http://dx.doi.org/10.5516/NET.01.2012.502

CHALLENGES AND PROSPECTS FOR WHOLE-CORE MONTE CARLO ANALYSIS

WILLIAM R. MARTIN
Department of Nuclear Engineering and Radiological Sciences
University of Michigan
*Corresponding author. E-mail : wrm@umich.edu

Invited March 23, 2011
Received January 09, 2012
Accepted for Publication February 19, 2012

The advantages for using Monte Carlo methods to analyze full-core reactor configurations include essentially exact representation of geometry and physical phenomena that are important for reactor analysis. But this substantial advantage comes at a substantial cost because of the computational burden, both in terms of memory demand and computational time. This paper focuses on the challenges facing full-core Monte Carlo for keff calculations and the prospects for Monte Carlo becoming a routine tool for reactor analysis.

KEYWORDS : Monte Carlo Analysis, Full-core Reactor Analysis, Parallel Processing, Fission Source Acceleration, Hybrid Monte Carlo, Multiphysics Feedback, NEA Full-core Benchmark Problem

1. INTRODUCTION

The advantages of the Monte Carlo method for reactor analysis are well known. Continuous energy Monte Carlo codes such as MCNP5 [1] are capable in principle of analyzing reactor configurations with arbitrary geometrical complexity – limited by the ability of the code (and patience of the user) to represent arbitrary shapes in a computational model; and arbitrary physics complexity – limited by the knowledge of cross sections that describe the physical phenomena being modeled. In addition to the flexibility of Monte Carlo to simulate the most complex geometry, continuous energy Monte Carlo treats neutron energy dependence correctly with essentially no approximations. In particular, Monte Carlo makes no "operator split" approximations in energy which are needed for deterministic transport in order to determine multigroup cross sections. Methods that use multigroup cross sections, whether deterministic or Monte Carlo, have intrinsic errors associated with the cross section processing due to the crude spatial model during this step of the overall calculation sequence. Every step down the path from evaluated cross section data to few-group cross sections involve trade-offs between spatial resolution and energy resolution, formally equivalent to operator splitting back and forth between space and energy. For example, the complexity of resonance processing for lattice physics codes is due to the need to have multigroup cross sections that represent phenomena on an energy scale far smaller than the width of the energy group, and this energy dependence depends on the spatial solution.

Moreover, previous work has demonstrated that Monte Carlo methods can be made to run efficiently on most if not all production computer architectures that have been introduced to date [2]. So why isn't Monte Carlo the method of choice for full-core reactor analysis? The primary reason for this state of affairs is the extreme computational burden to carry out a full-core Monte Carlo simulation, including prohibitive computational time and excessive memory demand. A number of factors contribute to the computational burden inhibiting full-core Monte Carlo, including:

- Prohibitive computational time for acceptable statistics
- Excessive demand on computer memory
- Slow convergence of the fission source
- Apparent versus true variance
- Accounting for multiphysics feedback
- Adapting to future computer architectures

These six (not necessarily independent) factors present challenges for the use of Monte Carlo for full-core reactor analysis and the remainder of this paper discusses these challenges and the prospects for overcoming them.

There are other challenges facing Monte Carlo such as uncertainty propagation due to depletion and multiphysics feedback or the correct treatment of resonance upscattering or unresolved resonances, but these topics (and others) are related more to the fidelity of the physics models rather than computational efficiency and will not be treated in this paper. However, it should be noted that computational time for Monte Carlo simulation is generally proportional to the fidelity of the physics, so it is likely that improvements in physics fidelity will not improve Monte Carlo performance although they may improve Monte Carlo predictions. The remainder of this paper discusses each of the six challenges, followed by a section on the development of a 3D full-core benchmark problem that will enable the measurement of progress towards the goal of making Monte Carlo a tool for routine design and analysis of full-core reactor configurations. The paper concludes with a summary of the challenges and prospects for achieving full-core Monte Carlo. The review is focused on full-core k_{eff} calculations with Monte Carlo, and does not address time-dependence or source calculations.

The following discussion is based on a number of papers, conference talks, and seminars over the past few years, including a special session, "Monte Carlo Methods in Reactor Physics: Current Status and Future Prospects," held at the 2011 ANS Summer Meeting [3-6].

2. PROHIBITIVE COMPUTATIONAL TIME FOR ACCEPTABLE STATISTICS

The computational time for Monte Carlo simulation is proportional to the number of neutron histories needed to obtain acceptable statistics for local flux/power estimates and the computational time per history. Using a Monte Carlo solution for an assembly as a basis, Smith [5] estimated that 100 billion histories would be needed to achieve 1% statistics on local power/flux tallies for a full-core calculation. Given the fact that local statistics are determined by the track length density and the total track length is roughly proportional to the size of the reactor configuration, it is unlikely that anything can be done to address the number of histories so the reduction in computational time will need to be accomplished with parallel processing. The only other recourse for reducing the computational time is to reduce the computational time/history, which will occur if progress is made in meeting the challenges that follow below.

3. EXCESSIVE DEMAND ON COMPUTER MEMORY

3.1 Estimate of Memory Demand

Monte Carlo simulation requires substantial memory for tallies, geometry data, and cross section data. Tallies represent the solution of a Monte Carlo simulation and sufficient tallies are needed to yield estimates of within-pin flux/power/isotopic distributions for all the pins in a reactor core, including radial, azimuthal, and axial dependence. Depletion is the key phenomenon that drives the number of tallies. Besides resulting in hundreds of isotopes that must be tracked and accounted for by the Monte Carlo analysis, the accurate estimation of isotopics requires accurate flux and power distributions, thus necessitating a refined tally mesh. The impact of depletion on the number of tallies is magnified greatly if one accounts for the "rim effect" [7] in high burnup UO_2 LWR fuel, which may necessitate tally regions on the order of 100 microns or less. Moreover, advanced fuel design features such as IFBA (integral fuel burnable absorber) coatings may require tally regions on the order of 10 microns or less [8]. The tally estimates reported in this paper do not account for either the rim effect or IFBA coatings.

