EE TIMES

PIVOTAL ENTERPRISES

WHAT'S YIELD GOT TO DO WITH IC DESIGN?

DESIGN FOR YIELD

CONTENTS

Introduction........3

Functional Yield........4

Parametric DFY, Part 1........6

Parametric DFY, Part 2........10

Parametric DFY, Part 3........15

Particle Defect DFY........23

DFY Economics........28

Chip/Die Yield Prediction........

DFY Benchmarking........

Conclusions/Observations........

DESIGN FOR YIELD

OVERVIEW

Welcome!

Welcome to a new and ambitious project for EE Times, both in content style and publishing technique. Nearly a year ago, we launched the Silicon Engineering news section and its monthly counterpart, Silicon Engineering In Focus, to peel back the onion on an emerging segment of electronics. Simply put, as the industry undergoes more disaggregation from the old vertically integrated design and manufacturing companies, the greater the reliance designers have on the back end. This reliance comes into full relief at 130nm design rules and below, where the physics of manufacturing silicon plays havoc on complex SoC designs.

So we created these two focus areas to report on what's happening at the front end and what's happening at the back and how increasingly intimate they're becoming. In short, if TSMC sneezes, designers all over the world can get nasty colds.

Shortly after launch, longtime designer and now design consultant Mark Rencher proposed a series of articles illuminating one of the core issues in silicon engineering: design-for-yield. He is compiling a monthly "chapter" if you will, excerpted as space allows, in the print version of EE Times and then compiled in full here in Adobe Acrobat format.

Rencher has more that 17 years of business development, strategy planning and marketing experience in the electronics and software industries. He has held senior management positions at Philips, Motorola and Cadence. Additionally, Rencher is internationally recognized for his technical achievements by invitations for journal publications, conference presentations and patents.

Rencher is also the chairman of Four Point Enterprises an Arizona corporation that is the holding company for Pivotal Enterprises and Three Skye Capital, which provides venture capital financing to small Arizona, companies.

He has an MBA, MS from Arizona State University in Computer Engineering and a BS from Brigham Young University in Computer Engineering.

We hope you'll find this compilation useful in your design endeavors.

Brian Fuller
Editor in Chief
EE Times Network

Yields can be improved via design techniques

BY MARK RENCHER

Profitability on semiconductor products is directly related to yield. But what can design teams do about yield? The answer often gets the tag "design-for-manufacturing." But I would like to argue that DFM is different from design-for-yield.

DFY is the management of the design's sensitivity to the manufacturing process, while DFM is the management of technology constraints (rules, lithography) applied to a design. For example, design rule checks define the minimum wire spacing (pitch) but do not mention anything about relaxing the spacing to avoid particle defect shorts. A common example of DFY in practice is to evaluate an analog circuit's sensitivity to process variation.

Measurable results have been achieved that overwhelmingly demonstrate that design yield can be improved with DFY methods. In this and subsequent articles, we will explore business, design and EDA aspects of DFY and examine real-world results and solutions. Experts throughout the industry will be contributing by sharing their experience and results on this critical topic.

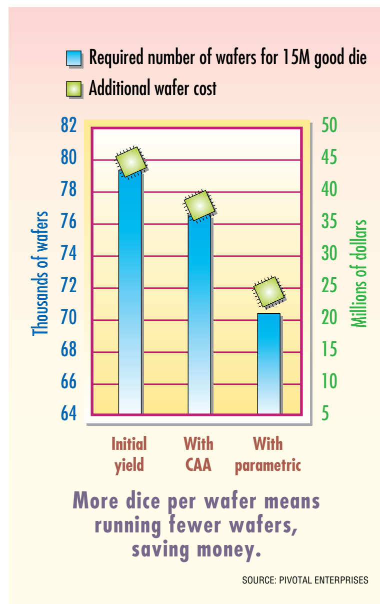
DFY should, at minimum, address the following issues:

• Parametric yield loss that is observed as parametric yield (process lot-to-lot and die-to-die variation);

• Parametric yield loss observed as functional yield (crosstalk, noise and on-chip timing contention);

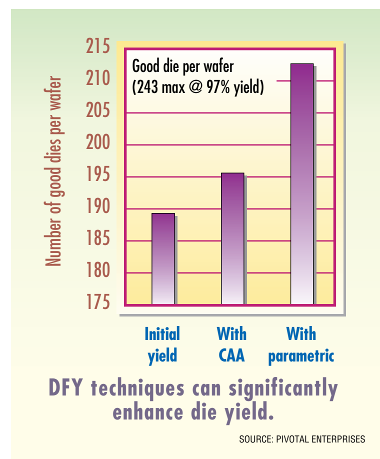
• Functional yield loss due to particles and contamination;

• Functional yield loss due to "intrinsic" marginalities (random occurrence of open contacts or vias); and



More dice per wafer means running fewer wafers, saving money.

SOURCE: PIVOTAL ENTERPRISES



DFY techniques can significantly enhance die yield.

SOURCE: PIVOTAL ENTERPRISES

• Test yield loss due to the product's not being designed to work in the test environment.

These classifications will be the framework of future articles. But perhaps the best place to start is with some definitions.

Particle defect yield is the sensitivity created by design elements such as interconnect, contacts or vias to creating a short or open from a defect particle.

Large system-on-chip designs can be sensitive to particle defects because of the density of the interconnect compared with the size of the particle defects and of the die. Structures such as stacked vias create addit-ional points of particle defect sensitivity.

Parametric yield is the design's sensitivity to process variations. For example, analog circuits can be sensitive to process variation, particularly when matching circuits and inter-connects are required. In addition, interconnect-dominated technologies below 0.18 micron are seeing increasing parametric signal delay. Therefore, new interconnect-modeling and simulation methods are required.

Neglected art

All these issues are well-known, but avoiding them is a neglected art among most design teams. The primary reason is that design yield has not been a critical product requirement: Manufacturing output has been measured by the number of wafers and of bad dice. The result is that design yield is left to the imagination and determination of the product and process engineer.

Other reasons have been the lack of commercially available design-automation tools, the lack of product yield specifications for the design engineer and the decision of foundries not to adopt good dice as their productivity measurement.

Academic organizations such as Carnegie-Mellon University and the University of Edinburgh (Scotland) have been developing

tools and methods to fill the gap left by the large EDA companies. Large semiconductor companies have also developed in-house software tools that have been successful in filling the existing EDA gap.

To amplify the economic impact of design yield on variable costs, consider the following simplified financial analysis:

- A new product is introduced into a mature 0.18-micron CMOS foundry.
- The committed product run rate to the customer is three years with a volume of 15 million parts.
- The initial wafer probe yield is at 78 percent (97 percent, or 243 good dice per wafer, is best in class).
- The cost per wafer is $2,500.

Applying critical-area analysis (CAA) EDA tools and wire spreading—two DFY techniques—a 3 percent yield improvement is achieved. Applying parametric yield analysis on top of CAA on the analog blocks results in an additional 7 percent improvement in yield.

Manufacturing yield becomes a predominant issue when fabs are running at or over capacity. During these conditions, it becomes clear that any increase in capacity is welcomed if it does not entail additional capital costs. Nevertheless, what is often overlooked is improving existng design yield as an alternative to purchasing new equipment and, in some cases, additional fabs.

The result is that DFY can assist in managing variable costs and can provide a direct bottom-line financial improvement when yield-verification methods are applied. By predicting the product yield before tapeout, IC design can have a direct dollar impact on the product's success.

In subsequent articles, with the help of leading academic experts, hands-on practitioners and EDA developers, we will explore the topic of DFY from business, design and EDA perspectives. The next installment will address functional yield—an area where the old rules no longer apply.

Old Rules No Longer Apply

BY DR. RIKO RADOJCIC AND MARK RENCHER

By Dr. Riko Radojcic
Director
PDF Solutions
San Jose, Calif.
and Mark Rencher
President
Pivotal Enterprises
Gilbert, AZ

Manufacturing yield is critical to the business success of products, and whole companies. Understanding the mechanisms behind the yield loss – thus enabling the rapid identification and fix of problems – is a fundamental component required for success. Predictability is also important, for production control, material management, timely product delivery, and other business factors.

Yield models are a principal business and operations planning tool. Accurate models mean a predictable business, and vice versa. This is especially true as the focal point of the industry evolves from production of standard products such as memories, to the production of ASIC, ASSP and SOC products that are characterized by a proliferation of customized IP and options within a product

Product Yields Stop Following the Old Rules

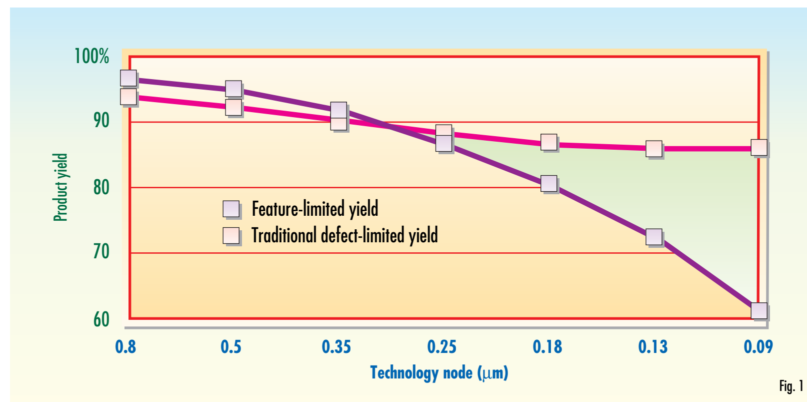
Experience through history indicates that functional yield, characterized by stuck-at faults at test, is driven by (a) defects, and/or by (b) systematic issues. Defects were typically particulates or other forms of gross contamination, which were assumed to be randomly distributed, and were modeled as a number of defects/cm2. Systematic issues were typically caused by aberrations in the process or design, and were typically managed through a set of rules and procedures.

Somewhere below the 0.18mm technology node, the world changed and the old rules-of-thumb based on the conventional wisdom ceased to apply. Product yields just stopped following the old trends. The particulate and defectivity contribution to the overall yield loss decreased, but the product yields did not show a corresponding improvement. Yield are now more limited by design content "features" than by random defects (Fig. 1), for three main reasons:

o Firstly, the sheer number of "features" – such as vias, contacts, lengths of metal wires, etc.. - has ballooned to a point where consistency of the order of failures per billion (fpb) is required in order to achieve any yield. For example, vias are critical features that limit overall yield, but their impact on yield cannot be realistically modeled by defects-per-area due to their relatively small open areas.

o Secondly, the tight tolerances associated with advanced technologies are making old 2nd-order issues be 1st-order problems. For example, plasma etch loading effects and corner rounding were always there, but were ignored as irrelevant 2nd-order phenomena. With advanced processes these mechanisms are killers and require OPC and/or complex rule sets.

o Thirdly, new materials and/or process technologies are precipitating a range of new systematic yield loss mechanisms. For example, in the move from aluminum to copper interconnects, newly required barrier materials such as tantalum-nitride exhibit new failure mechanisms associated with via formation



The modeled trends with defect-limited versus feature-limited yield with successive IC technology nodes. The defect-limited yield trend assumes constant failure rates for each successive technology node. The feature limited yield trend assumes a 50% improvement for each successive technology node, while the number of features grows >50% per node.

Given these trends, it is clear that some fundamental changes must occur to the rules for managing Functional Yield:

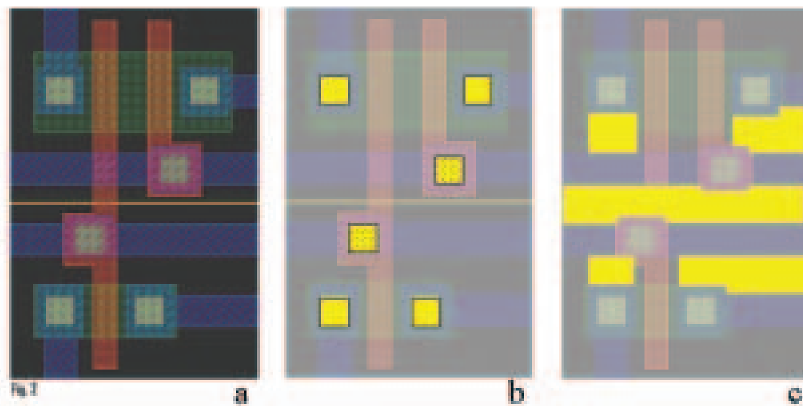
• Realistic cost and schedule constraints are such that engineering a

yield solution through design iterations is not an acceptable option for most products.

• To implement a design for yield, the yield attributes of a given design must be quantified. Managing the many failure mechanisms requires quantitative process characterization and design-process integration knowledge.

• A quantitative and credible yield prediction cannot be based on the traditional area based yield model, but must instead be based on models related to the design content, expressed in terms of the microscopic features that affect yield

The iterative hand-off between process development and design groups cannot work as it has in the past. Development of a set of design rules to describe the process capabilities, designing a product that is simply DRC clean, and then relying on product re-spins to fine tune the design or the process is no longer adequate or realistic. New methodologies and techniques must be adopted to deal with the current realities.



Schematics of A) a typical random sample layout, B) the 'features' in the same layout – in this case contacts and vias, and C) the 'critical areas' for metal shorts associated with the layout.

New Techniques, New Rules

The industry must trend towards a practice where a product is designed based on an optimum product-specific set of trade offs between performance, power, and cost (=yield) targets. A comprehensive feature-based methodology can be used to solve these problems.

• *Start with the Physics* – the various root cause failure mechanisms must be classified and understood, and the interactions between design attributes and process conditions catalogued.

• *Build a Model* – a catalog that relates failure mechanisms to product failure modes (shorts and/or opens) must be defined, based upon a suitable classification of mechanisms. Examples of model classifications include:

o Line-shorts and opens are controlled by the interaction between the distribution of defect sizes and the corresponding line widths and/or spaces (Fig. 2a). These phenomena can be modeled using the critical area concept (Fig. 2b)

o Other mechanisms, such as hole-opens, are characterized by the failure rates of a given feature (Fig. 2c) that corresponds to a given design lay out and neighborhood, and that captures the relevant interactions with other process and design attributes

o Application specific classifications to represent specific interaction with various design and process features. For example, models that describe antenna effects are affected by the ratio of the relevant areas and perimeters.

Note that the models must be quite granular to be useful. That is, in order to be useful to a designer, the models need to resolve the impact on yield of various lay out features.

• *Characterize the Process* – clearly the process characterization must be based on suitable test chips to give visibility into the individual failure mechanisms and to provide the observability into the failure statistics, such that the various models can be calibrated for a given process technology. Doing this correctly is not trivial, and must encompass the following aspects :

1. The individual yield failure mechanisms and corresponding models must be defined,
2. The parameters to be measured to characterize the given yield mechanism must be defined,
3. Structures that allow the desired measurement must be defined and sized appropriately to allow the desired statistical resolution,
4. The dependence of the mechanism on various design variables must be defined and described through a DOE of test structures,
5. A complete set of test structures for all target process layers and for all design dependencies must be placed on a single test chip reticle, such that they can be accessed and tested individually, and
6. All individual structures must be tested, on a statistically significant sample size, and the ensuing data reduced into a meaningful form for model calibration.

• *Characterize the Design* - design information for a given product must be extracted; statistics such as the critical areas, number of given features, characteristics of a given neighborhood effect, are required. Additional multi-layer effects, and information about connectivity or functionality may be required for some mechanisms.

• *Synthesize the Yield* - this 'bottom up' content-based approach produces yield predictions that are more accurate than the phenomenologically defined top-down area-based yield models. This is intuitively obvious, and is an approach that has been explored for many years. However, the approach is far more complex than the simple die area and defect density based models, and making it work requires methodologies that span multiple disciplines, along with comprehen-

sive technologies to make it all usable and accessible in practical terms.

New Perspectives in Yield Prediction

So how does it all come together into a useful solution? A vertical 'slice' of methodologies – spanning specialized Characterization Vehicle (CV) test chips, yield modeling, and design extraction –allows for an accurate yield simulation solution. It is then possible to construct a matrix that presents yield loss contributions, by individual process layers (rows), and by individual, user defined design blocks (columns), as illustrated in figure 3. An accumulation of the contributions – i.e. accumulation across all the rows and columns – provides the overall product yield prediction.

This bottom up yield synthesis really represents a paradigm shift in terms of how engineers relate to the yield predictions.

From the design perspective, the only useful information that could be derived out of the traditional area based yield models is that 'smaller is better'. Consequently all designers could do to ensure that a design is manufacturable is focus on making sure that the design is DRC clean and as small as possible. With the bottom up manufacturability approach, and with yield information presented in a form that is sufficiently granular to resolve the correlation between to the various design attributes and yield, the designer can understand and manage the manufacturability of his ICs. The degrees of freedom offered by this new approach encompass some of the following design activities:

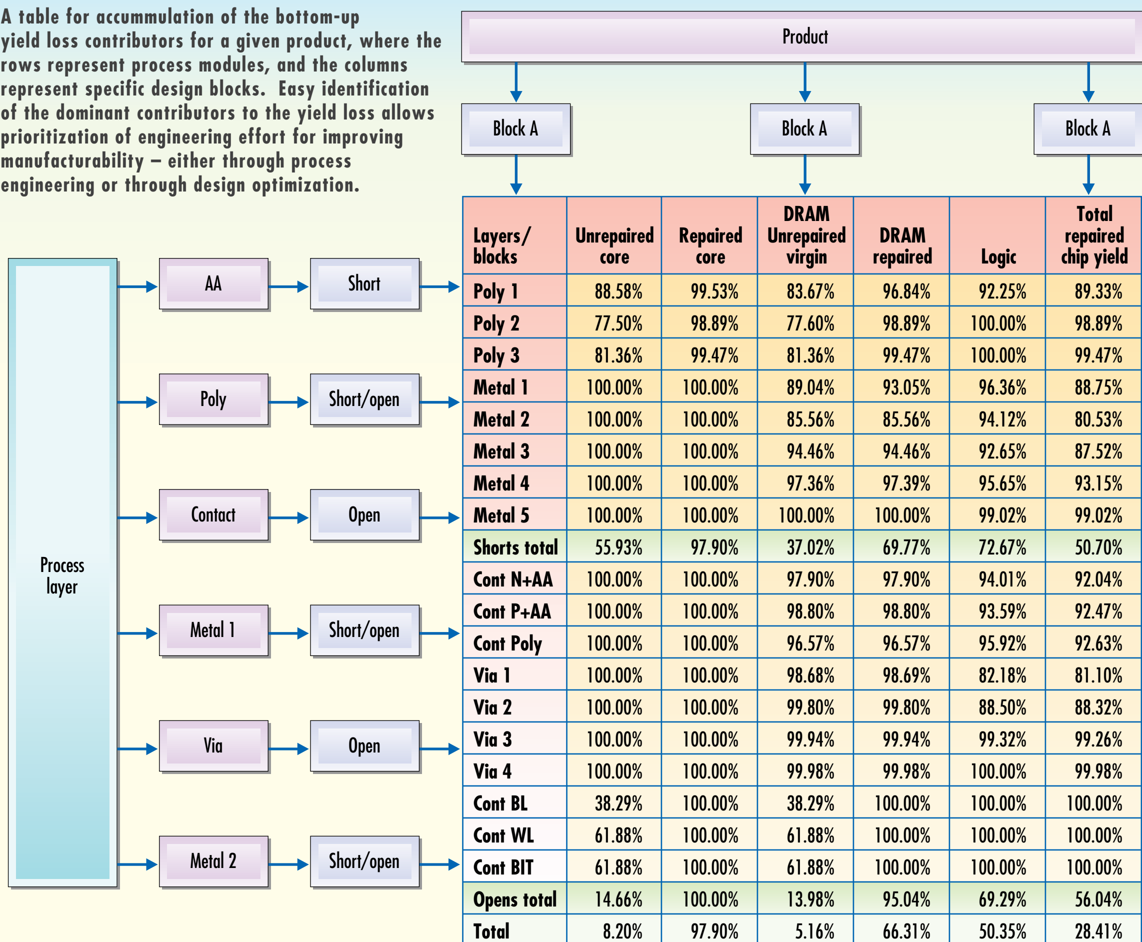
1. Design Planning: an understanding of the yield contribution from various individual blocks (functions) allow the designer to architect a chip to meet the target cost parameters.

