Cost of Silicon Viewed from VLSI Design Perspective

Invited

Wojciech Maly
Department of Electrical and Computer Engineering
Carnegie Mellon University , Pittsburgh , Pennsylvania 15213.

Abstract - This paper provides an overview of design/test/CAD silicon cost-related issues. All major factors contributing to the rapid growth of manufacturing costs are explained and a simple cost model is introduced to assess possible impact of cost growth on the VLSI arena.

I. INTRODUCTION

Design and test, and consequently CAD, cannot evolve independently of trends in IC manufacturing -- the main source of revenue supporting the entire microelectronics industry. The design-manufacturing interface has many components. The most visible and well understood are manufacturing constraints, given in the form of design rules and device characteristics of various kinds. There is, however, one more important but not as visible link between design and manufacturing -- the notion of silicon cost. This notion indirectly affects processes shaping evolution of the entire IC industry.

In the last twenty years silicon cost -- computed per single IC transistor -- has been constantly decreasing, making multi-million transistor systems on a single die both feasible and cost-effective. Consequently, for a long time design, test and CAD have focused solely on design complexity and time-to-market related issues while the cost of IC manufacturing has been viewed as relevant only for high volume production.

In the last couple of years the view on manufacturing costs has begun to change [1]. The main reason for this change is the exponentially growing cost of manufacturing facilities, which is estimated soon to reach 1 billion dollars per fabline [2,3,4]. Market conditions have changed as well. Increased competition has led to a decrease in previously lucrative profit margins [5].

How will the aforementioned changes affect microelectronics market? What will the new situation on the manufacturing-design interface look like? What kind of corrections to IC design and manufacturing strategies might be necessary? What are the needs to be addressed by new design strategies and CAD tools? These and many other questions of a similar kind should be discussed by the design/CAD community before any possible negative consequences of the manufacturing cost increase occur. The purpose of this paper is to provide an overview of the IC manufacturing cost domain in order to initiate discussion which addresses the above concerns.

II. ECONOMIC AND TECHNOLOGY TRENDS

To set the stage for the discussion presented in this paper it is useful to summarize basic facts characterizing electronics industry in general and microelectronics in particular. In the last 10 years the entire electronics industry has grown at an average rate of around 40 billion dollars of sales of electronic equipment a year, reaching a level of 700 billion dollars this year [6]. Such progress has been made possible by the progress of the IC industry which has accounted for around 10-13% of the total electronics market size. The progress in microelectronics, in turn, was enabled by the constantly decreasing size of an average integrated transistor. The rate of change of transistor size -- traditionally characterized through minimum length of a transistor channel, also called minimum feature size -- is shown in Fig. 1 [1,6,7,8].

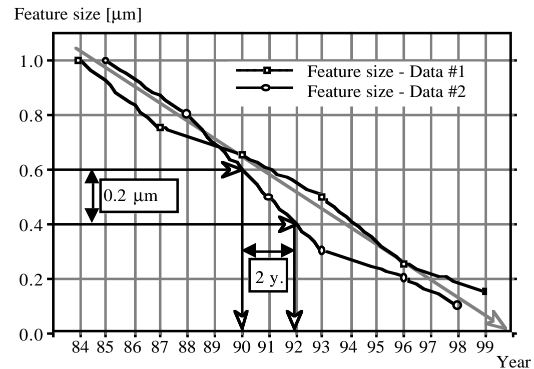


Fig. 1. Minimum feature size as a function of time.

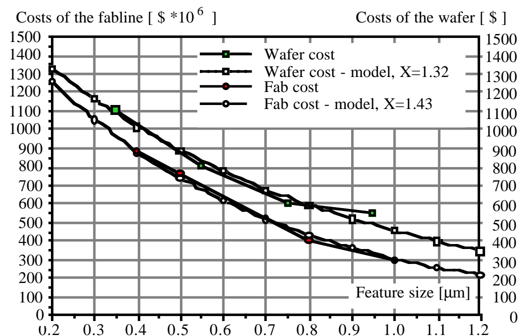


Fig. 2. Cost of a fabrication line and manufacturing wafer.

Such impressive progress has been facilitated mainly by the progress achieved in the design and construction of manufacturing equipment.

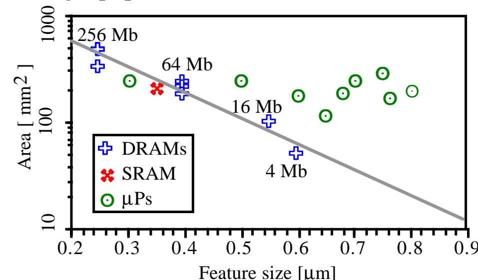


Fig. 3. Die size trend.

Hence, slightly simplifying the reality, one can conclude that manufacturing equipment industry decides the pace and direction of the entire electronics. However this industry, by accounting for less than 1.4% of the electronics market size, is

in turn fully dependent on the IC industry, or more precisely on high volume (commodity) IC producers.

One has to realize that the level of integration achieved has also resulted in: (a) the already mentioned increase in the cost of manufacturing facilities (Fig. 2), (b) an increase in die size (Fig. 3), (c) an increase of the process complexity and (d) a decrease in the required contamination levels [7,9] (Fig. 4).

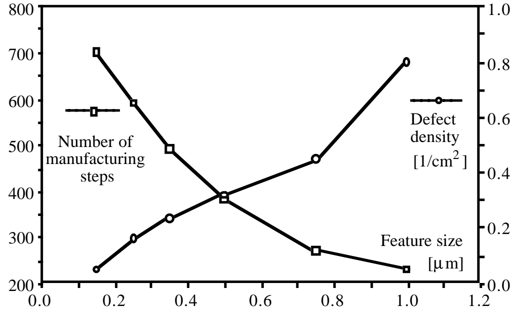


Fig. 4. Number of manufacturing steps and defect density required for subsequent generations of IC technologies.

