Adapting to Adaptive Testing

Erik Jan Marinissen¹ Adit Singh² Dan Glotter³ Marco Esposito³
John M. Carulli Jr.⁴ Amit Nahar⁴ Kenneth M. Butler⁴ Davide Appello⁵ Chris Portelli⁶

¹ IMEC vzw Kapeldreef 75 3001 Leuven Belgium
² Auburn University Electrical & Comp. Eng. Auburn, AL 36849 United States of America
³ OptimalTest 18 Einstein Street Nes Ziona 74140 Israel United States of America
⁴ Texas Instruments 13121 TI Blvd., MS366 Dallas, TX 75243 United States of America
⁵ ST Microelectronics srl Via C. Olivetti, 2 20041 Agrate Brianza Italy
⁶ ST Microelectronics 190 Av. Célestin Coq 13106, ROUSSET cedex France

erik.jan.marinissen@imec.be adsingh@auburn.edu dan.glotter@optimaltest.com marco.esposito@optimaltest.com

Abstract
Adaptive testing is a generic term for a number of techniques which aim at improving the test quality and/or reducing the test application costs. In adaptive tests, the test content or pass/fail limits are not fixed as in conventional tests, but dependent on other test results of the currently or previously tested chips. Part-average testing, outlier detection, and neighborhood screening are just a few examples of adaptive testing. With this Embedded Tutorial, we are offering an introduction to this topic, which is hot in the test community, to the wider DATE audience.

1 Adaptive Testing Basics
Adit Singh – Auburn University

1.1 Introduction
The wide variation in process parameters now being observed in advanced semiconductor processes requires that design specifications be tolerant of this manufacturing variability so as to achieve acceptable yield. This can sometimes cause the electrical impact of subtle manufacturing defects to remain within the acceptable range during testing, thereby masking defect detection. Such undetected manufacturing flaws can potentially cause functional failure for other input conditions not explicitly applied during the test; they can also behave as latent reliability defects that grow over time to cause early-life failure. Test methodologies that target low DPPM in complex devices and packages, particularly the zero-defect product quality requirements being pushed by the automotive industry, must address such test escapes by using innovative statistical test methods and adaptive test approaches because of the prohibitive costs of extensive functional and stress/burn-in testing of all parts. The basic idea here is similar to the strategy adopted by security screeners at airports to ensure ‘zero’ test escapes: parts that pass initial tests but appear to be suspect in some respect are separated out for more extensive testing. Also, based on the information available from the earlier test results, additional specialized tests can be adaptively added to the test flow to better target the suspected defect in a particular part. Non-suspect parts need not be so extensively tested, thereby minimizing the overall test costs.

1.2 Traditional vs. Adaptive Test Flow

Figure 1.1 shows a typical test flow for a high-end part. The flow includes wafer probe testing before the wafer is diced, tests before and after the bare die is packaged, post burn-in testing following the application of the reliability stress tests, and finally in-system testing. Traditionally, all parts are tested identically during each step (test insertion) in the test flow, using tests that are predetermined for that step. However, tests may be terminated early once a fault is discovered, so as to save test time using the so called ‘Stop-on-First-Fail’ strategy. Different tests are typically applied at the different test insertions. For example, in a digital IC, only slow scan-based stuck-at tests are often applied at wafer probe; at-speed functional timing tests are usually applied only after the die is packaged when more robust connections are available to the chip I/Os for accurate performance testing.

Figure 1.2 illustrates the information sharing required to support optimum adaptive testing across the test flow shown in Figure 1.1. Observe that the tests applied on each individual part at any test insertion here are now not fixed a priori, but are dynamically adapted for maximum cost-effectiveness. This adaptation can be based on the available test results for that particular part, as well as the test results for other related parts stored in the database. An important requirement for adaptive testing is an effective die-level traceability mechanism, such as a unique die ID for each part. This allows the device under test (DUT) during any test insertion to be uniquely identified, and permits access to earlier test results for that part, as well as test results for related parts as needed.
1.3 Test Adaptation Algorithms

Test adaptation algorithms aim to dynamically modulate the test intensity, test selection, and test limits used for testing each specific DUT at the different test insertions in the test flow. The goal is to optimize the overall cost versus defect detection probability, and the resulting DPPM levels. For example, it is now widely recognized that if all manufactured parts are identically tested, die from low yielding wafers, and low yielding regions within wafers, (i.e., die from bad neighborhoods), contribute disproportionately to test escapes and DPPM [1]. This suggests that such die be more extensively tested to minimize the test escapes [2]. An adaptive testing approach would increase test intensity, i.e. the test set size and coverage of the applied tests, for die from bad neighborhoods. Optimal test adaptation algorithms would match the test intensity applied to a DUT during any test insertion to the best yield estimate for the local region yield for that die based on the available test results for neighboring die from the wafer. Test intensity can also be adapted based on test results from the DUT itself. For example, the presence of defects in repairable memory arrays can indicate that the memory die may be from a wafer region with high defect density and should therefore be more extensively tested to ensure that all defects are reliably detected and repaired [3].

