UDC 681.32

Hardware of VX/VPP300/VPP700 Series of Vector-Parallel Supercomputer Systems

●Nobuo Uchida

(Manuscript received April 28, 1997)

This paper introduces the hardware of Fujitsu's VX/VPP300/VPP700 series of vector-parallel supercomputer systems for high-speed scientific calculations. The series can be configured with up to 512 processing elements (PEs) to achieve a performance of from 2.2 GFLOPS to 1.1 TFLOPS (the maximum performance of each PE is 2.2 GFLOPS). By using advanced CMOS technology, the series has been greatly improved in terms of cost-performance and physical characteristics such as power consumption and installation space. The series supports standard interfaces such as SCSI, ATM, FDDI, and HIPPI, making them very open machines.

1. Introduction

Recent, rapid advancements in computational sciences have made supercomputers indispensable not only in higher education and academic research, but also in industry. Also, there is a continuing demand for higher performance, larger storages, and higher input/output processing speeds.

Moreover, the processing capabilities of engineering workstations has recently improved, and the long economic recession is making it more important to obtain good returns on equipment investment. This situation makes the high cost-effectiveness, high-processing capability, superior installation conditions, and easy operation of supercomputers very attractive features.

Moreover, because supercomputers are in use by many companies, there is a strong call for an open architecture that can be linked with engineering workstations and for the use of de facto standard interfaces.

In response, Fujitsu used the latest technology to develop the high-performance, cost-effective, easy-to-use VX, VPP300, and VPP700 series of supercomputers. The VX series can be used as a high-performance calculation server that features high cost-effectiveness, easy usage, and relaxed installation requirements (installation is possible in an ordinary office). The VPP300 series can be used as a highly cost-effective, high-performance (up to 32.5 GFLOPS) central machine. The VPP700 series can be used as a high-end machine for large-scale science and technology calculations (up to 1.1 TFLOPS). Also, in February 1997, Fujitsu released the VX-E, VPP300E, and VPP700E series, which feature a single PE performance of 2.4 GFLOPS.

2. Aim of development

2.1 High performance by vector-parallel architecture

In the conventional parallel processing system, up to several tens of high-performance vector processors have been tightly connected through shared memory (shared-memory-connected parallel system). Alternatively, up to several thousand comparatively lower performance general-purpose scalar processors have been connected (massively parallel processor [MPP] system). In a shared-memory-connected parallel system, there are restrictions on the operation speed of each processor and on the hardware technology for connecting large numbers of processors requiring a large storage throughput. On the other hand, in an MPP system, the performance of the scalar processors depends on the cache memory of the main storage. Therefore, the wide-area-access data that is frequently found in large-scale programs cannot be processed efficiently. Moreover, it is not easy to construct a conventional multi-processor, high-speed network because of the high communication overhead.

To resolve the above problems, the VX, VPP300, and VPP700 series (**Fig. 1**) use a vector-parallel architecture in which several hundred high-performance vector processors are connected to a newly designed crossbar network. This vector-parallel architecture provides high performance by combining the following three parallel processing technologies (**Fig. 2**):

Fig. 1— VPP300 series.

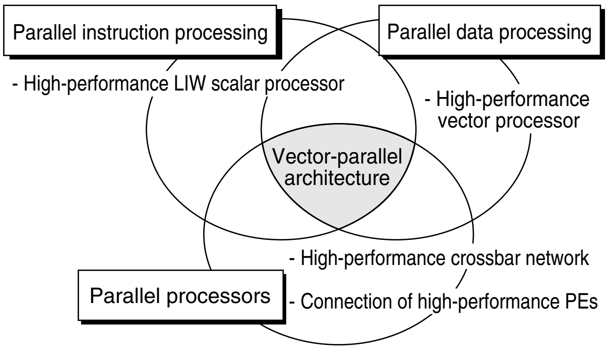

Parallel instruction processing
Parallel data processing
- High-performance LIW scalar processor
- High-performance vector processor
Vector-parallel architecture
- High-performance crossbar network
Parallel processors
- Connection of high-performance PEs

Fig. 2— Vector-parallel architecture.

