



UTM
UNIVERSITI TEKNOLOGI MALAYSIA

Faculty of
Electrical
Engineering



VeCAD

VLSI-Embedded Computing Architecture Design

VeCAD Annual Research Report 2023

Integrated Circuits and Systems

Foreward

It is with great pleasure and pride that I present to you the annual research book of the VLSI-Embedded Computing Architecture Design (VeCAD) 2023. As the head of this dynamic and dedicated group, I am honored to showcase the exceptional work and projects undertaken by our members in the fields of electronic systems design and computer engineering for the year 2022-2023.

This year's compilation reflects our commitment to advancing the frontiers of knowledge in diverse areas, including high-level modeling, system-on-chip (SoC), network-on-chip (NoC), FPGAs, and custom ICs. Some of the applications of these technologies include AI and machine learning, image processing, memory design, and stochastic computing. The collective efforts of our researchers have resulted in a rich innovative solutions, pushing the boundaries of what is possible in the ever-evolving landscape of electronic systems.

The world of technology is rapidly evolving, and at the heart of this transformation lies the ingenuity and dedication of our research group. The projects presented in this book represent months of rigorous exploration, experimentation, and collaboration. From conceptualization to implementation, each endeavor showcases the passion and expertise of our team members.

I would like to extend my heartfelt gratitude to every contributor who played a pivotal role in the success of this year's research initiatives. Your hard work, creativity, and perseverance have not only elevated the standing of our research group but have also contributed significantly to the advancement of knowledge in our field.



Ab Al-Hadi Ab Rahman
Head of VeCAD Research Group
January 2024

Editors:
Dr. Afiq Hamzah
Dr. Ab Al-Hadi Ab Rahman

Contents

Artificial Intelligence System-on-chip Muhammad Nadzir Marsono, Ab Al-Hadi Ab Rahman, Mohd Shahrizal Rusli, Shahidatul Sadiah, Afiq Hamzah	3
Compact Spectral-based Convolutional Neural Networks Ab Al-Hadi Ab Rahman, Shahriyar Masud Rizvi, Omid Aayat	4
Fast and Energy Efficient Optical Character Recognition with Spectral CNN Ab Al-Hadi Ab Rahman, Shahriyar Masud Rizvi, Ibrahim Yousef Alshareef, Nuzhat Khan	5
Design of Boosted 10 Transistor GC-eDRAM Processing in Memory (PIM) Cell for Multiply Accumulate (MAC) Operation Afiq Hamzah, Hui Qi Chung, Fauzan Fikri, and Izam Kamisian	6
Stochastic Computing Packet Framework Design of Sobel Edge Detection Optimized for Area and Energy Efficiency Afiq Hamzah, Omar Elsayed, and Izam Kamisian	7
Interleaved Incremental-Decremental Support Vector Machine for Embedded Applications Muhammad Nadzir Marsono, Nasir Shaikh Husain, Jeevan Sirkunan	8
Traffic-Aware Token-Based Medium Access Control Mechanism for Wireless Network-on-Chip Muhammad Nadzir Marsono, Mohd Shahrizal Rusli, Ayodeji Ireti Fasiku	9
Deep Pipeline Architecture for Fast Fractal Color Image Compression Utilizing Inter-color Correlation Abdul-Malik H. Y. Saad	10
Test Point Insertion for Testability Improvement Norlina Paraman	11
Stochastic Computing for Low Area and Low Power Applications Izam Kamisian	12
N-Gram Feature Extraction and Naïve Bayes Classifier for Malware Detection using FPGA Implementation Lee Ming Yi and Ismahani Ismail	13
Image Recognition using Capsule Network on FPGA Salim Ali Abdulrazziq Adrees and Ismahani Ismail	14
High Performance Programmable Logic Controller Processor using System on Chip of Ladder Rung Processor and RISC CPU Core Zulfakar Aspar	15

Artificial Intelligence System-on-chip

Muhammad Nadzir Marsono, Ab Al-Hadi Ab Rahman, Mohd Shahrizal Rusli, Shahidatul Sadiyah, Afiq Hamzah

Artificial Intelligence (AI) system-on-chips (SoC) will power intelligent sensing, hyper-automation, and edge computing applications. The global market for AI chips is forecasted to be approximately \$80 billion by 2025 according to a market survey. These cutting-edge technologies that are being developed and researched in both industry and academia around the world. This research project aims to further develop technologies and intellectual properties related to AI SoC, specifically in network-on-chip interconnect, SoC architecture exploration on FPGA, and physical design techniques for chip optimizations.

The program is expected to produce:

- High-value AI chip technology and IPs.
- Expansion of knowledge in AI SoC architecture.
- Nurturing Malaysians AI chip design experts.

- Foster collaborations with other experts from industries and academia.
- Joint research with industries and other agencies.

Opportunities and benefits to the industry:

- Semiconductor and chip industries are the main contributors to Malaysian GDP.
- IC Design companies in Malaysia drive the growth in this sector.
- We collaborate with three major design companies, Efnix, Skyechip, and Oppstar.
- Our collaboration with such firms can assist the universities in developing skilled and holistic workforce talent to propel Malaysia forward in the E&E thrust.