An estimate of the number of tallies need for full-core reactor analysis was also given by Smith [5]. He assumed a reactor core with 70,000 fuel pins with 10 radial and 100 axial meshes per pin, and 300 isotopes to be tracked. Allowing additional tallies per isotope for both absorption and fission reactions and for the variance estimation, Smith estimated 1 TB of memory would be needed, assuming 8 bytes per tally for over 100 billion tallies. This is well beyond the capacity of current compute nodes, which tend to be in tens of GB or less, even for multi-core nodes. A large number of tallies has implications for computational time as well as memory demand because compute cycles are needed to store and increment these tallies. The number of tallies is dictated by the resolution needed for flux/power/isotopic distributions and is unlikely to be reduced substantially.

Despite those extreme demands on memory due to tallies, there is evidence that progress is being made to overcome this challenge. The MC21 Monte Carlo code [9-11], developed as a joint project by the two Naval Reactors laboratories, Bettis Atomic Power Laboratory and Knolls Atomic Power Laboratory (KAPL), was designed from the outset to analyze full-core naval reactor configurations, including depletion and multiphysics feedback. A discussion of the performance of MC21 on a 3D full-core benchmark problem [10] will be discussed towards the end of this paper, but a brief discussion is given here of the MC21 results relating to the large number of tallies. The benchmark problem had ~ 7 million distinct regions where tallies were to be recorded. There were three tallies per region (energy-integrated flux, absorption rate, and fission rate) to be accumulated for the benchmark problem. The additional processing time to process the tallies for a large run (40 million histories) was less than 10%. When depletion was considered, the number of tallies increased to 100 billion, but the effect on the performance of MC21 was negligible for a relatively small simulation (2.5 million histories). These results, which

are likely to be updated and improved for PHYSOR12 [11] are very encouraging and indicate that it is possible for Monte Carlo developers to adapt their algorithms and methods to ameliorate the memory and computational burden of the huge number of tallies needed for full-core reactor analysis with depletion.

The geometry and cross section data also contribute to the memory demand. The cross section data increases sharply when depletion is included, due to the large number of fission nuclides. Depending on the number of isotopes that are tracked and the temperature range over which cross sections are required, the cross section data by itself may exceed the memory capacity of a single node. When tallies are included, realistic full-core configurations will not be possible on single computational nodes, including multicore nodes with shared memory, for the foreseeable future. Therefore, some type of decomposition will be needed to spread the data across multiple nodes. Two types of decomposition are being pursued – conventional spatial decomposition and data decomposition.

3.2 Spatial Decomposition

The conventional spatial decomposition method assigns spatial regions to specific processor nodes, and neutrons traveling from one spatial region to another must be explicitly communicated from one node to the other. This communication represents parallel overhead and reduces the parallel efficiency. A variation on spatial decomposition is overlapping spatial domains [6,12]. The motivation for overlapping domains was due to analyses with multiple assemblies where it was found that 76% of the neutrons born in an assembly exited the assembly but only 21% exited a domain of width ½ assembly surrounding the birth assembly. Moreover, less than 5% of the neutrons exited if the surrounding domain width was a full assembly. This is illustrated in Figure 1, where neutrons born in the shaded domain stay with its processor until it leaves the overlapping domain shown with dashed lines.

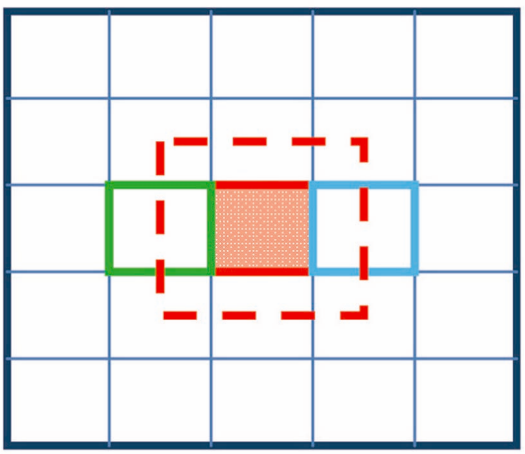

Fig. 1. Overlapping Domain Decomposition

A given processor only emits neutrons in its birth domain but can follow the neutrons into the overlapping domains. The width of the overlapping domain is a user option. Conventional spatial decomposition, including overlapping domains, may be described as static spatial decomposition because the data is assigned to processor nodes at run time.

3.3 Data Decomposition

An alternative to spatial domain decomposition is data decomposition [2]. Here neutrons are emitted in specific processor nodes and remain in that node until the end of the current cycle. As a neutrons moves through the problem geometry, data is served to the processor as needed to allow the random walk to proceed. This might be described dynamic spatial decomposition, because region-specific data is dynamically served to a specific processor node depending on where the neutron travels.

Siegel et al. [13] analyze the communication costs for domain decomposition with an idealized Monte Carlo simulation that represents a full-core reactor configuration with non-overlapping domain decomposition. They concluded that domain decomposition for full-core reactor configurations results in reasonable ratios of communications costs to computational work, hence domain decomposition appears to be a reasonable strategy for addressing the memory demand for full-core Monte Carlo.

