TAC No. MA5085 (NMP1)

Dr. Steven Penn

25 March 1999

Darl Hood
Office of Nuclear Reactor Regulation
United States Nuclear Regulatory Commission
Washington, DC 20555-0001

Dear Mr. Hood,

I have recently read your report "Safety Evaluation by the Office of Nuclear Regulatory Regulation Related to Extension of the Schedule for Reinspecting Core Shroud Vertical Welds Niagara Mohawk Power Corporation Nine Mile Point Nuclear Station, Unit 1" (Docket No. 50-220) issued 2 November 1998. I have also reviewed the studies issues by Niagara Mohawk Power Corporation (NMPC) which were submitted to the NRC to justify the operating extension authorized in the safety evaluation (SE). I am writing to alert you to several serious errors in the NMPC studies — errors which undermine the conclusions of the NMPC and the NRC that the crack growth rate may be safely assumed to be below 2.2×10^{-5} inches/hour. The most egregious errors were in the data analysis of the sensitization (EPR) measurements. The researchers who performed that measurement have not followed the scientific method and as a result have significantly overstated of the accuracy of their results. Once properly analyzed, the sensitization data does not infer the lower crack growth rate nor does it rule out irradiation assisted stress corrosion cracking (IASCC) as the cracking mechanism. In general, the tests results presented by NMPC do not clearly limit the effects of neutron irradiation on the crack growth. In fact, there is significant correlation between the fluence profile and crack depth profiles for both welds V-9 and V-10, suggesting that the fluence has a sizable contribution to the cracking mechanism.

Enclosure 1

I am also somewhat concerned about your review of the NMPC studies. In your SE I saw no mention of the errors present in the NMPC studies. Many of the errors in the NMPC report were rather blatant, and I am uncertain how these mistakes could have escaped your attention.

On 3 December 1998, I sent a letter to your office which detailed my review of the SE issued by your office on 8 May 1997 entitled "Safety Evaluation by the Office of Nuclear Regulatory Regulation Regarding Results of the Reinspection of the Core Shroud Niagara Mohawk Power Corporation Nine Mile Point Nuclear Station, Unit 1" (Docket No. 50-220). As of this writing, I have not received a reply to that letter. I trust that you are the appropriate person to whom I should be sending my evaluations and that you intend to answer the points I have raised. If that is not the case, I would appreciate a letter informing me of the correct person or office who will review and answer my concerns on these safety related matters and when I might expect that reply. In addition I ask that you send me a brief letter acknowledging receipt of this letter and informing me when and from whom I should expect a reply.

Below I present my review of the NRC SE dated 2 November 1998 of NMPC Nine Mile Point, unit 1 (NM1). If you have any clarifying questions regarding the material presented below, please feel free to contact me via telephone or electronic mail.

REVIEW OF NRC SE ISSUED 2 NOVEMBER 1998 REGARDING THE EXTENDED OPERATING CYCLE AT NMPC NM1.

- **Conflict of Interest and the Need for Independent Analyses**

As I discussed in my letter of 3 December 1998, General Electric (GE) and NMPC have a conflict of interest which prevents calculations performed by these corporations and their paid consultants from being viewed as truly objective research. Therefore all critical measurements and calculations performed by these corporations which were used as a basis for a safety evaluation should be repeated by the NRC or its paid consultants. Moreover, ALL data and methodology should be available for public review. I find it particularly disturbing that some of the supporting information for this SE have been withheld from public exposure by General Electric. For example, in attachment 5 of the NMPC report (#NMP1L 1290, 27 February 1998) many of the results and figures have been removed. As a result there was not sufficient data for me to check

the validity of the GE results. This closed-book research thwarts any objective public review – review processes which are a fundamental method for promoting accuracy and discouraging misconduct in the scientific community. Like any scientist, I have concerns about conclusions drawn from any data or methodology which is kept secret. The need for full disclosure is especially poignant in light of the errors found in the Altran Corporation EPR measurements presented in the same NMPC report (#NMP1L 1290). The NRC should not base their decisions upon any research where the complete data set and methodology are not available for public review. If GE and NMPC are unwilling to fully publish their data and their methodology then the NRC should sponsor independent and open research on the material in question.

- **Explicit Citation of Errors**

A point which I also raised in the 3 December letter, but which deserves repeating, is the necessity for explicit citation of errors for every critical value. When reviewing this SE I have no way of judging the validity of the results unless the error is cited as well. As I demonstrated in that previous letter, values which the SE authors had claimed were significance instead held little meaning when viewed in the context of their associated error. An example of this is given at the top of page 8 of my 3 December 1998 letter where I discuss the crack growth rates assumed from the Brunswick BWR data.

- **Sensitization Measurements of the Core Shroud Boat Samples**

On page 4 of the SE, the NRC presents its conclusion that the cracking in the core shroud is IGSCC and not IASCC. This conclusion is based upon an optical examination of the cracking features, and measurements of the material sensitization and yield strength. The optical examination of the material specimens, while useful in establishing the basic features of the cracking, is not very quantitative and can overlook microcracking or minute granular defects.

The material sensitization was measured using Electrochemical Potentiokinetic Reactivation (EPR) testing. These tests and the results are reported in the Altran Corporation Technical Report (TR 97181-TR-03). The analysis of the data, presented on page 6, is fundamentally flawed. The

researcher has discarded data without proven justification and in some cases with no more reason than the data was "out of line". The researcher's data analysis procedure was highly invalid. Such selective dismissal of the data undermines objectivity and incorporates the researcher's bias into the results.

To be clear, let me briefly review the proper methodology for handling experimental error. All experimental data is subject to both systematic effects and statistical background noise. A good experiment will minimize systematic effects which can influence the experimental results. If systematic effects remain which contribute in a significant way to the data, these effects should be characterized, the data should be corrected if necessary, and the systematic error should be calculated.

