Handling Partial Correlations in Yield Prediction

Sridhar Varadan
Dept of ECE
Texas A&M University

Janet Wang
Dept of ECE
University of Arizona

Jiang Hu
Dept of ECE
Texas A&M University

Presentation Outline

- **What is Yield?**
- Difficulties in Yield Prediction
- Previous Research
- Proposed Research
- Simulation Results
- Conclusion

What is Yield?

- Yield - Probability of any Manufacturing or Parametric spec satisfying its limits.
- Manufacturing Yield – for manufacturing specs.
- Parametric Yield – performance measures (timing, power etc.)
- Process variations affect yield prediction.
- Intra-die process variations no longer negligible.

Process Variations

- Chip manufacturing involves complex chemical and physical processes.
- Tighter pitches and bounds make process variations unavoidable.
- Types of process variations –
  1. Systematic process variations – layout dependent
  2. Random process variations –
     a. Inter-die Random variations – depend on circuit design
     b. Intra-die Random variations – dominant components
        (1) Independent random variations
        (2) Partially correlated random variations
  3. Overall intra-die variations at *n* locations –

$$p(n) = \mu(n) + \varepsilon(n)$$

where $\mu(n)$ – systematic intra-die variations
$\varepsilon(n)$ – random intra-die variations

CMP Yield

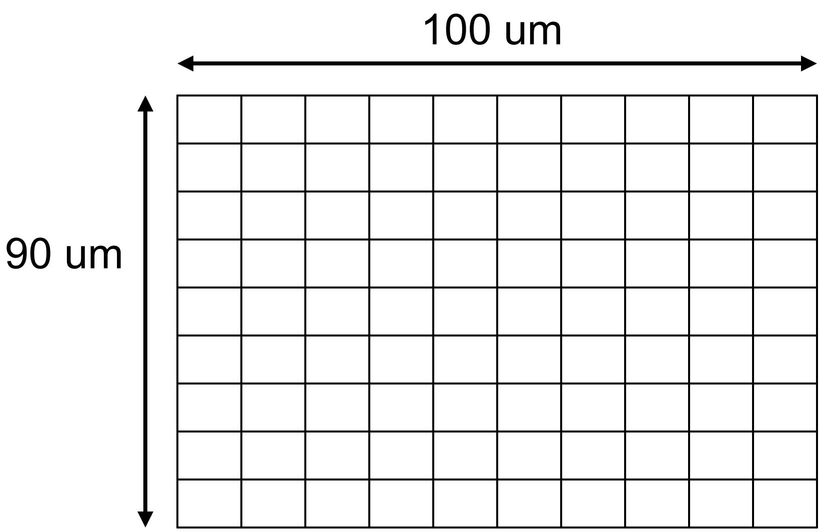
- Chemical Mechanical Planarization (CMP) – used in patterning Cu interconnects.
- CMP model – Yield is probability of thicknesses at all locations lying within the Upper and Lower thickness limits.
- For simplicity, a chip is meshed into a no. of tiles.
- Each tile is a location monitored for interconnect thickness.
- Meshing a chip into small tiles –

  Dimension – 100 μm x 90 μm.

