Volume 24, Number 3
Third Quarter 2024
ISSN 1531-636X

IEEE Circuits and Systems MAGAZINE

75 YEARS

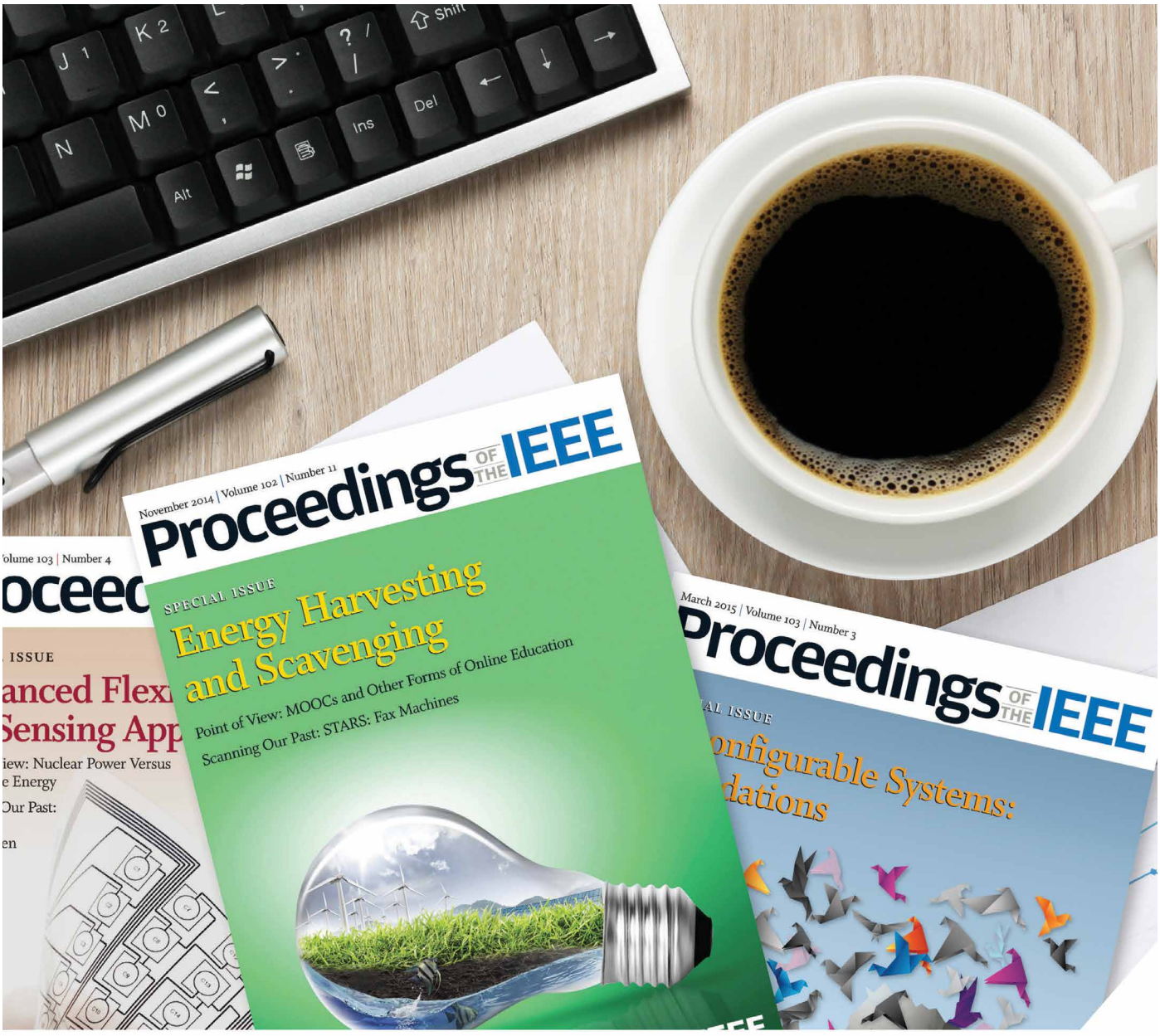
CAS

CELEBRATING THE 75TH Anniversary

In Memoriam: Belle Ananth Shenoi

CAS
IEEE CIRCUITS AND SYSTEMS SOCIETY

IEEE



One of the most influential reference resources for engineers around the world.

For over 100 years, *Proceedings of the IEEE* has been the leading journal for engineers looking for in-depth tutorial, survey, and review coverage of the technical developments that shape our world. Offering practical, fully referenced articles, *Proceedings of the IEEE* serves as a bridge to help readers understand important technologies in the areas of electrical engineering and computer science.

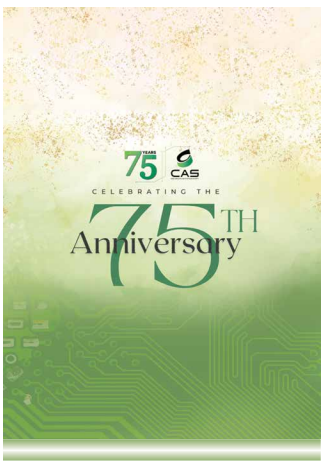
To learn more and start your subscription today, visit
ieee.org/proceedings-subscribe

IEEE

IEEE Circuits and Systems MAGAZINE

Volume 24, Number 3 Third Quarter 2024

Features



CIRCUITS—IMAGE LICENSED BY INGRAM PUBLISHING

IEEE Circuits and Systems Magazine (ISSN 1531-636X) (ICDMEN) is published quarterly by the Institute of Electrical and Electronics Engineers, Inc. Headquarters: 3 Park Avenue, 17th Floor, New York, NY, 10016-5997 USA. Responsibility for the contents rests upon the authors and not upon the IEEE, the Society, or its members. IEEE Service Center (for orders, subscriptions, address changes): 445 Hoes Lane, Piscataway, NJ 08854 USA. Telephone: +1 732 981 0060, +1 800 678 4333. Individual copies: IEEE members US$20.00 (first copy only), nonmembers US$159.00 per copy; US$7.00 per member per year (included in Society fee) for each member of the IEEE Circuits and Systems Society. Subscription rates available upon request. Periodicals postage paid at New York, NY, and at additional mailing offices. **Postmaster:** Send address changes to IEEE Circuits and Systems Magazine, IEEE Operations Center, 445 Hoes Lane, Piscataway, NJ, 08854 USA.
Printed in U.S.A.

Digital Object Identifier 10.1109/MCAS.2024.3407011

Departments

Magazine Deadlines

Final materials for *IEEE Circuits and Systems Magazine* must be received by the Editor on the following dates:

Issue	Due Date
First Quarter	December 1
Second Quarter	February 1
Third Quarter	May 1
Fourth Quarter	August 1

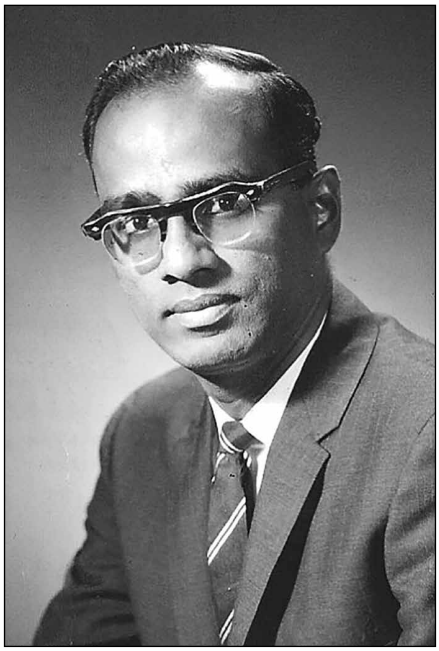
IEEE prohibits discrimination, harassment, and bullying. For more information, visit https://www.ieee.org/about/corporate/governance/p9-26.html.

Digital Object Identifier 10.1109/MCAS.2024.3407009

Obituary

Sanjit K. Mitra, *Life Fellow, IEEE*

Belle Ananth Shenoi (1929–2024)



It is with a heavy heart that I am writing this article in honor of Belle Ananth Shenoi who passed away on 19 April 2024 in Tampa, Florida, at the age of 94 with his family around him. He is survived by his wife Suman, son Hemanth, daughter Tara, and their families. I have known Ananth for a very long time because of our close association with the activities of the IEEE Circuits and Systems Society. He was one of my closest friends. I shall miss him very much.

Ananth was born in 23 December 1929 at Katapadi, India. He married Suman in December 1961. He received the B.Sc. degree in physics from the University of Madras in 1951, the D.I.I.Sc. degree in electrical engineering from the Indian Institute of Science, Bangalore, in 1955, and the M.S. and Ph.D. degrees in electrical engineering from the University of Illinois, Urbana, in 1958 and 1962, respectively. He joined the Faculty of electrical engineering at the University of Minnesota in September 1962. After 24 years at the University of Minnesota, he joined the Department of Electrical Engineering at Wright State University as a professor and chair in 1986. After joining the Wright State University, he organized the IEEE International Conference on Systems Engineering held at his university in September 1987. He continued organizing this conference which was held at his university in 1989 and 1991, respectively.

Ananth is internationally known for his research in active and digital filter theory and has published two books and also authored/coauthored more than 50 papers published in professional journals. He was elected an honorary professor of electrical engineering of the National Cheng-Kung University, Taiwan, in 1987. He was also elected a fellow of the Institute of Electronics and Telecommunication Engineers of India in 1988. During the four summers of 1968 to 1971, he was a Member of the Technical Staff of the AT&T Bell Laboratories, Murray Hill, New Jersey. Ananth has served as a Consultant for several industries between 1966 and 1986.

