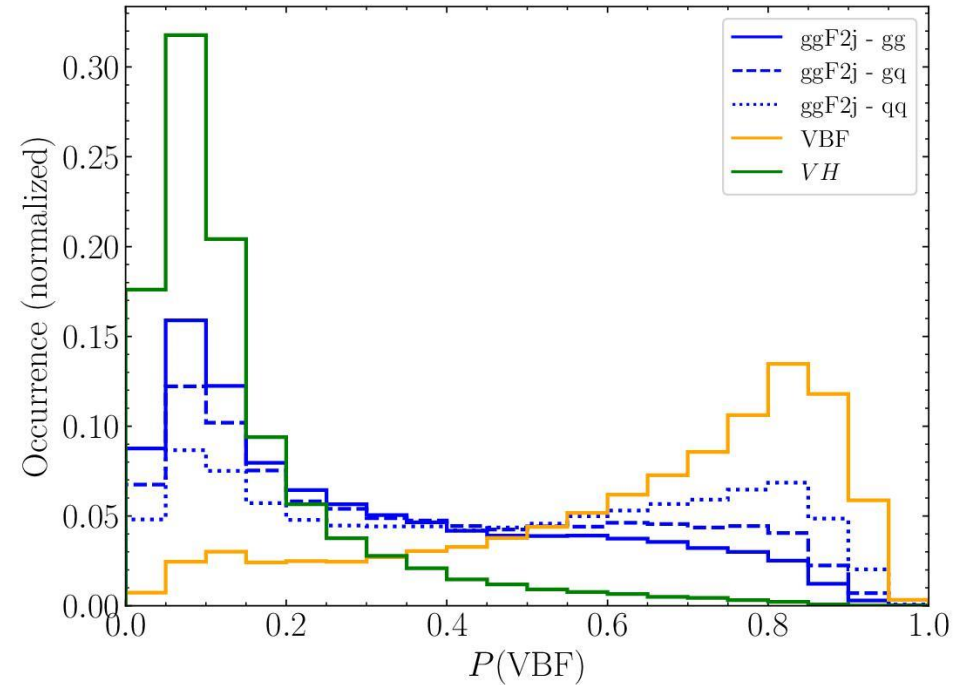
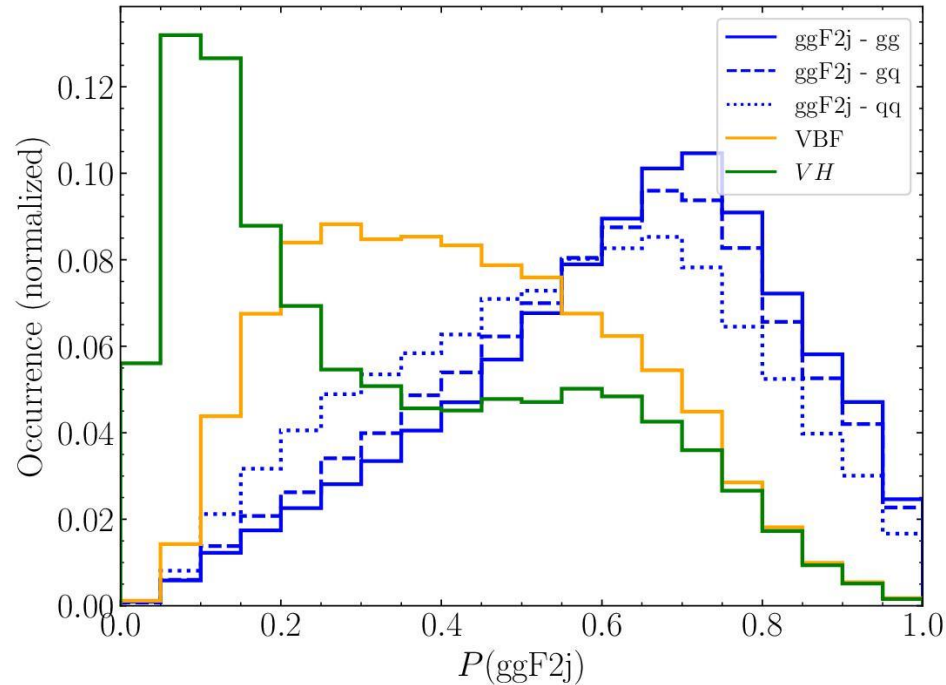
Hagiwara et al. '09 $\Delta\phi_{jj}$

Backup: Cutflow

Applied cut	Fraction of accepted events				
	ggF2j $ \mathcal{M}_{\text{even}} ^2$	ggF2j Interf.	ggF2j $ \mathcal{M}_{\text{odd}} ^2$	VBF	VH
Initial events	100%	100%	100%	100%	100%
$N_j \geq 2, N_\gamma \geq 2$	48.1%	50.8%	48.1%	62.6%	49.8%
$100\text{GeV} \leq m_{\gamma\gamma}$ $m_{\gamma\gamma} \leq 140\text{GeV}$	47.8%	50.5%	47.9%	62.0%	49.4%
$p_T^{\gamma_1} / m_{\gamma\gamma} \geq 0.35$ $p_T^{\gamma_2} / m_{\gamma\gamma} \geq 0.25$	39.4%	40.9%	39.8%	50.0%	40.5%
$p_T^{j_1} \geq 30\text{GeV}$ $p_T^{j_2} \geq 20\text{GeV}$	38.6%	40.2%	38.6%	49.7%	39.9%
$ \eta_j \leq 2.5$ $ \eta_\gamma \leq 2.5$	22.9%	21.5%	22.7%	39.8%	31.2%
$p_T^H \leq 200\text{GeV}^*$	18.6%	18.4%	18.3%	34.4%	26.8%

