Classifying the CP properties of the Higgs–gluon interaction

and the quest for interpretable ML

Henning Bahl

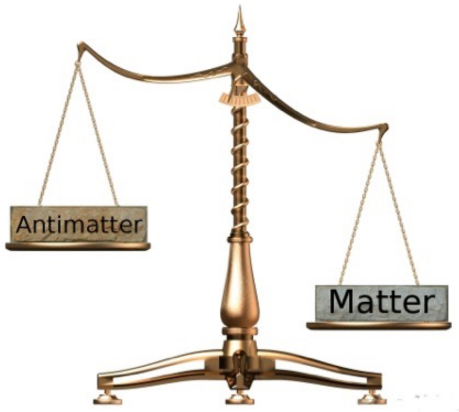
based on 2309.03146

in collaboration with E. Fuchs, M. Hannig, and M. Menen

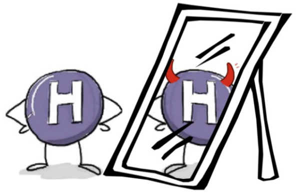
UNIVERSITÄT HEIDELBERG
ZUKUNFT SEIT 1386

ML4Jets, DESY Hamburg, 8.11.23

CP violation in the Higgs sector



- New sources of CP violation are necessary to explain the baryon asymmetry of the Universe.
- One possibility: CP violation in the Higgs sector.

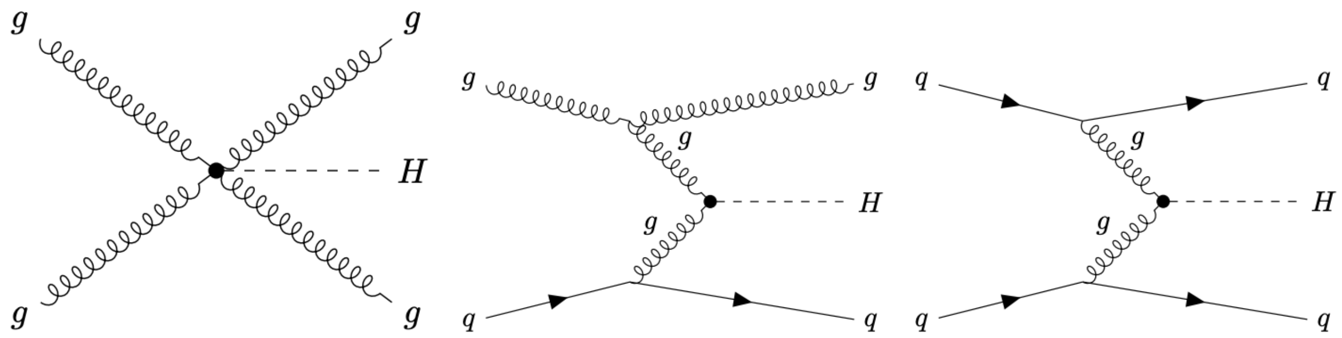


Is the Higgs boson a CP-admixed state?

- Why use ggF2j production for CP tests? [Hankele, Klamke, Zeppenfeld `06,`07, …]
  - Gluon fusion is the largest Higgs production channel → wealth of data.
  - Two additional jets in the final state allow to construct CP-odd observables.

  → CP sensitivity beyond total rate information.

ggF2j production



(a) gg-initiated (b) gq-initiated (c) qq-initiated

- Effective Lagrangian (after integrating out the top quark, SM: $c_g = 1$, $\tilde{c}_g = 0$):

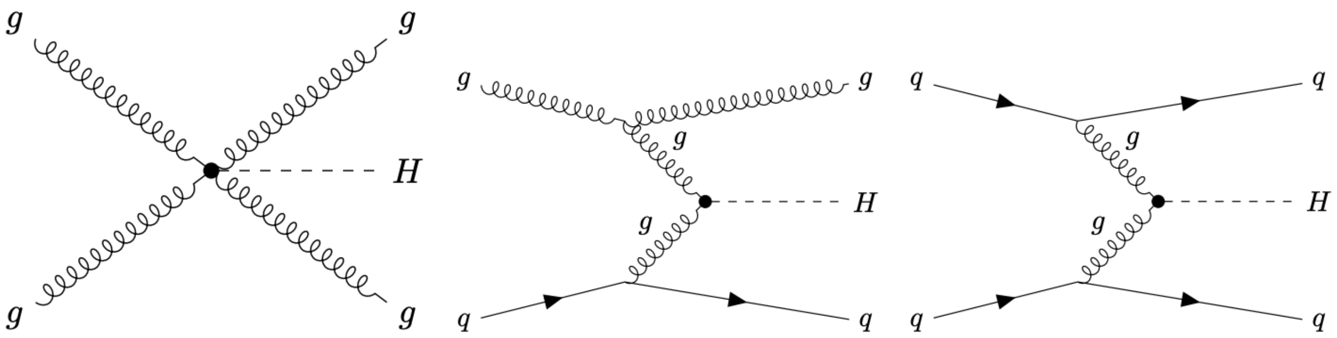
$$\mathcal{L}_{Hgg} = -\frac{1}{4v} H \left(-\frac{\alpha_s}{3\pi} c_g G^a_{\mu\nu} G^{a,\mu\nu} + \frac{\alpha_s}{2\pi} \tilde{c}_g G^a_{\mu\nu} \tilde{G}^{a,\mu\nu} \right)$$

(heavy top limit enforced by p_T cut)

- Amplitude splits up into three pieces:

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

ggF2j production



(a) gg-initiated (b) gq-initiated (c) qq-initiated

- Effective Lagrangian (after integrating out the top quark, SM: $c_g = 1$, $\tilde{c}_g = 0$):

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

(heavy top limit enforced by p_T cut)

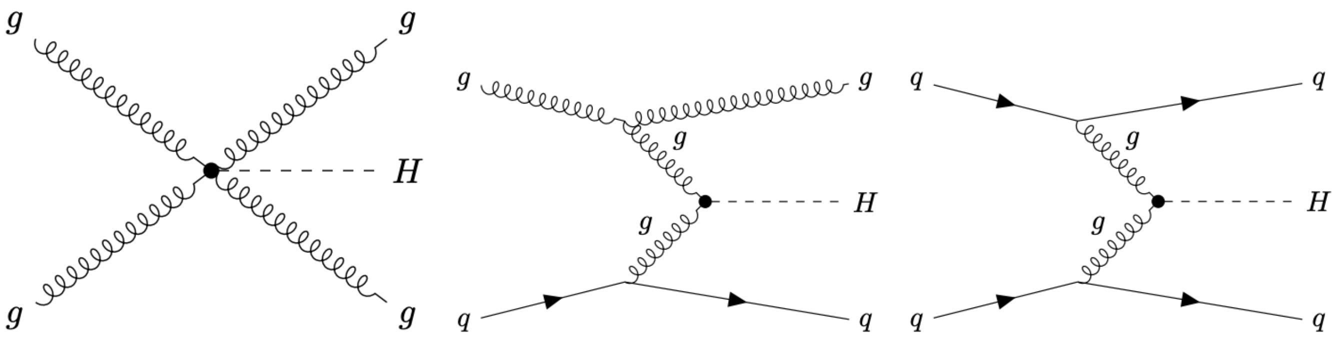
- Amplitude splits up into three pieces:

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

Assumption in the literature: [e.g., CMS `21, `22; ATLAS `21, `22]

- CP sensitivity highest for vector-boson-fusion (VBF) like kinematics, or
- azimuthal angle between the two jets $\Delta\phi_{jj}$ is the optimal observable.

ggF2j production



(a) gg-initiated (b) gq-initiated (c) qq-initiated

- Effective Lagrangian (after integrating out the top quark, SM: $c_g = 1, \tilde{c}_g = 0$):

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

(heavy top limit enforced by p_T cut)

- Amplitude splits up into three pieces:

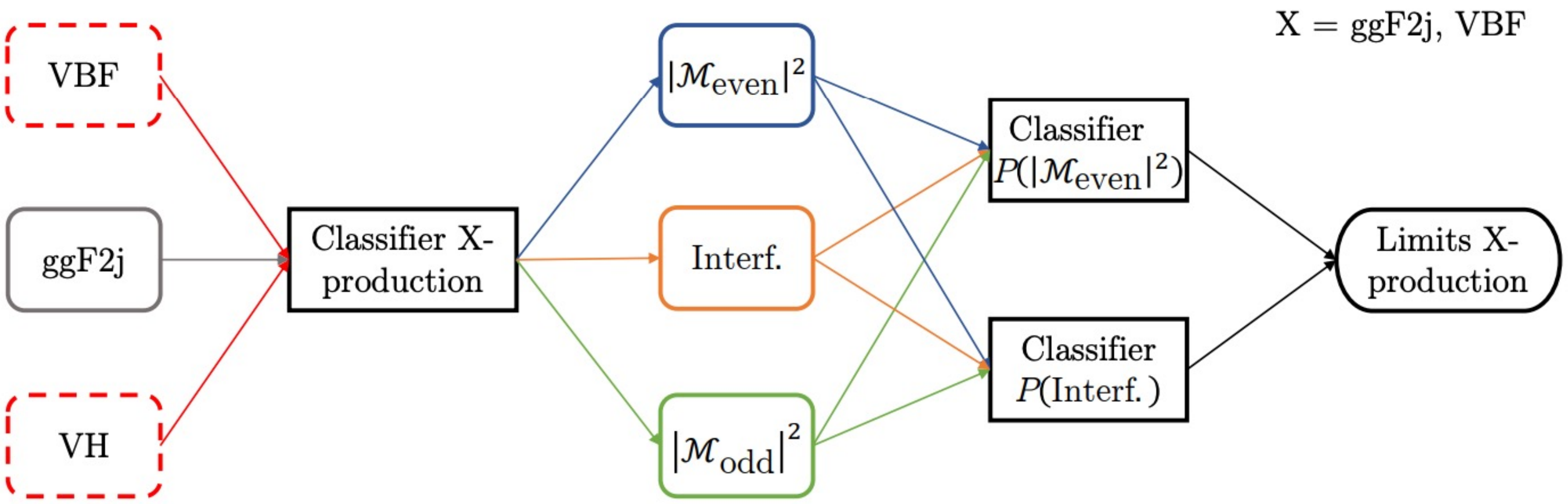
$$\left|\mathcal{M}_{\text{ggF2j}}\right|^2 = c_g^2 |\mathcal{M}_{\text{even}}|^2 + \underbrace{2 c_g \tilde{c}_g Re[\mathcal{M}_{\text{even}} \mathcal{M}_{\text{odd}}{}^*]}_{\text{interference}} + \tilde{c}_g^2 |\mathcal{M}_{\text{odd}}|^2$$

Assumption in the literature: [e.g., CMS `21, `22; ATLAS `21, `22]

- CP sensitivity highest for vector-boson-fusion (VBF) like kinematics, or
- azimuthal angle between the two jets $\Delta\phi_{jj}$ is the optimal observable.

