A Path Forward for NDE Reliability

Ward D. RUMMEL *
* D & W Enterprises, Ltd.,
8776 W. Mountainview Lane, Littleton, Colorado 80125, USA

Abstract. Nondestructive Evaluation (NDE) is an essential element in modern industry and to advanced technology challenges. The NDE role is diverse and ranges from quality assurance of products and systems produced to critical roles in risk analysis and structure integrity assurance in the life cycle management of modern structures and systems. The development and application of fracture mechanics was a revolution in the role and requirements for NDE applications. NDE technology knowledge and practices were challenged to provide quantified detection capabilities and supporting data for specific applications. I led a team to provide detection capabilities data for NDE procedures intended for application to the National Aeronautics and Space Administration (NASA), Space Shuttle. The metric and data developed were the origin of "Probability of Detection – POD". POD is now recognized and used throughout the world as a tool for quantifying NDE capability assessment and for support of NDE reliability advancements.

Previous workshops have addressed understanding NDE variables in detection, data collection and measurement analysis procedures, and alternative methods of estimating detection capabilities and applications including model assisted POD. A suggested path forward includes development of an NDE engineering protocol and a formalized protocol for "calibration" of NDE measurement tools. Requirements for "calibration" and validation of sensors for various applications includes sensors for "Structural Health Monitoring".

1. Introduction

Industrial development in the modern world was accompanied by the development of nondestructive evaluation methods and procedures used in process monitoring, quality assurance, and periodic maintenance. Design protocol generally included large margins (safety factors) to accommodate unknowns in materials properties, loading and anomaly (flaw) tolerance limits (toughness /damage) and flaw behavior (slow flaw growth) in service. In some cases, a first article, system proof test and /or test to failure was and is used for design / production validation. NDE was frequently used in quality assurance of test articles, but capabilities and results were not linked to fitness for purpose. Failures were addressed by adding margins and nondestructive inspection / evaluation / testing (NDE) to known and suspected failure initiation sites – and of course, blaming the NDE craftsman (modern term is "human factors"). Such trial and error procedures were incorporated as the designs, production and acceptance methods matured in traditional industrial products and product use. The design methods (including NDE) were validated by evolution and use. Traditional NDE methods remain and serve the needs of many industry applications as long as the structures design and requirements do not change.

2. The European-American Workshops on Reliability of NDE

The European-American Workshop Series on Reliability of NDT have focused on major contributors to NDE reliability in the field applications and has been a useful forum for the transition from traditional methods and practices to quantified and validated systems applications. In past workshops, a conceptual model was proposed [Mueller, 2000] in the form shown in Figure 1.

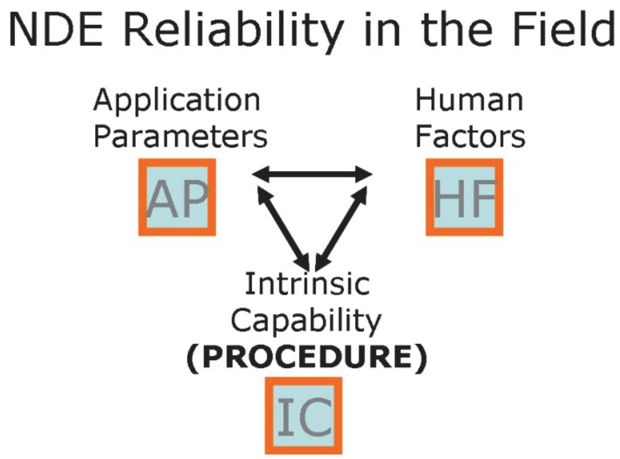

Figure 1. BAM Model for NDE Reliability in the Field

$$R \equiv f(IC) - g(AP) - h(HF)$$

Where R is the end item NDE system reliability

- f(IC) is a function of the "intrinsic properties" of the applied NDE system/ procedure
- g(AP) is a characteristic application parameter (conformance to the NDE system/ procedure), and
- h(HF) addresses variance in human factors (knowledge, skills and experience)

