April 29, 2011
EWI Project No. 50454GTH

Advanced Technologies and Methodology for Automated Ultrasonic Testing Systems Quantification

Other Transaction Agreement No. DTPH56-07-T-000002

Final Project Report

July 2007 through March 2011

Submitted to:

U.S. Department of Transportation
Pipeline and Hazardous Materials Safety Administration
Pipeline of Transportation Safety
Washington, DC

Final Report

Project No. 50454GTH

Other Transaction Agreement No. DTPH56-07-T-000002

on

Advanced Technologies and Methodology for Automated Ultrasonic Testing Systems Quantification

to

U.S. Department of Transportation
Pipeline and Hazardous Materials Safety Administration
Office of Pipeline Safety
Washington, DC

April 29, 2011

Roger Spencer, Evgueni Todorov, Perry White, and Nancy Porter (EWI)
Mark Lozev (BP)

Abstract

For automated ultrasonic testing (AUT) detection and sizing accuracy, this program developed a methodology for quantification of AUT systems, advancing and quantifying AUT systems image-capture capabilities, quantifying the performance of multiple AUT systems, establishing a guidance document, conducting field tests, and delivering a guidance document for future inclusion in reliability-based design and assessment (RBDA) standards. Improvements for AUT of pipeline girth welds were identified by investigating: imaging techniques for AUT data, modeling and simulation tools for AUT technique development and validation, and quantification methodologies for measuring the performance of AUT systems. Throughout the course of the project, many AUT scans were conducted using both zonal and non-zonal inspection techniques. The quantification methodology provides guidance for quantifying the performance of AUT systems with respect to probability of detection (POD) and accuracy of flaw sizing. Field trials and lab trials demonstrated that a sector sweep of beam angles provides alternative imaging capabilities that may enhance current zonal techniques. Current methodologies used for quantifying detection and sizing limits for AUT systems vary and produce great variability in detection and sizing results. Consequently, a standardized quantification approach is recommended to reduce this variability. This investigation evaluated six AUT systems; flaw heights were detected between 2 and 4 mm with a 90% probability and a confidence level of 95% (a90/95). The use of ultrasonic modeling and inspection simulation software provided a good tool for developing and evaluating AUT procedures; and can be used to evaluate changes in essential variables.

Acknowledgements

This research program consisted of a large international project team with multiple funding sources. Funding was provided by the U.S. Department of Transportation's Pipeline and Hazardous Materials Safety Administration (DOT PHMSA), UT Quality, Pipeline Research Council International (PRCI), ConocoPhillips, Petrobras, and Edison Welding Institute (EWI). In-kind contributions were provided by BP, PRCI, TransCanada, Applus RTD, ConocoPhillips, UT Quality, ExxonMobil, Chevron, Petrobras, and EWI.

Although large, the project team was quite active and provided contributions that exceeded our wildest expectations. The EWI project team wishes to acknowledge the extraordinary efforts of the following individuals and their organizations:

- **Jim Merritt** (DOT PHMSA), who attended numerous project meetings and provided guidance and insight throughout the project.

- **Frank Licari** (DOT PHMSA), who attended numerous project meetings, attended the field trials, and provided leadership and feedback throughout the project.

- **Mark Lozev** (BP), who started out as the Principal Investigator at EWI, then moved to BP where he continued to provide oversight and direction for the program.

- **Joe Kiefer** (ConocoPhillips), who attended numerous project meetings, provided feedback, support, guidance, and funding.

- **Kevin Hodgins** (UT Quality), who attended numerous project meetings, provided feedback, support, guidance, and funding.

- **Humberto Silva Campinho, Marcos Aiub de Mello, Agildo Badaro Moreira, Sérgio Damasceno Soares** (Petrobras) for participating in project meetings, providing guidance, support, and funding.

- **Max Toch** (PRCI), who provided assistance navigating funding cycles and contracting efforts at PRCI.

- **Evan Vokes** (TransCanada), who organized and facilitated the field trials on the Keystone Project in Canada. **David Hodgkinson** (TransCanada-retired), who helped design the program, lead the PRCI funding committee until his retirement from TransCanada, and continued to follow the project after his retirement.

- **Jan van der Ent** (Applus RTD), who assumed leadership of the PRCI funding committee (after Mr. Hodgkinson's retirement), attended project meetings and provided invaluable feedback.

- **John O'Brien** (Chevron) for providing support and feedback on the guidance document. **Luc Huyse** (Chevron) for attending our last project meeting to provide feedback on the statistical analysis.

- **Nathan Nissley** (ExxonMobil), who provided feedback, support, guidance, and equipment.

Disclaimers

U.S. Department of Transportation
Pipeline and Hazardous Materials Safety Administration
Pipeline Safety Research and Development Program

This research was funded in part under the Department of Transportation, Pipeline and Hazardous Materials Safety Administration's Pipeline Safety Research and Development Program. The view and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Pipeline and Hazardous Materials Safety Administration, or the U.S. Government.

Pipeline Research Council International, Inc.

This report is furnished to Pipeline Research Council International, Inc. (PRCI) under the terms of PRCI contract PR-185-084501, between PRCI and Edison Welding Institute (EWI). The contents of this report are published as received from EWI. The opinions, findings, and conclusions expressed in the report are those of the authors and not necessarily those of PRCI, its member companies, or their representatives. Publication and dissemination of this report by PRCI should not be considered an endorsement by PRCI of EWI, or the accuracy or validity of any opinions, findings, or conclusions expressed herein.

In publishing this report, PRCI and EWI make no warranty or representation, expressed or implied, with respect to the accuracy, completeness, usefulness, or fitness for purpose of the information contained herein, or that the use of any information, method, process, or apparatus disclosed in this report may not infringe on privately owned rights. PRCI and EWI assume no liability with respect to the use of, or for damages resulting from the use of, any information, method, process, or apparatus disclosed in this report.

Executive Summary

For automated ultrasonic testing (AUT) detection and sizing accuracy, the objective of the program was to develop a methodology for quantification with the goal of dramatically improving the predicted reliability of onshore pipelines.

The technical objectives of the program were to:

- Develop a methodology for quantification of AUT systems.
- Advance and quantify AUT systems image-capture capabilities.
- Quantify the performance of multiple AUT systems and establish a guidance document.
- Implement the quantification methodology in field tests and guidance document in reliability-based design and assessment (RBDA) standards.

The project focused on ways to improve AUT of pipeline girth welds by reviewing what is currently being done and looking at possible next steps. To accomplish this, the project looked at the following general areas:

- Imaging techniques for AUT data.
- Modeling and simulation tools for AUT technique development and validation.
- Quantification methodologies for measuring the performance of AUT systems.

Various imaging techniques were evaluated during the project. These primarily involved combining and merging of data to obtain composite views of the weld. By taking advantage of the beam steering capabilities of phased array (PA) it is possible to gather several angles of data for each probe position. The data can then be corrected for angular position differences and sound path distances so that the data can be merged and displayed as a single output. Use of such imaging techniques would allow multiple beam angles to be viewed in a similar strip chart-style format currently being used for single-angle viewing.

The project looked at both PA zonal and PA non-zonal techniques; no conventional AUT inspection techniques were evaluated. A zonal technique is commonly used during girth weld inspection. The weld is essentially divided into several through thickness “zones” at different depths (e.g., root, hot pass, fill pass 1, fill pass 2, fill pass 3, and cap). For this technique, a separate probe (or group of elements) is assigned to each zone. Conventional AUT systems require more probes, whereas PA AUT systems can provide the same coverage with fewer probes. If conventional AUT is being used, separate probes are assigned to each zone. If PA AUT is being used, a separate group of elements are assigned to each zone. The scanner rotates around the pipe and the zones are scanned simultaneously. One channel of the strip

chart is produced per zone. Alternately, a non-zonal technique can be used to inspect girth welds. With this technique, the weld is *generally* inspected as a whole and not organized into separate zones. For this type of inspection, either mechanical and/or electronic scanning is used for full coverage of the weld joint.

Modeling and simulation tools were used to develop a non-zonal AUT procedure that was used in field trials and compared to test results obtained using a typical zonal AUT procedure. Results from the field trials showed that the non-zonal AUT technique was able to detect all flaws that were detected by the AUT zonal technique being used for the AUT examination. Modeling and inspection simulation was also used to evaluate a typical AUT zonal procedure. Specifically, modeling was used to evaluate factors such as beam angle, probe frequency, PA aperture selection, exit point offset from weld centerline, and effects of guide band offset.

Throughout the course of the project, many AUT scans were conducted using both zonal and non-zonal inspection techniques. The girth welds were in 30-in.-diameter, X80 pipe having a 15.6-mm nominal wall thickness (WT). The weld bevel was a narrow groove J-prep with a 4-degree bevel. Most of the AUT data was collected on welds containing implanted flaws, which was part of the quantification portion of the project. These welds were later cross-sectioned and macro photographs were taken of the flaws. Images of flaw responses, along with metallographic cross-section images of flaws, were assembled into a reference library of flaw responses can be used for future training.

A quantification process was conducted during the project to benchmark different AUT systems. An AUT system consists of the following components:

- AUT equipment
    - Scanner
    - Instrumentation
    - Software
- AUT procedure
- AUT operator.

The quantification process during this project was also used as a basis for developing a guidance document for quantifying the performance of AUT systems with respect to probability of detection (POD) and accuracy of flaw sizing. Observations and lessons learned during the project were incorporated into quantification methodologies which were then used for development of the guidance document. Draft versions of the methodologies and guidance document were shared with the industry sponsors and participants throughout the project and their feedback was incorporated into the final versions included in this report.

Conclusions

Following are conclusions and observations from the project:

- It was noted that root flaws were not detected as well as expected. Only 77% of non-fused root flaws were detected during the blind trials by the PA zonal techniques. By comparison, PA non-zonal techniques detected all non-fused root flaws.

- One set of weld samples were machined in an effort to eliminate weld joint misalignment issues. This effectively decreased WT in some locations. These changes in WT did not have a statistically significant effect on AUT performance.

- The flaw-type affects the sizing accuracy especially for the height and depth. It was found that interbead lack-of-fusion (IB-LOF) flaws were easy to detect but were generally oversized in through wall height. The most accurate height sizing for IB-LOF flaws was achieved with time-of-flight diffraction (TOFD).

- In general, the tilted and skewed flaws did not have a statistically significant effect on the height sizing. However, a trend exists and if a larger sample of tilted and/or skewed flaws is used in narrow range of angles the effect might be significant.

- When flaws were in close proximity to each other, the UT beam hit multiple flaws simultaneously. The resultant flaw interaction affected all sizing estimates especially for the height and false-positive frequency (FPF).

- The FPF during the blind trials was relatively low at 1% or less.

- The systematic errors and standard uncertainty for AUT system blind trials were as follows:

    - Flaw height: systematic error = -1.04 mm; uncertainty = 1.48 mm
    - Flaw length: systematic error = -0.1 mm; uncertainty = 4.89 mm
    - Flaw depth: systematic error = -0.2 mm; uncertainty = 1.48 mm.

- During this project six AUT systems were evaluated. The resulting flaw heights that were detected with a 90% probability and a confidence level of 95% (a90/95) were generally between 2 and 4 mm. The best POD $a_{90/95}$ versus flaw height achieved for the blind trials was 3.02 mm, while the best open trial POD $a_{90/95}$ was 2.08 mm.

- Different POD functions and algorithms produced significantly different POD results for the same data. Results showed differences of up to approximately 40% in the

calculated POD versus flaw height. This data suggests the importance of knowing how the POD was calculated when making comparisons of POD values.

- The use of ultrasonic modeling and inspection simulation software provided a good tool for developing and evaluating AUT procedures during this project. In addition, the software proved useful for evaluating changes in essential variables which could help reduce the scope of practical trials.

- The current zonal approach to weld inspection works well; however, technique development and application of the inspection can vary between AUT companies. This is particularly true for PA probes where critical variables such as the number of elements in the aperture and the electronic focal depth need to be addressed.

- A PA probe contains multiple transducer elements that provide the capability to electronically sweep through a range of angles. This sweeping ability is not typically used for girth weld inspection. Field trials and lab trials conducted during this project show that using a sweep of beam angles provides alternative imaging capabilities that may enhance current PA zonal techniques.

Recommendations

Methodologies used for quantifying detection and sizing limits for AUT systems vary throughout industry. Quantification parameters such as fabrication of flawed sample welds, number and size of flaws, destructive evaluation techniques, essential variables, and statistical analysis methods can be quite different. This can lead to great variability in the detection and sizing results. Consequently, a standardized quantification approach is recommended to reduce variability in results caused by different quantification methodologies. It is further recommended that AUT quantification guidelines be considered for implementation into future revisions of RBDA standards.

Based on sample fabrication conducted during this project, it is recommended samples made by varying welding parameters (natural flaw samples) not be used for the quantification process. It is recommended that a study be performed to determine current practices and limitations for creating AUT flaw samples by varying welding parameters rather than implanting flaws. While implanted flaws generally are precise regarding location, type, and size; there is concern in the industry that implanted flaws are not representative of actual weld flaws, because the implanted flaws may be created with different welding processes and filler metals.

Consideration should be given to expanded use of PA sector scans to enhance current girth weld inspection techniques and provide alternative data display possibilities. In particular, PA sector scans should be included to improve detection of root and hot pass flaws.

It is recommended that future work should be done to determine the effect of essential variables on AUT flaw detection and sizing. These variables should include WT changes, flaw orientation and location, scanner guide band misalignment, beam angle deviations, surface roughness (e.g., due to common surface preparation techniques such as grinding), and surface geometry (e.g., typically due to angular distortion adjacent to welds).

Abbreviated Terms

AE	acoustic emission
ANOVA	analysis of variance
AUT	automated ultrasonic testing
CI	confidence intervals
CR	computed radiography
DNV	Det Norske Veritas
DOT PHMSA	U.S. Department of Transportation's Pipeline and Hazardous Materials Safety Administration
DS	downstream
DT	destructive testing
ECA	engineering-critical assessment
EDM	electric discharge machining
EMDC	ExxonMobil Development Company
ENIQ	European Network for Inspection Qualification
ET	eddy current testing
EWI	Edison Welding Institute
FPF	false positive frequency
FSH	full-scale height
HAZ	heat-affected zone
HP1	hot pass zone 1
IB-LOF	interbead lack of fusion
ID	inside diameter
LPA	linear phased array
MT	magnetic particle testing
NDE	nondestructive evaluation
OD	outside diameter
OTL	outliers
PA	phased array
PFP	probability of false positive
PO	purchase order
POD	probability of detection
POR	probability of rejection
PRCI	Pipeline Research Council International
PT	penetrant testing
RBDA	reliability-based design and assessment
RT	radiography testing
TJ	technical justification
TOFD	time-of-flight diffraction
US	upstream
UT	ultrasonic testing
UTQ	ultrasonic testing quality
VT	visual testing
WT	wall thickness

Contents

Contents (Continued)

Appendices

Contents (Continued)

Tables

Contents (Continued)

Contents (Continued)

Figures

Contents (Continued)

Contents (Continued)

Contents (Continued)

Page

Contents (Continued)

Page

Contents (Continued)

1.0 Introduction

Although automated ultrasonic testing (AUT) technology has improved significantly over the last 5-10 years, there is still conflicting evidence on the accuracy and limitations of current AUT methods for pipeline girth weld inspection.[1-4] Usually, for practical purposes the main criteria for determination of UT reliability are probability of detection (POD) and accuracy of sizing for the flaw height and length. A specific requirement for POD that is often used is that it should be at least 90%, when relating to allowable flaw sizes at a 95% confidence level (90/95 rule). An improved process of determining the AUT detection and sizing accuracy capabilities should dramatically improve predicted onshore pipeline reliabilities in the design stages.

The overall goal of the program was to improve the process of determining the AUT detection and sizing accuracy with the goal of dramatically improving the predicted reliability of onshore pipelines in the early design stage.

The technical objectives of the program:

- Develop a methodology for quantification of AUT systems
- Advance and quantify AUT systems image-capture capabilities
- Quantify the performance of multiple AUT systems and establish a guidance document
- Implement the quantification methodology in field tests and guidance document in reliability-based design and assessment (RBDA) standards.

The absence of enough publicly available data from independent investigators specific to girth weld AUT inspection tools leads to engineering critical assessment (ECA) and strain-based approaches being excessively or insufficiently conservative in assumed AUT uncertainty based on experiments with statistically significant data. The proposed solution and additional work needed to address the challenge were evaluated by focusing on following topic areas:

- Further enhancement of AUT technology to reliably deliver sizing accuracy throughout the entire wall thickness (WT) of a pipe and weld under a variety of conditions.

- The increasing use of AUT in pipeline construction, driven by the demands of ECA-based and strain-based design acceptance criteria highlights the need to develop a methodology and quantify detection capabilities and sizing accuracy of representative AUT systems currently available on the market.

- Incorporation of the enhancement, quantification methodology and benchmarking quantification results in Guidelines, Codes, and Standards.

This research area is related to reliability of AUT systems results and to two main quantities of the reliability: POD and accuracy of sizing of flaws in pipeline welds. The basic science and engineering principles of nondestructive evaluation (NDE) reliability is described elsewhere.[5] Advanced and quantitative NDE allows the use of acceptance criteria based on experimental destructive testing (DT) or fracture mechanics or other fitness-for-purpose assessments. DT or ECA approaches are used to determine allowable flaw sizes depending on material/weld properties and loading conditions during pipeline installation and operation. Typically, DT, or an ECA, gives allowable flaw sizes (smaller than critical flaw sizes, as safety factors are used) that can be expressed as an allowable flaw height as a function of flaw length.[1] In general, measurement of flaw height and length using NDE is uncertain, and the sizing errors encountered must be taken into account when calculating acceptable flaw sizes from allowable ones, as an assurance against accepting non-allowable flaws with a sufficiently high confidence.

NDE reliability determination is an integral part of the applied research efforts to validate the capabilities of new NDE techniques, instruments, and procedures for an intended application. Usually, for practical purposes, the main quantities for determination of NDE reliability are flaw POD and accuracy of flaw location, sizing, and type estimation. Determination of NDE reliability and demonstration of NDE technique performance include flaw detection and flaw characterization. Flaw detection is described by POD for the application and flaws in question. The most important flaw parameters to be determined during flaw characterization are size (height and length) location, and type for pipeline inspection applications. The log-logistics (log odds) function (nonlinear regression) has been known to be a good model for POD data.[5]

$$\mathrm{POD}(a) = \frac{\exp(\alpha + \beta \ln \alpha)}{1 + \exp(\alpha + \beta \ln \alpha)}$$

A typical POD curve based on hit/miss data from inspections for a variety of crack sizes and the lower 95% confidence bound are shown in Figure 1 (left).[1] An accuracy, or sizing plot [measurement error and measured ultrasonic testing (UT) height versus actual height], is also presented in Figure 1 (right).[3,4] False calls, or spurious indications, are critical NDE reliability quantities and can have a number of causes. When performing NDE of pipeline girth welds, a number of false calls may be observed and should be investigated. POD and sizing curves are specific for each material (metals or plastics, carbon or austenitic steel welds, welding procedures, etc.) and separate database for each material and weld are required in the Codes and Standards. Limited data is available in the open literature for AUT performance such as POD and sizing accuracy of flaw height using amplitude-based multi-probe and PA technology.

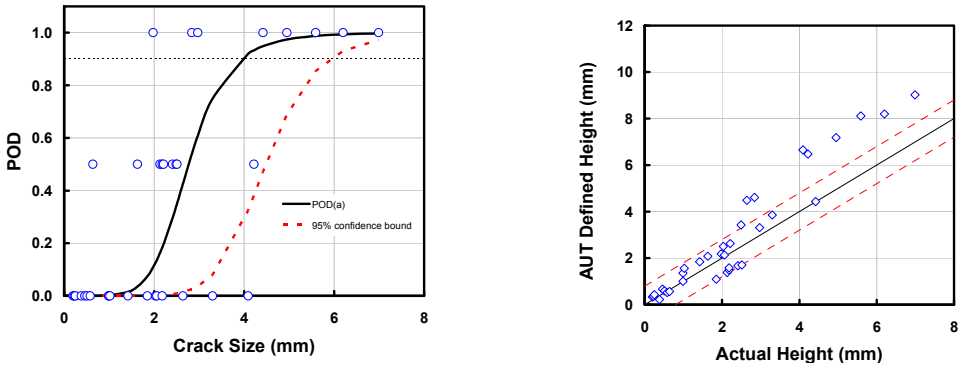


Figure 1. Examples of POD Curve (Left) and Sizing Plot (Right)

This program built on the results of previously successful projects funded by the private sector, State, and Federal agencies. It included the participation and support of potential end users of the deliverables and included a solid technology/knowledge transfer plan.

The deliverables for this program included a methodology to quantify imaging capabilities and AUT systems, POD and sizing accuracy curves for multiple representative systems, a guidance document for AUT capabilities, how to use the AUT sizing data for ECA/strain-based design approach, and technical justification (TJ) for modifications of the current requirements for AUT quantification trails demanded by the global practices of major companies and codes.

2.0 Methodology for Quantification of AUT Systems

EWI reviewed several report, industry standards, codes, and specifications to determine a quantification rationale that would work well for AUT of pipe girth welds. The reviews included documents such as ENIQ, ASTM, ASME, and API. Based on information gathered from these documents and previous experience by EWI and other industry partners, a set of methodologies were formulated to provide guidelines for quantification of AUT systems.[6-18] This quantification process provides a means to determine the POD and/or sizing accuracy of AUT systems for pipeline girth weld inspection. For critical AUT applications, quantification is often incorporated into AUT qualification programs as a means of measuring reliability of AUT systems. The guidelines developed during this project are intended to provide a standardized means of conducting the quantification process across the industry.

By way of definition, an AUT system consists of the AUT equipment, procedure, and operator. While it is common practice to quantify the AUT system as a whole, it is possible to quantify any of the three components making up the AUT system. Quantification is accomplished by conducting practical trials on welded samples containing a sufficient number of flaws to be statistically reliable. The only exception to this is when data from a previous quantification is applicable to another AUT application such as for a slightly different weld bevel geometry. In

this case, a previous quantification may provide sufficient TJ for the new application. Use of quantification results for different applications makes it important that the quantification process, as well as, the TJ process be standardized.

EWI produced initial drafts of methodologies early in the project and circulated them to industrial partners for review and comments. As the project progressed, lessons learned from the various stages of the quantifications process were combined with feedback from the industry partners to refine the methodologies. The methodologies are provided in the appendices as noted in Table 1. These documents are intended to serve as standardized procedures for conducting AUT quantification across the industry and improve the reliability of AUT girth weld examinations.

Table 1. Appendices Containing Methodologies

Methodology	Title	Appendix
1	Methodology for Contents and Review of Technical Justification	A
2	Methodology for Design, Fabrication, and Fingerprinting of Quantification Welds	B
3	Methodology for Practical Trials and Destructive Validation	C
4	Methodology for Data Analysis	D

3.0 Advance and Quantify AUT Systems Image Capture Capabilities

During this project, AUT imaging and AUT data-combining (data fusing) techniques were evaluated to determine their effectiveness for improving flaw detection and sizing. For quantification purposes, data display techniques will ultimately have a bearing on speed and accuracy of the inspection. Historically, pipeline girth weld inspection results have been viewed using strip chart outputs as shown in Figure 2. This type of output has the advantage of fast data acquisition and small file size; however, with advancements in computing power and file storage, the door is open for other data visualization tools that could possibly improve POD and sizing accuracy.

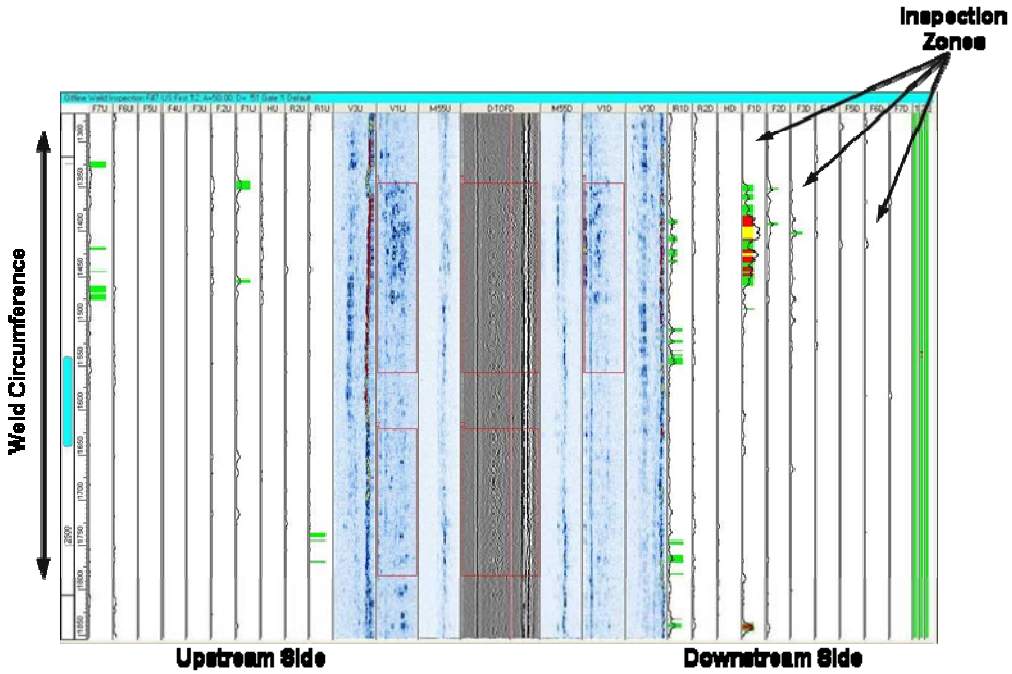


Figure 2. Typical AUT Strip Chart Output

The most fundamental aspect of an ultrasonic data file is the individual UT waveform data (A-scan). By using the signal amplitude versus time information from multiple A-scans, and combining it with probe position information, AUT data can be displayed in many formats. Figure 3 shows the concept of how AUT data can be constructed in a 3D array using A-scan and probe position information. Additional information regarding 3D imaging can be found in References 19-23. In Figure 3, the small blocks making up the cube represent individual AUT data points. For a 0-degree beam angle, the data rows and columns are the probe position and the "A-scan data" is the signal amplitude versus time-of-flight of the ultrasonic wave.

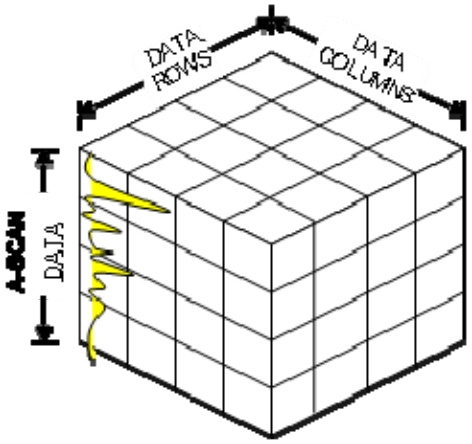


Figure 3. Arrangement of Scan Data

AUT of pipeline girth welds normally requires that sound is transmitted into the part at an angle. For complete inspection of a weld, several beam angles may be used to cover different zones within the weld. The sound beam is transmitted from the outside diameter (OD) surface and propagates through the material until it reaches the inside diameter (ID) surface. At this point the beam will reflect from the ID surface at an angle that is equal to the angle of incidence and continue to propagate back to the OD surface provided it does not encounter a flaw or a geometric feature. In order to image the data correctly, the beam entry point into the pipe, beam angles, part thickness, and material velocity must all be taken into account.

Once data arrays are acquired and arranged properly, the imaging software program can then present the data in a format that best describes the physical attributes, and conforms to established standards. Typical means of representation include A-, B-, C-, and D-scans, and in the case of PA systems, a sectorial scan (S-scan). It should be noted that B-, C-, and D-scans require position feedback information such as that obtained from an encoder. Figures 4 and 5 show examples of each data display format.

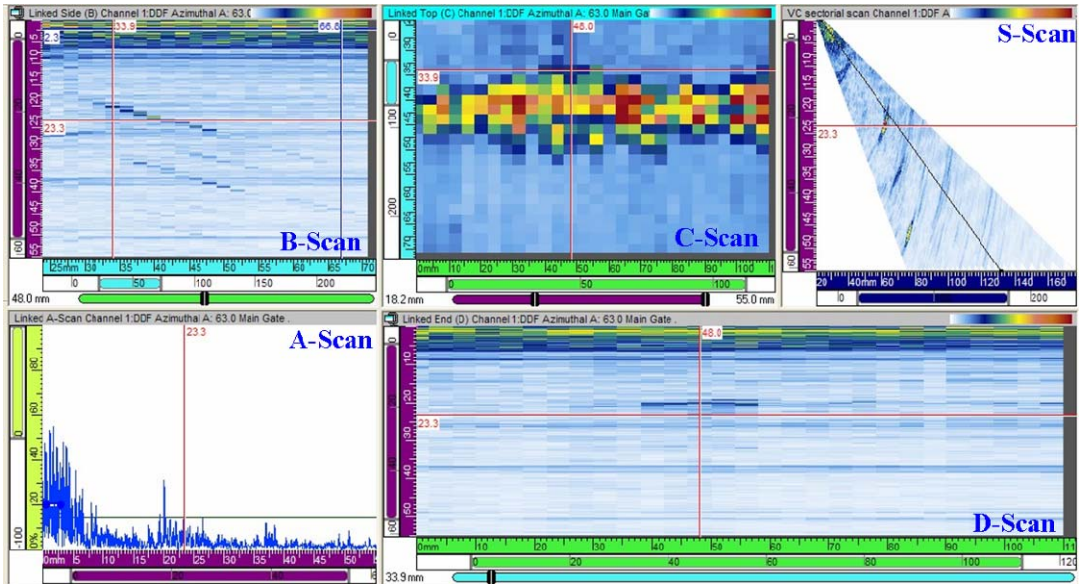


Figure 4. Example of Data Views

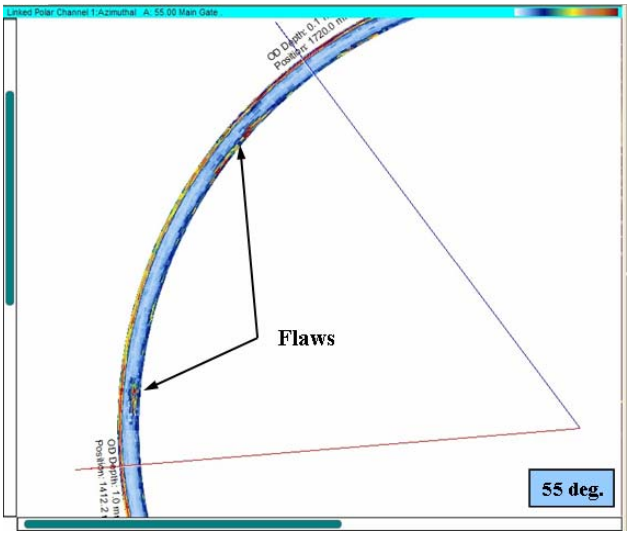


Figure 5. Example of Polar View

Following are brief descriptions of different data display outputs:

- **A-Scan** – is the basic ultrasonic waveform data displayed as amplitude of the reflected sound energy versus the time-of-flight of the ultrasonic wave.

- **B-Scan** – is a graphical representation of the ultrasonic data where the time-of-flight of the ultrasonic wave is plotted against the transducer travel position. For weld inspection it is normally a cross-sectional view transverse to the weld.

- **C-Scan** – is a graphical representation of the ultrasonic data where the amplitude of the reflected ultrasonic wave is plotted against the X-Y position of the transducer (a plan view). In weld inspection it is a view from the cap side.

- **D-Scan** – (same as B-scan except rotated 90 degrees) it is a graphical representation of the ultrasonic data where the time-of-flight of the ultrasonic wave is plotted against the transducer travel position. For weld inspection it is normally a cross-sectional view along the weld length.

- **S-Scan** – a graphical representation of the ultrasonic data where multiple angles of time-of-flight data is displayed from one transducer location. Very similar to fetal images obtained in medical ultrasound.

- **Polar View** – a graphical representation of UT data, obtained from inspection of cylindrical parts, which provides information of defect circumferential position, length, and depth. A polar view is essentially a D-scan displayed as a cylinder seen from the end. This is very useful for displaying pipeline girth weld data.

There are variations to data scan displays that can make the data easier to interpret. For example, B-, C-, and D-scans can be viewed with all the data combined or by selecting single slices of data. The concept of a single slice of data is shown in Figure 6 where a C-scan view is selected at a given depth within the material. This has the advantage of eliminating unwanted data caused by surface noise or part geometry. UT data files are typically arranged with groups of A-scan data in arrays so that all the A-scans are aligned in the order (in rows and columns) they were acquired. Therefore, a C-scan image at a certain depth is similar to taking a cross-section of the aggregate A-scan cube horizontally at a specified distance parallel to the part surface. This allows features at the same depth to be easily visualized as shown in Figure 7. Similarly, B- and D-scans allow data to be viewed in a cross-section manner in other planes as illustrated in Figure 8.

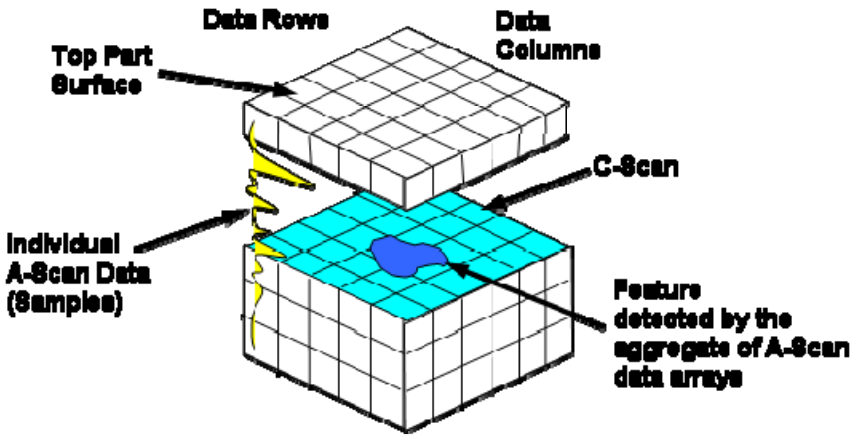


Figure 6. Basic Principle of C-Scan

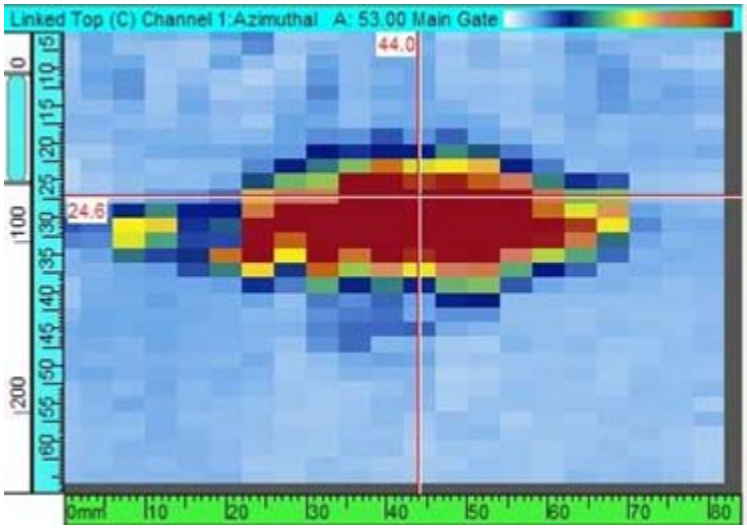


Figure 7. C-Scan Image of Weld Flaw

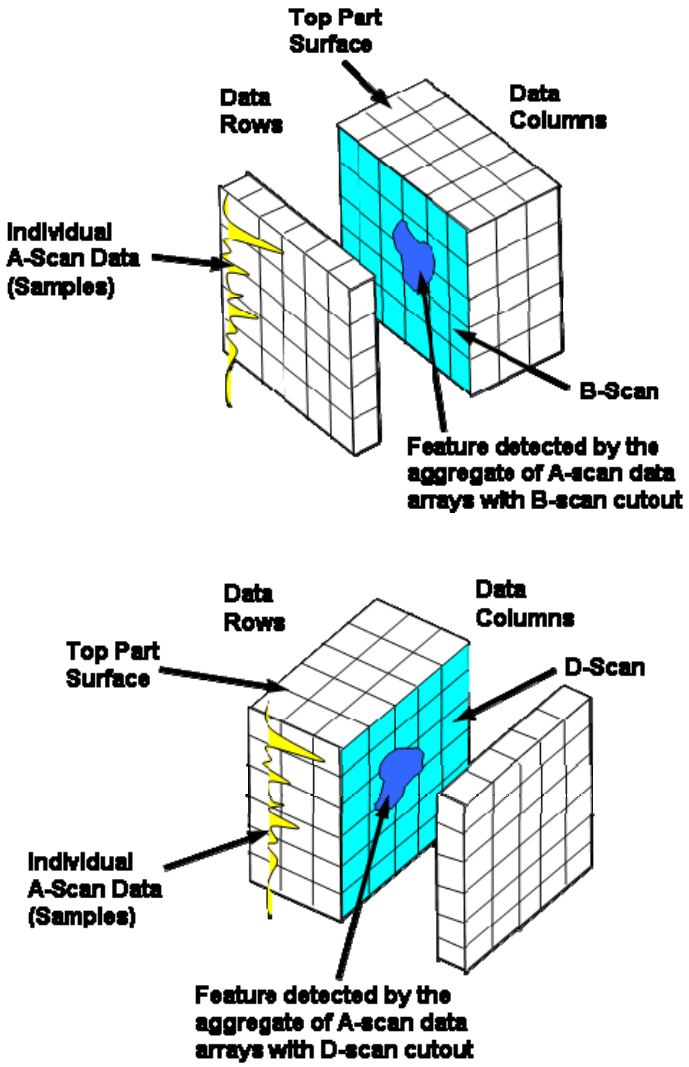


Figure 8. Principle of B- and D-Scans

3.1 Special Type of 2D Display – S-Scans (Sectorial Scans)

S-scans are a special type of 2D display of data that is only available with PA systems. The direct result of using electronic beam sweep is that a triangular fan of 2D area can be covered, and a cross-section can be created based on the data acquired. As a consequence, the S-scan map can provide operators with more information from a single probe location. This can be useful for sizing flaws in the through-wall direction and for detecting off angle flaws. The S-scan concept is depicted in Figures 9 and 10.

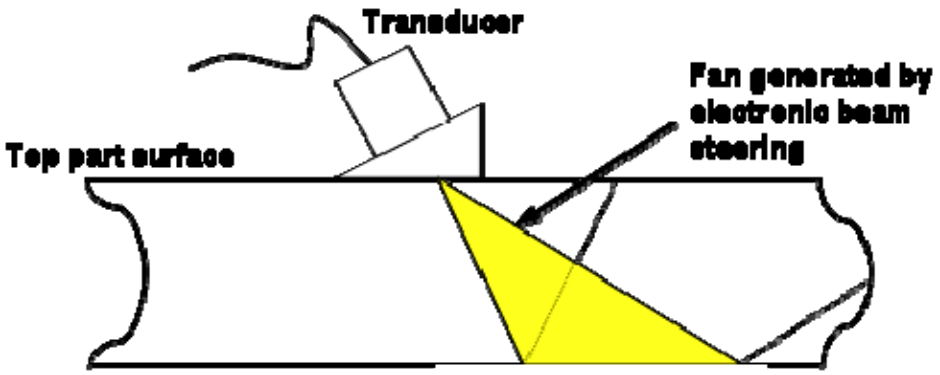


Figure 9. Electronic Scanning and Data to be Displayed

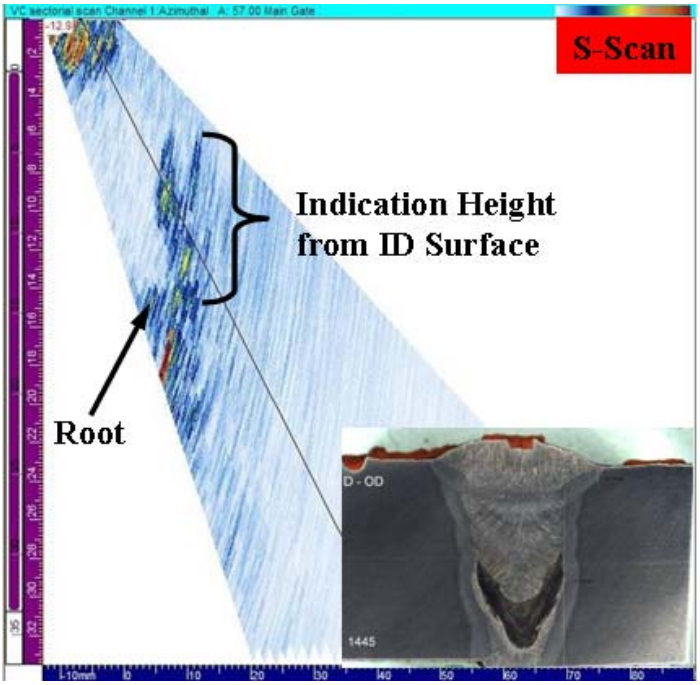


Figure 10. Typical S-Scan Showing Flaw Height Measurement

There are advantages as well as disadvantages for using electronic beam steering to generate S-scans. The advantages include the following:

- Advantages
    - User can gain volumetric information regarding a particular area of interest without moving the transducer, as the firing and receiving angles of a transducer array are steered electronically.
    - A vast amount of data can be generated without a significant amount of motion. For example, to create a 3D map of a certain area, only 1D mechanical motion is required (as the other two dimensions can be controlled electronically.
- Disadvantages:
    - A vast amount of data is available, which requires more powerful equipment (computer processing power, A/D card, storage, etc.).
    - Displaying the data becomes a non-trivial issue. Specifically, when a 2D surface is scanned, it is likely some data will overlap other data.

Displaying and interpreting AUT data is seldom straightforward. As data is combined to form different views, unwanted noise signals can be a hindrance to correct data interpretation. Because of the different element grouping combinations and beam angle combinations that can be produced with a given PA probe and wedge combination, noise in the ultrasonic data is always a possibility. Software routines that can filter unwanted noise without affecting relevant data can be beneficial. As an example, Figure 11 shows a relevant indication appearing at the same depth location as wedge noise signals. A software filtering routine was developed that effectively removed the unwanted data and made the flaw indication more visible. Figure 12 shows the filtered image beginning with the original data in the upper left and progressing in a counterclockwise direction to the final filtered image in the upper right of Figure 12.

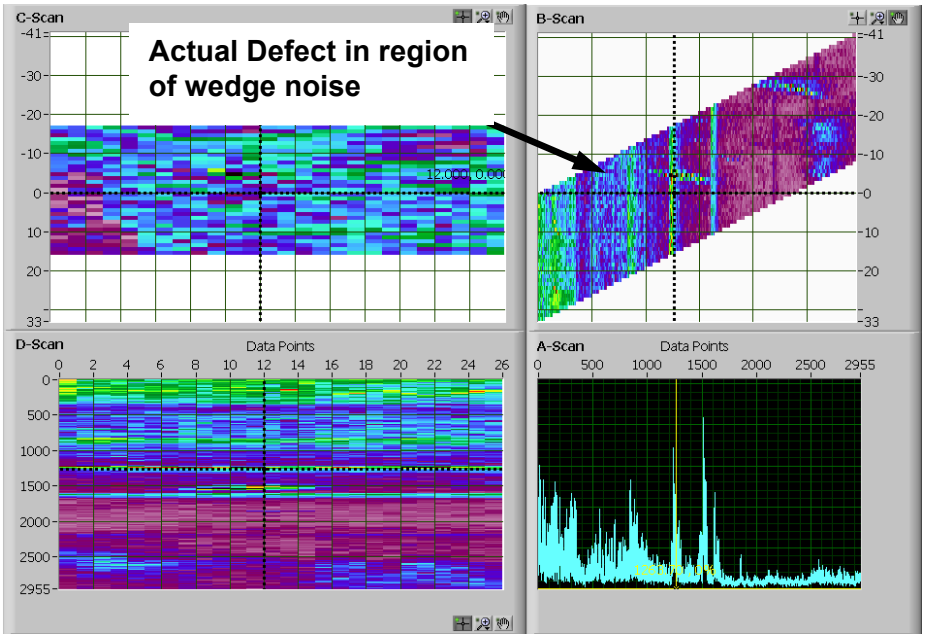


Figure 11. Region Where Wedge Noise Partially Masks Defects

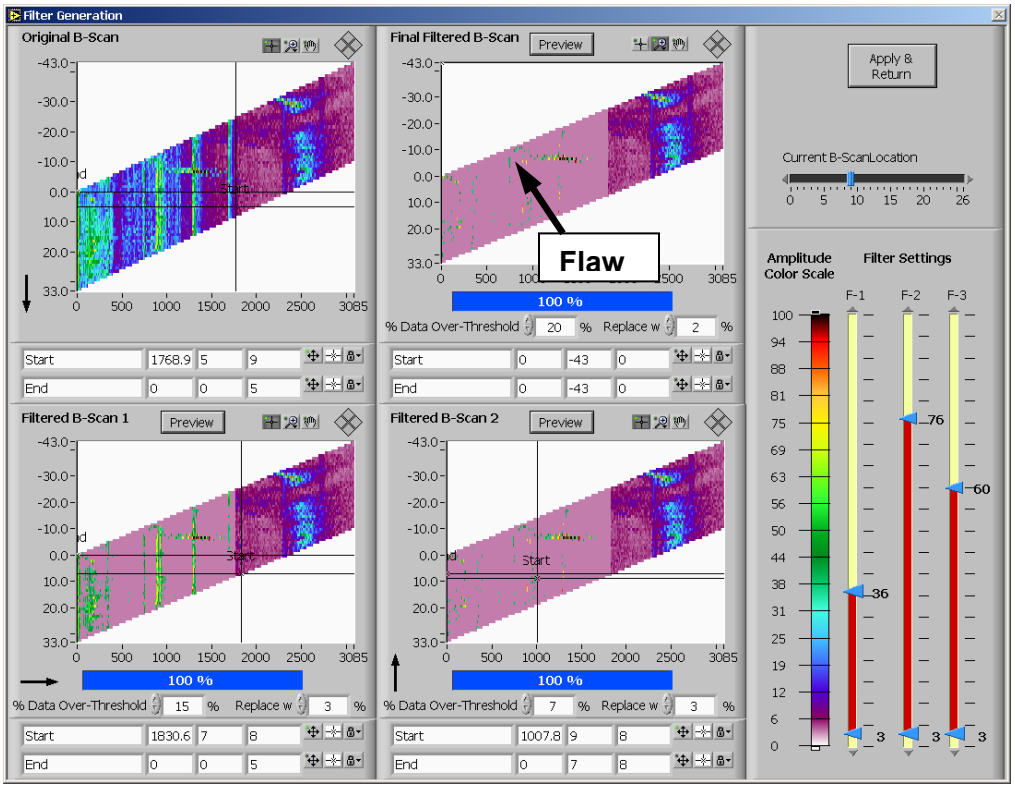


Figure 12. Screen Shot from Wedge Noise Removal Program

Other types of noise removal algorithms include (but are not limited to) localized averaging filters that focus on only a user-defined region (in terms of width and height on a particular B-scan), or filters that use the matrix-weighted average of nearby data values (use the nearby data values in the surrounding matrix (3 × 3, 5 × 5, 7 × 7, 9 × 9, or more) to perform weighted average to attain filtering). In addition, the same types of filtering can also be used with respect to C- or D-scans.

In order to image indications at the proper depth the material thickness must be taken into account. This will require AUT data to be normalized with regard to beam angle and sound travel distance as related to material thickness. For example, Figure 13 shows detection of an OD notch in a 25.4-mm-thick plate. Since the notch was detected on the second leg the notch appears at a depth of 50 mm after correcting for beam angle. As can be seen, the depth indicators show that the data acquired includes UT wave propagated through more than twice the material thickness (as indicated by 1- and 2T). Note that the data is clearly beyond the physical material thickness, thus some manipulation in the data is needed in order to represent the defect geometrically.

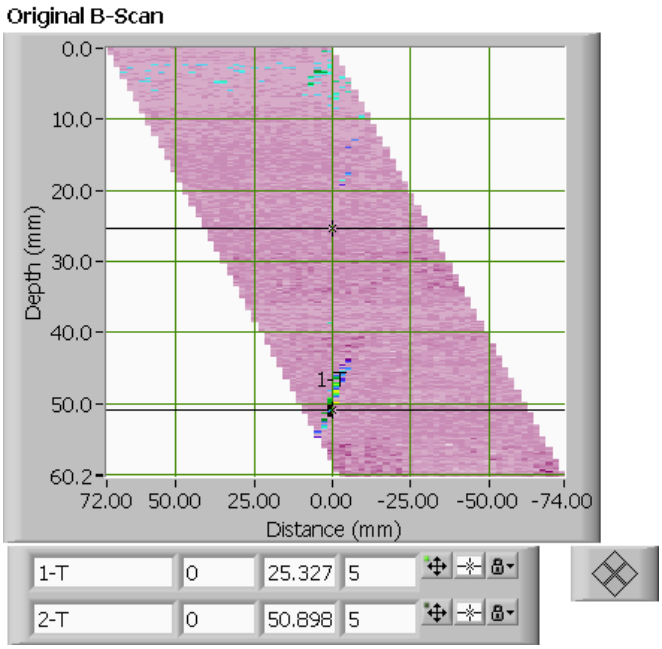


Figure 13. Angle-Corrected B-Scan of a Machined Notch

The simplest form of data manipulation is to “fold” the data according to the angle at which the UT beams are projected, e.g., to overlap the data back to where the data points are supposed to be physically located according to the focal laws that were used to project the UT beams into the material. The result can be seen in Figure 14.

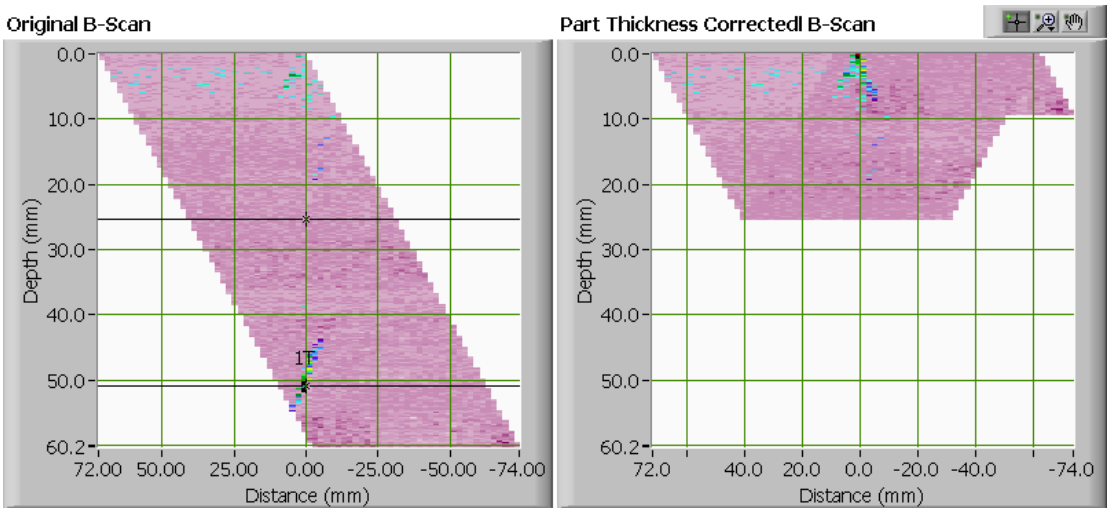


Figure 14. Fold-Back B-Scan Based on Ultrasonic Beam Angle

As can be seen, in the original angle-corrected B-scan, the defect appears to be in the bottom half of the scan, but in reality the defect should have been on the top surface. The “fold back” data manipulation scheme thus overlaps the data back to their “original” physical locations to re-construct the defect map shown in the right of the figure. If all of the B-scans contained in a file go through the same folding procedure and are then stacked end to end, the notch can be imaged in the correct location at the correct depth. This is illustrated graphically in Figure 15.