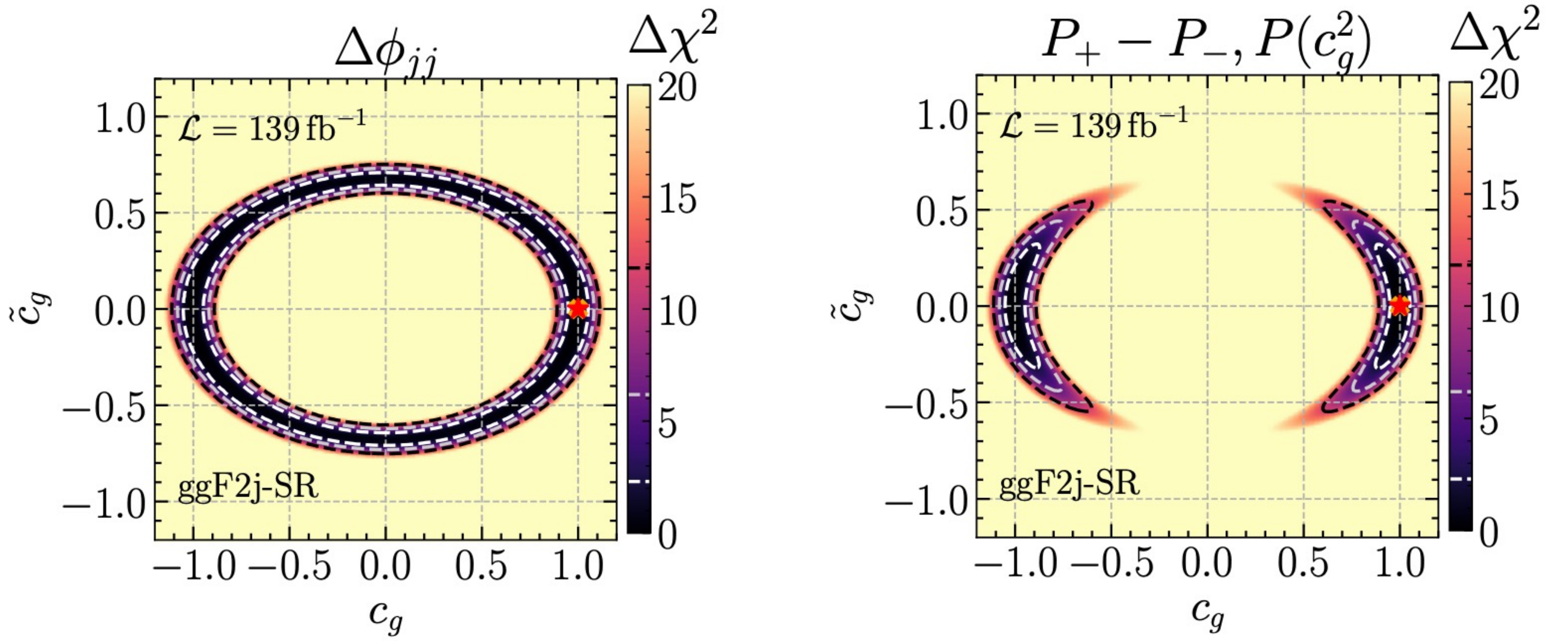
Can we do better?

Analysis flow



- Focus on $H \to \gamma\gamma$ decay channel.
- Two signal regions: **ggF2j-SR**, VBF-SR
- For each signal region: train signal–background classifier.
- Then, train two classifiers to distinguish $|\mathcal{M}_{\text{even}}|^2$ vs. $|\mathcal{M}_{\text{odd}}|^2$ and (positive intf.) vs (negative intf).
- Build two observables: CP-even $P(c_g^2)$ and CP-odd $P_+ - P_-$.

ggF2j signal region



- ggF2j signal region outperforms VBF signal region (not shown),
- $\Delta\phi_{jj}$ limit is significantly worse.

Which observables drive these constraints? → interpretable ML?!

Analog problem in game theory

Who is the highest-value player in a cooperative game?

Analog problem in game theory

Who is the highest-value player in a cooperative game?

- Example 1: Leibniz and Newton independently invented calculus which has the value of 100.

$$val(\{\text{Leibniz}\}) = 100,$$
$$val(\{\text{Newton}\}) = 100,$$
$$val(\{\text{Leibniz}, \text{Newton}\}) = 100$$

[shapleyvalue.com]

Analog problem in game theory

Who is the highest-value player in a cooperative game?

- Example 1: Leibniz and Newton independently invented calculus which has the value of 100.

$$val(\{\text{Leibniz}\}) = 100,$$
$$val(\{\text{Newton}\}) = 100,$$
$$val(\{\text{Leibniz}, \text{Newton}\}) = 100$$

Leibniz and Newton have same value: $\phi_{\text{Leibniz}} = \phi_{\text{Newton}} = 50$

[shapleyvalue.com]

Analog problem in game theory

Who is the highest-value player in a cooperative game?

- Example 1: Leibniz and Newton independently invented calculus which has the value of 100.

$$\left.\begin{aligned} val(\{\text{Leibniz}\}) &= 100, \\ val(\{\text{Newton}\}) &= 100, \\ val(\{\text{Leibniz}, \text{Newton}\}) &= 100 \end{aligned}\right\}$$

Leibniz and Newton have same value: $\phi_{\text{Leibniz}} = \phi_{\text{Newton}} = 50$

- Example 2: Netwon invented calculus. Leibniz, mad with envy, pretends that he also invented calculus.

$$\left.\begin{aligned} val(\{\text{Leibniz}\}) &= 0, \\ val(\{\text{Newton}\}) &= 100, \\ val(\{\text{Leibniz}, \text{Newton}\}) &= 100 \end{aligned}\right\}$$

[shapleyvalue.com]

Analog problem in game theory

Who is the highest-value player in a cooperative game?

- Example 1: Leibniz and Newton independently invented calculus which has the value of 100.

$$val(\{\text{Leibniz}\}) = 100, \quad val(\{\text{Newton}\}) = 100, \quad val(\{\text{Leibniz}, \text{Newton}\}) = 100$$

Leibniz and Newton have same value: $\phi_{\text{Leibniz}} = \phi_{\text{Newton}} = 50$

- Example 2: Netwon invented calculus. Leibniz, mad with envy, pretends that he also invented calculus.

$$val(\{\text{Leibniz}\}) = 0, \quad val(\{\text{Newton}\}) = 100, \quad val(\{\text{Leibniz}, \text{Newton}\}) = 100$$

Leibniz has no value: $\phi_{\text{Leibniz}} = 0$, $\phi_{\text{Newton}} = 100$

[shapleyvalue.com]

Analog problem in game theory

Who is the highest-value player in a cooperative game?

- Example 1: Leibniz and Newton independently invented calculus which has the value of 100.

$$\left.\begin{aligned} &val(\{\text{Leibniz}\}) = 100, \\ &val(\{\text{Newton}\}) = 100, \\ &val(\{\text{Leibniz}, \text{Newton}\}) = 100 \end{aligned}\right\} \text{ Leibniz and Newton have same value: } \phi_{\text{Leibniz}} = \phi_{\text{Newton}} = 50$$

- Example 2: Netwon invented calculus. Leibniz, mad with envy, pretends that he also invented calculus.

$$\left.\begin{aligned} &val(\{\text{Leibniz}\}) = 0, \\ &val(\{\text{Newton}\}) = 100, \\ &val(\{\text{Leibniz}, \text{Newton}\}) = 100 \end{aligned}\right\} \text{ Leibniz has no value: } \phi_{\text{Leibniz}} = 0,\ \phi_{\text{Newton}} = 100$$

[shapleyvalue.com]

⇒ Can we formalize this for more complex situations?

Shapley values

[Shapley, `51]

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

Shapley values

[Shapley, `51]

$$\phi_j(val) = \sum_{S \subseteq \{1,\ldots,p\} \setminus \{j\}} \frac{|S|!\,(p - |S| - 1)!}{p!} (val(S \cup \{j\}) - val(S))$$

Sum over all sets without player *j*

Shapley values

[Shapley, `51]

$$\phi_j(val) = \sum_{S \subseteq \{1,\dots,p\} \setminus \{j\}} \frac{|S|!\,(p - |S| - 1)!}{p!} (val(S \cup \{j\}) - val(S))$$

Sum over all sets without player j

Combinatorial factor

Shapley values

[Shapley, `51]

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

- Sum over all sets without player j
- Combinatorial factor
- Value of player j for given set

Shapley values

[Shapley, `51]

$$\phi_j(val) = \sum_{S \subseteq \{1,\dots,p\} \setminus \{j\}} \frac{|S|!\,(p-|S|-1)!}{p!} \left(val(S \cup \{j\}) - val(S)\right)$$

Sum over all sets without player j

Combinatorial factor

Value of player j for given set

Defining properties:

Shapley values

[Shapley, `51]

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

Sum over all sets without player j

Combinatorial factor

Value of player j for given set

Defining properties:

- Efficiency: $\sum_{j=1}^{p} \phi_j = val(\text{all players})$

Shapley values

[Shapley, `51]

$$\phi_j(val) = \sum_{S \subseteq \{1,\ldots,p\} \setminus \{j\}} \frac{|S|!\,(p - |S| - 1)!}{p!} \big(val(S \cup \{j\}) - val(S)\big)$$

- $\sum_{S \subseteq \{1,\ldots,p\} \setminus \{j\}}$: Sum over all sets without player j
- $\frac{|S|!\,(p - |S| - 1)!}{p!}$: Combinatorial factor
- $\big(val(S \cup \{j\}) - val(S)\big)$: Value of player j for given set

Defining properties:

- Efficiency: $\sum_{j=1}^{p} \phi_j = val(\text{all players})$
- Symmetry: $val(S \cup \{j\}) = val(S \cup \{k\})$ for all $S \subseteq \{1, \ldots, p\} \setminus \{j, k\} \Rightarrow \phi_j = \phi_k$

Shapley values

[Shapley, `51]

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

- Sum over all sets without player j
- Combinatorial factor
- Value of player j for given set

Defining properties:

- Efficiency: $\sum_{j=1}^{p} \phi_j = val(\text{all players})$
- Symmetry: $val(S \cup \{j\}) = val(S \cup \{k\})$ for all $S \subseteq \{1, \ldots, p\} \setminus \{j, k\} \Rightarrow \phi_j = \phi_k$
- Additivity: Shapley values for two games add up $\phi_j = \phi_j^{(1)} + \phi_j^{(2)}$

Shapley values

[Shapley, `51]

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

Sum over all sets without player j

Combinatorial factor

Value of player j for given set

Defining properties:

- Efficiency: $\sum_{j=1}^{p} \phi_j = val(\text{all players})$
- Symmetry: $val(S \cup \{j\}) = val(S \cup \{k\})$ for all $S \subseteq \{1, \ldots, p\} \setminus \{j, k\} \Rightarrow \phi_j = \phi_k$
- Additivity: Shapley values for two games add up $\phi_j = \phi_j^{(1)} + \phi_j^{(2)}$
- Dummy player: $val(S \cup \{j\}) = val(S)$ for all $S \subseteq \{1, \ldots, p\} \setminus \{j\} \Rightarrow \phi_j = 0$

How do we apply this to our physics problem?

Spelling out the analogy:

How do we apply this to our physics problem?

Spelling out the analogy:

- players ↔ physics observables

How do we apply this to our physics problem?

Spelling out the analogy:

- players ↔ physics observables
- value of player set ↔ separation achieved by classifier

How do we apply this to our physics problem?

Spelling out the analogy:

- players ↔ physics observables
- value of player set ↔ separation achieved by classifier

How do we compute the classifier score if certain observables are absent?

How do we apply this to our physics problem?

Spelling out the analogy:

- players ↔ physics observables
- value of player set ↔ separation achieved by classifier

How do we compute the classifier score if certain observables are absent?

- Retraining the classifier for every set is too expensive.

How do we apply this to our physics problem?

