A Fast Wafer-Level Spatial Variation Modeling Algorithm for Test Cost Reduction of Analog/RF Circuits

Hugo Gonçalves[1,2], Xin Li[1], Miguel Correia[2] and Vitor Tavares[2]
[1]ECE Department, Carnegie Mellon University, USA
[2]Faculdade de Engenharia, Universidade do Porto, Portugal

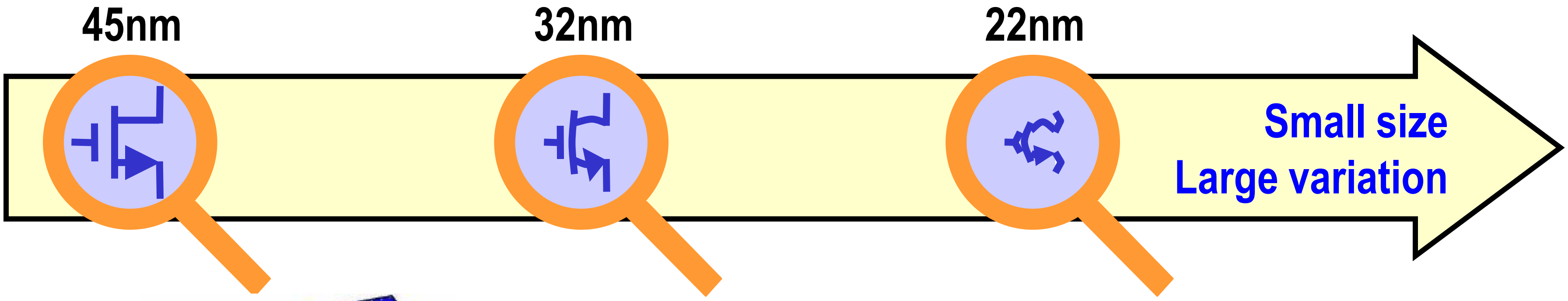
FAC '14

09/07/2014

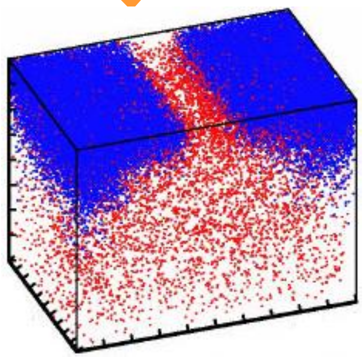
Outline

- **Motivation and background**
  - Virtual probe (VP)
- **Proposed approach**
  - Dual Augmented Lagrangian method (DALM)
  - Two-pass test flow
- **Experimental results**
- **Conclusions**

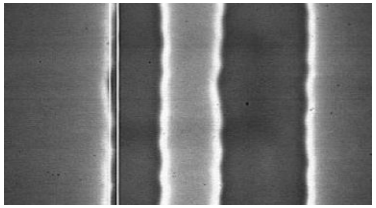
Process Variation



Small size
Large variation



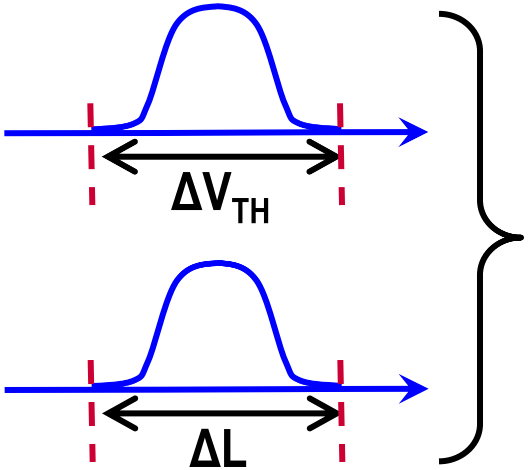
Doping fluctuation



Line edge roughness

$$V_{TH} = \overline{V}_{TH} + \Delta V_{TH}$$

$$L = \overline{L} + \Delta L$$



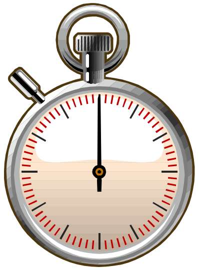
Parametric variations

Wafer Probe Test

- **Multiple test items must be measured for each die**

- **An industrial example of dual radio RF transceiver**
  - ~1 second testing time per die
  - ~6500 dies per wafer
  - ~ 2 hour testing time per wafer

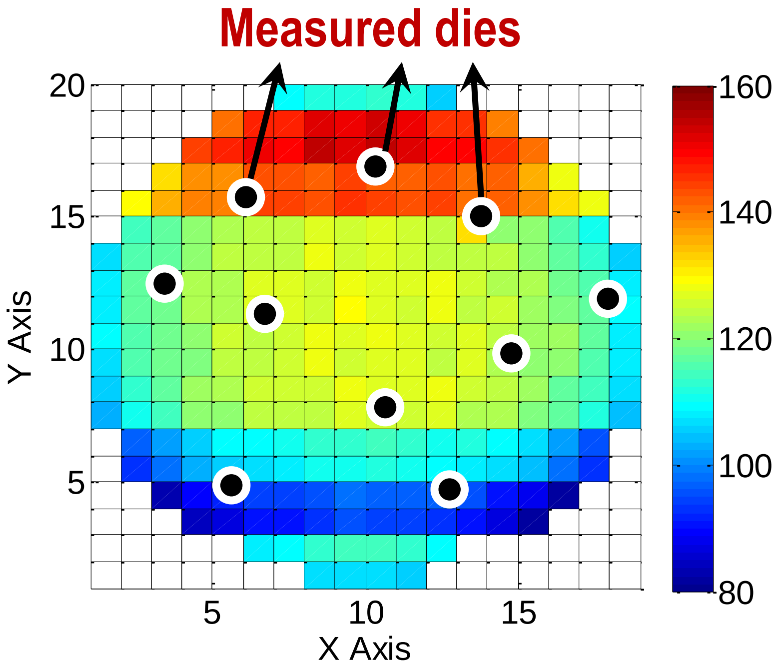
- **Measuring all test items is time-consuming**



