Design-Manufacturing Interface: Part II - Applications

W. Maly, H.T. Heineken[1], J. Khare[1], P.K. Nag, P. Simon[2] and C. Ouyang.

Dept. of ECE, Carnegie Mellon University, Pittsburgh, PA 15213
[1]Level One Communications, Product Manufacturing and Test, Sacramento, CA 95827
[2]Philips NV, Product Engineering Dept., Nijmegen, The Netherlands

Abstract

This paper illustrates via examples problems at the design-manufacturing interface that exist in the IC industry today, and the ability of the YAN/PODEMA framework [1] in solving these problems. The need for further development of the framework is also emphasized.

1.0 Introduction

In [1] and [2] it was argued that an efficient channel of communication between VLSI design and IC manufacturing is a must in today's rapidly changing IC markets and technology arenas. This requirement is well-recognized by product engineers and others responsible for the timely delivery of a given volume of ICs. However, it ranks low on the agenda of IC designers, and even lower among CAD developers - both of whom are used to “design rules” and “spice files” defining the domain of acceptable designs. As a consequence, manufacturability-oriented design aids and methodologies are under-developed. This causes design-manufacturing communication inefficiencies, which can be costly.

The discussion presented in [1] leads to the conclusion that there are a number of steps which should be undertaken to address the above concern. First, a clear definition of the design-manufacturing information exchange mechanism is needed. Second, after the first is accomplished, comes the development of the required logistic procedures and software tools.

The objective of this paper is to demonstrate that the design-manufacturing interface must and can be changed. To achieve this goal, this paper will begin with a very brief summary of the proposal for a new design-manufacturing paradigm suggested in [1]. The remainder of the paper will discuss a number of actual design-manufacturing interface problems, (some of them discussed in detail in [3]) encountered in industry. The descriptions of these problems (limited in detail by their proprietary nature) will be used (a) to illustrate the need for a change in the design-manufacturing interface paradigm, and (b) as examples of possible approaches to this change.

2.0 Yield Analysis (YAN) and Post-Design Manufacturability Assessment (PODEMA)

The discussion of the design-manufacturing interface presented in [1] can be summarized with the conclusion that the two key capabilities of this interface are:

(a) Yield Analysis (YAN) - which should assist designers/product engineers with the task of uncovering physical reasons for the observed yield loss, using attributes extracted from the designs and then correlating them with yield.

(b) Post Design Manufacturability Assessment (PODEMA) - which should predict the manufacturability of new designs before they are fabricated. This is done by extracting design attributes and using them in conjunction with yield models developed with YAN.

In [1], the structures of both YAN and PODEMA are discussed in detail. This paper describes how YAN and PODEMA-like tasks were solved in industrial settings using currently available tools. The examples are designed to not only illustrate the feasibility of YAN/PODEMA, but to also point out the inadequacy of existing tools, and emphasize the need for a robust framework at the design-manufacturing interface.

3.0 YAN/PODEMA Applications

3.1 YAN Applications

In this section three applications of YAN are discussed which deal with a variety of yield problems. It is important to note, though, that these examples illustrate only a fraction of design-sensitive yield loss mechanisms. In general, for a given combination of design and fabrication line, a large number of yield-related design attributes must be extracted and correlated to yield, to determine the dominant yield loss mechanisms [2].

3.1.1 Critical Area

In this sub-section, design attributes are used to model the dominant yield loss mechanism in mature stable manufacturing lines - i.e., random spot defects, which cause functional failures.

Example 1: Traditional yield loss models (e.g., Poisson model [4], Seeds model [5], negative binomial model [6]) or their variants are still the dominant models used in fabrication lines to estimate product yields. This is predominantly because of their simplicity. These models, which are based on the point defect model, have one key characteristic - as die area increases, yield decreases. This naturally leads designers to pack their designs in as small a die area as possible in order to reduce die cost. While these models may have worked well in the past, they do not necessarily work well in sub-micron manufacturing lines. Fig. 1 shows the yield vs area curve for eight products (with transistor counts ranging from 71K to 224K) manufactured in high volume in the same manufacturing line over a period of one year [7]. In this graph, yield *increases* with die area. No tra-

ditional yield model can account for the phenomenon. Clearly, better models are needed to explain such a trend.

