Copyright © 1985, by the author(s).
All rights reserved.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission.

A NEW, PROFIT MAXIMIZATION METHODOLOGY FOR STATISTICAL DESIGN OF INTEGRATED CIRCUITS
PART I: PROBLEM FORMULATION

by

David Riley and Alberto Sangiovanni-Vincentelli

Memorandum No. UCB/ERL M85/58

27 May 1985

A NEW, PROFIT MAXIMIZATION METHODOLOGY FOR STATISTICAL DESIGN OF INTEGRATED CIRCUITS

PART I: PROBLEM FORMULATION

by

David Riley and Alberto Sangiovanni-Vincentelli

Memorandum No. UCB/ERL M85/58

27 May 1985

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

A NEW, PROFIT MAXIMIZATION METHODOLOGY FOR STATISTICAL DESIGN OF INTEGRATED CIRCUITS

PART I: PROBLEM FORMULATION

by

David Riley and Alberto Sangiovanni-Vincentelli

Memorandum No. UCB/ERL M85/58

27 May 1985

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

A New, Profit Maximization Methodology for Statistical Design of Integrated Circuits - Part I: Problem Formulation

David Riley and Alberto Sangiovanni-Vincentelli

Abstract

A new methodology for the optimization of integrated circuit parameters which takes into consideration realistic statistical fluctuations occurring in their fabrication, is proposed. The methodology is based on a problem formulation which consists of maximizing a criterion of goodness never before used in statistical circuit design. The criterion is one that recognizes the ultimate economic goal of engineering design - profit expected from the production and sale of the product being designed. In Part I this criterion is motivated in the general context of engineering design and of statistical circuit optimization. The form of profit most appropriate to statistical circuit optimization is specified. The design parameters of the methodology are identified, and the dependence of profit on them modeled. In Part II, results in the study of computer efficient techniques for maximizing the profit criterion are presented.

The authors would like to acknowledge the support of Harris Semiconductor, and of the Joint Services Electronics Program at the Electronics Research Laboratory, contract number F 49620-84-C-0057.

1. Introduction

This two-part document presents a new methodology for the parametric design of integrated circuits. The methodology is based on optimization. The problem at hand is precisely formulated in terms of a single *objective* or *criteria* function to be maximized.

A major conclusion of this research has been that attempts to advance the state of the art in the statistical design of integrated circuits by finding new solution techniques for existing criteria would be pointless. The development of a new problem formulation has been a major direction of this research work. Part I presents the results of this work.

The new problem formulation does not in itself eliminate the prospect of prohibitively large cpu time requirements. Sophisticated techniques for minimizing cpu cost are needed, just as they are for previous problem formulations. Part II presents results of the study of algorithm considerations in the efficient optimization of the criterion function presented in Part I.

Specifying the new problem formulation requires specifying three types of information - the quantity which is to serve as the criterion, the quantities which are to serve as the independent variables, and the functional dependency of the former on the latter. Although none of these types of information can be thoroughly discussed without reference to the other two, some degree of decoupling has been possible herein. As the pieces of information making up the new problem formulation are presented, the reader may find it useful to be aware to which element or elements of the optimization problem the information is contributing. At any rate, note that the following three sections emphasize the development of an appropriate optimization criterion, while the sections after that are concerned jointly with specifying the problem variables and the functional dependency.

2. Treatment of Economic Considerations in Engineering Design

In free-market economies, the primary criterion for making decisions on matters having economic implications is the economic gain of the decision maker. (Whether the economic gain is measured as a profit, a return on investment, or some other measure is not material here.) Criteria relating to social conscience can sometimes be important secondary criteria. And laws and regulations sometimes introduce constraints on the decision-making process.

This document is concerned with a particular type of decision-making - detailed engineering design. For the overwhelming majority of the detailed design decisions which are made in the course of an engineering design project, such secondary criteria and legal constraints usually are not important. Even if the broad social implications of an engineering project are very controversial, usually many details of its implementation are socially neutral. For example, whether to build a chemical plant at a particular location may be quite controversial, but the sensitivity of social welfare to the flow rate of some valve within that plant is probably miniscule. This discussion will henceforth be concerned with engineering design methodologies in which economic gain is the sole criteria for making decisions. As has just been argued, use of such methodologies is appropriate for most examples of engineering design.

It is ironic that among the examples of engineering design methodologies of interest here, few use analyses which involve economic variables such as dollars of cost or revenue, or rate of production. Explaining this contradictory situation requires a closer look at the engineering design process.

Engineering design is a process of generating alternative feasible designs, then selecting from them the final design. Alternative designs may differ because of a

different selection either of some qualitative characteristic from a set of possibilities which is finite (or occasionally countably infinite), or of the value of a continuous parameter from among an uncountable infinity of possible values. In this document, the *design characteristics* of a particular product and a particular design methodology for that product, refers to the set of all characteristics of the product which its producer can directly implement in its manufacture, and whose selection is the purpose of the methodology. (*Design parameters* will refer to the subset of the design characteristics that are continuous parameters.) Let these design characteristics (including the qualitative ones) be represented in symbolic terms by the vector **d**.

It would seem that the heart of engineering design would be understanding how the economic gain expected from a product, G, say, depends on its design characteristics. Symbolically yet more concisely, this would involve exploring the function g in

$$G = g(\mathbf{d})$$

in the interest of making the best possible choice of final design. Yet, as suggested above, this is not an accurate characterization of what engineers typically do.

The engineer typically uses as design criteria a number of performance attributes **a**, say, that are of a technical rather than economic nature. He (or she) generally considers the heart of the design problem understanding how these technical attributes depend on design characteristics. This is tantamount to complete knowledge of the function **r** in

$$\mathbf{a} = \mathbf{r}(\mathbf{d})$$

However, economic aspects, as asserted above, cannot be ignored. The way they are typically handled is as follows.

For the purposes of this discussion, the economic implications of alternative engineering designs can be considered to be of two types - benefit implications (revenue, for the case of a product to be sold on the open market), and cost implications. On the economic benefit side, the engineer typically applies his own subjective evaluation of values of the elements of **a**, assuming the evaluations of most users of the product are nearly the same and coincides with his own. He uses these evaluations and comes up with possible design vectors **d**. His treatment of the cost side is generally limited to the use of rough estimates of cost differences between alternative designs.

Now suppose that for the design of a particular type of product engineers had the capability to acquire complete knowledge of the dependencies represented by the function **r**, for any given set of attributes **a**. It is of interest to consider some reactions the engineers would be likely to have to the following set of circumstances:

(1) The set of attributes in **a** is sufficiently complete that knowledge of them and of the design characteristics **d** uniquely determines the economic gain.

(2) the engineers have at their disposal this dependence (economic gain on **d** and **a**). In symbolic terms, this would be tantamount to having the function h in

$$G = h(\mathbf{d}, \mathbf{a})$$

Then the engineers would have at their disposal a set of relationships which, as suggested by the arguments above, ought to be ideal for designing the product. Symbolically, this conclusion is represented by the fact that the engineers have a

function playing the role of g, as results from combining the above equations to give

$$G = h(\mathbf{d}, \mathbf{r}(\mathbf{d}))$$

What is interesting is that most of the engineers would probably not consider this design methodology to be ideal. There are many rationales which might be given for this belief, including some might include the following:

(1) The relationships embodied in the function h are less elegant than those in the function **r**, and hence are less interesting and less worthy of his time.

(2) Using the function h is distasteful because as an engineer, he is not particularly interested in economic analyses.

(3) Models of economic costs and benefits are almost never more accurate than the relationships embodied in the function **r**, and therefore using them cannot possibly result in a better economic gain from the product.

(4) The increase in economic gain to be expected from incorporation of models of economic benefits and costs into the design process, when the accuracy these models typically have is considered, is not worth the effort required.

(5) The model of economic benefit depends on the behavior of individual users outside the organization conducting the design, and the information given these users regarding the product may be incomplete or misleading.

The first two rationales are not worthy of discussion, because they are self-serving. The third is an argument with a premise that is true in most design efforts, but the argument is invalid. Whether any of the last two rationales are legitimate depends on circumstances of the particular design effort.

The main purpose here is not to claim that a complete list of criticisms of the use of explicit economic models in engineering design has been presented and each criticism refuted. Indeed the fourth rationale is very relevant. Instead it is to support the point that there is no fundamental reason applying to all engineering design efforts, making it undesirable to incorporate explicit models of economic gain in the design process. Use of such models is to be desired unless there are legitimate reasons to the contrary.

In Part I of this document will be presented a model for one measure of economic gain as a function of design parameter values (the function $g(\mathbf{d})$, with **d** restricted to quantitative variables) for integrated circuits of arbitrary functional purpose. The reader can judge the desirability of the use of this model as the details of it are presented.

Note here that Part II will be concerned with algorithmic techniques specifically appropriate for efficient maximization of the ecomomic gain modeled in Part I.

3. Perspectives on Design Methodologies

3.1. Two Characteristics of Advanced Practical Design Techniques

One corollary of the arguments made above is that if a design methodology is truly the ultimate approach to a practical engineering problem, it should use explicit economic models. In other words the methodology should have developed to the point that rationale 4 above, the strongest of the rationales opposing use of explicit models, should not apply.