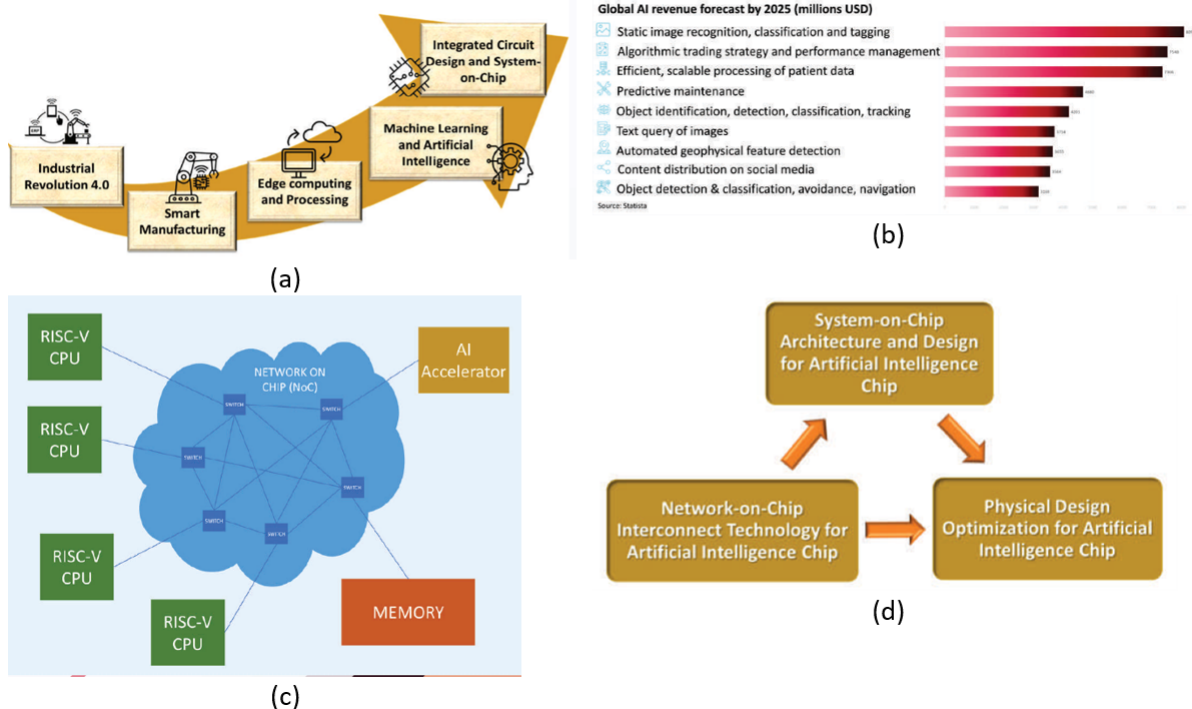


Figure 1: (a) SoC design as IR4.0 technology enabler, (b) Global AI revenue forecast, and (c) NoC architecture for AI SoC, (d) Scope of project

Compact Spectral-based Convolutional Neural Networks

Ab Al-Hadi Ab Rahman, Shahriyar Masud Rizvi, Omid Aayat

The convolutional neural network (CNN) has gained widespread adoption in computer vision (CV) applications in recent years. However, the high computational complexity of spatial (conventional) CNNs makes real-time deployment in CV applications difficult. Spectral representation (frequency domain) is one of the most effective ways to reduce the large computational workload in CNN models, and thus beneficial for any processing platform. By reducing the size of feature maps, a compact spectral CNN model is proposed and developed in this work by utilizing just the lower frequency components of the feature maps. Our proposed spectral CNN model takes an input feature map size of just 3×3 in spectral domain, and goes through

three CONV layers, i.e. C1, C2, and C3 (with the associated spectral pooling and activation layers). After layer C3, the feature map is converted back to the spatial domain via an IFFT. Here, the spectral ReLU, called SReLU is employed to prevent multiple domain switching, thus making the spectral CNN model compact and lightweight. The model is suitable to be implemented in embedded systems and FPGA with requirements for low power and high throughput. When compared to similar models in the spatial domain, the proposed compact spectral CNN model achieves at least $24.11\times$ and $4.96\times$ faster classification speed on AT&T face recognition and MNIST classification datasets, respectively.

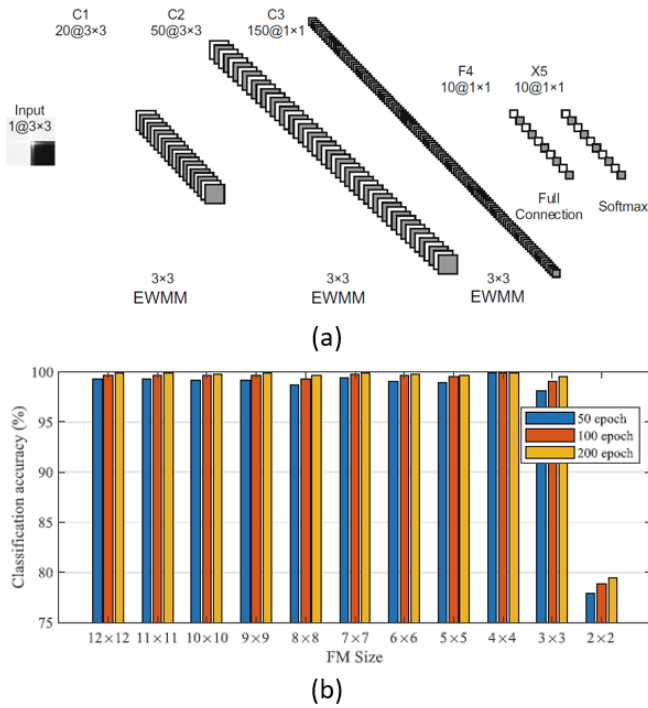


Figure 2: (a) Proposed Spectral CNN model, (b) Feature map size analysis, and (c) Results in terms of accuracy, (d) Results in terms of speed.

More info:

1. AH Awab, AAH Ab Rahman, MS Rusli, UU Sheikh, I Kamisian, GK Meng, HEVC 2D-DCT architectures comparison for FPGA and ASIC implementations, *Telkomnika*, 17(5), pp 2457-2464, 2019.
2. AH Awab, AAH Ab Rahman, I Kamisian, MS Rusli, VLSI Design of a Split Parallel Two-Dimensional HEVC Transform, *Innovations in Electrical and Electronic Engineering*, pp 431-440, 2021.

Fast and Energy Efficient Optical Character Recognition with Spectral CNN

Ab Al-Hadi Ab Rahman, Shahriyar Masud Rizvi, Ibrahim Yousef Alshareef, Nuzhat Khan

Spectral Convolutional Neural Networks (CNNs) distinguish themselves from traditional CNNs by incorporating spectral domain operations into the processing pipeline. This spectral approach streamlines the conventional spatial domain convolution operation into a pointwise Hadamard product, resulting in significantly improved speed and energy efficiency. These attributes make it particularly well-suited for mobile applications. What sets our approach apart is the integration of pooling and activation functions directly within the spectral domain, eliminating the need for costly domain transformations between CNN layers. Moreover, our innovative method reduces computational complexity

and memory access costs, yielding an exceptionally lightweight spectral CNN model that maintains high recognition accuracy. Our team has developed a comprehensive Optical Character Recognition (OCR) system, with the spectral CNN serving as the central machine learning model. It all begins with an input image containing text. We extract paragraphs and individual characters, transform them into the equivalent spectral domain, and then feed them into the spectral CNN model for recognition. The predicted characters are subsequently verified using our novel Character Verification Model (CVM). After verification, these characters are reconstructed into words, lines, and paragraphs.

