Capability of MLEM and OE to Detect Range Shifts With a Compton Camera in Particle Therapy

Nadja Kohlhase, Tilman Wegener, Moritz Schaar, Andreas Bolke, Ane Etxebeste, David Sarrut, Magdalena Rafecas

Capability of MLEM and OE to Detect Range Shifts With a Compton Camera in Particle Therapy

Nadja Kohlhase, *Student Member, IEEE*, Tilman Wegener, Moritz Schaar, *Student Member, IEEE*, Andreas Bolke, Ane Etxebeste, David Sarrut, and Magdalena Rafecas, *Member, IEEE*

***Abstract*—To identify range deviations by using Compton cameras (CCs), tomographic image reconstruction of CC data is needed. Within this context, image reconstruction is usually performed using maximum likelihood expectation maximization (MLEM), and more recently, the origin ensemble (OE) algorithm. In this article, we investigate how MLEM and OE affect the precision and accuracy of estimated range deviations. In particular, we focus on the effects of data selection, statistical fluctuations, and artifact reduction. The use of external information of the beam path through a hodoscope was also explored. Additionally, two methods to calculate range deviations were tested. To this aim, realistic proton beams were simulated using GATE and data from single spots as well as from seven contiguous spots of an energy layer were reconstructed. MLEM and OE reacted differently to the poor data statistics. In general, both algorithms were able to detect range shifts for single spots, particularly when multiple coincidences were also considered. Selection of events corresponding to the most relevant energy peaks decreased the identification performance due to the lower statistics. When data from several contiguous spots were jointly reconstructed, the accuracy of the results degraded significantly, and nonzero shifts were assigned when no shifts had occurred. The limited size of the cameras and the subsequent restriction in the orientation and aperture of detected cones, as well as in the number of detected events are major challenges. Future efforts should be devoted to noise regularization and compensation for data truncation.**

***Index Terms*—Compton camera (CC), image reconstruction, particle therapy, prompt gamma imaging, range verification.**

I. Introduction

In RECENT years, Compton cameras (CCs) have attracted attention for range verification in particle therapy [1]–[6]. Within this context, the goal of the CC is to detect prompt gamma-rays (PG) that are emitted during de-excitation of nuclei along the particle path within the body. Tomographic reconstruction techniques estimate the spatial distribution of the PG origins. The rationale is that the emission origins of the PG are indirectly correlated with the range of the beam in the patient. Hence, the difference between the theoretical and the actual beam range can be inferred by comparing the reconstructed image with a reference.

CCs usually consist of several detectors in coincidence, aimed to register the PG through one or more Compton interactions in the so-called *scatterer*, and photoelectric absorption of the scattered photon in the *absorber*. In conventional CC imaging, it is assumed that the scattered photon is absorbed through photoelectric effect after a first interaction in the scatterer. In this case, the initial photon energy, E_0, is simply equal to $E_1 + E_2$, being E_1 and E_2 the first and second interactions energy, respectively. Using Compton kinematics, the unknown emission point of the PG is constrained to the surface of a cone, whose vertex is located in the first Compton interaction. The half-angle of the cone, θ, corresponds to the Compton scattering angle

$$\cos(\theta) = 1 - m_e c^2 \cdot \frac{E_1}{E_0(E_0 - E_1)} \tag{1}$$

and the cone axis is the line connecting the two interaction points. The collection of cones calculated from the measured data is used as input for the tomographic reconstruction of the origins. However, in prompt gamma-ray imaging (PGI) the broad energy spectrum of the PG, up to several MeV [7] poses some challenges to the accurate calculation of θ. As E_0 is unknown, (1) can only provide the correct scattering angle when the photons are fully absorbed. Otherwise, a wrong cone is assigned to the event, contributing to blur the reconstructed image. This problem can be overcome when photons interact at least three times in the detectors [3]. In that case, the initial energy of the photon is calculated as

$$E_0 = E_1 + \frac{1}{2} \cdot \left(E_2 + \sqrt{E_2^2 + \frac{4E_2 m_e c^2}{1 - \cos(\theta_2)}} \right) \tag{2}$$

where θ_2 is the scattering angle of the second Compton interaction [8].

For CC, tomographic image reconstruction is mainly performed using the maximum likelihood expectation maximization (MLEM) algorithm for list-mode (LM) data [9]. The origin ensemble (OE) algorithm [10], [11] has also

Manuscript received February 26, 2019; revised June 30, 2019; accepted August 5, 2019. Date of publication August 27, 2019; date of current version March 3, 2020. This work was supported by the *Deutsche Forschungsgemeinschaft* through Research under Grant RA 2830/1-1. The capacities for computing the simulations were provided by the North-German Supercomputing Alliance (HLRN) under Project shp00016. *(Corresponding author: Magdalena Rafecas.)*

N. Kohlhase, T. Wegener, M. Schaar, A. Bolke, and M. Rafecas are with the Institute of Medical Engineering, Universität zu Lübeck, 23538 Lübeck, Germany (e-mail: rafecas@imt.uni-luebeck.de).

A. Etxebeste and D. Sarrut are with CREATIS, CNRS UMR5220, Inserm U1044, INSA-Lyon, Université Lyon 1, 69100 Villeurbanne, France.

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TRPMS.2019.2937675

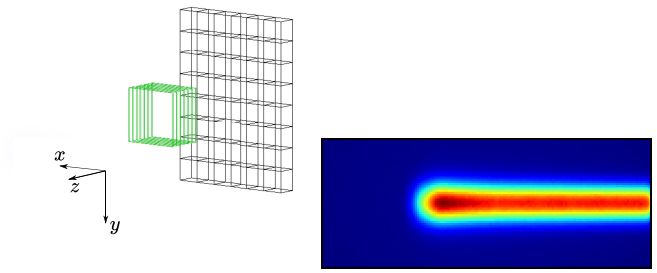

Fig. 1. Left: CC with seven scatterers (green) and the absorber (black). Right: PG distribution of a single proton beam of 120 MeV.

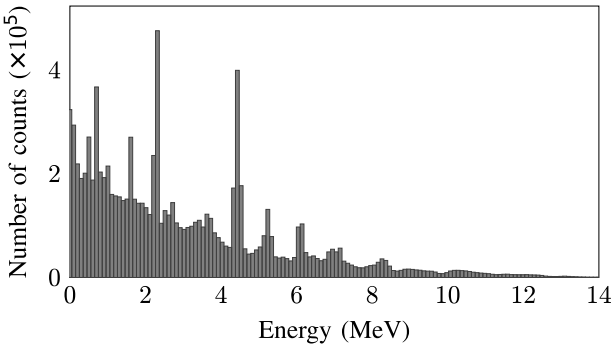

Fig. 2. Energy distribution of PG for a 120 MeV proton beam hitting a water phantom. After normalization, this spectrum was used as input for the simulation of the proton beams.

attracted attention for CC [12]. In particle therapy, image reconstruction of CC data faces several challenges, such as very low counts, incomplete sampling, and unknown initial photon energy. The limited spatial, time, and energy resolution of the detectors further impair the quality of the reconstructed images, which might become noisy, blurred, and even distorted by truncation artifacts. To cope with some of these effects, modified versions of OE have been proposed (e.g., [6] and [13]–[15]). Efforts have been also put in improving the quality of LM-MLEM images or its accelerated version Ordered Subset Expectation Maximization (OSEM) within the context of PGI, e.g., by better modeling the system response [16]–[21]. Data selection using geometrical or spectral considerations has also been considered for both LM-MLEM and modified OE approaches and promising results have been obtained [6], [22].

