**Estimation of Wire Bonding State by the Ensemble Based MT Method and Thin AE Sensor**

Shuichi Ishida*, Tatsuo Tabaru*, Wataru Iwasaki*, and Hiroyuki Miyamoto**

*Advanced Manufacturing Research Institute, National Institute of Advanced Industrial Science and Technology, 807-1 Shuku Tosu, Saga 841-0052, Japan

**Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, Fukuoka 808-0196, Japan

(Received July 5, 2016; accepted October 19, 2016, published November 21, 2016)

**Abstract**

With non-destructive, simple, and low-priced testing methods (which can be applied to manufacturing processes), we can assure high quality production by applying inspection across all products. Quality assurance in wire bonding is possible by using state diagnosis which utilizes detection of elastic waves close to the bonded joint. We succeeded in detecting elastic waves under difficult conditions and in harsh environments by using a thin AE (Acoustic Emission) sensor developed using AlN piezoelectric thin film technology. We attempted to conduct a wire bonding state estimation using the MT (Mahalanobis-Taguchi) method which is a simple and powerful pattern recognition technique. Two issues are raised when applying this method to the manufacturing process: 1) a large probability bias caused by an insufficient number of data samples, and 2) securing homogeneity of the Unit Space. In this paper, we present a state diagnosis method utilizing a real-world application which uses piezoelectric thin film sensing and an improved MT method to which ensemble learning is applied. Performance of this proposed method is also examined through a common benchmark dataset to assure efficient operation in any manufacturing process.

**Keywords:** Wire Bonding, Inspection Accuracy, MT Method, Ensemble Learning, Thin AE Sensor

1. **Introduction**

The acceptance/rejection decisions in wire bonding are usually performed through sampling and destructive testing in semiconductor manufacturing. Bond pull tests, bond peel tests, and bond share tests are commonly used in these destructive inspections. For example, the bond pull test evaluates breaking strength and mode of destruction by hooking a bonding wire and pulling it. However, the test results do not always reflect the actual bonding strength because of a shift in the pulling position or similar action. Regardless of optimization of sampling inspection, some distribution of a defective product is inevitable. To prevent this outflow of defective products, it is necessary to develop non-destructive inspection methods capable of performing total inspection.

Proposed methods of non-destructive inspection include imaging, and X-ray inspections.[1] These inspection methods decide upon acceptance and rejection of the wire bonding after the connection process is completed. The accuracy of inspection can be improved by evaluating effective information about ongoing process states and the device condition during the process itself. For example, the applied power and stability of ultrasonic vibrations have an effect on bonding quality in LSI manufacturing. There is a fair chance for extracting information related to diagnosis of the state by capturing emitted elastic waves during the bonding process. In accordance to above considerations, we focused on Acoustic Emission (AE) methods. In our previous research, we developed a thin AE sensor and reported on the possibility of application for state diagnosis of Al wire bonding in mass production.[2] The sensor developed has good heat resistance characteristics, having an operational temperature range of more than 250°C. The sensor is based on AlN piezoelectric thin film technology and has a thin detection unit of about 1 mm thickness. Sensing near to a bonded joint is difficult because of high temperatures and a small installation space. This difficulty is overcome by a thin AE sensor...
which can be set near to a bonding joint. We are aiming to perform a non-destructive, total inspection using AE signals which have excellent quality signal by using a thin AE sensor. It is particularly important to develop a suitable algorithm using any detected signals to achieve a high quality judgment.

