Yield: Statistical Modeling and Enhancement Techniques

Israel Koren

Dept. of Electrical and Computer Engineering

University of Massachusetts, Amherst

Supported in part by JPL and by NSF

Preliminaries – Defects

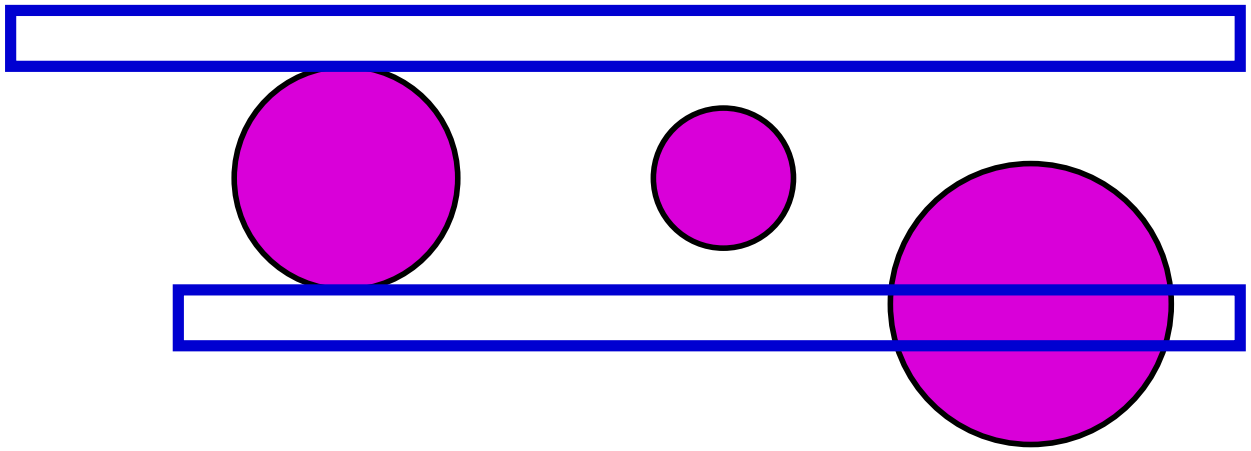
- Yield – fraction of operational ICs manufactured
- Yield losses – caused mainly by systematic defects and random spot defects
- Systematic defects – parametric defects, mask misalignments dealt with during the fabrication process
- Random spot defects – result of local contamination (dust particles) – cannot be eliminated
- Their effect can be minimized during the design process
- We will focus on spot defects – their statistical modeling and ways to minimize their harm

Defect Types

- Spot defects classification:
  - Missing material – may result in open circuits
  - Extra patterns – may result in short circuits
  - intra-layer defects– missing or extra material between adjacent wires
  - Inter-layer defects – missing or extra material between two separate layers
- Not all spot defects result in structural faults
- Only defects that turn into faults cause yield losses
- What is the percentage of defects that become faults?

Defect vs. Fault

- Assumption – A defect is circle shaped with diameter x (x is a random variable)
- Other geometrical shapes can be considered
- To become a fault, the defect needs to connect two disjoint conductors or disconnect a continuous pattern



Critical Area

- pof –
  Probability Of Failure: percentage of defects that cause faults – depends on the type i and diameter x of the defect
- A defect will cause a fault if its center is located "properly"
- $A_i^c(x)$ –
  the critical area for defects of type i and diameter x :
  the size of the area in which the center of a defect of type i and diameter x must fall in order to cause a circuit fault
- Assuming that the centers of the defects are uniformly distributed over the chip area, then

$$pof_i(x) = \frac{A_i^c(x)}{A_{chip}}$$

A_{chip} – area of the whole chip

Defect Size Distribution

- To compute the yield, critical area must be averaged over all defect sizes x and defect types i
- To average over x, $f_d(x)$ – the probability density function of the defect size x, is necessary
- Experimental data shows that $f_d(x)$ decreases as $1/x^p$ between x_0 – the resolution limit of the lithography process, and x_M – the maximum size of a defect
- p and x_M are determined empirically and may depend on the defect type i (Typically, $2 \leq p \leq 3.5$)
- Therefore, $f_d(x) = \begin{cases} k/x^p & \text{if } x_o \leq x \leq x_M \\ 0 & \text{otherwise} \end{cases}$

  where $k = (p-1)x_0^{p-1}x_M^{p-1}/(x_M^{p-1} - x_0^{p-1})$

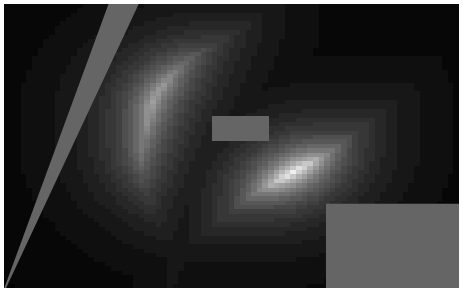
Averaging Critical Areas

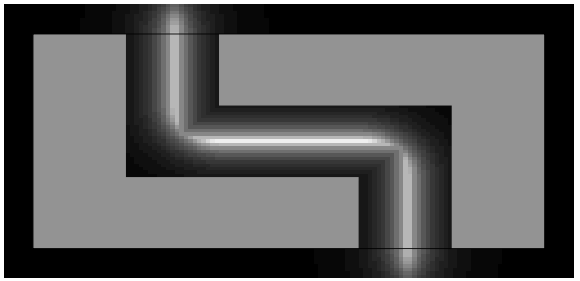
- Averaging over defect diameters x:
  - $pof_i = \int_{x_0}^{x_M} pof_i(x)\ f_d(x)\ dx$
  - $A_i^c = \int_{x_0}^{x_M} A_i^c(x)\ f_d(x)\ dx$
- Averaging over defect types i:
  - d_i – type i defect density – average number of defects of type i per unit area
  - λ – average number of circuit faults (of all types) on the chip
  - $$\lambda = \sum_i A_{chip} \cdot\ d_i \cdot\ pof_i = \sum_i A_i^{(c)} \cdot\ d_i$$