Spelling out the analogy:

- players ↔ physics observables
- value of player set ↔ separation achieved by classifier

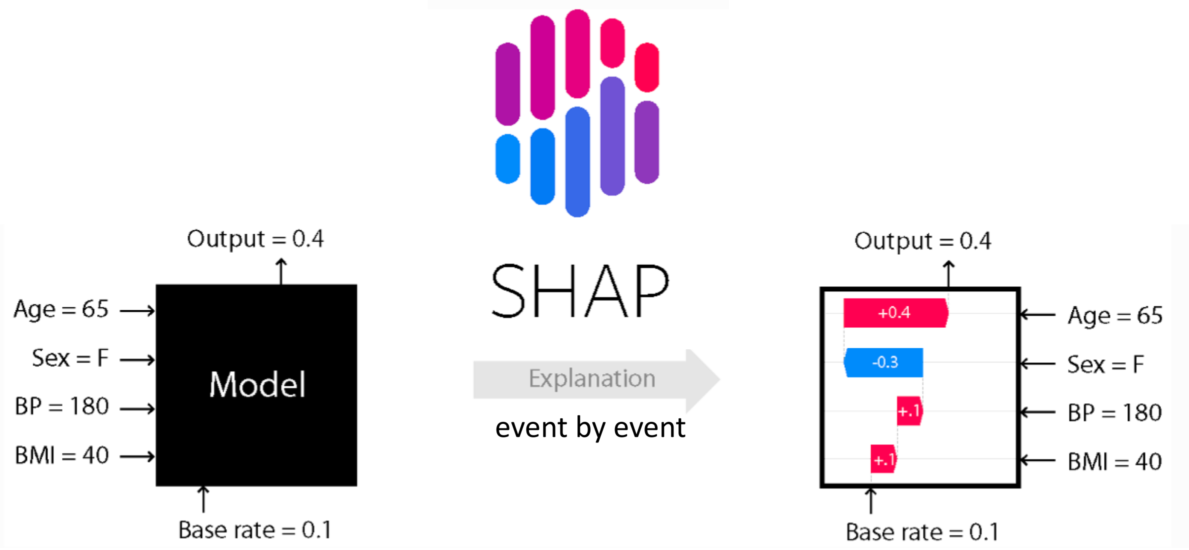
How do we compute the classifier score if certain observables are absent?

- Retraining the classifier for every set is too expensive.
- Use SHAP (SHapley Additive exPlanations) method instead: [Lundberg et al., 2020, github.com/shap]
    - calculate “local” event by event Shapley values,
    - “feature value is absent” ↔ “feature value is replaced by random feature value from data”,
    - sample dataset with larger weights for observables sets with almost no or almost all observables.

How do we apply this to our physics problem?

Spelling out the analogy:

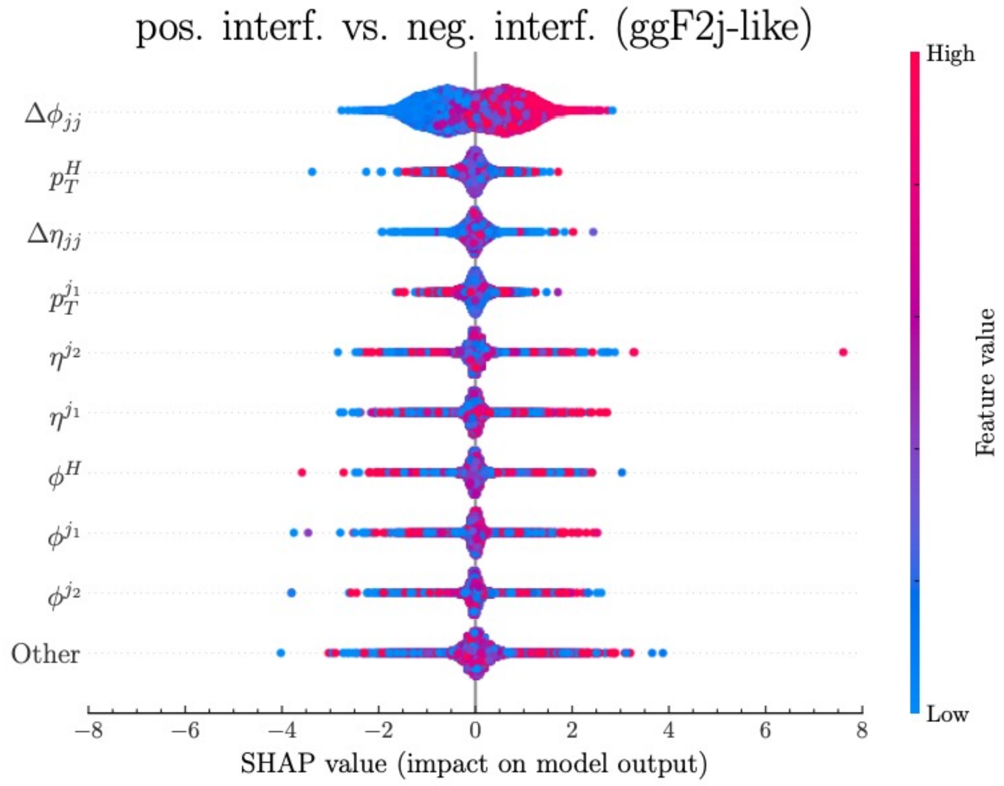
- players ↔ physics observables
- value of player set ↔ separation achieved by classifier



How do we compute the classifier score if certain observables are absent?

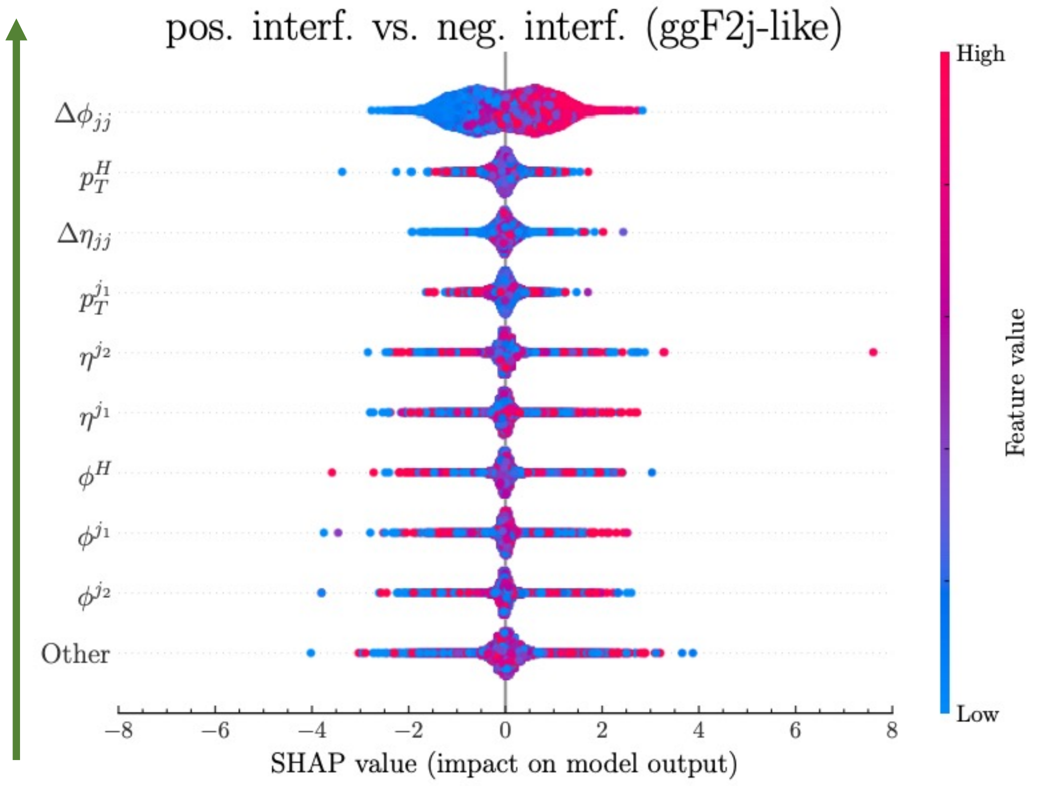
- Retraining the classifier for every set is too expensive.
- Use SHAP (SHapley Additive exPlanations) method instead: [Lundberg et al., 2020, github.com/shap]
  - calculate “local” event by event Shapley values,
  - “feature value is absent” ↔ “feature value is replaced by random feature value from data”,
  - sample dataset with larger weights for observables sets with almost no or almost all observables.

Results for interference classifiers



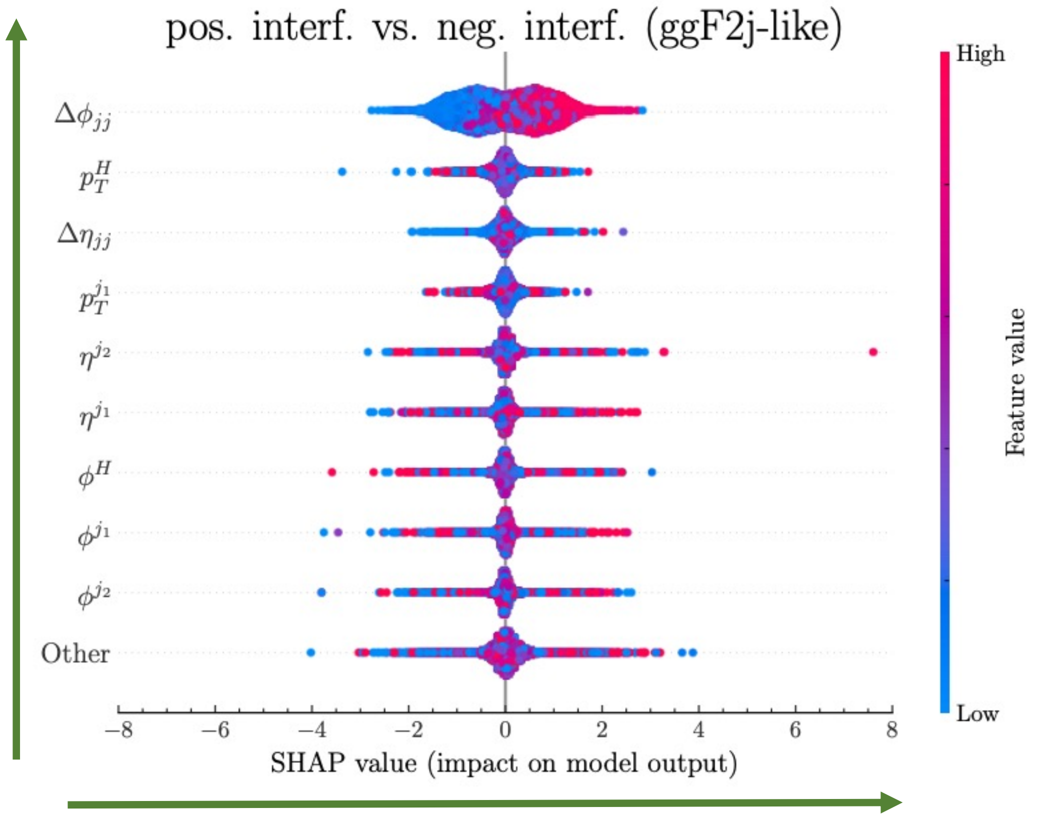
Results for interference classifiers

Importance of observable
($\sim \sum_{\text{events}} |\phi_j|$)



