The Monte Carlo approach for the dose calculation
Basis for particles transport inside matter

“Corso per l’utilizzo del codice Monte Carlo in campo medico”

Istituto Superiore di Sanità
Roma, 6-7 Dicembre 2007

Simulation of a Higgs Boson creation
DEFINITION OF MONTE CARLO METHOD

The Monte Carlo method is a numerical solution to a problem that models objects interacting with other objects or their environment based upon simple object-object or object-environment relationships. It represents an attempt to model nature through direct simulation of the essential dynamics of the system in question.

A solution is determined by random sampling of the relationships, or the microscopic interactions, until the result converges.

A. F. Bielajew, 2001

A technique that provides approximate solutions to problems expressed mathematically using random numbers and trial and error.

TechEncyclopedia, 2007

The Monte Carlo is a mathematical approach that use a sequence of random number to solve a problem.

If we are interested in a parameter of, i.e., an equation: we must construct a big number of this equations, using different random numbers, and estimate the parameter and its variance

A. F. Bielajew, 2001
Monte Carlo method

From Wikipedia, the free encyclopedia

A Monte Carlo method is a computational algorithm which relies on repeated random sampling to compute its results. Monte Carlo methods are often used when simulating physical and mathematical systems. Because of their reliance on repeated computation and random or pseudo-random numbers, Monte Carlo methods are most suited to calculation by a computer. Monte Carlo methods tend to be used when it is infeasible or impossible to compute an exact result with a deterministic algorithm.

The term Monte Carlo was coined in the 1940s by physicists working on nuclear weapon projects in the Los Alamos National Laboratory.
DEFINITION OF MONTE CARLO METHOD

Application areas
Areas of application include:
- Graphics, particularly for ray tracing; a version of the Metropolis-Hastings algorithm is also used for ray tracing where it is known as Metropolis light transport
- Modelling light transport in multi-layered tissues (MCML)
- Monte Carlo methods in finance
- Reliability engineering
- In simulated annealing for protein structure prediction
- In semiconductor device research, to model the transport of current carriers
- Environmental science, dealing with contaminant behavior
- Monte Carlo method in statistical physics, in particular, Monte Carlo molecular modeling as an alternative for computational molecular dynamics.
- Search And Rescue and Counter-Pollution. Models used to predict the drift of a life raft or movement of an oil slick at sea.
- In Probabilistic design for simulating and understanding the effects of variability
- In Physical chemistry, particularly for simulations involving atomic clusters
- In computer science
  - Las Vegas algorithm
  - LURCH
  - Computer Go
- Modeling the movement of impurity atoms (or ions) in plasmas in existing and tokamaks (e.g.: DIII-D).
- In experimental particle physics, for designing detectors, understanding their behavior and comparing experimental data to theory
- Nuclear and particle physics codes using the Monte Carlo method:
  - GEANT - CERN's simulation of high energy particles interacting with a detector.
  - CompHEP, PYTHIA - Monte-Carlo generators of particle collisions
  - MCNP(4) - LANL's radiation transport codes
  - EGS - Stanford's simulation code for coupled transport of electrons and photons
  - PEREGRINE - LLNL's Monte Carlo tool for radiation therapy dose calculations
  - BEAMnrc - Monte Carlo code system for modeling radiotherapy sources (LINAC's)
  - PENETRATOR - Monte Carlo for coupled transport of photons and electrons, with applications in radiotherapy
  - MONK - Serco Assurance's code for the calculation of k-effective of nuclear systems
  - Modelling of foam and cellular structures
  - Modelling of tissue morphogenesis
In some cases the microscopic interaction are not well known

For these MC can either provide a small correction to the theory or it can be employed directly to verify or disprove the theory

When the relation between experiment and theory is not so direct, MC enters in the picture
In some cases the microscopic interaction are well known (electromagnetic interactions, etc)

In this case MC is useful to predict the trajectories of high energy particles through detectors and other complex assemblies of materials

* D.W.O. Rogers 1996

Courtesy of E.DiMaio & S.Magni (INFN/Milano)
The most successful scientist will avail himself or herself of more than one avenue to solve a problem.

