An embedding technique to determine $\tau\tau$ backgrounds in proton-proton collision data

The CMS collaboration

E-mail: cms-publication-committee-chair@cern.ch

ABSTRACT: An embedding technique is presented to estimate standard model $\tau\tau$ backgrounds from data with minimal simulation input. In the data, the muons are removed from reconstructed $\mu\mu$ events and replaced with simulated tau leptons with the same kinematic properties. In this way, a set of hybrid events is obtained that does not rely on simulation except for the decay of the tau leptons. The challenges in describing the underlying event or the production of associated jets in the simulation are avoided. The technique described in this paper was developed for CMS. Its validation and the inherent uncertainties are also discussed. The demonstration of the performance of the technique is based on a sample of proton-proton collisions collected by CMS in 2017 at $\sqrt{s} = 13$ TeV corresponding to an integrated luminosity of 41.5 fb$^{-1}$.

KEYWORDS: Pattern recognition, cluster finding, calibration and fitting methods; Performance of High Energy Physics Detectors

ArXiv ePrint: 1903.01216
Contents

1 Introduction 1

2 The CMS detector 2

3 Event reconstruction 3

4 Simulation 6

5 Embedding procedure 6
   5.1 Selection of $\mu\mu$ events 7
      5.1.1 Selection requirements 8
      5.1.2 Expected sample composition 9
      5.1.3 Correction for the detector acceptance 10
   5.2 Removal of $\mu$ energy deposits from the reconstructed event record 10
   5.3 Simulation of tau lepton decays 11
      5.3.1 Post-processing of the simulated tau lepton decays 12
      5.3.2 Discussion of additional reconstruction effects 13
   5.4 Hybrid event creation 14

6 Validation of the method 14
   6.1 Validation using the $\mu$-embedding technique 17
   6.2 Validation using the $e$-embedding technique 19
   6.3 Validation using the $\tau$-embedding technique 19

7 Application of the $\tau$-embedding technique to data 23
   7.1 Correction factors 23
   7.2 Uncertainties 25
   7.3 Comparison to data 25

8 Summary 30

A Performance of the $\tau$-embedding method on data 32

The CMS collaboration 37
1 Introduction

An important background for many measurements at the CERN LHC is the decay of $Z$ bosons into pairs of tau leptons ($Z \rightarrow \tau\tau$). Among those measurements are studies of Higgs boson events in the $\tau\tau$ [1–5] and $WW$ [6, 7] decay channels, and searches for additional supersymmetric and charged Higgs bosons [3, 8–13]. This background can be estimated from observed events, using selected $Z$ boson events in the $\mu\mu$ final state ($Z \rightarrow \mu\mu$). Initially, the method was only used to model events originating from $Z \rightarrow \tau\tau$ decays, which are the most prominent source of $\tau\tau$ background events at the LHC. However, all statements made throughout this paper are equally true for other standard model (SM) background processes that decay into two tau leptons. The aim of this method is to model all such processes.

In the embedding technique, all energy deposits of the recorded muons are removed from the $Z \rightarrow \mu\mu$ events collected by CMS and replaced by the energy deposits of simulated tau lepton decays with the same kinematic properties for the tau leptons as for the removed muons. In this way, a hybrid event is created, comprised of information from both observed and simulated events. The parts of an event that are challenging to describe in the simulation, such as the underlying event or the production of additional jets, are taken directly from observed data. Only the tau lepton decay, which is well understood, relies on the simulation. In Higgs boson analyses, the small coupling strength of the muon with respect to the tau lepton guarantees a negligible contamination by signal events. The $Z \rightarrow \mu\mu$ selection thus serves as a sideband region for those analyses that rely on this technique, referred to as target analyses in the following. In this picture, the simulation of the tau leptons in place of the removed muons corresponds to the extrapolation into the signal region.

The method itself can be studied by applying the embedding technique to a reference sample of simulated $Z \rightarrow \mu\mu$ events and comparing the result to an independent validation sample of simulated $Z \rightarrow \ell\ell$ events, where $\ell = e, \mu, \tau$ stands for the embedded lepton flavor. All lepton flavors are embedded for the validation of the technique. The corresponding application is referred to as $e$-, $\mu$-, or $\tau$-embedding throughout the text. The $\mu$-embedding holds the special role of validating the technique itself. The $e$-embedding serves to validate the sophisticated electron identification in CMS, which relies on many detector quantities. Reconstruction efficiencies are determined from each application, using the “tag-and-probe” method, as described in ref. [14]. This monitors the level of understanding of the reconstruction of each lepton flavor, and allows us to derive residual correction factors for final use in the target analyses. Since these correction factors are derived for the simulated leptons that have been embedded into the event, they are expected to be similar to the correction factors obtained without the embedding technique. The branching fractions for $Z \rightarrow ee$, $Z \rightarrow \mu\mu$, and $Z \rightarrow \tau\tau$ are equal so the normalizations for all the decays are equal.

The embedding technique was implemented successfully for the first time by the CMS Collaboration in the search and analysis of Higgs boson events in the context of the SM and its minimal supersymmetric extension (MSSM) based on the data set obtained during the first operational run of the LHC between 2009 and 2013 (Run-1) [3–6, 9, 10]. The technique has been upgraded since then to cope with the new challenges of the most recent LHC data-taking periods that are related to the increased proton-proton ($pp$) collision rate. Further developments of the method include (i) the inclusion of other processes than $Z \rightarrow \tau\tau$; (ii) the estimate of the normalization of the corresponding background processes from data; (iii) and an improved description of the electron identification.
The upgraded embedding technique served as a cross-check of the estimate of the $Z \rightarrow \tau\tau$ background events from simulation in the first CMS search for additional Higgs bosons in the $\tau\tau$ final state at 13 TeV, in the context of the MSSM [15]. A similar technique was used during the LHC Run 1 data-taking period by the ATLAS Collaboration [1, 2, 8] and is described in ref. [16]. In this paper, the methodology, validation, and application of the embedding technique developed for the CMS experiment are described. The data sample used for the demonstration of the technique has been recorded in 2017 and corresponds to an integrated luminosity of $41.5\,\text{fb}^{-1}$. The validation of the method is based on event samples that have been simulated for the same run period.

In sections 2 and 3 the CMS detector and event reconstruction are introduced. The production of simulated events used for the validation of the technique is described in section 4. In sections 5 and 6 the technique itself and its validation are discussed. Section 7 contains a demonstration of the performance of the technique, when applied to data, for the selection and analysis of $Z$ or Higgs boson events in the $\tau\tau$ final state. The paper is concluded with a brief summary in section 8.

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 are 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 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.

The silicon tracker measures charged particles within the pseudorapidity range $|\eta| < 2.5$. It consists of 1440 silicon pixel and 15 148 silicon strip detector modules. For nonisolated particles with a transverse momentum of $1 < p_T < 10\,\text{GeV}$ and $|\eta| < 1.4$, the track resolutions are typically 1.5% in $p_T$ and 25–90 (45–150) $\mu\text{m}$ in the transverse (longitudinal) impact parameter [17]. The electron momentum is estimated by combining the energy measurement in the ECAL with the momentum measurement in the tracker. The momentum resolution for electrons with $p_T \approx 45\,\text{GeV}$ from $Z \rightarrow ee$ decays ranges from 1.7% for nonshowering electrons in the barrel region to 4.5% for showering electrons in the endcaps [18]. Matching muons to tracks measured in the silicon tracker results in a relative transverse momentum resolution, for muons with $p_T$ up to 100 GeV, of 1% in the barrel and 3% in the endcaps. The $p_T$ resolution in the barrel is better than 7% for muons with $p_T$ up to 1 TeV [19]. 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 remaining barrel photons have a resolution of better than 2.5% for $|\eta| \leq 1.4$. In the endcaps, the resolution of unconverted or late-converting photons is about 2.5%, while the remaining endcap photons have a resolution between 3 and 4% [20]. When combining information from the entire detector, the jet energy resolution typically amounts to 15% at 10 GeV, 8% at 100 GeV, and 4% at 1 TeV, to be compared to about 40, 12, and 5% obtained when the ECAL and HCAL calorimeters alone are used.

Events of interest are selected using a two-tiered trigger system [21]. 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 $\mu\text{s}$. The second level, known
as the high-level trigger, consists of a large array 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.

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. [22].

### 3 Event reconstruction

The reconstruction of the pp collision products is based on the particle-flow (PF) algorithm described in ref. [23], which combines the available information from all CMS subdetectors to reconstruct an unambiguous set of individual particle candidates. The particle candidates are categorized into electrons, photons, muons, and charged and neutral hadrons. A good understanding of the CMS lepton reconstruction is an important prerequisite for the assessment of the embedding technique. Therefore the reconstruction of electrons, muons, and decays of tau leptons to hadrons (\(\tau_h\)) from charged and neutral PF candidates is discussed in more detail in this section.

In 2017, the CMS experiment operated with a varying instantaneous luminosity with, on average, between 28 and 47 pp collisions per bunch crossing. Collision vertices are obtained from reconstructed tracks using a deterministic annealing algorithm [24]. The reconstructed vertex with the largest value of summed physics-object \(p_T^2\) is the primary collision vertex (PV). The physics objects for this purpose are the jets, clustered using the anti-\(k_T\) jet finding algorithm [25, 26], as described below, with the tracks assigned to the vertex as inputs, and the associated missing transverse momentum calculated as the negative vector \(p_T\) sum of those jets. Any other collision vertices in the event are associated with additional soft inelastic pp collisions called pileup (PU).

Electrons are reconstructed by combining energy deposits in the ECAL with tracks obtained from hits in the tracker [18]. Due to the strong curvature of the trajectory of charged particles in the magnetic field and the significant amount of intervening material, an average fraction of 33% (at \(\eta \approx 0\)) to 86% (at \(|\eta| \approx 1.4\)) of the electron energy is radiated via bremsstrahlung before the electron reaches the ECAL. All energy deposits above noise thresholds are combined into clusters, using different algorithms for the ECAL barrel and endcap sections. The clusters are further grouped into superclusters in a narrow window in \(\eta\) and an extended window in the azimuthal angle \(\phi\) (measured in radians). The energy and position of the superclusters are obtained from the sum of the energies and the energy-weighted mean of the positions of the building clusters. This way of clustering is complemented by an alternative clustering algorithm, based on the PF-reconstruction algorithm [23], resulting in an independent collection of PF clusters.

Hits in the tracker are combined into tracks, using an iterative tracking procedure as described in ref. [23]. To be efficient for the reconstruction of electrons, the track finding must include the additional bending of the particle trajectory due to the bremsstrahlung emissions. This is achieved by a dedicated Gaussian-sum filter algorithm [27]. Since this method of track reconstruction can be time consuming, it is initiated only on a selected set of electron track seeds, which are likely to correspond to electron trajectories. Two approaches are followed to determine these seeds. In the first approach, starting from the ECAL, the energy and position of the superclusters are used to extrapolate the electron trajectory to its origin. The intersections of this extrapolation with the innermost tracker layers or discs are matched to hits in the corresponding detectors. In the second
approach, starting from the tracker, reconstructed tracks obtained from a less efficient, but also less CPU intensive, algorithm are extrapolated to the ECAL surface and matched to PF clusters. The seeds of both approaches are combined to initiate the final electron track finding with an efficiency of \( \geq 95\% \) for electrons from Z boson decays.

The combination of the electron tracks with the ECAL clusters is achieved via a matching of the track extrapolated to the ECAL surface with the supercluster in \( \eta-\phi \) space with an efficiency of \( \approx 93\% \) for electrons from Z boson decays. Alternatively, the electron track is matched to a PF cluster, while at each intersection with a layer or disc of the tracker a straight line is extrapolated to the ECAL surface, tangent to the electron trajectory, to identify further PF clusters due to bremsstrahlung emission. This approach improves the reconstruction for low \( p_T \) electrons and electrons in jets. To increase their purity, the reconstructed electrons are required to pass a multivariate electron identification discriminant [18], which combines information on the quality of the differently reconstructed tracks, shower shape, and kinematic quantities. In the target analyses, for which the embedding technique is primarily foreseen, working points of this discriminant with an efficiency between 80 and 90\% are used to identify electrons.

Two main approaches are also pursued to reconstruct muons with the CMS detector [19]: in the initial steps tracks are reconstructed independently in the inner silicon tracker and the outer track detectors of the muon system. In the first approach inner and outer tracks are matched by comparing their parameters propagated to a common surface. If a match is found, a global-muon track is fitted combining the hits from both tracks. In a second approach, tracks from the inner tracker are extrapolated to the muon system taking into account the magnetic field, the average expected energy losses, and multiple Coulomb scattering in the detector material. If at least one muon segment (i.e., a short track stub made of drift tube or cathode strip chamber hits) matches the extrapolation, the corresponding track is identified as a muon track. The second approach improves the reconstruction efficiency for muons with \( p_T \leq 5 \) GeV, which are unlikely to traverse the entire muon system. For muons within the geometrical acceptance and with sufficiently high \( p_T \) to reach the muon system, the reconstruction efficiency reaches up to 99\%. It is supplemented by specialized algorithms for muons with a \( p_T \) of several hundreds of GeV. The presence of hits in the muon chambers already leads to a strong suppression of particles misidentified as muons. Additional identification requirements on the track fit quality and the compatibility of individual track segments with the fitted track can reduce the misidentification rate further. In the analyses for which the embedding technique is primarily foreseen, muon identification requirements with an efficiency of about 99\% are chosen.

