Constraining the Higgs potential with neural simulation-based inference for di-Higgs production

Radha Mastandrea,[1,2,*] Benjamin Nachman,[2,3,†] and Tilman Plehn[4,5,‡]

[1]*Department of Physics, University of California, Berkeley, California 94720, USA*
[2]*Physics Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA*
[3]*Berkeley Institute for Data Science, University of California, Berkeley, California 94720, USA*
[4]*Institut für Theoretische Physik, Universität Heidelberg, 69120 Heidelberg, Germany*
[5]*Interdisciplinary Center for Scientific Computing (IWR), Universität Heidelberg, 69120 Heidelberg, Germany*

(Received 7 June 2024; accepted 13 August 2024; published 3 September 2024)

Determining the form of the Higgs potential is one of the most exciting challenges of modern particle physics. Higgs pair production directly probes the Higgs self-coupling and should be observed in the near future at the High-Luminosity LHC. We explore how to improve the sensitivity to physics beyond the Standard Model through per-event kinematics for di-Higgs events. In particular, we employ machine learning through simulation-based inference to estimate per-event likelihood ratios and gauge potential sensitivity gains from including this kinematic information. In terms of the Standard Model Effective Field Theory, we find that adding a limited number of observables can help to remove degeneracies in Wilson coefficient likelihoods and significantly improve the experimental sensitivity.

DOI: 10.1103/PhysRevD.110.056004

I. INTRODUCTION

The discovery of the Higgs boson at the LHC in 2012 [1,2] completed the Standard Model (SM) and marked the beginning of an era of new measurements to characterize the Higgs boson properties. Measurements of Higgs boson production and decays are a powerful probe for beyond-the-Standard-Model (BSM) physics, and such searches are increasingly necessary given that the SM is unable to explain all known phenomena.

The most pressing fundamental question about the Higgs sector is the form of its potential. The structure of the potential is linked to fundamental questions like the stability of the Universe [3–5], the observed matter-antimatter asymmetry [6–8], and dark matter [9]. The SM assumes the simplest, renormalizable potential

$$V(\phi) = \mu^2(\phi^\dagger\phi)^2 + \lambda(\phi^\dagger\phi)^4, \tag{1}$$

for an SU(2) doublet ϕ. This potential describes trilinear and quartic self-couplings of the physical Higgs scalar h, with the coupling strength related to the Higgs vacuum expectation value ($v = 246$ GeV) and mass ($m_h = 125$ GeV) [10],

$$\lambda_{3h} = \frac{3m_h^2}{v} \quad \text{and} \quad \lambda_{4h} = \frac{3m_h^2}{v^2}. \tag{2}$$

Many BSM hypotheses shift these self-couplings and thus can be tested by measuring these shifts, either through quantum effects [11] or through direct multi-Higgs production.

ATLAS and CMS are capable of probing hh production and have performed many searches for both resonant and nonresonant di-Higgs boson production. In this way, they have placed limits on many BSM models affecting or extending the Higgs sector.

We focus on nonresonant Higgs-pair production [12–14], one of the main motivations for the High-Luminosity (HL) LHC [15,16] and for a future precision-hadron collider [17]. Effective field theory in terms of the Higgs doublet (SMEFT) [18–20] or the physical Higgs field (HEFT) [21–23] provides the appropriate framework for these analyses. Such effective theories rest on the fundamental assumption that any UV completion of the SM will not just modify the scalar Higgs-self coupling but induce all Wilson coefficients allowed by its underlying symmetry. A number of theoretical studies have explored this possibility [14,24–31], and recent experimental limits can be found in Refs. [32,33].

*Contact author: rmastand@berkeley.edu
†Contact author: bpnachman@lbl.gov
‡Contact author: plehn@uni-heidelberg.de

Given the small hh production rate, it will be essential to make the most of every candidate event. Many studies have relied on differential information to inform cuts to balance the number of hh-signal versus background events for a counting analysis. This is effective, but we know that there is much more information available in the full event kinematics [12,24,34].

For a fixed BSM hypothesis, the most powerful test statistic is the likelihood ratio with the SM. Collision events are independent, so the log-likelihood ratio factorizes into a rate term and a sum of per-event shape terms. Modern machine learning gives us access to the per-event likelihood ratios using neural simulation-based inference (NSBI) [35]. These likelihood ratios can be estimated with neural networks using simulated data at a variety of (B)SM parameter points. Our goal is to demonstrate that a multidimensional per-event analysis is a promising avenue to increase the sensitivity of current and future data to the hh signal.

Several studies have explored the effects of certain HEFT Wilson coefficients on the shapes of relevant kinematic distributions, particularly m_{hh} and $p_{T_{hh}}$ [36–39]. Such studies reflect a growing trend in particle physics toward using NSBI for exploring effective field theory (EFT) signatures, as also evidenced by the creation of several repositories that publicize useful code for carrying out these analyses [40–42].

In this paper, we explore how SMEFT Wilson coefficients associated with hh production can be better constrained by integrating per-event shape information, similar to earlier studies for associated Higgs production [43]. We consider the HL-LHC and a future 100 TeV hadron collider and attempt to constrain a set of three dimension-6 Wilson coefficients.

The paper is organized as follows. In Sec. II, we review relevant SM and SMEFT aspects for hh production at the two colliders. In Sec. III, we explain the event generation and selection procedure as well as provide an estimation of the expected background yields. In Sec. IV, we describe the rate and shape analyses used to constrain the Wilson coefficients. Numerical results are presented in Sec. V. The paper ends with an outlook in Sec. VI.

II. SMEFT SETUP

In this section, we briefly review some relevant theoretical background to hh production within the SMEFT framework and define the operators we choose to vary.

SMEFT is a consistent quantum field theory framework for parametrizing the effects of new physics on the known Standard Model fields. The SMEFT Lagrangian is written as

$$\mathcal{L}_{\text{SMEFT}} = \mathcal{L}_{\text{SM}} + \sum_i \frac{c_i}{\Lambda^{d_i-4}} \mathcal{O}_i, \tag{3}$$

where Λ is the matching scale to the UV theory, $\mathcal{O}_i$ is an operator of dimension d_i composed of Standard Model fields, and c_i is the Wilson coefficient governing the operator's coupling strength.

In Eq. (3), we see that, for instance, at dimension 6 there is a degeneracy in the way we assign a given effect on the Wilson coefficient c_i and the scale Λ. This degeneracy is only broken by matching to a full theory with given masses and couplings. Throughout this paper, we report SMEFT limits on c_i for the fixed scale

$$\frac{c_i}{\Lambda^2} \xrightarrow{\text{limits}} \frac{c_i}{(1\ \text{TeV})^2}. \tag{4}$$

At dimension 6, there exist 59 independent SMEFT operators, ignoring the flavor structure [19]. Global analyses of these operators is a realistic task for the LHC [44–46], while the 44,807 dimension-8 operators [47] are unlikely to be a realistic framework for global LHC analyses [48]. To eventually combine hh results with such a global analysis [49], we limit ourselves to dimension-6 operators.

For Higgs physics, there exist two complementary EFT descriptions. While the SMEFT treats the Higgs as part of a SU(2) doublet, the HEFT framework uses the physical Higgs boson and the Goldstones as the degrees of freedom. Different BSM models may be matched more easily to the SMEFT or the HEFT treatment, depending on the degree that the SM-Higgs is responsible for the gauge boson masses; a small comparison of models is given in Chap. 2 of Ref. [39] and Sec. 5.4 of Ref. [20]. We use the SMEFT framework, assuming that the SM-Higgs really forms an SU(2) doublet with the electroweak Goldstones. The event generation framework is implemented in MadGraph [50].

At the LHC, the main SM contribution to hh production is gluon-gluon fusion, with the Feynman diagrams presented in Fig. 1. The triangle diagram is sensitive to the trilinear Higgs vertex, and the box diagram can enhance this sensitivity through a cancellation at threshold [51]. When including SMEFT operators up to dimension 6, the SM coupling of Eq. (2) is modified to [52]

$$\lambda_{3h} = \frac{3m_h^2}{v}\left[1 - \frac{c_{\phi d}v^2}{\Lambda^2} - \frac{2c_\phi v^4}{\Lambda^2 m_h^2} + \frac{4c_{\phi d}v^2}{3\Lambda^2 m_h^2}\sum_{j<k}^{3}(p_j p_k)\right], \tag{5}$$

where c_ϕ and $c_{\phi d}$ are the Wilson coefficients associated with the operators

$$\mathcal{O}_\phi = (\phi^\dagger\phi)^3 \quad \text{and} \quad \mathcal{O}_{\phi d} = \partial_\mu(\phi^\dagger\phi)\partial^\mu(\phi^\dagger\phi). \tag{6}$$

We see that these two operators directly affect the trilinear Higgs coupling, but in different ways. In addition, $c_{\phi d}$ changes the field normalization of the physical Higgs and, with that, all physical Higgs couplings.

The complete set of operators contributing to Higgs pair production is given in Table I [53]. The operator $\mathcal{O}_{\phi D}$

FIG. 1. Interfering Feynman diagrams associated with hh production in the Standard Model.

contributes the same way as $\mathcal{O}_{\phi d}$, but it violates custodial symmetry and is therefore strongly constrained by electroweak precision data.

Because the two Feynman diagrams depend on the top Yukawa coupling differently, we include the modified top Yukawa coupling through

$$\mathcal{O}_{t\phi} = (\phi^\dagger\phi)\bar{Q}t\tilde{\phi} + \text{H.c.} \tag{7}$$

While a BSM-induced $\mathcal{O}_{\phi G}$ contributes non-negligibly to single-Higgs and double-Higgs production, the induced ggh and $gghh$ couplings are related through low-energy theorems, which means the BSM contribution to hh production will be strongly constrained by single Higgs production and structurally similar to the modified top Yukawa [54]. Similarly, while $\mathcal{O}_{tG}$ will produce a number of new Higgs vertices, the coefficient is most strongly bounded by single-Higgs and tt processes. The effect of our three operators on the hh Feynman diagrams is summarized in Fig. 2.

Current experimental constraints on the Wilson coefficients $c_{\phi d}$ and $c_{t\phi}$ have been found through global fits to LHC data [55,56]. For instance, Ref. [56] gives the current (profiled) 95% confidence bounds

$$c_{\phi d} \in [-2.23, 3.28] \quad \text{and} \quad c_{t\phi} \in [-3.56, 5.75], \tag{8}$$

where we note that the definition of $\mathcal{O}_{\phi d}$ used in Ref. [56] differs by the one used in the generating SMEFT@NLO [53] MadGraph model by a factor of $1/2$. In the future, tight bounds on $c_{t\phi}$ will come from the measurement of tth production.

TABLE I. Dimension-6 SMEFT operators contributing to the $gg \to hh$. For each operator, we provide the name of the Wilson coefficient in the SMEFT@NLO [53] model for MadGraph.

Operator	Explicit form	SMEFT@NLO ID
$\mathcal{O}_\phi$	$(\phi^\dagger\phi)^3$	cp
$\mathcal{O}_{\phi d}$	$\partial_\mu(\phi^\dagger\phi)\partial^\mu(\phi^\dagger\phi)$	cdp
$\mathcal{O}_{\phi D}$	$(\phi^\dagger D^\mu\phi)^\dagger(\phi^\dagger D_\mu\phi)$	cpDC
$\mathcal{O}_{t\phi}$	$\phi^\dagger\phi\bar{Q}t\tilde{\phi}$ + H.c.	ctp
$\mathcal{O}_{\phi G}$	$(\phi^\dagger\phi)G_A^{\mu\nu}G_{\mu\nu}^A$	cpG
$\mathcal{O}_{tG}$	$ig_s(\bar{Q}\tau^{\mu\nu}T_A t)\tilde{\phi}G_{\mu\nu}^A$ + H.c.	ctG

Current bounds on c_ϕ are much looser. Since hh production has not yet been observed at the LHC, the most recent limits come from the 2022 CMS and ATLAS summaries [32,33], which constrain the Higgs self-coupling modification $\kappa_\lambda \in [-1.24, 6.49]$ and $\in [-1.4, 6.1]$, respectively. Assuming a cutoff scale $\Lambda = 1$ TeV, the first interval can be converted to $c_\phi \in [-12.95, 5.39]$, but this limit is not based on a global EFT analysis and is neither model-independent nor conservative.

III. EVENT GENERATION

In this section, we provide details on the event generation procedure for the hh and background samples used in this analysis.

For the decay channel, we choose the process

$$pp \to hh \to (b\bar{b})(\gamma\gamma). \tag{9}$$

The $h \to b\bar{b}$ decay is ideal because it has the largest branching fraction at 58%; the $h \to \gamma\gamma$ channel has a much smaller branching rate of 0.227%, but it benefits from the excellent photon identification and resolution [13]. At the LHC, it is expected that a measurement of hh decay will be made through combining five channels [32,33]: $\bar{b}b\bar{b}b$, $b\bar{b}\tau\tau$ [57–59], $b\bar{b}\gamma\gamma$, $4l$, and $b\bar{b}ZZ$, with $b\bar{b}\gamma\gamma$ providing the best sensitivity along with $b\bar{b}\tau\tau$. Because of the large QCD backgrounds, the naively most promising $\bar{b}b\bar{b}b$ channel will likely be most useful for on-shell modifications of Higgs pair production.