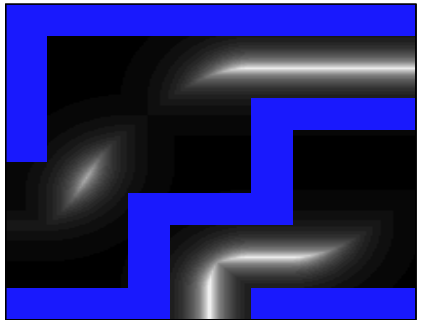
Calculating Critical Areas

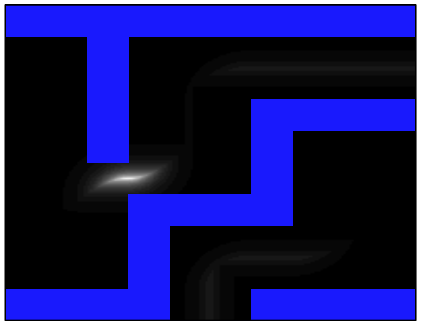
- Calculating either pof or critical area of chip – essential for yield prediction
- Several methods of calculation:
  - Geometry-based methods for calculating $A_i^c(x)$
  - Monte-Carlo methods for calculating $pof_i(x)$
- Geometrical methods – polygon expansion technique, etc.
- Monte Carlo approach – simulated circles representing defects of different sizes are placed at random locations in the layout. The fraction of defects of type i and diameter x which cause a fault is an estimate for $pof_i(x)$
- Commercial and academic tools for calculating critical areas exist

Critical Area Maps







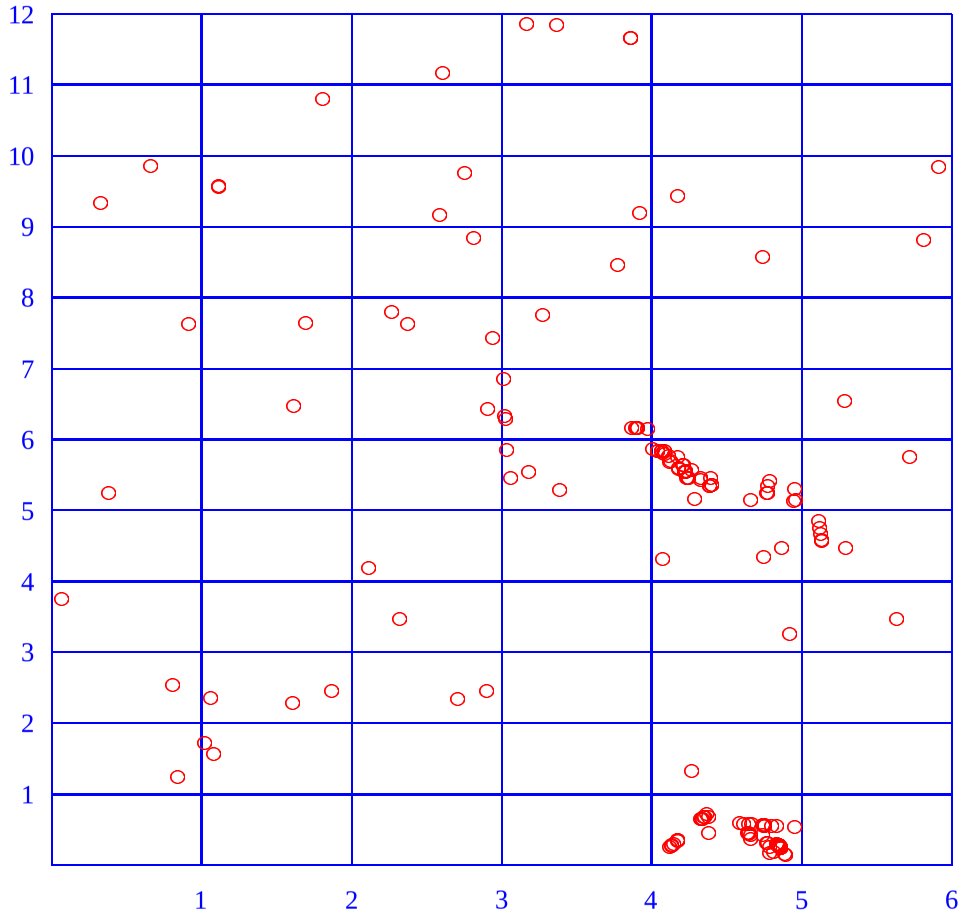


Source: I. A. Wagner and I. Koren, "An Interactive VLSI CAD Tool for Yield Estimation," *IEEE Trans. on Semiconductor Manufacturing,* Vol. 8, pp. 130-138, May 1995.

Yield Modeling

- λ – The average number of faults – main parameter in yield modeling
- In simplistic models – λ of the whole chip is used
- In more advanced models – λ_i for chip component i or for chip layer i
- Yield model – a probabilistic model describing the distribution of faults over the chip, using the parameter λ and possibly additional parameters
- Which probabilistic model should be used?

