

## 〔短期在外研究〕

研究者	理化学研究所 特別研究員 張 浩	2166002
共同研究者/ 所属・職名	Simon. W. Moore / ケンブリッジ大学 教授	
滞在期間	2016 年 6 月 13 日 ~ 8 月 16 日間	
滞在地	ケンブリッジ イギリス	
研究課題	Developing an Optimized Accelerator for Big Scale Deep Neural Network Computer 大規模ディープニューロン専用計算機に最適化した演算加速器の開発	

## 概 要 :

脳の構造に触発された学習アルゴリズム, ディープニューラルネットワーク (DNN) は, 他のアルゴリズムを圧倒する性能によって最近, 特に注目集めている。DNN の応用範囲は, 画像処理, 言語認識, 自然言語処理, バイオインフォマティクスなどの領域に及んでおり, これら様々な領域で既に最新の成果を生んでいる。現状よりも, さらに大規模のデータを処理する巨大な DNN の持つ可能性は計り知れないものがあり, 究極の人工知能の実現へと近づく可能性もある。しかし一方で, 処理データの大規模化や, DNN 自体の巨大化に伴って, 汎用 CPU, GPU を用いた標準的な実装では, その演算性能の限界が早速問題となってくる。そこでさらなる巨大な問題を扱うひとつの手段として, DNN に特化した専用計算機の創出は目指す意義が大きい。その主要構成要素となるディープニューロン演算加速チップの基本開発が本派遣の主目的であり, チップ内に集積するニューロン数の最大化, ニューロン間通信の最速化が重要な課題である。ケンブリッジ大の Moore 教授のグループはスパイクニューロンのシミュレータとして 1 チップに 1 千ニューロンを集積した実績を持ち, NOC の開発でも世界をリードしており, バッファレス NOC を用いた, ニューロン間通信の高速化とニューロンの高密度実装化という新機軸を追究するには理想的な環境である。本派遣によって得られる知見は, 直接的に本研究の強力な推進力となる。

一般に, 大規模 DNN では, “深い” 層を用いた計算が必要となり, 多数の “ニューロン” を実装する必要がある。我々は, FPGA 上に単一加速ノードを実装しているが, FPGA には論理回路とメモリ領域の制約がある。従って, この問題は, FPGA の最適化問題となる。我々は, バッファレスネットワークオンチップ (NOC) を用いることで, ルーター領域の面積を削減し, ルータに接続するプロセッシングエレメント (PE) ひとつあたりに実装できるニューロンの数を増加させることによって, この問題に対処した。本派遣において, 単一加速ノードに用いる RTL レベルモジュールの設計を完成させることが出来た。このモジュールには, ルーターモジュール, 演算単位モジュール, ニューロンモジュール, オンチップメモリモジュールが含まれる。また, チップ面積の有効活用のために, バッファレス機構を採用した。バッファレス機構のおかげで, 30% から 40% のチップ面積を削減することが出来, ニューロン数を 30 - 40% 増やすことが可能となった。