# Monte Carlo Methods

Slides by: Prof. Alok Shukla

Department of Physics, I.I.T. Bombay, Powai, Mumbai 400076

# Introduction

So far we used deterministic methods to determine values of integrals of different functions. Next we will discuss methods which are stochastic in nature and yet they can be used to evaluate various integrals. These methods are called Monte Carlo methods. Monte Carlo is a principality in France which is famous for gambling because of its numerous casinos. Since gambling involves basically stochastic processes, hence these stochastic numerical methods have been named Monte Carlo methods.

# Monte Carlo Integration

Foundations of the Monte Carlo approach to integration are based upon the mean value theorem valid for integrals. It states that

$$\int_a^b f(x)w(x)dx = f(\Theta)\int_a^b w(x)dx \tag{1}$$

where $\Theta \in (a,b)$ and $w(x) \geq 0$ for $x \in (a,b)$
From Eq. (1) it follows

$$f(\Theta) = \frac{\int_a^b f(x)w(x)dx}{\int_a^b w(x)dx} \tag{2}$$

in case $w(x)$ is some kind of a weight (i.e. probability) associated with $f(x)$, then clearly $f(\Theta)$ defined by Eq.(2) is the average of the function $f(x)$ when $x \in (a,b)$.

# Monte Carlo Integration (contd.)

Thus, if $\langle f(x) \rangle_w$ donates the weighted average value of $f(x)$, we have

$$\langle f(x) \rangle_w = \frac{\int_a^b f(x)w(x)dx}{\int_a^b w(x)dx} \tag{3}$$

or

$$\boxed{\int_a^b f(x)w(x)dx = \left\{ \int_a^b w(x)dx \right\} \{\langle f(x) \rangle_w\}} \tag{4}$$

Suppose $w(x)$ is a function which is easily integrable in the interval $(a, b)$ while the integral $\int_a^b f(x)w(x)dx$ is not easy to compute. Then one could use result (4) to compute $I = \int_a^b f(x)w(x)dx$ provided we could somehow computer the quantity $\langle f(x) \rangle_w$. In Monte Carlo integration one adaps this approach, i.e., first the average $\langle f(x) \rangle_w$ is computed by stochastic means and then the relation (4) is utilised to compute $I$.

# Monte Carlo Integration (contd.)

An interesting case emerges by setting $w(x) = 1$ in (4) and defining $L = b - a \equiv$ length of integration, we get from (4)

$$\int_a^b f(x)dx = L\langle f(x)\rangle \tag{5}$$

Thus, according to Eq. (5), if we could somehow compute the average of $\langle f(x)\rangle$ over the integral $(a, b)$, we could compute the integral by simply using Eq. (5). Results (4) and (5) can also extended to integration problems in higher dimensions, i.e.,

$$\iiint f(x_1, x_2, -x_n)\, w(x_1, x_2, -x_n)\, d^n x = \left\{\iiint w(x_1, x_2, -x_n)\, d^n x\right\} \\ \{\langle f(x_1 \ldots x_n)\rangle_w\} \tag{6}$$

and when $w(x_1, \ldots x_n) = 1$, we get

$$\iint_{\ldots}\int f(x_1, \ldots x_n)\, d^n x = V_n \langle f(x_1 \ldots x_n)\rangle \tag{7}$$

where $V_n \equiv$ volume of the $n$-dimensional region of integration

# Calculation of Averages - Random Sampling

- Monte Carlo(MC) method gets its name from the the way the averages $\langle f(x) \rangle_w$ are computed.
- The way the averaging is done is very simple. Let us suppose the at we want to compute integral of Eq. (7). $N$ n-tuples of random numbers $\left\{ \left( x_1^{(i)} - x_n^{(i)} \right), i = 1, N \right\}$ are generated and the value of the function $f(x_1 \ldots x_n)$ is computed for each $n$-tuple yielding $N$ values $\{f_i, i = 1 \ldots N\}$. The average is computed as

