Holistic Yield Improvement[tm]: A Comprehensive Methodology for Accelerated Yield and Performance Ramping

Kimon W. Michaels, Xiaolei Li, Mark F. Antonelli, Andrzej J. Strojwas
PDF Solutions, Inc.
San Jose, CA 95110

Biography

Kimon Michaels has received BS, MS, and PhD degrees from Carnegie Mellon Univ. He is a Vice President and co-founder of PDF Solutions where his responsibilities include statistical SPICE model characterization and algorithms for parametric yield analysis.

Xiaolei Li received a BS degree from Fudan Univ. (Shanghai, China), and MS and PhD degrees from Carnegie Mellon Univ. He is currently a Senior Consultant at PDF Solutions. His research interests include 3D photolithography simulation and application of simulation methods to yield improvement efforts.

Mark Antonelli has a BS from Cornell Univ. and an MBA from INSEAD (Fontainebleau, France). His experience includes strategy and business consulting for multinational firms and software product development. He is currently Director of Marketing at PDF Solutions.

Andrzej Strojwas has a PhD from Carnegie Mellon Univ. He has held positions at Harris Semiconductor, AT&T Bell Labs, Texas Instruments, NEC, Hitachi, and SEMATECH. He is a professor of electrical and computer engineering at Carnegie Mellon Univ. and serves as Chief Technologist at PDF Solutions.

Abstract:
With today's continuously changing market conditions, yield ramp has become the key driver of product profitability. Standard yield improvement practices focus too narrowly on defect elimination and use techniques that solve yesterday's problems, not today's problems. We present a comprehensive methodology for yield and performance improvement specifically designed for today's increased process complexity and process-design interaction.

Introduction:

Problems facing yield and performance ramping are increasing as manufacturing sensitivity to critical design and processing parameters increase. For example, within-chip variations of dielectric thickness after chemical mechanical planarization (CMP) are not random, but rather are dependent on design layout[1]. Process complexity is also increasing as vertical integration and multi-level interconnects drive the number of processing steps above several hundred[2].

While product complexity is increasing, market windows are shrinking. A delay in an IC introduction means decreased revenue as prices decline. For IC's destined for consumer markets, opportunity can be lost as seasonal demand changes. Today's short IC lifetimes mean minimal production at "final" yield. The result is that initial yield and yield ramp rate have replaced final yield as the key drivers of IC profitability (Figure
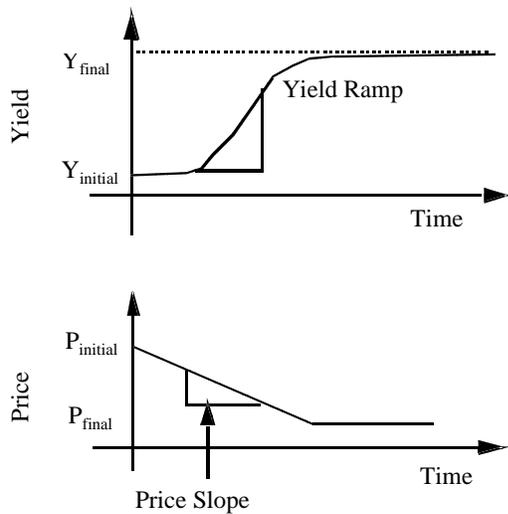
Figure 1. Typical yield vs. time and price vs. time relationships

Figure 2: Yield loss mechanisms

1). In other words, yield ramp rate is more important to profit margin than high-volume production yield.

Unfortunately, conventional yield improvement efforts focus primarily on defect elimination and do not solve today's problems. Today, a fab does not necessarily need more defect data. Instead, the answer lies in a "holistic" use of all existing data and simulation early in the production ramp to quickly form a comprehensive view of yield and performance learning.

A Holistic View:
IC yield loss mechanisms can be classified as either parametric or functional (Figure 2). Traditional yield improvement in the fab often refers to the reduction in contamination caused functional yield loss (e.g., a metal flake causing a short between two conductors). However, parametric variations (e.g., speed, power consumption) are increasingly important to final package yield. During the yield ramping phase, systematic and design-process margin yield
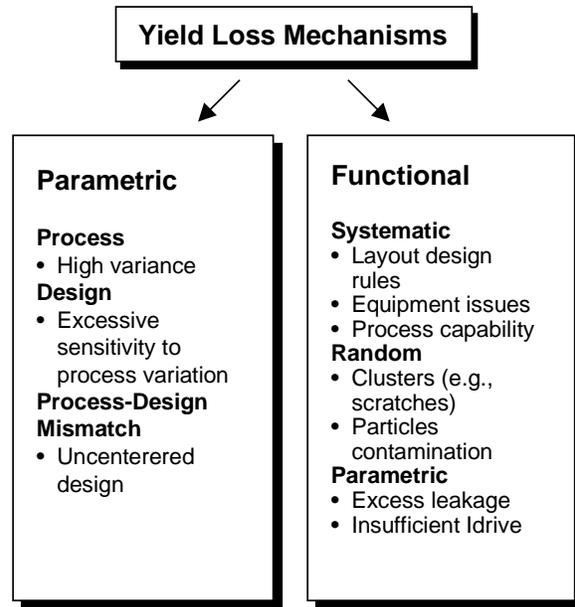
loss mechanisms typically dominate (Table 1).

