ESE534 Computer Organization

Day 6: February 1, 2012

Energy, Power, Reliability

Today

- Energy tradeoffs
- Voltage limits and leakage
- Variations

At Issue

- Many now argue **energy** will be the ultimate scaling limit
  - (not lithography, costs, …)
- Proliferation of portable and handheld devices
  - …battery size and life biggest issues
- Cooling, energy costs may dominate cost of electronics
  - Even server room applications

Preclass 1

- 1GHz case
  - Voltage?
  - Energy per Operation?
  - Power required for 2 processors?
- 2GHz case
  - Voltage?
  - Energy per Operation?
  - Power required for 1 processor?

Energy and Delay

$$E = \frac{1}{2}CV^2$$

$$\tau_{gd}=Q/I=(CV)/I$$

$$I_{d,sat}=(\mu C_{OX}/2)(W/L)(V_{gs}-V_{TH})^2$$

Energy/Delay Tradeoff

$$E = \frac{1}{2}CV^2$$

$$¿_{gd}=(CV)/I$$

$$I_{d,sat} \propto (V_{gs}-V_{TH})^2$$

- $E \approx V^2$
- $\tau_{gd} \approx 1/V$

- We can trade speed for energy
- $E \times (\tau_{gd})^2 \approx$ constant

Martin *et al.* *Power-Aware Computing*, Kluwer 2001
http://caltechcstr.library.caltech.edu/308/

Area/Time Tradeoff

- Also have Area-Time tradeoffs
  - HW2 spatial vs temporal multipliers
  - See more next week
- Compensate slowdown with additional parallelism
- …trade Area for Energy ➔ Architectural Option

Question

- By how much can we reduce energy?
- What limits us?

Challenge: Power

Origin of Power Challenge

- Limited capacity to remove heat
  - ~100W/cm² force air
  - 1-10W/cm² ambient
- Transistors per chip grow at Moore's Law rate = $(1/F)^2$
- Energy/transistor must decrease at this rate to keep constant power density
- $P/tr \propto CV^2f$
- $E/tr \propto CV^2$
  - …but V scaling more slowly than F

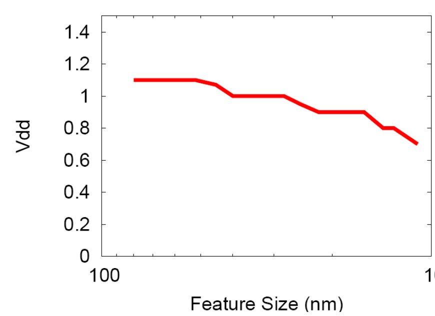
Energy per Operation

$$E = \frac{1}{2}CV^2$$

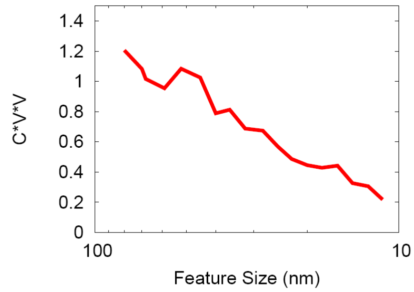
C_{total} = # transistors × C_{tr}

C_{tr} scales (down) as F
transistors scales as F^{-2}
…ok if V scales as F…

ITRS V_{dd} Scaling: More slowly than F

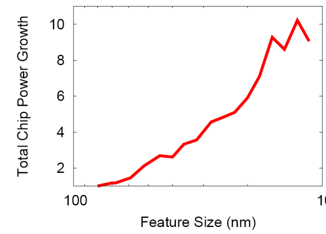


ITRS CV^2 Scaling: More slowly than $(1/F)^2$

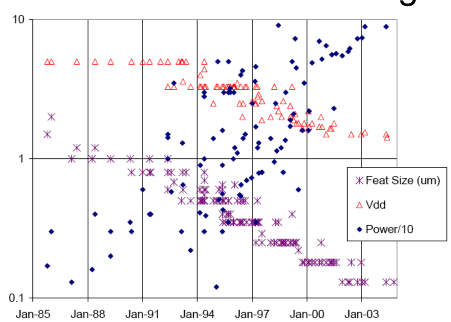


Origin of Power Challenge

- Transistors per chip grow at Moore's Law rate = $(1/F)^2$
- Energy/transistor must decrease at this rate to keep constant
- $E/tr \propto CV^2$

