

Cool Chips Tutorial

An Industrial Perspective on Low Power Processor Design

Srilatha Manne, Alpha Development Group, Compaq

Trevor Mudge, The University of Michigan, Ann Arbor

Dirk Grunwald, The University of Colorado, Boulder

In Conjunction with the 32nd Annual International Symposium on
Microarchitecture

November 15th, 1999

Dan Carmel Hotel, Haifa, Israel

These proceedings were printed with the
help and support of the Intel Israel
Communication Center.

Foreword

In recent years, power reduction has become a critical design goal for many microprocessors due to portability and reliability requirements. Most of the power savings is achieved through supply voltage reduction and process shrinks. However, there is a limit to how far the supply voltages may be reduced, and the power dissipated on chip is increasing even as process technology improves. Hence, solutions must be found which reduce power at all levels of the design process.

The Power-Driven Microarchitecture Workshop held at ISCA98 in Barcelona Spain (<http://www.cs.colorado.edu/~grunwald/LowPowerWorkshop/>), helped raise the awareness of the architecture community with respect to power issues. The workshop brought together members of industry and academia to explore architectural and compiler modifications for power reduction. It had over 25 papers and 50 registered attendees, and covered topics such as power estimation, architectural modifications for power, reliability issues, compiler techniques for power reduction, and voltage scaling.

This year, the goal remains the same as that of the Power-Driven Microarchitecture Workshop, although the approach is different. Our goal this time is to disseminate knowledge about low power microprocessor design to the architecture community. To this end, we have assembled a group of six speakers from leading microprocessor companies to give presentations on what they consider to be their critical low power issues now and in the future, and to propose some possible solutions to these problems.

Bobbie Manne
Trevor Mudge
Dirk Grunwald

Category	Time	Speaker	Topic
High Performance General Purpose Processors	9:00-9:45	Deo Singh	Power Challenges in the Internet World
		Manager & Principal Engineer, Low Power Design	
		Intel	
	9:45-10:30	Srilatha Manne / Kathy Wilcox	Alpha Processors: A History of Power Issues and A Look to the Future
		Alpha Development Group	
		Compaq	
10:30-11:00		Break	
Ultra Low Power Embedded Processors And DSP Processors	11:00-11:45	John Arends	The M•CORE Technology Center: Designing a Low-Power Architecture
		Design Manager, M*CORE Technology Group	
		Motorola	
	11:45-12:30	Mark Bickerstaff	Low Power DSPs for Wireless Infrastructure
		Senior ASIC Design Engineer	
		Lucent	
12:30-2:00		Lunch (Provided)	
System Level Power Issues And Power Tools	2:00-2:45	Frank Binns	The Power Managing OS Meets a Thermally Aware Processor
		VLSI Architect	
		Performance Microprocessor Division, Intel	
	2:45-3:30	George Z. N. Cai	Architectural Level Power/Performance Optimization and Dynamic Power Estimation
		Intel	
	3:30-4:00		Break
	4:00-4:30	Panel	Future directions

The Speakers

John Arends, Motorola:

Bio: John Arends is a member of the technical staff at Motorola, Inc. He received his BS and MS degrees from the University of Illinois in Electrical Engineering in 1987 and 1989, respectively. While at Motorola, John's area of focus has been RISC microprocessor design, including 88000 and PowerPC. He is currently design manager at the M*CORE Technology Center and has been working on the M*CORE architecture/microarchitecture for low-power, high-performance embedded applications and a low power design methodology.

Group Focus: The M*CORE Technology Center is focused on delivering low-power solutions to its customers which includes Motorola wireless, transportation, and consumer product groups as well as external customers. The M*CORE intellectual property is being design into almost all of Motorola's wireless products. The M*CORE technology center is focused on low power design techniques and tools, as well as architecture development.

Mark Bickerstaff, Bell Laboratories, Lucent Technologies:

Bio: Mark Bickerstaff is a member of technical staff in the Global Wireless Systems Research Department at Bell Laboratories, Lucent Technologies. He is based in Sydney, Australia. He received his BSc in Computer Science and BE in Electrical Engineering from the University of Sydney in 1985 and 1987, respectively. He received his PhD in Computer Science and Engineering from the University of New South Wales in 1994.

Prior to working at Bell Laboratories, Mark worked in the Electronics Department at Macquarie University, Sydney. Projects he has been involved in include a 1.3Gbps ATM cell switch fabric, the development of a standard cell generation system, the development of design flows for rapid layout of designs, an OFDM wireless modem, the development of structures for use in communications and the redesign of a low power calculator chip.

Mark is currently working on DSP systems for cellular mobile wireless infrastructure, in particular, investigating error correction systems.

Abstract: This session will attempt to overview digital signal processors, why they are useful, provide some examples, show how they have been optimized for their particular tasks, show how they apply to wireless communications systems, and show what future trends can be expected.

Frank Binns, Intel:

Bio: Prior to joining Intel, Frank held research positions with Marconi Research Laboratories and the Diamond Trading Company Research Laboratory both of the U.K. Frank has spent the last 15 years with Intel, initially holding technical management positions in the Development

Tool, Multibus Systems and PC Systems divisions. The last 6 years have been spent with the Performance Microprocessor Division as both a Staff Technical Marketing Engineer and most recently as a VLSI architect.

Group Focus: The Performance Microprocessor Division is responsible for the design and manufacture of processors positioned at the high performance end of the desktop and server market segments. PMD was responsible for the design of the P6 family architecture and development of derivatives now manufactured in high volume.

Abstract: This session will provide an overview of the features, requirements and capabilities of operating system power management as represented by the ACPI specification. The overall structure of ACPI and detail on the various sleep states that it defines will be discussed. The session also discusses the application of ACPI for both system and processor temperature control. The capabilities of ACPI with respect to processor temperature control will also be detailed.

George Z. N. Cai, Intel:

Bio: George Cai is working in the Intel Mobile and Handheld Product Group. He is working on microarchitecture and low power implementation for future high performance microprocessors. He has been in the microprocessor design and architecture fields for many years. He was in Honeywell Large Computer Division, IBM-Motorola Somerset Design Center, AMD, and TI for different mainframe computer and microprocessor projects. He received a BSEE from Jiao Tung University in China, and a MSEE and Ph.D. from University of Maryland at College Park.

Group Focus: Mobile computers, energy efficient architecture and low power implementation. Work with other Intel product and research groups to enabling power efficient design and implementation methods.

Abstract: The talk consists of two parts: (a) power/performance optimization is a new dimension of microprocessor architecture design; (b) dynamic power estimation for power/performance optimization.

In part (a), we use recent statistics and technology trends to identify the new challenges for microprocessor architects. High performance microprocessors are facing limits that cannot be solved with only traditional power reduction methods. New processing technology provides additional opportunities for high performance microprocessor to be energy efficient. We present several charts showing how interconnect correlates to transistor density and architecture decisions.

In part (b), we show that dynamic power estimation is critical for power/performance optimization. Dynamic power affects the performance and cost of high performance microprocessors in several ways. We use the SimpleScalar simulator as an example to show how to estimate the relative dynamic power consumption for microarchitecture studies. The fundamental difficulty of architectural level power/performance optimization is that architectural decisions have to be made early in the design process, while accurate power estimation are only available latter in the design process. Therefore, the relative power consumption estimation and

interactions among all design phases becomes very important. Power/performance optimization brings new challenges and new opportunities to microprocessor architects.

Deo Singh, Intel:

Bio: Deo manages the Low Power Design Technologies Group in the Intel DT division. He is responsible for developing technologies (micro-architecture, logic and circuit techniques, and tools) to reduce power in Microprocessor products. He has some 20 years of experience in CAD and design. During the last 7 years, he has been working on low power. He received Diplomas in EE and an MSc in CAD in the UK.

Group Focus: To reduce power in Intel Microprocessor products, including processors and chip-sets. The Low Power group is engaged in developing low power optimizations spanning circuit design through microarchitecture and software.

Abstract: Power is the performance limiter in the next generation of microprocessors. The talk will summarize the current situation in power and what has been accomplished to reduce power. This will be followed by a discussion on the new industries driving the Internet and the ramifications on power in Internet and high-performance servers. Finally, the talk will conclude with specific recommendations and areas of work for low power research and development.

Kathryn Wilcox, Compaq:

Bio: Kathryn (Kuchler) Wilcox is a Principal Hardware Engineer in the Alpha Development Group at Compaq Computer Corporation in Shrewsbury, MA. She joined the Alpha design team as a circuit designer on the Alpha 21064 microprocessor. On the Alpha 21264, she led the execution unit implementation team. Currently, she is leading the implementation team of the instruction mapper unit for the next generation Alpha microprocessor. She received the B.S. in electrical engineering from Cornell University in 1990.

POWER CHALLENGES IN THE INTERNET WORLD

Deo Singh

Vivek Tiwari

Low Power Design Technology

Intel Corp.

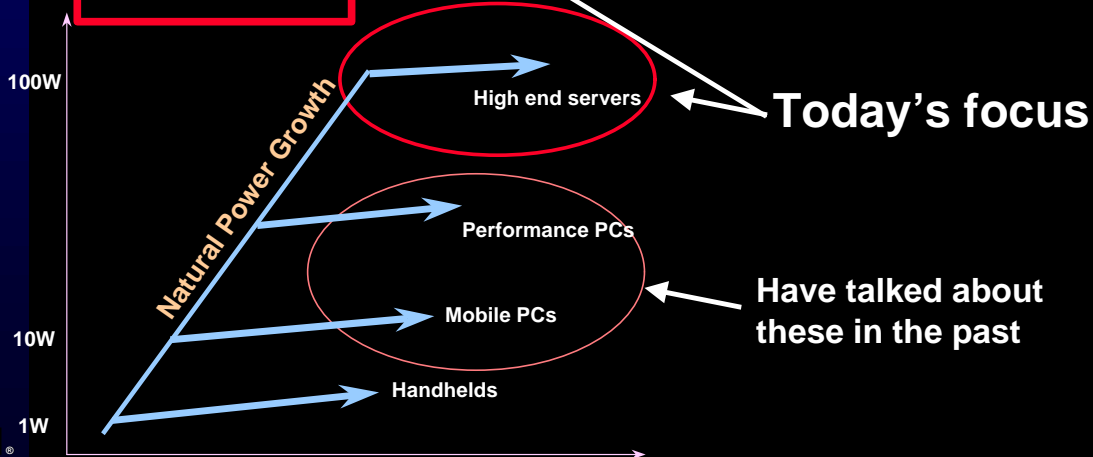
November, 1999



"Third party marks are the property of their owners"

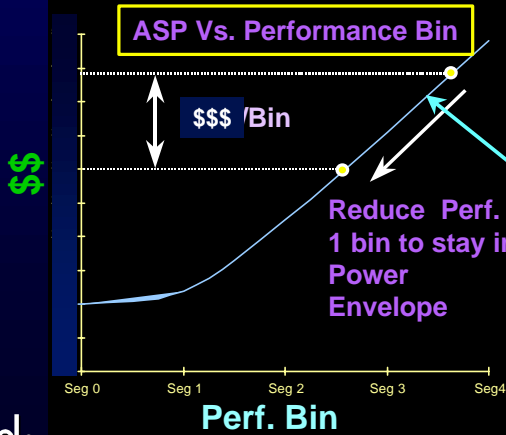
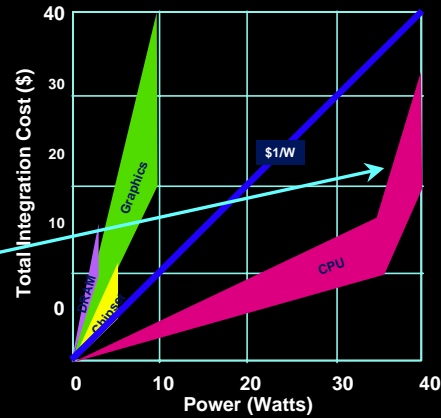
Power challenges per segment

	Servers	Desktops	Mobile	Handhelds
Power related system cost drivers	Thermal cost (\$)	Thermal cost (\$)	Thermal cost (\$)	Form factor (in ³)
	Delivery cost (\$)	Delivery cost (\$)	Delivery cost (\$)	Battery size (lbs.)
	Form factor (in ³)		Form factor (in ³)	Battery cost (\$)
Price drivers	Perf (SPEC, TPC-C)	Perf (SPEC, MHz)	Perf (SPEC, MHz)	Perf (MIPS, MHz)
	Perf/in ³	Perf/\$	Perf/lb.	Perf/battery hrs



What did we tell you before?

- The cost of power in Desktop PCs
 - Thermal & power dstbn cost
 - Every CPU Watt over 40 W = \$1/Watt



- Impact of power reduction
 - If reduce freq. to maintain 40W:
 - Loss of 1 Perf Bin => \$\$\$ on ASP

intel®

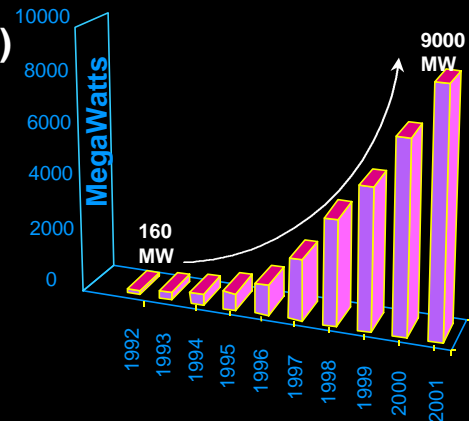
3

Environmental burden of CPUs!

- Total power consumption of CPUs in world's PCs:
 - 1992: 160 MWatts (87M CPUs)
 - 2001: **9,000 MWatts** (500M CPUs)
- That's 4 Hoover Dams!



Courtesy: United States Department of the Interior
Bureau of Reclamation - Lower Colorado Region

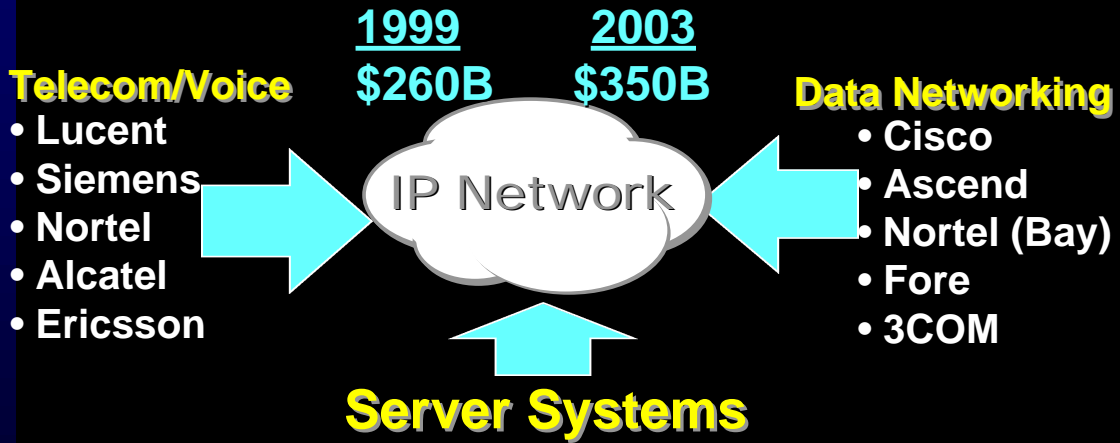


[Source: Dataquest (for installed base) + estimates for avg. installed CPU power]
Projected with PentiumII™ Power

intel®

Andy's vision: 1 Billion Connected PCs!

