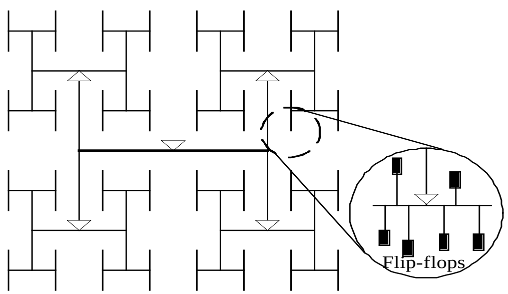
Accurate Clock Mesh Sizing via Sequential Quadratic Programming

Venkata Rajesh Mekala, Yifang Liu, Xiaoji Ye, Jiang Hu, Peng Li
Department of ECE
Texas A&M University

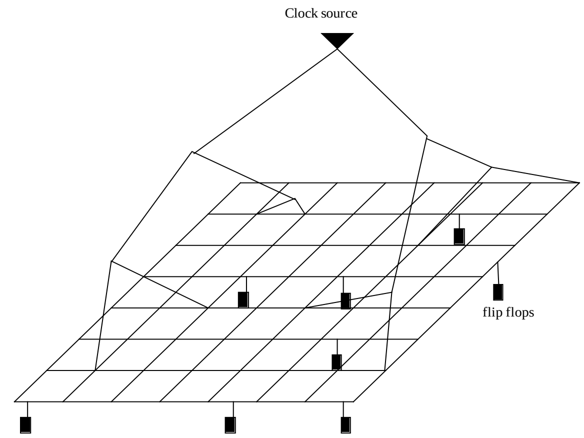
OUTLINE

- Introduction
- Previous Works
- Problem Formulation
- Algorithm Overview
- Results
- Conclusions

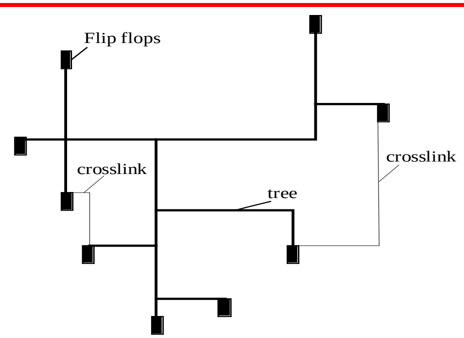
Clock Architectures



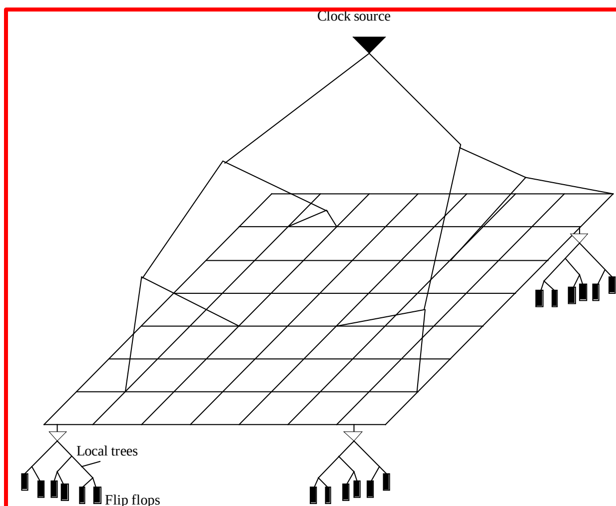
Clock Tree
- low cost (wiring, power, cap)
- higher skew, jitter than mesh
- widely used in ASIC designs
- clock gating easy to incorporate



Clock Mesh
- excellent for low skew, jitter
- high power, area, capacitance
- difficult to analyze
- clock gating not easy
- used in modern processors



Hybrid: tree + cross-links
- low cost (wiring, power, cap)
- smaller skew, jitter than tree*
- difficult to analyze



Hybrid: mesh + local trees

Clock Mesh

- Clock mesh architecture is very effective in reducing skew variation.
- Clock mesh is difficult in analyzing with sufficient accuracy.
- It dissipates higher power compared to other architectures.
- The challenge is to design the mesh with less power meeting the skew constraints.

