inst.eecs.berkeley.edu/~eecs151

EECS151 : Introduction to Digital Design and ICs

Lecture 16 – Wires, Energy

Bora Nikolić

Apple Announces M1 Pro & M1 Max: Giant New SoCs with All-Out Performance

October 18, 2021, AnandTech - The M1 Pro and Max both follow-up on last year's M1, Apple's first generation Mac silicon that ushered in the beginning of Apple's journey to replace x86 based chips with their own in-house designs. The M1 had been widely successful for Apple, showcasing fantastic performance at never-before-seen power efficiency in the laptop market. Although the M1 was fast, it was still a somewhat smaller SoC – still powering devices such as the iPad Pro line-up, and a corresponding lower TDP, naturally still losing out to larger more power-hungry chips from the competition.

Anandtech.com

1

Review

- Two delay components in logical effort:
  - Parasitic delay (p)
  - Effort delay (F)
    - Logical effort (g): intrinsic complexity of the gate
    - Electrical effort (h): load capacitance dependent
- To minimize the delay all stages should have the same effort (h)
- Ideal effort is 4

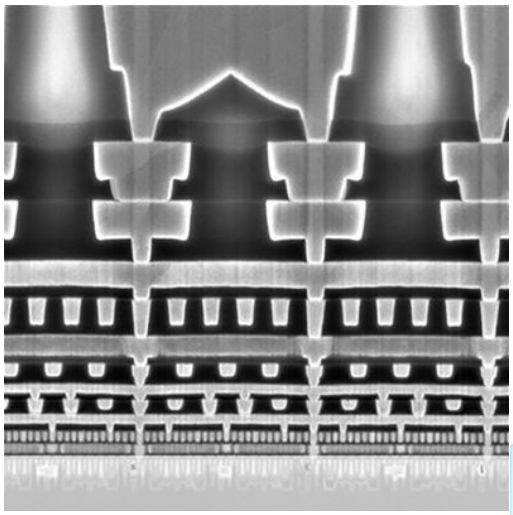
2

Wires

3

A modern technology is mostly wires

- Transistors are little things under the wires
- Many layers of wires
- Wires are as important as transistors
  - Speed and power



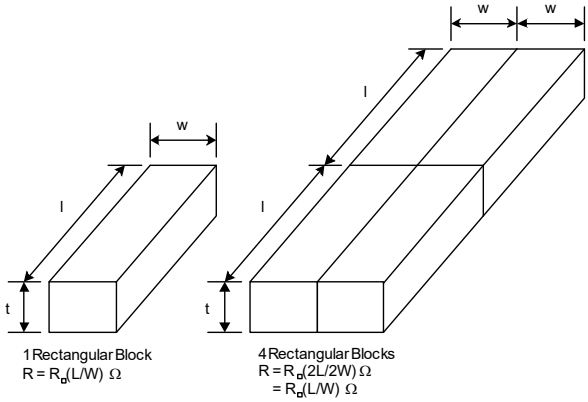
4

Wire Resistance

- ρ = *resistivity* (Ω*m)

$$R = \frac{\rho}{t}\frac{l}{w} = R_{\square}\frac{l}{w}$$

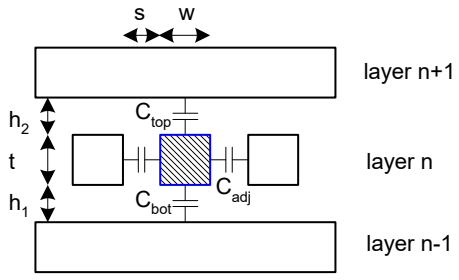
- $R_{\square}$ = *sheet resistance* (Ω/□)
  - □ is a dimensionless unit(!)
- Count number of squares
  - $R = R_{\square} * (\#\text{ of squares})$



5

Wire Capacitance

- Wire has capacitance per unit length
  - To neighbors
  - To layers above and below
- $C_{total} = C_{top} + C_{bot} + 2C_{adj}$



