Institut
Mines-Telecom

Digital CMOS Low Power Design

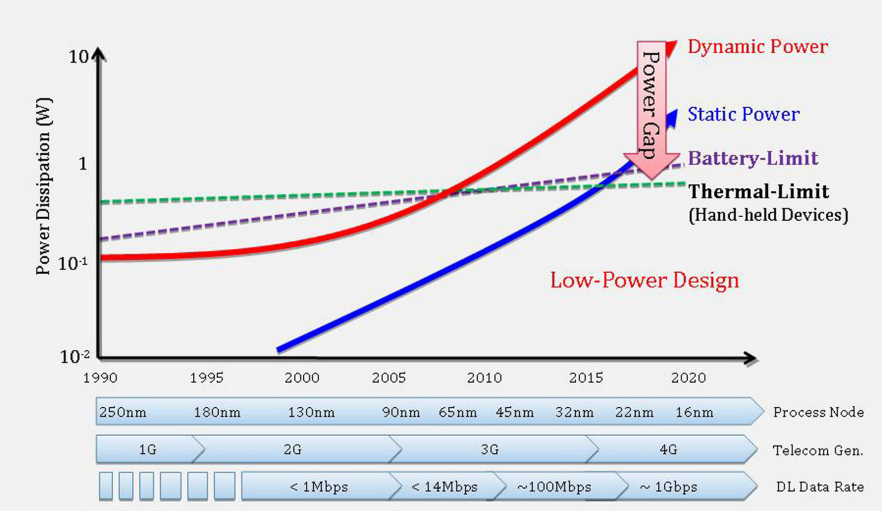
Chadi Jabbour

SE208 Electronics for embedded systems

Outline

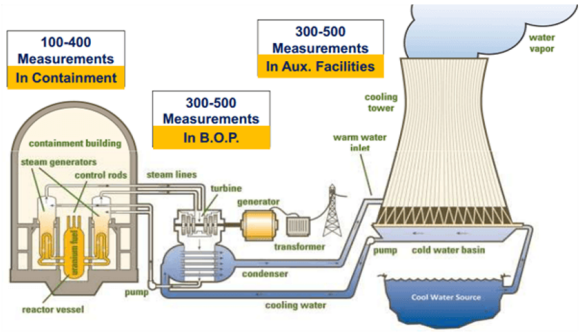
Outline

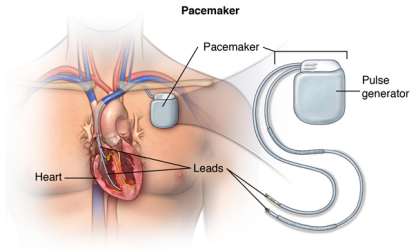
Why low power-1



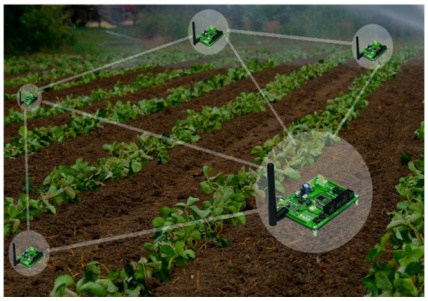
Battery life is not evolving as fast as silicon technologies!!!

Why low power-2





Charging or changing Batteries is not always a piece of cake

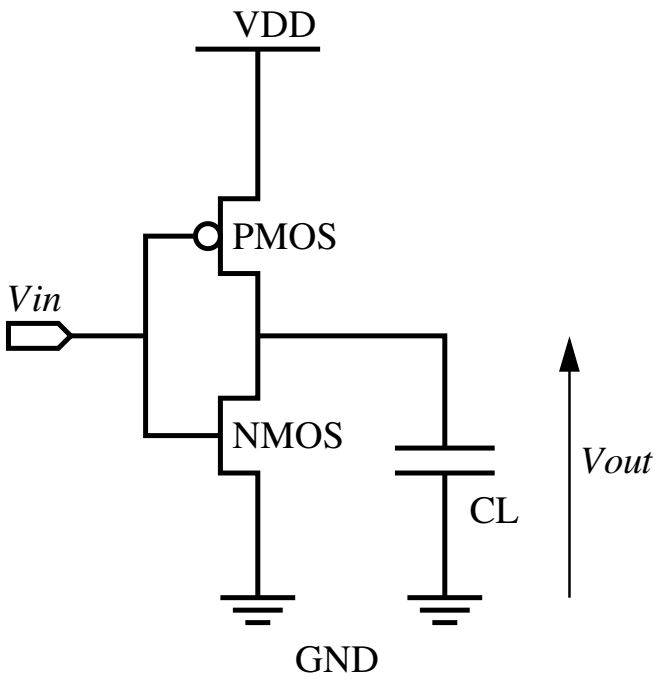


Why low power-3

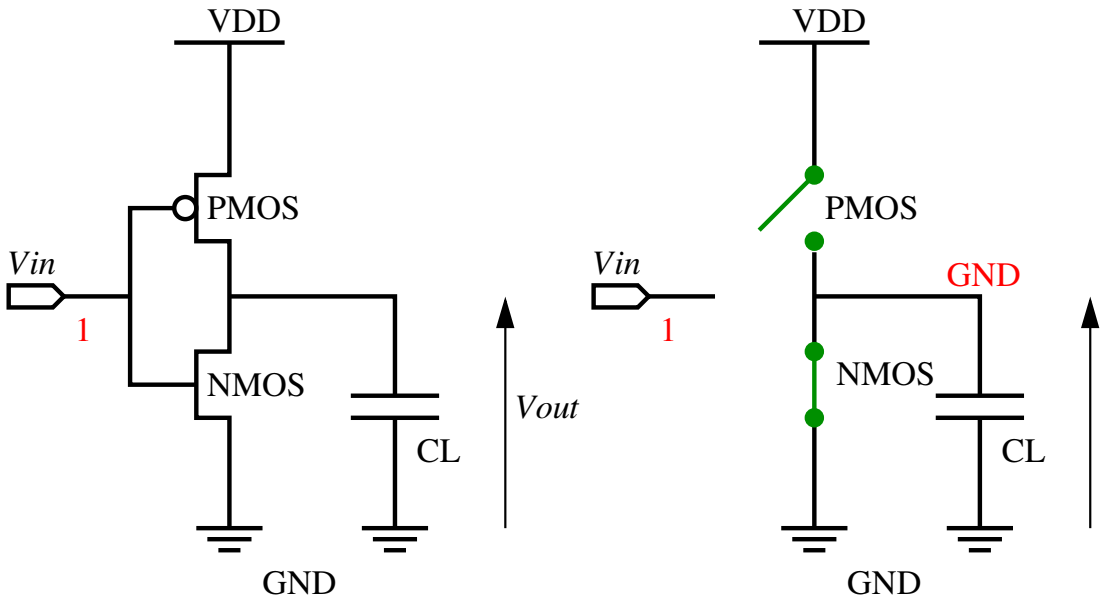
Power results in Heat $\Longrightarrow$ we need cooling $\Longrightarrow$ more power

