Lecture 6-A

Interconnect/Wires

C.K. Ken Yang
University of California at Los Angeles
yang@ee.ucla.edu

Overview

Reading
- W&H: 4.5
- Rabaey 8.1-8.4 – reference skimming

Introduction

Wires are very simple "devices". However, as we put millions on a chip, the entire network is very complex and poses many problems. What is the appropriate representation of a wire? The most simple (and common) is as a capacitor. But as resolution scales, it appears more like an RC. As speed scales, it appears more like an RLC. What makes matters even more complex is that C (and L) are very sensitive to geometry. The organization and density of wires pose a very challenging modeling and simulation problem for designers. Our goal is to understand some of the bounds of these models.
What Gates Drive

- Output of Gates drive not only Inputs of other Gates but also the Interconnect.
- The waveform at the input of a gate depends on the Gate driving the signal and the “load” that it drives.
- For instance, an inverter
  - $R_{DRV}$ and $C_{self} + C_{gate}$.
  - Delay of a gate depends on wire characteristics
    - Primarily consider C.
    - For longer wires, additionally consider R
    - For well-shielded low resistance wires, add L.

What Is Interconnect?

- Any layer used to connect between transistors can be considered as interconnect.
  - Early on, with few layers of processing: diffusion and poly.
  - Now, diffusion is only used between adjacent transistors, and poly used only within a Gate ($100\lambda \times 100\lambda$)
  - Primarily Aluminum and Copper.
    - Layers are connected by W

3 primary interconnect characteristics:
1. Capacitance
2. Resistance
3. Inductance
Interconnect (Wire) Capacitance

Basic Wire Capacitance

- Diffusion (junction capacitance – discussed earlier)
  - Nonlinear, depends on doping concentration.
- Metal and poly – mostly linear capacitor.
  - Terminals separated by oxide.
  - Depends on Wire Geometry
  - Width, Length, Height, $t_{ox}$
  - Not really parallel plate capacitance.
- Some earlier methods
  - Yuan & Trick split cap into Plate and Fringe
    - $C_{plate} = \varepsilon_{ox}L(W-0.5H/t_{ox})$
    - $C_{fringe} = C_{cylinder}$
  - Empirically fitted, i.e.
    - $C = \varepsilon_{ox}L(W/t_{ox}) + 1.06(W/t_{ox})^{0.25} + 1.06(H/t_{ox})^{0.5}$
    - Computational efficiency is important.

\[ C_{cylinder} = \frac{2\pi dL}{\ln(1 + \frac{2H}{T}(1 + \sqrt{1 + T/H}))} \]  

[Yuan, Trick]
Scaling Problem

- Scaling worsens the capacitance modeling.
  - Geometry of wire is taller than wide. (W/H is <1)
  - Spacing between wires (S/t_{ox} < 1)
- With cap mostly fringe, empirical equations or physically-based equations don’t work well.

Complexity of Modern Interconnect

- Wires have neighbors of arbitrary arrangement
  - Spacing varies.
  - Thickness is relatively fixed at 1µm between layers.
  - Contacts make matters even worse.
  - Capacitance to V\(_{\text{GND}}\)/V\(_{\text{DD}}\) AND to Signals.
- Solution involves solving 2D or 3D field solvers.
  - For various configurations and geometries.
Modeling Interconnect Capacitance

- Simple back of the envelope calculations – Zeroth order
  - 0.3fF/µm – for minimum width wires
- More accurate estimation uses capacitance tables
  - Determine a set of common geometries.
  - Width of wire(s), spacing, upper/lower neighbors
  - Run 2,3-D field solver for the geometries
  - Use simple equations to extrapolate in between configurations
- Capacitance "extraction" (2,2.5D Extractors)
  - Determine the capacitance of each net
  - Identify the geometry for each segment of wire and determine the capacitance for the entire net.
  - Computational cost is the criteria (LARGE # of nets)
    - Pattern matching (both node names AND geometry)
    - Table lookup and Extrapolation

Impact of Interconnect Capacitance

- Delay calculation is easy.
- Directly add interconnect capacitance to the node.
  - \( C_{total-load} = C_{self-loading} + C_{wire} + C_{gate-loading} \)
- Delay = RC_{total-load}.
- The difficulty is that \( C_{wire} \) is a sum of many capacitors.
  - Each may not be to ground.
  - Aside: similar approximation was made for gate/self loading.
  - Capacitance to other signals is a large fraction of total C.
    - Especially with scaling.
- So even when we have the capacitances, it is DIFFICULT to simulate accurately.
Signal Coupling

