Interconnect Aware VLSI Design

Vani Prasad

vani@ee.iitb.ac.in

Dept. of Electrical Engineering,
I.I.T. Bombay
12th Sept, 2K3
Outline

- Introduction
- Interconnect Aware VLSI Design
- Wire Architectures -- building blocks
- Summary
Introduction

- Technology scaling
- The interconnect problem in the SoC era
- Physical issues
  - Delay
  - Signal integrity
  - Area
  - Power distribution
  - Reliability issues

SoC - example
Interconnect Aware VLSI Design

GATE-CENTRIC

INTERCONNECT - CENTRIC

PERFORMANCE DRIVEN

PREDICTABLE

NON-ITERATIVE

INTERCONNECT AWARE DESIGN
How ???

(COMMUNICATION–BASED LAYERED APPROACH)

VLSI INTERCONNECTS ➔ HIERARCHY OF PHYSICAL BLOCKS

BASIC BUILDING BLOCKS:

1) P–2–P SINGLE LAYER CONNECTION
2) P–2–P MULTI–LAYER CONNECTION
3) P–2–P BUS (SINGLE LAYER AND MULTI–LAYER)
4) P–2–P MULTIPLEXED CONNECTION
5) BROADCAST CONNECTION
The Basic Building Blocks

(a) A Single Wire

(b) A Single Wire With External Noise

(c) A Set of $k$-wires Communication L-to-R

(d) $k$-wires from L-to-R and $m$-wires from R-to-L

(e) Resource Sharing Using Multiplexing
The Layered Approach

Upper layers

Communication based interconnect architecture (Network on Chip)

Physical layer

- Delay: Buffer insertion, wire sizing
- Crosstalk: Spacing, shielding, diff signaling
- Energy: Low swing signaling
- Area
(PHYSICAL LAYER)

MODULE PLANNING ← WIRE PLANNING

DELAY THROUGHPUT AREA CROSSTALK

INTERCONNECT NETWORK BACKPLANE

LOGIC MODULE

JUST PLUG IN THE DEVICES !!
The Network on Chip (NoC)

- Get the best interconnection architecture

- Architecture = topology + scheduling + switching + routing

- Analyze performance qualitatively and quantitatively

- Network models needed for such an analysis

[ Lalit Keshre – 5th Yr DD student ]
Where do we start??

- Single wire (single layer) characterization
- Single wire (multi-layer) characterization

Wires characterized for minimum delay, maximum throughput, minimum area, min crosstalk, and given latency

Interconnect Library

what more ???
Characterizing a P-2-P Connection

Express Delay as a function of length and sink capacitance
Delay Minimization Techniques

- **Buffer insertion**
  -- segment the wire to make the delay linear

- **Wire sizing**
  -- width of the wire is varied to optimize delay

- **Simultaneous buffer insertion and wire sizing strategies**
  -- adopt uniform wire sizing
The Interconnect Problem

Given

input capacitance, a sink load and source-sink distance of 'L'

Find

the optimal number of buffers, their sizes, placement and the wire width

Such that

the total delay from source to sink is minimum

Get simple models for the wire !!!
Optimization strategy

- Simultaneous buffer insertion and uniform wire sizing
- Delay equation formulated using Elmore delay model
- Delay is a posynomial in buffer and wire width
- Use convex optimization to arrive at the global minima
- Trends in delay and buffer widths for different lengths observed
Our Wire Architecture

The Pre-Mid-Post (p-m-p) Buffer Strategy

Variables to be calculated:
- # Buffers?
- Size of buffers?
- Wire size?

Variables:
- $R_b, C_b$
- $C_{sink}$
- $W, W_{int}$
- $L, L/n$
- $n_{pre}, n_{post}$
Our Contributions

- Same buffer and wire width across different lengths for a given metal layer
- The energy and delay models for a wire from the simplified pmp architecture:
  
  \[
  \text{Delay} = x_0(L-2000) + x_1 \log C_x \\
  \text{Energy} = y_0(L-2000) + y_1 C_x
  \]
  
  \[
  C_x = C_{si}/C_b ; x_0 = 0.13e-3; x_1 = 0.28; y_0 = 1.16e-3; y_1 = 6e-3
  \]
- Delay expressed as a function of wire-length and sink capacitance
- Unconstrained optimization performed to arrive at the optimal solution
- Simple and “near-optimal” pmp buffering strategy for wire delay minimization of a 2-pin net
On Chip Resource Sharing Scheme

Time division multiplexing (test chip in 0.25µ IHP tech)

-- for effective bandwidth utilization
-- reducing number of long buffered wires on chip

[ Vineet Kothari – DD student, Class of 2k3 ]
Multi-layer nets

Two layer nets

![Diagram of two layer nets with VIA, Layer 1 and Layer 2 connections, and discontinuity.]
Optimization technique

- Find the best solution by solving an unconstrained optimization

Unconstrained solution

- Approximate this solution with regular structures
Buffering Strategies

2-pmp solution

Spliced-solution -- 1
Buffering Strategies contd..

Spliced-solution -- 2

Spliced-solution -- 3
Results

- Delay model :-

\[
D_{m1m2} = 1.5 \times 10^{-4}L_1 + 10^{-4}L_2 + .081 \log(C_x)
\]

\[
D_{m4m1} = 9.5 \times 10^{-5}L_1 + 1.3 \times 10^{-4}L_2 + .081 \log(C_x)
\]

- Extension to multi-layer interconnections :-
  - Use inductive approximation
  - How does the sub-optimality change with increasing layers?
Summary

- Interconnects are critical for system performance

- Move from the traditional design to an interconnect aware design flow...

- A communication-based layered approach to be adopted

- Interconnect planning → Interconnect architecture

- Predictable and simple models required at the physical layer
The IAVD Team

Prof. Madhav P. Desai

N. C. Vasudeva

Vineet Kothari

Lalit Keshre