

CERN-EP-2022-095
2022/08/03

CMS-HIG-21-003

Search for the exotic decay of the Higgs boson into two light pseudoscalars with four photons in the final state in proton-proton collisions at $\sqrt{s} = 13$ TeV

The CMS Collaboration

Abstract

A search for the exotic decay of the Higgs boson to a pair of light pseudoscalars, each of which subsequently decays into a pair of photons, is presented. The search uses data from proton-proton collisions at $\sqrt{s} = 13$ TeV recorded with the CMS detector at the LHC that corresponds to an integrated luminosity of 132 fb^{-1} . The analysis probes pseudoscalar bosons with masses in the range 15–62 GeV, coming from the Higgs boson decay, which leads to four well-isolated photons in the final state. No significant deviation from the background-only hypothesis is observed. Upper limits are set on the product of the Higgs boson production cross section and branching fraction into four photons. The observed (expected) limits range from 0.80 (1.00) fb for a pseudoscalar boson mass of 15 GeV to 0.26 (0.24) fb for a mass of 62 GeV at 95% confidence level.

Submitted to the Journal of High Energy Physics

arXiv:2208.01469v1 [hep-ex] 2 Aug 2022

1 Introduction

In 2012, a Higgs boson (H) with a mass of 125 GeV was discovered by the ATLAS and CMS experiments [1–3] at the CERN LHC. Since its discovery, both collaborations have performed precision measurements of the spin, parity, width, and couplings of the Higgs boson in its various production and decay modes [4–14], all of which indicate that Higgs boson properties are compatible with the standard model (SM) predictions. However, data collected at $\sqrt{s} = 13$ TeV provide an upper limit on the branching fraction of the Higgs boson to undetected states of about 40% at 95% confidence level (CL) [15]. This leaves a large margin for beyond-the-SM (BSM) decays of the Higgs boson. Various theoretical models, such as 2HDM+S models [16], describe a BSM Higgs boson that decays to light bosons, which has not yet been excluded.

Multiple searches for exotic decays of the Higgs boson have been performed using the 8 TeV [17–20] and 13 TeV [21–27] data collected at the LHC. Decays of the type $H \rightarrow aa$, where a is a light pseudoscalar boson, are well motivated in various BSM scenarios [28–31]. In many scenarios, such as fermiophobic a decays, the branching fraction of the pseudoscalar bosons to a pair of photons is close to unity, which enables this search at the LHC. The final state, with four photons, provides an experimental signature that has very small contributions from SM processes and is thus an important channel for the search for light pseudoscalar bosons.

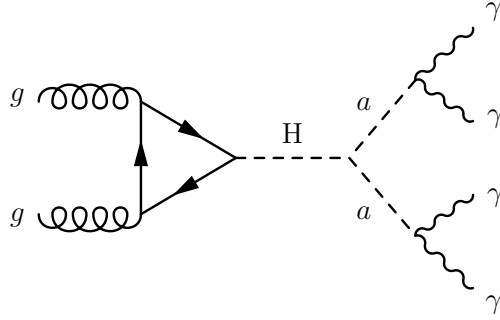


Figure 1: Feynman diagram for a BSM decay of the Higgs boson into a pair of light pseudoscalar bosons that subsequently decay into photons.

This paper presents a search for light pseudoscalar bosons that arise from the decay of a Higgs boson, with four photons in the final state: $H \rightarrow aa \rightarrow \gamma\gamma\gamma\gamma$. The event topology depends on the mass of the pseudoscalar boson being probed, which determines the opening angle between the photons for each pair. This analysis considers only events with four isolated photons in which the angular distance between the photons, for both photon pairs, is greater than 0.2. These requirements enable a search for pseudoscalar bosons that range in mass from 15 to 62 GeV and produce an experimental signature with four well isolated and fully reconstructed photons. The signal reconstruction efficiency increases as the mass increases. The dominant Feynman diagram contributing to this process is shown in Fig. 1.

A previous search for light pseudoscalar bosons in events with at least three photons was performed by the ATLAS Collaboration using the LHC data collected at $\sqrt{s} = 8$ TeV [32]. The first search of this type done by the CMS Collaboration is presented in this paper. Data collected by the CMS experiment from 2016 to 2018, which correspond to an integrated luminosity of 132 fb^{-1} , are used in this analysis.

This paper is organized as follows. A brief description of the CMS detector is given in Section 2. Details of the data and simulation used in this analysis are described in Section 3. The recon-

struction of $H \rightarrow aa \rightarrow \gamma\gamma\gamma\gamma$ events and the associated selection requirements are summarized in Section 4. An overview of the analysis strategy is given in Section 5. Section 6 describes the background estimation technique, which is used as input to a multivariate discriminator. The optimization procedure, which uses the output of this discriminator, is discussed in Section 7. Sections 8, 9 and 10 describe the statistical procedure and the signal and background modeling, respectively. The systematic uncertainties are discussed in Section 11. Finally, the results are presented in Section 12, and the paper is summarized in Section 13.

2 The CMS detector

The central feature of the CMS apparatus is a superconducting solenoid of 6 m internal diameter, providing a magnetic field of 3.8 T. Within the solenoid volume is a silicon pixel and strip tracker, a lead tungstate crystal electromagnetic calorimeter (ECAL), and a brass and scintillator hadron calorimeter (HCAL), each composed of a barrel and two endcap sections. Forward calorimeters extend the pseudorapidity (η) coverage provided by the barrel and endcap detectors. Muons are detected in gas-ionization chambers embedded in the steel flux-return yoke outside the solenoid.

Events of interest are selected using a two-tiered trigger system [33]. The first level, composed of custom hardware processors, uses information from the calorimeters and muon detectors to select events at a rate of around 100 kHz within a time interval of less than $4 \mu\text{s}$ [34]. The second level, known as the high-level trigger, consists of a farm of processors running a version of the full event reconstruction software optimized for fast processing, and reduces the event rate to around 1 kHz before data storage [35].

The particle-flow algorithm [36] aims to reconstruct and identify each individual particle in an event, with an optimized combination of information from the various elements of the CMS detector. The energy of photons is measured using the ECAL, as described in Ref. [37]. In the barrel section of the ECAL, an energy resolution of about 1% is achieved for unconverted or late-converting photons in the tens of GeV energy range. The energy resolution of the remaining barrel photons is about 1.3% up to $|\eta| = 1$, changing to about 2.5% at $|\eta| = 1.4$. In the endcaps, the energy resolution is about 2.5% for unconverted or late-converting photons, and between 3 and 4% for the photons that convert closer to the beam spot [37, 38].

A more detailed description of the CMS detector, together with a definition of the coordinate system used and the relevant kinematic variables, can be found in Ref. [39].

3 Data samples and simulated events

The proton-proton (p-p) collision data at $\sqrt{s} = 13 \text{ TeV}$ used in this analysis were collected in 2016, 2017, and 2018 and correspond, respectively, to integrated luminosities of 36.3, 41.5, and 54.4 fb^{-1} . This is 5.4 fb^{-1} less than the standard CMS detector collected luminosity [40–42], because a required trigger was missing for a short period. Events are selected using a high-level diphoton trigger, optimized for the low-mass diphoton Higgs boson search [43], with photon transverse momentum (p_T) thresholds of 30 GeV and 18 GeV. Additionally, calorimeter-based identification requirements, which use information such as the shape of the electromagnetic shower, the isolation of the photon candidate, and the ratio between the hadronic and electromagnetic energy deposit of the shower, are applied to the photon candidates at trigger level. Furthermore, the chosen trigger requires the diphoton candidate to have an invariant mass greater than 55 GeV in data collected during 2016–2017. Each event is required to contain at

least one diphoton candidate that satisfies these high-level trigger requirements.

The simulated signal samples were generated corresponding to a pseudoscalar boson mass, m_a , ranging from 15 to 60 GeV, in steps of 5 GeV, assuming a Higgs boson mass of 125 GeV. These samples were simulated considering only the gluon fusion production mode of the Higgs boson, using MADGRAPH5_aMC@NLO v2.4.2 [44].

The dominant backgrounds in this search are SM production of $\gamma\gamma + \text{jets}$, $\gamma + \text{jets}$, as well as multijet events, in which jets are misidentified as isolated photons. As in Refs. [45, 46], the background contributions are modeled entirely from data.