We consider two colliders to probe the Higgs self-coupling directly:

(i) The HL-LHC, $\sqrt{s} = 14$ TeV to 3 ab^{-1} [15].
(ii) A future hadron collider, $\sqrt{s} = 100$ TeV to 30 ab^{-1} [60].

A. SMEFT signal

The $gg \to hh$ production cross sections in the SM are known to next-to-leading order (NLO) in QCD [61–67], approximate next-to-next-to-leading order (NNLO) [68–70], N^{3}LO [71,72], and NLO including parton showers [73]. We use the NLO rate prediction 32.81(7) fb (+13.5%, −12.5%) for the HL-LHC and 1140(2) fb (+10.7%, −10.0%) at 100 TeV [28,74]. The combined branching ratio for $hh \to b\bar{b}\gamma\gamma$ is 0.262% [75]. The expected SM event yields are given in Table II.

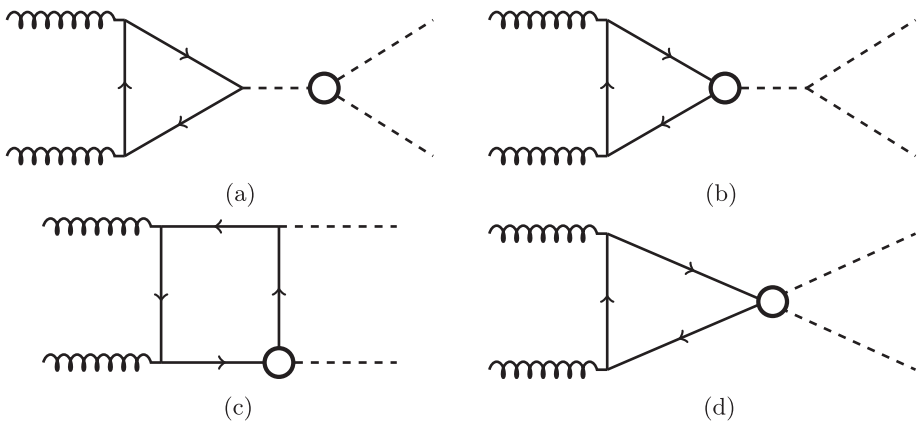

FIG. 2. Effect of the three dimension-6 operators considered in our analysis on the hh Feynman diagrams. The BSM vertices are denoted by open circles. For diagram (c), the BSM vertex can be on either of the tth junctions. (a) Contribution from $\mathcal{O}_\phi$ and $\mathcal{O}_{\phi d}$. (b) Contribution from $\mathcal{O}_{\phi d}$ and $\mathcal{O}_{t\phi}$. (c) Contribution from $\mathcal{O}_{\phi d}$ and $\mathcal{O}_{t\phi}$. (d) Contribution from $\mathcal{O}_{t\phi}$.

To generate hh events, we use MadGraph3.5.1 [50] with the SMEFT@NLO [53] model, which assumes a finite top quark mass and is leading order (LO) with respect to the ggh vertex. Note that SMEFT@NLO canonically normalizes the dynamical Higgs field, so the derivative correction to the trilinear coupling visible in the last term of Eq. (5) is not modeled. We use MadSpin [76] for the Higgs decays and PYTHIA8.306 [77] for the underlying event, parton shower, and hadronization. To simulate detector effects, we use DELPHES3.5.0 [78] with the HL-LHC card. The detector simulation uses a particle-flow-like reconstruction and clusters jets with the anti-k_t [79] algorithm with $R = 0.4$ using FastJet [80].

For SMEFT extensions, precision predictions for hh production exist to NLO with subleading operators [81], combined with parton shower [82], including truncation uncertainties [83], and approximate NNLO precision [84] As to be discussed in Sec. III C, we generate signal events at a set of 10 points in the three-dimensional (3D) Wilson coefficient space, such that event weights at an arbitrary parameter point can be calculated using the morphing basis. At the SM point, we generate 10^6 (1.5×10^6) events for the HL-LHC (100 TeV), and at each of the other nine points, we generate 5×10^5 (7.5×10^5) events.

After event generation, we carry out a number of analysis selections to guarantee the event acceptance and enhance the hh-signal rate relative to the background ratio. The choice of cuts leaves us a lot of freedom. Looser cuts will lower S/B, which is less ideal for a cut-and-count analysis but not a problem for machine learning-based analyses. We leave further explorations of this tradeoff to future studies. For the selections, we follow a strategy similar to that in Ref. [24]. We define four categories of selections:

(1) *b-quality* has at least two b-tagged jets.

(2) *Kinematics* has at least two jets and two photons with

$$p_T > 30 \text{ GeV} \quad \text{and} \quad |\eta| < 2.4. \tag{10}$$

The four main analysis objects must have an angular separation of $\Delta R > 0.4$. These cuts select events with high triggering efficiencies.

(3) *Higgs mass windows* have two mass windows,

$$|m_{b\bar{b}} - m_h| < 25 \text{ GeV}$$
$$|m_{\gamma\gamma} - m_h| < 3 \text{ GeV}, \tag{11}$$

TABLE II. Signal and background cut flows for the HL-LHC ($S/B = 0.30$) and the Future Circular Collider ($S/B = 0.83$). Both the signal and background event yields reflect the NLO prediction.

	HL-LHC, 14 TeV, 3 ab^{-1}				Future Circular Collider, 100 TeV, 30 ab^{-1}			
	Signal		Background		Signal		Background	
	Events	Retention (%)	Events	Retention (%)	Events	Retention (%)	Events	Retention (%)
Start	257	100	⋯	⋯	89,604	100	⋯	⋯
+ tagging and efficiencies	95	37.1	4.65×10^4	100	29,600	33.0	5.16×10^6	100
+ kinematic cuts	49	18.9	1.43×10^4	30.8	11,100	12.3	1.58×10^6	30.6
$+m_h$ windows	15	5.89	4.09×10^2	0.88	3,950	4.40	4.02×10^4	0.78
+ angular cuts	13	4.92	4.37×10^1	0.094	3,600	4.02	4.34×10^3	0.084

which strongly reduce the background events while minimally reducing the signal.

(4) *Angular cuts* tighten the loose kinematic cuts, setting the stricter angular separation thresholds." The item list looks fine otherwise

$$\Delta R_{b\gamma} > 1 \quad \text{and} \quad \Delta R_{\gamma\gamma} < 2. \tag{12}$$

This reduces the background by a factor of ∼7 and the signal by only a factor ∼1.2 [13].

After event selection, we are left with 129,000 (176,000) signal events for the HL-LHC (100 TeV), spread across the 10 morphing generation points.

B. Continuum backgrounds

The main backgrounds for the $hh \to b\bar{b}\gamma\gamma$ channel can be divided into three categories:

(1) *Continuum:* $gg \to b\bar{b}\gamma\gamma$ with two real b quarks and two real photons, but no intermediate resonances.

(2) *Mistags*: a light jet mistagged as a photon, or a light or a c jet mistagged as a b jet. Possible channels include $b\bar{b}j\gamma$, $b\bar{b}jj$, $jj\gamma\gamma$, and $c\bar{c}\gamma\gamma$, all without an intermediate Higgs.

(3) *Single Higgs:* a single-Higgs boson produced in association with other objects, including $b\bar{b}h$, $c\bar{c}h$, and Zh note that there is some overlap between this category and the mistags above.

To estimate the importance of each background, we run MadGraph at LO for each of the SM processes in Table III at leading order. Slightly looser kinematic selections to those from Sec. III A are enforced at the parton level during event generation, to speed up the analysis. When calculating the expected background yields, we do not run PYTHIA, and we simulate detector effects manually for computational tractability: p_T smearing, jet efficiencies, photon efficiencies, and mistagging rates are carried out following the HL-LHC DELPHES run card.

We calculate the NLO yields by multiplying the LO yields for each process by its corresponding K factor, given in Table IV. In is important to highlight that the K factors listed are for the total cross section. In reality, the K factors are scale-dependent functions of p_T and will modify the shapes of the distributions of kinematic observables. Since this shape modification is expected to be independent of the SMEFT Wilson coefficients, we do not consider this effect in this analysis.

Altogether, we generate 8×10^6 (7.6×10^6) background events for the HL-LHC (100 TeV). This large number is necessary to train the parametrized classifier for the shape analysis, described in Sec. IV A. After the selection given in

TABLE IV. $K_{\mathrm{NLO/LO}}$ factors for the cross section for the designated background process. Where we cannot find a value in the literature for the 100 TeV K factors, we use the 14 TeV value. For the 100 TeV K factor for the $jj\gamma\gamma$ background, there is a cut $p_{T_j} > 50$ GeV.

Process	$K_{\mathrm{NLO/LO}}$, 14 TeV	$K_{\mathrm{NLO/LO}}$, 100 TeV
$pp \to b\bar{b}\gamma\gamma$	1.36 [85] (13 TeV)	(14 TeV value)
$pp \to b\bar{b}j\gamma$	1.76 [86]	(14 TeV value)
$pp \to b\bar{b}jj$	1.50 [86]	(14 TeV value)
$pp \to jj\gamma\gamma$	1.54 [86]	1.3 [87]
$pp \to Zh$	1.19 [86]	(14 TeV value)

TABLE III. All background cross sections are given in pb. The $jj\gamma\gamma$ background includes the $c\bar{c}\gamma\gamma$ background. The $pp \to Zh$ cross sections and event yields are for the full decay $pp \to Z(\to b\bar{b})h(\to \gamma\gamma)$. Cross sections are calculated at leading order with loose kinematic cuts. All rows except the last represent the LO event yields; the last row multiplies by the process-specific K factor stated in the text to give the NLO event yields.

		$pp \to b\bar{b}\gamma\gamma$	$pp \to b\bar{b}j\gamma$	$pp \to b\bar{b}jj$	$pp \to jj\gamma\gamma$	$pp \to Zh$	Total
HL-LHC	Cross-section (LO) (pb)	0.009758	61.25	8833	1.946	5.773×10^{-5}	
	Events w/ loose kinematic cuts	2.92×10^4	1.84×10^8	2.65×10^{10}	5.83×10^6	1.73×10^2	2.67×10^{10}
	+ tagging and efficiencies	1.10×10^4	1.56×10^4	5.50×10^2	2.06×10^3	6.6×10^1	2.93×10^4
	+ kinematic cuts	4.55×10^3	3.96×10^3	1.75×10^2	5.48×10^2	4.3×10^1	9.28×10^3
	$+m_h$ windows	1.59×10^2	5.90×10^1	8.01×10^0	3.27×10^1	2.2×10^1	2.81×10^2
	+ angular cuts	1.29×10^1	6.51×10^0	1.43×10^0	2.19×10^0	7.7×10^0	3.07×10^1
	K factor adjusted	1.75×10^1	1.15×10^1	2.15×10^0	3.37×10^0	9.16×10^0	4.37×10^1
Future Circular Collider	Cross-section (LO) (pb)	0.09731	707	1.127×10^5	15.72	4.062×10^{-4}	
	Events w/ loose kinematic cuts	2.91×10^6	2.12×10^{10}	3.38×10^{12}	4.72×10^8	1.22×10^4	3.040×10^{12}
	+ tagging and efficiencies	1.10×10^6	1.83×10^6	7.24×10^4	2.54×10^5	4.61×10^3	3.26×10^6
	+ kinematic cuts	4.37×10^5	4.86×10^5	2.33×10^4	6.86×10^4	3.00×10^3	1.02×10^6
	$+m_h$ windows	1.47×10^4	6.57×10^3	1.10×10^3	3.98×10^3	1.50×10^3	2.79×10^4
	+ angular cuts	1.01×10^3	8.38×10^2	2.73×10^2	2.94×10^2	5.92×10^2	3.01×10^3
	K factor adjusted	1.37×10^3	1.47×10^3	4.10×10^2	3.82×10^2	7.04×10^2	4.34×10^3

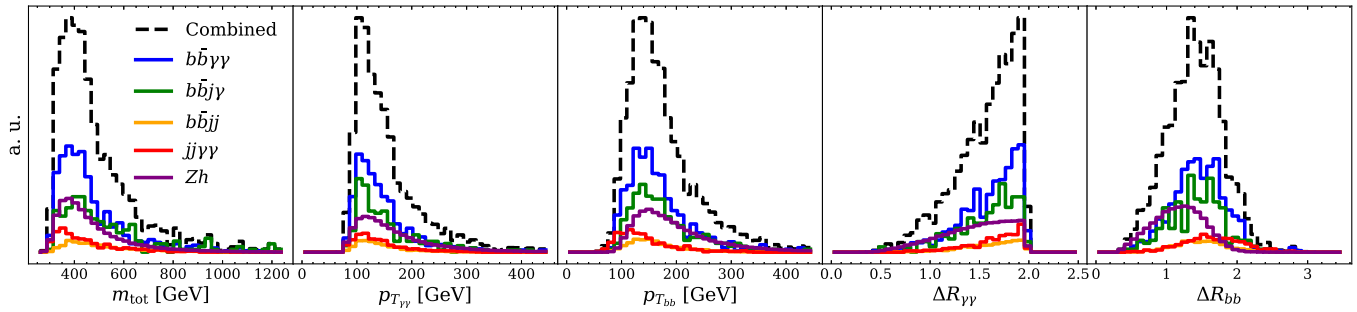

FIG. 3. Kinematic background distributions used in the shape analysis. We only show the HL-LHC predictions. The dashed line shows the total background from considering all processes in Table III, while the solid lines represent each individual process, scaled to their level of contribution. The vertical axis is linearly scaled. We defer definition and motivation of these observables to Sec. III D.

