EECS 427
Lecture 20: System-Level Power Reduction
Reading: 11.7

Last Time

- Memory reliability is crucial
  - Soft errors are a growing source of errors
  - Use redundant rows and columns combined with error correcting codes
- Memory power density is low but memory comprises a growing portion of total chip area
  - Leakage is primary because of low switching activity and large transistor width

Lecture Overview

- Power reduction techniques
  - Primary focus on static power
  - Earlier we discussed dual-Vdd design → aimed at dynamic power
    - Cost: Requires a 2nd power supply, must route a 2nd power grid, level conversion penalties
  - Many of these are cutting-edge; not commonly implemented in designs today
  - Important to differentiate between standby mode and active mode in leakage reduction techniques

Key to reducing power: exploit slack

• Timing slack is defined at each (output) node in the circuit

== (Required arrival time (RAT) – Latest arrival time (LAT))

Large positive slack means the gate driving the node can be slowed down a lot

Negative slack means the timing goal cannot be met (bad)

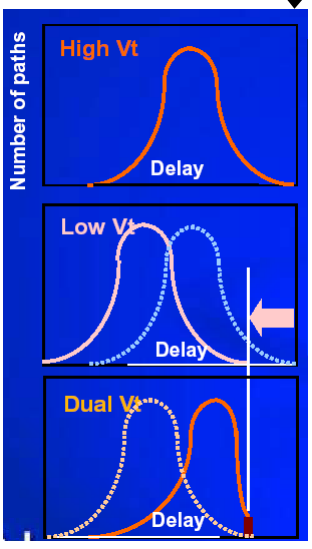
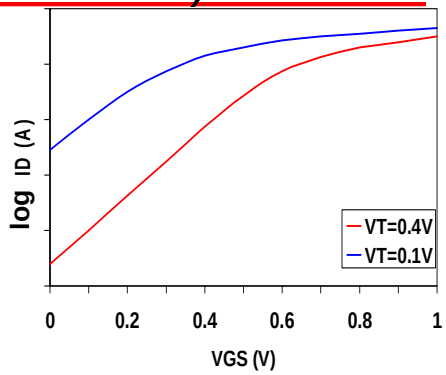


Fig courtesy Borkar,Intel

Leakage as a function of V_{th} (from Lecture 4)

- Reducing the V_{th} **increases** the sub-threshold leakage current (exponentially)
  - 90mV reduction in V_{th} increases leakage by 10X
- But, reducing V_{th} **decreases** gate delay (increases performance)

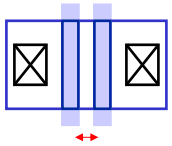


- Determine the critical paths during the design phase, use low V_{th} devices on those paths for speed
- Use a high V_{th} in rest of logic to control leakage
  - Can provide total leakage reduction of up to 80%

Thanks to Irwin/Narayanan

V_{th} Assignment Granularity

- V_{th} assignment can be at different levels of granularity
  - Gate level assignment
  - Pull up network / Pull down network based assignment (half gate)
    - Single V_{th} in pull up or pull down networks
  - Stack based assignment
    - Single V_{th} in series connected transistors
  - Individually assignment within transistor stacks
    - Possible area penalty (see right)

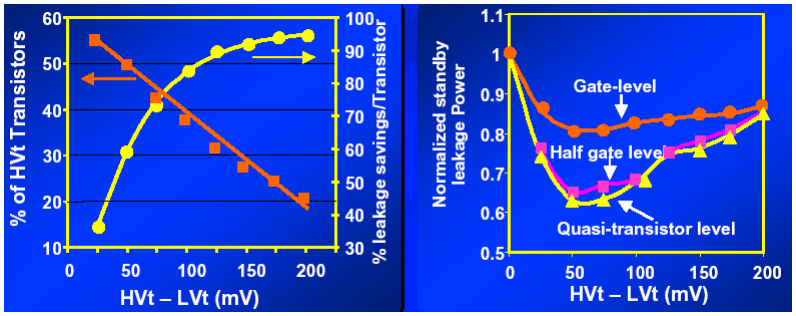


Design rule constraint for different V_t assignment

- Number of library cells increases with finer control
  - Better leakage / delay trade-off
  - Harder for synthesis tools to handle

