A Transport Process Approach to Understanding Monte Carlo Transport Methods

Thomas E. Booth

Los Alamos National Laboratory, Diagnostic Applications Group X-5,
Mail Stop F663, Los Alamos, New Mexico, 87545 USA email teb@lanl.gov

Abstract. Monte Carlo particle transport is usually introduced primarily as a method to solve linear integral equations such as the Boltzmann transport equation. This focus on solving integral transport equations gives rise to a number of common misconceptions about MCNP transport methods among many MCNP users.

A transport process approach that is often useful in understanding the variance reduction in MCNP focuses directly on the Monte Carlo sampling process itself.

1 Introduction

Many MCNP users understand Monte Carlo transport theory via linear integral equations. This is quite understandable as standard books in the field usually emphasize the connection of Monte Carlo transport and the transport equation. This connection has proven very useful both for teaching about Monte Carlo and for developing and analyzing many variance reduction methods. The success of the books also indicates that the readers have found the transport equation approach to understanding Monte Carlo transport useful. Indeed, the fact that so many Monte Carlo books emphasize the transport equation indicates that the experts writing the books have found the transport equation a very useful perspective both in practice as well as in teaching.

Monte Carlo books are written for many purposes, for example, as course textbooks and to illustrate the main ideas in the field. To my knowledge, none of the standard Monte Carlo transport theory books were written with the intent of covering MCNP variance reduction methods; the books are intended for a more general readership. It would therefore be unreasonable to expect to understand MCNP variance reduction based solely on standard Monte Carlo transport theory books. For example, there is no standard book that justifies the unbiasedness of arbitrary combinations of MCNP's variance reduction techniques.

This report has four main purposes

1. to explain the transport process approach to variance reduction methods in MCNP
2. to educate MCNP users about the sometimes significant differences between the theory that they typically read in the Monte Carlo books and what MCNP actually does

3. to caution MCNP users about interpreting Monte Carlo transport using the wrong transport equation
4. to illustrate some aspects of statistical weight that are often not considered by MCNP users

The "transport process" approach focuses on the Monte Carlo transport process itself to try and determine what the major sources of variance are in the simulation and what to do about these sources. The transport process approach is a general approach that is useful for all MCNP calculations. The transport process approach to variance reduction can be especially useful when the user has no integral transport equation results available, as is often the case for typical MCNP calculations.

Before proceeding with the transport process approach, note that understanding the transport process is not a substitute for understanding integral transport equations. Integral transport equations can sometimes produce results that would be very difficult, if not impossible, to produce by focusing solely on a transport process approach.

By examining the transport process as one sets up the variance reduction in MCNP, one often finds that particles contributing most to the variance are somewhat special. They may, for instance, tend to

1. have fewer collisions than typical particles
2. have large distances between collisions in one direction and short distances between collisions in another direction
3. have different energies or directions than typical particles
4. have very different weights in a region than typical particles

Once the MCNP user has analyzed what is special about the particles contributing most to the variance, he may be able to use variance reduction techniques to increase the sampling of these special particles.

MCNP users sometimes try simply tinkering with the values of MCNP variance reduction parameters without having examined what types of particles are contributing most to the variance. Unless the user's intuition and/or luck is very good, this approach is usually far from optimal. A few observations about MCNP users and variance reduction in the next two paragraphs may help illustrate the usefulness of adding a transport process approach to their variance reduction considerations.

An MCNP user's intuition is often based on some (sometimes good, sometimes perhaps very rough) notion of the importance function. This is useful knowledge about a particle's expected score. One can sometimes glean useful information about typical scoring particles. On the other hand, in many high variance situations the typical scoring particle may have little influence on the variance because the variance may be dominated by atypical scoring particles. It is often difficult to have an idea what causes the variance in a calculation by knowing what causes the mean. Stated very simply, when attempting variance reduction it pays to focus on a measure of the variance rather than the importance, which is a measure of the mean. MCNP users should understand that the importance function deals with the mean score, not the variance.

My experience dealing with MCNP users is that they tend to focus their efforts on getting lots of low weight particles to the detector region. The clearest example is that most (novice) MCNP users intuitively view a dxtran sphere as "a magnet for pulling low weight particles" to the sphere rather than as "a shield against high weight particles" trying to cross the sphere. Dxtran can serve both purposes, of course. Using dxtran as a magnet tends to focus on bringing lots of typical low weight particles into a detector region. The contribution to the mean is then dominated by these low weight particles. On the other hand, the contribution to the variance is then often dominated by a few high weight particles. Roughly speaking, viewing dxtran as a magnet is consistent with producing typical scoring particles whereas viewing dxtran as a shield is consistent with precluding the high weight particles that might otherwise dominate the variance.

2 Comments on "Analog" Monte Carlo and the Transport Process

Probably most Monte Carlo transport practitioners understand and use the term "analog Monte Carlo" in mostly the same way. That is, convenient probability densities are abstracted from the physical transport process. These convenient probability densities are then embedded in a transport code. For example, the neutron distance to collision is sampled conveniently from an exponential distribution without modeling the detailed interactions between the neutron and each nuclide along its path.

For this report, the term "analog" describes a direct sampling of these abstracted probability densities. For the most part, people have abstracted very similar probability densities from the physical transport process. Nonetheless, it is probably worthwhile to note that an analog sampling in the context of this report refers to the particular probability densities that MCNP has abstracted from the physical transport process. Roughly speaking, an analog Monte Carlo sampling of a neutron transport problem in MCNP is what one gets when no variance reduction techniques are used.