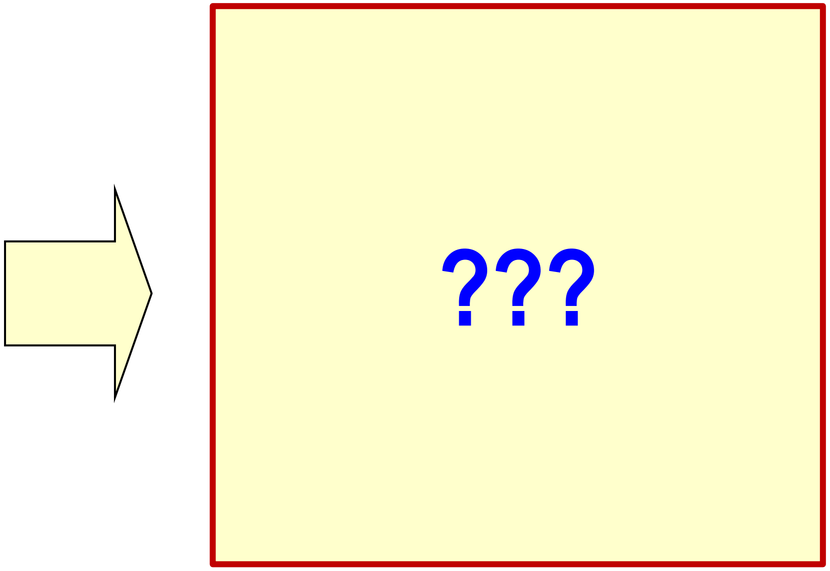
~2h per wafer

Test Cost Reduction by Spatial Variation Modeling

- Measure a small number of dies at selected spatial locations
- Recover the full wafer map by statistical algorithm



Measured delay values (normalized)
from 282 industrial chips

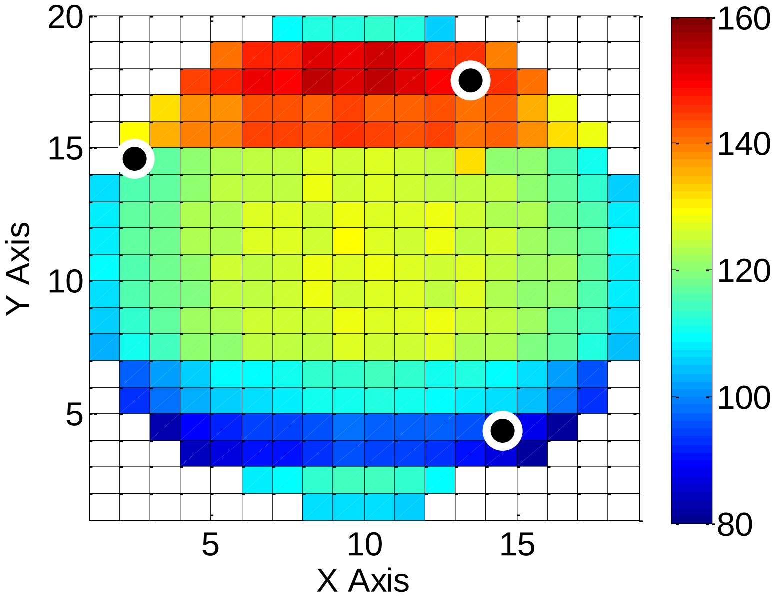


Recovered wafer
map

- [Chang11], [Kupp12], [Huang13], [Hsu13], etc.

Virtual Probe (VP)

- List a set of linear equations based on measurement data



Measured delay values (normalized)
from 282 industrial chips

DCT basis function

$$\begin{bmatrix} \times \\ \times \\ \times \end{bmatrix} = \begin{bmatrix} \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \end{bmatrix} \cdot \begin{bmatrix} \times \\ \times \\ \times \\ \times \\ \times \end{bmatrix}$$

Performance measurement

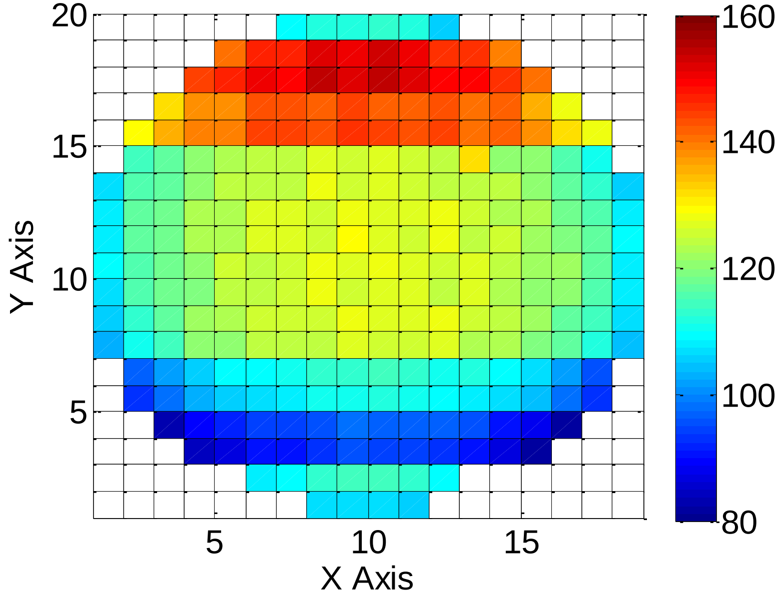
DCT coefficients

$$f(x,y) = \sum_{i=1}^{N} \left(b_i(x,y) \cdot \alpha_i \right)$$

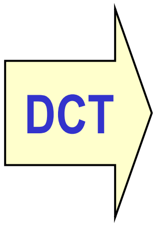
Results in an under-determined linear equation, since we have less measurements than unknown DCT coefficients

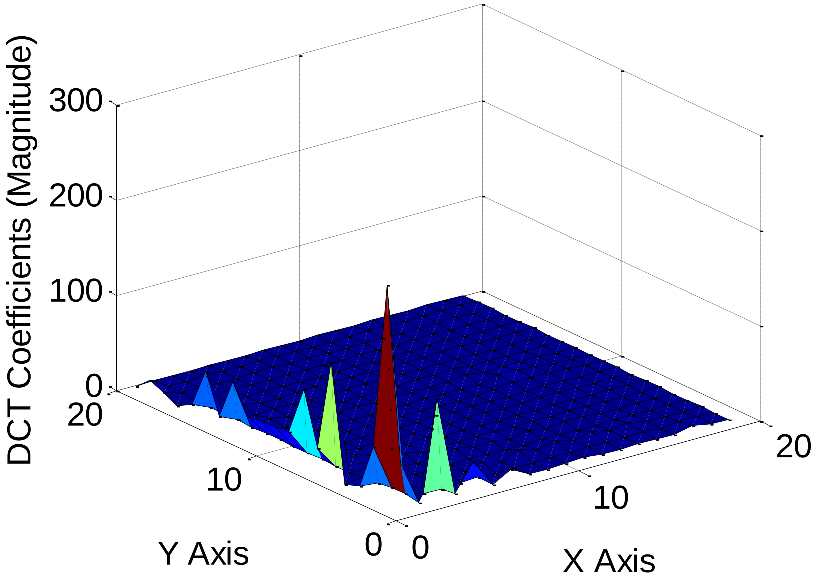
Virtual Probe (VP)

