UNITED STATES OF AMERICA
NUCLEAR REGULATORY COMMISSION

ATOMIC SAFETY AND LICENSING BOARD

Before Administrative Judges:
E. Roy Hawkens, Chair
Dr. Paul B. Abramson
Dr. Anthony J. Baratta

In the Matter of:	)	August 17, 2007
AmerGen Energy Company, LLC	)	
	)	Docket No. 50-219
(License Renewal for Oyster Creek Nuclear Generating Station)	)	

AMERGEN'S PRE-FILED REBUTTAL TESTIMONY PART 3 AVAILABLE MARGIN

I. WITNESS BACKGROUND AND CONCLUSIONS

Q. 1: Please provide the Licensing Board with your names and current titles. The Board knows that a description of your current responsibilities, background and professional experience was provided in Parts 1, 2 and 3 of AmerGen's Pre-Filed Direct Testimony on July 20, 2007, so there is no need for you to repeat that information here.

A. 1: (FWP) My name is Frederick W. Polaski. I am the Manager of License Renewal for Exelon.

(DGH) My name is Dr. David Gary Harlow. I am a Professor in the Mechanical Engineering and Mechanics Department at Lehigh University located in Bethlehem, Pennsylvania.

(JA) My name is Julien Abramovici. I am a consultant with Enercon Services, Inc. located in Mt. Arlington, New Jersey, but formerly worked for the Oyster Creek Nuclear Generating Station ("OCNGS").

(PT) My name is Peter Tamburro. I am a Senior Mechanical Engineer in the OCNGS Engineering Department.

(MEM) My name is Martin E. McAllister. I am an American Society of Mechanical Engineers ("ASME") Non-Destructive Examination ("NDE") Level III Inspector at Oyster Creek Nuclear Generating Station ("OCNGS").

Q. 2: Please summarize the purpose of your testimony and overall conclusions.

A. 2: (All) The purpose of our testimony is to respond to the Pre-Filed Direct Testimony of Dr. Rudolf Hausler that discusses available margin and statistical treatment of the ultrasonic testing ("UT") data taken from the drywell shell in the sand bed region. Our overall conclusions, as stated below, are that Dr. Hausler's statistical treatment of the UT data is inappropriate and that Citizens are using the wrong acceptance criteria for buckling.

<u>Internal UT Data Conclusions.</u> For the internal UT grid data – upon which AmerGen determines available margin – Dr. Hausler inexplicably ignores the averages of the data. For example, the average of the 49 UT measurements from grid 19A was 0.800" in 1992. Therefore, 0.800" is deemed to be representative of that 6" x 6" grid. Dr. Hausler, however, throughout his testimony focuses on the

lowest values from the 49 points and inexplicably assumes that those values are representative of the grid. There is no valid scientific support for this approach, which ignores reality. We believe that Dr. Hausler applies a type of "extreme value" statistics which is improper here because he uses extreme value statistics to look at the thinnest single points, whereas buckling is not a phenomenon that is dependent on very local thickness, but instead on the average thickness over a larger area. Thus, the averages of these data, not the thinnest extremes, are representative of each grid.

Dr. Hausler also argues that the internal grid data are not representative of the condition of the drywell shell in the sand bed region, and that the external single-point UT data should be used instead. (Citizens' Exhibit 12, at 3-4.) Dr. Hausler's argument is based on a comparison of internal, external, and trench UT data from Bay 17. (Citizens' Exhibit 12, at 3-4.) Whether on purpose or by error, his underlying calculation ignores an entire grid of 49 UT data points from Bay 17. (Citizens' Exhibit 12, at 3-4.) Dr. Hausler's argument falls apart when those data points are included. In other words, the internal UT data are indeed representative of the condition of the drywell shell in the sand bed region.

External UT Data Conclusions. Dr. Hausler also inappropriately statistically treats the external UT data. These data cannot represent the thickness of the drywell shell. First, there are too few of them for the points to be statistically representative of the shell as a whole. Second, they are biased toward the thin side (*i.e.*, they historically were selected as the thinnest locations). Dr. Hausler, however, ignores the limited number of data points and performs his calculations and computer "contouring" assuming that these external locations were selected at random and, thus, are representative of the condition of the drywell shell in the sand bed region. (Citizens' Exhibit 13, at 5-6, 9-11.)

Finally, Dr. Hausler relies upon an incorrect local buckling criterion. (Citizens' Exhibit 13, at 11-12.) He then improperly applies that criterion and the general buckling criterion to the single-point UT data collected from the exterior surface of the drywell shell to erroneously conclude that the drywell shell thickness currently is not in compliance with the ASME code.

Q. 3: What is your ultimate conclusion?

A. 3: (All) The bounding remaining available margin of the OCNGS drywell shell in the sand bed region for the period of extended operation remains 0.064".

II. BACKGROUND NEEDED TO UNDERSTAND CITIZENS' STATISTICAL ARGUMENTS

Q. 4: Please define the terms (a) "population mean," (b) population variance," (c) "sample mean," and (d) "sample variance" as used in the presented statistical analyses [Board Question 1].

A. 4: (DGH, JA, PT) In order to understand "population mean," you must first understand the term "population." "Population" is the set of all possible outcomes. In the case of the thickness of the drywell shell in the sand bed region, the "population" is a range that could be zero—if there was a hole in the shell—up to approximately 1.154", which is the nominal designed thickness.

(a) For the drywell shell thickness, the "population mean" can only be estimated, not actually measured. The more precise answer is that "population mean," which is symbolized by "μ", is the expected value for the population being considered. For random variables defined on real numbers, the technical definition is as follows:

$$\mu = \int_{-\infty}^{\infty} x f(x)\, dx,$$

where $f(x)$ is the probability density function that characterizes the randomness of the random variable. The "population mean" cannot be determined unless you know the probability of each of the values in the population.

(b) Variance is the amount of scatter that characterizes the randomness in the variable, for example, thickness of the drywell shell. The more precise answer is that "population variance," symbolized by "σ^2", is the expected value of the second moment about the population mean μ for the population being considered. For random variables defined on the real numbers, the technical definition is as follows:

$$\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx,$$

where $f(x)$ is the probability density function that characterizes the randomness of the random variable.

(c) "Sample" is the set of all observations, for example, UT measurements. The "sample mean," symbolized by "$\bar{x}$" or more appropriately the "sample average," is the arithmetical average of the physical measurements made from a population being considered. If the observations are $x_1, x_2, \ldots, x_n$, where n is the sample size or number of measurements, then the technical definition is as follows:

$$\bar{x} = \sum_{k=1}^{n} x_k / n .$$

This is analogous to measuring a limited amount of points over a 6" by 6" area (*i.e.*, 49 points), summing each measured value, and then dividing by the number of measurements that were taken. It is impossible to measure the thickness of the entire surface of the 6" by 6" area, or for that matter, the drywell shell, even by scanning the entire area. However, the more measurements that are taken, the better the sample average will approximate the population mean.

(d) The "sample variance," symbolized by "s^2" is the second arithmetical moment about the sample average $\bar{x}$ for the measurements from a population being considered. If the observations are $x_1, x_2, \ldots, x_n$, as above, where n is the sample size, then the technical definition is as follows:

$$s^2 = \sum_{k=1}^{n} (x_k - \bar{x})^2 /(n-1) .$$

This is analogous to measuring a limited amount of points over a 6" by 6" inch area (*i.e.*, 49 points), summing the square of the difference between each measured value minus the sample average, and then dividing by the number of measurements minus one. As above, it is impossible to measure the thickness of the entire surface of the 6" by 6" area, or for that matter of the drywell shell. However, the more measurements that are taken, the better the sample variance will approximate the population variance.

If you knew the population mean and the population variance for the drywell shell thickness, no measurements would be needed. Because they are not known, however, measurements are needed to estimate them. It should also be noted that the "standard deviation" for either the population σ or sample s is the square root of the variance.

Q. 5: Where does the term "uncertainty" fit into all this?

A. 5: (DGH, JA, PT) "Uncertainty" refers to the level of assurance that a measurement is accurate. Uncertainty is caused by things that are typically outside of your control. For example, the UT technicians are competent and qualified but cannot locate the *exact* measurement location each time; the accuracy of the UT equipment is excellent but still not 100%; and different technicians take the measurements in very slightly different ways.

