ABSTRACT

A challenge has been presented to the scientific community to apply "first principles" to the understanding, prediction, and control of nondestructive evaluation (NDE) processes and applications. The success of the program is evident in the attention of new researchers and in the diversity of scientific specializations that have been directed to NDE problems. Critique of the program has been in awareness of existing technology and in focus of resources. An approach based on "lessons learned" is suggested for meeting continuing challenges and projected challenges of the future. "Lessons learned" from NDE reliability assessment programs are reviewed. Quantitative NDE performance as a function of signal-to-noise ratio is discussed. Focus of future efforts on signal-to-noise improvements in production and maintenance environments is proposed.

INTRODUCTION

The greatest challenges that we have in our nation today are quality and productivity. Both challenges are addressed by implementation of inspection technologies to measure the quality and quantity of the goods and services that we produce. Inspection will be a key to the implementation of technologies that are emerging to address the productivity and reliability issues. Automation currently offers promise from the "re-industrialization of America." Automation demands understanding and reliability of the inspection tools that we use for implementation to improve our production quality and productivity. Automation without understanding and reliability will only enable us to screw-up at a much faster rate.
It is my task to bring some messages from the trenches—from the managers who are concerned about the technologies that we are implementing; the people on the floor who are concerned about the daily application of the technologies that we are implementing; NDE engineers who are responsible for implementing technologies; and the people in the field who are concerned with the use of the technologies in maintenance. The language, concerns and priorities of the respective concerned groups are somewhat different. The approach presented herein addresses some of the concerns and experiences from the user community.

Long-range planning and long-range commitment are required to understand and implement complex technologies. This conference reflects a significant long-range commitment to NDE technologies and is a tribute to the foresight of the sponsors and managers who assembled such an effort. As contributors, we have a responsibility to effect quality and productivity in our individual efforts and to effect rapid transfer and implementation of the knowledge that has been gained by our efforts. I am personally very pleased with the focus of the efforts of the program to address significant problems in NDE. We now have responsibility for productivity in execution and application of the "lessons learned" in addressing significant NDE problems.

THE ROLE OF THE NDE ENGINEER

NDE has evolved as an experience-based technology and thus has been primarily deterministic in nature. The direction of technology growth has been mostly that of applying "proven" techniques to new problems. Most frequently NDE has been applied as a rescue tool after a critical anomaly has been identified. As such, growth has been primarily "application" based, with little resource being devoted to in-depth understanding of the nature of the application or to "common denominators" for consistency in application.

Consistency in the application of NDE to critical problems has been recognized and a significant amount of effort has been devoted to addressing the consistency and reliability issues. Major efforts in effecting consistency have been directed toward generation of specifications and procedures for application. Although specifications and procedures can be useful guidelines for application of a new technique, they are not recipes for success and do not necessarily address major elements for understanding of the nature of the technology being applied. Contrary to popular belief, specifications and procedures are not ends in themselves but instead are departure points for focus in addressing "first principles" for defining and understanding the problems involved.

The historical role of the engineer has been to provide a vertical link between science and applied technology. The critical
need for NDE engineers is evident in efforts to refine applied NDE technologies and in the literature that has been generated that describes solutions to problems that have been solved and elegant solutions to trivial problems or to problems that do not exist. The near term critical role for the NDE engineer is in providing focus of problems that have a "common denominator" and focus on tools that have general applications. Focus on "common denominators" to problems is necessary to provide significant problems that can be addressed by the scientific community. The "common denominators" for NDE technology are the opportunity "windows" that Dr. Burte talked about in the preceding address. 1

THE PROBABILITY OF DETECTION (POD) CURVE

An engineering approach to the "common denominator" issue has been the development and application of the probability of detection or POD curve as a tool for assessing NDE capabilities and reliability. POD curves denoting inspection capabilities as a function of flaw size have provided a critical communication link between the designers in "what is attainable;" management in "what is controllable;" and NDE engineers in the "standards of performance for the technology." The POD curve provides a significant amount of information, but there is a lot of information that does not appear on the curve.

