Parallel Test Reduces Cost of Test More Effectively Than Just a Cheap Tester

Jochen Rivoir
Agilent Technologies R&D and Marketing GmbH & Co. KG
Herrenberger Str. 130, 71034 Boeblingen, Germany
jochen_rivoir@agilent.com

Abstract

Today's manufacturers of high-volume consumer devices are under tremendous cost pressure and consequently under extreme pressure to reduce cost of test. Low-cost ATE has often been promoted as the obvious solution. Parallel test is another well-known approach, where multiple devices are tested in parallel (multi-site test) and/or multiple blocks within one device are tested in parallel (concurrent test).

This paper shows quantitatively that parallel test is a much more effective test cost reduction method than low-cost ATE, because it reduces all test cost contributors, not only capital cost of ATE. It also shows that the optimum number of sites is relatively insensitive to ATE capital cost, operating cost, yield, and various limiting factors, but the cost benefits diminish fast, if limited independent ATE resources reduce the degree of parallelism and force a partially sequential test.

Keywords: Cost of test, test economics, multi-site test, ATE, reduced pin-count testing, low-cost ATE, probe-card, I/O bandwidth matching

1. Introduction

Advances in semiconductor process technology and a highly parallel production process allow manufacturing more and faster transistors at lower cost [1] [2]. Thousand dice, each containing millions of transistors are manufactured together on one wafer.

Today's test, however, is to a large degree sequential. Most commonly, multiple building blocks are tested sequentially, one die after the other. The consequence is that test does not scale as production does [1] [3]. DFT techniques, such as scan test and BIST bring some parallelism inside the chip, at least for the digital portion of the chip. So does concurrent test, where multiple analog and digital blocks are tested at the same time [4] [5].

While multi-site test, i.e. testing multiple dice at the same time, is predominant for memories test, it has only been partially adopted for SOC devices. Test cost reduction strategies based on multi-site test have been proposed with the goal to make test cost scale with technology progress [6] [7] [8] [9]. [10] [11] [12] describe the general test cost savings of multi-site test. Although the cost benefit of multi-site test is well known under ideal conditions, there is a common misconception that several negative effects neutralize the benefits fast. Instead, low-cost ATE is often the perceived solution to lower cost of test. The most-often stated arguments against multi-site test are:

- "The ATE will be more expensive."
- "The probe-cards will be more expensive."
- "Test time will be longer for bad dice, because test cannot stop on first fail anymore."
- "Lack of enough independent ATE resources for all sites will imply some sequential test and increases test time."
- "The change-over time between production lots will be added more often."
- "More contacting problems will cause more retest"
- "Under-utilization increases, if ATE is upgraded for multi-site test, but not all devices are tested in multi-site configuration."
- "Probe-cards will last longer than needed."

This paper describes an economic model that takes all of the above effects into account and quantifies their impact using a realistic example. This paper will show that parallel test, i.e. multi-site test and concurrent test, is a much more powerful test cost reduction method than the often-cited low-cost tester, even in the presence of the above-mentioned limiting factors. A sensitivity analysis identifies the parameters that are critical for a successful deployment of multi-site test and those that have only insignificant impact. Although the terminology and the numerical example targets parallel wafer test, the economic model can also be adapted to package test.

This paper is organized as follows: Section 2 gives a detailed description of a test cost model for parallel test. Section 3 states the numerical assumptions for an example that will be used for comparison later on. Section 4 quantifies the test cost benefits of multi-site test under nominal and under optimized conditions. The cost savings of concurrent test are quantified in section 5. In section 6 the combined effect is quantified and sensitivities to many parameters are shown. For comparison, the cost reduction with an absolutely free ATE is stated in section 7, before all above scenarios are compared in section 8. Conclusions are drawn in section 9.

2. Test cost model

This section describes a test cost model for parallel test that takes a multitude of effects into account, which might degrade its cost benefits. The expert reader can skip this section and just review the numerical assumptions in section 3 to find out which effects are modeled.

In this paper, the test cost per DUT c includes contributions from the depreciation of test cell capital cost c_{Cap}, equipment independent operating cost c_{Op}, purchase of probe cards c_{PC}, and the cost c_{Pkg} of packaging bad parts that test did not identify as being bad.

$$c = \underbrace{\left(c_{Cap} + c_{Op}\right)}_{c_{Testcell}} + c_{PC} + c_{Pkg} \qquad (1)$$

The capital related cost and the operating cost are combined to the test cell related cost

$$c_{Testcell} = c_{Cap} + c_{Op}, \quad (2)$$

which is usually the largest cost contributor. Costs for DFT, MTBF, and test development are not included here.

Throughout this paper c denotes test cost contributions per DUT, R is used for cost rates in \$ per time, C stands for (capital) cost in \$, p is used for percentages, k for unit-less correction factors, N for unit-less numbers, T for longer time durations, and finally t for times on the test cell.

2.1 Capital cost

The capital cost of the test cell C is depreciated over the depreciation period T_{Depr}, usually over 5 years. Maintenance, the number of shifts, reliability, calibration, and other factors reduce the usable time to a fraction p_{Util}, or, in total to $p_{Util} \cdot T_{Depr}$. A test that occupies the test cell for a time t_{Tot} carries a corresponding fraction of the overall capital cost. When, in average, $\overline{S}$ sites are tested in parallel, this fraction covers the cost for S devices under test.

$$c_{Cap} = \frac{C}{\overline{S}} \cdot \frac{t_{Tot}}{p_{Util} \cdot T_{Depr}} \quad (3)$$

With the capital related test cost rate for a single site

$$R_{Cap} = \frac{C}{p_{Util} \cdot T_{Depr}} \quad (4)$$

equation (3) can be written as:

$$c_{Cap} = R_{Cap} \cdot \frac{t_{Tot}}{\overline{S}} \quad (5)$$

For a 1M\$ test cell, a 65% utilization and a depreciation over 5 years, the test cost rate R_{Cap} is 0.98 ¢/sec. The test cell mainly consists of the prober and the ATE.

