Cluster-based acoustic emission signal processing and loading rate effects study of nanoindentation on thin film stack structures

Chen Liu a,*, Oliver Nagler b, Florian Tremmel b, Marianne Unterreitmeier b, Jessica J. Frick c, Radhika P. Patil a, X. Wendy Gu a, Debbie G. Senesky c

a Department of Mechanical Engineering, Stanford University, CA 94305, USA
b Infineon Technologies AG, Am Campeon 1 - 15, 85579 Neubiberg, Germany
c Department of Aeronautics and Astronautics, Stanford University, CA 94305, USA

ARTICLE INFO

Keywords:
- Loading rate effects
- Acoustic emission (AE)
- Nanoindentation
- Data clustering
- K-means
- Autoencoder

ABSTRACT

This paper presents a high-resolution, in-situ material testing system that integrates acoustic emission (AE) testing with a nanoindentation system for crack generation and detection in thin film stack structures. This is used to find the critical contact load during wafer probing of crack-sensitive backend-of-line (BEOL) structures in semiconductor integrated circuits. Scanning electron microscopy (SEM) and load–displacement curve analysis were used to confirm the formation and propagation of cracks in the multilayer structures. In order to improve the manual classification performance and understand the physical meaning of AE signals, this paper introduces a machine learning based signal processing approach based on a k-means clustering algorithm applied on collected AE signals. To obtain the optimal number of k-means clusters, Davies–Bouldin, Dunn, and Silhouette indices were calculated, and the individual ratings were cumulated based on a voting scheme. Multiple feature extraction methods, including raw time-domain AE signals, conventional AE extracted parameters, short-term signal energy, and representation features learned by the autoencoder, were used and evaluated by manually labeled clusters and binary confusion matrices. A supervised learning technique, the k-nearest neighbors algorithm, was also utilized on different AE signal datasets using different loading rates to further investigate the damage processes during nanoindentation and the physical meaning of different AE signals. The influences of loading rates on AE signals have been investigated, and loading rate effects on the critical load were observed – higher loading rates led to higher critical loads. This integrated test system and signal processing approach provides a high-resolution mechanical testing platform for studying and enabling automatic crack detection in wafer probing.

1. Introduction

In large-scale production of electronic devices, quality control is performed through electrical testing of individual devices fabricated on semiconductor wafers before further assembly and packaging [1–5], which requires physical contact between the device...
and a conductive elastic probe. During wafer testing, contact between the probe and the sample is analogous to the nanoindentation process if we assume the exclusively vertical movement of the probe tip. Nanoindentation is typically used to measure the hardness and elastic modulus of thin films and small specimens, which can be used to understand yield strength, fracture toughness, and other mechanical properties [6]. During the nanoindentation process, a rigid tip, frequently made of diamond, is pressed onto a testing sample, and both the load and displacement are monitored at the nanoscale. In cases of high contact force, there exists an increased risk of cracking the device’s brittle insulating layers [3–5], which, in turn, can lead to failure of the device due to short circuits or leakage and reduced yield. Therefore, in order to monitor the reliability of the semiconductor device testing procedure, suitable crack detection methods are essential.

Numerous researchers have studied the failure of the insulating layer in thin film stack structures during wafer probing and wire bonding [7–10]. Optical inspection and scanning electron microscopy (SEM) images of the focused ion beam (FIB) milled cross-sections have been used for crack detection in insulating layers. However, optical imaging is not in real-time, and FIB milling is destructive, time-consuming, and cost-intensive [11,12]. An attractive alternative to these techniques is acoustic emission (AE) testing

Fig. 1. (a) Schematic drawing and (b) photo of the experimental setup showing the integration of an AE sensor and the nanoindentation system.
AE testing is widely used in non-destructive material testing. Many studies have used AE testing to understand fracture behavior in different materials (e.g., composites, concrete, etc.) and structures (e.g., pipes, vessels, etc.) [15]. In addition, AE testing has been used to study the fracture behavior of small-scale samples at the nanoscale. To do this, AE sensors have been integrated with a nanoindenter to correlate high-resolution load–displacement curves with failure events [16–21]. These include one-off studies on a range of materials and structures, including bulk Fe$_3$Si, and zeolites [16]; Fe$_7$Pd$_3$ alloys [17]; and various brittle materials like soda-lime glass, SiC, and ceramic [18]. Although integrating AE sensors with a nanoindentation system can obtain additional information, the reproducibility, and thus reliability of this integrated system, is inhibited by several factors that affect a sample’s AE signal, including sample size and sensor-source distance.

In conventional, large-scale indentation testing with AE sensors, the samples are glued on top of a sensor; however, the reproducibility of AE signals on multiple sample tests is limited due to the sample size effect. For example, if the sample size is relatively small compared to the sensor size, the variation of sample position on the sensor and the glue thickness will generate irreproducible signals from different sample tests. Furthermore, it is unrealistic to glue each sample to the AE sensor for large-scale production. To alleviate the impact of sample size and sensor-source distance on AE signals, our team designed a sensor-indenter system to detect mechanical cracks in real-time, supported by analytical modeling and computer-aided simulations [22,23]. The sensor-indenter system consists of a piezoelectric sensor and an indenter, glued on top of the sensor, that includes a flat diamond tip with a rigid, cylindrical probe. Our current work presents an advancement of this setup through the use of a nanoindentation system to enable a higher resolution indentation damage testing system. This integrated nanoindentation AE sensing setup provides an in-situ, high-resolution damage testing platform for small-size specimens with high reproducibility and reliability. It is helpful in studying the critical loads and fracture behavior of thin-film stack structures and in developing a crack monitoring system in wafer probing.