$$\langle f(x_1 \ldots x_n) \rangle = \frac{\sum_{i=1}^{N} f_i}{N} \tag{8}$$

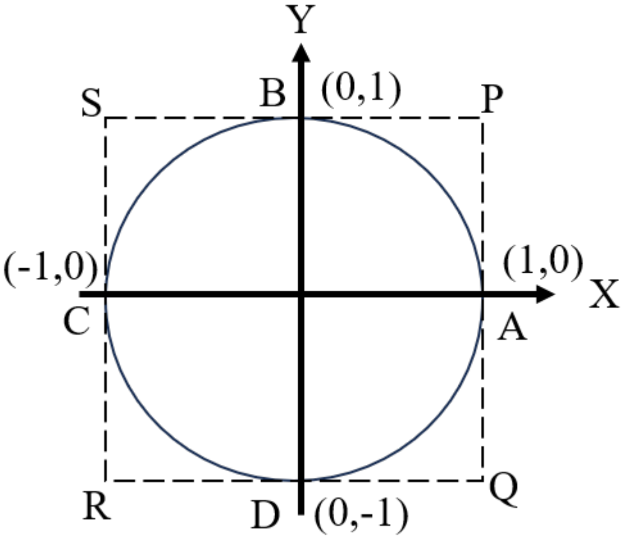
- subsequently this value is substituted in Eq. (7) to obtain the value of the integral.

# Calculation of Averages - Random Sampling

- If the $N$ sampling points used in the calculation are absolutely random (essentially an impossible task), the error associated with the integral is $\Delta I \propto \frac{1}{\sqrt{N}}$. The equivalent error associated with th trapezoidel rule in ID integration is $\triangle I \propto \frac{1}{N^2}$ which is much superior. However, in a general $n$-dimesind space $\Delta I_{\text{trap}} \propto \frac{1}{N^{2/d}}$, while the error of the MC method is uncharged.
- Thus for $d > 4$, the MC integration performs better than the trapezoidal rule.
- Thus the main utility of the MC lies in the integration in larger dimensions such as those involving many-particle systems. Of course, the fundamental assumption governing the accuracy of the MC approach is random sampling for a given value of $N$.
- Thus a highly accurate random number generator is absolutely essential for the success of the MC approach.

# Example - Evaluation of $\pi$ by MC Method

- It is clear that the area of a unit circle equal to $\pi$.



# Example - Evaluation of $\pi$ by MC Method

- Let us draw the unit circle enclosed inside the square of PQRS the length of whose sides is two. Clearly the ratio of the area of the circle to that of the square is $\pi/4$. In the MC approach to be described next we compute the ratio of these two cases and multiply that by 4 to compute $\pi$.
- In practice, we will generate $N$ random numbers $\{x_i, i = 1, \ldots N\}$ each satisfying $0 \leqslant x_i \leqslant 1$. Thus each point $x_i$ is confined to the first quadrant. After each point is generated, it is checked whether the point lies on or inside the circle by checking whether or not it satisfies $\sqrt{1 - x^2} \leqslant 1$. Let the number of such points be $N'$. Then clearly the ratio of one quarter of a circle and the quarter of the square it is enclosed inside is

$$\frac{N'}{N} = \pi/4$$

# Metropolis Algorithm

- Metropolis algorithm is very useful for computing integrals of the type of Eq. (4) which can be seen as related to the weighted average of $f(x)$ over the range of integration of $x$, with $w(x)$ as the weighting factor.
- It may be possible that the integral $w(x)f(x)$ varies wildly in the range of integration, however, $f(x)$ itself varies relatively slowly and $w(a) \leqslant 0$ in that range. In that case the integral $\int_a^b w(x)f(x)dx$ can be seen as a weighted average as argued earlier. If we try to directly evaluate the integral by random sampling, the result may have large amounts of errors.
- However, we do know that not all regions of $x$ are going to contribute equally to the integral - the regions of $x$ with greater value of $w(x)$ will contribute more than the ones with lesser values.

# Metropolis Algorithm (contd.)

- This fact is taken into account in the Metropolis approach by following the algorithm given below:
  1. An initial $x$ value $x_0$ is chosen and $w(x_0)$ and $f(x_0)$ are computed.
  2. Next value of $x$, $x_1$ is chosen according to the formula

     $$\begin{gathered} x_1 = x_0 + \Delta x \\ \text{where } \Delta x = \Delta(2n_i - 1) \end{gathered} \tag{9}$$

     where $\Delta$ is a suitably chosen step size and $n_i$ is a random number.
  3. In keeping with the interpretation that $w(x)$ represents the probability for the function to acquire the value $f(x)$, we make the following decision

     $$\begin{gathered} \text{If} \qquad \frac{w(x_1)}{w(x_0)} \geqslant w_i \\ \text{with} w_i \equiv \text{random number} \\ 0 \leqslant w_i \leqslant 1 \end{gathered} \tag{10}$$

# Metropolis Algorithm (contd.)

- the position $x_1$ is accepted while if it is not satisfied the new configuration is rejected and a new search is made.