Clock Distribution Networks

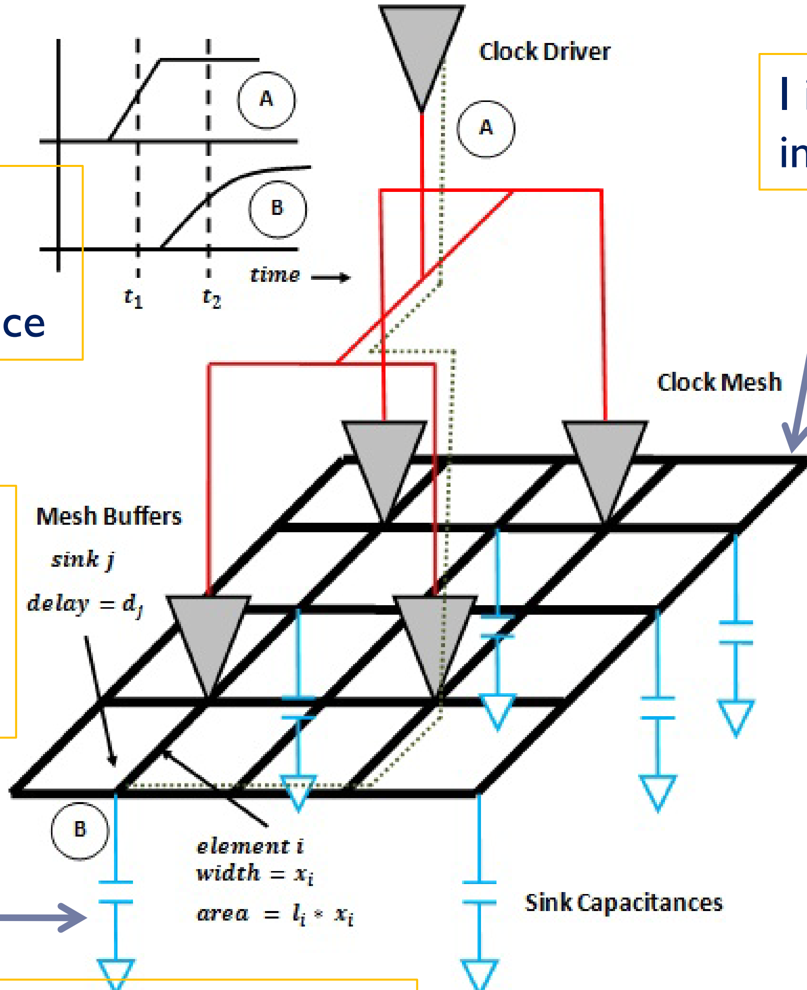
- **Clock Trees**
  - Pullela, Menezes and Pileggi
    Moment-sensitivity-based wire sizing for skew reduction
    1997
  - Wang, Ran, Jiang and Sadowska
    General skew constrained clock network sizing based on sequential linear programming
    2005
  - Guthaus, Sylvester and Brown
    Clock buffer and wire sizing using sequential programming
    2006
- **Crosslinks**
  - Rajaram, Hu and Mahapatra
    Reducing clock skew variability via crosslinks
    2006
  - Samanta, Hu and Li
    Discrete buffer and wire sizing for link-based non-tree clock networks
    2008
- **Clock Mesh**
  - Desai, Cvijetic and Jensen
    Sizing of clock distribution networks for high performance CPU chips
    1996
  - Venkataraman, Feng, Hu and Li
    Combinatorial algorithms for fast clock mesh optimization
    2006
  - Rajaram and Pan
    MeshWorks: An efficient framework for planning, synthesis and optimization of clock mesh networks
    2008

Motivation & Our Contributions

- Current-source based gate modeling approach to speedup the accurate analysis of clock mesh.
- Efficient adjoint sensitivity analysis to provide desirable sensitivities.
- Algorithm based on rigorous SQP.
- First clock mesh sizing method that does systematic solution search and is based on accurate delay model

Problem Formulation

μ is the average value of the sink delays and δ is the given maximum variance



I is the set of interconnects in the clock mesh

x_i ; i ϵ I is the width of element *i* in the interconnect set

w_i ; i ϵ I is the area of element *i* in the interconnect set

d_j ; j ϵ S the propagation delay of the signal from the root of the clock tree to sink j

D is the coefficient vector reflecting the linear size-area relation

S is the set of sinks or local trees

L_x and U_x represent the lower bound and upper bound vectors of the wires

Problem Formulation

total clock mesh area

Minimize: $$w = \sum_{i \in I} w_i = \mathbf{x}^T D,$$

s. t. $$\sigma^2 = \sum_{j \in S} (d_j - \mu)^2 \leq \delta,$$

$$L_{\mathbf{x}} \leq \mathbf{x} \leq U_{\mathbf{x}},$$

skew constraint in the variance form

lower bound, upper bound vectors of the wire widths

Higher wire area leads to a higher load capacitance for the clock buffers which in turn implies a higher power dissipation.

Constraint in the quadratic form is a differentiable function

Solving the Problem

- Lagrangian of the original problem:

$$\mathcal{L}(\mathbf{x}, \lambda) = \mathbf{x}^T D - \lambda(\delta - \sigma^2).$$

- Gradient vector of the Lagrangian function

$$\nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \lambda) = D + \lambda \nabla_{\mathbf{x}} \sigma^2.$$

$\nabla_{\mathbf{x}} \sigma^2$ is be obtained by circuit simulation and adjoint sensitivity analysis

Solving the Problem

- Lagrangian of the original problem:

$$\mathcal{L}(\mathrm{x}, \lambda) = \mathrm{x}^T D - \lambda(\delta - \sigma^2).$$

- Gradient vector of the Lagrangian function

$$\nabla_{\mathbf{x}} \mathcal{L}(\mathrm{x}, \lambda) = D + \lambda \nabla_{\mathbf{x}} \sigma^2.$$