III. Transistor Cost Model

The above summary leads to two seemingly contradicting observations. On one hand the complexity of the processes and manufacturing equipment are subject to rapidly growing costs. On the other hand the rate of progress seems to be absolutely unaffected by these growing costs. To explain the above contradiction one should focus on the manufacturing cost computed per single transistor - a good measure of the cost of an IC's functionality. This cost has always been decreasing [6] allowing the new functionality to be sold at very attractive profit margins. For this reason overall manufacturing cost increases did not really matter in the past .

Recently the situation has changed. There are some indications that the cost per transistor may no longer decrease [10], or at least the rate of the cost decrease may become slower [11]. Such change may have a dramatic impact on the entire IC industry since the transistor size decrease may not provide simultaneous performance and cost gains.

The purpose of this section is to conduct a transistor cost analysis in order to: (a) determine whether transistor cost trends known from the past will continue into the future and (b) demonstrate the complexity of the IC manufacturing cost problem.

The cost of a transistor, C_{tr} , in a functioning (fault free) IC can be expressed as:

$$C_{tr} = \frac{C_w}{N_{ch} N_{tr} Y} \qquad (1)$$

where: C_W - is the cost of manufacturing wafer, N_{ch} - is the number of IC dies per wafer, N_{tr} - is a number transistor per die and Y - is the manufacturing yield i.e. the probability that a fabricated and tested die functions according to its desired specifications. The remainder of this section discusses each of the above factors in detail.

A. C_W - Cost of the Wafer

The cost of manufacturing of the wafer can be analyzed from at least two different perspectives: the accounting perspective and the analytical perspective. The accounting perspective is simple - by dividing the total cost accumulated over a period of time by the total number of fabricated wafers in this period of time, one can determine C_W for the existing (or known from the past) processes and manufacturing facilities. (For instance, it was also estimated this way that the cost of 6 inch wafer fabricated in 1 µm CMOS was between $500 and $800 [12,13]. A wafer of the same size but fabricated with 3 metal layers and 0.8 µm feature size was reported to cost $1300 [14].)

The analytical perspective is much more complex. It requires that C_W is expressed in terms of all major relevant cost factors. The most important of them are as follows:

a. Manufacturing volume. The cost of manufacturing has two components: a variable cost growing as a linear function of volume [15] and fixed costs which include R&D costs, management overhead costs and all other of non-recurring costs. Consequently, the total cost computed per wafer, C_W, can be expressed by the following function:

$$C_W(V) = C'_W + \frac{C_{over}}{V} \qquad (2)$$

where C_{over} is overhead (fixed) cost and C'_W is true cost of manufacturing per wafer. Analysis of the overhead cost is very important but difficult. The main difficulty is of a subjective nature as different companies apply different overhead accounting policies. Furthermore different products require very different amounts of R&D and design effort. The reported numbers may vary between $100K for ASIC products up to $100M [14] for microprocessors. As a result it is much easier to handle the accounting aspect of C_W than C_{over}. In this paper C_{over} is not discussed in detail and the focus is on the "pure" cost of manufacturing which includes only investment, labor, material and operational costs. (Examples of approaches dealing with accounting aspects of C_W computations are presented in [16]). In the reminder of this section only C'_W is discussed and value of C_{over} is assumed to be known and given.

b. Minimum feature size. The relationship between cost and minimum feature size is very strong and complex. A full explanation of this relationship is beyond the scope of this paper. It must be, however, explained in some detail because it is the key to understanding the main conclusions of this section.

The decrease of the minimum feature size requires process modifications to overcome physically-based limitations and to enable an acceptable level of manufacturability. The hot electron phenomenon is the best example of a physically-based problem, which was solved by the introduction of LDD (lightly doped drain) structures. This solution, however, does not come for free - it requires an increase in the process complexity and consequently manufacturing cost.

Increases in process complexity which are due to manufacturability requirements are not as evident but they are rather costly. These requirements can essentially be translated into process stability and uniformity limits and contamination standards. (Smaller feature size demands thinner layers and fewer and smaller contaminations.)

Hence, minimization of the feature size is a major factor causing cost increase through: (a) an increase in the number of manufacturing steps which must be performed using

more expensive equipment and (b) extremely demanding contamination requirements. (Both factors have already been indicated by the curves in Fig. 4.)

Despite the complexity of the above relationship one can model the cost of the wafer, C'_W, as a function of minimum feature size, using a relatively simple relationship. Namely, it was found experimentally [10] that:

$$C'_W = C_0 \, X^{0.5\,(1 - \lambda)} \tag{3}$$

where: λ - is the minimum feature size in microns, C_O - is a reference value (in this paper C_O was assumed to be the cost of a 6 inch wafer fabricated with 1 µm feature size) and X - is a rate of the cost increase measured per single technology generation.

