MONTE CARLO MODELING IN EXTERNAL ELECTRON-BEAM RADIOTHERAPY—WHY LEAVE IT TO CHANCE?

Alex F Bielajew.
National Research Council,
Ottawa, CANADA.

Perhaps by the time we gather for ICCR XII and certainly by the time we gather for ICCR XIII, e^--beam treatment planning by Monte Carlo methods will be routine. As radical as this may seem, the argument is based upon the *current* state of hardware and software development.

Why Monte Carlo?

The Monte Carlo method, as applied to radiotherapy physics, is considered to have the following advantages and disadvantages:

Advantages

1. Algorithms are simple, coding and debugging efforts are minimised. Essentially, a Monte Carlo code couples ray-tracing algorithms to sampling algorithms of probability distributions describing physical interactions.
2. If the sampling algorithms are faithful, the accuracy of the code is determined by the accuracy of the fundamental cross section data.
3. The method is a microscopic one. Hence, boundaries between geometrical elements play no fundamental role in the mathematics except that ray tracing must be effected to boundaries and cross sections may change if the medium changes.
4. The geometries modeled may be arbitrarily complex and sophisticated— from a one element semi-infinite phantom to a complete electron linac with thousands of geometrical elements.

Disadvantages

1. Since the algorithms are microscopic, there is little theoretical insight derived in consideration of the macroscopic characteristics of radiation fields.
2. Monte Carlo calculations consume great amounts of computing resources, too much to be of practical use at this time in day-to-day radiotherapy, such as routine treatment planning.
3. Conventional $e\gamma$ Monte Carlo still relies on "condensed-history" algorithms. Some parts are still approximate (stopping power for cumulative low-energy events, multiple-scattering theory for cumulative small-angle events) and so there are still systematic errors.

The first disadvantage is largely a matter of perception. Monte Carlo is not a replacement for theoretical methods in radiation transport which provide a deep understanding of the behaviour of radiation fields. Rather, simulation methods should be considered to be placed somewhere between theory and measurement. Theorists tend to think of Monte Carlo methods as a quasi-experimental tool while experimentalists regard it as a quasi-theoretical analytical tool.
The second disadvantage is a weak one and will be addressed by the remainder of this report. To preview the discussion, computer hardware has become so fast and inexpensive and calculational techniques have been opti-

mised so that this weakness is greatly diminished.

The third disadvantage but true but only for applications that require an absolute dose to better than 1 or 2%. Much research is devoted to reducing these systematic errors. To be just, all aspects of radiotherapy have derived benefits from Monte Carlo methods, from determination of the particle output spectra of ^{60}Co-machines, linear accelerators and brachytherapy sources to the determination of dose in patients whose geometry and composition have been extracted from CT data. Simulations have played a crucial role in absolute dosimetry performed at Standards' Laboratories and relative dosimetry or machine calibrations as performed in cancer clinics. As the traditional standards based on exposure or air kerma in ^{60}Co-beams are being replaced by more clinically relevant high-energy absorbed dose standards, Monte Carlo modeling is essential since accurate results require accurate electron transport near detector interfaces. Arguably, Monte Carlo fulfills this latter role only marginally. This requirement is **driving** the current research into improving the accuracy of simulation methods. Therefore, it is necessary to ask:

Is Monte Carlo equal to the challenges of radiotherapy?

<u>Yes!</u>

External-beam radiotherapy is restricted to the range of several hundred keV to 50 MeV and the resolution required is of the order of 1 mm. The physics of this energy range and resolution are very well understood. Moreover, radiotherapy calculations are always referred to an absolute standard and require *relative* dose calculations only. Excellent relative agreement was demonstrated by benchmark experiments [1] where small air and aluminium cylinders were placed in water phantoms and the distributions of dose downstream from the heterogeneities using near-monoenergetic electron beams of two different qualities were measured.

These dose distributions exhibited remarkable short-range structure but the experimental results were reproduced by the Monte Carlo calculations providing that *all* aspects of the beam delivery system (accelerator exit window, scattering foil, collimator and intervening air) were modeled in the simulation as well. The results from similar calculations are shown in Figures 1 and 2.