The adjoint sensitivity analysis gives us the values of

$$\frac{\partial \sigma^2}{\partial R} \text{ and } \frac{\partial \sigma^2}{\partial C}$$

Solving the Problem

- Lagrangian of the original problem:

$$\mathcal{L}(\mathbf{x}, \lambda) = \mathbf{x}^T D - \lambda(\delta - \sigma^2).$$

- Gradient vector of the Lagrangian function

$$\nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \lambda) = D + \lambda \nabla_{\mathbf{x}} \sigma^2.$$

The sensitivities with respect to wire widths are calculated with the help of chain rule:

$$\frac{\partial \sigma^2}{\partial X} = (\frac{\partial \sigma^2}{\partial R} . \frac{\partial R}{\partial X}) + (\frac{\partial \sigma^2}{\partial C} . \frac{\partial C}{\partial X})$$

Solving the Problem

- Lagrangian of the original problem:

$$\mathcal{L}(\mathrm{x}, \lambda) = \mathrm{x}^T D - \lambda(\delta - \sigma^2).$$

- Gradient vector of the Lagrangian function

$$\nabla_{\mathbf{x}} \mathcal{L}(\mathrm{x}, \lambda) = D + \lambda \nabla_{\mathbf{x}} \sigma^2.$$

- Necessary conditions for any optimal point of the problem – KKT conditions

$$D + \lambda \nabla_{\mathbf{x}} \sigma^2 = 0,$$

$$\delta - \sigma^2 \geq 0.$$

Common way to solve this equation is by Newton's method.

$$\boxed{D + \lambda \nabla_{\mathbf{x}} \sigma^2 = 0}$$

Solving the Problem

- Let the Newton step in iteration k of solving the equation be:

$$\begin{bmatrix} \mathbf{p}_{\mathbf{x},k} \\ \mathbf{p}_{\lambda,k} \end{bmatrix} = \begin{bmatrix} \mathbf{x}_{k+1} \\ \lambda_{k+1} \end{bmatrix} - \begin{bmatrix} \mathbf{x}_k \\ \lambda_k \end{bmatrix}$$

x, λ are variables in the equation.
$p_{x,k}$ and $p_{\lambda,k}$ are the vectors representing change in width of wires and Lagrangian multiplier.

Solving the Problem

$$\boxed{D + \lambda \nabla_{\mathbf{x}} \sigma^2 = 0}$$

- Let the Newton step in iteration k of solving the equation be:

$$\begin{bmatrix} \mathrm{p}_{\mathbf{x},k} \\ \mathrm{p}_{\lambda,k} \end{bmatrix} = \begin{bmatrix} \mathrm{x}_{k+1} \\ \lambda_{k+1} \end{bmatrix} - \begin{bmatrix} \mathrm{x}_k \\ \lambda_k \end{bmatrix}$$

- Jacobian of the equation is:

$$\begin{bmatrix} \nabla^2_{\mathbf{xx}} \mathcal{L}(\mathrm{x}, \lambda) & \nabla_{\mathbf{x}} \sigma^2 \end{bmatrix}$$

- Hessian of the Lagrangian function:

$$H = \nabla^2_{\mathbf{xx}} \mathcal{L}(\mathrm{x}, \lambda)$$

- Newton step calculation implies that $p_{x,k}$ and $p_{\lambda,k}$ satisfy the following system:

$$\begin{bmatrix} H_k & \nabla_{\mathbf{x}} \sigma_k^2 \end{bmatrix} \begin{bmatrix} \mathrm{p}_{\mathbf{x},k} \\ \mathrm{p}_{\lambda,k} \end{bmatrix} = \begin{bmatrix} -D - \lambda_k \nabla_{\mathbf{x}} \sigma_k^2 \end{bmatrix}$$

Solving the Problem

- Newton step calculation implies that $p_{x,k}$ and $p_{\lambda,k}$ satisfy the following system:

$$\begin{bmatrix} H_k & \nabla_{\mathbf{x}}\sigma_k^2 \end{bmatrix} \begin{bmatrix} \mathrm{p}_{\mathbf{x},k} \\ \mathrm{p}_{\lambda,k} \end{bmatrix} = \begin{bmatrix} -D - \lambda_k \nabla_{\mathbf{x}}\sigma_k^2 \end{bmatrix}$$

- Adjusting the above equation gives us:

$$H_k \mathbf{p}_{\mathbf{x},k} + D + \lambda_{k+1} \nabla_{\mathbf{x}} \sigma_k^2 = 0$$

- This equation is solved by:
  - Minimize: $\frac{1}{2}\mathrm{p}_{\mathbf{x}}^T H \mathrm{p}_{\mathbf{x}} + D^T \mathrm{p}_{\mathbf{x}}$
  - Subject to: $\delta - \sigma^2 \geq 0$

Solving the QP sub-problem

- The QP sub-problem to be solved as a part of SQP is:

Minimize: $$\frac{1}{2}\mathbf{p}_{\mathbf{x}}^T H \mathbf{p}_{\mathbf{x}} + D^T \mathbf{p}_{\mathbf{x}}$$