An example of a typical POD curve is shown in Fig. 1. Such a curve is obtained by generating flaws in components in a spectrum of sizes and passing the components through an inspection process or repetitively through an inspection process. The responses are ordered in terms of actual flaw size. Responses are then grouped into statistically significant samples to allow calculation of a point estimate of detection for the sample group (successes divided by opportunities). The point estimate of detection is plotted as a function of flaw size for the sample group. Successive grouping and plotting of the point estimate for detection is used to generate the probability of detection POD curve. A large number of observations and a large quantity of data are required to generate a POD curve, and hence to describe the capability and reliability of the NDE technique being assessed. In like manner, POD curves may be generated to reflect other flaw parameters of interest, such as the flaw depth. 2

Experience in the generation and use of POD curves has enabled us to extract additional information concerning the inspection/evaluation process being addressed. For example, the shape of the POD curve provides a qualitative basis for assessment of the degree of control for a given data set and a criterion for grouping similar data sets. Figure 2 illustrates the POD curve for the data sets generated under varying conditions of control. Curve A is typical of an inspection process that is under control, and this is discriminatory
Fig. 1. Typical form of a probability of detection (POD) curve.

Fig. 2. Typical probability of detection (POD) curve under varying conditions of process control.

(specific) to the desired output. Curve B is typical of a process that is approaching control. The mode and type of variance denote the influence of factors not accounted for in the direct correlation of process performance with flaw size. Curve C is typical of a process that is out of control, but whose performance is influenced by flaw size. Curve A is worthy of further statistical rigor. Curve B is worthy of further analyses to improve the variances. Curve C is worthy of further analyses to improve the process or to provide a measure of inspection discrimination by sampling. Flaw size is a secondary variance in Curve C at the operating point for
inspection. Identification and control of the primary variance will change the nature of the data set, the specificity of the technique and the resultant POD curve.

Each POD curve is unique to the specificity of the inspection process, the degree of control effected in the inspection process and to the nature and distribution of the flaws being assessed. Rigorous use of the data in specific applications is limited to the specific processes, control and flaw distribution conditions used in data generation. The cost of generation, precision in data collection and the discipline required for specific applications have fostered many attempts to generalize and model POD curve prediction. To date, no satisfactory model has been developed and some modeling attempts have contributed to the confusion in application and data generation. Since many critical inspections are currently performed by skilled operators using manual techniques, human factors are most frequently cited as the source of unreliability. Although human factors are a primary contributor to unreliability, nondestructive test engineering and engineering management (selection of the right tool for the right job) are proposed to be greater sources of unreliability. An approach to the true boundary conditions for operation of the technique may be in understanding the "physics of the problem" and in applying "first principles" for modeling and solution of the problem. If the POD curve is a useful "common denominator" for communication of the engineering capabilities, it may provide the "window" for approach to modeling and general solutions to the NDE capability and reliability problem.

CONTRIBUTING FACTORS TO THE POD CURVE

Although the POD curve provides a graphical method for quantification and communication of NDE capabilities and reliabilities, several factors contribute to the curve that are not necessarily reflected by the curve. The POD curve reflects only the positive success rate for the inspection tool applied. An inspection process constitutes an exercise in conditional probability as opposed to joint probability due to the interdependence of inspection stimuli and inspection responses. A schematic presentation of such interdependence is shown in the following:

<table>
<thead>
<tr>
<th>STIMULI</th>
<th>POS a</th>
<th>NEG n</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M(Aa)</td>
<td>M(an)</td>
</tr>
<tr>
<td>POS</td>
<td>T.P.</td>
<td>F.P.</td>
</tr>
<tr>
<td>A</td>
<td>P(A,a)</td>
<td>P(A,n)</td>
</tr>
<tr>
<td>RESPONSE</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M(Na)</td>
<td>M(Nn)</td>
</tr>
<tr>
<td>NEG</td>
<td>F.N.</td>
<td>T.N.</td>
</tr>
<tr>
<td>N</td>
<td>P(N,a)</td>
<td>P(N,n)</td>
</tr>
</tbody>
</table>
The outcome of the inspection test may be:

TRUE POSITIVE (T.P.),
where \( M(Aa) \) is the total number of T.P. calls;
and \( P(A,a) \) is the probability of T.P. calls.

FALSE POSITIVE (F.P.),
where \( M(An) \) is the total number of F.P. calls;
and \( P(A,n) \) is the probability of F.P. calls.

FALSE NEGATIVE (F.N.),
where \( M(Na) \) is the total number of F.N. calls;
and \( P(N,a) \) is the probability of F.N. calls.

