ANL-03/14

Monte Carlo Criticality Source Convergence in a Loosely Coupled Fuel Storage System

by R. N. Blomquist and E. M. Gelbard

Nuclear Engineering Division

Nuclear Engineering Division
Nuclear Engineering Division
Nuclear Engineering Division
Nuclear Engineering Division
Nuclear Engineering Division
Nuclear Engineering Division
Nuclear Engineering Division
Nuclear Engineering Division
Nuclear Engineering Division
Nuclear Engineering Division
Nuclear Engineering Division
Nuclear Engineering Division
Nuclear Engineering Division
Nuclear Engineering Division

Argonne National Laboratory, Argonne, Illinois 60439
Operated by The University of Chicago for the
United States Department of Energy under Contract W-31-109-Eng-38

Nuclear Engineering Division
Nuclear Engineering Division
Nuclear Engineering Division
Nuclear Engineering Division
Nuclear Engineering Division

Available electronically at http://www.doe.gov/bridge

Available for a processing fee to U.S. Department of Energy and its contractors, in paper, from:

U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831-0062
phone: (865) 576-8401
fax: (865) 576-5728
email: reports@adonis.osti.gov

ANL-03/14

Argonne National Laboratory
9700 South Cass Avenue
Argonne, Illinois 60439-4801

MONTE CARLO CRITICALITY SOURCE CONVERGENCE IN A LOOSELY COUPLED FUEL STORAGE SYSTEM

by

R. N. Blomquist and E. M. Gelbard

March, 2003

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

MONTE CARLO CRITICALITY SOURCE CONVERGENCE IN A LOOSELY COUPLED FUEL STORAGE SYSTEM

Roger N. Blomquist and Ely M. Gelbard
Argonne National Laboratory
9700 S. Cass Ave.
Argonne, Il 60439-4842
RNBlomquist@anl.gov

ABSTRACT

The fission source convergence of a very loosely coupled array of 36 fuel subassemblies with slightly non-symmetric reflection is studied. The fission source converges very slowly from a uniform guess to the fundamental mode in which about 40% of the fissions occur in one corner subassembly. Eigenvalue and fission source estimates are analyzed using a set of statistical tests similar to those used in MCNP, including the "drift-in-mean" test and a new drift-in-mean test using a linear fit to the cumulative estimate drift, the Shapiro-Wilk test for normality, the relative error test, and the "1/N" test. The normality test does not detect a drifting eigenvalue or fission source. Applied to eigenvalue estimates, the other tests generally fail to detect an unconverged solution, but they are sometimes effective when evaluating fission source distributions. None of the test provides completely reliable indication of convergence, although they can detect nonconvergence.

I. INTRODUCTION

Loosely coupled systems provide special challenges for Monte Carlo analysts because of difficulties in convergence of the unaccelerated power method normally applied to the fission source iteration and because of inaccuracy of the usual variance computations [1,2,3]. Source convergence difficulties can be due to the inherent unsuitability of the unaccelerated power iteration method, because of undersampling, or because of other statistical problems. The fission source iteration acceleration methods available in deterministic codes do not always work well because of the noise inherent in the Monte Carlo method, although some mitigating algorithms are in use [e.g., 4,5,6]. The OECD Nuclear Energy Agency Expert Group on Fission Source Convergence in Criticality Safety Analyses has specified four source convergence benchmark problems [7], each of which has some combination of adverse stochastic and deterministic fission source convergence characteristics. The first of these problems, a loosely coupled spent fuel storage array, is studied in this work; it exhibits deterministic slow convergence and rather complex stochastic behavior in a particularly difficult combination.

Slow convergence is not an inherent problem in Monte Carlo or deterministic calculations unless its progress is so slow as to be mistaken for convergence. When this happens, the estimated fission distribution will be in error, and in some problems, the estimated eigenvalue can be substantially under-predicted, with adverse implications for criticality safety analysis. There are several obvious remedies: (1) perform additional calculations with more histories and/or more neutron generations used to converge the fission source, (2) perform similar calculations with a different initial source and compare the solutions, and (3) apply statistical tests that provide the analyst warnings when the fission source is insufficiently converged. In this report, we apply several tests that may be used in criticality calculations to assess their sensitivity and reliability, using as a test problem the checkerboard problem from the OECD/NEA benchmark set.

II. CHECKERBOARD CONVERGENCE BENCHMARK CONFIGURATIONS

The checkerboard benchmark problem is an array of spent fuel subassemblies stored in a rectangular 3x24 array. Half of the array cells are empty (water-filled), and the 36 subassemblies are arranged in the array in a checkerboard pattern, alternating with water channels (Figure 1). The top and bottom of the system are 30cm water reflectors, three of the sides are 40cm concrete reflectors, and the fourth (long) side is a 30cm water reflector. The atom densities are given in Table I.

In accordance with the benchmark Monte Carlo parameter specifications (Table II), the problem was run using the VIM Monte Carlo code [8] with 1000, 2000 and 5000 histories per generation, and, for every given number of histories per generation, results were computed for 500 active generations with 20, 40 and 100 generations skipped. In all the above problems, the starting source was selected randomly with uniform probability distribution throughout each fuel cell, and uniformly among the 36 fuel cells. The

checkerboard benchmarks were also run with starting sources only in certain individual cells. In our calculations, generations were collected into batches of 20, and, unless otherwise specified, all means and variances were computed on the basis of this batching. These calculations and those made at other institutions, reported elsewhere [9], are not analyzed here because a) probably no analyst would run the problems this way, and b) if any did it would quickly be seen from the evolving fission distributions that such runs were very far from converged in all cases. Instead, we will examine the source convergence behavior of a few of the specified cases and some simplified systems.

The NEA source convergence benchmarks are unusual in that it is not the exact solution of the physics problem that is of interest. Instead, the focus is on the rate and extent of convergence, the accuracy of uncertainty estimates, and the efficacy of statistical tests. Therefore, in our work, we used a simplified version of this problem both for the benchmark comparisons and for the additional analysis of its convergence properties and the statistical tests reported here. The subassembly pin lattice was homogenized by volume weighting (see Table I for the homogenized atom densities), but the surrounding water and structure in each cell in the pool array were retained to preserve the inter-cell coupling. The reflectors surrounding the storage array were also retained as specified. The homogenization no doubt changes the eigenvalue and perhaps the converged fission source distribution somewhat, but should not have a large effect on the rate of convergence or the utility of statistical tests.