  1. Step (3) is repeated n times and the latest configuration that passes condition (10) is considered to be the outcome of the cycle and the value of $f(x)$ corresponding to that value of $x = x_k$ is collected in the list.

     Such n-step Metropolis sampling cycles are repeated $N$ times fo complete the list $\{x_k, k = 1, N\}, \{f(x_k), k = 1, \ldots N\}$ and $\{w(x_k), k = 1, \cdots N\}$
  2. Finally, the average value of $f(x)$ is computed according to the formula:

$$\langle f(x) \rangle = \frac{\sum_{k=1}^{N} f(x_k)}{N}$$

# Metropolis Algorithm (contd.)

- It is clear for the way in which the sampling is done that more points in the sample will be accepted from the region where the weight $w(x)$ (hence the probability) is higher.
- The reason behind accepting one position from n attempted Metropolis steps is to reduce any sort of correlation-between successive steps.

# Applications to Statistical Mechanics (contd.)

- We presented the Metropolis algorithm as a very general method of evaluating integrals which can be seen as weighted averages. However, the origins of this algorithm lie in statistical mechanics, as also the majority of its applications.
- Let us consider the case of an $N$-particle system at finite temperature, obeying the laws of classical statistical mechanics. We can describe such a system within the framework of canonical ensemble. The thermal average of a property represented by an observable $A$ is then given by

$$\langle A \rangle = \frac{1}{z} \int A(R) e^{-u(R)/k_B T} dR \tag{11}$$

where $N$ is the total number of particles in the system which naturally needs 3N coordinates to be described. Thus

$$d\vec{R} = d\vec{r}_1 d\vec{r}_2 \ldots d\vec{r}_N$$

# Applications to Statistical Mechanics (contd.)

- The integral over $d\overrightarrow{R}$ represents an ensemble average which includes all possible configuration of $N$ particles weighted by the Boltzman factor $e^{-U(R)/k_B T}$. For a given temperature $T$, clearly, those configurations of the particles will contibute more to the average which have smaller potential energy. Thus this case is an ideal one for the application of the Metropolis algorithm. Of course, to perform calculations one needs to know the form of the potential energy function $U(R)$. One function used frequently to study liquids is the Lennard-Jones pair potential

$$U(R) = \sum_{i>j}^{N} v\left(r_{ij}\right) \tag{12}$$

with

$$V\left(r_i\right) = 4\varepsilon\left[\left(\frac{\sigma}{r_{ij}}\right)^{12} - \left(\frac{\sigma}{r_{ij}}\right)^{6}\right] \tag{13}$$

# Applications to Statistical Mechanics (contd.)

- where $\varepsilon$ and $\sigma$ are system dependent parameters and $r_{ij} \equiv$ distance between particles $i$ and $j$.
- One can actually apply the whole technique of the Metropolis algoritum by updating the coodinates of the particle according to

$$\begin{aligned} x_i^{(n+1)} &= x_i^{(n)} + \Delta_x (2\alpha_i - 1) \\ y_i^{(n+1)} &= y_i^{(n)} + \Delta_y (2\beta_i - 1) \\ z_i^{(n+1)} &= z_i^{(n)} + \Delta_z (2r_i - 1) \end{aligned} \qquad (14)$$

  where $(\alpha_i, \beta_i, r_i) \in$ random numbers.
- The potential energy is computed for this new configuration using Eq. (13) and then the probability check

$$\frac{w(R_{n+1})}{w(R_n)} = \frac{e^{-U\left(R^{(n+1)}\right)/kT}}{e^{-U\left(R^{(n)}\right)/kT}} \geqslant w_i \text{ (a random number)} \qquad (15)$$

  is made.

# Applications to Statistical Mechanics (contd.)

- If Eq. (14) is satisfied, then the new configuration of particles is accepted, otherwise it is rejected. Once $N$ such configurations are identified, the average of quantity $A$ is computed as

$$\langle A \rangle = \frac{\sum_{k=1}^{N} A(R_k)}{N}$$

- Condition (14) in this context is connected to the condition of detailed balance which is supposed to prevail in equilibrium. Let us examine the conditions under which Eq. (15) will be satisfied.