1) Parallel data processing by vector processing

Each processing element (PE) of the supercomputer performs vector processing so that many operations can be executed in time-parallel mode. Each PE can perform vector processing at a peak speed of 2.2 GFLOPS.

2) Parallel instruction processing using long instruction words (LIWs)

The PE scalar unit uses the LIW-type reduced instruction set computer (RISC) architecture for parallel processing of the instruction level. In one scalar unit, up to three instructions can be executed concurrently. This provides high-speed scalar processing of 428 million operations per second (MOPS).

3) Parallel processing using PEs

To improve the system performance, the distributed-storage-type parallel processing system is used. In this system, PEs with the above characteristics are connected through a crossbar. A VPP700 with the maximum system configuration achieves a system performance of up to 1.1 TFLOPS.

2.2 Large-capacity SDRAM main storage

Supercomputer main storage must have a high throughput; therefore, static RAM (SRAM) has been used because of its high-speed access. However, because the scale of simulations in science and technology calculations and therefore the amount of main storage required to perform them are increasing, SRAMs are looking less attractive because they cannot be highly integrated. Dynamic RAM (DRAM), on the other hand, can be highly integrated, but it has a lower access speed. Thus, it is difficult to achieve a high-throughput large-size storage.

VX, VPP300, and VPP700 use synchronous DRAM (SDRAM) to solve this problem. SDRAM has the same integration density as that of DRAM, and has a higher access speed because it employs clock synchronization. Using SDRAM, one PE can have a main storage of two gigabytes.

2.3 High-speed processing using parallel input/output processing

In the VX, VPP300, and VPP700, two high-speed input/output buses are mounted between a PE and a channel. These buses maximize the input/output performance per PE. Moreover, two or more IOPEs are used for parallel input/output processing. The operation performance can be improved scalably by adding new PEs. Because two or more IOPEs are mounted, the input/output performance is also improved scalably according to the operation performance.

2.4 High cost effectiveness using the latest complementary metal-oxide semiconductor (CMOS) technology

The CMOS used in workstations and PCs have been improved because of important developments in semiconductor process technologies. For example, the signal propagation delay, which was a weak point in the conventional technology, has been reduced; also the integration density has been increased, so products can be built with fewer parts and at lower cost.

Because of the above conditions, the VX, VPP300, and VPP700 series use COMS LSI for superior cost effectiveness. The devices used have a gate length of 0.35 μm, an average gate delay of 70 ps, and contain about 8 million transistors.

2.5 Superior installation conditions

The use of CMOS technology has sharply reduced the power consumption and installation space. **Figure 3** compares the weight, installation space, power consumption, and heat output of a VPP700 series and a previous series (VPP500) of the same performance. Conventional supercomputers have usually been installed in a dedicated computer room because of their installation requirements. However, the entry model of the VX series can be installed in an ordinary office because it takes only 0.6 m^2 of space, consumes only 2.3 kVA, and produces a maximum noise of only 49 dB.

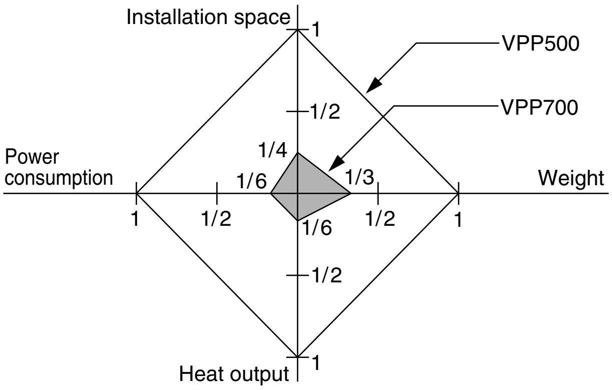


Fig. 3— Comparison between VPP700 and VPP500 of the same performance.

2.6 Open architecture

The VX, VPP300, and VPP700 series support the de facto standard SCSI, WIDE SCSI2, FDDI, HIPPI, and ATM interfaces for connecting open-interface devices.