This model supports known issues on the importance of human factors, and conformance to established NDT system and prescriptive procedure instructions. It assumes that procedures are capable of meeting acceptance and service life requirements. The "intrinsic properties" of the prescriptive procedures and applied systems are often based on beliefs, past experiences and practices and/or "engineering judgment" – in short - "business as usual". If an NDE system / procedure failed to meet expectations, shortfalls were attributed to "human factors" and / or "failure to follow" prescriptive procedures that had been used in the past. Those procedures that produced acceptable outputs were validated by incremental use in specific applications and industries. If changes occurred in application requirements, equipment or application parameters, validation by trial and error was repeated. Those procedures may be useful and remain valid if requirements and application parameters do not change. By those procedures, application performance or screening (flaw detection) levels are not known, but may be adequate when large service margins and similar application parameters are involved.

Human factors and replication of validates procedures are important in all NDE applications and may dominate end item output. Development and validation of NDE procedures is equally important and shortfalls in procedures may result in a failed inspection. The NDE craftsman deserves and expects to apply a procedure and cannot be expected to have confidence in a procedure that has not been validated. If a procedure is not supportable, an inspection by that procedure may be a very expensive inspection ceremony.

This paper focuses on considerations in characterizing the "intrinsic properties" of NDE procedures, on NDE procedure development, and on procedure optimization to improve both capabilities and reliability.

3. NDE Requirements

Implementation of fracture mechanics in design, production, fitness for purpose and life cycle management of structures and systems changed both the nature of requirements for NDE capability performance as well as the role of NDE. Damage tolerance requirements are quantified in terms of:

- Flaw type (damage condition)
- Flaw size
- Flaw location
- Flaw origination
- Nearest neighbor (combined conditions / indications)

Quantitative NDE was integrated and became a pivotal requirement many modern structures designs as shown schematically in Figure 2

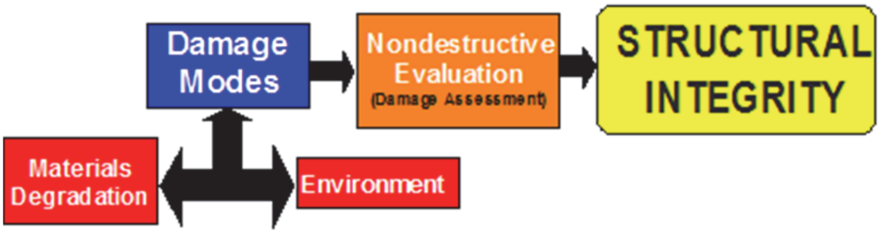

Figure 2. NDE is essential to structural integrity

In short, the requirement for quantitative NDE resulted in major changes in characterization in NDE capabilities, procedure development and applications. Note that all NDE procedures do not require rigorous characterization and quantification and traditional validation methods continue to serve many needs.

