Questions for comparison of clinical Monte Carlo codes

D.W.O. Rogers[1] **and Radhe Mohan**[2]

[1]National Research Council of Canada, Ottawa K1A OR6 (dave@irs.phy.nrc.ca), and [2]Medical College of Virginia, Richmond VA 23298-0058 (rmohan@hsc.vcu.edu)

Introduction

The use of Monte Carlo techniques for routine clinical treatment planning will soon be with us in the sense that various commercial producers of treatment planning systems are actively developing such systems. Since it has long been thought that Monte Carlo techniques represent the "ultimate" answer to the problem of accurate dose calculation, the commercialization of these techniques would appear to be highly desirable.

However we are now entering a particularly tricky stage in the development of routine clinical Monte Carlo techniques. The speed of the calculations is still an issue since if they take too long, Monte Carlo will never be used "routinely" in the clinic. A second issue concerns the accuracy of the calculations when actually implemented in a commercial system. Possible inaccuracies might come from two broad areas: (i) the calculation in the patient where compromises might be made to increase the speed or there might be bugs; and (ii) the specification or modelling of the clinical beam (including the patient specific shaping devices) which is essential for the overall calculation to be accurate.

We have arranged a session at the ICCR to which many of the major code developers have been invited and we have asked them some specific questions about their codes or algorithms. In the time permitted we have had to severely limit the topic and have chosen to emphasize the first two areas mentioned above, namely the speed issue and the accuracy issue. This does not imply that the beam modelling issue isn't important. On the contrary, it may prove to be the most difficult and important issue to resolve prior to useful clinical implementation. However, it is also the most difficult question to deal with in a short term comparative study.

In this paper we present the questions which the developers have been asked. These questions are by no means exhaustive and there are many other important questions to address but these could each be the focus of an entire session. For example: (i) How fast is fast enough? (ii) What level of accuracy and precision are clinically relevant and under what clinical circumstances? (iii) How do we prescribe dose for MC-based treatment planning? (iv) How do we evaluate MC treatment plans? (v) How do we calculate monitor units for MC-based treatments? (vi) How do we use MC for dose calculation for optimization of treatment plans? Time permitting we will discuss these and other questions in a panel discussion after the speakers.

Question I: Speed of photon calculations

To define a meaningful comparison of dose calculations between codes it is important to specify a case which everyone will use. There are many, many possible variations but the case selected here is meant to represent a "typical" case which is both easy to implement and comparable to present day clinical practice using other algorithms. By presenting this case in some detail it can provide a specific benchmark against which others can also do a comparison.

The phantom is 30.5 cm$\times$39.5 cm$\times$30 cm deep and filled with 5 mm^3 voxels. The odd dimensions are to ensure a voxel on the central axis but otherwise represent a realistic size. The voxels are to be filled randomly with one of 4 materials (water, aluminium, lung (ICRU, ρ=0.26 g/cm^3) and graphite) although if a particular algorithm's speed does not make use of voxels being the same material, then using water everywhere is acceptable. The incident beam is to be a 6 MV spectrum from a point source at 100 cm SSD and collimated to 10$\times$10 cm^2 at the phantom surface.

Statistical Uncertainties to be achieved

Any Monte Carlo timing comparison must specify what statistical precision has been achieved and the methods used to determine statistical uncertainties must be specified. If any smoothing procedure is applied it should be specified and its effects estimated but smoothing must not be used for the timing comparison.

Specification of precision is an evolving art. To avoid issues about the uncertainty on the uncertainty, a neutral way to specify precision is to sum in quadrature the estimated relative uncertainties in all voxels with a dose greater than some arbitrary lower dose limit, say $D_{max}/2$, and from this find the average relative uncertainty. For the timing comparison the precision sought is an average relative statistical uncertainty on these voxels of 0.02 or less.

I.a: How long?

The primary question is, how long does a code take to do the above calculation on an Intel P-III 500 MHz machine? To scale times from other machines the standard EGS4 timing benchmark can be used[1] as updated at
http://www.npl.co.uk/npl/rad/egs/bench/bench.html. If additional scaling is needed one should just scale a given chip architecture by its clock speed. This approach is subject to uncertainties of the order of at least 20% due to variations in compilers, memory size, cache size etc which also play some role. Any special "tricks" used by a particular

installation should be mentioned.

I.b: Physics Approximations?

One way to increase calculation speed is to make some approximations in the physics. These should be listed and an estimate made of the effects of these approximations (i.e. will they lead to 1%, 3% or 5% errors and under what circumstances). The techniques used to verify this accuracy should be described. Obviously the ideal situation is to find better ways of doing things which lead to no decrease in accuracy, at least for the in-patient calculation for the energy range of interest in radiotherapy.