The New World Order: Internet & Communication are merging



The Inflection Point Creates An Opportunity for The Computing Industry

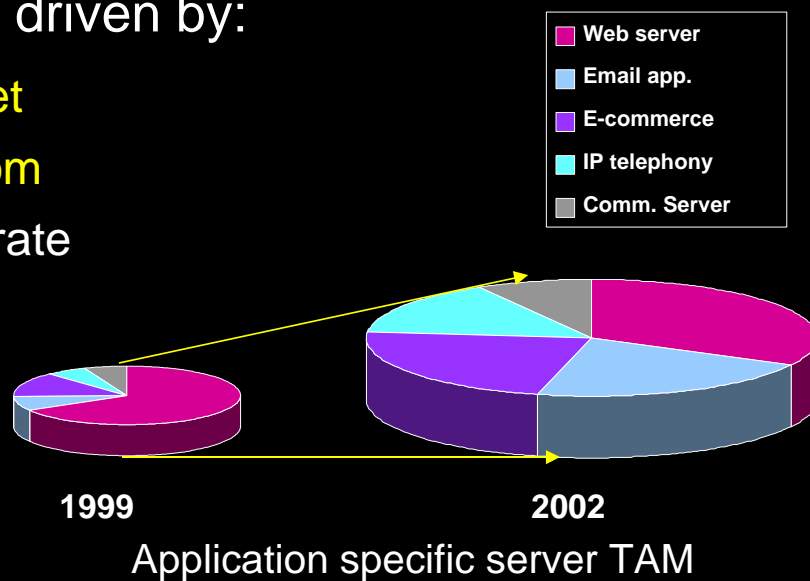
intel

Source: DataQuest, Piper Jaffrey

5

Server market segment trends

- Sever market Segment will show strong growth
- Strongly driven by:
 - Internet
 - Telecom
 - Corporate

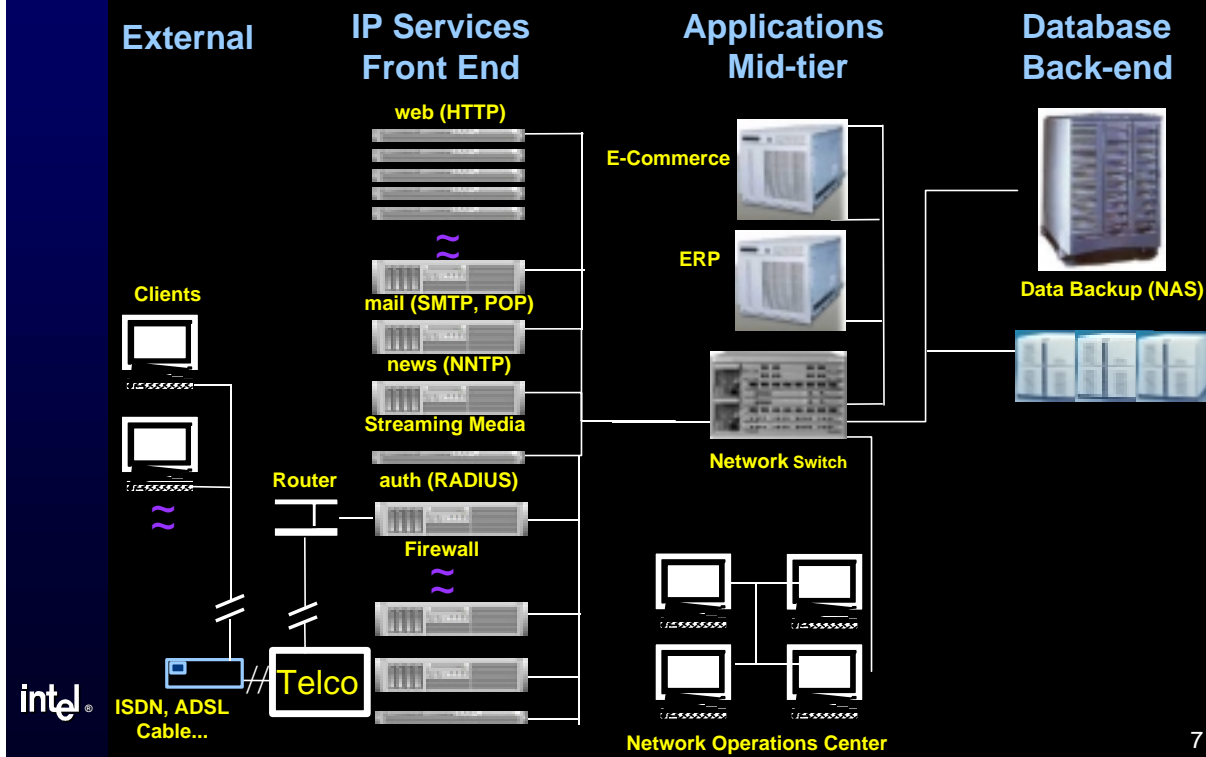


intel

(Intel Estimates)

6

New World Order Case-study: Large ISP server topology



7

What are the costs to an ISP?

- Internet data centers are like heavy-duty *factories*
 - E.g. Data Center 25,000 Sq. ft., ~8000 servers, ~2,000,000 Watts !!
- Want lowest *net cost per server per sq foot of data center space*
- Cost driven by:
 - Racking height
 - Cooling air flow
 - Power delivery
 - Maintenance ease (access, wt.)

“Behind all of this, power is the lead cost driver in the facility - about 25% of the total cost of a data center”

Data Center Facilities Mgr.

“They are concerned about power because it increases the weight of the node due to massive heat sinks. Weight is very critical for hot swaps.”

Customer quote

intel®

8

How does this drive Internet server requirements?

	IP Services Front End	Applications Mid-tier	Database Back-end
Application	Firewall, Web, Mail, News	Ecommerce, ERP	Database, Dir., NAS
ISP needs	Reliability Perf. (SpecWeb) Form factor (<1U) Power Cost Hot swap	Reliability Perf. (Spec,TPC-C) Cost Power Form factor	Reliability Perf. (TPC-C) Mem addr space Power Cost
Design challenge	Max performance for given form factor	Max performance at perf/size/cost balance	Most performance at leading edge tech.
<p>Key design challenges for CPU designers</p>			

intel®

9

Internet server examples

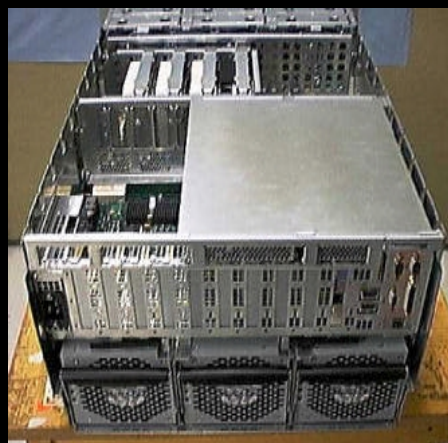
- Cobalt* High Density ISP Server

- *Front end server*
- 1U ff (*Up to 40 in one rack*)
- 250MHz 64-bit CPU
- 256MB mem, 16.8 GB disk
- Perf. metric: Web trans./sec
- **35 Watts**



- Intel AC450NX System

- *Back-end, Enterprise*
- 7U ff
- 1-4 Pentium III Xeon™ 500 MHz
- up to 8GB ECC mem
- Perf. metric: TPC-C
- **3 Power supplies - 420 Watts ea.**



intel®

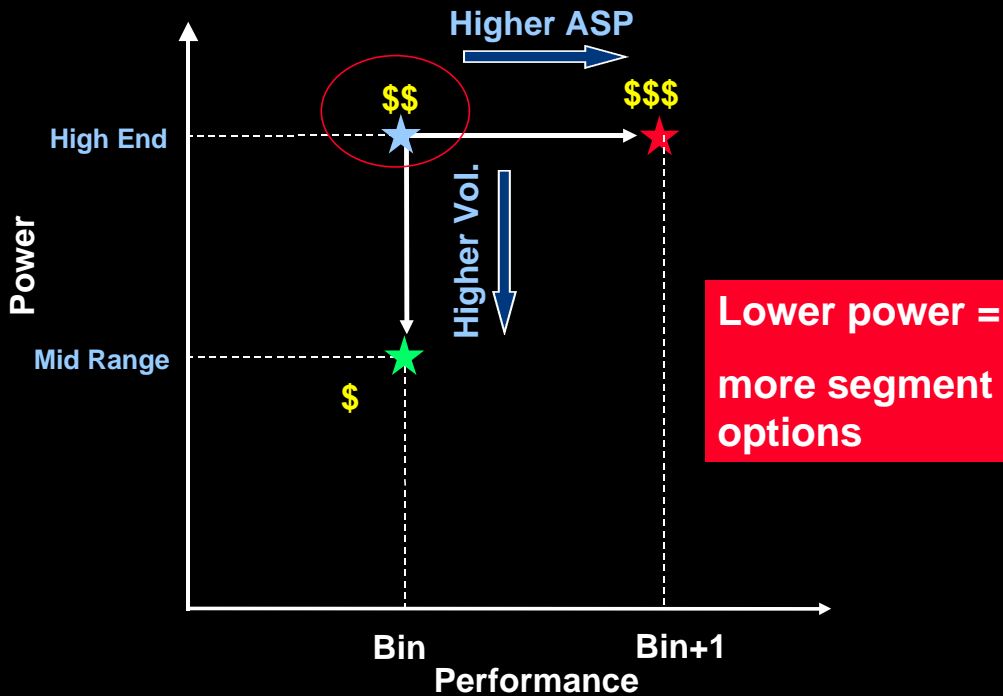
* Other brands and names are property of their respective owners

<http://intel.com/design/servers/AC450NX/>

10

Doing business in a power-limited world

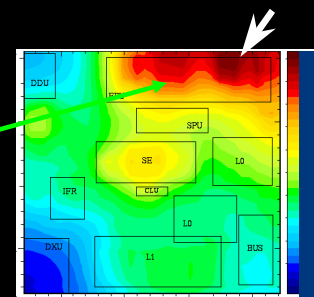
- Two vectors for higher revenue
 - Both require power reduction



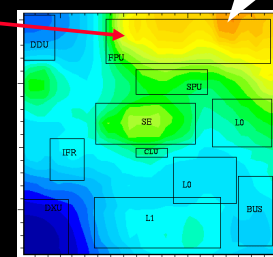
11

Summary - power reduction motivation and key message

- Traditionally measure ourselves by **perf** or **price-perf**, but now its also **perf/foot³**
- Power is THE key driver for **foot³**
- **Watts/foot³** matter
- Hot-spots also limit allowed perf
- Power reduction helps perf
- **Watts/mil²** matter



CPU thermal maps



*Power reduction is more critical than ever
We need your help!*

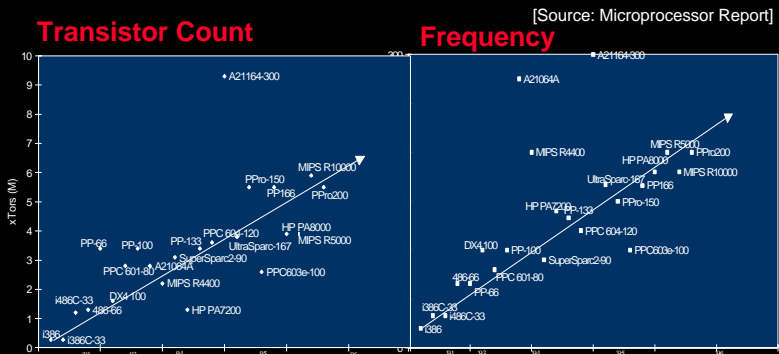
intel

12

13

Future Design Challenges

- The traditional path to perf.
 - 2x devices, 2x MHz per generation
 - **2x Power**



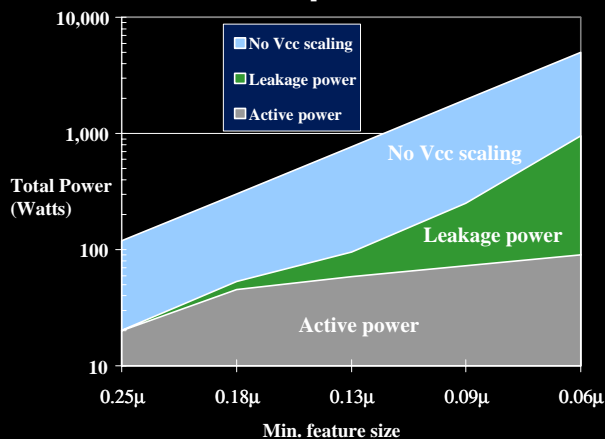
- To continue industry's performance ramp:
 - Need to adopt "Power-Aware" design in **all product segments** - not just in mobile/handhelds
 - Need to design for **power AND performance** at all levels - uArch to Ckts

Low Power Research IS High Performance Research



Key Research Opportunities: Fundamental circuit techniques

- Enable continued Vcc reduction
 - Enable development of ultra low voltage circuits
 - New logic families, multi threshold (Vt), dynamic Vt etc.
- Efficient leakage control
 - Leakage is catching up with switching power
- Ckt. families for future process technology limitations



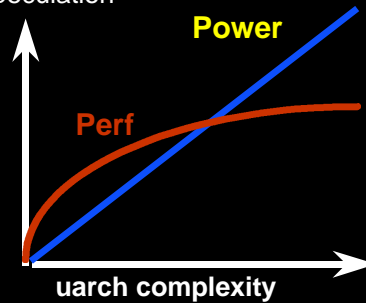
First order analysis, using constant field scaling.

Source: Shekhar Borkar, Intel Corp.