To distinguish AE signal types and understand the physical meaning of acoustic events, a machine learning based signal processing approach is proposed. AE testing is a powerful tool for detecting failure events, but the interpretation of AE signals and determination of signal origins are often complicated without additional analysis of recorded signals. Researchers have developed AE signal post-processing techniques to understand failure mechanisms better [24–32]. One of the significant challenges of AE analysis is correlating the AE signal features with damage modes and differentiating the AE signals according to their nature of damage events. In this paper, the signal processing approach is used to cluster the AE signals, and the corresponding damage modes are identified in SEM images of FIB milled cross-sections. Furthermore, different loading rates have been applied to study the influences on critical loads and assist in understanding the damage processes during nanoindentation. The ability to investigate different loading rates is a new capability of our setup, compared to a previous version of this setup which used one fixed loading rate [22,23]. The loading rate is important because wafer probing happens rapidly and dynamically. The fracture of silicon dioxide is rate-dependent because silicon dioxide experiences a slow crack growth process [33–37]. Thus, the strength of silicon dioxide increases with increasing loading rate. Additional aspects of rate-dependent failures, such as damage processes during nanoindentation, the existence of crack surface friction, and crack propagation, are unknown but can be understood using AE testing. To the best of the authors’ knowledge, nanoindentation of small specimens at different loading rates has never been combined with AE testing, which is the main contribution of this study.

2. Experimental setup

The experimental setup used in our study of thin film structures consists of a commercial nanoindentation system and a customized sensor-indenter system. A photo and a schematic drawing of the experimental setup are shown in Fig. 1(a) and (b). The indenter and sensor configuration in this work is based on the same configuration described by Unterreitmeier et al. [22,23]. Fig. 2(a) shows the thin film stack test structure, and Fig. 2(b) shows the SEM images of the test structure. The wafer was fabricated by Infineon on a silicon substrate. In order, the following layers were deposited on a silicon wafer (100): silicon dioxide (1780 nm), Cu (700 nm), silicon dioxide (380 nm), SiN (50 nm), Cu (700 nm), SiO$_2$ (1780 nm), and Si-Wafer (1 0 0).

![Fig. 2](image-url) (a) Schematic drawing of the thin film test structure with the indenter tip and (b) an SEM image of thin film test structure. Scale bar, 1 µm.
nitride (50 nm), and silicon dioxide (380 nm). The indenter includes a flat, 10 μm diamond tip supported on a rigid, cylindrical, steel metal probe that was glued on top of a piezoelectric sensor (Vallen Systeme VS900-M) with high sensitivity and a broad frequency range (100 kHz – 900 kHz) [38]. The combined sensor-indenter system was able to both generate cracks with controlled force and transmit the acoustic signals to a piezoelectric sensor. The AE signals were preamplified (Vallen Systeme AEP3N) to obtain a high signal-to-noise ratio with the preamplifier gain of 49 dB. A decoupling box (Vallen Systeme DCPL2) and a power supply were used to add a DC voltage to the preamplifier. Signals were continuously collected by a digital storage oscilloscope (PicosSope4000, Pico Technology) after the amplification. A voltage trigger was set beforehand above the noise level so that only signals above a certain voltage threshold (17.5 μV) were stored via a LabVIEW program. Each AE signal measured 2000 sampling points with a sampling rate of 1000 kHz. The testing sample was glued on the customized sample holder by graphite paste, and the sample holder was screwed into the electromagnetic actuator of the nanoindentation system. Nanoindentation tests were performed using a commercial instrument (iMicro, Nanomechanics, Inc.) with a force resolution of 6 nN and displacement resolution of 0.04 nm at room temperature. The iMicro software automatically detects the surface of the sample, and both load and displacement data are collected. A LabVIEW program was made to accomplish the time synchronizing of load/displacement data from iMicro software and AE signals from the oscilloscope by sending contact force values from iMicro software to LabVIEW.

All tests were performed under constant loading rate control. Each contact cycle started with a linear increasing loading process until the maximum contact force was reached and ended with the unloading process until zero force was reached. The description of the test parameters is listed in Table 1. After the nanoindentation process, SEM images of FIB prepared cross-sections were analyzed to confirm the formation of cracks.

### 3. Data clustering methodology

Signal processing methods were used to analyze AE signals corresponding to crack formation to determine the critical force and differentiate AE signals according to the nature of damage events (e.g., initial crack formation, plastic deformation). In this paper, the term ‘critical load’ is defined as the force at which the initial crack is formed.

To properly perform AE signal analysis, one must consider four necessary steps: (1) AE signal detection, (2) AE signal feature extraction techniques, (3) clustering algorithms, and (4) cluster evaluation methods. The AE signal detection technique used in this paper is to compare the AE signal against a voltage threshold (17.5 μV), which is set just above the noise. Conventional AE signal features include AE hits, rise time, duration time, amplitude, average frequency, rise angle value, and other time or frequency domain features. The literature provides numerous studies correlating the physical phenomenon with AE signal features on various structures [24–32]. For example, in 1998 Bahr et al. [24] discovered that signal rise time and frequency are related to sample geometry and that the signal’s energy scales with the elastic energy released during the damage event. More recently, Dai et al. [27] used the ratio between the rise time and peak amplitude, as well as signal counts over time, to describe the AE signals and classify the different fracture modes. For clustering algorithms, machine learning tools (both supervised [26] and unsupervised [29–32] clustering algorithms) have been widely used to correlate AE events with fracture and damage. Supervised learning, which includes support vector machines (SVM), neural networks (NN), etc., are used to infer a function from labeled training data. Conversely, unsupervised learning techniques, including Gaussian mixture model (GMM), k-means clustering, etc., do not need labels on every signal. Consequently, unsupervised learning techniques have been commonly used to distinguish AE signals. For cluster evaluation methods, Silhouette coefficient [39] and Davies-Bouldin [40,41] validation methods are most frequently used to investigate unsupervised clustering validity. Multiple clustering validity indices have been used together to evaluate the clustering results [42,30], Sause et al. [30] utilized the Davies-Bouldin index, Tou indices, Rousseeuw’s silhouette validation method, and Hubert’s Gamma statistics. These ratings are cumulated based on a voting scheme to evaluate the number of clusters for the best performance. In this study, we utilize a k-means clustering algorithm and compare the conventional AE signal feature extraction methods with the short-term signal energy method and autoencoder method. We would like to note that this study does not consider the temporal evolution of AE clusters; however, for more details on this consideration, we kindly direct the readers to these referenced works [31,43–45].