In general, test adaptation algorithms can exploit test results from the same step in the test flow, or from earlier test insertions. For example, industrial data also indicates that die from low yielding regions of the wafer display higher fail-out during burn-in tests [4]. Because wafer probe tests detect the vast majority of failing parts, fairly accurate yield data is available from this test step. This can be used to optimize stress testing and burn-in durations. In some cases, it may be possible to avoid burning in die from high-yielding populations, without a significant increase in the failure-in-time (FIT) rates in the shipped parts [5].

Adaptation can also be used to adjust the test selection based on the available test results. Recall that the testing of digital parts is an exponential problem: it is completely impractical to test all possible input conditions. Practical tests sets can activate and test only a very small fraction of the input space. Success of such a limited test strategy critically depends on including tests in the test set that target the likely faults. These can vary significantly during process excursions, and even from normal process variations. Thus if test results for the population of parts currently being tested displays a high incidence of a particular failure mode, say delay faults, then it can be cost effective to adaptively bias the test selection more towards delay tests, while deemphasizing tests for failure modes that are not being currently observed. In analog parts, test selection can be adaptively adjusted based on some initial tests applied to the DUT. For example, specifications of parts that exhibit electrical parameters close to the bounds of acceptable limits need to be extensively verified, whereas parts with more centered nominal parameters need mostly be tested for catastrophic defects. Furthermore, for analog measurements, available test results from other parts can also be used for dynamically setting the acceptable test limits [6]. In IDDQ testing for example, the nearest-neighbor residual has been shown to be an effective predictor of early life failure [7]. In dynamic parts average testing (PAT), acceptable limits on measured parameter values are statistically computed based on the mean and standard deviation of a moving window of most recently tested matched parts.

1.4 Statistical Post-Processing

In our discussion so far we have assumed that the tester makes a good/bad decision in real time based on the test response from the DUT; any part failing to meet the explicitly laid down test specifications is declared faulty. However, where the measured test parameters do not directly relate to functionality, but are only a (sometimes quite imprecise) indicator of the health of the chip, a final decision can be delayed to a post-processing step [8]. IDDQ tests applied to a digital part present a classic example, because IDDQ can show wide variability even in good circuits. If a DUT passes all the digital functional tests, but displays somewhat elevated IDDQ, it may be prudent to wait for IDDQ measurement results from other matched die before making a final good/bad decision. Statistical outlier detection approaches evaluate data from the DUT, in comparison to statistical data from other parts from the same manufacturing environment, before optimally making a final decision on shipping the die, also keeping in mind the DPPM and FIT rate requirements for the part. For example, if ensuring low DPPM is the highest priority, statistical outliers are more aggressively screened out, at the cost of some yield loss since not all the eliminated parts will fail in the field. Statistical post-processing algorithms attempt to optimize this tradeoff between yield and product quality.

1.5 Adaptive Test Support Tools

A practical challenge in experimenting with and implementing comprehensive adaptive test strategies has been the lack of easy access to test results across the different test insertions in a test flow. Traditionally, each of the test steps described in Figure 1.1 have been conducted completely independently, typically using different types of test equipment and test software, and often at physically distant locations. In such an environment, it is difficult to consolidate and share test results in a common easily accessible format; raw test response data for production parts can be overwhelming in volume and virtually impossible to interpret and process in the absence of a common format. Most critically, die IDs, which are essential to the tracing of individual parts through the test flow have only recently been widely adopted.

