

ARTICLE OPEN



A machine learning framework for damage mechanism identification from acoustic emissions in unidirectional SiC/SiC composites

C. Muir¹, B. Swaminathan¹, K. Fields², A. S. Almansour³, K. Sevens⁴, C. Smith³, M. Presby³, J. D. Kiser³, T. M. Pollock¹ and S. Daly²✉

In this work, we demonstrate that damage mechanism identification from acoustic emission (AE) signals generated in minicomposites with elastically similar constituents is possible. AE waveforms were generated by SiC/SiC ceramic matrix minicomposites (CMCs) loaded under uniaxial tension and recorded by four sensors (two models with each model placed at two ends). Signals were encoded with a modified partial power scheme and subsequently partitioned through spectral clustering. Matrix cracking and fiber failure were identified based on the frequency information contained in the AE event they produced, despite the similar constituent elastic properties of the matrix and fiber. Importantly, the resultant identification of AE events closely followed CMC damage chronology, wherein early matrix cracking is later followed by fiber breaks, even though the approach is fully domain-knowledge agnostic. Additionally, the partitions were highly precise across both the model and location of the sensors, and the partitioning was repeatable. The presented approach is promising for CMCs and other composite systems with elastically similar constituents.

npj Computational Materials (2021)7:146; <https://doi.org/10.1038/s41524-021-00620-7>

INTRODUCTION

Silicon carbide/silicon carbide ceramic matrix composites (SiC/SiC CMCs) are structural ceramics with advantageous mechanical properties for use in extreme environments. As a result, CMCs have been increasingly incorporated in applications such as high-pressure turbine (HPT) shrouds in the hot section of aircraft engines, HPT vanes, and combustor liners, and are being considered as fuel cladding in generation IV nuclear reactors^{1–4}. In these and other applications, CMC failure is costly and dangerous. A detailed understanding of the failure envelope is critical for predicting component lifetimes and optimizing material performance; to this end, structural health monitoring (SHM) techniques are well suited for characterizing the damage state in these high-impact materials.

Acoustic emission (AE) is one well-established SHM technique that is widely used for nondestructive monitoring of damage accumulation across a range of material systems. AE captures the elastic waves that are produced when strain energy is rapidly accumulated and released from damage sources⁵. In modern practice, AE is used to triangulate the locations of damage sources⁵, identify highly damaged areas in the bulk⁶, and assess the severity of incurred damage⁷. As a characterization tool, the spatial resolution of AE for triangulating surface cracking in situ has been shown to be $\pm 100 \mu\text{m}$ ⁷, and the temporal distance between resolvable AE events is $100 \mu\text{s}$ ⁸; these exceed those of other techniques, such as eddy current⁹, electrical resistance¹⁰, and X-ray computed tomography (XCT)¹¹, making AE uniquely powerful for capturing evolving damage⁸.

Despite the rich scope of information that AE provides, the full range of its utility remains to be explored. Akin to how humans interpret variations in sounds to distinguish a tuba from a trombone, one longstanding hypothesis is that AE waveforms contain signal-specific features that can be used to identify the damage

mechanism(s) that generated them^{12–15}. In continuous fiber reinforced composites, this signal-specific identification is equivalent to mapping an AE signal to the matrix crack, interfacial damage, or fiber failure event that emitted it. The ability to identify a damage mechanism from its AE event has far-reaching advantages both in SHM and in predictive modeling. It would allow for the non-destructive, in situ identification of damage sources, enable researchers to bypass expensive high-resolution methods (such as scanning electron microscopy and XCT^{7,8,11}), and produce information-rich datasets for large-scale statistical analyses of the effects of small-scale damage in full composite structures.

However, damage mechanism identification in CMCs is non-trivial and has remained an elusive goal^{16,17}. In order to assign a damage mechanism to an AE waveform, the identification of patterns between many features (i.e., the use of a high-dimensional representation) is necessary¹⁶, which is infeasible to perform manually. Instead, this objective requires the use of machine learning (ML) techniques, which are capable of finding structure in datasets where objects are described by high-dimensional feature vectors^{18,19}.

AE features can be understood in terms of the orthotropic model for wave propagation in solids, where the out-of-plane (w) and in-plane (u^0, v^0) displacements are governed by²⁰:

$$D_{11} \frac{\partial^4 w}{\partial x^4} + 2(D_{12} + 2D_{66}) \frac{\partial^4 w}{\partial x^2 \partial y^2} + D_{22} \frac{\partial^4 w}{\partial y^4} = -\rho^* \frac{\partial^2 w}{\partial t^2} \quad (1)$$

$$A_{11} \frac{\partial^2 u^0}{\partial x^2} + A_{66} \frac{\partial^2 u^0}{\partial y^2} + (A_{12} + A_{66}) \frac{\partial^2 v^0}{\partial x \partial y} = \rho^* \frac{\partial^2 u^0}{\partial t^2} \quad (2)$$

$$(A_{12} + A_{16}) \frac{\partial^2 u^0}{\partial x \partial y} + A_{66} \frac{\partial^2 v^0}{\partial x^2} + A_{22} \frac{\partial^2 v^0}{\partial y^2} = \rho^* \frac{\partial^2 v^0}{\partial t^2} \quad (3)$$

¹Materials, University of California-Santa Barbara, Santa Barbara, CA, USA. ²Mechanical Engineering, University of California-Santa Barbara, Santa Barbara, CA, USA. ³NASA Glenn Research Center, Cleveland, OH, USA. ⁴Materials Science and Engineering, University of Michigan-Ann Arbor, Ann Arbor, MI, USA. ✉email: samdaly@ucsb.edu

with A_{ij} and D_{ij} functions of the orthotropic elastic constants and plate thickness (h), with ρ^* defined as:

$$\rho^* = \int_{-h/2}^{h/2} \rho(z) dz \quad (4)$$

In this model, the waveform structure is dictated by the elastic constants and plate density. It follows that, when the elastic properties of the constituents are disparate, AE waveforms are distinct in both the time and frequency domains²¹.

Over the past 30 years, many AE-ML frameworks have been developed around material systems that obey the constraint of constituent elastic dissimilarity, such as polymer matrix composites or elastically dissimilar brittle matrix composites (BMCs)^{21–24}. These frameworks typically encode waveforms with time-domain parameters that are easily extractable by the commercial software used to record the waveforms. ML techniques are then applied to partition the encoded signals based on the source damage mechanism^{21,25,26}. As the ground truth of partition membership is rarely known, the results must be manually interpreted to ensure that the ML algorithm has indeed discriminated signals by damage mechanisms. This workflow is demonstrated by Kostopoulos et al., who recorded waveforms in elastically dissimilar BMCs. Waveforms were encoded with time-domain features and sorted with the k -means algorithm. Their findings showed clear differences in cluster activity, which corresponded with mechanistic expectations for their simplified composite geometry²¹, allowing them to assign damage mechanism labels to clusters.

Despite the success of frameworks developed for elastically dissimilar composites, damage mechanism identification in SiC/SiC CMCs is more nuanced. In this system, where the elastic properties of the fiber and matrix are similar, the microstructural landscape drives waveform structure in the time domain^{14,16,27–31}. Commonly used time-domain parameters, like rise time/amplitude value, are then dictated by waveform propagation pathways (both as a function of propagation distance and the accumulated damage state in the material bulk) rather than the generating damage mechanism³². This relationship is detrimental to the discriminating power of AE-ML frameworks, as waveforms are sorted based on stochastic waveform distortions rather than the source damage mechanism³³.