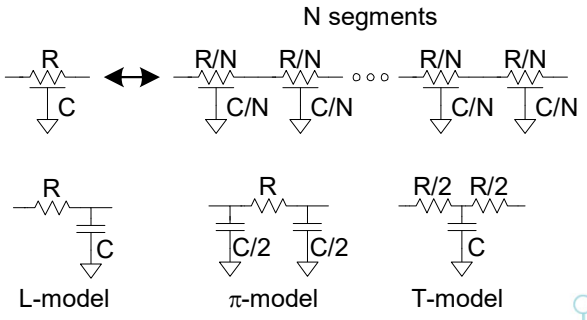
6

Wire Delay

7

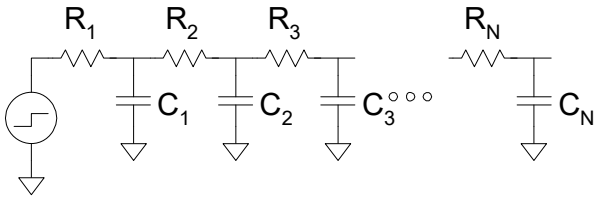
Wire RC Model

- Wires are a distributed system
  - Approximate with lumped element models
- 3-segment pi-model is accurate to 3% in simulation



8

Elmore Delay for RC Tree

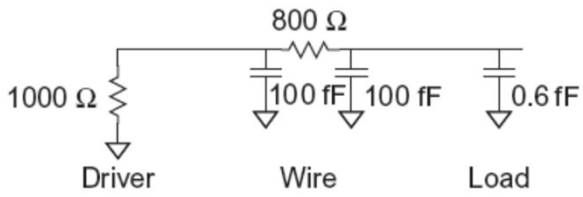


$$t_{pd} \approx \sum_{\text{nodes } i} R_{i-to-source} C_i$$

$$= R_1 C_1 + (R_1 + R_2) C_2 + \ldots + (R_1 + R_2 + \ldots + R_N) C_N$$

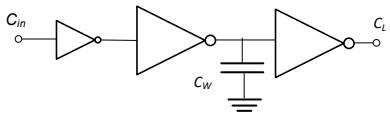
9

Example: RC Delay with Wire and Gate



10

Logical Effort with Wires



11

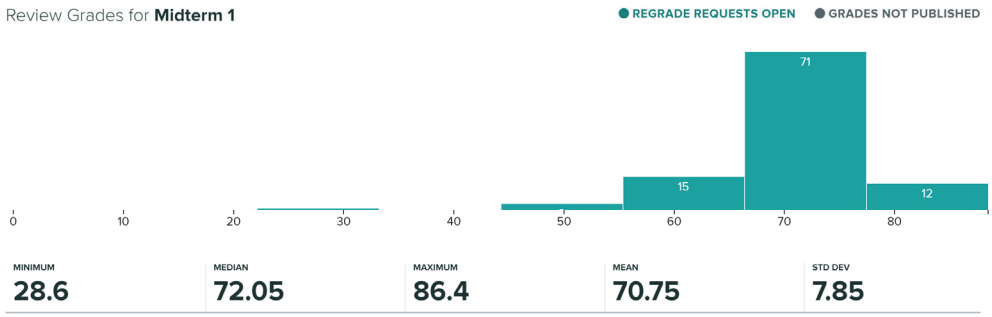
Administrivia

- Homework 6 due this week
  - Homework 7 next week
- All labs need to be checked off by next week!
- Projects (ASIC and FPGA) start this week
- Midterm 2 is on November 4 at 7pm
- Courses:
  - EECS251B will be offered in Spring (pending Campus approval)
  - EE194/290C SoC Design

12

Midterm 1 Scores

EECS151: Average: 69.7/84 (83%)
EECS251B: Average 86.4/88 (83%)

