LECTURE 11: CIRCUIT DESIGN FOR LOW POWER

Alan Drake, TSMC
Kevin Nowka, IBM Austin Research Laboratory

Agenda

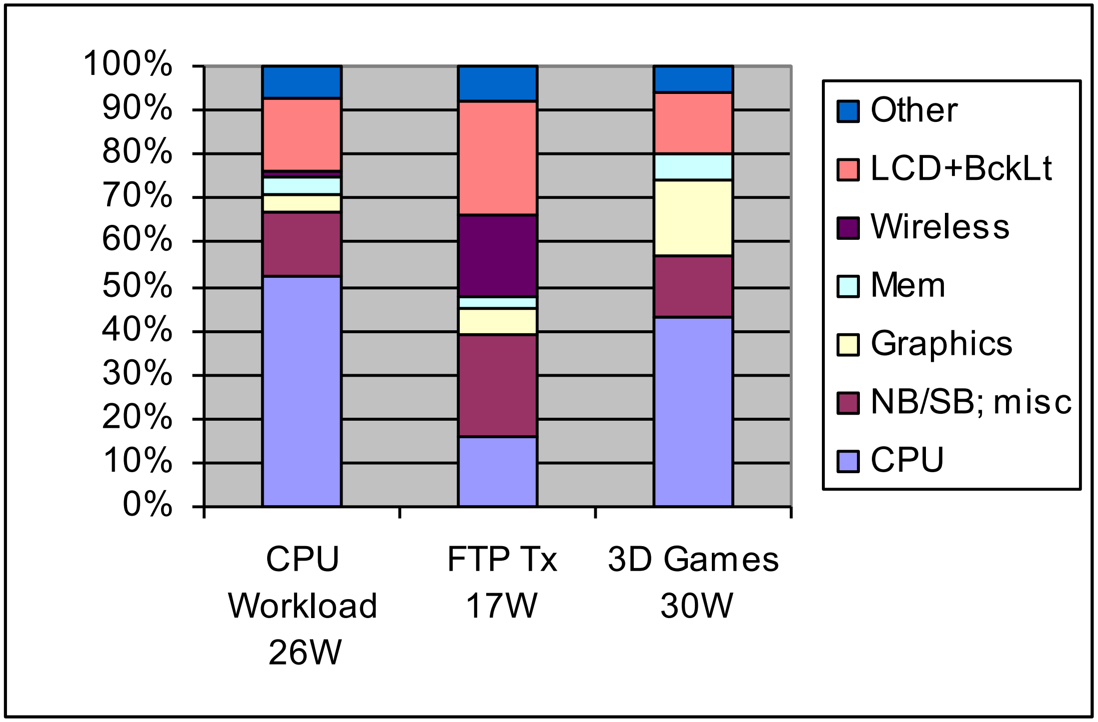
- Overview of VLSI power
- Technology, Scaling, and Power
- A look at the real trends and projections for the future
- Active power: components, trends, managing, & estimating
- Static power: components, trends, managing, & estimating
- Summary

iPhone Power

- Settings -> Battery -> Battery Usage
  - Shows percentage use of apps in past 24 hours to 7 days.

A quick look at the power consumption of a sort-of modern Laptop (IBM R40)

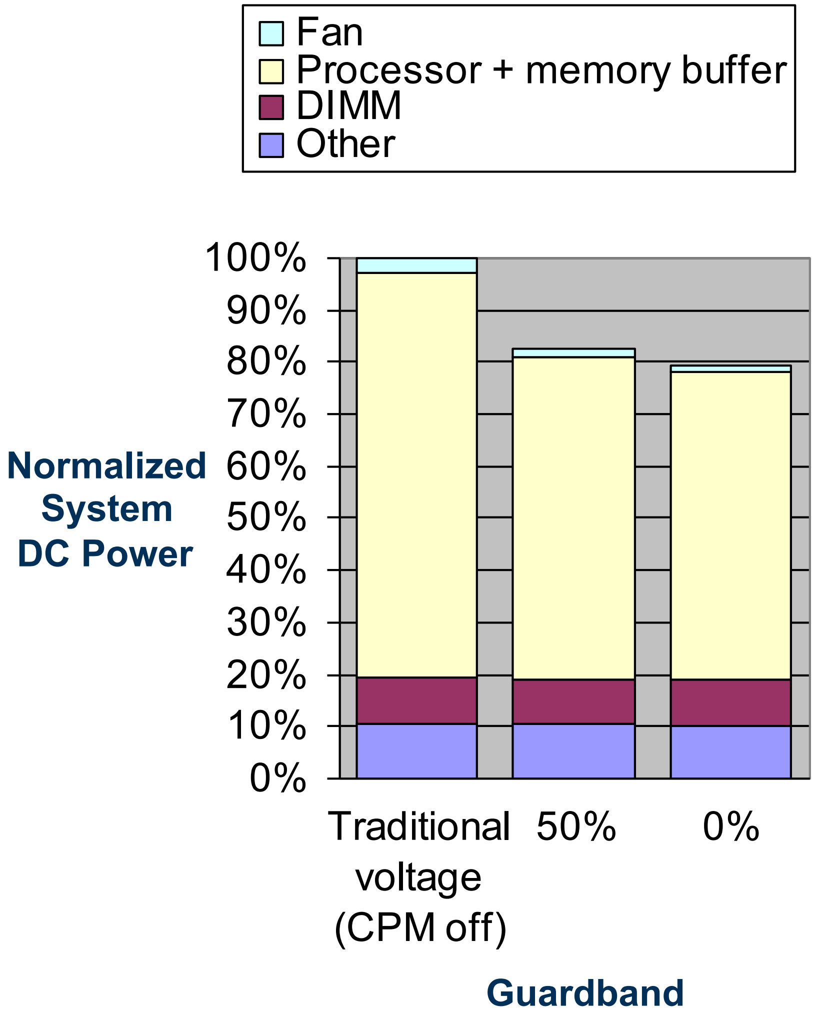
Power is all about the (digital) VLSI circuits…..and the backlight!



Src: Mahesri et al., U of Illinois, 2004

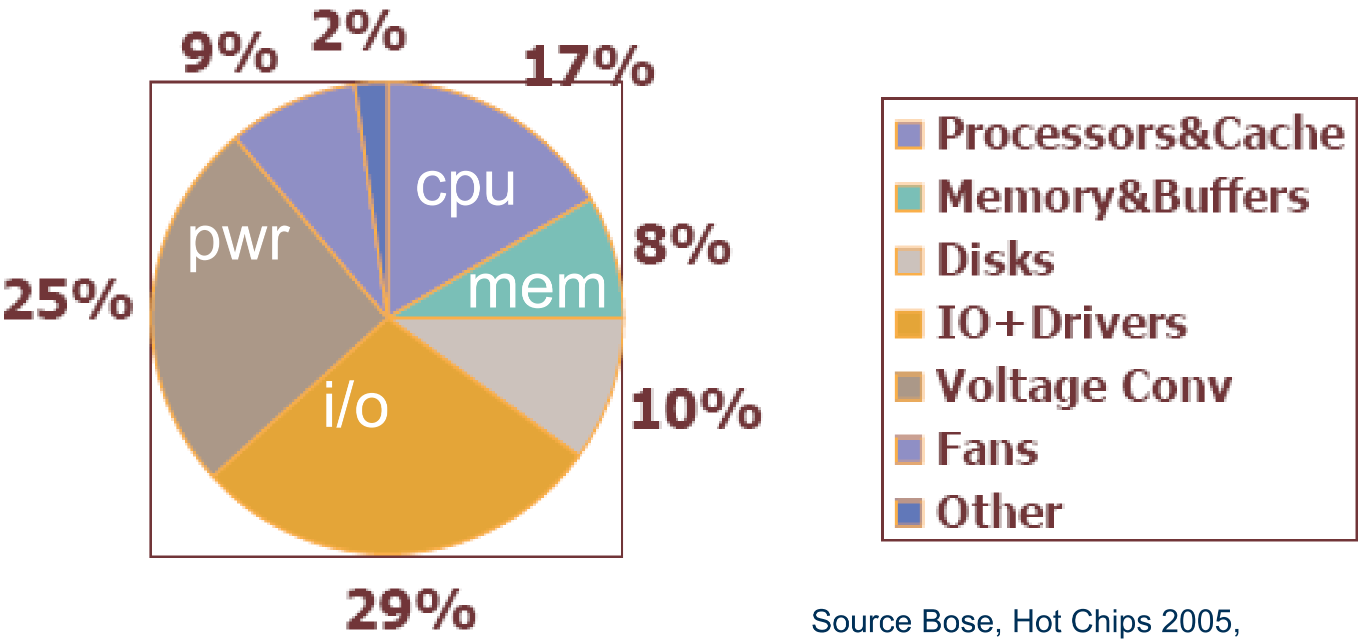
Power Consumption of an IBM Power 750 Express Server

- 4 POWER7 processors, 64 GB
  - Run SPEC CPU 2006 workloads
  - Frequency target = 3864 MHz (Turbo)
- Guardband reduced to 50%
  - 20% chip power reduction
    Voltage reduced 113 mV – 140 mV
  - 18% system power reduction
  - 50% fan power reduction
  - No change in performance
- Upper bound on power reduction (0% guardband)
  - 24% chip power reduction
  - 21% system power reduction



A quick look at the power consumption of a Server

Again, it’s a VLSI problem – but this time with analog!



