HOT Chips 2005

Circuit Design for Low Power

Kevin Nowka, IBM Austin Research Laboratory

Agenda

- Designing with power and energy limits
- Overview of VLSI power
- Technology, Scaling, and Power
  - Review of scaling
  - A look at the real trends and projections for the future
    - Active power – components, trends, managing active power
    - Static power – components, trends, managing static power
- Summary

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, 100W/sq-cm air cooled limit, 7.5kW 19” rack)
  - Device reliability (junction temp > 125C substantial reduction in 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
- Energy cost
  - Energy for IT equipment large fraction of total cost of ownership

Agenda

Designing with power and energy limits

Overview of VLSI power

Technology, Scaling, and Power

- Review of scaling
- A look at the real trends and projections for the future
  - Active power – components, trends, managing active power
  - Static power – components, trends, managing active power

Summary

CMOS circuit power consumption components

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

- Dynamic power consumption ($\tfrac{1}{2} C_{sw} V_{dd} \Delta V f + I_{st} V_{dd}$)
  - 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

Agenda

Designing with power and energy limits

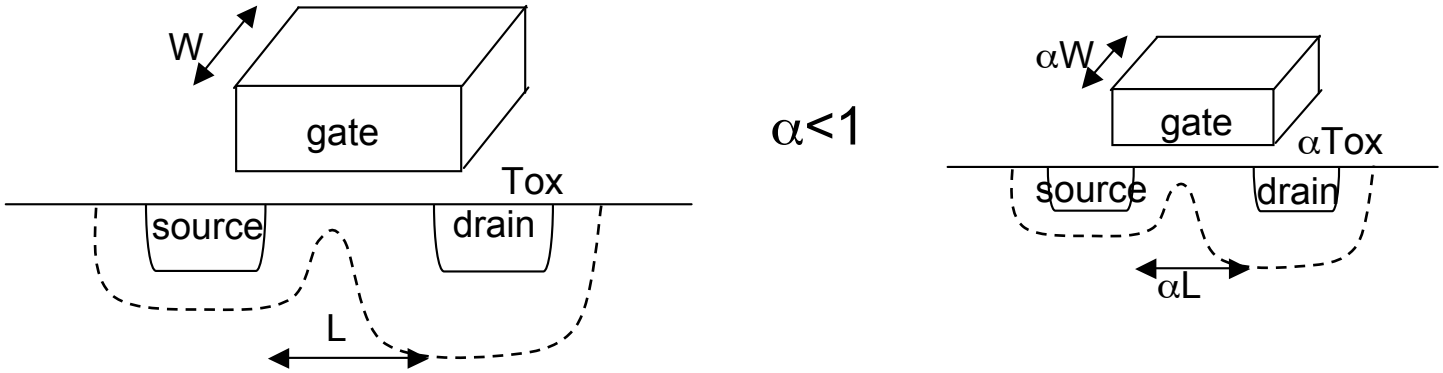
Overview of VLSI power

Technology, Scaling, and Power

- **Review of scaling**
- A look at the real trends and projections for the future
  - Active power – components, trends, managing active power
  - Static power – components, trends, managing active power

Summary

Review of Constant Field Scaling



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

What about Deltas?

Agenda

Designing with power and energy limits

Overview of VLSI power

Technology, Scaling, and Power

- Review of scaling
- **A look at the real trends and projections for the future**
  - **Active power – components, trends, managing active power**
  - Static power – components, trends, managing active power

Summary

CMOS Circuit Delay and Frequency

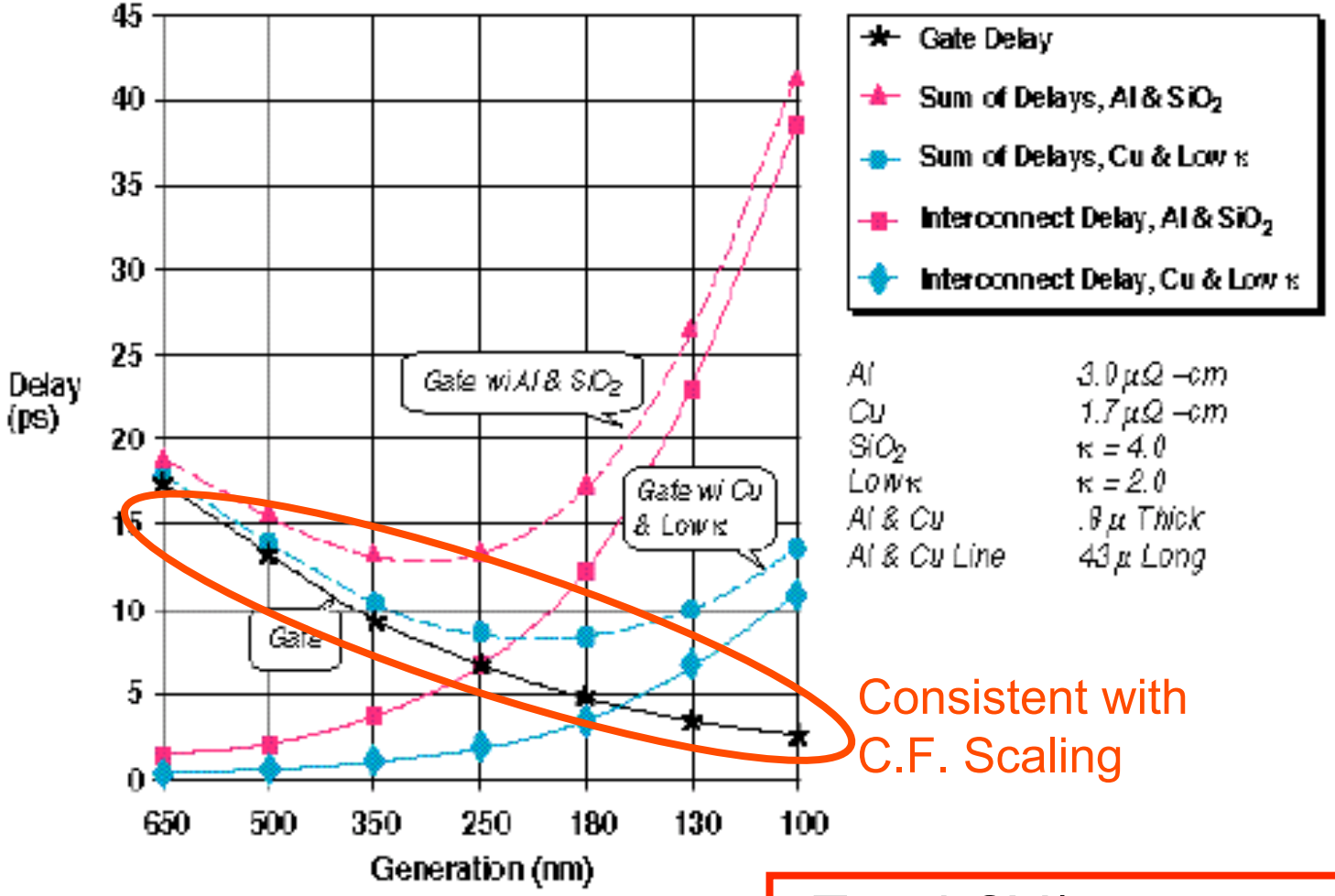
VLSI system frequency determined by:

- Sum of propagation delays across gates in "critical path" --
  - Each gate delay, includes time to charge/discharge load thru a FET and interconnect delay to distribute to next gate input.