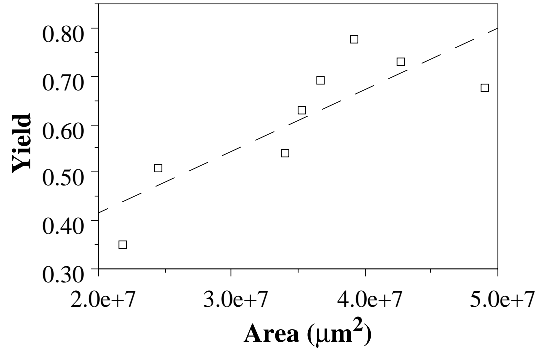

Fig 1. Yield vs. area curve for eight products in a sub-micron fabrication line [7].

To explain the above data, it was decided to use the spot defect model [8], which assumes defects to be disks of extra or missing material in conducting, semiconducting, or insulating IC layers. Defects of a particular type (i.e., an extra or missing material defect in some IC layer) are assumed to have as parameters (a) size distribution function $f_i(r)$ which specifies the variation in frequency between defects of different radii [9], and (b) density, D_{oi}, which specifies the frequency of occurrence of defects of type i (for details see [9], [10]).

The sensitivity of an IC to defects of various sizes and types is modeled by *critical area* [8], [11], [12], [13]. The critical area $A_{ci}(r)$ for a defect of radius r is defined as the area of a die where if the center of the defect is deposited a fault occurs in the circuit. The exact shape of the critical area function depends on the layout density in the given layer. Fig. 2 shows examples of critical areas for extra material defects in the metal 1 layer for five out of the eight designs in Fig. 1.

Based on the spot defect and critical area models, die yield in the presence of n defect types can be expressed as [8]:

$$Y = \prod_{i=1}^{n} \exp\left(-D_{0i}\int_{0}^{\infty} A_{ci}(r) \cdot f_i(r) \cdot dr\right) \qquad (1)$$

This model was applied to the data shown in Fig. 1 [7]. For the fabrication line under consideration, shorts were much more prevalent than opens (i.e., much more extra material defects). In addition, critical areas for extra material defects in poly and the metal layers were found to correlate well with each other. Hence, it was decide to model yields using critical area functions for extra material metal 1 defects only. These were extracted using the YAN/PODEMA engine, MAPEX-2 [14].

Based on (1) and the critical area functions, defect parameters D_0 *and* $f(r)$ were extracted to obtain the best least-squares fit of the yield data [7]. Fig. 3 shows the fit between computed and measured values. As can be seen, the yield model fits the data very well.

In addition, extracted defect parameters were used to estimate the yields of two other designs (D9 and D10) fabricated in the same line. As seen in Table 1, the model could predict the yields of new designs as well.

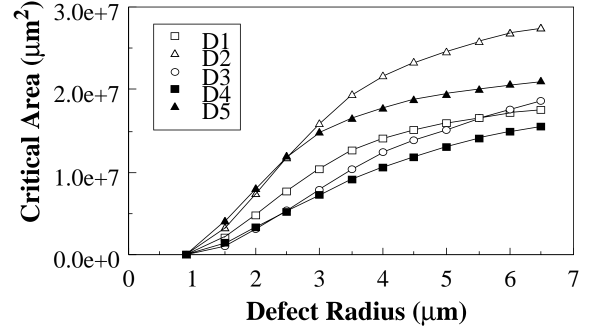

Fig 2. Metal 1 critical area for shorts vs. defect radius for five designs from Fig. 1.

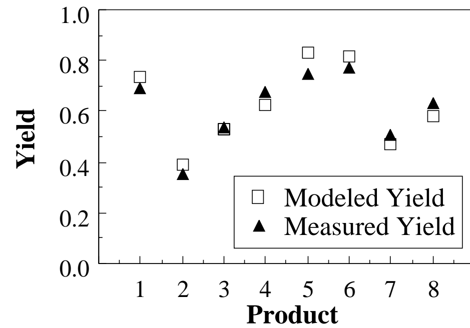

Fig 3. Measured and computed yields of the eight products in Fig. 1.

Design	Computed Yield	Measured Yield
D9	0.381	0.389
D10	0.621	0.693

Table 1. Computed and measured yields of two new designs based on yield modeling performed on eight designs in Fig. 1.