Since the breakdown of yield loss components is significantly different for yield ramping than for high volume production, different yield improvement strategies should be adopted at each stage. A defect reduction focus is insufficient during initial production when systematic and design-process matching issues are significant.

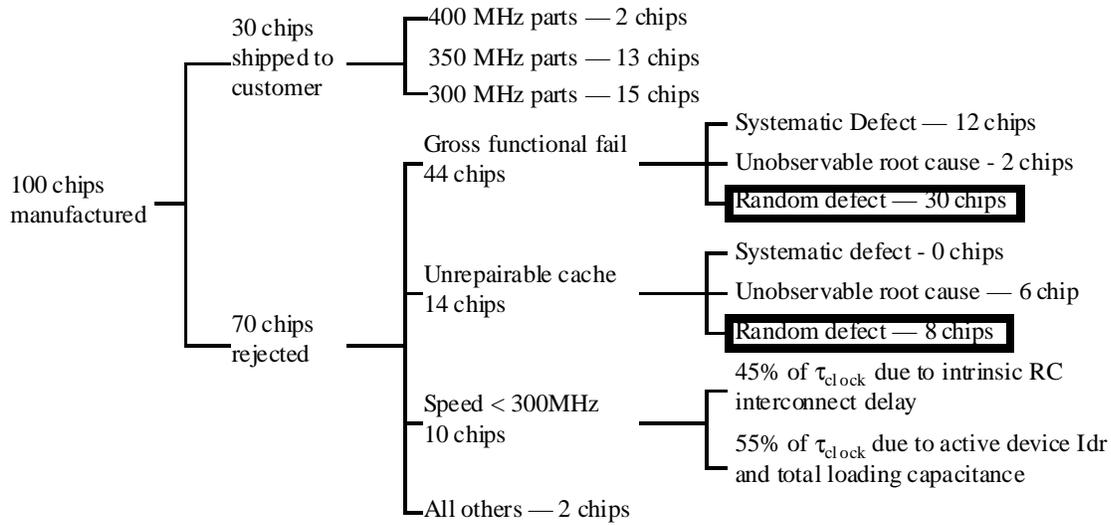| Probe Yield Problem	Impact on Final Yield	Impact on Yield Ramp
Contamination	60%	48%
Design-process interplay	14%	32%
Process variation	12%	10%
Photolith errors	10%	7%
Materials	4%	3%
Total	100%	100%

Table 1. Typical components of yield loss
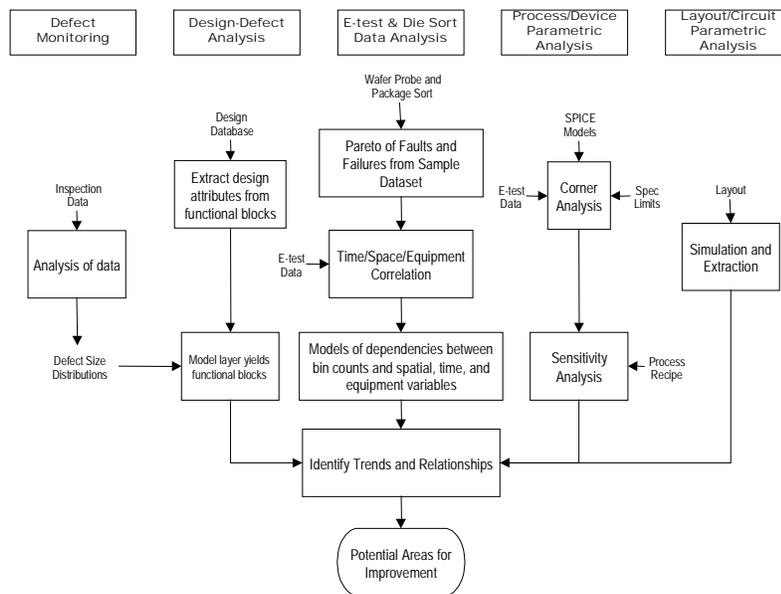
Figure 3. Example yield tree

Figure 4. Components of holistic yield analysis

Many issues which affect yield reside at the boundaries between traditional engineering groups. For example, organizational barriers typically exist between design and manufacturing groups.