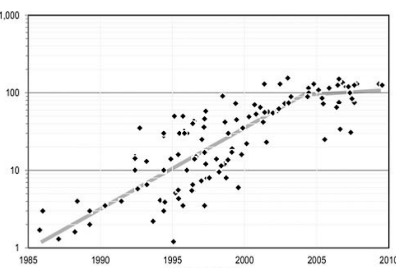


Historical Power Scaling



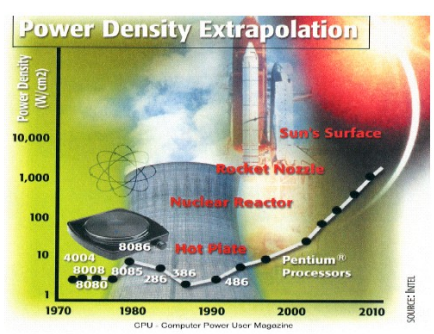
[Horowitz et al. / IEDM 2005]

Microprocessor Power Density

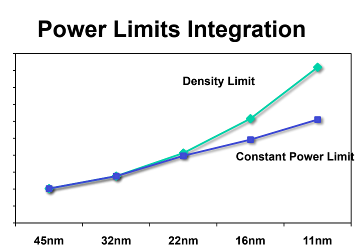


The Future of Computing Performance: Game Over or Next Level?
National Academy Press, 2011
http://www.nap.edu/catalog.php?record_id=12980

Intel Power Density



Impact



Source: Carter/Intel

Impact

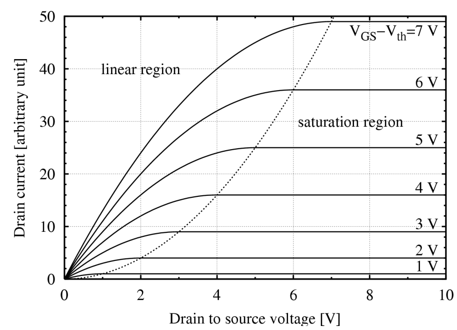
- Power density is limiting scaling
  - Can already place more transistors on a chip than we can afford to turn on!
- Power is potential challenge/limiter for all future chips.
  - Only turn on small percentage of transistors?
  - Operate those transistors as much slower frequency?
  - Find a way to drop V_{dd}?

How far can we reduce V_{dd}?

Limits

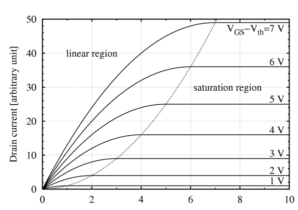
- Ability to turn off the transistor
- Parameter Variations
- Noise (not covered today)

MOSFET Conduction



From: http://en.wikipedia.org/wiki/File:IvsV_mosfet.png

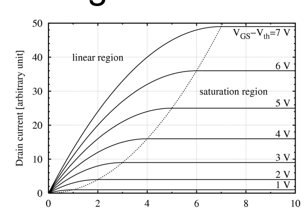
Transistor Conduction



- Three regions
  - Subthreshold ($V_{gs}<V_{TH}$)
  - Linear ($V_{gs}>V_{TH}$) and ($V_{ds} < (V_{gs}-V_{TH})$)
  - Saturation ($V_{gs}>V_{TH}$) and ($V_{ds} > (V_{gs}-V_{TH})$)

Saturation Region

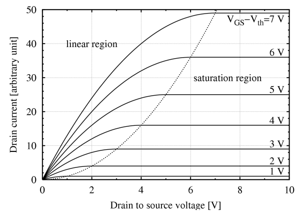
- ($V_{gs}>V_{TH}$)
- ($V_{ds} > (V_{gs}-V_{TH})$)



$$I_{ds,sat}=(\mu C_{OX}/2)(W/L)(V_{gs}-V_{TH})^2$$

Linear Region

- $(V_{gs} > V_{TH})$
- $(V_{ds} < (V_{gs} - V_{TH}))$



$I_{ds,lin} = (\mu C_{OX})(W/L)((V_{gs} - V_{TH})V_{ds} - (V_{ds})^2/2)$

Subthreshold Region

- $(V_{gs} < V_{TH})$

$$I_{sub} = I_{VT} \times 10^{((V_{gs} - V_{TH})/S)}$$

$$S = (\ln(10))\eta kT/q$$

[Frank, IBM J. R&D v46n2/3p235]

Operating a Transistor

- Concerned about I_{on} and I_{off}
- I_{on} drive (saturation) current for charging
  - Determines speed: $T_{gd} = CV/I$
- I_{off} leakage current
  - Determines leakage power/energy:
    - $P_{leak} = V \times I_{leak}$
    - $E_{leak} = V \times I_{leak} \times T_{cycle}$

