
ARTICLE

An Auto-Importance Sampling Method for Deep Penetration Problems

Junli LI^{1,2,3,*}, Chunyan LI⁴, Zhen WU⁴,
Zhi ZENG^{1,2,3} and Rui QIU^{1,2,3}

¹Department of Engineering Physics, Tsinghua University, Beijing, China

²Key Laboratory of Particle and Radiation Imaging (Ministry of Education), Tsinghua University, Beijing, China

³Key Laboratory of High Energy Radiation Imaging Fundamental Science, Beijing, China

⁴Nuctech Company Limited, Beijing, China

For many traditional radiation shielding problems as well as medical and nuclear well-logging applications, it has long been recognized that the analog Monte Carlo is inefficient, which are known especially in deep penetration problems. This arises from the fact that the particles have little chance of contributing to the solution far from the source when using analog procedures. Numerous variance reduction techniques have grown up to overcome this inefficiency. However, there are still some limitations. A further difficulty lies in the potential for biased results due to the misuse of variance reduction techniques or the statistical convergence of Monte Carlo results. Responding to these difficulties, a new variance reduction technique called “auto-importance sampling method” is proposed in this paper. In the new method, the geometry space is divided into sub-spaces by fictitious surfaces. On the fictitious surfaces, fictitious particles are created according to the importance distribution automatically and then they are transported. The particles’ number is increased on each fictitious surface. With this method, it is unnecessary to make pre-calculations to test different parameters for the variance reduction technique, or to make another calculation to get the importance distribution. Two examples are simulated with the new method. It can be seen that the new variance reduction technique is very easy to use and the computational efficiency achieved is very encouraging.

KEYWORDS: deep penetration, importance sampling, auto-importance sampling, fictitious particle

I. Introduction

The Monte Carlo method enables the accurate simulation of complicated radiation transport problems with complex geometry. Therefore in many traditional radiation shielding problems, the Monte Carlo method is usually used to make refined simulations of radiation shielding problems. The Monte Carlo method is to simulate the transportation of every particle, then obtain the average statistical behavior of the large number of particles. If the probability of the particle reaching the tally region is too small for the reason that either the particle needs to penetrate a very thick shielding block, or the particle must be scattered many times before arriving the tally region, a huge amount of source particles must be simulated in order to decrease the statistical error of the tally results. However due to the limitation of the computer’s calculation ability and the recycle period of the random variable, it is difficult to simulate too many particles in an acceptable CPU time. This is known as the deep-penetration problem.¹⁾

In order to solve this problem, many variance reduction techniques have been developed,²⁻⁴⁾ however, none of them can solve this problem completely. Among these variance reduction techniques, the most effective one is the importance sampling method, in which the importance of a particle is defined as its contribution to the final tally. Then the transport process of the particles of high importance are si-

mulated in detail, however, the particles’ behavior of low importance are simulated roughly. As a result, most of the CPU time is spent on the transport of high importance particles which will decrease the statistic error of tally results significantly, and reduce the total computation time at the same time. In the past, the wide application of importance sampling method has demonstrated that it can improve the simulation efficiency greatly and very good agreement has been shown between the simulation results and the true values.

However, compared with its high simulation efficiency, the importance sampling technique has a big disadvantage which makes this technique very difficult to be applied in practice. That is how to determine the importance of every particle whose location, energy and direction may vary significantly. The assignment of importance of each particle requires not only trial calculation, experience, but also intuition, and it is so time consuming.⁵⁻⁷⁾ If the parameters about particles’ importance are not set properly, the simulation results may be even worse. Booth and Hendricks (1984) state, “The selection (of parameters) is more art than science, and typically, the user makes a few short trial runs and uses the information these trials provide to better guess the parameters; that is, the user learns to adjust parameters on the basis of previous experience.”¹⁾

This paper proposes a new variance reduction technique, which can transport particles according to the importance distribution automatically, and the setting of the parameters

*Corresponding author, E-mail: lijunli@tsinghua.edu.cn

is much easier than that of the former variance reduction techniques.

II. Variance Reduction Techniques

In the Monte Carlo method, the estimated relative error R of the estimated mean \bar{G} is:

$$R = \frac{S_{\bar{G}}}{\bar{G}} = \frac{x\sigma_N}{\sqrt{N-1}} / \bar{G}, \tag{1}$$

where σ_N is the standard deviation, N is the number of particles simulated, x is usually set 2 which means that the corresponding confidence lever is 95%.