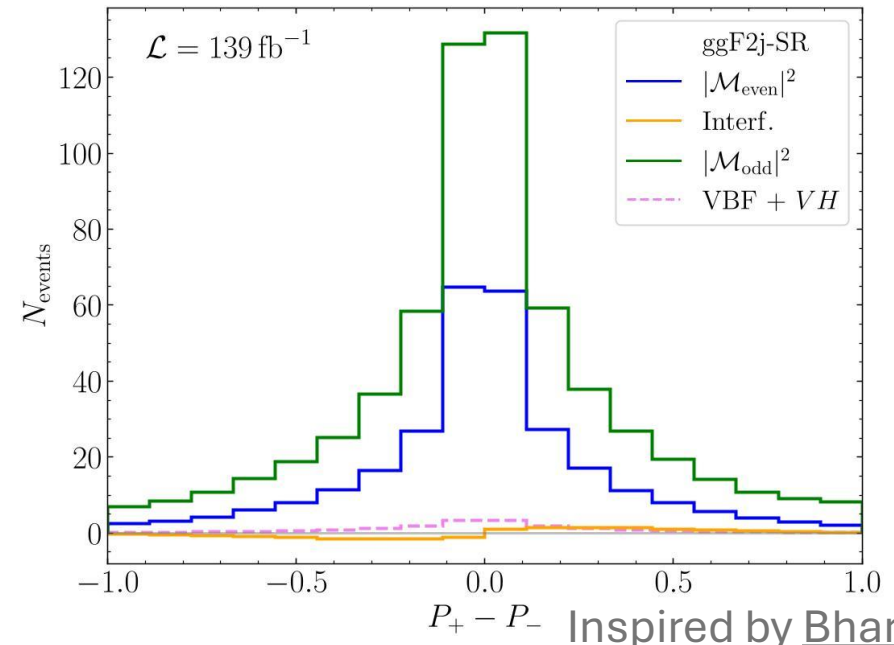
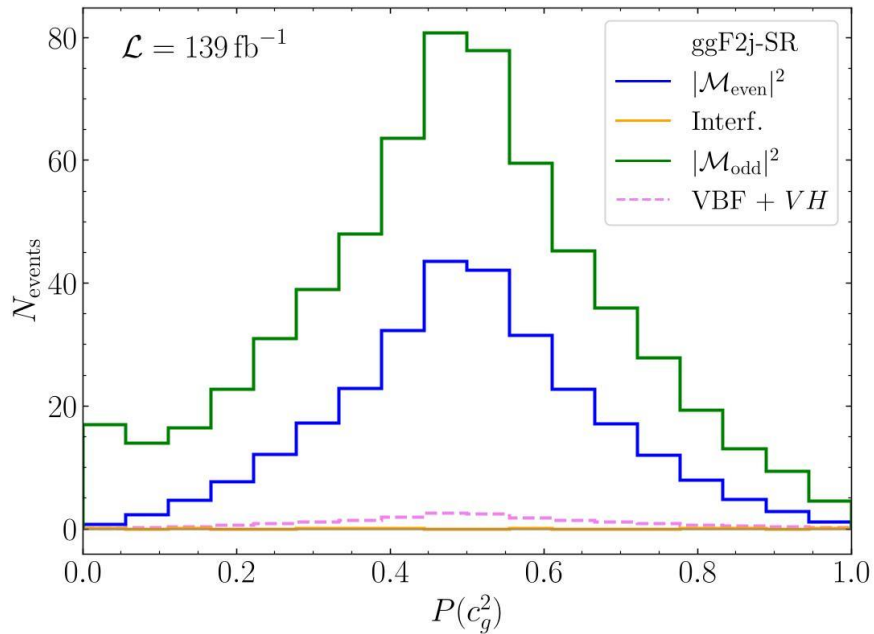
* It is possible to relax this cut using FT_{approx} → see [Maltoni et al. '14](#)

Backup: Signal regions



- Use $P(\text{signal}) > 0.5$ as a cut to define signal regions
- ggF2j with $q\bar{q}$ initial state are identified as VBF-like more often
- ggF2j interference events are also identified as VBF-like more often

