

## The NEC SX-4 at HWW

[The NEC SX-4 System Architecture](#)

[System Architecture](#)

[Processor Architecture](#)

[Memory Architecture](#)

[The SX-4 at HWW](#)

[System Software and Programming Models for the NEC SX-4](#)

[Resource Blocks](#)

[Vectorisation and Shared Memory Parallelisation](#)

[Message Passing](#)

[Compilers and Tools](#)

[Performance](#)

[RUS Loops](#)

[Performance in real Applications](#)

[Crash-30 Test with LS-DYNA3D](#)

[LAMTUR](#)

[Access](#)

[References](#)

---

World's most powerful Shared Memory Supercomputer starts Production

## The NEC SX-4 at HWW

*Alfred Geiger / Uwe Küster*

The NEC SX-4 which started production at Stuttgart on July 15th 1996, is jointly operated by university and industry in the framework of the HWW Betriebs GmbH. The major goal of HWW is to provide the users in academia and industry with the quality to solve larger problems, the capacity for better throughput and a broader diversity of usable architectures. The machine in Stuttgart is actually the only installed SX-4 equipped with 32 processors worldwide. In the Top 500 list of Summer 1996 [1], it is ranked as the most powerful system overall in Europe and the most powerful shared memory system worldwide. This article is a first description of the system and the HWW environment. It gives some performance numbers in comparison to the Cray C94D, which served Stuttgart University in the last years and makes some links to informations about how to access the system.

HWW (Höchstleistungsrechner für Wissenschaft und Wirtschaft Betriebs GmbH) was founded in 1995 with the purpose of operating supercomputers for science as well as for industry. The actual partners in this company are debis Systemhaus, Porsche, Stuttgart University and the state of Baden-Württemberg. HWW is highly focussed on achieving synergetic effects in supercomputing between industry and university.

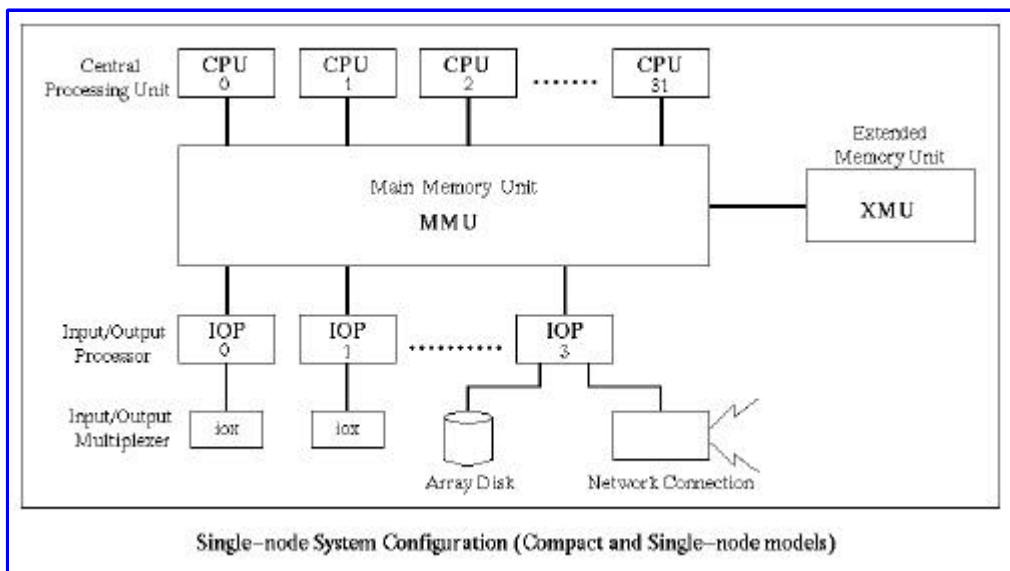
### The NEC SX-4 System Architecture

The SX-4 is NEC's newest generation of parallel vector supercomputers. It follows the trend of all major Japanese manufacturers to architectures which are on a path between the classical vector

machines with their technology at the limit of feasibility and MPPs with pure off-the-shelf technology. NEC tries to have the superior price/performance ratio of MPPs and the proven performance and broad software base of PVPs.

### System Architecture

From the philosophy of its designers, the NEC SX-4 is a two stage architecture. Up to 32 processors with an aggregate peak-performance of 64 GFLOP/s are coupled via a crossbar to 1024 banks with up to 16 GB of memory (SSRAM).



