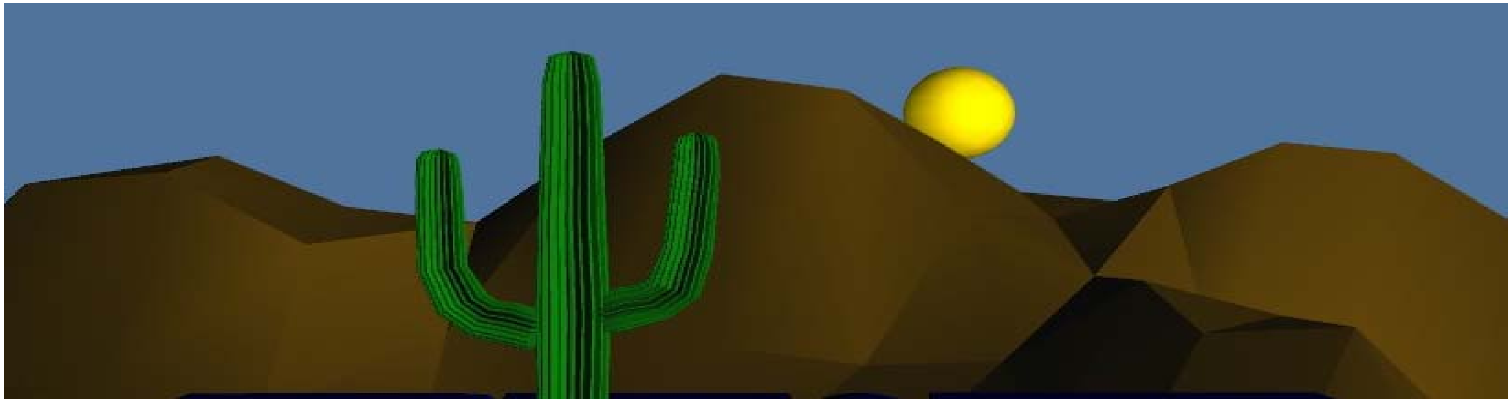
Principle Hessian Direction Based Parameter Reduction for Interconnect networks with Process Variation

Janet Meiling Wang

March 17, 2007

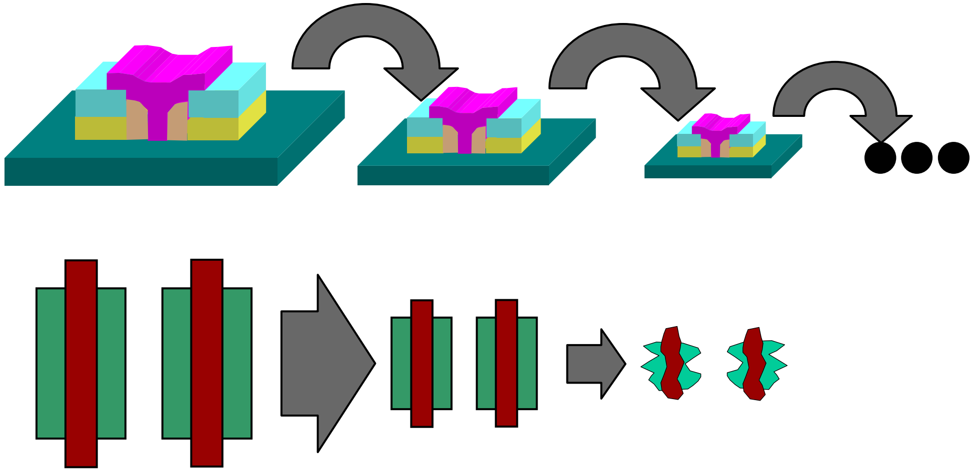


Outline

- Motivations
- Review of other approaches
- Principle Hessian Direction based method
- Experimental results
- conclusion

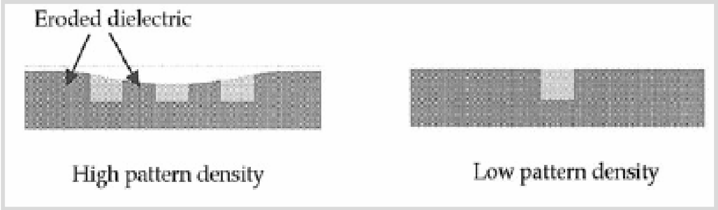
Motivations

- The continuously decreasing feature sizes provides high speed and high density but cause process variations

