Total Focusing Method used for flaw sizing in probability of detection determination

E. Ginzel $^1$, M.E. Bajgholi $^2$, M. Burrowes M. Guimarães $^3$, F. Foucher $^4$, Gilles Rousseau $^5$, Martin Viens $^2$

$^1$Materials Research Institute, Waterloo, Ontario, Canada  
e-mail: eginzel@mri.on.ca  
$^2$École de Technologie Supérieure, Montréal, QC H3C 1K3, Canada  
e-mail: mohammadebrahim.bajgholi.1@ens.etsmtl.ca, martin.viens@etsmtl.ca  
$^3$Federal University of Rio de Janeiro, Brazil  
e-mail: marianaburrowes@gmail.com  
$^4$Extende S.A, Massy, France  
e-mail: fabrice.foucher@extende.com  
$^5$Institut de Recherche d’Hydro-Québec (IREQ), Varennes, QC J3X 1S1, Canada.  
e-mail: rousseau.gillesa@hydroquebec.com (G.R.)  

2022.06.25

Abstract

Probability of Detection (POD) is used as an indicator of an inspection procedure’s reliability to detect flaws of concern. In non-destructive testing (NDT), POD is usually described using a cumulative distribution function that plots the random size variable on the horizontal axis against the probability or likelihood that a flaw of a given size will be detected. Standards describing the use of POD assume that the flaw size is known and no provision is made for sizing error. Test coupons made for inspection-procedure qualification using naturally occurring flaws (e.g., varying welding parameters to induce flaws) generally make unpredictably sized flaws; however, even intentionally fabricated flaws are often not the size and location that the manufacturer intended or documented them to be. This is particularly problematic for subsurface flaws that cannot be verified visually. NDT methods can be used to verify flaw sizes; however, this seems to be something of a conflict of interest when qualifying an inspection procedure by using another inspection method; e.g., verifying an ultrasonic weld inspection procedure by locating and sizing the same flaws with radiography. Such an approach seems to suggest that the verifying method (e.g., radiography) is trusted more than the method being qualified (e.g., ultrasound). When the purpose of the qualification is to establish reliability by POD, the apparent conflict can be lessened if the secondary NDT method is used solely for sizing. Since POD is highly dependent on accurately identifying a flaw size, there must be confidence in the sizing method. This paper considers the use of Total Focussing Method (TFM) as a tool to provide accurate flaw sizing for POD determinations.

Keywords: NDT, POD, TFM, sizing, reliability, flaws
1. Background

Several codes and standards [1,2,3,4,5,6] are available that provide information on the use of probability of detection (POD) for the purpose of assessing an inspection procedure’s reliability. The process of establishing the POD involves:

1. preparing flawed and unflawed samples,
2. applying the inspection procedure,
3. documenting response data of the inspection procedure to the flaws in the samples
4. applying statistical algorithms to assess the reliability of the inspection procedure

In most cases, the statistical algorithms will provide a probability of detecting a flaw of a certain size. POD information is often used in conjunction with fracture mechanics calculations that determine a critical flaw size for a component when in service. This is the basis for the concept of “Fitness for Purpose” or “Fitness for Service”. Over the past few decades, several standards relating to the application of POD have been written. These include:

- ASME Section V Art. 14
- ASTM E2862
- ASTM E3023
- DNVGL ST-F101
- ENIQ Recommended Practice #5
- MIL-HDBK 1823A

Typically, POD is represented as a curve of cumulative probabilities with probability increasing as the size of the flaw increases. Flaw size is plotted on the abscissa and probability on the ordinate axis. This implies that the known value is the flaw size and the calculated value is the probability.

In spite of the fact that flaw size is considered the known, or controlled parameter; how the flaw size (length or height) is known is, in most cases, not addressed in the standards used to derive POD.

Several approaches are used with respect to when fabricating flaws for POD applications:

- naturally made by adjusting the normal fabrication processes
- inserting shims, ceramic (e.g., spheres) or foil to prevent fusing of metal
- fabricating a flaw such as a fatigue crack in a block then welding it into the sample
- mill-cutting or EDM (electro-discharge machining)

When made naturally, the exact flaw size is completely unknown. However, even when fabricated, significant uncertainty may exist. Shims can be melted during the welding-in process, foils can be folded or tear, fabricated flaws in separate blocks of metal can have flaws or grain-structure changes occur as a “cloud” around the welded-in flaw and of course milled or EDM slots can be melted as subsequent molten metal is deposited over them or the electrode can deteriorate as the discharge process proceeds.

