Abstract—Most embedded systems of today are built using the SOC technology for large scale production. As technology advances delay of transistors and local interconnects scales down and in this way the local clock frequency is projected to increase rapidly. Scaling the transistors makes local interconnects shorter and results in smaller delay. However, since chip area is projected to increase, length of global interconnects increases and therefore their delay. For instance, the key solutions for reducing delay of global interconnects can be thought of as increasing the wire cross-sectional area and inserting repeaters while retaining a good wiring density. Wiring density on the other hand introduces signal integrity issues. Also, the communication between the SOC and memories outside the chip happens through off-chip interconnects, so an optimal partition for on and off-chip interconnects is required. These different tradeoffs necessary for SOC interconnect modeling are discussed in detail in this report.

I. Introduction

The wires linking transistors together are called interconnect and play a major role in the performance of modern systems. Fig. 1 shows a pair of adjacent wires. These wires have a width $w$, length $l$, thickness $t$, and spacing of $s$, from their neighbors and have a dielectric height $h$ between them and the conductive layer below. Earlier, low density SoC chips had relatively slower transistors and wider and thicker wires which had low resistance. Under those circumstances, wires could be treated as ideal equipotential nodes with lumped capacitance. But today, transistor switching frequencies are much higher and the wires have become narrower, driving up their resistance such that wire RC delay exceeds gate delay. Moreover, the wires are packed very closely together and thus a large fraction of their capacitance is to their neighbors, resulting in crosstalk. Also, on-chip inductance is now becoming a factor for systems with fast edge rates and closely packed buses. Considering all of these factors, circuit design is now as much about engineering the wires to extract optimal performance and reap the benefits of reducing feature size. This report is organized by introducing the interconnect modeling blocks in section II, then the issues with interconnect are discussed in section III followed by the methods of improving delay in section IV and concluded with the current and future trends and innovations in sections V, VI.

II. Building Blocks of Interconnect Modeling

Resistance

Resistance of a uniform slab of conducting material can be expressed as

$$ R = \frac{\rho}{t} \cdot \frac{l}{w} \quad (1) $$

where $\rho$ is the resistivity of the conductor and is a fixed property of the metal being used for the wire.
Capacitance
An isolated wire over the substrate (silicon layer over which transistors are built and is typically grounded) can be modeled as a conductor over a ground plane. The wire capacitance has two major components: the parallel plate capacitance of the bottom of the wire to the ground and the fringing capacitance arising from the fringing fields along the edge of a conductor with finite thickness. In the addition, a wire adjacent to a second wire on the same layer can exhibit capacitance to that neighbor. The classic parallel plate capacitance formula is

\[ C = \frac{\varepsilon_{\text{ox}}}{h} \cdot w \cdot l \]  

(2)

The fringing capacitance is more complicated to compute and requires numerical field solver for exact results. We see that capacitance interactions between layers is quite complex in today’s CMOS processes used to build SoC. A conservative upper bound on capacitance can be obtained assuming that the layers above and below the conductor of interest are solid ground planes. Similarly, a lower bound can be obtained assuming there are no other conductors in the system except the substrate. The upper bound can be used for propagation delay and power estimation while the lower bound can be used for contamination delay (minimum delay) calculations before laying out an SoC. A cross-section of the model used for capacitance upper bound calculation is shown in Fig. 2. The total capacitance of the conductor of interest is the sum of its capacitance to the layer above, the layer below and the two adjacent conductors. If the layers above and below are not switching, they can be modeled as ground planes and this component of capacitance is called \( C_{\text{gnd}} \). Theoretically, wires will have some capacitance to further neighbors, but in practice capacitance is infinitesimal and can be ignored because most electric fields terminate on the nearest conductors.

\[ C_{\text{gnd}} = C_{\text{bot}} + C_{\text{top}} \]

\[ C_{\text{total}} = C_{\text{gnd}} + 2 \cdot C_{\text{adj}} \]

(3)

Delay
Interconnect increases delay for two reasons. First, the wire capacitance adds loading to each gate. Second, long wires have significant resistance that contributes distributed RC delay or flight time. The distributed resistance and capacitance of a wire can be approximated with a number of lumped elements. Of the many approaches the \( \Pi \) – model is the most popular which gives an accuracy of up to 3%, the model is shown in Fig. 3,

If the resistance of the input buffer \( I \) is \( R_{\text{in}} \) and the output load offered by the output buffer \( O \) on the right hand side is \( C_{\text{out}} \), then Elmore Delay models says that delay of the circuit with interconnects is give by,
\[ \tau = \frac{C}{2} (R_{\text{in}}) + \left( \frac{C}{2} + C_{\text{out}} \right) (R_{\text{i}} + R) \quad \ldots \ldots \ (4) \]

This gives us an approximate hand calculation to estimate the delay of a circuit given the lumped capacitance and resistance values.

**Inductance**

Global signal and clock wires are routed with large widths and thicknesses at the top levels of the metal to minimize delays. This decreases the resistance of the wires, making their inductive impedance comparable to the resistive part. \( Z = R + j\omega L \), when \( \omega L \) is comparable to \( R \), inductive effects must be considered. As the clock frequency increases and the rise times decrease, electrical signals comprise more and more high-frequency components, making the inductance effects more significant. Example of Inductance Effects (a) Over/under-shoot edges, (b) \( L \frac{d}{dt} \) voltage drop, (c) Long range crosstalk, and (d) \( f \)-dependent \( R \).