Fig. 3 shows the flowchart diagram of the signal processing method used in this paper. An unsupervised clustering method, k-means, was used to cluster the AE events after applying feature extraction methods on signals. The number of clusters was chosen by internal clustering evaluation methods, which use the internal information of the clustering process. Three internal clustering evaluation indices were calculated, and a simple voting strategy was implemented to pick the optimal number of clusters. The performance of feature extraction methods was evaluated by an external evaluation scheme, which is based on previous knowledge about data, e.g., class labels. High energy AE signals which occur around the critical load could be labeled as one cluster which represents the initial crack formation. All AE signals were manually classified into two clusters: critical load cluster and others, and the criterion is given in Section 4.4.1. After manually signal classification, F-1 scores were calculated based on the binary confusion matrix to obtain the optimal feature extraction method. The resulting clusters were then compared to SEM images to determine if the clusters are due to noise, plastic deformation, initial crack formation, or other specific damage events. A supervised learning method, k-nearest neighbors

<table>
<thead>
<tr>
<th>Maximum Contact Force (mN)</th>
<th>Loading rate (mN/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>50, 100, 150, 200, 250, 300, 350, 400</td>
<td>5</td>
</tr>
<tr>
<td>400</td>
<td>5, 10, 50, 100, 500, 1000</td>
</tr>
</tbody>
</table>
algorithm (k-NN), was used to classify AE signals under different loading rates to not only verify the reproducibility of the testing system and signal processing approach but also further study the fracture behavior of the specimen. The feature extraction method and the training data used in supervised learning were obtained from previous unsupervised learning.

3.1. Feature extraction

Features of AE signals can be defined as parameters that are able to represent the underlying structure of signals. In this paper, AE wave transmission paths from the source to the acoustic sensor are different from the conventional AE method, where the test specimen is directly mounted to the sensor. Moreover, the eigenmodes of the frequency response of the sensor-indenter system were optimized by changing the geometry of the indenter to have the maximum oscillating amplitude to get the highest sensitivity for the acoustic signal detection [22,23]. Therefore, in addition to conventionally extracted AE signal features, three supplementary features were used in the clustering process: normalized time-domain raw data, short-term signal energy, and representation features learned by the autoencoder. The first 500 data points of each signal were used in signal processing to reduce the input dimensions since the rest of the data points only contained relatively low-amplitude data due to the long decay time.

For the conventional feature extraction method, seven parameters — amplitude, the duration time (the time period between first and last threshold crossing), the rise time (the time period between first threshold crossing and the peak of signal), the number of counts (cycles from signal start to end), the RA value (the ratio between rise time and amplitude), the peak frequency, and peak frequency amplitude — were extracted from the AE signals.

The first supplementary feature used in the clustering process, normalized time-domain raw data, is the most straightforward. These features were normalized AE signals with the dimension of 500 points and were directly set as the input into the clustering algorithm. Frequency domain analysis methods are not considered in this study since these methods give similar results as time-domain raw data, and the frequency domain is not expected to carry extra information because the sensor-indenter system used in this paper is designed to have optimized resonance frequency for better signal detection [22,23].

To reduce the input dimensions but maintain the time-domain amplitude information, the second supplementary feature, short-term signal energy, was used. Short-term energy is used in speech processing [46,47], and can be generally defined as:

\[ E_s = \sum_{m=n-N+1}^n |x(m)w(n-m)|^2, \]

\[ w(x) = \begin{cases} \cos^2\left(\frac{\pi x}{L}\right), & |x| \leq L/2 \\ 0, & |x| \geq L/2 \end{cases} \]

where \( x(m) \) is an AE signal, \( w(n-m) \) represents the windowing function of finite duration, \( n \) is the sample that the window is centered on, and \( N \) is the window length. The windowing function shown in Eq. (2), \( w(x) \), is the Hann function. The window length and the number of overlapped data points are two important parameters in applying the short-term signal energy method.

Finally, the third supplementary feature, representation features learned by the autoencoder, was used to reduce the input dimensions while keeping as much information as possible. Representation learning has been actively explored in AE signals processing,
both for supervised [43,48] and unsupervised [49] learning. An autoencoder is a type of artificial neural network that can learn features in an unsupervised manner automatically [50]. The autoencoder tries to learn a function \( h_w, b(x) \approx x \), where \( w \) and \( b \) are the parameters of the neural network, and \( x \) is the input vector. In other words, an autoencoder copies its inputs to its outputs, and the objective is to minimize the reconstruction error between the input and the output calculated by the network. By placing constraints on the network, such as by setting up the number of neurons in the hidden layers, salient representation features can be obtained for dimensionality reduction purposes. Fig. 4 shows the autoencoder network architecture used in this study, which is also a two-layer neural network (one hidden layer of four neurons). The encoder learns a function that maps the input data into a hidden low-dimension representation feature, and the decoder reconstructs the input by learning a function with representation features as input. To learn an autoencoder, a loss function needs to be minimized. The loss function used is a mean squared error function. The scaled conjugate gradient descent algorithm is utilized to minimize the loss function. To prevent overfitting, weight decay regularization is implemented. The training is stopped after 2000 iterations to make sure the loss function minimum is achieved. In this paper, the activation function is a sigmoid function: \( f(z) = \frac{1}{1 + e^{-z}} \). The performance of autoencoder feature extraction with different numbers of representation features was studied in this paper.

