Why IPF… and Why HP?
An Independent Analyst's View

Terry C. Shannon
Consultant and Publisher
Shannon Knows High Performance Computing

The Fine Print

- **The following text reflects legal requirements**
  - I am not nor ever have been an HP or Intel Employee
  - I am not being compensated to deliver a marketing pitch
  - This presentation reflects my opinion, not HP's opinion
  - No NDA material is contained in this presentation
  - I strive for accuracy, but offer no guarantees
  - Trust but verify: always get a second opinion
  - Make no purchasing decisions based on this session
  - Always check with HP before planning your purchases
  - Above all, enjoy the presentation
  - Comments? Email terry@shannonknowshpc.com

Session Agenda

- Processor Evolution from 1950 to Present
- From UNIVAC to Integrity
  - Current Superdome, PA-RISC and IPF Technology
  - Anticipated Improvements in the next few years
- Rumours of the death of Itanium are greatly exaggerated.
- Extensions are not all things to all people
- Why IPF? *Why not?*
- Why HP? *Who else!*

HP's hardware strategy

- Hardware evolution drives the rise of EPIC
- HP standardizes on Itanium for enterprise systems
- AMD Opteron and Intel Nocona enter the picture, delivering 64-bit support to ProLiant systems.
- Customer choice drives platform selection
- Intel Itanium hardware will form the basis of all HP enterprise systems
- This includes the next-generation Post-Superdome enterprise server due in ~2008 and new NSK systems that implement the NSAA.

A Chat About CPUs

- A topic with more misconceptions than facts
- Processor evolution keeps pace with technology
- Darwin was right… but…
- Beyond Darwin, there's always business decisions
- A 52-year microprocessor journey
  - UNIVAC 1952 – first commercial computer
  - VLSI, incompatibility, and poor marketing killed UNIVAC
  - IBM subsequently ruled the roost with a wide variety of machines based on a wide variety of architectures.
  - Incompatibility: scientific or commercial
  - IBM solves the problem in 1964

A chat about CPUs, ctd

- April 7, 1964: IBM debuts System/360 series
  - First commercial microprogrammed computer architecture
  - Floor wax and a dessert topping
  - First instantiation of a CISC architecture
  - Dominated computer design for two decades… even VAX/VMS… one of most successful CISC architectures
  - Mainframes are still alive and well
  - But evolution continues…

A Chat about CPUs, ctd…

- Improved measurement tools proved that most CISC programs relied a few simple instructions
- IBM sets out to solve the problem
  - 1975 – 1979: IBM invents the 801
  - Never commercialized, but the first exemplar of RISC
  - Developed by Joel Birnbaum, who later did PA-RISC
  - A derivative single-chip 801(ROMP) was used in IBM's first commercial RISC system, the PC/RT in 1986, which launched concurrently with PA-RISC.
  - The PC/RT was a commercial failure, but…
  - It set the stage for POWER and PowerPC
  - CISC was placed at risk by RISC

A Chat about CPUs, ctd…

- Computer Scientists Recognize RISC
  - Stanford computer scientists emulated the IBM example of relying on compiler optimization and pipeline efficiency and produced what became the MIPS architecture.
  - Berkeley focused on minimizing inherently slow calls to external memory with a register rich architecture adopted in 1987 by Sun Microsystems and, thus helped productize the SPARC processor
  - DEC developed and cancelled PRISM, a scalar/vector RISC architecture. Big mistake!
  - PRISM was supplanted by a DEC/MIPS joint venture, which ultimately failed. Bigger mistake.
  - DEC then developed Alpha, a superscalar RISC CPU

The Rise and Fall of Alpha

- HP acquired Alpha technology with the CPQ buy
  - For a variety of reasons, not the least of which was marketing malfeasance, Alpha never achieved critical mass or commercial success.
  - Alpha was very successful in real-time and HPTC
  - HP will phase-out Alpha after the EV7z release
  - PA-RISC will follow after final PA-8900 release in 2005
  - SPARC and MIPS are no longer relevant
- New choices: POWER, IPF, and x86 extensions
  - Today, RISC CPUs power the majority of servers and workstations, although IA and IBM's zSeries servers manage to disguise their RISC DNA quite well.

Intel gets inside with CICS

- While the workstation and server market focused on RISC, Intel had plans of its own.
  - Intel developed the first microprocessor in 1971, but waited 15 years to begin executing a plan to achieve microprocessor dominance.
  - Remember “Intel Inside” in 1988?
- Behind the scenes…
  - June 8, 1978: Intel launches 8086, a 16-bit CPU
  - First CPU in the x86 family… vestiges of which remain in the latest Pentium and Xeon processors.
  - Cost-reduced 8-bit 8088 powered IBM PC
- Enhancement acceleration begins

Intel processor proliferation

- First four Intel CPUs bore the 8x86 name
- Time for “Five-9’s”
  - 1980: 80186, 8 or 16 bits
  - 1982: 80286, 16 bits, 16MB RAM, VM, multitasking
  - 1985: 80386: 32 bits, 275M xsistors, 4GB memory
  - 1989: 80486: on-chip cache and FPU = 2x performance
- Numbers are out, names are in
  - 1983: Pentium 1 3.1M xsistors, 100MIPS
  - 1995: Pentium Pro: Intel’s 5.5M xsistor “October Surprise”
  - 1999 Pentium III and Xeon: competitive fear and loathing
  - 2000 Pentium 4 and Xeon double performance, again

What's Next?

- Intel increased speed and performance of 32b chips
- Meanwhile, the time had come for a post-RISC 64 bit processor that eliminated RISC limitations
- An EPIC Adventure Begins
  - In 1988, HP started to design a long-term RISC CPU
  - HP opted to partner with Intel on the effort
  - June 3, 1994: P7/Merced/Itanium is jointly announced
  - 2001 debut was four years late, performance lagged
  - June 25, 2001 – Compaq scuttles Alpha for Itanium

"History will prove that Compaq made the right decision at the right time to cancel the EV8 project and transition, over time, all its 64-bit enterprise platforms to Intel's Itanium technology."

