inst.eecs.berkeley.edu/~eecs251b

EECS251B : Advanced Digital Circuits and Systems

Lecture 22 – Power-saving techniques: Leakage optimization (multi-threshold, sleep)

Borivoje Nikolić, Vladimir Stojanović, Sophia Shao

Future computer chips may be made out of honey

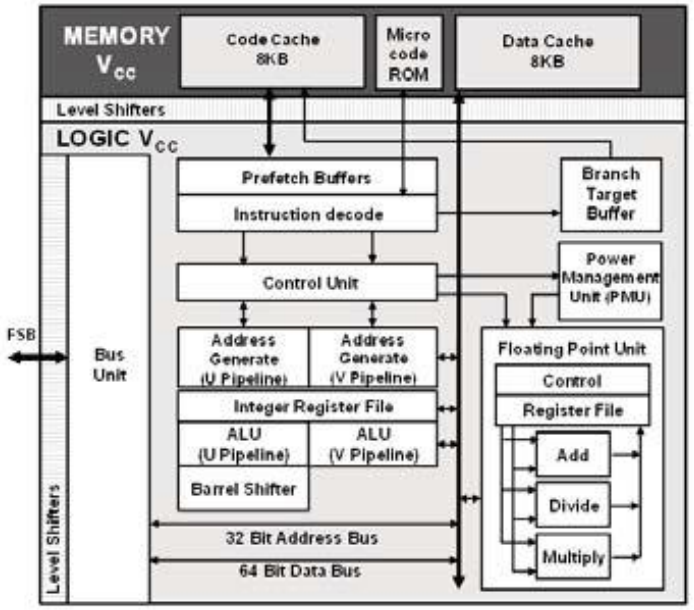
Digital Trends on WSU research

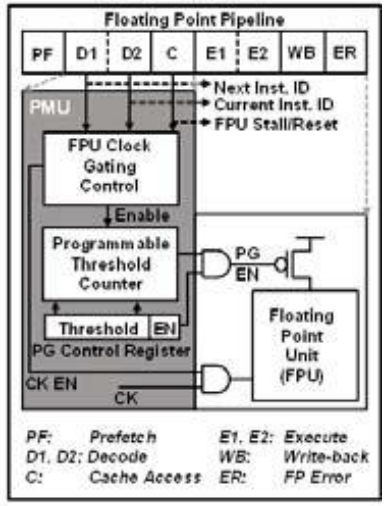
The wearables market, including smartwatches, smartbands, and smart glasses, together will generate more than $350 billion in cumulative revenues over the next five years

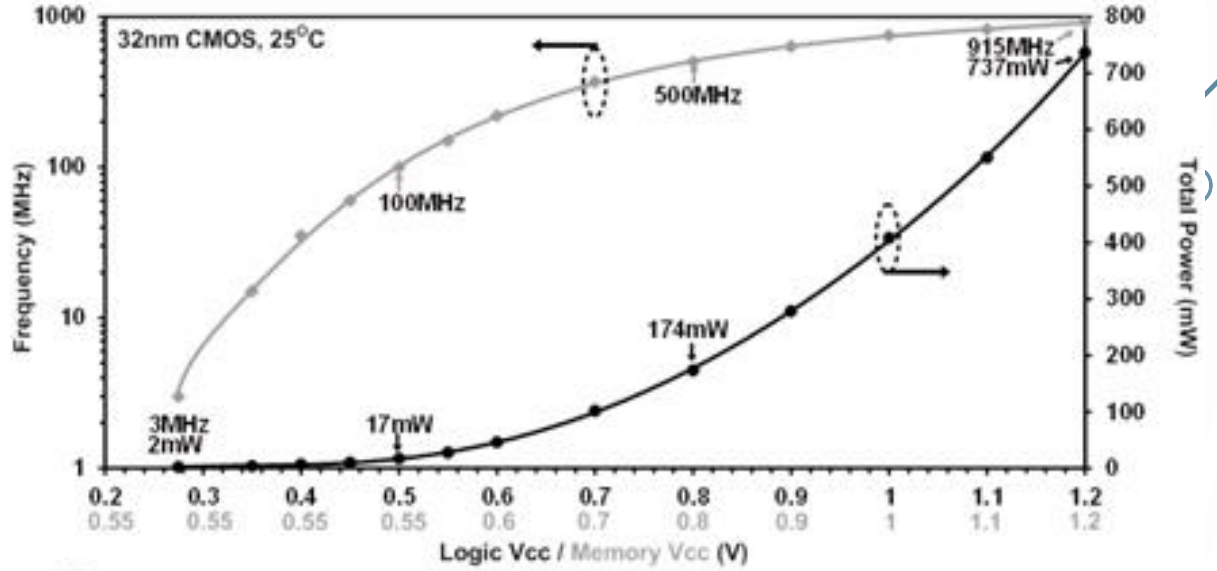
Announcements

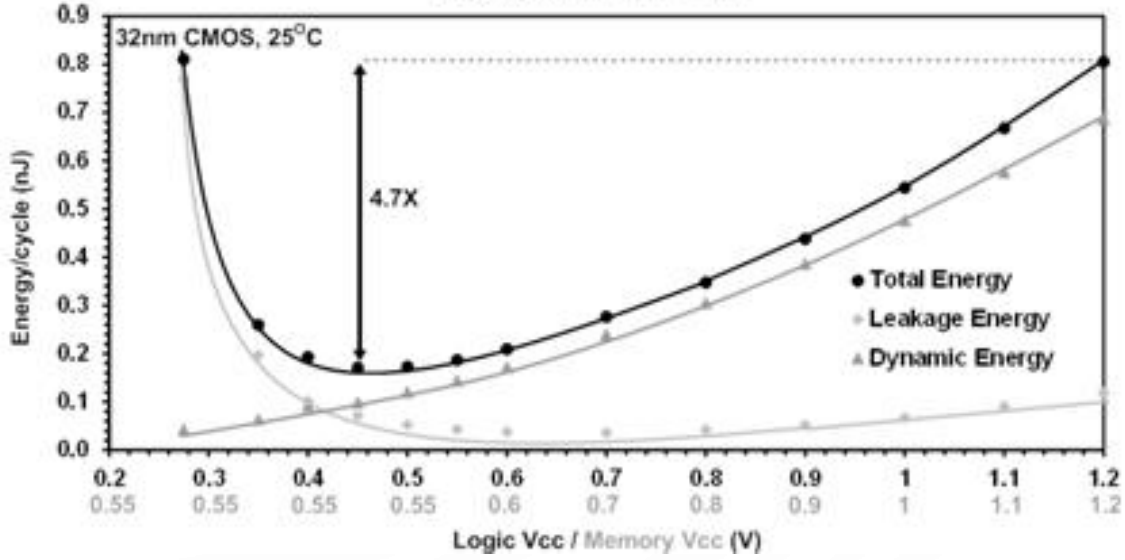
- Assignment 3 due Wed 4/13
- Assignment 4 out Thu 4/14
- Quiz 3 on 4/14 (SRAM)

Why optimize leakage?





