A MODEL FOR ENHANCED MANUFACTURABILITY OF DEFECT TOLERANT INTEGRATED CIRCUITS

Zahava Koren and Israel Koren*

University of Massachusetts at Amherst
Department of Industrial Engineering and Operations Research
*Department of Electrical and Computer Engineering
Amherst, MA 01003

Abstract

Many factors contribute to the cost of manufacturing integrated circuits. These include the yield of the designed IC, the complexity of its testing, the packaging cost, etc., and they all must be taken into account when designing a defect tolerant integrated circuit. We present in this paper a mathematical model which includes all major factors contributing to the cost of manufacturing ICs. This model allows the determination of the design which maximizes the expected profit rather than maximizing the yield. Numerical examples illustrating the proposed model are also presented.

Introduction

With the advances in integrated circuit (IC) technology and the trend towards Very Large Scale Integration (VLSI), a very large amount of research is being performed concerning the different aspects of IC manufacturing. However, most of the published research deals with one specific aspect of the general problem, and not enough effort has been put into viewing the picture as a whole. Lowering the cost of manufacturing integrated circuits (and consequently, increasing the profits) still remains one of the most important and most challenging problems of circuit manufacturing, and the need arises for a cost model which will take into account as many of the relevant aspects as possible, rather than concentrating on one factor of the general problem.

In addition to the various cost factors, such a model must include an objective function based on which the choice of an optimal design can be made. This objective function will, in general, differ according to the specific function that the chip must execute. However, in general terms, the objective of the ICs manufacturer is to

maximize the net profit, namely, the income from the sales of the operational chips minus the cost of manufacturing all the chips.

Due to the complexity of the fabrication process, manufacturing defects (caused by dust and undesired chemical particles) are unavoidable. Not all defects, however, result in an actual damage to the circuit [3]. A manufacturing defect which interferes with the proper operation of the chip (integrated circuit) is called a manufacturing fault. This paper deals with circuits which have some amount of fault tolerance incorporated into them.

The return from an operational chip depends on the type of the chip and is usually a function of its size, but in the case of fault-tolerant chips it may depend on the size of the fault-free portion of the chip. This reward may be monetary, or it may represent some other measure of the chip's performance (which eventually will be reflected in the selling price of the chip).

The number of operational chips is determined by the yield which depends on the complexity of the circuit, on it size, on the amount of redundancy incorporated in it, and clearly on the density and distribution of faults on the wafer.

The manufacturing cost has several ingredients, yet most of them (e.g., the cost of manufacturing the wafer's masks) are fixed (i.e., independent of the size of the chip), and are, therefore, not included in the objective function. The two main variable costs are the testing cost and the packaging cost. We assume that each manufactured chip is tested and, if found to be good, is then packaged. The testing cost depends on the function of the chip and on its size. The testing procedure is, however, not always complete and an existing fault may not be discovered. The faulty chip will in this case be packaged, and only after additional testing will be diagnosed as faulty. The packaging cost is determined primarily by the number of external connections that the chip has, which, in turn, depends on the function and design of the circuit, and for a given design can be viewed as a function of the total area of the chip.

The goal of this paper is the development and analysis of a mathematical model for enhanced manufacturability of fault tolerant integrated circuits. We suggest several objective functions, each suitable for a different type of circuit. This enables us to determine optimal values of chip size and redundancy for different integrated circuit structures, and to analyze the sensitivity of the design to the choice of the specific yield model and the set of system parameters.

The Mathematical Model

In our model, a chip consists of basic units named modules whose purpose is to execute the functions of the chip, and an auxiliary circuitry which supports the modules. For example, a memory chip consists of several storage cell arrays and address decoding circuitry. The area of a module is assumed to be the unit area, and all the other circuitry is measured in these units. The basic chip consists of N modules plus a support circuitry whose area is $S(N)$, where $S(N)$ is a non-decreasing function of N. Due to manufacturing defects, some of the modules in a chip may become faulty. To achieve a degree of fault-tolerance, R redundant modules are added to the basic number of N. The fault-tolerance capability requires an additional reconfiguration circuitry whose purpose is to restructure the fault-free modules into an operational chip, and whose area, denoted by $C(N,R)$, is a non-decreasing function of N and of R. Note that the auxiliary (support and reconfiguration) circuitry has usually no incorporated redundancy and must therefore be designed more conservatively to make it less prone to defects.

