Controlling Test Plans by Information-Content-Based Redundancy Analysis

Ivica Rogina[1], Hans Martin von Staudt[2], Gunther Karner[1]
[1] optimiSE GmbH, Karlsruhe, Germany
[2] Dialog Semiconductor, Kirchheim/Teck, Germany

Abstract

There is an increasing need to keep the cost of testing ICs under control especially as functional complexity rapidly grows. Testing of complex mixed signal devices creates a huge amount of test data when complete test logs are stored – information which is rarely exploited to identify and remove redundant tests. This paper describes an information-theoretical approach to determine test-plan redundancy to reduce time and cost, and then discusses a high-volume real-world application: power management and audio chips as used in today's mobile phones. The paper will illustrate a case which showed a redundancy potential of 30% using the method, which later translated to a 50% test time reduction without noticeably increasing the risk for undetected defects. A practical procedure for introducing and qualifying a "reduced" test plan is presented.

1 Introduction

The testing of ICs is a significant contributor to the overall cost in today's semiconductor manufacturing, making test cost reduction a key challenge [1]. Beyond the classic approaches of using faster hardware, conducting parallel testing or just letting the test engineer further tweak a program, techniques are required that enable reduction in the test plan.

Some approaches already pursued aim at rearranging the order of the tests to make the detection of faulty devices happen earlier in the test program (e.g. [2,3]). Such methods are limited for high yield products because only the time for testing faulty devices can be reduced. If parallel testing is applied even faulty devices will not save test time. Other efforts which try to reduce test time by omitting tests (e.g. [4,5]) are usually motivated by the analysis of the tests' error behavior. Various methods for selecting a subset of tests from an original test-plan are mainly driven by the exploitation of measurements from faulty devices or failing tests. Some approaches (e.g. [6]) are targeting the selection of a subset by trying to maintain a large measurement space, i.e. by omitting those tests that do not significantly contribute to the definition and size of the measurement space. A method that cuts down test time by only sampling a given set of tests depending on the statistical behavior is commercially available [7]. This approach focuses on dynamic C_p, C_{pk} judgments during test.

We perform analyses on the abundance of data from parametric tests of good devices and rarely or even never on fail data. As illustrated in figure 1 the devices under test are seen as a black-box that provides responses (test results) to a given stimulus (test). Redundant tests are identified by calculating the information content of their test responses in relation to the rest of the test-plan.

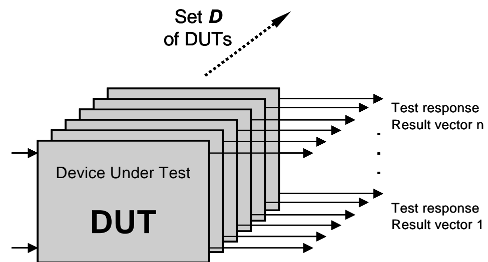

Figure 1: Black-box test response

Generally, redundancy in test-plans means that test t can be considered for omission because there exists a set of tests $S(t)=\{t_1,t_2,...\}$ such that whenever t reports a failure, at least one test in $S(t)$ reports a failure, too. For high-yield products, the amount of data from faulty devices is very low, and the analysis must focus on the data collected from the good parts: if $t(d)$ denotes the measurement of test t on device d, we can say that t has zero information, if the vector $t(d)$ is a linear combination of other vectors $\{t_1(d),t_2(d),...\}$. In reality, such ideal situations rarely occur with parametric tests. So we look at additional information to decide on the ability to omit t by incorporating knowledge about quality control measures, correlations, and sophisticated information content measures. An analysis declaring a test to (virtually) never find faulty chips but to be redundant can be used to review the test-plan and suggest reductions.

To keep the risk for undetected faulty parts virtually zero, we estimate the likelihood for failures in an omitted test (not found by other tests) by approximated integration of the tests' common probability density function. In addition, results of

omitted tests are estimated (reconstructed) from tests not omitted.

Despite tests being identified as redundant, they cannot be omitted unreflected because it is still perceived as quality risk by customers. Redundancy found by the analysis must have a technical reason to ensure correct classification of a test. In addition, methods of handling two test sets (standard = reduced, and extended = original) address quality concerns and form a vital part of our approach.