Results for interference classifiers

Importance of observable
$(\sim \sum_{\text{events}} |\phi_j|)$

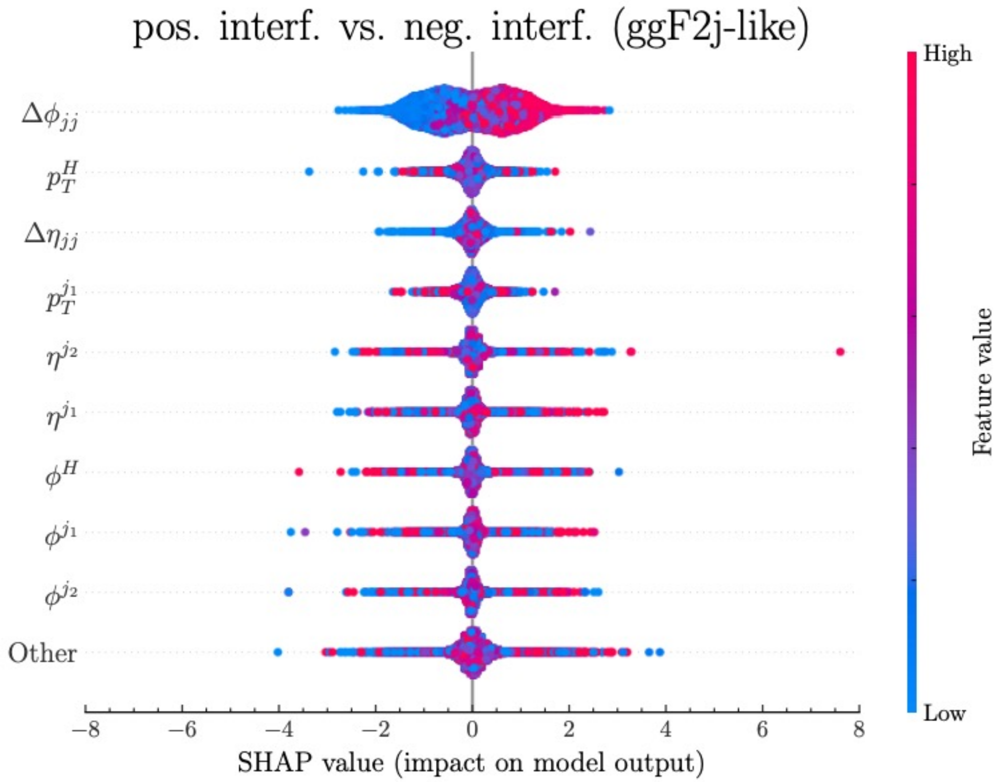


Shapley value for each event (shown as dot)

Results for interference classifiers

Importance of observable
($\sim \sum_{\text{events}} |\phi_j|$)

Value of observable

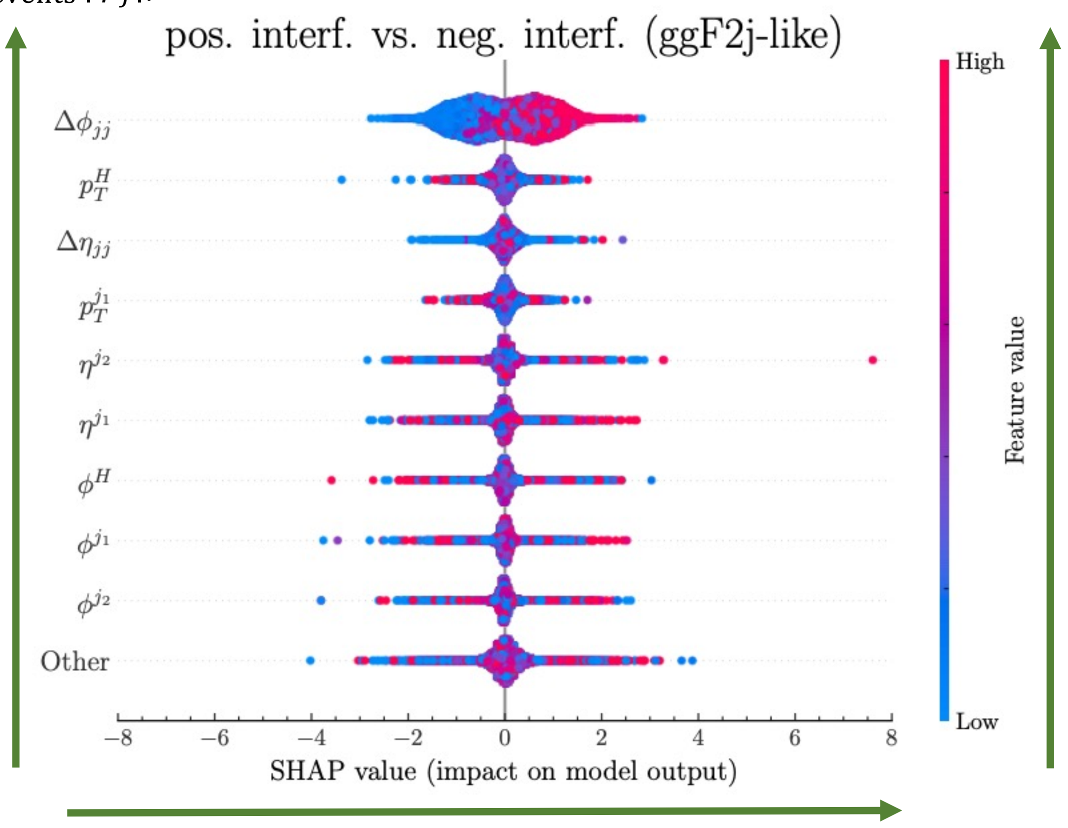


Shapley value for each event (shown as dot)

Results for interference classifiers

Importance of observable
($\sim \sum_{\text{events}} |\phi_j|$)

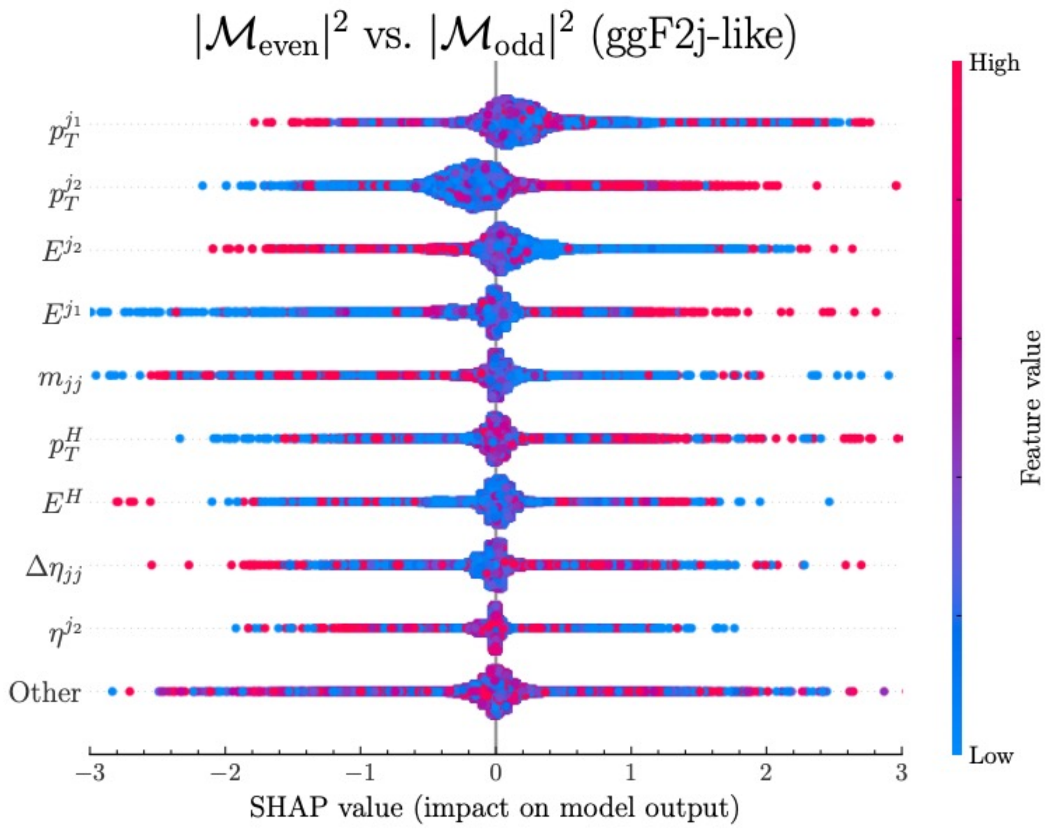
Value of observable



Shapley value for each event (shown as dot)

For the interference classifiers, as expected, the CP-odd $\Delta\phi_{jj}$ is most important.

Results for squared term classifiers



- p_T of jets/Higgs most important, $\Delta\phi_{jj}$ plays only subleading role.
- disadvantage: interplay between observables hard to judge.

Finding analytic expressions using symbolic regression

[work in progress, HB, Menen, Fuchs, Plehn]

Finding analytic expressions using symbolic regression

[work in progress, HB, Menen, Fuchs, Plehn]

Can we find an explicit analytic expression showing the interplay between the different features?

Finding analytic expressions using symbolic regression

[work in progress, HB, Menen, Fuchs, Plehn]

Can we find an explicit analytic expression showing the interplay between the different features?

→ Use symbolic regression!

[Schmidt&Lipson `09, Udrescu&Tegmark `19, Cranmer et al. `19, `20, `23]

Finding analytic expressions using symbolic regression

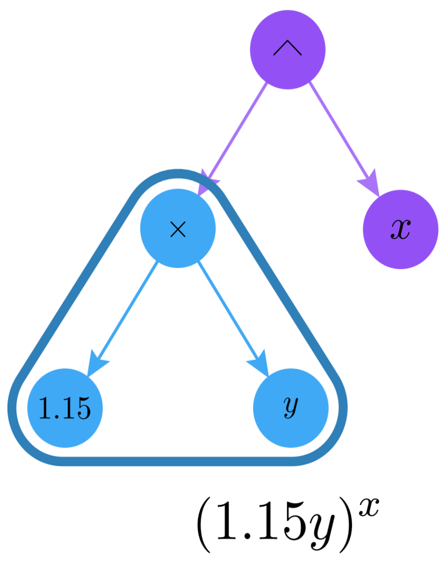
[work in progress, HB, Menen, Fuchs, Plehn]

Can we find an explicit analytic expression showing the interplay between the different features?

→ Use symbolic regression!

[Schmidt&Lipson `09, Udrescu&Tegmark `19, Cranmer et al. `19, `20, `23]

- Symbolic regression aims at fitting data using analytic equations.
- Analytic equations are implemented in terms of a tree-like structure.
- Uses multi-population evolutionary algorithm for optimization.
- Interplay between goodness-of-fit and complexity of equation.



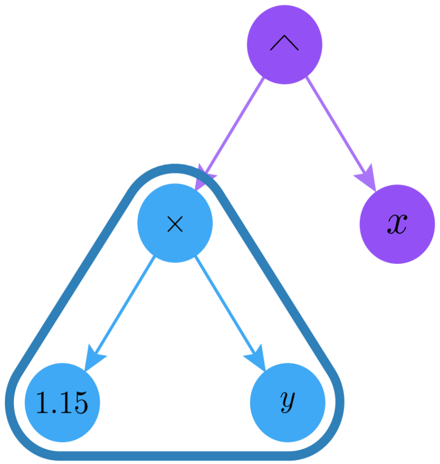
Finding analytic expressions using symbolic regression

[work in progress, HB, Menen, Fuchs, Plehn]

Can we find an explicit analytic expression showing the interplay between the different features?

→ Use symbolic regression!
[Schmidt&Lipson `09, Udrescu&Tegmark `19, Cranmer et al. `19, `20, `23]

- Symbolic regression aims at fitting data using analytic equations.
- Analytic equations are implemented in terms of a tree-like structure.
- Uses multi-population evolutionary algorithm for optimization.
- Interplay between goodness-of-fit and complexity of equation.



$(1.15y)^x$

Preliminary example: background discrimination VBF vs. ggF2j

Finding analytic expressions using symbolic regression

[work in progress, HB, Menen, Fuchs, Plehn]

Can we find an explicit analytic expression showing the interplay between the different features?

