Highlights

Highlighting Anomalies in Ultrasonic Scan Data by Shannon Information Processing

Jonathan Lesage, Mohammad Marvasti, Oliver Farla

- Shannon Information based processing scheme devised to represent ultrasonic scan data in terms of "unusualness" with respect to an un-flawed scan or scan region

- Processing preserves phase of signals allowing through wall positioning of flaws in ToFD signals and the formation of images by summation of processed signals

- Null distributions for Shannon Information processed signals and images rendered by summation of these signals are derived and used to set global significance based thresholds and dynamic range limits

- Shannon Information processing enables detection of near side crack in ToFD scan merged in lateral wave signal; between 6.4 and 16.7 dB increase in SNR observed compared with unprocessed data

- Unambiguous Detection of Corrosion Under Pipe Support flaws made possible by Shannon Information processing of Multiple-Skip Phased Array scans

- Noisy High Temperature Inspection via Delay and Sum Imaging with Shannon Information processed signals demonstrated to increase SNR (6.4 - 7.9 dB) compared with signal formed from unprocessed signals; detection of simulated flaw 2 mm from far side surface made possible by Shannon Information processing
Highlighting Anomalies in Ultrasonic Scan Data by Shannon Information Processing

Jonathan Lesage\textsuperscript{a,}*, Mohammad Marvasti\textsuperscript{b}, Oliver Farla\textsuperscript{b}

\textsuperscript{a}Acuren Application Development Group, 405 Maple Grove Road, Unit 1, N3E 1B6, Cambridge, Canada
\textsuperscript{b}Eclipse Scientific, 1260 2nd Avenue East, Unit 4, Owen Sound, N4K 2J3, Canada

Abstract

Analysis of ultrasonic nondestructive testing data can be viewed as an exercise in anomaly detection, whereby, signals/image features found to appear consistently throughout a scan are attributed to non-relevant reflections from the weld root/cap, back-wall, wedge base, etc., while isolated indications of sufficient amplitude (with respect to a calibrated reference level) are identified as potential defects requiring further characterization. Finding indications which correspond to flaws in this way can be challenging, particularly when they are low amplitude and/or are proximate to strong geometric reflectors. In this paper, a novel post-processing scheme is proposed to facilitate identifying defects in encoded ultrasonic scans. The algorithm effectively emphasizes indications which are unusual compared with an un-flawed reference scan (or region of a scan). This is accomplished by estimating the distribution of signal amplitudes in the reference scan and then using the reference distribution to compute the so-called "Shannon Information" associated with new data points. Shannon Information is a fundamental metric in Information Theory, which quantifies how surprised one should be to observe a given quantity - ultrasonic signal amplitude in this case - knowing how that quantity varies in general, as described by its probability distribution. The mathematical foundations of the technique are outlined in detail followed by a demonstration of the efficacy of Shannon Information processing on Time of Flight Diffraction (ToFD) and Phased Array (PA) scans featuring flaws which are difficult to discern by conventional means. The proposed method of presenting ultrasonic inspection data is observed to increase signal-to-noise ratio and highlight subtle perturbations in consistent, non-relevant scan features associated with presence of defects. In addition, in the absence of flaws, Shannon Information processing is shown to transform ultrasonic signals, projection views and Delay and Sum rendered images into stationary random processes with known null distributions which allows for global detection thresholds to be set according to a desired level of statistical significance.

Keywords: Anomaly detection, Inaccessible Corrosion, Time of Flight Diffraction, Multiple-Skip Phased Array, Shannon Information, Ultrasonic Testing, High Temperature Inspection, Noise Reduction

1. Introduction

Ultrasonic inspection of metallic industrial components is routinely performed after fabrication in order to identify any manufacturing flaws and later to detect defects which develop while the component is in service. In many cases, the exact process by which these inspections are to be performed (depending on the nature of the component) is directly specified in codes and standards published by ASME [1], ISO [2], etc. These codes and standards provide strict guidelines for performing a sensitivity calibration, whereby a reference signal amplitude is established with respect to reflectors of known size (side drilled holes, surface connected notches) and then specify an investigation threshold pegged to this reference amplitude (typically 20 % [1]) above which all indications must be considered potential flaws and carefully assessed by the technician/operator. In practice, analyzing ultrasonic inspection data involves considerable judgement by the analyst as many non-relevant signals passing the investigation threshold will be present in any given scan. Features occurring consistently throughout the scan are usually due to reflections from the component's boundaries or other benign sources, while localized scan features are attributed to defects. In other words, the analyst performs a kind of anomaly detection on the scan data where indications passing the code mandated investigation threshold are compared with the trend of signal amplitude along the scan axis - only indications which are unusual compared with the rest of the scan(s) are considered defects to be reported. This degree of operator discretion means that the number of defects identified by different personnel analyzing the same scan data can vary considerably. Additionally, certain types of flaws (e.g. porosity, slag inclusions, etc.) do not strongly reflect ultrasound and consequently can be easily missed by inexperienced analysts, despite passing the investigation threshold.

Over the years, several strategies have been attempted to mitigate operator dependence in the analysis of ultrasonic inspection data - these include: inspection specific training with performance demonstration requirements for personnel [1], capturing the variability in technician performance in Probability of Detection (PoD) studies [3] and the algorithm...
mic implementation of flaw detection via expert systems. Of these, completely automated flaw detection via expert systems is the most robust approach, as a properly implemented algorithm will find all indications passing a specified investigation threshold, however, for many applications (weld inspection in particular), a large number of non-relevant indications pollute the automatically generated inspection report requiring a human analyst to review all indications, a process that can be more time-consuming than completely manual analysis.

Efforts to completely automate the analysis process using Machine Learning (ML) have been detailed in the literature (see [4] and [5] for a comprehensive review of the state of the art). Training ML algorithms requires a large number of classified examples which may not be readily available given the relative paucity of naturally occurring defects [5]. The use of transfer learning together with supplementing experimental data sets with high fidelity simulated defect responses [6] can improve the classification accuracy of a given ML algorithm, though, at present, pure ML based approaches to UT data analysis appear to be viable for high volume production environments.

Recently, a method for applying a time dependent thresholds based on the extreme value distribution for a collection of A-Scans whose time dependent distribution parameters have been estimated from an un-flawed reference scan has been proposed by Song et al [7], [8], [9]. The technique shows promise in identifying sub-waveform simulated flaws in noisy materials, however, the approach requires that time gate chosen for analysis cannot feature any coherent scatterers. For general industrial inspections non-relevant reflections are difficult to robustly gate out, or are themselves to be monitored for drop-out or subtle variations in pulse shape. Song and co-authors dealt exclusively with automated immersion scanning of manufactured parts, where the only meaningful source of amplitude variation in scan amplitude is due to grain scattering/noise. In the field, scans are performed by coupling plastic wedges to rough or irregular surfaces spring-loaded scanning fixtures, robotic arms or human hands - taken together these represent additional sources of signal variation.

In this paper, a simple information theoretic approach to representing ultrasonic scan data which identifies anomalous scan features is proposed. The algorithm involves estimating the parameters of an assumed probability distribution of ultrasonic signal amplitudes for each point in a digitized ultrasonic signal using an un-flawed reference scan (or scan region). Inspection data is then evaluated based on how unusual the ultrasonic signal amplitudes are with respect to these estimated reference distributions. This is accomplished by performing a simple hypothesis test for each point in the digitized signal at each scan position, where the null hypothesis is that the signal’s amplitude is distributed according the reference distribution. Scan data is presented in terms of the amount Shannon Information against the null-hypothesis, which can be interpreted as how surprised one should be to observe a value at least as extreme as a given amplitude. Alternatively, the information theoretic presentation of scan data proposed in this paper can be thought of as the logarithm of the Probability of False Alarm (PoFA) [3], reported separately for each point in the digitized inspection record. The process of estimating reference distribution parameters and plotting ultrasonic scan data in terms of Shannon information is presented for both individual signals as well as projection views and Delay and Sum (DaS) images formed by the summation of multiple Shannon information processed signals (in analytic signal form). Next, the null distribution of Shannon Information processed A-scans and images rendered by summation of processed A-scans are derived and used to define appropriate dynamic range limits and significance based thresholds. Finally, the algorithm's general performance is demonstrated on Time of Flight Diffraction (ToFD), Phased Array Ultrasonic Testing (PAUT) and B-Mode Imaging scans featuring flaws which are difficult to detect without Shannon Information processing.