The term "transport process" is also used in the context of MCNP's abstraction of the physical transport process. An analog transport process is an analog simulation of the abstracted physical transport process. Similarly, a nonanalog transport process is a nonanalog simulation of the abstracted physical transport process.

A transport process approach focuses on the details of the simulation to try and determine what the major sources of variance are in the simulation and what to do about these sources.

3 Comments on Mathematics

Although many people have commented that particle transport can be simulated without reference to the transport equation, this does not mean that mathematical equations are irrelevant to MCNP. Except for the simple case of an analog simulation, mathematical equations are necessary in both the transport equation approach and the transport process approach. When the simulation deviates from an analog simulation, both approaches require mathematical equations to show that the mean estimates are preserved.

4 Caution on "The" Transport Equation

In most cases, Monte Carlo codes allow estimation of quantities for which the transport equations displayed in the literature do not apply. In particular, the typical transport equations totally ignore the correlation between particles. Thus any estimate, such as the pulse height tally in MCNP, that depends on the correlation between particles is typically ignored. Transport equations, of course, can be written to include correlation between particles, but authors typically choose not to display such equations. If an MCNP user wishes to use transport equations to analyze and/or improve his Monte Carlo calculation, it is important to understand what transport equations are relevant to the calculation. This last statement seems obvious, but people have sometimes talked about the pulse height tally in MCNP in the same breath as a transport equation that ignores the correlation between particles.

5 Variance Reduction for Difficult Problems

For many problems, Monte Carlo estimates can often be obtained to sufficient precision using little or no variance reduction. This report assumes that the transport problems under consideration are sufficiently difficult that the Monte Carlo user needs to get close to the most efficient calculation that can be run with MCNP.

Note that one of the answers to a transport problem is the calculational variance; it is not governed by the transport equation. At this point, two typical approaches are:

1. One uses one's intuition (often guided by some knowledge of the importance function) to set up the variance reduction and assumes that whatever variance results will be close to optimum.
2. One derives equations for the variance (or sometimes the product of the variance of the mean and the computer time $cost = \sigma_m^2 T$) as a function of some parameters. One then solves these equations (sometimes using results from short Monte Carlo runs) for the optimum parameters. See [5, chapter 7] for a number of examples.

The first approach often works well if one's intuition is good and one can arrange the sampling so that there is a relatively small spread in history scores. That is, each history contributes roughly the mean score. In this case, both the bulk of the mean and the bulk of the variance are produced by the same particles. That is, focusing on particles that contribute most to the second moment is similar to focusing on particles that contribute most to the mean. Indeed, in the limiting case of a zero variance solution all particles contribute exactly the mean score and there is no reason to consider the second moment at all.

The other side of the coin is that the available variance reduction techniques may not allow one to arrange the sampling so that there is a relatively small spread in history scores. Alternatively, the variance reduction techniques may allow one to arrange the sampling so that there is a relatively small spread in history scores, but the user may not a priori be able to guess how. In these cases, the set of particles contributing most of the variance may be very different from the set of particles contributing most of the mean. For instance, the set of typical particles that contribute 99 percent of the mean might only contribute 1 percent of the variance. In this case, focusing efforts on typical particles that score will not work very well because the typical particles are very different from the particles contributing most to the variance. One would do better to base the variance reduction on the second moment equation rather than the importance.

Concerning the second approach, if one views the mean transport (first moment) as resulting from a solution of the transport equation, then it is natural (and often useful) to derive a similar equation for the second moment (e.g. see[5, chapter 5]). Unlike the transport equation, note that an equation for the second moment depends on what variance reduction methods one uses. (At this point, it is almost universal practice to consider only transport processes that are independent of particle weight.) The optimum second moment is, of course, optimal only for the particular set of variance reduction methods considered. This second approach works well when the variance reduction methods considered in the optimization are a good match to sources of variance in the problem. For example, particle penetration of simple slabs often can be done reasonably well by optimizing "cell importances" (the geometry splitting and Russian roulette technique in MCNP.) If there is a streaming path, such as a duct running through the slab, then even optimum cell importances may not sufficiently reduce the variance to make the problem tractable with reasonable amounts of computing time.

6 Two Theoretical Results for the Transport Process Approach

Most of the theoretical results presented in standard Monte Carlo texts apply to only small subsets of the MCNP variance reduction methods. There are three main reasons why these results are inadequate for MCNP purposes.

1. The results are typically restricted to weight independent simulations in which the random walk sampling does not depend on the particle weight.
2. The results do not consider arbitrary combinations of variance reduction methods, typically not even arbitrary combinations of weight independent methods.
3. MCNP is an open code so that users can add their own variance reduction methods if they wish.

Consider the combination of variance reduction techniques. Until about 1990, the following was true[6]:

> "Nonanalog Monte Carlo techniques are essential to many calculations, and historically they have been developed one at a time as needed. Each new nonanalog technique usually, at best, has been proven to preserve the expected tallies (i.e., be unbiased) when used by itself. The techniques have not been proven to be unbiased in arbitrary combinations."

Reference [5] has one of the better treatments, proving via integral equations that the combination of splitting and biased kernels is unbiased. If, in addition to splitting and biased kernels, another variance reduction technique is added, then one has to rewrite the integral equations to include the new technique and prove unbiasedness via the new integral equation. MCNP relies on the fact that *any combination of Monte Carlo techniques is unbiased if the techniques are individually unbiased.* In statistical parlance, any combination of fair games is also a fair game. This is a general statistical result for any linear Monte Carlo process in any field, not just the transport field[6].

