Appendix 2

Clock Skew Model 2

Payman Zarkesh-Ha, Tony Mule', James D. Meindl [ZAR-98]

1. Introduction

Performance of high-speed synchronous digital systems is reduced significantly by clock skew of the clock distribution network. The concept of zero clock skew balanced networks has been proposed, however clock skew created by process parameter variations is still unavoidable. It is, therefore, imperative to characterize the clock skew components due to process parameter variations.

These authors describe a compact model to enable first-order estimation for on-chip clock skew as a function of device, interconnect and system parameter variations. Unlike previous models that describe qualitative behaviour of clock skew components, the new model provides a closed form expression for each clock skew component. This model provides a statistical expression (in function of both, process and design parameters, % of variation) for the clock skew in a balanced clock network (H-Tree), but this expression can be easily modified and applied to any clock network.

2. Clock skew components

Clock skew appears principally from unequal clock path lengths from the clock source to the clocked registers. To equalize line lengths, and thus reduce the clock skew, a common practice is the use of a balanced clock network. In this way the nominal value of skew becomes zero and clock skew reduces to the variations of the clock path from the clock generator to the registers.

These variations are originated by process and circuit parameter tolerances. We can distinguish the following parameter variations:

- ***Device Parameter Variations***

In the IC fabrication process, all device parameters are subject to deviations from their nominal values. Statistical models have been developed for transistor parameters such as **threshold voltage** (ΔV_T), **gate oxide thickness** (Δt_{ox}), and **effective channel length** (ΔL_{eff}).

- ***Interconnect Parameter Variations***

Interconnect width (ΔW_{int}) and **thickness** (Δt_{int}) and **interlevel dielectric thickness** (ΔT_{ILD}) variations are the main parameters of interest. As technology advances, the number of interconnect layers increases, and the interconnect lines become more non-uniform. This non-uniformity, which is caused by manufacturing processes, produces large variations of interconnect parameter values.

Chemical mechanical polishing (CMP) is the manufacturing process for planarization of metal and ILD layers that greatly reduces the ILD non-uniformity in multilayer structures. However, the CMP process still doesn't eliminate interconnect parameter variations completely.

- ***System Parameter Variations***

Besides process parameter variations, which are mainly the tolerances of device and interconnect physical parameters, system level fluctuations may create clock skew. **Power supply voltage fluctuation** (ΔV_{DD}), **temperature variations** ($\Delta \mathcal{T}$), and **non-uniform distribution of clocked registers** (ΔC_L) are considered as system level parameter variations.

3. Clock skew components

The goal of the derivation of a clock skew model is to understand the impact of process and system variations in an ideally zero skew clock network distribution. The most common strategy to ensure zero nominal clock skew, which is often used for distributing high frequency clock signals in digital systems, is a symmetric H-tree structure.

Although a model is derived here especially for the symmetric H-tree structure, the model can be easily modified for any balanced clock tree network.

3.1. Assumptions

Although the growing importance of on-chip transmission line effects has been predicted, the difficulty of modelling and simulating them, non-uniform transmission lines using existing CAD tools has prevented consideration of these effects in most cases. There are some methods, however, to reduce the inductance effects in actual design.

For example, there are clock distribution networks described, in which ground return path wiring has been implemented on the two metal levels above and below the clock wire to reduce inductance effects. To simplify the derivation of a clock skew model, good return path wiring has been assumed to surround the clock wiring network. Therefore, in this simplified study **the inductance effect is ignored**. Moreover, without lose of generality, it is assumed that **the clock network is a balanced H-tree structure**. This model, however, can be easily modified for any balanced clock tree network.

3.2. The Complete Clock Skew Model

Figure 1 shows a symmetric H-tree clock distribution with n=4 levels of H-tree branches. At the end of the 4th level, drivers are implemented to feed the clock signal to

all registers in the sub-blocks. It's important to underline that, in this model, the H-tree clock distribution system doesn't have any buffer at the split points where the i-level branch is divided in two (i+1)-level branches. Therefore, the wires width has to be designed so that i-level branch width has to be double as wide as (i+1)-level branch. It's necessary to avoid signal reflections in the split points.

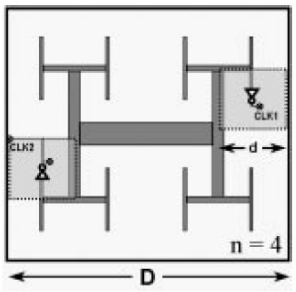

Figure 1: Symmetric H-tree structure.

The total clock skew, by definition, is the time difference between the maximum and minimum delays as illustrated in Figure 2.

