Performance-Driven Interconnect Optimization

Charlie Chung-Ping Chen
Publications

- Performance-Driven Buffered Clock Tree Optimization Based on Lagrangian Relaxation, DAC, 1996.
- Fast and Exact Simultaneous Transistor and Wire-Sizing by Lagrangian Relaxation, ICCAD, 98.
Outline

- Interconnect Optimization
- Thesis
  - Wire Sizing
  - Buffer Sizing
  - Buffer Insertion
  - Interconnect Simulation
Interconnect Delay Trend
Interconnect Delay Trend
Figure 3  Calculated Gate and Interconnect Delay versus Technology Generation
Wire-Sizing
Buffer Sizing

$X_1$ $X_2$
Buffer Insertion
Interconnect Model
(area capacitance)
Driver Model

\[ R_d \]
Elmore Delay Computation

\[ D = R_d \left( C_1 + C_2 + C_3 + C_4 + C_L \right) + R_1 \left( \frac{C_1}{2} + C_2 + C_3 + C_4 + C_L \right) + R_2 \left( \frac{C_2}{2} + C_3 + C_4 + C_L \right) + R_3 \left( \frac{C_3}{2} + C_4 + C_L \right) + R_4 \left( \frac{C_4}{2} + C_L \right) \]
Uniform Wire-Sizing
Non-uniform Wire-Sizing

\[ f(x) \]

\[ R_d \]

\[ y \]

\[ X \]

\[ C_L \]
Optimal Wire-Sizing Function: Exponential Tapering

\[ ae^{-bx} \]
Interconnect Model
(Fringing Capacitance)
Optimal Wire-Sizing Function: 
$W$-function tapering

$$f(x) = ae^{-bx}$$
Lambert’s W function

\[ w e^w = x \]
Optimal Wire-Sizing Function

\[ ae^{-bx} \]

\[ \frac{-c_f}{2c_0} \left(1 + \frac{1}{W\left(\frac{-c_f}{ae^{-bx}}\right)}\right) \]

\[ \sqrt{b-ax} \]
Constrained Wire-Sizing  \( L \leq f(x) \leq U \)
Relations among the six types of functions
Wire-Sizing for Routing Trees
Weighted Sink Delay Optimization

Minimize

\[ D(f) = \sum_{i=1}^{s} \lambda_i D_i \]

Subject to

\[ L_i \leq f_i \leq U_i, \quad 1 \leq i \leq n, \]

where

\[ \sum_{i=1}^{s} \lambda_i = 1. \]
**Theorem:**

For each $w_i$, the optimal wire-sizing formula for a single line can be used to optimally resize $w_i$ by the following transformation: $R_d = R_i, C_L = C_i, L = L_i, \sigma = \mu_i \sigma$. 
Minimizing Area with Delay Constraints

Minimize \( A \)

Subject to

\[ D_i(f) \leq B_i, \quad 1 \leq i \leq s, \]
\[ L_i \leq f_i \leq U_i, \quad 1 \leq i \leq n. \]
Minimizing Area with Delay Constraints

Minimize $A$

Subject to $D_i(f) \leq B_i$, $1 \leq i \leq s$,

$L_i \leq f_i \leq U_i$, $1 \leq i \leq n$.

Lagrangian Relaxation Subproblem

Minimize $D'(f) = A + \sum_{i=1}^{s} \lambda_i (D_i(f) - B_i)$

Subject to $L_i \leq f_i \leq U_i$, $1 \leq i \leq n$. 
Minimizing Area with Delay Constraints

Minimize \[ D'(f) = A + \sum_{i=1}^{s}(\lambda_i D_i(f) - B_i) \]
Subject to \[ L_i \leq f_i \leq U_i, \ 1 \leq i \leq n. \]

Theorem:
Solving this Lagrangian subproblem is equivalent to solving the weighted sink delay problem with driver resistance changed to \[ R_d + \frac{1}{\xi \lambda}, \] where \( \lambda = \sum_{i=1}^{s} \lambda_i. \)
Algorithm Framework

Adjust Lagrange Multipliers (Sink Weights)