Q. 6: The Board has asked the following question regarding uncertainty: "The SER lists ten sources of systematic error (SER at 4-53 to 4-55), but AmerGen's direct testimony does not appear to discuss all ten sources (AmerGen's Prefiled Direct Testimony Part 3, Available Margin at 21-23). Estimates and explanations for the all ten sources should be provided, or, if they are insignificant, it should be so stated." Please respond to this question. [Board Question 7]

A. 6: (PT, FWP) We provide each of the ten sources of systematic error (*i.e.*, uncertainty) below, with a brief explanation as to their significance.

a) **UT Instrumentation Uncertainties.** The uncertainty for each UT measurement is approximately +/- 0.010". However, as described below, this uncertainty is not significant for the internal UT grid data once these data are averaged over multiple sampling events.

b) **Actual Drywell Surface Roughness and UT Probe Location Repeatability.** The uncertainty associated with this factor is not quantifiable. It is not significant for the internal UT grid data due to the use of a template that constrains the UT probe and because these data are averaged.

c) **Actual Drywell Surface Roughness and UT Probe Rotation.** The uncertainty associated with this factor is not considered significant because inspection procedures require that NDE personnel performing the UT inspection place the probe in the same orientation.

d) **Temperature Effects.** The uncertainty associated with this factor is not considered significant. Significant temperature differences between inspections may result in a shift in the material thickness. Therefore, the inspection procedure ER-AA-335-004 requires that NDE personnel performing the inspection record the surface temperature and verify that the temperature is within manufacturer tolerances. The procedure also requires that the calibration block be within 25°F of the surface which is being inspected.

e) **Batteries.** The uncertainty associated with this factor is not considered significant. The inspection procedure requires the technician to install new batteries prior to each series of inspections.

f) **NDE Technician.** The uncertainty associated with this factor is not considered significant. Inspection specifications require that personnel conducting UT examinations be qualified in accordance with Exelon Procedure ER-AA-335-004.

g) **Calibration Block.** The uncertainty associated with this factor is not considered significant. Exelon Procedure ER-AA-335-004 requires that the UT technician use only calibration blocks that meet applicable specifications.

h) **Internal Surface Cleanliness** – The uncertainty associated with this factor is not considered significant. The interior UT grid locations are protected by grease between UT inspections. The failure to remove grease from the interior drywell shell surface may have affected the internal UT data measurements collected during the 1996 refueling outage. The UT inspection protocol at that time did not specify the removal of the grease prior to performing UT measurements. Therefore it is possible that the requirement to remove the grease was not communicated to the contractor, and that the contractor who performed the 1996 inspection may have not removed the grease. Tests performed in April and May of 2006 show that the presence of the grease could increase the readings as much as 0.012".

i) **UT Unit Settings**. The uncertainty associated with this factor is not considered significant. It is possible that the ultrasonic unit can be set in a "high gain" setting which may bias the machine into including the external coating as part of the thickness. AmerGen used modern "state of the art" UT units that do not have gain settings during the 2006 refueling outage, and intends to use the same or similar equipment for future inspections.

j) **Identification of the Physical Inspection Location.** The uncertainty associated with this factor is not considered significant. This is not an issue for the internal UT grid locations which are marked on the drywell itself. However, the external UT locations are identified by the area that was prepared (*i.e.*, ground) to make them suitable for UT measurements. The exact location within that prepared area is identified on the UT data sheets by X and Y coordinates from known plate welds, but locating the exact point within the prepared area over the uneven drywell surface is difficult.

Q. 7: Please explain why the systematic error (*i.e.*, uncertainty) is not significant for the internal UT grid data after those data are averaged over multiple sampling events (*i.e.*, 1992, 1994, 1996 and 2006).

A. 7: (DGH, PT, JA) The short answer is that systematic error is negligible for sufficiently large numbers of measurements collected over time. So the more measurements you have, for example, 49 points within a 6" x 6" area, and the more times you collect those measurements, the less significant systematic error becomes.

The more precise answer is that "systematic error" may be considered to be part of the overall uncertainty encountered in measuring the drywell thickness. Although it is not taken into account directly, it is considered indirectly as follows. Let x_k be the thickness measurement at position k, and let ε_k be the error associated with that position. Since ε_k is difficult to quantitatively characterize, the common practice is to assume that it is a normal random variable with mean zero and variance σ^2, which is typically small because the measurement error is minimized by constantly improving the techniques for observations. Thus, the average should be written as

$$\bar{x} = \sum_{k=1}^{n} (x_k + \varepsilon_k)/n$$
$$= \sum_{k=1}^{n} x_k / n + \sum_{k=1}^{n} \varepsilon_k / n,$$

where the last sum is the cumulative error per measurement. The Law of Large Numbers in probability theory implies $\sum_{k=1}^{n} \varepsilon_k / n$ approaches zero as n increases. Thus, the effect of the systematic error is negligible for sufficiently large numbers of measurements. Furthermore, assuming that the errors ε_k, for all k, are statistically independent, then the variance of $\sum_{k=1}^{n} \varepsilon_k / n$ is σ^2/n, which also approaches zero as n increases.

Consequently, the overall effect of systematic error is assumed to be negligible.

Q. 8: Please explain the relationship between "population mean and sample mean" and "population variance and sample variance." [Board Question 2]

A. 8: (DGH, JA, PT) The population mean (μ) and population variance (σ^2) cannot be computed explicitly. They must be estimated, *i.e.*, expressed by a function of the observations $x_1, x_2, \ldots, x_n$ from the population. There are several ways to estimate μ and σ^2; however, the best estimates statistically are the sample average and the sample variance, respectively. In technical jargon,

$$\hat{\mu} = \bar{x} \text{ and } \hat{\sigma}^2 = s^2,$$

where the carat (^) indicates estimate.

Most of the statistical analysis in this discussion focuses on the normal distribution which is completely characterized by two parameters μ and σ^2 which are the mean and variance of the normal distribution. It can be proven, using maximum likelihood estimation, that the best estimates for μ and σ^2 are

$$\hat{\mu} = \bar{x} \text{ and } \hat{\sigma}^2 = (n-1)s^2 / n.$$

It should be noted that if n is sufficiently large, $(n - 1)/n$ is essentially one.

Therefore, for 49 points that are normally distributed, the sample variance is essentially the best estimate for the population variance.

The confidence interval, defined below, for the population mean is a measure of how well the sample average estimates the population mean.

Q. 9: Please define "confidence" as used in the 41 Calc. [Board Question 3]

A. 9: (DGH, JA, PT) "Confidence," symbolized by "$(1 - \alpha)$" is the degree of assurance that a particular statistical statement is correct under specified conditions. The confidence in the data used for the statistical analyses in the 41 Calc is 0.95. However, as stated in A.10 and A.13 below, there is a difference between confidence in the data and a "confidence interval."

Q. 10: Please discuss "confidence interval" and how the interval relates to the sample and population and means and variances. [Board Question 4]

A. 10: (DGH, JA, PT) First, we note that the term "confidence interval" implies that you can statistically treat the data. If the data cannot be statistically treated—such as the external UT data from the drywell shell in the sand bed region—then you cannot determine a confidence interval for that data.

A confidence interval bounds an unknown parameter, such as the population mean μ, so that its probability is the desired level of confidence, $1 - \alpha$. Assuming a normal distribution, the interval is estimated by including the uncertainty and variability in the data. The more uncertainty and variability in the data, the greater is the range of the confidence interval for the parameter.

The technical answer to the question is as follows: Let $f(x; \theta)$ be the probability density function for a population where θ is a parameter in the density function which is unknown. In order to estimate θ observations $x_1, x_2, \ldots, x_n$ must be collected from the population. The statistics L and U, *i.e.*, functions of the samples $x_1, x_2, \ldots, x_n$, determine the $100(1 - \alpha)\%$ confidence interval (L, U) for the parameter θ, if $\Pr\{L \leq \theta \leq U\} \geq 1 - \alpha$. In order to compute the probability $\Pr\{L \leq \theta \leq U\}$ which defines the confidence interval, the probability density for the parameter θ must be known.