$$C_{Testcell} = C_{Prober} + C_{ATE} \quad (6)$$

To reflect the influence of device pin-count and the number of parallel test sites, the ATE capital cost is divided into the infrastructure cost for a 0-channel ATE $C_{ATE,0}$, the cost for all digital ATE channel resources, plus the cost for per-site resources, such as device power supplies and mixed-signal resources

$$C_{ATE} = C_{ATE,0} + \left(C_{1Ch} \cdot N_{Ch} + C_{1Site}\right) \cdot S, \quad (7)$$

where C_{1Ch} is the cost for one ATE channel, N_{Ch} is the number of device pins that require an ATE channel resource, C_{1Site} is the cost for the above-mentioned per-site resources for one site. S is the number of parallel test sites the ATE is equipped for.

Equation (3) suggests that multi-site effectively divides the capital cost C by the number of parallel tested sites $\overline{S}$. However, multi-site test also requires more tester resources, as shown in equation (7). Substituting equations (6) and (7) in (3) shows the complete picture

$$c_{Cap} = \left(\frac{C_{Infra}}{\overline{S}} + \frac{C_{1Ch} N_{Ch} + C_{1Site}}{S / \overline{S}} \right) \cdot \frac{t_{Tot}}{p_{Util} T_{Depr}} \quad (8)$$

with

$$C_{Infra} = C_{Prober} + C_{ATE,0}. \quad (9)$$

It can now be seen that multi-site test fully divides the infrastructure cost, i.e. the cost for the prober and the zero-channel ATE, whereas the channel cost and per-site cost remain unchanged as a best case, if all sites the ATE is equipped for are actually used, $S = \overline{S}$. Otherwise the latter cost contributor is increasing.

2.2 Operating cost

Operating cost covers the salaries for operating personnel, floor space in a building, cost for power, training, maintenance and so on. It is independent of the capital equipment cost and is therefore bundled into a fix cost rate R_{Op}, e.g. 35 \$/hour = 0.97 ¢/sec. When multiple devices are tested in parallel, the cost rate is shared across $\overline{S}$ sites:

$$c_{Op} = \frac{R_{Op} \cdot t_{Tot}}{\overline{S}} \quad (10)$$

It can be seen from the above equation that multi-site fully divides the operating cost portion.

2.3 Cost rate for test cell

Using equations (5) and (10), the test cell related cost as defined by equation (2), can now be combined to

$$c_{Testcell}(\overline{S}) = \frac{R_{Cap} + R_{Op}}{\overline{S}} \cdot t_{Tot} = \frac{R_{Testcell}}{\overline{S}} \cdot t_{Tot} \quad (11)$$

2.4 Total time on test cell

Equation (10) shows that the total time t_{Tot} on the test cell is another very important factor. This time consists primarily of the effective test time $\hat{t}_{test}$, the stepping time t_{Step} required by the prober to move to the next devices, and the changeover time t_{Lot} between production lots.

$$t_{Tot} = \hat{t}_{Test} + t_{Step} + t_{Lot} \quad (12)$$

The following subsections will show that the effective test time $\hat{t}_{test}$ is proportional to the single-site test time t_{Test} without any parallelism,

$$\hat{t}_{Test} = k \cdot t_{Test} \quad (13)$$

and that the proportionality factor k itself can be factored into several correction factors.

$$k = \frac{\hat{t}_{Test}}{t_{Test}} = k_{Conc} \cdot k_{Seq} \cdot k_{Fail} \cdot k_{Retest} \quad (14)$$

k_{Conc} models the test time reduction when multiple blocks in a device are tested concurrently, see paragraph 2.5. k_{Seq} reflects longer effective test time due to partially sequential multi-site test, see paragraph 2.6. k_{Fail} is a factor that models the reduced opportunity in multi-site test to stop a failing test early, see paragraph 2.7. Finally, k_{Retest} accounts for retest due to contacting problems, see paragraph 2.8.

2.5 Concurrent test

While multi-site test increases the parallelism during test by testing multiple devices at the same time, concurrent test covers multiple blocks in the same device at the same time [4] [5]. This concept is illustrated in the following figure, using an example with just 2 pin groups.

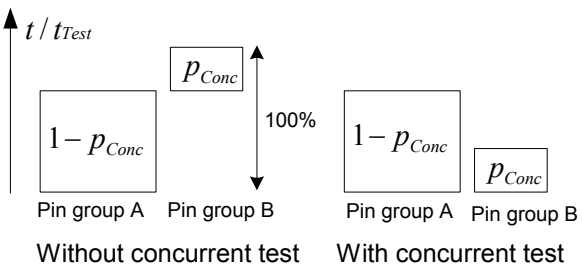

Figure 1: Concurrent test reduces test time

When p_{Conc} is the percentage of test time that can be hidden behind other tests, i.e. executed in parallel to other tests, the resulting test time $\hat{t}'_{Test}$ is reduced to a fraction

$$k_{Conc} = \frac{\hat{t}'_{Test}}{t_{Test}} = 1 - p_{Conc} \qquad (15)$$

of the original test time t_{Test}. This requires enough independent ATE resources for pin groups A and B, in particular clocks, sequencers and mixed-signal resources.

2.6 Partially sequential test, multi-site efficiency

When not enough independent ATE resources are available for all sites to run all test steps in parallel, some test steps must be executed sequentially, diminishing the benefits of multi-site test, see Figure 2.