Backup: ggF2j signal region

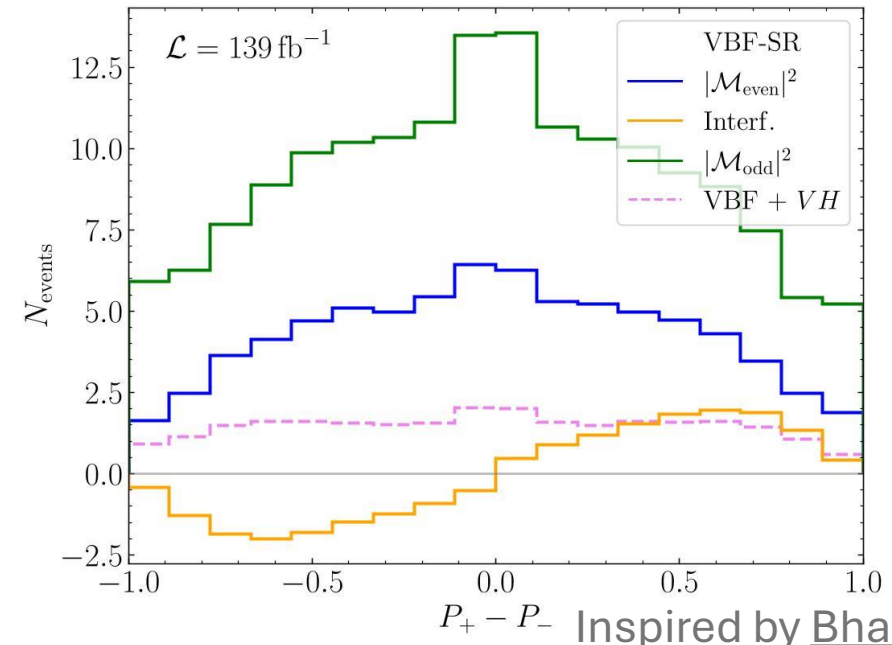
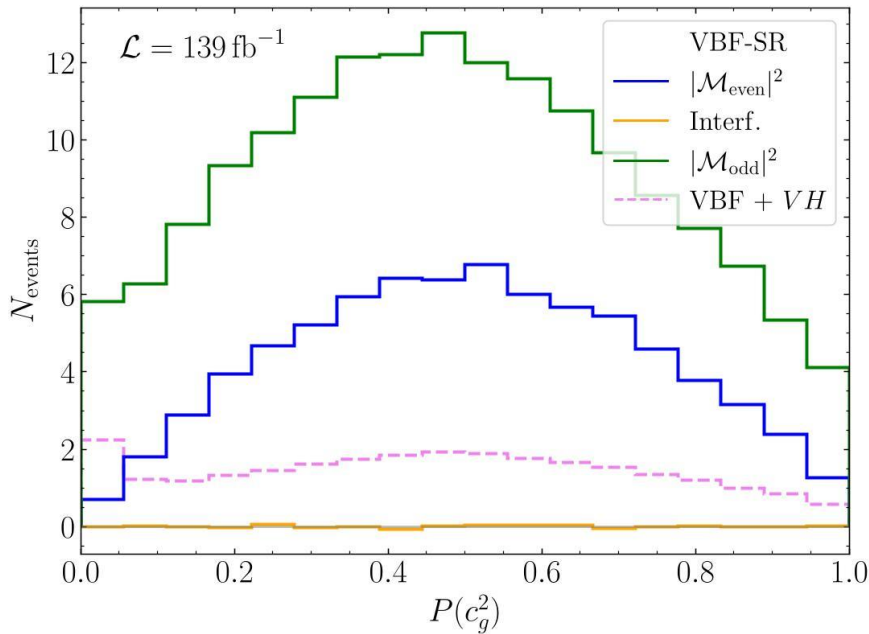


Inspired by Bhardwaj et al. `21

- $P(c_g^2)$ differentiates between $c_g^2 |\mathcal{M}_{\text{even}}|^2$ and $\tilde{c}_g^2 |\mathcal{M}_{\text{odd}}|^2$
- Kinematically very similar, but some separation in outer bins
- Interference term cancels out

- $P_+ - P_-$ differentiates between **positive and negative interference**
- Interference barely visible due to low cross section & looks more VBF-like
- CP-even terms are symmetric

Backup: VBF signal region



Inspired by Bhardwaj et al. `21

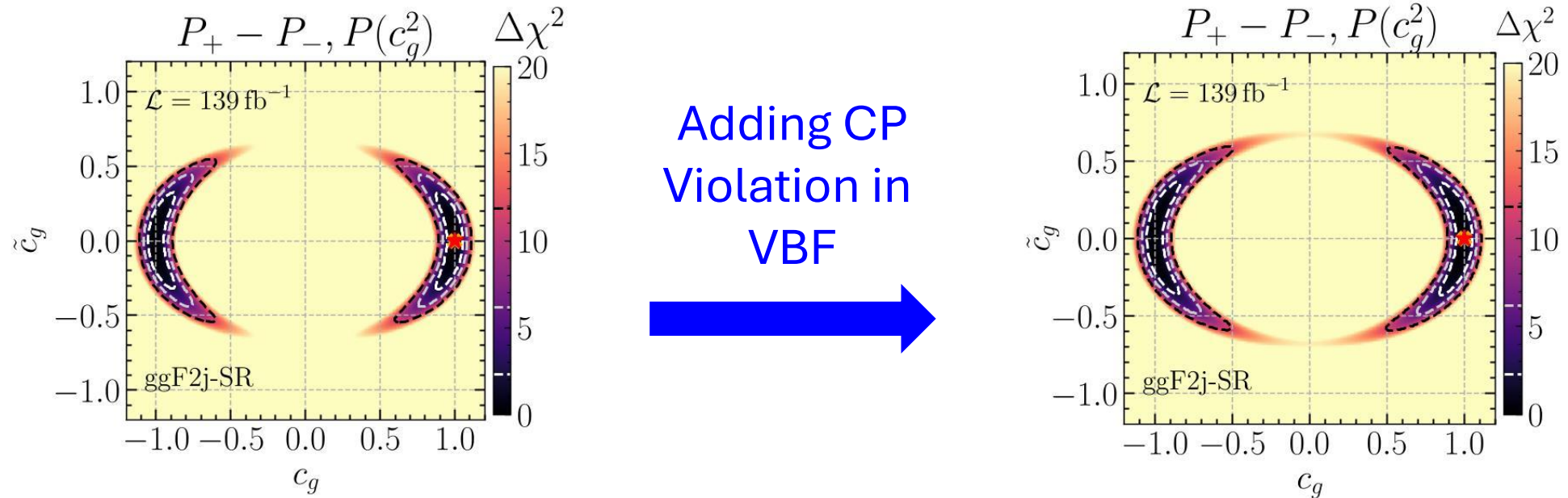
- $P(c_g^2)$ differentiates between $c_g^2 |\mathcal{M}_{\text{even}}|^2$ and $\tilde{c}_g^2 |\mathcal{M}_{\text{odd}}|^2$
- Less events than in ggF2j-SR, statistical fluctuations visible
- Wider peak around $P(c_g^2) = 0.5$