Statistical background noise is in many ways a simpler problem to address than systematic error. Background noise obeys a statistics which is usually Gaussian (the noise statistics for this experiment are most likely Gaussian, but the exact statistics in this case do not influence the conclusions of my argument). This noise statistics describes the distribution of the results from a series of measurements of an ideal value. Another way of looking at this distribution is that it describes the probability of recording a given result when trying to measure a given ideal quantity. This distribution implies that one can indeed make a valid measurement which is far from the mean (the expected value), but that the probability for making such a measurement is low. To determine that probability distribution, the researcher must perform a series of measurements. The distribution of the data set establishes the mean value and the standard deviation which approach their respective ideal values as the size of the data set approaches infinity. A more complete description of fundamental data analysis can be found in *Data Reduction and Error Analysis for the Physical Sciences* by P. R. Bevington. This book is an very valuable resource and is required reading for most physics graduates. I would suggest that NRC personnel who perform or contribute to NRC safety evaluations become familiar with the material covered in this book.

My point in describing systematic and statistical error is that the experimenter cannot discard data either because they do not "like" the value that they measured or because they think that the

data may have been influenced by a systematic effect which they have not characterized. Such exclusion of data improperly biases the data and violates the scientific method.

Now, let us review the data excluded in the EPR measurements as described on page 6 of the Altran Technical Report (TR 97181-TR-03). The purpose and method for the measurements are presented in the report. I will comment here only upon the aberrations in the analysis.

1) For calibration purposes, the researchers performed measurements on two standards, labeled "S" and "F". The researchers disregarded the data from run 1 of the standard "S" because they said the data was "out of line" with the other runs. No other reason was provided. However one cannot discard data without a demonstrated reason for doing so. Therefore the discarded data should be included in the data analyzed for run 1. Including this data yields a result for standard "S" of $IR/IA = (8.80 \pm 3.82) \times 10^{-3}$ and not $IR/IA = (7.00 \pm 1.41) \times 10^{-3}$. Both results are consistent with the standard's engraved reference value of 0.008. By comparison the result for standard "F", for which no data was excluded, is $IR/IA = 0.373 \pm 0.006$ which is significantly lower, 10σ or 14%, than the engraved reference value of 0.434.

2) The measurements performed on the V-9 sample were equally problematic. The author has thrown out measurement 1 saying that the surface was not correctly prepared. The author does not detail how the surface was incorrectly prepared nor how that altered procedure might be expected to affect the results. In fact the author cannot even say for certain that the sample surface was improperly prepared. The note accompanying the data states, "Run #1 was not valid since the operator acknowledged that the preparation of the test area *might not be appropriate*" (emphasis mine). Some proof that this measurement is indeed aberrant is required. We cannot reasonably ignore this data without a more detailed justification.

3) Similarly measurement 4 on the V-9 sample is said to be improper because it contained weld deposit and the test result would not accurately represent either the weld material or the SS plate material. No further justification is given. If measurements performed near the weld were known to yield inaccurate results, then the researchers should have cited a reference documenting this fact. Indeed if this problem was known about by the researchers, one wonders why the measurement was taken at this location. On the other hand, if this systematic effect was not known or understood, then the effect should have been documented with repeated measurements. The researcher could have clearly demonstrated that his presumption was justified. It sounds like a reasonable assumption, but it demands

proof. The updated results for V-9 are $IR/IA = 0.255 \pm 0.248$ for runs 1–3 and $IR/IA = 0.534 \pm 0.528$ if we include all the data.

4) For the measurements of sample V-10, the researcher once again attempts to discount data with the sole justification that he doesn't like its value. It is interesting to note that the researcher is willing to discount the data from Run #1 because it is a factor 2 larger than the "normal" values measured for runs #2 and 3. However the researcher does not seem to have a similar problem with the data taken for run #4 which is a factor 4 smaller than the data from runs 2 and 3. This research practice suggests that the researcher is looking for an answer and that he has introduced an unfair bias into his data analysis method.

5) The data for run #5 on sample V-10 is also noted for being unusual because IR was negative. The researcher notes "It is unclear why IR value for Run#5 was negative. This value was questionable until further clarification is obtained." No further clarification is provided in the report. Since IR was negative, the result for that run should have been $IR/IA = -0.122$, but the report lists the result as $IR/IA = 0.122$. Assuming that negative IR/IA values are meaningful, the actual result for sample V-10,including all the runs, is $IR/IA = 0.115 \pm 0.148$. Without any clarifying information, I believe it is prudent to exclude Run #5, in which case the result for V-10 is $IR/IA = 0.172 \pm 0.107$.

6) On page 6 of the report, the authors discuss why they chose not to use the sample of type 316L SS as a calibration standard. The sample had a level of sensitization midway between standards "F" and "S". The authors state that using the 316L sample as a standard was "inappropriate" due to the molybdenum content and they cite cautions from the DOS tester manual. However the manual only notes that in order to measure material with molybdenum content, such as type 316 SS, one must use a larger test area and a more concentrated electrolyte. In fact the measurement for the 316L sample, $IR / IA = 0.021$, was low when compared to the actual value, $IR / IA = 0.025$, by 16%. The measurements on standards "S" and "F" were also low by similar factors. Given the similarity in results, it is not clear that a problem with this measurement exists.

In table 1 I compare the original EPR results as reported by McDermott Corp. with the corrected results using all valid data as I have described above. These EPR results are taken using the double-loop method. I will refer back to these results in the next section when I discuss the conversion of these double-loop (DL) results to their equivalent single-loop (SL) values.