The contribution from nonprompt leptons to the electron (muon) selection is further reduced by requiring the selected leptons to be isolated from any hadronic activity in the detector. This property is quantified by a relative isolation variable

\[
I_{\text{rel}}^{\ell (\mu)} = \frac{1}{p_T^{\ell (\mu)}} \left[ \sum p_T^{\text{charged, PV}, i} + \max \left( 0, \sum E_{T,i}^{\text{neutral, PV}} - E_{T,i}^{\text{neutral, PU}} \right) \right], \tag{3.1}
\]

which uses the sum of the \( p_T \) of all charged and transverse energy of all neutral particles in a cone of radius \( \Delta R = \sqrt{\Delta \eta^2 + \Delta \phi^2} \) around the lepton direction at the PV, where \( \Delta \eta \) and \( \Delta \phi \) correspond to the angular distance of the particle to the lepton in the \( \eta \) and \( \phi \) directions. The chosen cone sizes are \( \Delta R = 0.3 \) and 0.4 for electrons and muons, respectively. The lepton itself is not included in
this calculation. To mitigate any distortions from PU, only those charged particles whose tracks are associated with the PV are included in the sum. The presence of neutral particles from PU around muons is estimated by summing the $p_T$ of charged particles in the isolation cone whose tracks have been associated with PU vertices and multiplying this quantity by a factor of 0.5 to account for the approximate ratio of neutral to charged hadron production, such that $E_T^{\text{neutral, PU}} = 0.5 \sum p_{T,l}^{\text{charged, PU}}$. For electrons, the FastJet technique [28, 29] is applied as described in ref. [18]. The energy of neutral particles from PU is estimated as $E_T^{\text{neutral, PU}} = \rho A_{\text{eff}}$, where $\rho$ is the median of the energy density distribution per area in the $\eta$-$\phi$ plane around any jet in the event and $A_{\text{eff}}$ is an effective area in $\eta$ and $\phi$. The value obtained is subtracted from the transverse energy sum, and the result set to zero in the case of negative values. Finally, the result is divided by the $p_T$ of the lepton to result in $I_{\text{rel}}^{(\mu)}$.

For further characterization of the event, all reconstructed PF candidates are clustered into jets using the anti-$k_T$ jet clustering algorithm as implemented in FastJet [25, 26] with a distance parameter of 0.4. To identify jets resulting from the hadronization of b quarks (b jets), a reoptimized version of the combined secondary vertex b tagging algorithm is used that exploits information from the decay vertices of long-lived hadrons and the impact parameters of charged-particle tracks in a combined discriminant [30]. A typical working point for analyses for which the embedding technique is foreseen corresponds to a b jet identification efficiency of $\approx 70\%$ and a misidentification rate for jets induced by light quarks and gluons of 1%. For the validation of the embedding technique, jets with $p_T > 20$ GeV and $|\eta| < 4.7$ and b jets with $p_T > 20$ GeV and $|\eta| < 2.5$ are used, unless otherwise indicated.

Jets are also used as seeds for the reconstruction of $\tau_h$ candidates. The $\tau_h$ reconstruction is performed by further exploiting the substructure of the jets, using the hadrons-plus-strips algorithm described in refs. [31, 32]. The decay into three charged hadrons, and the decay into a single charged hadron, accompanied by up to two neutral pions with $p_T > 2.5$ GeV, are used for the target analyses. The neutral pions are reconstructed as strips, i.e., clusters of electron or photon constituents of the seeding jet with stretched energy deposits along the azimuthal direction. The strip size varies as a function of the $p_T$ of the electron or photon candidate. The $\tau_h$ decay mode is then obtained by combining the charged hadrons with the strips. High-$p_T$ tau leptons are expected to be isolated from any hadronic activity in the event, as are high-$p_T$ electrons and muons. Furthermore, in accordance with its finite lifetime, the charged decay products of the tau lepton are expected to be slightly displaced from the PV. To distinguish $\tau_h$ decays from jets originating from the hadronization of quarks or gluons, a multivariate $\tau_h$ identification discriminant is used [32]. It combines information on the hadronic activity in the detector in the vicinity of the $\tau_h$ candidate with the reconstructed properties related to the lifetime of the tau lepton. Of the predefined working points given in ref. [32], the tight, medium, and very loose working points are used in the target analyses. These have efficiencies between $27\%$ (tight) and $71\%$ (very loose) for genuine tau leptons, e.g., from $Z \rightarrow \tau\tau$ decays, for quark/gluon misidentification rates of less than $4.4 \times 10^{-4}$ (tight), and $1.3 \times 10^{-2}$ (very loose). Finally, additional discriminants are imposed to reduce the misidentification probability for electrons and muons as $\tau_h$ candidates, using predefined working points from ref. [32]. For the discrimination against electrons these working points have identification efficiencies for genuine tau leptons ranging from $65\%$ (tight) to $94\%$ (very loose) for misidentification rates between $6.2 \times 10^{-4}$ (tight) and $2.4 \times 10^{-2}$ (very loose). For the discrimination against muons the typical $\tau_h$ identification efficiency is $99\%$ for a misidentification rate of $O(10^{-3})$. 


The missing transverse momentum vector \( \vec{p}_T^{\text{miss}} \), defined as the negative vector \( p_T \) sum of all reconstructed PF objects, is also used to characterize the events. Its magnitude is referred to as \( p_T^{\text{miss}} \). It enters the target analyses via selection criteria and via the calculation of the final discriminating variable used for the statistical analysis, which is usually correlated with the invariant mass of the \( \tau \tau \) system.

4 Simulation

For the validation of the embedding technique and to demonstrate its performance, simulated events are used to model the most important processes contributing after the event selections described in sections 5 and 7. The Drell-Yan production in the ee, \( \mu \mu \), and \( \tau \tau \) final states, and the production of W bosons in association with jets (W+jets) are generated at leading order (LO) precision [33] in the strong coupling constant \( \alpha_S \), using the MadGraph5_aMC@NLO 2.2.2 event generator [34].

To increase the number of simulated events in phase space regions with high jet multiplicity, supplementary samples are generated with up to four outgoing partons in the hard interaction. For diboson production MadGraph5_aMC@NLO is used at next-to-leading order (NLO) precision. For \( t\bar{t} \) and single top quark production samples are generated at NLO precision using POWHEG v2 [35–41]. For the generation of all processes the NNPDF3.0 parton distribution functions [42] are used. The simulation of the underlying event is parametrized according to the CUETP8M1 tune [43]. Hadronic showering and hadronization, as well as the \( \tau \) decays, are modeled using PYTHIA 8.212 [44]. For all generated events the effect of the PU is included by generating additional inclusive inelastic pp collisions with PYTHIA and adding them to the simulated events according to the expected PU distribution profile in data. Differences between this expectation and the observed PU profile are mitigated by reweighting the simulated events. All events generated are passed through a GEANT4-based [45] simulation of the CMS detector and reconstructed using the same version of the CMS event reconstruction software as used for the data.

5 Embedding procedure

The embedding procedure can be split into four steps:

- the selection of \( \mu \mu \) events from data (section 5.1),
- the removal of tracks and energy deposits of the selected muons from the reconstructed event record (section 5.2),
- the simulation of two \( \tau \) leptons with the same kinematic properties as the removed muons in an otherwise empty detector (section 5.3), and
- the combination of the energy deposits of the simulated tau lepton decays with the original reconstructed event record (section 5.4).

For validation purposes, electrons or muons can also be injected into the simulation to form an embedded ee or \( \mu \mu \) event, referred to as an e- or \( \mu \)-embedded event. A schematic view of the procedure is given in figure 1.
Figure 1. Schematic view of the four main steps of the \( \tau \)-embedding technique, as described in section 5. A \( Z \rightarrow \mu\mu \) candidate event is selected in data (“\( Z \rightarrow \mu\mu \) Selection”), all energy deposits associated with the muons are removed from the event record (“\( Z \rightarrow \mu\mu \) Cleaning”), and two tau lepton decays are simulated in an otherwise empty detector (“\( Z \rightarrow \tau\tau \) Simulation”). Finally all energy deposits of the simulated tau lepton decays are combined with the original reconstructed event record (“\( Z \rightarrow \tau\tau \) Hybrid”). In the example, one of the simulated tau leptons decays into a muon and the other one into hadrons.

5.1 Selection of \( \mu\mu \) events

In the first step of the embedding procedure, \( \mu\mu \) events are selected from data. Although the selected muons might not necessarily originate from \( Z \) boson decays, \( Z \rightarrow \mu\mu \) events are a natural target of this selection, which helps to identify genuine \( \mu\mu \) events. The selection should be tight enough to ensure a high purity of genuine \( \mu\mu \) events and at the same time loose enough to minimize biases of the embedded event samples. The selection of the muons defines the minimal selection requirements to be used in the target analyses that are discussed in more detail in section 5.3. Inefficiencies of the reconstruction and selection of the muons due to the geometrical acceptance of the detector are estimated, giving correction factors which are applied to the final distributions.

While strict isolation requirements help to increase the purity of prompt muons, e.g., from \( Z \rightarrow \mu\mu \) decays, in the selection, they introduce a bias towards less hadronic activity in the vicinities of the embedded leptons that will appear more isolated than expected in data. To minimize this kind
of bias, which cannot be corrected by a scale factor, isolation requirements are omitted as much as possible. At the same time the selected phase space is desired to be as inclusive as possible for the embedded event samples to be applicable for a variety of target analyses. The loose selection in turn leads to an admixture of other processes in addition to $Z \rightarrow \mu\mu$. This admixture and the consequences for the embedded event samples are carefully checked and assessed.

5.1.1 Selection requirements

At the trigger level, the events are required to be selected by at least one of a set of $\mu\mu$ trigger paths, with a minimum requirement between 3.8 and 8.0 GeV on the invariant mass of the two muons, $m_{\mu\mu}$. All trigger paths require $p_T > 17$ (8) GeV for the leading (trailing) muon, very loose isolation in the tracker, and a loose association of the muon track with the PV. Offline, the reconstructed muons are required to match the objects at the trigger level, their distance extrapolated to the PV is required to be $|d_z| < 0.2$ cm along the beam axis, and both muons are required to have $|\eta| < 2.4$. Their transverse momentum is required to be $p_T > 17$ (8) GeV for the leading (trailing) muon to match the online selection requirements. No additional selection requirements are imposed on the isolation of the muons to minimize any bias of the embedded event samples in this respect.

To form a $Z$ boson candidate, each muon is required to originate from a global-muon track. The muons are required to be of opposite charge with an invariant mass of $m_{\mu\mu} > 20$ GeV. If more than one $Z$ boson candidate is found in the event, the one with the value of $m_{\mu\mu}$ closest to the nominal $Z$ boson mass is chosen. This selection results in a total of more than 65 million events, with an average rate of about 1.5 million events per 1 fb$^{-1}$ of collected data. The expected event composition after these and several further selection requirements that will be specified in the following discussion is given in table 1. SM events composed exclusively of jets produced via the strong interaction are referred to as quantum chromodynamics (QCD) multijet production.

Table 1. Expected event composition after the selection of two muons, as described in section 5.1. The label “QCD” refers to SM events composed exclusively of jets produced via the strong interaction. The compositions after adding selections on $m_{\mu\mu} > 70$ GeV or on the number of $b$ jets in the event are shown in column 3 and 4 respectively. In the second column the fraction of events where the corresponding process has two genuine muons in the final state is given in parentheses. For $W$+jets events the second muon originates from additional heavy flavor production.

<table>
<thead>
<tr>
<th>Process</th>
<th>Inclusive</th>
<th>$m_{\mu\mu} &gt; 70$ GeV</th>
<th>$N(\text{b jet}) &gt; 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Z \rightarrow \mu\mu$</td>
<td>97.36 (97.36)</td>
<td>99.11</td>
<td>69.25</td>
</tr>
<tr>
<td>QCD</td>
<td>0.84</td>
<td>†</td>
<td>0.10</td>
</tr>
<tr>
<td>$t\bar{t}$</td>
<td>0.78 (0.60)</td>
<td>0.55</td>
<td>25.61</td>
</tr>
<tr>
<td>$Z \rightarrow \tau\tau$</td>
<td>0.74 (0.71)</td>
<td>0.05</td>
<td>0.57</td>
</tr>
<tr>
<td>Diboson, single t</td>
<td>0.20 (0.17)</td>
<td>0.17</td>
<td>2.35</td>
</tr>
<tr>
<td>$W$+jets</td>
<td>0.08 (0.01)</td>
<td>0.02</td>
<td>0.14</td>
</tr>
</tbody>
</table>

† Data-driven estimate, information not available.
Throughout the paper, this contribution is estimated from data using a background estimation method described in ref. [15]. The distributions of $m_{\mu\mu}$ and $p_T$ of the trailing muon for all selected events are shown in figure 2. Also shown are the contributing processes estimated by the simulation, to illustrate their kinematic distributions.

![Figure 2](image)

Figure 2. (Left) invariant mass, $m_{\mu\mu}$, of the selected dimuon Z boson candidates and (right) $p_T$ of the trailing muon after the event selection, as described in section 5.1.

### 5.1.2 Expected sample composition

In table 1, a relaxed selection of two muons compatible with the properties of a Z boson candidate already results in a sample of Z → $\mu\mu$ events with an expected purity of more than 97%. Smaller contributions are expected from Z → $\tau\tau$ events, mostly where both tau leptons subsequently decay into muons, and from QCD multijet, $t\bar{t}$, and diboson production.