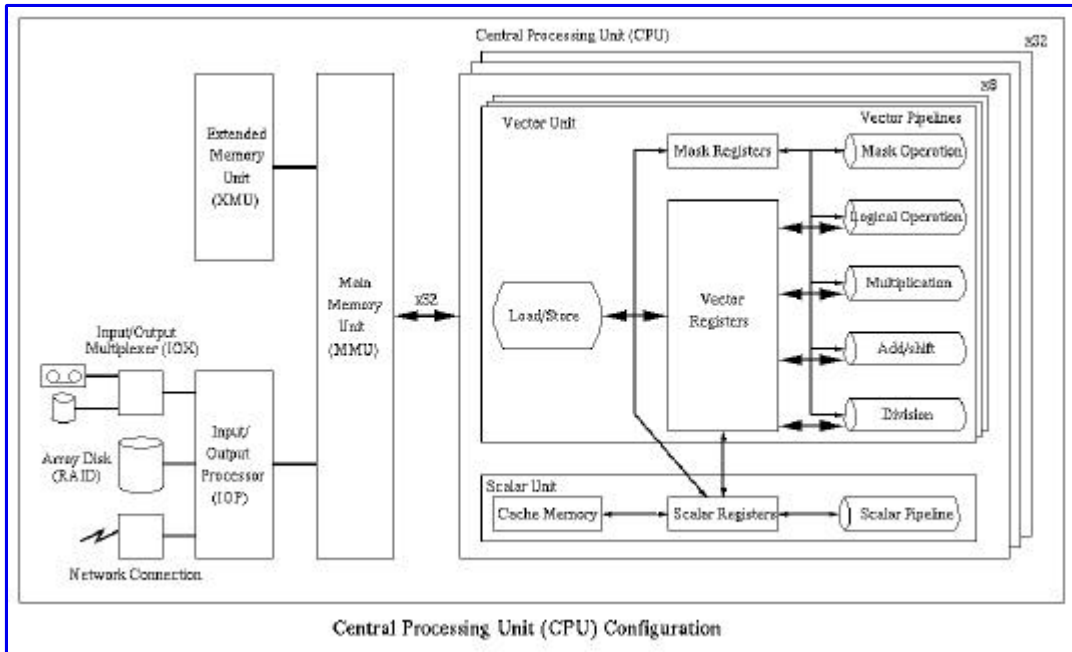
### NEC SX-4 Single-Node

In addition each node supports up to 4 extended memory units (XMU) with a total of up to 32 GB of standard DRAM and 4 I/O-processors (IOP).

On the second level, the architecture goes distributed memory. Up to 16 nodes (512 processors with an aggregate peak-performance of 1 TFLOP/s) are coupled via a crossbar. From the hardware point of view this is simply a cluster, the software however implements a single-system image for the user as well as for the system administrator. However, no system with more than one node was installed up to now.

### Processor Architecture

Each processor consists of a vector unit, a scalar unit and an instruction unit. In difference to earlier NEC machines and also to its most important direct competitor, the Cray T90, the whole CPU is realized in CMOS. The clock cycle is, with 8ns (125 MHz) relatively low, compared with actual ECL-machines.



NEC SX-4 Processor

The vector unit has eight-track pipes, giving a peak-performance of 2 GFLOP/s. The scalar unit has a standard superscalar design with the same clock rate as the vector-unit (Scalar peak rate: 250 MFLOP/s).

The SX-4 processor supports IEEE, Cray and IBM floating point formats and arithmetics at the user's choice at compile time. Although this is not important for academic users, who prefer IEEE arithmetics wherever possible, this solves a major issue for industrial users who hesitate in moving away from IBM or Cray platforms because they have validated their codes with these arithmetics over many years. For them the SX-4 offers a graceful step towards the IEEE standard.

The memory port is able to sustain 16 GB/s, giving one word per floating point operation from the memory and one word to the memory. This is of course not enough for real life and a drawback compared with Y-MP like architectures with their 3 paths to memory per processor.

### Memory Architecture

The memory is organized in 1024 banks of SSRAM (Synchronous Static RAM) memory modules. The main difference of the newly developed SSRAM against the classical SRAM is, that a memory request can be accepted during delivery of another. Address resolutions and memory requests are pipelined. This can be viewed as an extension of the model of vector processing to the memory subsystem. By this way, SSRAM doubles the sustainable bandwidth. The SX-4 is the first supercomputer based on SSRAM in the market, but others will switch to this technology in the near future. A similar technology for DRAM is also developed (SDRAM) and will show up in products pretty soon, allowing for memory bandwidths in the range of SRAM at the price of DRAM.

The processor/memory interface is realized by a full non-blocking crossbar with 512 GB/s sustainable bandwidth. Based on these features and additionally equipped with hardware support for multitasking, this memory subsystem is the reason for the very low de-gradation of the machine in a parallel workload. Even a fullblown node with 32 processors has only a few percent degradation in typical tests [2], much less than all competing systems.