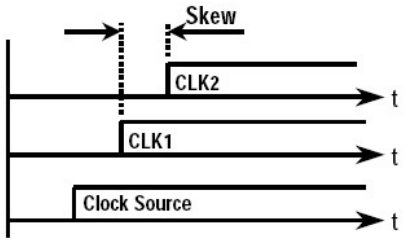

Figure 2: Clock skew between the points CLK1 and CLK2 (Fig 1).

Any H-tree circuit (Figure 1) can be simplified in the following equivalent circuit shown in Figure 3.

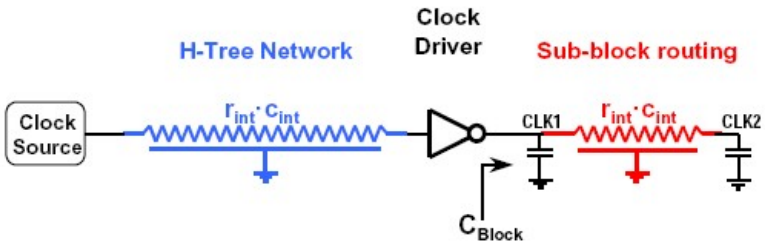

Figure 3: Equivalent circuit of clock H-tree network.

Using the equivalent circuit, the delay of the entire clock network of Figure 1 is divided into three parts:

- ***Interconnect delay from the clock source located at the centre of the chip through the H-tree to the driver***

Assuming that the H-tree network is driven by a single driver and the line capacitance of the H-tree network is much greater than the transistor input capacitance of a sub-block clock driver, then the interconnect delay expression for a distributed RC line using Sakurai's model (*50%* of time delay) is:

$$T_{H-tree} = 0.4 \cdot (r_{\text{int}} c_{\text{int}}) \cdot l^2 + \frac{\sqrt{\varepsilon_r}}{c_0} \cdot l \qquad (1)$$

where l is the length of the H-tree network, r_{int} and c_{int} are the distributed resistance and capacitance of the line, ε_r is the relative dielectric constant of the ILD material, and c_o is the speed of light in free space.

Because of the wires in clock distributions are often much wider than the minimum wire width, the fringing capacitance is negligible compared to parallel plate wiring capacitance. Using the expression for the length of the H-tree versus die size, D, and the number of H-tree levels, n, then (1) becomes:

$$r_{\text{int}} \cdot c_{\text{int}} = \frac{\rho \cdot \varepsilon_r}{t_{\text{int}} \cdot T_{ILD}}, \quad l = D \cdot \left(1 - \frac{1}{2^{n/2}}\right) \quad \Rightarrow$$

$$T_{H-tree} = 0.4 \cdot \left(\frac{\rho \cdot \varepsilon_r}{t_{\text{int}} \cdot T_{ILD}}\right) \cdot D^2 \cdot \left(1 - \frac{1}{2^{n/2}}\right)^2 + \frac{\sqrt{\varepsilon_r}}{c_0} \cdot D \cdot \left(1 - \frac{1}{2^{n/2}}\right) \qquad (2)$$

where t_{int} is interconnect thickness, T_{ILD} is interlevel dielectric thickness, and ρ is the line resistivity.

- ***Transistor delay of the sub-block clock driver***

The clocked registers within sub-blocks are assumed to be randomly placed and routed, therefore the delay expression for the time delay of the sub-block drivers is very simple (50% of time delay in Sakurai's Model):

$$T_{driver} = 0.7 \cdot R_0 \cdot C_L \qquad (3)$$

where R_0 is the average inverter resistance and C_L is the capacitive load. Using the expression for the output resistance of an inverter in saturation, expression (3) becomes:

$$T_{driver} = 0.7 \cdot \left(\frac{L_{eff} / W}{\mu \cdot C_{ox} \cdot (V_{DD} - V_T)} \right) \cdot C_L \qquad (4)$$

where, L_{eff} and W are transistor channel length and width respectively, μ is the mobility, C_{ox} is the gate oxide capacitance, and V_{DD} and V_T are supply and transistor threshold voltage respectively.

