Physics in Medicine & Biology

PAPER • OPEN ACCESS

UMC-PET: a fast and flexible Monte Carlo PET simulator

To cite this article: Pablo Galve *et al* 2024 *Phys. Med. Biol.* **69** 035018

IOP Publishing Phys. Med. Biol. 69 (2024) 035018 https://doi.org/10.1088/1361-6560/ad1cf9

Physics in Medicine & Biology

IPEM Institute of Physics and Engineering in Medicine

CrossMark

OPEN ACCESS

RECEIVED
16 March 2023

REVISED
20 December 2023

ACCEPTED FOR PUBLICATION
10 January 2024

PUBLISHED
30 January 2024

PAPER

UMC-PET: a fast and flexible Monte Carlo PET simulator

Pablo Galve[1,2,3], **Fernando Arias-Valcayo**[1,3], **Amaia Villa-Abaunza**[1], **Paula Ibáñez**[1,3] and **José Manuel Udías**[1,3]

[1] Grupo de Física Nuclear, EMFTEL & IPARCOS, Universidad Complutense de Madrid, CEI Moncloa, 28040 Madrid, Spain
[2] Université Paris Cité, Inserm, PARCC, F-75015 Paris, France
[3] Health Research Institute of the Hospital Clínico San Carlos (IdISSC), Madrid, Spain

E-mail: pgalve@ucm.es

Keywords: Monte Carlo simulation, positron emission tomography, GPU, UMC-PET

Abstract

Objective. The GPU-based Ultra-fast Monte Carlo positron emission tomography simulator (UMC-PET) incorporates the physics of the emission, transport and detection of radiation in PET scanners. It includes positron range, non-colinearity, scatter and attenuation, as well as detector response. The objective of this work is to present and validate UMC-PET as a a multi-purpose, accurate, fast and flexible PET simulator. *Approach.* We compared UMC-PET against PeneloPET, a well-validated MC PET simulator, both in preclinical and clinical scenarios. Different phantoms for scatter fraction (SF) assessment following NEMA protocols were simulated in a 6R-SuperArgus and a Biograph mMR scanner, comparing energy histograms, NEMA SF, and sensitivity for different energy windows. A comparison with real data reported in the literature on the Biograph scanner is also shown. *Main results.* NEMA SF and sensitivity estimated by UMC-PET where within few percent of PeneloPET predictions. The discrepancies can be attributed to small differences in the physics modeling. Running in a 11 GB GeForce RTX 2080 Ti GPU, UMC-PET is ~1500 to ~2000 times faster than PeneloPET executing in a single core Intel(R) Xeon(R) CPU W-2155 @ 3.30 GHz. *Significance.* UMC-PET employs a voxelized scheme for the scanner, patient adjacent objects (such as shieldings or the patient bed), and the activity distribution. This makes UMC-PET extremely flexible. Its high simulation speed allows applications such as MC scatter correction, faster SRM estimation for complex scanners, or even MC iterative image reconstruction.

1. Introduction

Positron Emission Tomography has benefited from Monte Carlo (MC) simulations for decades. Different MC simulation packages have been used for the development and optimization of modern scanners. They have also been used to include realistic physical models in the reconstruction process to improve image quality and help reducing artifacts. For example, the system response matrix (SRM) has been approximated with different approaches using MC simulations (Herraiz *et al* 2006, Gillam and Rafecas 2016, Wei and Vaska 2020), and image corrections such as scatter inside the patient body (Castiglioni *et al* 1999, Ma *et al* 2020), scatter inside detectors (Lee *et al* 2018, Peng *et al* 2018), or positron range modeling (Kraus *et al* 2012, Cal-González *et al* 2015, Cal-Gonzalez *et al* 2018) have been addressed with MC methods. Particle therapy also benefits from MC simulations during PET (and other imaging techniques, such as Prompt Gamma detection) for non-invasive dose monitoring. Nuclear activation during irradiation generates positron emitter fragments which may be used for range verification (Kraan 2015, Bauert *et al* 2019, Masuda *et al* 2020). MC simulations of PET signals are required to have a reliable estimation of detector response and proton range reconstruction in clinical scenarios (Jan *et al* 2013, Choi *et al* 2020, Onecha *et al* 2022).

In recent years, many specific MC simulation toolkits for medical physics have been developed. The most commonly used open source software for medical imaging and radiotherapy is GATE (Jan *et al* 2004, 2011, Grevillot *et al* 2020), which was created to facilitate simulations of medical systems in GEANT4 (Allison *et al* 2016).

GATE is developed in the core of the OpenGATE collaboration, gathering more than a hundred developers for more than 20 years (Sarrut *et al* 2022) that make possible its continuous adaptation to the state-of-the-art. In its current state, GATE is able to model optical photon tracking, silicon photomultipliers (SiPMs), Cerenkov-based time of flight (TOF), or Compton-camera modules, besides it has been used to simulate a wide range of PET and Single Photon Emission Computed Tomography (SPECT) scanners (Sarrut *et al* 2021). SimSET (Poon *et al* 2015) (an acronym for simulation system for emission tomography) is other Monte Carlo simulation package for emission tomography (PET and SPECT) based on variance reduction tools to enhance computational efficiency. Due to its high performance with voxelized geometries, SimSET was also combined with other simulation packages in the past, like MCNP (Du *et al* 2002), GEANT4 (Barret 2005), or GATE (Chen *et al* 2008, Lin *et al* 2014). Recently, SimSET has been integrated in a user-friendly platform called SimPET (Paredes-Pacheco *et al* 2021). A user-friendly adaptation of PENELOPE (NEA 2019) to PET systems has been implemented in PeneloPET (España *et al* 2009, Lopez-Montes *et al* 2019). PeneloPET is a simple-to-configure code for many different scanners since it works with a few simple input text files. However, it is more cumbersome to use in the case of complex geometries that are not based on classical cylindrical shape, since it is outside of the scope of its input files and code modifications might be needed. Even though GATE is known as the reference code for PET simulations due to its widely extended use, PeneloPET has shown to have similar accuracy and faster performance. Both are able to simulate a wide range of PET tracers, pixelated and monolithic detectors, different sources, shieldings, etc. The main difference between them is the underlying physical model: GEANT4 (based on C++) and PENELOPE (based on Fortran). Beyond the physics modeling, different authors compared the performance of PeneloPET against GATE and experimental data, both in preclinical (Vicente *et al* 2010, Popota *et al* 2015) and clinical scanners (Abushab *et al* 2016), and also for positron range tests (Cal-González *et al* 2013), showing reasonable agreement between both packages and reality. Popota *et al* (2015) also discussed that PeneloPET has a more precise modeling of the dead time parameters, though GATE was also able to obtain accurate results. Further information about development of MC codes dedicated to medical physics, and dedicated PET-SPECT codes can be found in Buvat and Lazaro (2006), Rogers (2006).

Even though in the literature there are many MC codes dedicated to medical physics, detailed MC calculations are computationally expensive, and long simulation times may be impractical for many applications. An effective solution to overcome long execution times is the use of general purpose graphical processing units (GPUs), which allow parallel computing in thousands of thread processors, thus increasing the overall code efficiency at the expense of higher programming effort. There are many examples of implementation of GPU parallelization in MC codes for different particle tracing scenarios that have been released along the last fifteen years. Among the literature, we can find independent Monte Carlo codes, such as the one developed by Alerstam *et al* (2008b), based on the White Monte Carlo developed in Alerstam *et al* (2008a), the code from Badal and Badano (2009), Badal *et al* (2021) for photon tracking in the energies between 50 eV to 1 GeV, the GPU Monte Carlo dose code for coupled photon-electron transport in the range 0.01–20 MeV (Hissoiny *et al* 2011), or gPMC for proton dose calculation (Jia *et al* 2012). Other authors relied on previously developed CPU-based MC simulation codes, such as GPU implementation of electron gamma shower (Lippuner and Elbakri 2011), the GPU versions of dose planning method (DPM) (Sempau *et al* 2000) so called gDPM (Jia *et al* 2010, 2011, Chi *et al* 2016), or the GPU Monte Carlo (GMC) (Jahnke *et al* 2012) and GPU accelerated Geant4 based Monte Carlo Simulation (GGEMS) (Bert *et al* 2013), both based on Geant4.

With respect to GPU based codes dedicated to PET, to the best of our knowledge, there are only three packages. The first one is a version of GGEMS adapted to account for the PET detectors (so called GGEMS-PET along this work) in the GPU photon tracking kernel (Ma *et al* 2020). The second one is gPET (Lai *et al* 2019), based on gDPM. The last one is MCGPU-PET[3], based on the previous work from Badal and Badano (2009), Badal *et al* (2021). All these codes share in common the transport of photons through a voxelized phantom, performed in parallel in a one-photon-per-thread trend in the GPU. One of the most relevant differences among them is the detector modeling. In gPET, the detectors are limited to single-layer cuboidal modules with additional boundary constraints based on parameterized surfaces, though gPET developers have announced that multi-layered detectors will be handled in an upcoming release of gPET. The detector parameterization in GGEMS-PET is not described in the literature. MCGPU-PET does not model the detectors, and the final output consists of a phase space file displaying the photon emission from the patient.

The main goal of this work is to present the Ultra-fast Monte Carlo PET simulator (UMC-PET), which has been partially introduced in conferences (Galve *et al* 2020a, 2021). UMC-PET presents a flexible and standardized framework to define the scanner geometry and the detectors; if you can voxelize it, you can simulate it. Within this framework, both the scanner geometry and the patient body are described with a voxelized geometry. For the detectors, their geometry is defined assuming they are composed of blocks, and a high-resolution description of each of these blocks is employed to provide detailed descriptions of the detector

[3] https://github.com/DIDSR/MCGPU-PET

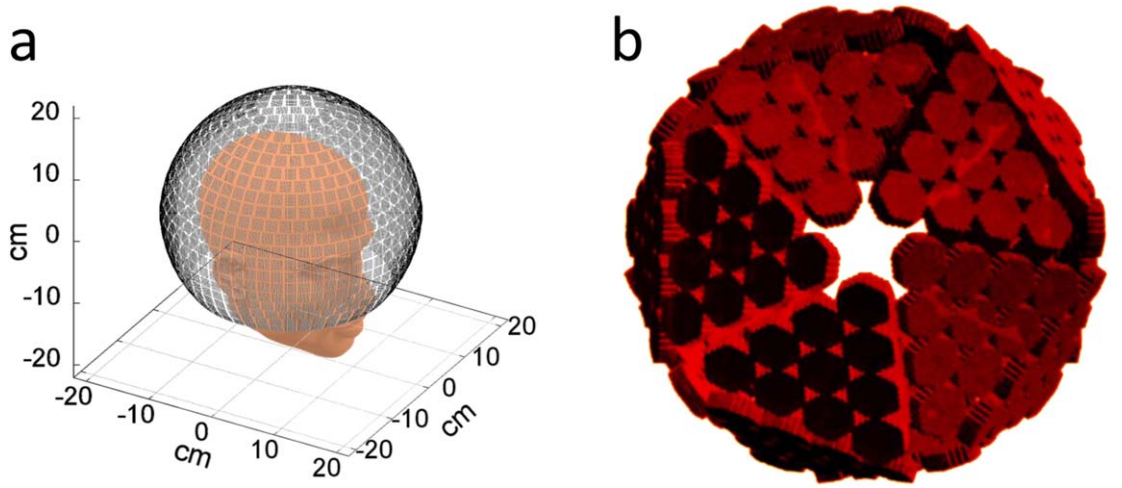

Figure 1. Examples of different non-traditional scanners and detector morphologies which can be implemented in the voxelized scheme of UMC-PET: brain-dedicated PET scanner with spherical geometry (a) (based on Catana (2019)), and icosahedrical geometry with 4π coverage proposed for a preclinical scanner with exagonal pixel detectors (b) (based on Perez-Benito *et al* (2018)).

