Efficient photon transport in positron emission tomography simulations using VMC++

Iwan Kawrakow\textsuperscript{1}, Krassimir Mitev\textsuperscript{2}, Georgi Gerganov\textsuperscript{2}, Jordan Madzhunkov\textsuperscript{2} and Assen Kirov\textsuperscript{3}

\textsuperscript{1} Ionizing Radiation Standards, NRCC, Ottawa, K1A 0R6, Canada
\textsuperscript{2} Faculty of Physics, Sofia University “St. Kliment Ohridski”, 1164 Sofia, Bulgaria
\textsuperscript{3} Department of Medical Physics, MSKCC, New York, NY 10021, USA

E-mail: iwan@irs.phy.nrc.ca

Abstract. vmcPET, a VMC\textsuperscript{++} based fast code for simulating photon transport through the patient geometry for use in positron emission tomography related calculations, is presented. vmcPET is shown to be between 250 and 425 times faster than GATE in completely analog mode and up to 50000 times faster when using advanced variance reduction techniques. Excellent agreement between vmcPET and EGSnrc and GATE benchmarks is found. vmcPET is coupled to GATE via phase-space files of particles emerging from the patient geometry.

1. Introduction
The Monte Carlo (MC) simulation of positron emission tomography (PET) is an area of active research [1]. Most early studies in the area were performed using specialized packages such as SimSET [2] and SIMIND [3]. Since its initial release in 2004, the Geant4 Application for Tomographic Emission (GATE) simulation package [4] has been continuously gaining in popularity and now has more than 400 members subscribed to the GATE mailing lists and discussion forums [1]. GATE, based on the Geant4 toolkit [5], is a powerful package permitting extremely detailed models of the scanning system and its detector response. The main drawback of GATE is its slow simulation speed. In a typical GATE simulation of a PET image acquisition most of the simulation time is spent in the voxelized phantom representing the patient geometry. For this part of the calculation GATE requires more than 1 hour of CPU time on a modern PC for one million decays. Considering that 15 mCi (~ 550 million decays per second) is a typical total activity in PET, it is clear that realistic PET simulations with GATE are only possible on a supercomputer with thousands of CPUs.

The purpose of this article is to present vmcPET, a fast MC tool for simulating photon transport in a voxelized patient geometry for use in PET calculations. vmcPET is based on VMC\textsuperscript{++}[6], a MC package primarily used in radiation treatment planning. The paper is organized as follows. In the next section a brief discussion of the necessary level of sophistication in modeling the underlying scattering and transport processes is given. Section 3 discusses analog simulations and simulations using simple variance reduction techniques (VRT). Section 4 is devoted to more advanced VRT. The coupling of vmcPET to GATE via phase space files and a few verification results are discussed in section 5. Summary and outlook is presented in section 6.
2. The physics of photon transport in the patient