Before leaving the subject of "combinations of fair games," note that [6] was controversial when it was reviewed. One reviewer initially sent in a two line review saying that the result was "well known" and "trivial" as well. When challenged to produce either a reference or supply a proof, the reviewer did neither. The paper was then accepted, although the reviewer emphasized that he still considered the result "trivial". Since publication, several knowledgeable Monte Carlo practitioners also have asserted that the fair games result can be proven easily. There is little reason to doubt that the proof in [6] is not the simplest possible proof. The MCNP documentation could be improved by inclusion of, or at least reference to, a simple proof that any combination of fair games is also a fair game. Please send any such proposed proofs (for the MCNP manual) to mcnp-forum@lanl.gov for assessment and comment by the MCNP community.

A second example of a general statistical result is the subject of zero variance methods. Zero variance solutions usually are derived solely in the context of importance biasing in the solution of an integral equation. To make matters even more specific, zero variance solutions are often derived only for last event estimators, instead of for general estimators. From much of the literature, one might erroneously conclude that there is only one way to get a zero variance solution. Most discussions in the literature ignore the fact that zero variance solutions can be obtained[3] using any other collection of variance reduction techniques in addition to importance biasing. The general rule to obtain a zero variance solution is to weight the sampling probability for each outcome to be proportional to the outcome's usual probability times the expected score generated if that outcome occurs. (Reference[3] shows one simple way to accomplish this sampling by expected score weighting the sampling of random numbers.) The usual derivation via importance sampling an integral equation is just one simple example of the general rule. Again note that this is a general statistical result for any linear Monte Carlo process in any field, not just the transport field.

7 A Caution on Optimal Sampling Claims for MCNP

A common mistake in verbal and written communications is to mix weight independent results from the literature together with the weight dependent transport processes in MCNP. Unless carefully qualified, such communications are often either false, misleading, or both.

A typical claim is that sampling from an importance weighted probability density minimizes the variance in an MCNP calculation. That is, if $f(P)$ is the true probability density for sampling the next phase space point and $I(P)$ is the importance function, then a typical claim is that sampling from the biased probability density

$$b(P) = \frac{f(P)I(P)}{\int f(P)I(P)\ dP}$$

minimizes the variance. Sometimes the claim is put forth as so obvious that no justification for the claim is supplied. Sometimes the claim is justified by reference to an importance sampling technique, despite the fact that the referenced importance sampling result was not derived in the presence of a weight dependent simulation.

Whenever one is using weight dependent MCNP techniques, one should be wary of assertions based on weight independent derivations in the literature. It is perhaps worthwhile to note that MCNP plays a number of weight dependent games by default. For example, if one wants an analog MCNP calculation then in addition to not explicitly requesting any variance reduction, one must explicitly turn off the default weight cutoff game. As a result, almost all MCNP calculations are weight dependent simulations and the standard results from weight independent theories do not apply.

8 Comments on Transport Equations and Variance Reduction in MCNP

Given the emphasis that standard books give to integral transport equations when discussing variance reduction, it is perhaps worthwhile to summarize the current state of affairs with regard to transport equations and MCNP. Three reasonable questions are:

1. Why does MCNP use weight dependent simulations?
2. Why not analyze weight dependent MCNP simulations using integral transport equations?
3. Why analyze MCNP simulations using a transport process approach?

Weight independent simulations can be viewed as a special case of more general weight dependent simulations in the same sense that $f(x) = constant$ can be viewed as a special type of function of x. Philosophically, it is not very surprising that selecting from a broader class of variance reduction techniques might allow for better variance reduction, but the primary reason that weight dependent simulations are allowed in MCNP is that they have proven useful in practice.

The answers to the second and third questions are a bit more difficult and the answers may change eventually depending on future developments in Monte Carlo transport theory. At the present time, here are some thoughts on why transport equation analysis might be useful for weight dependent simulations and why a transport process analysis is currently useful for weight dependent simulations.

1. Transport equations for the variance, usually via second moment equations, (e.g. [5, chapter 5]) are currently almost exclusively for weight independent simulations. There is apparently no essential difficulty deriving weight dependent transport equations for the variance. To date though, I know of no use ever made of such an equation. This may change in the future.
2. Note that integral transport equations often are concise and easily interpretable. The fact that integral transport equations are an average over the transport process can be a big advantage if one can effectively use a transport equation that averages over unimportant aspects of the transport process while preserving the important aspects of the transport process.
3. Although the transport process details have to have at least as much information as any integral equation average over the process, there is always the danger that useful general insights get lost in the details.
4. Inasmuch as nobody has figured out yet how to effectively use integral equations for the second moment in typical weight dependent MCNP calculations, it is difficult to assess what general insights (from the second moment equation) currently might be hidden in the transport process details.
5. A transport process approach to variance reduction in MCNP is somewhat of a necessity given the comments in items 1 and 4.
6. Many different fields have similar types of simulation processes so that techniques used for processes in one field are often useful techniques in other fields as well.

7. By examining in detail the particles having the largest contributions to the variance, it is often quite easy to identify the major source of variance still left in an MCNP problem.

9 Markov and Nonmarkov Processes and Random Walks

In MCNP's analog simulation of nature, the next step of a particle's random walk depends only on its current phase space location P. That is, MCNP's analog process is a Markov process.

Nonanalog simulations of particle transport depart, in one way or another, from the analog process. Nonanalog methods are also known as variance reduction methods because the intent of using nonanalog methods is to reduce the variance in the estimated mean for a given computer time. Note that nonanalog simulations need not be Markov processes.

