(B)

2014   2017

Self-learnable Analog-Digital-Mixed VLSI Processors for Smart
Human-Computer-Interaction

Self-learnable Analog-Digital-Mixed VLSI Processors for Smart
Human-Computer-Interaction

ZHANG, Renyuan

2,900,000

ACU

CNN

The analog-digital-hybrid computing processors have been developed for the
human-computer-interaction (HCI) relevant applications. The approximate computing technologies were
implemented in silicon to offer a different computing methodology from traditional binary processing
on the basis of "deductive computation". In this manner, the performances on computing speed and
energy consumption were greatly improved. The ACU processor was successfully developed as the
world-first programmable analog calculator for multi-operand calculation, which greatly speeds up
the widely applied HCI technologies such as CNN

Computer Architecture

Analog-Digital-Hybrid Approximate Computing Analog Calculator Multi-Valued-Logic

１．研究開始当初の背景
(1) The explosive development of HCI along with AI

During recent years, the human computer interaction (HCI) technologies have been widely applied due to explosive development (also demands) of artificial intelligence (AI). In many tasks of HCI, not only software programs on PCs but also specially designed VLSI processors are necessary. Thus, the efficient and well-performance VLSIs for computations are demanded.

(2) Next generation of computing technologies for post-Moore era

The road-map of Moore's Law has been approaching the end. Innovations of frontier computing technologies should be explored out of the main-stream of binary processing. Our efforts on the approximate computing impact not only HCI but more general-purpose applications for next generation of Post-Moore.

２．研究の目的

As a core part of HCI processors, the computational units are considered as one of most challenging technologies. In most cases, the computations of HCIs are very massive but the requirement on precision might not be very high. Thus, sufficiently compact and efficient computational units are demanded for the highly parallel processing.

On the other hand, the computations for HCIs are usually very flexible and unpredictable. Most of previously reported accelerators are application-domain specified, which means the functions are hardly reconfigured and re-programmed.

The goal of our project is to offer efficient and compact computational units with functional flexibility and programmability. Then, it becomes feasible to integrate massive computing cores in highly parallel (such as Google TPU, it was made by similar strategy to ours).

３．研究の方法

Entire project consists two types of efforts: one is the multi-valued-logic (MVL) processors; the other one is programmable analog calculation units.

[MVL-based computational units]

As the first half of this project, we developed several types of computational processors, memory systems, and relevant circuits for human-computer-interaction (HCI) data. From our previous investigations, it was found that the HCI data processing is different from general purpose calculations: it should be quite fast and efficient, but inaccuracy tolerant. Thus, we built many multi-valued-logic processors in the fashion of FPGA. During the previous year

of this project, we proposed several 4-valued logic (quaternary) FPGA with benefits of efficiency and interconnections. However, it is still not satisfying since quaternary data is too inaccurate for HCI processing. The benefit of efficiency is easily eaten up by designing complexity. Thus, we developed 16-valued logic (Hexadecimal) circuitries to compact the interconnections in further. We successfully verified the behavior of Hexadecimal processors, which greatly reduce the number of devices and interconnections. In HCI processing, image recognition for instance, Gaussian function is very important. Thus, we designed the hexadecimal FPGAs to implement functions such as Gaussian. From our experiments, the designed hexadecimal FPGA carried out all the expected functions reliably. As we expected, the hexadecimal processors greatly reduce interconnections (from four to one), which allows us to implement fully parallel computing for HCI tasks.

[Programmable Analog Calculation Unit]

From the investigations of MVL computing, we found it is even feasible to improve the efficiency by extending the MVL to pure analog. However, traditional analog calculators are usually function specified. Thus, we are developing programmable analog calculation units (ACUs). We retrieve the ``computation" by regression instead of traditional Boolean logic-based derivations.
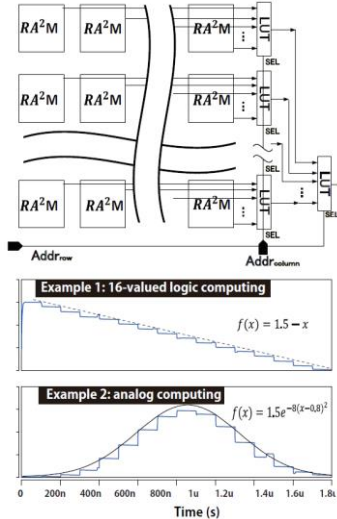
we are trying to construct a programmable network by hardware based on the regression algorithms. This network is considered as one ACU to behave as multi-input calculator. The demanded functions are known and fully listed as data-base samples to train the regression net-work, which is considered as "synthesis" of ACU. Compared to digital ALU, this ACU computes arbitrary vector functions in real-time (no clock cycles) with compacter hardware resource. The support vector regression (SVR) is adapted due to following: 1. It has well performance for high dimensional vector regressions; 2. It helps to reduce redundant samples, then shrink our circuit scale; 3. We hold experiences of circuitries for emulating and programming SVM node models. It is expected to build world-first programmable analog calculators, which has a similar bit-length with Google TPU but much smaller chip area.

Then, we are trying to extend the number of inputs to very high such as nine, which is usually two or three by traditional ALUs. The multi-input ACUs are used for the image processing algorithms such as CNN to improve the processing speed.

４．研究成果

[MVL-based computational units]

The MVL-based FPGA structure has been developed from the project. As seen in the sided figure, we designed the analog memory, hexadecimal RAM, and the hexadecimal addressing block to construct the FPGA-like architecture, which carries out arbitrary hexadecimal functions. Several examples are also given in this figure. By using this technology, the number of devices are reduced to **about 30% of traditional FPGAs** with the same precision (4-bit). At the same time, the interconnections are greatly reduced due to the hexadecimal representation.

