*International Conference on Mathematics and Computational Methods Applied to Nuclear Science and Engineering (M&C 2011)*
Rio de Janeiro, RJ, Brazil, May 8-12, 2011, on CD-ROM, Latin American Section (LAS) / American Nuclear Society (ANS)
ISBN 978-85-63688-00-2

# THE MONTE CARLO PERFORMANCE BENCHMARK TEST - AIMS, SPECIFICATIONS AND FIRST RESULTS

**J. Eduard Hoogenboom**
Faculty of Applied Sciences
Delft University of Technology
Mekelweg 15, 2629 JB Delft, The Netherlands
j.e.hoogenboom@tudelft.nl

**William R. Martin**
Nuclear Engineering and Radiological Sciences
University of Michigan
Ann Arbor, MI 48109-2104, USA
wrm@umich.edu

**Bojan Petrovic**
Nuclear and Radiological Engineering
Georgia Institute of Technology
Atlanta, GA 30332-0745, USA
Bojan.Petrovic@gatech.edu

## ABSTRACT

The Monte Carlo performance benchmark for detailed power density calculation in a full-size reactor core is organized under the auspices of the OECD NEA Data Bank. It aims at monitoring over a range of years the increase in performance, measured in terms of standard deviation and computer time, of Monte Carlo calculation of the power density in small volumes. A short description of the reactor geometry and composition is discussed. One of the unique features of the benchmark exercise is the possibility to upload results from participants at a website of the NEA Data Bank which enables online analysis of results and to graphically display how near we are at the goal of doing a detailed power distribution calculation with acceptable statistical uncertainty in an acceptable computing time.
First results are discussed which show that 10 to 100 billion histories must be simulated to reach a standard deviation of a few percent in the estimated power of most of the requested the fuel zones. Even when using a large supercomputer, a considerable speedup is still needed to reach the target of 1 hour computer time.
An outlook is given of what to expect from this benchmark exercise over the years. Possible extensions of the benchmark for specific issues relevant in current Monte Carlo calculation for nuclear reactors are also discussed.

*Key Words*: Monte Carlo, benchmark, power density, reactor, parallel computing.

## 1. INTRODUCTION

Notwithstanding the enormous capabilities of present-day Monte Carlo codes for neutron transport in reactor cores allowing all geometrical details of each individual fuel pin of each fuel

assembly to be modeled, it is not realistic to calculate local power densities (especially reaction rates for all isotopes of interest) for a large number of small regions of a fuel pin with acceptable statistical uncertainty within an acceptable computing time. Nonetheless, this is the requirement for current design calculations of a nuclear reactor core. A review of the literature over the last decade indicates it will take many years from now before we are at this desired stage of Monte Carlo computation.

Kord Smith [1] formulated the challenge for future Monte Carlo simulation as the calculation of the local power for each of the fuel pins in a fuel assembly when subdivided into 100 axial and 10 radial zones. The number of fuel pins in a typical fuel assembly of a PWR core is between 200 and 300 while the number of fuel assemblies in a reactor core is around 200. This results in the total of perhaps 40-60 million tallies. For an acceptable result, Smith specified that the standard deviation in each local power region should be 1 % or less. He estimated on the basis of Moore's law that it will be 2030 before such a full core Monte Carlo calculation could be done in less than one hour on a single workstation.

Bill Martin [2] analyzed the situation in some detail based on a more recent actual full core calculation. Assuming that Moore's law manifests itself as only more cores in a desktop computer, Martin estimated that it would be 2019 before a full reactor core calculation with 40,000 fuel pins and 100 axial regions (no radial subdivision) and 1 % statistical uncertainty for local power estimates could be accomplished in one hour calculation time. In this case, the desktop computer would have a 1500-core processor at that time assuming the computer technology was moving in the direction of multicore processors.

As there are several reasons why the advances in performance of full core Monte Carlo reactor calculations may be better or worse than estimated above it will be useful to monitor the performance of Monte Carlo full core calculations over the next 2 decade. To accomplish that in a systematic way a benchmark test is defined with sufficient detail for a full core realistic Monte Carlo calculation.

## 2. AIMS OF THE BENCHMARK EXCERCISE

The principal aim of the benchmark exercise is to monitor over the coming years the increase in performance of Monte Carlo calculations of a full-size power reactor core. To address the above mentioned issues, the key quantity to be estimated is the local power density in small fuel regions and especially its standard deviation in relation to the number of neutron histories and computing time. In this respect the total number of tallies, i.e. the total number of fuel regions for which the power density is calculated, has a strong effect on the overall execution time.

In addition to the number of neutron histories simulated and the resulting standard deviation in the power density at specific fuel volumes, the computing time is an important measure in this benchmark. Large scale Monte Carlo calculations will be done on computer systems with many nodes and/or a number of cores per node. For extrapolation to a very large number of cores in the future, the speedup as a function of the number of cores (i.e., parallel performance) is also very important.

To keep the benchmark exercise as clean as possible other issues as fission source convergence are not part of the benchmark at this stage, but may be addressed in future extended versions. Therefore, we now look only at active cycles for which source convergence is already achieved, no matter what it takes to reach that state. Moreover, we will not consider reactivity feedbacks (e.g., temperature dependence) or evolution of burnup. There will be, of course, differences in outcome of the local power densities due to differences in nuclear data, but such differences are not the subject of this benchmark, rather it is the efficiency of the calculation that is being measured.