Given the functions that the chip must execute, an appropriate objective function should be constructed to fit the specific requirements of this type of circuit. This objective function has N and R as decision variables and takes into account the four main factors, namely, the testing cost, the packaging costs, the yield, and the income from an operational chip.

Let $A(N,R)$ denote the total area of the chip. It can be calculated as follows,

$$A(N,R) = N + S(N) + R + C(N,R) \tag{1}$$

The cost of the initial testing of a chip is assumed to be a function $T_1(N,R)$ of N and R while the cost of testing the final product is a function $T_2(N)$ of N only. Similarly, for a given design, the packaging cost of one chip is assumed to be a function $K(N)$ of N only.

The return from one operational chip can be assumed to be a non-decreasing function $V(n)$ of the number n of fault-free functional (i.e., non-auxiliary) modules in the chip. Several special cases may be of interest here. We first make the distinction between two main types of applications. In the first, the chip must have N fault-free modules for proper operation, and any additional modules are just redundant spares. In this case

$$V(n) = \begin{cases} U(N) & for \quad n \geq N \\ 0 & for \quad n < N \end{cases} \tag{2}$$

where $U(n)$ is a non-decreasing function satisfying $U(n) > 0$ for $n > 0$. In the other type of applications (memory arrays, for example), any fault-free redundant modules can be used to enhance the chip operation rather than being only stand-by spares. In some of these applications, even a chip with less than N operational modules can be used, though to a lesser degree. In this case

$$V(n) = \begin{cases} U(n) & for \;\; L \leq n \leq N+R \\ 0 & for \;\; n < L \end{cases}$$

where L is the minimal number of modules required for operation of the "partially good chip."

The exact form of the function $U(n)$ depends on the application of the circuit. For memory chips, $U(n)$ is the usefulness of having n storage arrays, and can, therefore, be considered as being linear in n. Thus, $U(n) = u \cdot n$ for some constant u. For processors, $U(n)$ should reflect the speed-up obtained by n modules and is therefore some concave function of n, e.g., $U(n) = u \cdot log(n)$. If the chip's function is interconnection, then $U(n)$ should be some measure of the throughput obtained with n modules, and can be calculated once the design of the chip is known.

The design of the chip, combined with specific values for N and R determine the area $A(N,R)$, and the testing and packaging costs. The number of chips on a wafer, denoted by $I(N,R)$ is determined by

$$I(N,R) = \frac{W}{A(N,R)} = \frac{W}{N + S(N) + R + C(N,R)} \tag{3}$$

where W denotes the wafer area, measured in modules.

As opposed to all the deterministic factors mentioned above, the number of *operational* chips on a wafer is a random variable, since it is affected by the number and distribution of manufacturing defects which are random by nature. We will include in our model the yield of the chip, which is the expected value of this random variable.

Another element which should be considered random is the outcome of the initial testing of the chip. Let c_t denote the *coverage* of the initial test, defined as the probability that a faulty chip will be diagnosed as such. Then, $1 - c_t$ is the probability that a faulty chip will be considered good, and only after the packaging and the additional testing will be determined to be faulty. We assume that there are no "false positives", i.e., a non-existing fault will never be diagnosed and that the final testing has a coverage of 1.

We now proceed to construct an objective function which includes all the elements mentioned above, and represents the net profit obtained from one wafer for a given circuit design and specific values of N and R. Let $y^{(n)}(N,R)$ denote the probability that the chip can be restructured into an operational n-module chip. This implies that the auxiliary circuitry is fault-free (since it has no built-in redundancy) and that exactly n out of the $N+R$ functional modules are fault-free. Let $Y^{(n)}(N,R)$ denote the probability that the reconfigured chip has *at least* n functional modules, i.e.,

$$Y^{(n)}(N,R) = \sum_{i=n}^{N+R} y^{(i)}(N,R).$$

The *yield* of the chip, defined as the probability that it can be reconfigured into an operational chip with at least N modules, can now be expressed as

$$Y^{(N)}(N,R) = \sum_{i=N}^{N+R} y^{(i)}(N,R). \tag{4}$$

To obtain the expected net profit out of one wafer, note that although all manufactured chips are tested, only those passing the initial testing are packaged, and only those passing the final testing can be sold. The expected net profit obtained from one chip (denoted by $EP(N,R)$) can, therefore, be written in the form

