http://www.geant4.org

Detector Simulations

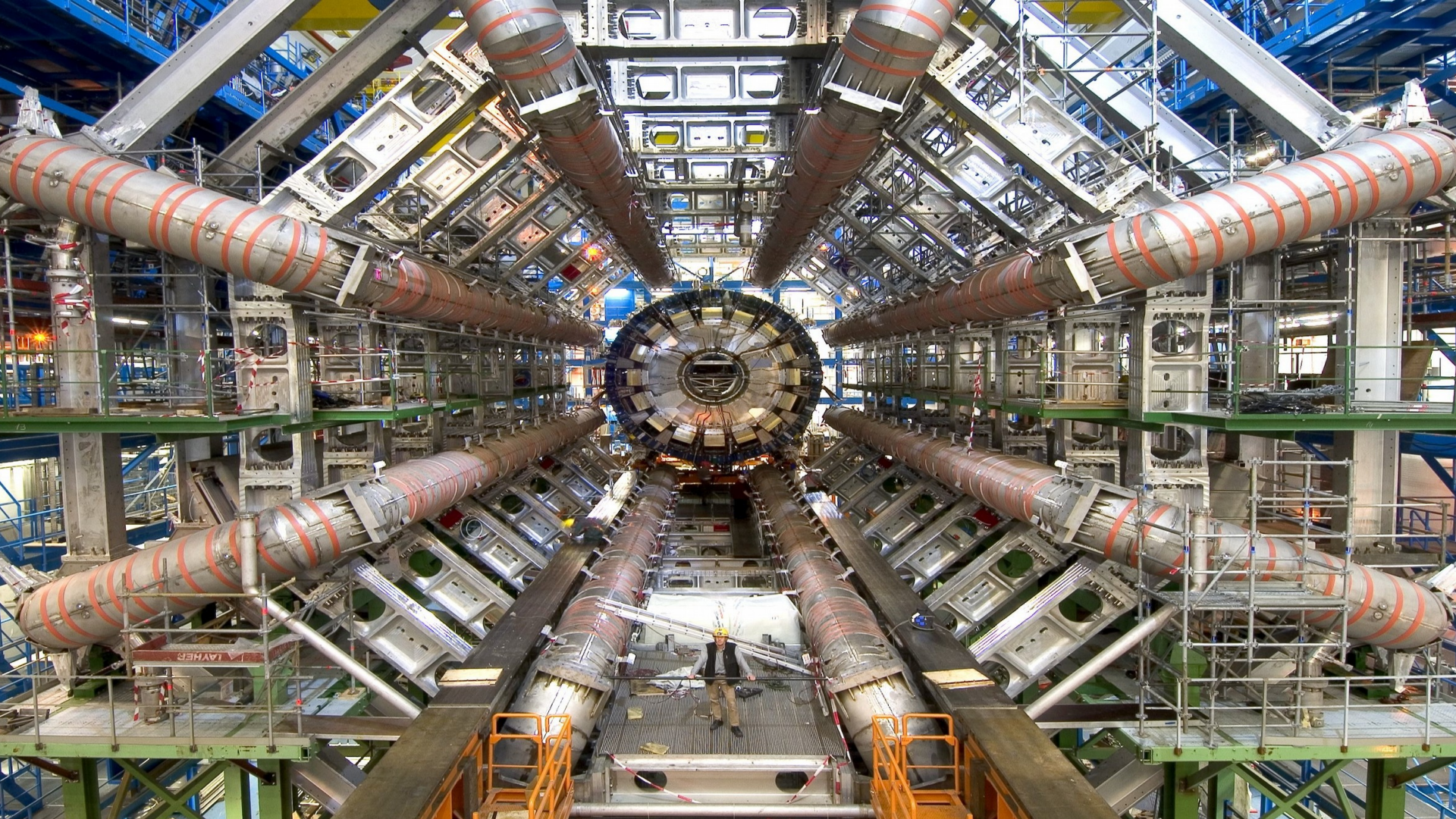
Introduction

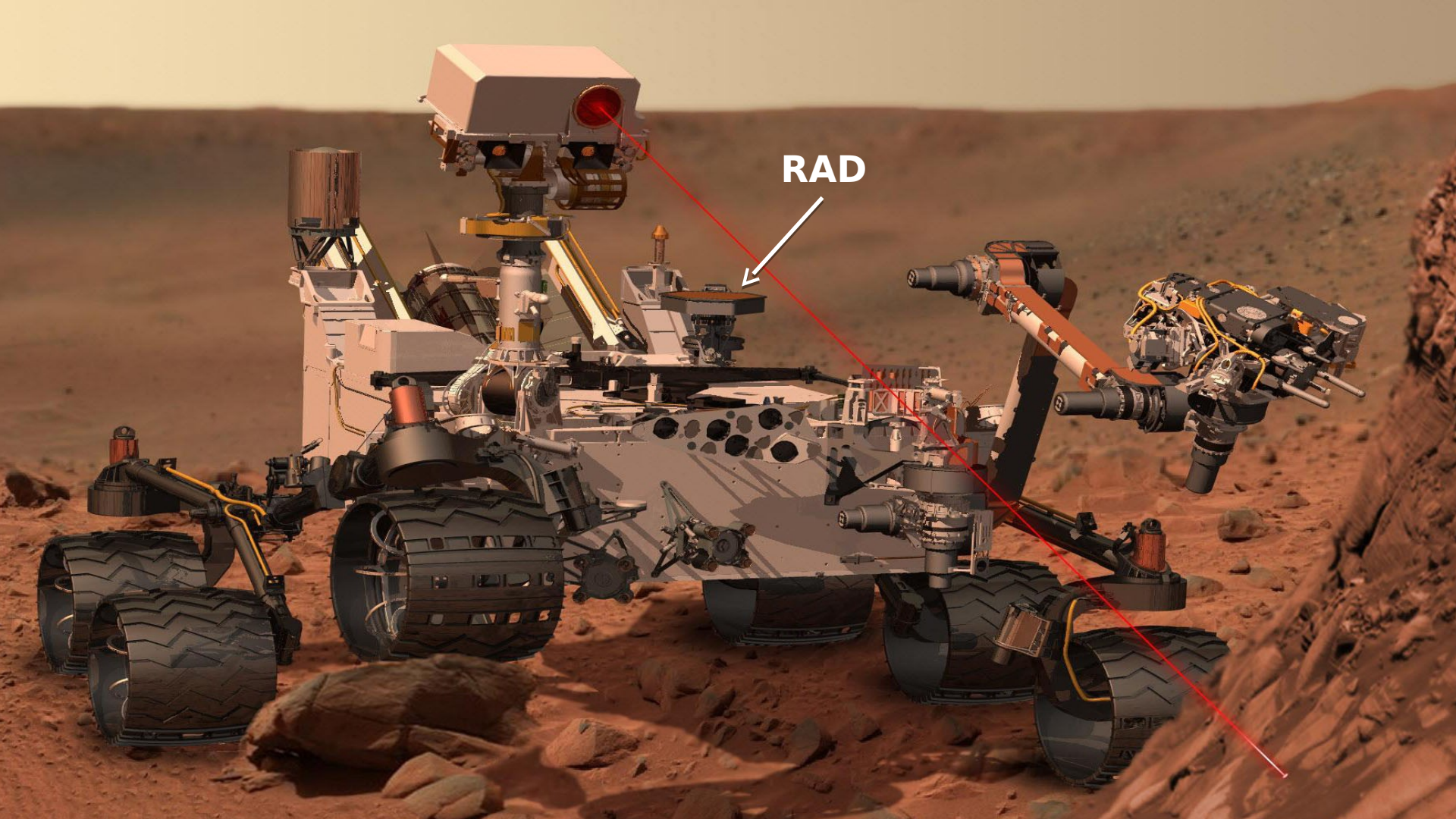
Andrea Dotti (adotti@slac.stanford.edu) ; SD/EPP/Computing

1st COFI Summer School – San Juan, PR – 11-17 July 2016

U.S. DEPARTMENT OF ENERGY
Office of Science

SLAC NATIONAL ACCELERATOR LABORATORY




RAD

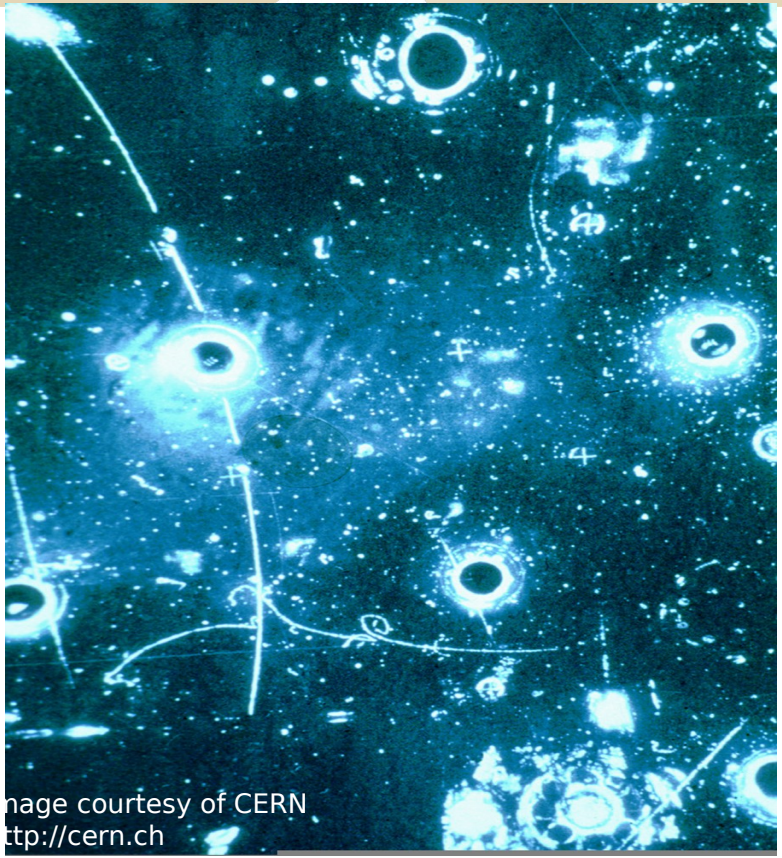
What do these applications have in common?

The study of the interaction of radiation (e.g. particles, x-rays) with matter has applications in several scientific areas:

- Basic research (e.g. at accelerators to discover new phenomena)
- Medical imaging (e.g. x-rays)
- Medical treatment (e.g. radio-therapy)
- Industrial (e.g. energy production, shielding)

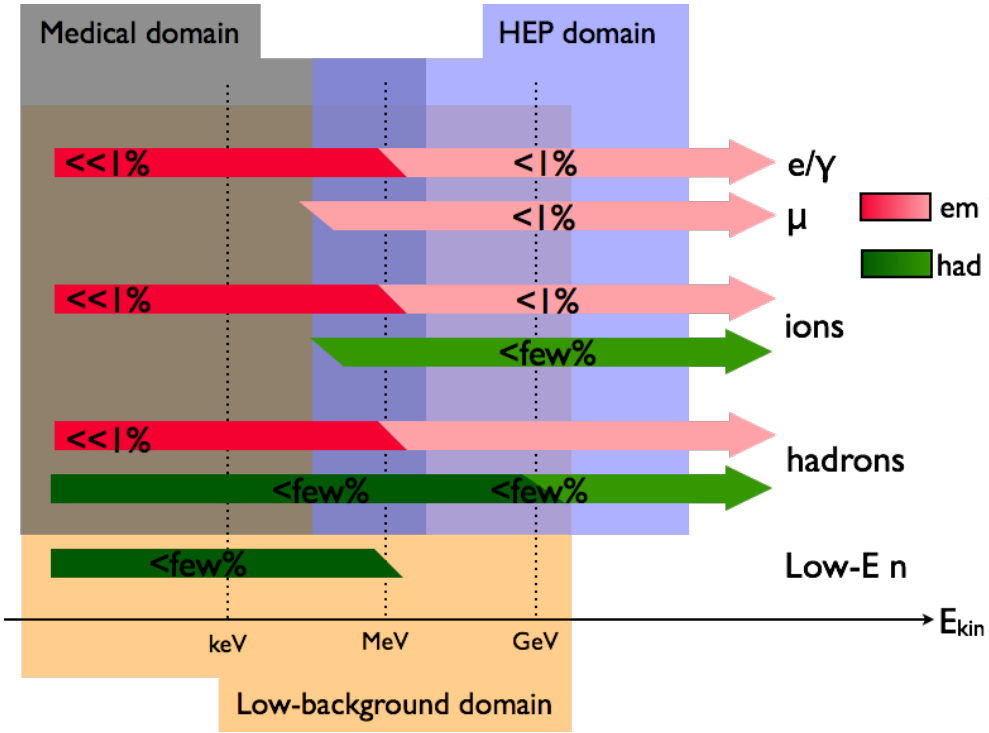
Essential tools in these fields are simulation programs. The most precise are based on Monte Carlo techniques

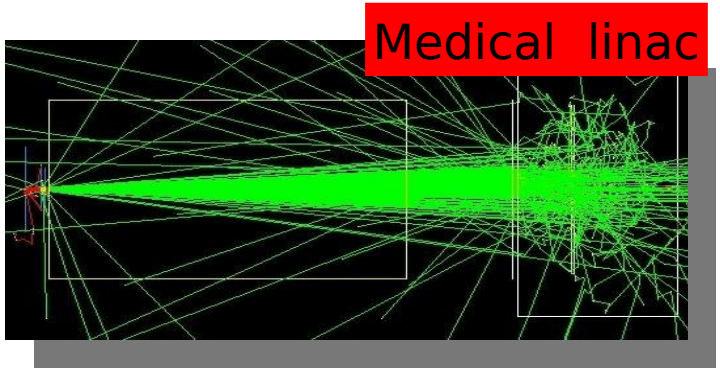
Several codes exists: Geant4 is one of them, the most widely adopted

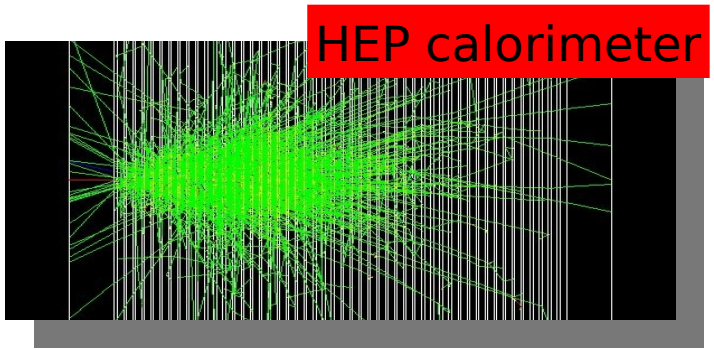


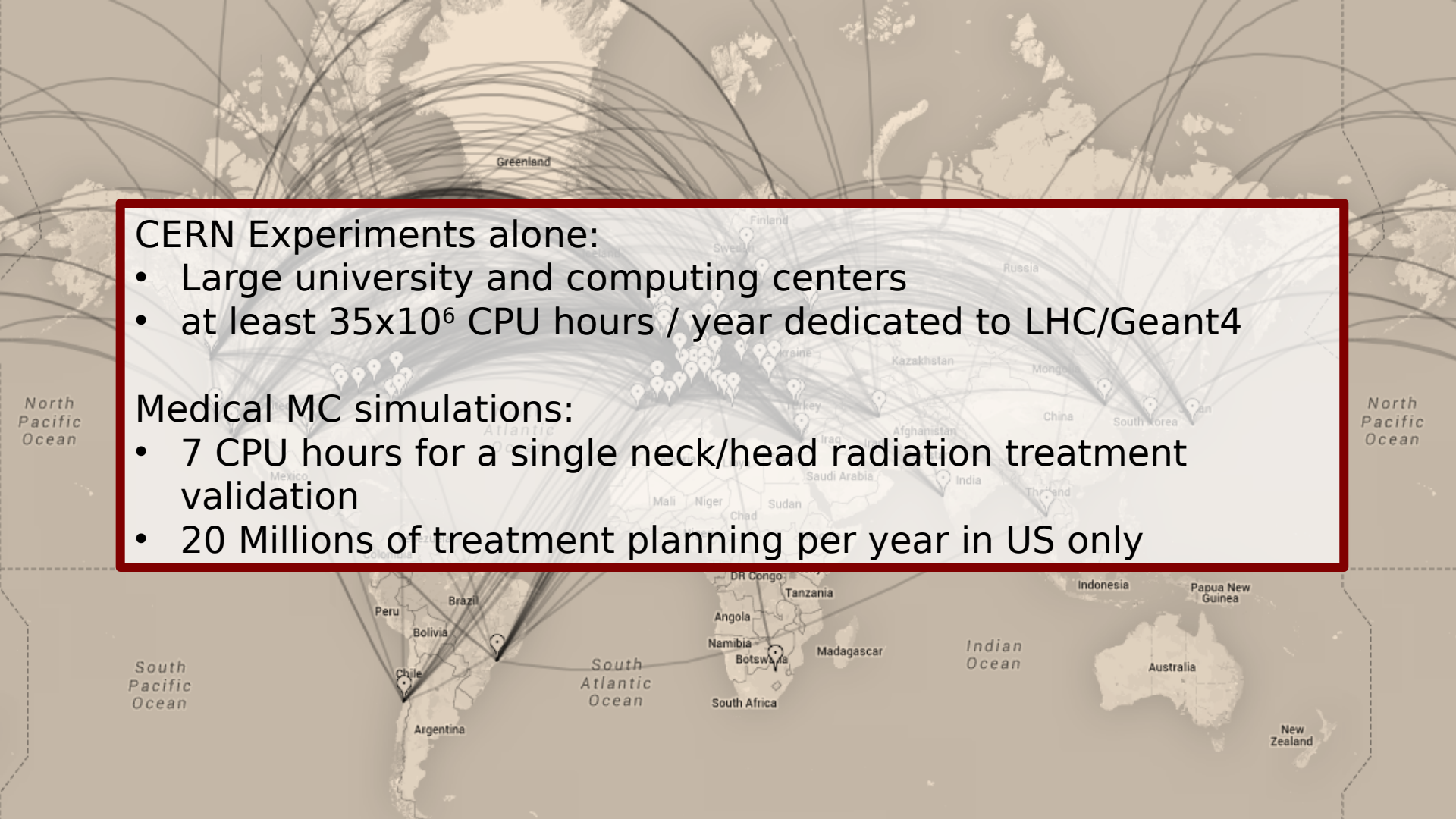
mage courtesy of CERN
ttp://cern.ch

Physics Requirements









CERN Experiments alone:

- Large university and computing centers
- at least 35×10^6 CPU hours / year dedicated to LHC/Geant4

Medical MC simulations:

- 7 CPU hours for a single neck/head radiation treatment validation
- 20 Millions of treatment planning per year in US only

Overview

SLAC

Part 1: Introduction

- Historical Notes
- Basics of Monte Carlo Method
- Basics of Monte Carlo for Radiation Transport
- Speedup of Simulations with variance reduction techniques

Part 2: Detector Simulations

- General Structure of a Detector Simulation package
- Geometry “ingredients”
- Physics “ingredients”

Part 1

Historical Notes: Monte Carlo Method

Monte Carlo method : definition

- The Monte Carlo method is a stochastic method for numerical integration.

- Generate N random "points" $\vec{x}_i$ in the problem space
- Calculate the "score" $f_i = f(\vec{x}_i)$ for the N "points"
- Calculate

$$\langle f \rangle = \frac{1}{N} \sum_{i=1}^{N} f_i, \quad \langle f^2 \rangle = \frac{1}{N} \sum_{i=1}^{N} f_i^2$$

- According to the Central Limit Theorem, for large N $\langle f \rangle$ will approach the true value $\bar{f}$. More precisely,

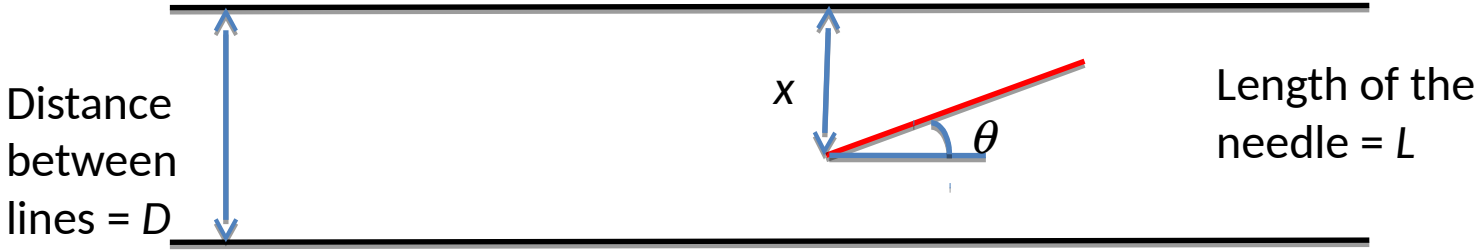
$$p(\langle f \rangle) = \frac{\exp\left[-(\langle f \rangle - \bar{f})^2 / 2\sigma^2 \right]}{\sqrt{2\pi}\sigma}, \quad \sigma^2 = \frac{\langle f^2 \rangle - \langle f \rangle^2}{N-1}$$

Buffon's needle

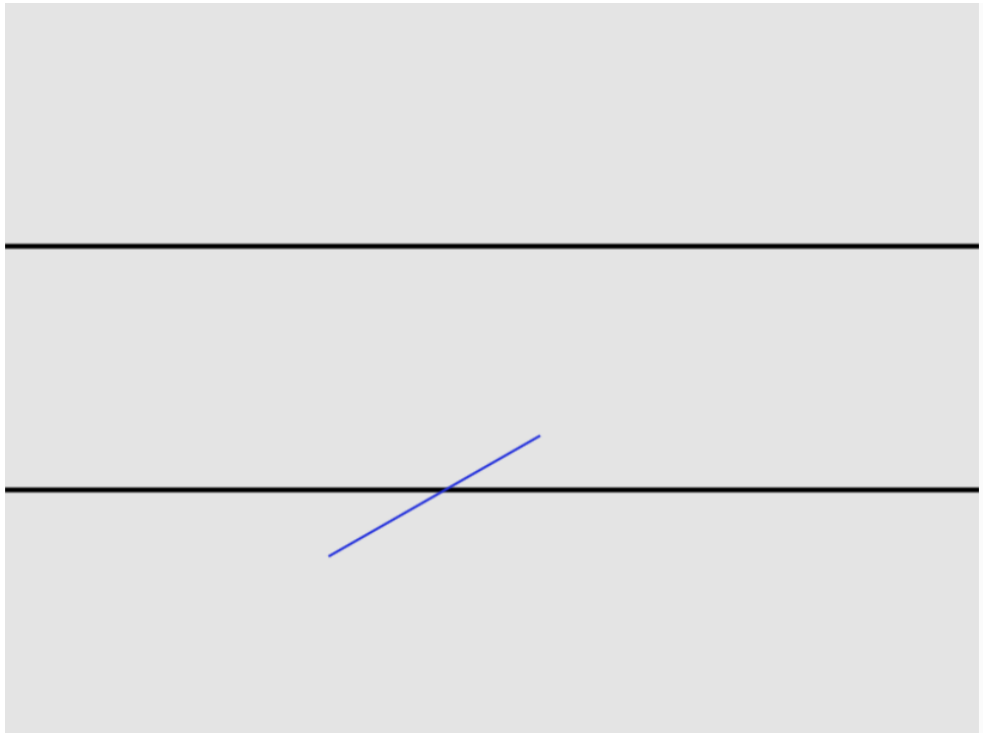
- Buffon's Needle is one of the oldest problems in the field of geometrical probability. It was first stated in 1777. It involves dropping a needle on a lined sheet of paper and determining the probability of the needle crossing one of the lines on the sheet. The remarkable result is that the probability is directly related to the value of π.
- The needle in the picture misses the line. The needle will hit the line if $x \leq L\sin(\theta)$. Assuming $L \leq D$, how often will this occur?

$$P_{cut} = \int_0^{\pi} P_{cut}(\theta)\frac{d\theta}{\pi} = \int_0^{\pi} \frac{L\sin\theta}{D}\frac{d\theta}{\pi} = \frac{L}{\pi D}\int_0^{\pi} \sin\theta\, d\theta = \frac{2L}{\pi D}$$

- By sampling P_{cut} one can estimate p.