Experiment 2: As an additional validation that critical areas of shorts are a good measure of yield, another experiment was performed [15]. Particle data from KLA equipment was used to extract defect parameters D_{0i} and $f_i(r)$ for the key layers of poly, metal 1 and metal 2. For these layers, the critical area functions for extra material defects were also extracted using MAPEX-2. (See example for one product in Fig. 4). Then the model in (1) was used to predict yield, using the data for these three layers. The results are shown in Fig. 5. As can be seen, predicted yield tracked

the measured yield relatively well, indicating that critical areas were indeed the key yield attributes for this high volume line. The difference between measured and calculated yield (less than 8% [15]) was perhaps due to mechanisms not captured by the critical area-based models.

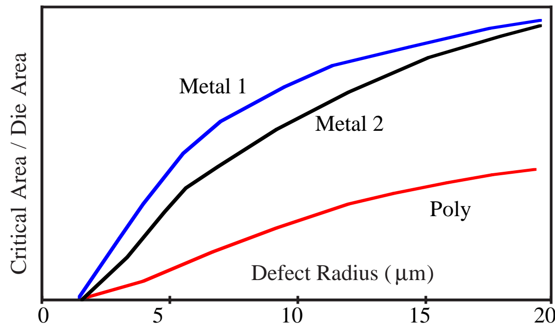

Fig 4. Critical area for poly, metal 1 and metal 2 layers for one of the products.

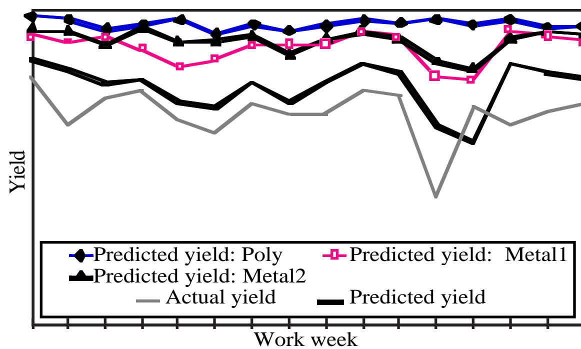

Fig 5. Yields (measured and predicted) as a function of time.

Thus, one can conclude that the critical area model is very effective in predicting yield loss in a mature, high volume manufacturing line. Any deviation from critical area-based yield would thus indicate the presence of other dominant yield loss mechanisms, which could then be further analyzed using YAN. An example of this is described in the next sub-section. In addition, since critical area is the yield determining factor instead of area, it is not necessarily obvious that designers should pack the layout as much as possible. It could be the case that a lower layout density may lead to higher yield, and more working die per wafer. The exact density can be determined by doing an analysis similar to this example (see [16]).

3.1.2 Metal Utilization

The YAN example in this sub-section uses the critical area model to extract defect-related yield loss out of the total yield loss. A systematic approach is then taken to isolate the remaining yield variations. The example involves three different products (A, B & C), manufactured in the same CMOS manufacturing line and thus exposed to similar manufacturing disturbances [17]. Product A was designed for 0.8µm CMOS technology with 2 metals, product B was designed for 0.5µm technology with 2 metals, and product C was designed for 0.5µm technology with 3 metals. Product C was designed primarily by reusing modules designed for a 2 metal technology. The interconnection between these modules and some global routing were done in the metal 3 layer.

Initial yield analysis was performed on product A, for which yield data was available for several test bins. This analysis (based on MAPEX-2 extracted critical areas and the yield model described in Section 3.1.1) indicated that yield loss was dominated by extra material spot defects in the two metal layers. Defect model parameters were then extracted using eq. (1) so as to match the measured yields of Product A. These parameters, along with critical areas of metals for product B, were then used to calculate the yield for product B. As seen in Fig. 6, the computed yield of product B was close to its measured yield, which confirmed that the defect parameters extracted for the 0.8µm process could be well extended for the 0.5µm process. In addition, in-line measurements using particle monitors for the 0.5µm process also matched the defect densities derived from extending those for the 0.8µm process.