There is another characteristic which almost always describes advanced methodologies for practical design: they take into consideration random processes present either in the manufacturing of the product or system, or in its application, or both. This can be argued much more easily than was the case for the use of economic models. It comes about because it is frequently not difficult to fully exploit all reasonably manageable deterministic relationships. When this is done, design choices are made which tend to promote random effects into the position of being the major class of effects limiting further advances in performance.

The evolution of integrated circuits is an ideal illustration of this point. First order deterministic considerations long ago showed that economic rewards generally increase as device dimension parameters decrease. So design practice immmediately moved into a world of design with "as small as possible" devices. But what has limited the rewards of shrinking device dimensions is not any set of deterministic relationships. It is the variations in realized device dimensions from circuit to circuit which results in variations in performance, which in turn degrades the economic value of the totality of circuits produced. Furthermore, the economic implications of this are very large, providing considerable incentives for advances in statistical circuit design techniques.

The model which will be presented in this document combines both of the above characteristics. It is a statistical model for economic gain in terms of design parameters over which the producer has control.

3.2. Previous Problem Formulations

Notwithstanding the validity and importance of the arguments made to this point, when the outlines of a methodology based on explicit economic models first began to emerge in our research group, over two-and-one-half years ago, they grew out of a practically-oriented look at previous problem formulations for the statistical design of (all types of) electronic circuits. It is useful to contrast these formulations with the one to be presented herein.

Statistical circuit design attempts to address variations in performance observed from unit to unit of the end product. Four assumptions are universal and will be used here as well:

(1) The electrical performance of any single unit of the product can be summarized by a finite number of real-valued numbers each of which represents a measured circuit response.

(2) Having the means to compute circuit responses (even if only by simulation) from given device parameters, it is always assumed there is some sort of statistical characterization either of these device parameters, or of a set of more fundamental (typically fabrication) parameters in terms of which the device parameters can be computed. For purposes of this discussion, the parameter set in question can be assumed specified completely by the mean vector and covariance of its multivariate distribution.

(3) The design goal (fortunately) does not include the infinite-dimensional problem of realizing a prescribed distribution of the performance responses of the circuit. Instead the implications of statistical variations in performance are communicated to the users of the circuits through the establishment of performance threshold levels (e.g., power consumption of 100 mw.), or, in standard terminology, product specs. The meaning of the product specs is that the producer offers to provide the

users with a collection IC's having performance responses no worse the values prescribed in the product specs.

(4) It is very desirable that a methodology for statistical circuit design be optimization-based. In particular the methodology should be capable of producing parametric designs arbitrily close to at least a locally optimum value. Accordingly, the phrase *statistical circuit design* will henceforth be replaced by *statistical circuit optimization*.

In addition, research published to date has assumed the following:

(5) The numerical threshold values for the above variables are given (and thus are fixed).

(6) the set of thresholds is the basis for categorizing each circuit as either passing, in which case it is considered to have full value, however much that might be, to the producer or user, or as failing, in which case it is considered to have no value. The fraction of circuits which are passing is defined to be the yield.

Beyond these common assumptions, there has been considerable diversity in proposed problem formulations. Some of this diversity is rationally based on differences in the economic considerations associated with different types of electronics. However, more to the point, there is also considerable diversity in proposed problem formulations for the class of electronics of interest here - integrated circuits. (See BRA81, section III).

(1) In optimal tolerance assignment, performance specifications, the yield, and the mean vector of the disturbances are held fixed, and variances of the presumed uncorrelated disturbances are determined which minimize production cost as a function of those variances.

(2) In design centering, performance specifications and the covariances of the uncertain parameters are fixed, and the mean vector is determined which maximizes the yield.

(3) In multiple-criterion optimization with yield maximization, the covariances of the uncertain parameters are fixed, and a mean vector is determined which results in a yield value and a set of performance values which are subjectively satisfying to the designer.

Some implications of the use of these formulations will be discussed in the next section, but first some digressions on the meaning of yield are called for. First, the yield defined above is more specifically the parametric yield, as opposed to the defect yield. In real IC fabrication various types of point defects appear, distributed randomly over the wafer. Each is capable of preventing the device or interconnect line where it is located from serving its purpose, so that the circuit is usually grossly non-functional. The defect yield is the fraction of fabricated circuits for which this has not happened. Among those circuits which are approximately functional, the fraction of circuits having performance meeting the specifications established for the circuit is the parametric yield.

Second, note that if the design parameters which are to be determined in a particular design methodology are assumed not to affect the defect yield, then it can be considered constant. Assuming this and certain other assumptions which may or may not be realistic, production cost is a (decreasing) monotonic function of the parametric yield for all values of the yield. Therefore the use of parametric yield as a performance attribute in the above problem formulations would seem to endow them with them a sophistication in their treatment of economic implications approaching that of explicit modeling of economic gain. It

is argued in the next section that this is not the case.

3.3. Evolution of the Economic-Gain-Maximizing Methodology

The interest in reformulating the IC statistial optimization problem resulted more fundamentally from an interest in conducting research sufficiently practical in its goals and execution that it might have a substantial impact on industrial practice. Examination of the previous formulations immediately resulted in the following two major conclusions, whose meaning is intended to be clear only after some elaboration.

(1) The motivations for the formulations contain an inconsistency which results in their missing a significant opportunity to maximize economic gain.

Explaining this is best accomplished using the kind of conceptual modeling equations which have aleady been put to use. Consider the design of a particular IC product. Any application of the previous formulations would have associated with it some choice of design parameters. Let the vector **u** represent these. This would almost certainly include nominal device dimensions, and possibly means and dispersion parameters needed to characterize the consequences of statistical fluctuations occurring in fabrication.

Now suppose a set of electrical performance parameters has been chosen such that for each one:

(1) its value is measured for each circuit produced.

(2) associated with it is a performance specification against which the parameter value for each circuit is compared

(3) any circuit not meeting the associated performance specification is a failed circuit (and is counted as such in yield assessment)

(4) the associated specification value is a formal specification of the product (values of the parameter are guaranteed to be no worse than the specification in communications from producer to user).

(5) the associated specification value has an effect on the desirability of the product to the user which is not negligible. Such specifications will be called salient specifications.

Let the elements of the vector **e** consist of the collection of these specification values. Let Y represent parametric yield. It depends not only on the design parameter vector **u**, but also on the specification vector **e**. With some abuse of notation (double use of Y), it can be written as,

$$Y = Y(\mathbf{u}, \mathbf{e})$$

If the defect yield is assumed independent of **u**, the cost can be written as

$$C = C(\mathbf{u}, Y(\mathbf{u}, \mathbf{e}))$$

An IC-producing organization may be either a firm which sells its products on the open market, or a division within a firm, which turns over its finished IC's to the firm for use in some electronic system produced by the firm. In either case, we assume there is some economic benefit, or value, v, say, per circuit, which the IC-producing organization expects to receive as compensation for supplying the circuit. Then the total benefit B from the product depends on the yield, since this affects how many circuits can be supplied to users, the performance specifications, since this determines how much the users desire the circuits, and the economic value assigned to the circuit. This can be written as,

$$B = B(Y(\mathbf{u}, \mathbf{e}), \mathbf{e}, \nu)$$

For the purpose of this discussion, take the economic gain G to be simply the difference between benefit and cost, that is,

$$G = B - C$$

Combining all these equations yields,

$$G = B(Y(\mathbf{u}, \mathbf{e}), \mathbf{e}, \nu) - C(\mathbf{u}, Y(\mathbf{u}, \mathbf{e}))$$

With this equation, the conclusion above can now be explained. The essential role that statistical circuit optimization techniques have is to provide an effective means to deal with the detrimental effect that inevitable fabrication fluctuations have on the ecomomic implications of a product. Clearly some capability to deal with these effects already exists in the industry, in the form of intuitive and trial-and- error methods, since catastrophic parametric yields cannot be tolerated. The development of an effective optimization methodology offers hope in two ways: the treatment of statistical fluctuations will require less total engineering effort; the methodology should result in somewhat greater economic benefit from the product.

The first statement is probably true. But it ignores the effort involved in the development of the methodology. The primary incentive for an effective optimization is really the second statement. But it is important to remember that not having an effective optimization methodology to use in the design of a product hardly implies the product will be a net economic loss, or that, for example, its yield will be zero. Continuing the discussion in terms of yield, for conciseness, suppose that typically the application of an effective optimization methodology would result in a yield 33% higher than would be otherwise achieved - say Y = .8 instead of Y = .6. Then one might say the purpose of statistical circuit optimization is to fine-tune whatever parameters can be considered design parameters, to squeeze out the most possible economic benefit from the product. But note that the parameters e and v are subject to control by the producer organization, and thus are candidates for design parameters. If it is worthwhile to develop a methodology to minimize the second term in the last equation above by adjusting **u**, perhaps it is worthwhile to develop a methodology to maximize G by adjusting **u**, e, and v.

Previous methodologies have ignored this. Application of such methodologies would presumably expend considerable effort and computer time to precisely adjust parameters based on a definition of what a "good" set of specifications which has, relatively speaking, been pulled out of the proverbial air.

The second, possibly more significant conclusion resulting from examination of previous formulations is the following.

(2) computer-efficient solutions to those formulations would be essentially useless for the statistical optimization of most IC products.

