Persistent Homology for Defect Detection
in Non-Destructive Evaluation of Materials

Jose F. CUENCA¹, Armin ISKE¹

¹Department of Mathematics, University of Hamburg, Germany
jose.fernando.cuenca@gmail.com, armin.iske@uni-hamburg.de

Abstract. Non-Destructive Testing (NDT) methods enable us to inspect, examine and evaluate an object, material or system without compromising its future use. The detection of welding defects in steel tubes is one of the most common applications of NDT methods, particularly for ultrasonic testing. We propose a novel concept for NDT, which relies on persistent homology, a recently developed tool for analyzing topological properties of manifolds from discrete data. This leads us to an automated interpretation of ultrasonic images in the NDT technique “Time of Flight Diffraction” (TOFD). The resulting techniques permit automated and fast classification of different types of defects, including crack propagations and pore defects, without human interaction.

1. Introduction
The Time of Flight Diffraction Method (TOFD) has extensively been used for automated weld inspection, especially in applications of steel industry, where welding processes are rather important to guarantee the quality of pipelines. There are several reasons that make TOFD a suitable NDT technique. Firstly, it provides sizing accuracy and characterizations of welding defects. Secondly, TOFD makes only a small rate of false indications. On the other hand, TOFD data interpretation and defect classification is often done manually by human operators. In order to reduce the production costs and to further reduce false indications, automated techniques are required to accelerate the defect classification. To this end, image processing methods are used for defect characterization [1, 2, 3] and other methods are used for the placement of probes in the material [4]. Moreover, in the interpretation of TOFD data, for defect detection and classification, Hough transforms are used [5, 6] to propose matched filtering for parabola detection [7]. Alternative approaches use mathematical morphology [8] or artificial neural networks [9, 10].

This paper proposes to use the novel concept of persistent homology for defect detection and classification. Persistent homology has quite recently gained enormous popularity and is regarded as a very powerful tool for analyzing topological features of big data sets [11, 12, 13], especially for noisy data [14]. Relevant applications are given in [15, 16, 17, 18, 19, 20, 21, 22]. The basic idea of persistent homology is to generate a filtration, i.e., a one-parameter family of simplicial complexes in Point Cloud Data (PCD), for the purpose of keeping track of the “birth” and “death” of topological features [15], such as e.g. connected components, holes, tunnels and their higher-dimensional relatives. There are several methods to create suitable filtrations, such as Vietoris-Rips complexes [23], alpha complexes [24], Lazy-Witness complexes [25], or Čech complexes [26]. Software packages to compute the persistent homology of a PCD include Perseus [27] (based on [28]), Javaplex [29], Dionysus [30], and Phom [31] (based on [13]).
The outline of this paper is as follows. In Section 2, we explain the TOFD method. Then, in Section 3 we introduce the basic concept of persistent homology, where we discuss their key objects and features, such as simplices, simplicial complexes, filtrations, persistent diagrams and barcodes. In Section 4 we propose a novel defect classification method relying on persistent homology, especially for the classification of defects in TOFD data. Numerical results are presented in Section 5, for two simulated cases and one real case of TOFD data. A final conclusion is given in Section 6.

2. The Time of Flight Diffraction Method (TOFD)

The TOFD is a NDT technique that was created in 1977 by Silk [32]. While most ultrasonic NDT methods rely on the echoes amplitude, TOFD is based on the travel time of the diffracted waves at a defect or discontinuity. The basic set-up of a TOFD machinery, as shown in Figure 1, works as follows. Two probes are used, one emitter and one receiver. These transducers are usually piezo-elements, i.e., the devices excite an ultrasonic wave to convert mechanical stress into electricity due to the piezoelectric effect (see [33] for details). Having placed the probes into the inspected weld, an ultrasound is sent through the material. If there is one defect inside, the signal will diffract to the receiver due to the Huygens principle (the bottom and top of the defect act as a source of diffracted waves).