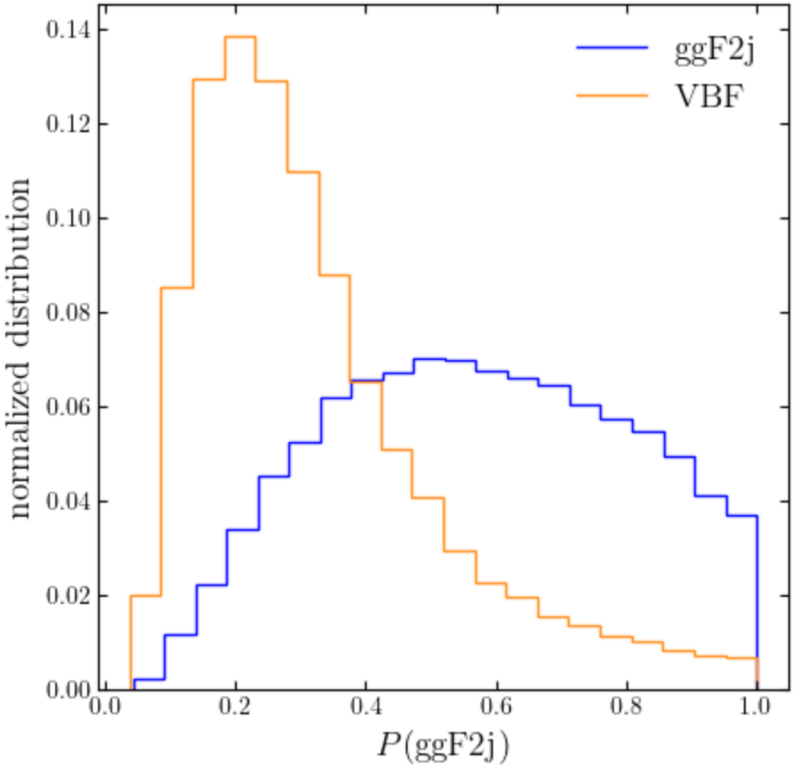
→ Use symbolic regression!

[Schmidt&Lipson `09, Udrescu&Tegmark `19, Cranmer et al. `19, `20, `23]

- Symbolic regression aims at fitting data using analytic equations.
- Analytic equations are implemented in terms of a tree-like structure.
- Uses multi-population evolutionary algorithm for optimization.
- Interplay between goodness-of-fit and complexity of equation.

Preliminary example: background discrimination VBF vs. ggF2j

⇒ symbolic regression → $P(\mathrm{ggF2j}) \sim \mathrm{Sigmoid}(p_{\mathrm{T},j_1} \log(|\Delta\eta_{jj}|))$



Conclusions

Conclusions

Summary:

- ggF2j production is a key process to probe the Higgs CP character nature.
- Existing analysis focus on VBF-like phase-space region and/or $\Delta\phi_{jj}$.
- Including full phase space information → significantly improved limits.
- Shapley values offer a mathematically well-defined way to understand feature importance.
- Shapley analysis shows that the traditional $\Delta\phi_{jj}$ observables is very sensitive to the interference term but not for distinguishing the squared terms.

Conclusions

Summary:

- ggF2j production is a key process to probe the Higgs CP character nature.
- Existing analysis focus on VBF-like phase-space region and/or $\Delta\phi_{jj}$.
- Including full phase space information → significantly improved limits.
- Shapley values offer a mathematically well-defined way to understand feature importance.
- Shapley analysis shows that the traditional $\Delta\phi_{jj}$ observables is very sensitive to the interference term but not for distinguishing the squared terms.

Outlook:

- Further optimize analysis using simulation-based inference,
- find analytic form of optimal observables using symbolic regression.

Conclusions

Summary:

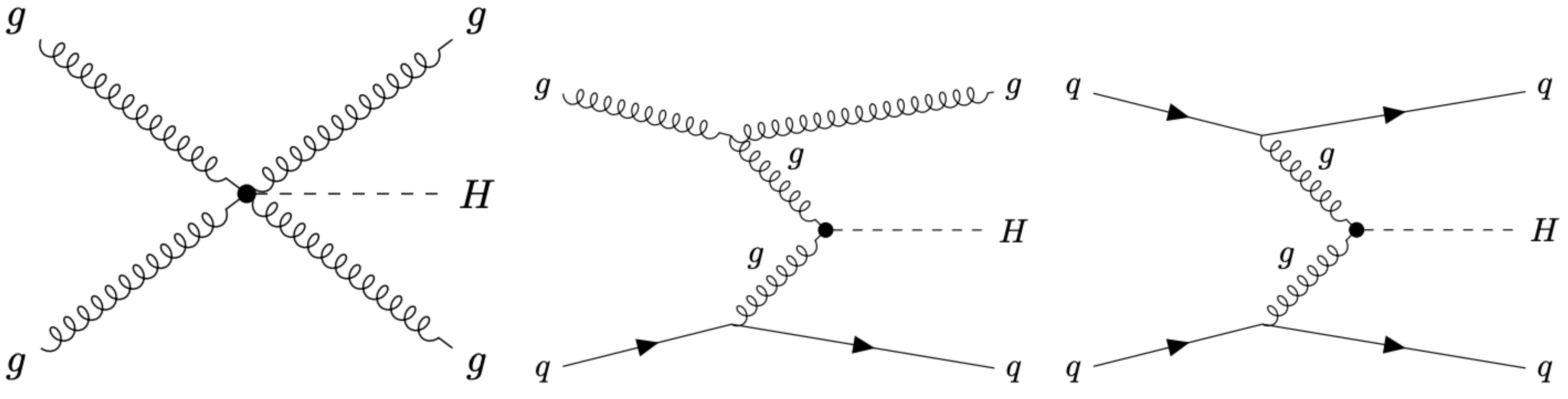
- ggF2j production is a key process to probe the Higgs CP character nature.
- Existing analysis focus on VBF-like phase-space region and/or $\Delta\phi_{jj}$.
- Including full phase space information → significantly improved limits.
- Shapley values offer a mathematically well-defined way to understand feature importance.
- Shapley analysis shows that the traditional $\Delta\phi_{jj}$ observables is very sensitive to the interference term but not for distinguishing the squared terms.

Outlook:

- Further optimize analysis using simulation-based inference,
- find analytic form of optimal observables using symbolic regression.

Thanks for your attention!

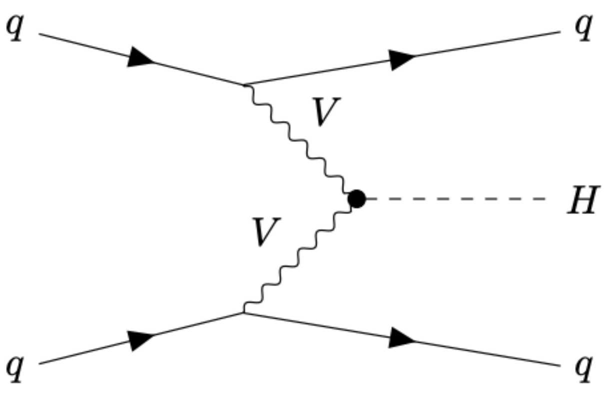
Conclusions



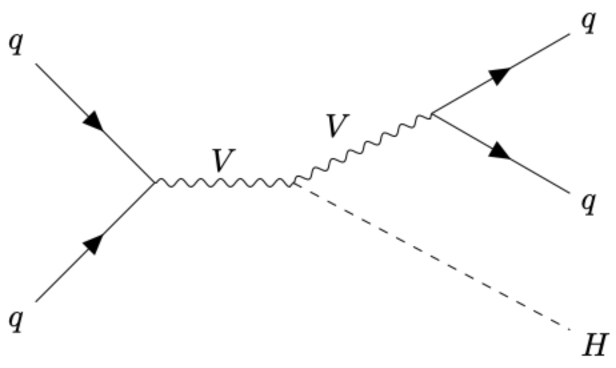
(a) gg-initiated (b) gq-initiated (c) qq-initiated