The proposed method of representing ultrasonic scan data can be readily extended to other types of Non-destructive Evaluation data (e.g. Ground Penetrating Radar, Eddy Current, etc) and is formalization of the heuristic anomaly detection typically performed by ultrasonic inspection technicians.

2. Mathematical Formulation

In the following section the mathematical formulation for the proposed method of representing ultrasonic inspection records in terms of statistical (Shannon) information is outlined in detail. Consider an ultrasonic inspection, where a volume of material is interrogated by scanning along the axis of a component, generating a series of L ultrasonic beams at each scan location. For generality, these beams can be assumed to have been generated by moving a single probe or pair of probes perpendicular to the scan axis (z), or, alternatively, by electronic sequencing of phased array probe(s). The transmitting probe/aperture can either be the same probe/aperture (pulse/echo configuration) or a different probe/aperture (pitch/catch configuration). The digitized inspection data is stored as an array, \( a_l(z_m, t_n) \), where, \( a_l(z_m, t_n) \) represents the A-Scan signal generated by the \( l^{th} \) beam at the \( m^{th} \) scan position at time \( t = t_n \). For example, a typical phased array sector scan of a weld is shown in Fig. 1, with the global coordinate system, beams and A-Scan signals defined.

Many industrial inspections involve looking for flaws/anomalies in a component which is effectively generated by extruding a nominal cross-sectional geometry. An ideal, un-flawed weld, for instance, should be identical regardless of position along the weld axis. It follows, therefore, that for an ultrasonic scan performed parallel to the axis of such an ideal, un-flawed component, \( a_l(z_m, t_n) \) would be independent of scan position, \( z \), that is \( a_l \) would be translation invariant in the scan direction, with \( a_l(z, t_n) = a_l(z + \delta z, t_n) \). In reality, factors such as surface conditions, grain noise and variations in the weld geometry will cause \( a_l \) to vary
randomly in the scan direction, even in the absence of defects.

Assume that in the absence of flaws, these combined sources of signal amplitude variation can be represented by a random variable, \( A_{0in} \), with probability density function, \( f_{A_{0in}}(u; \theta_{in}) \), where \( \theta_{in} \) is a vector of unknown distribution parameters for each combination of \( l \) and \( n \). A scan of an un-flawed reference specimen, then, can be used to estimate parameters, \( \theta_{in} \), using the maximum likelihood approach\(^1\):

\[
\theta_{in}^* = \arg \max_{\theta_{in}} \frac{1}{M_0} \sum_{m=0}^{M_0-1} \ln f_{A_{0in}}(a_{0}(z_m, t_n); \theta_{in})
\]

Where, \( a_{0}(z_m, t_n) \) represents the \( m^{th} \) sample for the \( l^{th} \) beam and \( n^{th} \) time from an \( M_0 \) point reference scan. Once the parameters have been estimated with Eqn. 1, the background distributions for each \( (l, n) \) pair can be used to perform a hypothesis test for each new digitized A-Scan amplitude. Consider the null hypothesis \( H_0 \): a random sample from \( A_{0in} \) has a value more extreme than \( a_l(z_m, t_n) \) by random chance. The probability that \( H_0 \) is true is given by the 2-tailed \( p \)-value which can be computed using the Cumulative Distribution Function (CDF) of \( A_{0in} \), \( F_{A_{0in}}(u; \theta_{in}) = \text{Pr}(A_{0in} \leq u) \):

\[
p_l(z_m, t_n) = \begin{cases} 2F(a_l(z_m, t_n); \theta_{ml}^*), & F(a_l(z_m, t_n); \theta_{ml}^*) \leq 0.5 \\ 2[1 - F(a_l(z_m, t_n); \theta_{ml}^*)], & F(a_l(z_m, t_n); \theta_{ml}^*) > 0.5 \end{cases}
\]

Eqn. 2 can be used directly to apply significance thresholds to ultrasonic data, though representing A-scan amplitudes in terms of \( p \)-values does not scale appropriately with the degree of "unusualness" associated with A-scan values which are well outside the expected range based on the null-distribution [10]. A more appropriate choice is to represent the A-scans in terms of so-called "Shannon Information" or "Self-Information", which quantifies how surprised one should be to observe a given value based on its probability of occurrence (the \( p \)-value in this case) [11]. The Shannon information associated with an event with probability, \( P \), is given by [11]:

\[ S = -\log_b P \]

Where, \( b \) is an arbitrary base for the logarithm, the choice of which varies according to context and determines the conventional units for information, i.e. bits for \( b = 2 \), Nats for \( b = e \), Hartleys for \( b = 10 \), etc [12]. By inspection of Eq. 3 it can be seen that the lower the probability of an event occurring,
the more "Information" observing the event provides, or put another way, the more "surprised" one should be to observe

\[
s_l(z_m, t_n) = \begin{cases} 
  \log_b \left[ 2F \left( a_l(z_m, t_n); \theta_m^* \right) \right], & F \left( a_l(z_m, t_n); \theta_m^* \right) \leq 0.5 \\
  -\log_b \left[ 2 \left( 1 - F \left( a_l(z_m, t_n); \theta_m^* \right) \right) \right], & F \left( a_l(z_m, t_n); \theta_m^* \right) > 0.5 
\end{cases}
\]

(4)

\[
\sigma_{ln} = \sqrt{\frac{1}{M_0} \sum_{m=0}^{M_0-1} \left( a_{ln}(z_m, t_n) - \mu_{ln} \right)^2}
\]

(7)

Song et. al. suggest that the scan increment be larger than the beam profile to ensure that grain noise estimates from adjacent scan positions are completely independent [8]. For industrial inspections where the random sources of amplitude variation are not restricted to grain noise an appropriate scan increment which ensures the statistical independence of samples from neighboring scan positions cannot be readily determined from the passive beam profile of the transducers. For the purposes of this paper, a scan increment of 1 mm (typical) is used for parameter estimation.

In many cases, ultrasonic inspection data is presented in rectified format. Since the underlying RF signal is normally distributed, then the full wave (FW) rectified signal (absolute value of the RF signal) is distributed according to a folded normal distribution [18] and the envelope rectified signal (modulus of the analytic signal representation of the RF signal) follows a Rice distribution [19]. When the RF signal has a non-zero mean (due to non-relevant scan features), the corresponding rectified distributions require iterative maximum likelihood estimation to obtain the distribution's parameters (FW) [18], (envelope) [20]. Estimating distribution parameters for the null hypothesis from the rectified signal in this way is undesirable and becomes impractical as the number of beams and sample points increases. Alternatively, the data can be recorded in RF format which allows the normally distributed background's parameters to be easily determined via Eqs 6, 7. The signed Shannon information representation of the signals can then be computed from Eq. 4 and if preferred, the signed Shannon information signals can be rectified by taking the absolute value of the result (FW rectification). This approach avoids the complications involved with estimating the parameters of Folded Normal or Rice distribution parameters.

Envelope rectification of Shannon Information processed signals can be obtained by computing the complex modulus
of $\tilde{s}_l(z_m, t_n)$ (the parameters for the background of $\tilde{s}_l$ can also be estimated from Eqns. 6 and 7 by substituting $\hat{a}_l$ for $a_l$), however, in doing so, the ability to directly set significance based thresholds to the resulting signal is not straightforward - the null-distribution for $\tilde{s}_l$ is not known. Fortunately, the null-distribution of the complex modulus of the sum of $\tilde{s}_l$ can be directly computed and used to set significance based thresholds and estimate the Root Mean Squared (RMS) noise level on images rendered by summation (see Appendix A).