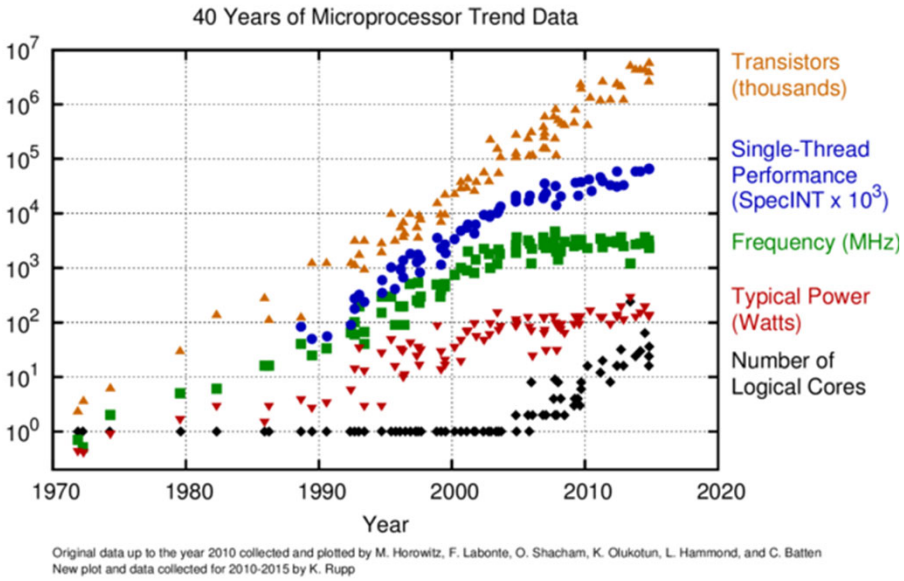


13

Energy

14

Processor Frequency Scaling



Original data up to the year 2010 collected and plotted by M. Horowitz, F. Labonte, O. Shacham, K. Olukotun, L. Hammond, and C. Batten
New plot and data collected for 2010-2015 by K. Rupp

15

Power and Energy

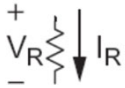
- Power is drawn from a voltage source attached to the V_{DD} pin(s) of a chip.
- Instantaneous Power: $P(t) = I(t)V(t)$
- Energy: $E = \int_0^T P(t)dt$
- Average Power: $P_{avg} = \frac{E}{T} = \frac{1}{T}\int_0^T P(t)dt$

16

Power in a Circuit Element

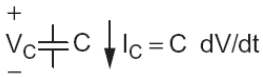
$$P_{VDD}(t) = I_{DD}(t) V_{DD}$$

$$P_R(t) = \frac{V_R^2(t)}{R} = I_R^2(t) R$$



$$E_C = \int_0^\infty I(t) V(t)\, dt = \int_0^\infty C \frac{dV}{dt} V(t)\, dt$$

$$= C \int_0^{V_C} V(t)\, dV = \tfrac{1}{2} C V_C^2$$



17

Sources of Power Dissipation

- $P_{total} = P_{dynamic} + P_{static}$
- Dynamic power: $P_{dynamic} = P_{switching} + P_{shortcircuit}$
  - Switching load capacitances
  - Short-circuit current
- Static power: $P_{static} = (I_{sub} + I_{gate} + I_{junct} + I_{contention})V_{DD}$
  - Subthreshold leakage
  - Gate leakage
  - Junction leakage
  - Contention current

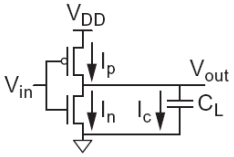
18

Dynamic Power

19

Charging and Discharging a Capacitor

- When the gate output rises
  - Energy stored in capacitor is $E_C = \frac{1}{2} C_L V_{DD}^2$
  - But energy drawn from the supply is



$$E_{VDD} = \int_0^\infty I(t) V_{DD} dt = \int_0^\infty C_L \frac{dV}{dt} V_{DD} dt$$

$$= C_L V_{DD} \int_0^{V_{DD}} dV = C_L V_{DD}^2$$

  - Half the energy from V_{DD} is dissipated in the pMOS transistor as heat, other half stored in capacitor
- When the gate output transitions HL
  - Energy in capacitor is dumped to GND
  - Dissipated as heat in the NMOS transistor

20

Dynamic Power Reduction

How can we limit switching power?