- Additional information is required to uniquely solve under-determined linear equation



Measured delay values (normalized) from 282 industrial chips



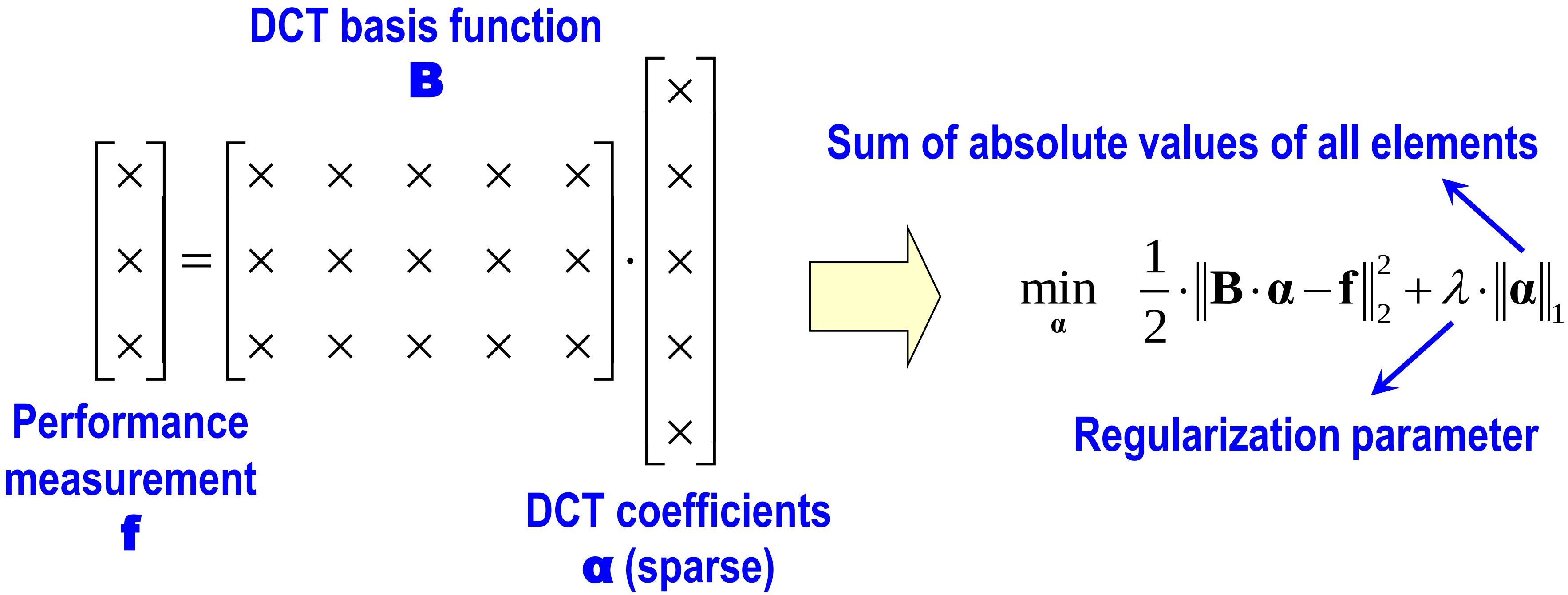


DCT coefficients (sparse)

If process variations are spatially correlated wafer maps show sparse patterns in frequency domain

Virtual Probe (VP)

- **Solve sparse DCT coefficients by L1-norm regularization**
  - DCT coefficients can be uniquely determined from a small number of measurements



DCT basis function

B

$$\begin{bmatrix}\times\\\times\\\times\end{bmatrix}=\begin{bmatrix}\times&\times&\times&\times&\times\\\times&\times&\times&\times&\times\\\times&\times&\times&\times&\times\end{bmatrix}\cdot\begin{bmatrix}\times\\\times\\\times\\\times\\\times\end{bmatrix}$$

Performance measurement

f

DCT coefficients

α (sparse)

Sum of absolute values of all elements

$$\min_{\boldsymbol{\alpha}} \quad \frac{1}{2}\cdot\left\|\mathbf{B}\cdot\boldsymbol{\alpha}-\mathbf{f}\right\|_2^2+\lambda\cdot\left\|\boldsymbol{\alpha}\right\|_1$$

Regularization parameter

Virtual Probe (VP)

Linear regression problem: $$\min_{\boldsymbol{\alpha}} \quad \frac{1}{2} \cdot \left\| \mathbf{B} \cdot \boldsymbol{\alpha} - \mathbf{f} \right\|_2^2 + \lambda \cdot \left\| \boldsymbol{\alpha} \right\|_1$$

- There is no closed-form solution
- A standard interior-point solver is not computationally efficient
- We aim to develop an application-specific solver to reduce computational time and, hence, testing cost

Outline

- **Motivation and background**
  - Virtual probe (VP)
- **Proposed approach**
  - Dual Augmented Lagrangian method (DALM)
  - Two-pass test flow
- **Experimental results**
- **Conclusions**

Dual Problem

Primal problem: $$\min_{\boldsymbol{\alpha}} \quad \frac{1}{2}\cdot\left\|\mathbf{B}\cdot\boldsymbol{\alpha}-\mathbf{f}\right\|_2^2+\lambda\cdot\left\|\boldsymbol{\alpha}\right\|_1$$

- **Key idea: form a dual problem to reduce the number of unknowns**

- **Primal problem**
  - \# of unknowns = # of DCT coefficients $\geq$ # of dies

- **Dual problem**
  - \# of unknowns = # of measurements

- **Since we have substantially less measurements than unknowns, solving the dual problem is significantly more efficient**