2.2. Combining Redundant Scan Data: Projection Views and Imaging

The signed Shannon Information representation of scan data has been defined for a collection of signals which can be interpolated onto a grid of points in the $y - z$ (Side View/B-Scan) or $x - z$ (Top View/C-scan) planes, providing projected images of the inspected volume [21]. Images formed by interpolation of the individual signals $a_l(z_m, t_n)$ overlap in space (at least partially) and therefore represent redundant images of the same underlying region. With phased array data, the images from different beams are combined by performing a point wise maximum, however, when the signals are in signed Shannon information form, the natural way to combine these redundant images is by summation:

$$V(y, z_m) = \left| \sum_{l=0}^{L-1} h_l[\tilde{s}_l(z_m, t_n)] (y, z_m) \right|$$

$$V(x, z_m) = \left| \sum_{l=0}^{L-1} g_l[\tilde{s}_l(z_m, t_n)] (x, z_m) \right|$$

where $h_l[..](y, z_m)$ and $g_l[..](x, z_m)$ represent functions which interpolate the $l^{th}$ Shannon processed signals (in analytic signal form) along the $x$ and $y$ coordinates, respectively (typically piece-wise linear or nearest neighbor, relating the time of flight and beam angle to $x$ or $y$). In addition, to the projection views, cross-sectional images in the $x - y$ plane can be generated at a given scan position by summing the partial images obtained by interpolation of the $l^{th}$ signal:

$$V(x, y, z_m) = \left| \sum_{l=0}^{L-1} q_l[\tilde{s}_l(z_m, t_n)] (x, y, z_m) \right|$$

Where, $q_l[..](x, y, z_m)$ denotes a function which interpolates the $l^{th}$ signal on a grid of points in the $x - y$ plane. Of particular interest is the case where the $L$ signals represent an FMC, Plane Wave Imaging (PWI) or other generic dataset to be rendered into a volume by a delay and sum (DaS) process (e.g. TFM). In this case, $q_l$ takes the following form [13]:

$$q_l[\tilde{s}_l(z_m, t_n)](x, y, z_m) = \tilde{s}_l(z_m, t_n = \tau_l(x, y))$$

Where, $\tau_l(x, y)$ is the combined transmission and reception delay required to focus the $l^{th}$ signal at $(x, y)$. Summations in Eqns. 9, 8 and 10, can be interpreted as determining how surprised one should be to observe a signal value more extreme than $a_l(z_m, t_n)$ (interpolated along the spatial axes) from all $L$ beams by random chance. In other words, if pixel values interpolated from the $L$ signals are assumed to be independent, then the logarithm of their joint p-value is simply equal to the sum of the logarithms of the individual p-values due to the product to sum identity of logarithms. The additive nature of information values from independent sources is an important property of measuring surprisal through the logarithm [11]. Additionally, since the direction of extremity is encoded in the sign of the Shannon Information processed signals (positive for above background, negative for below background), high values are expected to result from summing over $l$ when a plurality of the signals are consistently unusual (have the same sign). This works the same way as with unprocessed ultrasonic signals (points where signals are in phase sum coherently), though with the resulting image or projection view representing the amount of information against the hypothesis that no relevant feature is present at any given point in the inspection volume. Performing the summation on the analytic signal representation of the Shannon Information processed signals provides smoothly rectified outputs which can be readily thresholded based on a desired significance level.

2.3. Thresholding, Noise Estimation and Saturation Levels

In order to provide a framework for setting significance level based thresholds on signals represented as the Shannon information against a signal value being drawn from the null distribution for a given signal, the null distribution for signed Shannon Information is required. Similarly, thresholds on projection views and DaS images as defined in Eqns 9, 8 and 10 can be set knowing the null-distribution for the modulus of the sum of $L$ Shannon Information processed scans (in analytic signal form). Deriving these distributions is straightforward, though somewhat lengthy. Accordingly, the derivations of these PDFs and CDFs are provided in the appendix along with comparisons between the derived expression with histograms and empirical distribution functions obtained via simulation.

In summary, we define two random variables:

1. $S_0$ drawn from the null distribution of signed Shannon Information values, having a PDF and CDF defined by Eqns. 12 and 13, respectively.
2. $V_0 = \left| \sum_{l=0}^{L-1} \tilde{s}_0 \right|$ drawn from the null distribution of the modulus of the sum of $L$ samples each drawn from the null-distribution of signed Shannon Information (in analytic signal representation) - the PDF and CDF of $V_0$ are given in Eqns. 14 and 15, respectively - see the derivations in the Appendix for details.

$$f_{S_0}(u) = \begin{cases} \frac{\ln b}{2} b^{n} & u \leq 0 \\ \frac{\ln b}{2} b^{-u} & u > 0 \end{cases}$$

$$F_{S_0}(u) = \begin{cases} 1 - \frac{b^{-u}}{2} & u \leq 0 \\ \frac{b^{u}}{2} & u > 0 \end{cases}$$
Given a chosen significance level $\alpha$, the corresponding value of Shannon Information, $s_\alpha$, represents a value below which the Shannon processed signal is set to zero, i.e. $s_L(z_m, t_n) = 0, -s_\alpha \leq s_L(z_m, t_n) \leq s_\alpha$. The relationship between $s_\alpha$ and $\alpha$ is as follows:

$$a = \Pr \left[ S_0 \leq -\frac{s_\alpha}{2} \right] + \Pr \left[ S_0 \geq \frac{s_\alpha}{2} \right]$$

$$a = 2 \left[ 1 - F_{S_0}(s_\alpha) \right] = b^{-s_\alpha} \tag{16}$$

where, $F_{S_0}(\cdot)$ is the CDF of $S_0$ given in Eqn 13. As expected, Eqn 16 shows that the threshold on Shannon Information processed data is simply the Shannon Information associated with the chosen significance level.

Following the same process for determining a threshold for projection views and DaS images, a value $v_\alpha$ corresponding to $a$ is obtained as follows:

$$a = \Pr \left[ V_0 \geq v_\alpha \right]$$

$$a = 1 - F_{V_0}(v_\alpha) = e^{-\frac{-\ln^2(b)u^2}{4L}} \tag{17}$$

$$v_\alpha = \frac{2}{\ln b} \sqrt{-\ln a}$$

From Eqn. 17, it is observed that the threshold depends on both the significance level and the number of signals, $L$.

An interesting and useful feature of Shannon Information processing of ultrasonic signals is that variations in signal amplitude due to changes in beam profile are effectively equalized. Accordingly, Shannon Information processed signals are stationary which allows global thresholds to be set, as opposed to the time dependent thresholding approach taken by Song et al. [8]. Furthermore, this global threshold is directly tied to a measure of statistical significance.

In contrast to standard UT data, Shannon Information processed data does not have a limited dynamic range. Consequently, when analysing Shannon Information processed signals or images, the dynamic range of the data must be artificially limited to avoid a scenario where a highly unusual scan feature obscures less unusual features of significance. For conventional UT, PAUT and ToFD scans, the dynamic range is limited by the saturation level of the electronics - typically scaled to 100% (full screen height) [21], [22]. By increasing the reference gain, or performing a calibration, this dynamic range is scaled such that the sensitivity to a known reflector (lateral wave for ToFD, calibration target for pulse/echo UT) appears at a desired fraction of the full screen height (e.g. 80%) [22], [21]. In setting the gain level in this way, standard practice is to ensure that the noise level remains below 10% [1] (one tenth of the saturation level).

