Wafer-level Adaptive Testing Based on Dual-Predictor Collaborative Decision

Yuqi Pan[1] · Huaguo Liang[1] · Junming Li[1] · Jinxing Qu[1] · Zhengfeng Huang[1] · Maoxiang Yi[1] · Yingchun Lu[1]

Abstract

The growing complexity of integrated circuits (ICs) brings expensive manufacturing test cost. Adaptive testing becomes an important way to save test cost by predicting die quality to reduce the actual test content. However, reducing test items often results in unacceptable levels of test quality degradation. An adaptive testing method is proposed in the paper to reduce test cost while guaranteeing test quality. Two quality predictors are trained with a subset of test items and spatial information for subsequent decisions. The dies are clustered according to the prediction results, and the clustering results are graded. The distribution of the different grade classes determines the die quality of each grade. Experimental results using fabricated wafers and the associated test data show that the proposed method reduces more than 42% of test items, and can achieve better test quality, reducing test escapes and yield losses by more than 90% in the Circuit Probing test (CP).

1 Introduction

Due to the increasing quality expectations of IC testing, efficient testing has become crucial towards meeting the stringent requirements of the market [1]. Test cost limits the time available to test each chip. How to reduce the cost of IC testing under the premise of ensuring test quality has always been an issue of concern.

The parametric test is mainly based on measuring whether the circuit performances meet the specifications guaranteed by the manufacturer. It ensures good test quality but requires a longer test time, resulting in high testing cost [2]. Measuring the full test set during production test is redundant and results in unacceptably long test times [3]. To address this, test content is customized for each die under test during production to achieve a significantly better test time/test quality trade-off. By accepting a small number of test escapes and reducing the test set, considerable test cost reductions can be achieved [4].

To achieve low-cost and high-quality testing, adaptive testing has been proposed. It can predict die quality by mining test data, which uses partial test information to predict whether a die will pass a complete test. The standard test applies all the test items sequentially according to the test set, and if all items pass, the product is deemed good, otherwise, it fails [5]. Adaptive testing leverages prediction results to adjust test content, test sequence and test thresholds, thereby reducing test time and enhancing test quality. By exploring the trade-off between removing test items and prediction accuracy, production costs can be reduced without compromising quality testing accuracy. In the best-case

scenario, adaptive test simulation would significantly reduce test time without affecting test quality [6].

Loosening test specifications can cause test quality issues. For instance, although reducing test content can save the subsequent expensive test, incorrect prediction results can lead to a more significant loss of revenue [7, 8]. Since adaptive testing relies on the results of partial test items to predict product quality, there is a risk of degradation in test quality. The trade-off between prediction accuracy and test cost becomes an important consideration. Test quality is measured in terms of yield losses and test escapes, as shown in Fig. 1. The over or under testing and test equipment performance can lead to yield losses and test escapes. Test quality degradation from adaptive testing prediction errors is similar to actual testing. A good product is predicted to fail, resulting in yield losses, and when a bad product is predicted to pass brings about test escapes.

Adaptive testing primarily uses the test data correlations to predict die quality. In wafer testing, the common correlations are test item correlations and spatial correlations, with extensive research dedicated to these areas. Test item correlations occur because strongly correlated tests are highly likely to result consistent test conclusions, resulting in correlated performance in the test data. For example, a resistive open or short defect may show a highly consistent trend in the measured values of DC current and voltage parameter tests. Y. Makris et al. [9] established the strong correlation between the test criteria and the performance parameters of the circuit by the neural classifier for analog circuits. H. El Badawi et al. [2] explored the trade-off between test quality and test cost and proposed a two-tier test process. Krishnendu Chakrabarty et al. [10] from Duke University proposed a transfer learning algorithm to construct Bayesian network models to extract strong relationships among tests. C. -Y. Hsieh et al. [11] used deep learning, and Liu M et al. [12] utilized deep learning and online incremental learning techniques, respectively, to process real test data in a mass production environment.
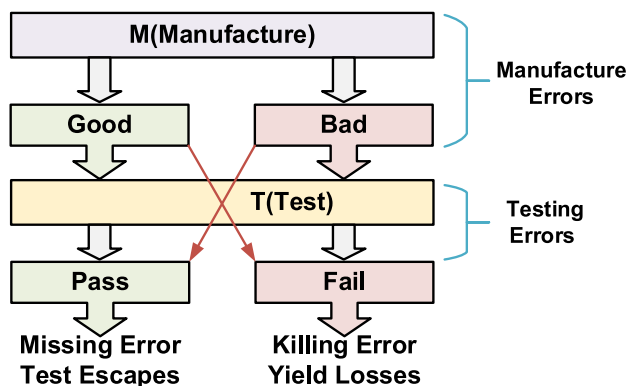
Since adjacent dies in a wafer are fabricated in a similar environment in semiconductor manufacturing, spatial correlations become a common data characteristic in wafer-level testing. During fabrication, process variations in space are similar between adjacent dies [13]. M. Li et al. [14] and Y. Makris et al. [15] predicted die quality by the density of the faulty die distribution in the region, and M. Eiki et al. [16] by building a novel wafer-level spatial correlation model. N. Chen et al. [17] predicted the variation of the die parameter response by capturing the spatial correlations. R. Jia et al. [18] proposed the LOF-KNN algorithm to detect whether the wafers are local outliers.

Isolated statistical data analysis may miss significant intra-process/test correlations. The predictor accuracy based on spatial correlations is greatly affected by the distribution of faulty dies in the wafer, especially in the region of sudden change quality change. Test item correlations can help to identify dies with unpredictable spatial information [19]. Combining wafer-level spatial correlations and test item correlations can train a predictor that performs better than training with only one correlation information. Recently, the method of identifying die quality by multi-correlation analysis has been gradually studied. X. Wang et al. [20] proposed a screening method to reduce chip test escapes using a multi-correlation analysis of parameters. Y. Makris et al. [1] independently assessed the validity of two types of tests through a machine learning approach. Currently, the majority of methods use a single predictor to assess die quality, with fewer methods combining test items and spatial information. Utilizing multiple predictors to determine whether a die passes the test is more complex, but can moderate test ambiguity to provide lower test escapes and yield losses. The different predictive information also complicates identifying whether the dies pass the test. The combination of intra-die and inter-die models to reduce test cost is described in references [13] and [19], but there are shortcomings. Both methods reduce test items that can be predicted by linear combinations of other tests [21, 22], but are not effective in reducing test items that are nonlinearly correlated. On the other hand, these methods rely on several assumptions that are generally applicable in practice. For instance, if the assumption of sparsity is not valid for the ground truth of a certain test item, finding a sparse solution will not be sufficient to recover the spatial pattern of the test item [23, 24]. Reference [13] uses test data from neighboring wafers for training spatial models. The systematic spatial variation may exhibit a radial shift across wafers [25], and there is a risk that cross-wafer information acquisition is not adapted to the wafer under test. Furthermore, reference [19] requires high computational costs, and the algorithms must be run in real time during the testing process [26].