The following sections will comment on four categories of random processes

1. Processes depending only on the current phase space location P.
2. Processes depending on all the random walk's past physical events, e.g. $P_0, P_1, P_2, \cdots, P_n$.
3. Processes depending both on the current phase space location P and the current statistical weight w.
4. Processes not included in the previous items.

10 Natural Markov Processes

When the sampling of the particle depends only on the current phase space position P, as it does in an analog MCNP modeling of nature, the sampling will be said to be a natural Markov process. Many of the common variance reduction techniques are naturally Markovian. For example:

1. Biasing the transport kernels.
   – exponential transform in MCNP (path length stretching)
2. Splitting techniques
   – geometry splitting and Russian roulette in MCNP
   – survival biasing in MCNP (split into absorbed and surviving parts)
   – forced collisions in MCNP (split into collided and uncollided parts)

(MCNP is not remarkable in having variance reduction techniques that are naturally Markovian; nor are the above techniques necessarily unique to MCNP. Transport codes sometimes differ both in their terminology and in their implementations of techniques having similar names. Specifying an MCNP technique makes the definition unambiguous, so that the categorization above is possible.)

There are variance reduction methods that are natural Markov processes and there are variance reduction methods that are not natural Markov processes. In general, the natural Markov processes are easier to study mathematically because the next step of a particle's random walk depends only on its current phase space location. Because of this simplicity, the pages of Monte Carlo transport theory literature devoted to natural Markov processes far exceeds the pages devoted to other Monte Carlo processes.

11 A Simple Nonmarkov Process

Reference[2, page 87] discusses a simple nonmarkov process in which the transport probabilities can depend on the current and all the previous phase space points, $P_1, P_2, \cdots, P_n$. On the same page, the book says " ... for those familiar with the term, we shall be dealing almost exclusively with Markov processes. Nonetheless, it seems worth pointing out that nonmarkov processes may be treated as well." Indeed, the remainder of the book almost exclusively considers transport problems with analog *and* nonanalog Monte Carlo simulations that depend solely on the current phase space point.

From an MCNP perspective, the reason that [2] usually does not apply to MCNP calculations is not so much that the previous phase space points are not considered, it is that the particle weight is not included in the current state of the particle. That is, the nonanalog Monte Carlo simulations in [2] only seem to use weight independent random walks.

Returning to the book's[2] notion of a nonmarkov random walk, note that including the previous phase space points is occasionally useful. For example, MCNP allows consideration of events before the current phase space point P via the cell and surface flagging options. If a cell is flagged, then the tally is partitioned into the part of the tally due to particles that have entered the flagged cell and the part of the tally due to particles that have not entered the flagged cell. The surface flag operates similarly. Although the production version of MCNP does not use different transport methods depending on the flag, it is easy and occasionally useful to modify MCNP to do so.

12 Weight Dependent Markov Processes

Many, probably most, of the common variance reduction techniques that are not natural Markov processes depend only on the the current phase space point P and the current weight w. Those processes whose transport depends only on (P, w) will herein be called weight dependent Markov processes to distinguish them from natural Markov processes.

Common weight dependent Markov processes in MCNP involve:

1. weight windows
2. weight cutoff (associated with the geometry splitting/roulette)

3. weight dependent roulette games associated with point detectors and dxtran
4. weight dependent secondary particle production

13 Weight Dependent vs Weight Independent Transport

Most theoretical Monte Carlo discussions assume that a particle's random walk is independent of the particle's weight. Under this assumption, a particle's score is directly proportional to its weight and the r^{th} score moment for a particle of weight w is w^r times the r^{th} score moment for a unit weight particle. [5, page 163]. To give some idea of the common appeal of this wide-reaching assumption, note that [5] first mentions this assumption in a footnote.

A cautionary note is perhaps worthwhile here. Because weight independent (natural) Markov simulations are more tractable mathematically, they account for almost all of the theoretical discussions in the Monte Carlo literature. (Two good exceptions can be found in [5, pages 178 and 186].) One should not be mislead into concluding that weight independent simulations are more important, better, or more widely used than weight dependent simulations. Many of the large production Monte Carlo codes allow weight dependent simulation. MCNP, which is probably the most widely used Monte Carlo transport code in the world, has always done weight dependent simulation as a default. (To the author's knowledge, the predecessor codes to MCNP, as far back as the 1950's, have always done weight dependent simulation as a default as well.)

There is often some distance between Monte Carlo theory and MCNP practice. Two examples are given below.

First, consider the weight window technique. The weight window is perhaps the most widely used variance reduction technique in transport Monte Carlo today, but it has received scant theoretical attention. (Fox's book [7, pages 213-233] gives an interesting discussion of the weight window.)

Second, consider Monte Carlo optimization techniques. There are numerous theoretical derivations on optimal parameters to minimize the variance; they almost always assume weight independent transport. A favorite problem for theorists is optimizing the exponential transform[8–11]. (The reference list is not exhaustive, see [5, page 487] for more.) Inasmuch as practical experience indicates that a weight window almost always improves the performance of the exponential transform, the usefulness of optimizing the exponential transform in the absence of a weight window is severely curtailed. Empirically,

1. The optimal transform parameter seems to be higher with a window than without a weight window.
2. An empirically optimized transform parameter used with a weight window can give very good results. When the weight window is removed

with the same transform parameter, the results are often disastrous. In one documented case [12, pages 54-56], the efficiency decreased by a factor of 100.

Because of item 2, MCNP issues a warning message if the exponential transform is used without a weight window.