Sec. III A, we are left with 290,000 (183,000) background events.

According to Table III, the $b\bar{b}\gamma\gamma$ continuum and Zh backgrounds contribute the most to the total background, followed by the $b\bar{b}j\gamma$ process. The $b\bar{b}jj$ background has by far the largest cross section but is effectively reduced by the low $j \to \gamma$ mistag rate and the m_h windows. The $b\bar{b}h$ and $c\bar{c}h$ backgrounds with their moderate cross sections, 0.012 and 0.040 pb at the HL-LHC and 0.22 and 0.34 pb at 100 TeV, respectively, are completely removed by the m_h window requirement and thus omitted from Table III. To simplify the analysis, we can look at the kinematic distributions of the combined and component backgrounds in Fig. 3. We see that the shapes of the full background distributions are very close to that those of the two most contributing processes $b\bar{b}\gamma\gamma$ and $b\bar{b}j\gamma$ (which, in turn, look similar to each other), justifying the computationally motivated choice to only generate continuum $b\bar{b}\gamma\gamma$ events for the shape analysis background.

We additionally make the assumption that none of the continuum backgrounds is significantly affected by SMEFT modifications. While this is not exactly true, we have checked the effect of $\mathcal{O}_{t\phi}$ on the Zh background and find that the change in the background prediction is negligible compared with that of the hh signal. Therefore, we do not consider SMEFT modifications to backgrounds in this analysis.

When comparing the number of expected background events with the number of expected signal events in Table II, we see that $S/B \sim 0.30$ at the HL-LHC and $S/B \sim 0.83$ at 100 TeV. Without the hard angular cuts, these would be 0.036 and 0.098, respectively. These values, as well as the background yields in Table III, are in agreement with Ref. [24].

Finally, we acknowledge that our treatment of the continuum and mistag backgrounds are simplified and require more detailed studies with full simulations and/or data for improved accuracy. In an actual analysis, these backgrounds would likely be estimated from data [88,89]. We prioritize showing a proof of concept of shape information constraining SMEFT coefficients, rather than a highly realistic collider analysis.

C. Morphing through parameter space

Parameter estimation with simulation-based inference compares data with simulations assuming different parameter values describing the underlying physics. Methods differ in how the real and synthetic data are compared (see Sec. IV), but they must all cover the full parameter space.

Generating new events at each parameter point is computationally prohibitive. Instead, one reuses events generated at one parameter point by leveraging the factorization between the parton-level physics governing Higgs production and decay and the long(er)-distance-scale physics governing fragmentation and subsequent simulation steps. This scale separation implies that the likelihood ratio between two parameter points is fully covered by an exchange of partonic matrix elements. We can reweight an event from one parameter value to another and in this way morph the set events for one parameter value to the set for another parameter point.

This idea is implemented in MadMiner [40]. In particular, MadMiner expresses the matrix elements for a given process as a polynomial over chosen Wilson coefficients, simulates events at a number of parameter points comprising the "morphing basis," and fits the polynomial to the morphing basis in to quickly generate event weights at an arbitrary parameter point. This works well when the new parameter point is close to the original one; when moving too far away, the weights can be far from unity, and the statistical power of the weighted sample is diminished [90]. An alternative, future direction could be to use differentiable simulations instead of surrogate models [91–93].

In our case, the model parameter space is composed of three-dimensional vectors

$$c = (c_\phi, c_{\phi d}, c_{t\phi}) \in \mathbb{R}^3. \tag{13}$$

For these three Wilson coefficients, we define a polynomial up to squared terms, leading to 10 fit parameters, which is

TABLE V. Morphing basis points, from MadMiner. This basis minimizes the sum of the squared weights out of a set of 10^7 random bases.

Generation point	c_ϕ	$c_{\phi d}$	$c_{t\phi}$
1 (SM)	0	0	0
2	5.710	0.354	4.604
3	−5.873	0.817	7.124
4	1.135	3.664	−2.754
5	−12.638	3.035	−2.288
6	1.281	−4.792	1.188
7	−15.854	−1.261	−1.477
8	−15.107	5.670	7.668
9	−5.265	4.612	5.967
10	0.221	−0.006	−5.613

also the standard procedure in global SMEFT analyses [46]. In our case, the corresponding 10 basis points are given in Table V. This set is chosen by MadMiner by minimizing the sum of the squared weights from 10^7 random instantiations. The locations of the generation points and the corresponding squared morphing weights are shown in Fig. 4. To generate the background scale, we scan over a dense grid of points c, compute the weights needed to convert from that point to each of the 10 benchmarks w_i, and display the sum $\sqrt{\sum_i w_i^2}$. Since higher weights are associated with larger uncertainties on the morphing basis, it is encouraging that the majority of the computed weights are $\lesssim 1$, or $\sqrt{\sum w_i^2} \lesssim \sqrt{10}$.

D. Kinematic observables

To study the main differences between the SM Higgs signal, the SMEFT effects, and the continuum backgrounds, we define a set of $N = 5$ observables, for which the sensitivity to the Higgs self-couplings and SMEFT effects is well known: (i) the reconstructed di-Higgs mass, (ii) the reconstructed transverse momentum of the photonically decaying Higgs, (iii) the reconstructed transverse momentum of the hadronically decaying Higgs, (iv) the the angular separation between the two photons, and (v) the angular separation between the two b jets from the Higgs decay,

$$N = 5 \qquad \{m_{\text{tot}}, p_{T_{\gamma\gamma}}, p_{T_{bb}}, \Delta R_{\gamma\gamma}, \Delta R_{bb}\}. \tag{14}$$

The reconstructed di-Higgs mass combines sensitivity to threshold cancellations with a test of the Higgs kinematics [12,24], and the Higgs transverse momenta test the momentum dependence of the production process and the top-loop threshold [58]. The angular separations are strongly correlated to the transverse momenta, but only for the signal and not for the continuum backgrounds.

All five observables are shown for the SM Higgs signal, two choices of the Wilson coefficient c_ϕ, and the continuum background in Fig. 5. The most striking feature is that the Higgs signals and the continuum background look very different, where the backgrounds contain much less energy per event and the reconstructed Higgs decays products are widely separated. An analysis searching for deviations from the SM hh production and for deviations from the SM signal and the continuum background will therefore be quite different [34].

IV. ANALYSES

Independent of the representation of our data D, the analysis goal is to determine if the data are more consistent with the SM prediction $c = 0$ or some finite values for the Wilson coefficients $c = (c_\phi, c_{\phi d}, c_{t\phi})$ defined in Eq. (13). The data can either be events or bins of an N-dimensional observable, $D = \{x_i\}_{i=1}^k$. For a fixed BSM model c, the Neyman-Pearson Lemma [94] states that the most powerful test statistic is the likelihood ratio mentioned already in

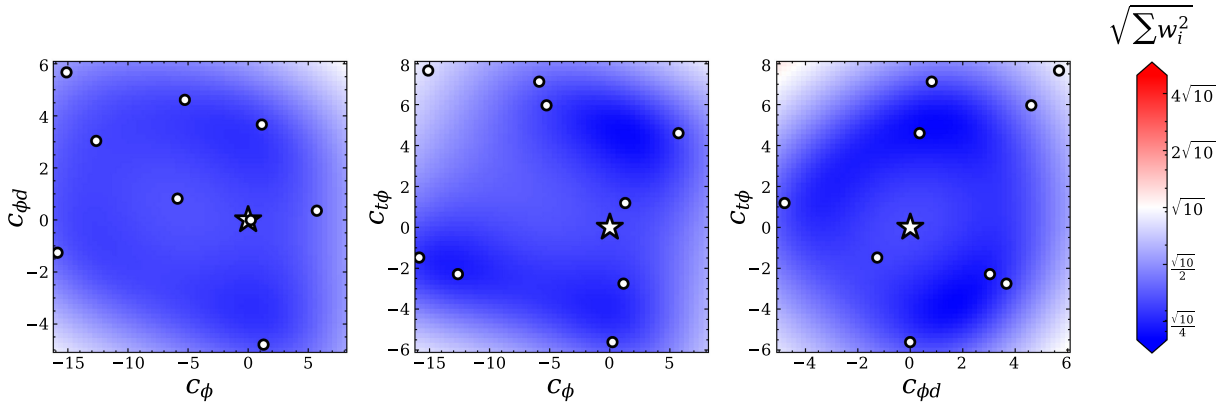

FIG. 4. Squared morphing weights over the chosen parameter ranges for the set of Wilson coefficients $c = (c_\phi, c_{\phi d}, c_{t\phi})$. The SM point is denoted by a star, and the other nine benchmarks given in Table V are denoted by circles. At each parameter point c, the weights needed to convert from that point to each of the 10 benchmarks are computed; the squared sum is then displayed. For numerical stability of the reweighting process, it is desirable that $\sqrt{\sum w_i^2} \lesssim \sqrt{10}$.

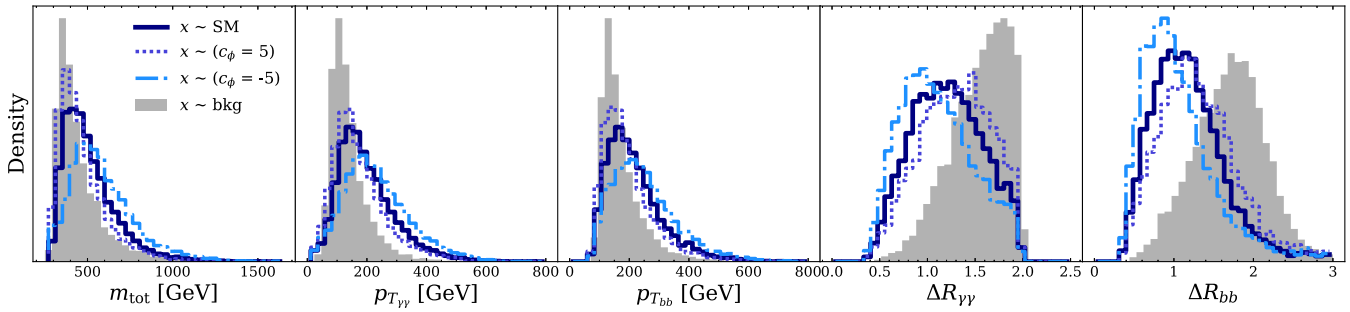

FIG. 5. Kinematic signal and background distributions, including two choices of c_ϕ. We only show HL-LHC predictions. The vertical axis is linearly scaled.

$$q(c|D) = -2\log\frac{p(D|c)}{p(D|c=0)}, \tag{15}$$

where the factor of 2 and natural logarithm imply that in the Gaussian approximation changes in q by one unit correspond to one standard deviation for a one-dimensional c.

If a production cross section strongly depends on a set of model parameters, a natural, but by no means optimal, first step is a rate-only analysis. This is the current approach used by CMS [32] and ATLAS [33]. This case corresponds to $N = 0$, and the probability density for D simplifies to a comparison of Poisson distributions for a given number of k events,

$$\begin{aligned} q_{\text{rate}}(c|D) &= -2\log\left(\frac{\text{Pois}(k|c)}{\text{Pois}(k|c=0)}\right) \\ &= -2\left(\bar{k}(0) - \bar{k}(c) + k\ln\left(\frac{\bar{k}(c)}{\bar{k}(0)}\right)\right). \end{aligned} \tag{16}$$

where $\bar{k}(c) = \sigma_{\text{tot}}(c) \times L$ is the predicted number of events for a given parameters c.

A. Shape analysis with classifiers

Collider events are statistically independent, which means that the full log-likelihood ratio factorizes into a rate term and a shape term, where the shape term is a sum over events or observable bins,

$$\begin{aligned} q(c|D) &= q_{\text{rate}}(c|D) - 2\sum_{i=1}^{k}\log\frac{p(x_i|c)}{p(x_i|c=0)} \\ &\equiv q_{c,\text{rate}}(c|D) + q_{c,\text{shape}}(c|D). \end{aligned} \tag{17}$$

We explore various scenarios, with N up to 5. The corresponding observables are given in Eq. (14), and their histograms are presented in Fig. 5. For our analysis, we use subsets of the five observables, namely,

$$\begin{aligned} N = 1 \quad & \{m_{\text{tot}}\} \\ N = 3 \quad & \{m_{\text{tot}}, p_{T_{\gamma\gamma}}, p_{T_{bb}}\}. \end{aligned} \tag{18}$$

Unlike in the rate-only case, we do not know $p(x_i|c)$ explicitly. To estimate the per-event likelihood ratio, we use the fact that trained classifiers $C(x) \in [0, 1]$ learn a monotonic function of the density ratio (see, e.g., Refs. [95–97]). A calibrated classifier then becomes

$$\frac{C(x)}{1 - C(x)} \approx \frac{p(x|c)}{p(x|c=0)}, \tag{19}$$

and the baseline configuration is known to work well [98]. Our key assumption here is that we can sample accurately and precisely from $p(x|c)$. Here, the simulation-based inference represents an ideal that may be achievable completely or partially by the time such an analysis is performed.