2. Design Architecture: with the quantitative yield information available to a designer, architecture level decisions can be made in an informed fashion, and various redundancy schemes can be evaluated and selected to optimize product cost.

3. Physical Design: with granular yield information available to the designer, the layout, both at the IP level and Chip Level, can be tuned to optimize product manufacturability and to select the best trade off between all design attributes – performance, power, area and yield.

DFY in Action

Yield simulation is really only a 'foundation technology' - as it, by itself, produces no value. The



Layers/ blocks	Unrepaired core	Repaired core	DRAM Unrepaired virgin	DRAM repaired	Logic	Total repaired chip yield
Poly 1	88.58%	99.53%	83.67%	96.84%	92.25%	89.33%
Poly 2	77.50%	98.89%	77.60%	98.89%	100.00%	98.89%
Poly 3	81.36%	99.47%	81.36%	99.47%	100.00%	99.47%
Metal 1	100.00%	100.00%	89.04%	93.05%	96.36%	88.75%
Metal 2	100.00%	100.00%	85.56%	85.56%	94.12%	80.53%
Metal 3	100.00%	100.00%	94.46%	94.46%	92.65%	87.52%
Metal 4	100.00%	100.00%	97.36%	97.39%	95.65%	93.15%
Metal 5	100.00%	100.00%	100.00%	100.00%	99.02%	99.02%
Shorts total	55.93%	97.90%	37.02%	69.77%	72.67%	50.70%
Cont N+AA	100.00%	100.00%	97.90%	97.90%	94.01%	92.04%
Cont P+AA	100.00%	100.00%	98.80%	98.80%	93.59%	92.47%
Cont Poly	100.00%	100.00%	96.57%	96.57%	95.92%	92.63%
Via 1	100.00%	100.00%	98.68%	98.69%	82.18%	81.10%
Via 2	100.00%	100.00%	99.80%	99.80%	88.50%	88.32%
Via 3	100.00%	100.00%	99.94%	99.94%	99.32%	99.26%
Via 4	100.00%	100.00%	99.98%	99.98%	100.00%	99.98%
Cont BL	38.29%	100.00%	38.29%	100.00%	100.00%	100.00%
Cont WL	61.88%	100.00%	61.88%	100.00%	100.00%	100.00%
Cont BIT	61.88%	100.00%	61.88%	100.00%	100.00%	100.00%
Opens total	14.66%	100.00%	13.98%	95.04%	69.29%	56.04%
Total	8.20%	97.90%	5.16%	66.31%	50.35%	28.41%

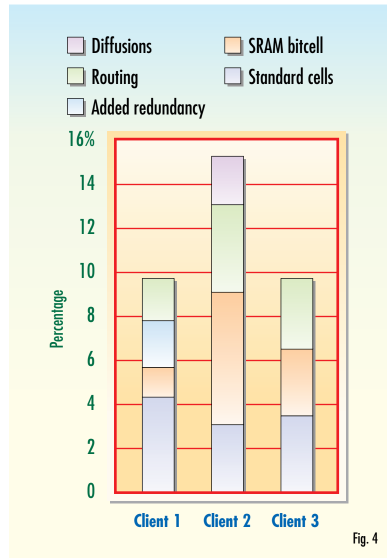
A table for accummulation of the bottom-up yield loss contributors for a given product, where the rows represent process modules, and the columns represent specific design blocks. Easy identification of the dominant contributors to the yield loss allows prioritization of engineering effort for improving manufacturability – either through process engineering or through design optimization.

value is in yield optimization, which can be achieved through process tuning and/or through design tuning, based on the quantitative information derived from the yield simulation 'foundation technology'.

The methodologies described here are the foundation for a comprehensive yield simulation technology developed by PDF Solutions Inc. Over the last several years this technology has been deployed and proven in numerous projects. Typically design optimization is deployed in concert with process optimizations, or as stand alone activities used in situations where the process is fixed and/or inaccessible. Design optimization includes:

• *Layout out optimization of standard cells and IP elements:* With quantitative and accurate yield prediction, the lay out of pre-designed elements used in product designs is optimized. Yield is optimized to meet fixed performance, power and/or area constraints, or may be traded-off versus other design requirements. Yield hardened elements can be swapped in place of standard elements, for pre-existing designs, or can be used in new designs for true Design-for-Yield process

• *Redundancy optimization of memories.* Use of optimum redundancy amount and scheme, especially for designs that contain large memory blocks, based on quantitative yield assessment of the overall die yield is very important. The yield benefit associated with use of redundancy can be very significant, especially during the process ramp phase, when the process technology is not fully matured and memories are not yielding to their full potential



The net yield improvements for 3 specific projects where PDF Solutions deployed some of the technologies described here. All 3 projects involved advanced and very complex SoC style ICs: A) 0.13 _m CMOS for telecommunications, B) 0.15 _m embedded-DRAM for processors, and C) 0.13 _m CMOS for graphics.

• *Bit Cell optimization of memories*: For small memory blocks use of redundancy is not appropriate. However, many small on board memory blocks are not performance critical and use of relaxed bit cells can add up to significant yield improvement for die with multiple small memory blocks. Again, the optimization involves a trade off between performance and yield, and cab be done only if quantitative yield information is available during the design phase

• *Interconnect Optimization:* The interconnect can be optimized, and if necessary yield characteristics can be traded off against the performance and/or density attributes. Use of wire-spreading and /or redundant vias have been well documented as opportunities for yield hardening, and may have a significant impact on product yield. Again these approaches involve a trade off between multiple yield loss contribution mechanisms, and the optimum is defined based on quantitative yield simulation.

Note that optimization for manufacturability of given product may require one, some, or all of these activities, depending on the design content, the targeted process technology, and the cost-benefits associated with each of the changes.

Figure 4 summarizes the experiences with such projects – where the yield simulation technology has been used purely to optimize designs (i.e. without process optimization). The typical net benefits routinely demonstrated are of the order of 10% to 15% improvement in yield, averaged over the product life cycle. When the technology is also used to optimize process, greater benefits are demonstrated.

Clearly, Design for Yield is a concept that has been around for a while. However, a paradigm shift is taking place now. The nature of the advanced process technologies, the complexity of SOC products, the costs of product development, and the accelerating market pressures, are all contributing towards a new reality that dictates the necessity for implementing Design for Yield practices. The risks of not dong it, and the liabilities associated with failing to be right the first time, are unacceptable.

It all begins and ends with Spice

BY MARK RENCHER And Colin McAndrew

BY MARK RENCHER,
President Pivotal Enterprises,
Gilbert, Arizona
AND COLIN MCANDREW ,
Motorola,
Tempe, Arizona

One important aspect of DFY is design for parametric yield. IC manufacturing processes are subject to statistical fluctuations in electrical parameters, and these must be accounted for during design. There are two parts to achieving predictable (statistical) design performance, the first is statistical design techniques, and second is the appropriate statistical models. Contributions in this article will include the current practices of statistical characterization at the SPICE level, linking the data to SPICE simulators and statistical logic modeling for nanometer technologies.

Zhihong from Cadence summarizes the importance of this effort. "While a lot of efforts have been put into the standardization of the compact model, it is time to fill the missing link between the ideal device behaviors vs. those inside a real design."

Dr. Sharad Saxena, of PDF Solutions, elaborated on this by explaining that the standard industry practices that focus on the accuracy of the nominal SPICE model are somewhat misplaced. "When all is said and done, a successful design must meet its specifications across the entire range of process variables, and the accuracy of the nominal model is not as relevant as the process variability. Typical process variability is somewhere of the order of 10% to 30% so that forcing the accuracy of the nominal model to better than a few % does not make sense".

THE BEGINNING

The use of the term "statistical model" although generically used in the industry does not accurately define the different types of statistical analysis techniques used during IC design, and the different types of analyses require different types of models.

There are two main classes of "statistical" models: enumerated, discrete case models (also called "skew" or "corner" models); and distributional models. There are two components of distributional models, commonly termed "global" variation and "local" variation. Local variation is often termed mismatch (parametric, not impedance). For modeling, simulation, and physical understanding, it is better to view these as components in two forms:

- *Correlated between devices* – the correlated component is independent of device layout attributes (length, width, orientation, surrounding topography, etc.),
- *Uncorrelated between devices* - the uncorrelated component is strongly dependent on device layout attributes, to first order increasing as the reciprocal of geometry.

Figure 1 shows data (from [1]) and models for the correlated (global), uncorrelated (local), and total variation in MOSFET threshold voltage in a 0.12mm CMOS technology. The uncorrelated (mismatch) component has historically been considered only to be important for analog circuits, however with decreasing geometries and increasing speeds it is becoming equally important for digital circuits.

Clearly the mismatch component can no longer be considered to be much smaller than the global variation. This has ramifications for both modeling and simulation, which are detailed below. Figure 2 provided by Philips illustrates this impact by contrasting supply voltage and threshold voltage variation as a function of feature size. Although the mismatch proportionality constant Avt improves with technology, the effective mismatch for minimum sized transistors, characteristic for SRAM cell design, still goes up. At the same time, signal space decreases because of the lower supply voltages, leaving virtually no dynamic range.

There are many approaches to statistical modeling, from fast and simple to slow and complex. One key observation needs to be factored in to all approaches to modeling process conditions or variation: statistical process capabilities are known imprecisely when case and distributional mod-

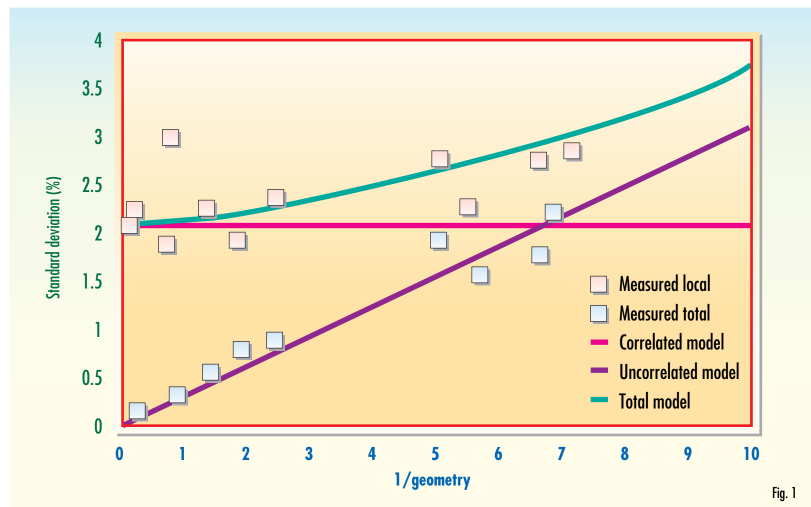


Fig. 1

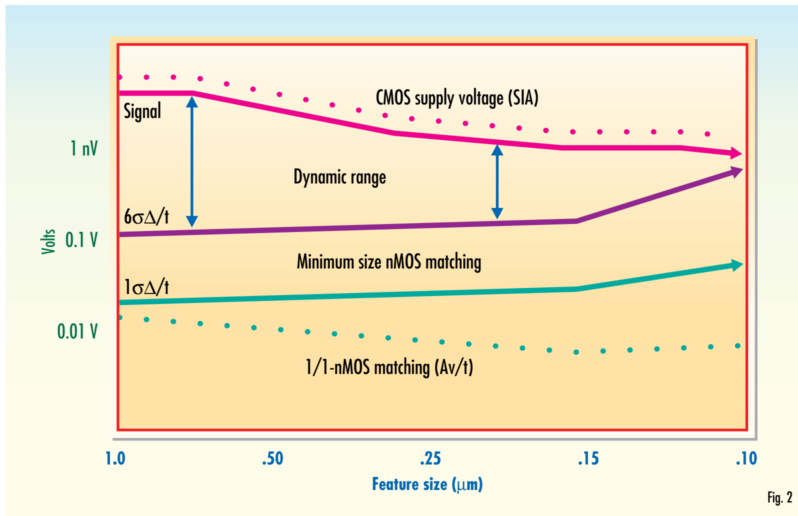


Fig. 2

els are required to be generated. Many designs require case and distributional modeling, along with statistical simulation, early in the technology lifecycle. It is unrealistic to provide accurate statistical data, over multiple manufacturing tools and sites, over all variants of a technology that is continually under development (e.g. a baseline CMOS process to which options for precision passives, various performance BJT modules, high voltage devices, etc. are added over the years) early in the technology lifecycle.

To address this issue of providing early case and distributional modeling and statistical simulation, the early statistical variations can be based on statistical limits. These limits are created from:

- limited lab and/or fab data,
- extrapolations from previous technology generations,
- product requirements,
- equipment and engineering knowledge,

but whether the long-term variation is 4.5% or 5.5% is not known precisely.

There are two main approaches to statistical modeling: numerical and physical.

- *Numerical* - This approach is based on extracting SPICE models (typically with 10's to 100's of parameters) from a set of sample die and wafers, and then using techniques like Principal Component Analysis (PCA). PCA determines the correlations between the SPICE model parameters and a reduced number (typically less than 10) of parameters. A major feature of PCA is that characterization groups do not need to know the intimate details of the technology to produce distributional models, the process is data driven.

In fact, "the choice of the appropriate statistical modeling methodology is a function of the process maturity, quality of the data, and the end application", says Dr. Sharad Saxena of PDF Solutions. Figure 3 illustrates the decision hierarchy used by PDF. For immature technologies, a TCAD device simulation environment is used to derive the statistical SPICE models, from the statistical process characteristics - even before a modeling-ready quality of material is available. This is a powerful methodology for combining concurrent engineering and statistical design methodologies. In cases where good device characterization data is available, Principal Component Analysis techniques are used to derive the correlation among the device parameters. These correlation coefficients are then embedded in the statistical SPICE models allowing accurate tracking of device characteristics as a function of principal process parameters.

- *Physical* - Most people doing SPICE models believe strongly they should be based on a small number of physical and uncorrelated parameters, commonly termed "process parameters" (such as lateral geometry variations, sheet resistances or doping levels, oxide and other layer thicknesses, flatband voltage, recombination/generation lifetimes, etc.). Not only does this simplify parameter extraction and allow efficient and reasonably accurate retargeting of models when a process change occurs, but it forms a natural and efficient basis for statistical modeling (Fig.4). This is because physically it is the manufacturing variations in these process parameters that, to first order, cause the variations in device electrical performances, and because these parameters can be formulated to be statistically independent. In addition, understanding of the physical source of variation can guide process improvement.

Many SPICE models are based on physical process parameters, but some, notably BJT models, are not. For the latter, model parameters need to be formulated as functions of process parameters,

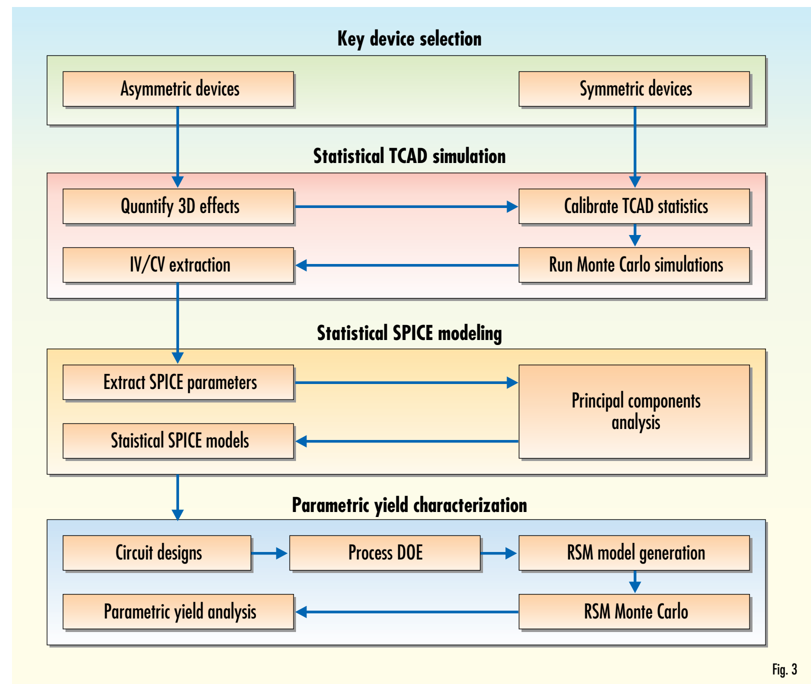


Fig. 3

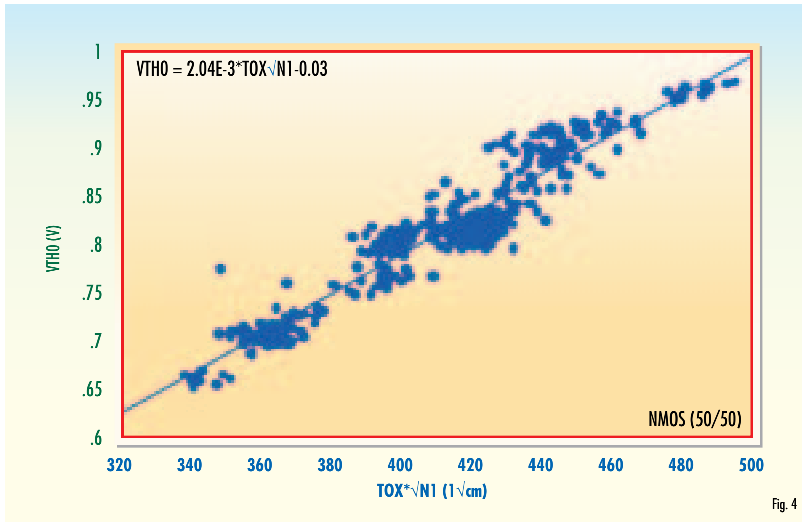


Fig. 4

by writing them as expressions rather than numbers in SPICE model files.