For the same transport problem, the standard deviation σ_N will change as the probability model is changed. From Eq. (1), we can see that reducing σ_N can decrease the estimated relative error and thus increase the reliability of the estimation. Therefore, there arises a number of probability models to reduce σ_N , which are called variance reduction techniques (σ_N is just the square root of variance).

1. Importance Sampling Method

One of the most important methodology used in variance reduction techniques is importance sampling. The objective of importance sampling method is to spend more time on sampling in important regions or important particles and less time on sampling in unimportant regions or unimportant particles, thus the simulating time is more spent on useful particles.

Take Eq. (2) as an example, which is to calculate the expected value of g .

$$E(g) = \int g(X)f(X)dX, \tag{2}$$

where $f(X)$ is probability density function of the random variable X .

The general way to calculate $E(g)$ with the Monte Carlo method is to sample large numbers of X_i from $f(X)$, then use the average of $g(X_i)$ as the estimator of $E(g)$.

The importance sampling method is as follows.

Introduce an importance function $I(X)$ and a new probability density function $f_{imp}(X)$:

$$f_{imp}(X) = \frac{f(X)I(X)}{\langle I \rangle}, \tag{3}$$

where $I(X)$ satisfies the conditions: $I(X) \geq 0$;

$$\langle I \rangle = \int I(X)f(X)dX.$$

Then $E(g)$ can be calculated with Eq (4):

$$E(g) = \int g_{imp}(X)f_{imp}(X)dX, \tag{4}$$

where $g_{imp}(X) = g(X) \frac{f(X)}{f_{imp}(X)} = \frac{g(X)}{I(X)} \langle I \rangle$.

That is to sample large numbers of X_i from $f_{imp}(X_i)$, use the average of $g_{imp}(X_i)$, not $g(X_i)$, as the approximation of $E(g)$. $I(X)$ has the importance information, and $f_{imp}(X)$ gives the importance distribution, so this method is

called importance sampling method.

2. Splitting/Roulette Technique

Splitting/Russian roulette is one of the oldest and most widely used techniques. Take geometric splitting/Russian roulette for example. The user needs to assign each cell an importance I in the input file. When a particle of weight W passes from a cell of importance I to a cell of higher importance I' , the particle is split into identical particles of the same weight. If I'/I is an integer $n(n \geq 2)$, the particle is split into n identical particles, each weighting W/n . If I'/I is not an integer, denoting $m = [I'/I]$ to be the largest integer in I'/I and $p = I'/I - m$. Then the particle is split into $m+1$ particles with probability p and split into m particles with probability $1-p$, each weighting $W \cdot I/I'$. On the other hand, if a particle of weight W passes from a cell of importance I to a cell of lower importance I' , the particle is killed with probability $1-I'/I$, or remains with probability I'/I weighting $W \cdot I/I'$.⁷⁾

III. Auto-Importance Sampling Method

As mentioned in Section I, the deep-penetration problem arises from the fact that the probability of the particles reaching the statistical region is too small. Therefore only if sufficient particles are simulated, there will be enough particles reaching the statistical region and thus we get a reliable statistical result. However, increasing the number of the source particles is limited by the computing level and the tolerant time. In this paper, a new variance reduction technique called ‘‘auto-importance sampling method’’ is proposed. In the auto-importance sampling method, the geometry space is divided into several sub-spaces by fictitious surfaces. In the process of transporting, fictitious particles are created and transported according to the importance automatically. The next event estimation method and weight splitting/Russian roulette method are employed.

Firstly, the geometry space is divided into several sub-spaces by introducing some fictitious surfaces, and the source particles are located in sub-space No. 1. The fictitious surfaces can be any type, such as plane, cylindrical surface, spherical surface, curved surface etc. Take four sub-spaces divided by three planes for example, which is illustrated in **Fig. 1**. The fictitious surface between sub-space No. 1 and sub-space No. 2 is called fictitious surface No. 1, and so on.