The Probability of Detection (POD) assessment methodology was developed to meet the challenges of damage tolerant design requirements and to quantify NDE performance capabilities. POD assessment addresses the multiple parameter nature of NDE procedures and incorporates probabilistic methods for output sampling and analysis. The first POD analyses were developed to support the NASA Space Shuttle Program [Rummel et al., 1974]. POD provided the methodology for assessing flaw detection capabilities and an output metric for integration into fracture control / damage tolerance requirements. Its use grew in multiple industries and applications as a method and protocol that is useful in structures design and service life management. The method involves assessment of a large number of test specimens, containing representative flaws of various sizes that are representative of the population of flaws (damage conditions) to be addressed in an application. The output is a probabilistic analysis of assessment results in the form of a plot of detection as a function of flaw size. Rigor in sampling methods supports a confidence level for the reported result. An example of an original POD output is shown in Figure 3 [Rummel et al., 1974]. By convention, the useful output result is reported for a 95% confidence for flaw detection at a 90% detection level (the 90 /95 metric). POD has evolved as a useful metric for characterizing an NDE procedure and application and for documenting and communicating NDE procedure performance. It has various uses in exploring detection capabilities to asses and quantify variances and the effects of variances into a single parameter (or group of parameters) that characterize the output of an NDE procedure t. It encompasses all of the factors in the workshop model. It is an end to end (snap shot) assessment of an NDE system and application at the time of the assessment. Extensive POD reference data has been documented for a variety of materials and applications and has been the reference, starting point for many applications [Rummel and Matzkanin, 1997]. However, POD is not a constant and varies with variations in the multiple parameters that characterize each NDE procedure application [NASA, 2008]. Prescriptive NDE procedures are available from various "document selling organizations" and may be useful for reference or for use in historical, quality assurance applications. Those procedures are not sufficient for most fracture control applications and rarely include supporting data or supporting validation.

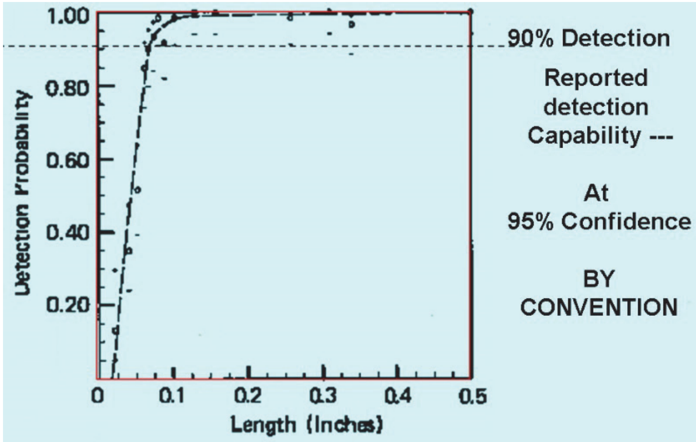

Figure 3. A POD Curve from the Original NASA Program

4. NDE Reliability

Quantitative NDE must be regarded as a measurement process. The output may be treated in terms of the NDE measurement systems; multiple measurement process parameters; and on variations in the test object and flaws / conditions being measured. To be useful, the output is expected to be reasonably the same each time the NDE procedure is applied. Some procedures include requirements for both flaw detection and flaw size measurement. Although some measurement characteristics and parameter controls are common to both flaw detection and sizing, the challenges are different and different approaches are required. This paper focuses on flaw detection (POD) and recognized contributing parameters.

Improvements in the output of multi parameter processes may be addresses by identifying and isolating the contributions and effects of sources of variance. POD is the product of variances in both controllable and non -controllable characteristics and parameters. Improvements can be made by reducing variances in controllable parameters and by bounding the effects of variances that are not controllable.

The reliability of an NDE procedure may be characterized in terms of its:

- Applicability
- Reproducibility
- Repeatability and
- Capability (POD)

A. Applicability - The applicability of a candidate NDE method may vary widely with material, industry, operator skills and prior experience. In general, the best method is that which provides the highest signal to noise at the target flaw / damage size. A starting point is usually a based on prior experience or literature search. Practical considerations, such as available equipment / materials; access; environment; operator skills; operator environmental exposure; and cost may constrain options. Beware of expediency. Failure to consider all aspects of a candidate application, including flaw behavior and life cycle, service may prove costly in application and risks. Selection and production of test artifacts is a critical step in development of an NDE procedure. Applicability is strongly dependent on artifacts that are representative of the population of flaws, damage conditions and the range of conditions to be addressed by a procedure. Artifacts of lesser fidelity may be suitable for initial screening, but representative artifacts (clones of intended test objects) are required to optimize procedures.