All simulated samples used in this analysis model QCD showering and hadronization with the PYTHIA 8.212 [47] event generator. The CUETP8M1 tune was used for data collected in 2016 and the CP5 tune was used for data collected in 2017–2018 [48, 49]. The response of the CMS detector is modeled using the GEANT4 [50] package. The simulated events also include additional p-p interactions within the same or nearby bunch crossings (pileup). The average pileup in the 2016 (2017–2018) datasets is 23 (32) vertices. The simulated events are weighted to reproduce the distribution of pileup in data.

4 Event reconstruction

The particle-flow algorithm [36] is used to reconstruct photon candidates from energy clusters in the ECAL that are not matched to any charged particle trajectories originating in the pixel detector. The energy of the photon candidates is calculated by applying in-situ measured calibrations to the reconstructed hits in the ECAL. A multivariate regression technique is employed to correct the photon energies measured by the ECAL. These procedures are described in Ref. [37].

Deposits from quark fragmentation and hadronization are clustered into hadronic jets. The energy of charged hadrons is determined from a combination of their momentum measured in the tracker and the matching ECAL and HCAL energy deposits, while the energy of neutral hadrons is obtained from the corresponding corrected ECAL and HCAL energy deposits. Jets are clustered using the anti- k_T jet finding algorithm [51, 52] with a distance parameter of 0.4. The missing transverse momentum vector $\vec{p}_T^{\text{miss}}$ is computed as the negative vector sum of the transverse momenta of all the PF candidates in an event, and its magnitude is denoted as p_T^{miss} [53].

Prompt photons are distinguished from jets, which could be misidentified as a photon, using a multivariate analysis (MVA) technique that uses information related to the photon's electromagnetic shower shape, isolation, energy, and η . This technique is detailed in Refs. [10, 46].

Since photons do not leave deposits in the tracker, the most probable primary interaction vertex in the event is identified using a boosted decision tree (BDT). The primary vertex BDT is trained on simulated $H \rightarrow aa \rightarrow \gamma\gamma\gamma\gamma$ events and uses input variables related to tracks recoiling against the four-photon system and information related to photons converted in the tracker material, similar to Ref. [4]. A separate training is performed for each data-taking year (2016–2018) to properly model the variation in detector conditions over the three years, in particular with respect to the pixel detector upgrade. The analysis identifies the vertex with the highest BDT score as the primary vertex, which improves the resolution of the invariant mass of the Higgs boson candidate by approximately 3%. It also increases the vertex identification efficiency, 80% in total, which is defined as selecting a vertex within 1 cm of the true generator vertex, by about 10%, with respect to the vertex chosen with the largest value of summed p_T^2 of

the tracks.

5 Analysis strategy and selection

Events considered in this analysis are required to contain at least one diphoton candidate where both daughter photon candidates pass the identification criteria, which are slightly more stringent than the trigger requirements. Additionally, events must contain at least four identified photon candidates in the ECAL and tracker fiducial region ($|\eta| < 2.5$). This excludes the ECAL barrel-endcap transition region ($1.44 < |\eta| < 1.57$) where the photon reconstruction is suboptimal. The four photon candidates are also required to pass p_T requirements. The p_T thresholds on the first- and second-leading photons are 30 and 18 GeV, respectively. These selections are in sync with the p_T requirements at trigger level for the two leading photons. The thresholds on the third- and fourth-leading photons are 15 GeV, since the BDT-based photon identification criteria are optimal for $p_T > 15$ GeV. When more than four photon candidates are found, the four candidates with the highest p_T are chosen. Each photon candidate must pass an electron veto, based on the presence of geometrically compatible hits in the pixel detector. The four photon candidates with the highest p_T are also required to have an invariant mass ($m_{\gamma\gamma\gamma\gamma}$) between 110 and 180 GeV. The lower bound is chosen to avoid $m_{\gamma\gamma\gamma\gamma}$ spectrum biases coming from the trigger selections. The Higgs boson candidate is constructed from the four photon candidates, which have passed all the previously described selection requirements.

To improve the sensitivity of the search, a 4-photon event classifier is trained to separate signal events from background events. The 4-photon event classifier utilizes the identification and kinematic information of the photons and pseudoscalar boson candidates. An optimized selection on the output of the event classifier is used to define the final signal regions for the analysis. The input variables for the classifier are uncorrelated with $m_{\gamma\gamma\gamma\gamma}$; therefore, the shape of the $m_{\gamma\gamma\gamma\gamma}$ spectrum is not affected by any selections placed on the 4-photon event classifier, as verified in simulation.

When all four photons from the decay of the pseudoscalar boson pair are within the acceptance criteria of the analysis, the $H \rightarrow aa \rightarrow \gamma\gamma\gamma\gamma$ signal will create a resonance in the $m_{\gamma\gamma\gamma\gamma}$ distribution at 125 GeV. The analysis performs an unbinned maximum likelihood fit of the signal and background models to the observed $m_{\gamma\gamma\gamma\gamma}$ distribution in data after a selection on the classifier output is applied. The signal model is constructed from a parametric fit to the simulated signal, while the background model is created using a data driven approach. As the impact on the 4-photon event classifier discrimination power is negligible if the other non-gluon-fusion Higgs boson production modes are added as signal, limits are set on the product of the Higgs boson inclusive production cross section and branching fraction into the four photon final state.

In this paper, “nominal signal hypothesis” refers to simulated signal samples corresponding to a particular m_a value. This assumes a branching ratio of $a \rightarrow \gamma\gamma$ equal to unity. In these hypotheses, m_a ranges from 15–60 GeV in steps of 5 GeV. The final results are reported with an m_a granularity of 0.5 GeV up to $m_a = 40$ GeV and 1 GeV for $m_a > 40$ GeV, where the signal models for the intermediate mass hypotheses are constructed by interpolating the signal models between the nominal masses.

6 Background estimation

Because of the low event selection efficiency on the background samples, it is difficult to model the background from simulation. Therefore, the expected background model, which is used to train the 4-photon event classifier, is estimated from data.

The method, referred to as event mixing (a simplified version of “Hemisphere mixing” [54]), does not rely on a control or sideband region, but instead aims to estimate the background contribution using the original dataset as input. This procedure begins with using data events that have passed the trigger selections, as described in Section 3, and replacing three out of the four reconstructed photons in each event with reconstructed photons from three consecutive events to create a “mixed” dataset. Photon candidates in the mixed dataset are required to pass the selections described in Section 5. The pseudoscalar boson candidates, a_1 and a_2 , are reconstructed considering all possible pair combinations among the chosen four photons, and taking the two pairs with the smallest value of the difference between the pair invariant masses. The shuffling of the reconstructed photons between the events not only constructs a dataset that is similar to the original data, but is also insensitive to the possible presence of a resonant signal. This procedure is performed separately for data collected during 2016, 2017, and 2018.

The events in the mixed dataset can potentially have different kinematic properties from those in the original data. To correct this, a multi-dimensional per-event weight is calculated by comparing events from the mixed dataset to the original data in the $m_{\gamma\gamma\gamma\gamma}$ sideband region, i.e. 110–115 or 135–180 GeV. The weight is computed using the following variables: angular distance between the two pseudoscalar boson candidates, defined by $\Delta R_{a_1 a_2} = \sqrt{(\eta_{a_1} - \eta_{a_2})^2 + (\phi_{a_1} - \phi_{a_2})^2}$ where ϕ is azimuthal angle; the transverse momenta of the two pseudoscalar bosons (p_{T,a_1} and p_{T,a_2}); and the difference between invariant mass of the two pseudoscalar boson candidates ($m_{a_1} - m_{a_2}$). This weight is applied to each event in the mixed dataset, and the reweighted events are then used to train the classifier.

The event mixing dataset is used only for training the 4-photon classifier and to optimize the selection on the classifier score. Since the background model used in the final maximum likelihood fit is obtained directly from data, any residual disagreement between data and event mixing in the background-like regions cannot induce any bias, and it could only result in sub-optimal performance of the classifier.