Outline

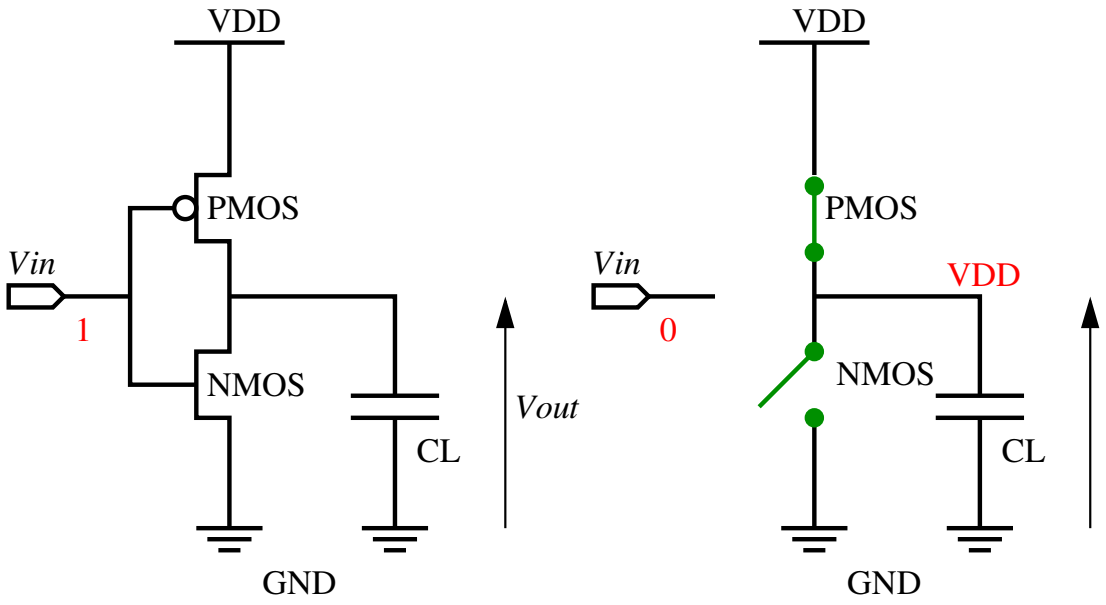
Example



Example



Example



CMOS Transistors state - reminder NMOS

OFF mode

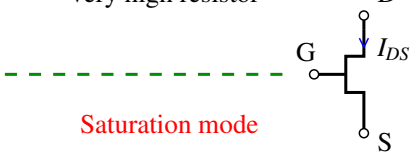
$V_{GS} = 0$
$I_{DS} \approx 0$

The transistor behaves as very high resistor

Sub-Threshold mode

$V_{GS} < V_{th}$
$I_{DS} = I_{D0}\frac{W}{L}e^{K_1.V_{GS}}$

The transistor behaves as exponential current source



Saturation mode

$0 < V_{GS} - Vth < V_{DS}$
$I_{DS} = \frac{1}{2}K_2\frac{W}{L}(V_{GS} - Vth)^2$

The transistor behaves as quadratic current source

Linear mode

$V_{GS} - Vth > V_{DS}$
$I_{DS} = K_2\frac{W}{L}(V_{GS} - Vth)V_{DS}$

The transistor behaves as resistor

V_{th}: Threshold Voltage - - - W : transistor width - - - L transistor length
I_{D0}, K_1 and K_2 are constants that depend on the technology and the transistor type

CMOS Transistors state - reminder PMOS

OFF mode

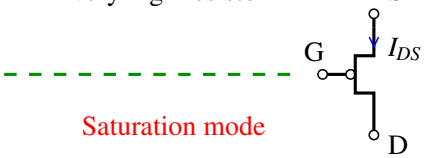
$V_{GS} = 0$
$I_{DS} \approx 0$

The transistor behaves as very high resistor

Sub-Threshold mode

$|V_{GS}| < |V_{th}|$
$I_{DS} = I_{D0}\frac{W}{L}e^{K_1.V_{GS}}$

The transistor behaves as exponential current source



Saturation mode

$0 < |V_{GS} - Vth| < |V_{DS}|$
$I_{DS} = \frac{1}{2}K_2\frac{W}{L}(V_{GS} - Vth)^2$

The transistor behaves as quadratic current source

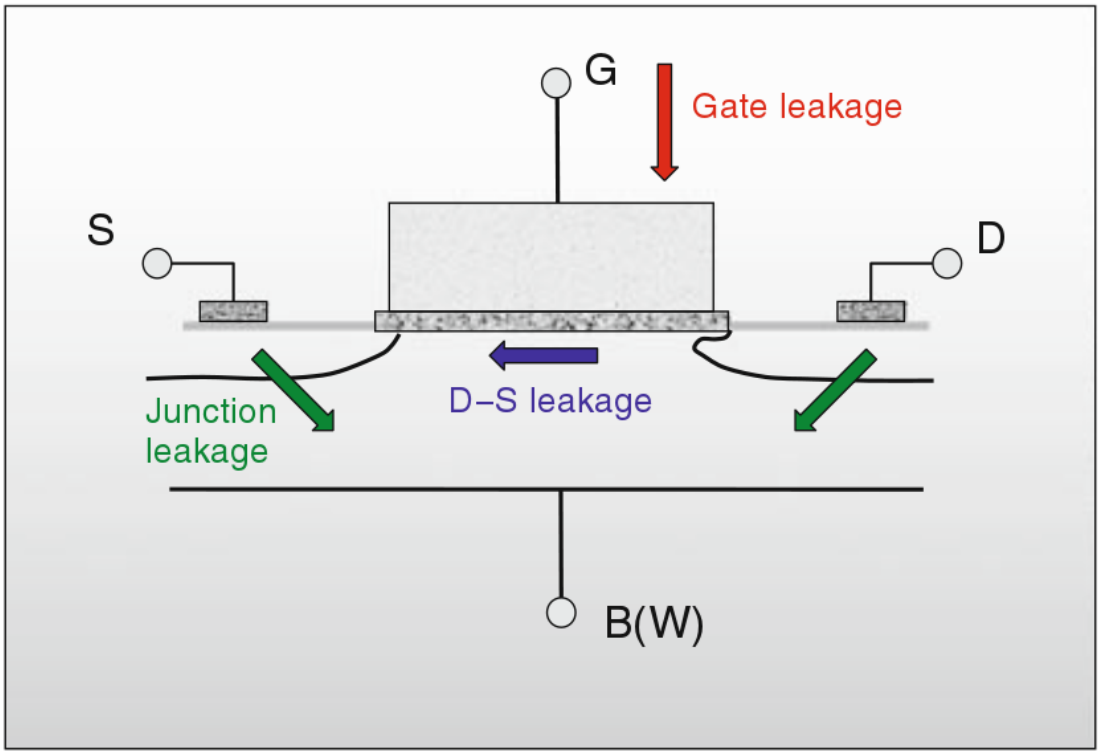
Linear mode

$|V_{GS} - Vth| > |V_{DS}|$
$I_{DS} = K_2\frac{W}{L}(V_{GS} - Vth)V_{DS}$

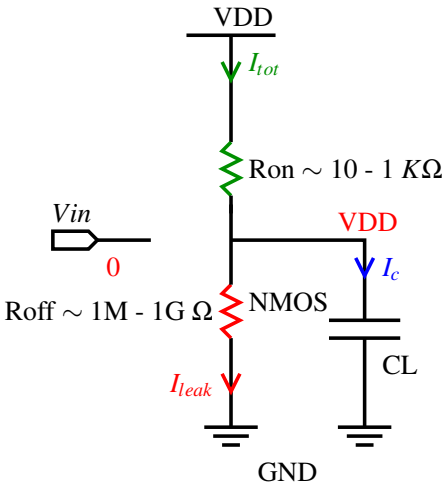
The transistor behaves as resistor

The behavior of a PMOS is similar to an NMOS except that V_{GS}, V_{DS} & $V_{th} < 0$
I_{D0}, V_{th}, K_1 and K_2 are not equal for PMOS and NMOS

Leakage Current



Power consumption calculation



1-Assuming *Roff* ∞, what is the energy needed to charge *CL* to VDD?