  Size of each tile – 10 μm x 10 μm

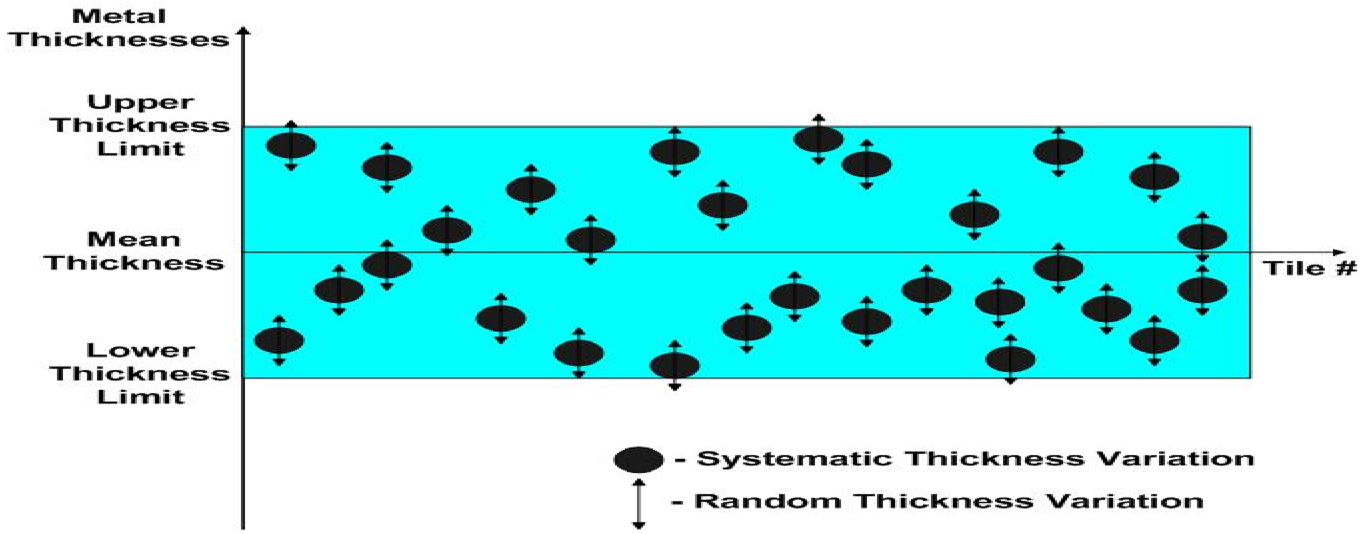
  Total no. of tiles – 90

  No. of locations monitored – 90



Illustrating a CMP Model

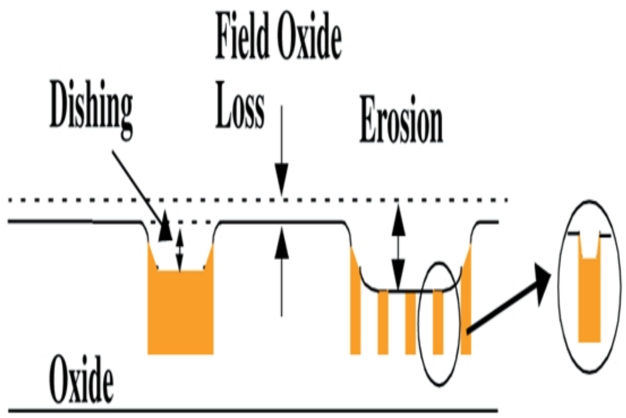
- Process variations in interconnect thicknesses at *n* locations –



- CMP Yield –Probability for thickness at *n* locations to lie in the shaded region.

Need for Predicting CMP Yield

- Factors making Yield Prediction important –
  1. Presence of Process Variations
  2. Shrinking feature sizes
- Dishing – Excessive polishing of Cu.
- Erosion – Loss in field oxide between interconnects.



- Potential open and short faults in interconnects.
- Predict Yield in circuit design stages to get Yield friendly design.

Equations for Yield Prediction

- Yield is obtained via numerical integration of a joint PDF -

$$Y = \int_L^U \int_L^U \ldots\ldots \int_L^U \phi(p).\,dp_1 dp_2 \ldots\ldots dp_n \qquad \ldots..(1)$$

$$\phi(p) = \frac{(\overrightarrow{p} - \overrightarrow{\mu})^T \Sigma^{-1} (\overrightarrow{p} - \overrightarrow{\mu})}{\sqrt{(2\pi)^n |\Sigma|}} \qquad \ldots..(2)$$

Where Σ - covariance matrix for the *n* variables – $\{p_1, p_2, \ldots, p_n\}$

U, L & *μ* - upper and lower thickness limits, & mean thickness value.

Yield equation *(1)* can be decomposed as –

$$Yield = Y_U + Y_L - 1 \qquad \ldots.(3)$$

Where Y_U *(High Yield)* - probability for thickness at all locations to stay below upper thickness limit.

Y_L *(or Low Yield)* - probability for thickness at all locations to stay above lower thickness limit.

Presentation Outline

- ✓ What is Yield?
- ➢ **Difficulties in Yield Prediction**
- ➢ Previous Research
- ➢ Proposed Research
- ➢ Simulation Results
- ➢ Conclusion

Difficulties in Yield Prediction

- Issues Affecting Yield Prediction –

1. Large number of locations to monitor (10^4–10^6).

2. Independent & partial correlations between locations.

3. Large memory requirements.

4. Complexity of numerical integration due to problem size.

Presentation Outline

- ✓ What is Yield?
- ✓ Difficulties in Yield Prediction
- ➢ **Previous Research**
- ➢ Proposed Research
- ➢ Simulation Results
- ➢ Conclusion

Previous Research

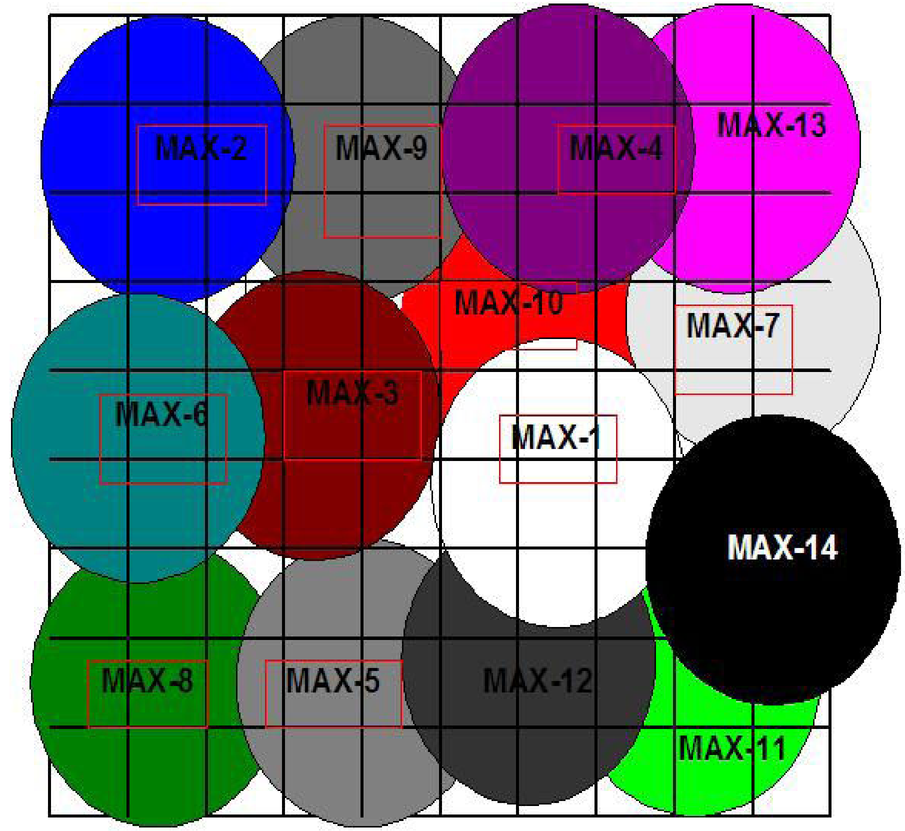
- Perfect Correlation Circles (PCC) approach – to reduce no. of tiles.
- Luo, et al., DAC 2006

Algorithm for PCC Approach

1. Find tile with maximum thickness MAX_1.
2. Form *PCC – CIRCLE*$_1$ (centre at MAX_1, pre-fixed radius).
3. Find tile with maximum thickness MAX_2 outside *CIRCLE*$_1$.
4. *Form PCC CIRCLE*$_2$ (centre at MAX_2).
5. Form similar *PCCs* until no tiles are left uncovered by *PCCs*.
6. Centers of PCCs (MAX_1, ….., MAX_m) form reduced set of variables.
7. Use Genz algorithm to compute yield.