![Figure 1. Time of Flight Diffraction method (TOFD)](image)

Moreover, there are other signals arriving at the receiver:

- Lateral wave: the fastest wave, travelling directly from emitter to receiver.
- Back wall echo: the reflection of the lower boundary of the component.
- Diffractions from inhomogeneities of the material (noise).
- Reflections from defect surface, of low amplitude; they usually disappear.

For every fixed probe position, an a-scan (an ultrasound representation) is obtained, giving a relation of amplitude versus time. Figure 2 shows a TOFD a-scan. To acquire more data, probes are moved in parallel to the weld and measurements are taken every time they are fixed. All combinations of a-scans generate a b-scan (see Figure 2), where existing defects may be indicated, which requires automated classification.

![Figure 2. Left to right: a-scan, b-scan, deconvolved b-scan (preprocessed data)](image)
Our classification method is based on the recognition of different shapes which defects may present in b-scans. More precisely, we focus on the distinction of two characteristic defects that appear in steel tubes: (a) \textit{planar flaws}, exhibiting the lack of fusion and crack propagations. These defects produce signals of linear shapes in the b-scan; (b) \textit{pore defects} or \textit{gas inclusions}. These defects produce signals of parabolic shapes in the b-scan (see Figure 2).

But TOFD data are usually very noisy. Therefore, b-scans need to be denoised in a preprocessing step. This will reduce the amount of data for the subsequent classification by persistent homology. We use the deconvolution method proposed in [34], where an \textit{Orthogonal Matching Pursuit} (OMP) algorithm is applied to ultrasonic testing. The aim of the deconvolution is to extract the strong signals or the peaks in every a-scan, thereby reducing the low amplitudes. The presence of remaining noise must be dealt with in a postprocessing step. Essentially, the defect classification can be viewed as a problem of shape recognition from noisy data (i.e., from the deconvolved b-scan). The processing steps are illustrated in Figure 2. For further information we refer to [34].

3. Persistent Homology

In this section, we explain basic ideas of persistent homology. For a more comprehensive account on topological data analysis we refer to [15, 35, 36].

Atomic objects in persistent homology are \textit{simplices}. Simplices are generalizations of triangles to higher dimensions, where an \textit{n-simplex} is an \textit{n}-dimensional polytope in $\mathbb{R}^n$, given by the convex hull of \(n+1\) vertices (see Figure 3). To be more precise, for a set $V = \{V_0, \ldots, V_n\} \subset \mathbb{R}^n$ of \(n+1\) points in $\mathbb{R}^n$, their \textit{n-simplex} $\Delta_n \equiv \Delta_n(V)$ is

$$\Delta_n = \left\{ V = \lambda_0 V_0 + \lambda_1 V_1 + \ldots + \lambda_n V_n : 0 \leq \lambda_j \leq 1 \text{ and } \sum_{j=0}^{n} \lambda_j = 1 \right\} \subset \mathbb{R}^n.$$

We will say that the simplex $\Delta_n$ has dimension $n$.

\begin{figure}[h]
\centering
\includegraphics[width=0.8\textwidth]{simplices.png}
\caption{0-, 1-, 2-, and 3- simplices: vertex, edge, triangle, tetrahedron}
\end{figure}

Now let us turn to \textit{simplicial complexes}. A simplicial complex $K$ is a finite set of simplices satisfying the following two conditions.

1. Every face of a simplex in $K$ belongs to $K$.
2. Any pair of two simplices in $K$ are either disjoint or they share a common face.

For $\tau \in K$, we call any subset $\sigma$ of $\tau$ a \textit{face} of $\tau$, $\sigma \preceq \tau$. Moreover, for any $d \geq 0$, we denote by $K_d$ the set of $d$-dimensional simplices in $K$. 
For any $N \in \mathbb{N}$, a filtration $F$ of a simplicial complex $K$ is given by a nested sequence of simplicial subcomplexes $F_nK \to K$, for $n \in \{0, \ldots, N\}$, satisfying

$$\emptyset = F_0K \to F_1K \to \ldots \to F_{N-1}K \to F_NK \to K,$$

where $N$ is said to be the length of the filtration.

