專 題 演 講

講 者： 葉宗泰 教授（交通大學資訊工程學系）

題 目： Scaling Performance and Energy Efficiency Through

Domain-Specific Accelerators

摘 要：

Machine Learning has captured tremendous successes to solve difficult problems but requires a high amount of computation to sustain the training accuracy in deep neural network (DNN) models. At the end of technology scaling, general-purpose processors are hard to accelerate DNN applications significantly because of their complicated operating overheads. As a result, people start to look to alternative computer architectures with lower overhead to continue scaling of performance and energy efficiency. For instance, Google has released TPU to accelerate its machine learning workloads.

In this talk, I will introduce the domain-specific accelerator (DSA), which is a hardware computing engine and specialized for a particular domain of applications. This DSA has successfully accelerated applications in deep learning, graphics, bioinformatics, image processing, etc.. Thus, in this talk, I will first characterize machine learning workloads and raise needs to apply the DSA in machine learning workloads. Second, I will delve into recent DSA architectures and present their limitations. Finally, I will discuss some possible methodologies to make DSA better.

簡 歷：

Tsung Tai Yeh is an assistant professor in the Department of Computer Science at National Chiao Tung University. He received a Ph.D. from the School of Electrical and Computer Engineering at Purdue University. He was the recipient of the Lynn Fellowship at Purdue University. His research results were also published in top-tier ACM and IEEE conferences and journals and filed as the patent. His research interests are focused on designs of computer architectures and systems such as domain-specific hardware accelerators, GPU, and memory systems. papers.