Source Bose, Hot Chips 2005,

Designing within limits: power & energy

- Thermal limits (for most parts self-heating is a substantial thermal issue)
  - Package cost (4-5W limit for cheap plastic package, 50-100W/sq-cm air cooled limit, 5k-7.5kW 19” rack)
  - Device reliability (junction temp > 125C quickly reduces reliability)
  - Performance (25C -> 105C loss of 30% of performance)
- Distribution limits
  - Substantial portion of wiring resource, area for power dist.
  - Higher current => lower R, greater dI/dt => more wire, decap
  - Package capable of low impedance distribution
- Energy capacity limits
  - AA battery ~1000mA.hr => limits power, function, or lifetime
  - Iphone 7+ has a 2900mAh battery (estimated)
- Energy cost
  - Energy for IT equipment large fraction of total cost of ownership

Agenda

- Overview of VLSI power
- Technology, Scaling, and Power
- A look at the real trends and projections for the future
- Active power: components, trends, managing, & estimating
- Static power: components, trends, managing, & estimating
- Summary

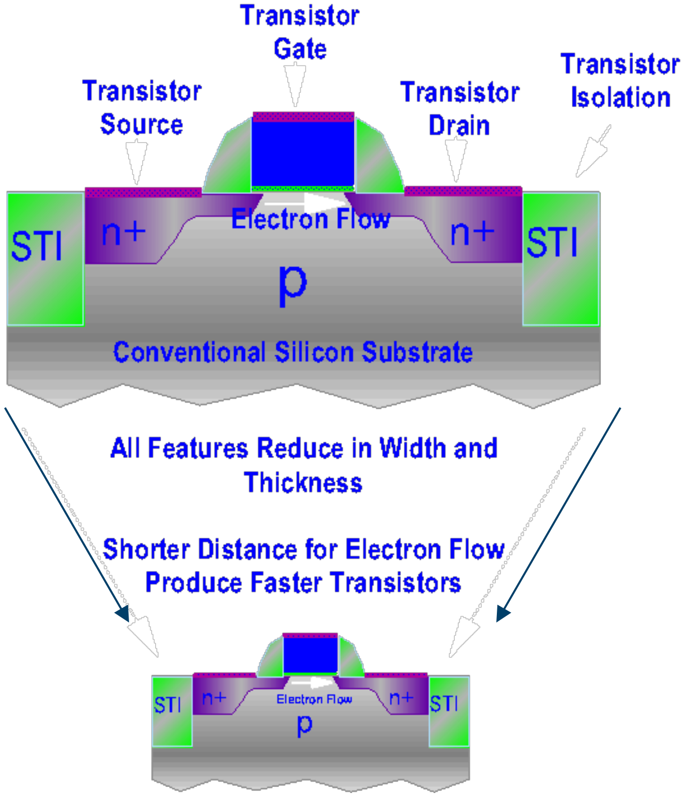
CMOS circuit power consumption components

- $P = ½\,C_{sw}V_{dd}\,\Delta V\,f + I_{st}V_{dd} + I_{static}V_{dd}$
- Dynamic power consumption ($½\,C_{sw}V_{dd}\,\Delta V\,f + I_{st}V_{dd}$)
  - Load switching (including parasitic & interconnect)
  - Glitching
  - Shoot through power ($I_{st}V_{dd}$)
- Static power consumption ($I_{static}V_{dd}$)
  - Current sources – bias currents
  - Current dependent logic -- NMOS, pseudo-NMOS, CML
  - Junction currents
  - Subthreshold MOS currents
  - Gate tunneling

Review of Constant Field Scaling

These are distributions … how do the σ s scale?

Parameter	Value	Scaled Value
Dimensions	L, W, Tox	αL, αW, αTox
Dopant concentrations	Na, Nd	Na/α, Nd/α
Voltage	V	αV
Field	E	E
Capacitance	C	αC
Current	I	αI
Propagation time (~CV/I)	t	αt
Power (VI)	P	α^2P
Density	d	d/α^2
Power density	P/A	P/A



Scale factor α<1

Agenda

- Overview of VLSI power
- Technology, Scaling, and Power
- A look at the real trends and projections for the future
- Active power: components, trends, managing, & estimating
- Static power: components, trends, managing, & estimating
- Summary

CMOS Circuit Delay and Frequency

$$P = \frac{1}{2} C_{sw} V_{dd} \Delta V F + I_{st} V_{dd} + I_{static} V_{dd}$$

VLSI system frequency determined by:

Sum of propagation delays across gates in "critical path" --

Each gate delay, includes time to charge/discharge load thru one or more FETs and interconnect delay to distribute the signal to next gate input.

$$\boxed{\begin{aligned} T_d &= kCV/I \\ &= kCV/(V_{dd}-V_t)^{\alpha} \end{aligned}}$$

Sakuri α-power law model of delay

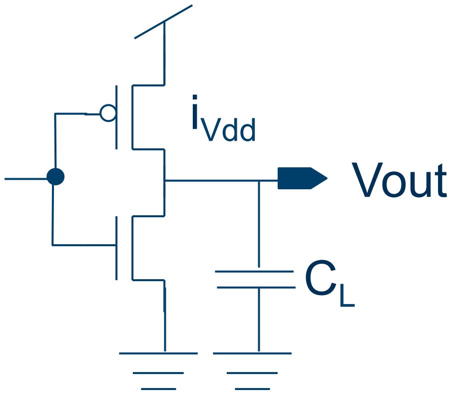
Dynamic Energy

$$E_{Vdd} = \int_{t=0}^{\infty} i_{Vdd}(t) V_{dd} dt = V_{dd} \int_{0}^{\infty} C_L \frac{dV_{out}}{dt} dt$$

$$E_{Vdd} = C_L V_{dd} \int_{Vout=0}^{Vdd} dV_{out} = C_L V_{dd}^2$$

$$Ec = \int_{t=0}^{\infty} i_{CL}(t) V_{out} dt = \int_{0}^{\infty} C_L \frac{dV_{out}}{dt} V_{out} dt$$

$$Ec = C_L \int_{Vout=0}^{Vdd} V_{out} dV_{out} = \frac{1}{2} C_L V_{dd}^2$$



Energy dissipated for either output transition consumes:

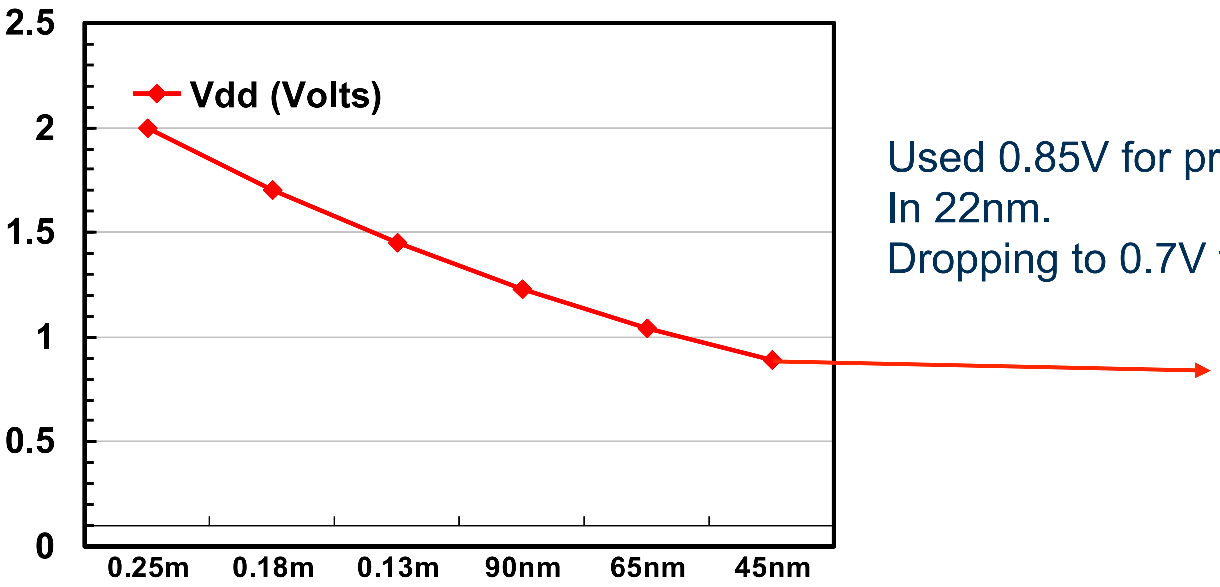
$\frac{1}{2} C_L V_{dd}^2$

P = $\frac{1}{2} C_L V_{dd} \Delta V f$ + IstVdd + IstaticVdd

Gate level energy consumption should improve as α^3 under constant field scaling, but….