Mathematical proofs exist demonstrating that MC is the most efficient way of estimate quantity in 3D when compared to first-order deterministic method, i.e. linear Boltzmann equation for particle transport.

Alex F Bielajew, 2001
With MANIAC: the first electronic digital computer
Fermi's work on pion-proton phase shift analysis

Fig. 4. A subprogram written by Fermi for calculating phase shifts by finding a minimum chi-squared in a fit to the data.
Two variables: $\theta$ and $D$

$$0 \leq \theta \leq \pi$$
$$0 \leq D \leq \frac{1}{2}$$

The needle will hit the line if the closest distance to a line $D$ is

$$D \leq \frac{1}{2} \sin(\theta)$$

NOTE
First application: scattering and absorption of neutrons.
These processes are intrinsically probabilistic and is simple the correspondence between the random numbers and the real result.
BUFFON EXPERIMENT OR THE ‘NEEDLE’ EXPERIMENT: MONTE CARLO EVALUATION OF $\pi$

The probability of an hit is the ratio of the blue area ($S_{\text{blue}}$) to the entire rectangle $R$

$$S_{\text{blue}} = \int_{0}^{\pi} \frac{1}{2} \sin(\vartheta) \, d\vartheta = 1$$

$$R = \frac{1}{2} \cdot \pi$$

$$\frac{S_{\text{blue}}}{R} = \frac{2}{\pi}$$

$N_0$ times the needle was shot
$N$ times the needle hit the line

$$\frac{N}{N_0} = \frac{2}{\pi}; \quad \Rightarrow \quad \pi = 2 \cdot \frac{N_0}{N}$$
BUFFON EXPERIMENT OR THE ‘NEEDLE’ EXPERIMENT: MONTE CARLO EVALUATION OF $\pi$
THE RESOLUTION OF INTEGRALS

....but,

Monte Carlo approach can be applied also for a deterministic problems like simply the solution of integrals, or, in a more complicated way, the transport of a particle (resolution of problems that can be written in for of definite integrals)

\[ \int_{a}^{b} f(x) \, dx = (b - a) \overline{f} \]

Theorem of the mean

\( \overline{f} \) is the average value of the function over the interval \([a,b]\)

One way to calculate this average is to evaluate the function at many random points inside the definition domain:

\[ \overline{f} \approx <f> = \frac{1}{n} \sum_{i=1}^{n} f(x_i) \]
By the large number law:

\[ \bar{f} = \lim_{n \to \infty} <f> \]

Since the \(<f>\) estimation is based on a finite number of points every integral will have an uncertainty.

To reduce the uncertainty repeated calculations (different batches) can be made of \(<f>\) and the associated variance derived.
THE INGREDIENTS OF MONTE CARLO

1. Random numbers and their generators
2. Uncertainty
3. Sampling techniques
4. Variance reduction

Isola Bella, Taormina (Messina)
One of the oldest and best-known pseudorandom number generator: the Linear Congruent Generator

\[ X_{n+1} = (aX_n + C) \mod m \]

- \( 0 < m \) the modulus
- \( 0 \leq a < m \) the multiplier
- \( 0 \leq c < m \) the increment
- \( 0 \leq X_0 < m \) the “seed” or start value

...appear random.....
• No correlation
• Long period
• Uniformity
• Reproducibility
• Velocity of production
While LCGs are capable of producing decent pseudorandom numbers, this is extremely sensitive to the choice of $c$, $m$ and $a$.

Historically, poor choices had led to ineffective implementations of LCGs. A particularly illustrative example of this is IBM RANDU which was widely used in the early 1970s and resulted in many results that are currently in doubt because of the use of this poor LCG

$$V_{j+1} \equiv (65539V_j) \mod 2^{31}$$

$$X_j \equiv V_j / 2^{31}$$

Points fall in 15 two-dimensional planes:
Marsaglia aplanes

The most efficient LCGs have an $m$ equal to a power of 2, most often $m = 2^{32}$ or $m = 2^{64}$. 
The Atari 8-bit family

Based on the MOS Technology 6502 CPU and were the first home computers designed with custom coprocessor chips.