Secondly, transport the source particles only in sub-space No. 1, and at every collision position (including the source position) create a fictitious particle on fictitious surface

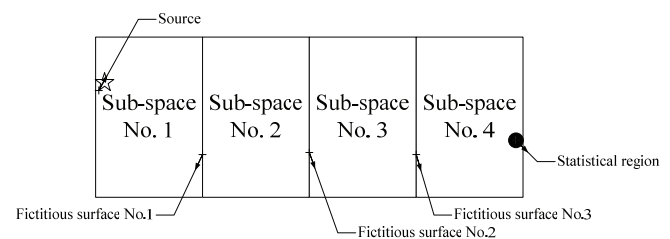


Fig. 1 The geometry space is divided with three fictitious plane

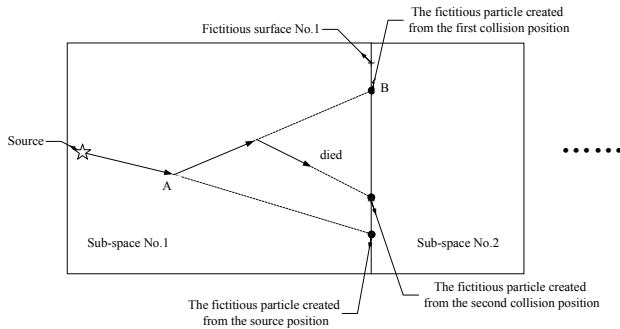


Fig. 2 The sketch map of the fictitious particles' creation

No. 1 with the next event estimation method. The fictitious particle is created only when the real particle travels toward the fictitious surface No. 1. If the real particle travels in a direction opposite to the fictitious surface No. 1, no fictitious particle is created. That means the fictitious particles are initially created at the collision position (including the source position) and deterministically transported toward the fictitious surface No. 1 in the direction of the real particle without any collision. **Figure 2** illustrates the fictitious particles' creation of one source particle. The real lines give the particle's track and the dashed lines describe the fictitious particles' creation. The particle has two collisions before killed and there are three fictitious particles created in all. Suppose the weight of the particle at the first collision position which is named 'A' is W_A , and the position of the fictitious particle created from position 'A' is named 'B', then the weight of the fictitious particle is $W_B = W_A \times e^{-\int_A^B \Sigma_t(s) ds}$, where $e^{-\int_A^B \Sigma_t(s) ds}$ is the attenuation between position 'A' and position 'B'. Once the particle is traversing the fictitious surface No. 1, it will be killed with no fictitious particle created.

Thirdly, after simulating all of the source particles in sub-space No. 1, automatically adjust the fictitious particles according to their importance, so as to realize auto-importance sampling of the fictitious particles. Let W_{avg} be the average weight of all fictitious particles, which is simply the arithmetic mean. The fictitious particles are adjusted one by one with the weight splitting/Russian roulette technique. If the weight of the fictitious particle is higher than W_{avg} , the particle will be split; if the weight is lower than W_{avg} , Russian roulette will be done. Either in splitting or roulette, the weight of the particle will be adjusted to W_{avg} .

Fourthly, adjust the number of the fictitious particles to be the number of the source particles. In order to reduce the estimated error, in the last sub-space there must be enough fictitious particles reaching the statistical region. And also to avoid too long CPU time, the fictitious particles cannot be too many. Therefore, the fictitious particles should be adjusted to be as many as the source particles. Let N_0 be the number of the source particles and N_{fic} be the number of the fictitious particles on a fictitious surface. Recalculate the average weight as $W_{avg}' = W_{avg} \times N_{fic} / N_0$. If $N_{fic} > N_0$, uniformly sample $N_{fic} - N_0$ fictitious particles from the N_{fic}

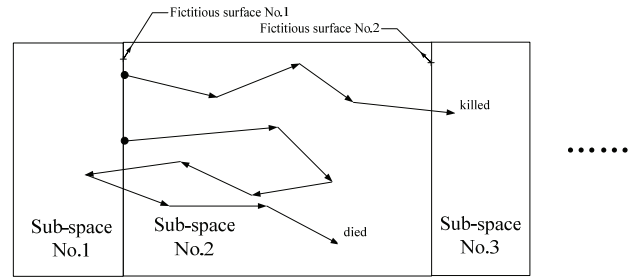


Fig. 3 The sketch map of the transportation of the fictitious particles on the fictitious surface No.1

fictitious particles and delete them, then there are just N_0 fictitious particles left. If $N_{fic} < N_0$, uniformly sample $N_0 - N_{fic}$ fictitious particles from the N_{fic} fictitious particles and duplicate them, then there are N_0 fictitious particles. The sample methods are both $(N_{fic} - 1) \times \xi + 1$, where ξ is a random number on the unit interval [0,1]. After the adjustment, the weights of the N_0 fictitious particles all become W_{avg}' .