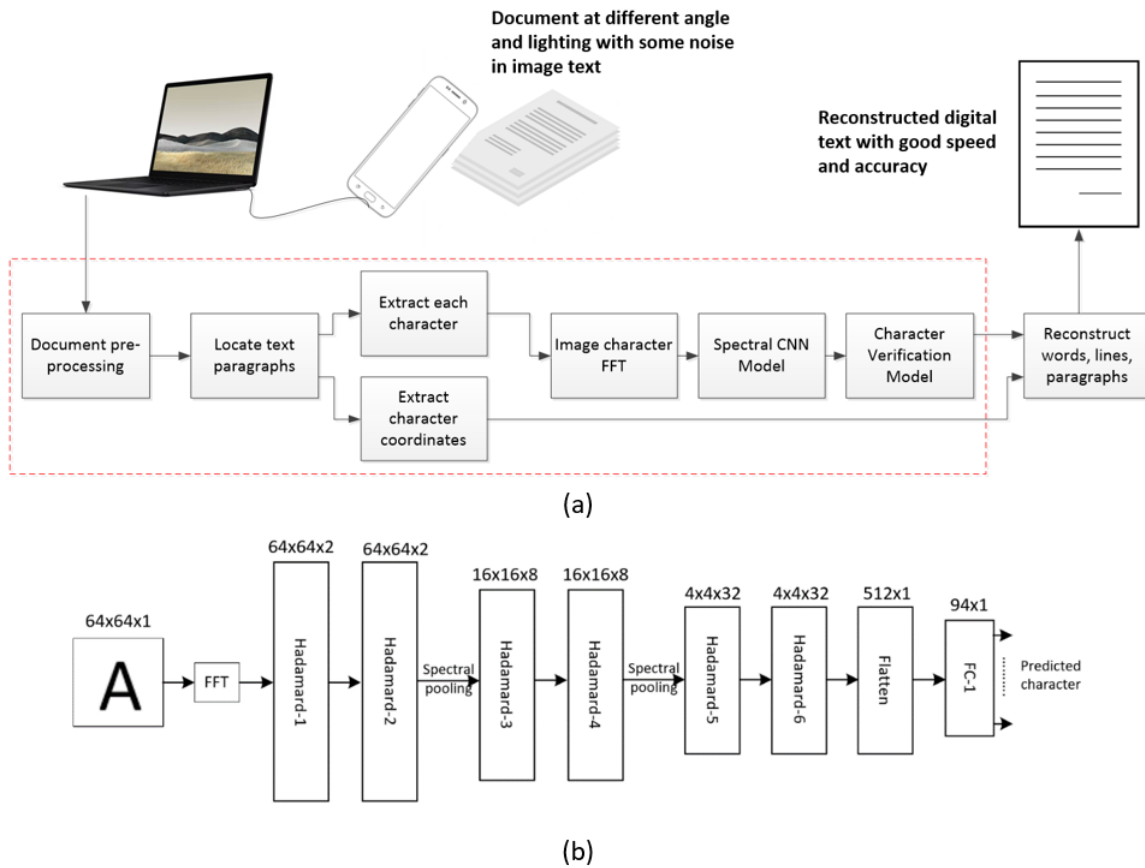


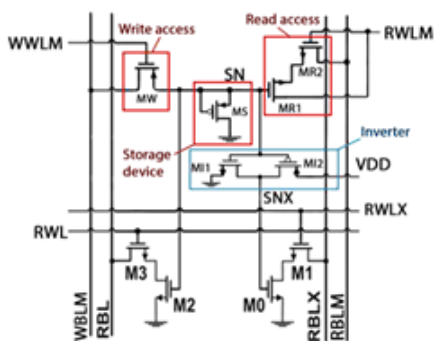
Figure 3: (a) OCR flow with spectral CNN and CVM, (b) Spectral CNN model.

Design of Boosted 10 Transistor GC-eDRAM Processing in Memory (PIM) Cell for Multiply Accumulate (MAC) Operation

Afiq Hamzah, Hui Qi Chung, Fauzan Fikri, and Izam Kamisian

Binarized Neural Network (BNN) undergoes a high workload computation, resulting in high power consumption in order to produce an accurate product. Consequently, memory computation operations play an important role in providing energy-efficient methods. Therefore, the computing method Processing in Memory (PIM), was introduced. This architecture performs memory and central processing unit (CPU) functions, increasing the data transfer speed and energy performance. It is characterized by its row-by-row storage of column data and its capability to perform MAC execution among all different pairs of datasets through its bit-by-bit data storage mechanism optimized with an optimal reference voltage. The ten-transistor (10T) SRAM PIM architecture, which combines six-transistor (6T) for memory mode and four transistors for PIM modes, was introduced to fulfill the high-performance requirement. However, 6T SRAM has a considerable amount of static power and requires a large portion of chip area, resulting in high design costs. On that account, the Gain Cell Embedded DRAM (GC-eDRAM) is

suitable memory technology to replace the 6T SRAM cell. GC-eDRAM has a denser memory architecture than SRAM, as it only requires a minimum of two transistors to store a single bit while SRAM needs six transistors. GC-eDRAM also employs a gate-controlled access mechanism, which reduces leakage power when compared to SRAM, making it more power-efficient. In this project, the boosted 10T GC-eDRAM PIM cell, a combination of three-transistor (3T) GC-eDRAM, an additional storage transistor, a transistor level inverter, and four PIM transistors, had been used to improve the issue of high-power consumption and large area design for PIM BNN macros. The boosted 10T GC-eDRAM PIM design is space-efficient, supports large-datasets, cost saving, and consumes less power than the 10T SRAM PIM architecture. For this project, a 1 Kb (32×32) PIM BNN macro block was proposed using 45nm CMOS technology was made using the Cadence-Virtuoso EDA tool. The simulated data generated from the proposed design was compared and listed with the previous works.



(a)

	[18]	[19]	[20]	[21]	[22]	[23]	This work
Technology	65nm	45nm	65nm	65nm	65nm	45nm	45nm
Macro size	16Kb	8Kb	4Kb	16Kb	2Kb	1Kb	1Kb
Cell Structure	10T	10T	Split-WL 6T	12T	8T1C	10T	10T
Memory	SRAM	Xcel-RAM	6T-SRAM	XNOR-SRAM	C3SRAM	SRAM CiM	GC-eDRAM
Input (bit)	6	1	1	1	1	1	1
Weight (bit)	1	1	1	1	1	1	1
Output (bit)	6	5	1	3.5	5	1	1
Operating frequency (MHz)	5	22.2	N/A	N/A	50	100	40
Throughput (GOPS)	8 (11.5)*	8.5	278 (400)*	614 (884)*	1638 (2358.7)*	204.8	81.92
Throughput density (TOPS/Kb)	0.5 (0.72)*	1.1	69.5 (100.08)*	38.4 (55.3)*	819 (1179.4)*	204.8	81.92

(b)

Figure 4: (a) 10T GC-eDRAM cell, (b) results comparison

Stochastic Computing Packet Framework Design of Sobel Edge Detection Optimized for Area and Energy Efficiency