A secondary aim of the benchmark test is to stimulate improvements in Monte Carlo codes and their implementation. This not only refers to improving the efficiency of generating random walks of neutrons in a full core nuclear reactor, but also to improving methods for dealing with large numbers of tallies and various issues in high performance computing like efficiency improvements in parallel computing using current and future computer architectures that will generally become available to researchers and engineers. Another issue is the quality of the algorithm and the programming for high performance computing, as well as the performance of the compiler. All these items are implicitly part of the benchmark test.

## 3. BENCHMARK SPECIFICATIONS

In this section only the main lines of the benchmark specifications are discussed. For complete specifications see Ref. 3.

The geometry and composition of the reactor core is based on a large PWR core without trying to model any specific existing reactor. As the main goal of the benchmark exercise is to estimate the standard deviation in the power density in a subvolume of a fuel pin, the details of the geometry and composition of the reactor core and fuel assemblies are not critical, as long as they require the Monte Carlo code to do realistic neutron history simulation with most of the features encountered in current reactor design. Therefore, a number of simplifications are made in geometry and composition of materials.

### 3.1. Core Configuration

The core consists of 241 identical fuel assemblies with arrangement shown in Fig. 1. The dimensions of a fuel assembly are 21.42 x 21.42 $cm^2$. The baffle plates and core barrel have been homogenized into a radial reflector region. The downcomer with “cold” water (corresponding to inlet water temperature) forms a cylindrical shell with inner radius 209 cm and outer radius 229 cm. The reactor vessel has an outer radius of 249 cm.

The active fuel length is 366 cm. To introduce an asymmetry in the vertical direction and to model in an approximate way the decreasing coolant density from bottom to top the water density in a fuel assembly and in the radial reflector is lower above the core midplane than below the core midplane. Three homogenized regions below the fuel of in total 46 cm in height include the bottom fuel assembly region, nozzle region and core plate region. Likewise there are three homogenized regions above the fuel of in total 40 cm.

Although the defined core lay-out shows octant symmetry in the *x-y* plane, it is the explicit requirement to model the full core and not take advantage of any symmetry in the geometry, mainly to deal with the large number of tally volumes for the power density and the standard deviation in all estimates. At this time, the apparent standard deviation is reported but it is understood that this may under-estimate the real standard deviation according to recent papers on this subject. [4, 5]