Appendix

Background processes

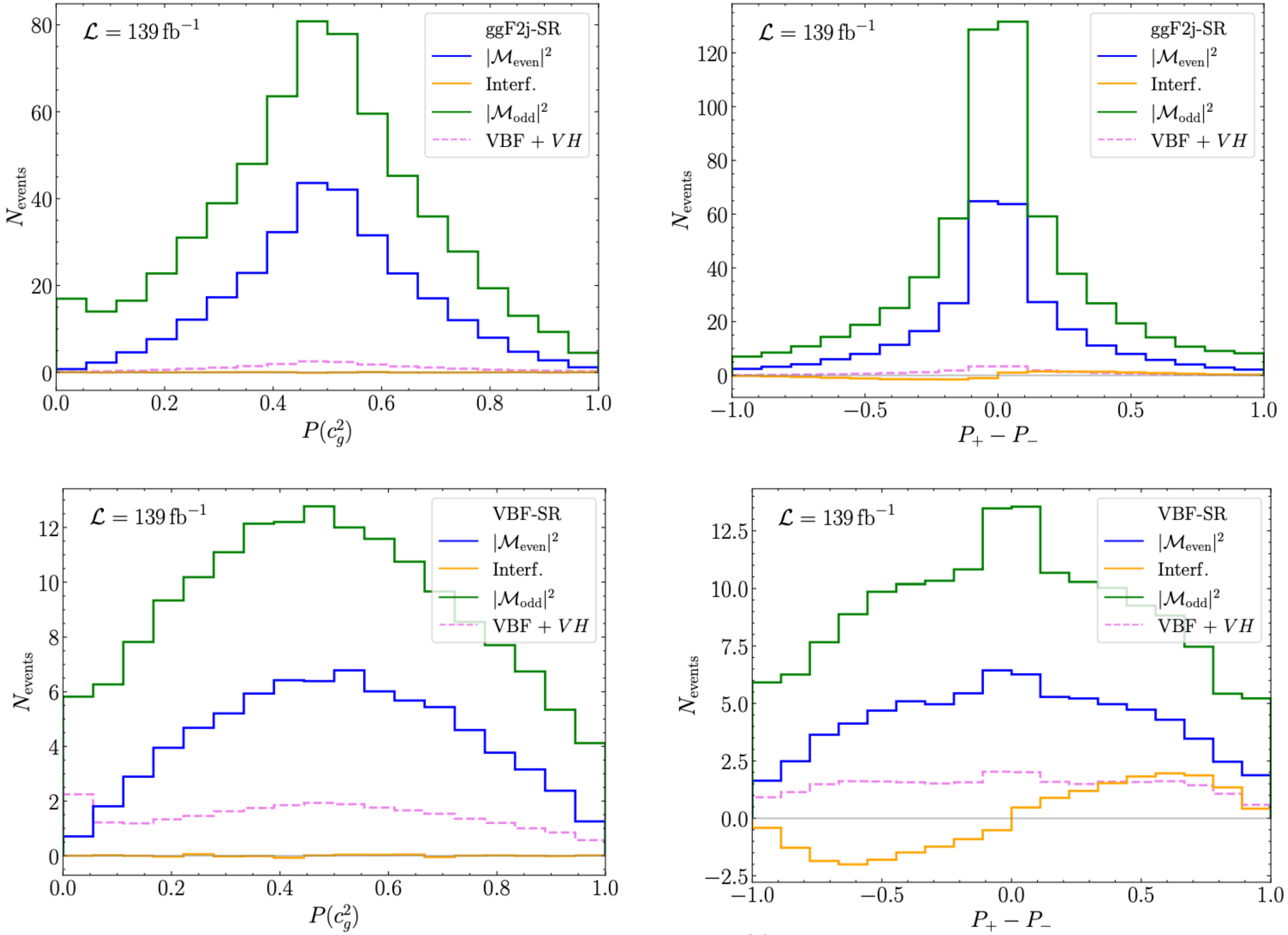


(a) VBF

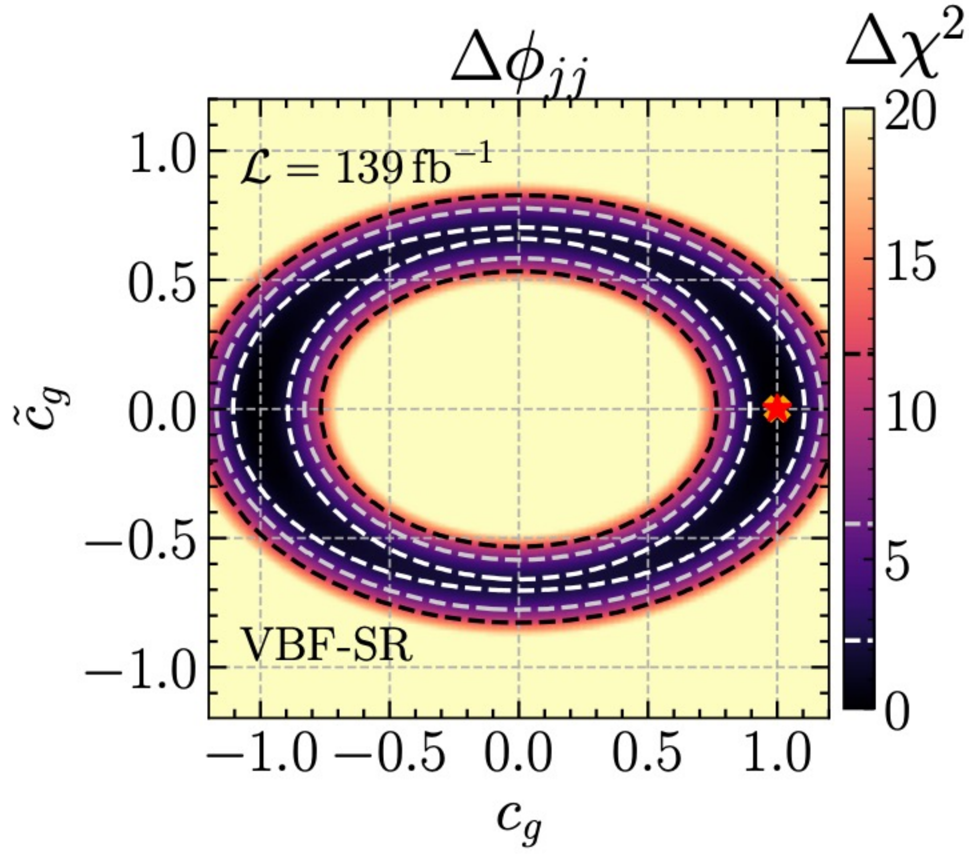


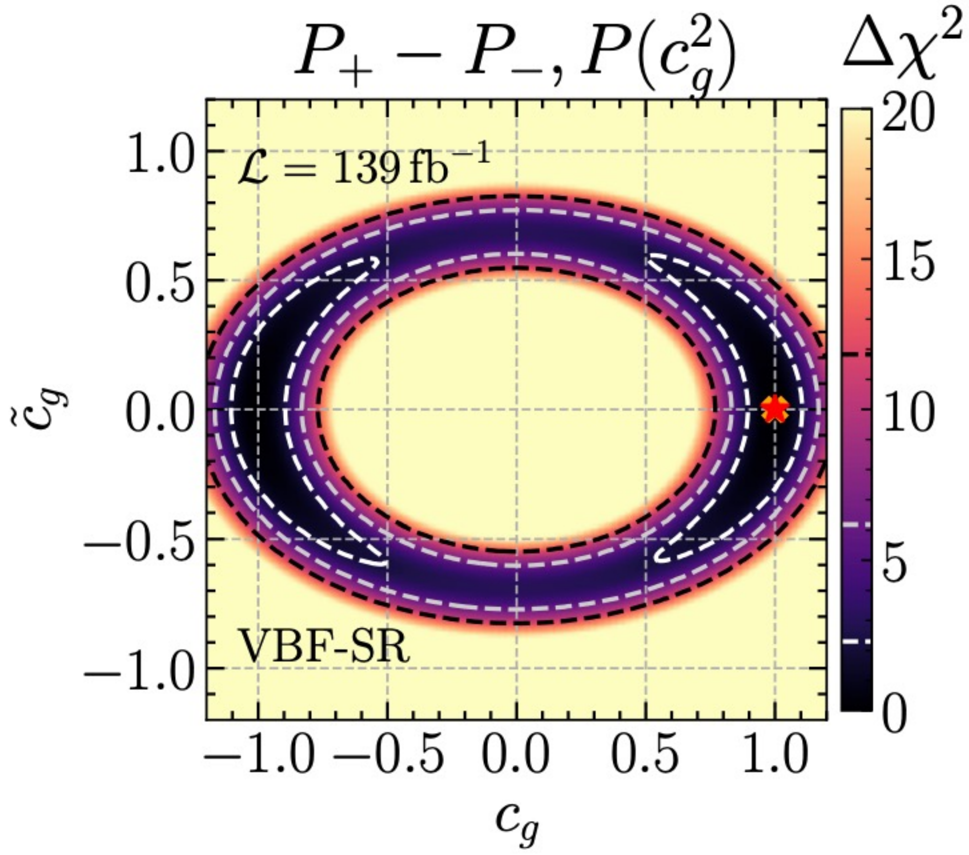
(b) VH

Classifier scores

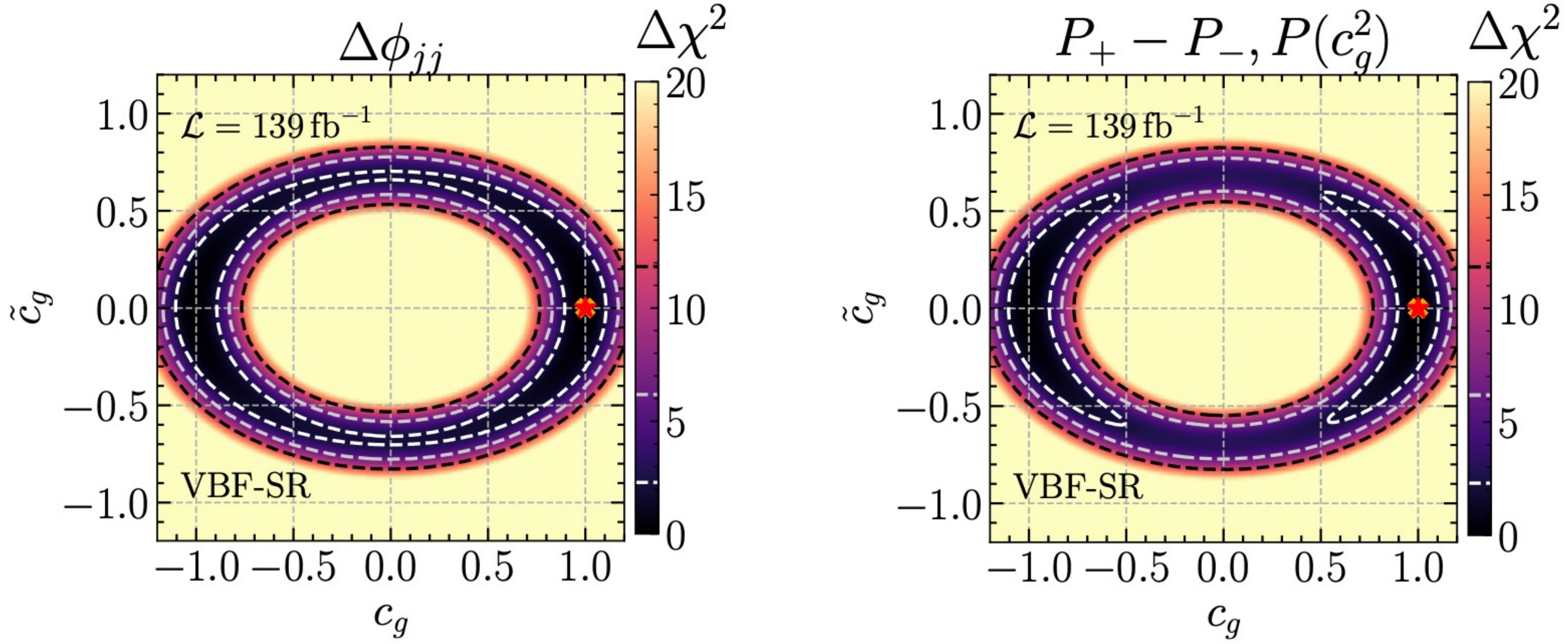


VBF signal region



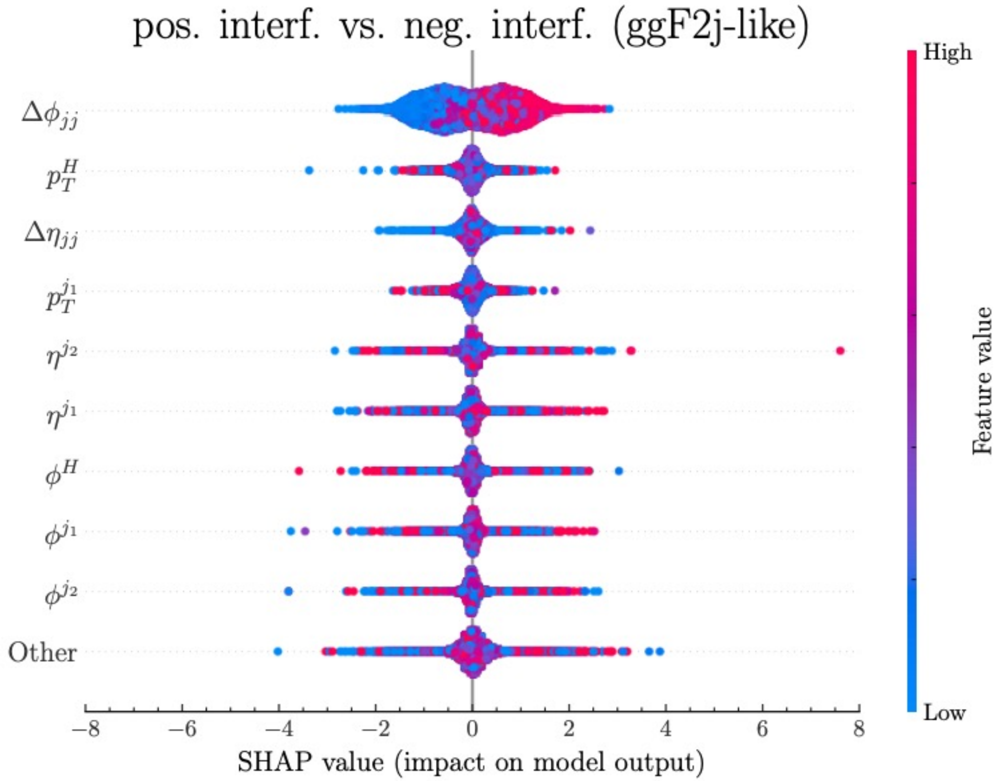


VBF signal region



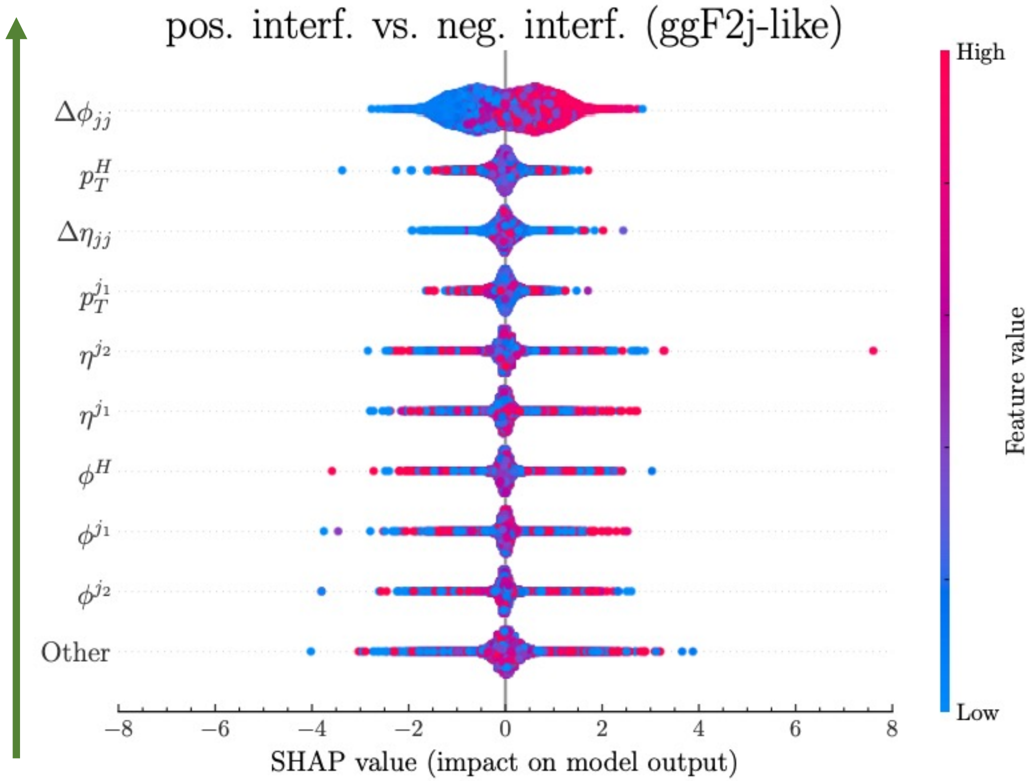
→ $\Delta\phi_{jj}$ limit only slightly worse than limit based on classifiers.

