UTM
UNIVERSITI TEKNOLOGI MALAYSIA

SCHOOL OF ELECTRICAL ENGINEERING
FACULTY OF ENGINEERING

VeCAD

VLSI-Embedded Computing Architecture Design

VeCAD Annual Research Report 2021

Integrated Circuits and Systems

Foreward

We are pleased to bring to you the 2021 the VLSI and Embedded Computing Architecture Design (VeCAD) Annual Research Report. The core mission of the VeCAD rsearch group is to foster research and teaching in diverse field in electronic systems and computer engineering, particularly in domain-specific computing architecture, VLSI/SoC design, and embedded systems.

On behalf VeCAD community, we thank every contributor to this year's VeCAD Annual Research Report who help to produce such a beautiful volume.

Muhammad Nadzir Marsono
Head of VeCAD RG
October 2021

Contents

Fast Parallel Two-Dimensional HEVC Transform Without Transpose Memory

Ab Al-Hadi Ab Rahman, Ainy Haziyah Awab, Izam Kamisian

There is currently a huge demand for high resolution and frame rate video as we move towards 4K TV and beyond. The major challenge is to support these requirements in real-time using the latest video compression standard, the High-Efficiency Video Coding (HEVC).

This project presents a hardware accelerator of a HEVC transform that features a highly parallel two-dimensional design, with a structure that does not require a transpose memory. It implements the full HEVC transform specification, i.e. for Discrete Sine Transform (DST) of size N=4, and the Discrete Cosine Transform (DCT) of size N={4, 8, 16, 32}. The inputs are sent to the units in parallel whereby a latency of 1 can be achieved for all transform unit types. This is in contrast to the conventional folded structure that requires a transpose memory between the first and second dimension of a NxN DST/DCT. In the proposed parallel structure, the transpose memory is eliminated by instantiating a second NxN DST/DCT unit, and perform a direct connection from the first to the second units such that it performs the matrix transposition function. In order to speed up the N-point DST/DCTs, the butterfly algorithm has been used to take advantage of the symmetrical property of the transform matrices. Split architecture is also proposed that allows parallel executions of different size transforms.

The combinations of all of these architectural optimizations result in a improvement of up to 33% and 300% in throughput and total cell count, respectively, as compared to the conventional folded structure, when implemented on SAED 32nm ASIC technology.

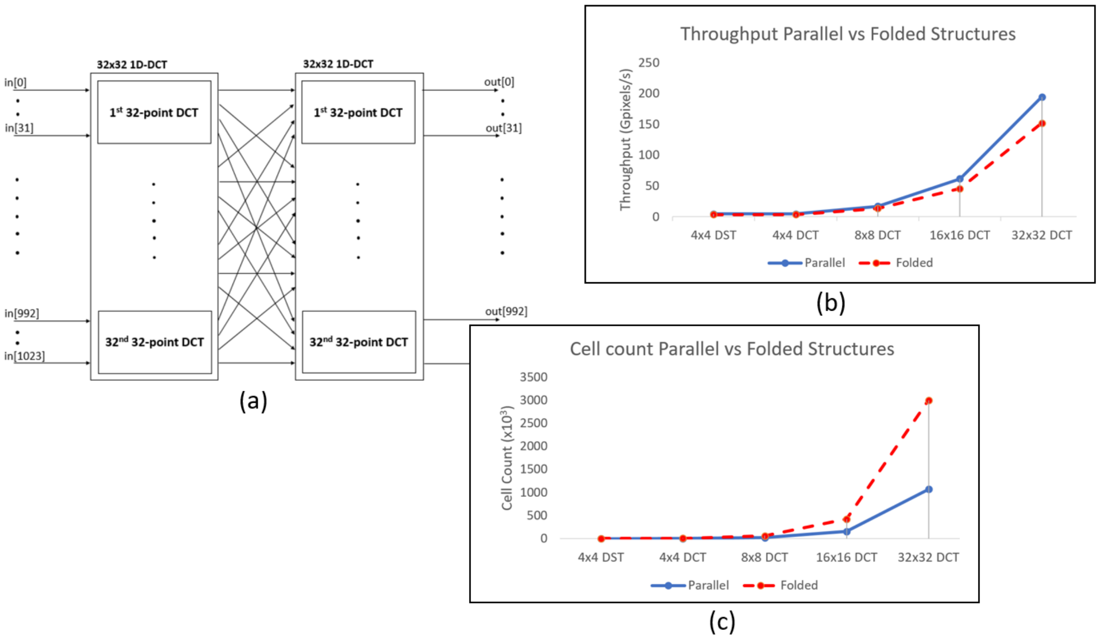


Figure 1: (a) System architecture, (b) throughput analysis, and (c) cell count analysis.

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.

Low Power Phase-domain ADC Based Demodulator for Bluetooh Low Energy

Ab Al-Hadi Ab Rahman, Ung Shen Jie, Norlina Paraman

Internet-of-Things (IoT) have driven the demand of wireless standard for short distance data exchange while using little energy. Bluetooth Low Energy (BLE) standard was developed for such operation, intended for small battery powered devices that could last for months. It is essential that the components of BLE, from the RF front-end to the application level, exhibit low power characteristics.

This project presents a low-power fully-digital implementation of the BLE demodulator. It utilizes the Gaussian Frequency Shift Keying (GFSK) type modulation operating at 2.4 GHz ISM radio band with a data rate of 1 Mbits/s. We have proposed to implement the Phase-Domain Analog to Digital Converter (Ph-ADC) based demodulator architecture due to its ability to handle both Zero-IF and Low-IF signals, as well as having a simple structure with very low power characteristics. Furthermore, it has also been shown to provide higher immunity to PVT variation, achieves higher accuracy, and consumes less space and power as compared to other structures such as the Amplitude ADC (AADC) or Limiter based demodulators. The proposed design includes noise filtering, preamble detection, parallel symbol synchronization and frequency offset compensation, as well as bit slicing and frame synchronization. The design conforms to the BLE standard and supports input bit size of 4- and 5-bits with sampling frequency range from 4 MHz to 16 MHz.

The design has been implemented in Silterra 180nm technology, and achieved best current consumption of only 50 μA, which translates to around 90 μW of power. This is significantly less compared to other reported works. Total chip layout area has also been found to be one of the smallest in literature, with only 0.049 mm^2.

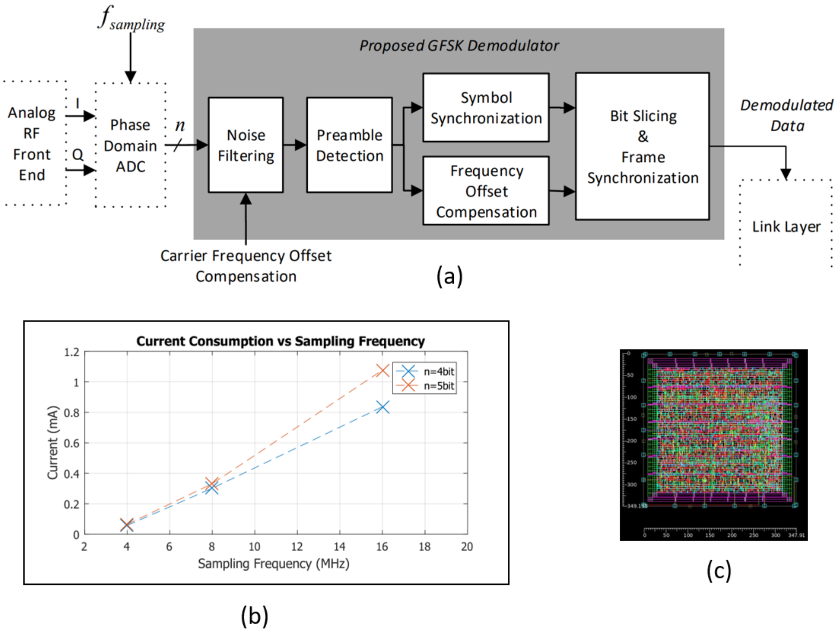


Figure 2: (a) System architecture, (b) Current consumption analysis, (c) Chip layout

More info:

1. SJ Ung, AAH Ab Rahman, A 88 μW digital phase-domain GFSK demodulator compatible with low-IF and zero-IF receiver with preamble detection for BLE, Turkish Journal of Electrical Engineering & Computer Sciences, 28(4), pp 2183-2199, 2020.