Without further correction, the presence of QCD multijet and Z → $\tau\tau$ events in the selected event sample leads to an overestimate of the Z → $\mu\mu$ event yield and a bias of the $m_{\ell\ell}$ and $p_T$ distributions of the embedded leptons towards lower values. This can be inferred from figure 2, where the accumulation of these events is visible for $m_{\mu\mu} < 70$ GeV and $p_T < 20$ GeV. The fraction of QCD multijet and Z → $\tau\tau$ events can be significantly suppressed by raising the requirement on $m_{\mu\mu}$ to be higher than 70 GeV, at the cost of a loss of $\approx 13\%$ of selected Z → $\mu\mu$ events. However, because of the low transverse momentum of the selected muons, these events have a low probability to end up in the final sample of $\tau$-embedded events, see section 5.3.

The contribution from $t\bar{t}$ and diboson events is distributed over the whole range of $m_{\mu\mu}$. Its relative contribution is larger at high values of $m_{\ell\ell}$, where the overall event yield is small, and in event selections with b jets, as shown in the last column of table 1. These conditions are met, e.g., in searches for additional Higgs bosons in models beyond the SM [15]. A large fraction of this contribution originates from events where the W bosons e.g., from both $t$ quark decays subsequently decay into a muon and neutrino ($t\bar{t}(\mu\mu)$). The contribution from $t\bar{t}$ and diboson production in all other modes is below the current accuracy requirements of the method. The substitution of the
muons by tau leptons provides an additional estimate for $t\bar{t}$ and diboson production with two tau leptons in the final state from data. This class of events needs to be removed from simulation in the target analyses to prevent double counting. For simplicity, all further discussion of the embedding technique will refer to the estimate of all genuine $\tau\tau$ events from either $Z \rightarrow \tau\tau$, $t\bar{t}$, or diboson production, unless explicitly stated otherwise.

### 5.1.3 Correction for the detector acceptance

As discussed above, inefficiencies in the reconstruction and selection of the $\mu\mu$ events lead to kinematic biases in the embedded event samples because of the limited detector acceptance. The global efficiency of the trigger selection in the kinematic regime where embedded event samples can be applied amounts to about 80%, the combined reconstruction and identification efficiency lies well above 95%. Both efficiencies are estimated differentially in a fine grid in muon $\eta$ and $p_T$, using the “tag-and-probe” method. They are then used to correct for the effects of the detector acceptance.

As a consequence, not only the kinematic distributions but also the yield of the estimated $\tau\tau$ events can be obtained directly via the embedding technique, assuming the same branching fraction of the Z boson into muons and tau leptons. This is achieved by correcting for the detector acceptance and selection efficiency of the $\mu\mu$ events and applying the reconstruction and selection efficiency from the $\tau$-embedded event sample. Residual corrections of these efficiencies with respect to the data, are discussed in section 7.1. When applied to the data this estimate renders uncertainties in the production cross sections and integrated luminosity irrelevant for the involved processes, as will be further discussed in section 7.2.

### 5.2 Removal of $\mu$ energy deposits from the reconstructed event record

In the second step, all energy deposits of the selected muons are removed from the reconstructed event record. This is done at the level of hits in the inner tracker and muon systems, and clusters in the calorimeters. Hits in the tracker are identified by their association to the fitted global-muon track. Clusters in the calorimeters are identified by the intercept of the muon trajectory interpolated through the calorimeters, as discussed in section 3. If an intercept matches with the position of a calorimeter cluster, an energy amount corresponding to a minimum ionizing particle is subtracted from the cluster. If the energy of the modified cluster drops below the noise threshold defined for the event reconstruction, the cluster is removed from the event record. By this procedure, all traces of the selected muons in the detector can be removed from the event reconstruction even in detector environments with additional hadronic activity in the vicinity of the selected muons.

Effects of the removal of energy deposits in the calorimeters can arise in cases where the energy deposit of the muon is not completely removed or leads to the split of a geometrically extended cluster into more than one piece. Such a removal may lead to the reconstruction of spurious photon or neutral hadron candidates. These additionally reconstructed objects are usually of low energy and low reconstruction quality, and play a negligible role in the target analyses. The removal of the energy deposits of the muons from the detector is illustrated in figure 3. In figure 3 (left), a selected $Z \rightarrow \mu\mu$ candidate event in the data set is displayed in the $\eta$–$\phi$ plane of the calorimeters, with the intercepts of the reconstructed muons with the calorimeter surface and clusters in the ECAL (HCAL) shown. One muon (with $p_T = 32$ GeV) in the upper and one muon (with $p_T = 59$ GeV) in the lower parts of the figure are visible. Several clusters in the calorimeters have been associated.
with the incident muon trajectories. In figure 3 (right) the same detector area is shown after the hits and energy deposits associated with the muons have been removed from the reconstructed event record. The HCAL clusters associated with each corresponding muon have been completely removed, whereas the energy of the ECAL cluster associated with the muon in the lower part of the figure has been reduced. The remaining ECAL cluster is identified as low-energy photon in the subsequent reconstruction.

5.3 Simulation of tau lepton decays

In the third step, the energy and momentum of the selected muons are either directly injected as electrons or muons into the detector simulation, for validation purposes, or used to seed the simulation of tau lepton decays via _pythia_, before entering the detector simulation. For this purpose an event record is prepared that contains only the information related to the kinematic properties of the two selected muons in an otherwise empty detector that is free of any other particles from additional jet production, underlying event, or PU. The invariant mass of the selected muons is fixed to the reconstructed value, as shown in figure 2 (left). Polarization effects are neglected in embedded events, since they are below the sensitivity of the target analyses.

To account for the mass difference between the muon and the tau lepton or electron (referred to by $\ell = e, \tau$), the four-momenta of the muons are boosted into the center-of-mass frame of the $\mu\mu$ pair, where the energy ($E_\ell^*$) and momentum ($p_\ell^*$) of each lepton, with mass $m_\ell$, are determined from

$$E_\ell^* = \frac{m_{\mu\mu}}{2} \; ; \; \quad |\vec{p}_\ell^*| = \sqrt{E_\ell^*^2 - m_\ell^2} \; ; \; \quad \ell = e, \tau \; .$$

The corrected values $\vec{p}_\ell^*$ and $E_\ell^*$ are then boosted back into the laboratory frame and used either for the electrons or to seed the tau lepton decays. The event vertex for the simulation of the embedded
leptons is set to the PV of the initially reconstructed $\mu\mu$ event. Four distinct samples of $\tau$-embedded events are produced from the same $\mu\mu$ event sample, for use in the most important final states of the target analyses, namely $e\mu$, $e\tau_h$, $\mu\tau_h$, and $\tau_h\tau_h$. This is achieved by enforcing the subsequent decay of the injected $\tau$ lepton pair in the simulation, with a branching fraction of 100%. It has been checked that the overlap of the resulting $\tau$-embedded event samples is small enough, such that even those distributions that are related to the part of the event that originates from the observed data, e.g. like jet distributions, are fully uncorrelated.

5.3.1 Post-processing of the simulated tau lepton decays

A significant amount of the energy and momentum of the tau lepton is not transferred to the visible decay products, but carried away by the neutrino(s) in the decay. As a consequence, the $A$ significant amount of the energy and momentum of the tau lepton is not transferred to the visible decay products, but carried away by the neutrino(s) in the decay. As a consequence, the

To increase the number of $\mu\tau$ events that can be used in the target analyses, the decay is repeated 1000 times for each tau lepton pair. This is done to give the decay products a higher probability to pass the eligibility requirements. Only the last trial that fulfills the kinematic requirements for the given final state is saved for the subsequent detector simulation. If at least one trial succeeds, the number of successful trials divided by 1000 times the branching fraction of the subsequent $\tau\tau$ decay is saved as an additional weight factor to the event. These weights take values below $10^{-4}$ at the kinematic thresholds of $20 \text{GeV}$ each is useless for an analysis with a trigger threshold of $p_T^{h} > 30 \text{GeV}$. To save computing time during the CPU-intensive detector simulation, a kinematic filtering is applied to the visible decay products, after the simulation of the tau lepton decay and before the detector simulation. The final-state-dependent thresholds of this filtering on the $p_T$ of the visible decay products (prior to the detector simulation) define the kinematic range of eligibility of the $\tau$-embedded event samples for later use in the target analyses. They are given in table 2.

Table 2. Kinematic range of eligibility for each $\tau$-embedded event sample in the $e\mu$, $e\tau_h$, $\mu\tau_h$, and $\tau_h\tau_h$ final states. The expression “First/Second object” refers to the final state label used in the first column. Also given are the probability of the simulated tau lepton pair to pass the kinematic filtering ($\epsilon_{\text{kin}}$), described in the text, and the equivalent of the integrated luminosity $L_{\text{int}}$, of the corresponding $\tau$-embedded event sample, in multiples of the data set, from which the embedded event sample has been created.

<table>
<thead>
<tr>
<th>Final state</th>
<th>First object</th>
<th>Second object</th>
<th>$\epsilon_{\text{kin}}$</th>
<th>$L_{\text{int}}/41.5 \text{ fb}^{-1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$e\mu$</td>
<td>$p_T^e &gt; 21 \text{ (10 GeV)}$</td>
<td>$p_T^\mu &gt; 10 \text{ (21 GeV)}$</td>
<td>0.58</td>
<td>60</td>
</tr>
<tr>
<td>$e\tau_h$</td>
<td>$p_T^e &gt; 22 \text{ GeV, }</td>
<td>\eta^e</td>
<td>&lt; 2.2$</td>
<td>$p_T^{\tau_h &gt; 18 \text{ GeV, }</td>
</tr>
<tr>
<td>$\mu\tau_h$</td>
<td>$p_T^\mu &gt; 18 \text{ GeV, }</td>
<td>\eta^\mu</td>
<td>&lt; 2.2$</td>
<td>$p_T^{\tau_h &gt; 18 \text{ GeV, }</td>
</tr>
<tr>
<td>$\tau_h\tau_h$</td>
<td>$p_T^{\tau_h &gt; 33 \text{ GeV, }</td>
<td>\eta^{\tau_h}</td>
<td>&lt; 2.2}$</td>
<td>$p_T^{\tau_h &gt; 33 \text{ GeV, }</td>
</tr>
</tbody>
</table>
eligibility. Depending on the $\tau\tau$ final state, the fraction of events that pass the kinematic filtering ranges between $\epsilon_{\text{kin}} = 27\%$ (in the $\tau_h\tau_h$ final state) and $58\%$ (in the $e\mu$ final state). In the $\tau_h\tau_h$ final state this means that $73\%$ of the $\tau$-embedded events that could in principle be used, according to the acceptance restrictions of the originally selected $\mu\mu$ events, are usually not accessible due to the stricter acceptance requirements in the target analyses.

Overall this procedure allows for the production of final-state-specific $\tau$-embedded event samples of approximately 5 to 60 times the size of the event sample of selected tau lepton pairs in the target analyses, independent of the integrated luminosity corresponding to this event sample. The efficiency of the kinematic filtering and the size of each $\tau$-embedded event sample are given in table 2.

In section 5.1.2, $Z \rightarrow \tau\tau$ events where both tau leptons subsequently decay into muons and the corresponding neutrinos are discussed as a potential source of bias of the $\tau$-embedded event samples. Of all $Z \rightarrow \tau\tau$ events in this final state a fraction of less than $0.25\%$ is expected to end up in the $\tau$-embedded event samples, in the given eligibility ranges. This corresponds to less than $2.8\%$ of the events indicated by the $Z \rightarrow \tau\tau$ contribution in figure 2, and a fraction far below the $1\%$ level in the initial event composition as given in table 1.

### 5.3.2 Discussion of additional reconstruction effects

Two more reconstruction effects arise in the discussion of the simulation step. First, the four-momenta of the selected muons correspond to already reconstructed objects, which are reinjected into the simulation of the detector response, effects due to the finite momentum resolution of the detector lead to a broadening, especially of the $p_T$ and $m_{\ell\ell}$ distributions of the embedded leptons. The distributions are corrected for this effect by an $m_{\mu\mu}$-dependent rescaling of the energy and

![Figure 4.](image-url)
momentum of the selected muons on an event-by-event basis, before using them to generate the simulated leptons for embedding. A simulated $Z \rightarrow \mu\mu$ sample is used to derive this $m_{\mu\mu}$-dependent rescaling. Figure 4 (left) shows the $m_{\mu\mu}$ distribution from a sample of simulated $Z \rightarrow \mu\mu$ events as well as the corresponding $\mu$-embedded event sample before and after the correction. In the lower panel of the figure, the ratio is given with respect to the simulated $Z \rightarrow \mu\mu$ sample. The $\mu$-embedded event sample without the correction reveals a slight broadening with respect to the simulated $Z \rightarrow \mu\mu$ sample, which is compensated by the correction.

A second effect can be attributed to the emission of photons from the initially selected muons, referred to as final-state radiation (FSR) in the following. When missed in the reconstruction, FSR leads to an additional broadening of the kinematic distributions and a systematic shift to lower values of the energy and momentum of the initially selected muons. This shift is subsequently transferred to the embedded leptons. Figure 4 (right) shows the $m_{\mu\mu}$ distribution of the $Z \rightarrow \mu\mu$ simulation sample for muons before and after FSR, to illustrate the effect. For the validation of $\mu$-embedded events, this effect can be eliminated by executing the simulation step of the embedding procedure without FSR. The $Z \rightarrow \mu\mu$ simulation sample and the corresponding $\mu$-embedded event data sample are then subjected to the same FSR effects during the initial simulation. For $e$-embedded events the effects of FSR are underestimated; for $\tau$-embedded events they are overestimated.

In the case of $\tau$-embedding, both effects that were discussed in this section are negligible compared to the energy and momentum fluctuations introduced by the undetected neutrinos in the decay, which already lead to a significant broadening of the related kinematic distributions. A more detailed discussion is given in section 6.