Fifthly, transport the fictitious particles on the fictitious surface No. 1 between sub-space No. 2 and sub-space No. 1. During transporting, create fictitious particles on the fictitious surface No. 2 and automatically adjust the fictitious particles according to their importance. And then adjust the number of the fictitious particles to be N_0 . **Figure 3** illustrates two examples: the transportation of two fictitious particles. One fictitious particle is going to traverse the fictitious surface No. 2 and is killed; the other fictitious particle is back scattered into sub-space No. 1. There are not any restrictions on the transportation of the particles, except that the particles will be killed if they traverse the fictitious surface on which the fictitious particles will be created.

Sixthly, transport the fictitious particles in the other sub-spaces until the last and the final tally result can be achieved.

In the new method without simulating a huge amount of source particles, there are particles as many as the source particles transported in every sub-space, so as to ensure enough particles to reach the statistical region. Furthermore, the user only needs to set the parameters of the dividing of the sub-spaces. The setting is much easier than that of the importance distribution needed in the old techniques.

IV. Application and Discussion

AIMS (Auto-Importance Sampling Monte Carlo) — which realizes the auto-importance sampling method, is developed based on MCNP code. Two examples are used to test this new method.

1. Slab Shielding Example

The slab shielding problem is a classic example in the radiation shielding calculation with the Monte Carlo method, in which the geometry is the simplest. So it is chosen to validate the new method. Suppose a concrete slab of an infinite

Table 1 The results with different variance reduction techniques

VRT	N	P	RE (%)	T (min)	FOM
No	1E9	1.59E-6	2.56	2502.93	0.61
AIS 2	1E5	1.99E-06	12.33	0.28	234.92
AIS 3	1E5	1.46E-06	4.89	0.40	1045.50
AIS 3	5E5	1.62E-06	2.14	3.13	697.63
AIS 4	1E5	1.61E-06	3.05	0.56	1919.61
AIS 5	1E5	1.59E-06	2.46	0.70	2360.65
AIS 6	1E5	1.63E-06	2.18	0.80	2630.25
AIS 7	1E5	1.62E-06	1.88	0.94	3009.93
AIS 8	1E5	1.62E-06	1.78	1.03	3064.24
AIS 9	1E5	1.67E-06	1.71	1.12	3053.44
AIS 10	1E5	1.58E-06	1.64	1.20	3098.35
AIS 11	1E5	1.59E-06	1.60	1.26	3100.20
AIS 12	1E5	1.58E-06	1.57	1.35	3005.16
AIS 13	1E5	1.64E-06	1.55	1.40	2973.09
AIS 14	1E5	1.61E-06	1.54	1.46	2888.06
AIS 15	1E5	1.60E-06	1.53	1.51	2829.05

section with 2-m thickness and the density of the concrete is 2.3 g/cm³. A point photon source of particles with 5 MeV is emitted from one side of the slab, and the cosine of the polar angle at the source point with respect to the axis normal to the slab is uniform distribution in [0,1]. The penetrating probability through the concrete slab is calculated.

The analog Monte Carlo method is also used, which simulated 10⁹ particles, while AISMC simulated only 10⁵ particles. In order to better understand the AIS method, fourteen different divisions of the geometry space are made with the number of sub-spaces from two to fifteen. In every division, the set of the fictitious surfaces is free, and the easiest way is to uniformly divide the geometry space. The results are shown in **Table 1**. The FOMs of the AIS method with different numbers of the sub-spaces when 10⁵ source particles are simulated are pictured in **Fig. 4** and the energy spectrums are shown in **Fig. 5**. In the table and figures, ‘VRT’ means variance reduction technique; ‘N’ means source particle number; ‘P’ means penetrating probability; ‘RE’ means estimated relative error; ‘T’ means CPU time; ‘FOM’ means a figure of merit; ‘No’ means simulating with no variance reduction technique, i.e., analog; ‘AIS’ means simulating with the AIS method; the digitals just after ‘AIS’ mean the number of sub-spaces divided.

Referring to Fig. 5, we can see that the energy spectrum simulated with the AIS method is better than that simulated with the analog method in terms of statistics. Actually, in the energy spectrum simulated with the analog method, the relative error of every energy bin is almost about 15% ~ 20%, some even 25%; while in the energy spectrum simulated with the AIS method, all of the relative errors of energy bins are less than 5%, although the CPU time of the AIS method is much less than that of the analog method.

