



# Photon Reconstruction in the Belle II Calorimeter Using Graph Neural Networks

F. Wemmer<sup>1</sup> · I. Haide<sup>2</sup> · J. Eppelt<sup>3</sup> · T. Ferber<sup>4</sup> · A. Beaubien<sup>5</sup> · P. Branchini<sup>6</sup> · M. Campajola<sup>7</sup> · C. Cecchi<sup>8</sup> · P. Cheema<sup>9</sup> · G. De Nardo<sup>10</sup> · C. Hearty<sup>11</sup> · A. Kuzmin<sup>12</sup> · S. Longo<sup>13</sup> · E. Manoni<sup>14</sup> · F. Meier<sup>15</sup> · M. Merola<sup>16</sup> · K. Miyabayashi<sup>17</sup> · S. Moneta<sup>18</sup> · M. Remnev<sup>19</sup> · J. M. Roney<sup>20</sup> · J.-G. Shiu<sup>21</sup> · B. Shwartz<sup>22</sup> · Y. Unno<sup>23</sup> · R. van Tonder<sup>24</sup> · R. Volpe<sup>25</sup>

Received: 7 June 2023 / Accepted: 25 October 2023 / Published online: 15 December 2023  
© The Author(s) 2023

## Abstract

We present the study of a fuzzy clustering algorithm for the Belle II electromagnetic calorimeter using Graph Neural Networks. We use a realistic detector simulation including simulated beam backgrounds and focus on the reconstruction of both isolated and overlapping photons. We find significant improvements of the energy resolution compared to the currently used reconstruction algorithm for both isolated and overlapping photons of more than 30% for photons with energies  $E_\gamma < 0.5$  GeV and high levels of beam backgrounds. Overall, the GNN reconstruction improves the resolution and reduces the tails of the reconstructed energy distribution and therefore is a promising option for the upcoming high luminosity running of Belle II.

**Keywords** Calorimeter · Photon reconstruction · Overlapping clusters · High background · Fuzzy clustering · Machine learning · Deep learning · Graph neural networks · End-to-end representation spaces

## Introduction

The Belle II experiment is located at the high-intensity, asymmetric electron-positron-collider SuperKEKB in Tsukuba, Japan. SuperKEKB is colliding 4 GeV positron and 7 GeV electron beams at a center-of-mass energy of

around 10.58 GeV to search for rare meson decays and new physics phenomena. Many of these decays include photons in the final state that are reconstructed exclusively in the electromagnetic calorimeter. The experimental program of Belle II targets a significantly increased instantaneous luminosity that ultimately exceeds the predecessor experiment

✉ T. Ferber  
torben.ferber@kit.edu

<sup>1</sup> <http://orcid.org/0000-0002-6475-0834>  
<sup>2</sup> <http://orcid.org/0000-0003-0962-6344>  
<sup>3</sup> <http://orcid.org/0000-0001-8368-3721>  
<sup>4</sup> <http://orcid.org/0000-0002-6849-0427>  
<sup>5</sup> <http://orcid.org/0000-0001-9438-089X>  
<sup>6</sup> <http://orcid.org/0000-0002-2270-9673>  
<sup>7</sup> <http://orcid.org/0000-0003-2518-7134>  
<sup>8</sup> <http://orcid.org/0000-0002-2192-8233>  
<sup>9</sup> <http://orcid.org/0000-0001-8472-5727>  
<sup>10</sup> <http://orcid.org/0000-0002-2047-9675>  
<sup>11</sup> <http://orcid.org/0000-0001-6568-0252>  
<sup>12</sup> <http://orcid.org/0000-0002-7011-5044>

<sup>13</sup> <http://orcid.org/0000-0002-8124-8969>  
<sup>14</sup> <http://orcid.org/0000-0002-9826-7947>  
<sup>15</sup> <http://orcid.org/0000-0002-6088-0412>  
<sup>16</sup> <http://orcid.org/0000-0002-7082-8108>  
<sup>17</sup> <http://orcid.org/0000-0003-4352-734X>  
<sup>18</sup> <http://orcid.org/0000-0003-2184-7510>  
<sup>19</sup> <http://orcid.org/0000-0001-6975-1724>  
<sup>20</sup> <http://orcid.org/0000-0001-7802-4617>  
<sup>21</sup> <http://orcid.org/0000-0002-8478-5639>  
<sup>22</sup> <http://orcid.org/0000-0002-1456-1496>  
<sup>23</sup> <http://orcid.org/0000-0003-3355-765X>  
<sup>24</sup> <http://orcid.org/0000-0002-7448-4816>  
<sup>25</sup> <http://orcid.org/0000-0003-1782-2978>

by a factor of 30. This increase in luminosity also leads to a significant increase in beam-induced backgrounds [1]. These background processes produce both high-energy particle interactions that could be misidentified as physics signals, but also energy depositions of low-energy particles that degrade the energy resolution of the electromagnetic crystal calorimeter. The electronics signals from the calorimeter are interpreted during a process called reconstruction to determine the properties of particles that created the signals.

In this paper, we describe a fuzzy clustering algorithm based on Graph Neural Networks (GNNs) to reconstruct photons. The term fuzzy clustering [2] refers to the partial assignment of individual calorimeter crystals to several clustering classes. In our case, these are potentially overlapping, different signal photons, but also a beam background class.

The paper is organized as follows: Sect. 2 gives an overview of related work on Machine Learning for calorimeter reconstruction. Section 3 describes the Belle II electromagnetic calorimeter. The event simulation and details of the beam background simulation are discussed in Sect. 4. The conventional Belle II reconstruction algorithm and the new GNN algorithm are described in Sect. 5. We introduce the metrics used to measure the performance of the GNN algorithm in Sect. 6. The main performance studies and results are discussed in Sect. 7. We summarize our results in Sect. 8.

## Related Work

Machine Learning is widely used in high energy physics for the reconstruction of calorimeter signals both for clustering [3, 4], energy regression [5, 6], but also particle identification [7, 8] and fast simulation [9–11]. Most of the recent work has been performed in the context of the high-granularity calorimeter (HGCal) at CMS [12, 13]. For Belle II, the use of machine learning utilizing the electromagnetic calorimeter is so far limited to image-based particle identification in the barrel [8, 14].

GNNs are now widely recognized as one possible solution for irregular geometries in high energy physics [15–17]. GNN architectures that are able to learn a latent space representation of the detector geometry itself [18, 19] are the basis of the work presented in this paper.

Previous work has focused on simplified and idealized detector geometries, often approximated as a regular grid of readout cells expressed as 2D or 3D images. Additionally, the presence of geometry changes and overlaps between barrel and endcap regions, large variations of cell sizes, and the presence of very high spatially non-uniform noise levels induced by beam background energy depositions are neglected.

For a complete list of works in particle physics that utilize machine learning, we refer to the review [20].

## The Belle II Electromagnetic Calorimeter

The Belle II detector consists of several subdetectors arranged around the beam pipe in a cylindrical structure that is described in detail in Ref. [21, 22]. We define the  $z$ -axis of the laboratory frame as the central axis of the solenoid. The positive direction is pointing in the direction of the electron beam. The  $x$  axis is horizontal and points away from the accelerator center, while the  $y$  axis is vertical and points upwards. The longitudinal direction, the transverse plane with azimuthal angle  $\phi$ , and the polar angle  $\theta$  are defined with respect to the detector's solenoidal axis.

The Belle II electromagnetic calorimeter (ECL) consists of 8736 Thallium-doped CsI (CsI(Tl)) crystals that are grouped in a forward endcap, covering a polar angle  $12.4^\circ < \theta < 31.4^\circ$ , a barrel, covering a polar angle  $32.2^\circ < \theta < 128.7^\circ$ , and a backward endcap, covering a polar angle  $130.7^\circ < \theta < 155.1^\circ$ . The crystals have a trapezoidal geometry with a nominal cross-sectional area of approximately  $6 \times 6 \text{ cm}^2$  and a length of 30 cm, providing 16.1 radiation lengths of material. While crystals in the barrel are similar in cross-section and shape, the crystals in the endcaps vary with masses between 4.03 kg and 5.94 kg [23]; crystals in the endcaps also have significantly more passive material in front of the crystals. Each crystal is aligned in the direction of the collision point with a small tilt in polar angle  $\theta$  to reduce detection inefficiencies from particles passing between two crystals. Crystals in the barrel additionally have a small tilt in azimuthal angle  $\phi$ . The scintillation light produced in the CsI(Tl) crystals is read out by two photodiodes glued to the back of each crystal. After shaping electronics, the waveform is digitized and the crystal energy  $E_{\text{rec}}^{\text{crystal}}$  over baseline and time  $t_{\text{rec}}^{\text{crystal}}$  since trigger time of the energy deposition are reconstructed online using FPGAs [24]. Waveforms of crystals with energy depositions above 50 MeV are stored for offline processing to allow for electromagnetic vs. hadronic shower identification through pulse shape discrimination (PSD) [25]. Available information from PSD is

- the fit type ID of a multi-template fit indicating which of the possible templates provides the best goodness-of-fit,
- the respective  $\chi^2$  value as an indicator of the goodness-of-fit,
- and the ratio of reconstructed hadronic and photon template energies, referred to as PSD hadronic energy ratio in the following.