A commonly used filtration is the Vietoris-Rips complex. We can explain the construction of a Vietoris-Rips complex as follows. First, a set $X = \{x_1, \ldots, x_N\}$ of $N$ vertices is fixed. Initially, the vertices in $X$ are not connected, and so $X$ can be viewed as a point cloud. Then, a set of $N$ balls, each centered at one vertex in $X$, is taken, where the radii of the balls are all equal. Initially, the radii of the balls are small enough, so that no pair of two balls intersect. Now, to construct the Vietoris-Rips complex from $X$, the (constant) radii of the balls are gradually increased. Whenever two balls intersect, a 1-simplex (line) is created by the connection of their two centers. As soon as three 1-simplices form a closed loop, one 2-simplex (triangle) is created, and so on for $n$-simplices of higher dimension (see Figure 4 for illustration).

![Figure 4. Construction of a Vietoris-Rips filtration. 3 steps with different radii.](image)

For further illustration, Figure 5 shows an example for a Vietoris-Rips complex for a larger point cloud $X$. Note that in the filtration of the Vietoris-Rips complex of Figure 5, holes in graph appear at smaller radii (top right), where the disappear later for larger radii (bottom left), at different steps of the filtration.

![Figure 5. 4 steps of a filtration, Vietoris-Rips complex, from random PCD [37].](image)

Next we explain how tools from algebra can be used to describe the topological invariants in a simplicial complex, in particular the number of connected components.
and the number of holes. For any simplicial complex $K$, we can calculate its homology groups, denoted by $H_n(K)$, where $H_n(K)$ gives the number of $n$-dimensional holes in $K$. Let $\phi : K \rightarrow L$ be a simplicial map, then homology produces group homomorphisms $\phi_n : H_n(K) \rightarrow H_n(L)$. To further explain this, refer to the simplicial complex in Fig. 6.

**Figure 6. A simplicial complex $K$ with 0-simplices $x, y, z$ and 1-simplices $a, b, c$.**

Now the $n$-dimensional chain group $C_n(K)$ of $K$ is given by all linear combinations of $n$-dimensional simplices. For example, in Figure 6 one element in $C_1(K)$ is

$$\alpha a + \beta b + \gamma c \quad \text{for } \alpha, \beta, \gamma \in \mathbb{R}.$$  

Therefore, a 0-chain is a set of vertices, a 1-chain is a set of line-segments.

The boundary operator of an $n$-dimensional simplex $\sigma$ is an $n-1$ dimensional chain that can be calculated by adding its faces:

$$\partial_n(\sigma) = \sigma_0 - \sigma_1 + \ldots + (-1)^n \sigma_n,$$

and so $\partial_n$ defines a linear transformation $C_n(K) \rightarrow C_{n-1}(K)$.

In the example of Figure 6, we find $\partial(x) = \partial(y) = \partial(z) = 0$ and

$$\partial(a) = y - x \quad \partial(b) = z - y \quad \partial(c) = x - z.$$

A cycle is, from a geometric viewpoint, a closed loop (e.g. $a + b + c$ in Figure 6). Algebraically, the cycle group $Z_n(K)$ is the subspace generated by the kernel of the boundary operator $\partial_n$, and the boundary group $B_n(K)$ is the subspace generated by the image of the boundary operator $\partial_{n+1}$.

For the example of Figure 6, we make the following observations.