$$T_d = kCV/I$$
$$= kCV/(V_{dd}-V_t)^{\alpha}$$

Sakuri α-power law model of delay

Gate Delay Trends



Consistent with C.F. Scaling

Each technology generation, gate delay reduced about 30% (src: ITRS '01)

$$T_d = kCV/I = kCV/(V_{dd}-V_t)^{\alpha}$$

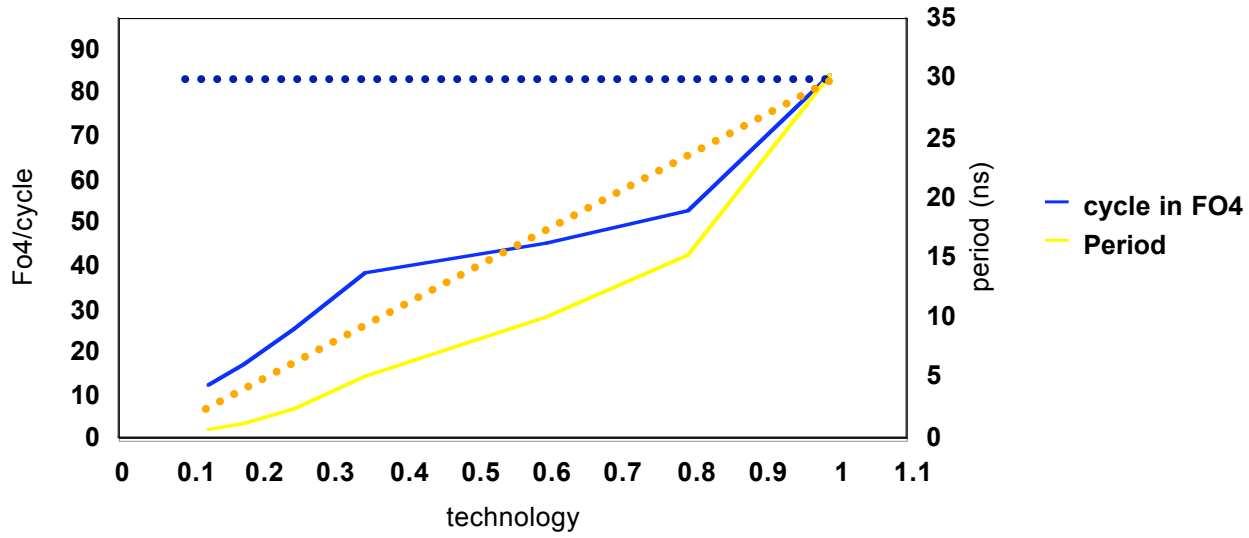
Microprocessor Frequency

In practice the trend is:

Frequency increasing by 2X (delay decreasing by 50%), not the 1.4X (30%) for constant field scaling (src: ITRS '01).

Why? decreasing logic/stage and increased pipeline depth.

Intel 32b (after Hrishikesh, et. al)



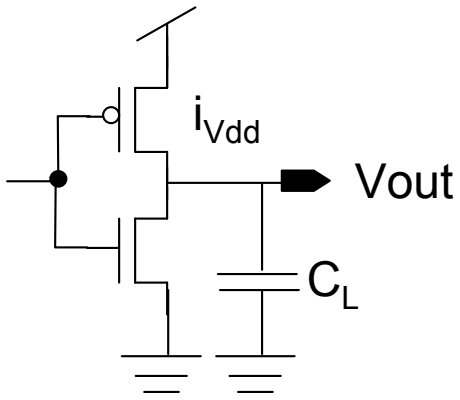
Dynamic Energy

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

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

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

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

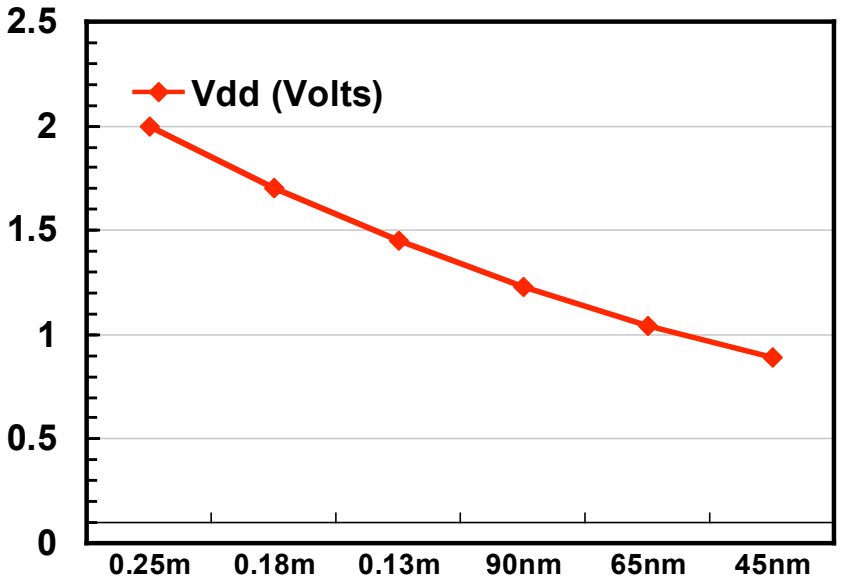


Energy dissipated for either output transition consumes:

$$½\ C_L\ V_{dd}^2$$

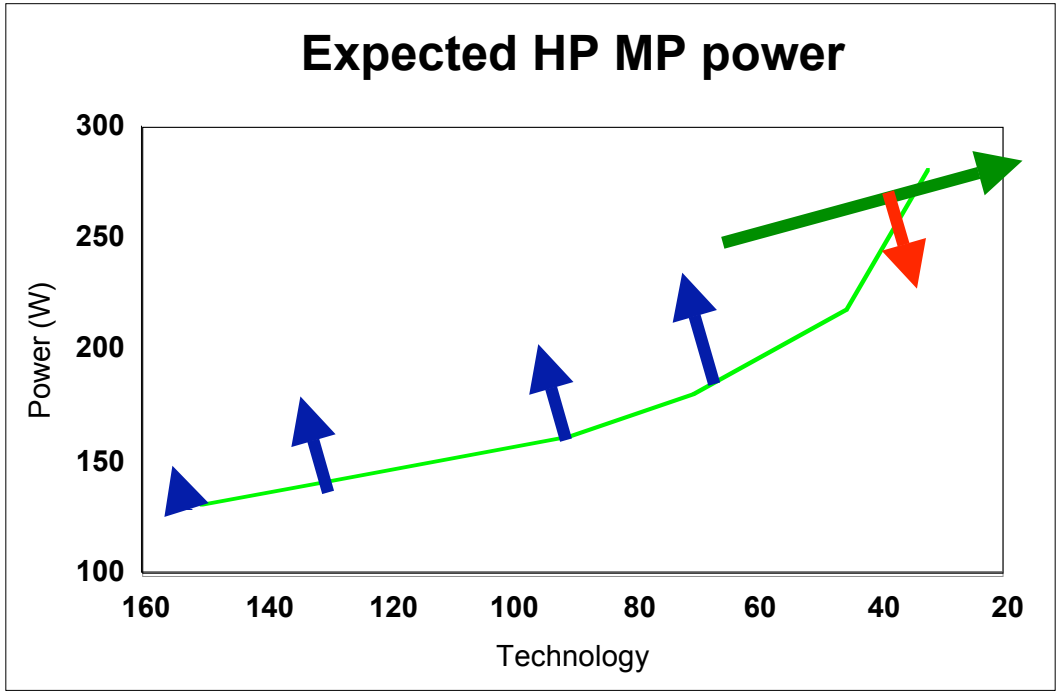
Gate level energy consumption should improve as α^3 under constant field scaling, but….