2-If *Vin* is a signal of frequency *freq* and a probability of toggling α, what will be the power consumption of our circuit?

Power consumption overall

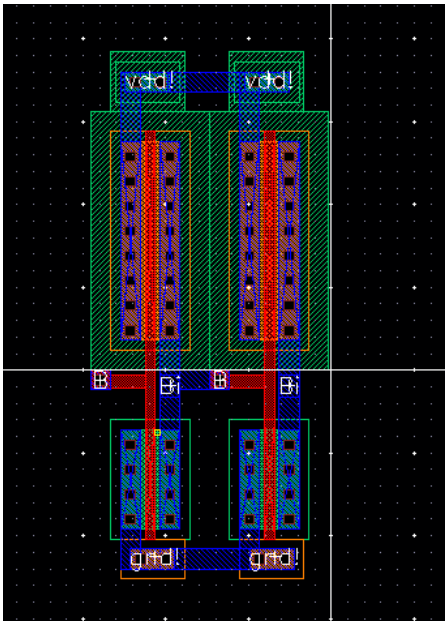
$$P = \underbrace{\frac{1}{2} \cdot \alpha \cdot CL \cdot freq \cdot VDD^2}_{Dynamic\ \ Power} + \underbrace{I_{leak} \cdot VDD}_{Leakage\ \ Power}$$

α is the activity or the probability to have a toggle in the gate
CL is the load
$freq$ is the operation frequency
VDD is the power supply
I_{leak} is the leakage current

Outline

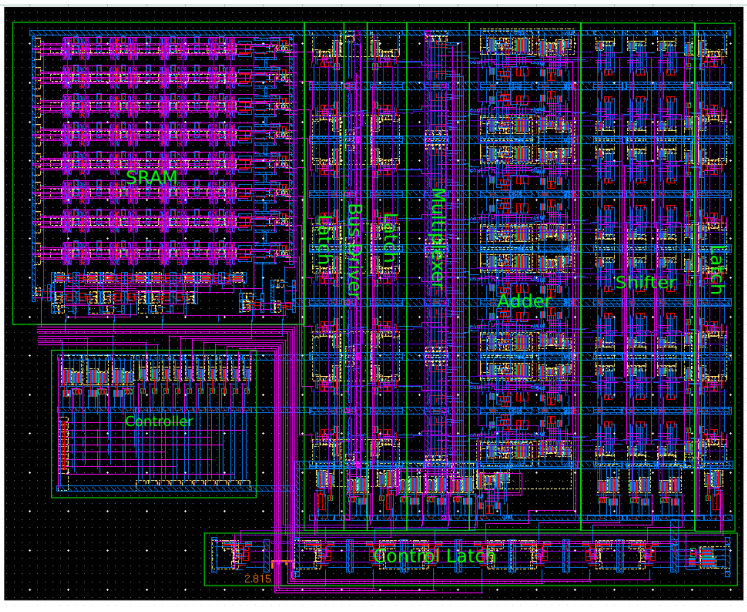
How to reduce *CL*

CL = Transistor capacitances + Interconnect capacitances



How to reduce *CL*

CL = Transistor capacitances + Interconnect capacitances



How to reduce CL

$$CL = \text{Transistor capacitances} + \text{Interconnect capacitances}$$

- The transistor capacitances are proportional to $W.L$ of the connected transistors
- The interconnect capacitances depend on the length and width of the wire.

How to reduce CL

$$CL = \text{Transistor capacitances} + \text{Interconnect capacitances}$$

- The transistor capacitances are proportional to $W.L$ of the connected transistors
- The interconnect capacitances depend on the length and width of the wire.

Hence, reducing CL can be done by using a smaller technology node (65 nm instead 130 nm, or 28 nm instead of 90 nm) but smaller technologies are:

- More expensive
- (often) more leaky

How to reduce α-encoding

Encoding consists in changing the representation of the information in a manner that reduces the average number of transitions.
Example: Grey Encoding

Binary		Gray-code	
State	No. of toggles	State	No. of toggles
000	-	000	-
001	1	001	1
010	2	011	1
011	1	010	1
100	3	110	1
101	1	111	1
110	2	101	1
111	1	100	1
000	3	000	1
Av. Transitions/clock = 1.75		Av. Transitions/clock = 1	

How to reduce α-encoding

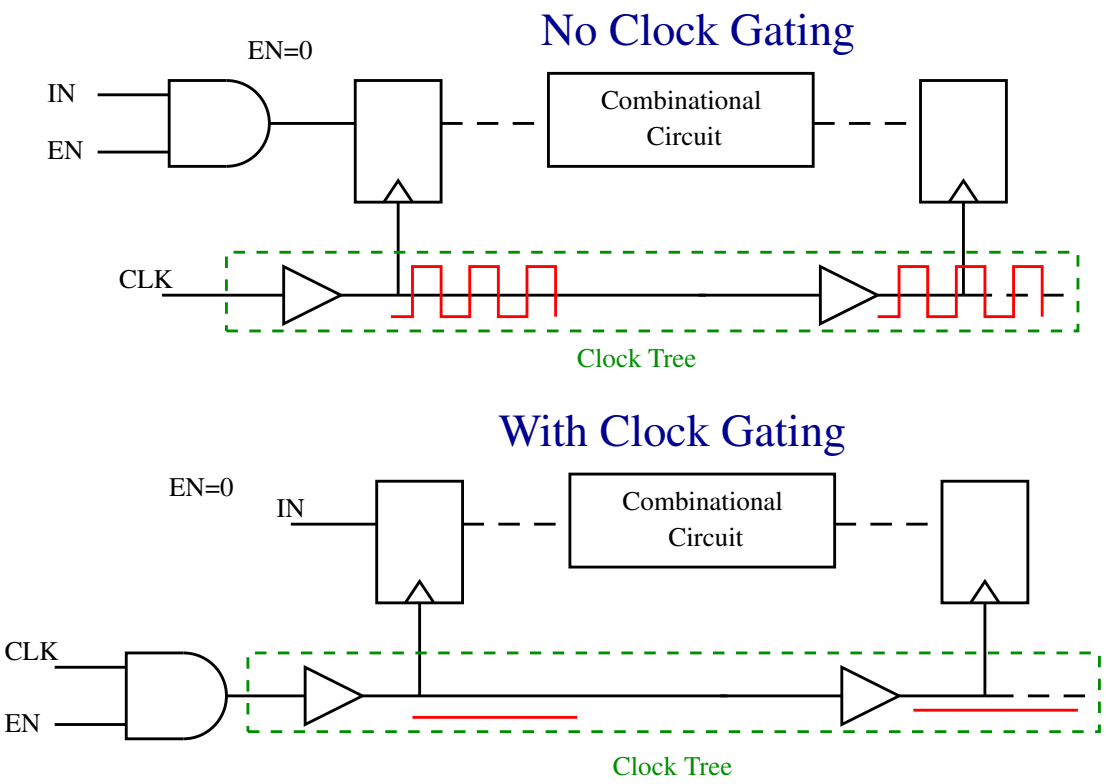
Example 2: Canonical Signed Digit (CSD)