For our own purposes we performed several other analyses. We reran the uniform source problem with 25,000 histories per generation, again with the same three alternative numbers of skipped generations, all with 500 active generations (identified as cases 37-39). Additional analysis was performed on an even more simplified problem set with a 2xM (M even) array with periodic boundary conditions in the x-y plane and reflection in z. These systems represent an infinite lattice, so the true fission fraction for each of the M subassembly locations is precisely $1/M$, and even if a non-uniform source guess is used, the salient feature of the Monte Carlo solution is statistical variation due to the Monte Carlo algorithm, exacerbated, as we shall see, by the very low coupling between nearest fuel bundles. Because the eigenvalues of these systems depend only on the flux shape inside a subassembly and not on its distribution among subassemblies, the eigenvalue should converge fairly quickly, even if the global fission source distribution does not.

The simplest test problem is one in which only two columns of the checkerboard are explicitly included. Thus in this problem the whole configuration contains only two fuel cells alternating with two water cells. The most complicated in this series involves 12 fuel cells alternating with 12 water cells in 12 columns of the 2 rows. All calculations were done with 500 active generations, skipping 20 and 100 generations. The net number of starters per generation, in each test, is the closest integer to N,

$$N = 5000\, M/36, \qquad (1)$$

so that the mean number of starters per fuel cell is nearly the same in all test problems, and nearly the same as in the original checkerboard benchmark with 5000 histories per

generation. In order to introduce flux-shape drifts, which turned out to be important in the checkerboard benchmark series, the guessed source was raised in the left half of each problem configuration and lowered in the right half, so that the ratio between the two halves was 1.5.

One of the important characteristics of the checkerboard benchmark problem is the influence of the concrete reflector on the fundamental mode. The geometrical arrangement of the fuel subassembly locations in the array is uniform, so one would expect a fairly uniform flux shape from one end of the array to the other. Initial observations of the flux distribution, however, indicated a flux peak in the only cell with two faces adjacent to concrete (1,3). To bound the effect of the surrounding materials, two calculations were completed for a single subassembly cell surrounded on four sides by 40cm of either concrete or water. Those eigenvalues, 0.8778 +/- 0.0012 and 0.8285 +/- 0.0012, respectively, confirm that cell (1,3) ought to contain a large peak reminiscent of that in the Whitesides problem [1], and that the fission fraction in that cell is the parameter that most directly indicates fission source convergence.

III. STATISTICAL TESTS AND RESULTS

Monte Carlo codes normally attempt to provide information sufficient for conclusions about the statistical validity of the Monte Carlo estimates, and some provide warnings when tests indicate computational results may be untrustworthy. In MCNP [10], for example, one set of these tests is intended for general use, while a second group is aimed at eigenvalue problems; but in principal, all the tests could be used for both eigenvalue and fixed source calculations. We discuss, here, five tests, all of them variants of tests in MCNP, and describe their weaknesses when applied to eigenvalue problems. It would seem on theoretical grounds that they can all be misleading when estimates from successive generations are very strongly correlated. This weakness is particularly troublesome in difficult problems (like the checkerboard benchmark problem) because so many eigenvalue problems are difficult precisely due to strong intergenerational correlations. Thus in using these tests it is necessary to proceed with caution, relying on a background of empirical studies.

Two tests will be considered in Section A below. Both involve separate consideration of Monte Carlo estimates over the first and second halves of a Monte Carlo run. Both are designed to determine whether Monte Carlo estimates are fluctuating randomly about a stable mean, or about a mean which has drifted significantly during the active generations of the Monte Carlo computation. In Section B the successive estimates of fission fractions and eigenvalues are subjected to the Shapiro-Wilk test for normality. Next, in Section C, we discuss the MCNP "*R* test", based on the size of the relative errors. Finally, in Section D, we consider a test we'll call "the 1/*N* test" in which one looks for a 1/*N* behavior of the relative standard deviation during *N* batches. All tests will be applied, here, to the checkerboard problems specified in the criticality safety benchmark series. In addition we examine their performance in simplified problem configurations described above. The quantities of interest in all of our studies are eigenvalues and fission fractions,

i.e. the ratios of the fission rates in each fuel cell to the total fission rates. In all tests the active generations were grouped into batches, each consisting of 20 consecutive generations. The basic data in all tests were averages of estimates over each of the 25 active batches. Variances in all tests will be estimated without regard for inter-batch correlation.

A. Two "Drift-in-Mean" Tests

This drift test is a variant of the"bottom half vs. top half" test in which the averages from each half of a calculation are compared to detect drift. In our variant, we compute a cumulative mean for each tested variable (i.e., eigenvalue or fission fractions) after each generation. Then we construct a linear least-squares-fit (with slope a) to the cumulative means in the last half of the Monte Carlo calculation. Strictly speaking (again because of intergenerational correlations) no statistical interpretation can be assigned to the fitting coefficients. Still, the fitting parameters do give us some indication as to the extent to which cumulative means are drifting. The expression

$$? F_{fit,i} = 13\, a_i / F_{i,25} \qquad (2)$$

is the relative drift in fission fraction; $F_{i,j}$ being the average of the fission fraction estimates in fuel cell i, taken over all active batches up through active batch j. Here, because there are 25 active batches, the halves into which the active batches are divide are the first 12 and the last 13. The same test is also applied to cumulative eigenvalue estimates.

In the second drift test we have copied one of the MCNP tests which examines drifts over all active generations. Here again the Monte Carlo run is divided into halves, but now means are computed, separately, over each half. If, for example, k_T is the mean of k over the first (top) half of the run, and k_B the mean over the last (bottom) half, define the eigenvalue difference in standard deviations, $? k_{TB} = (k_T - k_B)/s_{?k}$, where $s_{?k}$ is the standard deviation in this difference. $? k_{TB}$ is used, in this test, as the parameter whose value determines the acceptability of the eigenvalue estimate. In MCNP a computed mean is acceptable if the magnitude of $k_T - k_B$ is within the 99% probability limit, assuming that $k_T - k_B$ is normally distributed. This means here that $|k_T - k_B| < 2.6 s_{?k}$, where $2.6 s_{?k}$ is the approximate width of the 99% confidence interval. In the case of fission fraction in fuel cell i we use the corresponding parameter $? F_{TB,i}$.

