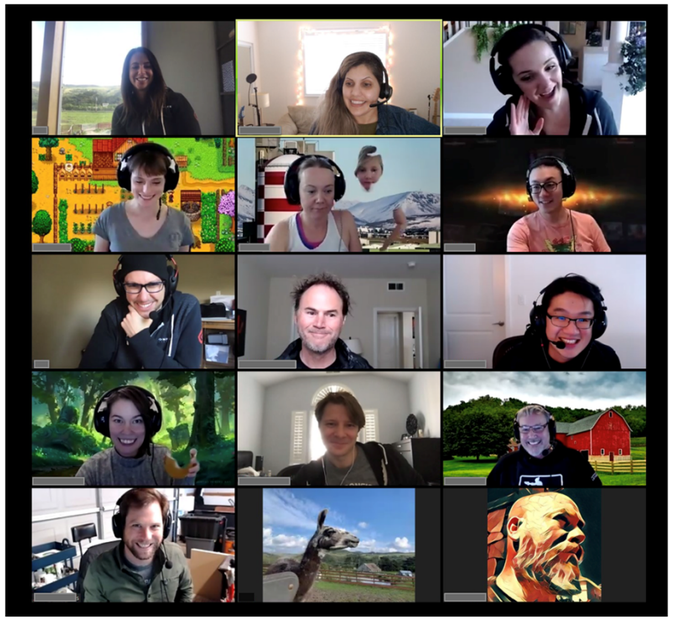
EE241B : Advanced Digital Circuits

Lecture 22 – Reducing Leakage

Borivoje Nikolić



Sweetfarm.org/goat-2-meeting:
Invite a goat or a llama to a zoom meeting

https://www.sweetfarm.org/goat-2-meeting

Announcements

- Assignment 4 due next Friday.

- Reading
  - Rabaey, LPDE, Chapter 8

Outline

- Module 5
  - Clock gating
  - Leakage reduction during design time and runtime

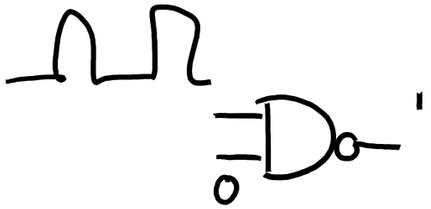
5.G Reducing Switching Activity Through Logic Design

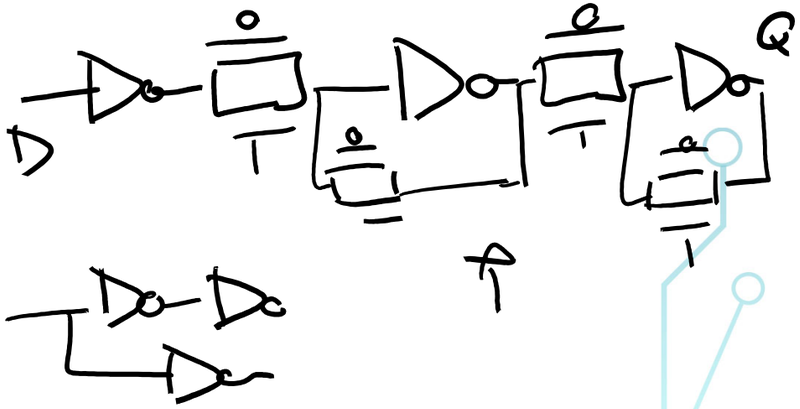
Power /Energy Optimization Space

	Constant Throughput/Latency		Variable Throughput/Latency
Energy	**Design Time**	**Sleep Mode**	**Run Time**
Active	Logic design ✓ Scaled V_{DD} Trans. sizing Multi-V_{DD}	Clock gating ✓	DFS, DVS
Leakage	Stack effects Trans sizing Scaling V_{DD} + Multi-V_{Th}	Sleep T's Multi-V_{DD} Variable V_{Th} + Input control	+ Variable V_{Th}