When a fraction p_{Seq} of the test time t_{Test} must be executed sequentially for each site, the total test time $\hat{t}'_{Test}$ is given by:

$$\hat{t}'_{Test} = \left(1 - p_{Seq}\right) \cdot t_{Test} + p_{Seq} \cdot t_{Test} \cdot \overline{S} \qquad (16)$$

The correction factor in equation (13) is

$$k_{Seq} = \frac{\hat{t}'_{Test}}{t_{Test}} = 1 + \left(\overline{S} - 1\right) \cdot p_{Seq} . \qquad (17)$$

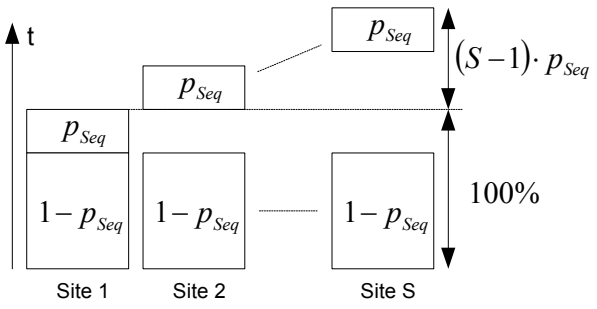

Figure 2: Partially sequential multi-site test

To reap most of the benefits of multi-site test, e.g. $k_{Seq} < 1.5$, the percentage of sequential test should be less than

$$p_{Seq} < \frac{1}{2\left(\overline{S} - 1\right)} \text{ for } k_{Seq} < 1.5 . \qquad (18)$$

An efficient use of 16 sites, suggests less than 3% sequential test, i.e. a multi-site efficiency of 97%. Multi-site efficiency can be degraded when not enough independent mixed-signal, RF, clock resources are available in the ATE and some tests must be executed sequentially. Upload time of results from digitizers or DC measurements to a central workstation can also cause undesired sequential test time overhead. An ideal ATE has an abundance of affordable mixed-signal test resources and processes digitizer results locally in the analog instruments to avoid time overhead altogether.

2.7 Stop on early fail

A single-site test can be stopped immediate when the first test step fails. This can be very early, because most defect devices show easily detectable catastrophic behavior. For bad devices the test time for the remaining test steps can be saved. When Y denotes the yield, i.e. the percentage of good devices, the percentage of bad devices is $1 - Y$. Those failing devices require only a fraction p_{Fail} of the full test time t_{Test}, while the fraction Y of good parts requires the full test time. The average test time for single site test is thus given by:

$$\hat{t}'_{Test} = \left(1 - Y\right) \cdot p_{Fail} \cdot t_{Test} + Y \cdot t_{Test} \qquad (19)$$

In multi-site test, all sites must contain bad devices to benefit from this test time reduction. If only one site is good, the tests have to be completed. With a yield of Y, the likelihood of all $\overline{S}$ devices being bad is only $\left(1 - Y\right)^{\overline{S}}$. I.e., for multi-site test, the average test time is

$$\hat{t}'_{Test} = \left(1 - Y\right)^{\overline{S}} \cdot p_{Fail} \cdot t_{Test} + \left(1 - \left(1 - Y\right)^{\overline{S}}\right) \cdot t_{Test} \qquad (20)$$

The correction factor for equation (13) is therefore

$$k_{Fail} = \frac{\hat{t}'_{Test}}{t_{Test}} = 1 - \left(1 - p_{Fail}\right) \cdot \left(1 - Y\right)^{\overline{S}} \qquad (21)$$

2.8 Retest rate

Due to contacting problems, a fraction p_{Retest} of the parts requires re-contacting and re-testing. Multi-site requires more contacts and therefore more retest. Retest is not needed, when all $\overline{S}$ sites are contacted correctly, i.e. with a likelihood of $\left(1 - p_{Retest}\right)^{\overline{S}}$. Otherwise the test is executed one more time. The effective test time is then

$$\hat{t}'_{Test} = \left(1 - p_{Retest}\right)^{\overline{S}} \cdot t_{Test} + \left(1 - \left(1 - p_{Retest}\right)^{\overline{S}}\right) \cdot 2t_{Test}$$

The test time correction factor for retest is:

$$k_{Retest} = \frac{t'_{Test}}{t_{Test}} = 2 - \left(1 - p_{Retest}\right)^{\overline{S}} . \qquad (22)$$

2.9 Changeover time between production lots

When a production lot starts with a different device under test, the test cell must be prepared for that new device, which requires changing the probe card and loading a new test program. During that time the test cell cannot be used for testing. When a lot includes N_{Lot} devices and $\overline{S}$ devices are tested in parallel, a lot needs $N_{Lot} / \overline{S}$ test cycles. In average, each test cycle is then prolonged by

$$t_{Lot} = \frac{T_{Lot}}{N_{Lot} / \overline{S}} = \overline{S} \cdot \frac{T_{Lot}}{N_{Lot}} , \qquad (23)$$

where T_{Lot} is the changeover time between lots. For a given lot size, the reduced number of test cycles in multi-site

test makes a given changeover time relatively more pronounced. When e.g. a changeover time of 15 minutes is shared across a lot of 24 200mm wafers, each containing about 15,000 dice of 50mm^2, the time penalty per test cycle t_{Lot} is 60ms in single site configuration or almost 1sec for 16 sites. Except for very large dice, this penalty will not be significant.

2.10 Partial deployment of multi-site test

A test cell must be equipped to cover a large variety of devices with differing ATE requirements. Test cells are therefore, in general, under-utilized for most devices. Multi-site test can improve the situation, but can also make it worse. When a tester is equipped with e.g. 1000 channels to handle some high-count devices, multiple smaller pin-count devices can be tested in multi-site to better utilize these 1000 tester channels. However, the under-utilization increases if testers are equipped with many resources to support multi-site test, but only some devices are actually tested in multi-site.