Since the null-distributions for Shannon Information processed signals and images are known, the corresponding RMS noise levels can be directly computed from Eqns. 12 and 14, as follows:

$$s_{\text{noise}} = \sqrt{\int_{-\infty}^{\infty} u^2 f_{V_0}(u) du} = \frac{\sqrt{2}}{\ln b} \tag{18}$$

$$v_{\text{noise}} = \sqrt{\int_{0}^{\infty} u^2 f_{V_0}(u) du} = \frac{2\sqrt{L}}{\ln b} \tag{19}$$

For the purposes of this study, the saturation level on all Shannon Information processed data will be set to 10 times the noise level as computed from Eqns. 18 and 19. Saturation here is only intended to scale the Shannon Information processed scans to a dynamic range consistent with standard NDT practices. In contrast to conventional ultrasonic scans, the saturation applied to Shannon Information processed scans is artificial and the range can be readily expanded to analyze the full waveform of highly unusual scan features.

3. Results and Discussion

In this section, comparisons between standard ultrasonic and Shannon Information processed data are presented for three inspection applications where traditional methods of analysis are complicated by the obscuring influence of non-relevant scan features, the lack of suitable calibration methods/detection criteria and/or high levels of noise. To this point, the units of Shannon information have not been specified. The authors favour Hartleys (Harts), corresponding to $b = 10$, which for an individual A-scan directly maps to the order of magnitude of statistical significance i.e. $s_{\alpha=0.01} = 1 \text{ Harts}$ for values in the $90^{th}$ percentile, $s_{\alpha=0.01} = 2 \text{ Harts}$ for values in the $95^{th}$ percentile, etc.

3.1. Weld Inspection with Time of Flight Diffraction

A welded test plate featuring 3 simulated weld flaws from FlawTech (Concord, North Carolina, USA) was scanned with a matched pair of 5 MHz, 3 mm diameter probes together with $70^\circ$ refracting wedges - both from Phenonix ISL (Wallington, UK). Data was collected via a TOPAZ64 Ultrasonic Pulser/Receiver from Zetec (Quebec City, QC, Canada). Fig. 2 shows the scan configuration, location and type of defect as well as unflawed regions used to estimate the parameters of the background distributions and the noise level (yellow boxes).

This particular test plate was chosen to evaluate the efficacy of Shannon Information processing as it contains flaws which are difficult to detect with ToFD, namely a non-planar flaw (slag inclusion), a top surface connected crack in the heat affected zone (offset from the weld centerline) and a side-wall lack of fusion (LOF) close to the top surface. The crack and LOF are within the nominal lateral wave dead-zone (the depth range obscured by the lateral wave signal marked by the dashed green line in Fig. 2 (b)). The ToFD setup
was calibrated according to [1] with the lateral wave amplitude set to 80%, ensuring the noise level between the lateral wave and longitudinal wave (LL) backwall signal remains below 10%. The probe center separation (PCS) was determined accordance with standard practice with the nominal crossover depth at \( \frac{2}{3} \) of the piece thickness [23]. The recorded time gate was chosen to encompass the lateral wave, direct longitudinal wave backwall reflection (LL Backwall) and the mode converted backwall signal (LT Backwall) - this allows direct L wave diffractions as well as various mode converted diffraction signals to be observed [22].

A comparison between the unprocessed and Shannon Information processed B-scans is shown in Fig. 3. A suitable unflawed reference region used to estimate the background distribution parameters was chosen to include all scan locations excluding the nominal flaw positions plus an additional 15 mm to account for beam divergence (yellow boxed regions shown in Fig. 2). This same scan region was used to compute the observed RMS noise values for the unprocessed and processed scans, taken from time gates excluding the lateral wave and backwall signals. For consistency of dynamic range between the processed and unprocessed scans, the digitized ToFD scan has been scaled such that the RMS noise level from the unflawed reference region is exactly 10%. The saturation
level for the Shannon Information processed scan was computed as 10 times the RMS noise value predicted from Eqn. 18 (rounded to the nearest integer) $s_{\text{saturation}} = 10 \times \frac{\sqrt{2}}{\ln(10)} \approx 6$ Harts.

From Fig. 3 (a) one can readily identify the Slag inclusion between the lateral wave and LL backwall signals (solid red box) and between the LL and LT backwall signals (dashed red boxes), though the direct and mode converted signals are relatively low SNR (typical for volumetric flaws in ToFD scans). The lack of fusion flaw is also readily observed in the unprocessed B-scan by a significant change in the lateral wave signal and an apparent tip signal partially merged into the lateral wave signal (solid blue box). The HAZ crack flaw, is more difficult to identify in the ToFD signal - a mode converted indication merging into the LT backwall signal can be discerned upon close examination (dashed green box), together with a very slight lateral wave signal shift/change (solid green box). From Fig. 3 (b) all three flaws appear above the “saturation" level and are much easier to detect, due to the apparent increase in SNR and near total suppression of the lateral wave and backwall signals. In particular, the small distortion in the lateral wave signal due to the HAZ crack (solid green box) is highlighted, which helps to confirm that the mode converted indication (also highlighted) close to the LT backwall is in fact due to a true flaw. In general, the Shannon Information processed scan data is observed to accentuate all subtle variations in the original ToFD data.

Figs. 4, 5, 6 show a comparison between the original ToFD A-scan signals and the Shannon Information processed A-scans for the Slag, HAZ Crack and Lack of Fusion defects, respectively. The Slag defect A-scan comparison confirms that Shannon Information processing preserves the relative phase and peak positions originating from the top and bottom of the flaw. Additionally, these indications are observed to be much higher SNR compared with the original ToFD signal, which has the side effect of making the top and bottom indications appear sharper/more distinct and thus somewhat easier to separate. The depth estimates (y - computed according to [22]) of the top and bottom of the flaw agree favourably with the through wall positions reported by the manufacturer of the plate (shown in Fig. 2) $^2$. The lateral wave and LL backwall signals are completely suppressed in the Shannon Information processed data (below the noise floor). The LT backwall is also largely suppressed, facilitating the detection of the second mode converted slag indication, however, two additional “signals” appear in the Shannon Information processed A-scan in the LT backwall timegate - these are due to the change in the mode converted backwall signal due to the presence of the slag defect which is clearly identified the processed B-scan (Fig. 3) (b).

From the HAZ Crack A-scan comparison (Fig. 5), the lateral wave indication is found to shift with respect to its nominal position due to the crack in the original ToFD signal (a), which manifests as a significant negative peak in the Shannon Information processed signal (b). The negative peak in 5 (b) occurs at the nominal lateral wave time of flight due to the amplitude of the lateral wave signal dropping well below the usual range of values caused by the shift in peak position associated with the interference between the lateral wave and the HAZ crack. Note that this indication is not a true signal from which the depth of the flaw can be estimated, rather, it is caused by the subtle disruption of the consistent lateral wave indication which is nearly imperceptible in the original ToFD data. The mode converted signal occurring in the unprocessed ToFD scan (partially merged into the LT backwall signal) would alert a skilled analyst to carefully scrutinize the lateral wave signal, to confirm the presence of a top surface connected flaw. Here, Shannon Information processing makes this process much more robust by enhancing the mode converted and lateral wave disruption signals. Unfortunately, minor changes in both the lateral wave and backwall signals which are apparently highlighted by Shannon Information processing are not always due to flaws. Variations in surface conditions, probe separation/tilt, weld root/crown morphology can all influence the lateral wave and/or backwall signals - though these sources of signal alignment can be partially mitigated by improvements to the scanning system and registration algorithms. Accordingly, detection of flaws cannot be solely based on small perturbations to the later wave and/or backwall signals and should be confirmed by another indication occurring at the same scan position.

Comparing the unprocessed and Shannon Information processed A-scan signals from the LOF defect (Figs. 6 (a) and (b), respectively), a marked increase in the bottom tip signal’s amplitude with respect to noise is observed in the processed signal. As with the HAZ crack, a significant negative spike can be seen in the processed signal at the nominal lateral wave time of flight indicating that there is a disruption in the lateral wave due to the near surface LOF. Unlike with the slag indication, the tip (bottom of LOF) signal’s peak position differs slightly between the original and Shannon Information processed data. This slight discrepancy is likely due to the fact that the lateral wave and tip signals are superimposed on one another. It is difficult to assess whether the apparent tip position in the original ToFD signal or its Shannon Information processed version is the correct tip location, however, the depth estimate obtained from the Shannon processed signal is closer to the through wall extent from the manufacturer’s specifications (6.0 mm - see Fig. 2).