To the best our knowledge, a comparison between MLEM and OE within the context of range verification in particle therapy using CC data has not been done yet. In this article, we investigate the capability of MLEM and OE to detect range shifts with a CC. One of our goals is to identify their weak points for the proposed applications in order to guide future developments. To this aim, we have simulated a CC based on the device being developed by the French collaboration CLaRyS [5]. This article partly focuses on the effects of data selection techniques and low-count measurements. Data selection and *a-priori* information helps to improve the quality of the data, but at the cost of decreasing the number of events used in the reconstruction. On the other hand, low-count measurements might lead to biased images, even if the information content of the data is accurate. As the goal of PGI is range verification, the capability of the algorithms to detect range shifts was quantified. For this purpose, data corresponding to realistic beams were reconstructed and two methods for range estimation were applied. Attention was also put on their ability to correctly reveal that no shifts occurred.

II. Methods

We have performed Monte Carlo simulations of PGI using realistic proton beams. To simulate interfractional changes different sources were used. Data were selected for reconstruction using various criteria. Afterward, two inherently different reconstruction techniques were applied to generate raw images that were further used as input for two range verification methods.

A. Monte Carlo Simulations

GATE 8.0 [23] was used with a new CC module[1] that is currently under development at CREATIS [24]. This module includes a coincidence sorter, so that the output of the simulation are coincidence events. The CC was simulated as shown in Fig. 1, left. It consists of seven silicon scatterers with $90 \times 90 \times 2$ mm^3 in size and a 1-cm distance between their centers. The bismuth germanate (BGO) absorber is $280 \times 210 \times 30$ mm^3 in size and placed 15 cm apart from the last scatterer (center-to-center distance). The first scatterer of the CC was located at a distance of 10 cm from the proton beams. Perfect spatial and energy resolution were simulated to better identify image degradation effects caused by data fluctuations and possible data truncation. Only Doppler broadening was included, as this is an intrinsic characteristic of the detector material.

This article focuses on identification of interfractional changes of the range using pencil beam spot scanning; for that reason, we simulated pencil proton beams. To reduce simulation time, the process of proton irradiation, prompt gamma-ray generation, and subsequent detection in the CC was split in two simulation steps. First, the spatial and energy distribution of the PG production was determined by irradiating a rectangular water phantom ($10 \times 30 \times 10$ cm^3) by a single proton beam using the pencil beam source available in GATE [25]. The beam spot parameters were chosen following [26], with a medium-spot size of $\sigma_{\text{spot}} = 5$ mm. To change the range, the following energies were considered: 120, 121, 122, and 123 MeV. The expected range shifts can be found in Table I. The resulting spatial and energy distribution of the PG production for an exemplary 120 MeV proton beam are shown in Fig. 1 (right) and Fig. 2, respectively.

In the next step, direct emission of gamma-rays was simulated based on the spatial and energy distributions obtained in the first step. The number of simulated PGs per spot was $3.5 \cdot 10^7$, which corresponds to the average PG emission produced in a GATE simulation of 10^8 protons (i.e., the number of protons for one spot in a 2 Gy irradiation in pencil-beam scanning mode [7]).

[1]The module will be available in a future GATE release.

TABLE I
RANGE SHIFT BETWEEN DIFFERENT PROTON ENERGIES AND A 120 MEV REFERENCE BEAM IN WATER [27]

Proton energy (in MeV)	Penetration depth in water (in mm)	Expected range shift (in mm)
120	106.6	0.0
121	108.2	1.6
122	109.8	3.2
123	111.4	4.8

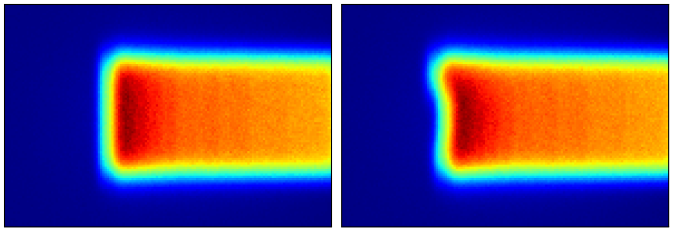

Fig. 3. PG distribution of a simulated phantom with seven proton spots, view parallel to the detector surface (*xy*-plane). Left: energy-layer phantom (reference), with all spots with 120 MeV proton energy. Right: extended phantom; (from bottom to top) two spots with 121 MeV, three spots with 120 MeV and two spots with 123 MeV.

Although the final goal of CCs is to reconstruct the PG activity from single spots, this might be a challenging task because several single spots are usually irradiated within a very short time and the number of detected coincidences per spot is very limited. On the other hand, the joint reconstruction of detected PG data corresponding to several spots might suffer from other degradation effects, due to the overlapping of blurred activity distributions. To study the feasibility of reconstructing data from several spots, we simulated an *energy-layer* phantom, i.e., the extended activity distribution obtained from seven contiguous spots (with an interspot distance of $1.5\sigma_{\text{spot}}$ [26]), for a beam energy of 120 MeV. We also simulated the *extended* phantom to evaluate the effects of range shifts. To this end, the energy-layer phantom was modified as follows: the energy of two proton beams was increased to 121 MeV, and to 123 MeV for two more beams. The PG activity distributions of the two phantoms is shown in Fig. 3. For each phantom, $7 \times 3.5 \cdot 10^7$ gamma-rays were simulated. Such extended activity sources can be of interest to investigate the performance of CC for treatments based on spread-out Bragg peaks (SOBP).

To account how statistical fluctuations might affect the range estimation, five independent simulations were performed for each single beam as well as for each phantom.

B. Reconstruction Algorithms

The algorithms were implemented in C++. A field of view (FOV) of size $40 \times 120 \times 40$ mm^3 was reconstructed for the single beams. For the energy-layer and extended phantoms, the FOV was enlarged to cover $80 \times 120 \times 40$ mm^3. A voxel length of 1 mm was chosen for all cases, as smaller voxels led to poorer image quality, especially for OE.

1) System Response Model: Both MLEM and OE require to calculate the elements t_{ij} of the system matrix. For LM data, t_{ij} describes the probability that the ith event of the list originated in voxel j. In this article, t_{ij} was calculated following [18] and [28] as:

$$t_{ij} = \frac{1}{\sqrt{2\pi}} K\left(\theta_{ij}|E_0\right) \frac{\left|\cos\left(\theta_{ij}\right)\right|}{l_{ij}^2} e^{-\frac{1}{2}d_{ij}^2} \tag{3}$$

where θ_{ij} is the angle between the vector connecting the center of voxel j to the location of the first interaction and the normal of the detector; $K(\theta_{ij}|E_0)$ is the Klein–Nishina differential cross section for the calculated initial energy; l_{ij} is the distance between voxel j and the first interaction in the detector; and d_{ij} is the normalized angular distance between the surface of cone i and the center of voxel j. The Gaussian was truncated to include only those voxels intersected the cone surface. The elements of the sensitivity were computed according to [18] as

$$s_j = \sum_{m=1}^{M} \frac{\left|\cos\left(\theta_{mj}\right)\right|}{l_{mj}^2} \tag{4}$$

where l_{mj} is the distance between the centers of voxel j and the mth detector element, for all detector elements M.

2) LM-MLEM: The iterative formula

$$\lambda_j^{(n+1)} = \frac{\lambda_j^{(n)}}{s_j} \sum_{i=1}^{I} \frac{t_{ij}}{\sum_{k=1}^{J} t_{ik}\lambda_k^{(n)}} \tag{5}$$

was implemented in a parallelized approach, where j corresponds to the jth voxel from a total of J image elements, and i refers to the ith event in the coincidence list, and I is the number of events. The image vector $\boldsymbol{\lambda}^{(n)}$ contains the spatial distribution of the PG, estimated at iteration n.