Example Showing Reduction

- Let the setup look like this after reduction →
- Reduction from 90 tiles to 14 variables (the centres of PCCs – MAX_1, ….., MAX_{14}.)
- PCCs are formed in a sequence –
  MAX_1 – $CIRCLE_1$,
  MAX_2 – $CIRCLE_2$,
  ……………………..,
  MAX_{13} – $CIRCLE_{13}$,
  MAX_{14} – $CIRCLE_{14}$.



- Compute Low Yield using similar procedure.

Pros and Cons of the PCC Approach

- Advantages –
  1. Reduction in problem complexity.
  2. Reduced run-time.

- Disadvantages –
  1. Yield Accuracy is affected.
     a. Large PCC radius → Heavy reduction in variables.
        (over-estimation in yield)

     b. Small PCC radius → Lesser reduction in Variables.
        → more accurate yield estimate
        (but larger run-time)

Presentation Outline

- ✓ What is Yield?
- ✓ Difficulties in Yield Prediction
- ✓ Previous Research
- ➢ **Proposed Research**
- ➢ Simulation Results
- ➢ Conclusion

Proposed Research

- Develop reduction methods to –
  1. Reduce problem complexity.
  2. Reduce effect on yield accuracy.

- Two new methods for predicting yield –

  1. *Orthogonal Principal Component Analysis (OPCA)*

  2. *Hierarchical Adaptive Quadrisection (HAQ)*

Yield Model used in this Work

- Let vector $\vec{p}$ be metal thicknesses at *n* locations –

$$\vec{p} = (p_1, p_2, \ldots, p_n)^T$$

- This vector an be decomposed as follows –

$$p_i = \mu_i + \delta_i \quad \text{and} \quad \mu_i = \mu + \Delta_i$$

where μ – nominal value

Δ_i – systematic variation

δ_i – random variation

Orthogonal Principal Component Analysis

- Objective – Transform correlated random variables to a reduced & uncorrelated set through an orthogonal base

- Procedure –

  1. Form initial thickness vector, correlation & covariance matrices.
  2. Perform Eigenvalue Decomposition.
  3. Transform into to set of uncorrelated variables through a mapping matrix.
  4. Discern unwanted eigenvalues to get reduced set of uncorrelated variables.

- Initial Setup for OPCA –

  Let the initial thickness variations at *n* locations be –

$$\vec{\delta} = \{\delta_1, \delta_2, \ldots\ldots, \delta_n\}^T \qquad \ldots.(1)$$

- Let Γ_{nxn} and $\sum_{nxn}$ be the corresponding correlation and covariance matrices. Let σ_i^2 be the variance.

$$\Gamma(\vec{\delta}) = (\Gamma_{ij})_{nxn} \text{ and } \sum(\vec{\delta}) = \Gamma(\vec{\delta})_{nxn} \cdot \sigma_i \cdot \sigma_j \qquad \ldots.(2)$$

Using Eigenvalue Decomposition

- Re-express covariance matrix using Eigenvalue Decomposition –

$$\sum(\vec{\delta}) = Q \cdot \Lambda(\vec{\delta}) \cdot Q^T \qquad ....(3)$$

where $\Lambda(\vec{\delta})$ – eigenvalue (diagonal) matrix

Q – corresponding eigenvector matrix

- The diagonal matrix $\Lambda(\vec{\delta})_{n \times n}$ will look like –

$$\Lambda(\vec{\delta}) = \begin{pmatrix} \lambda_1 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & \lambda_2 & \cdot & \cdot & \cdot & 0 \\ 0 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & 0 & \cdot & \cdot & \cdot & \lambda_n \end{pmatrix} \qquad ....(4)$$

such that $\lambda_1 \geq \lambda_2 \geq .... \geq \lambda_n$

- Eigenvalue decomposition gives dominant directions in covariance relationship between a correlated set of variables.

Mapping into a New Set of Variables

- Let $\mathcal{E}_{n \times 1}$ be the new set of uncorrelated variables such that –

$$\vec{\delta} = B \cdot \vec{\varepsilon} \qquad \text{....(5)}$$

- Without loss of generality, assume B follows a Gaussian Distribution –

$$\mu(\vec{\varepsilon}) = 0 \;\; \& \;\; \Lambda(\vec{\varepsilon}) = I \qquad \text{....(6)}$$

- The matrices $\delta_{n \times 1}$ and $\mathcal{E}_{n \times 1}$ are related as follows –

$$\Lambda(\vec{\delta}) = J \cdot \Lambda(\vec{\varepsilon}) \cdot J^T \qquad \text{....(7)}$$

where

$$J = \begin{pmatrix} \sqrt{\lambda_1} & 0 & \cdot & \cdot & 0 \\ 0 & \sqrt{\lambda_2} & \cdot & \cdot & 0 \\ 0 & 0 & \cdot & \cdot & 0 \\ 0 & 0 & \cdot & \cdot & 0 \\ 0 & 0 & \cdot & \cdot & 0 \\ 0 & 0 & \cdot & \cdot & \sqrt{\lambda_n} \end{pmatrix}$$

- Transforming through an Orthogonal Base – Let B be the mapping matrix –

$$B = Q \cdot J \qquad \text{....(8)}$$

Transforming through an Orthogonal Base …. Contd..