By far the usual assumption is that θ is well characterized by a normal distribution. It is for the normal distribution that formulae are given in textbooks for statistics. If any other distribution is operable for a parameter, then the standard textbook formulae are not applicable. Note that all of the internal UT grid data were normally distributed as analyzed in the 41 Calc.

Most often θ is to be taken as the mean μ. For the drywell statistics, this is the primary parameter for which a confidence interval is required. The first task was to establish that the data for drywell thickness were well characterized by a normal distribution for areas defined by the sampling grid. Furthermore, the Central Limit Theorem of probability theory indicates that the sample average can be characterized by a normal distribution for sufficiently large numbers of data. Thus, the confidence interval of concern is

$$\Pr\{L \le \mu \le U\} \ge 1 - \alpha.$$

Again, the population mean μ is not known. It is estimated by the sample average $\bar{x}$. Furthermore, the population variance σ^2 is unknown, and an estimate for it is also needed. Under these conditions the interval estimate for μ is computed by the following statistic:

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}},$$

where the statistic t has the t-distribution with $n - 1$ degrees of freedom. Specific values for the t-distribution are contained in standard statistical tables. The confidence interval for the statistic t is

$$\Pr\{-t_\alpha \le t \le t_\alpha\} \ge 1 - \alpha,$$

where $\pm t_\alpha$ are the two-tail α values, for the upper U and lower L interval values. Substituting for t and doing straightforward algebraic manipulation leads to the confidence interval for population mean μ when the population standard deviation σ is unknown. Thus,

$$\Pr\{\bar{x} - \frac{s t_\alpha}{\sqrt{n}} \le \mu \le \bar{x} + \frac{s t_\alpha}{\sqrt{n}}\} \ge 1 - \alpha,$$

$$\text{and } L = \bar{x} - \frac{st_\alpha}{\sqrt{n}}; U = \bar{x} + \frac{st_\alpha}{\sqrt{n}}.$$

Thus, L and U are the upper and lower confidence intervals.

Q. 11: What is a "standard deviation"?

A. 11: (DGH, JA, PT) A standard deviation is the square root of the variance. Confidence intervals for the mean μ for the normal distribution are determined as a multiple of the sample standard deviation. A standard deviation provides an estimate of the variability of readings within the measured UT grid. It does not provide a reasonable estimate of the uncertainty of the average of that grid, and it can not provide an estimate of the uncertainty or variability of the data outside the grid.

Q. 12: How does a 95% confidence interval relate to "standard deviation"?

A. 12: (DGH, JA, PT) Citizens refer to a 95% confidence interval for the mean μ (for example, in A.11). A 95% confidence interval is almost equal to two standard deviations divided by the square root of the sample size, *i.e.*, the standard error, defined below, higher and lower than the difference in the sample average and the population mean μ, assuming the data are normally distributed. We say *almost* equal, because 1.96 standard errors produce a 95% confidence interval; two standard errors produce a 95.5% confidence interval.

Q. 13: Is there a difference between a "confidence interval" and simply having "confidence" in the data?

A. 13: (DGH, JA, PT) Yes. For example, there is a difference between a 95% confidence interval for the population mean in UT data and the fact that 95% of a particular UT grid's data, when normally distributed, falls within +/- two standard deviations of the average. The latter 95% value is not a confidence interval and has nothing to do with statistical confidence interval estimation for the mean.

Q. 14: What is the student's "t distribution" and what is its significance relative to estimation of the mean thickness? [Board Question 5]

A. 14: (DGH, JA, PT) The significance is that this method is necessary if you are trying to calculate the confidence interval, and if you do not know the population variance (which we do not), you must use the "t test" to compute the confidence interval for the mean. The "student t-distribution" or simply "t-distribution" is the distribution function for the random variable $t = \frac{\bar{x} - \mu}{s / \sqrt{n}}$. It is used primarily for interval estimation of the population mean μ when the data are normally distributed and when the population variance σ^2 is unknown.

Specifically, for the drywell thickness the confidence is 0.95, and the degrees of freedom depend on the sample size. The most frequent sample sizes used in the analyses are grids of 49 and 7 points, so that the corresponding degrees of freedom are 48 and 6, respectively. The values of t_α for these cases are 2.010 and 2.447, respectively.

To illustrate this computation, let $\bar{x}$ = 800 mils, s = 62.4 mils, for 49 observations, then

$$\Pr\{\bar{x} - \frac{st_\alpha}{\sqrt{n}} \le \mu \le \bar{x} + \frac{st_\alpha}{\sqrt{n}}\} \ge 1 - \alpha$$

$$\Pr\{800mils - \frac{(62.4mils)(2.010)}{\sqrt{49}} \le \mu \le 800mils + \frac{(62.4mils)(2.010)}{\sqrt{49}}\} \ge 1 - 0.05$$

$$\Pr\{781.3mils \le \mu \le 818.7mils\} \ge 0.95.$$

Even though the population variance σ^2 is unknown, often investigators will use the two-tail α values z_α from the normal distribution, which are not dependent on sample size. For α equal to 0.05, z_α is 1.96. For practical purposes using a value of 2 is adequate except for small sample sizes where the degrees of freedom have a significant impact on the estimation of the confidence interval.

Q. 15: Is there a more reasonable estimate of the uncertainty of the average of the UT grid data than the standard deviation?

A. 15: (DGH, JA, PT) Yes. A more reasonable estimate (than standard deviation) of the variability of the average of the UT grid data is the "standard error." Assuming a normal distribution, the standard error estimates the variability of the average thickness by accounting for the standard deviation of the distribution *and* the number of samples. The standard error is calculated by dividing the standard deviation by the square root of the number of data points. Thus, the more data you have, the less the variability and the lower the standard error.

Q. 16: Can you provide an example?

A. 16: (DGH, JA, PT) Yes. An understanding of the UT grid averages over time can be developed by reviewing the standard error after the 1992 outage, when corrosion was arrested. At the bounding grid (19A), the 1992, 1994, 1996 and 2006 refueling outage averages (and standard errors) were 0.800" (0.0084"), 0.806"

(0.0099"), 0.815" (0.0096"), and 0.806" (0.0086"), respectively. This illustrates that the average thickness of this 6" by 6" grid has varied between 0.800" and 0.815" in four inspections over about 15 years, and the standard error has varied between 0.0084" and 0.0096".

But you can refine the sample variability even further, assuming no corrosion, through the standard error. AmerGen calculated the sample variability of the average of the data from this grid (through the standard error) over the four sampling events to achieve about +/- 0.005". (Applicant's Exhibit 25)

Q. 17: The Board requested that we provide a table of the location, mean thickness (by date), and the 95% confidence interval of the internal UT grid data. [Board Question 9]

A. 17: (PT, FWP) That table is provided as Applicant's Exhibit 25. Note, however, that AmerGen estimates the 95% confidence interval only for the internal UT grid data, and does so only for the 2006 data because the previous calculations (for 1992, 1994 and 1996) did not include these intervals.

Moreover, as explained above, the 95% confidence interval for each sampling event is not the best estimate of the uncertainty in the data. That is captured by the standard error, which is an estimate of the uncertainty corrected for multiple sampling events (referred to in the Table as the "Grand Standard Error"). Accordingly, AmerGen is also supplying the Grand Standard Error for each grid as calculated using the data from the 1992 through the 2006 refueling outages.

Q. 18: What is the "F statistic" used in the regression model of corrosion and its significance to the corrosion data? [Board Question 6]

A. 18: (DGH, JA, PT) The primary use of the "F statistic" is to test the ratios of two sample variances when it is reasonable to assume that (a) the population variances are equal and (b) the data are normally distributed. Specifically, the F statistic is

$$F = s_1^2 / s_2^2,$$

where s_1 and s_2 are sample standard deviations from the two samples with sample sizes of n_1 and n_2, respectively. Note that there are two degrees of freedom, one for each sample size. The specific values for the F distribution are found in standard statistical tables.

The application of the F test for the drywell is to determine if the variances from two samples of thickness measurements are equal.

Q. 19: Does AmerGen use the "F test," and if so, for what purposes?

A. 19: (PT, DGH, JA) AmerGen has only used the "F test" to evaluate potential corrosion rates. In the 41 Calc., AmerGen used the "F test" in an attempt to identify a corrosion rate. The data, however, failed that test because there were too few inspections (*i.e.*, only 1992, 1994, 1996, and 2006) and the data variability was too large.