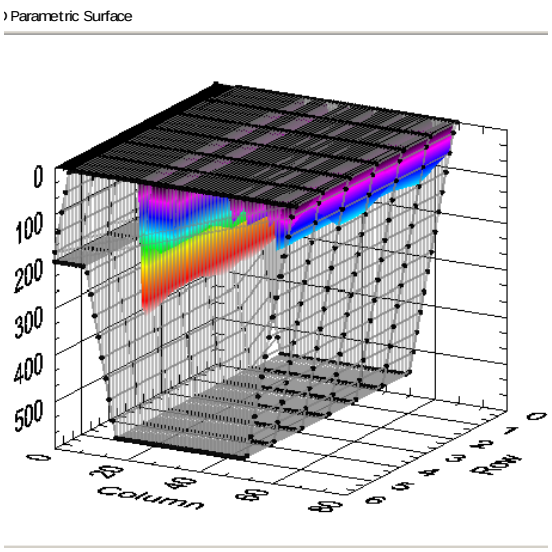


Figure 15. 3D Representation of a Defect Map Using Stacked B-Scans

A previous girth weld inspection study has shown the ability to combine data from different angles, views, and probes.[24] Figure 16 shows an example of fused D-scan views for five different beam angles (45, 50, 55, 60, and 65 degrees) along the entire weld length of a 61.0-cm (24-in.)-diameter pipe having a 7.8-mm (0.31-in.) WT. In Figure 16 the horizontal axis is distance along the weld and the vertical axis is material depth.

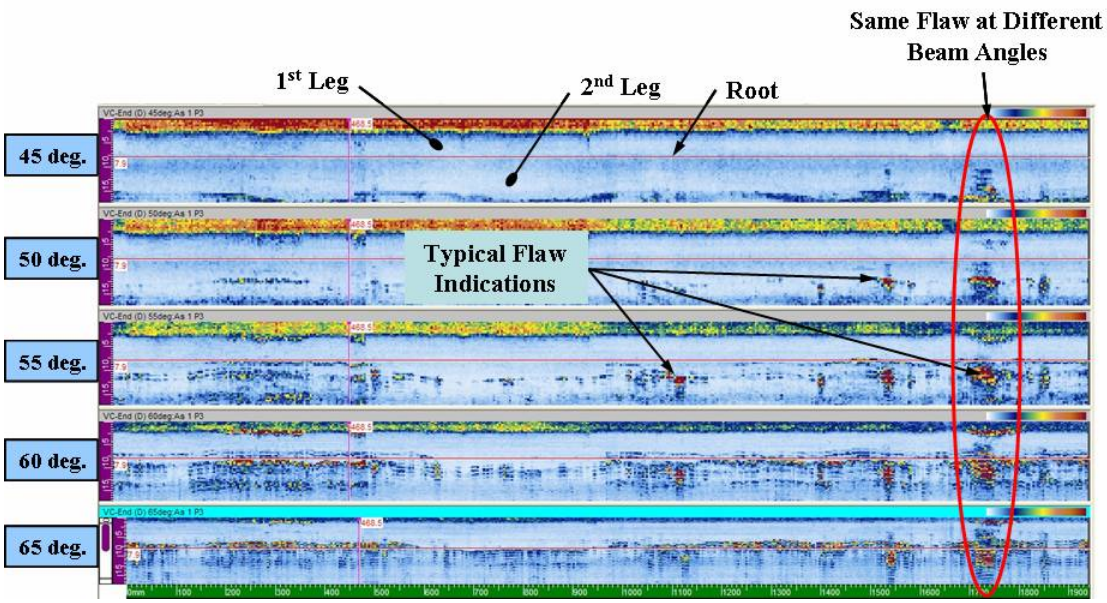


Figure 16. Fused D-Scan Views of 61.0-cm-Diameter Girth Weld

Figure 17 shows another example of combining data. In this example, multiple beam angles were used to construct a combined D-scan view from a through-wall fatigue crack. By using this technique, it was possible to get an idea of the overall shape of the crack which, in turn, was helpful for determining the length of the crack on both the OD and ID surfaces.

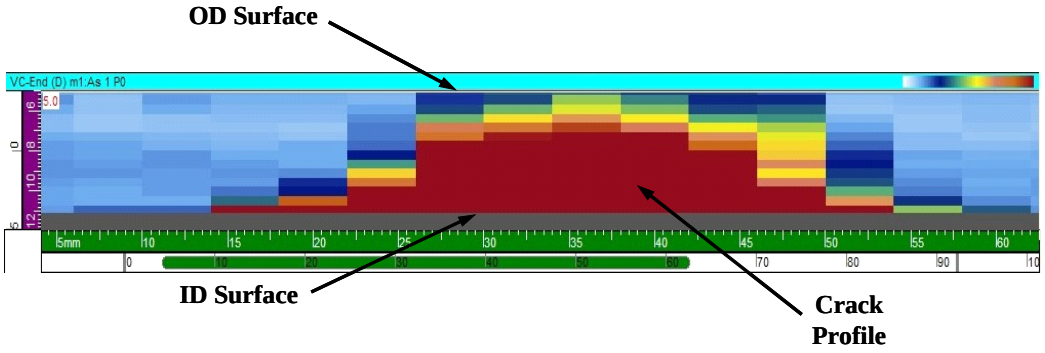


Figure 17. Multi-Angle D-Scan View of a Through-Wall Fatigue Crack in 11.7-mm (0.46-in.)-Thick Pipe Wall

Figures 18 through 22 show a progression of how a large amount of data can be fused and imaged. Figures 18 and 19 show multi-angle D-scan views of the upstream (US) and downstream (DS) scans, respectively. The first leg data is in the upper half of each view and the second leg data is in the lower half. When using these views, it is possible to determine which flaws were detected on the US scan and which were detected on the DS scan. In Figures 20 and 21 the second leg data for the same US and DS scans have been folded up onto the first leg data to give a combined view. Figure 22 goes one step further by merging the US and DS scan data together into a single view.

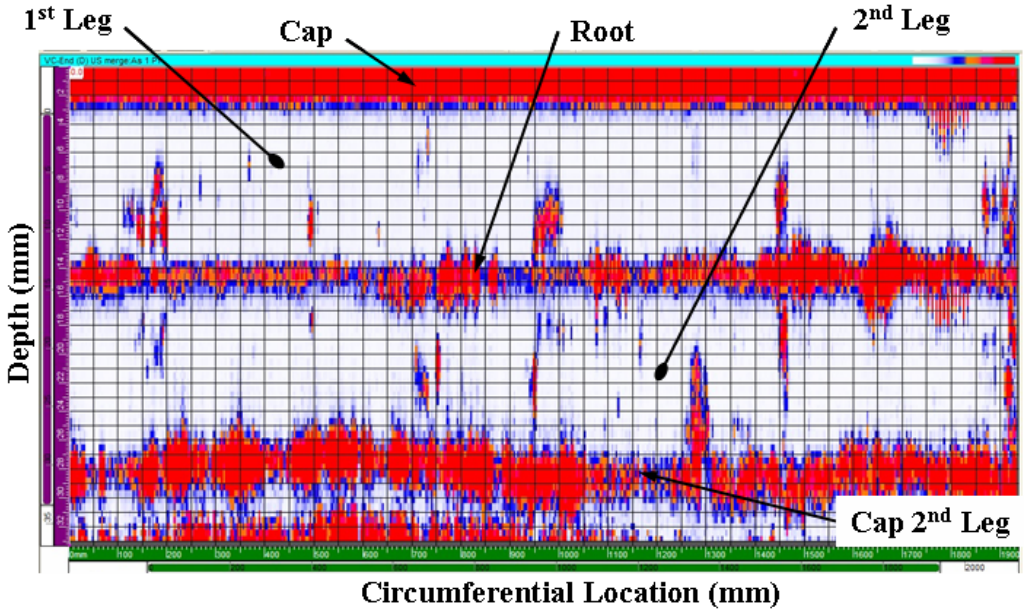


Figure 18. Multi-Angle D-Scan View for US Scan

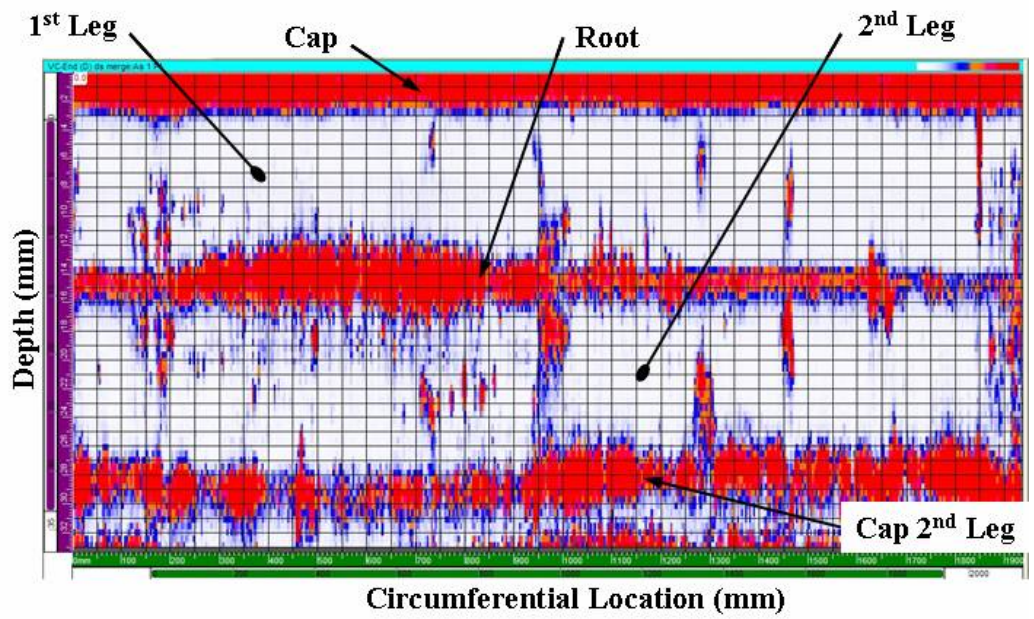


Figure 19. Multi-Angle D-Scan View for DS Scan

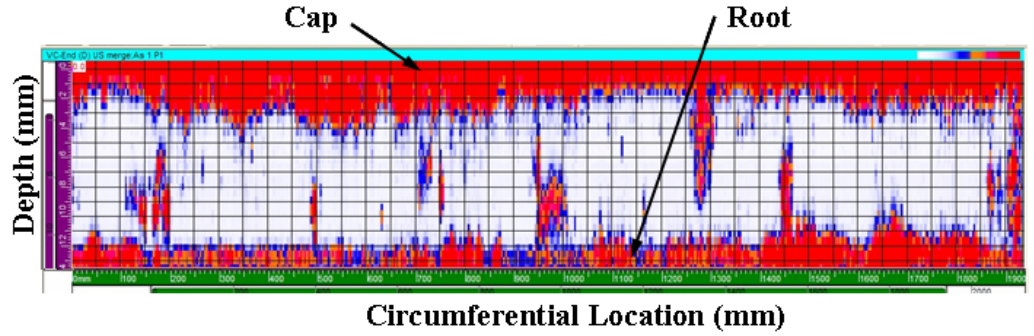


Figure 20. Multi-Angle D-Scan View for US Scan with Second Leg Folded up on First Leg

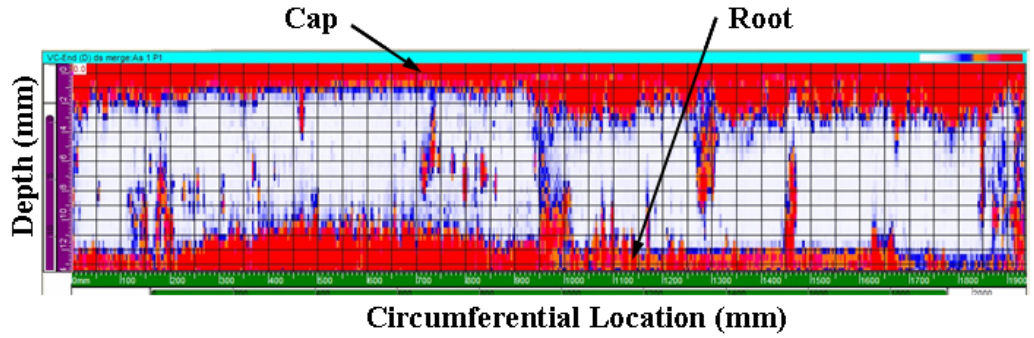


Figure 21. Multi-Angle D-Scan View for DS Scan with Second Leg Folded up on First Leg

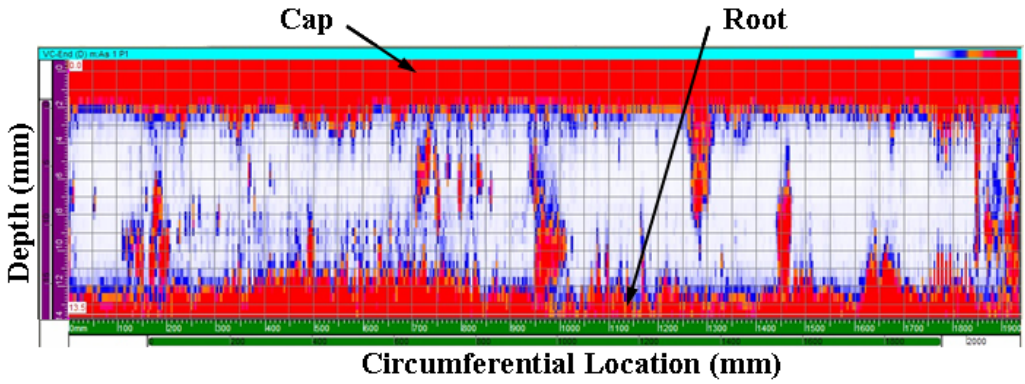


Figure 22. Multi-Angle D-Scan View with US and DS Data Fused and Second Leg Folded up on First Leg

AUT girth weld inspection could likely benefit from using electronic scanning and data merging. This could be accomplished by using multiple angles for each zone and then displaying the data in strip chart fashion that is familiar to AUT operators. Figure 23 shows this type of data display using data obtained from a girth weld calibration zed block. By using these techniques it may be possible to reduce the number of AUT zones and also better detect tilted flaws.

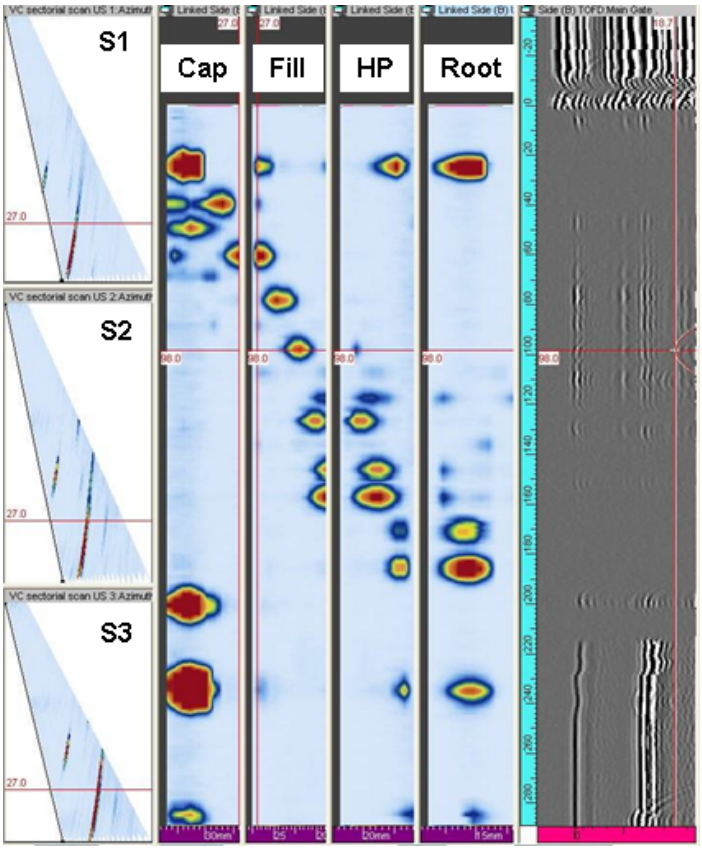


Figure 23. Strip Chart Display of Merged Angle Beam Data Obtained from a Girth Weld Calibration Block

In addition to imaging techniques, it is also possible to use software routines to detect and size flaws as shown in Figure 24. The image in Figure 24 was produced using software developed by EWI which automatically determines the boundary of flaws and outputs an approximate flaw size. While there are obvious challenges with automatic flaw detection and sizing, developments in computer technology, and PA technology makes it conceivable that such tools could be adapted for girth weld inspection within the next decade.

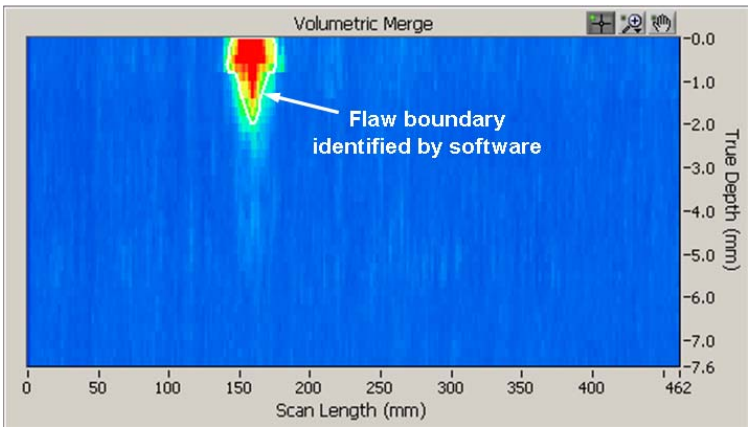


Figure 24. Example of Flaw Detection Using Software Routines on PA Data

3.2 Task Conclusions and Significance

Data imaging plays a crucial role in the accuracy of flaw detection and sizing and further research should be done to look at alternative imaging methods that can take advantage of PA technology. It is foreseeable that the use of filtering techniques and data merging techniques, demonstrated during this project, could provide beneficial assistance to AUT operators to improve the speed and accuracy of AUT girth weld inspection. Whatever type of imaging and data analysis tools are used, it is imperative that such tools be identified and evaluated as an integral part of any AUT qualification/quantification program. In addition, because of the numerous combinations of data display and filtering it is important that AUT companies provide details of how data will be displayed and evaluated in their AUT procedures.

4.0 AUT System Performance Quantification

4.1 Design and Fabrication of Test Welds

Two groups of test welds were designed and fabricated for this project. Each group consisted of girth welds in 30-in.-diameter X80 pipe having a 15.6-mm nominal WT. The first group consisted of four welds containing natural flaws that were created primarily by varying welding parameters to create flaws in a manner consistent with how actual flaws would be produced.

The second group consisted of six welds containing implanted flaws that were more controlled as to flaw type, location, orientation, and size. The circumference of all ten welds was divided into sixty sectors, each having a length of approximately 40 mm. This was done to assist in reporting and subsequent statistical analysis. Each sample was designed so that the weld contained approximately 20 to 25 intentional flaws that were at different depths within the pipe wall. Some sectors contained flaws while others did not. Specifications for through wall flaw heights ranged from 0.5 mm to approximately 12.0 mm, and lengths ranged from approximately 5 to 60 mm. Test welds W5 and W6 were designed with several flaws in close proximity to each other in order to evaluate the effects of flaw interaction.

Early fingerprinting and verification testing of the four natural flaw samples revealed that the intended flaws were generally more severe than anticipated and that several unintentional and undesirable flaws had been created during the fabrication process. The number and size of flaws made interpretation very difficult and time consuming for determining POD and sizing accuracy. Consequently, it was decided to use the implanted flaw samples for blind trial scans related to POD and sizing accuracy. Limited scans were conducted on the natural flaw samples for demonstration purposes only. Implanted flaw sample W1 was designated as an open trial sample to allow AUT operators to test out their equipment and procedures prior to performing the blind trials.

4.1.1 Fingerprinting

Following fabrication of the ten samples, EWI performed visual testing (VT), fluorescent magnetic particle testing (MT), radiography testing (RT), and PA ultrasonic scans on the welds to document both intentional and unintentional flaws. This process was referred to as fingerprinting the samples, and was later used in conjunction with destructive tests to determine actual flaw locations and dimensions. Visual and fluorescent MT was performed on the weld and heat-affected zone (HAZ) on both the OD and ID surfaces. RT was done using film and computed radiography (CR) phosphorus plates. PA UT scans were conducted using a raster scan technique to cover the entire weld circumference from both the US and DS sides of the weld. In addition to PA, the welds were also scanned with TOFD and transverse probes. The PA setup for fingerprinting was a sector scan technique where the ultrasonic beam was electronically steered through refracted shear wave angles of 40 to 70 degrees in 1-degree steps. The PA probe was a 60-element probe having a 7.5-MHz frequency. Instrument gain and distance settings were established using 2-mm-diameter flat-bottom holes

During fingerprinting it was noted that weld Samples W2 and W3 contained locations where the wall thickness had been reduced by grinding. This was done by the sample fabricator to facilitate fit up prior to welding. Estimates from the fingerprinting results showed the wall thinning to be as much as 2 mm. This equated to approximately 12.8 percent of the nominal wall thickness which was greater than what is allowable for wall thickness deviations per API

5L, “Specification for Line Pipe”. For pipe having an OD of 20 inches or greater API specification 5L stipulates a wall thickness tolerance of +19.5, -8% for X80 welded pipe used for this project. It was decided to use flaws in the thin locations to evaluate the effect of wall thinning on AUT results. If AUT results showed a significant effect, flaws in these locations would be excluded from the POD and sizing data.

4.1.2 Linear PA AUT Setups and Scanning

Five linear PA AUT setups were evaluated during this project. Rather than specifying a setup or procedure, each participant was simply asked to scan the welds and report flaws having a length greater than 5 mm. Calibration samples and test procedures were open to what each participant decided was appropriate for the application. The AUT setups used by the different participants were designated by EWI as Setups 1 through 5 and are summarized in Table 2. While the goal of the project was to evaluate AUT systems, it is well known that expertise of the AUT operator is a major factor in AUT of girth welds and is difficult to separate from evaluation of equipment and procedures.

Table 2. Summary of Linear PA AUT Setups

Setup No.	Zonal	Non-Zonal	No. of Fill Zones	PA Freq.	Cal. Sample(s)	Comments
1		X	--	7.5	2-mm FBHs (not Zed block)	3 staggered sector scans (45-65 degrees)
2	X		4	5.0	2-mm FBHs (Zed block)	
3		X		5.0	NAVSHIP S block	1.2-mm SDHs. Zed block used for setup verification. Single-sector scan (35-70 degrees).
4	X		6	5.0	2-mm FBHs (Zed block)	
5	X		4	5.0	2-mm FBHs (Zed block)	

The AUT setups reflected the background and previous experience of the different participants. For example, three of the five chose zonal girth weld inspection setups that are commonly used for this type of inspection and were in accordance with ASTM 1961. These three participants all had considerable expertise with zonal girth weld inspection. The two remaining participants chose to use electronic sector scanning, although in different ways. One setup (Setup 1) consisted of a series of three electronic sector scans from both the US and DS sides of the weld, combined with a mechanical line scan around the pipe circumference. The other sector scan setup (Setup 3) used a single electronic sector scan combined with a two axis, mechanical raster scan to cover the full weld.

All setups had at least one TOFD channel, and all setups, except for Setup 3, used scans for detection of transverse weld flaws. The transverse scans were conducted for demonstration purposes only and were not included in POD and sizing accuracy determinations. Following is a brief description of each AUT setup.

- **Setup 1:** Time-based technique having three electronic sector scans with refracted shear wave angles of 45 to 65 degrees in 1-degree steps. Each sector scan had a different element start position on the probe to provide full coverage of the weld without performing a mechanical scan perpendicular to the weld. This allowed the weld to be inspected using a single circumferential line scan similar to current girth weld inspection techniques. A TOFD scan was used as a supplementary technique. Mechanical scanning was accomplished using a band scanner. A detailed overview of Setup 1 is provided in Table 3.

- **Setup 2:** ASTM E1961 zonal technique using electronic steering and focusing to control the beam position and height for each zone. Zones consisted of root, cap, and two hot pass zones, as well as four fill zones. In addition, there were six volumetric channels used. TOFD was used as a supplementary technique. Calibration was accomplished using a "Zed block" containing 2-mm-diameter flat-bottom holes for fill channels and 1.5-mm-diameter flat-bottom holes for volumetric channels. Mechanical scanning was accomplished using a band scanner. A detailed overview of Setup 2 is provided in Table 4.

- **Setup 3:** Time-based technique having a single electronic sector scan with refracted shear wave angles of 35 to 70 degrees in 1-degree steps. A 2-axis mechanical scan was used to obtain full coverage of the weld. Setup 3 used TOFD as the primary inspection technique and PA sector scans as a supplementary technique. A detailed overview of Setup 3 is provided in Table 5.

- **Setup 4:** ASTM E1961 zonal technique using electronic steering and focusing to control the beam position and height for each zone. Zones consisted of root, cap, and two hot pass zones, as well as six fill zones. In addition, there were eight volumetric channels used. TOFD was used as a supplementary technique. Calibration was accomplished using a Zed block containing 2-mm-diameter flat-bottom holes for fill channels and 1.5-mm-diameter flat bottom holes for volumetric channels. Mechanical scanning was accomplished using a band scanner. A detailed overview of Setup 4 is provided in Table 6.

- **Setup 5:** ASTM E1961 zonal technique using electronic steering and focusing to control the beam position and height for each zone. Zones consisted of root, cap, and two hot pass zones, as well as, four fill zones. In addition, there were six volumetric channels used. TOFD was used as a supplementary technique. Calibration was accomplished using a Zed block containing 2-mm-diameter flat-bottom holes for fill channels and 1.5-mm-diameter flat-bottom holes for volumetric channels. Mechanical scanning was accomplished using a band scanner. A detailed overview of Setup 5 is provided in Table 7.

Table 3. Linear PA AUT Setup 1

Zone	Technique	Calibration Targets	Beam Angle (degrees)		No. of Elements		Leg	
			T	R	T	R	T	R
Sector 1	P/E	1 × 12 notch	45-65	--	13	--	Full skip	--
	P/E	2.0 FBH at 45 degrees						
	P/E	2.0 FBH at 55 degrees						
	P/E	2.0 FBH at 65 degrees						
Sector 2	P/E	1 × 12 notch	45-65	--	13	--	Full skip	--
	P/E	2.0 FBH at 45 degrees						
	P/E	2.0 FBH at 55 degrees						
	P/E	2.0 FBH at 65 degrees						
Sector 3	P/E	1 × 12 notch	45-65	--	13	--	Full skip	--
	P/E	2.0 FBH at 45 degrees						
	P/E	2.0 FBH at 55 degrees						
	P/E	2.0 FBH at 65 degrees						
TOFD OD	T/R	3.9 × 12 notch	60L	60L	--	--	1	1
TOFD ID	T/R	3.9 × 12 notch	60L	60L	--	--	1	1
TR OD	T/R	3.9 × 12 notch	60	60	--	--	2	2
TR ID	T/R	3.9 × 12 notch	60	60	--	--	1	1

Table 4. Linear PA AUT Setup 2

Zone	Technique	Calibration Targets	Beam Angle (degrees)		No. of Elements		Leg	
			T	R	T	R	T	R
Cap	P/E	1 × 10 notch	65	--	17	--	2	--
Fill 4	Tandem	2.0 FBH at 4 degrees	50	58	15	16	3	2
Fill 3	Tandem	2.0 FBH at 4 degrees	50	58	15	16	3	2
Fill 2	Tandem	2.0 FBH at 4 degrees	50	58	16	16	3	2
Fill 1	Tandem	2.0 FBH at 4 degrees	50	58	16	16	3	2
HP 2	P/E	2.0 FBH at 45 degrees	62	--	15	--	2	--
HP 1	P/E	2.0 FBH at 45 degrees	50	--	16	--	4	--
Root	P/E	1 × 10 notch	65	--	15	--	1	--
Vol. F4	P/E	1.5 FBH at 45 degrees	45	--	17	--	2	--
Vol. F3	P/E	1.5 FBH at 45 degrees	47	--	17	--	2	--
Vol. F2	P/E	1.5 FBH at 45 degrees	49	--	17	--	2	--
Vol. F1	P/E	1.5 FBH at 45 degrees	51	--	17	--	2	--
Vol. HP	P/E	1.5 FBH at 45 degrees	53	--	17	--	2	--
Vol. Root	P/E	0.5 × 7 notch	65	--	14	--	1	--
TOFD 1 OD	T/R	3 × 10 notch	60L	60L	--	--	1	1
TOFD 1 ID	T/R	1 × 10 notch	60L	60L	--	--	1	1
TOFD 2 OD	T/R	3 × 10 notch	70L	70L	--	--	1	1
TOFD 2 ID	T/R	1 × 10 notch	70L	70L	--	--	1	1
TR OD	None	--	--	--	--	--	--	--
TR ID	T/R	1 × 10 notch	60	60	--	--	1	1

Table 5. Linear PA AUT Setup 3

Zone	Technique	Calibration Targets	Beam Angle (degrees)		No. of Elements		Leg	
			T	R	T	R	T	R
Sector scan	P/E	1.2-mm SDH at a depth of 12.7 mm	35-70	--	16	--	Full skip	--
TOFD OD	T/R	Lateral wave at 60% full-scale height (FSH)	60L	60L	--	--	1	1
TOFD ID	T/R	Lateral wave at 60% FSH	60L	60L	--	--	1	1
TR OD	None	--	--	--	--	--	--	--
TR ID	None	--	--	--	--	--	--	--

Table 6. Linear PA AUT Setup 4

Zone	Technique	Calibration Targets	Beam Angle (degrees)		No. of Elements		Leg	
			T	R	T	R	T	R
Cap	P/E	1 × 10 notch	55	--	30	--	2	--
Fill 6	Tandem	2.0 FBH at 4 degrees	45	55	32	32	2	3
Fill 5	Tandem	2.0 FBH at 4 degrees	45	55	32	32	2	3
Fill 4	Tandem	2.0 FBH at 4 degrees	45	55	30	30	2	3
Fill 3	Tandem	2.0 FBH at 4 degrees	45	55	31	31	2	3
Fill 2	Tandem	2.0 FBH at 4 degrees	46	56	28	28	2	3
Fill 1	Tandem	2.0 FBH at 4 degrees	45	--	24	24	2	3
HP 2	P/E	2.0 FBH at 45 degrees	48	--	--	--	2	--
HP 1	P/E	2.0 FBH at 45 degrees	49	--	--	--	2	--
Root	P/E	1 × 10 notch	65	--	--	--	1	--
Vol. F6	P/E	1.5 FBH at 45 degrees	45	--	--	--	2	--
Vol. F5	P/E	1.5 FBH at 45 degrees	45	--	--	--	2	--
Vol. F4	P/E	1.5 FBH at 45 degrees	45	--	--	--	2	--
Vol. F3	P/E	1.5 FBH at 45 degrees	45	--	--	--	2	--
Vol. F2	P/E	1.5 FBH at 45 degrees	45	--	--	--	2	--
Vol. F1	P/E	1.5 FBH at 45 degrees	45	--	--	--	2	--
Vol. HP	P/E	1.5 FBH at 20 degrees	45	--	--	--	2	--
Vol. Root	P/E	3 × 10 notch	70	--	--	--	1	--
TOFD OD	T/R	3 × 10 notch	60L	60L	--	--	1	1
TOFD ID	T/R	1 × 10 notch	60L	60L	--	--	1	1
TR OD	TR	3 × 10 notch	60	60	--	--	2	2
TR ID	T/R	3 × 10 notch	60	60	--	--	1	1

Table 7. Linear PA AUT Setup 5

Zone	Technique	Calibration Targets	Beam Angle (degrees)		No. of Elements		Leg	
			T	R	T	R	T	R
Cap	P/E	1 × 5 notch	55	--	32	--	2	--
Fill 4	Tandem	2.0 FBH at 4 degrees	45	53	23	30	2	3
Fill 3	Tandem	2.0 FBH at 4 degrees	50	58	32	32	2	3
Fill 2	Tandem	2.0 FBH at 4 degrees	50	58	32	32	2	3
Fill 1	Tandem	2.0 FBH at 4 degrees	50	57	29	32	2	3
HP 2	P/E	2.0 FBH at 20 degrees	68	--	32	--	2	--
HP 1	P/E	2.0 FBH at 35 degrees	46	--	32	--	4	--
Root	P/E	1 × 5 notch	67	--	32	--	1	--
Vol. F4	P/E	1.5 FBH at 45 degrees	46	--	32	--	2	--
Vol. F3	P/E	1.5 FBH at 43 degrees	47	--	32	--	2	--
Vol. F2	P/E	1.5 FBH at 41 degrees	49	--	32	--	2	--
Vol. F1	P/E	1.5 FBH at 39 degrees	51	--	32	--	2	--
Vol. HP 2	P/E	1.5 FBH at 37 degrees	53	--	32	--	2	--
Vol. HP 1	P/E	1.5 FBH at 35 degrees	55	--	32	--	2	--
Vol. Root	P/E	0.5 × 5 notch	64	--	32	--	1	--
TOFD 1 OD	T/R	3 × 10 notch	60L	60L	--	--	1	1
TOFD 1 ID	T/R	1 × 10 notch	60L	60L	--	--	1	1
TOFD 2 OD	T/R	3 × 10 notch	70L	70L	--	--	1	1
TOFD 2 ID	T/R	1 × 10 notch	70L	70L	--	--	1	1
TR OD	T/R	1 × 10 notch	60	60	--	--	2	2
TR ID	T/R	1 × 10 notch	60	60	--	--	1	1

For each evaluation the AUT operators were asked to scan the open trial implanted flaw sample (W1) to check the operation of the systems and to make any adjustments. Each participant used their respective AUT setups to scan the open trial implanted flaw sample. This provided feedback that the setups were performing as designed on a weld of the same construction as those to be used for blind trial testing. After the setups were performing to the operator's satisfaction, the five blind trial welds were provided for scanning. The welds were scanned and the AUT data evaluated by the operators. During evaluation the operators recorded size and location of each flaw in a spreadsheet format that was later used for statistical comparisons.

4.1.3 Destructive Evaluation

Following completion of all AUT scanning activities, DT of the implanted flaw samples began. Five implanted flaw samples (W2 through W6), that were part of the blind trial study, were cut down to a manageable size using a combination of flame cutting and saw cutting. Flame cutting allowed a ring, containing the weld, to be removed from the pipe section. Flame cuts were made approximately 100 mm (4 in.) on each side of the weld. For each weld, the resulting ring was then saw cut into four smaller segments and excess material on each side of the weld was removed. As part of this process, saw cuts were made parallel to the weld on both the US and downstream DS sides of the weld. Cuts on the US side were approximately 25 mm (1 in.) from weld centerline while cuts on the DS side were approximately 13 mm (0.5 in.) from weld centerline as shown in Figure 25. This was done so that the US and DS sides were easily identifiable throughout the DT process.

Figure 25. Example of Saw Cut Segment

After saw cutting, the sample segments were then prepared for water jet cutting to remove weld cross-sections for subsequent metallographic preparation and measurement. In preparation for removing metallographic sections from the welds, tests were conducted with the water jet cutting system to determine the width (kerf) of the water jet cuts. In addition, previous data was reviewed to determine the amount of material that would be removed from each sample during grinding and polishing. Results of these evaluations showed that the water jet kerf was consistently 0.8 mm (0.03 in.) and that approximately 0.5 mm (0.02 in.) would be removed during metallographic processes related to grinding and polishing. Based on these results, it was decided to make water jet cuts at 3-mm (0.12-in.) intervals along the length of each flaw to be measured during the DT. After cutting, grinding, and polishing this resulted in weld cross-section samples having a thickness of approximately 2 mm (0.08 in.).

Cutting locations were selected based on a compilation of intended flaw location from the fabrication drawings and actual locations reported by multiple AUT scans conducted during the project. The number of cross-sections for each flaw was determined by the flaw length and any skewing of the flaw. For flaws with no skewing, the typical minimum number of cross-sections is shown in Table 8. When cross-sectioning skewed flaws, metallographic samples were generally removed every 3 mm along the length of the flaw.

Table 8. Number of Metallographic Cross-Sections for Non-Skewed Flaws

Flaw Length (mm)	Typical Number of Cross-Sections
< 8	1
8-12	2
>12	3

For flaws less than 8 mm (0.31 in.), one sample was removed at the flaw center. Generally, for flaw lengths of 8 to 12 mm (0.31 to 0.47 in.), a minimum of two samples were removed with one being near flaw center and the second near the beginning of the flaw. For flaw lengths greater

than 12 mm, at least three samples were typically removed with one being at flaw center and the other two taken 3 mm on each side of flaw center. Cross-sections were also removed from selected locations where UT indications were noted by multiple AUT systems, or where radiography indications were present. In addition, samples were also removed in locations where no flaws were planned or detected.

Circumferential weld positions on the samples were relative to the center of the water jet cut. As each sample was water jet cut, both faces of the thin samples were identified by weld number and circumferential weld position relative to the zero scan start position. Since water jet cuts were made at 3-mm (0.12-in.) intervals, the identification on opposing faces of the same sample differed by 3 mm. For example, Sample 2-100 would represent Weld No. 2, Location 100. This sample would be identified on one face as 2-100 and as 2-103 on the opposite face.

After removal of the metallographic samples by water jet, photographs were taken of the samples in the as-cut condition. It was thought that these photographs may help with flaw identification and measurements by allowing access to both sides of the thin samples. Review of these photographs revealed that most flaws were typically not visible in the photographs (Figure 26); consequently, the process of photographing the as cut surfaces was discontinued.

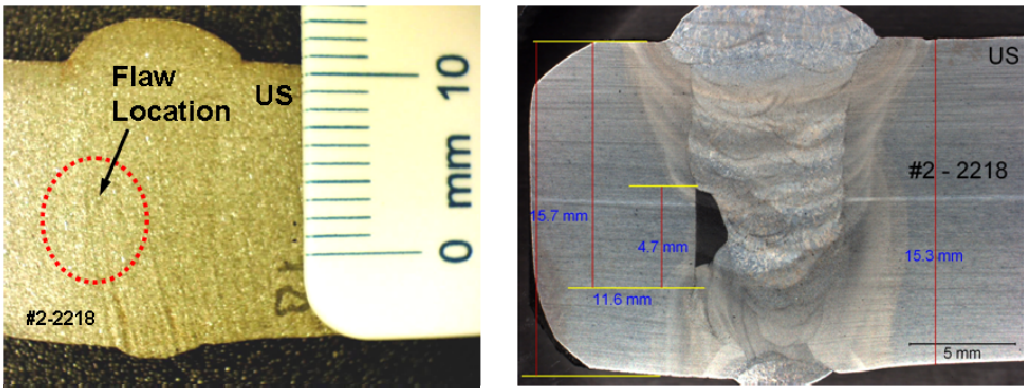


Figure 26. Photos Before (Left) and After (Right) Grind, Polish, and Etch

Prior to grinding and polishing, each sample was radiographed so that the flaw height could be seen through the 3-mm sample thickness. Radiography was conducted using CR to obtain a digital image (Figure 27). This technique helped provide good verification of flaw height measurements obtained by metallography, as well as detection and location of other unintended flaws.

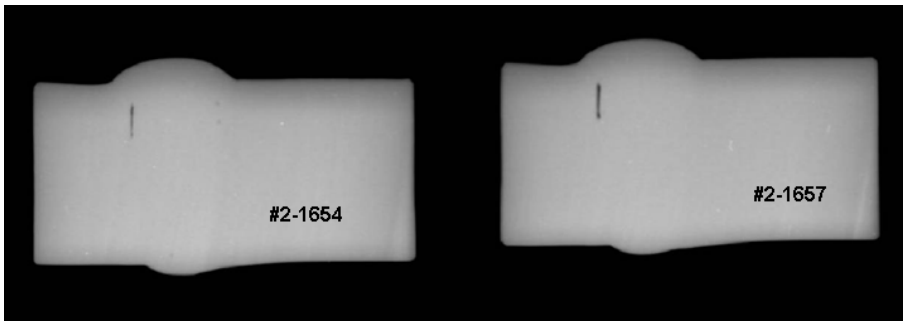


Figure 27. Example of CR on Un-Polished Water Jet Cut Samples

Metallography of the samples was accomplished by trimming the samples to fit within 38-mm (1.5-in.)-diameter mounts and oriented so that the sample face corresponding to the lowest circumferential weld position was polished for evaluation. Samples were polished to a final surface finish using 0.05-μm polishing compound and then lightly etched using a 2% nital etch solution. Photographs were then taken of each weld cross-section sample at a magnification of 5×. Linear measurements were made to determine the depth and height of flaws observed on the samples. Additional measurements of WT and misalignment were also obtained in locations where these parameters could have a possible effect on the AUT results. Figure 28 shows typical measurements performed on metallographic weld cross-sections.

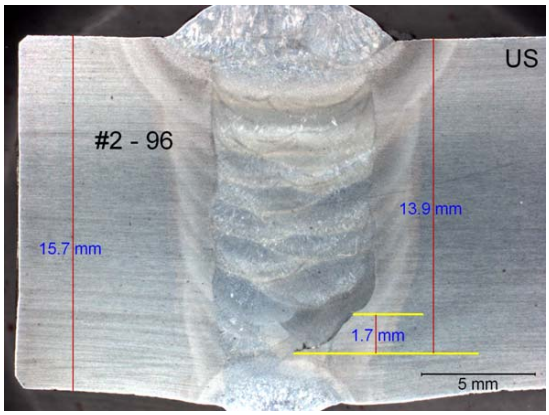


Figure 28. Polished and Etched Weld Cross-Section Showing Typical Measurements

As the metallographic weld cross-sections were evaluated, it became apparent that higher magnification was needed to determine the extent of some flaws. Lack of sidewall fusion flaws in particular were often difficult to view at 5× (Figure 29). This led to changing the measurement technique so that higher magnifications of up to 100× were used to find the extent of flaws prior to photographing at 5×.

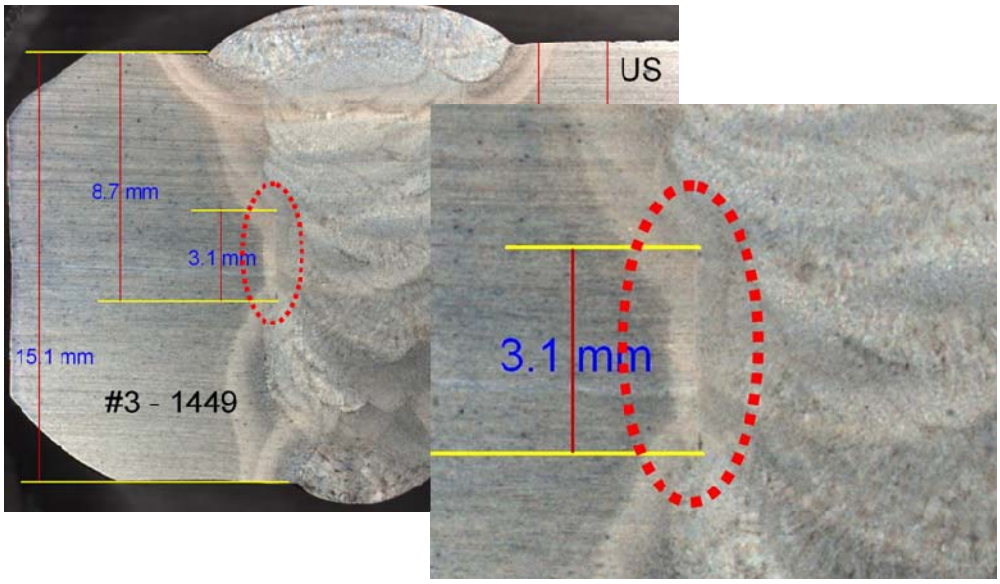


Figure 29. Example of Difficulty in Viewing Tight Flaws at 5×

Flaw dimensions were obtained from the macro photographs and were input into spreadsheet format for statistical analysis.

4.1.4 Task Conclusions and Significance

During the course of this task, some conclusions came to light and are summarized below:

- The process of making controlled flaws in the samples was difficult using changes in welding parameters alone. Samples made by this method resulted in relatively large flaws that were not useful for POD and sizing accuracy studies. It was found that implanted flaws could be controlled more accurately as to length, height, depth, and flaw type. The major disadvantage is that implanted flaws may not be produced by the welding process used during actual pipeline welding.

- There were a few attempts to make short flaws that would generally be shorter in length than those considered rejectable by most ECA acceptance criteria. When flaws became shorter than the ultrasonic beam width, ultrasonic sizing became dependant on both flaw length and flaw height. In order to adequately evaluate flaw height-sizing accuracy as an independent quantity, flaw lengths should be longer than the ultrasonic beam width.

- Fingerprinting proved useful as a means of comparing test results. In particular, the PA raster scans and TOFD scans helped to locate both intentional and unintentional flaws, as well as, providing additional flaw position information for the DT.

- During DT it was found that it was easier to keep track of US and DS sides of the weld by leaving the base metal longer on one side of the weld. For this project the US side was longer than the DS side.

- Typically, the ultrasonic beam size ("acoustical footprint") was at least 2-3 mm in through-wall height and 5-10 mm in width for each AUT technique used. This resulted in each AUT data point being an average of this acoustical footprint at each data collection step.

- During the salami sectioning, water jet cutting was used to make cuts approximately 3 mm apart. Other options such as EDM, milling, and grinding were considered but water jet cutting was preferred. EDM was not used because of higher cost and the higher possibility of destroying samples due to wire breaks. Since most cross-section locations contained flaws, it was recommended by EWI's EDM source that EDM not be used. Grinding and milling was a good option; however, the cross-sections could not have been preserved, which was an important consideration for this project.

- Some flaws were tilted and/or skewed which influenced uncertainties in AUT results. Based on observations from the data, smaller steps in the DT would not have significantly influenced these results.

- Radiographs of the water jet slices proved to be beneficial as a verification of flaw location, depth, and height.

- In several cases it was necessary to view the weld cross-sections at magnifications up to 100× to determine the full extent of the flaws. Since metallography is used as the reference measurement for "actual" flaw dimensions; the POD and sizing accuracy will only be as accurate as the metallography.

4.2 AUT Technique Evaluation Using Modeling and Inspection Simulations

Since the beam produced by PA UT is dependent on several parameters such as frequency, element size, active aperture size, and focal depth; modeling and inspection simulations were conducted to evaluate the effects of these parameters on flaw detection and sizing. This modeling and simulation approach has been successfully used in the past to perform similar evaluations.[3-4,25] For this project, one AUT procedure was selected for evaluation using the modeling tools. Later in the project, additional modeling was conducted to better understand the effects of various PA parameters on the overall inspection and in an effort to narrow parameter ranges that would be applicable for pipeline girth weld inspection. Due to budget constraints, only representative inspection zones were selected for modeling. A more thorough modeling effort would be needed to access all inspection zones.

4.2.1 Evaluation Approach

Using the selected AUT procedure four zones were initially evaluated using modeling and simulation. These four zones consisted of root, hot pass, fill zone, and cap zone. The modeling software used was a semi-analytical software called CIVA. The specific zones were identified in the AUT procedure as RT, HP1, F2, and Cap as shown in Figure 30. PA parameters and zone information is provided in Table 9. The pipe thickness and weld bevel were the same as those used later in the project for POD and sizing studies (15.6-mm WT; 4-degree bevel).

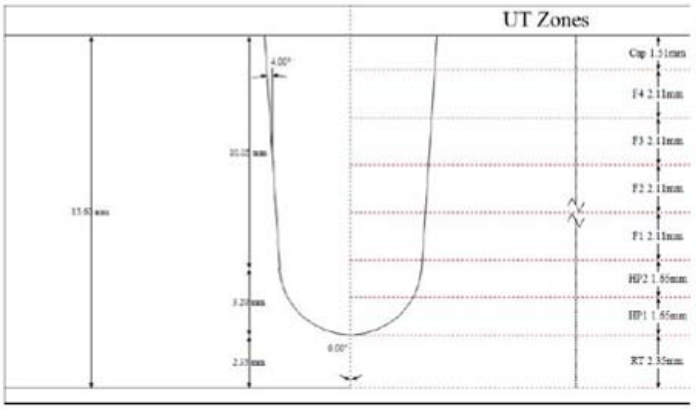


Figure 30. UT Zone Layout Described in Selected AUT Procedure

Table 9. AUT Procedure Parameters Used for Modeling

Zone	Zone Height (mm)	Depth (mm)	Focus Depth (mm)	Start Element	No. of Elements	Beam Angle (degrees)	Leg
Cap	1.51	1.51	30.4	1	32	63	2
F2 T	2.11	7.84	24.4	29	32	50	2
F2 R	2.11	7.84	37.6	2	32	58	3
HP1	1.65	13.25	18.3	30	32	55	2
RT	2.35	15.6	15.2	15	32	70	1

The procedure provided to EWI specified a 5-MHz, 64-element probe mounted on a wedge that produced a natural refracted beam angle of 55 degrees in steel. The wedge had a 100-mm radius in the secondary (passive) probe axis to mechanically focus the beam along the weld length. Beam models showed that the 100-mm radius provided a fixed focus at a sound path of approximately 42 mm. This sound path distance would result in a focus depth of approximately 25 mm at a beam angle of 55 degrees.

Beam models were calculated for each zone which provided information about the beam shape and size. Figure 31 shows an example of a cross-sectional view of a beam at a given beam

angle and sound path distance. These cross-sectional views allowed the beam height and width to be measured at the sound path for each zone. In cases were the beam was not perpendicular to the weld bevel, the actual beam projected on the weld bevel would be slightly larger than those shown in the cross-sectional views. Where possible, beam profiles were also calculated which showed the beam focal spot position relative to the zone target (Figure 32).