Ananth served as the President of the IEEE Circuits and Systems Society in 1991. In recognition of his many years of service in various capacities, Ananth received the Meritorious Service Award at its 1992 International Symposium on Circuits and Systems. He also served as a member of the Editorial Board for the Journal of Systems Engineering. He was elected a Life Fellow of the IEEE in 1995. Ananth was awarded the Golden Jubilee Medal by the IEEE Circuits and Systems Society in 1999 and the Third Millennium Gold Medal by the IEEE in 2000, in recognition of his research contributions to IEEE.

Prof. Suhash Chandra Dutta Roy, Professor Emeritus of the Indian Institute of Technology Delhi sent the following note after hearing the death of Prof. Shenoi: "My interactions with Shenoi" dates back to 1965. I had then submitted my Ph.D. thesis to the Calcutta University, and was waiting for the thesis defense examination. Within a few days, I received a letter from him from the University of Minnesota, where he was an Assistant Professor, inviting me to join him as a graduate student, supported by his NSF Project. To this I replied back stating that I had already completed my graduate studies and that I shall be happy to work with him, provided I get an offer of an Assistant Professorship. After a couple of months, I received a formal appointment from the Head of the EE Department which I duly accepted and joined the university in September 1965.

Shenoi had arranged for everything one needs if one comes to the U.S. for the first time and including an accommodation within walking distance from the university, and an office room in the Department. When we

Digital Object Identifier 10.1109/MCAS.2024.3401788
Date of current version: 15 August 2024

met, he gave me a long lecture on American customs. Within a few days of my arrival, he took me to his apartment and introduced me to his wife Suman.

Shenoi and I worked together on his NSF Project and published several papers in various journals. After three year stay, I decided to go back to India and joined IIT Delhi as an Associate Professor of EE in 1968. Shenoi and I met a few times until 1981. Since then, I have had correspondences with him off and on. For a long time, I had no direct contact with him until recently I got the sad news from Prof. Mitra. In Shenoi's passing away, I lost one of my very good friends, who was more than a friend to me. He was like an elder brother guiding me with good advice when I needed it. May his soul rest in peace."

Prof. Mostafa Kaveh of the University of Minnesota mentions: "I am very sad to hear that Belle Ananth Shenoi has passed away. I first met him when I joined the University of Minnesota in 1975 as a new Faculty Member in the areas of communications and radar. The discipline of digital signal processing (DSP) was evolving and, as a filter theorist, Belle had started to introduce students to 1-D and 2-D digital filter design. Our areas of specialty differed but converged over the years as signal processing expanded to include statistical methods and image processing among others. From the start, he was most welcoming, and played a significant role in providing mentorship and support in my formative years as faculty. His genuine interest in my work and development were instrumental in the trajectory of my academic and administrative career. I am grateful to have been Belle's colleague and friend."

Prof. Guoxiang Gu of the Louisiana State University wrote: "I am so sad to hear this news. Belle had been very nice to me when I was a graduate student at the University of Minnesota and when I was a Visiting Assistant Professor at Wright State University. Belle and his wife came to Baton Rouge once in middle of 1990s as their daughter was at LSU pursuing her MS degree. They stayed in our home for a few days. Belle was my mentor and has been very helpful to my early career, and extremely nice to me in both academic and personal relations. He will always be in my heart."

Dr. Raj Hegde, Broadcom Corporation, Austin, Texas said: "When I enrolled in a digital signal processing course with Prof. Shenoi during my first year of Graduate School at Wright State University, I had no idea that it would shape my career path in that field. His enthusiasm for signal processing and the underlying mathematics was very palpable in each of his lectures. While he held high expectations for dedication and hard work from his students, he was remarkably patient and deeply involved in their research work. I consider myself incredibly fortunate to have had him as an advisor, mentor, and lifelong friend. His presence will be sorely missed, but his impact will continue to resonate in the lives of those he touched."

Society News

Manuel Delgado Restituto, Past-President, IEEE Circuits and Systems Society

IEEE CASS Awards 2024

IEEE CAS Society is pleased to congratulate the following recipients of the 2024 IEEE Circuits and Systems Society Awards. The Society is especially proud and happy to honor their outstanding accomplishments and we look forward to their continued involvement and close contact with the CAS Society.

Society and Achievement Awards

IEEE Circuits and Systems Society Charles A. Desoer Technical Achievement Award

Wenwu Zhu

For contributions to visual signal processing and video communications technologies.

IEEE Circuits and Systems Society John Choma Education Award

Gordana Jovanovic Dolecek

For contributions to education in signal processing.

IEEE Circuits and Systems Society Meritorious Service Award (2 awards)

Victor Grimblatt

For his work promoting the microelectronics ecosystem both globally and regionally through CASS activities.

Yong Lian

For long-term outstanding services to the IEEE Circuits and Systems Society as the Vice President for Region 10, Vice President for Publications, and as the President.

Best Paper Awards

IEEE Transactions on Circuits and Systems Guillemin-Cauer Best Paper Award

Alex P. James and Leon O. Chua

For the paper entitled "Analog Neural Computing with Super-Resolution Memristor Crossbars," in *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 68, no. 11, pp. 4470–4481, Nov. 2021, doi: 10.1109/TCSI.2021.3079980.

IEEE Transactions on Circuits and Systems Darlington Best Paper Award

Mohammad Oveisi, Huan Wang, and Payam Heydari

For the paper entitled "A Study of a Millimeter-Wave Transmitter Architecture Realizing QAM Directly in RF Domain," in *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 70, no. 6, pp. 2243–2256, June 2023, doi: 10.1109/TCSI.2023.3255109.

IEEE Transactions on Circuits and Systems for Video Technology Best Paper Award

Zhenxun Yuan, Xiao Song, Lei Bai, Zhe Wang, and Wanli Ouyang

For the paper entitled "Temporal-Channel Transformer for 3D Lidar-Based Video Object Detection for Autonomous Driving," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 4, pp. 2068–2078, April 2022, doi: 10.1109/TCSVT.2021.3082763.

IEEE Transactions on Very Large-Scale Integration Systems Prize Paper Award

Kun-Chih (Jimmy) Chen, Yuan-Hao Liao, Cheng-Ting Chen, and Lei-Qi Wang

For the paper entitled "Adaptive Machine Learning-Based Proactive Thermal Management for NoC Systems," in *IEEE Transactions on Very Large-Scale Integration (VLSI) Systems*, vol. 31, no. 8, pp. 1114–1127, Aug. 2023, doi: 10.1109/TVLSI.2023.3282969.

IEEE Transactions on Biomedical Circuits and Systems Best Paper Award

Gawsalyan Sivapalan, Koushik Kumar Nundy, Soumyabrata Dev, Barry Cardiff, and Deepu John

For the paper entitled "ANNet: A Lightweight Neural Network for ECG Anomaly Detection in IoT Edge Sensors," in *IEEE Transactions on Biomedical Circuits and Systems*, vol. 16, no. 1, pp. 24–35, Feb. 2022, doi: 10.1109/TBCAS.2021.3137646.

Digital Object Identifier 10.1109/MCAS.2024.3411675
Date of current version: 15 August 2024

(continued on page 7)

Society News

Class of 2024 IEEE Fellows Who Are Members of the Circuits and Systems Society

Congratulations to the 2024 CAS Society Class of IEEE Fellows! Thank you for your many contributions to IEEE Circuits and Systems Society.

2024 Fellows Elevated by the CAS Society

Warren Gross

For contributions to the design of algorithms and integrated circuit architectures for communication systems.

Tsung-Yi Ho

For contributions to design automation and test of microfluidic biochips.

Liang Lin

For contributions to multimedia content analysis.

Huchuan Lu

For contributions to visual object tracking and salient object detection.

Jiwen Lu

For contributions to visual content analysis and recognition.

Siwei Ma

For contributions to video coding technologies and standards.

Xiaoning Qi

For leadership in open-source hardware of computer architecture and contributions to holistic interconnect system design.

Esther Rodriguez-Villegas

For contributions to low power biomedical circuits and systems for wearable medical applications.

2024 Fellows Elevated by other IEEE Entities

Yu-Gang Jiang

For contributions to large-scale video analysis and open-source datasets.

Vijay Raghunathan

For contributions to design of low power and energy harvesting embedded systems.

Jing Xiao

For contributions to the multiple modality knowledge mining technologies.

Digital Object Identifier 10.1109/MCAS.2024.3415243
Date of current version: 15 August 2024

Yang Tang

for contributions to hybrid multi-agent systems and complex networks.

Matthew Ettus

For contributions to software-defined radio products.

Dusan Stipanovic

For contributions to control of complex systems.

Wen-Huang Cheng

For contributions to intelligent multimedia computing and applications.

Shimeng Yu

For contributions to non-volatile memories and in-memory computing.

Claudio Adragna

For contributions to industrial usage of resonant converters in offline power supplies.

Deukhyoun Heo

For contributions to CMOS power amplifiers in multilayer packages and reconfigurable reactive components.

Ke-Horng Chen

For contributions to power management integrated circuits and system design.

Shuo-Wei Chen

For contributions to data converter architectures and clock generation techniques.

Seonghwan Cho

For contributions to time-domain circuits and applications.

Vladimir Stojanovic

For contributions to electronic-photonic design and system-on- chip integration.

Sri Parameswara

For contributions to embedded computer circuits and systems.

Qinru Qiu

For contributions to modeling and optimization of energy efficient computing systems.

Circuits and System Society Student Design Contest

The IEEE Circuits and Systems Society (CASS) held the 8th CASS Student Design Competition. The CASS Student Design Competition is a worldwide competition where undergraduate students suggest and execute projects

on electrical engineering and related areas. The focus is on finding a solution to a real-life problem based on circuits and systems. The competition takes place in three phases, the first at the chapter level, the second at the regional level, and the final at the world level.