5.4 Hybrid event creation

In a fourth and final step of the procedure, all energy deposits of the simulated electrons, muons, or tau lepton decays are combined with the original reconstructed event record, from which the energy deposits of the initially selected muons had been removed, to form a hybrid event that is mostly obtained from data and only relies on the simulation for the embedded lepton pair. This is done at the earliest possible reconstruction step to guarantee that all subsequent quantities for the lepton identification are based on the full event information and not only on parts of the event. The ideal way is to combine the reconstructed object collections at the level of tracker hits and energy deposits in the calorimeter crystals. However, in practice, the information is combined at the level of reconstructed objects (tracks, calorimeter clusters, and muons) rather than at the level of individual hits. This is to avoid complications with residual small differences between the simulation geometry and the real detector. The tracks of the embedded leptons are reconstructed based on the geometry used for the simulation, in the otherwise empty detector, of the simulation step. Since the detector in the simulation step is free from other particles, jet production, underlying event, or PU there may be a biased track reconstruction efficiency that must be checked and possibly corrected. Residual effects are discussed in section 6.

6 Validation of the method

Simulation-based closure tests are performed to test the validity of the embedding method. For this purpose, a validation sample for embedded events is created from simulated $Z \rightarrow \mu\mu$ events,
Figure 5. Comparison of μ-embedded events with exactly the same $Z \rightarrow \mu\mu$ events from simulation. Shown are the (upper left) $\eta$ and (upper right) $p_T$ distributions of the leading muon in $p_T$, (middle left) $p_T^{miss}$, (middle right) $m_{jj}$, (lower left) jet and, (lower right) b jet multiplicities, as described in the text.
Figure 6. Comparison of $\mu$-embedded events with exactly the same $Z \rightarrow \mu\mu$ events from simulation. Shown is the mean transverse momentum (energy) flux per muon, from all reconstructed particles with the distance $R$ from the muon, split by (upper left) charged hadrons from the PV and (upper right) PU vertices, (lower left) photons, and (lower right) neutral hadrons. The distributions are shown for the $\mu^-$ and for events with $m_{\mu\mu}$ close to the nominal $Z$ boson mass.

in which the embedding technique is applied in the same way as in the observed data: the selected muons are removed from the reconstructed event record and replaced with electrons, muons, or tau leptons. The embedded event data samples created in this way are compared to simulated events in the same final states. For $e$- and $\tau$-embedded events, this comparison is performed on statistically independent event samples. For $\mu$-embedded events, the comparison is performed on exactly the same simulated events, such that only the effects of the removal of energy deposits of the initially selected muons, and the reconstruction of the reinjected muons are tested.

For $e$- and $\tau$-embedded events, the normalization of the distributions is obtained from the yield of selected $Z \rightarrow \mu\mu$ events in the first step of the procedure, as described in section 5.1. For the $\tau$-embedded events, the yield of selected $\tau\tau$ events matches the yield of the simulated $Z \rightarrow \mu\mu$ sample within 1% with a statistical uncertainty of 0.5%. For the $e$-embedded events a similar agreement is achieved.
6.1 Validation using the $\mu$-embedding technique

The muon plays a special role in validating the embedding procedure itself. The broadening of the kinematic distributions of the embedded muons, due to the repeated reconstruction and the finite angular and $p_T$ resolution of the detector, and the effects of FSR, have already been discussed in section 5.3. For the following discussion, the simulation of FSR is switched off in the simulation step of the embedding procedure. In this way FSR is simulated only once, during the initial simulation of the validation sample, and all FSR effects are the same for the simulated and the embedded event.

Figure 5 shows the $\eta$ and $p_T$ distributions of the leading muon in $p_T$, the $p_T^{\text{miss}}$, the invariant mass of the two leading jets in $p_T$, $m_{jj}$, the number of jets with $p_T > 30$ GeV and $|\eta| < 4.7$, and the number of b jets with $p_T > 20$ GeV and $|\eta| < 2.5$. The blue dots correspond to the $\mu$-embedded event sample and the red histogram to the original simulation. The red-shaded bands represent the statistical uncertainty of the simulated event sample that is a reference for the comparison. All distributions are based on exactly the same events, so that the observed differences can exclusively be attributed to the removal and repeated simulation and reconstruction of the embedded muons. The uncertainty bands are added to facilitate the assessment of the observed differences between the compared samples. These differences are considered acceptable if they are compatible with the statistical uncertainty of the validation sample, which is chosen with 10 times more events than the expected number of events in the target analyses.

The kinematic distributions of the muons and jets, and the jet multiplicities are well reproduced. The structure in the distributions of the muon $\eta$ follows the geometry of the detector. The Jacobian peak corresponding to the Z boson decay is clearly visible in the $p_T$ distribution of the muon. A 5% effect in the ratio is visible for low values of $p_T^{\text{miss}}$, which is caused by the finite angular and $p_T$ resolution of the detector that can lead to small residual values of $p_T^{\text{miss}}$ for events with little or no $p_T^{\text{miss}}$. Corrections due to the finite momentum resolution of the detector, as described in section 5.3, are not propagated to the $p_T^{\text{miss}}$. For $\tau$-embedded events this effect is negligible compared to the kinematic fluctuations related to the neutrinos involved in the decays, as will be discussed in section 6.3. Another 5% effect in the ratio for $p_T^{\text{miss}} > 100$ GeV is explained by rare reconstruction effects, where muons of high $p_T$ may create additional track segments, e.g., due to multiple scattering in the outer tracker, which are not associated with the initially reconstructed global muon track. After the cleaning step of the embedding procedure, such track segments may be picked up in a different way and thus lead to a different assignment of $p_T^{\text{miss}}$. Since the validation is based on simulated $Z \to \mu\mu$ events, without genuine $p_T^{\text{miss}}$, it is clear that such events point to a poor reconstruction of the original event. The fact that this is a 5% effect only for a small fraction of events, and that the size of the effect is small compared to the statistical uncertainty of the validation sample, indicates that it is subdominant to the effect at low $p_T^{\text{miss}}$.

Figure 6 shows the mean transverse momentum flux per muon, $\langle \Delta p_T \rangle$, from all reconstructed particles within the distance $R$ from the muon, split by charged hadrons originating from the PV and PU vertices, photons, and neutral hadrons. It is defined as the average sum of the $p_T$ (transverse energy in case of neutral particles) of all corresponding particles between two cones with radii $R$ and $R+\Delta R$ in the distance $R$ from the muon, where $\Delta R$ corresponds to the widths of the histogram bins. All distributions are shown for the $\mu^-$ for events with $m_{\mu\mu}$ close to the nominal Z boson mass.
Figure 7. Comparison of e-embedded events with a statistically independent sample of simulated $Z \rightarrow ee$ events. Shown are distributions of the energy-weighted standard deviations of a $5 \times 5$ crystal array in (upper left) $\eta$, $\sigma_{\eta,i,j}$ and (upper right) $\phi$, $\sigma_{\phi,i,j}$, as described in the text, (lower left) the number $N_{\text{GSF}}$ of detector hits, used for the Gaussian Sum Filter algorithm [27] as described in section 3, and (lower right) the multivariate discriminator for the identification of electrons (electron-ID BDT). The black arrow, shown in addition to the electron-ID BDT distribution, indicates the working point with 80% efficiency in the displayed electron $\eta$ region. For better visibility, the statistical uncertainties of both samples, red-shaded band for simulated $Z \rightarrow ee$ events, and blue vertical bars for e-embedded events, are multiplied by 10 for the figures.

The figures indicate that in most cases no other particles are reconstructed in the spatial vicinity of the muon. For a uniform $p_T$ flux distribution, $\langle \Delta p_T \rangle$ is expected to increase linearly, because of the increasing area of the ring segments. This trend is roughly observed for all reconstructed particle types with a slope of 32 (550) MeV per unit of $R$ for $\langle \Delta p_T \rangle$ from charged hadrons originating from the PV (PU vertices), 110 MeV for photons, and 66 MeV for neutral hadrons. The larger slope for charged hadrons from PU vertices, photons, and neutral hadrons is related to the simulated PU profile and may vary in data. The displayed distributions are shown for the simulated PU profile between 40 and 70 additional inelastic pp collisions. For charged hadrons and photons, the progression from the simulation is well reproduced, apart from small regions close to the muon,
which show a small excess in \(\langle \Delta p_T \rangle\) for charged hadrons from the PV and photons, and a small deficit in \(\langle \Delta p_T \rangle\) for charged hadrons from PU vertices. A larger difference is observed for neutral hadrons, which is due to an incomplete removal of energy deposits of the muon in the HCAL, as discussed in section 5.2. When integrated over \(R\), and all reconstructed particle types, the additional hadronic energy in the predefined isolation cone adds up to less than 200 MeV.

### 6.2 Validation using the e-embedding technique

The identification of electrons in CMS is based on \(O(20)\) closely related detector variables that are combined into a multivariate discriminator [18]. As discussed in sections 5.3 and 5.4 the simulation of the embedded lepton pair takes place in an otherwise empty detector with no other particles from PU, underlying event, or additional jet production. The tight relation of the electron reconstruction and identification to closely related detector quantities poses an extra challenge to the embedding technique for this lepton flavor, which therefore requires a unique validation procedure. To monitor the success in simulating the distribution of this discriminator and its inputs, e-embedded events are created and compared to a statistically independent sample of simulated \(Z \rightarrow ee\) events. Figure 7 shows, for the leading electron in \(p_T\), the energy-weighted standard deviation of the position of a 5\times5 ECAL crystal array in \(\eta\) \((\sigma_{i\eta}).\) and \(\phi\) \((\sigma_{i\phi}).\) and \(N_{GSF}\) the number of detector hits used for the Gaussian Sum Filter algorithm [27] that is introduced in section 3. The quantities \(i\eta\) and \(i\phi\) are measured in integer crystal units, such that in a 5\times5 array a peripheral crystal can be one or two units away from the central crystal in the array. All quantities are in reasonable agreement given their high sensitivity to the exact geometry, intercalibration, and level of noise suppression of the detector. Also shown is the multivariate discriminator itself (output of the electron-ID boosted decision tree (BDT)), which, among others, has the discussed quantities as input. The vertical arrow added to figure 7 (lower right) corresponds to the 80% working point for the electron identification. Residual differences in the distributions of the electron-ID BDT are comparable to the differences between data and simulation. Correction factors for these differences are derived and applied to the \(\tau\)-embedded event samples, and are described in section 7.1. In figure 8, the distributions of \(m_{ee}\) and the \(p_T\) of the leading electron are shown. The observed differences are explained by differences in FSR, as discussed in section 5.3. Also shown is the effect of a variation of the electron energy scale by \(\pm 1\%\), which is usually applied to the target analyses and fully covers the effect.

### 6.3 Validation using the \(\tau\)-embedding technique

The main target of the embedding technique, the estimation of \(Z \rightarrow \tau\tau\) events is validated by comparing \(\tau\)-embedded events to a statistically independent sample of simulated \(Z \rightarrow \tau\tau\) events in each of the previously discussed \(\tau\tau\) final states. In figure 9 the \(p_T\) and \(\eta\) distributions of the electron, muon, and \(\tau\) candidate are shown using the \(e\mu\), \(e\tau\), and, \(\mu\tau\) final states. To increase the statistical significance of the validation results, the distributions of the purely lepton related quantities are shown for the combination of multiple final states. Figure 10 shows the distributions of the electron and muon isolation, \(I_{rel}(\mu)\), the multivariate \(\tau\) discriminant \((\tau\text{-ID BDT}), \(p_T^{miss}\), \(m_{jj}\), and the invariant mass of the visible decay products of the tau leptons, \(m_{vis}\) in the \(\mu\tau\) final state. The \(\tau\)-embedded event samples, by construction, have a larger size than the simulated validation sample and thus smaller statistical uncertainties, which becomes apparent from the smaller fluctuations, especially in the tails of the steeply falling distributions in the upper panels of the subfigures.
Figure 8. Comparison of the $e$-embedded events with a statistically independent sample of simulated $Z \rightarrow ee$ events. Shown are the distributions of (left) $m_{ee}$ and (right) $p_T$ of the leading electron in $p_T$. The blue vertical bars and red-shaded bands correspond to the statistical uncertainty of each sample. The effect of a variation of the electron energy scale of $\pm 1\%$ is also shown by the green lines.

In general, a good agreement is observed, within the statistical precision. Effects of FSR in the selection of the $\mu\mu$ event are not visible in the muon $p_T$ and $m_{vis}$ distributions. This is true for all $\tau\tau$ final states under investigation. Also shown for these distributions are the effects of a shift of the electron energy scale by $\pm 1\%$ and a shift of the tau lepton energy scale by $\pm 1.2\%$, corresponding to the uncertainties usually applied to the target analyses. Differences in the electron and muon $\eta$ are covered by the additional uncertainties in the correction for the geometrical $\mu\mu$ detector acceptance. Potential differences in the electron $p_T$ are small compared to the electron energy scale uncertainty usually applied to the target analyses, as discussed above. The effect of a corresponding shift in the electron energy scale is also shown in the corresponding subfigure. The same is true for the $p_T$ of the $\tau_h$ candidate. More pronounced deviations are visible in the $I_{rel}^\mu$ distribution. These are explained by an incomplete removal of the energy deposits of the initially selected muons. Integrated over the full isolation cone, the expected difference in $p_T$ amounts to less than 200 MeV, corresponding to the excess in $\langle \Delta p_T \rangle$, as observed in the context of the discussion of figure 6. The fact that similar effects are not visible in $I_{rel}^e$ can be explained by the different reconstruction of electrons that may associate parts of the remaining energy deposits of the initially selected muons in the calorimeters to the electron clusters, thus removing them from the objects taken into account for the calculation of $I_{rel}^e$. A 20% difference in the highest bin of the $\tau_h$-ID BDT distribution is explained by the reconstruction of tracks in the otherwise empty detector in the simulation step, for $\tau_h$ decays with one or three charged and no additional neutral hadrons. The overall effect on the identification efficiency is small and included in corresponding correction factors that are discussed in section 7.1.