The results of the drift tests described above are both listed in Table III for the simplest of this series of tests, computation of the system with only two fuel cells, skipping 20 generations (one batch). Here s_{Fi} is the Monte Carlo estimate of the standard deviation of the mean fission fraction, using all active batches. Interpreting $? F_{fit,i,}$ as the amount of drift along the fitted line we see that, over and above statistical fluctuations, the Monte Carlo estimates in the later batches of the iterative process have drifted by about 4.4% and –4.4%, respectively, in the two fuel cells. These drifts, by our standards here, are only slightly larger than the quoted standard deviations. Further we note that both values of $? F_{TB,i}$ lie in the acceptable range. We don't know what would happen if the

number of generations were increased, so it isn't possible to make a rigorous connection between the computed drift-parameters and the errors due to nonconvergence of the power iterations. The most we can hope for is that there will be a useful correlation between these parameters and the magnitude of such errors. In column six of Table III the $?F_{fit,i,}$'s give remarkably good estimates of the true errors in the Monte Carlo means, but one can't expect that this will be true in general. We note also that both values of $?F_{TB,i}$ lie in the acceptable range.

In the fitting process one computes the covariance between cumulative means and generation-numbers, and the corresponding correlation coefficient is a good measure of the goodness-of-fit. In fact suppose we denote the cumulative mean after *j* generations as X_j, and the variance of this running mean as $s^2(X_j)$. Let $y(j) = a*j + b$ be the equation of the linear fit to X_j, and $s^2(y)$ be the variance of this line. In other words $s^2(y)$ would be the variance of *X* if all the cumulative means lay exactly on the linear fit. The magnitude of the correlation coefficient is, then, the ratio of the standard deviation of y to the standard deviation of *X*; i.e. the correlation coefficient tells us how much of the standard deviation of cumulative means is attributable to linear drift.

It will be seen that the signs of $?F_{fit,i}$ and $?F_{TB,i}$ are the same, suggesting that each of the $F_{i,j}$'s have, in effect, drifted in a single direction throughout the run. Generally $?F_{fit,i,j}$ and $?F_{TB,i}$ give complementary information The latter parameter gives us information as to a general drift in $F_{i,j}$ throughout the run, while $?F_{fit,i}$ characterizes the behavior of the running averages, $F_{i,j}$, nearer the end of the run. Thus one may find that $?F_{TB,i}$ is relatively large while $?F_{fit,i}$ is small. In such a case it may be that the power-iteration process is close to convergence by the end of the run. On the other hand if $?F_{fit,i}$ is large, particularly if the magnitude of the correlation coefficient is close to one, then this process is probably still far from convergence.

Table IV lists the same data for the corresponding problem with twelve fuel cells, again skipping one batch. On comparing Tables III and IV what stands out most clearly is that (1) the quoted relative standard deviations in the computed fission fractions are much larger in Table IV than in Table III, and (2) that in Table IV the true errors are much larger than the standard deviations estimated by the Monte Carlo code. Both of these observations are consistent with our *a priori* assessment that the Monte Carlo process in the second problem should be much noisier and more slowly convergent than in the first. But we see also that (3) $?F_{fit,i,j}$, when it is much larger than $s_{Fi}/F_{i,25}$, might provide a useful warning that the computed mean may be grossly incorrect. Again we see that the signs carried by $?F_{fit,i}$ and $?F_{TB,i}$ tend to be the same, though they do differ in cell 6. Thus here a drift in the bottom half of the run again indicates an overall drift in the same direction. As before, we note that results of both tests depend on drifts over generations already run, and don't necessarily tell us what would happen if the run were continued. Both tests can be misleading as indicators of the reliability of means in individual cells. Thus in cell 4 the large values of $?F_{fit,i}$ and $?F_{TB,i}$ indicate, incorrectly, that the Monte Carlo standard deviation is much too small, while in fact it's reasonably accurate. Clearly both tests are misleading in this one cell, but taken as a whole they both

agree, and do tell us that the Monte Carlo power iterations are still far from convergence after 600 generations.

We note also that the correlation coefficients are very large, except in cell 6. This means that in most cells the generation-to-generation changes in fission fractions are not random but very systematic, and almost linear functions of the iteration number. Such behavior is not at all what one would expect of a set of means nearing convergence, and is an important indication that, in fact, that the Monte Carlo power iterations are far from convergence.

In Table V, we list similar results for eigenvalues in a range of simplified test problems, including a reference two-cell computation with 1000 active generations, 500 skipped and 5000 histories per generation. Generally, on examining both sets of test-results in Table V, we would be led to conclude that the eigenvalue computations are pretty well converged. Thus, for example, the magnitude of the net drift in k in the last half of the iterative process is everywhere less than twice the computed standard deviation. It seems surprising, therefore, to find such large correlation coefficients in cases 1 and 2 of Table V, since Table III suggests that the flux shape in case 1 is well converged. One might speculate that such strong drifts could be caused, somehow, by the tilt in the source-guess. We find, however, that the computed correlation coefficients remain just as high in magnitude when the tilted source-guess is replaced by a flat source-guess. At this point the high correlation coefficient is not understood. It will be seen, however, that the "Approximate Error" generally decreases as the net number of histories increases, again suggesting that the errors are primarily due to statistical effects, perhaps of the nature observed by Yamamoto [4], rather than to incomplete convergence of the power iterations. Again we see that the direction of drifts seems to persist throughout the calculation, and here we find that both tests agree everywhere.

We turn now to the original benchmark problem. In Table VI we list data similar to those listed in Table IV, but for the benchmark case 25, in which the source guess is uniform over all fuel cells, the computation is run with 5000 histories per generation, and there are 500 active generations after 20 generations skipped. Here fuel cells are grouped according to the rows in which they are located. Thus cells 1, 4, 7 and 10 are the left-most four fuel cells in row 1, etc. To avoid a very lengthy table we have listed data only for the 11 cells nearest to the left-hand boundary of the checkerboard. We hope it will become apparent later that this is a particularly interesting area of the checkerboard.