Recently, it is becoming possible to process large amounts of information due to the growth of Big Data processing capabilities. Over the years a number of laboratories have studied machine learning, such as Deep Learning, to extract useful knowledge from a huge amount of information.[3] On the other hand, it is still a major challenge to collect enough data during a manufacturing process. At the starting point of a production line operation, in particular, it is apparent that accuracy of anomaly detection is strongly influenced by an insufficient dataset. In such situations, it is desirable that anomaly detection by simple methods is available. The Mahalanobis-Taguchi (MT) method, which is a simple and powerful pattern recognition system, has been proposed in the field of quality engineering.[4, 5] Inspection, diagnostics, monitoring, estimation, prediction, and classification using the MT method have been extensively reported in many cases. A general diagnosis method can determine abnormal states from any general characteristics of abnormality obtained from previous research into abnormalities. However, there are limitations on the ability of systems to detect individual abnormalities since causes of abnormality are not always homogeneous. Conversely, the MT method can determine what a normal population is. The MT method performs quality determination using a similarity degree between a normal population referred to as the “Unit Space”, and a target sample.

The MT method has a rare and unique characteristic of defining the scale of abnormality. MT method does however have some limitations, such as its practical use and unsolved numerical instabilities. There have been no general solutions to this as of yet. In particular, two major shortcomings are raised when applying this method to manufacturing processes: 1) a large probability bias caused by an insufficient number of data samples, and 2) securing a homogeneous Unit Space. When the number of data samples in the Unit Space is not sufficiently large with respect to the dimension of any given variables, a significant degradation in determining performance has been observed.[6] Furthermore, individual different users judge sample data by different standards. Therefore, performance varies in consequence of differently constituted Unit Spaces.

In this paper, we present an inspection technology of wire bonding for achieving total inspection in a non-destructive way by utilizing piezoelectric thin film sensing and the “Ensemble based MT method” for manufacturing processes. The Ensemble based MT method was based on ensemble learning aimed at overcoming the above mentioned two major shortcomings. The performance of the proposed method is examined through an example consisting of common benchmark data which has been presented by Cortez et al.[7] and recognition of its accuracy is compared to conventional methods. In addition to this examination, we also conducted bonding quality determination using the Ensemble based MT method and piezoelectric thin film sensing to establish an efficient and effective inspection algorithm.

2. Proposal of the Ensemble Based MT Method

2.1 Description and drawbacks of the MT method

MT method defines a homogeneous normal population as the “Unit Space”, and calculates the distance between the center of the Unit Space and the target sample as the Mahalanobis distances (MDs). The calculated MD is compared to a predetermined threshold value to judge normal items and abnormal items. In quality determination, an item which belongs within the Unit Space is regarded as a good product. This section will summarize procedure of the MT method that is based on Woodall et al.[8] The procedure is categorized mainly into four stages.

Stage 1: Define the Unit Space and calculate fundamental statistic.
Stage 2: Calculate the MDs for the abnormal items.
Stage 3: Identify the most useful set of variables using S/N ratio and orthogonal arrays.
Stage 4: Calculate the thresholds using a quadratic loss function and provide a diagnosis.

The details are as follows.

Stage 1, collect the number of normal state samples and the number of variables for the Unit Space. The observation of the $i$th variable on the $j$th item is denoted as $V_{ij}$, $i = 1, 2, \cdots, p; j = 1, 2, \cdots, m$. Thus we obtain the standardized values $Z_{ij}$ by subtracting the mean of the variable, and dividing it by its standard deviation as

$$Z_{ij} = \frac{(V_{ij} - \overline{V}_i)}{S_i}, \quad (i = 1, 2, \cdots, p; j = 1, 2, \cdots, m),$$

where $\overline{V}_i = \frac{1}{m} \sum_{j=1}^{m} V_{ij}$ and $S_i = \sqrt{\frac{1}{m-1} (V_{ij} - \overline{V}_i)^2}$.

The MDs values of the $j$th item are calculated by
where \( z_{j}^{T} = [Z_{1j}, Z_{2j}, \cdots, Z_{p_{j}}] \), and \( \mathbf{S} \) is the sample correlation matrix calculated as

\[
S = \frac{1}{m} \sum_{j=1}^{m} z_{j} z_{j}^{T}.
\]

Here superscript \( T \) denotes the transpose.