crystals. This approach enables straightforward definition of any detector module, and given that the size of the space to describe a detector block is relatively small, very small voxel sizes can be employed to describe the detectors. Figure 1 illustrates two PET scanners and detectors with unconventional shapes, extending beyond the traditional cylindrical design and square crystal pixels. Our voxelized framework offers unlimited simulation capabilities for these configurations. The versatility, speed and accuracy of the code positions UMC-PET on par with the aforementioned PET simulation software, enabling accurate estimation of performance parameters of scanners (including spatial resolution or sensitivity), with application to scanner design (Galve *et al* 2020b), or successful improvement of image reconstruction with different approaches, such as scatter correction (Galve *et al* 2022), optimization of the SRM (Arias-Valcayo *et al* 2023), or direct implementation of the simulator in the projection step of the reconstruction process (Galve *et al* 2021). UMC-PET is also able to accurately estimate performance parameters of scanners such as spatial resolution and sensitivity, with application to scanner design. In this paper, the UMC-PET code is explained in detail, and several validations and benchmarks are presented.

2. Methods

We compare UMC-PET against PeneloPET using simulations of the scatter phantoms from the NEMA protocols (National Electrical Manufacturers Association 2007, 2008) in preclinical (6R-SuperArgus) and clinical scenarios (Biograph mMR). We also show a comparison against actual sensitivity values reported in the literature for the Biograph mMR.

2.1. Description of UMC-PET

The workflow of UMC-PET is shown in figure 2. As mentioned before, one of the key characteristics of the code is the voxelized representation of both scanner and object, which makes it possible to define arbitrary scanner geometries and detector shapes. We used NVIDIA CUDA and PGI CUDA Fortran Compilers, with separate CUDA kernels for particle generation, photon tracking, single events processing, and coincidence processing.

2.1.1. Input files

2.1.1.1. Scanner and object definition

We use a voxelized definition of the *world* inside the simulator. The scanner detectors and any additional object (such as shielding, the patient bed, or the patient body) are input as a 3D image (see figure 3). Additionally, we define the coincidence matrix that specifies the couples of detectors that are found in coincidence. Auxiliary programs to translate conventional scanner geometry files, such as the ones used by PeneloPET, to the ones required by UMC have been developed.

The detector blocks and eventually their pixel subdivisions inside are described by means of a voxelized image of a reference block detector. As this image is of relatively small size, a much smaller voxel size than the one

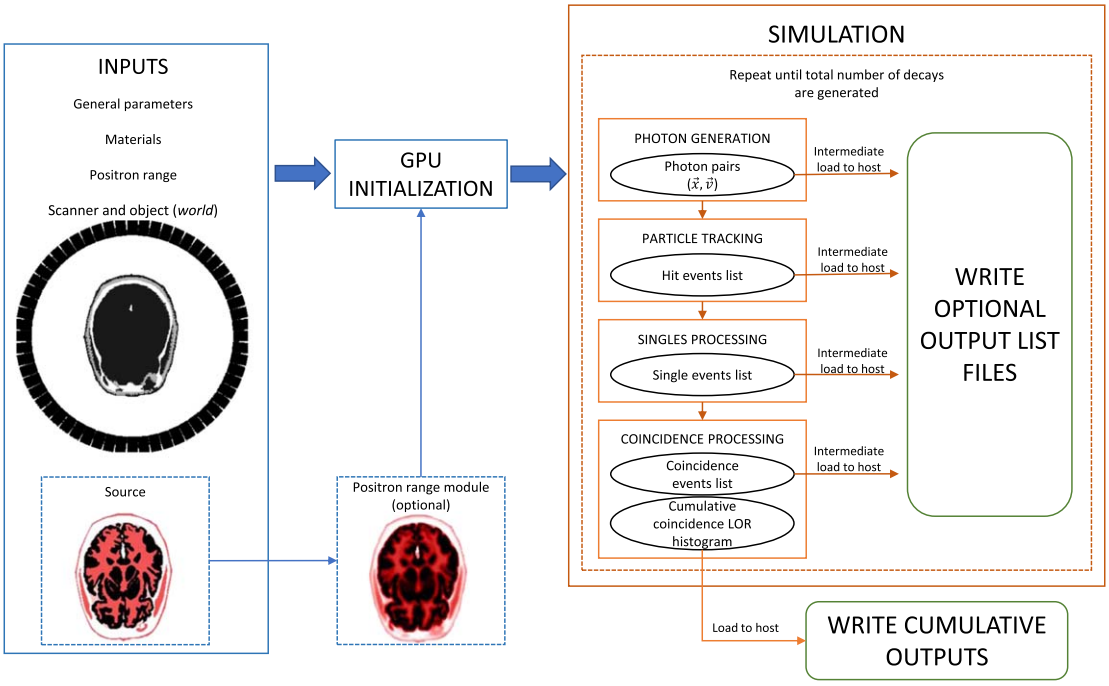

Figure 2. Schematic representation of the workflow of UMC-PET. First, all the inputs regarding computational performance and physical simulation are read. The Positron Range kernel is executed prior to the particle simulation. The photon simulation consists of four routines: photon generation, particle tracking, singles processing, and coincidence sorting. List files and cumulative histograms are generated during the simulation procedure.

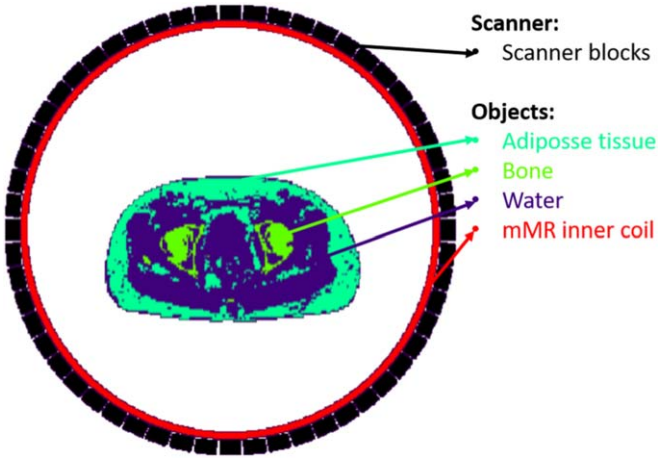

Figure 3. Voxelized image of the mMR scanner with the inner coil inside the PET bore and a pelvis CT segmented in adipose tissue, bone and water.

employed in the *world* image can be employed to define the crystal index and material. Once the reference detector block is described, the position and orientation of each block are given by the central coordinates of one of the faces of the block and three vectors for the block orientation. Figure 4 represents photon identification inside the detector, showing how voxel overlap between adjacent crystal pixels is avoided through the distinction between the *world* image and the *detector* image. This differentiation is crucial as it prevents the need for using higher-resolution world images, which would result in increased overall memory need.

This approach, can accurately represent traditional rectangular prism-shaped blocks, but it also accommodates complex block geometries and crystal distributions, such as the hexagonal prism pixel used by Perez-Benito *et al* (2018) (see figure 1), or multi-layered detectors (Wang *et al* 2006, Mohammadi *et al* 2017). The only limitation is the requirement to use a voxel size small enough to represent as fine details of the detectors as needed. For instance, reflector material or gap between adjacent crystals, which may be as small as 1/10 of the crystal pitch, can be considered this way. Memory size of the detector image may be a concern to describe very

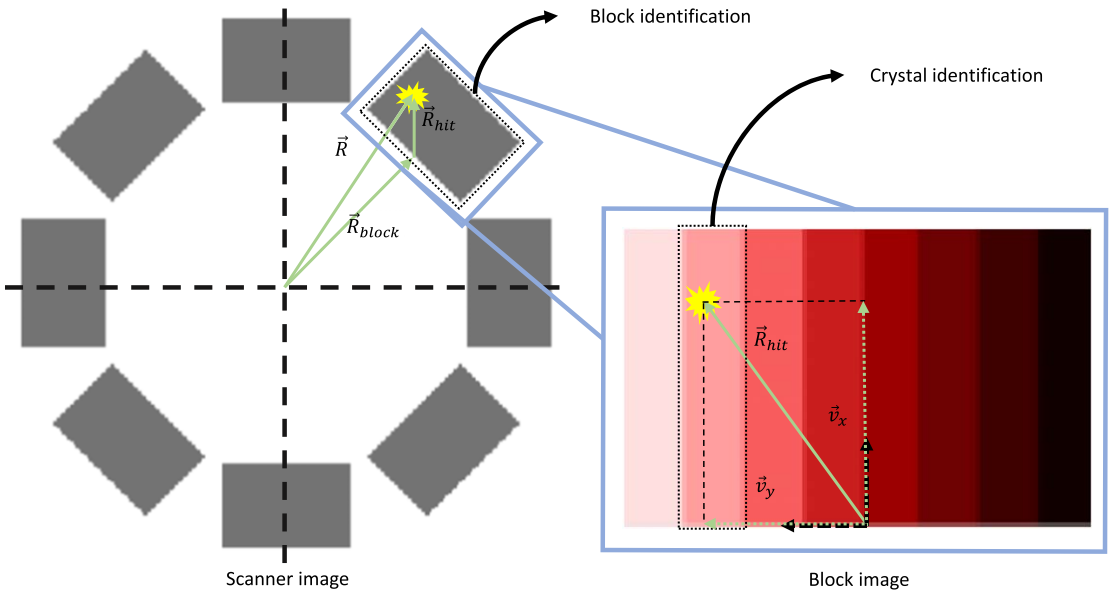

Figure 4. Scheme of the crystal identification after a hit occurs inside the detector. The position vector of a block is referred as $\vec{R}_{block}$, while the orientation vectors are referred as $\vec{v}_x$ and $\vec{v}_y$ (the third vector $\vec{v}_z$ is out of the 2D image). These vectors define the coordinate system to locate every hit ($\vec{R}_{hit}$) inside the detector image.

small details in the case of large detector blocks. In this case, the user may simply subdivide the blocks into smaller identical units, and use these sub-unit as new detector reference for the detector image, if possible. Alternatively, it is possible to define a material with equivalent properties to the combination of inter-crystal and scintillator materials, resulting in the same effective sensitivity of the block without need to define fine structures.