Used noise from an analog circuit to generate true random numbers.

It is not practice
The repeatability is essential for code debugging

Starting since 1979
What's this fuss about true randomness?

Perhaps you have wondered how predictable machines like computers can generate randomness. In reality, most random numbers used in computer programs are pseudo-random, which means they are generated in a predictable fashion using a mathematical formula. This is fine for many purposes, but it may not be random in the way you expect if you're used to dice rolls, roulette wheels and lottery draws.

RANDOM.ORG offers true random numbers to anyone on the Internet. The randomness comes from atmospheric noise, which for many purposes is better than the pseudo-random number algorithms typically used in computer programs. People use the numbers to run lotteries, draws and sweepstakes and for their games and gambling sites. Scientists use them for random sampling and as input to modelling and simulation applications. Artists use them to make art and music. The service has been operating since 1998 and was built and is being maintained by Mads Haahr who is a Lecturer in the School of Computer Science and Statistics at Trinity College, Dublin in Ireland.
1. Random numbers and their generators

2. Uncertainty

3. Sampling techniques

4. Variance reduction
TREATMENT OF UNCERTAINTIES

From the Comitè International des Poids et Measures (1981) for experimental methods that can also be used for Monte Carlo

Random and systematic now replaced in A and B categories

*Andreo P. and Fransson A.*
“Stopping power ratios and their uncertainties for clinical electron beam dosimetry”

*Physics in Medicine and Biology 34 1947-61, 1989*

Category A: objective, evaluated using statistical methods

Category B: subjective, estimated by any other method

Both are expressed in terms of standard deviations or variance
Uncertainties (i.e. MCNP) can be evaluated during the simulation of each history.

Mean

\[ \bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i \]

Variance

\[ s_x^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2 \]

Direct approach

Uncertainty

\[ s_{\bar{x}}^2 = \frac{s_x^2}{N} \]
A scoring variable or simply a “score” or a “tally” may be a complicated object or there may be many geometrical volume elements to consider. It may require a lot of effort to stop after each history and compute its mean and variance.

1. The total number of histories are divided in batches e.g. EGSnrc uses 10-30 batches

2. Standard deviations of the variables of interests (tallies) are calculated in each batch

→ Less accurate but more efficient

N/n histories in each batch, N histories in total, n = 30

\[ \bar{x} = \frac{1}{N} \sum_{j=1}^{N} \sum_{i=1}^{n} x_{ij} \]
Review article

Andreo P.,
“Monte Carlo techniques in medical radiation physics”

*Physics in Medicine and Biology, 36 N.37, 861-920, 1991*
THE INGREDIENTS OF MONTE CARLO

1. Random numbers and their generators
2. Uncertainty
3. Sampling techniques
4. Variance reduction

“I Faraglioni”, Aci Trezza (Catania)
How random number are used to generate a distribution describing a physical process?

- Direct approach Inverse sampling
- Rejection technique
- Mixed method

From a random number and a given probability distribution, describing a physical process, we can reproduce the reality.
The direct approach or invertible cumulative distribution functions

Let's look a typical probability distribution function

- a and b not necessarily finite
- Must be integrable (we can normalize it)
- Must be not negative (negative probability are difficult to interpret)
Now let proceed to construct its cumulative probability distribution function

\[ c(x) = \int_{a}^{x} p(x') \, dx' \]

(P(x) can be normalized)

The cumulative function is mapped on to the range of a random variable \( r \)

\[ 0 \leq r \leq 1 \]

\[ c(x) = r \]

\[ c^{-1}(r) = x \]
From the exponential attenuation distribution, the appropriate cumulative distribution can be evaluated and the distance $z$ between interactions in a medium (step length) is determined by

$$p(z)\,dz = \mu e^{-\mu z} \,dz \quad 0 \leq z \leq \infty$$

Can $p(z)$ be normalized?

Yes, it is already in the correct form:

$$\int_{0}^{\infty} \mu e^{-\mu z} \,dz = 1$$

$\mu$ is the interaction coefficient

Now we must calculate the cumulative distribution $c(z)$ that will be mapped in $r$ (random number).