B. Reproducibility – Significant improvements in NDE reliability can be achieved by attention to the reproducibility of the NDE measurement system. Variability in an applied NDE measurement system (materials, equipment, cables, sensors, etc.) is controllable. Attention to integrity, wear, functions, etc. of all components of an NDE measurement system is required. Note that quality assurance metrology and disciplines may require reinforcement to meet requirements of the NDE measurement system. A trip to the metrology laboratory may be a detriment unless requirements, allowable adjustments, etc. are clearly communicated and followed.

Various materials and components are assembled for each NDE measurement system. System level performance must be reproducible and repeatable. A system level "calibration" is required to assure continuing applicability for the detection procedure and a lesser system level "set-up and calibration" is required before each system use. Responses from artifact specimens are used to provide assurance that the NDE measurement system provides a reproducible and repeatable response. Artifact specimens have been developed for a multitude of NDE methods and applications. Responses from artifact specimens must be representative of the responses from flaws in service hardware in both magnitude and range of sizes / conditions. General purpose or specifically designed specimens (test blocks) may be used. The artifact specimens are an integral part of the measurement system and procedure.

Artifact specimens are produced commercially and support general procedures. Differences in NDE measurement response due to differences in the specimens may be a significant source of variance for precision work. Although physical measurements of reflector size, shape, etc. can be traceable, differences in NDE response is the important application parameter. It is difficult to produce and match responses from reference discontinuities. Commercial vendors use different processes to implant reference reflectors and some may have better process control and or be better suited for the intended application. Some differences have been discussed in recent literature [Caldwell et al, 2013]. Response repeatability may be assumed if the same artifact specimen is used whenever a validated procedure is used. If the procedure is used at more than one location or the artifact specimen is damaged, revalidation may be necessary. Use of a "master gage" artifact with off-sets documented for working artifacts is recommended and may be necessary for precision work. NDE response variance is left to the NDE engineer and the imposed rigor of procedure validation and use.

C. NDE System Set-up and "calibration" - Reproducibility in measurement systems applications are generally approached by rigorous attention to the measurement system set-up and "calibration". NDE detection is dependent on discrimination between signal and background noise responses. Typical signal and noise discriminations are schematically represented in Figure 4. Noise is not electronic noise, but is characteristic to the NDE method and test object material, shape, homogeneity, etc. If we view noise as a constant and the flaw size is decreased, the two signal distributions move together (merge) and discrimination is not possible at some (small) flaw size. Decreasing discrimination capabilities with decreasing flaw size are shown schematically in Figure

5. Figure 6 shows a state of overlapping signal and noise distributions that result in reduced signal discrimination (detection) and a source for “false calls” (flaws reported when no flaw is present).

