Leakage Reduction Techniques

Stefan Rusu Senior Principal Engineer Intel Corporation

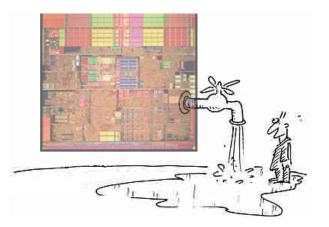
February 3rd, 2008



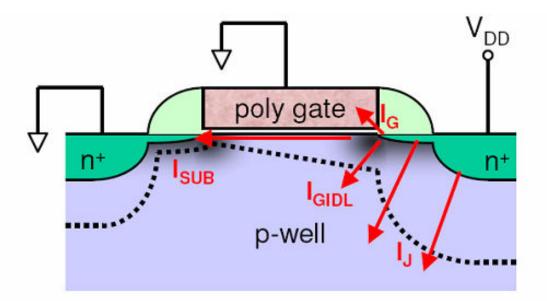


Outline

- Leakage mechanisms and trends
- Leakage reduction techniques
- Future process technology options
- Summary



Transistor Leakage Mechanisms

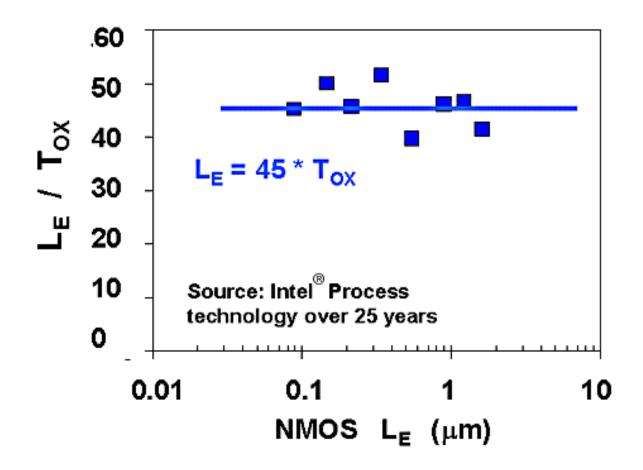


I_{SUB} Subthreshold leakage from source
I_{GIDL} Gate-induced drain leakage (GIDL)
I_J Junction reverse-bias leakage

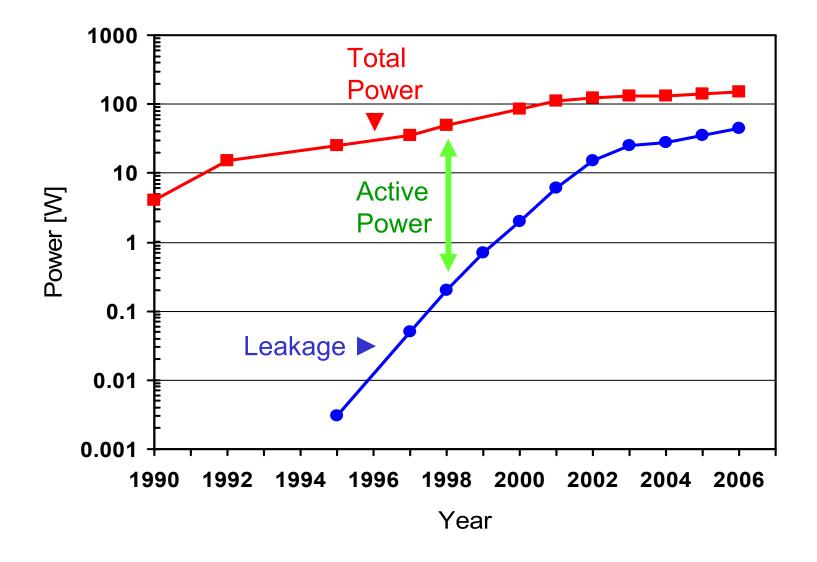
I_G Gate leakage (direct tunneling)

Technology Scaling Trends

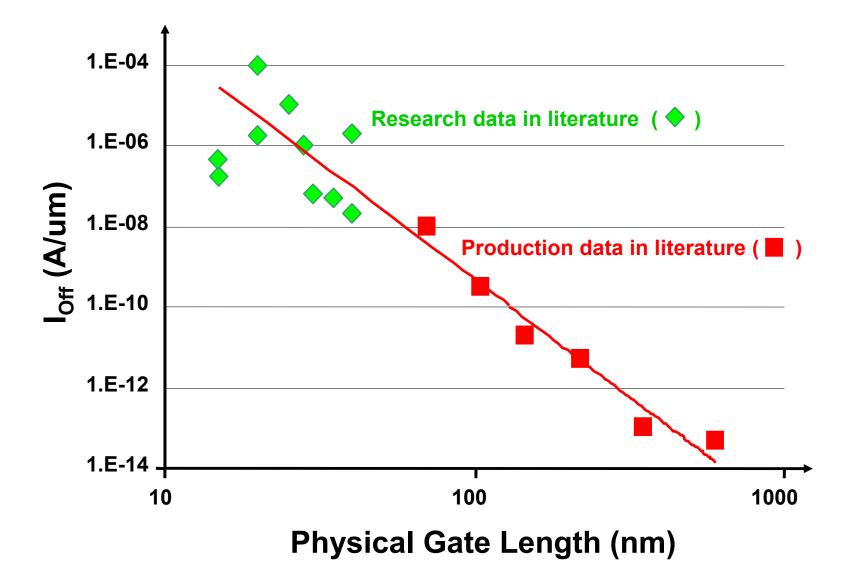
- Physics requires increased leakages at smaller dimensions
- · Failure to deal with leakage implies slower scaling



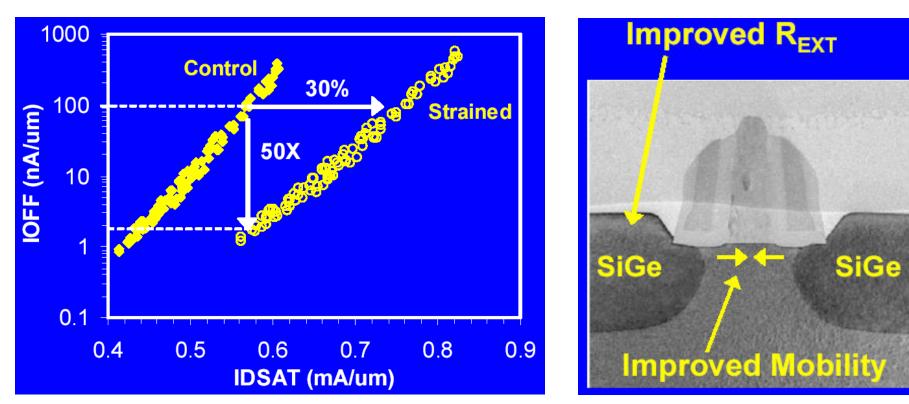
Server Processor Power Trends



Source/Drain Leakage (I_{off})



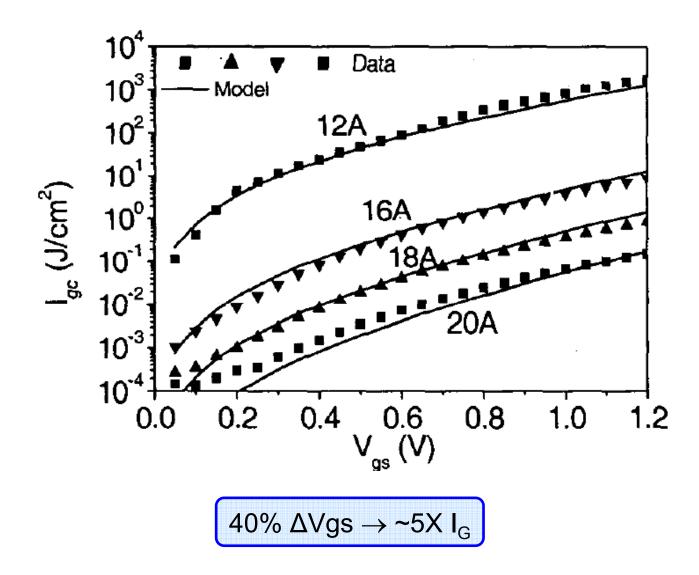
Strained Silicon



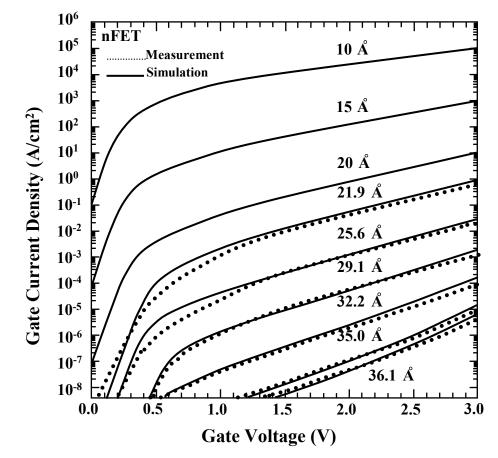
K.Mistry, 2004 VLSI

- 30% Idsat improvement at fixed loff
- 50x loff reduction at fixed ldsat
- Performance and leakage are intimately related

Gate Leakage Trends



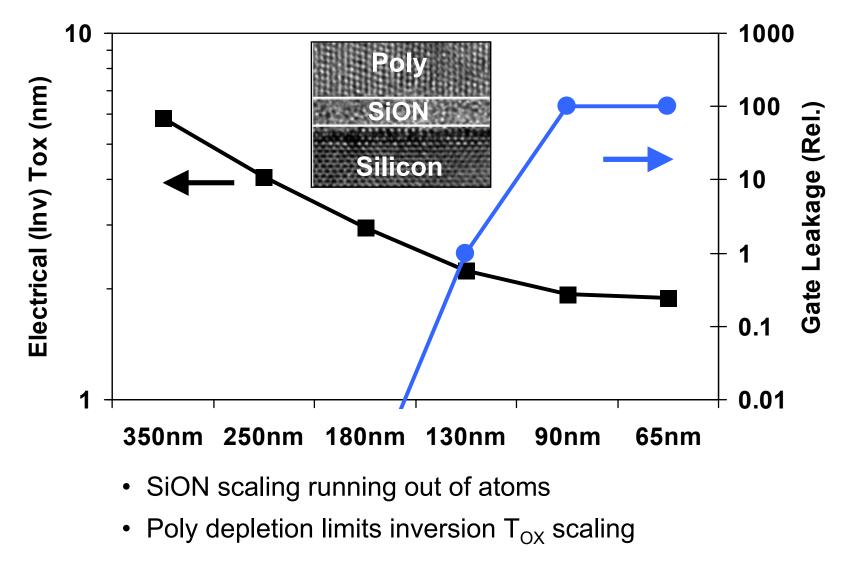
Gate Leakage Trends (cont)



 Leakage increases by 10x for every 2Å (SiO₂) gate thickness reduction

K. Itoh, Trends in Low-Voltage RAM Circuits, FTFC 2003

Gate Leakage Trends (cont)



[[]Mistry, et. al, IEDM 2007]

45nm High-K + Metal Gate Transistors

Metal Gate

- Increases the gate field effect

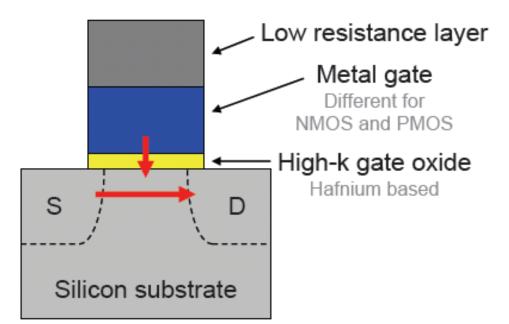
High-K Dielectric

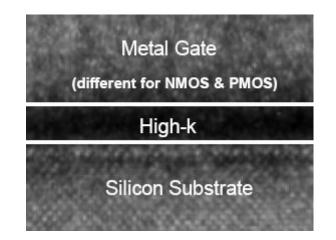
- Increases the gate field effect
- Allows use of thicker dielectric layer to reduce gate leakage

HK + MG Combined

- Drive current increased >20%
- Or source-drain leakage reduced >5x
- Gate oxide leakage reduced

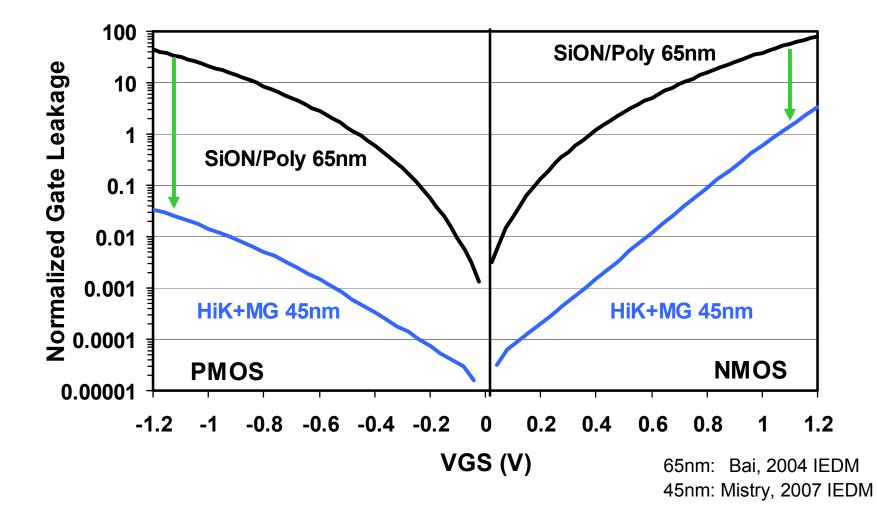
http://download.intel.com/pressroom/kits/45nm/ Press%2045nm%20107_FINAL.pdf