Terry C. Shannon

IT Consultant and Publisher, *Shannon knows High Performance Computing*

The timing of a processor transition

- CPQ wanted to use the EV7 launch to announce Itanium as an ”option,” Intel wanted a rapid closure.
  - In April 2001, Compaq realized 2FQ would be ugly. Finalizing the technology transfer by the end of the quarter would resolve the financial issues. 25 June was selected, contracts were signed, and a marketing effort commenced in June.
  - Compaq endured great wrath from irate Alpha customers. But with Alpha no longer the “other woman” in the Intel-Compaq relationship, that relationship improved and Intel got a big-name addition to the Itanium adoption list.

Alpha Retired, Itanium hired

- How and why did this happen?
- Darwin was right
- "The Origin of Species" sums it up in two lines
  - Evolution happens
  - Only the strong survive
- Microprocessor Evolution
  - Early specialized CPUs replaced by CISC
  - CISC supplanted by RISC
  - EPIC will likely supplant RISC

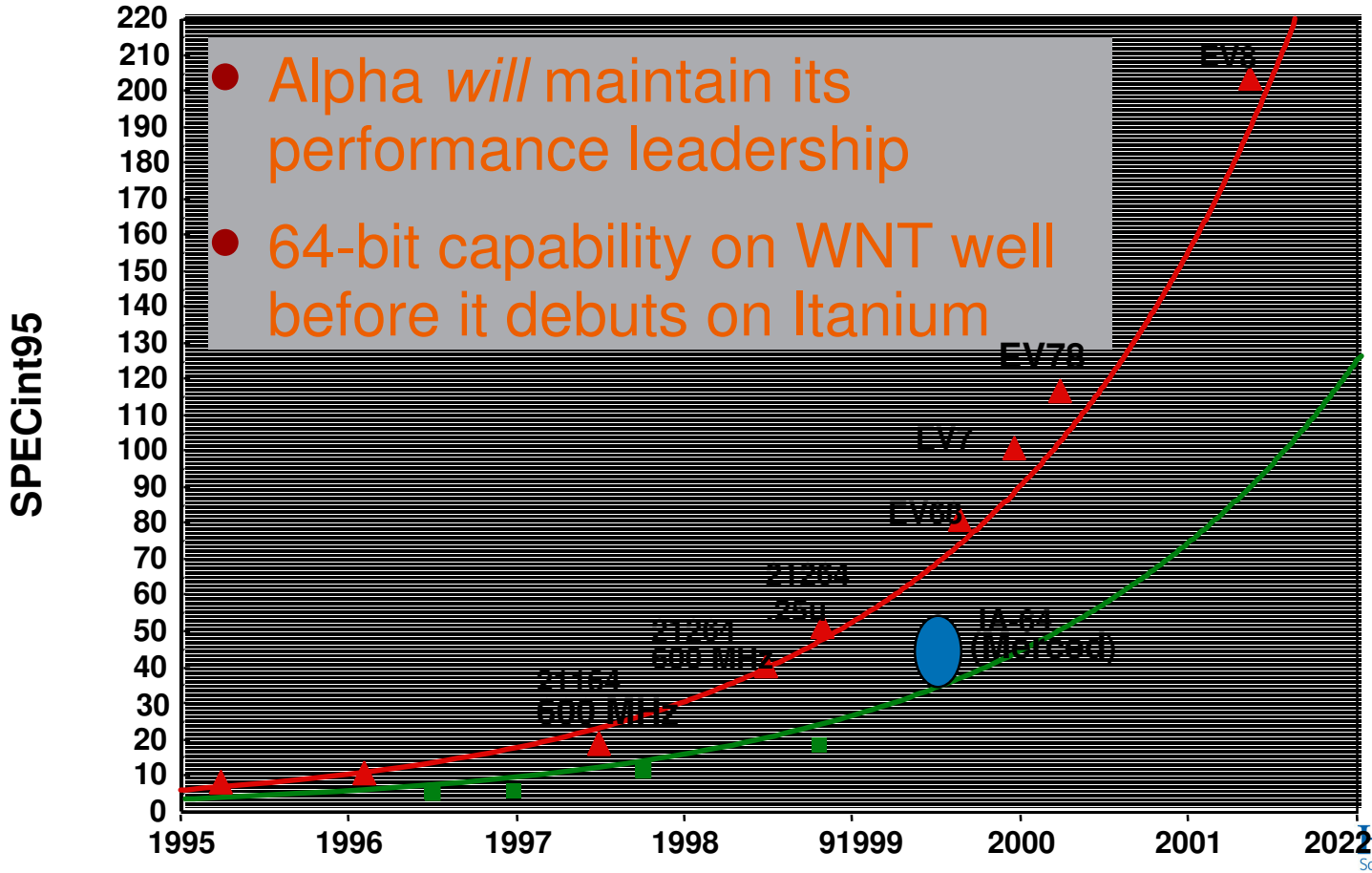
Alpha's omega

Business, technology, and corporate policy

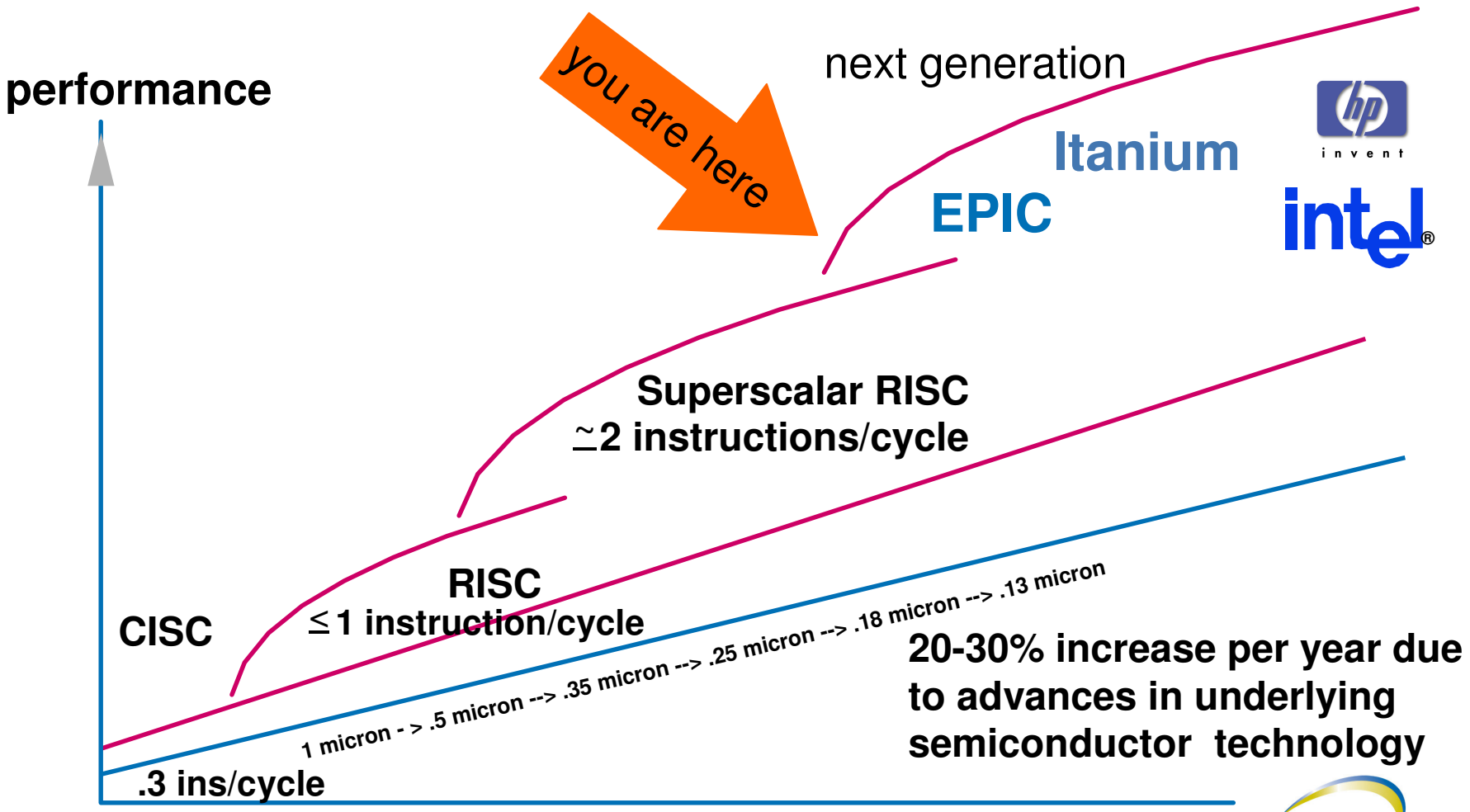
- Technology: Alpha as roadmap roadkill
- Alpha's differentiator was "twice the performance of Intel"
  - Itanium appeared on Alpha charts long before Intel delivered
  - Alpha CPU cartographers could better project performance
  - Intel could better project Itanium performance
  - So the following curve reflected Compaq's expectations for Intel
    - Intel invited Compaq to review the Itanium roadmap in 2000
      - Itanium would achieve performance parity in the EV7 era and rapidly extend its performance leadership
    - Without a value proposition, Alpha was toast
- Compaq used the same 1998 slide through mid-2001

Alpha/Itanium roadmap circa 1998

Alpha could only estimate Itanium performance



CPU evolution changed the curve



Acts of Omission in Prognostication?

- Back to the original chart:
- "It's difficult to predict, especially the future"
- Projecting the future performance of one's own product is feasible.
- Probability decreases as distance increases.