The vector-parallel architecture function of the operating system was extended and enhanced based on UNIX System V Rel.4.

3. Outline of system

3.1 System configuration

Figure 4 shows an example system configuration of the VX/VPP300/VPP700 series. The VPP series uses the crossbar unit designed for the VPP700 series. The VX and VPP300 series have a crossbar unit built into the PE. **Table 1** shows the main characteristics of these series.

3.2 Installation of system

The VX consists of one or two PEs, input/output adapters, built-in disk drives, and a service processor (SVP). These components are contained in a compact cabinet (690 mm wide × 840 mm deep × 1,400 mm high). A maximum of four PEs can be installed by installing a second cabinet.

The VPP300 consists of up to four PEs, input/output adapters, built-in disk drives, and an SVP. These components are contained in cabinets (820 mm wide × 850 mm deep × 1,800 mm high). The maximum configuration of four PEs consists of four cabinets.

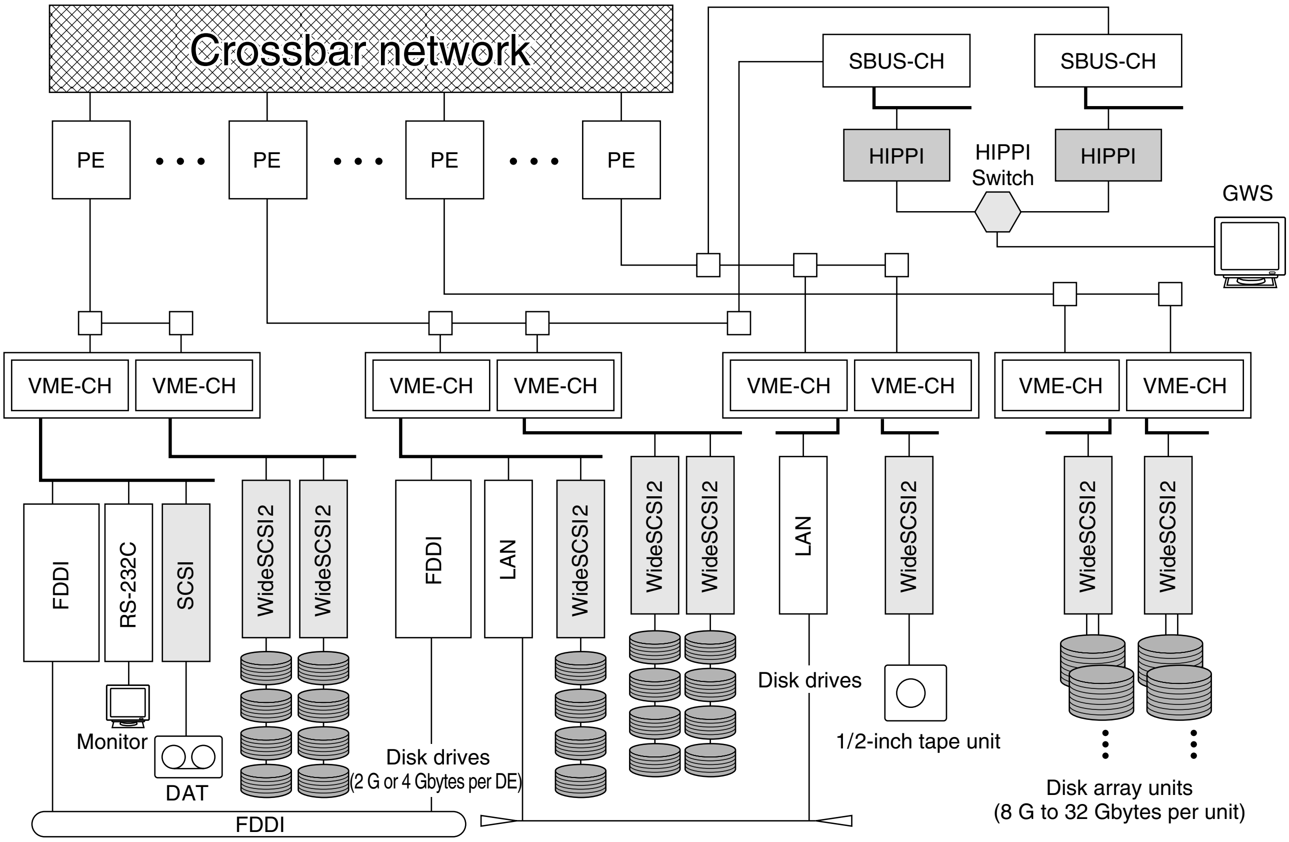


Fig. 4— Example of VX/VPP300/VPP700 system configuration.