- $P_+ - P_-$ differentiates between **positive and negative interference**
- Interference much more pronounced due to overall lower events and their VBF-like kinematic

Backup: CP violation in HVV – ggF2j-SR

BSM physics might introduce CP violation in multiple Higgs couplings
⇒ inject CP violation in VBF production, evaluate effect on c_g, \tilde{c}_g limits

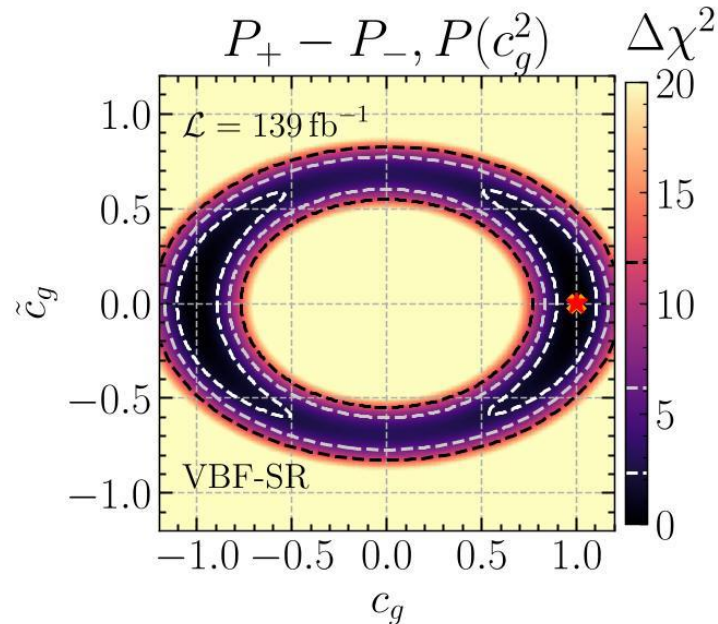
- Limits slightly weaken from $|\tilde{c}_g| \leq 0.32 @ 1\sigma$ to $|\tilde{c}_g| \leq 0.37 @ 1\sigma$
- Mainly a washout effect between SRs due to wrong BG assumption



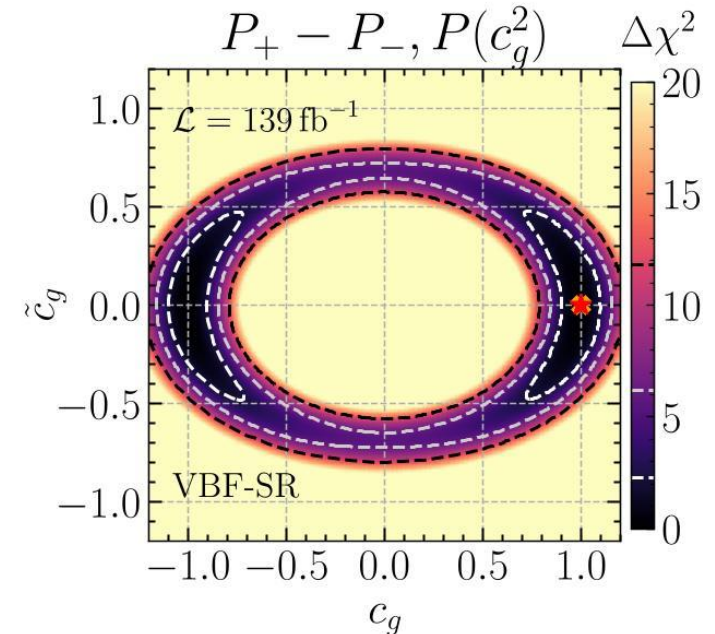
Backup: CP violation in HVV – VBF-SR

BSM physics might introduce CP violation in multiple Higgs couplings
⇒ inject CP violation in VBF production, evaluate effect on c_g, \tilde{c}_g limits

- Limits tighten from $|\tilde{c}_g| \leq 0.58 @ 1\sigma$ to $|\tilde{c}_g| \leq 0.48 @ 1\sigma$
- VBF-SR might “fake” stronger limits not originating from CP violation in ggH