In our case, we need to promote the likelihood ratio to be a function of c. This is accomplished by training a parametrized classifier [97,99] where c is promoted to a feature. This means that we train a classifier C acting on (x, c). For training, c is drawn from a prior $p_0(c)$, and then signal events are drawn from $p(x|c)$. For the background events, we assign values c so that the marginal distribution of c is p_0. In this way, c is not useful for the classifier, and

$$\frac{C(x,c)}{1 - C(x,c)} \approx \frac{p_{\text{BSM}}(x,c)}{p_{\text{SM}}(x,c)} = \frac{p(x|c)}{p(x|0)}. \tag{20}$$

This looks the same as Eq. (19), but now it is a continuous function of c.

In practice, we can further simplify the construction of q if we assume that we can neglect quantum interference between the hh signal and the continuum background. For our specific analysis, this is ensured by the small Higgs width, which suppresses all interference contributions. In that case, $p(x|c)$ can be approximated by a mixture model

$$p(x|c) \approx \mu(c)p_S(x|c) + (1 - \mu(c))p_B(x), \tag{21}$$

where the relative proportion $\mu(c)$ of hh to the total number of events is known from Sec. III C, and $p_{S,B}$ are the corresponding probability densities of the hh and the continuum events. As noted above, we have verified that

for the Wilson coefficients considered the variation of p_B with c is negligible. Thus, we can rewrite the log-likelihood ratio in terms of p_S and p_B [97],

$$\frac{p(x|c)}{p(x|0)} = \left[\frac{\mu(0)p_S(x|0)}{\mu(c)p_S(x|c)} + \frac{(1-\mu(0))p_B(x)}{\mu(c)p_S(x|c)}\right]^{-1} + \left[\frac{\mu(0)p_S(x|0)}{(1-\mu(c))p_B(x)} + \frac{1-\mu(0)}{1-\mu(c)}\right]^{-1}. \quad (22)$$

While this expanded form may look complicated, it allows us to break down the problem into three easier problems. In particular, the first three terms in Eq. (22) can be approximated with classifiers that each have an easier task than distinguishing samples drawn from $p(x|c)$ and $p(x|0)$ all at once:

(1) $\frac{p_S(x|0)}{p_S(x|c)}$ is learned by a parametrized classifier distinguishing SM from SMEFT hh events.
(2) $\frac{p_B(x)}{p_S(x|c)}$ is learned by a parametrized classifier distinguishing background from SMEFT hh events.
(3) $\frac{p_S(x|0)}{p_B(x)}$ is learned by a nonparametrized classifier distinguishing SM hh events from continuum background.

B. Training specifications

Event generation for the parameterized classifier is done within the MadMiner framework, whose morphing feature allows for quick generation of events at arbitrary parameter points c.

For classifier training, we generate sets of 10^7 events each for the SM hh signal, the BSM hh signal, and the continuum background. For the BSM sample, we generate the events by uniformly sampling 1000 values of c, which means p_0 in Sec. IV A is uniform. For the one-dimensional (1D) coefficient tests which will be shown in Fig. 6, we only allow the single scanned Wilson coefficient to be nonzero; for the two-dimensional (2D) coefficient tests, which will be shown in Fig. 7, only the two scanned Wilson coefficients are nonzero. The nonzero coefficients cover the prior ranges $c_\phi \in (-14, +4)$, $c_{\phi d} \in (-4, +5)$, and $c_{t\phi} \in (-5, +7)$. For the test sets, we generate a set of events following the expected event yields from Table II.

All classifiers are parametrized as relatively small, dense neural networks consisting of two layers with 32 hidden nodes. We use a batch size of 1024, a weight decay 10^{-4}, and an initial learning rate of 10^{-3}. The learning rate reduces by half if the validation loss does not decrease for five epochs. We train for up to 1600 epochs, stopping when the validation loss does not decrease for 20 epochs and evaluating the networks at the epoch of lowest validation loss. In practice, the classifiers trained on data for the 100 TeV collider often converged in fewer than 200 epochs. We use an 80:20 training-validation split. All networks are implemented in PyTorch [100] and optimized with ADAM [101]. All hyperparameters are optimized by manual tuning on a simplified version of the problem. This simplified problem refers to carrying out the 1D coefficient tests on pre-DELPHES samples in the zero-background case (i.e., we just train the classifier that discriminates SM hh signal from BSM hh signal). Performance was fairly robust with respect to classifier architecture and training hyperparamters, although we did find that a longer early stopping parameter produced better results.

To mitigate the stochastic nature of the network training, we ensemble the outputs of five networks with identical architectures and different initial random number generator seeds.

V. RESULTS

A. Coefficient recovery for the SM

We first assume that the observed data corresponds to the SM hypothesis. Using the pretrained network from the previous section, we compute the log-likelihood ratio for a reference dataset with all Wilson coefficients c_ϕ, $c_{\phi d}$, and $c_{t\phi}$ set to zero.

A 1D scan over one Wilson coefficient at a time is shown in Fig. 6 for both the HL-LHC and the 100 TeV collider setups. The resulting central parameter value is given by the maximum log likelihood, and the confidence intervals are determined based on the shape around the minimum. These uncertainties are indicated in Fig. 6 by vertical dashed lines as 1σ confidence intervals for the HL-LHC (so the y axis decreases by roughly one unit from the minimum) and 3σ for 100 TeV.

We consider three analysis methods: a rate-only analysis, one that incorporates shape information from the m_{hh} only, and one that incorporates shape information from the m_{hh}, $p_{T_{bb}}$, and $p_{T_{\gamma\gamma}}$ kinematic distributions. We provide equivalent plots for the one-dimensional test statistics in the Appendix for all five observables in Fig. 12, also including ΔR_{bb} and $\Delta R_{\gamma\gamma}$.

Starting with single Wilson coefficients in Fig. 6, all likelihood minima are consistent with zero within the reported uncertainty, and adding more kinematic information generally sharpens the peaks and leads to smaller uncertainty around the SM minimum. Away from the SM minimum, the additional but incomplete kinematic information can lead to features in the likelihood ratio dependence. Both effects are especially prominent for the classic trilinear Higgs coupling c_ϕ—in the rate-only analysis, there remains a degeneracy in the test statistic at $c_\phi = 0$ and $c_\phi = -11$, which is resolved by incorporating shape information. As a matter of fact, the rate-only analysis leaves a degeneracy for all three Wilson coefficients, but this degeneracy is effectively resolved by including single-Higgs production.

The fundamental assumption of effective field theories is that any high-energy model will induce all Wilson coefficients compatible with its symmetry. This means any

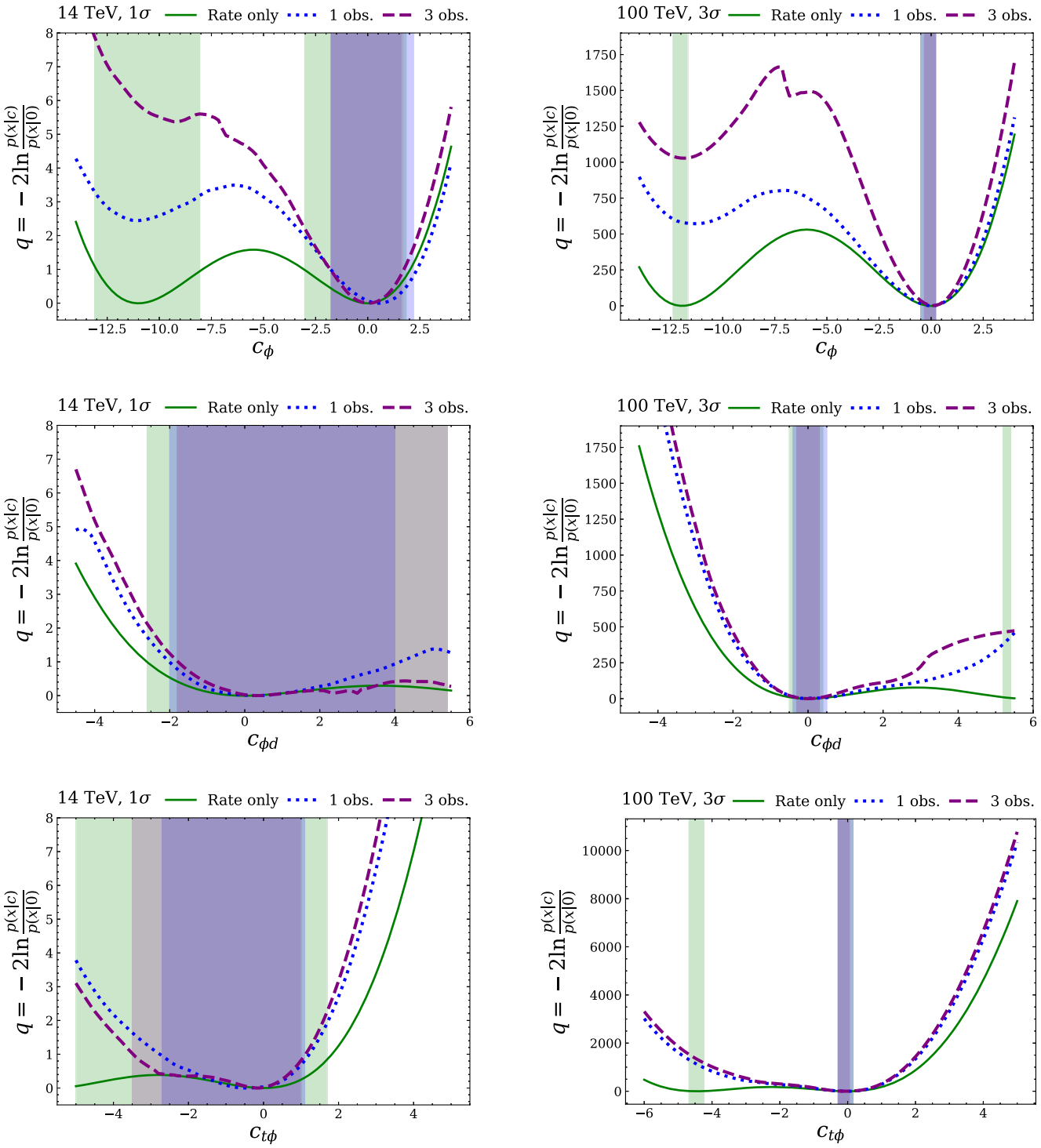

FIG. 6. Log-likelihood ratio test statistics in terms of 1σ (for HL-LHC) or 3σ (for 100 TeV) confidence intervals for one nonzero Wilson coefficient at a time. Data generation and test set size reflect the collider setup.

LHC signal will be affected by more than one operator, and the correlations between contributions from different operators will reflect the underlying theory. This motivates the variation of two Wilson coefficients at a time. The corresponding log likelihoods for the rate-only, one-observable, and three-observable test statistics are shown in Fig. 7. Equivalent plots for the five-observable test statistic are shown in the Appendix in Fig. 13. We overlay the 1σ and 3σ confidence regions for the HL-LHC and 100 TeV colliders, respectively. For the rate-only analysis, the double peak distribution from Fig. 6 now becomes an ellipse or annulus. For the statistically stable 100 TeV