4. SLOW CONVERGENCE OF THE FISSION SOURCE

4.1 Power Iteration

The conventional Monte Carlo fission source iteration is a simple power iteration for the fundamental (largest) eigenvalue k_0 and associated eigenvector ψ_0. It is well-known that the dominant error terms for k_0 decay faster than the dominant error terms for ψ_0, especially when the dominance ratio $\rho = k_1/k_0$ is close to 1. For this reason, the fundamental eigenvalue for a high dominance ratio problem will appear to converge much faster than the associated eigenvector, which will be contaminated by the higher harmonics. Since tallies cannot be accumulated until the fission source has converged and the active cycles have begun, slow fission source convergence results in additional time to solution.

The fission matrix approach [14] has been used to efficiently accelerate fission source convergence and a number of papers have been published on this over the years. This has been shown to reduce the number of inactive cycles considerably for large core configurations.

In recent years there has been progress in accelerating fission source convergence by using hybrid methods, where

deterministic transport methods are used to accelerate the Monte Carlo solution. The reported results for fission source convergence with hybrid Monte Carlo are based on using a Shannon entropy diagnostic during the Monte Carlo solution to assess fission source convergence [15]. There are two distinct types of hybrid methods that have been successfully used:

- Using weight windows determined by solving the adjoint transport equation to reduce the variance of the Monte Carlo solution
- Using a low-order operator (e.g., transport-corrected diffusion theory) to accelerate the high-order operator (Monte Carlo)

4.2 Importance Sampling

The first method is similar to conventional variance reduction techniques but which have traditionally been used for source problems, given the conventional wisdom that global eigenvalue problems do not have a preferred location for neutrons to travel to, such as estimating dose at specific points for a source problem. However, Wagner et al. [6,12] have developed the FW-CADIS (forward weighted, consistent adjoint driven importance sampling) method that has shown very promising results for full-core criticality problems. The FW-CADIS methodology has been demonstrated to substantially improve the fission source convergence [12].

4.3 CMFD Acceleration

The second method, the high-order/low-order iteration scheme, which has been very successful with deterministic transport methods [5], employs a low order method that preserves neutron balance over a given "coarse mesh" by expressing the neutron leakage as a modified diffusion-like expression with coefficients determined by the transport (high-order) solution. The following is a notional description of the method. Defining $J_{k+1/2}$ as the normal current at the surface between cells k and k+1, the following discretized leakage equation is assumed:

$$J_{k+1/2} = -D_{k+1/2}(\phi_{k+1} - \phi_k) + \hat{D}_{k+1/2}(\phi_{k+1} + \phi_k) \qquad (1)$$

where $-D_{k+1/2}(\phi_{k+1} - \phi_k)$ is a conventional (normalized) discretization of the neutron current, and $\hat{D}_{k+1/2}(\phi_{k+1} + \phi_k)$ is a "correction term" that preserves $J_{k+1/2}$ by solving for $\hat{D}_{k+1/2}$:

$$\hat{D}_{k+1/2} = \frac{J_{k+1/2} + D_{k+1/2}(\phi_{k+1} - \phi_k)}{\phi_{k+1} + \phi_k} \qquad (2)$$

The iteration strategy consists of a high-order solution to determine $\hat{D}_{k+1/2}$, followed by a low-order solution that includes Eq.(1) and the current value of $\hat{D}_{k+1/2}$. The low-order solution affects the high-order solution by renormalizing the high order solutions with the low-order solutions. This is repeated until convergence is obtained. Known as the coarse-mesh finite difference (CMFD) method from its roots in deterministic transport, this technique has been applied to Monte Carlo simulation by solving an independent low-order "diffusion" equation that includes Eq.(1) to couple the "coarse meshes". The low-order solution is then used to renormalize the fission source, thus nudging the global shape of the high-order solution in the direction of the low-order solution. Results reported by Lee et al. [16] are encouraging.

4.4 Functional Monte Carlo

Alternative approaches for the high-order/low-order iteration approach have been developed by Larsen, Yang, and Wolters, et al. [17-20]. Known as functional Monte Carlo (FMC), the idea is to define an exact low-order functional with coefficients determined by the high-order (Monte Carlo) solution. For example, with FMC, the equivalent equation to Eq.(2) might be:

$$\hat{D}_{k+1/2} = \frac{J_{k+1/2} - F_{k+1/2} + D_{k+1/2}(\phi_{k+1} - \phi_k)}{\phi_{k+1} + \phi_k} \qquad (3)$$

where $F_{k+1/2}$ is identically zero for the transport solution but has terms that tend to reduce the "noise" in the estimation of $J_{k+1/2}$ with Monte Carlo. The iteration strategy for FMC is similar to CMFD in that the Monte Carlo solution estimates the coefficients which are then used to solve the low-order equations. The low order solution is then propagated to the high order equation by renormalization of the fission source. By judicious choice of functionals to minimize the "noise" in the estimation of these coefficients with Monte Carlo, significant improvements in fission source convergence have been observed. Since Eq.(1) is a functional of the transport equation, the CMFD and FMC methods are closely related.

Not only do the CMFD and FMC hybrid methods improve the source convergence, they improve the estimation of the variance and appear to dampen the fluctuations in k_{eff} and flux/power tallies if the acceleration is continued during the active cycles [16,20]. However, there are issues with using acceleration during the active cycles and this is discussed in the next section.