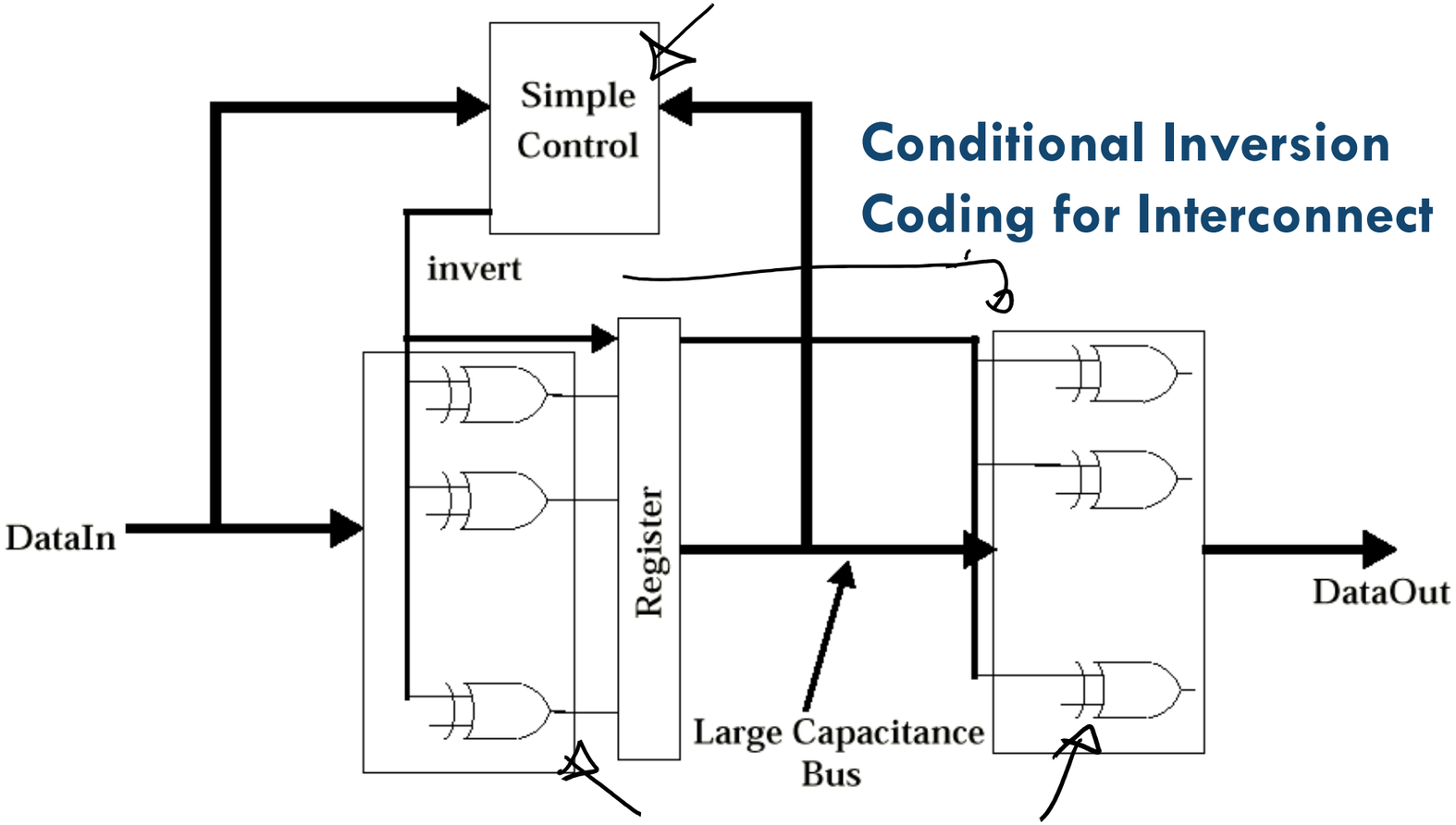
Basic Idea

- $E \sim \alpha CV^2$
- Reduce switching activity, α, through logic and architectural transformations
- Many options
  - Switching activity lower with deeper logic
  - Pipelining has significant effect
  - Reduce the number of clocked devices in a flip-flop
    - e.g. group generation of clk_b
  - A few logic ideas follow





Circuit-Level Activity Encoding



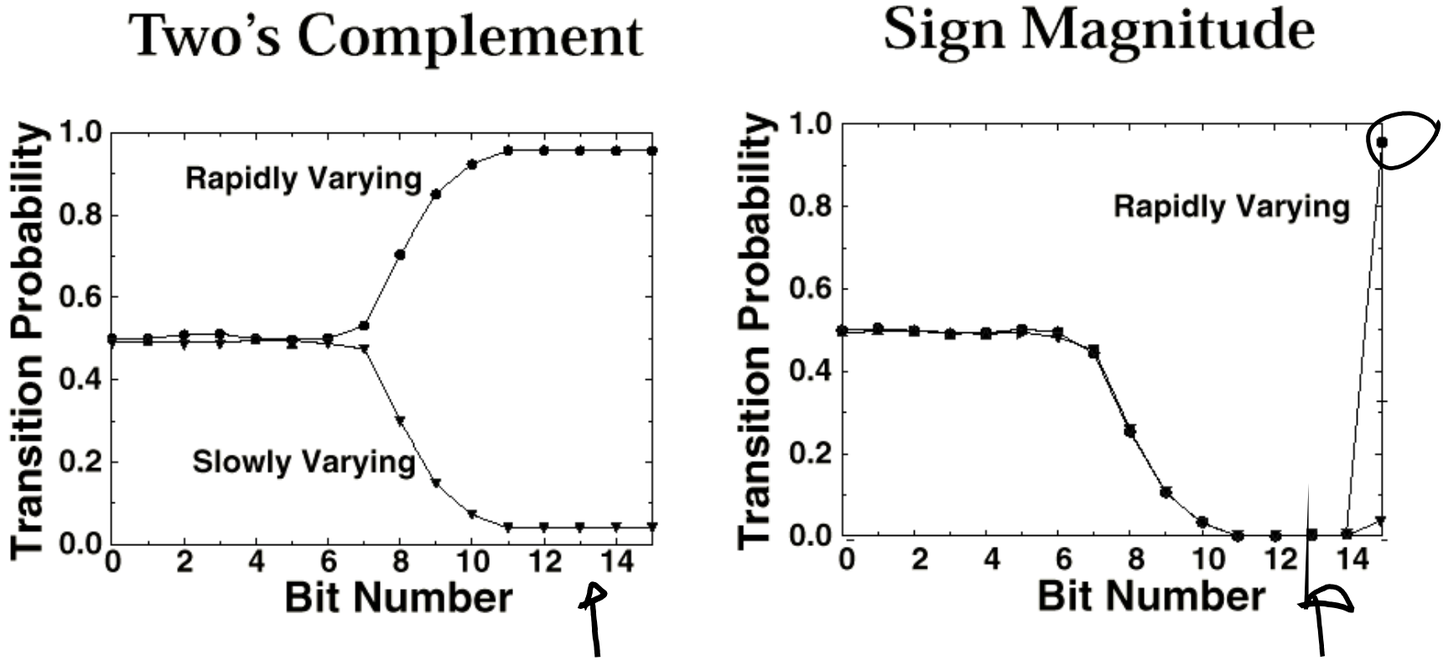
Conditional Inversion Coding for Interconnect

from [Stan94]

(*1994 International Workshop on Low-power Design*)

Number Representation

- Input signals are noise most of the time



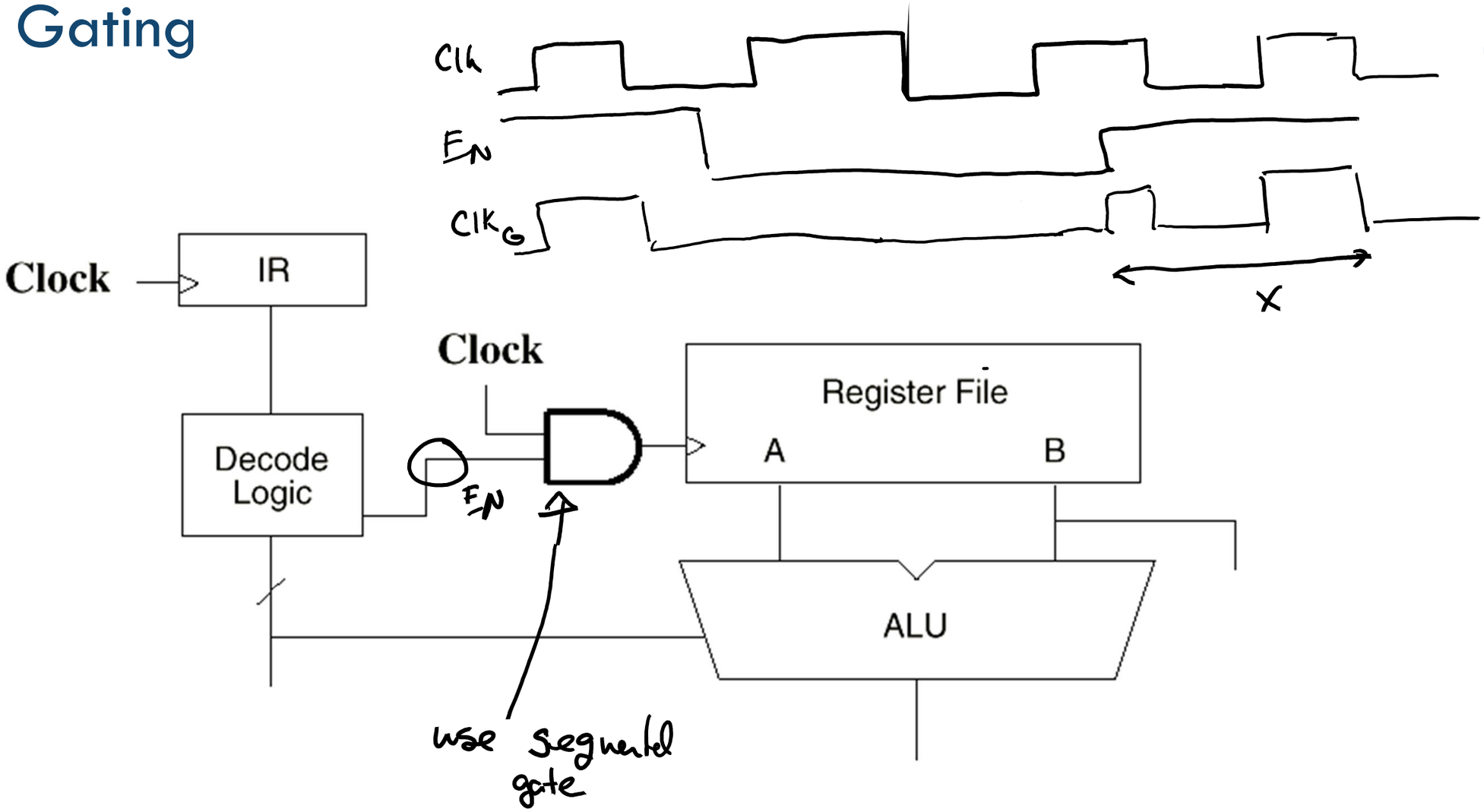
- Sign-extension activity significantly reduced using sign-magnitude representation