It has also been proposed that, given their elastically similar properties, the SiC/SiC matrix and fiber will fracture in similar manners, thus preventing discrimination between damage mechanisms altogether³⁴. This hypothesis was predicated on well-established theories of plate-wave propagation^{20,27,35} and further substantiated by experimentation that found little or no discriminating power between matrix cracking and fiber failure^{34,36}.

In this work, it is instead hypothesized that the historic lack of discriminating power between the dominant damage mechanisms in SiC/SiC CMCs is a result of both the encoding scheme and algorithm choice rather than their constituent elastic properties. In addition to the improper use of time-domain features for encoding schemes, the majority of previous ML frameworks^{21,22,34} use the k -means algorithm, which can only find isotropic clusters. Because of the restrictive assumptions of k -means, it is often outperformed by algorithms that are less limiting^{16,37,38}. This restriction motivates the exploration of alternate waveform feature representations in conjunction with alternate ML algorithms for damage mechanism identification from AE.

Here a frequency-based AE-ML framework capable of damage mechanism identification in SiC/SiC minicomposites is presented. The minicomposite architecture was chosen for this demonstration as it exhibits a limited number of damage mechanisms, in which a well-established damage chronology enables a straightforward evaluation of clustering results. AE was recorded with a four-sensor experimental configuration (Fig. 1a) in three SiC/SiC

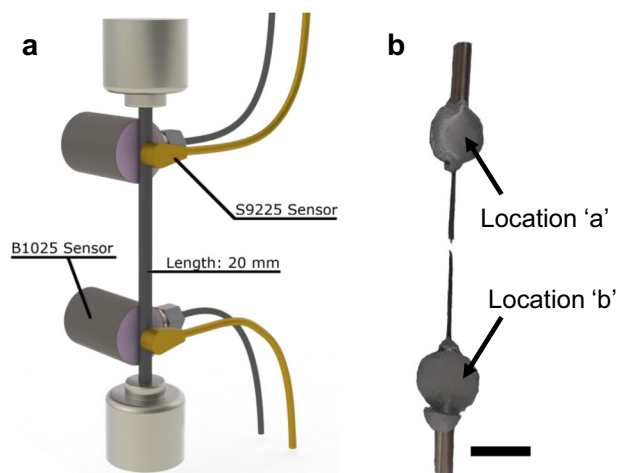


Fig. 1 Diagram of the experimental set-up. **a** Two sets of AE sensors were coupled to the minicomposite whose length is nominally 20 mm. When any sensor was triggered, all sensors began recording, ensuring that each AE event could be correlated between all sensors. **b** A photograph of a sample. Epoxy tabs used to mount AE sensors are denoted with arrows. Scale bar length is 10 mm.

specimens subjected to monotonic tensile loading. This configuration allowed for evaluation of framework precision, a metric of the labeling consistency. AE waveforms were encoded with a modified partial power scheme and partitioned according to their generating damage mechanism using the spectral clustering algorithm^{37,39}. We find distinct activity of clusters in the stress domain that follows the established damage chronology of minicomposites. This ML approach is shown to assign labels independently of the sensor used to record the AE signals. This sensor independence demonstrates that differences between the fundamental damage mechanisms drive label assignment, rather than the stochastic distortion of waveforms resulting from the experimental configuration. The framework developed herein can be more broadly applied to brittle composites whose constituents have elastically similar properties.

RESULTS

Unsupervised classification of acoustic spectra

AE data collection and preliminary filtering (described in “Methods”) represent Step 1 of the general unsupervised AE-ML framework, which proceeds as follows:

1. Experimentation: a number, n , waveforms are collected
2. Feature extraction: waveforms are represented in feature space by extracting d pertinent features (i.e., each waveform is represented by a d -dimensional vector)
3. ML algorithm: an algorithm is selected to partition waveforms into clusters that are representative of damage mechanisms
4. Labeling and error analysis: post-clustering analysis is performed to assign damage mechanism labels to clusters and assess the validity of results

At each step, considerations must be made to ensure the framework functions properly. The following section describes Steps 2–4. The code created for this investigation utilizes the Scikit-learn toolbox³⁹ and is available without restriction⁴⁰.

Feature extraction

After AE waveforms are collected, suitable representations to encode damage mechanism information must be determined. Appropriate waveform features are those that are more dependent on the generating damage mechanism than on extrinsic factors, such as propagation distance. Prior finite element analysis studies, supported by experimental evidence, indicate that partial power is one such feature, provided that signals are recorded in the near field^{31,41}.

Here AE waveforms are encoded with a modified partial power scheme, where the i th component of the feature vector is:

$$\text{Feature}_i = \frac{\int_{k_{i-1}}^{k_i} F[x(t)]dk}{\int_{k_0}^{k_d} F[x(t)]dk} \quad (5)$$

where $F[*]$ is the Fourier transform operator, $x(t)$ is the recorded signal, and d is the number of entries in the feature vector.

We set $k_0 = 200$ kHz and $k_d = 800$ kHz for all specimens. To determine the value of k_0 , a parametric sweep from $k_0 = 50$ to 350 kHz in increments of 50 kHz was conducted. The value-optimizing validity metrics was chosen. Including partial powers >800 kHz did not improve clustering quality. This was attributed to the fact that the power of a frequency spectrum >800 kHz approached zero and thus could not provide additional discriminating power.

The frequency bounds encompass the pre-amplifier bandpass on the digital wave system and the flat frequency response of the B1025 sensors. While the frequency range includes values outside the flat frequency response of the S9225 sensors, this does not impact discriminating power. Any partitions made by the ML algorithm result from differences between waveform characteristics. The only stipulation is that recorded waveforms should be clustered independently for each sensor to capture the shift in damage mechanism (i.e., the singular set of all waveforms are not clustered together).

Another parametric study was conducted to determine d , where d was swept from $d = 2$ to 45. It was found that, when all other parameters are fixed, $d = 26$ ($\Delta k = 23$ kHz) optimized validity metrics for all specimens.

Though previous investigations have included the partial power approach as part of their representation schemes^{15,42–45}, the representation scheme described here is unique in that it uses a comparatively much finer resolution and *only* uses partial power. Typically the partial power bandwidth used is 200–600 kHz, whereas the approach herein uses a width of 23 kHz.

Spectral clustering

Once AE data are properly represented, a suitable ML algorithm for clustering can be chosen. For this task, spectral clustering was used. This is an unsupervised learning technique that has been shown empirically to outperform k -means^{37,38} and is less restrictive with respect to assumptions about input data geometries.

Spectral clustering models the input dataset as a graph with nodes (data points) connected by edges whose weight is 1 if the nodes are nearest neighbors (NNs) and 0 if they are not. The algorithm finds the optimal place to remove edges and segment the original graph into a user-specified number of subgraphs (i.e., clusters)³⁷.