### The SX-4 at HWW

The configuration installed at HWW consists of one node, fully equipped with 32 processors. The actual memory size is 8 GB of SSRAM and an extended memory (XMU) of 16 GB DRAM. This is half of the maximum memory configuration of one node.

The I/O-configuration includes 350 GB of IPI-3 disks arranged in 5 RAID5s and the infrastructure for a later attachment of SCSI-2 disks. The network-I/O is realized via 16 HiPPI channels, some of them attached to gateways to FDDI and ATM.

## **System Software and Programming Models for the NEC SX-4**

The SX-4 runs under Super-UX, NEC's flavour of UNIX. Super-UX is a true 64-bit multiprocessor-UNIX with parallel kernel threads.

Additional features are:

- Resource management and allocation
- Single-system image in multinode systems
- Distributed development environment, mainly crosscompilers and tools for workstations
- Support for file and memory sizes beyond 2 GB.

### **Resource Blocks**

Parallel vector machines (PVP) are on the market since many years, in fact since the introduction of the Cray X-MP series. If however a closer look at the usage model of most of these machines is taken, it is obvious that the parallelism is only used to enhance the throughput capacity, but not for speeding up single applications. Benchmarks on dedicated machines are nearly the only exception from this observation.

There are mainly two reasons:

- The only method of scheduling in these machines was timesharing. Therefore the overhead of a parallel job on a non-dedicated machine was very high, because a job compiled for parallel operation dynamically allocated and lost CPUs depending on the overall load of the machine.
- The relation of memory size [GB] to peak-performance [GFLOP/s] was at least on some of the machines big enough (e.g. Cray C94D at RUS 2:1) to hold several jobs in memory, giving a better overall throughput and hardware usage in job parallel operation. Compared to this, the SX-4 at HWW has a relation of memory size to peak-performance of 1:8 if only the main memory is considered and of 3:8 if the XMU is included. More would have been unaffordable.

As a consequence of the growing gap between peak-performance and affordable memory size, NEC adapted techniques from the MPP world to the SX-4 to support the use of the machine for parallel applications: Resource blocks.

The usage of resource blocks is very similar to space-sharing on MPPs. The user accesses a system-administrator defined block with a fixed guaranteed number of CPUs and a fixed part of the memory for the duration of his job. The user himself is responsible for load-balancing and other methods of managing the allocated resources, thus keeping the operating system overhead at a minimum level. As a consequence, the allocated resources are accounted for the respective user during the allocation, independently of their usage.

### **Vectorisation and Shared Memory Parallelisation**

The SX-4 is binary compatible with its predecessors and is able to interpret directives for vectorisation and parallelisation. However, for copyright reasons, the syntax is slightly different to those of the Cray vector machines, although the functionality is the same. Scripts for an automatic exchange are available. This makes porting in these programming models very easy.

## **Message Passing**

There is no doubt, that message passing is the hardest way to parallelize existing codes, although the paradigm itself is very simple. However message passing is the smallest common denominator between all parallel platforms, including clusters of workstations or PCs. Therefore it is often the only attractive way for the designers of commercial co-des to parallelize them for message passing.

In contrary to many other shared memory parallel systems, the SX-4 has the basic properties to successfully run message passing codes:

- Resource blocks as operational model
- Availability of MPI, PVM and PARMACS
- High message passing bandwidth: 4.2 GB/s measured in MPI
- Acceptable latency: 38  $\times 10^{-5}$  s measured in MPI.

As soon as MPI-2 with it's one sided message passing will be available, things will be even more attractive, as the get/put operations between two processes involve only a copy, which is very fast on shared memory machines, and not the introduction of buffers as in standard message passing.

The use of the SX-4 at HWW for message passing will be encouraged especially for codes, which do not scale to hundreds of nodes and which can make use of vectorisation on the node. A very important example in this context is the LS-DYNA3D code from LSTC which is used for crashworthiness-simulations.

Tests have shown, that a code, parallelised with message passing is faster on most SMP machines than the same code parallelised with shared-memory constructs. The reason for this is, that the number of necessary synchronisations decreases with the coarseness of the parallelism.

## **Compilers and Tools**

Compilers are available for all programming languages which are relevant in the field of science and engineering: FORTRAN (77 and 90), C, C++. Furthermore High-Performance Fortran (HPF) will be available (December 1996) to guarantee portability also for distributed memory applications written in this language.

