Oral S06

Quantum Computing, Hardware Security, and Efficient ML Algorithm

Date/Time	8/2(三)　15:30-17:00
Chair(s)	賴以威／國立臺灣師範大學電機工程學系 陳元賀／長庚大學電子工程學系

S06.1 | 15:30—15:45

A 11.9M BPS, LOW AREA, FULLY SYNTHESIZABLE PHASE-LOCKED-LOOP-BASED TRUE RANDOM NUMBER GENERATOR

Li-Lian Deng, Chun-Heng You, Pao-Ying Cheng, Paul C.-P. Chao
Institute of Electrical Control Engineering, National Yang Ming Chiao Tung University

This study presents a validated and implemented fully synthesizable phase-locked-loop (PLL) based true random number generator (TRNG). The proposed TRNG utilizes randomness of timing jitter in a PLL, which is implemented using a digitally-controlled oscillator (DCO) via a CARRY4 cell and an inverter in an FPGA. The PLL, with manually placed DCO, achieves locked output clock generation with an input clock range of 15 MHz to 90 MHz on the AMD Xilinx ZYNQ-7000 platform. For entropy digitization, an inverse Muller C-element and a flip-flop are employed, while a two-bit Von Neumann's post-processing (VN2) is used to extract the true random bits. The hardware utilization of this design is 243 lookup tables (LUTs) and 84 flip-flops, without occupying any configurable PLL resources in the FPGA. Furthermore, the study evaluates the equivalent throughput of the proposed TRNG, considering its fully-entropy output. By subjecting the randomness of the proposed TRNG to NIST SP800-90B tests, it achieves a fully-entropy throughput of 11.9 Mbps at 70 MHz. These results position the proposed TRNG as a suitable choice for FPGA implementations with both sufficient throughput and low hardware utilization.

S06.2 | 15:45—16:00

Synthesis of Rotation Operations for Fault-Tolerant Quantum Computation

Tian-Fu Chen[1], Cheng-Han Liu[1], and Jie-Hong R. Jiang[1,2,3]
[1] Graduate School of Advanced Technology, National Taiwan University
[2] Graduate Institute of Electronics Engineering, National Taiwan University
[3] Department of Electrical Engineering, National Taiwan University

Rotation operations are essential ingredients in quantum algorithms. However, fault-tolerant quantum computation is intrinsically digital, based on a finite set of primitive gates, and thus cannot provide an analog rotation operation with arbitrary angles. It must resort to approximating a rotation operation using a basic universal gate set. Prior work on rotation synthesis either focuses on synthesizing individual rotations without considering a quantum circuit as a whole or is restricted to the specific quantum Fourier transform circuit. In this work, we develop a holistic and general approach to rotation synthesis taking all rotation operations of a circuit into account for quantum cost minimization. Experimental results demonstrate the superiority of our method compared to existing methods in reducing the cost of synthesized quantum algorithms with rotation operations.

S06.3 | 16:00—16:15

Design of Fast Truncated Polynomial Ring Multipliers for NTRU Algorithm

Chia-Yang Kang, Shiann-Rong Kuang, Meng-Wei Shen, and Xin-Han Liu
Department of Computer Science and Engineering, National Sun Yat-sen University

The development of quantum computers will produce great risks to current communication systems, which must integrate secure post-quantum cryptography to prevent attacks. The NTRU cryptosystem is one of the main alternatives for practical implementations of post-quantum public-key cryptography and the truncated polynomial multiplication is usually its performance bottleneck. Therefore, several hardware architectures have been proposed to detect the consecutive zero coefficients of blinding polynomial r(x) and skip the unnecessary polynomial multiplications to enhance the performance of NTRU cryptosystem. In this paper, we analyze in detail the expected value of consecutive zero coefficients for each parameter set in the standardized version of NTRU (IEEE 1363.1). As a result, befitting multiplexer size can be selected for each parameter set to avoid a large number of redundant polynomial multiplications. Compared with previous design, the proposed architecture can achieve 27% to 68% speed enhancement and 39% to 57% area-time product (ATP) improvement, respectively.

S06.4 | 16:15—16:30

Qubit Mapping for Trapped-Ion Systems Using Satisfiability Modulo Theories Approach

Wei-Hsiang Tseng[1], Yao-Wen Chang[1,2], and Jie-Hong Roland Jiang[1,2]
[1] Graduate Institute of Electronics Engineering, National Taiwan University
[2] Department of Electrical Engineering, National Taiwan University

Qubit mapping plays a crucial role in optimizing the performance of quantum algorithms for physical executions on quantum computing architectures. Many qubit mapping algorithms have been proposed for superconducting systems recently. However, due to their limitations on the physical qubit connectivity, costly SWAP gates are often required to swap logical qubits for proper quantum operations. Trapped-ion systems emerge as an alternative quantum computing architecture and gain much recent attention due to their relatively long coherence time, high-fidelity gates, and good scalability for multi-qubit coupling. However, the qubit mapping of the new trapped-ion systems remains an important research problem to be tackled. This paper proposes a coupling constraint graph to model the unique constraints and connectivity patterns in trapped-ion systems. To minimize the time steps for quantum circuit execution satisfying the coupling constraints for trapped-ion systems, we propose a new approach combining Satisfiability Modulo Theories and the divide-and-conquer technique to achieve efficient qubit mapping on trapped-ion quantum computing architectures. Experimental results demonstrate better scalability and effectiveness in our qubit mapping solutions compared to the previous work.

S06.5 | 16:30—16:45

Bit-Wise Quantization-Aware Training in Two's Complement Representation

Zih-Huang Cheng, Chih-Hung Kuo and Chun-Hao Chan
Department of Electrical Engineering, National Cheng Kung University

The von Neumann bottleneck becomes apparent for modern digital neural network accelerators. Researchers have explored the techniques of Compute-In-Memory (CIM) that stores weights in memory cells and perform in-situ computations to enhance the throughputs. Nonetheless, a large amount of multiply-accumulate (MAC) operations to compute for a deep neural network (DNN) requires the use of higher-resolution Analog-to-Digital Converters (ADCs), which in turn leads to high area cost and energy consumption. In this paper, we propose a bit-wise approach of Quantization Aware Training (QAT) for weight coefficients. This method enhances the sparsities in bit level and reduces the accumulated values of MAC computations, allowing for the use of lower-

resolution ADCs to save area and power consumption. Weights of filters can be trained as the format of two's complement for CIM operations. Experimental results demonstrate that the weights of VGG-16 model are quantized into 8-bit integers with the bit-level sparsity 97.91%, while a top-1 accuracy of 93.78% is achieved. With such sparsity, the required ADC resolution can be reduced from 6 bits to 1.25 bits in average, and hence significantly alleviates the power and area burden imposed on CIM accelerators.

S06.6 | 16:45—17:00

Implementation of an Annealing Chip for Advanced Quantum Computing Applications

Yu-Jie Yen[1], Yuan-Ho Chen[1,2]

[1] Dept. of Electronics Engineering, Chang Gung University
[2] Dept. of Radiation Oncology, Chang Gung Memorial Hospital

In recent years, due to the rise of the Internet of Things, the computing performance requirements of edge devices are also getting higher and higher. How to deal with complex combinatorial optimization problems in edge devices is regarded as a very important key. This paper applies the state of simulated quantum annealing to the calculation of the Ising model, and uses local calculations, multiple cycles of continuous relative annealing, and multi-threaded parallel computing methods to shorten the time to find the best solution and improve hardware performance. This paper uses a TSMC 90nm chip, which can achieve a power consumption of 36mW and a chip area of 2.02mm × 2mm at a frequency of 100MHZ.