Hence we’ll derive $z$
A PRACTICAL CASE: INTERACTION LENGTH FOR A PHOTON BEAM

\[ r = c(z) = \int_0^z p(z) = \int_0^z \mu e^{-\mu z} \, dz = \mu \left[ -\frac{1}{\mu} e^{-\mu z} \right]_0^z = 1 - e^{-\mu z} \]

\[ p(z) = \mu e^{-\mu z} \]
Now we can invert $c(z) = r$ to obtain $z$

$$r = c(z) = \int_{0}^{z} p(z) = \int_{0}^{z} \mu e^{-\mu z} dz = \mu \left[ -\frac{1}{\mu} e^{-\mu z} \right]_{0}^{z} = 1 - e^{-\mu z}$$

$$z = c^{-1}(r) = -\frac{1}{\mu} \log(1 - r)$$

This curve express the step length derived from a random number.
Even the direct method is in principle always possible, sometime can be very difficult to invert \( c(z) \):

1. From the probability distribution \( p(x) \) a normalized \( f(x) \) is generated

This is possible if \( p(x) \) does not diverge and it is simple to detect the maximum
2. Generate a random $x$ between $a$ and $b$

3. Choose a second random number $r_2$:

If $r_2 < \frac{p(x)}{p(x_{\text{max}})}$, then $x$ can be accepted

This permit to choose an $x$ weighting it on the distribution shape

Remember that our goal is at any time, to calculate the variable $x$ of a function $p(x)$
If the probability distribution function is too difficult to invert
….and it is “spiky” so that the rejection is inefficiently

The probability could be factorized in two functions: one invertible and containing the spikiness and the other flat but containing most of the mathematical complexity

\[ p(z) = f(z) \cdot g(z) \]
THE INGREDIENTS OF MONTE CARLO

1. Random numbers and their generators
2. Uncertainty
3. Sampling techniques
4. Variance reduction
CONDENSED OR MACROSCOPIC TECHNIQUES

…let’s go just a little bit toward our purpose….

There are two possible approach to transport particles: macroscopic or microscopic

Slowing down of an electron in aluminum from 0.5 MeV to 1 keV

\[ 10^4 \text{ collisions} \]

In the majority of the application is unrealistic to simulate all the interactions

*Berger M.J.,
“Monte Carlo calculation of the penetration and diffusion of fast charged particles”
Physical interactions of electrons are classified into groups which provide a ‘macroscopic picture’ of the process

\[ s_0, s_1, \ldots, s_n, \ldots \]
\[ E_0, E_1, \ldots, E_n, \ldots \]
\[ u_0, u_1, \ldots, u_n, \ldots \]
\[ r_0, r_1, \ldots, r_n, \ldots \]

s: distance traveled
E: energy
u: direction
r: position

The step size (or energy loss between two steps) must be kept as small as possible but such that multiple collision models and energy loss models are valid.
CONDENSED OR MACROSCOPIC TECHNIQUES

CLASS I

Class I scheme groups all the interactions and uses a predetermined set of paths lengths (MCNP, ETRAN)

- Constant length of each step
  or
- Constant energy loss of each step (step length decreases at each step)

CLASS II or mixed procedure

This groups together the small angles interaction while treats singularly the major or ‘catastrophic’ events
EGS4 uses a Class II algorithm and treat separately the catastrophic from the soft interaction

The same is for PENELlope, FLUKA and GEANT4

**CATASTROPHIC EVENTS**

- Large energy-loss Moller scattering ($e^-e^- \rightarrow e^-e^-$)
- Large energy-loss Bhabha scattering ($e^+e^- \rightarrow e^+e^-$)
- Hard bremsstrahlung emission ($e^\pm N \rightarrow e^\gamma N$)
- Positron annihilation \textit{in-flight} and at rest ($e^+e^- \rightarrow \gamma \gamma$).

**SOFT EVENTS**

- Low-energy Moller (Bhabha) scattering (modeled as part of the collision stopping power)
- Atomic excitation ($eN \pm \rightarrow eN^*$) (modeled as another part of the collision stopping power)
- Soft bremsstrahlung (modeled as radiative stopping power), and
- Elastic electron (positron) multiple scattering from atoms, ($e^\pm N \rightarrow e^\pm N$).