Several historical points in connection with exponential transform optimization and weight windows are worthwhile.

1. The references cited in the above paragraph generally predate the widespread use of weight windows, so that the optimization techniques were useful in their time. Additionally, they are still useful for Monte Carlo codes besides MCNP.
2. The author knows of no studies, theoretical or empirical, that demonstrate a benefit to using optimized transform parameters without a weight window.
3. The optimization of the exponential transform in combination with a weight window has not been attempted (except by empirical testing).
4. The author developed the weight window for MCNP after studying poorly behaved statistical results obtained while using the exponential transform. No integral equations were considered in developing the weight window. The weight window was developed after observing the random walk process for particles that produced the poor statistical results. That is, the analysis and subsequent corrective action was focused on the transport process and not integral equations. (Note, however, that proving that the weight window method is unbiased does require integrals to show the mean score is preserved.)
5. The weight window method then necessitated a way to obtain the weight windows. MCNP's weight window generator was then devised, again based solely on the transport process. No integral equations were considered in developing the weight window generator. (Although not necessary, note that the weight window generator concept also can be obtained via integral equations.)

14 Some Other Nonanalog Transport Methods

Standard Monte Carlo transport books understandably attempt to explain Monte Carlo variance reduction methods by focusing on a few powerful and relatively easily understood methods. The transport books are not intended to be encycopedias of all possible variance reduction methods, nor should they be. That said, many MCNP users unduly seem to have limited their views to the types of methods mentioned in the Monte Carlo transport books. This is unfortunate, because there is a large variety of possible unbiased Monte Carlo methods that are unlike the methods typically discussed.

Below are some possible nonanalog methods that are usually ignored. The comb method in item 1 is a practical nonanalog method that has been used for many years. The rest of the methods are just thought experiments and probably have never been implemented anywhere. No suggestion is being made that the methods are useful, only that they can be unbiased methods. Where possible, a rationale for the method is given to aid the reader's understanding of what might motivate one to consider such a method. Methods 6 and 7 are downright farcical from the standpoint of variance reduction, but they help indicate the generality of possible nonanalog simulations.

1. When the number of tracks associated with one source particle exceeds 100, then use an importance-weighted comb[13] to reduce the number to 50. Note that the comb uses the phase space location and weight of each particle, thus the random walk of each particle now depends on the weight and phase space locations of all the particles.
2. In the Monte Carlo literature, for example [2,4,5], the nonanalog transition kernel $\hat{K}(P, P')$ between collisions is almost always assumed to be independent of the particle weight. This may not be optimal. Consider two particles, with weights $w_1 > w_2$, penetrating a slab. Suppose that both particles are identical, except for their weights, and both are moving forward in the penetration direction. If w_1 is large enough, then modifying the distance to collision sampling by using an exponential transform will produce a transform modified weight w_{1t} that is larger than some minimum weight requirement at P. Conversely, if w_2 is small enough, then modifying the distance to collision sampling by using an exponential transform will produce a transform modified weight w_{2t} that is smaller than some minimum weight requirement at P and a roulette game will ensue. First reducing the particle weight via exponential transform and then playing a roulette game introduces an unnecessary fluctuation in the particle weight; this will generally lead to a higher variance simulation. Thus, it might make sense to employ the exponential transform only in those cases where the transform modified colliding weight will be above some minimum weight requirement at P. That is, it may make sense to use a weight dependent transition kernel $\hat{K}(P, P', w')$.
3. When a particle track bifurcates, either due to variance reduction or a physical process like fission or electron pair annihilation, the sampling of one branch can be made dependent on the sampling of another branch. For example, suppose the third photon collision produces two 0.511 MeV annihilation photons. Put the second branch aside ("save it to the bank") while the first branch is sampled. Note whether the first branch (or any of its progeny) reaches a detector cell. After finishing with the first branch, sample the second branch using an exponential transform if the first branch reached the detector; otherwise, sample the second branch normally. This will increase the number of times that both 0.511 MeV photon branches reach the detector and might thus be helpful in a pulse height tally calculation.
4. The randomness in an estimation process can depend on the randomness in a previous estimation process. For example, suppose that roulette

games are played in the estimator process that with probability p increases the contribution by $1/p$ or with probability $1 - p$ takes a zero contribution. The point detector estimator in MCNP plays a series of these roulette games[1, page 2-98] as it tracks a pseudoparticle from the collision site to the detector. Suppose that the particle has not moved far from its previous collision before it collides again. If the detector contribution from the previous collision lost one of the roulette games after a good deal of tracking work, then it might be reasonable to play some of these roulette games for the current collision before taking the computer time to track a pseudoparticle towards the detector.

5. The random walk of a particle can depend on the randomness in the estimator. As in item 4, consider a random point detector estimator. Change the particle's random walk conditional on the randomness in the detector.

   – If the pseudoparticle survives the point detector roulette games at the n^{th} collision, then the particle proceeds to sample the distance to the $n + 1^{st}$ collision.

   – If the pseudoparticle is rouletted in the estimation process, then roulette the particle with probability 1/2 before sampling the distance to the $n + 1^{st}$ collision.
6. Suppose there is a 2:1 split before the fourth collision. One can sample the post collision energy from the sixth collision of the second branch depending on the outcome of the 43rd collision on the first branch.
7. Note that the random numbers determining the outcome of the 7^{th} collision can be selected before the the sampling of the 4^{th} collision. The 4^{th} collision can then be sampled from a biased probability density dependent on the random numbers for the 7^{th} collision. This is not only not a Markov process, it violates some peoples' notions that events must be sampled in the order that they occur and that the sampling of earlier events cannot depend on later events.