Subject to:

$$\delta - \left(\sigma^2 + (\nabla_{\mathbf{x}}\sigma^2)^T \mathbf{p}_{\mathbf{x}}\right) \geq 0$$

and $$L_{\mathbf{x}} \leq \mathbf{x} \leq U_{\mathbf{x}}$$

Solving the QP sub-problem

- The QP sub-problem to be solved as a part of SQP is:

Minimize: $$\frac{1}{2}\mathbf{p}_{\mathbf{x}}^{T} H \mathbf{p}_{\mathbf{x}} + D^{T}\mathbf{p}_{\mathbf{x}}$$

Subject to:

$$\delta - (\sigma^2 + (\nabla_{\mathbf{x}}\sigma^2)^T \mathbf{p}_{\mathbf{x}}) \geq 0$$

and $$L_{\mathbf{x}} \leq \mathbf{x} \leq U_{\mathbf{x}}$$

the sensitivities with respect to wire widths are calculated with the help of chain rule:

$$\frac{\partial\sigma^2}{\partial X} = \left(\frac{\partial\sigma^2}{\partial R} \cdot \frac{\partial R}{\partial X}\right) + \left(\frac{\partial\sigma^2}{\partial C} \cdot \frac{\partial C}{\partial X}\right)$$

through sensitivity analysis we obtain the gradient.

Solving the QP sub-problem

- The QP sub-problem to be solved as a part of SQP is:

Minimize:

$$\frac{1}{2}\mathrm{p}_{\mathbf{x}}^{T} H \mathrm{p}_{\mathbf{x}} + D^{T}\mathrm{p}_{\mathbf{x}}$$

Subject to:

$$\delta - \left(\sigma^2 + (\nabla_{\mathbf{x}}\sigma^2)^T \mathrm{p}_{\mathbf{x}}\right) \geq 0$$