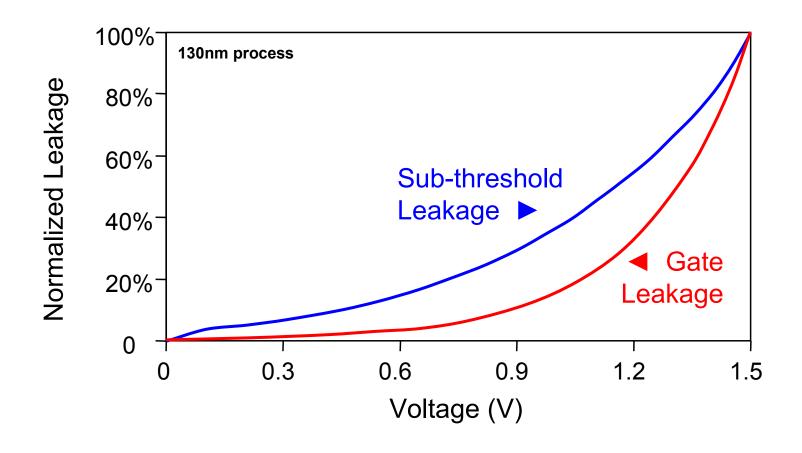


HK+MG Gate Leakage Reduction

• Gate leakage is reduced >25X for NMOS and 1000X for PMOS



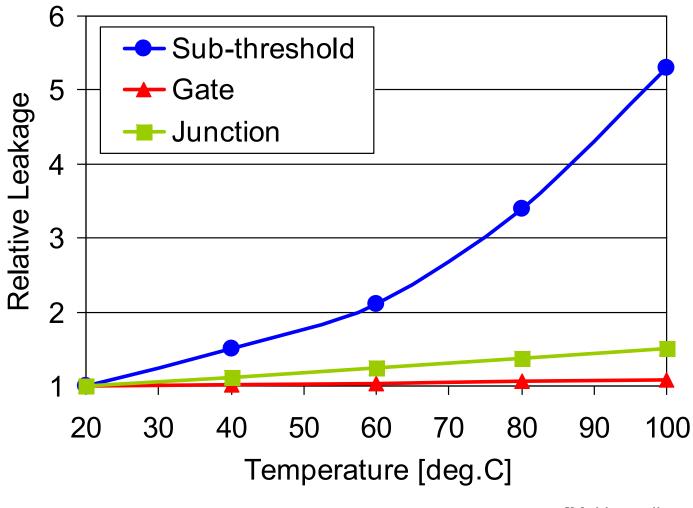
Voltage Dependence



Leakage components are a strong function of voltage

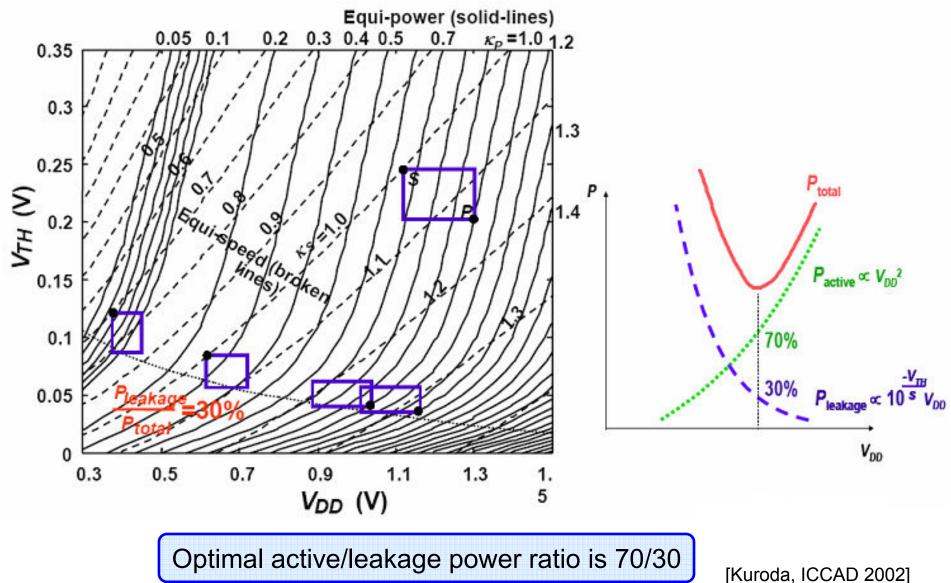
[Krishnamurthy, et. al, ASICON 2005]

Temperature Dependence



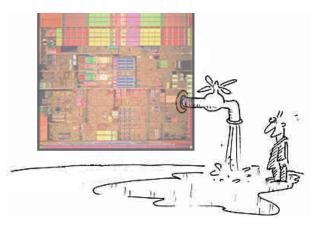
[Mukhopadhyay, et al., VLSI Symposium 2003]

Optimal Active / Leakage Power Ratio



Outline

- Leakage mechanisms and trends
- Leakage reduction techniques
- Future process technology options
- Summary

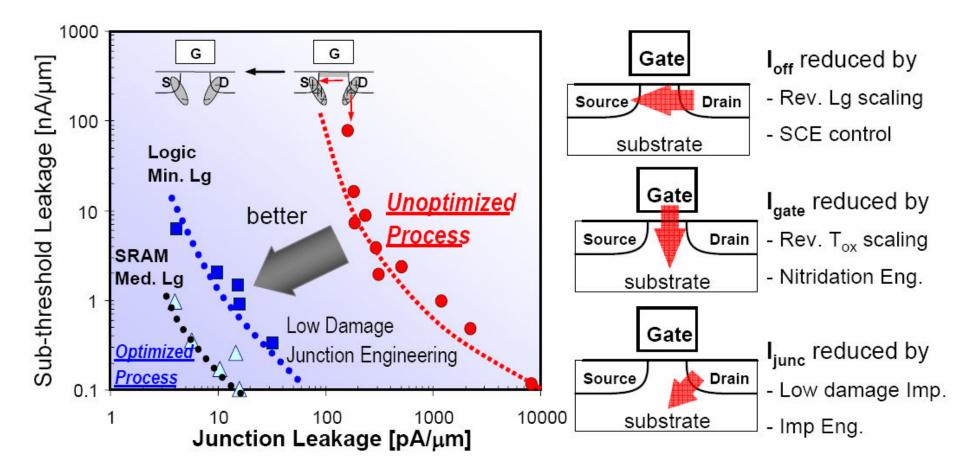


Leakage Reduction Techniques

Transistor level

- Longer transistor Le
- Higher-Vt devices
- Block level
 - MOS Threshold Voltage Control (MTCMOS, VTCMOS)
 - Gate Voltage Control (SCCMOS, BGMOS)
 - Transistor stacking methods
 - Cache leakage reduction
 - Sleep transistors
- Chip level
 - Multiple voltage supplies
 - Adaptive body bias

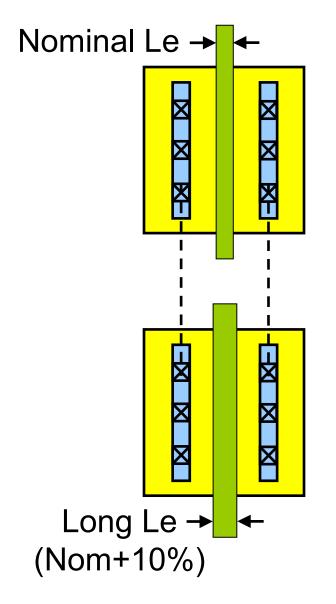
Transistor Leakage Reduction



Low damage junction engineering for <u>sub-threshold</u> and junction leakage reduction

[Wang, et. al, ISSCC, 2007]

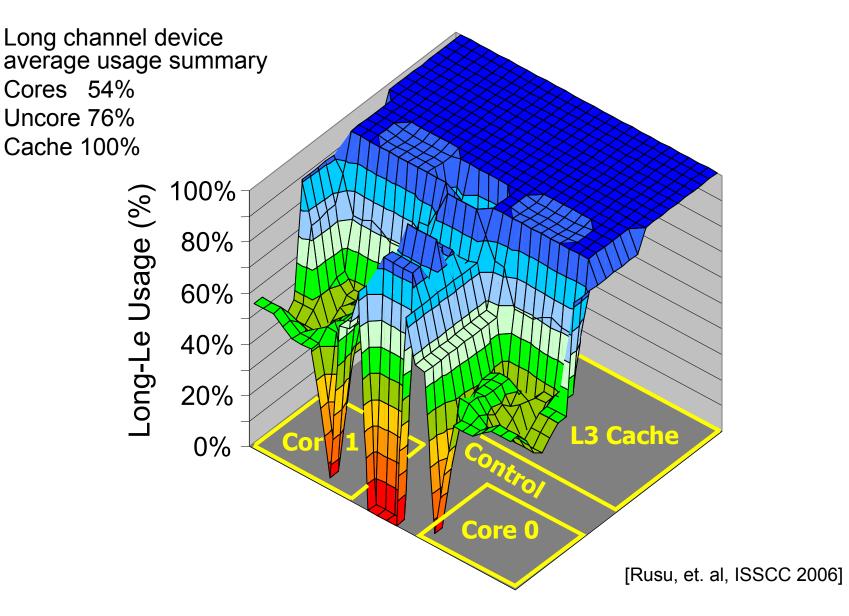
Long-Le Transistors

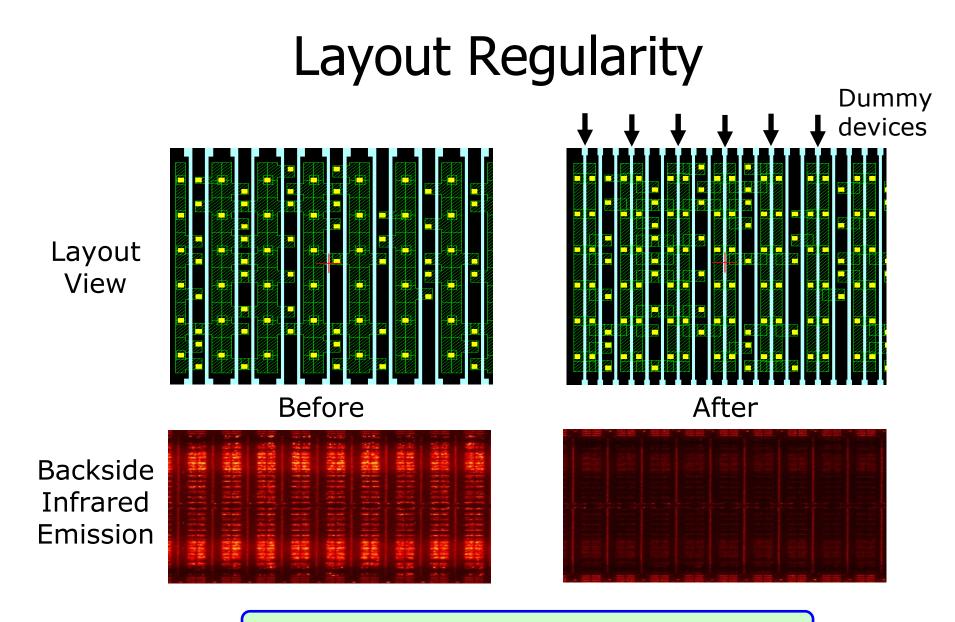


- All transistors can be either nominal or long-Le
- Most library cells are available in both flavors
- Long-Le transistors are ~10% slower, but have 3x lower leakage
- All paths with timing slack use long-Le transistors
- Initial design uses only long channel devices

[Rusu, et. al, ISSCC 2006]

Long-Le Transistors Usage



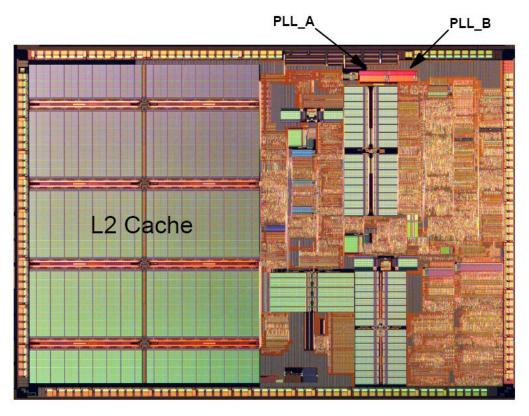


Regular layout reduces variability and leakage

[Rikhi and Rusu, IDF 2006]

Low-Leakage PowerPC 750

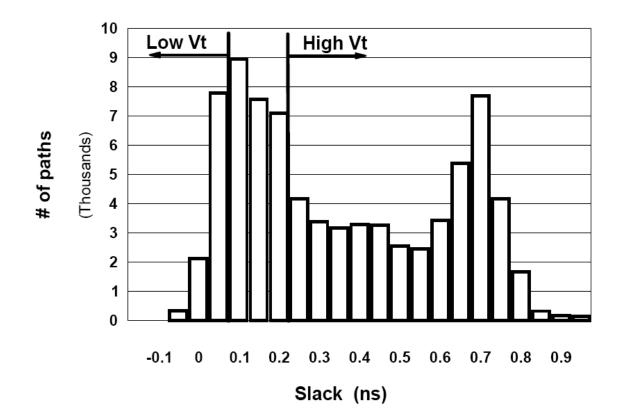
Processor Core	PPC750
Process Technology	0.13µm
Metal Levels	6 Layer Cu
Core Voltage	1.1V - 1.5V
I/O Voltage	1.2V, 1.5V, 1.8V, 2.5V, 3.3V
Lpoly	0.07µm
Тох	1.4nm / 4.0nm
SRAM cell area	2.16µm²
Chip size	34mm ²
Transistor Count	39.7 million



 0.13um SOI CMOS technology with copper interconnect, low-k dielectric, multi-threshold transistors, and dual oxide thickness FETs

[Geissler, et. al, ISSCC 2002]