- ***Internal wire routing delay within the sub-block from the clock driver to registers***

The wiring delay inside the sub-block is computed in a similar way from (1) except that the length of wire, *l*, is the distance from centre to the corner of the sub-block ($d = D/2^{n/2}$).

$$T_{Sub-Blk} = 0.4 \cdot (r_{\text{int}-sub} c_{\text{int}-sub}) \cdot d^2 + \frac{\sqrt{\varepsilon_r}}{c_0} \cdot l = 0.4 \cdot \left(\frac{\rho \cdot \varepsilon_r}{t_{\text{int}-sub} \cdot T_{ILD-sub}} \right) \cdot \frac{D^2}{2^n} + \frac{\sqrt{\varepsilon_r}}{c_0} \cdot \frac{D}{2^{n/2}} \qquad (5)$$

Because of, in general, the placement of clocked registers is not uniform, the routing length inside the sub-block is not the same always. For example, in Figure 1 the clock signal at the point CLK2 arrives later than CLK1. This delay, which is often called ***internal clock skew***, in the worst case is given by (5):

The overall delay of the entire clock distribution network, from the clock source to the clocked registers, is $T_{Delay} = T_{H\text{-}tree} + T_{Driver} + T_{Sub\text{-}Blk}$. Since the sub-block size is often much less than the chip size, the wiring delay within the sub-block, $T_{Sub\text{-}Blk}$, can be ignored. Therefore the total delay is given by:

$$T_{Delay} \approx T_{H-tree} + T_{driver} = 0.4 \cdot \left(\frac{\rho \cdot \varepsilon_r}{t_{\text{int}} \cdot T_{ILD}} \right) \cdot D^2 \cdot \left(1 - \frac{1}{2^{n/2}} \right)^2 + \frac{\sqrt{\varepsilon_r}}{c_0} \cdot D \cdot \left(1 - \frac{1}{2^{n/2}} \right) + 0.7 \cdot \left(\frac{L_{eff} / W}{\mu \cdot C_{ox} \cdot (V_{DD} - V_T)} \right) \cdot C_L \quad (6)$$

Equation (6) contains all device, interconnect and system parameters described previously. Assuming that these parameters have small variations compared to their nominal values, the clock skew, T_{CSK}, can be evaluated by:

$$T_{CSK}(x) = \Delta T_{Delay} \approx \left| \frac{\partial T_{Delay}}{\partial x} \right| \Delta x \quad (7)$$

where T_{Delay} is the complete delay function of (6), and x is any variation of clock skew components such as ΔV_T, Δt_{ox}, ΔL_{eff} , ΔH_{int}, ΔT_{ILD}, ΔV_{DD}, $\Delta \mathcal{T}$ and ΔC_L. Table 1 shows the closed form equations for each individual clock skew component by using (7):

Physical parameter and derivation used	**Clock skew component**
Threshold voltage fluctuation	$T_{CSK}(V_T) = 0.7 \cdot R_0 \cdot C_L \cdot \left(\frac{V_T}{V_{DD} - V_T} \right) \cdot \frac{\Delta V_T}{V_T}$
Gate oxide thickness tolerance	$T_{CSK}(t_{ox}) = 0.7 \cdot R_0 \cdot C_L \cdot \frac{\Delta t_{ox}}{t_{ox}}$
Transistor channel length tolerance	$T_{CSK}(L_{eff}) = 0.7 \cdot R_0 \cdot C_L \cdot \frac{\Delta L_{eff}}{L_{eff}}$
Wire thickness variation	$T_{CSK}(t_{\text{int}}) = 0.4 \cdot (r_{\text{int}} \cdot c_{\text{int}}) \cdot D^2 \cdot \left(1 - \frac{1}{2^{n/2}} \right)^2 \cdot \frac{\Delta t_{\text{int}}}{t_{\text{int}}}$
ILD thickness variation	$T_{CSK}(T_{ILD}) = 0.4 \cdot (r_{\text{int}} \cdot c_{\text{int}}) \cdot D^2 \cdot \left(1 - \frac{1}{2^{n/2}} \right)^2 \cdot \frac{\Delta T_{ILD}}{T_{ILD}}$

IR drop	$T_{CSK}(V_{DD}) = 0.7 \cdot R_0 \cdot C_L \cdot \left(\frac{V_{DD}}{V_{DD} - V_T} \right) \cdot \frac{\Delta V_{DD}}{V_{DD}}$
Non uniform register distribution	$T_{CSK}(C_L) = 0.7 \cdot R_0 \cdot C_L \cdot \frac{\Delta C_L}{C_L}$
Temperature gradient	$T_{CSK}(\mathrm{T}) = 0.7 \cdot R_0 \cdot C_L \cdot \left(\frac{E_g / q + V_T}{V_{DD} - V_T} \right) \cdot \frac{\Delta \mathrm{T}}{\mathrm{T}}$
Internal Clock skew	$T_{CSK}(sub) = 0.4 \cdot (r_{\mathrm{int}-sub} c_{\mathrm{int}-sub}) \cdot \frac{D^2}{2^n} + \frac{\sqrt{\varepsilon_r}}{c_0} \cdot \frac{D}{2^{n/2}}$

Table 1: Clock skew components.