7 Event classification

A dedicated 4-photon boosted decision tree (BDT) event classifier is trained to separate signal from background. The training sample is parameterized as a function of m_a in order to make the classifier output uniform and sensitive to the full range of signal hypotheses considered in the search [55]. In this approach, a parameter equal to the hypothesized pseudoscalar boson mass ($m_{a,\text{hyp}}$) is provided as input to the training. The set of variables is chosen such that $m_{\gamma\gamma\gamma\gamma}$ cannot be inferred from the inputs. This is done by verifying that their correlation with $m_{\gamma\gamma\gamma\gamma}$ variable is negligible and that the $m_{\gamma\gamma\gamma\gamma}$ spectrum is not distorted by applying a selection on the classifier output.

The parameterized classifier requires only a single training and is able to provide a smooth interpolation to pseudoscalar boson mass hypotheses not used for the training. The training signal sample is obtained by merging all generated signal samples of equal size with masses between 15 and 60 GeV in steps of 5 GeV. The value of the $m_{a,\text{hyp}}$ parameter is equal to the corresponding true mass for the nominal signal simulation. The event mixing dataset, as described

in Section 6, is used as the background in the training. For the background, the value of the parameter $m_{a,\text{hyp}}$ is randomly distributed as a flat function among all possible nominal m_a values in 5 GeV steps.

The variables used in the training help separate isolated photons originating from the decay of the pseudoscalar boson from those from prompt and non-prompt processes. Pseudoscalar boson candidates are constructed from the four photon candidates as follows. For every possible combination of two photon candidate pairs, the difference between the invariant masses of the paired photons (Δm) is evaluated. The pairs with the smallest value of Δm are chosen to reconstruct the pseudoscalar boson candidates.

The following discriminating variables are provided as input to the training:

1. Photon identification BDT score for all four photons.
2. p_T of the two pseudoscalar boson candidates, i.e., $p_{T,a1}$ and $p_{T,a2}$.
3. Difference between the reconstructed invariant mass of the pseudoscalar boson candidates, i.e., $m_{a1} - m_{a2}$.
4. Difference between the invariant masses of the pseudoscalar boson candidate and the $m_{a,\text{hyp}}$ parameter divided by $m_{\gamma\gamma\gamma\gamma}$, i.e., $(m_{a1} - m_{a,\text{hyp}})/m_{\gamma\gamma\gamma\gamma}$ and $(m_{a2} - m_{a,\text{hyp}})/m_{\gamma\gamma\gamma\gamma}$.
5. Angular distance $\Delta R_{a_1 a_2}$ divided by $m_{\gamma\gamma\gamma\gamma}$, i.e., $\Delta R_{a_1 a_2}/m_{\gamma\gamma\gamma\gamma}$.
6. Angle $\cos \theta_{a\gamma}^*$ in the pseudoscalar boson rest frame, between the leading pseudoscalar boson candidate and the leading photon produced from its decay, chosen in the laboratory frame. This variable is sensitive to the spin of the pseudoscalar boson object.

As part of this training procedure, simulated signal and background datasets from the three data-taking years (2016–2018) are scaled by their appropriate integrated luminosities and combined. This combination of datasets from three years provides large training statistics. Additionally, the signal and background samples are divided in half to create the training and testing samples.

The distributions of the four most highly ranked variables in the training: $(m_{a1} - m_{a,\text{hyp}})/m_{\gamma\gamma\gamma\gamma}$, $m_{a1} - m_{a2}$, and the photon identification BDT score of the third and fourth photon are shown in Fig. 2 for the event mixing dataset, data, and signal simulation for various pseudoscalar boson mass hypotheses. It shows the contributions from the event mixing dataset and the data from the $m_{\gamma\gamma\gamma\gamma}$ sidebands, satisfying either $m_{\gamma\gamma\gamma\gamma} = 110\text{--}115$ or $135\text{--}180$ GeV. The distributions of the event mixing dataset and data are found to be in reasonable agreement, and the residual disagreement does not induce any biases in the analysis since the final background model is derived directly from data.

A unique 4-photon BDT output is obtained for each pseudoscalar boson mass hypothesis. The difference between the correlations of the input variables used in the training leads to a disagreement in the output shape of the BDT between the event mixing model and data. This difference is addressed by reweighting the BDT output distribution for the event mixing model to match output distribution for data from the $m_{\gamma\gamma\gamma\gamma}$ sideband region.

In order to maximize the sensitivity of the analysis, events are categorized according to the output of the 4-photon BDT. The categorization is optimized by maximizing the approximate mean significance (AMS) [56] over all possible categories in a window covering the region:

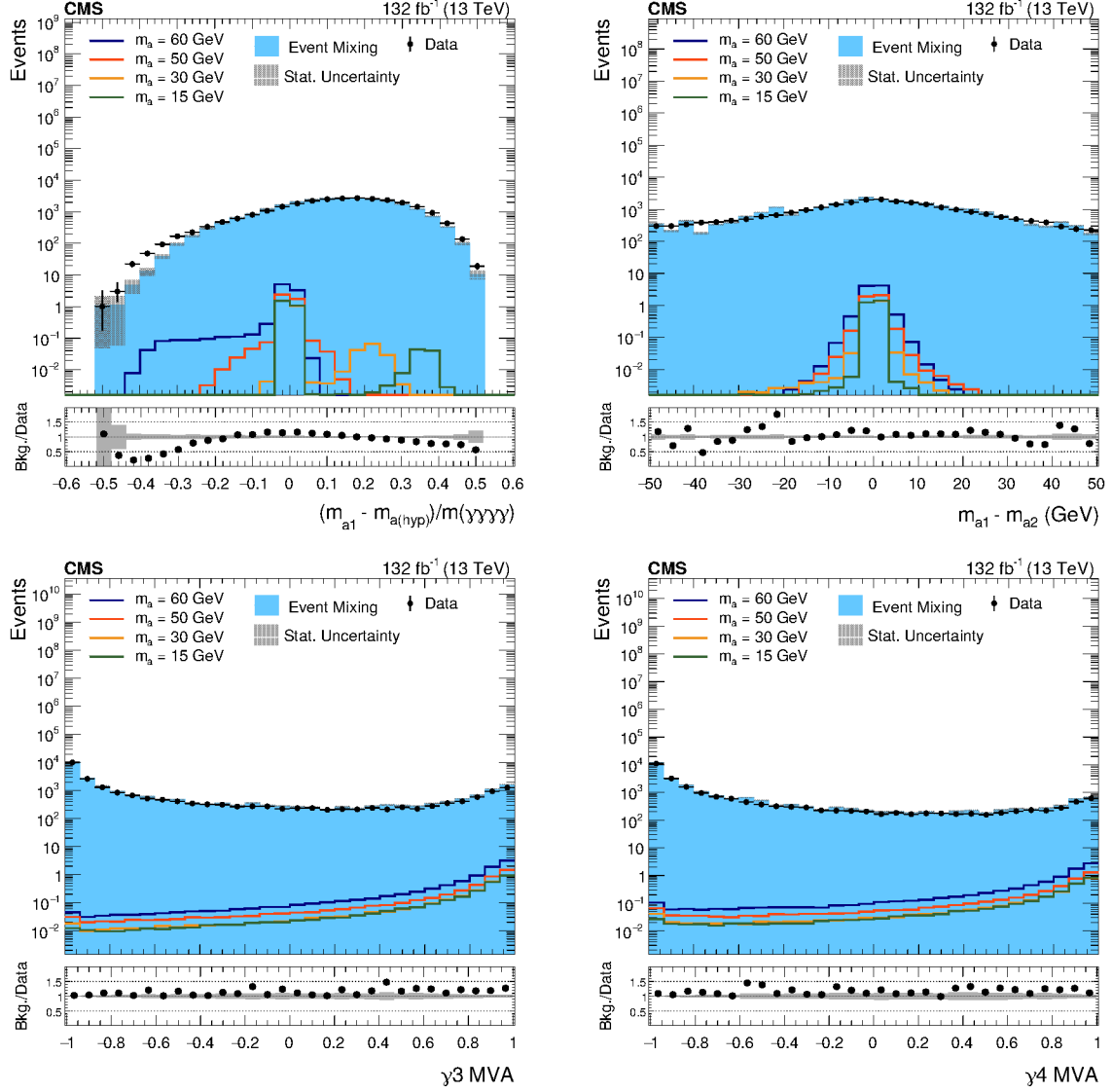


Figure 2: Distributions of the four most highly ranked discriminating variables: the difference between the invariant masses of the pseudoscalar bosons and the $m_{a,hyp}$ parameter, divided by the invariant mass of the four-photon system (upper left), with off-zero signal peaks from photon pairing mismatches; the difference between the invariant masses of the pseudoscalar boson (upper right); the photon identification BDT score of the third leading, γ_3 (lower left) and the fourth leading, γ_4 (lower right) photons. The events shown are selected from the $m_{\gamma\gamma\gamma\gamma}$ sidebands ($110 < m_{\gamma\gamma\gamma\gamma} < 115$ GeV or $135 < m_{\gamma\gamma\gamma\gamma} < 180$ GeV) for event mixing and data after fulfilling the selection criteria described in Section 4, while the signals are scaled with a cross-section of 1 pb.

