

## Power-aware Dynamic Allocation-control Technology for Maximizing Power Efficiency in a Photonic Disaggregated Computer

*Masashi Kaneko*

### Abstract

A photonic disaggregated computer developed as a computing platform of NTT's Innovative Optical and Wireless Network (IOWN) will consist of a variety of computing devices (energy-conserving central processing units, accelerators, etc.) distributed over a wide area. This article introduces power-aware dynamic allocation-control technology for maximizing power efficiency in a photonic disaggregated computer.

*Keywords: energy conservation, disaggregation, heterogeneous computing*

### 1. Power-related problems in server systems in the IOWN era

Internet Protocol traffic has been growing exponentially with expectations that it will reach 4000 times its current level by 2050. At the same time, the speed of improvements in central processing unit (CPU) performance on servers that process that traffic is slowing down. A trial calculation showed that the power needed for datacenter operation will likewise increase by 4000 times by 2050 given the power efficiency of today's server hardware. Improving the energy conservation and power efficiency of servers is therefore critical for future network operation.

With the Innovative Optical and Wireless Network (IOWN), our aim is to develop a photonic disaggregated computer as a new network-wide computer that can execute high-speed and efficient processing in a network via computing devices connected over the All-Photonics Network (APN), one of the main components of IOWN. A photonic disaggregated computer will introduce new devices and processing methods such as photonics-electronics convergence devices and memory-centric computing. Therefore, it

is expected to achieve a level of power performance surpassing the limits of current server hardware and solve power-related problems into the future.

### 2. Photonic disaggregated computer and power-aware dynamic allocation-control technology

Distributed heterogeneous computing can be achieved through the connection of diverse devices such as CPUs and accelerators (graphics processing units (GPUs), field-programmable gate arrays (FPGAs), etc.) as computing devices that make up a photonic disaggregated computer. When operating virtual network functions (VNFs) as a group of virtual functions making up a network on such a computing platform, a control mechanism is needed to allocate the software components configuring a VNF (in a conventional VNF, this would be virtual machines (VMs)) to each computing device. In a conventional VNF, a VM requiring an accelerator would have to be located in server hardware physically equipped with an accelerator. It is also common in standard VMs to use an accelerator in an exclusive

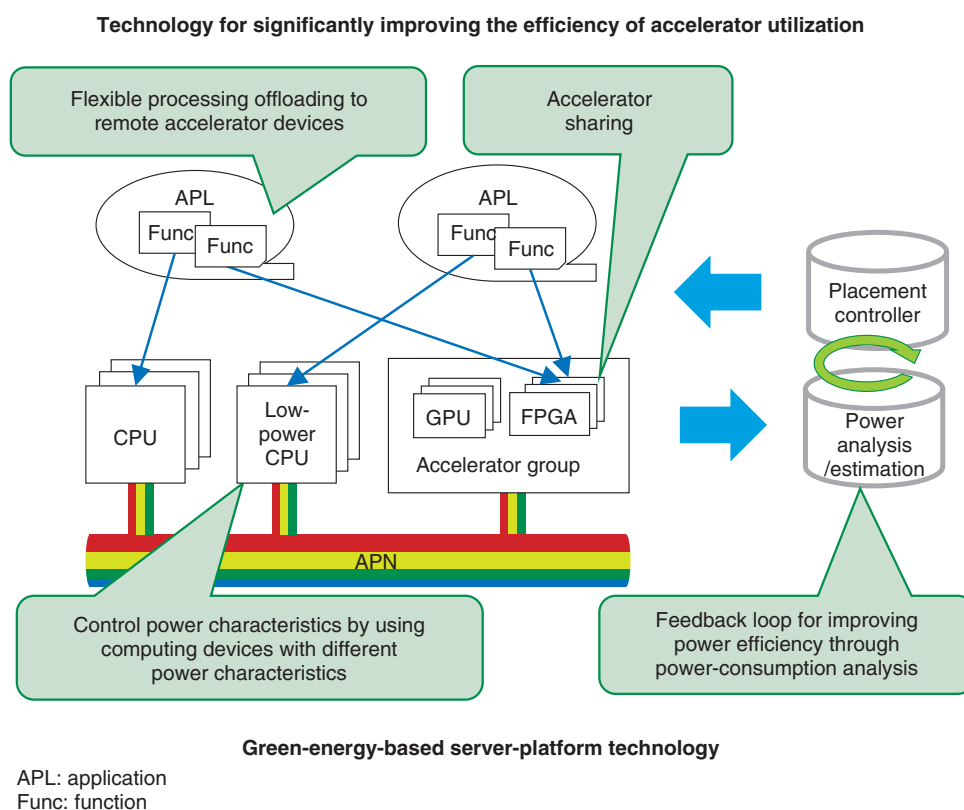


Fig. 1. Power-aware dynamic allocation-control technology.