Supply Voltage Trend

$$P = \frac{1}{2} C_L V_{dd} \Delta V f + IstVdd + IstaticVdd$$

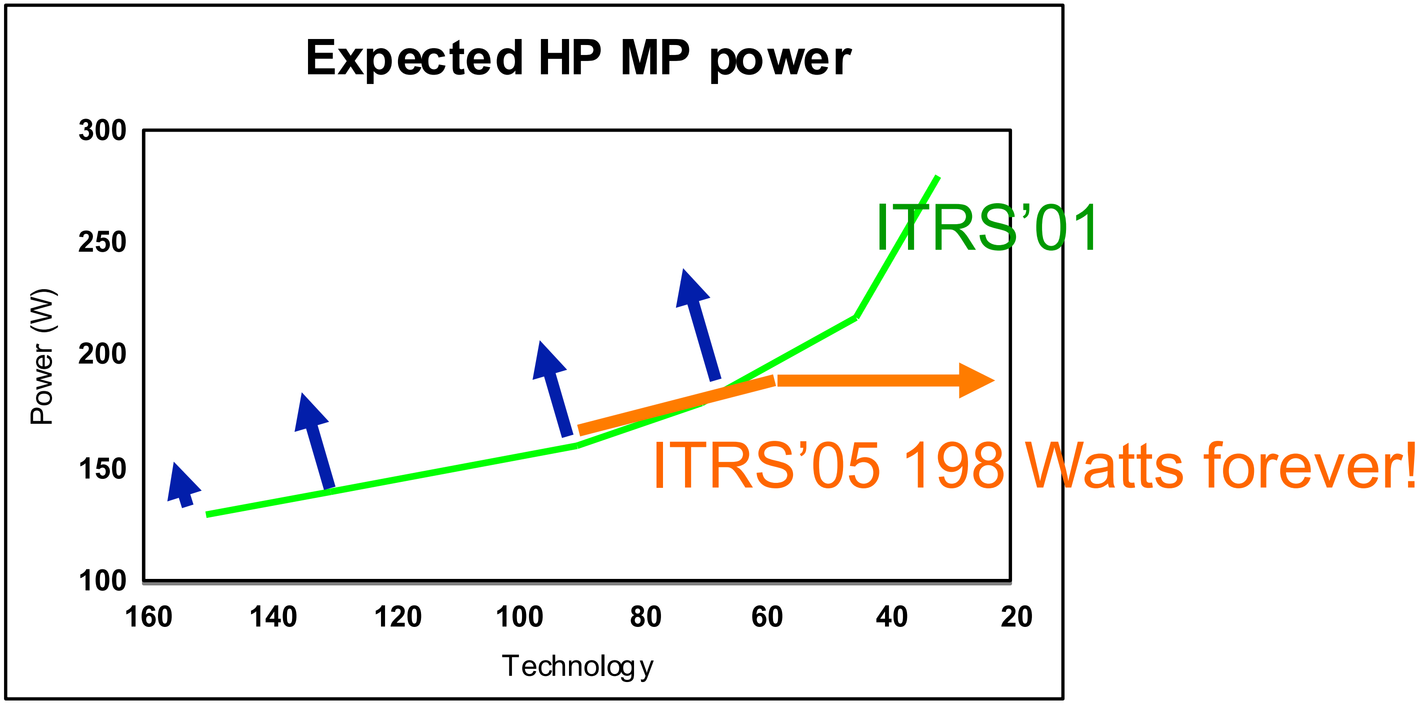


Used 0.85V for processors
In 22nm.
Dropping to 0.7V for 10nm+

With each generation, voltage has decreased 0.85x, not 0.7x for constant field.
Thus, energy/device is decreasing by 50% rather than 65%

Active Power Trend

$$P = \tfrac{1}{2} C_L V_{dd} \Delta V f + IstVdd + IstaticVdd$$



But, number of transistors has been increasing, thus

- a net increase in energy consumption,
- with freq 2x, active power is increasing by 50%

(src: ITRS '01-'05)

Intel Processor Roadmap, Extreme Desktop Processor

Codename	Technology	Cores	TDP (W)	Frequency (GHz)	
				Nom	Turbo
Kentsfield	65n	4	130	3	
Yorktown	45n	4	130	3.2	
Nehalem	45n	4	130	3.33	
Westmere	32n	6	130	3.46	
Sandy Bridge	32n	4	95	3.5	3.9
Ivy Bridge	22n	8	77	3.5	3.9
Haswell	22n	8	88	4	4.4
Broadwell	14n	4	65	3.3	3.7

Active-Power Reduction Techniques

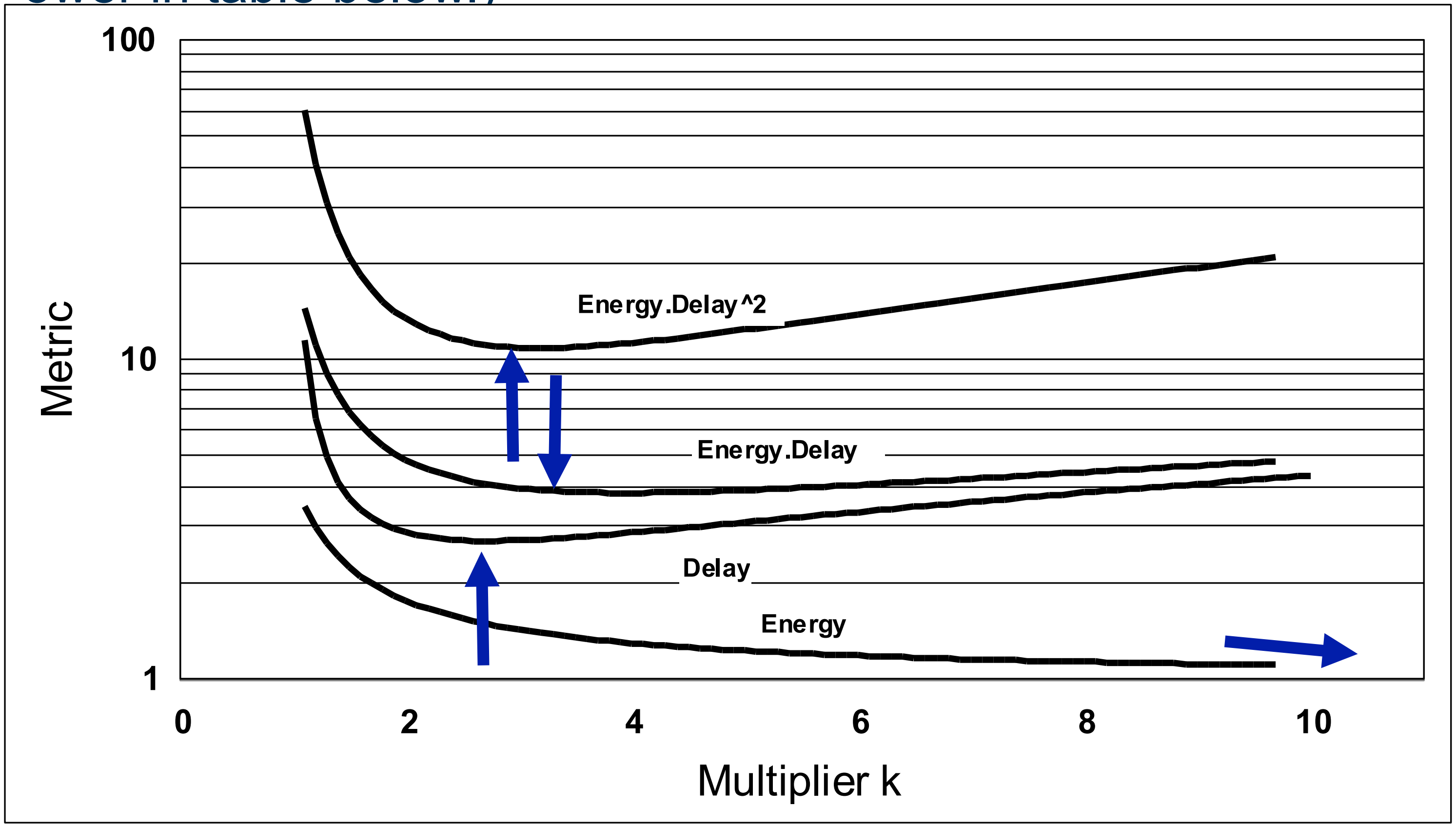
- $P = \frac{1}{2} C_L V_{dd} \Delta V f + I_{st} V_{dd} + I_{static} V_{dd}$
- Active power can be reduced through:
  - Capacitance minimization
    - Power/Performance in sizing
    - Clock-gating
    - Glitch suppression
    - Hardware-accelerators
    - System-on-a-chip integration
  - Voltage minimization
    - (Dynamic) voltage-scaling
    - Low swing signaling
    - SOC/Accelerators
  - Frequency minimization
    - (Dynamic) frequency-scaling
    - SOC/Accelerators

Capacitance minimization

- $P = ½\, C_{SW} V_{dd}\, \Delta V\, f + I_{st} V_{dd} + I_{static} V_{dd}$
- Only the devices (device width) used in the design consume active power!
  - Runs counter to the complexity-for-IPC trend
  - Runs counter to the SOC trend

Capacitance minimization

- Device sizing for power efficiency is significantly different than sizing for performance – eg. sizing of the gate size multiplier in an exponential-horn of inverters for driving large loads. (Energy should be replaced by Power in table below.)



Discussion on PDP, EDP: Gonzalez and Horowitz, JSSC, Sept. 1996, pp. 1277-1284

Functional Clock Gating

- $P = \frac{1}{2} C_{SW} V_{dd} \Delta V f + I_{st} V_{dd} + I_{static} V_{dd}$
- 25-50% of power consumption due to driving latches (Bose, Martinozi, Brooks 2001 50%)
- Utilization of most latches is low (~10-35%)
- Gate off unused latches and associated logic:
  - Unit level clock gating – turn off clocks to FPU, MMX, Shifter, L/S unit, … at clk buffer or splitter
  - Functional clock gating – turn off clocks to individual latch banks – forwarding latch, shift-amount register, overflow logic & latches, … qualify (AND) clock to latch
- Asynchronous is the most aggressive gating – but is it efficient?