15 Comments on Statistical Weight

From the standpoint of many, perhaps most, major Monte Carlo transport codes, weight is a particle attribute, like energy and position. That is, the weight is carried along with the particle, banked with the particle, and so forth. It is often convenient to interpret the weight as the number of physical particles represented by the computer particle. Heuristically, one expects that if the Monte Carlo process preserves the expected weight at each event, then the result will be an unbiased mean. For the most part, this is a very useful view of the Monte Carlo process, but it is perhaps useful to point out some cases for which this view needs some elaboration and/or modification. The purpose here is to illustrate some of the subtleties in the concept of "particle weight" that MCNP users may not have considered.

15.1 Preserving the Expected Weight is not Always a Sufficient Condition for an Unbiased Mean

Preserving the expected weight, by itself, will not ensure an unbiased estimate. The estimator must depend on weight in a correct way also. As an obvious example, if the number of particles crossing a surface is desired, then tallying "1" (regardless of weight) every time a particle crosses the surface will give the correct tally for an analog calculation, but will in general be wrong when variance reduction techniques change the weight.

For deterministic (nonrandom) estimators, unbiasedness is normally assured by making the tally function proportional to weight. Not all common estimators are deterministic. The point detector in MCNP[1, 3-106] is a random estimator because it plays roulette games when the optical path to the detector gets large. For random estimators, one requires that the expected tally (rather than the individual tally itself) be proportional to weight.

The conceptual mistake many people make is to separate the estimation process from the transport process. These two processes can be tied rather intimately in some unusual ways and one has to ensure that the combined process is unbiased. Consider estimating the number of particles that cross the cell shown in Fig. 1 without colliding.

```
                         Figure 1

                  Typical Estimation Process

           |<------ T ------->|
           |                  |
           |                  |
   w       |----------------->| probability p, no collision
 -------->|                   |
           |------->*         | probability 1-p, no collision
           |                  |
           |                  |
           |                  |
```

A typical transport and estimation procedure is: with probability $p = exp(-\sigma T)$ the particle crosses the cell without collision and tallies w and with probability $1 - p$ the particle collides and no tally is made. The particle is then followed from either the point where it crossed the surface or the point where it collided.

Another possible way to estimate the number of collisionless flights across the cell is shown in Fig. 2.

```
                        Figure 2

                  Another Estimation Process

            |<------ T ------->|
            |                  |
            |                  |
     w      |----------------->| probability p, no collision
   -------->|                  |
            |------->*         | probability 1-p, no collision
            |                  |
            |                  |
 pseudo-    |                  |
 particle   |                  |
            |----------------->| probability p, no collision
   -------->|                  |
            |---------->*      | probability 1-p, no collision
            |                  |
            |                  |
```

For Fig. 2, the tally is not dependent on whether the particle of weight w collides or not. Instead, the estimation is done using a "pseudoparticle" that only exists for the estimation procedure. (The pseudoparticle initially has the same phase space coordinates as the transported particle.) The pseudoparticle is sampled using the same probabilities as the transported particle, but the pseudoparticle is terminated after the estimation procedure is completed, it is not transported. Hence the term pseudoparticle, because it is not part of the transport. Transport then continues with the original particle. Thus the particle might not cross the surface without colliding, but it might contribute to the tally because the pseudoparticle did cross the surface without colliding. Note that if the estimation process used on the pseudoparticle is not correct, then the estimate can be erroneous despite the fact that the expected particle weight has been preserved. It is often so obvious how the pseudoparticle should be treated that the pseudoparticle's role in maintaining an unbiased estimate is not discussed. (A correct method tallies w when the pseudoparticle crosses the surface without collision.)

15.2 Preserving the Expected Weight is not Always a Necessary Condition for an Unbiased Mean

The previous subsection showed that preserving the expected weight is not always a sufficient condition for an unbiased mean. Now, it is shown that preserving the expected weight is not always a necessary condition for an unbiased mean. Experienced Monte Carlo practitioners correctly might suspect some legerdemain here. Consider Fig. 2 again. Inasmuch as the tally depends (for the current transport step) not on the particle's weight, but on the weight associated with the pseudoparticle, the particle weight can be set to any arbitrary value, provided the particle weight is returned to w when the particle collides or crosses the surface. Thus, preserving the expected weight is not necessary for this step in the transport process. With the tally not responding to the original particle, one possible interpretation is that the particle weight is zero for that step. Things will get even more curious in the next subsection.

15.3 Multiple Particle Weights

Particle weight is normally conceived of as a single value for each particle. Not only can one conceive of particles having multiple weights, multiple weights are used in some production transport codes. Before jumping to the practical uses of multiple weights, two simple examples are discussed.

Building on the previous two subsections, suppose that the code uses two *different* estimators for the number of particles crossing the surface. The first estimator uses the original particle as in Fig. 1 and the second estimator uses the pseudoparticle as in Fig. 2. In this case, the original particle should have weight w so that the first estimator is correct, but it can still have zero weight for the second estimator. That is, the particle can have a different weight for each estimator.

For another simple example, suppose that a particle of weight w reaches a surface as shown in Fig. 1. Upon crossing the surface, split the particle into two particles each of the original weight w. The total expected weight is not preserved by this split, but unbiased estimates can again be made by a bit of legerdemain with the estimators. Label the particles 1 and 2. Label the estimators with positive integers. Let the odd numbered estimators respond only to particle 1 and let the even numbered estimators respond only to particle 2. This can be viewed as follows. The presplit particle contributed to all tallies and thus can be considered to have a weight vector (w, w). After the split, particle 1 has weight vector $(w, 0)$ and particle 2 has weight vector $(0, w)$.