- Projecting the future performance of a rival product is more guesstimation than extrapolation
- And processor evolution plays a role
- If EPIC is the successor to superscalar RISC
- Projections may be even less accurate

The business of CPU death and life

- Alpha's fate was sealed by the Intel roadmaps
  - Alpha already was bound for Death Row because:
    - Sales volumes were decreasing
    - It cost Compaq $800USD for each processor it shipped
    - The EV8 effort required additional resources
  - Negotiations, Alphacide, and Itanium rising
    - Itanium was Compaq's chosen Alpha successor, Intel wanted Alpha out of the picture, negotiations were held.
    - The result: Compaq granted Intel nonexclusive rights to use and modify the Alpha architecture, and provide all Alpha IP. Alpha developers were offered jobs at Intel. Compaq agreed to adopt Itanium as the successor to Alpha when all EV7 work was done.

EPIC Addresses Alpha Limitations

Problem	Itanium® Architecture	Alpha	Comments
Branch Mispredicts	Predication	Conditional Move (CMOVE) Dual Algorithm Branch Predictor (EV6)	Branchless If-Then-Else at cost of increased I-Stream Bandwidth Reduced Branch Mis-Predicts
Memory Latency	Explicit Speculative Load	Explicit Memory Prefetch (Prefetch) Speculative Load	Allows compiler to move load to earlier block Allows speculative loads outside view of compiler
Explicit vs Implicit Parallelism	128 Integer Registers	32 Integer Registers + 48 Shadow Registers	Allows compiler to directly schedule all registers at cost of increased context switch Large register count without save/restore overhead. Shadow register count can grow in future implementations

EPIC: Basic Ideas for a New Architecture

- EPIC = Explicitly Parallel Instruction Computing
- Static hardware design
  - Compiler creates record of execution
  - Instructions arranged in “bundles”
  - The EPIC CPU “plays the record”
  - No runtime changes as with out of order execution
- High scalability of execution units
  - Very Long Instruction Word (VLIW) concept
  - Focus is parallelism
  - 6 instructions in parallel (2 bundles per cycle)
  - High number of execution units