Supply Voltage/Energy Trend



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



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)

* HP MP = High Performance Micro Processor

Active-Power Reduction Techniques

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

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

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

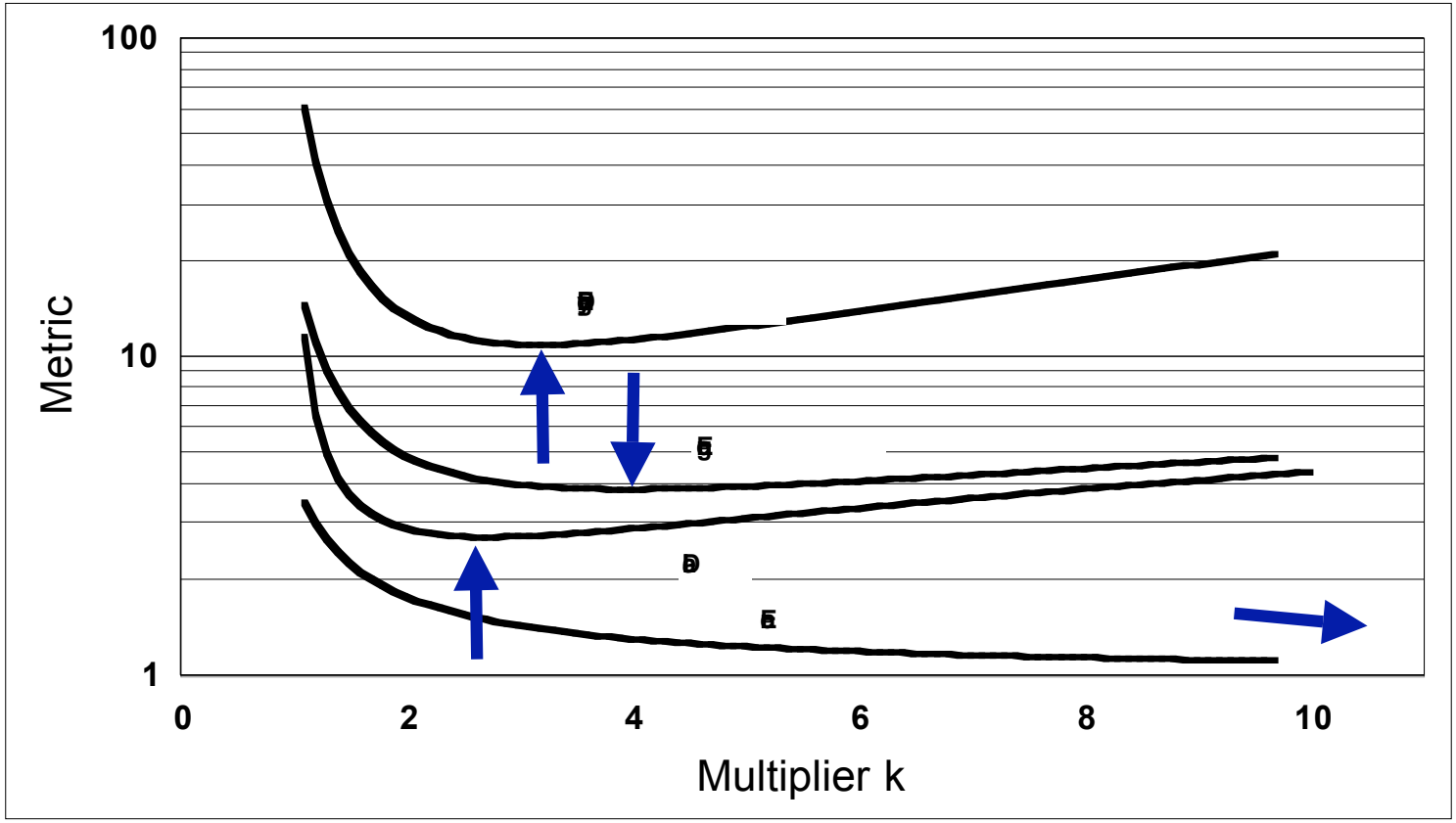
Only the devices (device width) used in the design consume active power!

- **Runs counter to the complexity-for-IPC trend**
- **Runs counter to the SOC trend**

Capacitance minimization

Example of managing design capacitance:

Device sizing for power efficiency is significantly different than sizing for performance – sizing of the gate size multiplier in an exponential-horn of inverters.



Functional Clock Gating

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

- 25-50% of power consumption due to driving latches.
- 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, …
  - Functional clock gating – turn off clocks to individual latch banks – forwarding latch, shift-amount register, overflow logic & latches, …
- Asynch is the most aggressive gating

Glitch suppression

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

- 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
  - 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?