NEC has it's own set of tools for performance analysis, tuning and parallel debugging (GUIDE environment). Portable tools like BBN's TotalView parallel debugger and the VAMPIR performance analysis tool will be available soon (December 1996) and will be supplied by RUS/HLRS on all HWW systems for academic use.

## **Performance**

The following results compare the NEC SX-4 with other machines at HWW, its predecessors and as a reference with a high-end workstation and a PC. All values are reported relative to the SX-4 (1.0).

## **RUS Loops**

The processor-performance is measured with some simple loops [3], of which the following table gives a selection. The values for the vector machines are asymptotic values and the values for the cache-based machines are only reported for problem sizes fitting into the cache. Thus the reported values are the best values achievable with the re-spective architecture, but as a consequence the problem sizes are different in this table.

	NEC SX-4	CRAY T90	CRAY J90	IBM RS/6000 590H	SNI Scenic Celsius-1
<b>Frequency (MHz) Peak Performance</b>	125.00 2000.00	450.00 1800.00	100.00 200.00	66.66 266.00	200.00 200.00
$a(i)=b(i)+c(i)$	1.00	0.95	0.08	0.13	0.10
$a(i)=\alpha*b(i)$	1.00	0.74	0.09	0.11	0.08
$a(i)=b(i)+\alpha*c(i)$	1.00	1.01	0.08	0.14	0.08
$a(i)=b(i)*c(i)+d(i)*e(i)$	1.00	0.86	0.08	0.10	0.06
$a(\text{index}(i))=b(i), \text{perm}$	1.00	1.96	0.20	0.15	0.11
$a(\text{index}(i))=a(\text{index}(i))+b(i), \text{noperm}$	1.00	0.88	0.10	0.11	0.12

When comparing vector and cache architectures on the base of this table, one has to take into account, that vector architectures typically get their best performance on large vector lengths, whereas cache based architectures have a good performance only for small vector lengths fitting into the cache.

### Performance in real Applications

As examples, the performance results in two real applications are presented for the NEC SX-4 in relation to the NEC SX-3 at DLR (which per processor has a much higher performance than the SX-4, but support only up to 4 processors), the Cray T93 at HWW and the Cray C94D at Stuttgart University.

### Crash-30 Test with LS-DYNA3D

The testcase used was provided by Audi within the framework of a VDI-workshop. It is a small example with about 25 000 elements. The table contains only single processor results for each machine.

	NEC SX-4	NEC SX-3	CRAY T90	CRAY C94D
<b>Relative Performance</b>	1.0	1.04	1.04	0.29

### LAMTUR

This code was developed at the Institute for Aerodynamics and Gasdynamics of Stuttgart University. It simulates laminar-turbulent transition problems. The table contains single processor results for all machines and parallel results for the SX-4, which were achieved on 9 CPUs using shared-memory parallelisation.

	NEC SX-4/1	NEC SX-4/9	NEC SX-3	CRAY T90	CRAY C94D
<b>Relative Performance</b>	1.0	7.99	1.67	0.81	0.33

### Access

There are three ways of accessing the NEC SX-4 at HWW:

- **For academic users:**
  - Allocation grants after scientific peer review
  - Accounting without review
- **For industrial users:**
  - Commercial marketing through debis Systemhaus.

Detailed informations are available on [/HLRS](#)

## References

[1] J. Dongarra, E. Strohmaier, The 1996 Top 500 Report

[2] A. Geiger, U. Küster, The NEC SX-4 at HWW - First Experiences in an Engineering Environment, in [7]

[3] A. Geiger, U. Küster, Ausschreibung Höchstleistungsrechner Stuttgart - Technisches Konzept und Auswahl, Internal Report, RUS Stuttgart, 1995

[4] A. Geiger, Parallelrechner - Architektur und Anwendung, RUS-19, ISSN 0941-4665, RUS Stuttgart, 1994

[5] A. Geiger, N. Kroll, Application Software for Supercomputers, SUPERCOMPUTER 60/61, Vol. XI, June 1995

[6] L. Nilsson, K. Schweizerhof, B. Waincott, J. O. Hallquist, Crashworthiness and Metal Forming: Simulations on Massively Parallel Computers using the Explicit Finite Element Program LS-DYNA3D, in Speedup Vol. 9, Number 2, December 1995

[7] H. Meuer (Editor), Supercomputer '96, K. G. Saur Verlag, 1996.

Dr. Alfred Geiger, NA-5719

E-Mail: [geiger@rus.uni-stuttgart.de](mailto:geiger@rus.uni-stuttgart.de)

Uwe Küster, NA-5984

E-Mail: [kuester@rus.uni-stuttgart.de](mailto:kuester@rus.uni-stuttgart.de)