Interconnect impact on circuit performances in Deep Sub-Micron technologies

Cristiano Forzan
Central R&D – NVM Design Platform

Bologna, June 2004

STMicroelectronics
Outline

- CMOS device and interconnect scaling overview
- Introduction to Static Timing Analysis
- Interconnect resistance effects
- References
CMOS technology scaling overview

1μm technology (1990) 0.13μm technology (2001)

800 KTx

1cm
Ideal CMOS scaling

<table>
<thead>
<tr>
<th>Scaled Parameter</th>
<th>Ideal Scaling Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>$W, L, \text{tox}$</td>
<td>$S$</td>
</tr>
<tr>
<td>$N_A$</td>
<td>$1/S$</td>
</tr>
<tr>
<td>$V_{dd}, V_{th}$</td>
<td>$S$</td>
</tr>
</tbody>
</table>
## Interconnect reverse scaling

<table>
<thead>
<tr>
<th></th>
<th>Local Wiring</th>
<th>Global Wiring</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linewidth &amp; Spacing</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>Wire Thickness</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>ILD Thickness</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>WireLength</td>
<td>S</td>
<td>1/sqrt(S)</td>
</tr>
<tr>
<td>Resistance</td>
<td>1/S^2</td>
<td>1/S^2</td>
</tr>
<tr>
<td>Capacitance</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>RC delay</td>
<td>1</td>
<td>1/S^3</td>
</tr>
<tr>
<td>Current density</td>
<td>1/S</td>
<td>1/S</td>
</tr>
</tbody>
</table>
DSM scaling impact on delay

- Gate delay scales
- Local interconnect delay scales
- Global interconnect delay dominates
  - Global interconnects do not scale with feature sizes

(from ITRS 99)
Quasi-ideal interconnect scaling

Keeping wire resistance low
- Increase the wire aspect ratio (AR = height / width)
- Use copper instead of aluminium
Interconnect capacitances

- **Metal 2**
- **Metal 1**
- **Substrate**
  - **CFRINGE**
  - **CAREA**
  - **CCOUPLING**
  - **CCROSSOVER**
Interconnect capacitance trends

- Interconnect average length and routing density increase
- Wire aspect ratio increases
- Coupling capacitance dominates over ground capacitance
- Interconnect capacitance dominates gate capacitance
Traditional top-down design flow

- Any top-down flow requires some predictive capabilities
- Traditional top-down flows are effective when performances can be predicted with the wireload model
Interconnect impact on design flow

- The evolutionary response to the DSM problem has been a **stronger integration** between the synthesis and the place&route process.

- This is what Physical Synthesis does today.

**Revolutionary breakthrough?**
Outline

- CMOS device and interconnect scaling overview

- Introduction to Static Timing Analysis
  - Timing verification
  - Delay calculation
  - Constraints

- Interconnect resistance effects

- References
Dynamic Timing Simulation

Advantages

- Can check asynchronous interfaces
- No need to specify false and multicycle paths, clock model, etc
- Can be very accurate (SPICE)

Disadvantages

- Analysis quality depends on stimulus vectors
- Non-exhaustive – virtually impossible to check every path
- Long run times
Static Timing Simulation

Advantages
- Fast – design is analyzed in one pass, for one clock cycle
- Exhaustive – checks all topological paths in design
- Does not require verification environment

Disadvantages
- Less accurate
- Must define timing requirements / exceptions
- Difficulty handling asynchronous designs, false paths
Principles of Static Timing Verification

- Netlist is represented as DAG – Directed Acyclic Graph
- Delay values associated with nodes (Cells) and links (Nets)
- Total path delay is the sum of Path delay values
Path-based STA

- Timing information is associated with topological paths
  - First, extract all possible topological paths
  - Next, for each path calculates its delay and compare it with endpoint (required) value

Path-based:

- $2+2+3 = 7$ (OK)
- $2+3+1+3 = 9$ (OK)
- $2+3+3+2 = 10$ (OK)
- $5+1+1+3 = 10$ (OK)
- $5+1+3+2 = 11$ (Problem!)
- $5+1+2 = 8$ (OK)
Timing Paths Types

Four types of Timing Paths

- Input to Register (Synchronous)
- Register to Register (Synchronous)
- Register to Output (Synchronous)
- Input to Output (Asynchronous)

Each path has:

- StartPoint (Input port or FF output)
- EndPoint (Output port or FF input)
- Calculated value for path delay
Outline

- CMOS device and interconnect scaling overview

- Introduction to Static Timing Analysis
  - Timing verification
  - Delay calculation
  - Constraints

- Interconnect resistance effects

- References
Path delay calculation

- The actual path delay is the sum of the Net and the Cell delays along the timing path
  - **Net Delay** – total time for charging/discharging all the parasitics of a given net. It is a function of:
    - Net capacitance
    - Net resistance
  - **Cell Delay** – delay arc between corresponding input and output ports of the cell. It is a function of:
    - Input transition time (Slew)
    - Total output load
    - Process parameters (Temperature, Power supply)
Net Delay Calculation (PreLayout)