Higher PLC modeling abstraction using SIPN while reusing existing LLD platform

Zulfakar Aspar

Ladder Logic Diagram (LLD) is widely being used to model a Programmable Logic Controller (PLC) based design. However, designs based on PLC are getting larger and complicated, many engineers resort to structured text programming such as C language. However, programming in a high level language are only adding design complexity, rising the need to have highly skills programmers, and more expensive tools.

This research is going to use Signal Interpreted Petri Net (SIPN) to model a PLC design at a higher level of abstraction than using LLD. Using higher level of abstraction, a PLC user does not need to worry about the detail logic of the model. Instead, the work can be focused on the flow and control of the system. SIPN is so powerful that in addition to model a PLC design graphically, it can also be done mathematically. Using mathematical model, a SIPN model can quickly being checked by a computer program for any design error, to suggest design improvement, to do model conversion from SIPN to LLD and much more. In addition, existing tools which are using LLD design entry and hardware implementation can still be reused.

The research team has also successfully designed a processor to speed up a LLD model execution known as Ladder Rung Processor (LRP). Most PLCs in the market are limited by the cyclic scan speed which is usually in millisecond in low and medium price PLCs which are usually based on general purpose microprocessors. To have a microsecond or faster cyclic scan in a large and complex system, system developers resort to distributed computing system (DCS) which is more expensive, more manpower and longer development time.

LRP was successfully implemented on an FPGA at 5 MHz operating speed with 12.64 us cyclic scan. The same architecture was implemented as an ASIC to achieve 2 GHz operating speed for estimated 31.6 ns cyclic scan. The high speed achievement has open the processor to be integrated with data processors for a complete high performance PLC processor.

The advantage of this research direction, existing PLC users can still use their knowledge on LLD modeling to use the new PLC processor. While a new knowledge on SIPN can be used to save their development time, easier to maintain the program and can verify the design in LLD while maintaining PLC robustness. Figure 3 shows an example of LLD and SIPN model equivalency for the same PLC hardware.

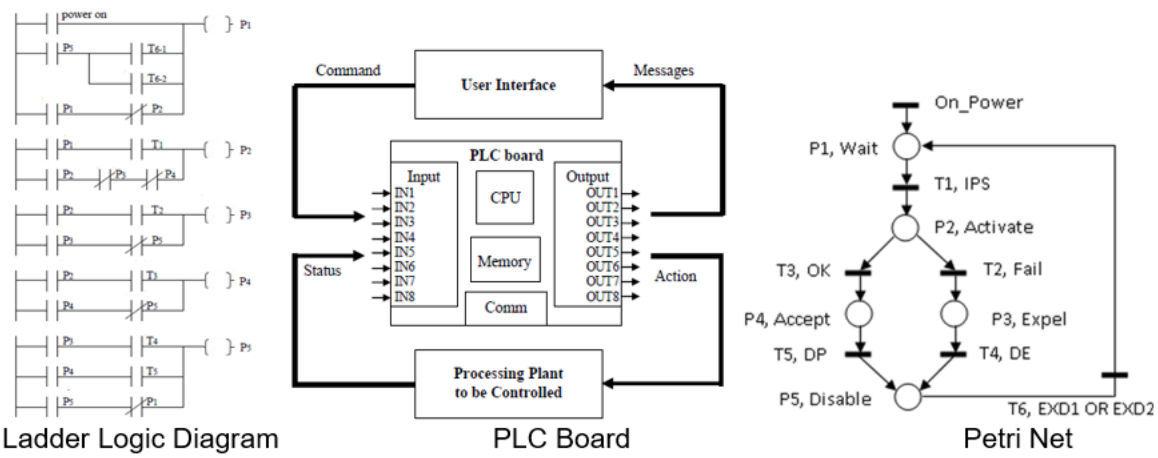


Figure 3: A LLD model equivalence to a SIPN model for the same PLC design.

High-Speed Fractal Image Compression Featuring Deep Data Pipelining Strategy

Abdul-Malik H. Y. Saad

Fractal image compression (FIC) is a very popular coding technique used in image/video applications due to its simplicity and superior performance. The major drawback with FIC is that it is very time consuming algorithm, especially when a full search is attempted. Hence, it is very challenging to achieve a real-time operation if this algorithm is implemented on general processors.

To overcome the speed problem, a new architecture based on deep data pipelining is proposed for coding high-resolution grayscale images in real-time. The general idea is to partition an image into overlapping range and domain blocks in which four range blocks constitutes one domain block. In this way, two matching operations can be performed simultaneously using two processor units (PUs). Further reduction in the encoding time is achieved by exploiting the inherently high degree of correlation among pixels in the neighborhood areas and restricting the search in the neighboring blocks only. The design is synthesized on Altera Stratix IV FPGA and optimized at circuit level in order to achieve a high-speed implementation.

The proposed architecture has been evaluated in terms of the peak signal-to-noise ratio (PSNR), the runtime, the memory utilization, and the compression ratio (CR). Experimental results suggest that the proposed architecture is able to encode a 1024×1024 size image in 10.8 ms with PSNR and CR averaging at 27 dB and 34:1 respectively. Meanwhile the energy dissipation is approximately 0.5 watt which is comparable to the state-of-the-art fractal processors.

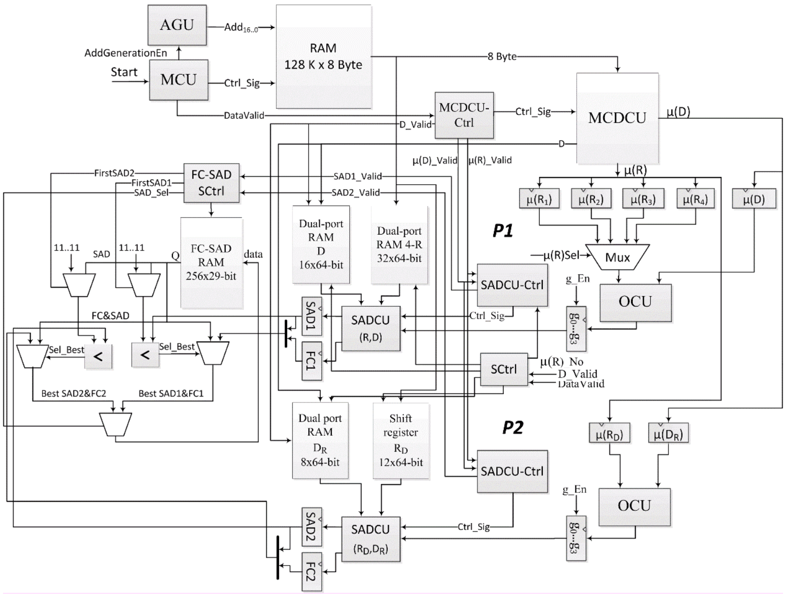


Figure 4: Overall hardware architecture of the proposed design.

More info:

1. Saad, Abdul-Malik HY, and Mohd Z. Abdullah. "High-Speed Fractal Image Compression Featuring Deep Data Pipelining Strategy." IEEE Access 6 (2018)

Network-on-Chip based Manycore System-on-Chip

Muhammad Nadzir Marsono, Nasir Shaikh-Husin, Ab Al-Hadi Ab Rahman, Norlina Paraman, Alireza Monemi, Tang Jia Wei, Mohamed Sultan Mohammed

Domain-specific computing architectures are massive and heterogeneous. Network-on-chip (NoC) interconnection has emerged as a viable alternative to traditional bus-based interconnections for manycore system-on-chip (MCSoC).

We develop Prototype-NoC (ProNoC) as an open-source integrated tool that automates design and validation of a complex NoC-based MCSoC. ProNoC supports rapid prototyping and validation of NoC-based MCSoC projects targeting FPGA devices. Our improved NoC architecture includes a delay-optimized arbiter, a low latency configurable router that combines virtual channel and switch allocations, improved fully adaptive routing strategy via speculative allocation.

We propose an effective application mapping strategy for NoC-based MCSoC to estimate the cycle-accurate application behaviour. Task-to-task communication delay model is proposed to analytically evaluate application mapping performance. An improved Genetic Algorithm (GA) technique is also proposed to speed up the mapping exploration process.

Rapid prototyping for design space exploration is not trivial as tasks have to be developed ground-up. We propose fast NoC-based MCSoC application prototyping called CAL2NoC by converting a high-level dataflow model of real-world applications to multiple executable C codes (tasks). CAL2NoC facilitates prototype-based design space exploration for real-world applications using emulated MPSoC platform for practical performance evaluation.