Buffon's needle



Number of drops (n) = 1

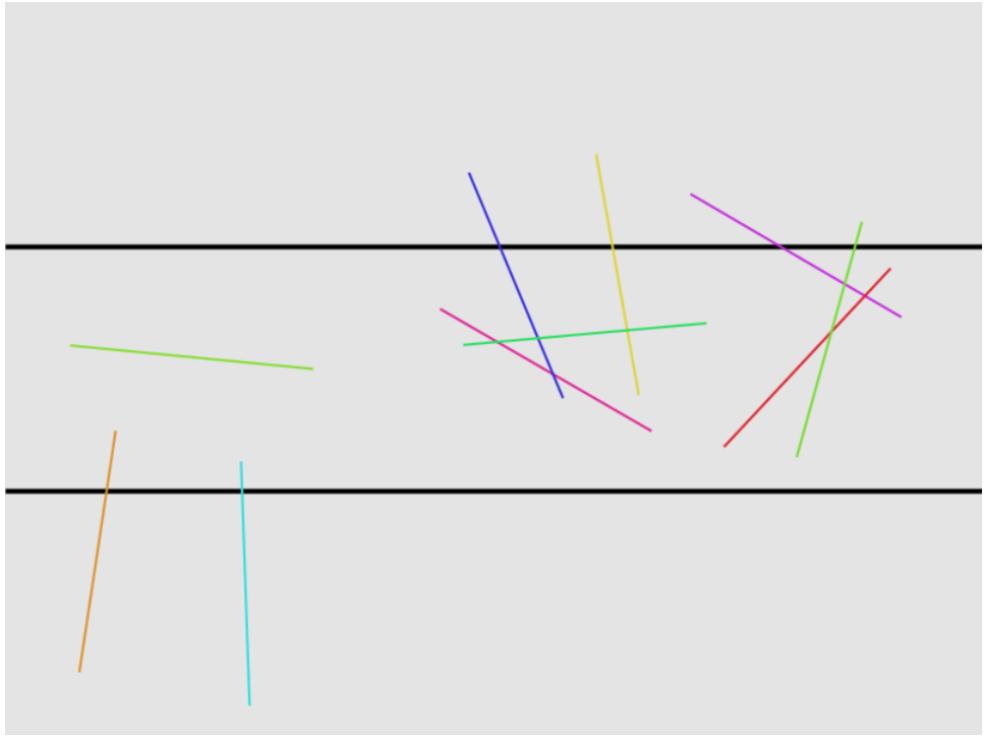
Number of hits (h) = 1

$P_{cut} = h/n = 1$

$\pi = (2\,L\,/\,D) * (\,n\,/\,h\,)$

$= 2\,/\,P_{cut} \sim 2$

Buffon's needle



Number of drops (n) = 10

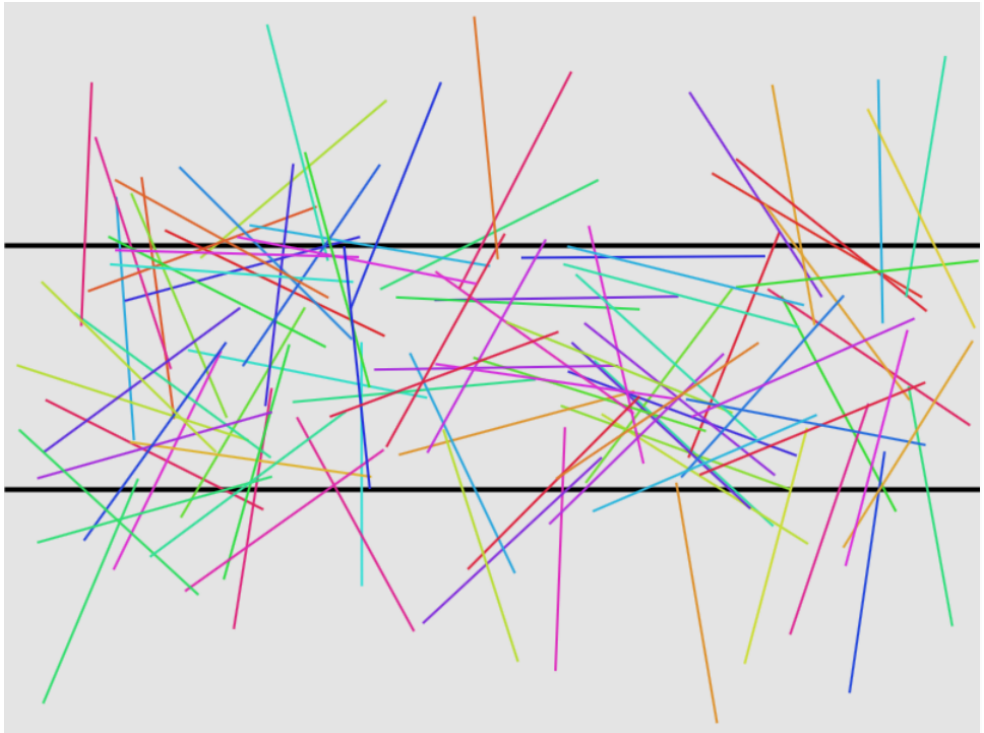
Number of hits (h) = 6

P_{cut} = h/n = 0.6

$\pi = (2\,L\,/\,D) * (\,n\,/\,h\,)$

$= 2\,/\,P_{cut} \sim 3.33333\ldots$

Buffon's needle



Number of drops (n) = 100

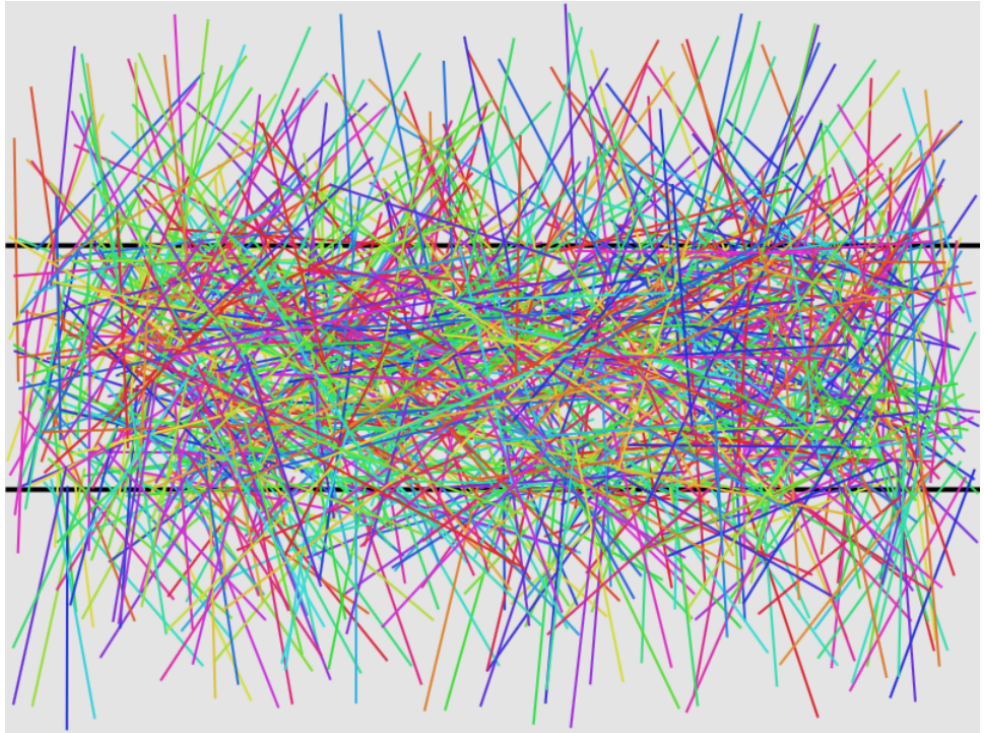
Number of hits (h) = 65

P_{cut} = h/n = 0.65

$\pi = (2\,L\,/\,D) * (\,n\,/\,h\,)$

$= 2\,/\,P_{cut} \sim 3.0769231$

Buffon's needle



Number of drops (n) = 1000

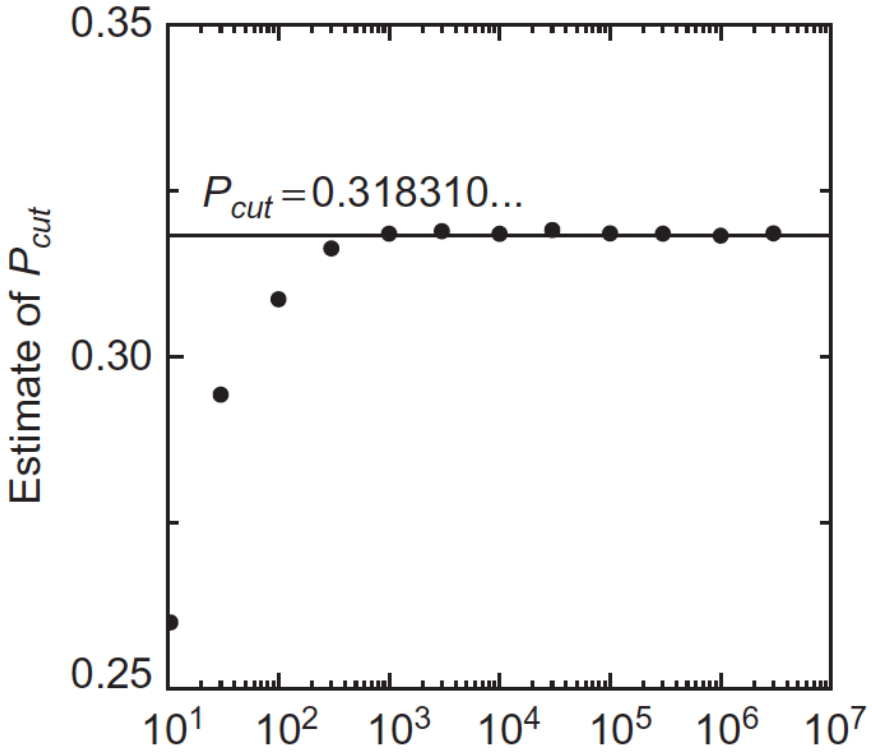
Number of hits (h) = 640

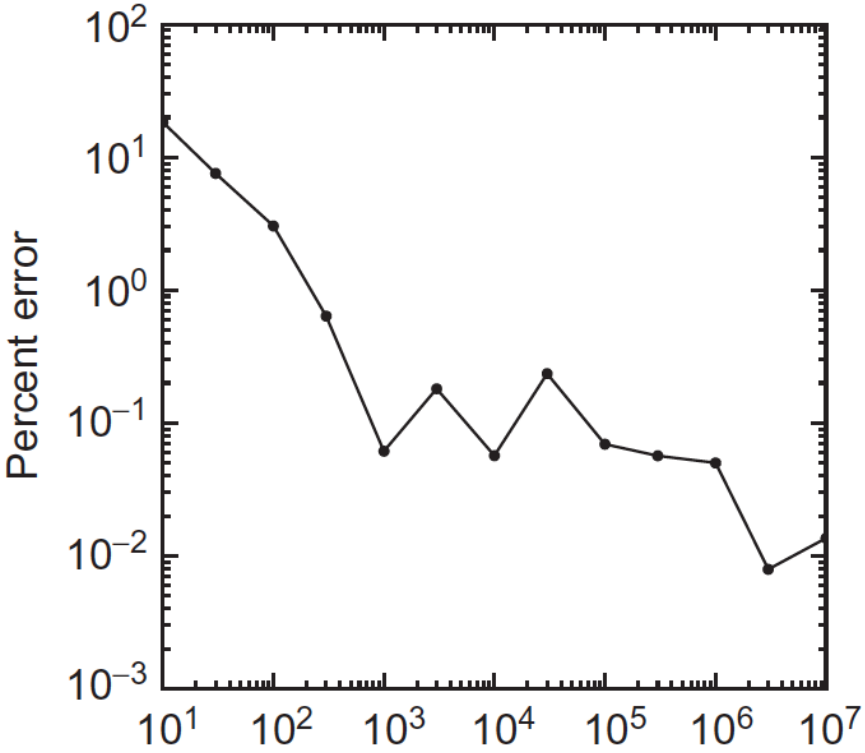
P_{cut} = h/n = 0.64

$\pi = (2\,L\,/\,D) * (\,n\,/\,h\,)$

$= 2\,/\,P_{cut} \sim 3.125$

Buffon's needle





P_{cut} = h/n = 0.318310... ➡π = (2 *L* / *D*) * (n / h) = 2 / P_{cut} ~ 3.14159...

Laplace's method of calculating π (1886)

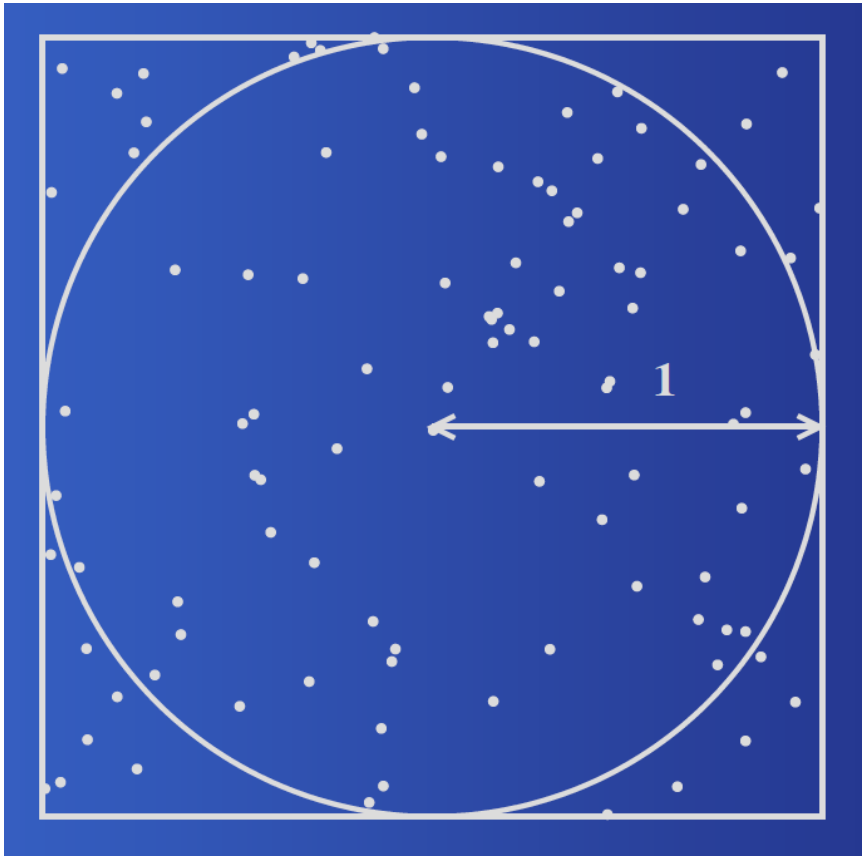
Area of the square = 4

Area of the circle = π

Probability of random points inside the circle = π / 4

Random points : *N*
Random points inside circle : *Nc*

$\pi \sim 4\, Nc / N$



Monte Carlo methods for radiation transport : brief history

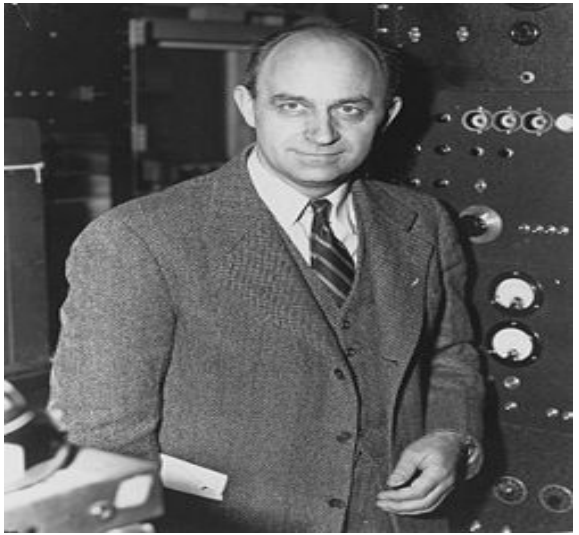
SLAC

- Fermi (1930): random method to calculate the properties of the newly discovered neutron
- Manhattan project (40's): simulations during the initial development of thermonuclear weapons. Von Neumann and Ulam coin the term "Monte Carlo"
- Field growth with the availability of digital computers
- Berger (1963): first complete coupled electron-photon transport code that became known as ETRAN
- Exponential growth since the 1980's

Pioneers of the Monte Carlo simulation method

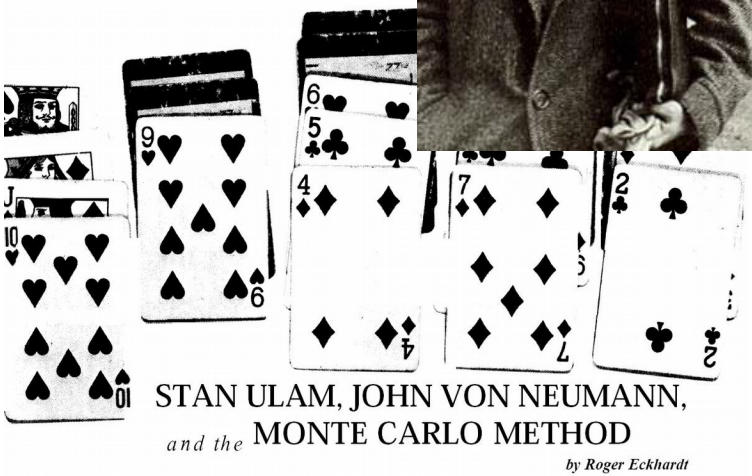
SLAC

- What is not so well known is that the Italian physicist Enrico Fermi (1901-1954), in the early 1930s, used statistical sampling methods, before the name Monte Carlo had been coined, in his studies of the slowing down of neutrons.
- According to Emilio Segré, Fermi's student and collaborator, Fermi wowed his Roman colleagues with his uncannily accurate predictions of their experimental results, which, unknown to them, were obtained by statistical sampling methods. The fact that Fermi had no computers to generate his random numbers and had to do his calculations with mechanical devices or even in his head, makes his accomplishments all the more remarkable.
- However, Fermi never published the results of these Monte Carlo calculations and, as a consequence, he often is neglected in the historical record of Monte Carlo.



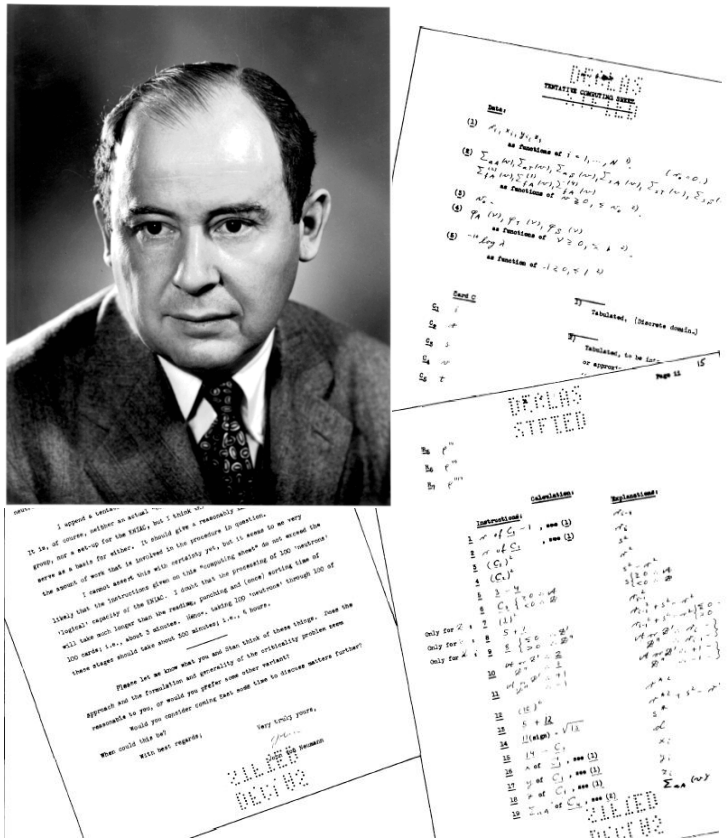
Pioneers of the Monte Carlo simulation method

Stanislaw Ulam (1909-1984) is a Polish mathematician who participated in the Manhattan Project and proposed the Teller–Ulam design of thermonuclear weapons. While in Los Alamos, he suggested the Monte Carlo method for evaluating complicated mathematical integrals that arise in the theory of nuclear chain reactions (not knowing that Enrico Fermi and others had used a similar method earlier). This suggestion led to the more systematic development of Monte Carlo by Von Neumann, Metropolis, and others.