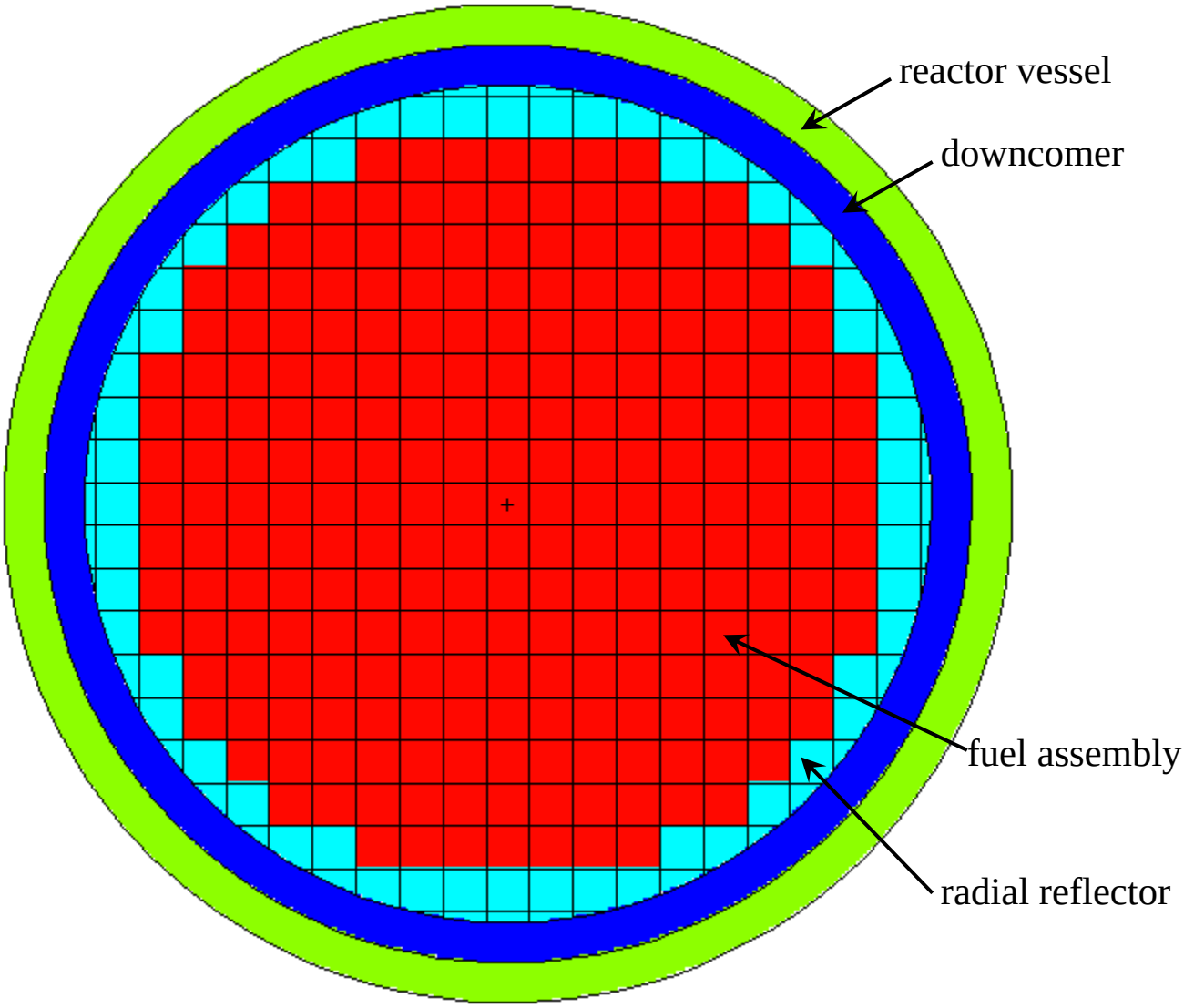


**Figure 1. Horizontal cross section of reactor core with fuel assemblies.**

### 3.2. Fuel Assembly Configuration

Each fuel assembly consists of 17 x 17 unit cells as shown in Fig. 2. The dimension of the unit cell is 1.26 x 1.26 $cm^2$. No spacers or other construction material for an assembly are modeled. No intra-assembly gap is assumed.

Out of 289 lattice locations, 24 symmetrically positioned unit cells are occupied by control rod instrumentation guide tubes, and the central location is filled with an instrumentation tube. No inserted control rods are modeled; thus, the guide tubes are filled with water. At the other unit cell positions fuel pins with cladding are present. Each fuel pin has an outer radius of 0.41 cm. The cladding has an outer radius of 0.475 cm. No gap is modeled.

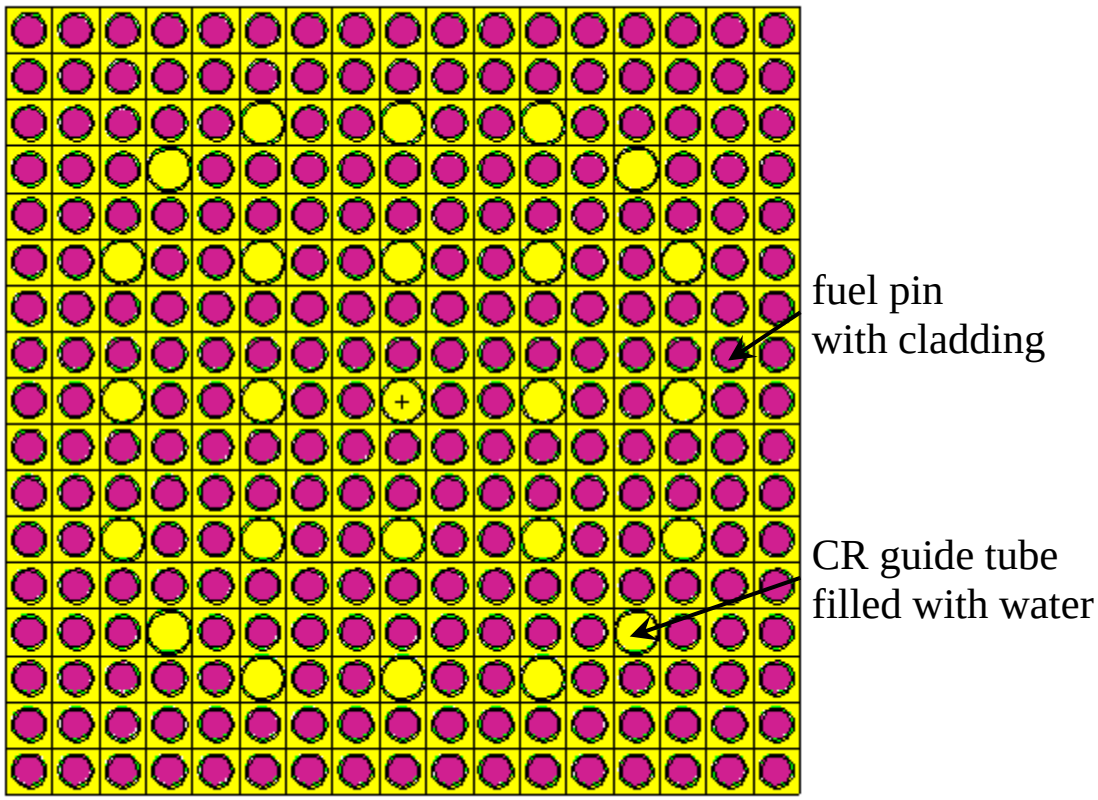


**Figure 2. Horizontal cross section of a fuel assembly.**

### 3.3. Material Composition

The effort needed to sample the nucleus a neutron collides with depends on the number of nuclides in a medium. In order to take into account a reasonable number of nuclides in the fuel, the fuel composition is derived from a certain burnup stage (roughly 24,000 MWd/ton) from which 17 actinides are selected together with 16 fission products, and oxygen.

For convenience the cladding material is taken as (natural) zirconium with a reduced density to account for smearing the gap between fuel and cladding. The coolant is water and is present at different densities below and above the axial core midplane to represent “cold” and “hot” water. Boron is added to the coolant and its concentration is chosen such that the reactor is near critical.