Stage 2, extract \( t \) items not belonging to the Unit Space and calculate the MDs based on fundamental statistic of the Unit Space. Thus we have \( \text{MD}_{j}, j = m + 1, m + 2, \cdots, m + t \), as defined in (2).

Stage 3, evaluate useful set of variables using the orthogonal arrays which most efficiently reduce the number of experiments. In the process of variable selection, evaluate effectiveness of each variable and decide by S/N ratio whether to use it or not. Many different S/N ratios are proposed in quality engineering such as static response (larger-is-better, smaller-is-better, nominal-is-best, and zero-is-best S/N ratio) and dynamic characteristics (dynamic type S/N ratio). For example, the Taguchi’s larger-is-better S/N ratio is defined as

\[
-10 \log_{10} \left( \sum_{j=m+1}^{m+t} \frac{1}{\text{MD}_{j}} \right).
\]

Stage 4, execute a diagnostic and a forecast using the most useful set of variables for “Signal Data” which were not selected for membership in the Unit Space as samples.

In this paper, we summarize the issues related to applying this method to a manufacturing process. MT method is based on the idea that the “Unit Space should be as homogeneous as possible”. The design procedure of the Unit Space is considered an important factor in the reliability of the recognition process. However, it is difficult to select data that fits a dense population in manufacturing processes. What are the criteria of designing a Unit Space? Is it desirable to select only the data that passes the quality inspection or has ideal characteristics? The ability to extract features which directly express the normal state also requires experience and technical knowledge. Hence it can be understood that the design procedure of the Unit Space depends to a significant degree on one’s versatility in MT method. Some researchers point out the following problems which can occur in the design procedure. If the number of data samples which are used to define the Unit Space is insufficient, MT method has a large probability bias. The number of the data samples need to be more than 10-times larger than the number of the variables even when the Unit Space follows a multivariate normal distribution. A method is proposed to solve these problems by applying the Feature Bagging.[9, 10] Although this method is effective in small sample data sets, the technique does not teach how to design the Unit Space. In this paper, the size of a small sample data is defined as one less than 10-times as much as the number of the variables. The Feature Bagging is described in detail in the following section.

It is very difficult to collect sufficient sample data for the Unit Space when applying MT method to a manufacturing process. Moreover, the performance degree caused by different choices of sample data and feature extracting depends on the user. Hence, it is difficult to utilize the performance of MT method to its full potential. We are not concerned in this paper with the problems of MT method’s theory, but other papers point out the following issues. As can be seen from (2), MT method requires calculating the inverse of a sample correlation matrix \( \mathbf{S} \). If the MD cannot be calculated due to a non-computability issue or instability of the inverse matrix, then multicollinearity exists.[11] Furthermore, it is pointed out that the MT method has limitations for analysis using nonlinear data.[12] Accordingly, we propose the “Ensemble based MT method” adopting ensemble learning in order to benefit from the excellent performance of MT method.

### 2.2 Description of the Ensemble based MT method

In this section, we provide a detailed overview of the Ensemble based MT method. Ensemble learning generates a prediction model by combining weak classifiers from a given training example set. In the case of classification or regression, any weak classifiers which are obtained are aggregated by majority voting or averaging. Before addressing the Ensemble based MT method, we briefly summarize an existing improved method based on Feature Bagging.[9, 10] Feature Bagging is one of the ensemble learning methods based on subsampling of the variables and contains the following steps.

Step 1: Obtain partial sets of variables by bootstrap sampling.

Step 2: Apply some kind of a classification algorithm, such as MT method to each dataset.

Step 3: Provide the final score by averaging for each partial set.

By contrast, our Ensemble based MT method is inspired by Feature Bagging, but is extended further on to subsampling of both sample data and its variables. The procedure is categorized mainly into five steps.