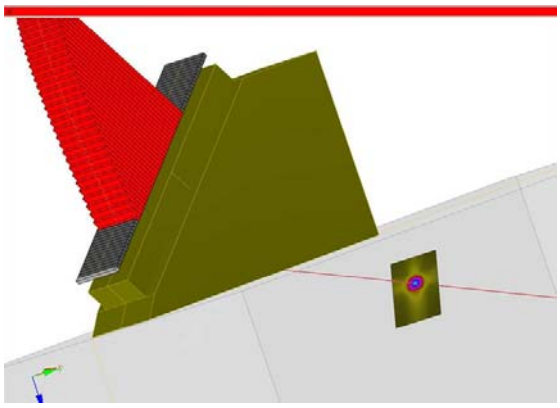


Figure 31. Cross-Sectional Beam Model

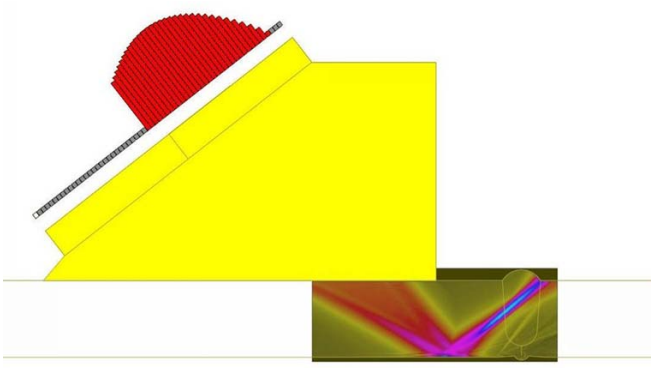


Figure 32. Beam Profile Model

After beam models were calculated for each zone, the beams were then used in flaw interaction models to evaluate how well the beam performed for flaw detection. Inspection simulations were conducted to look at the effect of flaw tilt and skew and also to determine the change in signal amplitude for planar flaws of different heights. Figure 33 provides examples of flaw tilt and skew conventions used for this work. Negative tilt occurred when the top of the flaws were tilted toward the probe while positive tilt occurred when the top of the flaws were tilted away from the probe. Due to the symmetry of skewed flaws, no direction of skew was used. All flaws used for tilt and skew simulations were 1- × 12-mm planar flaws. Typically, larger flaws will be more susceptible to tilt and skew effects than smaller flaws. Flaw tilt angles used in this analysis were deviations from the weld bevel angle at the center of the designated zone.

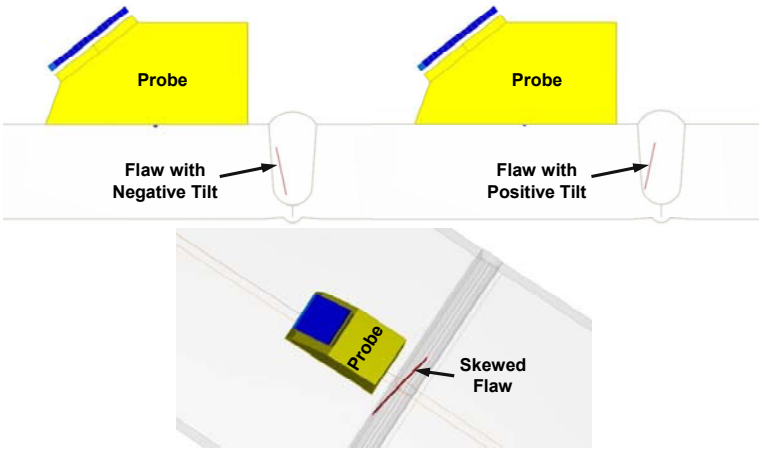


Figure 33. Flaw Tilt and Skew Convention

Initial flaw interaction models were performed using a 16-mm probe offset from weld centerline as called out by the procedure. However, in order to match the beam to the zone positions, the probe offset for modeling had to be changed to 19 mm. It is important to note that probe offsets and active aperture values contained in the AUT procedure were based on calculations and it is common for these values to change when the setup is refined using a calibration sample. The following paragraphs describe beam modeling and flaw interaction results for each zone using parameters specified in the AUT procedure. In some instances, parameters were modified to reflect parameters that would be closer to what the final AUT procedure would look like.

4.2.2 Root Zone

As shown in Figure 34, the beam cross-section model for the root showed the 6-dB through-wall beam dimension to be only about 0.4 mm less than the AUT procedure zone height of 2.35 mm for the root zone. Generally, the beam height should be approximately equal to or greater than the zone height to obtain good overlap coverage between zones. With this in mind, the root beam parameters were providing a beam very close to what would be desired.

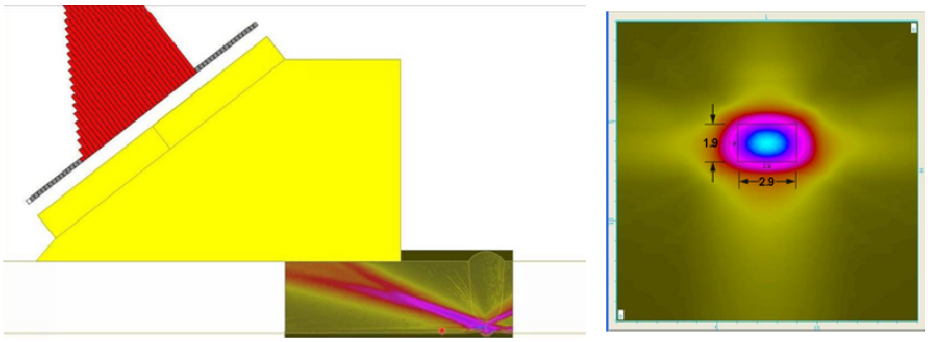


Figure 34. Beam Profile and Cross-Section for Root Zone

Flaw interaction modeling showed that flaws less than the reference notch height of 1 mm dropped off quickly in signal amplitude and would likely be greatly undersized. In fact, flaws less than approximately 0.7 mm in height would probably not be detectable as flaws since they were more than 12 dB less in signal amplitude than the reference notch. Simulations of flaws greater in height than the calibration target revealed that the tendency would be to slightly oversize when using the zonal amplitude technique (Figure 35). When the 1- ×12-mm root flaws were tilted, to simulate out of plane flaws relative to the root bevel preparation, the decrease in signal amplitude was less than 3 dB for negative tilt angles of 18 degrees or less and positive tilt angles less than 8 degrees. When root flaws were skewed, relative to the weld axis, the signal amplitude dropped off rapidly for skew angles greater than 2 degrees as shown in Figure 36.

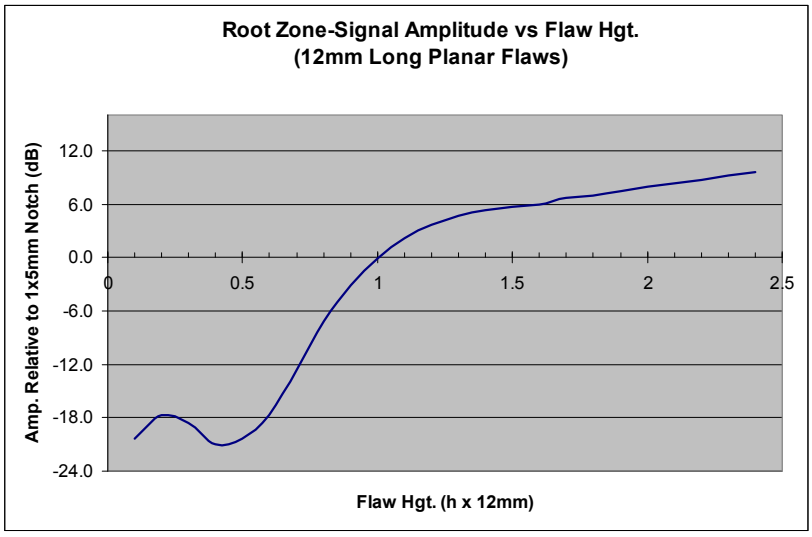


Figure 35. Root Zone Signal Amplitude vs Flaw Height

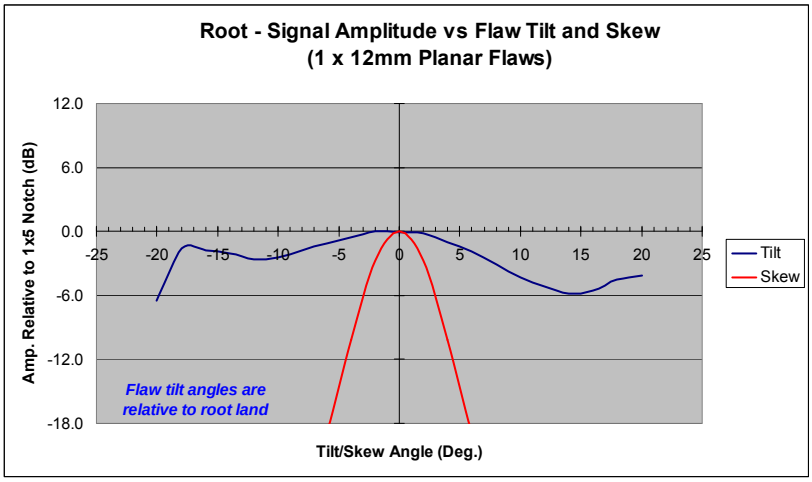


Figure 36. Modeling Prediction for Tilt and Skew of Root Flaws with a 70-Degree Beam Angle

4.2.3 Cap Zone

Beam models for the cap zone, provided in Figure 37, showed that the PA parameters produced a beam size that was about right for this zone. Flaw interaction models using 12-mm-long planar flaws of different heights showed that the signal amplitude dropped off rapidly when the through-wall height of the flaw was less than 0.5 mm. Results of signal amplitude versus flaw height for the cap zone are shown in Figure 38. It should be noted that the amplitudes shown in Figure 38 are relative to the reference reflector for the cap zone which was a 1- × 5-mm notch. Additional flaw interaction simulations, to look at tilt and skew effects, predicted less than a 6-dB drop in signal amplitude for flaws tilted up to 16 degrees (Figure 39). Skewed flaws, however, showed nearly a 9-dB drop in signal amplitude at skew angles of only 4 degrees.

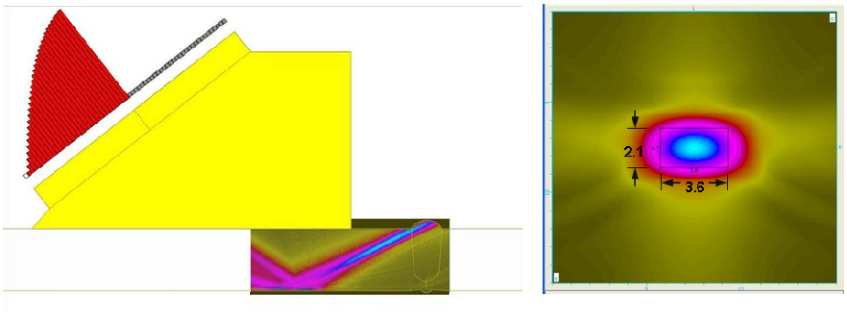


Figure 37. Beam Profile and Cross-Section for Cap Zone

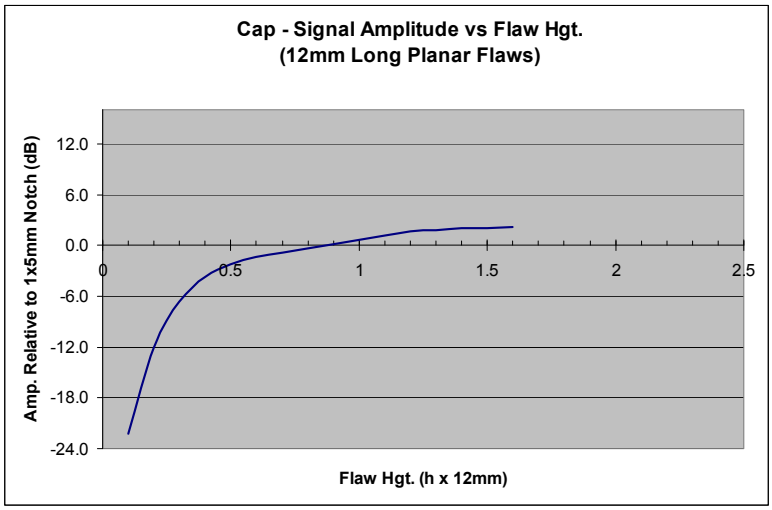


Figure 38. Cap Zone Signal Amplitude vs Flaw Height

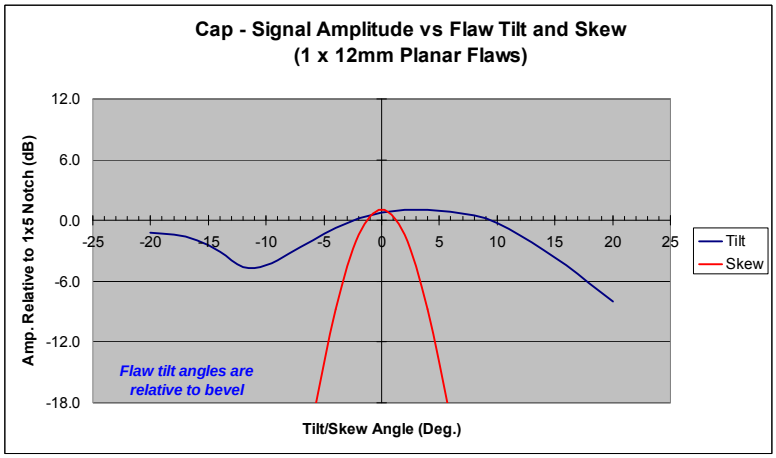


Figure 39. Modeling Prediction for Tilt and Skew of Cap Flaws with a 63-Degree Beam Angle

4.2.4 Hot Pass Zone 1 (HP1)

Beam models for HP1, in Figure 40, showed that the PA parameters produced a beam size that was slightly small for this zone. The vertical dimension of the beam was calculated to be 1.3 mm while the zone height for HP1 was 1.6 mm. Fewer elements in the aperture would produce a larger beam that, in turn, would provide more overlap (overtrace) for this zone and perhaps make the technique less susceptible to amplitude loss due to off-axis flaws. An inspection simulation of the UT technique predicted that the overtrace for HP1 would not meet the overtrace requirements of ASTM E1961 for adjacent zones as shown in Figure 41. ASTM E1961 states that the overtrace amplitude from adjacent zones should be between 6 and 14 dB less than the zone for which the beam is intended.

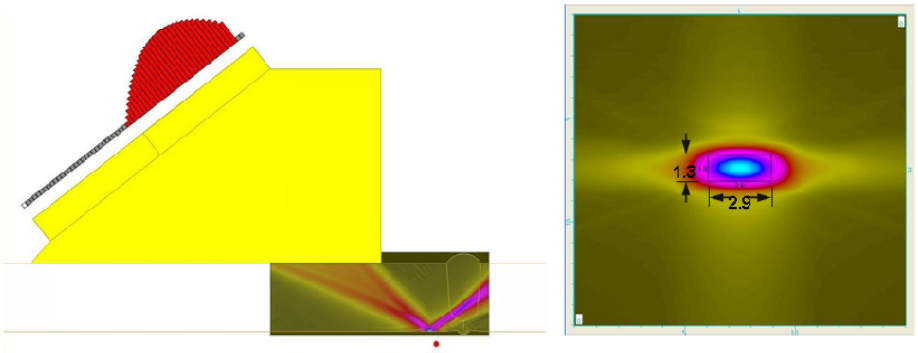


Figure 40. Beam Profile and Cross-Section for HP1

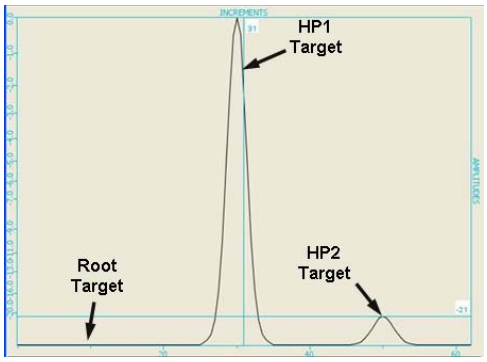


Figure 41. Inspection Simulation Showing Overtrace for HP1

During flaw interaction simulations, flaws in HP1 were oriented so that the flaws were tangent to the weld bevel radius at the mid-point of HP1 (12.4-mm deep). This resulted in flaws being tilted 47 degrees from vertical. The 2-mm-diameter flat-bottom hole calibration target hole for this zone was oriented 35 degrees from vertical as called out in the UT technique. This difference in tilt between the calibration target and the flaws resulted in low signal amplitudes for flaws in HP1. The results for different flaw heights can be seen in Figure 42. In this scenario all simulated flaws produced signals at least 4 dB less than the amplitude from the calibration target. This would result in all HP flaws being undersized with the given UT technique.

Results for flaw tilt and skew simulations for HP1 are provided in Figure 43. Negative tilting of the flaws resulted in significant decrease in signal amplitude, while positive tilts, greater than 5 degrees relative to the bevel, produced higher signal amplitudes. The non-symmetry of the tilt data in Figure 43 is due to a 12-degree difference between the angle of the calibration target and the actual bevel angle in the middle of HP1. Simulations predicted that the signal amplitude of skewed flaws drop by more than 6 dB when the skew angle was greater than 4 degrees.

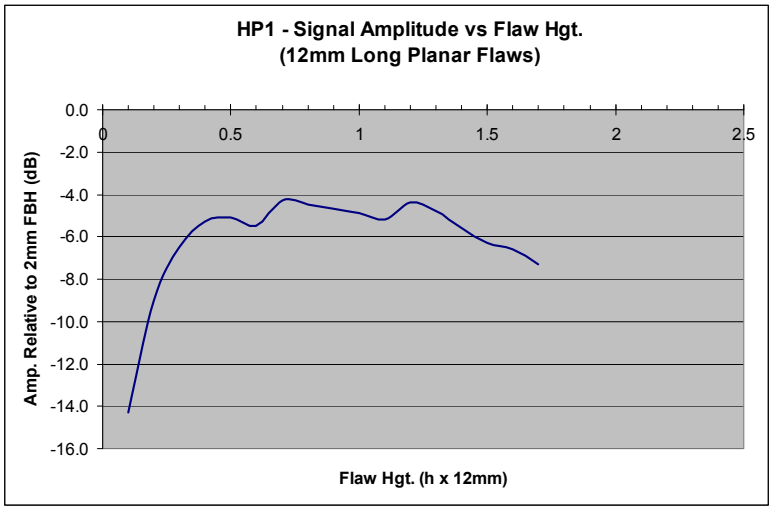


Figure 42. HP1 Signal Amplitude vs Flaw Height Relative to Reference Target

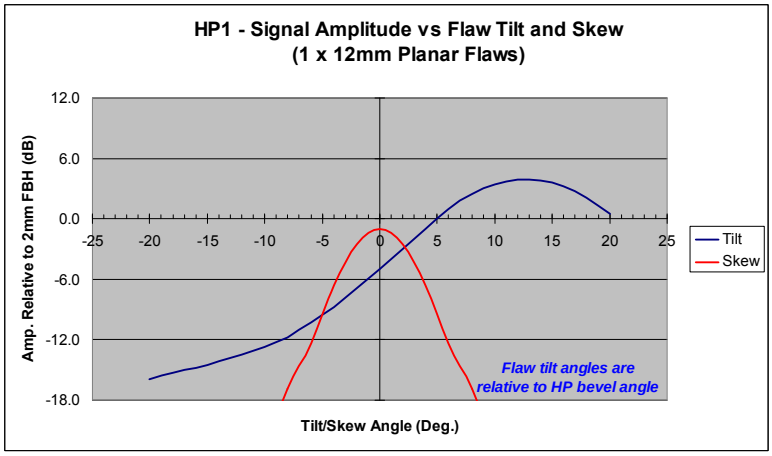


Figure 43. Modeling Prediction for Tilt and Skew of HP1 Flaws with a 55-Degree Beam Angle

4.2.5 Fill 2 Zone

Tandem pitch/catch beam models for the Fill 2 zone, in Figure 44, showed that the PA parameters produced beam sizes for the transmit and receive beams that were slightly small for this zone. The vertical dimension of the transmit beam was calculated to be 1.3 mm while the zone height for receive beam was 2.0 mm. Since the zone height was set at 2.1 mm, it was predicted that the overtrace for adjacent zones would be inadequate which was confirmed by simulation results contained in Figure 45. Additional simulations showed that aperture sizes of 17 and 21 elements for the transmit and receive apertures respectively, produced the desired overtrace (Figure 46).

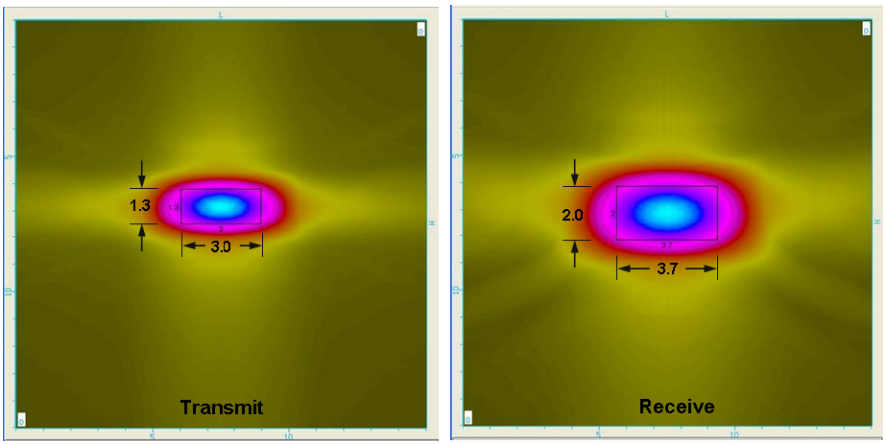


Figure 44. Beam Cross-Sections for Transmit and Receive Beams for the Fill 2 Zone

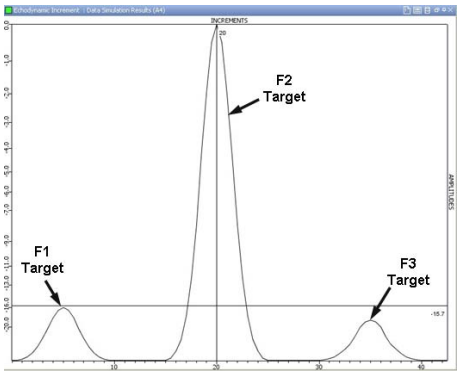


Figure 45. Inspection Simulation Showing Overtrace for Fill 2 Zone

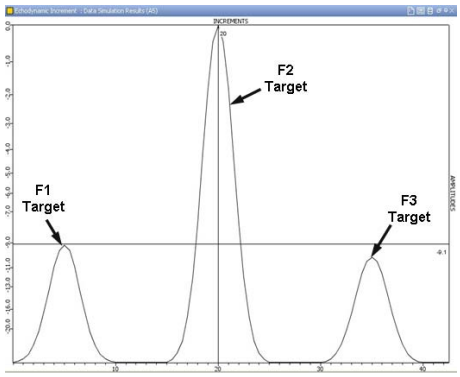


Figure 46. Inspection Simulation Showing Overtrace for Fill 2 Zone Using Reduced Aperture Sizes

Fill 2 zone flaw interaction models, using 12-mm-long planar flaws of different heights, showed that LOF flaws would generally be oversized in the height measurement. Since the zone height of Fill 2 zone was 2.11 mm, flaws exceeding the calibration target amplitude would generally be classified as full zone height. Modeling of incremental flaw height changes showed that LOF flaws greater than 0.6 × 12 mm would be sized as 2.1 mm in height or greater (Figure 47).

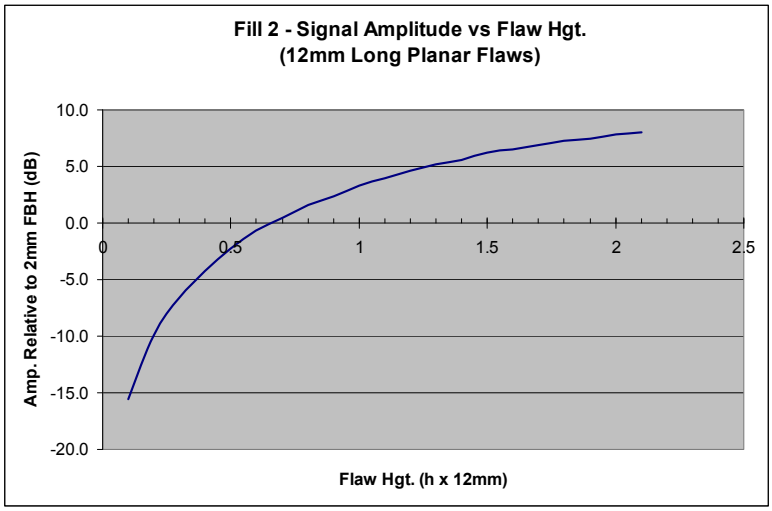


Figure 47. Fill 2 Zone Signal Amplitude vs Flaw Height Relative to Reference Target (Original Technique)

Tilt and skew modeling for Fill 2 zone using the original technique showed that positive tilt angles would result in decreased signal amplitude while negative tilt of up to 8 degrees from the weld bevel would still produce a good signal for detection. The simulations also showed that flaw skews of up to 5 degrees could still be detected. Tilt and skew results for the original Fill 2 zone technique are provided in Figure 48. These results do not appear to be representative of what would normally be expected because the transmit and receive beams did not intersect at the target zone when using the original technique. Additional simulations were performed using aperture sizes of 17 and 21 elements for the transmit and receive apertures respectively, while also correcting the beam index offset so that the intersection of the two beams was at the Fill 2 zone. Figures 49 and 50 show results from the modified Fill 2 zone technique which are considered to be a more accurate estimation of a typical fill zone tandem inspection technique. It can be seen from Figure 50 that signal amplitude resulting from flaw tilt is not symmetrical.

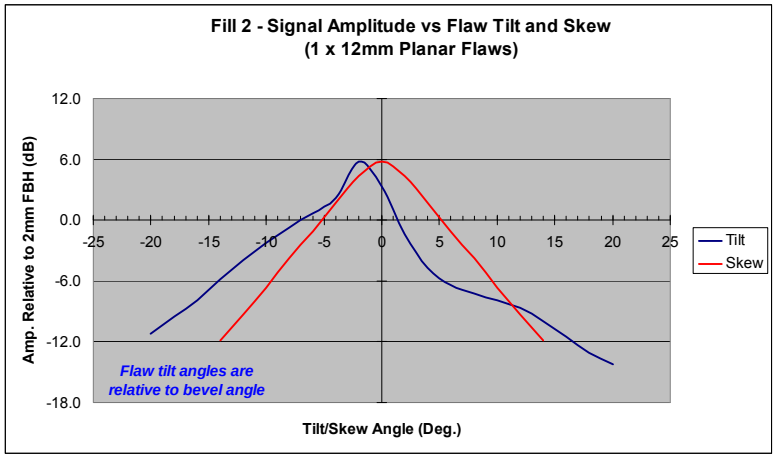


Figure 48. Modeling Prediction for Tilt and Skew of Fill 2 Zone Flaws (Original Technique)

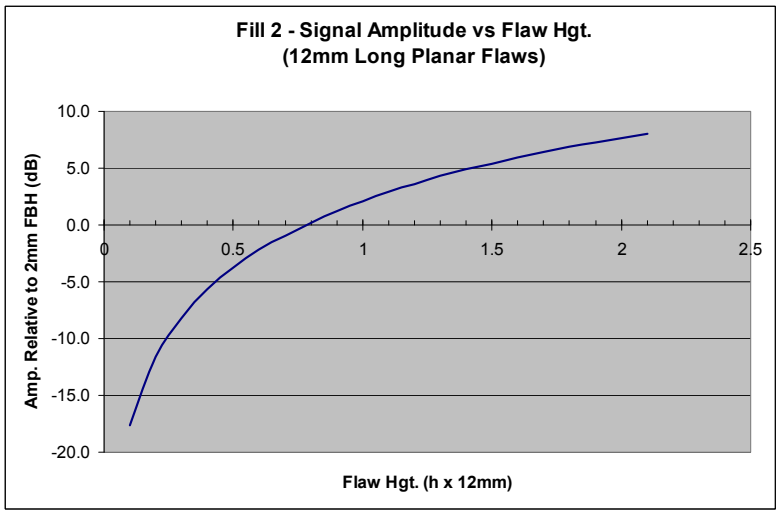


Figure 49. Fill 2 Zone Signal Amplitude vs Flaw Height Relative to Reference Target (Modified Technique)

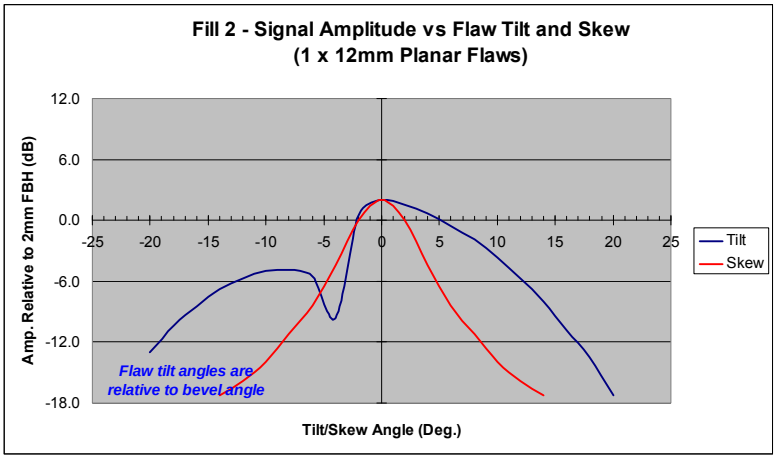


Figure 50. Modeling Prediction for Tilt and Skew of Fill 2 Zone Flaws (Modified Technique)

4.2.6 Observations from the AUT Technique Evaluation

- With the exception of the cap and root zones, too many elements were selected when defining the active probe aperture for an inspection zone. Modeling predictions show that insufficient overtrace of adjacent zones could be expected. This was especially noticeable in the hot pass and fill zones.

- Planar root zone flaws less than approximately 0.7 mm in height would be difficult to detect using the root zone technique specified.

- Planar cap zone flaws less than approximately 0.3 mm in height would be difficult to detect using the cap zone technique specified.

- Tilting of cap and root zone planar flaws did not have a significant affect on signal amplitude for tilt angles up to approximately 15 to 20 degrees. This is likely due to the corner trap effect contributing to signal amplitude.

- The tandem technique, typically used for fill zones, showed non-symmetrical amplitude behavior for flaw tilt.

- Due to a 12-degree angular difference between the calibration target and the actual bevel angle in the middle of HP1, simulations predicted that all flaws in HP1 would be undersized, if detected. Since all simulated flaw heights were at least 4 dB less than the calibration reference target amplitude, flaw detection could be an issue.

4.2.7 Evaluation of UT Parameters Using Modeling and Inspection Simulations

This section discusses beam modeling and inspection simulations that were conducted to compare changes in various AUT parameters that could affect performance. The selection of inspection parameters for an AUT inspection is critical for good flaw detection and sizing, but there is not a good understanding of the influence of these parameters on AUT results. While many other comparisons could be made, the scope and budget of the project did not permit additional work in this area.

4.2.8 Probe Frequency

Simulations were conducted for the root and HP1 to look at the affects of probe frequency, active aperture size, and scanner guide band location accuracy on flaw detection capabilities. Three probe frequencies commonly used for pipe girth weld inspection were evaluated: 4.0, 5.0, and 7.5 MHz. Other probe parameters such as element size, element pitch, and wedge design were held constant. For all evaluations the following parameters were used:

- PA element size: 1 × 10 mm
- PA element pitch: 1 mm
- Wedge angle: 38 degrees.

Modeling of the ultrasonic beams was performed to determine the beam spot size in the root zone and HP1 for different frequencies and aperture sizes. For the root zone, a beam angle of 60 degrees was selected with an electronic focal depth set at 16 mm. For the HP1 a beam angle of 55 degrees was selected with a focal depth of 12.8 mm on the second leg. Tables 10 and 11 show measurements for beam spot sizes obtained from beam cross-sections in the respective zones. From the data it can be seen that the beam height in the active probe axis becomes smaller as the frequency increases when all other parameters remain constant.

Table 10. Predicted Beam Spot Sizes at Root Using 60-Degree Beam Angle and Focal 16-mm Depth

Frequency (MHz)	Aperture Size (No. of Elements)	Spot Size (mm)	
		Height	Width
4.0	9	5.5	4.9
4.0	15	3.3	4.9
5.0	9	4.5	4.2
5.0	15	2.6	4.2
7.5	9	3.1	4.5
7.5	15	1.9	5.0

Table 11. Predicted Beam Spot Sizes at Hot Pass 1 Using 55-Degree Beam Angle and a Second Leg Focal Depth of 12.8 mm

Frequency (MHz)	Aperture Size (No. of Elements)	Spot Size (mm)	
		Height	Width
4.0	15	3.3	4.9
4.0	27	1.9	5.0
5.0	15	2.6	4.2
5.0	21	1.9	4.2
5.0	23	1.7	4.2
5.0	25	1.6	4.2
7.5	15	1.8	4.9

Initial simulations were conducted to look at the affects of frequency and aperture size on the detection of planar root flaws as the flaw height increased. To accomplish this, simulations were performed using two different aperture sizes for each probe frequency. Probe frequencies of 4.0, 5.0, and 7.5 MHz were selected with a 60-degree shear wave beam angle. Results showed that the amplitude response (echodynamic) pattern for root flaws of different heights were different for different probe frequencies. For a given frequency, however, the general amplitude response pattern was similar even when the aperture size changed. The amplitude responses are comprised of direct reflections from the flaws, as well as indirect reflections caused by the corner trap affect inherit to surface breaking flaws. These affects are shown graphically in Figure 51. The affect of aperture size became more apparent as the through-wall height of the flaws increased. As the flaw height became larger than the beam height, signal amplitude began to level off.

Models of the 4- and 5-MHz probes showed that root flaw heights of approximately 0.5 mm produced higher amplitude signals than flaw heights between 1 and 2 mm. The 7.5-MHz probe, on the other hand, displayed the lowest amplitude response for flaws 0.5 mm in through-wall height. Consequently, the choice of calibration notch depth could have an unexpected impact on test sensitivity. For example, the charts in Figure 51 show that probes having a frequency in the 4- to 5-MHz range produce a larger amplitude response for 0.5 mm-deep notches than for 1-mm-deep notches. This is especially true at a frequency of 4 MHz. This suggests that the use of a root reference notch of 0.5 mm in through-wall height, with a frequency of 4 to 5 MHz, could result in planar flaws approximately 1 mm in height being missed or undersized. In contrast, the 7.5 MHz results show that a root reference notch of 1 mm in through-wall height would be a good choice for calibration purposes.

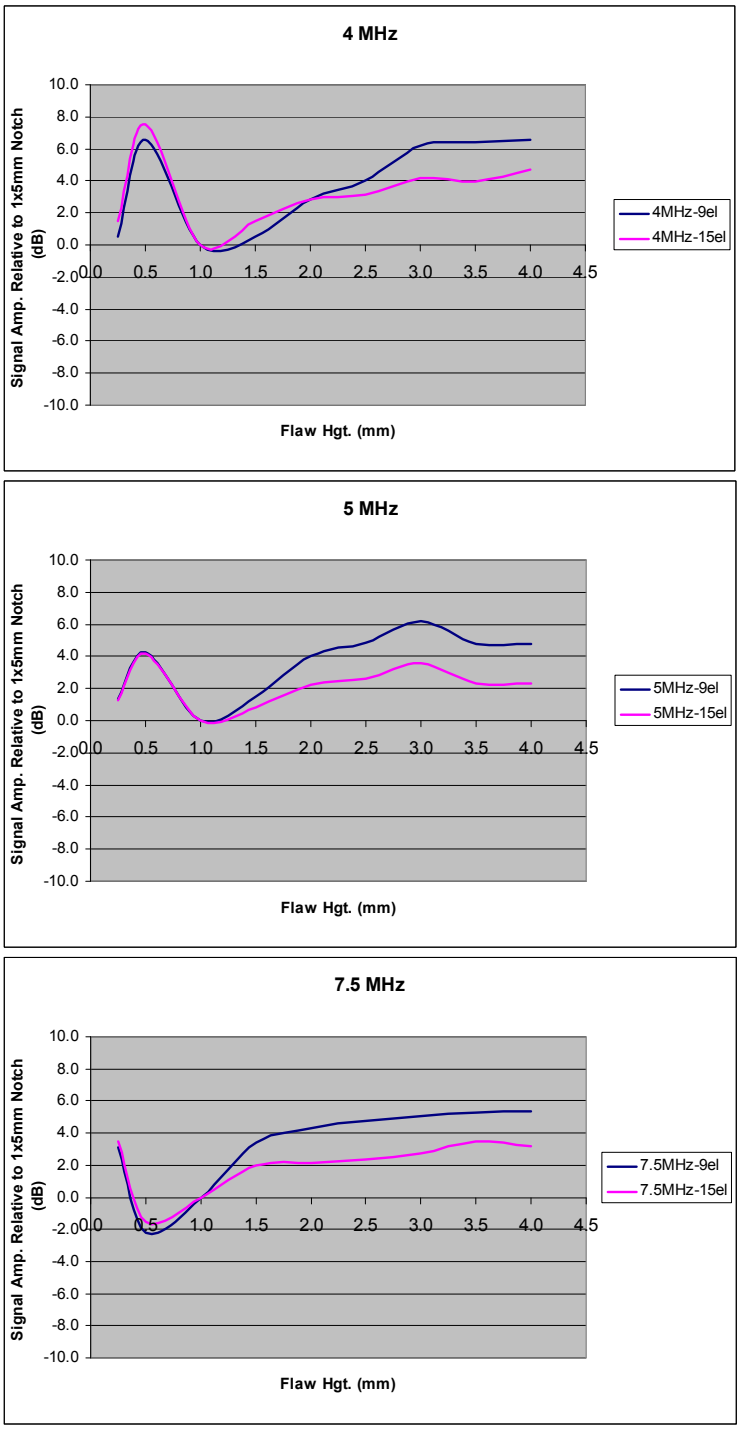


Figure 51. Affect of Probe Frequency and Aperture Size on Root Flaw Responses Using 4.0-, 5.0- and 7.5-MHz Probes (Top to Bottom)

In addition to root zone simulations, simulations were also performed to look at frequency affects in HP1. As with the root zone, probe frequencies of 4.0, 5.0, and 7.5 MHz were modeled. For HP1, however, a 55-degree shear wave beam was used to look at planar flaws oriented perpendicular to the beam. The results from HP1 simulations are plotted in Figure 52. Results showed that the amplitude response (echodynamic) pattern for HP1 flaws with different heights behaved similarly for the probe frequencies evaluated. The only significant differences observed were in the weaker signal amplitudes for smaller aperture sizes. However, when the aperture size was adjusted to produce equivalent beam heights at each frequency, the amplitude curves were nearly identical. This can be seen by comparing the larger aperture sizes for the 4- and 5-MHz probes to the 15-element aperture of the 7.5-MHz probe. From Table 11 it can be seen that these aperture sizes all result in nearly equivalent through-wall beam height of approximately 1.9 mm. Based on these results, similar AUT results could be expected for the hot pass zone with any of the three frequencies provided the beam spot sizes were similar.

Other simulations were conducted to look at the affects of flaw orientation on signal amplitude for different probe frequencies. The simulations were performed using 1.2-mm high × 12-mm long planar flaws centered in HP1. First, the flaws were tilted with no skew, and then the simulation was repeated with skew and no tilt. For these simulations, zero tilt and skew is when the flaw was perpendicular to the ultrasonic beam. Results of these simulations are provided in Figure 53 for the 5- and 7.5-MHz probe. These results show that the signal amplitude dropped quicker with the higher frequency probe even when the beam sizes were nearly identical. While orientation plays a big role in signal amplitude, the simulation results indicate that lower frequency probes would be less susceptible to flaw tilt and skew than higher frequency probes.

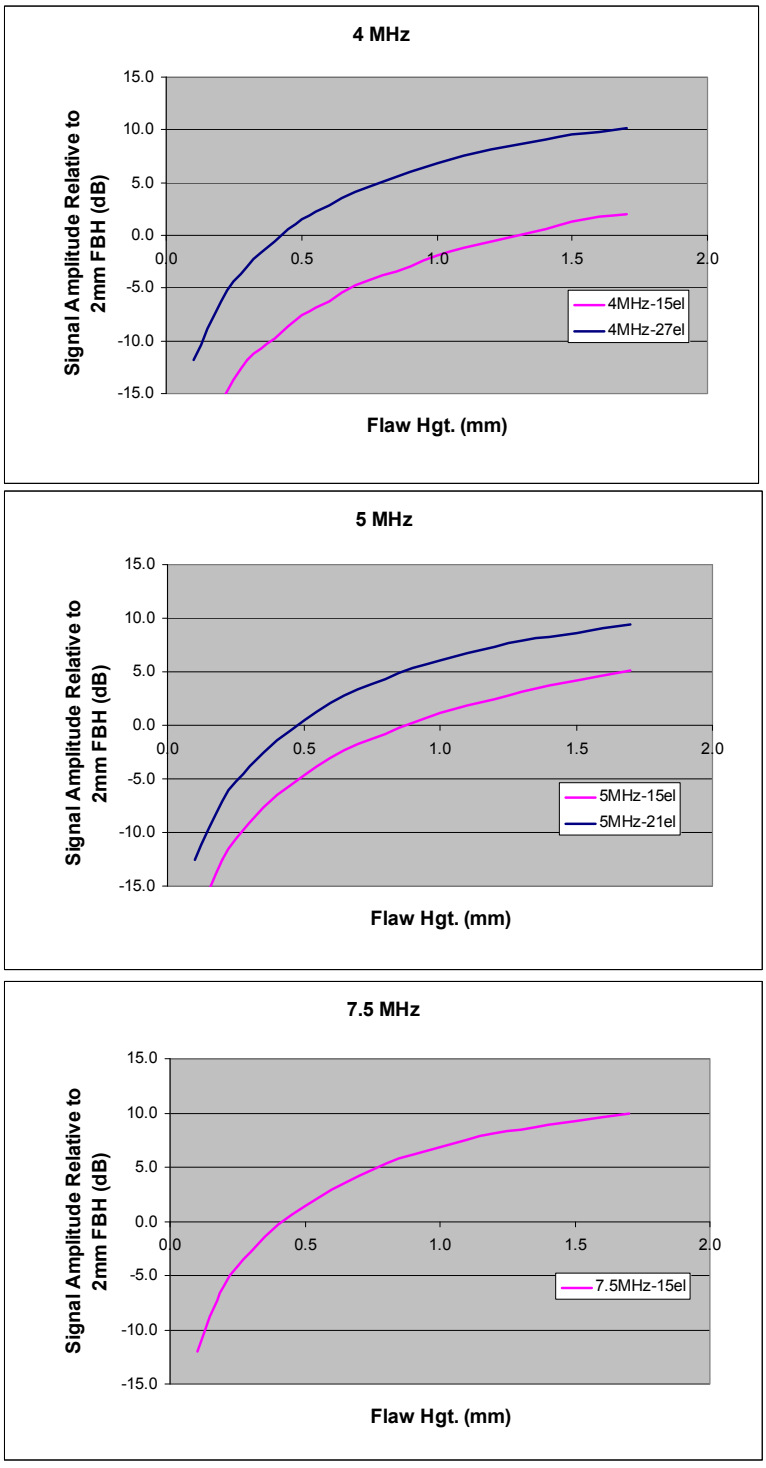


Figure 52. Affect of Probe Frequency and Aperture Size on HP1 Flaw Responses Using 4.0-, 5.0- and 7.5-MHz Probes (Top to Bottom) with Different Aperture Sizes

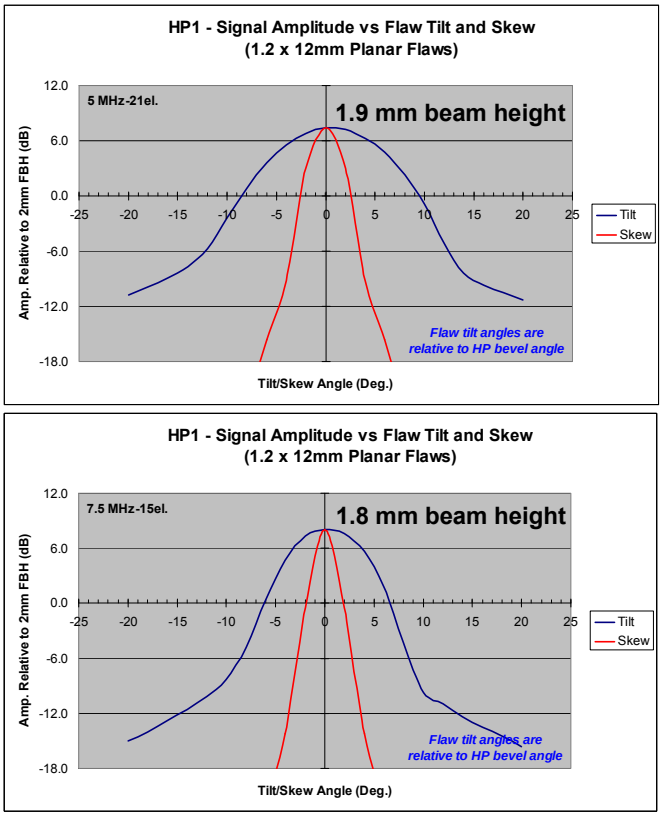


Figure 53. Affects of Flaw Tilt and Skew on Signal Amplitude for 5- and 7.5-MHz Probes with Similar Beam Sizes

4.2.9 Guide Band Offset

Simulations were performed to look at the affects of the guide band offset on signal amplitude. The software simulations were performed by offsetting the probe 2 mm in each direction from the optimum probe offset relative to weld centerline. For simulation purposes a negative band offset meant that the probe was offset away from weld centerline while a positive band offset meant the probe was offset toward weld centerline. Results provided in Figures 54 through 56 indicate that error in band offset had a greater affect when the beam spot sizes were small (refer to Tables 10 and 11). This was particularly evident for the small beam size shown in Figure 55. For all frequencies, any band offset resulted in a decrease in signal amplitude; however, for both the HP1 and root zones band offset away from the weld (negative offset) resulted in the greatest decrease in signal amplitude.

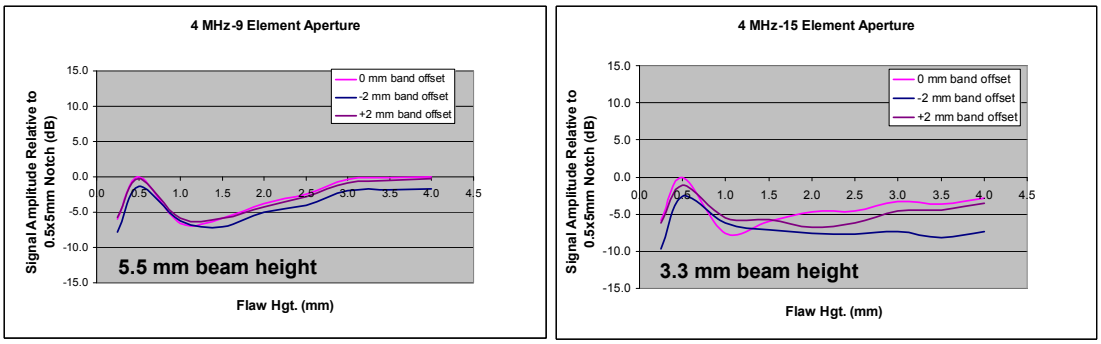


Figure 54. Band Offset Affect on Root Zone Signal Amplitude for Different Beam Sizes with 4-MHz Probe

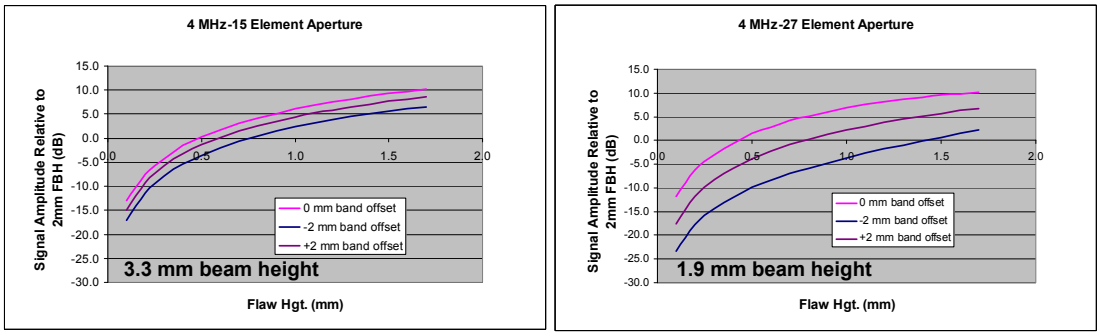


Figure 55. Band Offset Affect on HP1 Signal Amplitude for Different Beam Sizes with 4-MHz Probe

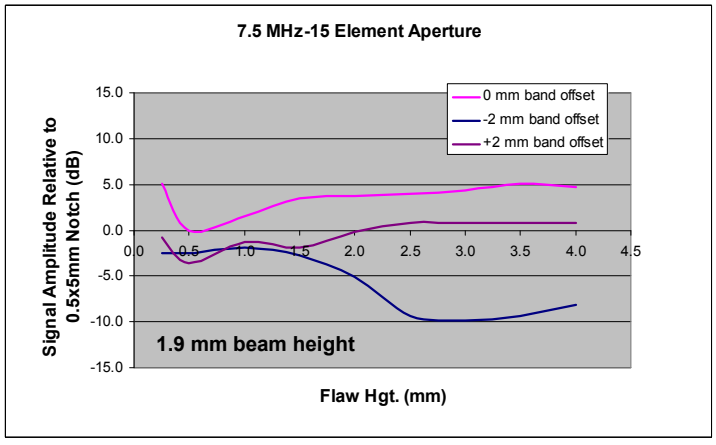


Figure 56. Band Offset Affect on Root Zone Signal Amplitude for Small Beam Size with 7.5-MHz Probe

4.2.10 Task Conclusions and Significance

Based on the modeling and inspection simulation results, the following conclusions can be made:

- Probe frequency influenced detection of root planar root flaws with a through-wall height less than 1 mm. Since this is the through-wall height range of typical root reference notches, it is important to select a reference notch compatible with the probe frequency.

- Using amplitude sizing in the root zone would result in over sizing flaws less than 1 mm in through-wall height and slight under sizing of flaws greater than 1 mm.

- For both the Root and HP1 zones, beam size in the through-wall dimension had a significant influence on signal amplitude when there was an error in the guide band placement. This error in band placement caused a decrease in signal amplitude for both positive (toward the weld) and negative (away from the weld) band offsets. For every modeling scenario of frequency and beam spot size combinations, flaw signal amplitudes dropped the most when the band offset error was away from the weld. The smaller beam sizes resulted in greater drops in signal amplitude as the band was offset. The drop in signal amplitude with band offset error appeared to be primarily a function of beam spot size rather than frequency.

