32way Itanium2 Server NX7700/i9510: High-Performance and Highly Reliable Infrastructure for Mission-Critical Enterprise Application

By Toshiteru SHIBUYA,* Shigemi MIKAYAMA,* Tetsuhide SENTA,* Masayuki KIMURA† and Hitoshi TAKAGI*

ABSTRACT NEC NX7700 Series Model i9510 is a high-performance and highly reliable server for mission-critical enterprise application. NX7700/i9510 provides suitable infrastructure for NEC's solution under the name of "Dynamic Collaboration."

KEYWORDS **Servers, Itanium, HP-UX, Windows, Linux, VALUMO, Autonomous operation, SystemGlobe GlobalMaster, Virtualization, High reliability, High performance, RAS**

1. INTRODUCTION

NEC Server NX7700 Series Model i9510 (**Photo 1**) provides the basis of the infrastructure that enables open, secure and non-stop business which NEC's Dynamic Collaboration advocates. NX7700/i9510 employs the leading edge de-facto standard CPU Intel Itanium2 and three industry standard OSes (HP-UX 11i V2.0, Microsoft Windows Server 2003, Linux) for broader solution availability. In addition to the standard component, NEC developed its original chip set that enables higher reliability and availability, at a level never before provided. By the cooperation with NEC's platform technology called "VALUMO," NX7700/i9510 can provide an autonomous operational environment. This paper summarizes the feature of NX7700/i9510 and benefits for users as the infrastructure for the Dynamic Collaboration.

2. HARDWARE ARCHITECTURE OF NX7700/i9510

NX7700/i9510 is an Intel Itanium2 based 32CPU server suitable for large data and transaction handling such as data center and mission-critical applications where high reliability and availability are most critical, as well as high performance and scalability. The system has a capacity of 32 CPUs, 512GB memory, 112PCI-X slots.

2.1 NEC's Original Chipset

To achieve high performance, NX7700/i9510 employs Intel's high-end enterprise CPU Itanium2 processor. Also, NEC has developed its original chip set to expand scalability up to 32CPU and 112PCI-X slots. NX7700/i9510 has contemporary design in modular architecture based on CPU Cell and PCI-X Cell (**Fig. 1**). Each CPU Cell and PCI-X Cell communicate with each other via a crossbar network. The interface between each CPU/PCI-X Cells and the crossbar network has 6.4GB/s per "port" and 12.8GB/s per "CPU Cell" bandwidth, which give the system aggregate bandwidth over 100GB/s. Each CPU Cell can house up to 4CPU and 32 DIMM slots, in other words, it supports up to 64GB memory per CPU Cell with 2GB DIMM (Dual In-line Memory Module). Memory system has total 12.8GB/s bandwidth to feed its data to both CPU bus and I/O systems. The PCI-X Cell has fourteen PCI-X slots, eight slots for 133MHz bus frequency and six slots for 66MHz frequency. The PCI-X

Photo 1 NX7700 Series Model i9510.

*Computers Division
†NEC Computertechno, Ltd.

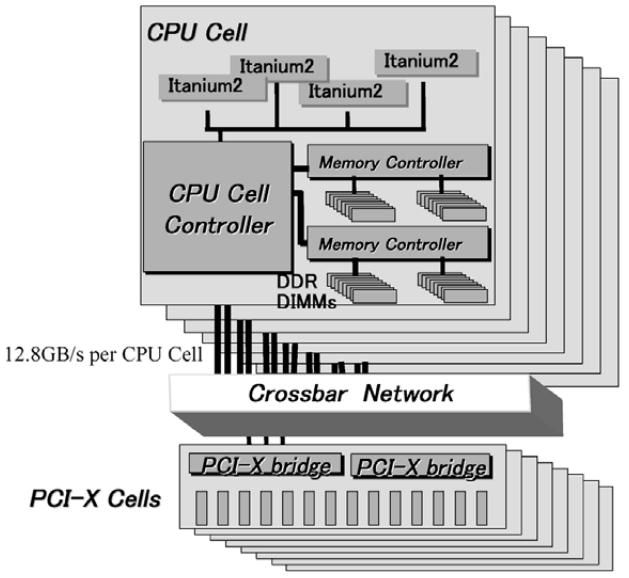

Fig. 1 NX7700/i9510 block diagram.