Leakage

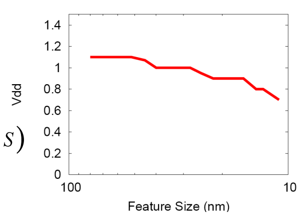
- To avoid leakage want I_{off} very small
- Switch V from V_{dd} to 0
- V_{gs} in off state is 0 $(V_{gs} < V_{TH})$

$$I_{sub} = I_{VT} \times 10^{((V_{gs} - V_{TH})/S)}$$

$$I_{off} = I_{VT} \times 10^{-((V_{TH})/S)}$$

Leakage

$$I_{off} = I_{VT} \times 10^{-((V_{TH})/S)}$$



- $S \approx 90$mV for single gate
- $S \approx 70$mV for double gate
- For lowest leakage, want S small, V_{TH} large
- 4 orders of magnitude $I_{VT}/I_{off} \rightarrow V_{TH} > 280$mV

Leakage limits V_{TH} in turn limits V_{dd}

How maximize I_{on}/I_{off}?

- Maximize I_{on}/I_{off} – for given V_{dd} ? $E_{sw} \propto CV_2$
- Get to pick V_{TH}, V_{dd}

$$I_{d,sat} = (\mu C_{OX}/2)(W/L)(V_{gs} - V_{TH})^2$$

$$I_{d,lin} = (\mu C_{OX})(W/L)(V_{gs} - V_{TH})V_{ds} - (V_{ds})^2/2$$

$$I_{sub} = I_{VT} \times 10^{((V_{gs} - V_{TH})/S)}$$

Preclass 2

- $E = E_{sw} + E_{leak}$
- $E_{leak} = V \times I_{leak} \times T_{cycle}$
- $E_{sw} \propto CV_2$

$$I_{sub} = I_{VT} \times 10^{\left(\left(V_{gs} - V_{TH}\right)/S\right)}$$

- $I_{chip\text{-}leak} = N_{devices} \times I_{tr\text{-}leak}$

Preclass 2

- $E_{leak}(V)$?
- $T_{cycle}(V)$?

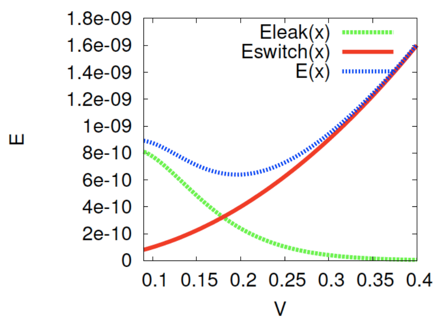
In Class

- Assign calculations
  - SIMD – each student computes for a different Voltage
- Collect results on board
  - Should go quick once students have time to calculate
- Identify minimum energy point and discuss

Values

V	T(v)	Esw(V)	Eleak(V)	E(V)
0.36	3.6E-09	1.296E-09	1.296E-11	1.30896E-09
0.27	0.000000027	7.29E-10	7.29E-11	8.019E-10
0.24	5.17064E-08	5.76E-10	1.24095E-10	7.00095E-10
0.21	9.74734E-08	4.41E-10	2.04694E-10	6.45694E-10
0.205	1.08137E-07	4.2025E-10	2.21682E-10	6.41932E-10
0.2	1.19897E-07	4E-10	2.39794E-10	6.39794E-10
0.19	1.4711E-07	3.61E-10	2.79509E-10	6.40509E-10
0.18	0.00000018	3.24E-10	3.24E-10	6.48E-10
0.15	3.23165E-07	2.25E-10	4.84748E-10	7.09748E-10
0.12	5.56991E-07	1.44E-10	6.68389E-10	8.12389E-10
0.09	0.0000009	8.1E-11	8.1E-10	8.91E-10

Graph for In Class

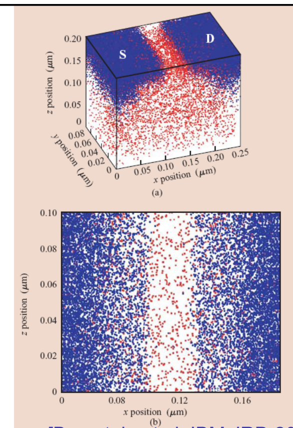


Impact

- Subthreshold slope prevents us from scaling voltage down arbitrarily.
- Induces a minimum operating energy.