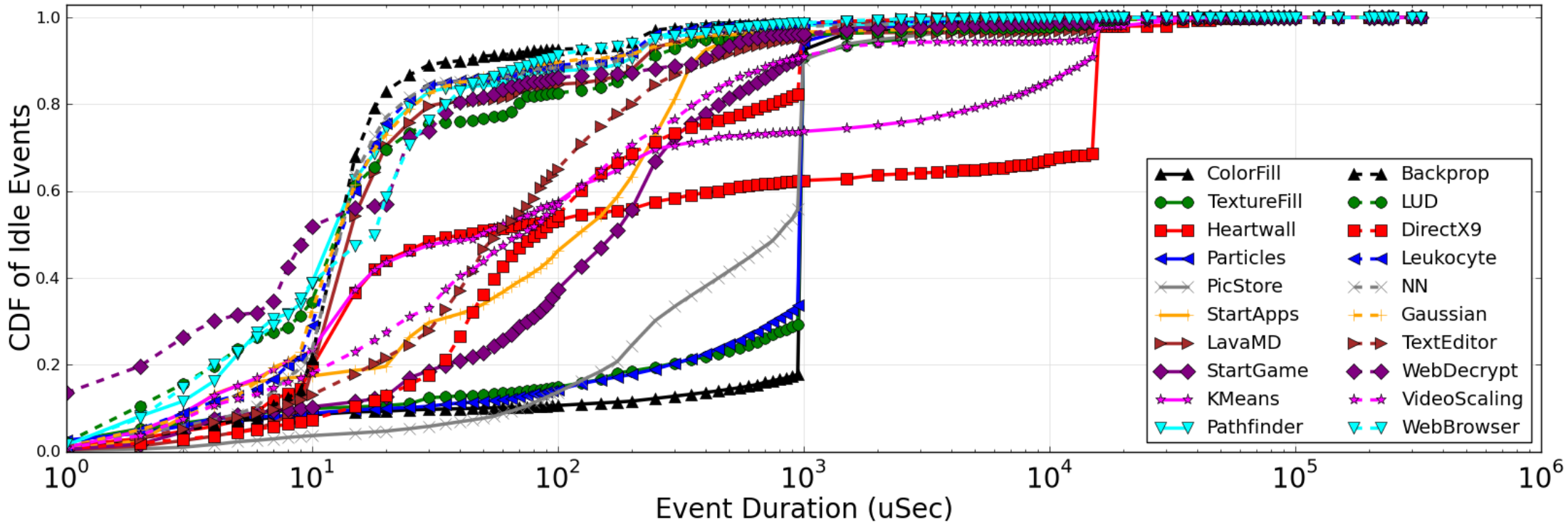




- Processors not active all the time
- Balancing the active and leakage energy in wide-range DVFS operation

Jain ISSCC'12

Application activity impact on power control strategy



Arora HPCA15

- The power control strategy depends on the statistics of idle events for typical workloads

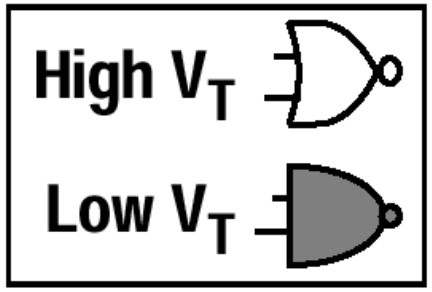
Lowering Leakage During Design: Multiple Thresholds