Afiq Hamzah, Omar Elsayed, and Izam Kamisian

Sobel Edge Detection (SED) is one of the common algorithms used in computer vision to locate edges in a grayscale image, mainly in applications of motion detection, object tracking, and object recognition. However, implementation of this algorithm is sensitive to noise, and is computationally expensive, which consequently consumes power. Therefore, there is a need for a design of SED, that is of lower power consumption, of lower area, and of higher performance. This project is based on Stochastic Computing (SC), a non-conventional technique that computes numbers as probability bitstreams, making it suitable for the lossy applications of SED. It uses simpler arithmetic blocks, compared to its Binary-Encoded (BE) counterparts. SC is expected to reduce area and power consumption, however it suffers from long bitstreams of encoding, and the need of random number generators (RNG) that slows down the system and takes up a large area. In this work, the stochastic packet framework (SPF) is proposed, a framework that splits up SC's long bitstreams into packets, that are concurrently pro-

cessed with area-efficient parallel RNGs. SPF depends on two parameters, namely the total length T of the bitstream, and the packet lengths P . SPF is then used to design the Stochastic SED (SSED) on ASIC, using different values for T and P . The designs are assessed and analysed, in terms of Power, Performance, and Area (PPA), and then are benchmarked against the Binary SED (SED). SPF also involves using parallel RNGs with sharing schemes, and doing Stochastic Computing Correlation (SCC) tests, in order to achieve the best improvement possible for PPA. The ASIC implementation results show that the proposed SPF is effective in reducing SED's overall area and power, of the algorithm. It is observed that for small, and even medium T and P values, SPF is able to achieve a decrease in area and power consumption by up to 54.7% and 56.9% respectively. SPF was capable of improving performance in few cases only, by 19.9%. All in all, these results show that it is possible to improve computer vision in edge devices that are low-power or small in area, provided that the use application is of lossy nature.

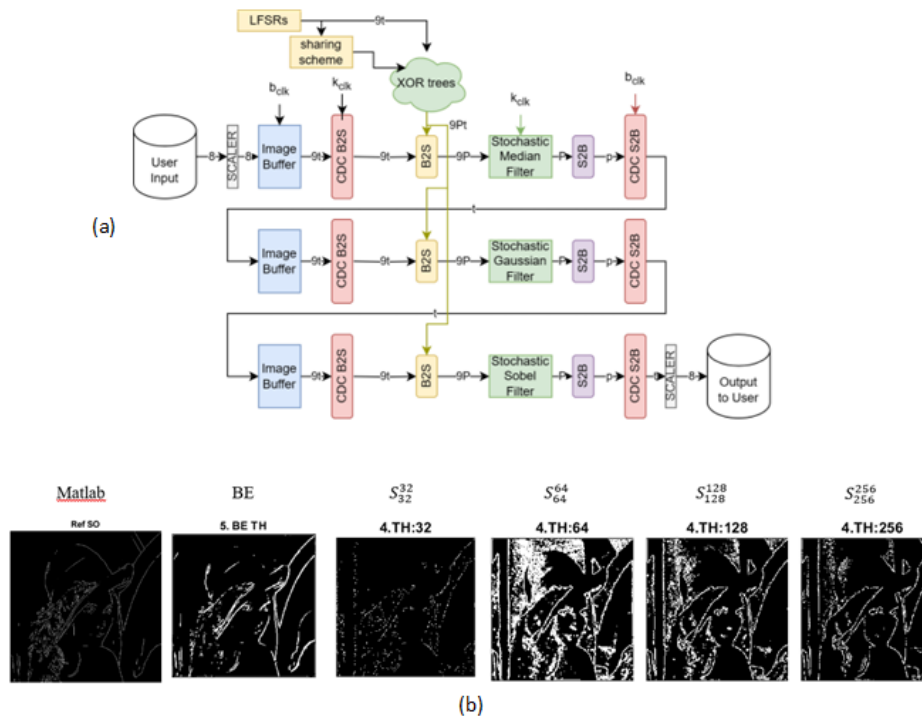


Figure 5: (a) The proposed stochastic packet framework (SPF), and (b) the results comparison of Lena's image between binary encoded design with various bitlength used in SPF

Interleaved Incremental-Decremental Support Vector Machine for Embedded Applications

Muhammad Nadzir Marsono, Nasir Shaikh Husain, Jeevan Sirkunan

Incremental Decremental Support Vector Machine (IDSVM) is one of the widely used incremental learning algorithms known for its high accuracy for data stream analytics and high computational complexity. One of the biggest problems of IDSVM is that the model scales with the input data set size that directly correlates with the computational and memory resources. In order to deploy IDSVM in an embedded system with limited memory, a moving window architecture is needed to limit the kernel sizes. However, this also increases the overall complexity of the algorithm since each data instance needs to be unlearned when exiting the window. This work proposes an Interleaved IDSVM (IIDSVM) algorithm that performs incremental and decremental learning concurrently. The interleaved method can reduce the overall kernel size and consume less memory. Our work also proposes a reduced-division IIDSVM algorithm that replaces the more complex division operations with simpler inverse multiplications. Certain IIDSVM tasks can be simplified by replacing most of the complex divisions with inverse multiplication that can achieve a similar outcome since only a single sample

variation value is used to update the weights. Finally, a Radial Basis Function (RBF) kernel, which is a widely used kernel in SVM, is proposed to be implemented as a hardware accelerator to speed up the computation time of the IIDSVM. Based on our experiments, the proposed IIDSVM achieved a speedup of 2.5 - 4.2 \times on computation time while producing similar accuracy as IDSVM and LIBSVM. Furthermore, the reduced-division IIDSVM can improve computation time up to 1.4 \times on a Nios II embedded platform for certain data sets. The RBF kernel's hardware implementation is analyzed on the Stratix V Field Programmable Gate Array (FPGA) platform. It can perform up to four orders of magnitude faster than the software implementation on the Nios II embedded processor for data sets with 8, 12, and 16 feature sizes. Besides that, the proposed architecture RBF kernel can maintain a maximum operating frequency of approximately 200Mhz for feature sizes 8, 12, and 16. Collectively the proposed works can improve the runtime of incremental SVM compute-intensive data stream analytics.

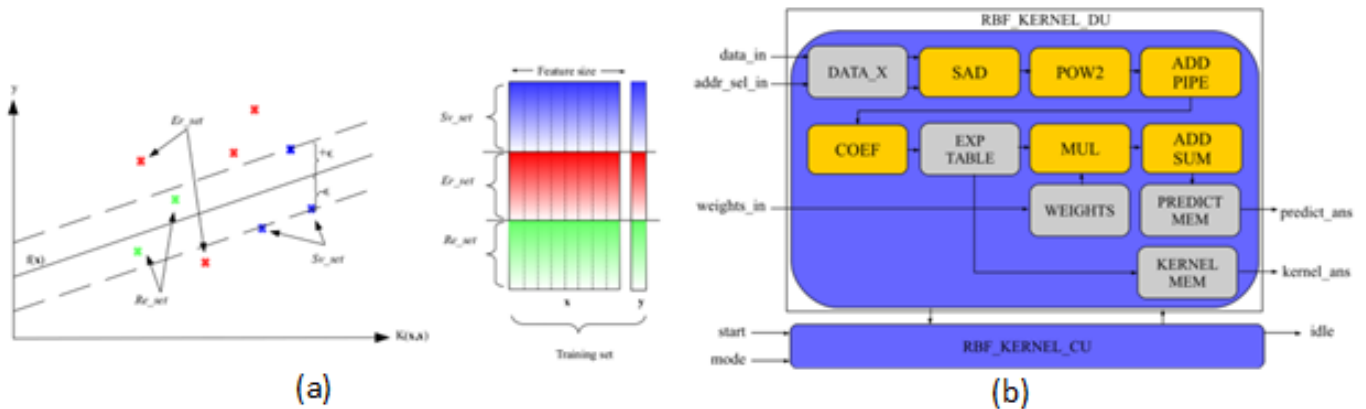


Figure 6: (a) Location of Support, Error and Remainder set in IIDSVM, and (b) Functional block diagram of *RBF_KERNEL* for IIDSVM.