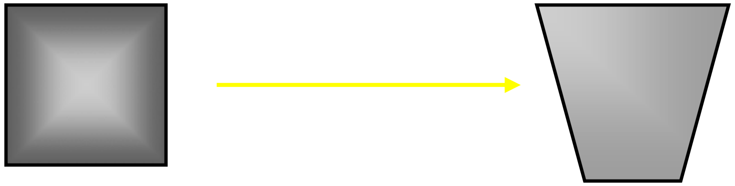


Process Variations

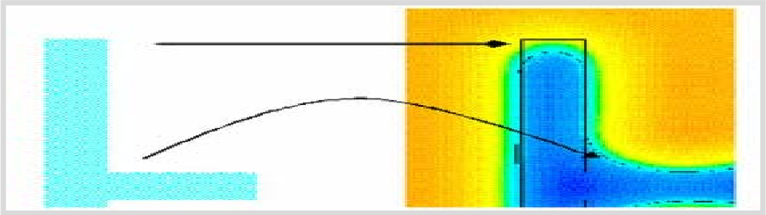
- Chemical-mechanical planarization

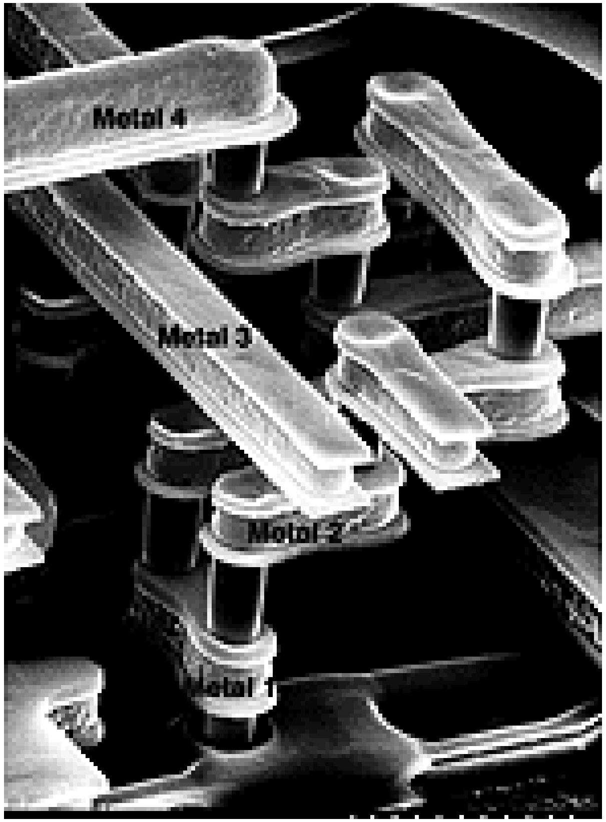


- Chemical etching



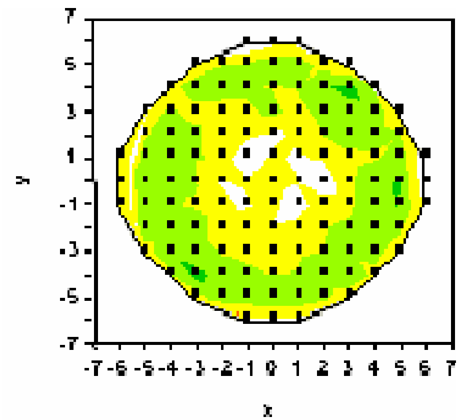
- Optical proximity effects

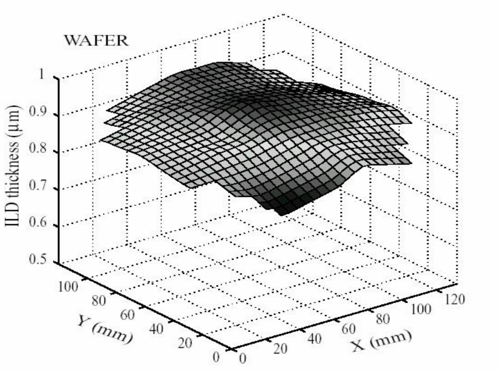


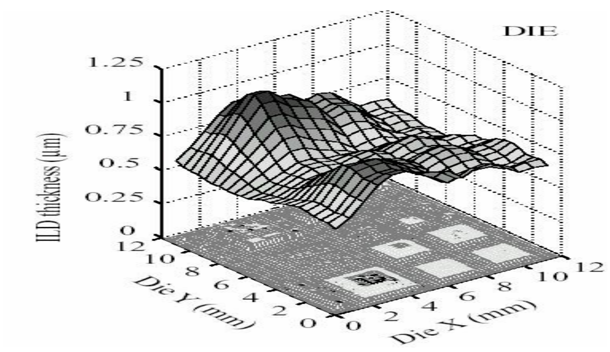


Sources of Variation

- Essential source of variation are the device spatial parameters.
- We consider the transistor width *Weff*, length *Leff* and oxide thickness *Tox*
- Global and local parameters related to spatial device characteristics affect the performance factors