$115 < m_{\gamma\gamma\gamma\gamma} < 135 \text{ GeV}$. In particular, AMS is computed for each category, and the total AMS, defined as the sum in quadrature of the AMS of each category, is maximized. The AMS of each category is defined as:

$$\text{AMS} = \sqrt{2 \left[(S + B) \ln \left(1 + \frac{S}{B} \right) - S \right]}. \quad (1)$$

In Eq. 1, S refers to the number of expected signal events and B refers to the number of estimated background events. In order to minimize the impact of statistical fluctuations on the optimization of the BDT selection, the output BDT distribution of the event mixing dataset is smoothed, using the super-smoothing technique [57, 58], prior to being used in this procedure. The distributions of the BDT output for data, simulated signal events, and event mixing dataset, after smoothing the BDT output distribution, is shown in Fig. 3 for $m_a = 15$ and 50 GeV. Events shown in this distribution are selected after passing the criteria described in Section 4 and are in the mass window $110 < m_{\gamma\gamma\gamma\gamma} < 180 \text{ GeV}$.

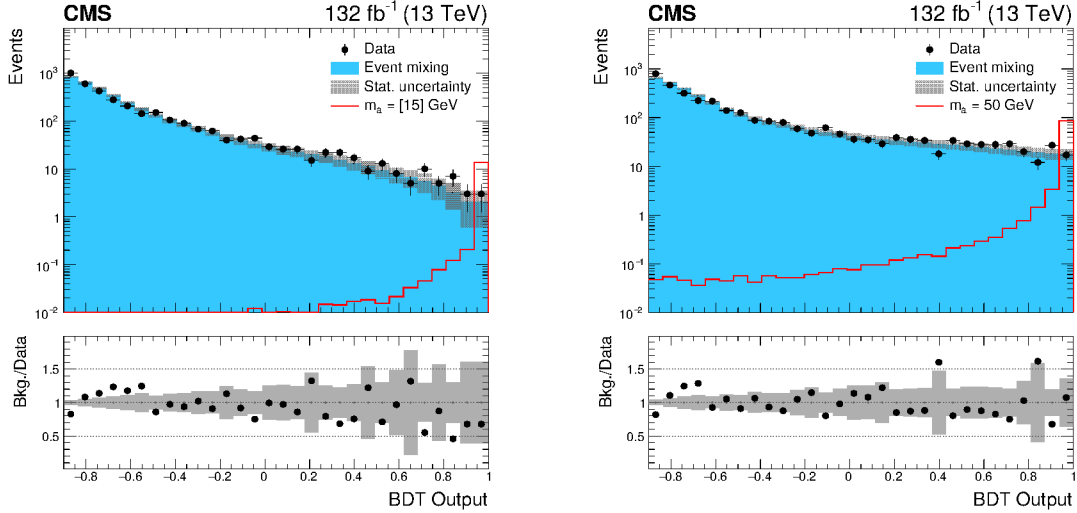


Figure 3: Distribution of the BDT output for $m_a = 15 \text{ GeV}$ (left) and 50 GeV (right) in data and simulated signal and event mixing (after smoothing) events. Events shown are selected after fulfilling the selection criteria described in Section 4 in the mass window $110 < m_{\gamma\gamma\gamma\gamma} < 180 \text{ GeV}$, while the signal is scaled with a cross-section of 1 pb .

A priori, the greater the number of categories, the better the significance is within the uncertainties. Therefore this optimization procedure was performed separately for each of the nominal signal hypotheses for up to five categories of the 4-photon BDT score. However, as an increase of less than 1% in the AMS value was observed when increasing the number of categories beyond one, only a single category based on the BDT output was created for each pseudoscalar boson mass hypothesis. Table 1 summarizes the minimum BDT selection and the efficiency of that selection obtained for each of the nominal signal hypotheses. For the intermediate m_a values, the BDT selection obtained from the closest nominal signal hypothesis is applied. For each signal hypothesis, events that pass the selection on the BDT output are used to obtain final results.

Table 1: Summary of the optimized BDT output threshold values and the efficiency with respect to a selection on this output for each of the nominal signal hypothesis.

m_a (GeV)	Minimum 4-photon event classification BDT output value	Signal efficiency of the selection on BDT
15	0.883	88%
20	0.891	87%
25	0.876	86%
30	0.897	84%
35	0.931	82%
40	0.945	78%
45	0.952	80%
50	0.958	80%
55	0.976	77%
60	0.987	71%

8 Statistical procedure

The statistical procedure used in this analysis is identical to that described in Ref. [59], as developed by the ATLAS and CMS Collaborations. Simultaneous unbinned maximum likelihood fits are performed to the $m_{\gamma\gamma\gamma\gamma}$ distributions of all analysis categories, with an m_a granularity of 0.5 GeV for $15 < m_a < 40$ GeV and 1 GeV for $40 \leq m_a \leq 62$ GeV. A likelihood function is defined for each analysis category using analytic models to describe the $m_{\gamma\gamma\gamma\gamma}$ distributions of signal and background events with nuisance parameters to account for the experimental and theoretical systematic uncertainties described in Section 11. The best fit values and confidence intervals for the parameters of interest are estimated using a profile likelihood test statistic:

$$q(\vec{\alpha}) = -2 \ln \left(\frac{\mathcal{L}(\vec{\alpha}, \hat{\theta}_{\vec{\alpha}})}{\mathcal{L}(\hat{\vec{\alpha}}, \hat{\vec{\theta}})} \right). \quad (2)$$

Where the quantities $\hat{\vec{\alpha}}$ and $\hat{\vec{\theta}}$ describe the unconditional maximum likelihood estimates for the parameters of interest and the nuisance parameters, respectively, whereas $\hat{\theta}_{\vec{\alpha}}$ corresponds to the conditional maximum likelihood estimate for fixed values of the parameters of interest, $\vec{\alpha}$. The value of twice the negative logarithm of the likelihood ratio, Eq. 2 is minimized when a fit of these functions is performed on the $m_{\gamma\gamma\gamma\gamma}$ distribution. A penalty term, equal to the number of parameters in the functions, is added to the $-2 \Delta \ln L$ to prevent the addition of unnecessary floating parameters in the fit.

9 Signal model

The signal shape for the $m_{\gamma\gamma\gamma\gamma}$ distribution, for each nominal signal hypothesis, is constructed from simulation. After all of the analysis selection criteria are applied, a unique signal model is built for each nominal signal hypothesis for each of the three data-taking years (2016–2018). The $m_{\gamma\gamma\gamma\gamma}$ distribution is modeled with a double-sided Crystal Ball (CB) function [60], which is a modified version of the standard CB function with two independent power-low tails. These signal models, scaled by the integrated luminosity for each year, are summed in order to construct the final model. The signal models for each year are shown in Fig. 4 for $m_a = 15$ GeV. The full width at half maximum (FWHM) and the effective standard deviation (σ_{eff}), defined as half the width of the smallest interval containing 68% of the $m_{\gamma\gamma\gamma\gamma}$ distribution, are also

shown.

Two factors need to be considered to build the signal models for the intermediate mass hypotheses: the shape of the $m_{\gamma\gamma\gamma\gamma}$ distribution and its yield. Since the shape of the $m_{\gamma\gamma\gamma\gamma}$ distribution is not found to change significantly within a 5 GeV window around the nominal mass hypothesis, only the yield of the signal model is parameterized as a function of m_a for the intermediate mass hypotheses. For each intermediate point, a signal model is constructed from the $m_{\gamma\gamma\gamma\gamma}$ shape of the nearest nominal mass hypothesis and the interpolated yield between the two nearest nominal mass hypotheses.