From Table 1 we can see that if only 10⁵ source particles are simulated, and the number of the sub-spaces is bigger than three, the results of the AIS method with different divisions are all in a good agreement with the analog method. It

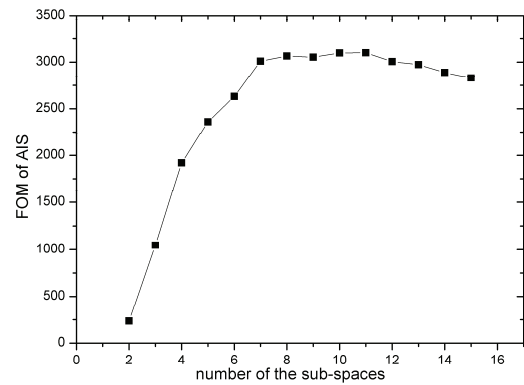


Fig. 4 The FOMs of the AIS method with different numbers of the sub-spaces when 10⁵ source particles are simulated

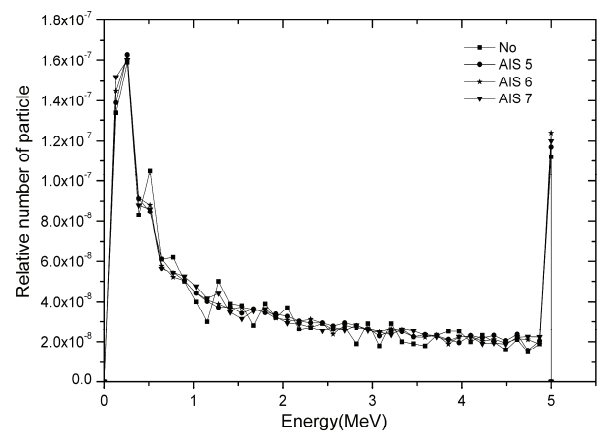


Fig. 5 The energy spectrums with different variance reduction techniques

is easy to understand that the number of the sub-spaces cannot be too small in order to ensure not too big estimated error in every sub-space. In this problem, if only 10⁵ source particles are simulated, the sufficient number of sub-spaces would be four and when the number of the sub-spaces are not less than four, the results are all good and are not affected by the way of division. Therefore, the parameters of the AIS method are much easier to be set than that of other importance sampling methods. From Table 1, it can also be seen that if 5×10⁵ source particles are simulated, three sub-spaces’ division can also give good result. It is obvious that as the number of the sub-spaces is increasing, the estimated relative error of the penetrating probability will decrease, but the computing time will increase at the same time, if the number of the source particles is not changed. And from the ‘RE’ and ‘T’ columns of Table 1 we can also see that. Figure 4 shows that the FOM of the AIS method increases fast with the number of the sub-spaces before it remains steady, then it decreases very slowly. From Table 1 we can see that the FOMs of the AIS method which have more than three sub-spaces are all four orders of magnitude better than that of the analog method, and the calculation of the AIS method takes only one second or so.