2 Information content

We define the normalized score $t'(d)$ of a test result $t(d)$ for a test t with limits $L(t)$and $H(t)$and a target value of $\mu(t)$ as:

$$t'(d) = \begin{cases} 1 & \text{for} \quad t(d) < L(t) \\ \left(\frac{\mu(t)-t(d)}{\mu(t)-L(t)}\right)^2 & \text{for} \quad L(t) \le t(d) \le \mu(t) \\ \left(\frac{\mu(t)-t(d)}{\mu(t)-H(t)}\right)^2 & \text{for} \quad \mu(t) \le t(d) \le H(t) \\ 1 & \text{for} \quad t(d) > H(t) \end{cases}$$

This is illustrated by the graph of figure 2

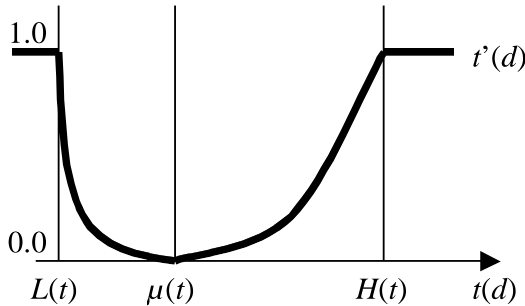

Figure 2: Result normalization

It is obvious that $t'(d)$ is 0 when the actual result $t(d)$ equals the target value $\mu(t)$ and $t'(d)$ is 1 for all result on or outside the test's limits. The square exponent in the above equation is used to emphasize results that are closer to the limits.

We now define a test t_1 to be more critical on device d than a test t_2 if $t_1'(d) > t_2'(d)$. If

$$\forall y \forall d \in D: t_x'(d) \le t_y'(d)$$

is true, then this means that test t_x is of no practical use on the device set D, because for every single measurement of t_x, there is some other measurement t_y which is more critical. This holds especially for tests which always produce target value results. Consequently, a test that always produces near-target results is a less critical test than one with results further off the target value.

Our approach assigns an information gain of 0.0 to a test t_0 which always produces target value measurements (i.e. $\forall d$: $t_0'(d)=0$). A test t_1 which produces only target values except for one device d, for which $t_1'(d)$ is r_1 ($0 \le r_1 \le 1$), has an information gain of r_1 over test t_0. Another test t_2, which is just like test t_1, except that $t_2'(d)$ is r_2, ($r_1 \le r_2 \le 1$), has an information gain of r_2-r_1 over test t_1. Generally, the information gain on a set D of devices under test (DUTs) of a test t_x producing the results $t_x(1),t_x(2),\ldots t_x(|D|)$ over a test t_y producing the results $t_y(1),t_y(2),\ldots t_y(|D|)$ is defined as:

$$\Delta I(t_x, t_y) = \sum_{d \in D} \min(t_x'(d) - t_y'(d), 0)$$

The information gain of a test t_x over a complete set of tests $T = \{t_1,t_2,..t_K\}$ producing the results $\{t_1(1),t_1(2), \ldots t_1(|D|), t_2(1), \ldots t_2(|D|), \ldots t_k(1), \ldots t_K(|D|)\}$ is correspondingly defined as:

$$\Delta I(t_x, \{t_1, t_2, \ldots t_K\}) = \sum_{d \in D} \min(t_x'(d) - \max_{k=1}^{K} t_k'(d)), 0)$$

Ranking the tests in the algorithm works as follows:

1. choose the test with the smallest individual information content as the first one
2. when K tests $t_1,t_2,..t_K$ have already been ranked, choose the next ranking test t_z such that

$$z = \mathrm{argmin}_{x \in \{1..k\}} \Delta I(t_x, \{t_1, t_2, \ldots t_k\})$$

3. continue with step 2 until all tests have been ranked

This approach ensures that the k-ranked test is the one with the smallest information gain over the set of tests ranked before k.

In an optimal test scheme, every test would produce approximately the same information gain. Note, that if two tests are absolutely identical, then one of them can contribute some information gain, the other one's gain will be exactly zero. Which one of two identical tests is chosen first (higher rank and some gain) and which one is chosen second (next rank and zero gain) is rather arbitrary.

3 Redundancy

We define a test to be redundant, when its added information content is less than a given threshold. Some additional properties of the test – like its C_{pk} or correlation values etc. – can be considered too in a rule-based mechanism to make the final decision whether the test is to be considered redundant or not. Therefore we use a linear combination

(knowledge function) of the information content and several other test parameters – depending on the analyzed product and production stage – to alter the test ranking in such a way that product-specific concerns can be incorporated. Note that this notion of redundancy is purely information-theoretic. Just like high correlations do not define "dependencies", redundancy does not define "omittability". Therefore a test engineer with profound knowledge of the actual test plan must eventually approve the suggested redundant tests for omission.