The most important parameter in (3) is X. In this paper it will be treated as a variable but it is very important to see the relation between X and the growth in process complexity. The value of X can also be estimated from data presented in the literature. According to Intel [14] X = 1.6, Mitsubishi [1] $X \in$ [1.6 - 2.4] and Hitachi [18] $X \in$ [1.5 - 2.0] and detailed study presented in [12] X = 1.79. (Value of X extracted from the data presented in Fig. 2 is between 1.2 - 1.4.) Of course these numbers must be treated as rough estimates. The investment increase may become much higher when existing contamination and failure analysis techniques reach their limits (which is likely) or lower when projects like TI's MMST [19] deliver promised cost/contamination control benefits.

c. Wafer size. An increase in the wafer size is highly desirable from a productivity point of view. (Larger wafers mean more chips). The problem is that larger wafer are more difficult to process. (Process uniformity and stability issues.) Larger wafers also require a significant investment in equipment retooling (microscopes, handlers, magazines etc.) We assume that any cost increase due to an increase in the wafer size is covered by the X factor.

d. Product mix. This term is used to describe the composition of the variety of the products manufactured simultaneously using the same manufacturing facility. This composition has a strong influence on the level of fabline utilization and consequently wafer costs. The essence of the problem is as follows. Fablines are designed to perform a certain process which requires specific amount of usage time of the necessary equipment. In the case of a mono-product high volume stable operation, a fabline can be designed such that each piece of equipment is utilized nearly to its full theoretical capacity. But in the cases when there are a number of products fabricated in small volumes, each requiring different composition of manufacturing floor, it is highly likely that the equipment load is non-uniform and demand for some operation is higher than the available capacity while some other equipment is idle but still consuming resources. (The cost of "ownership" for same equipment may be the same for "active" and "inactive" equipment usage.) Detailed study shows that the ratio of the cost of the wafer fabricated with low volume multi-product fabline and high volume mono-product environment may reach as high value as 7 [12].

e. Design complexity. There is no direct relationship between C_W and features of the fabricated IC if the cost of testing is not taken into account. If it is then C_W should be seen as a function of many test relevant attributes of fabricated IC which affect the length of the testing procedure. A detailed discussion of these attributes is again beyond the scope of this paper. However, it is necessary to mention that the cost of testing (both probe and final) will grow with a decrease of minimum feature and an increase in the die size. And in the extreme case the cost of testing a wafer may be comparable with the cost of manufacturing. It is also necessary to stress that adequate analytical relationships expressing cost of testing as a function of, even basic, design attributes do not exist and any available data is purely of an experimental nature.

B. Number of IC Dies Per Wafer and Transistors Per Die

It is easy to derive N_{ch} as a function of die dimensions, wafer radius and few other parameters characterizing wafer's geometry [20]. The function used in this paper is as follows:

$$N_{ch} = \sum_{i=0}^{\mathrm{Floor}\left[\frac{2R_w}{b}\right] - 1} \left\{ \mathrm{Floor}\left[\frac{2}{a/b} \mathrm{Min}\left(R_i, R_{i+1}\right) \right] \right\} \tag{4}$$

where: $R_j = \sqrt{R_w^2 - (j\,a\,b - R_w)^2}$, R_W - is radius of the wafer and a and b are dimensions of the chip.

The number of transistors per die, N_{tr}, is a function of the design density, minimum feature size and die area. This function takes the following form:

$$N_{tr} = \frac{A_{ch}}{d_d \lambda^2} \tag{5}$$

where: A_{ch} - is area of the chip (axb), d_d - is design density expressed in terms of a number of minimum feature size squares needed to draw single "average" transistors and λ is minimum feature size.

Table 1. Design densities for µPs functional blocks [22].

Funct. block	Area [mm ^{2}]	# of tr.	$d_d [\frac{\lambda^2}{\# tr.}]$
I - cache	33.2	1200k	43.2
D - cache	35.7	1100k	50.7
F. point unit	45.9	323k	222.3
Integer unit	38.3	232k	257.9
MMU	20.4	118k	270.5
Bus unit	12.7	50k	399.0

Table 2. Deign densities for a spectrum of ICs described in [23,24].

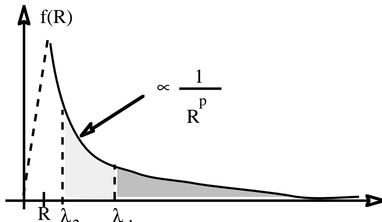
Type of IC	F. size [µm]	$d_d [\frac{\lambda^2}{\# tr.}]$
µP, BiCMOS, 3 M	0.3	907.95
µP, CMOS, 3 M, Alpha21064	0.68	250.13
µP, CMOS, 2 M, R4400SC	0.6	224.64
µP, CMOS, 3 M, PA7100	0.8	370.66
µP, BiCMOS, 3 M, Pentium	0.8	149.11
µP, CMOS, 4M, PowerPC601	0.65	102.28
µP, BiCMOS, 3 M, 2P, SuperSpark	0.7	168.53
µP, CMOS, 2M, 68040	0.65	249.23
1Mb SRAM, 2M, 2P	0.35	36.00
16Mb SRAM, 2M, 4P	0.25	17.80
64Mb, DRAM, 2M	0.4	22.29
256Mb DRAM, 3M	0.25	20.18
GateArray, 53Kg, BiCMOS, "50%"	0.8	507.66
GateArray, BiCMOS,	0.5	403.20
SOG, 177Kg, 35-70%, CMOS, 3M	0.8	249.44
SOG, 235Kg, 70%, CMOS, 3M	0.8	117.19
PLD, 1.2Kg, EEPROM, 2M, 2P	0.8	2631.04

Design density is a function of many attributes: the function implemented on the chip, design style, design tools and/or applied design library [21]. For the discussion presented in this paper it is enough to note that the large difference occurs between different designs as shown in Tables 1 and 2.

C. Manufacturing Yield - Y

Due to space limitations the discussion of manufacturing yield must also be limited to the presentation of basic facts. Yield loss occurs when there is an unacceptable mismatch between the expected and actual parameters of a fabricated IC [25]. The mismatch is a combination of process disturbances and/or non-optimal design decisions. These may cause either inadequate performance (e.g., excessive power consumption, too long delay) or functional failure. Typically inadequate performance is caused by a process-design mismatch producing ICs with an excessive sensitivity to "global process disturbances". Functional failures are due to "spot defects" which produce shorts or opens in the circuit's connectivity. Since both mechanisms are exclusive, yield loss, Y, is often represented by the product $Y_{fnc}Y_{par}$ where Y_{fnc} is the functional yield associated with spot defects and Y_{par} is the parametric yield associated with global process disturbances, respectively. In this paper we assume that parametric yield loss is not of primary importance and we focus on Y_{fnc}.