Wafer Defect Map



Defect Maps

- The previous picture showed a defect map
- Faults are a percentage of the defects
- Their distribution on the wafer will look similar, only with fewer points
- How do we describe the distribution of defects/faults in a probabilistic manner?

Basic Yield Models

- In the past, the spatial Poisson distribution was used to describe the defect and fault distribution
- It assumes a uniform distribution of defects and faults over the wafer resulting in

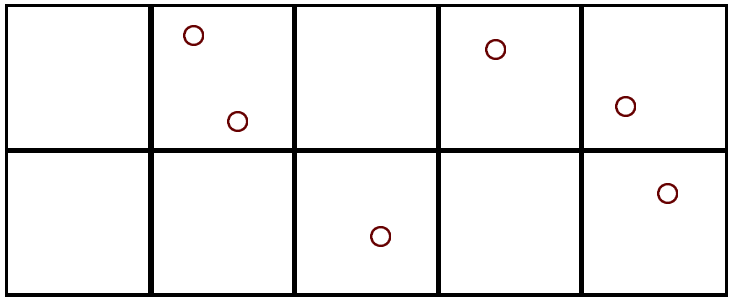
$$Prob(Number\ of\ Faults\ on\ Chip\ =\ k)\ =\ e^{-\lambda}\ \frac{\lambda^k}{k!}$$

$$Chip\ Yield = Prob(Number\ of\ Faults\ on\ Chip = 0) = e^{-\lambda}$$

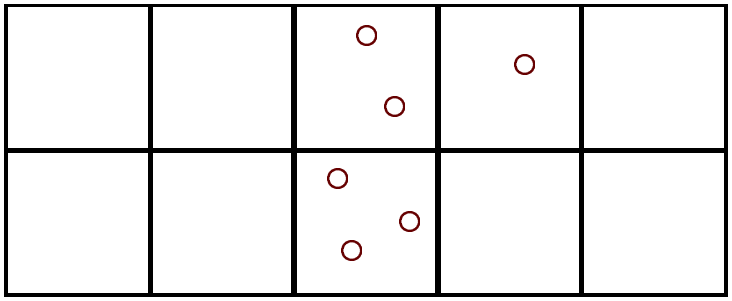
- This is a very simplistic model and does not fit empirical manufacturing data
- In practice, both defects and faults are more clustered than predicted by the pure Poisson distribution

Yield Models – Clustering

Effect of fault clustering on chip yield



(a) Non-clustered faults,
$\boldsymbol{Y_{chip} = 0.5}$



(b) Clustered faults,
$\boldsymbol{Y_{chip} = 0.7}$

Clustering increases the yield, and must be incorporated into the fault distribution model

Introducing Clustering

- Clustering is introduced by regarding λ – the average number of faults – as a random variable with a given probability density function
- Different density functions for λ result in different fault distributions
- The most common model – λ has a Gamma density function

Negative Binomial Yield Model

- Integrating

$$Chip\ Yield\ =\ e^{-\lambda}$$

over λ with regard to the Gamma distribution results in

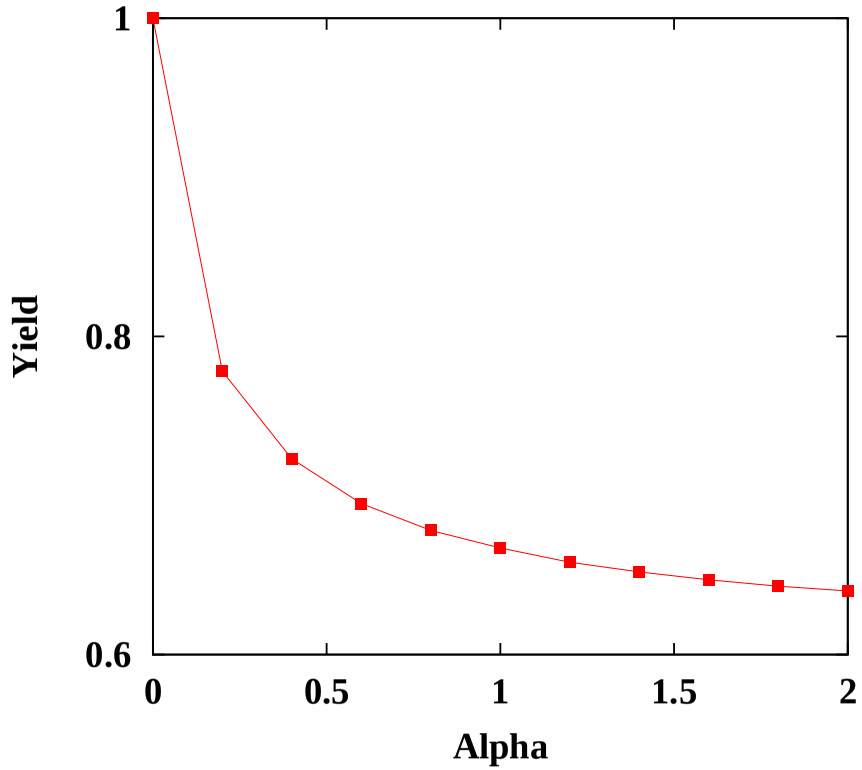
$$Chip\ Yield\ =\ \left(\frac{1+\lambda}{\alpha}\right)^{-\alpha}$$

- This is called the Negative Binomial model

Negative Binomial Model – cont.