Wire Load Model

- Net length is a function of net fanout and chip area
- For a given area, averages of R and C are estimated for different fanouts
- Net delay is calculated simply as R * C

<table>
<thead>
<tr>
<th>Capacitance as a function of fanout:</th>
<th>Resistance as a function of fanout:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1   0.015</td>
<td>1   0.012</td>
</tr>
<tr>
<td>2   0.030</td>
<td>2   0.016</td>
</tr>
<tr>
<td>3   0.046</td>
<td>3   0.020</td>
</tr>
</tbody>
</table>

For fanout = 3
Net delay = 0.046 * 0.020
Cell Delay Calculation

Cell delay is a function of
- Input Transition Time (calculated by previous gate)
- Total Output Capacitance (Net cap + sum of attached Pin Caps)
- Operating Conditions
Cell Delay and Slew Calculation

- Delay is calculated using 2-Dimensional Nonlinear Delay Model. Main calculation part is interpolation between nearest table values:

<table>
<thead>
<tr>
<th>Input Transition (ns)</th>
<th>Total Cload (fF)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.2  0.3  0.4  0.5</td>
</tr>
<tr>
<td>0.1</td>
<td>0    3    4.5   6    7</td>
</tr>
</tbody>
</table>

- A similar model is used for the Cell Slew calculation

- After calculation, delay is scaled by the operating conditions:
  - Final_delay = Table_delay * Kvoltage * Ktemp * Kprocess
Types of Cell Timing Arcs

- Combinatorial Cells: from all inputs to output:

- Sequential Cells: from Clock Input to outputs (propagation delay) and from Clock Input to Data Input (setup, hold checks):
Outline

❖ CMOS device and interconnect scaling overview

❖ Introduction to Static Timing Analysis
  – Timing verification
  – Delay calculation
  – Constraints

❖ Interconnect resistance effects

❖ References
Path Constraints Types

- Setup time (Input-to-Reg and Reg-to-Reg paths)
- Hold time (Input-to-Reg and Reg-to-Reg paths)
- Input delay (Input-to-Reg and Input-to-Output paths)
- Output delay (Reg-to-Output and Input-to-Output paths)
Setup and Hold Times

 SETUP: amount of time data must be stable at the data pin of FF before the clock capturing edge

 HOLD: amount of time data must remain stable at the data pin of FF after the clock capturing edge

\[
TD - T_{CLK} > T_{Hold}
\]

Hold Check refers to the same clock edge:

\[
(T_{CLK} + T_{Period}) - T_D > T_{Setup}
\]

Setup Check refers to the next Clock edge (add clock period):
Other constraints: Design Rules Check

- Design Rules are electrical checks performed on each gate. They are defined in the technology library.
- Checking min and max limits for:
  - Net / Port capacitance
  - Net Transition times
  - Net fanout

Max capacitance example:
Outline

- CMOS device and interconnect scaling overview
- Introduction to Static Timing Analysis
- Interconnect resistance effects
- References
Interconnect resistance impact

- For technologies up to 0.35µm delay estimation based on the wireload model can be accurate
- Few iterations for timing closure

- For DSM technologies (0.25µm and below) RC distributed effects become dominant
- Many iterations and no guarantee to achieve timing closure
Wire resistance impact on delay

\[ T_{\text{stage}(A-R)} = T_{\text{gate}(A-Z)} + T_{\text{wire}(Z-R)} \]

\[ f(T_{\text{IN}}, C_Z) \]
The propagation time across the gate decreases due to the shielding resistance effect.

The effective capacitance concept is introduced.
The driving cell can be represented by a Thevenin equivalent model.

The output waveform can be obtained via convolution of the driving point voltage waveform with the transfer function $H_{ZR}$.

SPICE: generally too much expensive.
Elmore delay

\[ T_D = \int_0^\infty t h(t) \, dt \]

- It is an upper bound on the delay of an RC tree response: \( T_D \geq \tau \)
Elmore delay calculation

\[ T_{D_i} = \sum_{k=1}^{N} R_{ki} C_k \]

\[ T_{D5} = R_1 C_1 + \]
\[ (R_1 + R_2)C_2 + \]
\[ (R_1 + R_2 + R_3)C_3 + \]
\[ (R_1 + R_2 + R_3 + R_4)C_4 + \]
\[ (R_1 + R_2 + R_3 + R_4 + R_5)C_5 + \]
\[ R_1 C_6 + \]
\[ R_1 C_7 \]

Elmore delay can be calculated very efficiently
- Linear time w.r.t the RC tree size
Model order reduction

\[ H(s) = \frac{a_0 + a_1 s + a_2 s^2 + \cdots + a_n s^n}{1 + b_1 s + b_2 s^2 + \cdots + b_m s^m} \]