TRUE NEGATIVE (T.N.),
where \( M(Nn) \) is the total number of T.N. calls;
and \( P(N,n) \) is the probability of T.N. calls.

Interdependence of the matrix quantities is denoted by:

\[
T.P. + F.N. = \text{Total opportunities for positive calls.}
\]

\[
F.P. + T.N. = \text{Total opportunities for negative calls.}
\]

Therefore, only two independent probabilities must be considered in alternative inspection/decision tasks.

The SPECIFICITY of the technique or the PROBABILITY OF DETECTION of flaws may be expressed as:

\[
\text{POD} = \frac{T.P.}{T.P. + F.N.} \quad \text{or} \quad \frac{\text{total positive calls}}{\text{opportunities for positive calls}}
\]

Likewise, the NONSPECIFICITY of the technique or the PROBABILITY OF FALSE ALARMS may be expressed as:

\[
\text{POFA} = \frac{F.P.}{T.N. + F.P.} \quad \text{or} \quad \frac{\text{total false alarms}}{\text{opportunities for false alarms}}
\]

Confidence limits for the probability of detection value may be calculated from standard tables for a given sample size and calculated value from experimental sample data. This technique establishes an estimate for performance at one flaw size value, calibration level and acceptance criteria level. Data of most interest to the design engineer, nondestructive inspection engineer and systems manager is plotted as a composite of the discrete values calculated for individual operating points.

In addition to the conditional nature of the inspection being performed and the potential for false alarms the POD curve does not directly reflect the nature of the calibration or the decision criteria used in the generation of the curve.
IMPACT OF SIGNAL AND NOISE RESPONSES ON POD

An approach to understanding the inspection reliability problem and to a general solution is in understanding the interaction of energy fields with flaws in materials and in quantifying the expected responses under varying experimental conditions. Much of this conference is directed to such understanding and quantification. It is well known that application of NDE in sizing a flaw yields results of considerable variance. Such variance is known to be due to the variation in response to the flaw under varying experimental conditions or to variation inherent to a flaw of a given size. This "third dimension" of analysis must be accounted for in an inspection model. Variation in response along a POD curve is shown schematically in Fig. 3.

SIGNAL/NOISE VARIATIONS AND THE POD CURVE

Consider a case where the response (signal) from a flaw is "Gaussian" in nature and where process noise is well separated from the signal (Fig. 4). Such an inspection has high specificity for discrimination of signals that are due to flaw responses from background or process noise signals that are inherent to the process.

Fig. 3. Interaction of the probability of detection (POD) curve with the distribution of flaw response.
Consider a second case where the response (signal) from a flaw is "Gaussian" in nature with process noise signals overlapping the flaw response envelope (Fig. 5). A threshold (signal) discrimination level may be set for this process to provide a degree of separation of flaw responses from inherent process noise. Some flaws will be missed by such a system and some false calls (rejections) will be inherent to the process. The lack of specificity will cloud the use of the process as a final discriminator.

Finally, consider a third case (where my inspections always seem to be), where the response (signal) from a flaw is coincident with the process noise signals (Fig. 6). Such a process provides a random discrimination of flaws and is not considered to be a valid process. Indeed, better separation is likely by simple coin flipping.

We want to make the right decision all the time. We do not want to reject parts falsely, so we must be very concerned about signal-to-noise ratios and that is one of the common denominators that we are looking for. If we design our evaluation such that it operates with a good separation of signal and noise, we can indeed have good separation or good specificity and good detection—high POD.

A POD curve typically reflects all of the variations in signal/noise response and discrimination levels as shown schematically in Fig. 7. A continuing variation in signal/noise response is reflected by variation in the discrimination level along the POD curve. The signal/noise response and the discrimination level appear to be "common denominators" for all inspection processes and hence all POD curves generated for respective processes.
Fig. 5. Signal/noise response for discrimination with overlapping signal and noise stimuli.

Fig. 6. Signal/noise response for coincident stimuli.