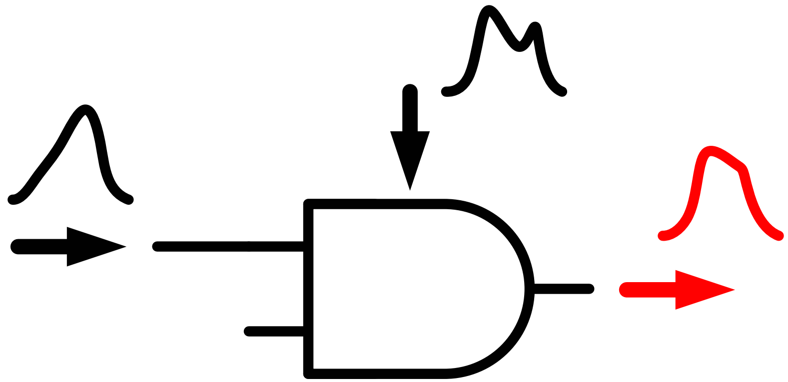


- Example: Intra-die and Inter-die spatial correlation*

*Figures are courtesy of IBM, Intel and TSMC

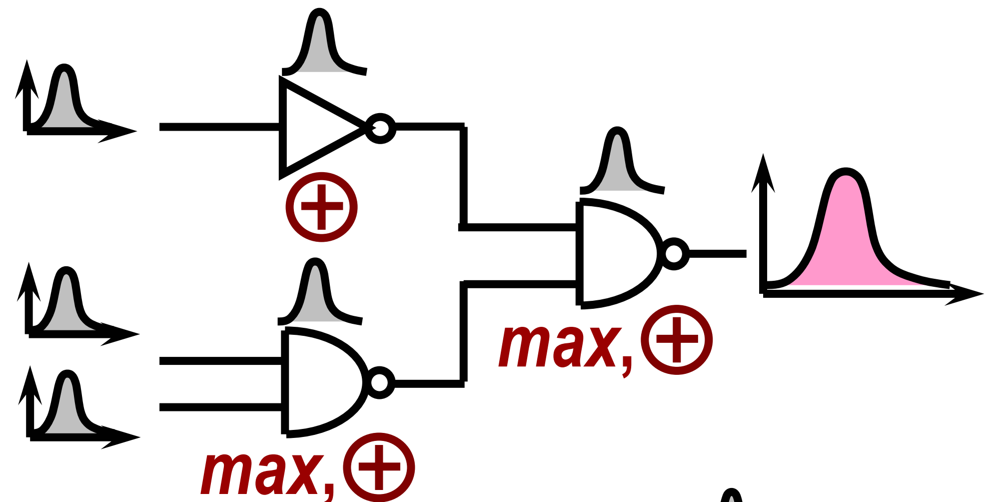
Statistical CAD tools

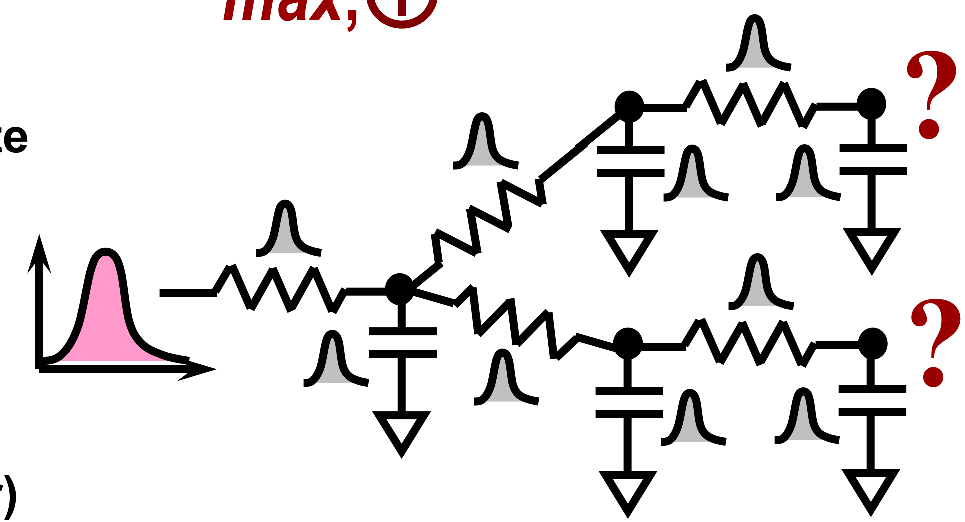
- Nanometer process technology cause circuit performance to deviate from their designed values
- In low cell level, output performance depends on both input and intrinsic uncertainties.
- The output performance deviation is approximated as polynomial with respect the variational sources



Variational Analysis

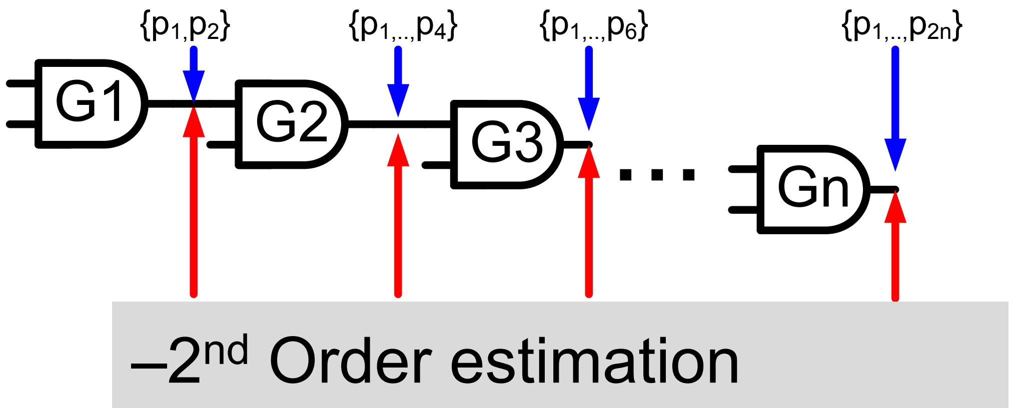
- **Statistical static timing analysis**
  - **Propagate correlated normal distribution**
  - **A limited number of operators: *sum* and *maximum***
- **Statistical *interconnect* timing analysis**
  - **Require a richer palette of computations**
  - ***Not* easy to represent statistics and push them through model reduction algorithms (Courtesy of Rutenbar)**