Once, a set T of tests has been analyzed, and a subset $R \subseteq T$ has been identified as redundant, two test-plans can be defined, a "regular" plan containing only "mandatory" tests $M = T \setminus R$, and an "extended" plan T. The extended plan can be used for problematic devices or device sets (e.g. identified by poor results from tests out of M, or – for wafer sort tests – by analysis of PCM-data [8]). The regular plan M, can be used in all other situations, saving the test time and cost for R.

What remains to be checked is the risk for undetected failures when running the test-plan M.

4 Risk management

Even when we have really huge amounts of test data, that go into millions of tested devices, and even if we have never observed a failure of a given test on all of these devices, it is usually not correct to say that the likelihood of this test producing an error is exactly 0.0. In most cases the test engineer in charge of a test program will not agree to omit a test that has already produced failures before (unless there is evidence that the failure would be caught by some other test). Without any history of errors, for those tests that are classified as redundant, it is difficult to give a meaningful estimate of the escape likelihood, i.e. of the likelihood that this test would detect a failure some day which will not be caught by any other test and which will lead to an undetected defective device.

As of today, there is no known analytical method to calculate this "escape risk" exactly. There are only numerical and approximate methods. Monte-Carlo methods that use a random sample generator are commonly used in some areas like medicine or economic sciences to estimate the integral below an also estimated parametric probability distribution. Such methods work well, when the domain of the samples is rather low dimensional (one or two). In our case the dimension of the samples domain has the dimension $|T|$, the number of analysed tests – usually several hundreds. Therefore, we use another analytical approximation. Consider figure 3.

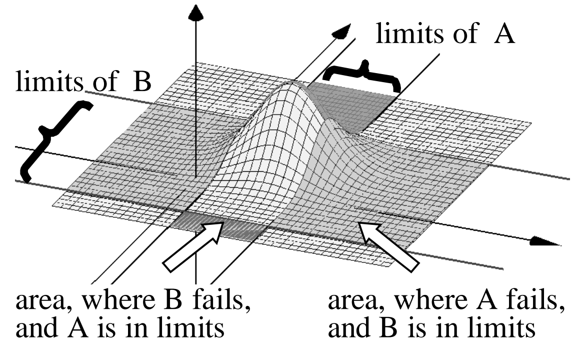

Figure 3: Two-test tesult-space

The graph shows the simultaneous result distribution of two tests. The lines parallel to the y-axis denote the limits of test A and the lines parallel to the x-axis denote the limits of test B. The shaded areas are those where test A is in limits while test B is out of limits. Clearly, the risk for undetected failures (escapes) that would occur if test A was omitted, is defined as the integral below the shaded area (provided that the test program consists only of tests A and B).

We have implemented a numerical estimation of this risk as follows. Let $p(a,b)$ denote the likelihood that tests A and B produce the results a and b on the same device. Let $L(A)$, and $L(B)$ be the tests' low limits and $H(A)$, $H(B)$ the corresponding high limits, then the actual risk for an undetected failure when omitting test A is:

$$\int_{-\infty}^{L(A)} \int_{L(B)}^{H(B)} p(a,b) \mathrm{d}b \mathrm{d}a + \int_{H(A)}^{+\infty} \int_{L(B)}^{H(B)} p(a,b) \mathrm{d}b \mathrm{d}a$$

and the risk for an undetected failure when omitting test B is:

$$\int_{L(A)}^{H(A)} \int_{-\infty}^{H(B)} p(a,b) \mathrm{d}b \mathrm{d}a + \int_{L(A)}^{H(A)} \int_{H(B)}^{+\infty} p(a,b) \mathrm{d}b \mathrm{d}a$$

These integrals cannot be computed analytically. Therefore, we approximate them by numerical methods.

This risk estimation procedure is purely static. It is obvious that the likelihood of a test failing increases with the size of the tested device set. In our experiments we have found that, although huge sets of DUTs produce sound stochastic figures, these results are becoming less meaningful. That is because several parameters that influence the test's behavior (like the overall quality of the wafers, the tidiness of the clean-rooms, etc.) can drift over time and therefore can make a test that has not failed on millions of parts suddenly become critical.

To cover such risks, we have enhanced our method by dynamic control mechanisms which track the tests' behavior over time and detect slow drifts or sudden changes in a set of several dozens of parameters like the tests' C_{pk} values, mean values,

standard deviations, etc. A tester controller can then be operated to switch from the regular test-plan (reduced subset) to the extended test-plan (full set of tests) for a period of time – at least until the parameters have returned back to normal or until new "normal" values for these parameters have been learnt [9].

5 Real world evaluation

A real-world evaluation of the described method taking a suitable evaluation vehicle has to prove whether the method described achieves any significant benefit.