Glitch suppression

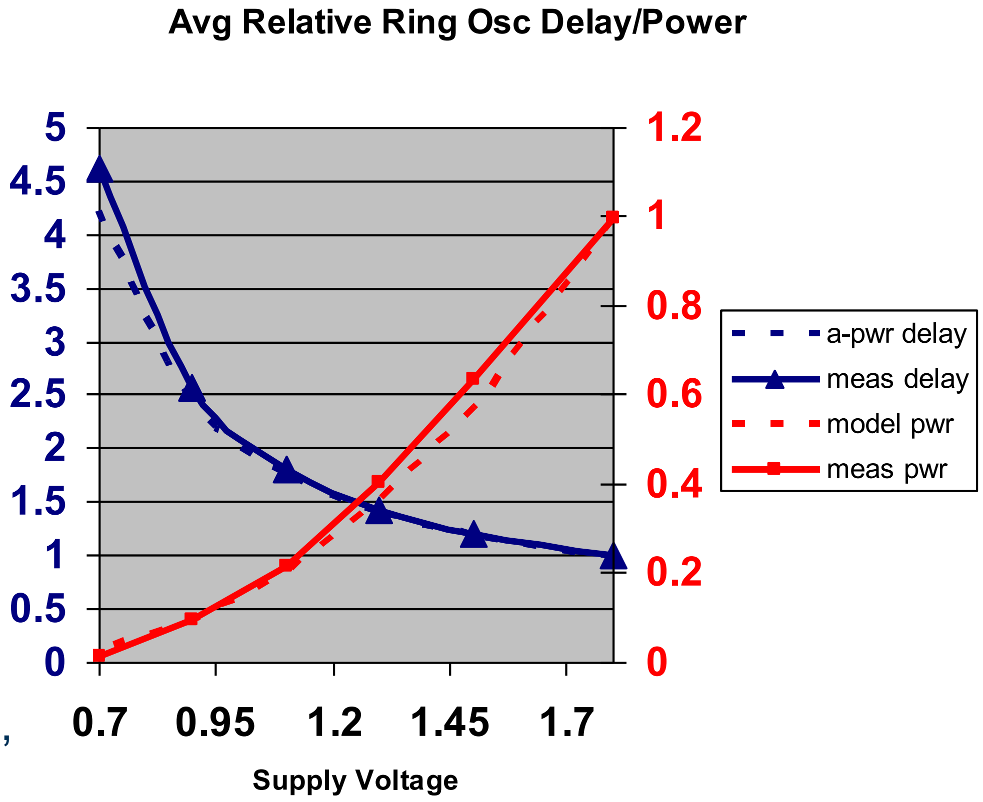
- $P = \frac{1}{2} C_{SW} V_{dd} \Delta V f + I_{st} V_{dd} + I_{static} V_{dd}$
- Glitches can represent a sizeable portion of active power, (up to 30% for some circuits in some studies)
- Three basic mechanisms for avoidance:
  - Use non-glitching logic, e.g. domino (drives up clock power)
  - Add redundant logic to avoid glitching hazards
    - Increases cap, testability problems
  - Adjust delays in the design to avoid
    - Shouldn't timing tools do this already if it is possible?

Voltage minimization

- $P = \frac{1}{2} C_{SW} V_{dd} \Delta V f + I_{st} V_{dd} + I_{static} V_{dd}$
- Lowering voltage swing, ΔV, lowers power
  - Low swing logic efforts have not been very successful (unless you consider array voltage sensing)
  - Low swing busses have been quite successful
- Lowering supply, Vdd and ΔV, (voltage scaling) is most promising:
  - Frequency ~V, Power $\sim V^3$

Voltage Scaling Reduces Active Power

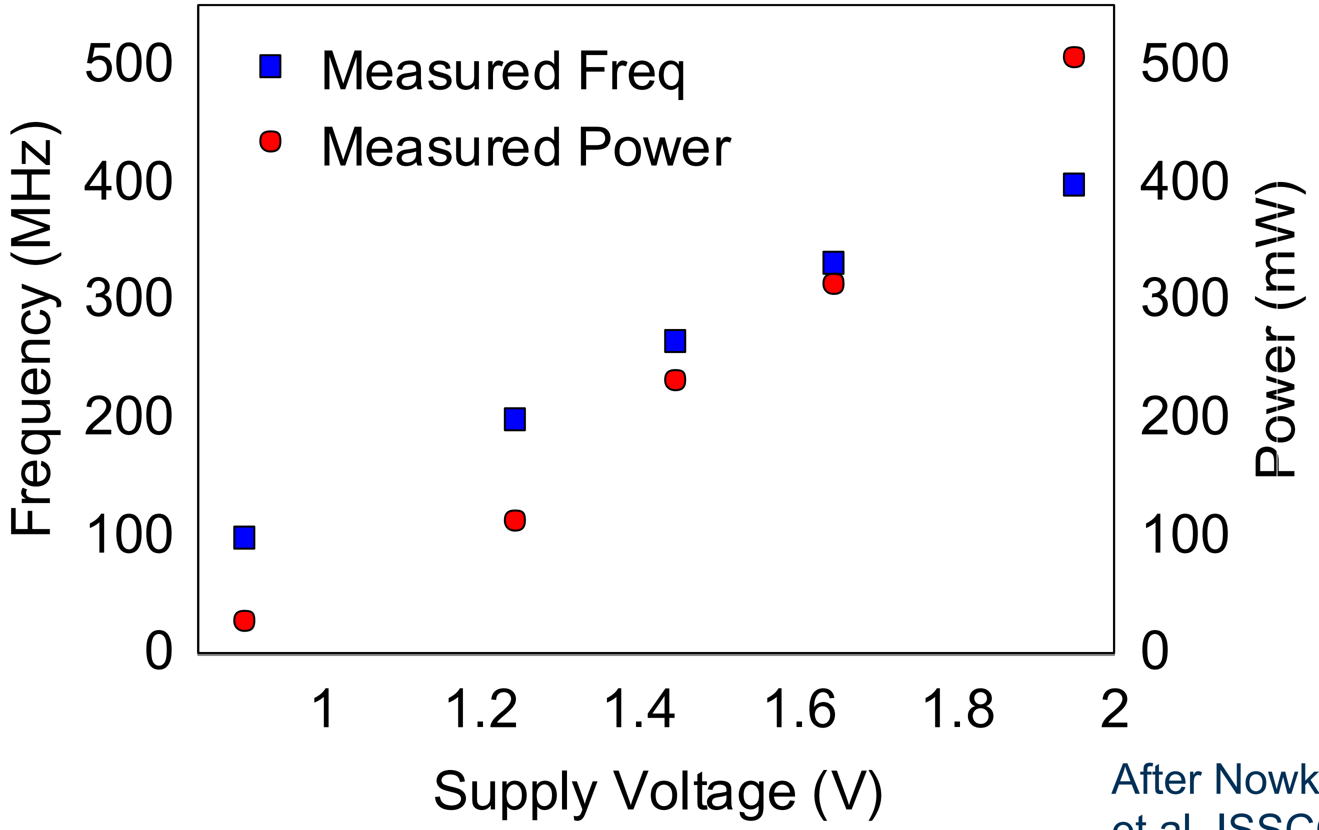
- Voltage Scaling Benefits
  - Can be used widely over entire chip
  - Complementary CMOS scales well over a wide voltage range => Can optimize power/performance (MIPS/mW) over a wide range
- Voltage Scaling Challenges
  - Custom CPUs, Analog, PLLs, and I/O drivers don't voltage scale easily
  - Sensitivity to supply voltage varies circuit to circuit – esp SRAM, buffers, NAND4
  - Thresholds tend to be too high at low supply
  - Variability becomes more prominent when scaling voltage.



After Carpenter, Microprocessor forum, '01

Dynamic Voltage-Scaling (e.g. XScale, PPC405LP)