In the next step the yield of product C was computed using the same procedure as for product B. The metal 3 defect density was assumed to be of the same order of magnitude as that for first and second metal layers (particle count in metal 3 was similar to that in the first two metals). However for product C, the measured yield was found to be much lower than the computed yield (Fig. 6). Hence, it was concluded that product C suffered not just from critical area-related yield loss, but from some other yield loss mechanism not affecting products A and B.

To find this mechanism, an extraction of design attributes of product C was carried out. This revealed an interesting fact - the third metal layer was found to occupy only about 6% of the die area, compared to about 40% of the area covered by metal 1 and metal 2 for all the three products. Further investigation confirmed that poor metal 3 utilization caused uneven etching rates in certain regions of the chip, leading to some metal 3 lines being more susceptible to opens. This effect, called etch-loading, is well known in manufacturing and can be observed in most modern semiconductor processes. To eliminate this yield loss mechanism, it was decided to add grounded dummy structures to the metal 3 layer to improve utilization.

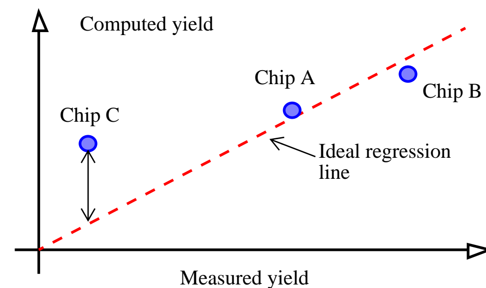

Fig 6. Computed vs. measured yields of the three products A, B, C.

3.1.3 Etch Rate Variation [3]

Yield loss is often caused by non-uniform distributions of process conditions across lots or across wafers. When this happens, the population of the manufactured dies can be divided into groups, each with a distinctive process history. (For example, dies at the edge of the wafer can be grouped in a separate set.) Yield loss figures can then be determined separately for each group. Such a separation can be very useful in determining the physical reason for the analyzed yield loss mechanism. The yield analysis study in this section illustrates the above observation.

In this analysis of a CMOS product, it was observed that yield tended to be lower at the center of the wafer than at the edges. The extent of this effect is shown in the wafer map in Fig. 7. (In the wafer map, each square represents one die. The size of the square indicates the yield for that die over a large number of wafers. Higher the yield, larger the square.)

For this product, yield data was also available in two bins - on-chip SRAM yield, and the yield of the rest of the logic. The wafer maps for the SRAM and logic yield losses are shown in Fig. 8. As seen from the figure (and from data in Table 2) the change in yield loss from edge to center was larger for the SRAM than for logic.

To explain the above yield loss pattern, a number of attributes were extracted for both the SRAM and the logic. It was found that the yield loss at the edge of the wafer for both, logic and SRAM, could be explained using critical areas of the various layers and defect statistics. The extra yield loss at the center, however, correlated well with another attribute - the density of deep contacts (contacts between metal 1 and active area) in the SRAM and logic (see Table 2 for values). It was hypothesized therefore that there was a problem with the uniformity of the contact etch process. At the center, the etch rate was smaller than at the edge. As a result, in areas with large densities of contact holes (for instance, SRAMs) the concentration of the etchant was not sufficient to open deep contacts.

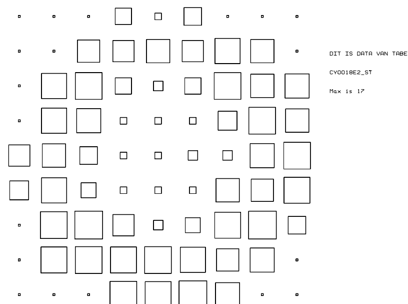

Fig 7. Stacked wafer map of yield for a product. (Each square represents a die. Larger square means higher yield at that location.)

This hypothesis was confirmed through failure analysis, and by the fact that the etch-rate pattern measured for the etch tool was similar to the wafer maps shown in Fig. 8.