The iteration process was stopped when the maximal voxel-wise intensity difference between two iterations was below 1 %, or when the number of iterations was larger than 80 to avoid long reconstruction times.

3) OE: The OE method is based on the Metropolis–Hasting algorithm [29], which is employed to estimate the expected value of the PG emission distribution conditioned on the measured data. To this end, the algorithm generates a sequence of samples of the emission distribution (*states*) following a Markov chain. To create the initial state, an emission origin is randomly assigned to each detected event. In our implementation, the origin of an event was generated following an uniform random sampling on the surface of the corresponding cone [12]. Next, for each event i a new origin is proposed, and this event is relocated to the new origin according to a certain acceptance probability.

The acceptance probability from state X to state X' was calculated as

$$A\left(X \to X'\right) = \min\left(1, \frac{t_{ij'}\ s_j\ \left(x_j - 1\right)^{x_j - 1}\ \left(x_{j'} + 1\right)^{x_{j'}+1}}{t_{ij}\ s_{j'}\ \left(x_j\right)^{x_j}\ \left(x_{j'}\right)^{x_{j'}}}\right) \tag{6}$$

where x_j denotes the number of counts in the current voxel j and $x_{j'}$ is the number of counts in the voxel which is proposed as new origin. Equation (6) describes the full acceptance probability as originally proposed in [10]. For high-count data, this formula can be simplified as pointed out in [11]; given the low number of emissions per voxel characterizing this article, the full formula was used. In any case, both the full and the simplified formulas introduced by Sitek take into account

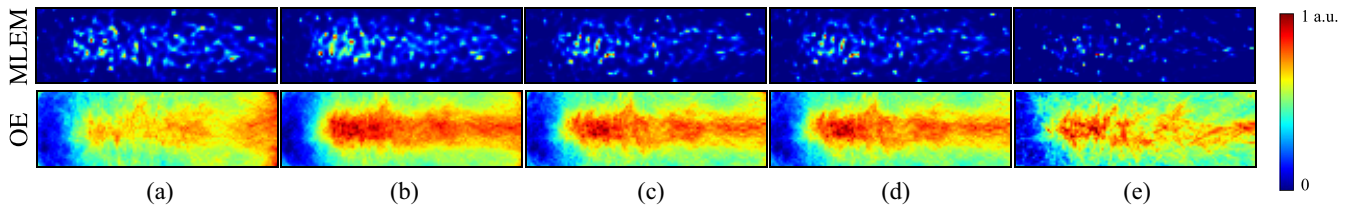

Fig. 4. Views (xy-plane) of the reconstructed images of a 120 MeV proton beam, summed along z-direction for MLEM (top row) and OE (bottom row). (a) Reconstruction of only two interaction coincidences, (b) two interaction coincidences plus multiple interaction coincidences, (c) like (b) plus an energy selection of $1-14$ MeV, (d) like (c) plus an event selection through considering a hodoscope, so that the beam path in yz-direction is known, and (e) selection of specific energy peaks. All images are normalized to their maximum.

the system response model through the terms t_{ij}, $t_{ij'}$, s_j, and $s_{j'}$. These terms are often assumed to be 1 in most works using OE or its modified versions for CC data (see [6], [12], [13], [15]). However, our preliminary studies shown that the inclusion of the system response improves image quality. For comparability, we use the same system matrix and sensitivity as for the LM-MLEM algorithm.

The pseudorandom number generator used was *std::mt19937*, which is based on the Mersenne-Twister algorithm. As the uniform sampling on the cone surface already takes into account the intersection of the cone surface and the related image voxel, the Gaussian term in (3) was not used. One iteration was completed when each event was proposed a new origin, which might have been accepted or not. The entropy of the image and the number of shifted events per iteration were used to inspect when the Markov chain reached the equilibrium. We calculated the final image from the average over 9000 states divided by the sensitivity, as described in [10]. The algorithm was implemented using the Armadillo linear algebra library [30].

C. Data Selection

Only those data whose cones intersected the FOV were considered. Different criteria were further used for event selection.

(a) *Classical Scenario:* We considered only coincidences composed of two interactions, the first one in the scatterer and the second one in the absorber. The cone half-angle was calculated assuming that E_0 was equal to $E_1 + E_2$. In PGI, this approach may lead to wrong cone apertures as E_0 is unknown and high-energy photons often scatter in the absorber.

(b) *Multiple Interaction:* In addition to the classic two-interaction coincidences, events made of three or more interactions in coincidence were also considered. Although recording multiple coincidences may require complex detector electronics, these events are useful to increase data statistics and the accuracy of the calculated cones. We examined further selection criteria to mitigate the artifacts caused by wrongly calculated cones, mainly because of wrongly assigned E_0.

(c) *Energy Cut:* Events with a calculated energy $E_0 <$ 1 MeV were discarded, as below 1 MeV Doppler broadening strongly deteriorates the angular resolution [31].

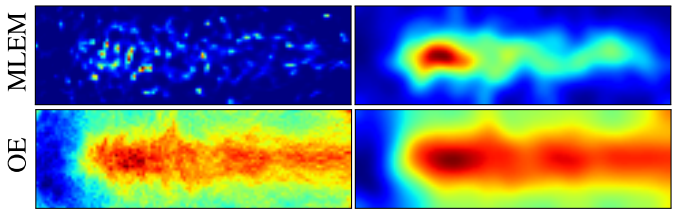

Fig. 5. Single 120 MeV proton beam reconstructed with MLEM (top) and OE (bottom) for scenario (c). On the left the raw images are shown and on the right the filtered images. All images are normalized to their maximum.

(d) *Hodoscope:* Together with criterion (c), *a-priori* information as the one obtainable from a hodoscope was used. A hodoscope measures the beam position and its spread in the yz-plane. This motivated us to select as valid events only those whose cones intersect a region-of-interest (ROI) defined by $x \in [-15\text{ mm}, 15\text{ mm}]$ and $z \in [-15\text{ mm}, 15\text{ mm}]$, for the single beam spots; and $x \in [-30\text{ mm}, 30\text{ mm}]$ and $z \in [-15\text{ mm}, 15\text{ mm}]$ for the two phantoms. Stronger constraints using *a-priori* information about the beam direction have been proposed in literature, e.g., [32] and [33]. In our case, we preferred not to assume that the beam is described by a perfect line given the non-negligible extension of the simulated beam and the resulting activity distribution.

(e) *Energy Peaks:* A common approach in PGI is to select events whose total deposited energy corresponds to certain transitions of the excited nuclei within the patient, reflected as peaks in the spectrum [28], [34]. Here, we selected the most prominent energy peaks that are seen in Fig. 2, i.e., 2.3, 4.4, 5.2, and 6.1 MeV.

D. Range Verification

Two approaches were implemented, both partly based on methods originally proposed for PET-based range verification.

