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 –
   $$p(n) = \mu(n) + \epsilon(n)$$
   where $\mu(n)$ – systematic intra-die variations
   $\epsilon(n)$ – 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.
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 -

\[ Y = \int_{L}^{U} \int_{L}^{U} \cdots \int_{L}^{U} \phi (p) \cdot dp \cdot dp \cdots dp \quad \cdots(1) \]

\[ \phi (p) = \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} (p - \mu)^T \Sigma^{-1} (p - \mu) \quad \cdots(2) \]

Where \( \Sigma \) - covariance matrix for the \( n \) variables – \( \{p_1, p_2, \ldots, p_n\} \)

\( U, L \) & \( \mu \) - upper and lower thickness limits, & mean thickness value.

Yield equation (1) can be decomposed as –

\[ \text{Yield} = Y_U + Y_L - 1 \quad \cdots(3) \]

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
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 $PCC$s until no tiles are left uncovered by $PCC$s.
6. Centers of $PCC$s ($MAX_1, \ldots, MAX_m$) form reduced set of variables.
7. Use Genz algorithm to compute yield.
Let the setup look like this after reduction →

Reduction from 90 tiles to 14 variables (the centres of PCCs – MAX₁, …., MAX₁₄.)

PCCs are formed in a sequence –
MAX₁ – CIRCLE₁,
MAX₂ – CIRCLE₂,
…………………..,
MAX₁₃ – CIRCLE₁₃,
MAX₁₄ – CIRCLE₁₄.

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 $\mathbf{p}$ be metal thicknesses at $n$ locations –
  $$\mathbf{p} = (p_1, p_2, \ldots, p_n)^T$$

- This vector can be decomposed as follows –
  $$p_i = \mu_i + \delta_i$$
  and
  $$\mu_i = \mu + \Delta_i$$

where $\mu$ – nominal value
$\Delta_i$ – systematic variation
$\delta_i$ – random variation
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 –
\[
\vec{\delta} = \{\delta_1, \delta_2, \ldots, \delta_n\}^T
\]
\[\ldots (1)\]

Let \( \Gamma_{nxn} \) and \( \Sigma_{nxn} \) be the corresponding correlation and covariance matrices. Let \( \sigma_i^2 \) be the variance.

\[
\Gamma(\vec{\delta}) = (\Gamma_{ij})_{nxn} \quad \text{and} \quad \Sigma(\vec{\delta}) = \Gamma(\vec{\delta})_{nxn} \cdot \sigma_i \cdot \sigma_j
\]
\[\ldots (2)\]
Using Eigenvalue Decomposition

- Re-express covariance matrix using Eigenvalue Decomposition -
  \[ \Sigma(\delta) = Q \cdot \Lambda(\delta) \cdot Q^T \]  
  ....(3)
  where \( \Lambda(\delta) \) - eigenvalue (diagonal) matrix
  \( Q \) - corresponding eigenvector matrix

- The diagonal matrix \( \Lambda(\delta)_{n \times n} \) will look like -
  \[
  \Lambda(\delta) = \begin{pmatrix}
  \lambda_1 & 0 & \cdots & 0 \\
  0 & \lambda_2 & \cdots & 0 \\
  0 & 0 & \cdots & 0 \\
  0 & 0 & \cdots & \lambda_n \\
  0 & 0 & \cdots & \lambda_n
  \end{pmatrix}
  \]
  ....(4)
  such that \( \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \)

- Eigenvalue decomposition gives dominant directions in covariance relationship between a correlated set of variables.
Mapping into a New Set of Variables

- Let $\mathbf{E}_{n \times 1}$ be the new set of uncorrelated variables such that:
  \[ \mathbf{\delta} = \mathbf{B} \cdot \mathbf{\varepsilon} \] ....(5)
- Without loss of generality, assume $\mathbf{B}$ follows a Gaussian Distribution:
  \[ \mu(\mathbf{\varepsilon}) = 0 \quad \text{&} \quad \Lambda(\mathbf{\varepsilon}) = I \] ....(6)
- The matrices $\mathbf{\delta}_{n \times 1}$ and $\mathbf{E}_{n \times 1}$ are related as follows:
  \[ \Lambda(\mathbf{\delta}) = \mathbf{J} \cdot \Lambda(\mathbf{\varepsilon}) \cdot \mathbf{J}^T \] ....(7)
  where
  \[ \mathbf{J} = \begin{pmatrix} \sqrt{\lambda_1} & 0 & \cdots & 0 \\ 0 & \sqrt{\lambda_2} & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & \sqrt{\lambda_n} \end{pmatrix} \]
- Transforming through an Orthogonal Base – Let $\mathbf{B}$ be the mapping matrix:
  \[ \mathbf{B} = \mathbf{Q} \cdot \mathbf{J} \] ....(8)
Correspondingly, we have –
\[ \delta = B \cdot \varepsilon = Q \cdot J \cdot \varepsilon \] ....(9)

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

\[ \sum(\delta) = Q \cdot \Lambda(\delta) \cdot Q^T = Q \cdot J \cdot \Lambda(\varepsilon) \cdot (Q \cdot J)^T \] ....(10)