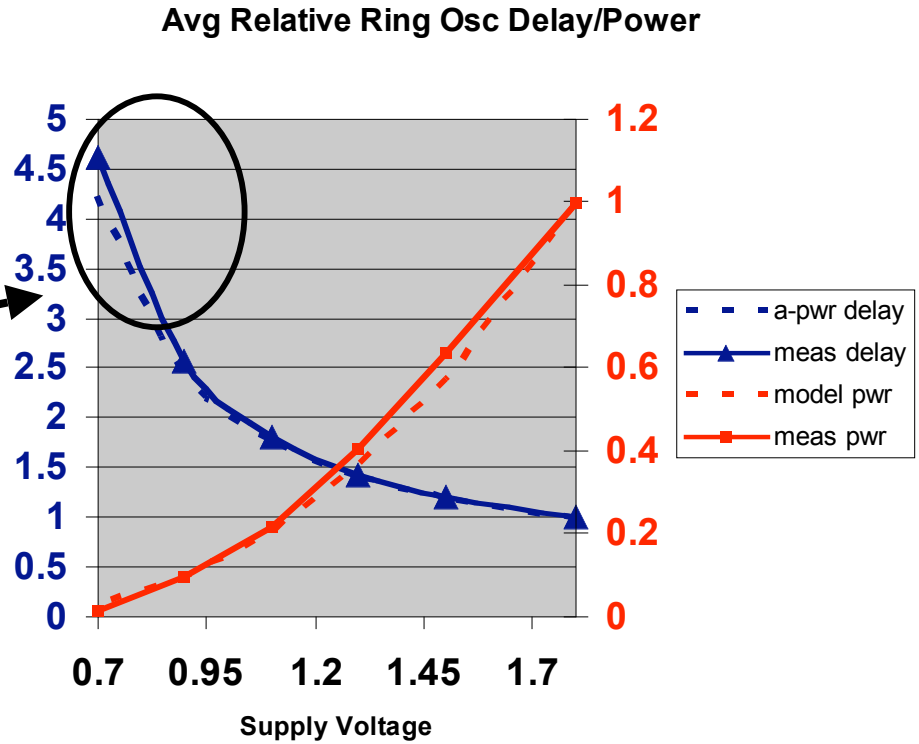
Voltage minimization

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

- 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 ~V^3

Voltage Scaling Reduces Active Power

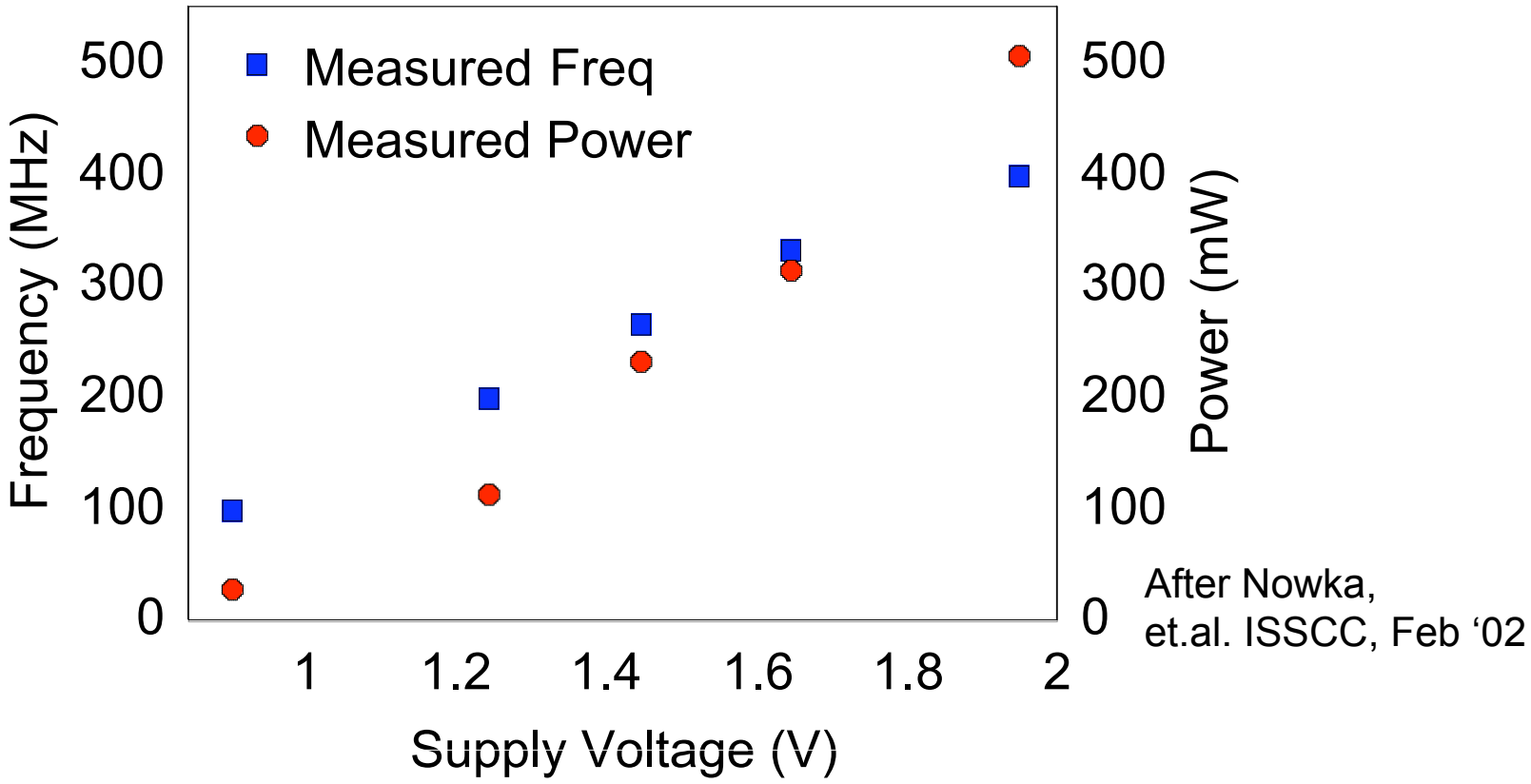
- Voltage Scaling Challenges
  - Custom CPUs, Analog, PLLs, and I/O drivers don't voltage scale easily
  - Sensitivity to supply voltage varies circuit to circuit – esp SRAM, buffers, NAND4
  - Thresholds tend to be too high at low supply
- Voltage Scaling Benefits
  - Can be used widely over entire chip
  - Complementary CMOS scales well over a wide voltage range
  - Can optimize power/performance (MIPS/mW) over a 4X range



After Carpenter, Microprocessor forum, '01

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

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



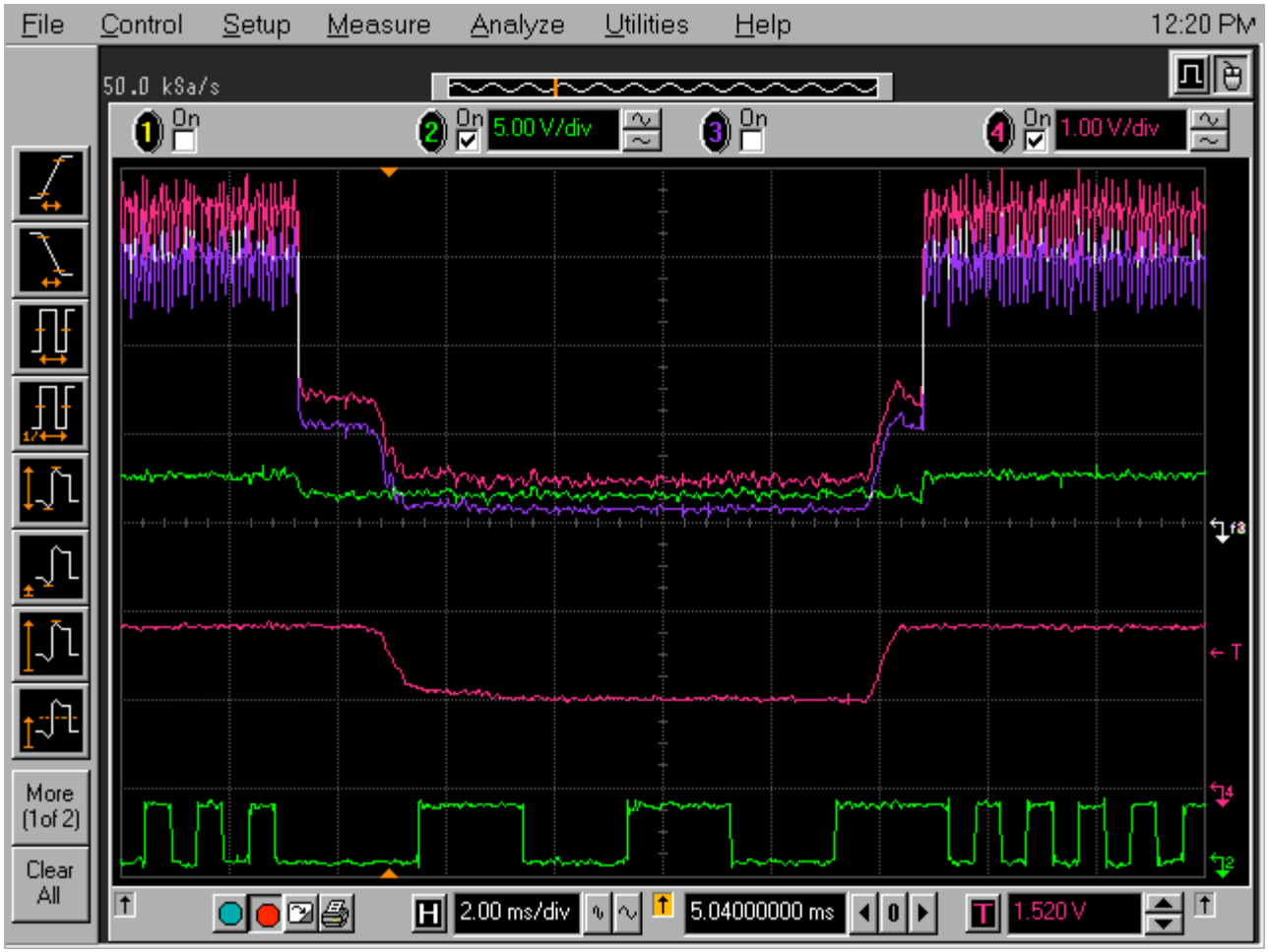
After Nowka, et.al. ISSCC, Feb '02

Frequency minimization

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

- 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

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

- 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

Agenda

Designing with power and energy limits

Overview of VLSI power

Technology, Scaling, and Power

- Review of scaling
- A look at the real trends and projections for the future
  - Active power – components, trends, managing active power
  - **Static power – components, trends, managing active power**