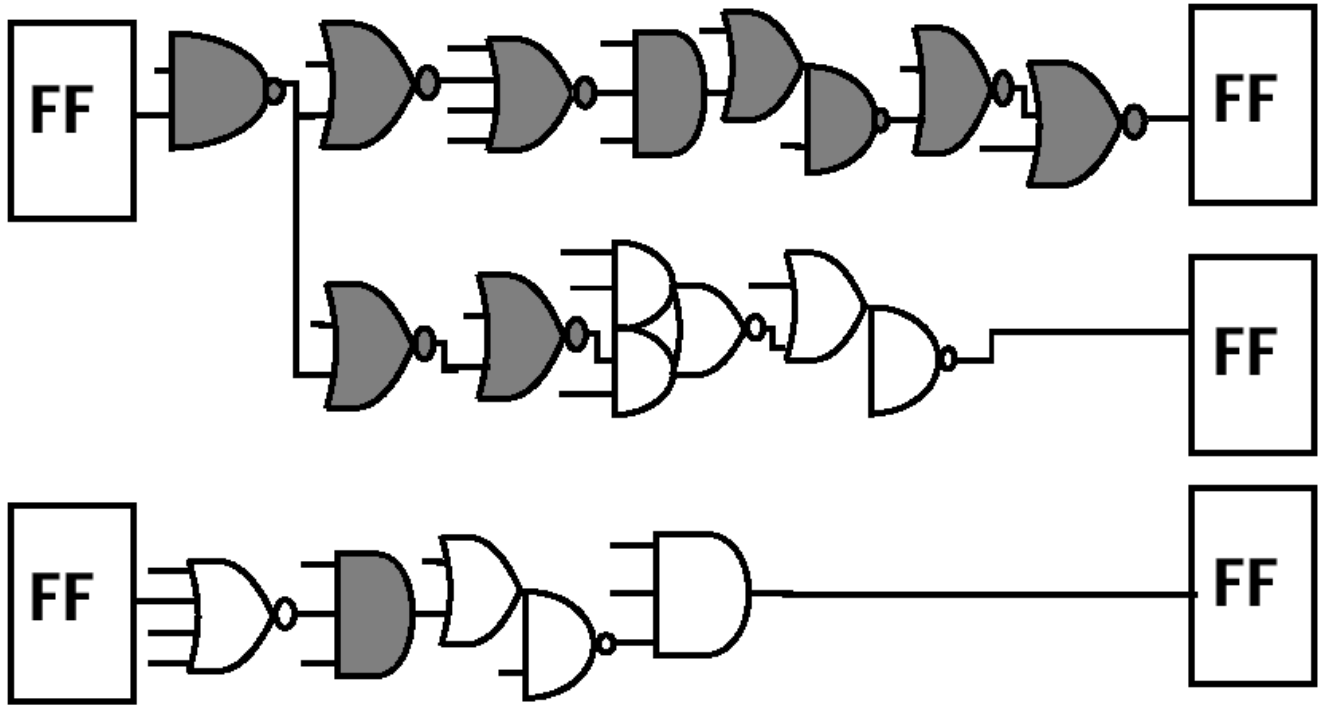
Power /Energy Optimization Space

	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}

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

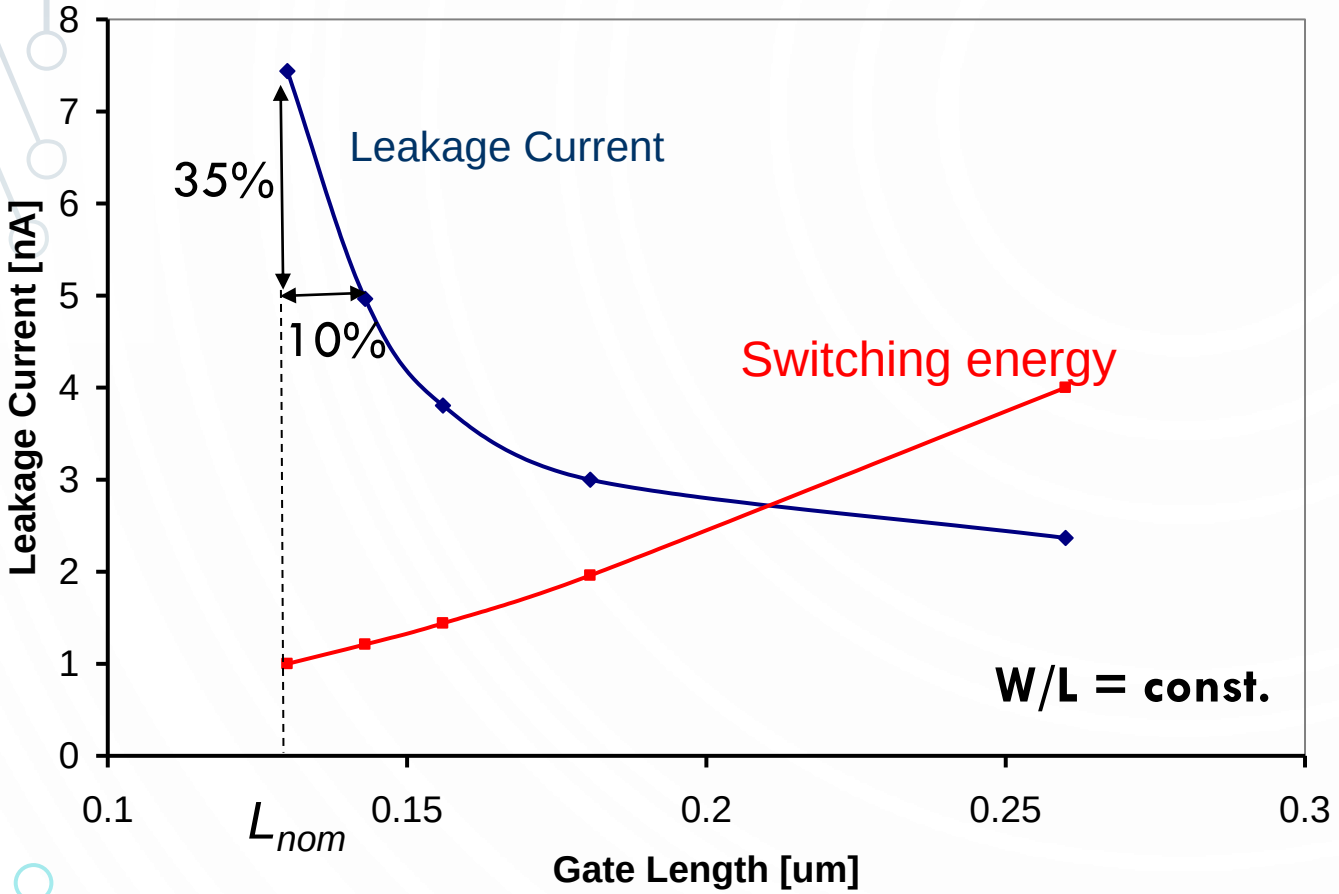
- 2-3 Thresholds
  - To choose from 4-6 in a node
  - In bulk and finfet, but not in FDSOI (unless doped)
- Threshold voltage diff ~5-10x in leakage

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

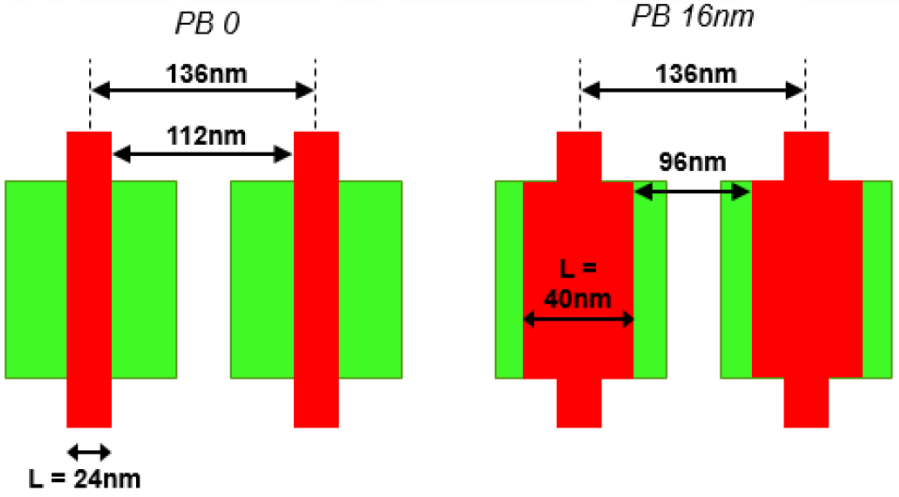


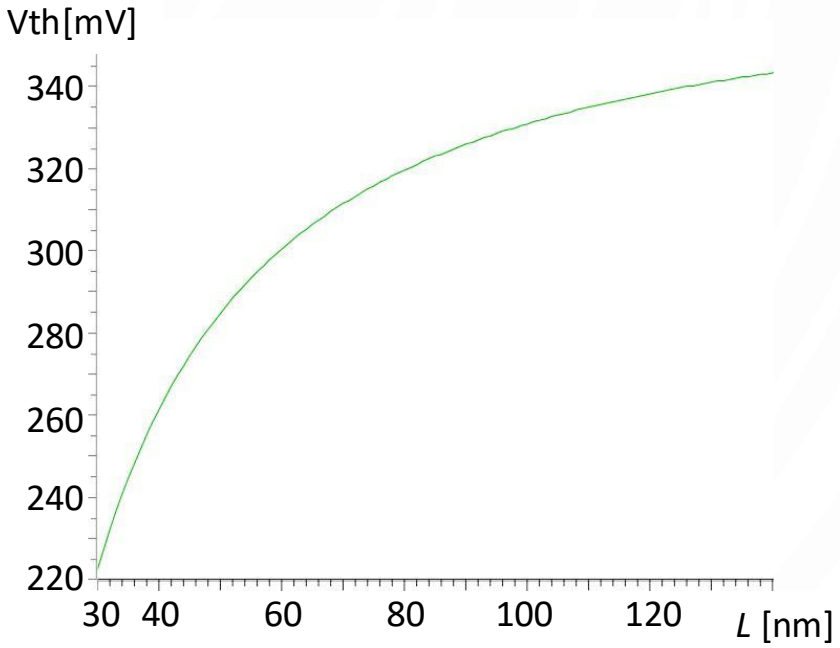
- 10% longer gates reduce leakage by 35% (in 130nm)
- Increases switching energy by 21% with W/L = const.