Pioneers of the Monte Carlo simulation method

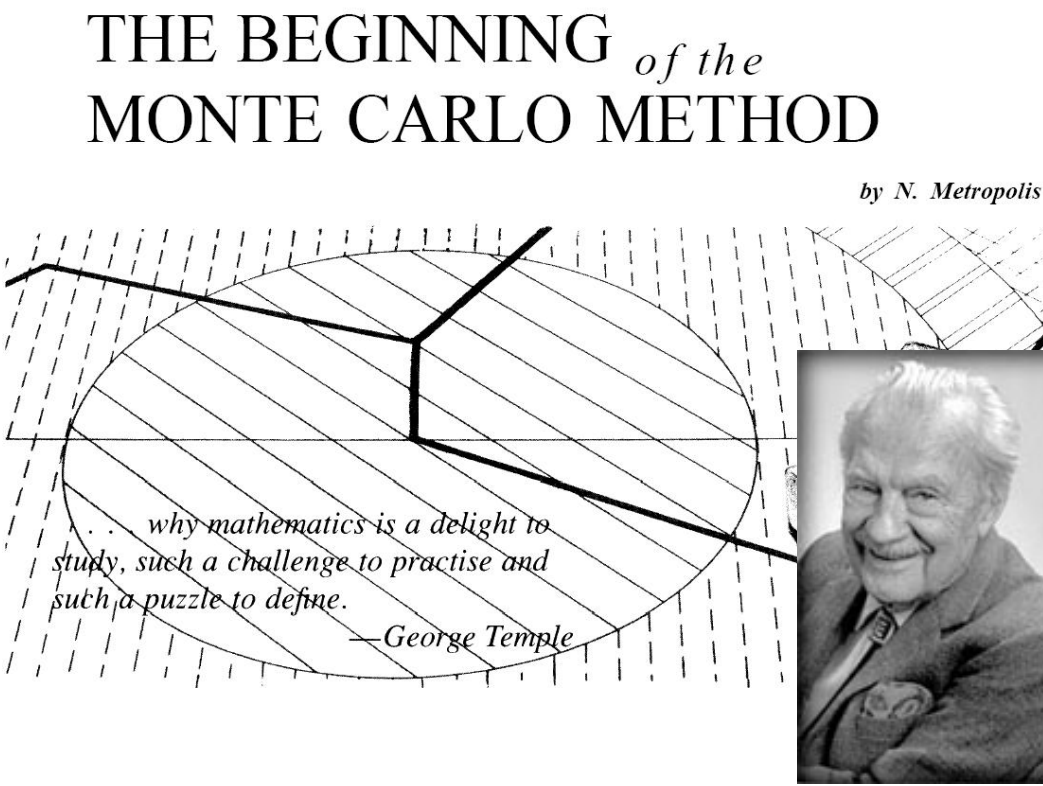
John Von Neumann (1903-1957) was taken by the idea of doing statistical sampling using newly developed electronic computing techniques. The approach seemed to him to be especially suitable for exploring behavior of neutron chain reactions in fission devices. In particular, neutron multiplication rates could be estimated and used to predict the explosive behavior of the various fission weapons then being developed. In March of 1947, he wrote about this to Robert Richtmyer, at that time the Theoretical Division Leader at Los Alamos.



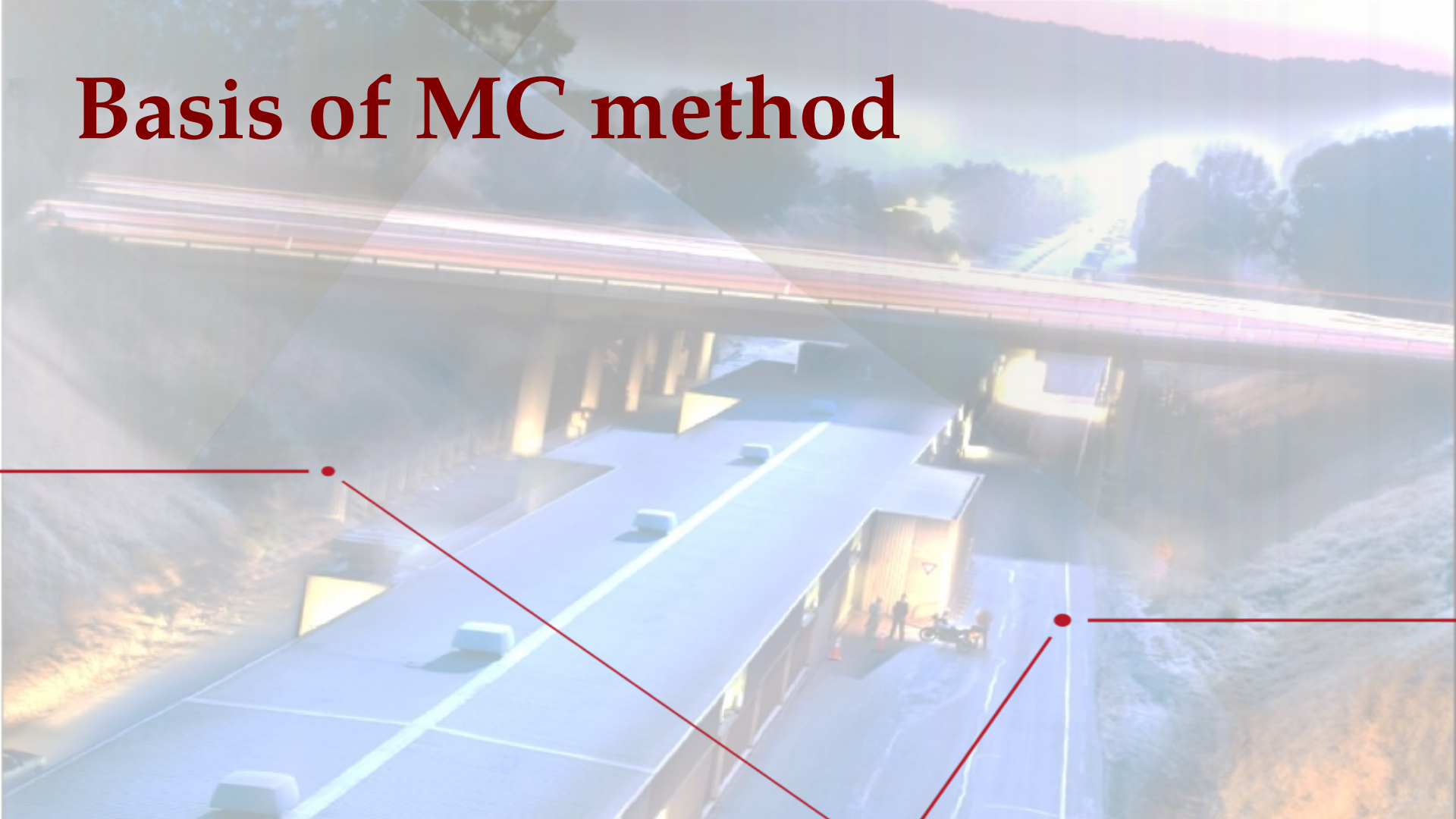
Pioneers of the Monte Carlo simulation method

SLAC

A team headed by Nicholas Metropolis (1915-1999) carried out the first actual Monte Carlo calculations on the ENIAC computer (the world's first electronic digital computer, built at the University of Pennsylvania) in 1948. The Metropolis algorithm, first described in a 1953 paper by Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and Edward Teller, was cited in Computing in Science and Engineering as being among the top 10 algorithms having the "greatest influence on the development and practice of science and engineering in the 20th century."



Basis of MC method



Probability Density Function (PDF) - 1

SLAC

- A variable is considered random (also called stochastic) if its value cannot be specified in advance of observing it.
  - Let x be a single continuous random variable defined over some interval. The interval can be finite or infinite.
- The value of x on any observation cannot be specified in advance because the variable is random. Nevertheless, it is possible to talk in terms of probabilities.
  - The notation Prob$\{x_i \leq X\}$ represents the probability that an observed value x_i will be less than or equal to some specified value X. More generally, Prob$\{E\}$ is used to represent the probability of an event E.
- A Probability Density Function (PDF) of a single stochastic variable is a function that has the following three properties:
  1) it is defined on an interval $[a, b]$
  2) it is non-negative on that interval, although it can be zero for some $x \in [a, b]$
  3) it is normalized such that $\int_a^b f(x)\,dx = 1$
     - Here, a and b represent real numbers or infinite limits (i.e., $a \to -\infty$ and/or $b \to \infty$).

Probability Density Function (PDF) - 2

- A PDF $f(x)$ is a density function, i.e., it specifies the probability per unit of x, so that $f(x)$ has units that are the inverse of the units of x.
- For a continuous random variable, $f(x)$ is not the probability of obtaining x.
  - There are infinitely many values that x can assume and the probability of obtaining a single specific value is zero.
- Rather, the quantity $f(x)\,dx$ is the probability that a random sample x_i will assume a value within x and $x+dx$.
  - Often, this is stated in the form

$$f(x) = \text{Prob}\{\, x \le x_i < x+dx \,\}$$

Cumulative Distribution Function (CDF)

- The integral defined by

$$F(x) \equiv \int_a^x f(x')\,dx'$$

  where $f(x)$ is a PDF over the interval $[a, b]$, is called the Cumulative Distribution Function (CDF) of f.

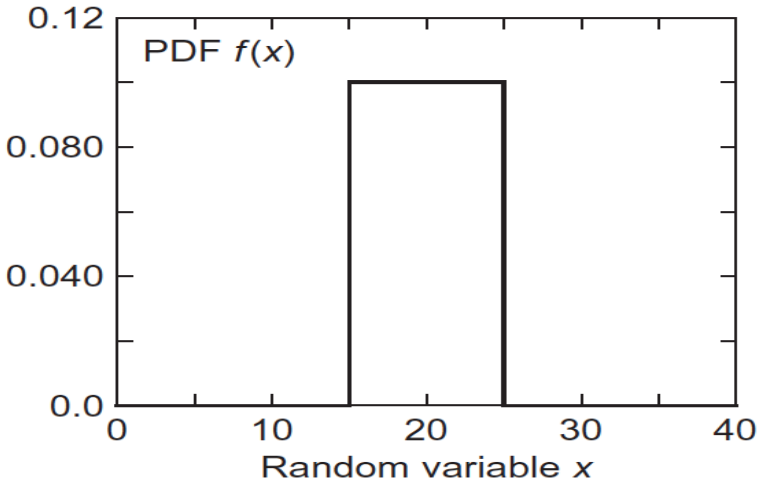
- A CDF has the following properties:
  1) $F(a) = 0.$, $F(b) = 1.$
  2) $F(x)$ is monotone increasing, as $f(x)$ is always non-negative.
- The CDF is a direct measure of probability. The value $F(x_i)$ represents the probability that a random sample of the stochastic variable x will assume a value between a and x_i, i.e., Prob$\{a \le x \le x_i\} = F(x_i)$.
- More generally, $\text{Prob}\{x_1 \le x \le x_2\} = \int_{x_1}^{x_2} f(x)\,dx = F(x_2) - F(x_1)$.

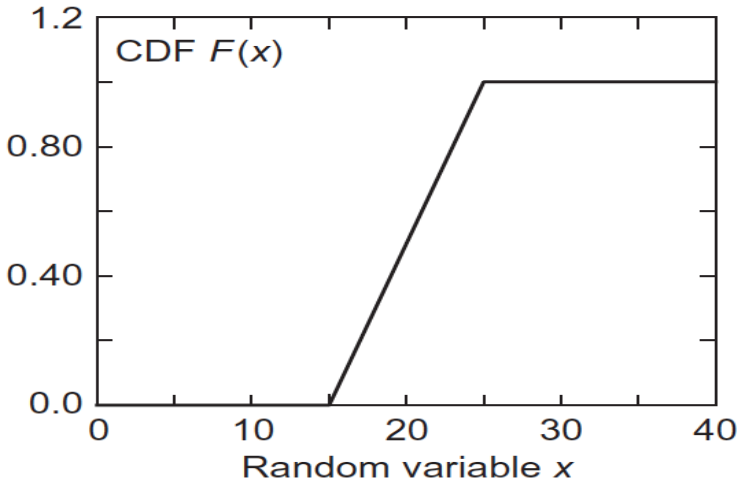
Some example distributions – Uniform PDF

The uniform (rectangular) PDF on the interval $[a, b]$ and its CDF are given by

$$f(x) = \frac{1}{b-a}$$

$$F(x) = \int_a^x \frac{1}{b-a}\,dx' = \frac{x-a}{b-a}$$



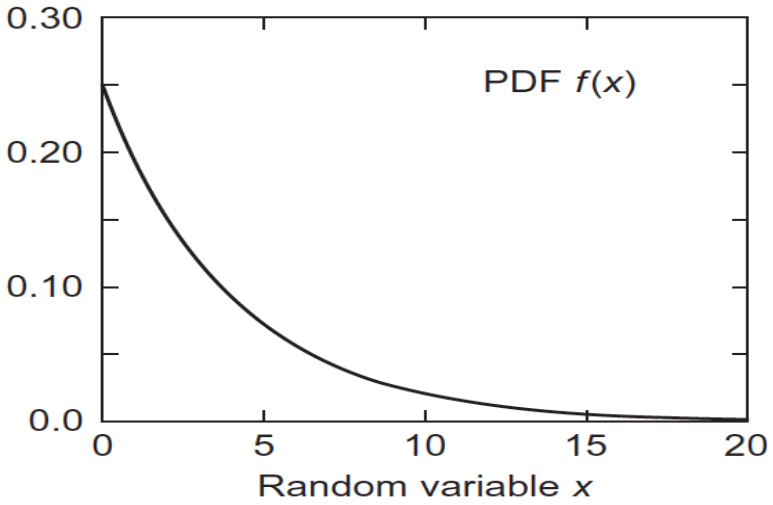


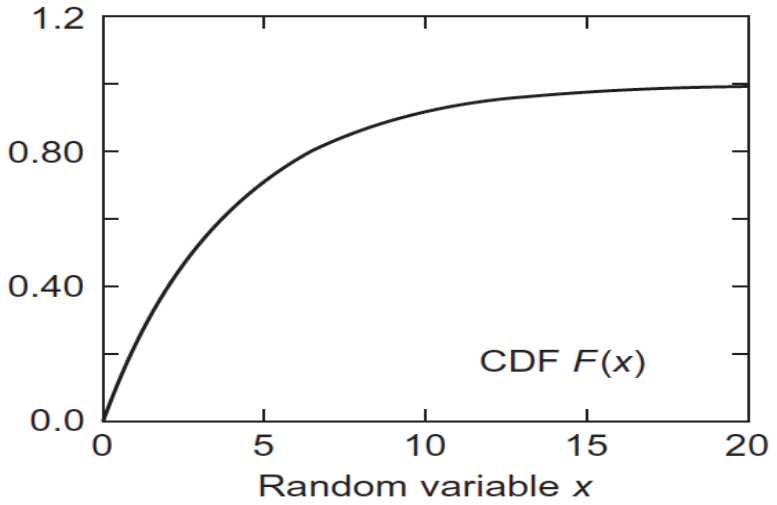
Some example distributions – Exponential PDF

The exponential PDF on the interval [0, ∞] and its CDF are given by

$$f(x) = f(x|\alpha) = \alpha e^{-\alpha x},$$

$$F(x) = \int_0^x \alpha e^{-\alpha x'}\, dx' = 1 - e^{-\alpha x}$$

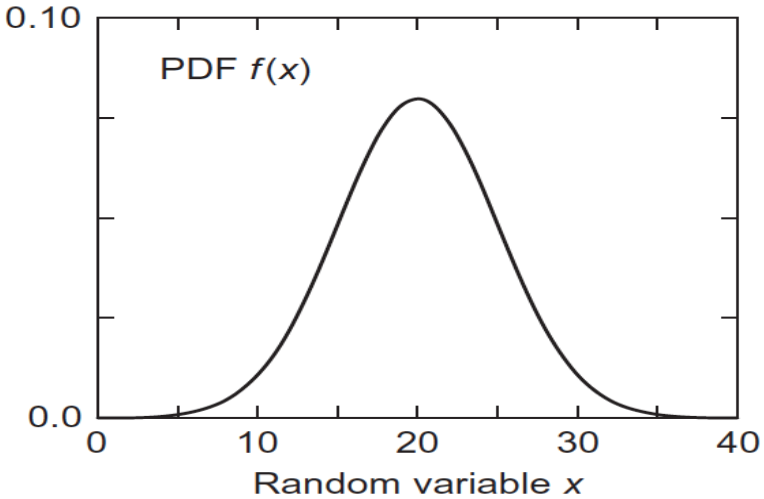


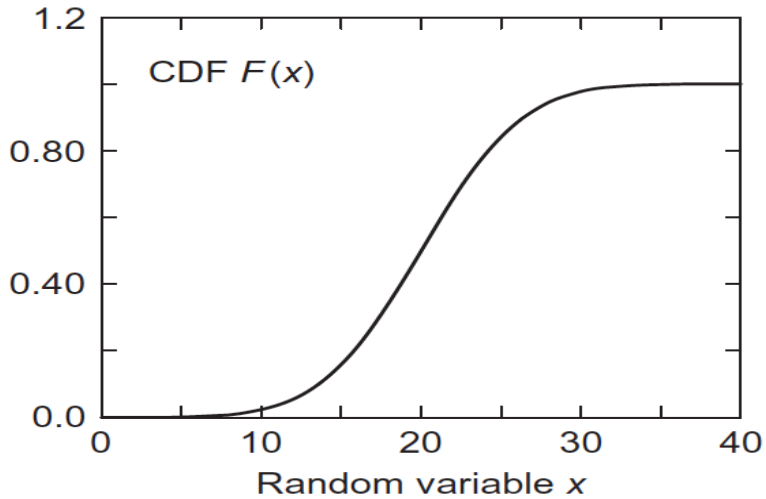


Some example distributions – Gaussian PDF

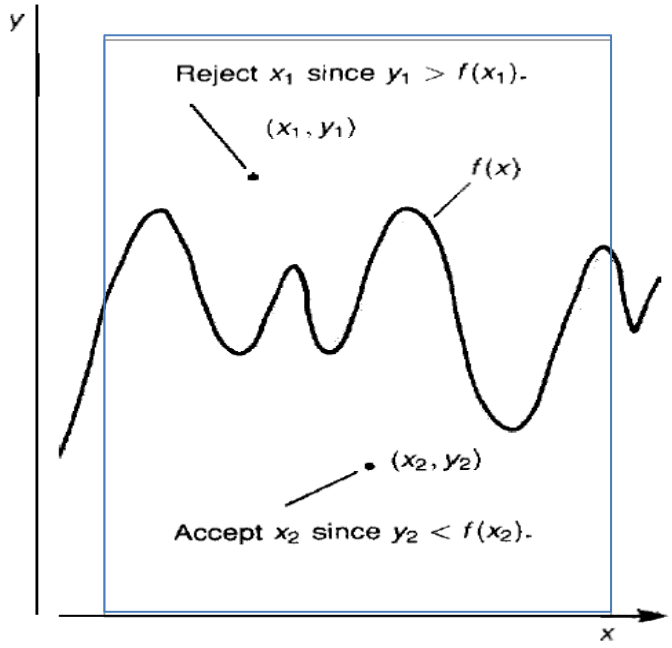
The Gaussian PDF on the interval [-∞, ∞] and its CDF are given by

$$f(x) = f(x|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \quad F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left[-\frac{(x'-\mu)^2}{2\sigma^2}\right] dx'$$





Acceptance-rejection method



- If the two numbers x_i and y_i are selected randomly from the range and domain, respectively, of the function f, then each pair of numbers represents a point in the function's coordinate plane (x, y).
- When $y_i > f(x_i)$ the point lies above the curve for $f(x)$, and x_i is rejected; when $y_i \leq f(x_i)$ the points lies on or below the curve, and x_i is accepted.
- Thus, the fraction of the accepted points is equal to the fraction of the area below the curve.
- This technique, first proposed by John Von Neumann, is also known as the acceptance-rejection method of generating random numbers for arbitrary Probability Density Function (PDF).

Random number generators (RNG)

- At the core of all Monte Carlo calculations is some mechanism to produce a long sequence of random numbers r_i that are uniformly distributed over the open interval [0,1). However, digital computers, by design, are incapable of producing random results.
- A true random sequence could, in principle, be obtained by coupling to our computer some external device that would produce a truly random signal. For example, we could use the time interval between two successive clicks of a Geiger counter placed near a radioactive source or use the “white noise” in some electronic circuit to provide truly random numbers.
- However, use of such a random number generator would not be practical!
  - First, the feasibility of having to couple some external device that generates a random signal to every computer we want to use for Monte Carlo calculations would be impractical.
  - But more important would be the impossibility of writing and debugging a Monte Carlo code if, on every run, a different sequence of random numbers were used. What is needed is a sequence of random numbers that is the same every time the program is run so that code errors can be found and so that the same results are produced when the same code is run on different computers.

Pseudo-random number generator

SLAC

An alternative to producing the same sequence of random numbers each time a program is run is to use a pseudo-random number generator. **Such a generator is a deterministic algorithm that, given the previous numbers (usually just the last number) in the sequence, the next number can be efficiently calculated**:

- $X_{n+1}=f(X_n,X_{n-1},X_{n-2},\ldots,X_0)$
- X_0 is usually called "seed"

Many pseudo-random number generators have been proposed and used over the years in a wide variety of Monte Carlo work. Designing better random number generators (the term "pseudo" is sometimes dropped from now on) is still an active area of research.

E.g. Mersenne Twister method (M. Matsumoto et al., 1998)
Fundamental ingredient of computing cryptography (e.g. financial transactions)

There have been many highly deficient random number generators used in many studies.

When you embark on Monte Carlo investigations, be well aware of the pedigree of the random number generator you are using.

Even more important, don't alter your random number generator unless you are very confident your changes are an improvement. The construction of random number generators is best left to the experts (try googling: RANDU).

pRNG: properties

A pRNG needs a seed to start a sequence. It will always produce the same sequence when initialized with that state.

The period of a pRNG is defined as the maximum, over all starting seeds, of the length of the repetition-free of the sequence. We want a large period (if the state of the pRNG is composed on n bits, the maximum period will be 2^n)

Most pRNG algorithms produce sequences which are uniformly distributed. We want an algorithm that does not show correlations on its output.

- Often the quality of a pRNG is judged performing a set of statistical tests.