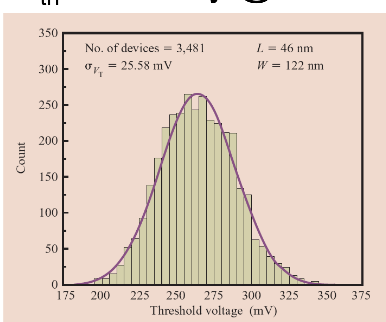
Challenge: Variation

Statistical Dopant Count and Placement



[Bernstein et al, IBM JRD 2006]

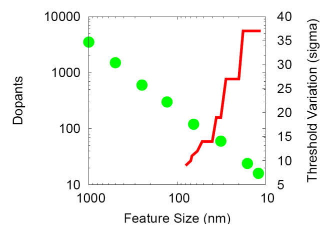
V_{th} Variability @ 65nm



[Bernstein et al, IBM JRD 2006]

Variation

- Fewer dopants, atoms → increasing Variation
- How do we deal with variation?



% variation in V_{TH}
(From ITRS prediction)

Impact of Variation?

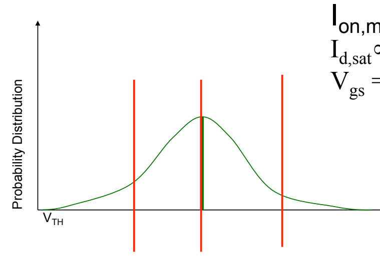
- Higher V_{TH}?
  - Not drive as strongly → slower
  - $I_{d,sat} \propto (V_{gs}-V_{TH})^2$
- Lower V_{TH}?
  - Not turn off as well → leaks more

$$I_{off} = I_{VT} \times 10^{-((V_{TH})/S)}$$

Variation

- Margin for expected variation
- Must assume V_{TH} can be any value in range

$I_{on,min}=I_{on}(V_{th,max})$
$I_{d,sat} \propto (V_{gs}-V_{TH})^2$
$V_{gs} = V_{dd}$



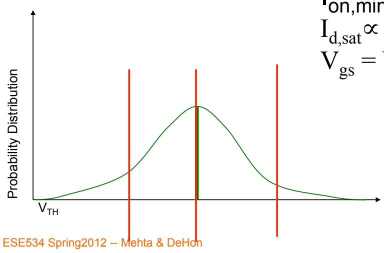
Margining

- Must raise V_{dd} to increase drive strength
- Increase energy

$I_{on,min}=I_{on}(V_{th,max})$

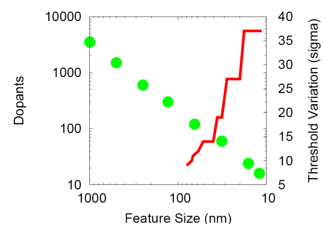
$I_{d,sat} \propto (V_{gs}-V_{TH})^2$

$V_{gs} = V_{dd}$

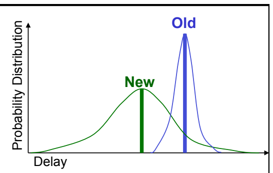


Variation

- Increasing variation forces higher voltages
  - On top of our leakage limits

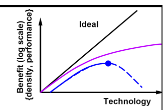


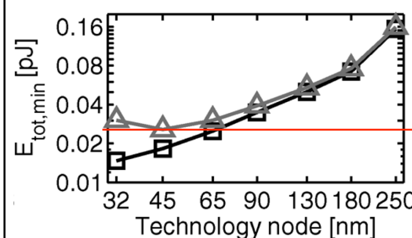
Variations



- Margins growing due to increasing variation
- Margined value may be worse than older technology?

End of Energy Scaling?





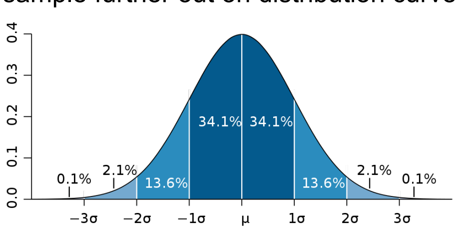
Black nominal

Grey with variation

[Bol *et al.*, IEEE TR VLSI Sys 17(10):1508—1519]

Chips Growing

- Larger chips (billions of transistors) → sample further out on distribution curve



From: http://en.wikipedia.org/wiki/File:Standard_deviation_diagram.svg

Admin

- Homework due Monday
  - Section 3.5 has changed
  - Please grab updated copy
- Reading for Monday on web
- André back on Monday

Big Ideas

- Can trade time for energy
  - … area for energy
- Variation and leakage limit voltage scaling
- Power major limiter going forward
  - Can put more transistors on a chip than can switch
- Continued scaling demands
  - Deal with noisier components
    - High variation
    - … other noise sources