Latency-Oriented Domain Specific Computing with Emerging Devices

川上, 哲志

（別紙様式 2）

氏　　名　：川上　哲志

論 文 名　：Latency-Oriented Domain Specific Computing with Emerging Devices
（新奇デバイスによるドメイン特化レイテンシ指向計算機に関する研究）

区　　分　：甲

論　文　内　容　の　要　旨

In current industrial sectors that play a central role in society such as energy, manufacturing, and medical care, new value creation including higher efficiency and cost reduction is required. There will be limits in terms of quality and volume on manually devoted efforts to meet these social requirements. Therefore, expectations for IoT (Internet of Things), that is a system for semi-automating a series of cycles such as "visualization of a current situation", "future prediction" and "feedback to the next countermeasures" are increasing. In this era of expectations of IoT, computing technology needs to live up to various demands such as real-time performance, robustness, and safety as well as higher speed and lower power consumption than ever.

On the other hand, processors, which is one of the core technologies of information processing, historically has achieved four times the degree of integration in three or two years and a 20% to 30% operating speed improvement or low power consumption based on Moore's Law and Scaling Rule. However, the shrinking of semiconductor size has been slowing down in recent years, and it is predicted that it will stop around 2020 according to ITRS (International Technology Roadmap for Semiconductors). Indeed, there are just a few companies which are building microscopic CMOS process technology less than 10nm in 2018.

Continuous computer performance improvement is indispensable to meet social expectations, and it is also requested to establish new computing technologies beyond the von-Neumann type computing method using conventional CMOS digital circuits. Research subjects of innovative computing technology to be addressed in the future involve all technical layers of applications, algorithms, architectures, circuits, and devices. However, rather than comprehensively implementing these technologies, it is crucial to select technology from each technology layer properly and to develop the technology by focusing on important applications from the viewpoint of efficiency of research, development and social implementation. In particular, when we focus on the edge computing at IoT, there are essential requirements such as real-time autonomous control and image recognition, and the establishment of an optimum computing technique that is conscious of each technology layer is crucial. In this paper, we propose a latency-oriented computing technology specialized for a domain and show its effectiveness and feasibility.

The first contribution of this paper is the proposal of a new low latency execution method for MPC (Model Predictive Control) application in the many-core processor. Since the many-core processor is a throughput-oriented computer utilizing parallelism, improvement of latency is inherently difficult. Hence, I devise a method of pre-executing the program with input value prediction before the true input data arrives by allowing calculation errors. Since the method does not rely on conventional thread/data level parallelism, it can be easily applied to such control systems without changing the algorithm in applications.

The second contribution is constructing an evaluation platform of nanophotonic computing based on its power-performance modeling for neural network applications. Nanophotonic computing which utilizes light waves as an information medium attracts considerable attention recently. Since optics has properties such as low delay, low power consumption, and high noise immunity, it will be suitable for edge computing. However, fundamental research has just begun worldwide, and the feasibility and the performance limit at the system level are veiled. In this paper, I construct an evaluation framework enabling various parameters searching from the perspective of neural network applications to the device in order to evaluate system-level performance.

The third contribution is to propose a gradient calculation method on optical devices to measure the difference using a minute phase difference occurring between wavelengths by focusing on the properties, that is wavelength dependency in the optical phase shifter which is the constituting fundamental element. This method enables to calculate with light speed because the phase difference generated as a physical phenomenon directly link to the calculation result, unlike the conventional digital signal processing.