Both the number of clusters and the number of NNs are considered hyperparameters (i.e., a set of user-selected parameters). One common method for estimating the number of clusters is through use of the eigengap heuristic, a measure of differences between successive eigenvalues of the graph Laplacian of the data⁴⁶. However, noisy data can reduce differences between successive eigenvalues, which is often the case for AE data. In this case, the eigengap heuristic is not sufficient for

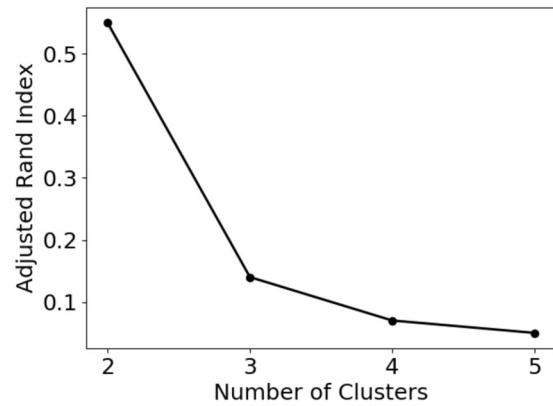


Fig. 2 Adjusted Rand Index as a function of the number of clusters. When more than two clusters are used to initialize spectral clustering, the steep drop in ARI corresponds to a decrease in precision. As such, when more than two clusters are specified, the spectral clustering algorithm is forced to find clusters, which are not correlated with damage mechanisms. This drop occurs in all experiments and is corroborated by results shown in Fig. 3.

determining the number of clusters. Instead, a parametric sweep from two to five clusters is performed and a drop in validity metrics is used to indicate the optimum number of clusters. A steep drop is observed after two (Fig. 2), corresponding with the hypothesis that matrix cracking and fiber failure events (the dominant damage mechanisms in SiC/SiC minicomposites) can be differentiated.

It is important to note there are other less dominant damage mechanisms active in minicomposites during loading (e.g., interfacial debonding and frictional sliding). It is well established from a micromechanics frame that, when matrix cracking occurs, there is simultaneous debonding and sliding in the crack wake; it is also understood that, when fibers fail, there is simultaneous fiber sliding and pullout^{47,48}.

When dominant and non-dominant mechanisms occur simultaneously, their waveforms become superimposed⁴⁹; when non-dominant mechanisms occur independently, they likely do so in quantities too small for recognition by the ML algorithm. As such, it is currently infeasible to isolate events resulting from damage to the boron nitride (BN) interphase or determine which AE features are characteristic of such damage. This is expected to be a source of error. Moreover, the inability to discriminate interfacial damage from other types of damage is reflected by the steep drop in validity metrics seen after two clusters in Fig. 2.

Similar to the determination of d and the number of clusters, a parametric study from $NN = 5$ to 20 was performed. The value of NN that optimized validity metrics slightly varied between studies (5, 5, and 7 for three experiments, respectively); however, there was a range of values for NN that produced acceptable results. To demonstrate the effectiveness of our approach, the number of NN is fixed to be 5 for all specimens resulting in suboptimal validity metrics for Experiment 3, particularly between sensors B1025-a and S9225-a (Table 3).

Validity metrics for error analysis

To evaluate the efficacy of our AE-ML framework, events at each sensor were clustered according to the steps described above, the results of which are called a *partition*. The desired outcome is that all partitions for AE data from a given specimen are the same and are independent of both the sensor model and sensor location.

To quantify partitioning success, the total matching rate is first considered, which is the percent of events assigned the same label by the clustering routine. As clusters have unbalanced sizes

(fewer fiber break events are expected than matrix crack events), it is possible to be unable to discriminate between damage mechanisms and retain high matching rates. For example, if 90% of AE events come from matrix cracks, then classifying every AE event as a matrix crack would yield a 90% matching rate, yet the ML framework would have no discriminating power.

Therefore, it is also useful to consider the permutation model of the adjusted Rand Index (ARI), which makes considerations for unbalanced cluster sizes^{50,51}. The ARI is an adjusted-for-chance version of the Rand Index (RI), a metric for comparing the similarity of a partition to the ground truth. First, the RI for two different partitions, (A, B) , is calculated as:

$$RI(A, B) = \frac{N_{11} + N_{00}}{\binom{N}{2}} \quad (6)$$

where N is the number of elements, N_{11} is the number of element pairs that are grouped into the same cluster in both partitions, and N_{00} is the number of element pairs that are grouped into different partitions in both A and B . The ARI is then calculated as³⁹:

$$ARI(A, B) = \frac{RI(A, B) - \mathbb{E}[RI(A, B)]}{\max[RI(A, B)] - \mathbb{E}[RI(A, B)]} \quad (7)$$

where $\mathbb{E}[RI(A, B)]$ is the expected value of the RI under a random model. The ARI is bound between 0 and 1, with 0 corresponding to random label assignments and 1 corresponding to perfectly matching labels.

This metric is useful for comparing a partition to the ground truth, and it is also useful for comparing two partitions that are assumed to be drawn from the same random model⁵¹. This makes it an effective tool to compare similarity between two partitions whose ground truth is not known a priori. ARI values exceeding 0.40 are correlated with high values of other classification metrics⁵².

T-distributed stochastic neighbor embedding (t-SNE)

A final, necessary step in this study was to confirm that the AE data forms identifiable clusters in the chosen feature space. To this end, t-SNE was employed. T-SNE is a manifold learning algorithm used to produce a low-dimensional visualization of high-dimensional data. Although t-SNE axes, inter-cluster separation, and cluster size have no intrinsic meaning, t-SNE has been empirically shown to be a powerful tool for the identification of natural cluster structures in high-dimensional data⁵³.

T-SNE maintains pairwise distances between the high- and low-dimensional representation of feature vectors, \mathbf{x}_i and \mathbf{y}_i , respectively⁵³. For a given feature vector, \mathbf{x}_i , t-SNE models pairwise distances in the high-dimensional representation according to a Gaussian probability distribution with standard deviation σ_i , centered at \mathbf{x}_i . Under this model, the conditional probability of finding another feature vector \mathbf{x}_j is then:

$$p_{j|i} = \frac{\exp(-|\mathbf{x}_i - \mathbf{x}_j|^2 / 2\sigma_i^2)}{\sum_{k \neq i} \exp(-|\mathbf{x}_i - \mathbf{x}_k|^2 / 2\sigma_i^2)} \quad (8)$$

and pairwise distances in the high-dimensional space are:

$$p_{ij} = \frac{p_{j|i} + p_{i|j}}{2N} \quad (9)$$

The pairwise distances in the high dimension are then translated to similar pairwise distances in the low dimension, q_{ij} , which follow a Student's t -distribution with a single degree of freedom:

$$q_{ij} = \frac{(1 + |\mathbf{y}_i - \mathbf{y}_j|^2)^{-1}}{\sum_{k \neq l} (1 + |\mathbf{y}_k - \mathbf{y}_l|^2)^{-1}} \quad (10)$$

If the low-dimensional representation has correctly maintained the same high-dimensional pairwise distances, then $p_{ij} = q_{ij}$ for all pairs i, j .

Table 1. Validity metrics for Experiment 1.

Sensor pair	Matching rate (%)	ARI
B1025-a, B1025-b	96.4	0.78
B1025-a, S9225-a	92.1	0.62
B1025-b, S9225-b ^a	81.0	0.0
S9225-a, S9225-b ^a	73.0	0.0

^aThis test had a misaligned sensor on its larger epoxy tab wherein it did not fully overlay the minicomposite and is included to show that the ARI metric can be used to detect experimental issues such as these. The ARI of zero for sensor pairs that included sensor S9225-b are the result of this improper placement (Section Results).

Table 2. Validity metrics for Experiment 2.

Sensor pair	Matching rate (%)	ARI
B1025-a, B1025-b	87.1	0.42
B1025-a, S9225-a	82.9	0.31
B1025-b, S9225-b	86.1	0.45
S9225-a, S9225-b	84.0	0.40

Table 3. Validity metrics for Experiment 3.