Key Research Opportunities: High-level (arch/uarch) design

- **Breakthrough machine organizations are needed**
 - *Diminishing performance improvements from uarch*
 - E.g. N-way superscalar does not give speedup of N, but power goes up by factor of N
 - Increasing levels of speculation: prefetches, out-of-order-issue, branch prediction, data-speculation



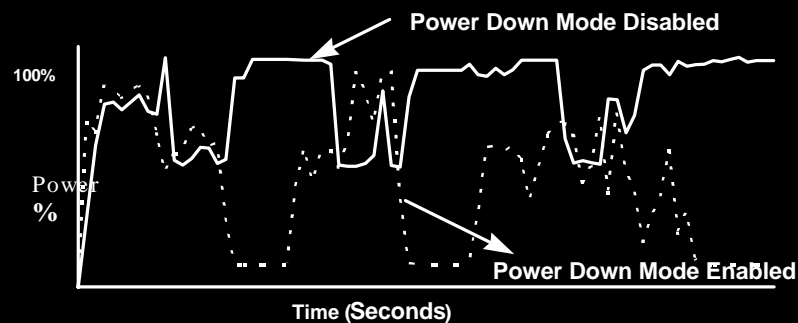
- **Best way to use a Billion xtors for power and perf?**
- **Enhanced modeling capabilities to quantify power-perf. Space**

intel®

15

Key Research Opportunities: Design for power delivery

- **Power delivery may be the limiter ahead of thermals**
 - *More devices, higher frequency, higher currents*
 - *Dynamic power mgmt and clock gating make things worse*



- **Modeling of the tech. parameters (L, di/dt, C) and costs**
- **Design at all levels (systems, pkg., uarch, logic, ckts)**

intel®

16

15



Alpha Processors: A History of Power Issues and A Look to the Future

Kathryn Wilcox and Srilatha
Manne
Compaq Computer Corporation



Alpha Processor Comparisons

	Power (Watts)	Freq. (MHz.)	Die Size (mm ²)	Vdd
Alpha 21064	30	200	234	3.3
Alpha 21164	50	300	299	3.3
Alpha 21264	90	575	314	2.2
Alpha 21364	>100	>1000	340	1.5
Alpha 21464	125-150	1000- 2000	NA	1.2

Compaq Corporation





21264 Microarchitecture

- Four-wide Instruction Fetch
- Line and Branch Predictor
 - Tournament prediction using both *local* & *global* history
 - Dynamic JSR/JMP prediction
- Out-of-Order Execution Pipelines
 - Quad-speculative-issue integer pipeline
 - Dual-speculative-issue floating-point pipeline
 - One ADD, One MULTIPLY
 - Also DIV (6 bits/cycle) and SQRT (2 bits/cycle)



Compaq Corporation



21264 Microarchitecture

- Memory System
 - Two out-of-order memory references per cycle
 - Up to 16 outstanding off-chip memory references
- 80 In-flight Instructions
- Registers: 80 Integer, 72 Floating Point
- Queue Entries: 20 Integer, 15 Floating Point



Compaq Corporation



21264 Microarchitecture

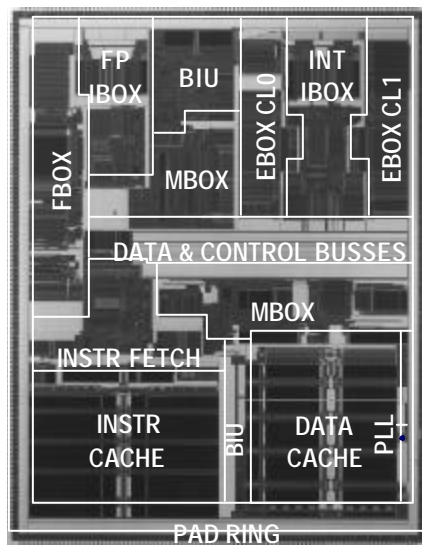
- 128-Entry Fully Associative ITB and DTB
- 2-Way 64KB L1 On-Chip Instruction and Data Caches
- Instruction extensions
 - Motion video (MVI)
 - FP square root
 - FPU integer register transfers



Compaq Corporation



21264 Die Photo & Floorplan



Compaq Corporation



Box Definitions

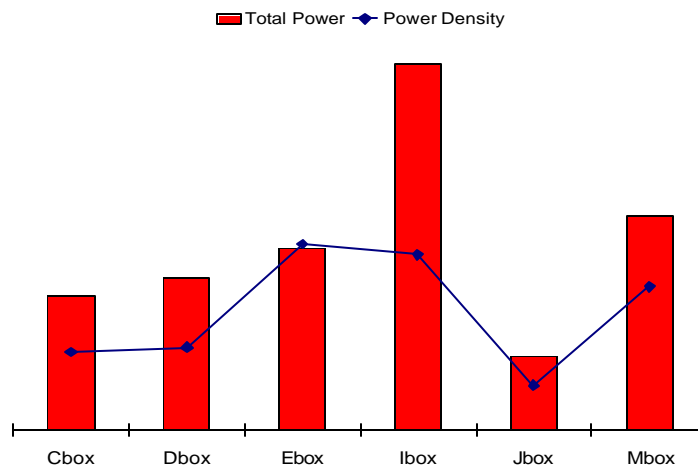
- CBOX: BIU, Data & Control Busses
- DBOX: Dcache
- EBOX: Integer Units (L, R)
- IBOX: Integer Mapper & Queue, FP Mapper and Queue, Instruction Data Path
- JBOX: Instruction Cache
- MBOX: Memory Controller (Load Queue, Store Queue, DTB, and Miss Address File)



Compaq Corporation



21264 Power Density/ Unit

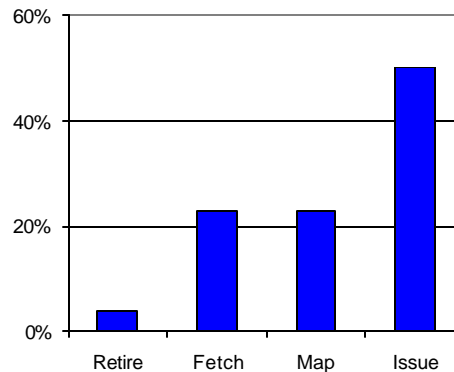


Compaq Corporation



21264 Ibox Power Breakdown

- ~50% of the Ibox power was consumed in the queue
- ~23% was consumed in the mapper and the early datapath



Compaq Corporation



21264 Queue Architecture

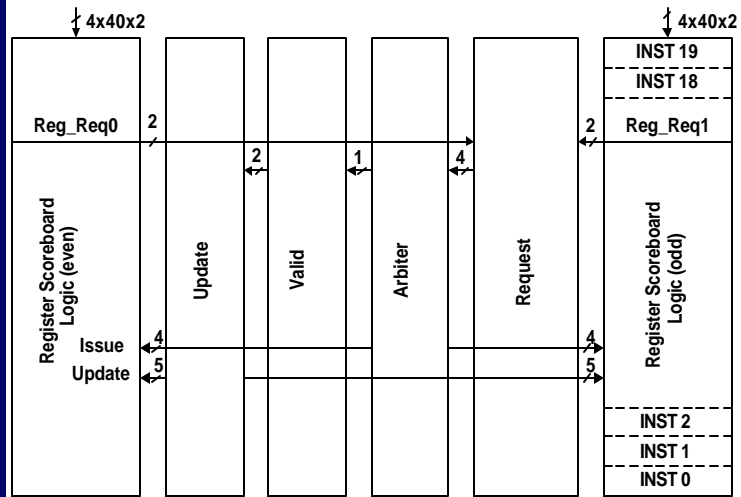
- 20 entry queue allows any instruction to issue to any execution unit
- Two pick-2 arbiters determine four issue signals in parallel
- Up to 4 instructions can enter queue each cycle
- Up to 4 instructions can issue from queue each cycle
- Queue entries are compacted such that the oldest entries are located at the bottom of the queue and the new instructions enter at the top of the queue



Compaq Corporation



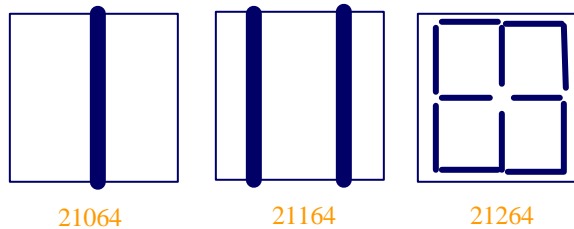
21264 Queue Datapath



Compaq Corporation



Alpha Clocking



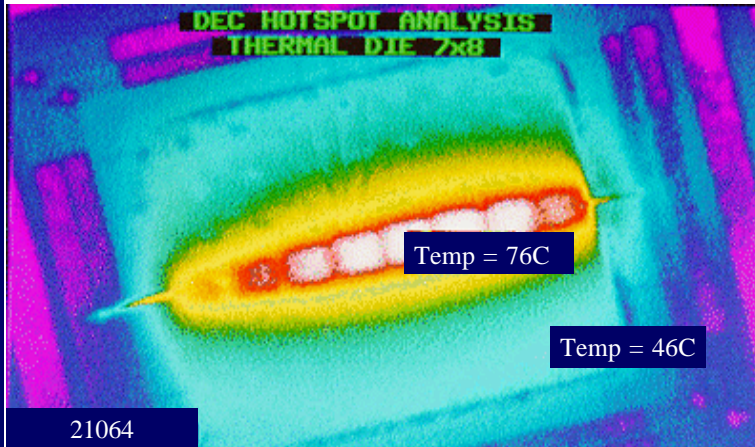
- Clock power is ~40% of the total power on the 21064, 21164, and ~30-40% of the total power on the 21264.
- One clock wire was generated from one or two clock spines.
- On the 21264, a window pane configuration was used to distribute the clock power across the chip.



Compaq Corporation



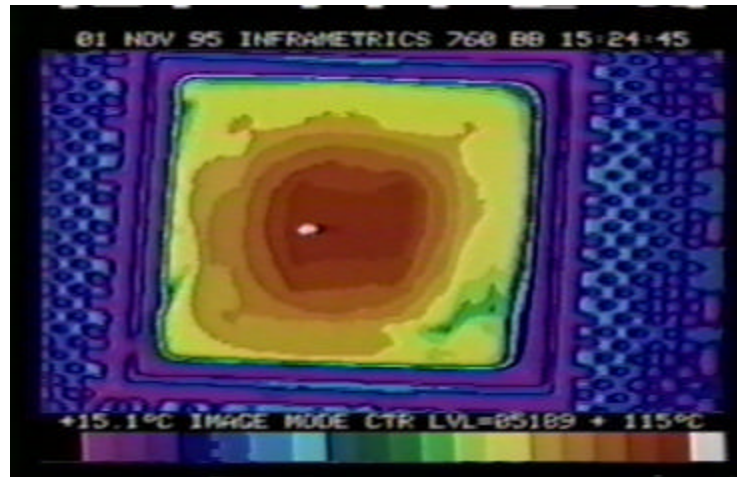
21064 Thermal Plot



Compaq Corporation



21164 Thermal Plot

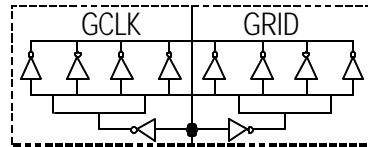


Compaq Corporation

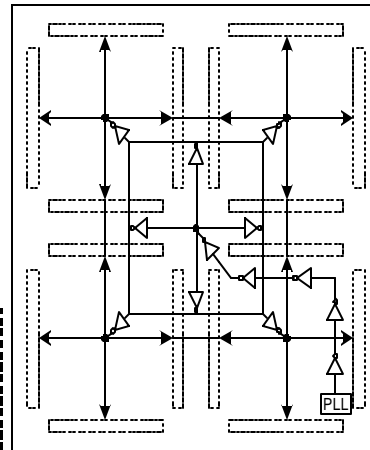


21264 Global Clock (GCLK) Distribution

- Reduced Power Supply Collapse
- Uniform die temperature



Final GCLK Driver
Compaq Corporation

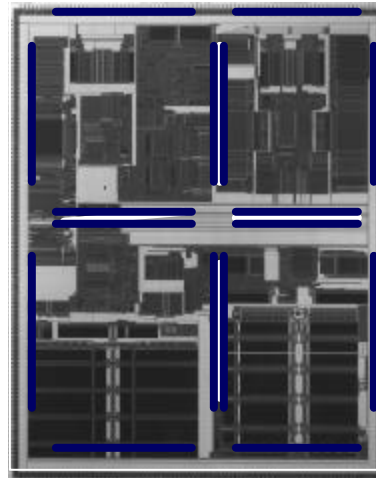


GCLK Distribution Network



21264 Global Clock Distribution – Results

- 3% Metal 3 and Metal 4 Usage
- GCLK Capacitance
 - 1.4nF – Interconnect
 - 0.6nF – Gate Loading
 - 0.2nF – Driver Parasitics
- GCLK Power @ 600MHz, 2V
 - Final Driver & Grid Only
 - 8.1W – Total
 - 2.8W – Crossover
 - GCLK Distribution Network
 - 12.9W – Total



Compaq Corporation Final GCLK Driver Locations





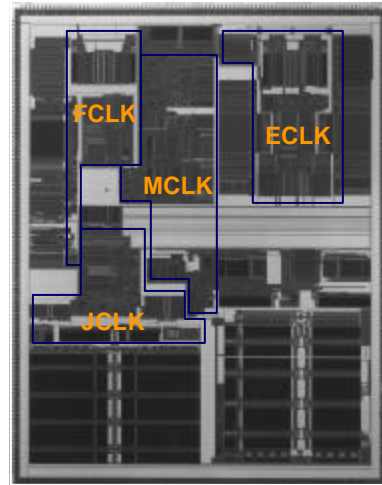
Global Clock (GCLK)

- *Conditioned Local Clocks*
- *Buffered Local Clocks*
- *Conditioned Section Clocks*
- *Buffered Section Clocks*
- *Conditioned Local Clocks*
- *Buffered Local Clocks*

Advantages

- Power Saved in Idle Functional Units
- More Clocking Options for Design
- Reduced Section Clock Skew
- Metal 3 & 4 Usage Reduced by > 3X

21264 Clock Hierarchy



Buffered Section Clock Regions



Better answers

Compaq Corporation



21264 Local Clocking Strategy

- Multiple major section clocks
 - 7 section clocks
 - double buffered off of the global clock
 - total cap ~ 3.0nF
- Local clocks
 - unconditional off of GCLK or major section clocks
 - ~1500 unconditional clocks
 - conditional
 - total cap ~ 6.0nF



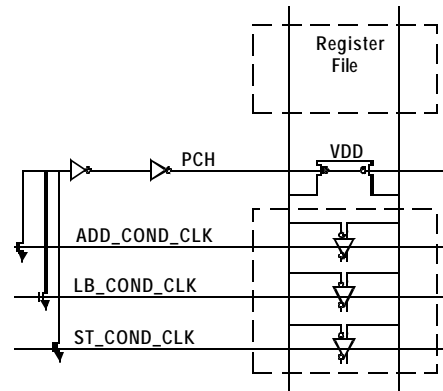
Better answers

Compaq Corporation



21264 Functional Units

- Conditional clocks were used widely in the integer and floating point units.
- Some additional power was burned to disable the units when not in use.



Integer Unit Conditional Clocking



Compaq Corporation



Signal busses

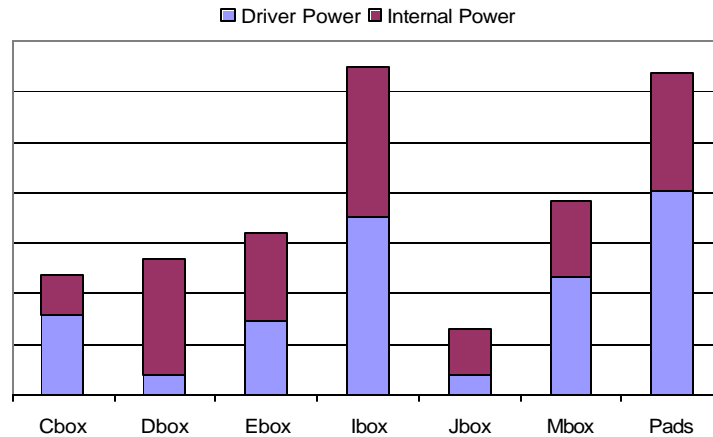
- Total signals: ~2.5M
- Total inter-schematic signals: 80K
- Global signals: 9K
 - crosses a layout partition or box partition boundary
- Used both full-swing and low-swing buses throughout the chip.