When a fraction p_{MS} of devices is tested in multi-site configuration with S sites, while the remaining fraction $1-p_{MS}$ is tested in single site, the average cost is:

$$\bar{c} = p_{MS} \cdot c(S) + (1-p_{MS}) \cdot c(1) \qquad (24)$$

For the test cell related cost, which is usually the largest cost contributor, $\bar{c}_{Testcell}$ can be simplified using equation (11) to:

$$\bar{c}_{Testcell} = \left[p_{MS} + (1-p_{MS}) \cdot S\right] \cdot \frac{R_{Testcell}}{S} \cdot t_{Tot} \qquad (25)$$

or

$$\bar{c}_{Testcell} = \frac{R_{Testcell}}{(k_{MS} \cdot S)} \cdot t_{Tot} = \frac{R_{Testcell}}{\bar{S}} \cdot t_{Tot}, \qquad (26)$$

where $\bar{S}$ is the effective number of test sites that has been used so far

$$\bar{S} = k_{MS} \cdot S \qquad (27)$$

and k_{MS} is the correction factor that models partial deployment of multi-site test:

$$k_{MS} = \frac{\bar{S}}{S} = \frac{1}{p_{MS} + (1-p_{MS}) \cdot S} \qquad (28)$$

2.11 Cost of probe-cards

More probe-card contacts make probe-cards for multi-site more expensive. On the other side, each probe-card can serve more devices before it is worn out, because each touchdown contacts multiple devices at a time. The inevitable spare units, however, are more expensive without bringing any benefits. These effects will be modeled now.

Probe-cards are device specific or specific to a family of devices. Because of their fragile nature and because of their limited lifetime in terms of maximum number of touchdowns, multiple probe cards N_{PC}, each costing C_{PC} must be purchased for the lifetime unit volume N_{LT} of a device. The cost contribution per device is thus given by:

$$c_{PC} = \frac{N_{PC} \cdot C_{PC}}{N_{LT}} \qquad (29)$$

After every N_{TD} touchdowns, each contacting $\bar{S}$ sites, the probe-card is worn out and a new one must be purchased. N_{Spare} additional spare units are needed so that production can be continued if some units are broken or must be cleaned. The total number of required probe cards is

$$N_{PC} = N_{Spare} + \left\lceil \frac{N_{LT}}{N_{TD} \cdot \bar{S}} \right\rceil, \qquad (30)$$

where $\lceil .. \rceil$ denotes rounding up to the next integer.

The probe-card cost C_{PC} is modeled with a cost per contact $C_{1Contact}$ to reflect the increased complexity and cost of a probe-card for multi-site. When the percentage p_{Ch} of all device pads that require an ATE channel equals N_{Ch} pads, the total number of contacts for each of the $\bar{S}$ sites is N_{Ch} / p_{Ch}.

$$C_{PC} = C_{1Contact} \cdot N_{Contacts} = C_{1Contact} \cdot \frac{N_{Ch}}{p_{Ch}} \cdot \bar{S} \qquad (31)$$

Combined, the probe-card cost is:

$$c_{PC} = \frac{\left(N_{Spare} + \left\lceil \frac{N_{LT}}{N_{TD} \cdot \bar{S}} \right\rceil \right) \cdot C_{1Contact} \frac{N_{Ch}}{p_{Ch}} \cdot \bar{S}}{N_{LT}} \qquad (32)$$

It can now be seen from the above equation that multi-site makes the N_{Spare} spare units more expensive but leaves the cost for worn out probe-cards basically unchanged, although the effect of rounding up to the next integer number of probe cards is emphasized because fewer probe-cards are needed. For a device with 256 test contacts, each 20 $, two 2 spare 16-site probe-cards cost 160 k$, or 1.6 ¢ for each of 10 million parts.

2.12 Packaging bad parts

When reduced pin-count testing (RPCT) is used to enable a higher degree of multi-site at relatively small incremental cost, the test coverage will decrease to some extend, increasing the number of bad parts that will not be identified as such, and that will be packaged unnecessarily. The associated package cost must be attributed to test and is given by

$$c_{Pkg} = (1-Y) \cdot (1-E) \cdot C_{Pkg}, \qquad (33)$$

where Y is the yield, i.e. the percentage of good parts that enter test; E is the test effectiveness, i.e. the percentage of bad parts that test is able to identify as bad; C_{Pkg} is the cost of a single package including the cost of package test.

3. Example

The cost benefits of parallel test will be quantified for a high-volume cell-phone base-band device with moderate pin-count and some mixed-signal content, which implies some (per-site) analog resources and some parametric yield loss. The assumptions are listed in Table 1.

Table 1: Assumptions for cost example

Device under test		
Number of channels during test	N_{Ch}	128
Signal (channel) to pad ratio	p_{Ch}	50 %
Yield	Y	80 %
Package cost	C_{Pkg}	$ 1
Lifetime unit volume	N_{LT}	25 Mio
Test method		
Test time (without parallelism)	t_{Test}	5 sec
Percentage of concurrent test	p_{Conc}	0 %
Test time for failing device	p_{Fail}	10 %
Test effectiveness	E	99 %
ATE		
Cost for zero-channel ATE	$C_{ATE,0}$	$ 400k
Cost per ATE channel	C_{1Ch}	$ 500
Additional cost per site	C_{1Site}	$ 50k
Equipped for number of sites	S	Varied
Sequential test time	p_{Seq}	10 %
Prober		
Cost of prober	C_{Prober}	$ 300k
Stepping time	t_{Step}	0.4 sec
Test cell		
Operating cost rate	R_{Op}	$ 35/h
Test cell utilization	p_{Util}	65 %
Depreciation time	T_{Depr}	5 yrs
Probe card		
Cost per contact	$C_{1Contact}$	$ 20
Number of touchdowns	N_{TD}	1 Mio
Number of spare probe cards	N_{Spare}	2
Test flow		
Device % tested in multi-site	p_{MS}	100 %
Retest rate for single site	p_{Retest}	3 %
Lot size in number of devices	N_{Lot}	15,000
Change-over time between lots	T_{Lot}	15 min