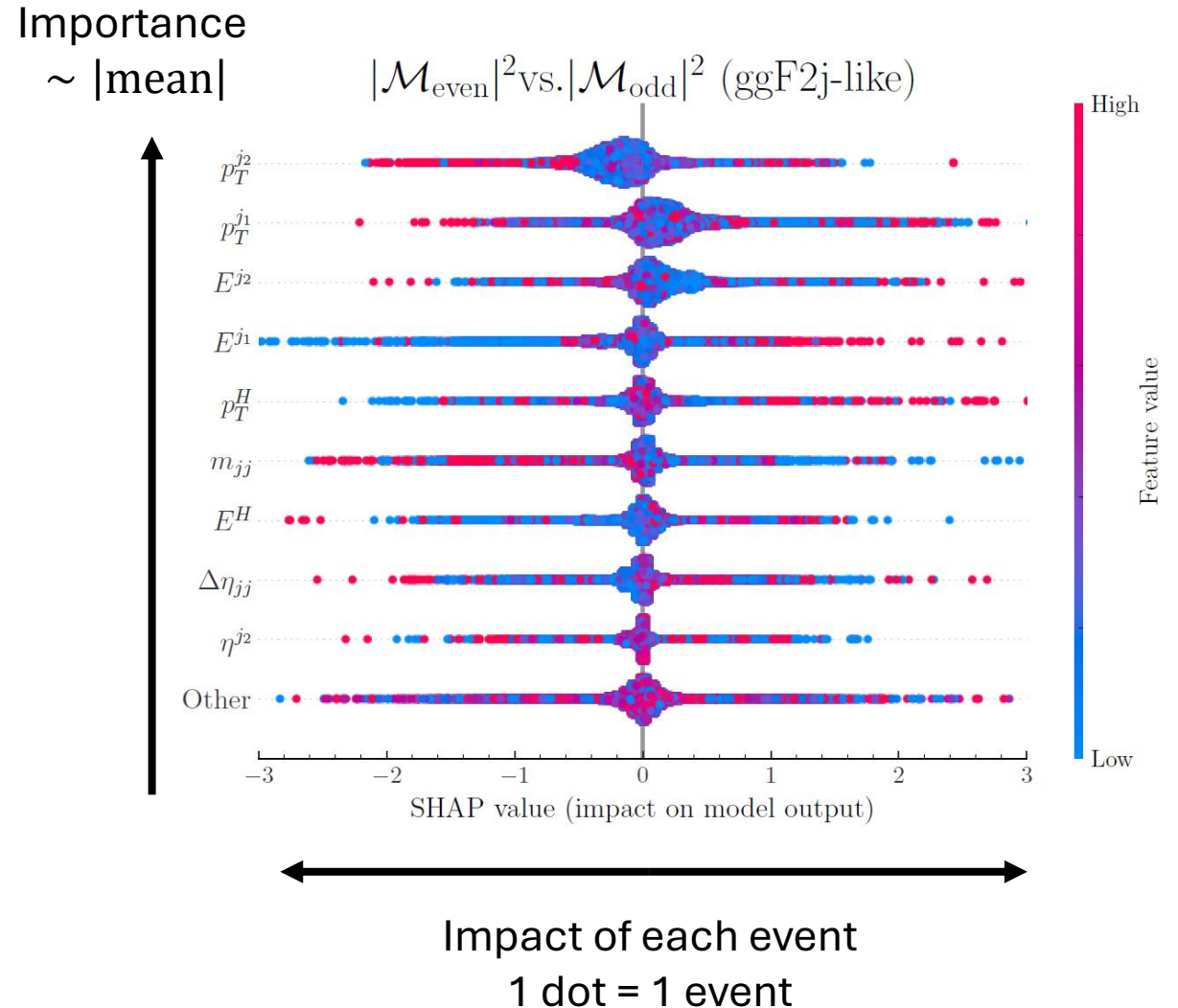
Adding CP
Violation in
VBF



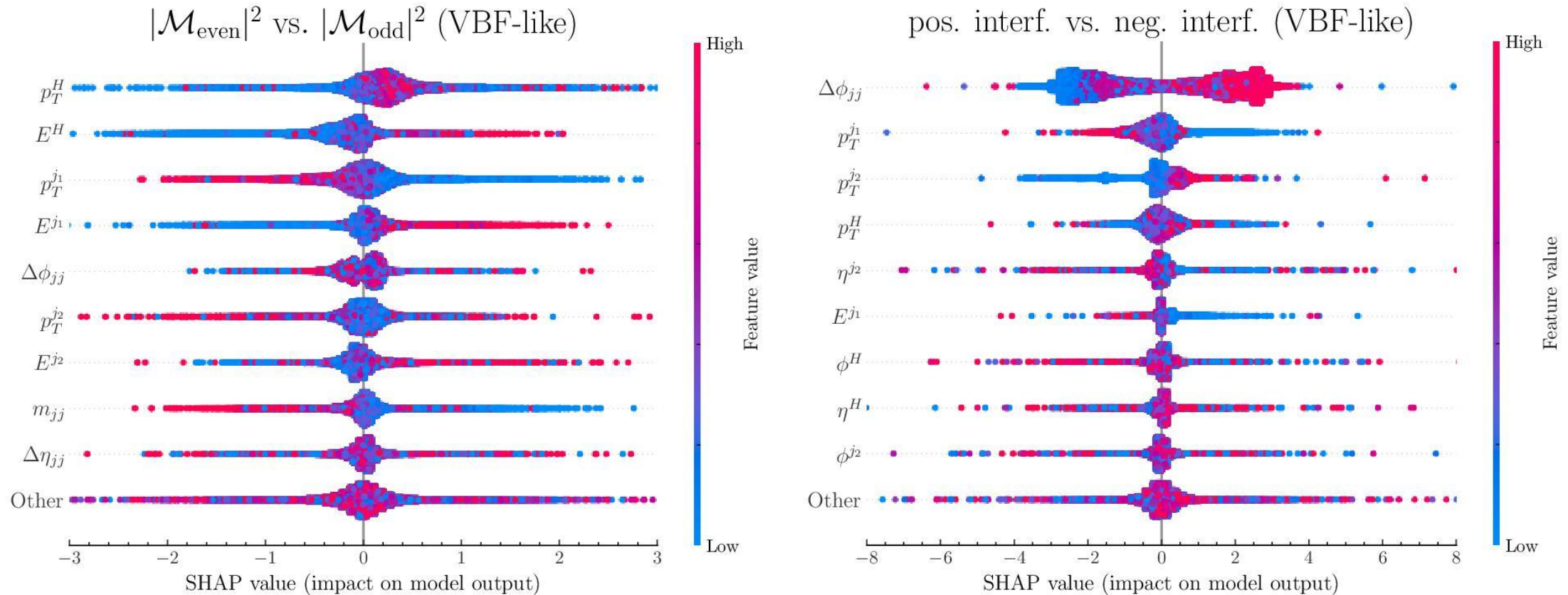
Backup: Feature importance – ggF2j-SR

CP-even classifier:

- $p_T^{j_1}$ and $p_T^{j_2}$ seem to be the most important variables
- Variable importance much less pronounced than in CP-odd case
- Hard to judge interplay
⇒ symbolic regression (WIP)

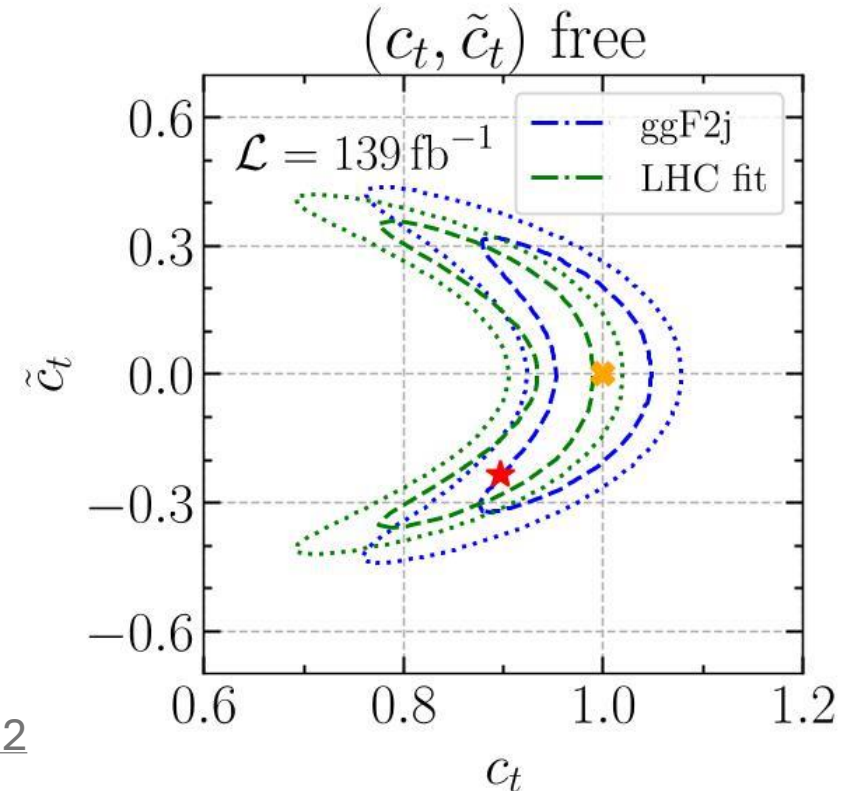


Backup: Feature importance – VBF-SR



Backup: Comparison to global fit

- Interpret c_g, \tilde{c}_g as top-Yukawa coupling modifiers c_t, \tilde{c}_t (heavy top limit / strong limits on colored BSM particles)
- Global fit based on experimental results for Higgs signal rates
- Performed with HiggsTools [Bahl et al. `22](#)
- Different form of ellipsis due to recent $ttH(bb)$ measurements [ATLAS `19](#), [ATLAS `22](#)
- Similar constraints from our analysis