	Standard "S"	Standard "F"	Sample V-9	Sample V-10
McDermott Results	$(7.00 \pm 1.41) \times 10^{-3}$	3.73 ± 0.006	0.080 ± 0.016	0.124 ± 0.050
Corrected Results	$(8.80 \pm 3.82) \times 10^{-3}$	3.73 ± 0.006	2.55 ± 2.48	1.72 ± 0.107

Table 1: Comparison of McDermott Double-loop EPR results to the Corrected Results

The experimental technique described in this section of the report is clearly problematic with 5 of the 18 data runs reporting some difficulty. By arbitrarily disregarding several data points, the researcher introduces a bias into the results. Further evidence that the researcher might not be approaching this work from an objective viewpoint comes on page 4 of the report where the researcher writes:

> "The degree of sensitization of a material influences IGSCC, and is one input parameter used to predict crack growth rates using the General Electric PLEDGE code (Reference 9). A quantitative measurement of the degree of sensitization associated with the core shroud vertical welds was needed **to confirm the material conditions that had been** assumed to estimate crack growth rates using the PLEDGE model. These predicted crack growth rates were used to estimate the progression of identified cracks in the V-9 and V-10 vertical core shroud welds at NMP-1 for the duration between restart and the next planned outage. Crack disposition has been detailed in the crack evaluation and the safety analysis reports submitted to the Nuclear Regulatory Commission (References 4 and 5). These reports assumed a moderately sensitized material condition. Austentic stainless steel material that is moderately sensitized should yield EPR test results between 2 and 15 coulombs/cm^2 in the single loop test. EPR testing was undertaken **to confirm these assumptions**:" (emphasis mine)

The researcher states that the purpose of the testing was *to confirm assumptions*. The purpose of the testing should have been to measure the true sensitization of the material, not to confirm assumptions. By phrasing the experimental motivation in this fashion and then inappropriately dismissing some of the data, the researcher calls into question the validity of his/her work.

- **Calibration of the Sensitization Data**

According to the authors, in order for the above mentioned double-loop EPR measurements to be incorporated into the PLEDGE model or compared to other published values, the double-loop

EPR values must be converted to single-loop EPR values. This conversion is discussed in Appendix 2 of the Altran Technical Report. The strategy for converting from double-loop (DL) to single-loop (SL) values was to measure the DL and SL values for two known standards, one with low sensitization (Standard "S") and the other with high sensitization (Standard "F"). The ratio of the measurements (SL/DL) would define the conversion factor. The "S" and "F" standards were provided for calibration purposed with the IHI test equipment, which was used by McDermott to performed the DL measurements discussed in the previous section. The GE CR&D lab has equipment capable of performing DL and SL measurements, and they made both measurements on the two standards.

This procedure, while being quite reasonable, was complicated by the fact that the DL results for the "S" and "F" standards obtained by McDermott and by GE CR&D were vastly different. I will review this data which is presented in Table 2 of Appendix 2 of the Altran report. (I would like to point out that the data is once again presented without error bars to qualify the accuracy of these measurements.)

When they provided the "S" and "F" standards, IHI listed the standard's DL value to be used as a reference for calibration. One might reasonably assume that these values are known with reasonably high accuracy, although the errors are not listed in the report. We can compare these IHI values with the corrected results for the McDermott data which are listed in Table 1 of this letter. The corrected result for Standard "F" ($IR/IA = 0.373 \pm 0.006$) is low by 10σ or 14% compared to the IHI value of 0.434. The corrected result for Standard "S" ($IR/IA = (8.80 \pm 3.82) \times 10^{-3}$) is in agreement with the IHI value of 0.008, but the error is large. It is interesting to note that the McDermott result for Standard "S" presented in table 2 is 0.009, the mean value when all five data points for Standard "S" are used, and not 0.007, which would have been the mean if the data from Run #1 was arbitrarily excluded as the Altran report had stated. The GE CR&D double-loop result for Standard "F" (0.239) is about 40% lower than the IHI value. It is not clear from the report what the true value is for the GE CR&D double-loop measurement of Standard "S". Table 2 of Appendix 2 of the Altran report lists (0.00016) which is an unbelievable factor 50 lower than the IHI value. However the value plotted in Figure 1 of that report is

approximately 0.00045 which is slightly less unbelievable, being low by a factor 18. The researchers dismiss this discrepancy. "It is clear that the IHI system tends to produce higher values. This could be expected due to the nature of the field process and the smaller area examined." However, since no attempt was made to explain this assertion, the reader is left questioning the validity of this statement. For example, why should we assume that the IHI measurement is high and not that the GE CR&D measurement is low? The IHI equipment was calibrated against the standards provided by IHI. How was the GE CR&D instrument calibrated? If the GE CR&D equipment was calibrated to a known and verifiable standard, why was that data not included? If the GE CR&D instrument was working properly, why were the double-loop results for the two standards so low compared to the IHI measurements and by such drastically different factors? The report leaves many unanswered questions about the GE CR&D data.

The ratio of GE results (SL/DL) define the DL-to-SL conversion factor. However since the authors have not presented any calibration data, one cannot be certain what is the true DL value for the two standards. If we assume that the GE data is correct, then the data taken with the IHI equipment must be renormalized to the GE DL results before being converted to SL values.

$$\begin{aligned}\text{Boat Sample SL result} &= \text{Boat Sample DL result} \times \frac{\text{GE Standard DL result}}{\text{IHI Standard DL result}} \times \frac{\text{GE Standard SL result}}{\text{GE Standard DL result}} \\ &= \text{Boat Sample DL result} \times \frac{\text{GE Standard SL result}}{\text{IHI Standard DL result}}\end{aligned} \qquad [1]$$

On the other hand, if we assume that the IHI DL results for Standards "S" and "F" are correct, then there is no need to renormalize the data and we can just use the DL-to-SL conversion factor.

$$\text{Boat Sample SL result} = \text{Boat Sample DL result} \times \frac{\text{GE Standard SL result}}{\text{GE Standard DL result}} \qquad [2]$$

In the Altran technical report, the researchers have assumed that the true DL values for standards "S" and "F" lie anywhere within the range of values between the IHI and GE CR&D results. This assumption implies that there is a range of DL-to-SL conversion factors which adds a additional factor to the uncertainty of the data. I am puzzled why this method was used rather than properly calibrating the instruments.