Case vs Statistical Modeling

Corner simulations and Monte Carlo simulations are two of the most common statistical design techniques. And these require case and distributional models, respectively. Often direct modeling of parameters is emphasized for statistical modeling, but it must be kept in mind that the real goal of statistical modeling and simulation is accurate representation of the statistical variation of circuit electrical performances.

Given that, the obvious way to generate accurate case files is to specify desired, often, variations in key electrical performances of devices that correlate to variations in circuit performance.

• ***Case Modeling*** - Generating and using, case files is not a simple matter. Fundamentally, a specific case file can only guarantee to target one figure of electrical performance for a device of one specific geometry at one set of bias conditions. If you are concerned about accurately bracketing arbitrary performances for all geometries and biases, case files guarantee nothing. So why are they so widely used? And why do they work reasonably well in many cases? Digital CMOS drives the industry, and circuit topologies and devices geometries (minimum length) are similar between many digital CMOS circuits. There are two key figures of merit of electrical performance, speed and power, and these are correlated, and are primarily controlled by drive current and capacitance ( ). Bracketing drive strength and capacitance in case files allows reasonable simulation of spreads for digital circuit speed and power, and for other quantities correlated to , for example op-amp slew rate. But it does not provide information about other figures of merit for other circuits that are not correlated to , for example Schmidt trigger hysteresis, op-amp input offset voltage, or op-amp phase margin. Techniques to generate specific case files for arbitrary circuits have been developed [2][3], but they are not widely used.

A common way to generate case files is to directly measure manufacturing variations in the process parameters, . These variations are then used directly for distributional models. For case files, specific variations, often , are introduced into the , in appropriate directions to make variations in electrical performances, , as extreme (“worst” and “best”, “slow” and “fast”) as possible. Although this is a widely used approach, it has some significant limitations.

First, by introducing in each of process parameters, the likelihood of this combination of parameters actually occurring in manufacturing is . This means the case models generated by this process are pessimistic, and will embody a much wider variation in device and circuit electrical performance than is observed in practice. Ring oscillator simulations (See figure 5) with extreme case MOSFET models generated by this approach typically predict a spread in gate delays that is about twice that observed in practice. Circuit designed to the case files limits will work, but because of the pessimism will be significantly under-using the capabilities of a process. This problem is often termed the “curse of dimensionality.”

• ***Distributional Modeling*** - Distributional modeling is used for some form of Monte Carlo simulation and is generally used to determine variations in circuit performances that do not correlate to case files (typically for analog circuits) (Fig. 6). These require distributional models for .

A common characterization method is to directly measure manufacturing variations in the process parameters, . This can sometimes be difficult and require special test structures and measurements, and different approaches can give different values, which is undesirable. It also does not directly address the prime goal of statistical modeling, to emphasize modeling of electrical performance over modeling of

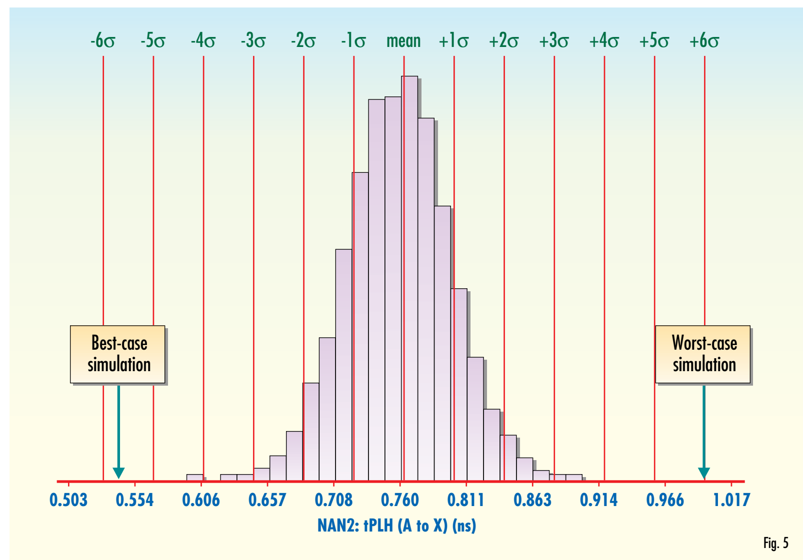


Fig. 5

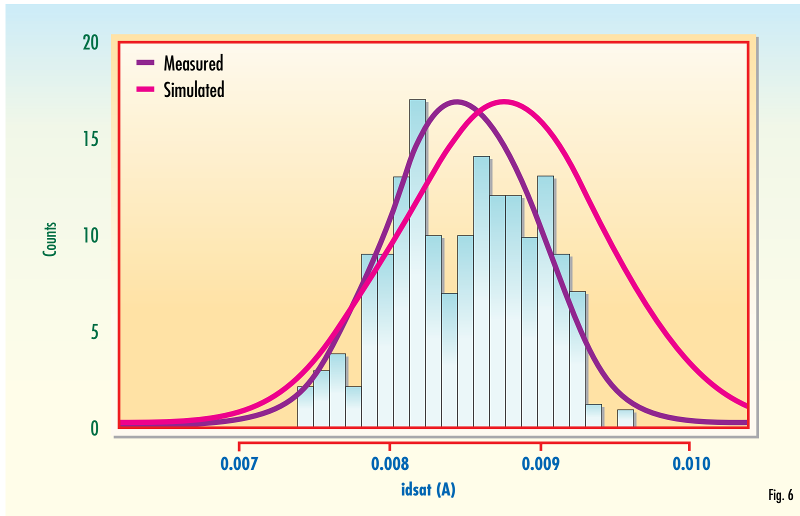


Fig. 6

parameters. If the assumptions of linearity or normality are not good approximations, transformations of the parameters may be necessary. For quantities that have a large variation, such as the non-ideal component of base current, an exponential transformation is used, and then results in a log-normal distribution.

Historically, the mismatch (uncorrelated) variation has been significantly smaller than the global (correlated) variation. This meant that the two types of variances could be characterized independently, typically the global variation first, and the mismatch variation was then done and added to the global variation.

Things are no longer that simple.

First, the electrical test data, for example of oxide thickness , inherently depends on geometry. This is typically not taken into account in defining test structures and measurements, and more critically is not taken into account in use of the data for modeling and simulation. is generally measured on large area structures (although this causes difficulties for modern thin and leaky gate dielectrics), and in use is then assumed to have the same variation for small devices. This means modeling and simulation underestimate the variation in for small devices.

Second, if the standard practice of independent, sequential characterization of global and local variations is followed, and the local variations are used additively on top of the global variations, then their contribution will be double-counted.

The entire concept of case files and corner simulation, which has been a corner stone of the industry for decades, becomes invalid.

RepresentingStatistical Models in Spice - Process Blocks

The flow diagram (figure 7), provided by Philips Semiconductors [4] illustrates the information linkage between design, manufacturing and test.

A key issue of linking the various data sources. This has issue has been solved by Philips and the use of their process blocks [5]. The parameter related part of DFY is concerned with the fit between product functionality and process variations, i.e. the fab's total process window. Statistical design tools (process blocks), timing accuracy and designed-in reliability are the three key words in this part of DFY.

The End – Case examples

"The entire concept of case files and corner simulation, which has been a corner stone of the industry for decades, becomes useless." To finalize this article on Parametric DFY – it begin and ends with spice, 4 analog and digital examples are included to demonstrate the success that has been achieved in predicting design yield.

The following CMOS op-amp example provided by Motorola confirms the modeling and simulation methods with silicon.

The following DC, AC, timing measurements were compared to spice simulations using statistical spice models. (Table 1)

9 bit DAC

An example provided by PDF has used the Statistical SPICE modeling techniques in optimizing the trade offs between performance and yield. A 4% yield improvement was achieved by characterizing and modeling the device mismatch (Table 2). More on this topic of design yield improvement is discussed in article 4.

Statistical Logic Timing Characterization

Recent articles in EEtimes by Ron Roher, Michael Santarini and Richard Goering outlines the need for statistical timing analysis. The following methodology specifically links process, voltage, temperature and loading conditions to the actual gate timing.

Using exactly the same distributional models for analog design, gate logical functions (rise, fall, delay, setup, hold, etc) are characterized using spice simulations.

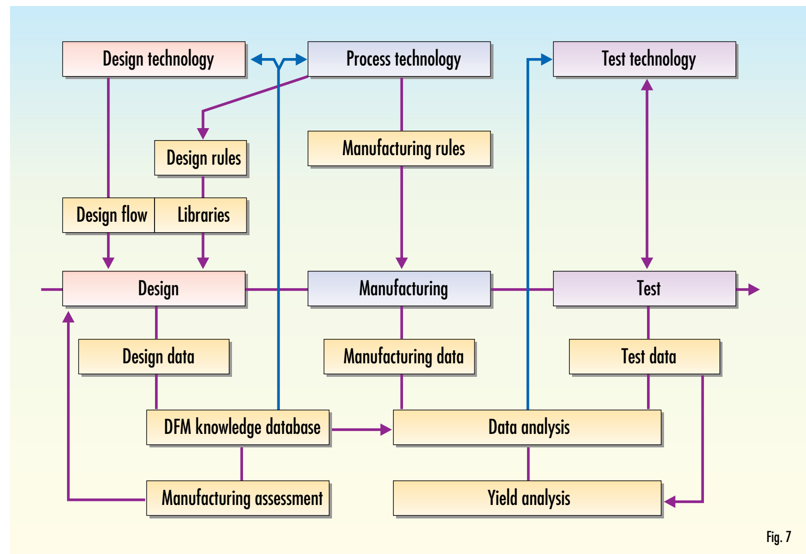


Fig. 7

Condition	Predicted	Measured
Original design	13%	15%
Corrected design	10%	11%

Table 2

(Table 3) The result is a polynomial model of gate logic functions as a function of Process, Voltage, Temperate and Loading (PVTL) variation. The form of the statistical gate model is tPLH = f(p1,p2,pn, temp, supply, tINPr,f).

This polynomial model is supported by Verilog, VHDL, AHDL, Spice, C and C++.

The final result is a statistical gate modeling methodology that links "real" process conditions to "real" gate performance and can be used by existing languages. Additionally, this method opens the door for mixed level simulation transistor & gate to be controlled by the same process parameters, ensuring continuity between the domains.

Table 4 provided by Motorola compare the timing results of a NOR2 and DFFP. Note the equivalency between the Monte Carlo, spice and statistical gate model. All are using the same process parameters and distributional models, while the ASIC timing model has significant error.

Function	Spice (typical)	Measured mean	Mean model	Measure standard deviation	Standard development model
psrr+ (dB)	93.35	94.61	93.31	9.48	12.99
psrr- (dB)	92.92	86.71	91.99	9.17	17.37
ugbw (MHz)	4.62	3.69	4.41	0.11	0.97
pM (deg)	51.24	42.0	50.48	2.35	1.791
avol (dB)	108.4	—	108.1	—	1.046
sr+ (V/in)	4.94	1.72	4.918	0.046	0.27
sr- (V/us)	4.67	3.67	4.63	0.252	0.95
vos (mV)	-0.132	0.244	-0.095	0.792	3.82
cmrr (dB)	119.0	83.6	85.85	9.82	15.64

Table 1

Delay	Poly	TOX	Supply	TempDC
T_{PHL} AtoX	1.5e-9	7.2e-12	-8.5e-10	6.5e-12
T_{PLH} BtoX	1.8e-9	9.6e-12	-6.2e-10	1.0e-11
T_{PHL} BtoX	-3.7e-10	-4.5e-12	5.9e-11	-3.1e-12
T_{PLH} BtoX	-1.7e-8	8.1e-12	-5.7e-10	1.3e-11

Table 3

NOR2 (ns)	t_{PLH} AtoX	t_{PHL} AtoX	t_{PLH} BtoX	t_{PHL} BtoX	t_{RISE} X
100 Monte Carlo	1.16	0.966	1.11	0.900	2.30
Nominal spice sim	1.19	0.984	1.14	0.906	2.34
Statistical gate model	1.18	0.981	1.12	0.904	2.37
ASIC timing model	0.957	1.048	.991	1.013	2.46

DFFP (ns)	t_{PLH} CKtoQ	t_{PHL} CKtoQ	t_{PLH} CKtoQB	t_{PHL} CKtoQB	t_{RISE} QB
100 Monte Carlo	1.47	1.48	1.63	1.65	1.18
Nominal spice sim	1.51	1.51	1.68	1.70	1.20
Statistical gate model	1.49	1.49	1.65	1.67	1.22
ASIC timing model	1.689	1.587	1.573	2.043	1.1

Table 4

References

[1] R. Difrenza, P. Llinares, S. Taupin, R. Palla, C. Carnier and G. Ghibaudo, "Comparison between matching parameters and fluctuations and the wafer level." Proc. IEEE ICMTS, pp. 241-246, 2002.

[2] S. R. Nassif, "Statistical worst-case analysis for integrated circuits," in Statistical Approach to VLSI, S. W. Director and W. Maly (Eds.), North-Holland, pp. 233-253, 1994.

[3] N. Lokanathan and J. B. Brockman, "Efficient worst case analysis of integrated circuits," Proc. IEEE CICC, pp. 237-240, 1995.

[4] Date2000 http://jamaica.ee.pitt.edu/Archives/ProceedingArchives/Date/Date2000/papers/2000/date00/pdffiles/01d_1.pdf

[5] Medea T651 http://medeaplus.net/medeanews/profiles/pp_t651.pdf

Making Heads and Tails of Statistical Design

BY MARK RENCHER AND JACK SIFRI

BY MARK RENCHER
President
Pivotal Enterprises,
Gilbert, Arizona
AND JACK SIFRI
Agilent EEsof EDA,
Westlake Village,
California

Statistical or six sigma design is establishing strong relationships between product defects and yields, reliability, cycle time, inventory, schedule, and cost.

Including Six-sigma statistical methods in the overall design process is of extreme importance in ensuring robustness, reliability, first pass success and higher yields, all of which translate into increased bottom line profitability.

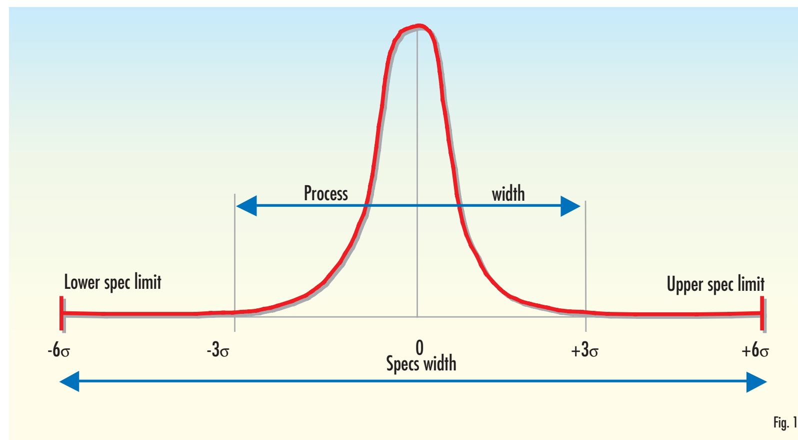
This article will discuss parametric DFY or statistical design methodologies and provide examples of the impact of using statistical design on actual designs.

What is parametric yield and how is it measured?

To achieve six-sigma (6s) quality, the product must first be designed so that the manufacturing processes are capable of yielding 99.99966% in the product. For example, assume the electrical design of an amplifier requires that the gain be greater than 12 dB. Both the manufacturing processes and the design must be capable of yielding 999,996.6 out of one million amplifiers that meet their requirements and only 3.4 out of the million amplifiers that do not meet their requirements, i.e., gain less than 12 dB.

So, six-sigma design process translates into perfecting the design process to ensure products that are at least 99.99966% defect free. As a result, the six-sigma design process will also:

- Decrease design and manufacturing cycle time



Six Sigma distribution

- Decrease inventory, scrap, and rework, and
- Increase profitability and market competitiveness.

We can see that achieving six-sigma quality depends on two design factors. The first factor is the customer requirements. Achieving six-sigma quality will be more difficult with tougher performance requirements. Using the amplifier design from the example above, higher yields will be achieved for the amplifier if the minimum requirement is 10 dB.

The second design factor is the anticipation of manufacturing process variations in the design process. If the design is very sensitive to component values, the gain distribution will be wider as component values vary during the manufacturing process. As a result, more amplifiers will fall out of the performance limits and the end product yield will decrease. On the other hand, if the amplifier is designed so that the output response is insensitive to component variations, gain variation will be minimized and yield will be maximized. Once this is completed, the sensitive elements remaining can be tightly controlled. This can be accomplished by either using tighter tolerance parts in the design or by monitoring the sensitive elements carefully during the manufacturing process.

To link the manufacturing control to design quality, six sigma design uses the "Process Capability Indices" (Cp and Cpk) . The conditions required for using Cp and Cpk are:

* The process must be in control
* Distribution of the process output must be normal or approximately normal.
* The process output must be independent.
* The process output must be a

stochastical parameter?

Process capability (Cp) is a measure of how closely and consistently the process can meet customer requirements. The customer requirements are indicated by the target value, the Upper Spec Limit (USL) and the Lower Spec Limit (LSL). This window is called the spec width. I mentioned above that the process is in control if all outputs lie within the upper and lower control limits. This process window is called the process width. If the UCL and LCL also happens to be the customer's target window (The USL and LSL), then Cp=1 using the following equation:

$$Cp = \frac{\text{Specification Width}}{\text{Process Width}} = \frac{(USL - LSL)}{\pm 3\sigma}$$

What does this mean? If Cp = 1, the overall design is achieving three sigma quality. In order to achieve six sigma quality, Cp must equal 2.0 or in other words, if the spec width is twice the process width, we should meet this requirement 99.99966% of the time (see figure 1).

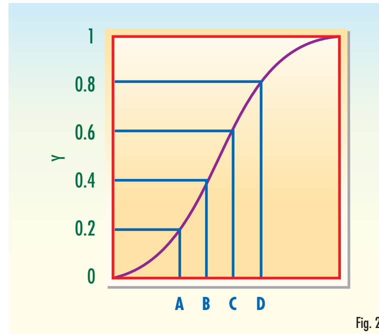
Statistical Design Methods

DFY article 3 discussed the development of relating process parameters to spice models. This important first step is essential in providing statistical design methods, that an IC designer can use, to measure Cp and Cpk.