- It is characterized by the two parameters:
  λ – the average number of faults per chip
  α – the clustering parameter
- Typically, $0.5 \leq \alpha \leq 5$
- The smaller the value of α , the more pronounced the clustering
- As $\alpha \longrightarrow \infty$, the Negative Binomial distribution approaches the Poisson distribution
- The use of the Negative Binomial distribution for yield projection is now the industry standard

Effect of Clustering on Yield



The effect of α on the chip yield ($\lambda = 0.5$)

Negative Binomial Distribution – Variations

Variations on the basic model:

- Including the effect of systematic defects in addition to spot defects

$$Chip\ Yield\ =\ Y_0 \left(\frac{1+\lambda}{\alpha}\right)^{-\alpha}$$

  Y_0 is the called the gross yield factor and is the probability that there are no systematic defects on the chip

- Dealing with varying fault densities such as for different chip components or for different chip layers

$$Chip\ Yield\ =\ \Pi \left(\frac{1+\lambda_i}{\alpha}\right)^{-\alpha}$$

Yield Enhancement

- All the models mentioned so far assume no redundancy in the chip –
  to be operational, the whole chip must be fault-free
- The yield of a chip can be enhanced through
  - Architecture choice (i.e., redundancy)
  - Decreasing the critical area at the design stage (during placement, routing and compaction)
  - Decreasing the defect density
- We will concentrate on the first two options

Effect of Redundancy on Yield

- Redundancy – effective for identical, replicated circuits
- Most common example – large memory chips
- Basic model: N modules needed, R spares added, all identical
- For proper operation - at least N out of the $N+R$ must be fault-free
- λ – average number of faults per module

$$Chip\,Yield = \sum_{i=N}^{N+R} \sum_{k=0}^{N+R-i} (-1)^k \binom{N+R}{i} \binom{N+R-i}{k} \left(\frac{1+(i+k)\lambda}{\alpha}\right)^{-\alpha}$$

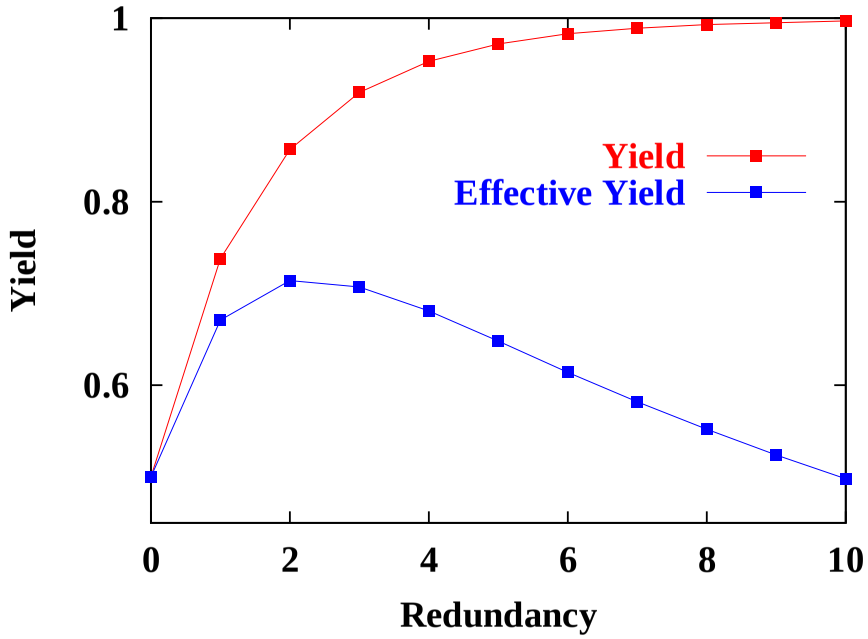
Effective Yield

- Adding spare modules increases the yield
- At the same time it also increases the chip area – less chips out of the wafer
- Even with a higher yield, we may end up with fewer operational chips per wafer
- Yield may not be the right measure for circuits with redundancy
- Effective Yield takes into account the increase in chip area

$$Y_{chip}^{eff} = Y_{chip} \frac{Area\ of\ Chip\ Without\ Redundancy}{Area\ of\ Chip\ With\ Redundancy}$$

Yield vs. Effective Yield

$N = 10,\ \lambda = 0.1,\ \alpha = 1,\ R = 0, 1, 2, ..., 10$



The maximum value of Y_{chip}^{eff} determines the optimal amount of redundancy to be incorporated into the chip

Yield of Defect-Tolerant Memories

- Large memories must have some redundancy for yield enhancement
- Conventional method of redundancy – spare rows and columns
- Memory chip includes some components (mainly control circuitry) which are uncorrectable
- $Chip\ Yield\ =\ Y_0 \cdot Y_c \cdot Y_{uc}$

  Y_0 – gross yield factor
  Y_c – yield of correctable part (with redundancy)
  Y_{uc} – yield of uncorrectable part