Over the past decade, several semiconductor companies have experimented with test adaptation using internally developed tools for at least parts of the test flow. More recently, a number of start-up companies have introduced software tools to support a common test results database, and also algorithms that can adaptively control the tests applied in any test environment. Typically these tools provide templates for this adaptive control, which can then be configured and customized for the specific target application.
The promise of adaptive testing is intriguing, but may be viewed as theory that is realizable exclusively for test time reduction (TTR). OptimalTest’s new generation of adaptive testing is not only a reality, but can deliver much more. In this paper, we will describe the features and benefits of OptimalTest’s adaptive test. In addition to reducing test time, it can significantly enhance quality, reliability, and yield learning. The key is to leverage the data: everything that is known about the device. Real time execution is its foundation. While the specifics of adaptive testing can be complex, the basics are to provide the customer a hassle-free solution with all the benefits. To that end, OptimalTest has leveraged the capabilities of today’s ATEs to provide an adaptive testing solution that is executed without touching the test program. The tests, the limits, the flows — everything is controlled and executed from the outside through OptimalTest’s Station Controller (OT-Box) or Proxy (OT-Proxy). Overall test times, and hence costs, are significantly reduced; the ROI is immense and the net effect is dramatic.

Test today should be viewed as leading-edge technology because of modern advances in software and networks. As test cell hardware has progressed, utilizing silicon integration and advanced packaging techniques, software applications and automation have also progressed — for example, the sophistication of 300mm wafer probe floors, impressive for their automation and integration. Similar kinds of automation and integration have emerged in the device test arena, bringing innovations to data mining and data integrity via new software tools, the latest in data base design and modern approaches to integration. OptimalTest (OT) has contributed to and leveraged these recent advances to provide comprehensive test management software with a suite of solutions that deliver significant improvement in test results and ROI from recovered yield, improved throughputs, reliability augmentation and top quality assurance with no compromise to test time reduction. These results are delivered via OT’s advanced adaptive test methodology which has been created by a team of test/software experts using innovative enabling technologies for superior real-time process control.

It is now feasible to reliably provide Data Feed Forward (DFF) and Data Feed Backward (DFB) across the enterprise to combine data from design models, litho/metrology, end of line test, e-test and previous test insertions with current real time results. Extracting the actionable data from each of these operations not only optimizes the test suite, but provides invaluable information for yield and reliability learning. It also accelerates feedback to design models and fab corrections. This is in fact the next generation of test process control: Intelligent Adaptive Testing. Test optimization is now done on a die-per-die basis using everything known about the device, neighborhood, wafer, lot, and batch. The benefits of a comprehensive adaptive testing solution go beyond optimizing test time to also deliver reliability augmentation and accelerated yield learning. With state of the art technologies and methodologies these objectives can be optimized in parallel. For instance using advanced adaptive testing (AAT) to change test limits will benefit quality assurance.

Let us start with considering the evolution of a typical test program. At the start of production there is an economic trade-off on the test content. Test, product, yield, and reliability engineers would prefer to have much greater test content to guarantee their objectives. However the reality of delivering a cost-competitive product means this is not possible. If we consider the initial production program to be the 100% test content, we have in fact a compromise where the engineers would like a 400% test content and the sales force would like a 50% test content for minimal cost. But from an ROI point of view it is agreed to use the 100% test content. Now over time, as production learning takes place, it is possible to perform test time reduction(TTR) where periodically the test content is reduced perhaps two or three times. While this is a benefit for the cost, this TTR impacts the yield base-line as these tests are lost for any future fab excursions or process shifts (see Figure 2.1). Let us consider how we can prevent impact to the yield base-line while reducing overall test costs.

![Figure 2.1: Traditional test time reduction.](image1)

![Figure 2.2: Reference dies/units.](image2)
dies creates and maintains a base-line that is critical for yield learning as well as providing feedback to update device models. At the same time manufacturing gets the productivity and cost advantages of performing test time reduction on the remaining die. Thus implementing intelligent adaptive testing with reference dies accelerates yield/reliability learning, optimizes throughput and maintains quality control of the test process.

Figure 2.3: Test time reduction with reference dies.

Now let us return to our example of starting production with a 100% test content program. This time with the reference dies being used, the yield base-line is not impacted as these die, which are the same on every wafer/lot for this product, are always tested first with the 100% program. Once again TTR can be implemented over time on the non-reference dies thus providing test cost savings with no impact to the yield base-line.

Figure 2.4: Quality and reliability with reference dies.

The reference dies can provide even greater benefits. The program that is run on these die can be expanded to include additional tests for quality, reliability, and yield learning. The test content of the reference dies can be expanded to two, three, or even four times the 100% program. Consequently those responsible for yield, quality, and reliability receive the data and results required for their objective. This is achieved using a relatively small number of die on each wafer. The non-reference dies can still see the original 100% program or, if appropriate, TTR can be implemented on the program to reduce cost. We have now met our primary goal of testing with high quality, improving yield and reliability learning, while reducing the test cost.