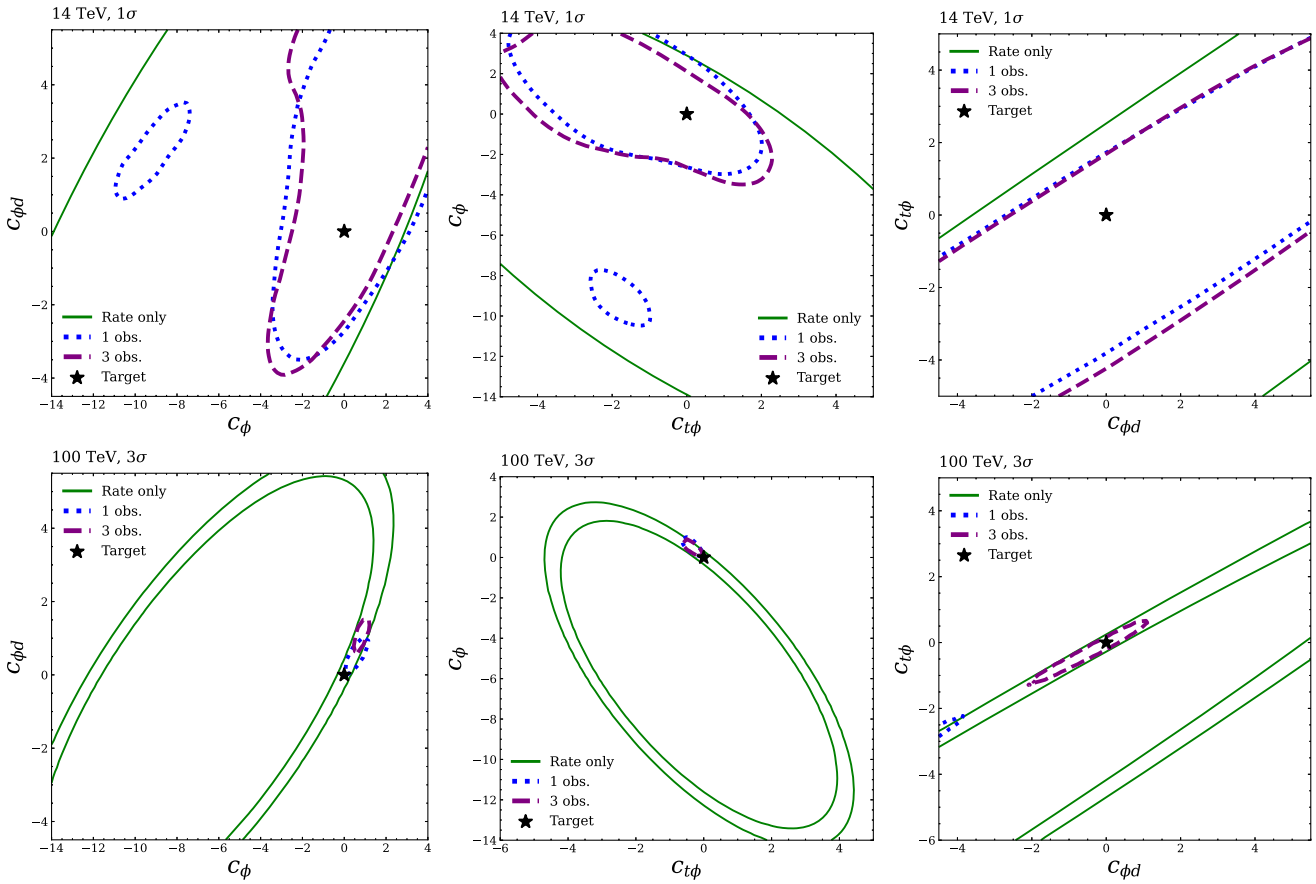

FIG. 7. Log-likelihood ratio test statistics in terms of 1σ (for HL-LHC) or 3σ (for 100 TeV) confidence contours for two nonzero Wilson coefficients at a time. Data generation and test set size reflect the collider setup.

setup, adding kinematic information can indeed break the degeneracy and improve the confidence contours to small regions on the correct side of the rate-only ellipse. While we show confidence contours for only a single realization of nature, the fact that the SM is not always contained within the confidence region for the 100 TeV machine is representative. With the high statistics of the 100 TeV machine, the precision required on the likelihood ratio estimation is much stricter than for 14 TeV. The confidence regions are qualitatively in the correct location, but achieving quantitatively precise results will require additional research (see Sec. IV A for some progress in other studies).

Going beyond two dimensions, it is difficult to visualize the full space. Since the neural networks are differentiable, it is possible to find the maximum likelihood estimate using gradient descent, and the Hessian can provide an estimate of the confidence interval. This is left to future studies to explore in more detail.

B. Coefficient recovery for BSM scenarios

To explore the sensitivity away from the SM value of $c = 0$, we also show the performance of recovering a nonzero c-vector. As a comprehensive scan of the accuracy and precision of the various approaches would be computationally demanding and difficult to visualize, we pick representative examples to study in detail.

In Fig. 8, we show the expected HL-LHC limits for the single Wilson coefficient c_ϕ for assumed true values away from the SM. The almost-perfect cancellation of the triangle and box diagrams is only true for a SM self-coupling. This means that, while the hh rate will increase for these points, the sensitivity will not be enhanced as much through the cancellation. Altogether, we find that the 1σ range for the three-observable analysis increases to

$$\begin{aligned} c_\phi = 0 \quad &\text{assumed}: \quad c_\phi \in [-1.8, +1.7] \\ c_\phi = -4 \quad &\text{assumed}: \quad c_\phi \in [-6.3, -1.2] \\ c_\phi = -6 \quad &\text{assumed}: \quad c_\phi \in [-6.8, -1.8]. \end{aligned} \tag{23}$$

In the HEFT basis assuming an EFT cutoff of 1 TeV and setting all other Wilson coefficients to zero, this is equivalent to

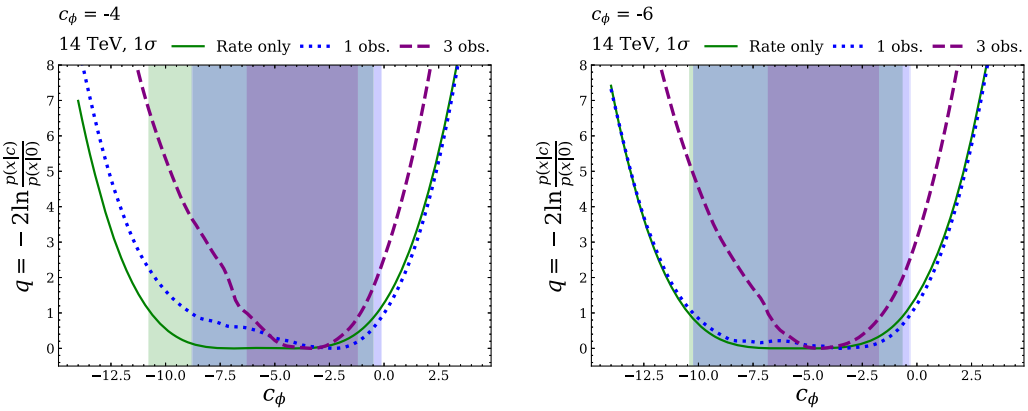

FIG. 8. Log-likelihood ratio test statistics in terms of 1σ for the HL-LHC, for one Wilson coefficient c_ϕ at a time. The central value of c_ϕ is chosen away from the SM. Data generation and test set size reflect the collider setup.

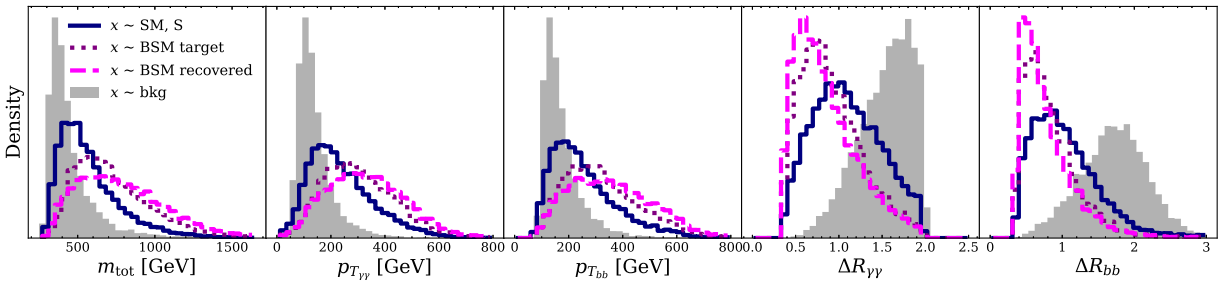

FIG. 9. Kinematic distributions from $b\bar{b}\gamma\gamma$ events from the 100 TeV collider setup. A test set generated with the Wilson coefficients $(c_\phi, c_{\phi d}, c_{t\phi}) = (-4, 0, 3)$, "BSM target," is close to degenerate in classifier output to a test set generated with $(0, 0, -5.5)$, "BSM recovered." The vertical axis is linearly scaled.

$$\begin{aligned} &\kappa_\lambda = +1 \quad \text{assumed}: \quad \kappa_\lambda \in [+0.2, +1.8], \\ &\kappa_\lambda = +2.9 \quad \text{assumed}: \quad \kappa_\lambda \in [+1.6, +4.0], \\ &\kappa_\lambda = +3.8 \quad \text{assumed}: \quad \kappa_\lambda \in [+1.8, +4.2]. \end{aligned} \tag{24}$$

For both choices, the rate-only measurement would not be able to distinguish these parameter points from the SM.

Second, we can scan the likelihoods landscape for a given set of events, for instance, at the 100 TeV collider, and identify parameter points far away in model space but close in likelihood. An example is the vectors of Wilson coefficients

$$\begin{aligned} &(c_\phi, c_{\phi d}, c_{t\phi}) = (0, 0, -5.5), \\ &(c_\phi, c_{\phi d}, c_{t\phi}) = (-4, 0, 3), \end{aligned} \tag{25}$$

which are indistinguishable from each other in a rate-only analysis and very similar in terms of our observables. In Fig. 9, we see that the observables are clearly distinguishable from the SM, given the high assumed statistics of the 100 TeV collider, but not from each other. This similarity is reflected in a shape analysis, seen in Fig. 10. Given a test set with the generating Wilson coefficient vector $c = (-4, 0, 3)$, the three-observable test statistic is doubly minimized, showing high likelihood for both the true underlying vector c and the similarly shaped $c = (0, 0, -5.5)$. The kinematic observables m_{tot}, $p_{T_{\gamma\gamma}}$, and $p_{T_{bb}}$ are indeed very similar to each other, especially when compared to the SM distributions; this similarity leads to the likelihood ratio degeneracy, which is somewhat broken when considering $\Delta R_{\gamma\gamma}$, which peaks in a distinct location for the true c and recovered c vectors. In fact, the five-observable classifier is able to recover values of c that are much closer to truth.

C. Statistical properties and coverage

In Fig. 7, we see that for a single realization of nature the confidence contours are not always ellipses and the SM is not always contained within the confidence region. This first of these is addressed with the higher statistics of the 100 TeV machine, while the latter issue is more acute in this setting. The problem is that for the HL-LHC there are around 100 total events expected, with around 10 of them from the hh signal.

This statistics limitation leads us to explore the accuracy and precision of the various parameter estimation approaches by computing confidence intervals for many

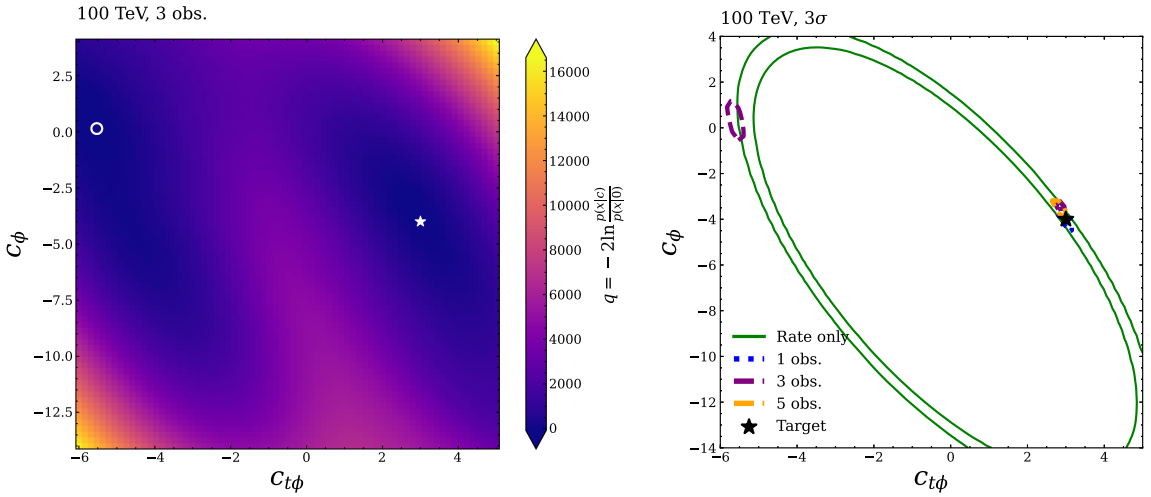

FIG. 10. Left: 2D test statistic for the three-observable test. The generating Wilson coefficient vector $c = (-4, 0, 3)$ is denoted by a star; the test statistic is optimized at $c \approx (0, 0, -5.5)$, which is denoted by an open circle. Right: 3σ confidence intervals for a variety of test statistics.

