Test Economics for Homogeneous Manycore Systems

Lin Huang† and Qiang Xu†‡
†CUhk REliable computing laboratory (CURE)
Department of Computer Science & Engineering
The Chinese University of Hong Kong, Shatin, N.T., Hong Kong
‡CAS-CUHK Shenzhen Institute of Advanced Integration Technology
Email: {lhuang,qxu}@cse.cuhk.edu.hk

Abstract

Homogeneous manycore systems that contain a large number of structurally identical cores are emerging for tera-scale computation. To ensure the required quality and reliability of such complex integrated circuits before supplying them to final users, extensive manufacturing tests need to be conducted and the associated test cost can account for a great share of the total production cost. By introducing spare cores on-chip, the burn-in test time can be shortened and the defect coverage requirements for core tests can be also relaxed, without sacrificing quality of the shipped products. The above test cost reduction is likely to exceed/compensate the manufacturing cost of the extra cores, thus reducing the total production cost of manycore systems. We develop novel analytical models that capture the above tradeoff in this paper and we verify the effectiveness of the proposed test economics model for hypothetical manycore systems with various configurations.

1 INTRODUCTION

Customers have high expectations for the quality and reliability of semiconductor products, and typically only a few hundred defective parts per million (DPPM) are allowed with a lifespan of several years. Manufacturing test is responsible for achieving this duanting objective and several types of testing are performed at different stages of the integrated circuit (IC) manufacturing process. On one hand, sophisticated automatic test-pattern generation (ATPG) techniques are needed to achieve adequately high defect coverage during the wafer and final test of the ICs. As technology advances, test patterns that target delay faults and many other kinds of subtle errors (e.g., signal integrity faults) are also essential to guarantee test quality, in addition to the traditional stuck-at vectors. The associated large number of test patterns not only require long testing time on the automatic test equipment (ATE), but also indirectly result in more false rejects and thus lower the manufacturing yield of the integrated circuits. On the other hand, accelerated testing methods such as burn-in test are used to screen out those chips with early-life failures to enhance product reliability. For ICs fabricated with latest technology, it is increasingly difficult to setup and control appropriate stress conditions for the circuits during the burn-in process [24], which makes it the bottleneck of the manufacturing process [30]. Due to the above, manufacturing test cost can account for a great share of the total production cost for complex ICs. In particular, burn-in cost may range from 5% to 40% of the total product cost, as pointed out in [23].

At the same time, advancements in semiconductor technology enable integration of a large number of cores on a single silicon die. Many state-of-the-art computing systems employ multiple identical cores on a single chip, known as multicore or manycore systems (e.g., multi-core DSPs [14], multi-core GPUs [18], multi-core processors [29]). They are shown to be much more power-efficient and therefore have become increasingly popular in the industry [6, 19]. To improve the manufacturing yield of such complex circuits, typically a few *yield-driven redundant cores* are placed on-chip and the system can be reconfigured to bypass those defective cores [31]. For instance, the 192-core Cisco Metro network processor [3] contains four spare cores while the 128-core nVidia GeForce 8800 GPU [18] can be degraded to a 96-core version.

With the defect-tolerance property of homogeneous manycore systems, it is possible to introduce one or more spare cores dedicated for test cost reduction (namely *test cost-driven redundant cores*) [8, 27], in addition to those yield-driven spares. Consider a manycore chip that functions with 16 cores, to guarantee that all 16 cores work well provided that they have passed manufacturing test, we need core tests with very high defect coverage to identify chips containing *killer defects* and sufficient burn-in to weed out chips with *latent defects*. If, however, we add two test cost-driven redundant cores on-chip, by contrast, since manufacturing test is responsible for 16 out of 18 cores (instead of all the 18 cores) to work after the so-called *infant mortality* period (the functioning cores can be identified using online testing techniques, if necessary), the defect coverage requirement for the core tests can be lowered while burn-in test can be also reduced or eliminated. If the associated test cost reduction exceeds the manufacturing cost increment for the spare cores, we are able to cut down the total production cost for the homogeneous manycore systems without sacrificing the required quality and reliability of the shipped products.

In this paper, we propose comprehensive analytical model to study the test economics for homogeneous manycore systems. We first consider the case of introducing spare cores for partial burn-in test (or the elimination of burn-in test), given

978-1-4244-4867-8/09/$25.00

fixed defect coverage of core tests. Next, we relax the defect coverage constraint and study the complex relationship among test coverage, test escapes and false rejects, partial/no burn-in test, yield-driven redundancy, test cost-driven redundancy, and product quality. Experimental results on hypothetical many-core systems with various configurations in terms of defect density and test cost distributions show the effectiveness of the proposed analytical model.

The remainder of this paper is organized as follows. Section 2 gives preliminaries and formulates the problems studied in this paper. The proposed analytical models to solve these problems are then detailed in Section 3 and Section 4, respectively. Next, experimental results are presented in Section 5 to show the effectiveness of the proposed analytical model. Finally, Section 6 concludes this paper.

2 PRELIMINARIES AND PROBLEM FORMULATION

2.1 Background

Integrated circuit fabrication is an extremely complex process. It is inevitable that some manufactured chips are defective. Prior work has proposed several methods to model the spatial distribution of defects on the wafer [13], and it was shown that negative-binomial distribution fits quite well with the actual defect distribution [10], as IC defects typically feature "clustering" effects. In terms of defect type, there are *killer defects* that result in direct chip failures and *latent defects* causing early-life reliability failures. Experiments conducted at Intel over a wide range of yield values showed that the two kinds of defects have a linear relationship, and typically, for every 100 killer defects present, one expects, on average, 1-2 latent defects [25].

IC products typically experience decreasing failure rate in their early lifetimes, followed by a nearly constant failure rate, and then an increasing one in their later life, known as the classical bathtub curve (as shown in Fig. 1). The first stage (caused by latent defects) is commonly referred as *infant mortality* period, and its reliability function follows Weibull distribution with shape parameter β and scale parameter θ [1], which reflect the core structural properties and usage-related factors (such as, operational voltage and frequencies), respectively [9]. Before burn-in process, latent defects are not significant enough to reveal themselves. By stressing the circuit at elevated temperature and voltage during the burn-in process, the number of latent defect-induced failures increases and these weak chips can be identified.

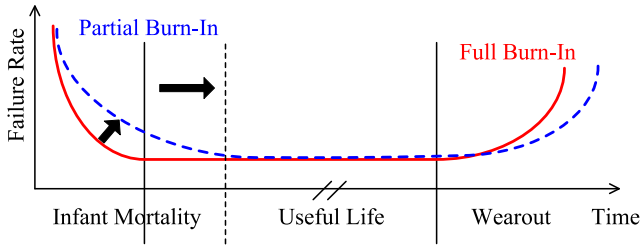


Figure 1. Bathtub Curve.