Finally we want the algorithm to be fast and use as little memory as possible

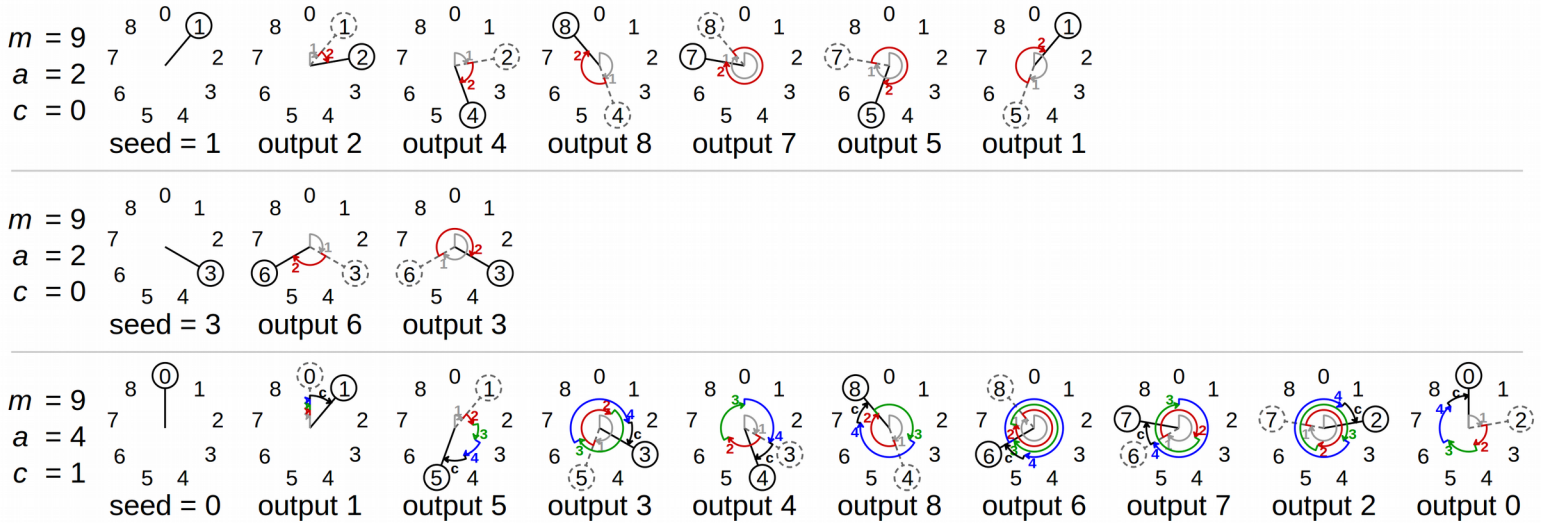
pRNG example: LCG

Linear Congruential Generator:

```
function lcg( Xn , a , c, m ):
    return (a*Xn+c) % m
```

The "art" is to find a good set of parameters: a,c,m

For example for GCC implementation (`rand()` in `stdlib.h`): $m=2^{31}$, a=1103515245 , c = 12345



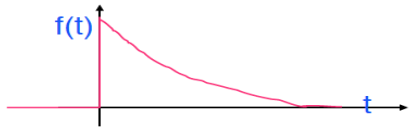
Basics of Monte Carlo Radiation Transport

Simplest case – decay in flight (1)

- Suppose an unstable particle of life time *t* is flying with initial momentum *p*.
  - Distance to travel before decay : $d = t\,v$
- The life time t is a random value with probability density function

$$f(t) = \frac{1}{\tau}\exp\left(-\frac{t}{\tau}\right) \quad t \geq 0$$

τ is the mean life of the particle



- *t* is determined in the general way that the cumulative distribution function itself is a random variable with uniform probability on [0,1)

$$r = F(t) = \int_{-\infty}^{t} f(u)du$$

- Thus, having a uniformly distributing random number *r* on [0,1), one can sample the value *t* with the probability density function *f*(*t*).

$$t = F^{-1}(r) = -\tau \ln(1 - r) \qquad 0 \leq r < 1$$

Simplest case – decay in flight (2)

- When the particle has traveled the $d = t\,v$, it decays.
- Decay of an unstable particle itself is a random process
  - For example:

    $\pi^+ \rightarrow \mu^+ \nu_\mu$ (99.9877 %)

    $\pi^+ \rightarrow \mu^+ \nu_\mu \gamma$ (2.00 x 10-4 %)

    $\pi^+ \rightarrow e^+ \nu_e$ (1.23 x 10-4 %)

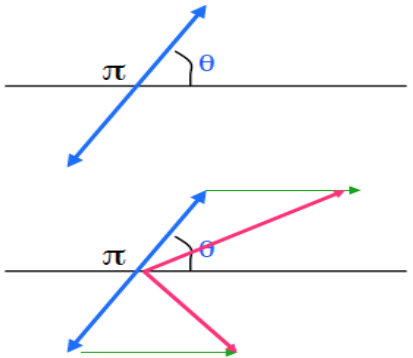
    $\pi^+ \rightarrow e^+ \nu_e \gamma$ (7.39 x 10-7 %)

    $\pi^+ \rightarrow e^+ \nu_e \pi^0$ (1.036 x 10-8 %)

    $\pi^+ \rightarrow e^+ \nu_e e^+ e^-$ (3.2 x 10-9 %)
- Select a decay channel by shooting a random number
- In the rest frame of the parent particle, rotate decay products in θ [0,π) and ϕ *[0,2π)* by shooting a pair of random numbers

$$d\Omega = \sin\theta\, d\theta\, d\phi$$

$$\theta = \cos^{-1}(r_1), \quad \phi = 2\pi \text{ x } r_2 \quad 0 \le r_1, r_2 <$$



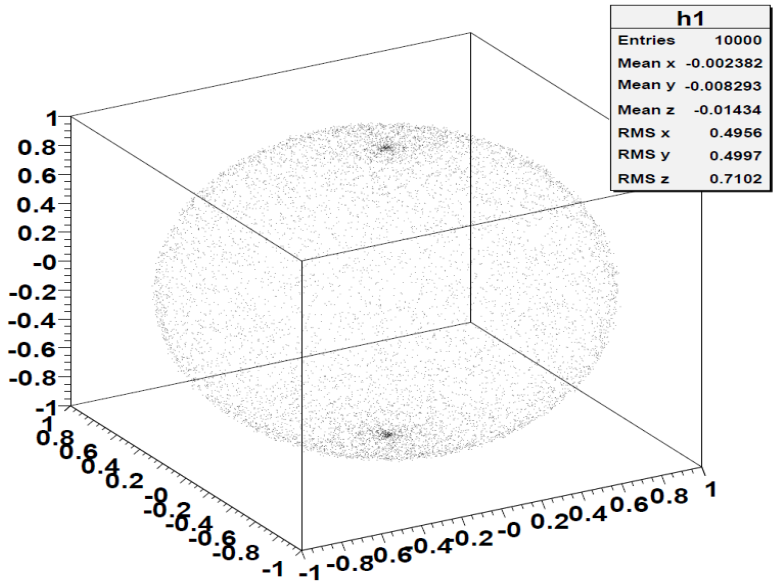
- Finally, Lorentz-boost the decay products
- You need at least 4 random numbers to simulate one decay in flight

Evenly distributed points on a sphere

SLAC

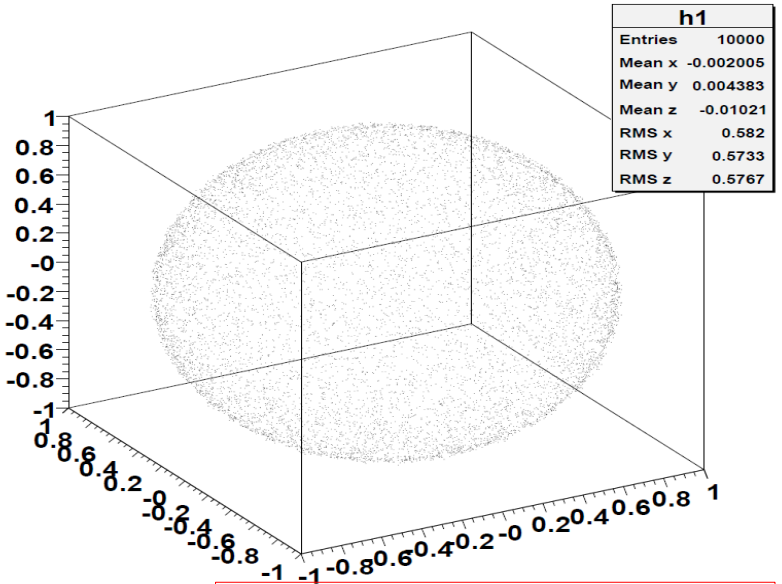
$\theta = \pi \times r_1, \phi = 2\pi \times r_2$

$0 \leq r_1, r_2 < 1$



$\theta = \cos^{-1}(r_1), \phi = 2\pi \times r_2$

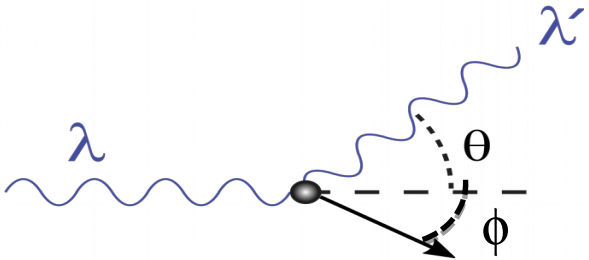
$0 \leq r_1, r_2 < 1$



$d\Omega = \sin\theta \, d\theta \, d\phi$

Compton scattering (1)

Compton scattering $e^- \gamma \rightarrow e^- \gamma$



Distance before Compton scattering, l, is a random value

Cross section per atom : $\sigma(E,z)$

Number of atoms per volume : $n = \rho N_A / A$

ρ : density, N_A : Avogadro number, A : atomic mass

Cross section per volume : $\eta(E, \rho) = n \sigma$

η is the probability if Compton interaction per unit length. $\lambda(E, \rho) = \eta^{-1}$ is the mean free path associated to the Compton scattering process.

The probability density function $f(l)$

$$f(l) = \eta \exp(-\eta l) = \frac{1}{\lambda} \exp\left(-\frac{l}{\lambda}\right)$$

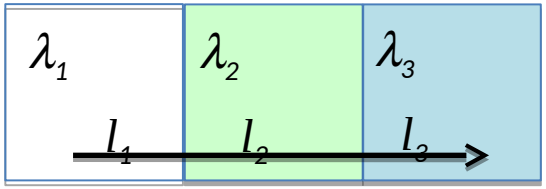
With a uniformly distributing random number r on [0,1), One can sample the distance l.

$$l = -\lambda \ln(r) \qquad 0 \le r < 1$$

Compton scattering (2)

- $\lambda(E, r)$ and l are material dependent. Distance measured by the unit of mean free path (n_λ) is independent.

$$n_\lambda = \frac{l_1}{\lambda_1} + \frac{l_2}{\lambda_2} + \frac{l_3}{\lambda_3} = \int_0^{end} \frac{dl}{\lambda(l)}$$



- n_λ is independent to the material and a random value with probability density function $f(n_\lambda) = \exp(-n_\lambda)$
  - Sample n_l at the origin of the particle $n_\lambda = -\ln(r) \quad 0 \le r < 1$
  - Update elapsed n_λ along the passage of the particle $n_\lambda = n_\lambda - l_i / \lambda_i$
  - Compton scattering happens at $n_\lambda = 0$

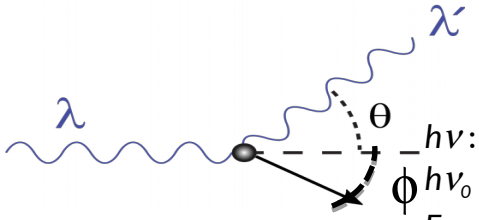
Compton scattering (3)

- The relation between photon deflection (q) and energy loss for Compton scattering is determined by the conservation of momentum and energy between the photon and

$$h\nu = \frac{h\nu_0}{1 + \left(\frac{h\nu_0}{m_e c^2}\right)(1 - \cos\theta)},$$

$$E = h\nu_0 - h\nu = m_e c^2 \frac{2(h\nu_0)^2 \cos^2\phi}{(h\nu_0 + m_e c^2)^2 - (h\nu_0)^2 \cos^2\phi},$$

$$\tan\phi = \frac{1}{1 + \left(\frac{h\nu_0}{m_e c^2}\right)} \cot\frac{\theta}{2},$$



$h\nu$: energy of incident photon
$h\nu_0$: energy of scattered photon
E : energy of recoil electron
m_e : rest mass of electron
c : speed of light

- For unpolarized photon, the Klein-Nishina angular distribution function per steradian of solid angle Ω

$$\frac{d\sigma_c^{KN}}{d\Omega}(\theta) = r_0^2 \frac{1 + \cos^2\theta}{2} \frac{1}{[1 + h\nu(1 - \cos\theta)]^2} \left\{1 + \frac{h\nu^2(1 - \cos\theta)^2}{(1 + \cos^2\theta)[1 + h\nu(1 - \cos\theta)]}\right\}$$

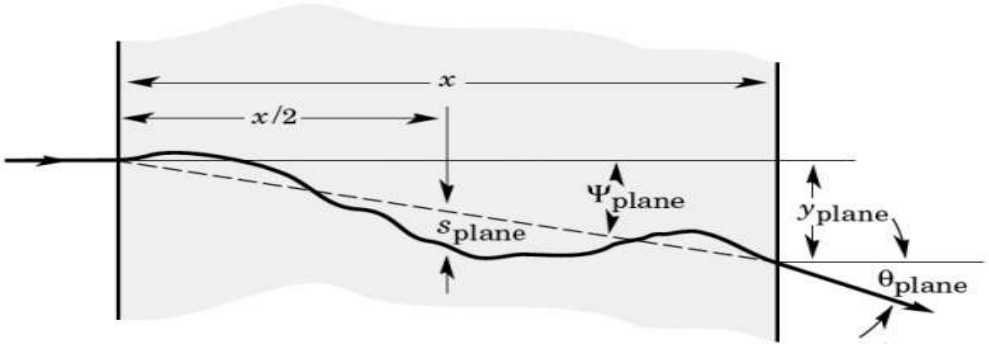
$$= \frac{1}{2} r_0^2 \left(\frac{k}{k_0}\right)^2 \left(\frac{k}{k_0} + \frac{k_0}{k} - \sin^2\theta\right) \qquad (cm^2 sr^{-1} electron^{-1}),$$

- One can use acceptance-rejection method to sample the distribution.

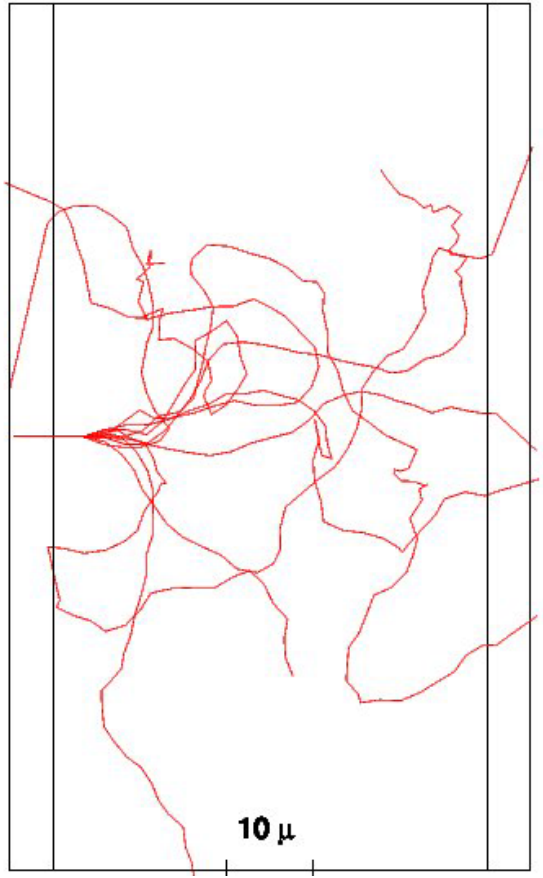
$$k_0 = \frac{h\nu_0}{m_e c^2}, \qquad k = \frac{h\nu}{m_e c^2}$$

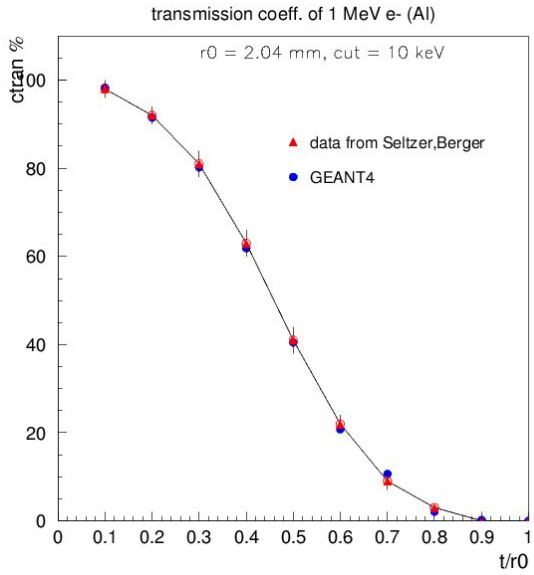
Multiple Coulomb scattering

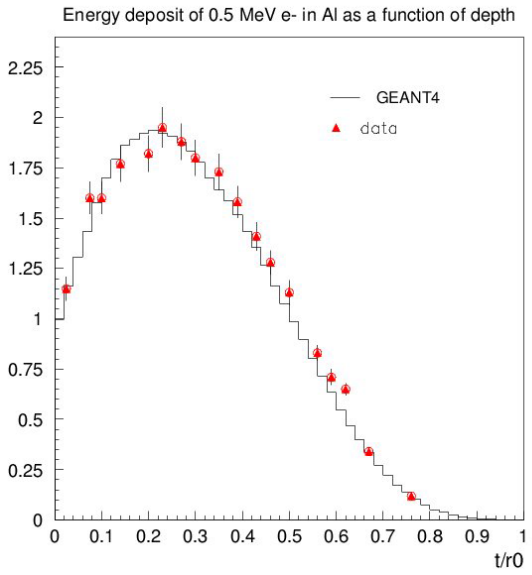
- Single Coulomb scattering : deflection of a charged particle in the Coulomb field of a nucleus.
  - Small deviation angle, practically no energy loss
  - In finite thickness, a particle suffers so many consecutive Coulomb scattering (> 10^6 interaction /mm in dense material).
- Simulate many interactions in a path segment of a given length s by a single computational step using the multiple-scattering approximation theories.
- Parameters:
  - Longitudinal displacement : x
  - Lateral displacement : r, ψ
  - True path length : τ
  - Angular deflection : θ, ϕ
- Theories:
  - Goudsmit-Saunderson theory
  - Lewis theory



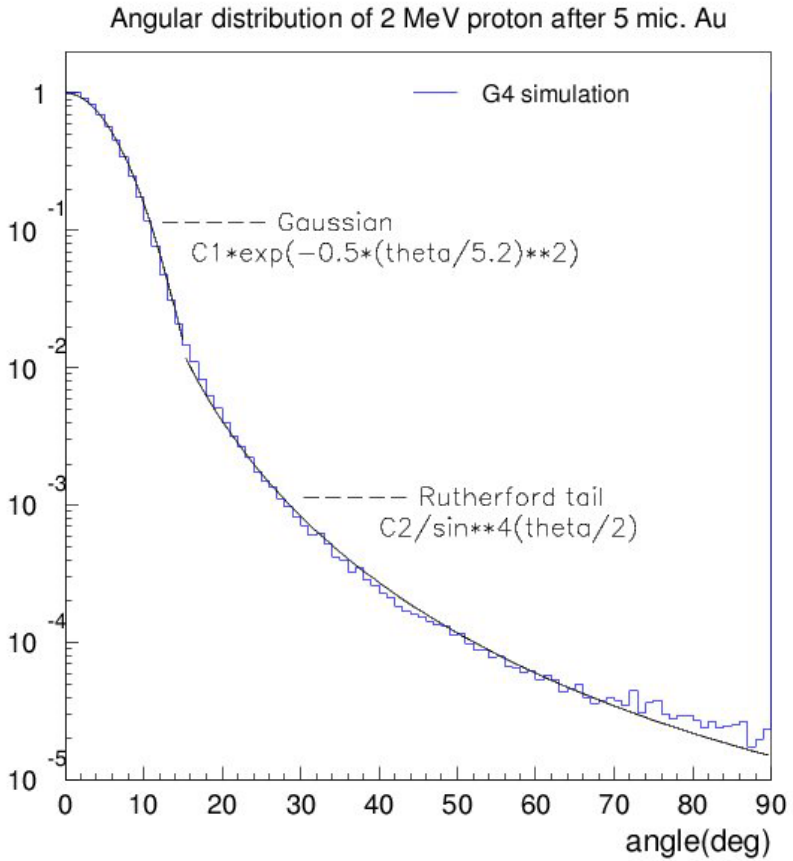
Multiple Coulomb scattering

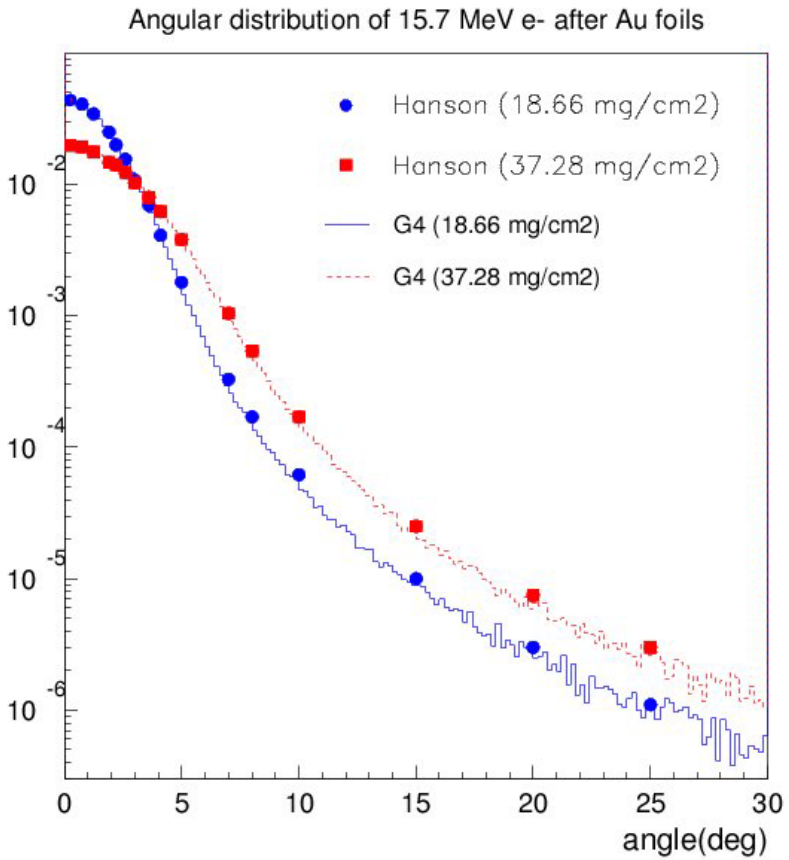






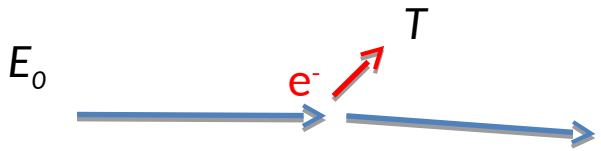
Multiple Coulomb scattering

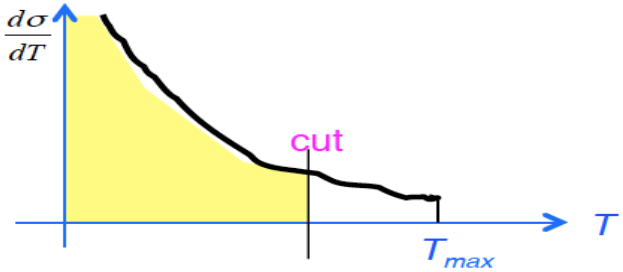




Ionization

- A charged particle collides with a quasi-free electron and knocks it off (δ-ray).