Detector readout is simulated using Anger logic (Anger 1969) considering optical reading at the back-end of the detector. The user defines a look-up table (LUT) and the centroid of each crystal inside the LUT (see section 2.1.2.4 for further details). The LUT is read as a 2D image that defines the crystal that corresponds to each final signal. Energy and time resolution are defined for every crystal index independently.

All the images are read as binary inputs, whose dimensions (number of voxels, voxel size, relative position or orientation) are specified in separate text files.

2.1.1.2. Source definition

UMC-PET simulates annihilations emitted from a voxelized 3D image of the radionuclide distribution. It is possible to choose the positron range (PR) blurring kernel for the radionuclide, that is applied to the source activity distribution. The PR kernel is based on the analytical implementation of Cal-González *et al* (2015), using the density map to estimate the electronic density of the medium. The PR kernel blurs the radionuclide distribution to generate the annihilation map (see figure 2).

2.1.1.3. Materials

Photon attenuation coefficients of a given list of predefined materials are obtained from PENELOPE (NEA 2019). More specifically, we extract the coefficients of coherent scatter (Rayleigh scatter), incoherent scatter (Compton scatter), and photoelectric absorption in the range from 1 keV to 1 MeV (pair production is forbidden below 1.022 MeV). To reduce calculations, the dispersion angle and energy deposited after each scatter event are randomly selected from a table built from simulated events binned every 1 keV.

2.1.1.4. GPU control parameters

The number of threads and blocks used in the GPU are left for optimization by the user, as well as the maximum number of hits/singles kept in the GPU memory per particle (seldom over more than ten interactions occur, although it depends on the simulation).

2.1.2. Main routines

In this section we describe the main routines involved in the PET simulation. As stated in figure 2, the routines for photon initialization (2.1.2.2), photon tracking (2.1.2.3), singles generation (2.1.2.4), and coincidence sorting (2.1.2.5) work sequentially, forwarding the output of each routine into the next one until all the

annihilations are simulated. Each subroutine works with a batch of decays defined by the number of threads and blocks.

2.1.2.1. Random number generator

The Fortran implementation of the subroutine *RANECU* (James 1990) is used for pseudo-random number generation. To guarantee the independence of the distributions among different threads, millions of seeds were precomputed, and transferred to each thread on the GPU (Ibáñez *et al* 2021).

2.1.2.2. Particle initialization

In this routine, all threads are initialized with a batch of antiparallel photons. The emission voxel is picked using the Walker's aliasing method (Salvat 1987), and the physical emission point is randomly selected with a uniform distribution inside the voxel volume. The emission direction is randomly selected with isotropic distribution. To model non-collinearity, one of the photons is rotated with a Gaussian angular distribution of given input FWHM. In this work, FWHM $= 0.5^{\circ}$ was used (Harrison *et al* 1999) for comparison against PeneloPET, although some authors suggest that FWHM $= 0.617^{\circ}$ better fit reality (Shibuya *et al* 2007).

2.1.2.3. Photon tracker

The two photons for every annihilation are tracked in a single thread until they are totally absorbed or they are out of the scanner. The Woodcock algorithm (Woodcock *et al* 1965, Carter *et al* 1972) for particle tracking has been implemented to simulate every particle step. To avoid using the mean free path of highly attenuating materials (usually scintillator materials) when the photons travel through body tissues, we also implemented subregions with higher reference mean free path to accelerate the simulation (Badal and Badano 2012, Behlouli *et al* 2018). In our implementation, we studied the vincinity of every voxel to define its reference material and longest step inside a closed sphere (pre-processing step). The material index is read from the object image, and material attenuation coefficients are read from the materials table. In case a detector is found, the particle position is localized inside the detector image, as explained in section 2.1.1.1. For objects defined with Hounsfield units (HU), the attenuation coefficient is linearly scaled between air and water or water and bone, using the bilinear conversion of Burger *et al* (2002). All interactions are saved in the global memory of the GPU device (up to the maximum number of hits) for further use in the singles generator routine.

2.1.2.4. Singles generator

Every thread analyzes independently the hits list for each annihilation photon to generate a list of singles. When the hit is located inside a detector block, the energy deposited and the hit TOF are blurred using respectively the energy and time resolution of the crystal. The energy resolution is given at 511 keV, and is scaled with the square root of the deposited energy for other energy values. To model the detector readout, crystal identification is derived from the energy-weighted center of gravity and the LUT. The Anger logic employed can be summarized as follows:

$$\begin{aligned} E &= \sum_{n=1}^{N} E_i \\ x &= \sum_{n=1}^{N} E_i \cdot x_i / E \\ y &= \sum_{n=1}^{N} E_i \cdot y_i / E., \end{aligned} \tag{1}$$

Where E_i is the energy deposited by hit i, and (x_i, y_i) are the centroid coordinates of the crystal hit inside the LUT. N is the total number of hit signals from a given photon inside the same detector. The (x_i, y_i) crystal centroid of each hit is energy-weighted and accumulated. The final (x, y) signal (averaged by the total energy deposited) is inputted to the LUT, giving the resultant crystal. If multiple detectors are hit, more than one single event can be generated by a single photon. The TOF is given by the earliest time signal in the detector after applying the timing resolution on every hit contributing to the single event. If multi-layer detectors are defined, we use an energy weighted method to determine which layer the single is attributed to. At the end of this routine, a list of single events is saved in GPU global memory.

The presented model for defining the detector readout and electronics provides a generalized description of the detector. It is important to note that while this model captures fundamental aspects, it may not encompass all detectors currently available in the market. The diversity of detector modules available today, including one-to-one and multiple coupling photomultipliers, double-side readout, various crystal reflectors, monolithic detectors, and others, renders it unfeasible to create a single model capable of encompassing all possibilities

within the scope of this research. We recommend the implementation of a separate postprocessing step to handle the list of hits events in the specific cases where the proposed model does not align with the detector readout and electronics.

2.1.2.5. Coincidence generator

This routine parses the single events list of every decay in a thread to determine whether to record the coincidence event, discarding couples of detectors out of the coincidence matrix and events out of the energy or time windows.

2.1.3. Output files

UMC-PET output files can be easily adapted by the user. The default outputs include events list files, line of response (LOR) histograms of coincidences, and other files related to general information about the simulation.

List of events chunks are moved from device memory to host memory after every call to the main routines, and the host CPU processes these list sequentially, discarding empty spaces while saving the events in a larger buffer (notice we work with chunks of particles with an independent mini-buffer of hits, singles, or coincidences for every particle/annihilation). Every time the buffer in the host is full, the code writes to disk the list of events and resets the buffer, thus reducing disk writing operations. Furthermore, these disk writing operations are done asynchronously, thus they are hidden 'behind' the main routines on the GPU, in terms of computing time. The list format can be easily tailored to the requirements of the experiment, and it may include information about the time-of-flight, energy, scatter information, crystal index, emission voxel, and many other parameters of interest. Accumulated outputs, such as LOR histogram of coincidences, emission image, hits image, or other parameters such as scatter fraction or sensitivity, are written at given intervals during the simulation and finally at the end of the run.

2.2. The 6R-SuperArgus scanner

The 6R-SuperArgus is a preclinical scanner based on the SuperArgus detector module (also present in the GE Healthcare eXplore Vista from General Electrics (Wang *et al* 2006), currently commercialized by Sedecal Medical Imaging). The scanner consists of 6 rings of 24 SuperArgus modules each, with 17 cm inner diameter and total 15 cm axial length for the six rings. Each module has a pixellated scintillator array of 13 $\times$ 13 crystals of 1.55 mm crystal pitch with a dual layer phosphor sandwich (phoswich) strategy for depth of interaction (DOI) information. This scintillator array is coupled to position-sensitive photomultiplier tubes (PMT). The phoswich array is made of lutetium-yttrium orthosilicate (LYSO) crystals of 7 mm length in the front layer, optically coupled to cerium-doped gadolinium orthosilicate (GSO) crystals of 8 mm length in the rear layer. The energy and time resolution has been chosen to match the properties of the actual scanner; 21% for the LYSO crystals, 33% for the GSO crystals, and 1.5 ns coincidence resolving time FWHM resolution. We use the 425–600 keV and 100–700 keV energy windows. A schematic representation of the 6R-SuperArgus is shown in the top side of figure 5.

2.3. The biograph mMR scanner

The Biograph mMR is a PET/MRI scanner for whole-body PET imaging (Delso *et al* 2011). The scanner consists of 8 rings of 56 detectors with 65.6 cm diameter, resulting in 59.4 cm transverse FOV and 25.8 cm axial FOV. The detector blocks are made of arrays of 8 $\times$ 8 lutetium oxyorthosilicate (LSO) crystals of 4 mm crystal pitch and 20 mm depth. Light readout is performed by an array of 3 $\times$ 3 avalanche photodiodes (APD). For the simulated scanner we chose the energy resolution of 14.5%, energy window of 430–610 keV (we also simulated 100–610 keV for some tests), time resolution is 2.93 ns, and coincidence window is of 5.86 ns, matching the actual scanner values. A schematic representation is shown in the bottom of figure 5.

2.4. NEMA scatter fraction phantoms

We have simulated rat-like and mouse-like phantoms described in NEMA NU 4-2008 (National Electrical Manufacturers Association 2008), and the clinical phantom described in the NEMA NU 2-2007 (National Electrical Manufacturers Association 2007) for scatter fraction (SF) assessment. In figure 5, we show a scheme of the phantoms. All the phantoms are cylinders with a line source parallel to the scanner at different distances off-axis. The mouse-like phantom is 25 mm diameter and 70 mm length, and the source is 10 mm off-axis and 60 mm long, whereas the rat-like phantom is 50 mm diameter and 150 mm length, and the source is 17.5 mm off-axis and 140 mm long. The clinical phantom is 203 mm diameter and 264 mm length, and the source is 45 mm off-axis. In all the cases the line source uses ^{18}F, it has a diameter of 3 mm, and the scatter cylinder is made of water.