To explain the essence of the functional yield loss mechanism let us assume, for the sake of simplicity, that the defect is a contamination-generated spot (disk) of extra conducting, semiconducting or insulating material embedded in a layer of the IC during the manufacturing process. Such a disk may (but does not have to), cause a fault (i.e. IC malfunction) and consequently yield loss. Whether a defect causes a fault or not depends on its size and location. Therefore, yield estimation must be based not only on defect density but also based on defect size distribution.



Fig, 5. Defect size distribution.

There are a number of functions which can be used to model a defect size distribution [23, 24]. The most widely accepted -- due to its simplicity -- is the function shown in Fig. 5 which decreases above certain value R_o as $1/R^p$ where R is defect radius and p is a parameter. Observe that the decrease in the minimum feature size rapidly increases the number of defects which may cause faults.

To handle the above effect one can modify the standard Poisson yield model:

$$Y = \exp (-A_{ch} \; D_o) \qquad (6)$$

by substituting D_o with the factor: D / λ^p, where D and p are defect characterization parameters and λ is minimum feature size in microns. One can also substitute A_{ch} with the product: $N_{tr} \, d_d \, \lambda^2$. Consequently, (6) takes the following form [26]:

$$Y = \exp[- \frac{N_{tr} \; d_d D}{\lambda^{p-2}}] \qquad (7)$$

Observe that Y is, as expected, a strong function of λ (p was found experimentally [26] to be in the range 4 - 5). Hence, an increase in the scale of integration accomplished through a decrease in λ requires a drastic decrease in defect density D if the yield is expected to be kept on the acceptable level.

D. Transistor Cost Model - Conclusions

The above presented brief overview clearly shows that the cost of a single transistor in a fault free IC, C_{tr}, strongly depends on the minimum feature size, manufacturing volume and the rate of the manufacturing cost increase. Moreover, transistor's cost can be decreased by a decrease in λ and/or an increase in the wafer size (assuming modest values of X) but under the condition that defect density is rapidly decreased to keep value of Y high. The next section explores this relationship in detail.

IV. COST OF THE TRANSISTOR - ESTIMATES

The constant decrease in the cost of an IC transistor has often been taken for granted. In some cases transistor cost has been even viewed as almost design/test irrelevant [27]. And such a view is understandable since in the past typical design/test objectives were focused on the IC performance only and manufacturing costs were determined through the choice of technology, design style and die size limits i.e., through arbitrary decisions which were beyond the typical design/test domain. In addition lucrative profit margins amplified the performance side of the design objectives and allowed for less cost effective design solutions. Now, as the situation may change and cost could become one of the designer's main concerns it is necessary to reexamine the transistor's basic cost trends and to analyze the design - cost dependency.

A. Transistor Cost Trends

The feature size trend shown in Fig. 1 (Sec. II) has been enabled by physics based innovation and the ability to "debug" subsequent generation of technologies to the level that assures acceptable yields. These two key abilities have allowed for a dramatic increase in the total number of transistors manufactured on a single wafer. Noting that in the past wafer cost increases have been kept on a reasonable level [11] one can clearly see the reasons for the transistor cost decrease.

To more quantitatively illustrate the above mechanism it is useful to define a generic manufacturing scenario (called here Scenario #1), which is based on the following assumptions:

<u>Assumption S1.1.</u> Value of X is between 1.1-1.3.

<u>Assumption S1.2.</u> Manufactured product is a 1 MB DRAM with appropriately designed redundant components.

<u>Assumption S1.3.</u> At the mature stage of each technology generation the yield is 100%.

<u>Assumption S.1.4</u> Production is conducted as a high volume low overhead ($C_{over} = 0$) operation.

This set of assumptions outlines a "typical" state-of-the-art manufacturing operation at the beginning of the 1990's and is especially adequate for technology and equipment driving segments of the IC market i.e. memories.

Observe now that using the above assumptions one can find, from (1) and (5), an approximation of C_{tr} :

$$C_{tr} = \frac{C'_w (\lambda) \, d_d \, \lambda^2}{A_w} \qquad (8)$$

where d_d and λ are as assumed earlier, $C'_W(\lambda)$ is given by (3) and A_W - is the area of the wafer. Using (8) one can compute curves shown in Fig. 6. Because the number of transistors per wafer increases faster than the wafer cost, C_{tr} goes down when feature size decreases which confirms the observation made earlier.

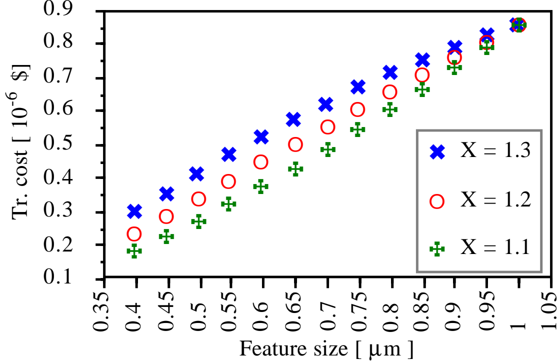


Fig. 6. Cost per transistor computed for X = 1.1, 1.2 and 1.3, respectively and C_O = \$500, d_d = 30 and R_W=7.5 cm.

It is important to stress, however, that the key premise of Scenario #1, i.e. the assumption that "somehow" it will be possible to achieve high yields for each subsequent technology generation while maintaining cost decrease on an acceptable level, is very optimistic. Never the less it seems to be accepted by almost entire industry (e.g. [2,28]) as adequate and many investment and R & D direction setting decisions of paramount importance are based on the premise of Scenario #1.