and

$$L_{\mathbf{x}} \leq \mathbf{x} \leq U_{\mathbf{x}}$$

we use quasi-newton (BFGS) method to approximate the hessian in each iteration

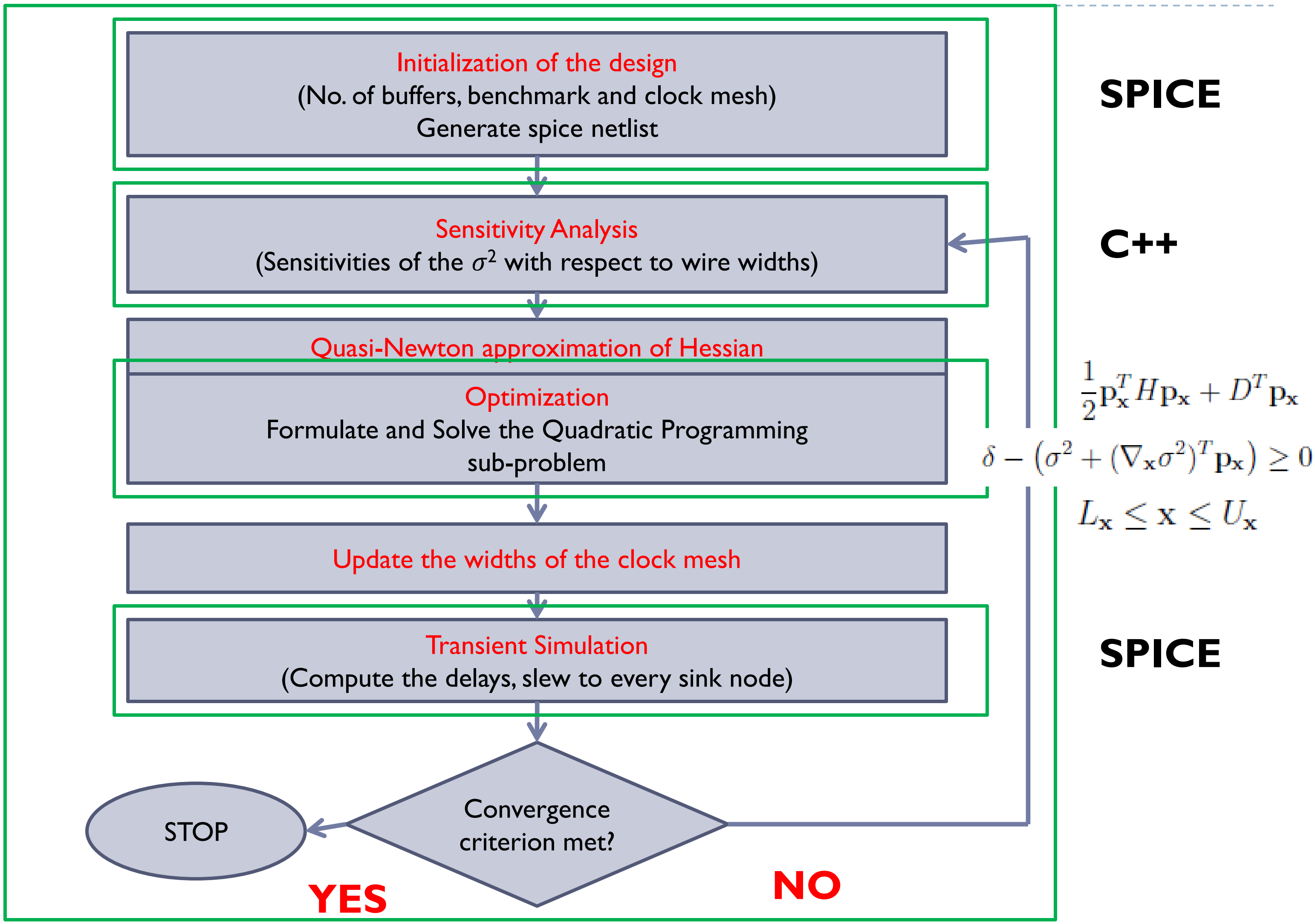
Sensitivity Analysis

- Sensitivity information of the original circuit obtained by convolution-like computation between transient waveforms of the original and the adjoint circuit.
- Compact gate model provides up to two orders of magnitude speedup over SPICE simulation while maintaining the same level of accuracy.

P. Li, Z. Feng and E. Acar. "Characterizing multistage nonlinear drivers and variability for accurate timing and noise analysis". In IEEE Trans. Very Large Scale Integration, pp 205 - 214, November 2007.

X. Ye and P. Li. "An application-specic adjoint sensitivity analysis framework for clock mesh sensitivity computation". In Proc. of IEEE International Symposium on Quality Electronic Design, pp 634 - 640, 2009.

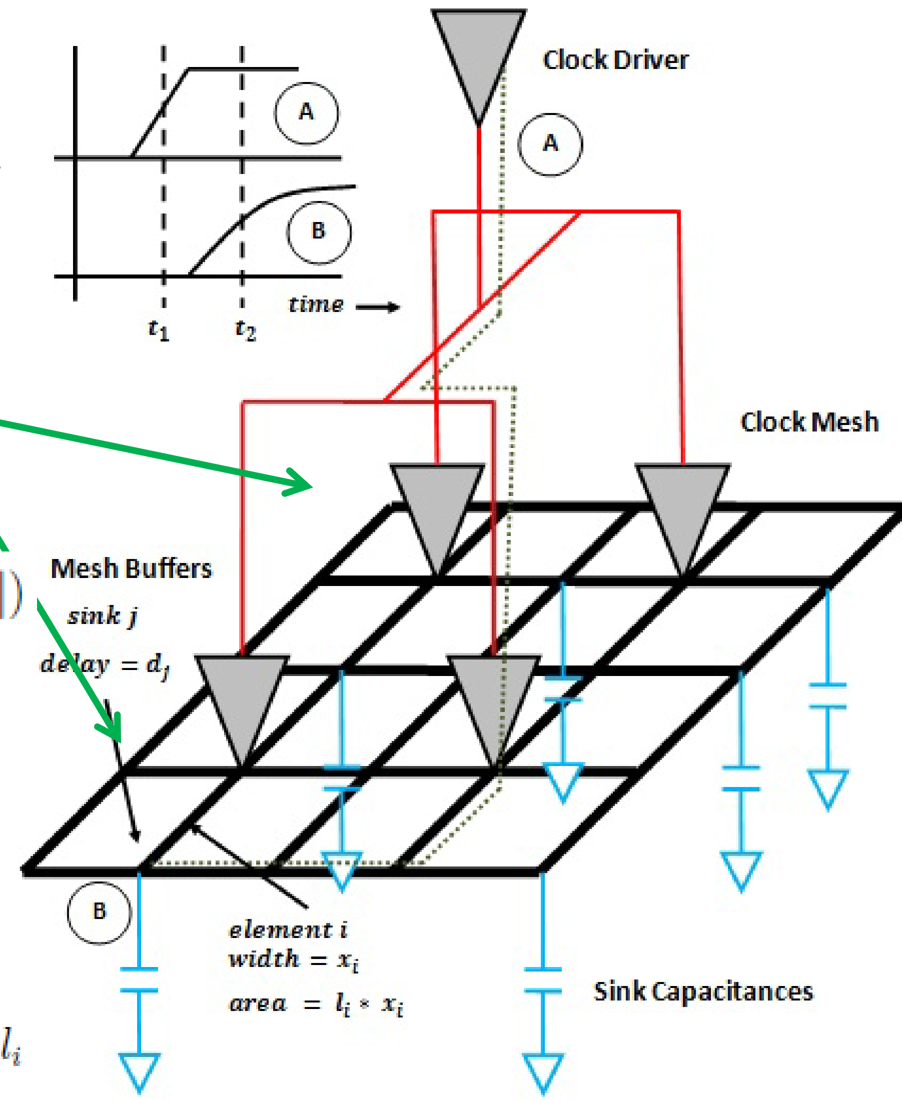
CMSSQP Framework



Results

- Experimental Setup
  - 65nm technology transistor models for the buffers
  - (m rows X n columns) mesh
  - Max skew $(\forall (i, j \in S)\ \mathrm{Max}|d_i - d_j|)$
  - Linux platform having two Intel Xeon E5410 quad-cores
  - ISCAS, ISPD benchmarks
  - Widths limited