Results for interference classifiers



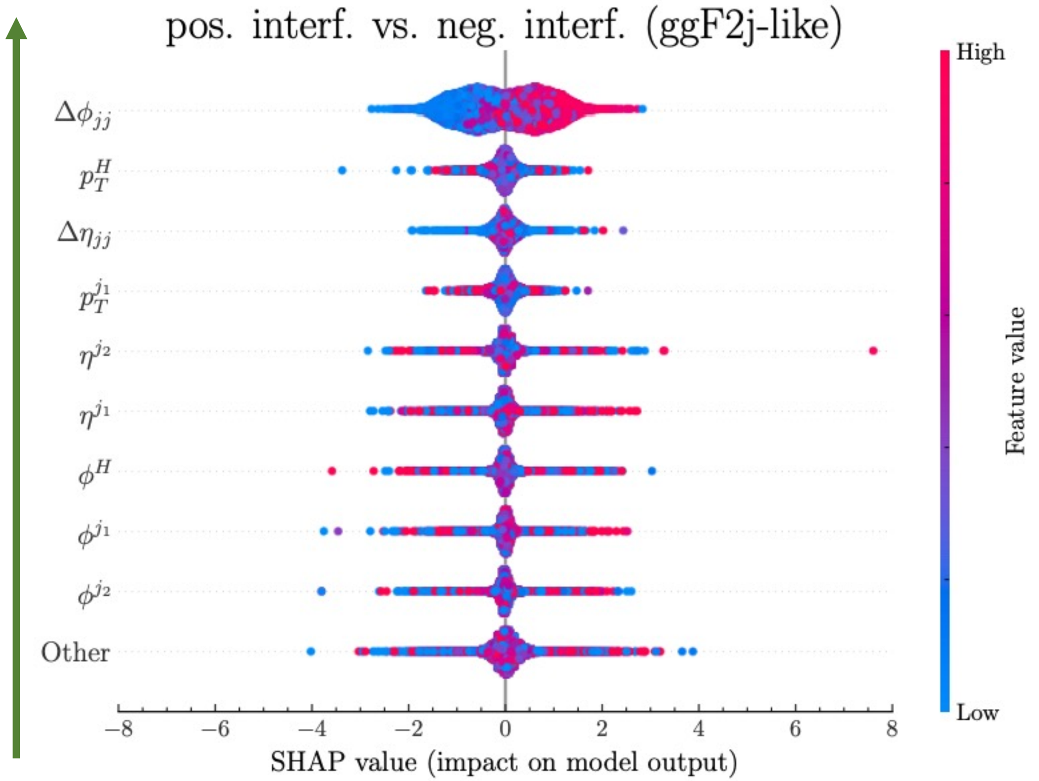
Results for interference classifiers

Importance of observable
($\sim \sum_{\text{events}} |\phi_j|$)



Results for interference classifiers

Importance of observable
($\sim \sum_{\text{events}} |\phi_j|$)

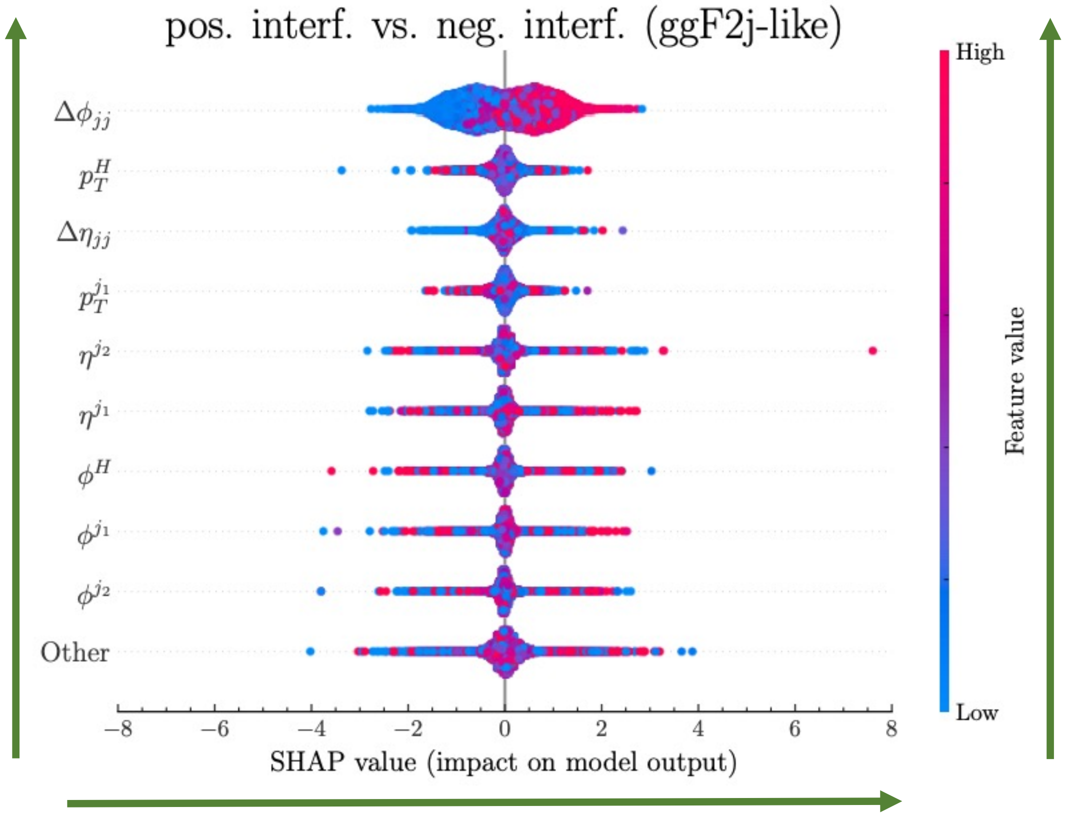


Shapley value for each event (shown as dot)

Results for interference classifiers

Importance of observable
($\sim \sum_{\text{events}} |\phi_j|$)

Value of observable

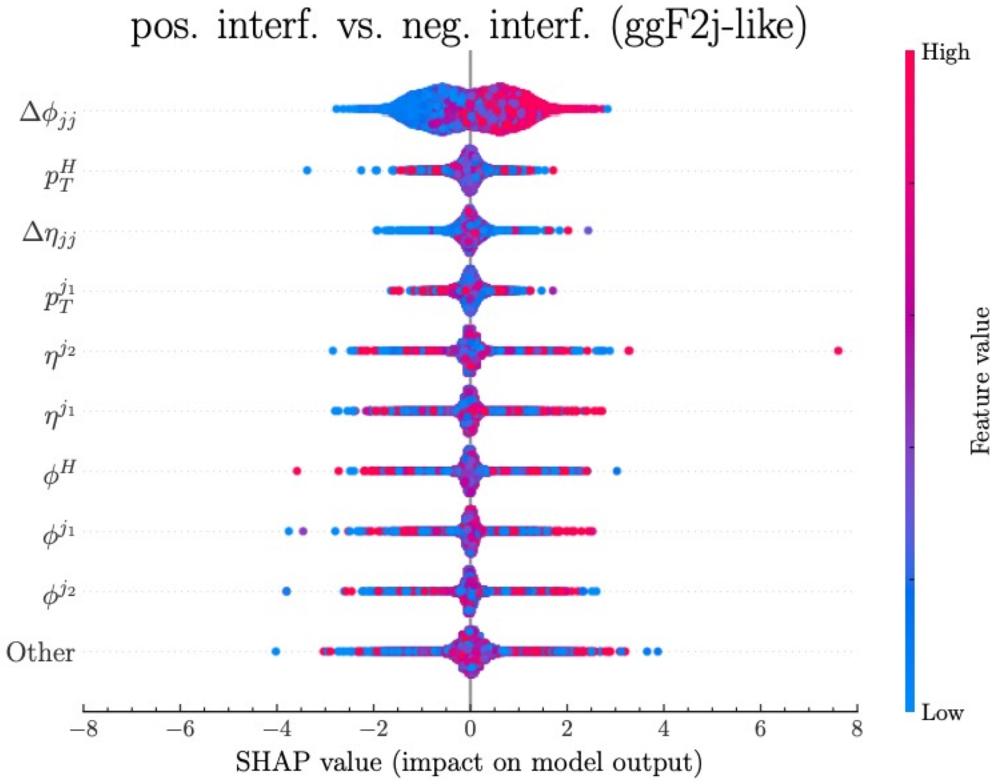


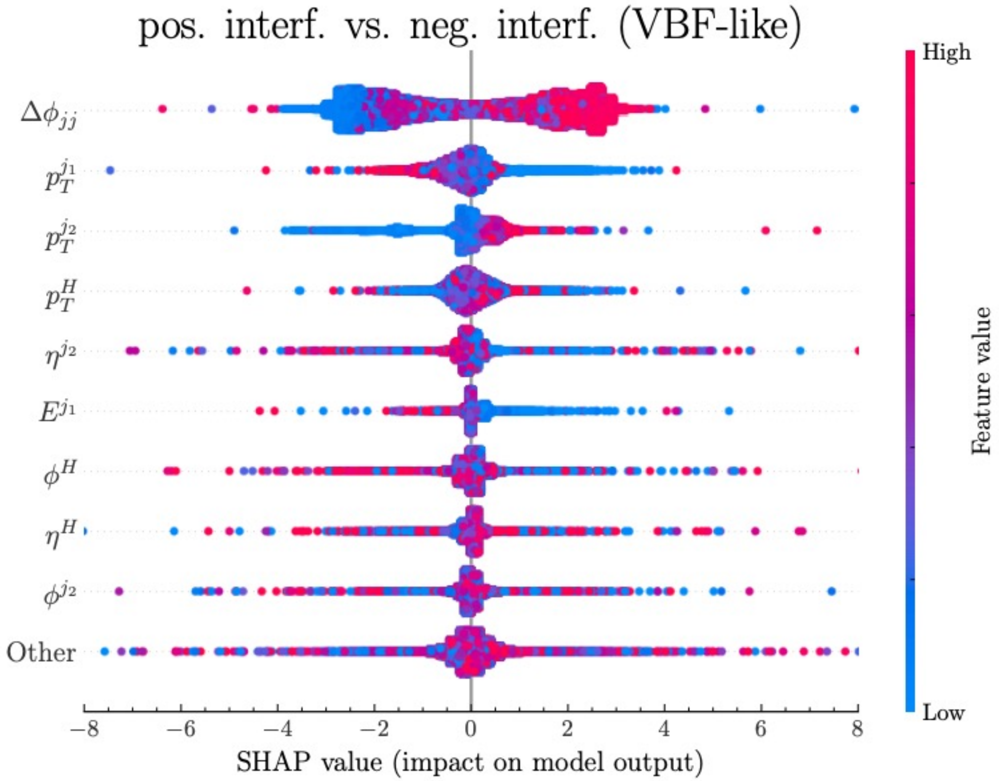
Shapley value for each event (shown as dot)