$$EP(N,R) = \sum_{i=L}^{N+R} y^{(i)}(N,R)U(i) - \left[1 - c_t \sum_{i=0}^{L-1} y^{(i)}(N,R)\right][K(N) + T_2(N)] - T_1(N,R)$$

$$= \sum_{i=L}^{N+R} y^{(i)}(N,R)U(i) - [K(N) + T_2(N)]\left[1 - c_t(1 - Y^{(L)}(N,R))\right] - T_1(N,R) \tag{5}$$

and the expected net profit for all $I(N,R)$ chips on a wafer is

$$Z(N,R) = I(N,R) \cdot EP(N,R). \tag{6}$$

Equation (5) is general, and can be applied to those types of circuits in which partially good chips are acceptable. In many applications, however, partially good chips cannot be used. For proper operation, a chip must have at least N fault-free modules out of the $N+R$ and any additional fault-free modules cannot be utilized but serve as spares. In this case, $V(n)$ has the form as in (2) and consequently,

$$EP(N,R) = \sum_{i=N}^{N+R} y^{(i)}(N,R)U(N) - [K(N) + T_2(N)]\left[1 - c_t(1 - Y^{(N)}(N,R))\right] - T_1(N,R)$$

$$= U(N)Y^{(N)}(N,R) - [K(N) + T_2(N)]\left[1 - c_t(1 - Y^{(N)}(N,R))\right] - T_1(N,R) \quad (7)$$

The design of the chip includes finding basic design rules which include N and R as parameters, and then choosing values for N and R which optimize the appropriate objective function $Z(N,R)$. Once the basic design and the objective function are determined, the maximization problem can be solved in two steps. First, an optimal R, $R^{(N)}$, is found for every value of N, and second, the value of N which maximizes $Z(N,R^{(N)})$ is calculated.

Further investigation of the function $Z(N,R)$ is required to determine which mathematical properties it possesses, so that the search for the optimal design of the circuit can be facilitated. If the function is concave or unimodal, then a single optimum exists and can be obtained by using differentials, differences, or some other known search method. The shape of the function Z and the optimal design depend on the main system parameters. These include the area of the auxiliary circuitry $S(N)+C(N,R)$, the yield $Y^{(N)}(N,R)$, the testing and packaging cost functions $T_1(N,R)$, $T_2(N)$ and $K(N)$, the test coverage c_t and the reward function $U(n)$. Since, in practical situations, we can obtain only estimates rather than accurate values of the above functions, we need to find how sensitive the solution is to the choice of the system parameters. Most of these parameters can be modified (at some cost). The auxiliary area can be decreased through a better design. The packaging cost can be decreased with a different technology, and there is a trade-off between the testing cost and the test coverage. The sensitivity analysis is a tool which helps in determining whether these modifications are cost effective.

Examples

We illustrate the application of (6) for calculating the optimal redundancy R^* through an example of a 16 bits defect-tolerant microprocessor similar to the design described in [1]. A bit-sliced design style is followed for the data path of the microprocessor, which enables the use of a straightforward redundancy scheme. One or more spare slices are incorporated in the implementation allowing the replacement of defective slices by good spares. Two choices for the design of these slices are investigated: one is single bit wide slices, the second is two bit wide slices. Having a two bit wide spare slice might prove to be more cost-effective than two single bit spare slices in the case of clustered faults. Often two adjacent (single) bit slices will be faulty, and the switching circuitry for a single two-bit spare slice is simpler and less area consuming

than that for two single-bit spare slices.

For the control part of the microprocessor we assume a (microprogram) control memory with spare rows and columns for defect-tolerance (unlike the PLA-based design in [1]). In this paper we examine, separately, the design alternatives of the control memory and those of the data path. A similar, but slightly more complex, analysis is required in order to find the combined optimal design of the system as a whole.