Total area of the clock mesh $= \sum_{i \in I} w_i l_i$



Initial clock mesh design

H-Spice Results					
Benchmark	No. of sinks	Size of mesh	Initial (before CMSSQP)		
			Max Skew (ps)	Max Slew (ps)	Area (μm^2)
ispd09f11	121	12X12	12.3	70.8	17160
ispd09f12	117	12X12	16.9	55.2	20192
ispd09f21	117	16X16	20.9	67.5	31590
ispd09f22	91	16X16	16.2	51.5	17264
s1423	74	6X6	14.4	49.8	12439
s5378	179	13X13	7.4	26.2	27189
s15850	597	24X24	14.9	37.4	62903
r1	267	16X16	12.3	35.8	198589
r2	598	30X30	22.3	59.2	499557
r3	862	30X30	12.3	34.8	520200
r4	1903	40X40	22.3	51.0	910821
r5	3101	32X32	25.0	59.0	828123
Average			16.4	49.9	262168.9

Table I. Summary of initial clock mesh design results

Results after executing CMSSQP

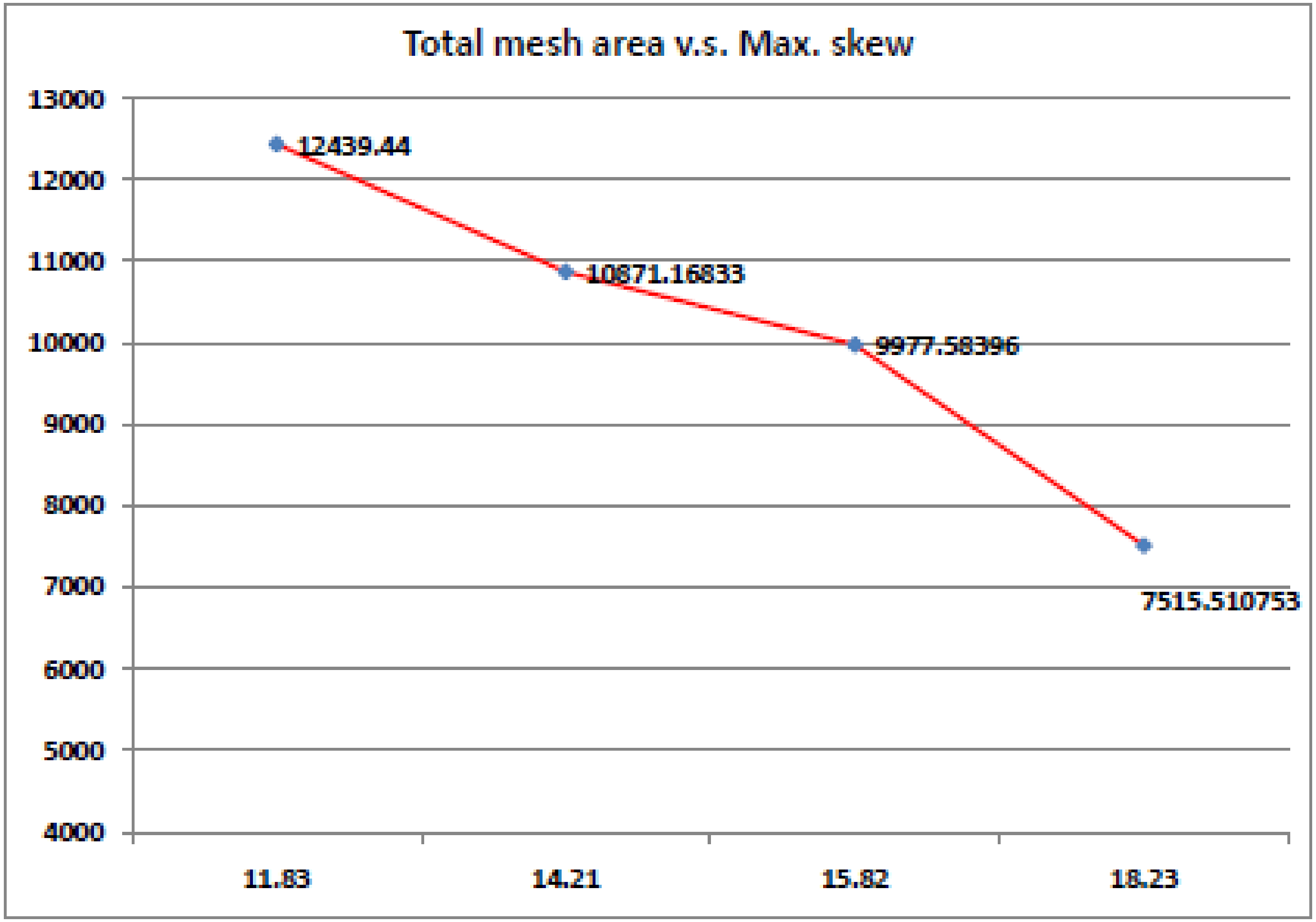
H-Spice Results					
Benchmark	No. of sinks	Size of mesh	Final (after CMSSQP)		
			Max Skew (ps)	Max Slew (ps)	Area (μm^2)
ispd09f11	121	12X12	12.2	71	9914
ispd09f12	117	12X12	17.4	52.1	11426
ispd09f21	117	16X16	22.9	67.3	21473
ispd09f22	91	16X16	19.9	51.1	14404
s1423	74	6X6	22	55.2	8614
s5378	179	13X13	9.9	25.4	18888
s15850	597	24X24	17.4	42.3	47150
r1	267	16X16	14.9	37.2	123931
r2	598	30X30	29.7	66.7	363002
r3	862	30X30	14.8	35	301505
r4	1903	40X40	29.9	61.4	552229
r5	3101	32X32	25.0	57.9	613754
Average			19.7	51.9	173857.5