Traffic-Aware Token-Based Medium Access Control Mechanism for Wireless Network-on-Chip

Muhammad Nadzir Marsono, Mohd Shahrizal Rusli, Ayodeji Irete Fasiku

Wireless network-on-chip (WiNoC) interconnection is one of the proposed techniques to enhance performance efficiency and overcome the limitations of the multi-hop nature of conventional network-on-chips (NoCs). WiNoC uses long-range, single-hop wireless links to connect distant cores that reduce multi-hop communication in conventional NoC. Fair and efficient medium access control (MAC) is critical to enhancing WiNoC performance. We propose a centralized token-based MAC (CMAC) mechanism to coordinate RHs activities in each maximum hold cycle (MHC). The proposed CMAC allocates tokens to the RH with the most packets and ensures fair sharing of access to the wireless channel. We also propose a load-balance congestion-aware routing algorithm for WiNoC (LCRAW) that dynamically load-balance packet distribution for the lower layer (wired) and the upper layer (wireless) in a WiNoC architecture. The lower wired

mesh network is partitioned into small subnets. The source-to-destination distance is used to decide the most optimized transmission path. Moreover, the proposed method also checks if the receiver data buffer (RDB) is sufficient to receive all packets transmitted. The proposed work was modeled in a Noxim cycle-accurate network simulator that shows that CMAC has higher network throughput of 31% and 19%, improves the latency by 20% and 10% and also save energy by 19% and 13% when compared with the baseline and a related radio access control mechanism (RACM) MAC, respectively. The proposed LCRAW with CMAC when compared with the XY-routing with CMAC improves the throughput, latency, and energy by 14%, 20%, and 18%, respectively. The proposed techniques improve the overall system performance and ensure maximum utilization of wireless resources in the WiNoC.

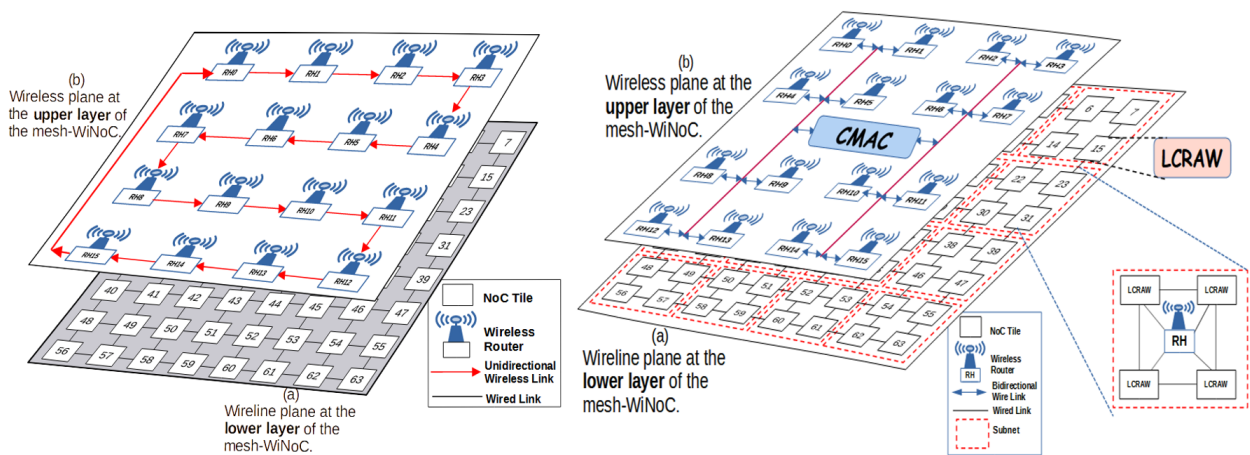


Figure 7: (a) WiNoC architecture (b) Architecture of the proposed WiNoC design.

More info:

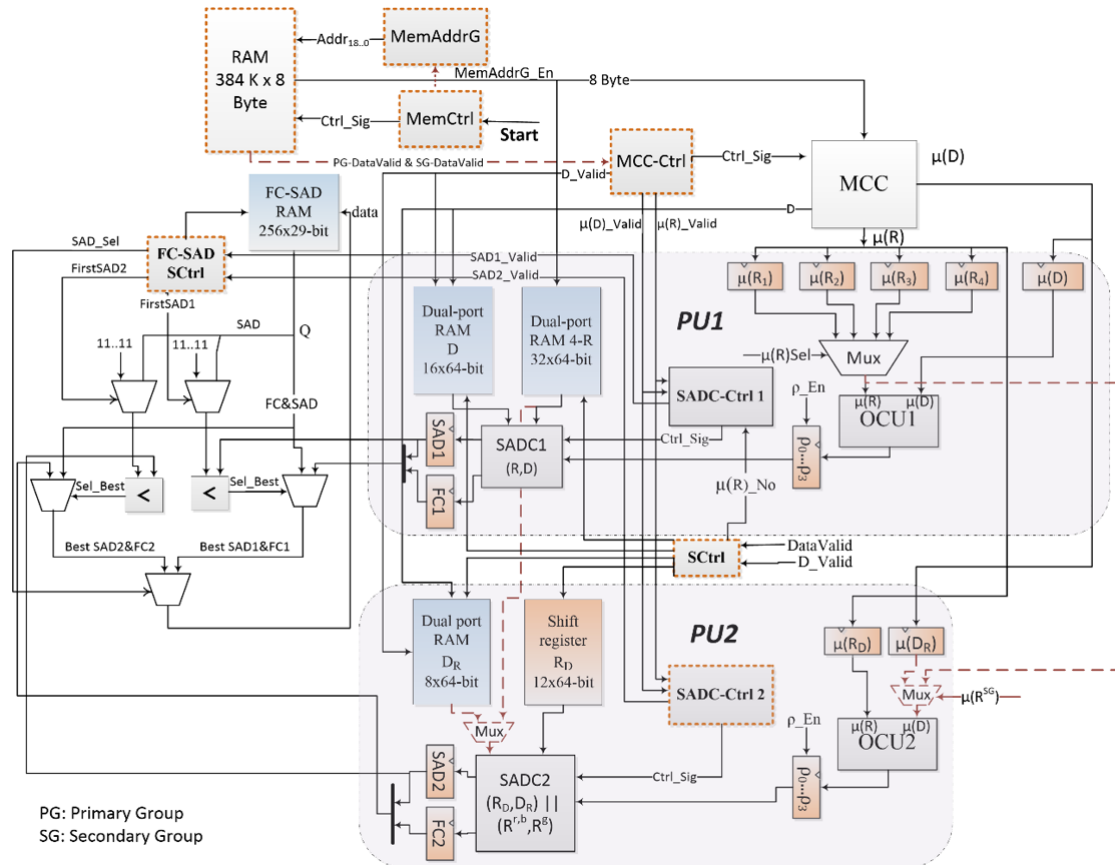
1. Fasiku, A. I., Oladokun, O., Rusli, S., & Marsono, M. N. (2021, June). A Centralized Token-based Medium Access Control Mechanism for Wireless Network-on-Chip. In 2021 International Conference on Artificial Intelligence and Computer Science Technology (ICAICST) (pp. 102-107). IEEE.
2. Fasiku, A. I., Rusli, S., & Marsono, M. N. B. (2020, September). Characterization of subnets, virtual channel and routing on wireless network-on-chip performance. In 2020 IEEE Student Conference on Research and Development (SCoReD) (pp. 117-121). IEEE.
3. Fasiku, A. I., Marsono, M. N. B., Numan, P. E., Lit, A., & Rusli, S. (2019). Wireless Network On-Chips History-Based Traffic Prediction for Token Flow Control and Allocation. ELEKTRIKA-Journal of Electrical Engineering, 18(3), 21-26.