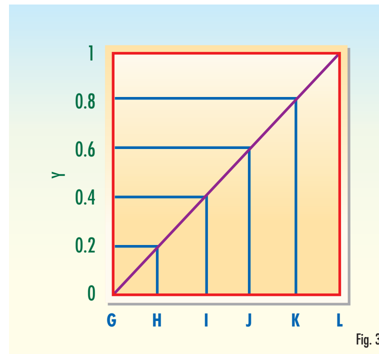
To simulate the process variation in circuit design, there are three prefered methods, corner, monte carlo and design of experiments/response surface methodology (DOE/RSM).

Corner analysis is so blatant in giving incorrect answers, it has no ability to predict accurate Cp and Cpk values and will not be discussed. See article 3 for a detailed justification why corners analysis is improper. Monte Carlo can predict Cp and Cpk, the simulation cost is very high due to the number of simulations (over 100) required to accurately predict the tails of the distribution.

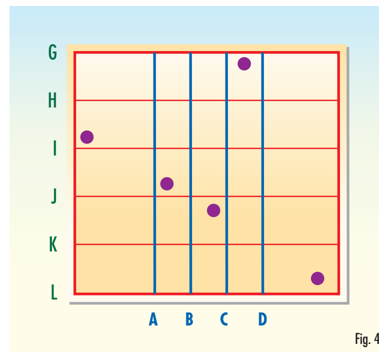
Design of Experiments (DOE)



LHS sample for non-linear distribution (X1)



LHS sampling of linear distribution (X2)



Complete LHS for X1 and X2

and Response Surface Methodology (RSM) is a well know statistical practice. In the early eighties publications by Motorola, Philips, Texas A&M, CMU documented successful application of DOE/RSM on transistor level circuits. The benefits of this method include:

- Process parameter sensitivity to circuit performance
- Reduction in number of simulations to achieve accurate results
- Polynomial model creation with error compared to simulations

Historical experience by Motorola has determined that the "best" experimental design for circuit simulation is the Latin Hyper Cube.

Latin Hyper Cube – accurate parameter distribution sampling

Latin hyper cube sampling was developed to address the need for uncertainty assessment for a particular class of problems. Consider a variable Y that is a function of other variables X1, X2, ..., Xk. This function may be very complicated, for example, a computer model. A question to be investigated is "How does Y vary when the Xs vary according to some assumed joint probability distribution?" Related questions are "What is the expected value of Y?" and "What is the 99 the percentile of Y?" The LHS software supports this approach for generating samples of the Xs. See Figures 2 and 3. The LHS output, for n Monte Carlo repetitions, is a set of n vectors of input variables (each such vector is k-dimensional). See Figures 4. Each input vector can then be evaluated by the function or program to generate n values of the result Y (Y may be any number of circuit performances of interest). This approach yields reasonable estimates for the distribution of Y if the value of n is quite large. Multiple or different measurements of Y can be completed at the same time. For example a logic gate timing characteristics can be measured in one LHS simulation run.

Latin hypercube sampling selects n different values from each of k variables X1, ... Xk in the following manner. The range of each variable is divided into n non-overlapping intervals on the basis of equal probability. One value from each interval is selected at random with respect to the probability density in the interval. The n values thus obtained for X1 are paired in a random manner (equally likely combinations) with the n values of X2. These n pairs are combined in a random manner with the n values. This is the Latin hypercube sample. It is convenient to think of this sample (or any random sample of size n) as forming an (n ´ k) matrix of input where the i th row contains specific

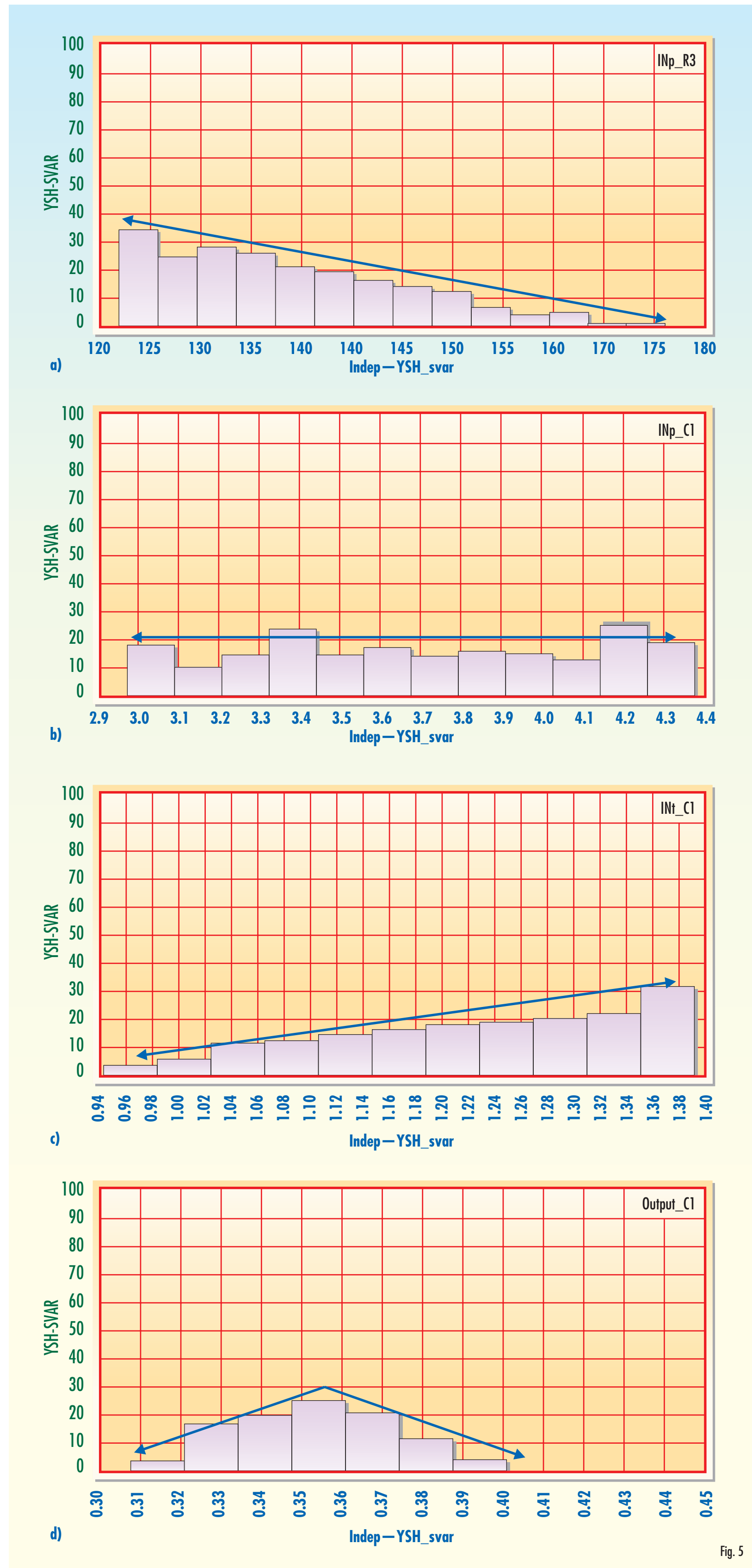
values of each of the k input variables to be used on the i th run of the computer model.

Yield Sensitivity Histograms

In order to find out which components in the design cause high variability to the output parameter, an easy to use method is Yield Sensitivity Histograms. Yield-sensitivity histograms are plots that determine which components in the design cause high variability to the output parameter for helping designers to locate and fix problematic and sensitive parts of the circuit. These histograms are very easy to use and they make it possible to examine the whole circuit with respect to the yield of each element in the design varying around its nominal value. These histograms are constructed from the stored data of the common Monte Carlo yield analysis (either from spice runs or evaluating a polynomial model). If the yield histogram is flat, it means that the design is insensitive to that component value and its variation. If the yield histogram is bell curved or mountain shaped, it means that the component is extremely sensitive and any small change from its nominal value results in a big drop in the yield. This is called a Red X component and has to be well-controlled in order to preserve a higher yield. If the yield histogram is sloping up, it means that the component nominal value needs to be set higher, and if it is sloping down, it means that the nominal value needs to be set lower.

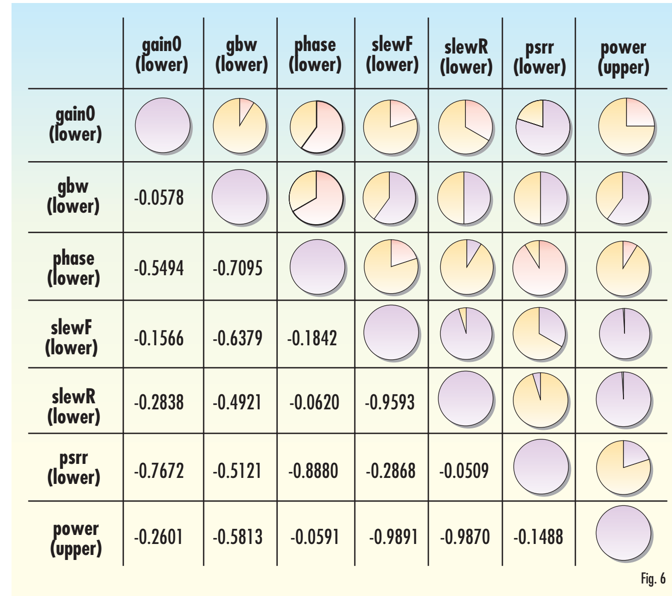
Note that the up and down sloping is usually taken care of and done automatically during the Yield Optimization (or) Design Centering stage. The usefulness of this tool is that it finds Red X components fast and easily. Designers then can take the appropriate steps to either control these red X components or eliminate them by choosing different topology.

Figure 5 shows the effects on an amplifier yield of four different elements: resistor Inp_R3 and capaci-



Yield sensitivity histograms

tors Inp_C1, Int_C1, and Out_C1. Figure 5(a) clearly indicates that the overall LNA yield would increase if the nominal value of resistor R3 were reduced. As mentioned above, this process, which moves the nominal values of the components right or left to increase the overall yield, can be automated using "design-centering" techniques. Figure 5(b) indicates that capacitor C1 in the input-matching network is not sensitive and its variation around its nominal value does not affect the yield. Figure 5(c) clearly indicates that the overall LNA yield will increase if the nominal value of C1 in the interstage network is increased. The histogram plotted in Figure 5(d) is especially interesting. It has a maximum value at the center, with the yield dropping to zero when the nominal value is moved either right or left. The plot shows that this component has a large effect on the yield of the design. These components are known as "Red X" components and they must not be allowed to vary. In a board-level design, this capacitor could be specified for a tighter tolerance. In IC designs, however, variation in a capacitor's value is determined by the semiconductor process. As a result, if variations in C1 are limiting IC yield, the matching network must be redesigned to accommodate the limitations of the process.



	gain0 (lower)	gbw (lower)	phase (lower)	slewF (lower)	slewR (lower)	psrr (lower)	power (upper)
gain0 (lower)							
gbw (lower)	-0.0578						
phase (lower)	-0.5494	-0.7095					
slewF (lower)	-0.1566	-0.6379	-0.1842				
slewR (lower)	-0.2838	-0.4921	-0.0620	-0.9593			
psrr (lower)	-0.7672	-0.5121	-0.8880	-0.2868	-0.0509		
power (upper)	-0.2601	-0.5813	-0.0591	-0.9891	-0.9870	-0.1488	

Fig. 6

Performance Correlations (provided by ChipMd)

Performance Correlation and Sensitivity Analysis

Additional analysis techniques using Monte Carlo simulations can produce correlation and sensitivity results. The computing of the performance correlation from a 300 sample Monte Carlo analysis are shown in Figure 6. The benefits of the correlation analysis include the determination strongly or weakly related and positively or negatively related performance criteria. It is easy to identify the very strong (.9593) positive correlation between the rising Slew Rate (SlewR) and the falling Slew Rate (SlewF). It can also be seen that there is a strong negative correlation (-.7095) between phase and bandwidth (gbw). From this fact the user will understand the impact of performance.

Figure 7 illustrates the viewing of performance (in this case gain) sensitivity to the input variation of the devices sizes. It can be seen that the width of transistor 5 (W5) has the largest influence on gain and that the influence is negative. In particular for every micron of width that W5 is increased, the gain will go down by .4112 db. Likewise the gain will go up by 1.193 db for every micron increase in L5 (channel length of transistor 5).

By comparing the effects of transistor sizing the user can make well-informed decisions about how to adjust transistor.

Correlation and sensitivity analyses are not limited to Monte Carlo simulations. Similar results are achieved by using the Response Surface Methods. Instead of performing a large number of Monte Carlo spice runs, the DOE/RSM approach creates mathematical expressions that represent the performance measurement variation with the variation to the inputs. Statistical analysis techniques are used to compute the output performance correlation and input sensitivities to the output performance. See Figure 9.

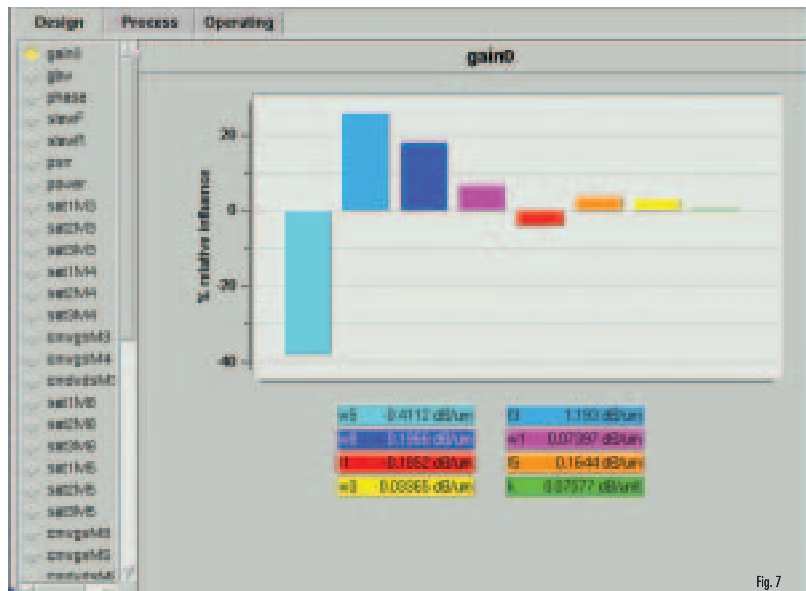
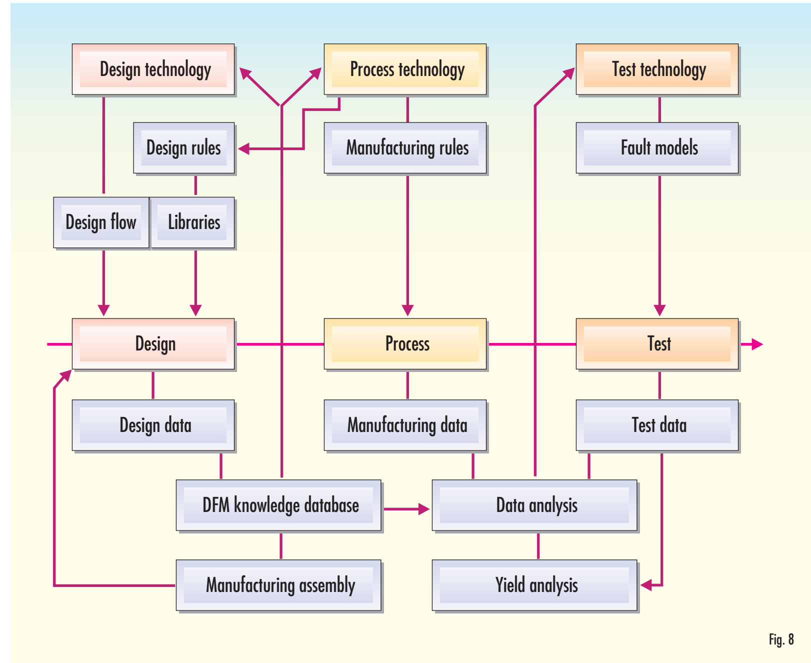


Fig. 7

Performance Sensitivities (provided by ChipMd)

Parametric Yield – Design flow

The flow diagram provided by Philips Semiconductors (Date 2000) illustrates the statistical design flow used to predict yield, Cp, Cpk and design sensitivity to process parameters. A key part of the flow is the information linkage between design, manufacturing and test. This has issue has been solved by Philips and the use of their process blocks (Medea T651). The parameter related part of DFM is concerned with the fit between product functionality and process variations, i.e. the fab′s total process window. Statistical design tools (process blocks), timing accuracy and designed-in reliability are the three key words in this part of DFM.



Philips Statistical Design flow

Parametric DFY Examples

The following examples are "real world" results of appling the techniques described earlier. Measured silicon is included to demonstrate the viability of the methods.

CMOS opAmp

The following 0.25 um CMOS opamp provides an in-depth analysis of statistical simulation and analysis at work. The CMOS opamp design is used on a data converter for wireless applications. LHS was used to create a polynomial model of the opAmp performance. Comparison between spice, measured data (5 wafers) and the polynomial model are provided.

The 3-D chart to the right concisely illustrates the sensitivity of the performance outputs to their inputs. Clustering of sensitivities can quickly be noticed. For example the DC measurements (VOS, CMRR, PSRR) are very sensitive to the mis-match between CMOS transistors. Knowing this information allows the designer to make correction in the mis-match (increase transistor size) and rerun the analysis to verify improvement.

Graphs of the key performance indicators are also indicating the measured distribution and the simulated distribution. It should be noted that simulated distributions would always provide a plus or minus six sigma spread.

Function	Spice (typical)	Measured mean	Mean model	Measure standard deviation	Standard development model
psrr+ (dB)	93.35	94.61	93.31	9.48	12.99
psrr- (dB)	92.92	86.71	91.99	9.17	17.37
ugbw (MHz)	4.62	3.69	4.41	0.11	0.97
pM (deg)	51.24	42.0	50.48	2.35	1.791
avol (dB)	108.4	—	108.1	—	1.046
sr+ (V/in)	4.94	1.72	4.918	0.046	0.27
sr- (V/us)	4.67	3.67	4.63	0.252	0.95
vos (mV)	-0.132	0.244	-0.095	0.792	3.82
cmrr (dB)	119.0	83.6	85.85	9.82	15.64

Table 1

X-band MMIC

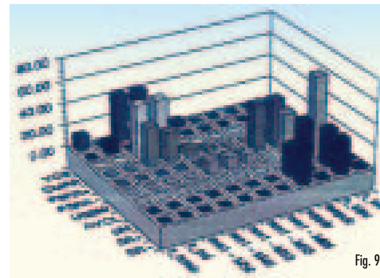
This is an example of two X-band MMIC amplifiers with the same specifications. One amp was designed using a standard design methodology and the other one used the six sigma design methodology. Both chips were placed adjacent to each other on a 50 reticles wafer and experienced the same manufacturing process variations during the fabrication process. On wafer testing, both chips were tested in the same test environment. Keep in mind that the only difference in the two chips is that they each have different topologies. The standard design method used an arbitrary standard topology that meets all the specifications at nominal conditions. The other design methodology used statistical analysis using various statistical tools and selected a robust and insensitive matching network topology.