5.H Clock Gating

Power /Energy Optimization Space

	Constant Throughput/Latency		Variable Throughput/Latency
Energy	**Design Time**	**Sleep Mode**	**Run Time**
Active	Logic design Scaled V_{DD} Trans. sizing Multi-V_{DD}	Clock gating	DFS, DVS
Leakage	Stack effects Trans sizing Scaling V_{DD} + Multi-V_{Th}	Sleep T's Multi-V_{DD} Variable V_{Th} + Input control	DVS, Variable V_{Th}

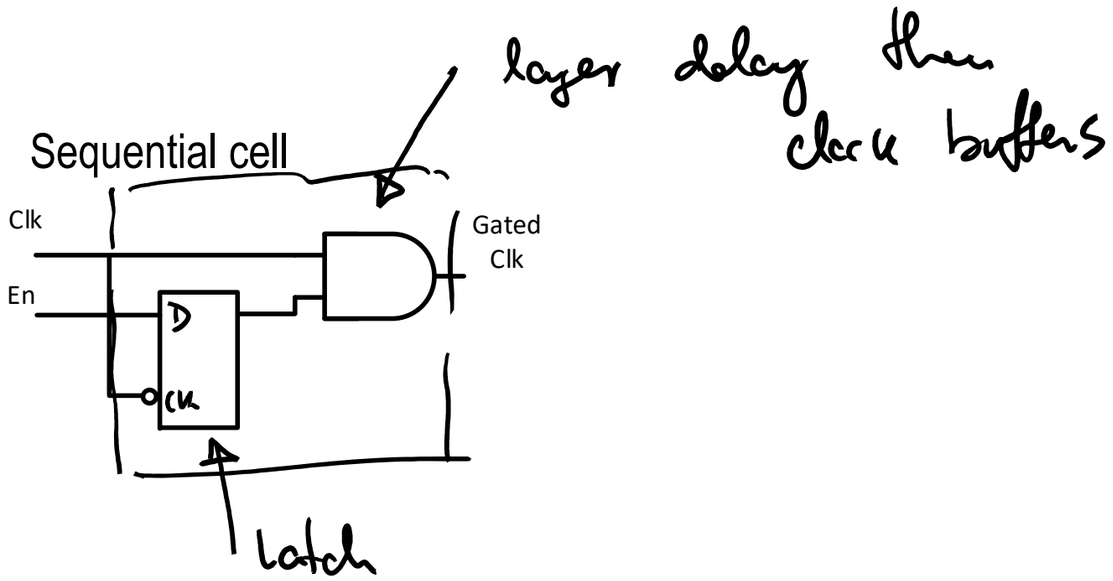
Clock Gating



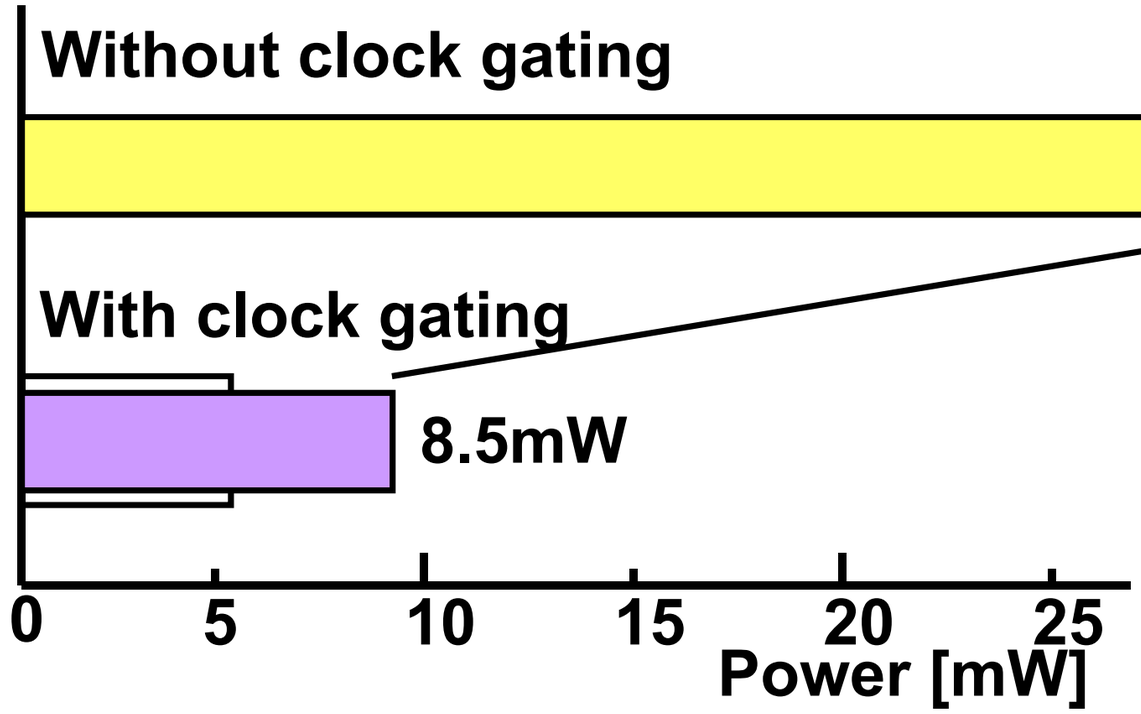
Requires a bit more complex gate ...
Well handled in today's EDA tools