1) Activity Depth Profile: This method relies on the analysis of distal profiles calculated from the reconstructed activity distributions [35]. In this article, the central slices of the 3-D image parallel to the CC detectors were summed up in z-direction. The range shift in the line phantoms was determined after image post-processing with a $3 \times 3 \times 3$ mm^3 Gaussian filter, and extraction of rectangular ROIs for each line. For each profile, reference points were calculated by averaging the depths for 25% and 50% maximum activity. Finally, the line profiles for the single proton beams

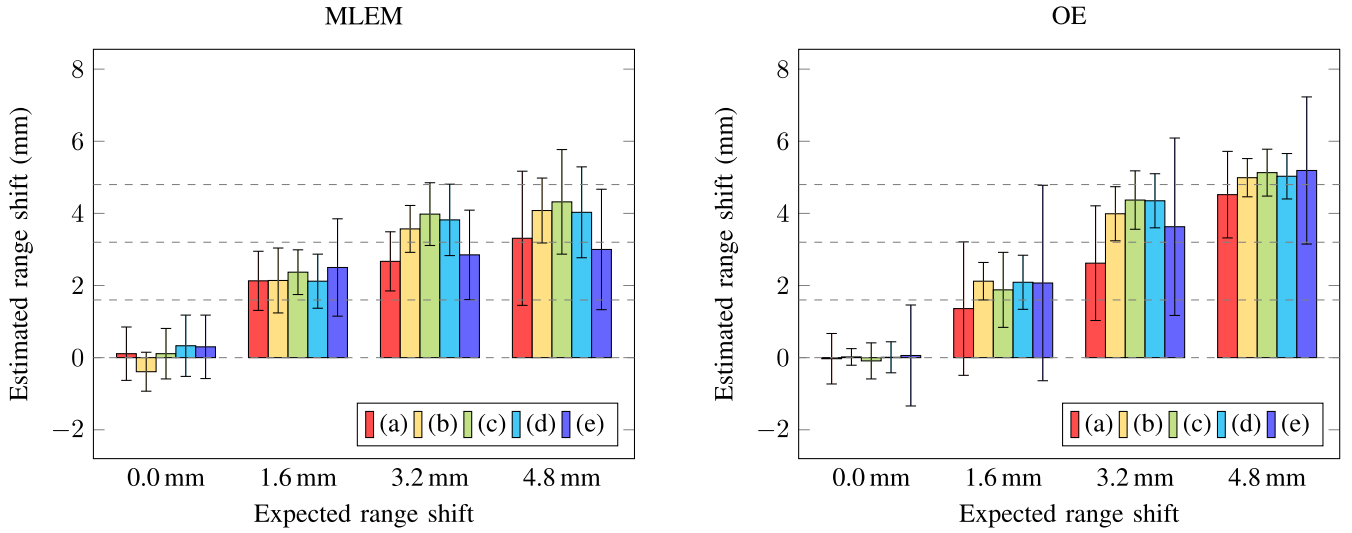

Fig. 6. Estimated range shift and standard deviation for the single beam spots for scenario (a) - (e) reconstructed with MLEM (left) and OE (right). Five independent simulations and reconstructions were performed. The expected range shift (marked as dashed lines) are (1) 0.0 mm (120 versus 120 MeV), (2) 1.6 mm (120 versus 121 MeV), (3) 3.2 mm (120 versus 122 MeV), and (4) 4.8 mm (120 versus 123 MeV).

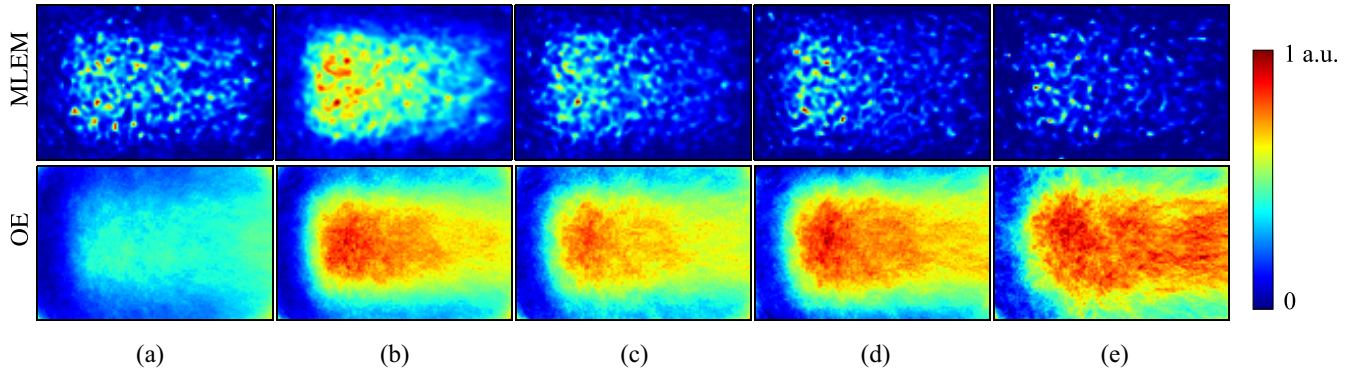

Fig. 7. Views (*xy*-plane) of the reconstructed images, summed along *z*-direction for MLEM (top row) and OE (bottom row). (a) Reconstruction of just two interaction coincidences, (b) two interaction coincidences plus multiple interaction coincidences, (c) like (b) plus an energy selection of $1 - 14$ MeV, (d) like (c) plus an event selection through considering a hodoscope, so that the beam path in *yz*-direction is known, and (e) selection of specific energy peaks. All images are normalized to their maximum.

and the two phantoms were compared to determine possible range shifts.

In spite of the filtering, this method can be very sensitive to statistical fluctuations, as local maxima might be wrongly used as reference to localize the distal falloff. Therefore, a second technique was implemented to evaluate range shifts between the energy-layer phantom and its extension.

2) Activity Contour: Possible range shifts were evaluated using the activity contours extracted from the reconstructed images. Compared to the former method, this technique is more appropriate for complex treatment plans and can accommodate data obtained from various fields. For this article, the method proposed in [36] was slightly modified to account for truncation artifacts in *z*-direction. First, the image was filtered with a $3 \times 3 \times 3$ mm^3 Gaussian filter. Next, the images were summed up in *z*-direction. Further filtering and contour extraction was performed as described in [36]. The contours from two images were then compared voxel-wise within the Bragg-peak region. To identify possible shifts, the Hausdorff distance d_H [37] was used; d_H is defined as the maximum value of all minimal distances found between the points of both contours.

For all methods and scenarios, mean values and standard deviation were computed using the various realizations.

III. Results and Discussion

In this section, the unprocessed raw images are analyzed and the resulting range shifts are quantified.

A. Single Proton Spots

In the following, the images are qualitatively evaluated, and the quantitative results for the estimated range shifts using the processed images are presented.

1) Image Quality: In Fig. 4, MLEM and OE reconstructed images for a single realization corresponding to a 120 MeV proton beam are shown. The images were obtained after summing along the *z*-direction, and show the view parallel to the camera. This view is the one used to identify range shifts. The average number of reconstructed coincidences varied with

TABLE II
MEAN NUMBER OF RECONSTRUCTED COINCIDENCES N (N IN $\times 10^4$ EVENTS) FOR THE SINGLE PROTON BEAMS (SPB) AND THE ENERGY-LAYER AND EXTENDED (ELP) PHANTOMS

Phantom \ Criterion	(a)	(b)	(c)	(d)	(e)
SPB	0.8	2.2	1.1	1.0	0.2
ELP	5.8	16.5	8.3	8.1	1.8

the level of selection (see Table II). Including multiple interactions significantly increased the number of valid events, even if this type of coincidences are less frequent than two-interaction events. This is probably due to two facts. On the one hand, several two-interaction events cannot provide the correct cone aperture due to incomplete absorption of the scattered photon. Part of these events are thus rejected since the assigned cones do not intersect the FOV. On the other hand, most multiple-interaction events allow for a better estimation of the cone aperture [see (2)], so the ratio of accepted events is much larger for multiple than for double coincidences. Furthermore, the seven scatterers favour multiple scattering when compared to other CC configurations with fewer scatterers.