The pressure vessel material is modeled as low-carbon steel with iron substituted for components with low weight fraction.

### 3.4. Temperatures

The different temperatures in an operating reactor generally complicate the preparation of the cross sections. As the aim of this benchmark is not to calculate physical quantities of an operating reactor it was decided to take all cross sections in all materials at room temperature. This will not affect the table lookup time of the cross sections and is supposed to have only a small effect on the simulation time of the neutron histories.

### 3.5. Cross sections

As in most current general-purpose Monte Carlo codes the energy dependence of the cross sections is treated in continuous energy (pointwise) mode, this is the preferred option for this benchmark. There is no specification of the underlying evaluated nuclear data file. Therefore, cross sections normally available with the Monte Carlo code can be used. As it will have an effect on the processing time of neutron histories, the thermal scattering in water should be treated by the $S(\alpha,\beta)$ thermal scattering law.

### 3.6. Fission Energy

The actual calculation of the local fission power is in fact a complicated process as it is composed of different contributions. Due to the contribution of prompt and delayed photons, amongst others, part of the fission energy will not be deposited at the site where the fission event took place. In the current benchmark we will consider neutron-only calculations assuming that all recoverable fission energy is deposited locally.

### 3.7. Expected Resuls

The aim of the benchmark test is to see whether and at what computational cost a detailed calculation of local power densities can be made. The basic volume for the calculation of the power density is an axial region of a fuel pin of 1/100th of the fuel length, which is 3.66 cm. Hence, the expected results are the power densities in all axial regions of all fuel pins and their standard deviation. As this constitutes a huge number of data, we selected a small number of regions to actually compare results. However, as the calculation of power densities in all regions seriously influences the calculation time, the aim is to include the tallies for the power density of all axial regions of all fuel pins. As there are 17 x 17 -25 = 264 fuel pins in an assembly and 241 assemblies, there are 264 x 241 x 100 = 6,362,400 power densities to be calculated.

Apart from general information about the participant, the Monte Carlo code and computer system used, the following results should be reported.

- The nominal number of neutron histories per cycle
- The number of inactive cycles to arrive at a converged source distribution
- The number of active cycles (batches)

- The execution time for the active cycles
- $k_{eff}$ and its standard deviation (absolute value)
- The produced energy and their standard deviation in the fuel regions specified in Table I. In this table the position of a fuel assembly in the core and of a fuel pin in an assembly is indicated by their position relative to the central fuel assembly or fuel pin, which is identified as (0,0) as shown in Fig. 3 for the fuel assemblies. The axial fuel regions are numbered from 1 at the bottom of the fuel pin to 100 at the top.
- Any other information relevant for the calculation; for instance modification to the original Monte Carlo code for this benchmark; special acceleration techniques or variance reduction techniques applied, etc.

**Table I. Expected energy production for selected fuel regions**

| Result No | fuel assembly | fuel pin | axial region | comment |
|---|---|---|---|---|
| 0 | all,integrated | all, integrated | all, integrated | whole core |
| 1 | (0, 0) | all, integrated | all, integrated | central FA |
| 2 | (3, 2) | all, integrated | all, integrated | another selected FA |
| 3 | (-3, 2) | all, integrated | all, integrated | another selected FA |
| 4 | (-3, -2) | all, integrated | all, integrated | another selected FA |
| 5 | (3, -2) | all, integrated | all, integrated | another selected FA |
| 6 | (-3, -8) | all, integrated | all, integrated | corner FA |
| 7 | (-6, 6) | all, integrated | all, integrated | corner FA |
| 8 | (6, 6) | all, integrated | all, integrated | corner FA |
| 9 | (3, 2) | (-8, -8) | all, integrated | pin at corner of FA |
| 10 | (3,2) | (2, 1) | all, integrated | another pin |
| 11 | (3,2) | (2,1) | 1 | region at core bottom |
| 12 | (3,2) | (2,1) | 50 | region at core midplane |
| 13 | (3,2) | (2,1) | 100 | region at core top |
| 14 | (6,6) | (-8,-8) | 100 | region at core top, another FA |

## 4. ORGANIZATIONAL ASPECTS

The current benchmark features a number of unique aspects. There is no restriction to participants other than registration at the NEA Data Bank for uploading results. Unlike most current and previous benchmark exercises there is not a final date for reporting results nor will

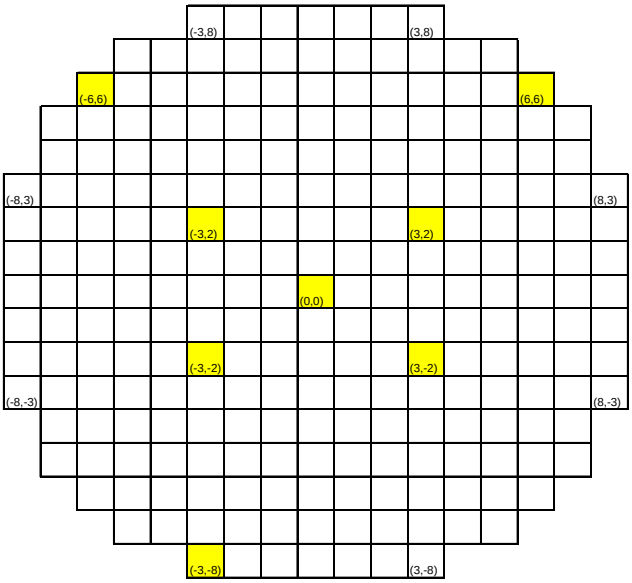


**Figure 3. Indication of fuel assemblies in the core. (assemblies from Table I highlighted in yellow)**

there be published a “final” report. The time span envisaged is much longer than normal, at least 10 years and possibly 20 years. Participants are encouraged to send in in the course of time results obtained from different Monte Carlo codes or code versions and/or obtained with different computer systems.