- Correspondingly, we have –

$$\vec{\delta} = B \cdot \varepsilon = Q \cdot J \cdot \vec{\varepsilon} \qquad \text{….(9)}$$

- This transforms the initial set of correlated random variables to an uncorrelated set through an orthogonal base.

$$\sum(\vec{\delta}) = Q \cdot \Lambda(\vec{\delta}) \cdot Q^T = Q \cdot J \cdot \Lambda(\vec{\varepsilon}) \cdot (Q \cdot J)^T \qquad \text{….(10)}$$

- Reducing the no. of uncorrelated variables –

  1. After reduction, if we have *k* variables, then matrices $\Lambda(\vec{\delta})$ and $J_{k \times k}$ are –

$$\Lambda(\vec{\delta}) = \begin{pmatrix} \lambda_1 & 0 & \cdot & \cdot & 0 \\ 0 & \lambda_2 & \cdot & \cdot & 0 \\ 0 & 0 & \cdot & \cdot & 0 \\ 0 & 0 & \cdot & \cdot & 0 \\ 0 & 0 & \cdot & \cdot & 0 \\ 0 & 0 & \cdot & \cdot & \lambda_k \end{pmatrix} \quad \& \quad J = \begin{pmatrix} \sqrt{\lambda_1} & 0 & \cdot & \cdot & 0 \\ 0 & \sqrt{\lambda_2} & \cdot & \cdot & 0 \\ 0 & 0 & \cdot & \cdot & 0 \\ 0 & 0 & \cdot & \cdot & 0 \\ 0 & 0 & \cdot & \cdot & 0 \\ 0 & 0 & \cdot & \cdot & \sqrt{\lambda_k} \end{pmatrix} \qquad \text{….(11)}$$

  2. The corresponding sizes of matrices B and Q become $n \times k$, *thus giving reduction.*

Hierarchical Adaptive Quadrisection

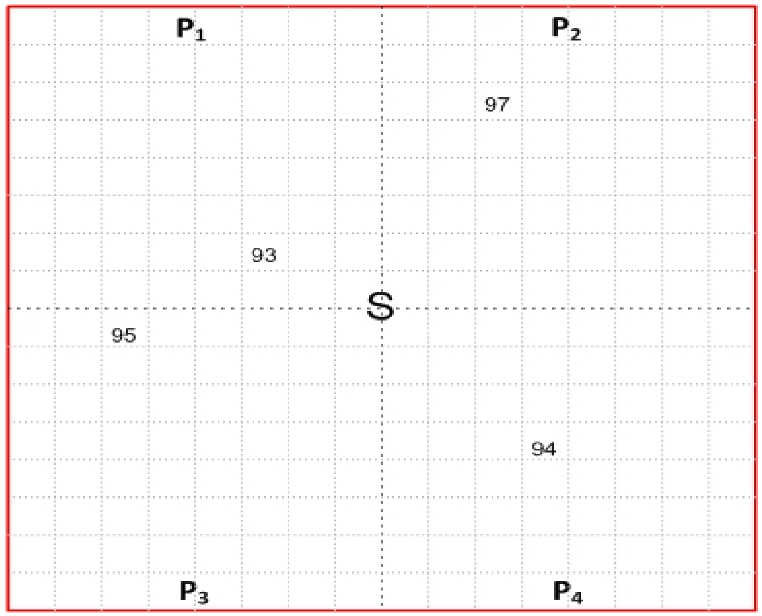
- Conquer and divide based clustering approach.
- Clustering done using *sub-regions* (similar to PCCs).
- Clustering in *sub-regions* is based on thickness variations.
- Sizes of clusters are not homogeneous.

Computing High Yield using HAQ

- Consider entire chip as one basic *sub-region S*.
- Sub-region *S* consists of tiles used in evaluating yield.
- Threshold thickness value θ decides possibility of clustering.
- Threshold θ tells on variations in thickness of tiles in a sub-region.

Working Model for Computing High Yield

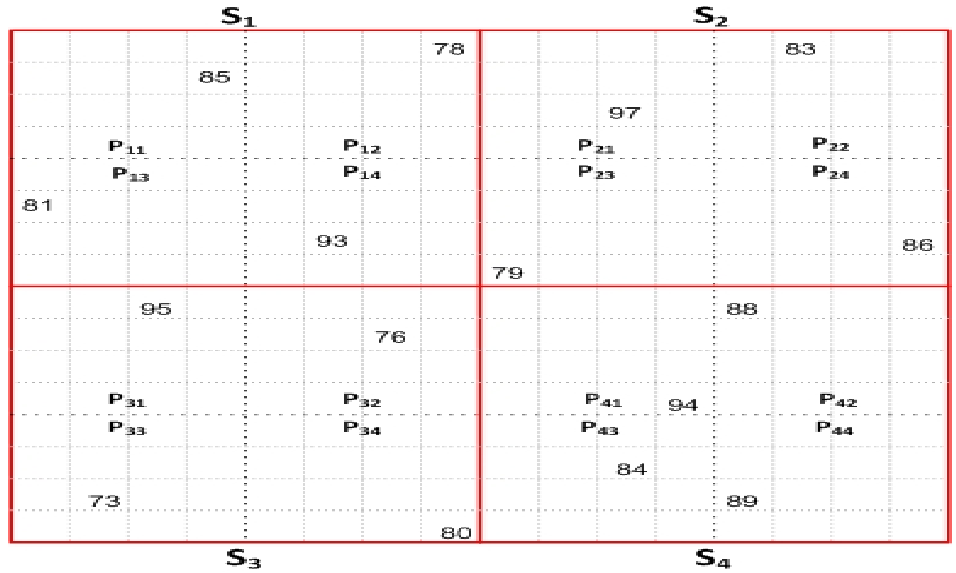
- Stage 1:- Sub-region S covers the entire chip. Let θ *be 10.*