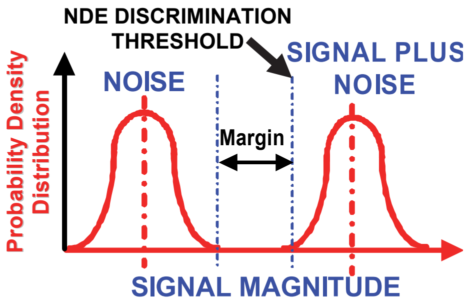

Figure 4. Clear Discrimination

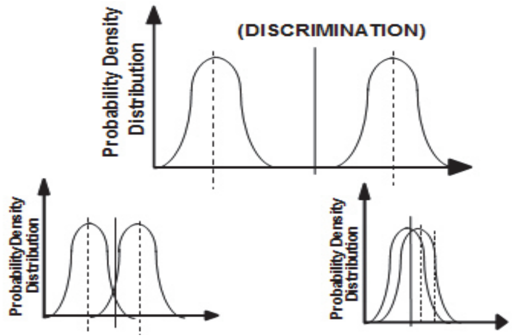

Figure 5. Reduced Discrimination at Smaller Flaw Sizes

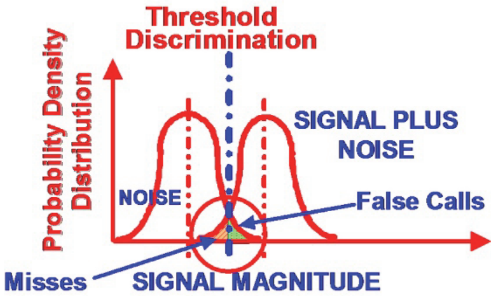

Figure 6. Overlapping Signal and Noise

The NDE threshold discrimination level is assumed to increase with increasing flaw size. Traditional NDE applications often use a single artifact for set up / “calibration” as a method of addressing measurement reproducibility. The set up artifact size is then used as a basis for setting the threshold discrimination level for a procedure. Unfortunately, a single artifact and set up on a single artifact does not assure reproducibility for detection or measurement precision. Figure 7 illustrates output variances that are possible using a single point “calibration”.

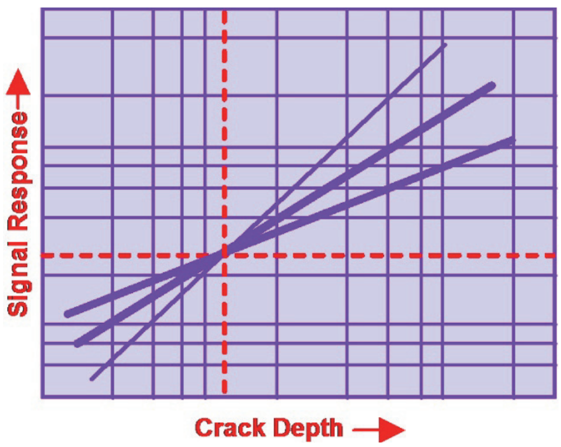

Figure 7. Signal Response Variances Using A Single Point “Calibration”

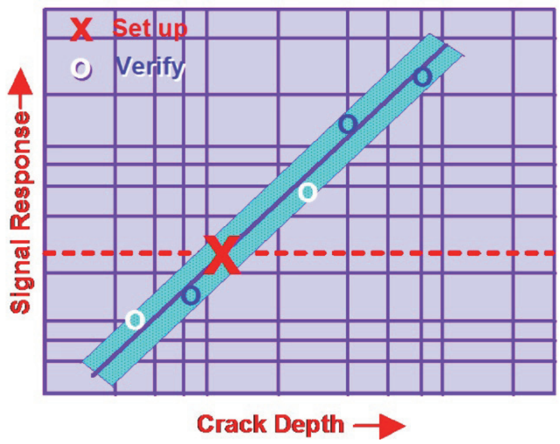

Figure 8. A Multiple Point “Calibration” and Response Verification

Significant improvement in detection reproducibility and reliability may be realized by the use of a multiple point "calibration" as shown in Figure 7. The multiple point "calibration" also provides an opportunity to verity component "noise" level and to quantify variances and off-sets that may better characterize and understand the properties of the component being assessed. A multipoint reference procedure is common to metrology and to many applications involving indirect measurements. Improvements in system reproducibility apply to both POD / damage tolerant and to traditional NDE procedures and applications.

D. Repeatability – We expect repeated applications of NDE procedures to produce the same results within characteristic variances of the measurement methods. An important part of developing an NDE procedure is in validating that it is capable of producing repeatable results. Indeed results produced by repetitive application of a procedure are useful in assuring that controllable variances have been identified and addressed. It may also be useful in identifying dominate variables and error sources in procedure application.

Repetitive output results for a procedure provide further data for procedure validation. Note that different test hardware and flaws are required. Repetitive results from known test sets, may be used as is a baseline for NDE craftsman training and proficiency demonstration. When sufficient data are available, POD may be used as a metric to confirm repeatability of a procedure in a specific application.