Probably the most unique aspect is that participants can upload their results at a specific webpage of the NEA Data Bank website (http://www.nea.fr/confdb/confdb/conf?id=55 ), which opens the possibility for on-line evaluation of results. In the future it may be possible to display results of all participants automatically in a graph and/or extrapolate results from various participants to predict when we will be able to calculate the detailed power distribution in a reactor with 1 % statistical uncertainty within one hour computation time, for instance based on Moore’s and Amdahl’s laws. This can be extrapolated for various types of computer systems (desk top PC, Linux cluster, supercomputer, etc.) using measured and/or projected parallel efficiencies.

Participants are also encouraged to upload their input file for a specific Monte Carlo code. This may facilitate the analysis of results, but more importantly, input files will become available to other registered participants who may benefit if they can start their contribution to the benchmark exercise from an existing input file for the Monte Carlo code they intend to use.

There are no meetings scheduled for participants to discuss problems and solutions, but that may be done in connection with international conferences on nuclear reactor analysis like M&C, PHYSOR and SNA+MC, which are organized every two or three years. Periodic overviews of results might be given at one or more of the above mentioned conference series.

## 5. POSSIBLE EXTENSIONS OF THE BENCHMARK

It is important not to change the specifications for a benchmark unless it is absolutely necessary because of serious shortcomings and at the moment there are no indications for that. However, there may be a need for researchers to use the basic geometry of the current benchmark to test other aspects of current or future Monte Carlo reactor calculations. For example, this might include consideration of:

- fission source distribution convergence
- determination of dominance ratio
- true variance estimation
- power distribution estimation for different rings in each fuel rod (see the "Kord Smith challenge")
- actual energy deposition in fuel and non-fuel regions including contribution from photons

Moreover, there may be a need to move to more realistic reactor models to validate design calculations. This may include

- different material composition per fuel assembly or groups of fuel assemblies
- thermal-hydraulic feedback leading to different temperatures (and thus cross section assignment) for all axial zones and for different fuel assemblies (or even fuel rods)
- burnup (leading to non-uniform fuel compositions in fuel assemblies, possibly including axial and radial zones in fuel pins and burnable poison pins)
- control rods

Although the present benchmark exercise is focused on the statistical error in the power density in a small fuel zone of a reactor in connection with the computer time necessary for such a calculation, we may define the specifications and quantities to be calculated for a number of the addressed problems if there appears to be a need for that.

## 6. ANALYSIS OF FIRST RESULTS

Although the current benchmark exercise was launched officially by the NEA Data Bank only in October 2010, several authors have already used this benchmark to produce results and report their experiences at international conferences. Together with the experience and results obtained by the present authors we give an analysis of first results.

Kelly et al. [6] made extensive investigations using the MC21 code. It should be noted that they used an earlier version of the benchmarks specifications, i.e., the first, preliminary proposal for the performance benchmark [7], which somewhat differs from the current, finalized specifications [3]. Therefore, their analysis' results of some specific reactor physics values, such as the power distribution or dominance ratio of this problem may not be fully applicable for the current benchmark. However, their results and analysis of the computational performance, including the achieved number of neutron histories per CPU-second, parallel performance and standard deviation for the power density in various zones of the fuel pins should be only minimally impacted by the change of the problem and are expected to remain relevant for the current benchmark. A brief summary of more prominent results follows, with details provided in the quoted work [6]:

(a) The authors examined the impact of tallies on the total CPU time. A detailed spatial tallying was employed with close to 7 million (6,964,900) fuel regions and ~7.2 million total regions. Combined with tallying of 3 physical quantities (flux, absorption rate, fission rate) it amounted to almost 22 million tallies. Nevertheless, the related overhead was very limited and fully acceptable, less than 10% for long reference simulations, and increasing to 45% for relatively short simulations with limited number of neutron histories.
(b) Simulations were performed on a single CPU as well as on a cluster with up to 400 cores. The parallel performance seems to scale well. Keeping the processor load of the same order of magnitude, the number of neutron histories per CPU per minute was reduced from ~150,000 per minute on a single CPU, to ~110,000 per minute on a 400-core cluster. This is a non-negligible reduction, which however still reflects quite acceptable performance and would correspond to a parallel fraction of 0.999 if fitted by Amdahl's law.
(c) The longest, reference simulation was performed with 10 million histories per cycle, for 250 inactive plus 1,000 active cycles (i.e., 10 billion histories in all active cycles), and took about 18 hours on the 400-core cluster.
(d) The estimated 1-sigma statistical uncertainty of power density of individual seven million spatial fuel regions was about 1% or less for those regions where the relative (normalized) power density was equal to or above unity, typically increasing for regions with lower power. Overall, 95% regions had the 1-sigma uncertainty below 3%.
(e) Additional results and analyses address issues beyond the scope of this benchmark, such as the fission source convergence and fuel depletion.