PowerPC 405LP measurements: 18:1 power range over 4:1 frequency range

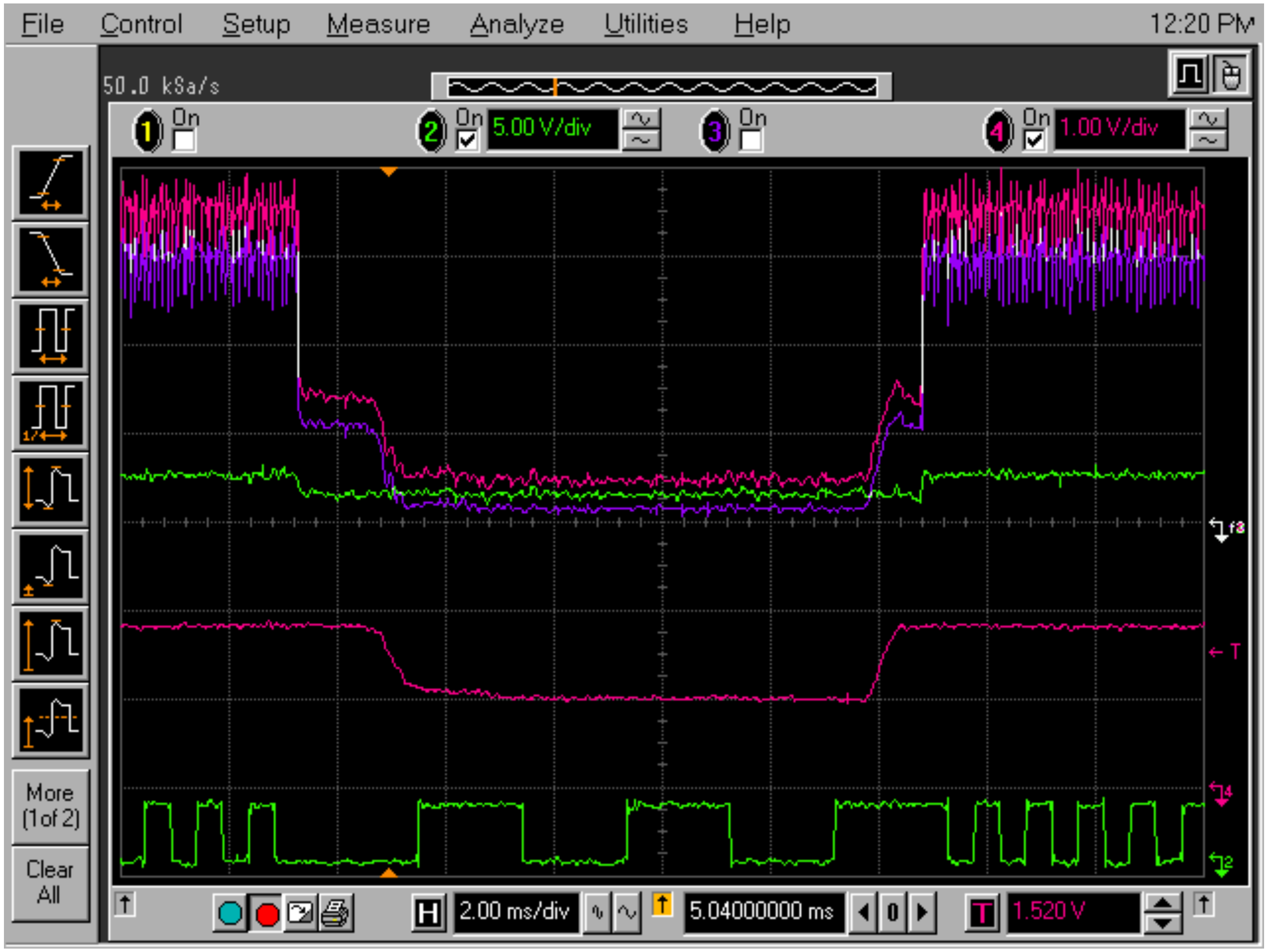


After Nowka, et.al. ISSCC, Feb ‘02

Frequency minimization

- $P = \frac{1}{2} C_{SW} V_{dd} \Delta V f + I_{st} V_{dd} + I_{static} V_{dd}$
- Lowering frequency lowers power linearly
  - DOES NOT improve energy efficiency, just slows down energy consumption
  - Important for avoiding thermal problems

Voltage-Frequency-Scaling Measurements PowerPC 405LP



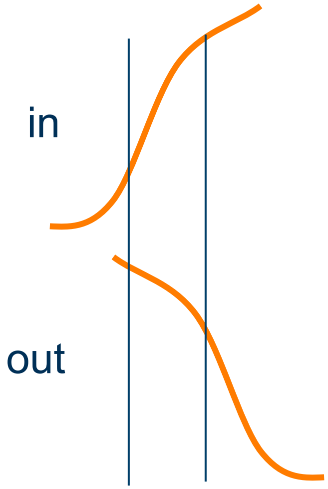
Freq Scaling

Plus DVS

Freq scale ¼ freq, ¼ pwr; DVS ¼ freq, 1/10 pwr

Src: After Nowka, et.al. JSSC, Nov '02

Shoot-through minimization



Both Pfet &
Nfet conducting

- $P = \frac{1}{2} C_{SW} V_{dd} \Delta V f + I_{st} V_{dd} + I_{static} V_{dd}$
- For most designs, shoot-thru represents 8-15% of active power.
- Avoidance and minimization:
  - Lower supply voltage
  - Domino?
  - Avoid slow input slews
  - Careful of level-shifters in multiple voltage domain designs
- What circuit can be used to eliminate shoot-through?
- Where is this practical?

Estimating Active Power Consumption

- $P = \frac{1}{2} C_{SW} V_{dd} \Delta V f + I_{st} V_{dd} + I_{static} V_{dd}$

- The problem is how to estimate capacitance switched
- Switch factor SF: $\frac{1}{2} C_{sw} = \Sigma_i SF_i Cnode_i$
  - Low level circuit analysis – spice analysis
  - Higher level: spreadsheet/back-of-the-envelope/power tools for estimation
    - Aggregate or node-by-node estimation of switch factors – 1.0 ungated clocks, 0.5 signals which switch every cycle, 0.1-0.2 for processor logic
    - These can be more accurately derived by tools which look at pattern dependence and timing
- Node Capacitance – sum of all cap: output driver parasitic, interconnect, load gate cap

Agenda

- Overview of VLSI power
- Technology, Scaling, and Power
- A look at the real trends and projections for the future
- Active power: components, trends, managing, & estimating
- Static power: components, trends, managing, & estimating
- Summary

Static Power

- $P = \frac{1}{2} C_{SW} V_{dd}\, \Delta V\, f + I_{st} V_{dd} + I_{static} V_{dd}$

- Static energy consumption ($I_{static} V_{dd}$)
  - Current sources – even µA bias currents can add up.
  - NMOS, pseudo-NMOS – not commonly used
  - CMOS CML logic – significant power for specialized use.
  - Junction currents
  - Subthreshold MOS currents
  - Gate tunneling

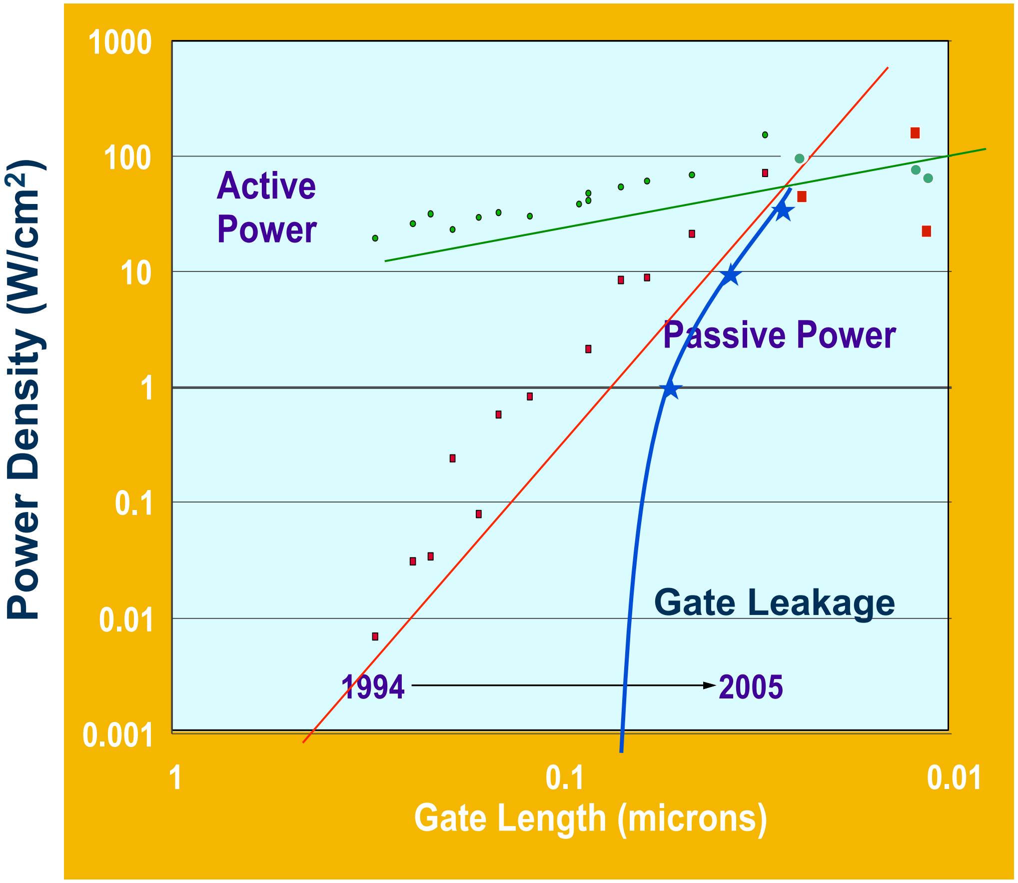
Subthreshold Leakage

- $P = KVe^{(Vgs-Vt)q/nkT} (1 – e^{-Vds\ q/kT})$
- Supplies have been held artificially high (for freq)
  - Threshold has not dropped as fast as it should (because of variability and high supply voltages) (FINFETs fix this for the time being).
  - We’d like to maintain Ion:Ioff = ~1000uA/u : 10nA/u
  - Relatively poor performance => Low Vt options
    - 70-180mV lower Vt, 10-100x higher leakage, 5-15% faster
- Subthreshold lkg especially increasing in short channel devices (DIBL) & at high T – 100-1000nA/u
- Subthreshold slope 85-110 mV/decade
- Cooling changes the slope….but can it be energy efficient?

Passive Power Continues to Explode

Leakage is the price we pay for the increasing device performance

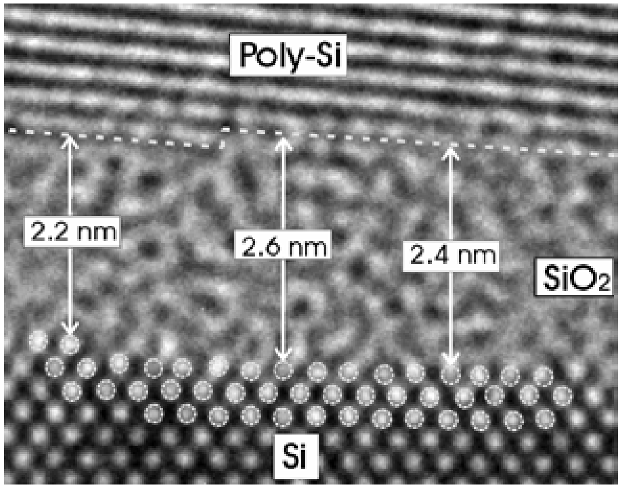
Fit of published active and subthreshold CMOS device leakage densities