Summary: Reduction in area

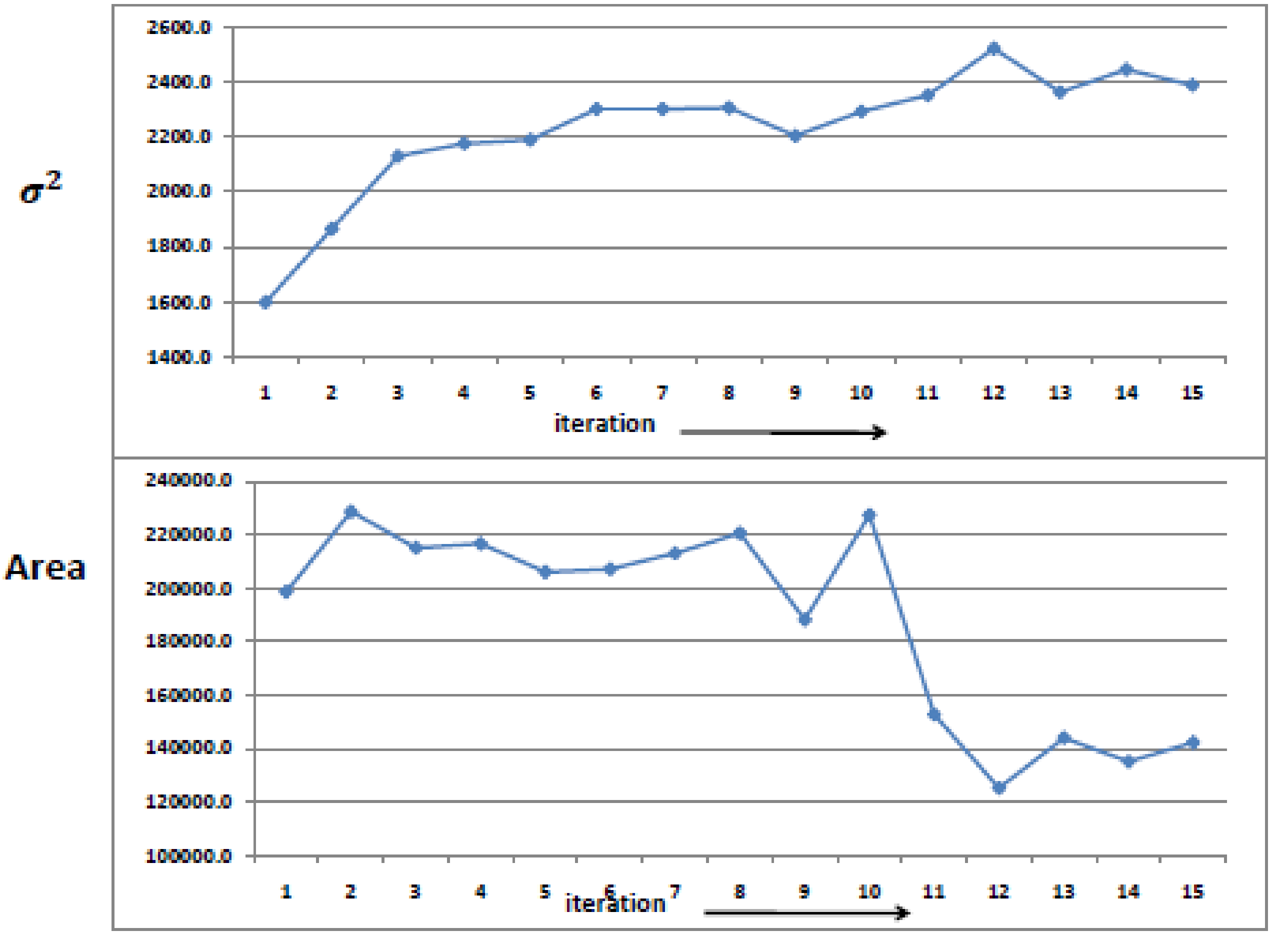
H-Spice Results				
Benchmark	No. of sinks	Size of mesh	Runtime (s)	Area reduction (%)
ispd09f11	121	12X12	465	42.2
ispd09f12	117	12X12	480	43.4
ispd09f21	117	16X16	640	32.0
ispd09f22	91	16X16	550	16.6
s1423	74	6X6	188	30.7
s5378	179	13X13	322	30.5
s15850	597	24X24	1430	25.0
r1	267	16X16	1197	37.6
r2	598	30X30	2954	27.3
r3	862	30X30	3115	42.0
r4	1903	40X40	10540	39.7
r5	3101	32X32	15440	25.9
Average			3110	33