Cell also houses six 3.5-inch hot-swappable disk drives. The PCI-X Cell supports ordinary standard adaptor cards such as Gigabit Ethernet, FC-AL, and SCSI. Also, the PCI-X Cell supports a Basic-IO card which provides essential IO system including keyboard and mouse, VGA display port, and serial interface.

The NEC original chip set has many unique features to provide data center and enterprise users with TCO (Total Cost of Ownership) reduction. Examples are: high availability chip set design which originated from NEC's experience in developing mainframe computers, system partitioning for flexible operation, hardware virtualization mechanism for the basis of autonomous operations in cooperation with software, and many others. Regarding high reliability and availability design, there are many design consideration acquired from mainframe computer design in developing the chip set such as ECC/parity protection on data paths, data integrity checks and command retry functions.

The Service Processor plays an important role in the platform RAS (Reliability, Availability, Serviceability) features. The firmware of the Service Processor implements all the functions to monitor, analyse, and take actions to all the events in the system. The Service Processor is connected to all CPU and PCI-X Cells, and the crossbar network modules via dedicated interface called DGI (Diagnostic Interface) to do numerous management tasks such as hardware initialization, hardware configuration, hardware diagnostics, and power control management. The Service Processor also performs error monitoring, logging, and reporting.

The N+1 power and cooling subsystem provides complete redundancy in case of failures. In addition to standard power supply modules or fans ("N" units), a redundant modules or fan ("+1" unit) ensures that system operation would not be affected when unexpected single failure in modules or fans. The Service Processors, clock modules, crossbar network modules and path between CPU Cell and PCI-X Cell can have redundant configuration.

2.2 Flexible Hardware Partitioning

In order to execute business in today's network connected society and as a result of rapidly changing business style, companies must revise their business strategy frequently. The outcome of the revision often shows the need for changes in IT infrastructure. The flexibilities in configuration of NX7700/i9510 provides users ease of set-up and reduction of the time when change to the new configuration is required.

One of these features is the hardware partitioning. NX7700/i9510 can divide itself into as many as eight "sub-servers," or partitions. Each of the sub-servers is a collection of the "CPU Cell" in which up to 4CPU and 64GB memory can reside (**Fig. 2**). Also, each sub-server runs its operating systems independently of other sub-servers within a cabinet (**Fig. 3**). This partition function provides server consolidation where a larger server replaces several smaller servers installed for many years and cost for management of these smaller servers becomes a major factor in IT operation.

The partitioning is done by hardware and all partitions are isolated from each other, therefore the transaction within each partition will be kept secret. Even if one of sub-servers goes down because of hardware or software failure, the remaining sub-servers in the system will stay unaffected by that failure. In this case, if a CPU Cell has a hardware failure, the failed parts can be swapped out while the entire system stays on-line. On another occasion, the numbers of CPU Cells in sub-servers can be adjusted when one of sub-servers is over-loaded. For those partitioning management, SystemGlobe GlobalMaster, a member of NEC platform middleware family called VALUMOware, gives users partition operation with ease of use graphical user interface as well as autonomous system operation described later in detail.

Although security among each sub-server is maintained, there is a way for the sub-servers to communicate with each other by accessing shared local memory space (**Fig. 4**). This feature reduces

communication overhead among the sub-servers within the system. Access to the shared local memory space can only be done by privileged mode software; therefore security at user application level is maintained.

2.3 Hardware Virtualization and Autonomous Operating Environment

Another characteristic of the partitioning in NX7700/i9510 is virtualization of hardware resources. Each of the sub-servers as well as the entire NX7700/i9510 is decoupled by the crossbar network into a part of CPU/Memory (CPU Cell) and an I/O (PCI-X Cell) part. The inter-relation between CPU/Memory and I/O can be programmable. This feature makes it possible to configure any sub-server with any population of CPU/Memory and I/O regardless of the position of CPU/Memory and I/O.