For the data path we chose the following cost parameters. First, exactly 16 operational bit slices are required. Any additional defect-free slices are useless. Also, a chip with 15 or less defect-free bit slices is unacceptable. Thus,

$$V(n) = \begin{cases} 16u & for \quad n \geq 16 \\ 0 & for \quad n < 16 \end{cases}$$

where u is a constant. The area of one bit slice is chosen as the unit area. The amount of support circuitry in a bit sliced microprocessor is linear in the number of bits and so is the packaging cost. Both are independent of the amount of redundancy added.

$$S(N) = sN, \qquad K(N) = kN\,.$$

The complexity of the reconfiguration circuitry is linear in the number of required bits but increases exponentially with the number of spare slices included in the design.

$$C(N,R) = c_1 N e^{c_2 R/l}$$

where l is the width of the slice.
The cost of the initial testing of the N bit slices and the R spare bit slices is assumed to be exponential in $(N+R)$, and the cost of the final testing is assumed to be exponential in N,

$$T_1(N,R) = d_1 e^{d_2(N+R)}, \qquad T_2(N) = f_1 e^{f_2 N}$$

Other variations of the above cost parameters can be analyzed.

The fault distribution model chosen is the negative binomial model under the large area clustering assumption [4] with an average of λ faults per unit area and a clustering parameter α. Using this model, the expressions for the yield become

$$y^{(n)}(N,R) = \binom{N+R}{n} \sum_{j=0}^{N+R-n} (-1)^j \binom{N+R-n}{j} \left(1 + \frac{[S(N)+C(N,R)+n+j]\,\lambda}{\alpha}\right)^{-\alpha}$$

and

$$Y^{(n)}(N,R)=\sum_{j=0}^{N+R-n}(-1)^j\binom{N+R}{n+j}\binom{n+j-1}{j}\left(1+\frac{[S(N)+C(N,R)+n+j]\,\lambda}{\alpha}\right)^{-\alpha}$$

The numerical values for the different parameters for the data path have been chosen as follows:
$u = 1$, $s = 0.3$, $k = 0.5$, $c_1 = c_2 = d_1 = d_2 = f_1 = f_2 = 0.1$, $c_t = 0.9$ and $\alpha = 0.25$.
We first calculated the expected net profit per wafer, Z, as a function of λ for three redundancy schemes: No redundancy ($R = 0$), $R = 1$ single bit slice, $R = 2$ single bit slices. The results are depicted in Figure 1. Clearly, the expected profit decreases with λ. For low values of λ the optimal redundancy is zero while for high values of λ it is $R = 2$ single bit slices.

Figure 2 shows the optimal redundancy in the data path as a function of λ for two values of the parameter α, $\alpha = 0.25$ and $\alpha = 2.5$. Here, less redundancy is required for lower values of α (which indicate higher clustering).

Figure 3 depicts the optimal redundancy in the data path as a function of λ (with $\alpha = 0.25$) for three values of the reconfiguration area coefficient c_2: $c_2 = 0.1$, $c_2 = 0.3$, and $c_2 = 0.5$. We see that the higher the value of c_2 (which indicates that the reconfiguration requires a higher area penalty), less redundancy is optimal. No redundancy is the optimum for the highest value chosen: $c_2 = 0.5$.

The required size of the control memory is denoted by $M_1 \times M_2$. In our numerical examples we assume a control memory of size $1K \times 32$ bits. Since $M_1 \gg M_2$, we assume that adding redundant rows is more cost-effective than adding redundant columns. Therefore, we restrict our analysis to the case where R redundant rows are added to the $1K$ required rows, and we chose the module, accordingly, to be a row. Most of the cost factors for the control memory are assumed to have the same functional form as before, except for the reconfiguration circuitry which is assumed to be linear, rather than exponential, in R, i.e., $C(N,R) = c_2 R$.

The numerical values for the different parameters for the control memory have been chosen as follows:
$u = 5$, $s = 0.1$, $k = 0.1$, $c_2 = d_1 = f_1 = 0.1$, $d_2 = 0.01$, $f_2 = 0.001$, $c_t = 0.9$ and $\alpha = 0.25$.
Figure 4 shows the optimal redundancy in the $1K \times 32$ bits control memory as a function of λ for two values of α. Here, unlike Figure 2, more redundancy is required for the lower value of α. This difference needs to be further investigated.