Results for interference classifiers

Importance of observable
($\sim \sum_{\text{events}} |\phi_j|$)

Value of observable



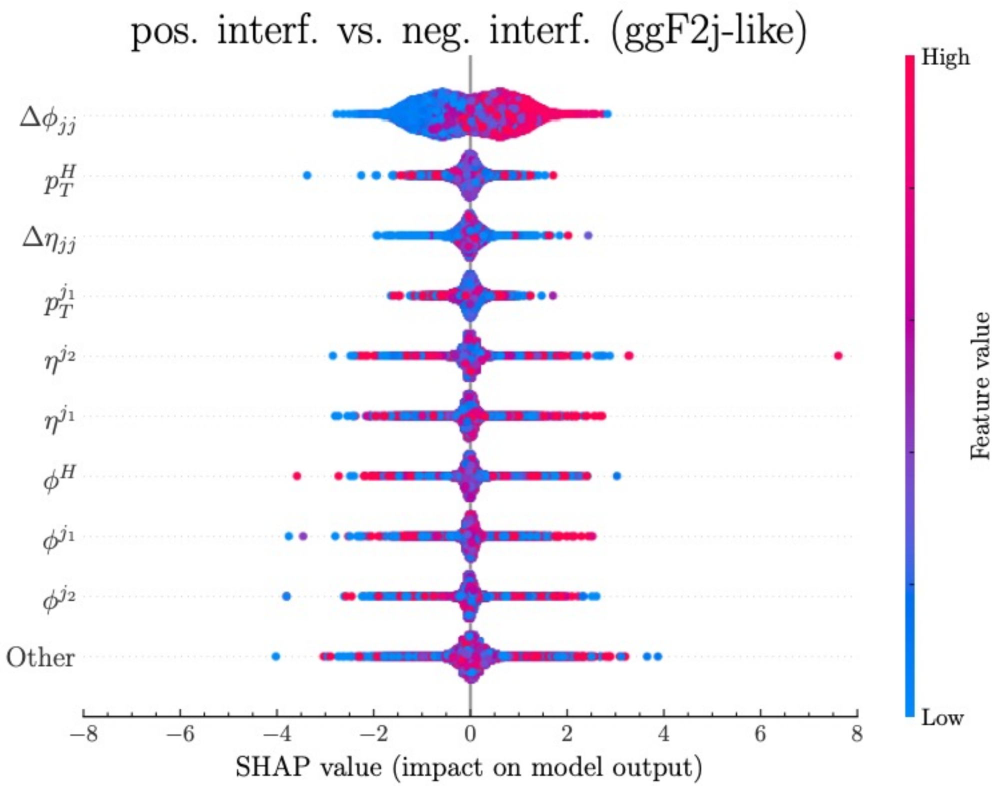


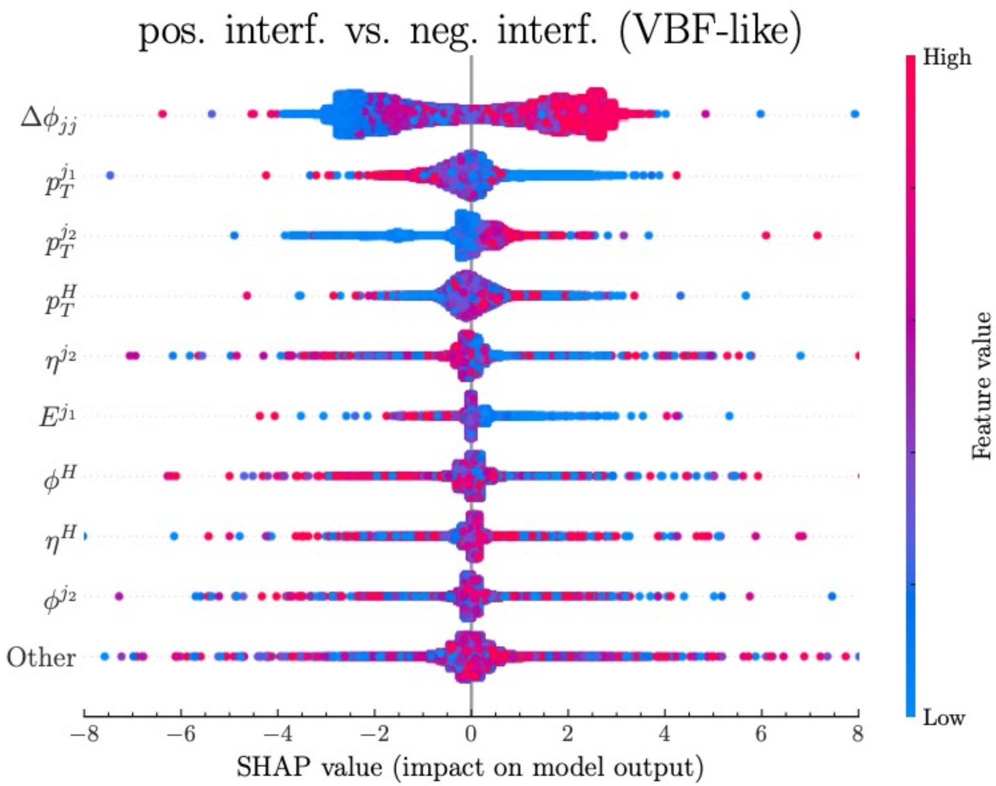
Shapley value for each event (shown as dot)

Results for interference classifiers

Importance of observable
($\sim \sum_{\text{events}} |\phi_j|$)

Value of observable

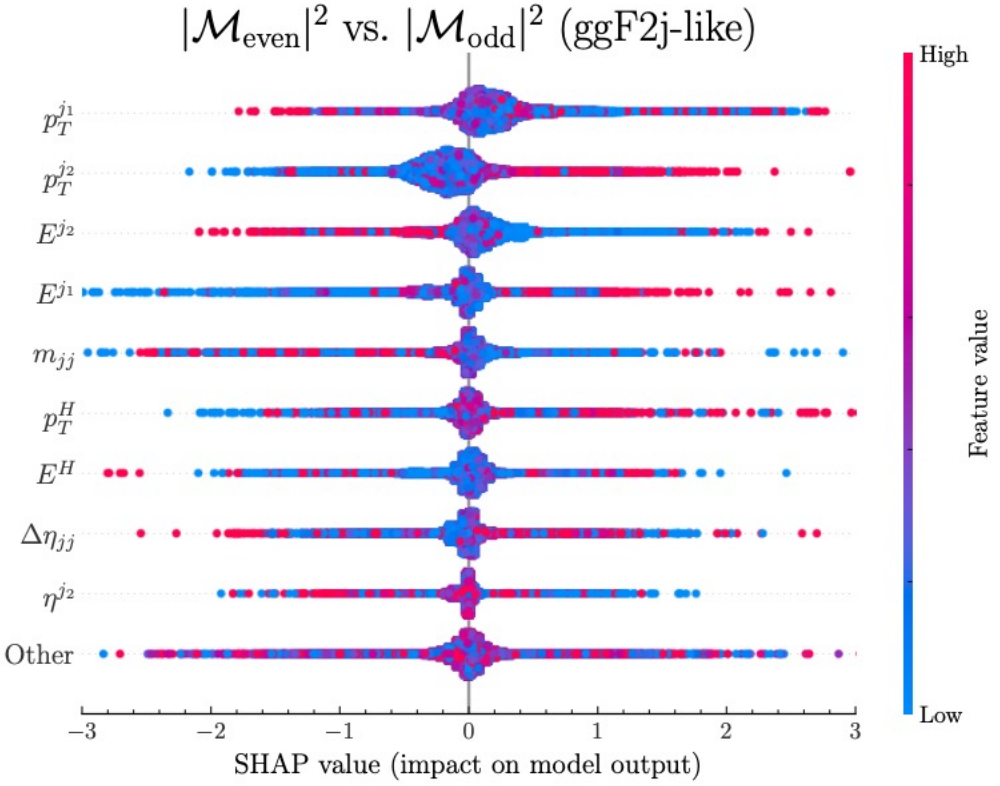


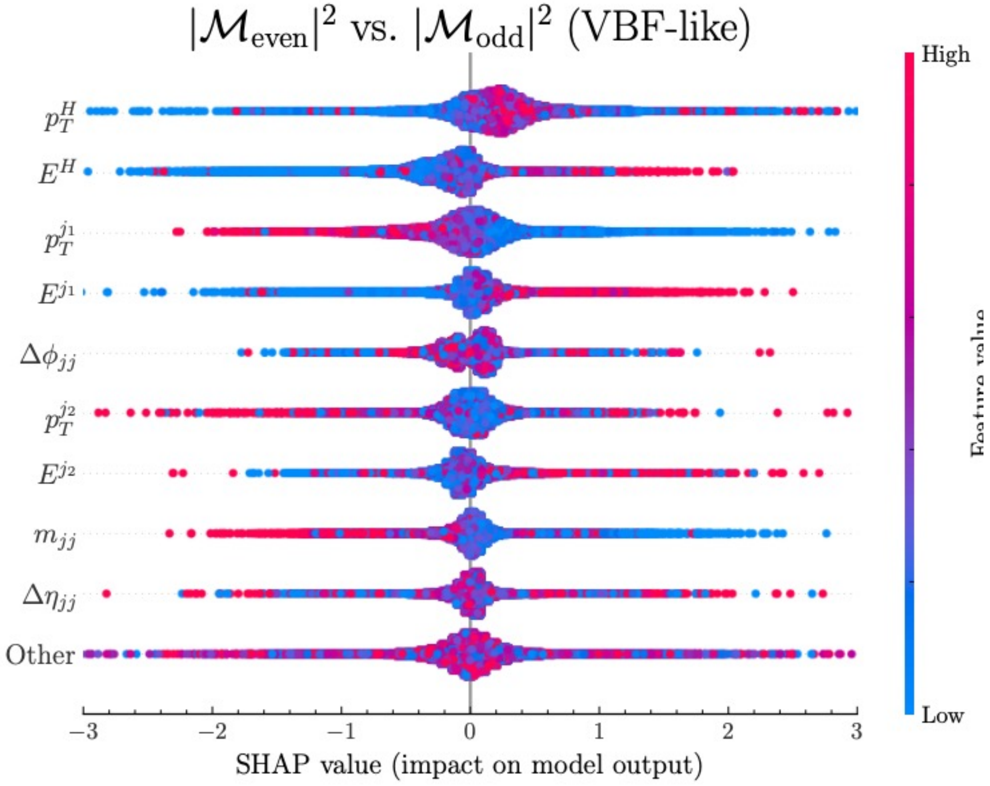


Shapley value for each event (shown as dot)

⇒ For the interference classifiers, as expected, the CP-odd $\Delta\phi_{jj}$ is most important.

Results for squared term classifiers





⇒ p_T of jets/Higgs most important, $\Delta\phi_{jj}$ plays only subleading role.
Disadvantage: interplay between observables hard to judge.