Step 1: Extract \( m \) sample data for the Unit Space and
Step 2: Obtain partial sets of both the sample data and the variables by bootstrap sampling.
Step 3: Apply MT method to each dataset.
Step 4: Evaluate the utility of each extracted partial set by calculating MD for Signal Data and perform determining the acceptance or rejection.
Step 5: In the case of classification, aggregate the weak classifiers that have met criteria by majority voting.

The particular feature of this proposed method is using subsampling not only for the variables but also for the sample data. Additionally, the final result is provided limited to the carefully selected individual partial set so the recognition accuracy is expected to be improved. Various methods are considered as adoption criteria. In this paper, to obtain the adoption criteria, conventional MT method is applied to the whole dataset which is taken before subsampling to discriminate the Signal Data. Thus, the adoption criterion is obtained as the accuracy rate of the result. By adopting the candidates which are superior to the Unit Space of conventional MT method, Ensemble based MT method can establish a prediction model which has several partial sets of the Unit Space. As a result, these solve the performance variation caused by different choices of sample data and feature extracting dependent on the user.

3. Performance Verification of the Ensemble Based MT Method

3.1 Problem setting
The validity of the proposed method is verified by applying a dataset “Wine Quality” which was discussed by Cortez et al.[7] Cortez et al. proposed a data mining technique to predict human wine taste preferences using Support Vector Machine (SVM). Details of the dataset are as follows.

The dataset is configured of 4,894 examples and 11 variables based on physicochemical tests: fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH, sulphates, and alcohol. Each sample was evaluated in 10 stages by sensory assessors. According to the relabeling method of Sato and Kuroki, a quality score of five or more is considered a high grade and the others a low grade.[12] By applying
the above procedure, we obtained 4,715 examples of high-grade and 183 examples of low-grade.

We compare our proposed method with conventional MT method, and an improved method based on Feature Bagging, using the Unit Space of high-grade wine samples. In the experiment, we established the following setting. Randomly select \( N_o \) (\( N_o = 2^5, 2^6, 2^7, 2^8, 2^9 \)) data samples chosen from the 4,715 high grade samples with 11 variables for the Unit Space. In this way 20 sets of \( N_o \) data samples for each condition are prepared. Signal Data are as follows.

Good: Good data is sampled by a prescribed number from the remaining high-grade samples of the Unit Space.

Defective: All 183 low-grade samples.

The number of the Unit Space is set to the maximum value \( 2^9 \) with the goal to use all the low-grade samples and to keep the composition rate as close as possible to the original one. The composition ratio of the produced Signal Data is 22.97 to 1, which is close to the original ratio of 25.76: 1.

### 3.2 Experimental results

In the experiments we compare accuracy rate for the Signal Data of each method by the increase of the data samples in the Unit Space. An accuracy rate is defined as 
\[
\frac{(A + D)}{(A + B + C + D)}
\]
in accordance with Table 1. The threshold level employed for the determination is set to \( \chi^2(11,0.05) \) since the MD follows the \( \chi^2 \) distribution with \( f = 11 \) degrees of freedom. The determined random sampling condition for the Ensemble based MT method is as follows.

Data samples: \( N_o / 2 \sim N_o - 1 \). Variables: 5-10. Thus, the number of data samples is always larger than the number of variables. In the proposed method, the number of extracted partial sets is set to a half of the data samples \( N_o \). The determined random sampling condition for the Feature Bagging is as follows.

Variables: 5-10. In the Feature Bagging, the number of extracted partial sets is set to 30. Then, the examination is repeated 20-times, with means and standard deviations of accuracy rate calculated. When the standard deviations of accuracy rate are low, users can construct the recognition system easily without technical know-how. The adoption criterion defines an accuracy rate of the conventional MT method. In the final result, the majority voting is performed between Good and Defective. According to the performance of the classifier, a polling score is weighted, and then, the aggregation is done.