To address the aforementioned problems, we propose an adaptive test flow for the parameter testing in the CP

Fig. 1 Test escapes and yield losses

stage. Two high-accuracy predictors are trained with the test items and wafer local space features for predicting die quality. By leveraging the complementary performance of these two predictors, the negative effects of abrupt quality change regions on the wafer can be mitigated. Additionally, a based on the DBSCAN (Density-Based Spatial Clustering of Applications with Noise) classification by grade method is proposed to resolve decision conflicts between the two predictors. The dies are graded based on quality and screened for failures. The proposed method effectively reduces test costs while maintaining test quality.

The rest of the paper is organized as follows: Section II provides an overview of the key techniques required for adaptive testing. Section III describes the proposed dual-predictor collaborative adaptive testing flow. Section IV introduces the results of applying the proposed method to a real production test. Section V concludes the paper.

2 Preliminary

2.1 Feature Selection

After test data is collected, feature selection is required to choose a small set of important test items from a standard test set as the basis for training predictor. This is because even though all test items in CP are carefully designed to detect die failures, a large portion of test items are not informative for the purpose of quality prediction [13]. While unprocessed test data provides a wealth of information, it can also present the "curse of dimensionality", the law that exponentially increases data sparsity by adding dimensions. This data sparsity can lead to learning algorithms with high-variance classification boundaries and poor generalization capability. The complexity of the problem grows exponentially with the number of dimensions. Feature selection improves classification accuracy by providing a subset of relevant features that best describe the dataset, while reducing predictor complexity and helping to prevent overfitting [27]. The goal of the predictor is to achieve good prediction quality, which can be measured by some loss functions on the constructed predictor.

2.2 Random Forest Predictor

Quality prediction is the key to adaptive testing, using the results of partial test items or test information from neighboring dies is used to train a quality predictor [28]. If the prediction result is good, the subsequent test items can be omitted; if the prediction result is bad, the dies are discarded directly. Some of the supervised learning methods that have worked well in the industry include random forest (RF), principal component regression (PCR), support vector machines (SVM), and so on. The random forest has become popular among them due to its best-of-class performance combined with its relative simplicity, an ensemble predictor consisting of multiple weak classifiers (decision trees) [29]. RF can effectively analyze datasets with a large number of samples and datasets with high dimensionality (each sample has a large number of variables) and can provide analysis about how similar the samples are to each other, which is useful in clustering and outlier identification. The flow of how random forest is applied to adaptive testing is shown in Fig. 2. Randomly selected some dies from the wafers are used as the training set to train the predictor, and the remaining dies are used as the test set. The test item responses and spatial locations of the training set can be used as features of the training predictor, and die quality information (pass/fail) as labels. The predictor can output the prediction quality of the test set by simply inputting information about the test set features.

2.3 Bad Neighbor Ratio

Bad neighbor ratio (BNR) indicates the percentage of bad dies in a region. As shown in Fig. 3, different colors represent different qualities of dies. Due to the similar process manufacturing environment, the quality of the dies in close proximity in the wafer tends to be consistent, and this feature can be used to predict the quality of the dies at the center.
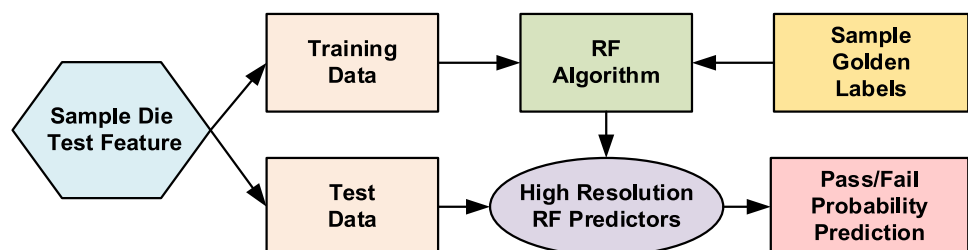
The BNR is calculated as follows:

$$BNR_{ij} = 1 - \frac{1}{n} \sum y_{ab}$$
$$i - 1 \le a \le i + 1, j - 1 \le b \le j + 1 \tag{1}$$

where BNR_{ij} denotes the BNR value of the die in row i and column j, and y_{ab} denotes the surrounding die quality (0 for

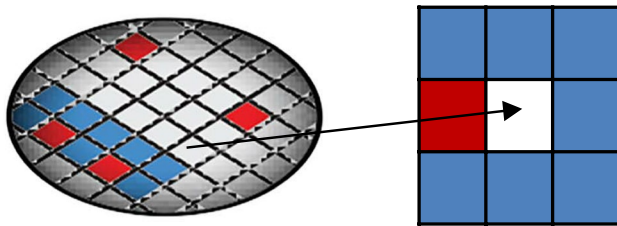**Fig. 2** Random forest predictor flow

Fig. 3 Wafer quality map

failure, 1 for pass), and n denotes the number of surrounding available dies. The relationship between BNR and die passing test probability was obtained after counting 2,946,004 dies in the reference [8], where $P_{negative}$ means the probability of a chip failing the test for the corresponding value of BNR. As shown in Table 1, the larger the BNR, the higher the probability that the die in the central region is failed.

2.4 Multi-predictor Collaborative Decision

The collaborative decision combines the results of multiple predictors to get a better quality result, which can reduce the risk of mis-selection and the possibility of the predictor falling into local traps, with higher accuracy and robustness. A single predictor is often mis-selected due to an oversized assumption space, which leads to degraded generalization capability. The common collaborative decision methods are weighted average score, voting method and learning method, among which learning method is widely used because it can be applied to many data types and can obtain better quality results.