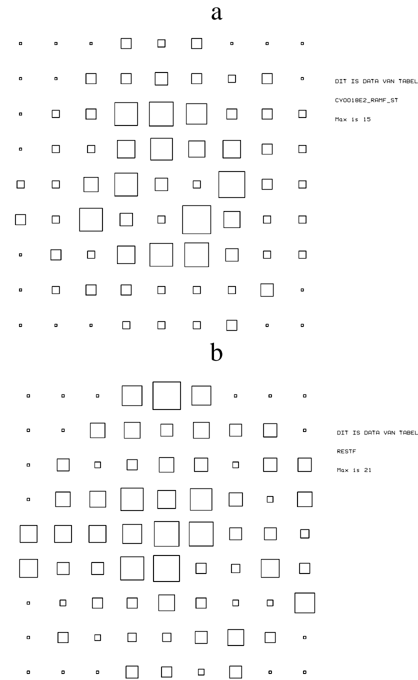

Fig 8. Stacked wafer map of the number of fails for (a) the SRAM and (b) the logic of the IC. (Each square represents a die. Larger square means more fails at the location.)

An interesting observation in this example was that to minimize the problem, action could be undertaken on both the design side and the manufacturing side. Either the etching recipe could be changed (longer etch time and more uniformity), or a sizing operation could be done on deep contacts (especially for SRAMs) to change their aspect ratio, thus allowing for a wider process window.

	Yield(%)		**Deep Contacts**	
	Edge of Wafer	**Center of Wafer**	**Number (Relative)**	**Density (Relative)**
SRAM	85	33	1.0	1.0
Logic	75	50	1.4	0.47

Table 2. Data on yield and deep contacts for the SRAM and logic.

3.2 PODEMA Applications

This section describes two applications of PODEMA. It is important to note that for PODEMA to be applied, it is necessary for yield models to have already been developed through YAN, and for the designer to have some knowledge of the dominant yield loss mechanisms in the fabrication line. If this is the case, only the relevant attributes need to be extracted from the design instead of a large list of generic yield-related attributes.

The two examples described here deal with (a) antenna effect, and (b) etch loading, both of which are common

problems in modern sub-micron processes.

3.2.1 Antenna Effect

In modern semiconductor manufacturing, chemical etch processes have been completely replaced by plasma-based etch processes. These processes are able to achieve very fine feature sizes. However, their one drawback is that their tendency to charge floating conducting regions (i.e., regions which are not connected to any diffusion areas, but only to gate areas). This charge accumulates on the gate, leading to high electric fields across the gate oxide. If the charge is sufficiently large, it can cause gate oxide breakdown or hot electron generation, leading to yield/reliability problems This effect is called the antenna effect [18], [19].

The magnitude of the gate charge depends on the ratio of the floating conducting region area to the gate area (called the antenna ratio). Typically, most fabrication lines specify an upper limit on the antenna ratio for the circuit to function properly. For modern manufacturing lines, this ratio is in the range of 100-300.

There are two methods which can be used to eliminate the antenna effect. The first method is to connect each gate to reverse biased p-n diodes. In this case, the plasma-induced charge is discharged through the diodes, leaving little charge on the gate. The second method is to take gate connections directly up to the top level metal through stacked vias. In this case, the gate will not see any antenna, since at the top metal all conducting regions get connected to at least one diffusion area.

Both the above methods have drawbacks. The use of antenna diodes can increase die area by as much as 4-6%. The use of stacked vias can lead to routing congestion, and hence also to an increase in die area. In addition, the use of a large number of additional stacked vias can potentially lead to reliability problems.

Another method which seems feasible is to route the design without considering the antenna effect. PODEMA could then be used for a post-route analysis to identify connections which exceed the recommended antenna ratio limit. These could then be ripped up and rerouted to achieve the final result. The increase in die area would not be as drastic as in the above mentioned cases. However, since the distribution of antenna ratios depends strongly on the place and route environment, it would then also become necessary to use an environment which minimizes antenna problems.

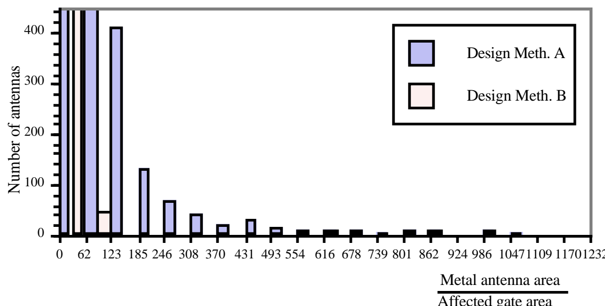

Fig 9. Metal 1 antenna area ratios for two different design environments.