Courtesy: Blaauw

Choice of optimal 2nd Vth



ΔVth usually chosen to be about 100mV

Too close together – not enough performance differential between the choices

Too far – high-Vth becomes too slow or low-Vth too leaky

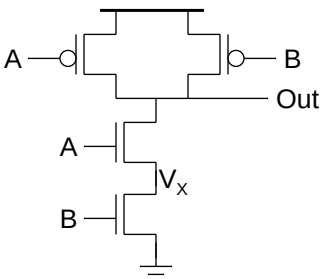
Ref: Wei, VLSI 2000

Stack Effect

- Leakage is a function of the circuit topology and the value of the inputs

$$V_{th} = V_{T0} + \gamma(\sqrt{|-2\phi_F + V_{SB}|} - \sqrt{|-2\phi_F|})$$

where V_{T0} is the threshold voltage at $V_{SB} = 0$; V_{SB} is the source-bulk (substrate) voltage; γ is the **body-effect coefficient**



A	B	V_X	I_{SUB}
0	0	V_{th} ln(1+n)	$V_{GS}=V_{BS}= -V_X$
0	1	0	$V_{GS}=V_{BS}=0$
1	0	$V_{DD}-V_{th}$	$V_{GS}=V_{BS}=0$
1	1	0	$V_{SG}=V_{SB}=0$

- Leakage is least when A = B = 0
- Leakage reduction due to stacked transistors is called the **stack effect**

Thanks to Irwin/Narayanan

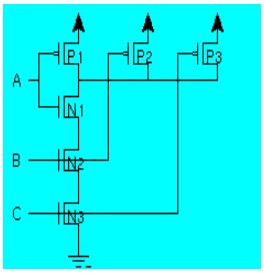
Short Channel Factors and Stack Effect

- Subthreshold leakage current depends on V_{GS},V_{BS} *and* V_{DS}
- V_{th} of a short-channel device decreases with increasing V_{DS} due to **DIBL** (drain-induced barrier loading)
  - Typical values for DIBL: a 50 to 120mV change in V_{th} per 1V change in V_{DS}
  - V_X reduces the drain-source voltage of the top NMOS, increasing its V_{th} and lowering its leakage
- V_X typically settles to ~50-100mV in steady state
  - $V_{GS} = V_{BS}$ = -100mV and $V_{DS} = V_{DD}$ -100mV, 20X reduction in leakage vs. $V_{GS} = V_{BS}$ = 0V and $V_{DS} = V_{DD}$

Thanks to Irwin/Narayanan

State-dependent leakage

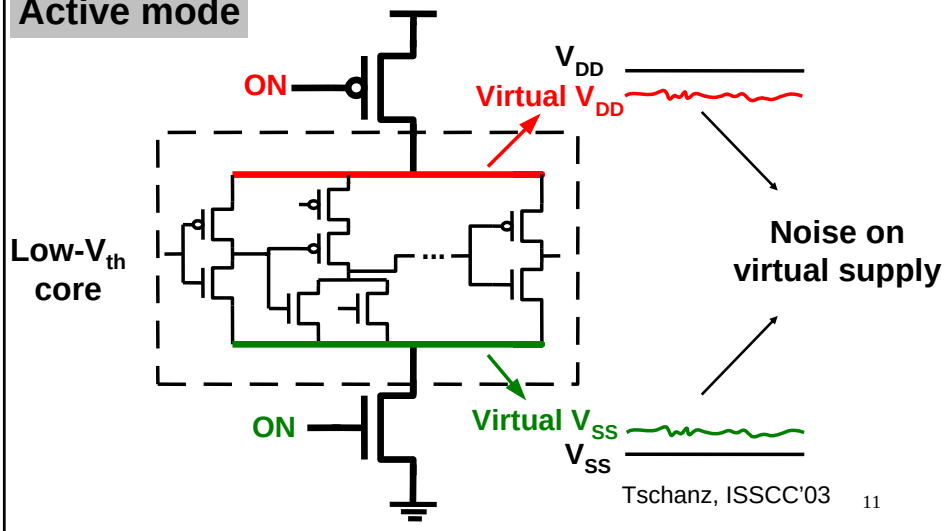
- # of states grows exponentially with # of gate inputs
- Only a few of the states have significant leakage
  - Dominant leakage states have only one transistor OFF in any path from V_{dd} to Gnd
- Exploit this by setting the entire circuit to a low leakage state in *standby* mode (MUX all registers)
  - Due to logical correlations, can't set all gates to their best states simultaneously