It is difficult to quantify the balance between performance and manufacturing yield. The trade-off between test cost vs. information content is not well understood, and it is difficult to know which in-line results affect performance and yield most. These barriers slow the rate at which yield and performance can be improved.

Holistic Yield-Ramp Methodology:
We present a novel yield and performance ramping methodology that leverages simulation and hypothesis driven statistical analysis to overcome the barriers discussed above. The combination of yield and performance prediction with statistically-based data analysis enables us to isolate where actual yield and predicted yield are inconsistent. This situation indicates that a systematic yield loss mechanims may be present. The result of this analysis is a "yield tree" which quantifies the yield loss mechanisms. An example is shown in Figure 3.

We also incorporate a hypothesis-driven work style to maintain alignment between analyses and the decision making process. The methodology combines data from design, manufacturing, test, and the process recipe. In addition to helping predict yield and performance limiters when few lots have been fabricated, simulation enables multiple solutions for each problem to be examined concurrently in a timely manner, reducing the time and expense of relying solely on lot split experiments.

The five components of the analysis are defect monitoring, design analysis, data analysis, process-device parametric analysis, and layout-circuit parametric analysis (Figure 4). Design and process-device analyses support yield prediction while data analysis isolates the exact signatures of the yield loss mechanisms that affect the IC. Defect monitoring and design analyses are combined to form the basis of predictive yield modeling. Statistical simulation supports the determination of performance variation and yield loss components.
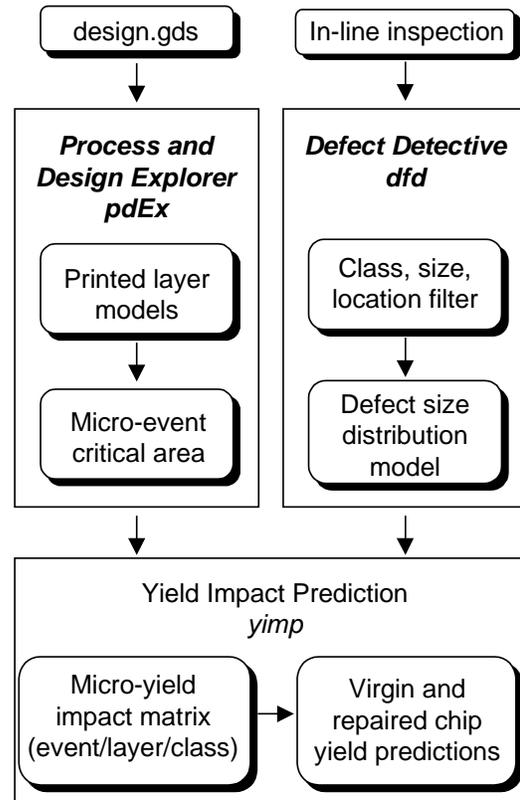

Figure 5: Defect-limited yield data flow

Limited Yield Modeling:
Predictive limited-yield modeling is indispensable to disaggregating the effects of individual root causes on final product yield. Multiple diverse yield-loss mechanisms, such as random defects, pattern-dependent effects, within-die process variations, and parametric process "miscentering," are typically present. By modeling final package yield as the product of limited yields, we can quickly identify and eliminate key yield detractors, and thus ramp yield more rapidly.

		Gate	Etch	M1		M2		M3		Total		
		particle	pattern	particle	pattern	particle	pattern	particle	pattern	particle	pattern	overall
yield after repair	chip	78%	80%	89%	81%	87%	89%	91%	91%	55%	52%	29%
	logic	82	83	95	87	91	92	93	93	66	62	41
	cache	95	96	94	93	96	97	98	98	84	85	71
virgin yield	chip	63	66	86	78	84	85	91	91	42	40	16
	logic	82	83	95	87	91	92	93	93	66	62	41
	cache	77	79	91	90	92	92	98	98	63	64	40

Table 2. Typical components of yield loss

As an example, we show the methodology for prediction of product- and process-specific "failure signatures" resulting from random defects[3]. The three components necessary are a design analyser (*pdEx*[4]), a measured data modeler (*Defect Detective*), and a defect limited-yield modeler and analyzer (*yimp*) (see Figure 5).