- The actual capacitance is sensitive to transition of coupled signal.
  - Same direction: effective $C_{\text{couple}} = 0$
  - Opposite directions: $C_{\text{couple}} = 2x$

- Noise injection
  - Capacitive coupling by an aggressor $\Delta V_1 = V_{DD}$ onto a victim
  - If capacitances are driven by resistances, then the R’s impact the amount of noise.
    - $R_{\text{aggressor}} < R_{\text{victim}}$
    - $R_{\text{aggressor}} > R_{\text{victim}}$

- Tools can help a lot but still too much to do so for every net.
  - Best if designers identify sensitive nets and use tools to treat them differently.

\[ \Delta V_2 = \frac{\Delta V_1 C_1}{C_1 + C_2} \]

[Charged conservation before and after the switching event.]

Interconnect Resistance
**Wire Resistance**

- Just like transistors, we treat resistances in terms of squares
  \[ R = \frac{\rho L}{A} = \frac{\rho L}{W \times H} \]
  Since height is constant, \( R_{\text{squ}} = \frac{\rho}{H} \).
  Unlike a transistor (where \( L \) is usually fixed, \( L_{\text{min}} \)), \( W \) and \( L \) changes, more often \( L \). So, \( R/\mu \text{m} \) is also commonly used.
- Diffusion interconnect
  - Large \( R \) (100's \( \Omega \)/square) but with silicide 1-5\( \Omega \)/square.
  - For adjacent gate interconnection \( R \) is not an issue
- Metal and poly interconnect
  - Poly has 100\( \Omega \)/square but also silicided
  - Metal has <0.1\( \Omega \)/square!
  - Good number to use for back of the envelop calculation!

---

**Distributing the Capacitance**

- The resistance and capacitance of an interconnect is distributed.
  - The resistance distributes the capacitance into many nodes.
- Delay is different from a lumped model.
  - The result is roughly \( \frac{1}{2} \) the delay.
  - For a step input:
    - \( T_{\text{delay, lumped}} = \ln(2) RC \)
    - \( T_{\text{delay, distributed}} = 0.38 RC \sim \ln(2) RC/2 \)
Example of Wire Delay

- Extract a wire’s R and C parameters
  - Simple example: R = 0.1Ω/square, C = 0.3fF/µm (for min W wire)
- Extract length and tree based on driving gate position and load position.

- Needed for global wires, busses, and large standard cell blocks.
  - But Elmore delay (only dominant pole) is not accurate for RLC models.
    - At least 3 moments and 2 poles are needed for an accurate delay estimate.

RC Model

- A lumped L-Model for RC wire has a large error
  - Distributed model uses N segments.
    - More accurate but computationally expensive
    - Number of nodes blows up.
- R-C Π Model is gives a more reasonable result.
  - Split each segment’s C into 2 C/2.

Distributed using multiple lumps of Π model of a single wire
Wire Loading and RC Tree

- Lump model uses 1 segment of Π to represent wire
  - Sufficient for most nets (point to point).
  - No different from distributed RC using Elmore Delay.
  - Roughly accounts for C/2
  - \( R_{\text{INT}} \) is the actual resistance

- Single gate can drive many gates creating a RC tree network.

A Slight Inconsistency

- If we calculate the time constant for a gate driving a distributed wire, with a step input,
  - \( \tau_{\text{Elmore}} = R_{\text{drv}}(C_{\text{SD}} + C_{\text{W}}/2) + (R_{\text{drv}} + R_{\text{W}})(C_{\text{W}}/2 + C_{\text{G}}) \)
  - Actual delay ~ ln(2)\( \cdot \)t (slight error for the wire)
- But, we already accounted for the ln(2) in the Rdrv.
  - To be accurate, we can use \( R_W = 0.76 \times \) to 1.0x of actual resistance.
  - To make our life easier, since \( RW \) is usually not very large, we can just use \( R_W = R_{\text{actual}} \).
When Is Resistance Important?

