State of the art processors

IA32 compatible: Intel P4/Xeon – AMD Opteron/Athlon
IBM Power4
IA64: Intel Itanium2
Vector processors: NEC SX6

Dr. Gerhard Wellein
G. Hager, T. Zeiser

HPC Services – Regionales Rechenzentrum Erlangen
University Erlangen-Nürnberg
Sommersemester 2005

IA-32 Architecture
A Little History

- IA-32 has roots dating back to the early 80s
  - **Intel's first 16-bit CPU: 8086 + 8087 math coprocessor (x86 is born)**
  - **Even the latest Pentium 4 is still binary compatible with 8086**
  - **loads of advances over the last 20 years:**
    - **addressing range (1MB → 16MB → Terabytes)**
    - **protected mode (80286)**
    - **32 bit GPRs and usable protected mode (80386)**
    - **on-chip caches (80486)**
    - **SIMD extensions and superscalarity (Pentium)**
    - **CISC-to-RISC translation and out-of-order superscalar processing (Pentium Pro)**
    - **floating-point SIMD with SSE and SSE2 (Pentium III/4)**
- Competitive in HPC: Pentium III
  - **Pentium 4 is today rivaling all other established processor architectures**
- First IA32 compatible 64-Bit processor: AMD Opteron (Besides Intel Itanium series)

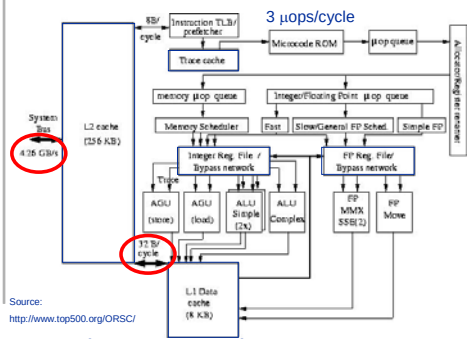
gerhard.wellein@rrze.uni-erlangen.de

IA-32 Architecture
Basics

- IA-32 has a CISC instruction set
  - **„Complex Instruction Set Computing"**
  - **Operations like „load value from memory, add to register and store result in register" are possible in one single machine instruction**
  - **Huge number of assembler instructions**
  - **Very compact code possible**
  - **Hard to interpret for CPU hardware, difficulties with out-of-order processing**
- Pentium Pro: CISC to RISC (called µOps here) on the fly
  - **„Reduced Instruction Set Computing"**
  - **Very simple instructions like „load value from memory to register" o „add two registers and store result in another"**
  - **RISC instructions are held in a reorder buffer for later out-of-order processing**

gerhard.wellein@rrze.uni-erlangen.de

IA-32 Architecture:
Pentium 4 Block Diagram



Source:
http://www.top500.org/ORSC/

IA32 Architectures
Intel Xeon: FP performance, Bandwidths &Balance

- FP performance
  - 1 Mult & 1 Add per cycle (double precision)
  - 2 Mult & 2 Add per cycle (single precision + SSE instructions)
- L1 caches
  - **16k bytes instruction and 8 kB data**
  - **4-way set associative (data)**
  - **4 word / cycle** 2 Word / Flop
- L2 cache
  - **Caches instruction and data streams**
  - **512 kB; 8-way set associative**
  - **4 word / cycle** 2 Word / Flop
- Main Memory
  - **1 Word@533 MHz**
  - **3 GHZ CPU** 0.09 Word / Flop
- Two way SMP nodes
  - **2*3 GHZ CPU** 0.045 Word / Flop

gerhard.wellein@rrze.uni-erlangen.de

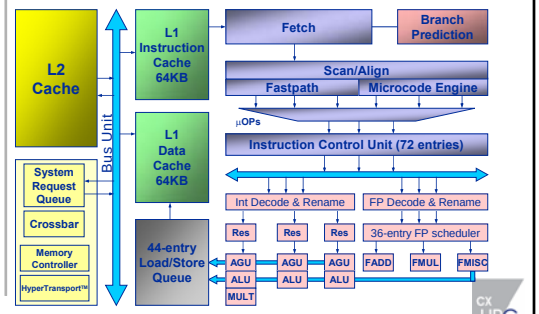
IA32 Architectures
Intel EM64T / AMD Opteron