No positron transport is considered in this investigation, i.e. the simulation begins with two back-to-back 511 keV photons. The inclusion of $e^+$ transport without a significant penalty in terms of simulation time is left for future publications. Because we are concerned with the transport in the patient geometry only, i.e. with particle transport in low-Z materials in the sub-511 keV energy range, bremsstrahlung emission by electrons set in motion by the interacting photons is completely negligible. Hence, the entire simulation can be done excluding electron transport and in fact, this is how emission tomography calculations have always been performed. Characteristic X-rays produced in low-Z materials have sub-1 keV energies. Hence, they don’t need to be transported and this makes the simulation of photo-absorption trivial: the photon simply disappears. For most of the energy range incoherent (Compton) scattering is the dominant process. Compared to Klein-Nishina (KN), binding corrections decrease the total cross section in water by $\sim 0.1\%$ at 511 keV and by $\sim 1.3\%$ at 100 keV. Coherent (Rayleigh scattering) has very little impact at 511 keV (a 0.2% increase of the total cross section in water), gradually increasing in importance at lower energies (2.8% increase of the total cross section at 100 keV). Given that the probability of detection within the energy window decreases with decreasing photon energy, one can surmise that ignoring Rayleigh scattering and modeling Compton interactions with the KN equation will produce adequate results. To test this hypothesis, a simple EGSnrc [7] user code was written that utilizes the EGSnrc C++ class library [8]. A point source emitting two back-to-back 511 keV photons is placed in the center of a $30 \times 30 \times 30$ cm$^3$ water box. The detecting surface is assumed to be a cylinder with radius 46.35 cm, height of 15.204 cm, divided into 672 x 18 crystals (radially x axially). Two simple detecting crystals are implemented: (i) an ideal detector, i.e. every photon generates a count equal to its energy (ii) a detector with a Gaussian energy response ($\sigma = 100$ keV) and 100% efficiency. In both cases the energy window is set to be between 350 and 550 keV. All parameters listed above are typical in PET. Two separate simulations are run, one including Rayleigh scattering and binding corrections plus Doppler broadening in Compton scattering, and one with no Rayleigh scattering and with Compton scattering modeled using the KN equation. Scattered photons are marked during the simulation so that the primary and scatter coincidences, as well as coincidences involving at least one Rayleigh interaction, could be calculated. Table 1 summarizes the results. It is evident that the total number of coincidences agrees within 0.1% or better. The $\sim 0.5\%$ difference in the total number of scatter coincidences is irrelevant for any practical application because it is comparable to the noise in measured sinograms for typical PET scans.

3. vmcPET: analog simulations and simple VRTs

The preceding discussion shows that the problem of the MC simulation of photon transport in the patient geometry for use in PET calculations is exceedingly simple: one has to deal with KN
scattering and the tracking of particles in a trivial geometry (rectilinear voxels). Thus, there are two equally viable options for tackling the problem:

- Develop a tool from scratch. This is the path most often taken in the early years of MC emission tomography simulations [1] and also in some more recent developments [9].
- Use an existing general purpose MC system but develop specialized photon transport methods to take full advantage of the simplicity of the physics. Although the use of general purpose codes has become increasingly more popular in recent years [1], GATE being a prime example, specialized photon transport algorithms have not been reported so far.

The second option is pursued in this investigation. The VMC++ [6] package, although not widely known as a general purpose code, is selected as a basis due to its object oriented design, optimized sampling and tracking algorithms, and the built-in capability to use multiple simulation threads, which is very useful on modern multi-core CPUs. VMC++ is similar to EGSnrc or Geant4 in the sense that it is a toolkit, i.e. a “user code” that utilizes the VMC++ library must be written for a specific application. The resulting application, referred to in the following as vmcPET, adds classes for simulating various activity distributions, specialized photon transport algorithms that incorporate several VRT (see next section), classes for the simulation of the 2 idealized detectors described in the previous section, and classes for computing a sinogram and extracting various other bits of information out of the simulation.

To better understand the CPU time cost of the various components of the simulation, vmcPET is run without any VRT, employing the generic photon transport algorithm provided by VMC++. The simulation time for 1 million decays and a point source positioned in the middle of a 30 × 30 × 30 cm³ water phantom divided into 0.4 cm cubic voxels is found to be 8.4 seconds on a 2.66 GHz QX6700 CPU using a single thread. This is about 2.5 times faster than the EGSnrc based simulation described in the previous section, and between 240 and 425 times faster compared to GATE (using identical physics, i.e. no electron transport and Rayleigh scattering), depending on computing platform and GATE/Geant4 version. There are approximately 7 million Compton interactions in this simulation, which require about 0.5 seconds of simulation time in vmcPET. About 0.3 seconds is spent for selecting random directions for the initial particles and recording coincidences, the remaining 7.6 seconds are needed for tracking the photons through the voxelized geometry. The factor of 2.5 gain in speed compared to EGSnrc is attributed to the various algorithm optimizations in VMC++ and to the fact that the GNU C++ compiler used in this investigation appears to produce faster code than GNU Fortran compiling the core EGSnrc components. In an attempt to understand the dramatic difference in simulation speed between vmcPET and GATE, a GATE simulation with “geantinos” (fictitious particles that do not interact) replacing the 511 keV photons is run. It is found that this simulation is about 60% faster indicating that 40% of the simulation time in GATE is spent on tracking particles. It is well known that some general purpose geometry packages are very slow when modeling voxelized geometries and the Geant4 package appears to be no exception in this regard. A factor of ~ 2 in speed could therefore be gained by simply providing a better voxel geometry implementation for GATE. However, the remaining difference is difficult to understand and indicates potentially very serious implementation issues in GATE and/or Geant4.