Table 1. Characteristics of VX, VPP300, and VPP700 series

		VX series	VPP300 series	VPP700 series
Number of PEs		1 to 4	4 to 16	16 to 512
Peak performance		2.2 G to 8.8 GFLOPS	8.8 G to 35.2 GFLOPS	35.2 G to 1,126 GFLOPS
Capacity of main storage		512 M to 8 Gbytes (512 M or 2 Gbytes per PE)	2 G to 32 Gbytes (512 M or 2 Gbytes per PE)	8 G to 1,024 Gbytes (512 M or 2 Gbytes per PE)
Throughput of main storage		18.2 G to 72.8 Gbytes per second	72.8 G to 291.2 Gbytes per second	291.2 G to 9,318.4 Gbytes per second
Capacity of built-in disks		4 G to 60 Gbytes	4 G to 188 Gbytes	4 G to 1,916 Gbytes
Number of channels	VME	1 to 8	1 to 32	1 to 320
	SBUS	up to 2	up to 8	up to 80
Performance of crossbar		570 megabytes per second × 2/PE	570 megabytes per second × 2/PE	570 megabytes per second × 2/PE

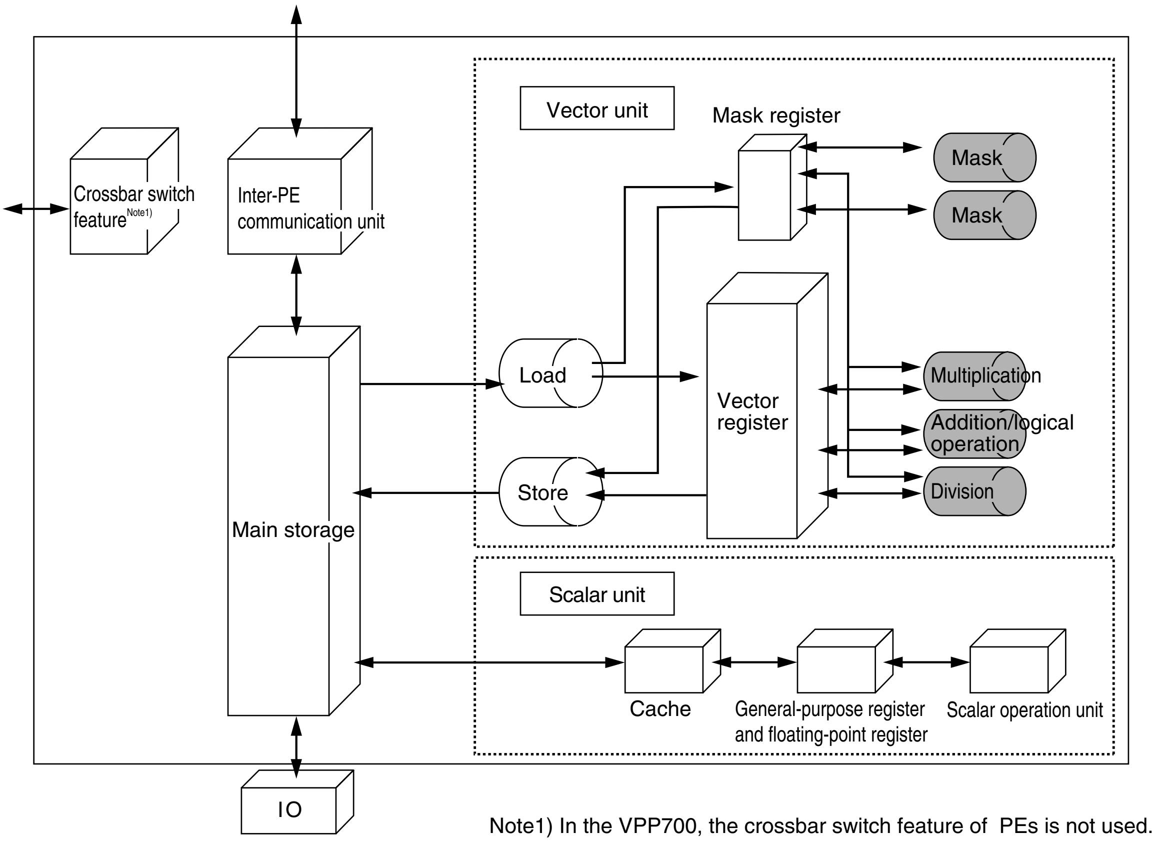
The VPP700 has three types of cabinets: (1) the same cabinet as for the VPP300, (2) a cabinet for up to eight PEs, and (3) a crossbar unit cabinet. (Cabinet (2) has the same dimensions as the VPP300 cabinets.) The VPP700 can have up to 512 PEs.