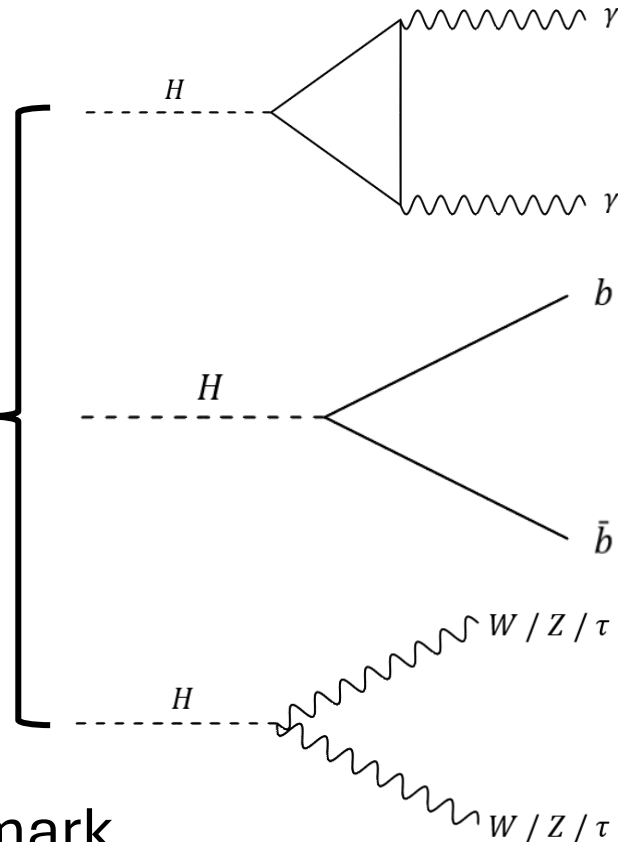
Backup: ttH event generation

Generate samples at LO
→ NLO via flat k-factor

$\alpha_t = 0^\circ$

$\alpha_t = 35^\circ$

$\alpha_t = 90^\circ$



Add (channel-specific):

- Acceptance factors for event selection
- Smearing of kinematics from limited resolution
- Uncertainties

Scale to $\alpha_t = 35^\circ$ as future benchmark

$$N(\alpha_t) = N_{0^\circ} \cos^2 \alpha_t + N_{90^\circ} \sin^2 \alpha_t$$

ATLAS '12, ATLAS '14, ATLAS '19, ATLAS '21
ATLAS '20, CMS '23, ATLAS '24

Backup: ttH significance evaluation

- Significance of excluding a BSM-hypothesis evaluated for all single variables and all 2-dimensional combinations at 300fb^{-1}
- Each bin is required to have > 2 events, otherwise bins are merged
- $H \rightarrow b\bar{b}$, $H \rightarrow \text{multilep}$. use no rate information to account for large uncertainty in the non-Higgs background normalization
- Additional uncertainties from statistics and shape:

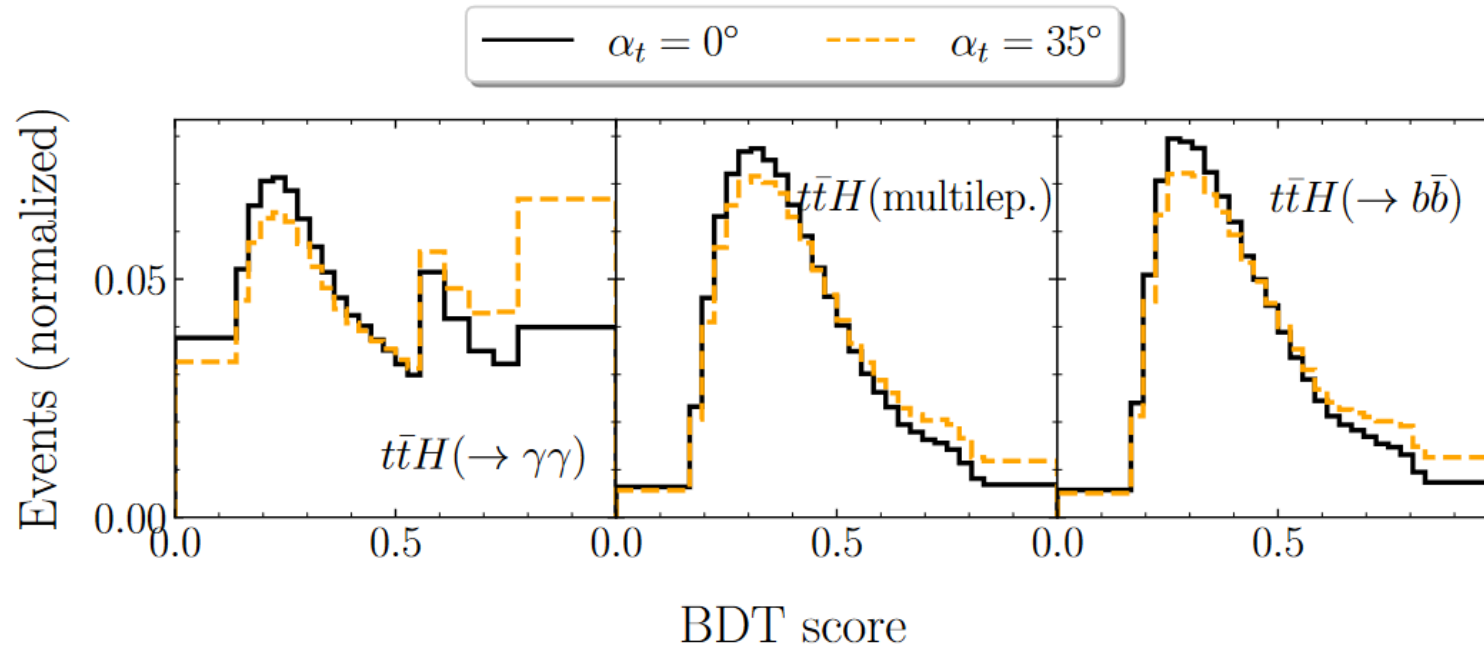
$$\sigma_i = \sqrt{\sigma_{\text{sys}}^2 + \sigma_{\text{stat}}^2} \quad \text{ATLAS '20, CMS '23, ATLAS '24}$$

