EE141-Fall 2012 Digital Integrated Circuits

Lecture 17
CMOS Scaling

Announcements

- Homework #7 due today
  - Project #1 out today, due next Thurs.
- Midterm 2 two weeks from today
- Let us know ASAP if you still don't have a project partner

CMOS Transistor Scaling

Goals of Technology Scaling

- Make things cheaper:
  - Want to sell more functions (transistors) per chip for the same money
  - Or build same products cheaper
  - Price of a transistor has to be reduced
- But also want to be faster, smaller, lower power…

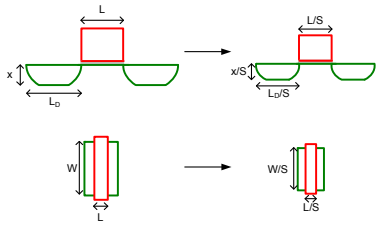
Technology Scaling

- Benefits of 30% "Dennard" scaling (1974):
  - Double transistor density
  - Reduce gate delay by 30% (increase operating frequency by 43%)
  - Reduce energy per transition by 65% (50% power savings @ 43% increase in frequency)
- Die size used to increase by 14% per generation (not any more)
- Technology generation spans 2-3 years

Technology Scaling Models

- **Full Scaling (Constant Electrical Field)**
  ideal model — dimensions and voltages scale together by the same factor *S*
- **Fixed Voltage Scaling**
  most common model until 1990's
  only dimensions scale, voltages remain constant
- **General Scaling**
  most realistic for today's situation —
  voltages and dimensions scale with different factors

Scaling



Full Scaling (Dennard, Long-Channel)

- W, L, t_{ox}: 1/S
- V_{DD}, V_T: 1/S

- Area: WL
- C_{ox}: 1/tox
- C_L: C_{ox}WL
- I_D: $C_{ox}(W/L)(V_{DD}-V_T)^2$
- R_{eq}: V_{DD}/I_{DSAT}

Full Scaling (Dennard, Long-Channel)

- W, L, t_{ox}: 1/S
- V_{DD}, V_T: 1/S

- t_p: $R_{eq}C_L$
- P_{avg}: $C_L V_{DD}^2/t_p$
- P_{avg}/A: $C_{ox}V_{DD}^2/t_p$

Scaling Relationships for Long Channel Devices

Parameter	Relation	Full Scaling	General Scaling	Fixed Voltage Scaling
W, L, t_{ox}		1/S	1/S	1/S
V_{DD}, V_T		1/S	1/U	1
N_{SUB}	V/W_{depl}^2	S	S^2/U	S^2
Area/Device	WL	$1/S^2$	$1/S^2$	$1/S^2$
C_{ox}	$1/t_{ox}$	S	S	S
C_L	C_{ox}WL	1/S	1/S	1/S
k_n, k_p	C_{ox}W/L	S	S	S
I_{av}	$k_{n,p}V^2$	1/S	S/U^2	S
t_p (intrinsic)	$C_L V / I_{av}$	1/S	U/S^2	$1/S^2$
P_{av}	$C_L V^2/t_p$	$1/S^2$	S/U^3	S
PDP	$C_L V^2$	$1/S^3$	$1/SU^2$	1/S

Table 3.1: Scaling Relationships for Long Channel Devices

Full Scaling (Dennard, Short-Channel)

- W, L, t_{ox}: 1/S
- V_{DD}, V_T: 1/S

- Area: WL
- C_{ox}: 1/tox
- C_L: C_{ox}WL
- I_D: $WC_{ox}v_{sat}(V_{DD}-V_T)^2/(V_{DD}-V_T-E_{crit}L)$
- R_{eq}: V_{DD}/I_{DSAT}

Full Scaling (Dennard, Short-Channel)

- W, L, t_{ox}: 1/S
- V_{DD}, V_T: 1/S

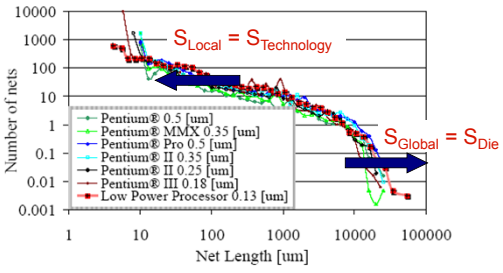
- t_p: $R_{eq}C_L$
- P_{avg}: $C_L V_{DD}^2/t_p$
- P_{avg}/A: $C_{ox}V_{DD}^2/t_p$

Transistor Scaling (Velocity-Saturated Devices)