In addition the researchers have mistakenly assumed that the range of allowed values for the converted SL data is solely dictated by the range in conversion factors. Rather they should have propagated the error in DL results into an error for the SL values. I will illustrate this point by performing the calculation.

In Table 2 I have listed the DL and SL data for standards "S" and "F". The GE data is taken from Appendix 2 of the Altran report. I have chosen to assume that the DL result for GE's Standard "S" measurement is 0.00045, as appears in the graph, rather than 0.00016 which appears in the table. I assume that the discrepancy was due to a typographical error. The choice has a negligible result on the final SL value. To avoid confusion over which of the McDermott DL results to use for standards "S" and "F" I have chosen to use the IHI reference values. I assume that the IHI values are more accurate, but again this assumption will not greatly effect the final SL results.

Testing Lab	Sample	DL EPR Value (Ir/Ia)	SL EPR Value (C/cm^2)
GE CR&D	Standard S	0.00045	0.0077
GE CR&D	Standard F	0.239	31.4
IHI	Standard S	0.008	Not Measured
IHI	Standard F	0.434	Not Measured

Table 2: Data for Conversion Calculation from Appendix 2 of Altran Report

Following equations [1] and [2] and given that the SL and DL values are related by a power law, we plot the data from Table 2 on a log-log plot (Figure 1). The shaded region is the range of conversion factors. The limiting curves are described by the conversion formulae:

Assuming IHI DL results "correct": $d_{IHI} = 0.0177\, s_{IHI}^{0.755}$ or $s_{IHI} = 209.6\, d_{IHI}^{1.325}$ [3]

Assuming GE DL results "correct": $d_{GE} = 0.0829\, s_{GE}^{0.480}$ or $s_{GE} = 177.7\, d_{GE}^{2.08}$ [4]

I have also plotted the corrected McDermott DL results (see table 1) for boat samples V-9 and V-10 with their error bars. The conversion limits establish the range of SL error. This range may also found using equations [3] and [4]. The converted SL data for the boat samples is listed in Table 3.

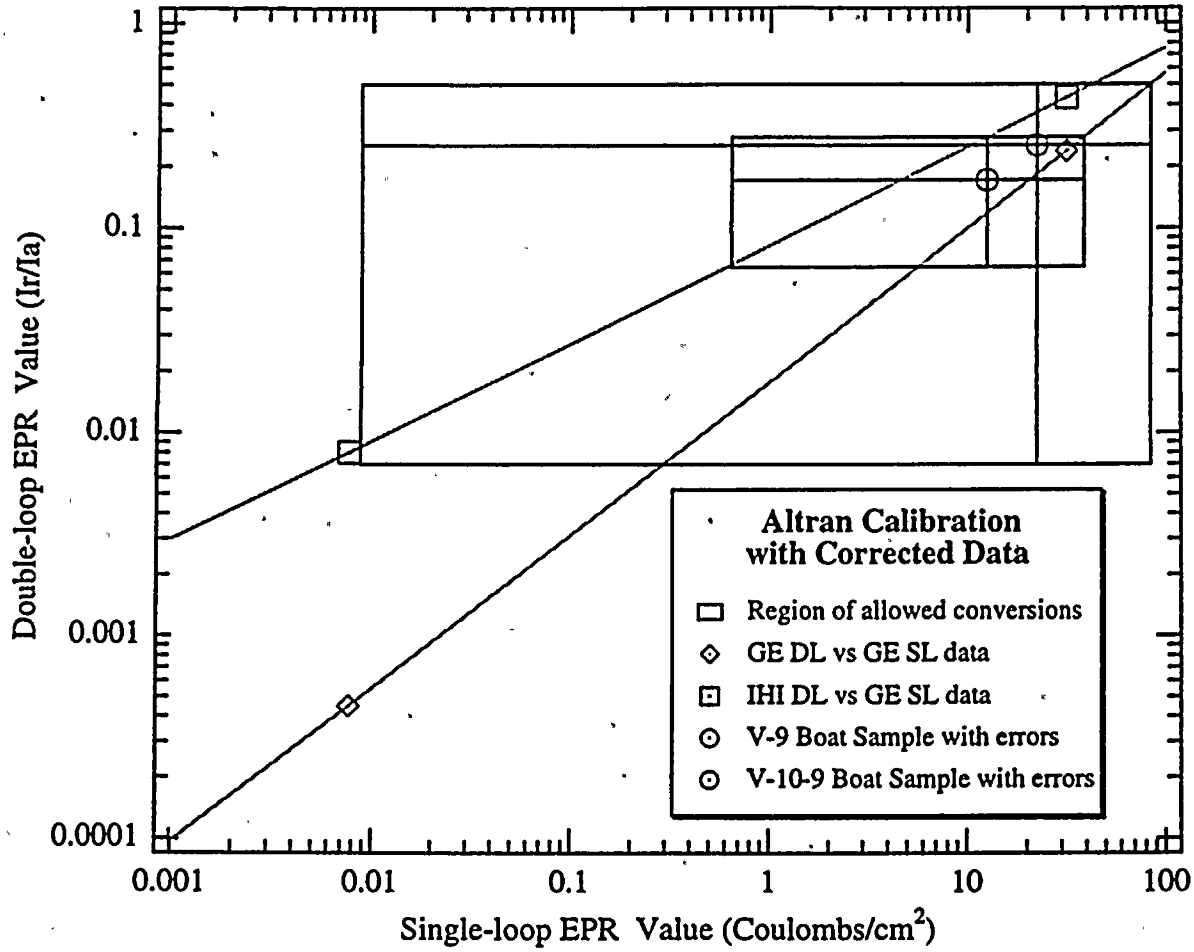

Figure 1: Corrected EPR Data Using Altran Conversion Method

Sample	DL EPR Value (Ir/Ia)	SL EPR Value (C/cm^2)	SL EPR Range (C/cm^2)
V-10	0.172 ± 0.107	12.44 (+26.1, -11.8)	38.5 – 0.603
V-9	0.255 ± 0.248	22.3 (+61.9, -22.3)	84.2 – 0.006

Table 3: Single-loop EPR results using Altran Conversion Method

Therefore, we find that when we correctly analyze the data and propagate the errors the results of the EPR measurements are more than a factor 2 greater than those listed in the Altran report and cited by NMPC. These SL results are also approximately proportional to the relative fluence at these locations. However , it is hard to draw any definitive conclusions since the error bars are quite large.