Minimize Weighted Sink Delays
Minimizing Maximum Delay

Minimize \[ D_{\text{max}} \]
Subject to \[ D_i(x) \leq D_{\text{max}}, \; 1 \leq i \leq s, \]
\[ L_i \leq f(x_i) \leq U_i, \; 1 \leq i \leq n, \]
\[ D_{\text{max}} > 0. \]

\[ \downarrow \]

Lagrangian Relaxation Subproblem

Minimize \[ D_{\text{max}} + \sum_{i=1}^{s} \lambda_i (D_i - D_{\text{max}}) \]
Subject to \[ L_i \leq f(x_i) \leq U_i, \; 1 \leq i \leq n, \]
\[ D_{\text{max}} > 0. \]
Minimize $\alpha D_{\text{max}} + \beta \text{Power} + \gamma \text{Area}$

Subject to $D_i(x) \leq D_{\text{max}}, \ 1 \leq i \leq s,$

$L_i \leq x_i \leq U_i, \ 0 \leq i \leq n + m,$

$D_{\text{max}} > 0.$
Simultaneous Buffer and Wire-Sizing
Uniform Wire-Sizing

\[ f_1 \]

\[ w_1 \]

\[ w_2 \]

\[ w_3 \]

\[ w_4 \]

\[ w_5 \]

\[ w_7 \]
Upperbound, Lowerbound, Skew
Buffer Insertion
Spec-based Buffer Insertion

- Given a routing tree, possible buffer insert location, required arrival times at receivers, max slope constraint, and polarity requirement
- A library of buffers: $B_1, B_2, \ldots, B_n$
- Insert buffers to satisfy the spec (maximum delay, delay bounds at each receiver and maximum slope)

Potential buffer location:
Problem formulation

- Combinations of the following:
  - Buffer Insertion
  - Buffer-Sizing
  - Wire-Sizing

- Goals and Constraints
  - Minimize the Maximum Delay
  - Satisfy delay constraints at each receiver
  - Repeater Insertion Location Constraints
  - Maximum Slope constraint
  - Polarity constraints.
Current Issues

• Previous solutions
  • Exhaustive enumeration-> Exponentially Growing
  • First Ginneken, and then, Lillis, suggested a dynamic programming approach which can get optimal solution for delay under the Elmore delay model.
  • Provides very useful information like power-delay curves
  • Problems: not accurate, doesn’t consider reliability issues, runtime and storage already high.
Brief Algorithm Description

- Algorithm:
  - Traverse circuit in a bottom-up manner
  - Enumerate all the possible solutions and prunes out sub-optimal solutions dynamically.

- How do we know which solution to kill?
  - Violate the constraints
  - If there is another solution cost less and achieve more in all aspects?
    - Number of buffers
    - Maximum delay
    - Polarity
    - Area, Power ...
Example
Problems

• How to calculate the gate and interconnect delay accurately and efficiently?
  • A naive approach: Repeatedly calculate the delay of the subtree by calling AWE or SPICE (causing $O(N)$ penalty for each solution). However, the runtime is already high (proportional to $N^2$).
  • An efficiently hierarchical delay computation method is needed

• How to include slope into consideration?
Accurate Load Model (Effective Capacitance)

- The total-net capacitance is no longer a valid load model -- the second-order $\pi$-load driving-point admittance approximation is more accurate.

![Diagram showing the comparison between different load models](image)

**too pessimistic, up 30% error**

- **Total capacitance load model**
  - 0.178 pF
  - 0.356 pF
  - 0.178 pF
- **Second-order $\pi$-load model**
  - 232.8 Ω
  - 232.8 Ω
  - 232.8 Ω
- **Effective capacitance load model**
  - 1.07 pF
  - 364.2 Ω
  - 0.76 pF

**equal average currents**
Accurate Repeater Model (Voltage Ramp)

- Several timing analyzers model the gate by a single resistor
  - Errors of up to 30% have been reported
- The proposed gate delay model is a fixed-resistor driven by a ramp voltage source

- Voltage ramp parameters, $t_0$ and $t_x$, are determined from the Gate characteristic equations