It will be seen that here the absolute values of ratios of the $?F_{fit,i,}$'s to the computed relative standard deviations are generally pretty large. In Table IV the maximum absolute value of this ratio is 3.6. Here this maximum is 4.7, taken on in cell 2. The average of all these absolute values is 2.2 in Table IV, 3.6 here. Our drift-in-mean criteria suggest that the case 25 results are still far from converged. Further we see that, in the lower half of the run the fission fractions are rising in cell 2 and its closest neighbors, and falling elsewhere. In this case results of both drift tests agree everywhere except in cells 5 and 11. We see from Table VI that in both cells the first test is right. Both tests, however, strongly suggest that the Monte Carlo iterations have not converged. Thus for

example, the first test indicates that the computed fission fractions are acceptable in only three out of 36 cells.

The same analysis has been performed for case 1, with only 1000 histories per generation. In this case we found, according to the first drift test, that fission fractions are acceptable in 5 out of 36 cells, again suggesting unsatisfactory convergence. Since case 1 flux shapes are probably no better converged than the shapes in case 25, probably the increase in numbers of results deemed acceptable by the drift test indicates deterioration in the test's performance due to the smaller sample size. Again we find that both drift tests give very similar results.

In Table VII, we show corresponding results for a "best estimate" calculation with 1000 active generations, 1000 skipped and 25000 histories per generation. Results shown in Tables VI and VII are more or less consistent, in the sense that fission fractions in cell 2 and its neighbors have risen while, except for cell 4, fission fractions in all other cells have dropped. On the other hand, it seems that our best-estimate computation has still not converged very well, particularly in cells far from corner-cell 2. It's a plausible guess that fission fractions in these cells are still too high, though the overall accuracy does appear to be much higher than those listed in Table VI.

Next, in Table VIII, we list eigenvalues and drift parameters for a few of the benchmark cases, now using the best estimate eigenvalue (0.85875) to compute an eigenvalue error. As one would expect, given the behavior of the flux shape, all eigenvalues in Table VIII are too low, and don't get substantially more accurate as the number of histories increases. Apparently the drifts in fission rates per cell, here as in the idealized checkerboards, make canceling contributions to the eigenvalues. It seems plausible that there is an upward drift in all the eigenvalues, corresponding to what appears to be a shift in fission source towards the most reactive fuel cell. The first test detects this drift in cases 13, 27, 37 and 39, while the second test sees the drift only in the last cases, cells 37 and 39. It seems that in cell 15 the drift reverses and tends upward in the lower half of the run. At any rate it's clear that neither drift test reliably indicates that the eigenvalues have not yet converged in cases 1, 3, 15 and 25. A user would have to take note of the drifts in source shape to infer that these eigenvalues might not have converged.

Concluding this section we note that here, in most cases, both drift tests give the same results, with the first test slightly better than the second. It will be recalled that, according to our first test, an estimate is unreliable if the "relative drift" is greater than twice the estimated standard deviation while, according to the second test, it is suspect if the drift-to-uncertainty ratio is greater than 2.6. It can be shown that the test ratios in both tests are closely related, and one might argue that the tests should be inter-compared with the same critical values of test-ratios in both. Where one of the tests fails it is usually not strict enough, so one might be inclined to reduce the critical drift-to-uncertainty ratio from 2.6 to 2. The test results for both tests would then become still closer. Generally the drift parameters do seem to give useful information as to the state of convergence of the

flux shapes. We see however, from Table VIII, that test results for eigenvalues are much less satisfactory.

B. Shapiro-Wilk Test

This test is intended to tell us whether or not a Monte Carlo estimate is normally distributed. There are two reasons for checking normality, both connected with confidence intervals. First, if an estimate isn't normally distributed the distribution may be pathological. It might, for example, have a very long, low tail, in which case one could run into scattered outliers (i.e. estimates very far from the mean), so that the true variance may be very difficult to estimate. But even if outliers aren't a problem, and even if individual samples are drawn from Gaussian distributions, the probability that the true answer is outside of the stated confidence interval may be underestimated when successive estimates are strongly positively correlated.

The Shapiro-Wilk test [11] takes a set of values of a random variable and computes a corresponding variate *W*. There is at least one variant of this test which accepts a sample-set of any size, but in its original form the test is intended for use with sample-sets of no more than 50 elements. This is no difficulty in our present application since the sets we treat will be, here, samples from 25 active batches.

For 25 gaussian samples the 95% limits for W are 0.918 and 0.985, and the expected value is about 0.95. Considering the eleven cells listed in Table VI we find 6 cells for which W is in this range. These are cell 2, with W = 0.95: cell 5, in which *W* = 0.92: cell 6, *W* = 0.98: cell 9 with *W* = 0.93; cell 10, *W* =0 .96, and: cell 11 with *W* = 0.94. Yet we have seen that for these cells the VIM standard deviations grossly overestimate the accuracies of the computed fission fractions. The Shapiro-Wilk test is, however, not without value in this case in that it rules out as unacceptable cells 1, 4 and 8, three of the four cells acceptable according to the drift tests.

In Table IV one finds that, by the Shapiro-Wilk test, the fission fraction estimates are unacceptable only in cells 8 and 11, though we know that the computed fission fractions are substantially incorrect in most cells. In Table VI all but the case 15 eigenvalue are acceptable, according to this test, though in fact they are all substantially too low.

One difficulty here, seems to be that the Shapiro-Wilk test (if one can judge from the discussion of Ref. 12) is intended for use with uncorrelated samples. *A priori* it can't be assumed to give us any information about correlations, or their effects on confidence intervals, but we see in Table V that the correlation coefficients are high in 5 of the 6 cells where the Shapiro-Wilk test is satisfied. In cells 9 and 11 the cumulative fission fractions in the lower 13 batches are very nearly linear functions of the batch number. It's easy to see what would be the value of *W* for samples exactly linear in batch number, and to show that this value would not depend on the slope or intercept of the line on which the samples lie. For 10 and 25 such samples we find that *W* = 0.97, and for 20 samples W = 0.96. Furthermore one can show that, if *x(i) = a*i + ?*, and if *?* is a normal random

variable, then W may well be in the acceptable range**.** We conclude that the Shapiro-Wilk test is of questionable value if successive iterates are strongly correlated. Here it should be noted that we have applied this test only to 25 batches of generations, not to the generations themselves, and the test may be used differently in MCNP.