- Intel P4/ Xeon uses 32-Bit addressing (address registers, address bus, etc) - can only address 2 GBytes of memory (at reasonable performance)
- Intel EM64T – Extend Memory 64 Technology
  - Extends address space to 64 Bit (uses 36-Bit for physical addresses)
  - Doubles the number of SSE registers
  - Available with up to 2 BM L2 cache and 3.6 GHz
- AMD Opteron: IA32 compatible with 64-Bit addressing (uses 40-Bit for physical addresses)
- AMD Opteron: Binary compatible with Intel Xeon/P4 series!!
- AMD Opteron: Extended Register Set

64-Bit Processors do not necessarily compute faster or with higher precision compared to 32-Bit processor.
It is only about addressing!

gerhard.wellein@rrze.uni-erlangen.de

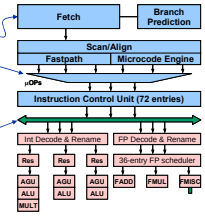
IA32 Architectures
AMD Opteron



gerhard.wellein@rrze.uni-erlangen.de

Processor Throughput

- Supply 16 instruction bytes to the decoder per cycle
- Convert x86 instructions to fixed length µOPs
- 24-entry integer scheduler can Dispatch 3 µOPs per cycle to integer/FP schedulers
- Instructions use one of two decoding pipelines
  - **Fastpath:** instructions which are decoded in to two or fewer mOPs are decoded by hardware and then packed into 3 dispatch positions
  - **Microcode:** x86 instructions which are decoded in to more than two mOPs, calculate microcode ROM entry point and fetch sequence from Microcode ROM



- Compared to AMD Athlon™ XP, more instructions use the Fastpath
  - Eg: Packed SSE is microcoded in AMD Athlon™ XP and Fastpath in AMD Opteron ™ processors
  - AMD Opteron™ has 8% fewer microcoded instructions for SPECint2000
  - AMD Opteron™ has 28% fewer microcoded instructions for SPECfp2000

gerhard.wellein@rrze.uni-erlangen.de

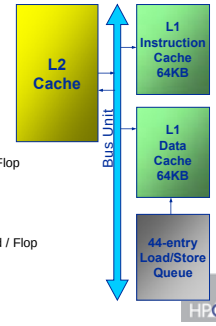
IA32 Architectures
AMD Opteron: Floating Point & Integer Performance

- **FPU Throughput**
  - SSE2, x87
    - ✓ Theoretical (double precision): (1 Mul + 1 Add)/cycle
    - ✓ Realized: 1.9 FLOPs/cycle
  - SSE, 3DNow!
    - ✓ Theoretical (single precision): (2 Mul + 2 Add)/cycle
    - ✓ Realized: 3.4+ FLOPs/cycle
- **32-bit Integer Throughput**
  - 1 add / clock cycle + 1 multiply / clock cycle
  - Multiply latency: 3 cycles
- **64-bit Integer Throughput**
  - 1 add / clock cycle +1 multiply every other clock cycle
  - Multiply latency: 4 cycles

gerhard.wellein@rrze.uni-erlangen.de

IA32 Architectures
AMD Opteron: Bandwidths and balance

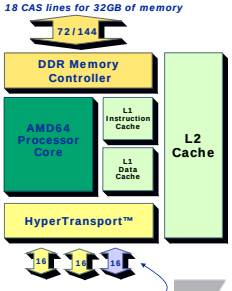
- L1 caches
  - **64k bytes instruction and data**
  - **2-way set associative**
  - **2 word / cycle** 1 Word / Flop
- L2 cache
  - **Caches instruction and data streams**
  - **16-way set associative**
  - **Victim Cache**
  - **2 word / cycle** 1 Word / Flop
- Main Memory
  - **On-Chip Memory Controller**
  - **2 Words@333 MHz**
  - **2 GHZ CPU** 0.17 Word / Flop



gerhard.wellein@rrze.uni-erlangen.de

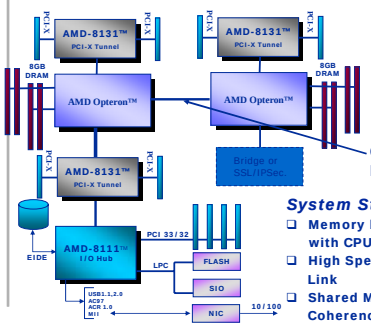
IA32 Architectures
AMD Opteron™ hypertransport

- Three 16-bit HyperTransport™ Links at max. 800 MHz (using DDR-RAM)
- Coherent HT links
  - 3.2 GB/s in each direction
  - shared memory &cache coher.
- One coherent link is used to built 2-way servers
- Three coherent links are used to build 8-way servers
- Memory bandwidth scales with CPU count!