- Due to the constant changing weld bevel radius in the HP zone, flaws along the weld bevel can be at different angles throughout the zone. This can lead to reduction in amplitude caused by the difference between beam angle and flaw angle. Modeling simulations showed that lower frequency probes had slightly less amplitude drop for tilted and skewed flaws in the HP zone when compared to higher frequency probes with a similar focal spot size. Flaws used for this comparison were 1.2 mm in height by 12 mm in length.

4.3 Statistical Analysis

4.3.1 Introduction

The need for NDE to become quantitative rather than qualitative arose in the period when the NDE became part of the structural integrity programs where risks or probability of failure were expected to be extremely low. Examples of such industries and structural integrity programs are aerospace military and commercial vehicles and nuclear power plants. One of the first programs for establishing the quantitative performance of NDE techniques for fatigue crack detection was conducted by W. Rummel in 1974.[26] A recommended practice[27] was published later specifying how to conduct the performance demonstration programs using binomial law to evaluate POD of flaws by various NDE methods. Since then, many studies were funded by the

U.S. Air Force and other organizations to research and improve the process of POD determination.[28] Later, the POD methodology was further developed[29] and a MIL-HDBK-1823 with current 2009 revision[30] provided comprehensive guidelines on POD estimates from NDE validation and quantification trials. Many examples of POD calculations are provided in the various revisions of the MIL-HDBK-1823 and other publications.[31]. Short summaries of various POD and NDT reliability studies are available in[32] up to year 2000.

A somewhat different approach is found in a methodology issued by the European Network for Inspection Qualification (ENIQ).[6] The POD is defined differently than the POD in the documents discussed earlier.[28-31] Further, a TJ can also be used for NDE qualification along with practical trials to determine the POD experimentally. The qualification of an NDE procedure with equipment and the qualification of an entire system consisting of procedure, equipment, and personnel can now be separate activities. Open trials (examiners know the flaw size and location) are usually recommended for the procedure and equipment qualification. On the other hand, blind trials (examiners do not know the flaw size and location) are used for NDE system (procedure, equipment and personnel) qualification. A similar approach is found in ASME, Section V.[33]

As far as the AUT of girth welds is concerned, Det Norske Veritas (DNV) code DNV-OS-F101[34] provides guidelines to conduct the AUT qualification for offshore applications and recommends the use of a Nordtest report[35] to evaluate the AUT performance in terms of POD. This methodology[35] recommends either point POD estimate using binomial law[27,33] or POD curves versus flaw size.[28-30] Additional data regarding the POD of different techniques for pressure vessels and welds can be found in the literature.[36,37] One should be careful interpreting the POD data from different sources because of different POD definitions and methodologies used to generate the POD estimates (e.g., Reference 36 vs Reference 37).

Advanced and quantitative NDE, such as AUT of girth welds, allow the use of acceptance criteria based on experimental DT or fracture mechanics, or other fitness-for-purpose/service assessments. The DT or the ECA approaches are used to determine an allowable flaw size depending on material/weld properties and loading conditions during pipeline installation and operation. Typically, a DT or an ECA gives allowable flaw sizes that are smaller than the critical flaw sizes, as safety factors are used. The allowable flaw size can be expressed as an allowable flaw height which, in turn, is a function of flaw length.[1] The flaw depth measurement is also required because different accept/reject criteria will apply for flaws with different depth through the pipe WT. An error is always present in the measurement of flaw height, length, depth, and flaw location along the pipe circumference using AUT and must be taken into account when calculating acceptable flaw sizes from allowable ones, as an assurance against accepting non-allowable flaws with a sufficiently high confidence.

The vast majority of NDE reliability studies dealt with the POD only[38] because “no flaw allowed” is the standard practice for aerospace engine and air frame applications. For girth weld applications, the AUT system flaw sizing accuracy or uncertainty must be quantified in addition to the detection capabilities expressed in POD.

False calls, or spurious indications, are critical NDE reliability quantities and can have a number of causes. When performing AUT of pipeline girth welds, a number of false calls may be observed and should be investigated and quantified for a specific AUT system or range of AUT systems before the actual field inspection.

Very limited data[39,40] is available in the open literature for assessment of AUT performance expressed in POD and accuracy of flaw sizing using amplitude-based multi-probe and PA technology. Sizing plots of measurement error and measured UT flaw height versus actual flaw height are found in Reference 41.

One of the deliverables of this project related to the statistical analysis was to determine the POD and flaw sizing accuracy of representative AUT systems for inspection of on shore pipelines.

4.3.2 Scope

Six weld specimens (W1 to W6) with implanted flaws were fabricated, to provide statistically significant number of flaws and sufficient areas without flaws. Open and blind practical trials were conducted with various AUT systems. The weld circumference was divided in units or sectors with a length of 40 mm to allow for better control of areas with and without flaws and simplify the analysis. Weld 1 (W1) specimen was used for training, open trials and additional experimentation with other techniques (Tomocar). W1 was not sectioned after completion of the practical trials. Specimens W2 through W6 were sectioned to establish the POD, flaw sizing uncertainty, and false positive rate. All six weld specimens were built from a typical carbon steel pipe with the same diameter and WT. The same weld bevel preparation and welding process was used to fabricate the samples.

For zonal AUT techniques, the equipment gain and signal threshold level for accept or reject decision will affect the POD and probability of false positive (PFP). Except for fingerprinting, system V1, V3 and V4 open trials, standard procedures for equipment setup and calibration used were representative of typical field procedures (reference AUT setups provided in Section 4).

Five AUT systems were used to determine the POD and flaw sizing accuracy of a typical AUT system. The vendors were asked to provide their best operators for the trials so that the

operator's effect on system performance would be minimized. Several of the systems used equipment from the same equipment vendor. One of the systems (V3) with lower level of automation was not a typical AUT system for pipeline girth weld inspection. The procedure for measuring and reporting the flaw depth used by the operator (V3) was not typical either.

4.3.3 Terms and Definitions

4.3.3.1 Sizing Definitions

The AUT system might be considered a measurement device because it is used to measure or size flaw dimensions such as height, length, depth, and position (start and stop) along the circumference. In this case, standard guidelines already developed and used for many years are followed to describe and express the uncertainty of the measurements.[42-44]

The measurand is the particular quantity subject to measurement. For AUT of girth welds, the measurand is the flaw height, length, depth, start, and stop position.

Any single AUT measurement provides an estimate $\hat{y}_i$ of the measurand consisting of its "true" value a_i and a measurement error ε_i in accordance with Eq. (1):

$$\hat{y}_i = a_i + \varepsilon_i \tag{1}$$

The error is usually assumed to be normally distributed with a mean and standard deviation.

For analysis purposes, the error ε_i consists of a systematic ($Sys\ \varepsilon_i$) and a random ($Ran\ \varepsilon_i$) component[35] shown in Eq. (2):

$$\varepsilon_i = Sys\ \varepsilon_i + Ran\ \varepsilon_i \tag{2}$$

The true value a_i of the measurand and the error ε_i are never known and can only be estimated. The true value estimate is provided by a metallographic test (DT) or other more accurate reference method (e.g., fingerprinting). An estimate of a single measurement error is the difference between the AUT estimate and the reference (true) measurand value Eq. (3):

$$\varepsilon_i = \hat{y}_i - a_i \tag{3}$$

An estimate of the systematic error component [$Est\ (Sys\ \varepsilon_i)$] in Eq. (2) is provided by averaging the individual errors of a large number of n measurements shown in Eq. (4) below:

$$\bar{\varepsilon} = Est(Sys\ \varepsilon_i) = \frac{\sum_{i=1}^{n} \varepsilon_i}{n} \tag{4}$$

An estimate of the random error spread or dispersion is provided by the standard deviation *s(ε)* error and variance *V(ε)* shown in Eq. (5). The standard deviation is also referred to as Standard Uncertainty.[42-44]

$$s(\varepsilon) = \sqrt{V(\varepsilon)} = \sqrt{\frac{\sum_{i=1}^{n} (\varepsilon_i - \bar{\varepsilon})^2}{n-1}} \tag{5}$$

One parameter that is used to describe the AUT performance is the 95% safety limit against undersizing (*95% LUS*)[41,45] shown in Eq. (6) where *k* is the so called coverage factor.[42-44] If the error is normally distributed, the coverage factor becomes the standardized normal deviate[46] with value of 1.645 for a large number (*n* > 120) of measurements. For normally distributed data, a parameter *t* from the Student's distribution with 95% (5% one tail) probability should be used as coverage factor where n <120. The value of the parameter *t* is determined from statistical tables or dedicated statistical software (e.g., Minitab®, MS Excel®).

$$95\%LUS = ks(\varepsilon) - \bar{\varepsilon} \tag{6}$$

The term *ks(ε)* is referred to as Expanded Uncertainty[42] defining an interval expected to encompass a fraction of the distribution of values attributed to the measurand at a specified confidence level.

Shall we incorporate prediction interval instead of confidence? We are interested to know where the next measurement will be based on what we see rather than obtaining uncertainty of the systematic/average error? (ET)

The (*95% LUS*) is also referred to as 5% error fractile or undersizing error tolerance that gives equal or less than 5% probability of undersizing.[34] The estimation of *95% LUS* requires knowledge of the distribution law (probability density function) so that a value of the coverage factor can be obtained for the desired confidence. It is usually assumed that the distribution is normal which may not be the case as it will be shown later.

The *95% LUS* estimation is needed when the undersizing is more critical than the oversizing as is the case for AUT flaw height and length measurements. However, an error estimation in either direction in the AUT depth measurements is equally important and the *95% LUS* may not be of such interest and importance. The standard form of expressing the results from each

depth measurements is then recommended[42] to be slightly modified to account for the presence of the systematic depth measurement error [Eq. (4)] estimated by comparison to the reference method. Accounting for the systematic error and expanded uncertainty interval, the actual depth is expressed as shown in Eq. (7):

$$d_i = \hat{d}_i - \bar{\varepsilon} \pm ks(\varepsilon) \tag{7}$$

The true depth value d_i is then expected to be in an interval as shown in Eq. (8):

$$\hat{d}_i - \bar{\varepsilon} - ks(\varepsilon) \le d_i \le \hat{d}_i - \bar{\varepsilon} + ks(\varepsilon) \tag{8}$$

As previously discussed, knowledge of the distribution law (probability density function) is required so that the value of the coverage factor can be obtained for the desired confidence. If the error is normally distributed, the coverage factor becomes the standardized normal deviate[46] with a value of 1.645 for a large number (n >120) of measurements and 90% confidence. The high confidence of 90% is recommended. It will have the same coverage factor as the 95% one-sided confidence used to estimate the *95% LUS* for normally or other symmetrically distributed data.

It is important to note that the error of a single height measurement ε_i was estimated by comparing the maximum of the true height a_i to the maximum of the AUT estimate $\hat{y}_i$ of the height. It is understood that if the average (several DT sections on each side of maximum) of the true height a_i was used instead of the maximum, smaller height error estimates might be associated with the AUT systems. However, the averaging approach might be less conservative for the purpose of the ECA where a conservative estimate of the AUT system performance is needed with regard to the maximum rather than the average flaw height. The depth error estimates were obtained by comparing the true to the AUT depth for the flaw location where the maximum flaw height was measured.

4.3.3.2 POD Definitions

Three parameters are usually obtained as a result of dedicated POD studies: a_{50}, a_{90}, and $a_{90/95}$. The interpretation or definition of each of these three parameters is as follows:

- a_{50} – flaw size with 50% POD. This means that 50% of the flaws with this size and larger will be detected.

- a_{90} – flaw size with 90% POD. This means that 90% of the flaws with this size and larger will be detected.

- $a_{90/95}$ – flaw size with 90% POD and 95% confidence. This is the most quoted parameter in the literature. It means that 90% of the flaws with this size and larger will be detected and this is true in 95% of the inspections under similar conditions (equipment, operators, environment, etc.).

The POD expressed as $a_{90/95}$ can be estimated for a single flaw size applying the so called "29-out-of-29" rule. It means that 29 out of 29 flaws at a given size must be detected to demonstrate $a_{90/95}$ at this size.[27,35] It is actually difficult to fabricate flaws with identical size and the actual flaws will be expected to cover an interval of sizes. The $a_{90/95}$ will then be the largest flaw size in the range. This approach might be applicable for NDE systems quantification where the $a_{90/95}$ is known to be well in the range of the NDE system capabilities.

As discussed earlier, there is little POD data in the open literature for AUT systems. In this case, the $a_{90/95}$ and the other POD parameters can be estimated by building a POD curve for a range of flaw sizes where the $a_{90/95}$ is expected to be. This approach is easier to implement and requires less number of flaws. There are two different techniques for building a POD curve – "*â vs a*" and "*hit/miss*".[28-30,35] When the *â vs a* technique is implemented, *a* is the flaw size (e.g., height, length) and *â* is the instrument response (e.g., milivolts, screen divisions, percent of screen height, and others) for the flaw with size *a*. The *hit/miss* technique requires that only two conditions of the instrument response are considered: *hit* (pass) – a flaw with size *a* was detected (instrument response coded as "1") and *miss* (fail) – a flaw with size *a* was missed (instrument response coded as 0). Another feature of hit/miss analysis is that the probability of hit or miss is analyzed (continuous function of flaw size) rather than the instrument response.

For AUT systems, the POD curve is usually built as a function of the flaw height or length. The *â vs a* was not possible to obtain for all systems used during this project. Where available, the signal amplitude exhibited large scatter and did not appear to correlate with flaw size. Consequently, the *hit/miss* technique and the *POD(a)* curves as a function of the flaw height *a* are reported in this study.

There are different link functions and probability transformations used for building a POD curve as shown below Eqs. (9) through (16):[30]

$$POD(a) = \frac{e^{Y}}{1+e^{Y}} \text{ logistic, log-odds or logit link} \tag{9}$$

$$Y = \log\left(\frac{p}{1-p}\right) \text{ logistic probability transformation} \tag{10}$$

$POD(a) = \Phi(Y)$ probit link (normal or log-normal if the flaw size is log transformed[28]) where the Φ( ) is the standard normal cumulative distribution function.[42] (11)

$$Y = \Phi^{-1}(p) \text{ probit probability transformation} \quad (12)$$

$$POD(a) = 1 - e^{-e^{Y}} \text{ complimentary log-log or cloglog link} \quad (13)$$

$$Y = \log[-\log(1-p)] \text{ cloglog probability transformation} \quad (14)$$

$$POD(a) = e^{-e^{Y}} \text{ loglog link} \quad (15)$$

$$Y = \log[-\log(p)] \text{ loglog probability transformation} \quad (16)$$

For the transformation and link functions above Eqs. (9) through (16), the probability of detecting the i-th flaw p_i is a function $POD(a_i)$ of the flaw size (e.g., height) a_i Eq. (17):[30]

$$p_i = POD(a_i) \quad (17)$$

The flaw size is either transformed (e.g., log) Eq. (18) or not Eq. (19):

$$X = \log(a) \quad (18)$$

$$X = a \quad (19)$$

A different ("special" in this report) function is recommended in the NORDTEST[35] report Eq. (20):

$$POD(a) = 1 - \frac{1}{1 + \left(\frac{x}{x_0}\right)^{b}} \quad (20)$$

where parameters x_0 and b are estimated from the data sample.

Another parameter used for AUT qualification and quantification that links sizing accuracy or uncertainty with detection capabilities is the probability of rejection (POR) of a flaw having a certain size.[34] This parameter is not addressed in this report, however, the raw data files

available on the project CD could be used to obtain the POR estimates for the flaw size range reported in this study.

4.3.3.3 Measurement-Sample-Group Designation

The format used throughout the report is as follows:

- Y(Y) v X - Ddd - (TT)

where Y(Y) – group of experimentally acquired measurements (response variable on sizing plots), v – versus, X – group of measurements or reference to which the experimental measurements Y(Y) are compared (predictor variable on sizing plots), Ddd – flaw dimension or position, (TT) – measurement technique as applicable.

The following abbreviations have been used for group description:

- F – Fingerprinting
- FA - Fabrication specifications
- D – Destructive
- Dpt – Depth
- Hgt – Height
- Len – Length
- PA – Phased array
- Sta - Circumferential start position
- Stp – Circumferential stop position
- TD – TOFD
- V1 through V5 - AUT System 1 through AUT System 5
- XRM - X-ray or magnetic particle.

A more detailed description of the data groups is available in the statistical presentation in Appendix E.

Flaw categories were also formed based on the DT. Following are abbreviations used for each category:

- 0Sk0TI - Implanted, planar, no skew and/or tilt.
- 0SkTI - Implanted, planar, non-intentional tilt. These flaws were initially specified as 0Sk0TI. The DT indicated that they had larger tilt angle than the required (zero). The flaws were pooled in one sample regardless of the actual tilt angle.
- IB - Implanted, IB-LOF.

- SkTI - Implanted, planar, intentional skew and tilt. These flaws had different skew and tilt angles. The flaws were pooled in one sample regardless of the actual skew or tilt angle.
- Natural - Natural, planar, non-intentional planar. All natural planar flaws were pooled in one sample regardless of the actual skew or tilt angle.
- Vol - Pores and other volumetric flaws
- 90Sk0TI - Implanted, planar, transverse.

4.3.3.4 Procedure for Data Analysis

The sizing error components of Eqs. (2) through (5) were estimated to compare the weld specimens, systems, processes, and to assess AUT system sizing performance and capabilities.

Each data sample was processed in accordance with the following procedure:[47]

- Obtain average
- Obtain standard deviation (variance) or uncertainty
- Plot histogram and compare to normal distribution
- Perform normality test (Anderson-Darling)
- Build and analyze box plots
- Obtain and analyze other statistics - Kurtosis, Skewness, Range, Confidence Intervals (CI), etc.
- Perform Equal Variance test
- Perform analysis-of-variance (ANOVA) or nonparametric tests as applicable to check whether a statistically significant difference exists between distributions
- Identify outliers
- Perform parametric Test 1 – “2-Sample t” for two or “F-Tukey's” for more than two distributions that are normally distributed and have equal variance
- Perform nonparametric Test 2 – “Mann-Whitney” for two or “Kruskal-Wallis” for more than two distributions that are not normally distributed and or do not have equal variance

The statistical significance for all tests (P-Value) was 0.05 (or 5%).

The sizing data was processed using Minitab® statistical software. Other dedicated software packages (e.g., Microsoft® Excel) may be used as applicable.

Examples of data analysis with Minitab are shown in Figures 57 through 59 for height errors obtained by comparison of fabrication specification with (versus) DT measurements (FAvD-Hgt). Major parameters as mean (average), standard deviation (uncertainty), variance, skewness,

kurtosis and others are shown in the summary plot (Figure 57) where data from all five welds (W2 through W5) was pooled. The distribution is considered normal because an error (p-Value) of 24.4% > 5% (significance level) would be made if the hypothesis of normality was rejected. Three outliers are also shown on the box plot. The distributions for each separate weld were tested in a similar way followed by tests (Test 1 in this case) for variance equality (Appendix E).

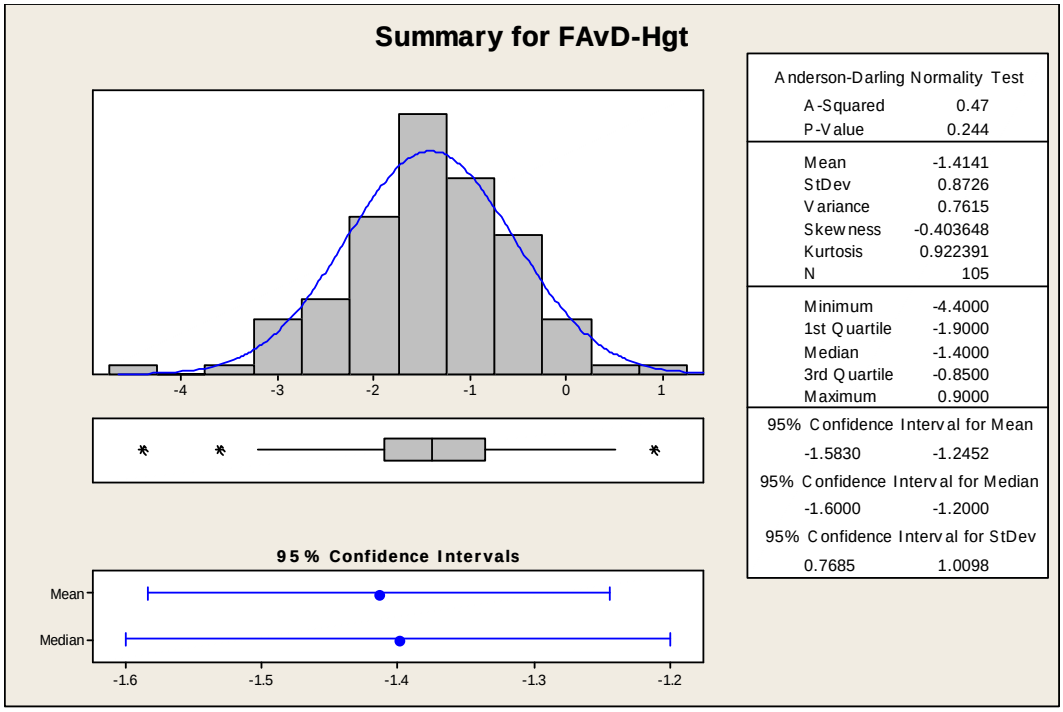


Figure 57. St1 Pooled Data Summary for All Five Welds, FAvD-Hgt

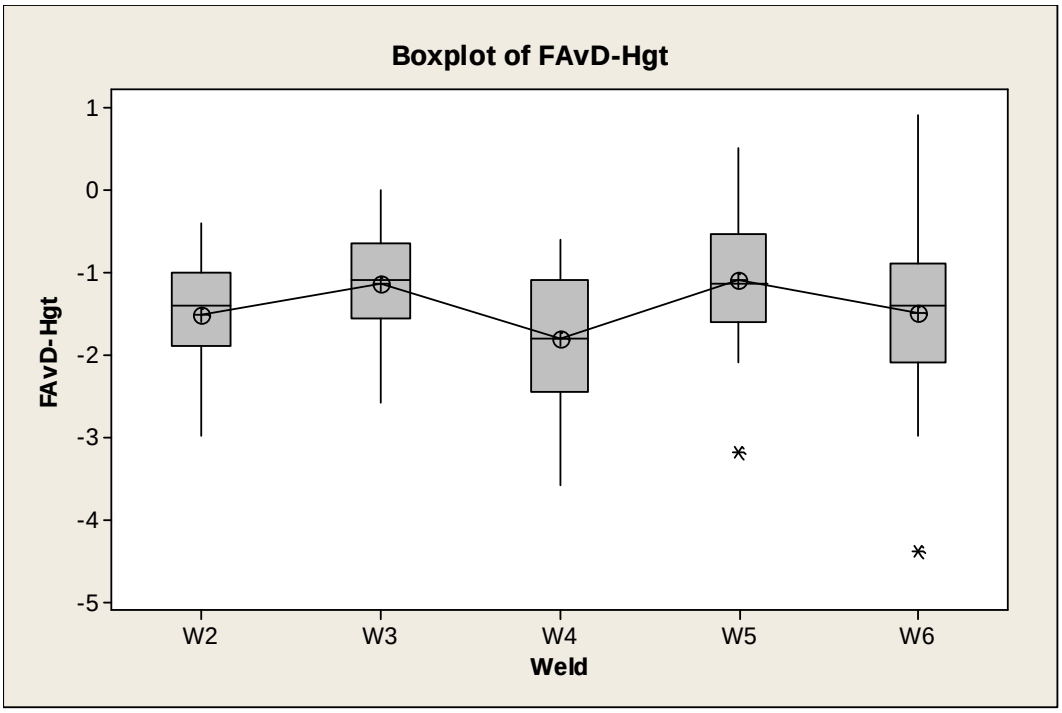


Figure 58. St1 One-Way ANOVA: FAvD-Hgt vs Welds - Boxplot

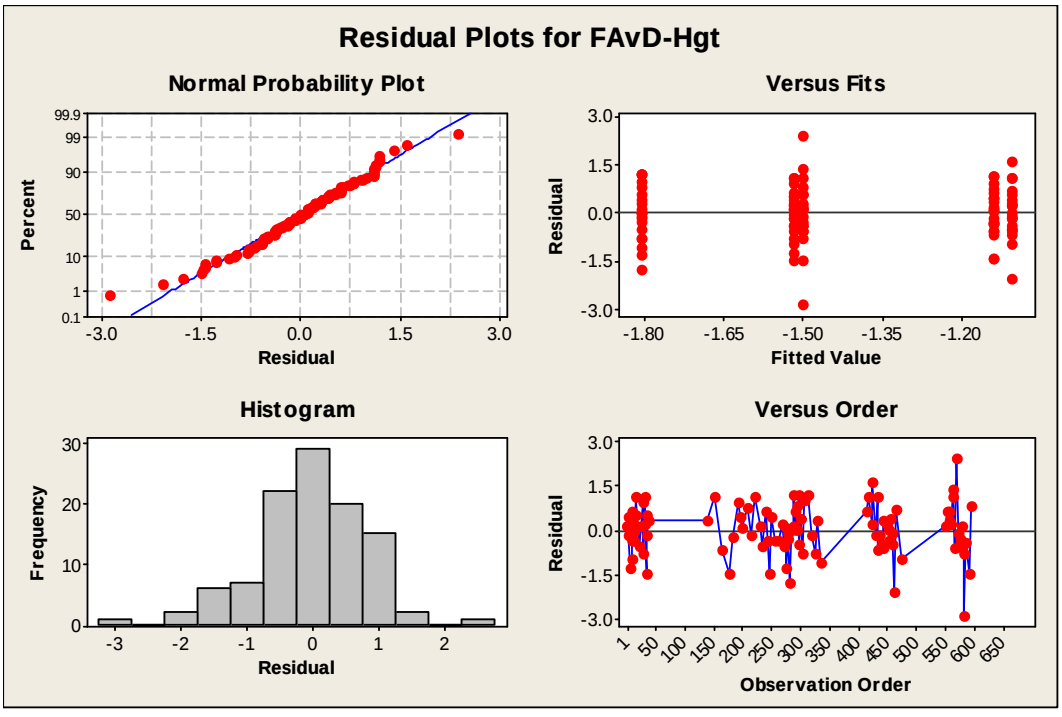


Figure 59. St1 One-Way ANOVA: FAvD-Hgt vs Welds - Residual Plots

Further, ANOVA was performed (Figures 58 and 59) to determine whether the average error would be different between the welds. W4 indicated some non-statistically significant difference

(Figure 58) on the box plot. Another summary plot (4 in 1) part of ANOVA shown in Figure 59 confirmed distribution normality (Normal Probability Plot) with only a few outliers and uniformity of residuals (difference between individual and average error) for all five specimens. More examples are shown in Appendix E.

The POD was estimated with off-the-shelf dedicated software packages mh1823 POD, PODv3 and STATUS. Most of the POD analysis was conducted with the mh1823 POD.[30] The different POD link and transformation functions used by the different packages would produce different results.[32] The mh1823 POD provides a criterion (minimum deviance) for selecting the best link function Eqs. (9) through (16) and whether a flaw size log transformation Eq. (18) or Eq. (19) is needed. Although mentioned on the diagnostic plot, no guidance is provided in mh1823 POD to determine whether the difference of the deviances for the different link functions producing significantly different $a_{90/95}$ is statistically significant. The MIL-HDBK-1823A[30] recommends similar link and transformation functions to be used when comparing similar sets (e.g., different AUT systems) of data. Other POD software packages do not provide any diagnostic tools for comparing "goodness of fit" of POD functions to the experimental data.

The statistical analysis consisted of two stages. During the first stage (St1), a minimum number of data points were removed and preliminary estimates of POD, average error and uncertainty were obtained. After removal of the pore (Vol) flaw category from the data, the second stage analysis focused on investigating statistical differences between different flaw categories, normality of data samples and obtaining final estimates of POD and sizing capabilities.

4.3.4 Results

A summary of all flaws is shown in Table 12 for each weld specimen and is divided into flaw categories. Flaws with intentional and unintentional skew and or tilt increased almost 4.6 times to 37 from the initially planned 8. The DT revealed that almost 1/3 (29 flaws) of the 90 initially planned flaws without skew or tilt had some tilt. It was not possible to reliably measure flaw skew angle during the specimen sectioning. The DT identified a significant number (108 flaws) of natural planar flaws in addition to the intentionally implanted flaws.

Table 12. Type and Number of Flaws for AUT Quantification

Flaw Type	W2[a]		W3[a]		W4[a]		W5[a]		W6[a]		Total		POD and Sizing Based on DT
	FA[b]	DT	FA	DT	FA	DT	FA	DT	FA	DT	FA	DT	
Implant No Skew and Tilt	19	11	16	13	16	10	20	14	19	12	90	60	59
Implant Unintentional Tilt	-	8	-	3	-	5	-	6	-	7	-	29	28
Implant Intentional Skew and Tilt	-	-	4	4	4	4					8	8	8
Implant Inter-Bead	4	4	2	2	2	2	-	-	-	-	8	8	8
Implant Transverse	1	1	1	1	1	1	-	-	-	-	3	3	-
Natural Planar	-	4	-	70	-	19	-	8	-	7	-	108	104
Natural Vol. (Pores)	-	7	-	42	-	18	-	4	-	3	-	74	74
Total	24	35	23	135	23	59	20	32	19	29	109	290	281

(a) Specimens W2 to W6
(b) FA – Fabrication specifications

The flaws from Table 12 used in the statistical sample for estimation of POD and sizing are shown in Figure 60 sorted by flaw type and weld specimen. Specimens W3 and W4 contained a relatively large number of Natural and Vol (pore) flaws indicating uncontrolled variability in the flaw fabrication process.

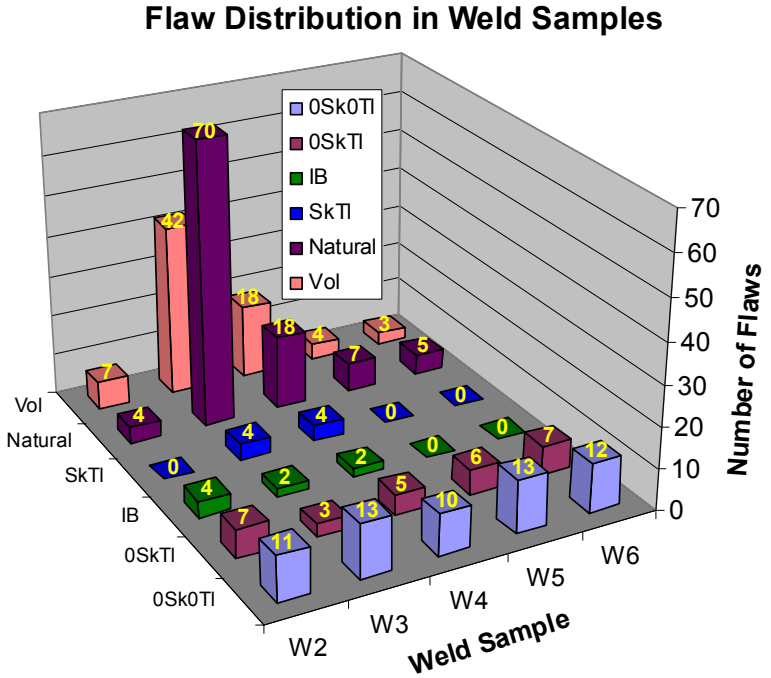


Figure 60. Flaw Distribution in Weld Samples

One of the flaw distributions of interest is the flaw height distribution shown in Figure 61. Most of the intentionally implanted flaws start from the interval 1.5 to 2.0 mm. The number of planar flaws decreases above 4.0-mm height. A significant number of Vol (68) and Natural (46) flaws are smaller than 0.5 mm. A large number of natural planar flaws is found in the height intervals 0.5 to 1.0 mm (35) and 1.0 to 1.5 mm (16).

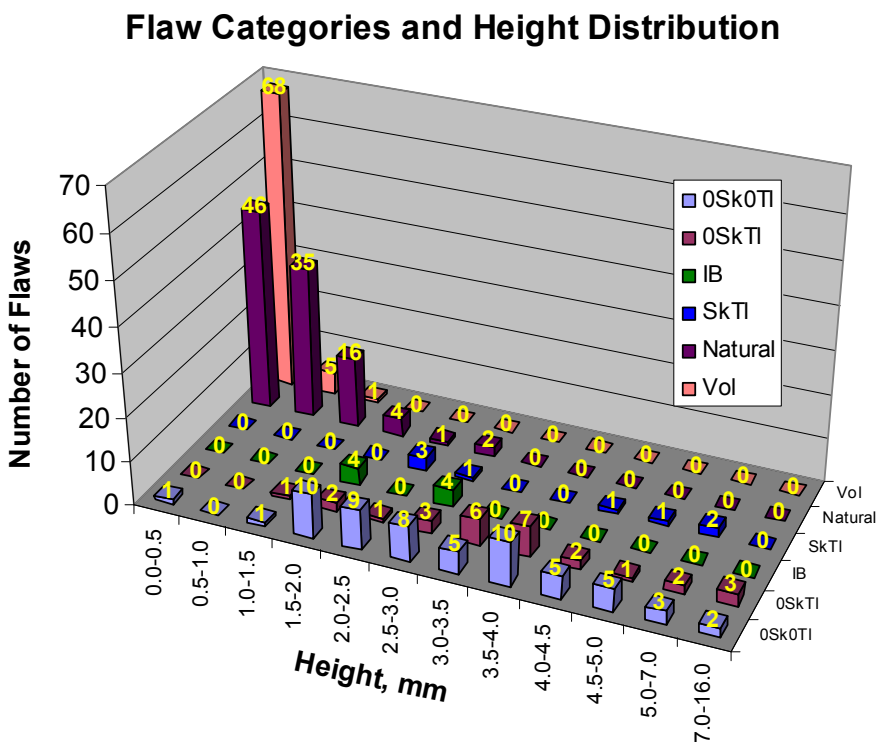


Figure 61. Flaw Categories and Height Distribution

Figure 62 shows the flaw length sorted into three length categories by flaw type as previously discussed. The largest number of flaws (85 out of 139) has a length larger than 12 mm. Only 16 flaws, mostly natural (12), are shorter than 6 mm. The total number (139) of flaws presented in Figure 62 is smaller than the total number of 281 used for POD and sizing shown in Table 12 because the ultrasonic fingerprinting (used for length measurement) did not detect all natural and Vol flaws. All 103 implanted planar flaws were detected during the fingerprinting.

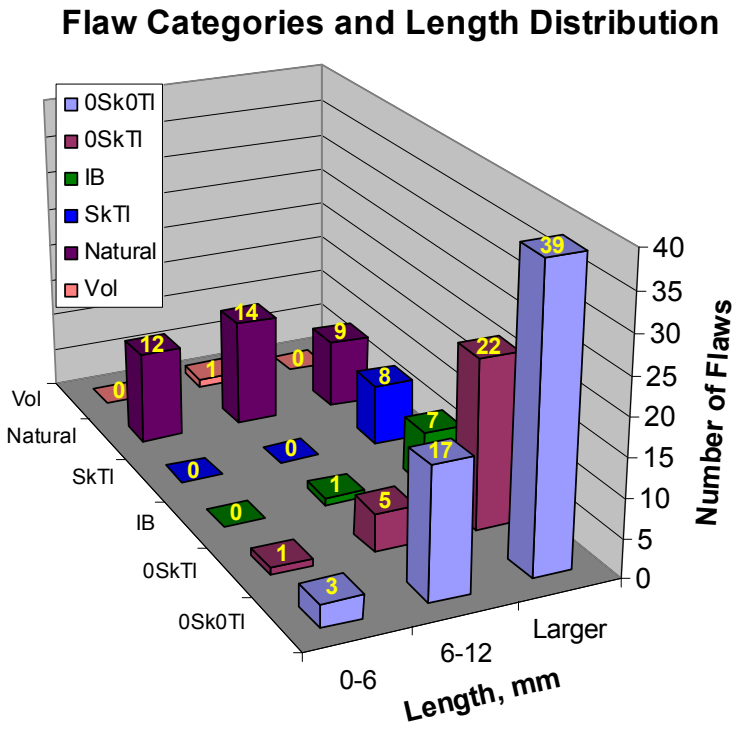


Figure 62. Flaw Categories and Length Distribution

The decision to use fingerprinting techniques for length reference measurements was driven by the significant cost of the DT if the length of all implanted and natural flaws was to be validated by DT with small increment of specimen section. In addition, radiography proved not to be reliable at detecting many planar flaws or the extent of the flaws. Only a limited number of flaws in the most challenging weld specimens W5 and W6 (as far as the length is concerned) were sectioned to validate the fingerprinting length measurement capabilities (Table 13). Thus, the length validation was conservative. The fingerprinting length measurements were in good agreement with the DT measurements.

The flaw depth distribution is shown in Figure 63. The cap zone is from 0 to 5 mm, the mid-wall is from 5 to 10 mm, and the root is from 10 to 15.6 mm. The flaw is considered to be in a certain zone when more than half of the flaw height is in the zone. The largest number of implanted flaws is in the weld root (44), followed by the mid-wall (35) and cap (24). The largest number of natural flaws is in the mid-wall (55). The root and cap have approximately the same number of natural flaws, 26 and 23, respectively.

Table 13. Validation of Fingerprinting Length Sizing with DT

Weld	Sector	Axial Posit.	FA[(a)] (mm)		F (mm)		Length (mm) (F)	Length (mm) (DT)	Note
			Start	Stop	Start	Stop			
W5	S25	C[(b)]	966	975	969	976	7	9 ± 3	Good agreement
W5	S25	US	980	995	980	990	10	15 ± 3	2 mm smaller than smallest possible
W5	S26	DS	1000	1012	1000	1013	13	15 ± 3	Good agreement
W6	S13	DS	483	498	485	494	9	12 ± 3	Good agreement
W6	S32	US	1240	1244	1239	1245	6	9 ± 3	Good agreement
W6	S32	US	1248	1258	1248	1259	11	9 ± 3	Good agreement
W6	S32	US	1264	1276	1265	1277	12	12 ± 3	Good agreement

(a) FA – Fabrication specifications
(b) C – Center

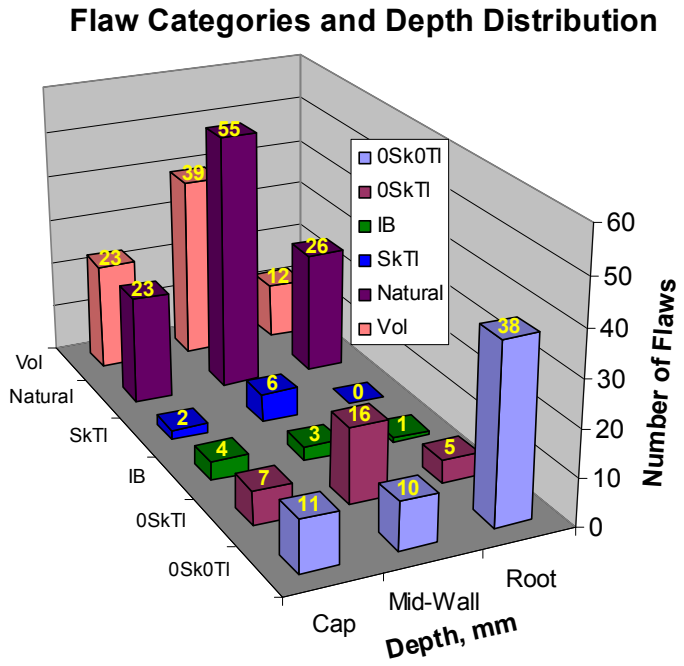


Figure 63. Flaw Categories and Depth Distribution

As discussed previously, some of the flaws were removed from the data sample initially. A list of flaws that were removed for all systems and the reason for removal is shown in Table 14. One of the implanted and field-generated natural-flaw features that will strongly influence flaw detectability is the closeness or tightness of planar flaws. The flaw then becomes partially or completely transparent to the sound energy and might be undetectable. It is difficult to estimate the ultrasonic transparency of a flaw by only visually analyzing a macrograph of a tight flaw. One technique suggested here is inspection of the macro surface with high resolution eddy current techniques. The eddy current techniques, if optimized, could penetrate deeper (~1 to 2

mm) in the specimen and provide estimate of whether the flaw is fused or not. This type of testing was not planned initially and off-the-shelf equipment and procedures were used. Two flaws removed from the data sample with eddy current scans (left) and flaw macro (right) are shown in Figures 64 and 65. The flaw with height of 6.5 mm according to the macro shown in Figure 64 was missed by the fingerprinting and all five AUT systems. This size of the flaw height is expected to be well in the range of the AUT system capabilities based on past experience. The flaw height measured with the eddy current technique was only 3 mm which indicated that the flaw might be partially fused. The eddy current scan of a macro from W4 shown in Figure 65 indicated almost complete fusion. The flaw was again missed by the fingerprinting and all AUT systems. More eddy current scans for flaws (total of 43 flaws) difficult to detect with the fingerprinting and AUT systems are available in Appendix F. The eddy current technique helped the decision process of reliably characterizing and removing large flaws that would negatively affect the POD estimates if left in the data sample. However, more work is needed to be done to optimize the eddy current technique performance for smaller flaws, better penetration and higher surface resolution.

Table 14. Flaws Removed from POD and Sizing Sample for All Systems

Weld	Sector	Flaw Type	Depth (mm)	Height (mm)	Reason
W2, W3, W4	As applicable	Implanted, transverse	--	--	Not used (industry practice)
W2	S52	Implanted, tilted	13.7	6.5	Partially fused (EC). Missed by all.
W4	S43	Natural, planar	11.9	2.4	Fused (EC). Missed by all.
W5	S14	Implanted, no skew or tilt	10.3	4.3	Multiple flaw interaction
W5	S20	Natural, planar	10.3	0.6	Shadowed by implanted
W6	S13	Natural, planar	7.5	5.8	Multiple flaw interaction
W6	S52	Natural, planar	11.8	9.9	Multiple flaw interaction

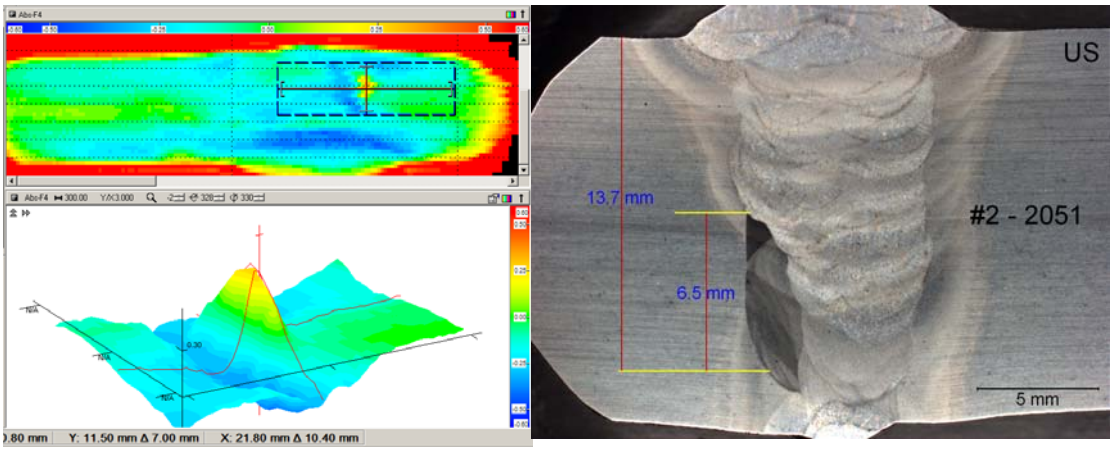


Figure 64. EC Scan, Flaw W2-2051 in S52 Removed from the POD and Sizing Sample

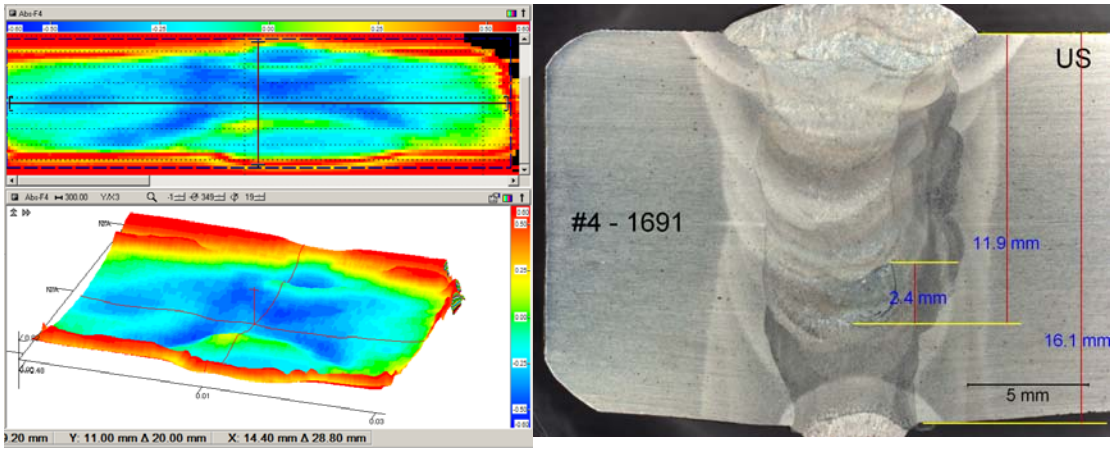


Figure 65. EC Scan, Flaw W4-1691 in S43 Removed from the POD and Sizing Sample

Macros of interacting flaws removed from the data (Table 14) are shown in Figure 66. Outliers related to length sizing with extremely large errors were also removed from the data samples for different AUT systems (Table 15).

Table 15. Outliers Removed from Sizing Sample for Some Systems – Length Sizing

Group	Weld	Sector	Error (mm)	Flaw Type	Reason
V3vF-Len-PA	W3	S47	-161	Nat. planar	Not representative (fingerprinting)
V5vF-Len-PA	W6	S3	55	Impl. no skew or tilt	Combined with flaw in S2 (system)
V5vF-Len-XRM	Same	Same	52.3	Same	Same
V4vF-Len-PA	W3	S29	40	Impl. skew and tilt	Not representative (system)
V4vF-Len-PA	W6	S40	--	Impl. No skew or tilt	US and DS reported with single length sum of both (system)

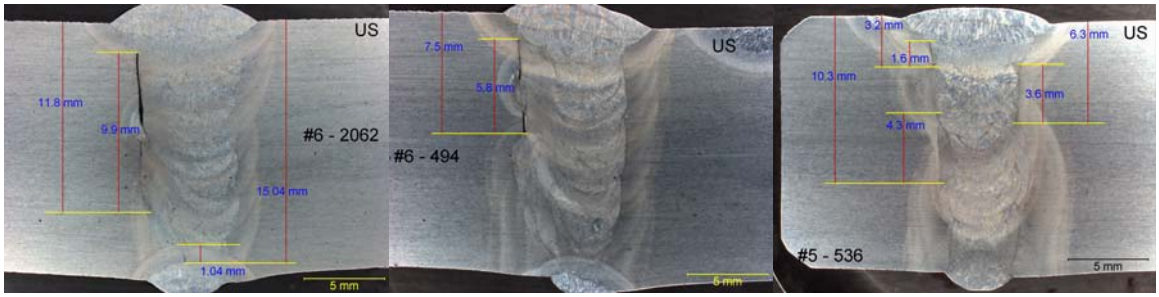


Figure 66. Interacting Flaws Removed from POD and Sizing Sample

As discussed, the W3 specimen had a large number of Natural and Vol flaws. This required a significant amount of the examiner's time to be dedicated to data analysis and reporting which interfered with the scheduling activities of the inspection providers participating in the trials.

This was the case with the V5 system where data from the inspection of the W3 specimen was not provided.

In summary, a sufficient number of 103 implanted flaws (60 minimum needed) shown in Table 12 and Figures 60 through 63 was available to estimate the POD and sizing capabilities of the AUT systems. Natural flaws were used or removed from the analysis depending on whether POD or sizing capabilities were estimated.

4.3.4.1 Stage 1

Following the data analysis procedure discussed earlier, data processing and analysis was conducted to obtain initial estimates of the detection and sizing capabilities, investigate data normality and determine whether statistically significant differences existed between the weld specimens. A full set of estimates for each weld specimen and joint (welds W2 through W6) distributions is available in Appendix E when either DT or fingerprinting was used as reference.

W2 and W3 had some areas where the WT was slightly smaller than the minimum thickness found in the pipe specifications (refer to Section 4). One question that needed immediate answer was whether the AUT system performance in the areas with slightly reduced WT was affected. A second set of data for each weld, designated as W2nt and W3nt, was formed where the flaw measurements in the areas with reduced thickness were removed. The samples with and without measurements in reduced WT were compared (W2nt vs W2 and W3nt vs W3) following the standard statistical procedures. No statistically significant difference was found. Therefore, data points obtained in areas with reduced WT were used for the AUT quantification. More details with quantitative results from the comparison can be found in Appendix E.

One important step of ensuring weld specimen quality is verification of the precision and control of the flaw fabrication process. The flaw fabrication process is evaluated first during the fingerprinting and later during the DT. According to the specimen fabrication specifications, the flaw dimensions and positions are expected to have an error that does not exceed ±1 mm. Preliminary data is unavailable regarding the error distribution during the flaw fabrication. One approach recommended in the literature[42] is to assume a uniform distribution of the error in the interval ±1 mm. This automatically produces an uncertainty (standard deviation) of 0.58 mm $[(1mm)/\sqrt{3}]$ for the uniform error distribution. Further, the average fabrication error is expected to be zero. This assumption will allow comparison to the error estimates (average and standard deviation) for the fabrication process generated during the fingerprinting and DT.

Several fingerprinting techniques were used in this study: ultrasonic with PA and TOFD, x-ray, and MT. The DT was conducted through specimen sectioning with step of 3 mm (refer to Section 4).

Joint (combined) distributions of the reference methods versus the fabrication specifications as a function of flaw height, length, and depth are shown in Figures 67 through 69. Results show that the flaws are fabricated with height larger (~ 1.5 mm) than the specifications (Figure 67). The scatter is relatively large ~ ±1.6 mm for 95% confidence. The systematic error is approximately constant or additive across the entire range. A larger number of flaws with larger height may actually change this trend because most of the flaws have heights between 0.5 and 3.5 mm. The systematic fabrication error for flaw length is relatively small and additive as shown in Figure 68. The scatter, though, is relatively large at ±10 mm for 95% confidence. The depth systematic error increases to ~ 1.5 mm with reduced depth which indicates a multiplicative pattern (Figure 69). The scatter of fabricated flaw depth error is ~ ±2 mm for 95% confidence.