Therefore, AmerGen modeled what corrosion rate would be required to pass the "F test" with the existing limited data and large variability. Based on these results, as stated in Applicant's Exhibit 3, page 6-17:

> AmerGen cannot statistically confirm that the sandbed region has a corrosion rate of zero. This is because of the high variance in UT data within each 49-point grid (standard within a range of

deviation 60 to 100 mils), the relatively limited number of data sets that have been taken and the time frame over which data has been collected since the sand was removed in 1992. The high variance in UT data within the grids is a result of the drywell exterior surface roughness caused by corrosion that occurred prior to 1992. However, AmerGen continues to believe that corrosion of the exterior surface of the drywell shell in the sandbed region has been arrested as evidenced by little change in the mean thickness of the 19 monitored (grid) locations and the observed good condition of the epoxy coating during the 2006 inspection.

Q. 20: Explain how systematic error is accounted for in estimating the thickness and corrosion rate. [Board Question 8]

A. 20: (DGH, JA, PT) Systematic error is not accounted for in estimating the thickness of the UT data for the reasons described above in Answer 7. Systematic error equals uncertainty. The ten sources of uncertainty were provided in Answer 6.

Q. 21: Please describe in detail how the term "reasonable assurance" has been defined and applied in the instant case. [Board Question 11]

A. 21: (All) AmerGen has demonstrated reasonable assurance through its aging management program for the drywell shell as a whole. For the UT inspection component of that program, AmerGen has demonstrated that: (a) the average, as an estimate of the mean, of the normally distributed UT data from each internal grid, is thicker than the general buckling criterion, (b) no grouping of external UT data points exceed the local buckling criterion, and (c) no single UT reading from either inside or outside the drywell shell exceeds the pressure criterion. AmerGen does not need to meet its burden to demonstrate reasonable assurance under 10 C.F.R. § 54.29(a) with 95% confidence.

ASME Code, Section XI, Subsection IWE, provides rules for inspection and evaluation of the drywell shell. The Code requires that UT measurements be taken in grids established by the Owner. There is no requirement that the data be evaluated using 95% confidence. The current approach was reviewed by the NRC Staff. The methodology is appropriate for UT data evaluation and is part of the current licensing basis.

Having said that, AmerGen has calculated the 95% confidence interval for the data collected from the internal UT grids in 2006. These intervals are presented in Applicant's Exhibit 25, in response to Board Question 9.

Q. 22: On page 28 of their Initial Statement, Citizens have interpreted the Board's July 11, 2007, Order as requiring AmerGen to demonstrate that "it currently has margin with 95% confidence." Dr. Hausler says the same thing in A.11. Alternatively, Citizens believe they can prevail "either by showing that at 5% confidence the drywell thickness is already below the established acceptance criteria, or that the thickness could go beyond any established margin within four years." Are Citizens correct?

A. 22: (DGH, JA, PT) Citizens are not correct. First, Citizens appear to be confused about what a confidence interval really does. The confidence interval does not provide any information about failure of a component, or compliance with a Code or regulation. Second, Citizens appear to be arguing that AmerGen is required to show that that it has 95% confidence that the drywell shell thickness meets acceptance criteria. (See A.11 "there is less than 95% confidence that the drywell shell currently meets the area acceptance criteria and other acceptance criteria.")

This is inappropriate. AmerGen is primarily interested in the data within a grid which are between ± two sigma about the sample average because this region accounts for 95% of normally distributed data. If there is relatively little scatter in these data, which has been demonstrated elsewhere, so that they are also reasonably close to the sample average, then the sample *average* is the quantity that should be used in comparison to the general buckling criterion. The 5% of the data outside ± two sigma about the sample average pose no threat to buckling; however, these data are considered relative to the pressure criterion.

Q. 23: Is there anything else you would like to add about these statistical issues?

A. 23: (All) Yes. AmerGen's statistical evaluations have been internally and externally reviewed by qualified people, in accordance with objective industry standards. The 41 Calc., for example, was reviewed internally by another senior mechanical engineer, and reviewed externally by consultants. This level of review provides a greater degree of certainty that the data are treated appropriately. Dr. Hausler's statistical treatment of the data does not appear to have been subject to any review, either internal or external, until now. And the many problems we will discuss later in this testimony demonstrate that Dr. Hausler has not treated the data appropriately.

III. DR. HAUSLER USES THE WRONG DATA AND THE WRONG METHODS TO EVALUATE THE INTERNAL UT GRID MEASUREMENTS

Q. 24: Citizens conclude that 0.064" is not the bounding available margin for the OCNGS drywell shell in the sand bed region. How do they arrive at that conclusion?

A. 24: (All) They appear to rely solely upon the opinion of Dr. Hausler, and Dr. Hausler reaches that conclusion only by manipulating the internal and external UT data in a manner that is not statistically appropriate. He also makes some mathematical errors.

Q. 25: Please explain how Dr. Hausler manipulates the data, and why his approach is inappropriate.

A. 25: (All) We will discuss the internal UT grid data first. In order to understand how Dr. Hausler manipulates the data, some background discussion is required. As we previously discussed in Part 3, Answer 12 of AmerGen's Direct Testimony, the internal UT data are collected from nineteen "grids" located throughout all ten drywell bays. Twelve of these grids are six inches square, each consisting of a total of forty-nine individual UT thickness measurement points. The remaining seven grids are rectangular—one inch by seven inches—consisting of a total of seven individual UT points.

As discussed in Part 3, Answer 24, the normally-distributed data from these grids are averaged and compared to the general buckling criterion of 0.736". As discussed in Part 3, Answer 31, the bounding margin of the drywell shell in the sand bed region of 0.064" is based on a 49-point grid in Bay 19 (19A), which had a general average thickness in 1992 of 0.800".

For the internal UT grid data – upon which AmerGen determines available margin – Dr. Hausler inexplicably ignores the averages of the data.

Q. 26: Can you provide some examples?

A. 26: (All) Yes. The average of the 49 UT measurements from grid 19A in 1992 was 0.800." The averages from this UT grid have varied little over time: 0.800" (1992), 0.806" (1994), 0.815" (1996) and 0.807" (2006). As part of the license renewal review process, AmerGen conservatively reported the smallest of these four values (0.800") to the Advisory Committee on Reactor Safeguards (ACRS) to document the minimum available margin in the sand bed region (*i.e.*, 0.800" - 0.736" = 0.064"). (Applicants' Exhibit 3, page 6-2)

Q. 27: Do Citizens agree?

A. 27: (All) No. Citizens claim that the remaining margin for buckling should not be 0.064" but rather 0.034". (Dr. Hausler Answer 16; Citizens Initial Statement at 2). They claim that AmerGen must subtract 0.030" from the measured average of 0.800" in grid 19A (0.064" – 0.030" = 0.034") in order for the average to be compared to the general buckling criterion (*i.e.*, 0.736"). Citizens derive the 0.034" value from an AmerGen response to an NRC Information Request in which AmerGen agreed to take action if the future average of any of the internal grid data collected during an outage was +/- 0.021" different than previous readings. (See Citizens' Direct Answer 16; Citizens' Initial Statement at 11 citing Ex. 10 at 2 and SER at 3-121). This 0.021" value was based on the standard deviation of internal UT data of 0.011" plus uncertainty associated with instrument accuracy of 0.010".

But Citizens believe this value is too low. They claim that 0.011" is based on only one standard deviation and that AmerGen is required to achieve two standard deviations (which, as explained above approximately equals 95% of the distribution for normally distributed data). Citizens conclude that the uncertainty should be approximately 0.030". Dr. Hausler's testimony does not show how he derived that value. We can only assume that Citizens derived this uncertainty as follows, (which would be the proper way to derive the uncertainty): assuming that the randomness in thickness and the measurement error are independent, then the overall standard deviation is $\sqrt{(0.011in)^2 + (0.01in)^2} = 0.0149in.$ Two standard deviations would be 0.0297", which Citizens appear to have rounded up to 0.030". To determine the lower limit of the 95% interval for the data, they argue that AmerGen must subtract 0.030" from the available margin of 0.064", thus concluding that only 0.034" remain.