Testing Lab	Sample	Technique	DL EPR Value (Ir/Ia)
McDermott	V-10	IHI Field Test	0.172 ± 0.107
McDermott	V-9	IHI Field Test	0.255 ± 0.248

Table 4: McDermott Double-loop EPR data with Errors

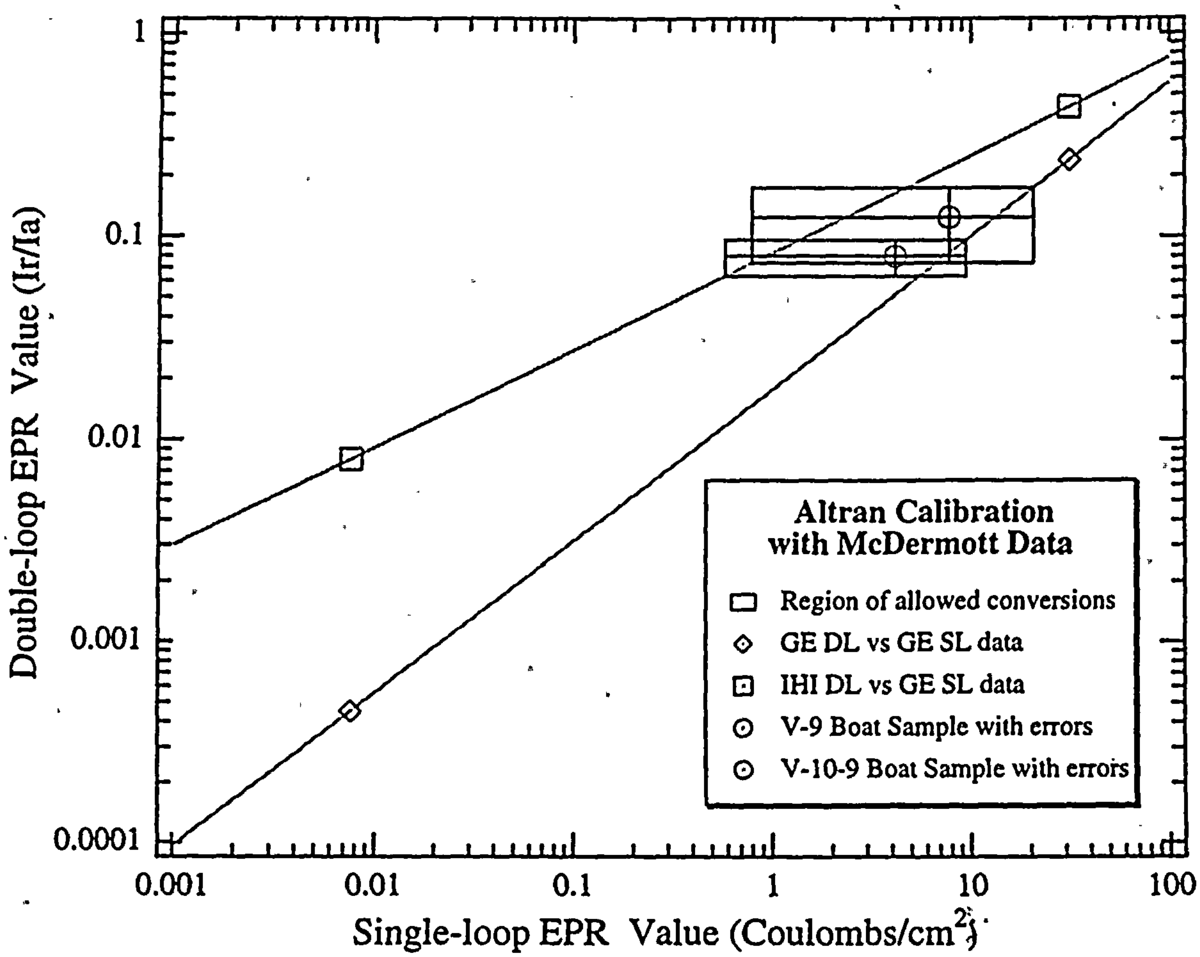

Figure 2: McDermott EPR data with Altran Conversion Method

Finally I will note that even if we restrict ourselves to using DL data which the McDermott Corp. researchers labeled as legitimate, the appropriate error bars should have been presented. In Figure 2 I have correctly plotted that data with errors. The McDermott DL results for V-9 and V-10 are listed in Table 1 of appendix 2 of the Altran report. Note that the data for Run #4 on V-10 was not reproduced in this table. There was no mention why that data was excluded. Oddly the data for Run #5, which was noted to have a problem because *IR* was negative, was included. I reproduce McDermott's data with error in Table 4. The single-loop results are presented in table 5.

Sample	DL EPR Value (Ir/Ia)	SL EPR Value (C/cm^2)	SL EPR Range (C/cm^2)
V-10	0.124 ± 0.050	7.74 (+12.9, -6.95)	20.6 – 0.790
V-9	0.080 ± 0.016	4.15 (+5.23, -3.57)	9.38 – 0.584

Table 5: Single-loop EPR data using McDermott DL data and Altran Conversion Method.