C. Relative Error (*R*) Test

The MCNP manual[10] suggests that, in a fixed-source problem, a Monte Carlo mean should be regarded as untrustworthy if *R*, the relative standard deviation, is greater than 10%. In difficult problems estimated standard deviations are often too small, and this is true both for fixed source and eigenvalue problems; but, in fixed source problems, as the number of sample particles increases the relative error in estimated standard deviation goes to zero. This will not be true in eigenvalue calculations if intergeneration correlations are strong enough. If intergenerational correlations are very strong and positive, relative standard deviations will continue to be strongly underestimated even as the number of starters per generation goes to infinity. For this reason the *R* test will tend to be much less useful for eigenvalue problems than for fixed source problems. Still there seems to be no reason why an estimate that fails the *R* test should not be flagged as unreliable. In our computations, however, the *R* test would have only one effect, i.e. it would lead us to label as questionable the cell-1 fission fraction in Table VI, ruled acceptable by both drift tests.

D. *1/N* Test

The 1/*N* test is used to see whether the estimated uncertainty is decreasing at a rate consistent with increasing sample size. In our implementation, it is applied only to the bottom half of the calculation to determine the behavior of the cumulative mean. Let N_T (14 here) be the batch-number of the first of the final 13 active batches of a Monte Carlo run that skips N_{skip} batches, and $N_B = N_{skip} + N$ be the number of the last of the active batches, where *N* (25, here) is the net number of active batches. It's the purpose of this test to determine whether, in the final batches, the relative standard deviations, *R(j)*, of the estimates *x(j)*, decreases like 1/sqrt(*N(j)*), where $N(j) = j + 1 - N_T$ is the number of batches in the range from N_T down to and including *j*.

For this purpose we first compute the cumulative mean of *x*, *X(j)*, where *j* ranges from N_T through N_B. Let $R_B(j)$ be the standard deviation of the mean of all *x*'s with batch-numbers in the range from N_T through *j*, and define the relative error $R(j) = R_B(j)/X(j)$. Define $C(j) = R(j)*\mathrm{sqrt}(j-N_T+1)$, which ought to be constant for normally distributed *x*'s produced from a converged fission source distribution. Now, given *C*(*j*), we fit the line

$$y(j) = a_C*(j-N_T+1) + b_C \qquad (3)$$

to the values of the *C*'s. Finally, we compute the relative change in *C(j)*, $?\,C_{TB} = a_C*13/C_{bar}$, over the range of the fitted line (the bottom half of the Monte Carlo calculation). The magnitude of $?\,C_{TB}$ is then to be taken as a measure of the fractional

change in C, and correspondingly, a measure of the magnitude of the uncertainty of R, due to the drift in C. Here, C_{bar} is the mean of C over the active batches.

It should be noted that, even if the computed value of R has converged, we still have no assurance that the converged value gives us a valid confidence interval. If successive values of x have a high correlation coefficient then, even if the MacMillan correction [12] is applicable (i.e. even if the x's are eigenvalue estimates), computed confidence intervals may be far from correct. In contrast, a converged mean can be notably incorrect only when Monte Carlo biases are significant. Although it's not completely clear when this will be true, there seems to be no indication that, in practice, such biases have affected neutronics eigenvalue calculations.

Here perhaps the worst failure of the preceding tests is their failure to detect nonconvergence in four of the eight eigenvalues listed in Table VIII. For this reason we list values of ? C_{TB} for these 8 cases in Table IX. We see that the observed relative drifts in C can be very small, certainly too small to warn the user that the stated confidence intervals for eigenvalues are grossly underestimated. Thus in case 1 the computed R, $R=s/k$, is incorrect by 0.74%: but the relative drift in R is only 0.02% of R. Such a small shift in the estimated relative error is much too small to warn the user that the computed eigenvalue may be in error by as much as 0.74%.

IV. CONCLUSIONS

We've seen above that, in dealing with eigenvalue problems, none of the tests we've discussed can tell us reliably that specific Monte Carlo computational results are trustworthy, though some tests can give us useful warnings when they're not. We are not interested here in difficulties due to weaknesses in modeling, or in Monte Carlo biases. Generally we mean by "untrustworthy results" results which, from the user's point of view, are unacceptably inaccurate, stated with misleadingly small confidence intervals.

Thus anomalies in source shapes are, here, often, associated with detectable drifts in local source densities. We see, however, from Table VIII that anomalies in eigenvalues do not necessarily reveal themselves through corresponding drifts in the computed eigenvalues. In Table VIII all the computed eigenvalues are too low, and are presumably drifting upwards as the computation progresses; but apparently the eigenvalue drift is often too slow to detect, being hidden by noise.

The computation of confidence intervals is generally based on the assumption that specified samples are drawn from a normal distribution. Since we are mainly concerned, here, with the validity of confidence intervals, it might be expected that the Shapiro-Wilk test for normality would be particularly valuable. This is very likely true in fixed-source problems where the samples are independent; but the Shapiro-Wilk test is applicable only to independent samples. It does not seem to be designed to detect correlations among the samples. Yet it's usually just such correlations that cause trouble in the estimation of statistical uncertainties in eigenvalue problems. For our purposes, then, the test's utility is

limited. Specifically, we find that it tends to accept erroneously small confidence intervals when, as is often the case, the submitted samples are highly positively correlated. It is, in other words, a test which is particularly prone to give false negatives. On the other hand there seems to be no reason to question the test results when the normality hypothesis is rejected.

As already pointed out, one weakness of the *R*-test in difficult eigenvalue problems is that, even as the number of histories per generation goes to infinity, the relative error in the computed R will often remain large and negative. Correspondingly the test often may fail to flag unreliable results. On the other hand, if a computational result doesn't pass this test it should be viewed with suspicion.

Finally, the *1/N* test may be misleading for two reasons. First, particularly in eigenvalue calculations the drift in C, *C=R*sqrt(n)*, may be too slow to see through the noise and, secondly, even if *C* has converged it may have converged to an erroneous value. Again, however, if the test detects an unacceptable result, i.e. an unacceptably large drift, the resulting warning ought to be taken seriously.

In the work above one type of test has not been examined. This is a test based on a comparison of various types of estimators (e.g. track-length estimators, collision estimators, etc.) of eigenvalues or reaction rates. These different estimators should not disagree significantly, i.e. their differences should be consistent with the standard deviation in the differences, computed with regard for correlations between estimators. Still one can expect that no foolproof test of this type will be discovered.