Given that in vmcPET most of the time is spent on following photons through the geometry, Woodcock tracking, a standard VRT also known as fictitious cross section, delta ray, or null collision method, can be employed. Use of Woodcock tracking in vmcPET makes the simulation time independent of the phantom resolution and, for the 0.4 cm voxel size considered throughout this paper, provides a factor of 2 gain in simulation speed for a total factor between 480 and 850 compared to GATE.
4. vmcPET: advanced VRTs

Woodcock tracking is a “simple” VRT, i.e. it does not modify the statistical weights of the photons and does not change the variance per simulated decay. Thus, the simulation time alone describes the gain in efficiency. With more advanced variance reduction techniques, one must look at the efficiency $\varepsilon = 1/(\sigma^2 T)$ of the simulation, where $\sigma$ is a measure of the statistical uncertainty and $T$ is the simulation time. For the purposes of this investigation, $\sigma^2$ is defined as the average relative uncertainty squared of the computed sinogram for a subset of the lines of response (LORs), i.e.

$$\sigma^2(x) = \frac{1}{N} \sum_{s_i > x s_{\text{max}}} \left( \frac{\Delta s_i}{s_i} \right)^2$$

where $s_i$ is the signal in the $i$th LOR, $\Delta s_i$ its statistical uncertainty, $s_{\text{max}}$ is the maximum signal in the entire sinogram, $N$ is the number of LORs with $s_i > x s_{\text{max}}$, and $x$ is taken as 2%, 10% or 50%. Using $\sigma(50\%)$ in the definition of $\varepsilon$ will emphasize the LORs dominated by primary coincidences, whereas $\sigma(2\%)$ will be more indicative for the average uncertainty in LORs with significant scatter fraction. If the purpose of the simulation is to subtract scatter coincidences from the measured sinogram before image reconstruction, $\sigma(2\%)$ is the most relevant measure of the average uncertainty of the simulation.

There are numerous papers in the literature dealing with VRTs for emission tomography simulations, the interested reader can follow the citations in Ref. [1] or the publications related to e.g. SimSET. Despite this wealth of VRT experience, it appears that existing algorithms are comparable in speed to a fully analog vmcPET simulation, thus indicating that the optimum set of VRT has not been found and motivating a detailed investigation within the framework of vmcPET. After exploring a number of techniques, including directional biasing and stratification used in SimSET[2] and YaPRA[9] but not in the final vmcPET algorithm, a combination of Russian Roulette (RR), interaction splitting, path-length transform and forced detection was found to result in the optimum efficiency. Before describing the algorithm in detail, it is worth pointing out that, once RR and interaction splitting are involved, a variable number of particles per incident photon will arrive at the detector ring. A more sophisticated coincidence calculation is therefore required. Consider for simplicity a single decay occurring within a collection time interval that has resulted in $N_1$ photons with statistical weights $w_1^{(1)}, w_2^{(1)}, \ldots, w_{N_1}^{(1)}$ from the first 511 keV photon and $N_2$ photons with statistical weights $w_1^{(2)}, w_2^{(2)}, \ldots, w_{N_2}^{(2)}$ from the second reaching the detector ring. This situation corresponds to $N_1 N_2$ coincidences between photon $i (i = 1, \ldots, N_1)$ from list 1 and photon $j (j = 1, \ldots, N_2)$ from list 2 with statistical weights $w_{ij} = w_i^{(1)} w_j^{(2)}$. This is easily generalized to the situation with $M$ decays per time interval (only needed if computing random coincidences): now one must build all pairs of descendents from initial photons $\mu$ and $\nu, \mu = 1, \ldots, M - 1, \nu = \mu + 1, \ldots, M$, and assign statistical weights $w_{ij}^{(\mu, \nu)}$ to them,