One of the main benefits of the Shannon Information representation of ultrasonic data is the ability to set global thresholds connected to measures of statistical significance. Figure 7 shows the effect of applying thresholds of increasing significance on the Shannon Information processed ToFD B-scans. Here, increasing the significance level was found to suppress more of the background noise, with a threshold corresponding the 99.9th percentile removing all but the re-
evant scan features. Selection of a significance level is always somewhat arbitrary, however, given the number of samples in a typical A-scan (∼ $10^3$) or DaS image (∼ $10^5 - 10^6$), amplitudes in the 99th percentile are expected which implies that
a higher significance level will be required to remove high values occurring by random chance. The approach taken by Song and his co-authors is more conservative - thresholds are determined based on the extreme value distribution of an ensemble of scans whereby the significance level is set for the maximum/minimum of the ensemble [8]. The extreme value distribution for Shannon information processed data may be derived from the null distribution and thresholds selected as in [8], however, Song’s approach effectively increases the threshold value for a given significance level based on the number of points in the ensemble (number of points × number of scan locations in our case). Here, the purpose is not to ensure that no non-relevant indications appear above threshold for any point in the scan at the expense of potentially missing smaller flaws of relevance (an appropriate strategy for auto-rejection of manufactured parts detailed in [8], [9]).
rather, we are primarily interested in highlighting flaws of significance which a human operator or possibly a trained ML algorithm can further characterize. On average, $100 \times 0.25\%$ of the recorded scan data are expected to pass the threshold by random chance, though, given that these are likely to be isolated points would likely not be flagged as relevant by a human operator or algorithmically disqualified by requiring a specified number of adjacent samples (determined from the transducer pulselength) to pass the significance threshold.

Having identified a reference region for estimating parameters of the background distributions with Eqn. 6 and 7 (see the yellow boxes marked on Fig. 2), the Shannon Information processed data in this unflawed region can now be used to experimentally verify the distribution of signed Shannon Information values under the null hypothesis. Fig. 8 compares the PDF and CDF determined empirically from the unflawed reference region to the corresponding PDF and CDF computed using Eqns. 11 and 13.

The empirical and theoretical PDFs and CDFs are found to agree favourably, implying that the assumption of normality for the background distribution is justified in the aggregate and that Shannon Information processed signals in the absence of flaws can be considered stationary random variables with a PDF and CDF defined by Eqs. 12 and 13, respectively. The RMS noise value computed from the reference region = 0.619 Harts is very close to the value predicted from Eqn. 18 (0.614 Harts) - this provides further evidence that the derived distribution for signed Shannon information processed signals under the null hypothesis is, in fact correct.

The authors acknowledge that it is very rare to have a pristine reference sample or known unflawed reference region from which to estimate background distribution parameters and that presenting results where a well-defined reference is assumed to exist may seem disingenuous. Fortunately, flaws are relatively rare such that in a given scan or collection of scans from an inspection campaign, the ratio of flawed to unflawed length is expected to be very low and as a result their presence should not greatly influence parameter estimates. In fact, even for the test plate examined in this section, where the total flawed length is nearly 15% of the total scan length, Shannon Information processing performed without specifying a reference region still provides an improvement over the original ToFD scan data. Fig. 9 shows the difference between a Shannon Information processed B-scan where a reference region has been specified and where the background distribution parameters have been estimated using the entire scan length.

Here, the B-scan obtained using a reference region is observed to feature improved lateral wave and backwall suppression as well as increased SNR compared with the B-scan where parameters were estimated from the entire scan, however, the relevant scan features are still highlighted. Fig. 9 (b) represents the “worst case scenario” in which a significant portion of the entire scan length is flawed and the background distribution's parameters are heavily skewed, which in turn results in relevant indications appearing as less "unusual" with respect to these skewed background estimates. It is important to note that for damage that can run the length of an entire scan; high temperature hydrogen attack (HTHA) or lack of fusion flaws resulting from improperly tuned robotic welding systems, for example, the flaws themselves will be suppressed as they will be "usual" features. For Shannon Information processing to be effective at highlighting these types of flaws, the background distribution's parameters must be estimated using an un-flawed mockup or calibration sample matching the geometry of the piece to be inspected.

3.2 Phased Array Inspection for Detection of Corrosion at Inaccessible Locations

Detection of corrosion and pitting defects at inaccessible locations such as under pipe supports or in the critical zone of large storage tanks, is of vital importance to petrochemical asset owners. A number of inspection techniques for detection of this type of corrosion have been proposed - these include: Guided Wave (S0, SH0 and SH1), creep head-wave (CHIME), Multi-skip (M-skip) and High Order Mode Clustering (HOMC). A detailed review paper summarizing available inspection technologies for detection and sizing of inaccessible corrosion flaws is provided in [24]. From the perspective of an NDT service company, the techniques described in [24] are specialized requiring equipment such as EMAT pulsers, large high angle wedges, through transmission alignment fixtures, etc, which are not generally available to technicians. Furthermore, field conditions such as support interference, and probe access restrictions often preclude their use altogether.

An alternative approach employed by many NDT service providers is to perform a type of reflection mode Multiple-skip technique with a phased array probe, where several shear wave beams skip multiple times between the component's surfaces, reflecting from localized pitting and sharp corrosion features. Often, low amplitude, diffuse reflections are produced by interaction with the rough texture of natural corrosion flaws which are difficult to unambiguously identify. Using a phased array probe to generate a range of refracted beams improves the chance of reflecting from a facet or surface of a rough corrosion flaw and provides redundant information which allows an analyst to confirm the presence of corrosion by cross-referencing signals from the different beams. Unfortunately, this approach is complicated by the fact that reflections from corrosion flaws are highly dependent on flaw morphology and are generally quite weak. Performing a time corrected gain calibration is not feasible due to the long travel path and would not capture the attenuating influence of repeated interaction with rough and/or fluid loaded surfaces and standard calibration targets (notches, side drilled holes) are not representative of corrosion. Consequently, the difference in reflection amplitude from the same corrosion feature at the beginning and end of the inspection range may be several orders of magnitude. As will be demonstrated in this section, Shannon Information processing has been found by the authors to effectively mitigate these challenges.
In order to assess the efficacy of Shannon Information processing of multiple skip phased array data, a demonstration pipe from Holloway NDT and Engineering (Georgetown ON, Canada) featuring two corrosion defects was scanned using a 5 MHz center frequency, 32 element probe (model 5L32-A31) from Evident Scientific (Quebec City, Canada) coupled to an Evident Scientific Rexolite shear wave refracting wedge (model SA31-N55S). Data was acquired via a Veo Phased Array Ultrasonic Pulser from Sonatest Ltd. (Quebec City, Canada). The corrosion features in the test pipe were produced by machining actual touch point corrosion flaws extracted by laser profilometry into a stock piece of NPS 8, schedule 40 pipe. The phased array sector scan included 41 beams (unfocused, 32 element aperture) spanning refracted...
angles between 50 and 70 degrees. A 700 mm scan covering the entire pipe’s circumference was performed - this includes both corrosion features and 375 mm of flawless length for reference - see Fig. 10.