There are few important things to observe in these measured output results. First, notice the manufacturing process was not well in control. The FET measured parame-

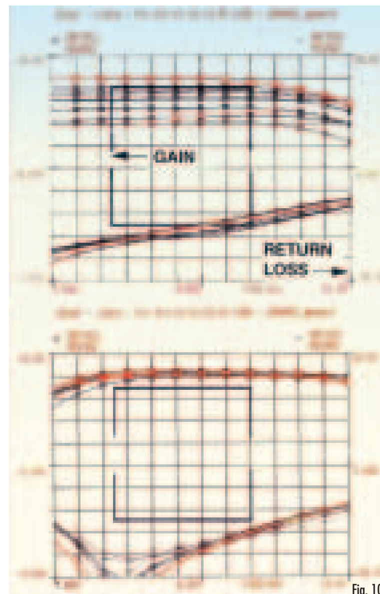
ters from this manufactured wafer were inserted back in the original design file. The corresponding re-simulated output is shown on the plots in Figure 10 overlaid with a red trace. The design original simulation was very well centered on the grid using the original device models, but in this run the process has shifted the response to the left (lower frequency) resulting in many failures in the return loss of the standard design. The same process shift was also applied to the six-sigma design but did not have as big of an affect as it did on the standard design. As a result, the six sigma design still ended up with a 100% pass. Next, regardless of the process shift, notice the amplifier gain results. In the standard design method, the gain variation from many sites of one wafer is very wide (more than 4 dB) due to extreme sensitivity of some elements in the design. In contrast, the six-sigma based design eliminated such sensitivity and resulted in a very tight gain variation. This type of gain variation relative to the amplifier specifications is what makes the capability index (Cp or Cpk) higher.

K-band Up-converter

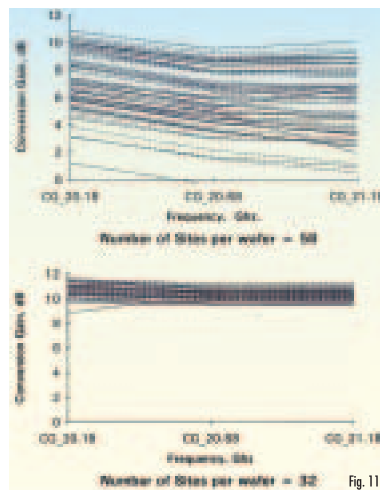
This is an example of two MMIC K-band Up-Converter Mixer Macrocells. The same X-band amplifiers discussed above in the X-band example were used along with Ku-band up-conversion mixers resulting in the K-band up-converted output. The six-sigma based amplifier was used with a six-sigma based mixer design, and the non six-sigma amplifier was used with a standard non six-sigma mixer design. The two Up-Converter Mixer Macrocells were processed together on the same wafer and they were measured under the same condition and environment. Figure 11 shows the results of the up-converted output K-band signal. The non six-sigma based design resulted with a 10 dB variation in its conversion gain, where as the six-sigma based design resulted with only a 2 dB tight variation.



Op-amp process, voltage, supply, matching sensitivities



Nominal and statistical analysis of X-band MMIC



Standard and statistical design results

K-band Up-converter

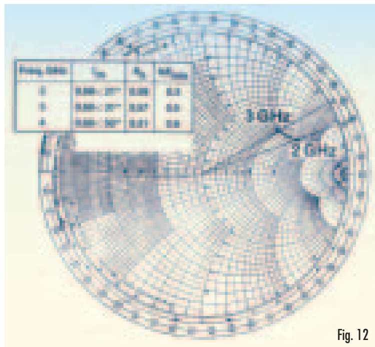
This is an example of two MMIC K-band Up-Converter Mixer Macrocells. The same X-band amplifiers discussed above in the X-band example were used along with Ku-band up-conversion mixers resulting in the K-band up-converted output. The six-sigma based amplifier was used with a six-sigma based mixer design, and the non six-sigma amplifier was used with a standard non six-sigma mixer design. The two Up-Converter Mixer Macrocells were processed together on the same wafer and they were measured under the same condition and environment. The Figure 11 shows the results of the up-converted output K-band signal. The non six-sigma based design resulted with a 10 dB variation in its conversion gain, where as the six-sigma based design resulted with only a 2 dB tight variation.

2.4 GHz LNA

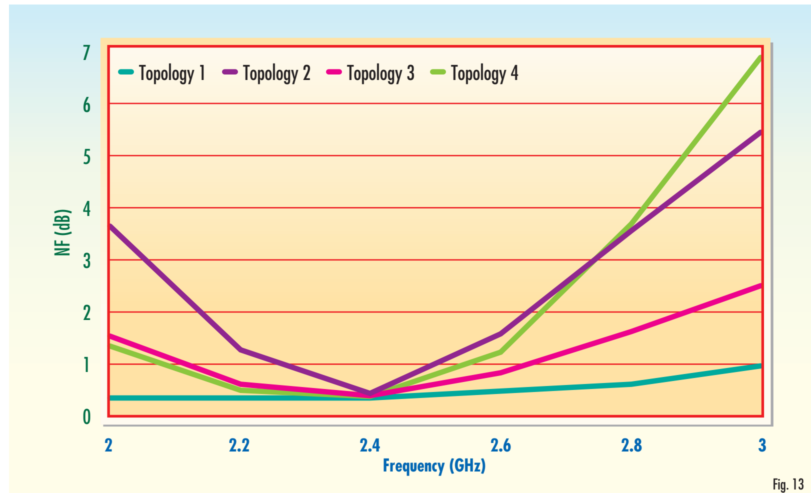
Figure 12 shows the locus of optimum load impedance between 2-3 GHz for optimum noise match. Using interpolation, we can find the optimum noise match point at 2.4 GHz, which lies on that curve. The designer has to come up with a matching network that would bring us to the 2.4 GHz optimum noise point in order to achieve the minimum .4 dB noise figure at 2.4 GHz.

For illustration purposes, four different matching networks were created. Topology 1 consists of just a single 120 ohm high impedance line. Topology 2 consists of a 50 ohm open shunt stub followed by a 104 ohm l/4 series line. Topology 3 uses a 50 ohm shorted shunt stub followed by a 104 ohm l/4 series line. Topology 4 uses a 50 ohm series line followed by a 50 ohm open shunt stub. Ideally and without any process variation, each of these four topologies would achieve the optimum .4 dB Noise figure at 2.4GHz because each of the four matching networks is designed to bring us to the optimum noise point at 2.4 GHz.

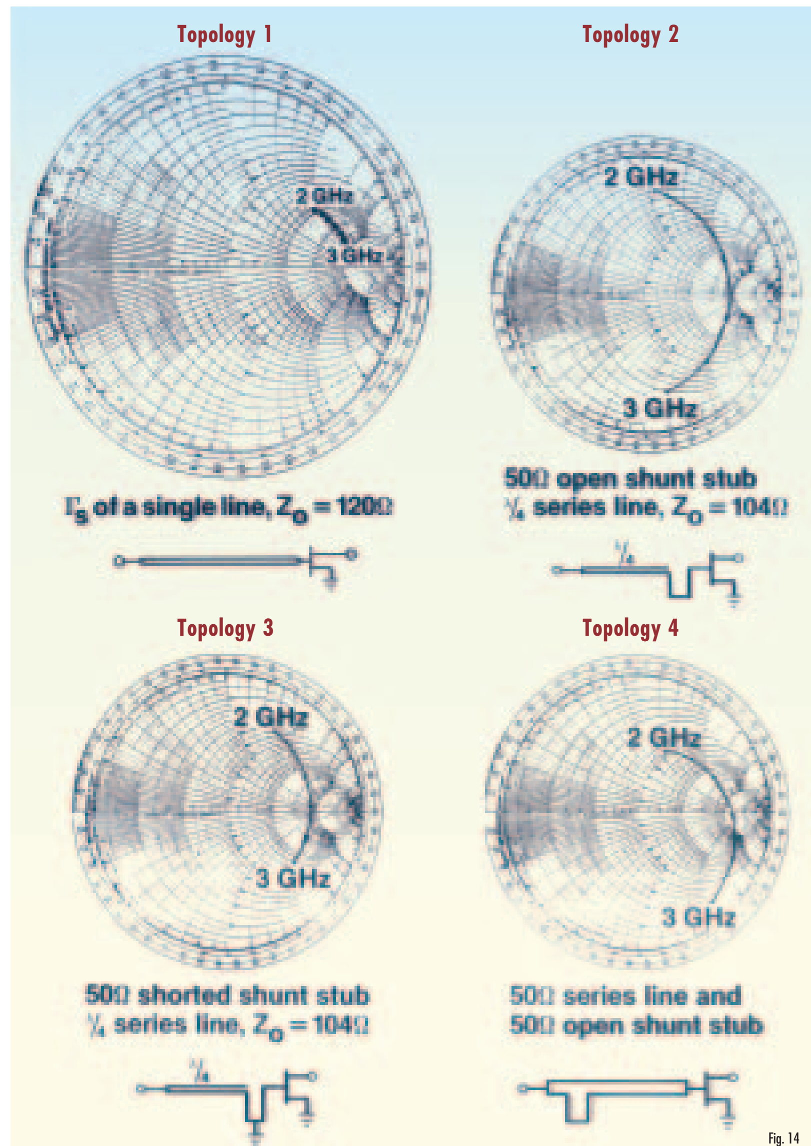
But in reality, process variation contributes variability in the topology and results in a deviation from the optimum NF point. Depending on which matching network we use, the resulting NF could be insensi-



NA optimum load



LNA matching network comparison



Topology configurations and performance variability

tive, slightly sensitive, or highly sensitive to the process variation.

One method to examine sensitivity is to simulate and plot the amplifier NF with each of the four different topologies as a function of frequency Figure 13 provides the graphs. Comparison of the topology sensitivity determines that the high impedance single-line matching network topology is the least sensitive (low Q) results. The other three topologies sensitivity increases as the NF moves away from the 2.4 GHz, due to any process variation.

To confirm the sensitivity results, each matching network topology is plotted on the Smith chart as a function of frequency from 2 to 3 GHz. Figure 14 provides the graphs. The spread of topology 1 is very tight, confirming the previous sensitivity. Its impedance points are very close to the optimum NF impedance of Figure 12 above. The other topologies constitute a much larger spread and go farther away from the optimum NF impedance of Figure 12. Topologies 2-4 resemble similar results from the sensitivity analysis graphed in Figure 13.

The conclusion of these examples is that in order to achieve robust and insensitive designs with high yield and minimum variation, statistical design tools such as Yield sensitivity histograms, Design of Experiments (DOE), and Yield Optimization or Design Centering are required. A complete description of each method can be found in EEtimes article "Statistical design can increase IC yield".

References

References

Date2000 http://jamaica.ee.pitt.edu/Archives/ProceedingArchives/Date/Date2000/papers/2000/date00/pdf-files/01d_1.pdf

Medea T651
http://medeaplus.net/medeanews/profiles/pp_t651.pdf

A User's Guide to LHS: Sandia's Latin Hypercube Sampling Software, Gregory D. Wyss and Kelly H. Jorgensen, Risk Assessment and Systems Modeling Department, Sandia National Laboratories, PO Box 5800, Albuquerque, NM 87185-0747, http://www.prod.sandia.gov/cgi-bin/techlib/access-control.pl/1998/980210.pdf

World Class Quality, by Keki R. Bhote

Company Recognition

Pivotal Enterprises provides marketing and business development consulting services to the DFY market place.

Agilent EEsof EDA is the leading supplier of Electronic Design Automation (EDA) software for high-frequency system, circuit, and modeling applications.

Motorola's Semiconductor Products Sector (SPS), headquartered in Austin, Texas, is the world's No. 1 producer of embedded processors. We create system-on-chip technologies for wireless communications, networking and automotive applications; and standard products that deliver embedded processing and connectivity capabilities for millions of devices throughout the world.

Philips Semiconductors, headquartered in Eindhoven, The Netherlands, employs over 32,000 employees in more than 50 countries. With sales of around $4.3 billion in 2002, Philips Semiconductors is one of the world's top semiconductor suppliers. We want to be the leading provider of semiconductor-based solutions for connected consumer and communications applications.

Why Interconnect and Lithography Modeling Impacts Yield

BY MARK RENCHER AND FRANK SCHELLENBERG

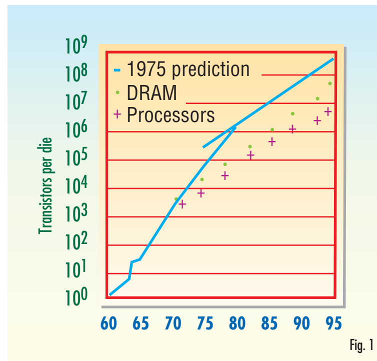
BY MARK RENCHER
President
Pivotal Enterprises,
Gilbert, Arizona
AND FRANK SCHELLENBERG
Mentor Graphics,
San Jose,
California

Moore's law and Failure Modes

In 1975, 10 years after Gordon Moore originally proposed his now very famous "Law" for the growth of transistor numbers in an IC, he revisited his "Law" and identified 3 drivers for the improvements he had observed since he originally saw the trend. These improvements were larger die size, technical advances (i.e. smaller line widths), and increasing "circuit and device cleverness". This "cleverness" we would now call design, and these techniques – for example, better isolation that allowed density to improve and the white space to be squeezed out of layouts - accounted for almost half of the trend. Moore said at the time that he thought 'cleverness' had reached it's maximum, and predicted a rolloff in the growth exponent of Moore's Law.

The rolloff happened even earlier than he predicted, as shown in Figure 1. And now, chip sizes are not growing anymore either, so the sole driver for Moore's Law going forward has become the technical achievements in fine patterning, mostly from lithography. Design has taken a back seat to the truly amazing improvements in process control that have allowed sub-wavelength features to be routinely fabricated, and features smaller than 100 nm to become a reality in 2003.

When ICs were all designed in vertically integrated firms, the ability for designers to be aware of process variations and eccentricities was, at least in theory, easier. Management could compel design groups and manufacturing groups to at least sit down at the table together, and both groups had a vested interest in the overall success of the company. With the evolution of the IC business to fabless design house and foundry, there is now a general agreement that process problems remain at the foundry, while designers will concern themselves with design issues.



Moore's Law roll-off due to lack of cleverness', predicted in 1975 (from Ref. 3).

This division of labor certainly allows each side to cultivate its expertise and strength, but it could not have come at a worse time. This is a time when many radical changes are occurring in fab processing. Aside from the new lithographic processes that are coming on line, there are new materials in the pipeline – low k dielectrics, oxide layers now only a few molecules thick, copper replacing aluminum for interconnects. What this means is that the intuition designers have cultivated over the years with more conventional silicon processing no longer applies.

Not only are the technologies for making the device elements changing, but their relative contribution to the overall IC performance is undergoing a sea change as well. We all knew that IC interconnect was eventually going to be the dominate contributor to signal performance. But what about yield? Consider this statistic: interconnect & white space contributed to 60 to 75% of the total available area on a 180 nanometer SOC. Perhaps there is room again for some more "cleverness" to squeeze designs. Again, the old rules don't apply.

The pareto chart in Figure 2 illustrates the contribution to yield loss from various effects for a typical CMOS process. All of these yield components together reduce the yield to roughly 70% - better than the 10% yield achieved in the initial stages of a process, but far from the 95% hoped for in mature manufacturing.

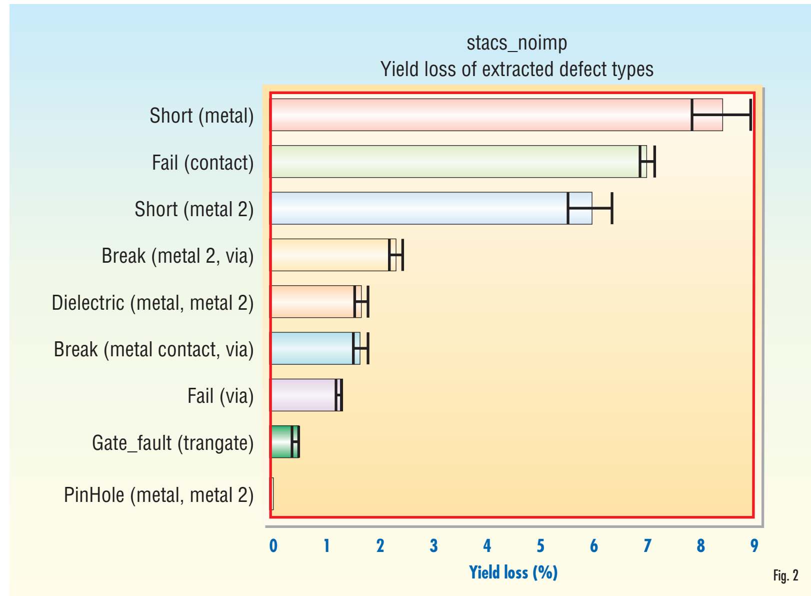
As the chart points out, the largest contributor to design yield loss is metal to metal shorted lines – 8% for Metal 1 and 6% additional loss for Metal 2. This can occur when normal linewidth variation for two nearby lines causes them to become too close together, allowing undesired bridging. It can also occur when a particle falls between two lines, causing undesired metal to remain. One remedy for this is to perform a Critical Area Analysis, analyzing the susceptibility to these modes of failure, and reduce this susceptibility to shorts by separating lines, adjusting the metal interconnect routing to locations that utilize the "white space" found on an SoC.

The second highest contributor to yield loss is the contact failure, at 7%. Contacts and vias have a simple task: as long as a good connection between layers is made, the contact or via does its job. . However, many subtle effects can lead to failure.

First, contact holes for new generations of ICs will be smaller than 100 nm themselves, and can be spaced just as close. If a contact, through normal statistical variation, pinches to be slightly narrower, the current density goes up. For some very small structures, the proper name for this is a fuse, and the contact can be destroyed. This can lead to failure of the chip. And there are a lot of contacts that can fail. A modern design may have 100 million contacts. Even if 6 sigma quality is achieved for this, there will still be 2400 random contact failures in this chip. Will a design be able to tolerate this number of failures?