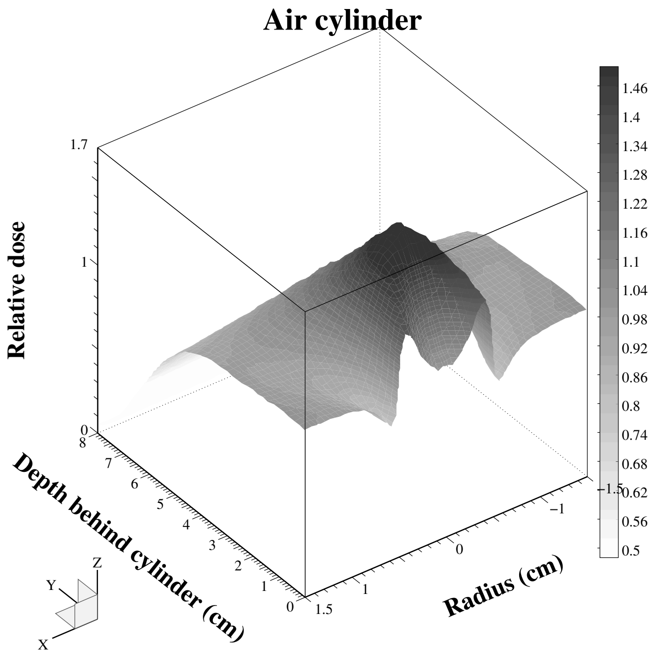

Figure 1: Dose distribution behind a 1 cm diameter, 2 cm long air cavity in a water phantom. The cylinder was aligned along the beam axis and placed at a depth of 2 mm. The incident electron spectrum is monoenergetic 20 MeV.

The accelerator employed in this work was a research-grade electron accelerator with energy resolution of the order of 100 keV at the exit window. Despite the excellent characteristics of the input particle spectrum, it was astonishing how sensitive the results were to

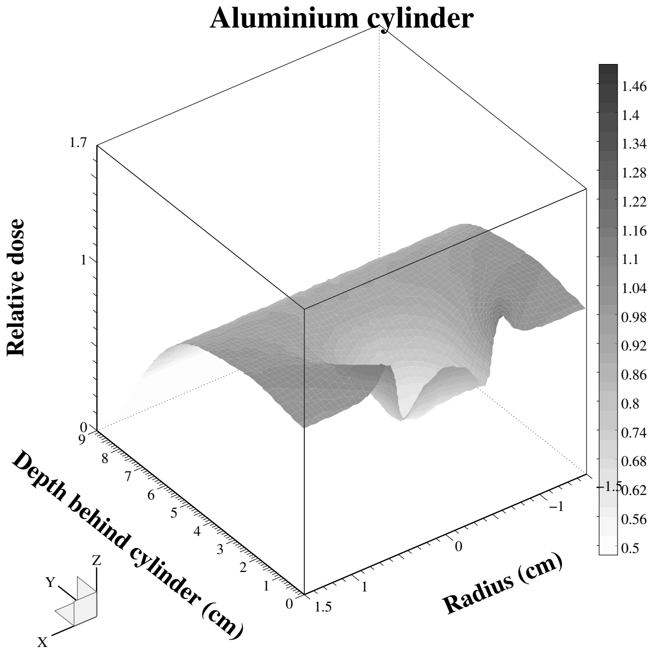

Figure 2: Similar to Figure 1 except that the cylinder was aluminium and 1 cm in length.

all the details of the accelerator. Therefore, this begs the question:

How exactly must clinical accelerators be modeled?

<u>In full detail!</u>

Clinical accelerators do not produce monoenergetic, monodirectional, and spatially uniform electron beams. At the point where the beam leaves the applicator plane and enters the patient, the particle spectrum has been given or has developed broad energy, spatial and angular spectra. A convincing demonstration of this was given by Udale [2] who investigated a Philips SL75-20 accelerator configured to produce 10 MeV electrons with various field sizes. Monte Carlo methods were employed in that work to study the effect of all beam-modifying accelerator head and applicator components—the accelerator exit window, primary collimators and assemblies, scattering foils, mirror, secondary collimators, applicator and its assembly. The simulation geometry is depicted in Figure 3. All of the components produce noticeable effects on some aspect of the particle spectrum at the patient plane.