Sensor pair	Matching rate (%)	ARI
B1025-a, B1025-b	88.4	0.55
B1025-a, S9225-a	81.2	0.27
B1025-b, S9225-b	94.4	0.73
S9225-a, S9225-b	91.7	0.53

An important consideration for t-SNE plots is the choice of perplexity. This hyperparameter estimates a global value of σ_i , as there is no single value of σ_i to describe all data points. Perplexity measures how much of the local structure is retained in the final low-dimensional map; as perplexity increases, local structure information is exchanged for global structure information⁵³. Typical values of perplexity range from 5 to 50; a perplexity value of 15 is chosen as this best shows the cluster structure within our data.

Framework precision

Matching rates and ARIs for AE data from three SiC/SiC specimens are presented (Tables 1–3) for partitions made from: (i) sensors of the same model fixed on opposite ends of the specimen gauge and (ii) sensors of different models fixed at the same gauge location. Specimen 1 was found to have a misaligned sensor on its larger epoxy tab wherein it did not fully overlay the minicomposite and is included to show that the ARI metric can also be used to detect experimental issues such as these. The ARI of zero for sensor pairs that included sensor S9225-b are the result of this improper placement.

From the high values of matching rates and ARIs obtained across specimens, we conclude that labels are not assigned randomly between sensors based either on model or location. A corollary of this observation is that stochastic experimental effects that are known to influence frequency spectra, such as source-to-sensor distance¹⁷ or proximity to a free surface⁵⁴, do not drive label assignment when using the approach presented herein.

The clustered events at each individual sensor exhibit distinct activities in the stress domain, which are strongly characteristic of

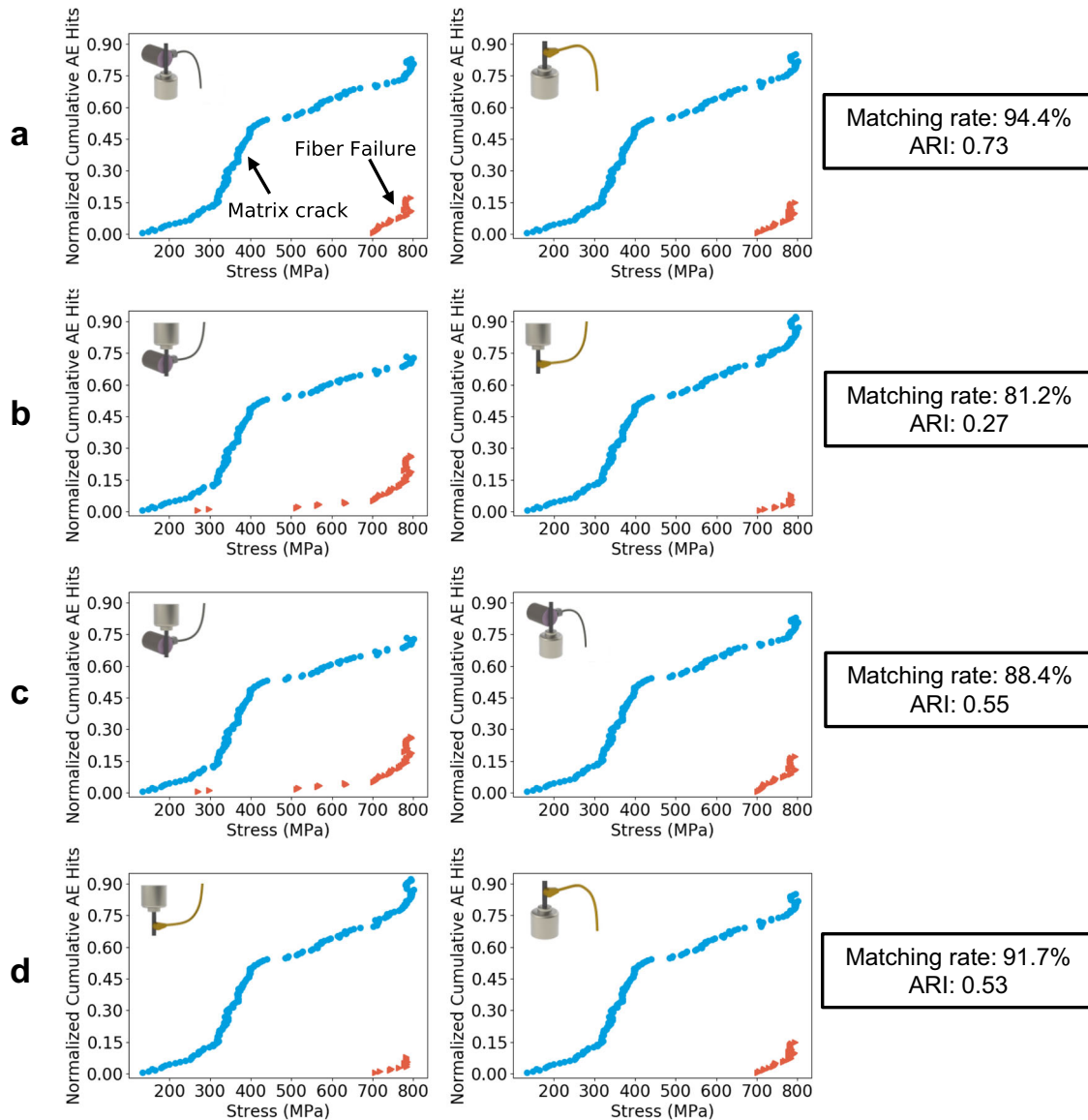


Fig. 3 Damage mechanism identification from acoustic emission (AE) signals. AE waveforms were generated by SiC/SiC ceramic matrix composites (CMCs) loaded under uniaxial tension and recorded by four sensors: **a** B1025-b and S9225-b, **b** B1025-a and S9225-a, and at different locations **c** B1025-a and B1025-b and **d** S9225-a and S9225-b. Cluster assignment of individual AE events is consistent across sensors. The resultant identification of AE events closely follows CMC damage chronology, wherein early matrix cracking is later followed by fiber breaks, even though the approach is fully domain-knowledge agnostic. The cluster that becomes active at $\approx 85\%$ of the UTS is labeled as fiber failure, consistent with experiment¹¹. Additionally, the partitions were highly precise across both the model and location of the sensors, and the partitioning was repeatable.

how damage progresses in a CMC (shown for Specimen 3 in Fig. 3). These clusters follow the established chronology of damage accumulation in CMCs, wherein cracks initiate in the SiC matrix early in the loading profile and evolve throughout the specimen lifetime. As such, the initially activated cluster is designated as matrix cracking. After the onset of major matrix cracking, fiber failure begins and fibers continue to fail until rupture⁵⁵. The secondary cluster becomes significantly active at 70–85% of the ultimate tensile strength (UTS), which agrees with experimental observations of fiber failure in SiC/SiC¹¹.

Partial power trends

To further explore the hypothesis that there are frequency trends which allow for discrimination between matrix cracking and fiber failure, it is useful to inspect the partial powers of AE signals (the input representation) as a function of stress. It was found that

select frequency bands exhibited similar behaviors, as shown in Fig. 4. Specifically, AE events occurring at stresses $>70\%$ UTS exhibited tighter distributions in partial power. A two-sample Kolmogorov–Smirnov test shows that the partial powers, for the selected frequency bands, sampled below 70% UTS come from a different distribution than the partial powers sampled above 70% UTS at a significance level of $\alpha = 0.01$ (Fig. 4).