The cumulative results of batteries of tests should help the Monte Carlo user avoid many serious mistakes. Further exploration of such batteries of tests, using different series' of test problems and development of new tests would be highly desirable. What is most disturbing in our results is that tests applied to eigenvalues, directly, have had so little success in detecting anomalies. After all, in criticality safety applications it's the eigenvalue that really counts. Our only recommendation, here, has been to infer the degree of convergence of eigenvalues from the degree of convergence of fission shapes, not the most satisfactory approach. Of course even such modest conclusions ought to be assessed through analysis of other sorts of difficult eigenvalue problems.

V. ACKNOWLEDGMENTS

This work was performed under the auspices of the U. S. Department of Energy, Office of Technology Support Programs, under contract W-31-109-ENG-38.

VI. REFERENCES

1. E. G. Whitesides, "A Difficulty in Computing the k-effective of the World", *Trans Am Nucl Soc*, **14** No. 2 (1971).

2. R. J. Brissenden and A. R. Garlick, "Biases in the Estimation of k_{eff} and its Error by Monte Carlo Methods," *Ann. Nucl. Energy*, **113**, No. 2, pp. 63-83 (1986).

3. E. Gelbard and Albert Gu, "Biases in Monte Carlo Eigenvalue Calculations," *Nucl. Sci. and Eng*., **117**, pp. 1-9 (1994).

4. T. Yamamoto, et al., "Fission Source of Monte Carlo Criticality Calculations in Weakly Coupled Fissile Arrays," *J. of Nucl. Sci. and Tech.*, **37**, No. 1, pp. 41-45 (2000).

5. A. Mohamed and E. Gelbard, "Stratified Source-Sampling Techniques for Monte Carlo Eigenvalue Analysis," *Proc. Int. Conf. On Physics of Nuclear Science and Technology*, Hauppage, N.Y., October 5-8,1998.

6. N. Smith, "The Unification of MONK – Extending the Monte Carlo Horizon", *Sixth International Conference on Nuclear Criticality Safety*, Vol I, Versailles, France, September 20-24, 1999.

7. R. N. Blomquist and A. Nouri, "The OECD/NEA Source Convergence Benchmark Program", *Trans Am Nucl Soc*, **87** (2002).

8. R. N. Blomquist, et.al., "VIM-A Continuous Energy Monte Carlo Code at ANL, A Review of the Theory and Application of Monte Carlo Method", *Proceedings of a Seminar-Workshop*, ORNL/RSIC-44, Oak Ridge, TN, April 1980.

9. M. J. Armishaw, "A Review of Benchmark 1"**,** *Trans Am Nucl Soc*, **87** (2002).

10. J. F. Briesmeister, Ed., "MCNP – A General Monte Carlo N-Particle Transport Code, Version 4A", Los Alamos National Laboratory (1993).

11. S. S. Shaprio and M. B. Wilk, "An Analysis of Variance Test for Normality", *Biometrika*, **52**, 591 (1965).

12. D. B. MacMillan, "Monte Carlo Confidence Limits for Iterated-Source Calculations", *Nucl. Sci. Eng.*, **52**, 591 (1965).