For IC's with memory, detailed analysis of block, row, bit, etc. failure signatures in probe-test bit maps have been used extensively in the past. However, such data-driven efforts are not suitable for predictive yield modeling. In addition, "macro" yield predictions, e.g., using whole chip critical area, are usually too coarse to determine the empirical relationship between failure event signatures and physical failure mechanisms. In contrast, the key step to our methodology is a prediction of "micro-yield loss events" which directly correspond to the "failure event signatures" observed in bitmap and bin test data.

The detailed defect-limited yield model developed includes calculation of yield losses per failure mode per defect type. This model is formulated in terms of critical areas per defect type and defect size density distributions. For geometries below 0.25μm, layout printability simulation is necessary to obtain more realistic representation of the actual semiconductor structure. Critical areas leading to micro-yield loss events are computed using geometrical manipulation[5]. All micro-yield predictions are guaranteed to be statistically independent by mapping each critical area polygon into a single event class.

By combining a hierachy of micro-yield events, a chip-level yield prediction is performed. Chip-, block-, and cell-level repair constraints and resources, as well as
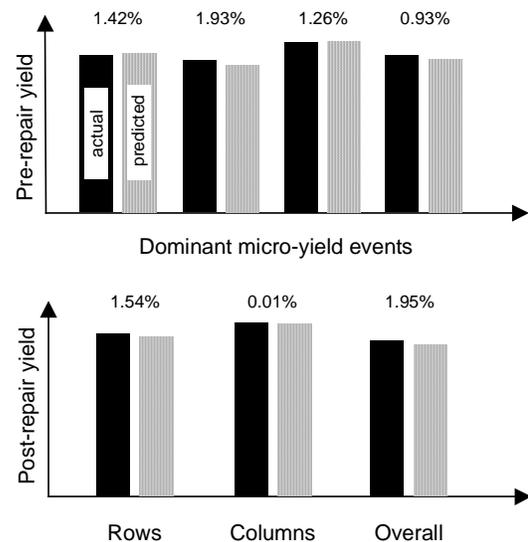

Figure 6. Actual vs. predicted pre-repair and repair yields

unrepairable interactions between these levels determine the model hierarchy.

The final step of the defect-yield prediction methodology calibrates the yield models

using available in-line inspection data, such as defects and critical dimensions. As with macro-yield predictions, micro-yield predictions require careful size-distribution analysis of in-line inspection data to accurately predict end-of-the-line yields. Data for such a calibration can be obtained from early test chips that are being used for verification of cell designs (e.g., the cell array test chips with minimum peripheral circuitry).

As an example of the possible accuracy for defect-limited yield predictions, a comparison of pre-repair and post-repair predicted yields with actual yields for a Flash memory product is shown in Figure 6.

The result of the defect-limited yield analysis is a yield impact matrix. A typical matrix for a high performance microprocessor is shown in Table 2. Based on this matrix, certain layers and defect types were chosen as targets for defect reduction efforts. In addition, through comparison with measured probe yields, a systematic cache yield loss mechanism was identified and quantified.

In addition to separating defect vs. systematic yield loss mechanisms, our defect limited-yield modeling has many applications to both design and manufacturing. Among the applications are layout design rule and redundancy optimization, determining defect budgets, inspection plan development, and even wafer disposition based upon in-line inspection results.

Modeling Performance and Process-Design Marginality:
Due to market pressures, performance improvement after design tape-out must come from process optimization rather than design or layout changes. Such performance

ramping requires the disaggregation of the impact of device and interconnect characteristics on overall performance (e.g., speed or power consumption). In addition, performance improvement cannot be assessed independent of the process-design marginality. Assessment of the statistical impact as well as nominal impact of process changes on performance and yield must be made.

Improving performance and process-design marginality in a timely manner requires a modeling methodology which combines simulation with measured data analysis, as illustrated in Figure 5. The methodology enables the space of possible process changes to be quickly explored for the impact on E-test and circuit performance.
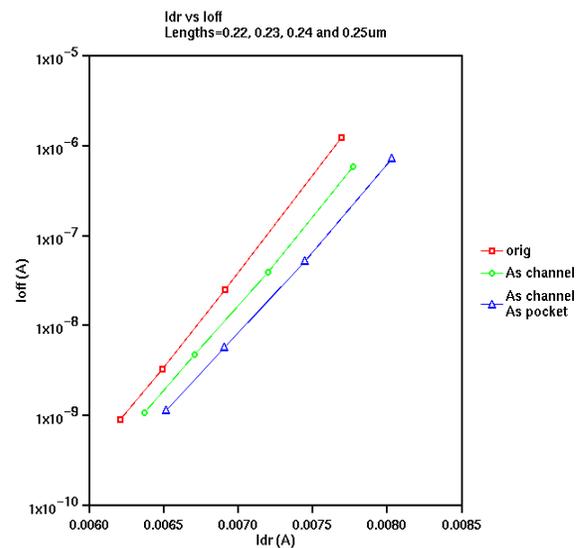