- The boundary of a vertex is 0, and so $Z_0$ is the space generated by the 3 vertices,

$$Z_0 = \ker(\partial_0) = C_0 = \langle x, y, z \rangle.$$

- the image of $\partial_1$ is the subspace generated by the boundaries of the edges $a, b, c$,

$$B_0 = \text{im}(\partial_1) = \langle y - x, z - y, x - z \rangle.$$

Now $Z_1 = \ker(\partial_1)$ can be computed by the solution of $\partial_1(\alpha a + \beta b + \gamma c) = 0$, i.e.,

$$\alpha(y - x) + \beta(z - y) + \gamma(x - z) = 0,$$
leading to $\alpha = \beta = \gamma$ and so $Z_1 = \ker(\partial_1) = \langle a + b + c > = \mathbb{R}$. Finally, $B_1 = \text{im}(\partial_2) = 0$, since there is no 2-simplex in the simplicial complex $K$.

Now the $n$-dimensional homology group is defined as the quotient of the cycle group and the boundary group,

$$H_n(K) = Z_n(K) / B_n(K).$$

For the example of Figure 6, the homology groups are

$$H_0 = Z_0 / B_0 = \langle x, y, z > / < y - x, z - y, x - z > = \mathbb{R},$$

$$H_1 = Z_1 / B_1 = \mathbb{R} / 0 = \mathbb{R}.$$

For each homology group, there is an associated Betti number, whose value is given by the dimension of the subspaces of each homology group. For the example of Figure 6, we have $\beta_0 = \beta_1 = 1$. Betti numbers measure the number of $n$-dimensional holes in a simplicial complex (e.g. connected components, tunnels, voids). For the example of Figure 6, where $\beta_0 = \beta_1 = 1$, the simplicial complex consists of one connected component and one 1-dimensional hole.

We remark that Betti numbers are sensitive w.r.t. perturbations in the simplicial complex, e.g., Betti numbers may change when adding only a single vertex (cf. Figure 7). To avoid such instabilities, persistent homology considers a filtration $F$ of a simplicial complex $K$. Betti numbers are essentially computed in each step of the filtration, keeping track of topological changes (e.g. for holes and connected components).

![Figure 7. By adding a vertex, the Betti Number is incremented by one.](image)

To explain this, a one-dimensional hole may be present in one step of the filtration. But in a later step, the hole may disappear. A collection of persistent intervals, monitoring the changes of the homology classes, when they appear (by “birth”) and disappear (by ”death”), leads us to a barcode representation of the filtration $F$ (see Fig. 8). The length of the interval measures the “lifetime” of a homology class across the filtration. We visualize this in a two-dimensional cluster of points, called persistent diagram, in which points close to the diagonal correspond to homology generators that disappear very soon across the filtration (e.g. noisy topological components), whereas points far off the diagonal have a long lifetime across the filtration and therefore correspond to stable features (see Fig. 8).

4. Persistent Homology and Classification of TOFD Defects

In this section, we explain how persistent homology can be used for the detection and classification of pore and crack defects from TOFD measurements.

Pore defects are represented by a parabolic shape in a TOFD b-scan. A typical case for a deconvolved b-scan with filtered back wall echo and lateral wave is shown in Figure 9a. To detect such parabolic features in noisy PCD, a horizontal line of points is
Figure 8. Barcode representation of $H_1(K)$ and persistent diagram of the PCD. Four steps of the filtration are shown. Stable holes correspond to long bars. The Betti numbers are equal to the number of bars intersecting the dashed line.

Figure 9. (a) Deconvolved b-scan including a parabolic defect; (b) a horizontal line is added and moved down; (c) small radii are needed to create a hole between the line and the parabola; (d) a large radius is needed to eliminate the hole.

added (see Figure 9b). Persistent homology is computed for the initial PCD. Then the line is moved down, and persistent homology is updated. This process is repeated until the whole image has been scanned. When the line intersects the parabola (see Figure 9b for $y = 8$), and when we start building the filtration by growing the balls around every vertex, a stable 1-dimensional hole will be created (see Figure 9c,d). This stable hole will be detected by persistent homology. As soon as the line has surpassed the parabola
(Figure 9a, for \( y = -2 \) or \( y = 10 \)), no stable holes will be detected. One-dimensional stable holes are generated by pore defects, whereas unstable holes are due to noise.