Deep Pipeline Architecture for Fast Fractal Color Image Compression Utilizing Inter-color Correlation

Abdul-Malik H. Y. Saad

Fractal compression technique is a well-known technique that encodes an image by mapping the image into itself and this requires performing a massive and repetitive search. Thus, the encoding time is too long, which is the main problem of the fractal algorithm. To reduce the encoding time, several hardware implementations have been developed. However, they are generally developed for grayscale images, and using them to encode colour images leads to doubling the encoding time 3x at least. Therefore, in this paper, new high-speed hardware architecture is proposed for encoding RGB images in a short time. Unlike the conventional approach of encoding the colour components similarly and individually as a grayscale image, the proposed method encodes two of the colour components by mapping them di-

rectly to the most correlated component with a search-less encoding scheme, while the third component is encoded with a search-based scheme. This results in reducing the encoding time and also in increasing the compression rate. The parallel and deep-pipelining approaches have been utilized to improve the processing time significantly. Furthermore, to reduce the memory access to the half, the image is partitioned in such a way that half of the matching operations utilize the same data fetched for processing the other half of the matching operations. Consequently, the proposed architecture can encode a 1024x1024 RGB image within a minimal time of 12.2 ms, and a compression ratio of 46.5. Accordingly, the proposed architecture is further superior to the state-of-the-art architectures.

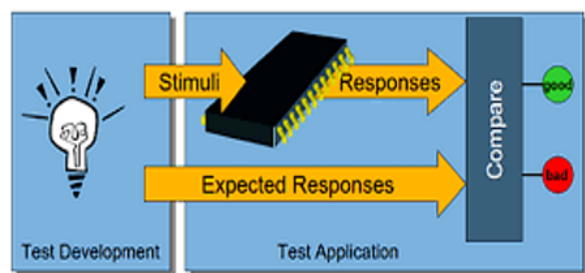
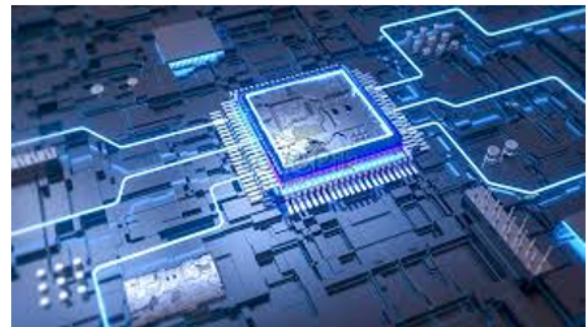


Test Point Insertion for Testability Improvement

Norlina Paraman

Test point insertion is an important design for testability technique used in the design and testing of Application-Specific Integration Circuits (ASIC). It involves the strategic placement of additional circuitry within the ASIC to enable observation and control of specific internal signals during the testing phase. Test points are essential for detecting faults and verifying correct functionality in ASICs, particularly in complex designs with many interconnected components. However, the insertion of test points can increase the size and cost of the ASIC, as well as impact its performance

and power consumption. Therefore, careful consideration and optimization are required to balance the benefits of test point insertion with the associated costs and design trade-offs. The positioning of the test point is a critical aspect of this project, as it needs to be placed strategically to provide the least possible visibility into the internal signals of the ASIC, while also making it easier to apply test patterns for fault detection. Reducing the number of test points can lead to lower power consumption and reduced area overhead, as excessive circuitry is avoided.



More info:

Stochastic Computing for Low Area and Low Power Applications

Izam Kamisian

Conventional binary computing (BC) is so far promising highly efficient, high quality, high accuracy and high-speed solutions. However, as the world now sees increasing need in mobile and embedded, edge computing, near-sensor computing and IoT, conventional BC faces difficulty in achieving the highly efficient, high quality, high accuracy and high-speed performances because of stringent low power, low energy, low area and error-tolerant requirements. Stochastic computing (SC) is a computing paradigm, which was first introduced by Gaines in the 1960s, as an alternative to the conventional binary computing (BC) technique. Stochastic computing (SC) is promising for lossy applications such as image processing, neural networks and filters and has been targeted to low energy, small size and high reliability applications. Stochastic computing requires very small area footprints and tolerate errors compared to conventional binary computing. Stochastic computing represents data as randomized serial bit streams that greatly reduce hardware complexity and power. The key trade-off is the bit stream length (B)

that affects stochastic computation accuracy (A) at the cost of longer latency (L) and energy consumption (E). The SC system consists of three main blocks which are Stochastic Number generator (SNG), Stochastic Operations and Stochastic to Binary Converter as shown in Figure 1(a). SNG converts conventional BC number (BN) to stochastic number (SN) as shown in Figure 1(b). SC block is the core block where the stochastic computation or processing are done. Stochastic to Binary Converter converts SN back to BN as shown in Figure 1(c). Research in SC can be divided into 3 main areas as shown in Figure 1 which are SNG, SC functions and Configurable SC. The main goal in SC research is to find optimal points depending on requirements based on the SC control knobs and A-L-E tradeoff. The trade-off provides SC with properties of variable precision, variable accuracy, variable latency and variable energy which can be configured easily by selecting the optimal point of the control knobs. However, this needs a lot of characterizations of the control knobs and A-L-E trade-off.