To summarize, lacking a calibration standard for their DL measurements, the Altran researchers have introduced an added factor of uncertainty into the EPR data by allowing a broad range of possible DL-to-SL conversion factors. The researchers did not present the errors for their DL data nor did they propagate these errors through the DL-to-SL conversion. They mistakenly listed the range of SL values for the mean of the DL data as the error in the SL result. Thus their published results do not correctly describe their measured data.

If one properly calculates the error for the DL data and then propagates that error through the DL-to-SL conversion formula, the corrected SL results are a factor two higher than the Altran results. In addition the errors are so large that the range of implied crack growth rates is well beyond the 2.2×10^{-5} inches/hour limit which NMPC claimed to have established.

• Tensile Tests on the Core Shroud Boat Sample Material

Tensile tests were performed on the V-9 and V-10 boat sample material. The yield strength and ultimate strength were measured. These results characterized the material hardness and susceptibility to cracking. The report, prepared by McDermott Technology, is denoted RDD:98:55863-004-000:01.

Overall this report was well done with a thorough documenting of the methodology and the data. I have only three comments regarding the analysis.

1) On the top of page 9 of the report, the author discusses the relationship between the measured yield strength, the bulk yield strength, and the cross sectional geometry. The author should have provided an estimate for the geometric correction factor for the irradiated samples. Certainly the correction is smaller for harder materials, but I doubt the contribution is truly negligible for these size samples.

2) At the bottom of page 9, the authors discuss the effect of radiation on yield strength. This point deserves much greater elaboration. The effect of radiation on metals is a field that has been extensively studied. I find it hard to believe that the author could not find more evidence to support his approximations than three data points from a single source. The data shows that the fluence estimates and the increase in yield strength are not in strong disagreement, but it is not a very rigorous argument.

3) Several factors can change the yield strength in a metal including heat treatment and irradiation. Samples V-9 and V-10 were exposed to different fluence levels and different heat treatment during welding. By themselves, these tensile tests do not allow us to separate the relative contributions to the increased yield strength. We are told that the V-10 sample was taken from the HAZ while the V-9 sample was outside the HAZ, so there should be different contributions to he change in yield strength due to the welding process. The authors need to estimate the change in yield strength that might arise from the welding process. In addition, while the ratio of the fluence at the V-9 sample to the fluence at the V-10 sample is independent of total fluence, the change in yield strength due to fluence is highly nonlinear in pure metals. I don't have ready access to the change in material properties for 304 stainless steel due to high energy neutron flux (nor is that material covered in the report) but I believe that the yield strength rises rapidly with fluence and then tapers off and asymptotically approaches a value below the ultimate strength. This relationship should be exploited in order to provide an upper limit on the fluence seen by the boat samples.

- **Assessment of Crack Growth Rates in NM1 Core Shroud Vertical Welds**

Attachment 5 of the NMPC report is a GE Nuclear Energy technical report (#GE-NE-523-B13-01869-113NP, Rev. 0) assessing the applicable crack growth rate in the NM1 core shroud vertical welds. The report makes several statements which are posed as supporting reasons for establishing a crack growth rate of 2.2×10^{-5} in/hr. I would like to touch on a few of those statements:

1) The authors state "The boat sample evaluations provide clear evidence that the cracks are intergranular stress corrosion cracking similar to that found in piping and in components that have been sensitized conventionally by the welding process." I believe the authors have overstated their conclusions, certainly beyond the level of accuracy of the experimental data. As shown above, the sensitization measurements are far from being "clear", and the authors cannot provide good quantifiable separation between the effects of welding and neutron fluence.

2) The authors state that the main cause for the core shroud cracking is the plant's history of poor water chemistry, but that hypothesis fails in the light of the core shroud cracking in other plants, such as NMPC's NM2 plant, which have maintained excellent water chemistry for the entire life of the plant.

3) At the end of the report the authors list tables with the results from the PLEDGE program which estimates the crack growth rate for several combinations of neutron fluence, stress intensity, and EPR value. Unfortunately, GE has chosen to remove enough of the results so that the reader is unable to estimate values beyond those listed in the tables. In this case I wanted to estimate the crack growth rate for the corrected EPR values for the boat sample and over the range of expected fluences. Although lacking sufficient data to make proper estimates, I feel it is safe to assume, given the trend in the data, that the allowed range of EPR values would not limit the crack growth rate below 5×10^{-5} in/hr, much less below 2.2×10^{-5} in/hr.

The NRC has noted that it has yet to certify PLEDGE as a reliable method for estimating the crack growth rate in reactor core shrouds. Nevertheless the PLEDGE estimates are frequently cited by NMPC and GE as supporting the lower crack growth estimates. Therefore it is important to note that these models will not yield the lower crack growth rate for the allowable range of EPR values.

- **Interpretive Metallurgical Report on Core Shroud Vertical Weld Boat Samples from NMP-1 (Altran Technical Report #97181-TR-02, Revision 0)**

The Altran report summarized the metallurgical studies performed on the boat samples (Altran Technical Report #97181-TR-02, Revision 0). On page 4 of the report the authors comment on the measurements of the depth of the V-9 and V-10 welds, noting that the "(a)ctual data are discrete measurements plotted as if they are continuous", and then later that "(e)ight short crack segments (parallel to the weld) were seen on the outer surface to the left side of weld V-10 as shown in Figure 2." First, the data should have been plotted as discrete points and the error bars for position and depth should have been shown as well. We can extract the error in these crack profile measurements from the NRC SE of NM1 dated 8 May 1997 and my letter to the NRC dated 3 December 1998 which corrected the calculations in the SE of 8 May 1997. The error in position is $\sigma_p = 1.164$ in. and the error in depth is $\sigma_d = 0.108$ in. If we overlay these error bars onto the graphs of the crack profiles we see that a significant portion of the crack depth variation may be