Compaq Corporation



21264 Power Distributions



Compaq Corporation



21264 Latches

- Our methodology used a schematic and layout library of edge-triggered latches.
- Advantage: 1 latch/cycle
- Disadvantage: power independent of data
- For the smallest sized latches, a lower power version of the latches existed.
- There were ~60,000 latches on the Alpha 21264.
 - ~70% of the total were library cell latches
 - ~25% of the total were low power latches



Compaq Corporation



21264 Conditional Latches

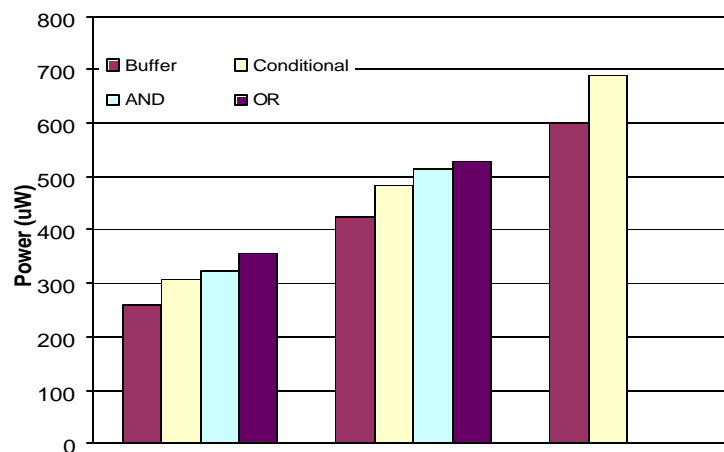
- Conditional latches were also included in the latch library.
- These latches burned slightly more power when they were enabled.
- Since these were added late in the design process, only ~3% of the latches were conditional.
- ~40% of the conditional latches were in the pads.



Compaq Corporation



Latch Power Comparison



Compaq Corporation



Short Circuit Current

- Another component of power dissipation is the cross-over current (short circuit current).
- Power Mill was run on some sections of the chip to determine the cross-over current.
- Some re-design of clocks was done where the power due to cross-over was deemed excessive.



Compaq Corporation



Decoupling Capacitor Ratios (di/dt)

- 21064
 - total effective switching cap = 12.5nF
 - 128nF of decoupling capacitance
 - dcap/switching cap ~ 10x
- 21164
 - 160nF of decoupling capacitance
 - 13.9nF of switching cap
- 21264
 - 320nF of decoupling capacitance
 - 34nF of effective switching cap



Compaq Corporation

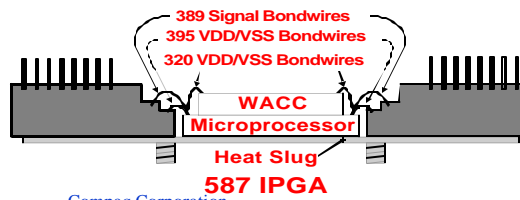


21264 Decoupling Cap

Design for D $I_{DD} = 25$ A @ $V_{DD} = 2.2$ V, $f = 600$ MHz

- Added 0.32- μ F of on-chip decoupling capacitance
 - Located beneath major busses and around major gridded clock drivers
 - Occupies 15-20% of die area
- 1- μ F 2-cm² Wirebond Attached Chip Capacitor (WACC) significantly increases “Near-Chip” decoupling
 - 160 VDD/VSS bondwire pairs on the WACC minimize inductance

389 Signal - 198 VDD/VSS Pins



Compaq Corporation

COMPAQ
Better answers



Packaging Issues

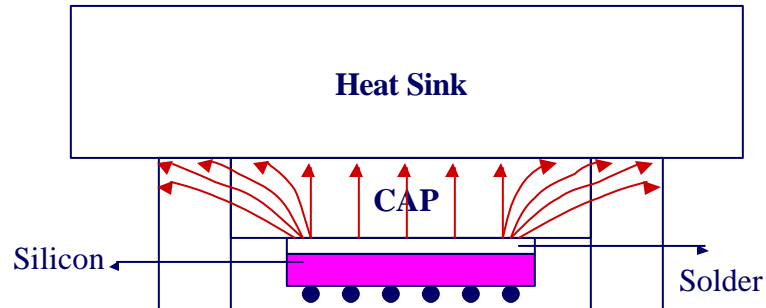
<i>Device</i>	<i>Power (W)</i>	<i>Package Cooling (C/W)</i>	<i>System Cooling (C/W)</i>	<i>Total Cooling (C/W)</i>
EV4	25-30	1.1	0.7-1.4	1.8-2.5
EV45	30-36	0.7	0.7-1.4	1.4-2.1
EV5, 56	50-60	0.4	0.4-0.8	0.8-1.2
EV6,67	70-95	0.3	0.3-0.5	0.6-0.8
EV7	>100	0.3	0.2	0.5

Compaq Corporation

COMPAQ
Better answers



Packaging Implication on Architecture



Compaq Corporation



Package Selection

- What does it mean when a package is rated to 100W?
- The package can sustain a power dissipation of 100W for up to ~100msec.
 - Short peaks are not a problem.
 - A cheaper package is possible if duration of peak power can be reduced.



Compaq Corporation



Package Summary

- Temperature is a function of power density.
- Chip can tolerate about 20C temperature differential.
- Want hot spots near the edge of the chip.
- Passive package can tolerate 125-150W maximum.
- EV9 packaging???



Compaq Corporation



Power based on Activity

- Chip power consumption is based on the actual code and data being run.
- Initially, each box determined their worst case power assuming that code existed that could create a worst case power scenario.
- Unfortunately, adding these numbers together doesn't indicate overall chip power.



Compaq Corporation



Activity Factors

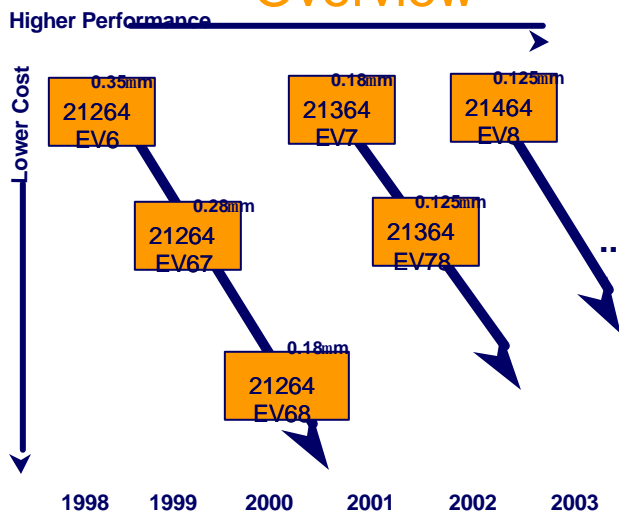
- A program was written to exercise as much of the chip as possible and target the worst case power.
- To determine signal activity factors, the model needs to be simulated and the toggle information extracted.
- The nodes from the model are mapped to the schematics to associate the activity factor with the capacitance.



Compaq Corporation



Alpha Microprocessor Overview



First System Ship
Compaq Corporation





21464 Architecture

- Enhanced out-of-order execution
- 8-wide superscalar
- Large on-chip L2 cache
- Separate L1 Inst and Data caches
- Direct RAMBUS interface
- On-chip router for system interconnect
 - for glueless, directory-based, ccNUMA
 - with up to 512-way multiprocessing
- 4-way simultaneous multithreading (SMT)
 - <5% additional area



Compaq Corporation



21464 Technology

- Power target: < 150 Watts tolerated peak
- Clock frequency range 1.0-2.0GHz
- Leading edge 0.125 μ m CMOS technology
 - SOI Compatible
 - Cu Interconnect
- ~1.2V Vdd
- ~250 Million transistors
- ~1100 signal pins in flip chip packaging



Compaq Corporation



Goals

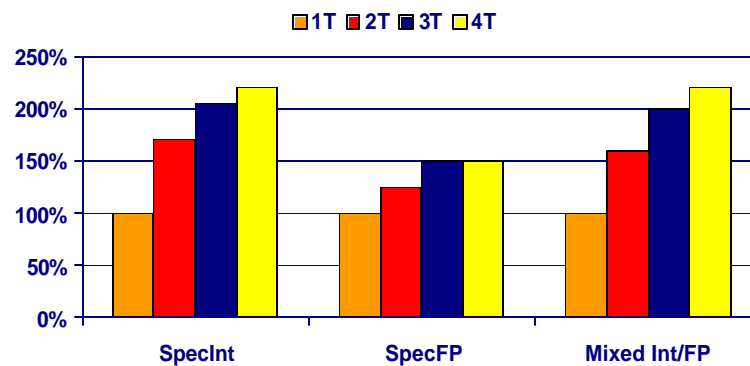
- Leadership single stream performance
- Extra multistream performance with multithreading
 - Without major architectural changes
 - Without significant additional cost



Compaq Corporation



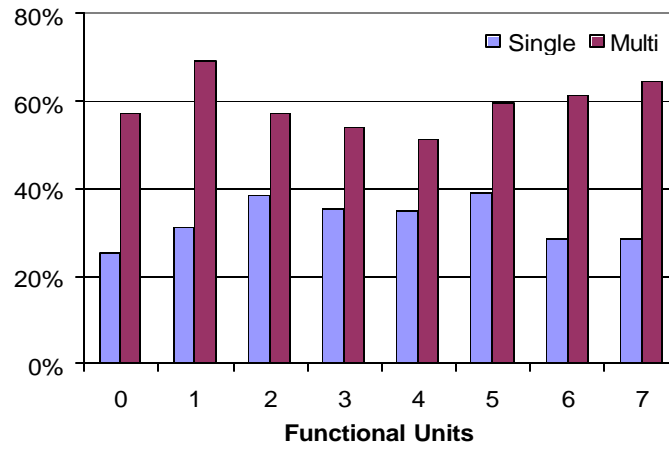
Multiprogrammed workload



Compaq Corporation



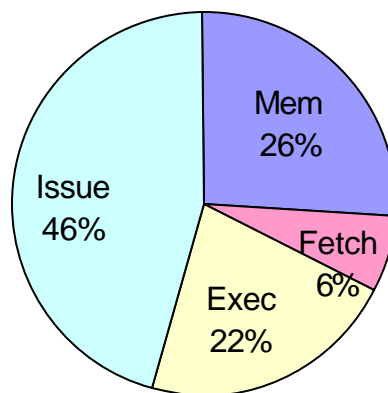
21464 Resource Utilization



Compaq Corporation



Estimated 21464 Power



Compaq Corporation



Future Trends

- Emphasis on reducing clock load since clock is a significant portion of the total power consumption
- More emphasis on a low power latch library
- Earlier understanding of switching activity based on real CPU activity



Compaq Corporation



Future Trends

- Improved tracking of power consumption throughout the project
- Integrating power tradeoffs into the early architectural feasibility
- Low power modes



Compaq Corporation



The M•CORE Technology Center: Designing a Low-Power **Solution**

John Arends, Bill Moyer,
Jeff Scott, Lee Hwang Lee

Motorola M•CORE Technology Center
{arends, billm, jscott, leahwang}@lakewood.sps.mot.com

MOTOROLA

 **DigitalDNA**
from Motorola



Presentation Outline

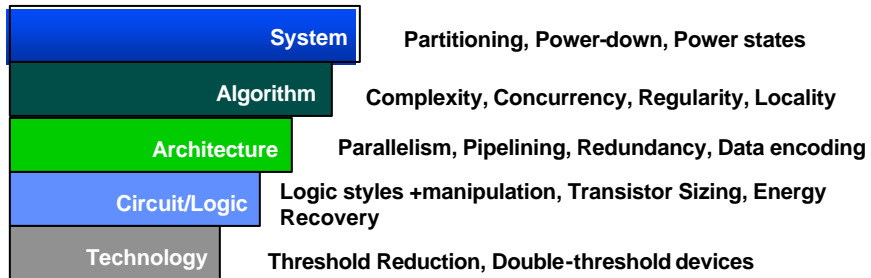
- **Introduction**
- Architecture and Microarchitecture for Low Power
- Memory System and Subsystems for Low Power
- Clock Gating and Clock Tree
- Process Technology and Circuit Techniques
- Tools for Low Power
- Conclusion

MOTOROLA

 **DigitalDNA**
from Motorola



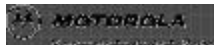
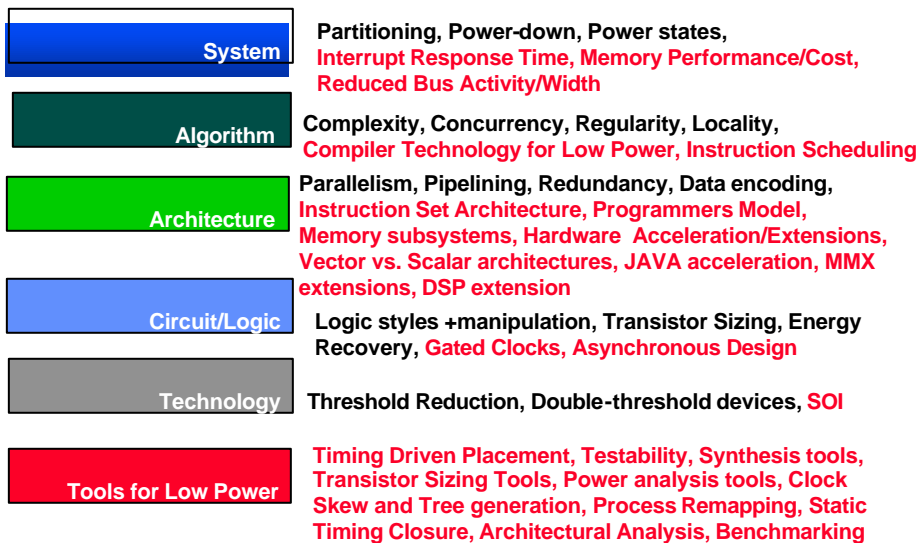
Introduction



“An integrated low power methodology requires optimization at **ALL** design abstraction layers - Jan Rabaey, Massoud Pedram, Paul Landman Low Power Design Methodologies, 1996.

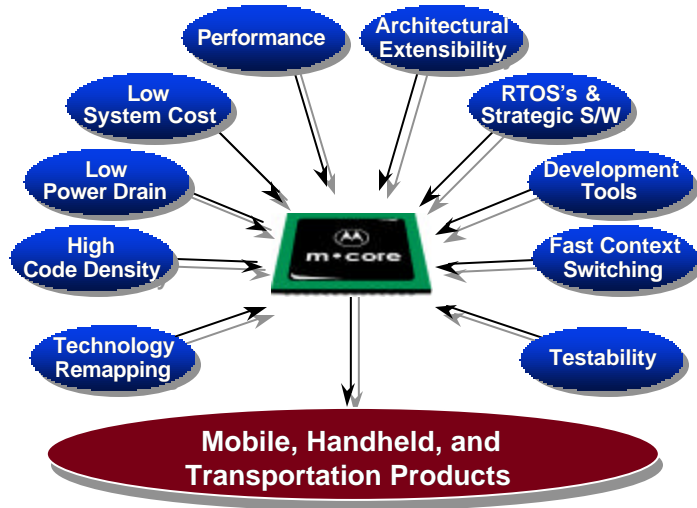


Introduction





Introduction



Presentation Outline

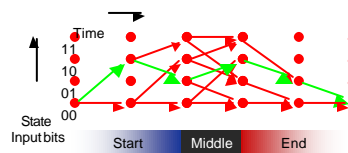
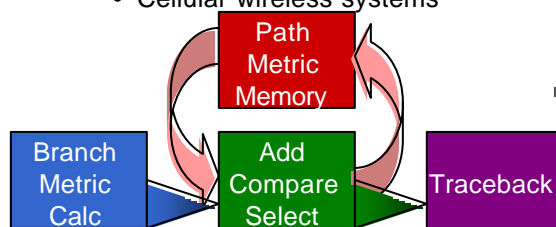
- Introduction
- **Architecture and Microarchitecture for Low Power**
- Memory System and Subsystems for Low Power
- Clock Gating and Clock Tree
- Process Technology and Circuit Techniques
- Tools for Low Power
- Conclusion



Reception - Decoding (1)



- Convolutional Codes/Decoders (Viterbi)
 - Convolutional codes provide a form of forward error correction (FEC) so that the integrity of data can be guaranteed to a particular level of limit.
 - Applications include:
 - Read/Write channels for hard disks
 - Satellite communications
 - Cellular wireless systems



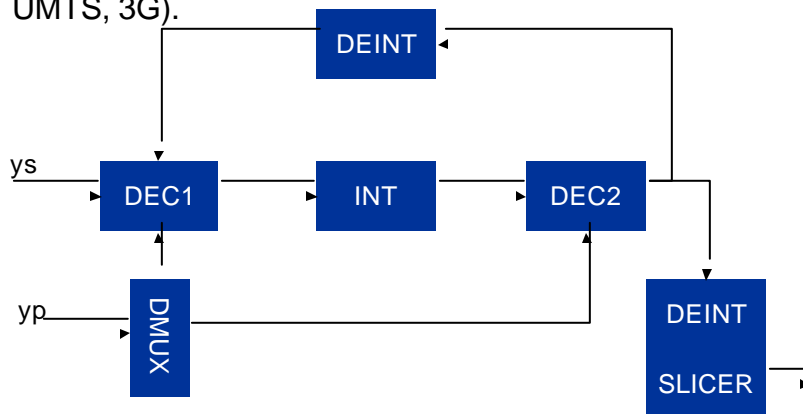
Micro 32 Cool Chips Tutorial - MAB

Bell Labs / Lucent Technologies

Reception - Decoding (2)



- Turbo
 - Satellite communications
 - Cellular data services for future implementations (e.g.: UMTS, 3G).