This decrease in partial power scatter coincided with the activation of the fiber failure cluster, but it is currently unclear whether this correlation is (i) characteristic of the damage mechanism or (ii) characteristic of the differential strain-at-failure of the constituents^{56,57} (e.g., a weak fiber failing at low strain is indistinguishable from matrix cracking at the same strain). The second option is possible while still allowing for the stepped activity observed as the strain-to-failure of the fiber and matrix are disparate. The former possibility is more consequential, as it would

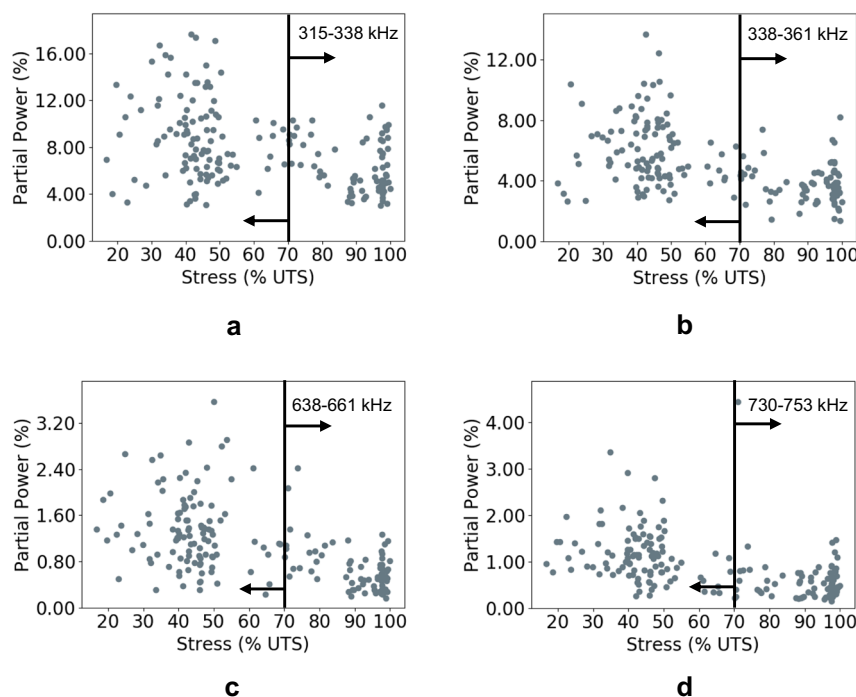


Fig. 4 Partial power as a function of stress for four selected frequency bands in Specimen 3. Events occurring below 70% UTS are sampled from a different distribution than events occurring $>70\%$ UTS at a significance level of $\alpha = 0.01$ in frequency bands (a) 315–338 kHz, (b) 338–361 kHz, (c) 638–661 kHz, and (d) 730–753 kHz. This is not predicted by the orthotropic model of wave propagation. We hypothesize that the shift in partial power is a result of a shift in active damage mechanism from matrix cracking to fiber failure; however, further experimentation and modeling is needed.

indicate that the ML algorithm is *learning* the differences between dominant damage mechanisms. Given that the homogeneous orthotropic model of wave propagation does not predict meaningful frequency trends for elastically similar material systems, additional experimentation and explicit modeling of AE in SiC/SiC minicomposite geometries is needed to understand the physical origins of the observed trends.

Feature vector cluster structure

The partitioning of each AE event into its associated cluster is the result of the intrinsic cluster-like structure and not an artifact of the chosen algorithm. This finding is evident in the visualization of the input feature vectors via *t*-SNE. In Fig. 5, the left-hand column shows the raw feature vectors plotted via *t*-SNE, and the right-hand column shows the same feature vectors subsequently colored according to the labels assigned by our ML framework. In Fig. 5a, b, the two clusters are sufficiently distinct to be visually identified in the unlabeled data. The unlabeled cluster structures in Fig. 5c, d are less readily discernible, yet still evident. This behavior is likely a result of dimension reduction rather than an intrinsic lack of cluster structure, as validity metrics show the events are given the same labels.

DISCUSSION

Damage mechanism identification in SiC/SiC CMCs from AE data is of interest for both lifetime prediction and SHM. While computational frameworks exist for damage mechanism identification, these are predominantly successful only when the composite constituents are elastically dissimilar. As such, differentiating between the dominant damage mechanisms in SiC/SiC architectures has remained a long-standing challenge.

In this work, we develop and evaluate an AE-ML framework to overcome these difficulties. A modified partial power representation scheme is adopted that allows inspection of local changes to

frequencies. This representation scheme is combined with the spectral clustering algorithm, which is well suited for partitioning AE data. This framework is then applied to waveforms collected by our unique four-sensor configuration, which allows us to draw the following conclusions:

1. Damage mechanism identification in elastically similar composite systems is possible, when salient features are chosen to represent AE waveforms and a suitable ML approach is applied.
2. Partial power is a salient feature for damage mechanism discrimination. It is not significantly perturbed by stochastic experimental factors, as is evidenced by consistent labeling between sensors, independent of both location and model. Moreover, *t*-SNE plots show that AE data intrinsically adopt cluster-like structures when represented by this scheme.
3. There are meaningful frequency trends, which are not predicted by the orthotropic model of wave propagation, which enable damage mechanism discrimination in SiC/SiC CMCs. Further investigation is needed to determine the physical basis of this.

This AE-ML framework is domain-knowledge agnostic, yet when contextualized with domain knowledge it is clear that cluster activities follow a domain-based assessment of accuracy, indicating promising robustness of this approach. Still, a full characterization of this framework's behavior is needed before it can be more broadly applied. While precision of clustering results is high in our model minicomposite system, it is unclear how precision will be affected when this framework is applied to more complex CMC architectures, where more damage mechanisms are active. Additionally, it is not known how specimen geometry, architecture, and scale influence AE features^{58,59}; this is an active area of research. Both of these questions will require testing large-scale composites in conjunction with in situ microstructural observations for error quantification and is the subject of future work.