In summary, in all investigated Drell-Yan final states, the agreement of the embedded event samples with the corresponding validation sample is observed to be compatible with the simulation. Most of the observed differences are within the statistical precision of the validation sample and smaller than the statistical precision of the target analyses in the $\tau\tau$ final state. Residual systematic
Figure 9. Comparison of $\tau$-embedded events with a statistically independent sample of simulated $Z \rightarrow \tau\tau$ events. Shown are the (left) $\eta$ and (right) $p_T$ distributions of the (upper row) electron in the $e\mu+\tau_h$ final states, (middle row) muon in $e\mu+\mu\tau_h$ final states, and (lower row) $\tau\eta$ candidate in the $e\tau_h+\mu\tau_h$ final states. The blue vertical bars and red-shaded bands correspond to the statistical uncertainty of each sample. The effect of a variation of the electron ($\tau_h$) energy scale of $\pm 1.0\%$ ($\pm 1.2\%$) is shown by the green lines.
Figure 10. Comparison of τ-embedded events with a statistically independent sample of simulated $Z\rightarrow\tau\tau$ events. Shown are distributions of (upper left) $I_{rel}^e$, (upper right) $p_T^{miss}$, (middle left) $I_{rel}^\mu$, (middle right) $m_{jj}$, (lower left) τh-ID BDT, and (lower right) $m_{vis}$, as discussed in the text. The black arrows indicate the working points usually used in the target analyses. The blue vertical bars and red-shaded bands correspond to the statistical uncertainty of each sample. The effect of a variation of the τh energy scale of ±1.2% is shown by the green lines.
trends have been checked to have negligible effects on the target analyses. No further measures are taken to improve the agreement of the embedded event samples with the simulation. Instead, correction factors for the reconstruction and identification of the simulated electrons, muons and tau leptons are derived from $e^-$-, $\mu^-$- and $\tau^-$-embedded events, in analogy to the correction factors usually provided for fully simulated events, as will be discussed in section 7.1.

7 Application of the $\tau^-$-embedding technique to data

The $\tau^-$-embedded event samples used for the target analyses are obtained using the $\mu\mu$ data event selection. They replace the simulation of all $Z \rightarrow \tau\tau$, $t\bar{t}(\tau\tau)$ and diboson($\tau\tau$) events in the $\tau\tau$ final states. To prevent double counting, $t\bar{t}(\tau\tau)$ and diboson($\tau\tau$) events are removed from background estimates that use simulation. Their selection must be performed on the undecayed tau leptons, at the stable particle level.

The $\tau^-$-embedded event sample, except for the $\tau$ decays, provides a data description better than the $Z \rightarrow \tau\tau$ simulation. The simulation can only reach an equivalent performance after a significant amount of tuning. This is true for the time-dependent PU profile of the data, the production of additional jets, especially in exclusive kinematic corners, like for multijet, multi b jet, forward jet, or vector boson fusion topologies and the underlying event. Other event quantities which are typically difficult to model in the simulation are the number of reconstructed primary interaction vertices, or $p_T^{\text{miss}}$. All quantities referring to the part of the event that is obtained from the data may be used in the target analyses without any further corrections. The time needed to produce the $\tau^-$-embedded event sample is of the order of time necessary to reprocess the collected $\mu\mu$ data set. The size of the $\tau^-$-embedded event sample is 5 to 60 times the size of the data sample used for the target analyses. These are advantages over the simulation that will become even more important for the planned High-Luminosity LHC upgrade, where typically between 140 and 200 PU collisions are expected.

The ability of the $\tau^-$-embedded event samples to describe the data is demonstrated below using a data set corresponding to an integrated luminosity of $41.5 \text{ fb}^{-1}$, collected with the CMS detector in 2017.

7.1 Correction factors

Residual differences between the $\tau^-$-embedded event samples and the data in individual control distributions, related to the simulated part of the event, can be adjusted by $p_T^{\tau^\pm}$ and $\eta$-dependent correction factors for the efficiencies of the selection and isolation requirements on each corresponding lepton. These correction factors map the efficiencies observed in the embedded event samples to the efficiencies observed in data. For electrons and muons they are obtained from a comparison of $ee$ ($\mu\mu$) selected events on the $e$ ($\mu$)-embedded event samples with the same event selection on data, using the “tag-and-probe” method [14]. They are provided as individual correction factors for the lepton identification and isolation efficiency, and the corresponding leg of the triggers used in the target analyses. The estimate of the reconstruction efficiency is included in the identification efficiency.

For the identification efficiency of the $\tau_h^-$ candidate, a global correction factor of $0.97 \pm 0.02$ is obtained from a likelihood fit to the yield of $Z \rightarrow \tau\tau$ events in the $\mu\tau_h^-$ final state in a control region. Figure 11 shows typical correction factors for the electron and muon identification and
isolation efficiencies in the central region of the detector, as function of the $p_T$ of the corresponding lepton. Clear turn-on curves are visible for the muon isolation and the electron identification and isolation efficiencies. In each case, a plateau is reached for each efficiency above a $p_T$ threshold of about 30 GeV, which is close to the 80% efficiency working point discussed in section 3 for the electron identification, and close to unity otherwise. In general, the correction factors differ from the efficiencies observed in data by less than 5% in the relevant kinematic regions, and they are smaller for the embedded event samples than for the simulated ones.

Figure 11. (Left column) muon and (right column) electron (upper row) identification and (lower row) isolation efficiencies as a function of the $p_T$ of the corresponding lepton in the central region of the detector. The black arrows indicate typical trigger thresholds of the target analyses. In the upper panel of each subfigure, the black dots correspond to the efficiencies obtained in data, the blue dots to the efficiencies obtained in the corresponding embedded event sample, and the red dots to the efficiencies obtained from the simulation. The lower panels show the ratios of the (blue) embedded event sample and (red) simulation, to the efficiency observed in data, which corresponds to the correction factors.
7.2 Uncertainties

When applied to the target analyses, the following uncertainties, which are specific to the method, should be applied:

- For the normalization of the $\tau$-embedded event samples, a global uncertainty of 2% should be assumed due to the insufficient knowledge of the unfolding corrections of the initially selected muons, as described in section 5.1. The 2% is chosen in accordance to the usual uncertainty in trigger leg efficiencies. This uncertainty should be applied per muon, resulting in an overall uncertainty in the normalization of 4%.

- For the simulated leptons, a variation of 1.2% in the $\tau_h$ energy scale, split by decay mode, as described in ref. [46], should be applied; a variation in the electron energy scale of 1% in the central detector and 2.5% in the endcaps of the ECAL should be applied.

- The uncertainty in the expected fraction of $t\bar{t}(\tau\tau)$ events in the embedded event samples is estimated from a 10% up and down variation of the expected fraction in simulation. The estimate is based on a study in a $t\bar{t}$-enriched control region. It includes the uncertainty in the number of $t\bar{t}$ events that do not contain muons in the final state (as given in table 1) and a general uncertainty in the $t\bar{t}$ event yield in the selected kinematic regime.

- The uncertainties in the correction factors for the trigger leg, identification, and isolation efficiencies are usually of the order of 2% in the kinematic regions relevant for the target analyses, which include the uncertainty in the removal of the energy deposits of the selected muon that primarily affects the isolation efficiency for muons.

- The effects of the finite angular and $p_T$ resolutions of the detector are checked and have negligible influence on the $\tau$-embedded events. They are covered by the variation in the $\tau_h$ and electron energy scale given above. This is also true for a variation of $p_T^{\text{miss}}$ within the observed discrepancies visible in figure 5 (middle left).

These uncertainties are usually a part of a more complex uncertainty model such as described in ref. [46]. For the simulated processes that are replaced by the $\tau$-embedded event sample, they replace uncertainties in the integrated luminosity, production cross sections, jet energy scale, $p_T^{\text{miss}}$ scale and resolution, as well as in the tagging and mistag rates of $b$ jets.

7.3 Comparison to data

To demonstrate how the embedding technique can help in a physics analysis on data, an inclusive event selection is performed for the $\tau\tau$ final states following typical selection requirements, as detailed in ref. [46].

The online selection for the $e\mu$ final state relies on a logical $or$ of two lower-threshold triggers that both require the presence of an electron and a muon in the event with $p_T > 23$ GeV for the higher-$p_T$ lepton and $p_T > 12 \,(8)$ GeV for the lower-$p_T$ electron (muon).

In the offline selection of the $e\mu$ final state, an electron with $p_T > 13$ GeV and $|\eta| < 2.5$ and a muon with $p_T > 9$ GeV and $|\eta| < 2.4$ are required. If the event passed only one trigger, the lepton identified with the higher-$p_T$ trigger object is required to have a $p_T > 24$ GeV, which guarantees
a trigger acceptance well above the turn-on of at least one of the triggers used. Both leptons are required to pass identification criteria and to be isolated according to $I_{\text{rel}}^{e(\mu)} < 0.15 (0.20)$. Events with additional electrons or muons fulfilling looser selection requirements than these are rejected.

The $e\tau_h (\mu\tau_h)$ final state is based on the presence of at least one electron (muon) with $p_T > 35 (27)$ GeV and $|\eta| < 2.1$ at the trigger level. In these final states, an electron (muon) with $p_T > 36 (28)$ GeV and $|\eta| < 2.1$ and a $\tau_h$ candidate with $p_T > 30$ GeV and $|\eta| < 2.3$ are required. The electron (muon) and the $\tau_h$ candidate must fulfill the identification requirements described in section 3. The $\tau_h$ candidate is required to pass the tight working point of the $\tau_h$ identification discriminant, the very loose working point of the discriminant against electrons and the loose (tight) working point of the discriminant to suppress muons and the loose (tight) working point of the discriminant to suppress muons in the $e\tau_h (\mu\tau_h)$ case. In addition, the electron (muon) is required to be isolated, according to $I_{\text{rel}}^{e(\mu)} < 0.10 (0.15)$. Events with additional electrons or muons fulfilling looser selection requirements are rejected.

In the $\tau_h\tau_h$ final state, a trigger decision based on the presence of two hadronically decaying tau leptons with $p_T > 35$ GeV and $|\eta| < 2.1$ is used. Furthermore, two $\tau_h$ candidates with $p_T > 40$ GeV and $|\eta| < 2.1$ are required. Both must pass the tight working point of the $\tau_h$ identification discriminant, the very loose working point of the discriminant against electrons and the loose working point of the discriminant against muons. Events with additional electrons or muons fulfilling looser selection requirements on identification, isolation, and $p_T$ than described for the $e\tau_h$ or $\mu\tau_h$ final state above are rejected.

In all cases, the decay products of the two tau leptons are required to be oppositely charged, separated by more than 0.5 units in $\Delta R$, and associated with the PV within a distance of 0.045 cm in the transverse plane for electrons and muons and 0.2 cm along the beam axis for all final-state particles. The vetoing of additional electrons or muons ensures that no event is used for more than one $\tau\tau$ final state. At most 0.8% of the selected events contain more $\tau_h$ candidates than required for the corresponding final state. In this case, the $\tau\tau$ pair with the most isolated final state products is chosen. In the $e\tau_h$ and $\mu\tau_h$ final state, the events are further selected according to the transverse mass,

$$m_T^{e(\mu)} = \sqrt{2p_T^{e(\mu)}p_T^{\text{miss}}(1 - \cos \Delta \phi)},$$

(7.1)

where $p_T^{e(\mu)}$ refers to the $p_T$ of the electron (muon) and $\Delta \phi$ to the difference in the azimuthal angle between the electron (muon) momentum and $p_T^{\text{miss}}$. In the $e\mu$ final state the events are further selected according to the event variable

$$D_\xi = p_T^{\text{miss}} - 0.85 p_T^{\text{vis}}; \quad p_T^{\text{miss}} = \vec{p}_T^{\text{miss}} \cdot \hat{\xi}; \quad p_T^{\text{vis}} = (\vec{p}_T^e + \vec{p}_T^\mu) \cdot \hat{\xi},$$

(7.2)

where $p_T^{e(\mu)}$ corresponds to the transverse momentum vector of the electron (muon) and $\hat{\xi}$ to the bisectional direction between the electron and the muon momenta in the transverse plane [47]. Events with $m_T^{e(\mu)} < 40$ GeV and $-10 < D_\xi < 30$ GeV are used for further consideration in each corresponding final state. Both $m_T^{e(\mu)}$ and $D_\xi$ quantify the size of $p_T^{\text{miss}}$ and how aligned it is with the momenta of the selected leptons. They are typical event variables to distinguish genuine $\tau\tau$ events from $W + \text{jets}$ and $t\bar{t}$ events.

In figure 12, the distributions of $p_T^{\text{miss}}, D_\xi, m_T^e,$ and $m_T^\mu$ are shown. In addition to the expectation using the $\tau$-embedded event samples, the overall expectation when using the simulation of $Z \rightarrow \tau\tau,$
Figure 12. Distributions of (upper left) $p_{\text{T}}^{\text{miss}}$ in the $\mu\tau_h$ final state, (upper right) $D_\tau$ in the $e\mu$ final state, (lower left) $m_e^T$ in the $e\tau_h$ final state, and (lower right) $m_\mu^T$ in the $\mu\tau_h$ final state. The distributions are shown prior to the maximum likelihood fit described in the text. For these figures, no uncertainties that affect the shape of the distributions have been included in the uncertainty model. The background estimation purely from the CMS simulation is shown as an additional red line.