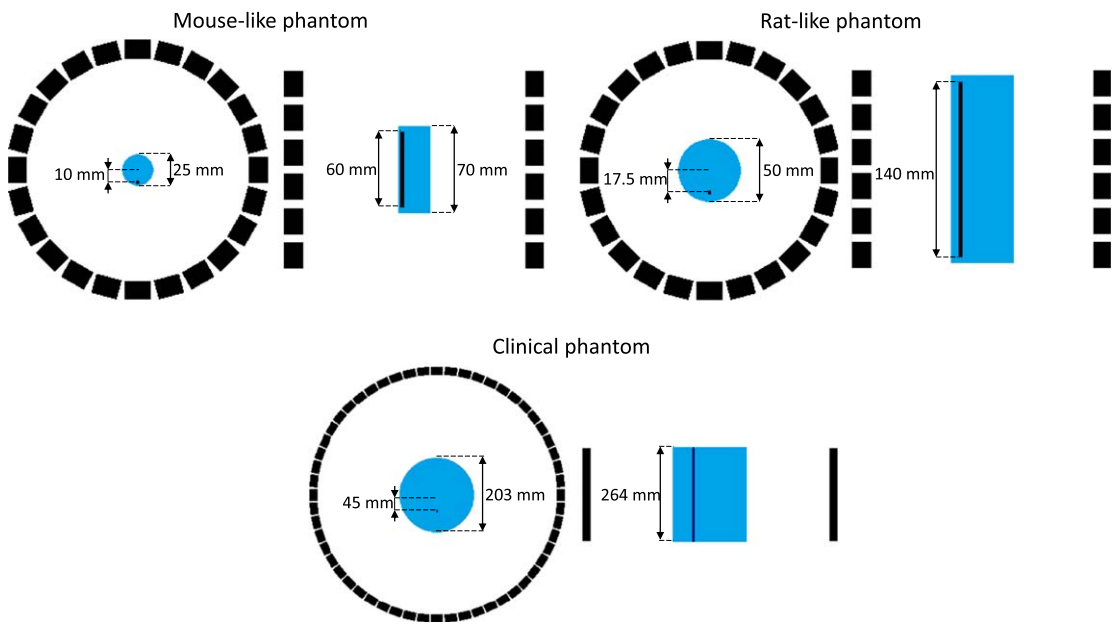

Figure 5. Scheme of the simulated NEMA NU 4-2008 mouse-like (top-left) and rat-like (top-right) phantoms inside the 6R-SuperArgus scanner, and the NEMA NU 2-2007 clinical phantom inside the Biograph mMR scanner. We show a central slice of the transverse view and sagittal view.

2.4.1. Simulation details

In table 1, we display the details of the 3D-volumes included in the UMC-PET simulations. In the preclinical simulations with the 6R-SuperArgus scanner, a total number of decays of 6.23×10^7 were simulated, whereas we simulated 4.47×10^8 decays for the Biograph mMR case.

2.4.2. Quantitative assessment

We compare the energy histograms, the SF following the NEMA protocols (National Electrical Manufacturers Association 2007, 2008), and the sensitivity of each simulation. In figure 6 we show a scheme of the scatter profile used to assess the NEMA SF, given by the ratio between the scatter events and the total events. The sensitivity is given by the ratio between the total events and the emitted decays.

2.5. Point source measurements for resolution assessment

To verify the accuracy of UMC-PET for image assessment, we compare the image resolution obtained in real acquisitions with the UMC-PET estimated resolution. A point source (less than 0.5 mm diameter encapsulated in a 3×3 mm diameter small epoxy cylinder) of ^{22}Na of 5 μCi was placed on the bed of the 6R-SuperArgus at different positions. We ran simulations at equivalent positions with 1×10^9 emissions per image. Both acquired and simulated sources were reconstructed using an OSEM algorithm without resolution modeling in order to retrieve the system resolution from the images (Iriarte *et al* 2016). We fitted the radial, transverse and axial profiles to a Gaussian function to evaluate the full width at half maximum (FWHM) in every case.

2.6. Biograph mMR NEMA measured sensitivity

We simulated the NEMA NU 2-2007 procedure (National Electrical Manufacturers Association 2007) to measure sensitivity and compared it with the results reported by Delso *et al* (2011) for the Biograph mMR. We have used the energy window of 430–610 keV. The inner coil of the magnetic resonance (MR) was modelled using the description given by Delso *et al* (2009), Aklan *et al* (2015): a hollow cyllinder of 10 mm depth made of glass fiber reinforced plastic (GRP) (33% carbon, 55% hydrogen and 13% oxygen with a density of 1.18 g cm^{-3}). Since we did not know the exact description of the patient bed, we modelled it as a simplified hollow prism of 37 mm $\times$ 5 mm $\times$ 264 mm with 5 mm depth, made of glass fiber (10% sodium, 5% calcium, 25% silicon and 60% oxygen with a density of 2.5 g cm^{-3}) (Delso *et al* 2009) to obtain equivalent absorption for the 511 keV photons. The source had 3.9 mm diameter and 700 mm length, and it has been simulated at the center of the simulation and 10 cm off-center. In figure 7 we show the simulation scheme. The voxelized volumes employed have the same details used in the table 1 for the mMR. We have sinogrammed the data using the single slice rebinning (SSRB), and the sensitivity per slice and total sensitivity is given.

Table 1. Details of the voxelized volumes used in the UMC-PET simulator.

Image	#Voxels	Voxel size (mm)	Volume size (mm)	Data format	Memory size (MB)
Scanner and object (6RSA)	269 × 269 × 208	0.775 × 0.775 × 0.775	208.5 × 208.5 × 161.1	Signed Short Integer (2B)	28.7
Detector (6RSA)	13 × 13 × 15	1.55 × 1.55 × 1.00	20.15 × 20.15 × 15	Signed Short Integer (2B)	4.95 (kB)
Source (mouse)	60 × 60 × 100	0.1 × 0.1 × 0.12	6 × 6 × 12	Float (4B)	1.37
Source (rat)	60 × 60 × 100	0.1 × 0.1 × 0.28	6 × 6 × 28	Float (4B)	1.37
Scanner and object (mMR)	360 × 360 × 66	2 × 2 × 4	720 × 720 × 264	Signed Short Integer (2B)	16.3
Detector (mMR)	8 × 8 × 1	4 × 4 × 20	32 × 32 × 20	Signed Short Integer (2B)	128 (B)
Source (clinical)	60 × 60 × 100	0.1 × 0.1 × 2.64	3.2 × 3.2 × 264	Float (4B)	391 (kB)

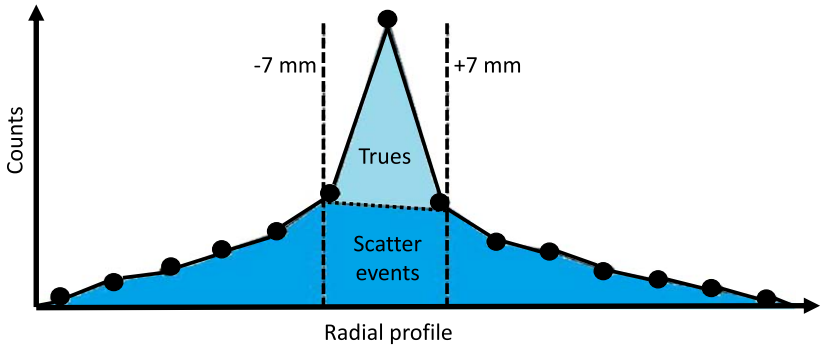

Figure 6. Scheme of the profile in the sinogram of the preclinical NEMA NU 4-2008 Scatter Fraction phantom, with SSRB, maximum pixel aligned in the center of each projection, and angular-axial projection collapsed. The dark-blue region defines the scatter area, whereas the light-blue region is the true events area. In the case of the clinical NEMA NU 2-2007, a strip of ±20 mm is used.

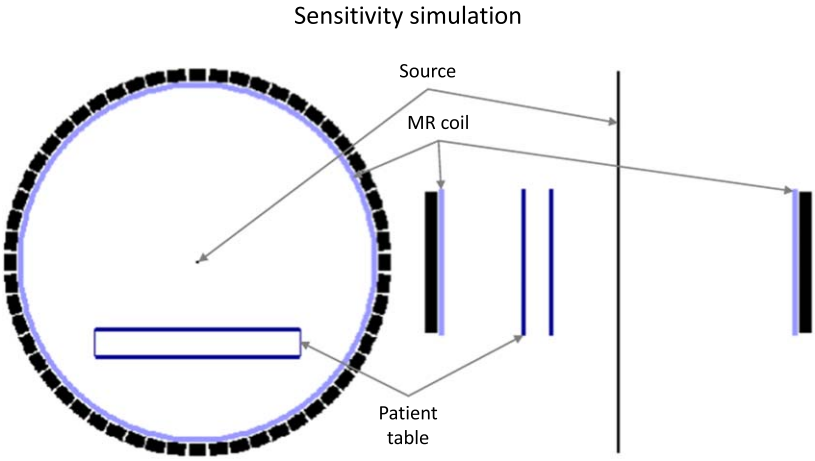

Figure 7. Scheme of the simulated NEMA NU 2-2007 sensitivity phantom in the Biograph mMR with the patient table and the inner MR coil. We show a central slice of transverse and sagittal view.

2.7. CPU and GPU devices

We benchmarked PeneloPET in a single core of an Intel(R) Xeon(R) W-2155 CPU @ 3.30 GHz. For UMC-PET, we ran the code in the same CPU model for the host computation, and a 11 GB GeForce RTX 2080 Ti GPU, with 4352 cores, as GPU for the multi-thread part of the calculations.

3. Results

3.1. NEMA scatter fraction and sensitivity

Figure 8 shows the energy histograms and SF radial profiles generated by UMC-PET and PeneloPET for the phantoms proposed. The energy histograms have been generated for the coincidence events of acquisitions without energy window.

In table 2, we show the SF calculated for the profiles using the NEMA protocol, and the sensitivity given by the simulation output information. The NEMA SF and sensitivity of both simulators are in good agreement, below 3% relative difference in all the cases.