With the exception of the DNVGL ST-F101, all of the standards identified above assume that the manufacturer's stated or intended flaw size is correct. Appendix E of DNVGL-ST-F101 describes their version of the only acceptable method for determining the Truth value for the flaw size of interest. It must be by salami cuts, with polished surfaces and then macro photographs taken to determine true flaw size.

---

1 See also ENIQ Report #41 Probability of Detection Curves: Statistical Best-Practices
From DNVGL-ST-F101

E.8.9.1 The reports from the AUT qualification testing shall be validated for accuracy in the determination of imperfection circumferential position, length, height and depth by reference destructive testing.

E.8.9.2 The testing shall be by cross-sectioning, preferably by the salami method, by making more cross sections around each location chosen. The imperfections as reported in the AUT reports shall be used when selecting the areas for cross-sectioning.

The other standards make no provision for the fact that errors in sizing may, and often do, exist.

ASME V Art. 14 states to use a “flaw of interest” in the component to be examined. The flaws are their locations, threshold detection size, and the critical size (that is, the fracture-mechanics determined critical size), orientation and shape determined. These details are to be used as a guide for developing the procedure. But at no point in ASME V does it instruct how to determine these details!

In ASME V Art 14 (Examination System Qualification) T-1442 it states:

All recordable indications shall be sized and located. The detection records shall note whether indications are located correctly.

Codes and standards imply that we are just supposed to “know” the details without knowing HOW to know the details.

Flaws used in qualifications using MIL-HDBK 1823A, usually use machined flaws and the size is not in question (with perhaps the exception of sub-surface flaws).

According to MIL-HDBK 1823A:

1.1 Departures from underlying assumptions – crack sizing and POD analysis of images

The software that accompanies this handbook, mh1823 POD, assumes that the input data is correct. That is, if the size is \( X \), then that is the true size. If the response is \( Y \), then that is the true response. In most situations this is reasonable. There are situations when this assumption does not hold and more advanced methods are needed.

a. Errors in \( X \) – Circumstances where target sizing is only approximate,

b. Errors in \( Y \) – Situations where the response cannot be easily categorized as either an amplitude, \( \hat{a} \), or a binary outcome, hit/miss, such as inspections that produce images.

ASTM E2862 and E3023 similarly assume that the documented flaw sizes are as stated:

E2862 states:

6.1.1 The analyst shall obtain the hit/miss data resulting from the POD examination, which shall include at a minimum, the documented known induced discontinuity sizes, whether or not the discontinuity was found, and any false calls.

E3023 states:

5.4 Prior to performing the analysis it is assumed that the discontinuity of interest is clearly defined; the number and distribution of induced discontinuity sizes in the POD specimen set is known and well documented; the POD examination administration procedure (including data collection method) is well designed, well defined, under control, and unbiased; the initial inspection system response is measurable and continuous in nature; the inspection system is calibrated; and the measurement error has been evaluated and deemed acceptable.

Although the ENIQ Recommended Practice 5 does not explain in any detail how to determine accurate flaw size Truth value, it has made provision in a more general way to “check” the flaws.

From ENIQ Recommended Practice 5:
4.7 Test piece quality checks

Test piece quality checks have to be carried out in principle by personnel of the qualification body before qualification starts. This may involve the following: provide detailed fabrication instructions to the test piece manufacturer, follow in detail the work done by the test piece manufacturer, review the documentation provided by the test piece manufacturer, etc. The test pieces should be examined using X-rays and ultrasonics and/or any other NDT method judged useful to ensure that the defects are as intended. It is also important to check that the volume around each defect does not contain significant indications that would make the defect unusable for qualification. An assessment of the accuracy of the defect sizes used as reference in order to assess the inspection results to be obtained during the practical trials, should be made available. Including two nominally identical defects and getting comparable results can be another way of ensuring test piece quality.

ENIQ alone has provided a statement that leaves an opening to use other NDT methods to produce a documented flaw size.

Perhaps the MIL-HDBK does not address it because they consider the X parameter not very significant in the end results; however, they too have provision for options stating, “There are situations when this assumption does not hold and more advanced methods are needed.”