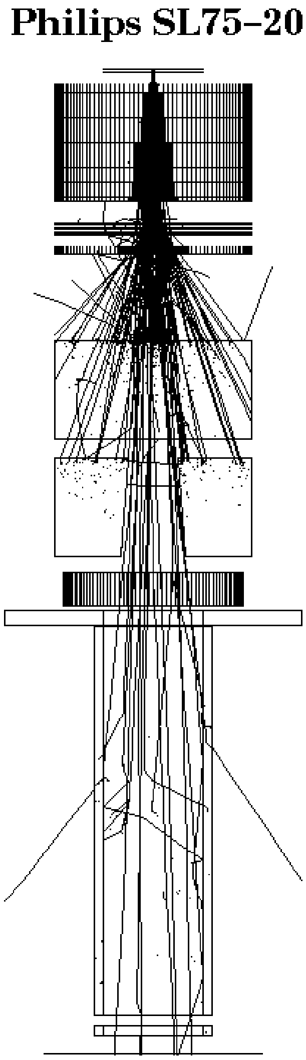

Figure 3: Simulation geometry of a Philips SL75-20 accelerator.

Recently, powerful software techniques have been developed [3] that enable the modeling of any accelerator to any level of complexity with relatively little effort. Starting from technical drawings of an accelerator treatment head and applicator geometries, a software mock-up can be constructed in hours compared to the weeks or months of special software coding with less powerful tools. Virtually any commercial or research accelerator can be modeled in minute detail. Agreement with experimental results are excellent [4] and powerful techniques have been developed for analysing the relative contributions of various accelerator components [5].

More than any other reason, the justification for employing Monte Carlo methods is to produce the characteristics of the spectrum of particles entering the patient. To do less is to make compromises. No matter how sophisticated the dose-calculation algorithm for the patient geometry, it depends upon high quality data describing the beam. If one accepts this argument then one is compelled to ask:

Is Monte Carlo fast enough?

Yes, starting today!

To answer this question, let us imagine a well-equipped modern-day radiotherapy clinic having 20 CPU's each of them having a computing power of 50 VUP's (Vax Unit of Power relative to a VAX 11/780)[1]. This can be in the form of single-CPU, or multiple-CPU desk-top workstations, a multi-CPU fileserver and, perhaps a multi-CPU compute server. It should be remarked that none of this hardware was purchased for Monte Carlo purposes. This technology is required for viewing CT-images, routine treatment planning, patients records and archives. These machines are all linked together on a network and the physicists are permitted to use the background CPU cycles at a low level of priority. Effectively, the Monte Carlo gets done "for free" as the hardware is required for other reasons.

If one employed all these CPU's in parallel on a full Monte Carlo-based treatment plan with a target accuracy of 2% starting the electron trajectories at the exit window of the accelerator, the calculation would require from 30 to 60 minutes. This figure is too long for routine treatment planning but short enough for special cases and for research into treatment planning methods.

To make this time shorter, the following strategies have been employed:

Divide and conquer

Rather than start the Monte Carlo calculations at the accelerator exit window, make use of the fact that the clinical accelerators are used typically in a fixed set of configurations, say 5 energies and 5 applicator assemblies and only the patient and patient-dependent paraphernalia (cutouts, immobilisation devices) change. Therefore, execute the accelerator-dependent part of the Monte Carlo once per machine configuration, store the accelerator output particle phase space in a file for repeated use on any patient. The storage requirements are modest, about 2.5 Gigabytes of data (1 Gigabyte using standard "lossless" compression algorithms) can represent all these 25 configurations of a single machine. This "divide and conquer" technique saves a factor of 4–6 reducing the time to execute a treatment plan to the range 5–15 minutes.

Optimisation of the Monte Carlo algorithm

There are standard "tricks" that can be applied to Monte Carlo calculations that take advantage of the fact that the perturbation by small heterogeneities is small in a global sense. While this may seem to be contradictory to the previous discussion on the effect of air and aluminium on dose deposition, it should be remarked that the perturbation was localised to the region near the heterogeneity and that the effect was cumulative and the global features several centimeters away from the homogeneity were left intact. Thus, one can pre-compute dose depo-

[1] In five years from the publication date of this report this statement will seem ridiculously out-of-date and will apply to a well-equipped home entertainment system.

sition within a homogeneous phantom once and for all to a high level of estimated precision. Then, with the patient representing a distribution of heterogeneity, a correction factor, $\mathrm{CF}(\vec{x}) = D_{\mathrm{het}}(\vec{x})/D_{\mathrm{hom}}(\vec{x})$, may be calculated with fewer histories to the same level of precision because of the correlation between D_{het} and D_{hom} [6]. Moreover, systematic errors in the calculational method tend to cancel out. The correction factor can then be applied to either a Monte Carlo calculated D_{hom} or a measured one! While this method is unproven for clinical applications, it is anticipated that the gain will be a factor of 1–5. This reduces the time for a treatment plan to the range 1–15 minutes.