Figure 7. Simulation of process changes Idrive-Ioff relationship

A critical component is the use of statistical process and device (TCAD) simulation combined with physically-based SPICE model extraction, as provided in pdFab[6,7].
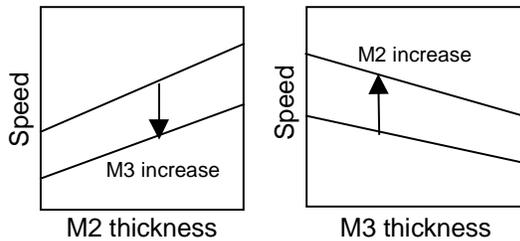
Figure 8. Analysis of interconnect changes on speed

Traditional SPICE parameter optimization is not appropriate for statistics, because the final parameter set is dependent upon the initial guess. The obtained correlations tend to be difficult to interpret and are often not physical, which may lead to incorrect design decisions and inaccurate design-process margin prediction.

As an example, we show the improvement of yield and performance for an advanced logic product. From data analysis, speed was determined to be a strong function of Idrive, and high Ioff was a source of functional yield loss due to dynamic gate failures.

To obtain speed and yield targets, the Idrive/Ioff ratio needed to be improved. Simple process changes, e.g., reducing Lpoly, were not sufficient as they do not change the Idrive-Ioff relationship.

Through simulation, we predicted that optimizing both the MOSFET channel profile and the addition of a pocket implant would lead to an improved Idrive/Ioff ratio. (Figure 7).

Analysis of speed as a function of interconnect properties indicated that decreasing interconnect delay could allow lower Idrive while achieving performance targets (Figure 8). By changing metal thicknesses and reducing the Idrive target,

functional yield loss due to leakage-related dynamic gate failures was further mitigated.

Verification lot splits confirmed the higher Idrive (resulting in improved performance) and tighter Ioff distribution (resulting in improved functional yield). Such an analysis can be performed in days rather than the weeks required for an experiment-based approach, while simultaneously enabling many more process changes to be explored.

Conclusion:
Traditional approaches to IC yield and performance improvement are time intensive and primarily defect focused, thus addressing only a fraction of today's yield-ramp challenges. We have developed a holistic yield improvement methodology that significantly reduces yield- and performance-ramp time by integrating process recipe, design information, in-line manufacturing data, and statistical simulation to identify and eliminate the full spectrum of yield loss mechanisms, including defects, parametric variation and systematic causes.

Using simulation where applicable and a hypothesis-driven work style, this approach delivers increased yield and performance in a fraction of the time of traditional methods. Geared towards today's leading edge ICs (i.e., <0.25-µm technology), the holistic yield improvement methodology is especially useful in ramping yield and performance of complex logic and system-on-a-chip ICs, as well as DRAM and Flash memory. Results have demonstrated a consistent doubling of the yield learning rate while achieving performance improvements of 10% to 20% over original specifications.

References:

[1]. E. Chang, et. al. "Using a Statistical Metrology Frame Work to Identify Systematic and Random Sources of Die- and Wafer-Level ILD Thickness Variation in CMP Processes," *1995 IEDM Conf.*

[2]. P.K. Chatterjee and G.B. Larrabee, "Gigabit Age Microelectronics and Their Manufacture*," IEEE Trans. On VLSI Systems*, vol. 1, no. 1, pp. 7-21, Mar. 1993

[3]. D. Ciplickas, et. al., "Predictive Yield Modeling for Reconfigurable Memory Circuits," Proc. *IEEE/SEMI 1998 ASMC*, Vol. 9, Sept 1998.

[4]. *pdEx* Users Manual, PDF Solutions Inc., 1998.

[5]. C. Ouyang and W. Maly, "Efficient Extraction of Critical Areas in Large VLSI IC's," *Proc. Int. Symp. Semiconductor Manufacturing (ISSM 96),* Tokyo, Japan, 1996, pp. 301-304.

[6]. *pdFab* Users Manual, PDF Solutions Inc., 1998.

[7]. D. Hanson, et. al., "Analysis of the Controllability of a Sub-Micron CMOS Process Using TCAD," *Proc. Int. Symp. Semiconductor Manufacturing (ISSM 94),* Tokyo, Japan, 1994, pp. 85-89.