The reconstructed activity distributions are either highly blurred (OE) or extremely noisy (MLEM) for all scenarios. Still, the distal edge is recognizable and correlates well with the ground truth (depicted in Fig. 1, right). This can be seen best in the filtered images (see Fig. 5). The appearance of raw MLEM and OE images highly differ, also after smoothing. MLEM images are characterized by the typical "night sky" noise which characterizes this algorithm when run until convergence without any regularization. However, useful information is recovered after smoothing with the Gaussian filter.

OE images are blurrier that MLEM ones. The differentiated behavior between the two algorithms possibly results from their very different nature, and how both cope with the limited information provided by a CC for the given FOV and low-statistic data. The MLEM solution is the image which maximizes the likelihood of the data, whereas the OE algorithm provides samples of the posterior probability of the number of emissions per voxel, conditioned on the measured data. In the current scenario, MLEM produces very noisy images to fit the statistically noisy data; given so few events in relation to the number of image voxels, OE yields a sequence of states which can much differ from each other, in spite of the similar value of the corresponding posterior.[2] Consequently, their average is a blurred image.

2) Range Verification: For the purpose of range verification, image quality is secondary as long as range shifts are calculated correctly. Fig. 6 shows the estimated range shifts with the *Activity Depth Profile* method for MLEM (left) and OE (right). Both reconstruction methods allowed for a correct identification of zero shifts with a standard deviation σ smaller than 0.9 mm for almost all scenarios, excepting scenario *Energy Peaks* in OE with $\sigma = 1.4$ mm. The true values of the nonzero range shifts lied within the error intervals, whereby scenario *Classical* and *Energy Peaks* have the highest standard deviation. The estimated range shifts for these scenarios also deviate the most from the expected range shift. The results for the *Classical* scenario are affected by the high level of noise introduced by the mispositioning of nonabsorbed photons. In the case of scenario *Energy Peaks*, the deviation is probably caused by the low number of events available for reconstruction despite the high degree of data selection that should minimize the noise. For the 1.6 mm and 3.2 mm range shifts, MLEM and OE perform similar. Both methods tend to slightly overestimate the small and the middle range shifts. This can be partially due to the voxel size and partial volume effects. The largest range shift is better estimated by OE, whereas MLEM underestimates it with a larger statistical uncertainty.

[2]In our imaging scenario, a single CC event can originate from a relatively large number of image voxels, i.e., the intersection between the cone and the 3-D FOV. As few events are available, very similar values of the posterior can be obtained from states of the Markov chain which correspond to very different distributions. Since the final image is calculated after averaging over all these states, the large variability between the states of the chain translates into the spread of the reconstructed activity over the FOV.

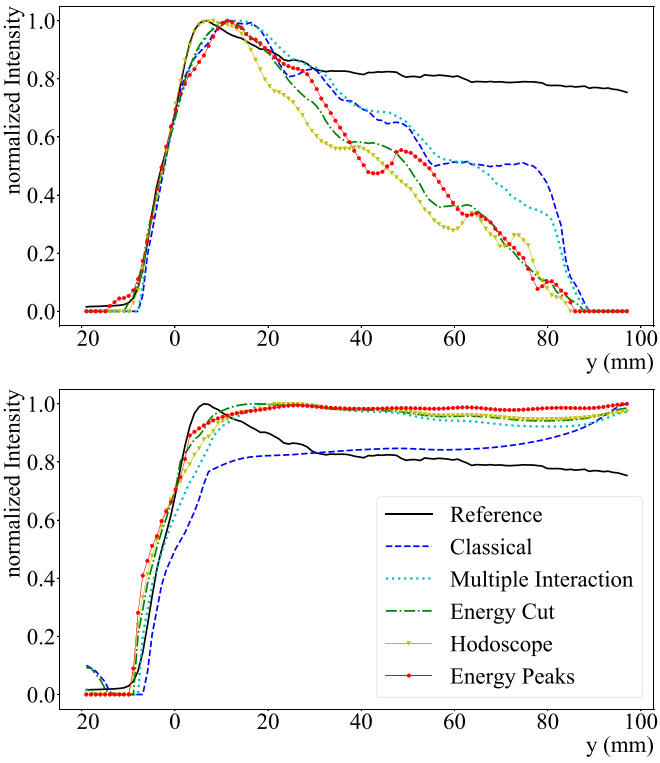

Fig. 8. Profiles along y-axis for the energy-layer phantom, reconstructed with MLEM (top) and OE (bottom) for scenario (a)–(e). All images are post-processed and normalized to their maximum.

B. Energy-Layer and Extended Phantoms

In this section, the interplay between the activity distribution originating from several spots within the same reconstructed layer is evaluated in the case of range shifts affecting some of the spots.

1) Image Quality: Raw reconstructed images with MLEM and OE for all considered scenarios are shown in Fig. 7. The

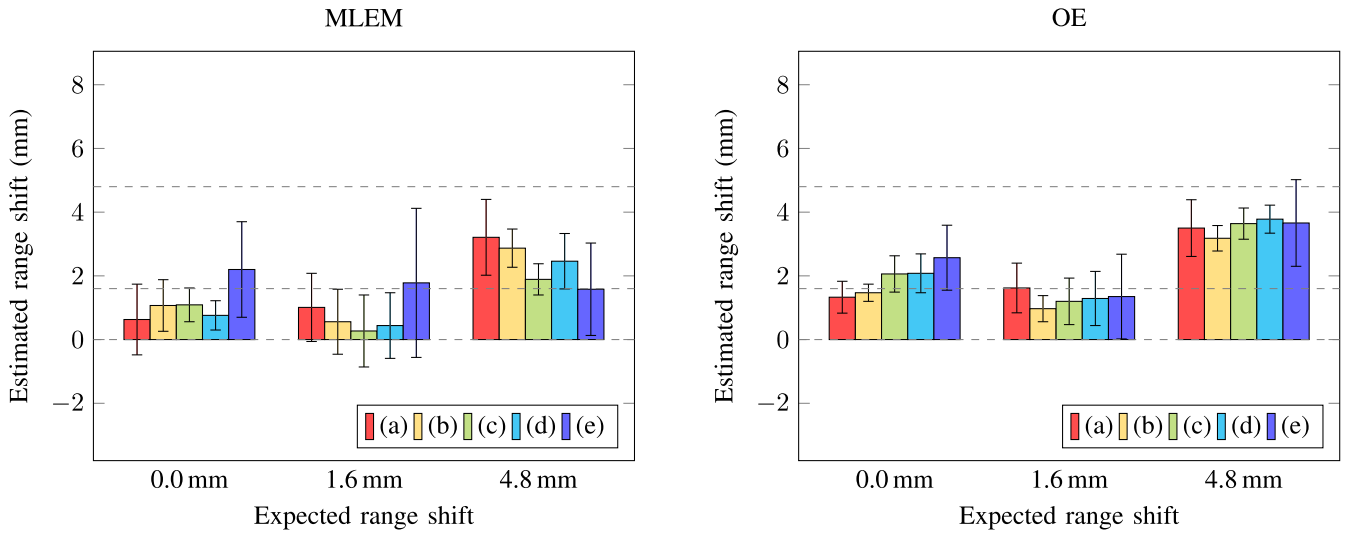

Fig. 9. Mean range shift and standard deviation for the extended phantom compared to the energy-layer phantom for scenario (a) - (e) reconstructed with MLEM (left) and OE (right). Five independent simulations and reconstructions were performed. The expected range shift (marked as dashed lines) are (1) 0.0 mm (120 versus 120 MeV), (2) 1.6 mm (120 versus 121 MeV), and (3) 4.8 mm (120 versus 123 MeV).

images were obtained after summing along the z-direction, and show the view parallel to the camera.