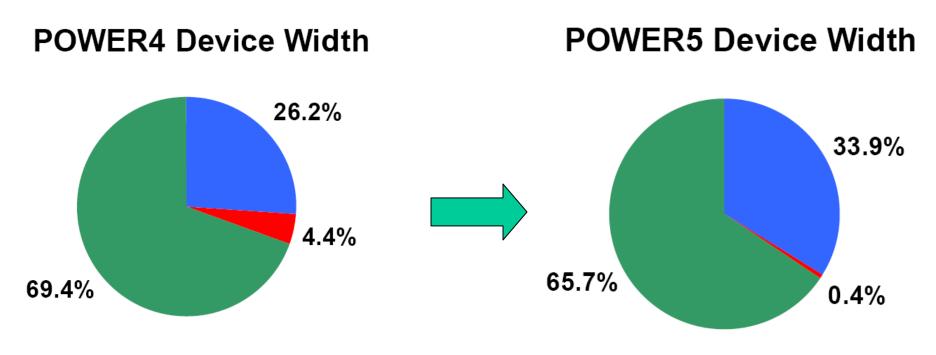
Triple Vt Usage in PowerPC 750



- Three threshold voltage devices are used for both the nFET and pFET, high Vt, standard Vt and low Vt devices
- Low Vt devices are used in frequency-limiting paths only
- High Vt devices are used in paths which are not frequencylimiting and in SRAM arrays

[Geissler, et. al, ISSCC 2002]

Power5 Leakage Reduction

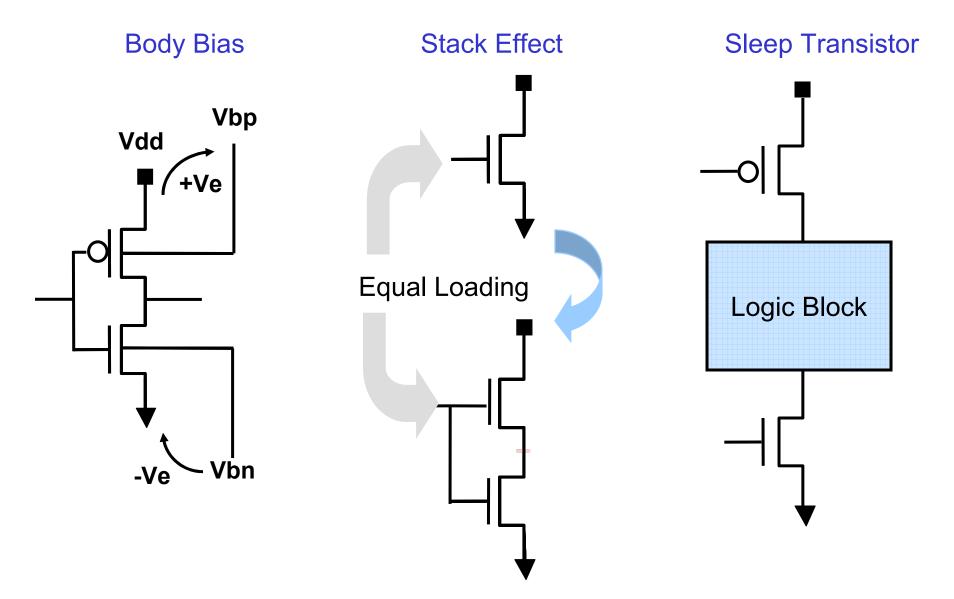


high Vt low Vt normal Vt

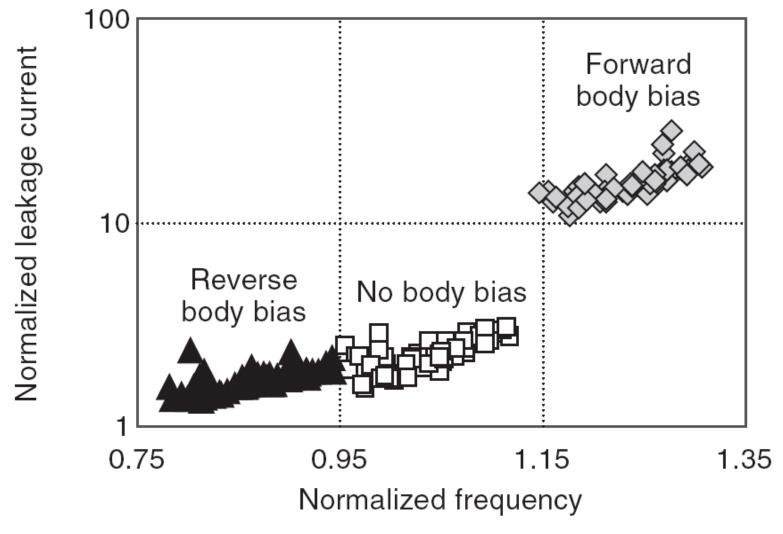
IBM's Power Processors are leveraging triple Vt process option

[Clabes, et al., ISSCC 2004]

Leakage Reduction Circuit Techniques



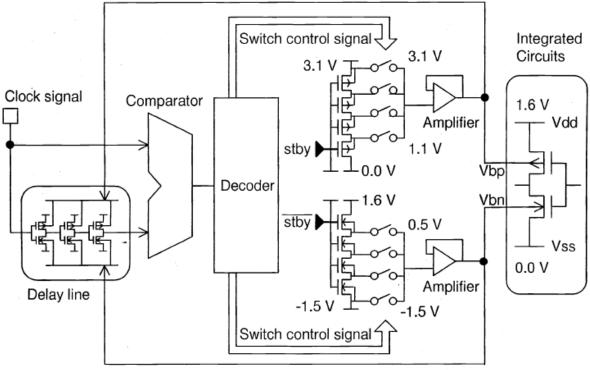
Body Bias Leakage Reduction



[Keshavarzi, et al., D&TC 2002]

Speed-adaptive Vt with Forward Bias

pMOS substrate bias

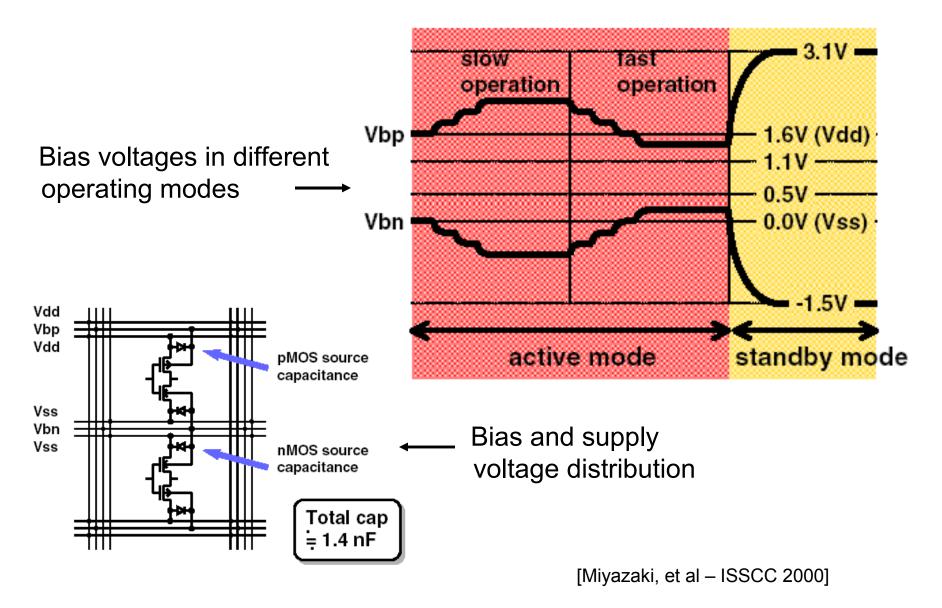


nMOS substrate bias

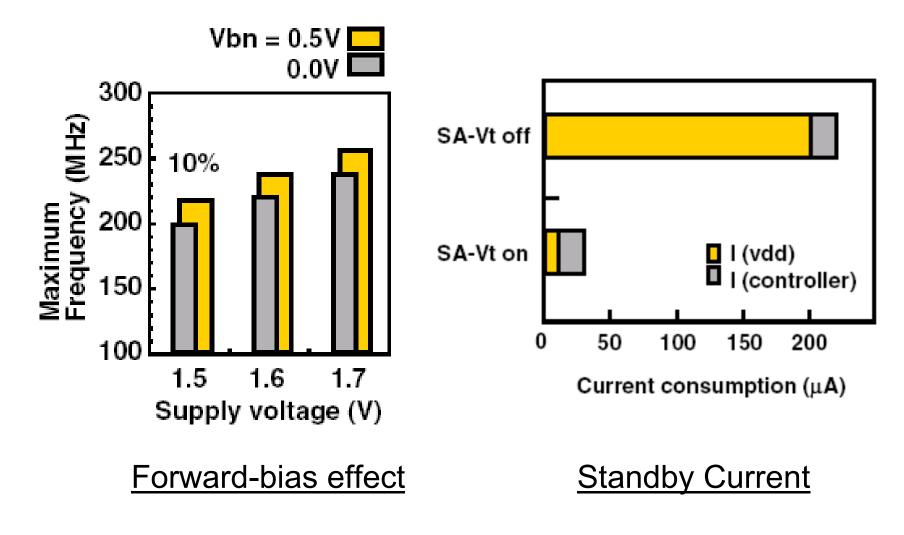
Supply voltage Clock frequency Power consumption Standby current 1.5 - 1.8V 220MHz 320 - 380mW 30μA Gate length - 0.2 um Oxide thickness - 4.5 nm Interconnect metal - 5 layers Well structure - triple well

[Miyazaki, et al – ISSCC 2000]

SA-Vt Implementation

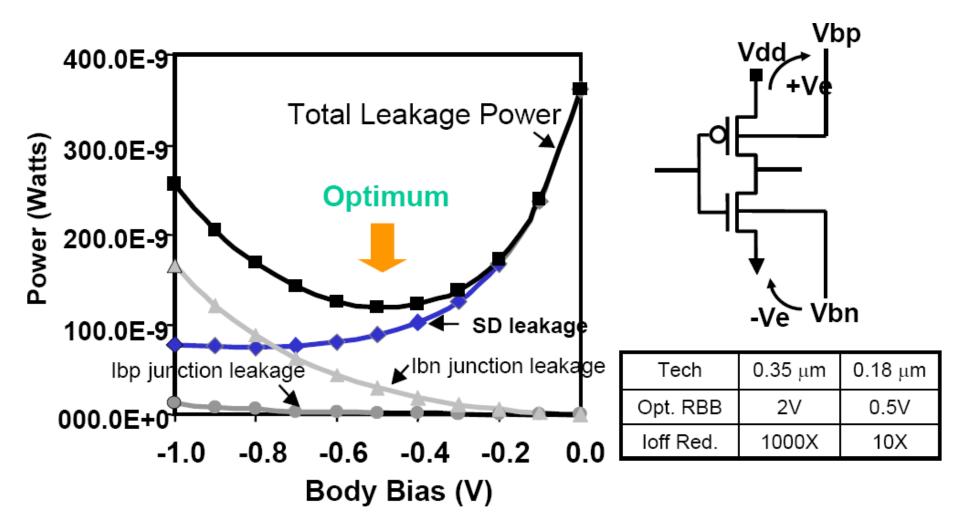


SA-Vt Experimental Results

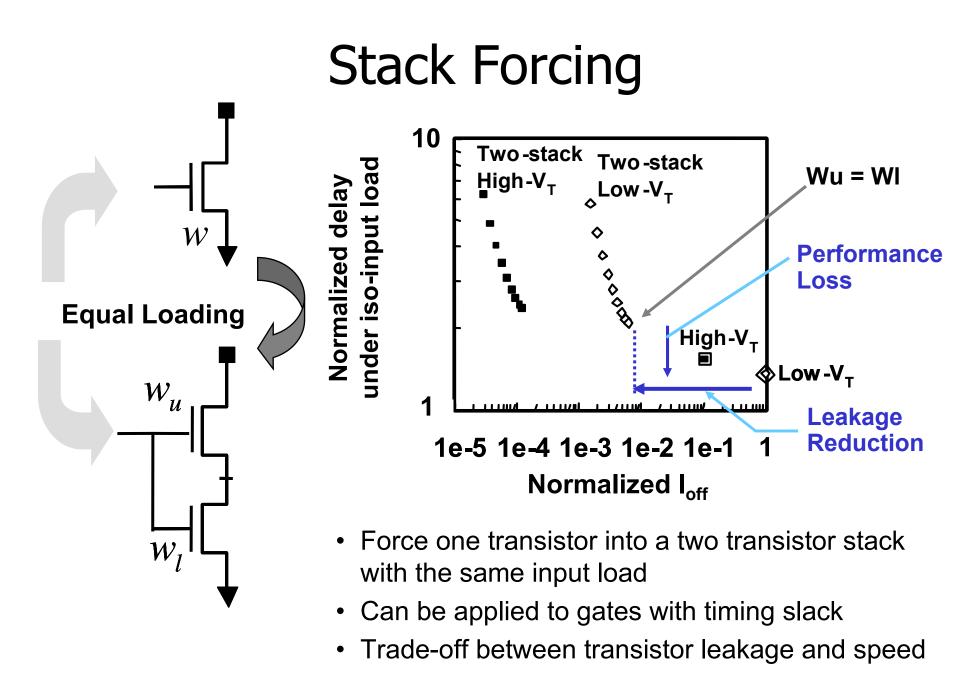


[Miyazaki, et al – ISSCC 2000]

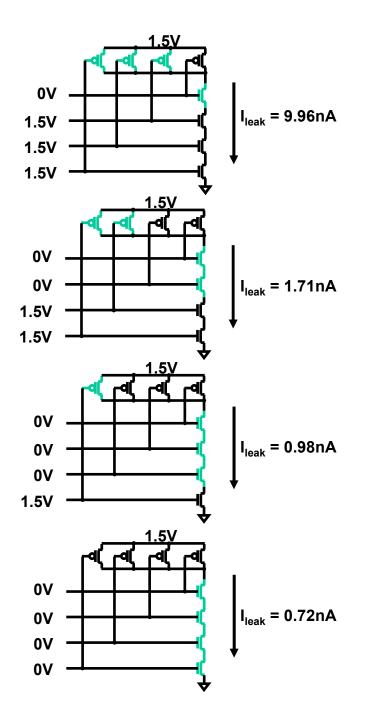
Scalability of Reverse Body Bias



Reverse Body Bias is less effective with technology scaling [Keshavarzi, et al – ISLPED 1999]