- 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

Poly Bias

- 28FDSOI example



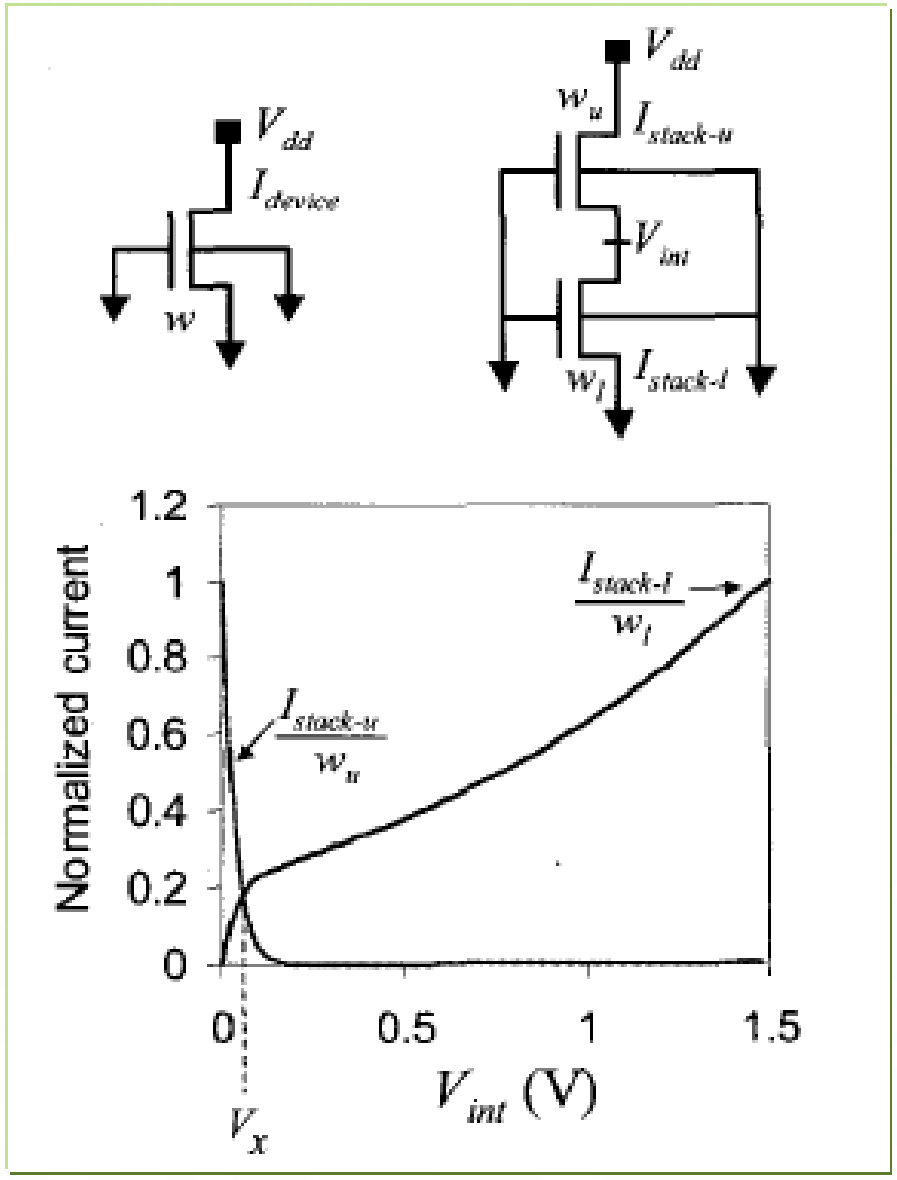


Lowering Leakage During Design: Transistor Stacking

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}

Stack Effect

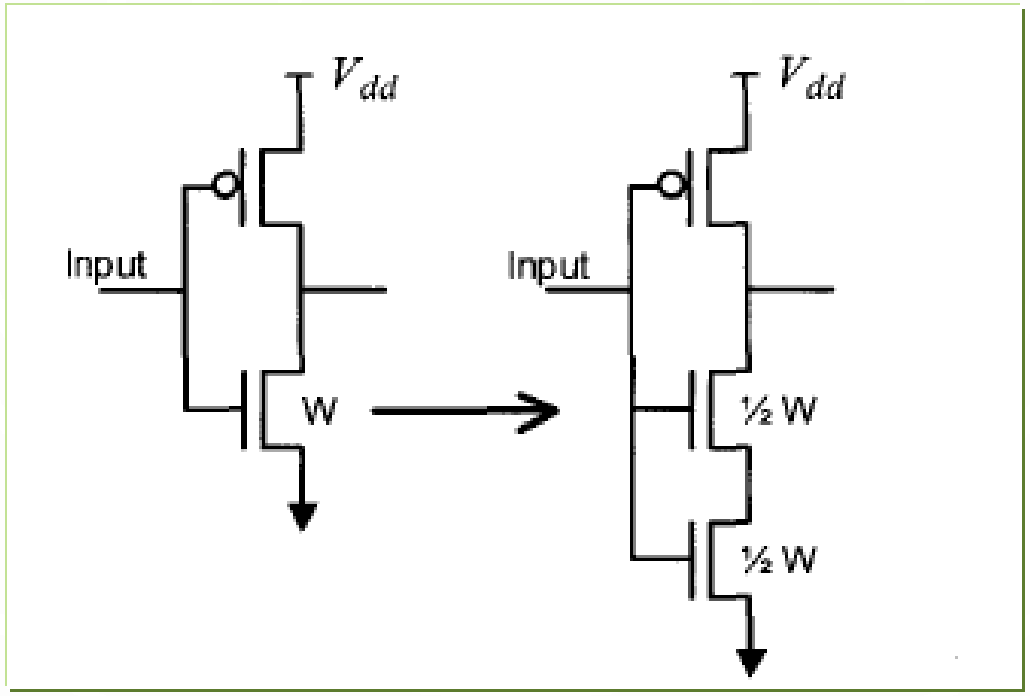


Narendra, ISLPED'01

Reduction (in 0.13μ):

	High V_t	*Low V_t*
2 NMOS	10.7X	9.96X
3 NMOS	21.1X	18.8X
4 NMOS	31.5X	26.7X
2 PMOS	8.6X	7.9X
3 PMOS	16.1X	13.7X
4 PMOS	23.1X	18.7X

Stack Forcing – Gate replacement



Tradeoffs:

- W/2 – 1/3 of drive current, same loading
- 1.5W – 3x loading, same drive current