As an example to demonstrate this, an experiment was performed [20]. A 900,000+ transistor CMOS design was routed in two different CAD environments (both with three layers of metal) to create two layouts, A and B. From these two layouts, antennas were extracted using the PODEMA engine MAPEX-2. As shown in Fig. 9., the design environment A creates more large antennas than B. Clearly, if the antenna effect was a significant problem in the fabrication line, A would be the preferred design.

3.2.2 Top Metal Utilization/Etch Loading

In Section 3.1.2, YAN was used to identify yield loss due to poor utilization in the top metal layer. Such poor utilization causes excess etching of the metal lines, leading to formation of opens. In order to prevent this, fabrication facilities usually specify a minimum utilization ratio for each metal layer.

In the previous example, poor utilization was caused by the third metal layer being used primarily for connecting re-used modules which had been designed with two metal layers. For designs done from ground up, this is typically not the case, since modern routers try to achieve uniform utilization in all metal layers. Examples of two modern designs routed using one CAD environment are shown in Table 3. As can be seen, utilization is similar for all four metal layers.

However, there are cases where even modern routers can create designs with poor top metal utilization. For example, a design re-routed in more metal layers to meet die size constraints (which is limited by the desired package's cavity size) can lead to such a scenario. Table 4 shows the same designs as in Table 3, but routed in 5 layers of metal. As shown, using an extra metal layer reduced die size. This reduction may be significant enough to reduce overall chip cost (especially if a smaller cheaper package can be used), which would justify using the extra layer. However, metal 5 utilization in both designs was less than 2.5% (compared to an average of 22% for other metals), resulting in susceptibility to etch loading-related yield loss.

Design	Metal Utilization [%]			
	Metal 1	**Metal 2**	**Metal 3**	**Metal 4**
Design A	43.8	13.9	16.3	11.4
Design B	38.8	12.5	18.2	13.8

Table 3. Metal utilizations of two designs routed with 4 metal layers.

Design	Area Saving	Metal Utilization [%]				
		Metal 1	**Metal 2**	**Metal 3**	**Metal 4**	**Metal 5**
A	4.13	44.8	14.4	19.4	10.3	1.7
B	8.66	41.7	14.4	19.6	11.4	2.2

Table 4. Metal utilizations of the designs in Table 1 routed with 5 layers of metal.

PODEMA can be used to identify such designs, and flag warnings. If metal utilization is unacceptably small, grounded dummy metal structures can be inserted to minimize yield loss due to etch loading.

4.0 Conclusions

This paper presents a number of yield analysis case studies conducted on actual VLSI products in a diverse range of manufacturing environments. Each analysis was performed using yield-relevant attributes such as critical areas, densities of contact cuts, number of antennas, etc. These attributes were extracted from the design and then used in a yield model or in a yield loss hypothesis which was verified using yield measurements. In each case, a new, useful result was obtained.

The analyses carried out for the case studies also indicated the following:

1. The design attribute extraction environment is a very useful element of both YAN and PODEMA.

2. The design attribute extraction environment must be flexible to be able to extract any needed set of design attributes.

3. Although tools exist to perform cross-correlational analysis in an ad-hoc fashion, a robust framework for YAN/PODEMA analysis is necessary.

In addition, all the results presented in the paper seem to support the conclusion that YAN and PODEMA are feasible and are potentially useful additions to the design-manufacturing interface. Such an addition should substantially enhance the accuracy and the rate of information exchange at this interface.

Acknowledgments

The authors would like to thank: E. Bruls from Philips NV, Product Engineering Dept., Nijmegen, L. Milor and Y.T. Lin from Advanced Micro Devices, Submicron Development Center, Sunnyvale, CA, S. Maturi and D. Ahrens from National Semiconductor, Product Development Automation, Santa Clara, CA, K. Estermaier, G. Muller-Liebler from Siemens A.G., CAD Dept., Munich, Germany, and D. Schmitt-Landsiedel from Technical University of Munich, for their technical and financial contributions to the research presented in this paper.

References

[1] W. Maly, H. T. Heineken, J. Khare and P. K. Nag, “Design-Manufacturing Interface: Part I - Vision,” *Proceedings of DATE*, Feb 1998.

[2] .H. T. Heineken and W. Maly, “Manufacturability Analysis Environment - MAPEX,” Proc. of Custom Integrated Circuits Conference, pp. 309-312, May 1994.