The final this year took place during the 2024 IEEE International Symposium on Circuits and Systems (ISCAS 2024) in Singapore. Four finalists, one for each region—Region 1–7 (USA and Canada), Region 8 (Europe, Middle East, Africa), R9 (Latin America), and Region 10 (Asia, Australia, Pacific) competed on the final stage! The results of the competition are listed below.

2023–2024 Student Design Competition—World Winner

A Wearable Cardiopulmonary Healthcare System For Real-Time Monitoring Of Multi-Modal Physiological Signals

Hanyu Shi, Yichen Long, Zhiyu Huang, Shuxun Li, Jiaxun Zhou, Zhengshi Liang, Yanchang Yang, Changyan Chen, and Yuhang Zhang

Shanghai Chapter

2023–2024 Student Design Competition—Regional Winners

Winner in North America

Ghost in the Machine: High-Performance FPGA Implementation for Deep Learning Acceleration

Maya Borowicz, James Ding, Winnie Fan, Zhongqi Gao, Davis Jackson, Zhengyi Lu, and Sophia Rohlfsen

Houston Chapter

Winner in Europe, Middle East, and Africa

FeatherTech

Victor Kimaru, Sheila Otuko, and Mercy Runo

Kenya Chapter

Winner in Latin America

Low Cost Interactive Embedded Device for Attention Training in Children with ADHD

Dennys J. Báez-Sanchez, María del Cisne Ortega-Cabrera, Luis J. Serpa-Andrade, Luis F. Guerrero-Vásquez, and Jorge O. Ordoñez-Ordoñez

Ecuador Chapter

Society News *(continued from page 5)*

IEEE Journal on Emerging and Selected Topics in Circuits and Systems Best Paper Award

Soyed Tuhin Ahmed, Kamal Danouchi, Christopher Münch, Guillaume Prenat, Lorena Anghel, and Mehdi B. Tahoori

For the paper entitled "SpinDrop: Dropout-Based Bayesian Binary Neural Networks with Spintronic Implementation," in *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 13, no. 1, pp. 150–164, March 2023, doi: 10.1109/JETCAS.2023.3242146.

IEEE Outstanding Young Author Award

Zunjin Zhao, Bangshu Xiong, Lei Wang, Qiaofeng Ou, Lei Yu, and Fa Kuang

For the paper entitled "RetinexDIP: A Unified Deep Framework for Low-Light Image Enhancement," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 3, pp. 1076–1088, March 2022, doi: 10.1109/TCSVT.2021.3073371.

Best Chapter Awards

Chapter of the Year Award

IEEE CASS Kerala Chapter

Chapter Chair: Sreelal S Pillai.

Regions 1–7 Chapter of the Year Award

IEEE CAS/EDS North Jersey Chapter

Chapter Chair: Durgamadhab (Durga) Misra.

Region 8 Chapter of the Year Award

IEEE CAS Kenya Chapter

Chapter Chair: Allan K. Koech.

Region 9 Chapter of the Year Award

IEEE Peru CAS Chapter

Chapter Chair: Walter Calienes Bartra.

Region 10 Chapter of the Year Award

IEEE CAS Tainan Chapter

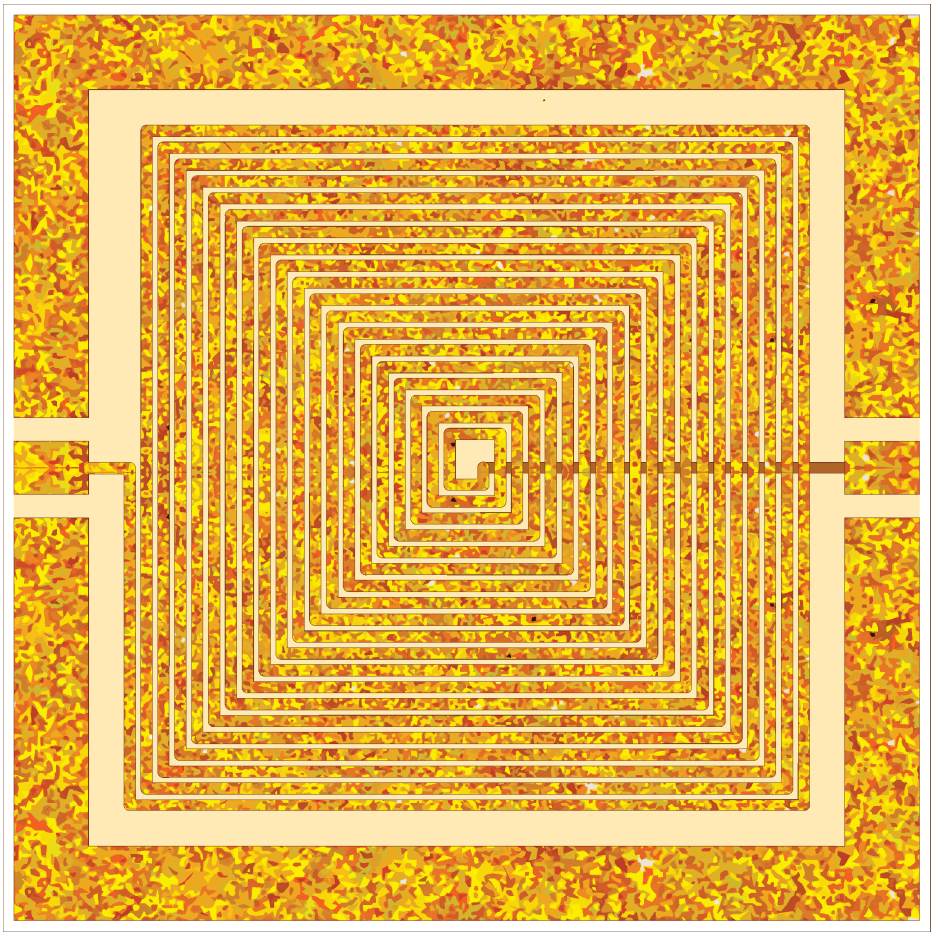
Chapter Chair: Cheng-Ta Chiang.

The following awards were not given in 2024.

IEEE Circuits and Systems Society Mac Van Valkenburg Award.

IEEE Circuits and Systems Society Industrial Pioneer Award.

IEEE Circuits and Systems Society Vitold Belevitch Award (not eligible in 2024).



A Compact Electronically Tunable Meminductor Emulator Model and Its Application

Pankaj Kumar Sharma, Rajeev Kumar Ranjan, *Senior Member, IEEE,* and Sung-Mo Kang, *Life Fellow, IEEE*

Digital Object Identifier 10.1109/MCAS.2024.3353151
Date of current version: 15 August 2024

Abstract

A compact MOSFET-C floating/grounded meminductor emulator (MIE) model is presented for high operating frequency and low power operation. The proposed MIE uses only 22 MOSFETs and two capacitors. Its performance is theoretically analyzed and rigorously verified using the Cadence Virtuoso software and hardware prototypes. The proposed MIE operates appropriately for a wide range of frequencies up to 5 MHz with 590 μW power consumption at a 180nm CMOS technology node and manifests important signature properties. The MIE layout area in 180 nm CMOS technology is 13107.5 μm^2. To analyze the effects of statistical variations in MIE elements, extensive Monte Carlo simulations have been performed to demonstrate the robustness of the proposed MIE. For experimental validation, hardware prototypes have been developed and tested successfully. An MIE-based adaptive learning neuromorphic circuit is presented to show that it can mimic the behavioral responses of amoeba under varying environments such as temperature.

Index Terms—Meminductor emulator model, pinched hysteresis loop, memory element, adaptive learning.