Restricting attention to IC's produced within essentially free-market economies, the portion of circuits manufactured that are designed in a production environment for which assumption 4 of the previous section holds, is quite small. This assumption can be called the single-grade assumption. Instead, typically two or more sets of thresholds are established for each product and, if the product is to be sold by the manufacturer, different prices are charged for IC's having performance in the different categories (grades) which result.

Furthermore there has been no suggestion as to how solutions (optimized design parameter values) of previous problem formulations for each grade taken one at a time might be combined to give the solution for that formulation for the multiple-grade product. Therefore computer-efficient solutions to those formulations are essentially useless for the statistical design of multiple-grade products. Hence the conclusion.

Although the problem formulation resulting from the modeling work presented herein differs in many ways from previous formulations, the major improvements result from the elimination of the two deficiencies which have been described in this section.

4. Further Definition of the Optimization Criterion

4.1. The Profit Measure of Economic Gain

The discussion to this point has attempted to identify and motivate an approach to statistical circuit optimization in which parameters are set in order to maximize some measure of economic gain expected from the product. There are a number of alternative measures which could be used. The measure which has been chosen to represent the gain G is the difference between the economic benefits and the costs, expected from the production of the the product being designed.

$$G = B - C.$$

Study of alternative measures of economic gain in industrial production and their appropriateness for the type of product of interest here has not been a major purpose of this research. Suffice it to say the following. Of the other measures which might be used, the most appealing is the return on investment which (according to some formula) can be attributed to the product being designed. The choice of the gain given by the above equation has been made as a result of the following rationale. Adequate models of return on investment would most probably have to be dynamic models of considerably greater complexity than the static models which are plausible for modeling benefits minus costs. Therefore use of a return on investment criteria would be attempting too large a step in the evolution of methodologies for statistical circuit optimization.

As has been mentioned earlier, an IC-producing organization may be either a firm which sells its products on the open market, or a division within a firm, which turns over its finished IC's to the firm for use in some electronic system produced by the firm. The former type of organization will be called an *IC house*, and the latter an *IC-supplying division* The methodology which is the subject here is considered to be potentially applicable to the case of the IC-supplying division. In particular the cost models presented in this document for the IC house should be applicable with at most minor modifications for the *IC-supplying division*. However the modeling of economic benefits from production of IC's might be significantly different in the two cases. Some comments about modeling economic benefits for the IC-supplying division are presented herein, but a full treatment of the subject has been left as a possible future extension of the methdology.

The economic benefit for the IC house is its revenue, which will be symbolized by R, and its economic gain therefore becomes its profit, which will be symbolized by π, as is conventionally done. The optimization criterion on which the methodology is based is therefore

$$\pi = R - C \tag{4.1}$$

4.2. Refinements in the Definition of Profit

Criteria for use in optimization must be precisely defined. The above equation is a precise definition of profit in terms of revenue and cost. But of course neither the revenue or cost are as yet precisely defined. Precisely defining profit in a way most appropriate for statistical circuit optimization can be thought of as the subject of the remainder of this document. Even this document in its entirety will leave some very detailed issues in the definition of profit unspecified. However, some of the most important issues can be addressed at this point.

The form of detailed expressions for profit is almost always that of a simple summation. Many of the issues involved in defining profit consist of whether a particular potential additive contribution to profit should in fact be included as a contribution. From (4.1), clearly if such issue are adequately addressed for revenue and cost separately, they are for profit as well. There are relatively few ambiguities in identifying appropriate components of revenue, partly because it is a flow of money across the usually well-defined "boundary" between the firm and the outside world. But determining whether or not to include various possible components of cost in the total is often plagued with difficulties. For the subject methodology, many such difficulties are avoided because of the general principle that addition of a constant (independent of the design parameters) to an optimization criterion has no effect on the optimal solution. Hence any additive contribution to cost (or, for that matter revenue) which does not depend on any design parameter can be omitted from the optimization criterion.

There are two potential cost terms which can be discussed at this point which among IC houses are generally very significant in magnitude - fixed costs of production, and payments to investors to provide a "normal" return on their investment. Fortunately, whether to include these as costs is an issue which can be decided using the optimization principle just stated. There is no reason to believe that either of these cost would change with changing design parameter values. Hence they can be ignored as a components of total cost. The components that remain are explicit costs incurred due to the production of the product being designed. These are somewhat more straightforward to identify.

There is another issue in the defining of profit which is a rather technical issue pertaining to the role of time. The issue must be made clear here even though it frequently is mistreated in elementary economics discussions. In the latter, profit quantities are sometimes labelled as having the dimensions of dollars, (in this document, dollars will quite arrogantly be taken as the generic unit of economic value), when quite clearly the dimensions should be dollars per unit time. For the purposes of this discussion, it is best to be unconventional and substitute the phrase *rate of profit* when the quantity in question is in fact in dollars per unit time.

Rates of profit fluctuate with time. A firm could if it wished, to give a concrete example, compute profit on a daily basis. There is no reason to suggest that this computation (dimensions of dollars per day) would yield the same result day in and day out. So rate of profit should be thought of as a function of time. Therefore, it would be a rare situation in which any firm would have a rate of profit quantity as its optimization criterion, since this would require the singling out of a special time or set of times as important while all other times are not. The way out of this is obvious. What is wanted by firms is a single real number

summarizing the profit status of the firm. Clearly the sensible criterion is the integral of the rate of profit over an appropriate time interval. This will be refered to as accumulated profit. The widely worshiped quarterly profit is such a criterion, although of course summation substitutes for integration in its computation. And if some form of profit is to be the criterion of choice for statistical circuit optimization, it should be an accumulated profit.

There remains the question of what time period should serve as the basis for the computation of accumuted profit in the subject methodology. At this point, suffice it to say that the time period should begin when the newly designed or newly redesigned product is introduced in the market, and, in the ideal application of the methodology, end when it is expected at least some of the models of the methodology will be updated and the product at least partially redesigned. This time period is called the design lifetime. The criterion in the methodology, then, is the expected profit associated with the product being designed, accumulated over the design lifetime.

The profit model of the subject methodology is a static model. There is no variable, either continuous or discrete, representing time, and running from beginning to end of the design lifetime. (Of course there is in general a time axis implicit in time-domain circuit simulations required by the methodology.) Instead, constant accumulated costs, revenues, and profit are considered to characterize the design lifetime as a whole. So there is no integration required to compute accumulated profit. A certain duration is chosen for the design lifetime, based on expected accuracy stability of the models, but no symbol for this duration is needed in the model equations. Note however profit depends directly on the total number of circuits which the firm starts in production during the design lifetime. The model does use a variable for this quantity. In fact, it is a design parameter in the methodology.

To reiterate, the type of profit used in the subject methodology is the profit expected to result from the initiation in production of a prescribed quantity of circuits, accumulated over the design lifetime of the product.

5. Some Global Assumptions and Notational Conventions

Modeling assumptions and notational definitions will be introduced as needed throughout the document. However some modeling assumptions and notational conventions are sufficiently global in their use that they are best presented here.

As has already been evident, vectors are in boldface.

Following conventional practice in statistics, random variables are denoted by upper-case letters, and their realizations by the corresponding lower-case letters.

If a random variable, say Y, is normally distributed with mean μ and variance σ^2, its density function will be denoted $n_Y(y;\mu,\sigma^2)$.

Since the topology of the circuit being optimized is given, the number of devices (sum of resistor, capacitor, and transistor counts) in the circuit is a known constant. And associated with the product being optimized is assumed a unique set of masks, which implies every wafer has the same number of die locations. Hence each wafer has the same number of potential chips. Let

n^d = number of devices in the circuit.

N^c = number of potential chips on each wafer.

Corresponding to these definitions, indexing variables are defined. Let

d = index of devices in the circuit.

c = index of potential chips on each wafer.

Also, let

n^w = number of wafers started in production during the design lifetime of the product.

Note that C and c always denote cost (with the units of dollars), while c used as a subscript indexes potential chips.

In the modeling of the costs of IC fabrication it has been found very advantageous to decompose the total variable cost C, of (4.1) into two major parts. Specifically, in IC fabrication there is one collection of manufacturing (and testing) operations performed on the circuits up to and including the die separation operation, and another distinct set of operations performed after die separation. In essential agreement with industry jargon, the pre-die-separation operations are called front-end (FE) operations, and the post-die-separation operations are called back-end (BE) operations. It is relatively straightforward to distinguish between costs incurred before and after die separation. Let

C^{FE} = total variable costs incurred in front-end operations, to produce n^w wafers.

C^{BE} = total variable costs incurred in back-end operations, to produce n^w wafers.

Then

$$C = C^{FE} + C^{BE}. \qquad (5.1)$$

6. Statistical Modeling of Process Disturbances

The disturbances in the fabrication of IC's can be statistically modeled using various types of parameters. The two major types which can be used are device parameters and fundamental process parameters. The use of device parameters has the serious drawback that correlations among them must be adequately modeled, a difficult if not impossible task. Until recent years, the use of fundamental process parameters has had the serious drawback that the solution of partial differential equations was necessary to compute the device parameters corresponding to the process parameters needed for the evaluation of circuit performance. For statistical circuit optimization, the cost of this computation, which in general needs to be repeated many times, is prohibitive. However, in 1981 the FABRICS IC process simulator was introduced, in [MAL81]. FABRICS represents an attempt to eliminate the need for costly partial differential equation solutions through the use of program parameters which are adjusted to the particular industrial fabrication process in which the product of interest is to be fabricated. It provides a means of calculating device parameters from fundamental process parameter with a cpu cost less than or comparable to the cost of computing circuit responses. As a consequence of these basic considerations, it was decided that the new methodology presented here should use fundamental process parameters to statistically model process disturbances, and should incorporate FABRICS code in its software implementation.