CRITERIA DISCRIMINATION RESPONSE AND THE POD CURVE

A second factor (common denominator) that may be shown to affect the mode and specificity of an inspection process is the criteria level selected. If the acceptance (discrimination) criteria level for this inspection (indicated by the vertical arrow) is set too high, then I am indeed going to miss some flaw that I should have seen by application of the process and "EVERYBODY WILL BE UNHAPPY." Consider an inspection process with a measurable separation in noise and flaw signal responses as shown in Fig. 8. If the acceptance criteria is set at a level that provides clear separation of noise signal from flaw signal, all flaws will be rejected, few false calls (rejections) will occur and "EVERYBODY WILL BE HAPPY." If the acceptance criteria is set too low, all flaws will be rejected, some false calls (rejections) will occur and "MANAGEMENT WILL BE UNHAPPY."
Fig. 7. Interaction of signal/noise discrimination with the probability of detection (POD).

Fig. 8. Influence of acceptance criteria level (vertical arrow) on process discrimination (specificity).
The process specificity and hence its POD curve may be affected by changes in the acceptance criteria level. Figure 9 illustrates the effects of varying levels of criteria discrimination levels on performance as denoted by the POD curve. It is important to note that the criteria discrimination level is a function not only of the rejection level imposed on an inspection process but also of the calibration reference standards and criteria used to set up and validate inspection process performance.

I am suggesting that some of the lessons that we have learned in running inspection reliability assessment programs in laboratory environments, in production environments, and in maintenance environments indicate that the signal-to-noise operating point established by the set up and calibration process and by the acceptance criteria established constitute "common denominators" for all inspections. I believe that the key is the signal-to-noise level response that is established and maintained for an inspection operation. If signal and noise can be established as a "common denominator," then we can write a specification that should be universally applicable. Not so—the application of the inspection process and knowledge of how the respective materials, flaws, surface conditions, etc. affect the signal-to-noise response is the all critical NDE engineering task for implementation. Misapplication of a technique, as I indicated in the beginning, is one of the primary reasons for poor
performance. We need to understand the interactions of the problem and the boundary conditions for application.

INSPECTION FACTORS AND SIGNAL/NOISE RESPONSE

Variations in the condition of the flaws to be interrogated and variations in inspection conditions will affect the signal/noise function of the inspection process and its resultant discrimination level. Figure 10 illustrates some known and projected variations in flaws and process applications on the signal/noise response. Experimental data on the effects of variation of a single parameter on the overall signal response have been documented by various investigators. It is now clear that documentation of the calibration technique and the process noise for the inspection is necessary to account for parameter variations in a predictive model.

PREDICTIVE MODELING OF INSPECTION PROCESS PERFORMANCE

Generation of POD curves and qualification of inspection processes are tedious, time consuming and expensive. At present, POD curves are unique to the inspection process and process application and cannot be used for a second process or process application. For critical applications, experimental qualification and validation are required and must be completed for each process and process application.
Current work is under way to approach predictive modeling based on "first principles" to calculate behavior and interaction of an energy field in a given application. The approach and emphasis of this important work will provide a prediction of the performance level (POD) for an inspection process for calibration and validation at a given signal/noise level. Ultrasonic and eddy current models have been initiated as first steps in providing the engineering tools for future nondestructive process applications.

Predictive modeling would be of significant advantage in both the qualification of additional inspection processes and in reconsideration of current processes. Consider the case of cracks emanating from a radius area in a slot as shown in Fig. 11. An eddy current inspection has been developed and qualified for cracks emanating from the center of the radius. The inspection consisted of inserting an eddy current probe, with a small ferrite core, into the radius area such that it touched the center of the radius in a plane passing through the center of curvature of the radius. After qualification and validation, crack initiation was discovered at both points of tangency of the radius area. Predictive modeling/analysis tools could have been used to calculate the size of cracks that could be reasonably detected by the center probe technique. Actual requalification and validation were necessary to establish the performance level with the analysis tools that are currently available.

Consider the problems of inspection for cracks growing from the radius area of a "T" stringer stiffened structure as shown in Fig. 12. Crack initiation and growth can be predicted to occur at the tangency point at the root radius. Indeed such a shape was designated as "fracture critical" and extensive work was completed to validate an eddy current technique. After the inspection was in place, a forging lap occurring anywhere around the radius of the

Fig. 11. Variation in location of service cracks in a slotted engineering hardware component.
juncture was identified to be a potential failure mode. The immediate question to be addressed was, "what size flaw would be reliably detected if the location was assumed to be anywhere around the radius?" Available analytical tools would have minimized the necessity to re-validate the technique for random flaw location. Modeling tools would have been of significant help in approaching this problem.