I. Introduction

Memristor (MR) was first proposed by L. O. Chua in 1971, and Di Ventra et al. [1] extended the MR concept to memcapacitor (MC) and meminductor (MI) in 2009. The MR, MC, and MI are memory elements (mem-elements) and thus can remember the last values and show memory effects. While MR shows a pinched hysteresis loop (PHL) in the voltage-current plane, MI shows PHL in the current-flux plane, and MC shows PHL in the charge-voltage plane [2]. The concept of a meminductor emulator (MIE) needs to be further investigated since MI device may appear as a commercial device in the near future. Several researchers have proposed MIE circuits [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19]. Many mutator circuits that can convert MRs into MIs have been described in the literature [3], [4], [5], [6]. In 2010, Pershin and Di Ventra [3] proposed a meminductor using memristor and op-amp. However, this emulation uses an approximately-equivalent circuit. In 2014, Yu et al. [4] proposed an MIE using a memristor, op-amp, and 2nd generation current conveyor (CCII). A universal mutator MIE is introduced in [5] and [6] that transfers MR into MIE. A voltage differencing current conveyor (VDCC) based mutator MIE circuit was proposed in [6], but it required a passive resistor and its maximum operating frequency is limited to 700 kHz. Later, many analog building blocks (ABBs)-based MIEs were proposed [7], [8], [9], [10], [11], [12], [13], [14], [15, [16], [17], [18], [19]. In 2020, Vista and Ranjan [7] proposed an MIE using two voltage differencing transconductance amplifiers (VDTAs), but it also requires a passive resistor and operates only up to 1 MHz. An operational transconductance amplifier (OTA)-based MIE is present in [8], but it operated only up to 10 kHz and required a multiplier. Raj et al. [9] proposed a multiplier-less meminductor using CCII and OTA, but it operated up to 700 kHz with power consumption of 14.36 mW. The MIEs proposed in [3], [4], [5], [6], [7], [8], and [9] require at least one passive resistor to emulate MI properties. In 2021, Raj et al. [10] proposed the first resistor-less MIE using three OTAs that operated up to 10 MHz, but its operation is limited to the grounded mode only. Kumar and Nagar [11] proposed an MIE using VDTA and OTA, which operated up to 3 MHz, and the capacitor value was 2 pF, which is practically unacceptable. Furthermore, the MIEs in [10] and [11] were limited to the grounded mode. In 2021, Bhardwaj and Srivastava [12] proposed an MIE using two VDTAs that operated up to 1.5 MHz. Yadav et al. [13] proposed an MIE using two OTAs and one current differencing buffered amplifier (CDBA), which operated up to 2 MHz and required a total of 37 MOSFETs and 6 current sources, which is not cost effective. Bhardwaj and Srivastava [14] proposed an MIE using one modified VDCC (MVDCC) and one OTA and it required a total of 38 transistors, one resistor, and two capacitors, and operated up to 300 kHz. A differential difference current conveyor (DDCC)-based MIE with hard and smooth switching behaviors was proposed in [15], but it required a multiplier and operated only up to 1.5 kHz. A multiple output VDTA (MO-VDTA)-based MIE was proposed in [16] that operated up to 50 MHz, but at 50 MHz frequency the capacitor value is only 5 pF. In [17], a single VDTA-based meminductor was introduced by Sharma et. al. which operated up to 25 MHz, but it worked in grounded mode only. In 2023, Korkmaz et. al. [18] proposed a grounded MIE-based on CCII and an analog multiplier that operates up to 700 kHz, but it requires a passive resistor and inductor. An OTA-based floating MIE was introduced by Aggarwal et. al. [19], which can operate up to 900 kHz. According to our literature search of MIEs, most of the MIE circuits described so far in the literature are complicated: for example, several MIEs [5], [8], [15], [18] use multipliers, some MIEs [5], [6], [7], [8], [9], [14], [15], [18] uses passive resistors, and some MIEs [8], [9], [10], [11], [15], [17], [18] works in grounded mode only.

Pankaj Kumar Sharma is with the Department of Electronics Engineering, Indian Institute of Technology (ISM) at Dhanbad, Dhanbad 826004, India, and also with the Department of Electronics and Communication Engineering, NIT Srinagar, Srinagar 190006, India (e-mail: sharma.prj@gmail.com).
Rajeev Kumar Ranjan is with the Department of Electronics Engineering, Indian Institute of Technology (ISM) at Dhanbad, Dhanbad 826004, India (e-mail: rkranjan2k@gmail.com; rajeev@iitism.ac.in).
Sung-Mo Kang is with the Department of Electrical and Computer Engineering, University of California, Santa Cruz, CA 95064 USA (e-mail: skang@ucsc.edu).

Meminductor-based circuit design and in-memory computing will attract more and more attention, giving rise to an ever-growing number of different computing approaches.

In this article, we present a compact, electronically tunable floating/grounded MOSFET-C-only MIE that uses only 22 MOSFETs and two capacitors. The proposed MIE is compact and its layout area in 180 nm CMOS technology is 13107.5 μm^2. This MIE operates up to 5 MHz, which is high compared to most of the available meminductors. The power consumption is only 590 μW. This article is organized as follows: an introduction in Section I followed by the MIE circuit design in Section II, simulation results, and verification in Section III. Section IV shows comparisons of our MIE with others available in the literature, followed by an application of our MIE to associative learning in Section V. Section VI concludes the article.

II. Proposed Meminductor Emulator

The generalized model of the meminductor emulator (MIE) is shown in Fig. 1 [7]. Considering the input current (i_{in}), equal to the current flowing through R, we can write a general meminductance equation in terms of flux ($\phi(t)$) and integration of flux ($\rho(t)$) as:

$$i_{in}(t) = \frac{1}{R}\left[\frac{a}{C_1} - \frac{a^2 b\rho(t)}{C_1^2 C_2}\right]\varphi(t) \tag{1}$$

where a and b are parameters that depend on the circuit topology and $\rho(t)$ is integral of flux ($\phi(t)$).

We propose an electronically tunable floating/grounded MIE composed of only 22 MOSFETs and two capacitors as shown in Fig. 2 with three subcircuit stages. The 1st stage consisting of M_1–M_5 transistors is a well-known simple single-ended transconductance stage with a differential input. The M_6–M_7 transistors form the 2nd stage, which is a simple transconductance stage having a single input and single output [20]. The 3rd stage consisting of M_8–M_{22} is a dual output operational transconductance amplifier (DO-OTA) [21]. The transconductances of the three stages are labeled as g_{m1}, g_{m2}, and g_m, respectively. The input is applied between terminals A and B. To operate as a grounded MIE configuration, terminal B is grounded. The input-output relationship of DO-OTA ports is:

$$I_{O+} = -I_{O-} = g_m(V_P - V_N) \tag{2}$$

where $g_m = K(V_B - V_{SS} - V_{th})$ is the transconductance of the 3rd stage, and V_{SS}, V_{th}, and V_B are the source voltage, threshold voltage, and the biasing voltage of the M_{22} transistor. $K = B\mu_n C_{OX}\sqrt{\frac{1}{2}\left(\frac{W}{L}\right)_{M8,9}\left(\frac{W}{L}\right)_{M22}}$ is a constant depending on the carrier mobility, the gate-oxide capacitance, and the aspect ratio of the MOSFETs; B is defined as a gain of the current mirror between $M_{10,\,11}$ and $M_{12,\,13}$ transistors. The input V_{in} is applied between terminals A and B. The output current of the 1st stage is g_{m1} times the input voltage, V_{in}. The output current passes through the capacitor C_1 and generates voltage V_{C1} as:

$$V_{C1} = V_N = \frac{1}{C_1}\int g_{m1}V_{in}dt = \frac{g_{m1}}{C_1}\varphi(t) = V_\varphi \tag{3}$$

where $\varphi(t)$ is flux and V_φ is the flux voltage. The same voltage is applied to the input of 2nd stage and the V_N port of DO-OTA. The current generated at an output terminal of the 2nd stage is g_{m2} times applied voltage V_{C1}. Due to the high impedance input terminal of V_B, the output current of the 2nd stage passes through C_2 and generates V_{C2}. The voltage V_{C2} has applied to biasing voltage V_B of DO-OTA and can be expressed as:

$$V_{C2} = V_B = \frac{1}{C_2}\int\frac{g_{m1}g_{m2}}{C_1}\varphi(t)dt = \frac{g_{m1}g_{m2}}{C_1C_2}\rho(t) = V_\rho \tag{4}$$

where $\rho(t)$ is integral of flux and V_ρ is the integral flux voltage. With the input port of DO-OTA, specifically the gate terminal of M_9, grounded, from Eqn. (2) and Eqn. (3), the output current of DO-OTA can be written as:

$$I_{O+} = -I_{O-} = g_m(V_P - V_N) = g_m\left(0 - \frac{g_{m1}}{C_1}\varphi(t)\right) \tag{5}$$

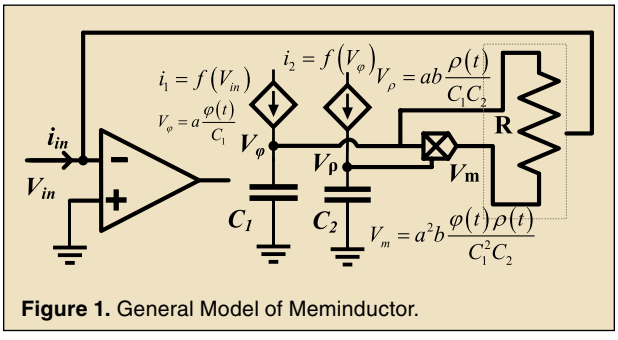


Figure 1. General Model of Meminductor.

Because the inductor's size is directly proportional to the inductance value, the current generation of on-chip inductors is constrained by relatively low inductances.

In Fig. 2, the current I_{O_+} of DO-OTA is equal to the current I_{in}, but it flows in the opposite direction. So, Eqn. (5) can be re-written by substituting the g_m as:

$$I_{\text{in}} = K[V_B - V_{\text{th}} - V_{\text{SS}}]\left(\frac{g_{m1}}{C_1}\varphi(t)\right) \tag{6}$$

By using V_B described in Eqn. (4), we can rewrite Eqn. (6) as:

$$I_{\text{in}} = \frac{Kg_{m1}}{C_1}\left[\frac{g_{m1}g_{m2}}{C_1C_2}\rho(t) - V_{\text{th}} - V_{\text{SS}}\right]\varphi(t) \tag{7}$$

Eqn. (7) can be simplified as:

$$\frac{I_{\text{in}}}{\varphi(t)} = \frac{Kg_{m1}}{C_1}[-V_{\text{th}} - V_{\text{SS}}] + \frac{Kg_{m2}g_{m1}^2}{C_1^2C_2}\rho(t) \tag{8}$$

Eqn. (8) has two parts, the first part is time-invariant, and the second part is time-variant. The transconductance of the 1st stage constituting of M_1–M_5 can be written as [20]:

$$g_{m1} = K_1[V_{\text{Tu}} - V_{\text{th5}} - V_{\text{SS}}] \tag{9}$$

where K_1 is the constant depending on the carrier mobility, the gate-oxide capacitance, and the aspect ratio of the MOSFET, V_{SS} is the source voltage, V_{th5} is the threshold voltage of M_5 and V_{Tu} is the external voltage applied at the gate of M_5. From Eqn. (9), it can be observed that g_{m1} can be adjusted with the voltage V_{Tu}. Therefore, the proposed floating/grounded MIE is electronically tunable and can be tuned by using V_{Tu}.