3.2. Unsupervised clustering and evaluation method

K-means clustering method, which is the simplest and most widely employed unsupervised clustering algorithm, was used in this paper. Subsequently, groups of 3, 4, ..., 8 clusters were calculated and evaluated by cluster evaluation in the following section. Euclidean distances from data points to centroids were calculated. Due to the clustering error resulting from the random initialization of the centroids, which represents the center of corresponding clusters, several sets of different initial centroids were applied to avoid finding the local optimum. Jain et al. [51] provide a detailed description of the k-means clustering algorithm.

Evaluation of the clustering algorithm is needed to determine the reasonable number of AE signal clusters and to evaluate the performance of different feature extraction methods. Both internal and external evaluation schemes are used for this procedure. To optimize the number of clusters, internal evaluation schemes were used to assess the quality of the clustering separation and homogeneity. Silhouette criterion [39], Dunn index [40,52], and Davies-Bouldin [40,41] validation methods were adopted in this paper for k-means performance evaluation. The Silhouette criterion value was used to evaluate the optimal number of data clusters by interpreting and validating the consistency within clusters of data. It compares the similarity of a signal to those within the same cluster with its dissimilarity to those in other clusters. A higher silhouette value describes a more appropriate clustering configuration. Dunn index evaluates the compactness of sets of clusters. It is defined as the ratio of the minimal distance between two signals of different clusters to the maximal distance between signals of the same cluster. A higher Dunn index indicates a better clustering result. Davies-Bouldin index is a function of the ratio of the sum of the distance between two signals within the same cluster to the separation between different clusters. A lower value indicates a better separation between clusters and better cohesion within clusters. After calculating each validity index, a simple voting strategy was implemented to increase the reliability of the evaluation method [30,53]. In principle, the number of clusters with better performance will have higher points. Then, the sum of the points of the three criterion indices was calculated and used to find the optimal number of clusters.

The external evaluation scheme, which requires the information not used for clustering, such as class labels, is used to evaluate the performance of different feature extraction methods. During the external evaluation process, all AE signals corresponding to 150 indents performed at a loading rate of 5 mN/s and a maximum load of 400 mN were manually classified into two clusters: critical load.
cluster and others. The detailed labeling criterion is given in Section 4.4.1. Then, the confusion matrix for a binary classifier, shown in Table 2, was used to evaluate the performance of different feature extraction methods. Two parameters extracted from the confusion matrix were used: recall and precision. Recall refers to the ratio of the number of correctly identified positive cases to all actual positive cases, while precision refers to the ratio of the number of correctly identified positive cases to all predicted positive cases. The F1 score was used to combine precision and recall value, which defines the harmonic mean of precision and recall, and gives a better evaluation of the model if the class distribution is imbalanced. Recall, precision, and the F1 score are defined as:

\[
\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \quad (3)
\]

\[
\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \quad (4)
\]

\[
F1 \text{ score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)
\]

### 3.3. Supervised learning

Using an unsupervised learning technique to identify a signal class requires all data to be available from the beginning until the end of the procedure. As additional data is obtained, the dataset will become larger and will strongly affect the speed of the unsupervised learning process. Therefore, a supervised learning technique is needed to build a framework for further AE signal analysis. K-NN is a non-parametric method used for supervised learning. It is a simple algorithm that stores all available cases and classifies new data based on the Euclidean distance [54]. In this paper, K-NN was used to verify the feasibility of applying supervised learning on different AE signal datasets and to further understand the fracture behavior by classifying signals from different loading rates. The extracted features and corresponding classes from the previous unsupervised learning method were used as the training data. The test sets were obtained by applying the optimal feature extraction method from unsupervised learning on AE signals at different loading rates. To assure that the loading condition, i.e., the loading rates, solely influences the specimens’ critical conditions and fracture behavior, rather than the AE signals, the loading rate effects on the critical loads of samples and AE signals patterns are discussed in Section 4.3.

### 4. Results

#### 4.1. Acoustic emission signals

The energy of collected discrete signals, \(x[n]\), is defined as

\[
E = \sum_{n=0}^{\infty} (x[n])^2
\]

Fig. 5(a) shows a representative contact cycle and its corresponding acoustic signal above the threshold at a loading rate of 5mN/s and a maximum force of 400 mN. The acoustic signal energy was calculated using Eq. (6) and plotted with respect to the event occurrence time. Two signals with high energy occurred at 177 mN and 295 mN. The time domain and frequency domain of the signals are shown in Fig. 5(b) and (c) at 177 mN and 295 mN, respectively. Peaks in frequency domain marked with ★ correspond to the resonance frequency of the sensor-indenter system which is used to obtain the highest sensitivity for the acoustic signal detection [22,23].