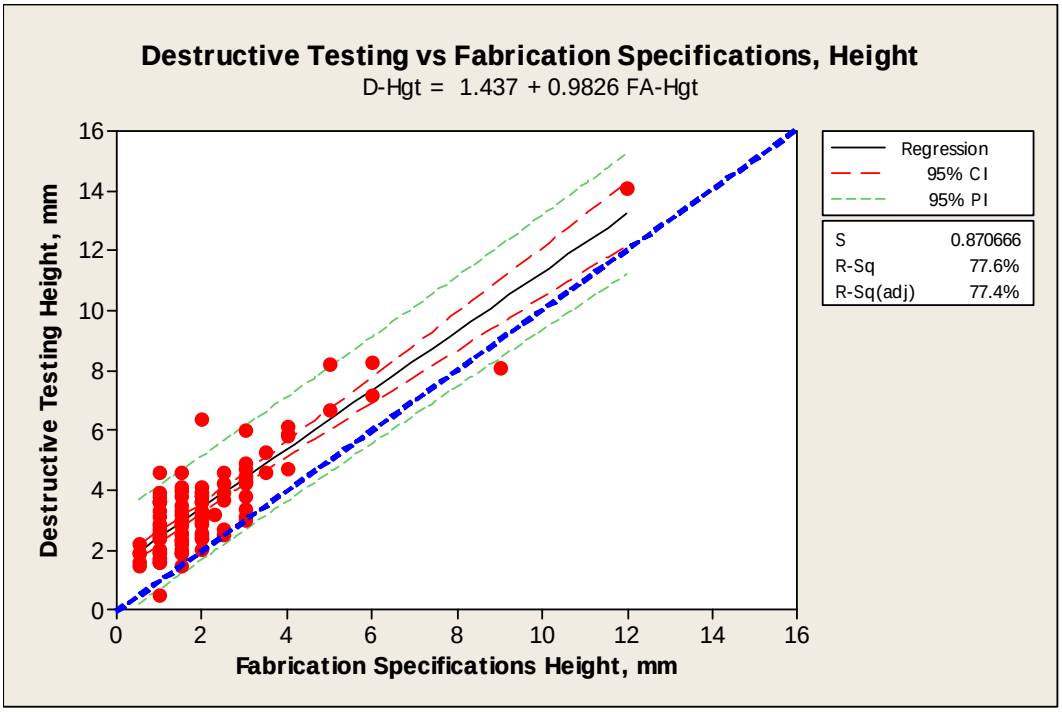


Figure 67. St1 DT vs Fabrication Specifications for Flaw Height

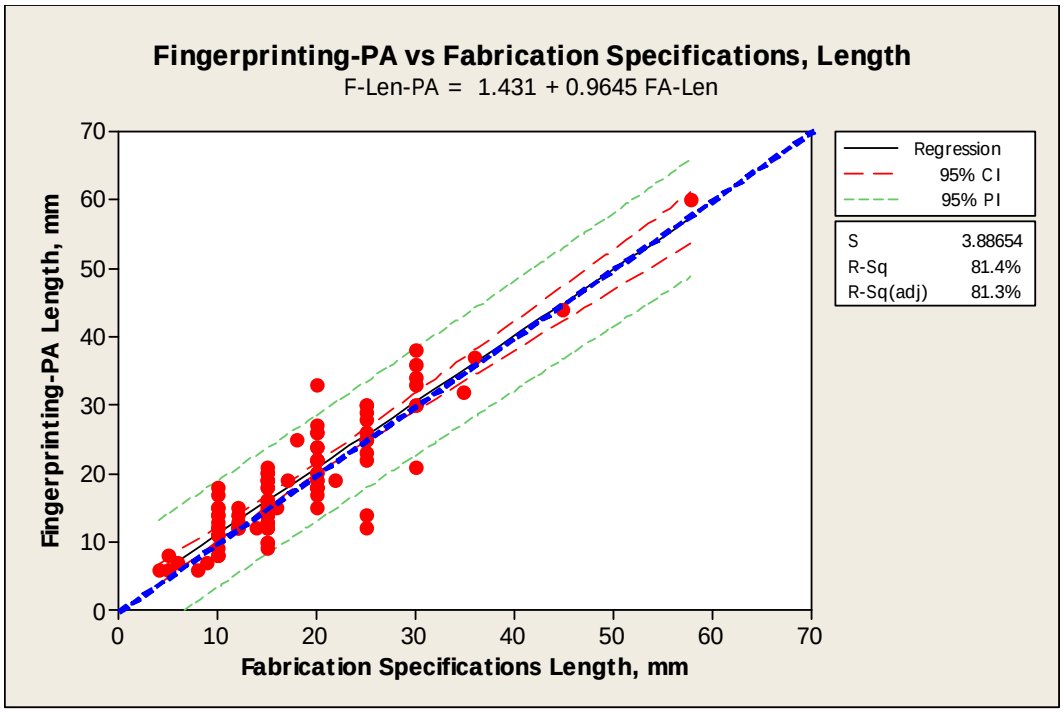


Figure 68. St1 Fingerprinting-PA vs Fabrication Specifications for Flaw Length

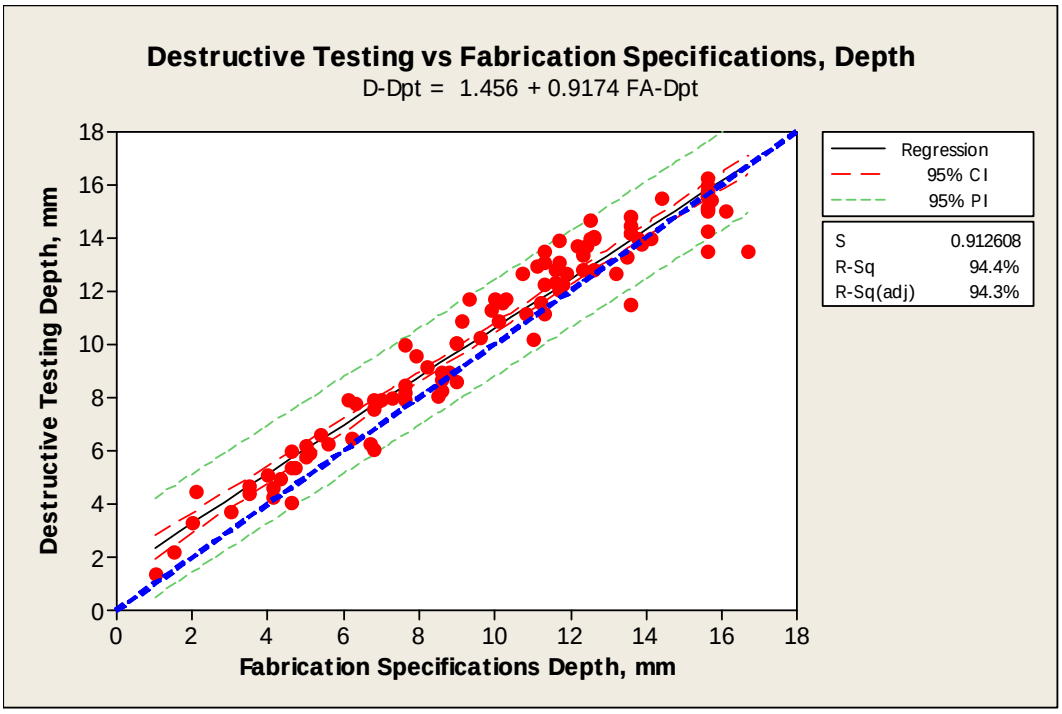


Figure 69. St1 DT vs Fabrication Specifications for Flaw Depth

The results from the fabrication process verification are summarized in Table 16. Many of the distributions are not normal. The systematic error estimate for flaw height is similar for the two reference methods, different from zero and outside of the specified interval. The uncertainty for the height error is also larger than 0.58 mm. The length and depth systematic errors are still different from zero but inside the allowed interval. The estimated length and depth uncertainties are larger (especially length uncertainty) than 0.58 mm.

The uncertainty of the start and stop positions is approximately 10 times larger than the expected 0.58 mm. The statistical analysis of the difference between the different welds indicated that W3 was different from the other specimens. The W3 specimen sector marking was done in an opposite direction initially and had to be corrected later. This possibly caused the large systematic errors and uncertainty for this specimen and the joint distribution. Specimen W3 also had the largest number of Natural and Vol unintentional fabrication flaws (Figure 60).

More quantitative data regarding the fabrication process verification for each weld is available in Appendix E.

Table 16. St1 Summary – Fabrication Specifications vs Fingerprinting and Destructive

Group	Sample	Normal Distribution (Y/N)	OTL[(a)]	s(ε) (mm)	Aver ε (mm)	Different Variances? (Y/N)	Different Distributions?	
							Test 1, Y/N	Test 2, Y/N
FAvD-Hgt	105	Y	3	0.87	-1.41	N	N	--
FAvF-Hgt-PA	104	Y	1	1.35	-1.44	N	Y (W2/5 and 6, W3/5, W4/5 and 6)	--
FAvF-Hgt-TD	74	Y	2	1.03	-1.59	N	N	--
FAvF-Len-PA	104	N	3	3.87	-0.77	N	--	N
FAvF-Len-XRM	50	N	5	3.46	0.9	Y	-- (W2/4)	Y
FAvD-Dpt	105	N	3	1	-0.66	N	--	N
FAvF-Dpt-PA	104	Y	2	1.31	-0.31	N	N	--
FAvF-Dpt-TD	78	Y	1	1.35	-0.82	N	N	--
FAvF-Sta	104	N	8	5.49	1.83	Y	-- (W3/others)	Y
FAvF-Stp	104	N	15	6.14	0.73	Y	-- (W3/others	Y

(a) OTL – Outliers

Initial estimates of the accuracy of height, length, and depth measurements were obtained at St1. The DT was used as reference for the fingerprinting and the AUT system quantification. For height and depth measurements, fingerprinting was used as a second reference to which the AUT system performance was compared. The effect of flaw skew and tilt on the measurements was analyzed.

The flaw categories used for the St1 analysis are different from the flaw categories described earlier. The St1 flaw categories are shown below:

- **0Sk0Tl -** Implanted, planar, no skew and/or tilt.
- **nSknTl -** Natural, planar, non-intentional skew and or tilt. The flaws were pooled in one sample regardless of the actual skew or tilt angle.
- **SkTl -** Implanted, planar, intentional skew and tilt. These flaws had different skew and tilt angles. The flaws were pooled in one sample regardless of the actual skew or tilt angle.
- **Vol -** Pores and other volumetric flaws.

The data was analyzed for each weld specimen and joint distributions (all welds). Tables and plots for the single weld specimens and joint distributions are available in Appendix E.

The fingerprinting PA height measurements versus DT for the pooled data from all five welds are shown in Figure 70. A clear trend of undersizing is indicated with the increased flaw height. The different flaw types did not affect the height measurement accuracy for the FvD-Hgt-PA as illustrated in Figure 71.

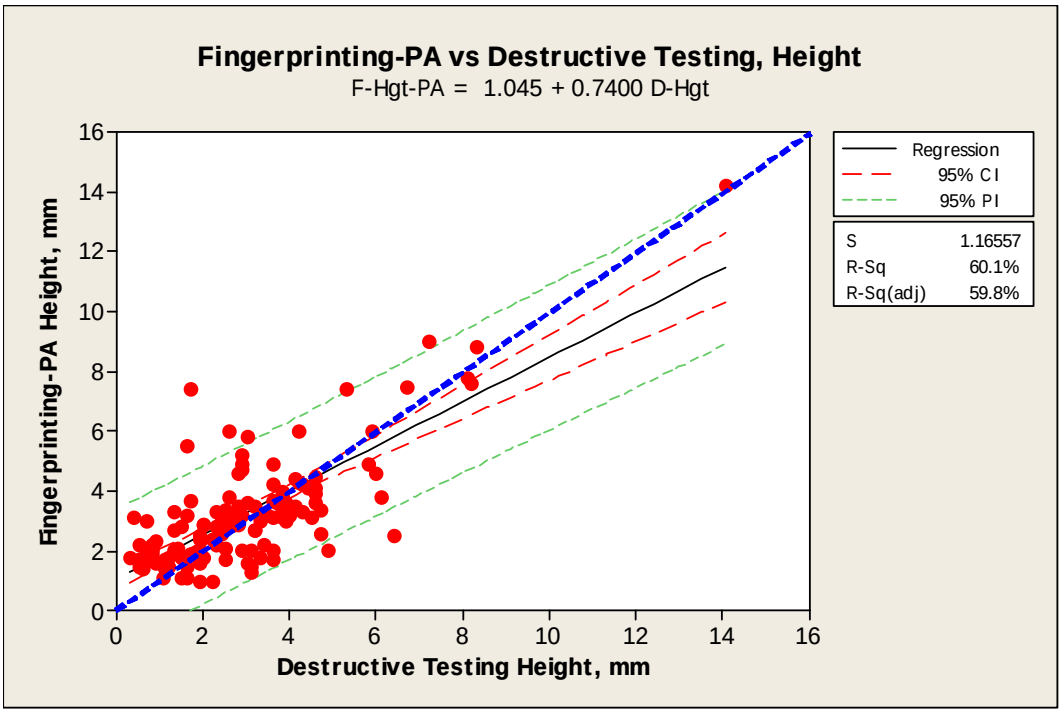


Figure 70. St1 Fingerprinting-PA vs DT for Flaw Height

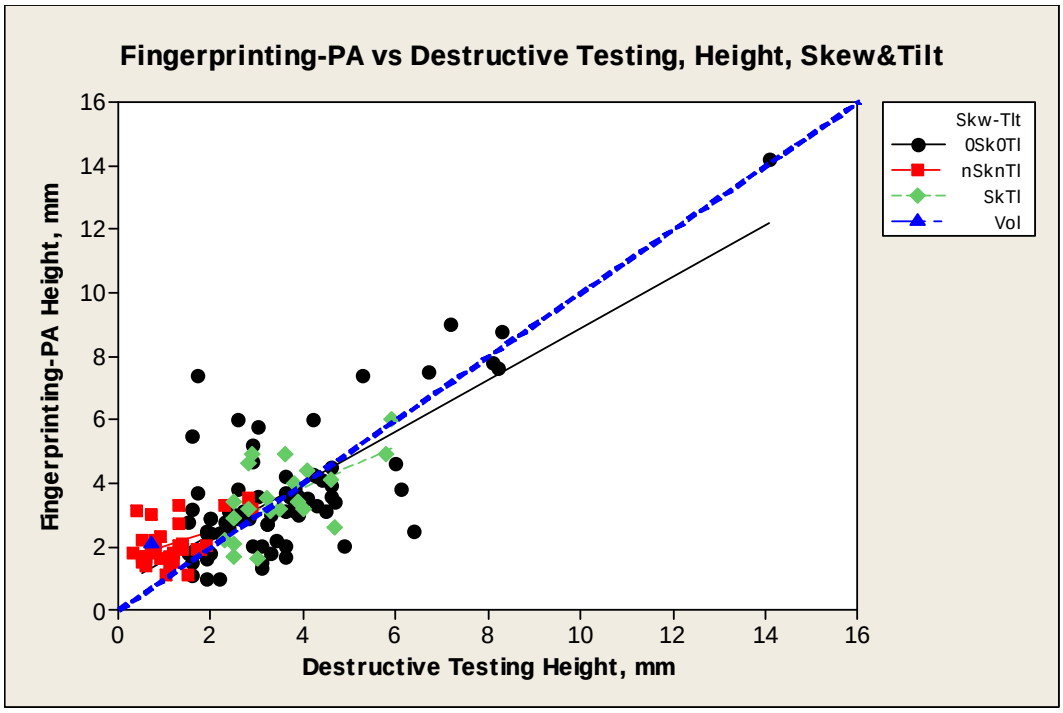


Figure 71. St1 Fingerprinting-PA vs DT for Flaw Height - Flaw-Type Effect

A typical plot of AUT system V4 versus DT for the flaw height measurements is shown in Figure 72. A similar trend (Figure 70) of undersizing with the increased flaw height is presented in Figure 72. The flaw skew and tilt, though, indicated some effect on the sizing trends (Figure 73). The spread of data with respect to the regression line was not affected by the flaw type. More data with skewed and tilted flaws in larger height ranges would help in obtaining more reliable estimates of the flaw-type effect.

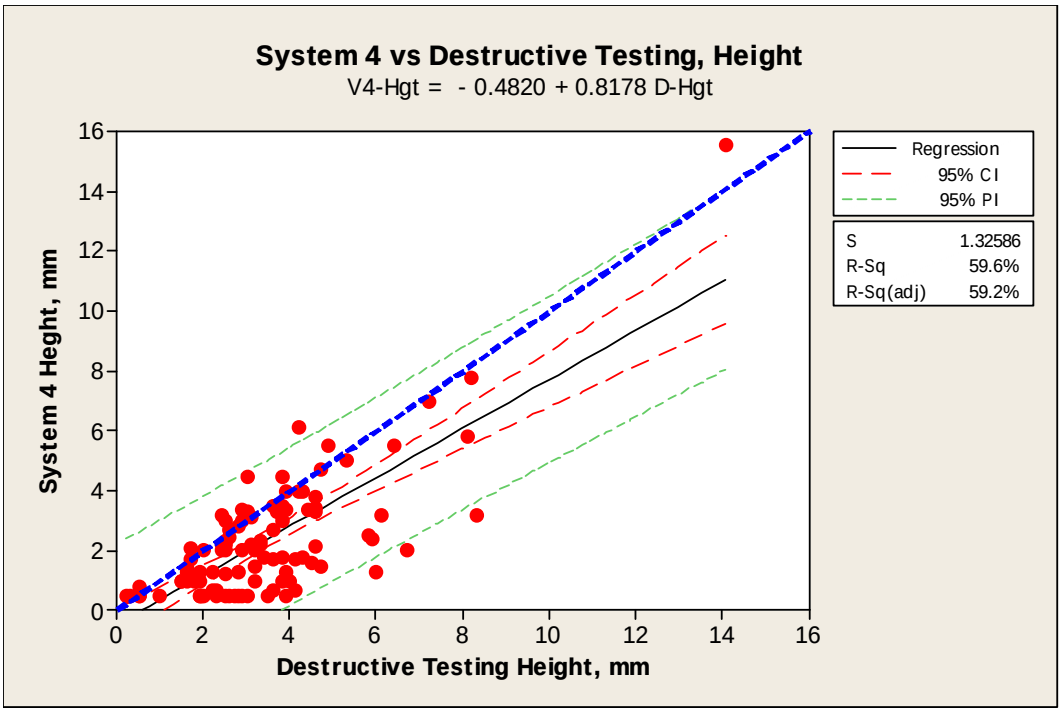


Figure 72. St1 System 4 vs DT for Flaw Height

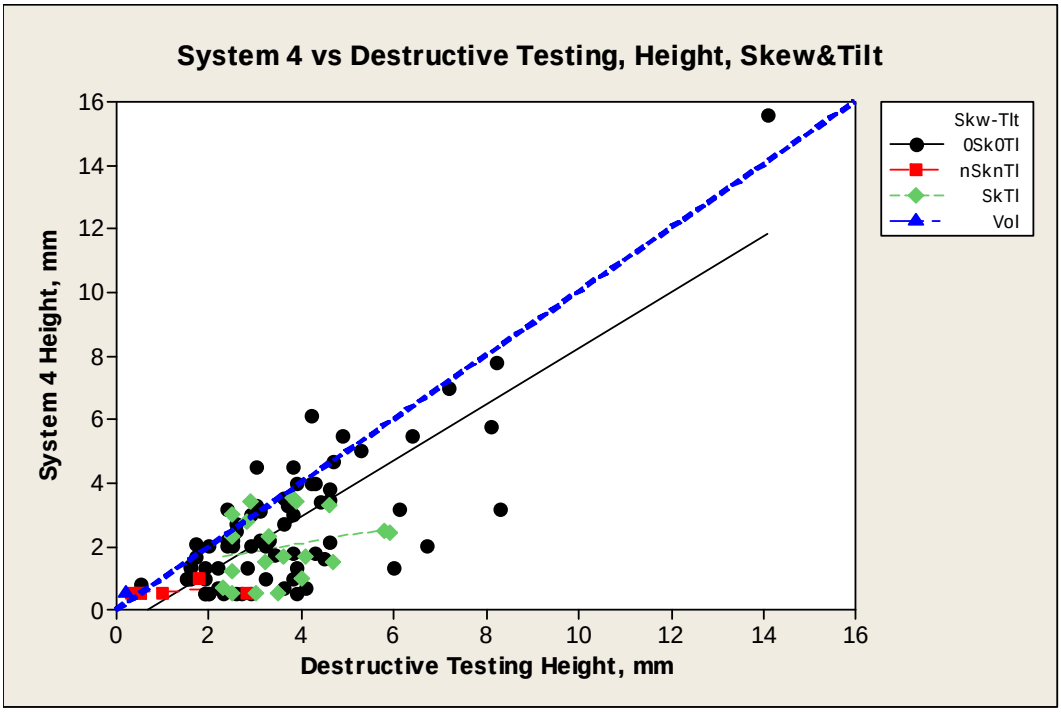


Figure 73. St1 System 4 vs DT for Flaw Height - Flaw-Type Effect

Additional plots in the format shown in Figures 70 through 73 for the other fingerprinting technique (TOFD) and the remaining four AUT systems regarding the height measurements are available in Appendix E.

The height error estimates and comparison of the distributions at St1 are summarized in Table 17 for the fingerprinting and all five AUT systems with the two reference techniques – DT and fingerprinting. For the case when the fingerprinting was used as reference for the AUT systems, the fingerprinting data was corrected based on the DT results to remove the fingerprinting systematic error component.

Table 17. St1 Summary of Height Measurements (Reference) for DT and Corrected Fingerprinting

Group	Sample	Normal Distribution (Y/N)	OTL[(a)]	s(ε) (mm)	Aver ε (mm)	Different Variances? (Y/N)	Different Distributions?	
							Test 1 (Y/N)	Test 2 (Y/N)
FvD-Hgt-PA	138	N	5	1.27	0.29	N	-- (W2/6, W3/5 and 6, W4/5 and 6)	Y
FvD-Hgt-TD	84	N	1	0.92	0.2	Y	-	N
V1vD-Hgt	92	Y	3	2.15	-0.86	N	Y(W2/4)	-
V2vD-Hgt	71	N	4	2.06	-0.52	N	--	N
V3vD-Hgt	63	Y	1	1.61	-0.92	N	N	-
V4vD-Hgt	100	N	1	1.47	-1.16	N	--	N
V5vD-Hgt	75	N	0	1.29	-0.75	N	--	N
V1vF-Hgt-PA	99	N	4	1.97	-0.83	N	-- (W2/4)	Y
V2vF-Hgt-PA	72	N	5	1.9	-0.59	N	-- (W2/5)	Y
V3vF-Hgt-PA	65	N	2	1.64	-0.97	Y	--	N
V4vF-Hgt-PA	100	N	5	1.42	-1.09	N	--	N
V5vF-Hgt-PA	79	Y	2	1.33	-0.78	Y	--	N
V1vF-Hgt-TD	72	Y	2	2.17	-0.79	N	Y(W2/4)	--
V2vF-Hgt-TD	56	Y	2	2.13	-0.34	N	N	--
V3vF-Hgt-TD	55	Y	1	1.46	-0.84	N	N	--
V4vF-Hgt-TD	74	Y	3	1.35	-1.11	N	N	--
V5vF-Hgt-TD	57	Y	1	1.4	-0.63	Y	--	N

(a) OTL – Outliers

The following observations can be made from the data presented in Table 17:

- The height is systematically undersized by all systems and slightly oversized during the fingerprinting.
- Many of the distributions (9 out of 17) are not normal especially for the DT and the PA as reference.
- The undersizing increased as the height increased. In other words, the systematic error is multiplicative.

- Trend lines on the plots indicated that the skew and tilt might have strong effect on the sizing. The spread of the data points for flaws with and without skew and tilt, however, are similar.
- More data with larger statistically significant number of skew and tilt flaws is needed to assess reliably the skew and tilt effect through physical experimentation.

The POD estimates are conducted with the following assumptions at St1:

- The POD is assumed to be a function of the flaw height (1 dimensional) only. This assumption may not be true for flaws that are smaller than the beam height at the flaw position.
- The data for the POD was interpreted using the *hit/miss* technique. The *â vs a* technique was not used as discussed earlier because the amplitude measurements were unreliable and unavailable for some AUT systems.
- Most of the data was analyzed with the mh1823 POD software using various link and transformation functions.
- The total number of flaws for all five welds was 282. Out of all 282 flaws, there were 74 pores smaller or equal to 1.1 mm missed by all AUT systems and the fingerprinting.
- Some parallel POD estimates were obtained with another software - PODv3.

The POD for the fingerprinting versus DT is shown in Figure 74. The three major POD estimates rounded to the second digit after the decimal point are as follows: a_{50} - 1.02 mm, a_{90} - 1.6 mm, and $a_{90/95}$ - 1.89 mm. The flaw height was not log transformed. It is important to note that the fingerprinting was conducted as open trials. A POD curve from the same data obtained with the PODv3 software is shown in Figure 75. The POD parameters are as follows: a_{50} - 0.87 mm, a_{90} - 1.72 mm, and $a_{90/95}$ - 2.04 mm. The flaw height had to be log transformed to obtain a solution for the POD curve parameters with the PODv3 software. The estimates obtained with the two software tools are similar.

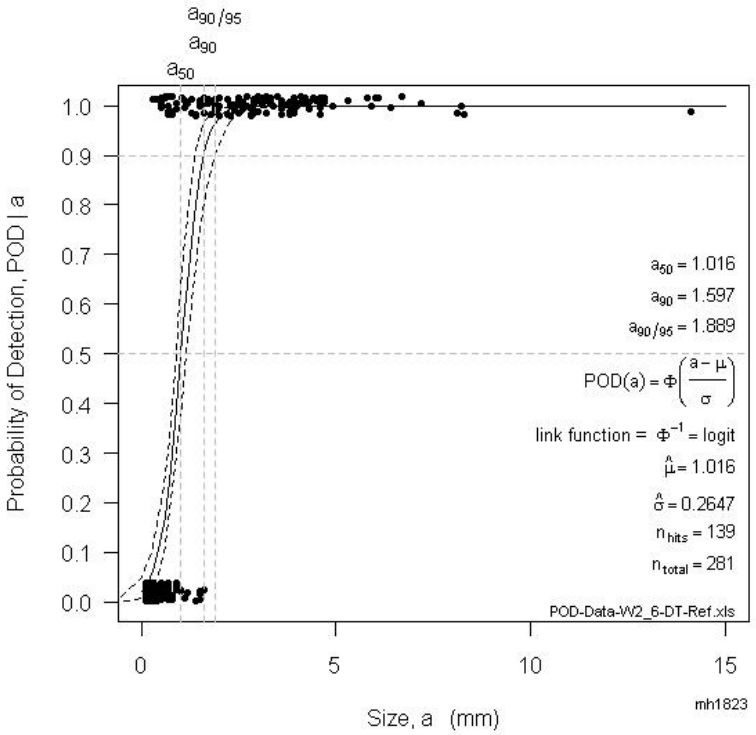


Figure 74. St1-R1 POD - Fingerprinting-PA vs DT for Flaw Height

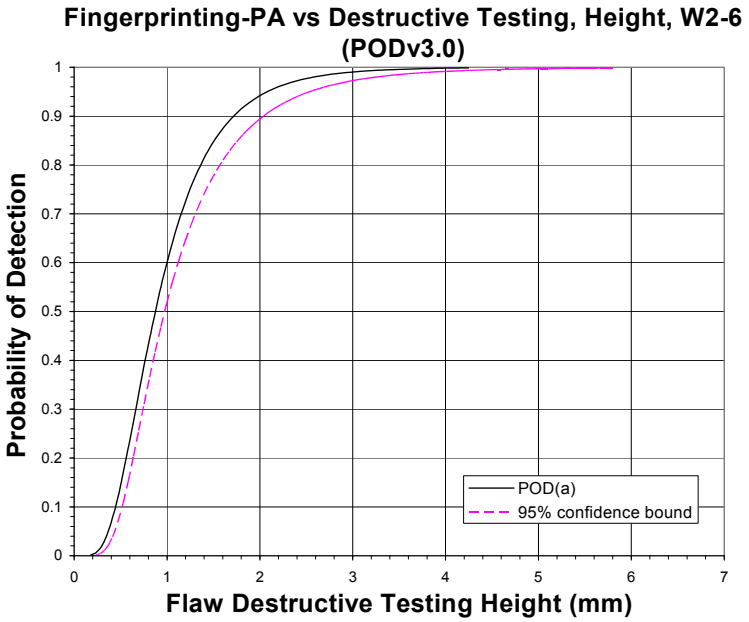


Figure 75. St1 POD - Fingerprinting-PA vs DT for Flaw Height - PODv3

One typical POD curve for the V1 AUT system is shown in Figure 76. The POD parameters for V1 are as follows: a_{50} - 2.02 mm, a_{90} - 3.03 mm, and $a_{90/95}$ - 3.44 mm. The estimates are larger than the fingerprinting POD (Figure 74). The trials with the AUT systems were blind. A combined plot for all five AUT systems is shown in Figure 77. The main POD parameter $a_{90/95}$ increased to 4.15 mm which is more than 2 mm larger than the $a_{90/95}$ obtained for the fingerprinting. The POD calculations for the combined data sample were repeated with the PODv3 software and the POD curve is shown in Figure 78. The POD parameters are as follows: a_{50} - 2.39 mm, a_{90} - 3.86 mm, and $a_{90/95}$ - 4.04 mm. The flaw height had to be log transformed to obtain a solution for the POD curve parameters with the PODv3 software. The estimates obtained with the two software tools are again very similar.

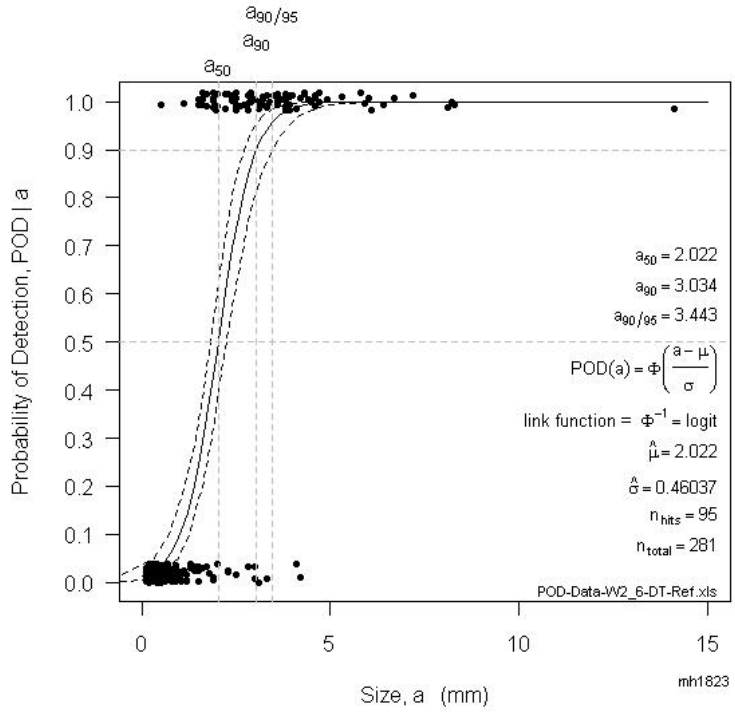


Figure 76. St1-R1 POD - System 1 vs DT for Flaw Height

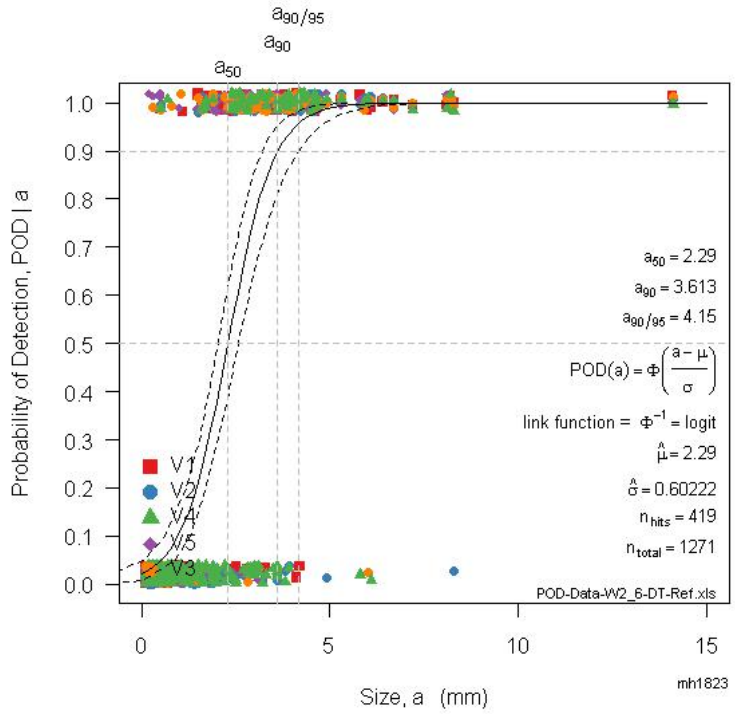


Figure 77. St1-R1 POD - All Systems vs DT, Height (W2-W6)

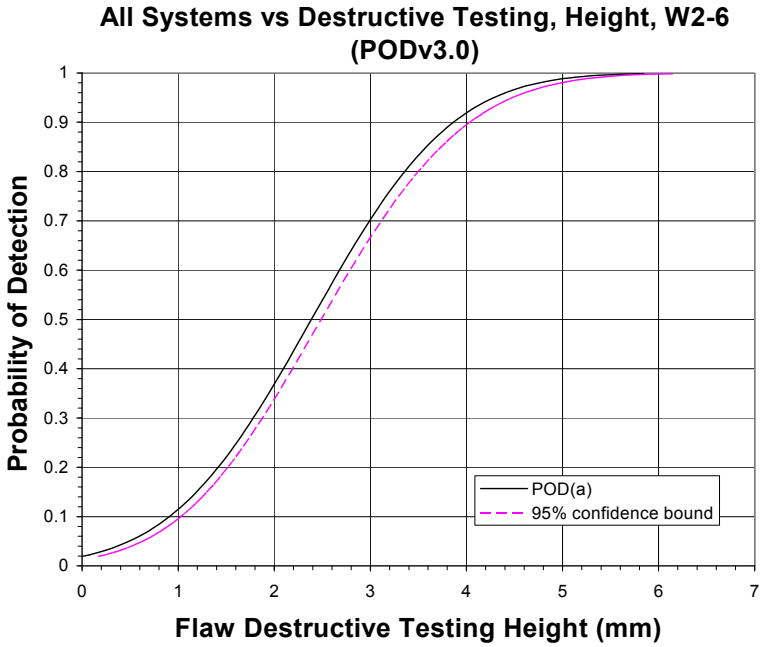


Figure 78. St1 POD - All Systems vs DT, Height (W2-W6), PODv3

More POD plots at St1 for the other systems are available in Appendix E.

The POD estimates obtained with the mh1823 POD for St1 are summarized in Tables 18 and 19. The flaw height was not transformed in Table 18 and only one link function (logit) was used to generate the estimates. The flaw height transformation and link function selection for the data presented in Table 19 was done using the minimum deviance criteria. With one exception, the logit link function was the best fit to the data regardless of whether log transformation of the height was done or not (Tables 18 and 19). For the fingerprinting, the logit (Table 18) with larger deviance provided very similar estimates compared to the cloglog (Table 19) with the minimum deviance. This agrees well with the results from a previous study[48] regarding the logit function as being the best function to fit the POD data in most of the cases.

Table 18. St1-R1 Summary POD for Flaw Height, All Systems vs DT, Logit Link, mh1823 POD

Group	a_{50} (mm)	a_{90} (mm)	$a_{90/95}$ (mm)	Link	Transf	Deviance
FvD-PA	1.016	1.597	1.889	Logit	No	133
V1vD	2.022	3.034	3.443	Logit	No	131
V2vD	2.632	4.126	4.739	Logit	No	166
V3vD	3.012	4.594	5.294	Logit	No	156
V4vD	1.839	2.794	3.185	Logit	No	132
V5vD	1.711	2.956	3.552	Logit	No	88
ALLvD	2.29	3.613	4.15	Logit	No	741

Table 19. St1-R1 Summary POD for Flaw Height, All Systems vs DT, Minimum Deviance, mh1823 POD

Group	a_{50} (mm)	a_{90} (mm)	$a_{90/95}$ (mm)	Link	Transf	Deviance
FvD-PA	1.153	1.66	1.999	Cloglog	Yes	128
V1vD	1.777	3.225	3.989	Logit	Yes	118
V2vD	2.3	4.959	6.649	Logit	Yes	150
V3vD	2.743	5.957	8.337	Logit	Yes	149
V4vD	Table 18	Table 18	Table 18	Logit	No	132
V5vD	1.335	2.968	4.082	Logit	Yes	79
ALLvD	1.952	4.218	5.573	Logit	Yes	694

Some of the plots (Figures 74, 76, and 77) and Tables 18 and 19 refer to a Revision 1 of the data as St1-R1. Only minor corrections and inconsistencies were detected and fixed while working with the raw data files. Neither the POD or sizing (presented earlier) estimates required recalculation. Summary tables with information regarding the POD estimates for the St1 data are available in Appendix E. The data is almost identical to the data presented in Tables 18 and 19.

An attempt was made as part of the initial analysis to estimate the POD for implanted flaws only by excluding the Natural and Vol categories. In a majority of the cases, the mh1823 POD software was unable to converge to a solution. The convergence problems were most likely caused by the relatively small number of missed implanted flaws in the POD sample. Additional estimates using the probit functions were also obtained. Summary tables with this additional data related to the POD for implanted flaws only and the probit POD function are available in Appendix E.

The flaw length measurements are the second major group which accuracy and consistency across the size range needs to be quantified for the AUT systems. The following notes apply to the length sizing quantification:

- The AUT systems length measurements are compared to the fingerprinting techniques only (PA and X-Ray) as reference.

- The same four flaw categories used for the height measurements were used for the length measurements as well:
    - 0Sk0Tl - Implanted, planar, no skew and/or tilt
    - nSknTl - Natural, planar, non-intentional skew and or tilt. The flaws were pooled in one sample regardless of the actual skew or tilt angle.
    - SkTl - Implanted, planar, intentional skew and tilt. These flaws had different skew and tilt angles. The flaws were pooled in one sample regardless of the actual skew or tilt angle.
    - Vol - Pores and other Volumetric flaws

One typical plot of system V2 length measurements versus fingerprinting is shown in Figure 79. Slight reduction of the systematic component is observed as the length increases. The flaw types are not affecting the trends across the measurement range (Figure 80).

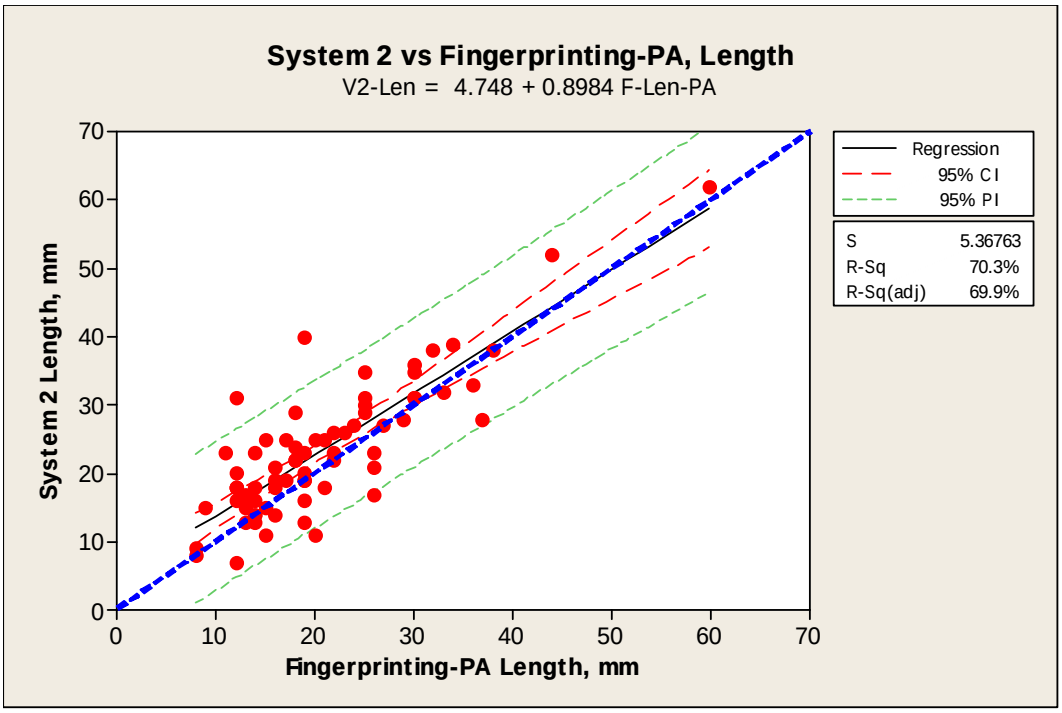


Figure 79. St1 System 2 vs Fingerprinting-PA for Flaw Length

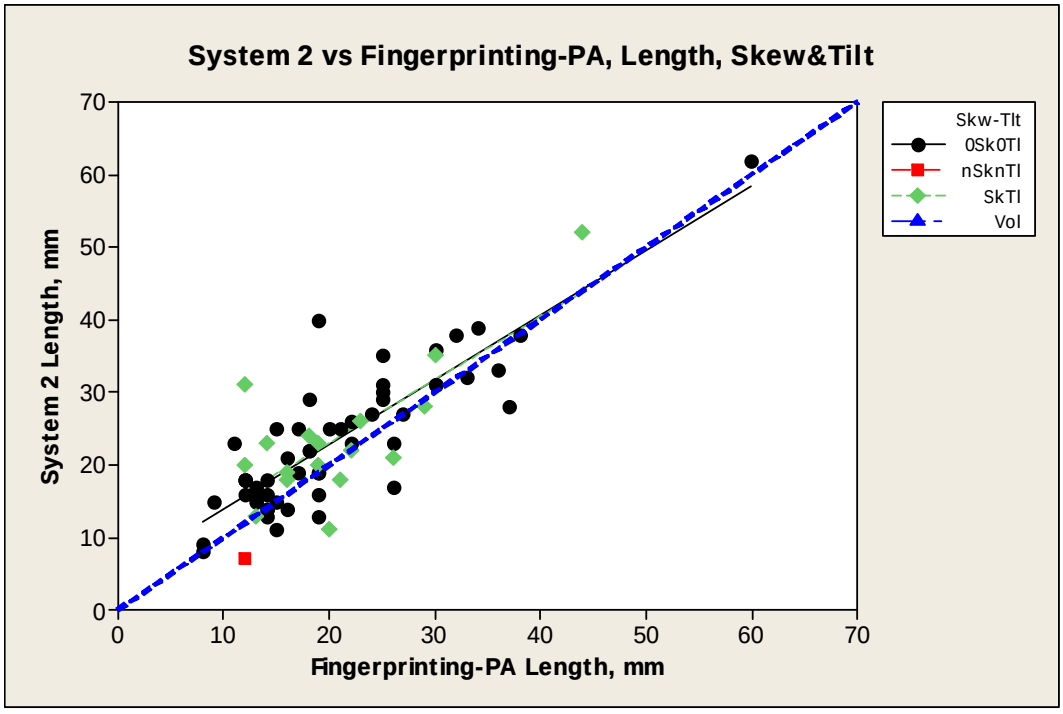


Figure 80. St1 System 2 vs Fingerprinting-PA for Flaw Length - Flaw-Type Effect

More plots of other AUT systems length measurements versus fingerprinting are shown in Appendix E.

The pooled data estimates such as average error and measurement uncertainty, and other statistical information such as data normality and comparison of weld sample distributions along with the number of outliers (OTL) is presented in Table 20.

Table 20. St1 Summary of Length Measurements (Reference) for Fingerprinting

Group	Sample	Normal Distribution (Y/N)	OTL[(a)]	s(ε) (mm)	Aver ε (mm)	Different Variances? (Y/N)	Different Distributions?	
							Test 1, Y/N	Test 2, Y/N
V1vF-Len-PA	97	N	6	5.89	-1.71	N	--	N
V2vF-Len-PA	68	Y	4	4.81	2.13	N	Y(W2/3)	--
V3vF-Len-PA	62	N	3	6.24	-3.19	Y	--	Y
V4vF-Len-PA	95	N	4	6.76	0.63	N	-- (W3/2 and 6)	Y
V5vF-Len-PA	72	N	3	4.64	0.56	N	-- (W2/5)	Y
V1vF-Len-XRM	43	N	4	5.99	1.91	N	--	N
V2vF-Len-XRM	35	Y	1	3.65	5.38	N	N	--
V3vF-Len-XRM	29	N	2	6.07	-1.1	N	--	N
V4vF-Len-XRM	42	Y	4	3.41	3.82	N	N	--
V5vF-Len-XRM	36	Y	0	3.59	3.91	N	Y (W5/4 and 6)	--

(a) OTL – Outliers

There are several observations that can be made from the summary data in Table 20:

- The V1vF-Len-PA data has many outliers (6).
- More than half of the joint distributions (6 out of 10) are not normal.
- A difference between the weld specimens (PA reference) is observed indicating inconsistent length measurements by the AUT systems when testing different welds.
- For some of the systems, the undersizing increases as the length increases (Appendix E).
- The flaw skew and tilt does not have effect on the length sizing.

The depth measurements and estimates are acquired in conditions similar to the height measurements:

- The AUT depth measurements are compared to two reference techniques – fingerprinting and DT.
- The effect of skew and tilt on the measurements is analyzed.
- Simple analysis of plots with the AUT measurements versus the reference technique is performed to evaluate the difference between the four flaw categories.

The fingerprinting depth measurements versus DT are shown in Figure 81. The systematic error was small across the depth range. The flaw-type effect on depth measurement is not very strong (Figure 82).

The AUT system V4 depth measurements versus DT are presented in Figures 83 and 84. Similar to the measurements shown in Figure 81, the systematic error of V4 depth estimates (Figure 83) is small and does not change with the depth. Some trend different from the overall trend is observed for the skewed and tilted flaws (Figure 84).

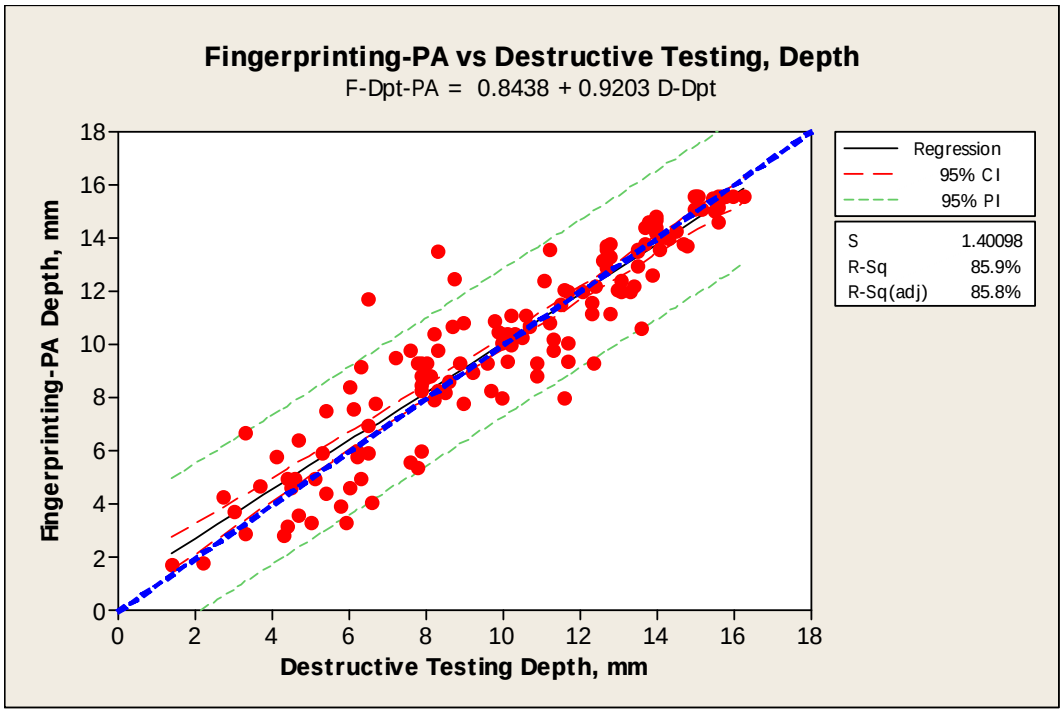


Figure 81. St1 Fingerprinting-PA vs DT for Flaw Depth

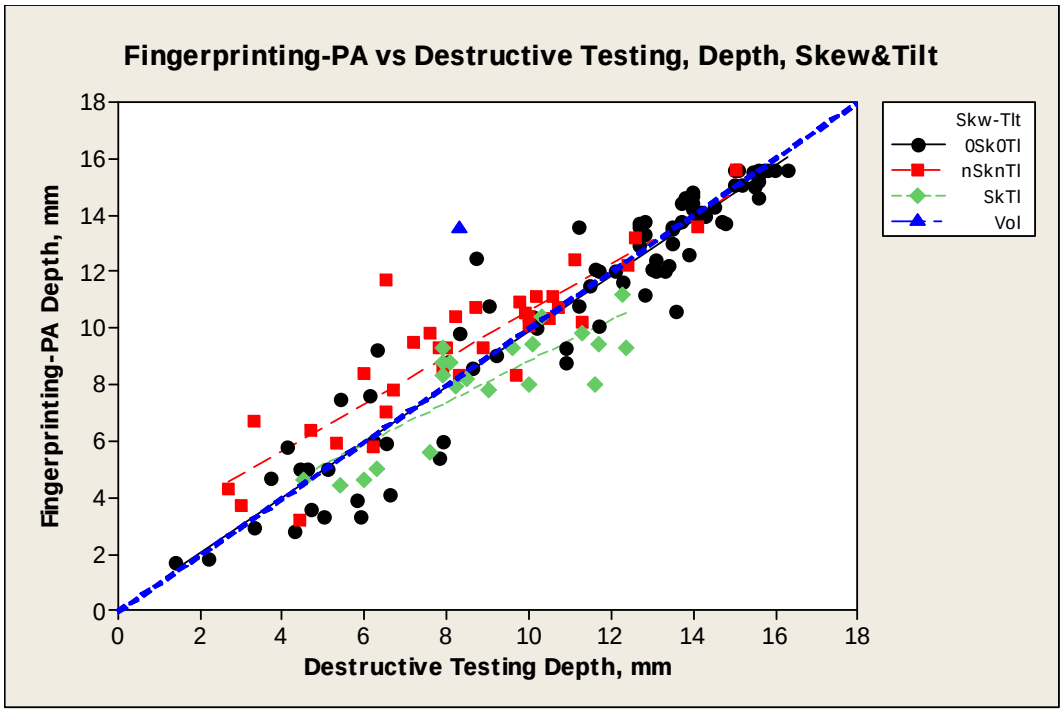


Figure 82. St1 Fingerprinting-PA vs DT for Flaw Depth - Flaw-Type Effect

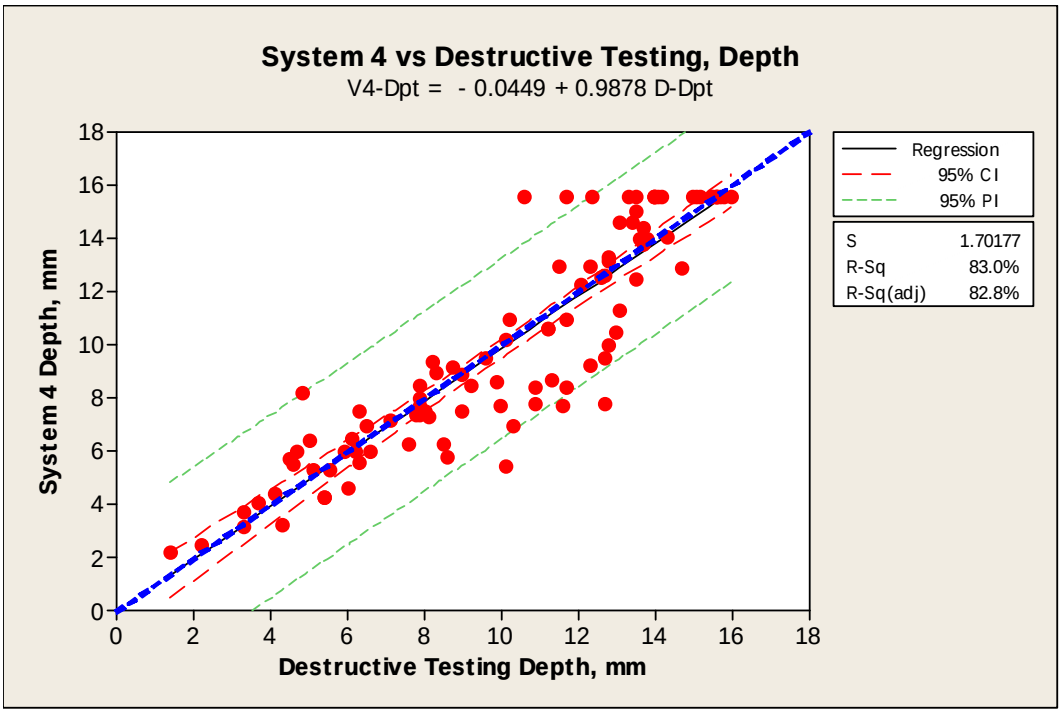


Figure 83. St1 System 4 vs DT, Depth

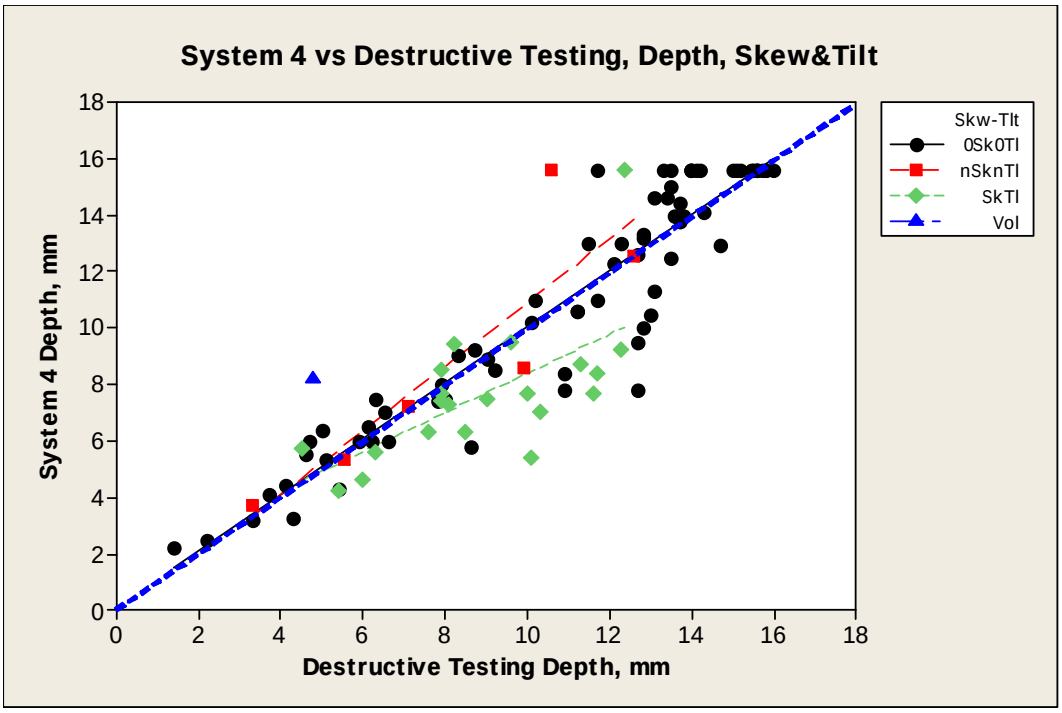


Figure 84. St1 System 4 vs DT, Depth, Flaw-Type Effect

Additional plots of fingerprinting with TOFD and other AUT systems depth measurements versus the DT are shown in Appendix E.