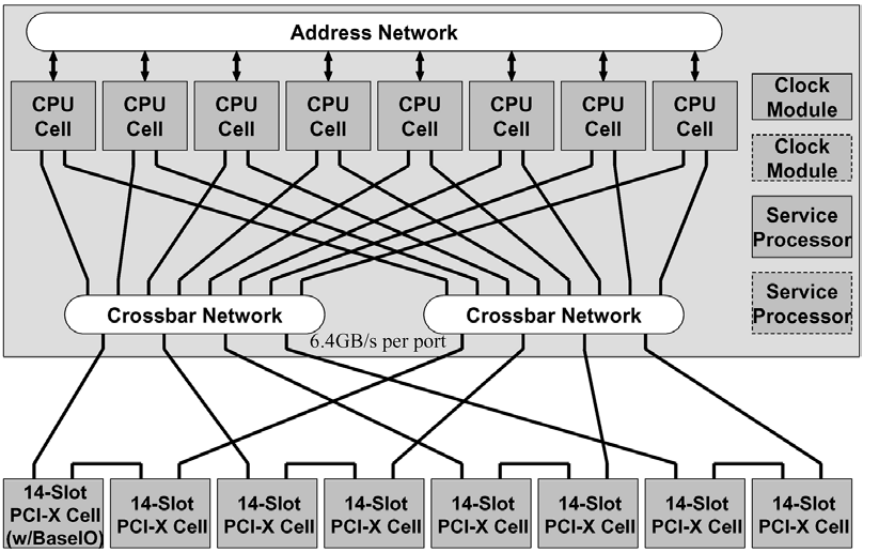

Fig. 2 Cell architecture (CPU Cells and PCI-X Cells).

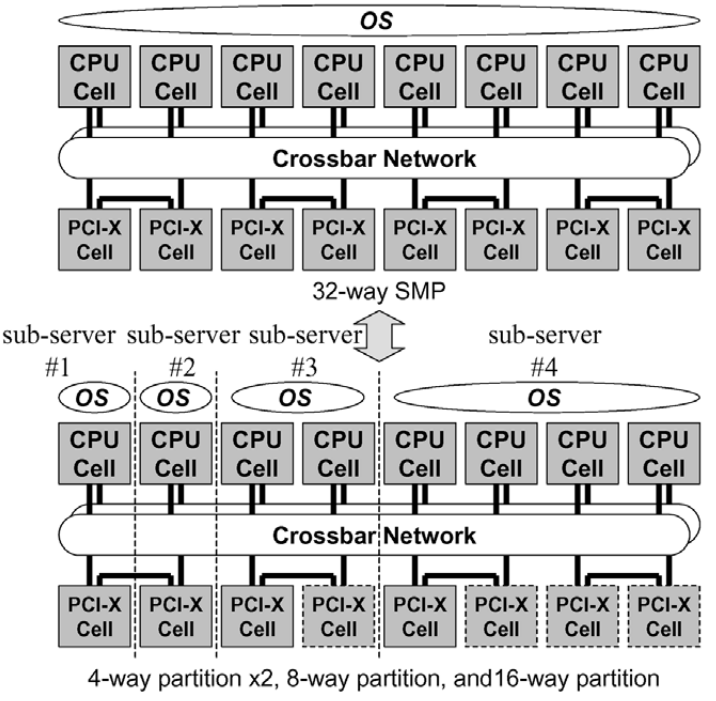

Fig. 3 Partitioning and sub-servers.

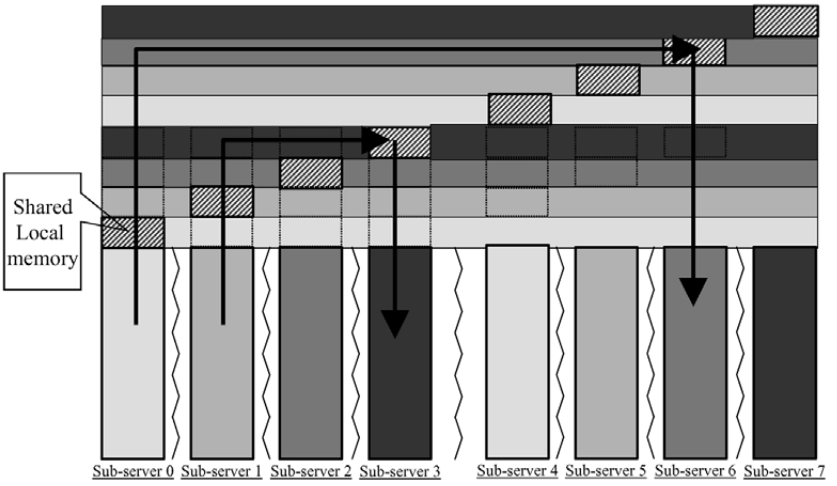

Fig. 4 Shared local memory.