One notable feature in IC testing is that most test patterns are applied to achieve the last several percentages of defect coverage. For example, according to the model presented in [5], to improve defect coverage from 99% to 99.9%, the number of test patterns increases for about 50%. In addition, testing with extremely high defect coverage leads to more false rejects and thus lower the IC's manufacturing yield. Consequently, if we are able to relax this coverage requirement, the manufacturing test cost can be dramatically reduced. While the demand for IC product quality constrains the above test cost reduction strategy for circuits with irregular structures, the emerging homogeneous manycore systems that contain a large number of structurally identical cores, being defect-tolerant, provide us such an opportunity. In addition, considering the high cost of burn-in test due to their lengthy testing time, introducing spare cores on-chip can also alleviate the burden for burn-in test to reveal all latent defects, which facilitates us to conduct partial burn-in or even eliminate burn-in test completely. Obviously, the manufacturing cost per fabricated chip increases with more redundancy, however, if the associated test cost reduction exceeds the manufacturing cost increment for the spare cores, we are able to cut down the total production cost for the homogeneous manycore systems. This has motivated the test economics model studied in this paper.

In [8], we proposed the concept of test cost-driven redundancy for homogeneous manycore systems and used a case study to show its potential advantage, but the detailed analysis is missing due to space limitation of the paper. Shamshiri *et al.* [27] presented a cost analysis framework for manycore systems with spares and advocated to introduce redundant cores to eliminate burn-in test. The authors considered the case to ship ICs with high DPPMs to customers and a large service cost is modeled for replacement. The test economics model presented in that work, however, has several limitations: (i). It neglected the correlation among parameters that are strongly related. For example, the manufacturing yield for cores are determined by the defect density and the "clustering" defect distribution parameter, but they are set as arbitrary values in [27]; (ii). The impact of defect coverage on testing cost is not considered in that work; (iii). This work only analyzed the case for either full burn-in or no burn-in, without considering partial burn-in; (iv). The difference between yield-driven redundancy and test cost-driven redundancy is not considered in [27].

Different from [27], we consider to introduce test cost-driven redundancy on-chip without sacrificing the required quality and reliability for shipped manycore products and it is hence not necessary to consider the service cost for replacement of defective chips. More importantly, our proposed model captures the complex relationship among test coverage, test escapes and false rejects, partial burn-in test, yield-driven redundancy, test cost-driven redundancy, and product quality.

It is important to note, effective online testing techniques (e.g., [2, 12, 21]) are essential to identify functioning cores in manycore systems for the success of the above test cost reduction methodology. With increasingly adverse reliability threats for nanometer-scale VLSI circuits, we believe such techniques are available in future large-scale manycore systems.

2.2 Problem Formulation

We model the test economics for homogeneous manycore processors progressively in this paper.

Firstly, with given defect coverage for core tests, we consider to introduce t redundant cores into a homogeneous manycore system that functions if no less than m cores are defect-free for partial/no burn-in. That is, in total we fabricate $u = m+t$ cores on a chip. As some latent defects are not detectable because of insufficient burn-in, only chips with all u cores pass test are sold to ensure product quality and reliability. Eventually we only need to guarantee m cores are defect-free at the end of infant mortality period. This problem can be formulated as:

Problem 1 [Partial Burn-In]: Given

- *design and material-related parameters*
- the required number of cores m for the homogeneous manycore chip to function;
- the wafer dimension d, and the silicon area of a core A_r;
- *defect distribution parameters*
- the average number of killer defects per core λ_K and the clustering parameter α, assuming defects to follow the negative-binomial distribution;
- the ratio between the average number of latent defects per core and that of killer defects per core γ;
- *test-related parameters*
- the product quality requirement for the maximum test escape percentage τ (e.g., $\tau = 0.0005$ for 500 DPPM);
- sufficiently high defect coverage of manufacturing test B_r to ensure product quality;
- *burn-in related parameters*
- infant mortality time with full burn-in that reveals all latent defects T_{IM};
- shape parameter of reliability function in infant morality duration β;
- *cost-related parameters*
- the ratio between ATE cost for applying manufacturing test patterns per core and its manufacturing cost ρ;
- the ratio between the burn-in cost per core and its manufacturing cost ξ;

Determine the number of burn-in driven spares t to achieve the minimum production cost per sold chip under product quality constraint and the associated burn-in time T.

Next, we consider to introduce n redundant cores to not only enable partial/no burn-in test but also relax the defect coverage for core tests in our test economics model. The imperfect manufacturing test process leads to both *test escapes* (i.e., bad chips pass the test) and false rejects (i.e., good chips fail the test, also known as *test overkill*), which is related to the effectiveness of the test decision criterion [20]. Generally speaking, the more patterns applied to the circuits (for higher defect coverage), the more false rejects occur. Therefore, by relaxing defect coverage, we can also achieve cost savings with less test overkills and this effect is considered in our model. Because some defects do not reveal themselves due to insufficient burn-in and some defects are not detected because of low test defect coverage, *n test cost-driven spares* are used to ensure the product quality and reliability for no less than m defect-free cores on-chip at the end of infant mortality period. In addition, we consider to have *s yield-driven spares* placed on-chip for yield enhancement. Consequently, we have totally $w = m+n+s$ homogeneous cores on-chip and only those chips containing no less than $m+n$ pass-test cores are shipped to the market.

Problem 2 [Partial Burn-In & Relaxed Defect Coverage]: Given all parameters as specified in Problem 1, except that the test-related parameter set becomes

- *test-related parameters*
- the effectiveness of the test decision criterion μ;
- the parameter v that sets the steepness of the fallout curve that shows the probability for defects to be detected when test patterns are gradually applied [20] (discussed later);
- the product quality requirement for the maximum test escape percentage τ (e.g., $\tau = 0.0005$ for 500DPPM);

Determine the number of test cost-driven spares n, the number of yield-driven redundant cores s, the defect coverage for core test B_r, and the burn-in time T such that the production cost per sold chip is minimized under the product quality and reliability constraints.

It should be noted that the cost for online testing used to identify defective cores in-field is not considered in our test economics model. This is because, it is essential to include circuitries for online testing in future manycore systems despite of whether test cost-driven redundancies are available or not, in order to detect not only defects introduced during manufacturing process but also soft errors and wearout-related effects occurred over the ICs' useful lifetime.

3 TEST ECONOMICS WITH PARTIAL/NO BURN-IN

In this section, we present our analytical model that captures the impact of introducing burn-in driven redundancy for partial/no burn-in on the total production cost of homogeneous many systems.

3.1 The Impact of Partial Burn-In

Traditional burn-in testing typically aims to achieve sufficient high defect coverage. Thus, similar to prior work, we assume that all latent defects reveal themselves after *full* burn-in time T_{IM}. It is possible that *partial* burn-in is performed, that is, the burn-in time T is shorter than T_{IM}. In order to examine the impact of partial burn-in on test economics, we need to know the percentage of the products passing burn-in test given the average latent defect quantity per core λ_L, which is equivalent to the probability that some latent defects of an IC product reveal themselves and referred to as the reliability induced by latent defects. Apparently, it is a function of burn-in time T and hence denoted by $R(T)$. The reliability function with respect

to time has been shown to follow Weibull failure distribution[1] with shape parameter β [1].

