Advanced Yield Learning Through Predictive Micro-Yield Modeling

Dennis J. Ciplickas, Andrzej J. Strojwas, Xiaolei Li, Rakesh Vallishayee, Wojciech Maly*
PDF Solutions, Inc., San Jose, CA, USA
*Department of Electrical and Computer Enginnering, Carnegie Mellon University, Pittsburgh, PA, USA

***Abstract*—**This paper describes a comprehensive methodology for predictive modeling of yield losses in deep submicron technologies. We also present a software framework pdEx which implements this methodology. We illustrate the versatility and accuracy of this modeling framework for high performance microprocessor and Flash memory products. We show that the extremely good prediction accuracy is achievable if the micro-level yield models are developed taking into account the available redundancy schemes, and the defect density and size distributions are properly extracted from the in-line inspection data.

INTRODUCTION

In the age of multi-billion dollar semiconductor fabrication facilities and increased time-to-market pressures, rapid yield learning is essential to achieve profitable production of integrated circuits. To be competitive, the cost per die must be minimized while quickly ramping the manufacturing yield to an economically acceptable level. Predictive yield modeling is an indispensable aid in this process, not only during the yield ramp phase, but also during technology and product development. This is especially true when multiple yield loss mechanisms may be present and include such diverse failure mechanisms as random defects, pattern-dependent within-die process variations and parametric process mis-centering. Furthermore, application of this methodology during technology or product development allows designers to evaluate certain types of yield loss and employ appropriate design optimizations.

This paper presents an accurate method for defect limited yield modeling and calibration. In a holistic yield improvement methodology developed at PDF Solutions, accurate defect limited yield prediction is a crucial step toward decomposing random and systematic components of measured probe yields. Figure 1 illustrates a typical “yield tree” resulting from such an analysis for a high performance microprocessor. As is evident from the breakdown shown in the figure, inaccurate predictions of random defect limited yield can lead to gross under or over estimations of systematic yield losses and consequent defocusing of yield improvement efforts. Other methods can be, and are, used to separate systematic from random yield loss [1,2], but their applicability is limited by their inconsistent accuracy and lack of insight into potential root causes. Both of these limitations are addressed in the algorithms described below.

The following section presents a detailed description of the defect limited yield modeling methodology used at PDF Solutions. This is followed by a presentation of the *pdEx* software framework implementing the methodology. We demonstrate the applications of *pdEx* to yield modeling in two examples taken from the state-of-the-art fabrication processes for microprocessor and Flash memory products. Finally, we conclude by outlining applications of PDF’s methodology and software to process/product development, yield ramping and volume manufacturing.

DEFECT LIMITED YIELD MODELING

The defect limited yield modeling methodology presented in this paper is very general and can be applied to both reconfigurable memory products and logic products with embedded memory. Detailed analysis of binmap and bitmap failure signatures has been used extensively in the past for yield improvement. These data driven efforts, however, have enjoyed comparatively little support from predictive yield models. In fact, "macro" yield predictions using the critical area of whole

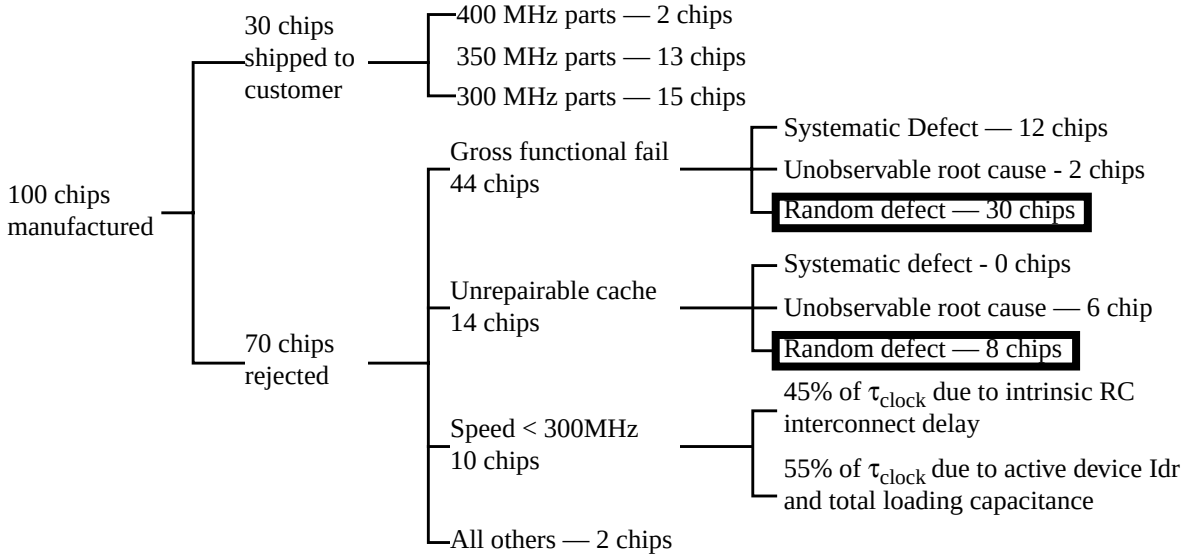

Figure 1. Yield Tree.