Summary

Static Power

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

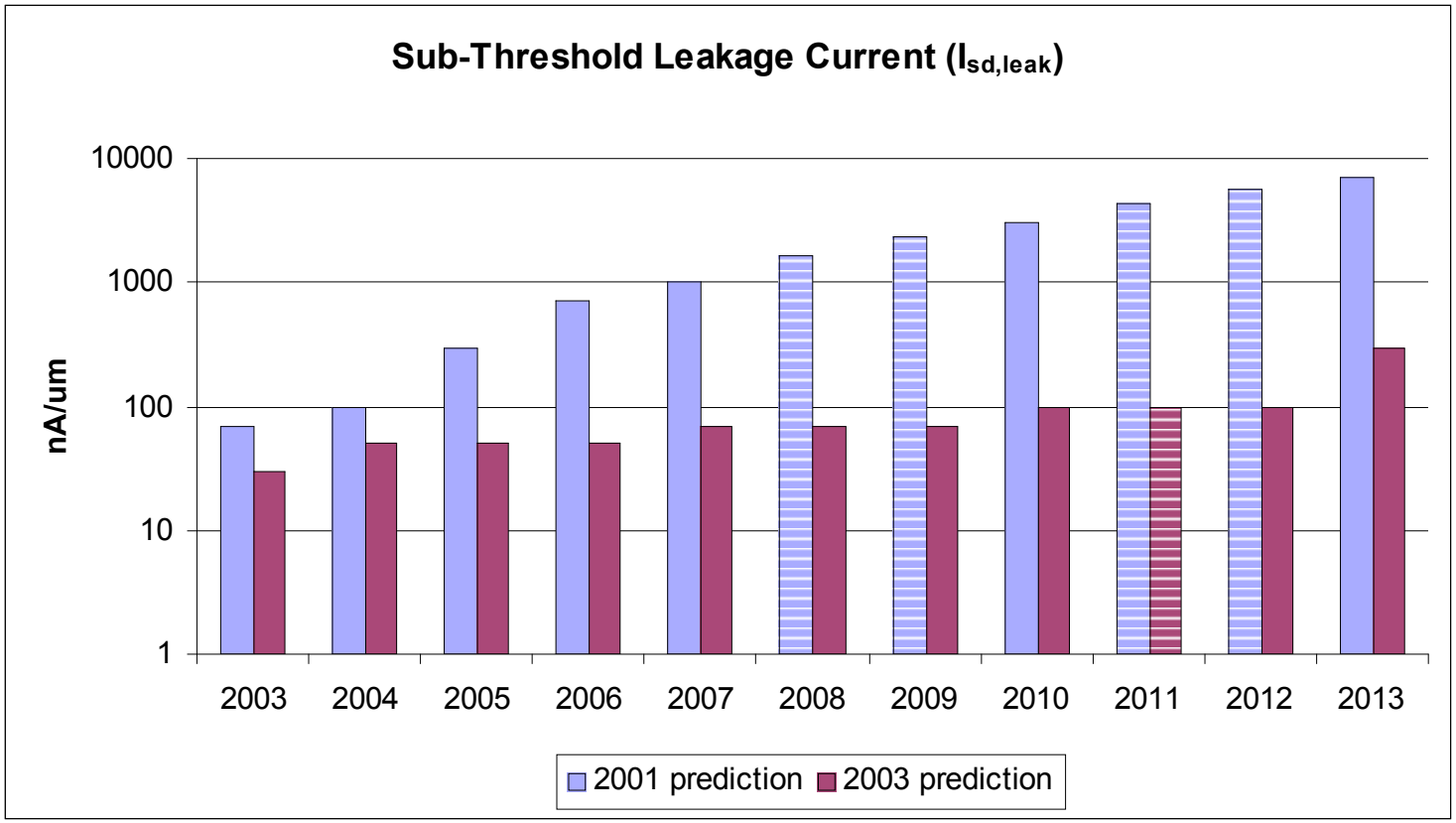
- Static energy consumption ($I_{static} V_{dd}$)
  - Current sources – even uA bias currents can add up.
  - NMOS, pseudo-NMOS – not commonly used
  - CMOS CML logic – significant power for specialized use.
  - Junction currents
  - Subthreshold MOS currents
  - Gate tunneling

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
  - Want 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 Ikg especially increasing in short channel devices (DIBL) & at high T – 100-1000nA/u
- Subthreshold slope 70-80mV/decade
- Cooling changes the slope….but can it be energy efficient?