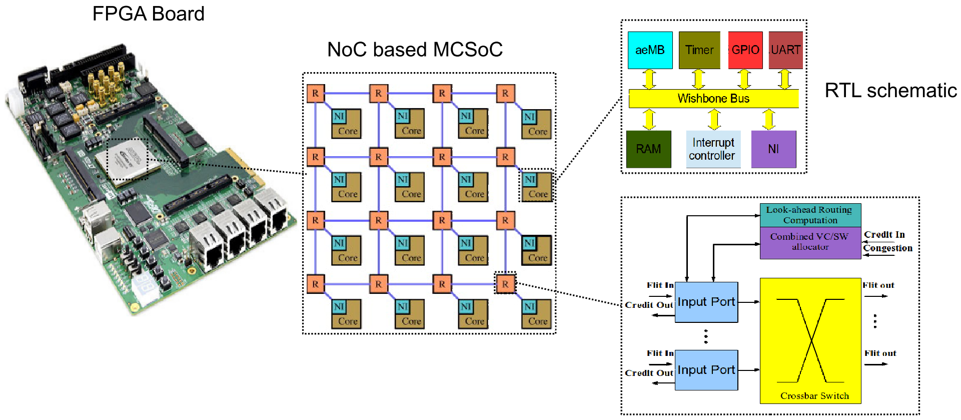


Figure 5: ProNoC: A low latency network-on-chip based many-core system-on-chip prototyping.

More info:

1. A. Monemi, J.-W. Tang, M. Palesi, M.N. Marsono, "ProNoC: A Low Latency Network-on-Chip based Many-Core System-on-Chip Prototyping Platform", Microprocessors and Microsystems 54, pp.60–74, 2017.
2. J.W Tang, Y.W Hau, N. Shaikh-Husin, M.N. Marsono, "Application Profiling and Mapping on NoC-based MPSoC Emulation Platform on Reconfigurable Logic", TELKOMNIKA 15(3), pp. 1040–1047, 2017.
3. M.S. Mohammed, J.W. Tang, A.A.-H. Ab Rahman, N. Paraman, M.N. Marsono, "Rapid Prototyping of NoC-based MPSoC Based on Dataflow Modeling of Real-World Applications", in ICSGRC2018, 2018, pp. 217-222.

Software-Defined Network Traffic Engineering

Muhammad Nadzir Marsono, Shahidatul Sadiah Abdul Manan, Mosab Hamdan, Usman Humayun, Hala Suliman, Azza Abdelkarim, Entisar Hassan

Traffic management in software-defined networks (SDNs) is critical for efficient bandwidth utilization and resource provisioning. Recent works on SDN load balancing (LB) have focused on identifying and rerouting elephant flows (EFs) for effective bandwidth usage.

We propose a flow-aware elephant flow detection applied to SDN. The proposed technique employs two classifiers, each respectively on SDN switches and controller, to achieve accurate elephant flow detection efficiently. Moreover, this technique allows sharing the elephant flow classification tasks between the controller and switches. Hence, most mice flows can be filtered in the switches, thus avoiding the need to send large numbers of classification requests and signaling messages to the controller.

An ant colony optimization-based technique is proposed for rerouting EFs while considering load-balancing in the SDN links. It obtains the global state of the SDN from which the most optimal paths for congested links are retrieved, and EF are redirected accordingly. The overall performance indicates that the proposed LB technique based on detection and rerouting of EFs can improve SDN's overall performance.

We also propose a classifier ensemble to accurately detect DDoS attacks. The mitigation module blocks malicious traffics and purges entries of malicious traffic from the switch flow table. The collaborative module shares DDoS detection and mitigation rules among multiple SDN controllers.

We are currently working on SDN flow prediction based on time-series analytics, service function chaining, and flow-table optimization, and multi-controller task migration.

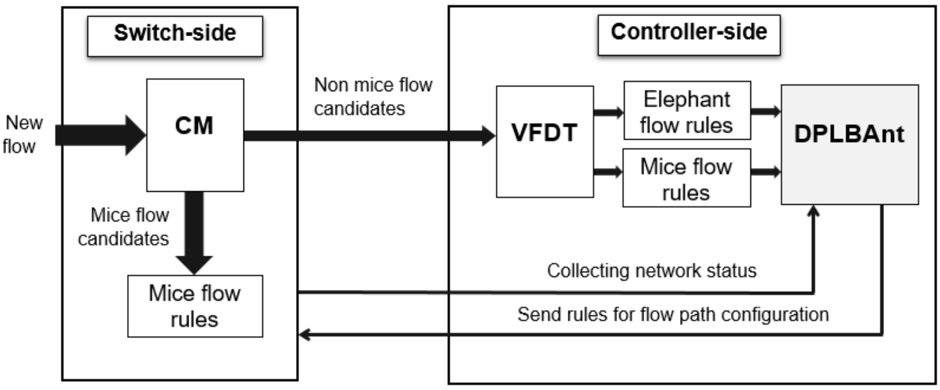


Figure 6: Flow-aware elephant flow detection and rerouting for software-defined networks.

More info:

1. M. Hamdan, B. Mohammed, U. Humayun, A. Abdelaziz, S. Khan, M. A. Ali, M. Imran, M. N. Marsono, "Flow-Aware Elephant Flow Detection for Software-Defined Networks", IEEE Access 8, pp. 72585–72597, 2020.
2. O.E. Tayfour, M.N. Marsono, "Collaborative Detection and Mitigation of DDoS in Software-Defined Networks", The Journal of Supercomputing, preprint, April 2021.

Stream Analytics for Fog Computing

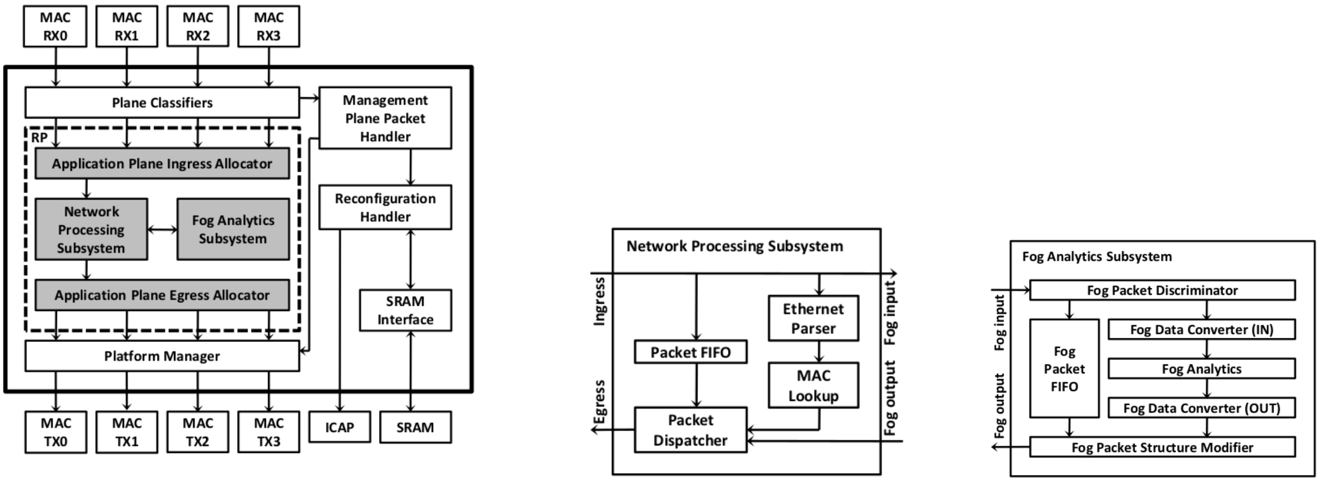
Muhammad Nadzir Marsono, Ooi Chia Yee, Nasir Shaikh-Husin, Tan Tze Hon, Loo Hui Ru, Jeevan Sirkunan

Field-programmable gate array (FPGA) is a versatile compute platform for fog analytics that can provide balanced processing throughput and architecturalflexibility. It can be dynamically program to adapt to application dynamicity.

We propose a standalone FPGA-based fog node architecture that allows remote functional update without service interruption. Network processing is non-reconfigurable while the analytics subsystem can be dynamically reconfigured. The Windowed Gaussian and KNN CAD are implemented as fog-based analytics case studies and both exhibit high energy efficiency, low latency, and high throughput.