$$101001\underbrace{111111} \Rightarrow 10101000000(-1)$$

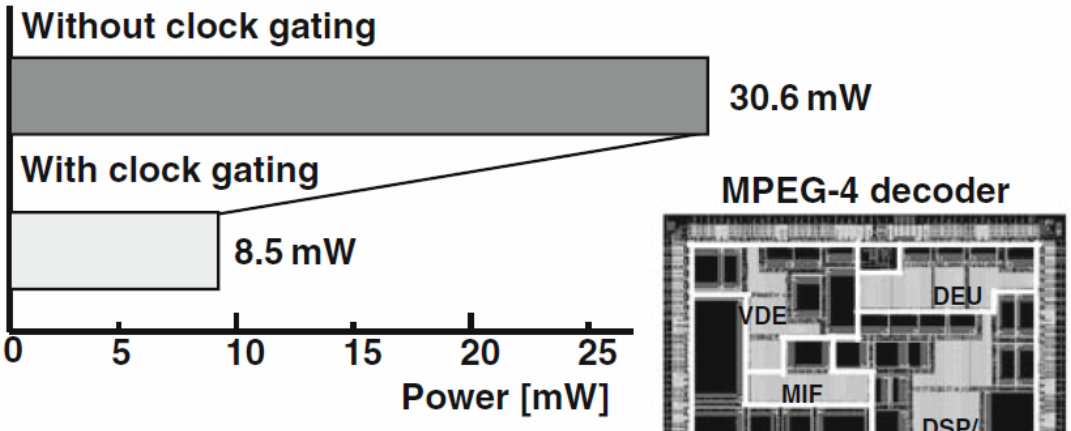
number	2's complement	CSD
3	011	$10\bar{1}$
2	011	010
1	001	001
0	000	000
-1	111	$00\bar{1}$
-2	110	$0\bar{1}0$
-3	101	$\bar{1}01$
-4	100	$\bar{1}00$

33% of non zero bits for CSD with respect to 50% for classical encoding

How to reduce α-Clock Gating



How to reduce α-Clock Gating Application



90% of FFs clock-gated.

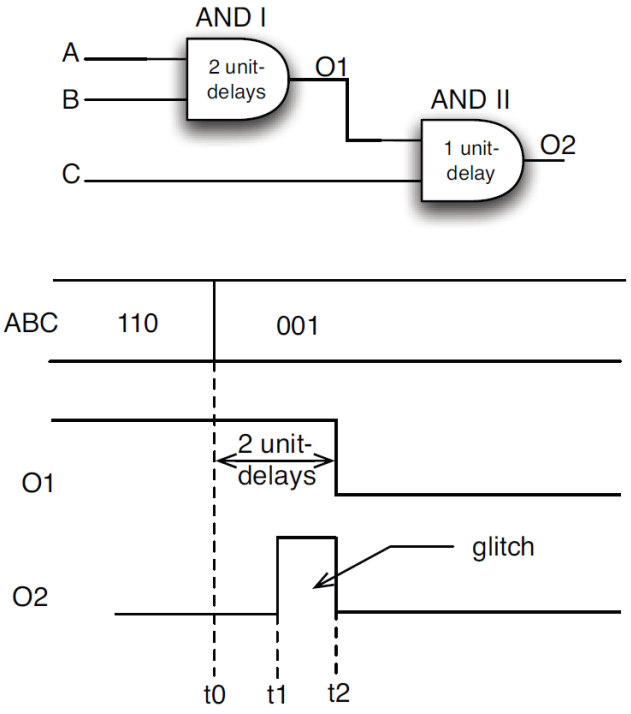
70% power reduction by clock gating alone.

[Ref: M. Ohashi, ISSCC'02]

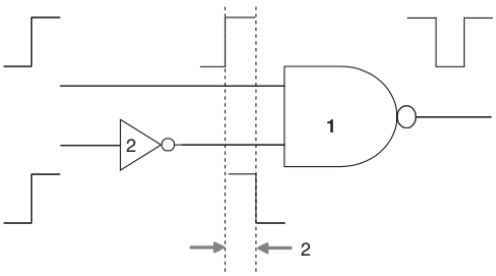
How to reduce α-Glitches

Glitches are temporary changes in the value of the output or unnecessary transitions

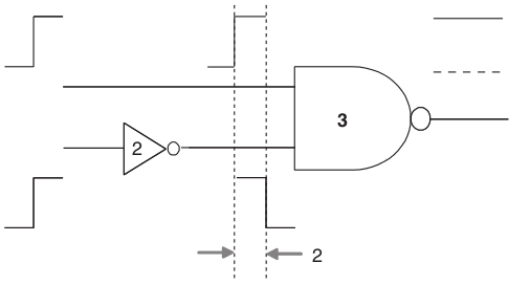
They are caused due to the skew in the input signals to a gate



How to reduce α-Reducing Glitches

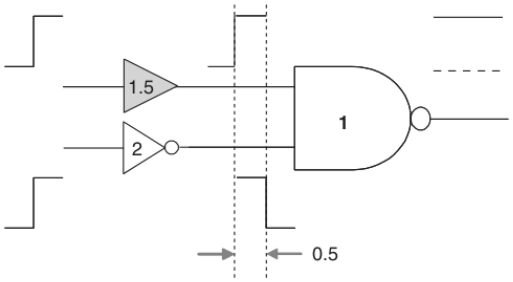


(a) Glitch at the output of one NAND gate



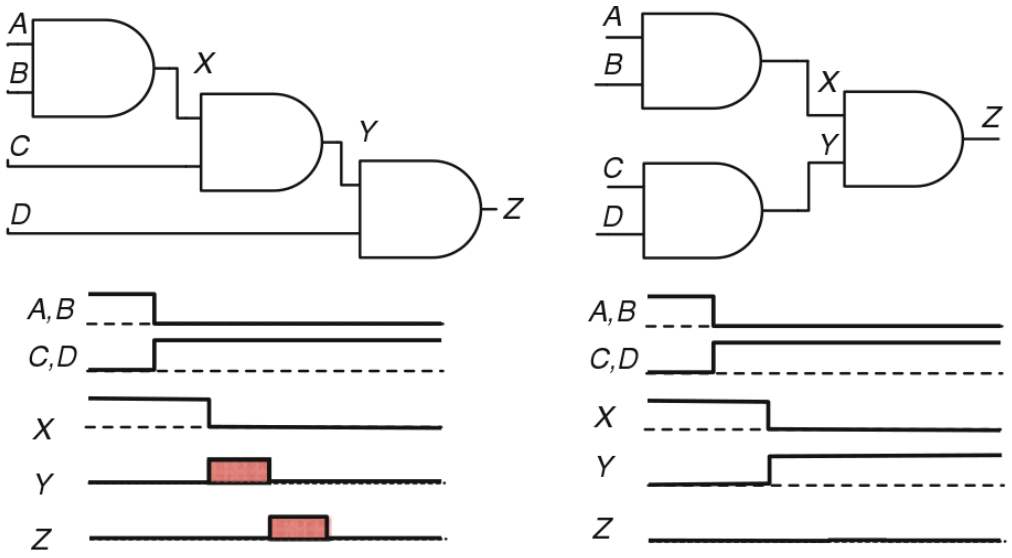
(b) Glitch elimination by hazard filtering

- b) A bad design does not necessarily lead to glitches
- c) Delay balancing can be very efficient to remove glitches but requires additional hardware



(c) Glitch elimination by path delay balancing

How to reduce α- Glitches Tree vs Chain



Uneven arrival times of input signals of gates due to unbalanced delay paths

Solution: balancing delay paths!