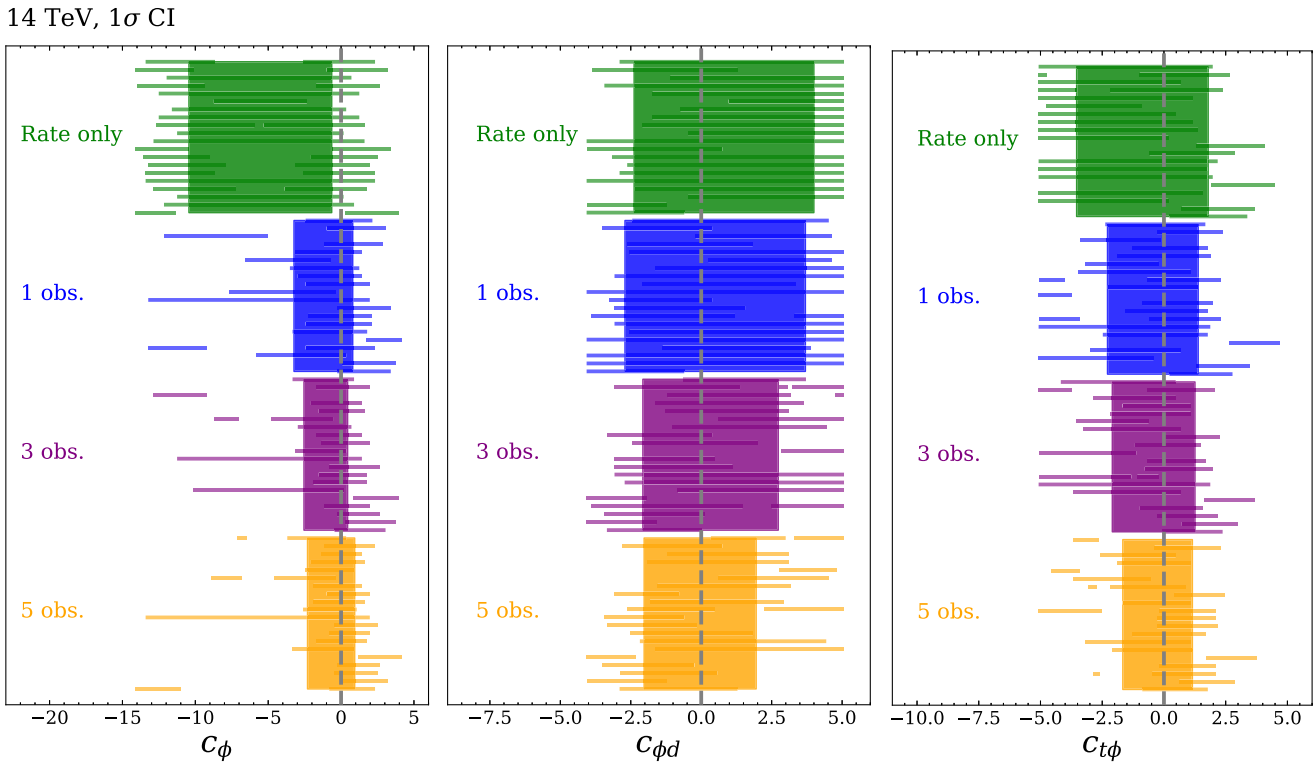

FIG. 11. Confidence intervals for log-likelihood ratio associated with a given Wilson coefficient. Data generation corresponds to the HL-LHC. Each horizontal band corresponds to a different event set. The shaded band shows the mean confidence interval for each test.

synthetic datasets. Analyzing different datasets is of particular interest for the HL-LHC, where there are only tens of events after the event selection. This is why we expect that the confidence intervals for the Wilson coefficients might change significantly between different tested event sets.

For ease of visualization, we focus on the 1D confidence intervals testing one nonzero Wilson coefficient at a time. In Fig. 11, we show expected 1σ confidence intervals for each of our Wilson coefficients for the HL-LHC. We use 20 different test event sets corresponding to the SM hypothesis, but with varying signal and background events.

For each test set, we allow the sizes of the signal and background sets to vary, following Poisson distributions with means given by the last row of Table II.

The shaded bars in Fig. 11 denote the mean coverage interval across the test sets, for each of the analysis techniques: rate only, one observable, and three observables. We also include results from a five-observable analysis strategy, adding ΔR_{bb} and $\Delta R_{\gamma\gamma}$.

Looking at the c_ϕ recovery at the HL-LHC, we see that for the rate-only analysis $(90 \pm 7)\%$ of the confidence intervals contain the correct value. For the one-observable analysis, $(80 \pm 9)\%$ contains the correct value; for the three-observable analysis, $(80 \pm 9)\%$ contains the correct value; and for the five-observable analysis, $(85 \pm 8)\%$ contains the correct value. For standard 1σ confidence intervals, we would expect around 68% of the intervals to give SM recovery. However, we almost always find that more than this percentage of the confidence intervals contain the SM, so our bounds are conservative. Given that for the rate-only analysis the likelihood ratio is exact (i.e., not estimated with neural networks), this may be due to the non-Gaussian nature of the test statistic.

A similar recovery is also achieved for the other two Wilson coefficients: for $c_{\phi d}$, the confidence intervals for the rate-only, one-observable, three-observable, and five-observable analyses return the SM value $(85 \pm 8)\%$, $(90 \pm 7)\%$, $(80 \pm 9)\%$, and $(55 \pm 11)\%$ of the time. For $c_{t\phi}$, the same numbers are $(75 \pm 10)\%$, $(65 \pm 11)\%$, $(70 \pm 10)\%$, and $(70 \pm 10)\%$ of the time.

A corresponding analysis can be carried out for the 100 TeV setup, where we evaluate 1D confidence intervals for 20 independent test sets. For c_ϕ recovery, all 3σ confidence intervals for all test statistics contain the SM value; for $c_{\phi d}$, the one-observable test statistic achieves SM recovery $(95 \pm 5)\%$ of the time and the other test statistics 100% of the time; for $c_{t\phi}$, the rate-only test statistic returns the SM value 100%, the one-observable test statistic returns it $(95 \pm 5)\%$, and the three- and five-observable test statistics return it $(90 \pm 7)\%$ of the time.

We may further contrast the test statistic types. The mean confidence interval is narrower when derived from the three-observable test statistic than when derived from the one-observable test statistic, and the mean one-observable test statistic is narrower than that of the rate-only analysis. Further, the test statistics that make use of kinematic observable information more often resolve the likelihood degeneracy seen in the c_ϕ recovery that is left ambiguous for the rate-only test statistic. This is consistent with our earlier findings that including shape information for kinematic distributions can place tighter constraints on Wilson coefficient bounds than rate-only analyses can. In addition to the known m_{tot}, the set of three observables is informative and relevant for the coefficients c_ϕ, $c_{\phi d}$, and $c_{t\phi}$.

Finally, it is worth noting that for the HL-LHC some test event sets do not resolve the degeneracy for the c_ϕ-coefficient or resolve it incorrectly by choosing a large negative value for this coefficient. The degeneracy is always correctly resolved, to 3σ, for the 100 TeV collider.

VI. CONCLUSIONS

In this work, we have explored the use of neural simulation-based inference to enhance the sensitivity to searches for pair production of Higgs bosons. For our example, we have simulated an analysis to place constraints on the SMEFT Wilson coefficients for a set of three dimension-6 operators associated with hh production: c_ϕ, $c_{\phi d}$, and $c_{t\phi}$. We have considered two collider setups in this report: a HL-HLC-like setup with $\sqrt{s} = 14$ TeV and 3 ab^{-1} of integrated luminosity and a future hadron collider setup with $\sqrt{s} = 100$ TeV to 30 ab^{-1}. We have shown that through parametrized machine learning tools we can augment more "standard" cut-and-count analyses with per-event shape information to increase constraining power for these hh-relevant Wilson coefficients.

In the idealized context of our study, we encountered a number of challenges that need to be addressed before these methods can be used in practice. Most importantly, it is difficult to achieve the level of precision required to produce accurate and precise confidence regions near the global minimum of the likelihood landscape. We have utilized a number of techniques to address this, such as factorizing the classifiers and using ensembling. It would be interesting to explore additional proposals for improving the likelihood-ratio estimation, and we hypothesize that additional methods are required, especially for the level of precision that will be afforded by future high-luminosity collider data.

There are a number of steps required to connect our idealized study to a real-life data analysis. In particular, we assumed that simulation will be used to estimate the background. It may be that this will be possible in the HL-LHC era, but the current state of the art is data-driven background estimates. It may be possible to combine such approaches with NSBI, which would be interesting to explore in the future. Additionally, we assumed that the signal and background are known with no systematic uncertainty. In practice, such uncertainties can be directly folded into the analysis protocol, although profiling over a large number of nuisance parameters may be challenging.

Going beyond the analysis presented here, it would also be interesting to explore how far we could push the dimensionality of the observable space and the parameter space. There may also be gains possible from a dedicated study of the tradeoffs between making restrictive selections and using per-event information for more events. In particular, we could relax the preselection to reduce the

starting significance of the signal but then recover (and ideally, exceed) the sensitivity through the per-event likelihood estimation. This approach will be limited in part by the ability of the neural networks to describe very low likelihood events. While we have focused on *hh* events, the tools and challenges are common to many NSBI analyses, and our study provides another important benchmark for refining and developing new methods.

All data used in this report are available on Zenodo at [102]. The analysis code is available at [103].

ACKNOWLEDGMENTS

B. N. and R. M. are supported by the U.S. Department of Energy (DOE), Office of Science under Contract No. DE-AC02-05CH11231 and Grant No. 63038 from the John Templeton Foundation. R. M. is additionally supported by Grant No. DGE 2146752 from the National Science Foundation Graduate Research Fellowship Program. This research used resources of the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231 using NERSC award HEP-ERCAP0021099. T. P. is supported by the Baden-Württemberg-Stiftung through the program *Internationale Spitzenforschung*, project *Uncertainties—Teaching AI its Limits* (BWST_IF2020-010), the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Grant No. 396021762—TRR 257 *Particle Physics Phenomenology after the Higgs Discovery*, and through Germany's Excellence Strategy EXC 2181/1–390900948 (the *Heidelberg STRUCTURES Excellence Cluster*).

APPENDIX: ADDITIONAL PLOTS

In Fig. 12, we show the one-dimension test statistic and confidence intervals for the three-observable (m_{tot}, $p_{T_{bb}}$, $p_{T_{\gamma\gamma}}$) and five-observable test statistics ($+\Delta R_{bb}$, $\Delta R_{\gamma\gamma}$). This plot serves as an extension to Fig. 6. In Fig. 13, we show the same for the two-dimension test statistic and confidence internal, as an extension to Fig. 7. We generally find that including more observables allows for tighter constraints on the given Wilson coefficient—this is unambiguous for the 100 TeV collider setup, while the 14 TeV setup does suffer from limited statistics (see, in particular, the c_ϕ limit for Fig. 12).

In Fig. 14, we show two examples of 2D coefficient recovery for BSM test sets for the 100 TeV collider setup. All test statistics that make use of kinematic observables greatly reduce the confidence limit areas when compared to the rate-only analysis. However, the displacement of the recovered areas from zero demonstrates the challenges associated with the high-precision requirement of the 100 TeV collider.

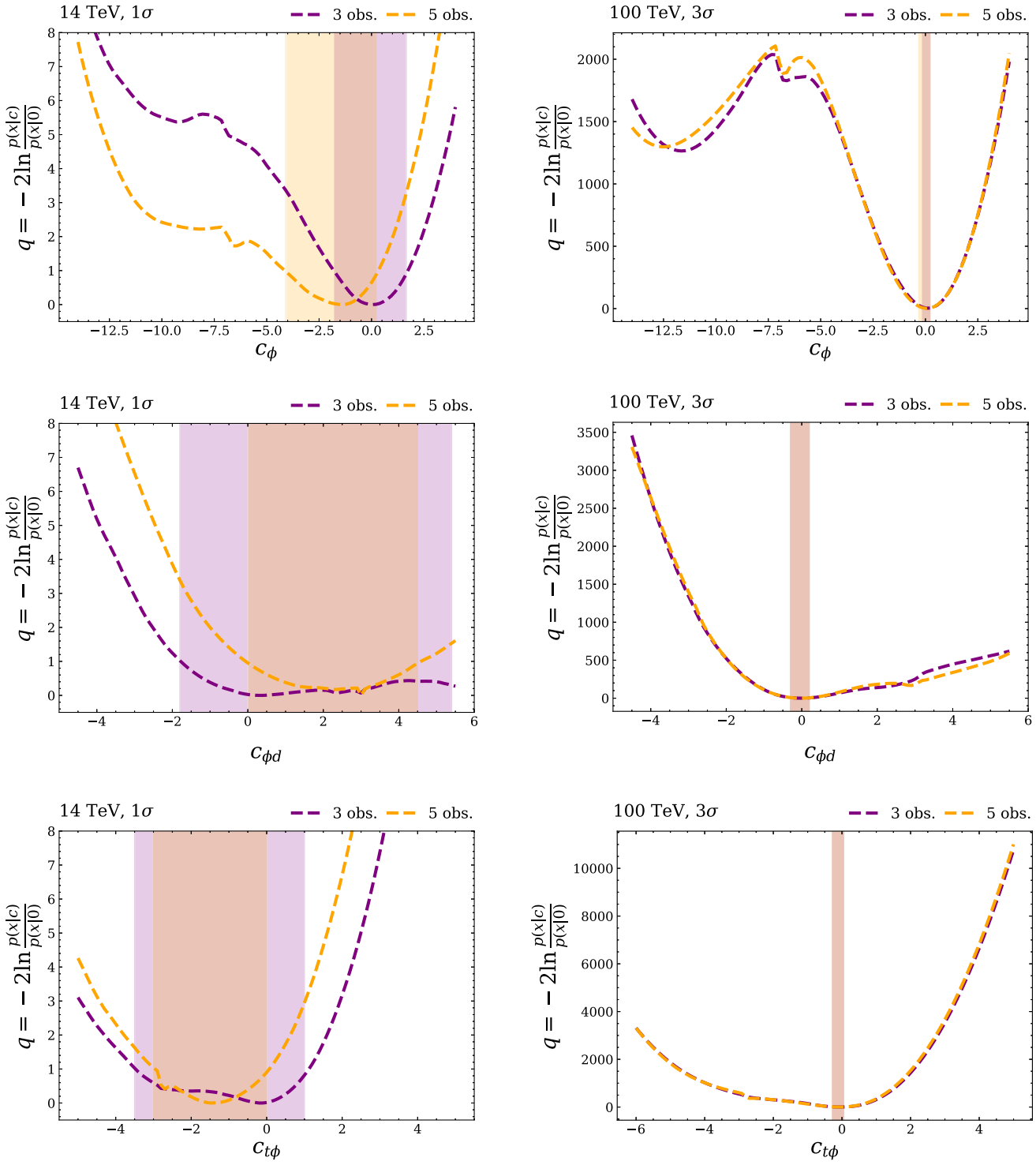

FIG. 12. Test statistics and 1σ (for HL-LHC) or 3σ (for 100 TeV) confidence intervals for the test statistic $q = -2\ln\frac{p(x|c)}{p(x|0)}$ associated with the given Wilson coefficient. Data generation and test set size reflect the labeled collider setup.