For this reason and also because some of the assumptions of Scenario #1 may soon become unrealistic it is worth analyzing, in more detail, the validity of these assumptions. The main problems are as follows:

S.1.1 The value of X in the future is likely to grow. According to the information presented in Sec. III.A X may be as high as 2.4 It is also important to remember that X may grow due to the wafer size increase, and at any juncture requiring quantum improvements in contamination control technology.

S.1.2. The assumed product has a number of attributes which do not adequately represent a number of important IC products. First of all only memories enjoy the benefits of redundancy. Secondly, the die size of an average IC product is rising. Finally, design densities of many products are very different from the d_d of DRAMs.

S.1.3 This is very critical assumption which is acceptable under the condition that contamination control and failure analysis are very effective. According to some indication published by manufacturing leaders [18] achieving high yields for subsequent technologies may be very difficult if possible at all.

S.1.4. This assumption is valid only for commodity ICs fabricated in a very high volume. All other IC including some µPs will be manufactured less efficiently and therefore for these products C_O will have to be higher than for memories.

Hence, in general one should conclude that "Scenario #1" is too "optimistic" and to obtain a better picture of the possible impact of the cost increase on the IC industry it makes sense to consider another more realistic manufacturing scenario ("Scenario #2") by assuming:

Assumption S.2.1. Value of X is in the range 1.8 - 2.4.

Assumption S.2.2 Manufactured product is a custom designed µPs with a number of transistors growing such that technology trends shown in Fig. 3 are followed.

Assumption S.2.3. The yield is 70% for each generation of technology for a 1 cm^2 die.

Assumption S.2.4 is the same as S.1.4.

To see the change in the transistor cost trend again (1) and (3) were used to derive the following formula:

$$C_{tr} = \frac{C'_w(\lambda)\, d_d\, \lambda^2}{A_w\, Y_0^{\left[\frac{A_{ch}(\lambda)}{A_0}\right]}} \tag{9}$$

where Y_0 = 70%, A_0 = 1 cm^2 and $A_{ch}(\lambda) = 16.5 \exp[-5.3\lambda]$ was extracted for the data in Fig. 3 Then (9) was applied to generate plots shown in Fig. 7.

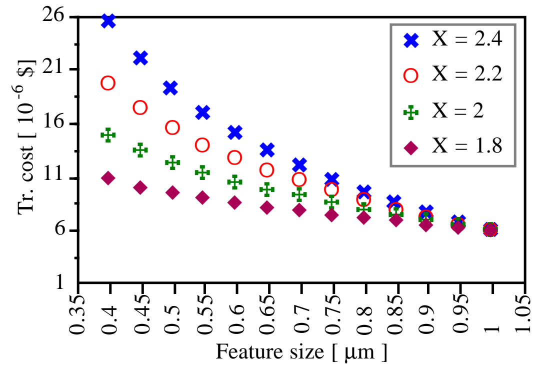


Fig. 7. Cost per transistor computed as a function of minimum feature size (C_O = \$500, d_d = 200 and R_W = 7.5 cm.)

As one can see this figure illustrates dramatically different situation. A decrease in the feature size causes an increase in the transistor cost! Keep in mind that this result was obtained using very realistic assumptions. So, one must conclude that an increase in cost per transistor is not unlikely and should be seen as a realistic scenario for some kinds of ICs.

The consequences of such possible change in the IC industry would be overwhelming. And all of them would be based in the fact that a continuation of the trend towards smaller feature size may become unhealthy or even damaging for some classes of ICs. Also as a result a growing portion of the industry would have to seek development strategies which would not be based in the expensive process innovation but could still provide equivalent cost/performance benefits.

B. Transistor Cost Optimization

To illustrate one such opportunity formulas (1), (3), (4) and (7) were used to compute transistor cost in terms of minimum feature size λ and the number of transistors per die, N_{tr}. It was assumed that X = 1.4, C_O = \$500, R_W = 7.5 cm, d_d = 152, D = 1.72 and p= 4.07. (These values were extracted from a real manufacturing operation [26]). The results are presented in Fig. 8. Each contour shown in this figure represents constant cost locations in the (λ x N_{tr}) space. Notice that there are a number

of local optima and that the cost C_{tr} changes considerably with the change of either N_{tr} or λ. This means that for each die size there is different λ^{opt} which minimizes the cost per transistor.

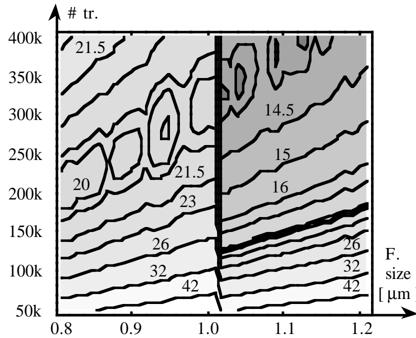


Fig. 8. Transistor costs contour plots.

In more general terms we can conclude, therefore, that by including in the IC system design process such variables as sizes of the system's partitions and minimum feature sizes of each partition one can minimize the overall system cost. It is important to note that the optimum solution may not call for the smallest possible (and expensive) feature size.

C. Cost Diversity

Another important observation which can be derived from the cost estimates discussed in this paper is illustrated by the data in Table 3. In this table the cost per transistor for a variety of product-manufacturing scenarios is presented. The first eight columns of this tables show input parameters for the cost model constructed of equations (1), (3), (4) and (7). Notice the large diversity of transistor's costs shown in 9-th column of the table.

Table 3. Cost per transistor.