Reducing the no. of uncorrelated variables –
1. After reduction, if we have \( k \) variables, then matrices \( \Lambda(\delta) \) and \( J_{k \times k} \) are –

\[ \Lambda(\delta) = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & \lambda_k \end{pmatrix} \quad \& \quad J = \begin{pmatrix} \sqrt{\lambda_1} & 0 & \cdots & 0 \\ 0 & \sqrt{\lambda_2} & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & \sqrt{\lambda_k} \end{pmatrix} \] ....(11)

2. The corresponding sizes of matrices \( B \) and \( Q \) become \( n \times k \), thus giving reduction.
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 $\theta$ decides possibility of clustering.
- Threshold $\theta$ tells on variations in thickness of tiles in a sub-region.
Stage 1:– Sub-region S covers the entire chip. Let $\theta$ be 10.

<table>
<thead>
<tr>
<th>Sub-region Monitored</th>
<th>Max Thickness</th>
<th>$C_d$</th>
<th>$C_d \leq \theta$</th>
<th>Next Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>97</td>
<td>93, 95, 94</td>
<td>2</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Stage 2 – After forming sub-regions $S_1$, $S_2$, $S_3$ and $S_4$. 

<table>
<thead>
<tr>
<th>Sub-region Monitored</th>
<th>Max Thickness</th>
<th>$C_d$</th>
<th>$C_d \leq \theta$</th>
<th>Next Action</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Critical</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$S_1$</td>
<td>93</td>
<td>85, 78, 81</td>
<td>8</td>
<td>Yes</td>
</tr>
<tr>
<td>$S_2$</td>
<td>97</td>
<td>83, 79, 86</td>
<td>11</td>
<td>No</td>
</tr>
<tr>
<td>$S_3$</td>
<td>95</td>
<td>76, 73, 80</td>
<td>15</td>
<td>No</td>
</tr>
<tr>
<td>$S_4$</td>
<td>94</td>
<td>88, 84, 89</td>
<td>5</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Stage 3 – Inside sub-regions \{S_{11}, S_{12}, S_{13}, S_{14}\} & \{S_{41}, S_{42}, S_{43}, S_{44}\}.

<table>
<thead>
<tr>
<th>Sub-region Monitored</th>
<th>Max Thickness</th>
<th>(C_d)</th>
<th>(C_d \leq \theta)</th>
<th>Next Action</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Critical</td>
<td>Non-Critical</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(S_{11})</td>
<td>85</td>
<td>72, 74, 79</td>
<td>6</td>
<td>Yes</td>
</tr>
<tr>
<td>(S_{12})</td>
<td>78</td>
<td>63, 65, 60</td>
<td>13</td>
<td>No</td>
</tr>
<tr>
<td>(S_{13})</td>
<td>81</td>
<td>70, 68, 66</td>
<td>11</td>
<td>No</td>
</tr>
<tr>
<td>(S_{14})</td>
<td>93</td>
<td>79, 77, 75</td>
<td>16</td>
<td>No</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sub-region Monitored</th>
<th>Max Thickness</th>
<th>(C_d)</th>
<th>(C_d \leq \theta)</th>
<th>Next Action</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Critical</td>
<td>Non-Critical</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(S_{41})</td>
<td>94</td>
<td>82, 78, 87</td>
<td>7</td>
<td>Yes</td>
</tr>
<tr>
<td>(S_{42})</td>
<td>88</td>
<td>75, 73, 67</td>
<td>13</td>
<td>No</td>
</tr>
<tr>
<td>(S_{43})</td>
<td>84</td>
<td>71, 66, 69</td>
<td>11</td>
<td>No</td>
</tr>
<tr>
<td>(S_{44})</td>
<td>89</td>
<td>86, 81, 78</td>
<td>3</td>
<td>Yes</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>HAQ Approach</th>
<th>PCC Approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heterogeneous cluster sizes</td>
<td>Homogeneous cluster sizes</td>
</tr>
<tr>
<td>Clustering based on variations and sensitivity inside sub-regions</td>
<td>No importance for sensitivity in variations for clustering</td>
</tr>
<tr>
<td>No. of Clusters in working model –</td>
<td>No. of Clusters in each stage of the working model</td>
</tr>
<tr>
<td>Stage–1 → 4</td>
<td>Stage–1 → 4</td>
</tr>
<tr>
<td>Stage–2 → 10</td>
<td>Stage–2 → 16</td>
</tr>
<tr>
<td>Stage–3 → 19</td>
<td>Stage–3 → 64</td>
</tr>
</tbody>
</table>
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 \(\mu\)m
     Upper thickness limit – 0.4580 \(\mu\)m
     Lower thickness limit – 0.2580 \(\mu\)m
     Standard deviation – 0.02 \(\mu\)m
Monte Carlo Simulations

- Correlation Equation: \(-3 \times 10^{-5} x + 0.9958\)
- Initial seed = 5
PCC Simulations