New Defect-Tolerant Memories

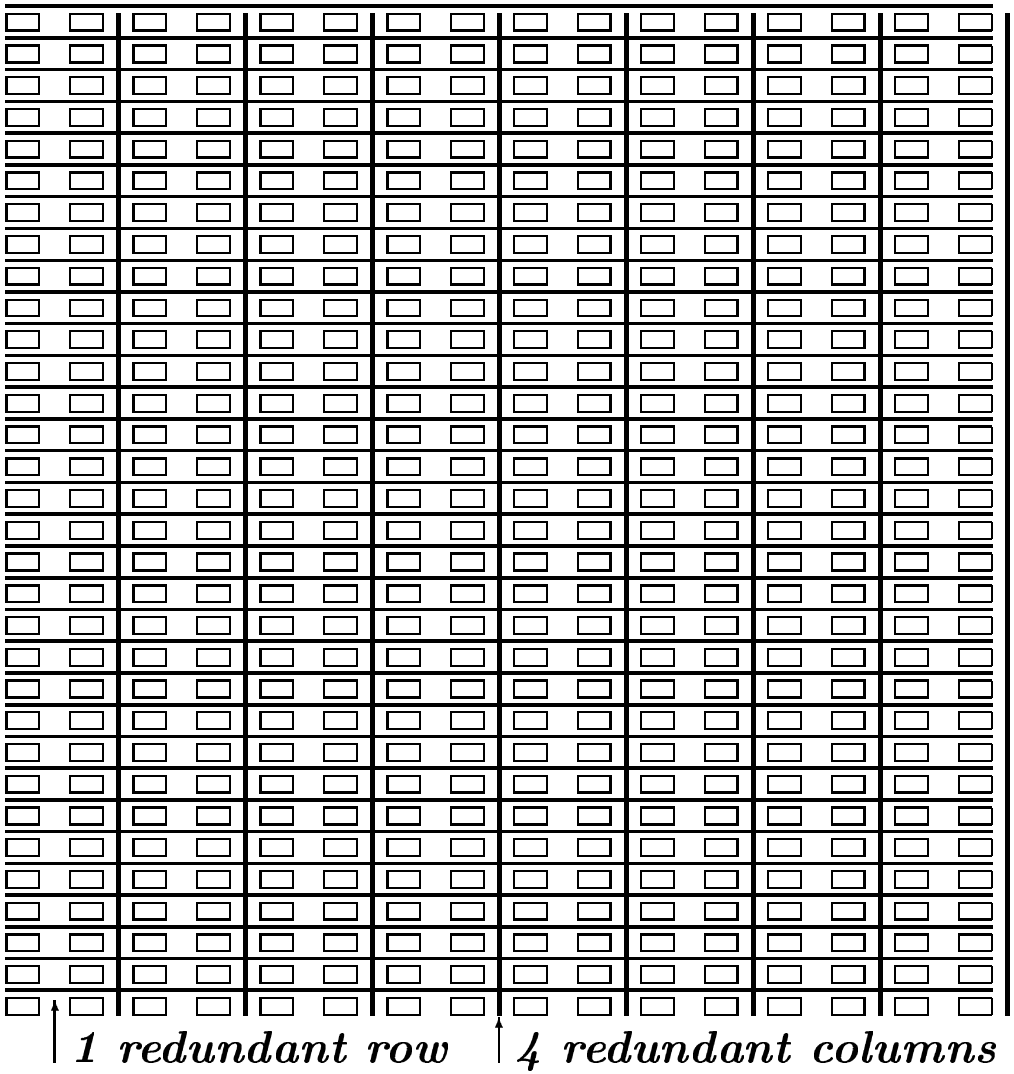
- Memory ICs have become very large
- Conventional redundancy of rows and columns is not sufficient
- Partitioning into sub-arrays is a must
  - Decrease the current
  - Shorten bit & word lines to reduce access time
- Disadvantages of conventional techniques
  - Inefficient use of redundant lines
  - Unable to deal with chip-kill defects
- New defect tolerance techniques are necessary

Memory with Redundant Blocks

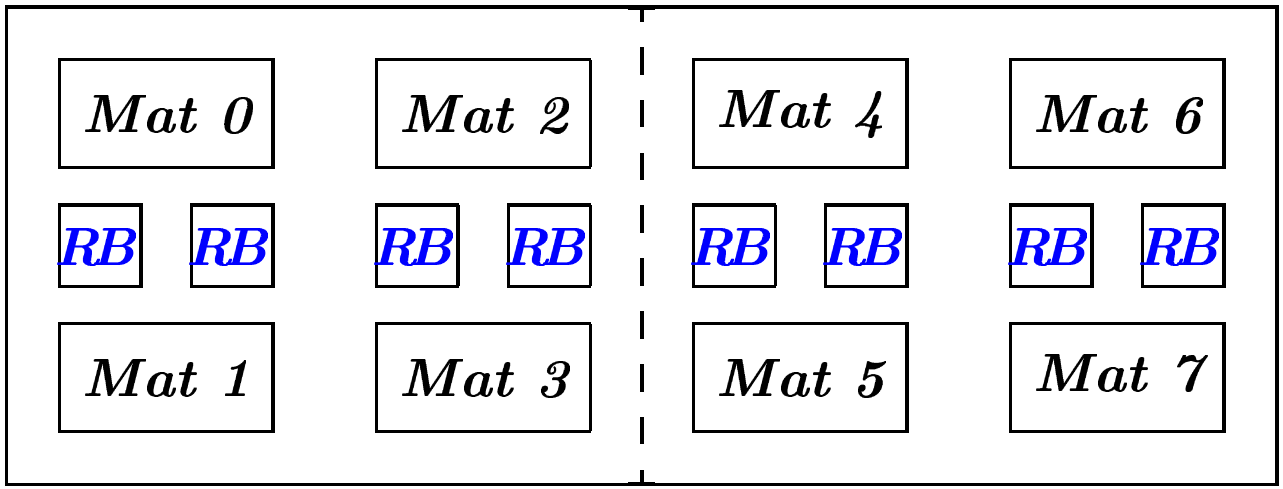
J-H. Yoo *et al.*, "A 32-Bank 1Gb Self-Strobing Synchronous DRAM with 1GB/s Bandwidth," *IEEE J. of Solid-State Circuits*, vol. 31, pp. 1635-1643, Nov. 1996.