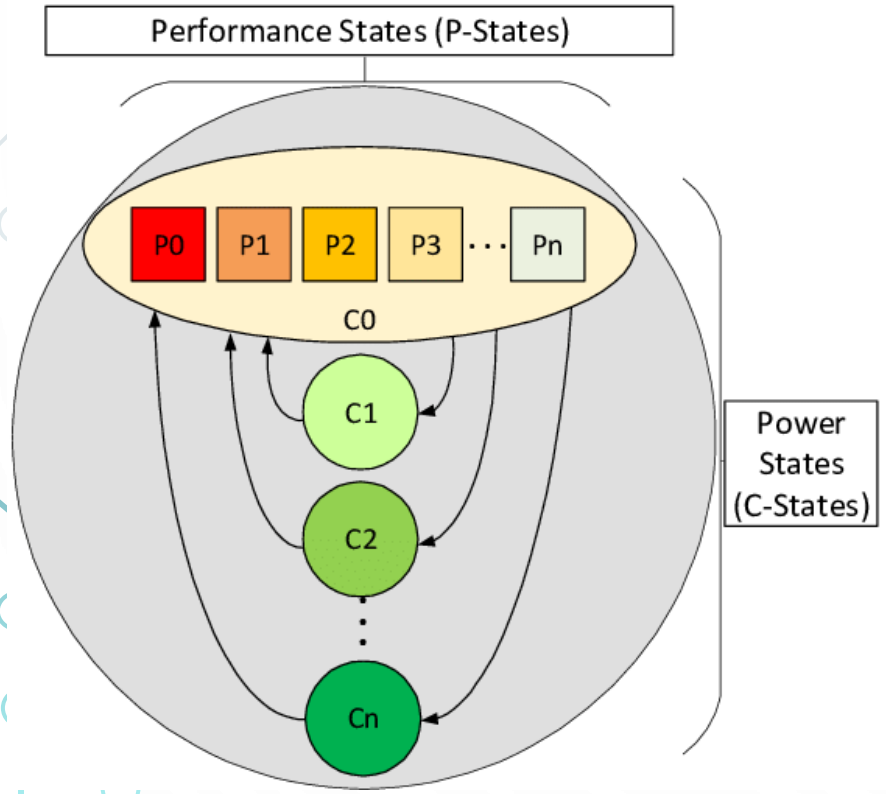
Narendra, ISLPED'01

Lowering Leakage: Sleep Mode

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}

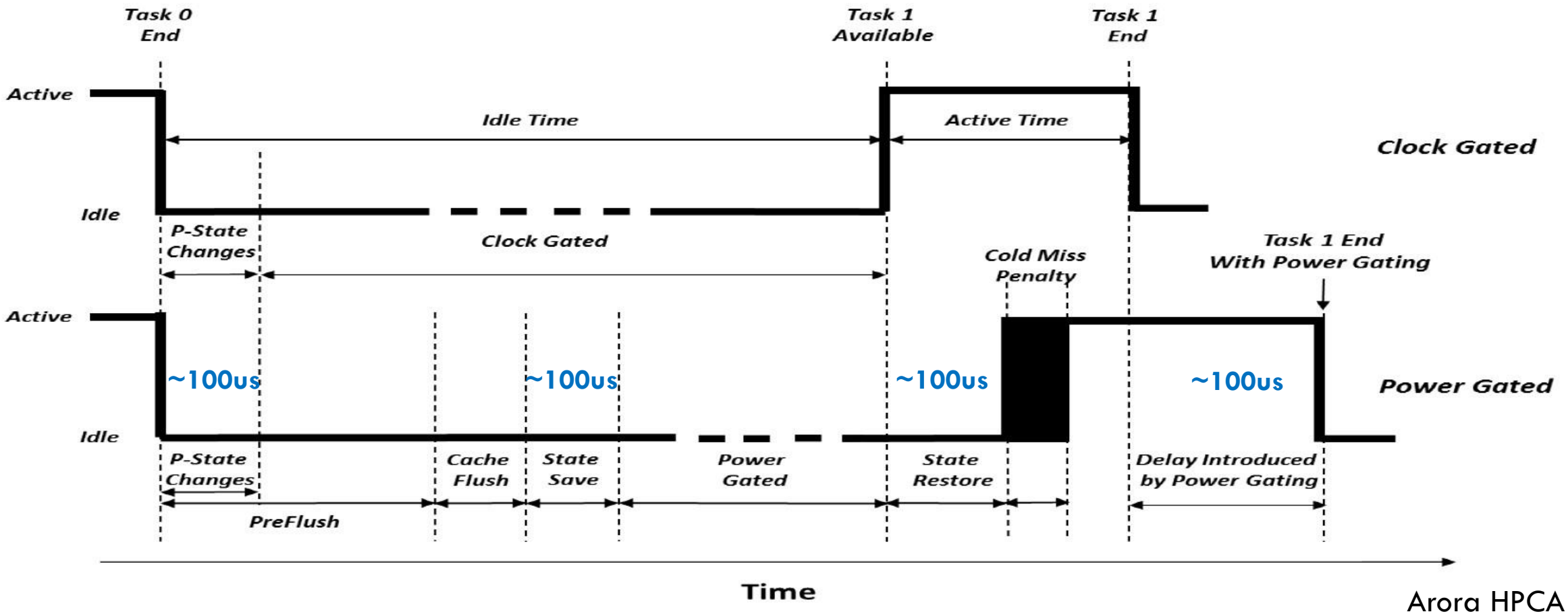
DVFS vs Gating



Software Impact to Platform Energy-Efficiency – Intel 2011

- The more resources are turned-off, the longer it takes to turn back-on and the more transition energy is spent

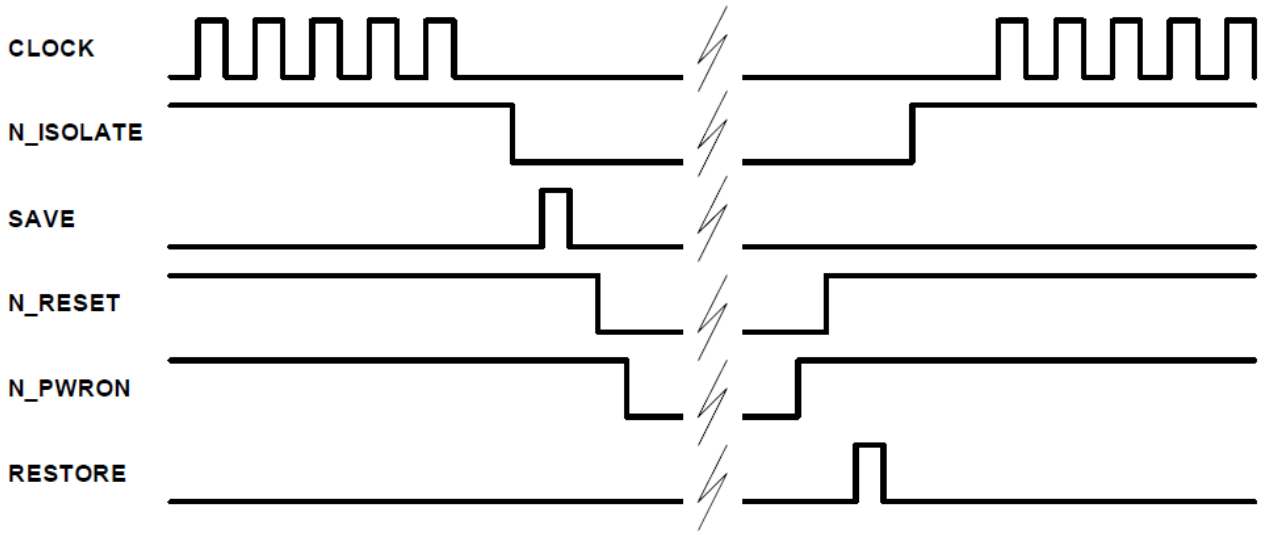
Putting the processor to sleep during idle events



Arora HPCA15

- Power-gating overheads (energy cost, delay) need to be less than the leakage savings to make it worthwhile