Performance based strategy

- **We calculate cell level performance using 2nd order polynomial function.**
- **Along each path/block, the performance measures the variation impact of all traversed cells.**
- **The output performance values is 2nd order polynomial with many variables**

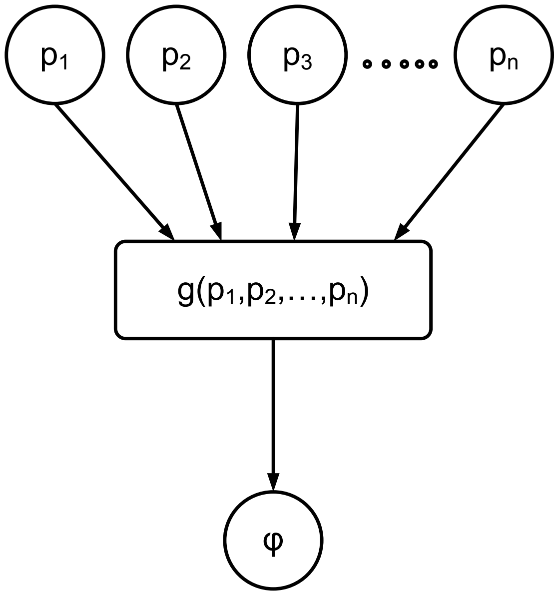


Performance factor approximation

- **The general regression model of the performance function:**

$$\varphi = m\left(v_1, v_2, \ldots v_k\right)$$

- **We seek for reduction of the input space** $\mathbf{p} = \left[p_1, p_2, \ldots, p_n\right]$ **in compact form along with keeping the statistical properties of the output**
- **One possible way is performing the Principle Component Analysis**



Principle Component analysis

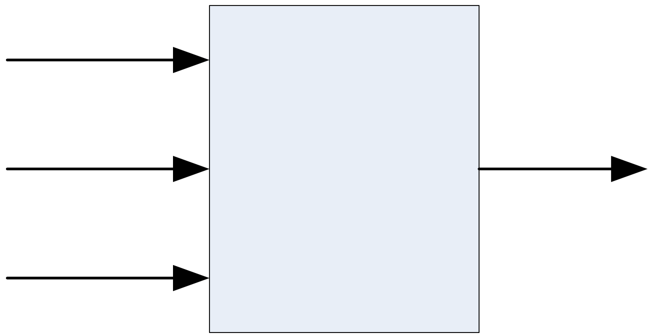
- Most traditional approaches for dimension reduction rely on PCA.
- The principle component is linear combination of all parameters corresponding to maximal resultant variance.

$$b_1 = \arg\max_{\|b\|=1} \left(b^T \Sigma_p b \right)$$

- However an additional reduction can be achieved by considering the output (performance) values

PCA problems

- Example1 : P1 and P3 are uncorrelated, P2=3P1
  - PCA returns 2 principle components
  - In fact the Output depends on one component



- Eample 2: g(x1, x2, x3, x4) and z(x1,x2,x3,x4), PCA leads to   g(x1,x3) and z(x1,x3).

ANOVA Based Approach

- What is ANOVA (Analysis of Variance)?
- As its name suggests – " Analyzes Variances"
- Main Idea - Decomposition of total variance

$$\sigma^2 = \sum_i \sigma_i^2$$

- Mean response due to a particular input - Keep that input constant and vary all other inputs

$$\hat{\mu}_i(\xi_i) \equiv \int \dots \int \hat{y}(\xi_1, \xi_2, \dots, \xi_n) d\xi_1 \dots d\xi_{i-1} d\xi_{i+1} \dots d\xi_n$$

ANOVA Based Approach

- Variance due to design variable ξ_i

$$\hat{\sigma}_i^2 = \int [\hat{\mu}(\xi_i) - \mu]^2 d\xi_i$$

- Statistical Significance parameter (F):

$$\frac{\int [\hat{\mu}(\xi_i) - \mu]^2 d\xi_i}{\sigma^2}$$

- We calculate the "F" parameter using ANOVA
- Another Important parameter found using ANOVA is: R^2
- Based on these parameters, the algorithm decides whether the input parameter is significant or not.