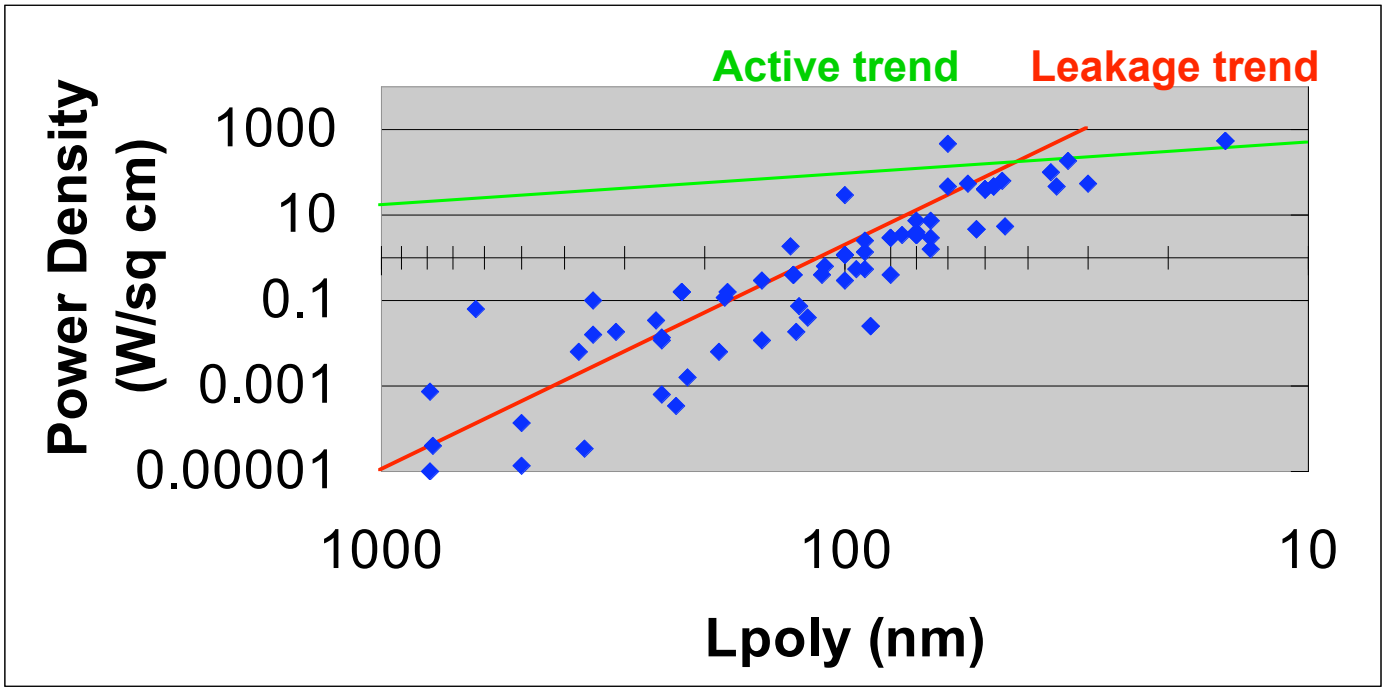
Projected Subthreshold Leakage Trends



Src: ITRS '01, '03
Note: Hatched bars are interpolated

Trends in Leakage Contribution to Power

Fit of published active and subthreshold leakage densities



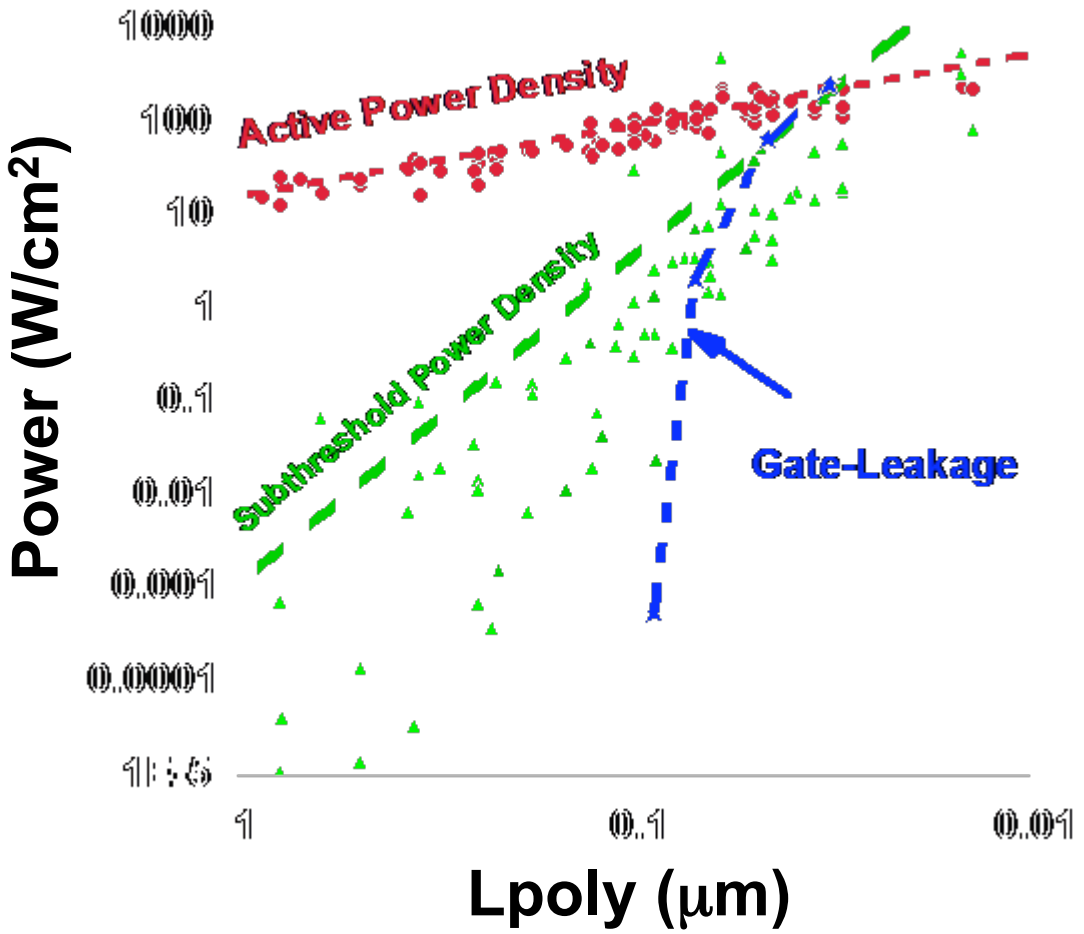
Src: Nowak, et al.

Gate Leakage

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

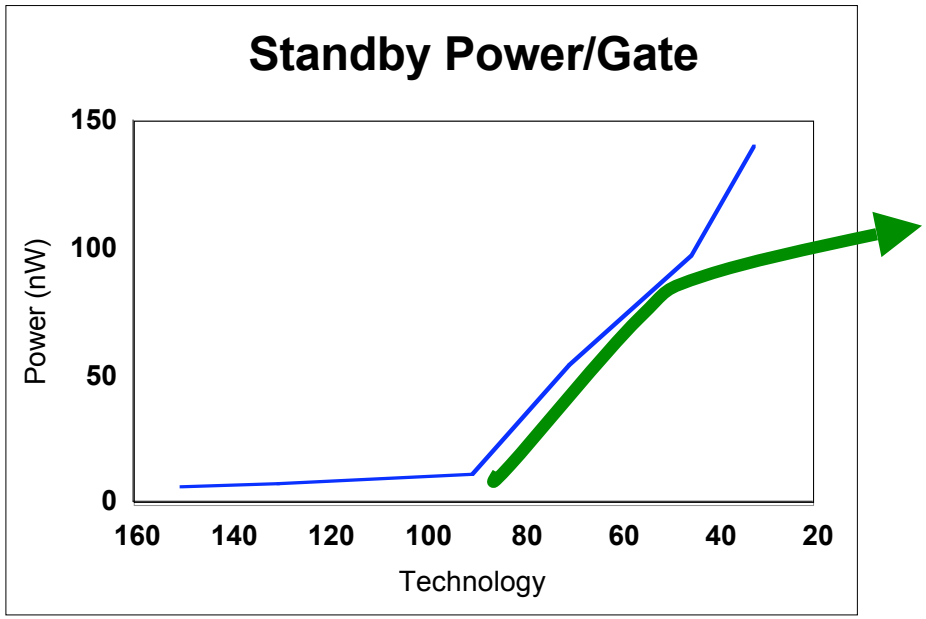
Gate Leakage Trends

Fit of published active, subthreshold, and gate leakage densities



After Nowak, et al.

Future Leakage, Standby Power Trends



Src: ITRS '01

And, recall number of transistors/die has been increasing 2X/2yrs
(Active power/gate should be 0.5x/gen, has been 1X/gen)

For the foreseeable future, leakage is a major power issue

Standby-Power Reduction Techniques

Standby power can be reduced through:

- **Capacitance minimization**
- **Voltage-scaling**
- **Power gating**
- **Vdd/Vt selection**