Figure 2 shows comparison of each method. The blue solid line, the green dash-dotted line, and the black broken line indicate Ensemble based MT method, Feature Bagging, and conventional MT method respectively. The horizontal axis indicates the number of data samples in the Unit Space and the vertical axis shows the accuracy rate. The error bar shows the standard deviation of the accuracy rate.

Firstly, we examine the performance of MT method. The accuracy rate increased with an increase in the number of data samples of the Unit Space. In the case of 512 data samples, the accuracy rate was 88.2%, an increase of about 23% percentage points when compared to 32 data samples. When there was more than 10-times the number of variables data difference, such as in the case of 128 and 512 data samples, the increase of accuracy rate was only about 3% points. It has been shown that a large probability bias is caused by an insufficient number of data samples. Moreover, the standard deviation decreased with an increase in the number of data samples of the Unit Space. Thus, the performance variation is caused by constituting the Unit Space. The known practical issues were confirmed by these experiments. When Feature Bagging was applied to MT method, the accuracy rate improved. In the

| Prediction | Good | Defective |
|------------|------|-----------|
| True       | A    | B         |
|            | C    | D         |

Table 1: Dividing table.

Fig. 2: Performance comparison ("Wine quality" benchmark).

Closed circle: Ensemble based MT method, Square: Feature Bagging, Diamond: conventional MT method.
In the case of 32 data samples, the performance increased for about 7.7% points compared to the conventional MT method. The standard deviation was not improved so much since the reduction value was only 0.1% points. On the other hand, the performance of Ensemble based MT method in comparison with the conventional MT method and Feature Bagging is clearly superior. When the number of data samples is 32, the accuracy rate of the Ensemble based MT method increased for about 16.2% points compared to the conventional MT method. Additionally, the standard deviation decreased by about 4.1% points. The Ensemble based MT method leads to suitable selection of the Unit Space from the candidate dataset. This indicates that the proposed method is appropriate for the case of small data samples in the Unit Space.

In the last part of this section, we explain the adoption criteria. Figure 3 shows the difference in accuracy rate between the classifier and the criteria. The data is sorted in descending order in the case of 32 data samples. If the difference is positive, the performance is better than the conventional MT method. When the criterion for each weak classifier is applied, 14 classifiers were adopted for construction of compositely recognizing space. As a result, the Ensemble based MT method shows excellent performance compared to conventional MT method, and to Feature Bagging. For further reference, the accuracy rate variation to cause the adoption criteria to worsen is given in Fig. 4.

The adoption criterion for the accuracy rate was set to mean + SD (Proposed (Mean + SD)) from mean (Proposed (Mean)). As can be seen from Fig. 4 and Table 2, the performance of Proposed (Mean+SD) is significantly improved relative to the Proposed (Mean). In particular, compared to the conventional MT method, when the number of data samples is 32, the accuracy rate of Ensemble based MT method is better for about 19.1% points. Additionally, the standard deviation is lower for about 4.6% points. When the number of the data size is small, the accuracy rate of each partial set is varied. On the other hand, the performance variation of partial sets become

### Table 2 Accuracy rate.

| Mean (SD): [%] | 32   | 64   | 128  | 256  | 512  |
|---------------|------|------|------|------|------|
| Conventional  | 64.8 (7.7) | 80.9 (3.1) | 85.3 (2.2) | 87.7 (1.8) | 88.2 (1.1) |
| Feature Bagging| 72.5 (7.6) | 84.0 (2.9) | 86.9 (1.7) | 88.1 (1.1) | 88.7 (0.9) |
| Proposed (Mean) | 81.0 (3.6) | 87.3 (1.0) | 88.7 (0.8) | 89.6 (0.7) | 89.6 (0.6) |
| Proposed (Mean+SD) | 83.9 (3.1) | 88.2 (1.0) | 90.1 (0.7) | 91.5 (0.6) | 90.8 (0.6) |

![Fig. 3 Difference of adoption criteria (dataset: 32, accuracy rate: 81.0%).](image)

![Fig. 4 Performance comparison (“Wine quality” benchmark).](image)
smaller as the number of data size rises. Thus, significant improvement of accuracy rate to conventional MT method is hardly obtained. When constructing a prediction model using a partial set which has superior judgment, the accuracy rate increased in small datasets and with less sensitivity to the variation caused by construction of the Unit Space was seen. However, further work is needed on the selection of appropriate adoption criteria.