- 1 Gb DRAM is partitioned into eight 128 Mb mats
  - 512 basic arrays of size 256Kbit (32×16 matrix)
  - 32 spare rows and 32 spare columns
  - Four spare rows are allocated to a 16Mbit portion of the mat
  - Eight spare columns are allocated to a 32 Mbit portion of the mat
- Eight redundant blocks of size 1Mbit each
  - Four basic 256Kbit arrays
  - Eight spare rows + four spare columns

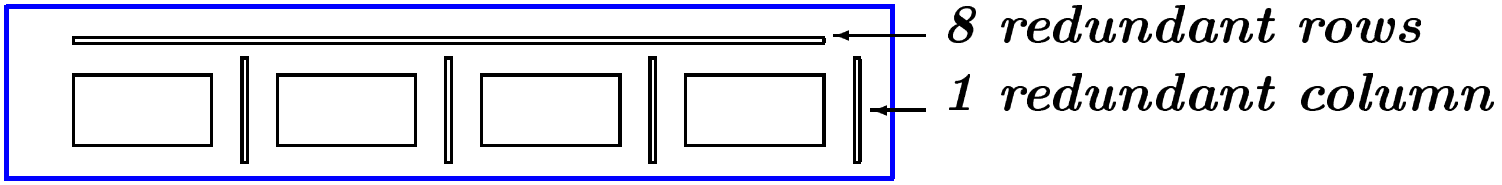
128Mb mat (32×16 256Kbit arrays)



Block Diagram

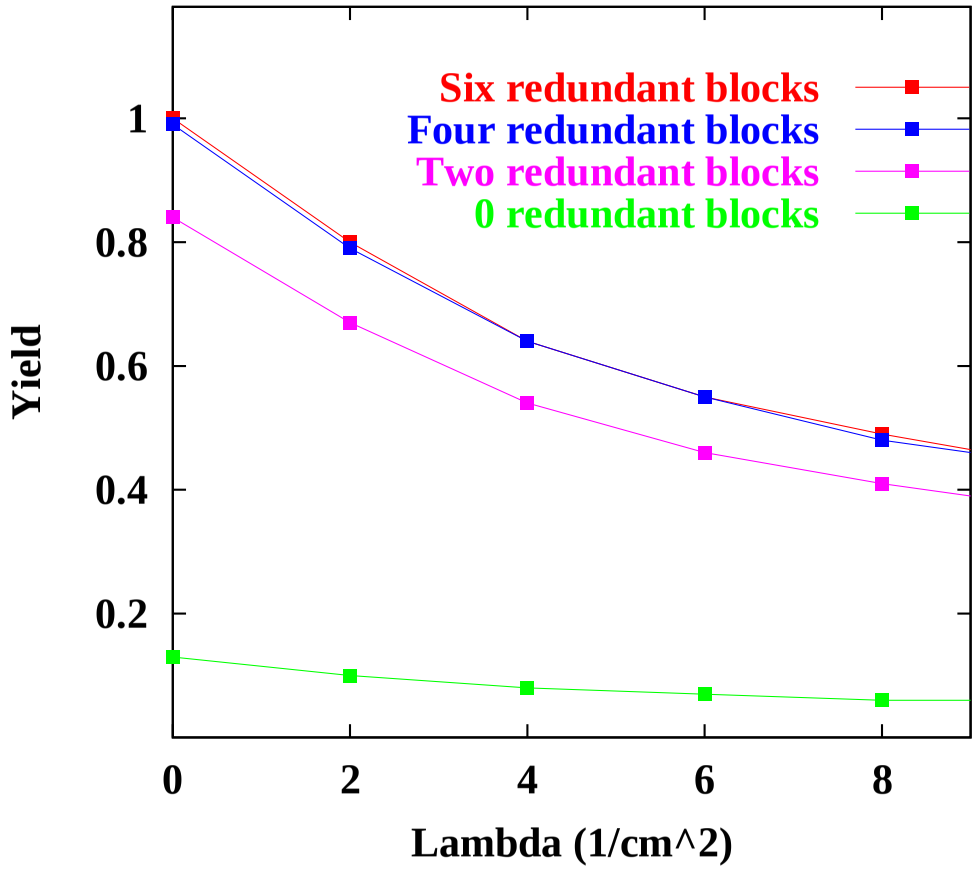


Eight mats (128Mbit each) + eight redundant blocks (*RB*) (1Mbit each).



A redundant block including four 256Kbit arrays, eight redundant rows and four redundant columns.