$\tau\tau$, and diboson($\tau\tau$) events is shown by an open histogram in the upper panel of the subfigures. For this comparison a series of corrections have been applied to the simulation, including a correction to match the pileup distribution in data, a reweighting of the $Z$ boson $p_T$ distribution of the LO simulation to that in $Z \rightarrow \mu\mu$ events observed in data, corrections for the electron and muon legs of the corresponding trigger paths, and for the electron and muon identification and isolation, and corrections of the $Z$ boson recoil, to mitigate differences in detector resolution, between the simulation and data, for the calculation of $p_T^{\text{miss}}$. For $\tau$-embedded events the corrections related to simulated leptons are applied, discussed in section 7.1. A generally good agreement between the expectation and the data is observed, within the applied uncertainty model. A better agreement is found when using the $\tau$-embedded event samples instead of the simulation. Fluctuations in the distributions of $m_e^T$ and $m_\mu^T$ originate mostly from the limited size of the sample of simulated.
Figure 13. Invariant mass distribution of the visible $\tau\tau$ decay products, $m_{\text{vis}}$, in the (upper left) $e\mu$, (upper right) $e\tau$, (lower left) $\mu\tau$, and (lower right) $\tau\tau$ final states, after a fit to the data exploiting a typical uncertainty model as discussed in ref. [46]. In the $e\tau$ final state a significantly larger contribution of $Z \rightarrow \ell\ell$ events is visible compared to the $\mu\tau$ final state. The reason for this is that high-$p_T$ electrons have a higher probability to be misidentified as $\tau$ decays than muons.

$W+\text{jets}$ events. In the target analyses, a large fraction of $W+\text{jets}$ and QCD multijet events are usually estimated from data, which implies that a fraction of up to 90% of the typical background expectation for the target analyses can be estimated from data.

In the target analyses the distributions of a variable related to the invariant mass of the $\tau\tau$ system are usually used as input for a maximum likelihood fit to extract the actual signal. This signal can be an SM process, such as the SM $Z$ or Higgs boson production in the $\tau\tau$ final state, or any other process of physics beyond the SM. In figure 13 the distributions of $m_{\text{vis}}$, as such a variable, in the $e\mu$, $e\tau$, $\mu\tau$, and $\tau\tau$ final states are shown, after the event selections, as described above, and after applying a maximum likelihood fit to the observation with the $\tau$-embedded event sample as signal. For this purpose, a likelihood model has been adapted from ref. [46]. It incorporates $O(100)$ uncertainties in form of nuisance parameters that may be correlated across the processes contributing to the input distributions, and across final states. Within a single input distribution
the nuisance parameters may allow for correlated shifts across bins, such as process normalization or energy scale uncertainties, and for shifts of individual bins, within the statistical precision of the template distributions used in the model. The ability of the model to describe the data can be quantified using a goodness-of-fit test, based on a saturated likelihood model (SAT) described in ref. [48], which corresponds to a generalization of a $\chi^2$ test including all systematic uncertainties of the model and their correlations. The SAT test indicates the overall statistical compatibility of the model with the observation, treating each bin of the input distributions as an independent measurement. Goodness-of-fit tests based on the empirical distribution function are usually more sensitive than a $\chi^2$-like test to small deviations that are correlated across several bins of a single histogram. A classical test of this kind that is mostly sensitive to deviations correlated across bins in the center of a given binned distribution is the Kolmogorov-Smirnov (KS) test [49, 50]. A variant of this test that gives more emphasis to the edges of the given input distribution is the Anderson-Darling (AD) test [51]. The $p$-values for each of these tests, split by final state, are shown in table 3. They have a one-to-one correspondence to the distributions shown in figure 13, with the small difference that these distributions are shown for a fit to the observation in all final states combined. The $p$-values are obtained from the comparison of the observed value for the corresponding test statistic with the outcome of pseudo-experiments based on the expectation. Their statistical precision is better than 0.5%. The actual values range from 17%, for the $p$-value of the AD test in the $e\tau_h$ final state, to 82%, for the $p$-value of the AD test in the $\mu\tau_h$ final state. All tests reveal good compatibility of the statistical model with the observation, which implies a successful description of the data with the given template distributions, especially with the $\tau$-embedded event samples. The fit to the observation in all final states combined reveals a $p$-value of 51% and a normalization of $1.00\pm0.05$ for the $\tau$-embedded event samples, which is in good agreement with the observations of ref. [46] that have been made on an independent data set. Also a good compatibility of the normalization across all final states is observed. The normalization of the $\tau$-embedded samples is obtained from the data. Figures showing distributions of more quantities relevant for the analysis of $\tau\tau$ events are given in appendix A.

### Table 3.
Normalization of the $\tau$-embedded event samples and $p$-values of the saturated model (SAT), Kolmogorov-Smirnov (KS) and Anderson-Darling (AD) test, as discussed in the text, separated by $\tau\tau$ final state, as introduced in section 5 and (where applicable) for all channels combined. The $p$-values have a statistical precision better than 0.5%.

<table>
<thead>
<tr>
<th>Final state</th>
<th>Normalization</th>
<th>SAT</th>
<th>KS</th>
<th>AD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$e\mu$</td>
<td>$1.02\pm0.05$</td>
<td>0.61</td>
<td>0.29</td>
<td>0.74</td>
</tr>
<tr>
<td>$e\tau_h$</td>
<td>$0.87\pm0.08$</td>
<td>0.69</td>
<td>0.35</td>
<td>0.17</td>
</tr>
<tr>
<td>$\mu\tau_h$</td>
<td>$0.96\pm0.07$</td>
<td>0.76</td>
<td>0.81</td>
<td>0.82</td>
</tr>
<tr>
<td>$\tau_h\tau_h$</td>
<td>$1.10\pm0.12$</td>
<td>0.71</td>
<td>0.54</td>
<td>0.30</td>
</tr>
<tr>
<td>Combined</td>
<td>$1.00\pm0.05$</td>
<td>0.51</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>
8 Summary

The $\tau$-embedding technique developed for the CMS experiment is described and its validation and relevant uncertainties are discussed. The 13 TeV proton-proton collisions collected by CMS in 2017 are used to demonstrate the performance of the technique with the data sample corresponding to an integrated luminosity of $41.5 \text{ fb}^{-1}$.

The main goal of the procedure is to estimate the background from $Z \rightarrow \tau\tau$ events using recorded $Z \rightarrow \mu\mu$ events. The estimate also includes events from $t\bar{t}$ and diboson production with two tau leptons in the final state. Recorded $\mu\mu$ events are selected, the muons are removed from the reconstructed event record, and replaced with simulated tau leptons with the same kinematic properties as the removed muons. In that way hybrid events are obtained, which rely on the simulation only for the decay of the tau leptons. Challenges in describing the underlying event or the production of associated jets in the simulation, as well as the costly simulation of PU events thus are avoided. The embedding technique decreases the uncertainties inherent in a typical simulation process, such as the uncertainties in the missing transverse momentum, jet energy scale and resolution, $b$ tagging efficiency, and misidentification probability.

A number of validation tests for $\mu$-, $e$-, and $\tau$-embedding, as well as several goodness-of-fit tests, show good agreement of embedded distributions with those obtained using simulated and recorded data events. The embedding technique avoids time-consuming simulations of events that becomes critical for the planned High-Luminosity LHC upgrade, where typical pileup of 140–200 collisions per bunch crossing is expected.

Acknowledgments

We congratulate our colleagues in the CERN accelerator departments for the excellent performance of the LHC and thank the technical and administrative staffs at CERN and at other CMS institutes for their contributions to the success of the CMS effort. In addition, we gratefully acknowledge the computing centers and personnel of the Worldwide LHC Computing Grid for delivering so effectively the computing infrastructure essential to our analyses. Finally, we acknowledge the enduring support for the construction and operation of the LHC and the CMS detector provided by the following funding agencies: the Austrian Federal Ministry of Education, Science and Research and the Austrian Science Fund; the Belgian Fonds de la Recherche Scientifique, and Fonds voor Wetenschappelijk Onderzoek; the Brazilian Funding Agencies (CNPq, CAPES, FAPERJ, FAPERGS, and FAPESP); the Bulgarian Ministry of Education and Science; CERN; the Chinese Academy of Sciences, Ministry of Science and Technology, and National Natural Science Foundation of China; the Colombian Funding Agency (COLCIENCIAS); the Croatian Ministry of Science, Education and Sport, and the Croatian Science Foundation; the Research Promotion Foundation, Cyprus; the Secretariat for Higher Education, Science, Technology and Innovation, Ecuador; the Ministry of Education and Research, Estonian Research Council via IUT23-4 and IUT23-6 and European Regional Development Fund, Estonia; the Academy of Finland, Finnish Ministry of Education and Culture, and Helsinki Institute of Physics; the Institut National de Physique Nucléaire et de Physique des Particules / CNRS, and Commissariat à l’Énergie Atomique et aux Énergies Alternatives / CEA, France; the Bundesministerium für Bildung und Forschung, Deutsche Forschungsgemeinschaft,
and Helmholtz-Gemeinschaft Deutscher Forschungszentren, Germany; the General Secretariat for Research and Technology, Greece; the National Research, Development and Innovation Fund, Hungary; the Department of Atomic Energy and the Department of Science and Technology, India; the Institute for Studies in Theoretical Physics and Mathematics, Iran; the Science Foundation, Ireland; the Istituto Nazionale di Fisica Nucleare, Italy; the Ministry of Science, ICT and Future Planning, and National Research Foundation (NRF), Republic of Korea; the Ministry of Education and Science of the Republic of Latvia; the Lithuanian Academy of Sciences; the Ministry of Education, and University of Malaya (Malaysia); the Ministry of Science of Montenegro; the Mexican Funding Agencies (BUAP, CINVESTAV, CONACYT, LNS, SEP, and UASLP-FAI); the Ministry of Business, Innovation and Employment, New Zealand; the Pakistan Atomic Energy Commission; the Ministry of Science and Higher Education and the National Science Center, Poland; the Fundação para a Ciência e a Tecnologia, Portugal; JINR, Dubna; the Ministry of Education and Science of the Russian Federation, the Federal Agency of Atomic Energy of the Russian Federation, Russian Academy of Sciences, the Russian Foundation for Basic Research, and the National Research Center “Kurchatov Institute”; the Ministry of Education, Science and Technological Development of Serbia; the Secretaría de Estado de Investigación, Desarrollo e Innovación, Programa Consolider-Ingenio 2010, Plan Estatal de Investigación Científica y Técnica y de Innovación 2013–2016, Plan de Ciencia, Tecnología e Innovación 2013–2017 del Principado de Asturias, and Fondo Europeo de Desarrollo Regional, Spain; the Ministry of Science, Technology and Research, Sri Lanka; the Swiss Funding Agencies (ETH Board, ETH Zurich, PSI, SNF, UniZH, Canton Zurich, and SER); the Ministry of Science and Technology, Taipei; the Thailand Center of Excellence in Physics, the Institute for the Promotion of Teaching Science and Technology of Thailand, Special Task Force for Activating Research and the National Science and Technology Development Agency of Thailand; the Scientific and Technical Research Council of Turkey, and Turkish Atomic Energy Authority; the National Academy of Sciences of Ukraine, and State Fund for Fundamental Researches, Ukraine; the Science and Technology Facilities Council, U.K.; the US Department of Energy, and the US National Science Foundation.

Individuals have received support from the Marie-Curie program and the European Research Council and Horizon 2020 Grant, contract No. 675440 (European Union); the Leventis Foundation; the A. P. Sloan Foundation; the Alexander von Humboldt Foundation; the Belgian Federal Science Policy Office; the Fonds pour la Formation à la Recherche dans l’Industrie et dans l’Agriculture (FRIA-Belgium); the Agentschap voor Innovatie door Wetenschap en Technologie (IWT-Belgium); the F.R.S.-FNRS and FWO (Belgium) under the “Excellence of Science — EOS” — be.h project n. 30820817; the Ministry of Education, Youth and Sports (MEYS) of the Czech Republic; the Lendület (“Momentum”) Program and the János Bolyai Research Scholarship of the Hungarian Academy of Sciences, the New National Excellence Program ÚNKP, the NKFIÁ research grants 123842, 123959, 124845, 124850 and 125105 (Hungary); the Council of Scientific and Industrial Research, India; the HOMING PLUS program of the Foundation for Polish Science, cofinanced from European Union, Regional Development Fund, the Mobility Plus program of the Ministry of Science and Higher Education, the National Science Center (Poland), contracts Harmonia 2014/14/M/ST2/00428, Opus 2014/13/B/ST2/02543, 2014/15/B/ST2/03998, and 2015/19/B/ST2/02861, Sonata-bis 2012/07/E/ST2/01406; the National Priorities Research Program by Qatar National Research Fund; the Programa de Excelencia María de Maeztu, and the Programa
A Performance of the $\tau$-embedding method on data

Distributions of more quantities relevant for the analysis of $\tau\tau$ events. The distributions are shown prior to the maximum likelihood fit discussed in section 7.3. In addition to the expectation using the $\tau$-embedded event samples, the overall expectation when using the simulation of $Z \rightarrow \tau\tau, t\bar{t}(\tau\tau)$, and diboson($\tau\tau$) events is shown by a red line in the upper panel of the subfigures. For this comparison a series of corrections have been applied as discussed in section 7.3. For $\tau$-embedded events the corrections related to the simulated leptons, discussed in section 7.1, have been applied. For these figures, no uncertainties that affect the shape of the distributions have been included in the uncertainty model.

Figure 14. Distributions of (left) $m_{jj}$ and (right) the number of reconstructed primary vertices $N_{\text{vtx}}$ in the $\mu\tau_h$ final state.
Figure 15. Distributions of the (left) jet and (right) $b$ jet multiplicity, as described in the text, in the $\mu\tau_h$ final state.