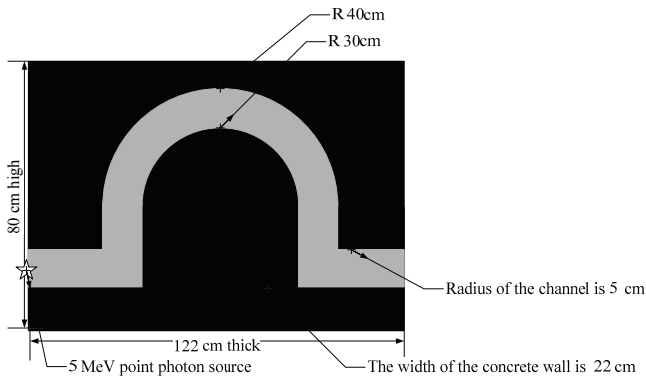


Fig. 6 The sketch map of the curved channel problem's section

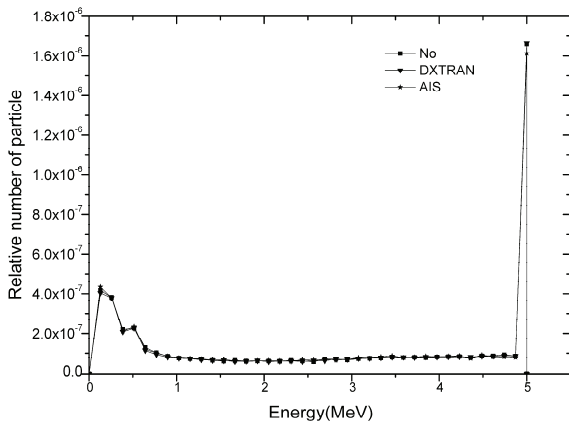


Fig. 7 The energy spectrums of the exit of the curved channel

Table 2 Target thickness and isotopic enrichment

VRT	N	P	RE (%)	T (min)	FOM
No	1E9	5.47E-6	0.57	1644.61	1.87
DXTRAN	1E8	5.40E-6	0.53	379.21	9.39
AIS	1E6	5.44E-6	0.50	15.11	264.73

2. Curved Channel Example – Photon

The curved channel problems are also very common in radiation shielding problems. To avoid the radiation leaking from the wire through holes directly, usually the holes are in slanting line or curved form. In this paper a curved channel problem is calculated, whose sketch map is shown in Fig. 6.

Firstly, a small curved channel problem is simulated. The concrete wall is 122 cm thick, 80 cm high and 20 cm wide, with the density of 2.3 g/cm³. The radius of the channel is 5 cm with the air inside. The inner radius of the curve is 30 cm and the outer radius is 40 cm. A point photon particles of 5 MeV is emitted at the entrance of the channel, with the cosine of the polar angle at the source point with respect to the axis normal to the concrete wall being uniform distribution in [0,1]. The flux of the exit of the curved channel is calculated.

Three methods are used, with no variance reduction technique, with the DXTRAN technique and with auto importance sampling method. The source particle numbers

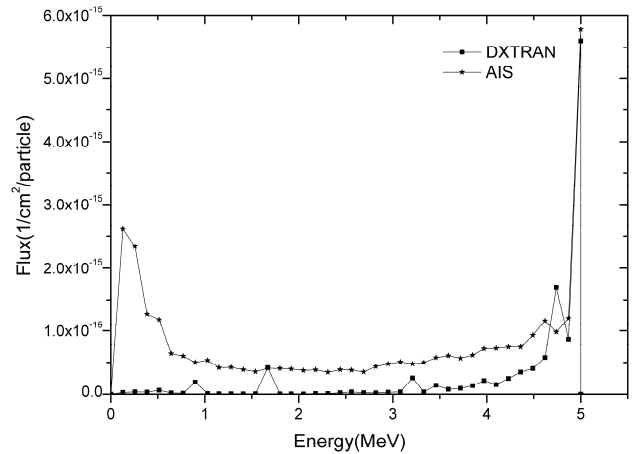


Fig. 8 The energy spectrums of the exit of the curved channel

Table 3 The results with different variance reduction techniques

VRT	N	P	RE (%)	T (min)	FOM
No	2E9	0	--	3915.50	--
DXTRAN	2E8	1.19E-14	7.54	4430.77	3.97E-3
AIS	1E6	3.24E-14	0.83	52.11	27.86

in the three methods are 10⁹, 10⁸, 10⁶, respectively. The results are shown in Table 2 and Fig. 7, in which 'DXTRAN' means simulating with the DXTRAN technique. The meanings of the other symbols are the same as in Table 1. In the AIS method, the geometry space is divided into eight sub-spaces.

From Table 2 and Fig. 7, we can see that the results of the three methods are very close. Though the computational efficiency of the DXTRAN technique is better than the analog method, the computational efficiency of the AIS method is two orders of magnitude higher than those two methods.

Secondly, a bigger curved channel problem is simulated. The concrete wall is 382 cm thick, 230 cm high and 20 cm wide. The inner radius of the curve is 160 cm and the outer radius is 170 cm. The others are all the same as the small curved channel problem described above. The same three methods are used, in which the source particle numbers are 2×10⁹, 2×10⁸, 10⁶, respectively. The results are shown in Table 3 and Fig. 8. In the AIS method, the geometry space is divided into twenty-one sub-spaces.

Referring to Table 3, it can be seen that with no variance reduction technique there is not any tally at all even simulating 2×10⁹ particles. With the DXTRAN technique though there are some particles reaching the statistical region, the tally result is not very good. However, from Fig. 8, we can see that the energy spectrum of the AIS method is smooth and physically correct. From Table 3 and Fig. 8, we can see that in very short time the AIS method can still achieve a very good result with small error.