Clock Gating

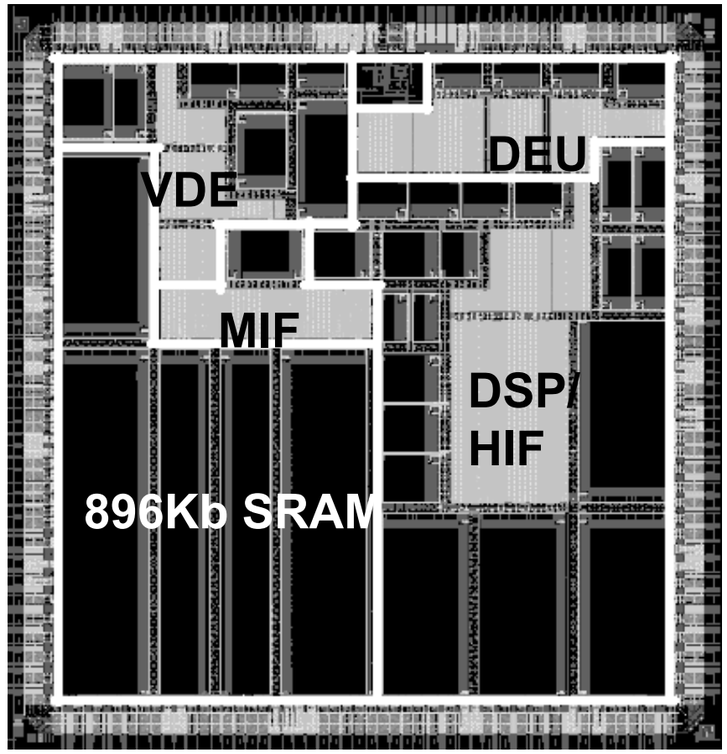
- Enabling clock needs to be synchronized



Clock Gating Efficiently Reduces Power



MPEG4 decoder



90% of F/F's were clock-gated.

70% power reduction by clock-gating alone.

Courtesy M. Ohashi, Matsushita, ISSCC 2002

Clock Gating

EN signals

- ARM Cortex-A9 Technical Reference Manual:

• core level gating signals
• System-level gating
↳ system manager

Dynamic high level clock gating activity

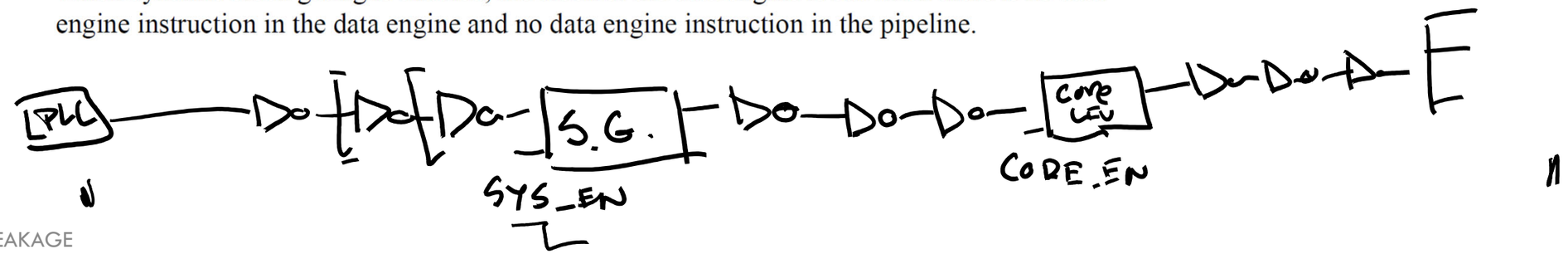
When dynamic high level clock gating is enabled the clock of the integer core is cut in the following cases:

- the integer core is empty and there is an instruction miss causing a linefill
- the integer core is empty and there is an instruction TLB miss
- the integer core is full and there is a data miss causing a linefill
- the integer core is full and data stores are stalled because the linefill buffers are busy.

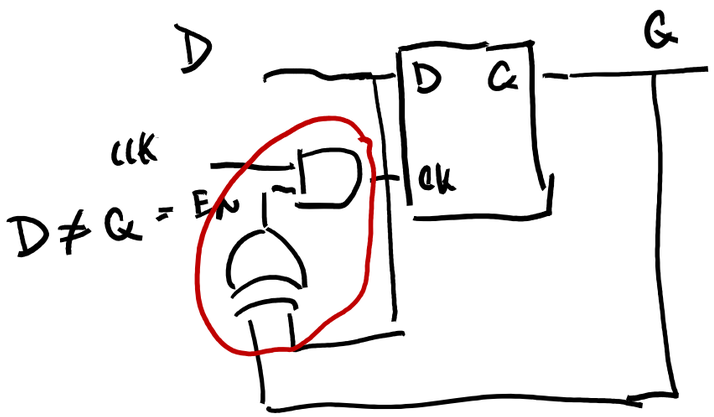
When dynamic clock gating is enabled, the clock of the system control block is cut in the following cases:

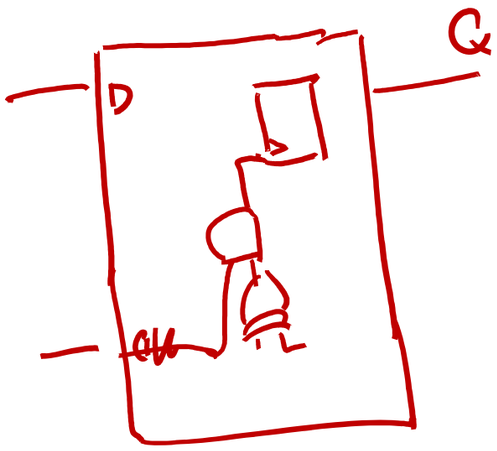
- there are no system control coprocessor instructions being executed
- there are no system control coprocessor instructions present in the pipeline
- performance events are not enabled
- debug is not enabled.

When dynamic clock gating is enabled, the clock of the data engine is cut when there is no data engine instruction in the data engine and no data engine instruction in the pipeline.