Since all latent defects reveal themselves after full burn-in, the reliability $R(T_{IM})$ is simply $\exp(-\lambda_L)$. With this relation and the linearly relation between killer defects and latent defects, we can express the reliability with partial burn-in time T as

$$R(T) = \exp\Big(-\lambda_L\big(\frac{T}{T_{IM}}\big)^{\beta}\Big) = \exp\Big(-\gamma\cdot\lambda_K\big(\frac{T}{T_{IM}}\big)^{\beta}\Big) \quad (1)$$

3.2 Product Quality and Chip Test Yield

To meet the product quality requirement, the probability that a sold chip actually functions (i.e., contains no less than m good cores) after all early infant mortality failures have been revealed should be higher than a threshold $(1-\tau)$. This conditional probability (given a chip is sold) is referred as product quality Q hereafter. Recall that a chip will be sold when all $(u=m+t)$ cores pass test. Let us use X_1 to represent the event that no less than m cores on a chip is defect-free at the end of infant mortality, and use X_2 to denote the event that all u cores on a chip pass manufacturing test after burn-in time duration T, the product quality is essentially the conditional probability of event X_1 given event X_2 occurs, i.e.,

$$Q = \Pr\{X_1|X_2\} = \frac{\Pr\{X_1X_2\}}{\Pr\{X_2\}} \quad (2)$$

Let us start with the calculation of $\Pr\{X_2\}$. Fig. 2 depicts several critical core sets. Among all u fabricated cores on a chip, there are j defect-free cores before burn-in process $(j \le u)$. Note that, probably some contain latent defects but the defects are not severe enough at this stage to cause failures. In other words, there are j cores without any killer defects on the chip. After partial burn-in with duration T, only i cores in that set remains defect-free $(i \le j)$. A core in this set, similarly, may contain latent defects since the burn-in time T can be less than T_{IM}.

To capture such complicated relations, we define a series of events. Let $M_{j,u}$ be the event that j-out-of-u cores are initially defect-free. Apparently, it is independent of burn-in time duration T. Also, we use $N_{i,j,T}$ to indicate that i cores in that set maintain defect-free after burn-in time T. When both events occur, there will be i defect-free cores out of u fabricated ones after partial burn-in. Thus, the event that a chip containing u cores passes manufacturing test given it contains i defect-free cores can be expressed as $[X_2|M_{j,u}\cap N_{i,j,T}]$. We assume that the occurrence of killer defects and latent defects are independent. By the theorem of total probability, we obtain

$$\Pr\{X_2\} = \sum_{j=0}^{u}\sum_{i=0}^{j}\Pr\{X_2|M_{j,u}\cap N_{i,j,T}\}\Pr\{M_{j,u}\}\Pr\{N_{i,j,T}\} \quad (3)$$

wherein $\Pr\{M_{j,u}\}$ can be derived according to negative-binomial defect distribution [11].

[1] Weibull failure distribution has reliability function $R(T) = \exp(-(\frac{T}{\theta})^{\beta})$, where θ and β are the scale parameter and shape parameter respectively.

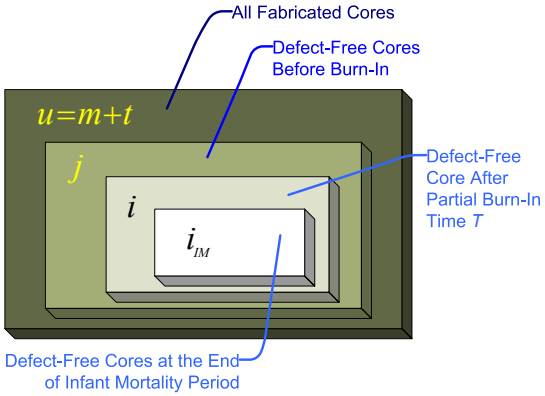


Figure 2. Defect-Free Core Sets at Various Time Points.

As mentioned before, the reliability induced by defects can be characterized by Weibull distribution. Specifically, within infant mortality period, the shape parameter $\beta \in [0,1]$ and thus the failure rate decreases with time. $\Pr\{N_{i,j,T}\}$ can be expressed in terms of reliability function $R(T)$ that is defined by Eq. (1):

$$\Pr\{N_{i,j,T}\} = \binom{j}{i}R^i(T)\big(1-R(T)\big)^{j-i} \quad (4)$$

Next, let us consider the calculation of the conditional probability $\Pr\{X_2|M_{j,u}\cap N_{i,j,T}\}$. For the sake of simplicity, we assume no false rejects (i.e., all good cores pass manufacturing test) for the time being (this assumption will be lifted later). Thus, all i defect-free cores after (insufficient) burn-in pass test. Note that, it is possible that some cores in this set contain unrevealed latent defects. Due to imperfect manufacturing test, q cores with revealed defects (including killer defects and revealed latent defects) out of $(u-i)$ also pass test, while the remaining $(u-i-q)$ are rejected. Since a chip is shipped to customers if all its $(m+t)$ cores pass test, we have $q = u-i$. Therefore, denoting by B_r the defect coverage of manufacturing test, we have

$$\Pr\{X_2|M_{j,u}\cap N_{i,j,T}\} = (1-B_r)^{u-i} \quad (5)$$

We then move to the computation of $\Pr\{X_1X_2\}$, which is more complicated. To be specific, this value is related to the situation of cores at the end of infant mortality period. Remind that i cores maintain defect-free after partial burn-in. Among them, some may still contain unrevealed latent defects. As a result, the number of cores retain defect-free at the end of infant mortality period can be less than i (i.e., $i_{IM} \le i$), as shown in Fig. 2.

Denoting by $D_{\{i,T\},\{i_{IM},T_{IM}\},u}$ the event that i-out-of-u cores do not contain revealed defects after burn-in time T and i_{IM} cores are defect-free at the end of infant mortality. As X_1 indicates no less than m cores on a chip are defect-free at T_{IM}, i_{IM} should be no less than m. In addition, because $T \le T_{IM}$, for the same chip we always have $i \ge i_{IM}$. By using the total probability theory, we have

$$\Pr\{X_1X_2\} = \sum_{i=m}^{u}\sum_{i_{IM}=m}^{i} \Pr\{X_1X_2|D_{\{i,T\},\{i_{IM},T_{IM}\},u}\} \cdot \cdot \Pr\{D_{\{i,T\},\{i_{IM},T_{IM}\},u}\} \quad (6)$$

Under the condition $D_{\{i,T\},\{i_{IM},T_{IM}\},u}$, by the same argument as Eq. (5), we express the conditional probability that all $(u-i)$ defective cores pass test after burn-in time T as $(1-B_r)^{u-i}$.

The event $D_{\{i,T\},\{i_{IM},T_{IM}\},u}$ can be divided into two sub-events: $M_{j,u}$, that has been introduced; and $P_{\{i,T\},\{i_{IM},T_{IM}\},j}$, meaning that i cores contain no revealed defects after burn-in time T (i.e., event $N_{i,j,T}$) and then i_{IM} cores are eventually defect-free at the end of infant mortality (i.e., event $N_{i_{IM},j,T_{IM}}$). Since j must be no less than i and $i \geq m$, we get $j \geq m$. That is,

$$\begin{aligned}\Pr\{D_{\{i,T\},\{i_{IM},T_{IM}\},u}\} &= \sum_{j=m}^{u} \Pr\{M_{j,u}\} \Pr\{P_{\{i,T\},\{i_{IM},T_{IM}\},j}\} \\ &= \sum_{j=m}^{u} \Pr\{M_{j,u}\} \Pr\{N_{i,j,T} \cap N_{i_{IM},j,T_{IM}}\}\end{aligned} \quad (7)$$