Explicit parallelism explained

- Instruction Level Parallelism (**ILP**) = ability to execute multiple instructions concurrently.
- Explicitly Parallel Instruction Computing (**EPIC**) allows the compiler or assembler to specify the parallelism
  - Compiler specifies **Instruction Group,** list of instructions with no dependencies that can be executed in parallel
  - Instructions are packed in **bundles** of 3 instructions each
    - Instruction bundle - two executed per cycle
- Massive resources on chip
  - Large number of registers to avoid register contention

Architecture limits – EPIC solutions

Today's Limits: complexity of multiple pipelines too great to allow effective on-chip scheduling for parallel operation

➔Solution: explicit parallelism

Compiler handles Scheduling and communicates this to the chip

Today's Limit: number of registers on chip limits parallelism

➔Solution: quadruple registers from 32 to 128 and increasing addressing from 5 bits to 7

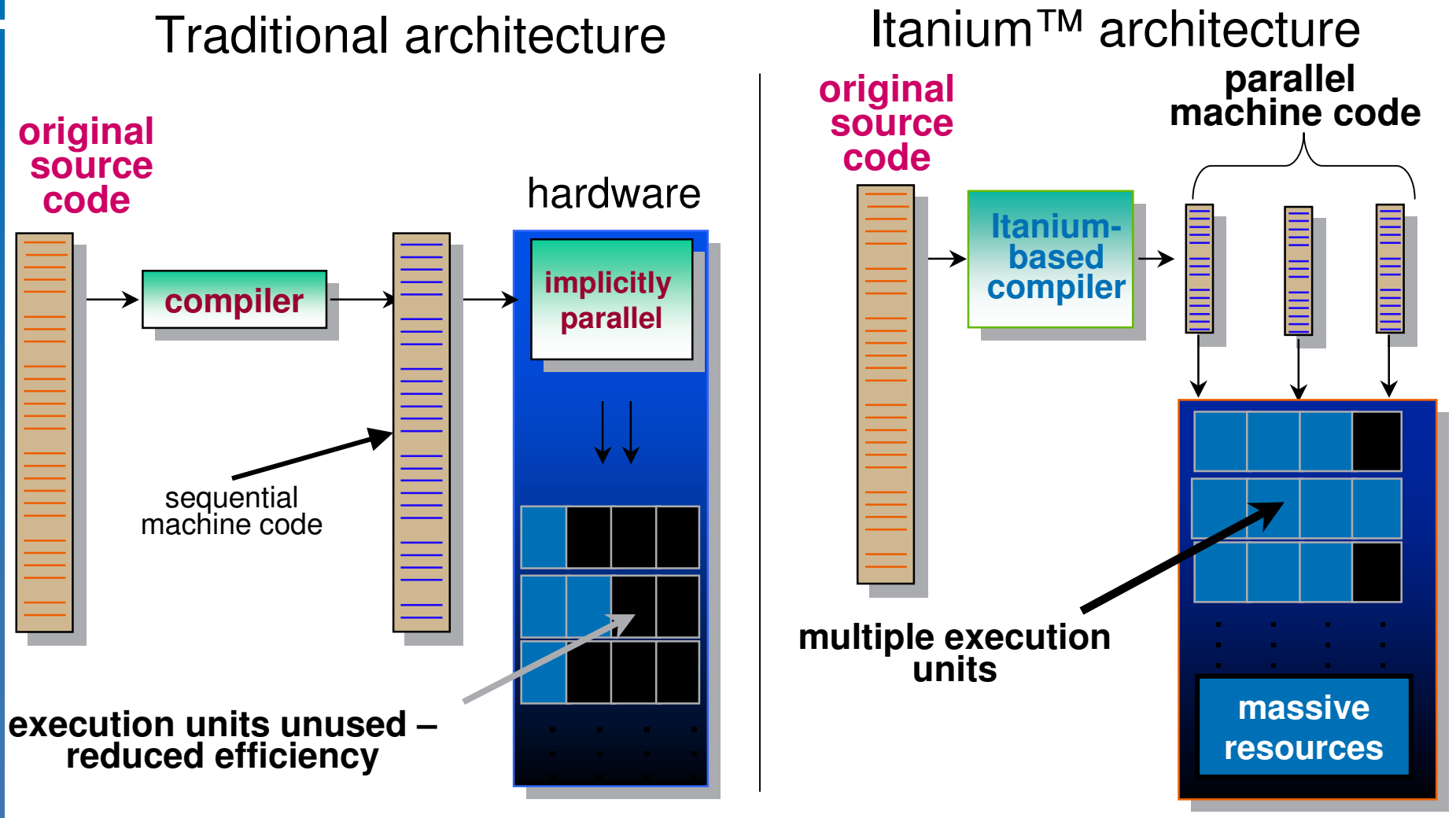
Today's Limit: Large (and growing) memory latency