5. Modeling Tools May be of Aid in NDE Procedure Development

Data analysis and plotting of the NASA data for first POD curves was completed using punch card inputs to a mainframe computer. Electronics and computing technology have expanded to provide tools and capabilities that were unimaginable at that time. In addition to great improvements in equipment, sensors and operating displays, Simulation models have been developed to aid in visualizing scan and wave propagation paths and reduced the time necessary to assure inspection / detection coverage (applicability) during procedure development. Popular models for NDE include:

- Ray tracing
- Data analysis (capability predication from measured data and data form)
- First principles capability predicative models
- Model aided POD based on controllable / measureable parameters

The Berens model [Berens, 1997] was one of the first developed to aid in POD data analysis. It provides a POD output with reduced data sets and has been widely used in many applications. It has evolved and is the basis for MIL STD 1823 [DOD, 1999]. Increasing NDE signal levels with increasing flaw size are assumed as shown schematically in Figure 9. It is consistent with the first data sets produced for NASA, where the flaw sizes were small with respect to the size of the sensing transducer / probe footprint. It is not, however, sufficient for analysis of NDE data that does not conform to the assumptions, data form and constraints that are inherent to the NDE procedure. The Berens model spawned development of a multitude of data analysis schemes and models. The assumptions, constraints, data form and data range for all models must be recognized and addressed to develop confidence in model applicability and use. For example, NDE data for applications where the flaw size is larger than the transducer / probe footprint do not conform to the basic requirements of the Berens model and adjustments or alternatives may be required. Response from probe / transducer interactions may vary and the variances must be integrated into both NDE measurements and POD analysis models. Figure 10 shows a typical variance for eddy current footprint / crack interaction with varying crack size. When the crack is smaller than the foot print of the interrogating field, the output is usually Gaussian as shown for smaller cracks. When the crack is larger than the foot print of the interrogating field, the output pattern varies along the scan path / along the crack length with a higher amplitude at the ends of the cracks. The data output and form varies with crack size.

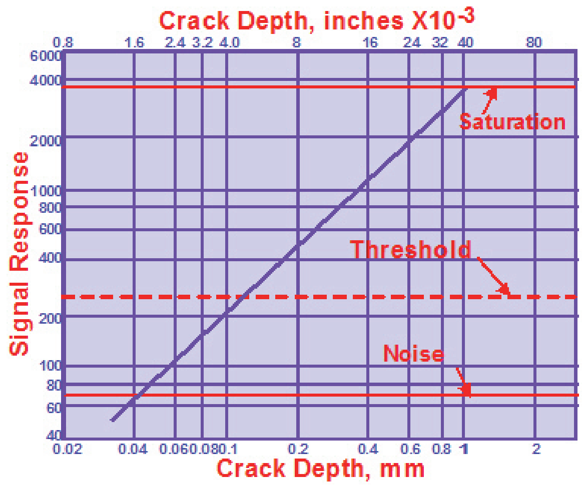

Figure 9. Data Form for the Berens Model

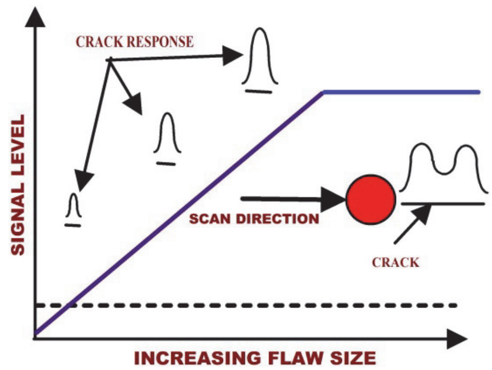

Figure 10. Eddy Current / Crack Interaction

It is self evident that data acquisition, data form and data analysis are specific to both the test object properties and configuration and to the NDE method being applied. Generalization is difficult. Useful guidance in design of experiment, data analysis and use of various models / analysis schemes in validating NDE data analysis procedures are available [Generazio, 2011]. The form of NDE data is specific to each application and most of the emerging models are specific to and constrained by the data, data form and data range. The user must assess the constraints and validity of each model and model use.