In addressing the problem of modeling, I cannot overemphasize the importance of communication between the analyst and the NDE engineer. I urge all of you to go out and touch the hardware; see what the problems are. Go see what the guy on the production floor has to deal with on a daily basis. One of the most frustrating parts of the implementation task is in addressing requirements that do not match the part, the production situation, or the part usage. Failure to touch the hardware results in lost communication for the task objectives that are the goals of the entire production team.

Interim approaches to tough problems are still a very important and welcome part of engineering. We do apply both experience and analytical tools so we do not have to wait for a perfect solution to benefit from analytical experience gained. Interim solutions are very welcome and can be very profitable as we apply them on a near-term basis to approach tough problems.

CONCLUSIONS

Flaw detection reliability assessment is indeed a complex process that requires consideration of many factors in both application processes and in assessment. Signal/noise response at a given discrimination (criteria) level have been introduced as a "common
denominator" to characterize inspection process performance. Performance levels may be summarized by plotting a probability of detection curve (POD) curve for the inspection process and application. The human factors contribution to the output must be minimized for "controlled data" that are used for assessment of the capability of a process.

Reliable flaw detection may be effected by knowledge of the nature and boundary conditions for signal response in a given inspection task. Such analysis is necessary to provide the nondestructive inspection engineering that is necessary to application to critical inspection processes.

Progress is being made for predictive inspection modeling based on "first principles" of energy interaction and scattering. Such techniques will be a primary tool for all future nondestructive inspection engineering analyses. Process modeling, together with human factors modeling, will provide the necessary tools for improvement of productivity and for automation of inspection processes in future applications.

I urge both open communications and documentation of "lessons learned" in both the analysis and applications worlds. I urge emphasis on NDE engineering and on interim approaches to engineering problems for early implementation of knowledge gained in the analytical world. Finally, I urge all of you to get out and touch the hardware, see what the problems are, and talk to the people who are addressing the problems. We can all benefit from vertical communication and can be most productive in solving significant problems as the problems are better defined and the boundary conditions are understood.

REFERENCES


DISCUSSION

M.H. Jacoby (Lockheed): You indicated that difficulty in inspection is improper application of techniques, and that's certainly true. But given right techniques--and you certainly have exquisitely sensitive methods--thermography, xerography, tomography, the weakest link is still the subjective evaluation of visual data by the inspector. To overcome this shortcoming, I really believe we need computer evaluation. You can spend a million dollars on the fluoroscopic situation that seems just perfect, and the readout is some poor inspector on a dimly-lit TV monitor trying to write down defects.

W.D. Rummel (Martin Marietta): I agree with that, but when I say proper NDE engineering, I don't mean to give the operator a dimly-lit TV monitor to make this decision. Once again, he needs to have the proper tool so he can discriminate between that which is significant and that which is not. That's the part of the NDE engineering I'm talking about. It is not just all the other things that go with automation, but it is to give the person who makes that decision the right tool or we can't expect him to make the right decision. It will still be very difficult, but it will be improved.

M.H. Jacoby: There's still variability but really, the only way I believe that we can improve it is to have computer link-up.

W.D. Rummel: I think it certainly does give us another dimension of improvement.

M. Nikoonahad (University College London): Could you say a few words about the effect of different techniques on the shape of the curve you showed, i.e., probability of detection versus flaw size? If you are using ultrasonic methods and/or eddy current methods, does the shape of the curve remain the same always?

W.D. Rummel: The shape essentially remains the same for surface-connected defects. Since my work has been on fatigue cracks, I know that shape will remain much the same for eddy current, ultrasonic, and very closely-controlled techniques. It obviously will not remain the same for x-ray radiography since it is one dimensional. One of the things that I normally emphasize is that we
are not looking for x-ray versus penetrant or eddy current versus ultrasonic. These are complementary techniques for making the right decision.

D. Green (Rockwell International Science Center): Is there any advantage in combining the estimation of flaw size with other properties so that you have combined the different types of criteria?

W.D. Rummel: There is certainly an advantage, and I would judge that fracture toughness would be one of the things that we might want to look at. The techniques that are applied are so different that we are not addressing the same issue in many cases. In order to assure performance on the piece of hardware that we are dealing with, all of the items must be plotted and we must have the entire problem under control. NDE is just one piece of that problem.