For frequency analysis, a sinusoidal signal: $V_{\text{in}}(t) = A_m \sin\omega t$ is applied to the proposed MIE. The integration of flux can be calculated by double integration of the input signal as:

$$\rho(t) = \iint V_{\text{in}}(t)dt = -\frac{A_m}{\omega^2}\sin\omega t \tag{10}$$

By substituting $\rho(t)$ in Eqn. (8) with Eqn. (10), we can obtain the meminductor value of the proposed MIE as:

$$\frac{I_{\text{in}}}{\varphi(t)} = \frac{Kg_{m1}}{C_1}[-V_{\text{th}} - V_{\text{SS}}] - \frac{KA_mg_{m2}g_{m1}^2}{\omega^2C_1^2C_2}\sin\omega t \tag{11}$$

Eqn. (11) reflects that the basic meminductor equation expressed in Eqn. (1) and has time-invariant and time-variant parts. The time-variant part is inversely proportional to the product of the frequency squared and the two capacitance values. The time-variant part shows that meminductance varies with frequency.

The parasitic model of the proposed MIE is shown in Fig. 3. The input terminals of all three stages have high impedance, so the parallel combination of resistors and capacitors appears as a parasitic component.

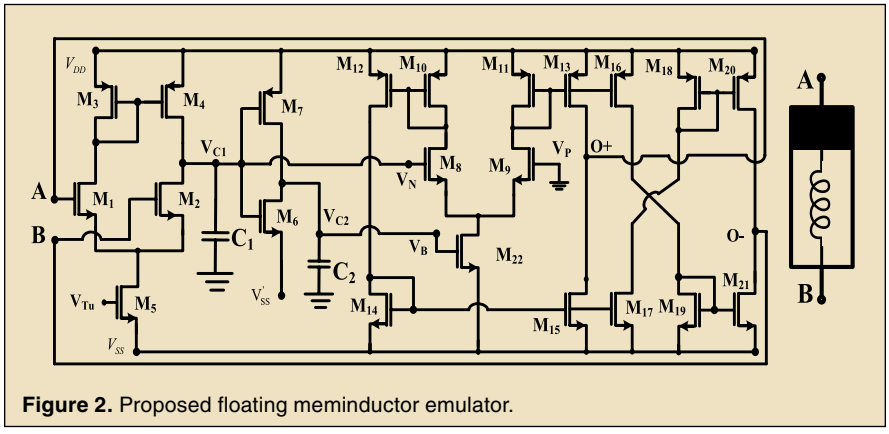


Figure 2. Proposed floating meminductor emulator.

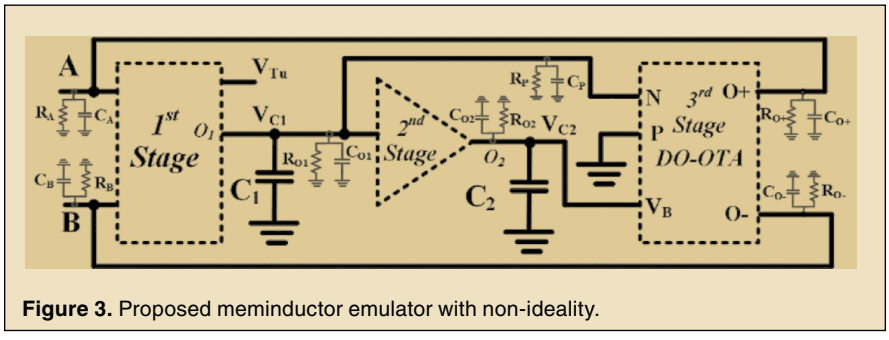


Figure 3. Proposed meminductor emulator with non-ideality.

Meminductive technologies are young and on their way to maturity. Hence, the path to maturity is no less long for the models describing them.

The shunt-connected parasitic resistors R_A, R_B, R_{O1}, R_{O2}, R_{O+}, and R_{O-} and the parasitic capacitances C_A, C_B, C_{O1}, C_{O2}, C_{O+}, and C_{O-} appeared at A, B, O_1, O_2, $O+$, and $O-$ terminals, respectively. At terminal O_1, the effective capacitance becomes the sum of C_1, C_{O1}, and C_P. Likewise, at terminal O_2, the effective capacitance is a sum of C_2 and C_{O2}. The effective resistive parasitic $R_A // R_{O+}$ at terminal A is much greater than the sum of capacitive parasitic $1/s(C_A // C_{O+})$, so the parasitic effect at these terminals can be neglected. Similarly, we can study the effect of the parasitics present at terminal B, which is small and can also be neglected. Due to the parasitics, the effective capacitance at terminal O_1 and O_2 are $C_{1\text{eff}} = C_1 + C_{O1} + C_P$ and $C_{2\text{eff}} = C_2 + C_{O2}$, respectively.

When we consider the influence of the non-idealities, the port relationships of the three stages can be revised as:

$$I_{O1} = \alpha g_{m1}(V_A - V_B) \tag{12}$$

$$I_{O2} = \beta g_{m2} V_{C1} \tag{13}$$

$$I_{O+} = -I_{O-} = \gamma g_m (V_P - V_N) \tag{14}$$

Here, α, β, and γ are tracking errors arising due to non-ideality present at three stages of the proposed MIE. By considering the parasitic and non-ideality, the meminductance of the proposed MIE can be modified as:

$$\frac{I_{\text{in}}}{\varphi(t)} = \frac{K\alpha\gamma g_{m1}}{C_{\text{leff}}}\left[(-V_{\text{th}} - V_{\text{SS}}) + \frac{\alpha\beta g_{m1} g_{m2}}{C_{\text{leff}} C_{2\text{eff}}} \int \varphi(t)dt\right] \tag{15}$$

Eqn. (15) replicates the MIE equation with non-ideality and parasitic effect and has time-invariant and time-variant parts.

III. Simulation Results and Verification

The analysis of the proposed MIE is verified using Cadence Virtuoso simulation with 180 nm CMOS parameters. The supply voltages used for simulation are $V_{\text{DD}} = -V_{\text{SS}} = 0.9$ V and $V'_{\text{SS}} = -0.8$ V. The V_{Tu} is taken as 150 mV and can be varied for tuning. The capacitances C_1 and C_2 are taken as 10 pF and 80 pF, respectively. The functionality of the proposed MIE is tested by applying a sinusoidal signal at different frequencies with a peak amplitude of 200 mV. The MIE operates properly up to 5 MHz. With a lower capacitance value, the proposed MIE can be operated at a higher frequency. The transient current, voltage, and flux plots with respect to time for five cycles are shown in Fig. 4(a) at 2.5 MHz frequency with C_1 = 10 pF and C_2 = 80 pF. A sinusoidal input voltage with 200 mV amplitude is applied at different frequencies to test the frequency-dependent behavior of MIE. The PHL at various frequencies are plotted in Figs. 4(b) and 5(a). Here, the PHL is obtained between the flux and the current in a meminductor (MI). For ease of plotting between flux and current, the voltage V_{C_1} is used instead of flux (ϕ). By using Eqn. (3), we can calculate flux. From Figs. 4(b) and 5(a), it can be concluded that the PHL area decreases as the applied frequency increases. The tunability of the MIE is tested in Fig. 5(b) by applying a varying tuning voltage (V_{Tu}) at 2.5 MHz. The meminductance of the MIE depends on the transconductance (g_{m1}) of the 1st stage, as described in Eqn. (11). By varying g_{m1}, the meminductance of the proposed MIE can be tuned. Therefore, with the voltage V_{Tu}, the transconductance (g_{m1}) can be tuned. For varying bias voltage, from –150 mV to +150 mV, the corresponding PHLs are plotted in Fig. 5(b). The PHLs for different input voltage and capacitance values are shown in Fig. 5(c) and (d), respectively. From

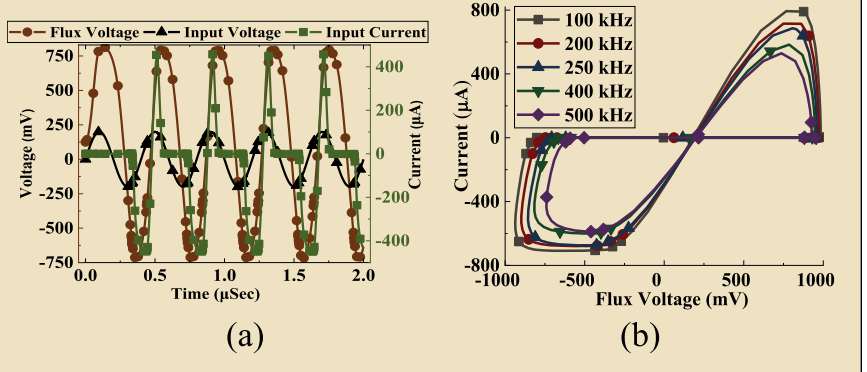


Figure 4. (a) Input voltage, flux voltage, and current for 5 cycles, (b) PHL curve at 100, 200, 250, 400, and 500 kHz.

One of the most important features of the meminductor emulator is the pinched hysteresis loop in the current-flux plane. The lobe area of the pinched hysteresis loop shrinks as the frequency increases, thereby confirming the fingerprints of a meminductor.

Fig. 5(c), we can observe that, with an increase in the input voltage, the PHL area increases as predicted in Eqn. (11). Similarly, Fig. 5(d) shows that the PHL behaviors largely depend on C_1 more than C_2, as shown in Eqn. (11).