Sub-region Monitored	Max Thickness		C_d	$C_d \leq \theta$	Next Action
	Critical	Non-Critical			
S	97	93, 95, 94	2	Yes	*Quadrisect*

Working model for High Yield ……. Stage 2

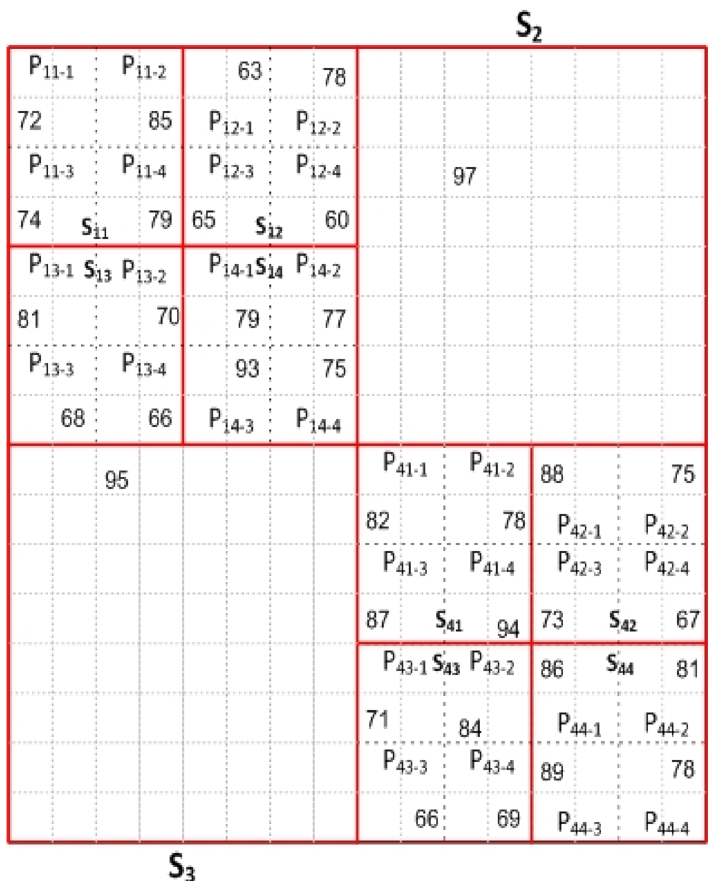
- Stage 2 – After forming sub-regions S_1, S_2, S_3 and S_4.



Sub-region Monitored	Max Thickness		C_d	$C_d \leq \theta$	Next Action
	Critical	Non-Critical			
S_1	93	85, 78, 81	8	Yes	*Quadrisect*
S_2	97	83, 79, 86	11	No	*Retain*
S_3	95	76, 73, 80	15	No	*Retain*
S_4	94	88, 84, 89	5	Yes	*Quadrisect*

Working Model for High Yield …. Stage 3

- Stage 3 – Inside sub-regions $\{S_{11}, S_{12}, S_{13}, S_{14}\}$ & $\{S_{41}, S_{42}, S_{43}, S_{44}\}$.

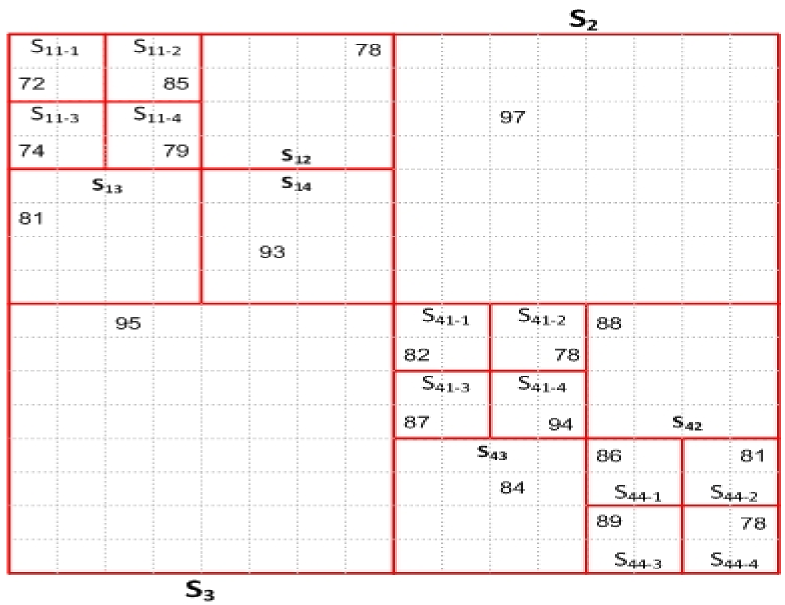


Sub-region Monitored	Max Thickness		C_d	$C_d \leq \theta$	Next Action
	Critical	Non-Critical			
S_{11}	85	72, 74, 79	6	Yes	*Quadrisect*
S_{12}	78	63, 65, 60	13	No	*Retain*
S_{13}	81	70, 68, 66	11	No	*Retain*
S_{14}	93	79, 77, 75	16	No	*Retain*

Sub-region Monitored	Max Thickness		C_d	$C_d \leq \theta$	Next Action
	Critical	Non-Critical			
S_{41}	94	82, 78, 87	7	Yes	*Quadrisect*
S_{42}	88	75, 73, 67	13	No	*Retain*
S_{43}	84	71, 66, 69	11	No	*Retain*
S_{44}	89	86, 81, 78	3	Yes	*Quadrisect*

Working Model for High Yield …. After Stage 3

- After Stage 3 in the HAQ algorithm, the setup will look like -



- Stage 3, the chip is covered by *19 basic sub-regions*.

- Further clustering based on thickness variations in new sub-regions.