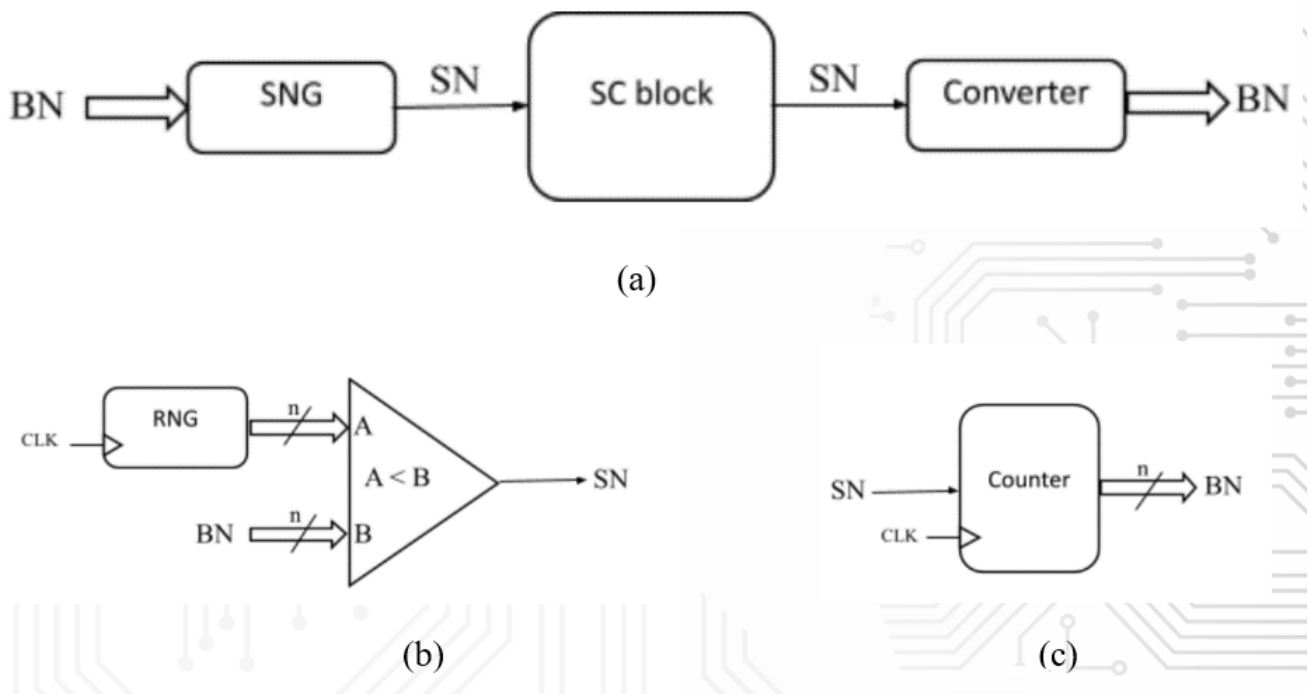


Figure 9: Stochastic computing functional block.

More info:

N-Gram Feature Extraction and Naïve Bayes Classifier for Malware Detection using FPGA Implementation

Lee Ming Yi and Ismahani Ismail

Malicious software, or malware as it is more generally called, now plays a crucial part in practically every network intrusion attack that aims to destroy the linked devices. Installing malware detection solutions is now much more essential in order to safeguard the network environment. The Naive Bayes classifier is a probabilistic supervised machine learning technique that may be used to tackle a variety of classification issues, including malware detection, on the majority of general-purpose machines. A competent feature extractor is crucial in addition to the classifier to boost the accuracy and dependability of the classifier model. The processing throughput of general-purpose devices is constrained for real-time applications. This research investigates the hardware implementation of the n-gram feature extractor and Naive Bayes classifier using

a field-programmable gate array (FPGA). By designing many processing units for the inference module to be implemented on the hardware, the parallel processing power of FPGA has been used to increase the throughput and latency of the malware detection process. The inference module is additionally created to be pipelined in six phases. In addition, this research employs hardware-friendly methods that implement base 2 logarithm transformation and floating-point to fixed-point conversion. On the test dataset, both software and hardware designs achieved an accuracy of 99.18%. Additionally, it is discovered that this design's higher n of parallel processing units results in higher malware detection throughput, resource utilization, power consumption, and energy efficiency.

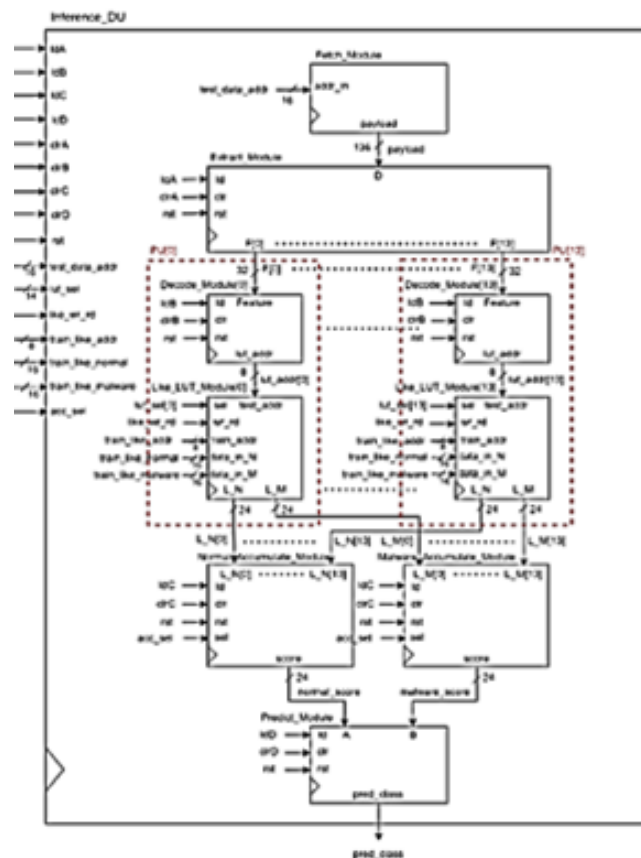


Figure 10: Datapath unit of the inference module.

Image Recognition using Capsule Network on FPGA

Salim Ali Abdulraziq Adrees and Ismahani Ismail

A new approach in artificial neural networks (ANN) called a capsule neural network (CapsNet) creates a better model hierarchical relationship. A capsule is performed by a collection of neurons. Each capsule produces a vector that displays an entity's details. On the Modified National Institute of Standards and Technology database (MNIST), CapsNet's performance on a graphics processing unit (GPU) has advanced to state-of-the-art levels, outperforming convolutional neural networks (CNN) at identifying highly overlapping numbers in images. To evaluate the speedup performance and compare it to the GPU, a CapsNet accelerator on a field-programmable gate array (FPGA) is investigated. High-level synthesis (HLS) is used in this study's de-

sign of the CapsNet model (accelerator) on an FPGA. Then, the speedup and accuracy of the performance of the FPGA and GPU are compared. The MNIST dataset is used to analyze and validate the behavioral module once it has been synthesized using HLS tools on an FPGA. The module is made to accept handwritten digits' feature vectors as input and process them via a number of layers to anticipate the result. Although FPGA accuracy is anticipated to be slightly lower than GPU, FPGA speed-up performance is anticipated to be higher than GPU. The module can be helpful in a variety of situations, including spotting license plates on moving cars.