To analyze the worst case and the robustness of the MIE against transistor mismatches and process variations, Monte Carlo (MC) and process corner simulations are performed at 2.5 MHz frequency. Fig. 6(a) shows MC-simulated PHL for 100 runs. Fig. 6(a) shows that all PHLs are similar and the meminductor retains its property. Process variation is one of the critical design considerations for monolithic integration. Process corner effects for NMOS and PMOS transistors are observed in PHL as fast-fast (FF), fast-slow (FS), normal-normal (NN), slow-fast (SF), and slow-slow (SS). The process corner outputs are shown in Fig. 6(b). It can be observed from Fig. 6(b) that, as expected the FF mode has the maximum current and the SS mode has the minimum current. Although all modes have different PHL areas, the proposed MIE design exhibits PHLs for all process corners. Moreover, no offset is observed there.

The functionality of the proposed MIE is investigated for different temperatures at 2.5 MHz in Fig 7(a) from 40 °C down to -25 °C. The PHL shown in Fig. 7(a) shows no sharp deviation in the proposed MIE with temperature variation. To perform the non-volatility test, we applied a train of pulses to the input terminal of the proposed MIE at C_1 = 10 pF, C_2 = 80 pF, and V_{Tu} = 150 mV. The pulse train has a 200 mV amplitude with a 50 ns pulse period and a 5 ns pulse width. The simulation results for the non-volatility test of the

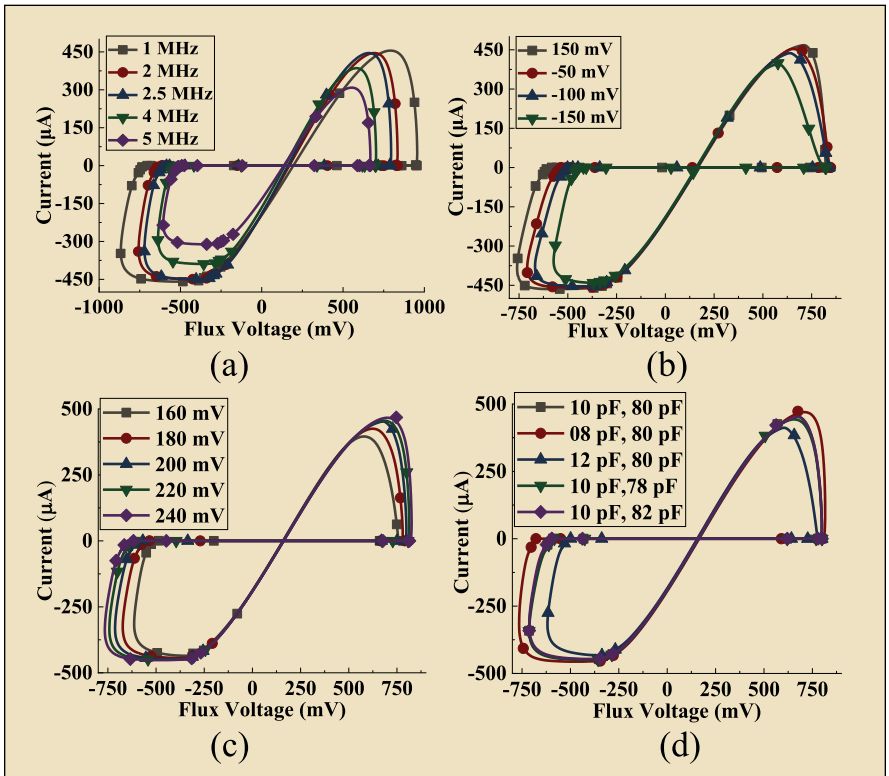


Figure 5. PHL at different (a) Frequencies, (b) V_{Tu}, (c) Input Voltage, (d) C_1 and C_2.

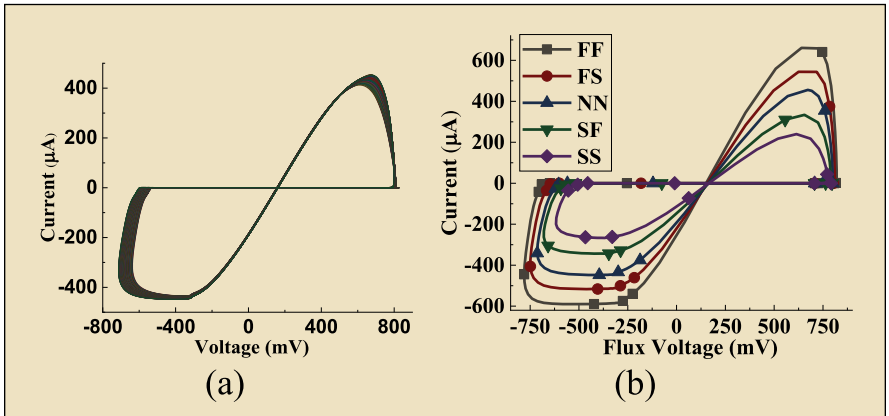


Figure 6. (a) Monte Carlo and (b) Process Corner analysis.

While physical devices are hard to access, awareness of potential challenges and thorough low-level simulations are crucial for a reliable design.

MIE circuit are shown in Fig. 7(b). In addition, Fig. 7(b) shows that the flux voltage virtually does not change when there is no input, but changes when input is applied. Thus, the non-volatile nature of the proposed MIE is verified.

The adaptability of the proposed MIE is tested by using the configurations shown in Fig. 8(a) and by comparing the PHLs of parallel, series-connected emulator circuits, with the same MIE with C_1 = 10 pF, C_2 = 80 pF, V_{Tu} = 150 mV operating at 2.5 MHz. Meminductors are nonlinear and exhibit inductor properties with memory. From Fig. 8(b), we can observe that the two parallel-connected MIEs have the lowest impedance and a higher current compared to a single MIE, and the serially connected emulators have a higher impedance and lower current compared to a single meminductor.

The layout of the MIE circuit is shown in Fig. 9(a). The total layout area of the proposed MIE is 13107.5 μm^2 (122.5 µm × 107 µm). The pre-layout vs. post-layout PHL at 2.5 MHz is shown in Fig. 9(b). From Fig. 9(b), we can observe there is a slight change in the PHL due to the parasitics present in the physical layout. The simulated power dissipation of the MIE is 590 µW.

The proposed MIE model is verified experimentally using commercially available integrated circuits (ICs). The MIE has three subcircuit stages: the first two stages are simple transconductance amplifiers and the third stage is a DO-OTA. Overall, the experimental prototype required four CA3080 ICs. Fig. 10(a) shows the experimental prototype of the MIE. Other than these ICs, some passive components are used to realize the experimental prototype.

The complete experimental setup required an additional digital storage oscilloscope (DSO), DC power supply, and function generator. During the experiment, we applied a sinusoidal signal using a function generator, and the output was observed through DSO. Fig. 10(b) shows the full experimental setup with transient voltages at 6 kHz. The DC power supply during the experiment is ±5 V. The resistance $R_1 = R_2$ and R_B values are 56 kΩ and 1 kΩ, respectively. The capacitances of C_1 and C_2 vary with frequencies. The PHL outputs at 8 kHz and 15 kHz are shown in Figs. 10(c) and 10(d), respectively. In the PHL, some DC shift is present due to non-idealities in discrete components used in the circuit.

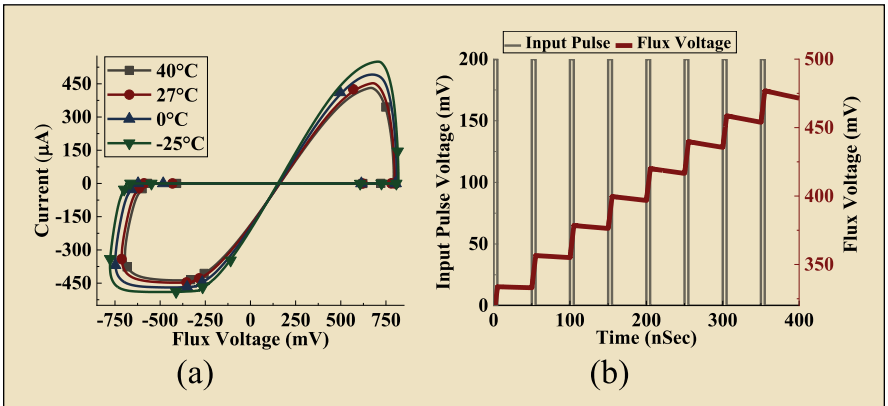


Figure 7. (a) PHL at a different temperature, (b) Non-volatility test.

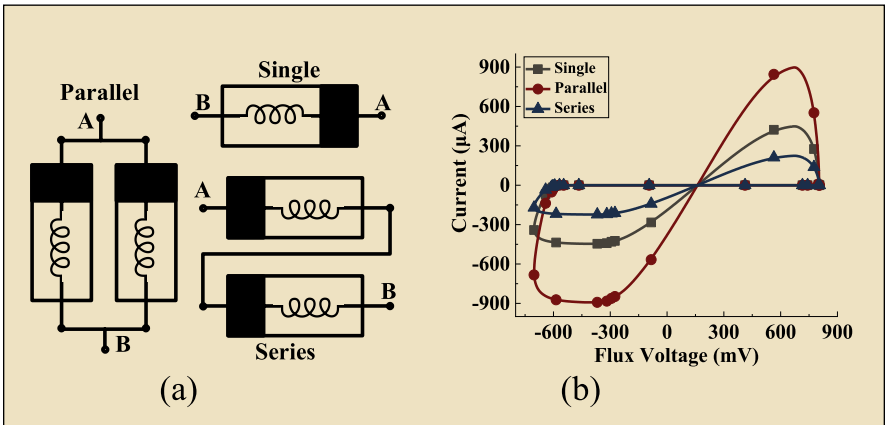


Figure 8. (a) Meminductor connection: Parallel, Single, and Series (b) Corresponding PHL.