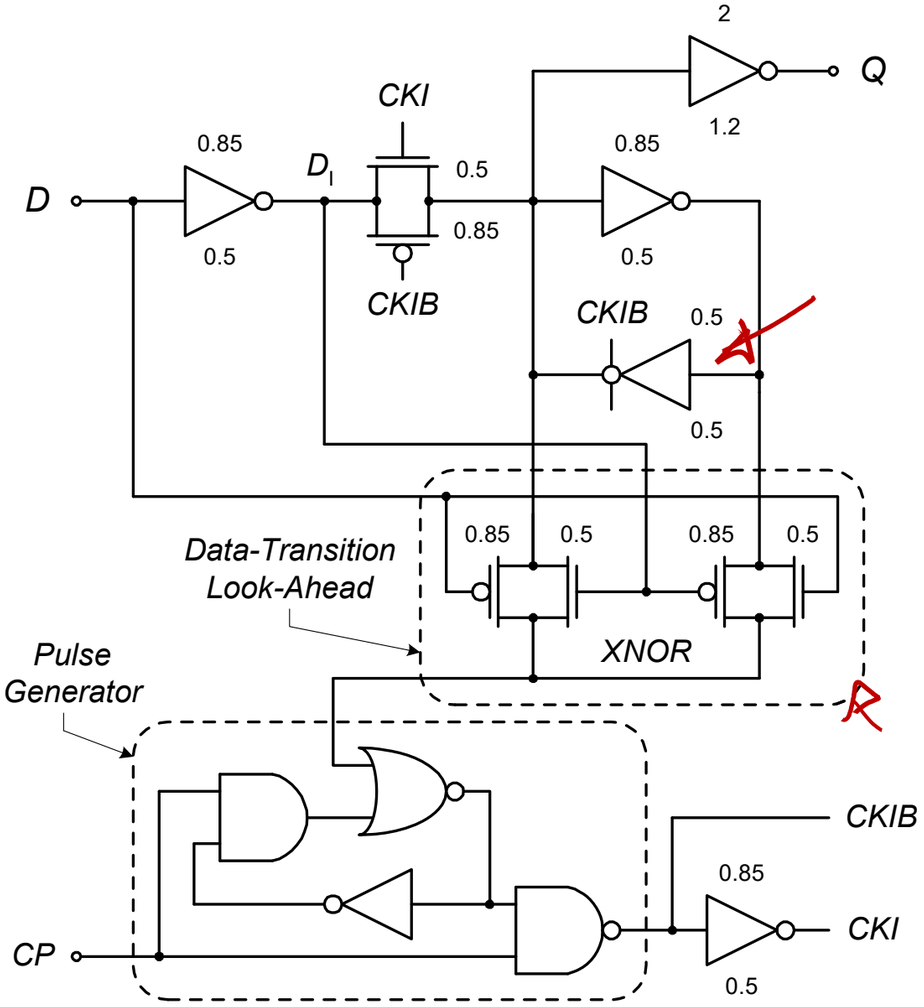
Local Clock Gating

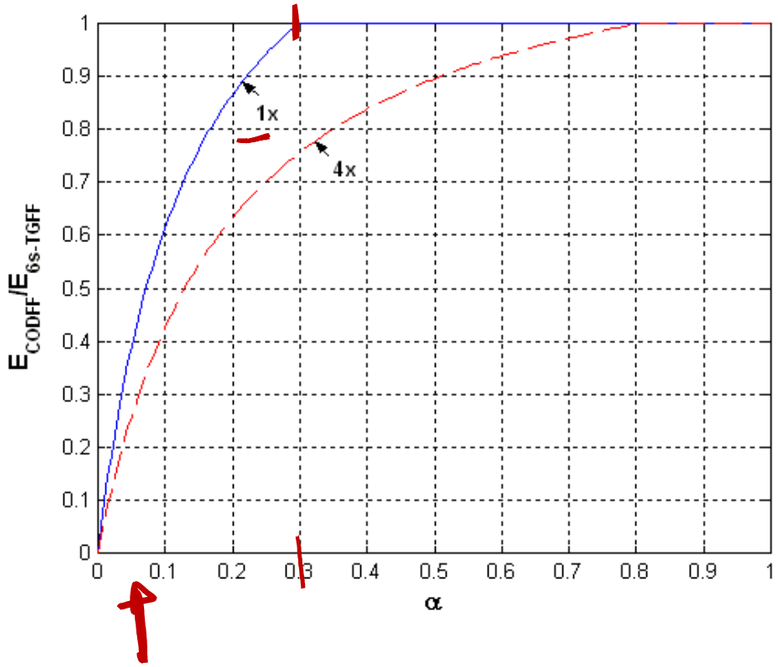




. VCD

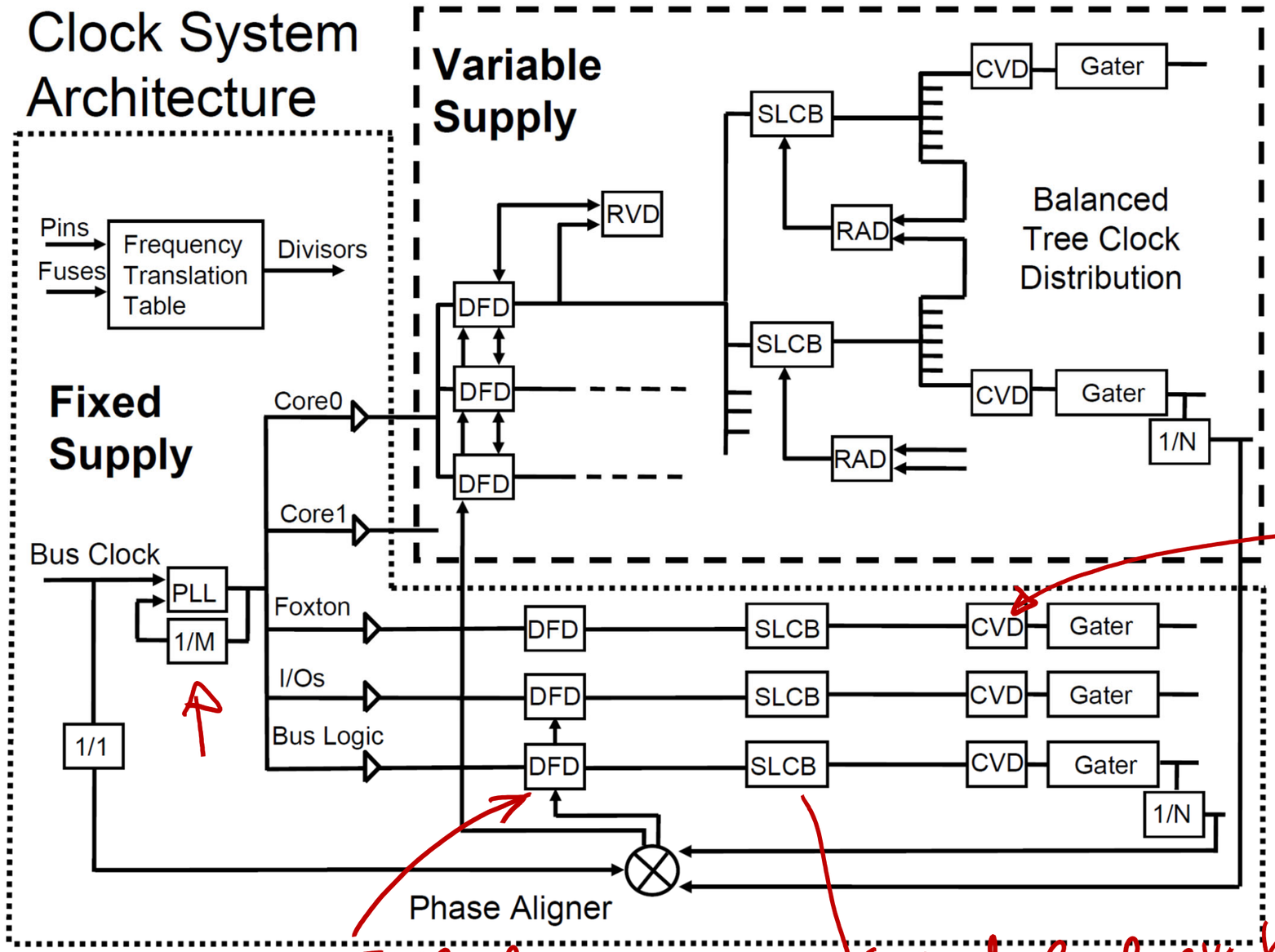
Local Clock Gating





'Clock on demand'
Flip-flop

Complex Designs



Fischer, ISSCC'05

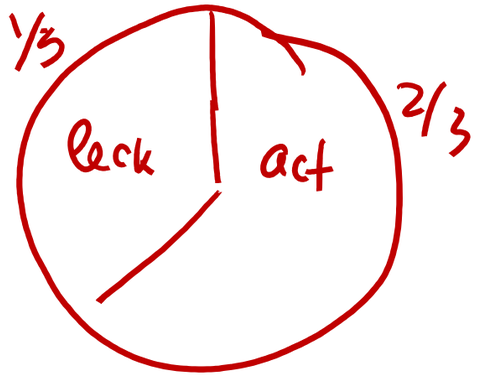
Power /Energy Optimization Space

	Constant Throughput/Latency		Variable Throughput/Latency
Energy	**Design Time**	**Sleep Mode**	**Run Time**
Active	Logic design Scaled V_{DD} Trans. sizing Multi-V_{DD}	Clock gating	DFS, DVS
Leakage	Stack effects Trans sizing Scaling V_{DD} + Multi-V_{Th}	Sleep T's Multi-V_{DD} Variable V_{Th} + Input control	DVS, Variable V_{Th}

Plan For the Rest of the Semester

- 4 more lectures (including today's):