3 Proposed Adaptive Testing Framework

The flow of the proposed method is shown in Fig. 4 and divided into the following three steps: 1. Training Single Predictor, 2. Dual-Predictor Collaborative Decision, 3. Test Selection. The adaptation of the proposed method is reflected in two aspects. 1. The wafer quality predictor is trained from the die test data in the respective wafer, which are more adaptable to reflect the wafers quality. It prevents differences between wafers from degrading the trained predictor performance. 2. The results of the predictor adjust the test content. For the dies with ambiguous predictor judgments, they can be retested, and the remaining dies can finish the test.

Table 1 Probability of die failure in different BNR intervals

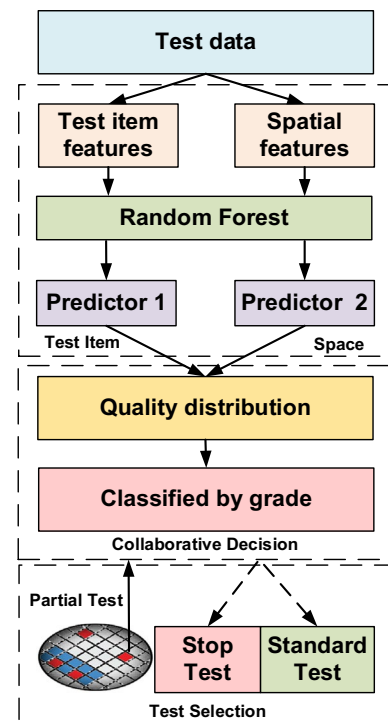| BNR	0%-20%	20%-40%	40%-60%	60%-80%	80%-100%
$P_{negative}$	0.63%	16.03%	51.12%	68.87%	95.24%

Fig. 4 The proposed adaptive testing flow

3.1 Training Single Predictor

(1) Training Test Item Predictor: The test item response is an important basis for reflecting die quality, and using it to train a die quality predictor can accurately distinguish most die qualities. However, not all test items are useful for training the predictor, and the test set needs to be filtered. The approach of predicting die quality with a partial test set gives an opportunity to save test cost.

A random sample of dies in a wafer is measured with a standard test set (apply all test items), and the resulting standard test data is used to train a quality predictor. In this paper, all test items are filtered using the Recursive Feature Elimination Cross Validation (RFECV) algorithm. The RFECV algorithm consists of two phases to implement test item selection. In the first phase, recursive feature elimination is performed to rate the importance of the test items. The sampled dies are categorized into training set and test sets. The test items in the test set are deleted sequentially, one test item at a time. The remaining test items are used as inputs to the Random Forest algorithm to train the quality predictor. A quality predictor is trained by removing one test item in each round. When the predictor has the highest accuracy, it means that the deleted test item has the least impact, so the lower the importance ranking. In the second phase, cross validation (CV) is performed to select the optimal number of test items. The dies are categorized multiple

times and the first phase of the test item importance ranking is repeated. Determine the number of feature test items when the predictor accuracy is highest.

The selected test items are used as features, and the binary pass/fail results of the CP are used as labels (the pass dies set to 1, and failed dies set to 0). The random-forest algorithm is used to learn from the combination of the features and the labels, and it trains a quality predictor. Each test item is considered as a node. Based on the sampled test data, the probability of a die passing the test is calculated for different test response. Multiple test items of a die constitute multiple nodes of a decision tree. Decision trees corresponding to large amounts of die test data can be trained as a random forest predictor. The random forest algorithm can output the parameter Score, which indicates the probability that the die is labeled 1, both the probability of passing the test. The predictor is a mapping relationship between partial test items as a function of die quality, as follows:where Q_i denotes the quality score of die i, and $item_a, \cdots, item_m$ indicates the feature test items filtered by RFECV. The dies to be tested are tested only with feature test items, which are used as input to the predictor to determine its quality.

$$Q_i = f\left(item_a, \ldots, item_m\right) \tag{2}$$

(2) Training Spatial Predictor: There is a clustered distribution of die quality in wafers, and the surrounding dies information can be used to predict center die quality. However, the region of abrupt change in die quality in the wafer can make the predictor less accurate. Therefore, the dies are graded, and then the die quality predictors are trained separately.

The dies in different quality regions exhibit different features, and the features test items of the faulty dies are filtered separately for each grade. The process is as follows:

1) The test item predictor in the previous subsection can describe the wafer quality prediction results, including the die's spatial location coordinates and quality prediction information.
2) The local spatial information (BNR) is calculated for each die to be tested and graded, as shown in Table 2. When BNR is close to 1 or 0, the dies are classified as easy to pass grade and easy to fail grade, respectively, because there is more certainty to judge the dies as pass or fail. The remaining dies are divided into ambiguous quality grades.
3) The test items of different grades of dies in the training set are filtered by applying the RFECV algorithm to the features separately. Based on the reference [8], dies with BNR in the interval [0,0.25] and [0.75,1] are considered as easy-to-pass grade and easy-to-fail grade, respectively, and dies in the BNR in the interval (0.25,0.75) are

Table 2 Training spatial graded predictor information

Chip Number	BNR	Spatial feature set	Quality label
1	Easy to pass	$\{item_c, \cdots, item_n\}$	P/F
2	Easy to fail	$\{item_e, \cdots, item_p\}$	P/F
⋮	⋮	⋮	⋮
n	Quality ambiguous	$\{item_j, \cdots, item_q\}$	P/F

considered as the ambiguous quality grade. The quality prediction results for each die are calculated based on the BNR input into the individual predictor.

The test item and BNR are used as features, and the binary pass/fail results of the CP are used as labels. The features and labels are fed into the random forest algorithm to train the quality predictor Q_j, as shown in Eq. (3).

$$Q_j = \begin{cases} g_1\left(BNR_j, item_c, \cdots, item_n\right), 0 \le BNR_j \le 0.25 \\ g_2\left(BNR_j, item_e, \cdots, item_p\right), 0.25 < BNR_j \le 0.75 \\ g_3\left(BNR_j, item_g, \cdots, item_q\right), 0.75 < BNR_j \le 1 \end{cases}$$

$$\tag{3}$$

3.2 Dual-Predictor Collaborative Decision

After the predictor evaluates die quality, the dies need to be classified based on the evaluation results. Setting simple thresholds to identify dies often does not achieve the expected accuracy. The distribution of the pass and failed dies is not strictly separated, which can lead to test escapes and yield losses. An improved algorithmic method based on DBSCAN clustering is proposed to distinguish die quality, which can identify irregular boundaries and outliers in clusters. Due to process fluctuations in quality during die manufacturing, the DBSCAN algorithm does not always classify the die in two classes. It clusters based on the distribution density of the predicted results, but is unable to determine the labels of the individual classes. The method is divided into the following three steps: (1) The dies are clustered into different clusters based on the predictor results. (2) The standard clusters are determined based on the distribution of clusters and ideals. (3) The remaining cluster die quality is determined based on the center of gravity distance between clusters.