Area-skew tradeoff by varying δ

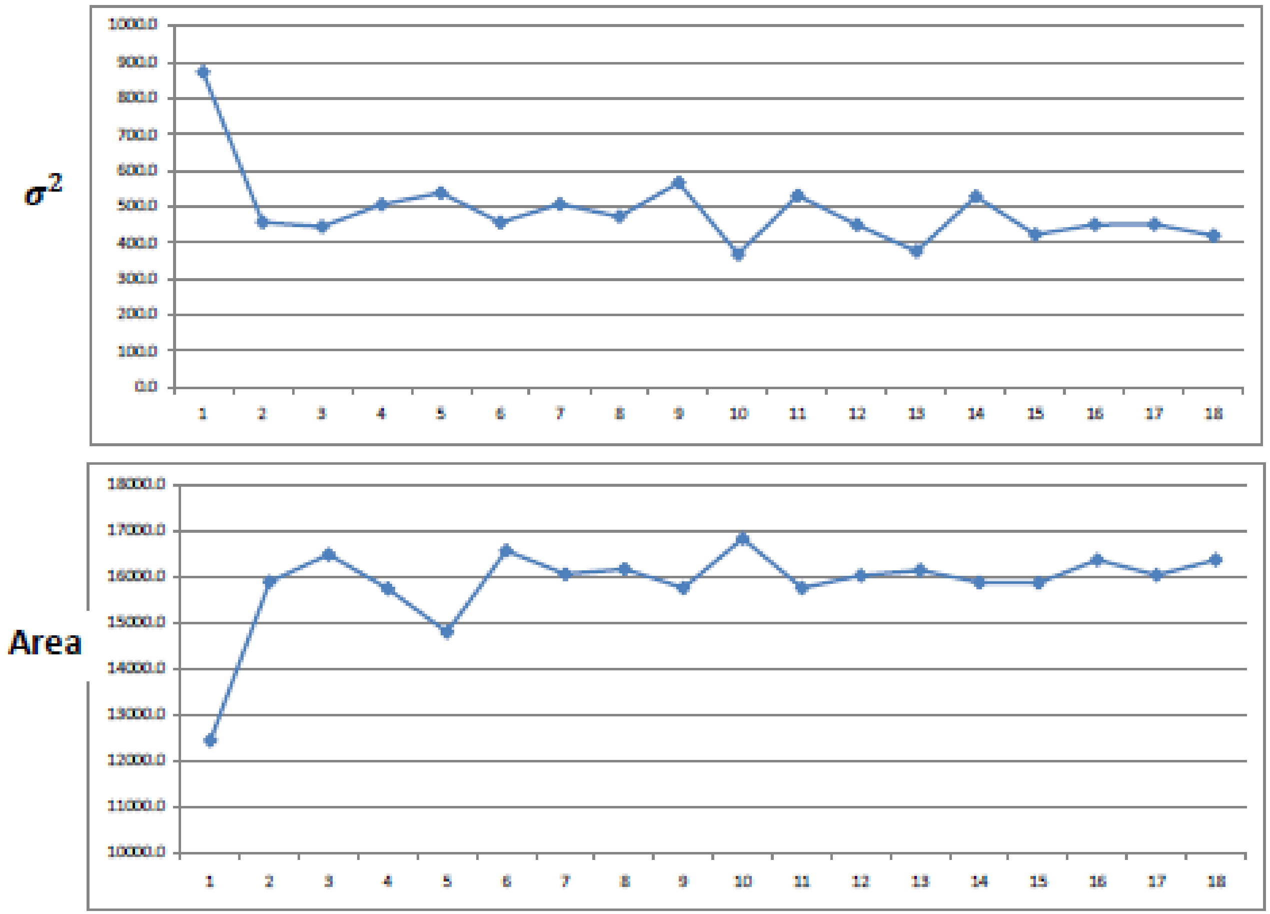
ISPD: ispd09f11



Case(a): ($\sigma^2 < \delta$), σ^2 , total clock mesh area in each iteration



Case(b): ($\sigma^2 > \delta$), σ^2 , total clock mesh area in each iteration



Conclusions & Future work

- Presented an algorithm for reduction of clock mesh area satisfying specified skew constraints in a clock mesh.
- Robust in dealing with any complex clock mesh network.
- First clock mesh sizing method that does systematic solution search and is based on accurate delay model.
- Experimental results achieved about 33% reduction in clock mesh area.
- Can be extended to size interconnects, mesh buffers simultaneously.

Thanks