As a result of this decision, some properties of the new methodology are dictated by properties of FABRICS. The two most important of these are the identities of the fundamental process disturbances that are modeled as random variables with specified distributions, and the special scheme by which these distributions are parametrized.

Regarding the fundamental process disturbances (fpd's for short), note that these are modeled statistically as a collection of independent normally distributed

random variables, and the term "fundamental" is meant to imply that every other random variable occuring within FABRICS is some function of at least one of the fundamental random variables. For the detailed identities of the fpd's, the reader is refered to the FABRICS documentation [NAS83, NAS84]. However this information is summarized here as follows. The approximately 40 fpd's include the following: line widths; diffusivities and segregation coefficients of impurities; implantation profile spread quantities; oxidation growth coefficients; oxide charge and fast state densities; substrate impurity concentration; poly resistivity and thickness; contact resistivity model parameters.

What is meant by the "special scheme for parametrization" of the fpd's needs clarification. FABRICS generates different values for each of the fpd's for each device in a circuit or collection of circuits. Furthermore, one goal of FABRICS is to accurately parametrize device-to-device variations in the fpd's. But variations between fpd values among devices belonging to the same chip are less than those among devices that belong only to the same wafer. And variations between fpd values among devices belonging to the same wafer are less than those among devices belonging to the same production run, and so on up the hierarchy of device groupings: circuit, wafer, run, production period. In order to account for this, and in recognition of the significance that these groupings have in judging the merits of a design, FABRICS uses a hierarchical system for generating values of fpd's. Let

$\mathbf{D}_{cd}$ = a vector for device d of chip c, with components the fpd random variables for that device.

In the generation of the set of vectors $\{\mathbf{D}_{cd}\}_{d=1\,c=1}^{n^d\,N^c}$, each component of the vectors is generated in the same way, so the discussion here will describe that system for one arbitrarily chosen component (but for all d and c). Call it D_{cd}. In FABRICS, realizations of the set of random variables $\{D_{cd}\}_{d=1}^{n^d}$, for all the devices on chip c are generated, using a random generator, to simulate a density for D_{cd} of $n(\cdot\,; m_c, v_c)$. However m_c and v_c, the mean and variance of the device disturbances for chip c, are not fixed. Instead they are random variables which must first be generated before the realizations of D_{cd} for all the devices on chip c can be generated. The mean random variable M_c is modeled as having density $n(\cdot\,; m_m, v_m)$, and the variance random variable V_c is modeled as having density $n(\cdot\,; m_v, v_v)$. In this two-level hierarchy, m_m, v_m, m_v, and v_v are fixed. Clearly values of D_{cd} can be generated in this manner, but it brings up two important theoretical issues. The first is whether the values so generated are realizations of some legitimate random variable. Fortunately the answer is yes, although the justification is very distant from the purposes of this document. The second issue is essentially how to compute expectation values. Suppose it is desired to compute the expectation of some function of the complete set of realizations $\{d_{cd}\}_{d=1\,c=1}^{n^d\,N^c}$, $g(\{d_{cd}\}_{d=1\,c=1}^{n^d\,N^c})$. This expectation is correctly given by,

$$Eg = \int\int \cdots \int \{\int\int \cdots \int g(\{d_{cd}\}_{d=1\,c=1}^{n^d\,N^c}) \prod_{c=1}^{N^c} \prod_{d=1}^{n^d} [n_{D_{cd}}(d_{cd}: m_c, v_c)\, \mathrm{d}d_{cd}]\}$$

$$\prod_{c=1}^{N^c} [n_{M_c}(m_c: m_m, v_m)\, n_{V_c}(v_c: m_v, v_v)\, \mathrm{d}m_c\, \mathrm{d}v_c]$$

where d (as opposed to d) denotes the differential element.

Note that FABRICS has the potential capability to generate random variables with four levels of hierarchy, not just two. But only two levels are needed in the methodology as presented herein.

In the subject methodology essentially all functions of interest are of course functions of more than one component of the vectors in the set $\{\mathbf{D}_{cd}\}_{d=1\,c=1}^{n^d\ N^c}$. Expectations of such functions can be expressed in a manner differing little from the above equation. The random variables D_{cd}, m_c, and v_c and their realizations, and the parameters m_m, v_m, m_v, and v_v, must be replaced by their vector analogs, printed here in boldface. Let there be n^f components of $\mathbf{D}_{cd}$, indexed by i. By way of example, and recalling the independence assumed among the n^f components, let

$$n_{\mathbf{D}_{cd}}(\mathbf{d}_{cd}; m_c, v_c)$$

stand for

$$\prod_{i=1}^{n^f} [n_{D_{cd,i}}(d_{cd,i}; m_{c,i}, v_{c,i})]$$

And, also by way of example, let the symbol $\mathrm{d}\mathbf{d}_{cd}$ stand for,

$$\prod_{i=1}^{n^f} \mathrm{d}d_{cd,i}$$

Finally, let H be the function of $\{\mathbf{D}_{cd}\}_{d=1\,c=1}^{n^d\ N^c}$ the expectation of which is desired. Then

$$EH = \iint \cdots \int \{\iint \cdots \int H(\{\mathbf{d}_{cd}\}_{d=1\,c=1}^{n^d\ N^c}) \prod_{c=1}^{N^c} \prod_{d=1}^{n^d} [n_{\mathbf{D}_{cd}}(\mathbf{d}_{cd}; \mathbf{m}_c, \mathbf{v}_c)\, \mathrm{d}d_{cd}]\}$$
$$\prod_{c=1}^{N^c} [n_{\mathbf{M}_c}(\mathbf{m}_c; \mathbf{m}_m, \mathbf{v}_m)\, n_{\mathbf{V}_c}(\mathbf{v}_c; \mathbf{m}_v, v_v)\, dm_c\, \mathrm{d}v_c]$$

There are two random variables of interest here, including most importantly the profit itself, which are functions not only of the processing disturbances modeled in FABRICS but of another scalar random variable specific to the methodology, and of design parameters. Although the meanings of these additional variables have yet to be described, in the interest of completeness, their proper role in the expectation integral is detailed here. The symbol for the unspecified design parameters is $\mathbf{z}$. The symbol for the additional random variable is Λ. Let the density function of Λ be f_Λ. Suppose the expectation of an integrand $I(\{\mathbf{d}_{cd}\}_{d=1\,c=1}^{n^d\ N^c}, \Lambda, \mathbf{z})$ is desired. It can be expressed as,

$$(EI)(\mathbf{z}) = \iint \cdots \int \{\iint \cdots \int I(\{\mathbf{d}_{cd}\}_{d=1\,c=1}^{n^d\ N^c}, \Lambda, \mathbf{z}) \prod_{c=1}^{N^c} \prod_{d=1}^{n^d} [n_{\mathbf{D}_{cd}}(\mathbf{d}_{cd}; \mathbf{m}_c, \mathbf{v}_c)\, \mathrm{d}d_{cd}]\} \tag{6.1}$$
$$\prod_{c=1}^{N^c} [n_{\mathbf{M}_c}(\mathbf{m}_c; \mathbf{m}_m, \mathbf{v}_m)\, n_{\mathbf{V}_c}(\mathbf{v}_c; \mathbf{m}_v, v_v)\, dm_c\, \mathrm{d}v_c] f_\Lambda(\lambda)\, d\lambda$$

The purpose of the remainder of this document is to define two integrand random variables which when inserted in the above integral form yield two expectations needed in the new methodology. As suggested above, one of the random variables is the profit random variable. The other must await section 11 for description and

motivation.

7. Front-end Cost Modeling

For the version of the methodology presented in this document, the front-end cost per wafer is basically a simple summation of constants representing costs of the various front-end operations. However, as mentioned earlier, device dimensions are among the designable parameters in the methodology, and there are two cost terms which in actuality vary significantly with these dimensions. One term is the cost of wafer probe testing and the other is the cost of die separation. These terms are typically relatively small, numerically. However modeling their dependence on device dimensions is worthwhile since most of the model elements needed to do so are required to model other more significant effects on profit. These model elements are presented first.

Let

$\mathbf{x}$ = the vector of device dimensions which are to be treated as design parameters for the circuit in question.

Each device in the circuit has an active area such that if a spot defect occurs in that area, the device is not likely to function normally. The exact definition of the active area for each device type is left to the user of the methodology. In any case, the circuit has a total active area which is the sum of the active areas of the devices having dimensions which are designable, and those treated as having fixed dimensions. Let

a = total active area of the circuit

Then a is some known function of $\mathbf{x}$, which with some abuse of notation, is denoted by $a(\mathbf{x})$.

Once the layout of a chip has been completed, it has some well-defined area (including border area). Let

A^c = total chip area.

Of course A^c depends on the details of the layout, and cannot adequately be modeled as a deterministic function of a. But there is no reason it cannot be modeled statistically. The layout process is thought of as a statistical experiment yielding a value for A^c. This value is not known at the time the methodology is applied. But it can be statistically modeled. The model proposed here is described as follows.