3.3 Configuration elements

The mainframe of the VX, VPP300, and VPP700 series consists of the following components:

1) Processing elements (PEs)

Figure 5 shows the hardware configuration of a PE. **Table 2** shows the main characteristics of a PE. A PE consists of the following units:


Crossbar switch feature[Note1)]
Inter-PE communication unit
Vector unit
Mask register
Mask
Mask
Load
Vector register
Multiplication
Addition/logical operation
Store
Division
Main storage
Scalar unit
Cache
General-purpose register and floating-point register
Scalar operation unit
IO

Note1) In the VPP700, the crossbar switch feature of PEs is not used.

Fig. 5— Hardware configuration of a PE.

Table 2. Main characteristics of a PE

PE			
PE	Peak performance per PE		2.2 GFLOPS
	Number of vector pipelines		7
	Registers	Vector register	128 Kbytes
		General-purpose register	32 (32 bits)
		Floating-point register	32 (64 bits)
	Cache		64 Kbytes
	Network throughput		570 megabytes per second × 2
Main storage	Storage capacity		512 M or 2 Gbytes
	Storage elements		SDRAM
	Throughput of main storage		18.2 gigabytes per second

i) Scalar unit (SU)

The SU executes scalar instructions and handles interrupts.

ii) Vector unit (VU)

The VU executes vector instructions at high speed. The VU has several instruction execution pipelines and a large-capacity vector register.

iii) Main storage unit (MSU)

The MSU is used to store programs and data. The MSU processes the large amount of storage accesses requested by the VU.

iv) Data transfer unit (DTU)

The DTU processes data communications between PEs through the crossbar network and synchronizes the data transfer.

2) Input/output processing element (IOPE)

The IOPE consists of controllers and adapters for connecting units i) to iv) above, the input/output-controlling channels, and various input/output devices.

3) Crossbar unit (XB)

The crossbar unit transfers data between PEs using the DTU.

4) Service processor (SVP)

The SVP is a computer system independent of the mainframe.

The SVP controls the power on/off sequence and diagnoses and maintains the system.

4. Outline of hardware

4.1 Scalar unit (SU)

The VX, VPP300, and VPP700 series use 1-chip processors having the long instruction word (LIW) architecture.

4.1.1 LIW architecture

The LIW is an instruction-level parallel processing system. One LIW instruction contains two or more fields for the operations to be executed. Operations are assigned to instruction words by the compiler. Because instructions are executed in series without modification, the hardware amount can be reduced and the processing speed increased. **Figure 6** outlines the LIW operation.

4.1.2 Speed increasing techniques

The main characteristics of the SU are as follows:

1) One to three scalar operations or one vector operation can be assigned to a 64-bit instruction word.