Micro 32 Cool Chips Tutorial - MAB

Bell Labs / Lucent Technologies

Low Power Design

Lucent Technologies
Bell Labs innovations



- ➔ Low power design is 'new' art - dating from '94
- ➔ Biggest gain from reduction in process supply voltage
- ➔ Radical approaches:
 - ultra-low supply voltage
 - adiabatic (energy restoring) circuits
- ➔ Conservative approaches:
 - full custom design
 - clock gating
 - minimizing data transitions
 - partitioned memory architectures
 - mixed voltage/threshold circuits

Micro 32 Cool Chips Tutorial - MAB

Bell Labs / Lucent Technologies

LP design techniques - arch & alg

Lucent Technologies
Bell Labs innovations



- Minimization of power consumption via hardware aware compilation. [Tiwari et-al]
- Complex instructions - improve code density - less chip accesses.
- Esoteric on-chip functional modules improve power and performance but are difficult to target by a compiler.
- Special modes in Programmable DSPs for filtering Algorithms (e.g. circular buffer addressing modes) = Tight loop control.
- Duplicate register file reduces cost of context switch
- In Booth-recoded MAC: $\text{Power}(A \times B) \neq \text{Power}(B \times A)$ [Kojima et-al: ISLPED96]
- Variable cycle multiplier optimizes for small operands

Micro 32 Cool Chips Tutorial - MAB

Bell Labs / Lucent Technologies

LP Design Techniques - Clock Gating

Lucent Technologies
Bell Labs innovations



- Instruction level clock gating (e.g. multiplier in 16210)
- Module level under program control (e.g. I/O blocks)
- Memory wait state disables processor clock (CPP)
- Power Control Register (e.g. 1609):

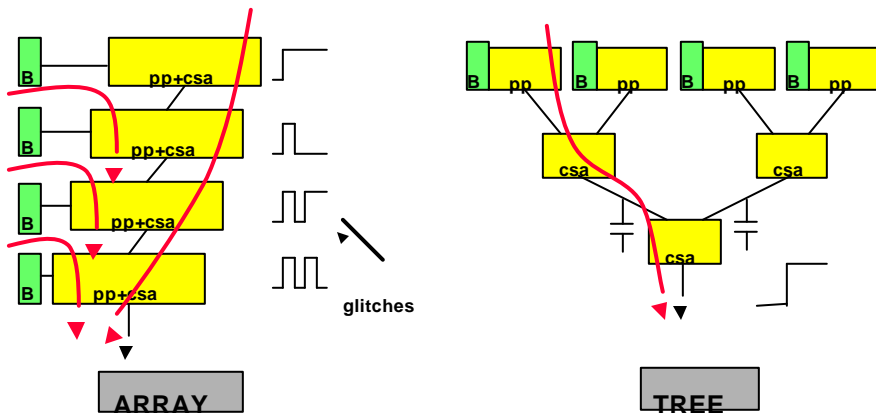
Operation Mode	Power (3.3V)
Normal (80 MHz)	120 mW
Standby (Halt)	21 mW
Slow Clock (16 kHz)	2.3 mW
StopClk	30 μ W

Micro 32 Cool Chips Tutorial - MAB

Bell Labs / Lucent Technologies

Minimizing Data Transitions

Lucent Technologies
Bell Labs innovations



Micro 32 Cool Chips Tutorial - MAB

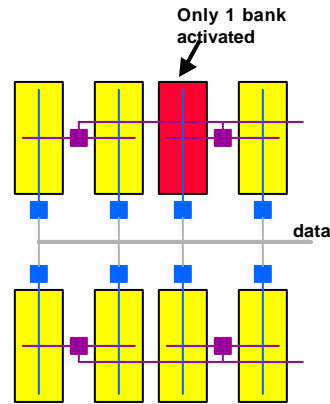
Bell Labs / Lucent Technologies

LP design techniques - SRAM

Lucent Technologies
Bell Labs innovations



- Sub-banking
- Address-driven gated clocking
- Hierarchical word-lines
- Hierarchical bit-lines
- Bit-line Isolation sense amps
- Reduced-swing data busses



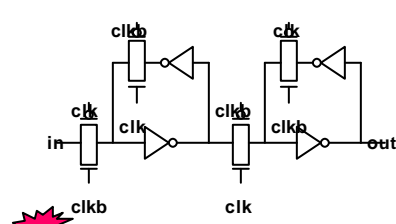
[Itoh, Survey in Proc. IEEE, May 1995]

Micro 32 Cool Chips Tutorial - MAB

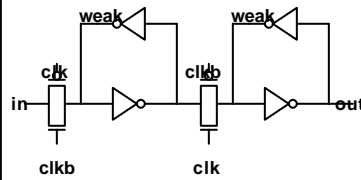
Bell Labs / Lucent Technologies

LP design techniques - flip-flops

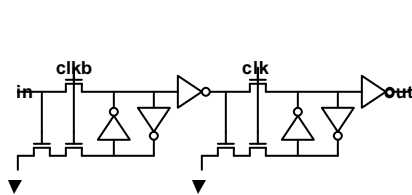
Lucent Technologies
Bell Labs innovations



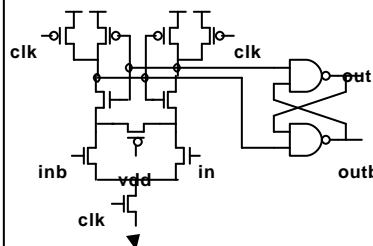
Standard CMOS Flip-Flop



"Low Area" Flip-Flop



Lee et-al, ISSCC97



Edge-Triggered Latch

Micro 32 Cool Chips Tutorial - MAB

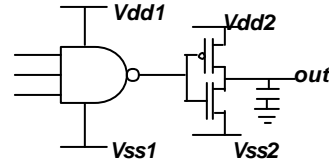
Bell Labs / Lucent Technologies

Technology & Voltage Scaling

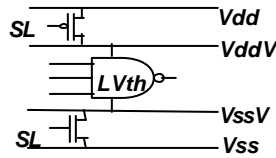
Lucent Technologies
Bell Labs innovations



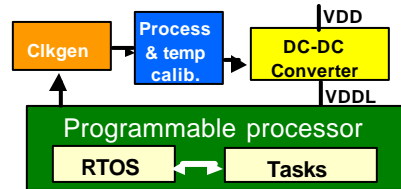
- Mixed supply voltage
 - e.g. Quadrail [CMU]



- Multiple thresholds
 - e.g. MTCMOS [NTT]



- Dynamic voltage scheduling



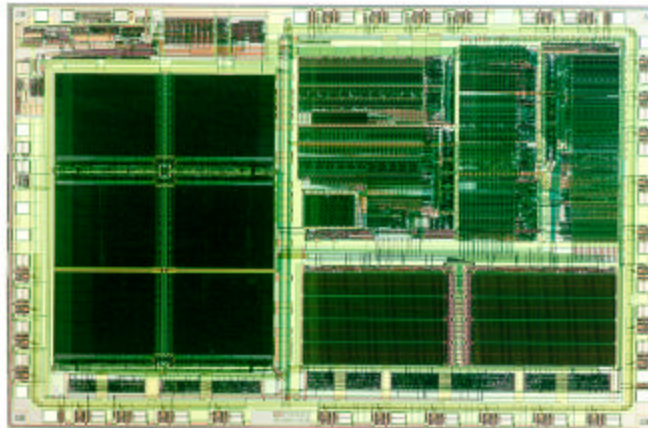
Micro 32 Cool Chips Tutorial - MAB

Bell Labs / Lucent Technologies

DSP 1600 Core

(Lucent - 1609 low cost consumer 16-bit)

Lucent Technologies
Bell Labs innovations



- 0.35 μ 3LM CMOS
- 80 M 16b MAC/s at 3.3V
- 1.4 mW/MHz at 3.3V
- 30 μ W stand-by power

Micro 32 Cool Chips Tutorial - MAB

Bell Labs / Lucent Technologies

What does the future hold?

Lucent Technologies
Bell Labs innovations



- DSP engineers are now taking notice of prior art in computer architecture field.
- Caches vs SRAMs
- SIMD structures
- MIMD structures
- Bus architectures
- Pipelining

Micro 32 Cool Chips Tutorial - MAB

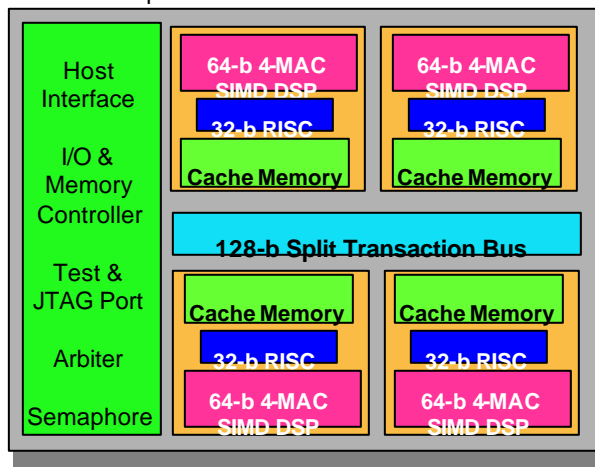
Bell Labs / Lucent Technologies

Single-Chip 1.6 Billion 16-b MAC/s Multiprocessor DSP

Lucent Technologies
Bell Labs innovations



Research Chip for 3G Wireless Base-stations



Chip Characteristics

Area 120mm²

Speed 100 MHz

Power 4W

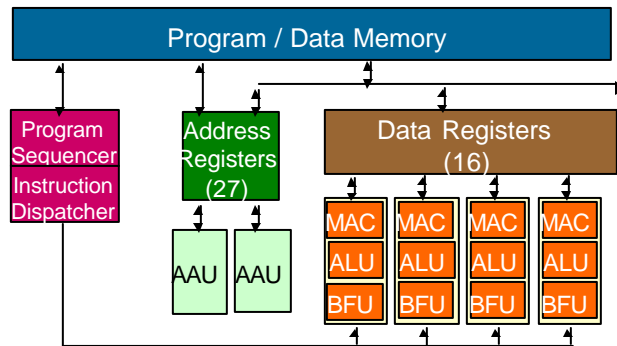
Tech 0.25um

- Capable of handling 1 GSM Carrier
- Low power due to hierarchical memory

Announced May 1999 - CICC99 San Diego

Micro 32 Cool Chips Tutorial - MAB

Bell Labs / Lucent Technologies



- 6-way VLIW with 128 bit (8x16) instruction fetch
- Prefix instructions for high performance without sacrificing code density
- Each execution set (parallel instructions + prefix) predicated
- 5 stage pipeline
- 1800 MIPS, 1200 MMACs @ 300 MHz, 0.1mA/MIP @ 1.5V

Conclusion



- DSPs developed from a need to handle a niche area.
- Their application has grown with realisation of their flexibility and performance.
- Low power consumption allows DSPs to be as pervasive as any embedded processors.
- The demand for more performance has caused designers to appreciate the prior art of general processor designers with obvious benefits.
- The holy grail for any system design is a software implementation that runs on a DSP that provides the performance required.
- The arrival of MIMD and VLIW architectures brings this goal closer.
- DSPs are evolving to meet the power/performance challenge.

The Power Managing OS meets A Thermally Aware Processor



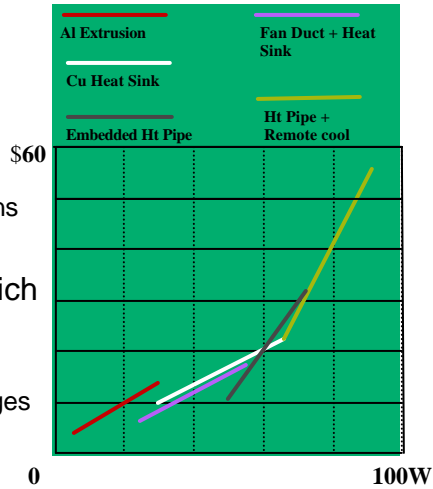
Agenda

- ◆ What is the motivation for this tutorial
- ◆ The Power Managing OS
 - OS Power Management
 - ACPI
 - OS response time
- ◆ A Thermally Aware Processor
 - System Impacts of High Power Processors
 - PowerPC 750*
 - Die Temperature Characteristics
- ◆ Conclusions



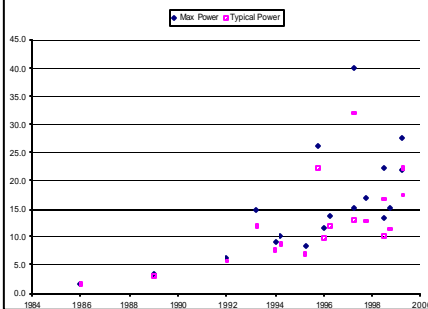
System Impact of High Power Processors

- ◆ It is clear that power dissipation adds to total system cost
 - Adds cost to power supply
 - Reduces the lifetime of the battery
 - Adds cost to the cooling solutions
- ◆ There is a processor power dissipation level beyond which cooling solutions add unreasonable costs
 - Fortunately the threshold changes with time & technology
 - Current threshold seems to be ~60W

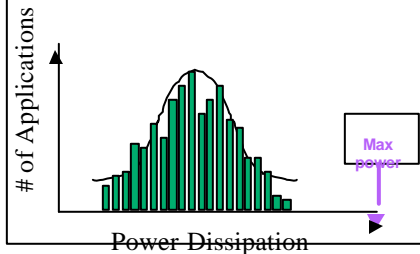


Power dissipation of existing processors

The gap between Max Power and Typical Application Power (TAP) is increasing



Application Traces reveal a broad distribution of TAP
Designing a system to be able to handle Max power of a leading edge processor can be expensive