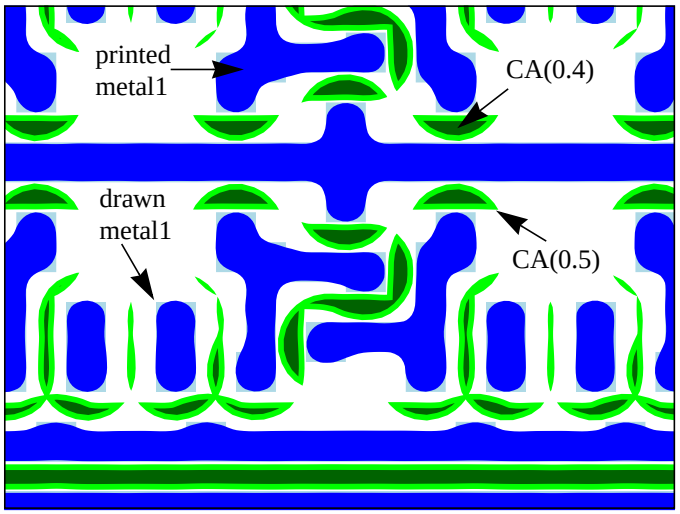

Figure 2. Critical areas of printed layout patterns in a 0.35μ SRAM.

chips are usually too coarse to provide insight into the physical mechanisms driving particular failure signatures. Consequently, significant quantities of in-line inspection and end-of-line test data must be gathered in order to determine empirical relationships between failure event signatures and physical failure mechanisms. In contrast, the key step in our methodology is a prediction of "micro-yield loss events" which directly correspond to "failure event signatures" observed in the probe test data. Examples of micro-yield loss events include individual logic block failures as well as memory array failures such as: two, three or more adjacent row shorts; two three or more adjacent column shorts; row/column shorts and lack of contact to a cell. This "bottom up" prediction of failure events not only allows for cross-verification of probe test yields but also provides an immediate correspondence between probe test results and individual failure mechanisms.

Micro-yield predictions are computed using a modified Poisson model:

$$Y_e = \prod_{\forall l} \exp\left(-\int_{x_0}^{\infty} [CA_{e,l}(x)][DSD_l(x)]dx\right), \qquad (1)$$

where x_0 is the minimum feature size in the technology, $CA_{e,l}(x)$ is the critical area of event e in layer l, and $DSD_l(x)$ is the density of defects in layer l with size x. Defect density functions $DSD_l(x)$ are estimated using in-line defect inspection data as discussed later. Critical area functions for each event $CA_{e,l}(x)$ are computed using a simulation of the printed layout patterns and a modification of the traditional oversizing algorithm for critical area extraction. During the extraction, each critical area polygon is categorized by the event type corresponding to the set of electrical nodes participating in each unique geometrical overlap. This concept is illustrated in Figure 2 and is described in more detailed in [3].

A chip-level yield prediction is achieved by combining a hierarchy of micro-yield loss events for the reconfigurable part of the product of interest. The overall yield is a product of the non-repairable and repairable block yields. In theory, it is possible to assign event classes to all possible combinations of nodes in an entire chip and generate a complete set of micro-yield events using the algorithm outlined above. These micro-yield events are then systematically combined to model all possible repair scenarios. In practice, however, this method is both infeasible and unnecessary. Depending on the repair options, a natural and sufficiently accurate yield model hierarchy often presents itself. The corresponding hierarchical yield model for reconfigurable memories was proposed in [3].

The final step of our methodology targets the crucial problem of calibrating the yield model using the available in-line inspection data. Just as with macro-yield predictions, micro-yield predictions require careful size distribution analysis of in-line inspection data from KLA213X to accurately predict the end-of-line yields. Histogram-based fitting methods were found to be unreliable and a more robust mean/variance matching method was used instead. A complete description of this fitting method is beyond the scope of this paper.