Finally, we compare (for the control memory) the optimal redundancy which max-

imizes the profit to that which maximizes the equivalent yield. The latter takes into account the additional area due to the redundant modules but ignores all other factors like testing and packaging costs. Figure 5 depicts these two optimal redundancies as a function of λ. The most important conclusion that should be drawn from this figure is that there are values of the manufacturing parameters for which maximizing the yield does not guarantee that the net profit is maximized. Therefore, it is worthwhile for IC manufacturers to employ a comprehensive model that includes all relevant factors affecting the cost of manufacturing an integrated circuit.

Conclusion

A mathematical model for enhanced manufacturability of defect tolerant integrated circuits has been described in this paper. Such a model allows the determination of the optimal redundancy that maximizes the expected net profit rather than the yield only. Numerical examples demonstrating the significance of the proposed model have been presented. Further investigation of the suggested model and its various cost factors is needed.

Acknowledgment: This work was supported in part by NSF under contract MIP–8805586.

References

[1] R. Leveugle, M. Soueidan, and N. Wehn, " Defect Tolerance in a 16 Bit Microprocessor," *Defect and Fault Tolerance in VLSI Systems*, Vol. 1, I. Koren (ed.), pp. 179-190, Plenum, 1989.

[2] I. Koren, Z. Koren and D.K. Pradhan, "Designing Interconnection Buses in VLSI and WSI for Maximum Yield and Minimum Delay," *IEEE J. of Solid-state Circuits*, pp. 859-866, June 1988.

[3] I. Koren and A.D. Singh, "Fault Tolerance in VLSI Circuits," *Computer*, Special Issue on Fault-Tolerant Systems, Vol. 23, pp. 73-83, July 1990.

[4] I. Koren and C.H. Stapper, "Yield Models for Defect Tolerant VLSI Circuits: A Review," *Defect and Fault Tolerance in VLSI Systems*, Vol. 1, I. Koren (ed.), pp. 1-21, Plenum, 1989.

[5] C.H. Stapper, F.M. Armstrong and K. Saji, "Integrated Circuit Yield Statistics," *Proc. IEEE*, Vol. 71, pp. 453-470, April 1983.