## Data Set

In this work, we use simulated events to train and evaluate the reconstruction algorithms. The detector geometry and interactions of final-state particles with detector materials are simulated using GEANT4 [26] combined with a dedicated detector response simulation. Simulated events are reconstructed and analyzed using the Belle II Analysis Software Framework (basf2) [27, 28]. We simulate isolated photons, with energy  $0.1 < E_{\text{gen}} < 1.5 \text{ GeV}$ , and direction  $17^\circ < \theta_{\text{gen}} < 150^\circ$  and  $0^\circ < \phi_{\text{gen}} < 360^\circ$  drawn randomly from independent uniform distributions in  $E$ ,  $\theta$ , and  $\phi$ . The generation vertex of the photons is  $x = 0$ ,  $y = 0$ , and  $z = 0$ . For events with two overlapping photons, we first draw randomly one photon with independent uniform distributions as outline above. We then simulate a second photon with an angular separation  $2.9 < \Delta\alpha < 9.7^\circ$  drawn randomly from uniform distributions in  $\Delta\alpha$  and in  $E$ . This angular separation covers approximately the distance needed to create two overlapping clusters. These two cases are typical calorimeter signatures in Belle II that describe the majority of photons. We note that the reconstructions of hadrons is a more difficult task not yet covered by our algorithm.

As part of the simulation, we overlay simulated beam background events corresponding to different collision conditions to our signal particles [1, 29]. The simulated beam backgrounds correspond to an instantaneous luminosity of  $L_{\text{beam}} = 1.06 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$  (called *low beam background*), and  $L_{\text{beam}} = 8 \times 10^{35} \text{ cm}^{-2}\text{s}^{-1}$  (called *high beam background*). Those two values approximately correspond to the conditions in 2021, and the expected conditions slightly above the design luminosity, respectively. The spatial distribution of beam backgrounds is asymmetric: They are much higher in the backward endcap than in the forward endcap, and they are slightly higher in the barrel than in the forward endcap. Additional electronics noise per crystal of about 0.35 MeV is included in our simulation as well.

The supervised training and the performance evaluation both use labeled information that relies on matching reconstructed information with the simulated *truth* information. For each of the four configurations, isolated and overlapping photons with low and high beam backgrounds, we use 1.8 million events for training and 200,000 events for validation. The performance evaluation is carried out on a large number of statistically independent samples simulated with various energies and in different detector regions.

We then study the performance of the GNN clustering algorithm in all four scenarios and compare it to the baseline basf2 reconstruction. Both reconstruction algorithms are described in detail in Sect. 5.

## Isolated Photon

To study isolated photons, we use the simulated events with a generated isolated photon only. For each event, we select a region of interest (ROI): We first determine the azimuthal angles of the fourth neighbour on either side of the local maximum (LM), and the polar angles of the fourth neighbours on either direction of the LM. We then include all crystals in that angular range. In the barrel this defines a regular  $9 \times 9$  array of crystals centered around a LM, while in the endcaps this array is not necessarily regular, but can contain a few crystals more or less. The LM is a crystal with at least 10 MeV of reconstructed crystal energy, and energy higher than all its direct eight neighbors. The LM must be the only LM in the ROI, and the matched truth particle must be a simulated photon responsible for at least 20% of the reconstructed crystal energy. Precisely, for the LM we require the ratio

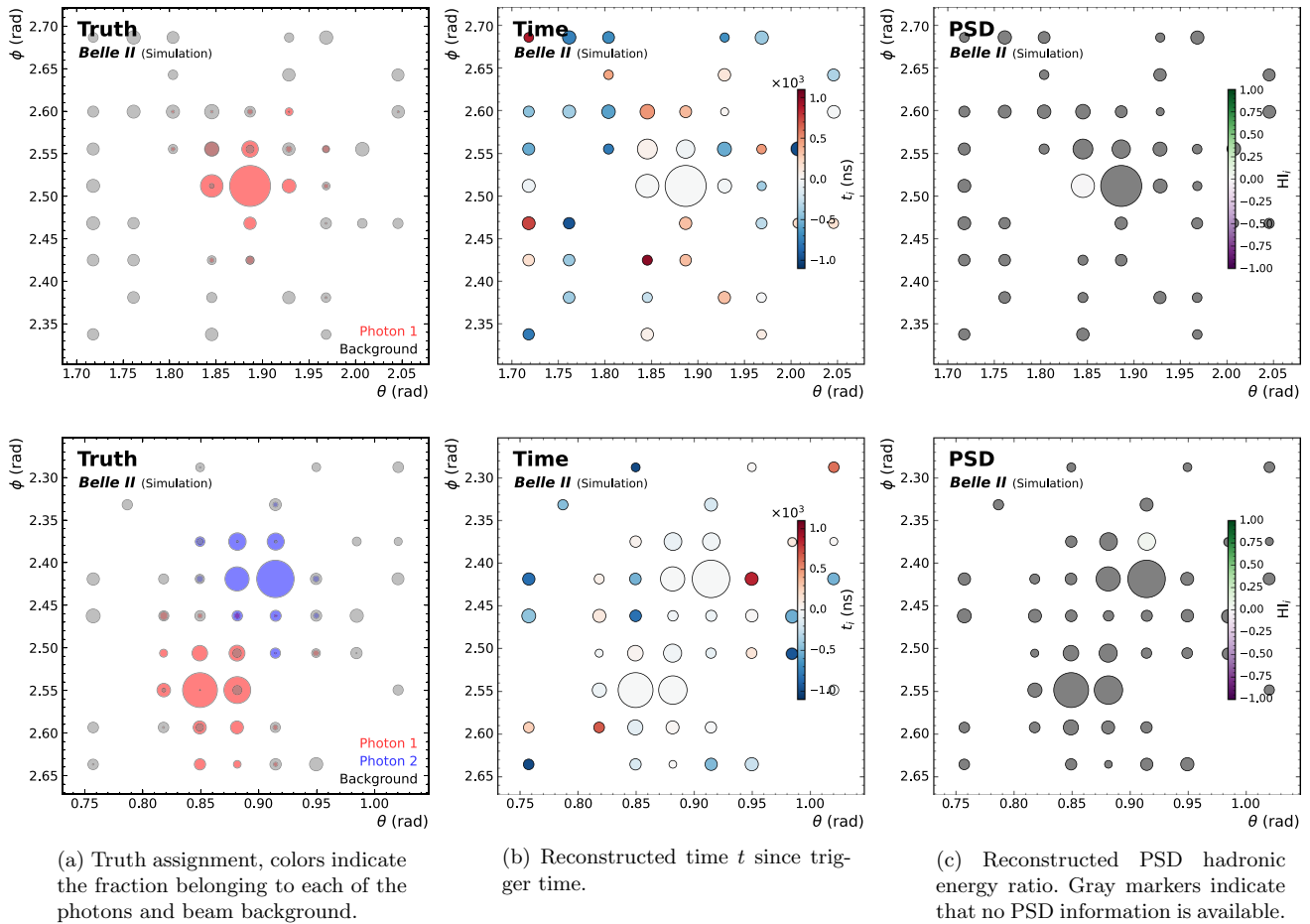
$$r_{\text{LM}}^{\gamma_1} = \frac{E_{\text{dep}}^{\gamma_1, \text{crystal}_{\text{LM}}}}{E_{\text{rec}}^{\text{crystal}_{\text{LM}}}} \geq 0.2. \tag{1}$$

here,  $E_{\text{dep}}^{\gamma_1, \text{crystal}_{\text{LM}}}$  denotes the truth energy deposition of photon 1 in the LM, and  $E_{\text{rec}}^{\text{crystal}_{\text{LM}}}$  the reconstructed crystal energy in the LM. The crystals contained in the ROI are considered for the clustering by the GNN algorithm and significantly extend the  $5 \times 5$  area considered by the baseline algorithm (Sect. 5). Furthermore, the ROI represents the area of the local coordinate system later used as an input feature, with the LM as the origin. Figure 1 (top) shows a typical isolated photon event with high beam background.

## Overlapping Photons

Two different photons that deposit some of their energy in identical crystals are referred to as overlapping photons. To study overlapping photons, we use the simulated events with two overlapping photons only. We select events that have exactly two LMs that must fulfill the following selection criteria:

- a) each LM must have reconstructed crystal energies greater than 10 MeV,
- b)  $r_{\text{LM}_1}^{\gamma_1} \geq 0.2$  and  $r_{\text{LM}_1}^{\gamma_1} > r_{\text{LM}_1}^{\gamma_2}$ ,
- c)  $r_{\text{LM}_2}^{\gamma_2} \geq 0.2$  and  $r_{\text{LM}_2}^{\gamma_2} > r_{\text{LM}_2}^{\gamma_1}$ .



**Fig. 1** Typical event displays showing (left) simulated truth assignments, (center) input variables time, and (right) PSD hadronic energy ratio for (top) isolated and (bottom) overlapping photons for two example events with high beam background. The marker centers indi-

cate the crystal centers, the marker area is proportional to the truth energy deposition for the left plots; it is proportional to the reconstructed crystal energy for the other plots

We refer to criteria a)-c) as *LM separation criteria* since they ensure that the particles form two separate LMs. Additionally, events must meet the *overlap criterion*:

- d) each of the two photons must deposit at least 10 MeV energy in shared crystals within a  $5 \times 5$  area around its respective LM.

