Sampling Approaches to Metrology in Semiconductor Manufacturing

Tyrone Vincent\textsuperscript{1} and Broc Stirton\textsuperscript{2}, Kameshwar Poolla\textsuperscript{3}

\textsuperscript{1}Colorado School of Mines, Golden CO
\textsuperscript{2}GLOBALFOUNDRIES, Austin TX
\textsuperscript{3}University of California, Berkeley CA
Outline

1. Approach
2. Modeling and Identification
3. Experimental Results
4. Conclusion
the progression for Advanced Process Control (APC) places increased demand on metrology

- lot level variation
  - wafer level variation
    - within-wafer variation
Goal

- decrease the number of measurements needed to determine key features on wafers
Approach

- develop a *correlation model* of metrology measurements for a particular process
- conduct an *a priori* analysis to determine the optimally informative sites that should be measured by minimizing an expected prediction error at the unmeasured sites.
- use a subset of measurement sites together with the correlation model to *predict* measurements at sites which are not measured
Prediction Process

*Identification Set*
- $m$ sites measured

*Prediction Set*
- $q$ sites measured
- $m - q$ sites predicted
Outline

1. Approach
2. Modeling and Identification
3. Experimental Results
4. Conclusion
**Sources of Variation**

\[ L_{\text{gate}}(f, d, k) = L_0 + L_a(k) + L_b(f, d, k) + L_c(f, k) + L_d(d, k) + L_e(d) + L_f(f, d, k) \]

- **“random” variation**
- **layout dependent variation**
- **field level systematic variation and bias**
- **wafer level systematic variation and bias**
Metrology Model 1

wafer sequence

spatial variation

\[
y_k(p) = \bar{y}(p) + \sum_i C_i(p)x_{i,k} + n_k(p)
\]

- parameters
  - \(y_k(p)\) - measurements at position \(p\) on wafer \(k\)
  - \(C_i(p)\) - variation basis function
  - \(x_{i,k}\) - scaling factor for variation function \(i\)
  - \(n_k(p)\) - measurement noise
Identification Process 1

- **vectorized model**
  - \( y_k = \begin{bmatrix} y_k(p_1) & y_k(p_2) & \cdots & y_k(p_m) \end{bmatrix}^T \)
  - \( x_k = \begin{bmatrix} x_{1,k} & x_{2,k} & \cdots & x_{n,k} \end{bmatrix}^T \)
  - \( n_k = \begin{bmatrix} n_k(p_1) & n_k(p_2) & \cdots & n_k(p_m) \end{bmatrix}^T \)

\[
y_k = \bar{y} + C x_k + n_k
\]

- **case 1**: \( x_k \) iid random sequence
  - identification process: \( C \) and covariance of \( x_k \) identified using Principle Component Analysis
Modeling and Identification

Metrology Model 2

wafer sequence

spatial variation

\[ y_k = \bar{y} + \sum_i C_i(p)x_{i,k} + n_k \]

\[
\begin{bmatrix}
  x_{1,k+1} \\
  \vdots \\
  x_{n,k+1}
\end{bmatrix}
= A
\begin{bmatrix}
  x_{1,k} \\
  \vdots \\
  x_{n,k}
\end{bmatrix}
+ w_k
\]
Identification Process 2

Case 2: $x_k$ time correlated

\[ x_{k+1} = Ax_k + w_k \]
\[ y_k = \bar{y} + Cx_k + n_k \]

- parameters $A$, $C$ and covariance of $w_k$ identified using Canonical Correlation Analysis
Prediction

- measured data

\[ y^M_k = \bar{y}^M + C^M x_k + n^M_k \]

- unmeasured values

\[ z^U_k = \bar{y}^U + C^U x_k \]

- common variable \( x_k \) can be estimated from \( y^M_k \) and used to predict \( z^U_k \)
Performance Prediction

- models include uncertainty in process variation and measurement error
- performance prediction possible

\[ S^U := \text{tr} \ \text{Cov} \left[ \hat{z}_k^U - z_k^U \right] \]
Measurement Sequencing

- greedy algorithm for minimizing prediction error

1. Set $\mathcal{M} = \emptyset$, $\mathcal{U} = \{1, \ldots, m\}$
2. For each element $i \in \mathcal{U}$, calculate $S_{\mathcal{U} - \{i\}}$.
3. Select the element $j$ for which $trS_{\mathcal{U} - \{j\}}$ is minimum.
4. Remove $j$ from $\mathcal{U}$ and add it to $\mathcal{M}$.
5. If $|\mathcal{M}| < q$ goto to step 2, otherwise quit.
Implementation

- model requires data - how do we get it?
- re-modeling may be needed - how do we check?
Implementation Process 1

initial data set

lower rate measurements to monitor performance

predicted features

measured features

(IMPACT)
Implementation Process 1

- Initial data set
- Predicted features
- Measured features

Model expires
Implementation Process 1

initial data set

\[ \text{predicted features} \]

\[ \text{measured features} \]

model expires

\[ \text{data set to renew model} \]

(IMPACT) Metrology Sampling
Implementation Process 1

- Predicted features
- Measured features

Initial dataset

Data set to renew model

Model expires

Model expires
Implementation Process 1

initial data set

predicted features

measured features

data set to renew model

model expires

data set to renew model

model expires
Implementation Process 2

- Predicted features
  - Initial data set
- Measured features
  - Data set to renew model
- Model expires
Outline

1. Approach
2. Modeling and Identification
3. Experimental Results
4. Conclusion
Will this work?

- performance is process specific! best case:
  - process variation is “low order”
  - process variation is “stationary”
- validation requires real process data
Real Process Data use for Verification

- poly-gate critical dimension (CD) process data from GLOBALFOUNDRIES Fab1
- 1789 wafers with common litho-etch equipment sequence
- Same feature measured on 18 die.
Experimental Results

Average wafer

![Graph of Average Wafer](image)
Experimental Results

PCA variance contribution

A plot of average prediction error vs. # of sites measured, for different model dimensions, \( n \).
First four basis elements

- PCA C matrix, column 1
- PCA C matrix, column 2
- PCA C matrix, column 3
- PCA C matrix, column 4
Data Splits for Validation

Full Data Set for Litho $i$/Etch $j$

Split: wafer $N_s$

Identification Set:
$N$ wafers
$m$ sites measured

Prediction Set:
$M$ wafers
$q$ sites measured
$m - q$ sites predicted

(Metrology Sampling)
Experimental Results

Prediction Error vs. Model Window

![Graph showing prediction error vs. model window]

- Prediction error with $M = 899$

- Required size of model set $\sim 200$ wafers
Prediction Error vs. Prediction Window

Prediction Error with $N = 200$.

- Performance degrades gracefully with prediction window
No improvement with more complex model, except for small number of measurements per wafer
Outline

1. Approach
2. Modeling and Identification
3. Experimental Results
4. Conclusion
Conclusions

- Intelligent strategies can reduce the expense of metrology without compromising effectiveness.
- Reduced sampling realized through models which predict data at unmeasured sites.

- The real challenge: SPC and fault detection.
- Why do we use metrology? To tell us if something goes wrong [also closed-loop process control].
- What are optimal sampling strategies that don’t compromise SPC or fault detection?
- Metrics: time to detect a process drift or failure, false alarm rates, etc.

Supported in part by NSF grant ECS 01-34132, Berkeley IMPACT program and DOE Programs DE-ZDO-2-30628-07 and DE-FG36-08GO88100.