A representative plot of load vs. displacement (contact force vs. depth of penetration) is given in Fig. 6(a). These experiments were performed under load control with a constant loading rate. As the load increases above a certain threshold, a subtle ‘pop-in’ event occurs [20]. This event is indicative of a sudden drop in stress, which usually correlates with fracture formation, crack initiation and growth, or onset of plastic deformation [55]. To detect the ‘pop-in’ event, the derivative of displacement was calculated. Fig. 6(b) and (c) contain zoomed-in plots of the load–displacement curve coupled with this calculation. By calculating the derivative of the depth, the observable ‘pop-in’ events are found at 177 mN and 295 mN, which are consistent with acoustic signals. This observation supports our conclusion that AE signals generated at 177 mN and 295 mN result from the formation or growth of cracks. Jungk et al. also suggested that ‘pop-in’ events during the loading process, which result from film crack growth, are associated with AE signals [19]. No additional ‘pop-in’ events were observed along the rest of the load–displacement curve, while additional AE signals were collected. This is due to displacement changes caused by other crack related events, e.g., crack surface friction, small crack formation, and crack propagation, which are too small to detect by plotting the derivative of depth.

Fig. 7 shows the scatterplot of burst signal energy as a function of contact force for all AE-hits above the threshold derived from 150 repeat hits in a 10-by-15 array on one sample. The repeat indents were positioned 100 μm apart to avoid interference. A cluster of high energy AE events is observed around 180 mN—this cluster is most likely associated with the initial crack formation. The signal energies.
before the first cluster peak are relatively small, which can most likely be attributed to the plastic deformation of the copper layer underneath the upper SiO$_2$ layer. Unterreitmeier et al. concluded that these small amplitudes AE signals are caused by the copper layer plastic deformation by comparing the AE tests carried out on the structures without a copper layer underneath the SiO$_2$ layer [23]. After the first cluster of high energy AE signals, subsequent lower energy AE events occur. These are most likely due to the crack growth, small crack formation, crack surface friction, or plastic deformation of the copper layer. The physical meaning of these lower energy AE events will be discussed in Section 4.4.2. At around 280 mN, there is another cluster of AE signals with relatively high energy. Interestingly, the signals' time domain pattern at 280 mN, shown in Fig. 5(b), is different from the time-domain pattern at ~180 mN, seen in Fig. 5(c). The difference in time domain pattern indicates that the source of this second, high energy signal cluster is different from the first high energy signal cluster. Further discussion on the origin of this second, high energy AE signal cluster will occur in Section 4.2.

4.2. SEM crack inspection

To verify the reliability of the acoustic method on detecting crack formation, SEM images of FIB-prepared cross-sections were prepared after nanoindentation at different maximum contact forces with a loading rate of 5 mN/s. During FIB preparation, electron beam assisted Pt deposition, followed by ion beam assisted Pt deposition, was performed to create a protective layer for subsequent cross-section milling. Fig. 8 shows a top view SEM image of an indent after a 400 mN force indentation. The FIB milling area is labeled as the black rectangle. Fig. 9 shows SEM images of wafer cross-sections under increasing maximum loads from 150 to 400 mN. The
bottom layer crack in Fig. 9 corresponds to the inner circle in Fig. 8, labeled by the blue arrow. The top layer crack corresponds to the outer circle, labeled by the red arrow.

The SEM images in Fig. 9 confirm the formation and propagation of cracks. Cracks were perceived in the top-most SiO$_2$ layer and the SiN layer, while no cracks were observed in the Cu layer, the second SiO$_2$ layer, or the SiO$_2$/Si interface. The vertical lines in the Cu layer represent the curtaining effect – an artifact of milling using the FIB technique. At a force level of 150 mN, no obvious AE signals were detected, and no cracks were observed in SEM cross-section images. Cracks in the nitride layer and the bottom and top of oxide layers could be seen after 200 mN, as shown in Fig. 9(b). The appearance of cracks at 200 mN force supports our claim that AE signals can detect crack formation occurring around 180 mN. From 200 to 250 mN, the top layer crack lengthens, which could be responsible for some of the AE signals at this load. Increasing to 300 mN, the crack opening on the top oxide layer widens, likely correlating with

![Fig. 6.](image1.png)

(a) A representative plot of load–displacement (contact force vs. depth). (b), (c) Zoomed-in load–displacement plots with the derivative of depth at 295 mN and 177 mN, respectively.

![Fig. 7.](image2.png)

Scatterplot of AE signal energy with contact force of 150 indents.
the second-high energy AE signal cluster shown in Fig. 7. The cracks on both the top and bottom of the oxide layer continue to grow from 300 mN to 350 mN. More significant crack growth occurs on the top layer crack, as seen in Fig. 9(d). Between 300 and 400 mN, additional, small cracks appear on the bottom oxide layer, seen in Fig. 9(d) and (e), labeled with green *. We attribute the AE signal events occurring between 300 and 400 mN in Fig. 7 to the formation of these cracks. Finally, under 400 mN loading, the opening of the bottom layer crack becomes larger, and the top layer crack penetrates through the entire silicon dioxide layer—corresponding to the last AE signal maxima around 400 mN in Fig. 7.

One last detail to note from our SEM data is that the top layer crack grows faster than the bottom layer crack. At 200 mN contact force, the bottom layer crack is larger than the top layer crack. But, as the contact force increases, the top layer crack becomes larger and wider than the bottom layer crack. One explanation could be that the top layer crack is under higher stress than the bottom layer cracks after crack formation. Another possible explanation is that silicon dioxide is highly susceptible to slow crack growth, which is generally associated with the chemical influence of water [33–35]. Therefore, it is possible that during the indentation process, water moisture from the air reached the top layer crack interface, accelerating the crack propagation.

