Acceleration of *SimSET* Photon History Generation

Robert L. Harrison, *Member, IEEE*, Somasekhar Dhavala, *Member, IEEE,* Prasanth N. Kumar, Yiping Shao, *Member, IEEE,* Ravindra Manjeshwar, *Member, IEEE,* Tom K. Lewellen, *Member, IEEE*, Floris P. Jansen

Abstract*-- *SimSET* (a Simulation System for Emission Tomography) is widely used for studying PET and SPECT. As emission tomography simulation has become a more mature field, the scope of the research being performed, and thus the complexity of the simulations required, has grown immensely. Researchers are increasingly interested in clinically realistic simulations, and in some cases need hundreds or thousands of realizations. To meet these needs, we are investigating methods for accelerating *SimSET*. *SimSET* has always incorporated importance sampling (IS). Early studies showed the use of IS led to efficiencies 10-100 times greater than those achieved using analog (conventional) simulation. However, as the simulation became increasingly realistic the assumptions underlying the IS algorithms were violated. The efficiency improvement fell as low as a factor of two for some simulations. We are addressing this loss of efficiency by updating *SimSET's* algorithms, code optimization, and by modifying the software to run on multiple processors. We hope, with the new IS, to be able to simulate a 3D PET FDG brain scan (300 million detected events) in 3 hours on a 2 GHz processor. This would be a factor of 20 speedup over the currently distributed software. To date we achieved a factor of 1.5-3 speedup by changing three algorithms and doing some code optimization. We have several more algorithm improvements and another round of code optimization planned. We have made significant progress on parallel processing. Prototype code based on the last distributed version of *SimSET* achieved a speedup very close to the number of processors used. The new software also allows for multiple realizations of the same simulation to be automatically generated on multiple processors.

I. Introduction

SimSET (a Simulation System for Emission Tomography) is a public domain software package that is widely used in the nuclear medicine research community. It provides photon tracking through the tomograph field-of-view (FOV), collimator and detector models, and user-configurable binned output [1].

Manuscript received Nov 13, 2002. This work was supported in part by the U.S. Public Health Service under Grant No. CA42593.

R. L. Harrison and T. K. Lewellen are with the University of Washington, Seattle, WA 98195 USA (e-mail: simset@u.washington.edu).

S. Dhavala and P. N. Kumar are with General Electric Global Research Center, Bangalore, India.

Y. Shao, R. Manjeshwar, and F. P. Jansen are with General Electric Global Research Center, Niskayuna, NY.

As emission tomography simulation has become a more mature field, the scope of the research being performed, and thus the complexity of the simulations required, has grown immensely. Researchers are increasingly interested in clinically realistic simulations, and in some cases need hundreds or thousands of realizations (e.g [2]). To meet these needs, we are investigating methods for accelerating SimSET. SimSET has always incorporated importance sampling (IS). Early studies showed the use of IS led to efficiencies 10-100 times greater than those achieved using analog (conventional) simulation. However, as the simulation became increasingly realistic the assumptions underlying the IS algorithms were violated. The efficiency improvement fell as low as a factor of two for some simulations.

We are addressing this loss of efficiency by updating SimSET's algorithms, doing code optimization, and by modifying the software to run on multiple processors. We hope, with the new algorithms, to be able to simulate a 3D PET FDG brain scan (300 million detected events) in 3 hours on a 2 GHz processor. This would be a factor of 20 speedup over the currently distributed software. To date we achieved a factor of 1.5-3 speedup by changing three algorithms and doing some code optimization. We have several more algorithm improvements and another round of code optimization planned.

We have made significant progress on parallel processing. Prototype code based on the last distributed version of SimSET achieved a speedup very close to the number of processors used. The new software also allows for multiple realizations of the same simulation to be automatically generated on multiple processors.

II. Acceleration Methods

Our acceleration work has two tracks: one accelerates the underlying code and algorithms, the other adapts the code for use on multiple processors.

Table I summarizes our current status and our goals for the algorithmic improvements. The first line gives a breakdown of simulation efficiency (for the voxel_man phantom [3] in a 3D PET tomograph using our latest SimSET release (2.6.2.3). It shows, with SimSET's IS options turned on, how many decays

0-7803-7636-6/03/$17.00

SimSET processed per CPU second (decays/sec), what fraction of the decays were detected (detected/decay), and what the quality factor (QF) and computational figure-of-merit (CFOM) [4] were, where

$$CFOM = \frac{decays}{\sec} * \frac{detected}{decay} * QF \qquad (1)$$

is a measure of the efficiency of the simulation. (Simulations were performed on a 375 MHz DEC Alpha.) The second line gives the improvements we have targeted to achieve a factor of 20 speedup.