2) Only a relative address of a PC can be assigned as a branch destination address for a conditional branch operation. This increases the speed of branch destination address calculation and branch destination instruction prefetch.

3) The asynchronous execution feature is supported to issue a subsequent instruction without waiting for completion of the preceding asynchronous operation (the preceding asynchronous operation requires at least two cycles). Therefore, the execution sequence of asynchronous operations can be changed provided the data dependence remains unchanged.

4) Instructions for trace scheduling are supported. Trace scheduling is a compiler technique that improves the performance by moving instructions across a no-branch instruction string called a "basic block."

4.2 Vector unit (VU)

The VU supports the vector processing method, which is a typical single instruction multiple data (SIMD) method. (When the SIMD method is used, one instruction executes two or more operations.)

4.2.1 Vector processing method

The VU receives a vector instruction from the SU and processes the vector instruction. Vector data is processed by the pipeline operation unit. There are seven instruction execution pipelines: the addition/logic-operation pipeline, multiplication pipeline, division pipeline, load pipeline, store pipeline, and two mask pipelines. Using these pipelines, two or more vector instructions are executed in parallel. The vector register has a capacity of 128 Kbytes. The mask register has a capacity of 2 Kbytes.

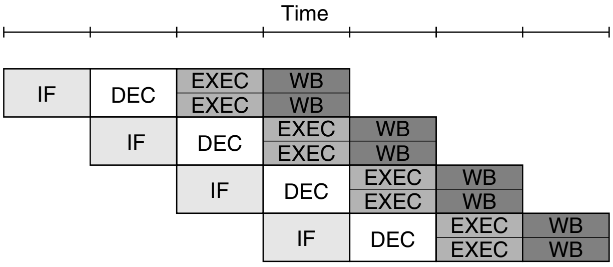
A series of vector processing is executed as follows. First, data in main storage is loaded into the vector register through the load pipeline. The data is then processed using the processing pipeline. The processing results are stored in main storage via the vector register and store pipeline.

4.2.2 RAS function

The vector register has the same ECC feature as the main storage. The ECC feature completely corrects 1-bit errors and completely detects 2-bit errors. It also corrects 1-bit errors in the main storage, load pipeline, store pipeline, and vector register of the VX, VPP300, and VPP700. Because the ECC feature corrects the hardware errors that caused machine checks in the conventional system, it greatly improves the reliability.

4.3 Main storage unit (MSU)

The storage elements of the MSU are SDRAMs with an access time of 60 ns (4 Mbits or 16 Mbits per chip). The SDRAMs are mounted on both sides of each RAM module board. One PE has 64 RAM



IF : Instruction fetch, DEC : Decode, EXEC : Execution, WB : Write back

Fig. 6— LIW operation.

modules. These RAM modules and the MCM containing the CMOS LSI are mounted on a high-density motherboard, which improves the data transfer between the LSI and SDRAMs. **Figure 7** shows a PE motherboard with memory modules mounted.

4.3.1 Capacity of main storage

The above mentioned method of mounting the high-density storage elements increases the storage capacity to up to 2 Gbytes per PE. By adding new PEs, the main storage capacity can be increased scalably. In the VPP700, a storage capacity of up to 1 Tbytes per system is supported.

4.3.2 Techniques for high-speed processing

The operation of the above SDRAM is synchronized with the system clock pulses. Therefore, the number of elements can be reduced to transfer accessed data at high speed. The total main storage is controlled by dividing it into 512 units that can be accessed independently. This provides a high throughput for handling large numbers of access requests from the VU.

4.3.3 RAS functions

The MSU has the following RAS functions for reliability improvement:

1) ECC feature

The MSU has an ECC feature for complete correction of 1-bit errors and complete detection of 2-bit errors. Because the vector register also has an ECC feature (explained above), the code required for data checking is stored, together with data, in both the MSU and vector register.

2) Patrol feature

The patrol feature is used for SDRAM recovery from an intermittent 1-bit error. If a 1-bit error is detected during data reading from main storage, the patrol feature corrects and rewrites the data.