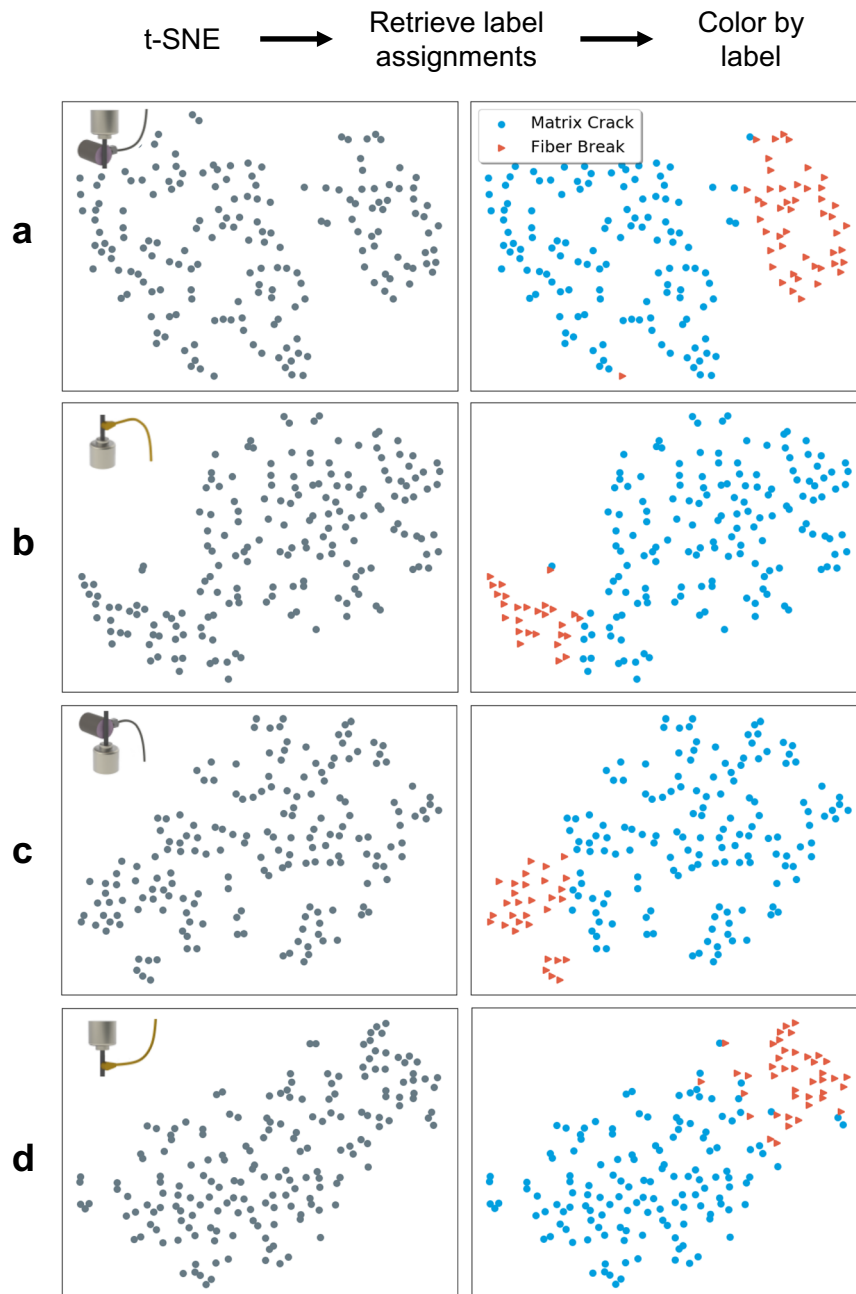


Fig. 5 Dimensionality reduction by t-SNE of Specimen 3 feature vectors. Plots for the raw feature vectors are shown for (a) B1025-a, (b) S9225-b, (c) B1025-b, and (d) S9225-a on the left and are subsequently colored on the right according to the labels given after spectral clustering as described in Fig. 3. While the t-SNE axes have no intrinsic meaning, well-formed clusters indicate that partitions were made according to similarity between feature vectors and are not an artifact of the clustering routine.

Finally, it is desirable to understand the drivers for observed trends in partial power in order to remove irrelevant frequency bands from the feature space and increase precision. Future investigations will aim to address this.

METHODS

Experimental configuration

SiC/SiC minicomposites (Rolls Royce High Temperature Composites, Cyprus, CA) were loaded under uniaxial tension using a custom load frame built in-house with a 220 N load cell and a cross-head displacement rate of 0.120 mm/min, corresponding to a nominal strain rate of $7.5 \times 10^{-5} \text{ s}^{-1}$. Specimens consisted of 500 Hi-Nicalon Type STM SiC fibers (HNS fibers), a BN interphase, and an overlayer of chemical vapor-infiltrated SiC matrix.

The microstructural and damage characteristics of these minicomposites are described in Swaminathan et al.⁷, where they are referred to as low fiber content (LFC) minicomposites. Three tension tests on LFC minicomposites are presented here to demonstrate the ML framework.

AE activity was recorded using a four-channel fracture wave detector acquisition system (Digital Wave Corporation, Centennial, CO). The threshold voltage was set to 0.1 V, the number of pre-trigger points was set to 256, and the total length of signal captured was 1024 points at a rate of 10 MHz. Two sets of AE sensor models were mounted (Fig. 1) on Duralco 132 epoxy tabs with vacuum grease (Fig. 1b). The tab thickness was 0.35 mm, as measured from the surface of the minicomposite to the sensor. Tabs were created using three-dimensional-printed molds of 10 mm in diameter.

The choice of coupling medium is significant for waveform transmission through an interface. The transmission coefficients are determined by how

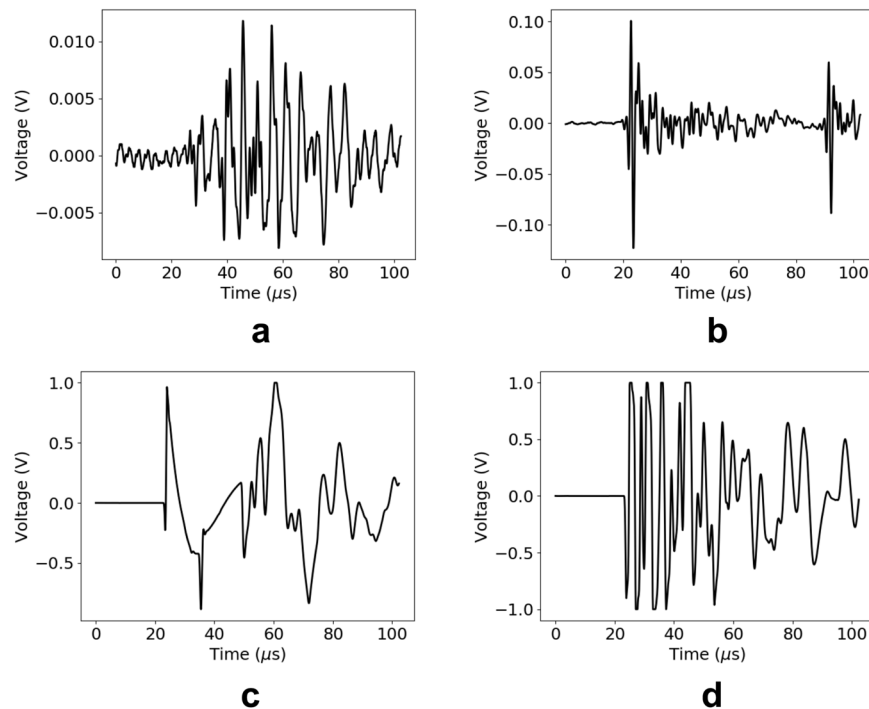


Fig. 6 Examples of removed waveforms. **a** Low signal-to-noise ratio, **b** two distinct damage events occurring in the same time window, and **c**, **d** events that saturated the AE sensor. The majority of removed waveforms were of types **a**, **c**, and **d**. Less than ten waveforms were of type **b** across the three specimens.

closely the acoustic impedance of the specimen and the couplant match. As the impedance mismatch between the two increases, the transmission coefficient decreases⁶⁰. Of the available medium choices examined, Duralco 132 exhibited the lowest impedance mismatch to SiC/SiC.

Additionally, the thickness of the epoxy tabs controls the quality of data collected. Previous work has reported that transmission coefficients decrease as a function of coupling medium thickness⁶⁰. Thinner epoxy tabs resulted in more consistent labeling, which is hypothesized to be a result of the higher transmission coefficients at all frequencies as compared to thick (>1.5 mm) epoxy tabs.