The tested dies are divided into three classes. The standard class dies are those for which the divergence of the two predictors is the smallest and closest to the ideal value. Both the edge class and the outlier dies are judged by the two predictors to be a gap away from the ideal value. The die

quality of the edge class is aggregated, and judging the die labels by class reduces the computation time.

(1) Dividing Clusters: Ideally, the predictor divides the die quality into two clusters (pass and fail). However, in practice, the die is classified into multiple clusters due to algorithm limitations and the diversity of failure dies. Therefore, the clusters must be distinguished first, as shown in Fig. 5(a). A two-dimensional Cartesian coordinate system represents the prediction results of the two predictors in the previous two subsections. The test item predictor (P1) results are used as horizontal coordinates, and the spatial predictor (P2) results are used as vertical coordinates. The coordinate value indicates the probability that the die is good, the closer it is to 1, the higher the probability that the die will pass the test. The two predicted results (probability of a die passing CP) of the test set dies are fed into the DBSCAN algorithm, which outputs the cluster number of each die and determines whether it belongs to the outlier class. The dies classified by the clustering algorithm only know which class they belong to, but it cannot tell which clusters are passed or failed.

(2) Determine the Standard Class: The location of the die cluster distribution reflects the die quality. The vertices of axes 0 and 1 indicate whether the die passed the test or not, respectively. The standard class is determined by calculating the distance of each cluster from the ideal point (0 or 1). The location of the cluster center of gravity represents the cluster distribution characteristics, and the center of gravity (X, Y) location is calculated as follows:

$$\begin{cases} X = \frac{1}{n} \sum_{i=1}^{n} X_i \\ Y = \frac{1}{n} \sum_{i=1}^{n} Y_i \end{cases} \quad (4)$$

where X_i and Y_i are the horizontal and vertical coordinates of die i in the cluster, respectively, n denotes the number of dies in the cluster, and X and Y are the horizontal and vertical coordinates of the center of gravity of the cluster on the coordinates, respectively. The average of the two predicted values of the die in each cluster is calculated separately, and the result is the center of gravity coordinate of

the cluster. Ideally, the center of gravity of the pass cluster is near the (1,1) point, indicating that both predictors predict the probability of the die is passed to be close to 100%. On the contrary, the failure dies are close to the point (0,0). The distance between the center of gravity of all clusters and the ideal points (0,0) and (1,1) is calculated. The cluster closest to (0,0) is the failure standard class, and the cluster closest to (1,1) is the pass standard class.

(3) Determine Die Quality: The remaining clusters determine their quality based on the distance of the center of gravity from the center of gravity of the standard class respectively. The dies show diversity in their qualities due to different defects in the manufacturing process, often more than two clusters after clustering. In addition to the standard class, the remaining clusters are named edge classes, and dies that do not belong to any cluster are named outlier classes. For the edge class, the distance between the center of gravity of the edge class and the center of gravity of the standard class is calculated using Eq. (5).

$$D = \sqrt{\left(X_{edge} - X_{standard}\right)^2 + \left(Y_{edge} - Y_{standard}\right)^2} \quad (5)$$

where X_{edge} and Y_{edge} are the horizontal and vertical coordinates of the cluster center of gravity, $X_{standard}$ and $Y_{standard}$ are the horizontal and vertical coordinates of the center of gravity of the standard class, and D is the distance between the cluster and the standard class. As shown in Fig. 5(b), clusters near the standard passed cluster are considered passed, and those near the standard failure cluster are considered failed. The outliers are generally located between the pass clusters and the failed clusters. The distance between the outlier dies and the center of gravity of all clusters needs to be calculated and classified as the closest neighboring class.

3.3 Test Selection

Collaborative decision-making guides adaptive testing, as shown in Algorithm 1. Input the feature test term and spatial information BNR_i of Die_i. The predictor results P_1, P_2 are computed and Die_i is classified. If the Die_i belongs to the standard class, the test stops. Calculate the distance D_0, D_1 of the center of gravity from (0, 0), (1, 1), respectively. If it is close to (0, 0), the Die_i is good. If the Die_i belongs to the edge class, the test stops. Die_i quality is consistent with the closest standard class. When Die_i belongs to the outlier class, retest with the standard test set if the user chooses a low risk test. If the predictor automatically classifies, the die is consistent with the closest class quality. Algorithm 1 outputs die quality Q_i and Stop test flag.
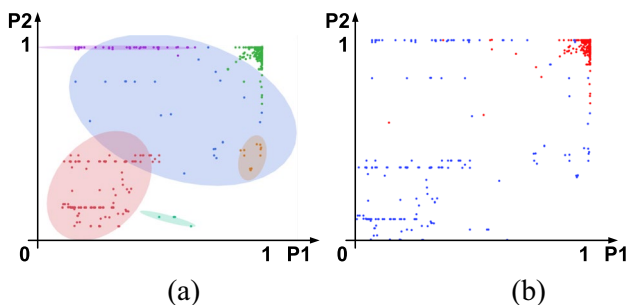
Algorithm 1. Adaptive Test Selection Process

Fig. 5 Distribution of model prediction results (**a**) Dies clustering results (**b**) Dies classification results

Input : Die_i feature test item results, BNR_i.

Output : Die_i quality，Stop test flag

1.   Calculate P_1, P_2 and classify Die_i.

2.   If Die_i is in the standard class, stop test.

3.   If $D_0 > D_1$, Q_i = good.

4.   Else, Q_i = bad.

5.   End if

6.   Else if Die_i is in the edge class, stop test.

7. $Q_i = Q_{closest\ standard\ class}$.