We also propose an online incremental semi-supervised algorithm based on incremental k-means that continuously update its model. A hardware classifier is incorporated in NetFPGA reference switch design. The proposed architecture can perform online classification at 1Gbps line speed without any flow loss.

We also propose an interleaved incremental/decremental support vector machine (IIDSVM) that performs simultaneous learning and unlearning. The IIDSVM is optimized to reduce the division operations and kernel is hardware accelerated for improved performance.



(a) Fog node architecture

(b) Network processing subsystem

(c) Fog analytics subsystem

Figure 7: FPGA-based fog node analytics with service-uninterrupted remote functional update.

More info:

1. TH Tan, CY Ooi and MN Marsono, ”An FPGA-Based Network System with Service-Uninterrupted Remote Functional Update”, International Journal of Electrical and Computer Engineering 11(4), pp. 3222–3228, 2021.
2. HR Loo, SB Joseph, MN Marsono, “Online Network Traffic Classification with Incremental Learning”, Evolving System 7(2), pp.129–143, 2016.
3. J Sirkunan , N Shaikh-Husin, MN Marsono, “Interleaved Incremental and Decremental Support Vector Machine for Embedded System”, in ISCAS 2019, 2019, pp. 1–5.

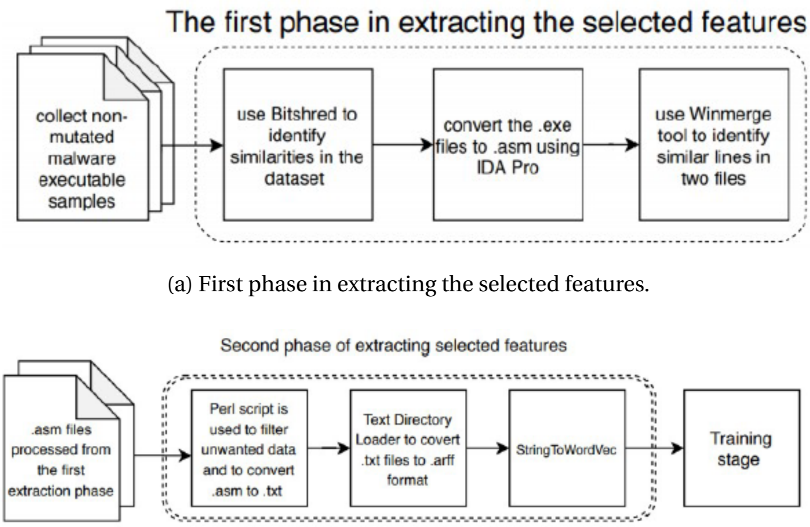
Metamorphic Malware Detection using Machine Learning

Ismahani Ismail and Mohammed Hasan Ali Ahmed Ali

Hackers make advantage of weaknesses in operating system architectures, web browsers, and online services, or exploit social engineering methods to motivate individuals to execute malicious software for the purpose of spreading them. Hackers utilize obfuscation methods such as reassigning registers, insertion of inexecutable code, reordering of the subroutine, transposition of code, integration of code, and substitution of instructions to avoid to be detected by conventional lines of defense such as gateways, firewalls, anti-virus software which relies on signature-based methods. This method demands the anti-virus vendors to supply a record of signatures of already analyzed malware samples to be then compared to possible attacks. Thus, commercially available anti-virus programs are incapable to detect formerly unobserved malevolent executable.

Techniques motivated by machine learning algorithms is categorized under non-signature-based malicious software detection techniques. This type of technique is capable of recognizing patterns found in the malicious software which will be utilized as a feature in order to create a generative classifier model for the purpose of malware detection.

In order to reduce the time and space complexity of the development of the machine learning classifier models, we propose similar code segments in the non-mutated malware samples to be identified through 2 phases of feature extraction process. The first phase to extract the informative textual strings from the non-mutated malware samples corpus with similar code. Then, the processed data files which in .asm format that contain undesirable contents such as automated comment section, unexplored code section, data section, white spaces, and empty lines is automated cleaned using Perl script and then convert the files extension to .txt to prepare the data to be analyzed using WEKA.


The first phase in extracting the selected features
collect non-mutated malware executable samples
use Bitshred to identify similarities in the dataset
convert the .exe files to .asm using IDA Pro
use Winmerge tool to identify similar lines in two files
Second phase of extracting selected features
.asm files processed from the first extraction phase
Perl script is used to filter unwanted data and to convert .asm to .txt
Text Directory Loader to covert .txt files to .arff format
StringToWordVec
Training stage

(a) First phase in extracting the selected features.

(b) Second phase in extracting the selected feature.

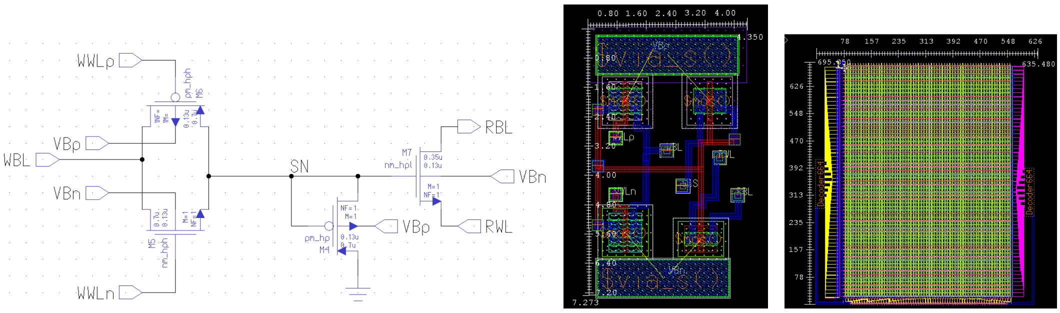
Figure 8: Feature extraction phases.

Implementation of Monolithic Gain-Cell eDRAM (GC-eDRAM) Memory Array in 130nm

Afiq Hamzah, Shi Rong Soo, Nurul Ezaila Alias and Izam Kamisian

The SRAM cell is the traditional choice to be implemented as cache memory owing to its high-speed write/read (W/E) operation. However, it suffers from high static power consumption owing to high leakage current in sub-micron technology. Gain-cell (GC) embedded dynamic random-access memory (eDRAM) provides high density, low leakage power, small size, and two-port functionality.

We develop the 4kb array of GC-eDRAMs namely the single-supply three-transistor (3T) GC-eDRAM and four-transistor (4T) GC-eDRAM with internal feedback. 3T GC-eDRAM is fully functional with 1V supply voltage. It has worst-case data retention time (DRT) of 9.2 µs and in 1.51 W/4 Kb total retention power. 4T GC-eDRAM with internal feedback is fully operated with an optimum supply voltage of 0.95 V and boosted voltage of 1.1 V, and exhibits a DRT of 25.27 µs and 0.597 W/4 Kb total retention power. The total power consumed by the GCs is approximately 20×–50× lower than the power consumed by conventional 6T SRAM.

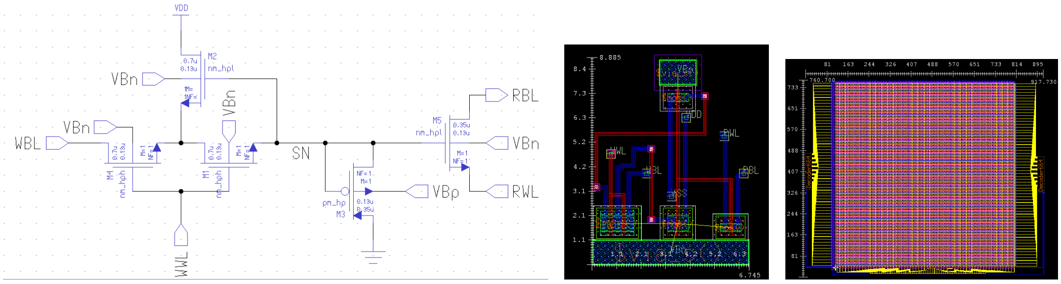


(a) Bit-cell schematic

(b) Bit-cell layout

(c) Memory array

Figure 9: Single-supply 3T GC-eDRAM



(a) Bit-cell schematic

(b) Bit-cell layout

(c) Memory array

Figure 10: 4T GC-eDRAM with internal feedback

More info:

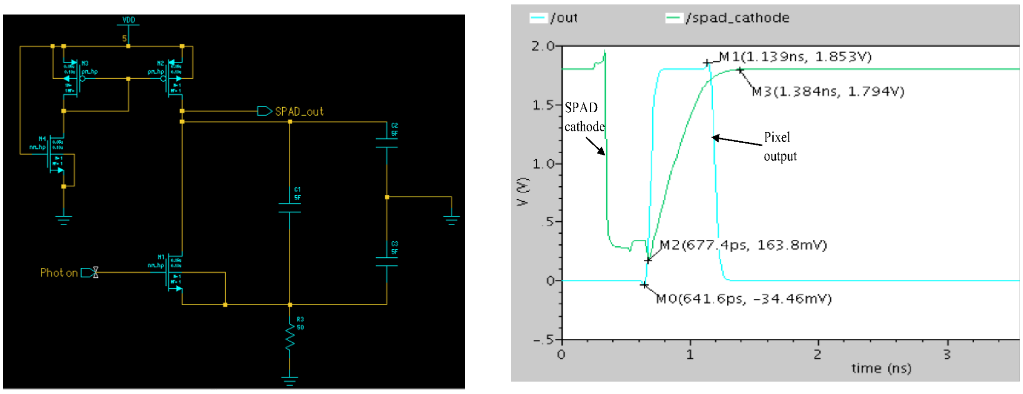
1. Hussien Abdelrauf Hussien Abdo, Nurul Ezaila Alias, Afiq Hamzah, Izam Kamisian, Michael Loong Peng Tan and Usman Ullah, “A 2 Kbit Memory Array of Mixed-VT 3T GC-eDRAM Implemented in 130nm Standard CMOS Technology”, in RSM 2021, 2019, pp. 1–4.

An improved Single Photon Avalanche Diode Simulation Model Using Low Voltage CMOS Technology

Suhaila Isaak

Since the past few decades, sensors capable of detecting single photons are required for low-light-level imaging. The development of Geiger mode sensor using low voltage CMOS technology has motivated I.C fabrication industry to design sensor array on chip including XFab, IMEC and TSMC. Detection at the single photon level provides the ultimate sensitivity possible, but low counting rates constrain image acquisition rates and dynamic range [1, 2]. The integrated circuit design of SPAD with associate passive quenching circuit in CMOS is highly desirable but requires photon detection at high rates. Therefore, there is an urgent need an improved SPAD simulation model to estimate the performance of the integrated passively quenched SPAD array and on chip counting system prior fabrication.

We are focusing on the computational modeling the behavior of Single Photon Avalanche Diodes (SPADs) and adapted the earlier simulation model, which have been implemented with passive and active quenching circuits [1], [2], [3]. A SPAD model by Suhaila et. al [4] based on UMC 180nm was successfully tested with passive quenching circuit. A simple current mirror circuit with ballast resistor (RB) current source configuration has been utilized as the quenching circuit in producing sufficient output threshold level Vout before the signal is converted to digital. An improvement has been made by replacing RB with NMOS to reduce the area consumption on die using Silterra 180 nm CMOS technology. The exhibited results from this modification have deadtime at 1ns and Vout = 0.63 which is doubled the volume obtained with UMC 180 nm CMOS technology. Our model followed the probabilistic Markovian model, which modeling the sensor response uncorrelated time photon detection. In the future, we aim to provide an accurate computational sensor model including the detection efficiency, deadtime rate, various quenching method and signal to noise ratio, to be used on top of physically-based transient light transport simulations to provide a promising low-cost and highly-efficient imaging technology.



(a) (b)

Figure 11: (a) Passively Quenched SPAD Circuit. (b) Characterization of SPAD model.

More info:

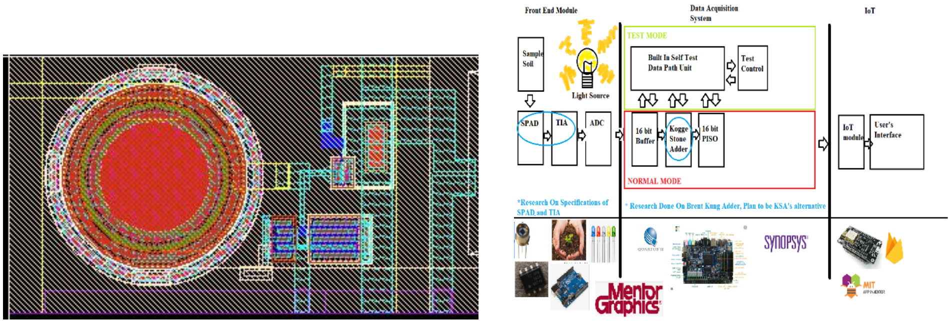
1. A. Rochas (2003), "Single Photon Avalanche Diodes in CMOS technology". EPFL,Switzerland
2. M.Dandin, N.Nelson., V. Saveliev, H. Ji and P. Abshire, "Single Photon Avalanche Detectors in Standard CMOS," IEEE Sensors, p. 585-588, Oct. 2007.
3. N. Faramarzpour, M.J.D., S. Shirani, Q. Fang (2008). “Fully integrated single photon avalanche diode detector in standard CMOS 0.18μm Technology”, IEEE Transactions on Electron Devices, Vol. 55 (3): pp. 760-767.
4. S. Isaak, S. Bull, M.C. Pitter, I. Harrison (2010), “Design and characterisation of 16x1 parallel outputs SPAD array in 0.18μm CMOS technology”, Proceedings of IEEE Asia Pacific Conference on Circuits and Systems (APCCAS 2010): pp. 979-982.

High Quantum Efficiency of 2-Dimensional Passively Quenched Photon Counting Array

Suhaila Isaak and Yusmeeraz Yusuf

The development of a SPAD device and its subsequent use in a passively quenched single photon counting imaging system, and was fabricated in a UMC 0.18 μm CMOS process. A SPAD is a low-doped p- guard ring (t-well layer) encircling the active area to prevent the premature reverse breakdown [1],[2]. The array is a 16×1 parallel output SPAD array, which comprises of a passively quenched SPAD circuit in each pixel with the current value being set by an external resistor. The SPAD I-V response, I(D) was found to slowly increase until breakdown voltage was reached at excess bias voltage, Ve = 11.03 V, and then rapidly increased due to avalanche multiplication. Digital circuitry to control the SPAD array and perform the necessary data processing was designed in VHDL and implemented on a FPGA chip. At room temperature, the dark count was found to be approximately 13 KHz for most of the 16 SPAD pixels and the dead time was estimated to be 40 ns. This integrated circuit was successfully fabricated and named as SPAD1 and SPAD with EuroPractice, Beligium. The experimental tests were carried out using the developed optical-electronics testbench by Suhaila et. al [2] in BIOS lab, University of Nottingham.

Currently, we are working with 16x2 aray, hence 2-dimensional design using Silterra 0.13μm CMOS Process to establish the circuit on chip fabrication for biosensing application. In addition, the electronic-optical testbench setup has been built to test the prototype sensor, where the DAQ system on FPGA [3]. The DAQ is equipped with built-in self-test module to trace count error for any two or more overlapped incoming photon signal. The designers are also motivated with some limitation on information exchange in agriculture area. In progress, the design of linear array SPAD has been implemented for soil macronutrient content as SPAD has the feature of very weak signal in visible range wavelength.



(a) (b)

Figure 12: (a) A layout of passively quenched SPAD pixel using UMC 0.18 μm CMOS Process. (b) Project modules on soil spectroscopy using 16×2 passively SPAD array using 0.13μm CMOS Process.

More info:

1. S. Isaak and W. Hasan, "Design and process simulation of p-well guard ring Si Avalanche Photodiode", Proceedings of 4th IEEE International Conference on Control System, Computing and Engineering, ICCSCE 2014: pp. 326-331.

2. S. Isaak, S. Bull, M.C. Pitter, I. Harrison (2010), "Design and characterisation of 16x1 parallel outputs SPAD array in 0.18μm CMOS technology", Proceedings of IEEE Asia Pacific Conference on Circuits and Systems (APCCAS 2010): pp. 979-982.