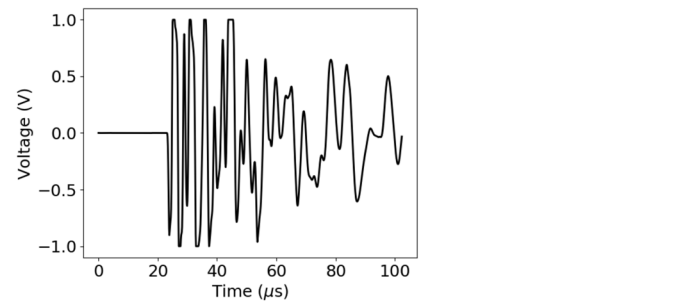
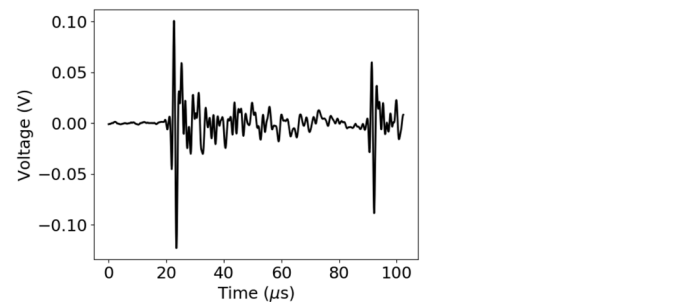
AE signals were recorded using two miniature S9225 piezoelectric AE transducers (Physical Acoustics, Princeton, NJ) with a broadband response of 300–1800 kHz and two B1025 transducers (Digital Wave Corporation, Centennial, CO) with a broadband response of 50–2000 kHz. The sensors were designated S9225-a, S9225-b, B1025-a, and B1025-b, where the I.D. letter corresponds to the position on the minicomposite (Fig. 1). The acquisition system was synchronized; when one sensor was triggered, all sensors recorded waveform data simultaneously.

Data processing

After acquisition, the raw AE data were cleaned to remove events not suitable as input to the ML framework. First, clearly identifiable noise events were removed (Fig. 6a). These events were either characterized by a single voltage spike in the time domain of the waveform or they presented with energies <0.001 V (low signal-to-noise ratio). Recorded waveforms that showed multiple damage events within the same time window (Fig. 6b) were also not considered; <1% of all recorded waveforms were of type (b). If the signal saturated the sensor (Fig. 6c, d), the event was not considered. The majority of waveforms removed were of types (a), (c), and (d). Then a location analysis was performed to remove out-of-gauge events²⁹ using:

$$\text{location} = \frac{x}{2} \left(\frac{\Delta t}{\Delta t_x} \right) \quad (11)$$

where x is the sensor separation distance, Δt is the difference in time-of-arrival, and Δt_x is the difference in time-of-arrival for an out-of-gauge event as a function of the damage parameter. This analysis ensured that only signals arising from in-gauge damage events were analyzed. During this



process, it was found that the majority of type (a) waveforms (>90%) came from out-of-gauge events. This cleaning removed approximately 35% of all recorded AE waveforms, including out-of-gauge events.

DATA AVAILABILITY

The datasets and code generated and analyzed in this study are available in the Muir-UCSB repository at https://github.com/Muir-UCSB/AE-ML_Framework.

Received: 18 May 2021; Accepted: 17 August 2021;

Published online: 13 September 2021

REFERENCES

- Zok, F. W. Ceramic-matrix composites enable revolutionary gains in turbine engine efficiency. *Am. Ceram. Soc. Bull.* **95**, 22–28 (2016).
- Lee, W. E., Gilbert, M., Murphy, S. T. & Grimes, R. W. Opportunities for advanced ceramics and composites in the nuclear sector. *J. Am. Ceram. Soc.* **96**, 2005–2030 (2013).
- Halbig, M. C., Jaskowiak, M. H., Kiser, J. D. & Zhu, D. Evaluation of ceramic matrix composite technology for aircraft turbine engine applications. In *51st AIAA Aerospace Sciences Meeting including the New Horizons Forum and Aerospace Exposition 2013* (AIAA, 2013).
- Steibel, J. Ceramic matrix composites taking flight at GE Aviation. *Am. Ceram. Soc. Bull.* **98**, 30–33 (2019).
- Morscher, G. N. & Godin, N. In *Ceramic Matrix Composites: Materials, Modeling and Technology* 1st edn (eds Bansal, N. P. & Lamon, J.) 571–590 (Wiley, 2015), 1 edn.
- Whitlow, T., Jones, E. & Przybyla, C. In-situ damage monitoring of a SiC/SiC ceramic matrix composite using acoustic emission and digital image correlation. *Composite Struct.* **158**, 245–251 (2016).
- Swaminathan, B. et al. Microscale characterization of damage accumulation in CMCs. *J. Eur. Ceram. Soc.* **41**, 3082–3093 (2021).
- Maillet, E. et al. Combining in-situ synchrotron x-ray microtomography and acoustic emission to characterize damage evolution in ceramic matrix composites. *J. Eur. Ceram. Soc.* **39**, 3546–3556 (2019).
- Sukhanov, D., Zavyalova, K. & Kadurina, A. Method for enhancement of spatial resolution of eddy current imaging. *Meas. Sci. Technol.* **30**, 065402 (2019).