TABLE I
CURRENT AND TARGETED EFFICIENCY

	Decays/sec	Detected/Decay	QF	CFOM
Current	1529	0.176	0.306	82.3
Targeted	2400	0.9	0.75	1620

We can improve the CFOM by improving any of the three factors comprising it. There is a lot of room for improvement in all three factors. With IS on we use 230,000 clock cycles per decay. This seems excessive. We seek to increase the decays/sec through the use of supervoxels and code optimization. The QF and detected/decay are also quite low. The goal of IS is to have both these factors near 1. We are not coming close for either metric, and the stratification and forced detection improvements we are implementing are aimed at raising both.

Our goal for the parallel processing is to achieve near optimal speedup. The resulting software should be transparent to no parallel-processor users, and have minimal chance of duplicating the same photon path.

A. Supervoxels

For anthropomorphic phantoms, the most time-consuming step in the current release of SimSET is tracking photons from voxel to voxel. When tracking photons through this array, the attenuation characteristics along the photon path are currently updated and manipulated each time the photon crosses a voxel boundary. The supervoxel (SV) algorithm attempts to reduce the number of updates and manipulations necessary by grouping the voxels into larger regions of uniform attenuation, a supervoxel. Such algorithms have been previously described in the literature. The MRR algorithm (Ogawa, IEEE TNS 44:4:1521-6, 1997) defines a SV as the largest uniform-attenuation rectangular box around a voxel.. Moore (IEEE TNS 48:3:720-4, 2001) noted that this box, in a torso phantom, was often long and skinny. This resulted in most photon paths leaving the box quickly. Moore modified MRR to limit the maximum SV length to reduce this tendency.

We introduce a further modification to MMR, the greatest cube (GC) algorithm. First we find the largest uniform-attenuation cube centered at a voxel. Then we expand a rectangular SV equally in the non-blocked directions until the SV can grow no further. This algorithm approximately maximizes the number of nearby voxels in a voxel's SV.

B. Stratification

SimSET uses stratification to choose the starting position/direction of decays/photons. The main goal of stratification is to increase the detected/decay column in Table 1, so more decays/photons are started in positions/directions that are likely to be detected, or "productive". To avoid bias, each photon is then given a weight that says how many real-world photons it represents. This weight is then changed anytime IS is used to reflect the probability of the chosen path. A secondary goal of stratification is that the chosen stratification interact with the succeeding IS in such a way that the resulting events have near equal weights. This improves the quality factor, the third column in Table 1.

We have made two modifications to our stratification algorithms. One addresses the first goal above by changing how we define when a photon is productive: this is the basis for how stratification will choose decay positions and starting photon directions. The other modification addresses the second goal above by stratifying a photon's transaxial starting direction: previously we had only stratified over the axial angle and z-position.

SimSET's original stratification was based on a simple tomograph model. A target cylinder was placed around the simulated object and any photon hitting the cylinder with axial angle less than a user-specified acceptance angle was 'productive'. All other photons were discarded. Once collimator simulations were added, only the productive photons were passed from the object to the collimator. As a result, the acceptance angle must be set to 90° to see all possible scatter events with Monte Carlo collimators. As the acceptance angle is a key to our IS, this makes the IS ineffective. The photons we choose to simulate are not that likely to be detected. This is a major factor in the detected/decay ratio in Table 1 being so low for the current software.

In our new stratification algorithm, we define a photon to be productive if it leaves the collimator through the detector-facing surface. This means most productive photons are in a position to be detected.

SimSET originally only stratified the z-cosine of photon directions. Azimuthal angles were chosen uniformly. In the 20 cm cylinders originally used to test SimSET this worked quite well. However, in torso phantoms choosing a uniform azimuthal angle results in a wide variation of photon weights: the photons are subsequently forced through the object to the target cylinder and have their weights adjusted for to compensate for the attenuation. SimSET produces too many photons, with too low weight, on lines with high attenuation, and too few photons, with too high weight, on lines with low attenuation. To address this, we have added a secondary azimuthal stratification based on an approximate transmission scan. More photons are started with lower weights in directions with low attenuation, and less photons with higher weights in directions with high attenuation. When the photons are forced through the attenuating material to the collimator, the weight reduction will be larger for the high weight/long attenuation path. This will even out the weight disparity between the photons.

C. Code Optimization

After profiling the SimSET software using *Rational quantify*, we identified the functions that took a lot of cpu time and were repeatedly called in the simulation. We modified two functions called during photon tracking. One computes voxel indices given an (x,y.z) position, the other determines whether two numbers are close to equal (i.e., equal up to floating point precision). We also made a cursory first pass through the rest of the tracking algorithm and removed a few obvious inefficiencies.

We plan to profile the software again after completing the new forced detection algorithms (described below in section *IIE*) and do further optimization.

D. Parallel Processing

A 52 node cluster of 1.7 GHZ P4 Xeon processors running LINUX was employed for the parallel processing prototype. Algorithms were carefully designed to distribute the computational burden equally among the nodes and minimize any possibility of duplication of photon tracking. The extendibility of the nodes was also considered in designing the prototype.