Strong Duality

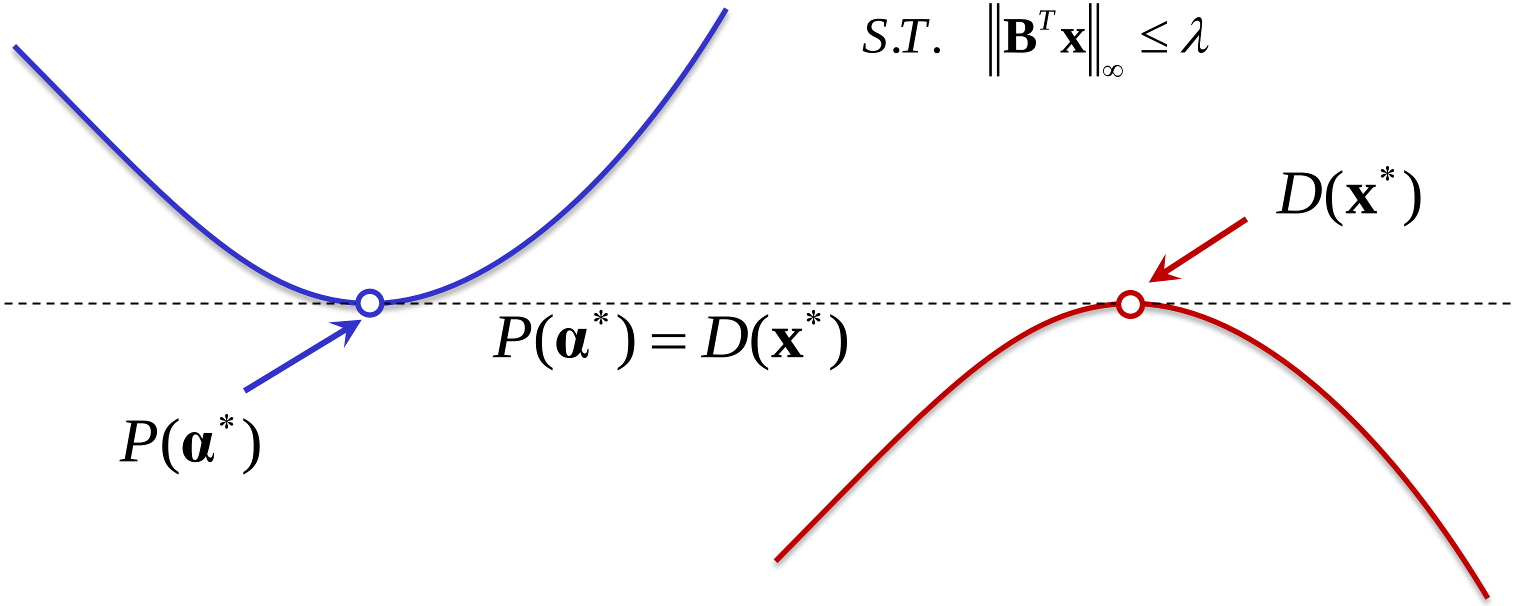
Primal function

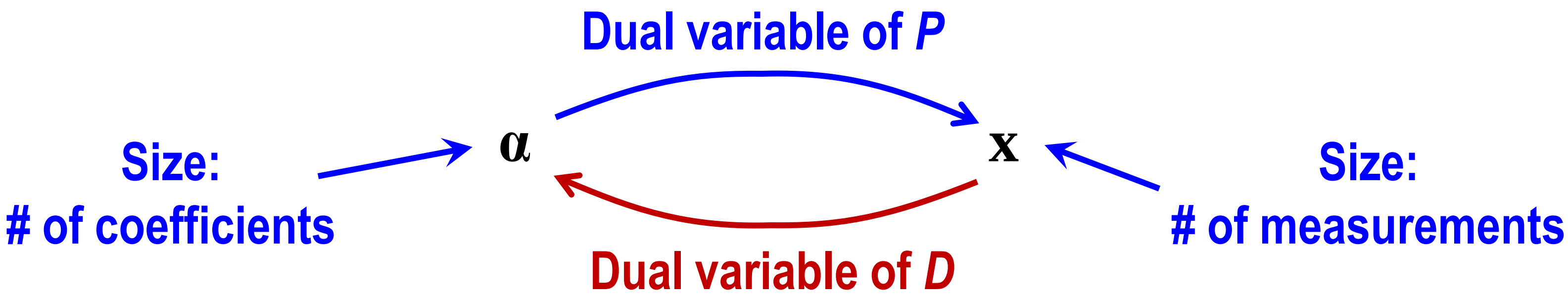
$$P(\boldsymbol{\alpha}) = \frac{1}{2}\left\|\mathbf{B}\boldsymbol{\alpha} - \mathbf{f}\right\|_2^2 + \lambda\left\|\boldsymbol{\alpha}\right\|_1$$

Dual function

$$D(\mathbf{x}) = -\frac{1}{2}\left\|\mathbf{x} - \mathbf{f}\right\|_2^2 + \frac{1}{2}\left\|\mathbf{f}\right\|_2^2$$

$$S.T. \quad \left\|\mathbf{B}^T\mathbf{x}\right\|_\infty \le \lambda$$




Dual variable of *P*

Size:
# of coefficients

$\boldsymbol{\alpha}$

$\mathbf{x}$

Size:
# of measurements

Dual variable of *D*

Dual Augmented Lagrangian

- **Define an auxiliary variable z to form an equality constraint**

Dual problem

$$\max_{\mathbf{x}} \quad D(\mathbf{x}) = -\frac{1}{2}\|\mathbf{x}-\mathbf{f}\|_2^2 + \frac{1}{2}\|\mathbf{f}\|_2^2$$
$$S.T. \quad \|\mathbf{B}^T\mathbf{x}\|_\infty \le \lambda$$

Dual problem w/ equality constraint

$$\max_{\mathbf{x},\mathbf{z}} \quad D(\mathbf{x}) = -\frac{1}{2}\|\mathbf{x}-\mathbf{f}\|_2^2 + \frac{1}{2}\|\mathbf{f}\|_2^2$$
$$S.T. \quad \mathbf{z} = \mathbf{B}^T\mathbf{x}$$
$$\|\mathbf{z}\|_\infty \le \lambda$$

- **Solve the augmented Lagrangian of the dual problem**

$$\max_{\mathbf{x},\mathbf{z}} \quad L_A(\mathbf{x},\mathbf{z},\boldsymbol{\alpha}) = -\frac{1}{2}\cdot\|\mathbf{x}-\mathbf{f}\|_2^2 + \frac{1}{2}\cdot\|\mathbf{f}\|_2^2 + \boldsymbol{\alpha}^T\cdot\left(\mathbf{z}-\mathbf{B}^T\mathbf{x}\right) - \frac{\eta}{2}\cdot\|\mathbf{z}-\mathbf{B}^T\mathbf{x}\|_2^2 - \delta_\infty^\lambda(\mathbf{z})$$

Primal variable
size = # of DCT coefficients

$$\delta_\infty^\lambda(\mathbf{z}) = \begin{cases} 0 & , \|\mathbf{z}\|_\infty \le \lambda \\ +\infty & , \|\mathbf{z}\|_\infty > \lambda \end{cases}$$

Alternating Direction Method

Augmented Lagrangian

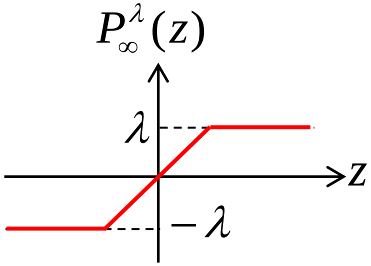
$$\max_{\mathbf{x},\mathbf{z}} \quad L_A\left(\mathbf{x},\mathbf{z},\boldsymbol{\alpha}\right) = -\frac{1}{2}\cdot\left\|\mathbf{x}-\mathbf{f}\right\|_2^2 + \frac{1}{2}\cdot\left\|\mathbf{f}\right\|_2^2 + \boldsymbol{\alpha}^T\cdot\left(\mathbf{z}-\mathbf{B}^T\mathbf{x}\right) - \frac{\eta}{2}\cdot\left\|\mathbf{z}-\mathbf{B}^T\mathbf{x}\right\|_2^2 - \delta_\infty^\lambda\left(\mathbf{z}\right)$$