Figure 2.5: Combination of test time reduction along with quality and reliability accompanied by reference dies.

While these reference dies provide an excellent representation of the health of the wafer, validation wafers can be used to monitor the health of the lot. Validation wafers consist of the first two or three wafers of each lot along with the 10th or so follow on wafers. These wafers are first tested with the complete test program that shows whether this lot is a ‘healthy’ lot and becomes the decision point for activation of adaptive test on the remaining wafers in the lot. The decision is based on monitoring a combination of yield, hard/soft bins, and parametric results, while comparing these measurements to expected results. Reference dies and validation wafers work in tandem to assure a high level of process and quality control via this innovative advanced adaptive testing methodology.

We have observed new benchmarks of results in all key success criteria from processing millions of parts using advanced adaptive testing with reference die. From a yield learning point of view, the yield baseline is maintained for the life of the product. Additional testing is performed on the reference dies to accelerate yield learning. The enabling technology of DFF/DFB is leveraged to not only optimize test time, but to enable smart data-logging where we know what to log on which chips. This significantly improves the cycle time for feedback to design models and fab corrections.

On another front, we have also improved the reliability learning with additional testing on the reference die. In many market segments, reliability is a critical factor. Advanced adaptive testing can also help improve this metric. An example is using DFF, where data from e-test can be used to trigger additional voltage stress to improve the reliability. Data can also be used from the current test insertion to adapt test limits based on the distribution of good die. This action will reduce the test time impact of performing outlier detection. Another example would be performing an analysis of the test results of die in a portion of the wafer. This neighborhood analysis may lead to down-grading a good die that is in a region of bad die to augment the reliability of the lot.

The latest generation of adaptive testing coupled with an innovative application of reference dies leads to significant improvements in the quality of test while delivering improved test floor productivity with accelerated yield and reliability learning. The test program no longer has to be a less than perfect compromise. The test content can be adapted automatically in real time to optimize test costs without sacrificing critical engineering information for design models and fab improvements. These results have been confirmed in worldwide manufacturing environments.
3 Adaptive Testing Applications and Results

John M. Carulli Jr, Amit Nahar, Kenneth M. Butler – Texas Instruments
Davide Appello, Chris Portelli – ST Microelectronics

3.1 Introduction

Adaptive control systems are not new to the semiconductor industry. They have received much attention in wafer processing for over twenty years. However, more recently, adaptive control systems are getting increased attention in the area of test.

In general, adaptive control systems are an extension of statistical process control (SPC). SPC is used to monitor a statistical measure of control of a process operation and signal when control is lost. This signal is used to indicate a need for a process adjustment to regain control to specified levels. Adaptive systems such as advanced process control (APC), run-to-run control, and model predictive control (MPC) use the SPC data and signals along with models and feedback-feedforward loops to further improve process control [9]. The move to more adaptive systems has become increasingly necessary to tighten variation in deep sub-micron processes.

The impact of technology scaling on process variation and chip complexity is also being felt in test. The primary challenge in test has been to attain required high quality at a competitive cost. From a quality perspective, very low defectivity requirements are now becoming standard across a broad spectrum of applications. The requirements may be stated in terms of initial quality or test escapes, as well as initial reliability or latent defects. Table 3.1 shows reliability requirements for different product segments; this table derived from ITRS [10].

<table>
<thead>
<tr>
<th>Product Segment</th>
<th>Market Driver</th>
<th>Technology Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Advances</td>
<td>Reliability</td>
</tr>
<tr>
<td>PDA &amp; wireless</td>
<td>function 2×</td>
<td>50-200 dppm</td>
</tr>
<tr>
<td></td>
<td>every 2 yrs.</td>
<td></td>
</tr>
<tr>
<td>computer</td>
<td>speed 2×</td>
<td>50-200 dppm</td>
</tr>
<tr>
<td></td>
<td>every 2 yrs.</td>
<td></td>
</tr>
<tr>
<td>internet</td>
<td>bandwidth 4×</td>
<td>99.999% up-time</td>
</tr>
<tr>
<td></td>
<td>every 3-4 yrs.</td>
<td></td>
</tr>
<tr>
<td>automotive</td>
<td>function</td>
<td>zero</td>
</tr>
<tr>
<td></td>
<td>ruggedness</td>
<td></td>
</tr>
</tbody>
</table>

Table 3.1: Product segments reliability requirement ranges.