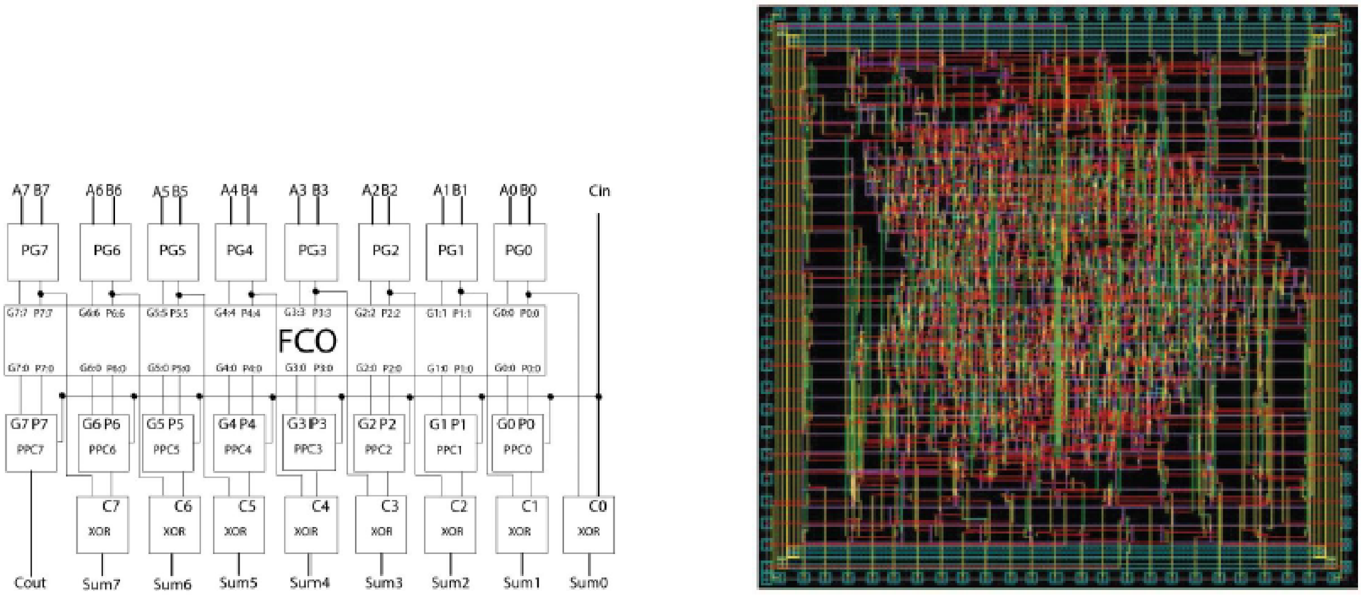
3. X. Y. Lau, C.H. Soo, Y. Yusof and S. Isaak, "Integrated Soil Monitoring System for Internet of Thing (IOT) Applications", Lecture Note on the 11th National Technical Seminar on Unmanned System Technology 2019 (NUSYS 2019): pp701-714.

VLSI Implementation of Kogge-Stone Parallel Prefix Adder

Muhammad Mun'im Ahmad Zabidi, Lee Mei Xiang, Ainy Haziyah Awab and Ab Al-Hadi Ab Rahman

The Kogge-Stone adder (KSA) is a look ahead parallel prefix form carry adder. It generates carry in $O(log_2 n)$ time and is commonly regarded one of the fastest adders. In KSA carries are computed quickly by computing them in parallel at the expense of larger silicon area.

We implemented the KSA using 0.18μm process technology. The implemented was validated through a comparison with other adder architectures including the standard ripple carry adder and the carry look ahead adder. Furthermore, our KSA adder is also compared with a default optimized adder from the Artisan standard cell library. The adders are compared for bit widths of 8, 16, and 32. The adders are designed using Verilog and synthesized using both front-end and back-end tools, with complete validation and verification stages, including analysis for performance, power, and area. Results show that in terms of performance, KSA results in the lowest propagation delay with almost constant delay for all bit widths, with up to 70% less delay as compared to all other architectures. Area and power penalty is found to also increase by roughly 59%. In terms of energy usage, the KSA adder results in up to 64% less. In the case when speed and energy are critical, this fast and energy efficient KSA adder can be readily integrated into custom VLSI designs.



(a)

(b)

Figure 13: (a) Architecture of 8-bit Kogge-Stone adder. (b) VLSI layout of 32-bit Kogge-Stone adder.

More info:

1. Xiang, L. M., Mun'im Ahmad Zabidi, M., Awab, A. H., and Ab Rahman, A. A. H, "VLSI Implmentation of a Fast Kogge-Stone Parallel-Prefix Adder", Journal of Physics: Conference Series. Vol. 1049. No. 1. IOP Publishing, 2018..

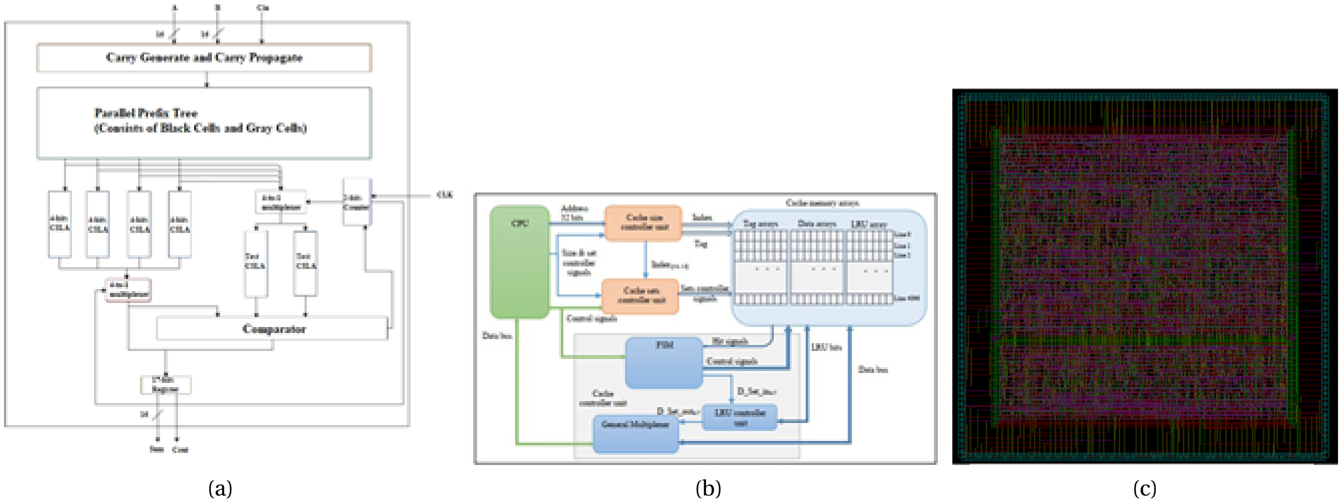
CMOS Processor Design for Biomedical Systems

Yusmeeraz Yusof, Suhaila Isaak, Chang Chin Kai, Ling Chung Yee, Man Kai Xian

The system for life-critical applications should be able to tolerate the faults coming from the soft error in harsh environment and impact due to the tremendous increase in the integration density on chip to avoid total system failure. The focus is to design a low power system on chip (SoC) with self-checking and self-repairing faults for the discrete wavelet transform (DWT) used in various biomedical systems. The DWT architecture consists of the adder, multiplier and memory elements to implement the function of predict, update, decomposition and co-efficient storage blocks.

We proposed a fault tolerant adder by mimic the triple mode redundancy techniques on the sparse kogge stone adder. This design is achieved by triplicate the summation unit with counter, register, multiplexer and spare sections. Fault signal detection and correction are achieved on the carry tree for carry generation and carry select adder for summation. Fault signals on nets will be detected and immediate replacement of the faulty section will be performed to return the system back to normal operation. The design is implemented in Verilog and synthesized using Silterra 0.18um CMOS technology. The faulting masking rate has been greatly improved at almost double the masking rate of the conventional adder. The overall ASIC implementation shows that the proposed adder can achieve optimum power dissipation and critical path delay for high performance DWT core applications.

We also work on the design of fault tolerant cache memory. A 64-bit cache memory with fast tag configuration protection circuit is implemented using single error correction (SEC) codes to tolerate the soft error occurred in cache memory. In the proposed architecture, the fast tag protection circuit requires only a SEC encoder regardless of the number of the cache way. This architecture has greatly reduced in the complexity of the circuitry and increase the reliability of the system. The design is implemented in Verilog and synthesized using Silterra 0.18um CMOS technology. The functionality of the proposed cache memory is tested using single stuck-at-fault model. The finding of this study shows that cache memory with fast tag protection circuit requires lesser area overhead and power consumption. The study is significant as it provides solution to improve the reliability of the cache memory while adding the least redundancy to the overall circuit.