manner in units of devices (GPU board, FPGA board, etc.). This means that the hardware resources of an accelerator cannot be effectively used on a server deploying an application having a low ratio of processing offloaded to the accelerator. In a photonic disaggregated computer, computing devices, such as CPUs and accelerators, are connected over optical paths, enabling high-speed, low-latency interaction between physically separated devices and an easing of configuration constraints caused by the physical form of a conventional server (e.g., the number of extension boards that can be mounted in a rack-mount server). Assuming that  $N$  is the number of accelerators used by the VNF and  $M$  is the number of VNFs using the accelerator, if the conventional 1-to- $N$  relationship between the VNF and accelerators can be made into a more flexible  $M$ -to- $N$  relationship, it will become possible to allocate the processing of multiple VNFs to a single accelerator, thus maximizing the use of an accelerator's hardware resources.

The power-aware dynamic allocation-control technology being developed at NTT Network Service Systems Laboratories will contribute to a reduction in

power consumption through software control of the server platform (Fig. 1). We are developing two key technologies toward this goal. The first is technology for significantly improving the efficiency of accelerator utilization to facilitate the sharing of accelerator devices and improve the availability rate of accelerators. This is accomplished by enabling the offloading of processing to accelerator devices distributed across the network and, in contrast to the conventional practice of exclusive access, by accepting parallel processing when software components running on multiple CPUs attempt to simultaneously access an accelerator. The second is green-energy-based server-platform technology to control power demand and make maximum use of power obtained from unstable renewable-energy power generation in an environment composed of diverse computing devices such as low-power CPUs and various types of accelerators. This will be achieved by analyzing power consumption in detail in units of devices and software components and reducing power by appropriately placing software components that can minimize power consumption and controlling energy conservation in