Q. 28: What are your concerns with how Dr. Hausler manipulated these data?

A. 28: (All) There are several problems with Dr. Hausler's manipulation of the data. First, Citizens miss the point of AmerGen's response to the NRC. AmerGen was identifying an action limit. If AmerGen had selected two standard deviations as Citizens suggest, then it would not take action until the difference in the average of data was approximately +/- 0.030". For an action limit, however, it is appropriate and conservative to assume only one standard deviation. Again, Citizens demonstrate that they do not understand basic information relevant to AmerGen's Aging Management Program.

Second, the actual standard error for grid 19A over time is about 0.005", not 0.030". The standard error for the grid 19A data is about 0.010" *each time* this 49-point grid was measured. (Applicant's Exhibit 25.) But AmerGen has four data sets to work with. If we assume no corrosion, then we can combine the four data sets for 1992, 1994, 1996 and 2006, which results in a standard error of about 0.005". Accordingly, the variability in the grid 19A data is an order of magnitude lower than cited by the Citizens (*i.e.*, 0.005" vs. 0.030"). That is no surprise, since the uncertainty that Citizens cite was taken out of context in the first place.

Q. 29: Doesn't Citizens' method ignore thicker metal that AmerGen has actually measured?

A. 29: (All) Yes. Subtracting 0.030" from the calculated grid average thickness ignores data. For example, the bounding grid (19A) had an average thickness of 0.800" in 1992. If you subtract 0.030" and conclude that the average is 0.770", then review of the 1992 data (41 Calc., Appendix 10, page 6) shows that Dr. Harlow ignores 32 of the 45 UT valid readings from that grid (because 32 were greater than 0.770"). (Four of the readings in 19A are located over a newer metal plug and are not considered valid for calculating the grid average).

The best confidence for the thickness is from the internal UT data. More specifically, it is the repetitive and consistent results for the internal grids in 1992, 1994, 1996 and 2006, and the known standard error which is an order of magnitude lower than that irresponsibly identified by Citizens.

Finally, the ASME Code and acceptance criteria do not require AmerGen to bound the condition of the drywell shell with 95% confidence. AmerGen has to determine a reasonable and conservative measure of the drywell and compare it to the Code-based criteria. By assuming that the bounding available margin is uniformly 0.800" thick, AmerGen has demonstrated that it has developed a conservative measure of the actual condition.

Q. 30: Does AmerGen ignore the lowest readings?

A. 30: (All) No. Each single point within the grid was compared with the pressure criterion to assure that it surpassed that test.

Q. 31: Is there anything else you would like to add before we move on to the topic of whether the internal UT data are representative of the drywell shell?

A. 31: (DGH) Yes. On page 7 of his April 25, 2007 memorandum, Dr. Hausler states that "if an average of ten measurements over a specific area results in a thickness of 0.750 inches with a variability (standard deviation) for the average of 0.03 inches, the lower 95% confidence limit for this average would be 0.690 (0.75 – 0.06)." In other words, Dr. Hausler concludes that the 95% confidence interval would be +/- 0.060".

I have attempted to replicate this value and can only do so if, within basic statistical equations, I fail to divide the standard deviation by the square root of $n = 10$. If Dr. Hausler had calculated the statistical equation properly, then the 95% confidence interval for the difference between the sample average and the population mean would have been approximately +/- 0.019", not 0.060". This means that the confidence interval in Dr. Hausler's example is much tighter than Dr. Hausler states.

IV. THE INTERNAL UT DATA ARE REPRESENTATIVE OF THE BOUNDING DRYWELL SHELL CONDITION IN THE SAND BED REGION

Q. 32: Dr. Hausler spends much of his April 25, 2007 memorandum alleging that the internal grid data are not representative of the condition of the drywell shell in the sand bed region, and that the external single-point UT data should be used instead. He compares the trench, internal grid, and external point data from Bay 17 to support his allegation. What is your response to that allegation?

A. 32: (All) Whether on purpose or by error, Dr. Hausler's underlying calculations ignore an entire grid of 49 UT data points from Bay 17. Dr. Hausler's argument falls apart when those data points are included. In other words, Dr. Hausler reaches his conclusion by conveniently ignoring data that contradict his position. Moreover, it is the omitted data that AmerGen relies upon for purposes of calculating the available margin in Bay 17. Accordingly, Dr. Hausler's calculations do nothing to undermine the fact that the internal UT data are indeed representative of the bounding condition of the drywell shell in the sand bed region.

Dr. Hausler's conclusion on page 4 of his April 25, 2007 memorandum (Citizens' Exhibit 12) states that "only the trench measurements and outside measurements come close to represent [sic] the most severe corrosion at the highest elevations." Dr. Hausler also concludes that the internal data are not representative of the worst corrosion in the sand bed. (Citizens' Exhibit 12, at 3-4.) Dr. Hausler's conclusion is based on evaluation of the data as presented in figures 3 and 4 on pages 15 and 16 of his memorandum. The figures attempt to show the relationship between the internal Bay 17 thickness data, the external Bay 17 data points of which there were only 10 points, and the Bay 17 trench data. All of these data were collected during the 2006 refueling outage.

Q. 33: What are the data that Dr. Hausler ignored that contradict his position?

A. 33: (PT FP) AmerGen routinely monitors only two internal grids that are entirely within Bay 17: 17A and 17D. 17A had a 2006 average thickness of 1.015". 17D had a 2006 average thickness of 0.818". Dr. Hausler uses the data from the 17A grid, but ignores the data from 17D.

Q. 34: What grid from Bay 17 does AmerGen use for license renewal?

A. 34: (PT FP) Oyster Creek considers grid 17D—not 17A—as the representative thickness value of the worst corrosion for Bay 17, and has used the average from that grid for purposes of license renewal. For example, the following values have been reported to the NRC and the ACRS as part of the license renewal process for grid 17D: 1992 – 0.817", 1994 – 0.810", 1996 – 0.848", and 2006 – 0.818" (page 94 of the January 18, 2007 ACRS Presentation – Applicant's Exhibit 26. The 1994 value of 0.810" was used in the ACRS presentation to document 0.074" of margin in Bay 17 (page 95 of the January 18, 2007 ACRS Presentation). It is also shown in Applicant's Exhibit 3 at 6-2 & Table 18. That value was achieved by subtracting the 0.736" general buckling criterion from 0.810".

Therefore, using Dr. Hausler's methodology and grid 17D supports the conclusion that this internal grid is representative of the worst corrosion in Bay 17. This should not be a surprise since the internal grids were originally selected based on a much more extensive set of UT inspections in the mid 1980's which identified the thinnest areas.

Q. 35: Before we move on to discuss the external UT data, there is one other issue that Citizens raise regarding the uncertainty of the internal UT data. Citizens claim that AmerGen uses an uncertainty for the internal UT data of 0.020", and that AmerGen "subtracted 0.020 inches before it compared the mean to the acceptance criterion." (Citizens' Initial Statement at 13.) Citizens cite to AmerGen's Exhibit 19, page 8, for support. Does AmerGen subtract 0.020" from the mean/average of the internal UT grids before comparing the mean to the general buckling criterion?

A. 35: (PT, FP) No. The document that Citizens rely upon (Applicant's Exhibit 19.) is Technical Evaluation AR A2152754 E09, which documented AmerGen's *preliminary* evaluation of the UT data collected in 2006 from the *internal* surface of the drywell shell in the sand bed region. The purpose of that Technical Evaluation was not to support license renewal. Rather, the Technical Evaluation documented why there was adequate margin of the drywell shell in the sand bed region to operate until the next refueling cycle in 2008, to support exiting the 2006 refueling outage.

Q. 36: Is this Technical Evaluation conservative in nature?

A. 36: (PT, FP) Yes. The Technical Evaluation reviewed the internal UT grid data as well as data collected from the two internal trenches. It was a preliminary analysis because we had not at that time had the opportunity to perform statistical analyses of those data. AmerGen, therefore, used extremely conservative factors, including an uncertainty of +/- 0.020", for its preliminary evaluation. Systematic error (*i.e.*, uncertainty) is not accounted for in estimating the final thickness of the UT data for the reasons described above in Answer 7.