➔Solution: speculative loads

Today's Limit: conditional and/or unpredictable branches

➔Solution: prediction and predication orchestrated by the compiler

Itanium Architecture – Basic Ideas



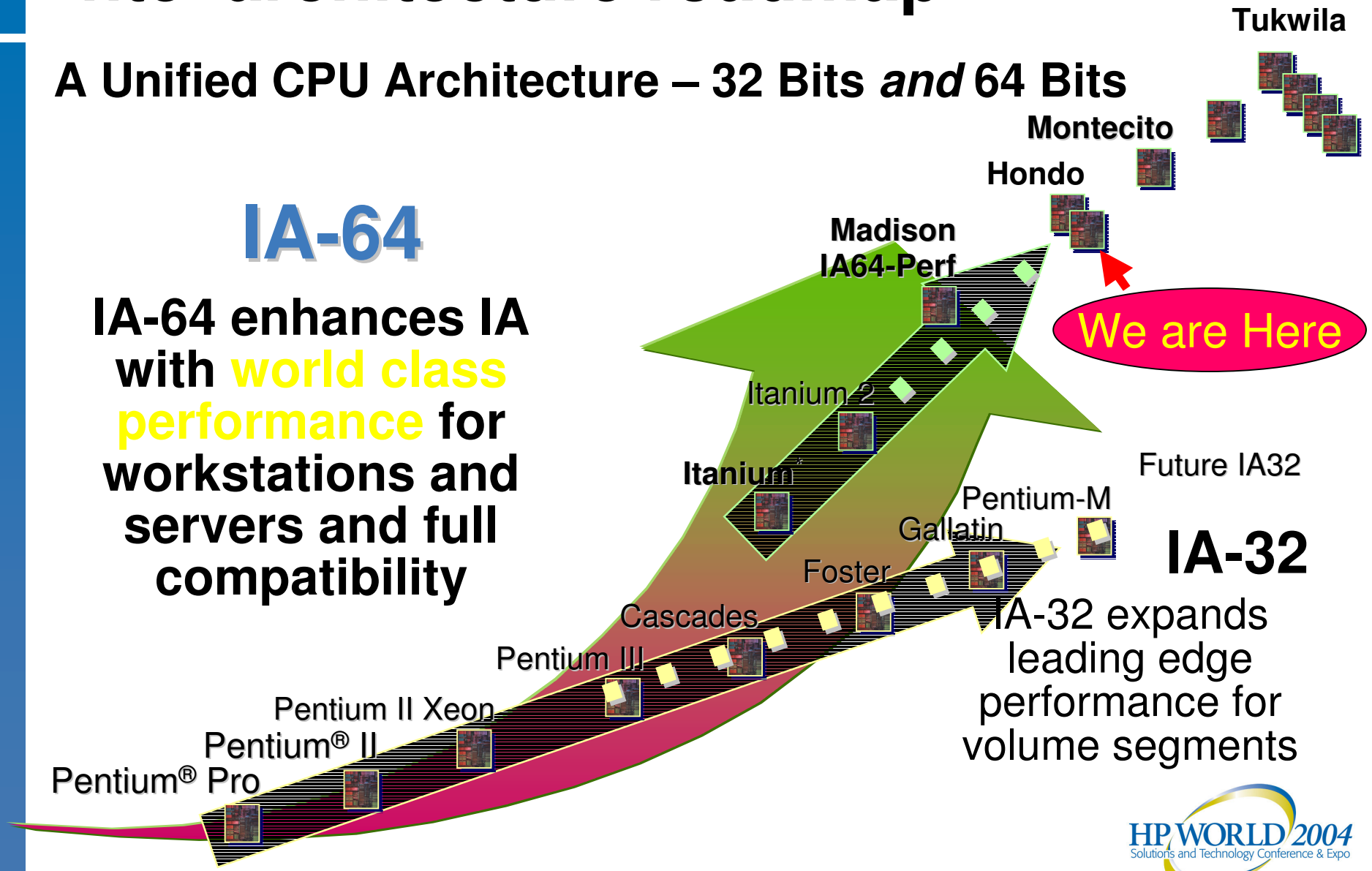
Increased parallelization – more throughput

Intel architecture roadmap

A Unified CPU Architecture – 32 Bits *and* 64 Bits

IA-64

IA-64 enhances IA with world class performance for workstations and servers and full compatibility



IA-32

IA-32 expands leading edge performance for volume segments

Things were looking good for Itanium

- Itanium finally arrives
  - Unfashionably late
  - Disappointing in performance
- McKinley successor arrives a year later
  - HP takes charge of compiler development
  - Higher performance, broader acceptance
  - Products begin to proliferate, HP Integrity thrust begins
- Madison debut imminent
  - Performance projections portended leadership
- Then along comes 18 February 2004

Itanium interruptus? No way!

- Breaking News from IDF
  - HP will offer AMD Opteron and Intel Nocona x86 CPUs with 64-bit extensions on select ProLiant models.
  - HP's commitment to Xeon & Itanium as strategic industry standard platforms for the future is unchanged.
    - Integrity for enterprise scalability and performance
    - Itanium NSK systems get NSAA performance gains
- Itanium still intact and on track

Fister and Robison

HP WORLD 2004
Solutions and Technology Conference & Expo

What is 64-bit Extension Technology?

+

Extended Memory Addressability
64-Bit Pointers, Registers

Additional Registers
8-SSE & 8-Gen Purpose

Double Precision (64-bit) Integer Support

=

With 64-Bit Extension Technology

Evolutionary IA-32 architectural enhancements to support extended memory starting in mid '04

And what do extensions give me?

- More choice in the ProLiant space
  - Compaq developed, but didn't ship, an Alpha ProLiant
  - Opteron option available now, Nocona next
  - Consistent design across all ProLiant product lines, so you won't see anything radically new in the new boxes.
  - Systems currently offered in three form factors
    - 100 series – economy-class HPC
    - 500 series - 4P performance for power-starved apps
    - BL series - B lades run cool, use less fuel are great for HPTC
  - Benefits both 32b and 64b users
    - Linux and Windows supported in 64-bit mode
    - 32-bit users can address as much as 64b of memory while gaining a performance boost.

What do else do extensions provide?

- 64-bit wide general purpose registers, instruction pointer and operations
- 48-bits of virtual addressing in initial product implementations
- Up to 40-bits physical addressing
  - Nocona & future Prescott 36-bits (64 GB)
  - Potomac 40-bits (1 TB)
- Flat virtual addressing space
- Instruction pointer relative addressing mode
- 8 new general purpose registers (GPR)
- 8 new SSE registers (128-bit)

So why bother with Itanium?