5. APPARENT VERSUS TRUE VARIANCE

The conventional fission source iteration for Monte Carlo criticality calculations is a sequence of "cycles", where the starting locations for the neutrons in cycle n are determined by the fission sites from cycle n-1. This induces an inter-cycle correlation that is stronger for high dominance ratio problems. This correlation does not affect the estimates of k_{eff} or the tallies, but it does affect the estimation of the variance in these quantities. Since the correlation is positive, this will result in an under-

estimate of the variance if the correlation is not accounted for. Ueki [21] introduced the term apparent variance, which is the expected value of the sample variance of the mean, and derived an expression for the bias in the apparent variance, which is simply the difference between the true variance and the apparent variance. This has been a longstanding problem [21-26] for two decades but more attention has been placed on this issue in recent years as numerous groups continue to develop and enhance their Monte Carlo codes for full core analysis. Brown [25,26] reported ratios of real to apparent standard deviations in assembly powers (2D full core PWR) that approached a factor of 5. Shim et al. [24] developed practical methods to estimate the true variance using a single Monte Carlo simulation. Lee et al. [16] reported results consistent with Brown for a 2D full core configuration, with ratios of real to apparent standard deviations in the range of 2-3. They estimated the true variance by internally partitioning the Monte Carlo simulation into 25 independent simulations and performing independent eigenvalue calculations for each of these runs.

Another study that supports the conclusion that the variance is under-estimated due to intercycle correlations are the results with MC21 reported by Kelly et al. [10]. In a thorough study that offers compelling evidence of variance underestimation and its cause, [10] presents results of an exhaustive series of Monte Carlo simulations of a 3D full-core benchmark problem that will be described later in this paper. By examining an ensemble of 120,000 tallies and their associated chi-square distributions for the sample variance estimated with 10 independent runs of 4M histories versus one run with 40M histories, they concluded that the variance estimated for a single Monte Carlo run was underestimated by a factor of .7, which is equivalent to the confidence intervals being too small by about 15%. To address this issue, the "batching" method has been examined with MC21 [11]. This method, suggested two decades ago by Gelbard and Prael [22], combines successive generations into batches and tallies are accumulated for these batches rather than for generations. The idea is that batches of generations will not be as correlated as generations. Preliminary results indicate this method works very well to remove the bias in the apparent variance [11].

The issue of under-predicting the variance takes an interesting twist when hybrid methods are used to accelerate the Monte Carlo simulation. Lee et al. [16], with the CMFD Monte Carlo method, and Wolters, et al. [20], who developed an alternative version of the FMC method of Larsen and Yang [17-19], noticed that the apparent variance was a better estimate of the true variance when CMFD or FMC were used to accelerate the Monte Carlo. In addition, it was found by both groups that the fluctuations in estimates of k_{eff} and flux/power tallies were considerably reduced when CMFD or FMC were employed during the active cycles. It appears that the renormalization of the fission source with the low-order solution helped to "pin" the eigenvector to the fundamental mode and minimize wobbling in the solution due to statistical fluctuations from one cycle to the next. This observation raises a red flag for Monte Carlo specialists because the active cycles are not independent and identically distributed (IID) samples, due to the fact that the low-order solution is modifying the fission source distribution in a different manner for every active cycle. This implies that error estimates based on the central limit theorem (irrespective of intercycle correlations) are not valid. The approach taken by Lee et al. [16] to partition the Monte Carlo simulation into independent runs would address this problem and allow one to assess the bias in the estimate of the variance using FMC or CMFD during the active cycles.

Finally, it should be noted that the implementation of the Wielandt method to the Monte Carlo power iteration method has been shown to reduce the under-prediction bias [27-28] and could be used to estimate more accurate variances.

6. ACCOUNTING FOR MULTIPHYSICS FEEDBACK

Coupling Monte Carlo neutronics to other physics modules such as thermal-hydraulics presents the following issues:

- Fluctuations
- Histogram solutions
- Temperature-dependent neutron cross-sections

Each of these issues is briefly discussed below with suggestions for resolution.

6.1 Fluctuations

Fluctuations are an inherent feature of Monte Carlo and cannot be eliminated. However, the analyst can reduce the fluctuations to arbitrarily small amounts by increasing the number of histories or using variance reduction techniques, or hybrid methods, to reduce the variance. One guideline for statistical resolution is the 1% criteria for local fluxes and powers over regions that are small enough to resolve important phenomena such as isotopic buildup/depletion or Doppler feedback.

An unanswered question is the extent to which statistical error will affect the convergence of the multiphysics feedback iteration. For example, Betzler et al. reports [29] results with temperature feedback (coupled MCNP5 and RELAP5 [30]) for the Fort St. Vrain high temperature gas reactor. For this application, the batch size had to be increased (compared to a case with a specified temperature distribution) by a factor of 5 (from 50,000 to 250,000 histories/batch) to converge the temperature feedback. The impact of statistical error on multiphysics feedback may be stronger with the modern coupling methods such as Jacobian-free Newton-Krylov (JFNK) [31-32] if the residual has to be explicitly estimated with Monte Carlo.

The uncertainty in this quantity may be substantial even if all the tallies are converged to 1%. However, hybrid methods may mitigate this issue because the multiphysics feedback can be applied to the low-order deterministic solution, which can then communicate the feedback to the Monte Carlo solution during the high-order/low-order iterations. Although the low-order solution will see some fluctuations, experience with hybrid methods indicates they are considerably diminished compared to a conventional Monte Carlo solution.

6.2 Histogram Solutions

Coupling discontinuous Monte Carlo results on a tally mesh to a continuum physics model described by a partial differential equation (PDE) may present some difficulties. While interfaces can handle discontinuities, the Monte Carlo discontinuities are an artifact of the fact that conventional tallies are histograms, even in regions that are homogeneous. Recent advances in developing continuous Monte Carlo tallies may address this issue. The functional expansion technique (FET) developed by Griesheimer [33] or the kernel density estimator (KDE) recently applied by Banerjee to reactor analysis [34,35] are two methods that allow continuous tallies. The KDE tally has the advantage that it is mesh-free, so can in principle adapt to any multiphysics mesh.