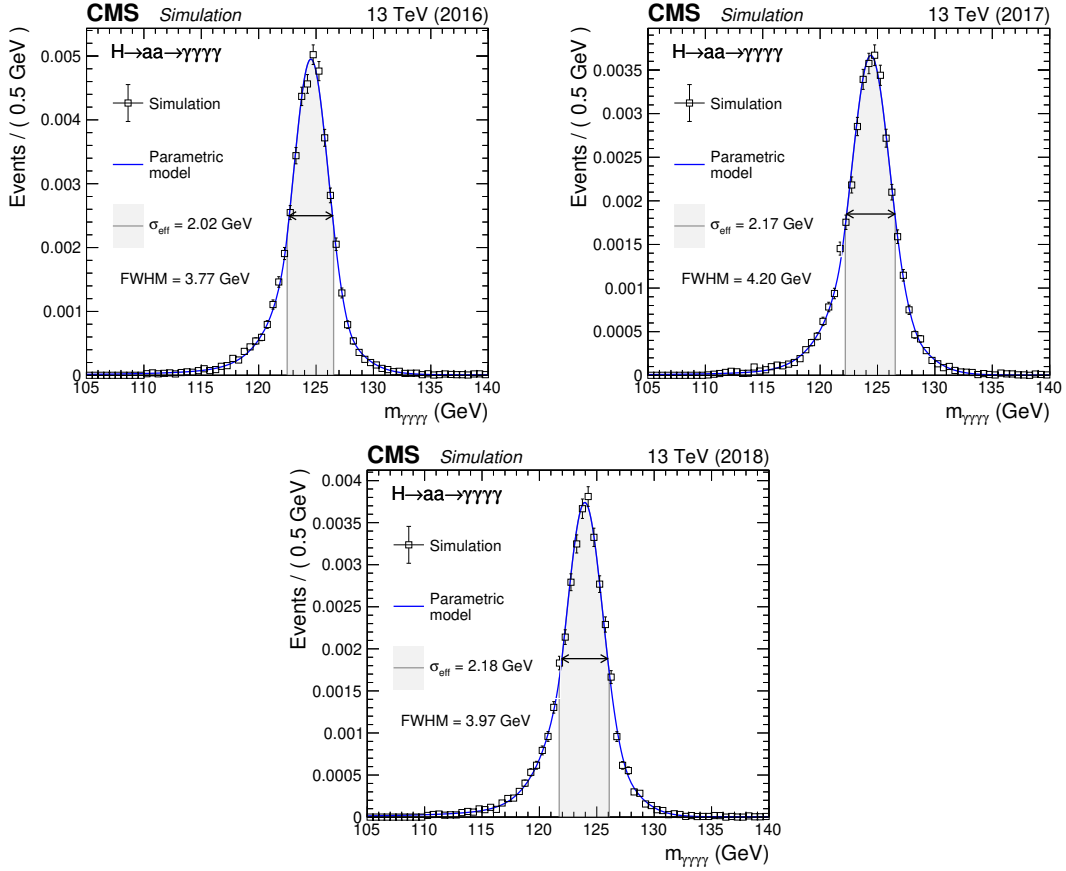


Figure 4: The parameterized signal shape for $m_a = 15$ GeV is shown for the 2016 (upper left), 2017 (upper right), and 2018 (lower) data-taking years. Separate signal models are built for each of the three data-taking years, which are then scaled by the appropriate luminosity and summed in order to construct the final signal model. The open squares represent simulated events and the blue line is the corresponding model. Also shown is the σ_{eff} value (half the width of the narrowest interval containing 68.3% of the invariant mass distribution), with the corresponding interval as a gray band and the FWHM, with the corresponding interval marked with a double arrow.

10 Background model

The background model is built to describe the shape of the $m_{\gamma\gamma\gamma\gamma}$ distribution that results from processes other than the signal process. Since the shape of this distribution is not known, different functional forms must be considered in the construction of the background model. The function chosen to describe the background can result in a different number of estimated

events in the signal peak and, as a result, affect the measured signal strength. This inherent uncertainty associated with choosing a background function is accounted for by the discrete profiling method [61]. This method describes background modeling performed using data as implemented in Ref. [45], and treats the choice of the background function as a discrete nuisance parameter in the likelihood fit to data.

As part of the background modeling procedure, a large set of candidate functions are considered that includes exponentials, Bernstein polynomials, Laurent series, and power law functions. A subset of functions from each family are used to build the background model. For each family, the maximum order of parameters is fixed by means of an F-test [62], and the minimum order is determined by applying a requirement on the goodness-of-fit.

The fits are performed over the range $110 < m_{\gamma\gamma\gamma\gamma} < 180 \text{ GeV}$, and the data from the three years (2016–2018) are combined in order to construct the background model. A unique background model is created corresponding to each m_a hypothesis. For each m_a hypothesis, an ensemble of pseudo-experiments was generated using the various background functions. Each pseudo-experiment was fitted using the discrete profiling method, and it was established that the chosen functional form, used to describe the background, does not introduce any potential bias in the signal strength measurement.

11 Systematic uncertainties

The systematic uncertainty associated with the background estimation is taken into account by the discrete profiling method, as described in Section 10. There are two kinds of systematic uncertainties that affect the signal model: those that modify the shape of the $m_{\gamma\gamma\gamma\gamma}$ distribution, and those that leave the shape of the $m_{\gamma\gamma\gamma\gamma}$ distribution unchanged but affect the overall normalization of the signal model.

The uncertainties that affect the shape of the $m_{\gamma\gamma\gamma\gamma}$ distribution, which are incorporated in the signal model as nuisance parameters, are described below. These uncertainties are typically related to the energy of the individual photons, and they affect the mean and width of the signal model.

1. *Photon energy scale and resolution:* Corrections are applied to the photon energy scale in data and to the energy resolution in simulation. The uncertainties related to these corrections are computed using $Z \rightarrow ee$ events [37]. The resulting uncertainty in the energy scale is 0.05–0.15%, depending on the photon p_T .
2. *Nonlinearity of the photon energy scale:* Any remaining differences in the linearity of the photon energy scale between data and simulation are covered by this uncertainty, which is estimated using Lorentz-boosted $Z \rightarrow ee$ events. The procedure for estimating this uncertainty is detailed in Ref. [4]. An uncertainty of 0.1% on the photon energy scale is assigned in this analysis, which accounts for the nonlinearity across the full range of photon p_T values.
3. *Shower shape corrections:* This uncertainty is associated with the imperfect modeling of shower shapes in simulation, and it is estimated by comparing the energy scale before and after any corrections are applied to the shower shape variables as described in Ref. [37]. This uncertainty in the energy scale is at most 0.01–0.15%, and it is dependent on the photon shower-shape and position in the detector.

4. *Nonuniformity of light collection:* Within a given ECAL crystal, there is an uncertainty associated with the modeling of the light collection as a function of the emission depth. This uncertainty is estimated by comparing simulation with the longitudinal shower profile estimates, and the procedure is detailed in Ref. [4]. The magnitude of this uncertainty is 0.07–0.25%, depending on the photon shower-shape.
5. *Modeling of material in front of the ECAL:* The behavior of electromagnetic showers is affected by the amount of material present in front of the ECAL. This behavior may not be well modeled in simulation, and thus special samples with variations in the amount of upstream material are used to compute the impact on the photon energy scale [45]. For most central photons, the magnitude of this uncertainty ranges 0.02–0.05% and increases to approximately 0.24% for photons in the endcaps.

The uncertainties that affect the normalization of the signal model are listed below.

1. *Integrated luminosity:* Uncertainties in the luminosity measurement are estimated to be 1.2% (2016), 2.3% (2017), and 2.5% (2018) by CMS luminosity monitoring [40–42]. The uncertainty in the total integrated luminosity of the three years together is 1.6%. The uncertainties for each dataset are partially correlated in order to account for the common sources in the luminosity measurement schemes.
2. *Photon identification BDT score:* The systematic uncertainty caused by the imperfect simulation of the input variables that are used to train the photon identification MVA is estimated by the procedure described in Ref. [37]. The average magnitude of the resulting uncertainty is below 0.25% across the full m_a range.
3. *Trigger efficiency:* The efficiency of the trigger selection is measured using a “tag-and-probe” technique on $Z \rightarrow ee$ events [63]. An additional uncertainty is introduced to account for a gradual shift in the timing to the inputs of the ECAL’s hardware level trigger in the region $|\eta| > 2.0$, which caused a specific trigger inefficiency during 2016–2017 data-taking [34]. The size of this uncertainty across the m_a range is around 0.5% for 2016 and 2018 data-taking and around 1.5% for 2017 data-taking.
4. *Photon preselections:* The systematic uncertainty on photon-based preselection is computed as the uncertainty on the ratio between the efficiency measured in data and in simulation. This is measured with the tag-and-probe technique using $Z \rightarrow ee$ events. The average magnitude of this uncertainty across the m_a range is about 5%.