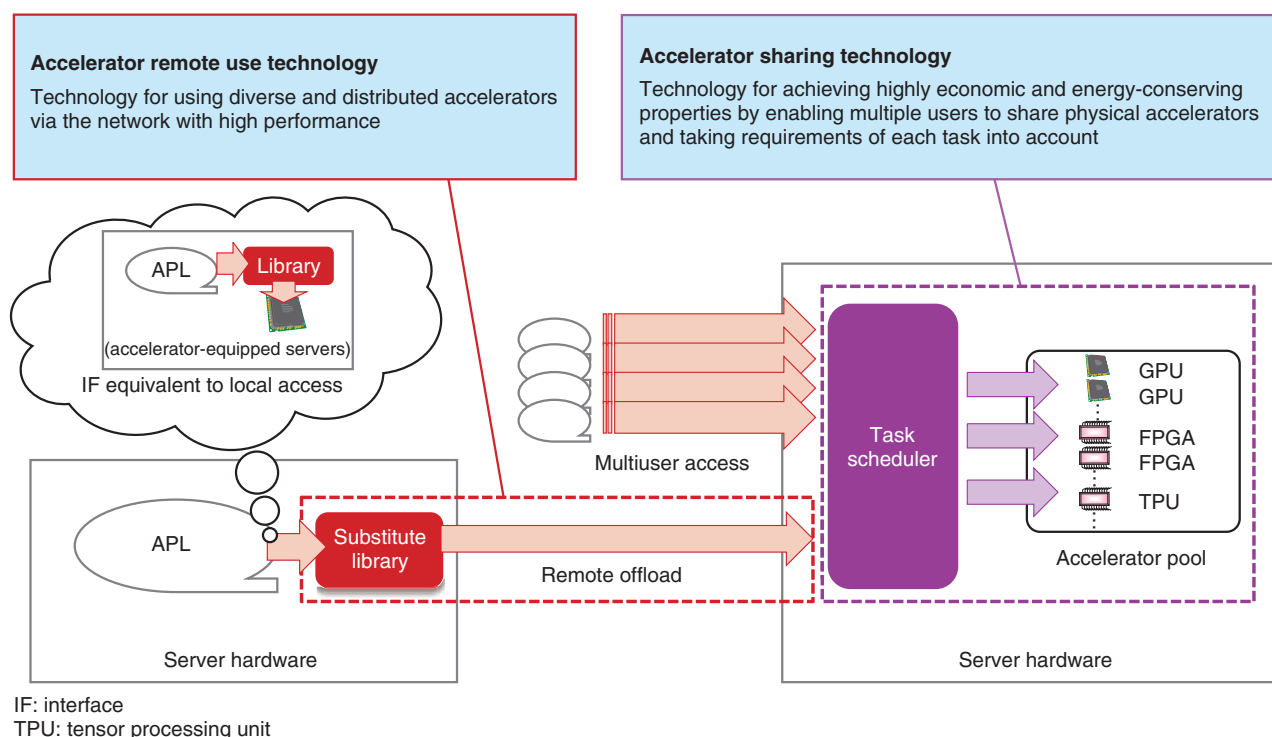


Fig. 2. Technology for significantly improving the efficiency of accelerator utilization.

individual devices. This could be done by adjusting power consumption on the server platform that takes into account the status of renewable-energy power generation in individual regions.

### 3. Technology for significantly improving the efficiency of accelerator utilization

The use of GPUs and FPGAs in addition to CPUs has been increasing in complex computational processing such as data analysis and machine learning. Accelerators, though usually weak in terms of general-purpose processing, can execute specialized processing more than 100 times more efficiently than CPUs. When using accelerators, it is a common practice to offload a part of a program running on a CPU to an accelerator via an application programming interface (API) such as OpenCL. However, such offloading cannot be done if an accelerator in idle state does not exist on the same server hardware as the CPU running the program.

We are developing a technology that significantly improves the efficiency of accelerator utilization. It does this in two ways. First, it enables high-speed, low-latency use of remote accelerators distributed

over the network with location transparency and low overhead just as if they were locally placed without being tied to a conventional physical-connection configuration. Second, it enables multiple users to share these accelerators (Fig. 2). In the offloading of processing to remote accelerators, location transparency can be achieved and the portability of existing applications can be improved by incorporating a substitute library that provides an existing offloading API (e.g., OpenCL API) to the offloading source and by providing an interface equivalent to conventional local access to the application.

### 4. Green-energy-based server-platform technology

Current server hardware has been designed assuming a continuous and stable supply of power, so a datacenter accommodating many servers must be able to provide a large amount of power in a continuous and stable manner. There has been a move to use renewable energy in the power feed to datacenters, but since current servers require a continuous and stable supply of power, as described above, it is difficult to achieve long-term stable operation only on