4. Estimation of Wire Bonding State

4.1 Detection of elastic wave by thin AE sensor

In this subsection, an actual dataset is prepared for the testing of wire bonding state diagnosis, and the effects of the Ensemble based MT method is verified. Bonding samples which are applied to the Au plated pad on an alumina substrate and an Au wire (25.4 μm diameter) are performed with a tabletop type manual wire bonder (model 7400D, WEST BOND, Inc.). The thin AE sensor can detect an elastic wave which is emitted during the bonding process, and we perform the state diagnosis using the detected signal.[13] Control parameters of the thermal ultrasonic bonding device used are load, temperature, ultrasonic vibration, and ultrasonic energy. The nominal frequency is 63 kHz. These parameters were set by the manufacturer as the default values when creating bonding samples. The wire bonding is performed by well-trained staff who are operationally adequate in order to prevent variations in positions, height, and shape of the wire bonding.

We focused on AE sensing since there is a fair chance for extracting information related to diagnosing the state by emitted elastic waves during the bonding process. Although an AE sensor is effective for detecting elastic waves, a general AE sensor cannot be used in the case of wire bonding. The reason for this is in the fact that the stage temperature rises to 200°C. Therefore, we developed and used a thin AE sensor which has good heat resistance characteristic, having an operational temperature range of more than 250°C. The experimental apparatus for AE sensing is shown in Fig. 5. To detect the elastic waves close to a testing substrate, the thin AE sensor is mounted on a heat block. The detected signals are sent through an amplifier to an analog input device (NI 9223, National Instruments Inc.) and collected as data in PC (personal computer) memory. The signals are obtained at a sampling rate of 1 MS/s, and a typical waveform is shown in Fig. 6. Figure 6 is indicative of two large amplitude parts consisting of 1st bonding and 2nd bonding. Features concerning connecting state are extracted from the individual bonding part by the following two methods: (1) AE method, and (2) Power spectrum calculating. In (1), the maximum amplitude, duration time, AE count, rise time, and RMS (root mean square) values are calculated as a total of 5 AE fea-
tures as shown in Fig. 7. In (2), by dividing the entire frequency band obtained by FFT processing with the power spectrum of each AE signal into a plurality of frequency bands (band 1 – band 8) of a fixed bandwidth, the mean and variance are found. Then, a total of 16 feature values are obtained. In total 42 features are obtained from both 1st and 2nd bonding parts. The principal component analysis is applied in order to compress dimensionality and to remove high correlation of variables. In the component combining process, 8 variables which have eigenvalues greater than 1 were newly obtained. A series of processing is performed by developed system with an algorithm developed in LabVIEW (National Instruments Inc).

4.2 Experiments

The Signal Data consisted of good and pseudo defective samples. Good data samples were collected from bonding samples which were performed under the recommended conditions, and also included only the samples which had passed a bond pull test (PTR-1102, Rhesca Co., Ltd). Pseudo defective samples were prepared under intentionally changed default values. The defective samples were made by assigning to each the following condition factors using the orthogonal tables based upon the experimental planning method in statistics and quality engineering.

- Load: 2 levels (Low (default), High),
- Heat, Ultrasonic vibration, and Ultrasonic energy: 3 levels (-30%, default, +30%).

According to the above procedure, the dataset consisting of 108 good samples and 18 pseudo defective samples was created.