A	B	C	Leakage Current	Leaking Transistors
0	0	0	0.095	N1, N2, N3
0	0	1	0.195	N1, N2
0	1	0	0.195	N1, N3
0	*1*	*1*	*1.874*	*N1*
1	0	0	0.185	N2, N3
1	*0*	*1*	*1.220*	*N2*
1	*1*	*0*	*1.140*	*N3*
1	*1*	*1*	*9.410*	*P1, P2, P3*

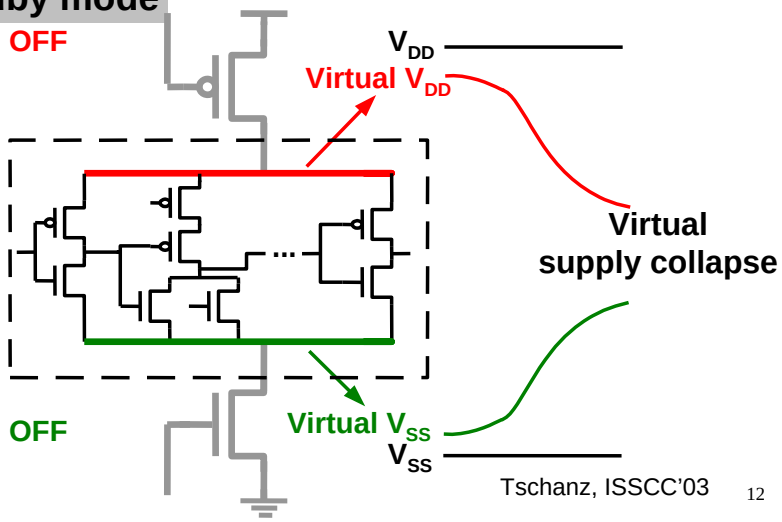
Dynamic sleep transistor

Active mode



Dynamic sleep transistor

Standby mode



How to Size the Sleep Transistor(s)?

- Circuits in active mode see the sleep transistor as extra power line resistance
  - The wider the sleep transistor, the better
- Wide sleep transistors cost area
  - Minimize the size of the sleep transistor for a tolerable ripple on virtual supply (e.g. 5%)
- Sleep transistor is not "free" – it will degrade performance in active mode
  - Typically by a few percent
- Charging and discharging of virtual rails consumes power as well

Sleep Transistor Results

A high-Vth sleep transistor has to be very wide for low resistance in linear region (MTCMOS in table)
A low-Vth sleep transistor (non-boosted sleep in table) needs much less area for the same resistance

	MTCMOS	Boosted Sleep	Non-Boosted Sleep
Sleep-TR size	**5.1%**	**2.3%**	**3.2%**
Leakage power reduction	**1450X**	**3130X**	**11.5X**
Virtual supply bounce	60 mV	59 mV	58 mV

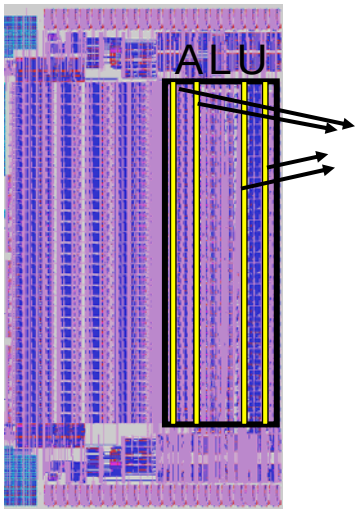
Size given as % area cost

Boosted → uses a larger gate voltage for the sleep signal

5% delay penalty

[R. Krishnamurthy]