The images illustrate the role of data statistics as well as the accuracy of the information provided by the data. The average number of reconstructed coincidences is provided in Table II. Compared to the ground truth (Fig. 3, left), the region of highest activity close to the distal edge is not clearly recognizable in the images obtained from the *Classical scenario* and the *Energy Peaks*, in agreement with Section III-A. To better visualize the differences, the y-profiles of the energy-layer phantom are shown in Fig. 8. In both methods the steep activity falloff at the distal edges can be observed clearly. MLEM profiles suffer from noise and intensity losses toward the entrance of the beam. For scenario (d), *Hodoscope*, the maximum intensity peak coincides with the ground-truth peak, whereas a slight shift of the maximum is visible for the other scenarios. The blurring of OE images translates into an uniform plateau and a falloff less steep than the ground-truth one. Accordingly, a poorer range-shift identification will be observed, especially for the lowest count scenario (e) (see the next section).

Similar trends as for the single-spot case were further observed: a notable improvement was reached by reconstructing both two-interaction and multiple coincidences, i.e., scenario *Multiple Interactions*, and MLEM images were affected by high statistical noise. OE was able to better reproduce the relatively uniform activity distribution of the energy-layer phantom, but at the cost of higher blurring. It must be noted that the sensitivity term used here only relies on geometrical considerations and ignores the role of the energy. We expect to improve image quality by applying more accurate system response models, as the ones introduced in [19] and [20].

2) Range Verification: The range shifts estimated with the *Activity Depth Profile* method for images reconstructed with MLEM (left) and OE (right) are shown in Fig. 9. Both

TABLE III
HAUSDORFF DISTANCE FOR THE ACTIVITY CONTOUR METHOD. ESTIMATED RANGE SHIFT BETWEEN ENERGY-LAYER AND EXTENDED PHANTOMS $d_{H,RS}$, AND BASE-LINE $d_{H,0}$ (IN MM)

Scenario	MLEM		OE	
	$d_{H,0}$	$d_{H,RS}$	$d_{H,0}$	$d_{H,RS}$
(a)	2.9 ± 0.5	5.0 ± 1.1	2.8 ± 0.9	3.0 ± 0.9
(b)	2.7 ± 0.5	3.9 ± 1.2	1.2 ± 0.4	2.6 ± 0.6
(c)	2.91 ± 0.09	4.7 ± 0.8	2.0 ± 0.6	2.8 ± 1.0
(d)	2.7 ± 1.1	4.2 ± 0.7	1.9 ± 0.7	2.9 ± 0.9
(e)	5.5 ± 1.6	5.9 ± 1.1	3.4 ± 0.8	4.8 ± 1.4

reconstruction approaches perform worse for the larger activity distributions of the phantoms as they did for single spots. Compared to the latter, the number of detected counts is larger but also the number of voxels which contained some activity and the reconstructed FOV. Additionally, some activity originating from neighboring spots overlaps, partly masking the shifts. MLEM-based estimated range shifts are, in general, affected by larger statistical deviations than OE-based ones. In contrast to the single-beam case, in many cases the true shifts were not contained within the error margins and the nonzero shifts were underestimated. The underestimation is more pronounced for MLEM. Additionally, "false positive" shifts arose, in particular when using OE. These effects might be caused by the aforementioned overlapping of the activity distributions. Since the spots with no shifts lie in the middle of the extended phantom, the activity from neighbouring spots is spilled over this region.

Selected results from the *Activity Contour* method are shown in Fig. 10. These images correspond to criterion *Multiple Interactions* (left column) as well as *Energy Cut* (right column), and illustrate the main features of the method. Because the statistical noise affecting MLEM images is not completely removed in the image processing step, irregular

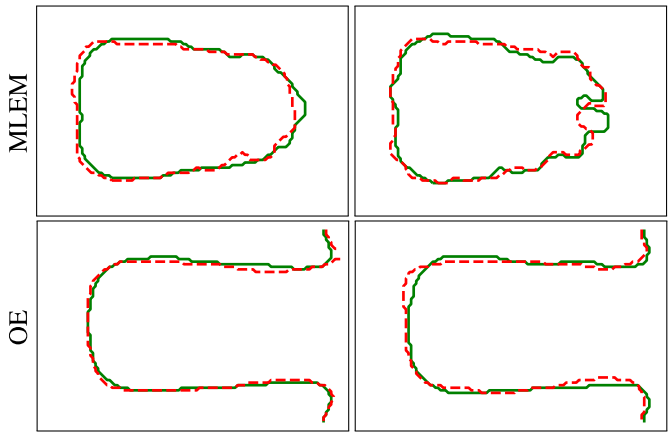

Fig. 10. Range-shift identification with the *Activity Contour* method for scenario *Multiple Interactions* (left) and *Energy Cut* (right), reconstructed with MLEM (top) and OE (bottom). The activity contours are shown in solid-green (energy-layer phantom) and dashed-red (extended phantom).

contours are produced. A higher order filter could overcome this appearance, but at the expense of a stronger edge blurring. Ideally, the Hausdorff distance should correspond to the maximum shift present, i.e., $d_{H,RS} = 4.8$ mm. For these particular realizations, $d_{H,RS}$ was around 4 to 5 mm for MLEM and criteria (a)–(d), whereas $d_{H,RS} \approx 3$ mm for OE images and criteria (a)–(d). Mean values of the Hausdorff distance and their standard deviation are listed in Table III. For reference, the Hausdorff distance was also calculated between the realizations of the energy-layer phantom. The corresponding values, $d_{H,0}$, are ideally equal to zero. In average, the range shifts were well detected using MLEM, particularly for scenario (c), and (d) in a lesser extent. OE underestimated the shifts excepting the noisiest case, (e). Interestingly, MLEM wrongly identified shifts larger than 2 mm when no shift was present (probably due to the high fluctuations, as exemplary shown with the contours in Fig. 10). In general, this effect was less pronounced for OE.

In principle, the *Activity-Contour* method is a promising alternative to deal with large range shifts, as it does not require any prior knowledge about the spot distribution. Smaller shifts might be masked by fluctuations in the data; nevertheless, this was also the case when the shifts were estimated by computing profiles from specific ROIs.

IV. Conclusion

In this article, we have evaluated two reconstruction algorithms, MLEM and OE, for CC-based range verification in particle therapy. Our goal was not to show the superiority of one over the other, but to identify their shortcomings and peculiarities in the proposed scenario toward future improvements.

Compared to other works using OE for CCs, our implementation includes a relatively accurate response model into the acceptance probability with t_{ij} and s_j, as used in MLEM. We also accounted for the stochastic nature of the algorithm by calculating the reconstructed image after averaging over many states, whereas other implementations of OE use one single state which is later smoothed.

The main sources of image degradation were the low number of counts, limited sampling, and incorrect estimation of E_0. The latter was caused by the Doppler effect as well as incomplete absorption of scattered photons. Both phenomena led to incorrect cone apertures which, in turn, resulted in misplaced events, mainly into wrong image slices. This kind of image degradation will be more prominent for a real CC, since perfect spatial and energy resolution were considered in our simulations. Some preliminary simulations including energy resolution for the scatterers and absorber show a non-negligible degradation of the shift estimates. To a certain extend, some compensation should be achieved by modeling the limited resolution into the reconstruction, either in the system matrix or in the selection of the cone parameters [14]–[17]. For the considered CC configuration with seven scatterers including multiple coincidences into the reconstruction helped to increase the statistical quality of the data and compensate for the effects of miscalculated E_0 introduced by two-interaction coincidences. Hence, efforts should be put on implementing a detector read-out able to process such events.