The pooled data estimates and other statistical information related to depth sizing capabilities are summarized and presented in Table 21.

Table 21. St1 Summary of Depth Measurements (Reference) for DT and Corrected Fingerprinting

Group	Sample	Normal Distribution (Y/N)	OTL[(a)]	s(ε) (mm)	Aver ε (mm)	Different Variances? (Y/N)	Different Distributions?	
							Test 1 (Y/N)	Test 2 (Y/N)
FvD-Dpt-PA	138	N	4	1.43	0.03	N	W3/5	Y
FvD-Dpt-TD	88	N	1	1.16	0.3	N	--	N
V1vD-Dpt	92	N	4	2.19	-0.78	N	--	N
V2vD-Dpt	71	N	0	1.86	-0.43	N	W2/3	Y
V3vD-Dpt	63	Y	0	2.49	-2.55	Y	--	N
V4vD-Dpt	101	N	6	1.71	-0.2	N	--	N
V5vD-Dpt	75	N	1	1.59	-0.15	N	--	N
V1vF-Dpt-PA	99	N	4	1.84	-0.61	N	--	N
V2vF-Dpt-PA	72	Y	1	1.97	-0.33	Y	W2/3	Y
V3vF-Dpt-PA	65	Y	1	2.44	-2.43	Y	--	Y
V4vF-Dpt-PA	101	N	4	1.84	-0.13	N	--	N
V5vF-Dpt-PA	78	Y	0	1.41	-0.2	N	N	-
V1vF-Dpt-TD	74	N	3	2.12	-0.7	N	--	Y
V2vF-Dpt-TD	60	N	3	1.79	-0.11	N	--	Y
V3vF-Dpt-TD	58	Y	1	2.28	-2.55	N	N	--
V4vF-Dpt-TD	77	Y	3	1.58	-0.31	N	N	--
V5vF-Dpt-TD	60	N	0	1.46	-0.29	N	--	N

(a) OTL – Outliers

The following observations can be made from the summary data in Table 21:

- The systematic depth undersizing for all systems is relatively small (except for V3) as opposed to the small oversizing during the fingerprinting.
- There are many (11 out of 17) distributions that are not normal.
- The system V3 has the largest uncertainty and systematic error. The procedure for the depth measurements used by V3 was not consistent with the other AUT procedures as previously discussed.
- With the exception of V3, the error distribution around the ideal sizing line is relatively uniform.
- Some of the effect of the flaw skew and tilt on depth sizing are similar to the effect on the height sizing discussed earlier.

Several conclusions can be drawn from this initial stage of data processing:

- Many not normal distributions (almost half in some cases) are observed.
- Many outliers are identified especially in start-stop error data.
- In many cases, the data from the separate welds indicates statistically significant differences caused by inconsistent fabrication, scanning, and others. This might partially

be explained with the small data samples acquired for each separate weld that would cause larger variability in the average error and uncertainty estimates.
- The change of WT for W2 and W3 does not have statistically significant effect on the sizing accuracy.

4.3.4.2 Stage 2

At stage 2 (St2), the analysis was performed under the following conditions:

- All pores were removed because only one pore (W3, S46) with height of 0.7 mm out of 74 was detected by the fingerprinting PA.
- The flaw category types were updated to reflect the findings at St1. The data shown in Table 12 and Figures 60 through 63 discussed earlier was sorted using these updated five categories. An explanation of the updated flaw categories is provided in the previous paragraph44.3.4.1 MEASUREMENT-SAMPLE-GROUP DESIGNATION and Appendix E.
- The measurements from all weld specimens were combined.
- The data with the DT reference only was analyzed.

The fingerprinting height-sizing estimates compared to the DT are shown in Table 22 using the redefined flaw categories. Most of the distributions are normal except for 0Sk0TI and the joint distribution of all flaw categories. As far as the joint distributions are concerned, the TOFD fingerprinting seems to perform slightly better (smaller uncertainty) than the PA, however, the TOFD sample size is ~40% smaller than the PA sample. The Test 1 also indicated statistically significant difference between the flaw categories. The box plots shown in Figure 85 and Appendix E illustrate the flaw category differences. While the IB flaws and in most cases Natural flaws are different from the other groups for the FvD-Hgt-PA (Figure 85), the Natural group is different in most cases for the FvD-Hgt-TD (Appendix E).

Table 22. St2 Fingerprinting vs Destructive – Height –Flaw-Type Effect

Group	Skew Tilt	Sample	Normal Distrib. (Y/N)	OTL[(a)]	s(ε) (mm)	Aver ε (mm)	Different Variances ? (Y/N)	Different Distributions?	
								Test 1 (Y/N)	Test 2 (Y/N)
FvD-Hgt-PA	0Sk0Tl	59	N	4	1.11	-0.1	--		--
	0SkTl	28	Y	0	0.94	-0.42	--		--
	IB	8	Y	0	1.49	2.80	--		--
	SkTl	8	Y	0	0.86	0.15	--		--
	Natural	34	Y	1	0.66	0.94	--		--
	Joint	137	N	5	1.27	0.28	N	Y	--
FvD-Hgt-TD	0Sk0Tl	38	N	2	0.97	0.13	--	--	--
	0SkTl	21	Y	0	0.79	0.03	--	--	--
	IB	8	Y	0	0.61	0.39	--	--	--
	SkTl	7	Y	0	0.50	-0.53	--	--	--
	Natural	10	Y	0	0.73	1.18	--	--	--
	Joint	84	N	1	0.92	0.20	N	Y	--

(a) OTL – Outliers

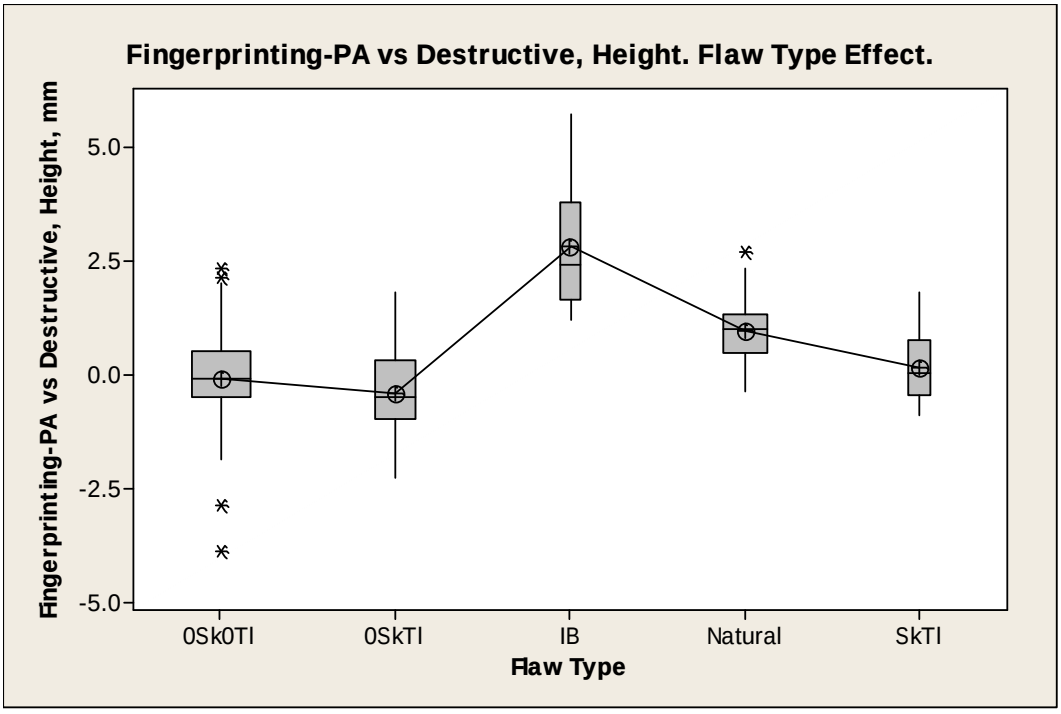


Figure 85. St2 One-Way ANOVA: FvD-Hgt-PA - Flaw-Type Effect

More plots and tables for the height, length and depth errors illustrating the flaw-type effect for each separate system are shown in Appendix E.

The height-sizing estimates for the joint (all welds) distributions of all five systems and fingerprinting are summarized in Table 23. Five out of the seven distributions are not normal. A systematic height undersizing is observed for all AUT systems as opposed to small oversizing during the fingerprinting. The IB flaws are statistically different from the other types for all systems and fingerprinting PA (Figure 85). Additional more conservative non-parametric Test 2 is conducted where distributions are not normal and or the variances are unequal. The Test 2 confirmed the results from the Test 1. With the exception of TOFD fingerprinting technique, the IB has a statistically significant effect on the height measurements.

Table 23. St2 Summary of Height Measurements (Reference) for DT

Group	Sample	Normal Distribution (Y/N)	OTL[(a)]	s(ε) (mm)	Aver ε (mm)	Different Variances (Y/N)	Different Distributions?	
							Test 1 (Y/N)	Test 2 (Y/N)
FvD-Hgt-PA	137	N	5	1.27	0.28	N	Y (IB, Natural)	Y
FvD-Hgt-TD	84	N	1	0.92	0.20	N	Y (Natural)	Y
V1vD-Hgt	92	Y	3	2.14	-0.90	N	Y (IB)	
V2vD-Hgt	72	N	4	1.97	-0.60	N	Y (IB)	Y
V3vD-Hgt	63	Y	1	1.61	-0.92	N	Y (IB)	
V4vD-Hgt	98	N	1	1.37	-1.12	Y	Y (IB)	Y
V5vD-Hgt	75	N	0	1.27	-0.82	N	Y (IB, Natural)	Y

(a) OTL – Outliers

The length sizing estimates for the joint distributions of all five systems are summarized in Table 24. With the exception of V2vF-Len-XRM, no statistically significant effect of flaw types on sizing uncertainty is observed. The PA fingerprinting detected almost two times more flaws than the other fingerprinting methods (X-ray and MPI). This results in a larger (more representative) sample size for the PA technique. Except for V3vF-Len-XRM, the PA fingerprinting demonstrates smaller systematic error. The PA fingerprinting reference technique and data were used to generate the final length sizing accuracy estimates.

Table 24. St2. Summary of Length Measurements (Reference) for Fingerprinting

Group	Sample	Normal Distribution (Y/N)	OTL	s(ε) (mm)	Aver ε (mm)	Different Variances? (Y/N)	Different Distributions?	
							Test 1, Y/N	Test 2, Y/N
V1vF-Len-PA	92	N	5	6.01	-1.52	N	N	Y (4.7%)
V2vF-Len-PA	72	N	4	4.70	2.21	N	N	N
V3vF-Len-PA	62	N	2	5.82	-2.79	N	N	N
V4vF-Len-PA	94	Y	3	5.48	0.36	N	N	
V5vF-Len-PA	74	N	2	4.60	0.78	N	N	N
V1vF-Len-XRM	43	N	4	5.99	1.91	N	N	
V2vF-Len-XRM	35	Y	1	3.65	5.38	N	Y (0SkTI)	
V3vF-Len-XRM	29	N	2	6.07	-1.10	N	N	
V4vF-Len-XRM	42	Y	4	3.41	3.82	N	N	
V5vF-Len-XRM	36	Y	0	3.59	3.91	Y	N	

OTL – Outliers

Table 25 summarizes the depth sizing estimates for the joint distributions for the fingerprinting and all five systems. Six out of the seven distributions are not normal. The IB, Natural and SkTI flaw types are different from the other types for some systems and the fingerprinting. The AUT system V3 has larger systematic error and uncertainty compared to other AUT systems and fingerprinting.

The final summary of the height systematic error and uncertainty is shown in Table 26. All the distributions are now normal and *95% LUS* is also provided in Table 26. As previously discussed, some single outliers and flaw groups (different distributions) are removed from the sample because the sizing estimates and distributions will be affected.

A list of single-point outliers with the highest impact that are removed from the height error distributions to achieve normality is presented in Table 27. It is remarkable that all of the single-point outliers are from specimens W5 and W6. These two specimens were added later to the program to investigate flaw interaction. This is an indication that interacting flaws might affect the height-sizing accuracy and have to be avoided in quantification trials if not specifically targeted. Further, the IB and Natural flaw types are removed from the distributions for all systems and the fingerprinting. One additional flaw-type group is removed from the V5vD-Hgt data sample. As already discussed, the removed flaw-type groups are statistically different from the other flaw types.

Table 25. St2 Summary of Depth Measurements (Reference) for DT

Group	Sample	Normal Distribution Y/N	OTL[(a)]	s(ε) (mm)	Aver ε (mm)	Different Variances? Y/N	Different Distributions?	
							Test 1 (Y/N)	Test 2 (Y/N)
FvD-Dpt-PA	137	N	4	1.36	-0.01	N	Y (IB, Natural)	Y
FvD-Dpt-TD	88	N	1	1.16	0.30	N	Y (Natural)	Y
V1vD-Dpt	92	N	3	2.20	-0.82	N	Y (IB)	Y
V2vD-Dpt	72	N	0	1.85	-0.41	N	Y (SkTI)	Y
V3vD-Dpt	63	Y	0	2.48	-2.56	N	Y (SkTI)	
V4vD-Dpt	99	N	5	1.66	-0.21	Y	Y (SkTI)	Y
V5vD-Dpt	75	N	1	1.59	-0.15	Y	Y (IB)	Y

(a) OTL – Outliers

Table 26. Final St2 Summary of Height Measurements (Reference) for DT – 95% LUS and Joint

Group	Sample	s(ε) (mm)	Aver ε (mm)	Coverage Factor (k) (df = n-1)	95% LUS = k.s(ε) (Aver ε) (mm)
FvD-Hgt-PA	95	1.05	-0.17	1.66	1.91
FvD-Hgt-TD	64	0.76	-0.05	1.67	1.32
V1vD-Hgt	82	1.95	-1.22	1.66	4.46
V2vD-Hgt	63	1.66	-0.94	1.67	3.71
V3vD-Hgt	55	1.53	-1.17	1.67	3.73
V4vD-Hgt	84	1.40	-1.27	1.66	3.59
V5vD-Hgt	44	1.04	-0.75	1.68	2.50
Joint V1 to V5	324	1.48	-1.04	1.645	3.47

Table 27. Final St2 Outliers Removed for Data Normality – Height

Group	Error (mm)	Weld	Sector	Type (Category)	D-Dpt (mm)	D-Hgt (mm)	System
FvD-Hgt-TD	3.1	W5	S14	0Sk0TI	4.7	3	NA
	2.3	W5	S14	0Sk0TI	6.6	3.8	NA
Joint V1 to V5	-6.7	W5	S53	0Sk0TI	15.7	14.1	V1
	-7.2	W6	S35	0SkTI	12.7	8.3	V1
	-6.2	W5	S37	0SkTI	14	8.2	V2
	-5.1	W6	S35	0SkTI	12.7	8.3	V4

Additional data tables are available in Appendix E illustrating the effect of flaw-type group removal on the height-sizing estimates.

A final comparison between the fingerprinting and the other systems performance in terms of height measurement is illustrated with the box plot in Figure 86. The fingerprinting (FvD-Hgt-PA and FvD-Hgt-TD) and one of the AUT systems (V5vD-Hgt) do not have statistically significant differences. All five AUT systems are not statistically different between themselves. However,

four of the AUT systems are different from the fingerprinting techniques. An additional box plot for the AUT systems only is available in Appendix E. The fingerprinting techniques have almost negligible systematic error and uncertainty that is smaller (except for V5vD-Hgt vs. FvD-Hgt-PA) but still in the same order of magnitude to the AUT systems uncertainty. All other systems systematically undersized the height by 1.04 mm with uncertainty of 1.48 mm (joint distribution) as shown in Figure 86 and Table 26.

A set of height-sizing plots with *95% LUS* for the five AUT systems, fingerprinting with PA technique and joint V1 through V5 data is shown in Figures 87 through 93. For each AUT system, the different flaw-type groups are shown with different data markers on the plots.

A set of POD curves for the PA fingerprinting technique, five AUT systems and joint V1, V4 and V5 data is shown in Figures 94 through 101. It is important to note that only pores are removed for the POD estimates. As already discussed, the removal of small flaws from the sample (mostly Natural in addition to Vol) that are usually missed during the testing would cause problems with the POD algorithms convergence. All plots shown in Figures 94 through 101 use the logit POD function without log transformation of the flaw height. Two plots are provided for the V4 AUT system – one for the blind (as other systems) and another for the open trials. The "open" trials for the V4 were not physically conducted. The equipment acquisition data files were reanalyzed at EWI for any flaw indications that could be associated with a known flaw regardless of the amplitude threshold setup. Fifteen flaws were added to the group of the detected flaws for the V4 open trials.

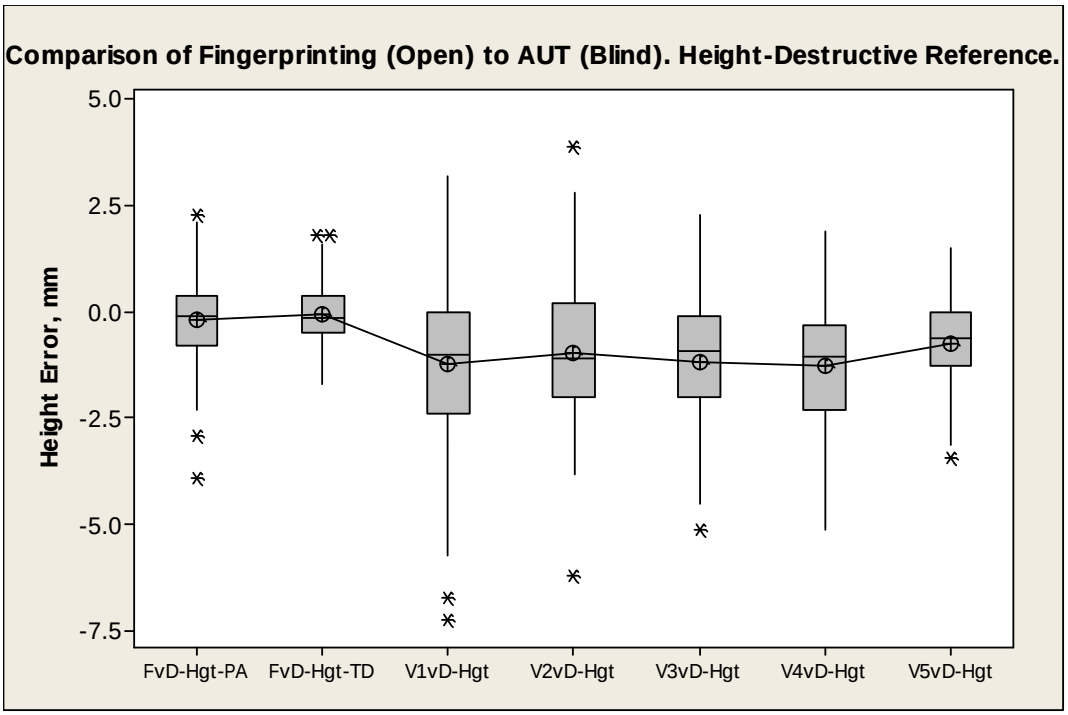


Figure 86. St2 One-Way ANOVA: Final Data, Fingerprinting and Other Systems, Height, DT Reference

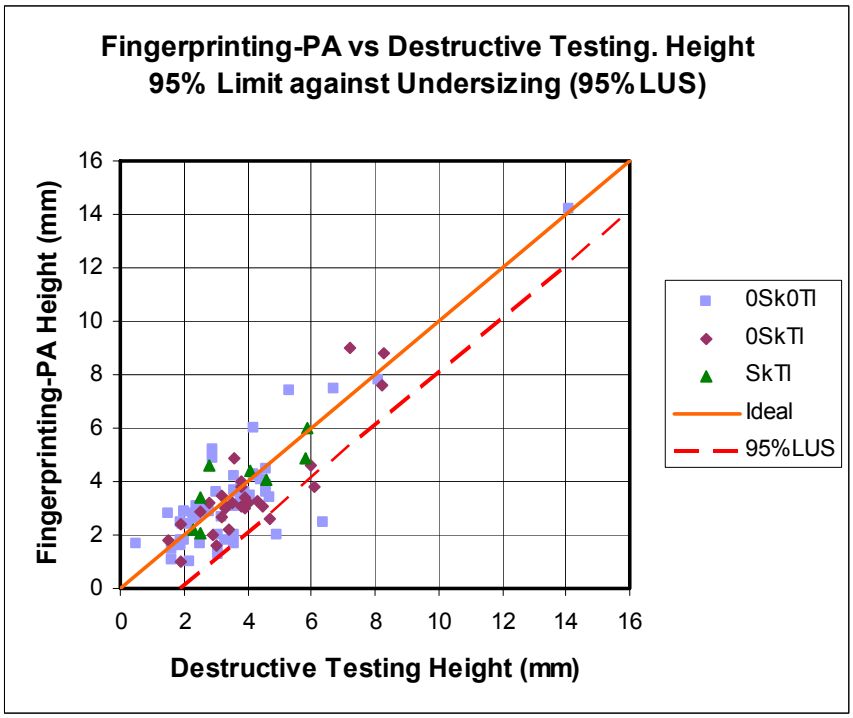


Figure 87. St2 Fingerprinting-PA vs DT, Height, 95% LUS

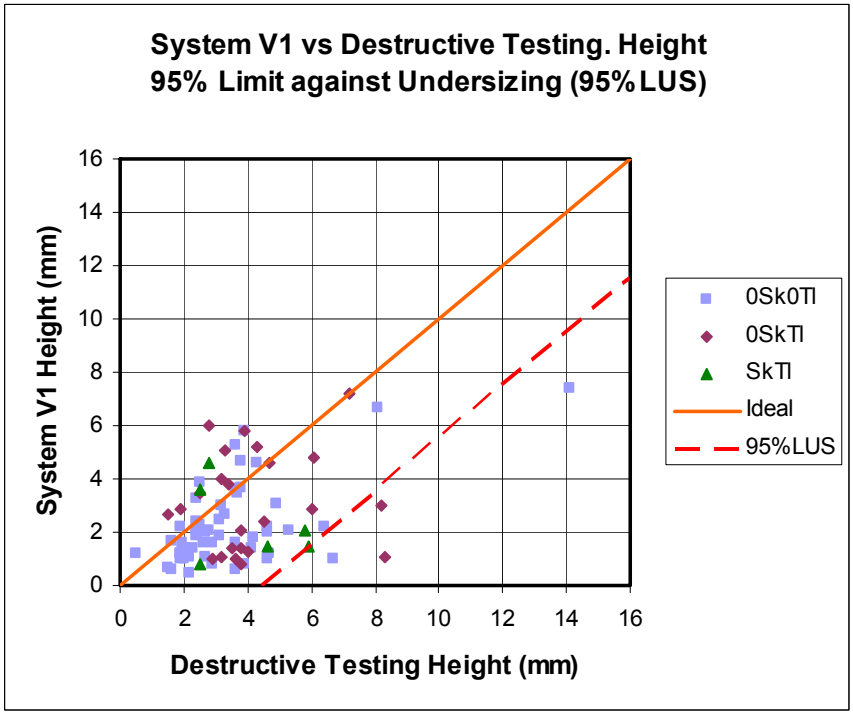


Figure 88. St2 System V1 vs DT, Height, 95% LUS

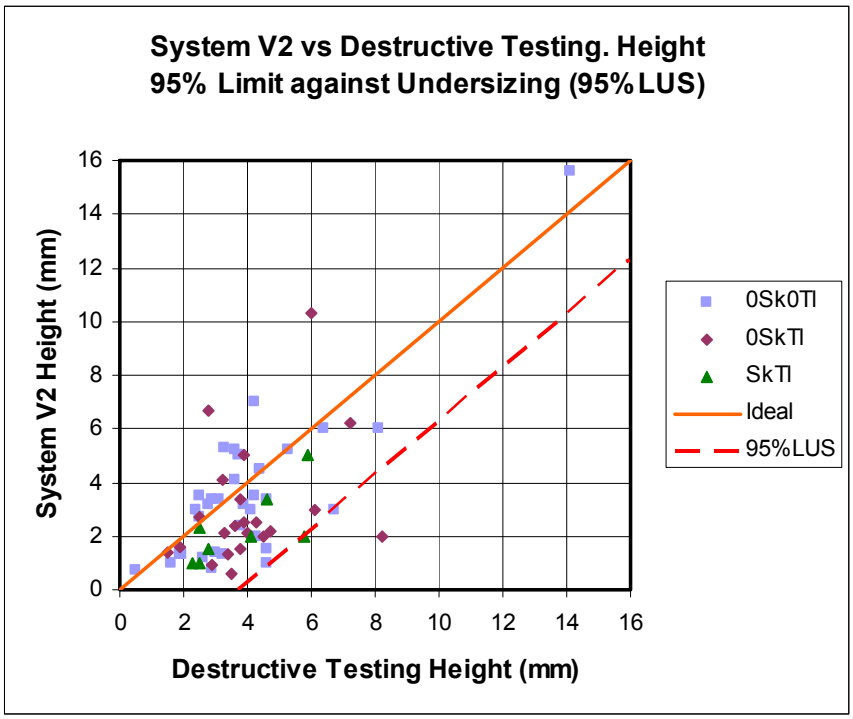


Figure 89. St2 System V2 vs DT, Height, 95% LUS

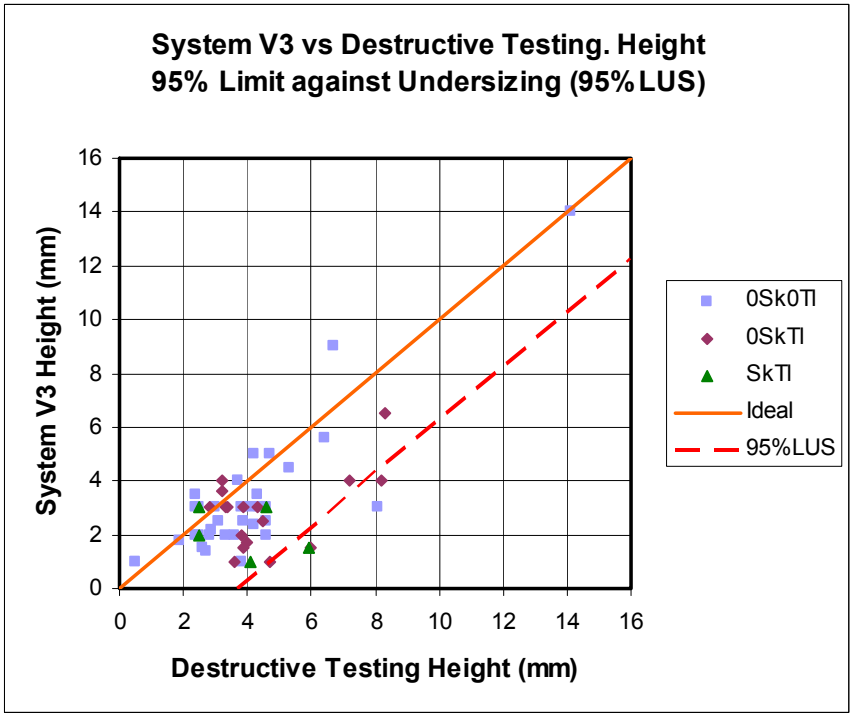


Figure 90. St2 System V3 vs DT, Height, 95% LUS

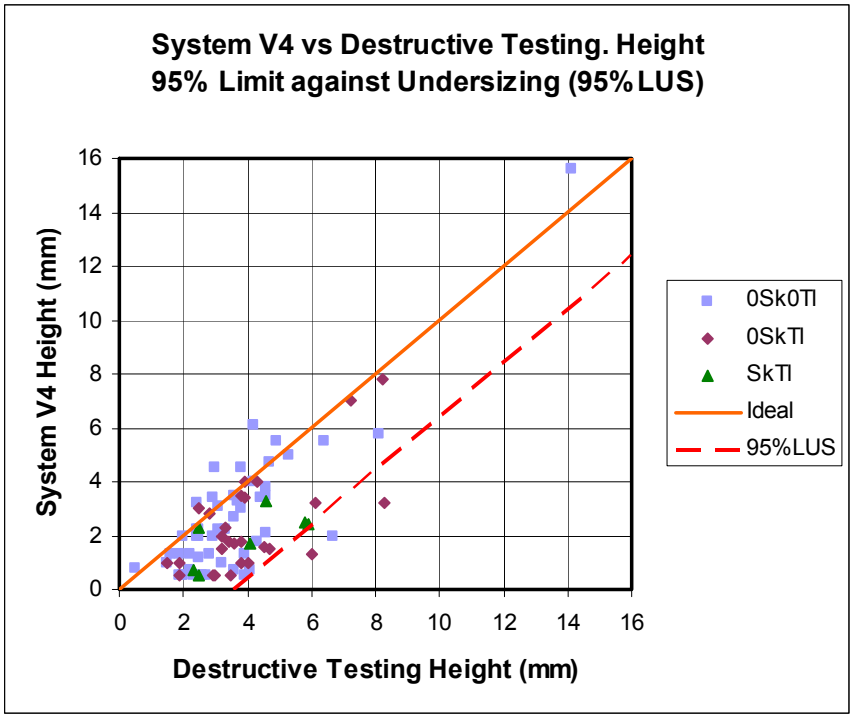


Figure 91. St2 System V4 vs DT, Height, 95% LUS

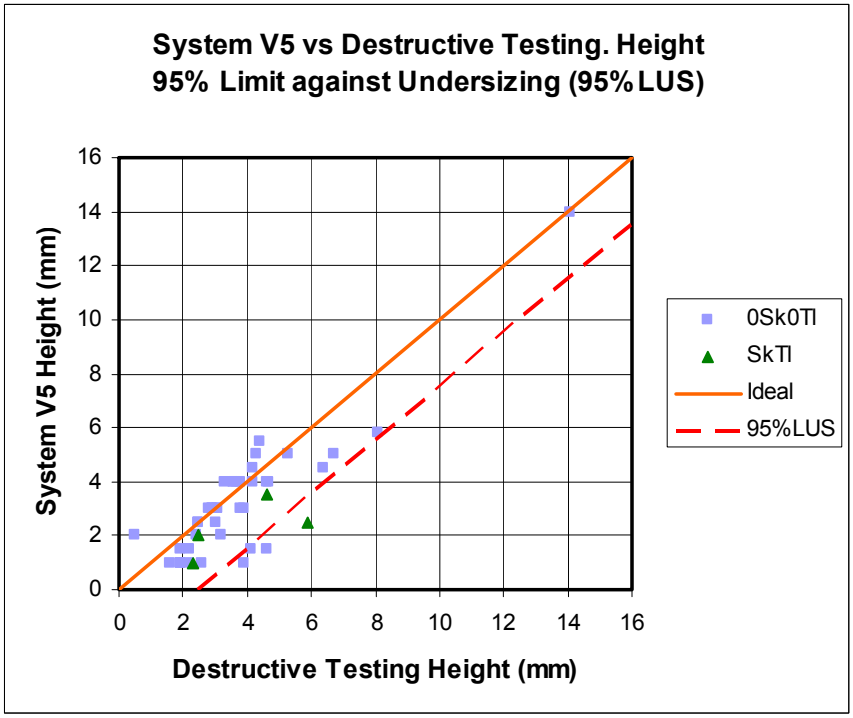


Figure 92. St2 System V5 vs DT, Height, 95% LUS

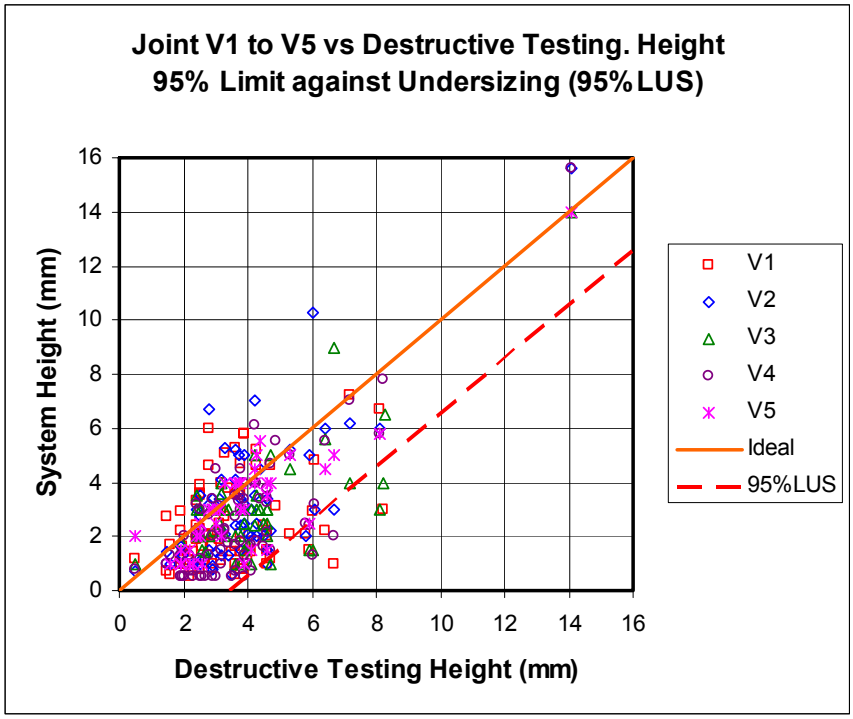


Figure 93. St2 Joint vs DT, Height, 95% LUS

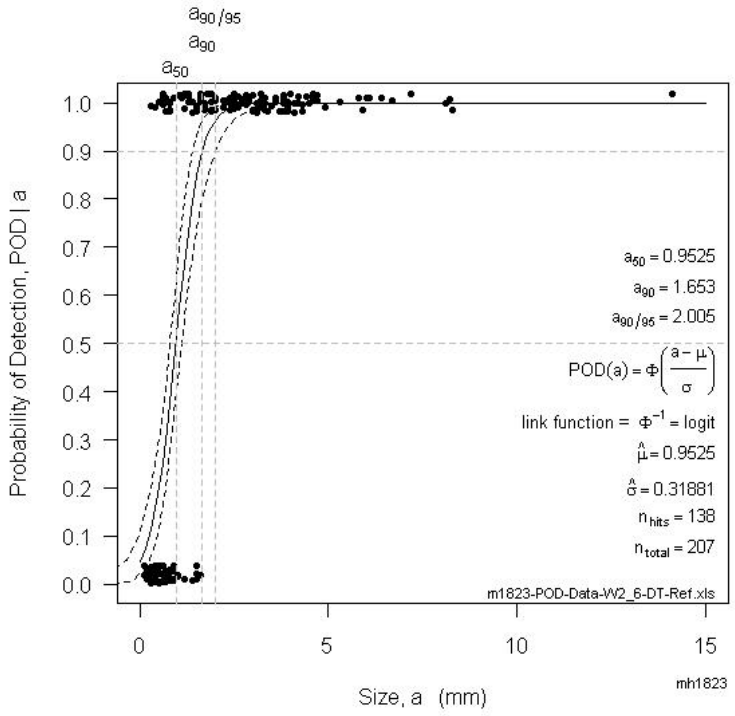


Figure 94. St2 POD, Fingerprinting-PA vs DT, Height, Logit

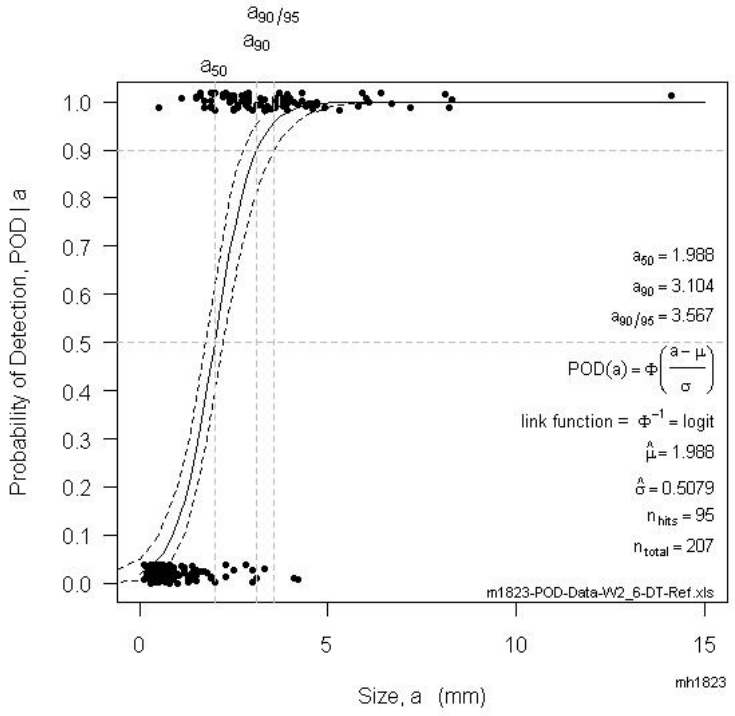


Figure 95. St2 POD, System V1 vs DT, Height, Logit

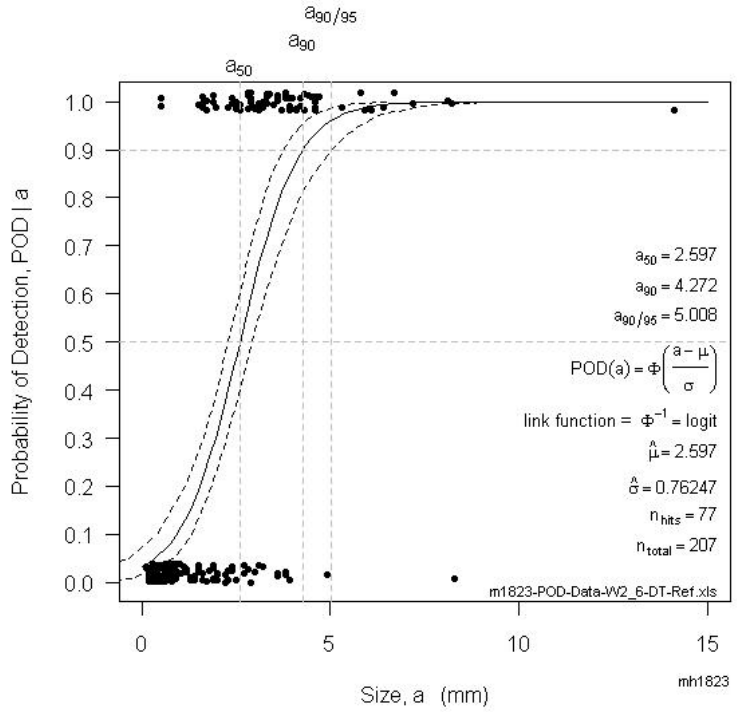


Figure 96. St2 POD, System V2 vs DT, Height, Logit

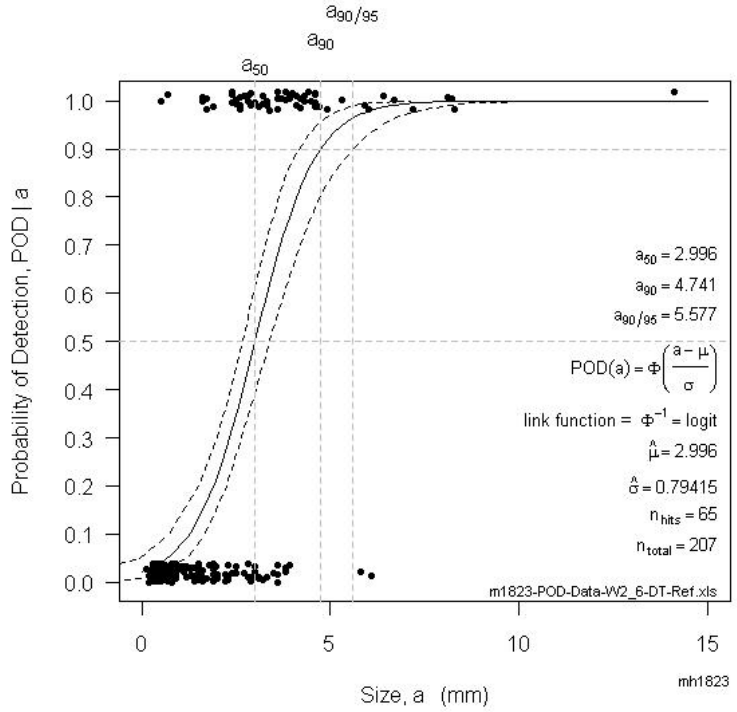


Figure 97. St2 POD, System V3 vs DT, Height, Logit

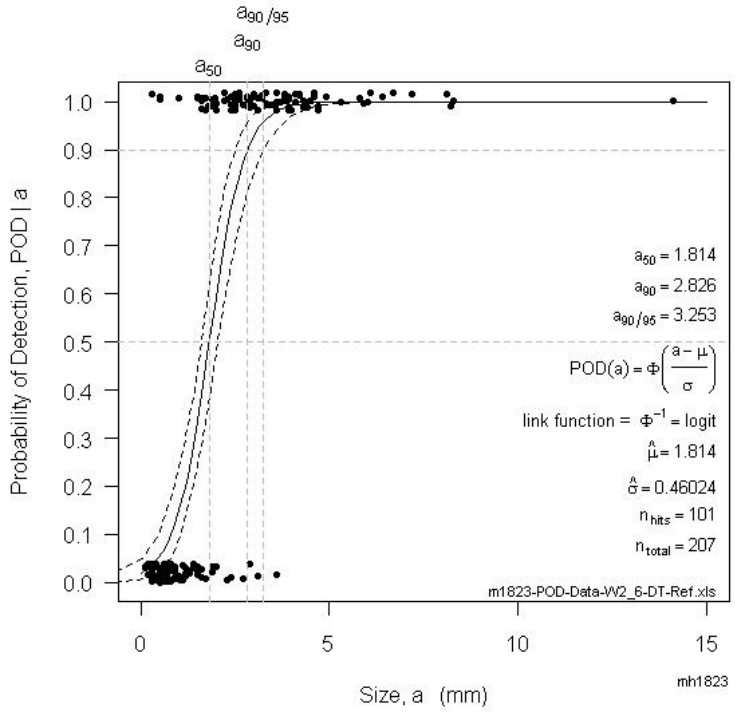


Figure 98. St2 POD, System V4 vs DT, Height, Logit, Blind

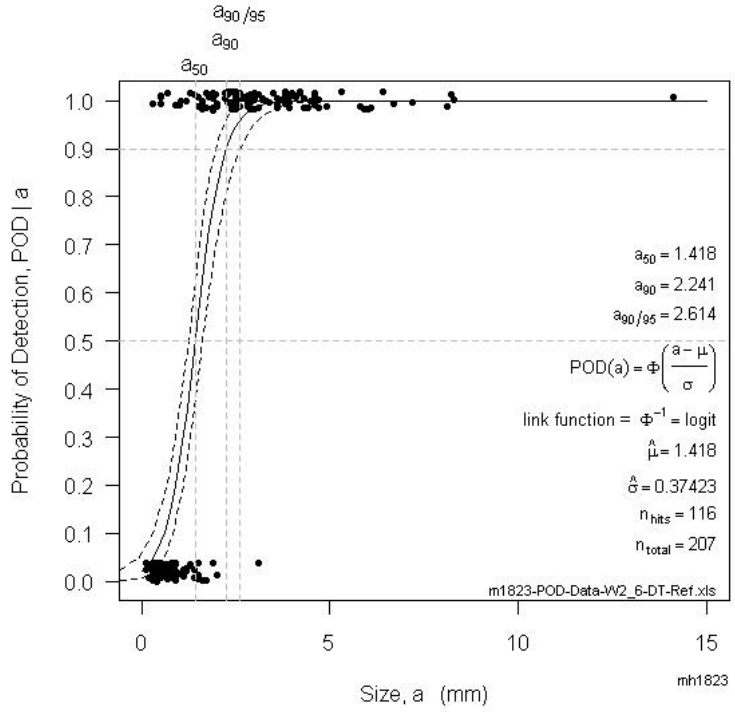


Figure 99. St2 POD, System V4 vs DT, Height, Logit, Open

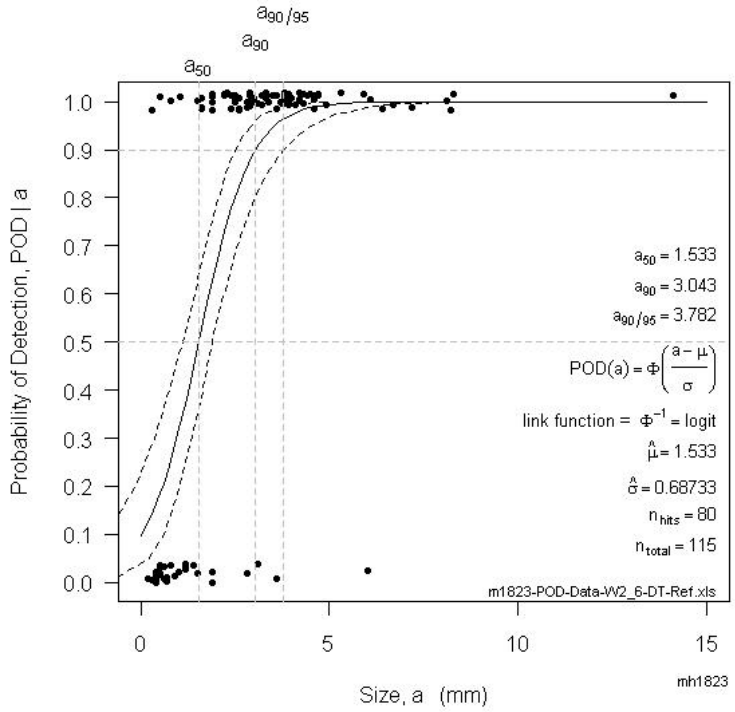


Figure 100. St2 POD, System V5 vs DT, Height, Logit

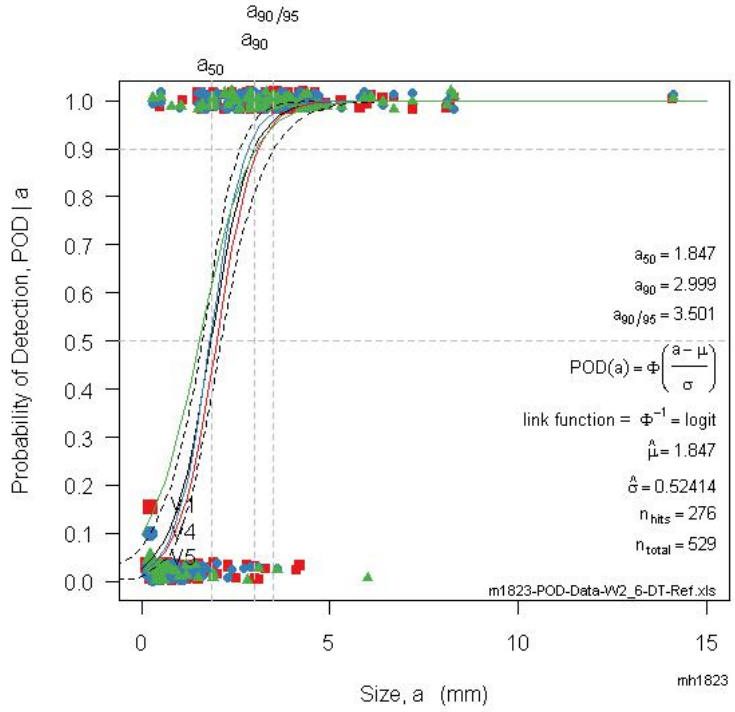


Figure 101. St2 POD, Systems V1, V4 (Blind) and V5 vs DT, Height, Logit

Quick comparison of the AUT system POD performance in terms of $a_{90/95}$ reveals that systems V2 and V3 have significantly larger $a_{90/95}$ than the other three systems (Table 28). The drop in the performance for V3 can be explained with the smaller degree of automation and examiner experience as discussed earlier. The drop in the performance for V2 may not be easily explained. The V2 system equipment was similar to the equipment used by several of the other systems and the examiner was experienced. Further study of the reasons for this discrepancy is outside of the scope of this project. This difference in the POD performance is used as a reason to ignore the V2 and V3 data when the average POD estimate (Figure 101) for the joint data is obtained. The $a_{90/95}$ for the AUT systems is in the range from 3.3 mm to 3.8 mm with average of 3.5 mm for the joint sample (Table 28).