In section 6 a random variable Λ was introduced to serve as a sort of mathematical place-holder pending precise description of its meaning in this section. The chip area is modeled using Λ according to,

$$A^c = \Lambda a \tag{7.1}$$

which reflects that as a tends to zero, so does A^c. The distribution proposed for Λ is normal, on the basis that the total area can be thought of as a sum of areas attributable to a number of subcircuits each of which make a small contribution to the total area. Estimates of the mean and variance of the distribution of Λ can be obtained by a study of the ratio of chip to active areas for layouts of circuits deemed to have similar layout considerations to those of the circuit of interest.

Different values of A^c would result in different values of N^c, the number of circuits (or, die locations) on a wafer. N^c is of course a positive integer and a discontinuous function of A^c, however the magnitude of the steps in the N^c versus A^c function are relatively small except for the larger VLSI circuits, to which the

methodology is not likely to be applied soon. Total errors that would result from the use of a continuous deterministic function to model the dependence of N^c on A^c would generally be less damaging than errors from other sources. Let $\hat{N}^c(A^c)$ denote the function which models the dependence of N^c on A^c. For convenience in the many equations to follow, N^c is given a second meaning - the composite function defined by

$$N^c(\Lambda, \mathbf{x}) = \hat{N}^c(\Lambda a(\mathbf{x}))$$

so that,

$$N^c = N^c(\Lambda, \mathbf{x}) \tag{7.2}$$

It should not be surprising that N^c has a major effect on the profit. This will be made explicit later, but for now attention returns to the modeling of the cost per wafer of wafer probe testing and die separation. By curve-fitting techniques which need not be extremely accurate, any IC house should be able to generate a reasonable continuous-function model for the dependence of each of these costs on N^c. Let $\hat{C}(N^c)$ denote the resultant model for the dependence of the sum of the wafer probe and die separation costs on N^c. And let $C^{\#}$ be the composite function,

$$C^{\#}(\Lambda, \mathbf{x}) = \hat{C}(N^c(\Lambda, \mathbf{x})) \tag{7.3}$$

Then if c^k denotes the portion of the front-end cost not attributed to the wafer probe or die separation operations, the front-end cost per wafer can be written as,

$$C^{FE/w} = c^k + C^{\#}(\Lambda, \mathbf{x}).$$

Since the profit quantity which forms the criterion for the methodology is the profit resulting from the fabrication of all n^w wafers produced during the design lifetime, the front-end cost quantity of interest is the total cost incurred to produce n^w wafers. Call this C^{FE}. This is then

$$C^{FE} = n^w[c^k + C^{\#}(\Lambda, \mathbf{x})]. \tag{7.4}$$

8. A Back-end Flow Notation System

Once the separation of the wafer dice has been carried out, those chips which performed well enough in wafer probe testing to avoid rejection are packaged in one of a number of package types. Then they are in general subjected to additional testing, and possibly other treatments, usually designed to increase customer confidence in the reliability of the product. (The prototypical example of the latter is the "burn-in" procedure, in which the circuits are operated under specified electrical and environmental conditions for some large specified number of hours, then performance tested.) The flow of material in most industrial manufacturing processes is a merging flow - one-piece parts are assemled into multi-piece parts, which in turn are assemled into sub-assemblies, and so on up a manufacuring tree until one final product is produced. The flow of material in the back-end processing of IC's is just the opposite. It is a splitting flow. As chips flow through the processing, they are placed in successively more narrowly defined categories, based on two basic types of conditions:

(1) which of the various electrical performance categories they belong in, as determined by their performance in tests.

(2) which of a number of discrete back-end treatments they have been subjected to, where possible treatments include various packaging options and other possibly reliability- related treatments.

Since in general finished IC's which are in different categories on whatever basis cost differently to produce, and sell at different prices, it is necessary to have a notational system to keep quantitative track of the splitting of the circuits into categories. The system which has been developed for the subject methodology will be presented for a particular example only. However, it is not difficult to extrapolate from the example to see what the system would be in its most general form.

The example is that of a hypothetical linear amplifier with specifications summarized in the table of Figure 1. *Variant* is a key word in the methodology which can be roughly translated as "version" (of the product). More precisely, two samples of a product represent different variants of the product if they are in different electrical performance categories, or have been subjected to different back-end treatment (including packaging). The entries in the variant column are simply numerical labels. v_{os} and t_s refer to appropriately defined offset voltage and settling times specifications of the product. A column for specifying the temperature at which the other electrical performance specifications are to hold is included even though in this example only one temperature value is used. It is assumed the product will be made available in two package types, named a and b. By *other treatment* is meant any procedure which might be performed on the circuits comprising the product variant, by the producer, because of an understanding between the producer and its customers that so doing increases the desirability of the product to the customer. The prototypical example of this is the burn-in procedure already mentioned. For this product the burn-in procedure is performed on variant 5 only.

In Figure 2 is a representation of the back-end flow of the product. It not only conveys the structure of the flow, but serves to define the notational system as well. The flow has a general tree structure. The chips undergo a sequence of splits which are of two types. The splits based on the conditions described in (1) above are called *test-outcome* splits. The splits based on the conditions described in (2) above are called *discretionary* splits.

The total collection of chips resulting from the die separation of a wafer begin at the root of the tree. On the basis of wafer probe tests they are split into three categories (a test-outcome split):

(2) those that are functional and have $v_{os} \leqslant 10$

(1) those that are functional and have $10 < v_{os} \leqslant 25$

(0) those for which neither of the above are true.

By functional is meant that the circuit shows to some approximation the behavior expected of the class of circuits of which it is a member, e.g. 4pst analog switch, voltage follower, 8-bit multiplexer. This is not necessarily the case for circuits which have some fatal defect, spot or otherwise. Y_2^{WP} and Y_1^{WP} are the fraction of chips from the wafer in categories (2) and (1) respectively. They are called *test-outcome split fractions.*

At the next level in the tree, to the right, is a discretionary split which would almost always be a packaging split. Consider the circuits in category (2) above. In what amounts to a random decision for each chip, the chips are split into two categories (a discretionary split):

(a) those to be packaged in package a.

(b) those to be packaged in package b.

s_{2a} and s_{2b} are the fractions of the circuits which are in package categories (a) and (b), respectively. They are called *discretionary split fractions*. c_{2a} and c_{2b} are the costs of packaging these circuits in packages a and b, respectively. (In most

cases, the cost of packaging a chip in a particular package would be independent of the wafer-probe test outcome of the chip. However, little additional notational complexity is required to avoid making this assumption.)

At the next level in the tree is a test-outcome split which would almost always be some portion of the final testing. Final testing here means simply post-packaging testing. Consider the circuits in categories (1) and (a) above. On the basis of the test in question in this case, the circuits are split into three categories (a test-outcome split):

(2) those that have $t_s \leqslant 5$.

(1) those that have $5 < t_s \leqslant 7$.

(0) those for which neither of the above are true.

$Y^{FT\,1}_{2a\,2}$ and $Y^{FT\,1}_{2a\,2}$ are the fractions of the circuits which are in categories (2) and (1) immediately above, respectively.

At the final level in the tree of this example is a discretionary split in which those circuits for which the fraction $Y^{FT\,1}_{2b\,1}$ applies, would be (randomly) split into those to be and not to be burned in. The only additional comment which should be necessary to make clear the notation is that $s_{2b\,1a}$ and $s_{2b\,1a}$ are the fractions of the circuits that are and are not burned in, respectively.

Note some properties of this back-end flow and notational system.

(1) The splits alternate between test-outcome and discretionary splits.

(2) The discretionary splits have costs associated with them, but the test-outcome splits do not.

(3) There are no branches in the tree for circuits which fail to meet minimal performance in a particulat test stage. Therefore, it is not the case that at each tree node representing a test-outcome split, the Y fractions sum to unity.

(4) There are no branches in the tree for circuits which fail to be successfully treated/packaged with treatment/package a or b, because their treatment/packaging is botched. The fraction for which this occurs is small, and the phenomenon can either be ignored, or, if the firm wishes, modeled as a constraint that the s fractions at a particular discretionary split node sum to some constant less than but close to one. If the phenomenon is ignored, then of course the s fractions at each discretionary split node would sum to unity.

(5) The rightmost nodes in the tree are refered to as *back-end outcomes*. Clearly the fraction of circuits having a particular back-end outcome is the product of the test-outcome and discretionary split fractions associated with the leafs of the tree between the root and the leaf associated with the back-end outcome in question (including the latter). For example, the fraction of circuits that are burned in is $Y_2 s_{2b} Y_{2b\,1} s_{2b\,1b}$.

9. Modeling of Relationships Pertaining to Back-end Processing

Having presented the notational system which has been found appropriate for describing the flow of chips in the back-end processing, it is now possible to discuss the role of some additional design parameters in the methodology, (adding to n^w and $\mathbf{x}$), and the functional relationships needed to model the dependence of profit on the design parameters which have been introduced. Specifically, the purpose of this section is to discuss the dependence of the split fractions on the disturbances and the design parameters.