I.c: Variance reduction used?

A strict definition of variance reduction techniques would be those techniques which are applied with no decrease in accuracy (e.g the use of forcing routines) but the term has been more generally used to include techniques which increase efficiency with negligible loss of accuracy (e.g. the use of range rejection to terminate a history when an electron cannot possibly escape from the current voxel). The variance reduction techniques used in a given code should be specified and the increase in efficiency ($= 1/(T_{\mathrm{cpu}}\sigma^2)$) from each type of variance reduction should be estimated.

I.d: Transport parameters used?

The transport parameters such as the energy cutoffs used in a Monte Carlo calculation play an important role in defining the speed and accuracy of the calculation. These cutoffs and other relevant transport parameters (eg ESTEPE step size limits or equivalent) should be specified and the methods used to determine that these parameters lead to accurate simulations should be specified.

I.e Time scaling with voxel size, energy etc?

The case specified here is for a given voxel size and beam energy. If the variation of calculation time with voxel size and beam energy are known they should be specified.

Question II: Speed of e^- calculations

If the code system being developed is also capable of calculations for incident electron beams, the above questions should be answered for mono-energetic electron beams with energies of 6 MeV and 20 MeV. In this case beam energy is thought to be a more critical parameter and hence the use of the two extreme energies.

Question III: Accuracy of calculations

The ultimate verification of code accuracy comes by comparison with experiment but this is very difficult to set up for comparison between codes because they require different inputs regarding beam characterization. As an initial step we propose that comparisons be against the EGS4/PRESTA code system. This provides a nominal way of demonstrating the accuracy of the in-phantom portion of the calculations and this well known code has been extensively benchmarked against experiment. The following cases have been set up to deliberately stress most codes without being too hard to implement. The results of the EGS4/PRESTA/DOSXYZ (ESTEPE=0.01(electrons) or default (photons), AE=0.521 MeV,ECUT=0.700 MeV) calculations are presented here graphically and are available in digital format on-line[4]. It is essential to show the absolute differences (or possibly ratios of results) between another calculation and the EGS4/PRESTA results because the expected differences are small.

Unlike the photon timing case where timing is not expected to be sensitive to the incident beam, here the incident beam is completely specified to avoid any differences due to different incident spectra.

To ensure fair comparisons, all parameters used for the timing comparisons in question I must be used for the accuracy comparisons in question III.

Both cases are for a 1.5×1.5 cm^2 beam from a uniform point source at 100 cm incident on the phantom and the dose is scored on the central axis in 5×5 mm voxels which are 2 mm thick in the beam direction.

III.a: Photon case

The phantoms are the same outer dimensions as in question I but they are now slab phantoms. From 0 to 3 cm is water, 3 to 5 cm is aluminium, 5 to 12 cm is lung and 12 to 30 cm is water. The voxels are 5 mm^2 in the x-y directions but only 2 mm deep to increase the resolution. The photon beam is a uniform 18MV beam from a realistic clinical accelerator as calculated at NRC using the BEAM code[2] and is available on-line[4]. Fig 1 presents the results as calculated with EGS4/PRESTA and these are also available on-line. The statistical precision of this reference calculation is $\pm0.3\%$.

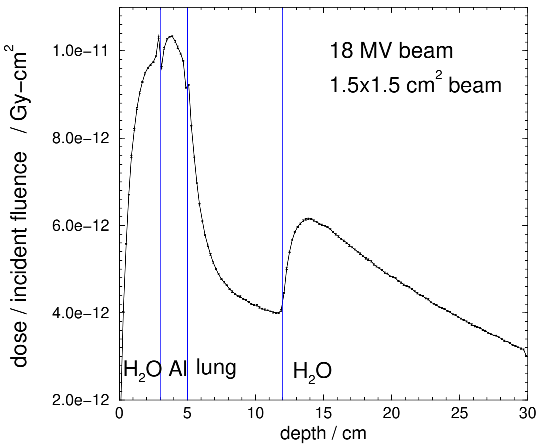

Figure 1: Depth-dose curves for the photon case as calculated by EGS4/PRESTA.

III.b: Electron case

The phantoms are very similar to those for photon beams although compressed in depth because of the limited range of the incident 20 MeV mono-energetic electron beam. From 1 to 2 cm is water, 2 to 3 cm is aluminium, 3 to 6 cm is lung material and 6 to 30 cm is water. Fig 2 presents the results of the EGS4/PRESTA calculation in this case. Typical precision is 0.2% of dose maximum. This figure also shows a direct comparison to an EGSnrc calculation[3] (with no spin effects).