V. DR. HAUSLER USES THE WRONG DATA AND THE WRONG METHODS TO EVALUATE THE EXTERNAL UT GRID MEASUREMENTS

Q. 37: Does AmerGen statistically treat external UT data for purposes of demonstrating compliance with the acceptance criteria?

A. 37: (All) No. As we testified in Direct Part 3 Answer 27, AmerGen does not statistically treat the external UT data for purposes of demonstrating compliance with the acceptance criteria. Rather, the raw UT data are compared against the relevant acceptance criteria without any statistical treatment.

Q. 38: Why?

A. 38: (All) Because AmerGen does not use the external UT data points to determine margin. AmerGen only uses that data to demonstrate compliance with the ASME Code. As stated in Part 3, A.29, the single-point UT measurements can tell you that you meet the applicable ASME Code, but not by how much. This is the case because there are an insufficient number of UT measurements over large areas to evaluate a representative average thickness over each area. So Citizens are performing statistical analyses on the external UT data that AmerGen does not perform.

Q. 39: Citizens claim in their response to AmerGen's Motion in Limine, however, that external UT data have in the past been used to estimate available margin.

Citizens cite to Applicant's Exhibit 17, p. 7, which is the original 24 Calc performed in 1993. What is your response to this allegation?

A. 39: (PT, FWP, JA) Citizens are taking that discussion out of context. The top of page 7 confirms that the external UT locations inspection "focused on the thinnest areas of the drywell . . . [thus] the inspection did not attempt to define a shell thickness suitable for structural evaluation." You cannot calculate available margin from a buckling perspective using biased thin points. Second, the evaluation *assumed* a *uniform thickness* of 0.800" for purposes of evaluation against the general buckling criterion. As stated on page 8, however, "In reality, the remainder of the shell is much thicker than 0.800" inches." This external UT data provide useful information that can help you determine that you meet the applicable ASME Code, but they cannot tell you by how much.

Q. 40: Please explain how Dr. Hausler manipulates the external UT data, and why it is inappropriate to do so.

A. 40: (All) As we will demonstrate below, Dr. Hausler statistically treats the external UT data in an inappropriate manner. These data cannot represent the average thickness of the drywell shell because there are too few of them and they are biased toward the thin side (*i.e.*, they historically were selected as the thinnest locations). Dr. Hausler, however, ignores the limited number of external data points and performs his calculations and computer "contouring" assuming that these external locations were selected at random and, thus, are representative of the condition of the drywell shell in the sand bed region. This is an improper assumption which necessarily leads to inappropriate conclusions. (Note that Dr.

Hausler does not appear to account for the UT thickness measurements from internal grids that overlap his contour map area. These are actual measurements that, if considered, would demonstrates that he has significantly underestimated the thickness of the shell).

We can best demonstrate Dr. Hausler's inappropriate techniques through an analogy. If you wanted to know the average weight of people walking along 5th Avenue in New York City, then you would make an inference that if you weighed enough people randomly from that street that their weights would be representative of all the people on that street (*i.e.*, you would have a statistically representative sample). You would not want to select only ten people (too few) or people who biased the sample population by, for example, purposefully selecting those who looked thin. You would then determine if you had a normal distribution of the individuals' weights. With a normal distribution, you would then calculate the average weight, which would be representative of the people on that street. You could then calculate the 95% confidence interval of those weights.

Dr. Hausler glosses over the fact that there are not enough UT measurements to statistically treat the external data in the first instance. He acknowledges there are not enough data when he states that "the paucity of data, particularly in the heavily corroded Bays makes definite conclusions very difficult and an assessment of the extent of the corroded areas somewhat intuitive," (July 18 memorandum at 2). We believe he goes beyond intuition, to speculation when he nevertheless statistically treats those data.

Q. 41: Are there any other reasons why Dr. Hausler is wrong?

A. 41: (All) Yes. Dr. Hausler also acknowledges, but then ignores the fact that the external UT data were selected because they were determined to be the thinnest points. For example, Citizens state on page 14 of their Initial Statement that "the best approach . . . is to regard the external readings as representative, even though they might actually be biased to the thin side by their method of selection." Dr. Hausler's rationale for this statement appears to be his April 25, 2007 memorandum on page 6: "I believe that when assessing the extent of severe corrosion, reviewers should assume that the measured points connect unless other measurements show this not to be the case."

Dr. Hausler then averages these thinnest points and improperly identifies a 95% confidence interval. He then focuses on the thinnest of these readings. Not surprisingly, he declares that the drywell shell, in some cases, already has exceeded the general and local buckling criteria.

Using our analogy, what Dr. Hausler does is similar to biasing the sample population from 5th Avenue by selecting too few people, and only those who are waif-like. Needless to say, it is statistically inappropriate to average biased thin measurements and treat them as representative of the population, whether it is the weight of people or the thickness of the drywell shell. These data simply are not representative of the average since the shell between these UT locations is thicker. It is similarly statistically inappropriate to take the thinnest of these biased thin areas (*i.e.*, the lower 2.5% of this biased sample) and claim that these extreme values could be representative of the average. Using our analogy, such statistics would lead to the absurd conclusion that only people with anorexic qualities walk on 5th Avenue.

Dr. Hausler is confusing extreme value behavior with averaging. If your sample population is biased thin, then the way to evaluate the data is through extreme value statistics. You would not use an averaging technique because averaging implies a normal distribution. Dr. Hausler argues that the average of the thinnest points is representative of the whole drywell shell, but it can only be representative of the extreme values.

Q. 42: What is the basis for your opinion that the external UT locations were selected because they were the thinnest locations?

A. 42: (JA, PT) During the 1992 refueling outage, OCNGS did not identify UT measurement points on the exterior of the drywell shell to identify the average thickness. Rather, it specifically looked for the thinnest areas. This is documented in Applicant's Exhibit 27 (TDR1108):

> The corroded vessel shell resembled a cratered golf ball surface. The areas where the heaviest corrosion had taken place appeared obvious from a visual inspection since the inside shell wall was relatively uniform. The GPUN metallurgist (S. Saha) identified on a sketch, areas to be prepared for UT readings. At a later time he reviewed the surface preparation and thickness data and identified additional locations to ensure that the thinnest areas were surveyed. [page 15]
>
> It was reasoned that since the inside surface of the vessel shell is smooth and not corroded, any thin area on the outer surface should represent the minimum thickness in that region. It was further reasoned that if six to twelve scattered spots, located in the area of worst corrosion, are ground smooth and the thickness of each spot is measured by UT method we will have a high level of confidence that we have identified the thinnest shell thickness for a bay. This approach is conservative since, (a) we are forcing

> a statistical bias in choosing only the thinnest areas and (b) grinding of the selected spots to obtain a flat surface for reliable UT readings will remove additional good metal. [page 16]

This is also discussed in other documents, including, Applicant's Exhibit 12 on p. 14, Applicant's Exhibit 16 on p.4, and Applicant's Exhibit 17 on page 7.

In addition, Dr. Hausler's own analysis has independently confirmed that these external points are biased thin. In Citizens' Exhibit 12 on page 4, Dr. Hausler states that "the average outside measurements are significantly lower at comparable elevations [than the interior measurements]. This is probably because the choice of location for the external measurements was deliberately biased towards thin spots."

The fact that the external UT locations are biased towards the thinnest locations is also demonstrated by comparison of those data to the data taken from the internal UT grids. Some of the external UT locations coincide with internal grid locations, as generally shown on the comprehensive map of all 2006 UT inspection results that AmerGen provided to the ACRS for a public meeting in February 2007. The map is located on Page 14 of AmerGen's presentation, which is attached as Applicant's Exhibit 28. We will refer to this map as the "2006 map" as we next discuss three illustrative examples.

Two of the thinnest external readings in Bay 19 (points 9 and 10) were 0.728" and 0.736", respectively, in 2006. The 2006 map shows that these points are located within inches of internal grids 19A and 19C, which had averages thicknesses of 0.807" and 0.824", respectively, in 2006.

Deleted: Three

Deleted: ,

Deleted: and 11

Deleted: ,

Deleted: , and 0.712"

Deleted: B

Deleted: 48

One of the thinnest external readings from Bay 13 was point 15 at 0.666" in 2006. The 2006 map shows that this external point is located within inches of internal grid 13C, which averaged 1.142" in 2006.