However, it is well known that the capture rate of the optical inspection tools is low for the smallest defect sizes. Hence, the reported defect density is smaller than the true value. To compensate for this equipment limitation, we have introduced a scaling function f(x) which measures the amount of extrapolation between the modeled and observed defect densities for a full range of defect sizes. Hence, the defect density used in yield model is given by:

$$DSD(x) = D_0 f(x) \frac{k}{x^p}, \qquad (2)$$

where D_0 is the measured defect density, $f(x)$ is the scaling function, and k/x^p is the size distribution function that integrates to 1.0 between x_0 and ∞. The distribution type is described by the parameter p which is extracted from the measured defect data for regions where the capture rate is sufficiently high. Then, this distribution type is used to extrapolate the defect sizes down to the minimum size of defect that can cause a chip failure.

It was found that using the alternate defect size definition $\sqrt{XY}$ provided more reliable yield predictions than the standard KLA "DSIZE" definition of $\min(X, Y, \sqrt{A})$, where X and Y are the horizontal and vertical dimensions of the bounding box containing an actual defect. This is a natural consequence of

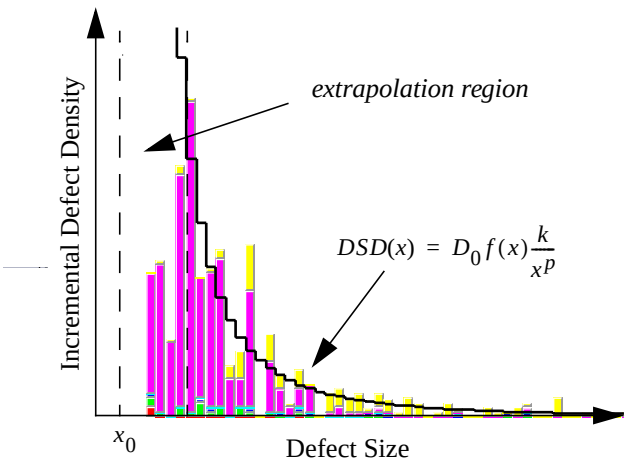

Figure 3. Typical DSD fit to KLA 213X data.

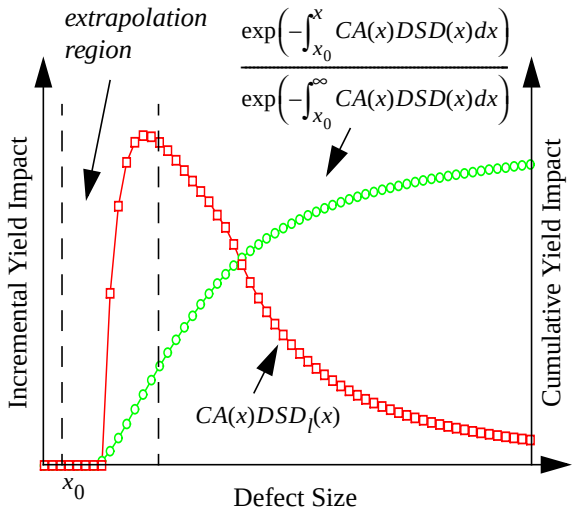

Figure 4. Incremental and cumulative yield impact curves calculated by *yimp* using critical areas from *pdex* and defect densities from *dfd*.

the assumptions present in critical area yield modeling. Namely, that all defects are both assumed to be symmetric, and that true defects tend to cause shorts along the maximum dimension rather than the minimum dimension.

A typical example of a fit to the defect size distribution (DSD) is shown in Figure 3. The extrapolated densities are typically much greater than the measured ones. Typical extrapolation factors (ratio of extrapolated to measured defect densities) may be in the range from 2 to 6, depending on the sensitivity of the inspection recipe and equipment used. More important than the value of the extrapolation factor, however, is the fraction of inferred yield impact. A typical yield impact curve corresponding to the DSD fit in Figure 3 is shown in Figure 4. As is evident in the figure, up to 25% of the predicted yield impact, including the peak yield impact region, may be accumulated in the extrapolated portion of the DSD curve. Furthermore, as shown below, the accuracy of the overall yield prediction indicates that such extrapolations are valid extensions of the measured defect density at larger defect sizes.

pdEx FRAMEWORK

A software framework has been created to implement this methodology [4]. Three components are used: a design analyzer (*pdEx*), a measured data modeler (*dfd*) and a defect-limited yield modeler and analyzer (*yimp*). Experience has shown that this division of labor provides the best trade-off in software efficiency and usability. The software architecture and dataflow is shown in Figure 5 below.

To perform an efficient defect limited yield modeling of memory products, *pdEx* was enhanced to perform the critical area categorization “on the fly” during a conventional Boolean AND operation. Since this Boolean AND operation is already required for “macro” critical area extraction, no execution time penalty is incurred for the micro-yield event extraction. Furthermore, *pdEx* allows seamless modeling of mask to topography transfer processing for increased accuracy in critical-area calculations.