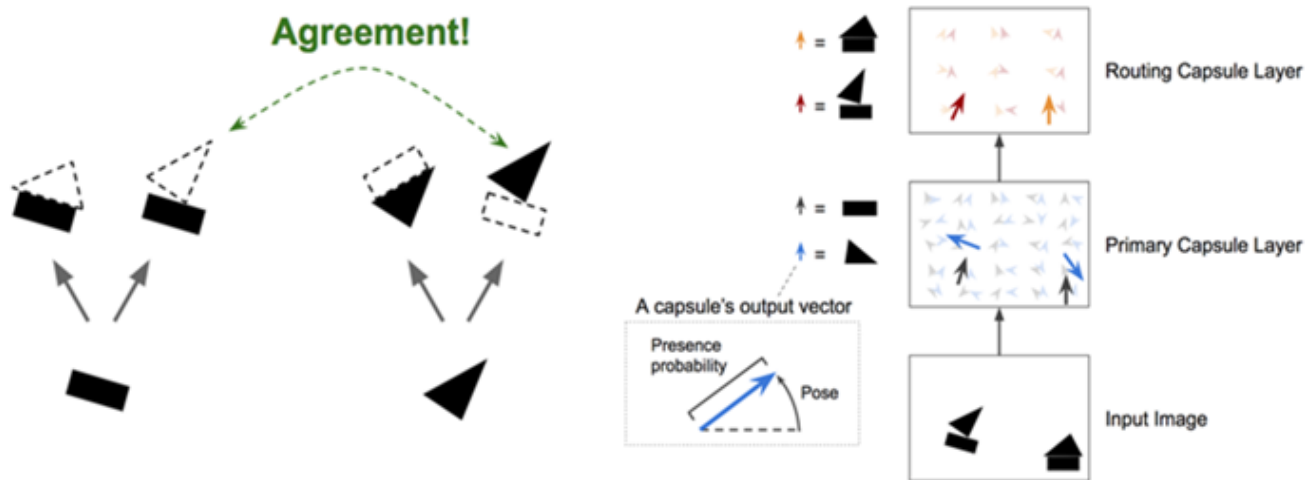


Figure 11: General idea of CapsNet.

High Performance Programmable Logic Controller Processor using System on Chip of Ladder Rung Processor and RISC CPU Core

Zulfakar Aspar

Most Programmable Logic Controllers (PLCs) in the market are microprocessor based systems. They are inefficient where it is only possible to get ms to μ s cyclic scan speed depending on the operating frequency. The cyclic scan depends on the number of step inputs, complexity of the input networks, number of rung's output and custom functions. A Ladder Rung Processor (LRP) was invented by the researcher to speed up Ladder Logic Diagram (LLD) model computation. It is a hybrid processor where a boolean equation or a rung is processed in parallel. While all the rungs are solved sequentially in a cyclic scan. So the cyclic scan depends on the number of rungs instead of input steps. LRP is proven run on a Field Programmable Gate Array (FPGA) at 5 MHz operating frequency and the cyclic scan is at 12.64 μ s which is very fast compared to the most PLCs in the industry. A pre-silicon IC design layout is done at 130 nm, and the simulation shows that the LRP can be operated up to 2 GHz while producing 31.6 ns cyclic scan. The computation core of the proposed LRP architecture as in Figure 1 is the Ladder Rung Block (LRB). The LRB module performs the time-critical ladder logic solving process. The Ladder Rung Configurator (LRC) holds the configuration data that is used to realize the LRB logic as defined by the LLD. At any one time, only a

single rung is realized on the LRB. Since, a LLD model typically consists of multiple rungs, the configuration data defining the different rungs are multiplexed, one rung at a time, into the LRB, in the sequential order of the scan. The Program Counter keeps track of the configuration to be mapped next. Essentially, the ladder logic network is solved in a cycle, and then the network is reconfigured for the next cycle, which is then solved. This process is repeated until the last rung is traversed. In addition to all basic peripherals as shown in Figure 1, the LRP is also built with the essential peripherals such as PWM, comparator, encoder etc. to complete a PLC as used in the PLC on FPGA board on the left of Figure 1. Look Up Table (LUT) can be used to speed up the process. These dedicated data processing are essential, but they are expensive especially in multi-axes operation where more than a single set of data processing peripherals are needed. To reduce the cost, a RISC CPU is integrated to do similar data processing operations with a small overhead processing speed penalty. By combining two different processors, slower speed data processing can be done by the RISC CPU while higher speed operations can be done by the dedicated data processing peripherals.

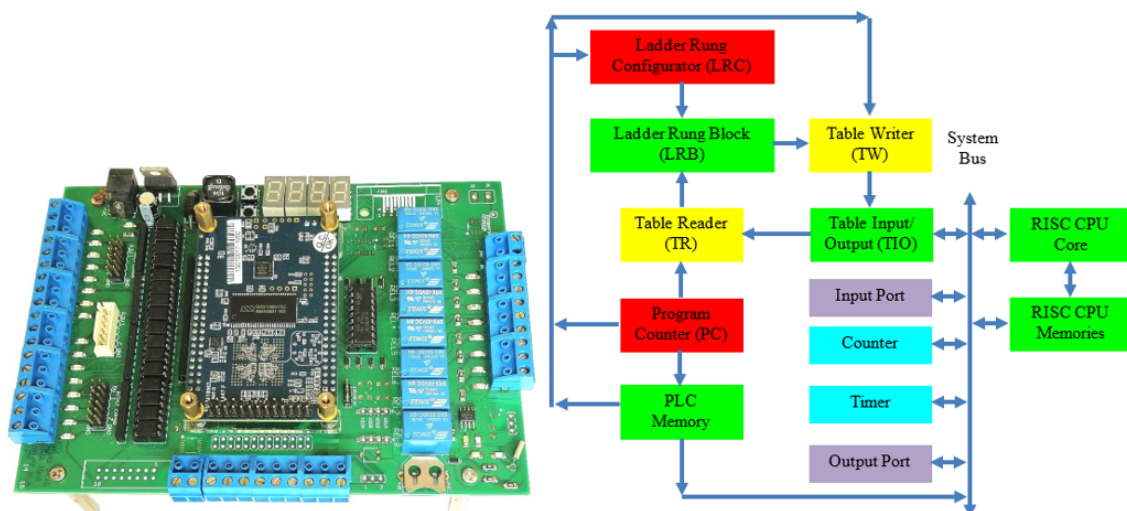


Figure 12: PLC on FPGA on the left using PLC processor architecture on the right.