An Example

Delay for a single RC segment of a global interconnect for 0.13um technology

$$\begin{aligned} delay = \; & 19.65 - 2.28\xi_1 - 0.9\xi_2 - 1.82\xi_3 - 0.32\xi_4 \\ & +0.28(\xi_1^2 - 1) + 0.1(\xi_2^2 - 1) + 0.12(\xi_3^2 - 1) \\ & +0.05(\xi_4^2 - 1) + 0.17(\xi_1\xi_2) + 0.03(\xi_1\xi_4) \\ & +0.2(\xi_2\xi_3) - 0.17(\xi_2\xi_4) + 0.17(\xi_3\xi_4)\ ps \end{aligned}$$

Mean = 19.62ps
Variance = 3.15ps

In this case, ANOVA gives us terms that are insignificant as follows:

$$\xi_4, \xi_2^2, \xi_4^2, \xi_1\xi_2, \xi_1\xi_3, \xi_1\xi_4, \xi_2\xi_4, \xi_3\xi_4$$

After removing these terms, the reduced equation is:

$$\begin{aligned} delay = \; & 19.65 - 2.28\xi_1 - 0.9\xi_2 - 1.82\xi_3 + 0.28(\xi_1^2 - 1) \\ & +0.12(\xi_3^2 - 1) + 0.2(\xi_2\xi_3)\ ps \end{aligned}$$

Mean = 19.64ps
Variance = 3.13ps

Still problems ??

- ANOVA may not reduce anything…
- How about represent the existing parameters in other parameters with a shorter list of new parameters ?
- The answer is YES, but you need to identify the transfer matrix B between the new and old parameters effectively

Effective Dimension Reduction (EDR)

- **Another way is through additional intermediate mapping the parameter space** p **to the output performance function** φ

- **The reduction is achieved if k<<n**

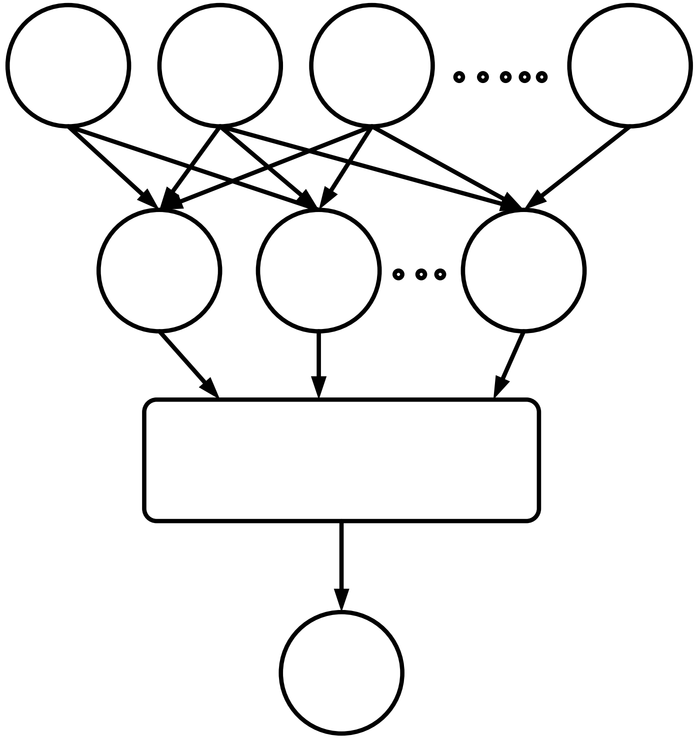
- Definition**: The space B generated by** $\mathbf{B}=[\beta_1,\beta_2,...\beta_K]$ **is called the EDR space. Any non-zero vector in the EDR space is called an EDR direction**.

Effective Dimension Reduction (EDR)

$$\varphi = m\left(v_1, v_2, \ldots v_k\right)$$

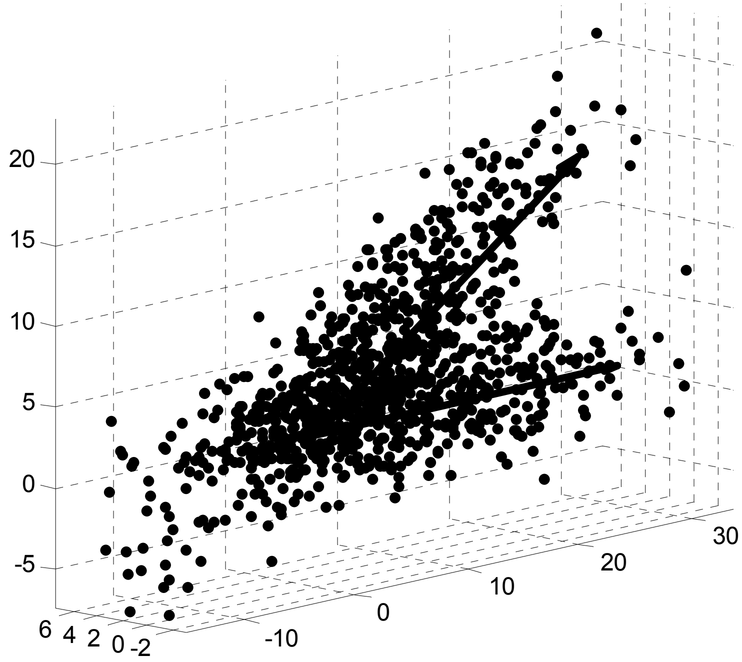
$$v_i = \beta_i\left[p_1, p_2, \ldots p_n\right]$$

$$k << n$$



Effective Dimension Reduction (EDR)