6.3 Temperature-dependent Neutron Cross-sections

Neutron cross sections are sensitive functions of temperature and need to be available for any temperature encountered throughout the reactor core. Typically, a cross-section processing code such as NJOY [36] will generate a cross section set for a given isotope at a given temperature and a Monte Carlo code will then use this cross section dataset when the region temperature equals the cross section dataset temperature. There are a number of ways to address this challenge:

- In-line cross sections at the correct temperature. Generate cross sections with an "in-line" cross section processing routine (e.g., run NJOY) at the exact temperature of the region. This option would be accurate but is not a serious alternative because of the prohibitive computing time to run NJOY for all the different temperatures and isotopes encountered in the core.
- Pre-generate the cross sections on a fine temperature mesh. Pre-generate cross sections on a relatively fine temperature mesh (e.g., 5 K) to cover the temperature range of interest and use the dataset generated at the closest temperature. Trumbull [37] has shown that cross sections may be required at 5 K or 10 K temperature increments in order to resolve the temperature effects without significant truncation error due to having cross sections at the wrong temperature. Since a temperature feedback calculation for an operating reactor may require temperatures ranging from room temperature to 2000 K or more, this could mean ~200 temperatures for each isotope, which when multiplied by the number of isotopes (~300) will easily overwhelm the memory available to a computational node.
- Pre-generate the cross sections on a coarse temperature mesh. Pre-generate cross sections on a relatively coarse temperature mesh (e.g, 50 K) to cover the temperature range of interest but obtain the actual cross sections at the region temperature by interpolating between the datasets at the temperatures that bound the temperature of interest. Conlin et al. [38] developed a "pseudo-material construct" where the same isotope at two different temperatures could be "mixed" to form an MCNP5 material that had weights as number densities, and the weights were simply the interpolation parameters for a $\sqrt{T}$ interpolation between the bounding dataset temperatures. The advantage of this approach is no changes are required to the MCNP5 code because the input processor handled the actual mixing with the interpolation weights. Since MCNP5 will decide statistically which nuclide to collide with, this method is equivalent to statistical interpolation between the two bounding datasets.
- "On-the-Fly" (OTF) Doppler broadening. Use "On-the-Fly" (OTF) Doppler broadened cross sections that are retrieved during the random walk from a regression model at the exact region temperature and neutron energy. The OTF Doppler broadening method was developed by Yesilyurt et al. [39]. In essence, the Monte Carlo code only needs to store 0 K cross sections for each isotope and the method will broaden the 0 K cross sections for any isotope in the library to any temperature in the range 77 K-3200 K for all incoming neutron energies up to 20 MeV or the unresolved resonance range. The methodology is based on a combination of Taylor series expansions and asymptotic series expansions. The comparison of the broadened cross sections using this methodology with the NJOY cross sections was excellent over the entire temperature (77 K-3200 K) and energy ranges. The computational overhead to utilize OTF cross sections is manageable, since it only involves the evaluation of a 13 term expansion. OTF Doppler broadening has the potential to both remove the cross section data as a significant demand on memory and eliminate temperature resolution of cross section datasets as a source of error for temperature feedback calculations. Effort is underway to incorporate OTF capability into MCNP5 and preliminary results are encouraging [40]. The success of OTF Doppler broadening will allow the analyst to account for temperature changes within a fuel pin to any level desired, although fine resolution may result in a prohibitive material and tally mesh since the material region will need a unique temperature assigned to it and the effect of temperature changes will affect the isotopics.

7. ADAPTING TO FUTURE COMPUTER ARCHITECTURES

Brown [2] recently summarized the status of Monte Carlo for advanced computers and presented a thorough discussion of the prospects for Monte Carlo on advanced computers including cell processors, graphical processing units (GPUs), general purpose GPUs, and greater than massively parallel architectures (e.g., millions of cores) that will be needed to achieve EXAFLOPS performance (10^{18} floating point operations per second). The following is a brief discussion of the issues but [2] should be referred to for a full discussion of the impact of advanced computer architectures on Monte Carlo performance.

To stay on the performance curve predicted by Moore's Law and adhered to by the computer industry over the past three decades, Monte Carlo codes must be adapted to run efficiently on new architectures. To date, Monte Carlo scales well on all production architectures that have been offered, including vector processors, vector-parallel architectures, massively parallel SIMD (single instruction multiple data) architectures, hypercube parallel architectures, Linux clusters, etc. For example, the MCNP5 code is parallelized with threads within a node (OpenMP) and message-passing (MPI) between nodes. However, the fact that Monte Carlo can be adapted to almost any architecture does not mean this will be easy. The physics for Monte Carlo particle transport implies that histories are independent within a fission source cycle or within a time step. Parallelizing across particle histories is natural and allows efficient load balancing (statistical) without a priori knowledge of the solution. However, for vector architectures, the history-based random walk algorithm must be turned inside out because the vectorized algorithm parallelizes across particle events (e.g., tracking, collision analysis) and histories are chopped up into segments corresponding to the events. Radical restructuring of the code data structures and the overall algorithm are required for a vectorized algorithm. The MVP code [41] is the only publicly available vectorized Monte Carlo code for particle transport, which in itself may indicate the difficulty of developing or adapting Monte Carlo codes for a vector architecture.

The high-end computing community is dependent on the commercial development of computers for the game and transactions industries because the market for high performance computing (HPC) is too small to support new computer architectures for the HPC community. Right now the trends seem to be in two directions – GPGPUs and multicore nodes. Both of these architectural trends may allow peak computer performance to continue to increase; however, the practical efficiency of these alternatives (e.g., turnaround time for a full-core Monte Carlo simulation) may be quite different. It is likely that production Monte Carlo codes will have considerable difficulty in adapting to GPGPUs but multicore architectures, while challenging with large numbers of cores, allow a programming model similar to today's computers.