$$w_{ij}^{(\mu, \nu)} = w_i^{(\mu)} \cdot w_j^{(\nu)} \prod_{\rho \neq \mu, \nu} w_0^{(\rho)}$$

where $w_0^{(\rho)}$ is the probability that the $\rho$'th initial photon does not reach the detector ring and where the product is over $\rho = 1, \ldots, M, \rho \neq \mu, \nu$. This coincidence counting logic is implemented in vmcPET and its correctness is checked by finding no statistically significant differences to sinograms computed in a fully analog fashion. The vmcPET algorithm then works as follows:

(i) Each “new” photon is checked if it is aimed at the detector ring (a “new” photon is both a 511 keV photon directly from the particle source or a photon just created in a photon

\footnote{A list can be found at http://depts.washington.edu/simset/html/simset\_main.html}
interaction) and if so, a copy is made to a list of photons that potentially needs to be tracked to determine the number of mean-free-paths (MFP) to the outside of the geometry.

(ii) Each “new” photon is then subjected to a RR game with survival probability $1/N_s$, with $N_s > 1$ a free parameter to be adjusted for optimum efficiency. The weight of surviving photons is increased by $N_s$.

(iii) The number of MFP $\lambda$ to the next interaction is sampled from $p(\lambda)$,

$$p(\lambda) = \frac{2\lambda_0^2}{(\lambda + \lambda_0)^3}$$

with $\lambda_0$ a free parameter to be adjusted for optimum efficiency. When a photon reaches an interaction site, its weight is modified by $\exp(-\lambda)/p(\lambda)$ to recover the actual exponential probability distribution for $\lambda$. This technique, known as path-length transform, is normally applied in deep penetration problems together with an exponential stretching of path-lengths. Although the form given in Eq. (3) is different, its purpose is the same: for $\lambda_0 > 1$ it leads to photons traveling longer distances between interactions and therefore to more interactions near the phantom surface.

(iv) Photons are transported to the interaction site using Woodcock tracking

(v) Each Compton interaction is split $N_s$ times, with the weights of the resulting “new” photons reduced by $1/N_s$.

(vi) When the transport of all photons has been terminated (by a RR game, photo-absorption, or by the photon leaving the geometry), the logic described above is used to check for coincidences and, if needed, the attenuation of photons participating in coincidences is computed and their weight adjusted accordingly (this is also known as forced detection).

The optimum values for the parameters $N_s$ and $\lambda_0$ depend on the quantity of interest. For the total number of coincidences, $\lambda_0 = 6$ and $N_s = 5$ results in optimum efficiency, which is about 20 times better than vmcPET with simple VRT. For the more relevant case of the average sinogram uncertainty $\sigma^2(x)$, optimum efficiency results for any $N_s \geq 20$ and $\lambda_0 \geq 1$. The efficiency gain is about a factor of 60, nearly independent on $x$. The overall speed gain compared to GATE is therefore somewhere between a factor of 29000 and 51000. A typical PET scan involves $\sim 15$ T

Table 2. CPU time in seconds to simulate 550 million decays using a modern quad-core CPU.

|       | GATE | vmcPET analog | vmcPET simple VRT | vmcPET full VRT |
|-------|------|---------------|-------------------|-----------------|
| mCi   |      |               |                   |                 |
| 2.5 $\times$ 10^6 | 825  | 412           | 6.9               |                 |

mCi total activity, i.e. about 550 million decays per second. Table 2 summarizes the CPU time in seconds needed by GATE and vmcPET to simulate these 550 million decays on a single quad-core CPU running at 2.66 GHz using 4 threads (or four processes in the case of GATE). Thus, only a modest cluster of 7 nodes would be needed to simulate all decays in real time using vmcPET with the VRT described in this section.