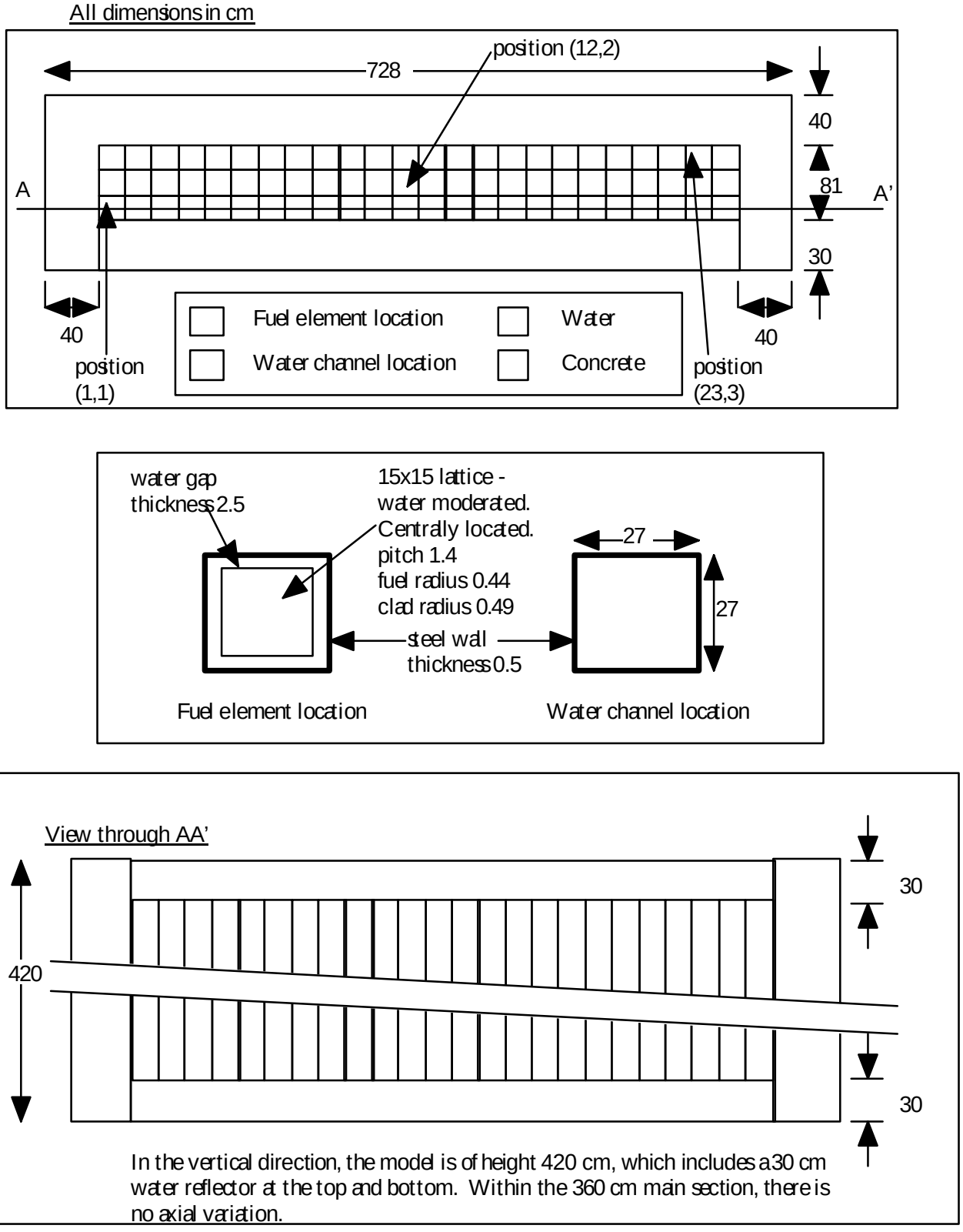

Figure 1. NEA Source Convergence Benchmark 1 Configuration

Table I.
NEA Checkerboard Benchmark Atom Densities

Material	Nuclide	Density
Fuel	^{238}U	2.2380E-02
	^{235}U	8.2213E-04
	O	4.6054E-02
Water	H	6.6706E-02
	O	3.3353E-02
Zirconium	Zr-90	4.2910E-02
Concrete	H	5.5437E-03
	C	6.9793E-03
	Si	7.7106E-03
	Ca	8.9591E-03
	O	4.3383E-02
Iron	Fe	8.3770E-02
Homogenized Subassembly	^{238}U	6.94479E-03
	^{235}U	2.55117E-04
	O	3.48084E-02
	H	4.10345E-02
	Zr-90	3.19820E-03

Table II.
NEA Checkerboard Benchmark Monte Carlo Parameter Specifications

Case	Starting Source	Skipped Generations	Starting source points
1	Uniform	20	1000
2	Uniform	40	1000
3	Uniform	100	1000
4	Location (1,1)	20	1000
5	Location (1,1)	40	1000
6	Location (1,1)	100	1000
7	Location (23,3)	20	1000
8	Location (23,3)	40	1000
9	Location (23,3)	100	1000
10	Location (12,2)	20	1000
11	Location (12,2)	40	1000
12	Location (12,2)	100	1000
13	Uniform	20	2000
14	Uniform	40	2000
15	Uniform	100	2000
16	Location (1,1)	20	2000
17	Location (1,1)	40	2000
18	Location (1,1)	100	2000
19	Location (23,3)	20	2000
20	Location (23,3)	40	2000
21	Location (23,3)	100	2000
22	Location (12,2)	20	2000
23	Location (12,2)	40	2000
24	Location (12,2)	100	2000
25	Uniform	20	5000
26	Uniform	40	5000
27	Uniform	100	5000
28	Location (1,1)	20	5000
29	Location (1,1)	40	5000
30	Location (1,1)	100	5000
31	Location (23,3)	20	5000
32	Location (23,3)	40	5000
33	Location (23,3)	100	5000
34	Location (12,2)	20	5000
35	Location (12,2)	40	5000
36	Location (12,2)	100	5000

Table III.
Fission Fractions for Two Fuel Cell System.

Cell No.	$F_{i,25}$	s_{Fi}	$s_F/F_{i,25}$ (%)	True Error ($F_{i,25}$ - 0.5) (%)	$?F_{fit,i}$ (%)	Correl. Coef.	$?F_{TB,i}$ (%)
1	0.478	0.013	2.8	-4.4	4.2	0.74	2.0
2	0.522	0.013	2.7	4.4	-3.8	0.74	-2.0

Table IV.
Fission Fractions for Twelve Fuel Cell System.

Cell No.	$F_{i,25}$	s_{Fi}	$s_F/F_{i,25}$ (%)	True Error ($F_{i,25}$ – 1/12) (%)	$?F_{fit,i}$ (%)	Correl. Coef.	$?F_{TB,i}$ (%)
1	0.111	0.0072	6.4	33	21	0.90	2.1
2	0.116	0.0051	4.4	39	16	0.98	4.6
3	0.110	0.0043	3.9	32	5.2	0.85	1.0
4	0.0860	0.0044	5.1	3.2	16	0.94	5.1
5	0.0853	0.0049	5.8	2.4	4.2	0.76	1.4
6	0.0654	0.0059	9.1	-21	-0.7	-0.08	-7.7
7	0.0667	0.0064	9.5	-20	-32	-1.0	-4.9
8	0.0699	0.0058	8.3	-16	-20	-0.97	-2.4
9	0.0650	0.0044	6.7	-22	-15	-0.95	-1.7
10	0.0619	0.0052	8.5	-26	-28	-0.91	-4.8
11	0.0724	0.0068	9.4	-13	-4.0	0.54	-0.68
12	0.0903	0.0058	6.5	8.4	14	0.84	0.48

Table V.
Eigenvalues in Idealized Checkerboard Calculations.

No. Fuel Cells	Generations Skipped	k_{eff}	s_k	s_k/k (%)	$?k_{fit,25}$ (%)	Correl. Coeff	Approx. Error (%)	$?k_{TB}$ (%)
2	500 (reference)	0.8608	0.00034	0.04	-0.02	-0.56	0.00	-0.08
2	20	0.8582	0.0025	0.29	-0.50	-0.94	-0.30	-1.9
2	100	0.8568	0.0024	0.28	-0.31	-0.88	-0.46	-2.0
4	20	0.8572	0.0015	0.17	0.10	0.83	-0.42	1.2
4	100	0.8585	0.0016	0.17	-0.05	0.86	-0.11	-0.57
8	20	0.8607	0.0013	0.19	-0.07	-0.43	0.00	0.17
8	100	0.8606	0.0013	0.15	-0.16	-0.83	0.00	-1.1
12	20	0.8613	0.0010	0.12	-0.13	-0.53	0.05	-1.8
12	100	0.8614	0.00097	0.11	-0.08	-0.78	0.11	-1.8

Table VI.
Benchmark Problem Fission Fractions for Uniform Starting Source

Cell No. (*i*)	$F_{i,25}$	$s_F/F_{i,25}$ (%)	$?F_{fit,i}$ (%)	Correl. Coeff.	$?F_{TB,i}$ (%)
1	0.018	18	34	0.70	2.0
4	0.0027	7.7	-14	-0.77	-0.5
7	0.031	11	-34	-1.0	-5.5
10	0.017	9.0	-26	-0.95	-5.4
3	0.047	7.9	30	0.96	3.6
6	0.055	4.0	2.1	-0.34	-1.2
9	0.042	10	35	1.00	-4.6
2	0.083	10	47	0.00	6.9
5	0.051	7.9	26	0.97	1.6
8	0.066	6.6	-7.3	-0.87	-1.5
11	0.028	8.6	-28	-0.89	-2.3

Table VII.
Benchmark Problem "Best Estimate" Fission Fractions.