# Applications to Statistical Mechanics (contd.)

- If $U\left(R^{n+1}\right) < U\left(R^{n}\right)$ then clearly $\frac{W(R_{n+1})}{W(R_n)} > 1$ and hence according to Eq.(14) the new configuration will be accepted.
- This is consistent with the physics of the situation where the system would always prefer configurations with lower energy. However, when $U\left(R^{n+1}\right) > U\left(R^{n}\right)$ then $\frac{w(R_{n+1})}{w(R_n)} < 1$ and the acceptance of the new configuration will be a stochastic event depending upon the value of th random number $w_i$. This solution criterion is consistent with the condition of detailed balance that the system in thermal equilibrium satisfy

$$W\left(R_n\right) T\left(\vec{R}_n \rightarrow \vec{R}_{n+1}\right) = W\left(R_{n+1}\right) T\left(\vec{R}_{n+1} \rightarrow \vec{R}_n\right)$$

# Applications to Statistical Mechanics (contd.)

- where $T\left(\vec{R}_n \rightarrow \vec{R}_{n+1}\right)$ specifies the transition rate of the system for configuration $\vec{R}_n$ to $\vec{R}_{n+1}$ so that

$$\frac{w\left(R_{n+1}\right)}{w\left(R_n\right)} = \frac{T\left(\vec{R}_n \rightarrow R_{n+1}\right)}{T\left(R_{n+1} \rightarrow R_n\right)}$$

$$\frac{T\left(R_n \rightarrow R_{n+1}\right)}{T\left(\bar{R}_{n+1} \rightarrow \bar{R}_n\right)} = e^{-(U(R_{n+1})-U(R_n)/k_B T}$$

if $U\left(R_{n+1}\right) \dot{<} U\left(R_n\right)$ we know that

$$T\left(R_n \rightarrow R_{n+1}\right) = 1$$

and the system makes the transition to the lower energy state.

# Applications to Statistical Mechanics (contd.)

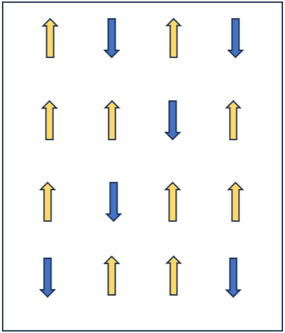
- However when $U(R_{n+1}) > U(R_n)$ then $T(R_{n+1} \longrightarrow R_n) = 1$ and weighted

  $$T(R_{n\to} \to R_{n+1}) = e^{-(u(R_{n+1}) - U(R_n))/k_B T}$$

  Thus whether the system will or will not make a transition to this higher energy state will be a random event caused by a thermal fluctuation. Thus, for such a case, the decision as to whether transition be made depends on the randan number $w_i$. Thus this Metropolis step mimics the interaction of the system with a thermal reservoir naturally.

# Applications to Ising Model

- Suppose we have a two dimensional lattice whose each side is occupied by a spin which can be either up or down



# Applications to Ising Model (contd.)

- The interaction between different spins is assumed confined only to nearest neighbor spins and is represented by by the Hamiltonian

$$H = -J \sum_{\langle ij \rangle} s_i s_j - B \sum_{i=1}^{N} s_i \qquad (16)$$

$$\begin{aligned} \langle ij \rangle &\equiv \text{ nearest neighbor sites} \\ s_i = 1 &\Rightarrow \text{ up spin} \\ s_i = -1 &\Rightarrow \text{ down spin} \\ B &\equiv \text{ external field} \\ J &\equiv \text{ exchange interation} \end{aligned}$$

- Such a system is believed to describe magnetic materials such as various transition metal and rare earth compounds consisting of localized spins.