**Treated singularly with big computing time**

**Here many interactions are regrouped and treated in statistical way**
A typical electron track simulation

Electron lost 4% of its energy in each step (Class I approach)

Catastrophic event
Catastrophic event

“No” means CLASS I and hence regrouped calculation
CONDENSED OR MACROSCOPIC TECHNIQUES: e- PHOTON

1. Place initial photon's parameters on stack

2. Pick up energy, position, direction, geometry of current particle from top of stack

3. Is photon energy < cutoff?
   - Yes: Terminate history
   - No: Sample distance to next interaction
     - Transport photon taking geometry into account

4. Has photon left the volume of interest?
   - Yes: Terminate history
   - No: Sample the interaction channel:
     - photoelectric
     - Compton
     - pair production
     - Rayleigh

5. Sample energies and directions of resultant particles and store parameters on stack for future processing
When we must treat very low-energy transfer (close to the electron binding energy)

Transport of low-energy electrons (say < 100 keV)

Geometrical regions (or distance to boundaries) are very small

The location of the threshold between catastrophic and statistical interaction can be chosen by the demands of physics of the problem and by the accuracy required
Techniques used to reduce the statistical error without increase the computation times

The CONDENSED APPROACH is a variance reduction technique in the class of AEIT (Approximate Efficiency Improving Techniques)

....to say better: it is the more important

• Problem dependent: not a general rule can be furnished

• Sometime spending man-hours and complicating programs to get a modest reduction in computation time may not be a good investment

• If we reduce the variance but time diverges, the efficiency is not reached

• It is important to note that we can reduce the statistical error of a quantity at the expense of increasing the uncertainties of other quantities
VARIANCE REDUCTION (…or reduce time with the same variance)

Variance reduction or efficiency increase?

\[ \mathcal{E} = \frac{1}{S^2 T} \]

S² variance

T computation time

We must avoid to make the calculation more efficient at the cost of computing results

It the same definition valid in simulation related to radiation treatment?
To decrease the Category A errors when a small number of interactions take place (i.e. photon interaction)

- **Geometry interrogation reduction**
- **Discard electron inside a zone**
- **PRESTA Algorithm**
- **Range rejection and Cut for secondary production**

**Interaction forcing**

**Splitting and Russian roulette**

**NOTE**

These methods sometimes refer to efficiency increase not necessarily to a variance reduction

Techniques mainly developed in the field of radiotherapy: combination of analytical techniques and Monte Carlo

- **Andreo P. And Nahum A.E.**
  “Stopping power ratio for a photon spectrum as a weighted sum of the values for monoenergetic photon beams”
  *Physics in Medicine and Biology* 31 1189-99, 1985

Monte Carlo

Analytically derived
VARIANCE REDUCTION (…or reduce time with the same variance)

Geometry interrogation reduction (or don’t calculate what you don’t really need)

The geometry routines are called only when necessary.

For example they are not called when the boundary will be not reached.

\[
\frac{\mathcal{E}_{\text{rejection}}}{\mathcal{E}_{\text{full}}} = 1.34
\]

This method is not necessary in the photon case when particle traverses long distance before interact.
Discard electron inside a zone

Electrons are neglected as we are interested in the energy deposited inside the zonal boundary.

The transport is stopped and the energy deposited “on the spot”

\[
\frac{\mathcal{E}_{\text{rejection}}}{\mathcal{E}_{\text{full}}} = 2.3
\]
PRESTA Algorithm

PRESTA has the geometry interrogation reduction

It makes improvement in the electron transport of electrons: it allows the use of large electron steps when electron is far away from boundaries

\[ \frac{\mathcal{E}_{\text{PRESTA}}}{\mathcal{E}_{\text{full}}} = 6.1 \]

A.F. Bielajew and D.W.O. Rogers, “PRESTA: the parameter reduced electron-step transport algorithm for electron Monte Carlo transport”
Range rejection

In this case we can discard the electrons that do not reach a sensitive volume.