- A question of how accurate do we want our simulations.
- If 10% error is permitted, then a $R_{\text{wire}} > 0.10R_{\text{drv}}$ is noticeable.
  - For discussion purposes, this is what we will use in this class.
    - Results in a fairly significant wire RC.
    - For example, wire width is usually $3\lambda$, and $R_{\text{DRV}}$ of a 0.25-µm device is $3k\Omega$-µm.
      - Assuming 2-µm devices, 1.5kΩ drive resistance.
      - Wire for 10% (150Ω) is approximately 5000λ (600µm).
    - Absolute distance shrinks with technology scaling

Skin Effect

- Switching frequency is getting higher.
  - Signal contains higher harmonics
- An interesting effect occurs at high frequencies
  - Current travels through the path of least impedance. At higher frequencies, reactive impedance becomes more significant and current crowds to the edges of the conductor.
- The thickness of conduction is known as skin depth, $\delta$, and depends on frequency.
  \[
  \delta = \sqrt{\frac{1}{\pi \mu \sigma f}}
  \]
- Area is effectively less, so $R$ increases.
Interconnect Inductance

Inductance

• When signal is coupled to a ground plane, the current loop has an inductance.
  – Approximate equation of trace on plane can be estimated (below).
  – The inductance is more apparent for upper layer metals (lower resistance) and longer lines.
  – Simple model (notice more nodes)

\[
L = \text{len} \frac{\mu_0}{2\pi} \ln\left(\frac{8h}{w + \frac{w}{4h}}\right)
\]
Impact of Inductance

- Inductance, makes the interconnect like a transmission line.
  - Propagates signal energy, with delay.
  - Results in sharper rise times and delayed signal

Additional Problem: Mutual Inductance

- Magnetic flux couples to many signals
  - Not just to immediate adjacent signals (unlike capacitors.)
  - Coupling over a larger range.
  - Inductive noise (v=Ldi/dt)
    - Large di/dt due to fast signal transitions.
  - Sensitive to geometry and configuration.
  - Much bigger lump model
    - Matrix not sparse (coupling coefficient)
- Very big problem if we have to do this with every single signal.
  - Computationally, we know how, but too intensive.
Inductor Modeling

- The error is not very large.
  - Ignored for most local nets.
- Care with “special” nets
  - Nets with precise delay needs
    - Especially over long distances
  - Nets with rise time requirements
    - Noise coupling is more severe with sharper edges.
- Common nets
  - Clock networks
    - Many latches work better with sharp clock edge ($t_{\text{setup}}$, data retention)
  - Wide and long data busses.
    - Noise coupling between busses.

Delay of RLC

- With a reasonably good transmission line
  - $R \ll$ than the line impedance.
    - Damping factor $\zeta \ll 1$
      $$\zeta = \frac{R}{2\sqrt{L/C}}$$
    - velocity $= \omega_0 = 1/\sqrt{LC}$
    - $t_{pd} = t_{\text{flight}} = \text{length}\times\sqrt{LC}$.
  - With $\zeta$ near 1, the estimate can have large error
    - An empirical equation [Ismail & Friedman]
      $$t_{pd} = \frac{\exp(-2.9\zeta^{1.35}) + 1.48\zeta}{\omega_n}$$
When Is Inductance Modeled

- Two simple rule of thumbs are:
  \[ 2 \pi f_{\text{transition}} L' > R' \]
  \[ t_{\text{rise(fall)}} < 4 \frac{L'}{R'} \]
  \[ t_{\text{rise(fall)}} < 2.5 t_{\text{FLIGHT}} = 2.5 \text{Length} \sqrt{L'C'} \]
  - When the propagated energy is greater than the dissipated, the effect becomes significant. \((L', C')\) are \(L, C\) per meter
  - \(f_{\text{transition}} \sim 1.6/t_{\text{rise}}\)
  - When time of flight is significant compared to rise/fall times.
  - Estimate max \(t_{\text{rise}}\) by the gate output without wire loading.

- Rewrite of the rules (from W&H)

\[ \frac{t_{\text{rise(fall)}}}{2 \sqrt{L'C'}} < \text{length} < \frac{2L}{R \sqrt{C}} \]

Approximate Range Where Inductance Matters

- Challenging problem to include into the analysis ALL nets.
  - Ignore inductance for most of this class (only when we talk about clock distribution)
- Solution – constrain the problem
  - Build fixed wire configurations for "special" nets (so it is easy to analyze).
  - Reduce coupling
    - Shielding with ground-plane below or non-signal adjacent signals.
    - Spacing out the wires.
Summary

- Interconnects are an increasingly important part of VLSI design.
  - Accounts for >25% of the critical path delay.
  - Accounts for >30% of power dissipation.
  - Most of the area is now wires.
- Including wires in the signal modeling is critical.
  - Analyze more than to delay/power. Signal integrity due to signal coupling noise (slew rate of signal)
- Must include capacitances.
- Include resistance for longer wires.
- Include inductance for low resistive wires and sharp signal transitions.
  - Transitions are getting sharper with scaling while distances are not shrinking!