Yield Comparison

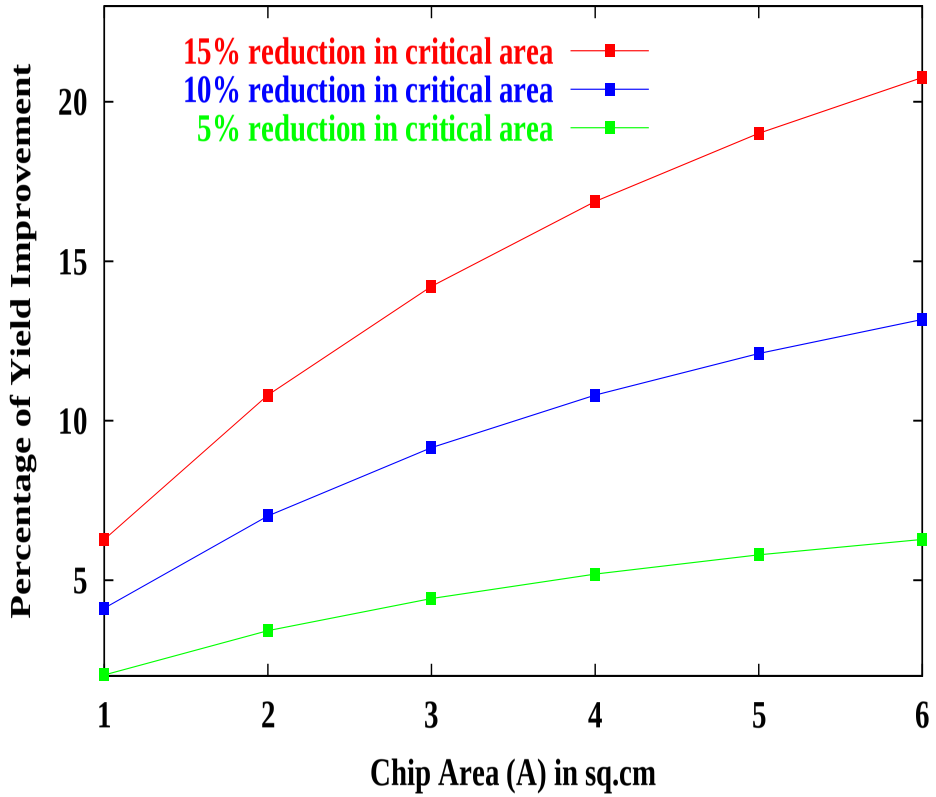


Source: **I. Koren and Z. Koren, "Analysis of a Hybrid Defect-Tolerance Scheme for High-Density Memory ICs,"** *Proc. of the 1997 IEEE Intern. Symp. on Defect and Fault Tolerance in VLSI Systems*, **pp. 166-174, Oct. 1997.**

Yield Enhancement At The Design Stage

- Adding redundancy increases the chip size
- Modifications at the design stage can decrease the critical area without increasing the chip area
- Decreasing the critical area increases the yield

Effect of Decrease in Critical Area



$$\text{Percentage of Yield Improvement} = \frac{\text{New Yield - Old Yield}}{\text{Old Yield}} \times 100$$

Source: V.K.R. Chiluvuri and I. Koren, "Layout Synthesis Techniques for Yield Enhancement," *IEEE Trans. on Semicond. Manufacturing,* pp. 178-187, May 1995

Critical Area Reduction During Compaction

Two approaches to yield enhancement at the compaction stage:

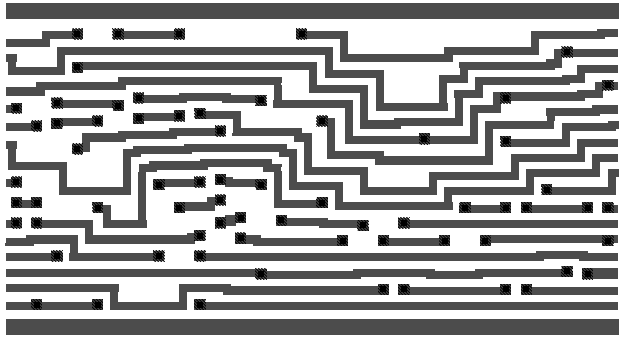
- Local modification in the layout as a post-compaction step
- Modification of the compaction algorithm for critical area reduction

Reductions in critical area of about 8%

Minimization of Short-Circuit Type Faults

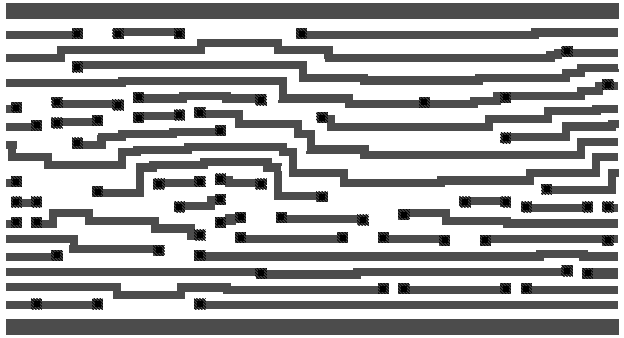
(a)

Original layout



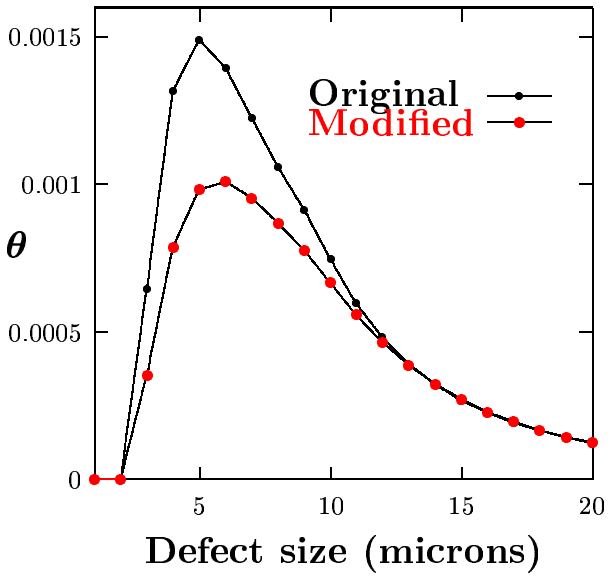
(b)