- Solve optimization with alternating direction method [Yang10]

Optimality conditions

Variable update

$$\frac{\partial L_A\left(\mathbf{x}^{(k)},\mathbf{z},\boldsymbol{\alpha}^{(k)}\right)}{\partial \mathbf{z}} = 0 \quad \Rightarrow \quad \mathbf{z}^{(k+1)} = P_\infty^\lambda\left(\boldsymbol{\alpha}^{(k)}/\eta + \mathbf{B}^T\mathbf{x}^{(k)}\right)$$



$$\frac{\partial L_A\left(\mathbf{x},\mathbf{z}^{(k+1)},\boldsymbol{\alpha}^{(k)}\right)}{\partial \mathbf{x}} = 0 \quad \Rightarrow \quad \mathbf{x}^{(k+1)} = \left(\mathbf{I}+\eta\cdot\mathbf{B}\mathbf{B}^T\right)^{-1}\left(\mathbf{f}-\mathbf{B}\boldsymbol{\alpha}^{(k)}+\eta\cdot\mathbf{B}\mathbf{z}^{(k+1)}\right)$$

AL step $\Rightarrow$ $\boldsymbol{\alpha}^{(k+1)} = \boldsymbol{\alpha}^{(k)} + \eta\cdot\left(\mathbf{B}^T\mathbf{x}^{(k+1)} - \mathbf{z}^{(k+1)}\right)$

Fast Matrix Inverse

$$\mathbf{x}^{(k+1)} = \underline{\left(\mathbf{I} + \eta \cdot \mathbf{B}\mathbf{B}^{T}\right)}^{-1}\left(\mathbf{f} - \mathbf{B}\boldsymbol{\alpha}^{(k)} + \eta \cdot \mathbf{B}\mathbf{z}^{(k+1)}\right)$$

- Since DCT basis functions are used, we have

$$\mathbf{B}\mathbf{B}^{T} = \mathbf{I}$$

- Hence, we do not need to explicitly calculate matrix inverse

$$\mathbf{x}^{(k+1)} = \frac{1}{1+\eta} \cdot \left(\mathbf{f} - \mathbf{B}\boldsymbol{\alpha}^{(k)} + \eta \cdot \mathbf{B}\mathbf{z}^{(k+1)}\right)$$

Fast Matrix-Vector Multiplication

$$\mathbf{z}^{(k+1)} = P_{\infty}^{\lambda}\left(\boldsymbol{\alpha}^{(k)}/\eta + \underline{\mathbf{B}^{T}\mathbf{x}^{(k)}}\right)$$

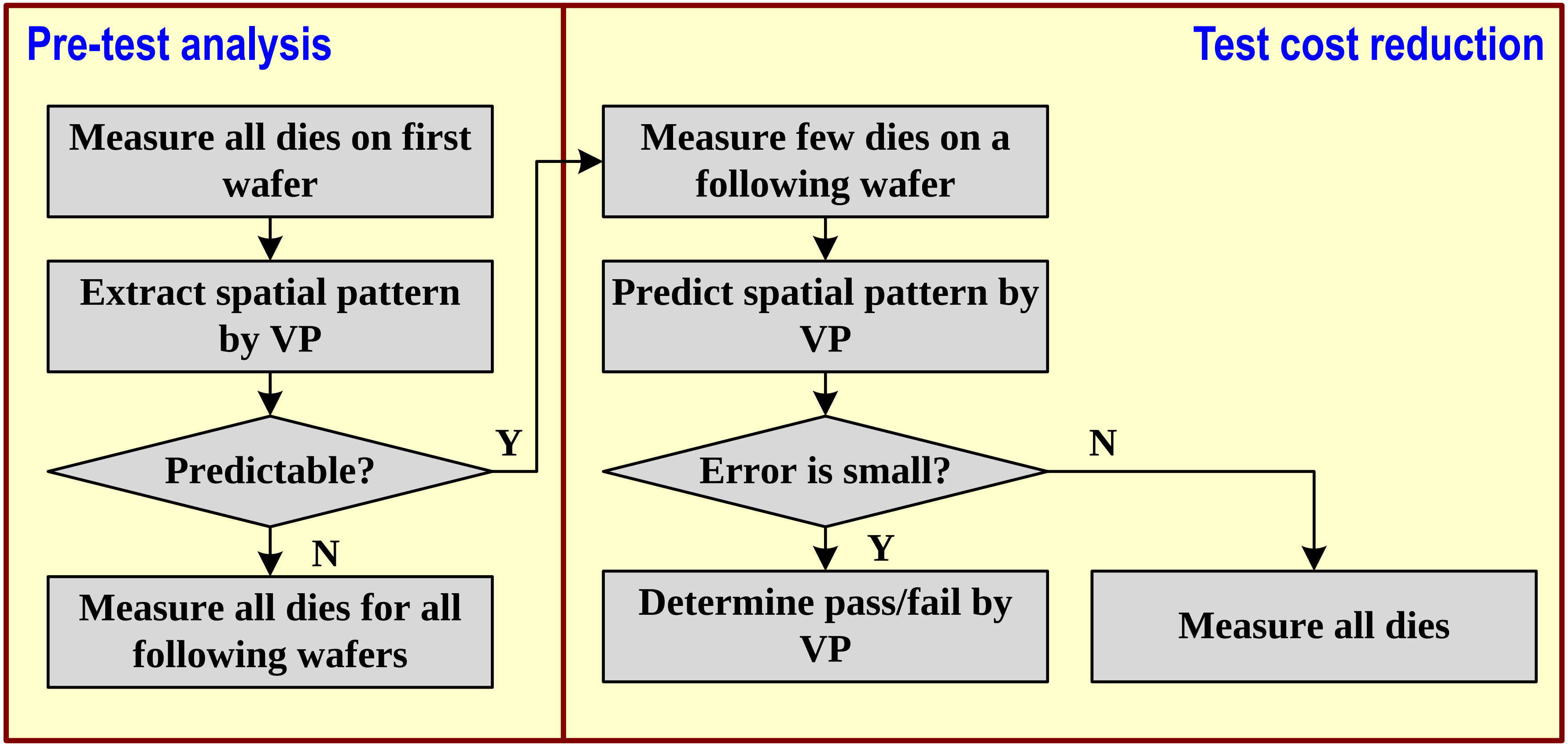
$$\mathbf{x}^{(k+1)} = \frac{1}{1+\eta}\cdot\left(\mathbf{f} - \underline{\mathbf{B}\boldsymbol{\alpha}^{(k)}} + \eta\cdot\underline{\mathbf{B}\mathbf{z}}^{(k+1)}\right)$$

$$\boldsymbol{\alpha}^{(k+1)} = \boldsymbol{\alpha}^{(k)} + \eta\cdot\left(\underline{\mathbf{B}^{T}\mathbf{x}}^{(k+1)} - \mathbf{z}^{(k+1)}\right)$$

- **Since DCT basis functions are used, we can calculate these matrix-vector multiplications by fast DCT or IDCT transform**

Two-Pass Test Flow

- Measure all dies on one wafer if its spatial pattern cannot be predicted by a number of pre-selected dies



Pre-test analysis

Test cost reduction

Measure all dies on first wafer

Extract spatial pattern by VP

Predictable?