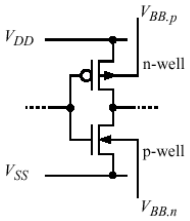
Sleep Transistor Layout



Area overhead	
PMOS	6%
NMOS	3%

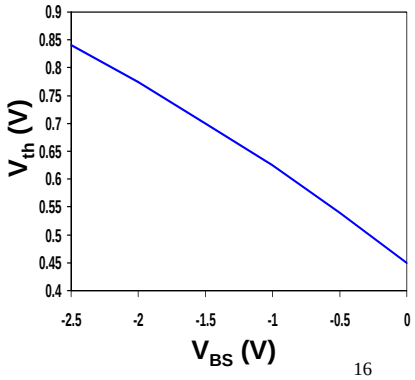
Tschanz, ISSCC'03

Body Biasing of Transistors



$$V_{th} = V_{T0} + \gamma(\sqrt{|-2\phi_F + V_{SB}|} - \sqrt{|-2\phi_F|})$$

- For NMOS, the substrate is normally tied to ground
- A negative bias causes V_{th} to increase from 0.45V to 0.85V
  - Requires bias generation circuitry
- Can adjust substrate bias based on workload (dynamically), process variation (static), or both
  - *Adaptive body-biasing* (ABB)

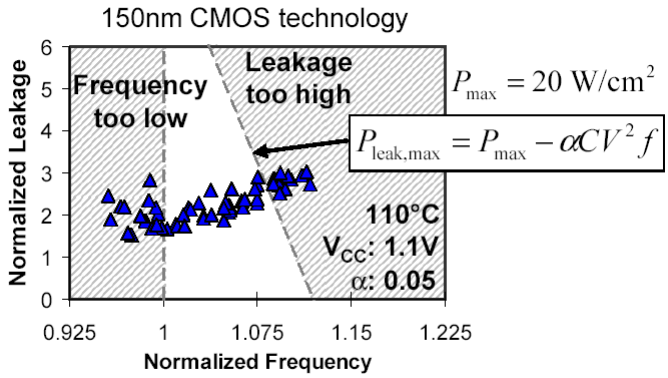


Thanks to Irwin/Narayanan

Body biasing pro/con

- Can employ forward or reverse bias
  - Improve yields in fast/slow dies (next slides)
- Limited range of Vth adjustability
  - Often ~100mV, or 10X leakage and 15% speed variability
- Energy cost of charging/discharging substrate/well caps
- Can't do in SOI technologies including future double-gate processes

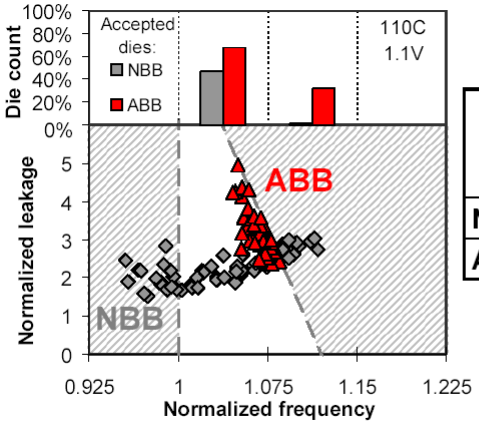
Substrate Biasing



Tschanz, JSSC 11/02

Effectiveness of Substrate Bias

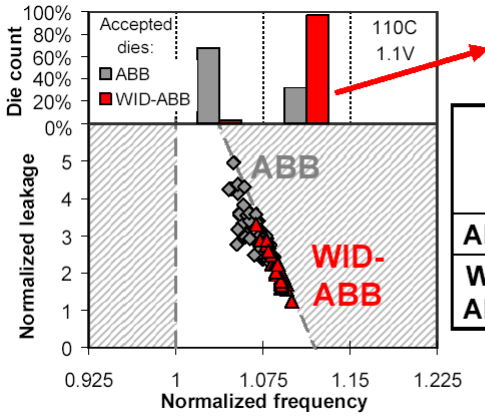
Die-to-die variations



	Frequency Variation σ/μ
NBB	4.1%
ABB	0.69%

Effectiveness of Substrate Bias

Within-die variations



	Frequency Variation σ/μ
ABB	0.69%
WID ABB	0.21%

Conclusions

- Lots of recent work on circuit and technology techniques to reduce static power
  - Standby mode leakage reduction can be orders of magnitude, may lose state, takes time to switch in and out of standby mode
  - Active mode leakage reduction is a tougher problem, smaller savings (<50% typically), must be ready for inputs to toggle at any time