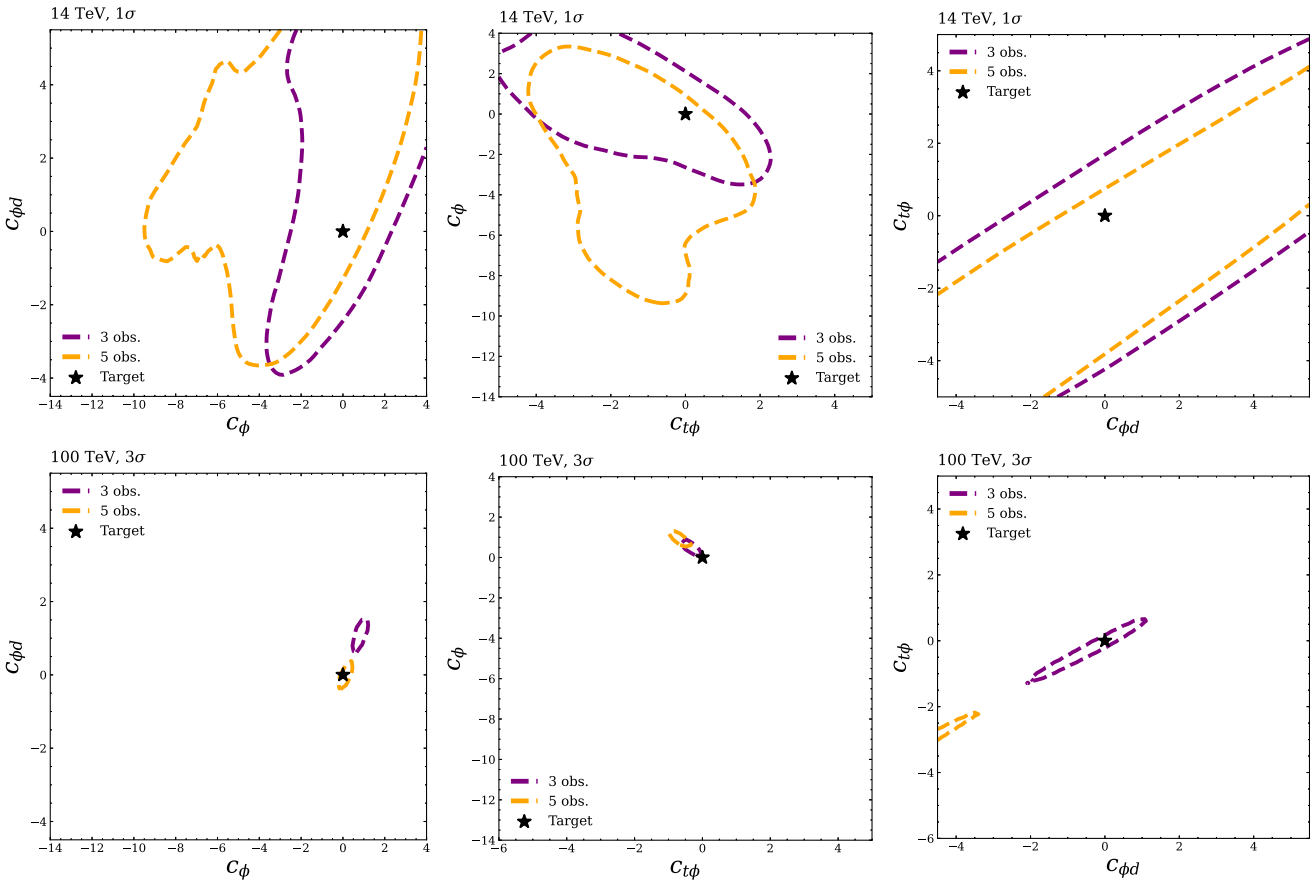

FIG. 13. Test statistics and 1σ (for HL-LHC) or 3σ (for 100 TeV) confidence contours for the test statistic $q = -2\ln\frac{p(x|c)}{p(x|0)}$ associated with the given Wilson coefficient. Data generation and test set size reflect the labeled collider setup.

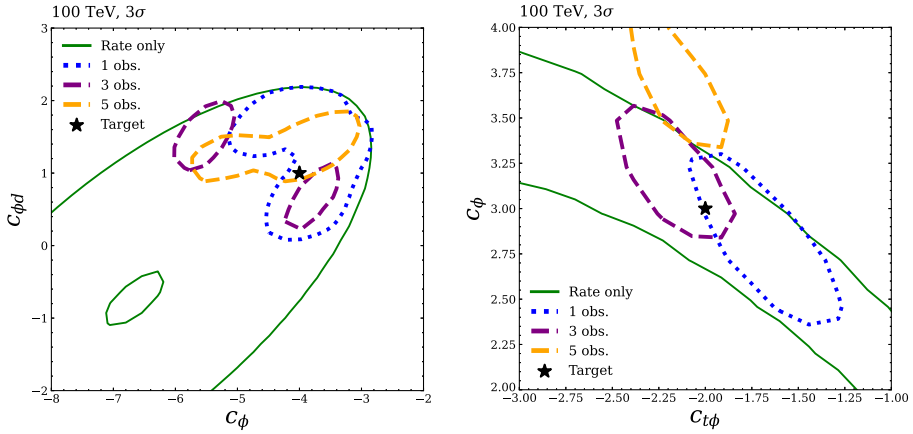

FIG. 14. Log-likelihood ratio test statistic in terms of 3σ confidence contours for two sets of Wilson coefficients with central values away from $c = 0$. The generating values are denoted by a star.

[1] G. Aad *et al.* (ATLAS Collaboration), Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC, Phys. Lett. B **716**, 1 (2012).

[2] S. Chatrchyan *et al.* (CMS Collaboration), Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC, Phys. Lett. B **716**, 30 (2012).

[3] G. Isidori, G. Ridolfi, and A. Strumia, On the metastability of the standard model vacuum, Nucl. Phys. **B609**, 387 (2001).

[4] G. Degrassi, S. Di Vita, J. Elias-Miro, J. R. Espinosa, G. F. Giudice, G. Isidori, and A. Strumia, Higgs mass and vacuum stability in the standard model at NNLO, J. High Energy Phys. 08 (2012) 098.

[5] A. Bednyakov, B. Kniehl, A. Pikelner, and O. Veretin, Stability of the electroweak vacuum: Gauge independence and advanced precision, Phys. Rev. Lett. **115**, 201802 (2015).

[6] C. Grojean, G. Servant, and J. D. Wells, First-order electroweak phase transition in the Standard Model with a low cutoff, Phys. Rev. D **71**, 036001 (2005).

[7] M. Reichert, A. Eichhorn, H. Gies, J. M. Pawlowski, T. Plehn, and M. M. Scherer, Probing baryogenesis through the Higgs boson self-coupling, Phys. Rev. D **97**, 075008 (2018).

[8] Anisha, L. Biermann, C. Englert, and M. Mühlleitner, Two Higgs doublets, effective interactions and a strong first-order electroweak phase transition, J. High Energy Phys. 08 (2022) 091.

[9] A. Biekötter, J. Fuentes-Martín, A. M. Galda, and M. Neubert, A global analysis of axion-like particle interactions using SMEFT fits, J. High Energy Phys. 09 (2023) 120.

[10] R. L. Workman *et al.* (Particle Data Group), Review of particle physics, Prog. Theor. Exp. Phys. **2022**, 083C01 (2022).

[11] S. Di Vita, C. Grojean, G. Panico, M. Riembau, and T. Vantalon, A global view on the Higgs self-coupling, J. High Energy Phys. 09 (2017) 069.

[12] U. Baur, T. Plehn, and D. L. Rainwater, Measuring the Higgs boson self coupling at the LHC and finite top mass matrix elements, Phys. Rev. Lett. **89**, 151801 (2002).

[13] U. Baur, T. Plehn, and D. L. Rainwater, Probing the Higgs self-coupling at hadron colliders using rare decays, Phys. Rev. D **69**, 053004 (2004).

[14] J. Chang, K. Cheung, J. S. Lee, and J. Park, Probing the trilinear Higgs boson self-coupling at the high-luminosity LHC via multivariate analysis, Phys. Rev. D **101**, 016004 (2020).

[15] I. Zurbano Fernandez *et al.*, High-luminosity large hadron collider (HL-LHC): Technical design report, 10.23731/CYRM-2020-0010 (2020).

[16] A. Faus-Golfe *et al.*, Accelerators for electroweak physics and Higgs boson studies, arXiv:2209.05827.

[17] A. Abada *et al.* (FCC Collaboration), FCC physics opportunities: Future circular collider conceptual design report volume 1, Eur. Phys. J. C **79**, 474 (2019).

[18] W. Buchmuller and D. Wyler, Effective Lagrangian analysis of new interactions and flavor conservation, Nucl. Phys. **B268**, 621 (1986).

[19] B. Grzadkowski, M. Iskrzynski, M. Misiak, and J. Rosiek, Dimension-six terms in the standard model Lagrangian, J. High Energy Phys. 10 (2010) 085.

[20] I. Brivio and M. Trott, The standard model as an effective field theory, Phys. Rep. **793**, 1 (2019).

[21] F. Feruglio, The chiral approach to the electroweak interactions, Int. J. Mod. Phys. A **08**, 4937 (1993).

[22] R. Alonso, M. B. Gavela, L. Merlo, S. Rigolin, and J. Yepes, The effective chiral Lagrangian for a light dynamical "Higgs particle", Phys. Lett. B **722**, 330 (2013); **726**, 926(E) (2013).

[23] G. Buchalla, O. Catà, and C. Krause, Complete electroweak chiral Lagrangian with a light Higgs at NLO, Nucl. Phys. **B880**, 552 (2014); **B913**, 475(E) (2016).

[24] D. Gonçalves, T. Han, F. Kling, T. Plehn, and M. Takeuchi, Higgs boson pair production at future hadron colliders: From kinematics to dynamics, Phys. Rev. D **97**, 113004 (2018).

[25] J. Chang, K. Cheung, J. S. Lee, C.-T. Lu, and J. Park, Higgs-boson-pair production $H(\to b\bar{b})H(\to \gamma\gamma)$ from gluon fusion at the HL-LHC and HL-100 TeV hadron collider, Phys. Rev. D **100**, 096001 (2019).

[26] S. Chang and M. A. Luty, The Higgs trilinear coupling and the scale of new physics, J. High Energy Phys. 03 (2020) 140.

[27] P. Agrawal, D. Saha, L.-X. Xu, J.-H. Yu, and C. P. Yuan, Determining the shape of the Higgs potential at future colliders, Phys. Rev. D **101**, 075023 (2020).

[28] M. L. Mangano, G. Ortona, and M. Selvaggi, Measuring the Higgs self-coupling via Higgs-pair production at a 100 TeV p-p collider, Eur. Phys. J. C **80**, 1030 (2020).

[29] K. Chai, J.-H. Yu, and H. Zhang, Investigating Higgs self-interaction through di-Higgs plus jet production at a 100 TeV hadron collider, Phys. Rev. D **107**, 055031 (2023).

[30] L. Da Rold, M. Epele, A. D. Medina, N. I. Mileo, and A. Szynkman, Double Higgs production at the HL-LHC: Probing a loop-enhanced model with kinematical distributions, J. High Energy Phys. 05 (2024) 072.

[31] A. Papaefstathiou and G. Tetlalmatzi-Xolocotzi, Multi-Higgs boson production with anomalous interactions at current and future proton colliders, J. High Energy Phys. 06 (2024) 124.

[32] A. Tumasyan *et al.* (CMS Collaboration), A portrait of the Higgs boson by the CMS experiment ten years after the discovery, Nature (London) **607**, 60 (2022).

[33] G. Aad *et al.* (ATLAS Collaboration), Constraints on the Higgs boson self-coupling from single- and double-Higgs production with the ATLAS detector using pp collisions at $\sqrt{s} = 13$ TeV, Phys. Lett. B **843**, 137745 (2023).

[34] F. Kling, T. Plehn, and P. Schichtel, Maximizing the significance in Higgs boson pair analyses, Phys. Rev. D **95**, 035026 (2017).

[35] K. Cranmer, J. Brehme, and G. Louppe, The frontier of simulation-based inference (Proceedings of the National Academy of Sciences, 2020), 10.1073/pnas.1912789117.