- Total significance evaluation from SM yield n_i and BSM yield m_i

$$S = \sqrt{2 \sum_i^{N_{\text{bins}}} \left(n_i \ln \left[\frac{m_i(n_i + \sigma_i^2)}{n_i^2 + m_i \sigma_i^2} \right] - \frac{n_i^2}{\sigma_i^2} \ln \left[1 + \frac{\sigma_i^2(m_i - n_i)}{n_i(n_i + \sigma_i^2)} \right] \right)}$$

ATLAS '20

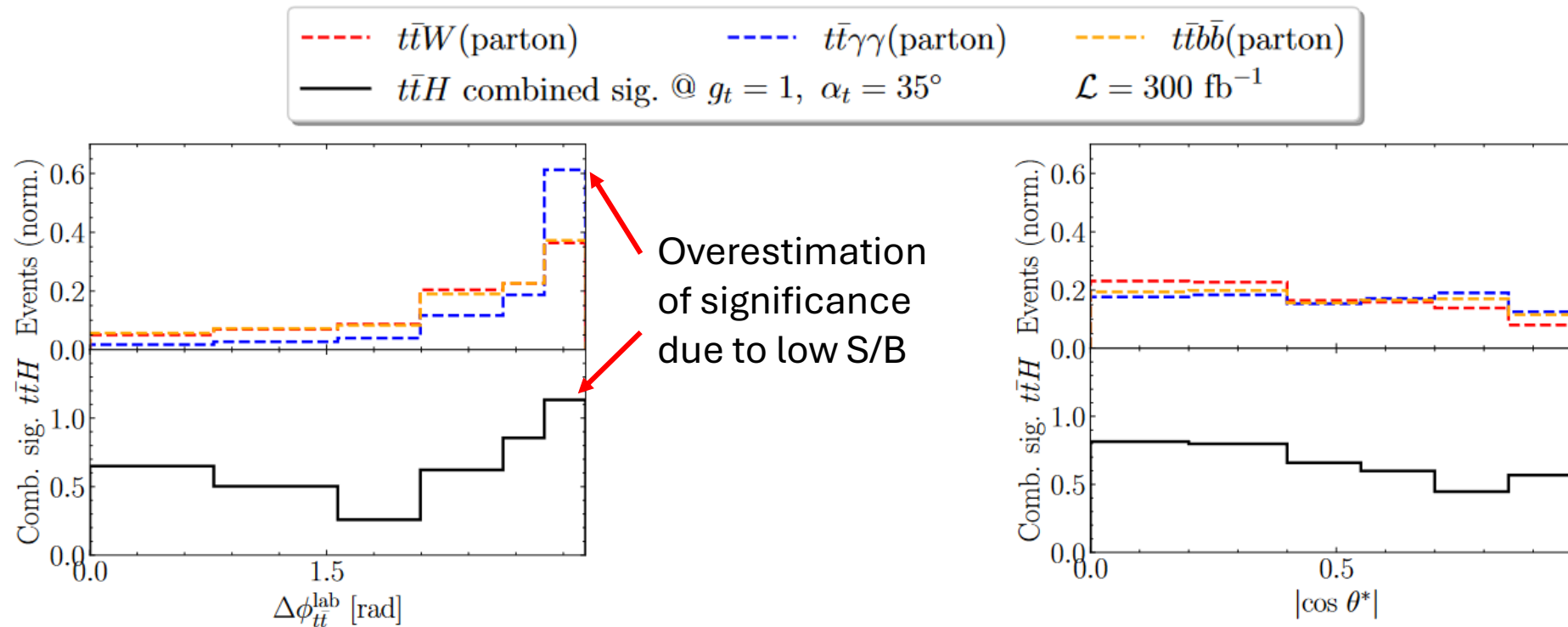
Backup: Comparison with BDT approach



Channel	$t\bar{t}H(\rightarrow \gamma\gamma)$	$t\bar{t}H(\text{multilep.})$	$t\bar{t}H(\rightarrow b\bar{b})$	Combined
Significance (2D)	1.57	0.94	0.55	1.91
Significance (BDT)	1.75	1.17	0.69	2.21

Backup: $t\bar{t}H$ background shape

- The shape of the background has been neglected so far by considering S/B to be equal in each bin but it could spoil the significances



We consider $\Delta\phi_{t\bar{t}}^{\text{lab}}$ and b_1^{lab} to be disfavoured by the background

Our final shortlist of variables includes b_2^{lab} , $\Delta\eta_{t\bar{t}}^{t\bar{t}}$ and $|\cos\theta^*|$