- $\frac{d\sigma(E_0,T)}{dT}$ is the cross-section of knocking-off an electron of energy T, and it has an infrared divergence.
- Maximum transferable kinetic energy $T_{max} = \frac{2 m_e c^2 \beta^2 \gamma^2}{1 + 2\gamma m_e / m_0 + (m_e / m_0)^2}$
  - e.g. Muon with E = 1.06 GeV (γ = 10) : T_{max} ~ 100 MeV
- Cross-section for δ-ray above cutoff : $\sigma(E_0, cut) = \int_{cut}^{T_{max}} \frac{d\sigma(E_0,T)}{dT} dT$

Continuous energy loss

- Ignore generating d-rays below T_{cut}. Energies taken by multiple sub-cutoff electrons are considered as continuous energy loss of the projectile particle.
- Mean energy of e- below cutoff : $\langle T_{soft}(E_0)\rangle_{cut} = \frac{1}{\sigma_{tot}} \int_0^{cut} \frac{d\sigma(E_0,T)}{dT} T dT$
- Mean energy loss by the projectile due to sub-cutoff e- : $\left(\frac{dE}{dx}\right)_{cut} = n_{at}\sigma_{tot}\langle T_{soft}\rangle_{cut}$
- Restricted Linear Energy Transfer : $L_{cut} = -\left(\frac{dE}{dx}\right)_{cut}$
- Stopping Power = Unrestricted Linear Energy Transfer : L_{Tmax}
- Bethe-Bloch Formula

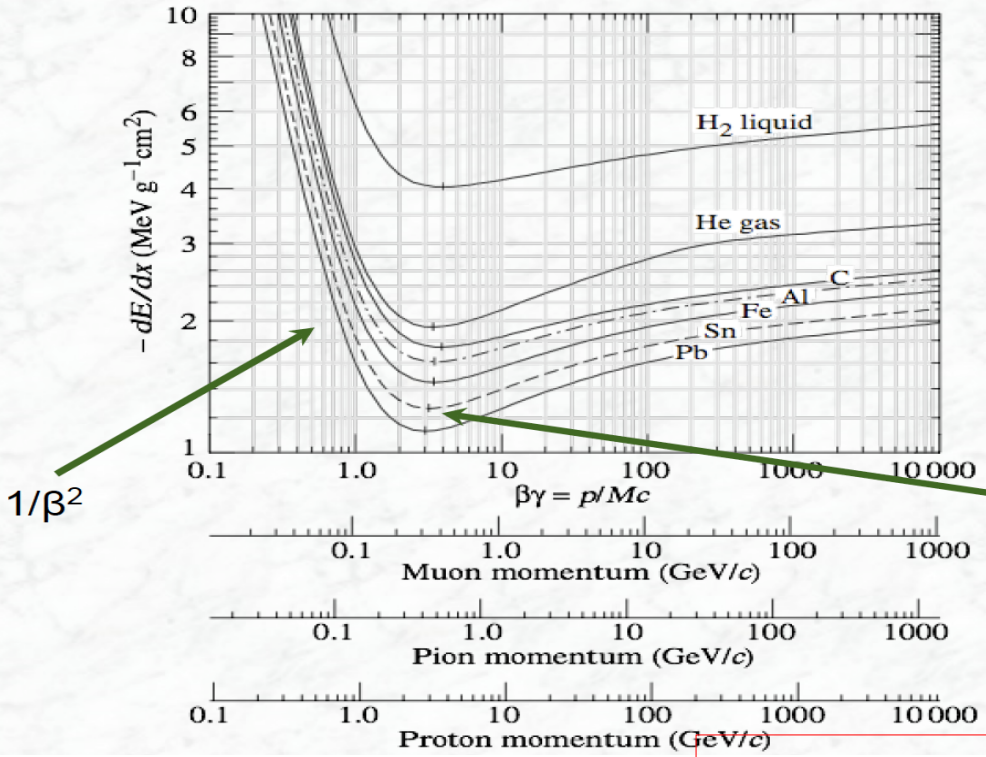
$$-\frac{dE}{dx} = Kz^2\frac{Z}{A}\frac{1}{\beta^2}\left[\frac{1}{2}\ln\frac{2m_e c^2\beta^2\gamma^2 T_{max}}{I^2} - \beta^2 - \frac{\delta(\beta\gamma)}{2}\right]$$

where : $K = 4\pi N_A r_e^2 m_e c^2 \sim 0.307075$ (MeV g^{-1} cm^2)

z : charge of incident particle Z : atomic number of target material

I : characteristic ionization constant (material dependent)

Mean energy loss



$1/\beta^2$

Bethe-Bloch formula:

$$-\frac{dE}{dx} = Kz^2\frac{Z}{A}\frac{1}{\beta^2}[\frac{1}{2}\ln f(\beta) - \beta^2 - \frac{\delta(\beta\gamma)}{2}]$$

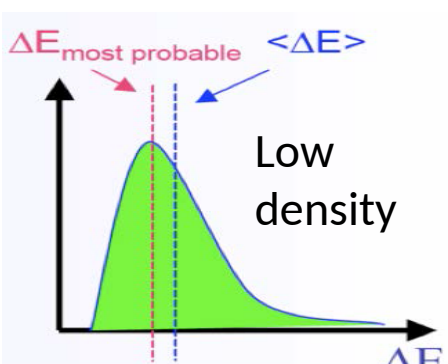
Except in hydrogen, particles of the same velocity have similar energy loss in different materials.

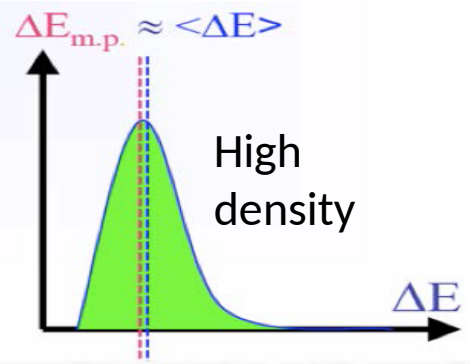
The minimum in ionisation occurs at βγ = 3.5 to 3.0, as Z goes from 7 to 100

Minimum Ionizing Particle : $\beta\gamma \sim 3.5$ ($\beta\gamma = p/m$)

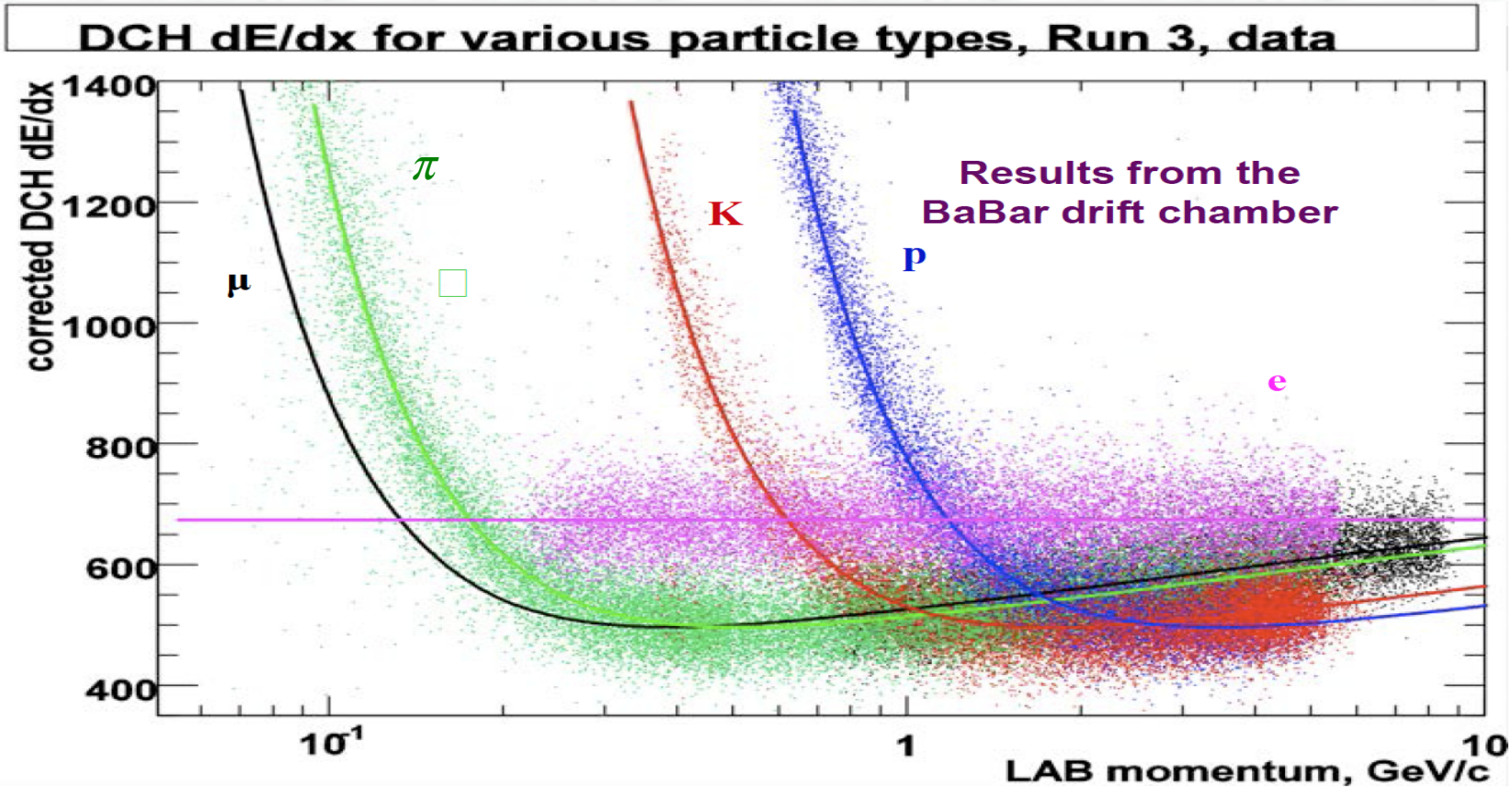
Fluctuations in energy loss

- A real experiment cannot measure <*dE/dx*>.
  - It measures the energy deposition ΔE in finite thickness Δx.
    - ➡Sampling of energy loss distribution
- For thin detector or low density material, the energy loss distribution shows large fluctuation with long tail (Landau tail).
  - e.g. silicon sensor 300 μm thick : ΔE_{mip} ~ 82 keV, $<\Delta E>$ ~ 115 keV
- For thick detector and high density material, energy loss distribution shows a Gaussian-like distribution.



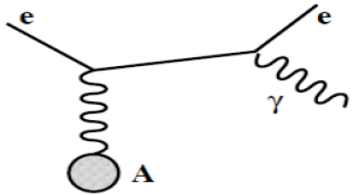


dE/dx measurements can be used to identify particle type



Energy loss due to bremsstrahlung

- High energy charged particles undergo an additional energy loss (in addition to ionization energy loss) due to bremsstrahlung, i.e. radiation of photons, in the Coulomb field of the atomic nuclei.



$$-\left.\frac{dE}{dx}\right|_{Brems} = 4\alpha N_A \left(\frac{e^2}{mc^2}\right)^2 \ln\frac{183}{Z^{1/3}} \frac{Z(Z+1)}{A} Q^2 E$$

- Most important characters of energy loss due to bremsstrahlung.
  - Proportional to $1/m^2$: light particle radiates more energy (e^- / μ^- ~ 40000)
  - Proportional to E : bremsstrahlung is significant in higher energy

- Critical energy : $$-\left.\frac{dE}{dx}\right|_{ion}(E_c) = -\left.\frac{dE}{dx}\right|_{brems}(E_c)$$

  - E_c for e^- ~ 19 MeV, E_c for μ^- ~ 1 TeV

Radiation length

- Radiation length (X_0) : $$\frac{1}{X_0} = 4\alpha N_A \left(\frac{e^2}{m_e c^2}\right)^2 \ln\frac{183}{Z^{1/3}} \frac{Z(Z+1)}{A}$$

- With radiation length, *dE/dx* due to bremsstrahlung for electron can be simplified.

$$-\left.\frac{dE}{dx}\right|_{Brems} := \frac{1}{X_0} E \quad \Rightarrow \quad E(x) = E_0 e^{-x/X_0}$$

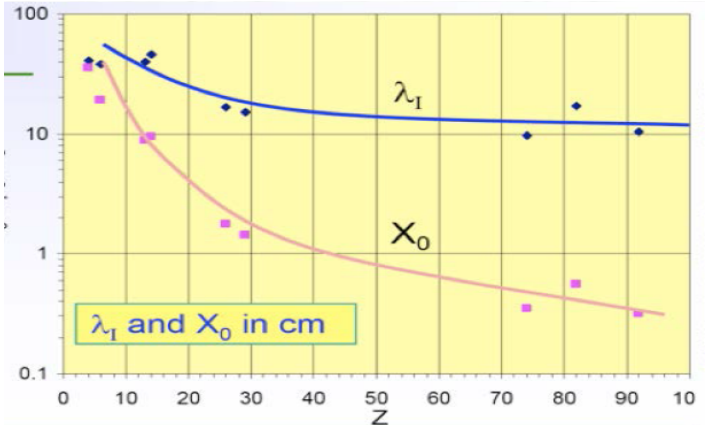
Material	X_0 (cm)	Material	X_0 (cm)
H_2 gas	700000	He	530000
C	18.8	Fe	1.76
Cu	1.43	W	0.35
Pb	0.56	Air	30000
SiO_2	12	Water	36

Overview of hadronic interaction

- Both charged and neutral hadrons can interact with material via the strong force (a.k.a. hadronic interaction).
- Hadronic interaction is characterized by the hadronic interaction length λ_{had}
  - Hadronic interaction length is a material constant.
  - Intensity of hadron beam attenuates in matter due to hadronic interaction.

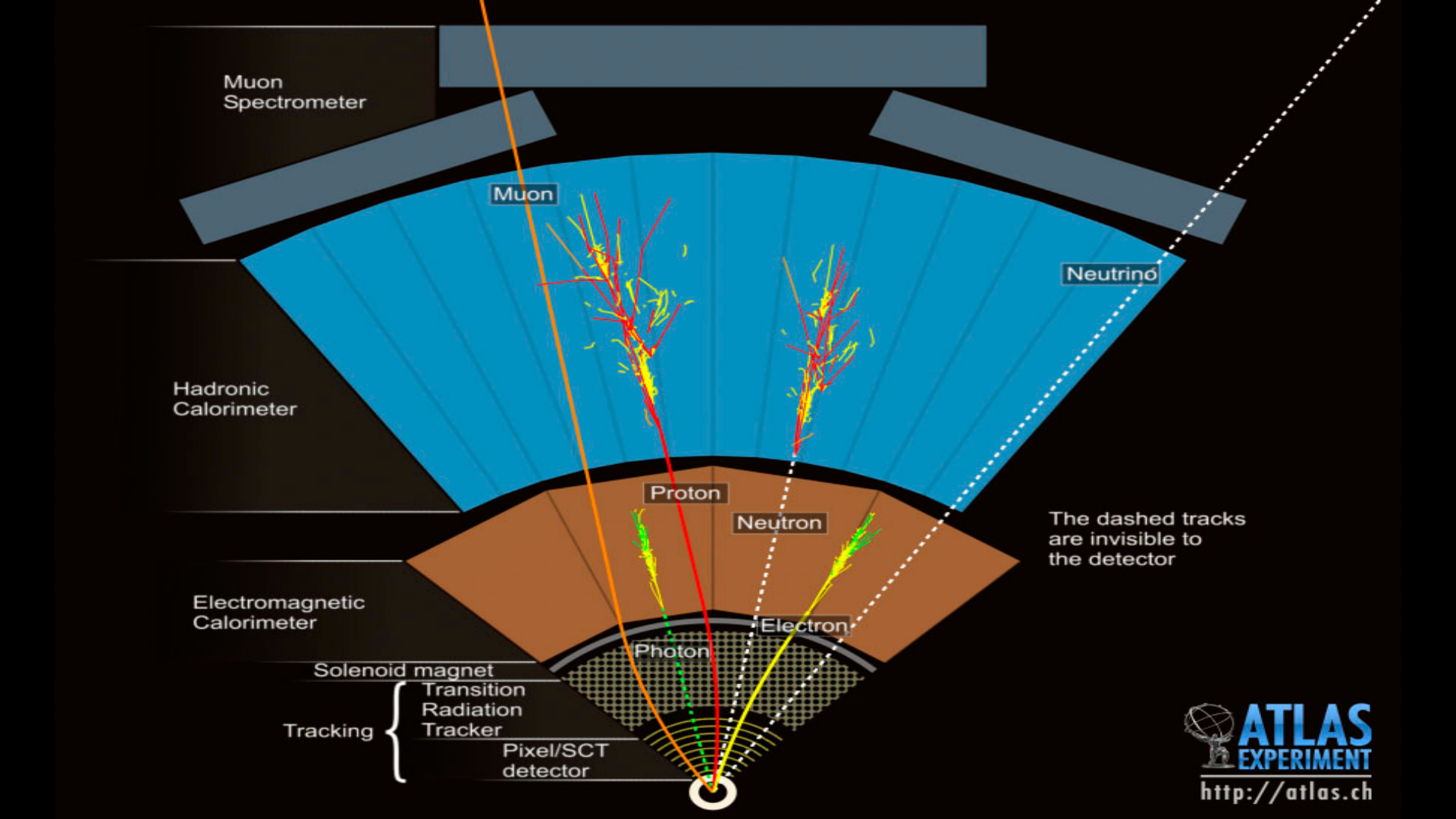
$$I(x) = I_0\, e^{-x/\lambda}, \quad 1/\lambda_{had} = \sigma_{\text{inel}}\, N_A\, \rho / A \quad (\lambda_{had} \sim 35\, A^{1/3}\,(\text{cm}))$$

Material	X_0 (cm)	λ_{had} (cm)
C	18.8	38.1
Fe	1.76	16.76
Cu	1.43	15.06
W	0.35	9.59
Pb	0.56	17.09
Water	36	91



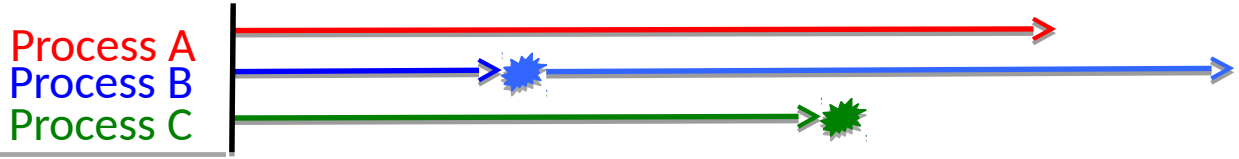
For high-Z material, λ is 10-30 times larger than X_0

➡more material is needed to stop hadrons compared to electrons

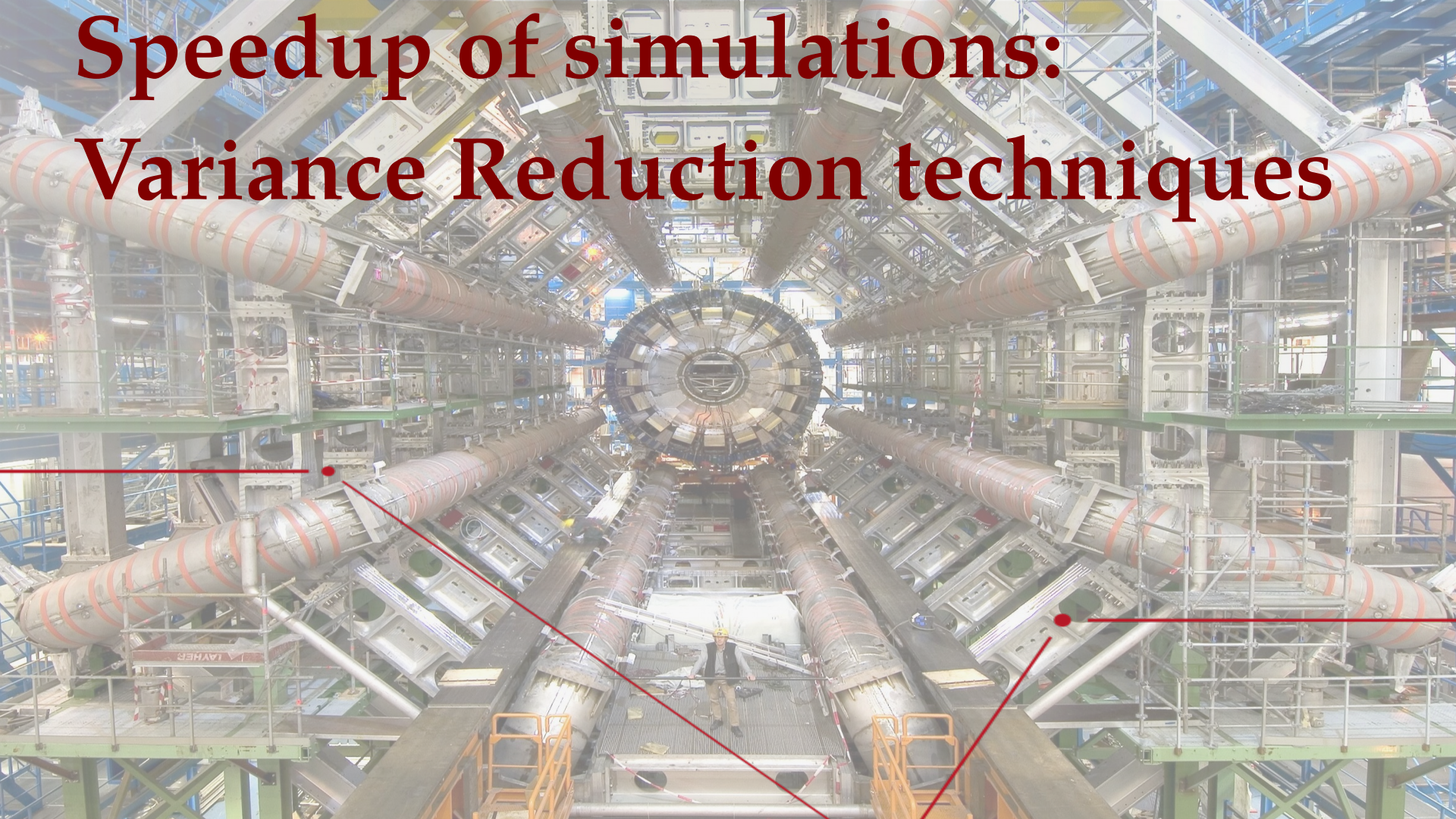

Muon Spectrometer
Muon
Neutrino
Hadronic Calorimeter
Proton
Neutron
The dashed tracks are invisible to the detector
Electromagnetic Calorimeter
Electron
Photon
Solenoid magnet
Tracking
Transition Radiation Tracker
Pixel/SCT detector
ATLAS EXPERIMENT
http://atlas.ch

Process competition

- “Ordinary” physics makes point-like interaction. Given many physics processes have chances to occur, one needs to make a fair competition among these eligible processes.
- Given PDF of each process, one can sample the path length normalized by mean free path (radiation length, hadronic interaction length, decay time, etc.) for each physics process.
- Compare the path lengths proposed by all physics processes. The process that proposes the shortest length occurs.
  - Given the length is normalized, competition should be made by the actual length (normalized length x mean free path of the material).
- Once the particle experiences an interaction by a physics process, the path length for that process is re-sampled, while proposed path lengths of other processes are reduced by the length traveled.
- Continuous processes (continuous energy loss, multiple scattering, Cherenkov radiation, etc.) are applied cumulatively.



Speedup of simulations: Variance Reduction techniques



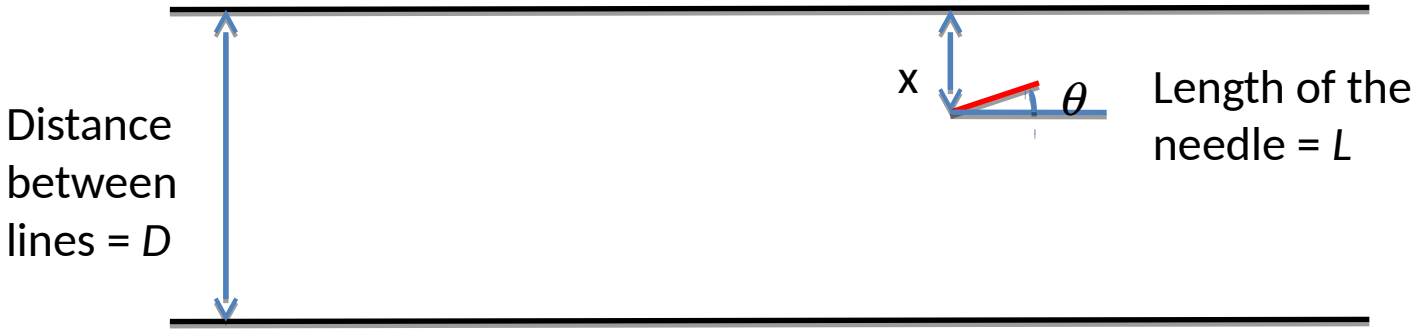
Buffon's needle – once again

- Suppose the distance between lines (D) is much larger than the length of the needle (L). In naïve simulation, the needle's location (x) is sampled uniformly over [0,D).