IV. Comparison Table

In this section, we have summarized and compared our meminductor emulator (MIE) design with the recently proposed designs in Table 1.

Although it may not accurately represent the behaviors of the proposed model, the current prototype can replicate its operation in a neuromorphic circuit.

Summary and comparisons are made based on the analog blocks used, number of CMOS transistors, floating/grounded configuration, the number of passive components, electronic tunability, power consumption, and the maximum operating frequency.

From Table 1, we can conclude that:

1) Except those in [10], [16], and [17], all other MIEs have operating frequencies less than 3 MHz.
2) The MIE proposed in [10] has a maximum operating frequency of 10 MHz, but it requires 33 MOSFETs while our proposed MIE requires only 22 MOSFETs.
3) MIE presented in [16] operates up to 50 MHz, but it requires 24 MOSFETs, small capacitance value of 5 pF, and its power consumption is 0.86 mW. In comparison, our design uses 22 MOSFETs, higher capacitance values and the power consumption is 0.59 mW.
4) MIE presented in [17] operated up to 25 MHz, but it worked in grounded configuration only.
5) MIEs introduced in [5], [6], [7], [8], [9], [14], [15], and [18] required a passive resistor, while our proposed MIE is resistor-less.
6) An additional analog multiplier is required to obtain a meminductive property in [5], [8], [15], and [18], whereas our proposed design is multiplier-less.
7) MIEs in [8], [9], [10], [11], [15], [17], and [18] are grounded types.
8) MIE in [18] uses an additional inductor.
9) Our proposed MIE design can be tuned by varying the external voltage.

V. Application in Neuromorphic Adaptive Learning

Meminductor can be extensively used in several fields and applications. Although the MIE does not have

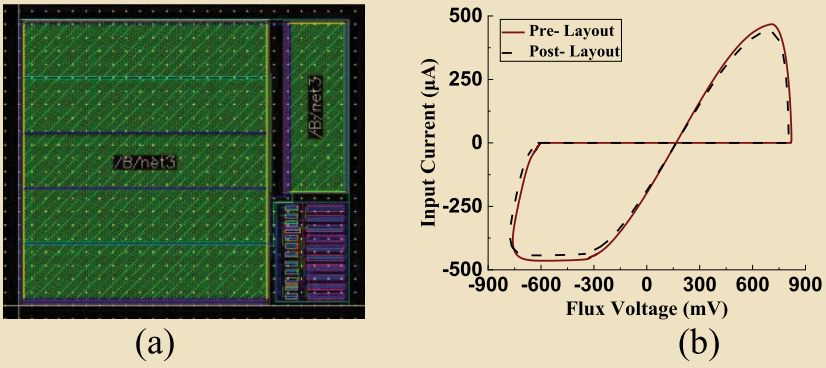


Figure 9. (a) The layout of the proposed MIE, (b) Pre and Post layout PHL at 2.5 MHz.

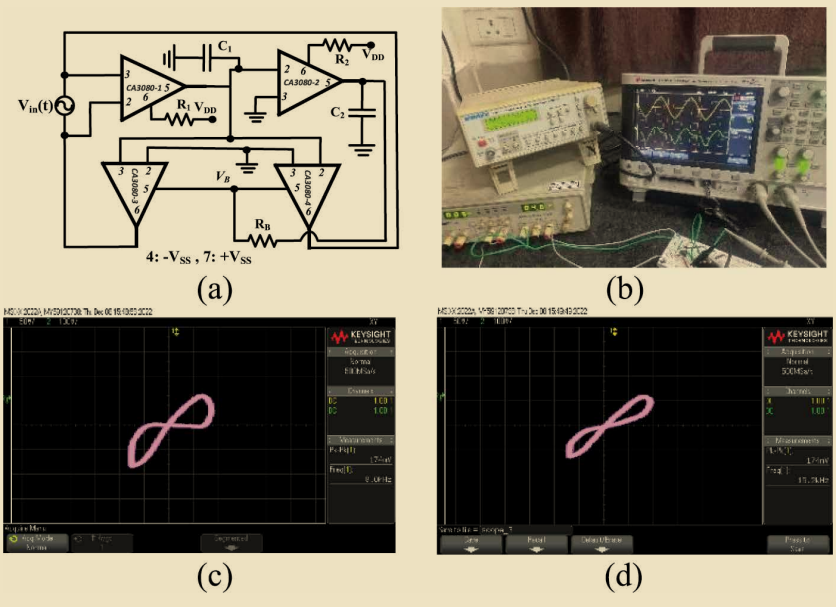


Figure 10. (a) The experimental prototype, (b) The complete setup, (c) PHL at 8 kHz, and (d) PHL at 15 kHz.

Table 1.
Summary of available meminductor emulators and comparison with proposed meminductor emulator circuit.

Ref. & Year	Active Comp. Used	No. of CMOS	Passive Comp.	Supply Voltage	Grounded/ Floating	Electronically Tunability	Process/ Technology	Power Consumption	Operating Frequency
[5] -2020	2 CBTA, 1 Mul[#]	-	2 R, 2 C	-	Both	Yes	180 nm	NA	250 kHz
[6] -2021	2 VDCC,	22 + 2 CS[*]	1 R, 1 C	±0.9 V	Both	No	180 nm	NA	700 kHz
[7] -2020	2 VDTA	32	1 R, 2 C	±0.9 V	Both	Yes	180 nm	NA	1 MHz
[8] -2020	2 OTA, 1 Mul[#]	-	2 R, 2 C	±1.25 V	Grounded	Yes	180 nm	2.25 mW	10 kHz
[9] -2020	2 CCII, 1 OTA	43	2 R, 2 C	±1.2 V	Grounded	No.	180 nm	14.36 mW	700 kHz
[10] -2021	3 OTA	33	2 C	±1.2 V	Grounded	Yes	180 nm	0.12 mW	10 MHz
[11] -2021	1 VDTA, 1 OTA	27	2 C	±1.2 V	Grounded	Yes	180 nm	NA	3 MHz
[12] -2021	2 VDTA	32	2 C	±0.9 V	Both	Yes	180 nm	NA	1.5 MHz
[13] -2022	2 OTA, 1 CDBA	37 + 4 CS[*]	2 C	±0.9 V	Both	Yes	180 nm	NA	2 MHz
[14] -2022	1 MVDCC, 1 OTA	39	1 R, 2 C	±0.9 V	Both	Yes	180 nm	NA	300 kHz
[15] -2022	2 DDCC, 1 Mul[#] 2 MOSFETs	-	2 R, 1 C	±1.5 V	Grounded	Yes	350 nm	NA	1.5 kHz
[16] -2022	1 MO-VDTA	24	2C	±0.9 V	Both	Yes	180 nm	0.86 mW	50 MHz
[17] -2023	1 VDTA, 2 MOSFETs	18	2 C	±0.9 V	Grounded	Yes	180 nm	5.93 mW	25 MHz
[18] -2023	1 DO-CCII, 1 CCII, 1 Mul[#]	-	2R, 2C, 1L	±12 V	Grounded	No	-	NA	700 kHz
[19] -2023	1 MO-OTA, 1 OTA	34	2 C	±0.9 V	Both	Yes	180 nm	NA	900 kHz
This MIE	**1 DO-OTA, 7 MOSFETs**	**22**	**2 C**	**±0.9 V**	**Both**	**Yes**	**180 nm**	**0.59 mW**	**5 MHz**

[#]Mul: Multiplier, [*]CS: Current Source

permanent memory like RAM, it offers short-term memory and is capable of in-memory computation. Therefore, this aspect of the meminductor can be harnessed in the field of adaptive neuromorphic circuits. The unicellular organism (amoeba) with its decision-making capability can adjust its locomotive speed depending on the ambient temperature of the surroundings.

An attempt to use the proposed MIE with a view of performing the brain-like functionality inspired by the amoeba adaptive behavior is explored in this application. The electronic analogous model of the neuromorphic adaptive learning circuit realized in Fig. 11(a) is derived from the behavioral response of amoeba [22]. The input voltage (V_{in}), that drives the amoeba's locomotion is analogous to the temperature and humidity, while the output voltage (V_{out}) across the capacitor is analogous to the locomotion speed of the amoeba. In an RLC circuit, the resistor R exclusively uses energy, and the capacitor C or the inductor L stores energy in the form of an electric and magnetic field, respectively. Within the circuit, energy can move from one form to another and this movement can be oscillatory, leading to resonance. The oscillations of the described RLC circuit will be used to emulate the actions of an amoeba.