- 64-bit computing is inevitable, and now it's sooner
- x86 with 64-bit extension technology will speed the development of a full 64-bit ecosystem
  - Day one 64-bit Windows and Linux support
  - Accelerating 32-bit to 64-bit apps upgrade rates
  - 32-bit apps freed 4GB addressing limit of 32-bit CPUs
  - And an unintended consequence for Itanium
    - Intel must increase the frequency and magnitude of Itanium price cuts to compete in volume market
    - Intel's loss is Itanium customers' gain

What? Extensions are not a panacea?

- X86 extensions are not all things to all people, except in press articles.
- There's plenty of areas in which Itanium adds value. One area reflects the EPIC architecture.
- EPIC Performance
  - Freedom from RISC limits deliver lower latency and higher throughput
  - High performance through parallelization
    - Up to 2x performance per clock cycle
  - Massive on-chip resources to boost apps performance
  - Lower memory latency gap with shorter pipelines

What else can Itanium do?

- Itanium can go places where x86 isn't an option.
- Enterprise support
- If you're an enterprise user who wants a system larger than 8P and you require support for VMS, NSK or Tru64 Unix, Itanium provides all of the above, as well as 64-bit Unix and windows.
  - Indeed, you can run four of these OSes concurrently in separate hard partitions on a Superdome.
  - Due to architectural necessity, NSK requires its own box, but it remains key component of HP's AE strategy.
- As a new architecture, EPIC is built to last…

Choose Your Weapon… or CPU

- IF… a 1P to 8P ProLiant running 64-bit Windows or Linux (or even Solaris) meets your needs…
- THEN… Opteron and Nocona are a good choice
- IF… you need a 1 to 128P enterprise system that can run HP-UX, NSK, or VMS as well as the commodity OSes listed above while delivering superior RAS and manageability capabilities…
- THEN: HP Integrity and Itanium are for you!

Some “pressing“ architectural issues

- Freedom of the press: upsides are many
- Downside (from some experienced in the trade)
  - During the past 20 years, the US trade press has shown a significant decline in quality
  - The job of a reporter is to report facts, not gather a few facts and, report their opinions as statements of fact.
  - Jumping to a conclusion is one thing, but any journalist who makes quantum leaps to conclusions needs firing.
- Conclusions
  - Failure to stick to the facts can injure a person or business or lead to misunderstandings among readers.

The press makes an EPIC assumption

- "X86 with extensions is equivalent to EPIC"
  - Completely inaccurate, but a good example of coming to a conclusion before checking all of the facts.
- A few facts the press failed to consider
  - Extending a 25-year-old 32-bit architecture to support extended memory addressing and 64-bit versions of two volume OSes does not render Opteron or Nocona the equal of Itanium.
  - Opteron and Nocona lack the scalability and RAS capabilities required by enterprise systems.
    - Opteron and Nocona do not support enterprise OSes (other than Solaris). There are no plans to port VMS or NSK or HP-UX to these platforms, and feasibility of doing so is unknown.

Is Itanium dead? I read this article…

- Was the article backed up by facts? Probably not!
  - No. Itanium's alive, the press has a problem distinguishing between news and opinion. Printing a news article proclaiming the death of Itanium is bad journalism generated by bad writers.
  - News is factual information backed by evidence and proof points. The debut of X86 extensions was news.
  - The impact of this news on Itanium remains uncertain and open to debate. The issue merited coverage on the editorial page, where opinions and assertions belong.

How do extensions affect HP?

- Does this change HP's industry standard server or Itanium processor strategy?
  - No. The ProLiant strategy is validated, and Itanium remains the 64-bit enterprise CPU of choice.
- Is Integrity and ProLiant positioning a big problem?
  - No. Differentiating low-end Integrity and ProLiant products is now an issue. Overlapping capabilities in the 1P to 8P space denies HP the ability to differentiate its low-end systems based on CPU count. Customers will choose the product that best fits their needs.

Itanium impact

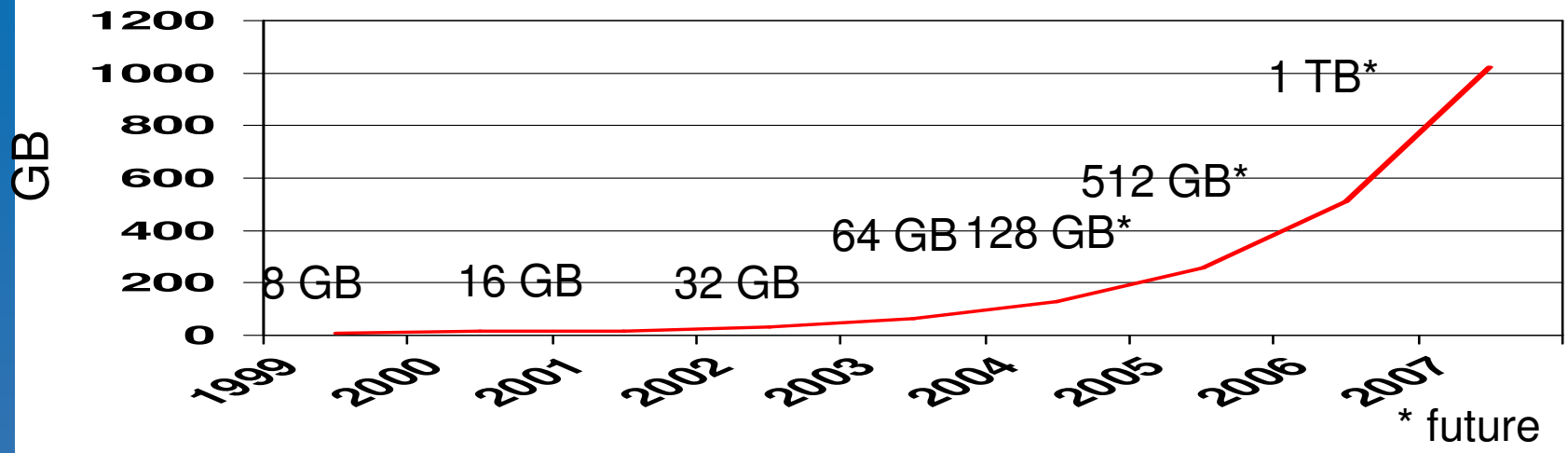
- Will this affect the acceptance of Itanium CPUs?
  - Slightly, but not to the extent that adoption rates will be impacted in Linux HPTC clusters and the low end market defined by CPU count and OS support. If an 8P system supporting 64-bit Linux and Windows meets a customer's needs, Opteron or Nocona is a cost-effective solution.
- What about the Itanium adoption rates?
  - They will not meet Intel's expectations in the near term, as Intel wanted the volume market. The only way Intel can enter this market is by accelerating the frequency and magnitude of its Itanium price cuts. Customers will get faster Itanium systems sooner