  - Finish low power (2 lectures)
  - Supplies, clocks and their interaction
- Homework 4 due on April 24th
  - Quiz 4 on April 28th
- Final on April 30th
  - 80 minutes, open everything
- Final presentations, May 4
  - Final reports due on May 4

5.1 Lowering Leakage During Design: Multiple Thresholds

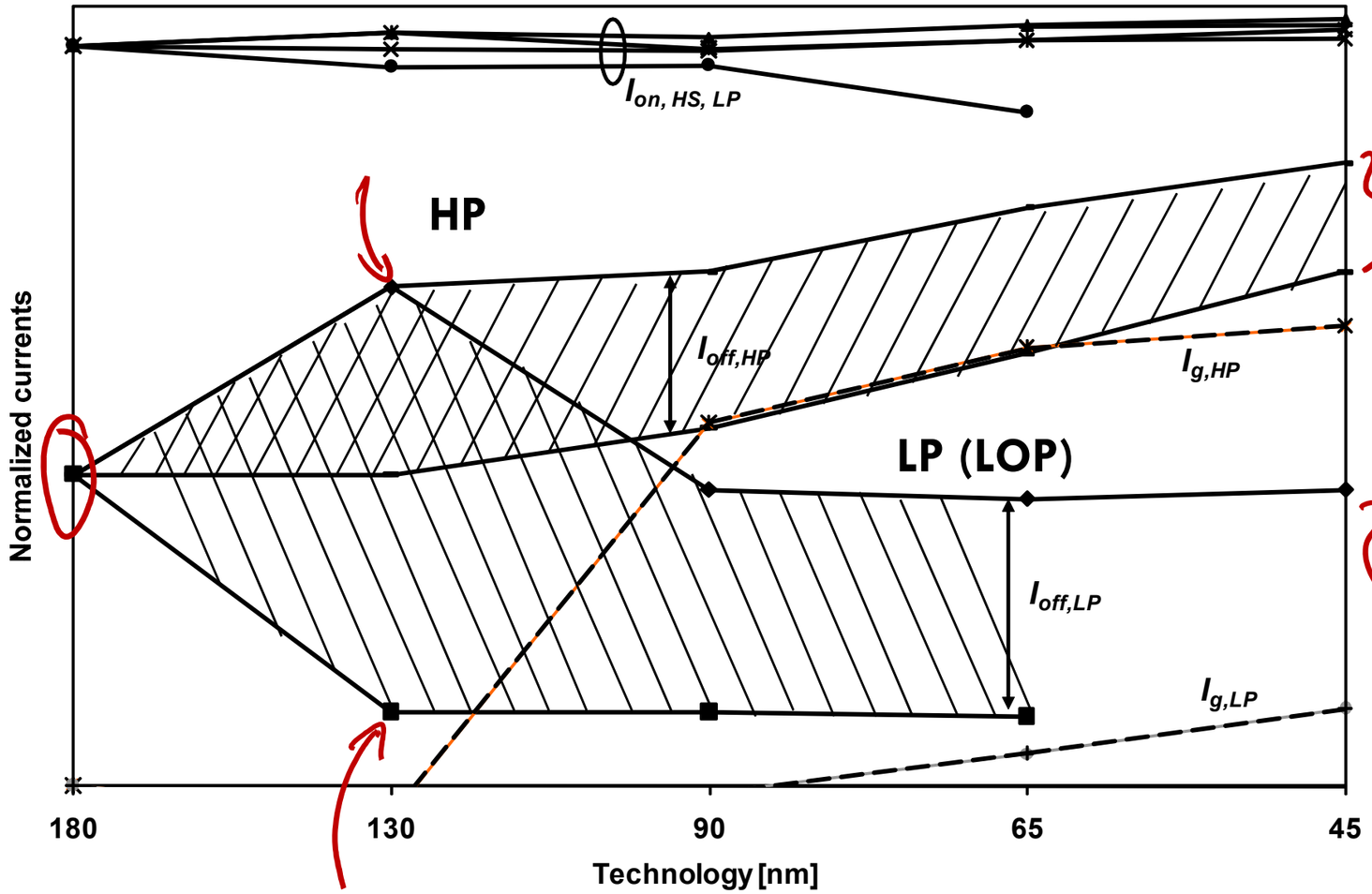


Power /Energy Optimization Space

	Constant Throughput/Latency		Variable Throughput/Latency
Energy	**Design Time**	**Sleep Mode**	**Run Time**
Active	Logic design Scaled V_{DD} Trans. sizing Multi-V_{DD}	Clock gating	DFS, DVS
Leakage	Stack effects Trans sizing Scaling V_{DD} + Multi-V_{Th}	Sleep T's Multi-V_{DD} Variable V_{Th} + Input control	DVS, Variable V_{Th}