[Programmable ACU]

From our previous experiments, it is feasible to design function-programmable analog calculation units (ACUs) with an accuracy loss of ~5%. So far, **two-operand ACUs was designed by 600 MOS transistors and perform one-cycle computations of arbitrary complex functions; 9-operand ACU with similar performance only consists of 5000 transistors,** which is greatly compact compared to binary ALUs with same accuracy (**the energy efficiency is increased to ~100x to traditional ALUs; the area is reduced to ~1/20x**). As seen in the sided figure, the regression technology (powered by SVM) is realized on-chip. This strategy solves a common problem of
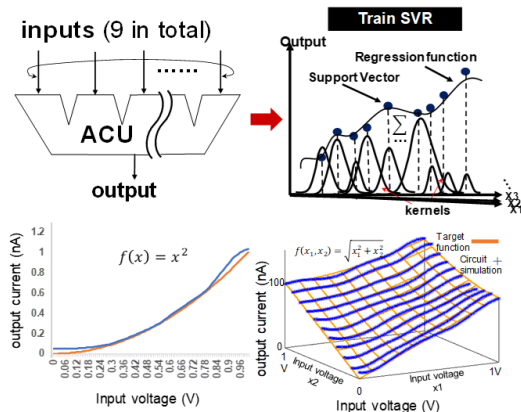
analog computational circuit: **programmability**. It has been found analog calculators perform very high speed, low power and extremely compact chip area, but suffer from the poor flexibility, which we solved. By using the soft-mapping methodology, our works explore a different direction of analog computing. **As the tentative exploration, we expanded the programmable ACUs to very highly dimensional vector computing. For instance, it is even feasible to carry out matrix-production within ONE clock cycle by ONE calculation unit.** Since the size of ACU could be much smaller than 8-bit ALU (referenced to Google TPU) but the computational capacity is higher (multi-operand), it can be additionally placed along with ordinary digital modules with slight overhead.

In general, we successfully developed various approximate computing units with compact area and high efficiency, which are suitable to apply in the HCI relevant processors.

５．主な発表論文等
〔雑誌論文〕（計１件）

R. Zhang, and M. Kaneko, "Robust and Low-Power Digitally-Programmable-Delay-Element Designs Employing Neuron-MOS Mechanism", ACM Tran. Des. Autom. Electron. Syst. (TODAES), Vol. 20, No. 4, Article 64, (19 pages), September 2015. DOI: 10.1145/2740963

〔学会発表〕（計９件）
①Noriyuki Uetake, Renyuan Zhang, Takashi Nakada, and Yasuhiko Nakashima: "A Programmable Analog Calculation Unit for Vector Computations", IEEE Symposium on Low-Power and High-Speed Chips 2018.

②R. Zhang, T. Nakada and Y. Nakashima, "A Feasibility Study of Programmable Analog Cal-culation Unit for Approximate Computing", The Fifth International Symposium on Compu-ting and Networking, (CANDAR), Aomori, Japan, Nov. 19-22, 2017 to appear (Outstanding Paper Award).

③R. Zhang, M. Kaneko, "A Random Access Analog Memory with Master-Slave Structure for Implementing Hexadecimal Logic", IEEE Int. Conf. System-on-Chip, (SOCC), Munich, Germany, Sept. 5-8, 2017, pp. 7-11.

④R. Zhang, and M. Kaneko, "A Feasibility Study of Master-Slave Flipflop Design for Hexadecimal Logic", IEEE Int. Conf. Industrial Electronics and Applications, (IEACon), Nov. 2016.

⑤R. Zhang, and M. Kaneko, "A 16-Valued Logic FPGA Architecture Employing Analog Memory Circuit", IEEE Int. Symp. Circ.s and Syst.s, (ISCAS), Montreal, Canada, May 22-25, 2016, pp. 207-212.

⑥R. Zhang, and M. Kaneko, "A Feasibility Study of Quaternary FPGA Designs by Implementing Neuron-MOS Mechanism", IEEE Int. Symp. Circ.s and Syst.s, (ISCAS), Lisbon, Portugal, May 24-27, pp. 942-945. 2015.

⑦R. Zhang, and M. Kaneko, "A Quaternary Master-Slave Flip-Flop with Multiple Functions for Multi-Valued Logics", the 19th Workshop on Synthesis and System Integration of Mixed Information Technologies, Yilan, Taiwan, March 16-17, pp. 193-198. 2015.

⑧R. Zhang, and M. Kaneko, "A Temperature and Process Variation Insensitive PDE Circuit Employing Neuron-MOS", IEEE/ACM Int. Conf. Computer Aided Design, Workshop on VMC, San Jose, USA, Nov. 2014.

⑨R. Zhang, and M. Kaneko, "A Feasibility Study on Robust Programmable Delay Element Design based on Neuron-MOS Mechanism", Great Lake Symposium on VLSI, pp. 21-26, May 2014.

〔図書〕（計　0　件）

〔産業財産権〕

○出願状況（計　0　件）

○取得状況（計　0　件）

〔その他〕
ホームページ等
http://arch.naist.jp/~zhang/pub.html

６．研究組織
(1)研究代表者
　ジャン　レンユアン（ZHANG RENYUAN）
　奈良先端科学技術大学院大学・情報科学研究科・助教
　研究者番号：00709131

(2)研究分担者
　なし

(3)連携研究者
　なし

(4)研究協力者
　なし