attributed to measurement error. The cracking in weld V-9 is characterized by a single fissure running along the left side of the weld. Weld V-10 has a crack along the right side and several small cracks along the left side. Examining the depth profile data, it appears that the density of measurements is higher for V-9 than V-10. It also appears that the depth measurements on the left side of V-10 were taken only in a few locations and that the researchers did not map the full profile of each small crack. Since V-9 has a single fissure, the crack profile represents the total level of cracking along that weld. On the other hand, the cracking along V-10 is the "sum" of the large fissure and the several small cracks. We could imagine creating a single "effective" crack depth profile for V-10, similar to V-9, but we cannot do so by simply adding the profiles from the two sides of V-10. Crack profiles are not linearly additive. Therefore while the V-9 crack depth profile can be used to represent the extent of cracking along weld V-9, the crack propagation along weld V-10 is not fully represented by either the left or right crack profiles for V-10.

This point is important when we consider the contribution of fluence to the rate of cracking along V-9 and V-10. In the report of NMPC and its consultants, there are several tests of the material properties but none which isolate the contribution of fluence to the level of cracking. The EPR tests are highly inconclusive due to the large measurement errors. The EPR results also depend on the level of heat conditioning in the material during the welding process and the level of Molybdenum in the plate material. These relative contributions of these effects are not established in the report. The tensile tests show that the yield strength has increased, but again we do not know the relative contributions of welding versus irradiation. The surface hardness tests also show an increase in surface hardness which could be attributed to irradiation or to cold working of the metal. Moreover, even if we knew the relative contributions of irradiation to the above mentioned effects, it has not been established that the GE irradiated PLEDGE model correctly represents the contributions from irradiation to the crack growth rate.