(a) (b) (c)

Figure 14: (a) 16-bit fault tolerant adder (b) 64-bit fault tolerant cache memory (c) Chip layout using 0.18um CMOS.

More info:

1. CC Kai, S Isaak, and Y Yusof, "16-bit Fault Tolerant Sparse Kogge Stone Adder using 0.18mum CMOS Technology," IEEE Int. Conf. Semicond. Electron. Proceedings, pp. 77–80, 2020.

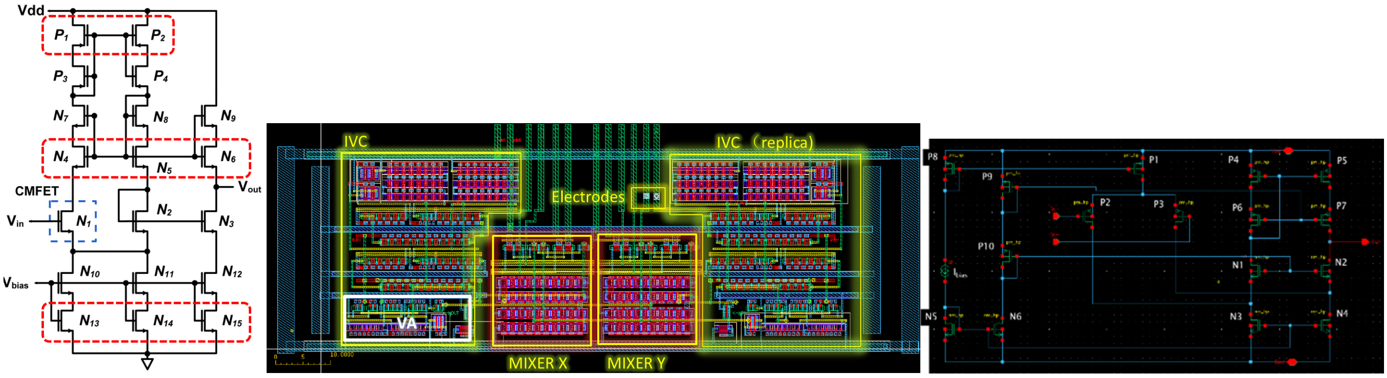
Analog Front-End Design for Point-of-Care Applications

Yusmeeraz Yusof, Wong How Hwan, Vinny Lam Siu Fan, Ow Tze Weng

Point-of-care (POC) testing brings the test conveniently and immediately to the patient. The advancement in CMOS technology allows the realization of a complete system that integrates the sensing unit and transducer element in the same device. The focus is to propose a circuit topology that is able to accurately detect a very low level biological signal and convert it into proportional voltage for further processing.

In view of the growing concerns about food security, health care, evidence-based care, infectious disease, and tailor-made medicine, a portable gene-based POC system is desired. We proposed a label-free electrical detection circuit based on two principles, which are potentiometric and impedimetric sensing. In the case of deoxyribonucleic acid (DNA) detection, the occurrence or nonoccurrence of specific DNA hybridization can be detected by the difference in charge since a nucleotide has a negative charge on the phosphate group. In the first approach, we proposed a detection circuit that consists of a self-cascode source-drain follower and a two-stage differential amplifier to improve the input voltage range and achieve the high gain. Charge-modulated field-effect transistors are used as sensing devices since it offers simplicity by eliminating the use of an external reference electrode and is compatible with the standard CMOS process. The input voltage range of the improved source-drain follower ranges from 0.104 V to 1.28 V within ± 5 mV of accuracy and the power consumption is as low as 1.8 nW. The achieved overall gain is 79.81 dB with a frequency range of 1.528 kHz. In the second approach, a current-to-voltage converter and two quadrature phase double-balanced Gilbert cell mixers are used to compute the impedance changes defined by the capacitive and resistive components of the on-chip microelectrodes signal. The proposed detection circuit can achieve a transimpedance gain of 166 dB and power consumption of 97.2 μW. All designs are translated into physical layout using 0.18 μm Silterra CMOS process.

We also propose an analog front-end circuit for electrocardiogram (ECG) monitoring system. While the demand of low power wearable cardiac monitoring is increasing in exponential way, the front-end ECG amplifiers are still suffering from flicker noise for low frequency (0.05 Hz to 250 Hz) cardiac signal acquisition, 50 Hz power line electromagnetic interference, and the large unstable input offsets. We use folded cascode topology in amplifier circuit to achieve high gain and low input referred noise design. Transistor sizing ratio are carefully done to optimize the CMRR and PSRR. The design is simulated using Cadence and able to response to cardiac signal characteristics range, where the input signal amplitude is from 5 μV to 10 mV. The total power consumption is only 3 μW and thus suitable to be implemented with further signal processing and classification back-end for low power POC applications.



(a) (b) (c)

Figure 15: (a) Potentiometric detection circuit (b) Impedimetric sensor layout (c) ECG amplifier design

Design and Implementation of Lightweight Encryption Algorithm Using Prince Cipher

Shahidatul Sadiah and Lee Jiah Shun

Internet of Things (IoT) has been a business enabler by leveraging the information shared in the network into the real world. However, when thousands of devices are connected and transferring data from each other, security challenges are inevitable. Therefore, encryption of data has become essential as a part of the countermeasure. Particularly, a lightweight cryptography has gained the attention since it is an encryption method that features a small footprint low computational complexity, which is appropriate for the IoT devices.

In 2019, a study reports on the comparison analysis of lightweight block ciphers using loop unrolled method to get a single cycle ASIC implementation [1]. In the study, low area and good throughput are achieved, but their practicality is limited by the long delay. The only algorithm that achieves low latency is reported to be PRINCE block cipher. The block size is 64-bit, and it has 11 rounds at the core, five forward rounds, one middle round, and five backward rounds which operate as substitution-permutation network to produce the ciphertext block. In this work, the ASIC implementation of PRINCE block cipher is analyzed in term of the latency, area, and power consumption performance metrics. The implementation is based on SAED 32nm Cell Library using Synopsis tools. In addition, analysis on the different method of design synthesis constraint is applied to improve the performance. Then, the analysis on different coding method is also added since the Verilog code construction matters as it heavily affects how the design tool understand the logic and the synthesize process. The results reveal that the performance of PRINCE block cipher in single-cycle implementation is improved with optimized constraints and RTL design. They are better in terms of latency and throughput but suffer at higher cell area and power consumption. With single cycle implementation, the required latency for a complete encryption process can be reduced to 3.87ns from 6.37ns that is measured from round-per-cycle implementation.

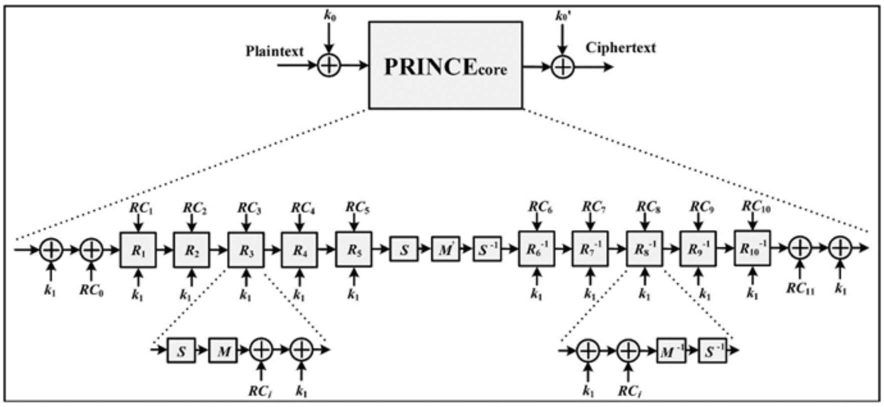

k_0
Plaintext
PRINCEcore
k_0'
Ciphertext
RC_1 RC_2 RC_3 RC_4 RC_5
RC_6 RC_7 RC_8 RC_9 RC_{10}
R_1 R_2 R_3 R_4 R_5 S M' S^{-1} R_6^{-1} R_7^{-1} R_8^{-1} R_9^{-1} R_{10}^{-1}
k_1 RC_0 k_1 k_1 k_1 k_1 k_1 k_1 k_1 k_1 k_1 k_1 RC_{11} k_1
S M RC_i k_1
M^{-1} S^{-1} k_1 RC_i

Figure 16: The PRINCE block cipher core.