One of the thinnest readings in Bay 17 (point 2) was 0.663" in 2006. This point is located within inches of internal grid 17A, in which the top half of the grid averaged 1.112" in 2006 and the bottom half of the grid averaged 0.935" in 2006.

One of the thinnest readings in Bay 1 in 2006 (point 5) was 0.680". This point is located within inches of internal grid 1D, which had an average thickness of 1.122" in 2006.

These data, from multiple bays, unambiguously demonstrate that the external locations are biased thin compared to their surroundings. To statistically treat these data as representative of the drywell shell in the sand bed region is, therefore, inappropriate.

Q. 43: But on Page 10 of their Initial Statement, Citizens discuss the measurements taken in 2006 from 0.25" around the coordinates for certain external UT points in Bays 7, 15, 17, and 19. They state that those measurements are thinner than the designated external UT data point. Are Citizens correct that these external measurement locations are, therefore, not the thinnest?

A. 43: (FP, PT, JA) No, they are not correct. They confuse the measured "points" with the "ground UT locations." The external measurement "point" is located within a 2-inch diameter area that was ground smooth during the 1992 refueling outage to allow for the UT probe to sit flat against the shell. Examples of these ground

Deleted: The
Deleted: of all the
Deleted: was
Deleted: (
Deleted: 7)
Deleted: 02
Deleted: D
Deleted: in
Deleted: the top half of the grid
Deleted: 047
Deleted: and the bottom half of the grid averaged 0.904" in 2006
Deleted:
Deleted: The thinnest reading in Bay 11 (point 1) was 0.700" in 2006. This point is located within inches of internal grid 11A, which has an average thickness of 0.822" in 2006.¶
Deleted: T
Deleted: 3
Deleted: 65

locations are shown in Applicant's Exhibits 29, which are two presentation slides from AmerGen's meeting with the ACRS in January 2007. These *locations* were selected because they were the thinnest locations in the sand bed region for each bay.

The coordinates on the UT data sheets direct the UT technician to a spot within a specific ground location. But that specific spot is not itself marked and UT data from that location is, therefore, not precisely reproducible from sampling event to sampling event. These nuances, however, in no way undermine that these ground *locations* are the thinnest locations in each bay. Indeed, the fact that UT readings 0.25" around the center reading were lower, further supports that these ground areas are the thinnest locations.

Q. 44: Did AmerGen ignore these thinner UT readings 0.25" around the center reading if they were lower?

A. 44: (PT) No. When I performed my evaluation of the external UT data, I used the thinnest UT value from each of the ground areas measured in 2006. This is shown in Rev. 2 of the 24 Calc. for data points from Bays 7, 15, 17, and 19.

Q. 45: Is there anything else wrong with Dr. Hausler's evaluation of the external UT data?

A. 45: (All) Yes. Dr. Hausler relies upon an incorrect local buckling criterion (*e.g.*, A.13). He compares the external UT data to a criterion consisting of a one square foot area with a thickness of 0.636", without any transition back to 0.736". The actual criterion—AmerGen's local buckling criterion—has a thickness of 0.536" in a tray configuration, with a transition back to 0.736". That criterion is shown on AmerGen's Exhibit 11. Using the wrong criterion compounds his errors, and affects his ultimate conclusions about whether the drywell shell thickness meets the ASME Code.

Q. 46: Dr. Hausler argues that there are severely corroded areas that are shaped "like long grooves" or are irregular in shape, that call into question AmerGen's use of a square-shaped, local buckling criterion. (A. 24) What is your response to this argument?

A. 46: (All) Dr. Hausler is wrong. This argument can only be based on Dr. Hausler's improper statistical treatment of the external UT data, and his assumption that "the measured points connect unless other measurements show this not to be the case." (April 25 memorandum, page 6) The bath tub ring is irregular in shape, but the corrosion in that ring is only relevant to buckling if the resulting thickness is less than 0.736". And AmerGen has evaluated as acceptable those locations within the bath tub ring with UT readings that are less than 0.736". Additionally, the thinnest average grid reading taken from inside the drywell is in the bath tub ring, supporting our position that there is adequate margin to buckling.

A. <u>Uncertainty in External UT Data</u>

Q. 47: Dr. Hausler claims that the uncertainty of each external point is approximately +/- 0.090". (A.15) The basis for this claim is from Section IV (page 3) and Section VII (pages 8 and 9) of his July 18, 2007 memorandum (Citizens' Exhibit 13). Is Dr. Hausler correct?

A. 47: (All) No. In order to understand why Dr. Hausler is wrong, you first need to understand how he derived his level of uncertainty. Dr. Hausler derives 0.090" as follows. He identifies locations in Bays 5, 15, and 19 where measurements were taken during the 2006 refueling outage in a 0.25"-diameter area around the designated external measurement point. (On Page 9 of his July 18 memorandum, Dr. Hausler refers to these measurement locations as "identical coordinates," when in fact, they were taken in an area 0.25" around the specified coordinate.)

He assumes that the external data are representative of the thickness of the shell in these three bays, so he averages the data from these locations. (See the last column of the table on page 9 of his July 18 memorandum.) He then assumes the external data are normally distributed, and calculates the standard deviations for each bay, arriving at 0.033", 0.050" and 0.043" for the points in Bays 5, 15, and 19, respectively. (Citizens' Exhibit 13, at 3.) He then inexplicably "pools" these three values to arrive at 0.045", which he argues applies as a representative thickness for all areas in all of the bays. He then doubles that value (0.045" x 2) to account for the two standard deviations required to identify the 95% confidence interval.

Q. 48: What is wrong with this use of the data?

A. 48: (All) In arriving at 0.090", Dr. Hausler completely ignores reality and proper statistical techniques. As discussed above, he ignores that the external data are biased thin and that the locations were deliberately chosen to be the thinnest locations in each bay; that the data are not normally distributed (as shown by Kurtosis of the three data sets); and that there are not enough data to establish a representative sample population of these very large areas. As to the last point, there are only eight external points in Bays 5 and 15, and nine in Bay 19, to represent three areas *each of which* is about 3.5 feet by 15 feet wide. He also conveniently ignores the Bay 7 standard deviation he calculates on the same table (page 9) which would have reduced his number from 0.090" to 0.075".

Dr. Hausler then assumes this 0.090" value can be applied globally to any one reading or set of readings throughout the sand bed region of the drywell shell. This is unsupported and suggests that Dr. Hausler's testimony in this area should be given little, if any, weight.

Using the analogy of people on 5th Avenue, what Dr. Hausler does by pooling these thin points is akin to selecting the thin-looking people from 1st Avenue, 3rd Avenue, and 5th Avenue, and concluding that everyone in New York City is underweight.

Q. 49: What do you mean by the use of the term "kurtosis" in your previous answer?

A. 49: (PT, DGH) For ease of discussion here, we have rescaled Kurtosis, so that it equals zero for a normal distribution. Distributions that are greater or less than zero are not normally distributed.

For Bay 5, the 2006 external points were 0.948, 0.955, 0.989, 0.948, 0.88, 0.981, 0.974, and 1.007 with a calculated Kurtosis of 2.43.

For Bay 15, the 2006 external points were 0.711, 0.777, 0.935, 0.791, 0.817, 0.715, 0.805, and 0.76, with a calculated Kurtosis of 1.65.

For Bay 19, the 2006 external points were 0.867, 0.85, 0.894, 0.883, 0.82, 0.721, 0.728, 0.736, and 0.721 with a calculated Kurtosis of -2.2.

B. Evaluation Thickness

Q. 50: On pages 6 and 7 of his July 18 memorandum, Dr. Hausler raises many allegations about the "Evaluation Thickness," which is discussed in the various revisions of the 24 Calc. He concludes on page 7 that, "We can, therefore, not accept the evaluation done by AmerGen using the 'evaluation thickness.'" Please explain what the "Evaluation Thickness" is and its use.