The test-outcome split fractions depend on a type of design parameter which has not yet been formally introduced. Consider the specification of the variants of a product as exemplified by the table introduced in the previous section to describe the hypothetical linear circuit. The information contained in such a table typically results from two types of decisions:

(1) discrete decisions about what might be called the "variant structure" of the product. These decisions pertain to such issues as how many package types to make available, how many variants are available in each package type, whether or not certain variants should have strictly better electrical specs than certain others, and other inequality constraints between spec values. An example of the latter is the constraint in the example presented, that the settling time specification for variants 3-5 should not be less than that for variant 2.

(2) selection of the numerical values in the chart.

The producing firm is free to implement whatever decisions of the type described in (1) it wishes, and the decisions would generally have a first-order effect on the economic aspects of the product. However, the decisions are discrete, and therefore not directly addressed by the subject methodology. The decisions of the type described in (2), on the other hand satisfy all three of the conditions that make them fair game in the methodology: the firm is free to implement them, they are significant to the economics of the product, and they are parametric. Therefore they are treated as design parameters in the methodology. The vector **e** is a vector whose components are the compressed set of electrical performance specification values used to specify the product. (Compressed here refers to the elimination of redundancy of variables that might tend to occur in setting up the **e** vector.) For example, the numerical value of the **e** vector for the linear circuit example would be given by,

$$\mathbf{e}' = [25\ 10\ 3\ 5\ 7]'$$

where $'$ denotes transpose.

The test-outcome split fractions depend on **x**, **e**, and the disturbance random variables. Before this can be discussed further, however, it is necessary to digress, and discuss the handling of defect yield effects in the methodology.

The analysis of the effects of spot defects on IC's is most definitely a statistical problem, and its ultimate simulation would treat the defect yield as a random variable, probably dependent on one or more other random variables useful in the modeling of defect phenomena. However, for simplicity, in this version of the methodology, only the expected value of the defect yield, denoted y^D, is modeled. The model is that y^D depends on a, the active area. That is,

$$y^D = y^D(a)$$

Since this research does not have improvement of techniques for defect yield modeling as one of its goals, the details of this fairly general model are left for implementation by the firm applying the methodology, which may well have developed its own functional form for the above dependence. Note that a depends on **x**, so that,

$$y^D = y^D(a(\mathbf{x}))$$

Returning to the discussion of the dependence of the test-outcome split fractions on design parameters, note that the design device dimensions together with the complete set of processing disturbances completely determine the performance of all

the circuits, assuming the circuits do not have catastrophic defects. Under this assumption, the dependence of the wafer-probe split fractions Y_2^{WP} and Y_1^{WP} on disturbances and design parameters can be written in terms of functions called $\hat{Y}_1^{WP}$ and $\hat{Y}_2^{WP}$ as,

$$Y_1^{WP} = \hat{Y}_1^{WP}(\{\mathbf{D}_{cd}\}_{d=1}^{n^d}{}_{c=1}^{N^c}, \mathbf{x}, \mathbf{e})$$

$$Y_2^{WP} = \hat{Y}_2^{WP}(\{\mathbf{D}_{cd}\}_{d=1}^{n^d}{}_{c=1}^{N^c}, \mathbf{x}, \mathbf{e})$$

In the methodology, the phenomena which are responsible for variations in parametric performance are considered independent from those which cause catastrophic failure. This leads to a simple multiplicative method of combining defect yield and wafer-probe split fractions. In the hypothetical circuit example, these fractions are written as,

$$Y_1^{WP} = y^D(a(\mathbf{x}))\, \hat{Y}_1^{WP}(\{\mathbf{D}_{cd}\}_{d=1}^{n^d}{}_{c=1}^{N^c}, \mathbf{x}, \mathbf{e})$$

$$Y_2^{WP} = y^D(a(\mathbf{x}))\, \hat{Y}_2^{WP}(\{\mathbf{D}_{cd}\}_{d=1}^{n^d}{}_{c=1}^{N^c}, \mathbf{x}, \mathbf{e})$$

Of course the test-outcome split fractions which are associated with final test outcomes contain no multiplicative defect yield factor since the effect is incorporated in the wafer-probe splits. In the hypothetical example, the dependence of Y_{2b1}^{FT1}, for example, on design parameters can be represented as,

$$Y_{2b1}^{FT1} = \hat{Y}_{2b1}^{FT1}(\{\mathbf{D}_{cd}\}_{d=1}^{n^d}{}_{c=1}^{N^c}, \mathbf{x}, \mathbf{e})$$

and the function on the right is computed via circuit simulations but does not depend in any way on defect yield.

Turning now to the discretionary split fractions, note that they themselves satisfy the three criteria mentioned above for electing to treat a parameter as a design parameter in the methodology. They are so treated. The vector $\mathbf{s}$ is comprised of discretionary split fractions. It does not, however include every such fraction appearing in the back-end flow diagram, because the fractions are subject to equality constraints on their sums as discussed earlier. Instead it includes an appropriately selected "basis set" of discretionary split fractions. By "basis set" is meant, somewhat as in vector space theory, a set which is independent yet sufficiently large to enable the calculation of all the fractions.

10. Back-end Cost Modeling

Most of the definitions and relationships needed in the modeling of back-end cost have already been presented. It is only necessary to add that aside from issues of distinguishing between fixed and variable costs of production, the costs of back-end operations are well modeled as constants. The expression for total back-end cost associated with the fabrication of n^w wafers is a summation which can be written in nested form. For the hypothetical circuit, this is,

$$C^{BE} = n^w N^c \{ y_1^{WP} c_{1a} + y_2^{WP} [s_{2a} c_{2a} + s_{2b} [c_{2b} + y_{2b1}^{FT1} s_{2b1b} c_{2b1b}]]\}$$

In the general case, the summation is notationally complex and is not presented. However, letting $\mathbf{Y}^{WP}$ be the collection of wafer-probe split fractions

assuming unity defect yield, and $\mathbf{Y}^{FT}$, the collection of final test split fractions, some detail of the dependence of back-end cost on quantities of interest is conveyed in,

$$C^{BE} = C^{BE}(n^{w}, y^{D}(a(\mathbf{x}))\, \mathbf{Y}^{WP}(\{\mathbf{D}_{cd}\}_{d=1}^{n^d}{}_{c=1}^{N^c}, \mathbf{x}, \mathbf{e}), \mathbf{Y}^{FT}(\{\mathbf{D}_{cd}\}_{d=1}^{n^d}{}_{c=1}^{N^c}, \mathbf{x}, \mathbf{e}), \mathbf{s}) \quad (10.1)$$

11. Revenue Modeling

As described in section 8, IC houses advertise and sell IC products in several variants. Total revenue from their sale can readily be expressed as a sum of revenues deriving from the variants. Let there be $\bar{v}$ variants of a the product being designed, indexed by v. Let

q_v = number of units of variant v sold during the design lifetime of the product.

p_v = average selling price for variant v.

And define $\mathbf{p}$ and $\mathbf{q}$ by,

$$\mathbf{p}' = [p_1\, p_2 \cdots p_{\bar{v}}]'$$

$$\mathbf{q}' = [q_1\, q_2 \cdots q_{\bar{v}}]'$$

Then it is not difficult to see that the revenue from the sale of all the variants of the product, during the design lifetime, can be written as,

$$R = \sum_{v=1}^{\bar{v}} p_v q_v (\mathbf{p}, \mathbf{q}) \quad (11.2)$$

Now note that the price and quantity vectors satisfy two of the three criteria used to select design parameters for the product: they are parametric, and they have a potential first-order effect on profit. Furthermore, the elements of the price vector satisfy the third criterion for design parameters - that the producing firm be free to set their values as it wishes. Hence the $\mathbf{p}$ vector is treated as designable in the methodology. The freedom of the producer to set the values of the elements of the quantity vector is a restricted freedom which can be described as follows. The producing firm is free to set them as small as it wishes, simply by limiting sales transactions. This freedom requires that the $\mathbf{q}$ vector also be treated as designable in the optimization formulation. However the producer is by no means free to set the quantities as large as he wishes. And since with nonzero prices, the revenue is a monotonically increasing function of the purchase quantities, it is important to include in the formulation, i.e., to model, all upper bounds on these quantities. There are two such types of upper bounds. One type has been given the name *demand constraints*, and the other, *supply constraints*.

Demand constraints arise from the principle that the purchase quantity of a particular variant cannot exceed the number of circuits of that variant which customers decide to buy during the design lifetime. Let

q_v^d = maximum quantity of circuits of variant v which potential customers of the design product are willing to purchase during its design lifetime, for $v = 1, 2, \ldots, \bar{v}$.