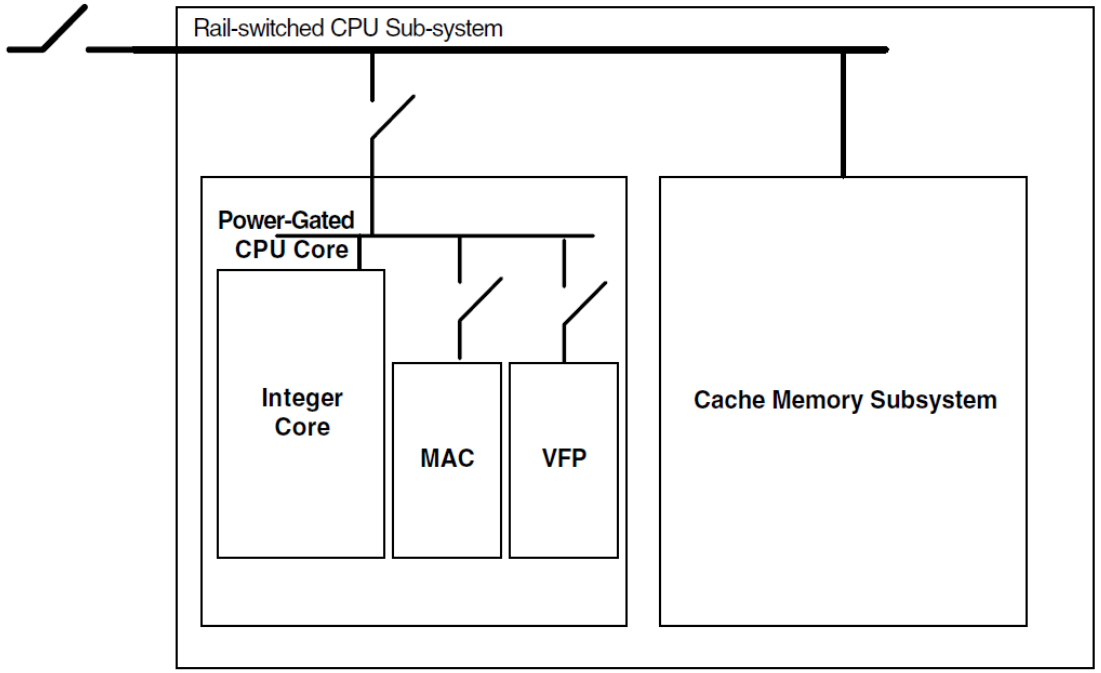
Gating Sequences

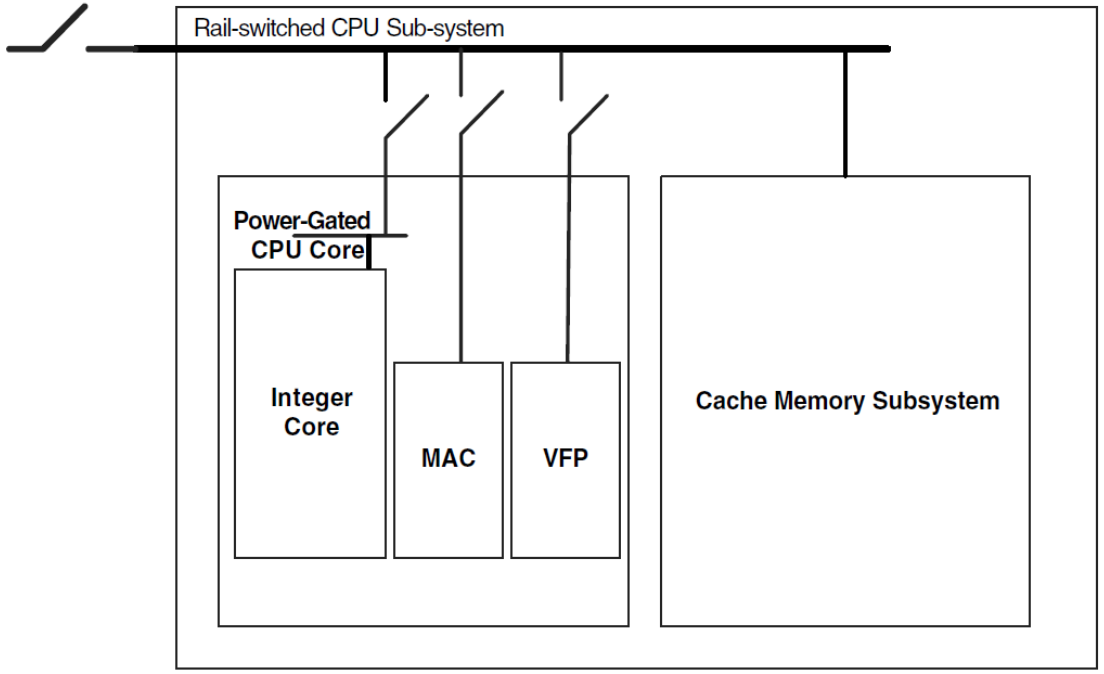


- Sequence of steps:
  - Gate clock
  - Isolate inputs
  - Save (scan out)
  - Reset
  - Gate power

Keating, et al, Low Power Methodology Manual, 2009.

Hierarchical Power Gating





Cache	CPU	MAC	VFP	Power State
(OFF)	(OFF)	-	-	Shutdown (Cache cleaned, VDDCPU off)
ON	OFF	-	-	Deep Sleep (Cache preserved)
ON	ON	OFF	OFF	Normal Operation
ON	ON	ON	OFF	DSP workload
ON	ON	OFF	ON	Graphics workload
ON	ON	ON	ON	Intensive multimedia mode

Cache	CPU	MAC	VFP	Power State
(OFF)	(OFF)	(OFF)	(OFF)	Shutdown (Cache cleaned, VDDCPU off)
ON	OFF	OFF	OFF	Deep Sleep (Cache preserved)
ON	ON	OFF	OFF	Normal Operation
ON	ON	ON	OFF	DSP workload
ON	ON	OFF	ON	Graphics workload
ON	ON	ON	ON	Intensive multimedia mode

Keating, et al, Low Power Methodology Manual, 2009.

Power Gating with Sleep Transistors

- Key components:
  - Power gates (& controller)
    - Leakage vs size
    - Switched capacitance
  - Slew-rate/rush current
  - State preservation
  - Energy overhead of sleep/wake-up transitions

How to Size the Sleep Transistor?

- Don't need both header and footer
- 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 and are slow to turn on/off
  - Minimize the size of the sleep transistor for given ripple (e.g. 5%)
- Need to find the worst-case vector
- Sleep transistor is not for free – it will degrade the performance in active mode
- Charging and discharging the virtual rails costs energy
- Need to sequentially wake up

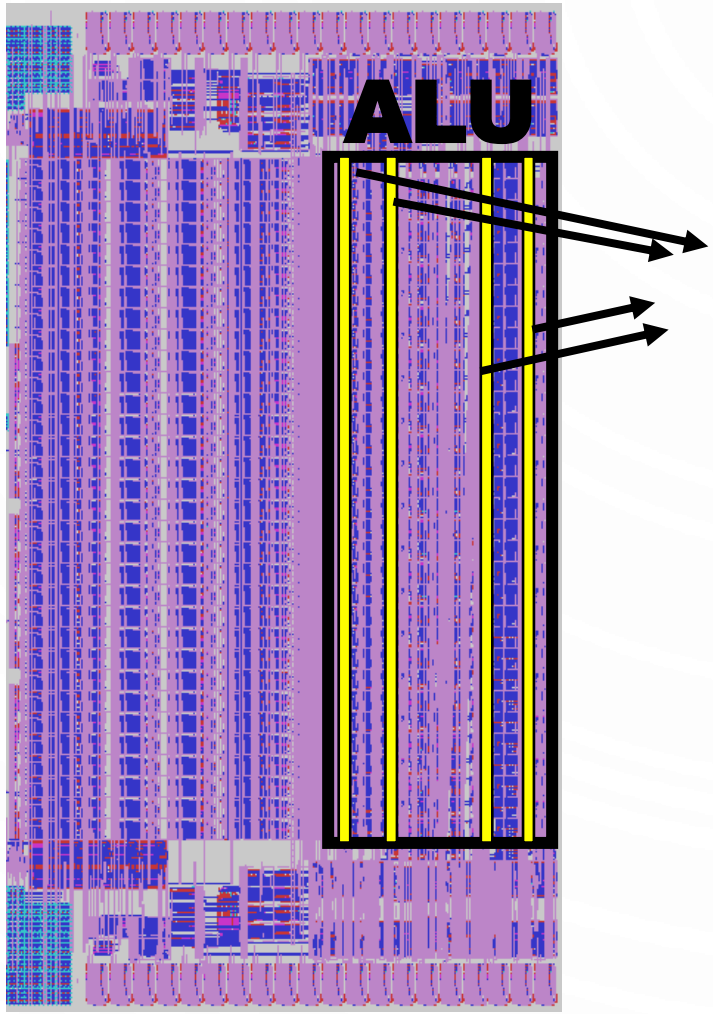
Sleep Transistor

High-VTH transistor (many in parallel) has to be very large for low resistance in linear region.
Low-VTH transistor 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