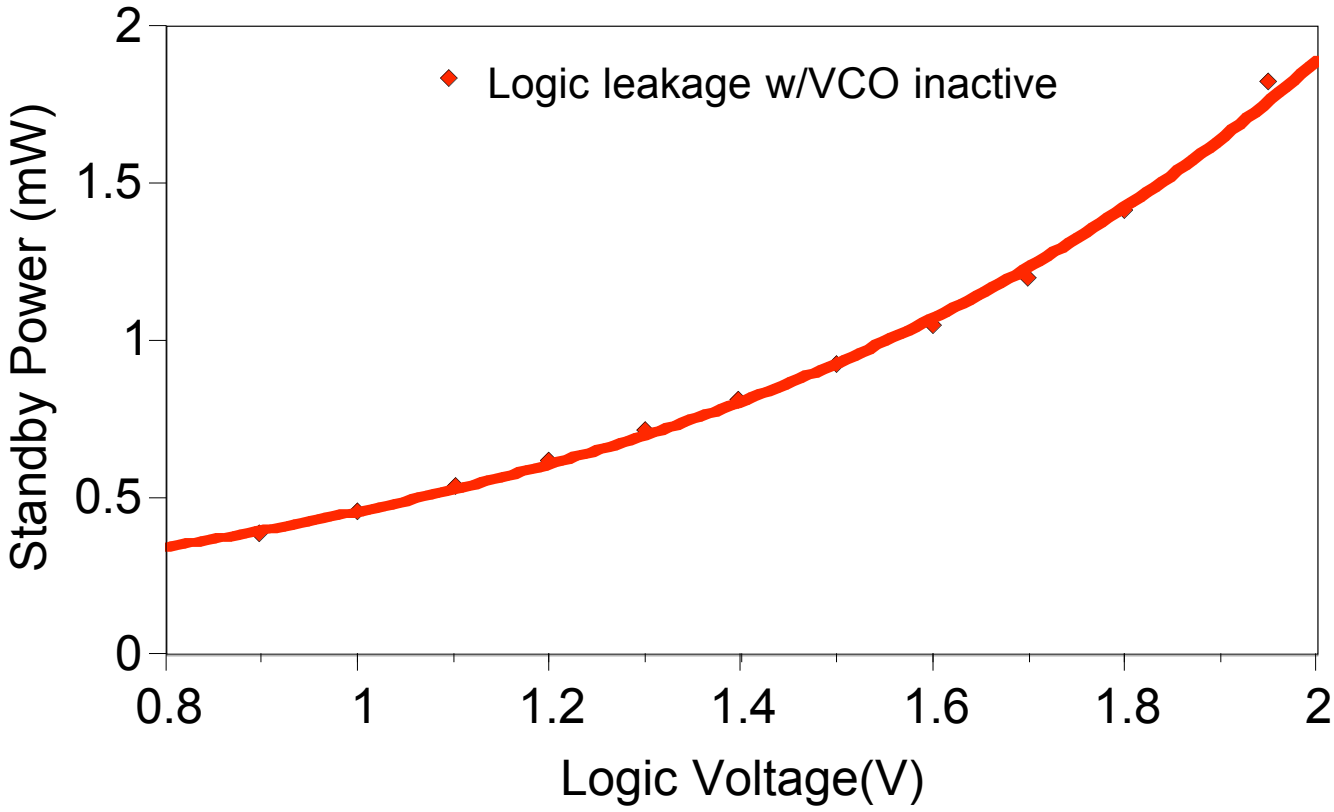
Capacitance minimization

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

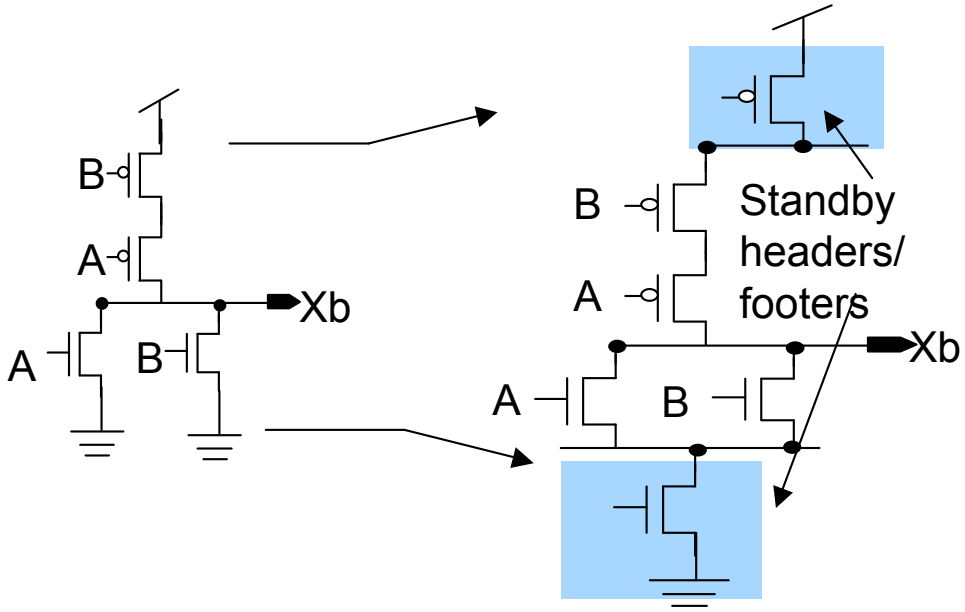
After Nowka, et.al. ISSCC '02

Supply/Power Gating

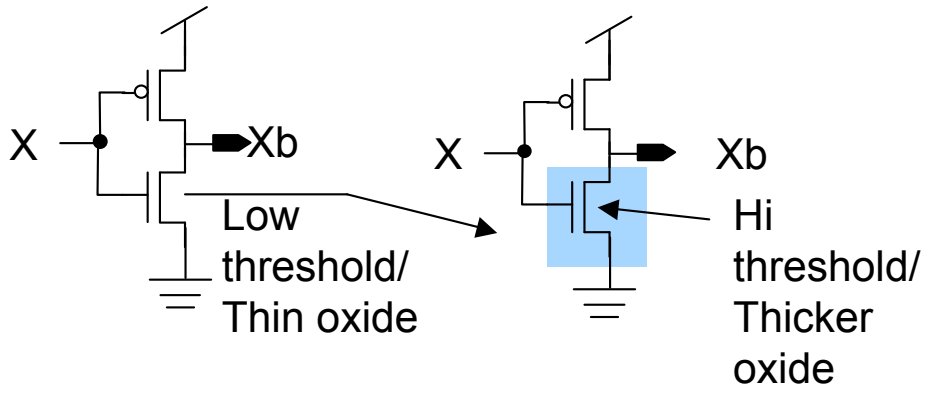
- Especially for energy constrained (e.g. battery powered systems). Two levels of gating:
  - "Standby, freeze, sleep, deep-sleep, doze, nap, hibernate": lower or turn off power supply to system to avoid power consumption when inactive
    - Control difficulties, hidden-state, entry/exit, "instant-on"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

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

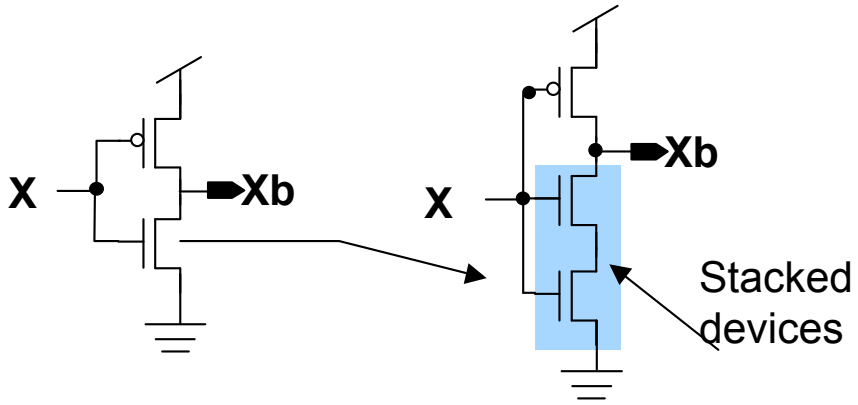


Vt / Tox selection



- 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 (Tox) 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