Figure 16. Distributions of the $p_T$ of the (left) leading and (right) trailing jet for events with more than one jet in the $\mu\tau_h$ final state.
References


[16] ATLAS collaboration, Modelling $Z \to \tau\tau$ processes in ATLAS with $\tau$-embedded $Z \to \mu\mu$ data, 2015 *JINST* 10 P09018 [arXiv:1506.05623].


[34] CMS collaboration, *Reconstruction and identification of $\tau$ leptons decaying to hadrons and $\nu_\tau$ at CMS*, 2016 *JINST* **11** P01019 [arXiv:1510.07488].


CDF collaboration, Search for neutral MSSM Higgs bosons decaying to tau pairs in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV, *Phys. Rev. Lett.* **96** (2006) 011802 [hep-ex/0508051].


A.N. Kolmogorov, Sulla determinazione empirica di una legge di distribuzione, Giornale dell’Istituto Italiano degli Attuari **4** (1933) 83.


The CMS collaboration

Yerevan Physics Institute, Yerevan, Armenia
A.M. Sirunyan, A. Tumasyan

Institut für Hochenergiephysik, Wien, Austria

Institute for Nuclear Problems, Minsk, Belarus
V. Chekhovsky, V. Mossolov, J. Suarez Gonzalez

Universiteit Antwerpen, Antwerpen, Belgium

Vrije Universiteit Brussel, Brussel, Belgium

Université Libre de Bruxelles, Bruxelles, Belgium

Ghent University, Ghent, Belgium

Université Catholique de Louvain, Louvain-la-Neuve, Belgium

Centro Brasileiro de Pesquisas Fisicas, Rio de Janeiro, Brazil

Universidade do Estado do Rio de Janeiro, Rio de Janeiro, Brazil
Universidade Estadual Paulista $^a$, Universidade Federal do ABC $^b$, São Paulo, Brazil
S. Ahuja$^a$, C.A. Bernardes$^a$, L. Calligaris$^a$, T.R. Fernandez Perez Tomei$^a$, E.M. Gregores$^b$, P.G. Mercadante$^b$, S.F. Novaes$^a$, SandraS. Padula$^a$

Institute for Nuclear Research and Nuclear Energy, Bulgarian Academy of Sciences, Sofia, Bulgaria
A. Aleksandrov, R. Hadjiiska, P. Iaydjiev, A. Marinov, M. Misheva, M. Rodozov, M. Shopova, G. Sultanov

University of Sofia, Sofia, Bulgaria
A. Dimitrov, L. Litov, B. Pavlov, P. Petkov

Beihang University, Beijing, China
W. Fang$^6$, X. Gao$^6$, L. Yuan

Institute of High Energy Physics, Beijing, China

State Key Laboratory of Nuclear Physics and Technology, Peking University, Beijing, China
Y. Ban, G. Chen, A. Levin, J. Li, L. Li, Q. Li, Y. Mao, S.J. Qian, D. Wang

Tsinghua University, Beijing, China
Y. Wang

Universidad de Los Andes, Bogota, Colombia
C. Avila, A. Cabrera, C.A. Carrillo Montoya, L.F. Chaparro Sierra, C. Florez, C.F. González Hernández, M.A. Segura Delgado

University of Split, Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture, Split, Croatia
N. Godinovic, D. Lelas, I. Puljak, T. Sculac

University of Split, Faculty of Science, Split, Croatia
Z. Antunovic, M. Kovac

Institute Rudjer Boskovic, Zagreb, Croatia
V. Brigljevic, D. Ferencek, K. Kadija, B. Mesic, M. Roguljic, A. Starodumov$^8$, T. Susa

University of Cyprus, Nicosia, Cyprus

Charles University, Prague, Czech Republic
M. Finger$^9$, M. Finger Jr.$^9$

Escuela Politecnica Nacional, Quito, Ecuador
E. Ayala

Universidad San Francisco de Quito, Quito, Ecuador
E. Carrera Jarrin
Academy of Scientific Research and Technology of the Arab Republic of Egypt, Egyptian Network of High Energy Physics, Cairo, Egypt
A. Ellithi Kamel\textsuperscript{10}, M.A. Mahmoud\textsuperscript{11,12}, E. Salama\textsuperscript{12,13}

National Institute of Chemical Physics and Biophysics, Tallinn, Estonia
S. Bhowmik, A. Carvalho Antunes De Oliveira, R.K. Dewanjee, K. Ehataht, M. Kadastik, M. Raidal, C. Veelken

Department of Physics, University of Helsinki, Helsinki, Finland
P. Eerola, H. Kirschenmann, J. Pekkanen, M. Voutilainen

Helsinki Institute of Physics, Helsinki, Finland
J. Havukainen, J.K. Heikkilä, T. Järvinen, V. Karimäki, R. Kinnunen, T. Lampén, K. Lassila-Perini, S. Laurila, S. Lehti, T. Lindén, P. Luukka, T. Mäenpää, H. Siikonen, E. Tuominiemi

Lappeenranta University of Technology, Lappeenranta, Finland
T. Tuuva

IRFU, CEA, Université Paris-Saclay, Gif-sur-Yvette, France

Laboratoire Leprince-Ringuet, Ecole polytechnique, CNRS/IN2P3, Université Paris-Saclay, Palaiseau, France

Université de Strasbourg, CNRS, IPHC UMR 7178, Strasbourg, France

Centre de Calcul de l’Institut National de Physique Nucleaire et de Physique des Particules, CNRS/IN2P3, Villeurbanne, France
S. Gadrat

Université de Lyon, Université Claude Bernard Lyon 1, CNRS-IN2P3, Institut de Physique Nucléaire de Lyon, Villeurbanne, France

Georgian Technical University, Tbilisi, Georgia
A. Khvedelidze\textsuperscript{9}
Tbilisi State University, Tbilisi, Georgia
Z. Tsamalaidze

RWTH Aachen University, I. Physikalisches Institut, Aachen, Germany

RWTH Aachen University, III. Physikalisches Institut A, Aachen, Germany

RWTH Aachen University, III. Physikalisches Institut B, Aachen, Germany

Deutsches Elektronen-Synchrotron, Hamburg, Germany

University of Hamburg, Hamburg, Germany

Karlsruher Institut fuer Technologie, Karlsruhe, Germany

Institute of Nuclear and Particle Physics (INPP), NCSR Demokritos, Aghia Paraskevi, Greece
National and Kapodistrian University of Athens, Athens, Greece
A. Agapitos, G. Karathanasis, P. Kontaxakis, A. Panagiotou, I. Papavergou, N. Saoulidou, K. Vellidis

National Technical University of Athens, Athens, Greece
G. Bakas, K. Kousouris, I. Papakrivopoulos, G. Tspirolis

University of Ioánnina, Ioánnina, Greece

MTA-ELTE Lendület CMS Particle and Nuclear Physics Group, Eötvös Loránd University, Budapest, Hungary

Wigner Research Centre for Physics, Budapest, Hungary

Institute of Nuclear Research ATOMKI, Debrecen, Hungary
N. Beni, S. Czellar, J. Karancsi, A. Makovec, J. Molnar, Z. Szillasi

Institute of Physics, University of Debrecen, Debrecen, Hungary
P. Raics, Z.L. Trocsanyi, B. Ujvari

Indian Institute of Science (IISc), Bangalore, India
S. Choudhury, J.R. Komaragiri, P.C. Tiwari

National Institute of Science Education and Research, HBNI, Bhubaneswar, India

Panjab University, Chandigarh, India

University of Delhi, Delhi, India
A. Bhardwaj, B.C. Choudhary, R.B. Garg, M. Gola, S. Keshri, Ashok Kumar, S. Malhotra, M. Naimuddin, P. Priyanka, K. Ranjan, Aasheq Shah, R. Sharma

Saha Institute of Nuclear Physics, HBNI, Kolkata, India

Indian Institute of Technology Madras, Madras, India
P.K. Behera, A. Muhammad

Bhabha Atomic Research Centre, Mumbai, India
R. Chudasama, D. Dutta, V. Jha, V. Kumar, D.K. Mishra, P.K. Netrakanti, L.M. Pant, P. Shukla, P. Suggisetti
Tata Institute of Fundamental Research-A, Mumbai, India

Tata Institute of Fundamental Research-B, Mumbai, India

Indian Institute of Science Education and Research (IISER), Pune, India
S. Chauhan, S. Dube, V. Hegde, A. Kapoor, K. Kothekar, S. Pandey, A. Rane, A. Rastogi, S. Sharma

Institute for Research in Fundamental Sciences (IPM), Tehran, Iran
S. Chenarani, E. Eskandari Tadavani, S.M. Etesami, M. Khakzad, M. Mohammadi Najafabadi, M. Naseri, F. Rezaei Hosseinabadi, B. Safarzadeh, M. Zeinali

University College Dublin, Dublin, Ireland
M. Felcini, M. Grunewald

INFN Sezione di Bari, Università di Bari, Politecnico di Bari, Bari, Italy

INFN Sezione di Bologna, Università di Bologna, Bologna, Italy

INFN Sezione di Catania, Università di Catania, Catania, Italy
S. Albergo, A. Di Mattia, R. Potenza, A. Tricomi, C. Tuve

INFN Sezione di Firenze, Università di Firenze, Firenze, Italy

INFN Laboratori Nazionali di Frascati, Frascati, Italy
L. Benussi, S. Bianco, F. Fabbri, D. Piccolo

INFN Sezione di Genova, Università di Genova, Genova, Italy
F. Ferro, R. Mulargia, E. Robutti, S. Tosi

INFN Sezione di Milano-Bicocca, Università di Milano-Bicocca, Milano, Italy
A. Benaglia, A. Beschi, F. Brivio, V. Ciriolo, S. Di Guida, M.E. Dinaro, S. Fiorendi, S. Gennai, A. Ghezzi, P. Govoni, M. Malberti, S. Malvezzi, D. Menasce

– 42 –
INFN Sezione di Trieste $^a$, Università di Trieste $^b$, Trieste, Italy
S. Belforte$^a$, V. Candelise$^{a,b}$, M. Casarsa$^a$, F. Cossutti$^a$, A. Da Rold$^{a,b}$, G. Della Ricca$^{a,b}$, F. Vazzoler$^{a,b}$, A. Zanetti$^a$

Kyungpook National University, Daegu, Korea

Chonnam National University, Institute for Universe and Elementary Particles, Kwangju, Korea
H. Kim, D.H. Moon, G. Oh

Hanyang University, Seoul, Korea
B. Francois, J. Goh$^{33}$, T.J. Kim

Korea University, Seoul, Korea

Sejong University, Seoul, Korea
H.S. Kim

Seoul National University, Seoul, Korea

University of Seoul, Seoul, Korea

Sungkyunkwan University, Suwon, Korea
Y. Choi, C. Hwang, J. Lee, I. Yu

Riga Technical University, Riga, Latvia
V. Veckalns$^{34}$

Vilnius University, Vilnius, Lithuania
V. Dudenas, A. Juodagalvis, J. Vaitkus

National Centre for Particle Physics, Universiti Malaya, Kuala Lumpur, Malaysia

Universidad de Sonora (UNISON), Hermosillo, Mexico
J.F. Benitez, A. Castaneda Hernandez, J.A. Murillo Quijada

Centro de Investigacion y de Estudios Avanzados del IPN, Mexico City, Mexico

Universidad Iberoamericana, Mexico City, Mexico
S. Carrillo Moreno, C. Oropeza Barrera, M. Ramirez-Garcia, F. Vazquez Valencia
Benemerita Universidad Autonoma de Puebla, Puebla, Mexico
J. Eysermans, I. Pedraza, H.A. Salazar Ibarguen, C. Uribe Estrada

Universidad Autónoma de San Luis Potosí, San Luis Potosí, Mexico
A. Morelos Pineda

University of Auckland, Auckland, New Zealand
D. Krofcheck

University of Canterbury, Christchurch, New Zealand
S. Bheesette, P.H. Butler

National Centre for Physics, Quaid-I-Azam University, Islamabad, Pakistan
A. Ahmad, M. Ahmad, M.I. Asghar, Q. Hassan, H.R. Hoorani, W.A. Khan, M.A. Shah, M. Shoaib, M. Waqas

National Centre for Nuclear Research, Swierk, Poland

Institute of Experimental Physics, Faculty of Physics, University of Warsaw, Warsaw, Poland

Laboratório de Instrumentação e Física Experimental de Partículas, Lisboa, Portugal

Joint Institute for Nuclear Research, Dubna, Russia

Petersburg Nuclear Physics Institute, Gatchina (St. Petersburg), Russia
V. Golovtsov, Y. Ivanov, V. Kim, E. Kuznetsova, P. Levchenko, V. Murzin, V. Oreshkin, I. Smirnov, D. Sosnov, V. Sulimov, L. Uvarov, S. Vavilov, A. Vorobyev

Institute for Nuclear Research, Moscow, Russia

Institute for Theoretical and Experimental Physics, Moscow, Russia
V. Epshteyn, V. Gavrilov, N. Lychkovskaya, V. Popov, I. Pozdnyakov, G. Safronov, A. Spiridonov, A. Stepennov, V. Stolin, M. Toms, E. Vlasov, A. Zhokin

Moscow Institute of Physics and Technology, Moscow, Russia
T. Aushev
National Research Nuclear University ‘Moscow Engineering Physics Institute’ (MEPhI), Moscow, Russia
M. Chadeeva\textsuperscript{43}, P. Parygin, E. Popova, V. Rusinov