Courtesy: R. Krishnamurthy, Intel

Sleep Transistor Layout



Area overhead	
PMOS	6%
NMOS	3%

Tschanz, ISSCC'03

Sleep in Standard Cells

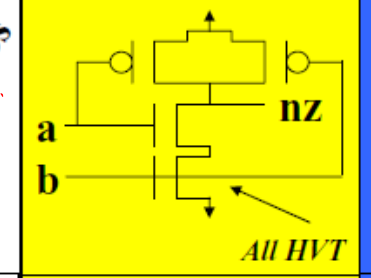
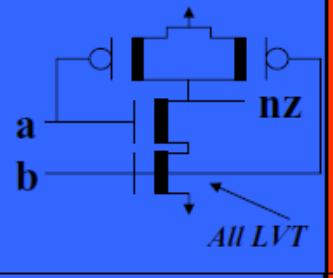
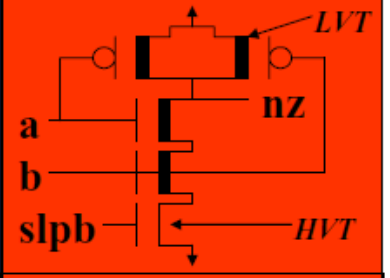
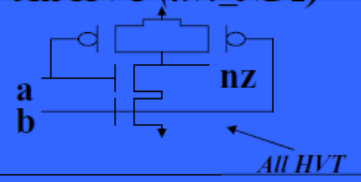
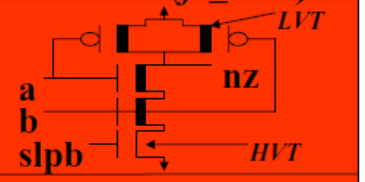
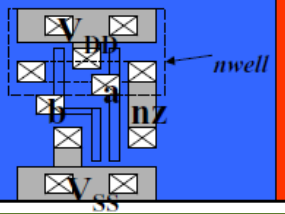
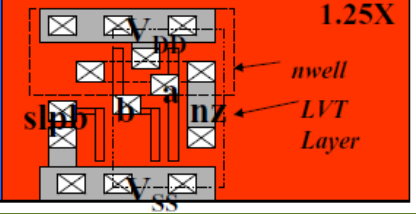
	All HVT *(hvt_ND2)*	All LVT *(lvt_ND2)*	Footswitch *(fs_ND2)*
Schematics			
Perf.	1X	1.5X - 2X	1.4X – 1.8X
Leakage	1X	70X - 100X	~ 1X
Area	1X	1X	1.25X

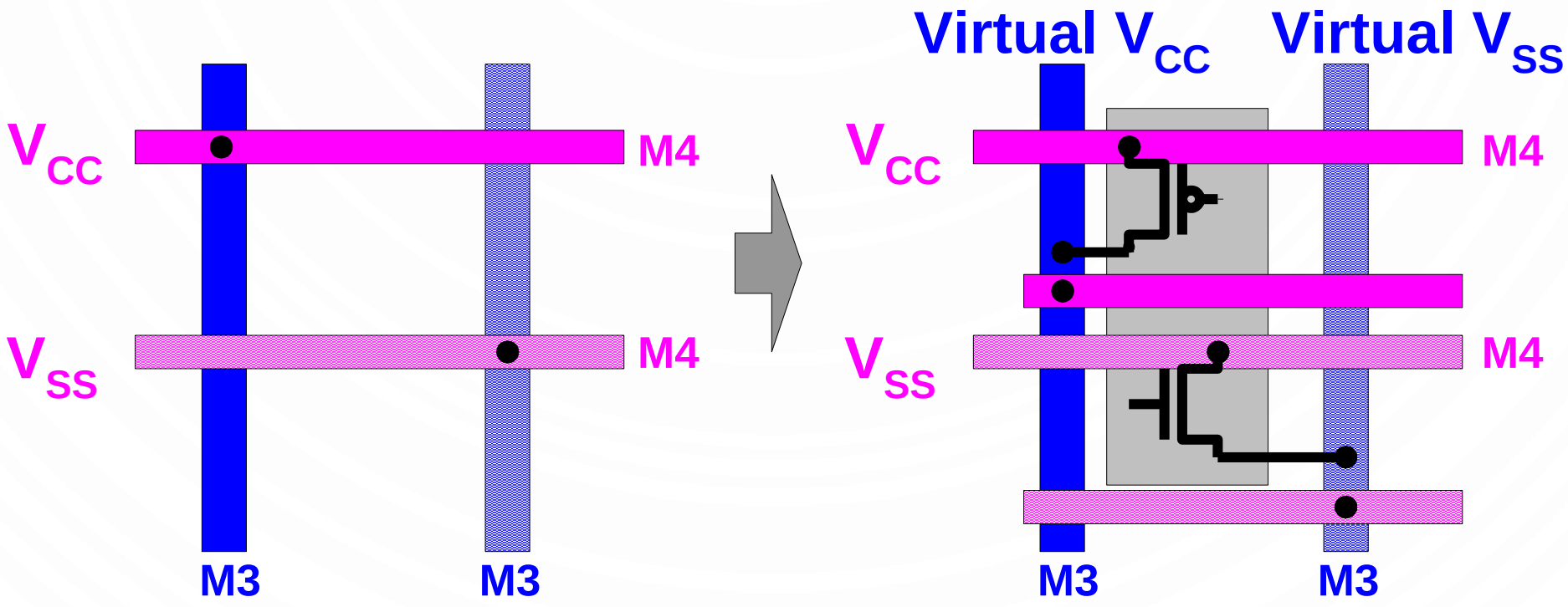
Diagram	All HVT *(hvt_ND2)*	Footswitch *(fs_ND2)*
SCHEMATICS		
LAYOUT		