It is possible to constrain the power dissipation of a processor without significant impact to application execution performance



The Power Managing OS



Operating System Power Management(OSPM)

Based on User preferences

Run in Performance mode or Quiet mode or Maximize Battery mode

Supported by Microsoft's desktop operating systems via APM - Advanced Power Management

OS/BIOS co-operation

When OS goes to idle condition it performs an access to a register that causes an SMI#

SMI handler puts system into low power state

APM required OS to trust the system BIOS



Current OSPM - ACPI

Advanced Configuration and Power Management Interface (ACPI)

- OS visible (SCI-based) as opposed to OS invisible (SMI-based)
- OS/drivers/BIOS are in sync regarding power states
- Individual device management w/o H/W traps and timers
- OS & drivers are better judges of system/device state

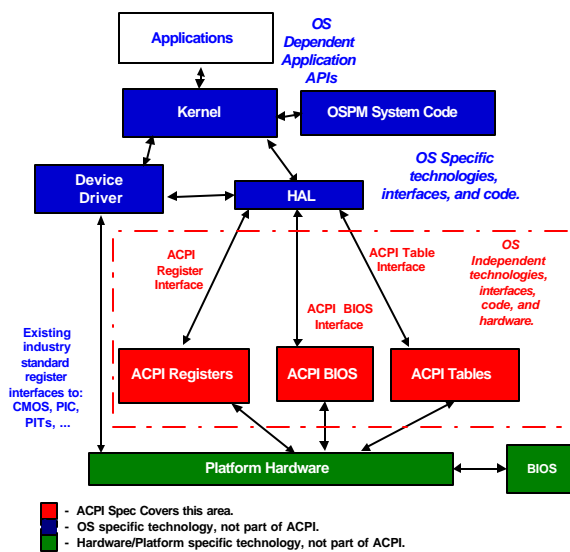
ACPI defines multiple sleep states

- Global states (G)
- CPU states (C)
- System states (S)
- Device states (D)
- Bus states (B)

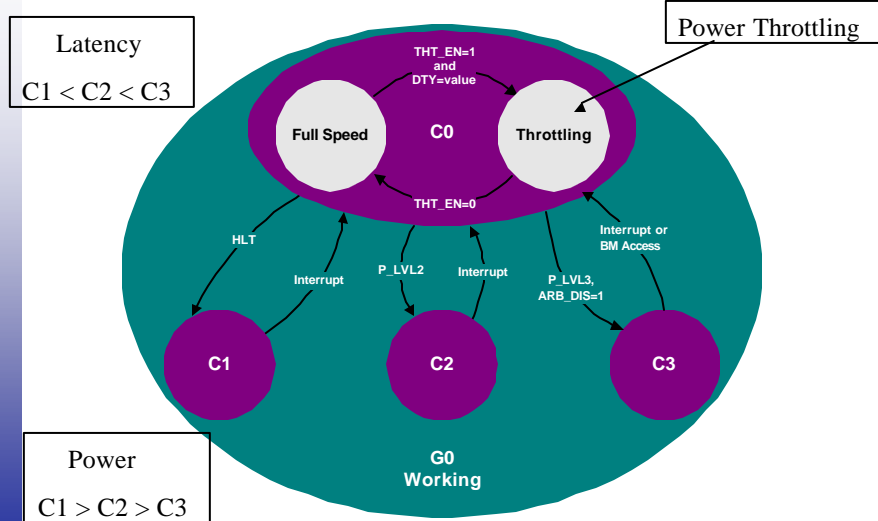
Thermal Management



ACPI System Architecture



ACPI Processor Power States



ACPI System States

State	CPU	Memory	Devices	Wake Up	Context Tracking
G0 Working	C0: Executing @ Full Speed C1-C3: Executing in PM state (ie Thermal Throttle/HLT)	Retained Power: ON Refresh: Normal	Powered Up & Down based on demand D0-D3		
S1 Sleeping	Not Executing Context Retained CPU CLK: OFF System CLK: ON Power: ON	Retained Power: ON Refresh: Normal	Devices Power down depending on wakeup & power requirements	Lowest Latency Restart @ CS:IP +1	H/W responsible for saving context of CPU, System I/O, & Memory
S2 Sleeping	Not Executing CPU/Sys Cache Context Lost CPU CLK: OFF System CLK: OFF Power: ON	Retained Power: ON Refresh: Standby / Auto	Devices Power down depending on wakeup & power requirements	Latency > S1 Restart @ Boot Vector	H/W responsible for saving context of System I/O & Memory OS responsible for saving CPU context
S3 Sleeping	Not Executing CPU/Cache Context Lost CPU CLK: OFF System CLK: OFF Power: OFF	Retained Power: ON Refresh: Standby / Auto	Devices Power down depending on wakeup & power requirements	Latency > S2 Restart @ Boot Vector	H/W responsible for saving Memory context BIOS restores Memory Controller Context. OS responsible for saving CPU & System I/O context
S4 S4BIOS Sleeping	Not Executing CPU/Cache Context Lost Everything OFF	Context Lost Power: OFF Refresh: N/A	Devices Power down depending on wakeup & power requirements	Latency > S3 Restart @ Boot Vector	OS(S4) / BIOS(S4bios) is responsible for saving and restoring all system context, including memory
G2/S5 Soft OFF	OFF	OFF	Devices are OFF. Power Button Press will wake up the system	Latency > S4 Restart @ Boot Vector	OS uses S5 to turn the machine off

NOTES:

- OS chooses the lowest supported sleep state in which all enabled wakeup devices still functions under the latency requirements from apps.
- ASL binds each Sx state to a SLP_TYP value, which based on platform design of power planes & clocking logic det what portions of the h/w power down.
- For each Device, ASL lists which power resources are needed to maintain a 'wakeup' capable state
- 'System I/O' refers to Motherboard Devices: PIT, PIC, DMAC, NMIState...OS saves & restores this stuff for S3



ACPI Timer

- ◆ Service provided to OS by chipset
 - Generates an interrupt every 2.34 seconds
 - 24-bit continuously running counter allows for fine granularity time measurement between events
- ◆ Hardware timer ensures correct OS timing algorithms in the face of variable processor and device execution speed



Processor Power Management

- ◆ OS manages the power dissipation of the processor
 - OS chooses Cx state based on idle time above a threshold
 - Above threshold, OS uses lower power Cx state
 - ACPI Timer used for idle timer detection
 - Timer sampled prior to and after exiting idle loop
 - Processor clock throttling
 - OS could scale processor duty cycle to match system usage
E.g 25% idle time, processor performance throttled to 75%
 - Clock throttling has a longer latency today than C1 implementation
- ◆ OS uses power dissipation of processor to manage zone temperatures
 - ACPI allows thermal zones
 - Different thermal characteristics are allowed per zone
 - OS will use processor clock throttling to control temperature of the zone that includes the processor



ACPI SW Concepts

- ◆ Dedicated ACPI Interrupt
 - SCI (System Control Interrupt)
 - RTC, ACPI Timer, Thermal sensor, Lid on Laptop, PCI device hot-plug etc
 - ACPI power/configuration events reflected to OS via SCI
 - Wake events cause system to transition to S0 state from a sleeping state, e.g. Wake on RTC
 - Runtime events - hot plug of PCI device, thermal sensor
- ◆ OS/BIOS Interaction
 - OS can generate SMI# by setting bit in Chipset
 - BIOS can generate SCI by setting a bit in Chipset



ACPI SW Concepts

- ◆ ACPI can execute “BIOS like” code
 - Code is created by BIOS/platform developers
 - ACPI Source Language (ASL) for writing methods
 - Compiled to ACPI Machine Language (AML) and merged w/ BIOS as part of ACPI tables
 - AML executed by OS



ACPI Thermal Management Methods

- ◆ Active Cooling
 - Turn on/Speed up system fan when system is hot
 - Turn off/slow down fan when system is cool
- ◆ Passive cooling
 - Reduces processor power dissipation when system is hot
 - Restricts power by modulating processors STPCLK# pin
 - STPCLK# duty cycle roughly proportional to reduction in CPU thermal dissipation
- ◆ Separate trigger points for Active versus Passive

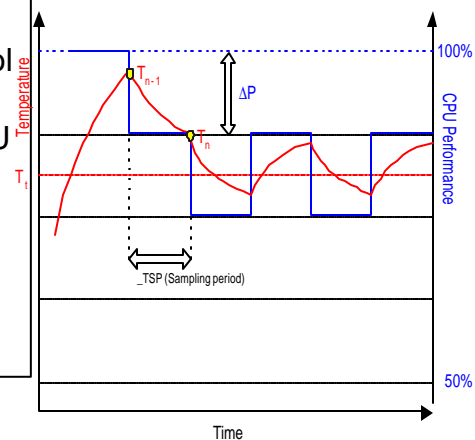


ACPI Thermal Model

For passive cooling the OS actively monitors the temperature in order to cool the platform.

The OS calculates the CPU performance change required to bring the temperature down

predefined equation with OEM supplied constants
OEM defined sampling period



Response Time

Operating system response times

SCI interrupt handler is NOT always the highest priority interrupt

SCI service routine potentially paged to disk

Results in a high latency between a thermal trigger point and OS induced response

responses may vary from small microseconds to several 10's of milliseconds



A Power Aware Processor



PowerPC 750*

- ◆ Is a low power and thermally aware processor
 - Contains a Thermal Assist Unit
 - Provides die temperature monitoring & measurement
 - $\pm 4^{\circ}\text{C}$ temperature resolution
 - Allows interrupt generation when temperature levels exceed preset thresholds
 - Implements reduced power operating modes
 - Implements Instruction Cache Throttling
 - Programmable delay inserted between cache fetch operations
 - Reduces maximum power dissipation
- ◆ With an ACPI compliant interface the processor die temperature could be controlled directly by an ACPI aware operating system



Simulated Die Thermal Rise Time

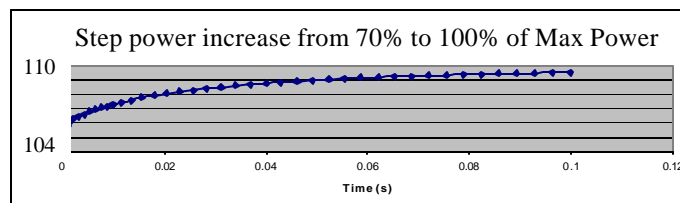
When a processor package is cooled to its Max Power the rise in die temperature over the last 1°C takes a “long time”

From a manufacturing perspective it is better to define max die temperature to be the lowest possible value

Direct impact on processor yield and/or frequency

When a processor is cooled to something less than its Max Power a 1°C temperature change can happen much quicker

Temperature increases from 105°C to 109°C in $\sim 60\text{ms}$



Conclusion

- ◆ Current OS ACPI implementations could not reliably control processor die temperature to an accuracy of $< 1^{\circ}\text{C}$
 - A processor die thermal rise-time of 1°C in 15mSec is less than SCI maximum response time
- ◆ OS can control system temperatures to an accuracy of $< 1^{\circ}\text{C}$
 - Thermal mass of system ensures that OS response times are adequate for effective closed loop control of system temperature
- ◆ Need other options for accurate temperature control



Closed Loop Control of Processor Die Temperature

Hardware could be added to a system to enable closed loop control of processor die temperature

Temperature sensor on die as in the PowerPC 750*

Closed loop control circuitry on the system baseboard

Adds cost to the basic system and dilutes the original "objective"

Closed loop control logic would be most effective if the complete circuit was added to the processor die

Any logic enabling closed loop temperature control should also be

SW accessible/controllable

Closed loop control is somewhat antagonistic to the goals of ACPI

The OS is no longer the best judge of when the processor should execute at $< 100\%$



Conflict of Processor Bandwidth Allocation Versus Temperature

- ◆ The constraint of processor performance to manage thermal dissipation or temperature would incrementally impact SW performance
 - Impact is incremental to the existing variables of interrupt rate, cache misses, I/O workload, available bus bandwidth etc
 - OS support for committed processor bandwidth allocation (real time) is currently unavailable
- ◆ Closed loop processor thermal control must be considered when determining total available processor bandwidth in real time allocation policy



Summary

- ◆ Cost can be removed from a system
 - reducing processor power dissipation
 - cooling a processor to less than its maximum power dissipation
- ◆ Processor yield and frequency can be helped by accurately controlling die temperature
 - To something less than $\pm 1^{\circ}\text{C}$
- ◆ Operating System Power Management as implemented today cannot respond quickly enough to achieve $\pm 1^{\circ}\text{C}$ control on higher power processors
 - On chip closed loop control can achieve better than $\pm 1^{\circ}\text{C}$





Architectural Level Power/Performance Optimization and Dynamic Power Estimation

- An Example from Simple Scalar Simulator Power Model

George Cai Chee How Lim

Intel Corp



Acknowledgements

Thanks to Vivek De and Shekhar Borkar for their very valuable statistical analysis, data collection, and many helps.

Thanks to Tosaku Nakanishi, Phil Wennblom, Krishnan Ravichandran, Shawn Searles, Tom Fletcher, Doug Carmean, Steve Gunther, and many of our colleagues at Intel.

Thanks to Professor Trevor Mudge, Brad Calder, Dean Tullsen, Wen-Mei Hwu, G. Gao, Dirk Grunwald and their research groups for their valuable feedback and encouragement.