The benefit of the virtualization is enhanced in cooperation with VALUMOware. Under the control of SystemGlobe GlobalMaster, policy-based autonomic operation is possible by itself and interactions with cluster management software in SystemGlobe.

For example, "autonomic restoration" minimizes the system down time, where the SystemGlobe GlobalMaster de-configures the defective CPU Cell and in turn configures the spare CPU Cell into the system (**Fig. 5**). Another example is "autonomic adjustment" that prevents the system from being disrupted by overloading, where SystemGlobe GlobalMaster adds the spare CPU Cell to the overload sub-server (**Fig. 6**).

In Figure 5, an error occurred in the left-most CPU Cell. Upon the Service Processor's notification, SystemGlobe GlobalMaster stops the defective sub-server. SystemGlobe GlobalMaster takes action according to the policy to restore the system to normal status. First, GlobalMaster orders Service Processor to remove the failed CPU Cell from the current configuration, and changes the configuration to connect the left-most PCI-X Cell and the spare CPU Cell. After the Service Processor finishes the change configuration, SystemGlobe GlobalMaster then directs the new sub-server to boot up OS. The failed CPU Cell can be swapped off from the system without bringing the entire system down.

Figure 6 shows an "autonomic adjustment" scenario. The right side sub-server is assumed to be over-loaded. SystemGlobe GlobalMaster is notified by performance and workload manager software on the situation. SystemGlobe GlobalMaster then directs the sub-server on the right to go down. After the sub-server goes down, SystemGlobe GlobalMaster directs the Service Processor to include the right-most CPU Cell to the sub-server on the right. Then SystemGlobe GlobalMaster directs the sub-server to boot the OS up.

Figure 7 illustrates the process of the failure discovery, re-configuration, and OS boot up for the new configuration.

3. NX7700/i9510 IN THE VISION OF DYNAMIC COLLABORATION

As described above, NX7700/i9510 fulfils the requirement for Dynamic Collaboration in the following aspects:

(1) An Open Platform

1) It employs Intel Itanium2 Processor which is industry de-facto standard. The development and manufacturing of CPU is a huge investment, and the companies which can do CPU business are very limited because they need comparable returns, namely larger CPU sales, on their investment. The Intel CPU has a far larger shipment volume than any other CPU, and therefore Intel is most likely to continue its supply of CPU like Itanium2. This assurance of CPU supply benefits end-users for their investment because the user will not re-invest the software development and maintenance. Also, Intel is neutral to all server vendors and software vendors. This assures that Itanium2 will be neutral and open CPU. Therefore, there will be a larger number of vendors who sell Itanium2 based servers than of traditional RISC servers. Also, there are larger numbers of

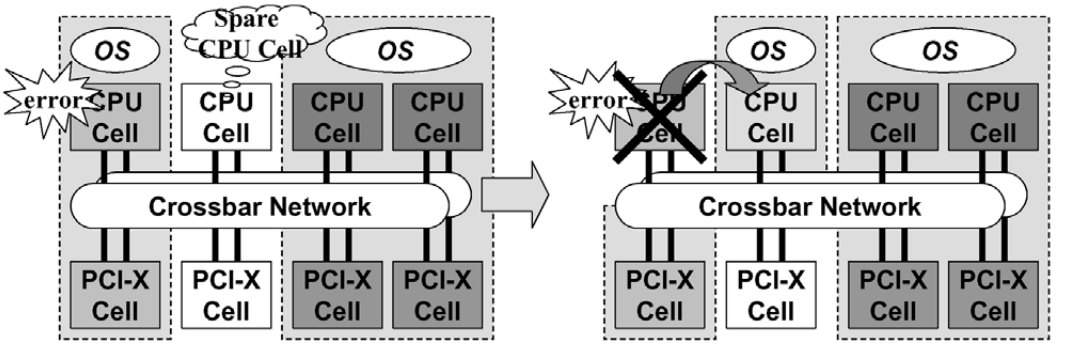

Fig. 5 An "autonomic restoration" scenario.