gerhard.wellein@rrze.uni-erlangen.de
Courtesy: W. Seyfried, AMD

IA32 Architectures
2P AMD Opteron™ 200 Server

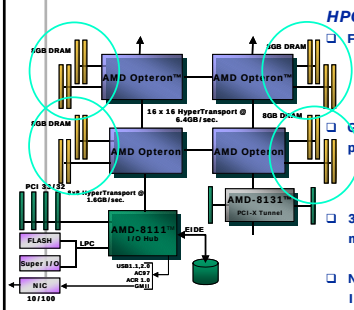


System Strengths

- **Memory bandwidth scales with CPU count**
- **High Speed Hypertransport Link**
- **Shared Memory & Cache Coherency**

gerhard.wellein@rrze.uni-erlangen.de
Courtesy: W. Seyfried, AMD

IA32 Architectures: AMD Opteron™ 800 HPC



HPC Strengths

- **Flat SMP like Memory Model:**
  - All four reside with the same 2^{48} memory map
  - Expandable to 8P NUMA
- **Glue-less Coherent multi-processing:**
  - low Latency and high Bandwidth ~1600M T/sec (6.4 GB/s)
- **32GB of High B/W external memory bus (>5.3GB/sec.)**
- **Native high B/W memory map I/O (>25Gbits/sec.)**

gerhard.wellein@rrze.uni-erlangen.de
Courtesy: W. Seyfried, AMD

The AMD Opteron™ processor

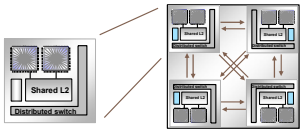
Project "Red Storm"

Sandia National Laboratories CRAY

- ✓ **Cray will build a 40+ teraflop super computer using x86-64 AMD Opteron™ processors for Sandia National Laboratories**
- ✓ **Will be used for advanced engineering simulations**
- ✓ **$90 million project will use more than 10,000 AMD Opteron™ processors**
- ✓ **Will feature a simple building block approach with HyperTransport™ technology that will enable easy implementation and reduce engineering, design, and component costs**

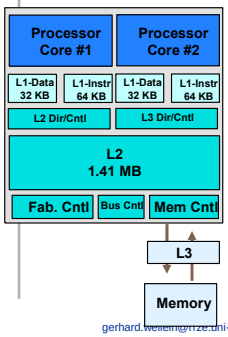
gerhard.wellein@rrze.uni-erlangen.de
Courtesy: W. Seyfried, AMD

Architecture of IBM Power4 and IBM p690 Series



gerhard.wellein@rrze.uni-erlangen.de

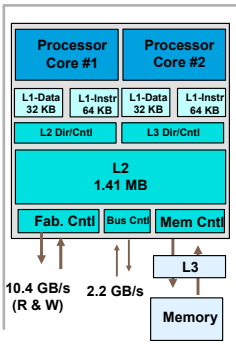
Current RISC architectures: IBM Power4



- each core:
  - Up to 1.7 GHz/ 6.8 GFlop/s
  - **8 Pre-fetch Streams from L2, L3, mem.**
  - **8 outstanding cache misses**
- L1
  - **64KB instruction, direct, 128B lines**
  - **32KB data, 2-way associative,**
  - **128B lines, store through**
  - **1 load/store per cycle**
- L2
  - **3 slices of .47 MB**
  - **8 way associative, 128B lines**
  - **1 load +1 store / cpu**
  - **3* 32B*1.3 Ghz= 124 GB/s**
  - **10-12 cycles latency (9.1-10.9 ns)**

gerhard.wellein@rrze.uni-erlangen.de
F. Krämer, IBM

Current RISC architectures: IBM Power4



- L3
  - 32 MB/chip **(external)**
  - **8-way associative,** 512B lines
  - **16 B 1:3= 6.9 GB/s read**
  - **16 B 1:3= 6.9 GB/s write**
  - **total BW/chip 13.8 GBs**
  - 92/100 cycles latency **(local/remote)**
- Memory
  - **up to 64/128 GB/ MCM**
  - **1 or 2 port memory cards:**
  - 6.4 GB/s read per port
  - 6.4 GB/s write per port
  - **252 cycles latency**

gerhard.wellein@rrze.uni-erlangen.de
F. Krämer, IBM

Current RISC architectures: IBM Power4

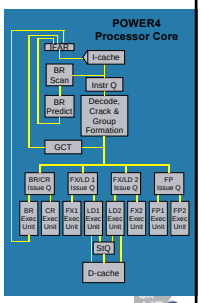
Speculative superscalar organization
- **out of order execution**
- **large rename pools**
- **8 instruction issue/ 5 complete**