	# tr.	λ [µm]	□ / # tr.	R_w	Y_0	C_0	X	C_{tr} [10^{-6}]	IC type
1	3.1M	.8	150	7.5	.9	700	1.4	9.40	BiCMOS, µP
2	3.1M	.8	150	7.5	.7	700	1.8	25.50	BiCMOS, µP
3	3.1M	.8	150	7.5	.6	700	2.2	49.30	BiCMOS, µP
4	1.7M	.8	190	7.5	.7	700	1.8	21.80	CMOS, µP
5	.85M	.8	370	7.5	.7	900	1.8	53.50	CMOS, µP
6	3.1M	.8	150	7.5	.7	700	1.8	25.50	BiCMOS, µP
7	2.8M	.65	102	7.5	.7	700	1.8	8.60	CMOS, µP
8	3.1M	.7	170	7.5	.7	900	1.8	32.60	BiCMOS, µP
9	1.2M	.65	250	7.5	.7	700	1.8	21.10	CMOS, µP
10	.91M	.8	400	7.5	.7	1500	1.8	115.00	BiCMOS,VSP
11	6.2M	.35	36	7.5	.9	500	1.8	0.93	SRAM, 1Mb
12	4.1M	0.6	35	7.5	.9	400	1.8	1.08	DRAM, 4Mb
13	264M	.25	29	7.5	.9	600	1.8	1.31	DRAM, 256Mb
14	264M	.25	29	10	.7	600	1.8	2.18	DRAM, 256Mb
15	40k	.8	500	7.5	.7	1200	1.8	43.10	G.A., 53kg
16	1.4M	.8	245	7.5	.7	1200	1.8	51.10	SOG, 177kg
17	7.2k	.8	2.6k	7.5	.7	1300	1.8	240.00	PLD, 1.2 kg

The results shown in Table 3 lead to at least two important conclusions. First is that the cost per transistor of a memory is very different and much lower than for all other IC types. Therefore, any discussion or decision made based on the memory cost data should not be extrapolated onto other types of ICs. The second conclusion is that there is a range of design/manufacturing scenarios which could be manipulated for cost minimization purposes. This opportunity should be actively explored, especially during IC/system design phase because possible gains are larger than one could anticipate (Compare for instance rows 4,7, 10 and 17).

D. Cost Estimation Conclusions

The overall conclusion of this section, and one of the most important messages of this paper, is as follows. Since memories are the largest "consumers" of the "process and equipment" R&D, the economic realties of memory production have decisive influence on the momentum of the entire IC industry. However, from the argument presented it is evident that what is cost effective for memories is not necessarily beneficial for non-memory products. It is also evident that under some circumstances continuation of the transistor cost decrease may not be possible for the redundancy free ICs. Consequently, one can expect dramatic changes in both the business and technology segment of the IC industry, which cannot benefit from high volume, mono-product, high yield manufacturing operations. And among these changes one can expect to see a growing importance of the design based manufacturing cost optimization.

V. Possible Cost Increase Consequences

There are a number of avenues the IC industry can take to handle cost increase problem. In this section we will focus only on one of them - the avenue which is probable but perhaps should be avoided. This avenue, which one could call "past-momentum-driven" can be outlined by the diagram in Fig. 9.

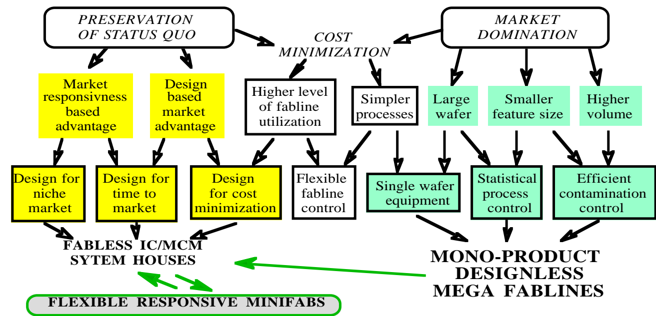


Figure 9. Objectives of two major sectors of IC industry.

It can also be described as an evolutionary process composed of the four phases:

Phase 1. (Taking place now.) In this phase two major trends will dominate the industry. The first one which is driven by "high volume" and "high profit" market winners will aim at smaller feature size and higher volume regardless of the required investment levels. The main premise of this trend seems to be in the "invest-now-to-dominate-later" strategy. Smaller silicon houses which can not spent 1 billion dollars to build a new fabline and have to "preserve the status quo" will participate in the second trend. The main objective of this trend will be to find niche markets and to cut cost in brute force ways ("down sizing", no money for research, open market but far from satisfactory CAD tools, consultants rather then employees, attention paid to the "bottom-line" only, etc.). Observe that the interest of these two trends are conflicting. The first trend, by the virtue of high manufacturing volume forces process R&D and the entire equipment arena to serve its interests. Consequently, producers of low volume, diverse,

redundancy free, niche oriented ICs have to accept the equipment designed and built with undesirable manufacturing requirements and objectives in mind. Worst of all, this equipment will have to be expensive. (The worst of the attributes of such equipment is, however, high throughput which indirectly leads to very low utilization levels in any diverse low volume manufacturing operation. See [12] and Sec. III)

Phase 2. *"True and smart cost cutting effort stage,"* In this phase winners of the race towards smaller feature size will be forced to maintain very high volume production to recover huge past investments. It will be done by expanding portfolio of products (sophisticated memory architectures [29] , variety of µPs versions, FPGA, etc.) and eventually renting superfluous fabline capacity. It will also lead to serious investment in process simplifications, CIM, contamination control , etc. The winners of the niche market competition will become fabless or will still attempt to survive by highly efficient use of obsolete fablines. They may also attempt to influence the equipment arena to provide more cost effective process/equipment solutions. They probably will also invest in such manufacturing cost cutting directions as computer aids in rapid yield learning, DFM and flexible fabline control.