Poor data statistics might remain the most challenging limitation. In our case, it translated into noisy (MLEM) or blurred (OE) images, and large uncertainty errors affecting the range estimations. In modified versions of OE, the lack of statistics is compensated by increasing the number of random points per event and cone [6], [15]. The results provided by this approach are very promising. However, it should be examined if by increasing the number of random points, some bias is introduced or the balance condition, on which the Metropolis–Hastings algorithm relies, is still ensured. For single beams, both MLEM and OE were able to detect interfractional range shifts; the level of accuracy and precision depended on the data selection method and the algorithm. Although no well-defined trends were recognizable, the higher information content of the data selected by the *Energy Peaks* criterion could not compensate for the related signal loss and the resulting increase of statistical fluctuations. In general, a better correlation between estimated and true range differences was observed for OE, which tended to overestimate the shifts. MLEM estimates were affected by larger statistical uncertainties due to the high level of noise of the underlying images. The noise regularization achieved by the Gaussian filter might not be sufficient for this application. As wider smoothing kernels might affect the accuracy of the estimation, other regularization techniques should be investigated. Please note that MLEM was run over a large number of iterations to assure convergence. If online reconstruction is desired, the slow convergence of MLEM will be an issue. The use of priors to regularize MLEM might both help to increase convergence and reduce noise, but at the cost of introducing an additional parameter whose optimal value also depends on the data. This can be problematic in the context of online range verification, as shown in [38] for in-beam PET. For OE, poor data statistics translated into blurred images; we believe that this behavior results from the large variability between the states averaged to compute the reconstructed image. OE results should improve by decreasing the number of image elements. However, a too strong reduction of the FOV size can be problematic, as it should be large enough

to fully cover the region where the PGI originate, and large voxels will hinder the detection of smaller range shifts. Also the possibility of using advanced priors within the framework of OE should be explored, as recommended in [39] for low-count scenarios. The noninformative flat prior distribution on which the current OE implementation relies could be substituted by more informative priors which can compensate for the little information provided by the measured data. Additionally, other capabilities of OE could be exploited, e.g., to estimate the confidence interval of the solution, or optimize the imaging system [40]. Further improvements in the performance of both algorithms might be achieved by optimizing FOV and voxel size. There is also much room for improvement in terms of computational speed, for example, using temporal subsets (e.g., OSEM); in the case of OE, alternative implementations include multithreaded execution and massive parallelization using GPUs [41], [42]. Concerning the role of instrumentation, image quality could be much improved by enlarging the angular coverage of the CC, or using two or more CCs (e.g., [43]) to obtain complementary spatial sampling.

Of particular interest was the evaluation of apparent range shifts when no shifts were present, as this aspect is obviated in many papers. Care should be taken in choosing an adequate threshold to discard "false positives." In our case, those scenarios including multiple coincidences combined with a mild data selection criterion applied to single spots should allow for separation between zero-shifts and 1.6 mm shifts. Smaller shifts might remain undetected. When several spots were reconstructed in one single image, apparent range shifts became a problem. At present, single-spot reconstruction is preferable. It remains open if other CC configurations and improvements in the algorithms might increase the accuracy and precision of the results for multiple-spot reconstruction. Such improvements will be essential if CC are to be used for treatments using an SOBP.

Here, we have used two methods to calculate the range shifts. In principle, the method based on depth profiles should allow single deviations to be calculated. This is theoretically feasible as long as the direction of the beams is known. For multiple-spot reconstruction, this method did not provide reliable results. For this case, the Hausdorff distance was a more appropriate metric to detect large range deviations. This method, based on activity contours, does not rely on *a-priori* information about the beam direction and can cope with PGs from several spots and fields. On the other hand, by definition the Hausdorff distance can only estimate the largest shift. Nevertheless, this metric combined with a threshold could be used to trigger a warning signal. Again, care should be put on selecting a threshold able to separate true shifts from apparent ones. To explore this possibility, further studies using realistic treatment plans would be needed.

In the conclusion, we believe that there is still room for improvement in CC-based PGI, both in software as well as in instrumentation.

ACKNOWLEDGMENT

The authors would like to thank Veronica Ferrero (INFN, Torino, Italy), Jorge Roser (IFIC, Valencia, Spain), and Voichiţa Maxim (CREATIS, Lyon, France) for providing helpful information.

REFERENCES

[1] M. Frandes, A. Zoglauer, V. Maxim, and R. Prost, "A tracking Compton-scattering imaging system for hadron therapy monitoring," *IEEE Trans. Nucl. Sci.*, vol. 57, no. 1, pp. 144–150, Feb. 2010.

[2] S. W. Peterson, D. Robertson, and J. Polf, "Optimizing a three-stage Compton camera for measuring prompt gamma rays emitted during proton radiotherapy," *Phys. Med. Biol.*, vol. 55, no. 22, pp. 6841–6856, 2010.

[3] G. Llosá *et al.*, "First images of a three-layer Compton telescope prototype for treatment monitoring in hadron therapy," *Front. Oncol.*, vol. 6, p. 14, Feb. 2016.

[4] F. Hueso-González, G. Pausch, J. Petzoldt, K. E. Römer, and W. Enghardt, "Prompt gamma rays detected with a BGO block Compton camera reveal range deviations of therapeutic proton beams," *IEEE Trans. Radiat. Plasma. Med. Sci.*, vol. 1, no. 1, pp. 76–86, Jan. 2017.

[5] M. Fontana *et al.*, "Large surface gamma cameras for medical imaging: Characterization of the bismuth germanate blocks," *J. Instrum.*, vol. 13, no. 8, 2018, Art. no. P08018.

[6] E. Draeger *et al.*, "3-D prompt gamma imaging for proton beam range verification," *Phys. Med. Biol.*, vol. 63, no. 3, 2018, Art. no. 035019.

[7] J. Krimmer, D. Dauvergne, J. M. Létang, and É. Testa, "Prompt-gamma monitoring in hadrontherapy: A review," *Nucl. Instrum. Math. A Accelerators Spectrometers Detect. Assoc. Equip.*, vol. 878, pp. 58–73, Jan. 2018.

[8] R. A. Kroeger, W. N. Johnson, J. D. Kurfess, B. F. Phlips, and E. A. Wulf, "Three-Compton telescope: Theory, simulations, and performance," *IEEE Trans. Nucl. Sci.*, vol. 49, no. 4, pp. 1887–1892, Aug. 2002.

[9] T. Hebert, R. Leahy, and M. Singh, "Maximum likelihood reconstruction for a prototype electronically collimated single photon emission system," in *Proc. Med. Imag.*. vol. 767, 1987, pp. 77–84.

[10] A. Sitek, "Representation of photon limited data in emission tomography using origin ensembles," *Phys. Med. Biol.*, vol. 53, no. 12, pp. 3201–3216, 2008.

[11] A. Sitek, "Reconstruction of emission tomography data using origin ensembles," *IEEE Trans. Med. Imag.*, vol. 30, no. 4, pp. 946–956, Apr. 2010.

[12] A. Andreyev, A. Sitek, and A. Celler, "Fast image reconstruction for Compton camera using stochastic origin ensemble approach," *Med. Phys.*, vol. 38, no. 1, pp. 429–438, 2011.

[13] D. Mackin, S. Peterson, S. Beddar, and J. Polf, "Evaluation of a stochastic reconstruction algorithm for use in Compton camera imaging and beam range verification from secondary gamma emission during proton therapy," *Phys. Med. Biol.*, vol. 57, no. 11, pp. 3537–3553, 2012.