Case of an ionization chamber with thick walls: energy released inside the wall is not of interest.

The efficiency can be improved of a factor 4.
Photon specific methods for the variance reduction

Interaction forcing: artificially increase the interaction probability of the process

Since the sampled step length tends to infinite there is a big probability that photon leave the geometry of interest without interaction

\[ z = c^{-1}(r) = -\frac{1}{\mu} \log(1 - r) \]
**Interaction forcing**

A common implementation of the method consist in the replacing the real mean free path $z$ with a shorter one $z_f$

This is equivalent to increasing the interaction probability per unit path length of a factor

$$\Gamma = \frac{z}{z_f} > 1$$

To maintain the simulation unbiased a correction factor must be introduced for every calculated quantity

$$w_p^1 = 1 \quad \text{For the primary and unforced particle}$$

$$w_p^2 = \frac{w_p^1}{\Gamma} \quad \text{For each particle produced in a forced interaction}$$
VARIANCE REDUCTION (….or reduce time with the same variance)

Splitting and Russian roulette methods

• A region of interest is defined
• Variance reduction is accomplished by modifying the weights of the particles
• If particle go towards the region of interest they are splitted
• If move away they are killed

Other user/code specific methods

• Use of the geometry to register particles
• Kill particle in specific regions
• Use the symmetry of the specific problem
• ..........
• as the ability of programmer permits!
VARIANCE REDUCTION (….or reduce time with the same variance)

Improve the performance of the simulation in terms of speed

✓ Parallelisation

✓ Access to distributed computing resources

Share with other institutes computing resources geographically distributed around the world
VARIANCE REDUCTION IN RADIATION TREATMENT

If one is interested in distributed quantity 3D dose distributions), it is better to define an overall uncertainty

This permit a more objective evaluation of the efficiency of a particular simulation algorithm

<table>
<thead>
<tr>
<th>Monte Carlo code</th>
<th>Time estimate (minutes)</th>
<th>% max. diff. relative to ESG4/PRESTA/DOSXYZ</th>
</tr>
</thead>
<tbody>
<tr>
<td>ESG4/PRESTA/DOSXYZ</td>
<td>42.9</td>
<td>0, benchmark calculation</td>
</tr>
<tr>
<td>VMC++</td>
<td>0.9</td>
<td>± 1</td>
</tr>
<tr>
<td>MCDOSE (modified ESG4/PRESTA)</td>
<td>1.6</td>
<td>± 1</td>
</tr>
<tr>
<td>MCV (modified ESG4/PRESTA)</td>
<td>21.8</td>
<td>± 1</td>
</tr>
<tr>
<td>RT_DPM (modified DPM)</td>
<td>7.3</td>
<td>± 1</td>
</tr>
<tr>
<td>MCNPX</td>
<td>60.0</td>
<td>max. diff. of 8% at Al/lung interface (on average ± 1% agreement)</td>
</tr>
<tr>
<td>Nomos (PEREGRINE)</td>
<td>43.3*</td>
<td>± 1*</td>
</tr>
<tr>
<td>GEANT 4 (4.6.1)</td>
<td>193.3**</td>
<td>± 1 for homogeneous water and water/air interfaces**</td>
</tr>
</tbody>
</table>

Sheik-Bagheri D. et al.,
Monte Carlo simulations: efficiency Improvement Techniques and statistical considerations
WHICH MONTE CARLO ALGORITHM?

It is difficult to compare different MC toolkits

- Different libraries (data or theoretical)
- Different generation of the primary beam (i.e. phase-space or exact reconstruction)
- Different approximation in the beam transport
- Different parameters (CUTs, STEPVs, etc)
- Statistical uncertainty not reported
OPEN ISSUES

• Which components of a MC toolkit (RNG, cross sections, cut parameters) must be known exactly?
• Which parameter should change the user in a MC TPS?
• Difference between simplified and full MC
• How can be compared in an objective way, different MC packages?

*Chetty I.J. et al.*
Issues associated with clinical implementation of Monte Carlo-based treatment planning: Report of the AAPM Task Group No. 105.
*Med Phys* (Submitted).
The End