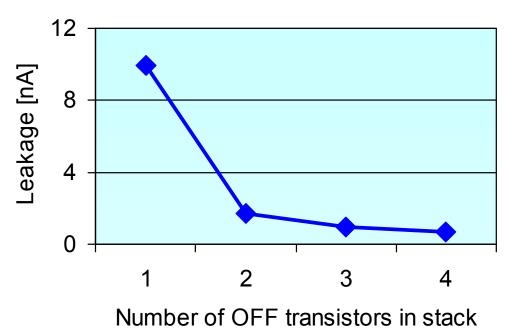


[Narendra, et al – ISLPED 2001]

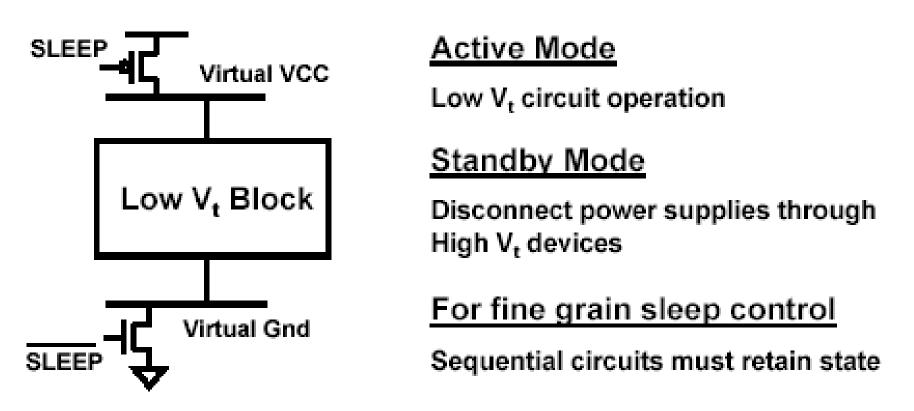


Natural Stacks

- Leakage reduced significantly when two transistors are off in a stack
- Educate circuit designers, monitor average stacking factor



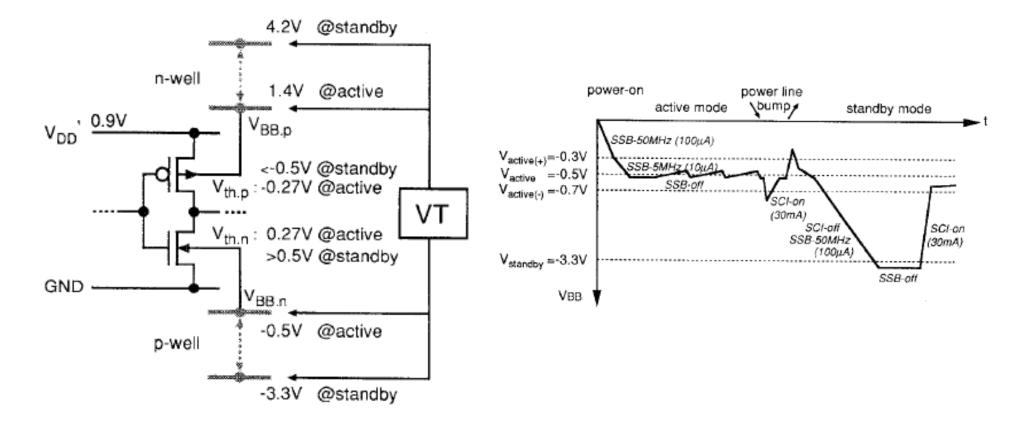
Multi-Threshold CMOS (MTCMOS)



Single polarity sleep device sufficient for combinational logic block

[S. Shigematsu, et al - VLSI Symposium, 1995]

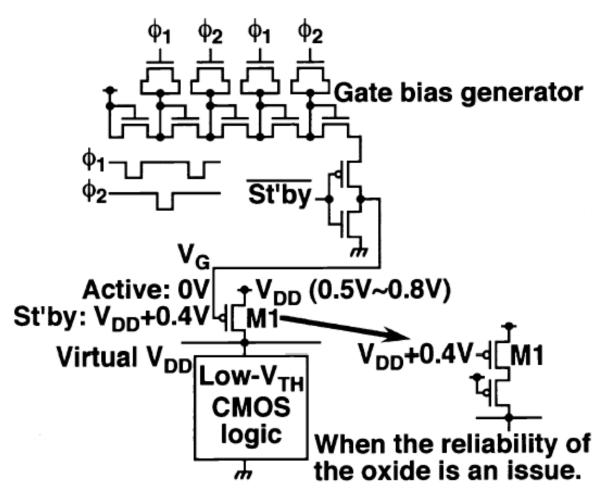
Variable Threshold CMOS (VTCMOS)



- In active mode, the VT circuit controls the active bias to compensate Vt fluctuations
- In standby mode, the VT circuit applies deeper substrate bias to cut off leakage

[T. Kuroda, et al - JSSC, Nov. 1996]

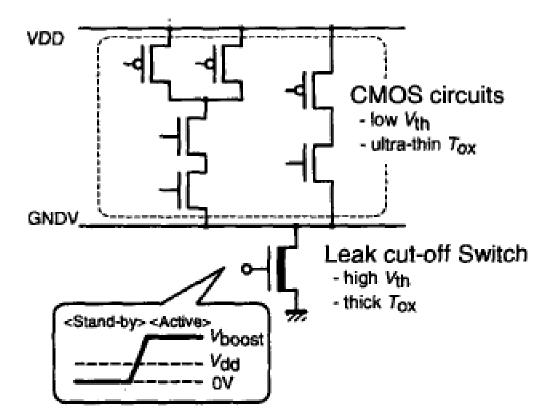
Super Cut-off CMOS (SCCMOS)



- On-chip voltage boost for the sleep control signal reverse biases the sleep PMOS device to suppress leakage
- Requires N-well separation and an efficient on-chip boost voltage generator
- Oxide reliability is another concern

[H. Kawaguchi, et al – ISSCC 1998]

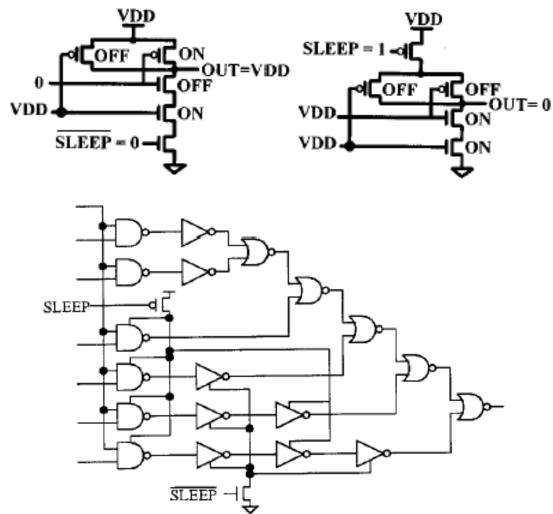
Boosted Gate MOS (BGMOS)



- Logic circuits use thin Tox and low Vt transistors
- Leakage cut-off switches use thick Tox and high Vt devices
- In active mode the leakage cut-off switches are driven by a boosted gate voltage to reduce the area penalty
- Requires dual supply voltages and dual Tox manufacturing process

[T. Inukai, et. al – CICC 2000]

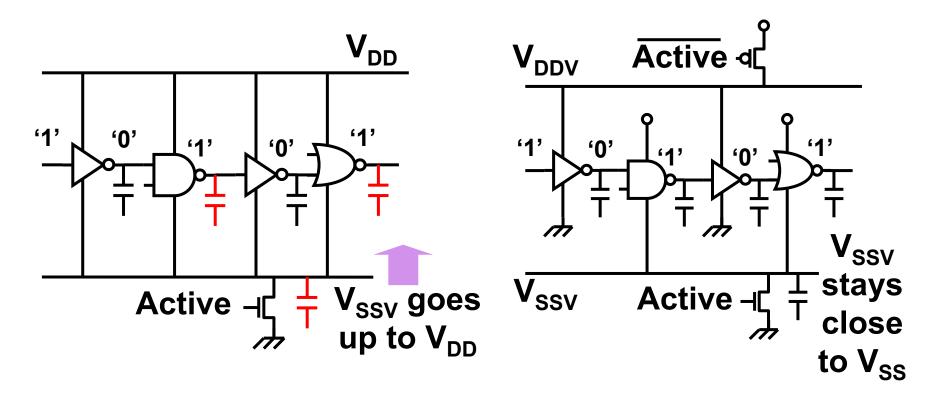
Stacking Transistor Insertion



- Identify first the circuit input vector that minimizes leakage
- For each gate in a high leakage state insert a leakage control transistor
- Requires intensive simulations for the optimal vector

[M. Johnson, et al – IEEE Trans. On VLSI, 2002]

Zigzag Cut-off CMOS

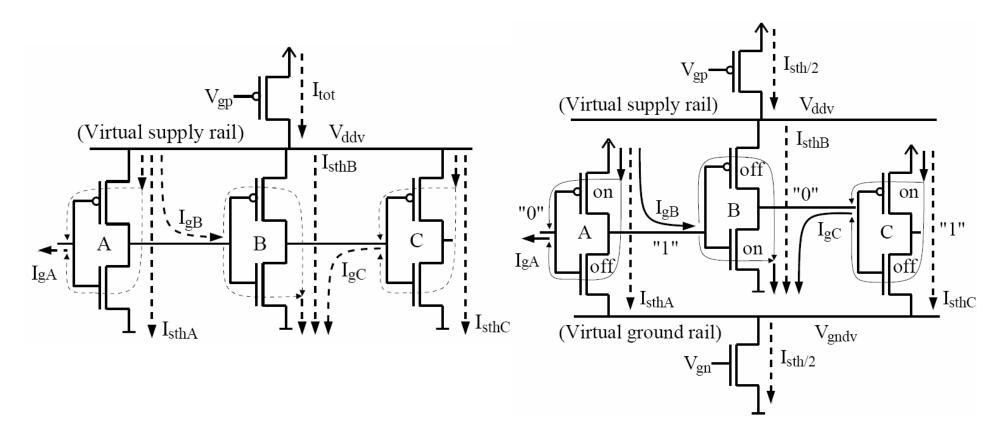


Conventional cut-off switch several clock cycles to wake-up

Zigzag scheme x10~100 wake-up speed

> [Min, et al., ISSCC 2003] University of Tokyo

Super Cut-Off vs. Zig-Zag

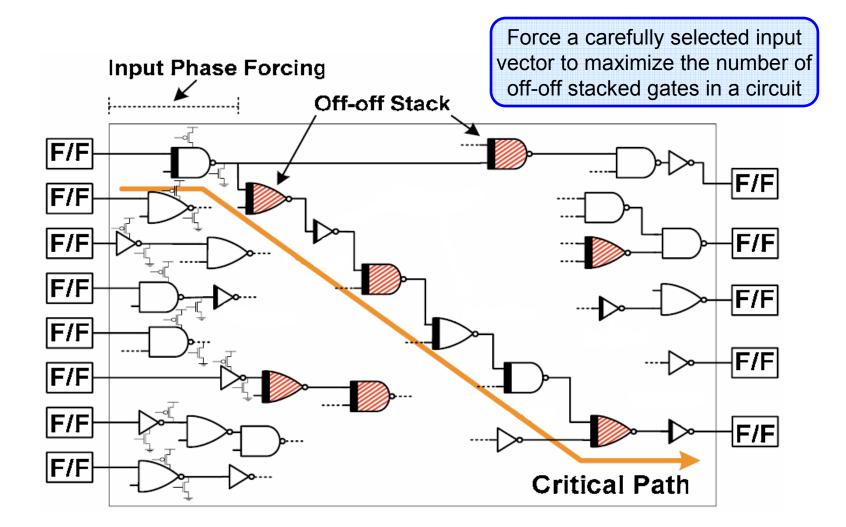


Super Cut-Off CMOS (SCCMOS)

Zigzag Super Cut-Off CMOS (ZSCCMOS)

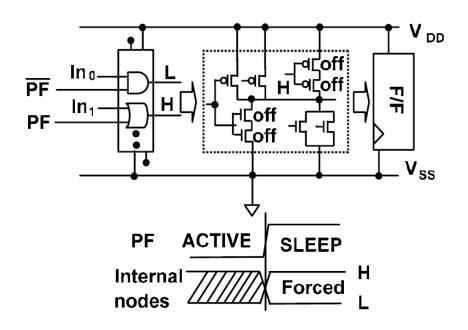
> [Drazdziulis, et al., ESSCIRC 2004] Chalmers University, Sweden