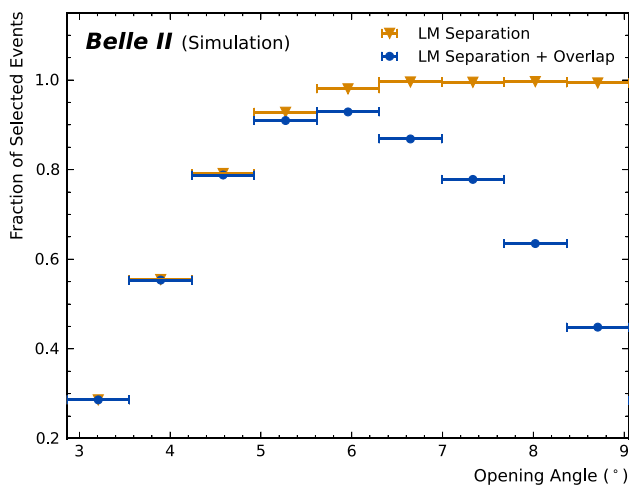
Figure 2 shows the fraction of events accepted by these selections as a function of the simulated opening angle. In the scope of this paper, we additionally require LMs to exclusively originate from simulated particles without additional LMs, e.g. from beam background, in the ROI, that is:

- e) the two LMs must be the only ones in the ROI and they must be truth-matched to the simulated photons.

Finally, we remove rare cases of small truth energy depositions and large backgrounds, by requiring:

- f) the crystal with the largest truth energy deposition of a photon must be within a  $5 \times 5$  area around its corresponding LM.

We then create a ROI centered at the midpoint between the two LMs, calculated using the shortest distance between two LMs projected onto the surface of a sphere. The crystal closest to the midpoint is defined as the ROI center. The LM positions for this are determined by interpreting the global LM coordinates of their associated crystals as latitude and longitude. Figure 1 (bottom) shows an overlapping photon event with high beam background.



**Fig. 2** Fraction of selected overlapping photon events in the barrel as a function of generated opening angle. The orange markers correspond to events fulfilling LM separation criteria (a–c); the blue markers correspond to events that additionally pass the overlap criterion (d) (see text for details)

The truth energy deposition per photon and the reconstructed crystal energy  $E_{rec}^{crystal}$ , crystal time  $t_{rec}^{crystal}$ , crystal PSD information (see Sect. 3), and the LM positions within the ROI are recorded for each event.

### Reconstruction Algorithms

Interactions of energetic photons in the Belle II ECL typically deposit energy in up to  $5 \times 5$  crystals. The task of the clustering reconstruction algorithms is to select a set of crystals that contains all the energy of the incoming photon, but no energy from other particles or from beam background. Low beam background results in approximately 17% of all crystals in the ECL having significant reconstructed energy  $E_{rec}^{crystal} \geq 1\text{MeV}$ ; for high beam backgrounds this number is expected to increase to about 40%. This increase in the number of crystals to consider in the clustering, adds to the complexity of the reconstruction.

### Baseline

The baseline algorithm is designed to provide maximum efficiency for cluster finding, contain all crystals from the incoming particle for particle identification, and select an optimal subset of the cluster crystals that provides the best energy resolution [21]. The clustering is performed in three steps. In the first step, all crystals are grouped into a connected set of crystals, so-called connected regions starting with LMs, as defined previously. In an iterative procedure all direct neighbors with energies above

0.5 MeV are added to this LM, and the process is continued if any neighbor itself has energy above 10 MeV. Overlapping connected regions are merged into one.

In the second step, each connected region is split into clusters, one per LM. If there is only one LM in the connected region, up to 21 crystals in a  $5 \times 5$  area excluding corners centered at the local maximum are grouped into a cluster. If there is more than one LM in a connected region, the energy in each crystal of the connected region is assigned a distance-dependent weight and can be shared between different clusters. The distance is calculated from the cluster centroid to each crystal center, where the cluster centroid is updated iteratively using logarithmic energy weights. This process is repeated until all cluster centroids in a connected region are stable within 1 mm.

In a third step, an optimal subset, including the  $n$  highest energetic crystals of all non-zero weighted crystals that minimize the energy resolution, is used to predict the cluster energy  $E_{rec}^{basf2}$ .  $n$  depends on the measured noise in the event, and on the energy of the LM itself. The noise level is estimated by counting the number of crystals in the event containing more than 5 MeV that have times  $t$  more than 125 ns from the trigger time.  $E_{rec}^{basf2}$  is also corrected already within `basf2` for possible bias using simulated events. This bias includes leakage (energy not deposited in the crystals included in the energy sum) and beam backgrounds (energy included in the sum that is not from the signal photon).  $E_{rec}^{basf2}$  is the estimator for the generated energy of a particle.

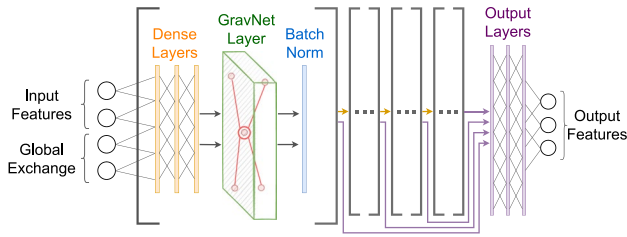
The `basf2` clustering algorithm also returns a cluster energy  $E_{rec,raw}^{basf2}$  that is not corrected for energy bias.  $E_{rec,raw}^{basf2}$  is the estimator for the deposited energy of a particle.

### Graph Neural Network Architecture

GNN architectures have shown that they are powerful network types to deal with both irregular geometries and varying input sizes. In this work, all crystals of an ROI with an energy deposition above 1 MeV are interpreted as nodes in a graph, which leads to variable input sizes and is thus a good use case for GNNs. The implementation of this GNN is done in `PyTorch Geometric` [30].

The input features consist of crystal properties and crystal measurements: The global coordinates  $\theta$  and  $\phi$  of each crystal, the local coordinates  $\theta'$  and  $\phi'$  with respect to the ROI center, the crystal mass, and the LM(s) (in one-hot encoding) represent crystal properties. The crystal energy  $E_{rec}^{crystal}$  in GeV, the time  $t_{rec}^{crystal}$  in  $\mu\text{s}$ , and the PSD fit type, PSD  $\chi^2$ , and PSD hadronic energy ratio are crystal measurements used as input features. Pre-processing scales the input uniformly before further processing with the GNN: All features are min-max normalized to an interval of [0, 1] with





**Fig. 3** An illustration of the GNN architecture. Each pair of gray, square brackets represents one GravNet block consisting of dense layers, a GravNet layer and a batch norm layer. The input features describe the feature vector of one node. The global exchange denotes appending the average each input features over all nodes in the ROI

the exception of  $t_{rec}^{crystal}$  and the PSD hadronic energy ratio which are both normalized to the interval  $[-1, 1]$ . The global coordinates and the crystal masses are normalized based on the range of coordinates and masses of all crystals in the detector instead of only the ones in the ROI. Additionally, we average each input feature over all nodes in the ROI and concatenate the averaged input features as additional inputs, thus enabling a global exchange of information.

As displayed in Fig. 3, our model is built out of four so-called GravNet [19] blocks of which the concatenated outputs are passed through three dense output layers with a final softmax activation function. Each GravNet block features three dense layers at the beginning of the block, the initial two of which with ELU [31] activation functions and the last one with a tanh activation function. The dense layers feed into a GravNet layer and the overall GravNet block is concluded by a batch normalization layer [32]. The GravNet layer is responsible for the graph building and subsequent message passing between the nodes of the graph. It first translates the input features into two learned representation spaces: one representing spatial information  $S$  while the other, denoted  $F_{LR}$ , contains the transformed features used for message passing. In the second step, each node is connected to its  $k$  nearest neighbors defined by the Euclidean distances in  $S$ , thus creating an undirected, connected graph. For each node, the input features of connected nodes are then weighted by a Gaussian potential depending on the distance in  $S$  and aggregated by summation. The resulting features are concatenated with the GravNet input features and, after batch normalization, passed to the next GravNet block and to the dense output layers.

The implementation in the present work follows the concept of fuzzy clustering which refers to the partial assignment of individual crystals to several clustering classes. Consequently, the GNN predicts weights  $w_i^X$  that indicate the proportion of the reconstructed energy  $E_{rec}^{crystal_i}$  in a crystal  $i$  that belongs to a clustering class  $X$ . For models used

with isolated photons,  $X \in \{\gamma_1, \text{background}\}$ , for models with overlapping photons  $X \in \{\gamma_1, \gamma_2, \text{background}\}$ . As a loss function, we then use the Mean Squared Error (MSE) between the true and predicted weights summed over all classes and crystals. The training is stopped when there has been no improvement for 15 epochs in the optimization objective. For low beam background models that objective is the MSE loss on the validation data set, whereas the high beam background models employ the more high-level  $FWHM_{dep}$  (Sect. 6) on the validation data set.

Hyperparameters have been chosen through a hyperparameter optimization using Optuna [33]. The optimization is done with respect to the  $FWHM_{dep}$  (Sec. 6) instead of the loss function. We optimize the two models trained for high beam backgrounds and use the respective hyperparameters also for the corresponding low beam background models. The final hyperparameters for both the isolated photon models and the overlapping photon models are shown in Table 1.

The learning rate, the number of dense layers in each GravNet block, and all dimensions of the output layers have been manually optimized by testing a reasonable range of values. The learning rate is set to  $5 \times 10^{-3}$  and is subject to a decay factor of 0.25 after every five epochs of stagnating validation loss. We did not observe significant over-training and as a consequence, we do not use dropout layers or other regularization methods but rely on the large data set.

