White Paper
Insights to innovation of Deep Learning Unit
An Interview with the Developer

Among the AI technologies, Deep Learning has received particular attention. Its main feature is that it can be applied to a very wide range of objects by learning actual data, and high recognition accuracy can be obtained. On the other hand, it is necessary to operate a system platform with high processing performance for a long time, since complex calculations are repeated on huge data for learning. Fujitsu is releasing Deep Learning Unit hereinafter called DLU that is an AI processor designed for Deep Learning in 2019 to solve this issue. DLU's target performance per watt is 10 times of a major processor for learning platform.

Introducing domain specific processor for Deep Learning into FUJITSU AI Zinrai Deep Learning System

Fujitsu systemizes the knowledge and technology on AI that has been cultivated since the 1980s under the name of "FUJITSU Human Centric AI Zinrai", hereinafter called Zinrai, and provides it as a number of solutions and services that make it easy to introduce and use AI. Fujitsu will provide an on-premises environment "FUJITSU AI Zinrai Deep Learning System", hereinafter referred to as "Zinrai Deep Learning System", as system platform for deep learning. This enebles users to quickly build high-performance deep learning platform and easily start working on using AI. And DLU will be included in this new Zinrai deep learning system lineup, available for users. It is expected that the learning time can be shortened, the power consumption can be reduced by excellent power saving, and more users can use deep learning.

DLU is developed by Fujitsu's development team rich with experience in developing supercomputer processors, like K computer processor etc. Takumi Maruyama who leads the processor development team as a Senior Director gave an interview and showed DLU's latest status and his thought on its development.

Takumi Maruyama
Senior Director
AI Platform Business Unit
Fujitsu Limited

Processors for specific purpose

DLU is more specialized in specific purpose than existing processors

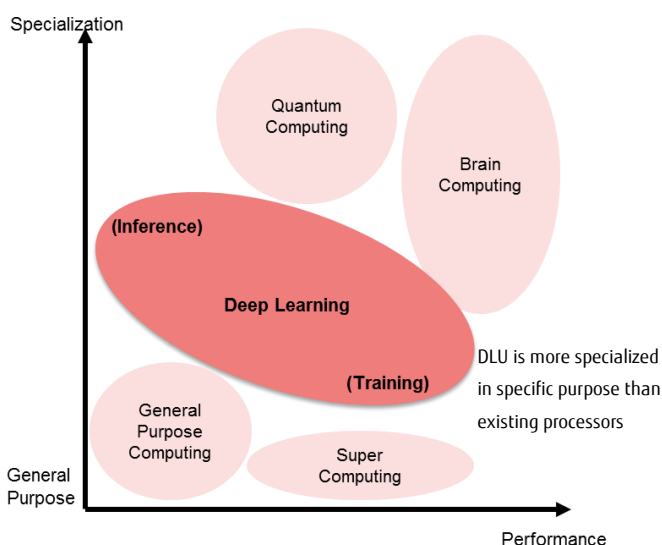### Challenges facedin this new field by the development team

----- " First of all, please tell us about the history and achievements of Fujitsu's processor development and the characteristics of the processor development team."
Mr. Maruyama: Our processor development has a history of over 60 years. Starting from mainframes, we have worked on UNIX servers, supercomputers, and so onaiming for the world's number one performance. It is a single team of experts . By sharing the basic micro-architecture and designing to develop variations according to the purpose, it is easy to apply technologies developed in one area to other areas. For example, our processor is enhancing reliability by adding a unique check function to the computing unit itself and this design is common to the "K computer" processor. Therefore high reliability is also achieved in supercomputers.

----- "How did DLU development start?"
Mr. Maruyama: DLU was born out of a brainstorm: taking advantage of our competencies to create some new value. About three years ago. Based on the market movement that is not only general-purpose but also special-purpose processors will expand in the future, we thought that we would like to contribute to society by applying the processor technology we have cultivated to a new field, Artificial Intelligence. In addition, I thought that I wanted to make something from Japan, so I decided to take on the challenge.