The validity of the Ensemble based MT method is verified by application to this dataset and the experiments were conducted in the same manner as in the preceding section. In the experiment, we first set the following setting. Randomly select \( N_0 \) (\( N_0 = 18, 27, 36, 45, 54, 63, 72, 81, 90 \)) data samples which were chosen from the 108 good samples for the Unit Space. 20 sets of data samples for each condition were prepared. The number of extracted partial sets is set to half of that of the data samples namely to \( N_0/2 \). The Signal Data was as follows.

Good: Good data are 108 sampled by the remaining good samples of the Unit Space.
Pseudo Defective: All 18 pseudo defective samples.

The examination was repeated 20 times, and the means and standard deviations of the accuracy rates were calculated. The adoption criteria were set at the accuracy rate of the conventional MT method in the same manner as above (Proposed (Mean)). For further reference, we changed the adoption criteria from Mean to Mean + SD in order to be more accurate (Proposed (Mean + SD)). The results are given in Fig. 8. The accuracy rate increased with an increase in the number of data samples of the Unit Space. Furthermore, the performance of both proposed methods was excellent compared with the conventional method.

When the number of data samples was small, (in particular at 18 samples), the accuracy rates were: Conventional: 78.1% (SD: 7.7), Proposed (Mean) 87.1% (SD: 5.2), and Proposed (Mean + SD) 89.9% (SD: 3.8). Thus, the accuracy rate of Proposed (Mean + SD) method is better by about 11.8% points compared to the Conventional method. Additionally, the standard deviation decreased by about 3.9% points. Thereby we found that the bond pull test and AE signals are correlated. Furthermore, the accuracy rate can be improved by Ensemble based MT method.

5. Conclusions

In this paper, for cases when data samples are inefficient (at the operation starting point of a production line, for instance), we proposed the “Ensemble based MT method”, adopting ensemble learning in order to benefit from the excellent performance of MT method. The validity of the Ensemble based MT method was verified by applying a dataset “Wine quality” compared with conventional MT method. When the number of data samples was small, the accuracy rate of the proposed method increased at best by about 19.1% points, and the standard deviation decreased at most by about 4.6% points compared to the conventional MT method. We showed that the proposed method is effective in achieving an improvement of the major shortcomings, which are large probability bias and unsatisfactory performance. In addition to the experiments on a common benchmark dataset, we also executed tests to
clarify the effectiveness of piezoelectric thin film sensing and Ensemble based MT method for wire bonding state diagnosis. The experimental results confirmed an increase in the accuracy rate compared to the conventional method, particularly when the number of data samples is small. It is necessary to verify the effectiveness of the proposed method in more detail, and to provide a theory of MT method, under conditions similar to an actual manufacturing environment. In the future, applications not only for wire bonding but also for other various tasks in manufacturing processes also need to be examined.

**References**

[1] D. B. Perng, C. C. Chou, and S. M. Lee, “Design and development of a new machine vision wire bonding inspection system,” The International Journal of Advanced Manufacturing Technology, Vol. **34**, pp. 323–334, 2007.

[2] S. Ishida, T. Tabaru, M. Akiyama, C. Kubo, and C. Miyazaki, “Detection of Elastic Wave during Al Wire Bonding by Thin AE Sensor,” IEEJ Transactions on Industry Applications, Vol. **134**, No. 9, pp. 840–841, 2014 [in Japanese].

[3] A. Krizhevsky, I. Sutskever, and G. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” Proc. Neural Information and Processing Systems, 2012.

[4] G. Taguchi and R. Jugulum, “The Mahalanobis-Taguchi Strategy: A Pattern Technology System,” Wiley, 2002.

[5] S. Teshima, Y. Hasegawa, and K. Tatebayashi, “Quality Recognition and Prediction: Smarter Pattern Technology with the Mahalanobis-Taguchi System,” Momentum Press, 2012.

[6] M. Miyakawa, K. Tanaka, T. Iwasawa, and H. Nakanishi, “The Error Rate in Mahalanobis-Taguchi System,” Journal of the Japanese Society for Quality Control, Vol. **37**, No. 1, pp. 101–106, 2007 [in Japanese].