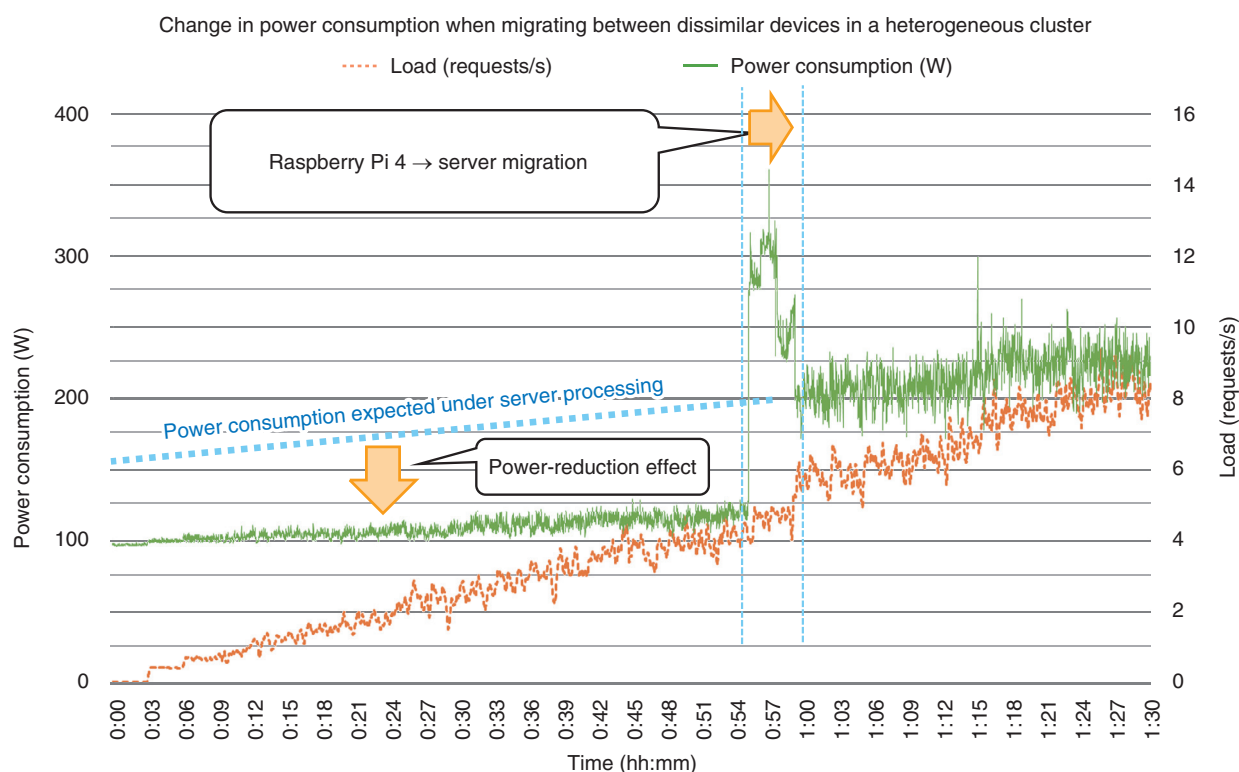


Fig. 3. Example of energy-conservation control in a heterogeneous cluster.

the basis of renewable energy in which the amount of power generated is unstable, as in the case of solar power generation. Green-energy-based server-platform technology controls power in units of computing devices and controls the placement of software components that make up the system by analyzing and estimating the power consumed by servers. This enables energy conservation over the entire software platform and operation control corresponding to the amount of power being supplied.

Standard server hardware in a low-load state still consumes about 60–70% of the power consumed in a high-load state. A system on a chip (SoC) installed in smartphones decreases as much power consumption in a low-load state as possible by controlling operation in accordance with power-supply conditions (battery storage), as in the smartphone “power-saving mode.” This includes measures such as CPU power state control and frequency control tailored to the load and mounting of CPU cores particularly suitable for low-power operation. On a server platform, as well, if power demand can be adjusted by varying performance on the basis of the amount of supplied power, it should be possible to achieve a server plat-

form that can use ever-fluctuating renewable energy in a non-wasteful manner. As an alternative approach, power efficiency could be improved by configuring a heterogeneous cluster of servers that combines multiple types of servers with different levels of power performance and dynamically selecting optimal servers in accordance with operating conditions.

To give an example, power improvements could be achieved at low loads by configuring a heterogeneous server cluster that combines standard server hardware and computing devices that, while having low maximum performance, excel in power performance such as the Raspberry Pi computer, selecting optimal servers in accordance with load conditions, and executing system migration dynamically. The graph in **Fig. 3** shows power consumption when configuring a Raspberry Pi cluster consisting of 15 Raspberry Pi 4 devices and one 1U (unit) rack-mount server and executing software-component migration on the basis of load. From these results, processing by the Raspberry Pi cluster in a low-load state (1U rack-mount server is OFF) can be achieved at a level of power, even lower than that of the server in idle state (about 150 W). Therefore, power efficiency can be improved

across a wide range of loads by combining multiple computing devices having different power characteristics and controlling the placement of software components making up the system on the basis of load.

---

## 5. Future developments

---

The power consumed by information technology services is growing and expected to become an increasingly serious problem. Going forward, we will

continue to investigate power-aware dynamic allocation-control technology to solve this problem by efficiently controlling a photonic disaggregated computer. We will also promote technology development to contribute to the creation of a low-carbon society. To this end, we will pursue the use of renewable energy in combination with our quest for higher levels of power efficiency in addition to solving performance and cost issues.



**Masashi Kaneko**

Senior Research Engineer, Server Network Innovation Project, NTT Network Service Systems Laboratories.

He received an M.E. from the University of Electro-Communications, Tokyo, in 2004. He joined NTT Network Service Systems Laboratories the same year and studied network server platform technologies including web-telecom service convergence, and a sharding method of telecom systems. From 2015 to 2017, he engaged in the development of commercial NFV/software-defined wide area network services at NTT Communications Corporation. He is currently studying photonic disaggregated computers.

---