8.   Else if Die_i is in the outline class,

9.   If user selects the low-risk test, retest.

10.   Else, $Q_i = Q_{closest\ class}$, stop test.

11.   End if

12.   End if

13.   Return Q_i, $Flag_{stop}$

4 Experiment and Result Analysis

In this section, the effectiveness of the proposed method is experimentally verified using multiple lots of products. All case studies are implemented with actual industrial production test data from the CP stage. All experiments are performed on a computer with Intel i7-7700 CPU and 16GB memory. In the experiment, the algorithm used is implemented in Python 3.7 to simulate the actual test. We believe that the standard test set can detect all faults and a die that passes all test items is considered good. Five wafers from different products and one lot level wafers are selected, and the number of dies is 53,836 and 684,755, respectively. Sample wafer yields include multiple quality levels of 55%, 74% and 98%, etc. The data were collected from an actual semiconductor test plant. For confidentiality reasons, only some cases can be shown without sensitive information. Circuit types under test include battery charger chips and other products. The test items include open and short circuit test, function test, frequency test, current test under special working conditions, etc. The number of product test items includes 73 items, 45 items, 94 items and so on.

4.1 Single Predictor Performance

The quality predictor is the basis of the proposed method. In the experiment, 20 percent of the dies in the wafer are randomly selected as the training set, which is tested with standard test set. The residual part is used as the test set to verify the effectiveness of method.

(1) Performance of Test Items Predictor: The test item response is the most direct information on die quality. The training set data is used as the input to the RFECV algorithm for sorting and filtering. As shown in Fig. 6, taking wafer 1 as an example, based on the Recursive Feature Elimination (RFE) output, the test items are selected in turn for training predictors. It is not the case that the more test items are trained, the higher accuracy is obtained. When a high level of accuracy is reached, increasing the number of test items trained does not result in a significant increase in accuracy. After cross validation (CV), it is optimal to determine the number of test items when the accuracy is maximum. The feature test items output from RFECV are input to the random forest, and the trained predictor can achieve more than 95% accuracy.

(2) Performance of Spatial Graded Predictor: In order to prevent the predictor from being invalidated by abrupt quality change regions on the wafer, the BNR information is used to grade the dies and train a quality predictor based on the grading results separately. Considering the computational complexity and the reliability of the neighborhood information, a 3×3 window is chosen to calculate the BNR, using the percentage of bad dies in the surrounding eight dies to represent the BNR of the central dies. Based on Eq. (3), the dies are classified into three grades. The dies in the three grades are filtered with feature items using the RFECV algorithm and input into a random forest for training. Figure 7 shows the performance of the predictor on five wafers. The left vertical coordinate indicates the predictor accuracy trained for the dies in each BNR class, which can reach over 98.4% for each predictor. The right vertical coordinate and the line graph indicate the total accuracy of the predictor in different wafers, and the stability of the spatial
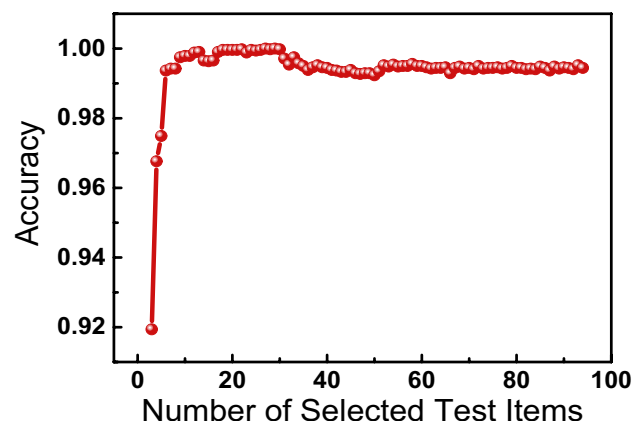
Fig. 6 Variation of predictor accuracy for training different number of test items
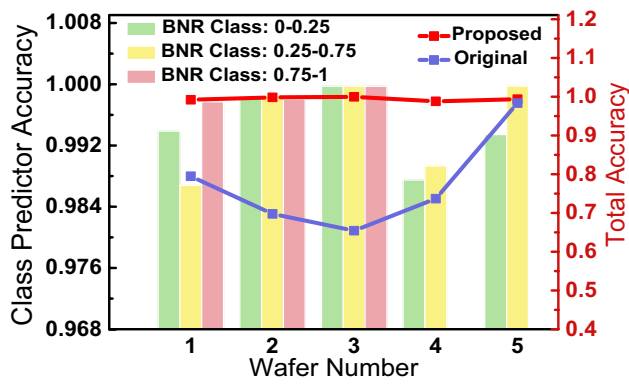
Fig. 7 Prediction performance of spatial graded predictors

Fig. 8 Accuracy of different quality dies in three classes

graded predictor is better than the conventional spatial predictor. The conventional spatial predictor only uses the two-dimensional coordinates of the die as features, and the wafer quality dramatically affects the accuracy. The accuracy of the conventional spatial predictor is above 77.34%, and the accuracy of the spatial graded predictor can reach 99.44%, an improvement of more than 22%.

4.2 Dual-Predictor Collaborative Decision

Each die is judged to pass or fail based on the quality classified by grade method as described in Section III. The standard class is closest to the ideal value, and the predictor should perform most accurately. Outlier classes are far from each cluster, and the predictor should be most prone to errors. Test escape rate and yield loss rate are used to measure the prediction quality, which are calculated as shown in Eqs. (6) and (7), respectively:

$$TER = N_{TE}/N_{Total} \tag{6}$$

$$FLR = N_{FL}/N_{Total} \tag{7}$$

where N_{Total} denotes the total number of dies in this class, N_{TE} denotes the number of dies that test escaped, and N_{FL} denotes the number of dies lost in yield.