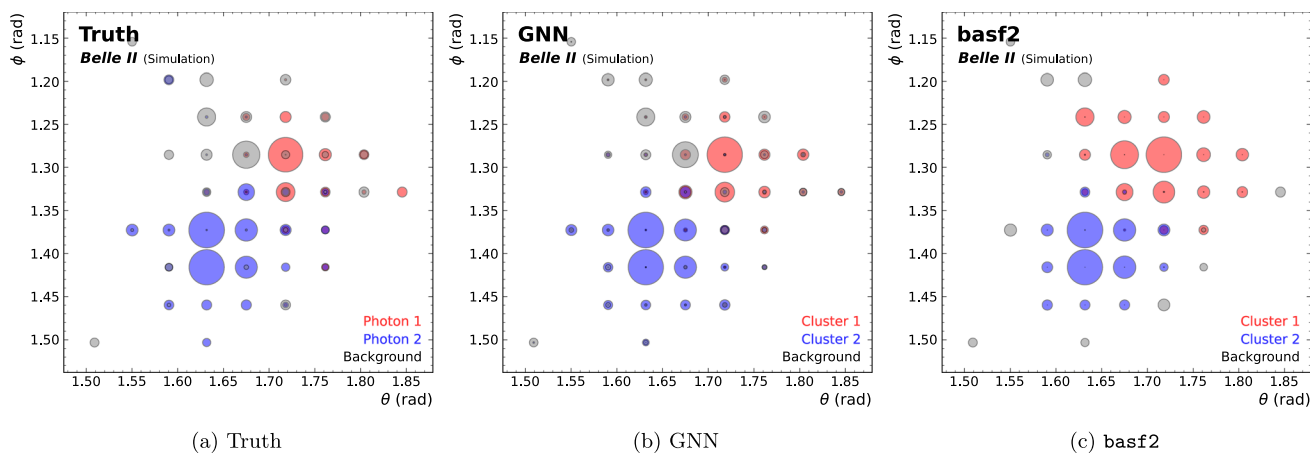
The GNN algorithm yields the weights  $w_i^X$  per crystal for all crystals in the ROI with an energy deposition above 1 MeV. In order to reconstruct the total cluster energy  $E_{rec}^{GNN}$  associated with a certain particle, we then sum over all specific weights multiplied by the reconstructed energies per crystal,  $E_{rec}^{GNN} = \sum w_i^X E_{rec}^{crystal_i}$ .

Figure 4 shows how the GNN and the `basf2` algorithms behave in clustering a typical case of overlapping photons.

**Table 1** Optimized hyperparameters of the isolated photon, and overlapping photon GravNet models

Hyperparameter	Isolated photon models	Overlapping photon models
Width of the dense layers, $F_{IN}, F_{OUT}$	22	24
Feature space dimension $F_{LR}$	16	16
Spatial information space dimension $S$	6	6
Connected nearest neighbors $k$	14	16
Batch norm momentum	0.01	0.4
Stacked GravNet blocks	4	4
Batch size	1024	512

The hyperparameters are the result of an optimization of the  $FWHM_{dep}$  on the respective high background validation data set



**Fig. 4** Comparison of (a) truth energy fractions, (b) reconstructed energy fraction by the GNN, and (c) reconstructed energy fraction by *basf2* for an example event with high beam background. Colors indicate the fractions belonging to each photon or background. The

marker centers indicate the crystal centers, the marker area is proportional to the truth or reconstructed (GNN, *basf2*) energy deposition respectively

### Metrics

For performance evaluation, the reconstructed energy of a particle is compared with two different truth targets: the total deposited truth energy  $E_{dep}$  per photon in the ROI, and the generated truth energy  $E_{gen}$  per photon. This results in two variants of relative reconstruction errors. The reconstruction error on the deposited energy

$$\eta_{dep}^{basf2} = \frac{E_{rec,raw}^{basf2} - E_{dep}}{E_{dep}} \quad \text{and} \quad (2)$$

$$\eta_{dep}^{GNN} = \frac{E_{rec}^{GNN} - E_{dep}}{E_{dep}}$$

gives access to the energy resolution ignoring leakage and other detector effects. It is a direct evaluation of the clustering performance of an algorithm.

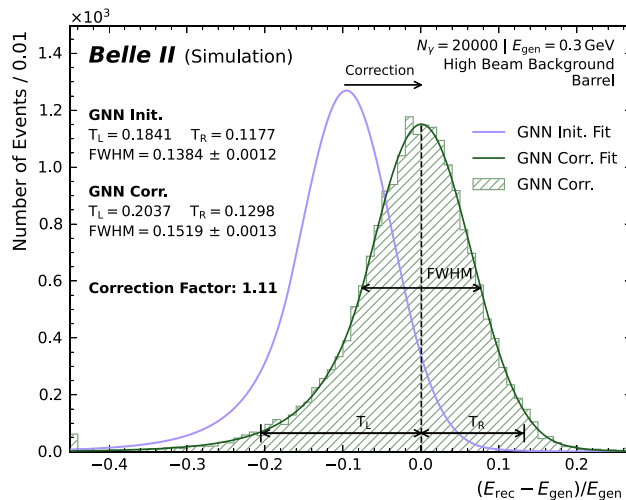
On the other hand, the reconstruction error on the generated energy

$$\eta_{gen}^{basf2} = \frac{E_{rec}^{basf2} - E_{gen}}{E_{gen}} \quad \text{and} \quad (3)$$

$$\eta_{gen}^{GNN} = \frac{E_{rec}^{GNN} - E_{gen}}{E_{gen}}$$

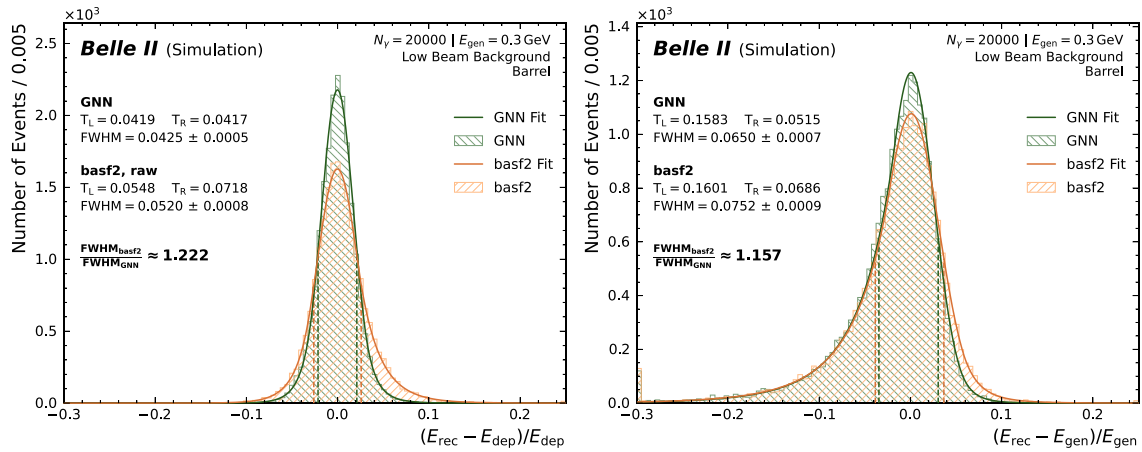
factors in all detector and physics effects and quantifies how much of the improvements to the underlying clustering carry over to downstream physics object reconstruction.

Evaluating both algorithms on a large number of simulated photons yields peaking distributions in both reconstruction errors  $\eta_{dep}$  and  $\eta_{gen}$ . Both distributions



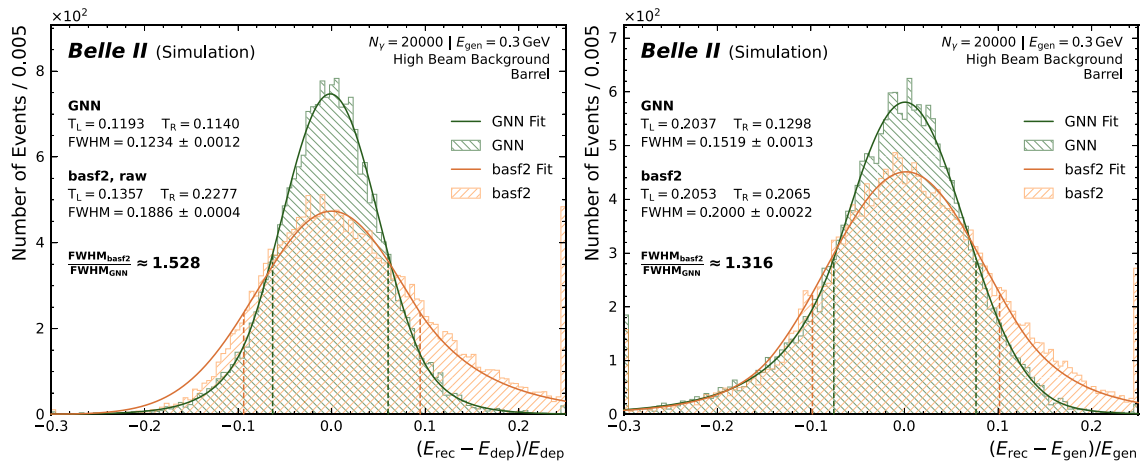
**Fig. 5** Example distribution of the relative reconstruction error  $\eta_{gen}$  of the generated energy and illustration of the bias correction, the FWHM, and the tail ranges

are potentially biased because of energy leakage and the presence of beam backgrounds (see Sect. 5.1). We perform a binned fit using a double-sided crystal ball [34, 35] function as probability density function (pdf) with the *kafe2* [?] framework. We shift all reconstruction error distributions independently by a multiplicative factor to correct the difference between the fitted peak position and zero (Fig. 5). Since  $\eta_{dep}$  and  $\eta_{gen}$  are asymmetric distributions, we repeat this procedure until the difference between the fitted peak position and zero is less than 0.002. This procedure usually converges within two or three iterations.