P.N. Lebedev Physical Institute, Moscow, Russia
V. Andreev, M. Azarkin, I. Dremin\textsuperscript{40}, M. Kirakosyan, A. Terkulov

Skobeltsyn Institute of Nuclear Physics, Lomonosov Moscow State University, Moscow, Russia
A. Belyaev, E. Boos, V. Bunichev, M. Dubinin\textsuperscript{44}, L. Dudko, A. Ershov, A. Gribushin, V. Klyukhin, O. Kodolova, I. Lokhtin, S. Obraztsov, M.Perfilov, V. Savrin

Novosibirsk State University (NSU), Novosibirsk, Russia
A. Barnyakov\textsuperscript{45}, V. Blinov\textsuperscript{45}, T. Dimova\textsuperscript{45}, M. Kirakosyan, A. Terkulov

Skobeltsyn Institute of Nuclear Physics, Lomonosov Moscow State University, Moscow, Russia
A. Belyaev, E. Boos, V. Bunichev, M. Dubinin\textsuperscript{44}, L. Dudko, A. Ershov, A. Gribushin, V. Klyukhin, O. Kodolova, I. Lokhtin, S. Obraztsov, M.Perfilov, V. Savrin

Skobeltsyn Institute of Nuclear Physics, Lomonosov Moscow State University, Moscow, Russia
A. Belyaev, E. Boos, V. Bunichev, M. Dubinin\textsuperscript{44}, L. Dudko, A. Ershov, A. Gribushin, V. Klyukhin, O. Kodolova, I. Lokhtin, S. Obraztsov, M.Perfilov, V. Savrin

Novosibirsk State University (NSU), Novosibirsk, Russia
A. Barnyakov\textsuperscript{45}, V. Blinov\textsuperscript{45}, T. Dimova\textsuperscript{45}, M. Kirakosyan, A. Terkulov

Institute for High Energy Physics of National Research Centre ‘Kurchatov Institute’, Protvino, Russia

National Research Tomsk Polytechnic University, Tomsk, Russia
A. Babaev, S. Baidali, V. Okhotnikov

University of Belgrade: Faculty of Physics and VINCA Institute of Nuclear Sciences
P. Adzic\textsuperscript{46}, P. Cirkovic, D. Devetak, M. Dordevic, P. Milenovic\textsuperscript{47}, J. Milosevic

Centro de Investigaciones Energéticas Medioambientales y Tecnológicas (CIEMAT), Madrid, Spain

Universidad Autónoma de Madrid, Madrid, Spain
C. Albajar, J.F. de Trocóniz

Universidad de Oviedo, Oviedo, Spain

Instituto de Física de Cantabria (IFCA), CSIC-Universidad de Cantabria, Santander, Spain

University of Ruhuna, Department of Physics, Matara, Sri Lanka
N. Wickramage
CERN, European Organization for Nuclear Research, Geneva, Switzerland

Paul Scherrer Institut, Villigen, Switzerland

ETH Zurich — Institute for Particle Physics and Astrophysics (IPA), Zurich, Switzerland

Universität Zürich, Zurich, Switzerland

National Central University, Chung-Li, Taiwan
T.H. Doan, C.M. Kuo, W. Lin, S.S. Yu

National Taiwan University (NTU), Taipei, Taiwan

Chulalongkorn University, Faculty of Science, Department of Physics, Bangkok, Thailand
B. Asavapibhop, N. Srimanobhas, N. Suwonjandee

Çukurova University, Physics Department, Science and Art Faculty, Adana, Turkey
Middle East Technical University, Physics Department, Ankara, Turkey
B. Isildak, G. Karapinar, M. Yalvac, M. Zeyrek

Bogazici University, Istanbul, Turkey
I.O. Atakisi, E. Gülmez, M. Kaya, O. Kaya, Özçelik, S. Ozkorucuklu, S. Tekten, E.A. Yetkin

Istanbul Technical University, Istanbul, Turkey
M.N. Agaras, A. Cakir, K. Cankocak, Y. Komurcu, S. Sen

Institute for Scintillation Materials of National Academy of Science of Ukraine, Kharkov, Ukraine
B. Grynyov

National Scientific Center, Kharkov Institute of Physics and Technology, Kharkov, Ukraine
L. Levchuk

University of Bristol, Bristol, United Kingdom

Rutherford Appleton Laboratory, Didcot, United Kingdom

Imperial College, London, United Kingdom

Brunel University, Uxbridge, United Kingdom
J.E. Cole, P.R. Hobson, A. Khan, P. Kyberd, C.K. Mackay, A. Morton, I.D. Reid, L. Teodorescu, S. Zahid

Baylor University, Waco, U.S.A.
K. Call, J. Dittmann, K. Hatakeyama, H. Liu, C. Madrid, B. McMaster, N. Pastika, C. Smith

Catholic University of America, Washington, DC, U.S.A.
R. Bartek, A. Dominguez

The University of Alabama, Tuscaloosa, U.S.A.
A. Buccilli, O. Charaf, S.I. Cooper, C. Henderson, P. Rumerio, C. West
Boston University, Boston, U.S.A.

Brown University, Providence, U.S.A.

University of California, Davis, Davis, U.S.A.

University of California, Los Angeles, U.S.A.

University of California, Riverside, Riverside, U.S.A.

University of California, San Diego, La Jolla, U.S.A.

University of California, Santa Barbara — Department of Physics, Santa Barbara, U.S.A.

California Institute of Technology, Pasadena, U.S.A.

Carnegie Mellon University, Pittsburgh, U.S.A.

University of Colorado Boulder, Boulder, U.S.A.

Cornell University, Ithaca, U.S.A.
Fermi National Accelerator Laboratory, Batavia, U.S.A.


University of Florida, Gainesville, U.S.A.


Florida International University, Miami, U.S.A.

Y.R. Joshi, S. Linn

Florida State University, Tallahassee, U.S.A.


Florida Institute of Technology, Melbourne, U.S.A.


University of Illinois at Chicago (UIC), Chicago, U.S.A.


The University of Iowa, Iowa City, U.S.A.


Johns Hopkins University, Baltimore, U.S.A.


The University of Kansas, Lawrence, U.S.A.

Kansas State University, Manhattan, U.S.A.

Lawrence Livermore National Laboratory, Livermore, U.S.A.
F. Rebassoo, D. Wright

University of Maryland, College Park, U.S.A.

Massachusetts Institute of Technology, Cambridge, U.S.A.

University of Minnesota, Minneapolis, U.S.A.

University of Mississippi, Oxford, U.S.A.
J.G. Acosta, S. Oliveros

University of Nebraska-Lincoln, Lincoln, U.S.A.
E. Avdeeva, K. Bloom, D.R. Claes, C. Fangmeier, L. Finco, F. Golf, R. Gonzalez Suarez, R. Kamalieddin, I. Kravchenko, J.E. Siado, G.R. Snow, B. Stieger

State University of New York at Buffalo, Buffalo, U.S.A.

Northeastern University, Boston, U.S.A.

Northwestern University, Evanston, U.S.A.

University of Notre Dame, Notre Dame, U.S.A.
The Ohio State University, Columbus, U.S.A.

Princeton University, Princeton, U.S.A.

University of Puerto Rico, Mayaguez, U.S.A.
S. Malik, S. Norberg

Purdue University, West Lafayette, U.S.A.

Purdue University Northwest, Hammond, U.S.A.
T. Cheng, J. Dolen, N. Parashar

Rice University, Houston, U.S.A.

University of Rochester, Rochester, U.S.A.

Rutgers, The State University of New Jersey, Piscataway, U.S.A.

University of Tennessee, Knoxville, U.S.A.
H. Acharya, A.G. Delannoy, J. Heideman, G. Riley, S. Spanier

Texas A&M University, College Station, U.S.A.

Texas Tech University, Lubbock, U.S.A.

Vanderbilt University, Nashville, U.S.A.
University of Virginia, Charlottesville, U.S.A.
M.W. Arenton, P. Barria, B. Cox, R. Hirosky, M. Joyce, A. Ledovskoy, H. Li, C. Neu, Y. Wang, E. Wolfe, F. Xia

Wayne State University, Detroit, U.S.A.

University of Wisconsin — Madison, Madison, WI, U.S.A.

†: Deceased
1: Also at Vienna University of Technology, Vienna, Austria
2: Also at Skobeltsyn Institute of Nuclear Physics, Lomonosov Moscow State University, Moscow, Russia
3: Also at IRFU, CEA, Université Paris-Saclay, Gif-sur-Yvette, France
4: Also at Universidade Estadual de Campinas, Campinas, Brazil
5: Also at Federal University of Rio Grande do Sul, Porto Alegre, Brazil
6: Also at Université Libre de Bruxelles, Bruxelles, Belgium
7: Also at University of Chinese Academy of Sciences, Beijing, China
8: Also at Institute for Theoretical and Experimental Physics, Moscow, Russia
9: Also at Joint Institute for Nuclear Research, Dubna, Russia
10: Now at Cairo University, Cairo, Egypt
11: Also at Fayoum University, El-Fayoum, Egypt
12: Now at British University in Egypt, Cairo, Egypt
13: Now at Ain Shams University, Cairo, Egypt
14: Also at Department of Physics, King Abdulaziz University, Jeddah, Saudi Arabia
15: Also at Université de Haute Alsace, Mulhouse, France
16: Also at CERN, European Organization for Nuclear Research, Geneva, Switzerland
17: Also at RWTH Aachen University, III. Physikalisches Institut A, Aachen, Germany
18: Also at University of Hamburg, Hamburg, Germany
19: Also at Brandenburg University of Technology, Cottbus, Germany
20: Also at Institute of Physics, University of Debrecen, Debrecen, Hungary
21: Also at Institute of Nuclear Research ATOMKI, Debrecen, Hungary
22: Also at MTA-ELTE Lendület CMS Particle and Nuclear Physics Group, Eötvös Loránd University, Budapest, Hungary
23: Also at Indian Institute of Technology Bhubaneswar, Bhubaneswar, India
24: Also at Institute of Physics, Bhubaneswar, India
25: Also at Shoolini University, Solan, India
26: Also at University of Visva-Bharati, Santiniketan, India
27: Also at Isfahan University of Technology, Isfahan, Iran
28: Also at Plasma Physics Research Center, Science and Research Branch, Islamic Azad University, Tehran, Iran
29: Also at Italian National Agency For New Technologies, Energy And Sustainable Economic Development, Bologna, Italy
30: Also at Centro Siciliano Di Fisica Nucleare E Di Struttura Della Materia, Catania, Italy
31: Also at Università degli Studi di Siena, Siena, Italy
32: Also at Scuola Normale e Sezione dell’INFN, Pisa, Italy
33: Also at Kyung Hee University, Department of Physics, Seoul, Korea
34: Also at Riga Technical University, Riga, Latvia
35: Also at International Islamic University of Malaysia, Kuala Lumpur, Malaysia
36: Also at Malaysian Nuclear Agency, MOSTI, Kajang, Malaysia
37: Also at Consejo Nacional de Ciencia y Tecnología, Mexico City, Mexico
38: Also at Warsaw University of Technology, Institute of Electronic Systems, Warsaw, Poland
39: Also at Institute for Nuclear Research, Moscow, Russia
40: Now at National Research Nuclear University ‘Moscow Engineering Physics Institute’ (MEPhI), Moscow, Russia
41: Also at St. Petersburg State Polytechnical University, St. Petersburg, Russia
42: Also at University of Florida, Gainesville, U.S.A.
43: Also at P.N. Lebedev Physical Institute, Moscow, Russia
44: Also at California Institute of Technology, Pasadena, U.S.A.
45: Also at Budker Institute of Nuclear Physics, Novosibirsk, Russia
46: Also at Faculty of Physics, University of Belgrade, Belgrade, Serbia
47: Also at University of Belgrade, Belgrade, Serbia
48: Also at INFN Sezione di Pavia a, Università di Pavia b, Pavia, Italy
49: Also at National and Kapodistrian University of Athens, Athens, Greece
50: Also at Universität Zürich, Zurich, Switzerland
51: Also at Stefan Meyer Institute for Subatomic Physics (SMI), Vienna, Austria
52: Also at Gaziosmanpasa University, Tokat, Turkey
53: Also at Beykent University, Istanbul, Turkey
54: Also at Istanbul Aydin University, Istanbul, Turkey
55: Also at Mersin University, Mersin, Turkey
56: Also at Piri Reis University, Istanbul, Turkey
57: Also at Adiyaman University, Adiyaman, Turkey
58: Also at Ozyegin University, Istanbul, Turkey
59: Also at Izmir Institute of Technology, Izmir, Turkey
60: Also at Marmara University, Istanbul, Turkey
61: Also at Kafkas University, Kars, Turkey
62: Also at Istanbul University, Faculty of Science, Istanbul, Turkey
63: Also at Istanbul Bilgi University, Istanbul, Turkey
64: Also at Hacettepe University, Ankara, Turkey
65: Also at Rutherford Appleton Laboratory, Didcot, United Kingdom
66: Also at School of Physics and Astronomy, University of Southampton, Southampton, United Kingdom
67: Also at Monash University, Faculty of Science, Clayton, Australia
68: Also at Bethel University, St. Paul, U.S.A.
69: Also at Karamanoğlu Mehmetbey University, Karaman, Turkey
70: Also at Purdue University, West Lafayette, U.S.A.
71: Also at Bingol University, Bingol, Turkey
72: Also at Sinop University, Sinop, Turkey
73: Also at Mimar Sinan University, Istanbul, Istanbul, Turkey
74: Also at Texas A&M University at Qatar, Doha, Qatar
75: Also at Kyungpook National University, Daegu, Korea
76: Also at University of Hyderabad, Hyderabad, India