Figure 8 illustrates the prediction accuracy of the pass and failed dies in different clusters. It can be seen that the standard class has the highest prediction accuracy, with 99.92% accuracy in predicting as the pass dies and complete correct prediction of failed dies. The failed dies in the edge class were still correctly predicted, but the pass die prediction accuracy decreased to 89.92%. The outlier class has the lowest accuracy, with a test escape rate of 17.97% and a yield loss rate of 8.66%. It is shown that the proposed method can distinguish the dies suitable for prediction and accurately predict the majority of die quality.
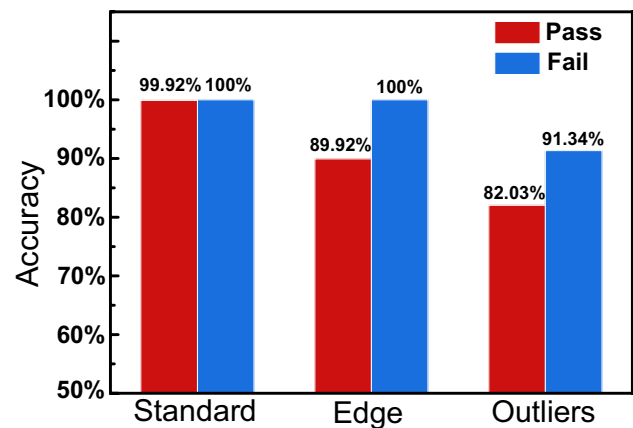
4.3 Experimental Comparison

Single wafer data and wafer lot data are used to demonstrate the effectiveness of the proposed method. This experiment is compared with the dynamic part average testing (DPAT) method proposed in 2018 [6] and the indirect test method proposed in 2021 [2]. References [6] and [2] are both effective ways to reduce test cost. DPAT adjusts the test content based on sampled die test data on each wafer [6]. By calculating the Cpk value for each test item, which reflects the validity of the test item, it is decided whether the die skips some of the test items. Indirect testing uses some of the test items to train the regression model, which predicts the remaining test items for the purpose of reducing the number of test items [2].

(1) Impact of Removing Test Items: In order to illustrate the effect of removing the test item effect, five wafers with different yields are selected for the experiment. Due to the small number of dies in a single wafer and the limited number of training sets, the ratio of the training set to the test set in a single wafer experiment is 1:4. Figure 9 shows the probability of die misclassification (test escapes + yield losses) for the three methods when using different percentages of test items in the standard test set. Experiments from five wafers with different yields show that the proposed method (outlier classes without standard tests) has a lower misclassification rate, in the case of training the same number of test items. The optimum point of test economics is to achieve a minimum test error state by testing the fewest test items. With increasing test items, the proposed method converges faster to a state with low test escapes and yield losses.

(2) Test Performance Comparison: Lot-level wafer test data is selected to verify the proposed method's effectiveness in high-volume industrial production. To comprehensively compare the performance of the methods in terms of test quality and test cost reduction, in addition to the number

Fig. 9 Misclassification comparison test results
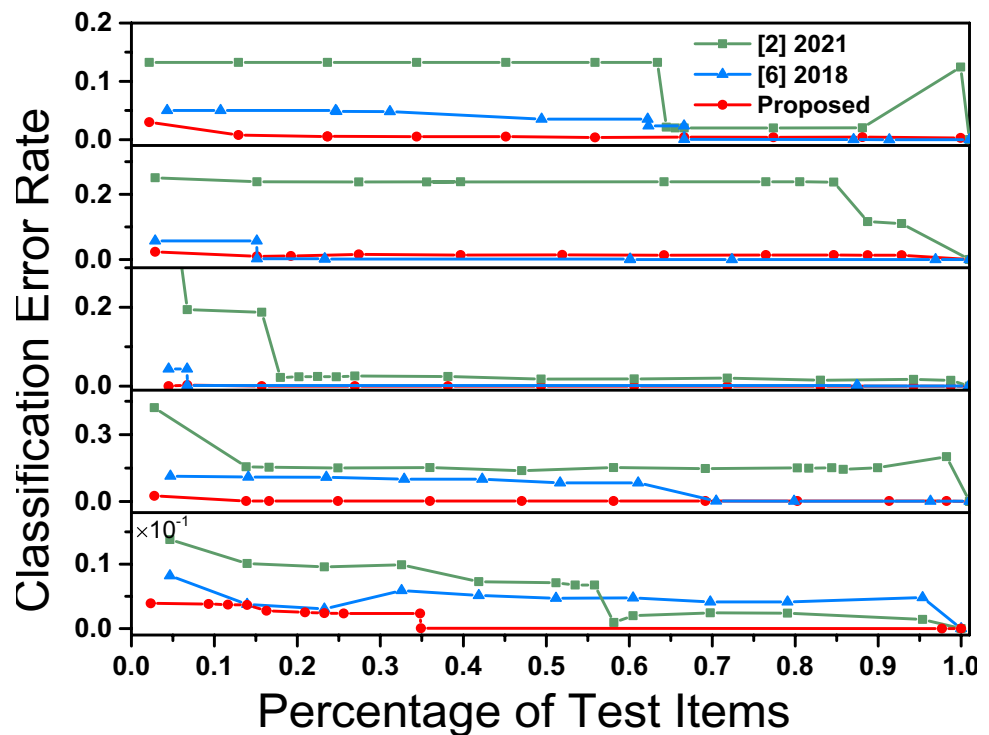
of test escapes (TE) and yield losses (YL), the following two metrics are used for evaluation. Test item reduction rate (TIRR): the rate of reduction in the number of actual test items compared to the standard test. The calculation formula is as follows:

$$TIRR = \frac{N_{actual} - N_{standard}}{N_{standard}} \times 100\% \qquad (8)$$

where N_{actual} denotes the number of test items actually used and $N_{standard}$ denotes the number of test items used for the standard test. Stability: the standard deviation of prediction error (test escapes + yield losses) is calculated to measure the stability of the effect of the method in different wafers. The smaller the value indicates a more stable method. The calculation formula is as follows:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})}{n}} \qquad (9)$$

where n denotes the number of wafers, x_i denotes the number of mispredicted dies in $wafer_i$, and $\bar{x}$ denotes the average number of mispredicted dies for all wafers. Running time: the computation time of each predictor on the same experimental platform, reflecting the resource overhead of the method.

Due to the large number of dies at the lot level, enough data can be used for training the predictor. The ratio of the training set to the test set is 4:1 in the lot-level wafer experiments. The number of test items used for training was determined using

the respective method algorithm, as shown in Table 3. The proposed method achieves a good trade-off between test cost and test quality. When using the same test items, the proposed method can reduce test escapes by more than 95% compared to predictors trained by common algorithms (Bayesian Regression, Logistic Regression and Random Forest). The stability of the performance on different wafers is also substantially improved. The common algorithm python modules are implemented in the scikit-learn package. The single predictor performance is also shown in the Table 3. It is worth noting that the spatial predictor is less stable and it is highly susceptible to the die quality distribution. The two predictors are processed with collaborative decision and show a great improvement in test quality and stability, indicating the effectiveness of the proposed method. The other methods show test item reduction

Table 3 Lot-level wafer comparison test results

Methods	TIRR	TE	YL	Stability	Running time(s)
Bayesian Regression	42.00%	3119	0	43.78	120.3
Logistic Regression	42.00%	1642	9	43.70	124.8
Test Item Predictor	42.00%	219	0	49.09	148.0
Spatial Predictor	42.00%	1188	78	214.05	106.9
[6] 2018	40.00%	125	0	4	164.9
[2] 2021	55.33%	649	2	17.86	141.3
Proposed	42.00%	7	2	1.72	276.4

rates that are similar to or slightly better than the proposed method, but the test quality is significantly degraded with a large number of test escapes. The proposed method also has the most stable performance on all wafers.