Leppänen [8] recently selected the current benchmark problem to test performance of the Serpent code. In his model, there were ~6.4 million fuel tallies. He examined the issue of source convergence, and used the problem to assess effectiveness of two available flux estimators (based on conventional surface tracking, or Woodcock delta tracking). For the portion of the analysis most relevant for this benchmark, he performed simulations using a total of 100 billion histories (40,000 active cycles with 2.5 million source neutrons each). Over 90% of tallies had 1-sigma uncertainties below 2%. These calculations required a total of 3,538 CPU-hours, or just over 21 wall-clock days on 7 CPUs. Interestingly, that would correspond to ~470,000 histories per processor per minute, about 3 times more than observed in reference [6]. However, the CPU clock was 25% faster (3 GHz vs 2.4 GHz), and the simulations were performed on 7 independent CPUs, i.e., essentially without any parallel overhead penalty. After accounting for these differences, the number of histories would likely be within a factor of two. Moreover, a Serpent simulation using the traditional surface-tracking method roughly doubled the CPU time, thus approximately equating the performance when measured consistently. Without further specifics of each simulation it is meaningless to directly compare their efficiency; nevertheless, it is fair to state that rater similar "throughput" has been achieved, providing more confidence if an extrapolation into future performance is to be attempted.

M.-J. Lee et al., [9] started with the preliminary version of the benchmark, and modified it into a 2D problem to study CMFD formulation in multigroup framework for variance reduction. While not directly comparable to performance parameters of our interest, it demonstrates utility of having a well defined realistic benchmark.

In the initial benchmark proposal, Hoogenboom and Martin [7] presented preliminary MCNP5 analysis pointing to potential bottlenecks. For that version of MCNP, there is a significant overhead associated with tallying, which would make using simulations with 7 million tallies completely impossible. Moreover, a decrease of efficiency to 82% was observed for a parallel calculation with 24 CPU cores.

Recent analysis by the authors of this paper generally confirms the initial findings. In one series of calculations, a 184-core Linux computer cluster has been used with results for speedup and parallel efficiency shown in Fig. 4. On the left, the straight line represents the perfect speedup, equal to the number of CPUs. The lower curve shows the actual speedup, which not only starts to significantly lag below the straight line for more than 20 processors, but levels off for more than 60 processors, and even decreases when more than 90 processors are used. The ratio of the observed speedup to ideal speedup, i.e., parallel efficiency, is shown in the right figure. It drops to 60% for 32 CPUs, and below 40% for 60 CPUs. These results confirm that the parallel speedup may deteriorate when using more than 30 processors on such a cluster.

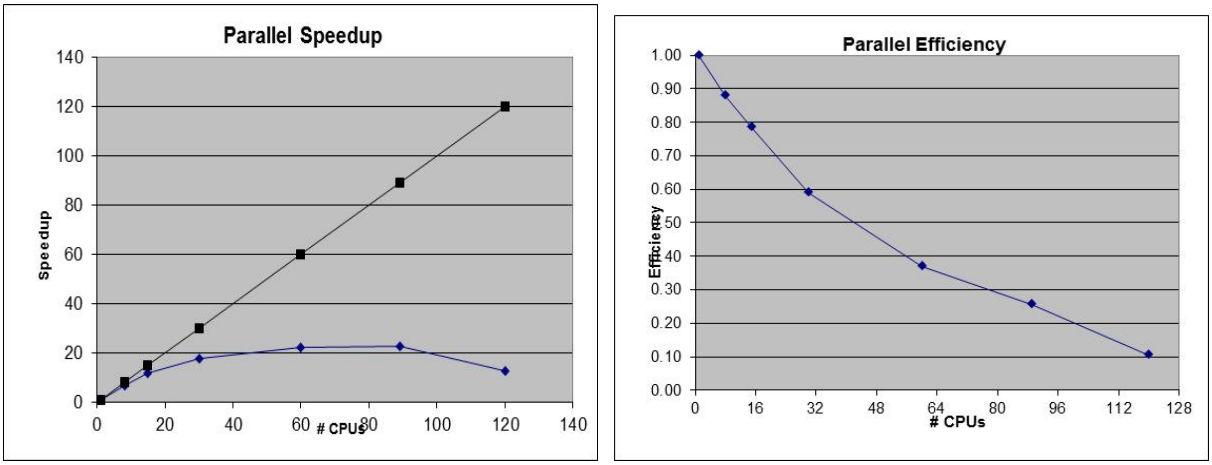


**Figure 4. Parallel speedup (left) and parallel efficiency (right) as a function of number of processors obtained from a Linux computer cluster**

MCNP5 was used by the present authors with the so-called FMESH tally, which allows to estimate the power in all about 7 million requested axial fuel zones. It was also found that MCNP5 sets a rendez-vous point for all processors in a parallel calculation after a maximum of 100,000 processed histories at each processor. Removing this limit requires a (small) modification of the source code, but seems to have a considerable effect on the parallel efficiency when using very large number of histories per cycle.

While it would still take many days on a single processor, this work demonstrated that modern Monte Carlo codes, designed with specific focus on reactor-physics simulations, has potential to reach the objectives on a mid-size cluster (several hundred cores) within days.