8. MEASURING PROGRESS TOWARDS ACHIEVING FULL-CORE MONTE CARLO

In order to assess progress towards achieving full-core Monte Carlo, a benchmark problem was developed by Hoogenboom et al. [42] to measure Monte Carlo performance for a realistic 3D reactor configuration. The specifications for the benchmark problem are available at the Nuclear Energy Agency (NEA) website. The NEA website allows the analyst to upload their results for the benchmark problem and this will allow for convenient intercomparison of results. The NEA Monte Carlo benchmark problem is a revision of the problem originally proposed by Hoogenboom and Martin at the 2009 M&C Conference in Saratoga Springs [43], and this was based on talks by Smith at the 2003 Gatlinburg M&C Conference [44] and Martin at the 2007 Monterey M&C Conference [45]. The benchmark configuration is representative of a 3D PWR with explicit fuel pins and a mixture of fuel and burnable poison pins. The key requirement of the benchmark exercise is the prediction of local powers with 1% relative errors. Depletion is not included in this version of the NEA benchmark problem.

Hoogenboom et al. [46] presented a detailed analysis of the results reported by Kelly et al. [10] at PHYSOR10 and the results reported by Leppänen at SNA+MC2010 in Tokyo [47]. Both efforts are approaching the criterion of 1% relative errors on all local power densities but the run times are well in excess of the 1 hour criterion suggested by Smith [44]. Brief summaries are given below; the reader should consult the original reports for details:

- Kelly et al. [10] reported results which indicated that 95% of the tallies had standard deviations less than 3% for 10 billion histories. The wall clock time was 18 hours on 400 cores.
- Leppänen [47] reported that 90% of the local powers had standard deviations less than 2% for 100 billion histories. The wall clock time was 21 hours on 7 CPUs.

These results are encouraging and indicate that it may not be long before a Monte Carlo code simulates the NEA benchmark problem on a cluster and achieves 1% statistics within 1 hour wall clock time. In fact, Sutton [11] hints that MC21 will report substantially improved results at PHYSOR12.

9. SUMMARY AND CONCLUSIONS

The challenges and prospects for full-core Monte Carlo are summarized below.

- Prohibitive computational time for acceptable statistics – this will be met by faster and cheaper multicore nodes,

algorithmic advances with hybrid Monte Carlo methods, and innovative variance reduction methods. The fact there is a benchmark problem specifically focused on Monte Carlo performance will encourage progress in this area.

- Excessive demand on computer memory – this will be addressed by innovative decomposition schemes and the increasing memory capacity of multicore nodes and decreasing cost of memory.
- Slow convergence of the fission source – this has been successfully addressed with CMFD and FMC methods. However, improving the fission source convergence simply moves the problem to the active cycles, since there have to be enough to obtain satisfactory statistics (eg., 1% on local powers).
- Apparent versus true variance – multiple realizations provide assurance that the estimated variance is correct but more analysis is needed to allow valid variance estimation with a single Monte Carlo simulation when CMFD/FMC is turned on during the active cycles to pin down the fission source. Recent results with MC21 [11] indicate that the batching method of Prael and Gelbard [22] may remedy this problem.
- Accounting for multiphysics feedback – more analysis and studies are needed to understand the impact of statistical errors on convergence with multiphysics feedback. This area is just beginning to be explored. The use of OTF Doppler broadening is a potential game changer for cross section generation for Monte Carlo codes.
- Adapting to future computer architectures – this is perhaps the most uncertain challenge. The direction that computer architectures take is dependent on where the gaming and transaction industries go. In general, a trend to larger multicore nodes would be acceptable but a trend to GPUs would be challenging for production Monte Carlo codes.
- Progress towards full-core Monte Carlo – the results that have already been reported [10,47] are very promising and there is reason to believe that results reported at PHYSOR12 will continue in the direction of achieving the 1% relative standard error within 1 hour wall clock time. It appears that the 1% error criterion will soon be satisfied since both [10] and [47] could achieve this without prohibitive increase in the number of histories. However, the one hour wall clock criterion will be harder to meet and may require a cluster with thousands of cores. Whether this meets the spirit of the original "Kord Smith Challenge" will be a topic of lively discussion. It may be useful to consider extensions or new benchmark problems to account for depletion, to assess source convergence, and to estimate the true variance.

ACKNOWLEDGEMENTS

This review paper is focused on the challenges and prospects for achieving full-core Monte Carlo and is not intended to be an exhaustive survey of Monte Carlo methods development. It is based on an invited talk presented at the workshop on "Current Status and Future of the Nuclear Data, Reactor Physics and Computational Science" that was embedded in the 2011 Autumn Meeting of the Korean Nuclear Society October 26-29, 2011. Special thanks are given to the workshop chairman Dr. Jae Man Noh for the invitation to present the lecture and to Professor Han Gyu Joo for his hospitality and assistance during the trip to Korea.

REFERENCES

[1] X-5 Monte Carlo Team, "MCNP – A General N-Particle Transport Code, Version 5 – Volume I: Overview and Theory," LA-UR-03-1987, Los Alamos National Laboratory (April, 2003).

[2] F. B. Brown, "Recent Advances and Future Prospects for Monte Carlo," *Proc. Joint International Conference on Supercomputing in Nuclear Applications and Monte Carlo 2010 (SNA + MC2010)*, Tokyo, Japan, October 17-21, 2010.

[3] F. B. Brown "Monte Carlo Methods in Reactor Physics: Current Status & Future Prospects," presented at the M&C Division Computational Roundtable, American Nuclear Society Summer Meeting, Hollywood, FL, June 27, 2011.

[4] D. Griesheimer "Monte Carlo Methods in Reactor Physics: Current Status and Future Prospects - In-Line Feedback Effects," presented at the M&C Division Computational Roundtable, American Nuclear Society Summer Meeting, Hollywood, FL, June 27, 2011.

