A Qualification Process for Phased-Array UT using DNV RP-F118 Guidelines

Edward A. GINZEL 1, Ryan THOMSON 2, Robert K. GINZEL 2

1 Materials Research Institute, Waterloo; Canada
Phone: +519 886 5071 e-mail: eginzel@mri.on.ca
2 Eclipse Scientific Products; Waterloo, Canada

Abstract
A phased array UT procedure was qualified to permit fracture-mechanics-based acceptance criteria to be used on a specific Tee shaped weld. The process used the DNV RP-F118 [1] as a guideline for the detection and sizing assessments required. The qualification process is described and the results are summarised along with modelling confirmation examples.

Keywords: Keywords: Phased-array UT, Qualification, PoD, Sizing, modelling

1. Introduction

In Europe non-destructive testing is generally done in accordance with the European norms. For ultrasonic testing this is usually in accordance with EN-1714 [2] and using the acceptance criteria in EN-1712 [3]. These norms are like many others around the world in that they were written with manual UT in mind. Provisions in EN-1714 do exist that can allow the use of automated UT equipment with agreement between the contracting parties.

Advanced UT methods, Time of Flight Diffraction (TOFD) and phased-array pulse-echo ultrasonic-testing techniques (PAUT), can offer several advantages over manual techniques. These have been mentioned by others in previous years;

- A permanent record is made
- The initial file data record is auditable
- A high degree of repeatability exists so the initial record can be used as a baseline for in-service inspections
- Speed of inspection is generally faster than manual UT so the initial equipment costs can soon be offset with the increased rate of production

Another advantage of automated systems is their ability to use the inspection results with fitness-for-purpose acceptance criteria. This concept has long been part of pipeline construction on the inspections of pipeline girth weld. But the same principles can be applied to other applications.

Fracture engineers now generally use the document BS-7910 [4] to determine the allowable flaws size in welded structures. This requires determination of the metallurgical properties, the stresses the component may undergo, the environment of the weld when in service and knowledge of the location and size of any flaws introduced during fabrication.

For simple butt welds the use of traditional “workmanship” acceptance criteria may often be adequate. However, in more complex geometries there can be concern for unnecessary repairs being made. When simple symmetry does not exist in a component, the repair process can introduce significant stresses on the surrounding material when weld metal is removed to effect the repair. A risk exists that, in the process of repairing a flaw, the situation could be made worse than leaving the flaw in place.
Parameters required in developing fitness for purpose acceptance criteria were mentioned above. Material properties are easily enough identified in a materials lab and can be obtained from the welding qualification samples. Environmental stresses during service are already available in most engineered projects as they comprise part of the design calculations. There remains only the detection and sizing capability of the inspection system to be determined. Assessment of NDT procedure effectiveness is now a fairly common expectation in some industries. Most such methodologies are found in pressure vessel applications (ASME Sec. V Art 14, ASME Sec III Appendix VIII, ASME Sec VIII CC2235 [5], ENIQ [6]).

This paper examines the unique aspects of a Tee weld that was to be assessed. Since the components to be examined were structural, as opposed to pressure vessel, meant that there were no codified methods to qualify the detection and sizing capabilities of the NDT procedure. The applicable Euro-code for the fracture mechanics computations is BS 7910 but this does not provide an applicable method to assess the NDT procedure. It was therefore agreed to use a method adapted from the new DNV RP F-118 that provides a detailed method of procedure qualification for pipeline girth weld inspections.

2. The Qualification Process

DNV RP118 outlines the following main components for a procedure qualification:

A qualification programme for any AUT system shall comprise the following stages:

1. review of the technical documentation of the AUT system
2. review of the operating methodology for the AUT system
3. review of the Quality Assurance system for development, verification, maintenance and operation of the AUT system
4. review of available performance data for the AUT system (detection abilities and defect sizing accuracy)
5. evaluations and conclusions based on the information made available
6. identification and evaluation of significant parameters and their variability
7. planning and execution of a repeatability test programme
8. planning and execution of a reliability test programme
9. documentation of results from the repeatability and reliability test programmes
10. supplementary NDT and destructive testing
11. analysis of data from the deriving from the above activities. Establish PoD and sizing accuracy.