Using the impact point and refracted angle for each beam, the corresponding signal was interpolated along the x direction, for each of the $M = 700$ scan positions in order to obtain $L = 41$ C-scan images ($x - z$ plane), representing a top view of the component. Prior to interpolation, each A-scan signal was scaled such that the RMS noise value computed from the unflawed reference scan positions marked in Fig. 10 was 10% full screen height. To generate the Shannon Information processed C-scans, the scaled A-scans were transformed into analytic signal format and then the background distribution parameters for the real and imaginary parts of the analytic signal were estimated using the unflawed reference region. The Shannon information processed analytic signals were interpolated along the x axis at each scan po-

Figure 8: ToFD Scan Data: Comparison Between Estimated and Theoretical Null Distributions; (a) Null Probability Density Functions for Shannon Information Processed Signals, (b) Null Cumulative Distribution Functions for Shannon Information Processed Signals

Figure 9: Comparison Between Shannon Information Processed B-Scans with Background Distribution Parameters Estimated From: (a) The Unflawed Reference Region, (b) The Entire Scan Length
Figure 10: Multiple Skip Phased Array Scan of Inaccessible Corrosion Flaws - Locations of Flaws and Unflawed Reference Regions (Yellow Boxes) Annotated; (a) Top View, (b) Side View

sition. Finally, the unprocessed combined C-scan projection view was formed by taking the maximum intensity over the 41 C-scan images interpolated on $x - z$ using the individual beams. The processed combined C-scan projection view was formed by taking the modulus of the sum of the 41 Shannon information processed C-scan images interpolated on $x - z$ using the individual beams (Eqn. 8). The wall loss for each corrosion feature was measured using a laser profiling system from Creaform (Laval Canada) - these wall loss maps are plotted together with the unprocessed and Shannon Information processed C-scans in Fig. 11.

Here the dynamic range of the Shannon Information processed C-scan data was limited by saturating values above 10 times the RMS noise value estimated from Eqn. $19 = 10 \times \sqrt{\frac{24}{41}} = 56$ Harts. From Fig. 11, the deepest part of the first corrosion feature (red box) is readily identified in the Shannon Information processed C-scan (c), while this same feature is almost completely invisible in the unprocessed data. Similarly, the most severe corrosion on the second feature (yellow box) is clearly detected on the Shannon Information processed C-scan while only faintly visible in the unprocessed data. Some of the pitting on the second corrosion feature (purple box) which is closer to the probe is visible in the unprocessed C-scan, however, these same pits are more clearly defined on the Shannon Information processed scan. The weak reflections highlighted by Shannon Information processing must, of course, be present in the unprocessed data, however, it would be difficult to identify them in any systematic way. Note that reflection amplitude is not well correlated with the depth of corrosion: the first corrosion feature (smooth) barely registering in the unprocessed scan despite extending nearly through wall, while the shallower, more sharply textured second corrosion feature reflects more strongly. It is therefore essential that small reflections can be reliably detected as they are potentially associated with serious corrosion flaws.
Similar to the Shannon Information processed ToFD data, the processed Multiple-Skip PA scan shows progressively less background noise when the significance level threshold is increased - see Fig. 12 - with only the relevant scan features remaining in the 99.9\textsuperscript{th} percentile (α = 0.001).

Referring to the laser profiled remaining wall map in Fig. 11 (a), the areas of corrosion on both features above 20% of the nominal thickness are readily identified on the Shannon Information processed data, thresholded in the 99.9\textsuperscript{th} percentile (Fig. 12 (c)), despite the fact that the regions of the simulated corrosion profiles are relatively smooth (Feature 1 in particular). The minimum wall loss percentage detectable by the various techniques described in [24] is 20%. The proposed Multiple Skip PA technique appears to provide a similar level minimum detection level as the these other techniques when the scan data is Shannon Information pro-
cessed. For corrosion under pipe supports (CUPS) inspections in particular, an unflawed reference region is readily available - the portion of the pipe’s circumference off of the support is unlikely to feature any corrosion. Shannon Information processing can also be applied to other pulse/echo techniques such as HOMC/A1 in order to highlight weak corrosion reflections.

The null-distribution of Shannon Information processed A-scans and C-Scan projection view estimated from the unflawed reference scan length are compared with the PDFs and CDFs in Fig. 13.

Here, the PDF and CDF estimated from the unflawed region are observed to differ slightly from the theoretical values computed from Eqns. 12 and ?? - see Figs 13 (a) and (b). This is likely due to the fact that the background noise itself appears to be bimodally distributed with distinct bands of higher and lower intensity, as seen in Figs. 11 (b) and (c). This non-random amplitude variation may be due to non-uniformity in material properties associated with the manufacturing process of seamless pipe [25]. Deviation from the null-distribution for Shannon Information processed signals is apparently compounded in forming the C-scan projection view. Figures 13 (c) and (d) show that the sample taken from the unflawed reference region differs appreciably from the
null distribution for the modulus of the sum of Shannon Information processed analytic signals with PDF and CDF defined by Eqns. 14 and 15, respectively. In any case, the theoretical and empirical distributions tend to converge on the right tail which is important for computing the significance thresholds $v_\alpha$. Furthermore, the RMS noise value computed from the unflawed reference sample was very close to the value predicted from Eqn. 19 ($5.14 \approx 2\sqrt{\frac{\ln 10}{\ln 10}} = 5.56$ Harts). Therefore, at least for the purposes of defining an appropriate threshold and dynamic range for C-scans generated from Shannon Information processed analytic signals, Eqns. 17 and 19 appear to be sufficiently accurate, despite a noticeable difference between the measured and theoretical null-distributions attributed to non-random sources of amplitude variation.

### 3.3. High Temperature Ultrasonic Inspections

Many critical components requiring periodic inspection, particularly in the petrochemical industry, operate at elevated temperatures ($> 200^\circ$ C). As a result, asset owners will typically perform inspections on these components during scheduled maintenance shutdowns which may add to the total time that the components are out of service and restricts the time between inspections to once or twice per year. If problems arise between scheduled shutdowns, an emergency shutdown may be required to perform an inspection which can be especially costly. In order to avoid these issues, ultrasonic inspections may be performed at elevated temperatures by using specialty designed wedges, probes and couplant [26]. A low SNR complicates detection of any type of damage - this should be especially true for subtler forms of damage, such as small cracks or High Temperature Hydrogen Attack (HTHA). The increased levels of background noise present in scans performed at elevated temperatures are due to random sources and therefore should be effectively mitigated by Shannon Information processing.

In order to investigate whether Shannon Information processing can facilitate the interpretation of ultrasonic data acquired at elevated temperatures (high background noise), a 101.6 mm (4 inch) thick carbon steel test piece, featuring 1.5 mm flat bottom hole (FBH) targets drilled to various depths was heated 280 $^\circ$ C and then scanned using a custom designed, high temperature wedge - see Figure 14. The wedge material is a special heat-resistant polymer called polybenzimidazole (tradename Duratron) [26]. A dual linear configuration was employed to keep the travel path in the wedge to a minimum without introducing strong wedge multiples. Reducing the travel path in the wedge helps to reduce amplitude losses due to high material attenuation in the wedge at high temperature. A matched pair of 60 element, 7.5 MHz center frequency probes with a 1 mm pitch from Olympus/Evident Scientific (Quebec City, Canada) are mounted at a slight angle (5.3 degrees with respect to the z axis) such that their beams overlap in the passive plane (scan direction). The acquisition sequence involved pulsing all elements on the transmitting probe together to generate a plane wave which can reflect from discontinuities in the test piece - these are recorded by the individual elements on the receiving probe. Once acquired, the $L = 60$ elementary A-scans are processed using a delay and sum approach (B-mode imaging). Prior to scanning the block was placed on a hot plate set to 280 $^\circ$ C and left until the top surface temperature reached a steady state ($255^\circ$ C). The scan was performed using a TempMaster high
temperature scanning system from Eclipse Scientific (Owen Sound, Canada). Data was acquired at 1 mm increments with a Zircon ultrasonic phased array system from Zetec (Quebec City, Canada). Ultrasonic velocities in the wedge and piece materials were estimated according to [26] and [27], respectively.