Note that, the event that i_{IM} cores are defect-free at T_{IM} is not independent of the event that i cores contain no defects at T, where $T \leq T_{IM}$. By the multiplication rule, we are able to obtain

$$\Pr\{N_{i,j,T} \cap N_{i_{IM},j,T_{IM}}\} = \Pr\{N_{i,j,T}\} \Pr\{N_{i_{IM},j,T_{IM}}|N_{i,j,T}\} \quad (8)$$

where, $\Pr\{N_{i,j,T}\}$ has been expressed by Eq. (4). To compute the conditional probability $\Pr\{N_{i_{IM},j,T_{IM}}|N_{i,j,T}\}$, we start from the conditional reliability of a single core given it contains no revealed defects after partial burn-in. With Eq. (1), it is given by

$$R_c(T_{IM}|T) = \frac{R(T_{IM})}{R(T)} = \exp\Big(-\gamma \cdot \lambda_K \cdot \frac{T_{IM}^{\beta} - T^{\beta}}{T_{IM}^{\beta}}\Big) \quad (9)$$

With this notation, the conditional probability can be computed by

$$\Pr\{N_{i_{IM},j,T_{IM}}|N_{i,j,T}\} = \binom{i}{i_{IM}} R_c^{i_{IM}}(T_{IM}|T)\big(1-R_c(T_{IM}|T)\big)^{i-i_{IM}} \quad (10)$$

By using these equations, we conclude that

$$\Pr\{X_1X_2\} = \sum_{j=m}^{u}\sum_{i=m}^{j}\sum_{i_{IM}=m}^{i} (1-B_r)^{u-i} \cdot \Pr\{M_{j,u}\} \Pr\{N_{i,j,T}\} \cdot \cdot \Pr\{N_{i_{IM},j,T_{IM}}|N_{i,j,T}\} \quad (11)$$

Substituting Eq. (3) and (11) into (2) results in the expression for product quality. Note that, it should be no less than the predefined threshold $(1-\tau)$, namely, $Q \geq 1-\tau$.

Test yield indicates the probability that no less than $(m+t)$ cores on the manycore chips pass test[2], i.e., $Y_{test} = \Pr\{X_2\}$. These chips will be shipped to customers as quality products.

[2]Test yield reveals the influence of test quality on manufactured chips, while true yield, having been well studied in the literature, reflects the probability that less than m cores on a chip contain defects. True yield in our case can be simply computed as $Y_{true} = \sum_{i=m}^{u} \Pr\{C_{i,u,T_{IM}}\}$.

With this analytical model, we first consider full burn-in case, that is, setting $T = T_{IM}$. As no burn-in driven redundancy is introduced, t is set to zero. Then, we gradually reduce T, with which some latent defects do not reveal themselves before functioning in-field and then cannot be detected during the manufacturing test process. Without introducing burn-in driven redundancy, the product reliability decreases with the reduction of burn-in time. To meet product quality constraint, some burn-in driven redundant cores need to be introduced on-chip.

3.3 Proposed Cost Model

Various cost models for IC manufacturing and test have been presented in the literature (e.g., [15–17, 28]). Different from the above models that involve a great amount of manufacturing and test parameters, we present a simple yet effective cost model to capture the key impact of introducing burn-in driven redundancy into homogeneous manycore systems. That is, instead of obtaining concrete values for different cost factors, the input to our model is unified ratio parameters among various cost factors.

The production cost per sold chip[3] can be calculated by the following equation to evaluate different redundancy configurations

$$C_{prod}^{chip} = \frac{(C_{manu}^{core} + C_{ATE}^{core} + C_{burn-in}^{core}) \cdot (m+t)}{Y_{test}} \quad (12)$$

where, C_{manu}^{core}, C_{ATE}^{core}, and $C_{burn-in}^{core}$ indicate the manufacturing cost, ATE cost for applying test patterns, and burn-in cost per fabricated core, respectively. Note that, test cost includes both ATE cost C_{ATE}^{core} and burn-in cost $C_{burn-in}^{core}$.

3.3.1 Manufacturing Cost

We set the manufacturing cost of each core for manycore chips without redundancy to be *1 unit* and we normalize manufacturing cost for chips with redundancy to this base value accordingly.

Manufacturing cost per fabricated chip can be simplified as fabrication cost per wafer F divided by the gross die per wafer N. We use superscript *wo* to distinguish the "without redundancy" case from the "with redundancy" case. Given all $(u = m+t)$ cores of a chip are fabricated on the same die, N can be modeled as a function of the number of on-chip cores [7]

$$N = \frac{\pi(d/2)^2}{A_r \cdot u} - \frac{\pi \cdot d}{\sqrt{2 \cdot A_r \cdot u}} \quad (13)$$

where, A_r is the area of each core, d is the dimension of the wafer. We obtain the manufacturing cost per fabricated chip C_{manu}^{chip} accordingly.

As stated earlier, we normalize the manufacturing cost of each on-chip core for manycore systems without redundancy to be unit 1. That it, we first substitute the condition $t = 0$

[3]Other cost factors (e.g., R&D cost) are excluded without loss of the model's accuracy as they are independent to the redundancies introduced on-chip.

into Eq. (13) to achieve the number of cores on a chip without redundancy N^{wo}. Then, similar to the computation for "with redundancy" case, the manufacturing cost per chip without redundancy $C_{manu}^{chip,wo}$ can be expressed as dividing fabrication cost per wafer F by N^{wo}. Since there are m cores on a chip in total, the manufacturing cost per core in this "without redundancy" case is given by

$$C_{manu}^{core,wo} = \frac{C_{manu}^{chip,wo}}{m} = \frac{F}{N^{wo} \cdot m} \tag{14}$$

With this equation, we set $C_{manu}^{core,wo} = 1$ to obtain the normalized fabrication cost per wafer F (measured in units) and finally the manufacturing cost per fabricated core of the chip in the "with redundancy" case:

$$C_{manu}^{core} = \frac{C_{manu}^{chip}}{u} = \frac{F}{N \cdot u} = \frac{N^{wo} \cdot m}{N \cdot u} \tag{15}$$

3.3.2 ATE Cost

As mentioned before, the defect coverage of manufacturing test has been assumed to be a fixed value B_r. Recall that the manufacturing cost per core in "without redundancy" case is normalized to be unit 1. We simply set the ATE cost per fabricated core as ρ units, where ρ is a unified ratio parameter between the ATE cost and the manufacturing cost. That is,

$$C_{ATE}^{core} = \rho \cdot C_{manu}^{core,wo} \tag{16}$$

3.3.3 Burn-in Cost

The burn-in cost is assumed to be proportional to the burn-in time T. Thus, by normalizing the cost of fully burn-in process as ξC_{manu}^{core}, we model the burn-in cost with duration T as

$$C_{burn-in}^{core} = \xi \cdot C_{manu}^{core,wo} \cdot \frac{T}{T_{IM}} \tag{17}$$

3.4 Case Study for Partial/No Burn-In

Consider a homogeneous manycore system that functions with no less than 32 defect-free cores (i.e., $m = 32$). To meet the product quality requirement $\tau = 500DPPM$ for the manycore system without any redundancy given full burn-in process, we set defect coverage for core test as a sufficient high value 99.9%. In addition, the killer defect density $\lambda_K = 0.05$, latent-to-killer defect density ratio $\gamma = 0.02$, ATE cost ratio $\rho = 10\%$, and burn-in cost ratio $\xi = 20\%$. Other parameter setups are provided in Section 5.1. The experimental results for this case study are shown in Fig. 3.