3. Curved Channel Example – Neutron

Take another curved channel problem which is of neutron

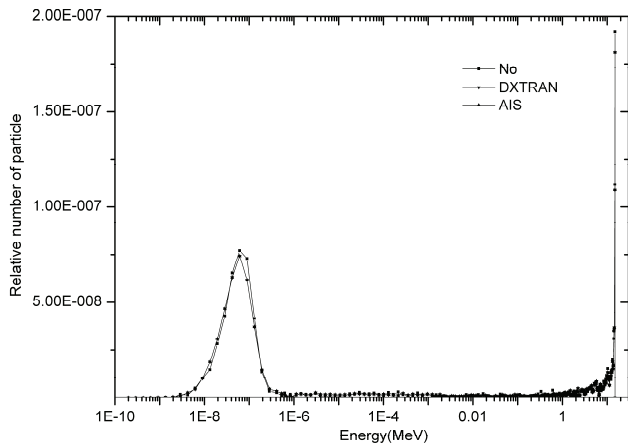


Fig. 9 The energy spectrums of the exit of the curved channel

Table 4 The results with different variance reduction techniques

VRT	N	P	RE (%)	T (min)	FOM
No	1E9	1.33E-6	1.12	2157.97	3.69
DXTRAN	1E8	1.31E-6	1.81	423.90	7.20
AIS	1E6	1.32E-6	0.43	75.21	719.10

source for example. The geometry is the same as that of the small curved channel problem described above. The wall is made up of polyethylene and the energy of the point neutron source particles is 15 MeV. The others are the same as the small curved channel problem described above. The same three methods are used, in which the source particle numbers are 10^9 , 10^8 , 10^6 , respectively. The results are shown in **Table 4** and **Fig. 9**. In the AIS method, the geometry space is divided into six sub-spaces.

Referring to Table 4 and Fig. 9, it can be seen that the results of the three methods are very close and the AIS method can achieve a very good result with small error in very short time. So the AIS method is also applicable to neutron problems.

V. Conclusion

This paper describes a new variance reduction technique for deep-penetration problem which is referred to as the auto-importance sampling method. In the new method, the

geometry space is divided into several sub-spaces by fictitious surfaces. The particles' number is increased on each fictitious surface by creating fictitious particles according to the importance distribution automatically, so as to ensure sufficient particles to reach the statistical region. Because the particles' number will be increased in the process of simulating, there needn't be a large amount of source particles which is of difficulty to simulate. With the new method, the setting of the parameters is much easier than that of the former variance reduction techniques. It is unnecessary to make many pre-calculations to test different parameters, or to make another calculation to get the importance distribution. And from the two examples it can be seen that the auto-importance sampling method can greatly increase the computational efficiency, about three orders of magnitude compared to the analog Monte Carlo method.

Acknowledgment

This paper is supported by National Natural Science Foundation of China (project number: 10775084).

References

- 1) T. E. Booth, J. S. Hendricks, "Importance estimation in forward monte carlo calculations," *Nucl. Technol./Fusion*, **5**[1], 90-100 (1984).
- 2) H. C. Gupta, "Importance biasing scheme for use with the expectation estimator in deep-penetration problems," *Ann. Nucl. Energy*, **11**[6], 283-288 (1984).
- 3) A. Haghghat, J. C. Wagner, "Monte Carlo variance reduction with deterministic importance functions," *Prog. Nucl. Energy*, **42**[1], 25-53 (2003).
- 4) H. P. Smith, J. C. Wagner, "A case study in manual and automated monte carlo variance reduction with a deep penetration reactor shielding problem," *Proc. ANS M&C Topical Meeting: Nuclear, Mathematical, and Computational Sciences: A Century in Review—A Century Anew*, April 6–10, 2003, Gatlinburg, Tennessee, American Nuclear Society, 555 N. Kensington Avenue, La Grange Park, IL 60525 (2003).
- 5) A. Dubi, "General statistical model for geometrical splitting in Monte Carlo-I," *Transp. Theory Stat. Phys.*, **14**[2], 167-193 (1985).
- 6) A. Dubi, "General statistical model for geometrical splitting in Monte Carlo-II," *Transp. Theory Stat. Phys.*, **14**[2], 193-221 (1985).
- 7) X-5 Monte Carlo Team, *MCNP—A General Monte Carlo N-Particle Transport Code, Version 5*, LA-UR-03-1987, Los Alamos National Laboratory (LANL) (2003).