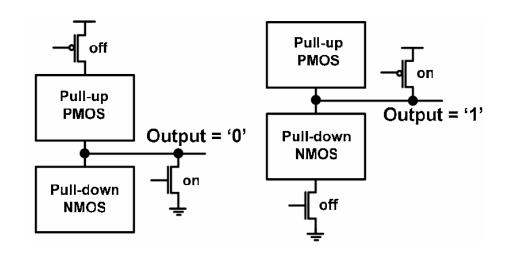
Input Phase Forcing



[Choi, et al., VLSI Symp. 2005] University of Tokyo

Input Phase Forcing Circuit

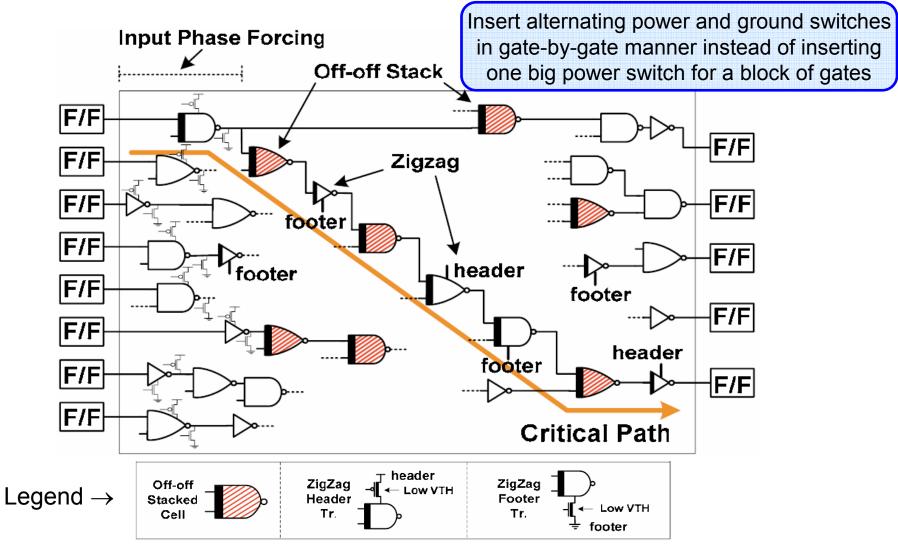




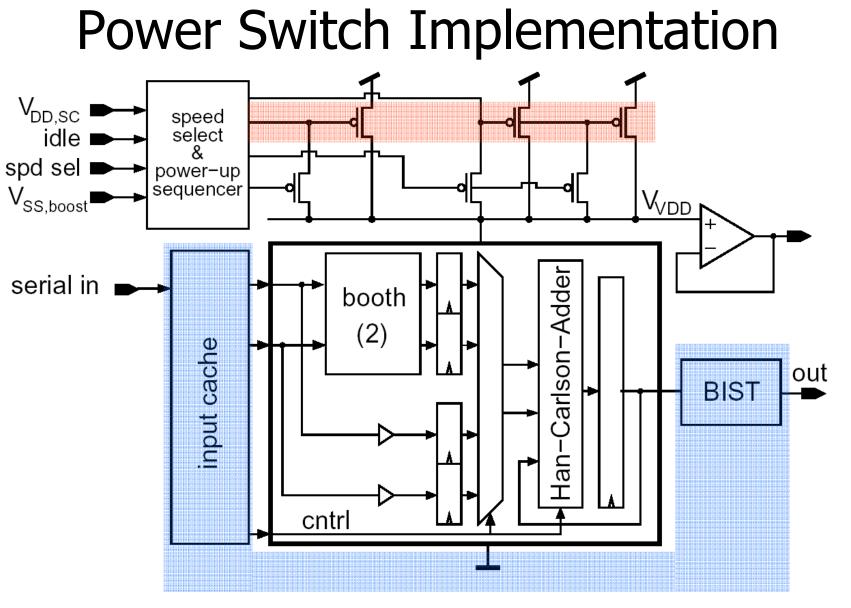
- Input gates are modified to drive the required input vector
- A random-based Monte Carlo search is used to identify the optimal input vector
- Circuit modifications for input phase forcing
- Total delay overhead is less than 2% on average

[Choi, et al., VLSI Symp. 2005]

Optimal Zigzag Power Gating

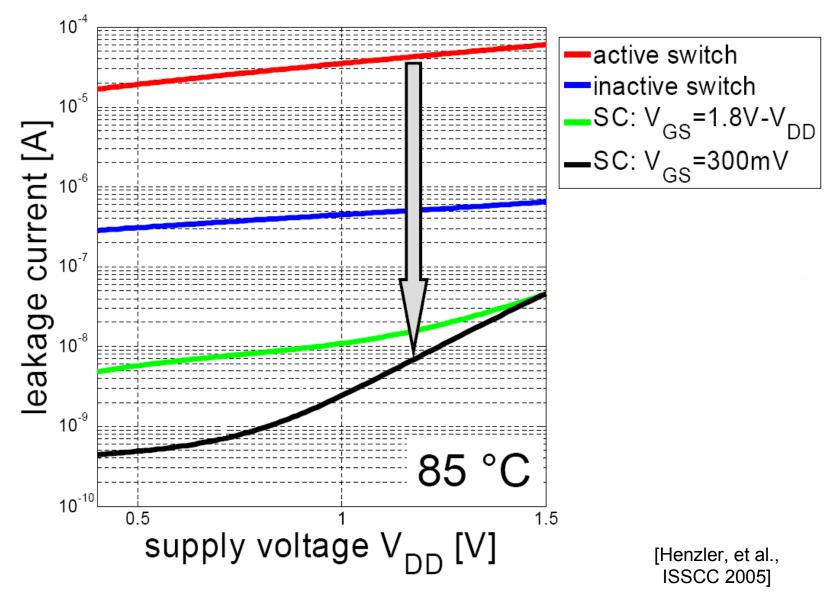


[Choi, et al., VLSI Symp. 2005]

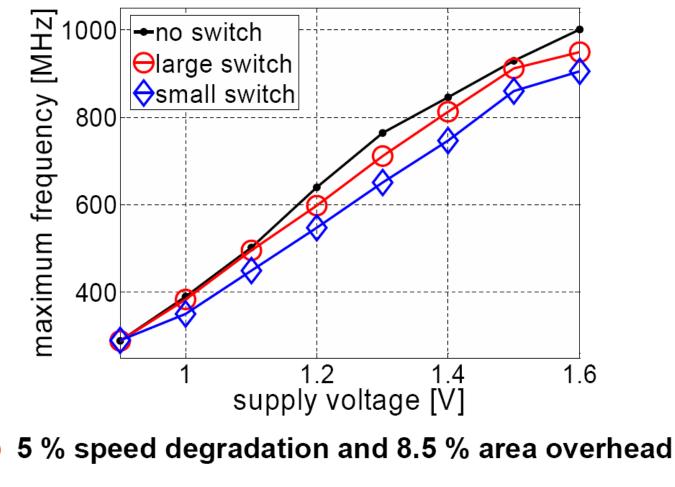


• 16-bit MAC in 130nm process technology [Henzler, et al., ISSCC 2005] TU Munich + Infineon Technologies

Power Switch Benefit

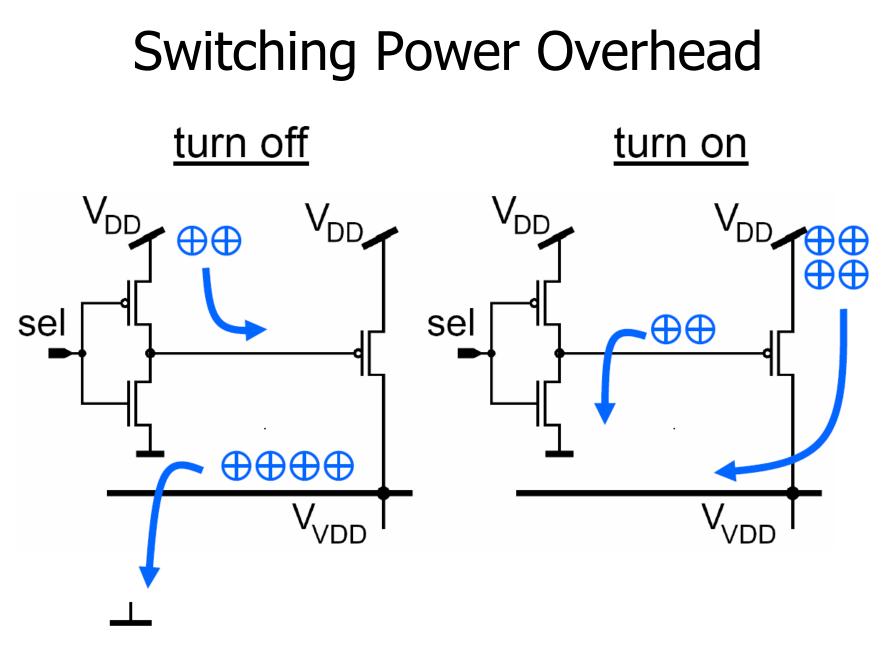


Power Switch Overhead



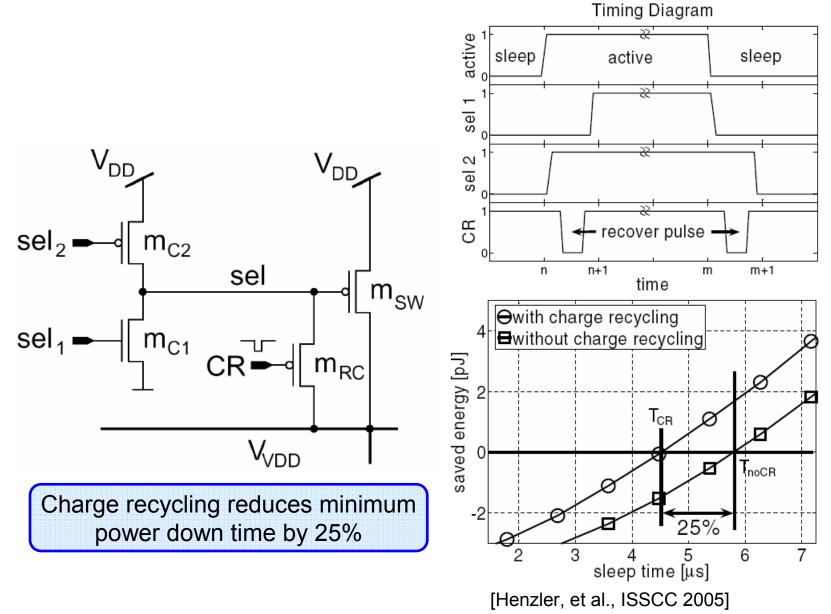
9.5 % speed degradation and 2.8 % area overhead

[Henzler, et al., ISSCC 2005]

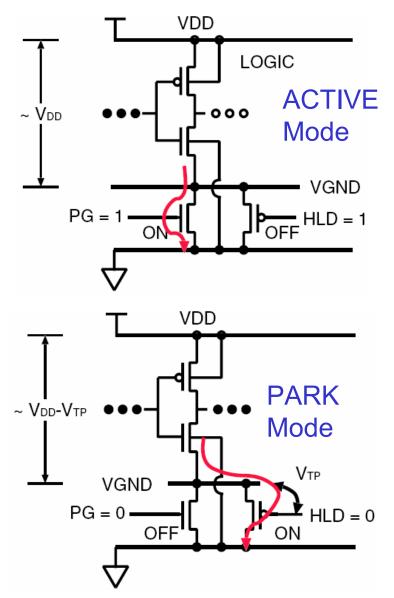


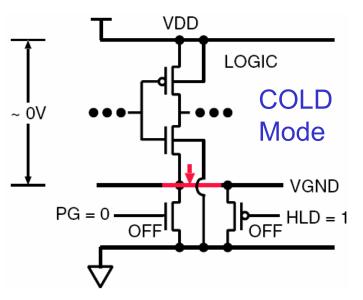
[Henzler, et al., ISSCC 2005]

Charge Recycling Scheme



Intermediate Power Switch State

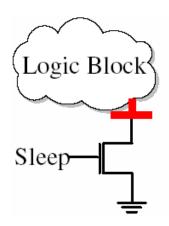




- PMOS power switch enables the PARK mode with $V_{\text{DD}}\text{-}V_{\text{TP}}$ across logic circuit
- Fixed and large V_{TP} step

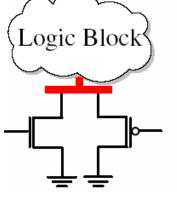
[Kim, et al., ISLPED 2004]

Fine Grained Multi-Threshold CMOS



- Virtual Ground rail has two levels
 - $\circ \sim V_{DD}$ (standby mode)
 - o ~Gnd (active mode)

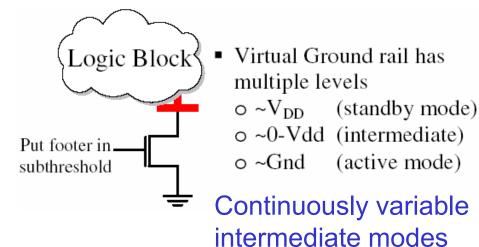
Classic MTCMOS



 Virtual Ground rail has only three levels

 ~V_{DD} (standby mode)
 ~V_{th_p} (intermediate)
 ~Gnd (active mode)

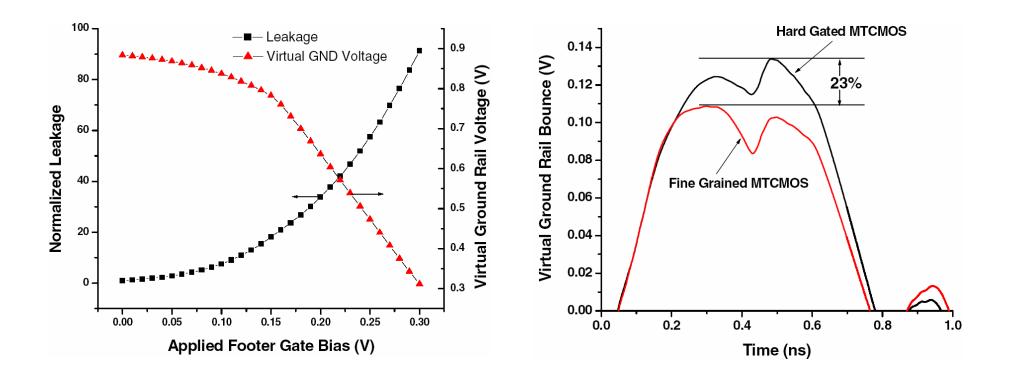
MTCMOS with intermediate mode



- Place the footer device in sub-threshold operation to achieve soft gating
- Tailor virtual ground voltage to maximize leakage savings