Phase 3. *"End of the technology race and beginning of stagnation stage"*. At this phase the market will be served by many fabless IC/MCM houses and a few "designless" mono-product mega fablines. But "one-size-fits-all" technologies will not be optimal for some applications. As a result the need for custom/niche processes will become apparent.

Phase 4. *"New beginning."* New software/system/circuit/process co-synthesis driven design/manufacturing operations will emerge as a response to the growing gap between fabless design houses and designless silicon vendors.

Of course, it is likely that some of above phases will overlap. Observe also that many of the R&D objectives listed in Fig. 10 are already seen as key and have bee identified as R&D goals for research consortia and universities [17].

The most important conclusion which perhaps could be derived from the above vision is that one should expect a period of time in which design/test/CAD based market advantage will not be seen as adequate to survive. The attention will be paid to the business side of IC design and manufacturing rather then to the innovation and research. Consequently, if the above vision materializes in its worst form, one can expect that rapid changes in the design manufacturing interface will inflict damage in the form of unnecessary unemployment among the best, an irreversible change in research and academia and consequently stagnation in the entire industry.

VI. COST RELATED DESIGN/TESTING/CAD R&D OBJECTIVES

The key message presented so far in this paper is that the escalating cost of IC manufacturing must be "contained" and that it is in the best interest of design/test/CAD to help the IC industry in the coming cost driven transformation. Such a need has not been addressed thus far properly because of two basic reasons: lack of integration of system testing and IC-manufacturing objectives and lack of adequate cost models. One can provide a number of examples supporting such a point.

The best of them is the situation in the MCM arena. MCMs are still seen as packaging domain and not as a cost saving system/IC design opportunity. For instance, one can demonstrate that by applying active silicon substrate (i.e. very expensive substrate) one can build a smart substrate system [30] which can minimize the overall system cost [31] by performing self testing and enabling cost savings impossible with cheaper but passive substrates. But traditional MCM strategies focus on the cost of the substrate itself. Consequently, overall system level cost gains are not clearly identified and typical MCMs are seen as more expensive way to package small and medium size systems.

There are many more examples which clearly show that the lack of integration between design, test and manufacturing is, on the one hand, a reason for unnecessary cost increases and, on the other hand, a great cost saving opportunity. Consider, for instance, testing. DFT and BIST techniques exist to minimize cost and complexity of test generation. But designers are wary to allocate the resources (such as silicon area, and/or performance) required to employ these techniques. The problem is lack of adequate procedure which quantifies the benefit (such as the associated decrease in test generation cost, increase in overall test quality, or impact on time to market) which any BIST of DFT technique would provide in return.

The core of the problem is that the system level cost minimization is possible if, and only if, cost modeling strategy, integrating in a single model such quantities as: yield of the system's components, expressed in terms of all strategic design variables (λ, N_{tr} etc.), cost of testing as a function of the probability of fault escapes [32] and many others, is available.

Fig. 10 depicts all major elements of such a strategy by listing cost models needed for overall system level cost optimization.

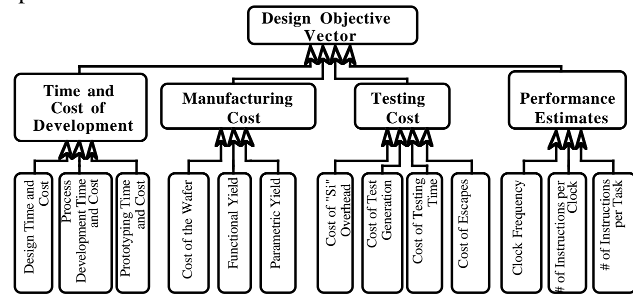


Fig. 10. Cost models needed for system/IC level cost minimization purposes.

The elaborate discussion of each model mentioned in Fig. 10 is beyond the scope of a single paper. It is necessary, however, to point out a couple of problems which are within the expertise of the VLSI design/testing/CAD domain and have not received adequate attention even if they are of a very important nature. The first is very low level of "manufacturability awareness" among CAD tools developers - especially those who design high level synthesis tools. The second problem is lack of design quality - productivity trade off models. The third problem is the lack of adequate testing cost models. Finally, there is a need for an increase of the "communication bandwidth" between all levels of design

abstraction. And from the discussion presented in this paper it should be evident that all of the above problems should be addressed by VLSI community with adequate attention and resources.

VII. CONCLUSIONS

The goal of this paper was to provide an overview of the silicon cost related issues and to initiate the discussion analyzing possible consequences of the anticipated dramatic increase of the IC manufacturing facility. The presented facts indicate that VLSI design domain may be strongly affected by the silicon cost changes. One of the consequences of such a change will be a necessity to refocus VLSI design/test/CAD on the cost related problems, i.e. on (a) development of cost models which are of analytical (not accounting) nature , (b) development of tools performing system level design cost optimization and (c) development of tools and strategies helping to identify and then minimize cost of the manufacturing (yield learning, process simplifications etc.).

The diagram in Fig. 11 attempts to graphically summarize areas of cost minimization opportunities which should be explored in a response to the growing costs of manufacturing.

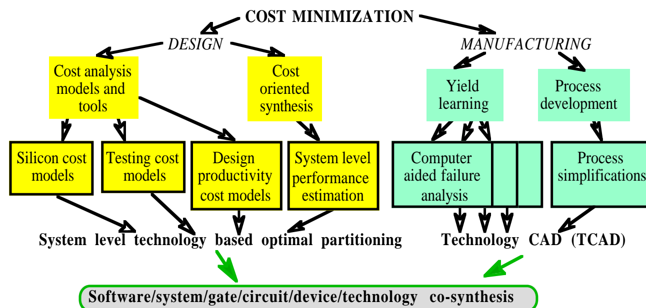


Fig. 11. Cost minimization tasks.