As discussed, more general purpose Monte Carlo codes (including MCNP) may not exhibit the same level of parallel performance. However, this benchmark may prove useful to pinpoint to and resolve the bottlenecks, thus leading to their improved performance.

In general, several research groups have already used the benchmark with different Monte Carlo codes ran at different computer systems and produced results that allow meaningful comparison.

## 7. OUTLOOK

As the benchmark addresses items that are considered very relevant for current and future Monte Carlo calculations for nuclear reactors, the present authors and initiators of this benchmark expect a considerable number of contributions over the years, including different contributions from the same participant with different Monte Carlo codes or code versions and/or different

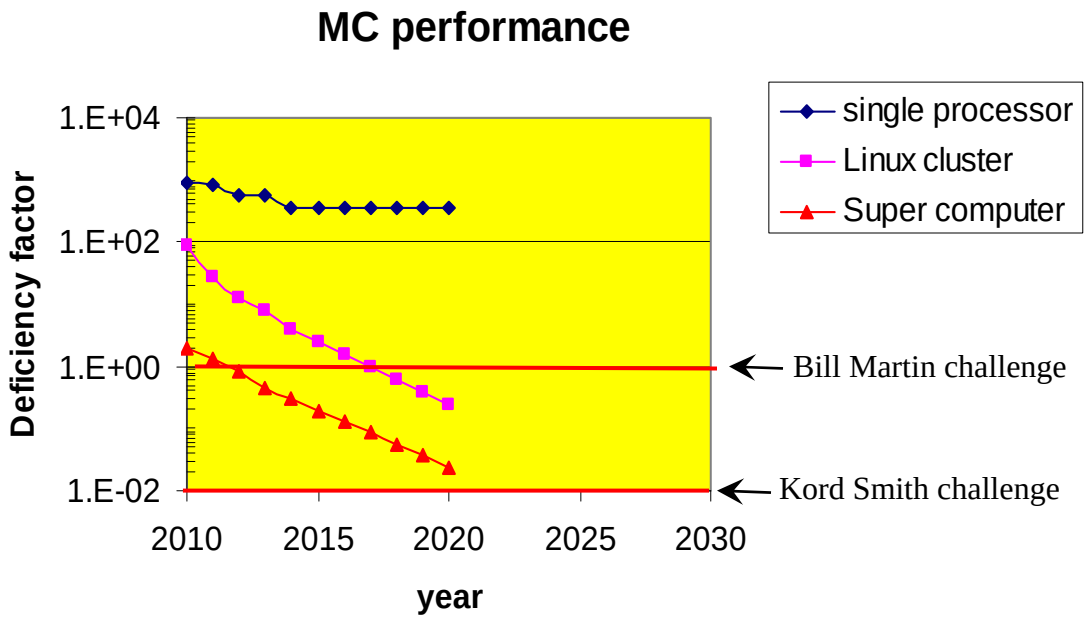


Figure 5. Possible results of the benchmark exercise for the deficiency factor for different computer systems to reach the target (purely fictious data).

computer systems. This will hopefully result in a clear tendency when the detailed power density calculation in a full-size reactor can be done in an acceptable time with acceptable statistical uncertainty. Figure 5 shows an example of how this may look like through the years. This graph displays the factor in the figure of merit that must be overcome to reach the goal of at most 1 % statistical uncertainty in the power estimate in all axial volumes within 1 h computing time. This factor is called the deficiency factor as it indicates the deficiency at a certain time with respect to the final goal. Note that this graph contains purely fictitious data and is intended only to give an idea what can be expected. The deficiency factor will depend of course on the computer system used, especially how many processors are used effectively. Therefore, fictitious data are shown for a single processor (assuming some increase in efficiency of the Monte Carlo simulation code, the tallying process and due to the application of variance reduction methods), a Linux cluster

and a supercomputer (both with an increasing number of processors over the years according to Moore's law).

Of course there will be different forms of these computer types, such as general purpose graphical processing units (GPGPUs), and one should actually expect clouds of data points from different contributions using different computer systems. For the use of supercomputers with huge numbers of processors it will be interesting to see whether the efficiency of parallel processing can be maintained as the number of processors increases.

Notwithstanding the enormous increase in computer power that we can expect over the next decade(s), this should not be a reason to just wait for more powerful computers. In fact one may also expect improvements in the Monte Carlo codes for more efficient generation of neutron histories, but also in the tallying process when large numbers of tallies are involved and one may expect that advanced variance reduction methods will be further developed and implemented in general purpose Monte Carlo codes. All those developments will undoubtedly also lead to completely new Monte Carlo codes taking full advantage of the best Monte Carlo simulation and tallying techniques and of new computer architectures. This benchmark exercise will register the results of all these developments.

Making use of the current widespread internet communication it may be useful to open an internet forum on this benchmark exercise where participants can raise specific problems, discuss improved methods to enhance the Monte Carlo efficiency of a calculation, make suggestions, etc., which will result in best practices for the type of Monte Carlo calculations necessary for a full-size reactor core. This may also result in the definition of extensions of the current benchmark test as discussed in Sect. 4.

As the benchmark exercise will span a long period of 10 to 20 years, it is thinkable that the coordination and analysis is taken over by the total group of participants communicating via internet and working together as a kind of open source project or start a benchmark Wiki. It will be interesting to see whether such a construction will work. This will solve the problem that the current initiators of the benchmark exercise will not be able to coordinate the activities over such a long time.