[Deogun, et al., A-SSCC 2005] Univ. of Michigan + IBM Research

Fine Grained MTCMOS Simulation Results



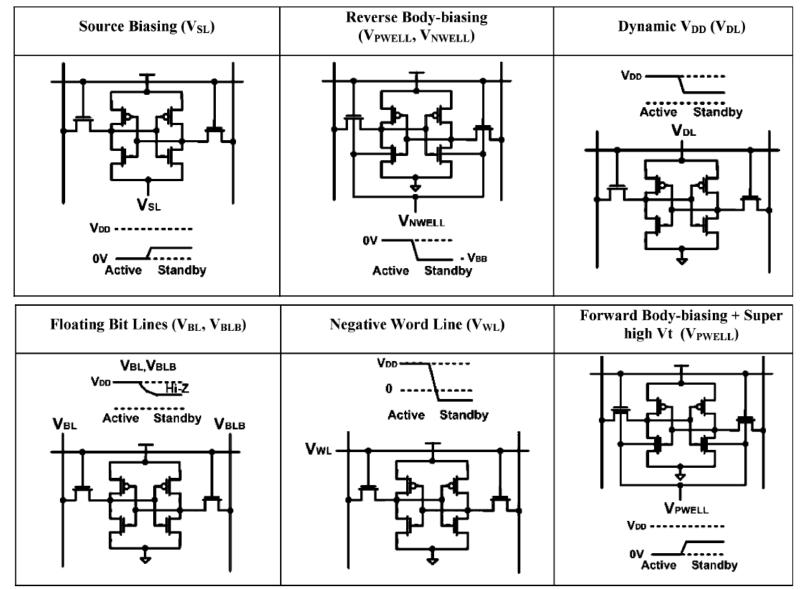
 Impact of applied footer V_{gs} on leakage and virtual ground rail potential

65nm SOI, 64-bit CLA adder, 0.9V, 85°C

Ground bounce reduced by 23%

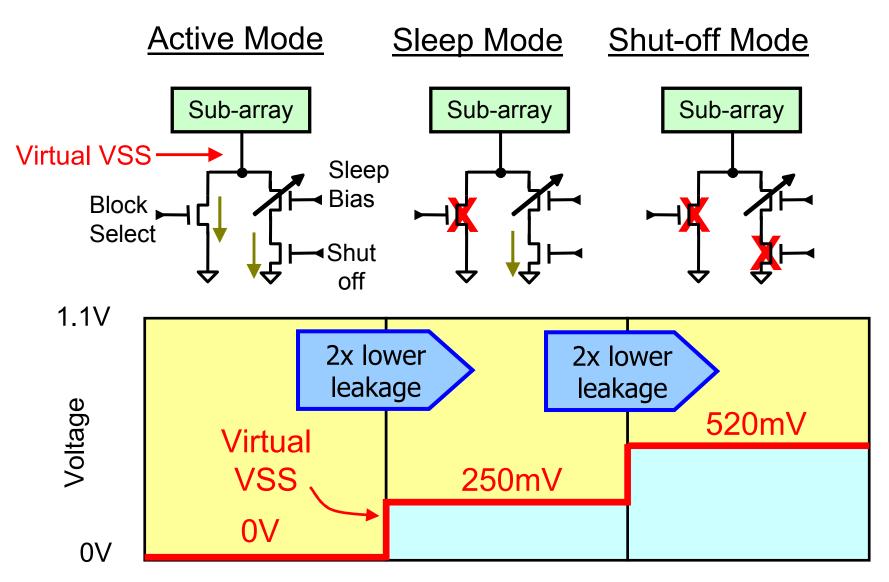
[Deogun, et al., A-SSCC 2005]

Cache Leakage Reduction Techniques



[Kim, et al., IEEE Trans. VLSI Sys., 2005]

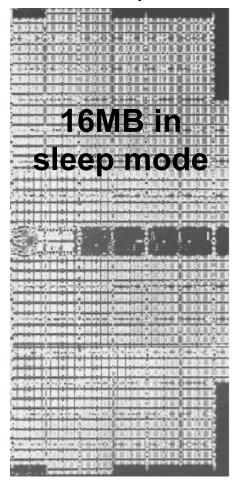
Cache Sleep and Shut-off Modes



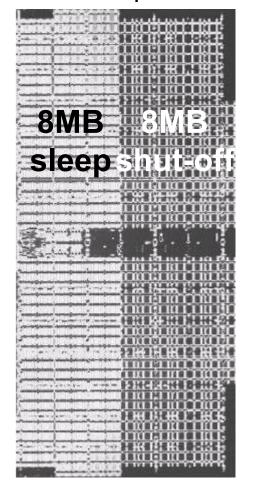
[Rusu, et al., ISSCC 2006]

Leakage Shut-off Infrared Images

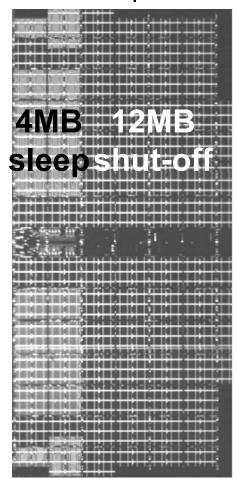
16MB part



8MB part



4MB part

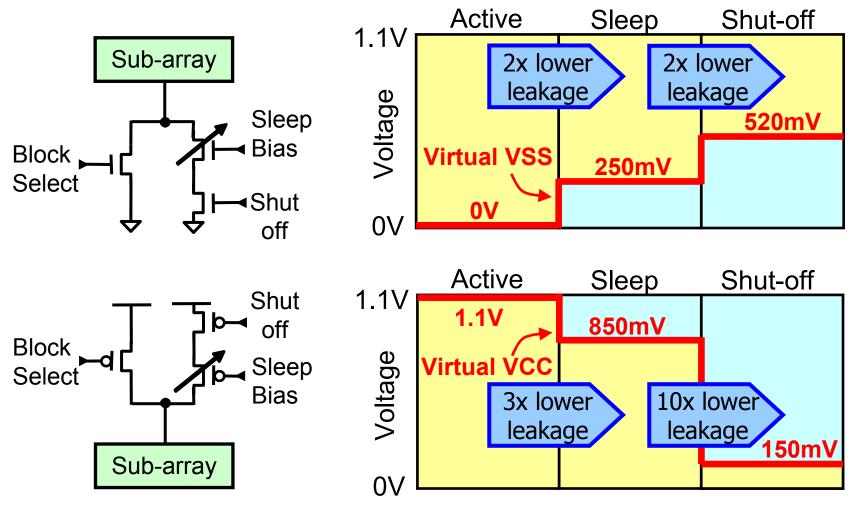


Leakage reduction ► 3W (8MB)

5W (4MB)

[Rusu, et al., ISSCC 2006]

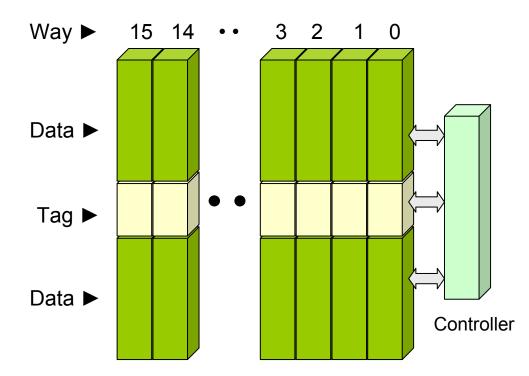
Shut-off Mode Scaling Trends



PMOS reduces junction leakage and has better shut-off

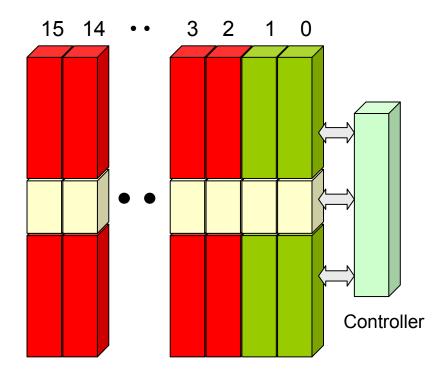
US Pat App 20070005999, 6/2005

Cache Dynamic Shut-off



Normal Operation

 In the full-load state, all 16 ways are enabled (green)

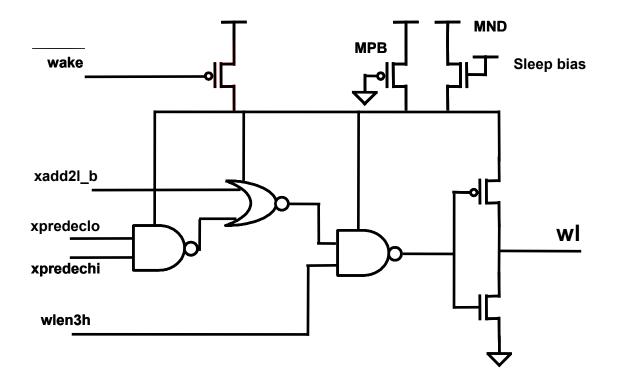


Cache-by-Demand Operation

 Under idle or low-load states, cache ways are dynamically flushed out and put in shut-off mode (red)

[Sakran, et al., ISSCC 2007]

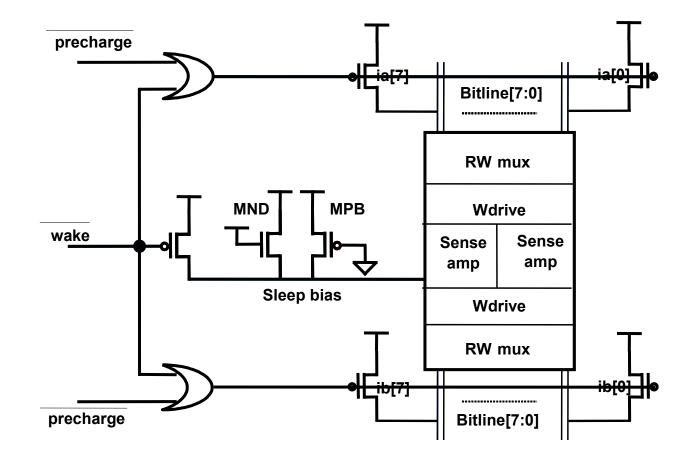
Cache Word-Line Sleep Transistor



- P-MOS sleep device for the word-line driver and final decoder
- N-MOS diode (MND) limits the virtual Vcc voltage drop to ensure word-lines have proper logic values during the sleep mode

[Chang, et al., VLSI Symposium, 2006]

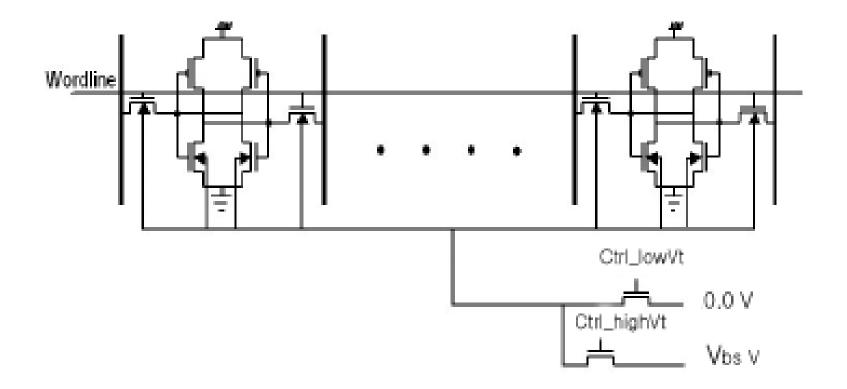
Cache I/O Sleep Transistor



• PMOS I/O sleep transistor cuts cache I/O leakage by about 3x

[Chang, et al., VLSI Symposium, 2006]

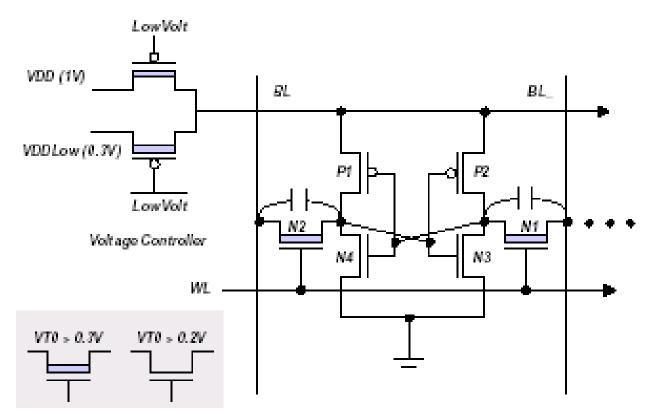
Dynamic V_{th} SRAM



 Uses body bias to raise V_{th} for inactive subarrays and lower the V_{th} for frequently accessed ones

[C. Kim, K. Roy – ISLPED 2002]

Drowsy Cache

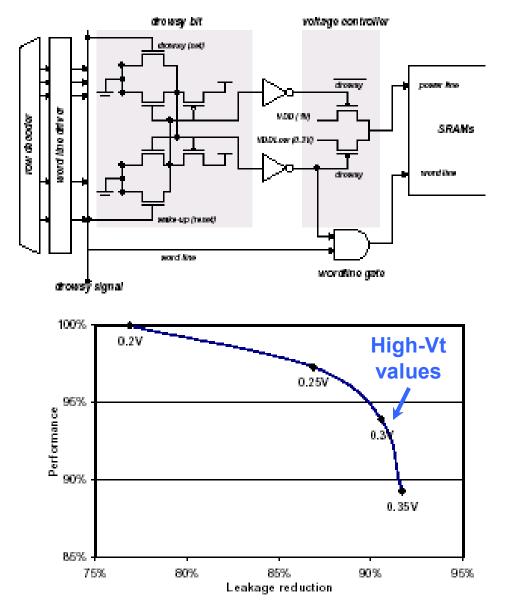


- Faster switching than body-bias scheme
- Increased susceptibility to soft error upsets

[K. Flautner, et al – ISCA 2002]