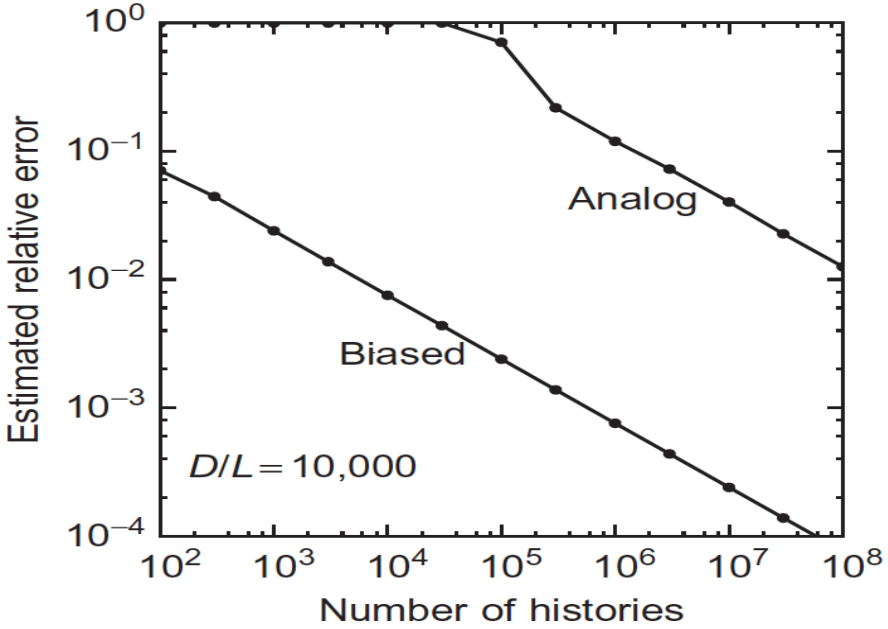
  $\pi \sim (2\,L\,/\,D) * (\,h\,/\,n\,)$

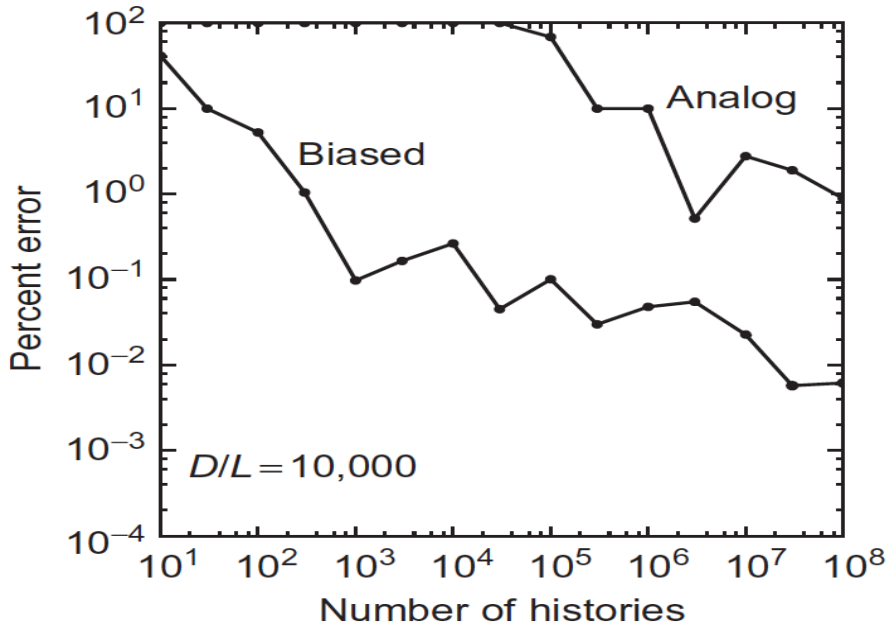
- However, the needle has no chance to hit lines for $L < x < D - L$. Also, symmetry shows the probability of hitting lines for [0,D/2) is equal to [D/2,D).
- Once can estimate π by sampling x over [0,L), while, given the probability of $0 \leq x < L$ is L/(D/2), each successful count should be multiplied by the weight L/(D/2).

  $\pi \sim (2\,L\,/\,D) * (\,h * (D\,/\,2\,L)\,/\,n\,)$



Buffon's needle – case of D / *L* = *10,000*





Analogue simulation / non-analogue simulation

SLAC

- The real power of Monte Carlo is that the sampling procedure can be intentionally biased toward the region where the integrand is large or to produce simulated histories that have a better chance of creating a rare event, such as Buffon's needle falling on widely spaced lines.
  - "Analogue simulation" : follows the natural PDF
  - "Non-analogue (a.k.a. biased) simulation" : biased sampling
- Of course, with such biasing, the scoring then must be corrected by assigning weights to each history in order to produce a corrected, unbiased estimate of the expected value.
- In such non-analog Monte Carlo analyses, the sample variance $\sigma^2(z)$ of the estimated expectation value z is reduced compared to that obtained by an unbiased or analog analysis.
  - In the Buffon's needle example, biasing a simulation of problem with widely spaced lines to force all dropped needles to have one end within a needle's length of a grid line was seen to reduce the relative error by two orders of magnitude over that of a purely analog simulation.

Mean and variance

- Two important measures of ant PDF $f(x)$ are its mean μ and variance σ^2.
- The mean μ is the expected or averaged value of x defined as

$$\langle x \rangle \equiv E(x) \equiv \mu(x) \equiv \int_a^b x f(x)\, dx.$$

- The variance describes the spread of the random variable x from the mean and defined as

$$\sigma^2(x) \equiv \langle [x - \langle x \rangle]^2 \rangle = \int_a^b [x - \langle x \rangle]^2 f(x)\, dx.$$

$$= \int_a^b \left[x^2 - 2x\langle x \rangle + \langle x \rangle^2 \right] f(x)\, dx$$

$$= \int_a^b x^2 f(x)\, dx - 2\langle x \rangle \int_a^b x f(x)\, dx + \langle x \rangle^2 \int_a^b f(x)\, dx$$

$$= \langle x^2 \rangle - \langle x \rangle^2$$

Note: $\int_a^b x^2 f(x)\, dx = \langle x^2 \rangle \qquad \int_a^b f(x)\, dx = 1$

Variance reduction techniques

- The aim of Monte Carlo calculation is to seek an estimate $\langle z \rangle$ to an expected value $\bar{z}$. The goal of variance reduction technique is to produce a more precise estimate than could be obtained in a purely analogue calculation with the same computational effort.

$$\langle z \rangle \equiv \int_V z(\mathbf{x}) f(\mathbf{x})\, d\mathbf{x} \simeq \bar{z} = \frac{1}{N} \sum_{i=1}^{N} z(\mathbf{x}_i) \qquad s(\bar{z}) = \frac{1}{\sqrt{N-1}} \sqrt{\overline{z^2} - \bar{z}^2}$$

- Because both $\overline{z^2}$ and $\bar{z}^2$ are always positive, the sample variance s^2 can be reduced by reducing their difference.
- Thus, the various variance reduction techniques are directed, ultimately, to minimizing the quantity $\overline{z^2} - \bar{z}^2$
  - Note that, in principle, it is possible to attain zero variance, if $\overline{z^2} = \bar{z}^2$ which occurs if every history yields the sample mean. But this is not very likely. However, it is possible to reduce substantially the variance among histories by introducing various biases into a Monte Carlo calculation.

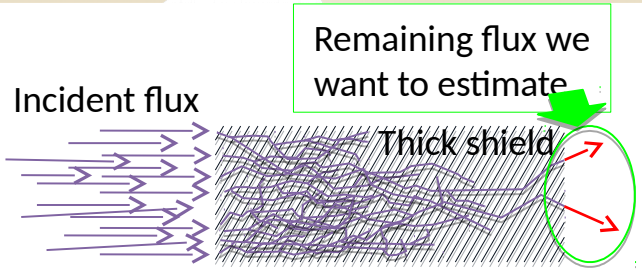
Variance reduction techniques

There are simulation problems where the event we are interested in is very rare due to physics and/or geometry.
Over the years, many clever variance reduction techniques have been developed for performing biased Monte Carlo calculations.
The introduction of variance reduction methods into Monte Carlo calculations can make otherwise impossible Monte Carlo problems solvable. However, use of these variance reduction techniques requires skill and experience.

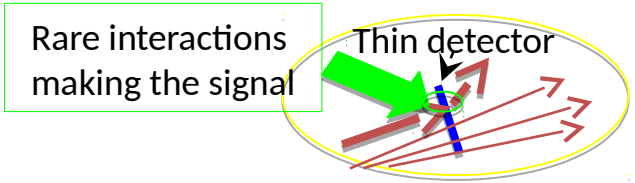
- Non-analog Monte Carlo, despite having a rigorous statistical basis, is, in many ways, an “art” form and cannot be used blindly.

Example use-cases of variance reduction

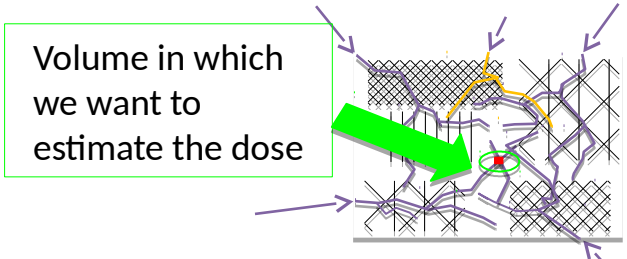
- Efficiency of a radiation shielding
  - E.g. large flux entering to a thick shield
  - Lots of interactions : compute intensive
  - Very few particles escape



- Response of thin detector
  - E.g. compact neutron detector
  - Most of particles pass through without interaction
  - Signal is made by the interaction



- Dose in a very small component in a large setup
  - E.g. an IC chip in a large satellite in cosmic radiation environment
  - Most of the incident radiation do not reach to the IC chip



Some variance reduction schemes - 1

Truncation Methods:

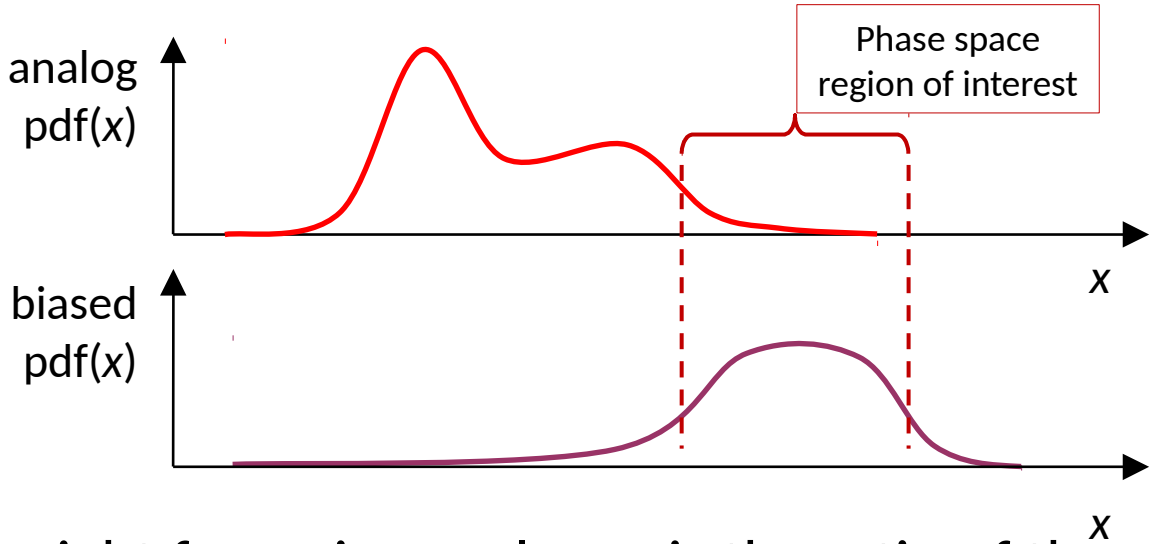
- This is the simplest of all variance reduction techniques. In this approach, the underlying physical models from which random samples are taken are simplified so that each history or sample takes, on average, less computer time.
  - For example, some detail in the geometry far from the scoring region generally has little effect on z and so needs not be modeled.
- Equivalently, a detailed model of the low probability tail of the PDF sampling distribution is not needed if $z(x)$ is small in the tail region.
- Model simplification is a “brute force” approach, because quadrupling the number of histories only halves the standard deviation. Moreover, physical insight into a particular problem is required to introduce simplifications that reduce the calculation complexity for parts of sampling space that have little effect on the expected value.
- The only way to verify that some simplification has negligible effect on the estimate of z is to perform the calculation with and without the simplification.

Transform the Problem:

- In some calculations, problem symmetry or other features of the problem can be used to create an equivalent problem that has the same expected value but that can be treated by Monte Carlo much more efficiently. Often such a transformation produces far better variance reduction than any other method.
- However, it takes considerable understanding of the fundamental processes governing the problem to find such a transformation. A good example is importance sampling for a radiation shielding problem.

Importance Sampling

In an « importance sampling » technique, the analog PDF is replaced by biased PDF:



$$w(x) = \frac{\text{analog-pdf}(x)}{\text{biased-pdf}(x)}$$

The weight for a given value x, is the ratio of the analog over the biased distribution values at x.

Transforming Buffen's needle

- Let's take $z = D/L$. One can write the probability of hitting lines (p) as such.

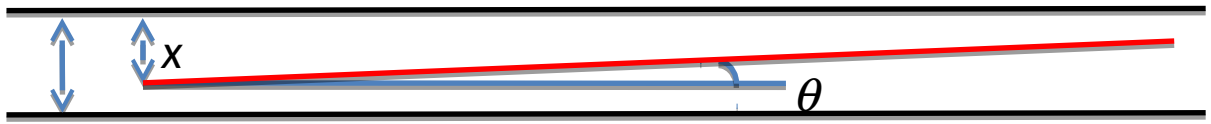
$$p = \frac{2}{\pi z}, \quad \text{if } z \geq 1$$

$$p = \frac{2\left[1 + z\arccos z - \sqrt{1 - z^2}\right]}{\pi z}, \quad \text{if } z < 1$$

$$\approx 1 - \frac{z}{\pi}\left(1 + \frac{z^2}{12} \pm \cdots\right), \quad \text{if } z \ll 1$$

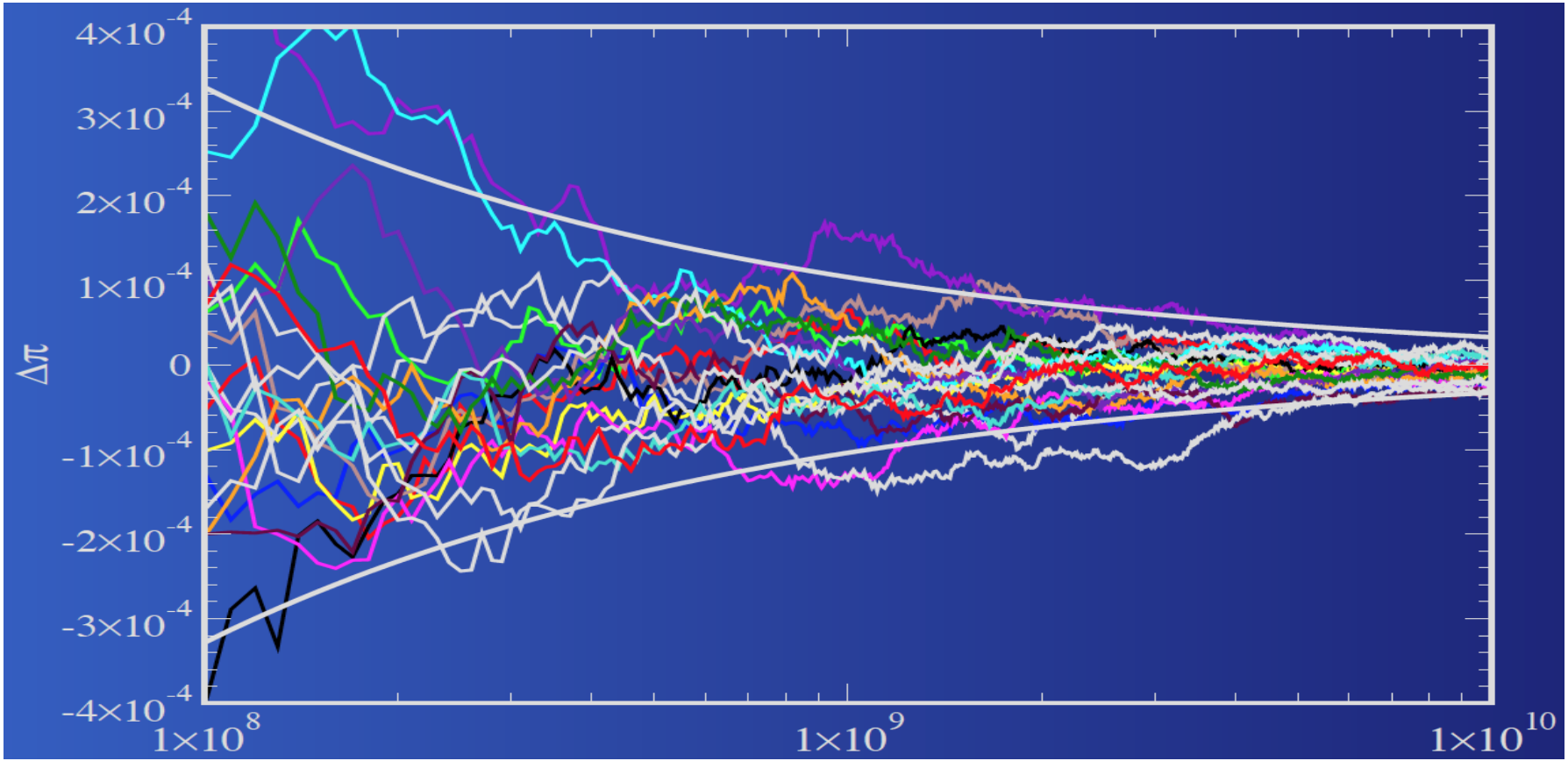
- Obviously $z << 1$ gives the best efficiency.

Distance between lines = D

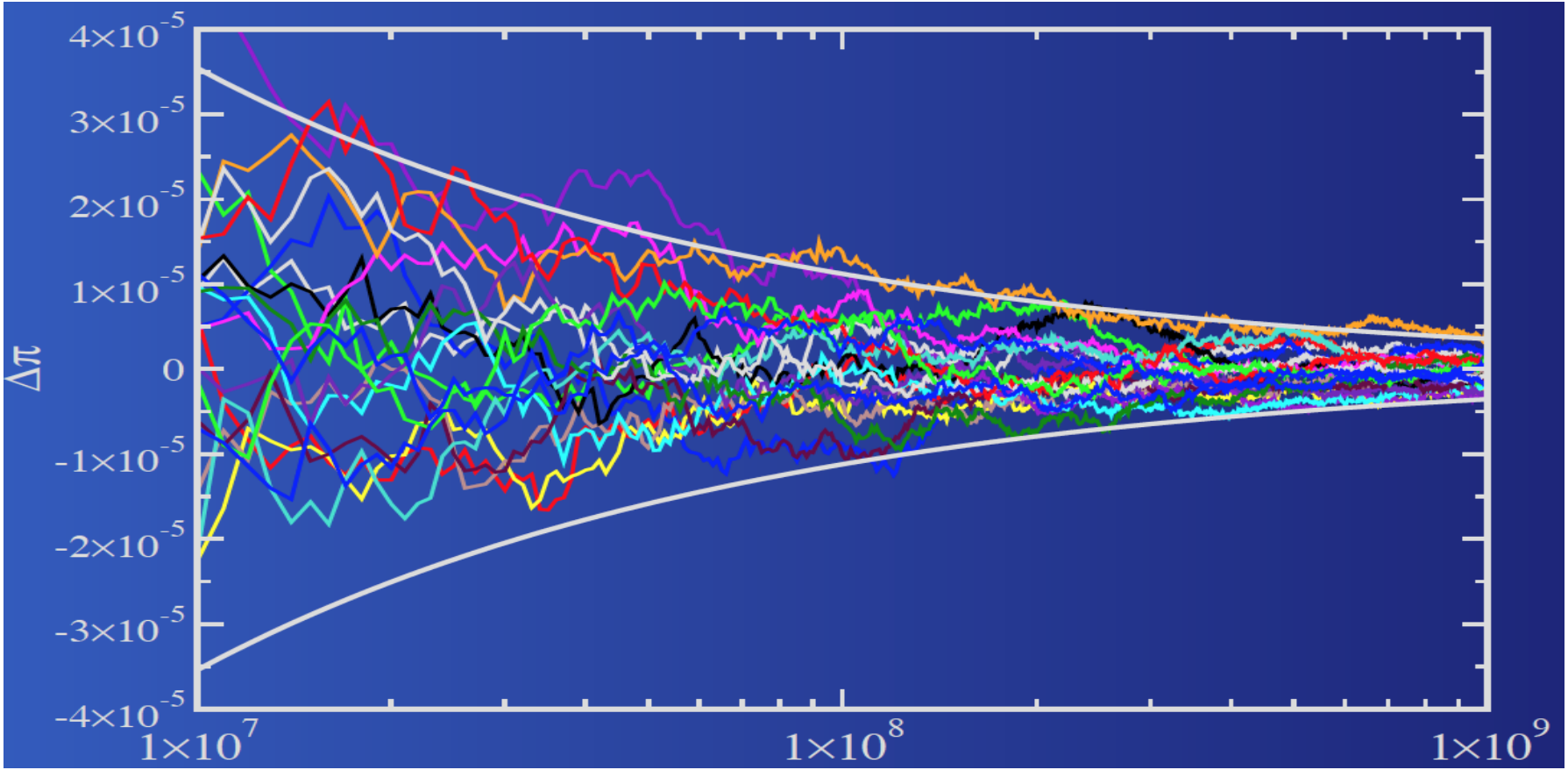


Length of the needle = L

$z > 1$



$z << 1$



Some variance reduction schemes - 2

SLAC

Modified Sampling Methods:

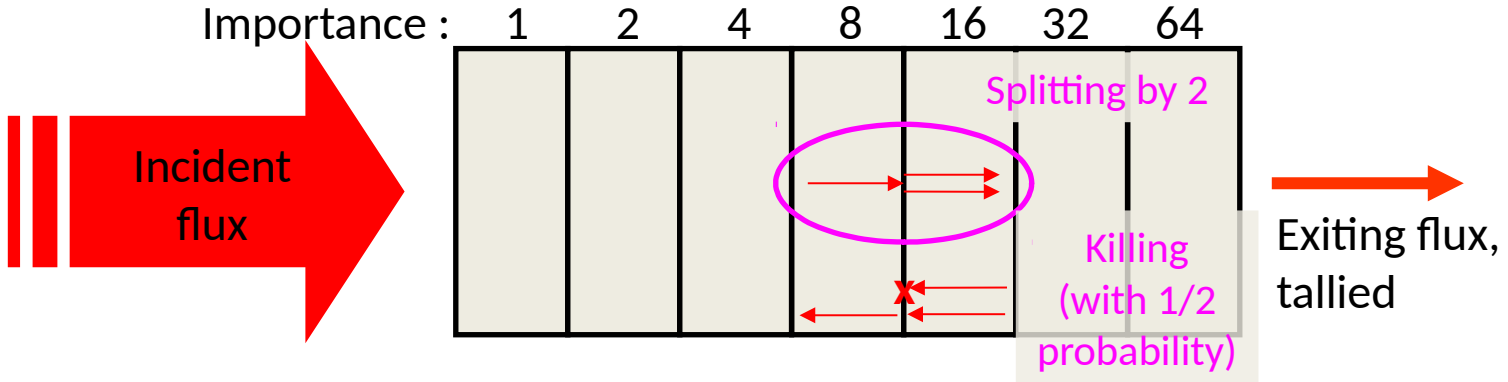
- Most variance reduction techniques modify the underlying sampling PDFs to bias the sampling. Such biasing is used to increase the likelihood that a history produces a non-zero score, i.e., a non-zero $z(x_i)$.
- Biased sampling can still produce an unbiased score provided each history is assigned a “weight” that is adjusted to compensate for any biases introduced at the various steps in the history. By recording a history’s weight in the tally, an unbiased value of is achieved.