[5] K. Smith "Monte Carlo for Practical LWR Analysis: what's needed to get to the goal?, " presented at the M&C Division Computational Roundtable, American Nuclear Society Summer Meeting, Hollywood, FL, June 27, 2011.

[6] J. Wagner "Hybrid and Parallel Domain-Decomposition Methods Development to Enable Monte Carlo for Reactor Analyses," presented at the M&C Division Computational Roundtable, American Nuclear Society Summer Meeting, Hollywood, FL, June 27, 2011.

[7] Hj. Matzke, "On the rim effect in high burnup UO_2 LWR fuels," *J. Nucl. Materials*, **189**, 141-148 (1992).

[8] M. Rosa, J. S. Warsa, J. H. Chang, and R. S. Baker, "Discrete Ordinate Calculation of the k-Eigenvalue of an IFBA Pin Using Unstructured Meshes in 2D," to be presented at the Summer Meeting of the American Nuclear Society, Chicago, June 2012.

[9] T. M. Sutton, T. J. Donovan, T. H. Trumbull, P. S. Dobreff, E. Caro, D. P. Griesheimer, L. J. Tyburski, D. C. Carpenter, and H. Joo, "The MC21 Monte Carlo Transport Code," *Proc. ANS Mathematics & Computation Division Topical Meeting, M&C2007*, Monterey, CA, USA, April 2007.

[10] D. J. Kelly, T. M. Sutton, T. H. Trumbull, and P. S. Dobreff, "MC21 Monte Carlo Analysis of the Hoogenboom-Martin Full-Core PWR Benchmark Problem," *Proc. ANS Reactor Physics Division Topical Meeting*, PHYSOR10, Pittsburgh, May 2010.

[11] T. Sutton, "Progress in Monte Carlo for Reactor Design and Analysis," presentation at the University of Michigan, October 6, 2011.

[12] J. C. Wagner, D. E. Peplow, S. W. Mosher, and T. M. Evans,

"Review of Hybrid (Deterministic/Monte Carlo) Radiation Transport Methods, Codes, and Applications at Oak Ridge National Laboratory," *Proc. Joint International Conference on Supercomputing in Nuclear Applications and Monte Carlo 2010 (SNA + MC2010)*, Tokyo, Japan, October 17-21, 2010.

[13] A. Siegel, K. Smith, P. Fischer, and V. Mahadevan, "Analysis of Communication Costs for Domain Decomposed Monte Carlo Methods in Nuclear Reactor Analysis," accepted for publication, J. Comp. Phys. (December 2011), http://dx. doi. org/10. 1016/j. jcp. 2011. 12. 014.

[14] T. Kitada and T. Takeda, "Effective Convergence of Fission Source Distribution in Monte Carlo Simulation," *J. Nucl. Sci. Technol*., **38**, 324-329 (2001).

[15] T. Ueki and F. B. Brown, "Stationarity Modeling and Informatics-Based Diagnostics in Monte Carlo Criticality Calculations," *Nucl. Sci. Eng*., **149**, 38-50 (2005).

[16] M. -J. Lee, H. G. Joo, D. Lee, and K. Smith, "Multigroup Monte Carlo Reactor Calculation with Coarse Mesh Finite Difference Formulation for Real Variance Reduction," *Proc. Joint International Conference on Supercomputing in Nuclear Applications and Monte Carlo 2010 (SNA + MC2010)*, Tokyo, Japan, October 17-21, 2010.

[17] E. W. Larsen and J. Yang "A Functional Monte Carlo Method for k-Eigenvalue Problems," *Nucl. Sci. Eng*., **159**, 107–126 (2008).

[18] J. Yang and E. W. Larsen, "Application of the 'Functional Monte Carlo' Method to Estimate Continuous Energy k-Eigenvalues and Eigenfunctions," *Proc. ANS Mathematics & Computation Division Topical Meeting, M&C2009*, Saratoga Springs, New York, USA, 2009.

[19] J. Yang and E. W. Larsen, "Calculation of k-Eigenvalues and Multi-Group Eigenfunctions Using the Hybrid 'Functional Monte Carlo' Method," *Proc. ANS Reactor Physics Division Topical Meeting, PHYSOR10*, Pittsburgh, May 2010.

[20] E. R. Wolters, E. W. Larsen, and W. R. Martin, "Generalized Hybrid Monte Carlo – CMFD Methods for Fission Source Convergence," *Proc. ANS Mathematics & Computation Division Topical Meeting, M&C2011*, Rio de Janeiro, Brazil, May 8-12, 2011.

[21] T. Ueki, T. Mori, and M. Nakagawa, "Error Estimation and their Biases in Monte Carlo Eigenvalue Calculations," *Nucl. Sci. Eng*., **125**, 1-11 (1997).

[22] E. M. Gelbard and R. Prael, "Computation of Standard Deviations in Eigenvalue Calculations," *Prog. Nucl. Energy*, **24**, 237 (1990).

[23] T. Ueki, "Intergenerational Correlation in Monte Carlo k-Eigenvalue Calculation," *Nucl. Sci. Eng*., **141**, 101–110 (2002).

[24] H. J. Shim, Y. Kim, and C. H. Kim, History-Based Batch Method for a Real Variance Estimation in Monte Carlo Eigenvalue Calculations, *Trans. Am. Nucl. Soc*., **100**, 300 (2009).

[25] F. B. Brown, "A Review of Monte Carlo Criticality Calculations – Convergence, Bias, Statistics," *Proc. ANS Mathematics & Computation Division Topical Meeting, M&C2009*, Saratoga Springs, NY, May 3-7, 2009.