Y

N

Measure all dies for all following wafers

Measure few dies on a following wafer

Predict spatial pattern by VP

Error is small?

Y

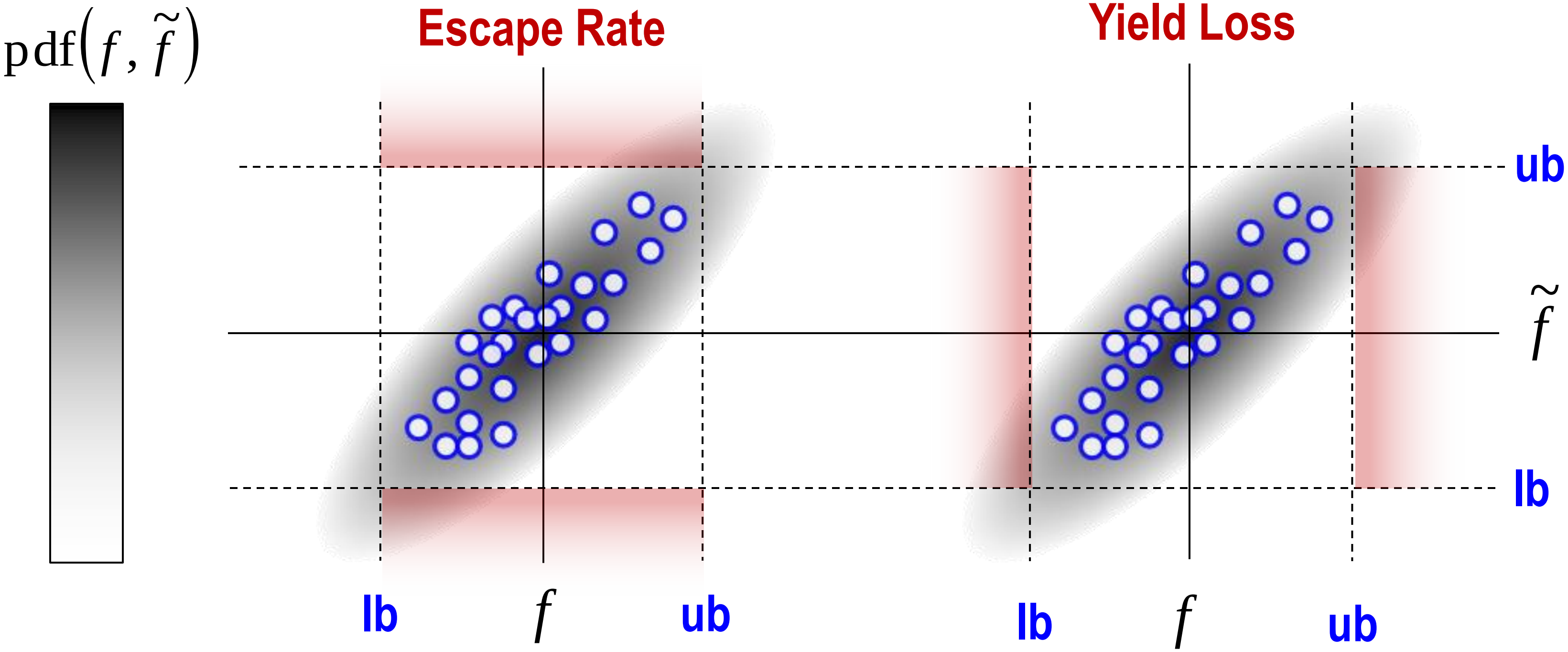
N

Determine pass/fail by VP

Measure all dies

Error Estimation

- Modeling error by VP must be sufficiently small to ensure small escape rate and yield loss



$$ER = \int\limits_{\substack{lb \le \tilde{f} \le ub \\ f \ge ub \cup f \le lb}} \mathrm{pdf}(f, \tilde{f}) \cdot df \cdot d\tilde{f} \qquad YL = \int\limits_{\substack{lb \le f \le ub \\ \tilde{f} \ge ub \cup \tilde{f} \le lb}} \mathrm{pdf}(f, \tilde{f}) \cdot df \cdot d\tilde{f}$$

$\tilde{f}$ **expected values from training** f **measured values from current wafer**

Outline

- **Motivation and background**
  - Virtual probe (VP)
- **Proposed approach**
  - Dual Augmented Lagrangian method (DALM)
  - Two-pass test flow
- **Experimental results**
- **Conclusions**

Experimental Setup

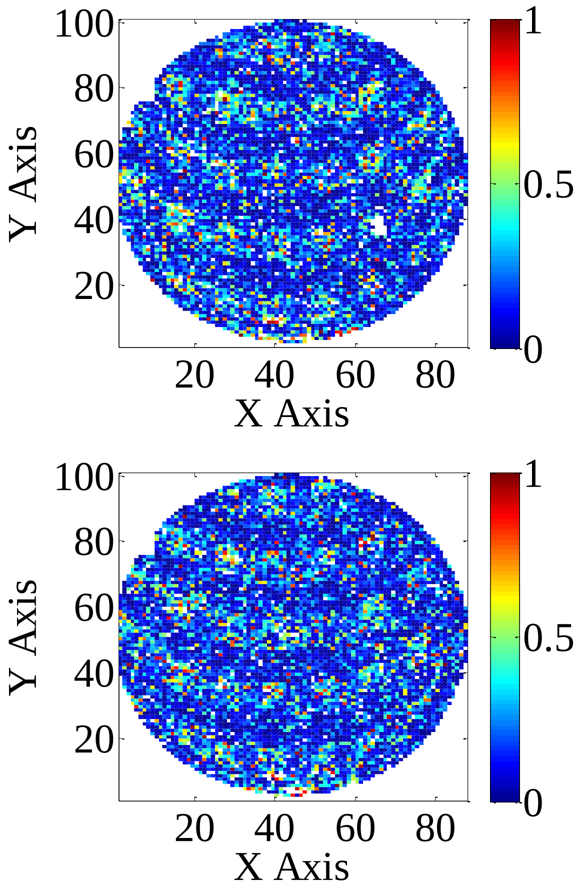
- **Production test data of an industrial dual radio RF transceiver**
  - 9 lots and 176 wafers in total
  - 6766 dies per wafer and 51 test items per die – test items were selected by [Chang11]
  - 1,089,120 good dies and 101,696 bad dies

