IDEC Chip Design Contest

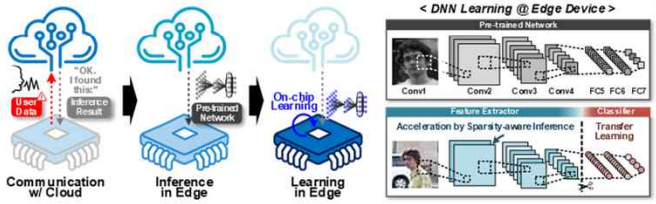
DF-LNPU: A Pipelined Direct Feedback Alignment based Deep Neural Network Learning Processor for Fast Online Learning

Donghyeon Han, Jinsu Lee, and Hoi-Jun Yoo

School of Electrical Engineering
Korea Advanced Institute of Science and Technology
Daejeon, South Korea

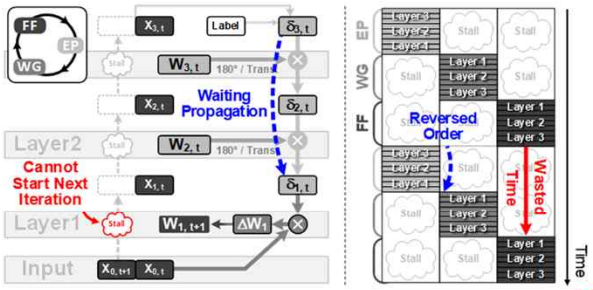
Motivation & Introduction

On-chip Learning Examples in the Edge Devices



- **Transfer Learning:** for User-specific Classification & Domain Adaptation
- **Object Tracking:** Object Detection with DNN Online Learning

Online DNN Training based Tracking Flow[1]

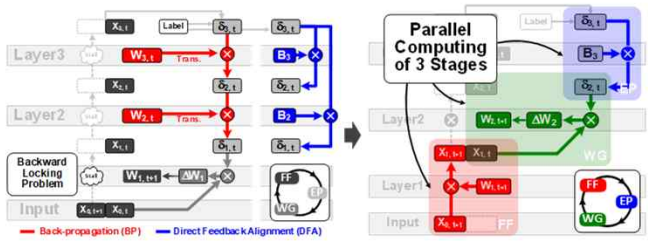


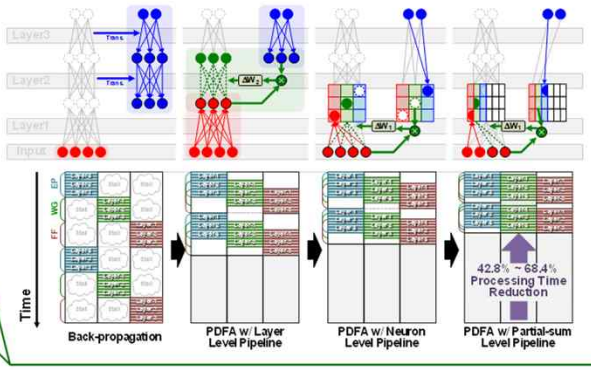
- **Sequential Computing Characteristic → Impossible to accelerate with parallel processing → Reason for Slow Training**

Pipelined Direct Feedback Alignment

Basic of Direct Feedback Alignment (DFA) & Proposed DFA

Algorithm: Pipelined Direct Feedback Alignment

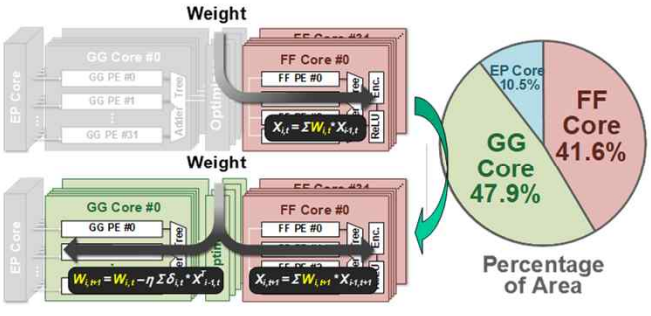


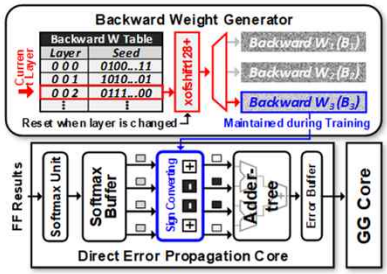


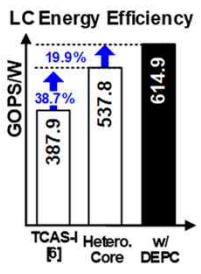
KAIST SSL

Heterogeneous Training Core

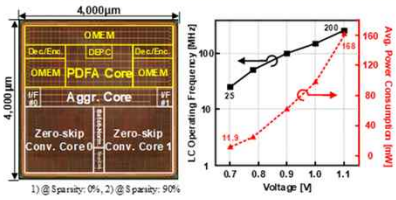
Efficiency Improvement w/ Heterogeneous Core Architecture







Implementation Results



Specification		
Technology	65 nm Logic CMOS	
Die Area	4.00 mm × 4.00 mm (16.00 mm²)	
Online Learning Core (LC)		
Die Area	3.57 mm × 1.50 mm (5.36 mm²)	
SRAM	168 KB	
Precision (A/W)	FXP 13b / FXP 16b	
Operating Condition	200MHz @1.1V	25MHz @0.7V
MAX. Throughput	155.2 GOPS	19.4 GOPS
Power Consumption	252.4 mW	17.9 mW
Energy Efficiency	614.9 GOPS/W	1083.8 GOPS/W
Zero-skip Convolution Core (ZCC)		
Die Area	1.70 mm × 1.66 mm × 2 core (5.64 mm²)	
SRAM	84.5 KB / core	
Precision (A/W)	FP 8b / FP 8b	
Operating Condition	200MHz @1.1V	50MHz @0.78V
MAX. Throughput	300~1002.1 GOPS	75~250.5 GOPS
Power Consumption	171.6[1]~78.5[2] mW	20.2[1]~9.3[2] mW
Energy Efficiency	1,748[1]~12,766[2] GOPS/W	3,722[1]~27,072[2] GOPS/W