5.1 Evaluation vehicle

Power management and audio chips as found in mobile phones increasingly integrate most of the analog base band functionality. This usually comprises a charger, various linear and switching regulators, the audio functions like voice and HiFi CODECs and microphone and speaker amplifiers. These chips can be seen as a conglomeration of mixed-signal functionality that for various reasons is not integrated into the digital base band processor chip.

The test plan for such devices is made up of around 500 parametric tests; more highly integrated devices, which are in development, target an even higher number. Hence such devices constitute a perfect vehicle for a method claiming to calculate information content and to identify redundancy when analyzing the data of parametric tests.

The vehicle for proving the method's claim is a well established and mature chip which, although not the latest generation, is still shipping in significant volumes. The original test plan was made up of 358 parametric tests.

5.2 Result of redundancy analysis

As highlighted in the abstract, storing each and every test result creates a huge amount of data. The device was tested on a Teradyne Catalyst such that the format of choice was STDF (Standard Test Data Format, [10]). A single batch consisting of 70,000 devices creates an STDF file of 3 GByte size.

Two studies were carried out. An initial analysis took 13,000 devices into account. In a second step the analysis algorithm was fed with test data of 180,000 devices. Not surprisingly the identified redundancy diminished the more devices were included to the analysis as, for instance, the number of defects increases in general..

Algorithm results of test data	
	Redundancy Potential
13k study	27%
180k study	17%

5.3 Process of test classification

The transition from mathematical and statistical results to engineering decisions is most critical. Besides technical considerations, "psychological" aspects - like the need for safety – must also be addressed. As the device in question is an ASIC product the test specification is negotiated between test engineering and customer.

In short the process can be described as follows:

1. The method points to test(s) in a test group to be redundant.
2. Analyze whole group of tests
    a. Is test defect oriented? If yes: keep.
    b. Identify root cause of redundancy
    c. Choose most appropriate test(s) to be omitted based on engineering insight on topology and test program flow.

Step 2a illustrates the first limitation to the redundancy removal process. A typical test program contains a number of tests that although being parametric are not testing against parametric fails but are checking for manufacturing defects. A typical example is a leakage test which normally gives measurement values close to the target value, very wide of the limits. Even if the analyzed quantity does not show a leakage fail for a specific pin it cannot be omitted. Sometimes a simple functional test can replace the parametric test that the analysis claimed to be redundant.

Step 2b is the decisive step, establishing the link between the "black box" findings of the information theoretical approach and the actual implementation of the functionality on the chip. If this link can be established, the case for omitting tests becomes easier.

Step 2c enables engineers to omit even tests that the described method had declared mandatory. The reason being that the mathematical approach makes arbitrary assumptions about which of two nearly identical tests to choose. Another practical limitation that 2c overcomes are the fails that are not silicon, but test machine induced. Tests producing fails will always be classified as mandatory by the algorithm as the information content of fails is very high. As long as analysis is done on stop-on-fail data there is no mathematical way to prove a failing test redundant. This can only be achieved with the test and design engineers' knowledge.

Results of test classifications	
Total number of tests	358
Tests classified redundant	158
Tests classified mandatory	200
Redundancy test time related	31 %
Test time reduction incl. manual improvements	50 %

5.4 Qualification and introduction of reduced test plan

Even with the best-proven mathematics and a clear reasoning about topology given redundancy, a customer still will perceive omittance of a test to be a risk. Further effort has to be invested before a reduced test plan is accepted and released for mass production.

The first step is to qualify the reduced (now called "regular") test plan against the original test plan (now called "extended"). In our example more than 100,000 devices out of three different lots were first tested with the reduced test program. All pass devices were then tested again with the extended test program. This flow is shown in figure 4. The expectation was not to find any fails. Such a quantity is sufficient to push any remaining potential quality risk below 10 ppm. If lower quality margins are required, larger quantities need to be used for qualification.

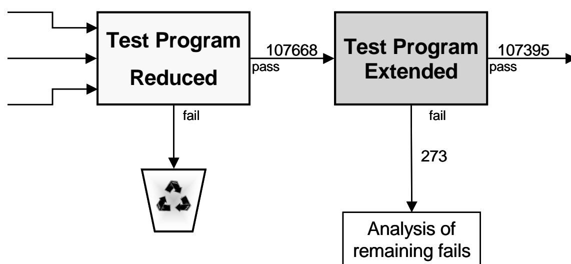

Figure 4: Qualification experiment flow

In fact 273 fails were found that had to be analyzed thoroughly. Analysis revealed the following results:

Results of qualification experiment	
No of Devices	Reason for fail
197	Marginal fails, no further action required
14	Handling errors
4	2 test program weaknesses
58	2 classifications wrong

The vast majority of fails are marginal devices as test limits of both versions were identical. Of most interest are the two wrong classifications. In one case a defect mechanism was found that make the corresponding test to be defect oriented and hence mandatory. The other classification was a simple misjudgement that needed correction. This finding proves that such a qualification process is required before releasing a reduced test plan.