8 execution pipelines
- **2 load/store units**
- **2 FX units**
- **2 DP multiply-add units => 4 flops/cycle**
- **1 branch resolution unit**
- **1 CR execution unit**

Hardware pre-fetching
- **8 streams per cpu**

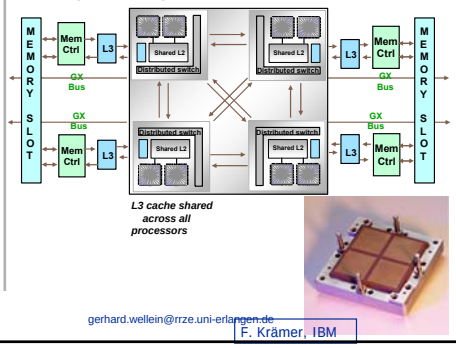
Outstanding cache misses
- **8 on data / 3 on instruction**

Aggressive branch prediction
- **target address and outcome prediction**
- **static prediction + branch hints used**



gerhard.wellein@rrze.uni-erlangen.de
F. Krämer, IBM

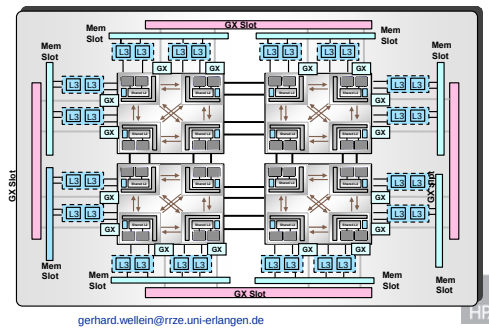
Current RISC architectures: IBM Power4 Multi-Chip Module



F. Krämer, IBM

gerhard.wellein@rrze.uni-erlangen.de

Current RISC architectures: IBM Power4 SMP System Design



gerhard.wellein@rrze.uni-erlangen.de

Itanium: Intel´s new 64 Bit Architecture: What´s so new about it?

- **It´s not RISC:**
  Compiler generates bundles with three instructions each

Length: 41 Bit each

- **It´s not VLIW:**
  Compiler generates a 5-Bit template field containing information about instruction level parallelism

- **This is called Explicity Parallel Instruction Computing (EPIC)**

gerhard.wellein@rrze.uni-erlangen.de

Itanium: Intel´s new 64 Bit Architecture Principles of EPIC

- Template field provides information about
  - the instruction type / execution unit of each instruction
  - independent instruction groups (instruction level parallelism)
- Template field allows only a limited number of instruction combinations (12) in each bundle, e.g.:
  - a maximum of one Floating-Point-Instruction per bundle
  - a maximum of two Memory-Instruction per bundle
- Template field may include up to two stop operations (;;) to mark the end of an independent instruction group:

Template	I0	I1	I2
0E	M	M	F
0F	M	M	F ;;
0B	M ;;	M	I ;;

gerhard.wellein@rrze.uni-erlangen.de

Itanium: Intel´s new 64 Bit Architecture Principles of EPIC

- If instruction sequence does not fit one of the templates the compiler will pack one or two nop instructions into the bundle
- No correlation between (independent) instruction groups and beginning or end of bundles

gerhard.wellein@rrze.uni-erlangen.de

Itanium: Intel´s new 64 Bit Architecture
Principles of EPIC

- Itanium1/Itanium2 can execute two bundles at each processor cycle
  → maximum of 6 instructions per cycle
  → 6 way "superscalar"
- No out-of-order execution
- Ability of compiler to determine (independent) instruction groups is crucial for Itanium processors
  - Use appropriate algorithms
  - Do not hide independence of operations
  - Use compiler directives (!DEC$ IVDEP)
  - Enable software pipelining (-opt_report provides information about compiler opts.) (force (disable SWP) !DEC$ (NO)SWP)

gerhard.wellein@rrze.uni-erlangen.de

Itanium2: Intel´s current 64 Bit processor
Processor Specifications

- Processor frequencies (today): 1 – 1,6 GHz
- Functional units:
  - 6 INTEGER
  - 3 BRANCH
  - 2 FLOATING POINT (FMA) → max. 4 Flop/Cycle
- Registers:
  - 128 Floating Point (82 Bit)
  - 128 Integer (64-Bit)
  - 64 Branch
  - 64 Predicate
  - Rotating Register / Register Stack