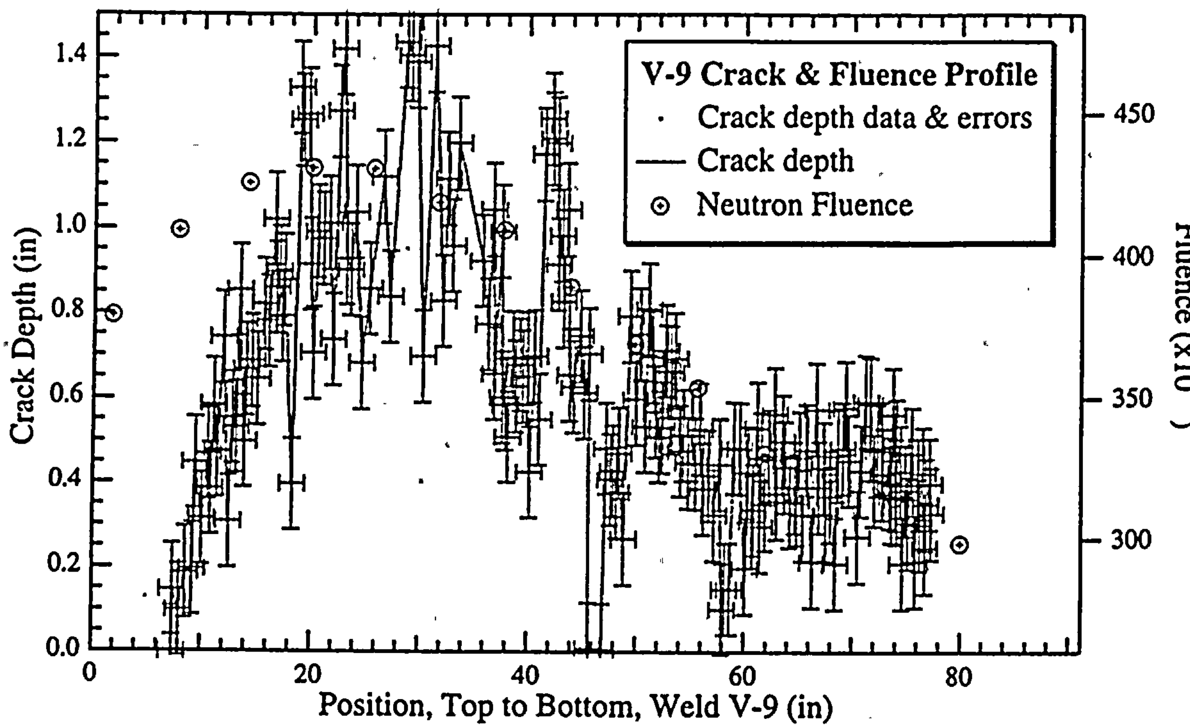

Figure 3: Weld V-9 Crack Profile overlaid with Fluence Profile along V-9

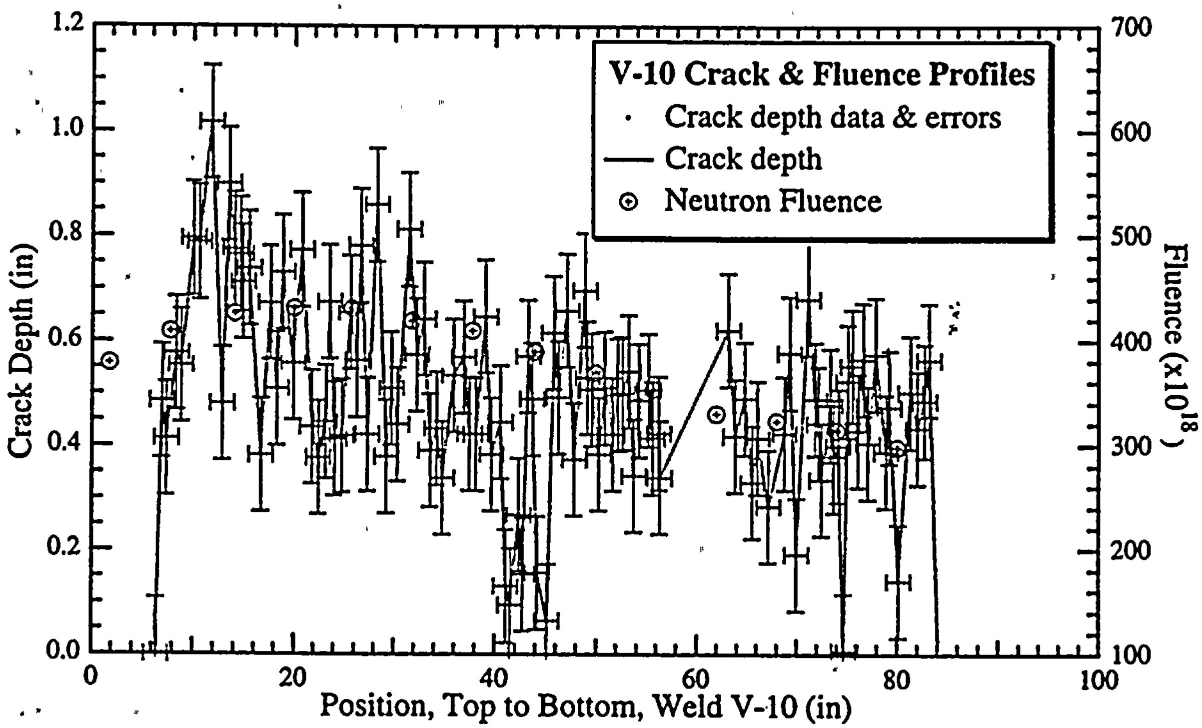

Figure 4: Weld V-10 Crack Profile overlaid with Fluence Profile along V-10.

The best indication of the contribution of fluence to the crack growth rate is to make a direct comparison — to compare the fluence profile with the total "effective" crack depth profile for each weld. In Figure 4 I overlaid the fluence predictions and the crack depth profiles, both taken from the NRC SE. When studying this graph we should bear in mind that the point-to-point fluctuation is largely due to the measurement error. For V-9 we see that the fluence profile and the crack depth profile are well correlated suggesting a causal relationship. For V-10 we do not have a true effective crack depth profile, as explained above. However, noting this limitation, we may still compare the fluence profile with the depth profile for the crack to the right of V-10. In this case we also see significant correlation though not to the same extent as witnessed in V-9. Taken together, these correlations suggest that the fluence is a significant driving force behind the vertical weld cracking. On the final page of the 2 November 1998 SE, the NRC noted the correlation between the fluence profile and the V-9 crack profile. The NRC should have taken the fluence profile and the crack depth data, with errors, and performed a weighted correlation calculation. A significant correlation would have indicated a causal relationship between the fluence and the vertical weld cracking.

In closing let me review the major points of concern raised by your safety evaluation.

1) The NRC should implement a more stringent review of the calculations and studies which contribute or influence the NRC SE. The improper research procedures, lack of error analysis, and overstated conclusions are quite worrisome. In addition the conflict of interest of the authors of industry studies demands a more careful review than it appears the NRC currently performs.

2) Every important measured value should include its error. Without the error, the value holds little weight. Without a published error, the NRC has no way to evaluate the validity of the industry's experimental results. If the NRC does make a judgement without knowledge of the error, it in turn is following poor procedure.

3) The EPR measurements performed by McDermott Corp. and overseen by Altran Corp. is notable for its improper scientific method. The researchers have discarded data without proven justification and in so doing have biased their results and underestimated their error. The DL-to -SL conversion method has introduced an additional uncertainty factor, and the DL errors were not correctly propagated to SL errors. Once corrected the EPR results are a factor 2 higher than those presented by Altran Corp. The NRC made no mention in the SE

(2 Nov 1998) of these serious problems and accepted the EPR test results as presented in the NMPC report.

4) The tensile tests, while performed reasonably well, fell short of performing any analysis which may have separated the relative contributions of fluence and welding on the change in the material yield strength.

5) The fluence profile and the crack depth profiles appear to be reasonably well correlated, suggesting that the neutron fluence is a significant driving force behind the core shroud cracking. The NRC noted this correlation and then dismissed it on the seeming lack of correlation for the V-10 weld. The matter should have been (and should be) investigated fully. In addition, if the neutron irradiation is a strong component of the cracking mechanism, how does this effect the estimates for crack growth rate?

I thank you for your efforts in addressing these concerns. I regret that I was unable to present my analysis to you at an earlier date. I am hopeful that my efforts in reviewing this material will lead to a higher level of safety for everyone.

I am aware that NM1 will be starting its refueling outage in the very near future. I know that during that refueling outage the core shroud is scheduled to be inspected and the cracks along the welds measured. Does there exist a public copy of NMPC's plan for that inspection and, if so, how I might receive a copy? I assume that the NRC will have to issue a SE authorizing the restart. I would like to know how I might get a copy of the NMPC crack inspection results, a copy of NMPC's report to the NRC justifying the restart, and a copy of the NRC's SE as each becomes available. Is there a procedure by which I might review the supporting documents and submit my analysis of that material before the NRC issues its final SE? If so I would like details on how I might participate in that review.

I have several questions regarding the upcoming inspection of the core shroud. Will all the welds, both horizontal and vertical, be inspected for cracks, and will all the cracks be measured? Will the shroud surface outside the HAZ be inspected for cracks and, if found, will those cracks be measured? Will the other internals also be inspected for age degradation and cracking? Will the tie-rod assembly be inspected? Is there any plan to reinforce the vertical welds, especially V-9 and V-10? Does the NRC issue its SE before the plant restarts? Will there be a public comment period to review the inspection data and provide feedback on those results?

Thank you once again for your prompt attention to these matters.

Sincerely,

Steven Penn

Dr. Steven Penn

Cc: David Lochbaum, Union of Concerned Scientists

Nuclear Information and Resource Service