Defects cannot be neglected, either. For vias, nanoscale voids in the metal (and especially true for copper) can migrate and aggregate, causing what was initially a functional connection to fail. Tolerance of via failure must therefore also be built into the interconnect layers as well, or the same statistical nightmare will emerge there..

It is easy for designers to simply assign blame and responsibility for yield to their foundry. Although easy, this is not entirely productive. Slightly more productive is a better understanding of the statistics of processing and process control on the part of designers. This can be facilitated if a foundry is willing to share statistical data on processes with their customers. Although sometimes awkward, this will become essential if accurate calibrations of parasitic extraction and timing are to be made. Foundries and designers can work towards an interactive model, whereby essential behavior models can be passed back and forth without revealing competitive process secrets.



Pareto of yield losses due to various effects for a 0.25um CMOS process. Courtesy: Predictions Software.

WHAT YOU SEE IS NOT WHAT YOU GET

The recent advances in lithography provide some guidance in this direction. When IC lithography was printing features larger than a wavelength, imaging distortions could be generally ignored. With sub-wavelength features, distortions dominate the imaging process. One example if what is called MEEF – mask error enhancement factor. As a process is operating near the failure mode, small changes in the mask layout can result in large changes on the corresponding wafer. Although predictable, this can indicate when control should be applied and when it can be relaxed. Optical proximity effects, caused by diffraction as lines move closer together, also alter the predicted line widths.

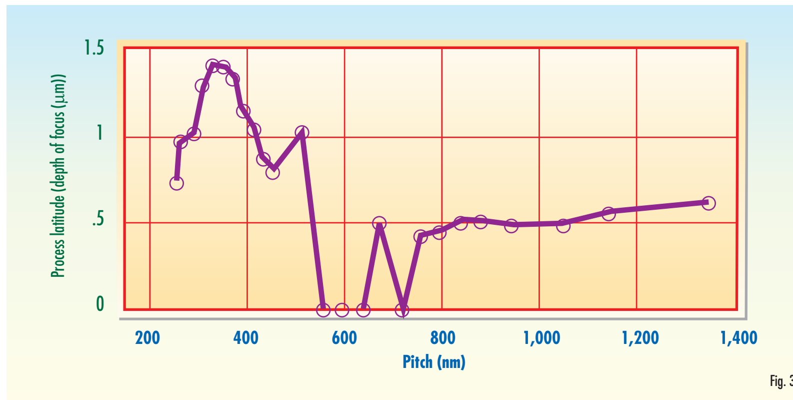
To correct for these, OPC (Optical and process correction) software alters the IC layouts for each critical layer of lithography, tailoring the pattern on the mask to the exact lithography conditions. To do this, advanced OPC programs have a process modeling and calibration routine that systematically measures the distortions, and produces a succinct set of parameters that can be generically called a process behavior model.

Unlike a rigorous TCAD model, in which exact material dimensions must be revealed and material compositions known, this OPC model functions more as an empirical transfer function, calibrated to a particular lithography procedure. When an IC layout needs to be corrected for a particular process, the OPC software refers to the behavior model as the layout alterations are made. However, since the confidential physical details of the process are not encoded, just the final behavior, the model can be shared by a foundry with its design customers.

This model-based OPC methodology has been partly responsible for the success of the 130 and 90 nm IC developments, and may be used for as many as 8 layers in a contemporary IC. Each layer requires a different process characterization, which exists as reference files to be called by the OPC software for each respective layer. As processes change, the reference files are updated, placing a premium on rapid model building and calibration.

Applying this approach to the broader issues of process variation for interconnects will not be trivial, however. First, the lithographic behavior can be modeled because CD SEM tools and mask inspection tools exist to actually measure, with statistical certainty (if you have the patience) the CD and variation produced by lithography tools. Although CD-SEMs may be able to estimate the roughness at the edges of nearby lines and vias, the 3D orientation of various features maters to the electrical properties of formed lines. to better estimate the parameters needed for extraction, only breaking a wafer and viewing in cross section allow dimensions such as via wall angle, or copper wire dimensions, to be determined. No matter how automated an SEM may be, it is impossible to break enough wafers and mount the fragments to get enough data for a statistically meaningful result. In this regard, new tools such as AFMs that generate quasi-cross sectional data, may be adaptable for this application.

OPC is one of a class of techniques called resolution enhancement techniques (RET) that have been applied to control the amplitude, phase, and direction of a wavefront emerging from the lithography mask. This control is vital in the world of sub-wavelength lithography. To implement these changes, the illumination system may be modified through a



'Forbidden Pitches' for 130 nm lines using Quasar illumination. 130 nm lines 130 nm apart will print well, while 130 nm lines ~500 nm apart will not print at all. Ref. 8.

number of techniques called generally called off-axis illumination (OAI). The phase can be controlled by providing structures on the mask that change the transmitted phase, either through material properties (generally implemented as attenuated phase shifting masks (attPSM) or by etching topography patterns in the mask (also called alternating phase shifting masks (altPSM)). Changes the polygon shapes for OPC control the amplitude of the transmitted light.

These can be implemented singly, but are generally implemented in combinations, and the proper selection of RET options is the subject of many conferences and books. Our goal here is to see if the adoption of RET can serve as a template for the future models and calibrations that may need to be applied more generally to the interconnect layers, let's take a closer look at some of the gremlins that have come home to stay in this sub-wavelength world.

OPC is used with a variety of RET solutions, notably special off-axis illumination systems that enhance the diffraction of certain spacings in the mask (at the expense of others). Odd phenomena have emerged as these combinations of RET are used. A good example is the phenomenon of a 'forbidden pitch'.

Although design rules assume things below certain minimum dimensions become impossible, they also inherently assume anything larger is possible. This was based on old assumptions of how illuminators and lenses for lithography work. With the new illumination schemes, the angle of the light works constructively with diffraction from some pitches, but destructively with other diffractive orders, and reduces their transmission through the imaging system. Hence the term "forbidden pitch".

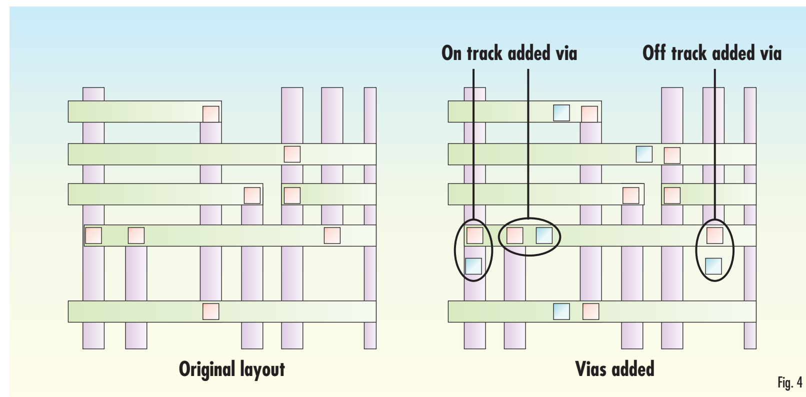
This is illustrated in Figure 3. Here, 130nm line-and-space pairs (a pitch of 260 nm) produce an excellent image using Quasar Illumination in an ASML Stepper. The same line in a more isolated environment, separated by ~500 nm from the nearest neighbor, produces no image at all. Only when the line is completely isolated (separated by greater than a micron) does the process latitude recover.

If an entire design contains nothing but dense or very large features (which may actually be the case for DRAM chips), this might work. But, as can be seen in the figure, a rule set that accurately describes which pitches are allowed and which are forbidden for various dimensions will be very complex. Although design rules may evolve to encompass a full description of these, only a model with good predictive power can identify them in the random orientations found in complex IC layouts. We would expect interactive phenomena from combinations of effects, such as mutual inductance between elements, will lead to surprises in interconnect modeling as well.

To increase the effectiveness of lithography, phase-shifting masks have also been employed. Phase masks allow interference of light to occur in the dark portions of the image, effectively boosting contrast. Because the dark fringe is formed by phase interference, the sensitivity to mask variations is very small, and the MEEF can effectively become 0. The additional control this provides makes phase masks very attractive for certain applications, such as the formation of thin gates.

However, this comes at a price. A phase mask requires several different writing steps, and there is no effective repair technology. Mask yields are generally low, so a mask shop must write several plates to get one that will meet specifications. This is all reflected in the higher price that mask shops must charge, and is part of the reason mask sets are beginning to approach $1M for a high performance ASIC. And, so it may be with interconnects, where exceptional control and predictability in line dimensions and edge roughness control may be achievable, but at a price that is barely affordable.

The most fundamental problem confronting any model-based process is that the metrology tools for process calibration simply do not exist. CD-SEMs can measure with a certain precision, which normally has an uncertainly of 3-4 nm. For a 90nm line, that's half the total process error budget. If you were doing well or poorly, how would you know? Oh, and by the way, NIST doesn't make a linewidth standard smaller than 1 um. There is therefore no reference for calibration. The ability to get precise and accurate meas-



An example of Redundant Via Placement. (Courtesy: Predictions Software)

urements for interconnect structures will be a problem, as well, and will add uncertainty to all the models that are generated.

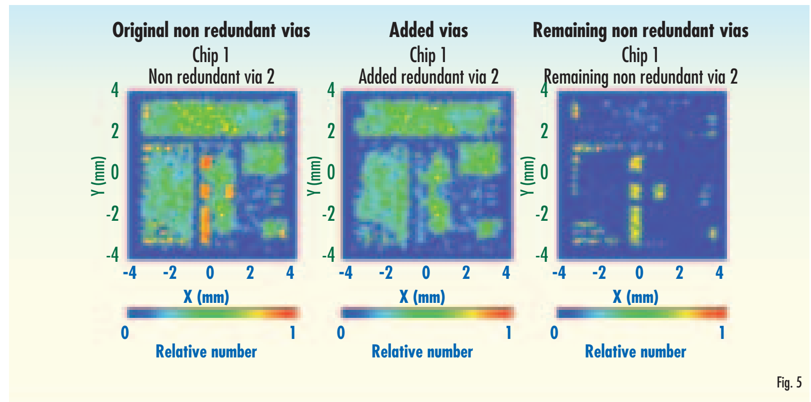
Is this Really a Design Issue? Real Problems – Real Solutions

The new world of RET for devices below 100 nm, with strange process behavior, costly masks, and fuzzy metrology, makes the negotiation between design and fab that much more complex. This is where the designer may be called upon to help out. If designs follow certain guidelines, many of these problems may become far more manageable for the fab.

For example, the contact holes that now appear in such large numbers, (and will statistically be bound to fail in much smaller but still significant numbers) can be made more manageable by the counterintuitive solution of adding more contacts. The technique, illustrated in Figure 4, involves replacing a single via with two vias whenever possible. These serve to back each other up, so that the statistical probability of having both fail is far far smaller than having a single contact fail. There is safety in numbers.

This technique can be implemented entirely in the design phase, and can even be done using conventional EDA tools, although specialized tools might make the process more efficient. Figure 5 shows how effective this can be in reducing the chance for error when systematically applied throughout a layout.

For metal lines, the counterintuitive solution of adding more features can also help overcome the problem of shorts. Here, however, lines are added not to the IC design, but to the pattern on the mask. The addition of sub-resolution assist features (SRAF) near isolated features places them in an optically dense environment.



Example of the reduction in susceptibility to via failure through via redundancy.

They will then interact with the light as if they were dense features, even though the sub-resolution features will not end up forming a printable image on the wafer.

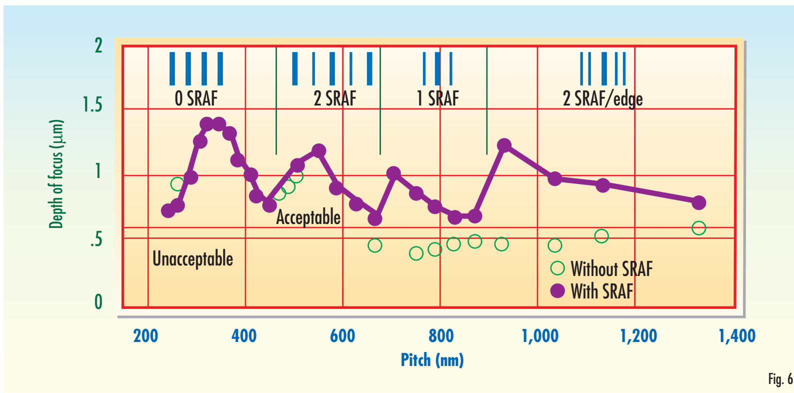
This gives a solution to the Forbidden pitch problem. Figure 6 shows an example of the application of SRAFs to eliminate the forbidden pitch problem of Figure 3, producing more uniform printing of features regardless of pitch. With the width of lines better under control, the susceptibility to shorts will drop, and yield will improve.

These solutions both alter the layout, but this really a designer's problem? Historically, the answer was "Yes" – if the polygons are being changed, there may be an impact on timing and performance, and design wanted to know about it. But features such as SRAFs never appear in the circuit, only on the mask. Should an IC designer be concerned about these alterations to the layout?

Lars Liebmann, of IBM, has explored in depth what is required to implement alternating phase shifting masks throughout the design flow, and believes this to be clearly a design issue. He concludes: "The practical implementation of altPSM design for logic requires a great deal of concerted effort in the areas of design rule definition, DRC tools, tools for the automatic generation of altPSM shapes, layout migration, layout synthesis, and hierarchical design construction. In addition, designer education and the development of flexible yet comprehensive altPSM design methodologies for varying circuit and design styles are critical. Without these two elements, no strategies for altPSM design, regardless of the complexity or sophistication of the design tools, will succeed."

The complexity of problems these cross-functional teams encounter is significant. Figure 7 shows a layout from a Metal 1 layer for an IC using 130nm technology at Motorola. Here, the RET recipe chosen was a blend of attenuated phase shifting for improved contrast, and model based OPC for image fidelity. This should enhance the process yield, and it typically does. However, as Kevin Lucas explains, "Sidelobe defects, which are spurious patterns caused by phase shifting, can form in complex circuit locations even when not observed in process test patterns. These sidelobes can lead to metal shorts, damage to underlying transistors or significantly increased capacitance. Full chip defect verification is therefore needed on product designs to ensure high yield. Fortunately, fast and accurate full chip sidelobe simulation and detection is now possible with leading OPC/EDA software."

Harry Levinsonof AMD, and author of several books on lithographic process control, echoes the need for vigilance and an intelligent control plan for the adoption of OPC. "The implementation of OPC definitely brings along a number of new process control issues. Physics drives the

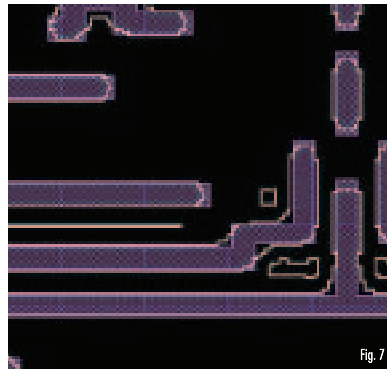


Solution of the Forbidden pitch problem by the application of SRAFs. Ref. 10.

need for OPC, and the physical phenomena that necessitate OPC represent new sources of process variation. For example, line widths can vary because we didn't get the OPC design right, or we may not be exactly on target on the mask simply because of the design grid or writing grid. Of course, fine detail in the OPC may be of no value, because the lens of the exposure tool filters mask details. With OPC, because the relationship among mask, design and reticle is no longer what-you-see-is-what-you get, accounting for variation on the wafer is more challenging. Real CD variation on the mask may be a good thing - because OPC could be reducing wafer linewidth variation though intentional mask CD variation. Distinguishing poor reticle linewidth control from inadequate OPC is not straightforward. The use of strong resolution enhancement techniques, such as off-axis illumination, enhances the sensitivity to illumination quality and lens aberrations."

Conclusions

The implementation of RET has allowed Moore's Law to continue to progress through improvements in lithography, but the additional problems that are introduced are not trivial. Automated EDA software to address these problems can make the task much more manageable, but these cannot be applied carelessly, and are an aid, but not a substitute, for qualified engineering.



Ideal layout (purple) and the wafer image predicted (white outline) for an RET mask. The image fidelity and contrast are improved, but unwanted sidelobes between metal lines can form, which can increase capacitance and lead to metal shorts. Courtesy: Motorola.

For the larger problems of yield and design, the model of the cross-functional team, armed with statistical data about the process from the fab and software tools with predictive power, will emerge as the dominant approach to yielding success. This has worked well for the adoption of RET, and the lessons learned from the integration of process models into verification products such as Mentor Graphics' Calibre can be expanded to incorporate more design for yield (DFY) and design for manufacturing (DFM) solutions as well.

Joe Sawicki, VP of D2Si at Mentor Graphics, sees opportunity in this situation. "With this new interest in DFM and DFY, we are seeing demands to coordinate different kinds of data that EDA traditionally never had to deal with. We saw the same thing when RET first came on the scene, but we adapted. Now these capabilities have become fully integrated into the heart of the Calibre product suite, and we expect the same to happen with DFM and DFY capabilities. The key is managing design and process data in a way that will get yielding products to market faster, and that does not require a forced reeducation of designers. The EDA companies that don't help this process will be left behind."

As the adoption of newly proposed lithography techniques appears to be delayed, this may be what is needed to bring some "cleverness" back to the table and allow the progression of Moore's Law to continue.

Company Recognition

Pivotal Enterprises provides marketing and business development consulting services to the DFY market place.

Mentor Graphics Corporation (Nasdaq: MENT) is a world leader in electronic hardware and software design solutions, providing products, consulting services and award-winning support for the world's most successful electronics and semiconductor companies. Established in 1981, the company reported revenues over the last 12 months of about $600 million and employs approximately 3,500 people worldwide.

References

1 Gordon Moore, "Cramming more components onto integrated circuits", Electronics Vol. 38(8), pp. 114-117, 1965.

2 Gordon Moore, "Progress in digital integrated electronics", Technical Digest of the 1975 International Electron Devices Meeting, pp. 11-13, (IEEE, New York, 1975)

3 Gordon Moore. "Lithography and the future of Moore's Law", Proc. SPIE 2440 pp. 2-17, 1995.

4 Laura Peters "Exploring Advanced Interconnect Reliability", Semiconductor International, 7/1/2002.

5 W Maurer, K Satoh, D Samuels, T Fischer. "Pattern Transfer at k1=0.5: Get 0.25_m lithography ready for manufacturing." Proc SPIE 2726:113-124, 1996.

6 For example, Optical Microlithography XVI, A. Yen, Editor. Proc. SPIE Vol. 5040, 2003 (Proceedings of the Spring SPIE Microlithography Conference).