Since DNV prefers to use naturally induced flaws, the idea of cutting the samples open and sectioning them seems to make good sense and is typically considered the “Gold Standard” for flaw size determination. But even salami cutting can have a degree of uncertainty. If the cuts are made at 2mm intervals, there is a chance that the maximum extent of a flaw lies between two slices, or it may be possible that the true extent of the flaw is so fine that it could only be seen in the grain structure of a micro-graph photo. In both of these cases, the flaw might be slightly larger than documented.

A recent tool that has the potential to be even more accurate than destructive testing (salami cuts) is micro-computed tomography X-ray imaging (micro-CT). Micro-CT can provide flaw sizing with better resolution than the destructive salami test. Micro-CT, for example, has the measurement certainty approximately 100 microns, which is the length of the voxel (considering linear measurement, such as crack length or height).

2. Uncertainties

In many cases the documented height size provided by flaw manufacturers has a tolerance allowed of +/-2mm (+/-0.080”). Yet when the critical flaw size as determined by fracture mechanics is calculated, this often indicates a flaw on the order of 3mm in height for many projects involving pressure vessels and high-pressure piping. POD analyses should then use coupons with a scattering of flaw sizes both smaller and greater than the critical flaw size (e.g., typically half to double the critical flaw size). However, with uncertainties of the flaw provided by the manufacturer, the POD uncertainty could be significant. Even the sizing by salami cuts and macro-photography is considered to have a tolerance (based on the possibility

---

2 Anecdotally, one of the authors has purchased flaw samples from reputable companies only to find that the documentation is simply wrong. E.g., A pipe weld was documented to have incomplete penetration at a specific location but it was actually over 100 mm away from the reported location. On another sample, the intended flaw was significantly altered by being melted to a much smaller height than indicated in the report. In that case, the documented flaw height of 9 mm could never be seen to be more than 4mm as determined when using 10 MHz TOFD and pulse-echo tip-diffraction. It should be noted that a 9mm high flaw would usually put it at the far-right side of any POD curve whereas a 4mm high flaw would typically be close to the centre of a typical flaw size range on the horizontal axis. Even EDM notches are not always the depths indicated by the manufacturer. E.g., an EDM notch that was documented to be 15x1.5 is seen by the rubber-replica to be a maximum of 1.3mm deep and that for less than a third of its length.
that the slice missed the actual maximum or that the flaw exists to a greater extent in the grains at the ends of the flaw).

Compared to destructive testing, micro-CT-X-ray could provide improved tolerances, but there are limits to both these technologies. Destructive testing requires that the samples be destroyed to determine the true size of flaws and a micro-CT-X-ray cabinet may not be large enough to contain the sample, so both options can be costly.

Since there are uncertainties in flaw sizing, even when carefully fabricated flaws are used, it must be accepted that the true sizes used in PODs will not be absolutely precise; however, effort should be made to minimise the error to something less than the full scale of the POD sizing range which might occur if using the tolerances provided by some flaw manufacturers.

British Standard 7910-2019 [7] has added an informative annex on Guidance on the use of NDT with ECA (engineering critical assessment). It clearly discriminates “detection” and “sizing” as two separate processes. Tables T1 to T3 in that annex provide guidance on the flaw sizing estimates that can be estimated (in fine-grained steel) for various NDT methods. Focussed phased-array, zonal ultrasonic testing and TOFD are considered the best options for length and height determination at ±1.5mm for height and ±7mm for length tolerance. When considering the Total Focussing Method (TFM) as a tool for flaw sizing for POD purposes, the expectations could be even better than these values.

3. Examples of TFM Sizing and Accuracy

The ultrasonic Total Focussing Method (TFM) has been demonstrated to provide reasonable accuracy in many cases [8,9,10,11,12,13,14,15,16,17,18,19]. When the NDT test method is being evaluated for reliability (and the test method being evaluated is not TFM \(^3\)), it may be feasible to use a TFM-determined flaw size as the truth value for POD purposes. One of the first to propose TFM as the basis of flaw sizing for POD was Bajgholi [20 &21].

Several studies have been carried out suggesting that the flaw size estimated by TFM is accurate.

- Rachev [8] sized notches over a range of 1mm to 8 mm in height. By suitable selection of aperture (larger was better) and using a Plane Wave Imaging (PWI) acquisition, they achieved better than 0.5 mm accuracy for all but the largest notch.