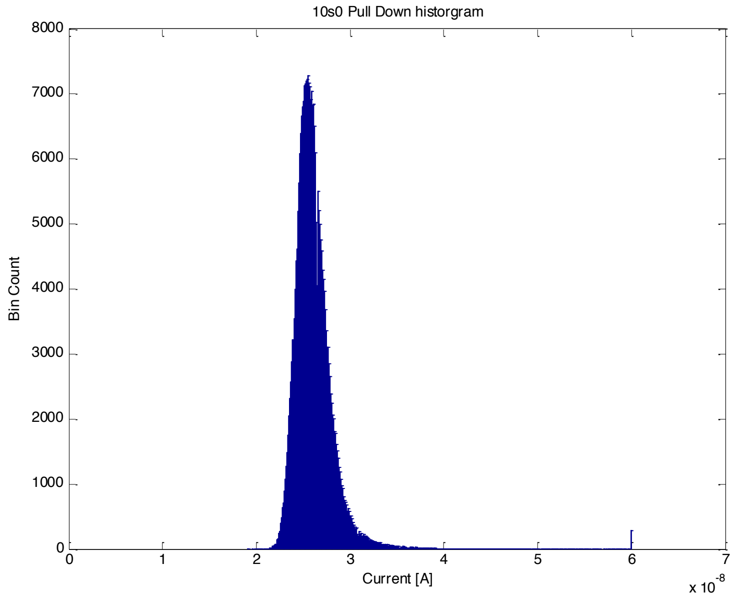
Src: Nowak, et al

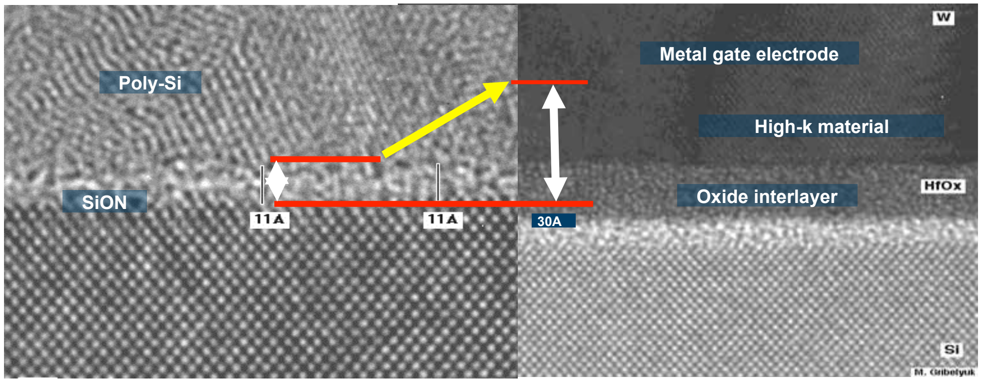
Gate Oxide & Leakage

- **Gate tunneling becoming dominant leakage mechanism in very thin gate oxides**
  - **Exponential effect on gate tunneling currents**
    - **Current exponential in oxide thickness**
    - **Current exponential in voltage across oxide**
  - **Atomistic variation in oxide thickness**
- **Reduction techniques:**
  - Lower the field (voltage or oxide thickness)
  - New gate ox material



1.1nm oxide is ~6 atomic layers. Across a 300mm wafer >10^9 atomic layers





Standby-Power Reduction Techniques

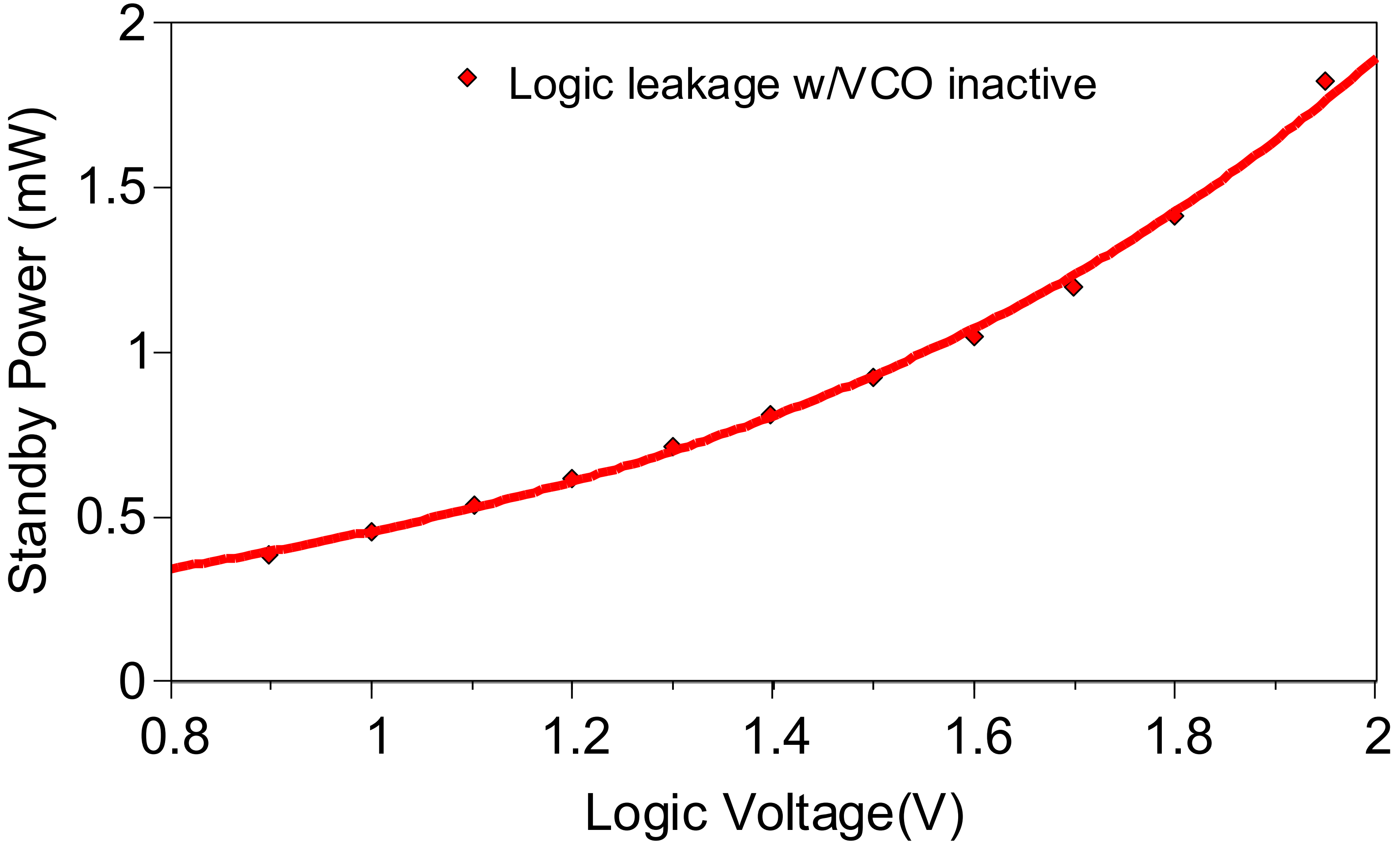
- Standby power can be reduced through:
    - Capacitance minimization
    - Voltage-scaling
    - Power gating
    - Vdd/Vt selection

Capacitance minimization

- Only the devices (device width) used in the design leak!
    - Runs counter to the complexity-for-IPC trend
    - Runs counter to the SOC trend
    - Transistors are not free -- Even though they are not switched they still leak

Voltage Scaling Standby Reduction

Decreasing the supply voltage significantly improves standby power



Subthreshold dominated technology

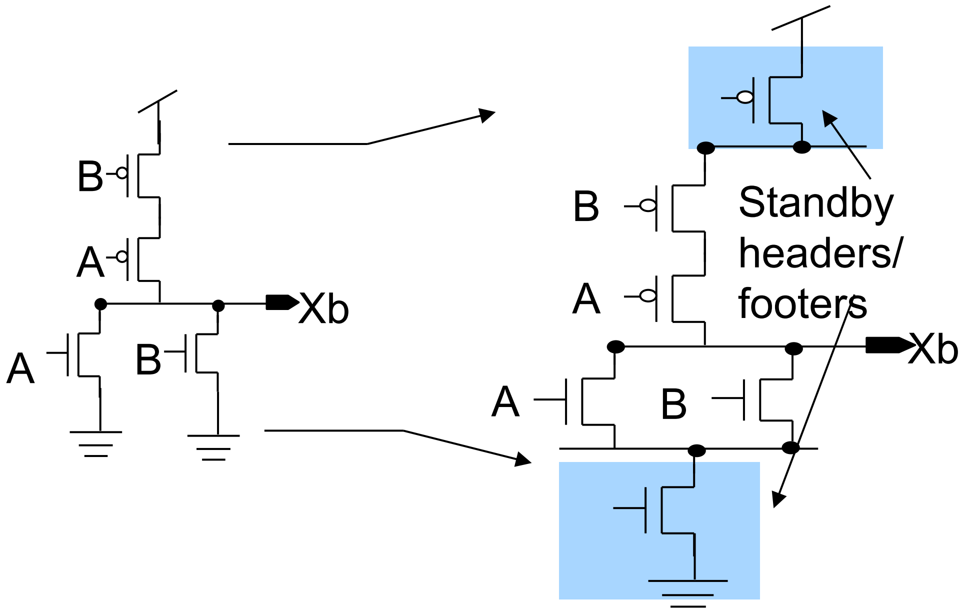
After Nowka, et.al. ISSCC '02

Supply/Power Gating

- Especially for energy constrained (e.g. battery powered systems). Two levels of gating:
  - “Standby, freeze, sleep, deep-sleep, doze, nap, hibernate”: lower or turn off power supply to system to avoid power consumption when inactive
    - Control difficulties, hidden-state, entry/exit, “instant-on” or user-visible.
  - Unit level power gating – turn off inactive units while system is active
    - Eg. MTCMOS
    - Distribution, entry/exit control & glitching, state-loss…
  - How do you determine if this is effective?