Agenda

- Challenges to microprocessor architecture
Power/performance optimization is a new dimension of microprocessor architecture
- Dynamic power estimation technique for power/performance optimization
An example from Simple Scalar Simulator Power Model

Power/Performance Optimization

-new dimension of microprocessor architecture

- **Power and thermal limitation impacts on very high frequency implementation**
- **Difficulties of traditional voltage scaling for power reduction**
 - Transistor scaling difficulties
 - Complexity of high speed circuit scaling increases rapidly
 - Power/perf. tradeoff between leakage and multiple V_t
 - Architectural solution of soft error for reliability at low V_{cc}
- **Rapid increase in the number of transistors for non-computational logic blocks on chip**

Power efficient architecture must be emphasized



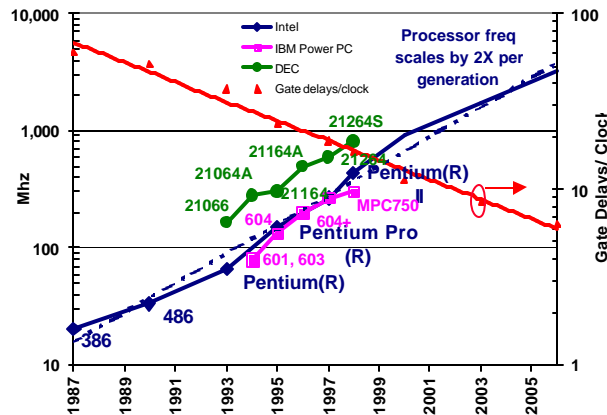
Power Consumption And Thermal Requirement Restricting High Frequency Implementation

- Gate_delays/clock_cycle reduction of 25% per generation
- Clock frequency doubles every generation
- Deeper and deeper pipeline for higher frequencies
- The better the power efficient architecture, the higher the implementation frequency within given thermal budget
- The better dynamic power behavior the microarchitecture have, the higher performance and lower cost the CPU can achieve

High frequency architecture must be power efficient



Processor frequency trend



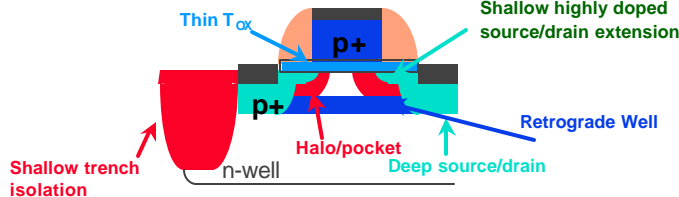
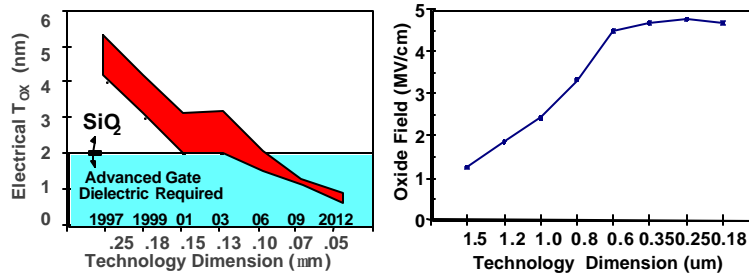
Frequency doubles each generation
Number of gates/clock reduce by 25%

Challenges of Voltage Scaling For Power Reduction

- Voltage scaling has been the most effective way to reduce power consumption
 - Reduce gate_delay 25% per generation
 - Double the transistor density per generation
 - Reduce energy per transition by 30% -65% per generation
- Traditional voltage scaling faces the challenge
 - Difficult to scale transistor oxide aggressively
 - Transistor scaling becomes more difficult than it was

Equivalent Architectural Scaling Method Should Be Enabled

Transistor scaling trends - T_{ox}





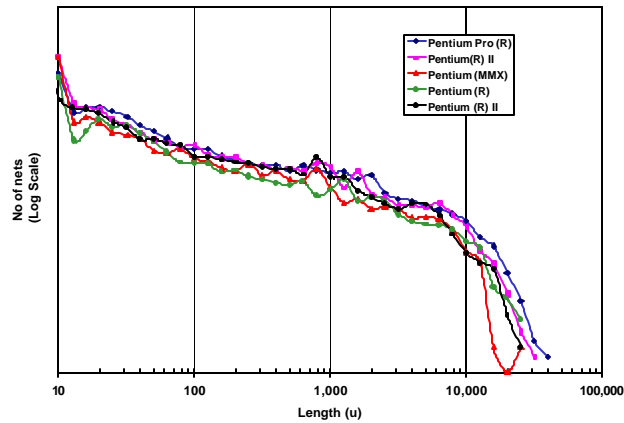
Complexity Of High Speed Circuit Scaling Increasing Rapidly

- Self timed circuits gate delay and IC delay sensitivities to Vcc are different
- Feedback circuits: N-mos in keeper is often non-minimum L different L MOS has different Vcc sensitivity
- I/O circuits: I/O timing is relative to external clock, voltage translator delay sensitivity to Vcc is non linear
- Interconnect scaling difficulties
 - High parasitic (R&C) and micro-architectural complexity
 - Complex interconnect distribution reducing transistor density

Microarchitecture must consider its scalability



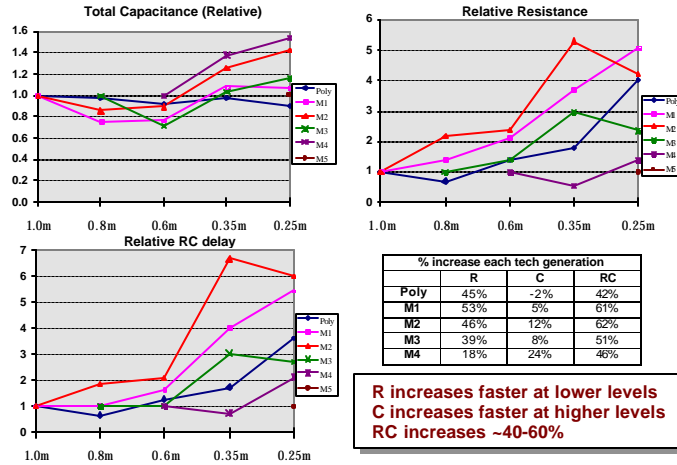
Interconnect distribution



Interconnect distribution does not change significantly



Interconnect performance



Leakage Current Important To Performance And Power Consumption

- Leakage I_{off} increasing 5X when V_t scaling 15% in future microprocessors
- Driving thermal runaway!
- Cooling systems have high cost
- Increasing total power consumption and negative impact on high end server product performance
- Critical to battery life of mobile computing



Power/performance Tradeoff Between Leakage And Dual Vt

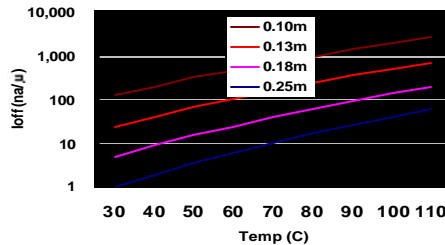
Dual Vt microarchitecture for power/performance optimization

- Low Vt for performance critical microarchitecture and data paths
- High Vt for power critical microarchitecture
- Evaluate each class of dual Vt circuits for systematically applying them to appropriate microarchitecture to achieve the best power/performance optimization
- Need architecture/circuit/cad tool/process cooperation

Power/performance optimization for Dual Vt CPUs



Lower V_t → higher drain Leakage



Starting with 0.25m technology, assume:

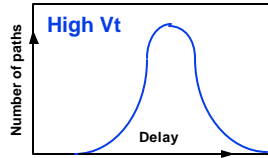
- V_t 450 mV
- I_{off} at 30 C 1 na/μ
- Subthreshold slope at 30 C 80 mv/decade
- Subthreshold slope at 100 C 100 mv/decade

Reference:
Mark Bohr, et al
IEDM, 1996

- V_t scaling per generation 15%
- I_{off} increase at 30 C 5X

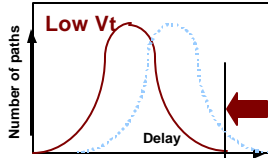


Dual V_t design--a leakage control technique

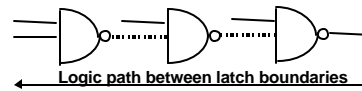


Technology provides two V_t
 ⚡ High V_t with nominal I_{off} (lower performance)
 ⚡ Low V_t with ~10X higher I_{off} (higher performance)

Employing high V_t everywhere yields lower performance, and lower leakage (1X)



Employing low V_t everywhere yields higher performance, but higher leakage (10X)



Selective usage of low and high V_t yields higher performance, yet low leakage between 1X, and <<10X



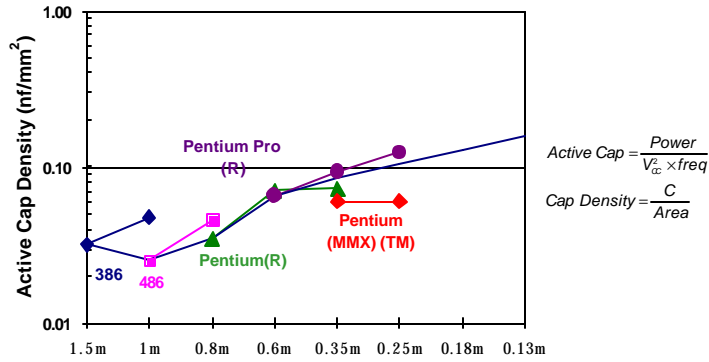
Power And Architectural Power Efficiency

Rapidly increasing non-computational logic used transistors on chip

- Active capacitance (power/[V_{dd}²Frequency]) grows by 35%
- Die size grows 25%
- 2X frequency
- V_{dd} scaling down 30%
- Misprediction penalty higher and higher because we predict almost everything!
- If trends continue (per generation)

2000 watts for supply voltage scaled microprocessors

Active capacitance density



$$Active\ Cap = \frac{Power}{V_{cc} \times freq}$$

$$Cap\ Density = \frac{C}{Area}$$

Active capacitance grows 30-35% each technology generation

How To Achieve Power/Performance Optimization

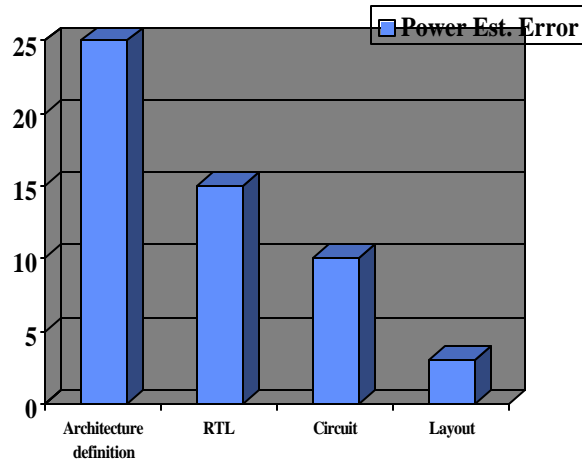
- Architectural definition stage
- RTL implementation stage
- Circuit implementation stage
- Physical design stage
- Processing and manufacture stage

Fundamental difficulty for new microarchitecture:

- Important architecture and design decisions must be made at early design phase, such as architecture definition stage
- Accurate power estimation is obtained at late design phase

Leads to multiple phase optimization and many iterations among all phases

Power Estimation Errors vs. Microprocessor Design Phases



Power Consumption Correlation Studies And Power Estimation Error Analysis

- Correlation between power estimations from low level design and high level architecture power simulator
- Critical paths based power estimation correlation analysis between architectural simulation and low level design
 - Circuit type based analysis
 - Typical activity factor based analysis
- Thermal imagines based correlation analysis
 - Hottest spot locations, coolest spot locations
 - Temperature differences, temperature distribution
- Micro-benchmarks based power estimation correlation analysis
 - Low level design (circuit, schematics, post layout)
 - Silicon correlation (max, average, min. I/O di/dt, thermal analysis)

Dynamic Power Modeling And Estimation

- Architectural and design partition
- Dynamic power behavior measurability and controllability
- Power density, activity, and effectiveness
- Architecture, circuit, layout, process impacts
- Statistical estimation and **error analysis**
- **Relative** value and absolute value

Modeling Parameters For Dynamic Power Estimation

(1)

- **di/dt threshold (DT):** *the threshold of the supply current difference during a unit clocking time;*
- **power threshold (PT):** *the threshold of the microprocessor dynamic power consumption during its execution;*
- **dynamic power monitor (DPM):** *a group of runtime counters and procedure calls that monitor the microprocessor runtime dynamic power behaviors including di/dt and max power violations and violation distribution;*
- **effective activity factor (EAF):** *a scaling factor that appropriately scales architecture activity factors and dynamic power monitor variables to applied logic and its possible layout area for power impact measurement;.*

Modeling Parameters For Dynamic Power Estimation (2)

- **effective area (EA):** *the scaled circuit area of several categorized circuits.*
- **active power density (APD):** *the power consumption per unit circuit area within a functional block during the functional block implementation. It is one of the most important power parameters for dynamic power estimation;*
- **inactive power density (IPD):** *the power consumption per unit circuit area within a functional block during the functional block inactive, such as sub-threshold leakage current. It is one of the most important power parameters for future microprocessor average power estimation;*
- **average power (AP):** *an average power consumption of microprocessor during an given execution time or an given performance benchmark.*

Assumption of Dynamic Power Modeling And Estimation Example

- Activity-sensitive power based on functional blocks
- Functional block activity derived from SimpleScalar
- Functional blocks comprised of selected circuit types:
 - Static, dynamic, SRAM, clock, programmable logic array (PLA), synthesizable and custom design
- Each circuit type dissipate power through:
 - Active power: P_{dynamic} is dominant
 - Inactive power: P_{leakage} is dominant
- Power can be statistically estimated from “reference” circuit designs
- Power = $\sum_i \{EAF * \sum_m (EA * APD) + (1-EAF) * \sum_m (EA * IPD)\}$
 - where $i = \text{\#cycles}$; $m = \text{circuit types}$

Simple Scalar Power Model Overview

(Characteristics)

- Activity-sensitive power simulation
- Block-level
 - $\text{power (active)} = \text{activity} * (\text{circuit_type} * \text{area} * \text{active_power_density})$
- Block-level
 - $\text{power (inactive)} = (1 - \text{activity}) * (\text{circuit_type} * \text{area} * \text{inactive_power_density})$
- Basic Simplescalar architecture is partitioned into 32 physical blocks for power estimation.
- This partition and area estimation are done based on microprocessor design experience.
- Circuit power density is estimated from SPICE simulations (the circuit structures -SRAM, dynamic, static, PLA, clock - can be obtained from textbooks).

Added Power Simulation Variables

- **ppres_blockname** = present cycle power contribution of blockname
- **pprev_blockname** = previous cycle power contribution of blockname
- **pdidt_blockname** = the change from present to previous power (dynamic power)
- **pres_blockname** = present cycle activity contribution of blockname
- **prev_blockname** = activity contribution of blockname up to previous cycle
- **count_blockname** = activity contribution of blockname up to present cycle
- **blockname.ckt_pda** = active power density of circuit ckt for block blockname
 - where ckt = {dyn,sta,mem,pla,clk}
- **blockname.ckt_pdi** = inactive power density of circuit ckt for block blockname
 - where ckt = {dyn,sta,mem,pla,clk}
- **blockname.ckt_a** = circuit area

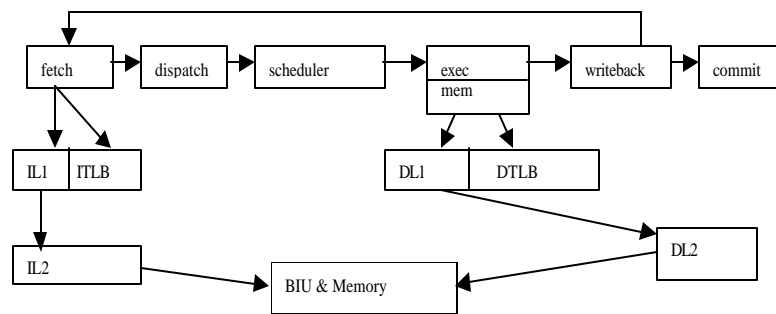


Names of 32 Functional Blocks

- **npclog** = Next-pc generation logic
- **btblog** = BTB logic
- **btbcac** = BTB cache
- **rsbcac** = RSB
- **itlbcac** = Instruction TLB
- **dtlbcac** = Data TLB
- **pmhlog** = Page miss handler
- **il1log** = L1 inst cache logic
- **il1tag** = L1 inst cache tag
- **il1cac** = L1 inst cache array
- **dl1log** = L1 data cache logic
- **dl1tag** = L1 data cache tag
- **dl1cac** = L1 data cache array
- **dispatchq** = Dispatch queue
- **decodepla** = Inst decoder
- **decodemisp** = Misprediction handling logic
- **decodestall** = Decoder stall logic
- **reorder** = Reorder table
- **ratarr** = Rename table
- **ruuarr** = Re-order buffer
- **lsqarr** = Load/Store queue
- **ruurdyq** = Re-order buffer ready queue
- **lsqrdyq** = Load/Store queue ready queue
- **ruuarb** = Re-order buffer arbitration logic
- **lsqarb** = Load/Store queue arbitration logic
- **ruuwb** = Re-order buffer writeback scheduler
- **lsqwb** = Load/Store queue writeback scheduler
- **fuint** = Integer execution unit
- **fufp** = Floating point execution unit
- **ul2log** = Unified L2 logic
- **ul2tag** = Unified L2 tag
- **ul2cac** = Unified L2 cache
- **biu** = Bus/IO buffer