A summary of all the systematic uncertainties considered in this analysis is given in Table 2. The impact of systematic uncertainties on the expected limit is about 1% across the m_a range, and so the analysis sensitivity is primarily limited by the expected signal event yields.

12 Results

The data collected by the CMS experiment in 2016, 2017, and 2018 are combined for the fit. The data and the signal-plus-background model that was fit to the $m_{\gamma\gamma\gamma\gamma}$ distribution are shown in Fig. 5 for $m_a = 15$ and 50 GeV.

No significant deviation from the background-only hypothesis is observed. Upper limits are set at the 95% confidence level (CL) on the product of the production cross section of the Higgs boson and the branching fraction into four photons via a pair of pseudoscalars, $\sigma_H \mathcal{B}(H \rightarrow$

Table 2: Summary of the systematic uncertainties considered in this analysis.

$m_{\gamma\gamma\gamma\gamma}$ distribution shape	2016–2018		
Photon energy scale and resolution	0.05–0.15%		
Nonlinearity of the photon energy scale	0.10%		
Shower shape corrections	0.01–0.15%		
Nonuniformity of light collection	0.07–0.25%		
Modeling of material in front of the ECAL	0.02–0.05% (EB) and 0.24% (EE)		
Signal model normalization	2016	2017	2018
Integrated luminosity	1.20%	2.30%	2.50%
Photon identification	0.25%	0.25%	0.25%
Trigger efficiency	0.50%	1.50%	0.50%
Photon preselections	5.00%	5.00%	5.00%

$aa \rightarrow \gamma\gamma\gamma\gamma$). This is done using the modified frequentist approach for CL_s , with the LHC profile likelihood ratio used as the test statistic [59, 64]. The observed (expected) limit, shown in Fig. 6, ranges from 0.80 (1.00) fb for $m_a = 15$ GeV to 0.26 (0.24) fb for $m_a = 62$ GeV. For comparison, the Higgs production cross section for all channels combined is 52 pb [65].

The results presented in this section are provided in a tabulated form in the HEPDATA record [66] for this analysis.

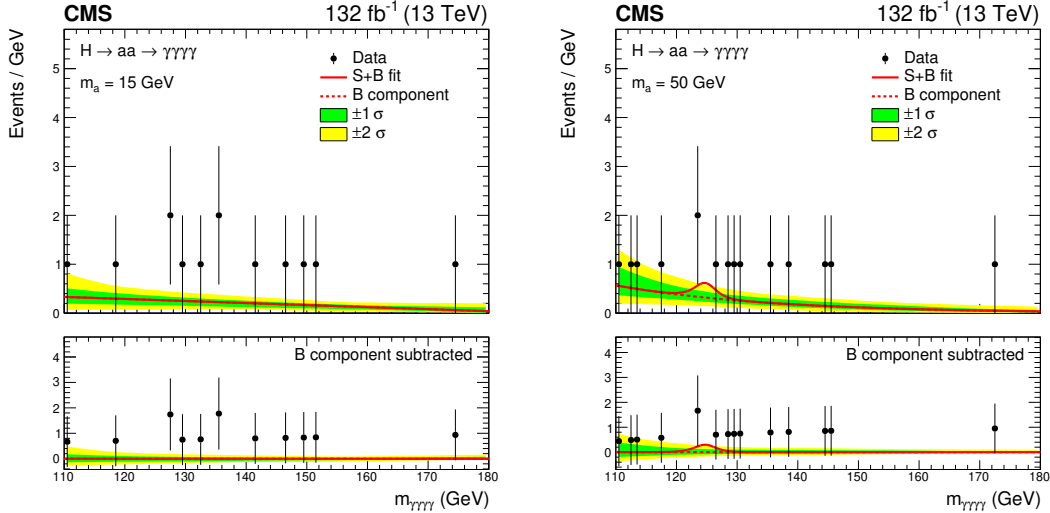


Figure 5: The invariant mass distribution, $m_{\gamma\gamma\gamma\gamma}$, for data (black points) and the signal-plus-background model fit is shown for $m_a = 15$ GeV (left) and $m_a = 50$ GeV (right). The solid red line shows the total signal-plus-background contribution, whereas the dashed red line shows the background component only. The lower panel in each plot shows the residual signal yield after the background subtraction. The one (green) and two (yellow) standard deviation bands include the uncertainties in the background component of the fit.

13 Summary

A search for a pair of light pseudoscalar bosons produced from the decay of the 125 GeV Higgs boson, which subsequently decay into photons, is presented. The analysis is based on proton-proton collision data collected at $\sqrt{s} = 13$ TeV by the CMS experiment at the LHC in 2016, 2017, and 2018, which corresponds to a total integrated luminosity of 132 fb^{-1} . The analysis probes

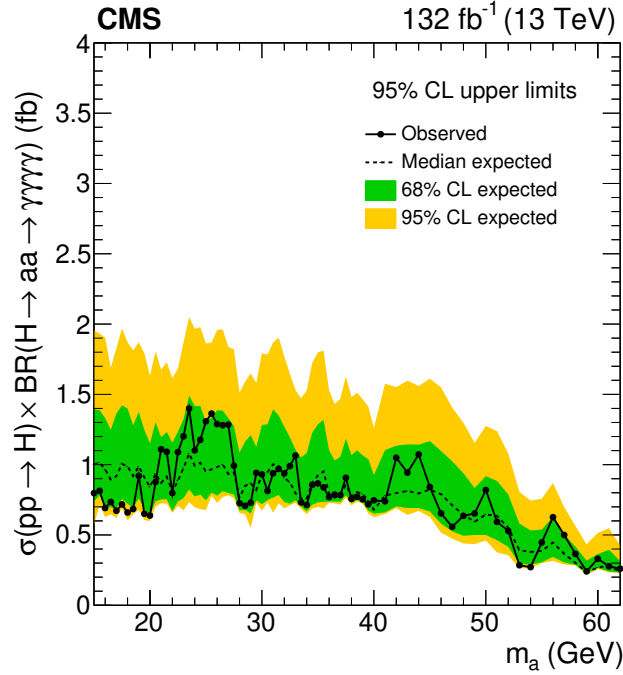


Figure 6: Expected and observed 95% CL limits on the product of the production cross section of the Higgs boson and the branching fraction into four photons via a pair of pseudoscalars, $\sigma_H \mathcal{B}(H \rightarrow aa \rightarrow \gamma\gamma\gamma\gamma)$, are shown as a function of m_a . The green (yellow) bands represent the 68% (95%) expected limit CL intervals. The fluctuation between individual points is due to the statistical limitation of the data sample and the result of individual BDT training networks utilized for each individual mass point scenario.

pseudoscalar bosons ranging in mass (m_a) from 15 to 62 GeV. No significant deviation from the background-only hypothesis is observed. Upper limits are set at 95% confidence level on the product of the production cross section of the Higgs boson and the branching fraction into four photons via a pair of pseudoscalars, $\sigma_H \mathcal{B}(H \rightarrow aa \rightarrow \gamma\gamma\gamma\gamma)$. The observed (expected) limits range from 0.80 (1.00) fb for $m_a = 15$ GeV to 0.26 (0.24) fb for $m_a = 62$ GeV.

References

- [1] 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** (2012) 1, doi:10.1016/j.physletb.2012.08.020, arXiv:1207.7214.
- [2] CMS Collaboration, “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”, *Phys. Lett. B* **716** (2012) 30, doi:10.1016/j.physletb.2012.08.021, arXiv:1207.7235.
- [3] CMS Collaboration, “Observation of a new boson with mass near 125 GeV in pp collisions at $\sqrt{s} = 7$ and 8 TeV”, *JHEP* **06** (2013) 081, doi:10.1007/JHEP06(2013)081, arXiv:1303.4571.
- [4] CMS Collaboration, “Measurements of Higgs boson production cross sections and couplings in the diphoton decay channel at $\sqrt{s} = 13$ TeV”, *JHEP* **07** (2021) 027, doi:10.1007/JHEP07(2021)027, arXiv:2103.06956.