Parameter	Relation	Full Scaling	General Scaling	Fixed-Voltage Scaling
W, L, t_{ox}		$1/S$	$1/S$	$1/S$
V_{DD}, V_T		$1/S$	$1/U$	1
N_{SUB}	V/W_{dep}^2	S	S^2/U	S^2
Area/Device	WL	$1/S^2$	$1/S^2$	$1/S^2$
C_{ox}	$1/t_{ox}$	S	S	S
C_{gate}	$C_{ox}WL$	$1/S$	$1/S$	$1/S$
k_n, k_p	$C_{ox}W/L$	S	S	S
I_{sat}	$C_{ox}WV$	$1/S$	$1/U$	1
Current Density	I_{sat}/Area	S	S^2/U	S^2
Ron	V/I_{sat}	1	1	1
Intrinsic Delay	$R_{on}C_{gate}$	$1/S$	$1/S$	$1/S$
P	$I_{sat}V$	$1/S^2$	$1/U^2$	1
Power Density	P/Area	1	S^2/U^2	S^2

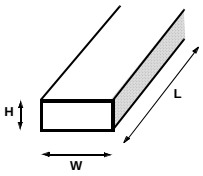
Interconnect Scaling

Interconnect Length Distribution



From Magen et al., "Interconnect Power Dissipation in a Microprocessor"

Resistance Scaling (local)



Scale W, H, and L:

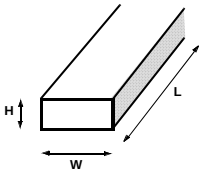
$R_w = \rho L/(WH)$

R_w α (1/S) / [(1/S) (1/S)]

→R_w α S

(R/□ α S)

Resistance Scaling (global)



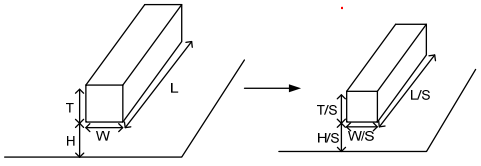
Scale W, H, constant L:

$R_w = \rho L/(WH)$

R_w α 1/[(1/S) (1/S)]

→ R_w α S^2

Local Wire Scaling (Scenario 1)

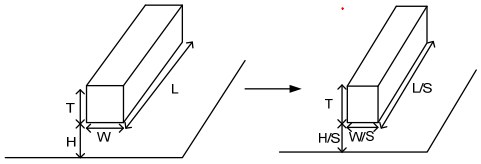


C_{pp} α WL/H
C_{fringe} α ~L
R_w α L/(WT)
t_{pwire} α R_wC_w

C_{pp}' α 1/S
C_{fringe}' α 1/S
R_w' α S
t_{pwire}' const.

Bad news: gates speed up by S…

Local Wire Scaling (Scenario 2)

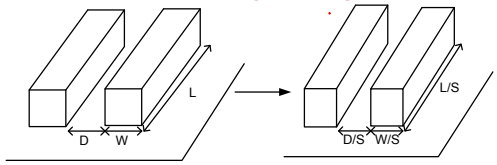


C_{pp} α WL/H
C_{fringe} α ~L
R_w α L/(WT)
t_{pwire} α R_wC_w

C_{pp}' α 1/S
C_{fringe}' α 1/S
R_w' const.
t_{pwire}' α 1/S

Better (wire RC tracks inverters), but…

Scenario 2: Intralayer Capacitance

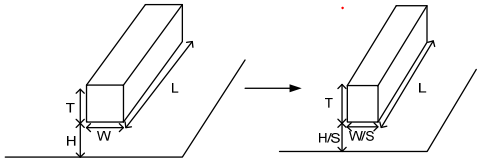


$C_{pp,side}$ α LT/D

$C_{pp,side}$' const.

- $C_{pp,side}$/Length increases
  → Crosstalk, coupling issues get worse
- Aspect ratio limited – eventually have to scale T
  - Different metal layers have different T

Global Wire Scaling (Scenario 2)



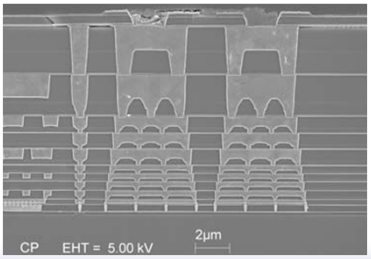
C_{pp} α WL/H
C_{fringe} α ~L
R_w α L/(WT)
t_{pwire} α R_wC_w

C_{pp}' const
C_{fringe}' ~const
R_w' α S
t_{pwire}' α S

Very bad: wire delay S^2 worse than gates

Modern Interconnect

- 90nm process



Next Lecture

- Ratioed and Pass Transistor Logic