The outlier class of the proposed method contains only 27 dies, with 5 test escapes and 2 yield losses, accounting for 71.4% of the total test escapes and all yield losses, respectively. It shows that the proposed method effectively classifies the majority of dies, and the stability of test quality is improved. The proposed method has more running time than the other methods, but the calculation process can be completed offline without taking up the working time of the automatic test equipment (ATE).

5 Conclusion

In this paper, a dual-predictor collaborative decision adaptive testing method is proposed to reduce test cost and ensure test quality. The method uses test items in the CP stage and spatial information to train a test item predictor and a spatial graded predictor to predict die quality, respectively. The correlation information in the standard test data is more fully utilized. The DBSCAN algorithm clusters the results of the two predictors, and then the dies are classified by grade using the center of gravity position between clusters, which solves the problem that the clustering algorithm cannot distinguish between passed and failed dies. Experimental results using multiple lots of CP test data show that the proposed method can reach a better test economics point. Test escapes and yield losses can be reduced by more than 90% when the number of test items is reduced by 42%.

Data Availability Statement The datasets generated and analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Competing Interests 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.

References

1. Neethirajan D, Niranjan VA, Willis R, Nahar A, Webster D, Makris Y (2022) Machine learning-based overkill reduction through inter-test correlation. Proc IEEE VLSI Test Symp 1–7. https://doi.org/10.1109/vts52500.2021.9794170

2. El Badawi E, Azais F, Bernard S, Comte V, Kerzerho V, Lefevre F (2021) Evaluation of a two-tier adaptive indirect test flow for a front-end RF circuit. J Electron Test-Theory Appl 37(2):225–242. https://doi.org/10.1007/s10836-021-05934-4

3. Ahmadi A, Nahar A, Orr B, Pas M, Makris Y (2016) Wafer-level process variation-driven probe-test flow selection for test cost reduction in Analog/RF ICs. Proc IEEE VLSI Test Symp 1–6. https://doi.org/10.1109/VTS.2016.7477263

4. Yilmaz E, Ozev S, Butler KM (2013) Per-device adaptive test for Analog/RF circuits using entropy-based process monitoring. IEEE Trans Very Large Scale Integr (VLSI) Syst 21(6):1116–1128. https://doi.org/10.1109/tvlsi.2012.2205027

5. Stratigopoulos HG (2018) Machine learning applications in IC testing. European Test Workshop, Proc, pp 1–10. https://doi.org/10.1109/ETS.2018.8400701

6. Stratigopoulos HG, Streitwieser C (2018) Adaptive test with test escape estimation for mixed-signal ICs. IEEE Trans Comput-Aided Design Integr Circuits Syst 37(10):2125–2138. https://doi.org/10.1109/tcad.2017.2783302

7. El Badawi H, Azais F, Bernard S, Comte M, Kerzerho V, Lefevre F, Gorenflot I (2020) Implementing indirect test of RF circuits without compromising test quality: a practical case study. IEEE Latin-Am Test Symp LATS 1–6. https://doi.org/10.1109/LATS49555.2020.9093666

8. Yang CH, Yen CH, Wang TR, Chen CT, Chern M, Chen YY, Lee JN, Kao SY, Wu KC, Chao MCT (2021) Identifying good-dice-in-bad-neighborhoods using artificial neural networks. Proc IEEE VLSI Test Symp. https://doi.org/10.1109/vts50974.2021.9441055

9. Stratigopoulos HGD, Makris Y (2005) Non-linear decision boundaries for testing analog circuits, IEEE Trans. Comput-Aided Design Integr Circuits Sys 24(11):1760–1773. https://doi.org/10.1109/TCAD.2005.855835

10. Pan RJ, Zhang ZB, Li X, Chakrabarty X, Gu L (2021) Black-box test-cost reduction based on bayesian network models. IEEE Trans Comput-Aided Design Integr Circuits Syst 40(2):386–399. https://doi.org/10.1109/tcad.2020.2994257

11. Tsai TH, Lee YC, Hsieh CY (2019) Enhancing the data analysis in IC testing by machine learning techniques. Proc Tech Pap Int Microsystems Pack Assem Circuits Technol Conf IMPACT 183–186. https://doi.org/10.1109/IMPACT47228.2019.9024981

12. Liu MY, Chakrabarty K (2021) Adaptive methods for machine learning-based testing of integrated circuits and boards. Proc IEEE Int Test Conf (ITC) 153–162. https://doi.org/10.1109/itc50571.2021.00023

13. Huang K, Kupp N, Carulli JM, Makris Y (2013) On combining alternate test with spatial correlation modeling in Analog/RF ICs. Proc European Test Workshop 1–6. https://doi.org/10.1109/ETS.2013.6569358

14. Li KSM, Chen LLY, Cheng KCC, Liao PYY, Wang SJ, Huang AYA, Chou L, Tsai NCY, Lee CS (2022) TestDNA-E: Wafer defect signature for pattern recognition by ensemble learning. IEEE Trans Semicond Manuf 35(2):372–374. https://doi.org/10.1109/tsm.2022.3145855

15. Xanthopoulos C, Neckermann A, List P, Tschernay KP, Sarson P, Makris Y (2020) Automated die inking. IEEE Trans Device Mater Reliab 20(2):295–307. https://doi.org/10.1109/tdmr.2020.2994291

16. Shintani M, Mian RUH, Inoue M, Nakamura T, Kajiyama M, Eiki M (2021) Wafer-level variation modeling for multi-site RF IC testing via hierarchical gaussian process. Proc IEEE Int Test Conf (ITC) 103–112. https://doi.org/10.1109/itc50571.2021.00018

17. Wang R, Zhang LM, Chen N (2019) Spatial correlated data monitoring in semiconductor manufacturing using gaussian process model. IEEE Trans Semicond Manuf 32(1):104–111. https://doi.org/10.1109/tsm.2018.2883763