Reduce VDD - Impact

Reducing *VDD* has an impact of the delay of the gates

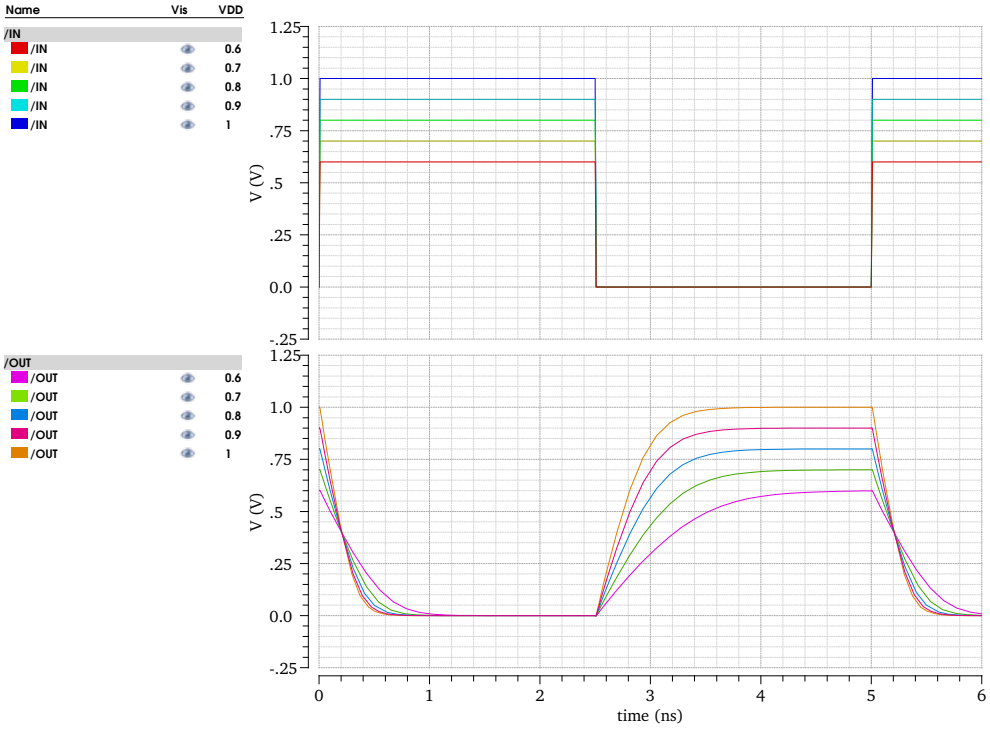
$$delay \propto \frac{VDD}{(VDD - V_{th})^K}$$

- K is a variable that depends on many parameters such as the technology, VDD, V_{th}.
- Its value is typically between 1.5 and 2.

Reducing VDD increases the delay and therefore the systems should be adapted to keep the same performance

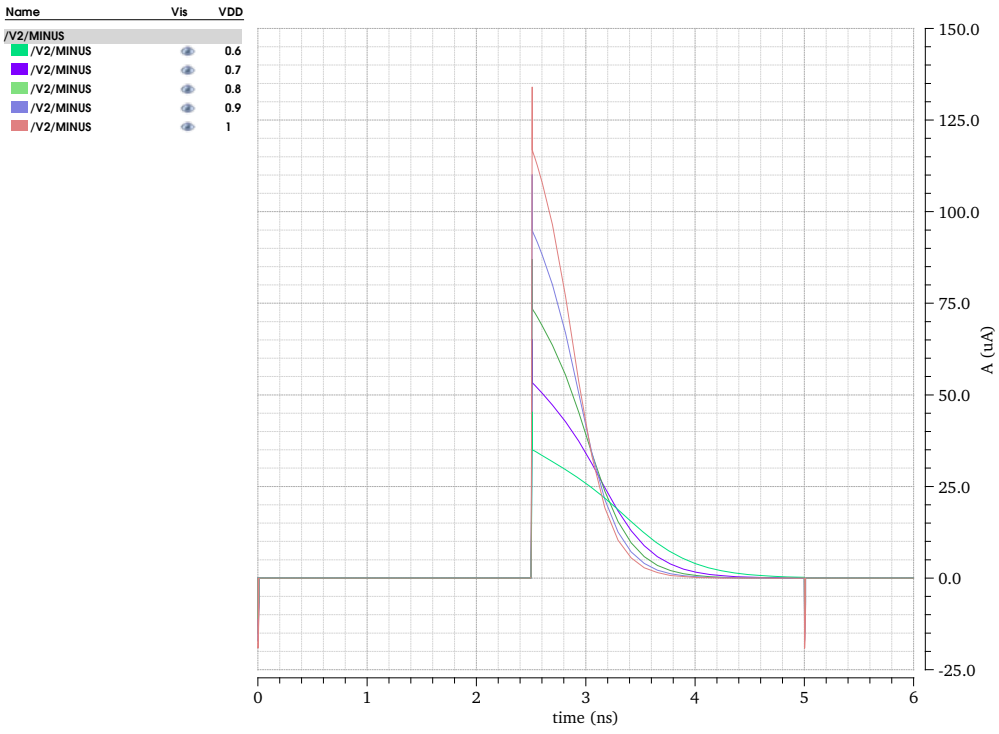
Reduce VDD - Impact

Inverter simulation in 65 nm CMOS

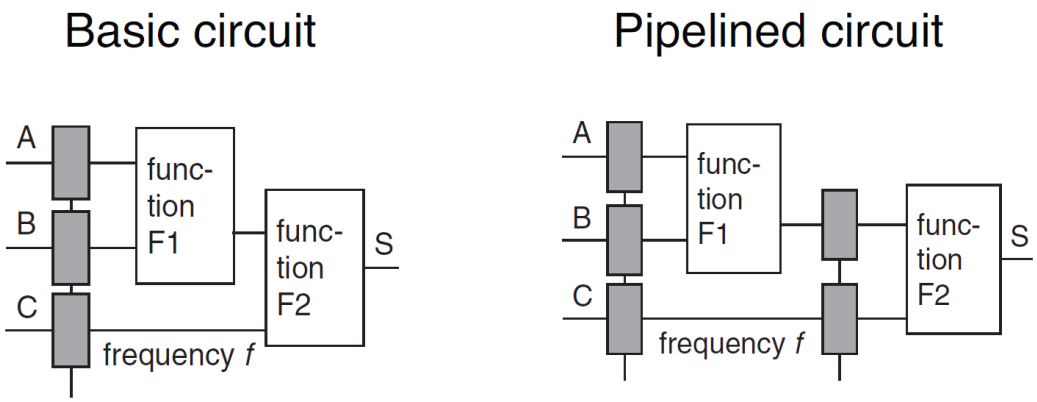


Reduce VDD - Impact

Inverter simulation in 65 nm CMOS



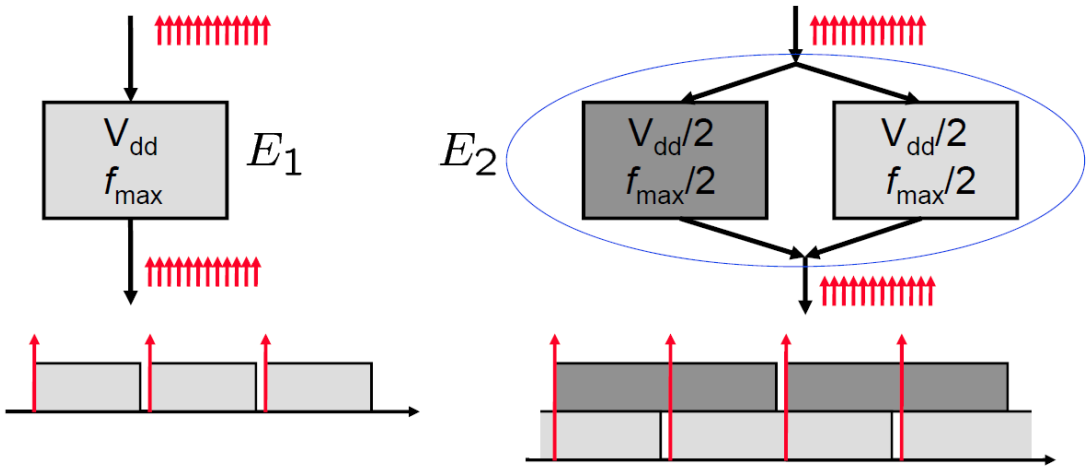
How to reduce VDD - pipelining



Pipelining allows to reduce the length of the critical path and thus to reduce VDD