**III. Issues due to increased Package Density and Feature Size Reduction**

**Signal Integrity and Crosstalk**

The most important Signal integrity problems are: 1) crosstalk, 2) Overshoot (signal rising momentarily above the power supply voltage or below ground), 3) reflection (echoing back a portion of a signal when it reaches the end of interconnect), 4) signal skew (the difference in arrival time of one source signal to different receivers). The occurrence of crosstalk faults is attributed to the fact that a wire not only serves as a conductor of electrons but also includes parasitic resistor (at low frequencies), capacitor (at mid-range frequencies), inductor (at high frequencies) and antenna (at very high frequencies). From Fig.2 we may see that wires A, B and C affect each other through coupling capacitance which is also called the Miller capacitance. If both a wire and its neighbor are switching, the direction of the switching affects the amount of charge that must be delivered and the delay of the switching. The charge delivered to the coupling capacitor is, \( Q = C_{\text{adj}} \Delta V \), where \( \Delta V \) is the change in voltage between A and B. \( \Delta V \) might change depending on whether the switching in the same direction or different or idle. Aggressor and victim models of crosstalk determination have been developed to deal with the computation of crosstalk. In essence it is important to remember that the delay is improved/worsened by addition of crosstalk from neighboring wires.

There are many possible design and fabrication solutions to margin and minimize crosstalk problems between interconnects. These solutions include: 3-D layout modeling and parasitic extraction, accurate RLC simulation of on-chip power grid, using decoupling capacitors to limit the maximum \( \frac{dV}{dt} \), inserting repeaters/buffers on interconnects and shielding wires.

**IV. Improving Interconnect Delay**

**Repeaters**

We see that due to the increase in the length of the delay lines the signal of interest get attenuated and might result in an undesired transition/result. To cope with this problem an optimal number of repeaters have to be inserted along the length of the wire to maintain signal integrity. Repeaters if made of good buffers also help in elimination of crosstalk effects by driving a in a direction that is functionally desired.

\[ T_{pd} = \frac{C_{\text{i}} l}{2N} (R_{\text{in}}) + \left( \frac{C_{\text{i}} l}{2N} + C_{\text{out}} \right) (R_{\text{i}} + R_{l}/N) \quad \ldots \ldots \ (5) \]
The idea is to minimize delay by inserting the optimal number of repeaters, so by partial differentiation of the above result with respect to $N$, i.e., $\delta T_{pd}/\delta N = 0$, we get,

$$l/N = \left[\frac{2R_{in}C_{out}}{RC}\right]^{1/2} \quad \text{(6)}$$

The above result is very important for quick hand computation of diving a given length of a wire segment to get the minimal delay.

**On and off-chip Interconnects**

Scaling the transistors increases overall chip area and the length of global interconnects increases and therefore their delay. The key solutions for reducing delay of global interconnects are increasing the wire cross-sectional area and inserting repeaters. However, using fat wires decreases the wiring density. Off-chip interconnects usually have very large cross-sectional area and therefore, they have negligible loss. In this way some of the long global interconnects may be routed through the printed wiring board (PWB). However, since PWB wiring density is small, very few interconnects may be routed through the board. An optimal partition between on-chip and off-chip interconnects is required which results in highest performance. It has been heuristically determined that interconnects for of length $l$ such that, $l_{max} < l < 2D_{chip}$, (where $l_{max} = 0.75.2D_{chip}$, $D_{chip}$ being the edge size of the chip) should be on-PWB to have the highest global clock frequency.

**V. Current Tools Trends in Interconnect Modeling**

- Place and Route (P&R) tools like Magma’s Mantle and Synopsys’s Astro provide data to the timing analyzer by back-annotating the laid out SoC’s capacitance and resistance in lumped (delay modeling) or coupled (crosstalk modeling) format through IEEE Standard Parasitic Extraction File (SPEF), Standard Delay Format (SDF) or the binary SPEF.
- Cadence’s Allegro is one of the advanced tools that is used to model chip-to-board interconnects. Allegro allows platform users to trace a signal as it moves from an IC's I/O buffers and through redistribution layers, traversing die bump pads, package substrates, connectors and pc boards.

**VI. Future Trends and Innovations**

- Parallel Repeater-Insertion Methodologies for SoC Interconnects – Work till now has been restricted to serial repeaters operating when interconnect inductive effects are not significant. It is proven in that parallel repeaters outperform serial repeaters in terms of delay, power and silicon area when regenerating signals in SoC interconnects.
- Interconnect Delay Aware RTL Bus Architecture – A quorum of researches in Georgia Institute of technology demonstrated a methodology to generate a custom bus architecture using accurate estimations of interconnect delay. To improve bus delay accuracy, they altered the bus Verilog register transfer level (RTL) specification based on interconnect delay estimations and used early in the design phase.

**VII. Conclusion**

Due to the limitations posed by interconnects; designers today are not able to efficiently reap the benefits of reducing feature sizes and frequency scaling. Towards this end many advanced models are being researched and deployed concurrently to model the higher order effects in interconnects that arise in lower device geometries. Interconnect delay effects are being incorporated and taken into account in the early design and architecture phase by introducing delay headroom. Future of interconnect modeling calls for design of noise-canceling interconnects which is an active field of research wherein the wires are design to reject (to the least minimize) effect of neighboring high frequency nets. Mixed signal SoC call for more stringent design constraints on the way digital interconnects are laid over analog nets. A possible research area to increase wiring density would be to make high frequency resistant analog nets (with inbuilt low pass filters) to reject digital noise of adjacent nets. Many such advanced design techniques can be researched, validated, designed and deployed to be able to maximize the benefit and leverage on technology progresses.
VIII. References