7 A.K.K. Wong, Resolution Enhance-ment Techniques in Optical Lithography (SPIE Press Vol. TT47, Bellingham, WA, 2001)

8 Socha et al. "Forbidden Pitches for 130 nm lithography and below", Proc. SPIE Vol. 4000, pp. 1140-1155, 2000.

9 F.M. Schellenberg, O. Toublan, L. Capodieci, and B. Socha, "Adoption of OPC and the Impact on Design and Layout" in Proceedings of the 38th Design Automation Conference DAC 2001, pp. 89-92 (June 2001).

10 L. Liebmann, J. Lund, F.L. Heng, and I Graur, J. Microlith. Microfab. Microsyst. 1, pp. 31-42 (2002).

11 H.J. Levinson, Lithography Process Control, (SPIE Press Vol. TT28, Bellingham, WA, 1999); and Principles of Lithography (SPIE Press, Bellingham, WA, 2001).

12 D. Lammers and M. LaPedus, "Intel's litho move sparks a scramble", EETimes, Issue 1272, p. 1&110 (June 2, 2003).

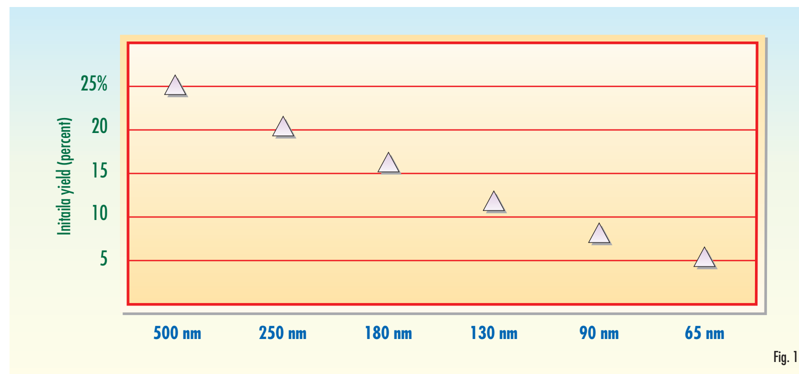
GDSII yield signoff methods

BY MARK RENCHER AND GERARD ALLAN

By Mark Rencher
President
Pivotal Enterprises,
Gilbert, Arizona
and Gerard Allan
President
Predictions Software

Physical defects have always played a role in IC yields. However, in today's nanometer technologies design sensitivity to these physical elements has continued to increase. The result, a continued reduction in initial product yields, increased yield learning times and increasing significance of the physical effects on yield.

Aside from the traditional particle defect, which will be discussed in this article, lithography and process variations all contribute to poor yields. Due to the complexities of these defects, the faulty behaviors frequently present themselves as transient in terms of current, well-known logic stuck-at, timing delay or IDDQ fault models. This Design for Yield article discusses the methods of predicting and reducing the impact of physical yield defects.



Historical initial yield.

Silicon Defects

To accurately predict the contribution to particle defect yield as it relates to the design, classification of defects types is required. Defects need to be classified according to their electrical fault types. That is shorts, breaks, pinholes, dielectric failures etc., so that they can be matched to the critical area measurements and subsequent tests. All defects that cause the same kind of electrical fault in a layer can be grouped together or if they they can be separately classified (e.g., dome shorts, pancake shorts) counted in a sub-group. This way the yield loss associated with a defect classification can be identified.

Electrical Fault Types

The classification of particle defects is required to associate the electrical faults types with the defect particle types. To make the best use of yield enhancement efforts, it is important to know what the yield loss associated with each particle defect type is, not just the defect count. For example, the yield loss from shorts might be associated with dome and pancake shaped particles.

The electrical fault types created by particle defects include:

Shorts

Defect material that creates an electrical connection between 2 nodes. The figure below shows an interconnect short, which causes an electrical fault.

Contact Failure

A contact failure is due to a particle defect creating an electrical break. Typically the failure is due to the contact coverage is reduced, which causes a higher resistance and reliability failures. Other contact failures may not be due to particles, but may instead be process related. These failures can cause the same fault type as a particle fail.

Breaks

An interconnect break is shown in the following figure. Although obvious, the classification is real.

Pin Hole

Pinhole faults are caused by defects in the dielectric separating two conducting layers. These defects are very small and are usually modeled as having zero radius. The regions of a layout susceptible to these defects are those where two conductors that belong to separate electrical nodes overlap. This includes transistor gate faults caused by pin holes. The figure below shows the pinhole critical area for a small layout.

Other non zero sized dielectric faults can also be modeled in a similar way to extra material critical area.

Dielectric Fault

Particulate defects can also be present in the dielectric between two layers. These differ from pinhole defects in that they are not small enough to be modeled as having zero radius. The source of these defects is different to the source of pinhole type defects. The critical area is associated with a dielectric shorts between for example metal1 and metal2. That is only inter level shorts between metal1 and metal2 are considered and not intra level shorts

between metal1 and metal1 or between metal2 and metal2.

Where the defect type does not create an electrical fault, it is considered cosmetic blemish.

Classification of electrical faults can be broken down into defect categories. For example, there may be two defect types that cause shorts (classified as dome and pancake defects) and two defect types that cause contact/via fails (resist bubbles and straight forward "defects"). The following figure illustrates how faults are classified in a pareto chart.

To predict the electrical failures on a physical design and increase design yields, physical DFY techniques are applied as the next step.

Physical DFY Techniques

A common physical DFY approach is Critical Area Analysis (CAA) which identifies the fault regions on an IC. The reduction of the fault regions includes methods such as wire spreading and/or contact/via redundancy.

Critical Area Analysis

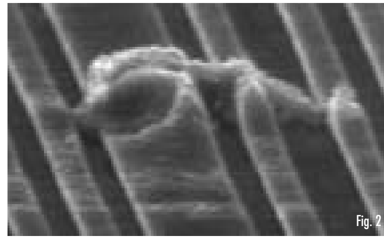
The critical area of a design layout is a measure of its susceptibility to defects. It is important to recognize that not all electrically active defects cause faults.

Critical area defines the region of the layout where a defect must fall in order to cause an electrical fault.

Traditionally these defects where viewed as random and yield loss was divided into systematic and random yield. However, critical area techniques can be used to measure the impact of both random and layout related systematic yield loss. For example end of line vias may suffer from yield loss as a result of end of line metal draw-back. Where this an important yield loss mechanism such vias should be identified and included in the “critical areas”.

Critical Area Analysis Types

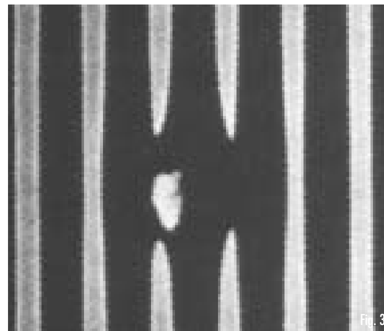
As there are a number of different defect types there are also a number of different types of critical area. For example the critical area for transistor gate pinhole defects can be easily identified using Boolean operations (active AND poly). Other critical area such as extra and missing material critical area are extracted using specialized algorithms.



Particle short.

Extra Material Critical Area

The critical area for extra material defects is defined as the region in which the center of a circular defect of a given size must fall in order to produce a fault. The figure below shows the extra material critical area for the defect size shown. The area of the polygons that define such a region is a measure of the probability of the fault occurring in the circuit. In order to characterize the layout fully this operation must be repeated for a range of defect sizes.



Material break.

Missing Material Critical Area

In exactly the same way the critical area for missing material defects is defined as the region in which the center of a circular defect of a given size must fall in order to produce a fault. However, here the critical area region covers the tracks. The figure below shows the critical area for a simple layout. Again the critical area must be generated for a range of defect sizes.

Other measurements as critical area

Other measures of critical area may also be important, depending on the defects present within the process. For example, if via yield is low, a count of the number of vias, or perhaps more importantly the number of non redundant vias in the design becomes a “critical area” of the design too.

Extracting critical area from Layout

Extracting and plotting the critical area of a design can be helpful when attempting to optimize a design for yield. This is particularly useful for library cells and SRAM designs. For single cells or small groups of cells it is possible to extract the critical area interactively in a layout editor.

Critical area analysis is computationally intensive, one successful method of reducing the computational effort is to use sampling. Yield prediction by sampling is based on the statistics of survey sampling [1]. Properties of a population can be estimated with bounds on the error of estimation by taking a number of random samples from the population.

For example, the estimate of the whole population mean, is given by the sample average (1)

$$\hat{\mu} = \bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$$

The estimated variance of (2)

$$\hat{V}(\bar{y}) = \frac{s^2}{n}\,\frac{N-n}{N}$$

where N is the population size, n is the number of samples and s is the sample variance,

$$s^2 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n-1} \quad (3) \Rightarrow$$

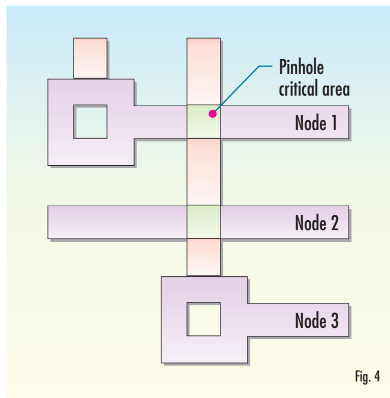
$$s^2 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n}$$

The bound on the error of estimation:

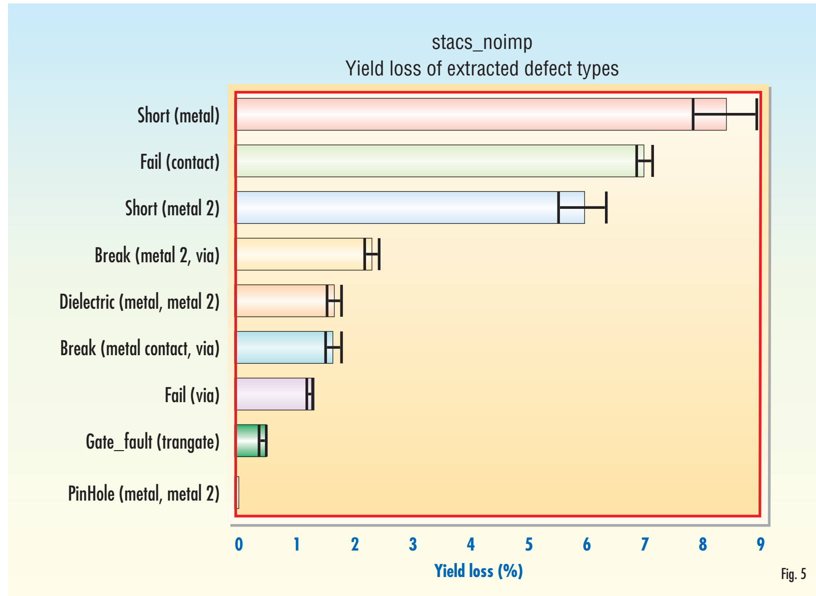
$$2\sqrt{\hat{V}(\bar{y})} = 2\sqrt{\frac{s^2}{n}\left(\frac{N-n}{N}\right)} \quad (4) \Rightarrow$$

$$2\sqrt{\hat{V}(\bar{y})} = 2\sqrt{\frac{s^2}{n-1}\left(\frac{N-n}{N}\right)}$$

An IC can be viewed as a population of regions each with varying susceptibility to a range of defect types. The electrical fault sensitivity for the chip as a whole can be estimated by randomly sampling the layout [2]. Survey sampling is particularly suited to IC yield prediction since, for large populations (large in comparison to



Pin hole illustration.



Yield loss by electrical type.

the number of samples), the error bound on estimatesdoes not depend on the population size but on the variance of the population. This implies that the number of samples required to characterize very large chips does not significantly change with chip area or even the chips complexity but only with the variation of the fault sensitivity over the chip. It is this attribute that enables sampling to be used for even the largest layouts, because the variation of layout fault sensitivity within a chip is in general not related to its area.

The sampling method also has the advantage of obtaining not a single value for critical area or yield but a whole range of values that are generated from across the whole chip. Plotting these values gives a map of the chip critical area and shows where within the chip the design is most susceptible to faults. The figure below shows a map of combined critical area for a chip converted into yield values. The red areas with the lowest yield are SRAM, with other areas of the chip consisting of mostly standard cells and routing.

Yield Synthesis

While the measurement of critical area can be used simply to predict the yield of a chip, it can also be used to drive design for yield efforts. Reducing the critical area will increase the circuit yield and may also help with reliability.

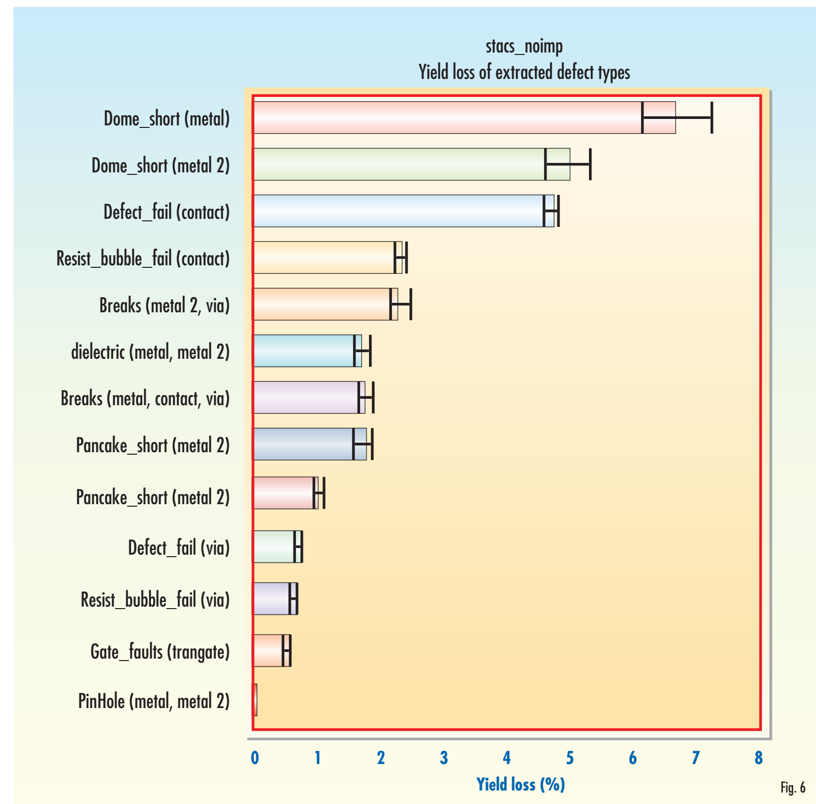
Contact and Via Redundancy

One way to increase the yield of contacts and vias is to use more than one where existing space with the layout makes this possible.

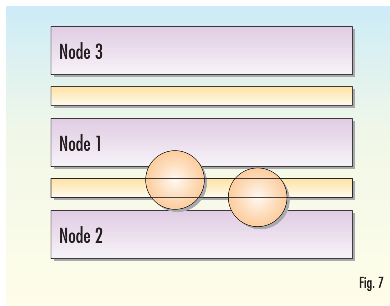
Typically within routing there are many opportunities to add extra vias. To reduce the extra metal that will also be required vias should be added preferentially on existing metal lines, but they can also be added where this is not possible in empty space beside the existing via. In typical routing it would not be uncommon for more than 50% of existing vias to be partnered in this way.

Wire Spreading

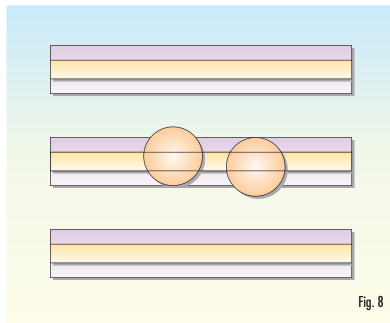
One of the major sources of yield loss is shorts between wires. This failure mode is made more likely by the habit of designers and design tools of running interconnect wires at or near the minimum metal pitch. This may make sense at the time,



Yield loss by a finer classification of fault type.



Extra material critical area.



Missing material critical area.

conserving valuable space for more wires later. But, this can leave the finished design with tightly packed wiring next to empty space. Techniques for spreading these wires into the empty space can result in significant reductions in extra material critical area and increases in yield.

Techniques for Wire Spreading

A simple technique that involves minimal changes to the design is to search the layout for interconnect wires that are next to some empty or white space.

Two layers can then be "added" to define a region to be removed and a corresponding region to be added. These two regions define a small displacement of the wire, reducing the critical area between the interconnect.

Figure 13 illustrates the results before using DFY. The items of interest are labeled for comparison with the results after using DFY.

Area A has a high concentration of vias which left unattended have a higher probability of creating a fault. This is a typical example of channel routing techniques of having different metal layers associated with horizontal and vertical traces. The result is a significant number of vias. Area B highlights is where there is available "white space" for wire spreading. Figure 14 illustrates the physical design impact after using physical DFY techniques.

Area A with its high concentration of vias has been adapted by via doubling without any increase in area or electrical behavior. Wire spreading also occurred on the design with a movement of the interconnect. Area B illustrates that even with extra vias present the wire spreading can be accomplished.

The best practices sequence for via redundancy and wire spreading is to first move or spread the wires and then add the additional vias.

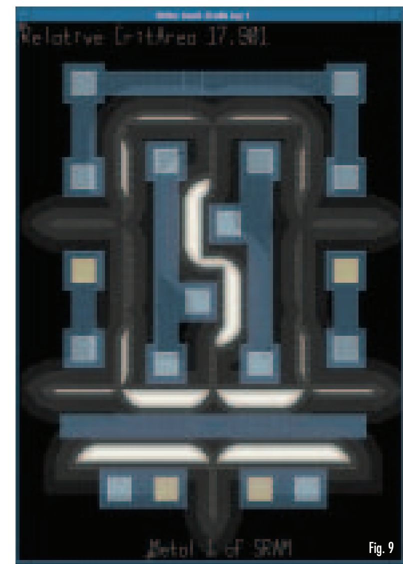
The overall impact of reducing the critical on the yield is shown in Figure 15. Here we have plotted the value of the product A(r)D(r) as function of the particle diameter r, for the critical area A(r) of metal layer 4 of the product, and a typical defect size distribution D(r).

Financial Impact

Predicting IC yield with the associated particle defects is important to assess the real cost of defects and help target defect reduction efforts where they will be most cost effective.

This allows the ability to apply a cost amount to the yield loss of a particular defect type. A yield prediction will more accurately indication of the scale of the problem and can be used to estimate ROI from defect reduction measures targeted at those defects. Figure 16 provides the critical area model (Courtesy of Philips Semiconductors). The model inputs are the D0 and the die size. The output is the %yield improvement. D0 is the defect density rating determined by the foundry. A complete description is found in Appendix A.