[36] A. Carvalho, M. Dall'Osso, T. Dorigo, F. Goertz, C. A. Gottardo, and M. Tosi, Higgs pair production: Choosing benchmarks with cluster analysis, J. High Energy Phys. 04 (2016) 126.
[37] G. Buchalla, M. Capozi, A. Celis, G. Heinrich, and L. Scyboz, Higgs boson pair production in non-linear effective field theory with full m_t-dependence at NLO QCD, J. High Energy Phys. 09 (2018) 057.
[38] M. Capozi and G. Heinrich, Exploring anomalous couplings in Higgs boson pair production through shape analysis, J. High Energy Phys. 03 (2020) 091.
[39] L. Alasfar *et al.*, Effective field theory descriptions of Higgs boson pair production, arXiv:2304.01968.
[40] J. Brehmer, F. Kling, I. Espejo, and K. Cranmer, MadMiner: Machine learning-based inference for particle physics, Comput. Software Big Sci. **4**, 3 (2020).
[41] K. Kong, K. T. Matchev, S. Mrenna, and P. Shyamsundar, New machine learning techniques for simulation-based inference: InferoStatic nets, kernel score estimation, and kernel likelihood ratio estimation, arXiv:2210.01680.
[42] R. Gomez Ambrosio, J. ter Hoeve, M. Madigan, J. Rojo, and V. Sanz, Unbinned multivariate observables for global SMEFT analyses from machine learning, J. High Energy Phys. 03 (2023) 033.
[43] J. Brehmer, S. Dawson, S. Homiller, F. Kling, and T. Plehn, Benchmarking simplified template cross sections in WH production, J. High Energy Phys. 11 (2019) 034.
[44] J. Ellis, M. Madigan, K. Mimasu, V. Sanz, and T. You, Top, Higgs, diboson and electroweak fit to the standard model effective field theory, J. High Energy Phys. 04 (2021) 279.
[45] J. J. Ethier, G. Magni, F. Maltoni, L. Mantani, E. R. Nocera, J. Rojo, E. Slade, E. Vryonidou, and C. Zhang (SMEFiT Collaboration), Combined SMEFT interpretation of Higgs, diboson, and top quark data from the LHC, J. High Energy Phys. 11 (2021) 089.
[46] N. Elmer, M. Madigan, T. Plehn, and N. Schmal, Staying on top of SMEFT-likelihood analyses, arXiv:2312.12502.
[47] S. Borowka, C. Duhr, F. Maltoni, D. Pagani, A. Shivaji, and X. Zhao, Probing the scalar potential via double Higgs boson production at hadron colliders, J. High Energy Phys. 04 (2019) 016.
[48] C. W. Murphy, Dimension-8 operators in the standard model effective field theory, J. High Energy Phys. 10 (2020) 174.
[49] A. Biekötter, D. Gonçalves, T. Plehn, M. Takeuchi, and D. Zerwas, The global Higgs picture at 27 TeV, SciPost Phys. **6**, 024 (2019).
[50] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H.-S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations, J. High Energy Phys. 07 (2014) 079.
[51] T. Plehn, M. Spira, and P. M. Zerwas, Pair production of neutral Higgs particles in gluon-gluon collisions, Nucl. Phys. **B479**, 46 (1996); **B531**, 655(E) (1998).
[52] T. Plehn, *Lectures on LHC Physics* (Springer, Cham, 2015).
[53] C. Degrande, G. Durieux, F. Maltoni, K. Mimasu, E. Vryonidou, and C. Zhang, Automated one-loop computations in the Standard Model effective field theory, Phys. Rev. D **103**, 096024 (2021).
[54] M. Buschmann, D. Goncalves, S. Kuttimalai, M. Schonherr, F. Krauss, and T. Plehn, Mass effects in the Higgs-gluon coupling: Boosted vs off-shell production, J. High Energy Phys. 02 (2015) 038.
[55] J. Ellis, M. Madigan, K. Mimasu, V. Sanz, and T. You, Top, Higgs, diboson and electroweak fit to the standard model effective field theory, J. High Energy Phys. 04 (2021) 279.
[56] I. Brivio, S. Bruggisser, N. Elmer, E. Geoffray, M. Luchmann, and T. Plehn, To profile or to marginalize—A SMEFT case study, SciPost Phys. **16**, 035 (2024).
[57] U. Baur, T. Plehn, and D. L. Rainwater, Examining the Higgs boson potential at lepton and hadron colliders: A comparative analysis, Phys. Rev. D **68**, 033001 (2003).
[58] M. J. Dolan, C. Englert, and M. Spannowsky, Higgs self-coupling measurements at the LHC, J. High Energy Phys. 10 (2012) 112.
[59] J. Baglio, A. Djouadi, R. Gröber, M. M. Mühlleitner, J. Quevillon, and M. Spira, The measurement of the Higgs self-coupling at the LHC: Theoretical status, J. High Energy Phys. 04 (2013) 151.
[60] A. Abada *et al.* (FCC Collaboration), FCC-hh: The hadron collider, Eur. Phys. J. Spec. Topics **228**, 755 (2019).
[61] S. Dawson, S. Dittmaier, and M. Spira, Neutral Higgs boson pair production at hadron colliders: QCD corrections, Phys. Rev. D **58**, 115012 (1998).
[62] J. Grigo, J. Hoff, K. Melnikov, and M. Steinhauser, On the Higgs boson pair production at the LHC, Nucl. Phys. **B875**, 1 (2013).
[63] F. Maltoni, E. Vryonidou, and M. Zaro, Top-quark mass effects in double and triple Higgs production in gluon-gluon fusion at NLO, J. High Energy Phys. 11 (2014) 079.
[64] S. Borowka, N. Greiner, G. Heinrich, S. P. Jones, M. Kerner, J. Schlenk, and T. Zirke, Full top quark mass dependence in Higgs boson pair production at NLO, J. High Energy Phys. 10 (2016) 107.
[65] S. Borowka, N. Greiner, G. Heinrich, S. P. Jones, M. Kerner, J. Schlenk, U. Schubert, and T. Zirke, Higgs boson pair production in gluon fusion at next-to-leading order with full top-quark mass dependence, Phys. Rev. Lett. **117**, 012001 (2016); **117**, 079901(E) (2016).
[66] J. Baglio, F. Campanario, S. Glaus, M. Mühlleitner, M. Spira, and J. Streicher, Gluon fusion into Higgs pairs at NLO QCD and the top mass scheme, Eur. Phys. J. C **79**, 459 (2019).
[67] J. Baglio, F. Campanario, S. Glaus, M. Mühlleitner, J. Ronca, M. Spira, and J. Streicher, Higgs-pair production via gluon fusion at hadron colliders: NLO QCD corrections, J. High Energy Phys. 04 (2020) 181.
[68] D. de Florian and J. Mazzitelli, Higgs pair production at next-to-next-to-leading logarithmic accuracy at the LHC, J. High Energy Phys. 09 (2015) 053.
[69] J. Grigo, J. Hoff, and M. Steinhauser, Higgs boson pair production: Top quark mass effects at NLO and NNLO, Nucl. Phys. **B900**, 412 (2015).
[70] D. de Florian, M. Grazzini, C. Hanga, S. Kallweit, J. M. Lindert, P. Maierhöfer, J. Mazzitelli, and D. Rathlev,

Differential Higgs boson pair production at next-to-next-to-leading order in QCD, J. High Energy Phys. 09 (2016) 151.
[71] L.-B. Chen, H. T. Li, H.-S. Shao, and J. Wang, Higgs boson pair production via gluon fusion at N^{3}LO in QCD, Phys. Lett. B **803**, 135292 (2020).
[72] L.-B. Chen, H. T. Li, H.-S. Shao, and J. Wang, The gluon-fusion production of Higgs boson pair: N^{3}LO QCD corrections and top-quark mass effects, J. High Energy Phys. 03 (2020) 072.
[73] G. Heinrich, S. P. Jones, M. Kerner, G. Luisoni, and L. Scyboz, Probing the trilinear Higgs boson coupling in di-Higgs production at NLO QCD including parton shower effects, J. High Energy Phys. 06 (2019) 066.
[74] J. Baglio, F. Campanario, S. Glaus, M. Mühlleitner, J. Ronca, and M. Spira, $gg \to HH$: Combined uncertainties, Phys. Rev. D **103**, 056002 (2021).
[75] D. de Florian *et al.* (LHC Higgs Cross Section Working Group), Handbook of LHC Higgs cross sections: 4. Deciphering the nature of the Higgs sector, 10.23731/CYRM-2017-002 (2016).
[76] P. Artoisenet, R. Frederix, O. Mattelaer, and R. Rietkerk, Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations, J. High Energy Phys. 03 (2013) 015.
[77] C. Bierlich *et al.*, A comprehensive guide to the physics and usage of PYTHIA 8.3, SciPost Phys. Codebases **2022**, 8 (2022).
[78] J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens, and M. Selvaggi, DELPHES 3: A modular framework for fast simulation of a generic collider experiment, J. High Energy Phys. 02 (2014) 057.
[79] M. Cacciari, G. P. Salam, and G. Soyez, The anti-k_t jet clustering algorithm, J. High Energy Phys. 04 (2008) 063.
[80] M. Cacciari, G. P. Salam, and G. Soyez, FastJet user manual, Eur. Phys. J. C **72**, 1896 (2012).
[81] G. Heinrich and J. Lang, Combining chromomagnetic and four-fermion operators with leading SMEFT operators for $gg \to hh$ at NLO QCD, J. High Energy Phys. 05 (2024) 121.
[82] G. Heinrich, S. P. Jones, M. Kerner, and L. Scyboz, A non-linear EFT description of $gg \to HH$ at NLO interfaced to POWHEG, J. High Energy Phys. 10 (2020) 021.
[83] G. Heinrich, J. Lang, and L. Scyboz, SMEFT predictions for $gg \to hh$ at full NLO QCD and truncation uncertainties, J. High Energy Phys. 08 (2022) 079; 10 (2023) 086(E).
[84] D. de Florian, I. Fabre, G. Heinrich, J. Mazzitelli, and L. Scyboz, Anomalous couplings in Higgs-boson pair production at approximate NNLO QCD, J. High Energy Phys. 09 (2021) 161.
[85] D. Fäh and N. Greiner, Diphoton production in association with two bottom jets, Eur. Phys. J. C **77**, 750 (2017).
[86] D. Kim, S. Lee, H. Jung, D. Kim, J. Kim, and J. Song, A panoramic study of K-factors for 111 processes at the 14 TeV LHC, J. Korean Phys. Soc. **84**, 914 (2024).
[87] M. L. Mangano *et al.*, Physics at a 100 TeV pp collider: Standard model processes, 10.23731/CYRM-2017-003.1 (2016).
[88] A. M. Sirunyan *et al.* (CMS Collaboration), Search for nonresonant Higgs boson pair production in final states with two bottom quarks and two photons in proton-proton collisions at $\sqrt{s} = 13$ TeV, J. High Energy Phys. 03 (2021) 257.
[89] Z. Jia (ATLAS Collaboration), Studies of new Higgs boson interactions through nonresonant HH production in the $b\bar{b}\gamma\gamma$ final state in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector, Proc. Sci. EPS-HEP2023 (**2024**) 422.
[90] S. Chatterjee, S. Rohshap, R. Schöfbeck, and D. Schwarz, Learning the EFT likelihood with tree boosting, arXiv:2205.12976.
[91] S. Carrazza, J. Cruz-Martinez, M. Rossi, and M. Zaro, MadFlow: Towards the automation of Monte Carlo simulation on GPU for particle physics processes, EPJ Web Conf. **251**, 03022 (2021).
[92] L. Heinrich and M. Kagan, Differentiable matrix elements with MadJax, J. Phys. Conf. Ser. **2438**, 012137 (2023).
[93] B. Nachman and S. Prestel, Morphing parton showers with event derivatives, arXiv:2208.02274.
[94] J. Neyman and E. S. Pearson, On the problem of the most efficient tests of statistical hypotheses, Phil. Trans. R. Soc. A **231**, 289 (1933).
[95] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, Springer Series in Statistics (Springer, New York, New York, 2001).
[96] M. Sugiyama, T. Suzuki, and T. Kanamori, *Density Ratio Estimation in Machine Learning* (Cambridge University Press, Cambridge, England, 2012).
[97] K. Cranmer, J. Pavez, and G. Louppe, Approximating likelihood ratios with calibrated discriminative classifiers, arXiv:1506.02169.
[98] S. Rizvi, M. Pettee, and B. Nachman, Learning likelihood ratios with neural network classifiers, J. High Energy Phys. 02 (2024) 136.
[99] P. Baldi, K. Cranmer, T. Faucett, P. Sadowski, and D. Whiteson, Parameterized neural networks for high-energy physics, Eur. Phys. J. C **76**, 235 (2016).
[100] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, PyTorch: An imperative style, high-performance deep learning library, arXiv:1912.01703.
[101] D. P. Kingma and J. Ba, ADAM: A method for stochastic optimization, arXiv:1412.6980.
[102] https://zenodo.org/records/11222924.
[103] https://github.com/rmastand/nsbi_for_dihiggs.