Lot ID	1	2	3	4	5	6	7	8	9
Wafer #	25	9	23	25	25	25	17	25	2

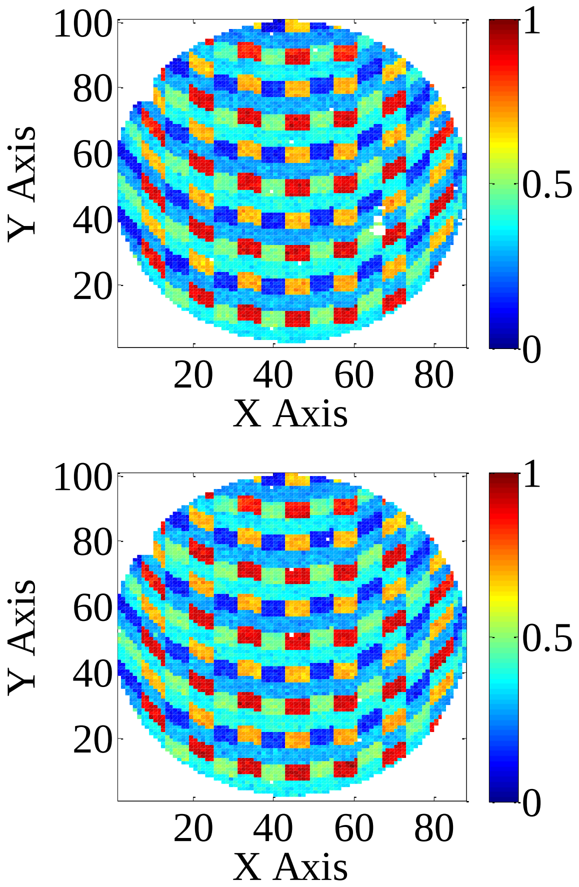
[Chang11]: H. Chang, K. Cheng, W. Zhang, X. Li and K. Butler, “Test cost reduction through performance prediction using virtual probe,” ITC, 2011

Spatial Pattern Examples

- **Spatial pattern is observed for a subset of test items, but not all test items**



Test item #1



Test item #48

Wafer Map Prediction

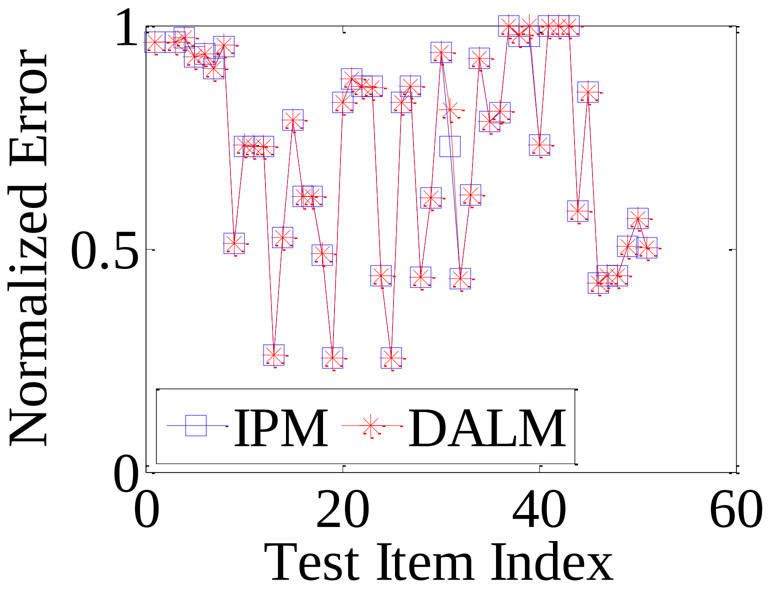
- **Two different solvers are implemented for comparison purpose**
  - IPM: interior-point method
  - DALM: dual augmented Lagrangian method

Number of Dies	IPM	DALM		
	Runtime (Sec.)	Runtime (Sec.)	Iteration #	Speed-up
100	48.3	12.2	7027	3.96×
250	62.7	10.3	5664	6.07×
500	84.7	8.9	5083	9.52×
1000	119.9	8.1	4504	14.88×
2000	171.2	7.3	3922	23.56×
4000	255.2	6.7	3580	37.86×

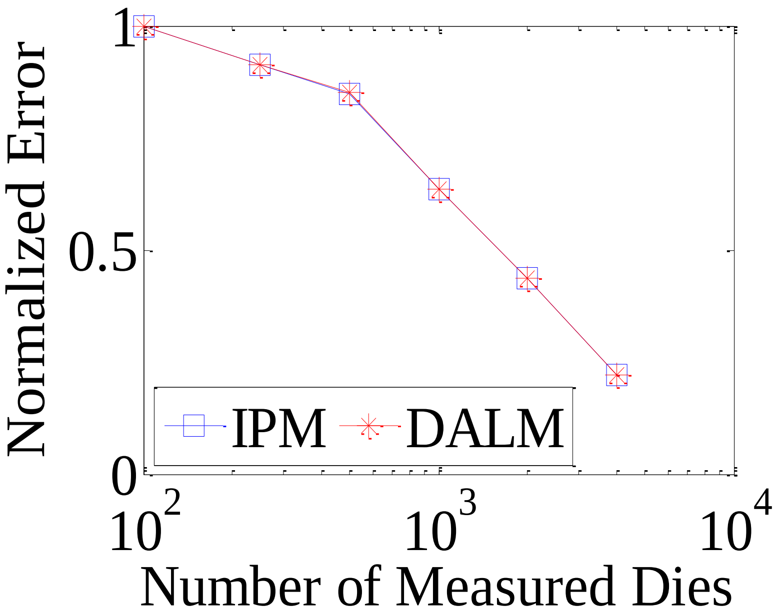
- **DALM achieves up to 37× runtime speedup in this example**