Cell No. (*i*)	$F_{i,25}$	$s_F/F_{i,25}$ (%)	$?F_{fit,i}$ (%)	Correl. Coeff.	$?F_{half,i}$ (%)
1	0.091	1.8	-5.6	-0.97	-12.
4	0.053	2.2	0.04	0.11	-1.5
7	0.015	3.5	-12	-0.98	-25
10	0.0043	11.	-20	-0.84	-40
3	0.16	1.1	3.1	0.96	6.6
6	0.048	2.2	-3.8	-0.85	-10.
9	0.011	8.1	-22.	-0.98	-48.
2	0.46	0.08	2.3	0.96	5.0
5	0.11	1.6	5.4	0.98	9.7
8	0.028	5.6	-16.	-0.94	-34
11	0.0050	7.0	-6.2	-0.85	-23

Table VIII.
Selected Benchmark Series Eigenvalues

Case No.	Hists/ Gen.	Gens. Skip-ped	k_{eff}	s_k	s_k/k (%)	$?k_{fit,25}$ (%)	Correl. Coeff.	$?k_{TB}$ (%)	Approx Error (%)
1	1000	20	0.8524	0.0014	0.16	0.19	0.34	1.8	-0.74
3	1000	100	0.8529	0.0015	0.18	0.28	0.17	0.81	-0.69
13	2000	20	0.8516	0.0010	0.12	0.32	0.06	0.90	-0.83
15	2000	100	0.8518	0.00091	0.11	0.22	0.09	-0.57	-0.82
25	5000	20	0.8529	0.00069	0.11	0.14	0.22	0.94	-0.68
27	5000	100	0.8538	0.00063	0.08	0.18	0.59	1.8	-0.58
37	25000	20	0.8527	0.00030	0.04	0.11	0.51	3.8	-0.71
39	25000	100	0.8532	0.00032	0.04	0.10	0.34	3.4	-0.65

Table IX.
Relative drifts $?C_{TB}$, in $C = R$*sqrt(N), for computed eigenvalues.

Case No.	Approx. Error (%)	$?C_{TB}$ (%)	R (%)
1	-0.74	-0.02	0.16
3	-0.69	-1.1	0.18
13	-0.83	-2.1	0.12
15	-0.82	3.8	0.11
25	-0.68	-4.4	0.11
27	-0.58	0.35	0.08
37	-0.71	28	0.04
39	-0.65	21	0.04

Distribution for ANL-03/14 Internal:

J. I. Sackett, OTD-ER, 208
S. K. Battacharyya, OTD, 360
P. J. Finck, OTD, 208
J. Helt, OTD-ER, 208
D. P. Weber, NE, 208
H. S. Khalil, NE, 208
E. K. Fujita, NE, 208
R. N. Hill, NE, 208
J. Roglans, NE, 208
C. Grandy, NE, 208
R. N. Blomquist, NE, 208
G. Palmiotti, NE, 208
C. G. Stenberg, NE, 208
T. A. Taiwo, NE, 208
K. N. Grimm, NE, 208
J. Stillman, NE, 208
R. Hwang, NE, 208
R. M. Lell, NE, 208
J. A. Morman, NE, 208
C. H. Adams, NE, 208
B. J. Toppel, NE, 208
R. D. McKnight, NE, 208
J. E. Cahalan, NE, 208
K. L. Derstine, NE, 208
T. Wei, NE, 208
R. N. Hwang, NE, 208
T. H. Fanning, NE, 208
J. Deen, NE, 360
N. Hanan, NE, 360
Y. Gohar, NE, 360
R. W. Schaefer, NT-AW
J. Carpenter, PNS, 360
IPD-203

External:

ANL-East Library
ANL-West Library
DOE – Office of Scientific and Technical Information
NE Division (30)
F. B. Brown, LANL
A. Forster, LANL
R. Little, LANL
T. Ueki, LANL
R. MacFarlane, LANL
M. Chadwick, LANL
Rick Anderson, LANL

Tom McLaughlin, LANL
M. Westfall, ORNL
C. Parks, ORNL
M. DeHart, ORNL
Lester Petrie, ORNL
Hamilton Hunter, ORNL
Calvin Hopper, ORNL
Luiz Leal, ORNL
Tim Valentine, ORNL
Ian Gould, ORNL
B. Kirk, ORNL
J. Wagner, ORNL
Hans Toffer, ?????
David Heinrichs, LLNL
Song Huang, LLNL
Rich Evarts, LLNL
J. Blair Briggs, INEEL
Enrico Sartori, OECD/NEA
Ali Nouri, OECD/NEA
Nigel Smith, Serco Assurance - Winfrith
Malcolm Armishaw, Serco Assurance – Winfrith
Olivier Jacquet, IRSN
Joachim Miss, IRSN
Dennis Mennerdahl, EMS
John Wagner, ORNL
Nobutoshi Shirai, JNC
Yaushi Nomura, JAERI
Toshihrio Yamamoto, JAERI
Yoshitaka Naito, NAIS
Greg O'Connor, BNFL?
Jim Gulliford, BNFL
Mikail Kalugin, MCU Kurchatov Institute
Dmitri Oleinik, MCU Kurchatov Institute
Kiyosumu Takahide, JAERI
Hadosy Gabor, KFKI Budapest
Adolf Garcia, DOE Idaho
Hoyt Johnson, DOE-EM-21
Michael Thompson, DOE-NA-117
James Felty, DOE-DP
Kim Loll, DOE
John Evans, DNFSB
Jerry McKamy, DOE-EH
Matthew Hutmaker, DOE-DP
Ray Schwartz, DOE
Richard Stark, DOE-EH
William Lake, DOE-RW

Ivon Fergus, DOE – OA-50
David Crandall, DOE-NA-11
Burt Rothleder, DOE-EH
Tom Reilly, SRS-EH
C. Dunford, BNL (NNDC)
P. Oblizinsky, BNL (NNDC)
Robert Wilson, DOE-RF
Todd Taylor, BWXT-Idaho