- Volf [11] used TFM imaging to monitor crack growth in a fatigue test and compared the TFM depth of the crack to the values indicated by the clip gauge. During the test TFM appeared to be oversizing the crack by 1-2 mm compared to the clip gauge. However, after the test was completed and the sample opened to expose the crack profile along the length, the final depth estimate from TFM was seen to be within 0.06 mm of the physical measurement. What appeared to be overestimation was in fact due to a curved crack front where the tip of the crack had propagated deeper that was indicated by the clip gauge.

- Peng [12] ran a similar test of TFM on a fatigue crack in aluminium. They too found an apparent error between the optical image of the crack on the surface and the TFM image was attributable to the curved crack profile. This test demonstrated sizing errors less than 11% for surface breaking fatigue cracks 1.95 to 2.81 λ at 5 MHz (i.e., 2.50–3.6 mm) and 2.81–7.34 λ at 10 MHz (i.e., 1.80 – 4.70 mm). 11% of 3.6 mm for the 5MHz condition would be 0.39 mm and 11% of 4.7 mm for the 10MHz condition is 0.52 mm.

\(^3\) It would be philosophically inappropriate or unethical to use the sizing performance of a test method to validate itself.
• Holloway [9,10] implemented the summation of TFM images using a commercialised software package. This was used to overcome the limitations identified by Sy [18] where certain combinations of modes imaged flaws better depending on the flaw location and orientation. Using this technique, the standard deviation of sizing error on the real crack was 0.22 mm and a maximum oversizing error of 0.6 mm.

• In Bajgholi’s [22] work on assessing flaws for fitness-for-service of turbine runners, they carried out sizing assessments of side-drilled holes (SDH) using TFM. Depending on the probe used, they obtained mean sizing errors of 0.88 for probes focussed in the passive axis and 0.55 mm for unfocussed probes.

4. Developing Guidance on TFM Sizing

Although TFM has been suggested as an option for detection of flaws, the acquisition of data from phased-array FMC (Full Matrix Capture) typically results in weaker signals than that provided by focussed pulse-echo phased-array.

In TFM analysis, both sensitivity and accuracy of sizing are functions of the set-up and the nature of the flaws detected. Probe positioning relative to the inspection volume, probe aperture, probe frequency, TFM reconstruction mode and the voxel grid size in the region of interest (ROI) can have a bearing on the results. TFM imaging can provide improved resolution of tip diffractions; which, for planar flaws, is probably the most accurate sizing technique available. When diffraction tips cannot be resolved, TFM sizing resorts to a beam boundary technique.

Several companies now provide software to map the relative sensitivity that can be obtained for a specific set of equipment parameters and mode combinations (e.g. TT, LTL, TTT, etc.). Civa simulation software maps sensitivity regions in the ROI for defined modes and combinations of modes as well as the type of flaw. Figure 1 illustrates the sensitivity maps in Civa for a point diffractor and a planar diffractor at 0° orientation. Individual maps are available for the different modes and combinations of modes.

4 Civa simulation software https://www.extende.com/
Eclipse Scientific Beamtool has added a feature that provides maps of several aspects of a TFM setup, including index and depth resolution maps. These will aid in assessing potential sizing capabilities. Figure 2 illustrates that the resolution capabilities for a 10MHz 64-element probe can be better than 1mm for both depth and index (standoff), when suitably positioned over the weld.

With the exception of work by Kurz et al [23,24] there are very few documented studies where TFM sizing has been compared to documented sizes of flaws. Kurz’s work did not describe the techniques used to determine true and ultrasonic sizing and the sizing comparison was limited to only flaws in austenitic material samples.

Until more results are made available, comparing TFM sizing to destructive testing or micro-CT-X-ray results, the most practical alternative for assessing TFM sizing capabilities would be to use simulated flaws that TFM algorithms can be used on.

Variation on TFM algorithms exist; however, their underlying principles are essentially the same. Civa simulation software can be used to generate a wide variety of “raw data” files (e.g., simulations using FMC

---

5 Beamtool software [https://www.eclipsescientific.com/index.html](https://www.eclipsescientific.com/index.html)
and PWI acquisition formats). Upon generating “raw data’ from different setups (geometries, probes, flaw types and sizes, etc.), the sizing capabilities of TFM can be assessed by analysis of the simulated data.