# Applications to Ising Model (contd.)

- It is of considerable interest to predict the magnetization properties of various systems using the Ising model. One is particularly interested in finding out whether or not the system exhibits a phase transition at $B = 0$, as a function of temperature. If yes, then what is the critical temperature $T = T_C$. In order to obtain the value of magnetization as a function of temperature one has to calculate the following thermal average:

$$
\begin{gathered}
m = \frac{1}{z} \sum_{\sigma} S_{\sigma} e^{-H_{\sigma}/k_B T} \\
\text{when } S_{\sigma} = \frac{S_{\sigma}}{N} \text{ and } S_{\sigma} = \sum_{i} S_i \\
\text{and } z = \sum_{\sigma} e^{-H_{\sigma}/k_B T}
\end{gathered}
\tag{17}
$$

# Applications to Ising Model (contd.)

- In Metropolis algorithm we initialize the system with a spin configuration by randomly assigning spin values ±1 to different sites.
- Search for a new configuration is initiated by flipping one of the spins at random and then computing the energy difference $E_n - E_{n+1} = \Delta t$ and checking whether $e^{-\Delta E/k_B T} \geqslant w_i$ where $w_i$ is a random number
- Clearly if the new configuration is a lower energy configuration, it will definitely be selected. However, if this configuration is higher in energy than the previous one, whether or not it will be selected will be a stochastic event depending on the value of the random number $w_i$.
- Finally, after collecting $N$ data points we compute

$$\boxed{m = \frac{\sum S_\sigma}{N}} \tag{18}$$

# Applications to Ising Model (contd.)

- In addition to spontaneous magnetization, the other quantities of interest are total energy

$$\boxed{\begin{aligned}\langle E\rangle &= \frac{1}{N}\sum_{\sigma} E_{\sigma}\\ \langle E^2\rangle &= \frac{1}{N_m}\sum_{\sigma} E_{\sigma}^2\end{aligned}} \tag{19}$$

and the specific heat

$$\boxed{C = \frac{(\Delta E)^2}{k_B T^2}} \tag{20}$$

where

$$\boxed{(\Delta E)^2 = \langle E^2\rangle - \langle E\rangle^2} \tag{21}$$

# Applications to quantum Mechanics

- Next we will illustrate as to how the Monte Carlo approach can be used to obtain the ground state energies of various one -particle quantum mechanical problem. This method is called the Variational Quantum Monte Carlo approach(VQMCA).
- Its implementation in the context of one particle problems is straightforward, however, when extended in many-electron problems, the method has many difficulties of principle. Suppose we have a system described by the Hamiltonian

$$H = -\frac{\hbar^2}{2m}\nabla^2 + V(\vec{r}) \tag{22}$$

- If $\phi$ is the wave function of the ground state which not normalized, the energy associated with the wave function is

$$\boxed{E = \frac{\int \phi^* H \phi d\tau}{\int \phi^* H \phi d\tau}} \tag{23}$$

# Applications to quantum Mechanics (contd.)

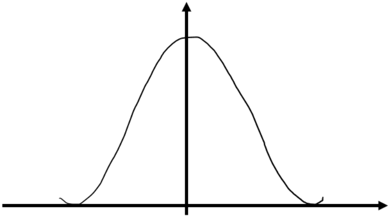
- Let us consider the implementation of VQMC approach to a one-dimension of simple Harmonic oscillator,i.e.,

$$H = -\frac{\hbar^2}{2m}\frac{d^2}{dx^2} + \frac{1}{2}kx^2 \tag{24}$$

- We know that the exact ground state energy for this problem is

$$E_0 = \frac{\hbar\omega}{2}$$

  and the exact wave function looks like



# Applications to quantum Mechanics

- However, in a Metropolis algorithm based search for the ground state we proceed as follows:
  1. We restrict our search for the true wave function over a fixed region i.e., say in this case $x \in (-10, 10)$
  2. We divide the search interval into bias of size $\Delta x = 0.1(say)$. The total # of bias $n = \frac{x_{\text{max}} - x_{\text{min}}}{\Delta x}$.
  3. We assume a trial wave function $\phi^{(0)}$ of the following simple type $\phi(0) = c$ for $x \in (-1, 1)$ say and $\phi^{(0)} = 0$ outside, where $c$ can be determined by the normalization condition, i.e.,

     $$\int \phi^{(0)^*} \phi^{(0)} dx = 1$$

     $E^{(0)}$ is computed for $\phi^{(0)}$ using Eq. (23)
  4. Now we pick a bias at random say at $x = x_i$ and change the wave function at $x = x_i$ by an amount chosen randomly in the range it $\pm d\phi$.
  5. The energy for the modified wavefuction $E^{(1)}$ is computed according to Eq.(23). If $E^{(1)} < E^{(0)}$ the new wavefunction is accepted, else it is rejected.
  6. Process is continued until we have achieved convergence, i.e. minimum of the energy has been obtained.