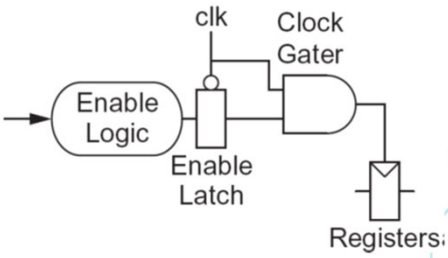
- Try to minimize:
  - Activity factor
  - Capacitance
  - Supply voltage
  - Frequency

$$P_{\text{switching}} = \alpha C V_{DD}^{\ 2} f$$

Reduce Activity Factor

$$P_{\text{switching}} = \alpha C V_{DD}^{\ 2} f$$

- Clock gating
- The best way to reduce the activity is to turn off the clock to registers in unused blocks
  - Saves clock activity (a = 1)
  - Eliminates all switching activity in the block
  - Requires determining if block will be used



Reduce Capacitance

$$P_{\text{switching}} = \alpha C V_{DD}^{\ 2} f$$

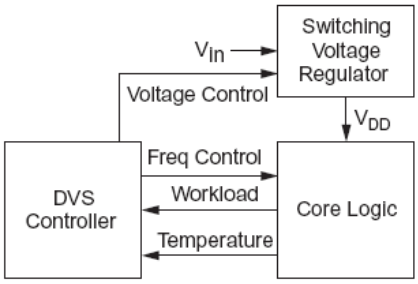
- Gate capacitance
  - Fewer stages of logic
  - Smaller gate sizes
- Wire capacitance
  - Good floorplanning to keep communicating blocks close to each other

23

Reduce Voltage/Frequency

$$P_{\text{switching}} = \alpha C V_{DD}^{\ 2} f$$

- Run each block at the lowest possible voltage and frequency that meets performance requirements
- Voltage domains
  - Provide separate supplies to different blocks
- Dynamic voltage/frequency scaling
  - Adjust V_{DD} and f according to workload

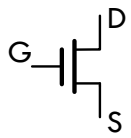


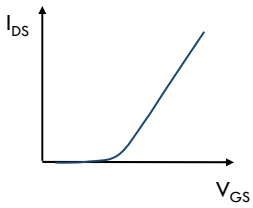
24

Leakage Power

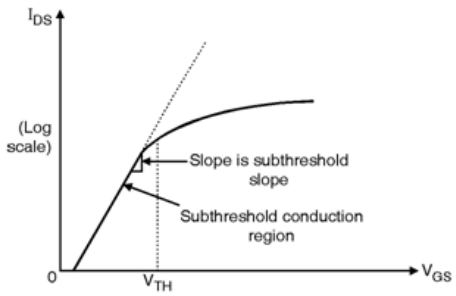
25

Subthreshold Leakage





Nearly linear
$I_{DS} \sim K(V_{GS}-V_{Th})$

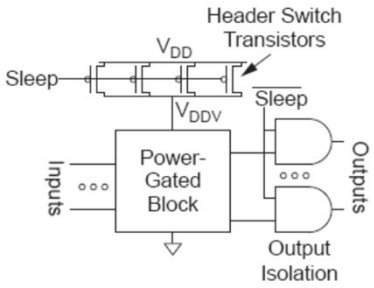


I_{DS} Vs V_{GS} characteristics in log scale

26

Power Gating

- Turn OFF power to blocks when they are idle to save leakage
  - Use virtual V_{DD} (V_{DDV})
  - Gate outputs to prevent invalid logic levels to next block
- Voltage drop across sleep transistor degrades performance during normal operation
  - Size the transistor wide enough to minimize impact
- Switching wide sleep transistor costs dynamic power
  - Only justified when circuit sleeps long enough



27

Example: Power Management

- Power states

	C0 HFM	C0 LFM	C1/C2	C4	C6
Core Voltage	▮	▮	▮	▮	▁
Core Clock	⎍	⎍	OFF	OFF	OFF
PLL	⎍	⎍	⎍	OFF	OFF
L1 Caches	≡	≡	Flushed	Flushed	OFF
L2 Caches	≡	≡	≡	Partial Flush	OFF
Wake-Up Time	active	active	< 1 µs	< 30 µs	< 100 µs
Power	▮	▮	▁	▁	▁

28

Summary

- Wire contributes to delay, especially in modern technology
- We can use RC model to capture wire delays
- Energy becomes an increasingly important optimization goal
  - Dynamic energy
  - Static energy