When we perform full burn-in process with adequately high defect coverage, it is not necessary to introduce any burn-in driven redundancy (i.e., $t = 0$). With the shortening of burn-in time T, introducing more burn-in driven redundant cores becomes a must for meeting the product quality requirement. To be specific, when T is in the range from $90\%T_{IM}$ to $10\%T_{IM}$

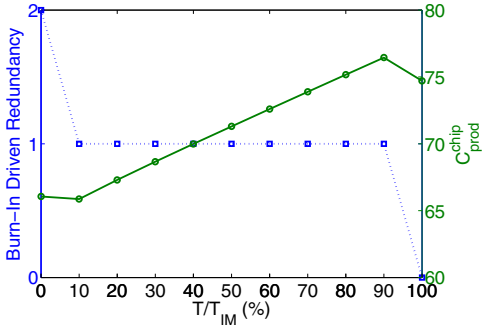


Figure 3. Production Cost with Partial/No Burn-In Test.

one burn-in driven redundant core is enough, while if no burn-in test is provided (i.e., $T = 0$) the system needs one more redundant core to guarantee product quality (See the dotted line).

If we introduce one burn-in driven redundancy but not reduce burn-in time much (i.e., $T = 90\%$ or $80\%T_{IM}$), the production cost does not decrease. This is because the manufacturing cost increment caused by burn-in driven redundancy exceeds the burn-in cost reduction. But if the burn-in time is further shortened, we observe significant benefits in terms of total cost. In this experiment, the minimum production cost is achieve at $T = 10\%T_{IM}$. The cost reduction compared with full burn-in case is close to 10%. When the burn-in time drops from $T = 10\%T_{IM}$ to 0, more redundant cores are introduced (i.e., t increases). As a result, the increment of manufacturing cost and ATE cost per fabricated chip is 1.1 units, while the burn-in cost reduction is only 0.66, which increases the total production cost.

4 TEST ECONOMICS WITH PARTIAL MANUFACTURING TEST

Similar to partial burn-in process, insufficient manufacturing test can lead to product quality decrease but the quality loss could be recovered by introducing some redundant cores. As can be observed in Section 3.4, if the test cost reduction exceeds the manufacturing cost increment, the total production cost can be reduced. We therefore study the impact of introducing redundancy for relaxed defect coverage requirement. As both partial/no burn-in and partial manufacturing test result in product quality decrease and the corresponding spares are used for recouping the product quality loss, we do not distinguish these two types of redundancy deliberately in the rest of this paper. That is, instead of representing them separately, we use n to denote the total number of test cost-driven redundant cores.

4.1 The Impact of Test Decision Criterion

The primary objective of manufacturing test is to obtain low test escapes in order to ensure the quality of the shipped products, and a limited number of false rejects are considered as acceptable loss. With the ever advancement in semiconductor technology, however, it has been reported that the number of false rejects has dramatically increased [22, 26]. The associated test yield loss (i.e., the fraction of chips that fail manufacturing tests but would work in application) may have signifi-

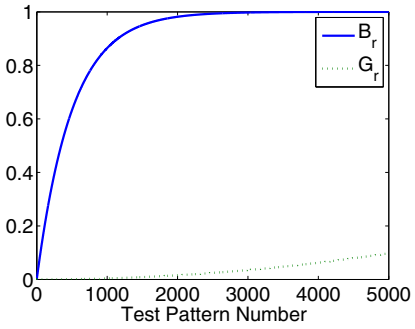


Figure 4. Test Escapes vs. False Rejects.

cant impact on manufacturing cost. We therefore examine the influence of false rejects and test escape in this section.

To capture the above effects, let us use G_a, G_r, B_a, B_r to denote the conditional probability that a defect-free core passes test, that a defect-free core is rejected, that a core containing defects escapes from the manufacturing test, and that a bad core is rejected, respectively. Apparently, $G_a + G_r = 1$ and $B_a + B_r = 1$. According to [5], the fraction of bad cores to be detected after applying k test patterns can be modeled as

$$B_r = 1 - e^{-\nu k} \tag{18}$$

Also, depending on the effectiveness of the decision criterion μ, the correlation between B_r and G_r can be expressed as [20]

$$B_r = 1 - e^{-\mu\sqrt{G_r}} \tag{19}$$

Combining Eq. (18) and Eq. (19), we can express G_r, the probability for false rejects, in terms of the number of applied test patterns k as

$$G_r = (\frac{\nu k}{\mu})^2 \tag{20}$$

Ideally, a prefect manufacturing test is able to reject all bad cores while accept all defect-free ones, i.e., $B_r \equiv 1$ and $G_r \equiv 0$. It can be achieved by taking the limit of the above equations as μ goes to ∞. In reality, because of various challenges in decision-making (e.g., the overlap between the good and the bad populations [20]), μ is a finite value. Fig. 4 shows a typical G_r and B_r versus the test pattern count, where the decision criterion μ is set to be 32 and ν is set to be 0.002 [5]. Generally speaking, the better the decision method, the more square the plot of B_r versus G_r.

4.2 Product Quality with False Rejects

With test cost-driven redundancy, the problem comes down to determine s and n values such that the production cost for sold chips is minimized under the product quality constraint. In total $(w = m+n+s)$ homogeneous cores are fabricated on the chip. Among them, if no less than $(m+n)$ cores pass test after burn-in time T, the chip will be shipped to the market. We need to guarantee that the probability that a sold chip contains no less than m defect-free cores at the end of infant mortality is higher than the given threshold τ. We therefore redefine X_2 as the event that no less $(m+n)$ cores among all w cores pass the partial manufacturing test given burn-in time T, denoted by $\widetilde{X}_2$. Again, by the total probability theorem, we compute $\Pr\{\widetilde{X}_2\}$ in a divide-and-conquer manner, that is,

$$\Pr\{\widetilde{X}_2\} = \sum_{j=0}^{w}\sum_{i=0}^{j} \Pr\{\widetilde{X}_2 | M_{j,w} \cap N_{i,j,T}\} \Pr\{M_{j,w}\} \Pr\{N_{i,j,T}\} \tag{21}$$

In this equation, $\Pr\{N_{i,j,T}\}$ has been defined by Eq. (4) and $\Pr\{M_{j,w}\}$ can be computed according to negative-binomial defect distribution [11]. Given both $M_{j,w}$ and $N_{i,j,T}$ occur (that is, i-out-of-w cores are defect-free after burn-in time T), taking both false reject and test escape into account, event $\widetilde{X}_2$ is the union of a series of mutually exclusive events $[A_{p,i} \cap B_{q,w-i}]$, meaning that exactly p good cores and q bad cores pass test, where $A_{p,i}$ represents the event that among i good cores on a chip, p pass the test while $(i-p)$ fail the test, and $B_{q,w-i}$ is the event that q-out-of-$(w-i)$ bad cores pass test. To explore all possible combinations for $[\widetilde{X}_2 | M_{j,w} \cap N_{i,j,T}]$ event to be true, we need to determine all possible values for p and q. Apparently, due to false rejects and the fact that the number of good cores that pass test cannot exceed that of good cores i, p can be $0, \cdots, i$. As for q, a sold chip should satisfy two conditions: (i) $q = 0, \cdots, w-i$; (ii) $p+q \geq m+n$, that is, only the chips contain no less than $(m+n)$ pass-test cores are sold. We therefore have $q = \max\{0, m+n-p\}, \cdots, w-i$. For the ease of discussion, let $\omega \equiv \max\{0, m+n-p\}$. Based on the above, we have

$$\Pr\{\widetilde{X}_2 | M_{j,w} \cap N_{i,j,T}\} = \sum_{p=0}^{i}\sum_{q=\omega}^{w-i} \Pr\{A_{p,i} \cap B_{q,w-i} | M_{j,w} \cap N_{i,j,T}\} \tag{22}$$