Computing Low Yield using HAQ

- Clustering based on minimum thickness variations in sub-regions.

Comparing HAQ and PCC approaches

HAQ Approach

- Heterogeneous cluster sizes
- Clustering based on variations and sensitivity inside sub-regions
- No. of Clusters in working model –
  - Stage-1 → 4
  - Stage-2 →10
  - Stage-3 →19

PCC Approach

- Homogeneous cluster sizes
- No importance for sensitivity in variations for clustering
- No. of Clusters in each stage of the working model
  - Stage-1 → 4
  - Stage-2 → 16
  - Stage-3 → 64

Presentation Outline

- ✓ What is Yield?
- ✓ Difficulties in Yield Prediction
- ✓ Previous Research
- ✓ Proposed Research
- ➢ **Simulation Results**
- ➢ Conclusion

Simulation Results

- Experiments simulated –

  1. Monte Carlo (MC) Simulations
  2. PCC method
  3. OPCA method
  4. HAQ method

- Yield evaluated for three cases of correlation – $(-\alpha \times 10^{-5} x) + 0.9958$

  where $\alpha = \{2, 3, 4\}$ and x– distance between centres of different tiles.

- Simulation Inputs –

  1. Input thickness –

     Mean thickness value – 0.3580 μm

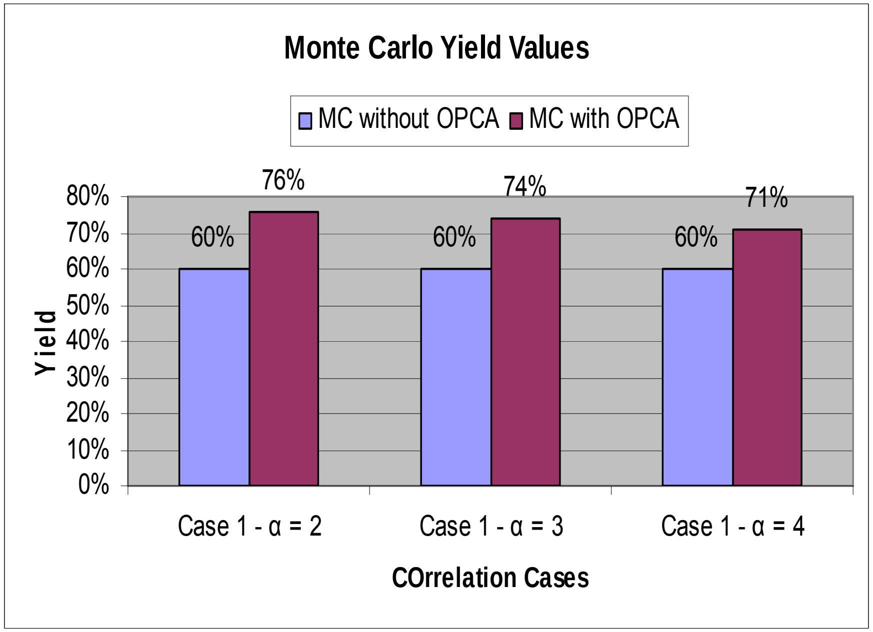
     Upper thickness limit – 0.4580 μm

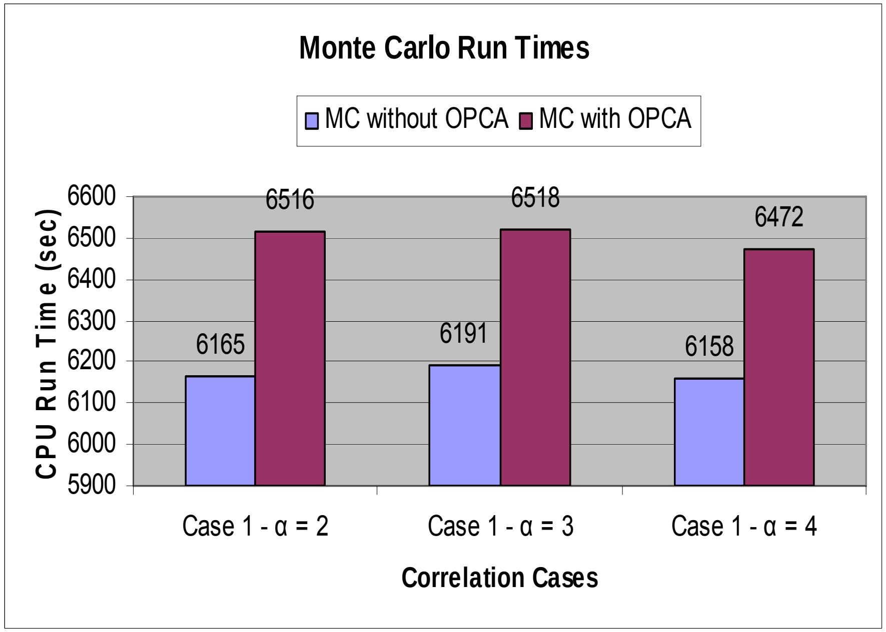
     Lower thickness limit – 0.2580 μm

     Standard deviation – 0.02 μm

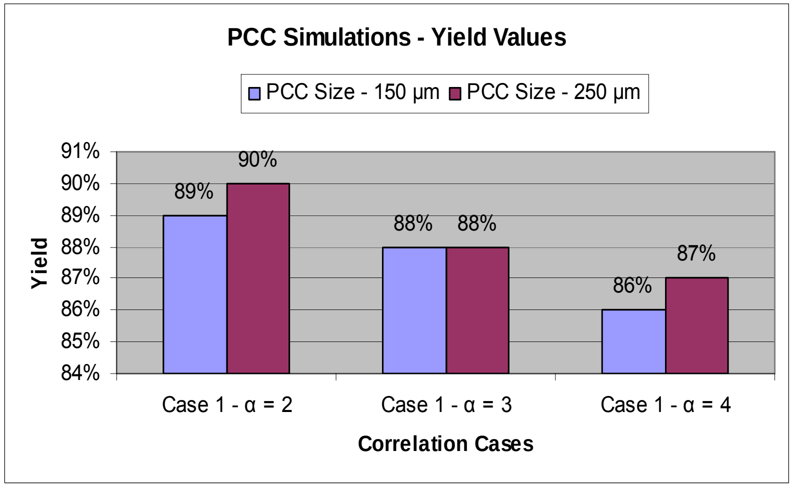
Monte Carlo Simulations

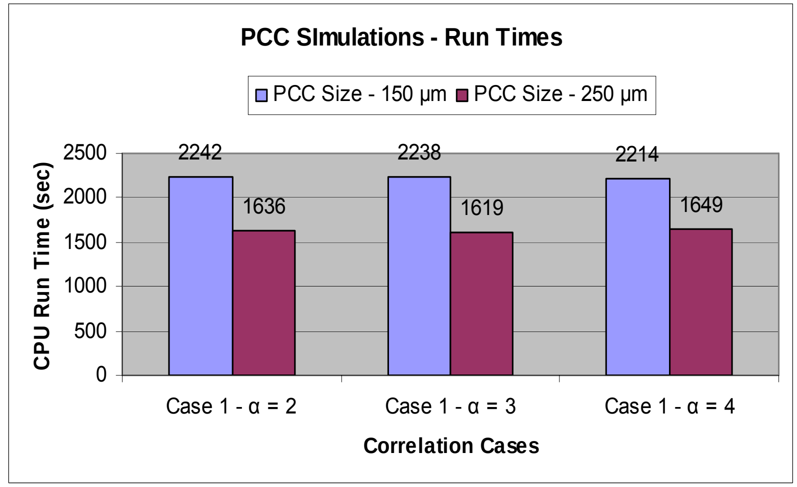
- Correlation Equation: $-3\times10^{-5}x+0.9958$
- Initial seed = 5