But it increases the delay between the input and the output and it requires additional material (Flip-Flops)

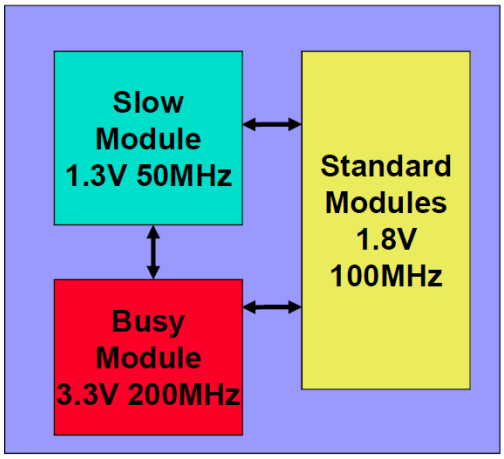
How to reduce the VDD-Parallelizing



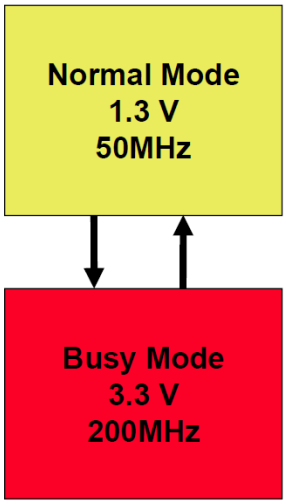
Parallelism allows to reduce the frequency per channel/path thus to reduce VDD

The circuit is duplicated $\Longrightarrow$ higher area, higher leakage (Maybe)

Multiple and Dynamic VDD



Not all components require same performance.

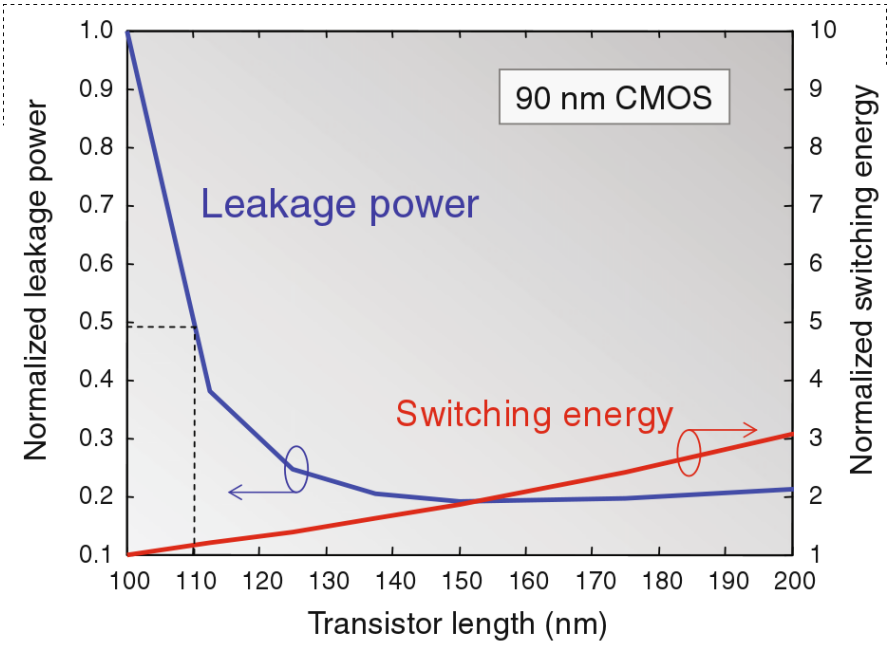


Required performance may change over time

Using multiple and Dynamic VDD reduces the power consumption
It increases the complexity of the power generation block

Outline

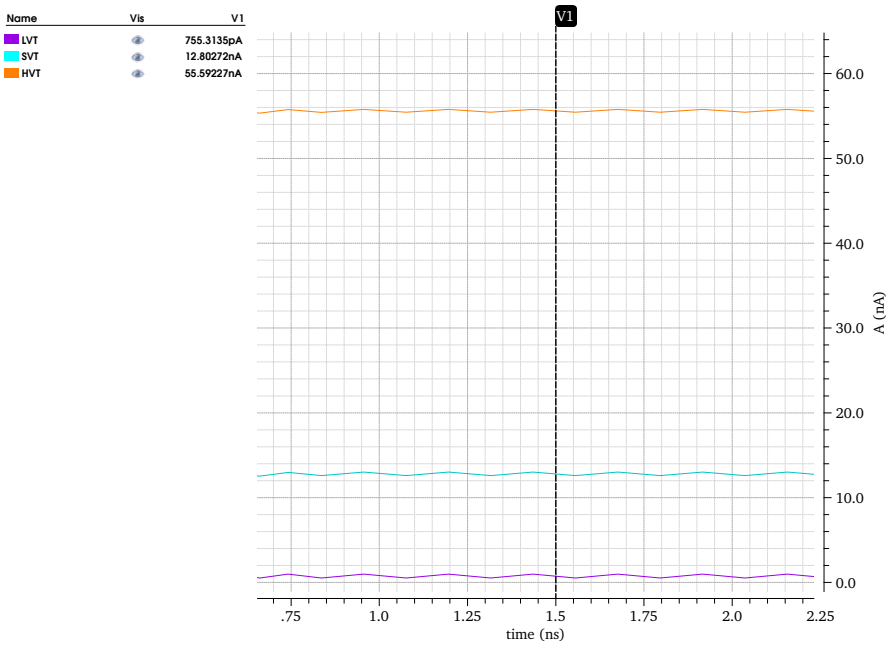
How to reduce I_{leak} - Use wider transitor