- [5] CMS Collaboration, “Measurements of production cross sections of the Higgs boson in the four-lepton final state in proton-proton collisions at $\sqrt{s} = 13$ TeV”, *Eur. Phys. J. C* **81** (2021) 488, doi:10.1140/epjc/s10052-021-09200-x, arXiv:2103.04956.
- [6] CMS Collaboration, “Measurements of properties of the Higgs boson decaying into the four-lepton final state in pp collisions at $\sqrt{s} = 13$ TeV”, *Eur. Phys. J. C* **11** (2017) 047, doi:10.1007/JHEP11(2017)047, arXiv:1706.09936.
- [7] CMS Collaboration, “Measurement of the Higgs boson production rate in association with top quarks in final states with electrons, muons, and hadronically decaying tau leptons at $\sqrt{s} = 13$ TeV”, *Eur. Phys. J. C* **81** (2021) 378, doi:10.1140/epjc/s10052-021-09014-x, arXiv:2011.03652.
- [8] CMS Collaboration, “Measurements of $t\bar{t}H$ production and the CP structure of the Yukawa interaction between the Higgs boson and the top quark in the diphoton decay channel”, *Phys. Rev. Lett.* **125** (2020) 061801, doi:10.1103/PhysRevLett.125.061801, arXiv:2003.10866.
- [9] CMS Collaboration, “Measurement of the inclusive and differential Higgs boson production cross sections in the leptonic WW decay mode at $\sqrt{s} = 13$ TeV”, *JHEP* **03** (2021) 003, doi:10.1007/JHEP03(2021)003, arXiv:2007.01984.
- [10] CMS Collaboration, “A measurement of the Higgs boson mass in the diphoton decay channel”, *Phys. Lett. B* **805** (2020) 135425, doi:10.1016/j.physletb.2020.135425, arXiv:2002.06398.
- [11] ATLAS Collaboration, “Measurement of the Higgs boson mass in the $H \rightarrow ZZ^* \rightarrow 4\ell$ and $H \rightarrow \gamma\gamma$ channels with $\sqrt{s} = 13$ TeV pp collisions using the ATLAS detector”, *Phys. Lett. B* **784** (2018) 345, doi:10.1016/j.physletb.2018.07.050, arXiv:1806.00242.
- [12] ATLAS Collaboration, “Combined measurement of differential and total cross sections in the $H \rightarrow \gamma\gamma$ and the $H \rightarrow ZZ^* \rightarrow 4\ell$ decay channels at $\sqrt{s} = 13$ TeV with the ATLAS detector”, *Phys. Lett. B* **786** (2018) 114, doi:10.1016/j.physletb.2018.09.019, arXiv:1805.10197.
- [13] ATLAS Collaboration, “Measurement of the Higgs boson coupling properties in the $H \rightarrow ZZ^* \rightarrow 4\ell$ decay channel at $\sqrt{s} = 13$ TeV with the ATLAS detector”, *JHEP* **03** (2018) 095, doi:10.1007/JHEP03(2018)095, arXiv:1712.02304.
- [14] ATLAS and CMS Collaboration, “Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV”, *JHEP* **08** (2016) 045, doi:10.1007/JHEP08(2016)045, arXiv:1606.02266.
- [15] CMS Collaboration, “Combined measurements of Higgs boson couplings in proton-proton collisions at $\sqrt{s} = 13$ TeV”, *Eur. Phys. J. C* **79** (2019) 421, doi:10.1140/epjc/s10052-019-6909-y, arXiv:1809.10733.
- [16] D. Curtin et al., “Exotic decays of the 125 GeV Higgs boson”, *Phys. Rev. D* **90** (2014) 075004, doi:10.1103/PhysRevD.90.075004, arXiv:1312.4992.
- [17] ATLAS Collaboration, “Search for Higgs bosons decaying to aa in the $\mu\mu\tau\tau$ final state in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS experiment”, *Phys. Rev. D* **92** (2015) 052002, doi:10.1103/PhysRevD.92.052002, arXiv:1505.01609.

-
- [18] CMS Collaboration, “A search for pair production of new light bosons decaying into muons”, *Phys. Lett. B* **752** (2016) 146, doi:10.1016/j.physletb.2015.10.067, arXiv:1506.00424.
- [19] CMS Collaboration, “Search for a very light NMSSM Higgs boson produced in decays of the 125 GeV scalar boson and decaying into τ leptons in pp collisions at $\sqrt{s} = 8$ TeV”, *JHEP* **01** (2016) 079, doi:10.1007/JHEP01(2016)079, arXiv:1510.06534.
- [20] CMS Collaboration, “Search for light bosons in decays of the 125 GeV Higgs boson in proton-proton collisions at $\sqrt{s} = 8$ TeV”, *JHEP* **10** (2017) 076, doi:10.1007/JHEP10(2017)076, arXiv:1701.02032.
- [21] ATLAS Collaboration, “Search for the Higgs boson produced in association with a W boson and decaying to four b-quarks via two spin-zero particles in pp collisions at 13 TeV with the ATLAS detector”, *Eur. Phys. J. C* **76** (2016) 605, doi:10.1140/epjc/s10052-016-4418-9, arXiv:1606.08391.
- [22] ATLAS Collaboration, “Search for Higgs boson decays to beyond-the-Standard-Model light bosons in four-lepton events with the ATLAS detector at $\sqrt{s} = 13$ TeV”, *JHEP* **06** (2018) 166, doi:10.1007/JHEP06(2018)166, arXiv:1802.03388.
- [23] ATLAS Collaboration, “Search for Higgs boson decays into pairs of light (pseudo)scalar particles in the $\gamma\gamma j j$ final state in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”, *Phys. Lett. B* **782** (2018) 750, doi:10.1016/j.physletb.2018.06.011, arXiv:1803.11145.
- [24] ATLAS Collaboration, “Search for the Higgs boson produced in association with a vector boson and decaying into two spin-zero particles in the $H \rightarrow aa \rightarrow 4b$ channel in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”, *JHEP* **10** (2018) 031, doi:10.1007/JHEP10(2018)031, arXiv:1806.07355.
- [25] ATLAS Collaboration, “Search for Higgs boson decays into a pair of light bosons in the $b b \mu \mu$ final state in pp collision at $\sqrt{s} = 13$ TeV with the ATLAS detector”, *Phys. Lett. B* **790** (2019) 1, doi:10.1016/j.physletb.2018.10.073, arXiv:1807.00539.
- [26] CMS Collaboration, “Search for an exotic decay of the Higgs boson to a pair of light pseudoscalars in the final state with two b quarks and two τ leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV”, *Phys. Lett. B* **785** (2018) 462, doi:10.1016/j.physletb.2018.08.057, arXiv:1805.10191.
- [27] CMS Collaboration, “Search for an exotic decay of the Higgs boson to a pair of light pseudoscalars in the final state of two muons and two τ leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV”, *JHEP* **11** (2018) 018, doi:10.1007/JHEP11(2018)018, arXiv:1805.04865.
- [28] S. Chang, P. J. Fox, and N. Weiner, “Visible Cascade Higgs Decays to Four Photons at Hadron Colliders”, *Phys. Rev. Lett.* **98** (2007) 111802, doi:10.1103/PhysRevLett.98.111802, arXiv:hep-ph/0608310.
- [29] B. A. Dobrescu, G. L. Landsberg, and K. T. Matchev, “Higgs boson decays to CP odd scalars at the Tevatron and beyond”, *Phys. Rev. D* **63** (2001) 075003, doi:10.1103/PhysRevD.63.075003, arXiv:hep-ph/0005308.