In table 3, we present the computing time required by each simulation, comparing the rate of decays and coincidences simulated. Simulations conducted using UMC-PET required an additional time of less than 6 s for loading all input files into GPU memory, precomputation of necessary arrays for the Walker's method, and optimizing the reference attenuation material in the subregions for the Woodcock algorithm. We did not account for the preprocessing time required to estimate decay and coincidence rates, as this remains constant, irrespective of the number of simulated decays. We observe an acceleration factor of ~2000 in all the preclinical simulations, and ~1500 for the clinical simulation. This difference is attributed to the larger simulated 3D volumes in the clinical case, which increases the average number of Woodcock steps executed per thread (5.50 and 5.94, respectively, for the mouse and rat preclinical scenarios and 11.1 for the clinical one). The time

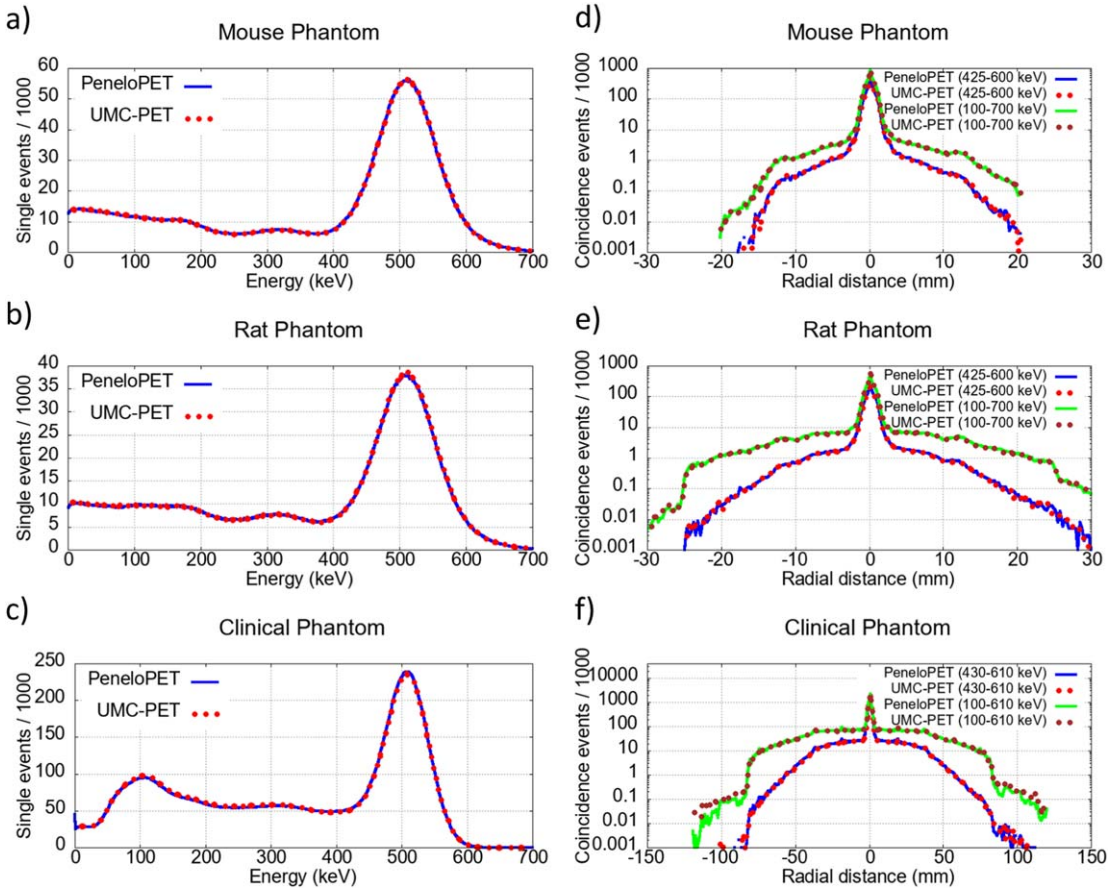

Figure 8. Energy histograms (a, b and c) and radial profiles for the NEMA SF (d, e and f) simulated in PeneloPET and UMC-PET with the mouse (top), rat (center) and clinical (bottom) phantoms. The radial profiles have been angular-axially collapsed and maximum-centered for two different energy windows (100–700 keV and 425–600 keV, for the mouse and rat phantom, and 100–610 keV and 430–610 keV for the clinical phantom) for PeneloPET and the UMC-PET simulator.

Table 2. Estimated SF using the NEMA protocol and sensitivity for the mouse, rat and clinical phantoms using UMC-PET and PeneloPET.

Mouse phantom			
	keV	PeneloPET	UMC-PET
NEMA SF (%)	425–600	4.51	4.46
	100–700	7.48	7.38
Sensitivity (%)	425–600	2.96	3.02
	100–700	6.32	6.41
Rat phantom			
	keV	PeneloPET	UMC-PET
NEMA SF (%)	425–600	11.5	11.2
	100–700	24.9	24.6
Sensitivity (%)	425–600	1.89	1.92
	100–700	4.93	5.00
Clinical phantom			
	keV	PeneloPET	UMC-PET
NEMA SF (%)	430–610	33.4	33.1
	100–610	62.6	62.7
Sensitivity (%)	430–610	1.20	1.16
	100–610	4.10	4.13

Table 3. Computation time (simulation time, number of simulated decays per second, and coincidences generated per second) for the mouse, rat, and clinical phantoms using UMC-PET and PeneloPET. Last column shows the acceleration factor obtained with UMC-PET. The values presented in this table for the UMC-PET did not account for the time needed for input reading and data loading in the GPU (less than 6 s for each simulation). PeneloPET was executed in a single core of an Intel(R) Xeon(R) W-2155 CPU @ 3.30 GHz, whereas UMC-PET used a 11 GB GeForce RTX 2080 Ti GPU, with 4352 cores for the multi-thread part of the calculations.

Mouse phantom				
	keV	PeneloPET	UMC-PET	Sped up ratio
Simulation time (s)	425–600	2999	1.31	2288
	100–700	2912	1.35	2162
Decays/s	425–600	2.08×10^4	4.75×10^7	2288
	100–700	2.14×10^4	4.62×10^7	2162
Coincidences/s	425–600	6.14×10^2	1.41×10^6	2297
	100–700	1.35×10^3	2.93×10^6	2167
Rat phantom				
	keV	PeneloPET	UMC-PET	Sped up ratio
Simulation time (s)	425–600	2989	1.42	2098
	100–700	2990	1.47	2041
Decays/s	425–600	2.08×10^4	4.37×10^7	2098
	100–700	2.08×10^4	4.25×10^7	2041
Coincidences/s	425–600	3.93×10^2	8.25×10^5	2100
	100–700	1.03×10^3	2.10×10^6	2040
Clinical phantom				
	keV	PeneloPET	UMC-PET	Sped up ratio
Simulation time (s)	430–610	2.45×10^4	15.5	1580
	100–610	2.48×10^4	15.7	1581
Decays/s	430–610	1.83×10^4	2.89×10^7	1580
	100–610	1.80×10^4	2.85×10^7	1581
Coincidences/s	430–610	2.19×10^2	3.40×10^5	1555
	100–610	7.41×10^2	1.18×10^6	1592

Table 4. Computing time per billion decays of the main routines (decay initialization, photon tracker, singles generator, and coincidences generator) in the UMC-PET simulator. The total time is given in table 3. *The energy window in the case of the clinical phantom is 430–610 kev and 100–610 keV.

	keV *	Mouse		Rat		Clinical	
		(s/10^9 dec.)	(%)	(s/10^9 dec.)	(%)	(s/10^9 dec.)	(%)
Decay init.	425–600	0.38	1.8	0.38	1.7	0.51	1.5
	100–700	0.38	1.8	0.38	1.6	0.52	1.5
Photon tracker	425–600	17	83	19	84	29	85
	100–700	17	80	19	82	29	84
Sing. generator	425–600	3.0	14	3.0	13	3.7	11
	100–700	3.5	16	3.5	15	4.1	12
Coin. generator	425–600	0.22	1.1	0.21	0.93	0.32	0.94
	100–700	0.32	1.5	0.32	1.3	0.44	1.3

required for every routine in UMC-PET is presented in table 4, showing that the simulator spends most of its time in the photon tracker routine. Remaining time is dedicated to CPU general operations (<1%).

3.2. Image resolution assessment with point sources

In table 5 we show the measured FWHM from acquired and simulated point sources at different positions. In all the cases simulations and actual data are with differences below 4%.

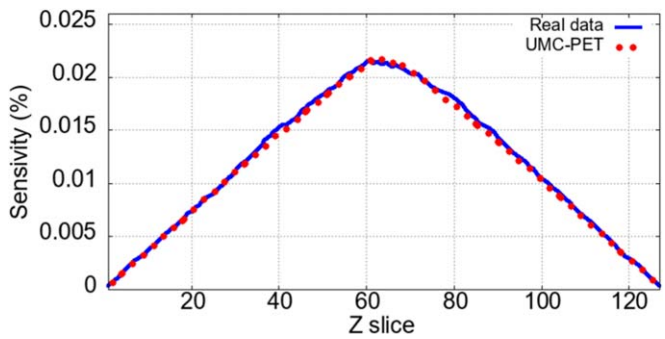

Figure 9. Axial profiles of the NEMA sensitivity simulated with UMC-PET for the Biograph mMR at the center of the scanner. Real data taken from figure 2 of (Delso *et al* 2011).

Table 5. Radial, tangential and axial FWHM measured on point sources acquired and simulated at different positions of the 6R-SuperArgus. The images were reconstructed using the OSEM algorithm without resolution modeling.

Source position		Radial FWHM (mm)		Tangential FWHM (mm)		Axial FWHM (mm)	
z (mm)	x (mm)	Acquisition	UMC-PET	Acquisition	UMC-PET	Acquisition	UMC-PET
0	15	1.59	1.53	1.32	1.29	1.34	1.35
0	35	2.06	2.03	1.76	1.71	1.50	1.48
0	60	2.55	2.53	1.91	1.96	1.62	1.59
5	15	1.41	1.44	1.31	1.28	1.63	1.61
5	35	1.95	1.89	1.58	1.60	1.53	1.49
5	60	2.70	2.65	1.84	1.83	1.79	1.82

Table 6. Estimated sensitivity using the NEMA protocol in the Biograph mMR for the measured values reported in the literature (Delso *et al* 2011) and the simulated values with the UMC-PET.

	NEMA sensitivity (%)	
	Measured value	UMC-PET
0 cm off-center	1.50	1.48
10 cm off-center	1.38	1.43

3.3. Comparison with the biograph mMR measured NEMA sensitivity

In figure 9 we show the axial sensitivity profiles at the center of the Biograph mMR scanner measured on an actual scanner and simulated with UMC-PET. Table 6 compares the measured sensitivity reported in the Biograph mMR scanner with the same parameter derived from UMC-PET simulations. Measured and simulated sensitivity values agree within a few percent.

4. Discussion

Among the few existing GPU-based MC codes dedicated to PET we may find in the literature, UMC-PET stands out for its extremely flexible framework for defining scanners and detectors. In table 7, we provide a concise overview of the key capabilities of other GPU-based PET simulation packages, including GGEMS-PET, gPET, MCGPU-PET, and UMC-PET. The discussed simulation packages exhibit both advantages and limitations. The selection of the most suitable simulator should be contingent upon the user's specific requirements.

In the introduction, we have outlined the distinctions in detector modeling among each software, emphasizing the innovative voxelized approach adopted by UMC-PET for precise detector geometry definition. In contrast to gPET, UMC-PET can simulate a variety of detector configurations, including non-cuboidal detectors, such as the hexagonal crystal pixel proposed by Perez-Benito *et al* (2018), figure 1(b)) or complex multi-layered detectors[4] (Mohammadi *et al* 2017), which represent the current state-of-the-art in mitigating

[4] The developers of gPET have indicated that multi-layered detectors will be handled in a future upgraded release.