### 4.3. Loading rate effects

Various loading rates have been applied to study the influence of loading rate on the critical load of SiO\textsubscript{2}/Cu/SiO\textsubscript{2}/Si-wafer samples and AE signals. Fig. 10 shows a scatterplot of burst signal energy as a function of contact force at different loading rates, each based on 75 repeat tests. The total number of detected signals is different for each loading rate. The average number of AE signals in each indent cycle are 12.9, 4.9, and 3.22 for the loading rate of 5 mN/s, 100 mN/s, and 1000 mN/s, respectively. The number of signals collected decreases as the loading rate increases. This could be caused by the limitation of the speed of AE signal storage at higher loading rates, which is constrained by the performance of the desktop and LabVIEW. The minimum time interval between two consecutive AE signals is 0.06 s when obtained by continuous AE signal collection. This means that at our fastest loading rate, 1000 mN/s, the resolution of signals is only 60 mN (i.e., 1000 mN/s * 0.06 s = 60 mN). To normalize the data, signals occurring within 60 mN of each other (starting at 0 mN) were eliminated from data collected at loading rates of 5 and 100 mN/s. Table 3 lists the normalized average number of AE signals: 6.36, 3.33, and 3.22 for loading rates of 5, 100, and 1000 mN/s, respectively. From this normalization, we can conclude that the initial decrease in the number of signals with an increase in loading rate from 100 mN/s to 1000 mN/s is most likely because of the limitation in data storage speed. However, the persistent discrepancy of the average number of signals present in the normalized data of 5 mN/s loading rate leads us to conclude there is indeed a change in fracture behavior of the specimen under different loading rates. Detailed information on how fracture behavior changes with different loading rates will be discussed in Section 4.4.2.

Fig. 10 also reveals that the loading rate has no significant effect on the energy of the AE signals. Other AE parameters (i.e., rise time and duration time) have been studied, and there is not a clear relationship between these AE parameters and loading rates. The different values of these AE parameters result from the different types of AE events. As shown in Fig. 11, the loading rate did not affect the AE signal’s time domain resulting from the initial crack formation. Figs. 10 and 11 suggest signals themselves are independent of the equipment setup and loading rate. Therefore, the differences in scatterplots of AE energy vs. force result from the different fracture behavior of the specimen under various loading rates.

The first high energy AE signal of each indent was treated as the critical load signal. Fig. 12 plots the critical load at six different
Fig. 9. SEM images of indent cross-sections after indentation under a force of (a) 150 mN, (b) 200 mN, (c) 250 mN, (d) 300 mN, (e) 350 mN, (f) 400 mN.

Fig. 10. Scatterplot of AE signal energy with contact force at different loading rates.
loading rates. It is observed that the critical load increases with the increased loading rate. At low loading rates, the large error bars prevent us from assigning any relationship between an increase in loading rate and the corresponding critical load. However, at higher loading rates, there is a clear increase in critical load with increased loading rate, which supports a conclusion of loading rate dependence of this sample type. The time-dependent behavior of copper could explain this behavior [56]. For the specimen used in this paper, the insulating layer acts as a beam in the bending process. And as the compressive deformation of the copper layer increases, the upper insulating layer will experience higher stresses. A previous study indicates that a larger loading rate indicates smaller imprint displacement at a given contact load [57], which would result in a smaller critical load. But, as shown in Fig. 13, there is no obvious trend in the load–displacement curve at different loading rates. To verify that effects of loading rate do not result from the copper layer, experiments on stack structures without copper layer needs to be done in the future. Another possible explanation is that brittle materials, like silicon dioxide, have slow crack growth behavior, which has been confirmed as a rate-dependent mechanism in crack growth [35–37]. These studies indicate that the materials’ strengths and critical loads increase with increasing loading rates. The results presented in Fig. 12 demonstrate this positive correlation, especially at higher loading rates.

Table 3
Initial and normalized average number of AE signals at different loading rates.

<table>
<thead>
<tr>
<th>Loading rate</th>
<th>Average number of signals</th>
<th>Initial</th>
<th>Normalized</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>12.9</td>
<td>6.36</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>4.9</td>
<td>3.33</td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>3.22</td>
<td>3.22</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 11. Time domain of AE signals that represents the initial crack formation at 5 mN/s and 1000 mN/s. The signal collected at 1000 mN/s has been offset for clarity.

Fig. 12. Plot of critical loads vs. loading rate. The error bars were acquired from the standard deviation of 75 repeat tests for each loading rate.
4.4. Signal processing

4.4.1. Unsupervised learning

Fig. 14 shows an example visualization of external evaluation of the k-means clustering using representative features with a dimension of seven learned by the autoencoder. Fig. 14(a) shows three indexes—Silhouette Criterion (red), Dunn Index (black), and Davie-Bouldin (blue), and Fig. 14(b) plots the voting points. The optimal number of clusters is 5. After defining the number of clusters by the three indexes, and using a voting scheme, Fig. 15 plots AE signal energy vs. contact force using four different types of AE parameters: (a) conventional extracted features, (b) normalized time-domain raw data, (c) short-term signal energy, and (d) seven representative features learned by the autoencoder. The number of optimal clusters in four cases are 4, 4, 3, and 5 as shown in legends of Fig. 15. Overall, Cluster C contains the low energy signals which we attribute to the copper layer plastic deformation studied by Unterreitmeier [23]. The author reported that these low amplitude signals were not present when indenting the specimen without a Cu layer. Cluster A represents the first peak cluster in all scatter plots, i.e., the initial crack formation signals. After obtaining the scatter plots and identifying the first peak cluster, F1 scores were calculated based on the external evaluation scheme. Cluster A is treated as “true,” while the other classes are treated as “false” in the confusion matrix. Table 4 shows an example of a confusion matrix using seven representative features learned by the autoencoder. Of the 2130 total signals, 180 signals are manually labeled as “true.” The manually labeled criterion selects the AE signals with a similar shape to the signal in Fig. 5(b), i.e., signals with high signal-to-noise ratio, low decay time, and smooth change in the envelope. The F1 score calculated from this confusion matrix is 0.685.