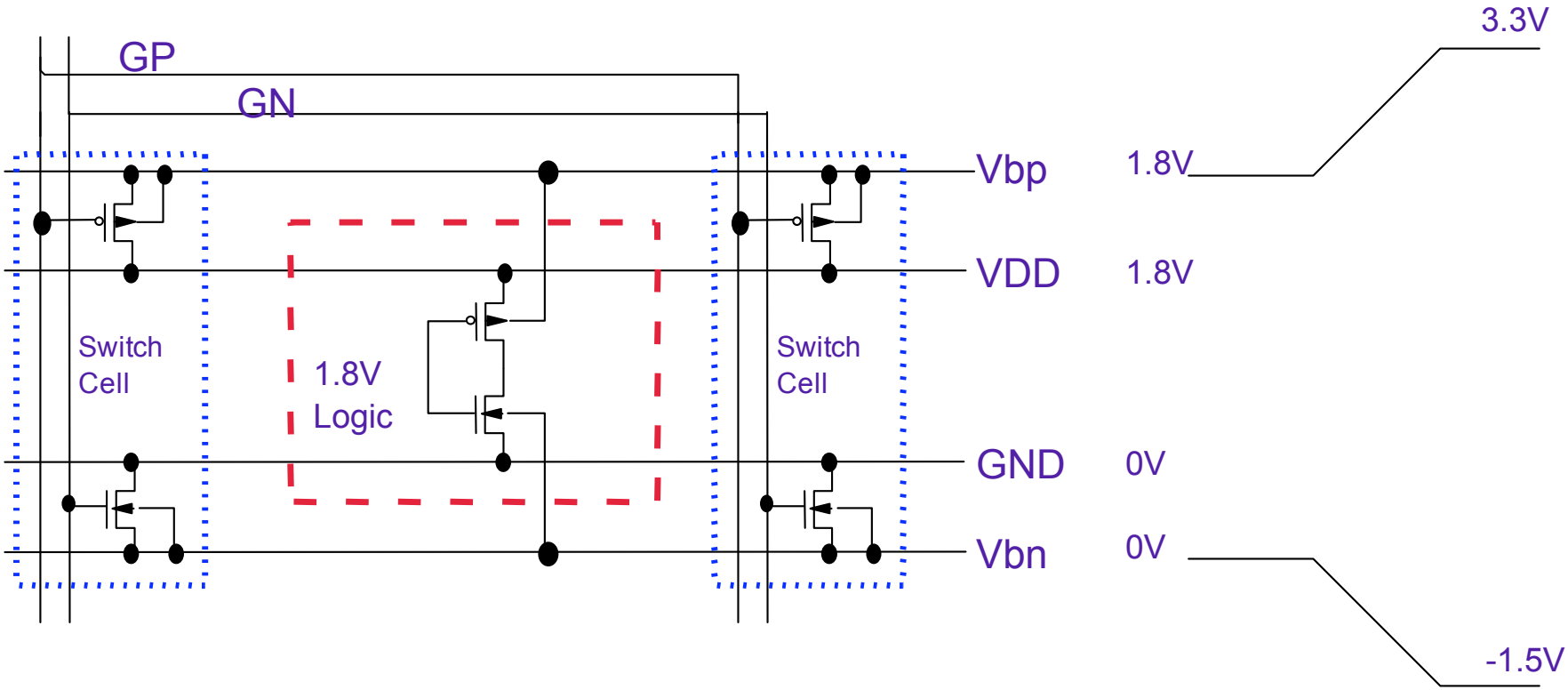
Vt or/and Vdd selection (cont'd)

- 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

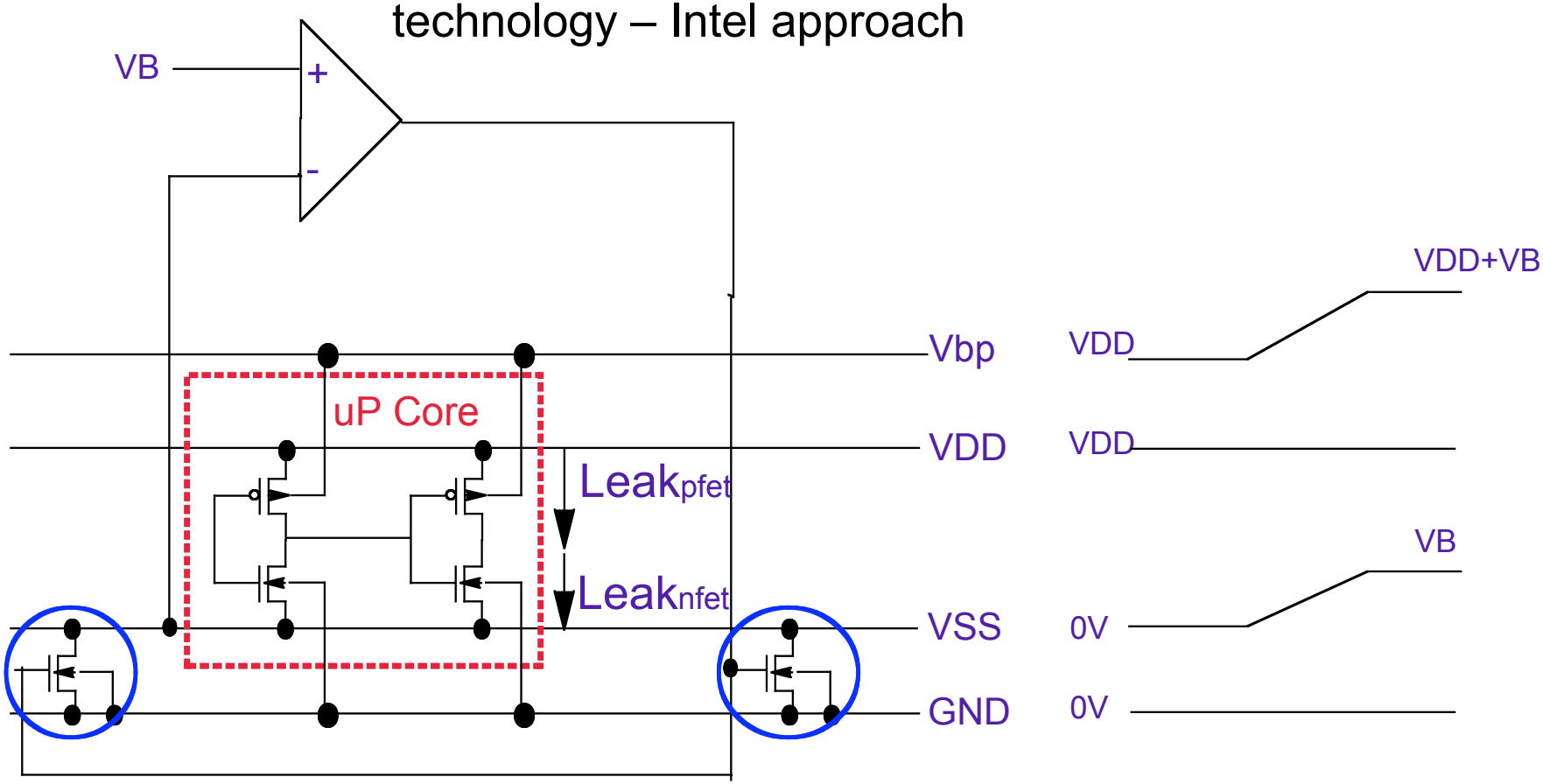
Hitachi-SH4 leakage reduction

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



Agenda

Designing with power and energy limits

Overview of VLSI power

Technology, Scaling, and Power

- Review of scaling
- A look at the real trends and projections for the future
  - Active power – components, trends, managing active power
  - Static power – components, trends, managing active power

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

- 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
- **Power/Low Power**
  - – SIA, International Technology Roadmap for Semiconductors, 2001,2003 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.