*Crack defects* are represented by a directed line. Figure 10 shows one example, where the crack is marked by blue dots. To detect the crack, we add data points lying on a rectangle, see Figure 10, left. When the rectangle does not intersect with the crack (e.g. Figure 10, left), persistent homology detects only one big stable hole. But as soon as rectangle arrives at a crack defect (see Figure 10, right), two stable holes are created.

![Figure 10. Two steps with adding points on rectangle. Left: one stable generator is detected; Right: two stable generators are detected by persistent homology](image)

5. Numerical Results

We have analyzed three different cases of TOFD defects. In two cases the data was simulated with the model proposed in [34]. To compute the persistent homology of the PCD we have used the software packages Perseus [27] and the Phom [31].

5.1. Test Case 1: Simulated TOFD Crack

A component of 20 mm thickness is simulated with the weld at \( z = 0 \). A crack is located at \((x = 40 - 75, y = -11, z = 0)\). Two probes, one emitter and one receiver, are moving along the weld, see Fig. 12b. The angle of aperture of the emitter is 12 degrees, other parameters of ultrasound are: bandwidth: 30 MHz, frequency: 5 MHz, shift = \( \pi/4 \). A-scans are taken every 0.5 mm, 300 in total. The velocity of ultrasound in steel tube is known and defined \((c_l = 5.92 \text{ mm/s} \text{ and } c_t = 3.23 \text{ mm/s})\). Gaussian noise is added to the geometrical model at a variance of 0.035. The deconvolved b-scan is shown in Figure 10 left (blue dots). The crack is detected very well, see Figure 11. The entire detection process takes less than 3 seconds CPU time only.

5.2. Test Case 2: Real TOFD Data with Lack of Fusion

The data (see Figure 13a) was obtained from a sample of a pipe with outer diameter 1.166 m and wall thickness 23.3 mm. The weld seam was tested with a 10 MHz transducer, 6 mm diameter (Olympus C563-SM). The angle of incidence is 70 degrees. The data exhibits a lack of fusion at the end of the pipe. A-scans were taken every 0.5 mm. Again, 300 a-scans were performed. Barcodes for the first homology group are shown in Fig. 14. The defect is detected very well.
5.3. Test Case 3: Simulated TOFD Data with Two Pore Defects

We used the same weld as in test case 1, but now with two pore defects, see Fig. 12a. We chose bandwidth 50 Mhz, and set the variance of the noise to 0.02. The pore defects are at $(0, -13.5, 0)$ and $(16, -12, 0)$. Our results are shown in Fig. 15. The two pore defects are detected at the jump from 0 to 2 stable holes in the barcode representation. This happens when the horizontal line intersects with the parabolic defects. The total CPU time for all computations is less than 3 seconds only.
Figure 14. Barcode representation of persistent homology from PCD for a lack of fusion. (b) no extra data is added, therefore no stable hole is detected; (c) blue rectangle of points added: one stable hole; (d) green rectangle of points added: one stable hole; (e) red rectangle of points added: two stable holes. The jump in the number of stable one-dimensional holes (from one to two) detects the defect.

Figure 15. Barcode of persistent homology from PCD for two pore defects. (C)(D): jump in the number of stable holes from 0 to 2. Short bars for holes due to noise.

6. Conclusion

We have proposed a new method for the automated detection and classification of defects in TOFD measurements, where persistent homology is used to detect pore and crack defects from deconvolved b-scans. As supported by our numerical examples, both types of defects are well-detected by the proposed method at small computational costs.
References


[27] Vidit Nanda. Perseus, the Persistent Homology Software. URL http://www.sas.upenn.edu/~vnanda/perseus
[34] Bossmann, F. Model Based Defect Reconstruction in Ultrasonic Non-Destructive Testing. Dissertation zur Erlangung des mathematisch-naturwissenschaftlichen Doktorgrades (University of Göttingen, 2013)