History Control Methods:

- Two very powerful variance reduction techniques are to use Russian roulette and/or splitting to alter, on-the-fly, the number of histories that can potentially produce a non-zero tally. As a history moves through phase space, it is possibly killed if it enters a sub-region from which it is probably producing zero score, or it is possibly split into multiple histories if a history leaving this sub-region is likely to yield a nonzero score.
- Similarly, in biased sampling, Russian roulette and splitting can be used to eliminate histories that have too small a weight or split if the history obtains too large a weight. This technique is often used in combinations with other variance reduction methods.

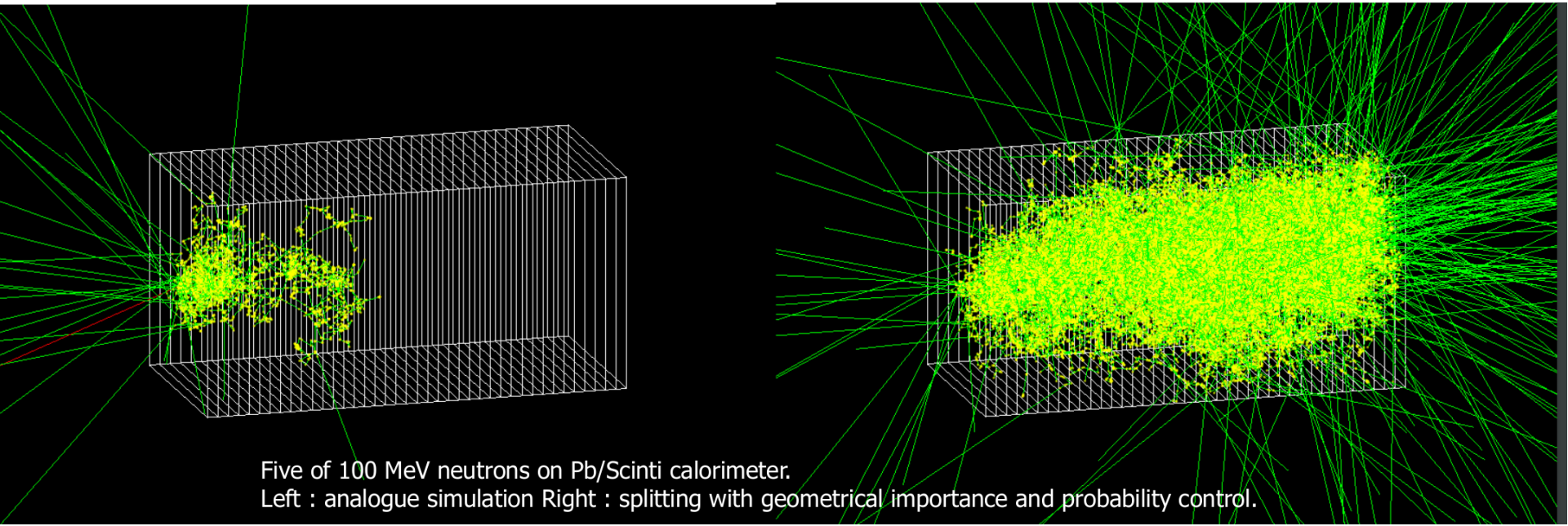
Splitting / Russian Roulette with geometrical importance

- Slice a thick volume and weigh importance to each slice.
  - Typically weight is power of 2.
- When a particle proceeds to more important slice, the particle is duplicated with half weight.
- When a particle proceeds to less important slice, the particle is killed with 50% probability. If the particle survives, its weight is doubled.



Splitting / Russian Roulette with geometrical importance

SLAC



Five of 100 MeV neutrons on Pb/Scinti calorimeter.
Left : analogue simulation Right : splitting with geometrical importance and probability control.

Weight window

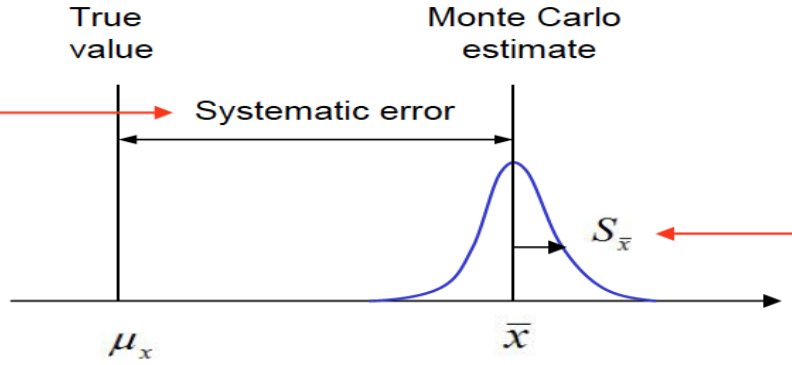
- Watch the weight of each particle.
- If the weight becomes too high;
  - It makes too large contribution to the estimation and slow the conversion
  - Split the particle with half the weight
- If the weight becomes too low;
  - It does not contribute to the estimation and thus it's a waste of CPU
  - Russian Roulette with 50% of probability and double the weight if survives

Accuracy, precision, relative error and figure of merit

Accuracy

$$\mu_x - \bar{x}$$

is a measure of how close the expected value <x> is to the true physical quantity μ_x being estimated. The difference between them is called the systematic error, which is seldom known.



Precision

$$S_{\bar{x}} = \sqrt{\frac{\overline{x^2} - \bar{x}^2}{N}} \propto \frac{1}{\sqrt{N}}$$

refers to the uncertainty of the Monte Carlo estimate and not to the accuracy. It is possible to calculate a highly precise result that is far from the physical truth because nature has not been modelled faithfully.

Relative error

is the measure of the calculation (statistical) precision of the Monte Carlo result.

$$R = \frac{S_{\bar{x}}}{\bar{x}} \propto \frac{1}{\sqrt{N}}$$

Guidelines for Interpreting the Relative Error:

Range of R	Quality of the result
0.5 – 1.0	Garbage
0.2 – 0.5	Factor of a few
0.1 – 0.2	Questionable
< 0.1	Generally reliable except for point detector
< 0.05	Generally reliable for point detector

Figure of Merit

$$FOM = \frac{1}{R^2 T} \approx const$$

Here T ~ N is the computer time needed to sample N histories.

FOM serves:

- as the reliability indicator for tally (it must be constant except for small statistical variations),
- as the measure of the efficiency of the Monte Carlo calculation (the higher FOM the better the efficiency), and
- as a useful tool for estimating the time needed to achieve given statistical precision.

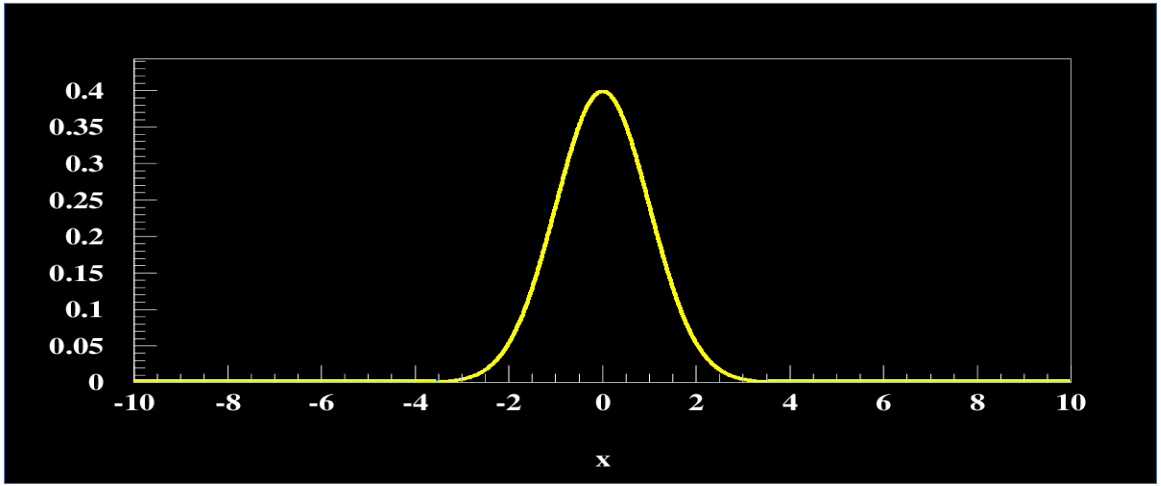
Convergence

A « toy » example of a bad sampling:

We want to estimate the mean value of the unknown « yellow » distribution.

As we don't know where it is, we try to sample it with the « blue » distribution

And we know how to compute the weight



Convergence

SLAC

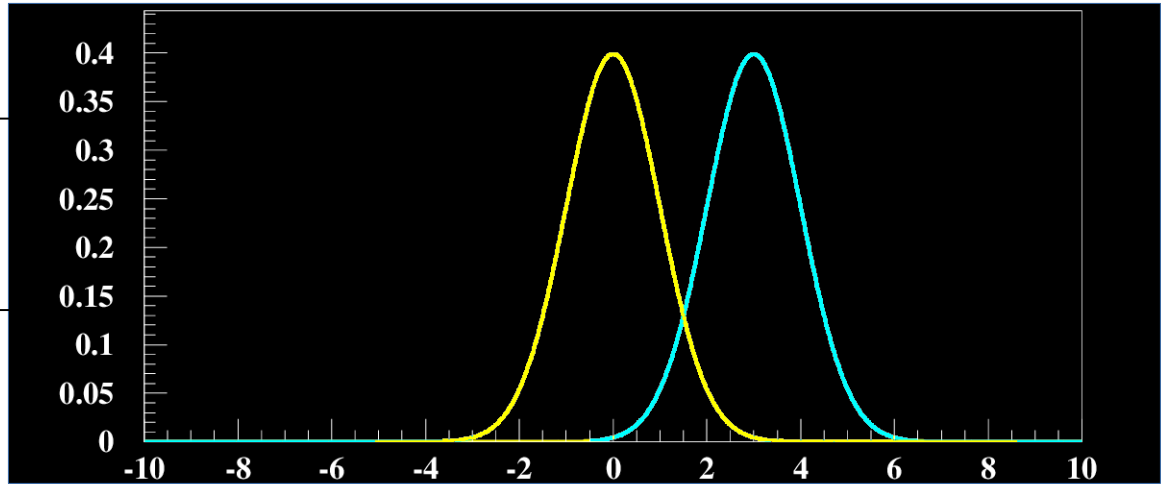
A « toy » example of a bad sampling:

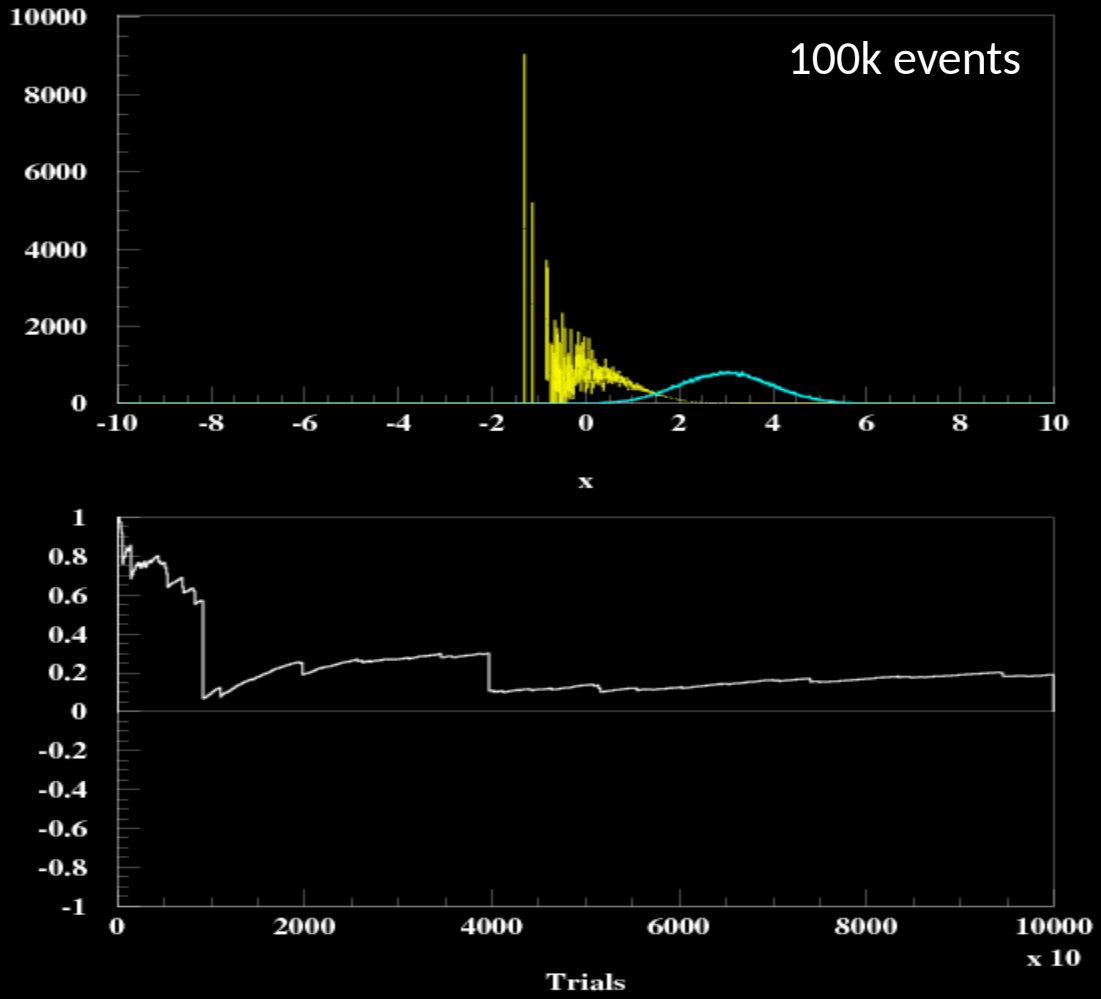
We want to estimate the mean value of the unknown « yellow » distribution.

As we don't know where it is, we try to sample it with the « blue » distribution

And we know how to compute the weight

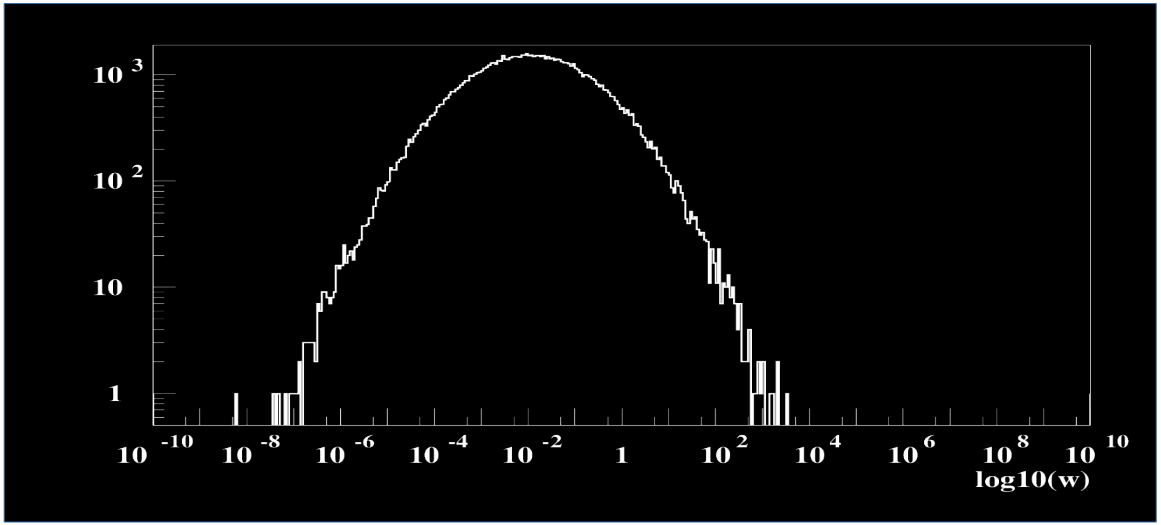
$$w(x) = \frac{\text{analog-pdf}(x)}{\text{biased-pdf}(x)}$$



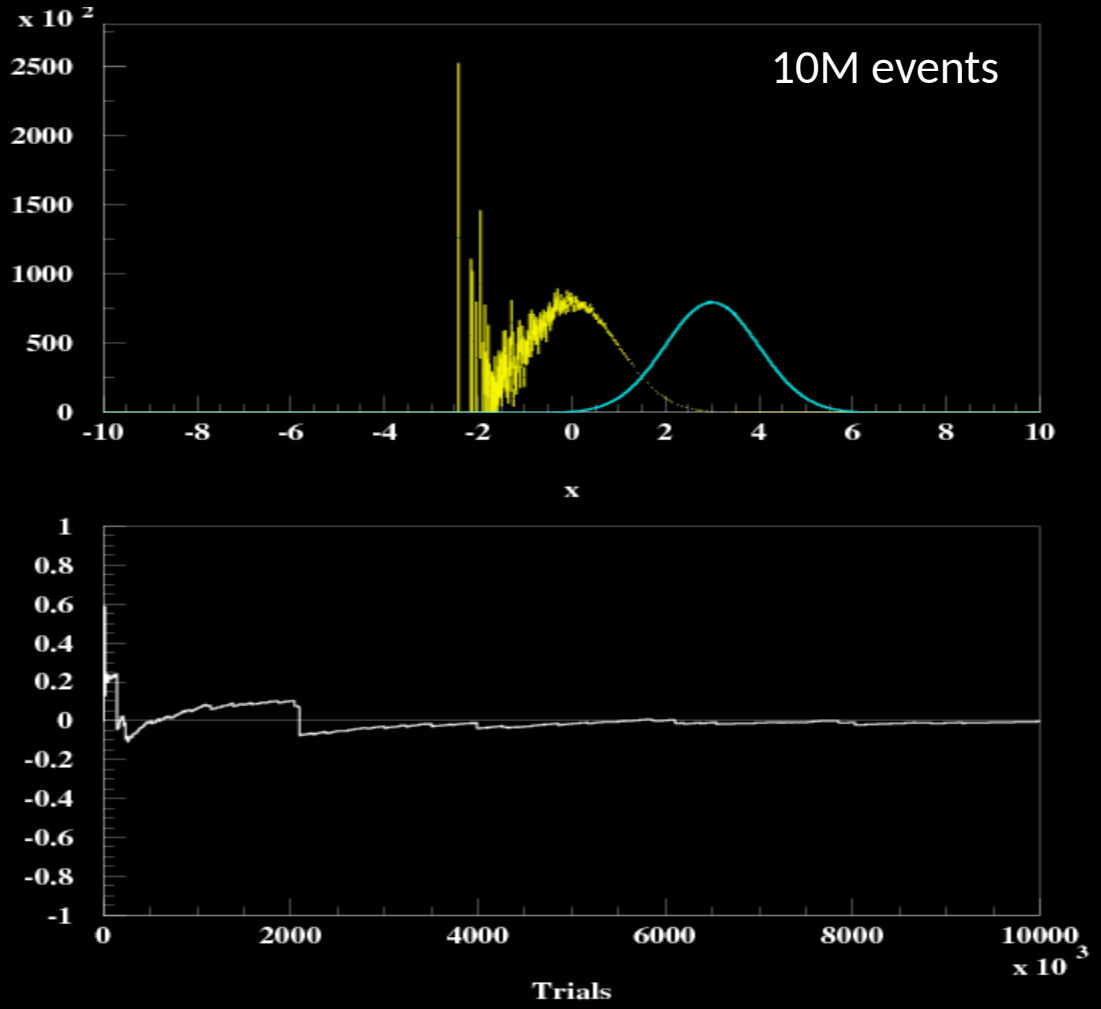

100k events
10000
8000
6000
4000
2000
0
-10
-8
-6
-4
-2
0
2
4
6
8
10
x
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0
2000
4000
6000
8000
10000
x 10
Trials

Weight distribution…

Lots of tiny weights…



Few huge ones…


x 10 2
2500
2000
1500
1000
500
0
-10
-8
-6
-4
-2
0
2
4
6
8
10
x
10M events
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0
2000
4000
6000
8000
10000
x 10 3
Trials

Some qualitative observations

SLAC

Observations that can help spotting the inappropriate sampling:

- We have a wide variety of weights
- Many tiny ones: Waste of time to determine the mean value
- Few huge ones: Responsible for jumps
  - Such problemw –if can not be improved- could at least be alleviated with a weight window technique

We have huge weights from time to time:

- Temptation would be to "dismiss" these events
- But in our case, these events bring down the mean value to the correct value
- These are not wrong per-se, but their presence is a sign of a bad sampling

We observe monotonic increase of the mean value

- We are sampling only "one side" of the problem

Part 2

Basic components

A detector simulation program requires at least the following three components:

- Geometry description module: to describe the experimental setup in terms of shapes, materials, relative positioning
- Physics modules: to cover all particles, energies and interaction types of interest
- Primary definition/generation: to describe what are the properties (species, 4-momenta) of the first particles that "appear" in the setup (can be provided by an external tool, e.g. a *generator* -PYTHIA, HERWIG,...- for HEP)

A system (user-hooks) to interact with the simulation and extract the physics quantities must be provided (e.g. scorers that record energy deposits in specified regions)

Alternative, useful components:

- Analysis tools to create histograms and store data in files
- Visualization drivers (to display geometry and possibily tracks and doses)
- System integration tools: MPI interfaces to submit jobs on clusters, scripting/macro systems

Semi-classical Approach

SLAC

In Geant4, a particle that flies through a detector is **treated as a classical particle**, i.e. not a wave function, but a point-like object which has a well-defined momentum at each instant:

- Space-time position (t, x, y, z)
- Energy-momentum (E, px, py, pz)

This is a reasonable approximation, given that in most practical situations particles are seen as "tracks" in macroscopic detectors

Geant4 is based on a semi-classical approach, because the particles are treated classically, but **their interactions** - cross sections and final states - **often take into account the results of quantum-mechanical effects**

Simulation in Steps

SLAC

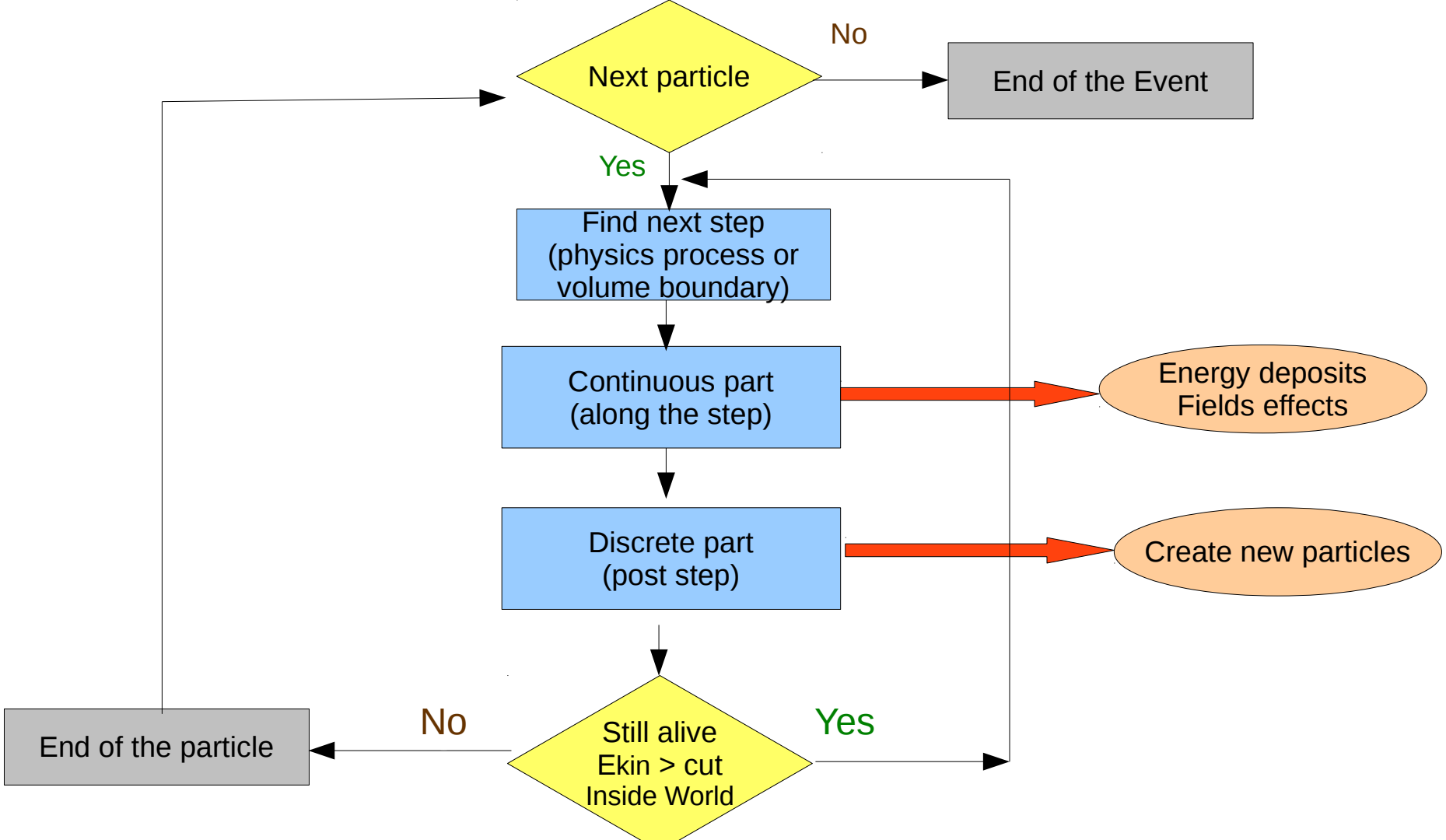
Treat **one particle at the time**
Treat a particle **in steps**

For each step

1) the step length is determined by the cross sections of the physics processes and the geometrical boundaries;
2) if new particles are created, add them to the list of particles to be transported;
3) local energy deposit; effect of magnetic and electric fields;
4) if the particle is destroyed by the interaction, or it reaches the end of the apparatus, or its energy is below a (tracking) threshold, then the simulation of this particle is over;

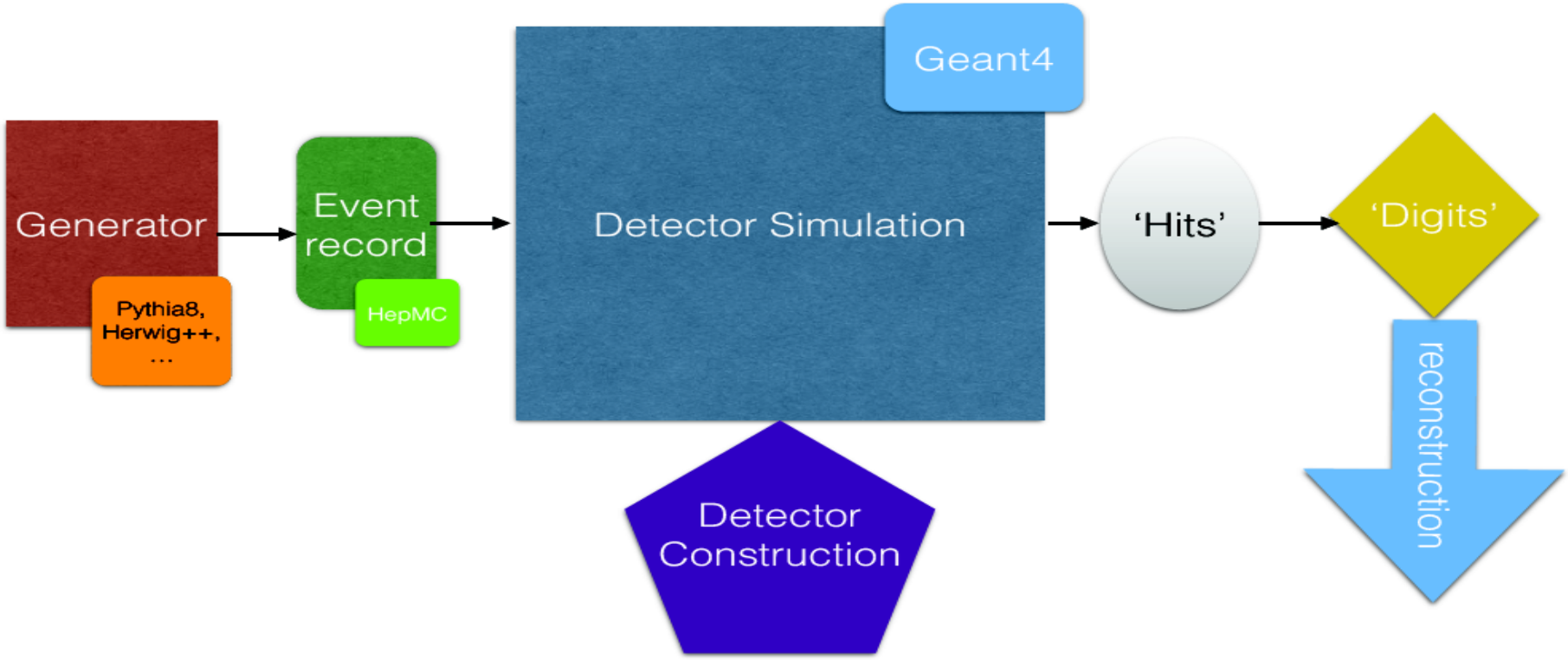
Repeat for next step or for a new particle

Outputs: - new particles created
- local energy deposits throughout the detector

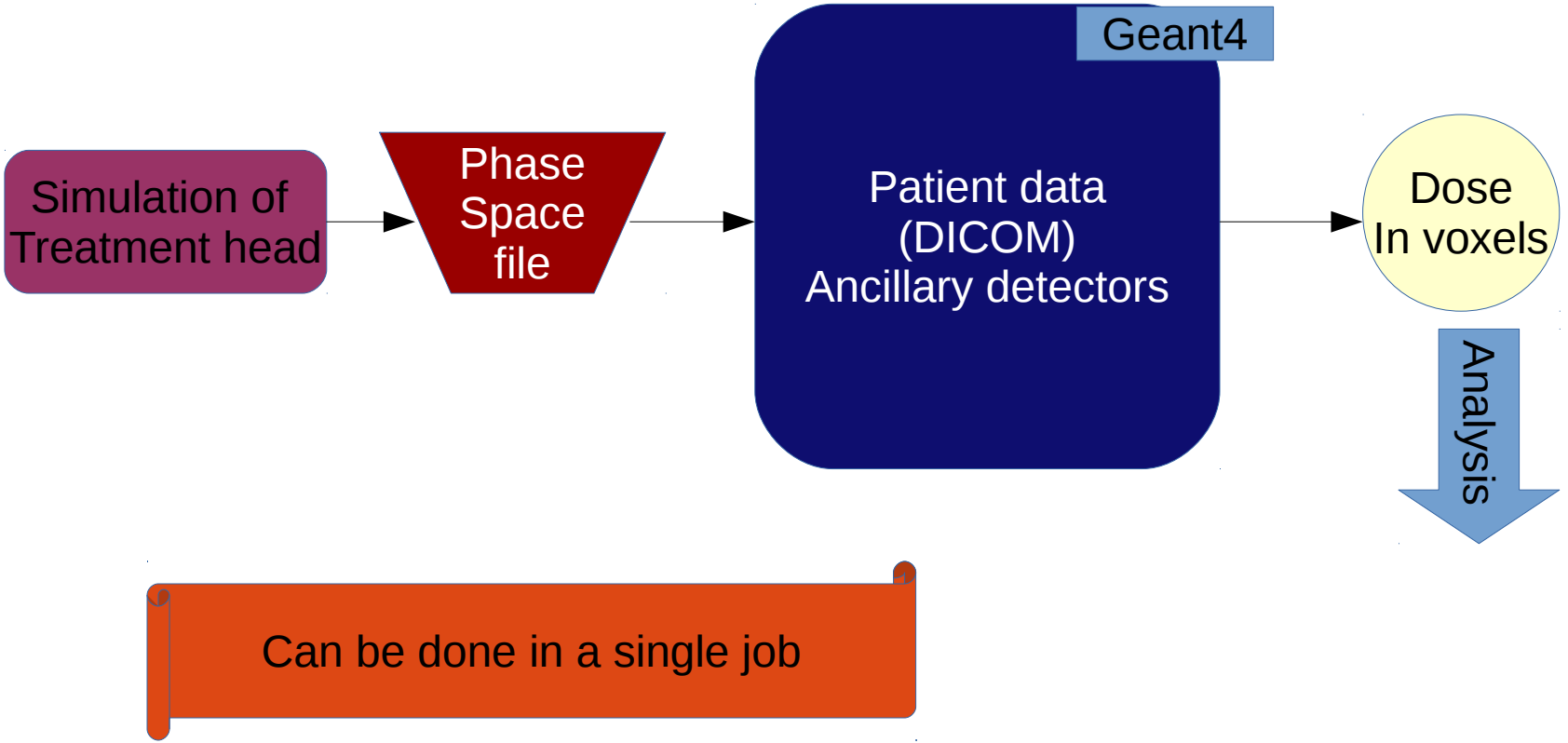

Next particle
No
End of the Event
Yes
Find next step
(physics process or
volume boundary)
Continuous part
(along the step)
Energy deposits
Fields effects
Discrete part
(post step)
Create new particles
Still alive
Ekin > cut
Inside World
No
Yes
End of the particle

Simulation Chain for HEP Experiments

SLAC



Simulation chain for Treatment Planning (medical)



Non-exhaustive list of MC codes

- EM physics
  - ETRAN (Berger & Seltzer; NIST)
  - EGS4 (Nelson, Hirayama, Rogers; SLAC)
  - EGS5 (Hirayama *et al.*; KEK/SLAC)
  - EGSnrc (Kawrakow & Rogers; NRCC)
  - Penelope (Salvat *et al.*; U. Barcelona)
- Hadronic physics / general purpose
  - Fluka (Ferrari *et al.*, CERN/INFN)
  - **Geant4** (Geant4 Collaboration)
  - MARS (James & Mokhov; FNAL)
  - MCNPX / MCNP5 (LANL)
  - PHITS (Niita *et al.*; JAEA)

http://www.geant4.org/

S. Agostinelli et al.

Geant4: a simulation toolkit

NIM A, vol. 506, no. 3, pp. 250-303, 2003

J. Allison et al.

Geant4 Developments and Applications

IEEE Trans. Nucl. Sci., vol. 53, no. 1, pp. 270-278, 2006

http://top25.sciencedirect.com/journal/01689002/

Top 25 Hottest Articles

Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment

October to December 2015

RSS Blog This! Print **Show condensed**

1. **Geant4-a simulation toolkit**
   Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, Volume 506, Issue 3, Pages 250-303
   Agostinelli, S.; Allison, J.; Amako, K.; Apostolakis, J.; Araujo, H.; Arce, P.; Asai, M.; Axen, D.; Banerjee, S.; Barrand, G.; Behner, F.; Bellagamba, L.; Boudreau, J.; Broglia, L.; Brunengo, A.; Burkhardt, H.; Chauvie, S.; Chuma, J.; Chytracek, R.; Coope
   Cited by Scopus (7384)

2. **The gas electron multiplier (GEM): Operating principles and applications**
   Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, Volume 805, Pages 2-24
   Sauli, F.

3. **The Super-Kamiokande detector**
   Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, Volume 501, Issue 2-3, Pages 418-462
   Fukuda, S.; Fukuda, Y.; Hayakawa, T.; Ichihara, E.; Ishitsuka, M.; Itow, Y.; Kajita, T.; Kameda, J.; Kaneyuki, K.; Kasuga, S.; Kobayashi, K.; Kobayashi, Y.; Koshio, Y.; Miura, M.; Moriyama, S.; Nakahata, M.; Nakayama, S.; Namba, T.; Obayashi, Y.; Okada, A
   Cited by Scopus (331)

4. **Development of a low-cost-high-sensitivity Compton camera using CsI (Tl) scintillators**

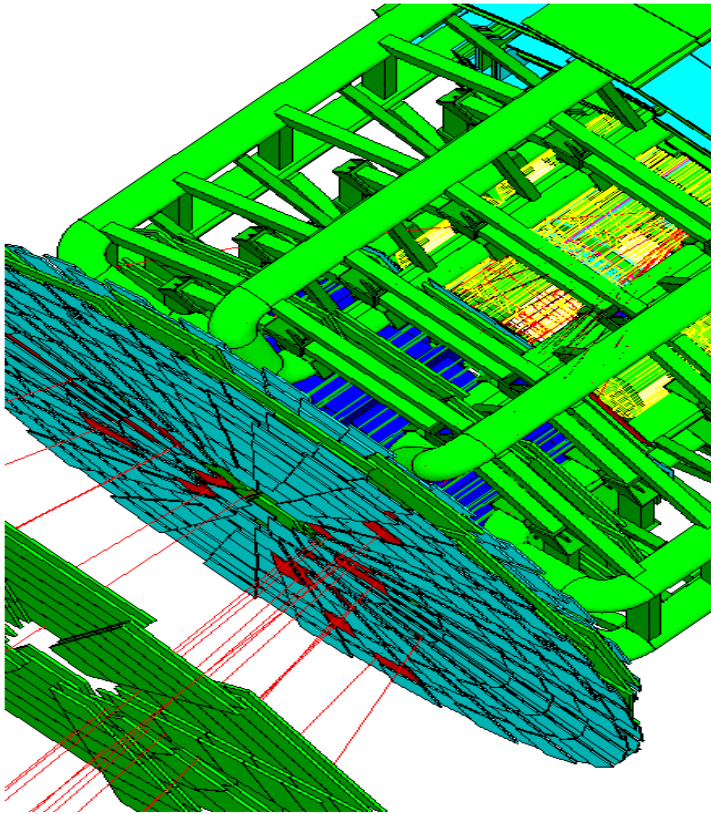
Geant4 history

SLAC

- Dec ’94 - Project start
- Apr ’97 - First alpha release
- Jul ’98 - First beta release
- Dec ’98 - First Geant4 public release - version 1.0
- …
- Nov 30th, ’12 – Geant4 version 9.6 release
  - Feb 4th, ’15 - Geant4 9.6-patch04 release ← Version frozen
- Dec 6th, ’13 – Geant4 version 10.0 release
  - Mar 6th, ’15 - Geant4 10.0-patch04 release ← Retroactive patch release
- Dec 5th, ’14 – Geant4 version 10.1 release
  - Feb 5th, ’16 - Geant4 10.1-patch03 release ← Retroactive patch release
- Dec 4th, ’15 – Geant4 version 10.2 release
  - Feb 26th, ’16 - Geant4 10.2-patch01 release ← Current version

Key Geant4 functionalities

- Geant4 is a general purpose Monte Carlo simulation tool for elementary particles passing through and interacting with matter. It finds quite a wide variety of user domains including high energy and nuclear physics, space engineering, medical applications, material science, radiation protection and security.
- Geant4 offers most, if not all, of the functionalities required for the simulation of elementary particle and nucleus passing through and interacting with matter.
  - Kernel
  - Geometry and navigation
  - Physics processes
  - Scoring
  - GUI and Visualization drivers



Geant4: the first impact

Ok, you go to http://geant4.org and download G4, then what? How do you start it?

Geant4: the first impact

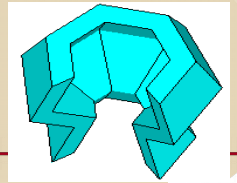
Ok, you go to http://geant4.org and download G4, then what? How do you start it?

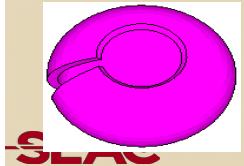
… well, you don't!

Geant4 is a toolkit, not an application, it is a set of libraries and header files to build your own application

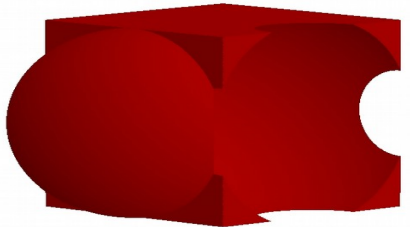
- Unfortunately this mean that it has a quite steep learning curve
- But many (~100) examples exist from which you can start
- Best way: take part to one of the 1-week G4 courses around the world

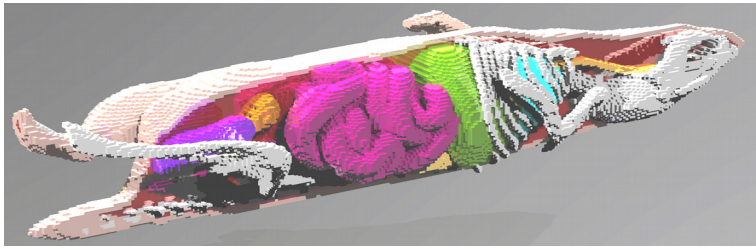
Key geometry capabilities

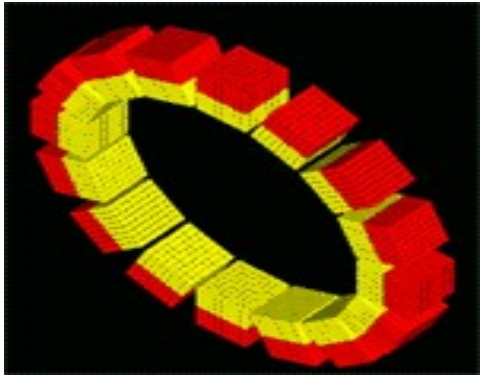




- Richest collection of shapes
  - CSG (Constructed Solid Geometry), Boolean operation, Tessellated solid, etc.
  - The user can easily extend
- Describing a setup as hierarchy or ‘flat’ structure
  - Describing setups up to billions of volumes
  - Tools for creating & checking complex structures
  - Interface to CAD
- Navigating fast in complex geometry model
  - Automatic optimization
- Geometry models can be ‘dynamic’
  - Changing the setup at run-time, e.g. “moving objects”

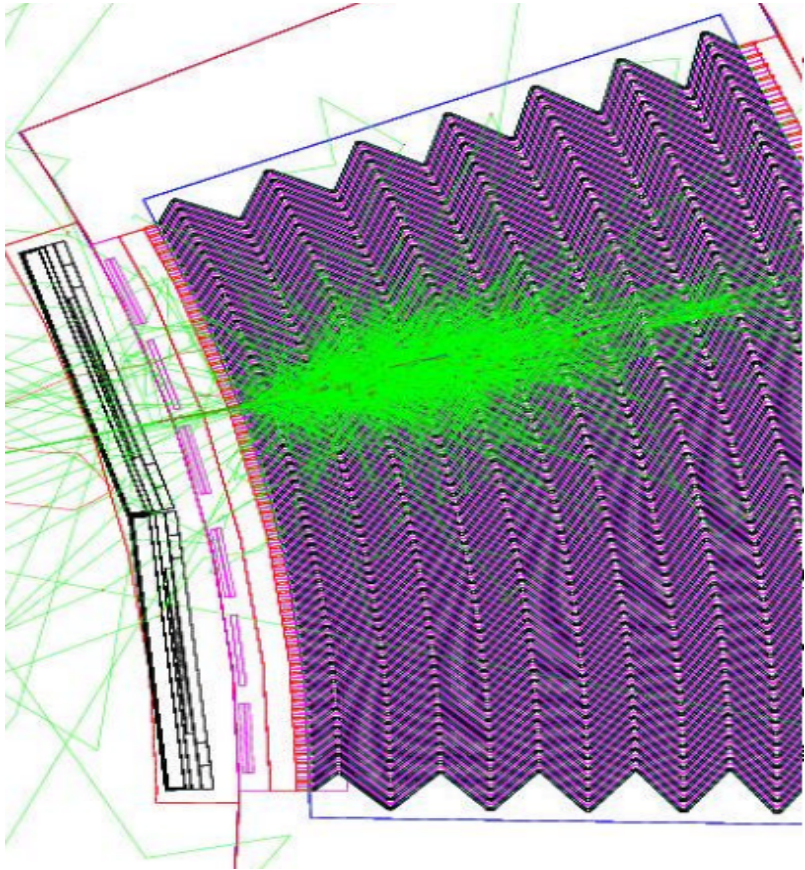






Physics models in Geant4

- Geant4 offers
  - Electromagnetic processes
  - Hadronic and nuclear processes
  - Photon/lepton-hadron processes
  - Optical photon processes
  - Decay processes
  - Shower parameterization
  - Event biasing techniques
  - And you can plug-in more
- Geant4 provides sets of alternative physics models so that the user can freely choose appropriate models according to the type of his/her application.
  - For example, some models are more accurate than others at a sacrifice of speed.




SLAC