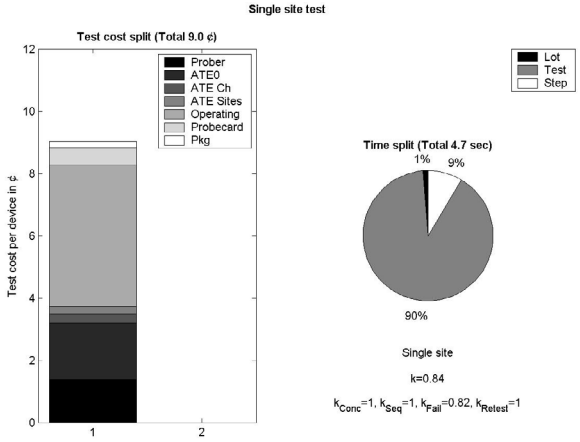

Figure 3: Test cost split for single site test

For single-site test, which will be used as the baseline, the total test cost, 9¢ per device, is dominated by operating cost and capital depreciation, while the total time of 4.7 sec is clearly dominated by the effective test time, see Figure 3.

4. Multi-site test

Because operating cost and capital for infrastructure account for about 75% of the test cost, we can expect improvements from multi-site test. Indeed, Figure 4 shows that multi-site test reduces test cost by almost 50%, with 7 sites, but increases again for more than 7 sites. Although 7 is not a very practical number, it will be used for comparison.

The test time split for 7 sites in Figure 5 shows a test time penalty of +60% for sequential test, ($k_{Seq} = 1.6$; 10% for each of 6 additional sites). This penalty makes larger number of sites unattractive.

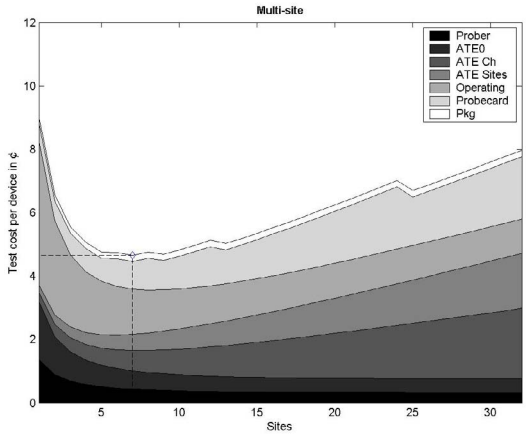

Figure 4: Multi-site reduces test cost by ≈50%

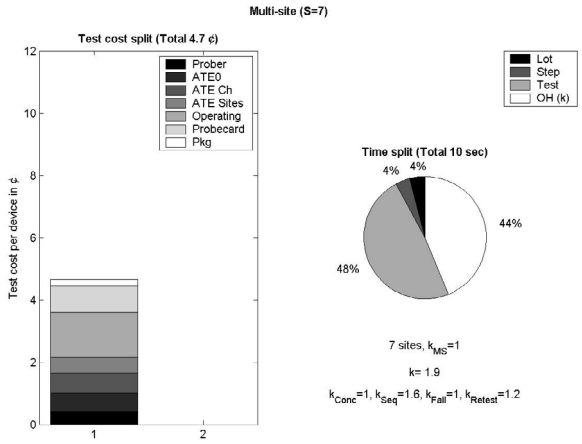

Figure 5: Test cost split for 7 sites

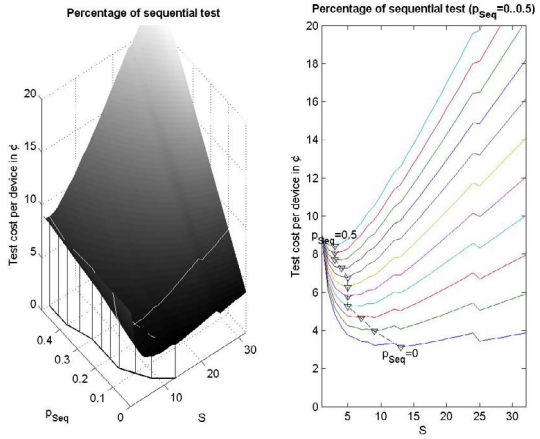

Figure 6: Sensitivity to multi-site efficiency, measured in percentage p_{Seq} of sequential test

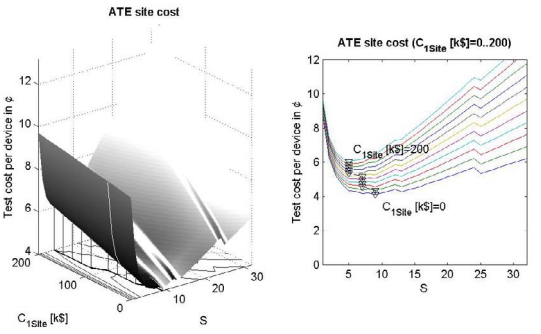

Figure 7: Sensitivity to per-site resource cost, e.g. for mixed-signal instrumentation and device power supplies

Figure 6 shows the strong sensitivity to the percentage p_{Seq} of test that must be executed sequentially in a multi-site test with S sites. The projected solid line below the surface shows the optimum number of sites for each value of p_{Seq}. As expected, multi-site test loses its benefits with increasing sequential test. For truly parallel test, $p_{Seq} = 0$, the optimum number of sites increases to 12.

True parallel test requires an abundance of affordable mixed-signal resources and device power supplies in the ATE. The sensitivity to per-site ATE resource cost, C_{1Site}, shown in Figure 7, is clearly visible, but not as strong as one might expect. When per-site resource cost ranges from \$10k to \$200k, the optimum number of sites varies only from 8 to 4.

Comparing the cost split for single site test in Figure 3 with the one for multi-site test in Figure 5 shows that both the probe-card contribution and the ATE channel resource contribution were increasing.