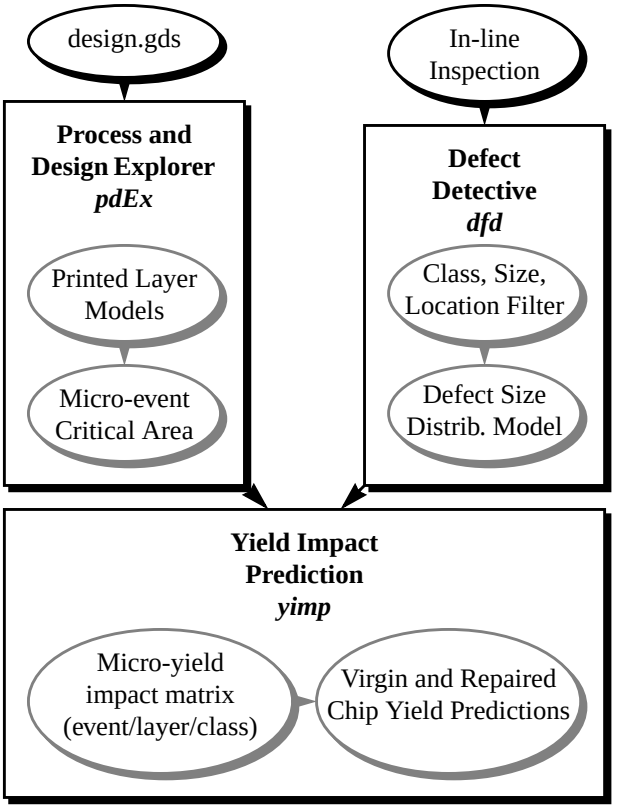

Figure 5. Software implementation and dataflow.

Since the analysis of KLA213X data is not a turn-key process, *dfd* is implemented to allow flexible application of defect filters, size definitions, model fitting methods and fit ranges. Taken together, these settings allow accurate assessment and encapsulation of in-line defect data for yield prediction.

Finally, the yield impact tool *yimp* has been implemented to facilitate rapid calculation, review and interpretation of yield predictions across design blocks, micro-yield events and processing layers. The modeling and display methods (e.g., Figure 4 and Table 1) have been proven in the field to be invaluable for focusing defect reduction efforts and separating systematic vs. random yield losses.

pdEx APPLICATIONS

The micro-yield predictions and methods described above have been validated through comparison with measured bin sort and bitmap yields.

First, atypical yield impact matrix resulting from an analysis for a high performance microprocessor is shown in Table 1. Extensive defect classification analysis and embedded memory redundancy analysis are key elements of a successful yield prediction. Various conclusions were drawn from this matrix. For example, certain layers and defect types were chosen as target for defect reduction efforts. Second, through comparison with measured probe yields, a systematic cache yield loss mechanism was identified and quantified.

Second, a state-of-the-art Intel’s Flash memory product was analyzed. A detailed description of this study was presented in [3]. Critical area curves were extracted for the *poly1*, *poly2*,

TABLE 1. Typical yield impact matrix produced by *pdex*, *dfd* and *yimp* for a microprocessor product.

YIMP Matrix		gate etch		m1		m2		m3		total		
		particle	pattern	particle	pattern	particle	pattern	particle	pattern	particle	pattern	overall
yields after repair	chip	0.78	0.80	0.89	0.81	0.87	0.89	0.91	0.91	0.55	0.52	0.29
	logic	0.82	0.83	0.95	0.87	0.91	0.92	0.93	0.93	0.66	0.62	0.41
	cache	0.95	0.96	0.94	0.93	0.96	0.97	0.98	0.98	0.84	0.85	0.71
virgin yields	chip	0.63	0.66	0.86	0.78	0.84	0.85	0.91	0.91	0.42	0.40	0.16
	logic	0.82	0.83	0.95	0.87	0.91	0.92	0.93	0.93	0.66	0.62	0.41
	cache	0.77	0.79	0.91	0.90	0.92	0.92	0.98	0.98	0.63	0.64	0.40

metal1 and *metal2* layers from the Flash device in this study. For increased accuracy in predicting coplanar, intra-layer shorts, the *contact* layer and *poly2* layer were combined for the *poly2* extraction. Examples are shown in Figure 6. Finally, micro-yield predictions for each event type (*e.g.*, single bit, double row) were formed by combining individual layer yields according to (1).