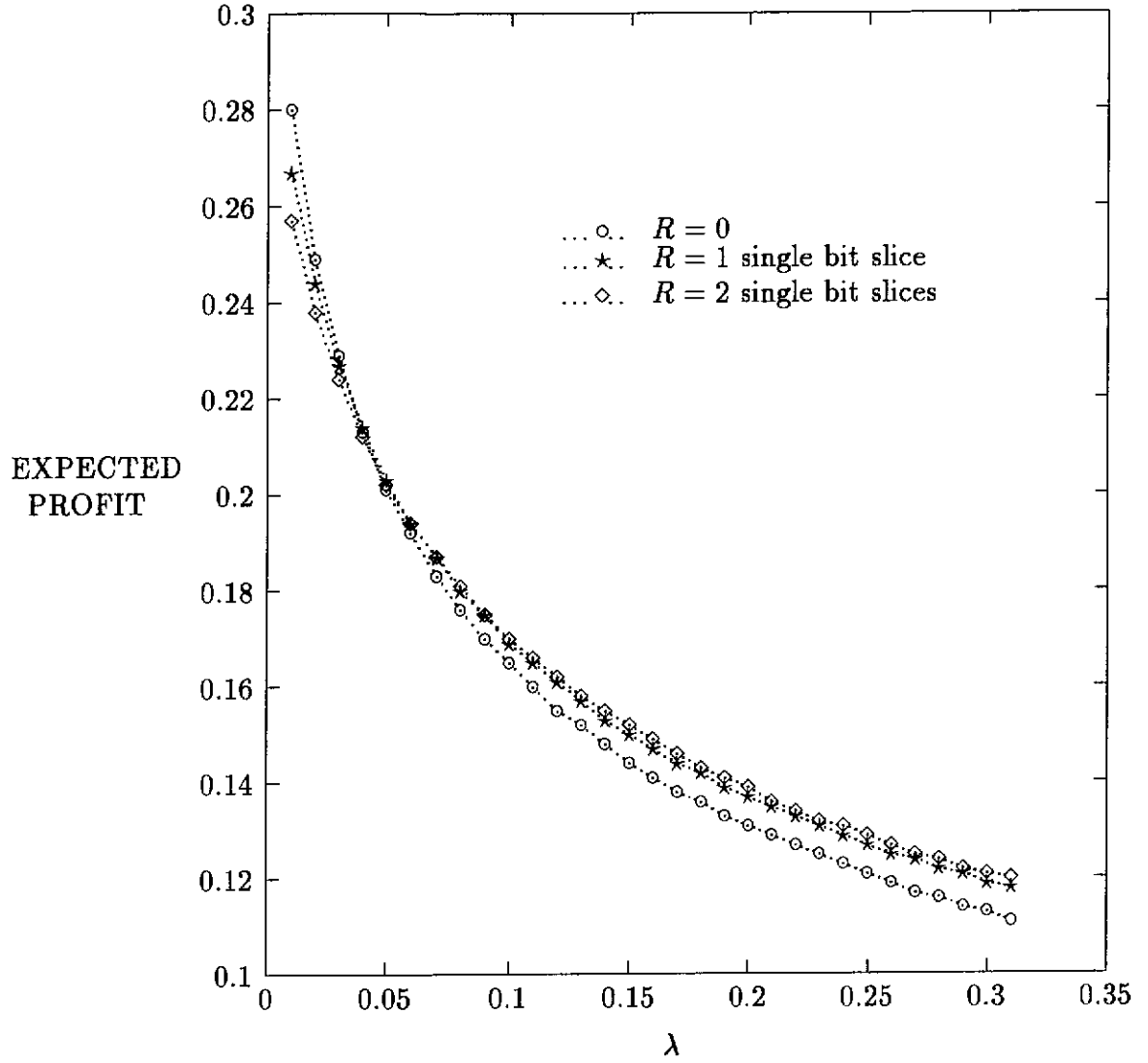

Figure 1: The expected net profit vs. λ for three redundancy schemes ($\alpha = 0.25$).

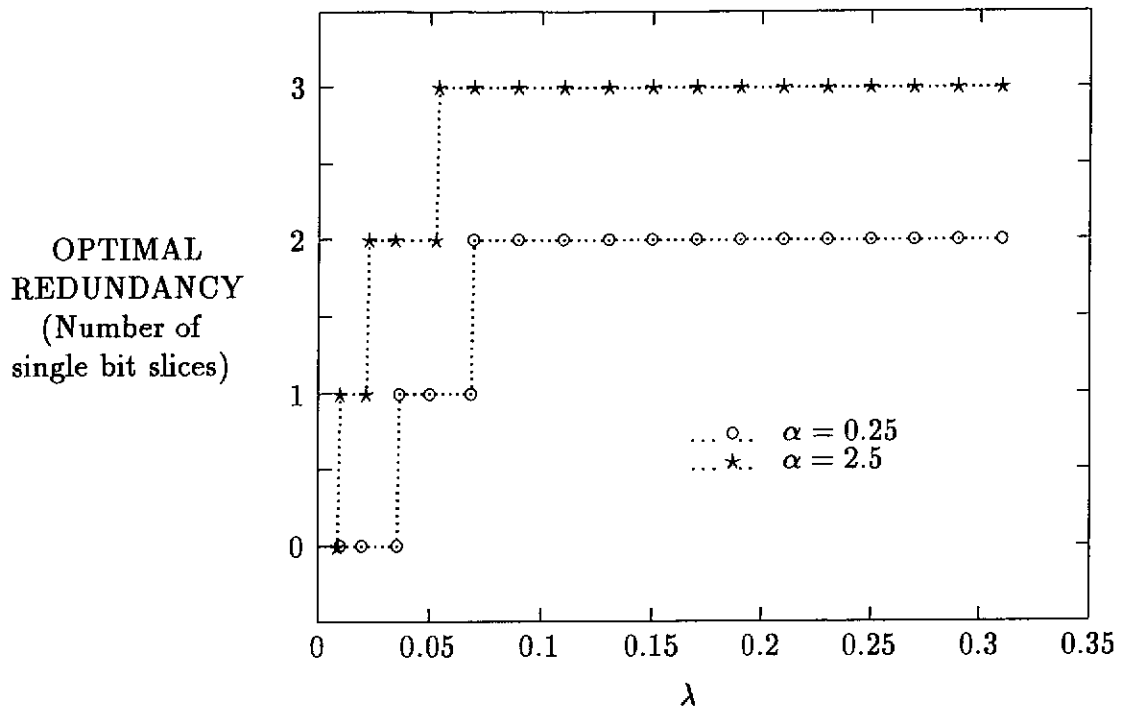

Figure 2: The optimal data path redundancy vs. λ for two values of α.