A number of "first principles" predictive models have been developed and are useful in predicting data form. This is the basis for a "model assisted" POD protocol [Thompson, 2008]. Although models are useful aids, their output must be anchored to NDE measurements and data. [Aldrin, 2001], [Maleo, 2010]. Each is specific to a method and controllable constraints of an application. Bounds for applicability of model and measurement outputs may be imposed by variances due to uncontrollable parameters. If the data outputs, form and NDE acceptance levels can be shown to be similar to that of a more extensive data set, a projected POD capability may be assumed [Rummel,1999], [Smith, 2005].

6. NDE Engineering/Linking the Tools and Disciplines for Reliable NDE Quantification

One of the most difficult challenges in developing, validating and applying NDE procedures is that of obtaining and assessing test specimens that are representative of the population of flaws and flaw conditions to be addressed. Lower fidelity specimens may be useful in assessing data form and linking system measurement data to model output, however . The ultimate test / validation is that of linking data output to the population of defects that must be addressed in application.

In short, much progress has been made in identifying and predicting interactions of NDE system variables, test object variables, application variables and first principles NDE measurement output variances. Little work has however been done to characterize the interactions and contributions of flaw variances on NDE measurements and output [Corbly,1969], [Wooldridge, 1979], [Wooldridge, 1980]. Material damage and flaw variances may be considered to be uncontrollable, but may be dominant in specific applications. These include flaw angles; state of loading (compression or tension / flaw closure); flaw surface morphology; state of oxidation; presence of fluids; load history and materials variances due to loading, environment or service use. Materials and part service history are often known and may provide descriptive inform for predicting potential flaw condition. Although service history data may not be quantified, it can be useful in bounding the expected NDE outputs. Flaw state and interrogating NDE field interaction must be addressed on a case by case basis. Considerations may include:

- NDE signal reduction due fluids in a crack (a common condition)
- NDE signal reduction due to flaw closure (flaw under compression or tension)
- NDE signal reduction due to flaw surface roughness (for example surface texture of a fatigue crack as opposed to that of an overload tear
- NDE signal variance due to flaw inclination or surface contour
- NDE signal reduction due to flaw shape that may be linked to recent loading history
- NDE signal variations due to flaw geometry (crack length and depth)

Few reference examples or engineering protocol information are available for general use. Some have been approached on a case by case basis. In most cases, NDE margins are used to accommodate unknown and uncontrolled variances.

7. SUMMARY

A multitude of opportunities are available to provide improvements in NDE procedures development and applications. Some improvements are straight forward and available for immediate use. Traditional NDE and NDE practices are embedded and many have stood the test of time. Resistance to change will continue to slow implementation. Some improvements require in depth understanding to address multiple sources of variance. New understanding and integration of the available tools are new challenges to implementing improvements. The path forward is clear. The timeline to meet the challenges and reap the benefit is less clear.

Damage tolerant design and management needs have resulted in challenges to NDE capabilities and needs. New NDE challenges require new knowledge and skills that are beyond those used in traditional NDE procedure development, validation, application and management. The demands for NDE and the challenges to NDE are likely to continue with increasing demands for engineering structures and systems. Parallel demands are increasing as engineering structures and systems age and remain in service. The integration of new science, knowledge, practices, skills and proficiency extends beyond traditional NDE practices technology advancements. The required / integrated expertise may be classified as NDE engineering.

A new world of NDE engineering embraces classical metrology and quality control, together with the physics of NDE and NDE measurements; statistical analyses; materials engineering; design engineering; process engineering and process modeling; and systems engineering disciplines and protocol. NDE is on the threshold of an ever increasing and useful future. This workshop continues to provide a forum for identifying challenges, idea exchanges and review of accomplishments of works in progress. Much work remains

References

Aldrin, J.C., “Overview of Mathematical Modeling in Nondestructive Evaluation (NDE),” R*eport NT-SP-O1-03*, Computational Tools, Gurnee, Illinois, November 2001.