- Therefore the intermediate function m capture the contribution of the parameter set to the output performance function
- Key point is to finding the smallest effective dimension reduction space and such of space is unique [Cook 1998]



Proposed approach

- **Link process variation parameters with performance by weighed sum strategy**

Performance

mean

$$\Sigma_{\phi pp} = (\phi - \bar{\phi})(\mathbf{p} - \bar{\mathbf{p}})(\mathbf{p} - \bar{\mathbf{p}})^t$$

Performance-Parameter Covariance matrix

Hessian Matrix

- Hessian Matrix for performance is a symmetric matrix with 2nd order derivative and defined as

$$\mathbf{H}_\phi(\mathbf{p}) = \begin{bmatrix} \frac{\partial^2 \phi}{\partial p_1 \partial p_1} & \frac{\partial^2 \phi}{\partial p_1 \partial p_2} & \cdots & \frac{\partial^2 \phi}{\partial p_1 \partial p_n} \\ \frac{\partial^2 \phi}{\partial p_2 \partial p_1} & \frac{\partial^2 \phi}{\partial p_2 \partial p_2} & \cdots & \frac{\partial^2 \phi}{\partial p_2 \partial p_n} \\ \frac{\partial^2 \phi}{\partial p_n \partial p_1} & \frac{\partial^2 \phi}{\partial p_n \partial p_2} & \cdots & \frac{\partial^2 \phi}{\partial p_n \partial p_n} \end{bmatrix}$$

- Each entry is a function of p, so we have $\bar{\mathbf{H}}_\phi(\mathbf{p}) \doteq E\left[\mathbf{H}_\phi(\mathbf{p})\right]$

Relationship between p and Hessian Matrix

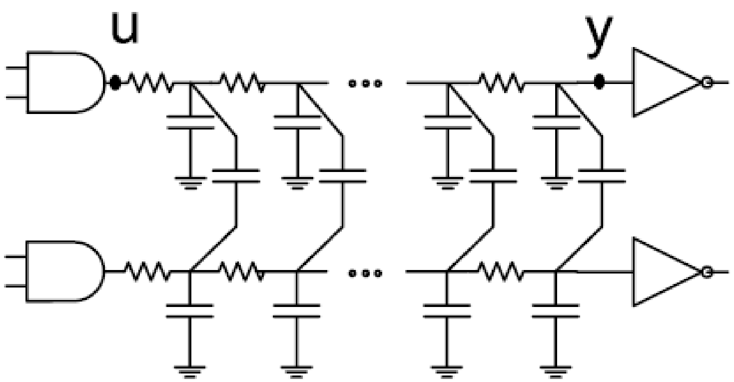
Theorem 1 Let matrix Σ_p be $\mathbf{p}$'s covariance matrix. Assume $\phi(\mathbf{p})$ has second order derivative with regard to $\mathbf{p}$. Then the following equation is true:

$$\Sigma_{\phi pp} = \Sigma_p \bar{\mathbf{H}}_\phi \Sigma_p$$

eigen-decomposition

$$\bar{\mathbf{H}}_\phi \Sigma_p b_j = \lambda_j b_j$$

RC network analysis



$$G(\mathbf{p}) = G_0 + G_1\mathbf{p} + G_2\mathbf{p}^t\mathbf{p}$$
$$C(\mathbf{p}) = C_0 + C_1\mathbf{p} + C_2\mathbf{p}^t\mathbf{p}$$

$$\begin{aligned} G(\mathbf{p})\mathbf{x} + sC(\mathbf{p})\mathbf{x} &= Bu \\ y &= L\mathbf{x}, \end{aligned}$$