Table 5 shows a summary of the results from four types of parameters used in this paper. Revealed in Table 5, conventional extracted features have a low F1 score with 7-dimension input, and different clusters are mainly distinguished by the signal energy, as displayed in Fig. 15(a). The conventional AE signal features are clearly not suitable for this application. The raw data parameters have a relatively high F1 score and can distinguish different clusters that are not purely based on signal energy, as seen in Fig. 15(b). However, the raw data parameters’ 500-dimension input is so large that it slows down the K-mean clustering and makes it impractical for future application in automatically monitoring systems. Short-term signal energy parameters reduce the input dimensions but result in a lower F1 score with the highest recall value and lowest precision value. The low F1 score of this method is likely due to the
small number of clusters, which is obtained from the voting points that try to find the optimal cluster. As seen in Fig. 15 (c), the short-term signal energy method is able to distinguish different clusters based on AE signal patterns, rather than pure signal energy, which means better performance can be obtained by tuning parameters such as window length and the number of overlapped samples, or by combining this method with other feature extraction methods. Lastly, the autoencoder representation features not only reduce the input dimensions but also increase the F1 score. The cluster pattern of autoencoder features seen in Fig. 15 (d) is similar to that of the raw data from Fig. 15 (b), except for the total number of clusters, which suggests that the autoencoder method is able to reduce the input dimensions with less information lost.

**Fig. 15.** AE signal energy vs. contact force of up to 5 clusters using 4 different types of AE parameters: (a) conventional extracted features, (b) normalized time-domain raw data, (c) short-term signal energy, and (d) seven representation features learned by the autoencoder.

**Table 4**
An example confusion matrix using seven representative features learned by the autoencoder.

<table>
<thead>
<tr>
<th>Actual Labeled signals = no</th>
<th>Predicted Labeled signals = no</th>
<th>Predicted Labeled signals = yes</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual Labeled signals = no</td>
<td>1911</td>
<td>39</td>
<td>1950</td>
</tr>
<tr>
<td>Actual Labeled signals = yes</td>
<td>66</td>
<td>114</td>
<td>180</td>
</tr>
<tr>
<td>Total</td>
<td>1977</td>
<td>153</td>
<td>2130</td>
</tr>
</tbody>
</table>

**Table 5**
Summary of performance of four AE parameters.

<table>
<thead>
<tr>
<th>Input dimension</th>
<th>Optimal number of clusters</th>
<th>Recall</th>
<th>Precision</th>
<th>F1 score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conventional extracted features</td>
<td>7</td>
<td>4</td>
<td>0.539</td>
<td>0.678</td>
</tr>
<tr>
<td>Normalized time-domain raw data</td>
<td>500</td>
<td>4</td>
<td>0.633</td>
<td>0.699</td>
</tr>
<tr>
<td>Short-term signal energy (Frame: 128, Overlap: 72)</td>
<td>7</td>
<td>3</td>
<td>0.992</td>
<td>0.416</td>
</tr>
<tr>
<td>Representation features by the autoencoder</td>
<td>7</td>
<td>5</td>
<td>0.633</td>
<td>0.745</td>
</tr>
</tbody>
</table>
To further improve the F1 score, the short-term signal energy feature extraction method and autoencoder were used simultaneously to create the combination method. The dimension of raw time-domain AE signals was first reduced by calculating the short-term signal energy to remove the high-frequency information, and then the autoencoder algorithm was used to get the low-dimension representation features. Different values of three parameters, i.e., the window length, the number of the overlapped samples of short-term energy, and the number of autoencoder representation features, were optimized to obtain the highest F1 score by looping over these parameters. The window length is 64, and overlapping is 20; then after the first short-term signal energy dimension reduction, the number of inputs to the autoencoder is 10. The representation features dimension is two, so the number of the inputs to k-means is two.

Fig. 16(a) shows the scatter plot of AE signals vs. contact force using the combination method with the highest F1 score. The optimal number of clusters is 7, and the F1 score is 0.763 with the recall of 0.839 and the precision of 0.699. Fig. 16(b) shows the scatter plot of two representative features and corresponding clusters. The discussion of the physical meaning of each cluster based on classification results using this combination method is in the next section.

4.4.2. Supervised learning on different loading rates

To further understand the physical meaning of different clusters of AE signals, supervised learning on different loading rates was used. Fig. 17(a) and (b) show the scatter plots of AE signals vs. contact force and two representation features, respectively, using supervised learning – k-NN, at a loading rate of 100 mN/s. Fig. 17(c) and (d) show the same information but at a loading rate of 1000 mN/s. The features were extracted by the combination method with the same parameters as that of Fig. 16. Different clusters were classified using the k-NN. The training instances and corresponding labels of k-NN were the extracted features and their labels, separately, from previous unsupervised learning, and the input data were collected using a loading rate of 5 mN/s. Fig. 17(a) and (c) reveal that the combination feature extraction approach successfully distinguishes the same signal patterns at different loading rates.