The traditional application of static test content and limits has become less effective at screening defectivity and more costly in the presence of increased variation. Hence alternate methods to improve test screening efficiency have been pursued. These methods have been broadly categorized as ‘adaptive testing’.

3.2 Adaptive Test Definition

“Adaptive Test is a broad term used to describe methods that change test conditions, test flow, test content and test limits based on manufacturing test data and statistical data analysis” [11]. The benefits are observed in many areas such as lower test costs, better quality and reliability, higher yields, and rapid yield learning. Many of the adaptive test applications to date have been focused on outlier screening for reliability improvement and reduced test cost.

As an example, in Figure 3.1 below, adaptive test limits are used in the post-processing of wafer test data. Only outlier material at risk for lower quality may undergo disposition for burn-in (BI) stress screening. Failures from burn-in stress can be evaluated for outlier parameter screen effectiveness and for shifts in signatures. Outlier classification results can be fed back or fed forward as needed to meet customer requirements or to drive continuous improvement activities for design, process, test, or system.

![Figure 3.1: Adaptive test flow example.](image)

3.3 Example and Results

An example of the above adaptive test flow is discussed next in the context of a 90nm SOC from Texas Instruments. The primary goal was to identify outlier screens at wafer probe which could predict later fails from burn-in stressing or from the customer. The results also had to minimize the use of burn-in test and the associated costs.

Burn-in testing was performed on this device as part of product characterization and bring-up. A number of post burn-in fails were obtained over a large sample of initial production material. The initial fail fraction along with the product EFR model suggested burn-in was required to meet the quality expectations with a static test program and operations flow.

Wafers containing the burn-in fails were identified for statistical outlier screening assessment. This was enabled by electrical die traceability. The Location Averaging algorithm and a robust linear regression model were used to indentify outliers on all of the parameters collected in the test program [12]. The Location Averaging template adapts to the spatial variation on each wafer and for each measurement parameter as shown in the example in Figure 3.2. The limits are set as a confidence level (CL) which will adapt to the variance of the residual of each wafer as shown in Figure 3.3.

![Figure 3.2: Example Location Averaging template of die ranks.](image)

The same wafer and parameter data is used for Figures 3.3 and 3.4. In the residual data plot, the burn-in fail is identified as an outlier beyond the adaptive 99% CL limits. In the raw data plot, the burn-in fail is in the middle of the distribution and cannot be screened with static limits.

After identifying outliers for every parameter, Chi-square and p-value statistics were calculated for every parameter to evaluate the statistical correlation between the outlier-classified population and the burn-in fail population [12]. An iterative approach was used to select statistically significant and efficient outlier screens. The result retained 15 parameters out of approximately 400 parameters. This process can be repeated to monitor changes in fail signatures from burn-in and adapt the list of optimal screening parameters.
In this case, by screening 55.3% of burn-in fails at probe, the customer DPPM requirements for this device were met at a level of 2.5% outliers. These outlier screens were deployed in production. The burn-in stressing and associated post-BI test were eliminated on the non-outlier screens. These outlier screens were deployed in production. The burn-in stressing and associated post-BI test were eliminated on the non-outlier screens. The burn-in stressing and associated post-BI test were eliminated on the non-outlier screens.

Figure 3.3: Adaptive limits at 99% CL. The BI fail is an outlier.

Figure 3.4: Static limit in raw distribution cannot screen BI fail.

In a case from ST Microelectronics, depicted in Figure 3.5, embedded traceability was exploited. The test data collected at wafer sort, at package test, and burn-in was stored in a single repository. During wafer sort tests, traceability data was stored on-chip. At package test, the traceability data was retrieved from the chip and saved in test data-logs along with test results.

Post-processing allowed for the reconstruction of a virtual wafer. Spatial analysis normally done on bin maps generated at wafer sort was now also available at package test. Multiple optimizations were possible thanks to the views extracted by this data, including adaptation of test limits at wafer sort, feedback to process, and the study of reliability behavior.

3.4 Conclusion

The case shown is just a small sample of adaptive testing. It is clear that leveraging more data in a statistical manner with adaptive feedback and feed-forward control can deliver improved quality and cost. There are many other creative areas where adaptive testing is applied – adapting test searches based on process information [13], adapting test content or test flows based on the wafer fail signature [14], adapting test data collection for diagnosis based on the yield signature [15], and the list goes on. Much of this work is already done in static studies. The challenge moving forward is to allow for improved infrastructure to enable adaptive test in an automated production environment.

References