Since the two events $A_{p,i}$ and $B_{q,j}$ are conditionally independent given both $M_{j,w}$ and $N_{i,j,T}$ occur, we have [4]

$$\begin{aligned}\Pr\{A_{p,i} \cap B_{q,w-i} | M_{j,w} \cap N_{i,j,T}\} &= \Pr\{A_{p,i} | M_{j,w} \cap N_{i,j,T}\} \cdot \\ &\cdot \Pr\{B_{q,w-i} | M_{j,w} \cap N_{i,j,T}\}\end{aligned} \tag{23}$$

Combining Eq. (21), Eq. (22), and Eq. (23) results in

$$\begin{aligned}\Pr\{\widetilde{X}_2\} = \sum_{j=0}^{w}\sum_{i=0}^{j}\sum_{p=0}^{i}\sum_{q=\omega}^{w-i} &\Pr\{A_{p,i} | M_{j,w} \cap N_{i,j,T}\} \cdot \\ &\cdot \Pr\{B_{q,w-i} | M_{j,w} \cap N_{i,j,T}\} \Pr\{M_{j,w}\} \Pr\{N_{i,j,T}\}\end{aligned} \tag{24}$$

To calculate $\Pr\{A_{p,i} | M_{j,w} \cap N_{i,j,T}\}$ and $\Pr\{B_{q,w-i} | M_{j,w} \cap N_{i,j,T}\}$, for the sake of simplicity, we assume that the event that good cores rejected by test and that bad cores accepted by test follow Poisson distribution [11]. This assumption is reasonable because of the "clustering" effect of defects. Therefore, we have

$$\Pr\{A_{p,i} | M_{j,w} \cap N_{i,j,T}\} = \binom{i}{p} (1-G_r)^p \cdot G_r^{i-p} \tag{25}$$

and

$$\Pr\{B_{q,w-i}|M_{j,w}\cap N_{i,j,T}\} = \binom{w-i}{q}(1-B_r)^q \cdot B_r^{w-i-q} \quad (26)$$

where G_r and B_r are the functions of the applied number of test patterns k (see Eq. (18) and Eq. (20)). Note that, by the definition, $\Pr\{\widetilde{X}_2\}$ can also be viewed as the test yield of fabricated chips (i.e., $\widetilde{Y}_{test}$).

To compute the product quality $\widetilde{Q}$, it is also necessary to redefine Eq. (11). By similar argument, we have

$$\begin{aligned}\Pr\{X_1\widetilde{X}_2\} = \sum_{j=m}^{w}\sum_{i=m}^{j}\sum_{i_{IM}=m}^{i}\sum_{p=0}^{i}\sum_{q=\omega}^{w-i}\Pr\{A_{p,i}|M_{j,w}\cap N_{i,j,T_{IM}}\}\cdot \\ \cdot\Pr\{B_{q,w-i}|M_{j,w}\cap N_{i,j,T_{IM}}\}\Pr\{M_{j,w}\}\cdot \\ \cdot\Pr\{N_{i,j,T}\}\Pr\{N_{i_{IM},j,T_{IM}}|N_{i,j,T}\} \quad (27)\end{aligned}$$

and therefore the product quality constraint is written in terms of $\Pr\{X_1\widetilde{X}_2\}$ and $\Pr\{\widetilde{X}_2\}$ accordingly.

4.3 Cost Model

With the analytical model of product quality and test yield, we move to discuss the impact of partial manufacturing test and test cost-driven redundancy on the total production cost. The total production cost for the manycore system is given by

$$\widetilde{C}_{prod}^{chip} = \frac{(C_{manu}^{core} + \widetilde{C}_{ATE}^{core} + C_{burn-in}^{core})\cdot(m+n+s)}{\widetilde{Y}_{test}} \quad (28)$$

In the above equation, C_{manu}^{core} and $C_{burn-in}^{core}$ remain the same as derived in Section 3.3.

For $\widetilde{C}_{ATE}^{core}$, the ATE cost per fabricated core is determined by the number of test patterns applied on ATE, which is constrained by the test quality requirement. We calculate this value as follows.

Given the ATE cost ratio parameter ρ, we set the ATE cost per fabricated core with an arbitrary defect coverage η as ρ units. The actual ATE cost, since its corresponding defect coverage may not be η, is normalized to the reference case. To be specific, we first compute the test pattern count for achieving defect coverage η by Eq. (18) as

$$k_\eta = \frac{\ln(1-\eta)}{-v} \quad (29)$$

Thus, we obtain the normalized average cost for applying a single test pattern, that is,

$$\widetilde{C}_{ATE}^{pattern} = \frac{\rho\cdot C_{manu}^{core,wo}}{k_\eta} = \frac{\rho}{k_\eta} \quad (30)$$

Similar to Eq. (29), the test pattern count for achieving defect coverage B_r is given by

$$k = \frac{\ln(1-B_r)}{-v} \quad (31)$$

The ATE cost for each core is therefore

$$\widetilde{C}_{ATE}^{core} = \widetilde{C}_{ATE}^{pattern}\cdot k \quad (32)$$

5 EXPERIMENTAL RESULTS

5.1 Experimental Setup

To evaluate the effectiveness of the proposed strategy, we perform extensive experiments for the production cost of a homogeneous manycore system that functions with no less than 32 defect-free cores (i.e., $m = 32$), varying the number of test cost-driven redundancy n and burn-in time T. In our work, the best n, T and s combination in terms of production cost is determined by exploring the possible n, T and s solution space. Note, this is not a time-consuming process because the computation time for each single configuration is quite small. Also, a large value of n and/or s will increase the production cost due to the extra silicon area and hence the possible combinations to be explored are quite limited.

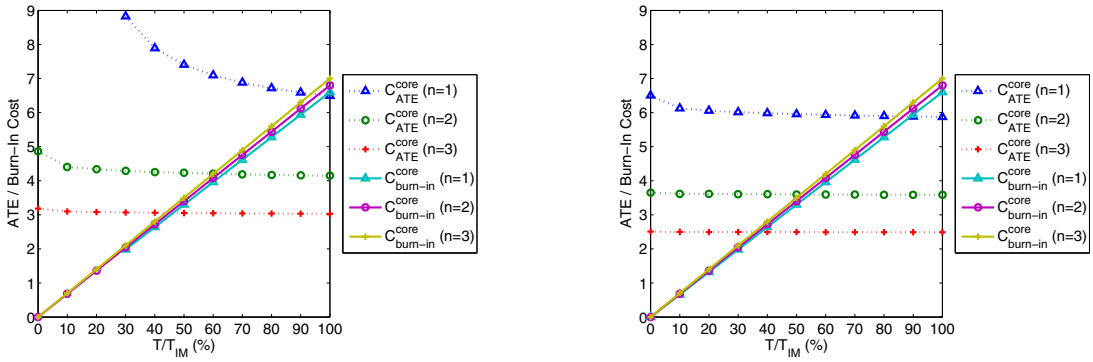
We set the system parameters based on prior work as follows: $v = 0.002$ [5], $\mu = 18$ [20], $\alpha = 0.3$ [13], $\beta = 0.3$, $\xi = 0.2$ [1], $d = 300mm$, $A_r = 10mm^2$, $\eta = 95\%$. The product quality requirement is set to $500DPPM$ (i.e., $\tau = 5\times10^{-4}$).