The proposed qualification programme was submitted to DNV and the details agreed upon. After the programme details were agreed upon the flaw samples and calibration blocks were prepared. It was agreed that although for surface indications EDM fabricated flaws would be acceptable, all embedded flaws would need to be made using a more “natural” process such as welding parameter adjustments and welded-in flaws as opposed to ceramic inserts.

Scanning restrictions existed in the field conditions so low-profile probes and extended probe holders on the magnetic-wheeled scanner were designed. Probe placement on the weld was modelled by ray-tracing as illustrated in Figure 1.
Figure 1 illustrates that there is some redundancy in the process. Two S-scans are used with approximately 90% of the weld being inspected with at least 2 angles passing through each voxel (volume element). In addition, each probe is used with 2 groups of focal laws. One group is devoted to the S-scan while the other is a single focal law generated with all the elements forming a 0° compression pulse that serves as a coupling check channel for each probe.

Figure 2 illustrates preparation of the 75cm length sections of welded Tee component.
Qualification scanning consisted of multiple passes on 8 mock-ups, each 75cm long. In the field only a single side of the plate would be accessible for the test. However, because the weld samples used for the mock-ups were symmetrical there was opportunity to scan the plate from both sides.

Scans carried out for the qualification included:
- Scan from Side 1 at reference sensitivity
- Scan from Side 1 at reference sensitivity plus 6dB
- Scan from Side 2 at reference sensitivity
- Scan from Side 2 at reference sensitivity plus 6dB
- Repeating scans from both sides at both sensitivities

Scanning from both sides of the plate provided an opportunity to obtain increased statistical sampling. Since the flaws were being approached from completely different angles the amplitude data would be different from side to side for the same flaw. Therefore, a single flaw could provide two data points for the detection and sizing statistical assessments.

To indicate repeatability of the scanning system, specimens were scanned twice and some a third time. Using the results of multiple scans made from the same surfaces allowed comparison of the flaw amplitude responses and assessed lengths. This provided opportunity to illustrate the repeatability of the system.

3. Data Analysis

Data analysis was carried out to identify a table of flaw locations where macro images were to be made. Flaw positions were marked on the specimens based on the ultrasonic responses and the samples shipped to an independent metallurgical laboratory for sectioning and macro
EDM flaws were replicated using a rubber replicating process and macros of the replicas were used to assess the flaw depths and orientations. Seventy two surface flaws and 46 subsurface flaws were identified for the statistical assessments. Salami cuts (i.e. sequential sectioning of a single flaw separated by 2mm along the weld) were made on all welding flaws selected. Over 300 macro-photo images were made for the assessments.

Analysis of the UT data involved identifying all flaws over an amplitude threshold relative to reference amplitude. These flaws were then assessed for maximum amplitude, start and end locations, side from which they were tested, and assessed depth from the test surface. In addition, algorithms were established to estimate flaw height.

Once the metallurgical lab had completed the replicas and polishing, dimensions from the macro photos could be used to complete the tabulated information for the flaws identified.

Examples of the embedded and surface macros are seen in Figure 3 and 4. The corner and tip signals are seen overlaid on one of the surface interacting flaws in Figure 5.

![Figure 3](image1.png)

**Figure 3**  
Example macro image and dimensions used for PoD analysis (embedded flaw)

![Figure 4](image2.png)

**Figure 4**  
Example macro image and dimensions used for PoD analysis (surface flaw)  
EDM rubber replica showing length profile and cross-section depth and angle profile
Size “estimates” are a critical aspect of the development of fracture-mechanics based acceptance criteria. In some codes such as ASME Code Case 2235-9 and Code Case 181 where a Tier 2 acceptance criteria is provided based on generic materials properties, no assessment of UT sizing performance is required. Detection in the ASME case is merely “demonstrated” using 3 or 4 “typical” flaws being detected over the evaluation amplitude. DNV RP-F118 requires sufficient statistical evidence of the system to demonstrate detection. 3-4 examples are not adequate to demonstrate confidence in any statistical sample. A minimum of 29 samples is required to have statistical “confidence” that flaws of a critical size will be detected (e.g. PoD of 90% with 95% confidence). RP F-118 requires significantly more than the minimum 29 samples. For double V submerged arc welds, RP F-118 recommends a minimum of 91 samples be used.

