

Session Oral 7 (8/7 Wed. 15:30 – 17:00)

Session Topic: Neural Network Accelerators

Session Chair: Chi-Chia Sun (National Formosa University) and Pei-Jun Lee (National Chi Nan University)

Room: 6F 茗廳

1. 15:30 – 15:43 (S0208) A bio-potential and piezoelectric sensing system with intelligent real-time computing, compressing and communication (Best Paper Candidates)

Ing-Jer Huang, Hsu-Kang Dow, and Shih-Jung Pao

National Sun Yat-sen University

We present a wearable bio-signal sensing system that monitors electromyography (EMG), electrocardiogram (ECG), vibration, and temperature for elderly care. The system contains analog front-end circuits (AFE) for bio-potential signal sampling, amplifying and digitization, a phase-lock loop

(PLL) circuit, a digital signal controller for AFE control and signal calibration, an 32-bit microprocessor for compression and communication. The prototype is implemented with a small 5x5cm development board which is then attached to smart T-shirt and arm bands as a light-weight wearable system. A tablet-based game has been implemented to engage elder people to exercise by interacting their muscles with the game character while monitoring their muscle fatigue and heart rate variation.

2. 15:43 – 15:56 (S0177) An Electrocardiogram Classification System with Neural Network Hardware

Implementation (Best Paper Candidates)

Yu-Yi Liao, Peng-Wei Huang, and Shuenn-Yuh Lee

National Cheng-Kung University

This paper presents a real-time identification system for electrocardiogram (ECG) classification with the neural network (NN) classifier. The identification flow of the proposed system is described as following

step by step: 1. Collecting ECG lead II signal. 2. Filtering original signals by wavelet transform. 3. Calculating twenty feature values and normalizing these features. 4. Using principal component analysis (PCA) to reduce feature number. 5. Classifying the normal beat, premature atrial complex (PAC) and premature ventricular contraction (PVC) by classifier. The accuracy of the proposed method are evaluated using different normal and abnormal ECG signals taken from the standard MIT-BIH arrhythmia database. The proposed system is verified on software design. The software part is designed by python and tested by Matlab, and the hardware is implemented by the chip fabricated with TSMC 0.18um CMOS technology. All machine learning processors, including preprocessing, feature extraction, and classifier, are implemented on a chip. The training data and testing data are independent each other. In other words, the person included in training data set never appears in testing data set for blind test. The accuracy of the proposed system is about 95.45% by the verification on the software. It reveals the proposed architecture is effective for ECG classification.

3. 15:56 – 16:09 (S0072) An Acoustic DSP Processor with CNN-FFT Accelerators for Speech Enhancement

(Best Paper Candidates)

Yu-Chi Lee (1), Tai-Shih Chi (2), and Chia-Hsiang Yang (1)

(1) National Taiwan University and (2) National Chiao Tung University

This paper proposes an acoustic DSP processor with a neural network core for speech enhancement. Accelerators for convolutional neural network (CNN) and fast Fourier transform (FFT) are embedded. The CNN-based speech enhancement algorithm is adopted in this work. An array of multiply-accumulator (MAC) and coordinate rotation digital computer (CORDIC) engines are deployed to efficiently compute linear and nonlinear functions. Hardware sharing is applied to reduce hardware area by leveraging the high similarity between CNN and FFT computations. The proposed DSP processor chip is fabricated in a 40-nm CMOS technology with a core area of 4.3 mm². The chip's power dissipation is 2.17 mW at an operating frequency of 5 MHz. The speech intelligibility can be enhanced by up to 41% under low SNR

conditions.

4. 16:09 – 16:22 (S0159) Low-Complexity Neural Network-Based Digital Pre-Distortion of 5G Wideband

Power Amplifiers with Hybrid Architecture Design

You-Cheng Lu, Ching-Chun Liao, Sin-Sheng Wong, and An-Yeu (Andy) Wu

National Taiwan University

Digital pre-distortion (DPD), is exploited for power amplifier (PA) linearization. Due to the ultra-high linearity required modulations employed in 5G communications, the performance of PA linearization based on conventional polynomial-based (MP) DPD becomes limited. Besides, despite of superior linearization performance, the complexity of recently proposed deep learning-based DPD is too high to be efficiently implemented in hardware. In this paper, a low-complexity hybrid neural network (NN)-based DPD of 5G wideband PA is proposed. The linearization error can be jointly compensated by a

hybrid architecture design with the proposed NN compensation model and a coarse-grained MP linearizer. The simulation results show that, with comparable linearization performance, the total parameters can be reduced by 80% compared with the state-of-the-art NN model.

5. 16:22 – 16:35 (S0160) Recurrent Neural Network-based Equalizer with Utilization of Coding Gain in

Advance

Chieh-Fang Teng, Han-Mo Ou, and An-Yeu (Andy) Wu

National Taiwan University

Recently, deep learning has been exploited in many fields with revolutionized breakthroughs. In the light of this, deep learning-assisted communication systems have also attracted much attention in recent years and have potential to break down the conventional design rule for communication systems. In this work, a recurrent neural network-based equalizer is proposed, which not only eliminates channel fading,

but also exploits the code structure with utilization of coding gain in advance. The equalizer in conventional block-based design may destroy the code structure and degrade the capacity of coding gain for decoder. On the contrary, our proposed approach can increase the overall utilization of coding gain with more than 1.5 dB gain.

6. 16:35 – 16:48 (S0118) A New and Efficient SVM Accelerator Design

Jian-Jhang Chen, Jer-Min Jou and Ming-Han Shieh

National Cheng Kung University

Support vector machines (SVMs) are widely used in various artificial intelligence (AI) applications. Due to AI applications' high computation complexity and real-time requirement, it is critical to speed up the SVM operation efficiently. The most part of the SVM computation is the kernel functions, which dominate the overall SVM speed and need to be implemented with special hardware. In this paper, we designed a new SVM hardware accelerator that speeds up efficiently the calculation of kernel functions

by changing the form of the decision function and by tiling the loops in it. And, we had also designed a new efficient fixed-width multiplier with very low errors for use in this SVM accelerator. Therefore, our SVM accelerator has a significantly improved detection speed compared to others, and the fixed-width multiplier has the lowest errors than other approximate multipliers.

7. 16:48 – 17:01 (S0105) CNN Training Acceleration Solution Based on the FloatSD Technique and Its

Systolic Array FPGA Implementation

Mu-Kai Sun, Chu-King Kung, and Tzi-Dar Chiueh

National Taiwan University

In this paper, we propose a CNN training acceleration system design based on FPGA implementation for the floating-point signed digit (FloatSD) number representation and update method [1]. The FloatSD technology exploits the imprecision tolerant characteristic of neural network training and adopts only a couple of non-zero digits in a neural network weight, reducing the convolution multiplication to addition

of two shifted partial products. Furthermore, the mantissa field and the exponent field of neuron activations and gradients during training are also quantized. In addition, we describe in detail how the overall system operates with cooperation between software and acceleration hardware in FPGA. Finally, we present the design of a systolic array based PE Cube for execution of convolution based on FloatSD arithmetic. The measure power consumption results indicate that the FloatSD MAC is more than 20 times energy efficient than the counterpart FP32 MAC.