Correlation Equation | PCC Size | No. of Variables |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$- 2 \times 10^{-5} x + 0.9958$</td>
<td>150 µm</td>
<td>431/435</td>
</tr>
<tr>
<td></td>
<td>250 µm</td>
<td>305/310</td>
</tr>
<tr>
<td>$- 3 \times 10^{-5} x + 0.9958$</td>
<td>150 µm</td>
<td>432/427</td>
</tr>
<tr>
<td></td>
<td>250 µm</td>
<td>305/310</td>
</tr>
<tr>
<td>$- 4 \times 10^{-5} x + 0.9958$</td>
<td>150 µm</td>
<td>429/425</td>
</tr>
<tr>
<td></td>
<td>250 µm</td>
<td>307/308</td>
</tr>
</tbody>
</table>
OPCA Simulations - Yield Values

OPCA Simulations - CPU Run Times

Case 1 - $\alpha = 2$
Case 1 - $\alpha = 3$
Case 1 - $\alpha = 4$

After OPCA - 300 Variables
After OPCA - 200 Variables
HAQ Simulations

HAQ Simulations - Yield Values

<table>
<thead>
<tr>
<th>Correlation Cases</th>
<th>Yield 70%</th>
<th>Yield 72%</th>
<th>Yield 74%</th>
<th>Yield 76%</th>
<th>Yield 78%</th>
<th>Yield 80%</th>
<th>Yield 82%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case 1 - $\alpha$ = 2</td>
<td>80%</td>
<td>77%</td>
<td>75%</td>
<td>76%</td>
<td>78%</td>
<td>80%</td>
<td>82%</td>
</tr>
<tr>
<td>Case 1 - $\alpha$ = 3</td>
<td>77%</td>
<td>79%</td>
<td>76%</td>
<td>78%</td>
<td>80%</td>
<td>82%</td>
<td></td>
</tr>
<tr>
<td>Case 1 - $\alpha$ = 4</td>
<td>75%</td>
<td>76%</td>
<td>78%</td>
<td>80%</td>
<td>82%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

HAQ Simulations - CPU Run Times

<table>
<thead>
<tr>
<th>Correlation Cases</th>
<th>CPU Run Time (sec) 372</th>
<th>CPU Run Time (sec) 239</th>
<th>CPU Run Time (sec) 361</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case 1 - $\alpha$ = 2</td>
<td>372</td>
<td>312</td>
<td>361</td>
</tr>
<tr>
<td>Case 1 - $\alpha$ = 3</td>
<td>239</td>
<td>221</td>
<td>361</td>
</tr>
<tr>
<td>Case 1 - $\alpha$ = 4</td>
<td>221</td>
<td>316</td>
<td>361</td>
</tr>
</tbody>
</table>

Correlation Equation

<table>
<thead>
<tr>
<th>Correlation Equation</th>
<th>$\theta$</th>
<th>No. of Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>$-2 \times 10^{-5} x + 0.9958$</td>
<td>$0.09 \mu m$</td>
<td>175/178</td>
</tr>
<tr>
<td></td>
<td>$0.075 \mu m$</td>
<td>153/155</td>
</tr>
<tr>
<td>$-3 \times 10^{-5} x + 0.9958$</td>
<td>$0.09 \mu m$</td>
<td>80/79</td>
</tr>
<tr>
<td></td>
<td>$0.075 \mu m$</td>
<td>61/61</td>
</tr>
<tr>
<td>$-4 \times 10^{-5} x + 0.9958$</td>
<td>$0.09 \mu m$</td>
<td>172/170</td>
</tr>
<tr>
<td></td>
<td>$0.075 \mu m$</td>
<td>148/143</td>
</tr>
</tbody>
</table>
Observations in Results

- Monte Carlo without OPCA –
  - Neglecting correlation under-estimates yield.

- OPCA –
  - Less variable reduction $\rightarrow$ better accuracy, yield is closer to Monte Carlo.

- PCC –
  - Larger PCC sizes $\rightarrow$ more reduction $\rightarrow$ over-estimated yield value
  - Smaller PCC sizes $\rightarrow$ improves accuracy in yield $\rightarrow$ longer run time

- HAQ –
  - Higher threshold values $\rightarrow$ less reduction (fine-grained grid)
    $\rightarrow$ improved accuracy
  - Smaller threshold values $\rightarrow$ over-estimated yield
Comparisons in Results

- Comparing yield accuracy and algorithm run time -

<table>
<thead>
<tr>
<th>Correlation Equation</th>
<th>Method</th>
<th>Yield Error</th>
<th>Speedup</th>
</tr>
</thead>
<tbody>
<tr>
<td>$-3 \times 10^{-5} x + 0.9958$</td>
<td>PCC OPCA HAQ</td>
<td>18.9% 2.7% 4.1%</td>
<td>1x 4.6x 9.4x</td>
</tr>
<tr>
<td>$-4 \times 10^{-5} x + 0.9958$</td>
<td>PCC OPCA HAQ</td>
<td>21.1% 2.8% 5.6%</td>
<td>1x 4.7x 6.2x</td>
</tr>
<tr>
<td>$-2 \times 10^{-5} x + 0.9958$</td>
<td>PCC OPCA HAQ</td>
<td>17.1% 1.3% 5.3%</td>
<td>1x 4.7x 6x</td>
</tr>
</tbody>
</table>
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