Table 7. Main details of the GPU-PET simulation packages available: GGEMS-PET, gPET, MCGPU-PET, and UMC-PET.

	GGEMS-PET	gPET	MCGPU-PET	UMC-PET
Platform	NVIDIA CUDA	NVIDIA CUDA	NVIDIA CUDA	NVIDIA CUDA FORTRAN
Detector parameterization	Not specified	Cuboidal repetitive structures	Phase Space	Voxelized volumes
Multi-layered detectors	Not specified	No (expected in an upcoming release)	N/A	Yes
Detector response	Yes	Yes	N/A	Yes
Photon transport	Voxel-wise step	Woodcock algorithm	Woodcock algorithm	Woodcock algorithm
Cross sections data base	Tables from Biggs and Lighthill (1988)	gDPM	PENELOPE	PENELOPE
Positron range	No	Yes	Yes (external software)	Yes
Non-collinearity	No	Yes	Yes	Yes
Non-pure beta emitters	No	No	Yes	No
Time evolution	Yes (CPU)	Yes (GPU)	Yes (GPU, no singles sorting)	Yes (external software)

depth of interaction effects. For GGEMS-PET, the specific methodology for detector definition remains unspecified. MCGPU-PET exclusively considers a cylindrical phase space surrounding the phantom, while the detector response in this code accounts just for energy resolution and energy window. gPET includes additional spatial blurring effects inside detector blocks, and it allows different readout schemes inside a block. In GGEMS-PET each block of the scanner shares the same readout electronics. In UMC-PET, the hit processing routine can be adapted to a particular read-out/multiplexing scheme, albeit a center of energy algorithm for crystal identification is employed by default.

When comparing photon tracking methodologies, all codes track a single photon per thread, with the exception of UMC-PET, which tracks the two photons coming from the same decay per thread, a difference of minor relevance. GGEMS-PET employes a voxel-wise approach in its photon tracking kernel, while the other codes implement the Woodcock algorithm. The Woodcock algorithm enables faster simulations, regardless of the number of voxels employed to describe the phantom. Both GGEMS-PET and gPET use a specific photon transport module to track inside the detectors, once the photons reach the front face of the detector. On the other hand, UMC-PET does not differentiate between phantom, detectors, and other scanner materials (bed, shieldings); all objects are defined within a unified volume. This design choice provides greater flexibility, particularly in experiments where a PET insert is situated inside the field of view (FOV) in proximity to a region of interest (as proposed by Qi *et al* (2011), Grkovski *et al* (2015)). Furthermore, this framework is easily adaptable to various scanner geometries. Spherical geometries are gaining relevance among brain dedicated PET scanners (Catana 2019, Yoshida *et al* 2020) and small animal scanners (Perez-Benito *et al* 2018) due to their enhanced sensitivity. In Galve *et al* (2020b), simulation results of a spherical brain dedicated PET scanner with the UMC-PET simulator were presented. Either gPET or GGEMS-PET are adaptable for non-cylindrical geometries as well, but other simulators such as PeneloPET only allow for relatively simpler block detector configurations.

When addressing the physics aspects, all these codes employ cross-sections (or attenuation coefficients) derived from physics databases that have been thoroughly validated with experimental data. In the case of UMC-PET, the scattering angles and energies are randomly selected from a precomputed sample using PENELOPE. Typically, the samples from different materials at different energies will be mixed, preventing poor statistics while reducing the need for calculations on-the-fly. All simulators accounted for photoelectric effect, Compton scatter, and Rayleigh scatter, with the exception of GGEMS-PET, which did not consider Rayleigh scatter. Positron range and non-collinearity are two significant factors in PET imaging. Positron tracks are not simulated in any of the packages, but other solutions are implemented. In GGEMS-PET, neither of these factors is clearly described, and the original version of GGEMS did not incorporate positron transport, leading us to infer that the code does not address these aspects. In gPET, the positrons are simulated using the energy distribution from GATE (Jan *et al* 2004) and the semi-analytical method developed by Harrison *et al* (1999). In the case of MCGPU-PET, positron range estimation is not included, but the authors recommend utilizing penEasy (Sempau *et al* 2011) externally to convert the emission map into an annihilation map before simulation. In UMC-PET, we incorporated a blurring kernel based on the parametric formulation developed by Cal-González *et al* (2015). Non-colinearity is modeled using a Gaussian FWHM of 0.5 degrees in gPET, MCGPU-PET and UMC-PET. It is worth mentioning that MCGPU-PET is the only one that allows additional gamma emissions

from non-pure beta emitters. This feature was successfully utilized to eliminate spurious background signals from triple coincidences by Pratt *et al* (2023). It is noteworthy that several other physical processes are not accounted for in any of the GPU-optimized PET simulators, such as the lutetium activity within detectors, the impact of magnetic fields on positron range, or the simulation of optical photons within the detectors.

One of the primary concerns in GPU computing lies in efficiently handling time evolution calculations in parallel. gPET employs sorting routines for individual pulses prior to coincidence sorting. In contrast, GGEMS-PET delegates this aspect to the CPU, where a global timestamp is added to each decay following particle tracking computations. MCGPU-PET, on the other hand, estimates timestamps independently for each voxel in the emission map in parallel. Consequently, the resulting list of single events is not sorted in time when coming out of the GPU. Given that one of the key goals of UMC-PET was to facilitate image reconstruction through rapid simulations (e.g. scatter correction estimation, as discussed in Galve *et al* (2022), or the implementation of the simulator in the projection step of the reconstruction, as presented in Galve *et al* (2021)), we decided to rely the sorting step of the code outside of the GPU. This choice avoids sorting coincidences in the GPU. Even though coincidence sorting is possible in the GPU, it is more efficiently performed in one or more CPU cores, working asynchronously with the GPU. Thus, in UMC-PET a list of hit events or single events, stamped with the decay index, are computed in the GPU and passed by to the CPU for further processing. Post-processing with these external codes will assign a global timestamp for each decay and address any timing-related effects (e.g. dead-time, pile-up, time coincidence windows and random events). This approach enables the use of the same event list for various activity rates or electronics modeling, largely speeding up the estimation of singles and coincidences rates, random coincidences, pile-up events, and other phenomena. As said before, one or more cores of the CPU can be used to post-process a bunch of decays after photon transport in the GPU, while the GPU is computing the next bunch, in this way hiding this post-processing computation time. Lai *et al* (2019) reported a 50%–50% computation time split between photon tracking and digitizer routines in gPET, with each routine taking 0.4 s per million decays. In the examples presented in table 4, the photon tracker in UMC-PET accounted for 80% to 85% of the total simulation time (in this case the CPU post-processing is performed sequentially to the GPU task), supporting the approach to keep these post-processing calculations in the CPU. Time evolution and other post-processing routines required for the typical situations, will be bundled with UMC-PET.

We have observed a relative slowdown of UMC-PET performance compared with PeneloPET for the larger physical volumes simulated, i.e. when we move from preclinical to clinical scanners, due to the additional Woodcock steps needed to escape these larger simulated volumes. Since one of the current fields of interest in PET is large axial FOV (LAFOV) scanners (over 1 m in length), often referred to as total-body scanners (Filippi *et al* 2022), we conducted a brief assessment of simulation speed for a Quadra Vision Biograph scanner (Prenosil *et al* 2022), around 4 times larger in the axial direction than the Biograph mMR, simulating an equivalent NEMA phantom as shown in figure 5. The UMC-PET took 28.3 s to simulate the same number of decays in the Quadra, while requiring 15.5 s in the Biograph mMR. That is, 1.82 times slower for the Quadra Vision Biograph, which aligns with the expected performance degradation resulting from the larger physical volume. This is just a modest penalty which makes UMC-PET truly applicable for LAFOV simulations.

Some authors discussed the need of optimization schemes to avoid thread divergence in the GPU due to different particle types in every single thread (Hissoiny *et al* 2011, Jia *et al* 2011, Li *et al* 2022). In our case, we are simulating just one particle type (gamma photons), but we could find divergence caused by different fates and consequently different lifetimes of photons in each thread. While it may be of interest to study the impact of thread divergence in UMC-PET performance and to develop ways to reduce it, the truth is that modern GPUs are increasingly tolerant to thread divergence and further, even medium size GPUs outperform a 16-core CPU by two-orders of magnitude, and this performance gap increases every year. This suggests that employing a GPU-based MC package like UMC-PET is advisable, even if further optimization is possible.

After reviewing the strengths and weaknesses of each simulator, it is important to recognize that several factors can impact their performance. These factors include choices such as whether to employ a voxel-wise or Woodcock strategy in the photon tracker, the computational requirements imposed by distinct physical models, or the way in which the singles and coincidences are processed on the CPU or GPU. Regarding the photon tracker, which constitutes the most computationally intensive piece, we expect similar performance in gPET, MCGPU-PET, and UMC-PET, as they share similar algorithms. As for GGEMS-PET, it is likely that its performance depends largely on the number of voxels employed in the phantom. When comparing the results presented in table 3 with those reported by Ma *et al* (2020) (1.5×10^6 decays/s) and Lai *et al* (2019) (ranging from 4.0×10^5 to 5.8×10^5 decays/s, depending on the specific case), we should acknowledge that drawing definitive conclusions about whether UMC-PET is faster than other GPU-based simulators is not feasible due to variations in GPU hardware configurations and simulation scenarios across different studies. Conducting a comprehensive comparison of equivalent simulations in the same GPUs among these simulators is beyond the scope of this work. We would like to emphasize that any of the mentioned packages, including UMC-PET, is

bound to be much faster than similar calculations, even on a cluster of multi-core CPUs. Furthermore, any relatively minor differences in speed expected among the existing packages are by far less relevant than the substantial differences in their capabilities, ease of use, and realistic implementation of typical scenarios.

The accuracy of UMC-PET was validated against PeneloPET in section 3.1. The results shown in figure 8 and table 2 show good agreement between both simulators. The relative errors in the measured NEMA SF, which represents the actual distribution of events in the scanner, were below the ~3% in all the cases. We assume that the discrepancies in the simulators performance is caused by differences in the code scheme, like the voxelized representation of the volumes in UMC-PET against the parametrized volumes of PeneloPET, or different implementations of the detector response. Image assessment was shown in table 5, comparing the resolution obtained in real point sources measurements against simulated ones. In section 3.3, we validated the NEMA sensitivity simulated in UMC-PET for the Biograph mMR scanner against the reported values by Delso *et al* (2011). A better description of the patient table could help to improve these measurements and it is not a limiting factor for UMC-PET, but the authors did not have access to the real dimensions.