The A-scans were converted to analytic signal representation and then Shannon Information processed prior to rendering the DaS images at each scan location. Due to the relatively large area over which the transmitter and receiver beams spread overlap and the short length of the test block (block dimensions were minimized to facilitate heat transfer from the hot plate into the piece), there was no significant scan length observed to be completely free of target indications (for all y). Consequently, no reference region was specified in estimating the parameters of the background distributions, rather data from the entire scan length was used. For the volume rendered from unprocessed A-scans, amplitude was scaled such that the RMS noise in a reference region between the entry surface reflection and the deepest FBH (10 mm ≤ y ≤ 75 mm, −30 mm ≤ x ≤ 30 mm, 25 mm ≤ z ≤ 125 mm) - see the yellow box in Fig. 14 - was 10 %. For the volume rendered using Shannon Information processed A-scans, the saturation level was set to $10 \times \text{noise} = \frac{2 \sqrt{10}}{\ln 10} \approx 67$ Harts (using Eqn. 19).

A comparison between the scans rendered with and without Shannon Information processing applied to the A-scans is shown in Fig. 15 - these represent a slice through the volume along the (0, y, z) plane (beneath the center of the array). The DaS images (x, y) plane at scan positions where the FBH indications were maximized are shown in Fig. 16.

From Figs. 15 and 16, Shannon Information processing is observed to almost completely remove all indications appearing consistently through the scan axis, namely, the entry surface, wedge multiples (not entirely suppressed by the dual wedge design) and backwall indications. Backwall suppression allows the detection of the shallowest FBH target, which cannot be separated from the backwall in the unprocessed data. Detection of small flaws close to specimen boundaries is of great practical interest. The authors have previously suggested using Vector Coherence Imaging (VCI), a Shannon Information derived metric of the Phase Coherence Imaging technique originally described by Camacho et. al. [28], [29]), to highlight tip diffracted/scattered signals with respect to the backwall [30]. The VCI approach tends to highlight omnidirectional scatterers while partially suppressing specular reflections, whether they are relevant or not - consequently, VCl's of large planar flaws can be distorted and de-emphasized [30]. The proposed Shannon information processing scheme appears to be a promising alternative method of identifying small flaws close to consistent boundary reflections, which avoids the risk of also suppressing relevant specular reflections.

As was the case with the ToFD and multiple skip phased array data, the SNR of all targets was found to improve considerably through application of Shannon Information processing - this, in spite of the fact that an unflawed reference region was not available to estimate background distribution parameters. Furthermore, from Fig. 16, the FBH indications are found to be equally well resolved and positioned when the A-scans are Shannon information processed prior to DaS rendering. The SNR and height sizing accuracy are quantitatively compared between the processed and unprocessed scans in Table 3.5. Here, flaw heights were estimated by subtracting the $y$ coordinate position where the FBH is maximized from the peak position of the backwall (from the unprocessed data), $y = 102.5$ mm. For the purposes of computing the SNR values for the Shannon information processed images, the background noise was estimated to be 6.53 Harts from the yellow boxed region (for consistency with the unprocessed results) shown in Fig. 14, which is very close to the value computed from Eqn. 19 (6.73 Harts).

From Table 3.3, the flaw height estimates for the unprocessed and processed data were found to be highly consistent. As was seen in Figs. 15 and 16, the SNR values were found to be much higher when Shannon Information processing is applied to the A-scans prior to DaS imaging. In fact, without Shannon information processing, the 3:1 (9.5 dB) minimum SNR standard for detection [1] is only met for 3 of the 5 targets.

Identification of the FBH targets is made easier by applying a significance threshold to the Shannon information processed data, as shown in Fig. 17. As was observed with the ToFD and Multiple-Skip PA scans, retaining only scan features in the 99.9th percentile removes nearly all background noise and irrelevant indications.

Comparing empirically derived PDFs and CDFs to Eqn. 12, 13, 14 and 15 (see Fig. 18), the theoretical distribution is found to agree favourably with the variation in the experimental data. The samples used to obtain the empirical distributions shown in Fig. 18 span the entire inspected volume, excluding depths where the FBH reflections are observed, that is, for $V_{ub}$, the sample was taken as bound by: $-30 \text{mm} \leq x \leq 30 \text{mm}, 25 \text{mm} \leq z \leq 125 \text{mm}, 0 \text{mm} \leq y \leq 75 \text{mm}$ or $101 \text{mm} \leq y \leq 110 \text{mm}$ and for $S_0$ the sample was taken as bounded by $0 \leq t \leq 60, 25 \text{mm} \leq z \leq 125 \text{mm}, 0 \mu s \leq t \leq 25.42 \mu s$ or $34.24 \mu s \leq t \leq 37.28 \mu s$. This region was chosen to encompass grain noise as well as the consistent scan features (entry surface, wedge reverberations, backwall) in the Shannon information processed samples.

Interestingly, the scan of the block performed at elevated temperature does not feature any unflawed length and yet only 34 % of the imaged volume is above the theoretical RMS noise value predicted by Eqn. 19. Development of an appropriate compression scheme for Shannon information processed scan data is outside the scope of this work, however,
the fact that a large percentage of the digitized scan data can be safely discarded (no relevant scatterer can be justifiably be considered detectable if it registers below the noise floor) suggests that Shannon information processed data is highly compressible.
<table>
<thead>
<tr>
<th>35 mm</th>
<th>53 mm</th>
<th>73 mm</th>
<th>93 mm</th>
<th>115 mm</th>
</tr>
</thead>
</table>

Wedge
Backwall

<table>
<thead>
<tr>
<th>Entry Surface</th>
<th>Backwall</th>
<th>Entry Surface</th>
<th>Wedge Reverbs</th>
</tr>
</thead>
</table>

Figure 15: Scan of Test Block at 280°C: Side View Comparison on Plane (0, y, z); (a) DaS Image Rendered with Unprocessed A-Scans, (b) DaS Image Rendered with Shannon Information Processed A-Scans. The FBH Indications are Boxed.
Figure 16: Scan of Test Block at 280°C: Comparisons of DaS Images Rendered with and without Shannon Processing of A-Scans; (a) Unprocessed - (x, y, 35), (b) Unprocessed - (x, y, 53), (c) Unprocessed - (x, y, 73), (d) Unprocessed - (x, y, 93), (e) Unprocessed - (x, y, 115), (f) Processed - (x, y, 35), (g) Processed - (x, y, 53), (h) Processed - (x, y, 73), (i) Processed - (x, y, 93), (j) Processed - (x, y, 115)
Figure 17: Scan of Test Block at 280°C: Side View Comparison on Plane (0, y, z) with Different Significance Thresholds; (a) $v_{x=0.1} = 10.2$ Harts (90th Percentile), (b) $v_{x=0.01} = 14.4$ Harts (99th Percentile), (c) $v_{x=0.001} = 17.7$ Harts (99.9th Percentile)
Figure 18: Scan of Test Block at 280°C: Comparison Between Estimated and Theoretical Null Distributions; (a) Null Probability Density Functions for Shannon Information Processed Signals, (b) Null Cumulative Distribution Functions for Shannon Information Processed Signals, (c) Null Probability Density Functions for DaS Images (Rendered with Shannon Information Processed Signals), (d) Null Cumulative Distribution Functions for DaS Images (Rendered with Shannon Information Processed Signals)
4. Conclusion

A simple information theoretic method of representing ultrasonic scan data in terms of "unusualness" or "surprisal" with respect to an unflawed reference has been proposed. So-called "Shannon Information" processing was found to facilitate detection of small flaws, particularly when they appear close to consistent, non-relevant scan features and generally increase SNR - this was demonstrated using several examples of defects which would otherwise be challenging to detect. Shannon Information processed signals (in analytic signal representation) from beams overlapping spatially (redundant sources) can be combined by summation with the result representing how surprised one should be to observe all signal amplitudes by random chance. This forms the basis upon which projection views and Delay and Sum images computed using Shannon Information processed signals can be interpreted. Null distributions for Shannon Information processed signals and the modulus of the sum of Shannon Information processed signals have been derived as a means of setting global significance based thresholds, estimating RMS noise values and determining appropriate dynamic range limits for analysis.

In addition to the reported benefits of applying Shannon information processing, the authors view the technique as a formalization of the implicit anomaly detection process performed by NDT technicians and further speculate that the Shannon information representation may in fact be a more sensible way to display and analyze ultrasonic inspection data than the common practice of directly interpreting absolute signal amplitude, which is somewhat inflexible and highly reliant on the distance amplitude correction/time corrected gain calibration process.