Requirements specific to "special-purpose" and ingenuity of DLU design

----- "The name DLU comes from the "Deep Learning Unit" and is positioned as a processor specialized for deep learning. What is the difference with other processors?"

Mr. Maruyama: The processor we have worked on until now is a general purpose processor, which basically has various circuits to "speed up a wide variety of applications". An easy-to-understand example is cache memory, in which hardware manages data by predicting software behavior.
On the other hand, deep learning is a new field, and it is not necessary to be conscious of existing applications. We focused on the hardware performance itself and took an approach to extract that performance
by software control. In other words, by eliminating hardware black-boxing, software is able to draw 100% of the power of hardware. For example, the function equivalent to cache is not provided daringly but realized by software managing a huge register file on hardware.
This is a design philosophy to achieve the best performance by combining it with a dedicated library. Frameworks are important for AI, so there is no problem if layers above frameworks can be used as they are. These frameworks are included in the Zinrai deep learning system.

Integer operation for DL: Deep Learning Integer

Data Size and Precision
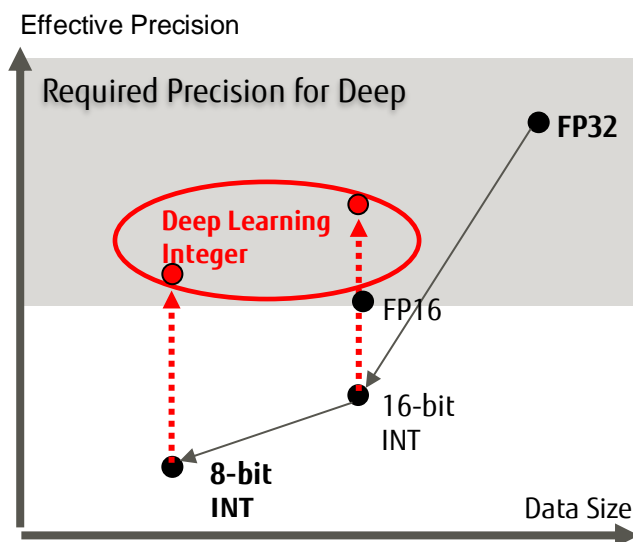
Circuit technology "Deep Learning Integer" developed by Fujitsu to increase power efficiency in deep learning. 8-bit data size achieves the accuracy needed for DL with power saving compared to FP32
Source: ISC High Performance 2017 "Fujitsu HPC and AI Processors"

----- "What specific design improvements do you use in DLU to increase hardware performance?"

Mr. Maruyama: One key requirementof Deep Learning is that it repeats the same type of operation over an over. So, for example, in the case of an SIMD (applying one instruction to multiple data) type instructions, it is common to handle at most about 512 bits at a time, whereas in DLU it is 4096 bits. In addition, our company's unique circuit technology "Deep Learning Integer" is adopted, which suppresses the decrease in calculation accuracy while reducing the number of input / output bits to the computing unit.
DLU incorporates various other design ideas to greatly improve the processing capability per operating clock. This keeps the operating clock frequency low and the power consumption low.

----- "In the AI field, GPUs (graphics processing processors) are already mainstream at present, and development of new dedicated processors that are likely to become a competitor to DLU is also in perhaps in progress. How competitive do you expect DLU is compared to them?"

Mr. Maruyama: GPUs are high-performance, but there are also issues such as high power consumption. Focusing on deep learning applications will enable more optimization. Currently, there are not many customer options yet, but the AI market will continue to grow and processor competition will continue to intensify. We believe that we can beat this competition by leveraging our years of processor development experience.

DLU supports a wide range of users and extends the scope of AI

----- "What kind of users do you want to use DLU?"

Mr. Maruyama: DLU should be useful for big users such as platformers and general users who want to use AI in their own business. Some of the people who are interested in deep learning may be hesitant because of the image that they have to wait for months of learning and hundreds of servers are required.
We believe that DLU will benefit those people to start using AI today. We also look forward to the market response. When we announced "K computer" once, it became a topic not only for the processor industry but also for the general public. In the same way, I would be happy if it had a major impact on society.