Figure 3 illustrates a typical weld with a flaw simulating lack of fusion. The scanning acquires data in a PWI mode (5 steps from 50° to 70°) which is then displayed as an uncorrected S-scan. After data collection, post-processing using the TFM option for TT (direct transverse) mode is used to produce the image on the weld overlay. In the example in Figure 3, only tip echoes are seen so the operator must know, prior to applying TFM, that the probable origin of the signal is a planar flaw.

![Figure 3 - PWI scanning with subsequent data collection and processing using TFM T-T mode](image)

For branching or volumetric flaws, tip signals will be weak or absent and variations on TFM can be used to enhance the flaw image to improve sizing.

In the case of a V bevel, the lack of fusion flaw can often be oriented so that a more specular reflection occurs in one direction than in another. For a flaw of sufficient size that the tip diffracted signals can be discerned, both the direct and the reflected path modes can provide tip echo signals that can be used for sizing. Figure 4 illustrates the separate T-T and T-T-T-T modes indicating the dimension of the lack of fusion in a V-bevel weld. Summing of the TFM maximum amplitude signals from the two modes can provide improved identification of the tips by seeing the points where the two sets of tip echoes cross and adds the benefit of identifying components of specular reflection. The summation of TFM data was described by Holloway [10] and is one of the options provided in CIVA.
Figure 4  PWI scanning with subsequent data collection and processing using TFM T-T and T-T-T-T mode and then the summation of the TFM modes

Figure 5  TFM processing where no signal is detected on T-T_T-T and only specular on T-T

However, not all flaws present useful tip echoes for sizing. Volumetric flaws with smaller vertical extent and poorly oriented are likely to require some form of beam-boundary (dB drop) technique to estimate size.
An example is provided in Figure 5 where a small 3x10mm lack of fusion occurs near the root where the direct path (T-T) generates a strong specular reflected signal and the reflected path (T-T-T-T) is weak and hindered by the root protrusion.

A flaw’s proximity to the closest surface (ligament) is another challenge for TFM analysis. Souad [13] made two important observations regarding planar surface-breaking flaws;

1) Using "corner echo mode" or "specular echo" mode, the flaw extent will be visible in the reconstruction to confirm the nature of the flaw but also potentially to size it with -6dB drop technique

2) A surface breaking flaw would also lead to a break in the continuity of the backwall echo on the TFM image which would then identify one of the end point of the flaw.

Souad also cautions that the 6dB amplitude drop may be inaccurate when the defect dimensions are larger than the ultrasonic beam for the mode considered for reconstruction, resulting in underestimating the flaw height.

For a given application, where a POD is required, it should be feasible to configure a TFM sizing qualification procedure of good practice. Preparation of a sizing procedure could take advantage of the TFM optimisation software described in Figures 1 and 2, and this would address equipment requirements (probe frequency and aperture) and setup parameters (probe placement to optimise sensitivity and resolution). Additionally, limitations would need to be identified, such as the use of beam boundary techniques if tip echoes are not clearly visible. Sizing accuracy is also expected to be a function of the nature of the materials tested (e.g., coarse-grained materials will require lower frequency probes with their inherent reduction in resolution).

5. Conclusions and Future Work

Until more evidence from destructive tests or micro-CT X-ray sizing compared to TFM sizing is available, it will be difficult to assess the tolerance that should be used for the true flaw size for POD reliability analyses. When a POD is being used to assess the reliability of an NDT inspection procedure (other than TFM), TFM may be a viable option to destructive testing to assess the true flaw size.

When clearly defined tip-diffraction signals can be identified, sizing expectations on the order of +/-0.5mm should be achievable using TFM. Where tip-diffraction signals cannot be resolved, a form of beam-boundary sizing based on relative amplitude may still provide reasonable sizing estimates if acquisition parameters are suitably selected.

The NDT community has long been aware of sizing uncertainty in samples used for reliability studies (POD). In spite of this, there seems to have been no attempt to incorporate the sizing uncertainty into the derivation of the POD calculations. Conversely, uncertainties in POD calculations (confidence bounds) are regularly incorporated into the acceptance criteria derived from POD curves.

When TFM is used as the sizing tool for a POD qualification of an NDT procedure, a documented sizing procedure is recommended that can help to ensure consistent and accurate results.

Future work to assess the potential accuracy of TFM sizing is planned by using a round robin test on simulated data with a variety of flaw types and geometries.
References


5. ENIQ Recommended Practice #5, Guidelines for the design of test pieces and conduct of test piece trials, Publications office of the European Union, Netherlands, 2011.