10. Smith, C. & Gyekenyesi, A. Detecting cracks in ceramic matrix composites by electrical resistance. In *Proc. SPIE The International Society for Optical Engineering* 7983 79830N (SPIE, 2011).
11. Hillmas, A. M., Sevens, K. M. & Halloran, J. W. Damage evolution in SiC/SiC unidirectional composites by x-ray tomography. *J. Am. Ceram. Soc.* **103**, 3436–3447 (2020).
12. Elsley, R. K. & Graham, L. J. Pattern recognition in acoustic emission experiments. *Pattern Recognit. Acoust. Imaging* **0768**, 285 (1987).
13. de Groot, P. J., Wijnen, P. A. & Janssen, R. B. Real-time frequency determination of acoustic emission for different fracture mechanisms in carbon/epoxy composites. *Compos. Sci. Technol.* **55**, 405–412 (1995).
14. Johnson, M. & Gudmundson, P. Broad-band transient recording and characterization of acoustic emission events in composite laminates. *Compos. Sci. Technol.* **60**, 2803–2818 (2000).
15. Sause, M. G., Gribov, A., Unwin, A. R. & Horn, S. Pattern recognition approach to identify natural clusters of acoustic emission signals. *Pattern Recognit. Lett.* **33**, 17–23 (2012).
16. Muir, C. et al. Damage mechanism identification in composites via machine learning and acoustic emission. *npj Comput. Mater.* **7**, 1–15 (2021).
17. Maillat, E., Baker, C., Morscher, G. N., Pujar, V. V. & Lemanski, J. R. Feasibility and limitations of damage identification in composite materials using acoustic emission. *Compos. Part A Appl. Sci. Manuf.* **75**, 77–83 (2015).
18. Jain, A. K., Murty, M. P. & Flynn, P. J. Data clustering: a review. *ACM Comput. Surv.* **31**, 264–323 (1999).
19. Jain, A. K. Data clustering: 50 years beyond k-means. *Pattern Recognit. Lett.* **31**, 651–666 (2010).
20. Gorman, M. R. & Ziola, S. M. Plate waves produced by transverse matrix cracking. *Ultrasonics* **29**, 245–251 (1991).
21. Kostopoulos, V., Loutas, T. & Dassios, K. Fracture behavior and damage mechanisms identification of SiC/glass ceramic composites using AE monitoring. *Compos. Sci. Technol.* **67**, 1740–1746 (2007).
22. Chelliah, S. K., Parameswaran, P., Ramasamy, S., Vellayaraj, A. & Subramanian, S. Optimization of acoustic emission parameters to discriminate failure modes in glass-epoxy composite laminates using pattern recognition. *Struct. Health Monit.* **18**, 1253–1267 (2019).
23. Moevus, M. et al. Analysis of damage mechanisms and associated acoustic emission in two SiCf/Si-B-C composites exhibiting different tensile behaviours. Part II: Unsupervised acoustic emission data clustering. *Compos. Sci. Technol.* **68**, 1258–1265 (2008).
24. de Oliveira, R. & Marques, A. T. Health monitoring of FRP using acoustic emission and artificial neural networks. *Comput. Struct.* **86**, 367–373 (2008).
25. Godin, N., Huguet, S. & Gaertner, R. Integration of the Kohonen's self-organising map and k-means algorithm for the segmentation of the AE data collected during tensile tests on cross-ply composites. *NDT E Int.* **38**, 299–309 (2005).
26. Marec, A., Thomas, J. H. & El Guerjouma, R. Damage characterization of polymer-based composite materials: Multivariable analysis and wavelet transform for clustering acoustic emission data. *Mech. Syst. Signal Process.* **22**, 1441–1464 (2008).
27. Gorman, M. R. & Prosser, W. H. AE source orientation by plate wave analysis. *J. Acoust. Emiss.* **9**, 283–288 (1991).
28. Surgeon, M. & Wevers, M. Modal analysis of acoustic emission signals from CFRP laminates. *NDT E Int.* **32**, 311–322 (1999).
29. Morscher, G. N. Modal acoustic emission of damage accumulation in a woven SiC/SiC composite. *Compos. Sci. Technol.* **59**, 687–697 (1999).
30. Hamstad, M., O'Gallagher, A. & Gary, J. A wavelet transform applied to acoustic emission signals: part 2: source location. *J. Acoust. Emiss.* **20**, 62–82 (2002).
31. Sause, M. G. & Horn, S. Simulation of acoustic emission in planar carbon fiber reinforced plastic specimens. *J. Nondestructive Eval.* **29**, 123–142 (2010).
32. Aggelis, D. G., Shiotani, T., Papacharalampopoulos, A. & Polyzos, D. The influence of propagation path on elastic waves as measured by acoustic emission parameters. *Struct. Health Monit.* **11**, 359–366 (2012).
33. Muir, C. et al. Large scale statistical methods to evaluate damage accumulation in SiC/SiC composites. In *SEM Annual 2020 Virtual XIV International Congress (SEM, 2020)*.
34. Maillat, E. et al. Damage monitoring and identification in SiC/SiC minicomposites using combined acousto-ultrasonics and acoustic emission. *Compos. Part A Appl. Sci. Manuf.* **57**, 8–15 (2014).
35. Graff, K. F. *Wave Motion in Elastic Solids* (Ohio State University Press, 1975).
36. Kostopoulos, V., Loutas, T. H., Kontsos, A., Sotiriadis, G. & Pappas, Y. Z. On the identification of the failure mechanisms in oxide/oxide composites using acoustic emission. *NDT E Int.* **36**, 571–580 (2003).
37. Ng, A., Jordan, M. & Weiss, Y. On spectral clustering: analysis and an algorithm. In *Proc. 14th International Conference on Neural Information Processing Systems: Natural and Synthetic*. 849–846 (NIPS, 2001).
38. Rodriguez, M. Z. et al. Clustering algorithms: a comparative approach. *PLoS ONE* **14**, 1–34 (2019).
39. Pedregosa, F. et al. Scikit-learn: machine learning in python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
40. Muir, C. Spectral clustering for AE. https://github.com/Muir-UCSB/AE-ML_Framework (2021).
41. Prosser, W. Advanced AE techniques in composite materials research. *J. Acoust. Emiss.* **14**, S1–S11 (1996).
42. Sause, M. G. & Horn, S. Quantification of the uncertainty of pattern recognition approaches applied to acoustic emission signals. *J. Nondestructive Eval.* **32**, 242–255 (2013).
43. Kempf, M., Skrabala, O. & Altstädt, V. Reprint of: Acoustic emission analysis for characterisation of damage mechanisms in fibre reinforced thermosetting polyurethane and epoxy. *Compos. Part B Eng.* **65**, 117–123 (2014).
44. Wisner, B. et al. Acoustic emission signal processing framework to identify fracture in aluminum alloys. *Eng. Fract. Mech.* **210**, 367–380 (2019).
45. Guel, N. et al. Data merging of AE sensors with different frequency resolution for the detection and identification of damage in oxide-based ceramic matrix composites. *Materials* **13**, 1–22 (2020).
46. Afzalan, M. & Jazizadeh, F. An automated spectral clustering for multi-scale data. *Neurocomputing* **347**, 94–108 (2019).
47. Marshall, D. B. & Oliver, W. C. Measurement of interfacial mechanical properties in fiber-reinforced ceramic composites. *J. Am. Ceram. Soc.* **70**, 542–548 (1987).
48. Callaway, E. B., Christodoulou, P. G. & Zok, F. W. Deformation, rupture and sliding of fiber coatings in ceramic composites. *J. Mech. Phys. Solids* **132**, 103673 (2019).
49. Swaminathan, B. et al. Interpreting acoustic energy emission in SiC/SiC minicomposites through modeling of fracture surface areas. *J. Eur. Ceram. Soc.* **41**, 6883–6893 (2021).
50. Hubert, L. & Arabie, P. Comparing partitions. *J. Classification* **2**, 193–218 (1985).
51. Gates, A. J. & Ahn, Y. Y. The impact of random models on clustering similarity. *J. Mach. Learn. Res.* **18**, 1–28 (2017).
52. Santos, J. M. & Embrechts, M. In *Artificial Neural Networks – ICANN 2009* (eds Alippi, C., Polycarpou, M., Panayiotou, C. & Ellinas, G.) 175–184 (Springer Berlin Heidelberg, 2009).
53. van der Maaten, L. & Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
54. Hamstad, M. A. Acoustic emission signals generated by monopole (pencil-lead break) versus dipole sources: finite element modeling and experiments. *J. Acoust. Emiss.* **25**, 92–106 (2007).
55. Curtin, W. A., Ahn, B. K. & Takeda, N. Modeling brittle and tough stress-strain behavior in unidirectional ceramic matrix composites. *Acta Mater.* **46**, 3409–3420 (1998).
56. Almansour, A., Maillat, E., Ramasamy, S. & Morscher, G. N. Effect of fiber content on single tow SiC minicomposite mechanical and damage properties using acoustic emission. *J. Eur. Ceram. Soc.* **35**, 3389–3399 (2015).
57. Callaway, E. B. & Zok, F. W. Strengths of ceramic fiber bundles: theory and practice. *J. Am. Ceram. Soc.* **100**, 5306–5317 (2017).
58. Sause, M. G. Acoustic emission source identification in large scale fibre reinforced composites. In *Czech Society for Nondestructive Testing 32nd European Conference on Acoustic Emission Testing*. 125–136 (Czech Society for Nondestructive Testing, 2016).
59. Ospitia, N., Aggelis, D. G. & Tsangouri, E. Dimension effects on the acoustic behavior of TRC plates. *Materials* **13**, 955 (2020).
60. Ono, K. Through-transmission characteristics of AE sensor couplants. *J. Acoust. Emiss.* **34**, 1 (2017).

ACKNOWLEDGEMENTS

C.M. and B.S. gratefully acknowledge financial support from the NASA Space Technology Research Grant Program (Grants: 80NSSC19K1164 and 80NSSC17K0084). S.D. and T.M.P. gratefully acknowledge financial support from the National Science Foundation (Award: 1934641) as part of the HDR IDEAS² Institute. The authors thank Abed Musaffar for creating the CAD schematic in Fig. 1a and thank Dr. Neal Brodnik for a detailed introduction to t-SNE.

AUTHOR CONTRIBUTIONS

C.M., B.S. and K.F. performed experiments and data collection. C.M. and B.S. were responsible for writing the manuscript. All authors discussed scientific content and were responsible for reviewing, editing, and ensuring the correctness and completeness of the statements made herein.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Correspondence and requests for materials should be addressed to S. Daly.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing,

adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021