As previously discussed, Cluster A represents the initial crack initiation, Cluster B is most likely the top layer crack opening, and Clusters C, F, and G are low energy signals which may not be associated with cracks and treated as the copper layer plastic deformation signals and noise signals. To investigate the physical meaning of Clusters D, Table 6 shows the percentage of each class, except for the low energy clusters (C, F, and G). Cluster D shows a significant decrease at higher loading rates. The hypothesis is that they are related to the crack surface friction. The crack surfaces are rough in texture with sharp peaks and valleys [58,59]. The crack surface friction induced by applying load can generate AE signals. Joseph. et al. concluded that rubbing and clapping of rough crack surfaces generated AE signals [60,61]. This hypothesis is supported by observing that the signals from Cluster D mainly occur between Cluster A (the initial crack formation) and Cluster B (top layer crack opening) shown in Fig. 16(a). After the crack opening expands, the friction phenomena decrease, which is consistent with the observation that the number of Cluster D signals decreases after Cluster B. From the comparison between the percentage of Cluster D at different loading rates and what has been discussed in Section 4.3 – the total number of AE signals decreases from 5 mN/s to higher loading rates, we conclude that as loading rates increase, the crack surface friction effects decrease.

Cluster E is adjacent to Clusters B, D, and F, as seen in the representation features plots, i.e., Figs. 16(b), 17(b), and (d). Some signals in Cluster E have relatively high energy, which happened around Cluster B, and some signals in Cluster E have low energy, as seen in Figs. 16(a), and (c). K-means algorithm clusters the data into spherical shapes, and it tends to produce equal-sized clusters [62]. Therefore, k-means fails to learn a meaningful Cluster E using signals at 5 mN/s loading rate. Therefore, Cluster E is the combination of the copper layer plastic deformation, noise signals (Cluster F), crack surface friction signals (Cluster D), and crack opening signals (Cluster B). K-means is the simplest clustering method with limitations. Further investigation on clustering method is needed in the future for better clustering results.

![Fig. 16. Scatter plot of (a) AE signals vs. contact force using the combination method, and (b) two representation features obtained by the autoencoder.](image-url)
5. Conclusions

In this study, a novel integrated system composed of a commercial high-resolution nanoindentation system and an acoustic emission sensor/indenter system, as well as a cluster-based signal processing method for crack classification is described. The suitability of using the acoustic emission method for crack detection in thin film stack structures (test structures) was verified by “pop-in” phenomena of load–displacement curves and SEM images of indent cross-sections. The influence of loading rates on the crack formation was observed. At higher loading rates, the critical load increases with increasing loading rate. A signal processing approach was applied to the acoustic events to classify the initial crack formation signals, the copper layer plastic deformation signals, and other crack related signals. Unsupervised learning (k-means clustering) was used to select the feature extraction methods and obtain the training dataset for supervised learning. Four feature extraction methods used were (a) conventional extracted features, (b) normalized time-domain raw data, (c) short-term signal energy, and (d) representation features learned by the autoencoder, and the performance of these methods was evaluated. The combination method that uses short-term signal energy parameters with representation features learned by the autoencoder has fewer input dimensions to k-means and the best F1 score. The k-NN method, a supervised learning technique, was applied to AE signals of different loading rates to learn the physical meaning of different clusters and build a framework for further data processing. The SEM images have been used in comparison with the AE signal clusters for better understanding of damage mechanisms in this type of thin film stack structure. In addition, the comparison between AE signal datasets of different

Fig. 17. Scatter plot of (a) AE signal energy vs. contact force, and (b) two representation features at a loading rate of 100 mN/s. (c) AE signal energy vs. contact force, and (d) two representation features at a loading rate of 1000 mN/s. These data were classified using supervised learning.

Table 6
The percentage of Clusters A, B, D, and E at different loading rates.

<table>
<thead>
<tr>
<th>Loading Rate (mN/s)</th>
<th>Cluster A</th>
<th>Cluster B</th>
<th>Cluster D</th>
<th>Cluster E</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>22.5%</td>
<td>22.7%</td>
<td>28.2%</td>
<td>26.5%</td>
</tr>
<tr>
<td>100</td>
<td>44.4%</td>
<td>33.8%</td>
<td>6.0%</td>
<td>15.8%</td>
</tr>
<tr>
<td>1000</td>
<td>39.75%</td>
<td>27.3%</td>
<td>8.1%</td>
<td>24.8%</td>
</tr>
</tbody>
</table>
loading rates is helpful in differentiating AE signals due to the crack surface friction. This signal processing approach has the promising ability to distinguish AE events associated with crack formation, crack propagation, crack surface friction, the copper layer plastic deformation and low energy noise signals. Further failure analysis and optical inspection, such as SEM confirmation, are needed to verifiably label the source of the clusters through signal classification and supervised learning.

**CRediT authorship contribution statement**

**Chen Liu:** Conceptualization, Methodology, Data curation, Validation, Formal analysis, Investigation, Writing - original draft, Visualization. **Oliver Nagler:** Conceptualization, Methodology, Resources, Data curation, Writing - review & editing, Supervision, Project administration. **Florian Tremmel:** Software, Data curation. **Marianne Unterreitmeier:** Conceptualization, Methodology, Resources. **Jessica J. Frick:** Writing - review & editing. **Radhika P. Patil:** Resources. **X. Wendy Gu:** Resources, Writing - review & editing, Supervision. **Debbie G. Senesky:** Resources, Writing - review & editing, Supervision, Funding acquisition.

**Declaration of Competing Interest**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

**Acknowledgements**

This work was supported by Stanford SystemX Alliance. The hardware used is supported by Infineon Technologies AG. Part of this work was performed at the Stanford Nano Shared Facilities (SNSF), supported by the National Science Foundation under award ECCS-2026822.

**References**