MTCMOS

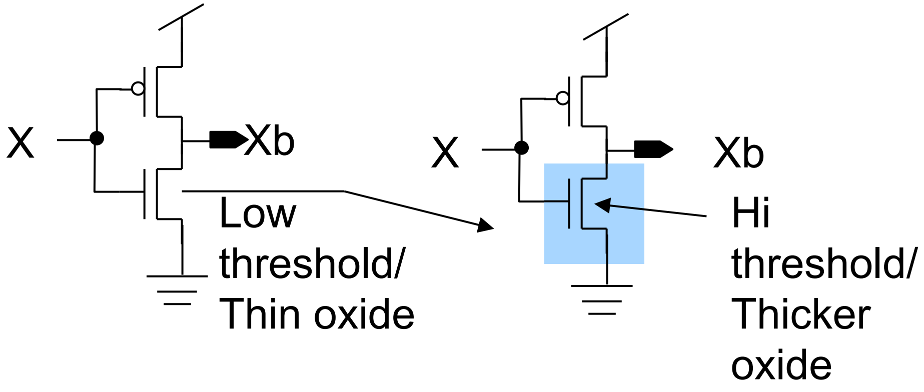
- Use header and/or footer switches to disconnect supplies when inactive.
- For performance, low-Vt for logic devices.
- 10-100x leakage improvement, ~5% perf overhead
- Loss of state when disconnected from supplies
- Large number of variants in the literature



Standby headers/ footers

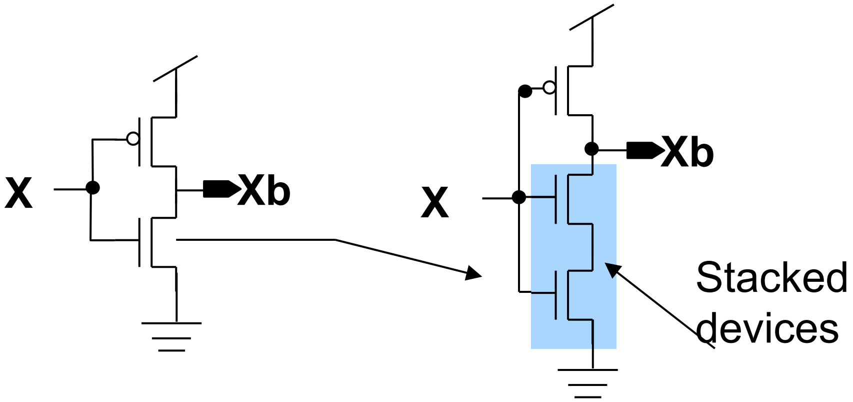
Vt / Tox selection

- Low Vt devices on critical paths, rest high Vt
- 70-180mV higher Vt, 10-100x lower leakage, 5-20% slower
- Small fraction of devices low-Vt (1-5%)
- Thick oxide reduces gate leakage by orders of magnitude



Device Stacking

- Decreases subthreshold leakage
- Improvement beyond use of long channel device
- 2-5x improvement in subthreshold leakage
- 15-35% performance penalty



Vt or/and Vdd selection

- Design tradeoff:
  - Performance => High supply, low threshold
  - Active Power => Low supply, low threshold
  - Standby => Low supply, high threshold
- Static
  - Stack effect – minimizing subthreshold thru single fet paths
  - Multiple thresholds: High Vt and Low Vt transistors
  - Multiple supplies: high and low Vdd

Vt or/and Vdd selection (cont’d)

- Design tradeoff:
  - Performance => High supply, low threshold
  - Active Power => Low supply, low threshold
  - Standby => Low supply, high threshold
- Static
  - Stack effect – minimizing subthreshold thru single fet paths
  - Multiple thresholds: High Vt and Low Vt Transistors
  - Multiple supplies: high and low Vdd
  - Problem: optimum (Vdd,Vt) changes over time, across dice

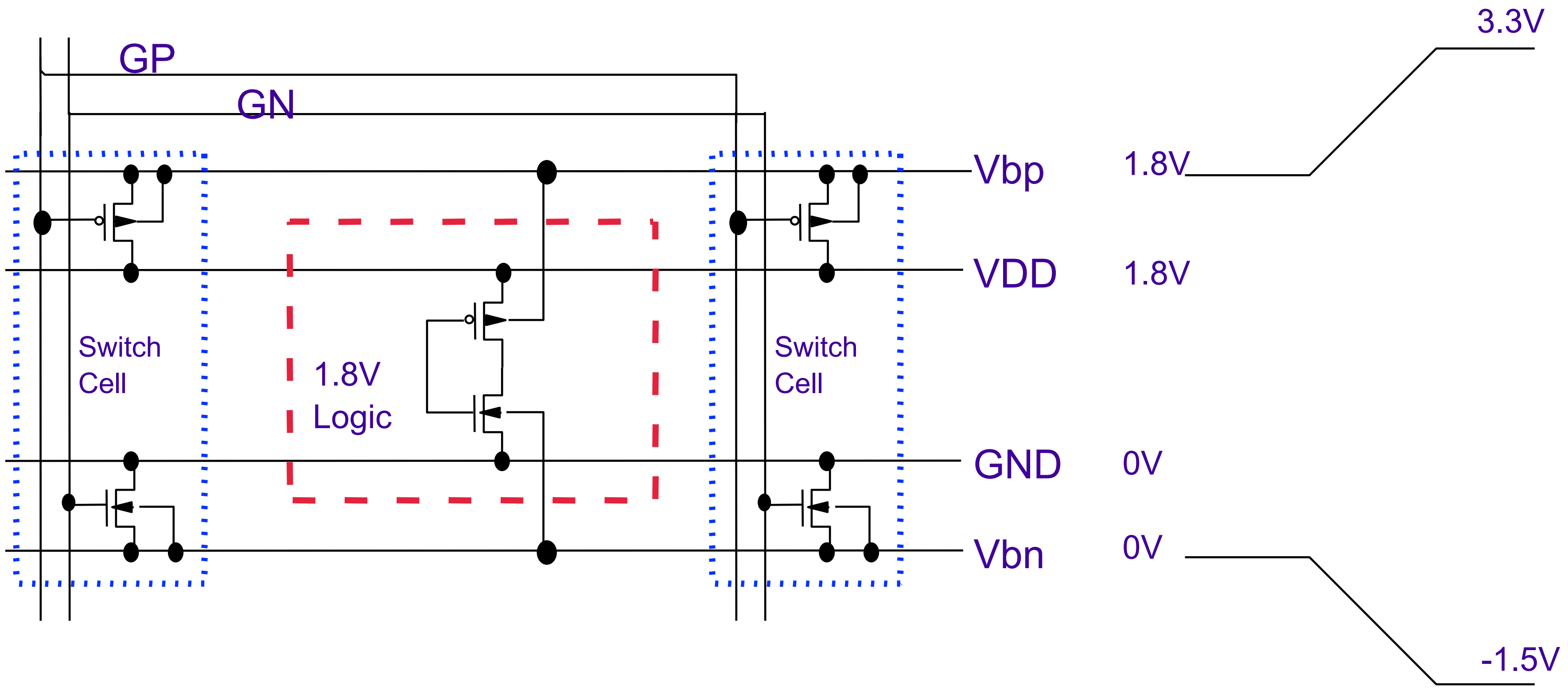
- Dynamic (Vdd,Vt) selection
  - DVS for supply voltage
  - Dynamic threshold control thru:
    - Active well
    - Substrate biasing
    - SOI back gate, DTMOS, dual-gate technologies