![Diagram](image-url)
Accuracy of Voltage Ramp Model

- Model voltage source
- Model output
- Actual output

Error histogram of 16000 Merced nets
How to calculate $\pi$-load hierarchically

- There exists a simple way to calculate the $\pi$ load (actually it can handle arbitrary higher order approximation) hierarchically.

\[ Y_{1}(s) \]
\[ Y_{2}(s) \]
\[ R(s) \]
\[ Y_{eq}(s) \]

\[ Y_{new}(s) = \frac{1}{R(s) + Y_{1}(s) + Y_{2}(s)} \]

Taylor Expansion:

\[ Y_{eq}(s) = y_{1}s + y_{2}s^{2} + y_{3}s^{3} + y_{4}s^{4} + y_{5}s^{5} + y_{6}s^{6} \]
What about wires?

\[ V_2(s) = V_1(s) \]

**Taylor Expansion:**

\[ H(s) = m_0 + m_1 s + m_2 s^2 + m_3 s^3 + m_4 s^4 + m_5 s^5 + m_6 s^6 \]

**Transfer function computation**
Many Stages

\[ V_3(s) = \frac{Y_1(s)}{Y_1(s) + Y_{eq}(s)} \]

Transfer function computation
What about Trees?

Keep track the worse sink’s transfer function
Hierarchical moment computation -- REX

- Assume in general
  - \( H(s) = 1/[b_0 + b_1s + b_2s^2 + b_3s^3 + b_4s^4 + b_5s^5] \)
  - \( Y(s) = y_1s + y_2s^2 + y_3s^3 + y_4s^4 + y_5s^5 + y_6s^6 \)

- Across a capacitor
  - \( H'(s) = H(s) \)
  - \( Y'(s) = Y(s) + Cs \)

- Across a resistor
  - \( H'(s) = H(s)/[1 + R Y(s)] \)
  - \( Y'(s) = Y(s)/[1 + R Y(s)] \)

- Base case at the receiver
  - \( H(s) = 1 \)
  - \( Y(s) = C_L s \)
  - \( C_L \)
With buffer inserted and slope consideration?

Interpolate and extrapolate the delay at receivers
Results

- Sample net -- 10000 µm line on m1pm2 broken into 40 segments
- Delay before optimization -- 2405 ps
- Time for optimization -- 27 seconds on RS6K

<table>
<thead>
<tr>
<th>Stages</th>
<th>OUR</th>
<th>SPICE</th>
<th>% error</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2467</td>
<td>2405</td>
<td>2.5</td>
</tr>
<tr>
<td>1</td>
<td>1736</td>
<td>1761</td>
<td>-1.4</td>
</tr>
<tr>
<td>2</td>
<td>1388</td>
<td>1404</td>
<td>-1.1</td>
</tr>
<tr>
<td>3</td>
<td>1267</td>
<td>1301</td>
<td>-2.6</td>
</tr>
<tr>
<td>4</td>
<td>1218</td>
<td>1274</td>
<td>-4.3</td>
</tr>
</tbody>
</table>
Cost-Performance Curves

# of repeaters vs max delay (ps)

- **OUR**
- **SPICE**
Case Study:

Roses report:  
Delay = 1529 ps

Cse report:  
Delay = 1548 ps
Case Study: Manual Result (8 buffer)

Roses report:
Delay = 1047 ps

Cse report:
Delay = 1053 ps

max slope: 294

slope: 100
Case Study: Our Result (5 buffers)

Roses report:
Delay=953 ps

SPICE report:
Delay=1017 ps
# of wire segment vs maximum delay

Maximum delay

# of repeaters
**Conclusion**

- Buffer model provides about 5~7% accuracy relative to SPICE
  - The total net capacitance is no longer a valid load approximation
  - Using accurate models aid
- Hierarchical moments computation of RC delays and slopes
  - New Moment-matching methods provide efficient and accurate delay calculation for RC nets especially for hierarchical moment computation
- Dynamic programming approaches applied to buffer insertion
  - Hierarchical moment methods for efficient RC delay computation