- Linear, time-invariant circuit
- The transient response is generally dominated by a small number of poles
- Set of dominant poles is obtained by using moment-matching techniques (AWE, PRIMA, Arnoldi …)
Moment matching technique (I)

- The goal of model order reduction techniques is to obtain a set of \( q \) approximate dominant poles efficiently:

\[
H^*(s) = \frac{k_1}{s-p_1} + \frac{k_2}{s-p_2} + \cdots + \frac{k_q}{s-p_q}, \quad q << m
\]

\[
h^*(t) = k_1 e^{p_1 t} + k_2 e^{p_2 t} + \cdots + k_q e^{p_q t}
\]

- The \( i \)-th moment of the function \( f(t) \) is defined as:

\[
m_i = \frac{(-1)^i}{i!} \int_0^\infty t^i f(t) dt
\]

\[
m^*_i = \frac{k_1}{p_1^{i+1}} + \frac{k_2}{p_2^{i+1}} + \cdots + \frac{k_q}{p_q^{i+1}}
\]
The moments of $H^*(S)$ are forced to match the actual circuit moments:

\[
\frac{k_1}{p_1} + \frac{k_2}{p_2} + \cdots + \frac{k_q}{p_q} = m_0
\]

\[
\frac{k_1}{p_1^2} + \frac{k_2}{p_2^2} + \cdots + \frac{k_q}{p_q^2} = m_1
\]

\[
\vdots
\]

\[
\frac{k_1}{p_1^{2q}} + \frac{k_2}{p_2^{2q}} + \cdots + \frac{k_q}{p_q^{2q}} = m_{2q-1}
\]
Moment matching technique (III)

- Once the reduced order model is obtained, the output waveform is calculated via convolution with the input signal.
  - Example: saturated ramp

\[
\forall \text{ For time interval } t_0 \leq t \leq t_1 \\
\nu_Z(t) = \frac{V_{DD}}{t_1 - t_0} \sum_{k=1}^{q} r_k \left( e^{p_k(t-t_0)} - 1 - p_k(t-t_0) \right)
\]

\[
\forall \text{ For time interval } t \geq t_1 \\
\nu_Z(t) = \frac{V_{DD}}{t_1 - t_0} \sum_{k=1}^{q} r_k \left( e^{p_k(t-t_0)} - e^{p_k(t-t_1)} - p_k(t_1 - t_0) \right)
\]
Asymptotic Waveform Evaluation

- A linear time-invariant network can be represented by the state equation:
  \[ \dot{x} = Ax + Bu \]

- In the frequency domain:
  \[ sX(s) - x(0) = AX(s) + BU(s) \]

- The network transfer function is expressed by:
  \[ H(s) = \frac{X(s)}{U(s)} = (sI - A)^{-1} B \]
Asymptotic Waveform Evaluation (II)

- The transfer function is the Laplace transform of the impulse response:

\[ H(s) = \int_{0}^{\infty} e^{-st} h(t) \, dt \]

- The coefficients of the MacLaurin series expansion are the moments of the impulse response

\[ H(s) = \sum_{k=0}^{\infty} s^k \frac{(-1)^k}{k!} \int_{0}^{\infty} t^k h(t) \, dt = \sum_{k=0}^{\infty} s^k m_k \]

- From the previous representation:

\[ H(s) = (sI - A)^{-1} B = -A^{-1} (I + A^{-1} s + A^{-2} s^2 + \cdots) B \]
Moment determination

- The moments are computed recursively

\[ m_0 = A^{-1}B \]

\[ m_1 = A^{-2}B = A^{-1}m_0 \]

\[ m_q = A^{-(q+1)}B = A^{-1}m_{q-1} \]

- State matrix is inverted only once
Moment computation for RC trees

- Nodal Analysis instead of State Analysis
  - Based on KCL and KVL
  - It is not a general methodology but it is easier than State Equation

- Nodal Analysis formulation:

\[ \dot{v} = C^{-1}Gv + C^{-1}Bu \]

- $u$ is a unit step input voltage
- $C$ is the diagonal capacitance matrix
- $G$ is the conductance matrix
Nodal Analysis

- First moments generation
  - Replace input drivers with their final value
  - Perform dc analysis
  - Moments are the voltages across the capacitors

- Successive moments
  - Set input drivers to zero
  - Replace each capacitor with a current source \( I = m_{k-1}C \)  
    \( (V = m_{k-1}L) \)
  - Perform dc analysis
  - Moments are the voltages across the current sources
Outline

- CMOS device and interconnect scaling overview
- Introduction to Static Timing Analysis
- Interconnect resistance effects
- References
References (I)

References (II)

Our works on this field (I)


- D. Pandini, P. Scandolara, C. Guardiani “Reduced order macromodel of coupled interconnects for timing and functional verification of sub half-micron IC designs”, Asia and Pacific DAC 1997
Our works on this field (II)


- B. Franzini, C. Forzan “Crosstalk Aware Static Timing Analysis Environment”, ESNUG 2001