5.2 Results and Discussion

First of all, we demonstrate the tradeoff between burn-in cost and ATE cost under the product quality constraint. As both partial burn-in process and partial manufacturing test may scarify some product quality, we introduce a few test cost-driven redundant cores to recap this loss. From another point of view, given certain test cost-driven spares, to meet the same product quality requirement, if we shorten the burn-in time, the test coverage must be increased and hence the ATE cost increases if we save burn-in cost. Fig. 5(a) illustrates this trend for the cases when $n = 1$, $n = 2$, and $n = 3$, with high defect density ($\lambda = 0.05$, $\gamma = 0.05$) and high ATE cost ratio ($\rho = 20\%$). No yield-driven spares are introduced in this experiment (i.e., $s = 0$).

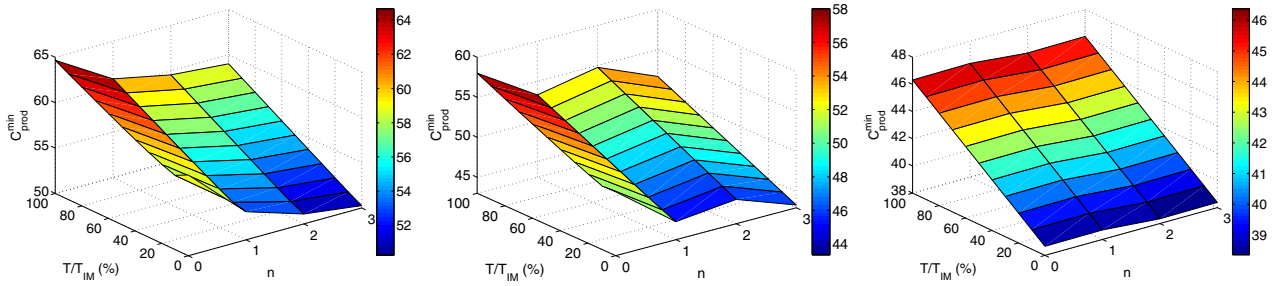
As shown in this figure, with the shrinking of burn-in time, the ATE cost increases dramatically when $n = 1$, but if more redundant cores are added (that is, $n = 2$ and $n = 3$), the ATE cost only increases slightly. This is because, the responsibility for manufacturing test can be significantly relaxed by introducing more than one test cost-driven redundant cores, in spite of the large number of fabricated cores on a chip ($m = 32$). Thus, although a great percentage of latent defects cannot be revealed because of partial burn-in test, we still do not need very high manufacturing test coverage. However, if only one test cost-driven redundancy is involved (i.e., $n = 1$), the product quality requirement is not relaxed much. As a result, we need to trade a small percentage of burn-in time reduction with a significant test pattern count increment. Another interesting observation is that, with the decrease of burn-in time, ATE cost grows sharply. We attribute this phenomenon to the slowdown of failure rate decrease with respect to the last of burn-in time. Also, we observe that if the burn-in time is shorter than $30\%T_{IM}$, we cannot satisfy the product quality constraint by increasing manufacturing test coverage.

Fig. 5(b) shows the results when the defect density is relatively low ($\lambda = 0.02$ and $\gamma = 0.02$). In this case, the test pattern



(a) $\lambda = 0.05, \gamma = 0.05$ (b) $\lambda = 0.02, \gamma = 0.02$

Figure 5. The Tradeoff between ATE Cost and Burn-In Cost.



(a) $\lambda = 0.05, \gamma = 0.05, \rho = 20\%$ (b) $\lambda = 0.02, \gamma = 0.02, \rho = 10\%$ (c) $\lambda = 0.01, \gamma = 0.01, \rho = 10\%$

Figure 6. Minimum Production Cost.

count and hence the ATE cost increment with respect to burn-in time decline is very small. We therefore tend to achieve better results in terms of production cost by reducing burn-in time. If the ATE cost ratio is even lower than $\rho = 20\%$, since the test pattern count does not vary with ρ, more benefits can be obtained. This observation can also be used to explain the phenomenon shown in Fig. 3.

Next, we compare the traditional strategy that does not introduce any test cost-driven redundancy (i.e., $n = 0$) but includes a few yield-driven spares (s can be more than zero), and the proposed approach with both redundancies, for various burn-in time T. For fair comparison, given T, and n, we vary s to find the minimum production cost per sold chip, and record this value and the corresponding s.

Fig. 6(a) shows the minimum production cost given various burn-in time T and test cost-driven redundancy n, for killer defect density $\lambda = 0.05$, latent-to-killer defect density ratio $\gamma = 0.05$, and ATE cost ratio $\rho = 20\%$. Introducing test cost-driven spares results in significant cost reduction when compared with the traditional approach with full burn-in time without such redundant cores. In particular, the maximum cost reduction is as high as 22.28%, obtained when $n = 3$ and $T = 0$. With the shrinking of burn-in time, the production cost gradually decreases. We take a closer observation and consider $n = 1$ case as an example. In this case, in spite of the variation of burn-in time T, the number of yield-driven redundant cores s that results in the minimum production cost remains the same value. Thus, the manufacturing cost does not increase with the decrease of burn-in time. Since the burn-in cost reduction is more significant than the ATE cost increment, the total production cost keeps declining. A special case is observed when $n = 0$. When the burn-in time drops from $T = 10\%T_{IM}$ to 0, the total production cost increases because of the diminishing test yield and the increment of ATE cost.

We observe even more production cost reduction when $\lambda = 0.02$ and $\gamma = 0.02$ (see Fig. 6(b)). We achieve up to 25.26% production cost reduction by using three test cost-driven spares (i.e., $n = 3$) and no burn-in test. This is mainly because, due to the low defect density in this experimental setup, the test coverage requirement and the associated ATE cost do not increase much with the shortening of burn-in time (see Fig. 5(b)). Therefore, with the decrease of T, the test cost (including both ATE cost and burn-in cost) reduces dramatically.

Fig. 6(c) presents a case that the minimum production cost occurs when two test cost-driven redundant cores are involved (i.e., $n = 2$). For certain burn-in time T, only modest variation in terms of production cost can be observed. This is because when n increase from 0 to 1, and then to 2, the number of fabricated cores on a chip (i.e., $m+n+s$) in these cases remains the same. Thus, the manufacturing cost and burn-in cost per chip are fixed. Since the defect density is quite low ($\lambda = 0.01$ and $\gamma = 0.01$), there is no significant ATE cost reduction with respect to the product quality relaxation. If more test cost-driven redundant cores are introduced, as it might result in the increment of number of fabricated cores on chip, we cannot obtain more benefits.