Apply convolution to contaminant photons

Typically, there are about 4 times as many photons as electrons emerging from the treatment plane of an accelerator running in electron mode. These photons are a "contaminant", represent a small fraction of the dose, and deposit energy both near and far from the treatment volume. Rather than track these photons in full Monte Carlo detail, the photon spectrum at the patient plane can be employed as a source to a photon convolution algorithm which executes much more quickly than Monte Carlo. (The convolution "kernels" are also Monte Carlo-based.) Photons generated within the patient may be treated with Monte Carlo or convolution. These techniques can lessen computing time by a factor $1\frac{1}{2}$–2 reducing treatment planning time to the range 30 seconds – 10 minutes.

Hybrid transport algorithm

Conventional Monte Carlo techniques are usually "broad spectrum" in character. For example, the EGS4 code [7] can execute $e^{\pm}\gamma$-transport in any geometry, resolutions from microns in water to kilometers in the upper atmosphere, any material, and any quantity can be tabulated as scoring routines can be called just before or just after any particle phase-space change. Clearly, such powerful methods carry unnecessary overheads for routine dose determination in patient-like geometries. Techniques have been developed to "condense" the algorithm to its essentials. These "macroscopic" Monte Carlo methods [8] reduce the computing time by another factor of 10. Thus, the final span of execution times to execute a Monte Carlo-based treatment plan is reduced to the range 3 seconds – 1 minute.

These developments and adaptations have all been accomplished, at least in a research framework, in the OMEGA-project (**O**ttawa **M**adison **E**lectron **G**amma **A**lgorithm) [9]. The hardware and software capabilities are developed and it remains to fully integrate this capability into a working clinic. Even the most pessimistic turn-around time of 60 seconds is fast enough for routine use and even fast enough for the final stages of an optimisation algorithm.

The Future?

Based on recent developments the outline for the near future is clear. Aside from the recent software development and software optimisation, the most remarkable force behind this development is the amount of computing per unit cost. This has increased by a factor of about 1200 in the last 10 years, doubling annually. There are certainly constraints on this continuing forever but for the immediate future we can expect this trend to continue. Within 5 years, one should be able to do Monte Carlo-based treatment plans in less than one second.

Another important hardware development is communications. Once communication is fast enough, the previous model of a well-equipped modern-day clinic with in-house computing no longer applies. Small clinics with more limited resources can connect together over distances long enough to make national and even international co-operatives realisable. The amount of data required to fully specify a patient geometry is small in comparison to perceived demands of interactive home entertainment. Connecting to this network would be no more expensive than subscribing to a cable television service.

Perhaps by the time we meet for ICCR XV (maybe this will be a "virtual conference" made possible by "virtual reality" interfaces) Monte Carlo treatment planning will be commonplace. Computers will be so fast and inexpensive that our efforts to refine and optimise to save the factor of 60–600 discussed herein may seem to have been a great deal of wasted effort. Some will be prompted to ask, "Why did they just not wait until computers got fast enough?". The answer to this is that we learned a lot of physics along the way, learned about the effect of different accelerator designs and how heterogeneities perturb electron-beam dose-deposition patterns. By making something work when it is just marginally possible makes it work better and more reliably for the next generation of tools that come available.

Acknowledgements.

The author is grateful for the advice of Drs Bruce Faddegon, Charlie Ma, Rock Mackie and Dave Rogers. This work is supported by NCI grant R01 CA52692.

References.

1. Shortt *et al.*, 1986. Phys. Med. Biol. **31**, 235 – 249.
2. Udale, 1988. Phys. Med. Biol. **33**, 939 – 954.
3. Faddegon *et al.*, 1992. Med. Phys. **19** 1138 (abstract).
4. Ding *et al.*, 1993. Med. Phys. **20** 933 (abstract).
5. Wei *et al.*, 1993. Med. Phys. **20** 934 (abstract).
6. Holmes *et al.*, 1993. Phys. Med. Biol. **38**, 675 – 688.
7. Nelson *et al.*, 1985. Stanford Linear Accelerator Center Report SLAC-265.
8. Neuenschwander *et al.*, 1993. Med. Phys. **20** 900 (abstract).
9. Mackie *et al.*, 1990. Med. Phys. **17** 732 (abstract).