gerhard.wellein@rrze.uni-erlangen.de

Itanium2: Intel´s current 64 Bit processor
Processor Specifications

- L1 cache (instruction)
  - 16 kB; 4-way set associative; 64 byte cache line
- L1 cache (data)
  - 16 kB; 4-way set associative; 64 byte cache line
  - no floating point data (L1-cache bypass)
  - Latency: 1 cycle / write through
  - 4 references per cycle
- L2 cache (unified)
  - 256 kB; 8-way set associative; 128 byte cache line
  - Latency: 5 cycles (LD) ; 7 cycles (ST) / write back
  - 2 LD & 2 LD/ST per cycle

gerhard.wellein@rrze.uni-erlangen.de

Itanium2: Intel´s current 64 Bit processor
Processor Specifications

- L3 cache (unified)
  - 1.5 MB or 3 MB ,or 6 MB on die!
  - 6-way or 12-way set associative; 128 byte Cache line
  - Latency: 7 Cycles (ST) ; 12 cycles (LD) / write back
- System/memory bus
  - 128 Bits wide, running at 400 MHz ("double-pumped" 200 MHz)
  - Bandwidth: 6.4 GB/s
  - Max. of 18 outstanding Bus requests per CPU
- Addressing
  - 50 Bit physical / 64 Bit virtual
  - Max. page size: 4 GB (current Linux Kernel limit: max. 64 kB)

gerhard.wellein@rrze.uni-erlangen.de

Vector processors

- Vector processors and vector computers are technologies mainly designed to achieve high sustained performance in scientific simulations
- Vector processors show best balance but are extremely expensive
- Vector technology has almost vanished in the past 10 years (less than 3% of TOP500 computers are vector computers) but now starts to recover ……

gerhard.wellein@rrze.uni-erlangen.de

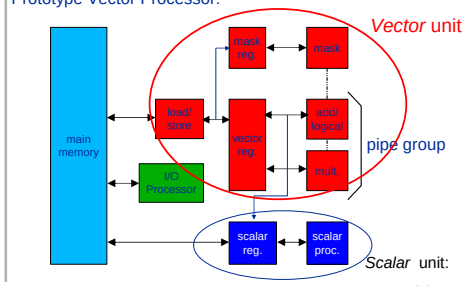
Key Features of a Vector Processor

- Vector instruction (one single machine instruction) performs the instruction on a whole vector/array, e.g.
  VADD: Add two vector/arrays
- Length of vector/array: "Vector length" (fixed !)
- Register to register machine, however a single vector register always contains a vector
- Vector pipelines deliver one result per clock cycle
- Pipelines can be chained so that even more is possible (e.g. 1 Mult & 1 Add)
- Multi-track pipelines (8 or 16 way) increase parallelism even further, e.g. a 8-way Multiply pipeline delivers 8 results per clock cycle
- Non-vectorizable operations (e.g. a(i)=a(i)-a(i-1)) are performed on a separate scalar unit
- Scalar units are quite slow:
  Use vector units as far as possible!

gerhard.wellein@rrze.uni-erlangen.de

Vector Processor

Prototype Vector Processor:



Scalar unit: e.g. RISC

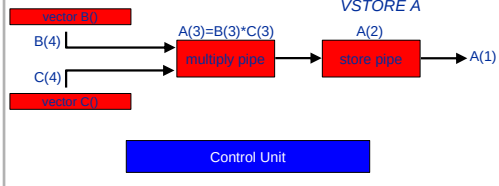
gerhard.wellein@rrze.uni-erlangen.de

Vector Processor Chaining

A vector processor has separate functional units for all steps of a vector-like operation!