In keeping with conventional economic parlance, q_v^d is called the *demand for variant* v. With this notation, the demand constraints are simply,

$$q_v \leqslant q_v^d, v = 1, 2, \ldots, \bar{v}. \tag{9.5}$$

Defining the *variant demand vector* $\mathbf{q}^d$ by

$$\mathbf{q}^d = [q_1^d \; q_2^d \; \cdots \; q_{\bar{v}}^d]'.$$

allows these inequalities to be written as,

$$\mathbf{q} \leqslant \mathbf{q}^d. \tag{9.6}$$

It remains to discuss the determination of the variant demands. They depend on the autonomous behavior of potential customers. Among the determinants of q_v^d are certainly p_v and those elements of $\mathbf{e}$ which pertain to variant v. However, since the variants of a product compete with each other for the interest of the customer who intends to buy some of the product, q_v in general must be considered to depend on all the elements of $\mathbf{p}$ and $\mathbf{e}$. Furthermore, it is clear that the variant demands do not depend on any of the other design parameters, n^w, $\mathbf{x}$, $\mathbf{s}$, and $\mathbf{q}$, which have been introduced. Hence,

$$q_v^d = q_v^d(\mathbf{p}, \mathbf{e}), v = 1, 2, \ldots, \bar{v}.$$

or, in vector notation,

$$\mathbf{q}^d = \mathbf{q}^d(\mathbf{p}, \mathbf{e}), \tag{9.7}$$

and it remains only to determine these demand functions. As a group they constitute an element of the profit model that is refered to as the *demand model*. This is standard economics terminology, although many demand models in economics are dynamic models, rather than static, and most which include the effects of product attributes on demand involve less well-defined attributes. Note that a particular variant v competes for the interest of customers not only with other variants of the design product, but also with all other variants of any other products on the market which fit their applications. The latter may include products of both the IC house applying the new methodology to the design product, and other IC houses. Hence the demand model requires price and characteristic information pertaining to all of the other variants of all of these products.

Development of a methodology to determine this static demand function for IC products taken one at a time has been a major goal of this research. And at the time of this writing, most of the details of this methodology have been fixed. However, the detailed results of this component of the research are to be presented elsewhere. But two characterizations of the resultant class of demand models are needed here. The first is that while the demand is subject to significant random effects, the modeling methodology yields demand functions which are the expected value of the demand (with respect to the random variables present in the detailed stochastic model), and this, symbolized by q_v, is what is used in the version of the methodology presented here. Second, the demand modeling methodology yields a class of demand functions which are differentiable functions of their arguments.

The other type of upper bounds needed, the supply constraints, arise from the principle that the purchase quantity of a particular variant cannot exceed the expected number of circuits resulting from the starting in production of n^w wafers and having characteristics meeting or exceeding those which define the variant. It is not difficult to see that the supply constraints form a bridge between the revenue and cost models. In order to model them, the concept of number of IC's must first be made more precise. Let there be n^o back-end outcomes for the product in question. And let,

F_o = fraction of circuits from the wafer with back-end outcome o, $o = 1,2, \cdots n^o$.

The back-end flow representation of Figure 2 shows an appropriate assignment of F components to back-end outcomes for the hypothetical circuit (at the right side of the tree). Note that the *number* of circuits in back-end outcome category o, resulting from n^w wafers started in production is just $n^w E\{F_o N^c\}$. Back-end outcome fractions can be expressed as products of back-end split fractions. Again for the hypothetical circuit, these expressions are,

$$F_1 = Y_1^{WP} s_{1a} Y_{1a\,1}^{FT\,1} s_{1a\,1a}$$

$$F_2 = Y_1^{WP} s_{1a} Y_{1a\,2}^{FT\,1} s_{1a\,2a}$$

$$F_3 = Y_2^{WP} s_{2a} Y_{2a\,1}^{FT\,1} s_{2a\,1a}$$

$$F_4 = Y_2^{WP} s_{2a} Y_{2a\,2}^{FT\,1} s_{2a\,2a}$$

$$F_5 = Y_2^{WP} s_{2b} Y_{2b\,1}^{FT\,1} s_{2b\,1a}$$

$$F_6 = Y_2^{WP} s_{2b} Y_{2b\,1}^{FT\,1} s_{2b\,1b}$$

With this type of relationship, the quantities of circuits having the various back-end outcomes has been related to familiar quantities.

Next, the correspondence between variants and back-end outcomes must be understood. It has probably already been noticed that it is not a one-to-one correspondence. In particular, there are back-end outcomes the circuits of which can be sold as more than one variant. In the hypothetical example, the circuits having back-end outcome 4, for example can be sold either as variant 1 or variant 3. An effective tool for understanding the correspondence in question, as well as the set relationships between the grades is the Venn diagram showing the sets of circuits which can be sold as the various variants of the product. For the hypothetical product, this diagram is shown in Figure 3. The reader can verify that the variants bear the correct set relationship to each other. For example, the diagram displays the fact that no circuit which can be sold as variant 4 or 5 can be sold as any one of the first three variants, because of the difference in packages between these two groups of variants. The back-end flow diagram has been so constructed that it has a back-end outcome for each "region" of the Venn diagram representing non-discarded circuits. The back-end fractions associated with each "region" are labelled in the figure. The meaning of "region" here can be made clear with the comment that F_1 is the fraction of circuits which can be sold as variant 2 but not as variant 1 or variant 3.

With the insight provided by this representation, the constraints on quantities which are needed in the methodology, for the hypothetical circuit can be written as,

$$q_1 \leqslant n^w E\{F_2 N^c\}$$

$$q_3 \leqslant n^w E\{[F_3 + F_4] N^c\}$$

$$q_1 + q_2 \leqslant n^w E\{[F_1 + F_2 + F_4] N^c\}$$

$$q_1 + q_2 + q_3 \leqslant n^w E\{[F_1 + F_2 + F_3 + F_4] N^c\}$$

$$q_5 \leqslant n^w E\{F_6 N^c\}$$

$$q_4 + q_5 \leqslant n^w E\{[F_5 + F_6]N^c\}$$

where E denotes the expectation operation.

Note that in general there may be as many as all $\bar{v}$ components of the $\mathbf{q}$ vector appearing on the left of the quantity inequalities.

In the general case, it is not difficult to see that there exists a matrix A of 1's and 0's and a column vector $\hat{F}$ with components the sum of back-end fractions, and depending on all the random variables introduced in the modeling and $\mathbf{x}$, $\mathbf{e}$, and $\mathbf{s}$, such that,

$$A\,\mathbf{q} \leqslant E\{\hat{F}(\{\mathbf{D}_{cd}\}_{d=1}^{n^d}{}_{c=1}^{N^c}, \mathbf{x}, \mathbf{e}, \mathbf{s})\, n^w\, N^c(\Lambda, \mathbf{x})\}$$

Equivalently, this is,

$$\{A\,\mathbf{q} - \hat{F}(\{\mathbf{D}_{cd}\}_{d=1}^{n^d}{}_{c=1}^{N^c}, \mathbf{x}, \mathbf{e}, \mathbf{s})\, n^w\, N^c(\Lambda, \mathbf{x}) \leqslant 0\}. \qquad (11.3)$$

12. The Integrated Profit Model

Substituting (5.1) into (4.1) gives the three major terms in the profit written in the "chronological order" in which the passage of material through production and sale affects profit,

$$\pi = -C^{FE} - C^{BE} + R.$$

Substituting (7.4), (10.1), and (11.2) into this gives,

$$\pi = -n^w[c^k + C^*(\Lambda, \mathbf{x})]$$
$$- C^{BE}(n^w, y^D(\mathbf{x})\mathbf{Y}^{WP}(\{\mathbf{D}_{cd}\}_{d=1}^{n^d}{}_{c=1}^{N^c}, \mathbf{x}, \mathbf{e}), \mathbf{Y}^{FT}(\{\mathbf{D}_{cd}\}_{d=1}^{n^d}{}_{c=1}^{N^c}, \mathbf{x}, \mathbf{e}), \mathbf{s})$$
$$+ \sum_{v=1}^{\bar{v}} p_v q_v\,(\mathbf{p}, \mathbf{q})$$

This profit random variable is one of the two integrands which when inserted into the integral of (6.1) yields an expectation needed in the methodology. In this case it yields the expected value of profit. The resulting expression is,

$$(E\pi)(n^w, \mathbf{x}, \mathbf{e}, \mathbf{s}, \mathbf{p}) =$$
$$\int\!\!\int\cdots\int\!\{\int\!\!\int\cdots\int -n^w[c^k + C^*(\lambda, \mathbf{x})]$$
$$- C^{BE}(n^w, y^D(\mathbf{x})\mathbf{Y}^{WP}(\{\mathbf{d}_{cd}\}_{d=1}^{n^d}{}_{c=1}^{N^c}, \mathbf{x}, \mathbf{e}), \mathbf{Y}^{FT}(\{\mathbf{d}_{cd}\}_{d=1}^{n^d}{}_{c=1}^{N^c}, \mathbf{x}, \mathbf{e}), \mathbf{s})$$
$$+ \sum_{v=1}^{\bar{v}} p_v q_v\,(\mathbf{p}, \mathbf{q}) \qquad (12.1)$$
$$\prod_{c=1}^{N^c}\prod_{d=1}^{n^d} [n_{\mathbf{D}_{cd}}(\mathbf{d}_{cd}; \mathbf{m}_c, \mathbf{v}_c)\, \mathrm{d}d_{cd}]\}$$
$$\prod_{c=1}^{N^c} [n_{\mathbf{M}_c}(\mathbf{m}_c; \mathbf{m}_m, \mathbf{v}_m)\, n_{\mathbf{V}_c}(\mathbf{v}_c; \mathbf{m}_v, v_v)\, \mathrm{d}m_c\, \mathrm{d}v_c] f_\Lambda(\lambda)\, d\lambda$$

As might be surmised, the other integrand to be inserted in (6.1) is the expression inside the outer curly brackets of (11.3). This yields

$$\iint \cdots \int \{ \iint \cdots \int \{ A\ \mathbf{q} - \hat{F}(\{\mathbf{d}_{cd}\}_{d=1}^{n^d}{}_{c=1}^{N^c}, \mathbf{x}, \mathbf{e}, \mathbf{s})\, n^w\, N^c(\lambda, \mathbf{x}) \}$$

$$\prod_{c=1}^{N^c} \prod_{d=1}^{n^d} [n_{\mathbf{D}_{cd}}(\mathbf{d}_{cd} ; \mathbf{m}_c, \mathbf{v}_c)\, \mathrm{d}d_{cd}]\} \tag{12.2}$$

$$\prod_{c=1}^{N^c} [n_{\mathbf{M}_c}(\mathbf{m}_c ; \mathbf{m}_m, \mathbf{v}_m)\, n_{\mathbf{V}_c}(\mathbf{v}_c ; \mathbf{m}_v, \nu_v)\, \mathrm{d}m_c\, \mathrm{d}v_c] f_{\Lambda}(\lambda)\, d\lambda \ \leqslant\ 0.$$

In summary, the objective of the methodology is to maximize the expected profit of (12.1) with respect to the design parameters n^w, $\mathbf{x}$, $\mathbf{e}$, $\mathbf{s}$, and $\mathbf{p}$, and $\mathbf{q}$, subject to the supply constraint of (12.2). As stated in the introduction, results from the development of computer-efficient techniques to solve this optimization are the subject of Part II.