The second step to secure acceptance of the method is to handle both test plans - the regular and the extended one - in one program. This enables two scenarios:

1)
The extended test plan is carried out every 50 or 100 devices. This does not significantly reduce throughput but still produces statistical data on the "removed" tests. Any abnormal behavior over time (drift of mean, increase of sigma) can be monitored and analyzed even though the test is not carried out for most devices.

2)
The extended test plan is used for any other purpose than volume production. In particular, applying the extended test plan to conformance tests (sample probe) makes sure that constant monitoring of the omitted tests happens. Any deviation from the expected behavior immediately comes to the attention of the product engineer, thus triggering further analysis.

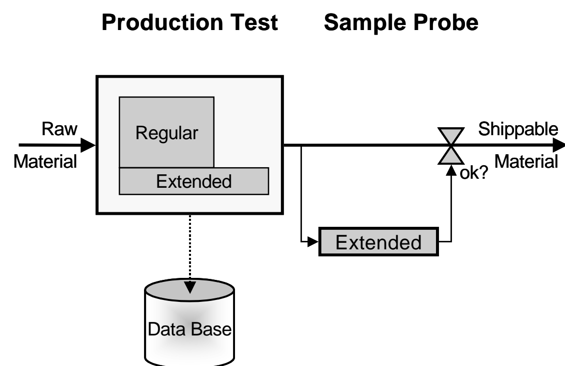

Figure 5: Production flow

The procedures as described above and as shown in figure 5 were established at Dialog Semiconductor together with the release of the reduced test program. Results so far are convincing:

Production results of reduced test program	
Total number tested devices according to the reduced test plan	2.8 M
Sample probe failing due to omitted test	0
Field returns due to omitted test	0

6 Conclusion and outlook

The presented method of applying the idea of information content to parametric test data was proven to be practical for the application of testing of mixed signal power management and audio chips. In the given example, it was possible to reveal and remove significant redundancy in the test plan thus harvesting a significant test cost saving.

Four key elements make the method applicable in an industrial environment.

1. An information theoretical approach, not imposing the prerequisite of Gaussian distributions
2. Providing a link between "black-box" findings and circuit topology
3. Qualification
4. Constant monitoring

The test plan that resulted from applying the method is still static as it is the outcome of a negotiation and a technical judgment between test engineer and customer. Further work has to be undertaken to establish an automated on-line surveillance for "omitted" tests. The process of dynamic reaction upon unexpected behavior of an omitted test is favorable compared to reacting on the failure of a sample probe.

References

[1] The ITRS: "International Technology Roadmap for Semiconductors 2003", online at http://public.itrs.net/

[2] L. Milor and A. L. Sangiovanni-Vincentelli, "Minimizing production test time to detect faults in analog circuits," *IEEE Trans. Computer-Aided Design,* vol. 13, pp. 796–813, June 1994.

[3] L. Milor and A. L. Sangiovanni-Vincentelli, "Optimal test set design for analog circuits," in *Proc. ICCAD,* 1990, pp. 294–297.

[4] G. N. Stenbakken and T. M. Souders, "Test-point selection and testability measures via QR factorization of linear models," *IEEE Trans. Instrum. Measur.,* vol. 36, pp. 406–410, June 1987.

[5] G. N. Stenbakken and T. M. Sourders, "Linear error modeling of analog and mixed-signal devices," in *Proc. Int. Test Conf.,* 1991, pp. 573–581.

[6] E. Felt and A. L. Sangiovanni-Vincentelli, "Testing of analog systems using behavioral models and optimal experimental design techniques," in *Proc. ICCAD,* 1994, pp. 672–678.

[7] J. Bibbee "Method, apparatus and product for evaluating test data", US Patent US6711514, 2004

[8] I. Rogina, "Exploiting Redundancy Information in Logs from IC and PCM Tests", 5th European AEC/APC Conference, Dresden April 14-16, 2004

[9] W.Gehring, I.Rogina, G.Karner, "Dynamic and Online Control of Flying Probe and Wafer Testers" Technical Report 2005/01, optimiSE, Karlsruhe, Germany, 2005

[10] "The Standard Test Data Format STDF", Teradyne, Inc.. 321 Harrison Avenue, Boston, MA 2118-2238. USA