Since the PAUT system will also be used to size flaws that are “detected” the system must be assessed on how accurately it sizes flaw length and height. Sizing in this qualification used a combination of techniques for vertical sizing, depending on the data available to the operator. Tip echo assessments were the preferred vertical sizing tool, but most of the flaws in SAW processes are volumetric-like (slag inclusions) so a version of “MaxAmp” was developed for embedded flaws with that had no visible tip signals. Surface flaws with no tip echoes were sized using a version of ring-time assessment. Length sizing used the 6dB drop method.

Analysis of the data collected provides quantitative values for these critical parameters. Once the destructive test results (macro photo evidence) were available and tabulated, a statistical analysis could be carried out. In such an analysis macro results are considered the “Truth” value against which the UT estimates are compared.

Analyses generated for the qualification included:
- Repeatability comparison
- Amplitude versus height responses
- Sizing comparison between macro and AUT estimates

These data allowed the following quantitative assessment;
Repeatability
Repeat scan variations over 100 data points.
- Average Length difference: 3.7 mm
- Average Amplitude difference in % FSH: 2.5%
- Average Amplitude difference in dB: 0.002
- Average Depth difference: 1.4 mm
- Average height Estimate difference: 0.7 mm

Amplitude versus Height
By now it is well known that, for real weld flaws, there is no linear relationship between flaw size and echo amplitude. A plot of amplitude versus flaw height illustrates this fact. But this type of plot is critical to understanding the amplitude level needed to ensure detection of critical flaw sizes for evaluation. Figure 6 shows some larger flaws can have relatively small amplitude response while some small flaws provide disproportionately large amplitudes.

Figure 6 Scatter plot of Amplitude versus Flaw Height
**Probability of Detection**

PoD is the indicator of the effectiveness of the procedure and technique to ensure that flaws of a critical size are detected. Several methods can be used but DNV uses the guidelines outlined in Nordtest NT Techn. Report 394 (Guidelines for NDE Reliability Determination and Description, Approved 1998-04).

Using the Nordtest methods, the data collected in the qualification process generated the PoD for the combined surface and subsurface flaw detections seen in Figure 7.

![POD vs Flaw Height](image)

**Figure 7**  
*PoD for Qualification Project Data*

Using the combined detection performance of the system for embedded and surface flaws at 20% screen height, the 90% PoD at 95% confidence level using the MLE (maximum likelihood estimator) option was obtained for flaw height value of 1.16mm.

The advantage of using PoD data is that an amplitude threshold can be established based on assessing the system performance at different levels. Re-running the PoD using a 25% threshold for evaluation indicates that the 90% PoD at 95% confidence level could be achieved for a flaw of 1.8mm. Raising the evaluation threshold could provide adequate assurance of detection of flaws considered to be of critical size if the critical size was assessed as 1.8mm.
**Sizing accuracy**

Sizing error can be derived in several ways. The least useful method of assessing error is by the average. This may provide a trend towards over-sizing or under-sizing. But the scatter of error is more critical for fracture mechanics. RMS error and standard deviation of error are two of the more common options (ASME quotes RMS in Appendix VIII of Section XI and other standards (e.g. old BS PD 6493) use standard deviation).

**Length**

Flaw lengths were established using the EDM notches as they provided a length for every sample. Several salami cuts were made to confirm sizing of embedded flaws. These were very accurate (generally within 1mm of the actual ends).

- Average Length Sizing error: 1.2mm oversizing
- Standard deviation of length size error: 2.97mm

**Height**

A summary of the relevant sample statistics and flaw dimensions in the combined samples are tabulated below.