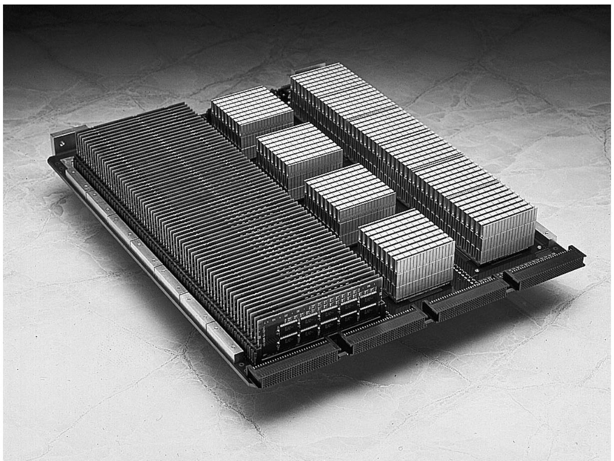


Fig. 7— PE motherboard.

4.4 Data transfer unit (DTU)

A DTU is installed in each PE. The DTUs execute inter-PE communication through the XB independently of operation to provide a high parallel-processing efficiency. Each DTU consists of a data processing unit and an inter-PE synchronization processing unit. The main DTU characteristics are described below.

4.4.1 Data transfer

1) Data sending can be executed in parallel with data reception. The data transfer rate is 570 megabytes per second.

2) Storage access operations during data transfer can be classified into four patterns: continuous pattern, equal-interval pattern, partial-array pattern, and indirect pattern. The inter-PE transfer efficiency can be improved by selecting the appropriate access pattern.

3) The DTU has an address translation feature for translating a transfer-data memory address and a transfer-destination PE address. This provides virtual PE numbers and virtual memory addresses.

4.4.2 Inter-PE synchronization

Each DTU has an inter-PE synchronization feature to synchronize two or more PEs with each other (**Fig. 8**). This feature broadcasts the program progress status information (about each PE) to all PEs, and notifies the system that the PEs are synchronized after it receives confirmation that the PEs have received the broadcast information.

Each PE contains a mask register that indicates which PE group is to be synchronized. Using the mask register, a program can be executed in a group of arbitrary PEs. This enables two or more parallel programs to be executed efficiently.

4.5 Crossbar unit (XB)

The XB is connected to the DTU of each PE to control data communication between PEs. In the VPP700, the XB is mounted in a separate cabinet from the PE cabinet. In the VX and VPP300, the XB is integrated into each PE.

In the VX, VPP300, and VPP700, the XB enables communication between any two PEs at a speed of 570 megabytes per second. In the VPP700, the throughput of the maximum network of 512 PEs is 291 gigabytes per second. The main characteristics of the XB are described below.

1) Unless a remote PE is in a communication, contention rarely occurs because a crossbar switch is used for communications (**Fig. 9**).

2) The distance between PEs is always the same. Therefore, even if processors are arbitrarily selected and grouped, the characteristics of the network do not change. This enables two or more parallel programs to be operated effectively.

4.6 Input/output channels, controllers, and adapters

Two types of channels (direct VME channel and SBUS channel) can be connected to the IOPE. Various types of controllers and adapters having the standard interface can be connected to the bus of each channel (**Fig. 10**).

4.6.1 VME channel

The VME channel controls the connection between a PE and a channel. Five types of controllers for input/output control must be connected to the VME channel. The main VME channel characteristics are described below.

1) Support of a 64-bit block transfer function equivalent to the function of VME64
2) Standard support of the interrupt handler and bus arbitration function
3) Standard support of the VME bus time monitoring feature and bus parity check feature