Acknowledgement

This work was supported by Acuren Industrial Services. The CUPS sample scans were performed by Paul Holloway and Nolan Jarv of Holloway NDT and Engineering. Editorial input/guidance by Ed Ginzel of the INDT Lab and the University of Waterloo. Software access and support was provided by Sonatest Ltd. Special thanks to Mike Matheson of Eclipse Scientific for his continuing commitment to NDT research.

Appendix A. Derivation of Null Distributions

Consider a random variable, \( P_0 \) representing the one tailed p-value under the null hypothesis, then \( P_0 \sim \text{Unif}[0,1) \) [31]. From Eqn. 4 we can get a relationship between the signed Shannon information under the null hypothesis, \( S_0 \), and \( P_0 \):

\[
S_0 = \begin{cases} 
\log_b [2P_0], & S_0 \leq 0 \\
\log_b [2(1-P_0)], & S_0 > 0 
\end{cases}
\] (A.1)

For \( S_0 \leq 0 \), then, the CDF of \( S_0 \) is derived as follows:

\[
F_{S_0}(u) = \Pr [S_0 \leq u] = \Pr \left[ P_0 \leq \frac{u}{2} \right]
\] (A.2)

\[
F_{S_0} u = \int_0^{b u / L} f_{P_0}(p) dp = \frac{b u}{2}
\] (A.3)

Similarly, for \( S_0 > 0 \), the CDF of \( S_0 \) is given by:

\[
F_{S_0}(u) = \Pr [S_0 \leq u] = \Pr \left[ P_0 \leq 1 - \frac{u}{2} \right]
\] (A.4)

From Eqn. 15, the PDF of \( S_0 \) is obtained by taking the derivative of \( F_{S_0}(u) \) with respect to \( u \) [17]:

\[
f_{S_0}(u) = \frac{\log_b [b u]}{2} \frac{1}{b u}, \quad S_0 \leq 0
\] (A.5)

Following the same process for the signed Shannon information representation of the Hilbert transform, starting with Eqn. 5, gives the CDF and PDF for \( \hat{S}_0 \) under the null-distribution, \( \hat{S}_0 \):

\[
F_{\hat{S}_0}(u) = \begin{cases} 
\frac{\log_b [b u]}{2}, & \hat{S}_0 \leq 0 \\
\frac{\log_b [b u]}{2}, & \hat{S}_0 > 0 
\end{cases}
\] (A.6)

\[
f_{\hat{S}_0}(u) = \begin{cases} 
\frac{\log_b [b u]}{2} b u, & \hat{S}_0 \leq 0 \\
\frac{\log_b [b u]}{2} b u, & \hat{S}_0 > 0 
\end{cases}
\] (A.7)

To verify the derived PDF and CDF for \( S_0 \), a large sample of p-values was simulated and used to generate random samples from \( S_0 \) using Eqn. A.1, from which a histogram and empirical distribution function are obtained and plotted against Eqs. A.5 and A.4 in Fig. A.19.

The simulated histograms and empirical distribution functions shown in Fig. A.19 confirm the derived form of the PDF and CDF for \( S_0 \).

Now we define, \( V_{0L} \) as follows:

\[
V_{0L} = \left| \sum_{i=0}^{L-1} S_0 + \frac{1}{L} \sum_{i=0}^{L-1} S_0 \right|
\] (A.8)

Where, \(| \cdot |\) denotes the complex modulus of the argument. The random variable, \( V_{0L} \), represents the variation in projection view or DaS image amplitudes rendered by the summation of Shannon information processed signals in analytic signal form. By the central limit theorem, the sum of \( L \) values drawn from \( S_0 \) and the sum of \( L \) values drawn from \( \hat{S}_0 \) are normally distributed for large \( L \) [17]:

\[
\frac{1}{L} \sum_{i=0}^{L-1} S_0 \sim \mathcal{N}(E[S_0], LE[(S_0 - E[S_0])S_0])
\] (A.9)
The modulus of a complex random variable whose real and imaginary components are independent and identically distributed according to $\mathcal{N}(0, \sigma^2)$ is itself distributed according to a Rayleigh distribution [32]. Therefore, $V_0L \sim \text{Rayleigh}(\sigma = \sqrt{\frac{2L}{\ln b}})$, where the PDF and CDF of $V_0L$ have the following forms [32]:

$$f_{V_0L}(u) = \frac{\ln^2(b)u}{2L} e^{-\frac{\ln^2(b)u^2}{4L}}$$  \hspace{1cm} (A.13)

$$F_{V_0L}(u) = 1 - e^{-\frac{\ln^2(b)u^2}{4L}}$$  \hspace{1cm} (A.14)

Simulated histograms and empirical distribution functions shown in Fig. A.20 confirm the derived form of the PDF and CDF for $V_0L$. 

Figure A.19: Comparison of Simulated and Theoretical Distribution of $S_0$ ($b = 10$): (a) Simulated Histogram vs. PDF from A.5, (b) Simulated Empirical Distribution Function vs. CDF from A.4

$$\sum_{i=0}^{L-1} \hat{S}_0 \sim \mathcal{N} \left( \mathbb{E} \left[ \hat{S}_0 \right], L \mathbb{E} \left[ \left( \hat{S}_0 - \mathbb{E} \left[ \hat{S}_0 \right] \right)^2 \right] \right)$$  \hspace{1cm} (A.10)

Where, $\mathbb{E}[\cdot]$ is the expectation operator. From Eqns. A.5 and A.7, the expectations defining the location and scale of these normal distributions are calculated as follows:

$$\mathbb{E} [S_0] = \mathbb{E} [\hat{S}_0] = \int_{-\infty}^{\infty} u f_{S_0}(u) du = 0$$  \hspace{1cm} (A.11)

$$\mathbb{E} \left[ (S_0 - \mathbb{E} [S_0])^2 \right] = \mathbb{E} \left[ (\hat{S}_0 - \mathbb{E} [\hat{S}_0])^2 \right]$$

$$= \int_{-\infty}^{\infty} u^2 f_{S_0}(u) du$$

$$= 2 \int_{0}^{\infty} u \frac{\ln b}{2} b^u du$$

$$= \frac{2}{\ln^2 b}$$  \hspace{1cm} (A.12)

The modulus of a complex random variable whose real and imaginary components are independent and identically distributed according to $\mathcal{N}(0, \sigma^2)$ is itself distributed according to a Rayleigh distribution [32]. Therefore, $V_0L \sim \text{Rayleigh}(\sigma = \sqrt{\frac{2L}{\ln b}})$, where the PDF and CDF of $V_0L$ have the following forms [32]:

$$f_{V_0L}(u) = \frac{\ln^2(b)u}{2L} e^{-\frac{\ln^2(b)u^2}{4L}}$$  \hspace{1cm} (A.13)

$$F_{V_0L}(u) = 1 - e^{-\frac{\ln^2(b)u^2}{4L}}$$  \hspace{1cm} (A.14)

Simulated histograms and empirical distribution functions shown in Fig. A.20 confirm the derived form of the PDF and CDF for $V_0L$. 

The simulated histograms and empirical distribution functions shown in Fig. A.20 confirm the derived form of the PDF and CDF for $V_0L$. 

25
Figure A.20: Comparison of Simulated and Theoretical Distribution of $V_0_L$ ($b = 10$): (a) Simulated Histogram vs. PDF from A.13 - $L = 40$, (b) Simulated Empirical Distribution Function vs. CDF from A.14 - $L = 40$, (c) Simulated Histogram vs. PDF from A.13 - $L = 60$, (d) Simulated Empirical Distribution Function vs. CDF from A.14 - $L = 60$, (e) Simulated Histogram vs. PDF from A.13 - $L = 80$, (f) Simulated Empirical Distribution Function vs. CDF from A.14 - $L = 80$
References