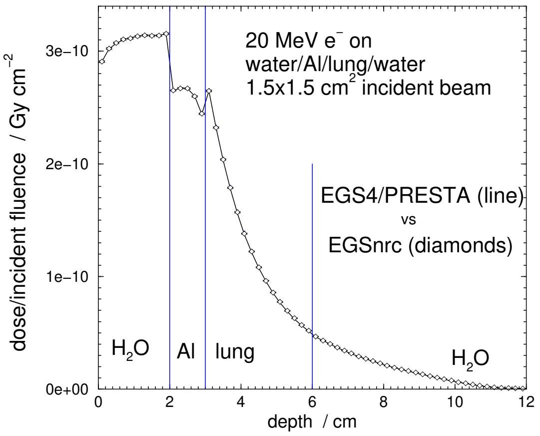

Figure 2: Comparison of depth-dose curves for the electron case as calculated by EGS4/PRESTA and EGSnrc (courtesy of Iwan Kawrakow).

The direct comparison in Fig 2 is not very informative since the differences are small compared to the scale whereas Fig 3 presents 2 other comparisons of the data. The absolute difference as a fraction of dose maximum shows that even for these two very slow codes there are some differences but they are limited to a few tenths %. Plotting the ratio of the two calculations vs depth shows that the codes diverge but this is for such small doses that it is not important. The presentation of the absolute difference in doses as a fraction of the total dose seems the most informative way to do the comparison.

Question IV: Beam models

As discussed above, beam modelling may prove to be one of the most difficult problems to handle in implementing Monte Carlo techniques into routine clinical practice. Within the tight space limitations, the authors were asked to give a very brief overview of how they are handling this.

Question V: CT conversion

Another important area, especially for electron beams, is the issue of converting CT numbers to materials and densities for the Monte Carlo calculations. The contributors were asked to briefly describe the techniques they use.

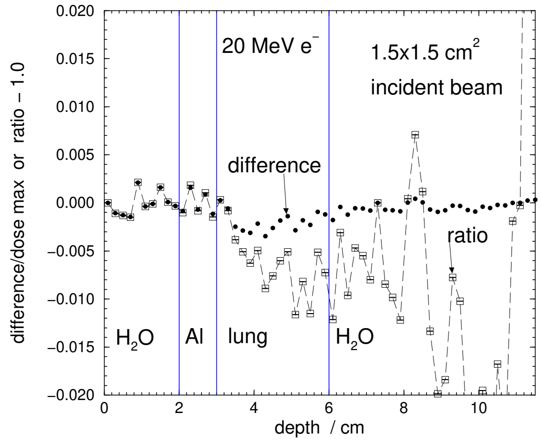

Figure 3: Difference in dose vs depth calculated by EGS4/PRESTA for the electron case less that calculated by EGSnrc as a fraction of the dose maximum (filled circles). Ratio of EGSnrc dose/EGS4-PRESTA dose -1.0 (open joined squares).

Question VI: Commercial Links

At their discretion, the authors were asked to mention any links they have with commercial treatment planning companies.

Summary/Acknowledgments

As the implementation of Monte Carlo techniques into routine clinical treatment planning proceeds, it will be necessary to develop standard tools for comparing various approaches and for assessing the speed of the calculations in a meaningful way. This is a first step in that direction and we thank the authors who have agreed to participate in this session at the ICCR. It must be emphasized that this is only a first step and there are many other important issues involved with the implementation of Monte Carlo into routine treatment planning in the clinic. We would like to thank Iwan Kawrakow for doing the EGSnrc calculations and for suggesting the criterion for statistical precision.

References

[1] A. F. Bielajew and D. W. O. Rogers A standard timing benchmark for EGS4 Monte Carlo calculations. *Medical Physics*, 19:303 – 304, 1992.

[2] D. W. O. Rogers, B. A. Faddegon, G. X. Ding, C. M. Ma, J. Wei, and T. R. Mackie. BEAM: A Monte Carlo code to simulate radiotherapy treatment units. *Med. Phys.*, 22:503 – 524, 1995.

[3] I. Kawrakow. Accurate condensed history Monte Carlo simulation of electron transport. I. EGSnrc, the new EGS4 version. *submitted to Med. Phys.*, 1999.

[4] http://www.irs.inms.nrc.ca/inms/irs/papers/iccr00/iccr00.html