[3] W. Maly, H.T. Heineken, C. Ouyang, A.J. Strojwas, E. Bruls, P. Simon, L. Milor, Y.T. Lin, S. Maturi D. Ahrens, K. Estermaier, G. Muller-Liebler, D. Schmitt-Landsiedel, J. Khare and P.K. Nag, “Design-Manufacturing Interface - Problems and Solutions” in *Design for Manufacturability for the Next Decade*, W. Maly Ed., Research Report: CMUCAD-97-10.

[4] R. M. Warner Jr., “Applying a Composite Model to the IC Yield Problem,” *IEEE Journal of Solid State Circuits,* vol. SC-9, no. 3, pp. 86-95, June 1974

[5] R. B. Seeds, “Yield, Economic, and Logistic Models for Complex Digital Arrays,” *IEEE International Convention Record*, pp. 60-61, Mar 1967.

[6] C. H. Stapper and R. Rosner, “A Simple Method for Modeling VLSI Yields,” *Solid-State Electronics*, vol. 25, no. 6, pp. 487-489, June 1982.

[7] H. T. Heineken, J. Khare and W. Maly, “Yield Loss Forecasting in the Early Phases of the VLSI Design Process,” *Proceedings of CICC-96*, May 1996.

[8] W. Maly and J. Deszczka, “Yield Estimation Model for VLSI Artwork Evaluation,” *Electronics Letters*, vol. 19 no. 6, pp. 226-227, Mar 1983.

[9] A. V. Ferris-Prabhu, “Defect Size Variations and Their Effect on the Critical Area of VLSI Devices,” *IEEE Journal of Solid-State Circuits*, vol. 20 no. 4, pp. 878-880, Aug 1985.

[10]J. Khare, W. Maly and M. Thomas, “Extraction of Defect Size Distributions in an IC Layer using Test Structure Data,” *IEEE Trans. on Semiconductor Manufacturing*, vol. 7, no. 3, pp. 354-368, Aug 1994.

[11]C. H. Stapper, “Modeling of Integrated Circuit Defect Sensitivities,” *IBM Journal of Research and Development*, vol. 27 no. 6, pp. 549-557, Nov 1983.

[12]A. V. Ferris-Prabhu, “Modeling the Critical Area in Yield Forecasts,” *IEEE Journal of Solid-State Circuits*, vol. SC-20, no. 4, pp. 874-878, Aug 1985.

[13]W. Maly, “Modeling of Lithography Related Yield Losses for CAD of VLSI Circuits,” *IEEE Trans. on CAD of Integrated Circuits and Systems*, vol. 4 no. 4, pp. 166-177, July 1985.

[14]H. T. Heineken, J. Khare, W. Maly, P. K. Nag, C. Ouyang, and W. Pleskacz, “CAD at the Design- Manufacturing Interface,” *Proceedings of DAC-97*, June 1997.

[15]R. K. Nurani, A. J. Strojwas, W. Maly, C. Ouyang, W. Shindo, R. Akella, M. McIntyre and J. Derret, “In-line Yield Prediction Methodologies Using Patterned Wafer Inspection Information,” *Proceedings of the 1996 Int. Symp. on Semiconductor Manufacturing*, pp. 243-250, Oct 1996.

[16]H. T. Heineken and W. Maly, “Performance-Manufacturability Tradeoffs in IC Design”, *Proceedings of DATE*, Feb 1998.

[17]Klaus Estermaier, *Investigation of VLSI Layout-to-Yield Relationship*, Diplomarbeit, Technische Universitaet Muenchen, Germany, Mar 1996.

[18]T. Watnabe and Y. Yoshida, “Dielectric Breakdown of Gate Insulator due to Reactive Etching,” Solid State Technology, vol. 26, no. 4, p. 263, 1984.

[19]S. Fang and J. McVittie, “Thin-Oxide Damage from Gate Charging During Plasma Processing,” IEEE Electron Devices Lett., vol. 13, no. 5, pp. 288-, May 1992.

[20]W. Maly, C. H. Ouyang, S. Ghosh, and S. Maturi, “Detection of an Antenna Effect in VLSI Designs,” in Proc. of the 1996 Int. Symp. on Defect and Fault Tolerance in VLSI Systems.