Figure 1. Specification summary for a hypothetical linear circuit.

variant number	v_{os} (mv.)	t_s (μs.)	Temp (°)	package type	other treatment
1	25	3	25	*a*	-
2	25	5	25	*a*	-
3	10	7	25	*a*	-
4	10	7	25	*b*	-
5	10	7	25	*b*	burn-in

Figure 2. Representation of the back-end flow for the hypothetical linear circuit.

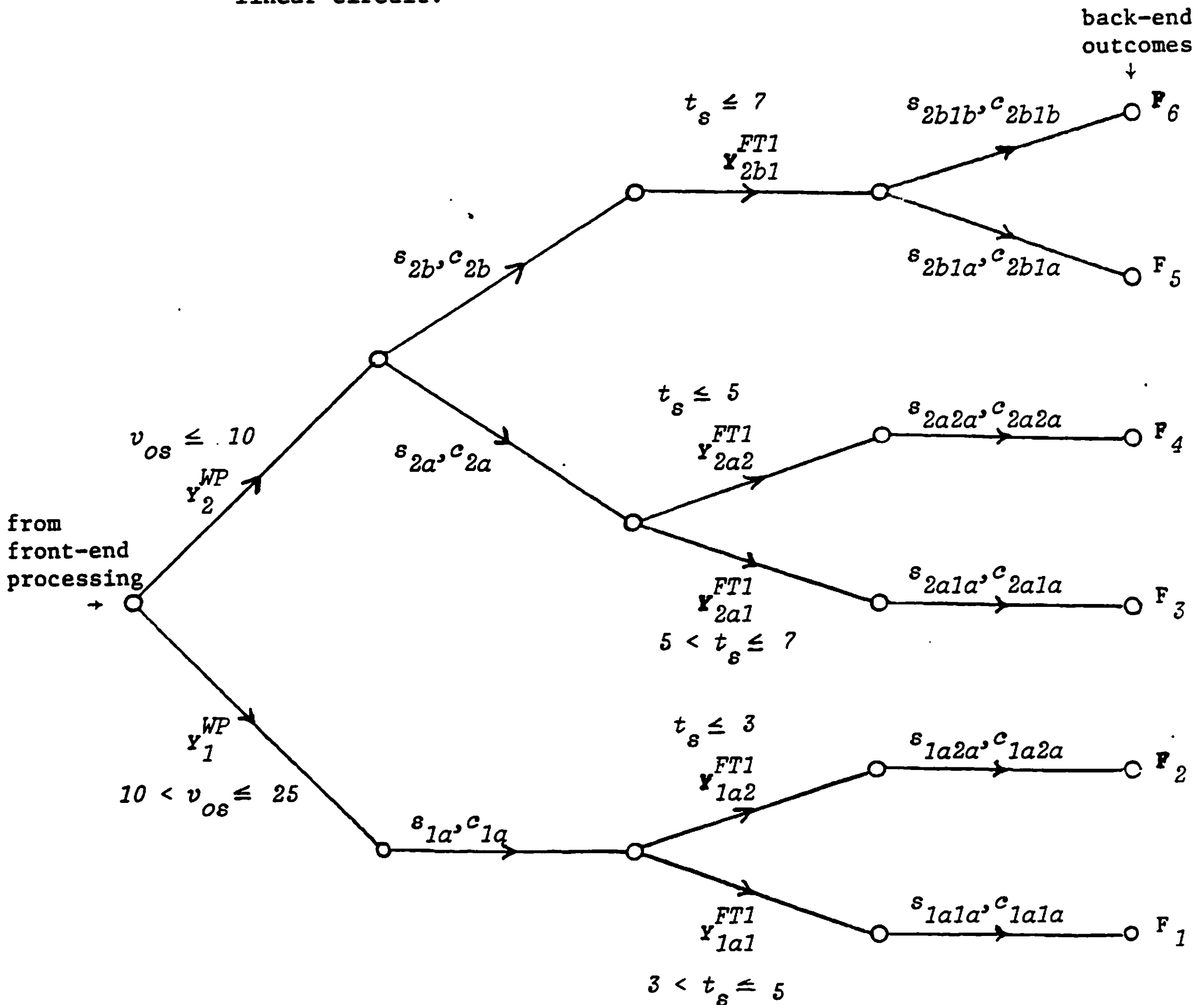
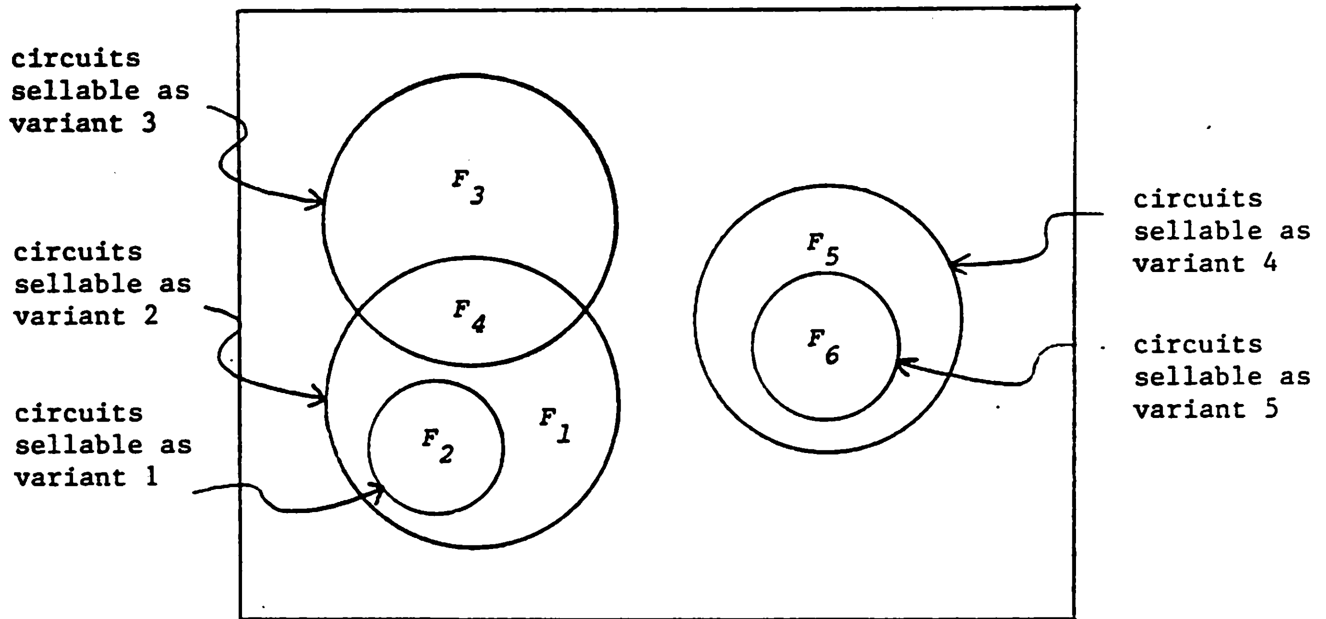


Figure 3. Venn diagram representing the set relationship between the variants of the hypothetical product.



References

[BRA81] R. K. Brayton, G. D. Hachtel, and A. L. Sangiovanni-Vincentelli, "A survey of optimization techniques for integrated-circuit design," *Proc. of IEEE*, vol.69, no. 10, pp.1334-1362, Oct. 1981.

[MAL81] W Maly, A. J. Strojwas, and S. W. Director, "Fabrication based statistical design of monolithic IC's," in Proc. ISCAS, pp. 135- 138,Apr. 1981.

[NAS83] S. R. Nassif, A. J. Strojwas, and S. W. Director, *FABRICS II, A Statistical Simulator of the IC Fabrication Process: Users Manual*, SRC-CMU Center for Computer-Aided Design, 1983.

[NAS84] S. R. Nassif, A. J. Strojwas, and S. W. Director, "FABRICS II, A Statistically Based IC Fabrication Process Simulator", *IEEE Transactions on Computer-Aided Design*, vol. CAD-3, No. 1, January, 1984.