UMC-PET parallel implementation in the GPU clearly outperformed PeneloPET, obtaining a speed-up factor of ~2000× in the 6R-SuperArgus simulations and ~1500× in the mMR simulations. It is important to mention that PeneloPET (and any CPU-based simulator in general) can be easily parallelyzed in multiple CPU cores, achieving an acceleration equivalent to the number of cores (common computers use 4 or 8 cores, thus the acceleration gain thanks to the GPU usage is still very remarkable).

5. Conclusions

We have presented the UMC-PET simulator, a fast, versatile and accurate Monte Carlo code for PET simulation with GPU. UMC-PET has been developed with a primary focus on enhancing image reconstruction, including scatter correction and SRM estimation, as well as supporting the design of PET scanners. Its approach to define scanners and detectors in a voxelized manner simplifies its application and makes it possible to consider various geometries and the majority of detector configurations currently in use or planned for. The incorporation of precomputed tables for scattered photons and attenuation coefficients, along with a factorized scheme for the main physic principles underlying the technique, has simplified the code without adversely affecting simulation accuracy when compared to PeneloPET. The authors are working on the development of a platform to run the software on demand, to collect feedback before making the source code available.

Acknowledgments

We acknowledge support from Spanish Governement and Next GenerationEU Recovery and Resilience Facility (RRF): PID2021-126998OB-I00, CPP 2021-008751 NEWMBI; European Union as part of the European Innovation Council's Pathfinder Open Programme: RETIMAGER, 101099096; Comunidad de Madrid: S2022/BMD-7434-ASAP-CM; Instituto de Salud Carlos III: PT20/00044; Spanish Government and Next GenerationEU/PRTR: DI2M: Smart detectors for molecular imaging (2021/C005/00147498); Complutense University of Madrid funded by the European Union-Next-GenerationUE: CT18/22 Margarita Salas Fellowship.

Data availability statement

The data cannot be made publicly available upon publication because no suitable repository exists for hosting data in this field of study. The data that support the findings of this study are available upon reasonable request from the authors.

ORCID iDs

Pablo Galve https://orcid.org/0000-0003-4507-898X
Fernando Arias-Valcayo https://orcid.org/0000-0002-3814-0207
Amaia Villa-Abaunza https://orcid.org/0000-0003-1479-5230
Paula Ibáñez https://orcid.org/0000-0001-8612-5303
José Manuel Udías https://orcid.org/0000-0003-3714-764X

References

Abushab K M, Herraiz J L, Vicente E, Cal-Gonzalez J, Espana S, Vaquero J J, Jakoby B W and Udias J M 2016 Evaluation of PeneloPET simulations of biograph PET/CT scanners *IEEE Trans. Nucl. Sci.* **63** 1367–74

Aklan B, Jakoby B W, Watson C C, Braun H, Ritt P and Quick H H 2015 GATE Monte Carlo simulations for variations of an integrated pet/mr hybrid imaging system based on the biograph mMR model *Phys. Med. Biol.* **60** 4731–52

Alerstam E, Andersson-Engels S and Svensson T 2008a White Monte Carlo for time-resolved photon migration *J. Biomed. Opt.* **13** 041304

Alerstam E, Svensson T and Andersson-Engels S 2008b Parallel computing with graphics processing units for high-speed Monte Carlo simulation of photon migration *J. Biomed. Opt.* **13** 060504

Allison J *et al* 2016 Recent developments in Geant4 *Nucl. Instrum. Methods Phys. Res.*A **835** 186–225

Anger H O 1969 Scintillation camera and multiplane tomographic scanner, Lawrence Berkeley National Laboratory, LBNL Report #: UCRL-18705. Retrieved from https://escholarship.org/uc/item/4j65h22h

Arias-Valcayo F, Galve P, Herraiz J L, Desco M, Vaquero J J and Udías J M 2023 Reconstruction of multi-animal PET acquisitions with anisotropically variant PSF *Biomed. Phys. Eng. Express* **9** 065018

Badal A and Badano A 2009 Accelerating Monte Carlo simulations of photon transport in a voxelized geometry using a massively parallel graphics processing unit *Med. Phys.* **36** 4878–80

Badal A and Badano A 2012 A GPU-optimized binary space partition structure to accelerate the Monte Carlo simulation of CT projections of voxelized patient models with metal implants *2012 9th IEEE Int. Symp. on Biomedical Imaging(IEEE ISBI)* (IEEE) pp 634–7

Badal A, Sharma D, Graff C G, Zeng R and Badano A 2021 Mammography and breast tomosynthesis simulator for virtual clinical trials *Comput. Phys. Commun.* **261** 107779

Barret O, Carpenter T A, Clark J C, Ansorge R E and Fryer T D 2005 Monte Carlo simulation and scatter correction of the GE Advance PET scanner with SimSET and Geant4 *Phys. Med. Biol.* **50** 4823–40

Bauert J, Tessonnier T, Debus J and Parodi K 2019 Offline imaging of positron emitters induced by therapeutic helium, carbon and oxygen ion beams with a full-ring PET/CT scanner: experiments in reference targets *Phys. Med. Biol.* **64** 225016

Behlouli A, Visvikis D and Bert J 2018 Improved Woodcock tracking on Monte Carlo simulations for medical applications *Phys. Med. Biol.* **63** 225005

Bert J, Perez-Ponce H, Bitar Z E, Jan S, Boursier Y, Vintache D, Bonissent A, Morel C, Brasse D and Visvikis D 2013 Geant4-based Monte Carlo simulations on GPU for medical applications *Phys. Med. Biol.* **58** 5593–611

Biggs F and Lighthill R 1988 *Analytical Approximations for X-Ray Cross Sections III* (National Technical Information Service)

Burger C, Goerres G, Schoenes S, Buck A, Lonn A and von Schulthess G 2002 PET attenuation coefficients from CT images: experimental evaluation of the transformation of CT into PET 511 keV attenuation coefficients *Eur. J. Nucl. Med. Mol. Imaging* **29** 922–7

Buvat I and Lazaro D 2006 Monte Carlo simulations in emission tomography and GATE: An overview *Nucl. Instrum. Methods Phys. Res.*A **569** 323–9

Cal-González J, Herraiz J L, España S, Corzo P M G, Vaquero J J, Desco M and Udias J M 2013 Positron range estimations with PeneloPET *Phys. Med. Biol.* **58** 5127–52

Cal-González J, Pérez-Liva M, Herraiz J L, Vaquero J J, Desco M and Udías J M 2015 Tissue-dependent and spatially-variant positron range correction in 3D PET *IEEE Trans. Med. Imaging* **34** 2394–403

Cal-Gonzalez J, Vaquero J J, Herraiz J L, Pérez-Liva M, Soto-Montenegro M L, Peña-Zalbidea S, Desco M and Udías J M 2018 Improving PET quantification of small animal [^{68}Ga]DOTA-Labeled PET/CT studies by using a ct-based positron range correction *Mol. Imaging Biol.* **20** 584–93

Carter L L, Cashwell E D and Taylor W M 1972 Monte Carlo sampling with continuously varying cross sections along flight paths *Nucl. Sci. Eng.* **48** 403–11

Castiglioni I, Cremonesi O, Gilardi M C, Bettinardi V, Rizzo G, Savi A, Bellotti E and Fazio F 1999 Scatter correction techniques in 3D PET: a Monte Carlo evaluation *IEEE Trans. Nucl. Sci.* **46** 2053–8

Catana C 2019 Development of dedicated brain pet imaging devices: recent advances and future perspectives *J. Nucl. Med.* **60** 1044–52

Chen C-L, Wang Y, Lee J J S and Tsui B M W 2008 Integration of SimSET photon history generator in GATE for efficient Monte Carlo simulations of pinhole SPECT *Med. Phys.* **35** 3278–84

Chi Y, Tian Z and Jia X 2016 Modeling parameterized geometry in GPU-based Monte Carlo particle transport simulation for radiotherapy *Phys. Med. Biol.* **61** 5851–67

Choi H J, Jang J W, Shin W G, Park H, Incerti S and Min C H 2020 Development of integrated prompt gamma imaging and positron emission tomography system for in vivo 3-D dose verification: A Monte Carlo study *Phys. Med. Biol.* **65** 105005

Delso G, Furst S, Jakoby B, Ladebeck R, Ganter C, Nekolla S G, Schwaiger M and Ziegler S I 2011 Performance measurements of the siemens mmr integrated whole-body PET/MR scanner *J. Nucl. Med.* **52** 1914–22

Delso G, Martinez M-J, Torres I, Ladebeck R, Michel C, Nekolla S and Ziegler S I 2009 Monte Carlo simulations of the count rate performance of a clinical whole-body MR/PET scanner *Med. Phys.* **36** 4126–35

Du Y, Frey E C, Wang W T, Tocharoenchai C, Baird W H and Tsui B M W 2002 Combination of MCNP and SimSET for Monte Carlo simulation of SPECT with medium- and high-energy photons *IEEE Trans. Nucl. Sci.* **49 I** 668–74

España S, Herraiz J L, Vicente E, Vaquero J J, Desco M and Udias J M 2009 PeneloPET, a Monte Carlo PET simulation tool based on PENELOPE: features and validation *Phys. Med. Biol.* **54** 1723–42

Filippi L, Dimitrakopoulou-Strauss A, Evangelista L and Schillaci O 2022 Long axial field-of-view PET/CT devices: are we ready for the technological revolution? *Expert Rev. of Med. Devices* **00** 1–5

Galve P, Arias-Valcayo F, Izquierdo-Garcia D and Udias J M 2022 Accurate and practical Scatter Correction in PET with Ultra fast Monte Carlo *2022 IEEE Nuclear Science Symp. and Medical Imaging Conf.*

Galve P, Arias-Valcayo F, Lopez-Montes A, Villa-Abaunza A, Ibanez P, Herraiz J L and Udias J M 2020a Multi-purpose Ultra-fast Monte Carlo PET simulator *2020 IEEE Nuclear Science Symp. and Medical Imaging Conf. (NSS/MIC)*

Galve P, Arias-Valcayo F, Lopez-Montes A, Villa-Abaunza A, Ibanez P, Herraiz J L and Udias J M 2021 Ultra-fast Monte Carlo PET Reconstructor *16th Virtual Int. Meeting on Fully 3D Image Reconstruction in Radiology and Nuclear Medicine* pp 152–6

Galve P, Herraiz J L, Catana C and Udias J M 2020b GPU based fast and flexible iterative reconstructions of arbitrary and complex PET scanners: application to next generation dedicated brain scanners *2020 IEEE Nuclear Science Symp. and Medical Imaging Conf. Proc. (NSS/MIC)*

Gillam J E and Rafecas M 2016 Monte-Carlo simulations and image reconstruction for novel imaging scenarios in emission tomography *Nucl. Instrum. Methods Phys. Res.*A **809** 76–88