Berens, A.P., “NDE Reliability Data Analysis, Nondestructive Evaluation and Quality Control, Quantitative Nondestructive Evaluation,” *ASM Metals Data Book*, Vol. 17, 5th ed., ASM International, Materials Park, Ohio, December 1997, pp. 689–701.

Caldwell, Jeff, Michael Moles, Luke Robertson, Allan Hull and Anthony Sinclair, “Reference Discontinuities for Small Surface-breaking Cracks,” *Materials Evaluation*, Vol. 71, No. 8, August 2013, pp. 929–935.

Corbly, D.M., P.F. Packman and H.S. Pearson, “The Accuracy and Precision of Ultrasonic Shear Wave Flaw Measurements as a Function of Stress on the Flaw,” *Proceedings of the American Society for Nondestructive Testing Fall Conference*, October 1969.

DOD, *MIL-A-83444, USAF Damage Tolerant Design Handbook:Guidelines for the Analysis and Design of Damage Tolerant Aircraft Structures*, Department of Defense, Washington, D.C., 1974.

DOD, *MIL-HDBK-1823, Nondestructive Evaluation (NDE)System Reliability*, Department of Defense, Washington, D.C., 30 April 1999.

Generazio, E.R., “Validating Design of Experiments for Determining Probability of Detection Capability for Fracture Critical Applications,” *Materials Evaluation*, Vol. 69, No. 12, December 2011, pp. 1399–1407.

Maleo, N., “Project Overview,” PICASSO (Improved Reliability Inspection of Aeronautical Structure through Simulation Supported POD), Moissy-Cramayel, France, 2010, www.picasso-ndt.eu/project-overview.

Mueller, Christina (Nockemann), Matt Golis, Tom Taylor, ”Basic Ideas of the American-European Workshops 1997 in Berlin and 1999 in Boulder”, **15th World Conference on Nondestructive Testing, Roma (Italy), 2000,** http://www.ndt.net/article/wcndt00/papers/idn733/idn733.htm

NASA, *NASA-STD-5009, Nondestructive Evaluation Requirements for Fracture Critical Metallic Components*, National Aeronautics and Space Administration, Washington, D.C., 2008.

Rummel, W.D. and G.A. Matzkanin, *Nondestructive Evaluation (NDT) Capabilities Databook*, 2nd ed., Nondestructive Testing Information Analysis Center, November 1997.

Rummel, W.D., P.H. Todd, Jr., R.A. Rathke and W.L. Castner, “The Detection of Fatigue Cracks by Nondestructive Test Methods,” *Materials Evaluation*, Vol. 32, No. 10, October 1974, pp. 205–212.

Rummel, W.D., “Transfer of POD Performance Capabilities from Simple Shapes to Complex Shapes,” *Review of Progress in Quantitative Nondestructive Evaluation*, Vol. 18, 1999, pp. 2305–2310.

Smith, K.D., “POD Transfer Function Approach,” *Presentation to Model Assisted Probability of Detection meeting*, Center for Nondestructive Evaluation, Iowa State University, 4 February 2005.

Thompson, R.B., “A Unified Approach to the Model-assisted Determination of Probability of Detection,” *Materials Evaluation,* Vol. 66, No. 6, June 2008, pp. 667–673.

Wooldridge, A.B., “The Effects of Compressive Stress on the Ultrasonic Response of Steel-steel Interfaces and Fatigue Cracks,” *Report NW/SSD/RR/42/79*, Central Electricity Generating Board, London, United Kingdom, April 1979.

Wooldridge, A.B. and G. Steel, “The Influence of Crack Growth Conditions and Compressive Stress on the Ultrasonic Detection and Sizing of Fatigue Cracks,” *Report NW/SSD/RR/45/80*, Central Electricity Generating Board, London, United Kingdom, April 1980.