Example of Simple Scalar Simulator Partition (1)



Example of Simple Scalar Simulator Partition (4)

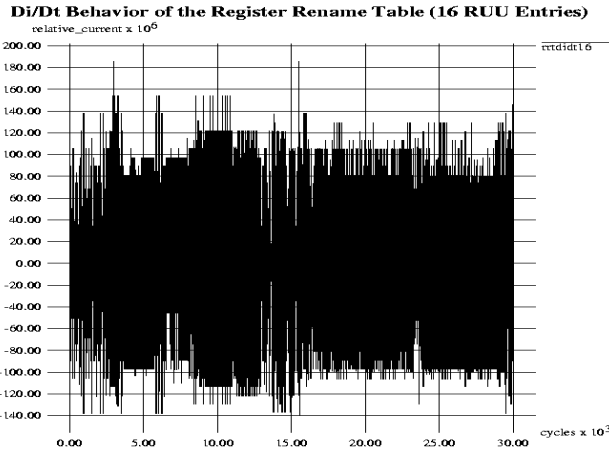
UNIT	BLOCK	ARCHITECTURE FEATURE	ACTIVITY
SCHEDULER	SCRUIWBLOG	Writeback logic	ruwwb + ruwbq + ruwet
SCHEDULER	SCRUIRYQUE	Reg Dep. Chk Cam	ruurdyqcam
SCHEDULER	SCRUIABLOG	Writeback Arbitration	ruarab
SCHEDULER	SCLSQUALARR	LSQ array, 8 entry	lsqarr + lsqrdyqsch + lsqrec + lsqret
SCHEDULER	SCLSQWBLOG	Writeback logic	lsqwb + lsqwqb + lsqret
SCHEDULER	SCLSQRRYQUE	Mem Dep. Chk Cam	lsqrdyqcam
SCHEDULER	SCLSQABLOG	Writeback Arbitration	lsqarab
EXECUTE	EXIEUCPLOG	Int execution logic, 4 ALU, 1 MULT/DIV	fuint
EXECUTE	EXFEUCPLOG	Fp execution logic, 4 ALU, 1 MULT/DIV	fufp
MEMORY	METLBDACAC	DTLB, 32 set, 4KB page, 4 way, LRU	dtlbacc + dtlbrep + dtlbwbk + dtlbinv
MEMORY	MEDL1DACAC	L1 data cache, 128 set, 32B block, 4 way, LRU, 1 cycle hit latency	d1facc + d1frep + d1fwbk + d1finv
MEMORY	MEDL1DATAG	L1 data cache tag	d1facc + d1frep + d1fwbk + d1finv
MEMORY	MEDL1DALOG	D1L1 decode, LRU	d1facc + d1frep + d1fwbk + d1finv
MEMORY	MEUL2IDCAC	Unified L2 cache, 1024 set, 64B block, 4 way, LRU, 6 cycle hit latency	ul2acc + ul2rep + ul2wbk + ul2inv
MEMORY	MEUL2IDTAG	Unified L2 cache tag	ul2acc + ul2rep + ul2wbk + ul2inv
MEMORY	MEUL2IDLOG	UL2 Decode, LRU	ul2acc + ul2rep + ul2wbk + ul2inv
MEMORY	MEPMHIDLOG	Page Miss Handler, 30 cycle latency	itbmis + dtbmis
BUS	BUBUSIOBUF	Bus interface logic, 8B bus width, 18.2 cycle latency	ul2mis + ul2rep + ul2wbk + ul2inv

Running the Simulator

- `> pow-outorder3 -tlb:dtlb dtlb:4:4096:64:1 ../spec95-big/compress95.ss < inputs/compress/test.in`
- The statement above will run the pow-outorder3 simulator for compress benchmark with specific data TLB configuration (refer to Todd Austin's manual).
- If you want to recompile ...
 - Makefile:
 - * Ensure that pow-outorder3.c and main1.c are listed
 - * Point the BINUTILS_INC and BINUTILS_LIB to appropriate areas



Dynamic Power Behaviors of Register Rename Table (16 RUU Entries)



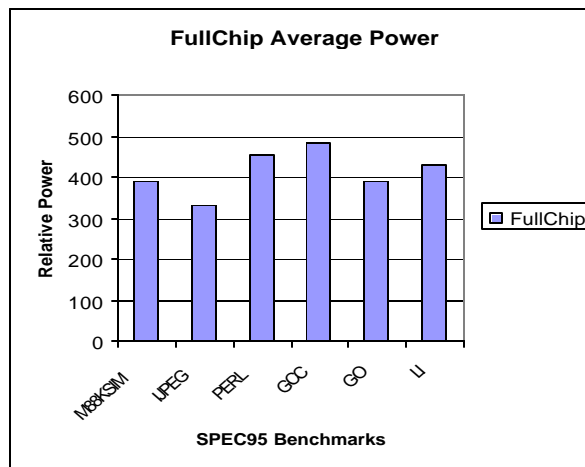
Architectural Level Power/Performance
Optimization and Dynamic Power Estimation

ACM/IEEE Micro32
Nov. 15, 1999 Haifa, Israel

Page 33



Full Chip Average Power Estimation (Relative Value)



Architectural Level Power/Performance
Optimization and Dynamic Power Estimation

ACM/IEEE Micro32
Nov. 15, 1999 Haifa, Israel

Page 34

Conclusion

- Power efficiency has severe impact on microprocessor performance, function, and cost
- New architecture and implementation challenges from power/performance optimization
- New technology providing new opportunities for power efficient architecture and power/performance optimization
- Dynamic power behavior observation and power estimation are important for power/performance optimization

Backup Foils

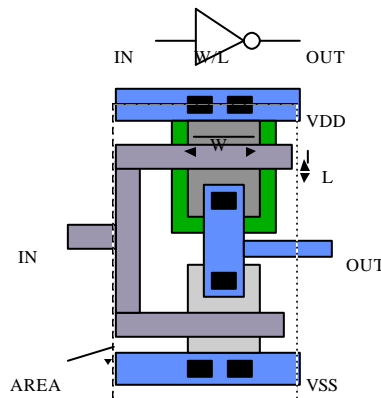
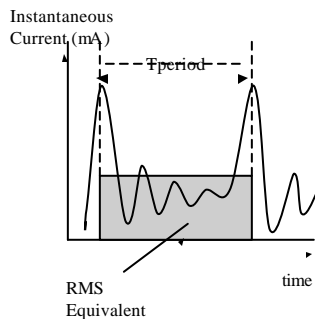
- Environment concerns
- Measurement entities
- Statistical estimation
- Circuit design style categorization
- Design variables
- Reference design boundary

Environment Concerns

- Active Power Case (Mostly P_{dynamic})
 - Fast Process Skew (low V_t , low L_{eff} , low T_{ox})
 - Low Temperature
 - High Voltage
- Inactive Power Case (Mostly P_{leakage})
 - Fast Process Skew (low V_t , low L_{eff} , low T_{ox})
 - High Temperature
 - High Voltage

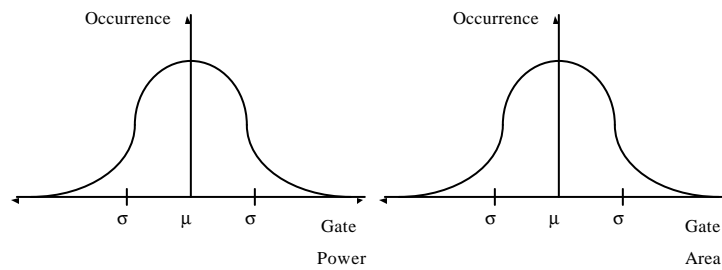
Measured Entities

- Root-Mean-Square (RMS) current consumption ($\times V_{\text{supply}}$ to get power)
- Area occupied



Statistical Estimation

- Determine distribution weights of each sim. point
 - for simplicity, assume equal weights
- Determine mean and standard distribution



Architectural Level Power/Performance
Optimization and Dynamic Power Estimation

ACM/IEEE Micro32
Nov. 15, 1999 Haifa, Israel

Page 39

Statistical Estimation

- Determine “confidence” criteria (e.g. 1 sigma)
- Power Density = $\text{Power}_M / \text{Area}_N$ (in $\text{W}/\mu\text{m}^2$)
 - $M, N = \{-\sigma, \mu, +\sigma\}$
 - Example:

$$\text{PD}(\text{worst-case}) = \text{Power}_{+\sigma} / \text{Area}_{-\sigma}$$

$$\text{PD}(\text{best-case}) = \text{Power}_{-\sigma} / \text{Area}_{+\sigma}$$

Architectural Level Power/Performance
Optimization and Dynamic Power Estimation

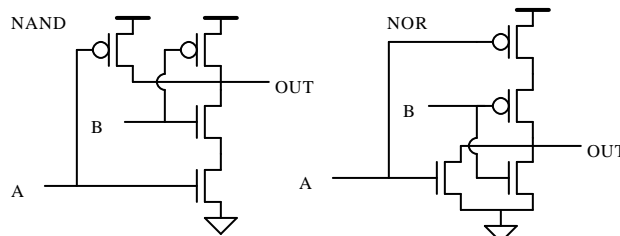
ACM/IEEE Micro32
Nov. 15, 1999 Haifa, Israel

Page 40



Circuit Types

- Static Circuit
 - Simulate NAND/NOR gates with different fanin/fanout
 - 2/3/4 fanin NAND with fanout of 2/3/4
 - 2/3/4 fanin NOR with fanout of 2/3/4
 - INV with fanout of 2/3/4



Architectural Level Power/Performance
Optimization and Dynamic Power Estimation

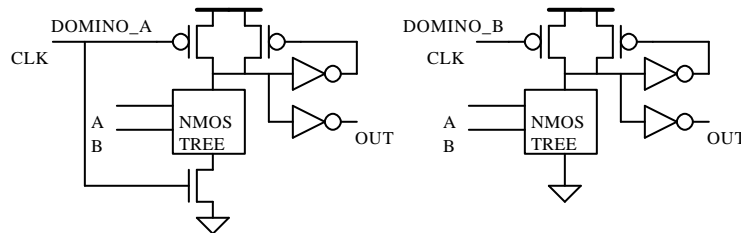
ACM/IEEE Micro32
Nov. 15, 1999 Haifa, Israel

Page 41



Circuit Types

- Dynamic Circuit
 - Simulate NAND/NOR gates with different fanin/fanout
 - 2/3/4 fanin NAND with fanout of 2/3/4
 - 2/4/8/16/32 fanin NOR with fanout of 2/3/4
 - Domino_A and Domino_B



Architectural Level Power/Performance
Optimization and Dynamic Power Estimation

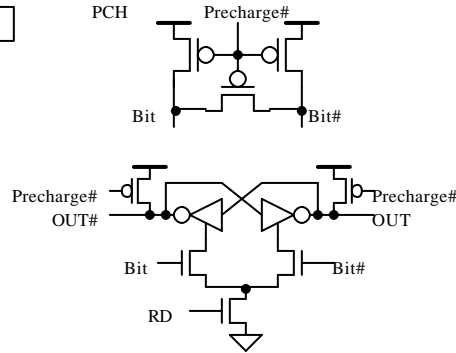
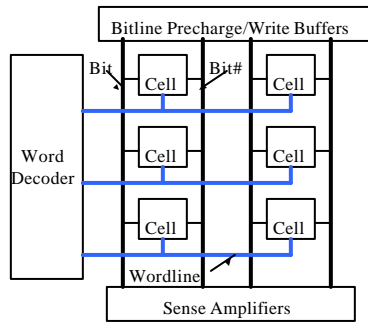
ACM/IEEE Micro32
Nov. 15, 1999 Haifa, Israel

Page 42



Circuit Types

- SRAM Circuit
 - Simulate READ/WRITE cycles (e.q. M x 32bits)
 - RD: Wordline(Domino), Bitline(Pch), Sense Amplifier(SA)
 - WR: Wordline(Domino), Bitline(Pch), Write Buffer(Inverter)



Architectural Level Power/Performance Optimization and Dynamic Power Estimation

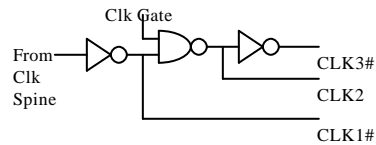
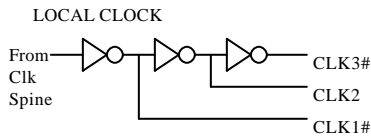
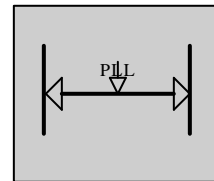
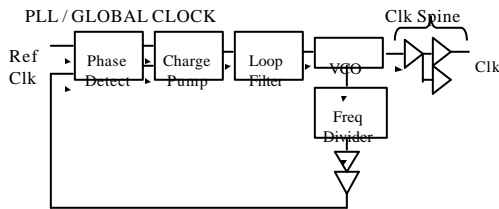
ACM/IEEE Micro32
Nov. 15, 1999 Haifa, Israel

Page 43



Circuit Types

- Clock Buffer
 - Simulate Global Distribution (H tree/Grid) and Local Clock Generation (Buffers/Choppers)



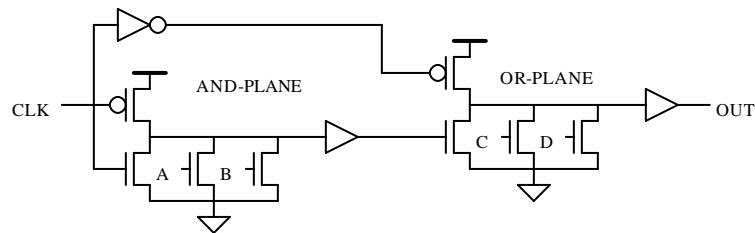
Architectural Level Power/Performance Optimization and Dynamic Power Estimation

ACM/IEEE Micro32
Nov. 15, 1999 Haifa, Israel

Page 44

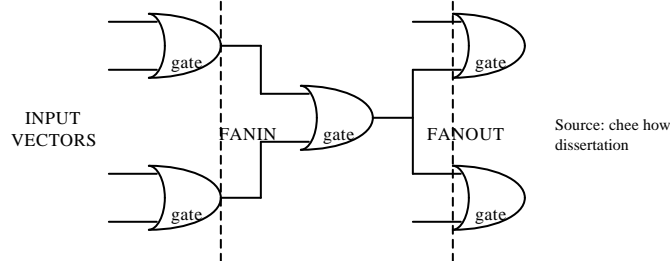
Circuit Types

- Programmable Logic Array
 - Simulate AND-OR Plane (e.g. $M \times N$ matrix)
 - Implement with Dynamic NOR-NOR circuit



Design Variables

- Input vectors
 - Random, Bimodal, Gaussian, Favor-high, Favor-low
- Circuit Fanin (2/3/4/5 etc. inputs)
- Circuit Fanout (2/3/4/5 etc. output loads)
- Circuit Sizes (2/4/6/8 etc. μm widths)



Reference Design Boundary

- Reference design requirement
 - Number of gates per pipestage (e.g. 8 gates/stage)
 - Pipestage period = $1 / F_{max}$
 - $F_{max} = 1 / (T_{skew} + T_{clkq,stage1} + T_{logic} + T_{su,stage2})$
 - Gate Delay $\approx T_{logic} / 8$