(a) Relative reconstruction error  $\eta_{dep}$  of the deposited energy. (b) Relative reconstruction error  $\eta_{gen}$  of the generated energy.

**Fig. 6** Distribution of relative reconstruction errors (a)  $\eta_{dep}$  and (b)  $\eta_{gen}$  for isolated clusters for low beam backgrounds. The first bin contains all underflow entries; the last bin contains all overflow entries



(a) Relative reconstruction error  $\eta_{dep}$  of the deposited energy. (b) Relative reconstruction error  $\eta_{gen}$  of the generated energy.

**Fig. 7** Distributions of relative reconstruction errors (a)  $\eta_{dep}$  and (b)  $\eta_{gen}$  for isolated clusters for high beam backgrounds. The first bin contains all underflow entries; the last bin contains all overflow entries

We then determine the full width half maximum (FWHM) of the final shifted distributions in  $\eta_{dep}$  and  $\eta_{gen}$ , yielding  $FWHM_{dep}$  and  $FWHM_{gen}$  respectively. The uncertainty on the FWHM is calculated from the uncertainties of the fit parameters. In addition to the FWHM, we determine the tails of the reconstruction error distribution. The left and right tails  $T_{L,R}$  are calculated as the 95th percentile when ranking the unbinned events on the respective side of the peak position, as given by the fit parameters, in ascending order ( $T_R$ ) and descending order ( $T_L$ ) respectively. Propagating the uncertainty on the peak position as given by the fit yields the uncertainty on  $T_{L,R}$ .

## Results

The first sections of the results focus on detailed studies of isolated clusters. Section 7.4 then introduces overlapping clusters and their effects on the performance. Figure 6 shows examples for the distributions of both reconstruction errors  $\eta_{dep}$  and  $\eta_{gen}$ , as well as the fit results for events with low beam background. Figure 7 shows the equivalent distributions for events with high beam background.



The  $\eta_{\text{gen}}$  distributions are wider because the reconstruction error includes the effects of leakage which result in missing energy with respect to the generated photon energy. This only affects the left-side tails.

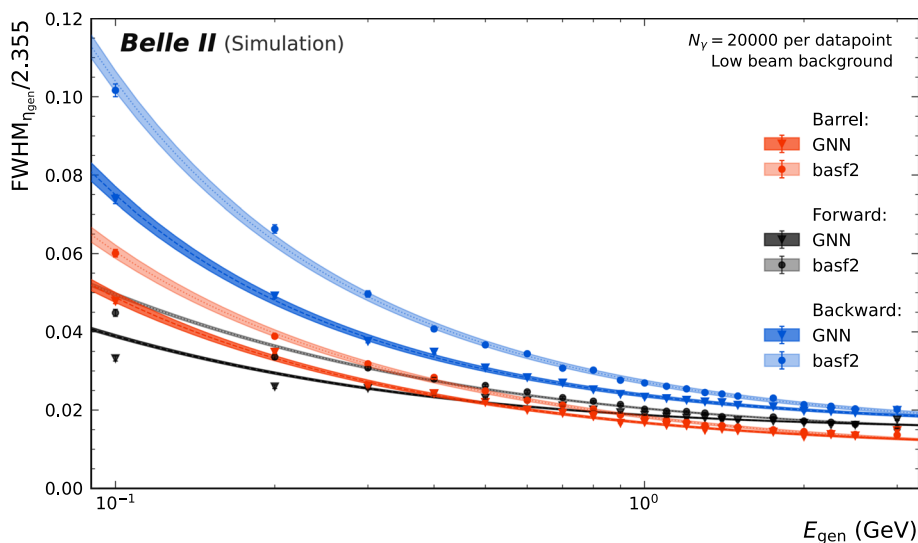
In the following subsections, we are comparing the performance of the GNN and the *basf2* reconstruction algorithms for different detector regions for low and high beam backgrounds by evaluating the energy resolution  $\text{FWHM}_{\text{gen}}/2.355$  and the tail parameters. We then analyze the GNN in more detail by testing the input variable dependencies and the robustness against differences in beam background levels between training and evaluation.

### Energy Resolution and Energy Tails

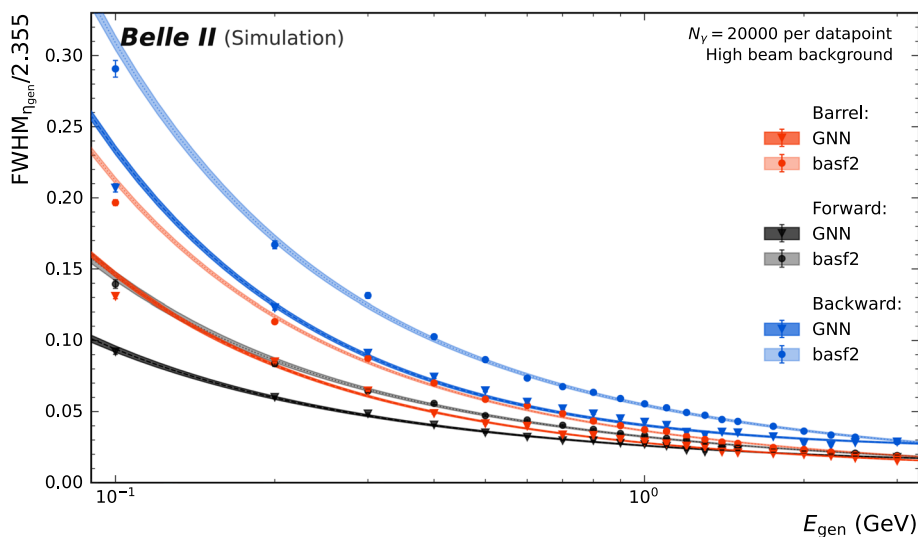
The three detector regions barrel, forward endcap, and backward endcap described in Sect. 3 differ in crystal geometry, levels of background, and amount of passive material before and in between crystals. The following section studies the variations in the energy reconstruction performance that arise as a direct result of these differences.

In order to access the energy dependence of the resolution and tail parameters we simulate test data sets of photons at various fixed energies. The FWHM for each simulated data set is then determined according to Sect. 6. Plotting the

**Fig. 8** Resolution  $\text{FWHM}_{\text{gen}}/2.355$  of the GNN and *basf2* as function of the simulated photon energy  $E_{\text{gen}}$  for both endcaps and the barrel for (a) low and (b) high beam background. Each color is associated with one detector region; the light color indicates *basf2*, the dark color the GNN. The bands indicate the uncertainty of the fits, see text for details. The fit parameters are summarized in Table 2



(a) Low beam background.



(b) High beam background.

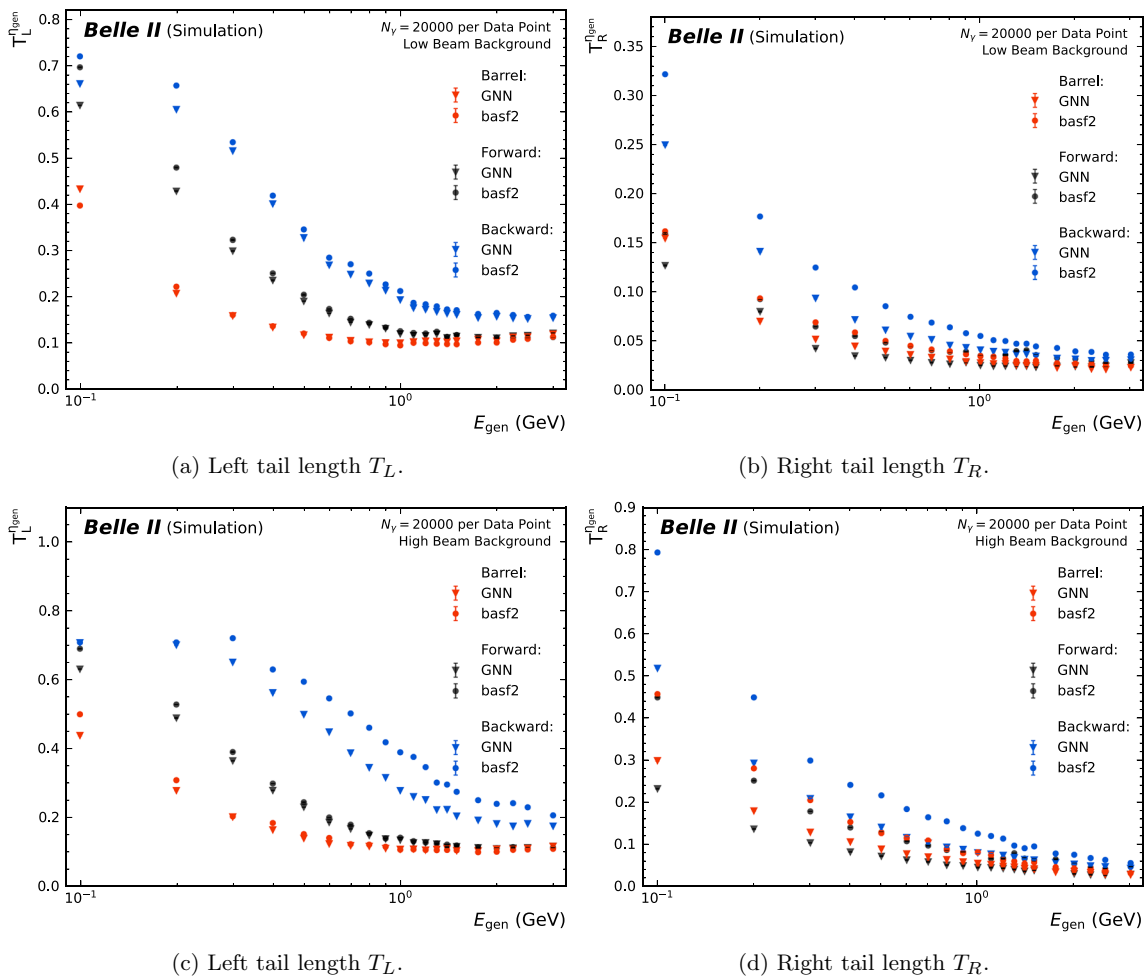
**Table 2** Fit results  $(a/E_{\text{gen}} \oplus b/\sqrt{E_{\text{gen}}} \oplus c)$  of the fits shown in Fig. 8