3.3. Clock Skew for Temperature Variation

The clock skew due to temperature gradient on a chip, in general, is more complex since there are three main parameters that vary with temperature: resistivity of interconnect $\rho(\mathcal{T})$, threshold voltage $V_T(\mathcal{T})$, and mobility $\mu(\mathcal{T})$. Assuming that the variation of threshold voltage is greater than that of mobility and resistivity of lines, then the clock skew due to temperature difference is given by:

$$T_{CSK}(\mathrm{T}) = \Delta T_{Delay}(\mathrm{T}) = \left| \frac{\partial T_{Delay}}{\partial V_T} \cdot \frac{\partial V_T}{\partial \mathrm{T}} \right| \cdot \Delta \mathrm{T} \qquad (8)$$

where $\Delta\mathcal{T}$ is temperature difference of two points in the chip. The first expression, $\partial \mathrm{T}_{\mathrm{Delay}} / \partial \mathrm{V}_{\mathrm{T}}$, is computed from (6) as:

$$\frac{\partial T_{Delay}}{\partial V_T} = 0.7 \cdot R_0 \cdot C_L \cdot \frac{1}{V_{DD} - V_T} \qquad (9)$$

Also the second expression, $\partial\, \mathrm{V_T} / \partial\mathcal{T}$, is:

$$\frac{\partial V_T}{\partial \mathrm{T}} = \frac{1}{\mathrm{T}} \cdot \left(2 - \frac{Q_B}{2 C_{ox} \varphi_f} \right) \cdot \left(\varphi_f + \frac{E_g}{2q} \right) \qquad (10)$$

where Q_B is the depletion-region charge, C_{ox} is the gate oxide capacitance, φ_f is the Fermi level potential, E_g is the energy gap of Si, and q is the charge of electron. In order to simplify (10), the threshold voltage can be written as $V_T \approx \varphi_{Si} - Q_B/C_{ox}$. Moreover assuming that the substrate doping concentration is relatively high, then the surface potential of MOSFET transistor is given by $\varphi_{Si} = \varphi_f \approx 2(E_g/2)$. Therefore the first order approximation of (9) is given by:

$$\frac{\partial V_T}{\partial \mathrm{T}} = \frac{E_g / q + V_T}{\mathrm{T}} \qquad (11)$$

where $Eg/q = 1.12\ V$ is the energy gap of Si in volts, and $\mathcal{T}$ is the temperature in degrees Kelvin. Equation (8) along with the results of (9) and (11) give:

$$T_{Delay}(\mathrm{T}) = 0.7 \cdot R_{tr} \cdot C_L \cdot \left(\frac{E_g / q + V_T}{V_{DD} - V_T} \right) \cdot \frac{\Delta \mathrm{T}}{\mathrm{T}} \qquad (12)$$

4. Example

In order to illustrate the new clock skew model, a design example for 0.18 μm technology has been studied using the design parameters illustrated in Table 2. The H-tree clock distribution is assumed to be routed by the 4th metal level shielded with the 3rd and 5th metal levels as shown in Figure 4. This wiring structure ensures minimal inductive effect.

Parameters		Values
Process parameters	L_{eff}	0.18 μm
	V_T	0.32 V
	V_{DD}	1.8 V
	$r_{int}c_{int}$	115 ps/cm^2
Design parameters	R_0	12.0 Ω
	C_L	6.25 pF
	D	2.0 cm
	n	4

Table 2: Process and design parameters.

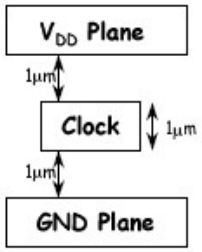

Figure 4: Clock distribution wiring structure.

Equation (1) is, therefore, valid for interconnect propagation delay with $r_{int}c_{int}=115$ *ps/cm*^{2}, assuming that Cu/SiO_2 materials are used in wiring network. In this example, the total number of clocked registers is set to *20,000*. With an input register capacitance of *5fF*, the total capacitance of each sub-block (2^4 = 16 in total), on average, is computed as $C_L=6.25$ *pF*. A driver with $R_0=12.0\Omega$ output resistance is selected to ensure fast rise time for the total sub-block loading capacitance.

Parameters	% of variation	Clock Skew Component [ps]
V_T	5 %	0.55
t_{ox}	1.2 %	0.64
L_{eff}	5 %	2.65
H_{int}	3 %	3.12
T_{ILD}	3 %	3.12
V_{DD}	10 %	6.40
C_L	20 %	10.60
τ	8 %	4.08
Internal Clock Skew [ps]		61.7
Total Clock Skew [ps]		92.9

Table 3: Clock Skew Components.