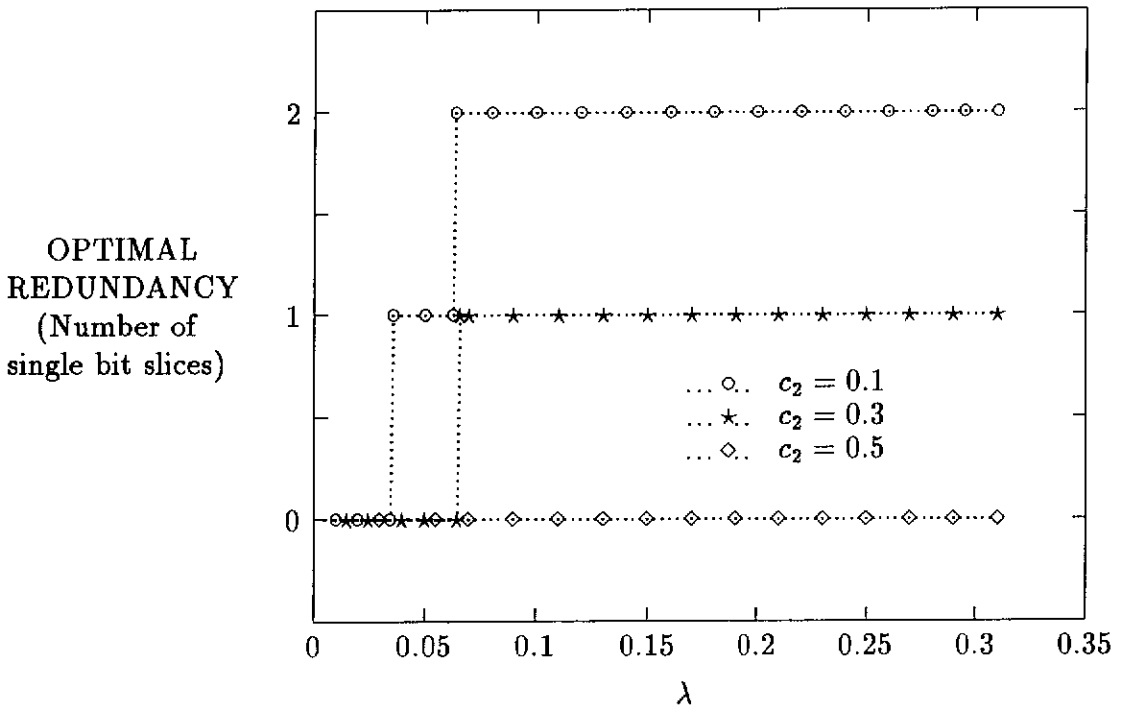

Figure 3: The optimal redundancy vs. λ for three values of the reconfiguration coefficient c_2.