## 8. CONCLUSIONS

From the contributions reported in the literature we can conclude that the benchmark is already successful before the official launching by the NEA Data Bank. A number of unique aspects have been mentioned, among which the facility to upload contributions for all participants and the possibilities to analyze and process results in a more or less automated way in order to provide on-line insight in the state of the art at any time to reach the final goal of being able to calculate the detailed power density in a full core reactor in an acceptable time with acceptable statistical uncertainty. Unique features that may be implemented in the future are a forum on the website for discussing results and prospects and developing extensions of the current benchmark for testing other relevant issues with regard to the Monte Carlo calculation of realistic full-core reactors.
From the results reported up to now we can conclude that

- the benchmark is feasible; no problems with interpretation of the specifications are reported
- various Monte Carlo codes and computer systems have been used
- the benchmark proved useful to identify certain issues with parallel performance
- first results indicate that 10 to 100 billion histories must be simulated to get the standard deviation below a few percent in most of the 7 million tally volumes.
- for meeting the Bill Martin challenge (1 % standard deviation within 1 hour) a gain in efficiency of a factor 20 to 200 on a supercomputer (compared to the 400 processors used by Kelly et al. [6]) is still needed, using more processors keeping up the same parallel efficiency and/or improved tallying and simulation in the Monte Carlo code.

The initiators of the benchmark exercise expect a considerable number of future participants and contributions in the course of time which will not only show when the final goal of the benchmark will be reached, but will also stimulate new developments in Monte Carlo simulation techniques, development of new Monte Carlo codes more tailored to detailed calculation of full-size reactor cores and taking maximum advantage of current and new computer architectures and parallel processing.

It may also lead to a new form of organizing and conducting benchmark exercises in the nuclear field where a kind of open source project is used on the internet where a larger group of interested people determine together the specifications and perform the necessary analyses.

## ACKNOWLEDGMENTS

The authors gratefully acknowledge the efforts of the NEA Data Bank to realize the logistics for this benchmark exercise.

## REFERENCES

1. Kord Smith, "Reactor Core Methods," Invited lecture at the *M&C 2003 International Conference*, April 6-10, 2003, Gatlinburg, TN, USA (2003).
2. William R. Martin, "Advances in Monte Carlo Methods for Global Reactor Analysis," Invited lecture at the *M&C 2007 International Conference*, April 15-19, 2007, Monterey, CA, USA (2007).
3. J.E. Hoogenboom, W. R. Martin and B. Petrovic, "Monte Carlo Performance Benchmark for Detailed Power Density Calculation in a Full Size Reactor Core," Bechmark Specifications Revision 1.1, June 2010, http://www.nea.fr/dbprog/MonteCarloPerformanceBenchmark.htm (2010).
4. F. B. Brown, "A Review of Monte Carlo Criticality Calculations – Convergence, Bias, Statistics," *Proceedings M&C 2009*, Saratoga Springs, NY, USA, May 3-7, 2009, on CD-ROM, American Nuclear Society, LaGrange Park, IL, USA (2009).
5. H. J. Shim, C. H. Kim, "Real Variance Estimation in Monte Carlo Wielandt Calculations," *Proceedings of the International Conference on the Physics of Reactors "Nuclear Power: A Sustainable Resource"* (*PHYSOR 2008*), Interlaken, Switzerland, September 14-19, 2008.
6. D. J. Kelly et al., "MC21 Monte Carlo Analysis of the Hoogenboom-Martin Full-Core PWR Benchmark Problem," *Proceedings of PHYSOR 2010 – Advances in Reactor Physics to*

*Power the Nuclear Renaissance*, Pittsburgh, Pennsylvania, USA, May 9-14, 2010, on CD-ROM, American Nuclear Society, LaGrange Park, IL (2010).

7. J. Eduard Hoogenboom and William 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", *Proceedings M&C 2009*, Saratoga Springs, NY, USA, May 3-7, 2009, on CD-ROM, American Nuclear Society, LaGrange Park, IL, USA (2009).
8. J. Leppänen, "Use of the Serpent Monte Carlo Reactor Physics Code for Full-Core Calculations" *Proceedings of the Joint International Conference on Supercomputing in Nuclear Applications and Monte Carlo 2010 (SNA + MC2010)*, Tokyo, Japan, October 17-21, 2010.
9. M.-J. Lee et al., "Multigroup Monte Carlo Reactor Calculation with Coarse Mesh Finite Difference Formulation for Real Variance Reduction," *Proceedings of the Joint International Conference on Supercomputing in Nuclear Applications and Monte Carlo 2010 (SNA + MC2010)*, Tokyo, Japan, October 17-21, 2010.
10. B. Petrovic, "MCNP Performance Evaluation for the NEA 3D PWR Benchmark," *Proceedings of the Joint International Conference on Supercomputing in Nuclear Applications and Monte Carlo 2010 (SNA + MC2010)*, Tokyo, Japan, October 17-21, 2010.
11. D. van Veen and J. E. Hoogenboom, 'Efficiency Improvement of Local Power Estimation in the General Purpose Monte Carlo Code MCNP,' *Proceedings of the Joint International Conference on Supercomputing in Nuclear Applications and Monte Carlo 2010 (SNA + MC2010)*, Tokyo, Japan, October 17-21, 2010.