Region	Algorithm	Low beam background			High beam background		
		a ( $\times 10^{-2}$ )	b ( $\times 10^{-2}$ )	c ( $\times 10^{-2}$ )	a ( $\times 10^{-2}$ )	b ( $\times 10^{-2}$ )	c ( $\times 10^{-2}$ )
Barrel	GNN	$0.23 \pm 0.02$	$1.32 \pm 0.02$	$1.00 \pm 0.01$	$1.25 \pm 0.02$	$2.39 \pm 0.02$	$0.75 \pm 0.03$
	basf2	$0.35 \pm 0.02$	$1.54 \pm 0.02$	$0.91 \pm 0.02$	$1.88 \pm 0.02$	$3.11 \pm 0.03$	$0.31 \pm 0.10$
Forward	GNN	$0.00 \pm 0.14$	$1.11 \pm 0.01$	$1.49 \pm 0.00$	$0.61 \pm 0.03$	$2.23 \pm 0.02$	$1.20 \pm 0.02$
	basf2	$0.00 \pm 0.37$	$1.51 \pm 0.01$	$1.38 \pm 0.01$	$1.11 \pm 0.03$	$2.92 \pm 0.03$	$0.84 \pm 0.03$
Backward	GNN	$0.50 \pm 0.02$	$1.69 \pm 0.03$	$1.59 \pm 0.02$	$2.18 \pm 0.03$	$2.51 \pm 0.05$	$2.28 \pm 0.02$
	basf2	$0.78 \pm 0.03$	$2.12 \pm 0.04$	$1.50 \pm 0.03$	$2.72 \pm 0.05$	$4.64 \pm 0.05$	$0.91 \pm 0.08$

resolutions  $\text{FWHM}_{\text{gen}}/2.355$  over the generated photon energies  $E_{\text{gen}}$  reveals a characteristic relationship that is parameterized by the function  $a/E_{\text{gen}} \oplus b/\sqrt{E_{\text{gen}}} \oplus c$ , where  $\oplus$  indicates addition in quadrature.

Both the GNN as well as the baseline algorithm perform differently in regards to the energy resolution in all three detector parts, as can be seen in Fig. 8a for low beam

background and as Fig. 8b for high beam background. Table 2 reports the parameters of the fitted parameterization of the resolution. We attribute these difference to the large spread of both shape and size of crystals in the endcaps, the asymmetric distribution of beam backgrounds, and the different amount of passive material in front of the different detector regions.



**Fig. 9** 95% left- and right tail lengths  $T_L$  and  $T_R$  of  $\eta_{\text{gen}}$  for the GNN and basf2 as function of the simulated photon energy  $E_{\text{gen}}$  for both endcaps and the barrel for (a and b) low and (c and d) high beam background. Each color is associated with one detector region

Overall, the energy resolution of the GNN algorithm is significantly better than the baseline algorithm for all photon energies. The GNN energy resolution is better by more than 30% for photon energies below 500 MeV which is the energy range of more than 90% of all photons in *B*-meson decay chains. The higher the beam background, the larger the difference between the GNN and the baseline algorithm. The difference between the two algorithms decreases with energy because the relative contribution of beam backgrounds to the photon energy resolution decreases.

The shape of the left-side tails is dominated by passive material and is hence expected to be different in the different detector regions. The left-side tails are almost independent of beam backgrounds as can be seen by comparing Fig. 9a for low beam background and Fig. 9c for high beam background. The GNN and the baseline algorithm both show the smallest tail length for the barrel region with decreasing tail lengths for increasing energy. The left-side tails are largest in the backward endcap due to the highest ratio of passive to active material as expected. The right-side tails are mostly originating from beam background being wrongly added to photon clusters. The GNN produces shorter tails than the baseline algorithm for all energies and for both low and high beam backgrounds, with the performance difference increasing for lower energies and higher beam backgrounds.

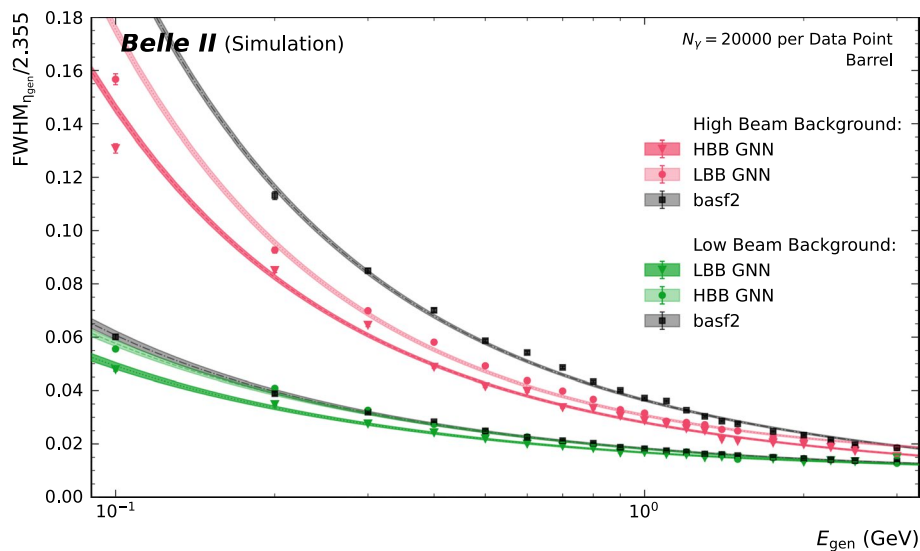
### Beam Background Robustness

The beam background levels are changing continuously during detector operations. Ideally, reconstruction algorithms at

Belle II are insensitive to such changes. The `basf2` baseline algorithm achieves robustness against increasing beam backgrounds by adaptively including fewer crystals in the energy sum calculation. Since our GNN is trained with a large number of events with event-by-event fluctuations of beam backgrounds, we expect robustness against varying beam backgrounds if the GNN generalizes well enough. We test the robustness of our GNN by comparing GNNs trained and tested on the same backgrounds, against GNNs trained and tested on the two different beam backgrounds (Fig. 10, parameterization in Table 3). While the GNNs trained on the same beam backgrounds achieve a better resolution than the ones trained on different beam backgrounds, the GNN still outperforms the baseline algorithm even for networks trained on the different beam backgrounds. This demonstrates a promising generalization with respect to different levels of beam backgrounds.

### Input Parameter Dependency

As discussed in Sect. 3, multiple input features are available for the GNN, while the `basf2` algorithm uses crystal position and energy only. This section presents a study of the influence of the input features on the FWHM. For that, the architecture described in Sect. 5.2 is trained on isolated photon events with low or high beam backgrounds using different combinations of input features. The 200,000 events from the respective validation data set, as described in Sect. 4, are used for inference. The data set covers an energy range of  $0.1 < E_{\text{gen}} < 1.5$  GeV and the full detector range



**Fig. 10** Resolution  $\text{FWHM}_{\text{gen}}/2.355$  as a function of the simulated photon energy  $E_{\text{gen}}$  for the GNNs trained with low beam background (LBB GNN) and high beam background (HBB GNN) in the barrel. The color is associated with the evaluation on either beam background; the dark color indicates the model trained with the beam

background identical to the evaluation, and the light color indicates the model trained with the respective other beam background. The bands indicate the uncertainty of the fits, see text for details. The fit parameters are summarized in Table 3. The resolution of the `basf2` algorithm is shown for comparison

**Table 3** Fit results ( $a/E_{\text{gen}} \oplus b/\sqrt{E_{\text{gen}}} \oplus c$ ) of the fits shown in Fig. 10 for the GNN trained with low beam background (LBB GNN) and high beam background (HBB GNN)

Region	Algorithm	Low Beam Background			High Beam Background		
		a ( $\times 10^{-2}$ )	b ( $\times 10^{-2}$ )	c ( $\times 10^{-2}$ )	a ( $\times 10^{-2}$ )	b ( $\times 10^{-2}$ )	c ( $\times 10^{-2}$ )
Barrel	LBB GNN	$0.23 \pm 0.02$	$1.32 \pm 0.02$	$1.00 \pm 0.01$	$1.59 \pm 0.02$	$2.27 \pm 0.03$	$1.32 \pm 0.02$
	HBB GNN	$0.28 \pm 0.02$	$1.58 \pm 0.01$	$0.85 \pm 0.02$	$1.25 \pm 0.02$	$2.39 \pm 0.02$	$0.75 \pm 0.03$

The values for the LBB GNN inferred on low beam background test samples, and for the HBB GNN inferred on high beam background are identical to the ones reported in Table 2

**Table 4** Comparison of the performances of GNN models with different additional input features, and the performance of the `basf2` baseline