Wafer Map Prediction

- **IPM and DALM result in identical modeling errors**



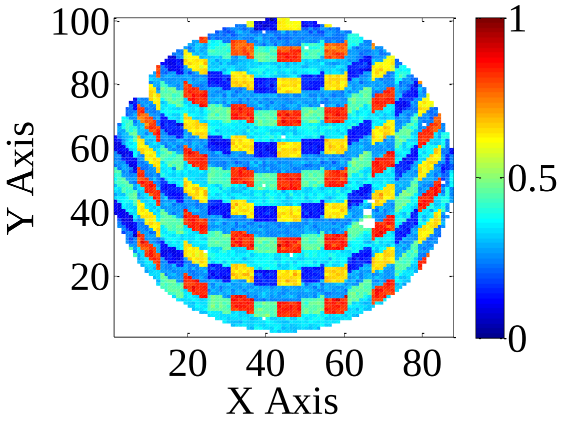
Modeling error with 2000 measured dies per wafer



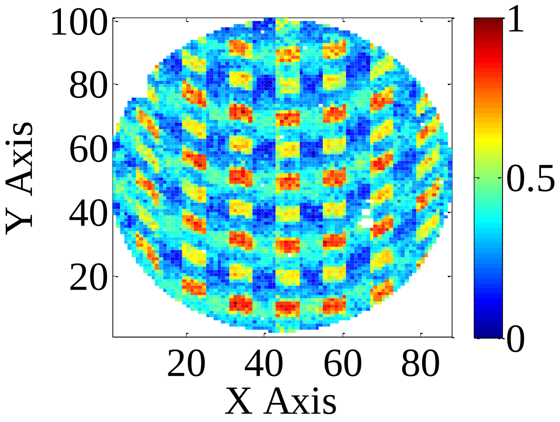
Modeling error for test item #48

Wafer Map Prediction

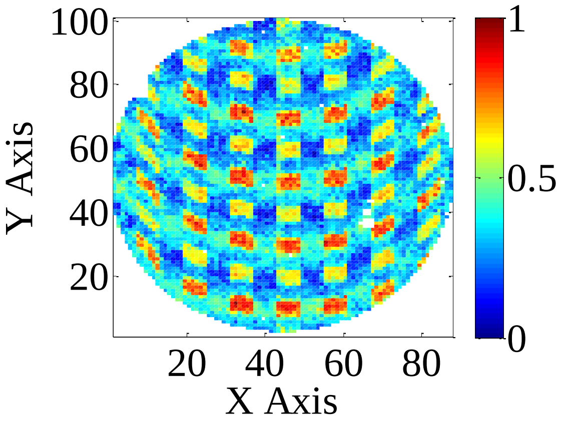
- **IPM and DALM predict identical wafer maps**



Actual wafer map



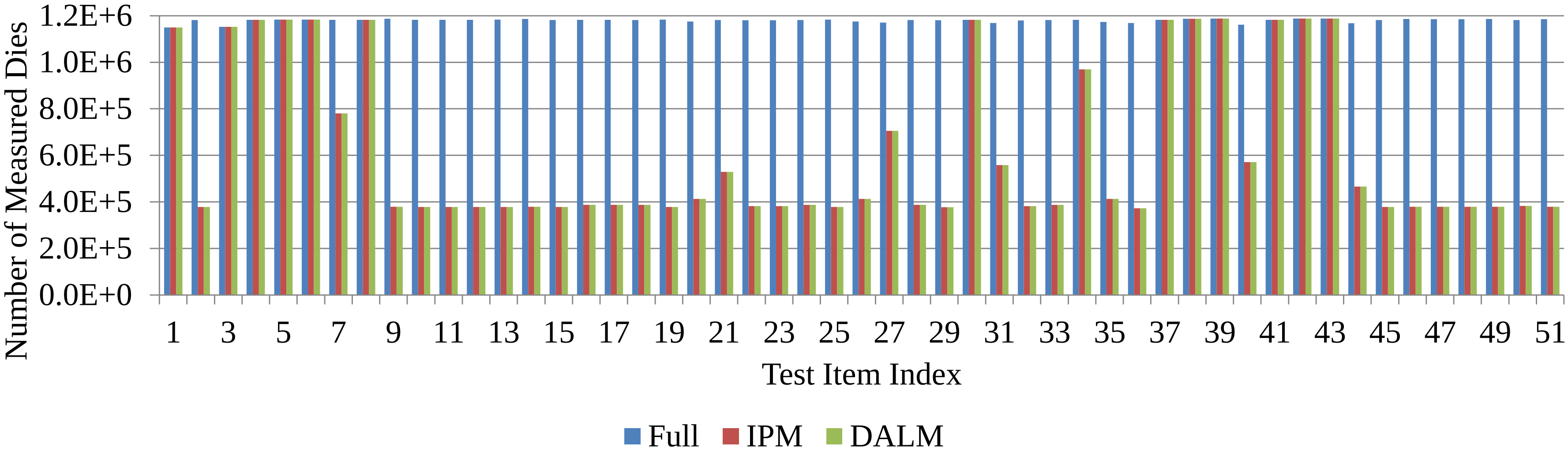
Predicted wafer map by IPM



Predicted wafer map by DALM

Test Cost Reduction

- **Total number of measured dies for each test item**
  - IPM and DALM yield identical results



	Full	IPM	DALM
Overall test cost	60M	32M	32M
Test cost reduction	—	1.9×	1.9×
Escape rate	—	1.2×10^{-3}	1.2×10^{-3}
Yield loss	—	2.0×10^{-3}	2.0×10^{-3}

Conclusions

- **Reducing test cost is a critical task for nanoscale integrated circuit design and manufacturing**
  - Virtual Probe (VP) is an efficient method for test cost reduction based on wafer-level spatial variation modeling

- **Propose an efficient Dual Augmented Lagrangian method (DALM) to reduce the computational cost of VP**
  - Achieve up to 37× runtime reduction over the conventional interior-point solver

- **The proposed DALM solver can be further applied to a number of other analog CAD problems related to sparse approximation**
  - E.g., analog performance modeling, analog self-healing, etc.

References

- [Chang11]: H. Chang, K. Cheng, W. Zhang, X. Li and K. Butler, “Test cost reduction through performance prediction using virtual probe,” ITC, 2011
- [Kupp12]: N. Kupp, K. Huang, J. Carulli and Y. Makris, “Spatial estimation of wafer measurement parameters using Gaussian process models,” ITC, 2012
- [Huang13]: K. Huang, N. Kupp, J. Carulli and Y. Makris, “Handling discontinuous effects in modeling spatial correlation of wafer-level analog/RF tests,” DATE, 2013
- [Hsu13]: C. Hsu, F. Lin, K. Cheng, W. Zhang, X. Li, J. Carulli and K. Butler, “Test data analytics - exploring spatial and test-item correlations in production test data,” ITC, 2013
- [Yang10]: J. Yang and Y. Zhang, “Alternating direction algorithms for l1-problems in compressive sensing,” Technical Report, TR09-37, Rice University, 2010