The Response of the RC network

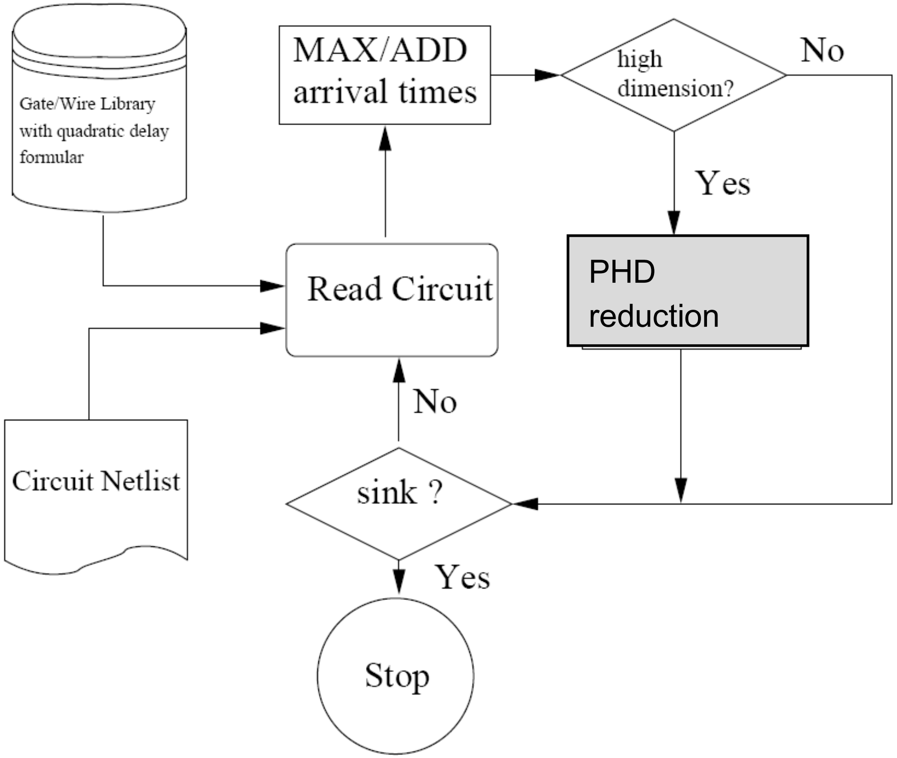
$$H(s) = L(G(\mathbf{p}) + sC(\mathbf{p}))^{-1}B \approx M_0(\mathbf{p}) + M_1(\mathbf{p})s + M_2(\mathbf{p})s^2$$

$$\begin{aligned}
M_0(\mathbf{p}) &\approx m_{00} + m_{01}\mathbf{p} + m_{02}\mathbf{p}^t\mathbf{p} \\
M_1(\mathbf{p}) &\approx m_{10} + m_{11}\mathbf{p} + m_{12}\mathbf{p}^t\mathbf{p} \\
M_2(\mathbf{p}) &\approx m_{20} + m_{21}\mathbf{p} + m_{22}\mathbf{p}^t\mathbf{p}
\end{aligned}$$

This is the performance

PHD based algorithm for parameter reduction

- **Static Statistical Timing Analysis (SSTA) flow with Sliced Inverse Regression based reduction**



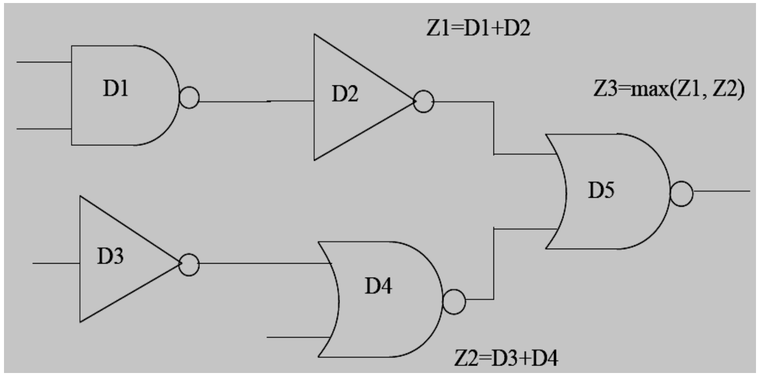
Numerical examples

- Consider the timing block-wise SSTA analysis with regard to the gates D2, D4 and D5.
- Assume that all delays are in quadratic form and at each gate the variational sources are not the same (at least correlated).
- The propagation delay model is expressed in space of 9 deign parameters and we seek for dimension reduction

$$Z_1 = y_1(x_1, x_2, x_3)$$
$$Z_2 = y_4(x_4, x_5, x_6)$$
$$Z_3 = y_5(x_7, x_8, x_9)$$

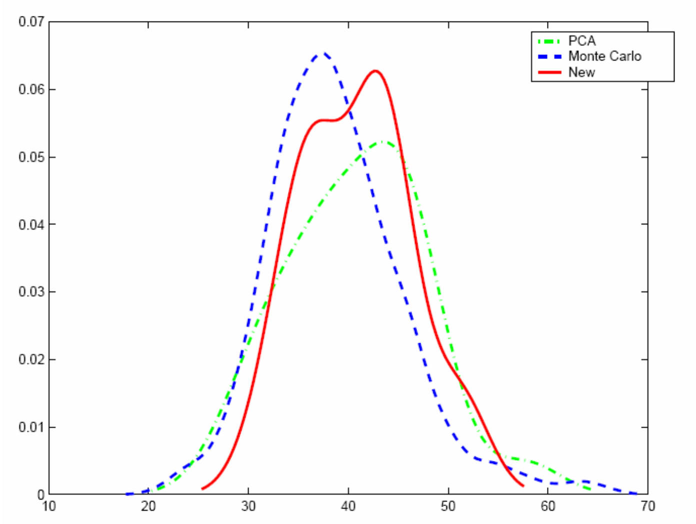
- Arrival time at D5 is calculated as:

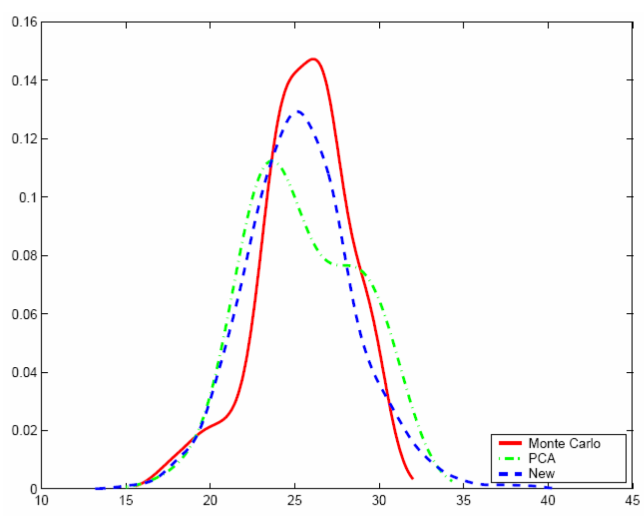
$$y = \max\left(\mathrm{sum}(Z_1, Z_3), \mathrm{sum}(Z_2, Z_3)\right)$$



Numerical examples

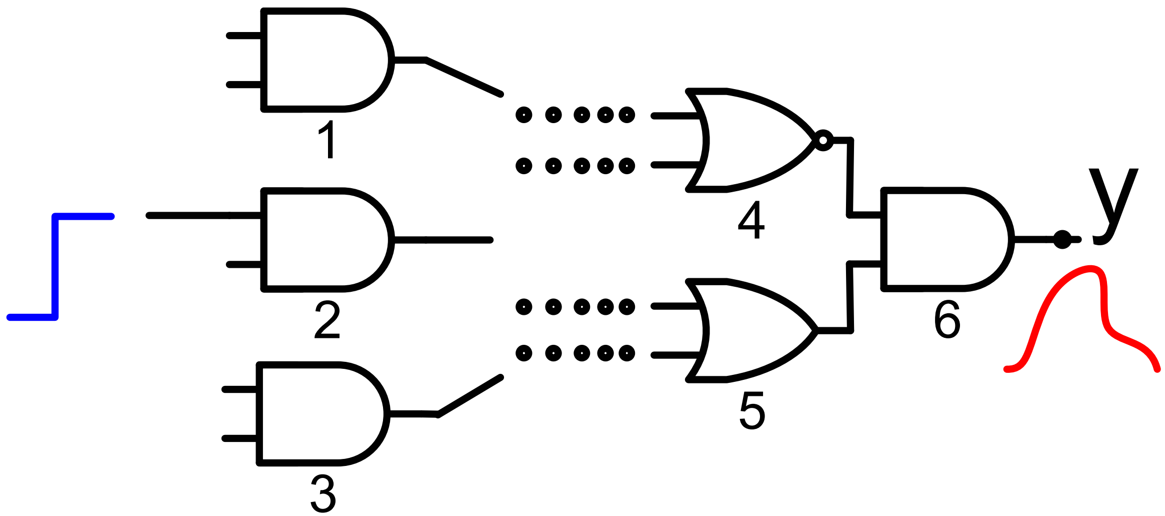
- The new reduction algorithm reduces the original 9 variable design space to 3, a reduction of 60%.
- If PCA is applied, the reduced space will consider the 6 variables, leading to only 34% reduction.
- The reconstruction of the new function $\bar{m}(\hat{\beta}_1^t x, \hat{\beta}_2^t x, \cdots, \hat{\beta}_K^t x)$ is carried through least square approximation





Numerical examples

- **ISCASS'85 benchmark circuits**
  - **We determine different input-output paths and estimated the arrival time with Monte Carlo, SSTA + PCA, SSTA+PHD**
  - **Assume deterministic input for tested paths**
  - **All transistors and interconnects are affected by process variation**
  - **The expression for intermediate arrival times are reduced with PHD at most after 3 gates in series.**



Numerical examples

- **The results show clearly the advantage of using the proposed reduction scheme over PCA: new method can achieve 20% to 50% parameter reduction with only less than 8% error on average..**

Circuit	Number of Gates	PCA delay mean error (%)	New delay mean error (%)	PCA delay variance error (%)	New delay variance error (%)	PCA reduction (%)	New reduction (%)
C17	6	0.5	0.4	6.5	4.3	25	42
C432	160	2.3	1.9	9.2	7.1	37	61
C499	202	5.4	2.2	11.9	8.5	23	39
C880	383	3.0	3.0	6.0	6.0	44	44
C1355	546	3.2	3.0	5.5	6.0	30	38
C1908	880	1.1	1.0	6.7	7.2	10	41
C2670	1193	3.1	2.9	5.0	4.5	20	30
C3540	1669	5.4	4.0	8.0	8.0	25	35
C5315	2307	5.0	5.0	8.3	8.3	35	35
C6288	2416	1.6	2.6	6.6	7.0	23	36
C7552	3512	2.7	2.2	4.0	7.4	20	53

Conclusions

- We propose a new way of reducing the statistical variations.
- The new approach creates an effective reduction subspace and provides a transformation matrix by using the mean and variance of the response surface.
- With the generated transformation matrix, the proposed method maps the original statistical variations to a smaller set of variables with which we process variability analysis.
- The computational cost due to the number of variations is greatly reduced.