More info:

1. P.Maene, "Single-Cycle Implementations of Block Ciphers," Lightweight Roots of Trust for Modern Systems on-Chip, pp. 47-64, 2019.
2. S Sadiah, Lee Jiah Chun, “Design and Implementation of Lightweight Encryption Algorithm Using Prince Cipher”, January 2021.

Grounded Active Inductor using Carbon Nanotube Field Effect Transistors

Nasir Shaikh-Husin, Muhammad I. Masud, Iqbal A. Khan, Abu Khari A'ain

Inductors are important components of many high frequency analog signal processing circuits. For example, they are utilized in low noise amplifiers (LNAs), voltage controlled oscillators, filters, frequency dividers, impedance matching networks, and phase shifters. An on-chip spiral inductor requires a large die area of an integrated circuit (IC), resulting in higher fabrication cost. It also suffers several disadvantages such as low quality factor (QF), fixed inductance value and incompatibility with semiconductor fabrication process. These limitations motivate IC designers to design active inductors (AIs) that can be more easily fabricated together with the rest of IC components.

An AI can have tunable inductance, has high QF, and it consumes smaller area. Current AI implementations use large number of active and passive devices, thus require large area and consume considerable power. Their high frequency performance is also limited due to low self-resonance frequency. Most of contemporary AIs are implemented in complementary metal oxide semiconductor (CMOS) technology. To facilitate very high frequency (approaching 100 GHz) and wide bandwidth (tens of GHz), we proposed a grounded AI (GAI) using carbon nanotube field effect transistor (CNTFET) technology. The circuit is based on a gyrator-C topology utilizing positive transconductance element and negative transconductance element as active building blocks. CNTFET is realized by replacing the CMOS field effect transistor (MOSFET) conventional channel with an array of isolated and aligned single wall carbon nanotubes (CNTs). The CNTs behave as the medium of conduction between the drain and source terminals. Like MOSFET, CNTFET also works as a voltage controlled active device. CNTFET gate is coupled capacitively with the underneath channel that utilizes one or more CNTs. In comparison to MOSFET, advantages of CNTFET are higher temperature resilience, larger transconductance, larger driving current, one-dimensional ballistic transport capability, near ideal subthreshold slope and lower intrinsic capacitances [1].

The proposed GAI offers a tunable inductance with a very wide inductive bandwidth, high QF and low power dissipation. The tunability of the realized circuit is achieved through CNTFET based varactor. The proposed topology achieves inductive behavior in the frequency range of 0.1 GHz – 101 GHz with maximum QF of 9,125. The GAI operates at 0.7 V and consumes 0.338 mW. To demonstrate the performance of the GAI, a broadband LNA circuit was designed utilizing the GAI. The realized LNA provides high frequency bandwidth (17.5 GHz – 57 GHz), low noise figure (< 3dB), a flat forward gain of 15.9 ± 0.9 dB, a reverse isolation less than -63 dB and input return loss less than -10 dB over the entire frequency bandwidth. The proposed CNTFET based GAI and LNA circuit designs were verified using HSPICE simulations with Stanford CNTFET 16 nm technology node model parameters.

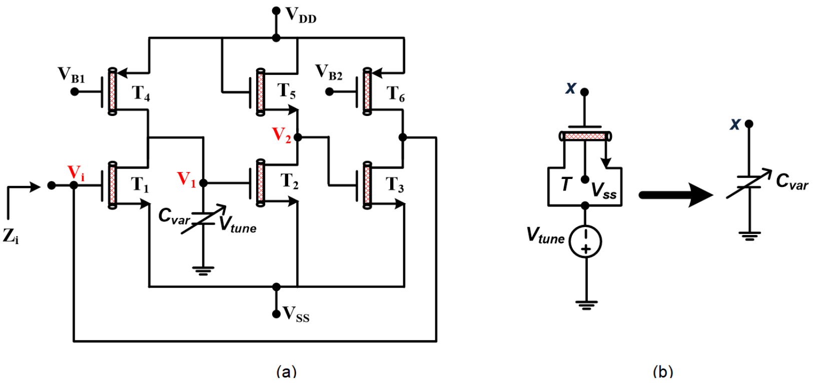


Figure 17: CNTFET GAI (a) proposed circuit (b) varactor circuit and its symbol.

More info:

1. Muhammad I. Masud, Abu A'ain, Iqbal Khan, and Nasir Husin, "Design of Voltage Mode Electronically Tunable First Order All Pass Filter in ±0.7 V 16 nm CNFET Technology", Electronics, vol. 8, no. 1:95, 2019.

Low-Area and Accurate Inner Product Based on Stochastic Computing

Nasir Shaikh-Husin, Hamdan Abdellatef, Mohamed Khalil-Hani, Sayed Omid Ayat

Stochastic computing (SC) is an emerging computing paradigm, which was first introduced by Gaines [1] in the 1960s, as an alternative to the conventional binary-encoded (BE) deterministic computing technique. In SC, data are represented by bit-streams (referred to as stochastic numbers (SNs)), and the value of the data is encoded as the probability of 1s appearing in the bit-stream. The main advantage of an SC element is its low hardware cost (e.g., AND gate is an SC multiplier) and high tolerance for soft errors. SC has been adopted in a wide range of embedded solutions for image processing, neural networks, and digital filters. A key function for all applications mentioned above is the inner product (multiply-accumulate) function.

Research has shown that this function, when implemented in SC domain, can result in significant reduction in area cost and power consumption compared to its equivalent counterpart in the conventional BE deterministic computing. However, existing designs of SC inner product are disadvantaged due to high BE-SC conversion circuits, hence high overall area cost. They also suffer from correlation-induced errors that affect their accuracy performance. We propose a novel inner product design method for the SC domain that has high accuracy, low area cost, and most importantly, the circuit is correlation-insensitive thus correlation manipulation circuit is not necessary [2]. In this work, we use SC one-line bipolar representation. Experimental results show that the proposed design on average reduces 85.7% of hardware footprint when compared to its equivalent BE counterpart. It outperforms current state-of-the-art SC designs in terms of area savings, both in computation and conversion costs. Furthermore, it achieves better (or comparable) accuracy performance compared to existing works, especially in designs having large number of inputs with low SN lengths.

The proposed circuit to calculate the inner product z of two vectors x and h with N. X and H are matrices that contain SNs representing the values in vectors x and h, respectively, where the dimension of these matrices is $N \times L$. X_i and H_i are SNs of length L that encode the values x_i and h_i, respectively. Z is the output SN that encodes the inner product result z.

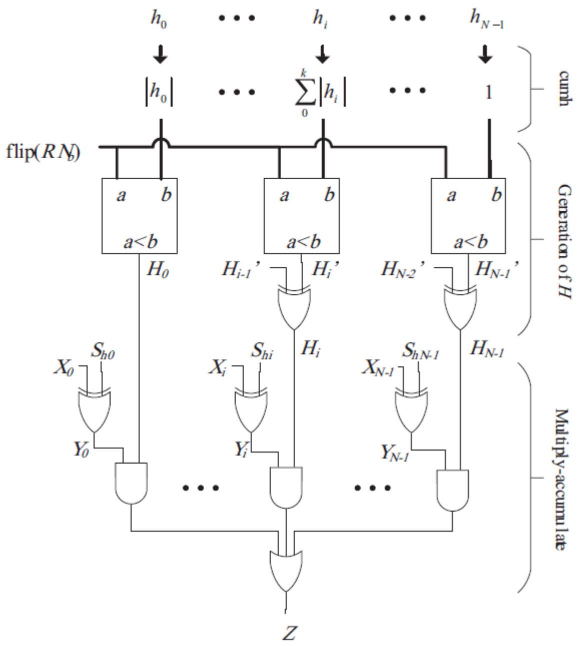


Figure 18: Proposed stochastic inner product circuit.

More info:

1. B. R. Gaines, “Stochastic Computing”, in ACM Proc of the Spring Joint Computer Conf, pp. 149 – 156, 1967.
2. Hamdan Abdellatef, Mohamed Khalil-Hani, and Nasir Shaikh-Husin, “Accurate and Compact Stochastic Computations by Exploiting Correlation”, Turk J Elec Eng & Comp Sci, vol. 27, pp. 547 – 564, 2019.