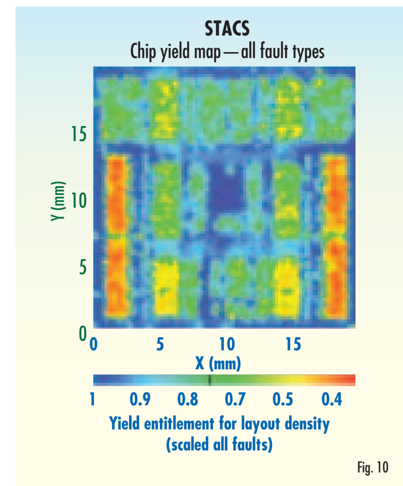
Report by Philips Semiconductors confirms that Physical DFY techniques increase yields between 3 and 6 percent.



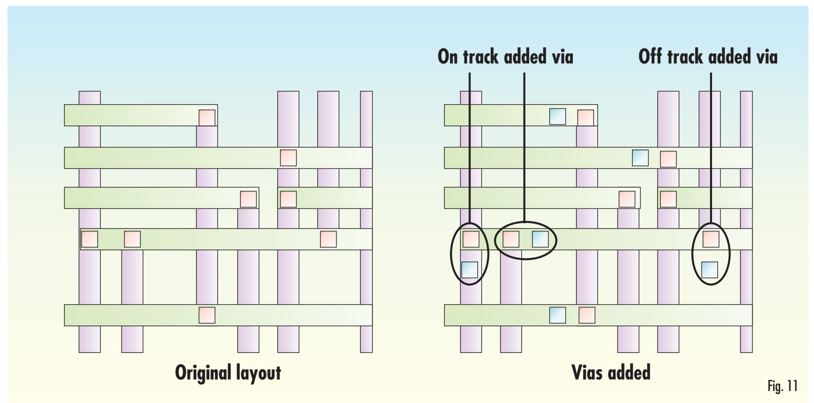
Cell critical area.

Table 1 confirms the yield improvement for 2 lots.

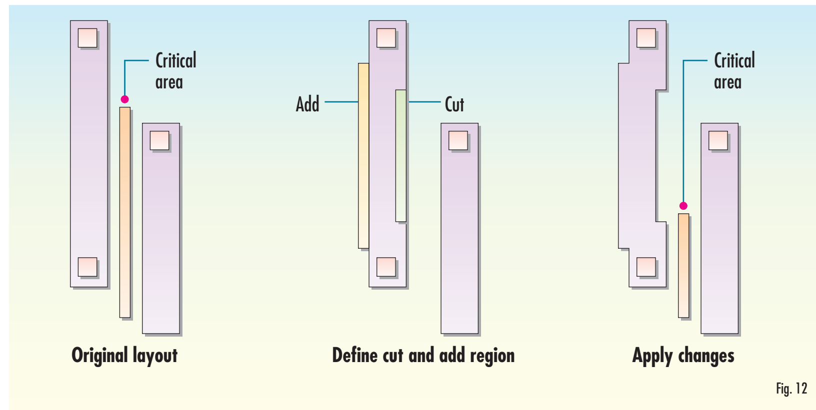
Referring to article 1 in the DFY series, a production example was cited. Using



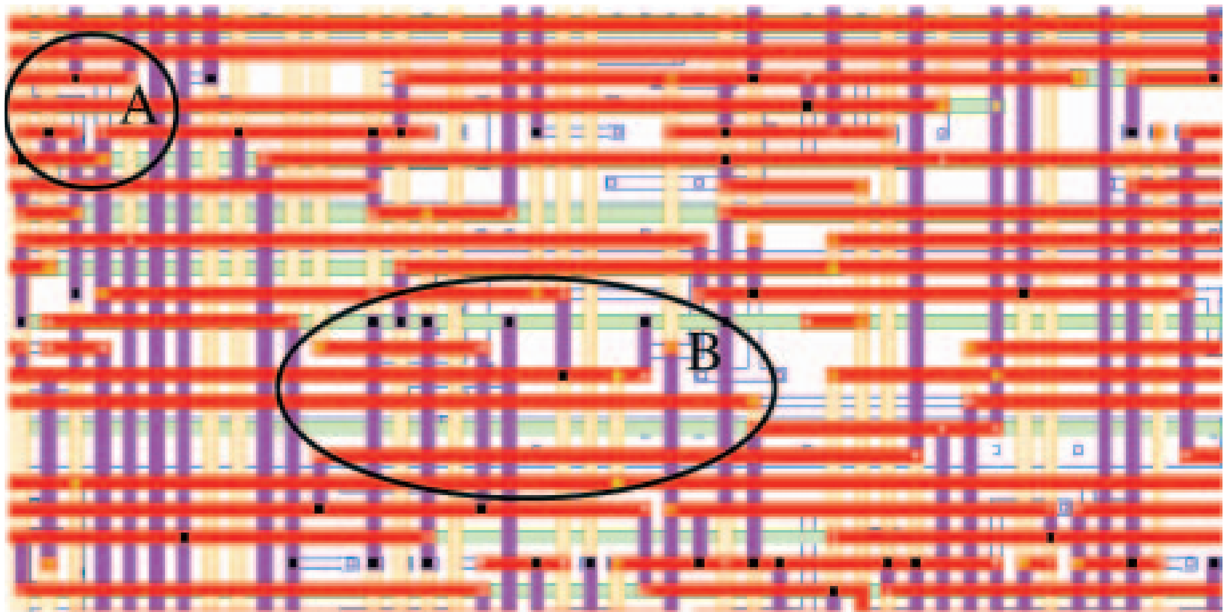
Yield map.



Via/contact redundancy

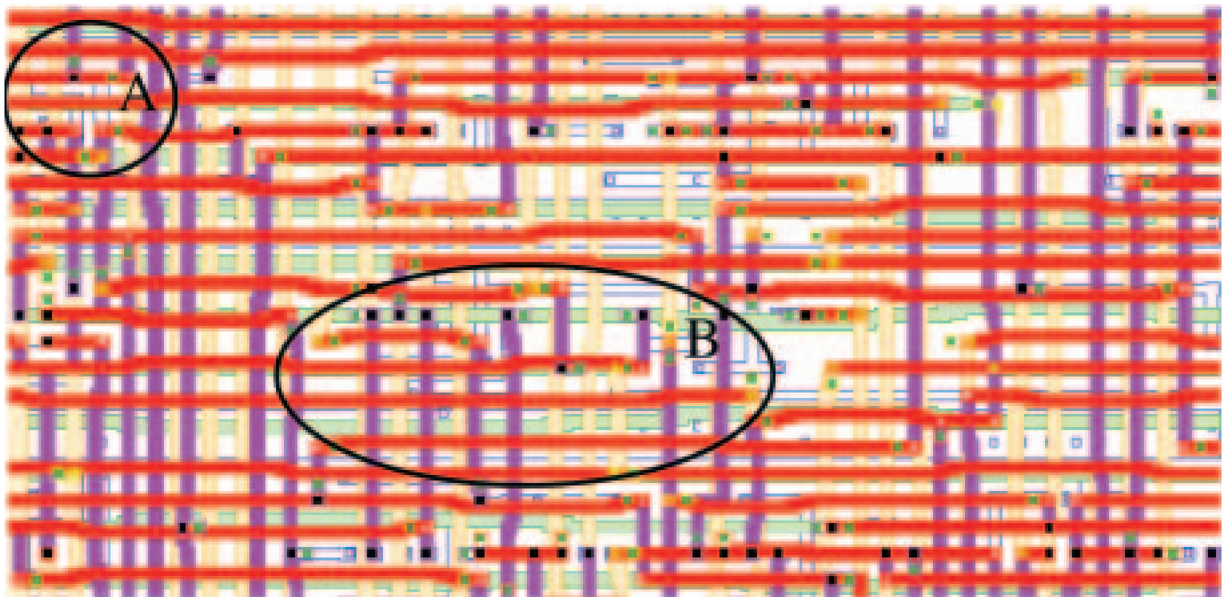


Wire spreading techniques.



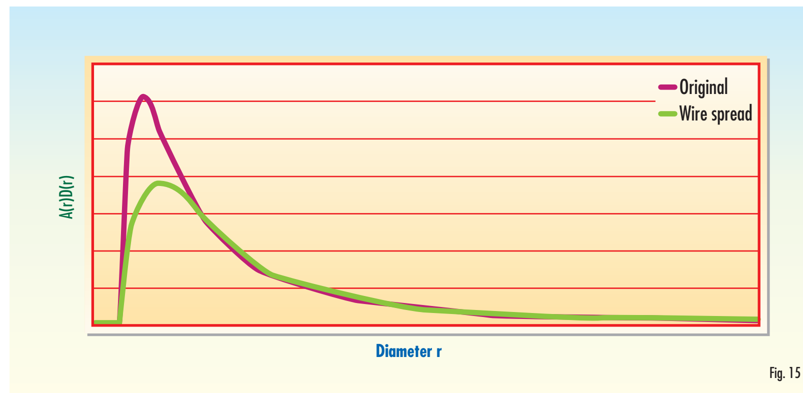
Before applying DFY techniques.

Fig. 13

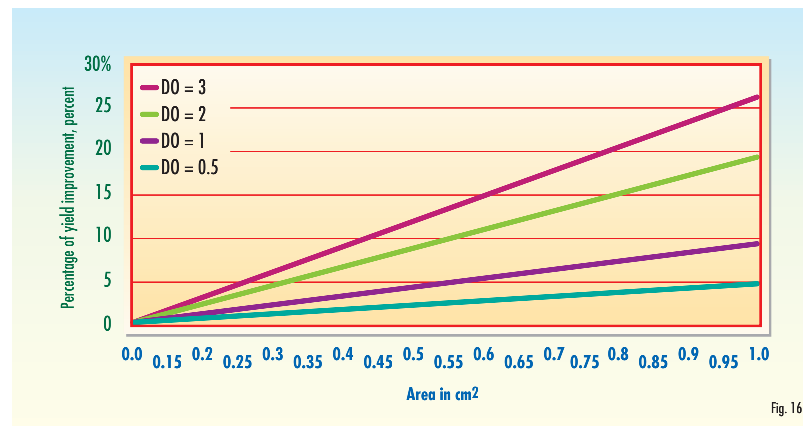


After applying DFY techniques.

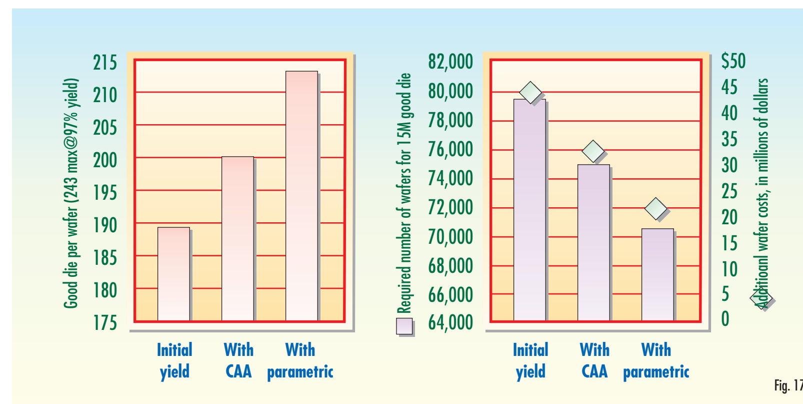
Fig. 14



Critical area reduction using wire spreading.



Critical area yield improvement model.



Yield improvements using critical area.

LOT_ID	Original yield	Wire-spread yield	Difference	Relative
cz6053	58.00%	61.00%	3.00%	5.17%
cz8279	66.40%	68.70%	2.30%	3.46%
Both	61.60%	64.30%	2.70%	4.38%

Table 1

Summary of yield values from wire spreading.

LOT_ID	Original	Wire-spread	Total
cz5659	3	3	6
cz5660	19	19	38
cz6380	7	7	14
cz8694	12	12	24
cz6895	2	2	4
cz6981	14	14	28
Total	57	57	114

Table 2

Number of wafers per lot.

the same information, the actual financial impact is calculated by the following:

- A new product is introduced into a mature .18 micron CMOS foundry.
- The committed product run rate to the customer is three years with a volume of 15 million parts.
- Initial wafer probe yield at 78%. (97% is best in class)
- 243 good die per wafer @ 97% yield
- $2500 wafer cost

Because of the 4.38% relative yield improvement, there were 2109 less wafers required to reach the 15 million production quota. The 4.38% yield improvement directly translates to a $5,272,500 savings or ($2500 wafer cost * 2109 saved wafers). Figure 17 illustrates adjustments from article 1 original data.

Physical DFY Example

Yield improvements on a channel routed design are presented below. The DFY techniques applied were wire spreading and contact/via redundancy.

The design, is the core of the 16 bit REAL processor with a boundary scan chain connecting all data inputs, data outputs, all dangling memory data inputs, memory data outputs, and memory address outputs. Its area is 0.76cm.

To verify the yield improvements in the design, 6 lots were run. The number of wafers for each lot is enumerated in Table 2. The number of die for each wafer was 206, for a total number of parts evaluated to be 23484.

The measured yield results from the 23484 die are computed in Table 3.

Table 3. Summary of yield data for the RDMR experiment.

LOT_ID	Original yield	Wire-spread yield	Difference	Relative
cz5659	44.00%	44.90%	0.90%	2.05%
cz5660	75.60%	79.60%	4.00%	5.29%
cz6380	20.90%	29.40%	8.5%	40.67%
cz8694	71.70%	70.80%	-0.30%	-0.42%
cz6895	71.70%	82.30%	11.9%	15.75%
cz6081	65.30%	63.70%	-1.60%	-2.45%
All	63.20%	65.90%	2.70%	4.72%

Table 3

Summary of yield data for the RDMR experiment.

Conclusions

Integrated circuit yield can be predicted very accurately using critical area techniques, To achieve these high levels of accuracy detailed knowledge of the process defect types and levels is required. In some cases information on layout related systematic faults will also be required, The steps required to build these accurate yield models enable not only a yield prediction to be made but can also help focus effort towards improving yield, Traditionally this focus has been on the process itself using the information on yield loss of specific defect types to concentrate on the most costly defects. However, the yield of a chip is also a function of the chip layout. This layout can be modified to improve the yield of the chip. Layout related systematic yield loss should be addressed but also the use of DFY techniques to reduce the sensitivity of the layout to defects. Real world results show that DFY techniques can increase yield from 2 to 25% with significant cost savings.

References

[1] Mendenhall, L. Ott, and R. L. Scheaffer. “Elementary Survey Sampling,” Wadsworth PublishingCompany Inc, Belmont, California, 1991.

[2] G. A. Allan. “Yield prediction by sampling IC layout,” IEEE Trans Comp. Aided Design, vol 19(3), Mar 2000, pages 359-371.

[3] D Maynard et al., "Wafer Line Productivity Optimization in a Multi-Technology Multi-Part-Number Fabricator," in Proceedings of the 9th Annual Advanced Semiconductor Manufacturing Conference and Workshop (Piscataway, NJ: Institute of Electrical and Electronics Engineers, 1998), 34--42.

Company References

Pivotal Enterprises provides marketing and business development consulting services to the DFY market place.

Predictions Software is a new startup. The company supplies software tools to predict semiconductor yield from design layout and tools to enhance IC layout to make it more manufacturable. Our software has been used for over 5 years by major European and US fabs. Find out more at www.icyield.com.

Philips’ semiconductor division, headquartered in Eindhoven, The Netherlands, employs over 32,000 employees in more than 50 countries. With sales of around $4.3 billion in 2002, Philips Semiconductors is one of the world's top semiconductor suppliers. Philips wants to be the leading provider of semiconductor-based solutions for connected consumer and communications applications. Find more at www.semiconductors.philips.com.

Appendix A –Defect Density Description

The number of defects observed in an area of size A units is often assumed to have a Poisson distribution with parameter A x D, where D is the actual process defect density (D is defects per unit area). In other words:

$$P(\#\text{Defects} = n) = \frac{(AD)^n}{n!}e^{-AD}$$

We assume that AD is large enough so that the normal approximation to the Poisson applies (in other words, $AD > 10$ for a reasonable approximation and $AD > 20$ for a good one). That translates to

$$P\{\#\text{Defects} < n\} = \Phi\left(\frac{n - AD}{\sqrt{AD}}\right)$$

where Φ is the standard normal distribution function. If, for a sample of area A with a defect density target of D_0, a defect count of C is observed, then the test statistic

$$Z = \frac{C - AD_0}{\sqrt{AD_0}}$$

can be used exactly as shown in the discussion of the test statistic for fraction defectives in the preceding section.

For example, after choosing a sample size of area A (see below for sample size calculation) we can reject that the process defect density is less than or equal to the target D_0 if the number of defects C in the sample is greater than C_A, where

$$C_A = \sqrt{AD_0}Z_\alpha + AD_0$$

and $Z\alpha$ is the upper 100x(1-α) percentile of the standard normal distribution. The test significance level is 100x(1-α). For a 90% significance level use $Z\alpha = 1.282$ and for a 95% test use $Z\alpha = 1.645$. α is the maximum risk that an acceptable process with a defect density at least as low as D_0 "fails" the test. In order to determine a suitable area A to examine for defects, you first need to choose an unacceptable defect density level. Call this unacceptable defect density $D_1 = kD_0$, where $k > 1$.

We want to have a probability of less than or equal to β is of "passing" the test (and not rejecting the hypothesis that the true level is D_0 or better) when, in fact, the true defect level is D_1 or worse. Typically β will be .2, .1 or .05. Then we need to count defects in a sample size of area A, where A is equal to:

$$A = \frac{k}{D_0}\left(\frac{\frac{Z_\alpha}{\sqrt{k}} - Z_{1-\beta}}{k-1}\right)^2$$

Suppose the target is $D_0 = 4$ defects per wafer and we want to verify a new process meets that target. We choose $\alpha = .1$ to be the chance of failing the test if the new process is as good as D_0 (α = the Type I error probability or the "producer's risk") and we choose $\beta = .1$ for the chance of passing the test if the new process is as bad as 6 defects per wafer (β = the Type II error probability or the "consumer's risk"). That means $Z\alpha = 1.282$ and $Z1\text{-}\alpha = -1.282$.

The sample size needed is A wafers, where:

$$A = \frac{1.5}{4}\left\{\frac{\frac{1.282}{\sqrt{1.5}} - (-1.282)}{1.5 - 1}\right\}^2$$

which we round up to 9. The test criteria is to "accept" that the new process meets target unless the number of defects in the sample of 9 wafers exceeds

$$C_A = \sqrt{AD_0}Z_\alpha + AD_0 = \sqrt{36}\times 1.282 + 36 = 43.7.$$

In other words, the reject criteria for the test of the new process is 44 or more defects in the sample of 9 wafers.