Example: **A(1:N)=B(1:N)*C(1:N)**

VLOAD C
VLOAD B
VMULT B, C -> A
VSTORE A



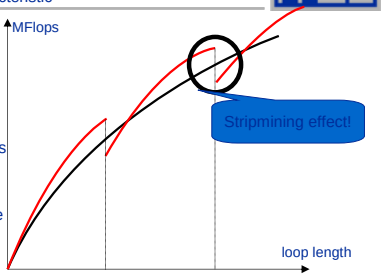
Control Unit

Chaining: Several pipelines can run at the same time and deliver their results to other pipelines (via vector registers)

gerhard.wellein@rrze.uni-erlangen.de

Vector Processor Performance Characteristic

- Non-negligible overhead in setting up vector registers and filling the pipeline!
- Fixed vector length: Large arrays have to be handled by stripmining
- Max. performance for large data sets!
- Memory bus runs at processor speed



gerhard.wellein@rrze.uni-erlangen.de

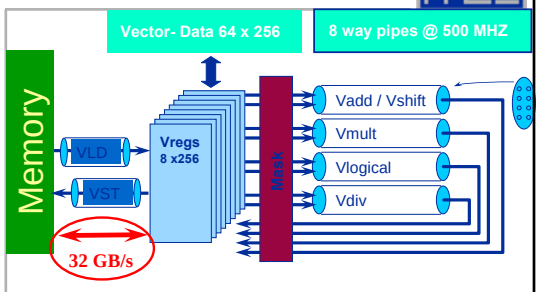
Vector Processor History

- *Famous series of CRAY vector machines (1976 – 1995):*
  - *CRAY 1 (80 MHz; 1976)*
  - *CRAY YMP (167 MHz; 1988)*
  - *CRAY T90 (450 MHz; 1995)*

  → 2 or 3 memory operations per processor cycles; ECL technology
- *New CRAY vector machine available since 2003: CRAY X1 (CMOS; RD-RAM!)*
- *Fujitsu built vector systems from 1983 (VPP400) till 2000 (VPP5000). Fujitsu moved from ECL to CMOS in 1990s.*
- *NEC: SX1 (1993; ECL) – SX4 (1994; CMOS) – SX6 (2001; single chip) - SX-8 (2005)*

gerhard.wellein@rrze.uni-erlangen.de

Vector Processor NEC SX6 – schematic view



gerhard.wellein@rrze.uni-erlangen.de

Vector Processor NEC SX6: Vector-unit

- Vector pipelines: MULT / ADD / DIV / LOAD / STORE / MASK
- 8-track pipelines ⇒ Max. 8 results per pipeline and cycle
- Two (MULT , ADD) arithmetic pipelines can work simultaneously
- LOAD or STORE can only be done
- All pipelines (and main memeory) run at 500 MHz

Peak Performance: 2 * 8 * 500 MFlops = 8 GFlops

(Peak Performance Pentium 4 @ 3GHz: 6 GFlops)

- Balance: (8 Words / cycle) / (2 * 8 Flop / cycle) = 0.5 Words / Flop
- Balance of P4@3GHz: 0.17 Words / Flop

gerhard.wellein@rrze.uni-erlangen.de

Vector Processor
NEC SX6: Memory access

- 64 Vector register with 256 entries each - > vector length = 256
- Memory capacity per processor: 8 – 16 GBytes
- 1 LOAD or 1 STORE Pipeline (8-track,)
- Memory bus: 500 MHz (=Processor frequency!)

Memory bandwidth: 1 * 8 * 8 *500 MByte/s = 32 GBytes/s

(Pentium 4: 4,3 GByte/s)

- Contiguous memory access!
- 4096 independent memory modules/ banks in one 8-way SMP nodes
  -> Minimize bank conflict and achieve high sustained memory bandwidth

gerhard.wellein@rrze.uni-erlangen.de

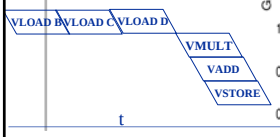
Vector Processor
NEC SX-6 single node system

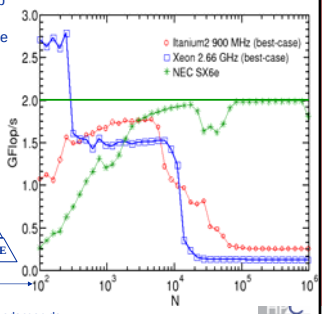
- Microprocessor-based SMP
- Vector supercomputer
- 8 processors: 8 GFlop/s each
- Maximum 64 GFop/s per node
- Maximum 64 Gigabyte memory
- Ultra-high bandwidth shared memory subsystem: 256 GB/s

gerhard.wellein@rrze.uni-erlangen.de

Vector Processor
NEC SX6 performance: vector triad

- Balance of SX6: 0.5 W/Flop
- Balance of triad: 2.0 W/Flop
- Est. performance limit 0.5 / 2.0 * Peak Performance = 2 GFlop/s
- Max. Performance: (1/4)*8 GFlop/s = 2 GFlop/s





gerhard.wellein@rrze.uni-erlangen.de