Itanium distinguishing features

- Performance through Parallelization
  - Advantage: Up to 2x performance per clock cycle
  - Proof point: Best SpecFP
- Massive On-chip Resources
  - Advantage Imaging, rich data, voice, encryption
  - Proof point: Superior Linpack results
- Business Critical Availability
  - Advantage: Machine check architecture
    - Inherent protection and security features
- Shorter pipelines beat the memory latency gap
  - Advantage: SMP complex workloads: OLTP, data mining
    Proof point: TPC-C, TPC-H benchmarks

64-bit computing becomes pervasive

4p server memory will exceed 1 TB by 2007



Applications are a driving force to 64-bit computing

- Real time security - Biometrics, encryption, virus scanning
- Huge data warehousing (database and data mining)
- Rich data types, complex technical workloads

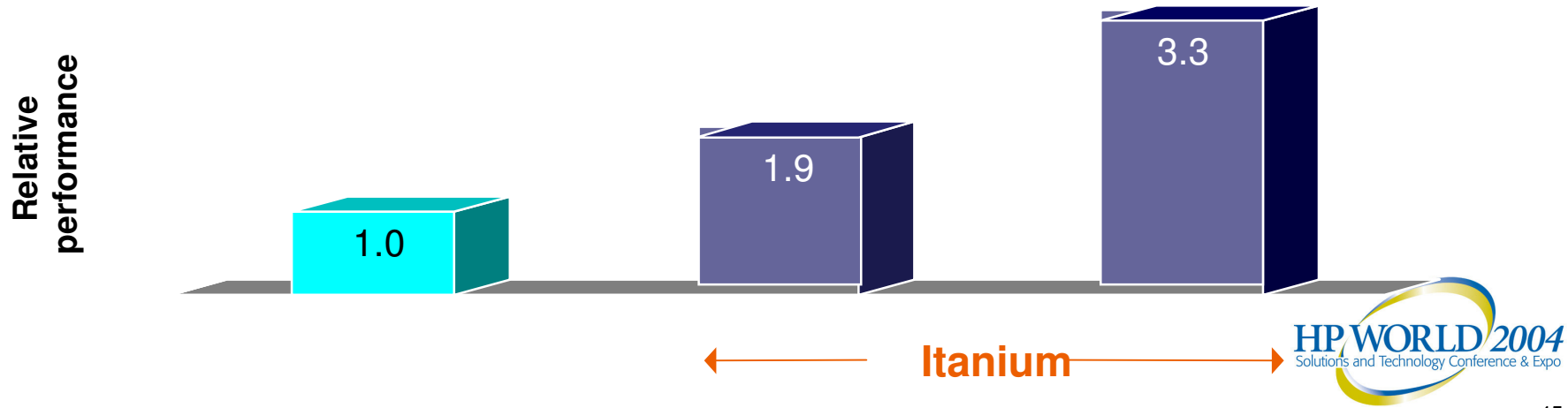
Itanium: THE Technology for mission critical computing at HP

Itanium's Unique Advantages

- **Built-in instruction-level parallelism**
- **Massive on-chip resources**
- **up to 2X instructions/clock cycle**
- **CPU clock and compiler maturity curve**
- **Fewer memory loads/stores on complex workloads**

Customer Benefits

- **Higher performance in FP-intensive and complex technical workloads: MCAE, Transforms**
- **2X performance of x86, at any clock speed, for faster:**
  - **Image manipulation**
  - **Voice encoding/recognition**
  - **Encryption**



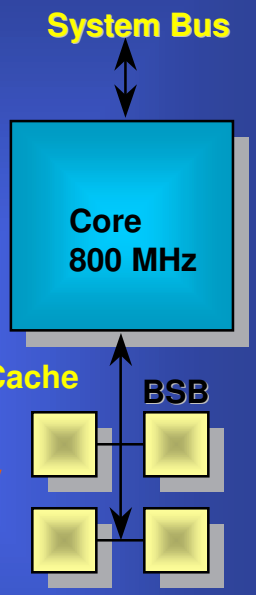
Itanium Processor

System Bus
64 bits wide
133MHz/266 MT/s
2.1 GB/s

Width
2 bundles per clock
4 integer units
2 load or stores per clock
9 issue ports

Caches
L1 – 2X16KB - 2 clock latency
L2 – 96K – 12 clock latency
L3 - 4MB external –20 clk
11.7 GB/s bandwidth

Addressing
44 bit physical addressing
50 bit virtual addressing
Maximum page size of 256MB



Itanium2 / McKinley / Madison

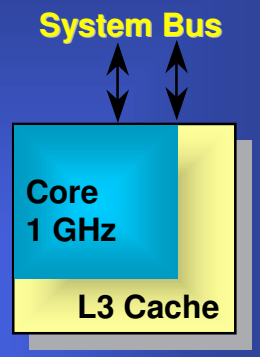
System Bus
128 bits wide
200MHz/400 MT/s
6.4 GB/s

50% Increase in Clock Rate

Width
2 bundles per clock
6 integer units
2 loads and 2 stores per clock
11 issue ports

Caches
L1 – 2X16KB - 1 clock latency
L2 – 256K – 5 clock latency
L3 - 3MB / 6MB – 12 clk
32 GB/s bandwidth

Addressing
50 bit physical addressing
64 bit virtual addressing
Maximum page size of 4GB



At HP, Itanium and standards rule!