5. Verification and coupling to GATE

All comparisons between the total and scatter coincidence count computed with vmcPET and with the simple EGSnrc user code described in section 2 showed agreement within statistics (in all cases 0.1% or better). This is not surprising considering that exactly the same physics is used
by both codes (KN scattering and photo-absorption with total cross sections from the XCOM tabulations).

The implementation of coincidence computation in GATE directly resembles a real PET scanner. Thus, GATE can not currently handle non-analog simulations. The vmcPET to GATE coupling is therefore based on vmcPET with Woodcock tracking only and uses phase-space files containing all particles emerging from the patient geometry written by vmcPET. A new source is added to GATE, which returns particles on a specified surface. This source reads a vmcPET phase space file and adds timing information based on the total activity and the number of vmcPET simulated decays between subsequent particle entries in the phase space file. The particles are then passed to GATE for a simulation of the transport through the detection system. In this way the same simulation can be run using only GATE or using vmcPET for the transport in the patient geometry and GATE for the transport through the detectors, permitting an independent check of vmcPET. Comparisons performed so far show 1% or better agreement for the number of singles and coincidences, which is comparable to the statistical uncertainties (due to the much lower speed of GATE only tests with 3 million decays were run).

6. Summary

vmcPET, a VMC++ based fast code for simulating photon transport through the patient geometry for use in PET related calculations, is presented. vmcPET is 250 to 425 times faster than GATE in completely analog mode and up to 50000 times faster when using the variance reduction techniques described in section 4. At the same time vmcPET shows excellent agreement with EGSnrc and GATE benchmark calculations.

vmcPET is coupled to GATE via the use of phase-space files and a new particle source for GATE. It is planned to attempt a direct vmcPET to GATE interface in the near future.

The addition of $e^+$ transport to vmcPET would also be very valuable. Although VMC++ is known to be capable of simulating charged particle transport much faster than general purpose codes such as EGSnrc and Geant4, the inclusion of $e^+$ transport would nevertheless result in a significant increase of the overall vmcPET simulation time. It will therefore be necessary to combine $e^+$ transport with non-analog VRT such as splitting and directional biasing. Thus, the capability to handle non-analog simulations in GATE is a required pre-requisite for this addition.

Further investigations will involve the characterization of detector response, which would permit a realistic direct computation of coincidences within vmcPET.

References

[1] Buvat I and Lazaro D 2006 Nucl. Inst. Meth. A 569 323–329
[2] Lewellen T, Anson C, Haynor D, Harrison R, Bice A, Schubert S, Miyaoka R, Gillespie S and Zhu J 1988 J. Nucl. Med. 29 871
[3] Ljungberg M and Strand S 1989 Comp. Meth. Prog. in Biomedicine 29 257–272
[4] S Jan et al 2004 Phys. Med. Biol. 49 4543–4561
[5] S Agostinelli et al 2003 Nucl. Inst. Meth. A 506 250 – 303
[6] Kawrakow I 2001 Advanced Monte Carlo for Radiation Physics, Particle Transport Simulation and Applications: Proceedings of the Monte Carlo 2000 Meeting Lisbon ed Kling A, Barao F, Nakagawa M, Távora L and Vaz P (Springer, Berlin) pp 229 – 236
[7] Kawrakow I 2000 Med. Phys. 27 485 – 498
[8] Kawrakow I 2005 EGSnrc C++ class library Technical Report PIRS–898 National Research Council of Canada Ottawa, Canada
[9] Rehfeld N and Alber M 2006 Med. Phys. 33 3498–3507