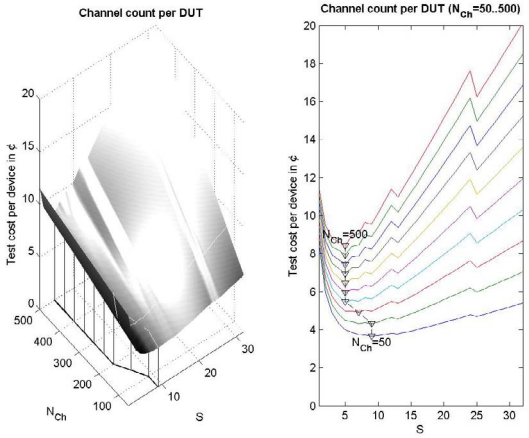

Figure 8: Fewer channels during test reduce both ATE and probe-card cost

Reduced pin-count testing (RPCT) [13] reduces both the number of ATE channels and the number of probe-contacts, however at the expense of some test coverage, because not all I/O are fully tested anymore. Techniques like I/O wrap can be used to mitigate this effect. An increased test time can be avoided by using I/O bandwidth matching techniques [3] [7] [8] [13].

This scenario, optimized for multi-site, assumes no sequential test, reduced pin-count test with 32 ATE channels per site, a 1% coverage loss resulting from less I/O test, and a reduced cost of per-site resources. The assumed differences compared to the baseline in Table 1 are summarized in Table 2.

Table 2: Changes for optimized multi-site test

Sequential test time	p_{Seq}	0 %
Number of channels during test (RPCT)	N_{Ch}	32
Test effectiveness	E	98 %
Additional cost per site	C_{1Site}	\$ 20k

Under these assumptions, the total test cost decreases from 9¢ to 1.6¢, while the optimum number of sites increases from 7 to 25. At this point, about half of the test cost can be attributed to probe-card cost and cost of packages due to yield loss of RPCT.

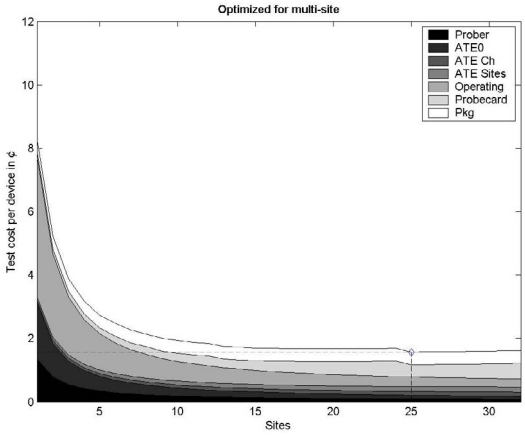

Figure 9: Optimized multi-site test reduces test cost by a factor of 5

5. Concurrent test

Both, the test time split for single site test, Figure 3, and for multi-site test, Figure 5, show that the effective test time is 90% of the total time. Changeover times between production lots and stepping time are both insignificant.

Since concurrent test reduces the 90% portion, major savings can be expected. Figure 10 shows the effect under the assumptions of Table 1.

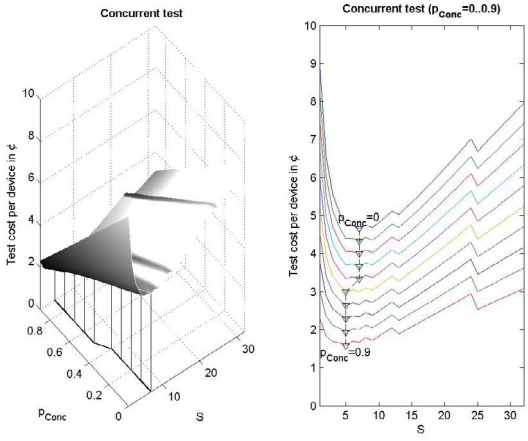

Figure 10: Concurrent test reduces test cost significantly

When $p_{Conc} = 40\%$ of the test time can be hidden behind other tests, the test cost reduces from originally 9¢ with 7 sites to about 3.3¢ with also 7 sites.

6. Parallel test

When optimized multi-site test, see Table 2, is combined with concurrent test ($p_{Conc} = 40\%$) to a fully parallel test, test cost reduces slightly to 1.4¢ for 25 sites, see Figure 11.

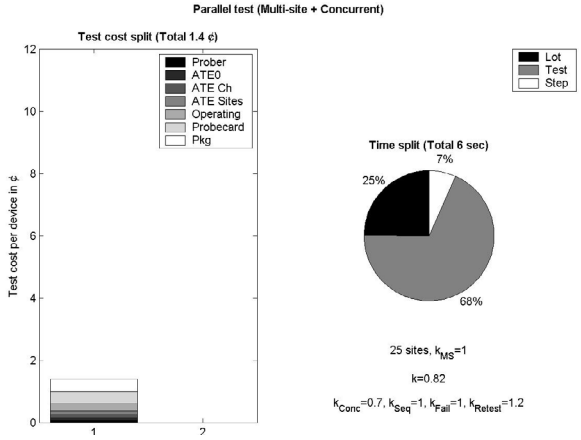

Figure 11: Lowest test cost with parallel test

For this scenario, the sensitivities against parameter variations will be investigated now.

Multi-site efficiency is a very sensitive parameter as has been demonstrated before. The solid projected line under the 3D plot in Figure 12 shows that test cost and the optimum number of sites depend strongly on the percentage p_{Seq} of sequential test.