Grevillot L *et al* 2020 Technical Note: GATE-RTion: a GATE/Geant4 release for clinical applications in scanned ion beam therapy *Med. Phys.* **47** 3675–81
Grkovski M *et al* 2015 Evaluation of a high resolution silicon PET insert module *Nucl. Instrum. Methods Phys. Res.*A **788** 86–94
Harrison R L, Kaplan M S, Vannoy S D and Lewellen T K 1999 Positron range and coincidence non-collinearity in SimSET *1999 IEEE Nuclear Science Symp. Conf. Record. 1999 Nuclear Science Symp. and Medical Imaging Conf. (Cat. No.99CH37019). IEEE* vol 3 1265–8
Herraiz J L, España S, Vaquero J J, Desco M and Udías J M 2006 FIRST: Fast Iterative Reconstruction Software for (PET) tomography *Phys. Med. Biol.* **51** 4547–65
Hissoiny S, Ozell B, Bouchard H and Després P 2011 GPUMCD: a new GPU-oriented Monte Carlo dose calculation platform *Med. Phys.* **38** 754–64
Ibáñez P, Villa-Abaunza A, Vidal M, Guerra P, Graullera S, Illana C and Udías J M 2021 XIORT-MC: a real-time MC-based dose computation tool for low- energy X-rays intraoperative radiation therapy *Med. Phys.* **48** 8089–106
Iriarte A, Marabini R, Matej S, Sorzano C O S and Lewitt R M 2016 System models for PET statistical iterative reconstruction: a review *Comput. Med. Imaging Graph.* **48** 30–48
Jahnke L, Fleckenstein J, Wenz F and Hesser J 2012 GMC: a GPU implementation of a Monte Carlo dose calculation based on Geant4 *Phys. Med. Biol.* **57** 1217–29
James F 1990 A review of pseudorandom number generators *Comput. Phys. Commun.* **60** 329–44
Jan S, Frisson T and Sarrut D 2013 GATE simulation of ^{12}C hadrontherapy treatment combined with a pet imaging system for dose monitoring: a feasibility study *IEEE Trans. Nucl. Sci.* **60** 423–9
Jan S *et al* 2004 GATE: a simulation toolkit for PET and SPECT *Phys. Med. Biol.* **49** 4543–61
Jan S *et al* 2011 GATE V6: a major enhancement of the GATE simulation platform enabling modelling of CT and radiotherapy *Phys. Med. Biol.* **56** 881–901
Jia X, Gu X, Graves Y J, Folkerts M and Jiang S B 2011 GPU-based fast Monte Carlo simulation for radiotherapy dose calculation *Phys. Med. Biol.* **56** 7017–31
Jia X, Gu X, Sempau J, Choi D, Majumdar A and Jiang S B 2010 Development of a GPU-based Monte Carlo dose calculation code for coupled electron–photon transport *Phys. Med. Biol.* **55** 3077–86
Jia X, Schümann J, Paganetti H and Jiang S B 2012 GPU-based fast Monte Carlo dose calculation for proton therapy *Phys. Med. Biol.* **57** 7783–97
Kraan A C 2015 Range verification methods in particle therapy: underlying physics and monte carlo modeling *Front. Oncol.* **5** 1–27
Kraus R, Delso G and Ziegler S I 2012 Simulation study of tissue-specific positron range correction for the new biograph mMR whole-body PET/MR system *IEEE Trans. Nucl. Sci.* **59** 1900–9
Lai Y, Zhong Y, Chalise A, Shao Y, Jin M, Jia X and Chi Y 2019 gPET: a GPU-based, accurate and efficient Monte Carlo simulation tool for PET *Phys. Med. Biol.* **64** 245002
Lee M S, Kang S K and Lee J S 2018 Novel inter-crystal scattering event identification method for PET detectors *Phys. Med. Biol.* **63** 115015
Li Y, Sun W, Liu H, Ding S, Wang B, Huang X and Song T 2022 Development of a GPU-superposition Monte Carlo code for fast dose calculation in magnetic fields *Phys. Med. Biol.* **67** 125002
Lin H-H, Chuang K-S, Lin Y-H, Ni Y-C, Wu J and Jan M-L 2014 Efficient simulation of voxelized phantom in GATE with embedded SimSET multiple photon history generator *Phys. Med. Biol.* **59** 6231–50
Lippuner J and Elbakri I A 2011 A GPU implementation of EGSnrc's Monte Carlo photon transport for imaging applications *Phys. Med. Biol.* **56** 7145–62
Lopez-Montes A, Herraiz J L, Galve P, Espana S, Vicente E, Cal-Gonzalez J and Udias J M 2019 PeneloPET v3.0, an improved multiplatform PET Simulator *2019 IEEE Nuclear Science Symp. and Medical Imaging Conf. (NSS/MIC)* (IEEE) pp 1–3
Ma B *et al* 2020 Scatter correction based on gpu-accelerated full monte carlo simulation for brain PET/MRI *IEEE Trans. Med. Imaging* **39** 140–51
Masuda T, Nishio T, Sano A and Karasawa K 2020 Extension of the ML-EM algorithm for dose estimation using PET in proton therapy: application to an inhomogeneous target *Phys. Med. Biol.* **65** 185001
Mohammadi A, Inadama N, Yoshida E, Nishikido F, Shimizu K and Yamaya T 2017 Improvement of crystal identification performance for a four-layer composed of crystals segmented by laser processing *Nucl. Instrum. Methods Phys. Res.* A **866** 29–35
National Electrical Manufacturers Association 2007 *NEMA Standards Publication NU 2-2007 Performance Measurements of Positron Emission Tomographs* (National Electrical Manufacturers Association)
National Electrical Manufacturers Association 2008 *NEMA Standards Publication NU 4-2008 Performance Measurements of Small Animal Positron Emission Tomographs* (National Electrical Manufacturers Association)
NEA 2019 *PENELOPE 2018: A Code System for Monte Carlo Simulation of Electron and Photon Transport* (Paris: OECD publishing)
Onecha V V, Galve P, Ibáñez P, Freijo C, Arias-Valcayo F, Sanchez-Parcerisa D, España S, Fraile L M and Udías J M 2022 Dictionary-based software for proton dose reconstruction and submilimetric range verification *Phys. Med. Biol.* **67** 045002
Paredes-Pacheco J, López-González F J, Silva-Rodríguez J, Efthimiou N, Niñerola-Baizán A, Ruibal Á, Roé-Vellvé N and Aguiar P 2021 SimPET—An open online platform for the Monte Carlo simulation of realistic brain PET data. Validation for 18 F-FDG scans *Med. Phys.* **48** 2482–2493
Peng P, Judenhofer M S, Jones A Q and Cherry S R 2018 Compton PET: a simulation study for a PET module with novel geometry and machine learning for position decoding *Biomed. Phys. Eng. Express* **5** 015018
Perez-Benito D, Chil R, Udias J M, Desco M and Vaquero J J 2018 SiPM-based PET detector module for a 4π span scanner *Nucl. Instrum. Methods Phys. Res.*A **936** 18–21
Poon J K, Dahlbom M L, Casey M E, Qi J, Cherry S R and Badawi R D 2015 Validation of the SimSET simulation package for modeling the Siemens Biograph mCT PET scanner *Phys. Med. Biol.* **60** N35–45
Popota F D, Aguiar P, España S, Lois C, Udias J M, Ros D, Pavia J and Gispert J D 2015 Monte Carlo simulations versus experimental measurements in a small animal PET system. A comparison in the NEMA NU 4-2008 framework *Phys. Med. Biol.* **60** 151–62
Pratt E C *et al* 2023 Simultaneous quantitative imaging of two PET radiotracers via the detection of positron-electron annihilation and prompt gamma emissions *Nat. Biomed. Eng.* (https://doi.org/10.1038/s41551-023-01060-y)
Prenosil G A, Sari H, Fürstner M, Afshar-Oromieh A, Shi K, Rominger A and Hentschel M 2022 Performance characteristics of the biograph vision quadra pet/ct system with a long Axial Field of View Using the NEMA NU 2-2018 standard *J. Nucl. Med.* **63** 476–84
Qi J, Yang Y, Zhou J, Wu Y and Cherry S R 2011 Experimental assessment of resolution improvement of a zoom-in PET *Phys. Med. Biol.* **56** N165–74
Rogers D W O 2006 Fifty years of Monte Carlo simulations for medical physics *Phys. Med. Biol.* **51** R287–301
Salvat F 1987 Algorithms for random sampling from single-variate distributions *Comput. Phys. Commun.* **46** 427–36

Sarrut D *et al* 2021 Advanced Monte Carlo simulations of emission tomography imaging systems with GATE *Phys. Med. Biol.* **66** 10TR03
Sarrut D *et al* 2022 The OpenGATE ecosystem for Monte Carlo simulation in medical physics *Phys. Med. Biol.* **67** 184001
Sempau J, Badal A and Brualla L 2011 A PENELOPE -based system for the automated Monte Carlo simulation of clinacs and voxelized geometries-application to far-from-axis fields *Med. Phys.* **38** 5887–95
Sempau J, Wilderman S J and Bielajew A F 2000 DPM, a fast, accurate Monte Carlo code optimized for photon and electron radiotherapy treatment planning dose calculations *Phys. Med. Biol.* **45** 2263–91
Shibuya K, Yoshida E, Nishikido F, Suzuki T, Tsuda T, Inadama N, Yamaya T and Murayama H 2007 Annihilation photon acollinearity in PET: volunteer and phantom FDG studies *Phys. Med. Biol.* **52** 5249–61
Vicente E, Herraiz J L, Canadas M, Cal-Gonzalez J, Espana S, Desco M, Vaquero J J and Udias J M 2010 Validation of NEMA NU4-2008 scatter fraction estimation with 18F and 68Ga for the ARGUS smallanimal PET scanner *IEEE Nuclear Science Symp. Medical Imaging Conf.(IEEE NSS MIC)* (IEEE) pp 3553–7
Wang Y, Seidel J, Tsui B M W, Vaquero J J and Pomper M G 2006 Performance evaluation of the GE healthcare eXplore VISTA dual-ring small-animal PET scanner *J. Nucl. Med.* **47** 1891–900
Wei S and Vaska P 2020 Evaluation of quantitative, efficient image reconstruction for VersaPET, a compact PET system *Med. Phys.* **47** 2852–68
Woodcock E, Murphy T, Hemmings P and Longworth S 1965 Techniques used in the GEM code for Monte Carlo neutronics calculation *Proc. Conf. Applications of Computing Methods to Reactors, ANL-7050* pp 557–79
Yoshida E, Tashima H, Akamatsu G, Iwao Y, Takahashi M, Yamashita T and Yamaya T 2020 245 ps-TOF brain-dedicated PET prototype with a hemispherical detector arrangement *Phys. Med. Biol.* **65** 145008