Modified layout

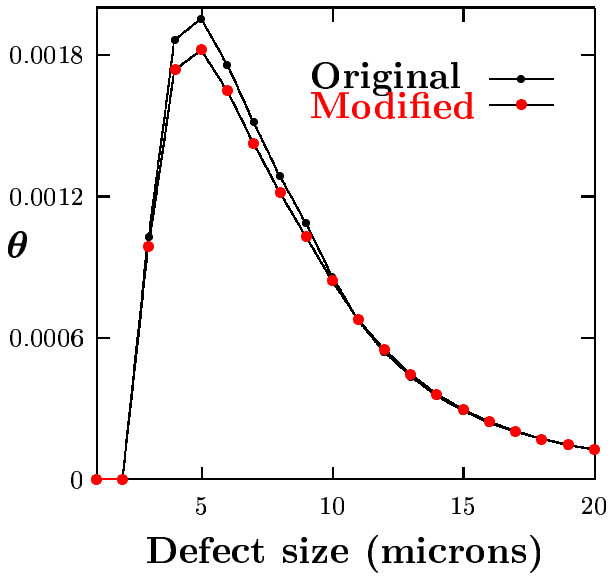


Effect of Uniform Spacing

Probability of failure (θ) vs. Defect size



(a) Short-circuit faults



(b) Open-circuit faults

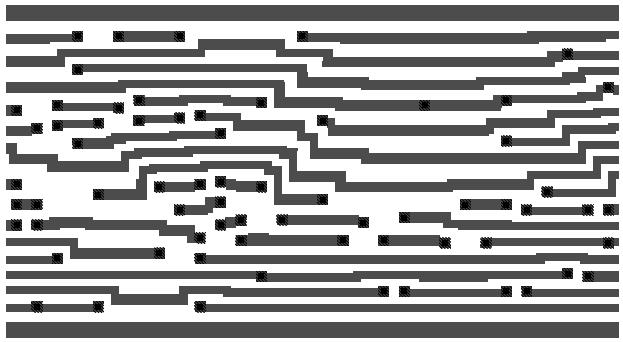
Minimization of Open-Circuit Type Faults

(a)

Modified for same defect density

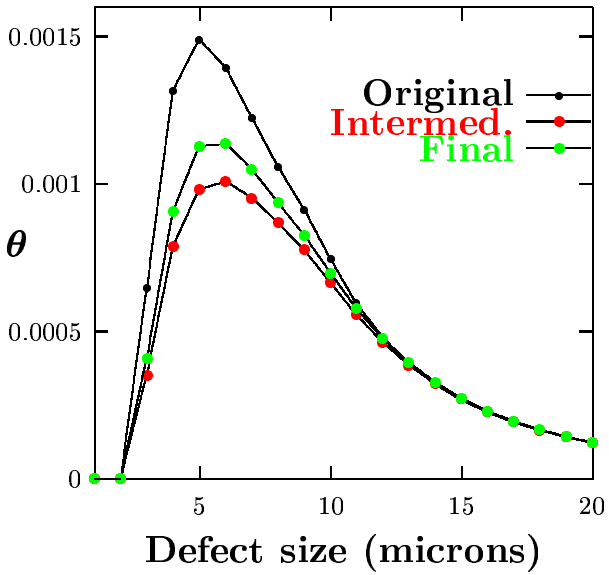
(b)

Modified for higher short-circuit defect density

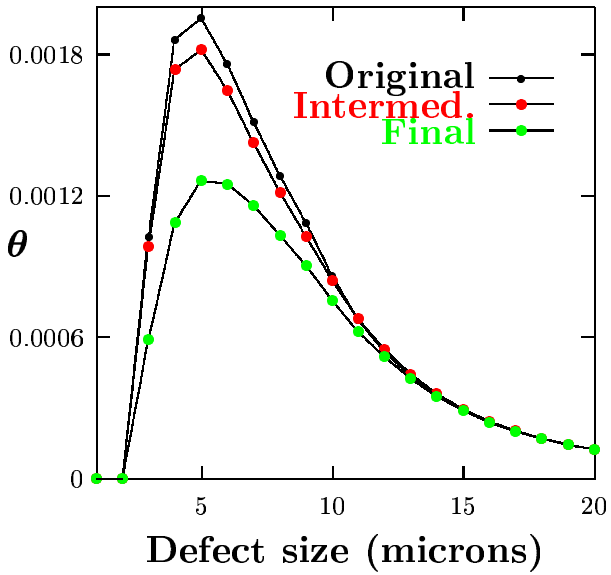


Effect of Wider Wires

Probability of failure (θ) vs. Defect size



(a) Short-circuit faults



(b) Open-circuit faults

Critical Area Reduction During Routing

The effectiveness of enhancements during compaction depends on the previously obtained layout

Two approaches to yield enhancement at the routing stage:

- Local modification in the layout as a post-routing step
- Modification of the routing algorithm for yield enhancement

Reductions in critical area of about 6%

Conclusions

- As density and size of integrated circuits increase, so do the importance and difficulty of achieving high yield
- Accurate yield projection at every stage of the design is of utmost importance
- Yield enhancement efforts must be part of the physical design stage