The four micro-yield events which dominated yield losses were the intra-layer shorts at two poly levels and two metal levels. Since their occurrence in the array was observable in the sort test results, we were able to compare directly our model accuracy for both the virgin (raw) and repaired yield results. These comparisons were performed for a development fab (Fab A) and a volume production fab (Fab B). In both cases, the accuracy of the model was extremely good, which is demonstrated in Figure 7 for the pre-repair yield for the four dominant events. The accuracy of the overall yield prediction for the memory array (2 planes) was also excellent, namely 0.4% in both Fab A and Fab B. To verify our redundant yield model accuracy, we have compared the post-repair yields for the row and column events by grouping the appropriate poly and metal level events and taking into account the redundancy scheme in the actual Flash memory product. The results for Fab A are shown in Figure 7 together with the overall repaired yield prediction error (the accuracy for the latter was 1.95%). For Fab B, similarly good results were obtained and the overall redundant yield prediction error was also less than 2%).

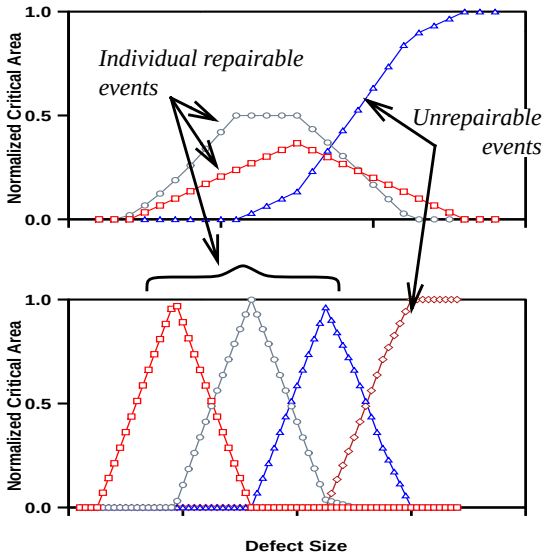

Figure 6. Typical critical areas for micro-yield events extracted by *pdEx*.

CONCLUSIONS

This paper presented a new comprehensive methodology for predictive modeling of yield losses in deep submicron fabrication processes. The *pdEx* software framework was shown to be extremely accurate in predicting the defect limited yield for Flash memory products, as well as being crucial for separating random vs. systematic yield in both memory arrays and high performance microprocessors. As we have shown, *pdEx* goes far beyond traditional critical area-based yield modeling tools. Its accuracy is due to the realistic modeling of micro-yield loss events and available redundancy, but also due to sophisticated post-processing of the raw defect inspection data. The resulting yield models can be used for derivation of optimal design rules, local optimization of memory array and microprocessor core layouts, and evaluation of redundancy and ERC needs. Moreover, defect targets per layer and type can be derived and transferred to the volume production fablines.

REFERENCES

[1] P. Mullenix, J. Zalnoski and A. J. Kasten, "Limited Yield Estimation for Visual Defect Sources", *IEEE Trans. on Semiconductor Manufacturing*, vol. 10, no. 1, pp. 17-23, February 1997.

[2] C. H. Stapper, F. M. Armstrong and K. Saji, "Integrated Circuit Yield Statistics", Proc. IEEE, Vol. 71, pp. 453-468, April 1983.

[3] D. Ciplickas, X. Li, R. Vallishayee, A. Strojwas, R. Williams, M. Renfro, R. Nurani, "Predictive Yield Modeling for Reconfigurable Memory Circuits", Proc. IEEE/SEMI 1998 ASMC, Vol. 9, Sept 1998.

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

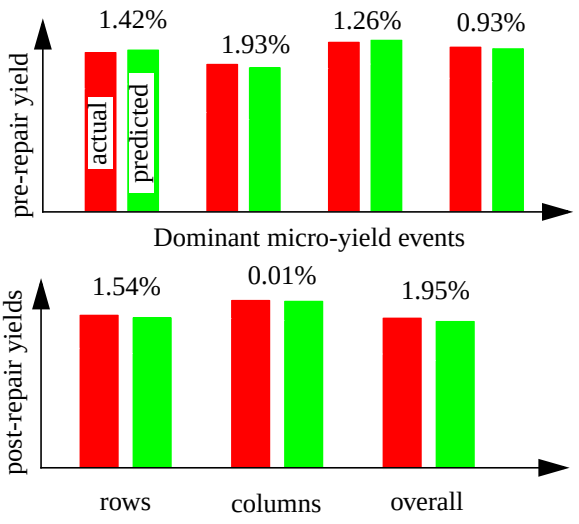

Figure 7. Actual vs. predicted pre-repair and post-repair yields. The numbers above the bars represent the absolute errors in yields.