Turning to practical uses of multiple weights, note that perturbation and correlated sampling methods use different weights for the reference system and the perturbed system. For example, Ref. [5, page 307] explicitly uses a weight vector in the discussion of correlated sampling.

The dxtran method in MCNP is very similar to the second example above. Upon surviving a collision, a particle is partitioned into two particles. The "dxtran particle" represents the uncollided particles that arrive on a user specified dxtran sphere. The "nondxtran particle" represents the remainder of the particles. Note that the nondxtran particle has the original weight, w, at the collision exit point and the dxtran particle has a nonzero weight. Thus, the total particle weight is always larger than w at the collision exit point. The trick here is that the dxtran particle has zero weight for any tallies made before crossing the dxtran sphere and appropriate weight for any tallies afterward. Conversely, the nondxtran particle has weight w for all tallies made before crossing the dxtran sphere and zero weight for any tallies afterward.

Multiple weights can also be used to get low variance estimates for multiple tallies. Consider a particle with a single weight in a slab penetration problem. Suppose the numbers of particles exiting the slab in the three energy ranges 1.00 to 1.01, 1.01 to 1.02, and 1.02 to 1.03 MeV are desired. Note that a zero variance sampling for the energy range 1.00 to 1.01 MeV means that every particle has to exit the slab within this energy range. This means that no particles exit in the other two energy ranges. Thus, a random walk process that gives a zero variance estimate for one energy range gives an infinite variance estimate for all other energy ranges. Most of the sampling to get a zero variance solution in one interval is going to be very similar to the sampling to get a zero variance solution in either of the other two intervals. It seems ridiculous that a zero variance solution in one interval forces an infinite variance in the other intervals. Reference [14] shows that it is possible to get zero variance solutions in all three intervals at once using particles that carry three weights. The method works by simultaneously applying several different importance functions, one for each tally, in a correlated way. Although zero variance estimates are impractical because the importance functions are not known exactly, low variance solutions are possible with approximate importance functions. The method in [14] follows a single particle with multiple nonzero weights until the correlation between the importance functions decreases enough that the particle must, statistically, execute different random walks for different tallies. (Curiously, Monte Carlo theories seem to focus on a single importance function despite the fact that multiple estimates are usually sought.)

16 Practical Variance Reduction

Designing practical variance reduction methods using solely an integral equation approach is often problematical. First, the methods designed via an integral equation approach almost always are weight independent methods because weight dependent methods are not usually analyzed by integral equations. Second, minimization of the variance in the transport process is usually limited to making approximate use of zero variance biasing or optimizing a fixed set of parameters associated with the method (e.g. see [5, chapter 7]). At the end of this optimization, one typically has some roughly optimized set of parameters that minimize the variance in a weight independent simulation. What one does not usually have is an understanding of the remaining *sources* of variance in the simulation after the optimization.

It may be that the variance reduction method, though optimized for the particular set of parameters, is not treating some important source of variance in the simulation. Consider, for example, neutron penetration of an iron slab. One can optimize the geometry splitting/roulette parameters in MCNP, but the calculation may still have a large source of variance associated with inadequate sampling of the iron cross section window at 24 kilovolts.

As a practical matter, it is usually important to understand the source of any remaining variance in the problem. Unless one understands the source of the variance, it is difficult to know if any of a code's standard variance reduction methods attack the source of the variance. Stated another way, once the source of the remaining variance cannot be attacked by any of the code's standard variance reduction techniques, then either

1. the user's variance reduction efforts should end, or
2. a new variance reduction method that attacks the source of the variance must be implemented in the code

Fortunately, understanding the source of variance often is not too difficult. One finds the source particles that contributed most to the tally (MCNP does this automatically for the largest contribution) and one looks at these source particles in detail either via a printout for the particle (an "event log" in MCNP) or via a debugger. If one cannot find anything indicating poor sampling (e.g. hitting an iron window, streaming up a duct, or excessively high weights) then the simulation should be reasonably efficient. On the other hand, if the particle having the largest tally is rare in the sense that it sampled an important pathway that almost all other particles miss, then the user can examine the variance reduction methods in the code that might increase the sampling frequency for this rare pathway.

Note that many of the deficiencies in Monte Carlo variance reduction techniques can be corrected by introducing weight dependent games. In the past, the exponential transform at Los Alamos was often described as a "dial an answer technique" because the sample mean was sometimes extremely unstable and often seemed to depend on what transform parameter was used. The weight window easily corrected this problem. The stratified splitting game suggested in [4] sometimes was found to have a higher variance than the standard weight window splitting in MCNP. From a theoretical point of view, it was difficult to understand why the stratified splitting game was worse than the unstratified weight window splitting. After examining the stratified splitting transport process by following a few particles around, the situation became clear. An analysis of the *cause* of the higher variance associated with the stratified splitting technique pointed the way toward a small (weight dependent) modification of the technique that made the stratified splitting better than the weight window splitting[15].

As a practical matter, people need to pay attention to the sources of variance in the transport process when considering variance reduction methods.

17 Comments on Theory and Tinkering

A person confronted with solving a Monte Carlo transport problem today has a variety of variance reduction techniques that can be applied in the various transport codes. The Monte Carlo literature tends to focus on variance reduction techniques that have been analyzed using integral equations. Unfortunately, only a small fraction of MCNP calculations fit the cases described in the literature. The usual culprit, as indicated in numerous instances herein, is the presence of weight dependent games in the MCNP calculations. More theory is needed for these weight dependent games. Weight dependent games in MCNP seem to allow more efficient simulations than the weight independent games described in the literature.