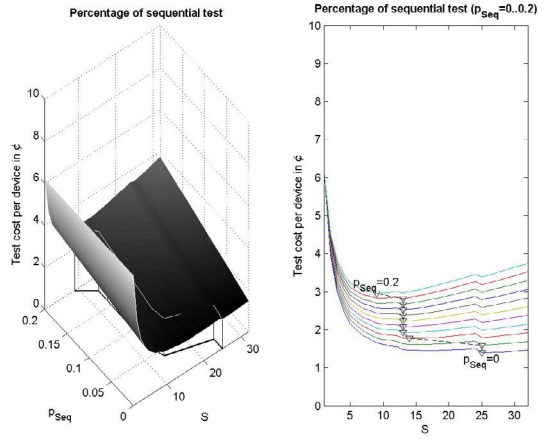

Figure 12: Sensitivity to multi-site efficiency measured in percentage of sequential test

When ATE has been equipped with enough resources for multi-site, it is important to also use the resources in a multi-site test configuration, see paragraph 2.10. For a high-volume device that can keep one or more ATE busy, this is usually not a constraint, since the ATE can then be configured to the specific needs of that device. With a 5 sec test time it takes 4 million devices per year to keep an ATE busy all year round, at 65% utilization.

For cases where the ATE configuration is not optimized for the device, Figure 13 shows the sensitivity to under-utilized ATE resources, where the ATE is equipped to support S sites, but only a fraction p_{MS} is tested with S sites, the rest is tested in single-site configuration. According to the right graph in Figure 13, it is still better to use equipment,

which can support 10 sites, for 25% of the devices in multi-site configuration, rather than always use single-site equipment.

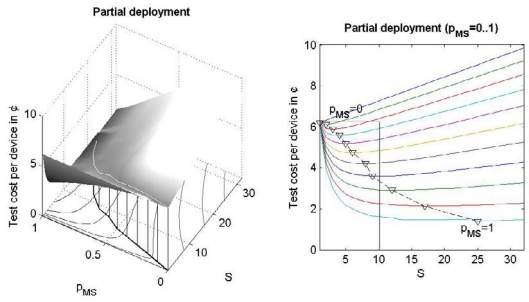

Figure 13: Sensitivity to partial deployment of multi-site test

Figure 14 shows that the sensitivity to the lifetime unit volume.

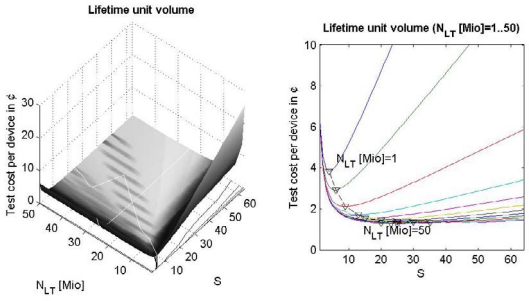

Figure 14: Sensitivity to lifetime unit volume

For low volumes, the high cost of probe-cards suggests fewer sites, although the effect is minimal for more than 5 million devices. The often talked about 'stop on early fail', see paragraph 2.7, has almost no visible influence beyond a few sites, as shown in Figure 15.

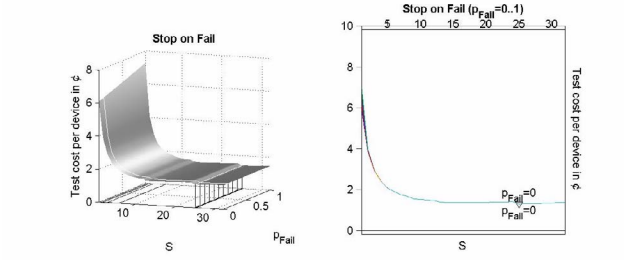

Figure 15: Minimum sensitivity of 'Stop on Fail'

The sensitivity to 'retest' and ATE channel cost are similar.

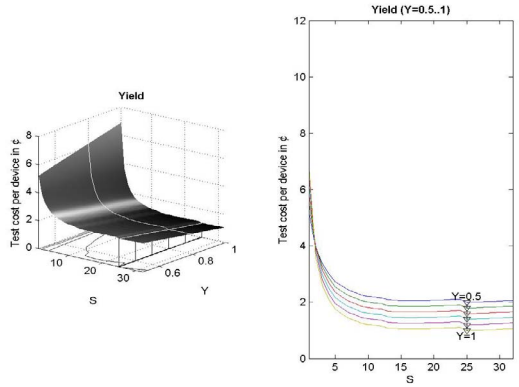

Figure 16: Yield has no influence on best number of sites

Figure 16 shows that the product yield influences the overall test cost, but does not change the optimum number of sites.

The sensitivity charts for test effectiveness, operating cost, and test cell utilization show the same behavior. They show no influence on the optimum number of sites.

7. Low-cost ATE

So-called 'low cost testers' are often the intuitive answer to reducing cost of test. Although reduced ATE capital does obviously reduce test cost, the effect is often over-emphasized.

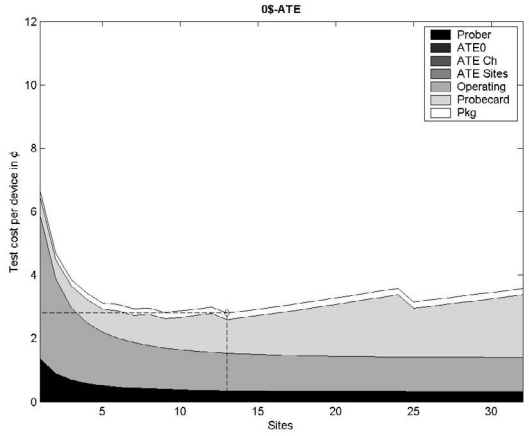

Figure 17: Low-cost ATE reduces test cost, but less than optimized muti-site test

Table 3: Assumptions for "0$-ATE"

Cost for zero-channel ATE	$C_{ATE,0}$	$ 0k
Cost per ATE channel	C_{1Ch}	$ 0
Additional cost per site	C_{1Site}	$ 0k

To stress this point, parallel test will now be compared to a '0$-ATE', see Table 3. According to Figure 17, the lowest test cost is 3.4¢ per device with 13 sites.

8. Comparison

Figure 18 compares the total test cost of all above-mentioned scenarios as a function of the number of sites. Figure 19 shows the test cost split for the best number of sites for each scenario.