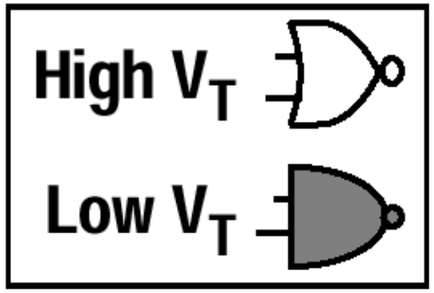
Technology Options

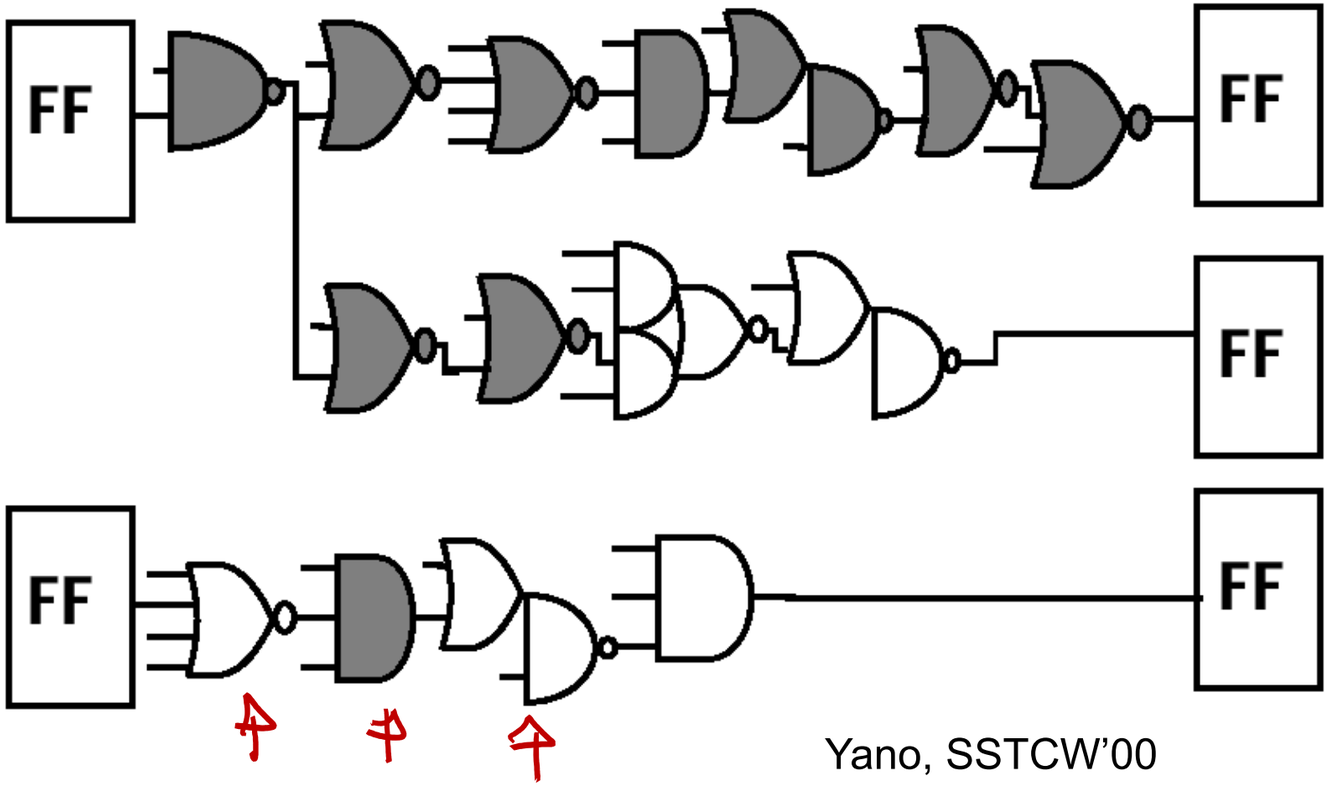
- Multiple thresholds, each spaced 50-100mV apart (5-10x less leakage)

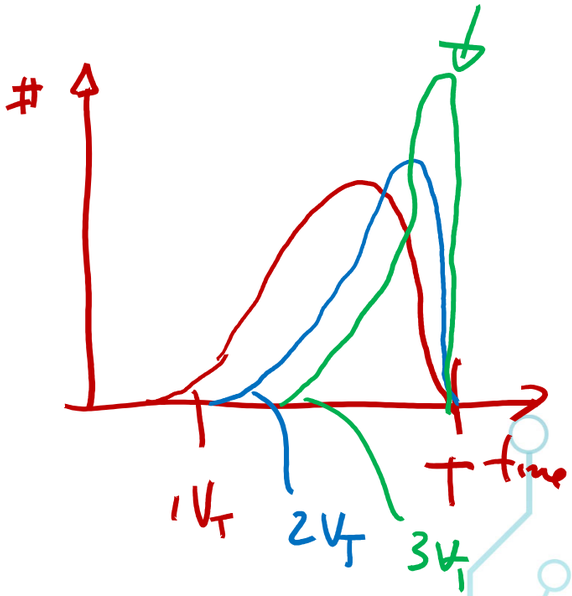


Using Multiple Thresholds

- Cell-by-cell V_T assignment (not block level)
- Allows us to minimize leakage
- Achieves all-low-V performance







Yano, SSTCW'00

Typical Technologies

- 2-3 Thresholds
  - To choose from 4-6 in a node
  - In bulk and finfet, but not in FDSOI (unless doped)
    - 22FDX
- Threshold voltage diff ~5-10x in leakage

$\Delta V_{th} = 70mV \Rightarrow$ 10x I_L

$\frac{\Delta D}{\Delta V_{th}} = 10\%$ (~30%)

$\frac{\Delta P}{\Delta V_{th}} = \frac{\Delta P_{act} + \Delta P_{leak}}{\Delta V_t}$ ($\Delta P_{act} \to 0$, $\Delta P_{leak} \to 0.1 P_{leak,orig.}$)