# Generation of Random Numbers

- We saw that the implementation of the Monte Carlo approach depended crucially on the supply of random numbers. In this chapter we will briefly discuss the generation of uniform deviates, i.e., uniformly distributed random numbers which lie within a specified range. In particular, we will concentrate on what are called the linear congruential generators (LCG). The LCGs are based upon the following recursion relation

  $$R_k = (aR_{k-1} + c) \bmod \mathrm{m} \tag{25}$$

  where $a, c$, and $\mathrm{m}$ are preselected constants. $a$ is called the multiplier, $c$ is the increment, $m$ is the modulus, and $R_0$ is the seed.

# Generation of Random Numbers

- The quality of the generator depends strongly upon the choice of these constants. Obviously $0 < R_k < m$. Thus, in order to obtain uniform deviates $x_k \in (0,1)$ all we need to do is

$$x_k = R_k/m \tag{26}$$

- Some of the choices for the constants are:
  1. $m = 2^{31}, a = 1103515245, \quad c = 12345, R_0 = 12345$ used in ANSIC
  2. $m = 2^{31}, a = 2^{16} + 3 = 65539, c = 0, R_0 = 1$ RANDU
  3. $m = 2^{31} - 1, a = 630360016, c = 0, R_0 = 1$ SIMSCRIPT
  4. $m = 2^{35}, a = 5^{15}, c = 7261067085, R_0 = 0$ BCSLIB
  5. The most popular among the LCGs is the minimal standard generenter of Park and Miller corresponding to parameters

$$a = 16807, m = 2^{31} - 1 = 2147483647$$

$$c = 0, \text{ and } R_0 = 1.$$

# Generation of Random Numbers

The number chosen for $m$, i.e., $2^{31}-1$ is called Merssene prime number named after its discoverer.

- The problem with most of the LCGs then the multiplication step $aR_{k-1}$ can occasionally lead to an integer overflow in 32-bit machines because of the large numbers involved. This problem com be avoided by using the concept of Schrage's factorization based upon the approximate factorization of $m$ :

$$m = aq + r \tag{27}$$

with

$$q = [m/a], r = m \bmod a \tag{28}$$

then

$$\text{then } az \bmod \mathrm{m} = \begin{cases} a(z \text{ mod } q) - r[z/q] \text{ if } > 0 \\ a(z \text{ mod } q) - r[z/q] + m \text{ otherwise} \end{cases} \tag{29}$$

# Generation of Random Numbers

- It is obvious that different terms in Eq. (29) will be small enough to avoid overflow problem on 32 -bit machines.
- Next we prove the results stated in Eq. (29)

**Proof of Schrage's Factorization**:

- Our aim is to obtain an expression for the quantity

$$x = az \bmod m \tag{30}$$

Now

$$az = \left[\frac{az}{m}\right] m + x \tag{31}$$

or

$$x = -\left[\frac{az}{m}\right] m + az \tag{32}$$

# Generation of Random Numbers

- Let us write

$$m = ag + r \tag{33}$$

where $q = [\mathrm{m/a}]$ and $r = [m \bmod a]$

now $\left[\frac{az}{m}\right] = \left[\frac{z}{[m/a]}\right] = \left[\frac{z}{q}\right]$

$$x = az - \left[\frac{z}{q}\right] m$$

$$x = az - \left[\frac{z}{q}\right](aq + r)$$

$$x = az - aq\left[\frac{z}{q}\right] - r[z/q]$$

$$x = az - aq[z/q] - r[z/q]$$

$$x = a\{z - q[z/q]\} - r[z/q]$$

$$x = a(z \bmod q) - r[z/q] > 0$$

$$x = x + m$$

# Generation of Random Numbers

$$\begin{aligned} x &= az - [z/q]m \\ &= az - [z/q](aq + r) \\ x &= az - aq[z/q] - r[z/q] \\ x &= a\{z - q[z/q]\} - r[z/q], \text{where } z - q[z/q] = z \text{ mod } q \\ x &= a(z \text{ mod } q) - r[z/q] \end{aligned}$$

- if $< 0$

$$x = a(z \bmod q) - r[z/q] + m$$

- this is valid only if

$$r < q$$

- for MSG

$$q = 127773$$
$$r = 2836$$