The higher the L, the lower the leakage current but the slower the transistors

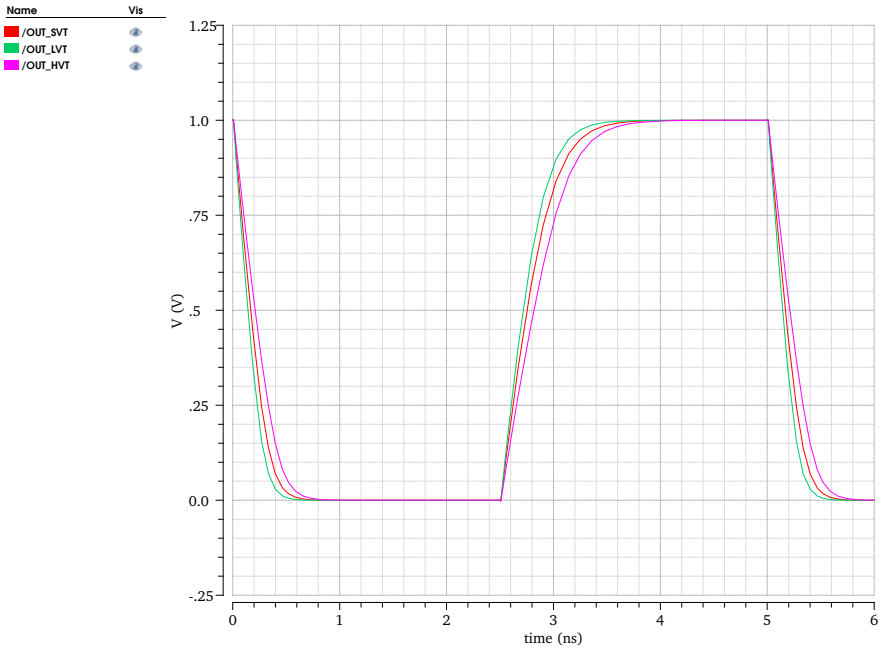
How to reduce I_{leak} - Change Threshold Voltage

Inverter simulation in 65 nm CMOS



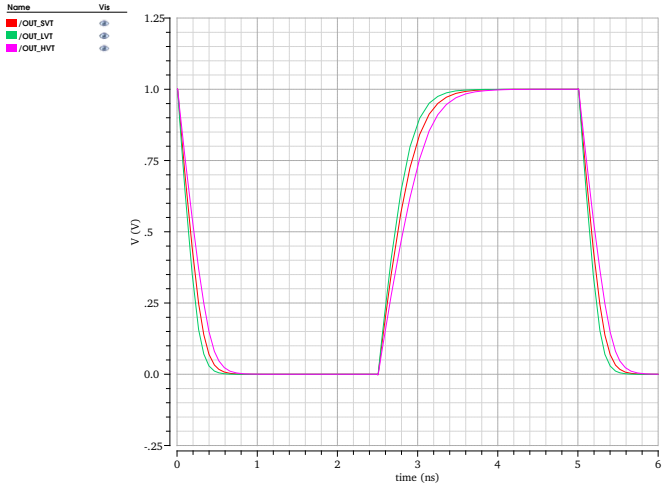
How to reduce I_{leak} - Change Threshold Voltage

Inverter simulation in 65 nm CMOS



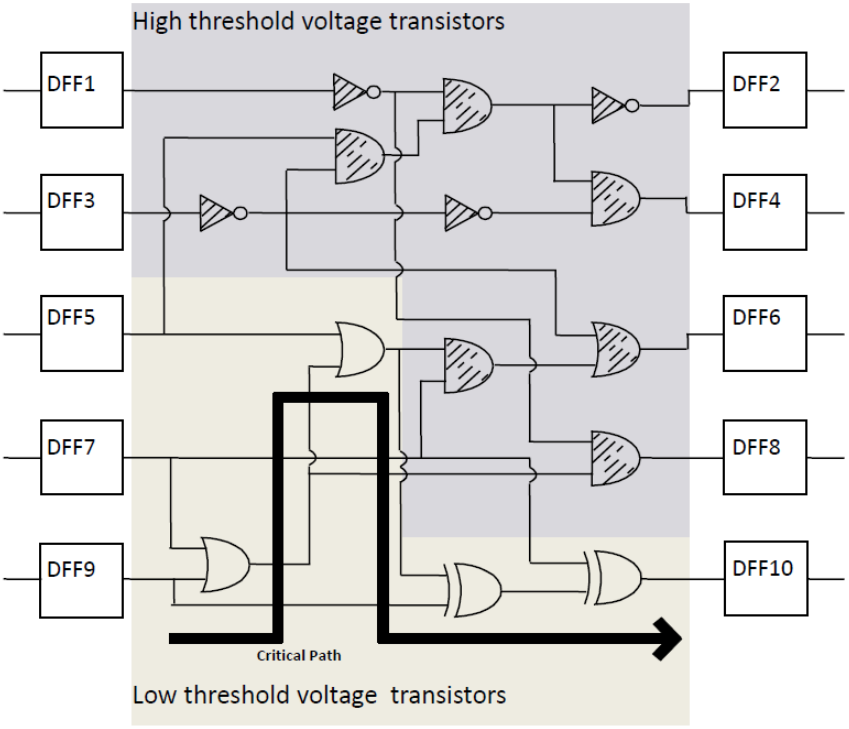
How to reduce I_{leak} - Change Threshold Voltage

Inverter simulation in 65 nm CMOS



The higher the *Vth*, the lower the leakage current but the slower the transistors

How to reduce I_{leak} - Multiple Threshold



Dark: High *Vth* Light: Low *Vth*

How to reduce I_{leak} - Multiple Threshold

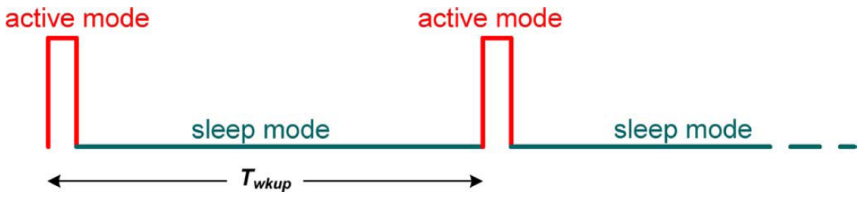
	High-V_{TH} Only	Low-V_{TH} Only	Dual-V_{TH}
Total Slack	–53 ps	0 ps	0 ps
Dynamic Power	3.2 mW	3.3 mW	3.2 mW
Static Power	914 nW	3873 nW	1519 nW

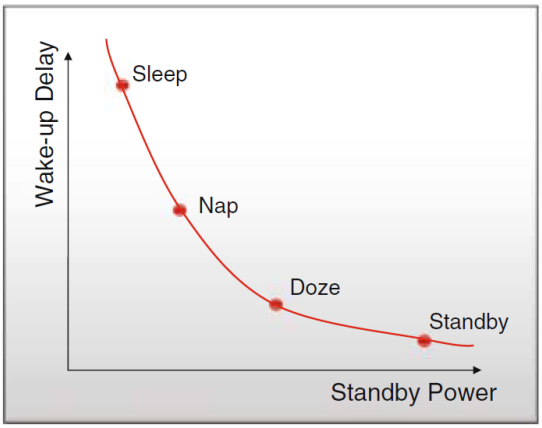
All designs synthesized automatically using Synopsys Flows

[Courtesy: Synopsys, Toshiba, 2004]

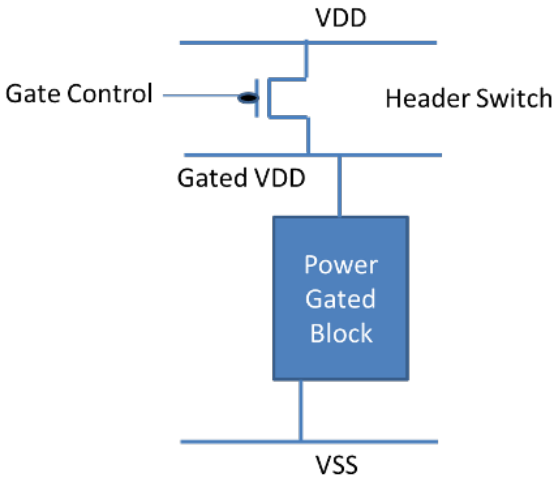
Fais dodo?