Drowsy Cache (cont)

- Drowsy cache line implementation requires:
 - Drowsy bit
 - Voltage controller
 - Word-line gating circuit

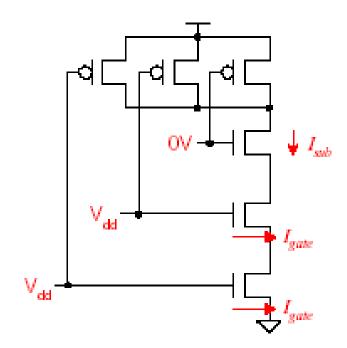


 High-Vt threshold selection is a trade-off between performance and leakage reduction

[K. Flautner, et al – ISCA 2002]

Pin Reordering

- Key difference between the state dependence of ${\rm I}_{\rm sub}$ and ${\rm I}_{\rm gate}$ for combinatorial gates:
 - I_{sub} primarily depends on the number of OFF in stack
 - $-I_{qate}$ depends strongly on the position of ON/OFF transistors

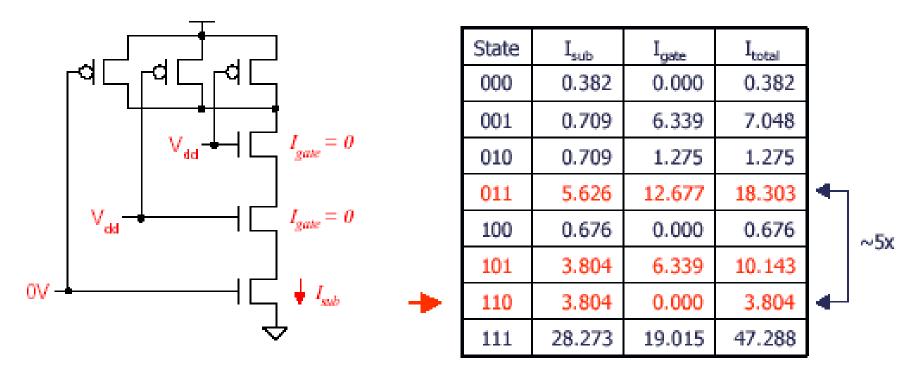


State	I_{sub}	Igate	I_{total}
000	0.382	0.000	0.382
001	0.709	6.339	7.048
010	0.709	1.275	1.275
011	5.626	12.677	18.303
100	0.676	0.000	0.676
101	3.804	6.339	10.143
110	3.804	0.000	3.804
111	28.273	19.015	47.288

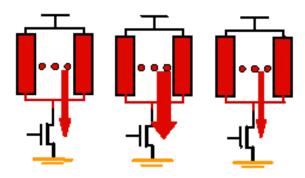
[D. Lee, et al - DAC 2003]

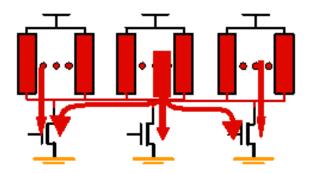
Pin Reordering (cont)

- Combine pin re-ordering and state assignment for standby (sleep) mode:
 - State assignment is utilized for reducing Isub
 - Pin re-ordering is targeted at Igate reduction: place off-transistor at the bottom of the stack



Distributed Sleep Transistors





Cluster-based design

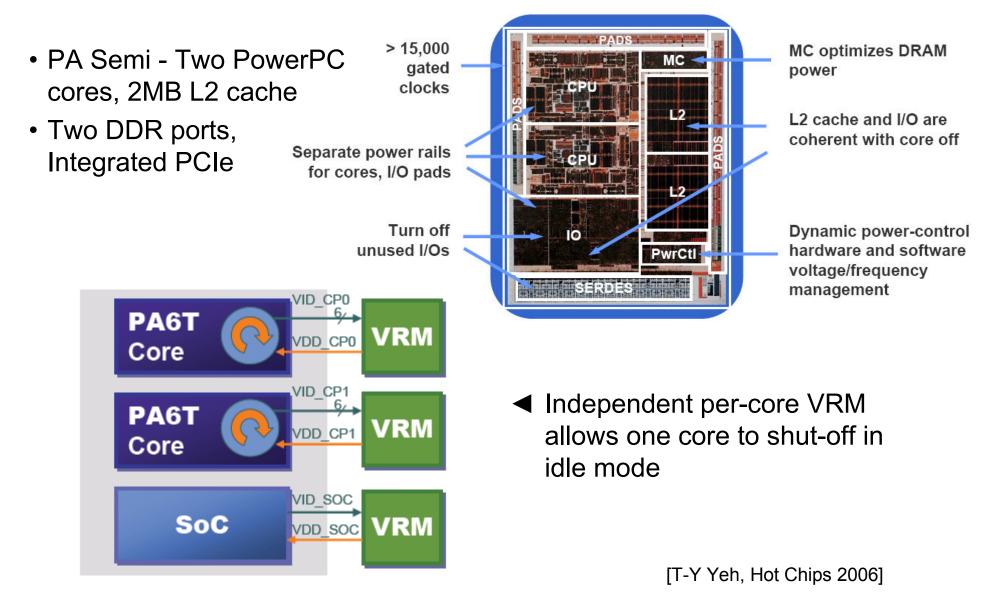
Distributed sleep transistor

Parameters	Without Sleep tr.	Cluster-based	DSTN
Leakage Current (nA)	59.80	5.72	1.23
Critical path delay (<i>nS</i>)	1.66	1.79	1.68
ST Area(µm²)	0	1449.6	212.2
Chip Area(µm²)	11960.0	13892.0	12880.0

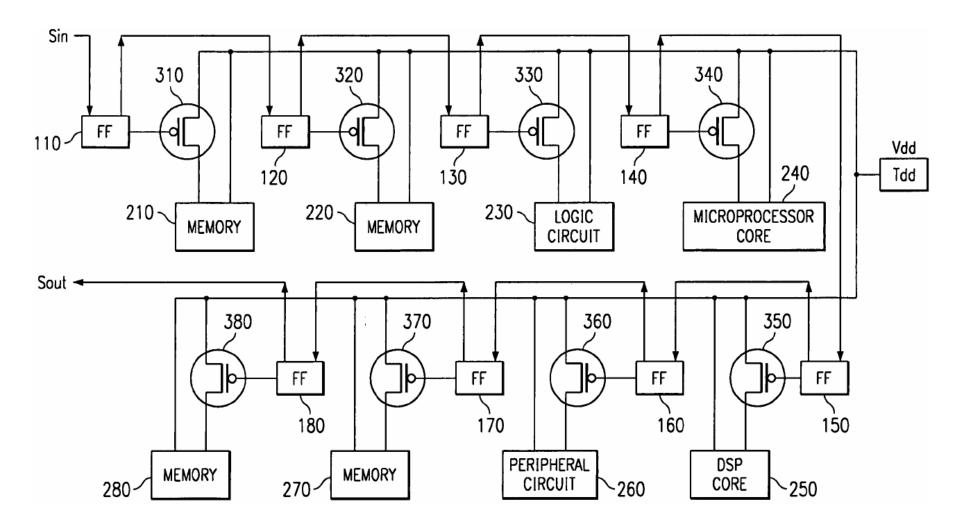
Distributed sleep transistor network (DSTN) reduces leakage by 50x vs. conventional design and 5x compared to cluster-based design

[Long, et al – DAC 2003]

Core-Level Voltage Control

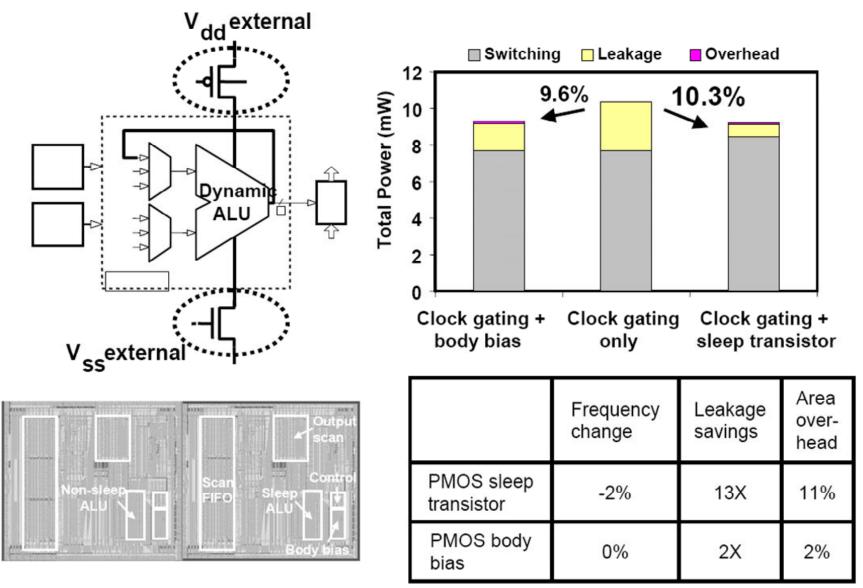


Power Switches Everywhere!



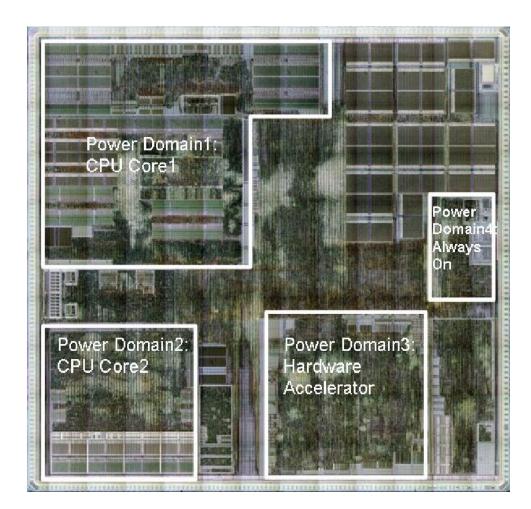
Texas Instruments, US Patent 6,864,708 "Suppressing the leakage current in an integrated circuit"

Sleep Transistor



[Tschanz, et al., ISSCC 2003]

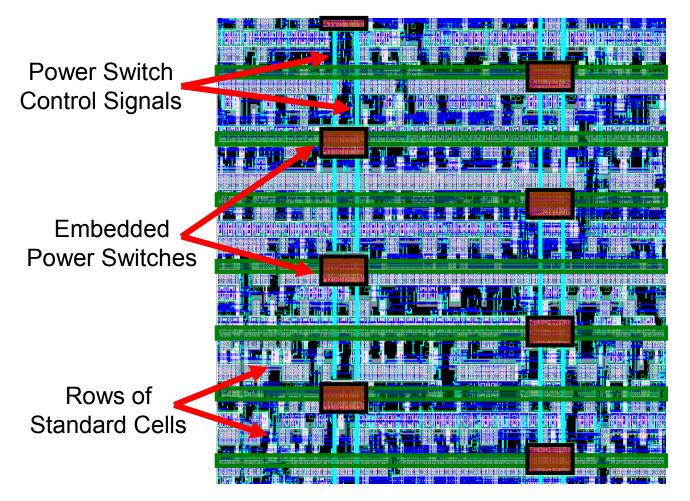
TI's 90nm OMAP Processor



- 90M transistors
- 90nm technology
- Single voltage supply
- Five power domains 1) MCU Core
 - 2) DSP Core
 - 3) Graphic Accelerator
 - 4) Always On logic
 - 5) Rest of chip

[Royannez, et al., ISSCC 2005] Texas Instruments

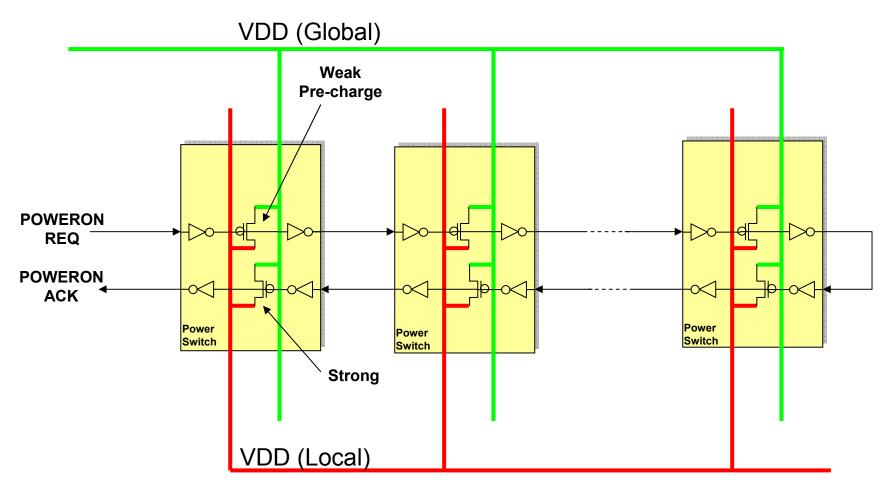
Embedded Power Switches



- Each power switch cell is a 90um PMOS
- A 1.3M gates power domain uses 4k switch cells

[Royannez, et al., ISSCC 2005]