This paper indicates also that ultimately, VLSI design/test/CAD should be directed towards development of the system/circuit/device/layout/process co-synthesis strategy which is needed [33] to support the concept of a cost effective highly flexible fablines - the only alternative to mega-fablines with mega-dollars price tags.

Acknowledgments

The author would like to thank his students: Anne Gattiker, Hans Heineken, Jitendra Khare and Tom Marchok for help in the preparation of this paper. He also would like to acknowledge the support of SRC, Sematech, National Semiconductor, Intel, Philips, Siemens Corporations and Deutsche Forschungsgemeinschaft.

References

[1] M. Komiya, "Future Technological and Economic Prospects for VLSI, " ISSCC 93, pp. 16-19.

[2] C. R. Barrett, "Microprocessor Evolution and Technology Impact," 1993 Symposium on VLSI Technology, May 17-19, Kyoto, pp. 7-10.

[3] W. J. Spencer, "National Interests in Global Semiconductor Industry," Keynote Speech, IEDM-93, Washington D.C., Dec. 5-8, 1993.

[4] Word News, Solid State Technology, September 1993.

[5] "The Siege of Intel," The Economist, February 12, pp. 63-64.

[6] "Status 1993 - A report on the Integrated Circuit Industry," Integrated Circuits Engineering Corporation.

[7] C. R. Deninger, "Fabs of the 2000," Techcon 90, San Jose, Oct. 16-18, 1990 pp. 276-278.

[8] T. Masuchara, K. Itoh, K. Seki and K. Sasaki, "VLSI Memories: Present Status and Future Prospects," IEICE Trans. Vol. E 74. No 1, January 1991.

[9] V. Menon, "Microcontamination, Defect Measurement and Control in ULSI Manufacturing," Techcon 90, San Jose, Oct. 16-18, 1990 pp. 224.

[10] W. Maly, "Evolution of Microelectronics from the Cost of Manufacturing Perspective," Research Report No. CMUCAD-91-16, September 1991.

[11] Y. Tarui, et al., "New DRAM pricing trends: The Bi rule," IEEE Circuits & Design, March 1991, pp. 44-45.

[12] W. Maly, H. Jacobs and A. Kersch, " Estimation of Wafer Cost for Technology Design," IEDM-93, Washington D.C., Dec. 5-8, 1993, pp. 35.6.1 - 35.6.4.

[13] I.A. Saadat, M.T. Thomas, A. Gattiker and W. Maly," Wafer Cost Modeling and Analysis for Strategy Planning Purposes," Eight Annual SRC/ARPA CIM-IC Workshop, Aug. 1993.

[14] L. Gwennap, "Estimating IC Manufacturing Costs," Microprocessor Report, Aug. 2, 1993, pp. 12-16.

[15] J. N. H. E. Weste and K. Eshragian," Principles of CMOS VLSI Design: A system Perspective," Addison-Wesley Pub. Co., Second edition, 1993.

[16] E. Neacy, et. al., "Cost Analysis for Multiple Product/Multiple Process Factory: Application of SEMATECH's Future Factory Design Methodology," 4-th Annual ASMC, Boston, Oct. 18-19, pp. 212-219

[17] SIA Technology Road Map - Workshop Conclusions; November 1993.

[18] M. Ogirima, "Process Innovation for Future Semiconductor Industry," 1993 Symposium on VLSI Technology, May 17-19, 1993, Kyoto, pp. 1-5.

[19] M. Moslehi, et al., "Microelectronics Manufacturing Science and Technology (MMST) : Single Wafer RTP-Based 0.35 mm CMOS Fabrication", IEDM-93, Washington D.C., Dec. 5-8, 1993, pp. 27.1.1 - 27.1.4.

[20] A.V. Ferris-Prabhu, "Parameters for optimization of device productivity at wafer level," *IBM Burlington Technical Bulletin*, TR 19.90488, Nov. 1989.

[21] H.T. Heineken and W. Maly, "Manufacturability Analysis Environment - MAPEX", to appear in Proc. of CICC-94.

[22]. F. Abu-Nofal, et. al. "A Three Million-Transistors Microprocessor ", 1993 ISSCC, Feb. 1993, pp. 108-109.

[23] IEEE Spectrum, Dec. 1993, pp. 20-25.

[24] Proc. of ISSCC 1991, 1992, 1993 and CICC of 1989.

[25] W. Maly, "Computer-Aided Design for VLSI Circuit Manufacturability," Proc. of IEEE, Vol. 78, No. 2, Feb. 1990, pp. 356-392.

[26] W. Maly, H.T. Heineken and F. Agricola," Yield Model for Manufacturing Strategy Planning and Product Shrink Applications," to be published.

[27] T. W. Williams," Test Technology 20 Years and Beyond", ITC1989, Washington D.C. p.7.

[28] P.K. Chatterjee and G.B. Larrabee, " ULSI Research and Development in the United States: Status and Outlook," Proc. of the Fourth International Symposium on ULSI Science and Technology, 1993, pp. 1-19.

[29] B. Prince, "Memory in the Fast Lane", IEEE Spectrum, Feb. 1994, pp. 38-41.

[30] W. Maly, D. Feltham, A. Gattiker, M. Hobaugh, K. Backus and M. Thomas, "Multi-Chip Module Smart Substrate System," Accepted for publication in IEEE Design and Test Magazine, June 1994.

[31] A. Gattiker, W. Maly and M. Thomas, "Are There Any Alternatives to Known Good Die?", to appear in Proceedings of IEEE Multi-Chip Module Conference, March 1994.

[32] T.E. Marchok and W. Maly, "Automatic Synthesis and the Cost of Testing", to appear in Proc. of CICC-94.

[33] Z.J. Lamnios, Beyond MMST: The virtual factory," Solid State Technology, Feb. 1994, pp. 25-26.