6 CONCLUSION

In homogeneous manycore systems, by introducing spare cores on-chip, the burn-in test time can be shortened and the defect coverage requirements for core tests can be also relaxed, without sacrificing quality of the shipped products. This paper presents novel analytical models that capture the above tradeoff and the effectiveness of the proposed strategy is verified for hypothetical manycore systems with various configurations.

7 Acknowledgements

This work was supported in part by the General Research Fund CUHK417406, CUHK417807, and CUHK418708 from Hong Kong SAR Research Grants Council (RGC), in part by National Science Foundation of China (NSFC) under grant No. 60876029, in part by a grant N_CUHK417/08 from the NSFC/RGC Joint Research Scheme, and in part by the National High Technology Research and Development Program of China (863 program) under grant no. 2007AA01Z109.

References

[1] T. S. Barnett and A. D. Singh. Relating Yield Models to Burn-In Fall-Out in Time. In *Proceedings IEEE International Test Conference (ITC)*, pages 77–84, 2003.

[2] P. S. Bhojwani and R. N. Mahapatra. A Robust Protocol for Concurrent On-Line Test (COLT) of NoC-based Systems-on-a-Chip. In *Proceedings ACM/IEEE Design Automation Conference (DAC)*, pages 670–675, 2007.

[3] Cisco. Cisco and IBM Collaborate to Design and Build World's Most Sophisticated, High-Performance 40Gbps Custom Chip. http://newsroom.cisco.com/dlls/partners/news/2004/pr_prod_06-09.html.

[4] A. P. Dawid. Conditional independence in statistical theory. 41(1):1–31, 1979.

[5] F.-F. Ferhani, N. R. Saxena, E. J. McCluskey, and P. Nigh. How Many Test Patterns are Useless? In *Proceedings IEEE VLSI Test Symposium (VTS)*, pages 23–28, 2008.

[6] D. Geer. Chip Makers Turn to Multicore Processors. *IEEE Computer*, 38(5):11–13, May 2005.

[7] J. L. Hennessy and D. A. Patterson. In *Computer Architecture: A Quantitative Approach*. Morgan Kaufmann, fourth edition, 2006.

[8] L. Huang and Q. Xu. Is It Cost-Effective to Achieve Very High Fault Coverage for Testing Homogeneous SoCs with Core-Level Redundancy? In *Proceedings IEEE International Test Conference (ITC)*, poster 18, 2008.

[9] L. Huang, F. Yuan, and Q. Xu. Lifetime Reliability-Aware Task Allocation and Scheduling for MPSoC Platforms. In *Proceedings Design, Automation, and Test in Europe (DATE)*, 2009.

[10] I. Koren, Z. Koren, and C. H. Stapper. A unified negative-binomial distribution for yield analysis of defect-tolerant circuits. *IEEE Transactions on Computers*, 42(6):724–733, June 1993.

[11] I. Koren and C. H. Stapper. Yield Models for Defect-Tolerant VLSI Circuits: A Review. In *Proceedings International Workshop on Defect and Fault Tolerance in VLSI Systems*, pages 1–22, 1988.

[12] N. Kranitis, A. Merentitis, N. Laoutaris, and G. Theodorou. Optimal Periodic Testing of Intermittent Faults In Embedded Pipelined Processor Applications. In *Proceedings Design, Automation, and Test in Europe (DATE)*, pages 65–70, 2006.

[13] W. Kuo and T. Kim. An overview of manufacturing yield and reliability modeling for semiconductor products. *Proceedings of the IEEE*, 87(8):1329–1344, August 1999.

[14] Y.-H. Lee and C. Chen. A two-level scheduling method: An effective parallelizing technique for uniform nested loops on a dsp multiprocessor. *Journal of Systems and Software*, 75:155–170, 2005.

[15] J.-M. Lu and C.-W. Wu. Cost and Benefit Models for Logic and Memory BIST. In *Proceedings Design, Automation, and Test in Europe (DATE)*, pages 710–715, 2000.

[16] S.-K. Lu and C.-Y. Lee. Modelling economics of dft and dfy: A profit perspective. *IEE Proceedings, Computers and Digital Techniques*, 151(2):119–126, March 2004.

[17] P. K. Nag, A. Gattiker, S. Wei, R. D. Blanton, and W. Maly. Modeling the economics of testing: A dft perspective. *IEEE Design & Test of Computers*, 19(1):29–41, January/February 2002.

[18] Nvidia. Geforce 8800 graphics processors. http://www.nvidia.com/page/geforce 8800.html.

[19] K. Olukotun, B. A. Nayfeh, L. Hammond, K. Wilson, and K. Chang. The Case for a Single-Chip Multiprocessor. pages 2–11, 1996.

[20] P. M. O'Neill. Statistical Test: A New Paradigm to Improve Test Effectiveness & Efficiency. In *Proceedings IEEE International Test Conference (ITC)*, pages 1–10, 2007.

[21] A. Paschalis and D. Gizopoulos. Effective software-based self-test strategies for on-line periodic testing of embedded processors. *IEEE Transactions on Computer-Aided Design*, 24(1):88–99, January 2005.

[22] J. Rearick. Too Much Delay Fault Coverage Is A Bad Thing. In *Proceedings IEEE International Test Conference (ITC)*, pages 624–633, Nov. 2001.

[23] A. W. Righter, C. F. Hawkins, J. M. Soden, and P. Maxwell. CMOS IC Reliability Indicators and Burn-In Economics. In *Proceedings IEEE International Test Conference (ITC)*, pages 194–203, 1998.

[24] W. C. Riordan, R. Miller, and E. R. S. Pierre. Reliability Improvement and Burn In Optimization through the Use of Die Level Predictive Modeling. In *Proceedings IEEE International Reliability Physics Symposium*, pages 435–445, 2005.

[25] W. C. Riordan, R. Miller, J. M. Sherman, and J. Hicks. Microprocessor Reliability Performance as a Function of Die Location for a 0.25μ, Five Layer Metal CMOS Logic Process. In *Proceedings IEEE International Reliability Physics Symposium*, pages 1–11, 1999.

[26] J. Saxena, K. Butler, V. Jayaram, and S. Kundu. A Case Study of IR-Drop in Structured At-Speed Testing. In *Proceedings IEEE International Test Conference (ITC)*, 2003.

[27] S. Shamshiri, P. Lisherness, S.-J. Pan, and K.-T. Cheng. A Cost Analysis Framework for Multi-Core Systems with Spares. In *Proceedings IEEE International Test Conference (ITC)*, pages 1–8, 2008.

[28] K. Sundararaman, S. Upadhyaya, and M. Margala. Cost Model Analysis of DFT Based Fault Tolerant SOC Designs. In *Proceedings International Symposium on Quality of Electronic Design (ISQED)*, pages 465–469, 2004.

[29] Tilera. Tile64 processor family. http://www.tilera.com/products /processors.php.

[30] M. F. Zakaria, Z. A. Kassim, M. P.-L. Ooi, and S. Demidenko. Reducing burn-in time through high-voltage stress test and weibull statistical analysis. *IEEE Design & Test of Computers*, 23(2):88–98, March 2006.

[31] L. Zhang, Y. Han, Q. Xu, and X. Li. Defect Tolerance in Homogeneous Manycore Processors Using Core-Level Redundancy with Unified Topology. In *Proceedings Design, Automation, and Test in Europe (DATE)*, 2008.