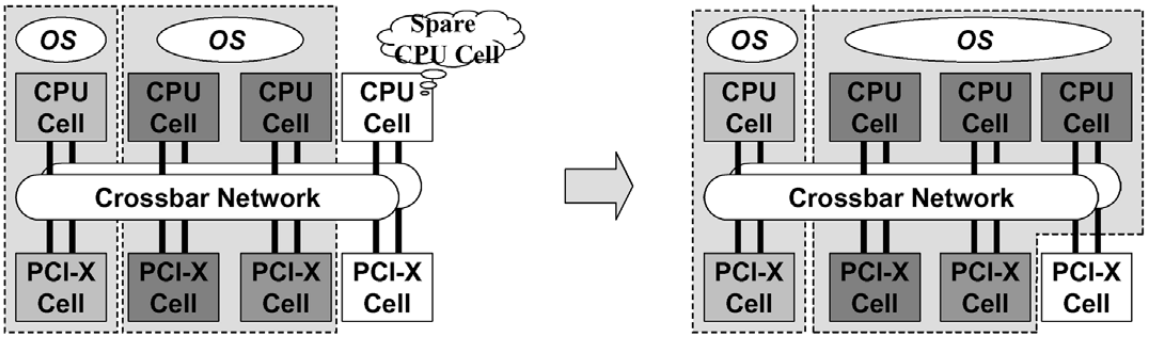

Fig. 6 An "autonomic adjustment" scenario.

software products for Itanium2 than traditional RISC servers.

2) It complies with many industry standards such as DIG64 (Developers Interface Guide for IA64), ACPI (Advanced Configuration and Power Interface), IPMI (Intelligent Platform Management Interface) and PCI-X. These standard conformances assure NX77000/i9510's inter-operability with other platforms.

3) It supports three major operating systems such as Hewlett-Packard HP-UX11i v2.0 for mission critical application, Microsoft Windows Server2003 for platform continuity from desk tops to servers, and Linux for from manufacturing aid solutions to research and development. NEC has been extensively conducting assurance tests for NX7700/i9510 to HP-UX environment in cooperation with Hewlett-Packard. The certification process assures that virtually all HP-UX applications on Itanium servers can be run on the NX7700/i9510 without modifications. Windows Server2003 is supported by the sister Microsoft machine of NX7700/i9510, called Express5800/1320Xd whose hardware is identical to NX7700/i9510. Express5800/1320Xd is certified as Windows Server and the certification assures compatibility of applications available to Windows on Itanium Servers. Moreover, Linux is supported by another sister machine of NX7700/i9510, called TX/i9510. The Linux on the TX7/i9510 is enhanced by NEC to accommodate 32CPU scalability and large IO configuration, both of which are the weak points of the current version of Linux. There are many major technical and scientific applications validated on TX7/i9510 such as MSC.NASTRAN, LS Dyna, Gaussian, STAR-CD, and Fluent. The availability of three OSes and associated application gives users a broader range of selection of their applications.

(2) Business Continuity

1) NEC's original chip set includes many mainframe computer-originated design considerations and enhances high reliability to the system level such as ECC on all data paths, command retry function on