<table>
<thead>
<tr>
<th>Result Statistics – Combined Embedded &amp; EDM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Samples Used:</td>
</tr>
<tr>
<td>Avg Size of PAUT Flaw Height: (mm)</td>
</tr>
<tr>
<td>Avg Size of DT Flaw Height: (mm)</td>
</tr>
<tr>
<td>Max AUT Flaw Height: (mm)</td>
</tr>
<tr>
<td>Max DT Flaw Height: (mm)</td>
</tr>
<tr>
<td>Min AUT Flaw Height: (mm)</td>
</tr>
<tr>
<td>Min DT Flaw Height: (mm)</td>
</tr>
<tr>
<td>Standard Dev of PAUT Flaw Height: (mm)</td>
</tr>
<tr>
<td>Standard Dev of DT Flaw Height: (mm)</td>
</tr>
<tr>
<td>Avg Vertical Sizing Error: (mm)</td>
</tr>
<tr>
<td>RMS Error in Height Sizing: (mm)</td>
</tr>
<tr>
<td>Standard Deviation of Height Error (mm)</td>
</tr>
</tbody>
</table>

Table 1 Flaw sample statistical summary

The table illustrates the importance of average error versus Standard Deviation of error. The average error is almost zero (0.2mm). Whereas the scatter of sizing shows that the uncertainty of the sizing is better indicated by the standard deviation or RMS values (1.7mm). Tolerance for sizing error would be added to the computations for allowed flaw lengths by the engineers adding the tolerance to the flaw heights in their tables when developing the acceptance criteria.

**Modelling Confirmation**

Automated UT scanning systems typically incorporate a means for confirming coverage. As well it is typically used with a system that can carry out periodic checks on sensitivity without the need to break down the probe components to verify this on the traditional SDH calibration blocks.
A method for such dynamic calibration was developed for the Tee weld inspections. This involved a series of targets in a Tee section that can be used with run-off plates. The design used was configured using the CIVA modelling programme. Scans of the plate allowed the operator to verify probe placement and system response continuity from one weld inspection to the next. Figure 8 illustrates the block designed in CIVA and the target locations.

Figure 8  Dynamic calibration block designed using CIVA

Responses to the targets can be compared to measured distances to the various targets in a merged scan. The poorly oriented small flat bottom holes provide only weak tip diffracted signals less than 20% screen height at reference plus 6dB. Figure 9 illustrates the target detection using merged data projected in the vertical plane (lower) compared to the modelled positions (upper).

Figure 9  Comparing calibration block designed positions to scanned positioning
Civa modelling predicts the poor responses obtained. Figure 10 indicates the predicted response from the hemispherical target at 14dB below reference whereas the measured response is 16dB below reference.

Figure 10  Comparing Civa predicted response to actual response

SDHs are the same 3mm diameter used for setting the reference sensitivity. However, because some of the focal laws detecting the SDHs in the dynamic calibration scan do so after a skip off a portion of the weld fillet surface, the response is not expected to be identical to that established on the TCG block.
Conclusions

Results of the analysis indicate several conclusions may be made;

- The procedure’s level of sensitivity provides a good PoD (flaws 1.6mm and greater are detected 90% of the time with 95% confidence)
- The sizing estimate provides a reasonable RMS error or Standard Deviation of Error which can be used to provide provision for uncertainty when developing the acceptance criteria.
- Repeatability of the scanning system is effective for length and detection of flaws with some variation expected due to small differences that could occur due to placement of the scanner at the start reference location. Good repeatability is necessary if the initial scan is to be compared with subsequent scans for in-service monitoring.
- The qualification process gives the opportunity and confidence to make decisions on how to treat assessed indications (i.e. repair or monitor using a periodic inspection programme).
- The process defined in DNV RP-118 also provides an effective method to assess AUT system performance on applications other than pipeline girth welds.

Acknowledgements

We would like to thank Pål Erik Gjerde at DNV (Stavanger, Norway) who was very helpful in preparing this programme of qualification. We also would like to thank Håkon Stokka-Hasting (DNV Stavanger) for his review comments and help with the statistical assessments.

References

1. Pipe Girth Weld AUT System Qualification and Project Specific Procedure Validation, Recommended Practice DNV RP F118, 2010
2. BS EN 1714 Non-destructive testing of welds - Ultrasonic testing of welded joints, 1998
3. BS EN 1712 Non-destructive testing of welds Ultrasonic testing of welded joints Acceptance levels
5. ASME Boiler and Pressure Vessel Code, American Society for Mechanical Engineers