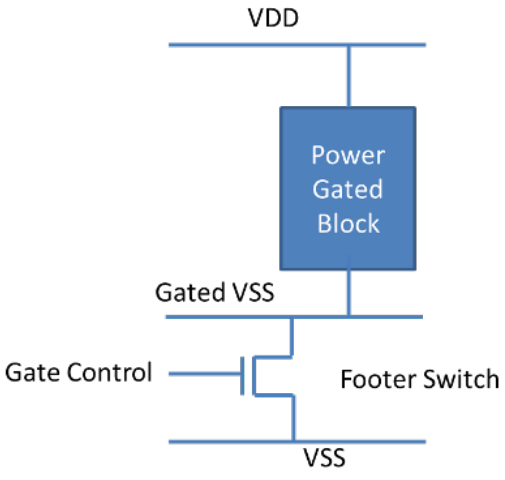
For low speed applications such as sensors, the system can be sent to sleep (Nap/Doze) to minimize leakage:





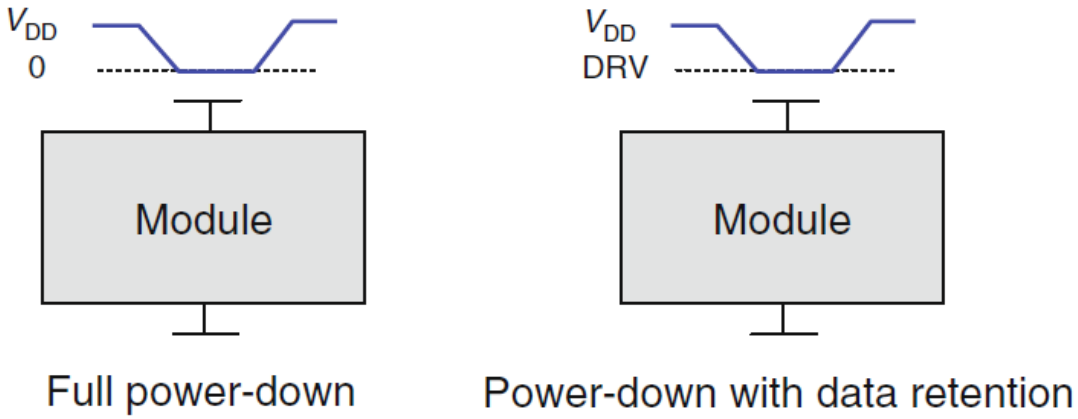
How to reduce I_{leak} - Power gating





- Power gating can be done either on VDD or the ground
- The switch is implemented using a high V_{th} transistor with a high length in order to reduce the leakage

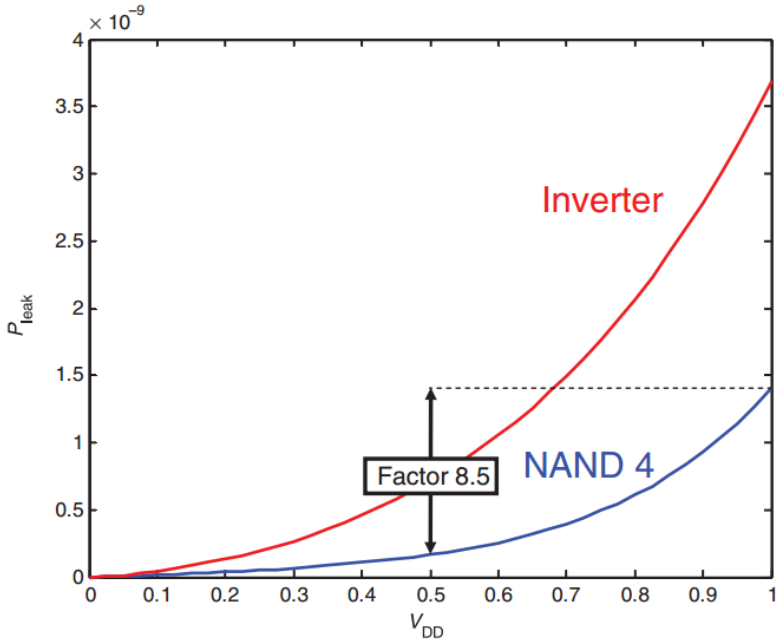
How to reduce I_{leak} - Supply voltage Ramping



- Supply voltage ramping is the most efficient approach to reduce leakage but it requires a controllable voltage regulator
- The supply voltage can be reduced down to "0" if no data retention is needed.

How to reduce I_{leak} - Supply voltage Ramping

Simulation done with CMOS 90 nm



Outline

Conclusion

- Minimizing power consumption is very important in embedded systems in order to
  - Save energy and ecological reasons
  - Reduce cost of changing or charging batteries difficult to access (Human body, satellites, agriculture, ...)
  - Avoiding heating the system
- Dynamic Power consumption can be minimized by:
  - $\searrow$ Capacitor: Smaller technologies, better layout, slower speed
  - $\searrow$ Activity: Clock Gating, glitches suppression, encoding, ...
  - $\searrow$ Vdd (dynamic): Pipelining and parallelism, ...
- Static or leakage power consumption can be minimized by:
  - $\searrow$ I_{leak}: Multiple V_{th}, use lower VDD, ...
  - $\searrow$ VDD (static): Power gating, Supply voltage ramping, ...

References

- A. P. Chandrakasan *et al.*, "Low-power cmos digital design," *IEEE JSSC*, vol. 27, no. 4, pp. 473–484, Apr 1992.
- J. T. Kao *et al.*, "A 175-mv multiply-accumulate unit using an adaptive supply voltage and body bias architecture," *IEEE Journal of Solid-State Circuits*, vol. 37, no. 11, pp. 1545–1554, Nov 2002.
- K. Roy *et al.*, "Leakage current mechanisms and leakage reduction techniques in deep-submicrometer CMOS circuits," *Proceedings of the IEEE*, vol. 91, no. 2, pp. 305–327, Feb 2003.
- J. Rabaey, *Low Power Design Essentials*. Springer, 2009.
- P. Ranjan *et al.*, *Power-efficient System Design*. Boston, MA: Springer US, 2010, ch. Basic Low Power Digital Design, pp. 11–39.
- M. Slimani, "Low energy design of digital circuits," Ph.D. dissertation, Telecom Paristech, 2012.
- M. Alioto, "Ultra-Low Power VLSI Circuit Design Demystified and Explained: A Tutorial," *IEEE TCAS I*, vol. 59, no. 1, pp. 3–29, Jan 2012.
- T. Austin, "On the rules of low power design," 2013. [Online]. Available: https://www.youtube.com/watch?v=jpkiFRRX1g8
- A. Pal, "Low power vlsi circuits and systems," 2015. [Online]. Available: https://www.youtube.com/watch?v=bKhs2Wv3Kb4

Methodology - Design example: Decimation filter

- Application level :
  - Is there anyway to relax my specifications?
  - Should my system be ON all the time?
  - ...
- Architecture and system level :
  - What is the best architecture for my system?
  - Should i do the decimation in 1 step, 2 steps ...?
  - How many coefficients for each of my filters?
  - On how many bits should i code my coefficients?
  - Should i use any type of encoding?
  - ...
- Gate and circuit level :
  - Technology, gate sizing
  - Pipelining, parallelism
  - Multiple V_{th}, Lower VDD, multiple VDD
  - Design custom gates to improve power
  - ...