Hitachi-SH4 leakage reduction

Triple Well Process

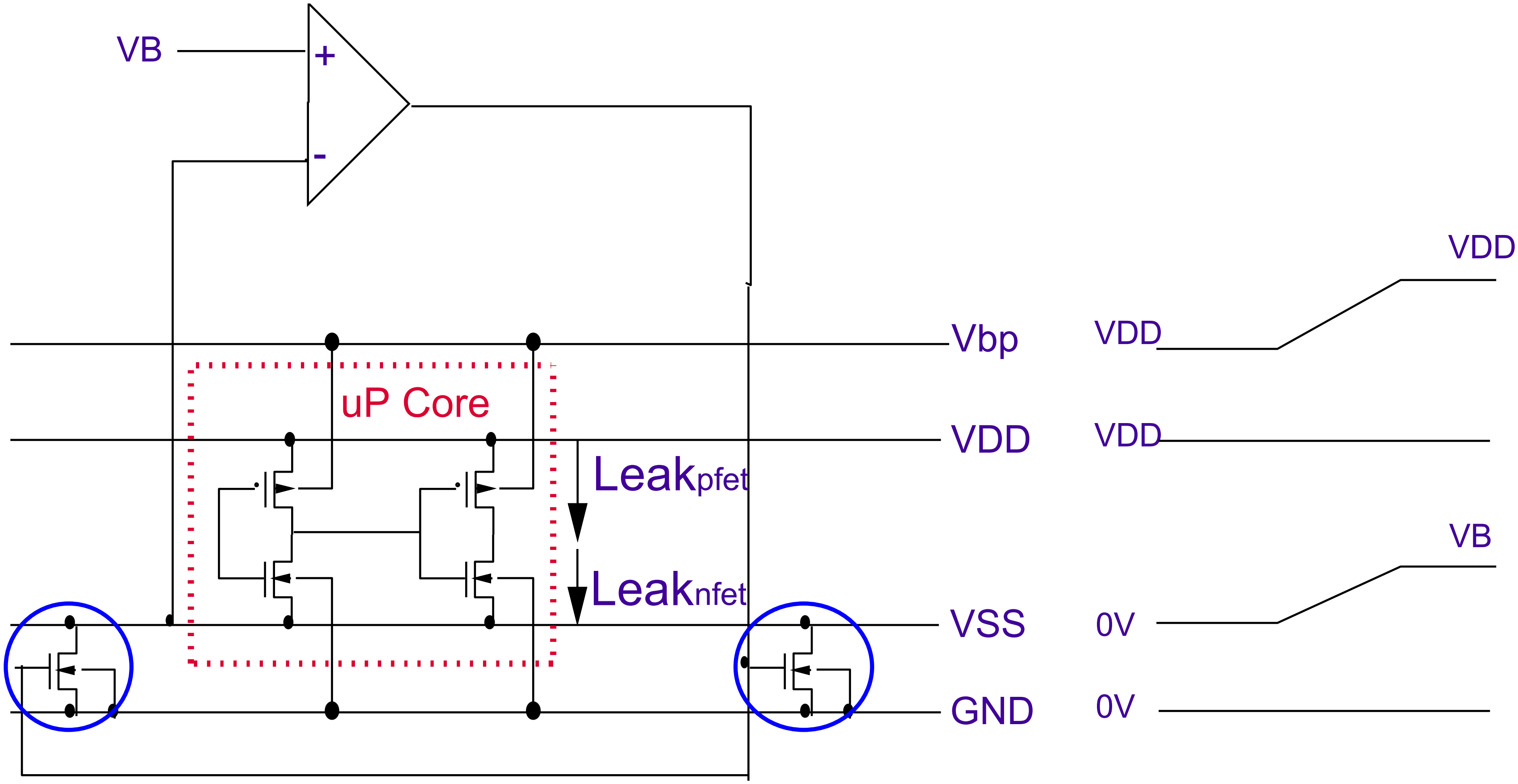
Reverse Bias Active Well –

can achieve >100x leakage reduction



Nwell/Virtual Gnd Leakage Reduction

Similar technique for Nwell/Psub technology
– Intel approach



Estimating Leakage Power Consumption

- $P = ½ C_{SW}V_{dd} \Delta V f + I_{st}V_{dd} + I_{static}V_{dd}$

- The problem is how to estimate the leakage current
- Estimating leakage currents
  - Low level circuit analysis – spice analysis
  - Higher level: spreadsheet/back-of-the-envelope/power tools for estimation
    - Subthreshold: Estimates based on the fraction of the device width leaking. Usually evaluated for some non-nominal point in the process and higher temperature. Aggregate or node-by-node estimation of derating factors – fraction of devices with field across the SD device ~1/3 for logic.
    - Gate leakage: Estimates based on the fraction of the device area leaking. Aggregate or node-by-node estimation of derating factors – fraction of devices with field across the gate of the device.

Agenda

- Overview of VLSI power
- Technology, Scaling, and Power
- A look at the real trends and projections for the future
- Active power: components, trends, managing, & estimating
- Static power: components, trends, managing, & estimating
- Summary

Low Power Circuits Summary

- Technology, Scaling, and Power
- Technology scaling hasn't solved the power/energy problems.
- So what to do? We've shown that,
- Do less and/or do in parallel at low V. For the circuit designer this implies:
  - supporting low V,
  - supporting power-down modes,
  - choosing the right mix of Vt,
  - sizing devices appropriately
  - choosing right Vdd, (adaptation!)

References

- Power Metrics

  - T. Sakurai and A. Newton, “Alpha-power law MOSFET model and its applications to CMOS inverter delay and other formulas”, IEEE Journal of Solid State Circuits, v. 25.2, pp. 584-594, Apr. 1990.
  - R. Gonzalez, B. Gordon, M. Horowitz, “Supply and threshold voltage scaling for low power CMOS” IEEE Journal of Solid State Circuits, v. 32, no. 8, pp. 1210-1216, August 2000.
  - Zyuban and Strenski, “Unified Methodology for Resolving Power-Performance Tradeoffs at the Microarchitectural and Circuit Levels”,ISPLED Aug.2002
  - Brodersen, Horowitz, Markovic, Nikolic, Stojanovic “Methods for True Power Minimization”, ICCAD Nov. 2002
  - Stojanovic, Markovic, Nikolic, Horowitz, Brodersen, “Energy-Delay Tradoffs in Combinational Logic using Gate Sizing and Supply Voltage Optimization”, ESSCIRC, Sep. 2002

References

- Power/Low Power

  - SIA, International Technology Roadmap for Semiconductors, 2001,2003, 2005 available online.
  - V. Agarwal, M.S. Hrishikesh, S.W. Keckler, and D. Burger. "Clock Rate Versus IPC: The End of the Road for Conventional Microarchitectures," 27th International Symposium on Computer Architecture (ISCA), June, 2000.
  - Allan, et. al., “2001 Tech. Roadmap for Semiconductors”,IEEE Computer Jan. 2002
  - Chandrakasan, Broderson, (ed) Low Power CMOS Design IEEE Press, 1998.
  - Oklobdzija (ed) The Computer Engineering Handbook CRC Press, 2002
  - Kuo, Lou Low voltage CMOS VLSI Circuits, Wiley, 1999.
  - Bellaouar, Elmasry, Low Power Digital VLSI Design, Circuits and Systems, Kluwer, 1995.
  - Chandrakasan, Broderson, Low Power Digital CMOS Design Kluwer, 1995.
  - A. Correale, “Overview of the power minimization techniques employed in the IBM PowerPC 4xx embedded controllers” IEEE Symposium on Low Power Electronics Digest of Technical Papers, pp. 75-80, 1995.
  - K. Nowka, G. Carpenter, E. MacDonald, H. Ngo, B. Brock, K. Ishii, T. Nguyen, J. Burns, “A 0.9V to 1.95V dynamic voltage scalable and frequency scalable 32-bit PowerPC processor “, Proceedings of the IEEE International Solid State Circuits Conference, Feb. 2002.
  - K. Nowka, G. Carpenter, E. MacDonald, H. Ngo, B. Brock, K. Ishii, T. Nguyen, J. Burns, “A 32-bit PowerPC System-on-a-Chip with support for dynamic voltage scaling and dynamic frequency scaling”, IEEE Journal of Solid State Circuits, November, 2002.

References

- Low Voltage / Voltage Scaling

  - E. Vittoz, “Low-power design: ways to approach the limits” IEEE International Solid State Circuits Conference Digest of Technical Papers, pp. 14-18, 1994.
  - M. Horowitz, T. Indermaur, R. Gonzalez, “Low-power digital design” IEEE Symposium on Low Power Electronics Digest of Technical Papers, pp. 8-11, 1994.
  - R. Gonzalez, B. Gordon, M. Horowitz, “Supply and threshold voltage scaling for low power CMOS” IEEE Journal of Solid State Circuits, v. 32, no. 8, pp. 1210-1216, August 2000.
  - T. Burd and R. Brodersen, “Energy efficient CMOS microprocessor design ” Proceedings of the Twenty-Eighth Hawaii International Conference on System Sciences, v. 1, pp. 288-297, 466, 1995.
  - K. Suzuki, S. Mita, T. Fujita, F. Yamane, F. Sano, A. Chiba, Y. Watanabe, K. Matsuda, T. Maeda, T. Kuroda, “A 300 MIPS/W RISC core processor with variable supply-voltage scheme in variable threshold-voltage CMOS” Proceedings of the IEEE Conference on Custom Integrated Circuits Conference, pp. 587 –590, 1997
  - T. Kuroda, K. Suzuki, S. Mita, T. Fujita, F. Yamane, F. Sano, A. Chiba, Y. Watanabe, K. Matsuda, T. Maeda, T. Sakurai, T. Furuyama, “Variable supply-voltage scheme for low-power high-speed CMOS digital design” IEEE Journal of Solid State Circuits, v. 33, no. 3, pp. 454-462, March 1998.
  - T. Burd, T. Pering, A. Stratakos, R. Brodersen, “A dynamic voltage scaled microprocessor system ” IEEE International Solid State Circuits Conference Digest of Technical Papers, pp. 294-295, 466, 2000.

References

- Technology and Circuit Techniques

  - E. Nowak, et al., “Scaling beyond the 65 nm node with FinFET-DGCMOS” Proceedings of the IEEE Custom Integrated Circuits Conference, Sept. 21-24, 2003, pp.339 – 342
  - L. Clark, et al. “An embedded 32b microprocessor core for low-power and high-performnace applications”, IEEE Journal of Solid State Circuits, V. 36, No. 11, Nov. 2001, pp. 1599-1608
  - S. Mukhopadhyay, C. Neau, R. Cakici, A. Agarwal, C. Kim, and K. Roy, “Gate leakage reduction for scaled devices using transistor stacking” IEEE Transactions on Very Large Scale Integration (VLSI) Systems, Aug. 2003, pp. 716 – 730
  - A. Bhavnagarwala, et al., “A pico-joule class, 1GHz, 32 Kbyte x 64b DSP SRAM with Self Reverse Bias” 2003 Symposium on VLSI Circuits, June 2003, pp. 251-251.
  - S. Mutoh, et al., “1-V Power Supply High-Speed Digital Circuit Technology with Multi-Threshold Voltage CMOS,” IEEE Journal of Solid State Circuits, vol. 30, no. 8, pp. 847-854, 1995.
  - K. Das, et al., “New Optimal Design Strategies and Analysis of Ultra-Low Leakage Circuits for Nano- Scale SOI Technology,” Proc. ISLPED, pp. 168-171, 2003.
  - R. Rao, J. Burns and R. Brown, “Circuit Techniques for Gate and Sub-Threshold Leakage Minimization in Future CMOS Technologies” Proc. ESSCIRC, pp. 2790-2795, 2003.
  - R. Rao, J. Burns and R. Brown, “Analysis and optimization of enhanced MTCMOS scheme” Proc. 17th International Conference on VLSI Design, 2004, pp. 234-239.