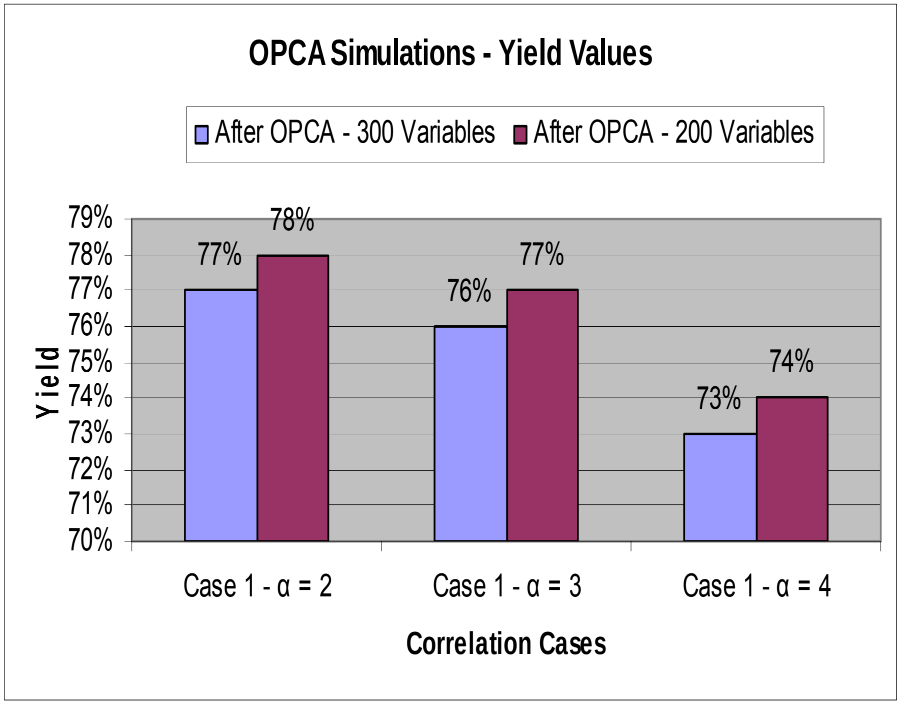
PCC Simulations

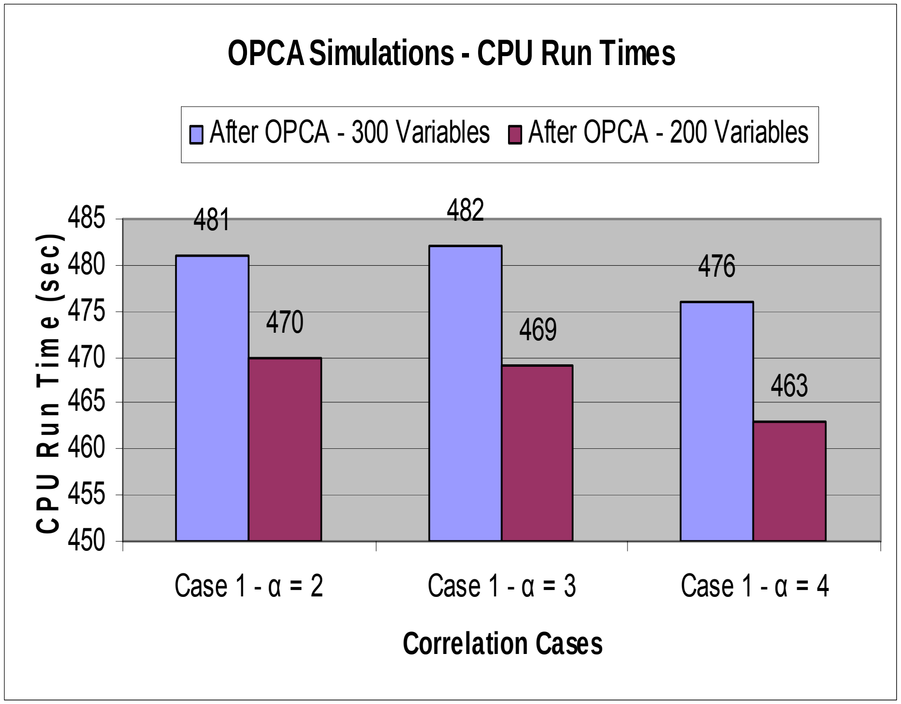




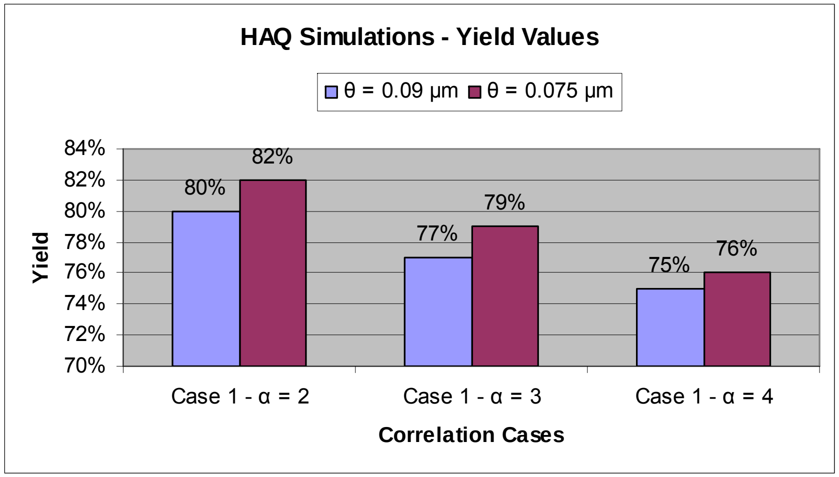
Correlation Equation	PCC Size	No. of Variables
$-2\times10^{-5}x+0.9958$	150 µm 250 µm	431/435 305/310
$-3\times10^{-5}x+0.9958$	150 µm 250 µm	432/427 305/310
$-4\times10^{-5}x+0.9958$	150 µm 250 µm	429/425 307/308

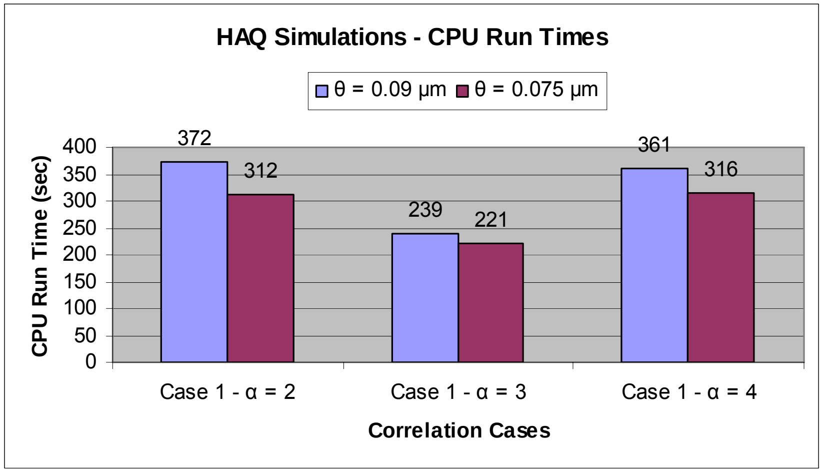
OPCA Simulations





HAQ Simulations





Correlation Equation	θ	No. of Variables
$-2\times10^{-5}x+0.9958$	0.09 µm 0.075 µm	175/178 153/155
$-3\times10^{-5}x+0.9958$	0.09 µm 0.075 µm	80/79 61/61
$-4\times10^{-5}x+0.9958$	0.09 µm 0.075 µm	172/170 148/143

Observations in Results

- Monte Carlo without OPCA –

  Neglecting correlation under-estimates yield.

- OPCA –

  Less variable reduction → better accuracy, yield is closer to Monte Carlo.

- PCC –

  Larger PCC sizes ➔ more reduction ➔ over-estimated yield value

  Smaller PCC sizes ➔ improves accuracy in yield ➔ longer run time

- HAQ –