As expected, the V4 $a_{90/95}$ of 2.6 mm for the open trials is better (~ 20%) compared to 3.3 mm for the blind trials (Table 28).

The results from another study investigating the effect of different POD functions on the POD estimates are shown in Table 29 for the fingerprinting open trials. Five different algorithms found in the three off-the-shelf POD software packages are used to produce the estimates with the same data sample. The $a_{90/95}$ is in the range from 2 to 2.3 mm for the various options. The difference of 0.3 mm is a relatively small (~15%) compared to the smaller estimate of 2 mm. The POD plots for the various POD algorithms and software packages are shown in Appendix E.

An approach similar to the approach shown in Table 29 is used to investigate the effect of different POD algorithms on the POD estimates for the V4 blind trials. The results are presented in Table 30. A somewhat larger difference of 1.3 mm (~40%) between the smallest (3.2 mm) and the largest (4.5 mm) estimate of the $a_{90/95}$ is observed compared to the difference in Table 29. The Option 1 (Opt 1), Opt 2 (MIL 1823) and Option 1 (PODv3) provide the smallest and similar POD estimates without log transformation of the height. The Opt 3 and 4 (MIL 1823), Opt 2 (PODv3) with log transformation and STATUS options provide larger and similar $a_{90/95}$ estimates in the range from 3.9 to 4.5 mm.

Some project contributors were interested in investigating how the performance of an AUT system would be affected if the examiner influence was reduced and all indications (not only reported) visible on the AUT scans were accounted. The open trials provide this opportunity as already discussed. A complete set of POD estimates with all available software packages and options is summarized in Table 31 for the V4 open trials. The observed difference of 0.7 mm (~28%) between the smallest (2.6 mm) and the largest (3.3 mm) estimate of the $a_{90/95}$ is smaller compared to the difference for the blind trials in Table 30. Similar to the blind trials, the Opt 1 and Opt 2 (MIL 1823) and PODv3 provide smaller and similar POD estimates. The Opt 3 and 4 (MIL 1823) with log transformation and STATUS options provide larger and similar $a_{90/95}$ estimates in the range from 3.1 to 3.3 mm.

Table 28. POD for AUT Systems - Blind and Open (V4 Only) Trials

Group	a_{50} (mm)	a_{90} (mm)	$a_{90/95}$ (mm)	Link	Transf	Deviance
V1vD	1.988	3.104	3.567	Logit	No	127
V2vD	2.597	4.272	5.008	Logit	No	160
V3vD	2.996	4.741	5.577	Logit	No	152
V4vD blind	1.814	2.826	3.253	Logit	No	120
V4vD open	1.418	2.241	2.614	Logit	No	110
V5vD	1.533	3.043	3.782	Logit	No	81
V1, V4bl, and V5	1.847	2.999	3.501	Logit	No	335

Table 29. POD for Fingerprinting-PA - Open Trials

Software Option	a_{50} (mm)	a_{90} (mm)	$a_{90/95}$ (mm)	Link	Transf
Opt 1 (MIL 1823)	0.9525	1.653	2.005	Logit	No
Opt 2 (MIL 1823)	0.9641	1.659	1.993	Probit	No
PODv3	0.815	1.707	2.057	Log normal	Yes
Nordtest (Status)	--	1.79	2.17	Special	No
Hit/Miss (Status)	--	1.71	2.34	Probit	

Table 30. POD for System V4 - Blind Trials

Software Option	a_{50} (mm)	a_{90} (mm)	$a_{90/95}$ (mm)	Link	Transf	Deviance
Opt 1 (MIL 1823)	1.814	2.826	3.253	Logit	No	120
Opt 2 (MIL 1823)	1.837	2.914	3.32	Probit	No	121
Opt 3 (MIL 1823)	1.583	3.175	4.048	Logit	Yes	127
Opt 4 (MIL 1823)	1.517	3.296	4.206	Probit	Yes	130
Opt 1 (PODv3)	1.837	2.914	3.212	Normal	No	
Opt 2 (PODv3)	1.517	3.296	3.947	Log normal	Yes	
Nordtest (Status)		3.38	4.04	Special		
Hit/Miss (Status)		3.3	4.52	Probit		

Table 31. POD for System V4 - Open Trials

Software Option	a_{50} (mm)	a_{90} (mm)	$a_{90/95}$ (mm)	Link	Transf	Deviance
Opt 1 (MIL 1823)	1.418	2.241	2.614	Logit	No	110
Opt 2 (MIL 1823)	1.434	2.31	2.67	Probit	No	111
Opt 3 (MIL 1823)	1.227	2.438	3.095	Logit	Yes	117
Opt 4 (MIL 1823)	1.186	2.476	3.121	Probit	Yes	119
PODv3	1.186	2.476	2.95	Log normal	Yes	
Nordtest (Status)	--	2.65	3.18	Special		
Hit/Miss (Status)	--	2.48	3.34	Probit		

The POD plots for the various options used to obtain the V4 blind and open trials POD estimates summarized in Tables 30 and 31 are shown in Appendix E.

The results from this comparison of various POD functions (Tables 29 through 31) support the approach recommended in the literature[30] and implemented in this study that one function (if possible) should be used when comparing the POD performance of various systems.

A short study was conducted to investigate whether any of the link functions would be capable of providing POD estimates when the Natural flaw category was removed in addition to pores (Vol). In this case, the POD sample will consist of the implanted flaws only. Few implanted flaws were missed by the various AUT systems and all of the implanted flaws were detected by the fingerprinting. The POD algorithms have convergence problems when only a few flaws are missed, as discussed, and they cannot be used if all of the flaws are detected. The special link function Eq. (20) only was capable of providing POD estimates for the AUT systems. The POD plots for the V4 implanted flaw sample are shown in Appendix E. The $a_{90/95}$ is 3 mm for the blind trials and 2.1 mm for the open trials. This is smaller but not drastically smaller than the best $a_{90/95}$ estimates of 3.2 and 2.6 mm (Tables 30 and 31), respectively, when the Natural flaws were accounted for.

It seems that the $a_{90/95}$ of approximately 2 mm (see Tables 29 and 31) is the best POD performance of the equipment and procedures (minimized examiner influence) that might be expected for this type of inspection.

For many field inspections, the accuracy of flaw length measurement is also critical. Similar to the height measurements, the flaw length undersizing should be estimated and accounted in the ECA process. A summary of the AUT systems performance in terms of systematic error, uncertainty and *95% LUS* is shown in Table 32. A list of the outliers removed to improve the distribution normality is shown in Table 33. Except for the pores (Vol), all other flaw-type groups were kept in the sample because they did not have statistically significant effect on the length sizing accuracy (Table 23). A comparison of the AUT systems is shown in Figure 102. The ANOVA test indicates that Systems V2 and V3 are different from the other systems. Estimates for the two joint distributions are presented in Table 32 with and without removal of V2 and V3 data. The systematic error, uncertainty and *95% LUS* are similar. The benefit of the V2 and V3 removal, however, is visible in Table 33. Only two outliers were removed from Joint V1, V4, and V5 distribution while nine outliers needed to be removed from the Joint V1 to V5 distribution to become normal.

Table 32. Final St2 Summary of Length Measurements (Reference) for Fingerprinting PA, All Welds (W2-W6), 95% LUS and Joint

Group	Sample	s(ε) (mm)	Aver ε (mm)	Coverage Factor (k) (df = n-1)	95% LUS = k.s(ε) (Aver ε) (mm)
V1vF-Len-PA	90	5.25	-1.08	1.66	9.80
V2vF-Len-PA	70	4.10	2.13	1.67	4.72
V3vF-Len-PA	61	5.31	-2.48	1.67	11.35
V4vF-Len-PA	94	5.48	0.36	1.66	8.74
V5vF-Len-PA	73	4.10	0.53	1.67	6.32
Joint V1, V4 and V5	255	4.89	-0.1	1.645	8.14
Joint V1 to V5	385	4.88	-0.18	1.645	8.21

Table 33. Final St2 Outliers Removed for Data Normality, Length

Group	Error (mm)	Weld	Sector	Type (Category)	Length (mm)	D-Dpt (mm)	D-Hgt (mm)	System
V1vF-Len-PA	-17.7	W6	S15	0Sk0Tl	14.3	10.2	3.1	NA
	-25	W6	S35	0Sk0Tl	3	7.9	12.7	NA
V2vF-Len-PA	19	W2	S26	0SkTl	31	79	3.9	NA
	-9	W3	S29	SkTl	11	6	2.8	NA
V3vF-Len-PA	-22	W6	S15	0Sk0Tl	10	10.2	3.1	NA
V5vF-Len-PA	19	W2	S26	0SkTl	31	7.9	3.9	NA
Joint V1, V4, V5	-16	W4	S43	SkTl	3	7.6	2.3	V4
	16	W2	S48	0Sk0Tl	34	15.6	1.9	V4
Joint V1 to V5	-17.7	W6	S15	0Sk0Tl	14.3	10.2	3.1	V1
	-25	W6	S35	0SkTl	3	7.9	12.7	V1
	19	W2	S26	0SkTl	31	7.9	3.9	V2
	16	W6	S13	0Sk0Tl	25	7.9	3.3	V3
	-22	W6	S15	0Sk0Tl	10	10.2	3.1	V3
	16	W2	S48	0Sk0Tl	34	15.6	1.9	V4
	17	W4	S32	0SkTl	32	11.5	3.8	V4
	-16	W4	S43	SkTl	3	7.6	2.3	V4
	19	W2	S26	0SkTl	31	7.9	3.9	V5

Table 34. Validation of Fingerprinting Length Sizing with DT

Weld	Sector	Axial Posit.	FA[(a)] (mm		F (mm)		Length (mm) (°F)	Length (mm) (DT)	Note
			Start	Stop	Start	Stop			
W5	S25	C[(b)]	966	975	969	976	7	9 ± 3	Good agreement
W5	S25	US	980	995	980	990	10	15 ± 3	2 mm smaller than smallest possible
W5	S26	DS	1000	1012	1000	1013	13	15 ± 3	Good agreement
W6	S13	DS	483	498	485	494	9	12 ± 3	Good agreement
W6	S32	US	1240	1244	1239	1245	6	9 ± 3	Good agreement
W6	S32	US	1248	1258	1248	1259	11	9 ± 3	Good agreement
W6	S32	US	1264	1276	1265	1277	12	12 ± 3	Good agreement

(a) FA – Fabrication specifications
(b) C – Center

A set of length sizing plots with *95% LUS* line for the five AUT systems and joint V1, V4, V5 data is shown in Figures 103 through 108. For each AUT system, the different flaw-type groups are shown with different data markers on the plots.

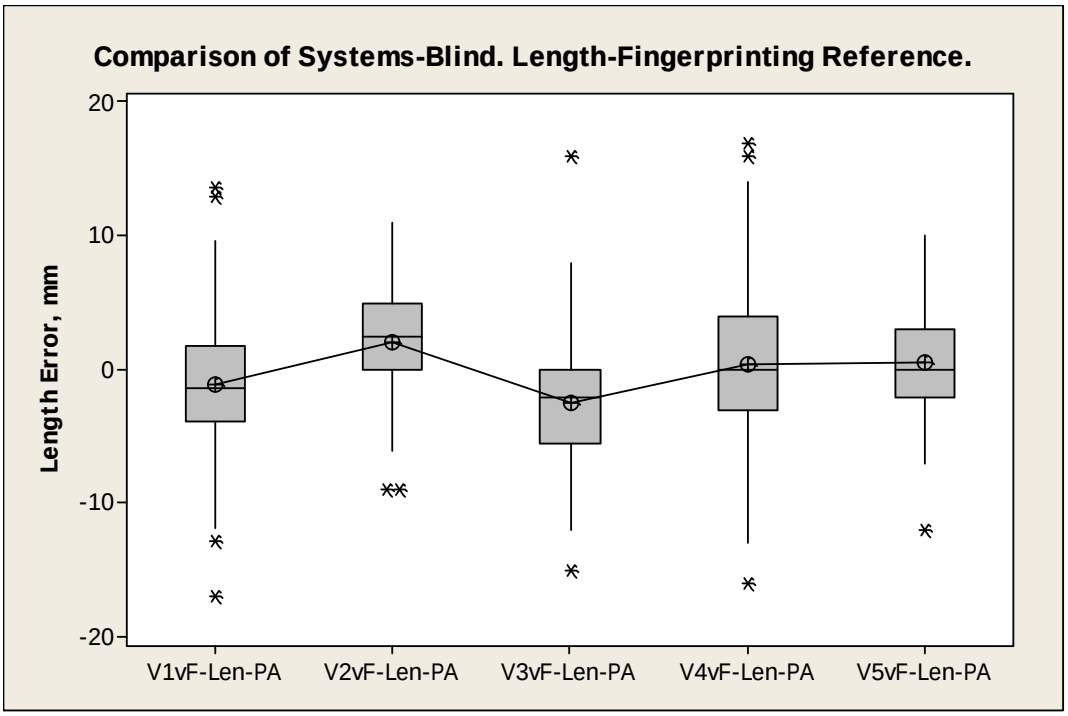


Figure 102. St2 One-Way ANOVA: Final Data, Comparison of Systems, Length, Fingerprinting PA Reference

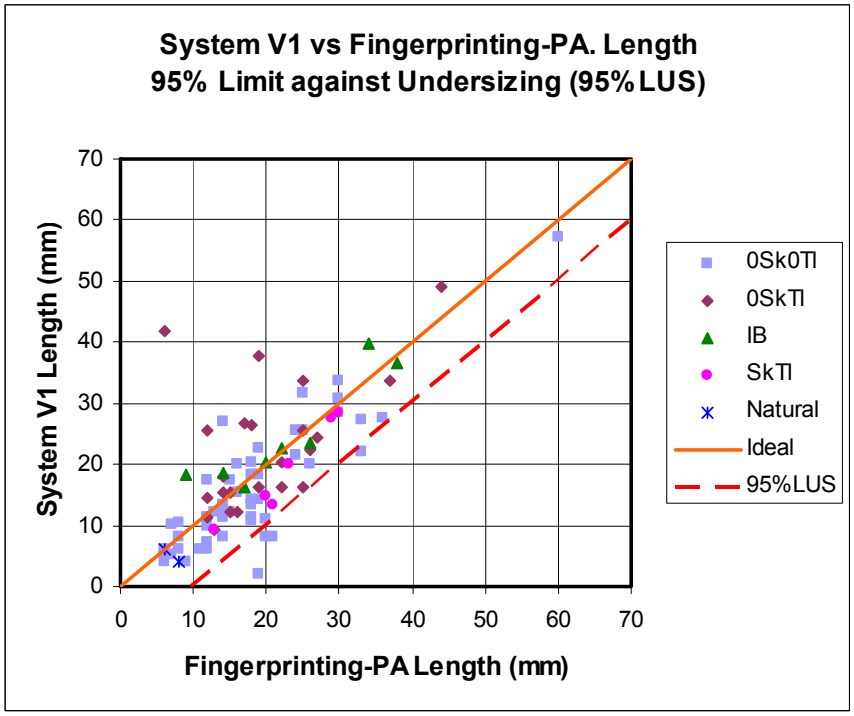


Figure 103. St2 System V1 vs Fingerprinting PA – Length – 95% LUS

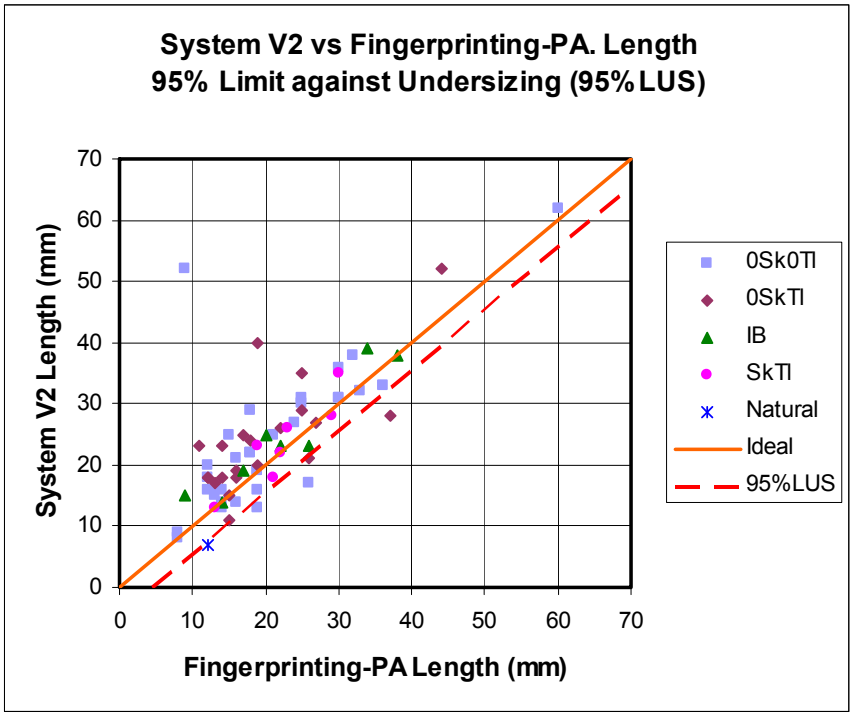


Figure 104. St2 System V2 vs Fingerprinting PA – Length – 95% LUS

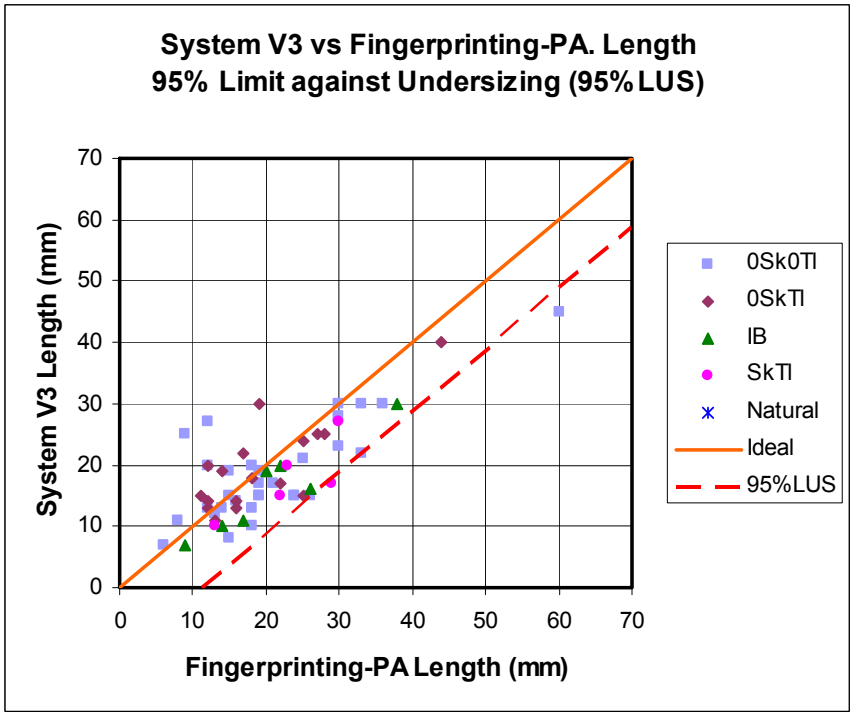


Figure 105. St2 System V3 vs Fingerprinting PA, Length, 95% LUS

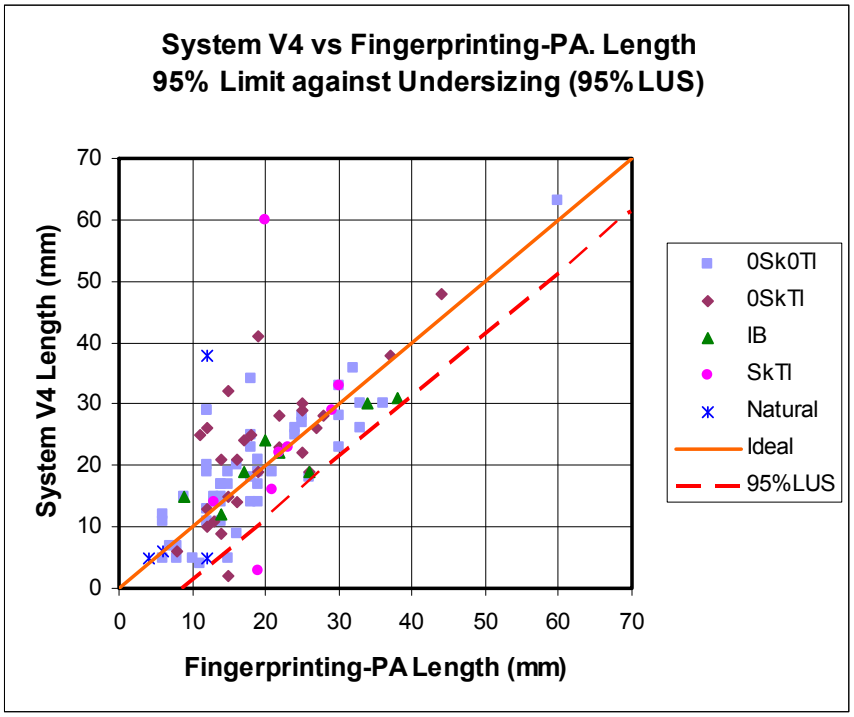


Figure 106. St2 System V4 vs Fingerprinting PA, Length, 95% LUS

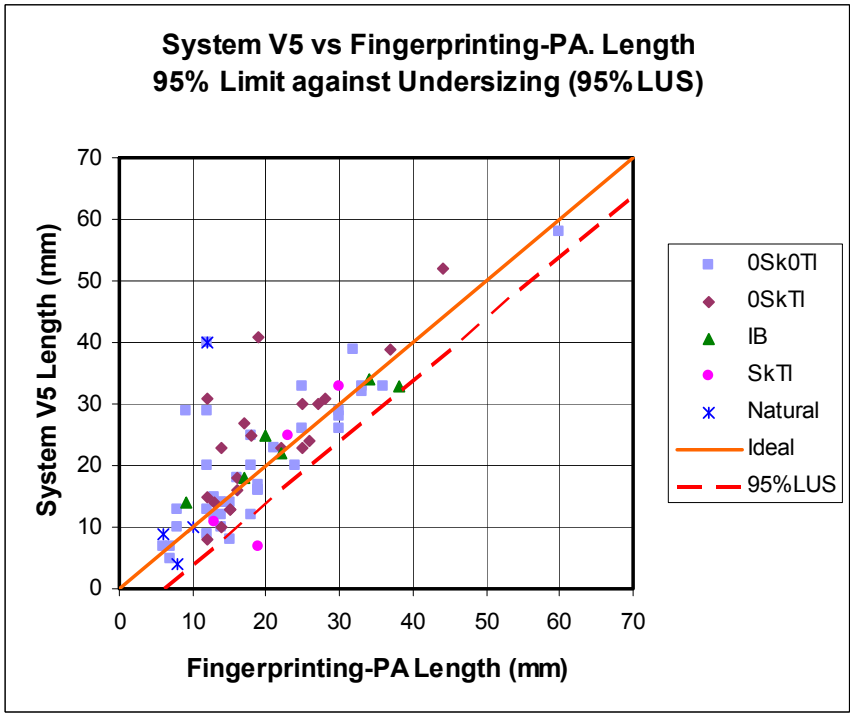


Figure 107. St2 System V5 vs Fingerprinting, Length, 95% LUS

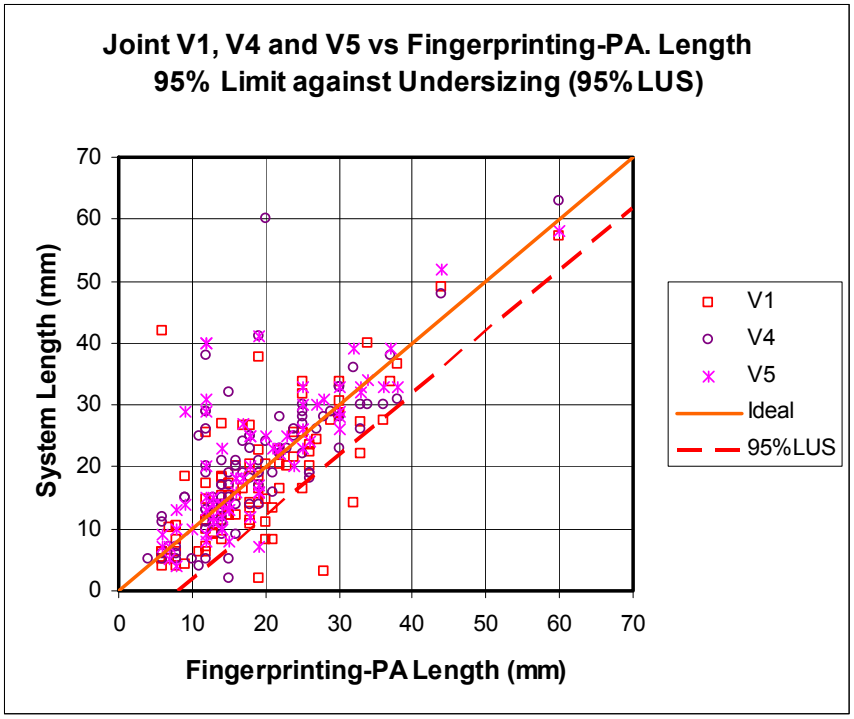


Figure 108. St2 Joint V1, V4, and V5 vs Fingerprinting PA, Length, 95% LUS

The final summary estimates for the depth measurement errors are shown in Table 35. As discussed earlier, a standard format of reporting the result from a single depth measurement is used [Eq. (7)]. The table provides the values of the average (systematic) error and the uncertainty multiplied with the coverage factor for the required 90% confidence.

In addition to pores, three more flaw types were removed from all samples – Natural, IB, and SkTI. Further, the 0SkTI group was removed from the V4vD-Dpt sample. Additional box plots and tables are available in Appendix E illustrating the flaw-type effect on depth measurements. The skew and tilt have statistically significant effect on the depth measurements for all systems while the height and especially length measurements are less affected.

As far as the data normality is concerned, a list of removed outliers required to improve the normality is shown in Table 36. Out of 16 unique outliers, more than 50% (9) were in W5 and W6. A comparison of corrected-normal depth-sizing-error distributions is shown in Figure 109. It clearly indicates that V3vD-Dpt is different from the other distributions. To reiterate, the V3 depth measurement procedure was different from the other procedures. The joint distribution estimates with the V3 data removed are shown in Table 35. The joint distribution is not normal and the coverage factor is not provided. Compared to the error-height joint distribution (Table 26), the error-depth joint distribution has the same uncertainty (1.48 mm) and very small systematic undersizing (-0.2 mm). The height measurements are usually obtained as a difference in the through-wall position of the flaw upper and lower (depth) tips. One possible explanation of the difference in the systematic errors between the height and depth measurements is that the error in the upper flaw-tip through-wall position measurement is higher than the error in the lower flaw-tip through-wall position measurement.

Table 35. Final St2 Summary of Depth Measurements (Reference) for DT, 90% Confidence Bounds for Expression of Single Depth Measurement d_i - $d_i = \hat{d}_i - \bar{\varepsilon} \pm ks(\varepsilon)$

Group	Sample	s(ε) (mm)	Aver ε (mm)	Coverage Factor (k) (df = n-1)	k.s(ε) for 90% Confidence (mm)
FvD-Dpt-PA	80	0.88	-0.32	1.66	1.46
FvD-Dpt-TD	63	0.98	0.14	1.67	1.64
V1vD-Dpt	73	1.56	-0.71	1.67	2.61
V2vD-Dpt	54	1.67	-0.18	1.67	2.79
V3vD-Dpt	50	2.21	-2.36	1.68	3.71
V4vD-Dpt	47	1.06	0.35	1.68	1.78
V5vD-Dpt	57	1.33	-0.03	1.67	2.22
Joint V1, V2, V4, and V5	231	1.48	-0.20	Not normal	Not normal

Table 36. Final St2 Outliers Removed for Data Normality, Depth

Group	Error (mm)	Weld	Sector	Type (Category)	D-Dpt (mm)	D-Hgt (mm)
FvD-Dpt-PA	-3.08	W3	S52	0Sk0Tl	12.36	2.51
	1	W4	S24	0Sk0Tl	12.7	5.3
	-2.5	W5	S14	0Sk0Tl	6.6	3.8
	-3	W5	S31	0Sk0Tl	13.6	3.6
	-2.4	W5	S31	0Sk0Tl	7.8	4.2
	1	W6	S14	0Sk0Tl	3.7	3.1
	-3.6	W2	S56	0SkTl	11.6	4.7
V1vD-Dpt	-8.3	W5	S53	0Sk0Tl	15.7	14.1
	-6.1	W6	S35	0SkTl	10.9	6.1
	-8.8	W6	S52	0SkTl	11.7	6
V2vD-Dpt	-3.3	W5	S14	0Sk0Tl	4.7	3
V4vD-Dpt	-3.2	W3	S24	0Sk0Tl	12.7	4.6
	-2.53	W4	S57	0Sk0Tl	13	3.9
	-2.8	W6	S26	0Sk0Tl	8.6	8.1
V5vD-Dpt	-3.1	W4	S4	0Sk0Tl	13.1	4.6
	-4.3	W2	S13	0SkTl	11.3	4
	-3.1	W2	S56	0SkTl	11.6	4.7
	-2.9	W6	s35	0SkTl	10.9	6.1

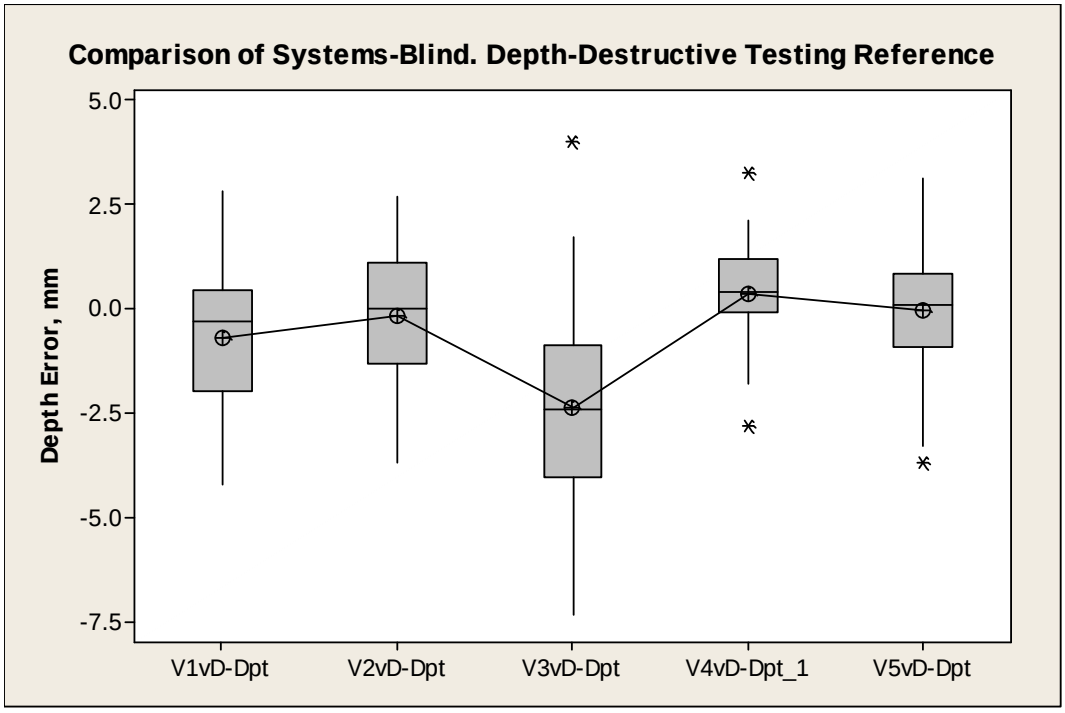


Figure 109. St2 One-Way ANOVA: Final Data, Comparison of Systems, Depth, DT Reference

The final estimate required to determine the AUT system reliability is the FPF. It can be determined by dividing the number of the sectors where the flaw indications were not confirmed by the reference method to the total number of the sectors.[30] Another approach is to provide the number of false indications per unit of scanned length for the entire scanned weld length.[35]

The length of one weld is 2393.9 mm divided into 60 sectors. The total scanned length for the five welds is 11969.5 mm (~12 m) or 300 sectors. The system V5 did not provide data for the W3. Consequently, the scanned length for the V5 is 9.6 m and 240 sectors, respectively.

The systems with false indications, indications per unit length and FPF are as follows:

- FvD-Hgt-PA – 1 in W6 - 0.083 per 1 m and 0.33% FPF
- V3vD-Hgt – 3 in W6 - 0.25 per 1 m and 1% FPF
- V5vD-Hgt - 1 in W5 and 1 in W6 - 0.21 per 1 m and 0.83% FPF

All false positive indications were in the weld specimens with flaw interaction (W5 and W6). The V3 system has the highest FPF.

4.3.5 Conclusions

The following conclusions can be drawn from the statistical analysis:

- From the implanted flaw sample characterization
    - Some of the natural planar flaws (in W3 especially) are used in the POD quantification.
    - The implanted flaws are fabricated with slightly larger depth and larger height (~1.4 mm) than required.
    - Some unusually large outliers and acoustically transparent/fused flaws are removed – the average error and uncertainty estimates are unreliable if large outliers are present in the samples.

- The WT change (W2 and W3) does not have statistically significant effect on the performance.

- Except for depth, the unintentionally tilted (0SkTI) flaws do not have statistically significant effect on the performance either.

- The flaw-type affects the sizing accuracy especially for the height and depth.

- To address the flaw-type effect, the final height and depth sizing accuracy estimates are obtained with Vol (pores), Natural, and IB categories removed from all data samples. Additional categories (SkTI and 0SkTI) are removed for the depth sizing estimates.

- In general, the tilted and skewed flaws do not have statistically significant effect on the height sizing. However, a trend exists and if a larger sample of tilted and/or skewed flaws is used in narrow range of angles the effect might be significant.

- It is challenging and expensive to study the flaw type and skew/tilt effect with physical specimens only.

- Some system data samples (e.g., V3) are removed from the joint distributions because of statistically significant differences with the others.

- Many not normal distributions require removal of few outliers (in most cases) to become normal.

- The flaw interaction in W5 and W6 also affects all sizing estimates especially for the height and FPF.

- The pores are difficult to detect and size. All pores (Vol category) are removed to obtain the final POD and sizing estimates.

- The FPF is relatively low 0.33% for one of the fingerprinting techniques (PA) and two of the systems - 1 and 0.83%, respectively. The other fingerprinting technique (TOFD) and three AUT systems did not have any false-positive indications compared to DT.

- The height sizing and detection performance for the two fingerprinting techniques PA and TOFD is as follows:
    - Sizing: -0.17 and -0.05 mm systematic error, 1.05 and 0.76 mm uncertainty.
    - POD $a_{90/95}$ with various link functions and software packages: from 1.99 to 2.34 mm.

- The height sizing and detection performance for the AUT systems is as follows:
    - Sizing
        - Separate AUT systems: from -0.75 to -1.27 mm systematic error and from 1.04 to 1.95 mm uncertainty.
        - Joint V1 to V5: -1.04 mm systematic error and 1.48 mm uncertainty.
    - POD $a_{90/95}$ with logit link without size log transformation
        - Separate AUT systems: from 3.25 to 5.58 mm.

        - Joint V1, V4, and V5: 3.5 mm.
    - POD $a_{90/95}$ for the best performer V4 with various link functions and software packages
        - Blind trials: from 3.21 to 4.52 mm, for Vol and Natural removed – 3.02 mm.
        - Open trials: from 2.61 to 3.34 mm, for Vol and Natural removed - 2.08 mm.

- The length sizing performance for the AUT systems is as follows:
    - Separate AUT systems: from 2.13 to 0.36 mm (minimum) to -2.48 mm systematic error and from 4.1 to 5.48 mm uncertainty.
    - Joint V1, V4, and V5: -0.1 mm systematic error and 4.89 mm uncertainty.

- The depth sizing performance for the fingerprinting and the AUT systems is as follows:
    - Fingerprinting with PAs and TOFD: -0.32 and 0.14 mm systematic error, 0.88 and 0.98 mm uncertainty.
    - Separate AUT systems: from 0.35 to -0.03 mm (minimum) to -2.36 mm systematic error and from 1.06 to 2.21 mm uncertainty.
    - Joint V1, V2, V4, and V5: -0.2 mm systematic error and 1.48 mm uncertainty.

- The minimum $a_{90/95}$ obtained during the open trials (examiner influence minimized) is approximately 2 mm for the fingerprinting and the best performer.

4.3.6 Recommendations

The following recommendations can be made from the statistical analysis:

- Better and more reliable techniques are needed for the fabrication of flaws with height smaller than 2 mm.

- Unless specifically targeted, IB-LOF, natural and interacting flaws should not be used for quantification of the AUT system due to the significant and difficult to control effect on the sizing performance.

- Validated (through comparison with reasonable number of specimens) computer modeling techniques are needed as part of the TJ and where needed to conduct studies related to different flaw type and skew/tilt effects on the sizing performance.

- Better techniques are needed (e.g., high-resolution eddy current) to determine whether a flaw is fused or transparent to the acoustic energy on a macro.

- If comparison of the AUT system POD capabilities is required, one software package or link function should be used (if possible) to eliminate the effect of variability between the different computational techniques when used with the same set of data.

- Provided the systematic error is accounted for and compensated, the uncertainty of the fingerprinting techniques although smaller are still in the same range of magnitude as the AUT systems uncertainties (Table 26) and cannot be ignored. Although considered the standard practice, the metallographic analysis is not perfect either. The POD and sizing estimates will be affected if the errors of the fingerprinting (reference) techniques are comparable with the AUT system errors.[30] Additional study is needed to investigate and justify the use of the fingerprinting techniques to replace the DT as the primary reference technique.

5.0 Field Tests

5.1 Field Trials

The objective of the field trials was to validate and compare current zonal discrimination PA inspection techniques to linear phased-array (LPA) non-zonal inspection techniques on production pipeline girth welds. EWI developed a LPA field test procedure and evaluated the performance of an LPA AUT system under field conditions with the help of UT Quality (UTQ). As an in-kind cost share, project team member TCPL provided access to a pipeline construction site in Hardisty, Alberta, Canada for field trials from June 5, 2009 through June 18, 2009. The field tests were conducted on a 30-in. OD seam welded, carbon steel pipeline with a 9.8-mm WT. EWI conducted field trials for this project concurrently with DOT project *Enhanced Defect Detection and Sizing Accuracy Using Matrix Phased Array Ultrasonics Tools Program* (DTPH56-08-T-000002); EWI Project No. 50854GTH.

5.2 Field Trial Location

EWI performed the field trials on a section of the Keystone pipeline project operated by TCPL located in Hardisty, Alberta, Canada. A photograph of the field trial site in Hardisty is shown in Figure 110. The 2,148-mile Keystone Pipeline will transport crude oil from Hardisty, Alberta, Canada to U.S. Midwest markets in Wood River (Patoka, Illinois) and to Cushing, Oklahoma Figure 111. The Keystone Pipeline varies in size and WT.

Figure 110. Field Trial Site

Figure 111. Map of Keystone Pipeline Project

5.3 EWI Inspection Procedure

The non-zonal inspection procedure used during the field trials was developed by EWI using information gained during other tasks of this project. The procedure is contained in Appendix G and is primarily composed of two LPA probes (US and DS) carried by a band scanner. Instead of using tandem zonal techniques for fill passes, the EWI procedure consisted of three sectorial scans (S-scans) for each probe, using different element groupings. This arrangement provided complete coverage of the weld zone with multiple angles. A screen display showing the output from this technique is shown in Figure 112.

For the non-zonal approach, the idea was to identify flaws and determine their through-wall height and length, regardless of signal amplitude. Consequently, the gain settings for each S-scan were set to maintain an acceptable signal-to-noise ratio. The calibration sample was used as a verification that the holes and notches could be detected and correctly measured.

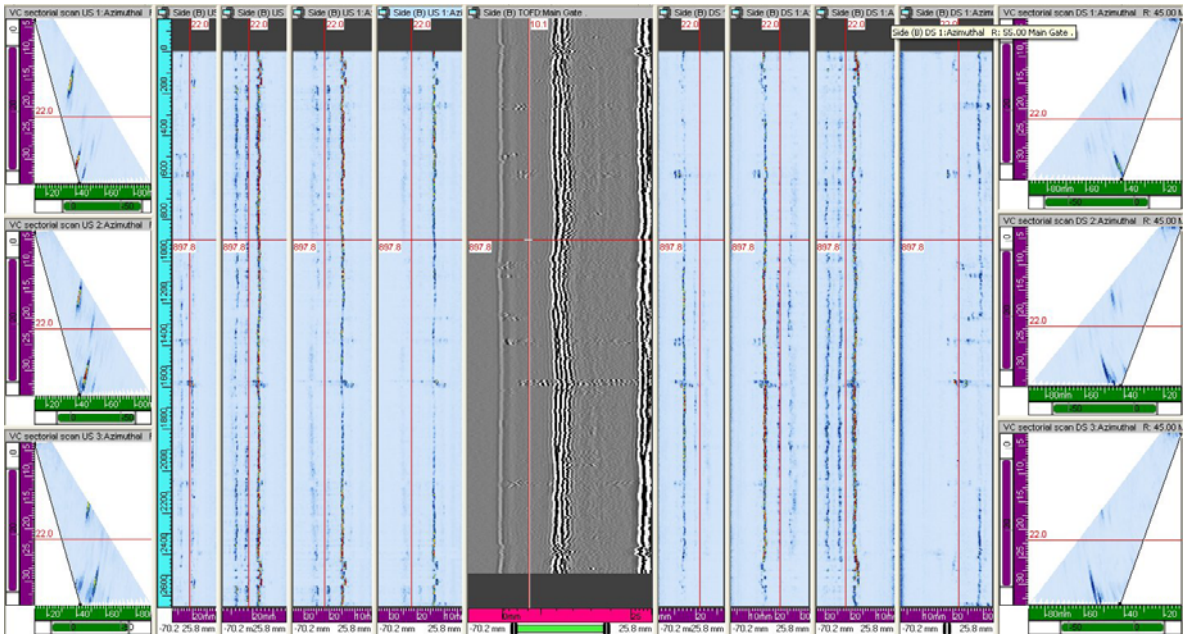


Figure 112. Non-Zonal Sector Scan Screen Display

5.4 Calibration Blocks

Early in the project, UTQ provided a calibration drum for use by all project team members who performed scanning with their equipment at EWI. This drum had a 15.6-mm calibration sample also owned by UTQ that was designed for a single U-groove weld joint with an 8-degree included angle for a 15.6-mm-wall pipe (illustrated in Figure 113). The calibration drum itself was designed to mount on a bracket at the rear of a truck; therefore, EWI decided to use the UTQ calibration drum for the field trials. The UTQ calibration block drawings are located in Appendix H.

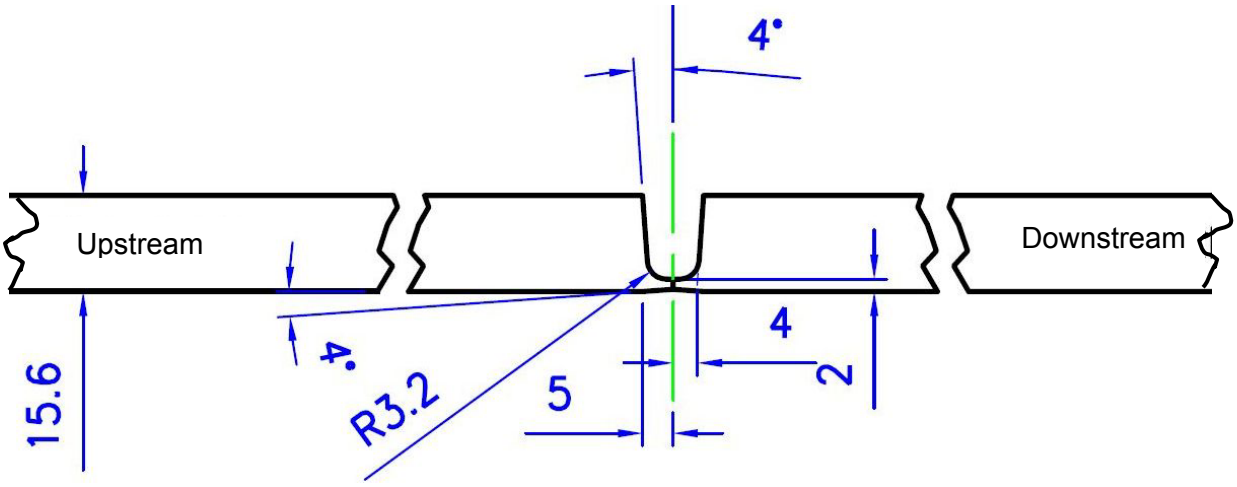


Figure 113. U-Groove Joint for Calibration Block

When field trial preparations were under way, EWI was informed that the pipe for the field trials would have a 10-degree included angle single U-groove weld preparation geometry with a 15.6-mm pipe WT. The inspection procedure in Appendix G was designed to inspect this joint/material thickness combination.

As mentioned previously, the Keystone Pipeline varies in diameter size and WT. Shortly before the field trials, EWI was informed that the pipeline at the Hardisty site would be a 10-degree included angle single U-groove weld preparation geometry with a 9.8-mm WT. Based on this geometry/material thickness combination, EWI ordered a new calibration block to be delivered to the UTQ facility where the EWI scanning equipment was being delivered/assembled prior to the field trials. The new EWI calibration block drawings are located in Appendix I. Figure 114 shows the new EWI 9.8-mm-thick calibration block mounted on the calibration drum attached to the truck provided by UTQ.

Figure 114. UTQ Calibration Drum and 15.6-mm Calibration Sample

A small problem was encountered mounting the EWI prototype scanner on the calibration block. The EWI prototype band scanner requires 230 mm of distance from the edge of the band to the weld centerline. Due to the position of the calibration blocks, band placement and drum size, this 230-mm distance was not achievable. To overcome this issue, the DS probe was relocated to the next forward rail and an offset measure between both PA probes was added to the set up.

When EWI arrived at the field trial site, the weld joints to be inspected consisted of a 20-degree included angle K-groove weld joint geometry for 9.8-mm pipe WT (per TCPL Welding Procedure Specification LD-A-WPS1). Figure 115 is a sketch of this weld joint geometry. While this was different than the bevel angle in the EWI calibration sample, the non-zonal technique was readily adaptable since the technique did not depend on absolute signal amplitude and the same S-scan beam angles could be used for the K-groove weld.

The approach for the field trial was simple. When UTQ discovered a significant weld defect, EWI was to step in and do a LPA scan (for DTPH56-07-T-000002). When that scan was complete, EWI was to change the scanner equipment to do a MPA scan (for DTPH56-08-T-000002). This was anticipated to be the best approach to gather data in the most efficient manner with minimal interruption to the TCPL production schedule. Figure 116 shows the EWI truck (provided by UTQ) following the UTQ truck with the crew that was performing the AUT inspections for TCPL.

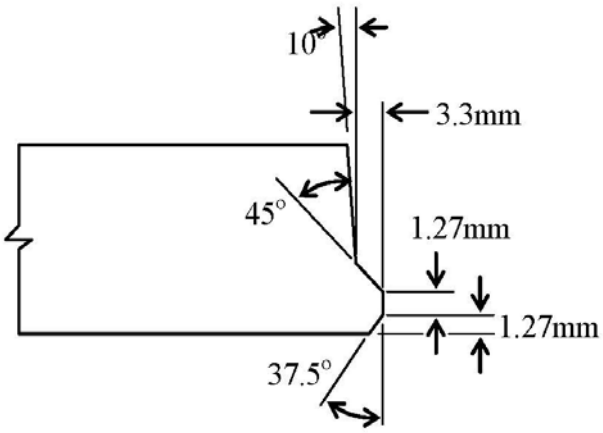


Figure 115. Field Trial Weld Joint Geometry

Figure 116. EWI Truck Following UTQ Truck Performing AUT Scans for TCPL

Using the non-zonal technique under field conditions was beneficial from a couple of points. First, the technique could be tested under actual conditions to access the ability to keep up with production, and second, it provided a direct comparison of flaw detection with current zonal approach techniques. Figures 117 through 122 show example comparisons made during the field trials. Further comparisons and a daily activity log is provided in Appendix J.