[7] P. Cortez, A. Cerdeira, F. Almeida, T. Matos, and J. Reis, “Modeling wine preferences by data mining from physicochemical properties,” Decision Support Systems, Vol. **47**, pp. 547–553, 2009.

[8] W. H. Woodall, R. Koudelik, K. L. Tsui, S. B. Kim, Z. G. Stoumbos, and C. P. Carvounis, “A Review and Analysis of the Mahalanobis-Taguchi System,” Technometrics, Vol. **45**, No. 1, pp. 1–30, 2003.

[9] M. Ohkubo and Y. Nagata, “Consideration to Apply the Mahalanobis-Taguchi Method to Small Data,” Journal of Japan Industrial Management Association, Vol. **66**, No. 1, pp. 30–38, 2015 [in Japanese].

[10] A. Lazarevic and V. Kumar, “Feature Bagging for Outlier Detection,” Proceedings of the 11th ACM SIGKDD International Conference on Knowledge discovery in data mining, pp. 157–166, 2005.

[11] M. Miyakawa and Y. Nagata, “On Countermeasures against Multicollinearity in Mahalanobis-Taguchi System,” Journal of the Japanese Society for Quality Control, Vol. **33**, No. 4, pp. 77–85, 2003 [in Japanese].

[12] N. Sano and M. Kuroki, “The Kernel Based MT Method and Its Application,” Journal of the Japanese Society for Quality Control, Vol. **42**, No. 1, pp. 127–138, 2012 [in Japanese].

[13] Y. Kasashima, T. Tabaru, Y. Kobayashi, and M. Yasaka, “Detection of Micro-Arc Discharge using ESC Wafer Stage with built-in AE Sensor,” IEEE Transactions on Semiconductor Manufacturing, Vol. **26**, No. 3, pp. 350–354, 2013.
Shuichi Ishida received the B.S., M.S. and Ph.D. degrees in engineering from Kyushu Institute of Technology, Fukuoka, Japan in 2008, 2010, and 2013, respectively. He is currently a researcher at National Institute of Advanced Industrial Science and Technology (AIST), Japan. His current research interests are anomaly diagnosis in manufacturing process and machine learning.

Tatsuo Tabaru received the B.S., M.S. and Ph.D. degrees in engineering from Tohoku University, Miyagi, Japan, in 1992, 1994, and 1997, respectively. He is a leader of a research group at Advanced Manufacturing Research Institute, the National Institute of Advanced Industrial Science and Technology (AIST), Saga, Japan. In 1999, he joined the Kyushu National Industrial Research Institute, Agency of Industrial Science and Technology (currently AIST), Saga, Japan. He has also been a Guest Associate Professor with Graduate School of Science and Engineering, Saga University, Saga, Japan, since 2006. His current research interests include anomaly detection in manufacturing process, analysis of material degradation, and aluminum nitride piezoelectric film.

Wataru Iwasaki received the B.S., M.S. and Ph.D. degrees in engineering from Kyushu University, Fukuoka, Japan in 2008, 2010 and 2013, respectively. He is currently a researcher at National Institute of Advanced Industrial Science and Technology, Japan. His research interests are fabrication of MEMS devices and biochemical sensors for animal health monitoring.

Hiroyuki Miyamoto received the B.S., M.S. and Ph.D. degrees in biophysical engineering from Osaka University, Osaka, Japan, in 1985, 1987 and 1998, respectively. From 1994 to 1996, he was a member of research staff of Department 3, ATR Human Information Processing Research Laboratories, Kyoto, Japan. From 1996 to 2001, he was a member of research staff of the Kawato Dynamic Brain Project, ERATO, JST, Kyoto, Japan. Since 2001, he has been an associate professor of Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology, Kitakyushu, Japan. His main interest is in the area of brain-like intelligent machines.