Using the expressions of Table I, the complete set of clock skew components are evaluated as shown in Table 3. The second column contains percentage of variations, which are technology and design dependent. It means the following: if % of variation of V_T is *5%*, then ΔV_T is $0.05V_T$.

The tolerances of interconnect and ILD thickness is roughly *3%* for a well-controlled CMP process. Statistical modelling extracted from the measurement data of

0.18μm technology shows that the fluctuations of threshold voltage, MOS effective channel length, and gate oxide thickness are about *5%*, *5%* and *1.2%* respectively. The tolerance of supply voltage is usually limited to *10%* of total supply voltage. We are going to suppose a *30ºC* temperature gradient over the entire chip (as the thermal image of the Alpha microprocessor), which gives a temperature variation of about *8%*. The variations of loading capacitance of clock drivers highly depend on the uniformity of architecture. A specific investigation of a microprocessor design shows that about *20%* variation exists on the sub-block loading capacitance.

The clock skew components are evaluated based on the amount of variations in the third column of Table 3. Figure 5 illustrates a graphical view of clock skew components. A skew of *10.6ps* is created just by clock driver load mismatch. Also, IR drop, temperature gradient, interconnect and ILD thickness variations, and MOS channel length tolerance create *6.4ps*, *4.08ps*, *3.12ps* and *2.65ps* respectively.

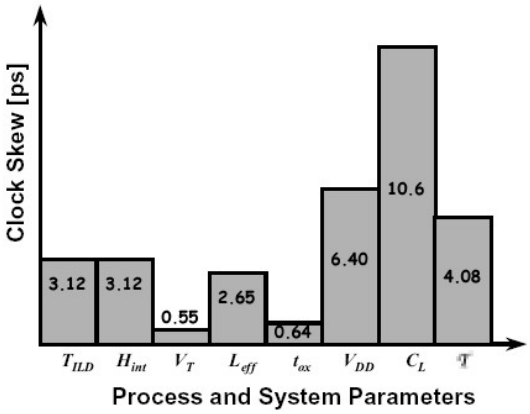

Figure 5: Clock skew components.

5. Conclusions

- The model provides a statistical expression for the clock skew in a H-Tree clock network, where there aren't intermediate buffers at the split points.

- Its easily applicable to any clock network, modifying the expressions obtained.

- It's a statistical model because the total clock skew is obtained in function of both, process and design parameters, % of variation.

- To ignore the inductance effects between wires, system design have to include a good return path wiring surrounding the clock-wiring network.

- To apply the model it's necessary to know:

  - Process parameters:
    - Interconnection parameters: $r_{int}c_{int}$
    - Threshold voltage of inverters: V_T
    - Power supply voltage: V_{DD}
    - Transistors energy gap: E_g

  - Design parameters:
    - Output resistance of driving buffers: R_0
    - Die size: D
    - H-tree levels: n
    - Capacitive load of sub-blocks: C_L
    - Relative dielectric constant of ILD material (only to calculate sub-block delay): ε_r

  - Parameter variations (in %):
    - Threshold voltage: σ_{VT}
    - Gate oxide thickness: σ_{tox}
    - Transistor effective channel length: σ_{Leff}
    - Wire thickness: σ_{tint}
    - ILD thickness: σ_{TILD}
    - Power supply voltage: σ_{VDD}
    - Load capacitance: σ_{CL}
    - Temperature: $\sigma_{\mathcal{T}}$

- The model equations can be easily modified to be more similar to other models, where the Sakurai's expressions are used with 90% of time delay. It only supposes to change the coefficients of $T_{H\text{-}tree}$, T_{Driver} and $T_{Sub\text{-}Blk}$.

  - $T_{H-tree}: \quad 0.4 \rightarrow 1.02 \Rightarrow T_{H-tree} = 1.02 \cdot (r_{\text{int}} c_{\text{int}}) \cdot l^2 + \frac{\sqrt{\varepsilon_r}}{c_0} \cdot l$
  - $T_{driver}: \quad 0.7 \rightarrow 2.3 \Rightarrow T_{driver} = 2.3 \cdot R_0 \cdot C_L$
  - $T_{Sub-Blk}: \quad 0.4 \rightarrow 1.02 \Rightarrow T_{Sub-Blk} = 1.02 \cdot (r_{\text{int}-sub} c_{\text{int}-sub}) \cdot d^2 + \frac{\sqrt{\varepsilon_r}}{c_0} \cdot l$