[26] F. B. Brown, "K-Effective of the World and other Concerns for Monte Carlo Eigenvalue Problems," *Proc. Joint International Conference on Supercomputing in Nuclear Applications and Monte Carlo 2010 (SNA + MC2010)*, Tokyo, Japan, May 3-7, 2009.

[27] B. C. Kiedrowski and F. B. Brown, "Using Wielandt's method to eliminate confidence interval under prediction bias in MCNP5 criticality calculations," *Trans. Am. Nucl. Soc*., **99**, 338-340 (2008).

[28] H. J. Shim and C. H. Kim, "Tally Efficiency Analysis for Monte Carlo Wielandt Method," *Ann. Nucl. Eng*., **36**, 1694-1701 (2009).

[29] B. R. Betzler, E. E. Sunny, J. C. Lee, and W. R. Martin, "Coupled Nuclear-Thermal- Hydraulic Calculations for Fort St. Vrain Reactor," *Proc. 14th International Topical Meeting on Nuclear Reactor Thermal-Hydraulics (NURETH-14)*, Toronto, Canada, September 25-29, 2011.

[30] RELAP5-3D Code Development Team, "ATHENA Code Manual," INEEL-EXT-98-00834, Rev. 2. 2, Idaho National Engineering and Environmental Laboratory (2003).

[31] D. A. Knoll and D. E. Keyes, "Jacobian-free Newton-Krylov methods: a survey of approaches and applications," *J. Comp. Phys*., **193**, pp. 357-397 (2004).

[32] H. Park, D. A. Knoll, D. R. Gaston, and R. C. Martineau, "Tightly Coupled Multiphysics Algorithms for Pebble Bed Reactors," *Nucl. Sci. Eng*., **166**, 118–133 (2010).

[33] D. P. Griesheimer, W. R. Martin, and J. P. Holloway, "Convergence Properties of Monte Carlo Functional Expansion Tallies," *J. Comp. Phys*., **211**, 129-153 (January 2006).

[34] K. Banerjee and W. R. Martin, "Kernel Density Estimated Monte Carlo Global Flux Tallies," *Proc. ANS Mathematics & Computation Division Topical Meeting, M&C2009*, Saratoga Springs, NY, May 3-7, 2009.

[35] K. Banerjee and W. R. Martin, "Kernel Density Estimation Method for Monte Carlo Global Flux Tallies," accepted for publication, *Nucl. Sci. Eng*. (2011).

[36] R. E. MacFarlane and D. W. Muir, "NJOY99.0 Code System for Producing Pointwise and Multigroup Neutron and Photon Cross Sections from ENDF/B Data," PSR-480/NJOY99.00, Los Alamos National Laboratory, Los Alamos (1999).

[37] T. H. Trumbull, "Treatment of Nuclear Data for Transport Problems Containing Detailed Temperature Distributions, " *Nucl. Technol*., **156**, 75-86 (2006).

[38] J. L. Conlin, W. Ji, J. C. Lee, and W. R. Martin, "Pseudo Material Construct for Coupled Neutronic-Thermal-Hydraulic Analysis of VHTGR," *Trans. Am. Nucl. Soc*., **92**, 225-227, San Diego, CA (June, 2005).

[39] G. Yesilyurt, W. R. Martin, and F. B. Brown, "On-the-Fly Doppler Broadening for Monte Carlo Codes," accepted for publication, *Nucl. Sci. Eng*. (2011).

[40] F. B. Brown, W. R. Martin, G. Yesilyurt, and S. Wilderman, "Progress with On-The-Fly Neutron Doppler Broadening in MCNP," to be presented at the Summer Meeting of the American Nuclear Society, Chicago, June 2012. Also LA-UR-12-00423, Los Alamos National Laboratory (2012).

[41] Y. Nagaya, K. Okumura, T. Mori, and M. Nakagawa, "MVP/GMVP II: General Purpose Monte Carlo Codes for Neutron and Photon Transport Calculations based on Continuous Energy and Multigroup Methods, " JAERI 1348 (2004).

[42] J. E. Hoogenboom, W. R. Martin, and B. Petrovic, "Monte Carlo Performance Benchmark for Detailed Power Density Calculation in a Full Size Reactor Core," Benchmark Specifications Revision 1. 1, http://www. nea. fr/dbprog/

MonteCarloPerformanceBenchmark. htm (2010).
[43] J. E. Hoogenboom and W. R. Martin, "A Proposal for a Benchmark to Monitor the Performance of Detailed Monte Carlo Calculation of Power Densities in a Full Size Reactor Core, " *Proc. ANS Mathematics & Computation Division Topical Meeting, M&C2009*, Saratoga Springs, NY, May 3-7, 2009.
[44] K. Smith, "Reactor Core Methods," Invited lecture, ANS Mathematics & Computation Division Topical Meeting, M&C2003, Gatlinburg, TN, USA, April 2003. URL: http: //www.nea.fr/html/dbprog/documents/ MC03Smith.pdf.
[45] W. R. Martin "Advances in Monte Carlo Methods for Global Reactor Analysis," Invited lecture, ANS Mathematics & Computation Division Topical Meeting, M&C2007, Monterey, CA, USA, April 2007. URL: http://www.nea.fr/ html/dbprog/documents/M&C07Martin.pdf.
[46] J. E. Hoogenboom, W. R. Martin, and B. Petrovic, "The Monte Carlo Performance Benchmark Test – Aims, Specifications, and First Results," *Proc. ANS Mathematics & Computation Division Topical Meeting, M&C2011*, Rio de Janeiro, Brazil, May 8-12, 2011.
[47] J. Leppänen "Use of the Serpent Monte Carlo Reactor Physics Code for Full-Core Calculations" *Proc. Joint International Conference on Supercomputing in Nuclear Applications and Monte Carlo 2010 (SNA + MC2010)*, Tokyo, Japan, October 17-21, 2010.