The job splitting among the nodes is performed in two steps. In the first step, a pre-processor splits the activity objects into multiple non-overlapping sub-objects such that each node processed approximately the same number of decays. Next, multiple realizations of the same simulation are run on each node. This second step minimizes the loss of data in the event a particular node on the cluster fails. It also facilitates the production of multiple realizations for observer studies.

E. Further Importance Sampling Changes

We are not yet finished with this project, and do not expect to approach our target efficiencies until several further algorithmic changes have been made. In particular, we plan several changes and additions to our forced detection algorithms.

The biggest change will be the addition of forced detection in the collimator. The new stratification described above leads to a much wider photon weight range, and thus to a lower quality factor. A photon path that hits a septum is not very productive, so photons that do so will have a very high weight. Without forced detection, most of these photons would be discarded, but a few would scatter or penetrate and be detected without any weight change; with forced detection many of these photons will be detected, but with their weight substantially reduced.

Forced detection of scatter in the object will be changed in two ways. SimSET will use productivity information to help choose scatter locations and will use the approximate transmission table to help choose the outgoing scatter direction.

Currently SimSET chooses the location for a forced scatter somewhere within the part of the object visible from the target cylinder within the acceptance angle. However, with the acceptance angle set to 90° this becomes the entire object. As a result, regions far outside the FOV are just as likely to be used for a scatter location as those inside. We plan to change probability distribution from which the forced scatter location is chosen to include the productivity information used for stratification.

The algorithm for choosing outgoing scatter direction for forced scatters will be modified to use the approximate transmission tables described above in the stratification section. The purpose is the same as that mentioned above: to take into account the subsequent step of forcing the photon through the attenuating material on the chosen photon path, with the goal being to reduce weight variation in the outgoing photons.

Finally, we plan to extend SimSET's forced detection to include coherent scatter. Currently SimSET does not allow the use of forced detection when coherent scatter is being simulated in the object. We plan to include coherent scatter in the new forced detection algorithms.

III. Results

A. Algorithmic Changes

Tables II and III show the efficiency improvements for simulations of an anthropomorphic torso phantom (Zubal's voxel_man) using two different tomograph configurations, 3D PET and 2D PET. The tomograph definitions were roughly based on the GE Advance. Table II also shows the target values we set for each column for comparison.

TABLE II
3D PET EFFICIENCY

	Decays/sec	Detected/Decay	QF	CFOM
Current	1529	0.176	0.306	82.3
Targeted	2400	0.9	0.75	1620
New version	1571	0.419	0.204	134.3

TABLE III
2D PET EFFICIENCY

	Decays/sec	Detected/Decay	QF	CFOM
Current	1447	0.029	0.495	20.8
New version	1576	0.185	0.185	53.9

B. Parallel Processing

Our tests on the speed-up that can be achieved with parallel processing indicate that near optimal speed-up can be achieved. Simulations of a 350*350*180 voxel NEMA2001 IQ Phantom were performed on a single P4 Xeon processor and on 20 nodes operating in parallel. The overhead of the pre-processing for job parsing was negligible: the single processor processed 2.7e3 decays/sec while 20 nodes operating in parallel processed 5.4e4 decays/sec.

IV. Discussion

We are still at a very early stage in the algorithmic work. It is too soon to know if our goal of a factor of twenty speedup is hopelessly optimistic. The rise in the number of detections per decay is smaller than we had hoped for to this point, but several of the not-yet-implemented algorithms should increase this number even if we don't get to our goal of 0.9. The quality factor has fallen to this point, but this is to be expected with the more aggressive stratification we implemented. This should be balanced out by the more aggressive forced detection we are planning.

Raising the number of decays processed per second as far as we hoped may not be possible. Our code optimization to date achieved a small speedup, and we expect to get some further improvement, but the 50% improvement we projected does not appear to be attainable without a major code restructuring.

To this point, the parallel processing has been an unqualified success. When these changes are added to the hoped-for algorithmic improvements, even a small parallel processing cluster could run simulations at an impressive rate. It should be possible to produce thousands of data sets a month.

V. References

[1] T. Lewellen, R. Harrison, and S. Vannnoy, "The SimSET Program, in Monte Carlo Calculations," *Monte Carlo Calculations in Nuclear Medicine*, M. Ljungberg, S. Strand, and M. King, Editors. 1998, Institute of Physics Publishing: Philadelphia. p. 77-92.

[2] S. D. Wollenweber, S. G. Kohlmyer, and T. K. Lewellen, " Evaluation of model-based scatter correction accuracy using Monte Carlo calculated phantom inputs," 2000 *IEEE Nuclear Science Symposium and Medical Imaging Conference*; Lyon, France, 2000 (proceedings on CD)..

[3] I. Zubal, and C. Harrell, Computerized three-dimensional segmented human anatomy*, Med. Phys*, 1994, pp. 299-302.

[4] D. Haynor, R. Harrison, T. and Lewellen, " The use of importance sampling techniques to improve the efficiency of photon tracking in emission tomography simulations," *Med Phys*, 1991, pp. 290-300.