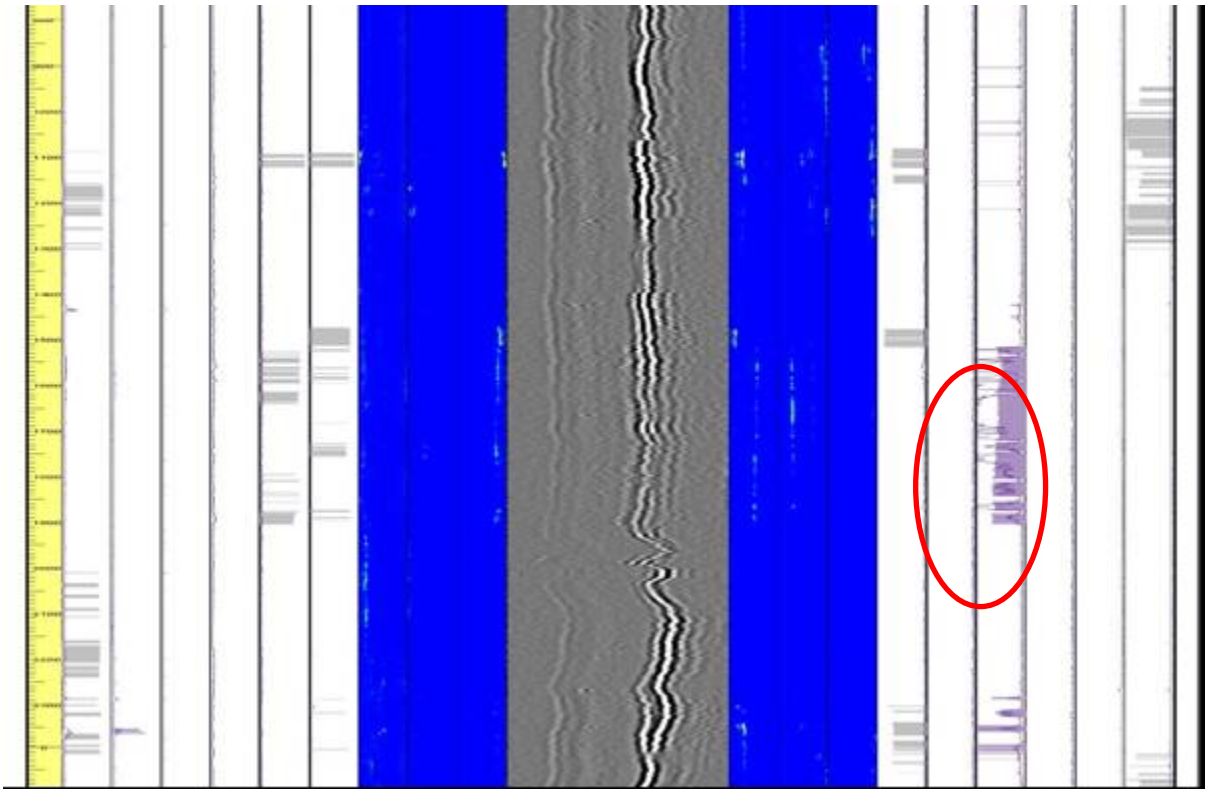


Figure 117. Weld MLA 4530 - UTQ Scan Screen Capture

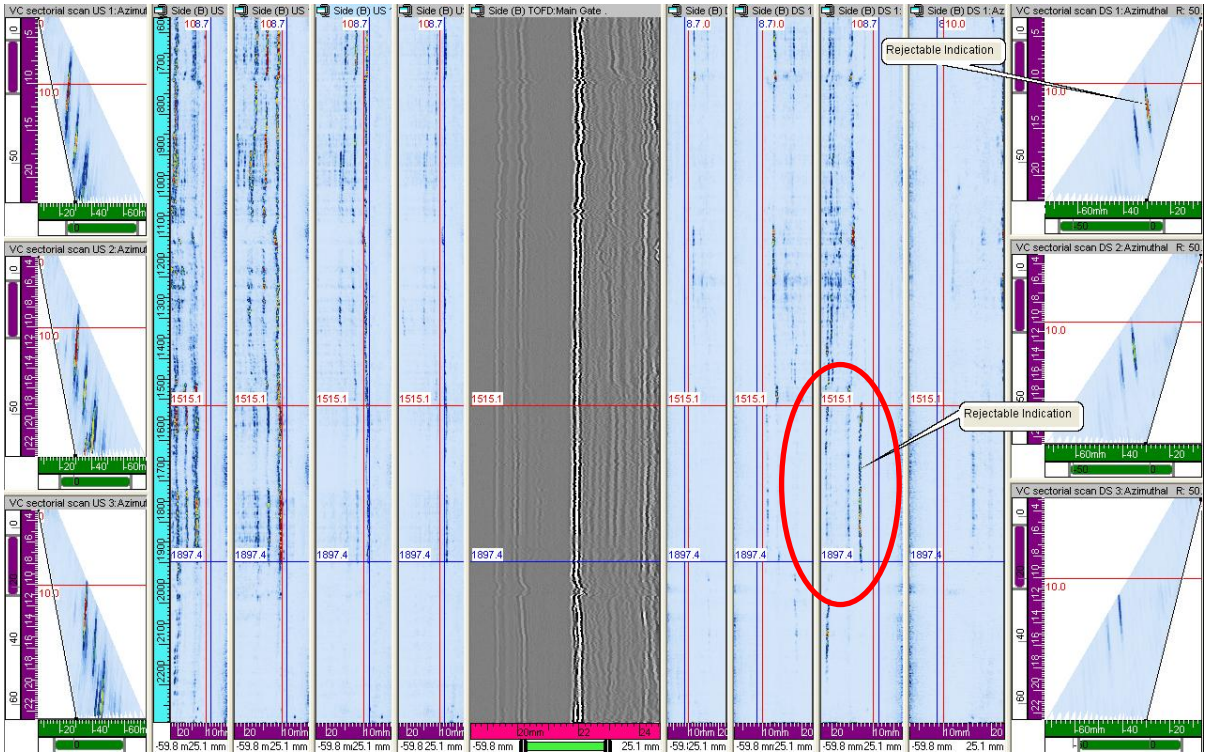


Figure 118. Weld MLA 4530 - EWI Scan Screen Capture

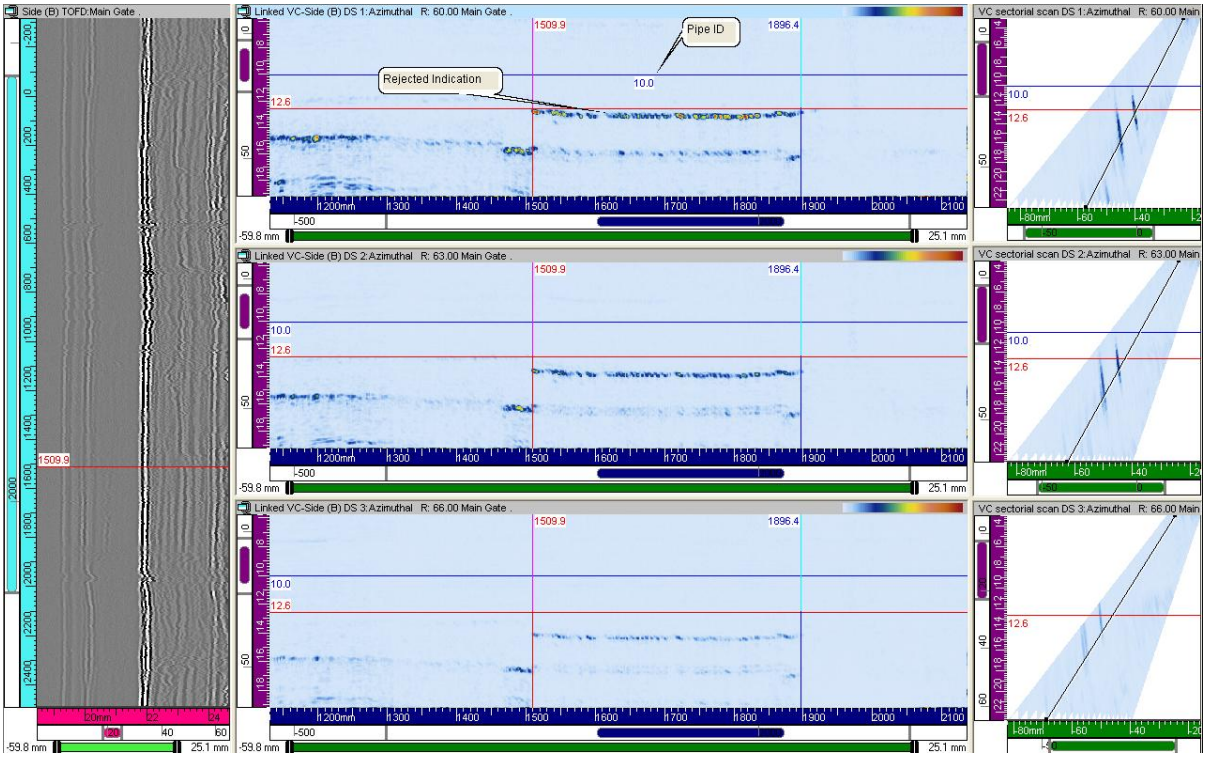
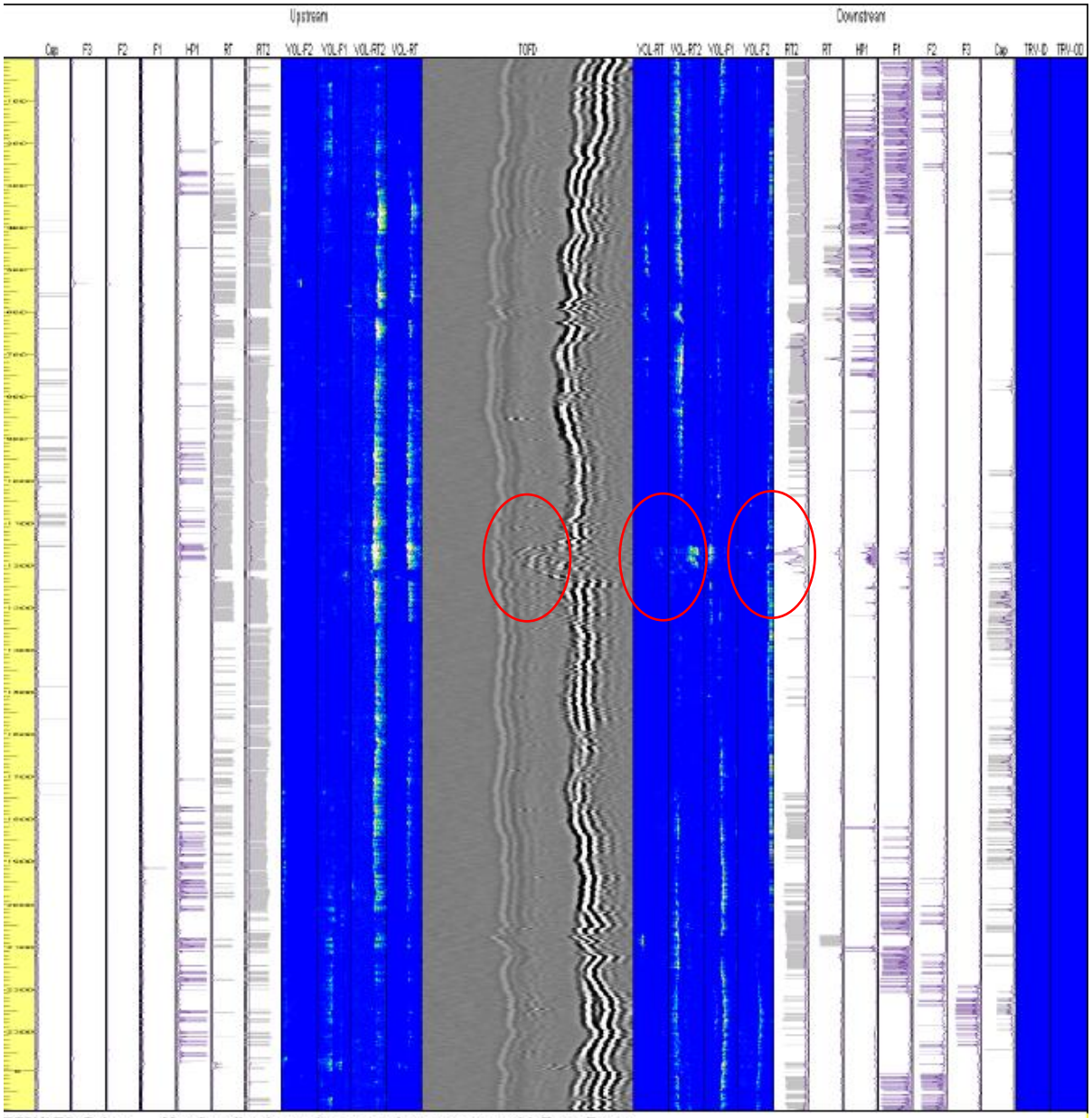


Figure 119. Weld MLA 4530 - EWI Scan Screen Capture - DS Zoomed View



No.	Seq.	Label.	Start(mm)	End(mm)	Length	Height	Depth	TD(mm)	Ampl(%)	Type	Result
1	14	RT2_D	1158	1199	41	2.60	9.80	2.50	99	Incomplete Fusion	Reject
2	25	TOFD	1134	1237	103	2.00	9.80	2.22	102	Incomplete Fusion	Reject

Figure 120. Weld PB 0996 - UTQ Scan Screen Capture

Figure 121. Weld PB 0996 - EWI Scan Screen Capture

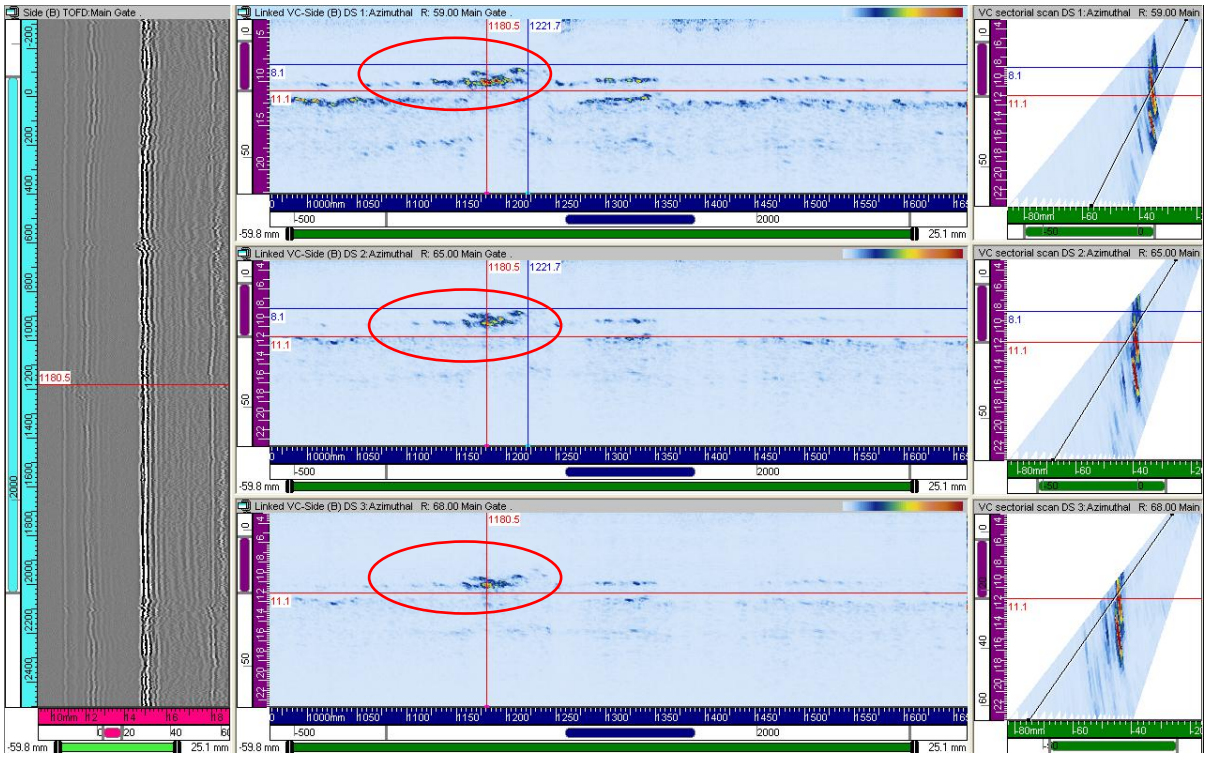


Figure 122. Weld PB 0996 - EWI Scan Screen Capture - DS Zoomed View

5.5 Task Conclusions and Significance

The non-optimized EWI LPA non-zonal procedure was able to detect the same indications that the fully optimized zonal discrimination technique detected. Although the EWI LPA non-zonal procedure was calibrated/optimized for a 10-degree included angle U-groove for 15.6-mm WT, it detected indications on a 20-degree included angle K-groove weld joint geometry for 9.8-mm wall well within the current inspection code criteria. Zonal discrimination techniques are not capable of inspecting weld joint geometries for which they are not calibrated/optimized.

The LPA non-zonal technique had more data display options due to the amount of data gathered with each scan. The technique was able to display flaws in a much clearer format (S- and B-scan) as compared to the zonal discrimination scan format.

The LPA non-zonal technique and equipment performed well in the field trials and was able to provide a full weld scan in approximately 2.5 min with a data file size of approximately 160 MB. The zonal technique was faster and produced small data files. The zonal technique scan time was approximately 45 s and produced data files approximately 9 MB in size. These are factors that would need to be considered when selecting the non-zonal technique.

Current inspection codes (e.g., API 1104 and ASTM 1961) allow PA zonal discrimination techniques. Based on the results of the field trials, it is recommended that LPA non-zonal PA techniques be considered for code approval.

Future DOT PHSMA projects should feature field trials as a means to evaluate the applicability of new technologies for the pipeline industry.

6.0 Reference Library

6.1 AUT Library

Throughout the course of the project, many AUT scans were conducted using both zonal and non-zonal inspection techniques. Most of the AUT data was collected on pipe welds containing implanted flaws, which was part of the quantification portion of the project. These welds were later cross-sectioned and macro photographs were taken of the flaws. Images of flaw responses, along with metallographic cross-section images of flaws, were assembled into a *reference library* of flaw responses can be used for future training and reference.

For the reference library, all girth weld examples were 30-in. diameter, X80 carbon steel pipe with a 15.6-mm nominal WT. The weld bevel was a J-bevel narrow groove weld with a 4-degree bevel angle.

The reference library is contained in Appendices K-T of this report. Table 37 is a list of reference library images and the appendix where they are located.

Table 37. Reference Library Appendices

Appendix	Weld No.	Scan Technology
K	2	Non Zonal
L	2	Zonal Tech
M	3	Non Zonal
N	3	Zonal Tech
O	4	Non Zonal
P	4	Zonal Tech
Q	5	Non Zonal
R	5	Zonal Tech
S	6	Non Zonal
T	6	Zonal Tech

6.2 Advanced NDE Training

A training class was developed for representative AUT systems that featured project results and was sponsored by ExxonMobil Development Company (EMDC) as cash cost share for the DOT program. The attendees received an in-depth theoretical and practical overview of the current advanced NDE methods and techniques that are used in oil and gas industry including current project results. The training event was held at EWI in Columbus, OH, from March 24-28, 2008 and covered the following topics.

- Introduction to conventional [vision testing (VT), penetrant testing (PT), magnetic particle testing (MT), eddy current testing (ET), radiography testing (RT), and ultrasonic testing (UT)] and advanced NDE [acoustic emission (AE), automated ultrasonic testing (AUT), and computed radiography (CR)] of materials and welds with limited PT, MT, CR demonstrations – 8 hours.
- Ultrasonic testing (UT) and AUT practical demonstrations and training – 8 hours.
- Optimization of UT/AUT procedures using UT modeling and simulation tools with practical demonstrations and training; UT/AUT procedures validation and qualification – 8 hours.
- ET practical demonstrations and training – 8 hours.
- Advanced ET practical demonstrations and training; ET/advanced ET procedures validation and qualification – 8 hours.

The first day of the training was mainly theoretical with limited demonstrations. During the last 4 days of the training, EMDC staff were divided into two groups. Parallel training was conducted with a student to teacher ratio of 4:1 to better expose and involve participants in

practical demonstrations and applications related to AUT and ET. The training schedule and syllabus are described below:

- Day One – Welcome and Theoretical Introduction to Advanced NDE Techniques
- Day Two and Three – UT/AUT practical
- Day Four – RT/CR and MT/PT practical
- Day Five – ET practical

Day One

Group A and Group B – Lecturer Dr. Mark Lozev

- Advanced NDE Overview
- Advanced Ultrasonics and Acoustic Emission
- Break
- Advanced Radiography
- Lunch Break
- Advanced Penetrant and Magnetic Particles
- Break
- Advanced NDE/AUT Quantifications/Qualifications

Day Two

Group A - Lecturer Mr. Roger Spencer

- UT Weld Inspection Review
- Break
- UT Weld Inspection Introduction Demos
- Lunch Break
- CIVA Modeling and Simulations
- Break
- CIVA AUT Optimizations

Group B - Lecturer Mr. Perry White

- UT Weld Inspection Demos with MasterScan on SDH block
- Break
- UT/AUT Omniscan Weld Inspection Demos
- Lunch Break
- UT/AUT weld inspection on first set of selected samples

Day Three

Group A - Lecturer Mr. Roger Spencer

- PA General
- Break
- AUT
- Lunch Break
- AUT Interpretation

Group B - Lecturer Mr. Perry White

- Weld scanning using Focus LT and Focus on selected samples
- Lunch Break
- Demo of UT Scan/Weldstar system

Day Four

Group A - Lecturer Mr. Perry White

- PT Weld Inspection Demo and PT on selected samples
- Break
- MT Weld Inspection Demo and MT on selected samples
- Lunch Break

Group B - Lecturer Mr. Kevin Clear

- Film RT Weld Inspection Demo and Film RT on selected samples
- Break
- CR Weld Inspection Demo and CR on selected samples
- Lunch Break

Day Five

Group A and Group B – Lecturer Dr. Evgueni Todorov

- EC Techniques and Applications
- Break
- Modeling of EC Techniques and Procedures
- Break
- Advanced EC Techniques
- Lunch Break
- Demonstration of Typical EC Techniques
- Break
- Demonstration of Advanced EC Techniques

Following is a list of the EMDC staff that participated in the training event.

- Nathan E. Nissley, Ph.D.
- Joshua Sleigh
- Brandon Henneke
- Wentao Cheng, Ph.D.
- Karl C. Henize
- Juan Orphee
- Russell E. Tanner
- Po Yan Ho

Each attendee was given a CD with PowerPoint presentations for all NDE techniques covered during the training.

7.0 Guidance Document

During the course of the project, a guidance document was formulated which was based on lessons learned during the project and feedback from project participants. The document was designed to provide general guidance for performing AUT quantification activities in a standardized fashion and allow results to be more transferrable. The document is provided in Appendix U.

The quantification process provides a good understanding of AUT capabilities for a given set of parameters. The resulting POD and accuracy of sizing information, obtained from the AUT quantification process, is intended to be used in conjunction ECA-based acceptance criteria for girth welds such as API 1104; Appendix A. For example, API 1104; Appendix A provides three options for determining acceptable imperfection size for girth welds. All three options require that inaccuracies in NDE measurements of imperfection height be taken into account when determining the ultrasonic acceptance criteria. Certain inaccuracies are assumed in the options provided, but other values (larger or smaller) may be used if demonstrated through inspection trials on flawed samples. In API 1104, the assumed inaccuracy in flaw height measurements is the lesser of 1.5 mm or 8% of the WT.

In general, the quantification process is designed to provide the user with two critical pieces of information. One is POD, which allows the user to determine if the ECA calculated flaw size can be detected at the probability and confidence levels needed. The second important piece of information is the 95% limit against under sizing (95% LUS), which shows the accuracy for determining flaw height using the AUT equipment, procedures, and personnel. The 95% LUS result can then be subtracted from the ECA size to produce AUT acceptance criteria that accounts for errors in AUT sizing.

8.0 Standards Implementation

During this project a guidance document was developed for conducting AUT quantifications. The document is provided in Appendix U. The eventual goal is for the guidance document to be implemented into future revision(s) of API 1104, Appendix A (RBDA standard) and ASTM 1961. As the document was being developed, consideration was given to current practices and ways to harmonize the guidance document to smooth the incorporation into standards. During this task, standards and industry documents, including DNV-OS-F101, were reviewed to make certain the resulting guidance would fulfill requirements of industry. Project sponsors and participants provided valuable input as to current practices and possible improvements. Some

of the sponsors are members of committees that can influence decisions to incorporate results of this project for improving AUT reliability.

9.0 Financial Reporting

This report section summarizes the details of the project financials, including the status of the contributions by the Team Participants. It includes a final accounting of budgeted and actual expenditures, including cost share amounts. It also explains any major deviations.

A high-level summary of project funding is shown in Table 38. Slightly more cost share ($113,796) was obtained as compared to the plan ($1,637,928 was obtained; $1,524,132 was planned).

Table 38. Summary of Actual Project Funding

Funding Type	Funding ($)
Government	612,682
Cash Cost Share	907,098
In-Kind Cost Share	730,830
Total	**1,637,928**

9.1 Government Funding

DOT Agreement No. DTPH56-07-T-000002 was fully executed on July 2, 2007; the effective date is July 1, 2007. DOT is providing a total of $612,682 for Tasks 1-8. All expenses for the DOT-PHMSA funded portion of this project were tracked per task in the automated EWI accounting system (Solomon) via EWI Project No. 50454GTH. Budget and cumulative costs are shown in Table 39.

Table 39. DOT Funding per Project Task

Task	Budget ($)	Cumulative Expenses ($)
1	82,549	81,987
2	66,690	66,990
3	90,512	90,831
4	56,243	56,295
5	45,895	45,935
6	76,936	76,926
7	70,987	70,298
8	122,870	123,420
Totals	**612,682**	**612,682**

9.2 Team Cost-Share Funding

As documented in proposal Section 3.2.7 (Cost Sharing/In-House Contributions, and Joint Ventures), the project was projected to benefit from industry cost sharing of **$910K**, which was comprised of **$280K** of direct cash funding and **$630K** of in-kind support. ConocoPhillips, Chevron, BP, Heerema, TCPL, GE Inspection Technologies, UTTechnology/UTQuality, Mechanical Integrity, and ISQ committed to cost sharing the project when the proposal was written. Report Section 9.2.1.1 is a description of the planned cash cost-share and Section 9.2.1.2 is a description of the planned in-kind support.

Also included in proposal Section 3.2.7 was "additional" cash/in-kind contributions from EMDC, EWI and Petrobras that are expected during the project; however, at the time the proposal was written, these contributions were not yet quantified. Report Section 9.2.2.1 is a description of the additional cash cost-share and Section 9.2.2.2 is a description of the additional in-kind support.

9.2.1 Planned Cost-Sharing (As Quantified in Proposal)

9.2.1.1 Planned Cash Cost-Share

Planned cash cost-share contributions were provided via parallel EWI R&D projects funded by individual cost-share partners. When the proposal was written, a total of $280K in cash cost-share was committed. These projects featured a unique agreement between EWI and the cost-share partner and a unique project number against which EWI expenses were tracked. Table 40 contains a summary of all planned cash cost-share contributions Table 40 (budget and cumulative).

Table 40. Status of Planned Cash Cost-Share

Team Participant	EWI Project No.	Task	Budget ($)	Cumulative ($)
ConocoPhillips	50600CSP	3, 5	200,000	200,000
UTTechnology and UTQuality	50416CSP	3	80,000	80,000
Total			**280,000**	**280,000**

ConocoPhillips is a leading U.S. pipeline/offshore operating company, who will provide a cash cost-share contribution of **$200K** in support of Task 3 and 5 activities. It will be tracked as a lump sum by EWI Project No. 50600CSP in the EWI accounting system. This funding will be used to fabricate weld samples with flaws and to destructively validate AUT detection and sizing capabilities. The agreement between ConocoPhillips and EWI was fully executed on August 23, 2007. **EWI received all $200K from ConocoPhillips as of January 2009.**

UTTechnology (UTT), located in Canada and the U.S., is a leading manufacturer of industrial AUT multi-probe and PA equipment and UTQuality (UTQ) is a leading NDE inspection company. Combined, these two organizations will provide a cash cost-share contribution of **$80K** in support of Task 3 activities. It will be tracked as a lump sum by EWI Project No. 50416CSP in the EWI accounting system. This funding will be used to fabricate weld samples with flaws and to destructively validate AUT detection and sizing capabilities. The agreement between UTT/UTQ and EWI was fully executed on August 7, 2007. **EWI received all $80K from UTT as of June 2010.**

9.2.1.2 Planned In-Kind Support

When the proposal was written, a total of **$630K** in in-kind cost-share was committed. In-kind cost share was provided to EWI in accordance with as applicable, *49 CFR 18 - Uniform Administrative Requirements for Grants and Cooperative Agreements to State and Local Governments, Section 18.24, Matching or Cost Sharing*, or *49 CFR 19 - Uniform Administrative Requirements for Grants and Agreements with Institutions of Higher Education, Hospitals, and Other Non-Profit Organizations, Section 19.23, Cost Sharing or Matching*. EWI captured and maintained an Excel spreadsheet with a running summary of planned in-kind cost-share contributions as reported by team participants. Budget and cumulative costs are shown in Table 41.

Table 41. Status of Planned In-Kind Support

Team Participant	Task	Budget ($)	Cumulative ($)
Chevron	1,3,5,6,7	40,000	0[*]
GE Inspection Technologies	2,3	100,000	105,486
~~ISQ~~[†]	~~2,3~~	~~100,000~~	~~0~~
Mechanical Integrity	3	100,000	100,000
TCPL	4,5,6	100,000	100,000[‡]
UTQuality	3	40,000	74,094[§]
UTTechnology	3	150,000	150,000
	Totals	**530,000**	**529,580**

Chevron is a leading U.S. pipeline/offshore operating company. Chevron provided an in-kind contribution in the form of an US NDE expert in support of Tasks 1, 3, 5, 6, and 7. The value of

[*] John O'Brien began contributing to the project in the quarter ending March 31, 2009. Luc Huyse attended a project meeting at EWI on July 29, 2010.

[†] ISQ officially withdrew from the project during the quarter ending on December 31, 2008, as their equipment will not be ready to scan welds before the project deadline to scan the samples with implanted flaws.

[‡] On February 22, 2010, Evan Vokes reported (via phone call) that the TCPL cost share has been met in full. When David Hodgkinson retired, he (Evan) didn't capture the costs correctly as he wasn't familiar with the accounting system when he took over.

[§] On June 22, 2009, UTT updated their in-kind cost share amount to include all funding spent to date.

this in-kind effort was estimated at **$40K**. Chevron staff* provided support virtually and attended a at project meeting at EWI; however, Chevron did not report a report a dollar value for its in-kind contributions. **This in-kind cost share contribution was provided, but no dollar value was assigned to it.**

GE Inspection Technologies is a leading manufacturer of industrial AUT multi-probe and PA equipment in Canada and the U.S. GE Inspection Technologies provided an in-kind contribution in the form of equipment leases valued at **$100K** in support of Tasks 2 and 3. **This cost share was provided in full.**

ISQ is the largest R&D welding and NDE organization in Portugal. ISQ planned to provide in-kind contributions in three forms: direct labor of inspector's time, selected results of a parallel R&D program funded at ISQ, and welds. The estimated value of this in-kind cost-share was **$100K** and would have been in support of Tasks 2 and 3. **In Quarter 6, ISQ withdrew from the project as their scanning equipment will not be operational before the project ends.** In Table 41, this contribution is shown in strike through text.

Mechanical Integrity is a leading NDE inspection company located in Canada and the U.S. Mechanical Integrity provided direct labor of inspector's time as an in-kind contribution valued at **$100K** in support of Task 3. **This cost share was provided in full.**

TransCanada Pipeline Ltd. (TCPL) is one of the largest transmission pipeline companies in Canada. TCPL provided **$100K** in the form of direct in-kind support to coordinate Field Testing of AUT systems for girth weld inspection using zonal (supplied by TCPL) and non-zonal (supplied by EWI) approaches that is proposed in Task 4 of the project. Field trials were hosted in Hardisty, Alberta (Canada) from June 5-18, 2009. On February 22, 2010, Evan Vokes reported (via phone call) that the TCPL cost share has been met in full. When David Hodgkinson retired, he (Evan) did not capture the costs correctly as he wasn't familiar with the accounting system when he took over. **This cost share was provided in full.**

UTQuality (UTQ) is a leading NDE inspection company located in Canada and the U.S. UTQ provided personnel and procedural support valued at **$40K** in support of Tasks 2 and 3. **This cost share was provided in full and more.**

UTTechnology (UTT) is a leading manufacturer of industrial AUT multi-probe and PA equipment in Canada and the U.S. UTT provided in-kind contribution in the form of AUT system/imaging system valued at **$150K** in support of Task 3. (UTT is the parent company that owns UTQ.) **This cost share was provided in full.**

9.2.2 Additional Cost-Sharing (NOT Quantified in Proposal)

In Section 3.2.7 of the proposal to DOT, additional cash/in-kind contributions from EMDC, DNV, EWI and Petrobras were noted as "expected" during the duration of the project; however, at the time the proposal was written these contributions were not yet quantified. These cost-share contributions are not only above and beyond the proposal, they are also above and beyond the 50/50 government/industry cost-share requirement as dictated by law. This report section documents additional cost-share as promised and delivered.

9.2.2.1 Additional Cash Cost-Share

Additional cash cost-share contributions were provided via parallel EWI R&D projects funded by individual cost-share partners. These projects featured a unique agreement between EWI and the cost-share partner and a unique project number against which EWI expenses were charged/tracked. In the case of EWI purchasing some PA equipment in support of Task 3, the EWI purchase orders were used to track this contribution. Table 42 contains a summary of all additional cash cost-share contribution** (budget and actual).

Table 42. Status of Additional Cash Cost-Share

Team Participant	EWI Project No.	Task	Budget ($)	Actual** ($)
EWI	50717GTO	3	90,000	94,983
EWI (Equipment Purchase - PA Instrument)	08-0173**	3	59,607	59,607
EWI (Equipment Purchase - Scanner)	08-0078**	3	24,525	24,525
EMDC	51180CTR	3	25,000	22,983
Petrobras††	50907CSP	3	200,000	200,000
PRCI	51219CSP	3	225,000‡‡	214,946
		Totals	***624,132***	***617,044***

EWI provided a cash cost-share contribution of **$94,983** in support of Task 3 activities. It was tracked as a lump sum by EWI Project No. 50717GTO in the EWI accounting system. This project was initiated on July 1, 2007. **This cost share contribution was provided in full.**

** This is an update EWI purchase order.

†† An updated proposal was submitted to Petrobras December 8, 2008.

‡‡ PRCI was asked to provide a cash cost-share of $225K in support of the project. On June 29, 2010, EWI received a contract for the remaining $85,200.

EWI provided a cash cost-share contribution of **$84,132** in the form of purchasing a pipe scanner and a PA instrument in support of Task 3 activities. The scanner tracked via EWI Purchase Order (PO) No. 08-0078 and was quoted at $24,525. The scanner was purchased and received. The PA instrument was tracked via EWI PO No. 08-0173 and was quoted at $59,607. The PA instrument was purchased, received and used for this project. **This cost share contribution was provided in full.**

EMDC is the world's largest publically traded international oil and gas company. EMDC was originally asked to provide a cash cost-share contribution of approximately **$110K** in support of Task 2 and 5 activities. Since then, EMDC has implemented a major change in the direction of its R&D investments. The proposed additional cash-cost share (formerly listed as EWI Project No. 51043CSP for approximately $110,000 in Table 4 of the Quarter 2 report), was not funded. EMDC's contributed cost-share contribution was in support of Task 3 in the form of a training class that they sponsored in March 2008, which was tracked as a lump sum by Project No. 51180CTR in the EWI accounting system. **This cost share contribution was provided in full.**

Petrobras is an integrated company that performs in oil and oil byproduct exploration, production, refining, marketing, and transportation, both in Brazil and abroad. Petrobras was asked to provide a cash cost-share contribution of **$200K** in support of Task 3 activities. The agreement between EWI and Petrobras was fully executed on April 22, 2009. As of June 2, 2010, EWI has received all $200K from Petrobras. **This cost share contribution was provided in full.**

PRCI is a collaborative technology development organization of, by, and for the energy pipeline industry. PRCI was asked to provide a cash cost-share of **$225K** in support of the project (this was tracked as EWI Project No. 51219CSP in Table 42. This funding was used to fabricate weld samples with flaws and to destructively validate AUT detection and sizing capabilities in support of Task 3. Every fall, PRCI members will vote on whether to provide the next increment of cost-share funding for this project for the following calendar year. PRCI provided $74,500 in cost-share funding for calendar year 2008. PRCI provided $65,300 for calendar year 2009. In the fall of 2009, PRCI members voted to not provide the remaining $85,200 in cost share. As a result, EWI was forced to scale back the sample cross-sectioning effort by approximately 50%. On November 10, 2009, the PRCI Materials Technical Committee provided EWI a list of their concerns. On December 11, 2009, EWI submitted a response to PRCI, which addressed their concerns and dispelled many misconceptions about the project in an effort to persuade PRCI to restore their cost-share funding. On February 22, 2010, PRCI informed EWI that they intend to restore the remaining cost share by the end of March. On June 29, 2010, EWI received a PRCI contract to reinstate the remaining $85, 200 in cash cost share. **This cost-share contribution was provided in full.**

9.2.2.2 Additional In-Kind Support

Additional in-kind cost share was provided to EWI in accordance with as applicable, *49 CFR 18 - Uniform Administrative Requirements for Grants and Cooperative Agreements to State and Local Governments, Section 18.24, Matching or Cost Sharing*, or *49 CFR 19 - Uniform Administrative Requirements for Grants and Agreements with Institutions of Higher Education, Hospitals, and Other Non-Profit Organizations, Section 19.23, Cost Sharing or Matching*. EWI captured and maintained an Excel spreadsheet with a running summary of additional in-kind cost-share contributions as reported by team participants. Budget and actual costs are shown in Table 43.

Table 43. Status of Additional In-Kind Support

Team Participant	Task	Budget ($)	Actual ($)
BP	1,3,5,6,7	TBD	110,000
ConocoPhillips	3	TBD	1,250
~~Heerema~~	~~3~~	~~TBD~~	~~0~~
Imperial Oil (via the MGP)	3	10,000	10,000
EMDC[§§] (51043CSP)	All	80,000	80,000
	Total	**90,000**	**201,250**

Small Issue: BP and ConocoPhillips representatives are very busy and have not had time to determine the dollar value for their in-kind activities; however, they have been very active in project activities and have each provided an estimated minimum of $10K in in-kind support. Overall the project team exceeded the budget for anticipated additional in-kind support.

BP is a leading U.S. pipeline/offshore operating company, who provided an in-kind contribution in the form of an NDE Expert/Welding Engineer in support of Tasks 1, 3, 5, 6, and 7; no value has been estimated for this contribution. Dr. Mark Lozev provided this in-kind contribution through his active participation in the project and estimated his contribution at $110,000+ over 3 years. **This in-kind cost-share contribution was provided in full.**

ConocoPhillips provided an in-kind contribution of technical/industry oversight in support of the program; no dollar value was estimated for this contribution. Joe Kiefer actively participated in the project. **This in-kind cost share contribution was provided, but no dollar value was assigned to it.**

[§§] New in-kind contribution commitment fully executed during the quarter ending December 31, 2008.

Heerema is a leading pipeline laying offshore contractor in the Gulf of Mexico, who committed to providing an in-kind contribution in the form of equipment leases and technical/industry oversight: no dollar value was estimated for this contribution. Heerema did not participate in the project; therefore, their contribution in Table 43 is listed in strike out text. **This cost share was removed.**

Imperial Oil is one of the largest producers of crude oil in Canada and the country's largest refiner and a leading marketer of petroleum products. Imperial Oil, via the Mackenzie Gas Project (MGP), provided **$10K** in line pipe for Task 3 weld sample fabrication. **This cost-share contribution was provided in full.**

EMDC is the world's largest publically traded international oil and gas company. In Quarter 6, EMDC agreed to provide a cost-share equivalent of **$80K** of direct in-kind support contribution by making available through loan one automated weld inspection system, including one AUT PA/eddy current array instrument and related equipment, which was located at its research center in Houston, TX. This system loan was tracked as EWI Project No. 51043CSP in Table 43. On November 21, 2008, EMDC fully executed agreement 51043CSP. The system was received in January 2009 and resided at EWI for the duration of the project for use in the stages of procedure optimization, procedure validation, operators training, performance determination, systems capabilities quantification, and such other uses as agreed by project participants. **This cost-share contribution was provided in full.**

9.2.3 Cost-Share Funding Summary

Table 44 contains a detailed list of actual cash and in-kind cost-share contributions for the project.

Table 44. Cost Share Funding Detail

Organization	Cash ($)	In-Kind ($)	Totals ($)
ConocoPhillips	200,000	1,250	201,250
UTTechnology and UTQuality	80,000	224,094	304,094
BP	0	110,000	110,000
GE Inspection Technologies	0	105,486	105,486
Mechanical Integrity	0	100,000	100,000
TCPL	0	100,000	100,000
EWI	179,115	0	179,115
PRCI	225,000	0	225,000
Petrobras	200,000	0	200,000
EMDC	22,983	80,000	102,983
Imperial Oil	0	10,000	10,000
Totals	**907,098**	**730,830**	**1,637,928**

10.0 References

1. Forli, O., "Qualification of AUT for Offshore Pipelaying and the Role of NDE Reliability", 3rd European-American Workshop on Reliability of NDE and Demining, Berlin, September 2002.

2. Morgan, L. P., Nolan, P., Kirkham, A., and Wilkinson, P., "The Use of Automated Ultrasonic Testing (AUT) in Pipeline Testing", *Insight*, Vol. 11, 2003.

3. Lozev, M., Hodgkinson, D., Spencer, R., and Grimmett, B., "Validation of Current Approaches for Girth Weld Defect Sizing Accuracy by Pulse-Echo, Time-of-Flight Diffraction, and Phased-Array Mechanized Ultrasonic Testing Methods", *Materials Evaluation*, Vol. 63, No.5 (2005).

4. Lozev, M., Spencer, R., Huang T. and Patel P., "Improved Inspection and Assessment Methods for Pipeline Girth Welds and Repair Welds", EWI Final Report, DOT Agreement No. DTRS56-03-T-0012 (February 2006).

5. Berens, Alan P., "NDE Reliability Data Analysis," *Metals Handbook*, 9th Edition, Vol. 17 (1989).

6. European Methodology for Qualification of Non-destructive Testing, Issue 3, European Network for Inspection Qualification (ENIQ) Report No. 31, EUR EN 22906, published by the European Commission, Brussels-Luxembourg, August 2007.

7. QA Programme for the ENIQ Pilot Study, European Network for Inspection Qualification (ENIQ) Report No. 8, EUR EN 18118, published by the European Commission, Brussels-Luxembourg, November 1998.

8. API 1104, Welding of Pipelines and Related Facilities, American Petroleum Institute (API).

9. ENIQ Recommended Practice 2: Recommended Contents for a Technical Justification, Issue 1, European Network for Inspection Qualification (ENIQ) Report No. 4, EUR EN 18099, published by the European Commission, Brussels-Luxembourg, July 1998.

10. Technical Justification Pre-Trials, European Network for Inspection Qualification (ENIQ) Report No. 10, EUR EN 18114, published by the European Commission, Brussels-Luxembourg, November 1998.

11. Offshore Standard OS-F101, Submarine Pipeline Systems, Appendix E: Automated Ultrasonic Girth Weld Testing, Det Norske Veritas, January 2000.

12. ENIQ Recommended Practice 6: The Use of Modeling in Inspection Qualification, Issue 1, European Network for Inspection Qualification (ENIQ) Report No. 15, EUR EN 19017, published by the European Commission, Brussels-Luxembourg, December 1999.

13. ASTM E 1961, Standard Practice for Mechanized Ultrasonic Examination of Girth Welds Using Zonal Discrimination with Focused Search Units, Section 3 – Metals Test Methods

and Analytical Procedures, Vol. 03.03 Nondestructive Testing, Annual Book of ASTM Standards

14. Qualification Procedure for the ENIQ Pilot Study, European Network for Inspection Qualification (ENIQ) Report No. 9, EUR EN 18117, published by the European Commission, Brussels-Luxembourg, November 1998.

15. ENIQ Recommended Practice 5, Guidelines for the Design of Test Pieces and Conduct of Test Piece Trials, European Network for Inspection Qualification (ENIQ) Report No. 14, EUR EN 18686, published by the European Commission, Brussels-Luxembourg, February 1999.

16. 2007 ASME Boiler & Pressure Vessel Code, Section XI – Rules for Inservice Inspection of Nuclear Power Plant Components, Appendix VIII – Performance Demonstration for Ultrasonic Examination Systems, The American Society of Mechanical Engineers, July 1, 2007, 2007 Edition.

17. MIL-HDBK-1823 (2007), Nondestructive Evaluation System Reliability Assessment, Department of Defense Handbook.

18. Berens, A., "Probability of Detection (POD) Analysis for the Advanced Retirement for Cause (RFC)/Engine Structural Integrity Program (ENSIP) Nondestructive Evaluation (NDE) System Development. Vol. 1 POD Analysis", Report AFRL-ML-WP-TR-2001-4010, Air Force Materiel Command, Wright-Paterson Air Force Base, OH 45433-7750.

19. Savadjiev, P., Ferrie, F. P., and Siddiqi, K., "Surface Recovery from 3D Point Data Using a Combined Parametric and Geometric Flow Approach", *Proceedings Energy Minimization Methods in Computer Vision and Pattern Recognition*, McGill University, Quebec, Canada, 2003.

20. Amenta, N. and Kil, Y. J., "The Domain of a Point Set Surface", *Proceedings Eurographics Symposium on Point-Based Graphics*, University of California at Davis, 2004.

21. Pauly, M., Gross, M., and Kobbelt, L. P., "Efficient Simplification of Point-Sampled Surfaces", *Proceedings of the Conference on Visualization*, ETH Zurich, pp. 163-170, ISBN: 0-7803-7498-3, 2002.

22. Lee, I. and Schenk, T., "3D Perceptual Organization of Laser Altimetry Data", *Proceedings Int'l Archives of Photogrammetry and Remote Sensing*, Vol. XXXIV-3W4, The Ohio State University, October 2001.

23. Garland, M. and Heckbert, P. S., "Surface Simplification Using Quadric Error Metrics, *Proceedings SIGGRAPH '97*, Carnegie Mellon University.

24. Lozev, M, Spencer, R., Huang, T., and Patel, P., "*Improved Ultrasonic Inspection and Assessment Methods for Pipeline Girth Welds and Repair Welds*", EWI Final Report 46997GTH, DOT Agreement No. DTRS56-03-T-0012, February 2006.

25. Lozev, M., "Define, Optimize and Validate Detection and Sizing Capabilities of Phased-Array Ultrasonic Technique to Inspect Joints in Polyethylene Pipes", DOT Agreement DTRS56-06-T-002.

26. Rummel. W. D., Todd Jr., P. H., Frecska, S. A., and Rathke, R. A., “The Detection of Fatigue Cracks by Nondestructive Testing Methods”, CR-2369, National Aeronautics and Space Administration, February 1974.

27. Rummel. W. D., “Recommended Practice for a Demonstration of Nondestructive Evaluation (NDE) Reliability on Aircraft Production Parts”, *Material Evaluation*, Vol. 40, No. 9, pp. 922-932, 1982.

28. Berens, A. P., “NDE Reliability Data Analysis”, *ASM Metals Handbook*, Vol. 17, Nondestructive Evaluation and Quality Control, 9-th Ed., American Society of Metals International, Metals Park, OH, 1989.

29. A Recommended Methodology for Quantifying NDE/NDI Based on Aircraft Engine Experience, AGARD Lecture Series 190, Advisory Group for Aerospace Research & Development (AGARD), North Atlantic Treaty Organization (NATO), 1993.

30. MIL-HDBK-1823A, Nondestructive Evaluation System Reliability Assessment, Department of Defense Handbook, 2009.

31. Herberich, J., "Applying MIL-HDBK-1823 for POD Demonstration on a Fluorescent Penetrant System", *Materials Evaluation*, Vol. 67, No. 3, March 2009.

32. Singh, R., “Three Decades of NDI Reliability Assessment”, Report No. Karta-3510-99-01, Submitted to AF NDI Office Contract F41608-99-C-0404, Karta Technologies, Inc., May 2000.

33. 2007 ASME Boiler & Pressure Vessel Code (BPVC), Section V – Nondestructive Examination, Article 14 – Examination System Qualification, The American Society of Mechanical Engineers, July 1, 2007, 2007 Edition. Addendum 2008a, BC03-1552.

34. Offshore Standard DNV-OS-F101, Submarine Pipeline Systems, Appendix E: Automated Ultrasonic Girth Weld Testing, Det Norske Veritas, October 2007.

35. Nordtest Report NT TECHN REPORT 394, Approved 1998-04.

36. Visser, W., “POD/POS Curves for Non-Destructive Examination”, Offshore Technology Report 2000/018, Prepared by Visser Consultancy Ltd. For the Health and Safety Executive, 2002.

37. Georgiou, G. A., “Probability of Detection (POD) Curves: Derivation, Applications and Limitations”, Research Report 454, Prepared by Jacobi Consulting Ltd. For the Health and Safety Executive, 2006.

38. Matzkanin, G. A. and Rummel, W. R., “NDE Capabilities Data Book”, NTIAC-DB-97-02, 1997.

39. O’Donnell, J. R., Buitrago, J., and Hoyt, D. S., “A Methodology to Verify Inspection Reliability of Fracture-Critical Girth Welds”, Deep Offshore Technology Conference, Vitoria, Brazil, November 2005

40. Forli, O., “How to Develop Acceptance Criteria for Pipeline Girth Weld Defects?”, European-American Workshop Determination of Reliability and Validation Methods of NDE, Berlin, Germany, June 1997.

41. Lozev, M., Hodgkinson, D., Spencer, R., and Grimmett, B., “Validation of Current Approaches for Girth Weld Defect Sizing Accuracy by Pulse-Echo, Time-of-Flight Diffraction, and Phased-Array Mechanized Ultrasonic Testing Methods”, *Materials Evaluation*, Vol. 63, No.5, 2005.

42. ANSI/NCSL Z540-2-1997 (R2007) “U.S. Guide to the Expression of Uncertainty in Measurement”, American National Standard for Calibration, NCSL International.

43. Taylor B. N. and Kuyatt, C. E., “NIST Technical Note 1297, 1994 Edition, Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results”, United States Department of Commerce, Technology Administration, National Institute of Standards and Technology (NIST).

44. Guide 98-3:2008, “Guide to the Expression of Uncertainty in Measurement (GUM)”, International Organization for Standardization (ISO), International Electrotechnical Commission (IEC).

45. Ginzel, E., “Automated Ultrasonic Testing for Pipeline Girth Welds: A Handbook”, Olympus NDT Canada, 2005.

46. Box, G.E.P., Hunter, J. S., and Hunter, W. G., “Statistics for Experimenters: Design, Innovation, and Discovery”, Second Edition, 2005, John Wiley (New York, N. Y.).

47. Minitab Statistical Software, Minitab Help.

48. Berens, A. P. and Hovey, P. W., AFWAL-TR-81-4160, Vol. 1, “Evaluation of NDE Reliability Characterization”, University of Dayton Research Institute, Materials Laboratory, Wright-Paterson AFB, 1981.

49. SNT-TC-1A (ASNT Recommended Practice), “Personnel Qualification and Certification in Nondestructive Testing.