Input features	Low beam background		High beam background	
	$\text{FWHM}_{\text{dep}} \times 10^{-2}$	$\text{FWHM}_{\text{gen}} \times 10^{-2}$	$\text{FWHM}_{\text{dep}} \times 10^{-2}$	$\text{FWHM}_{\text{gen}} \times 10^{-2}$
Energy	$2.17 \pm 0.01$	$5.25 \pm 0.02$	$5.05 \pm 0.03$	$8.08 \pm 0.04$
Energy, local coordinates	$2.11 \pm 0.02$	$5.19 \pm 0.02$	$5.04 \pm 0.04$	$8.04 \pm 0.04$
Energy, local coordinates, PSD	$2.19 \pm 0.01$	$5.20 \pm 0.02$	$5.06 \pm 0.03$	$8.07 \pm 0.04$
Energy, local coordinates, time	$1.72 \pm 0.01$	$4.85 \pm 0.02$	$4.52 \pm 0.03$	$7.63 \pm 0.03$
Energy, local coordinates, time, PSD	$1.72 \pm 0.01$	$4.85 \pm 0.02$	$4.51 \pm 0.03$	$7.62 \pm 0.03$
<code>basf2</code>	$2.32 \pm 0.02$	$5.13 \pm 0.02$	$6.73 \pm 0.05$	$8.97 \pm 0.07$

Shown are the  $\text{FWHM}_{\text{dep}}$  and  $\text{FWHM}_{\text{gen}}$  (see Sect. 6), for 200 000 events in the validation data sets (see Sect. 4) with low and high beam background. The data sets cover an energy range of  $0.1 < E_{\text{gen}} < 1.5 \text{ GeV}$  and the full detector range  $17^\circ < \theta_{\text{gen}} < 150^\circ$  and  $0^\circ < \phi_{\text{gen}} < 360^\circ$ , each of which in uniform distribution. The uncertainties of the FWHM in each column are correlated since they use the same simulated events. The input features are described in detail in Sect. 3

$17^\circ < \theta_{\text{gen}} < 150^\circ$  and  $0^\circ < \phi_{\text{gen}} < 360^\circ$ , each of which in uniform distribution. The FWHM of  $E_{\text{gen}}$  and  $E_{\text{dep}}$  is calculated as described in Sect. 6. All GNNs use the global crystal coordinates, the LM position, and the crystal mass as input features. A comparison of the FWHM for the different additional input features is shown in Table 4. The results show,

that even for the minimal set of input variables, the GNN’s FWHM is smaller than `basf2`’s for both the deposited and the generated energy in both beam background scenarios. Adding local coordinates leads to small improvements and using time information brings significant improvement in the GNN performance. PSD information has almost no effect on

**Table 5**  $\text{FWHM}_{\text{gen}} \times 10^2$  of one photon with photon energy  $E_\gamma^{(1)}$  in dependence of the second photon energy  $E_\gamma^{(2)}$  for low beam background for the full detector (barrel and endcaps combined)

$E_\gamma^{(1)}$ (GeV) ↓	$E_\gamma^{(2)}$ (GeV) →	[0.1, 0.2]	[0.2, 0.5]	[0.5, 1.0]	[1.0, 1.5]
[0.1, 0.2]	GNN	$11.04 \pm 0.79$	$11.98 \pm 0.40$	$11.94 \pm 0.31$	$13.25 \pm 0.34$
	<code>basf2</code>	$12.72 \pm 0.80$	$13.93 \pm 0.55$	$14.32 \pm 0.41$	$15.16 \pm 0.48$
	Improvement	15.2%	16.3%	20.0%	14.4%
[0.2, 0.5]	GNN	$7.38 \pm 0.18$	$7.57 \pm 0.12$	$8.23 \pm 0.09$	$8.38 \pm 0.12$
	<code>basf2</code>	$8.48 \pm 0.22$	$8.30 \pm 0.14$	$8.84 \pm 0.12$	$8.96 \pm 0.12$
	Improvement	14.9%	9.7%	7.5%	7.0%
[0.5, 1.0]	GNN	$5.22 \pm 0.08$	$5.43 \pm 0.05$	$5.69 \pm 0.04$	$5.89 \pm 0.04$
	<code>basf2</code>	$5.58 \pm 0.10$	$5.71 \pm 0.06$	$5.85 \pm 0.05$	$6.17 \pm 0.05$
	Improvement	6.7 %	5.1 %	2.8%	4.9%
[1.0, 1.5]	GNN	$4.24 \pm 0.06$	$4.43 \pm 0.04$	$4.67 \pm 0.03$	$4.77 \pm 0.03$
	<code>basf2</code>	$4.55 \pm 0.07$	$4.58 \pm 0.04$	$4.74 \pm 0.04$	$4.85 \pm 0.04$
	Improvement	7.3%	3.4%	1.4%	1.8%

The uncertainties of the FWHM for the two algorithms are correlated for each energy interval since they use the same simulated events. The improvement over the `basf2` baseline algorithm is stated in percent for each energy interval

**Table 6** FWHM<sub>gen</sub> × 10<sup>2</sup> of one photon with photon energy E<sub>γ</sub><sup>(1)</sup> in dependence of the second photon energy E<sub>γ</sub><sup>(2)</sup> for high beam background for the full detector (barrel and endcaps combined)

E <sub>γ</sub> <sup>(1)</sup> (GeV) ↓	E <sub>γ</sub> <sup>(2)</sup> (GeV) →	[0.1, 0.2]	[0.2, 0.5]	[0.5, 1.0]	[1.0, 1.5]
[0.1, 0.2]	GNN	24.77 ± 0.83	24.10 ± 0.76	24.02 ± 0.60	24.72 ± 0.63
	basf2	33.12 ± 1.08	32.82 ± 1.38	31.28 ± 0.79	32.42 ± 0.88
	Improvement	33.7%	36.2%	30.3%	31.1%
[0.2, 0.5]	GNN	13.16 ± 0.30	13.96 ± 0.20	14.17 ± 0.16	14.17 ± 0.16
	basf2	17.73 ± 0.47	17.56 ± 0.31	17.62 ± 0.24	16.88 ± 0.23
	Improvement	34.8%	25.8%	24.3%	19.1%
[0.5, 1.0]	GNN	8.07 ± 0.12	8.56 ± 0.08	8.71 ± 0.06	8.84 ± 0.06
	basf2	10.53 ± 0.19	10.77 ± 0.12	10.75 ± 0.09	10.73 ± 0.08
	Improvement	30.6%	25.8%	23.4%	21.4%
[1.0, 1.5]	GNN	6.05 ± 0.08	6.33 ± 0.05	6.42 ± 0.04	6.54 ± 0.04
	basf2	7.52 ± 0.12	7.56 ± 0.07	7.60 ± 0.06	7.68 ± 0.06
	Improvement	24.2%	19.6%	18.3%	17.4%

The uncertainties of the FWHM for the two algorithms are correlated for each energy interval since they use the same simulated events. The improvement to the basf2 baseline is stated in percent for each energy interval

the FWHM. Since the main purpose of the PSD information is to differentiate electromagnetic and hadronic interactions per crystal, this is expected. In anticipation of future extensions of the GNN to hadronic interactions as well, the PSD information is kept throughout this work.

### Overlapping Photons

When discussing overlapping photon events, it is important to note that the FWHM of the photon energy distribution not only depends on its own properties but also on the properties of the second photon present. To account for that, the evaluation is split in energy bins of [0.1, 0.2], [0.2, 0.5], [0.5, 1.0], and [1.0, 1.5] GeV for both photons respectively. We report the FWHM of the first photon for different simulated energies of the second photon for low beam backgrounds (see Table 5) and high beam backgrounds (Table 6).

The GNN provides a better FWHM for all combinations, but the improvement is most significant if the photon is low energetic. For low beam backgrounds, the GNN improves the FWHM by up to 20% for photons with simulated energies between 0.1 < E<sub>gen</sub> < 0.2 GeV. For high beam backgrounds, the GNN improves the FWHM by more than 35% for photons with simulated energies between 0.1 < E<sub>gen</sub> < 0.2 GeV.

The result shows that the significant performance improvement observed for isolated photons can also be achieved for the more complicated overlapping photon signatures.

### Conclusion and Outlook

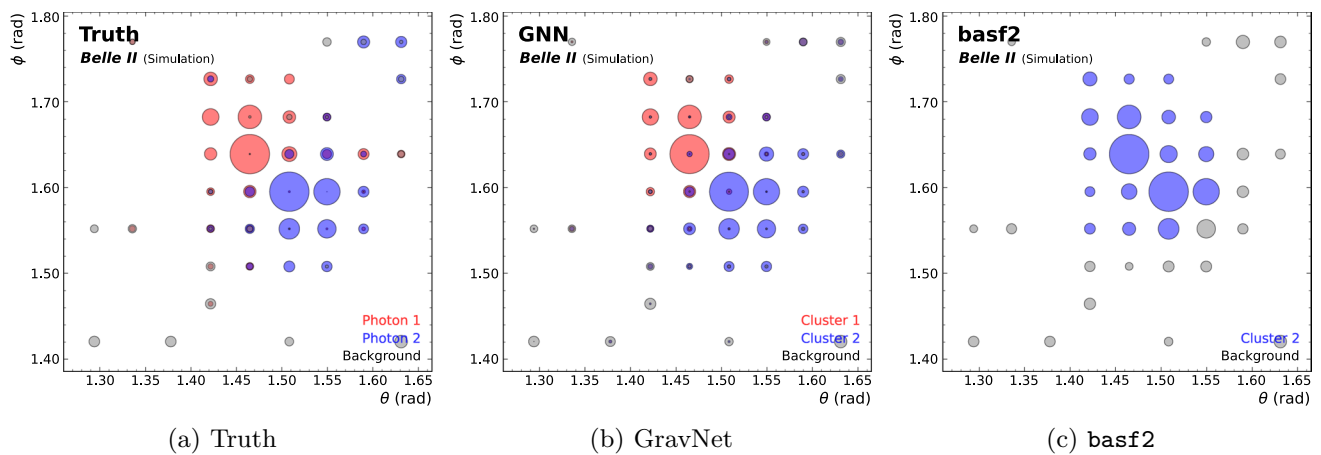
In this work, we have presented a complete study of a GNN-based fuzzy clustering algorithm for the Belle II electromagnetic calorimeter. We have been using a

realistic full detector simulation and simulated beam background for low and high luminosity conditions of Belle II. The GNN algorithm has been compared to the currently used basf2 baseline algorithm. We find a significantly improved resolution of more than 30% for high beam backgrounds, but also improved performance in reducing the right-side tails of the reconstruction errors that are caused by beam background. Such significant improvements in photon reconstruction performance directly improve the physics reach of Belle II for almost all final states with photons, but also analyses that use missing energy information [21]. We also trained different GNNs to separate energy depositions of overlapping photon clusters. The improvement of the energy resolution is up to 30% for the low energy photon in asymmetric photon pairs. Any improvement in overlapping photon reconstruction has direct implications for the reconstruction of boosted π<sup>0</sup> mesons or axion-like particles with couplings to photons [36].