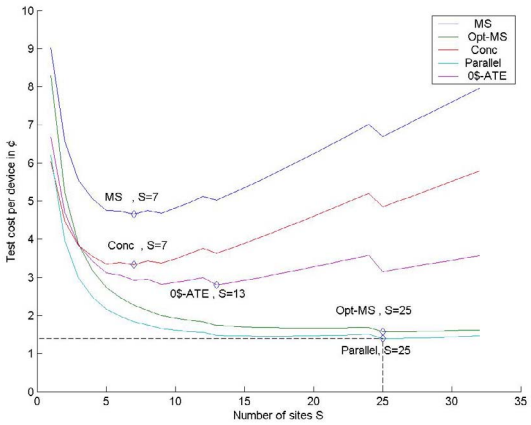

Figure 18: Comparison of scenarios

It can be seen that optimized multi-site test 'Opt-MS' and fully parallel test 'Parallel' achieve by far the lowest test cost, even significantly lower than with a free tester '0S-ATE'. This shows that an effective cost reduction strategy must be based on exploiting parallelism rather than just trying to reduce capital cost.

The benefits of single-site concurrent test are significant, but diminish when being combined with multi-site test. However, a combined parallel test strategy is less sensitive to parameter variations.

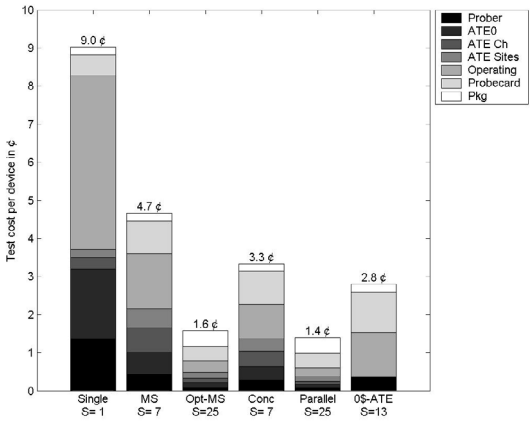

Figure 19: Comparison of scenarios for best number of sites

9. Conclusions

Both parallel test approaches, multi-site test and concurrent test, are very powerful test cost reduction methods. Optimized multi-site test achieves lower cost than just low-cost ATE, even than a free ATE.

Multi-site test works best for high-volume devices with moderate pin-count, or when reduced pin-count testing can be used to reduce the ATE channel and probe-card cost. Multi-site requires an ATE with enough digital channels and device power supplies. For mixed-signal devices, a large number of independent and affordable mixed-signal resources and local result processing are needed to maximize multi-site efficiency. Typically, the return for multi-site test starts to diminish beyond about 16 sites. The main reasons are more expensive probe-cards, and a higher risk of resource under-utilization. Time overheads for retest, changeover times between lots, and less opportunity to stop on first fail have only minor impact beyond a few sites. Yield, test effectiveness, ATE utilization, operating cost influence test cost, but have no impact on the optimum number of sites.

Also concurrent test can lead to significant test cost reduction when multiple device ports can be tested at the same time. This requires an ATE with independent resources per port, in particular independent clocks and independent sequences. A per-pin ATE architecture is ideally suited for this purpose.

Although a combined parallel test strategy does not bring the added benefits of multi-site test and concurrent test, it reduces the sensitivity to various parameters.

Acknowledgments

My thanks go to Walter Benger, Hendrik Volkerink, Klaus-Dieter Hilliges, and Holger Engelhard of Agilent Technologies for many stimulating and insightful discussions.

References

1. Semiconductor Industry Association (SIA), "*International Technology Roadmap for Semiconductors, 2003 Edition",* http://public.itrs.net
2. G. Moore, "No Exponential Is Forever: But "Forever" Can Be Delayed", *IEEE Solid-State Circuits Conference,* 2003
3. A. Khoche, J. Rivoir, "I/O Bandwidth Bottleneck for Test: Is it Real?", *1st IEEE Workshop on Test Resource Partitioning,* 2000
4. K. Hilliges, A. Khoche, D. Armstrong, H. Volkerink, "Test Resource Partitioning for Concurrent Test", *2nd IEEE Workshop on Test Resource Partitioning",* 2001
5. K. Hilliges, "Concurrent Test", *www.agilent.com,* 2001
6. W. Radermacher, J. Rivoir, "An Evolution to a DFT-Centric Test Paradigm that Scales with Technology Progress", *European Test Workshop,* 2001
7. J. Rivoir, H. Volkerink, A. Khoche, "Reduced Pincount Multisite Test with I/O Bandwidth matching – Why and How?", *2nd IEEE Workshop on Test Resource Partitioning,* 2001
8. H. Volkerink, A. Khoche, J. Rivoir, K. Hilliges, "Test Economics for Multi-Site Test with Modern Test Cost Reduction Techniques", *VLSI Test Symposium,* 2002
9. H. Volkerink, A. Khoche, J. Rivoir, "Modern Test Techniques: Tradeoffs, Synergies, and Scalable Benefits", *Special Issue of Journal of Electronic Testing: Theory and Applications, Kluwer Academic Publishers,* 2002
10. B. Davis, "The Economics of Automatic Testing", *London, McGraw-Hill,* 1994

11. A.C Evans, “Applications of Semiconductor Test Economics, and Multi-site Testing to Lower Cost of Test”, *IEEE International Test Conference,* 1999
12. H. Volkerink, A. Khoche, L. Kamas, J. Rivoir, H. Kerkhoff, “Tackling Test Tradeoffs from Design, Manufacturing to Market using Economic Modeling”, *IEEE International Test Conference,* 2001
13. H. Vranken, T. Waayers, H. Fleury, D. Lelouvier, “Enhanced Reduced Pin-Count Test for Full-Scan Designs”, *European Test Workshop,* 2001
14. A Khoche, R. Kapur, D. Armstrong, T.W. Williams, J. Rivoir, “A New Methodology for Improved Tester Utilization”, *IEEE International Test Conference,* 2001