Uvieghara, ISSCC'04

Sleep Transistor Grid

No sleep transistor

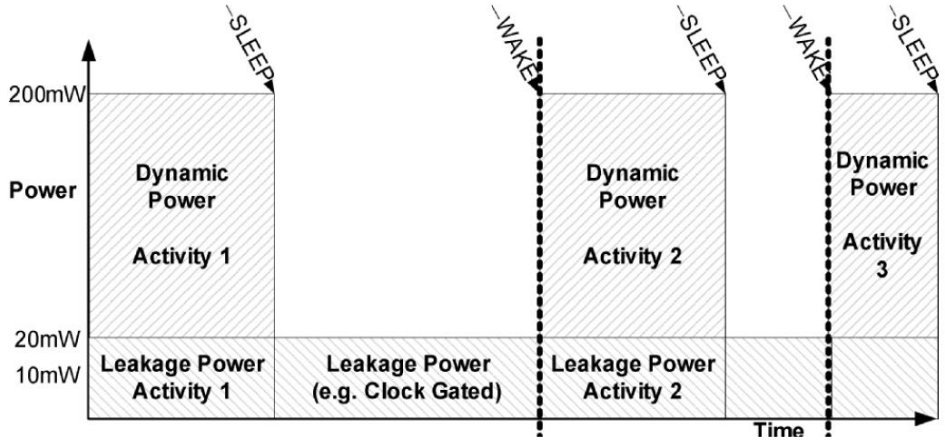
PMOS & NMOS sleep transistors

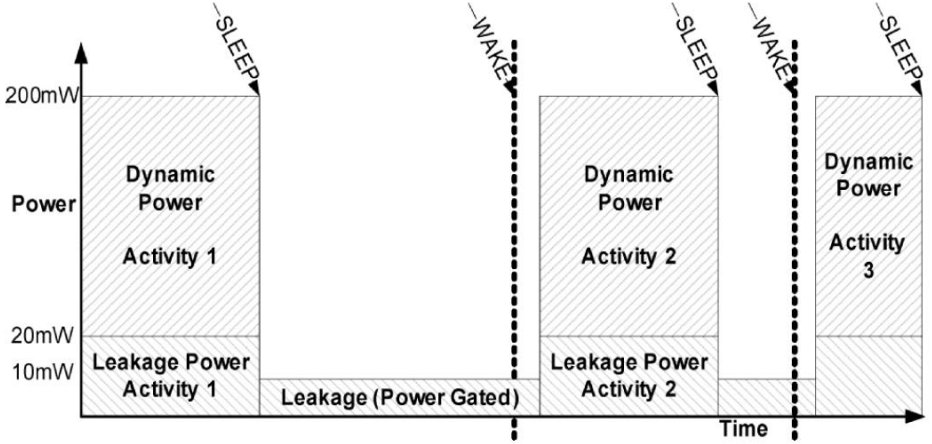


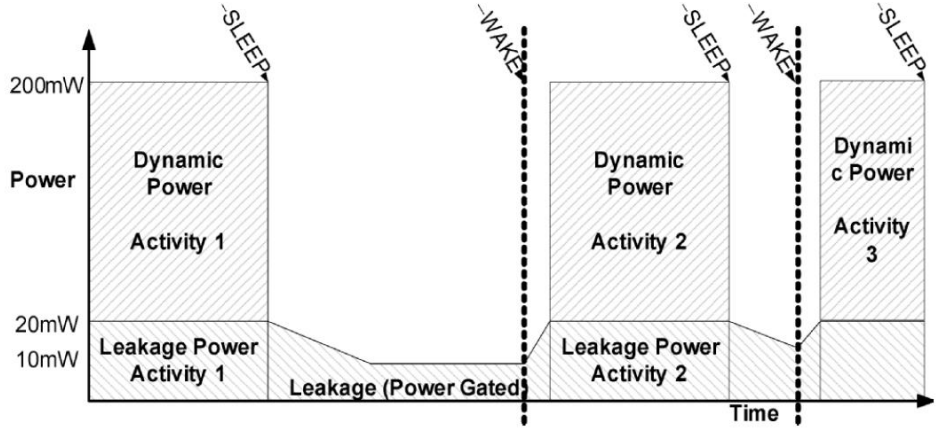
Tschanz, ISSCC'03

Power Gating

- No power gating
- "Ideal" power gating transient
- Realistic profile







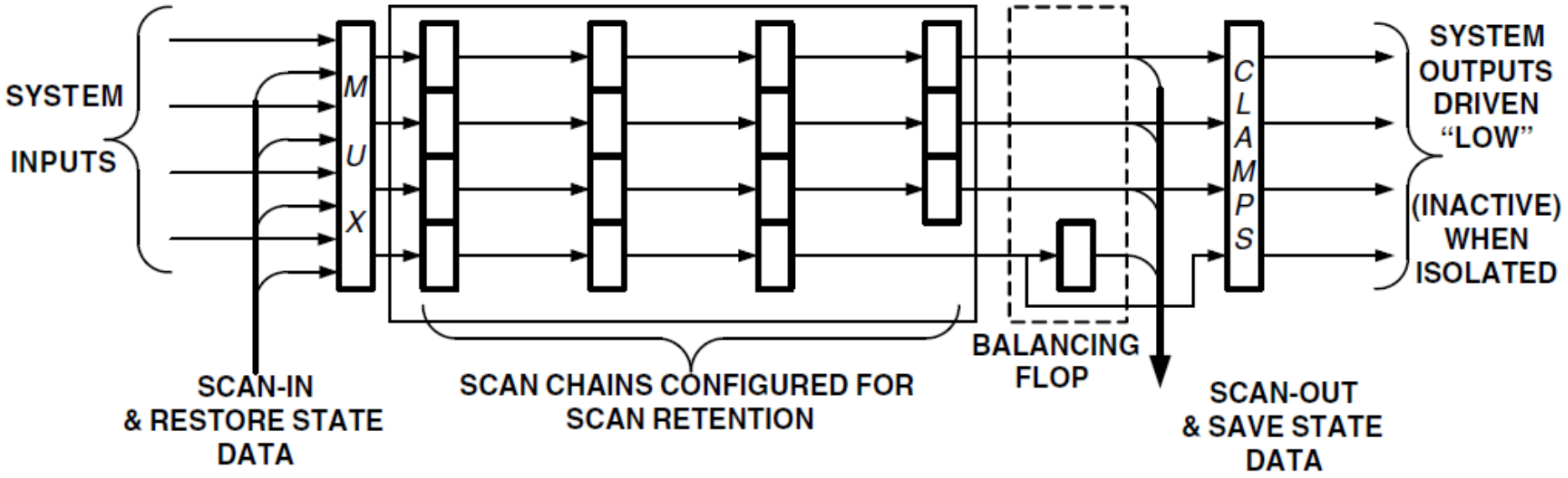
Keating, et al, Low Power Methodology Manual, 2009.

Preserving State

- Virtual supply collapse in sleep mode will cause the loss of state in registers
- Putting the registers at nominal VDD would preserve the state
  - These registers leak
  - The second supply needs to be routed as well
- Can lower VDD in sleep
  - Some impact on robustness, noise and SEU immunity
- State preservation and recovery

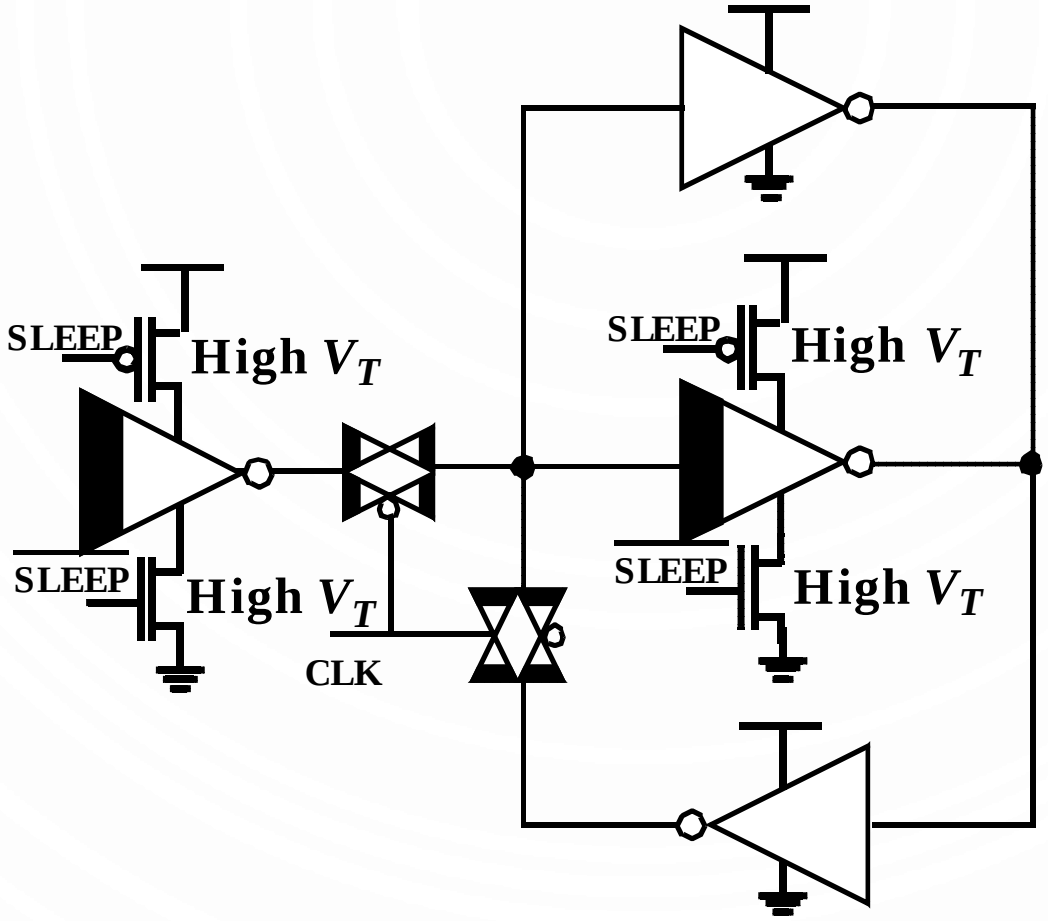
Scan-Based Retention

- Scan-out/scan-in state to preserve/restore state



Keating, et al, Low Power Methodology Manual, 2009.

Retention Register Design



[Mutoh95]

Next Lecture

- More power-saving techniques:
  - Dynamic thresholds
- Energy optimality (Vdd, Vth)