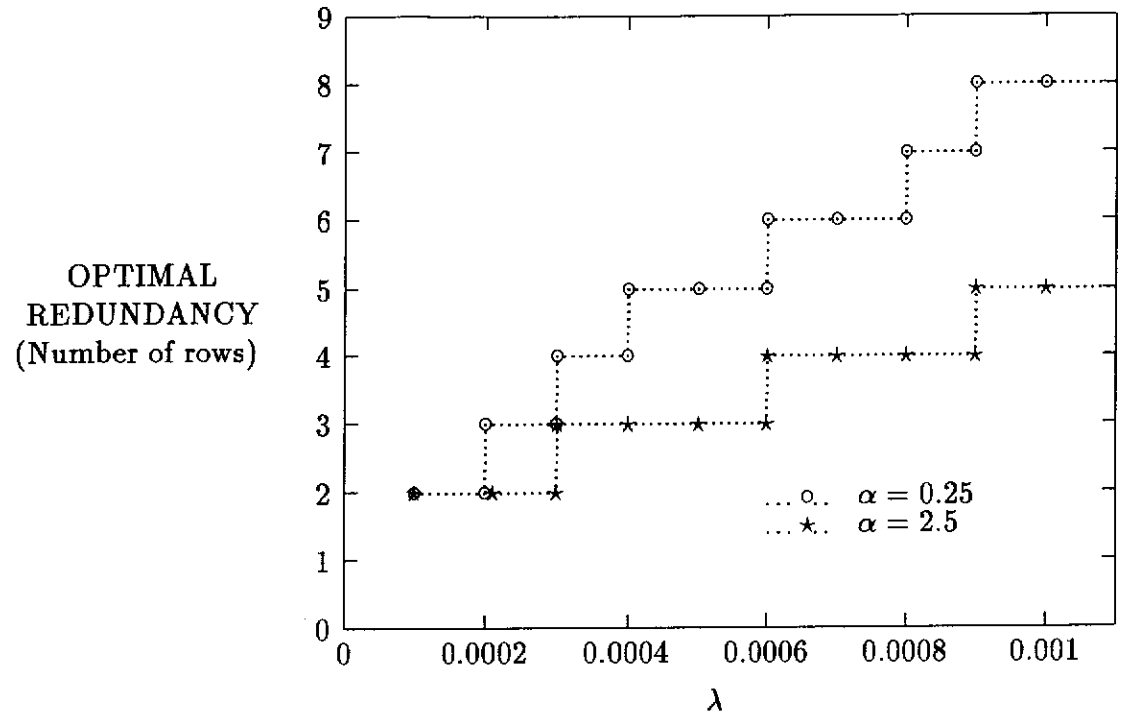

Figure 4: The optimal control memory redundancy vs. λ for two values of α.

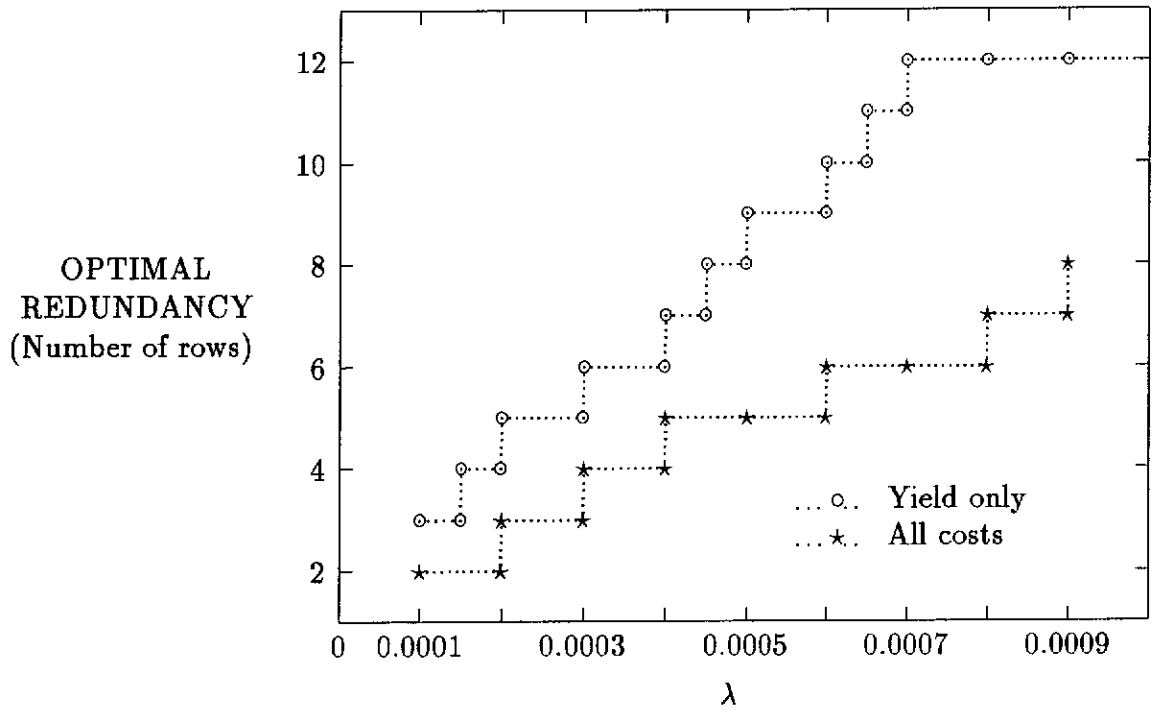

Figure 5: The optimal redundancy vs. λ for maximizing the net profit or the yield only.