- HP has the most complete portfolio of any vendor offering Itanium solutions
- For example:
  - Hardware
  - Applications
  - Migration tools
  - Assessment services
  - Online assistance
- And it's working: HP Integrity sales are growing

Recent Integrity Announcements

- 3 November 2003: HP Fleshes Out Integrity Line…
- New products include…
  - Itanium-based HP Integrity rx4640 4P server
  - Itanium-based HP Integrity rx7620 8P server
  - Itanium-based HP Integrity rx8620 16P midrange servers
- The rx7620 and rx8620 servers feature the Superdome cellular architecture, and thus are "Superdome Juniors." CPU capacity can be doubled on cellular systems with the new Hondo module.

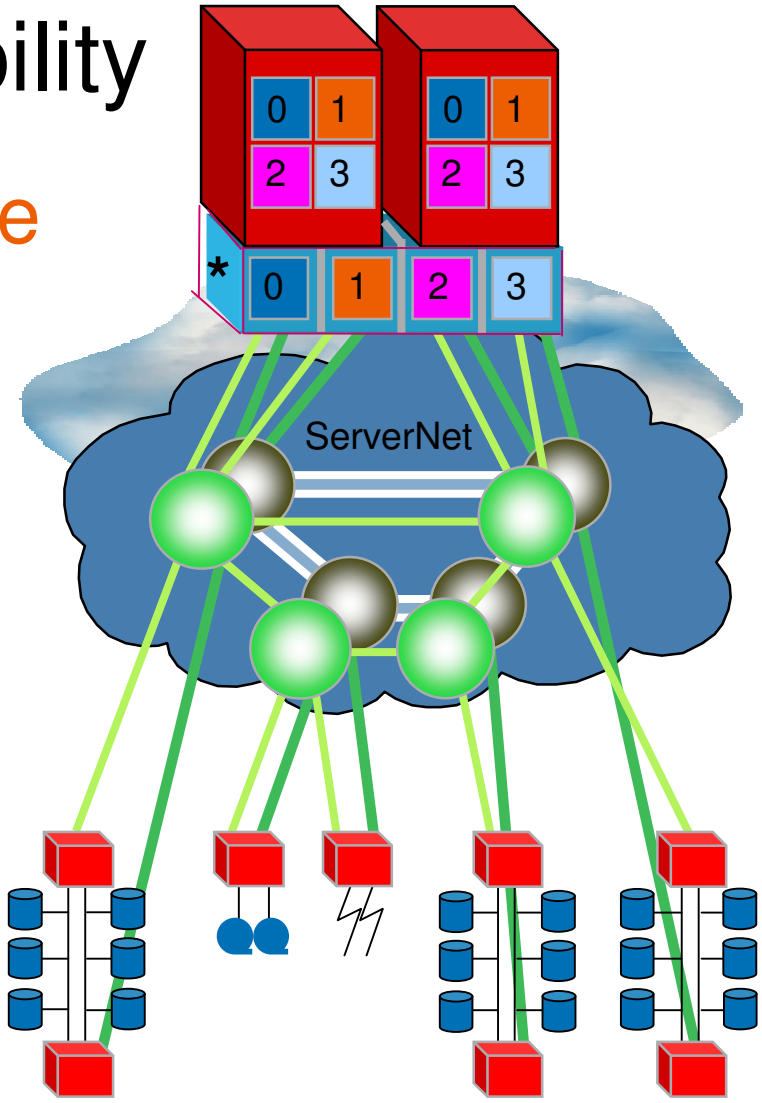
More Recent Announcements

- Also announced was the cost-reduced HP ProLiant 100 server series, including its first member, the ProLiant 140, then the Opteron-based 145 and 585.
- And HP addressed the HPTC community with the
  - HP XC6000 cluster,based on the HP Integrity rx2600 systems
  - HP XC3000 cluster is based on HP ProLiant servers.
  - Both clusters support up to 512 nodes
  - Both support high-speed interconnects from Quadrics LTD and Myricom.

NSK Users: New Itanium-based HP NSAA Delivers Greater Availability

HP Integrity architecture is the foundation for the industry's leading fault-tolerant solution

- Intel Itanium processors and Integrity board sets enable new design that delivers better fault tolerance
- Individual microprocessors on a four-processor "slice" run as separate CPUs
- Recovers transparently from multiple hardware or software faults



* = logical synchronization units

Why IPF… And Why HP?

- IPF is the logical RISC successor
- Intel and HP co-developed Itanium
- Both firms depend on Itanium
- Itanium will get annual “speed bumps”
- HP has the industry’s broadest portfolio of
  - IPF systems and future IPF systems
  - Applications
  - OS support
  - Customer support
- **So why NOT IPF… and why NOT HP?**

Just my opinion, and I could be wrong…

- After 30 years in the business, I believe the current state of affairs will rank on the Top Ten List of IT Industry "Tales of the Unexpected."
- If I was responsible for damage control, I would…
  - Assert that x86 extensions are not inherently evil
  - Emphasize customer choice
  - Separate fact from fiction
  - Answer the inevitable questions
  - Stress Itanium's strong points
  - Show where HP intends to take Itanium in the future
- I hope I've made some progress in this direction!

Thanks for your time and attention…

- And thanks to HP, Interex, and the “folks behind the curtain” and the organizers who made HP World 2004 a success.

- Questions or comments? Email me at
- terry@shannonknowshpc.com
- Enjoy the rest of the show, fill out your evaluation forms, and I hope to see you all at HP World 2005!

HP WORLD 2004
Solutions and Technology Conference & Expo

HP WORLD 2004

Solutions and Technology Conference & Expo

Co-produced by:

interex

shared knowledge • shared power

encompass

AN HP USER GROUP