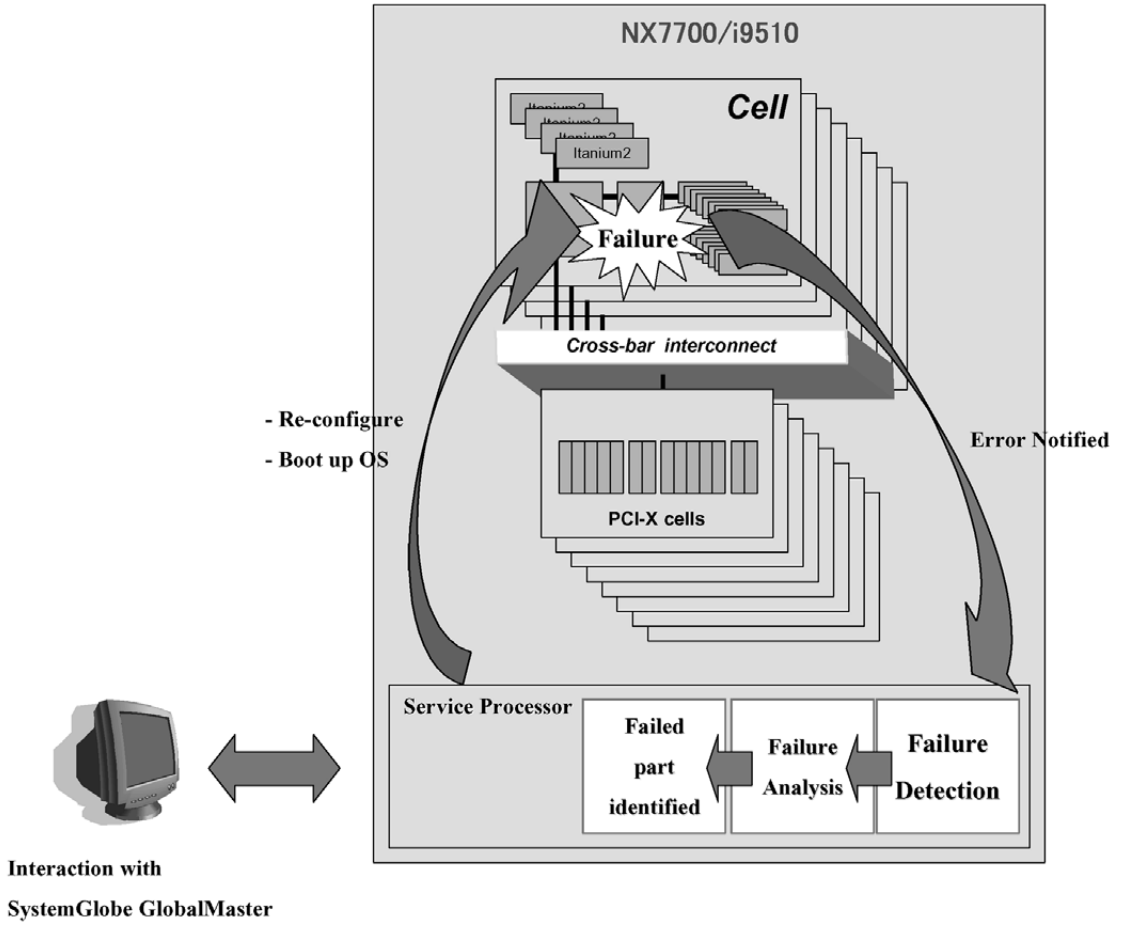

Fig. 7 Process of the failure discovery, re-configuration and OS boot up.

crossbar network data exchanges, memory patrol and duplicated parts.

2) The hardware partitioning provides isolation among sub-servers where no system event including sub-server's system down propagates to the rest of the system, and this gives users great confidence in the system security and higher availability.

3) All the field replaceable parts can be hot swapped after being de-configured from the system. This reduces the planned and/or unplanned down time.

4) Hardware virtualization mechanism provides autonomous system operation in cooperation with VALUMOware and minimizes business disturbance. The virtualization also gives the system autonomous adjustment capability.

4. CONCLUSION

NX7700/i9510 provides a robust and reliable system and therefore provides the one and only infrastructure for the Dynamic Collaboration System. This comes from NEC's original added-value such as original chipset design and cooperation with VALUMOware. These technologies are applied to the lower models NX7700/i9010 and NX7700/i6010, which shares the same cell-based architectures. NEC is striving to enhance this concept of infrastructure design and provide users with higher value.

REFERENCE

[1] T. Senta, et al., "Itanium2 32-way Server System Architecture," *NEC Res. and Develop.*, **44**, 1, pp.8-12, Jan. 2003.

Received March 30, 2004

* * * * * * * * * * * * * * * * *

Toshiteru SHIBUYA received his B.E. and M.E. degrees from Saitama University in 1981. He joined NEC Corporation in 1981, and is now Chief Manager of Computers Division. He is engaged in the development of high-end open server products.

Shigemi MIKAYAMA received his B.E. degree from University of Tokyo in 1981. He joined NEC Corporation in 1981, and is now Chief Manager of Computers Division.

Mr. Mikayama is a member of the Information Processing Society of Japan (IPSJ).

Tetsuhide SENTA received his M.E. degree from the University of Tokyo in 1985. He joined NEC Corporation in 1985, and is now Department Manager of 3rd Engineering Department, Computers Division. He is engaged in the development of open server products.

Masayuki KIMURA is Department Manager at the 1st Computer Engineering Department of NEC Computertechno. He received his ME degree and BE degree from Yokohama National University, and joined NEC in 1985.

Mr. Kimura is a member of the Information Processing Society of Japan (IPSJ).

Hitoshi TAKAGI is Manager of the Product Engineering Department, Computers Division. He received his M.E. degree from Nihon University and joined NEC Corporation in 1983.

Mr. Takagi is a member of the Computer Society of the Institute of Electrical and Electronics Engineers (IEEE).

* * * * * * * * * * * * * * * *