A. 50: (FP, PT) As explained on pages 17-19 of Rev 2 of the 24 Calc. (AmerGen's Exhibit 16), the Evaluation Thickness is a representative average thickness in an area of 2" in diameter surrounding the external points that were less that 0.736" as measured by UT in 1992. During the 1992 refueling outage, micrometer readings were taken in a 2" diameter area around each external UT point that measured less than 0.736" (*i.e.*, about 20 points). This uniform depth was generated from actual measurements which had surface roughness variability of 0.200" from the micrometer readings for the two thinnest points in Bay 13 (see 24 Calc, Rev 2, p. 19). The Evaluation Thickness method is the UT thickness reading, plus the average depth of the area relative to its surroundings, minus 0.200" (referred to in the Evaluation Thickness method as "T roughness").

Dr. Hausler assumes the Evaluation Thickness method is to "correct for the fact that due to the roughness the UT probe may not have 'coupled' well with the metal surface and therefore detect less metal (thinner wall) than was actually there." (July 18 memorandum, page 7). He also assumes that "T-roughness" was to correct for roughness under the UT probe, and that it therefore should *not* have been used in 2006 when the epoxy coating would have created a smooth surface for the probe.

Q. 51: Is Dr. Hausler correct?

A. 51: (PT, FP) Dr. Hausler is wrong. The purpose of the method—as stated in Applicant's Exhibit 16— is to evaluate a 2-inch diameter area around the UT location, and estimate the average thickness of that 2-inch diameter area, not to account for the ability of the UT probe to properly couple. The purpose of "T-roughness" is to account for the roughness under the micrometer's straight edge, not roughness under the probe.

In addition, Dr. Hausler does not understand the implication of his argument. If AmerGen had *not* used T-roughness in 2006, as Dr. Hausler suggests, then the value would have been *thicker* by 0.200", which would not have been conservative.

Q. 52: On page 7 of his July 18 memorandum, Dr. Hausler quotes a document that you, Mr. Tamburro, wrote in 2006, suggesting that the Evaluation Thickness ought not to be used. Can you please respond to this?

A. 52: (PT). Yes. I did indeed submit a document to the OCNGS corrective action system (Citizens Exhibit 3), raising a concern with Rev 0 of the 24 Calc. (Applicant's Exhibit 17). However, my concern was limited to inadequate documentation. I identified approximately 11 items that required additional documentation in that calculation. All of the items were related to *documentation* of assumptions, methods, and data. This included an item about documentation of the methodology and justification for the Evaluation Thickness method. In other words, the deficiencies could be resolved with additional documentation. My concern about the Evaluation Thickness method was properly and thoroughly resolved through AmerGen's corrective action process and pages 17-19 of Rev 2 of the 24 Calc. document the resolution of the deficiency that I had identified.

I believe the method is appropriate to use, and I employed that method to evaluate data from the 2006 refueling outage.

VI. AMERGEN'S EVALUATION OF THE LOCAL BUCKLING CRITERION IN THE 24 CALC. IS APPROPRIATE

Q. 53: Dr. Hausler calls into question AmerGen's evaluation of the external UT data in Rev. 2 of the 24 Calc by challenging AmerGen's assumptions about the size of the historically corroded areas. (A. 23) Please respond to this.

A. 53: (PT) I performed the evaluations that are documented in Rev. 2 of the 24 Calc., and am very familiar with the prior revisions. For Rev. 1 (which he calls the second revision), he states that AmerGen "assumed, contrary to the visual observation, that all the severely areas measured were less than 2" in diameter." Dr. Hausler does not cite a specific page in the calculation so I cannot determine what precisely he is referring to. However, he is not correct. AmerGen identified the thinnest areas within the severely corroded areas, and then ground the metal around those points for a 2" diameter.

Dr. Hausler also states that, for Rev. 2 (which he calls the third revision), "AmerGen has taken an approach of drawing squares by eye on plots of the external data points." (A.23). On page 5 of his July 17 memorandum, he states that this was a "one-dimensional analysis." These too are incorrect. I did not draw squares by "eye on plots." I entered each of the external UT points using the x and y coordinates provided on the UT data sheets into Microsoft Excel. I then used Excel to create a 36" x 36" square, to represent the boundaries of the tray configuration that comprises the local buckling criterion. For points that measured less than 0.736" in 2006, I used Excel to move the square around to ensure that it encompassed, *in three dimensions*, the external points that were thinner than 0.736". Some of the points that measured less than 0.736" were evaluated using the Evaluation Thickness method described above.

Q. 54: Please address the following Board question, "This Board understands that UT thickness measurements are commonly used to determine pipe wall thickness and plate thickness in other industries (see, e.g., Attachment to Citizens Answer (Selected Papers by Dr. Hausler)). To enhance the Board's general understanding and thereby enable it to make a more informed decision, the parties should discuss other applications of UT thickness measurement and identify the best practices recommended by National Association of Corrosion Engineers or other professional organizations, if any, with particular attention to the determination of the thicknesses of corroded plates and the rate of corrosion. The discussion should include use of mean versus extreme value statistics and the Analysis of Variance used in these cases." [Board Question 10]

A. 54: (MEM, PT, JA) The Board's understanding that UT thickness measurements are commonly used is correct. For power plant applications, UT inspection has been the predominant technique used to measure wall thickness and flaws in pressure vessels, piping, tanks and heat exchanger shells and tube sheets. It is the most widely used method in the power industry as well as the nuclear industry. Recommended practices are provided in codes and standards such as ASME Code Section V (NDE) and ASTM E797: Practice for Measuring Thickness by Manual Ultrasonic Pulse-Echo Contact Method.

The ASME codes used in power plants, ASME Section III (Nuclear), Section VIII (Unfired Pressure Vessels), and Section XI (Inservice Inspections) specify UT as the examination method of choice for thickness, particularly for operating plants. In a similar fashion, other codes such as American Petroleum Institute (API) also predominantly use the UT technique to determine thickness and flaws. National Association of Corrosion Engineers (NACE) in its "Corrosion Basics" publication identifies ultrasonics as a method to measure "metal losses caused by corrosion and erosion" and states that "the measurements can be made from the outside of the vessels or pipelines during operation."

In general, these codes and standards do prescribe rigid UT inspection methodology, but do not prescribe data evaluation methodology (including whether to evaluate the data using the mean, extreme values, or analysis of the variance). Rather, they recommend that the owner specify the methodology on a case-by-case basis. To our knowledge, NACE does not require or suggest that the data be statistically evaluated using any particular method.

Typical power plant applications of UT include:

- Evaluation of Degraded Piping. Evaluation Methodology is prescribed by ASME Section XI, and applicable code cases (such as Code Case N513). UT measurement and subsequent evaluations focus on the average thickness of the degraded areas and the size of the degraded areas and not on extreme thickness values.

- Erosion-Corrosion (FAC) Prone Piping. Inspection practices were developed to identify the problems in regard to Erosion/Corrosion monitoring programs as they relate to NRC Bulletin 87-01, "Thinning of Pipe Wall in Nuclear Power Plants" and NRC Generic Letter 89-08 "Erosion/Corrosion-Induced Wall Thinning, and EPRI TR-106611." Components are examined both to ensure equipment reliability and personnel safety. EPRI has developed software (TR-106611), and workgroups have been established to incorporate the best practices and to share industry experience and technology development. UT measurements and evaluations use grids of points to determine the average thicknesses of the piping. The average of these grid readings is used for evaluation and determination of corrosion rates.
- Pressure Vessel Shell Inspection. Components are examined in accordance with ASME Section VIII to identify degradation of the vessel shells in order to ensure both equipment reliability and personnel safety. Inspection practices for feedwater heaters, for example, are developed to identify the degraded area due to steam impingement wear. In this case, UT measurements and subsequent evaluation focus on the average thicknesses of pressure retaining sections of the Feedwater Heater Shell.

- Tanks. Inspection practices are developed to identify degraded tank walls and floors. Components are examined in accordance with ASME Code Section XI and/or API 650 and 653. UT measurements and subsequent evaluation focus on the average thicknesses of degraded areas and not extreme values.

Q. 55: Does this conclude your testimony?

A. 55: (All) Yes.

In accordance with 28 U.S.C. § 1746, I state under penalty of perjury that the foregoing is true and correct:

Frederick W. Polaski	Date
Dr. David Gary Harlow	Date
Julien Abramovici	Date
Peter Tamburro	Date
Martin E. McAllister	Date