18. Zhang JL, You HL, Jia RX (2019) Reliability hazard characterization of wafer-level spatial metrology parameters based on LOF-KNN method. Proc Int Symp Phys Failure Anal Integr Circuits IPFA 1–4. https://doi.org/10.1109/IPFA47161.2019.8984814

19. Hsu CK, Lin F, Cheng KT, Zhang WY, Li X, Carulli JM, Butler KM (2013) Test data analytics - exploring spatial and test-item correlations in production test data. Proc IEEE Int Test Conf (ITC) 1–10. https://doi.org/10.1109/TEST.2013.6651900

20. Zhang JL, You HL, Jia RX, Wang XW (2022) The research on screening method to reduce chip test escapes by using multi-correlation analysis of parameters. IEEE Trans Semicond Manuf 35(2):266–271. https://doi.org/10.1109/tsm.2022.3144283

21. Herrera AEH, Stoyanov S, Bailey C, Walshaw C, Yin C (2019) Data analytics to reduce stop-on-fail test in electronics manufacturing. Open Computer Science 9. https://doi.org/10.1515/comp-2019-0014

22. Hinojosa A, Stoyanov S (2018) Data driven predictive model to compact a production stop-on-fail test set for an electronic device. Proc - Int Conf Comput Electron Commun Eng iCCECE 59–64. https://doi.org/10.1109/iCCECOME.2018.8658941

23. Hsu CK, Sarson P, Schatzberger G, Leisenberger F, Carulli J, Siddhartha S, Cheng KT (2016) Variation and failure characterization through pattern classification of test data from multiple test stages. Proc IEEE Int Test Conf (ITC) 1–10. https://doi.org/10.1109/TEST.2016.7805845

24. Zhang S, Li X, Blanton RD, da Silva JM, Carulli JM, Butler KM (2014) Bayesian model fusion: Enabling test cost reduction of analog/RF circuits via wafer-level spatial variation modeling. Proc IEEE Int Test Conf (ITC) 1–10. https://doi.org/10.1109/TEST.2014.7035328

25. Huang K, Kupp N, Xanthopoulos C, Carulli JM, Makris Y (2015) Low-cost Analog/RF IC testing through combined intra- and inter-die correlation models. IEEE Des Test 32(11):53–60. https://doi.org/10.1109/MDAT.2014.2361721

26. Gonçalves H, Li X, Correia M, Tavares V, Carulli J, Butler K (2015) A fast spatial variation modeling algorithm for efficient test cost reduction of analog/RF circuits. Proc Des Autom Test Eur DATE 1042–1047. https://doi.org/10.7873/DATE.2015.0690

27. Tello G, Al-Jarrah OY, Yoo PD, Al-Hammadi Y, Muhaidat S, Lee U (2018) Deep-structured machine learning model for the recognition of mixed-defect patterns in semiconductor fabrication processes. IEEE Trans Semicond Manuf 31(2):315–322. https://doi.org/10.1109/tsm.2018.2825482

28. Krishnan S, Kerkhoff HG (2013) Exploiting multiple mahalanobis distance metrics to screen outliers from analog product manufacturing test responses. IEEE Des Test 30(3):18–24. https://doi.org/10.1109/mdt.2012.2206552

29. Kang S, Cho S, An D, Rim J (2015) Using wafer map features to better predict die-level failures in final test. IEEE Trans Semicond Manuf 28(3):431–437. https://doi.org/10.1109/tsm.2015.2443864

Yuqi Pan received the B.S. degree in electronic science and technology from Guangdong University of Technology in 2018. He is currently pursuing Ph.D. degree in integrated circuits and systems with the School of Microelectronics, Hefei University of Technology. His research interests include manufacturing test, adaptive testing, machine learning, and data analytics.

Huaguo Liang received the Ph.D. degree in computer science from the University of Stuttgart, Germany, in 2003. From 1998 to 2003, he worked as a Research Fellow with the Department of Computer Science, University of Stuttgart. He is currently a full professor and a Ph.D. Supervisor with the School of both Computer and Information, and Microelectronics, HFUT, Hefei, China. His research interests include built-in-self-test, design automation of digital systems, ATPG algorithms, and distributed control. He served as the General Chair for the organizing committee of the IEEE Asian Test Symposium in 2018.

Junming Li received the B.S. degree in microelectronics science and engineering from Hefei University of Technology in 2022, where he is currently pursuing the M.S. degree in integrated circuits and systems with the School of Microelectronics. His current research interests include IC test and Design for testability.

Jinxing Qu received the B.S. degree in Electronic Information Science and Technology from Wenzhou University in 2021. He is currently pursuing the M.S degree in Electronic Science and Technology with the School of Microelectronics, Hefei University of Technology, in 2021. His current research interest includes Integrated Circuit adaptive testing.

Zhengfeng Huang received the Ph.D. degree in computer engineering from Hefei University of Technology in 2009. He is a full professor and a Ph.D. Supervisor with the School of Microelectronics, HFUT, Hefei, China. His current research interests include design for soft error tolerance/mitigation. He is a member of Technical Committee on Fault Tolerant Computing which belongs to China Computer Federation. He worked as a visiting scholar at the University of Paderborn, Germany from 2014 to 2015. He served on the organizing committee of the IEEE European Test Symposium in 2014. He served as a program cochair of Asian Test Symposium in 2018.

Maoxiang Yi (Associate Member, IEEE) received the B.S. degree in semiconductor devices and the M.S. degree in microelectronics from the Hefei University of Technology, Hefei, China, in 1986 and 1989, respectively, the Ph.D. degree in computer application technology from the Hefei University of Technology, Hefei, China, in 2010. From 2002 to 2003, he was a Visiting Scholar with the Institute of Physical Electronics, University of Stuttgart, Stuttgart, Germany. He is currently a full professor and a Master's Supervisor with the School of Microelectronics, HFUT, Hefei, China. His research interest includes very large-scale integrated circuit design for testability and reliability.

Yingchun Lu received the B.S. degree in microelectronics and the M.E. degree in microelectronics and solid state electronics from the Hefei University of Technology, Hefei, China, in 2002 and 2005, respectively, received the Ph.D. degree in integrated circuits and systems from the Hefei University of Technology in 2021, where he is currently an associate professor. His research interests include Hardware Security and Anti-radiation design of integrated circuit.