- [30] F. Larios, G. Tavares-Velasco, and C. P. Yuan, “Update on a very light CP odd scalar in the two Higgs doublet model”, *Phys. Rev. D* **66** (2002) 075006, doi:10.1103/PhysRevD.66.075006, arXiv:hep-ph/0205204.
- [31] R. Dermisek and J. F. Gunion, “The NMSSM Solution to the Fine-Tuning Problem, Precision Electroweak Constraints and the Largest LEP Higgs Event Excess”, *Phys. Rev. D* **76** (2007) 095006, doi:10.1103/PhysRevD.76.095006, arXiv:0705.4387.
- [32] ATLAS Collaboration, “Search for new phenomena in events with at least three photons collected in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector”, *Eur. Phys. J. C* **76** (2016) 210, doi:10.1140/epjc/s10052-016-4034-8, arXiv:1509.05051.
- [33] CMS Collaboration, “The CMS trigger system”, *JINST* **12** (2017) P01020, doi:10.1088/1748-0221/12/01/P01020, arXiv:1609.02366.
- [34] CMS Collaboration, “Performance of the CMS Level-1 trigger in proton-proton collisions at $\sqrt{s} = 13$ TeV”, *JINST* **15** (2020) P10017, doi:10.1088/1748-0221/15/10/P10017, arXiv:2006.10165.
- [35] CMS Trigger, Data Acquisition Group Collaboration, “The CMS high level trigger”, *Eur. Phys. J. C* **46** (2006) 605, doi:10.1140/epjc/s2006-02495-8, arXiv:hep-ex/0512077.
- [36] CMS Collaboration, “Particle-flow reconstruction and global event description with the CMS detector”, *JINST* **12** (2017) P10003, doi:10.1088/1748-0221/12/10/P10003, arXiv:1706.04965.
- [37] CMS Collaboration, “Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC”, *JINST* **16** (2021) P05014, doi:10.1088/1748-0221/16/05/P05014, arXiv:2012.06888.
- [38] CMS Collaboration, “Performance of photon reconstruction and identification with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV”, *JINST* **10** (2015) P08010, doi:10.1088/1748-0221/10/08/P08010, arXiv:1502.02702.
- [39] CMS Collaboration, “The CMS Experiment at the CERN LHC”, *JINST* **3** (2008) S08004, doi:10.1088/1748-0221/3/08/S08004.
- [40] CMS Collaboration, “Precision luminosity measurement in proton-proton collisions at $\sqrt{s} = 13$ TeV in 2015 and 2016”, *Eur. Phys. J. C* **81** (2021) 800, doi:10.1140/epjc/s10052-021-09538-2, arXiv:2104.01927.
- [41] CMS Collaboration, “CMS luminosity measurement for the 2017 Data-Taking period at $\sqrt{s} = 13$ TeV”, CMS Physics Analysis Summary, CMS-PAS-LUM-17-004, 2018.
- [42] CMS Collaboration, “CMS luminosity measurement for the 2018 Data-Taking period at $\sqrt{s} = 13$ TeV”, CMS Physics Analysis Summary, CMS-PAS-LUM-18-002, 2019.
- [43] CMS Collaboration, “Search for a standard model-like Higgs boson in the mass range between 70 and 110 GeV in the diphoton final state in proton-proton collisions at $\sqrt{s} = 8$ and 13 TeV”, *Phys. Lett. B* **793** (2019) 320, doi:10.1016/j.physletb.2019.03.064, arXiv:1811.08459.

-
- [44] R. F. J. Alwall et al., “The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations”, *JHEP* **07** 79 (2014) doi:10.1007/JHEP07(2014)079, arXiv:1405.0301.
- [45] CMS Collaboration, “Observation of the diphoton decay of the Higgs boson and measurement of its properties”, *Eur. Phys. J. C* **74** (2014) 3076, doi:10.1140/epjc/s10052-014-3076-z, arXiv:1407.0558.
- [46] CMS Collaboration, “Measurements of Higgs boson properties in the diphoton decay channel in proton-proton collisions at $\sqrt{s} = 13$ TeV”, *JHEP* **11** (2018) 185, doi:10.1007/JHEP11(2018)185, arXiv:1804.02716.
- [47] T. Sjöstrand et al., “An introduction to PYTHIA 8.2”, *Comput. Phys. Commun.* **191** (2015) 159, doi:10.1016/j.cpc.2015.01.024, arXiv:1410.3012.
- [48] CMS Collaboration, “Event generator tunes obtained from underlying event and multiparton scattering measurements”, *Eur. Phys. J. C* **76** (2016) 155, doi:10.1140/epjc/s10052-016-3988-x, arXiv:1512.00815.
- [49] CMS Collaboration, “Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements”, *Eur. Phys. J. C* **80** (2020) 4, doi:10.1140/epjc/s10052-019-7499-4, arXiv:1903.12179.
- [50] GEANT4 Collaboration, “GEANT4 – a simulation toolkit”, *Nucl. Instrum. Meth. A* **506** (2003) 250, doi:10.1016/S0168-9002(03)01368-8.
- [51] M. Cacciari, G. P. Salam, and G. Soyez, “The anti- k_T jet clustering algorithm”, *JHEP* **04** (2008) 063, doi:10.1088/1126-6708/2008/04/063, arXiv:0802.1189.
- [52] M. Cacciari, G. P. Salam, and G. Soyez, “FastJet user manual”, *Eur. Phys. J. C* **72** (2012) 1896, doi:10.1140/epjc/s10052-012-1896-2, arXiv:1111.6097.
- [53] CMS Collaboration, “Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s} = 13$ TeV using the CMS detector”, *JINST* **14** (2019) P07004, doi:10.1088/1748-0221/14/07/P07004, arXiv:1903.06078.
- [54] P. D. C. Manzano et al., “Hemisphere Mixing: a Fully Data-Driven Model of QCD Multijet Backgrounds for LHC Searches”, *2017 European Physical Society Conference on High Energy Physics (EPS-HEP 2017)* **EPS-HEP2017** (2017) 370, doi:10.22323/1.314.0370, arXiv:1712.02538.
- [55] P. Baldi et al., “Parameterized neural networks for high-energy physics”, *Eur. Phys. J. C* **76** (2016) 235, doi:10.1140/epjc/s10052-016-4099-4, arXiv:1601.07913.
- [56] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, “Asymptotic formulae for likelihood-based tests of new physics”, *Eur. Phys. J. C* **11** (2011) 1554, doi:10.1140/epjc/s10052-011-1554-0, arXiv:1007.1727. [Erratum: *Eur.Phys.J.C* **73**, 2501 (2013), 10.1140/epjc/s10052-013-2501-z].
- [57] J. H. Friedman, “SMART User’s Guide”, Stanford University Department of Statistics Technical Report LCS.01, 1984.
- [58] J. H. Friedman, “A variable span scatterplot smoother”, Stanford University Department of Statistics Technical Report LCS.05, 1984.

- [59] CMS Collaboration, “Precise determination of the mass of the Higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8 TeV”, *Eur. Phys. J. C* **75** (2015), no. 5, 212, doi:10.1140/epjc/s10052-015-3351-7, arXiv:1412.8662.
- [60] T. Skwarnicki, “A study of the radiative CASCADE transitions between the Upsilon-Prime and Upsilon resonances”. PhD thesis, Cracow, INP, 1986.
- [61] P. D. Dauncey, M. Kenzie, N. Wardle, and G. J. Davies, “Handling uncertainties in background shapes: the discrete profiling method”, *JINST* **10** (2015) P04015, doi:10.1088/1748-0221/10/04/P04015, arXiv:1408.6865.
- [62] R. A. Fisher, “On the Interpretation of χ^2 from Contingency Tables, and the Calculation of p”, *J. Royal Stat. Soc* **85** (1922) 87, doi:10.2307/2340521.
- [63] CMS Collaboration, “Measurement of the inclusive W and Z production cross sections in pp collisions at $\sqrt{s} = 7$ TeV”, *JHEP* **10** (2011) 132, doi:10.1007/JHEP10(2011)132, arXiv:1107.4789.
- [64] CMS Collaboration, “Combined results of searches for the standard model higgs boson in pp collisions at $\sqrt{s} = 7$ TeV”, *Phys. Lett. B* **26** (2012) 710, doi:10.1016/j.physletb.2012.02.064, arXiv:1202.1488.
- [65] LHC Higgs Cross Section Working Group, “Handbook of LHC Higgs cross sections: 4. deciphering the nature of the Higgs sector”, CERN (2016) doi:10.23731/CYRM-2017-002, arXiv:1610.07922.
- [66] HEPData record for this analysis, 2022. doi:10.17182/hepdata.113445.