While the basf2 algorithm strictly reconstructs one cluster for each LM, the GNN algorithm only uses the LMs to center the ROI. The GNN algorithm can therefore in principle also be used to reconstruct overlapping photons that only produced one LM (Fig. 11). The extension of the GNN algorithm to such overlapping signatures as well as to charged particles and neutral hadrons will be the focus of follow-up work. Future work is also going to address robustness against varying beam backgrounds explicitly, for example by introducing features that are directly sensitive to beam-background levels.

This is the first application of a GNN-based clustering algorithm at Belle II for a realistic detector geometry and realistic and high beam backgrounds. This is also the first time that an algorithm has shown to improve the





**Fig. 11** Comparison of truth energy fractions (a), the reconstructed energy fraction by the GNN (b), and the reconstructed energy fraction by *basf2* (c) for one example event with only one local maxi-

mum. Colors indicate the fractions belonging to each photon or background. The marker centers indicate the crystal centers, the marker area is proportional to the reconstructed energy in each crystal

performance of the photon reconstruction by explicitly including timing information on clustering level at Belle II.

**Acknowledgements** The authors would like to thank the Belle II collaboration for useful discussions and suggestions on how to improve this work. The authors would like to thank Jan Kieseler for helpful discussions. The training of the models was performed on the TOPAS GPU cluster at the Steinbuch Centre for Computing (SCC) at KIT. This work is funded by Helmholtz (HGF) Young Investigators Group VH-NG-1303 and BMBF ErUM-Pro 05H23VKKBA. I. Haide is supported by the Landesgraduiertenförderung Baden-Württemberg.

**Funding** Open Access funding enabled and organized by Projekt DEAL.

**Data Availability** The datasets generated during and analysed during the current study are property of the Belle II collaboration and not publicly available. The instructions and code to replicate the studies in this paper are available at [37, 38].

## Declarations

**Conflict of Interest** The authors declare that they have no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Natochii A et al (2022) Beam background expectations for Belle II at SuperKEKB. [arxiv:2203.05731](https://arxiv.org/abs/2203.05731)
2. Dunn JC (1973) A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. *J Cybern* 3(3):32–57
3. Canudas NV et al (2023) Graph clustering: a graph-based clustering algorithm for the electromagnetic calorimeter in LHCb. *Eur Phys J C* 83:02
4. Valsecchi D (2023) Deep learning techniques for energy clustering in the CMS ECAL. *J Phys: Conf Ser* 2438(1):012077
5. Simkina P (2022) Machine learning techniques for calorimetry. *Instruments* 6:47
6. Belayneh DT et al (2019) Calorimetry with deep learning: particle simulation and reconstruction for collider physics. *Eur Phys J C* 80
7. Boldyrev A, Chekalina V, Ratnikov F (2020) Machine learning approach to boosting neutral particles identification in the LHCb calorimeter. *J Phys: Conf Ser* 1525(1):012096
8. Charan AN (2023) Particle identification with the Belle II calorimeter using machine learning. *J Phys: Conf Ser* 2438(1):012111 [arxiv:2301.11654](https://arxiv.org/abs/2301.11654)
9. Paganini M et al (2018) CaloGAN: simulating 3D high energy particle showers in multilayer electromagnetic calorimeters with generative adversarial networks. *Phys Rev D* 97(1):014021
10. Buhmann E et al (2021) Getting high: high fidelity simulation of high granularity calorimeters with high speed. *Comput Softw Big Sci* 5(1):13
11. Deep generative models for fast photon shower simulation in ATLAS. 10 (2022) [arxiv:2210.06204](https://arxiv.org/abs/2210.06204)
12. Bhattacharya S et al (2023) GNN-based end-to-end reconstruction in the CMS phase 2 high-granularity calorimeter. *J Phys: Conf Ser* 2438:012090
13. Grasseau G et al (2020) A deep neural network method for analyzing the CMS high granularity calorimeter (HGCAL) events. *EPJ Web Conf* 245:02003

14. Novosel A et al (2023) Identification of light leptons and pions in the electromagnetic calorimeter of Belle II. In 11th International Workshop on Ring Imaging Cherenkov Detectors , 01. [arxiv:2301.05074](https://arxiv.org/abs/2301.05074)
15. Shlomi J, Battaglia P, Vlimant J-R (2021) Graph neural networks in particle physics. *Mach Learn Sci Technol* 2(2):021001
16. Duarte Javier, Vlimant Jean-Roch (2020) Graph neural networks for particle tracking and reconstruction, 12. [arxiv:2012.01249](https://arxiv.org/abs/2012.01249)
17. DeZoort Gage, Battaglia Peter W, Biscarat Catherine, Vlimant Jean-Roch (2023) Graph neural networks at the Large Hadron Collider. *Nat Rev Phys* 5(5):281–303
18. Wang Y et al (2019) Dynamic graph CNN for learning on point clouds. *ACM Trans Graph* 38(5):10
19. Qasim SR et al (2019) Learning representations of irregular particle-detector geometry with distance-weighted graph networks. *Eur Phys J C* 79(7):608 [arxiv:1902.07987](https://arxiv.org/abs/1902.07987)
20. HEP ML Community. A living review of machine learning for particle physics. <https://iml-wg.github.io/HEPML-LivingReview/>
21. Kou E et al (2019) The Belle II Physics Book. *PTEP*, 2019(12):123 C01. [arxiv:1808.10567](https://arxiv.org/abs/1808.10567)
22. Abe T et al (2010) Belle II Technical Design Report. Technical report, Belle-II, 11. [arxiv:1011.0352](https://arxiv.org/abs/1011.0352)
23. Ikeda H (1999) Development of the CsI(Tl) calorimeter for the measurement of CP violation at KEK B-Factory. PhD thesis, Nara Women's University
24. Aulchenko V et al (2017) Time and energy reconstruction at the electromagnetic calorimeter of the Belle II detector. *J Instrum* 12(08):C08001–C08001
25. Longo S et al (2020) CsI(Tl) pulse shape discrimination with the Belle II electromagnetic calorimeter as a novel method to improve particle identification at electron-positron colliders. *Nucl Instrum Meth A* 982:164562
26. Agostinelli S et al (2003) GEANT4: a simulation toolkit. *Nucl Instrum Meth A* 506:250–303
27. Kuhr T et al (2019) The Belle II Core Software. *Comput Softw Big Sci.* 3(1)
28. Belle II Collaboration. Belle II Analysis Software Framework (basf2). <https://doi.org/10.5281/zenodo.5574115>
29. Liptak ZJ et al (2022) Measurements of beam backgrounds in SuperKEKB Phase 2. *Nucl Instrum Meth A* 1040:167168 [arxiv:2112.14537](https://arxiv.org/abs/2112.14537)
30. Fey M, Lenssen JE (2019) fast graph representation learning with PyTorch geometric. In ICLR workshop on representation learning on graphs and manifolds
31. Clevert D-A et al (2016) Fast and accurate deep network learning by exponential linear units (ELUs). In 4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2–4, 2016, Conference Track Proceedings
32. Ioffe S (2015) Batch normalization: accelerating deep network training by reducing internal covariate shift. In Francis Bach and David Blei, editors, Proceedings of the 32nd International Conference on Machine Learning, vol. 37, pp 448–456, Lille, France, 07. PMLR
33. Akiba T et al (2019) Optuna: a next-generation hyperparameter optimization framework. In Proceedings of the 25rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining
34. Gaiser J (1982) Charmonium spectroscopy from radiative decays of the  $J/\psi$  and  $\psi'$ . PhD thesis, Stanford University
35. Skwarnicki T (1986) A study of the radiative cascade transitions between the Upsilon-Prime and Upsilon resonances. PhD thesis, Cracow, INP
36. Abudínén F et al (2020) Search for axion-like particles produced in  $e^+e^-$  collisions at Belle II. *Phys Rev Lett* 125(16)
37. Wemmer F et al (2023) Photon reconstruction in the Belle II calorimeter using graph neural networks. [https://github.com/JonasEppelt/gnn\\_photon\\_clustering\\_in\\_belleII\\_ecl](https://github.com/JonasEppelt/gnn_photon_clustering_in_belleII_ecl)
38. Wemmer F et al Photon reconstruction in the Belle II calorimeter using graph neural networks. <https://zenodo.org/record/8409638>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.