$I_D \sim K(V_{DD} - V_t)$

$\frac{70}{400} \to 20\%$

700mV → 10%

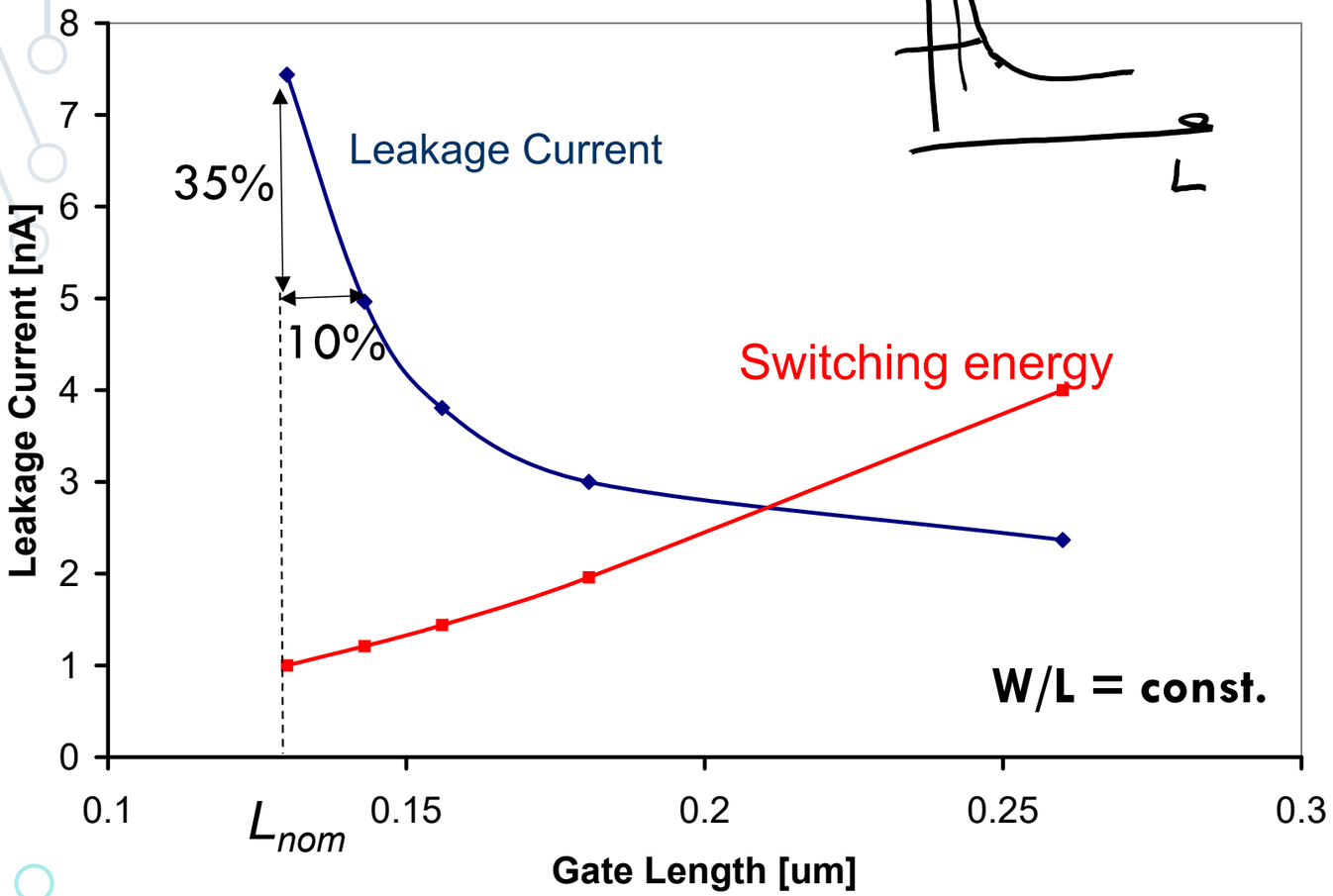
ΔV_{th} 70mV

5.1 Lowering Leakage During Design: Longer Channels

Power /Energy Optimization Space

	Constant Throughput/Latency		Variable Throughput/Latency
Energy	**Design Time**	**Sleep Mode**	**Run Time**
Active	Logic design Scaled V_{DD} Trans. sizing Multi-V_{DD}	Clock gating	DFS, DVS
Leakage	Stack effects Trans sizing Scaling V_{DD} + Multi-V_{Th}	Sleep T's Multi-V_{DD} Variable V_{Th} + Input control	DVS, Variable V_{Th}

Longer Channels



Multiple L_S ≡ "Poly bias"

- 10% longer gates reduce leakage by 35% (in 130nm)
- Increases switching energy by 21% with W/L = const.

$\frac{\Delta D}{\Delta L}$ | W/L = const | $\Delta L = 10\%$

$\frac{\Delta P}{\Delta D} = \frac{\Delta P_{act} + P_{2nm}}{\Delta D}$ | W/L = const | $\Delta D = 10\%$

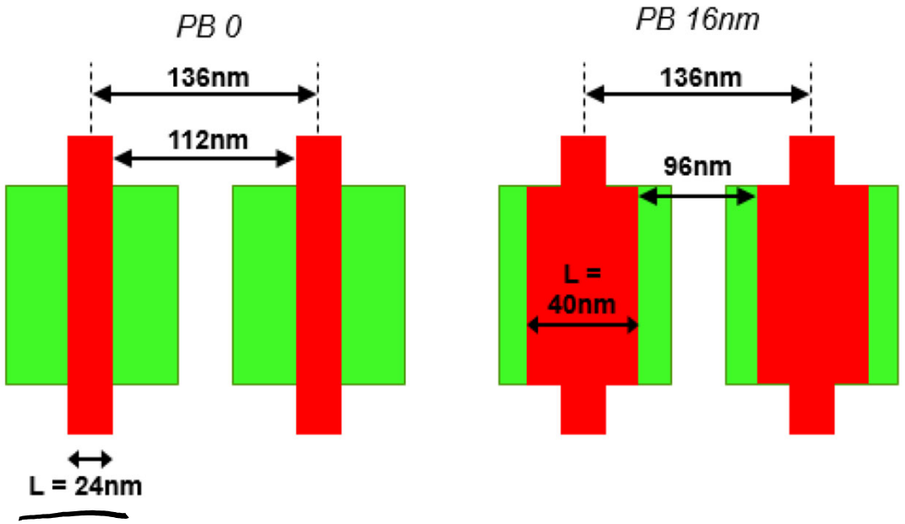
- Attractive when don't have to increase W (memory)
- Doubling *L* reduces leakage by 3x (in 0.13um)
- Much stronger effect in 28nm!
- Effect improves with shorter channel devices

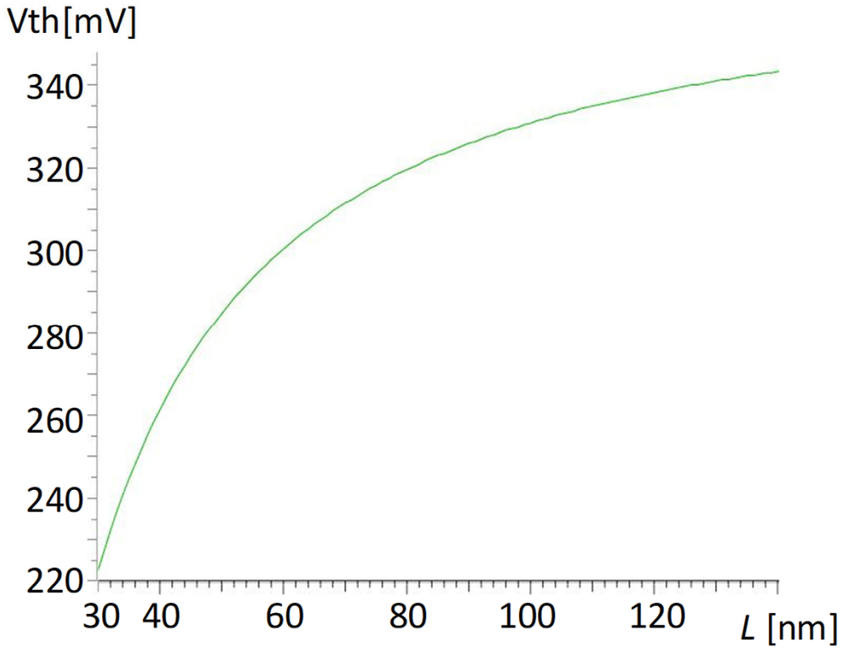
HVT, SVT, LVT, ULVT

PB: PB0, PB2, PB4, ...

Poly Bias

- 28FDSOI example





Longer Channels