With the important exceptions of choosing weight windows or MCNP cell importances, which can be obtained by stochastic methods (e.g. MCNP's weight window generator) or sometimes deterministic methods (e.g. discrete ordinates[16]), very little of the variance reduction is automatic. In practice, some MCNP users attempt variance reduction by tinkering. They tinker without looking at the transport process to determine the source of the variance. They tinker with different methods and they tinker with the parameters of the methods. Quite often, the tinkering is ineffective, for reasons explained below.

With enough effort, the user empirically can optimize a set of parameters, but the optimum is only over the methods and sets of parameters chosen. Without investigating particles to determine what is causing the variance, the tinkering is essentially done in the dark, with the hope that the optimum parameter selection will lead to an efficient calculation. Suppose, for instance, that the user attempts to do a slab penetration problem using solely the exponential transform. With enough tinkering, the user will presumably arrive near the optimum transform parameter obtainable by references [8–11]. As mentioned in section 13, practical MCNP experience indicates that the exponential transform without a weight window gives a very inferior result compared to the exponential transform with a weight window. Whether the optimum is theoretically derived or empirically derived, the calculation is missing weight control as a key factor in controlling the weight fluctuations introduced by the exponential transform technique. A quick look at the particle history contributing the largest tally highlights the problem almost immediately.

18 Summary

This report has noted some significant differences between the way Monte Carlo transport theory is normally presented and what MCNP actually does. Additionally, this report has tried to show that there is a significant value in a transport process approach to variance reduction in MCNP. That is, MCNP users should understand not only the integral equation approach to variance reduction, but the transport process approach as well. Note that the two approaches are not mutually exclusive. This report gives two examples of useful techniques (i.e., stratified splitting and the exponential transform) that were developed based on transport equation considerations and then were improved by transport process considerations.

As a practical matter, MCNP users need to pay attention to the sources of variance in the transport process when considering variance reduction methods. The cause of a high variance simulation in MCNP is not often apparent from a look at the transport equation; in contrast, the cause is often apparent after examining the largest scoring particle histories.

Finally, this report (especially sections 14 and 15) encourages MCNP users to take a broad view of Monte Carlo variance reduction.

Acknowledgement

Though they do not necessarily agree with everything in this mansuscript, the author wishes to acknowledge helpful comments and discussions with Malvin Kalos and Jerome Spanier.

References

1. X-5 Monte Carlo Team, "MCNP-A General Monte Carlo N-Particle Transport Code, Version 5," Los Alamos National Laboratory Report LA-UR-03-1987, April 24, 2003
2. Jerome Spanier and Ely M. Gelbard, "*Monte Carlo Principles and Neutron Transport Problems*," Addison-Wesley Publishing Company (1969).
3. Zero-Variance Solutions for Linear Monte Carlo, Thomas E. Booth, Nuclear Science and Engineering: 102, 332-340 (1989)
4. Malvin H. Kalos and Paula A. Whitlock, "*Monte Carlo Methods Volume I: Basics,*" John Wiley and Sons (1986).
5. Iván Lux and László Koblinger, "*Monte Carlo Particle Transport Methods: Neutron and Photon Calculations*," CRC Press, Inc. (1991).
6. Unbiased Combinations of Nonanalog Monte Carlo Techniques and Fair Games, Thomas E. Booth and Shane P. Pederson, Nuclear Science and Engineering: 110, 254-261 (1992)
7. Bennett L Fox, "*Strategies for Quasi-Monte Carlo,*" Kluwer Academic Publishers 1999

8. An Analytic Approach to Variance Reduction, J. Spanier, SIAM J. Appl. Math. 18 172 (1970)
9. A New Multistage Procedure for Systematic Variance Reduction in Monte Carlo, J. Spanier, SIAM J. Appl. Math. 18 172 (1970)
10. Optimal Choice of Parameters for Exponential Biasing in Monte Carlo, A. Dubi and Donald J. Dudziak, Nuclear Science and Engineering: 70, 1-13 (1979)
11. Prediction of Statistical Error and Optimization of Biased Monte Carlo Transport Calculations, P. K. Sarkar and M. A. Prasad Nuclear Science and Engineering: 70, 243-261 (1979)
12. A Sample Problem for Variance Reduction in MCNP, Thomas E. Booth Los Alamos National Lab. Report: LA-10363-MS, 1985 (available electronically via http://www-xdiv.lanl.gov/x5/MCNP/thedocumentation.html)
13. A Weight (Charge) Conserving Importance-Weighted Comb for Monte Carlo T. E. Booth 1996 Radiation Protection and Shielding Division Topical Meeting, American Nuclear Society, N. Falmouth, MA, April 1996
14. Simultaneous Monte Carlo Zero-Variance Estimates of Several Correlated Means, Thomas E. Booth, Nuclear Science and Engineering: vol 129, June 1998
15. Insights into Stratified Splitting Techniques for Monte Carlo Neutron Transport, Thomas E. Booth, Los Alamos Report LA-UR-03-5955, submitted to Nuclear Science and Engineering
16. Wagner, J.C. and Haghighat, A., "Automated Variance Reduction of Monte Carlo Using the Discrete Ordinates Adjoint Function," Nuclear Science and Engineering, Vol. 128,1998, pp. 186-208.