[14] A. Andreyev, A. Celler, I. Ozsahin, and A. Sitek, "Resolution recovery for Compton camera using origin ensemble algorithm," *Med. Phys.*, vol. 43, no. 8, pp. 4866–4876, 2016.

[15] Z. Yao, Y. Xiao, Z. Chen, B. Wang, and Q. Hou, "Compton-based prompt gamma imaging using ordered origin ensemble algorithm with resolution recovery in proton therapy," *Sci. Rep.*, vol. 9, no. 1, p. 1133, 2019.

[16] J. E. Gillam *et al.*, "Simulated one pass listmode for fully 3-D image reconstruction of Compton camera data," in *Proc. IEEE Nucl. Sci. Symp. Med. Imag. Conf. Rec. (NSS/MIC)*, 2012, pp. 3298–3305.

[17] S. M. Kim *et al.*, "Resolution recovery reconstruction for a Compton camera," *Phys. Med. Biol.*, vol. 58, no. 9, pp. 2823–2840, 2013.

[18] V. Maxim *et al.*, "Probabilistic models and numerical calculation of system matrix and sensitivity in list-mode MLEM 3-D reconstruction of Compton camera images," *Phys. Med. Biol.*, vol. 61, no. 1, pp. 243–264, 2015.

[19] S. Schoene *et al.*, "An image reconstruction framework and camera prototype aimed for Compton imaging for *in-vivo* dosimetry of therapeutic ion beams," *IEEE Trans. Radiat. Plasma. Med. Sci.*, vol. 1, no. 1, pp. 96–107, Jan. 2017.

[20] E. Muñoz *et al.*, "Study and comparison of different sensitivity models for a two-plane Compton camera," *Phys. Med. Biol.*, vol. 63, no. 13, 2018, Art. no. 135004.

[21] M.-L. Jan, M.-W. Lee, and H.-M. Huang, "PSF reconstruction for Compton-based prompt gamma imaging," *Phys. Med. Biol.*, vol. 63, no. 3, 2018, Art. no. 035015.

[22] H.-M. Huang, C.-C. Liu, M.-L. Jan, and M.-W. Lee, "A low-count reconstruction algorithm for Compton-based prompt gamma imaging," *Phys. Med. Biol.*, vol. 63, no. 8, 2018, Art. no. 085013.
[23] D. Sarrut *et al.*, "A review of the use and potential of the GATE Monte Carlo simulation code for radiation therapy and dosimetry applications," *Med. Phys.*, vol. 41, no. 6, 2014, Art. no. 064301.
[24] A. Etxebeste, Y. Feng, J. M. Létang, V. Maxim, E. Testa, and D. Sarrut, "An extension of the GATE Monte Carlo simulation toolkit to model Compton camera systems," in *Proc. 3rd Geant4 Int. User Conf.*, 2018, p. 1.
[25] L. Grevillot, D. Bertrand, F. Dessy, N. Freud, and D. Sarrut, "A Monte Carlo pencil beam scanning model for proton treatment plan simulation using GATE/GEANT4," *Phys. Med. Biol.*, vol. 56, no. 16, pp. 5203–5219, 2011.
[26] A. C. Kraan, N. Depauw, B. Clasie, M. Giunta, T. Madden, and H. M. Kooy, "Effects of spot parameters in pencil beam scanning treatment planning," *Med. Phys.*, vol. 45, no. 1, pp. 60–73, 2018.
[27] M. J. Berger, J. S. Coursey, M. A. Zucker, and J. Chang. (2005). *ESTAR, PSTAR, and ASTAR: Computer Programs for Calculating Stopping-Power and Range Tables for Electrons, Protons, and Helium Ions (Version 1.2.3)*. Accessed: Sep. 13, 19. [Online]. Available: http://physics.nist.gov/Star
[28] E. Hilaire, D. Sarrut, F. Peyrin, and V. Maxim, "Proton therapy monitoring by Compton imaging: Influence of the large energy spectrum of the prompt-γ radiation," *Phys. Med. Biol.*, vol. 61, no. 8, pp. 3127–3146, 2016.
[29] W. K. Hastings, "Monte Carlo sampling methods using Markov chains and their applications," *Biometrika*, vol. 57, no. 1, pp. 97–109, 1970.
[30] C. Sanderson and R. Curtin, "Armadillo: A template-based C++ library for linear algebra," *J. Open Source Softw.*, vol. 1, no. 2, p. 26, 2016.
[31] A. Zoglauer and G. Kanbach, "Doppler broadening as a lower limit to the angular resolution of next generation Compton telescopes," in *Proc. X-Ray Gamma-Ray Telescopes Instrum. Astronomy*, vol. 4851, Mar. 2003, pp. 1302–1310. [Online]. Available: https://doi.org/10.1117/12.461177
[32] J. E. Gillam *et al.*, "Hodoscope coincidence imaging for hadron therapy using a Compton camera," in *Proc. IEEE Nucl. Sci. Symp. Conf. Rec.*, 2011, pp. 3508–3513.
[33] E. Draeger, S. Peterson, D. Mackin, H. Chen, S. Beddar, and J. C. Polf, "Feasibility studies of a new event selection method to improve spatial resolution of Compton imaging for medical applications," *IEEE Trans. Radiat. Plasma. Med. Sci.*, vol. 1, no. 4, pp. 358–367, Jul. 2017.
[34] A. Koide *et al.*, "Precision imaging of 4.4 MeV gamma rays using a 3-D position sensitive Compton camera," *Sci. Rep.*, vol. 8, no. 1, p. 8116, 2018.
[35] C. H. Min, X. Zhu, K. Grogg, G. El Fakhri, B. Winey, and H. Paganetti, "A recommendation on how to analyze in-room PET for *in vivo* proton range verification using a distal PET surface method," *Technol. Cancer Res. Treatment*, vol. 14, no. 3, pp. 320–325, 2015.
[36] V. Ferrero *et al.*, "Double-field hadrontherapy treatment monitoring with the INSIDE in-beam PET scanner: Proof of concept," *IEEE Trans. Radiat. Plasma. Med. Sci.*, vol. 2, no. 6, pp. 588–593, Nov. 2018.
[37] F. Hausdorff, *Gesammelte Werke Band II: Grundzüge der Mengenlehre*. Berlin, Germany: Springer, 2002.
[38] V. Ferrero *et al.*, "Evaluation of in-beam PET treatment verification in particle therapy with different reconstruction methods," *IEEE Trans. Radiat. Plasma. Med. Sci.*, to be published.
[39] A. Sitek, "Data analysis in emission tomography using emission-count posteriors," *Phys. Med. Biol.*, vol. 57, no. 21, pp. 6779–6795, Nov. 2012. doi: 10.1088/0031-9155/57/21/6779.
[40] A. Sitek and S. C. Moore, "Evaluation of imaging systems using the posterior variance of emission counts," *IEEE Trans. Med. Imag.*, vol. 32, no. 10, pp. 1829–1839, Oct. 2013.
[41] M. C. Lyon, A. Sitek, S. D. Metzler, and S. C. Moore, "Reconstruction of multiple-pinhole micro-SPECT data using origin ensembles," *Med. Phys.*, vol. 43, no. 10, pp. 5475–5483, 2016.
[42] A. Zheng, Z. Yao, and Y. Xiao, "GPU accelerated stochastic origin ensemble method with list-mode data for Compton camera imaging in proton therapy," *IEEE Trans. Radiat. Plasma Med. Sci.*, to be published.
[43] A. Kishimoto *et al.*, "First demonstration of multi-color 3-D *in vivo* imaging using ultra-compact Compton camera," *Sci. Rep.*, vol. 7, no. 1, p. 2110, 2017.