  Higher threshold values ➔ less reduction (fine-grained grid)

  ➔ improved accuracy

  Smaller threshold values ➔ over-estimated yield

Comparisons in Results

- Comparing yield accuracy and algorithm run time –

Correlation Equation	Method	Yield Error	Speedup
$-3\times10^{-5}x+0.9958$	PCC OPCA HAQ	18.9% 2.7% 4.1%	1x 4.6x 9.4x
$-4\times10^{-5}x+0.9958$	PCC OPCA HAQ	21.1% 2.8% 5.6%	1x 4.7x 6.2x
$-2\times10^{-5}x+0.9958$	PCC OPCA HAQ	17.1% 1.3% 5.3%	1x 4.7x 6x

Presentation Outline

- ✓ What is Yield?
- ✓ Difficulties in Yield Prediction
- ✓ Previous Research
- ✓ Proposed Research
- ✓ Simulation Results
- ➢ **Conclusion**

Conclusion

- Yield prediction is complex –
  1. Large number of locations monitored
  2. Partial & independent correlations between locations
- New methods used in yield prediction –
  1. Orthogonal Principal Component Analysis
  2. Hierarchical Adaptive Quadrisection
- Both reduce complexity & have less impact on Yield Accuracy.

Scope for Future Work

Extend same methods to predict timing yield in sequential circuits.

Thank You