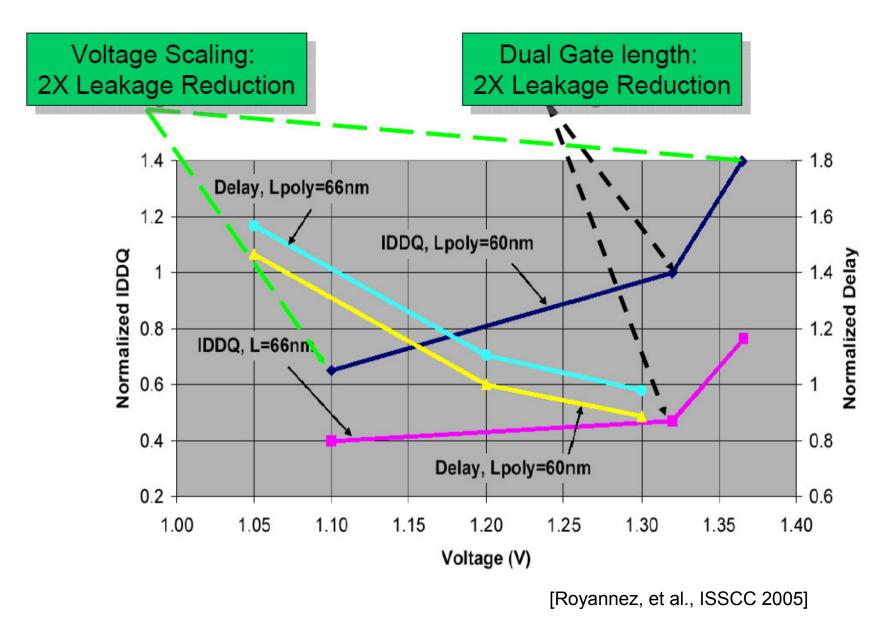
Power Gating Control



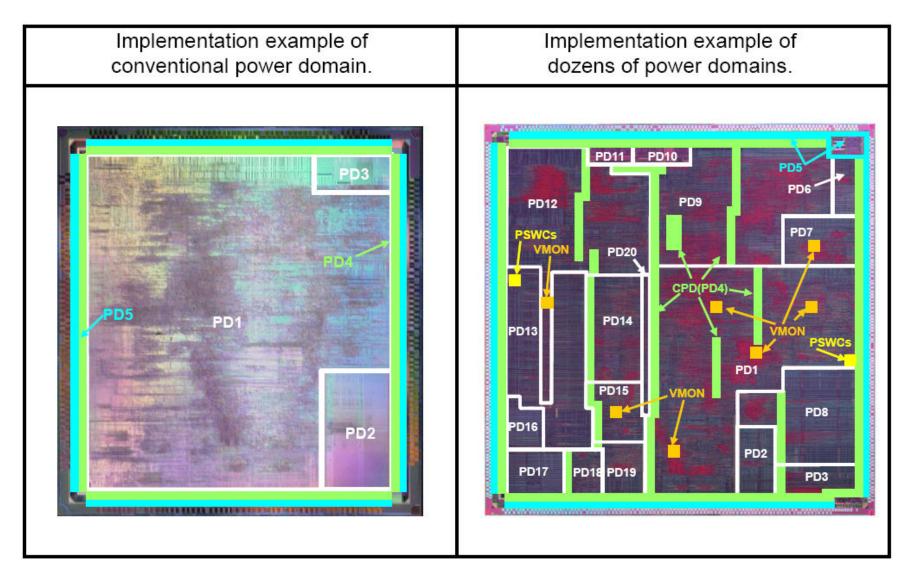
• Two-pass turn-on mechanism: weak PMOS for power restore and strong PMOS for normal operation

[Royannez, et al., ISSCC 2005]

Leakage Reduction Techniques

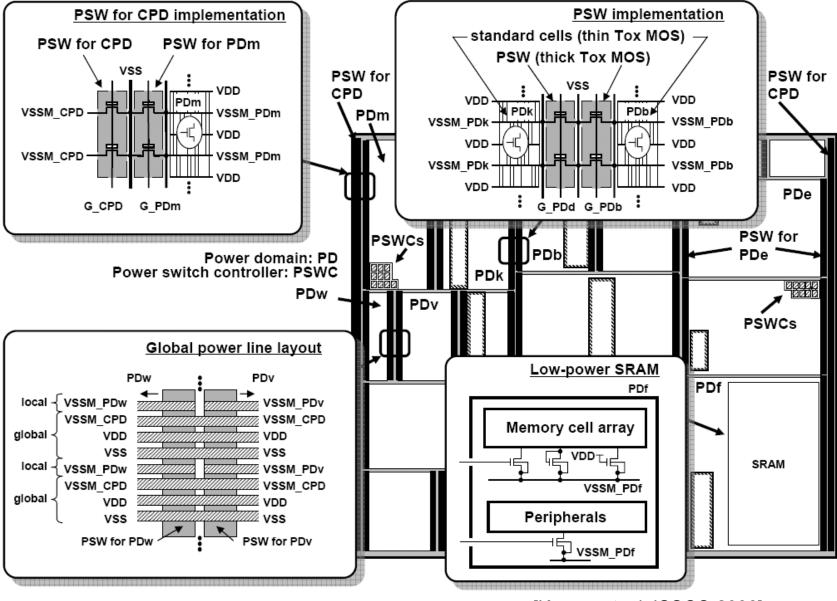


Multiple Power Domains



[Kanno, et. al, ISSCC-2006] Hitachi + Renesas

Power Switch Implementation



[Kanno, et. al, ISSCC-2006]

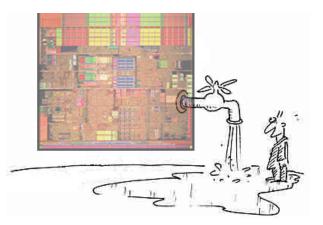
Power Domains Activation Examples



[Kanno, et. al, ISSCC-2006]

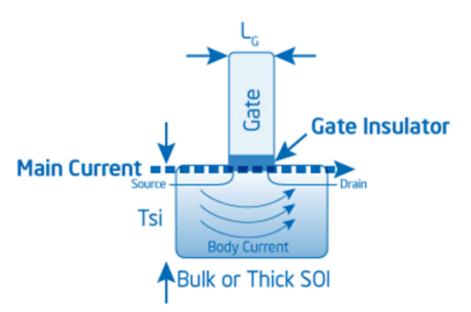
Outline

- Leakage mechanisms and trends
- Leakage reduction techniques
- Future process technology options
- Summary



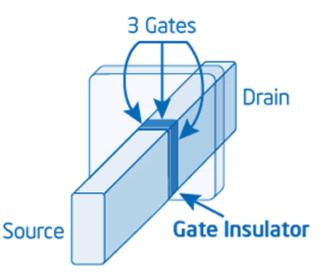
Tri-Gate Devices

Power Leakage on a Planar Transistor



 In planar devices the gate can only control the surface of the channel. Leakage paths, indicated by the semi-circular arrows, cause unwanted power consumption.

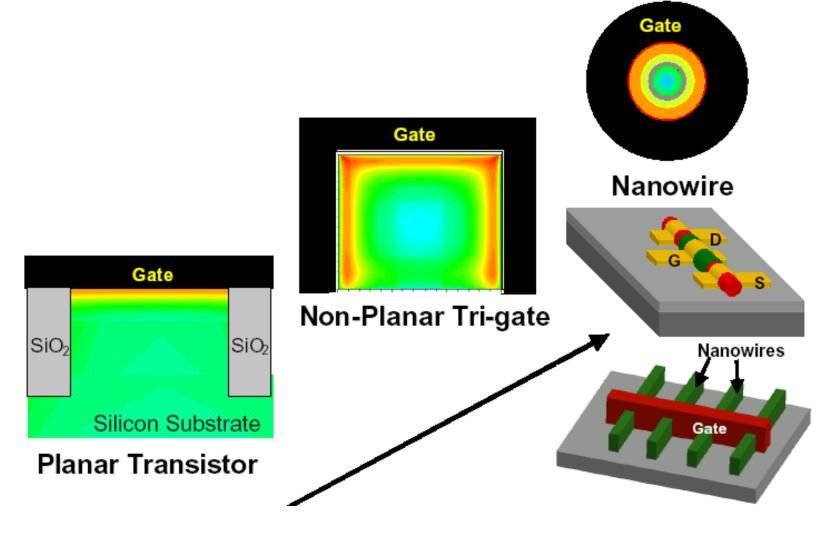
Tri-Gate: Surrounding the Channel



 An ideal transistor would have a gate surrounding a very thin channel of gate insulator. This gives the highest on-to-off current ratio and therefore the highest power efficiency.

http://www.intel.com/technology/silicon/tri-gate-demonstrated.htm

Device Structure Evolution



[R. Chau, et al - June 2003]

Summary

- Leakage will continue to grow due to device scaling, although at a lower pace
- There is no silver bullet for leakage reduction
- Low leakage design requires contributions from:
 - Transistor level strained silicon, high-K gate dielectrics, long-Lg devices, higher-Vt transistors, tri-gate devices
 - Block level sleep transistors, stack forcing
 - Chip level power switches, voltage islands, body bias
 - Platform level lower operating temperatures, multiple voltage power delivery
- Leakage reduction design techniques are becoming a way of life at all levels of chip design !

References

- S. Rusu, et. al, "A Dual-Core Multi-Threaded Xeon Processor with 16MB L3 Cache", ISSCC 2006, pp. 315 324
- J. Chang, et. al, "The 65nm 16MB On-Die L3 Cache for a Dual Core Multi-Threaded Xeon Processor" VLSI Symposium 2006, pp. 126 – 127
- T. Kuroda, "Optimization and control of VDD and VTH for low-power, high-speed CMOS design", ICCAD 2002, pp. 28 34
- S. Mukhopadhyay, K. Roy, "Accurate modeling of transistor stacks to effectively reduce total standby leakage in nano-scale CMOS circuits", VLSI Symposium 2003, pp. 53 56
- S. Geissler, et. al, "A low-power RISC Microprocessor using dual PLLs in a 0.13µm SOI Technology with Copper interconnect and low-k BEOL Dielectric", ISSCC 2002, pp. 148 – 149
- J. Clabes, et. al, "Design and implementation of the POWER5 microprocessor", ISSCC 2005, pp. 56 57
- S. Narendra, et. al, "Scaling of stack effect and its application for leakage reduction", ISLPED 2001, pp. 195 200
- S. Shigematsu, et. al, "A 1-V high-speed MTCMOS circuit scheme for power-down applications", VLSI Symposium 1995, pp. 125 – 126
- T. Kuroda, et. al, "A 0.9V, 150MHz, 10mW, 4mm², 2-D discrete cosine transform core processor with variable threshold-voltage (VT) scheme", IEEE Journal of Solid-State Circuits, Nov. 1996, pp. 1770 1779
- H. Kawaguchi, et. al, "A CMOS scheme for 0.5V supply voltage with pico-ampere standby current", ISSCC 1998, pp. 192 193
- T. Inukai, et. al, "Boosted Gate MOS: Device/circuit cooperation scheme to achieve leakage-free giga-scale integration", CICC 2000, pp. 409 412
- M. Johnson, et. al, "Leakage Control With Efficient Use of Transistor Stacks in Single Threshold CMOS", IEEE Transactions On VLSI, 2002
- K. Min, et. al, "Zigzag super cut-off CMOS (ZSCCMOS) block activation with self-adaptive voltage level controller: an alternative to clock-gating scheme in leakage dominant era", ISSCC 2003, pp. 400 502
- M. Drazdziulis, et. al, "A power cut-off technique for gate leakage suppression", ESSCIRC 2004, pp. 171 174
- R. Krishnamurthy, et. al, "High-performance and low-voltage challenges for sub-45nm microprocessor circuits", International conference on ASIC (ASICON), 2005, pp. 283 286

References (cont)

- K. Choi, et. al, "Optimal zigzag (OZ): an effective yet feasible power-gating scheme achieving two orders of magnitude lower standby leakage", VLSI Symposium 2005, pp. 312 315
- S. Henzler, et. al, "Sleep transistor circuits for fine-grained power switch-off with short power-down times", ISSCC 2005, pp. 302 600
- S. Kim, et. al, "Experimental measurement of a novel power gating structure with intermediate power saving mode", ISLPED 2004, pp. 20 25
- C. Kim, et. al, "A forward body-biased low-leakage SRAM cache: device, circuit and architecture considerations", IEEE Trans. On VLSI Systems, 3/2005, pp. 349 357
- N. Sakran, et. al, "The Implementation of the 65nm Dual-Core 64b Merom Processor", ISSCC 2007, pp. 106 590
- C. Kim, et. al, "Dynamic V_t SRAM: a leakage tolerant cache memory for low voltage microprocessors", ISLPED 2002, pp. 251 254
- K. Flautner, et. al, "Drowsy caches: simple techniques for reducing leakage power", ISCA 2002, pp. 148 157
- D. Lee, et. al, "Analysis and minimization techniques for total leakage considering gate oxide leakage", DAC 2003, pp. 175 180
- C. Long, et. al, "Distributed sleep transistor network for power reduction", DAC 2003, pp. 181 186
- T.-Y. Yeh, "The Low-Power High-Performance Architecture of the PWRficient Processor," Hot Chips 18 http://www.hotchips.org/archives/hc18/2_Mon/HC18.S2/HC18.S2T1.pdf
- Y. Kanno, et. al, "Hierarchical Power Distribution with 20 Power Domains in 90-nm Low-Power Multi-CPU Processor", ISSCC 2006, pp. 2200 2209
- M. Miyazaki, et. al, "A 1000-MIPS/W microprocessor using speed-adaptive threshold-voltage CMOS with forward bias", ISSCC 2000, pp. 420 421, 475
- R. Chau, et. al, "Silicon nano-transistors and breaking the 10nm physical gate length barrier", 61st Device Research Conference, June 2003, pp. 123 – 126
- Y. Wang, et. al, "A 1.1GHz 12uA/Mb-leakage SRAM design in 65nm Ultra-Low-Power CMOS Technology with Integrated Leakage Reduction for Mobile Applications", ISSCC 2007
- A. Keshavarzi, et. al, "Leakage and process variation effects in current testing on future CMOS circuits", IEEE Design & Test of Computers, Sept.-Oct. 2002, pp. 36 43
- J. Tschanz, et. al, "Dynamic-sleep transistor and body bias for active leakage power control of microprocessors", ISSCC 2003, pp. 102 103
- K. Mistry, et. al, "A 45nm Logic Technology with High-k + Metal Gate Transistors, Strained Silicon, 9 Cu Interconnect Layers, 193nm Dry Patterning, and 100% Pb-free Packaging", IEDM 2007, pp. 247 251