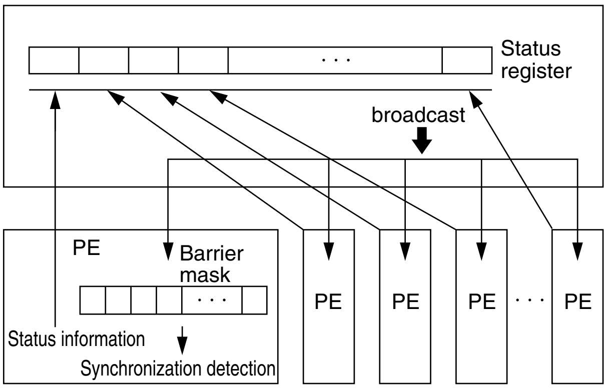
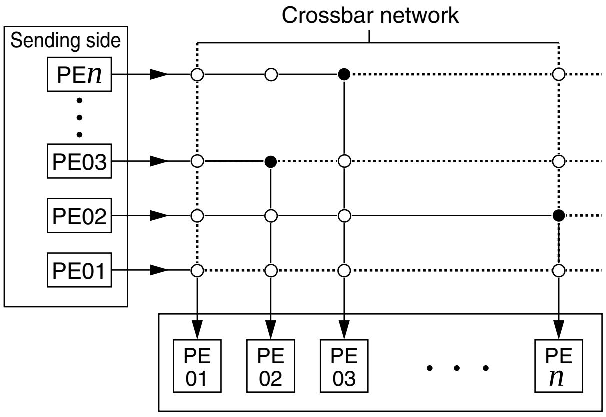


Fig. 8— PE synchronizing feature.



● : Switch-on state (The two intersecting lines are connected with each other.)
○ : Switch-off state (The two intersecting lines are not connected with each other.)

Fig. 9— Crossbar network.

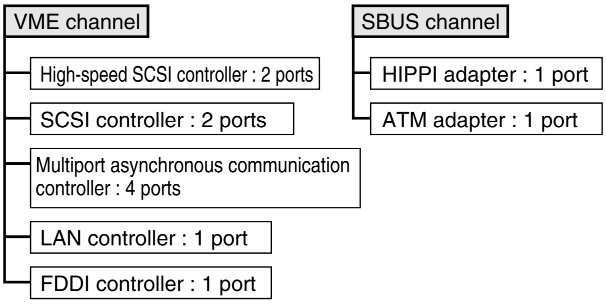

VME channel
High-speed SCSI controller : 2 ports
SCSI controller : 2 ports
Multiport asynchronous communication controller : 4 ports
LAN controller : 1 port
FDDI controller : 1 port
SBUS channel
HIPPI adapter : 1 port
ATM adapter : 1 port

Fig. 10— Channels, controllers, and adapters.

4.6.2 SBUS channel

Two types of adapters for input/output control must be connected to the SBUS channel. The main SBUS characteristics are described below.

1) Support of the 64-bit extended transfer function
2) Support of the parity check function
3) Support of the DMA transfer function (burst transfer of up to 64 bytes)

4.7 Service processor (SVP)

The SVP performs various operation functions through interfaces with the mainframe system hardware and communication lines. The main SVP characteristics are described below.

1) Power control

The SVP turns the power of the system on and off. In the VPP700, hot-system maintenance can be performed in cabinet units by turning off their power.

2) Configuration control

If PEs fail, the system configuration can be changed without stopping the system by disconnecting the faulty PEs with a command from the operating system.

3) Monitoring of mainframe system

The operation status of the mainframe system is monitored. For quick recovery when an error is detected, the error information is logged and passed to the maintenance center through a communication line.

4) Automatic operation function

The automatic operation function is supported for initial program loading of the operating system according to a previously set command and for turning the system power off using a command from the operating system.

5. Conclusion

This paper explained the aim of VX, VPP300, and VPP700 development, explained the system configuration, and outlined the hardware functions.

The high performance, superior cost-effectiveness, and open architecture of these supercomputers were achieved using a vector-parallel architecture, CMOS technology, and standard interfaces.

Supercomputers will be used in more and more fields, and the demand for higher processing speeds will continue. We will therefore strive to increase operation speeds, enhance input/output performance, and improve cost effectiveness.

Nobuo Uchida received the B.E. degree in Electrical Engineering from Waseda University, Tokyo in 1982. He joined Fujitsu Ltd. in 1982, where he is currently engaged in development of supercomputer hardware.