In this application, a signal at a specific frequency ($f_r = 1/2\pi\sqrt{LC}$) is extracted from a more complex signal using RLC circuits. The resistance (R) causes damping in the RLC circuit and determines the resonance nature. Some dissipation and impedance are believed to be present inside amoeba; otherwise, input signals would pass instantaneously without any attenuation. Apart from resistor (R) and capacitor (C), a flux-controlled meminductor emulator (MIE) is used to scan a range of

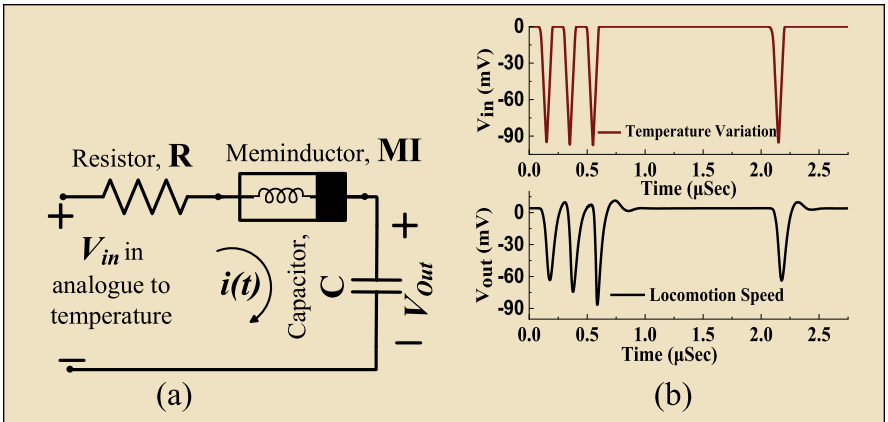


Figure 11. (a) Neuromorphic circuit, (b) Locomotion adaptive behavior with input temperature variation.

input frequencies. The inductance (L) of the meminductor is flux controlled ($L(\int \varphi dt)$). The RLC circuit scans a range of frequencies and generates resonance frequency:

$$f_r = 1/2\pi\sqrt{LC} = 1/2\pi\sqrt{L\left(\int \varphi dt\right)C} \tag{16}$$

An MIE-based circuit has a variable time constant $\sqrt{L\left(\int \varphi dt\right)C}$, which naturally traces the applied stimuli based on the delayed switching effect. When the variable resonance frequency (f_r) equals the (temperature) stimulus frequency (f_s), a resonance will be triggered. At this point, the circuit is locked and the learning of a regular event is accomplished.

$$f_r = 1/2\pi\sqrt{L\left(\int \varphi dt\right)C} = f_s \tag{17}$$

The response of the adaptive learning circuit shown in Fig. 11(a) is expressed by the following set of equations:

$$V_{\text{out}}(t) + \frac{d[L(\varphi)\cdot i(t)]}{dt} + i(t)\cdot R = V_{\text{in}}(t) \tag{18}$$

$$C.\frac{dV_{\text{out}}(t)}{dt} = i(t) \tag{19}$$

where $V_{\text{out}}(t)$, $V_{\text{in}}(t)$, and $i(t)$ are the output voltage, input voltage, and total current, respectively.

The simulation results for locomotion adaptive learning behavior with respect to changes in temperature are shown in Fig. 11(b). From Fig. 11(b), we can observe that at each drop in temperature, the locomotion speed also drops accordingly. Initially, the learning process of slowing down the locomotion speed takes time and the temperature drops three times during that period, after which the movement becomes slow. Interestingly, due to the learning capability of amoeba, in the next incident, the locomotion speed starts to slow down immediately after a temperature change. This shows that the neuromorphic circuit using a meminductor emulator can learn adaptively.

VI. Conclusion

In this article, a compact and novel MIE comprising only 22 MOSFETs and two capacitors is introduced and its functionality is fully validated analytically, by simulation and through experiment. This MIE offers the following advantages: simple and novel design; operational for both grounded and floating configurations; no need for any analog multiplier; high-speed operation in the MHz range; electronically tunable; and low power consumption. The proposed MIE is compact in terms of the number of circuit elements, namely MOSFETs and capacitors. It operates in the high MHz range. An adaptive learning circuit is designed using the proposed MIE to mimic closely the behavioral response of the unicellular organism amoeba over various environmental conditions.

Acknowledgment

This work was supported in part by the Department of Electronics Engineering, IIT (ISM), India, through Projects SICRG, under Grant MHRD [Shastri Institutional Collaborative Research Grant (SICRG)]/2020-2021/740/ECE; in part by the MeitY (Ministry of Electronics and Information Technology) entitled "Ultra-Low Power Neuromorphic Spiking Architecture for Assistive Smart Glass" under Grant MietY/2023-2024/1012/ECE; and in part by a Council of Scientific & Industrial Research (CSIR) project entitled "Memristor Based Multilayer Neural Networks (MNN) and its application in Neuromorphic System" under Grant 70/0081/23/EMR-II (Institute Project no. CSIR(40)/2023-2024/1045/ECE). Corresponding author: Rajeev Kumar Ranjan (e-mail: rkranjan2k@gmail.com; rajeev@iitism.ac.in).

Pankaj Kumar Sharma was born in India, in 1989. He received the Bachelor of Technology (B.Tech.) degree in electronics and communication engineering from Nilamber Pitamber University Jharkhand, India, in 2013, the Master of Technology (M.Tech) degree in VLSI design automation and techniques from the Department of Electronics and Communications Engineering, National Institutes of Technology Hamirpur, India, in 2016, and the Doctor of Philosophy (Ph.D.) degree in electronics engineering from the Indian Institute of Technology (Indian School of

Mines), Dhanbad, India, in 2023. He has been associated as a Faculty with the Department of Electronics and Communication Engineering, National Institute of Technology Srinagar, India, since October 2023. He has published over ten articles in esteemed journals and conferences in these areas. His research interests include microelectronics and VLSI design, current mode circuit, mem-elements (memristor, memcapacitor, meminductor) emulator design, memristive systems, and MEMS design.

Rajeev Kumar Ranjan (Senior Member, IEEE) was born in India, in 1979. He received the Master of Technology (M.Tech) and Doctor of Philosophy (Ph.D.) degrees from the Indian Institute of Technology (Indian School of Mines), Dhanbad, India, in 2007 and 2016, respectively. He is currently an Associate Professor with the Department of Electronics Engineering, Indian Institute of Technology (Indian School of Mines). He has authored more than 40 articles in esteemed journals and more than 15 papers in esteemed conferences. His current research interests include mem-elements (memristor, memcapacitor, meminductor) emulator design, and memristive systems, neuromorphic computing, low-power VLSI design, and VLSI signal processing circuits and systems. He currently serves as an Associate Editor for *Journal of Circuit Systems and Computer.*

Sung-Mo (Steve) Kang (Life Fellow, IEEE) received the B.S. degree (summa cum laude) from Fairleigh Dickinson University, Teaneck, NJ, USA, the M.S. degree from The State University of New York, Buffalo, and the Ph.D. degree from the University of California at Berkeley. He is currently a Distinguished Professor Emeritus and a Research Professor of the Jack Baskin School of Engineering, University of California at Santa Cruz, Santa Cruz. He was the 15th President of Korea Advanced Institute of Science and Technology (KAIST), the 2nd Chancellor of the University of California at Merced, Merced, the Dean of Engineering of the University of California at Santa Cruz, and the Head of the Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign. He holds 16 U.S. patents, has authored ten books, and published over 500 journals and conference papers. His current research interests include memristors and memristive systems, neuromorphic computing, low-power VLSI design, and computer-aided design. He is a fellow of ACM and AAAS.

References

[1] M. Di Ventra, Y. V. Pershin, and L. O. Chua, "Circuit elements with memory: Memristors, memcapacitors, and meminductors," *Proc. IEEE*, vol. 97, no. 10, pp. 1717–1724, Oct. 2009.

[2] P. K. Sharma et al., "Charged controlled mem-element emulator and its application in a chaotic system," *IEEE Access*, vol. 8, pp. 171397–171407, 2020.

[3] Y. V. Pershin and M. Di Ventra, "Memristive circuits simulate memcapacitors and meminductors," *Electron. Lett.*, vol. 46, no. 7, p. 517, 2010.

[4] D. Yu et al., "A universal mutator for transformations among memristor, memcapacitor, and meminductor," *IEEE Trans. Circuits Syst. II, Ex. Briefs*, vol. 61, no. 10, pp. 758–762, Oct. 2014.

[5] Z. G. Ç. Tas¸kıran et al., "A new universal mutator circuit for memcapacitor and meminductor elements," *AEU- Int. J. Electron. Commun.*, vol. 119, May 2020, Art. no. 153180.

[6] A. Singh and S. K. Rai, "VDCC-based Memcapacitor/Meminductor emulator and its application in adaptive learning circuit," *Iranian J. Sci. Technol., Trans. Electr. Eng.*, vol. 45, no. 4, pp. 1151–1163, Dec. 2021.

[7] J. Vista and A. Ranjan, "High frequency meminductor emulator employing VDTA and its application," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 39, no. 10, pp. 2020–2028, Oct. 2020.

[8] M. Konal and F. Kacar, "Electronically tunable meminductor based on OTA," *AEU-Int. J. Electron. Commun.*, vol. 126, Nov. 2020, Art. no. 153391.

[9] N. Raj et al., "Mem-elements emulator design with experimental validation and its application," *IEEE Access*, vol. 9, pp. 69860–69875, 2021.

[10] A. Raj, K. Kumar, and P. Kumar, "CMOS realization of OTA based tunable grounded meminductor," Anal. Integr. Circuits Signal Process., vol. 107, no. 2, pp. 475–482, May 2021.

[11] K. Kumar and B. C. Nagar, "New tunable resistorless grounded meminductor emulator," *J. Comput. Electron.*, vol. 20, no. 3, pp. 1452–1460, Jun. 2021.

[12] K. Bhardwaj and M. Srivastava, "New electronically adjustable memelement emulator for realizing the behaviour of fully-floating meminductor and memristor," *Microelectron. J.*, vol. 114, Aug. 2021, Art. no. 105126.

[13] N. Yadav, S. K. Rai, and R. Pandey, "New high frequency memristorless and resistorless meminductor emulators using OTA and CDBA," *Sa¯dhana¯*, vol. 47, no. 1, pp. 1–8, Mar. 2022.

[14] K. Bhardwaj and M. Srivastava, "New grounded passive elements-based external multiplier-less memelement emulator to realize the floating meminductor and memristor," *Anal. Integr. Circuits Signal Process.*, vol. 110, no. 3, pp. 409–429, Mar. 2022.

[15] K. Orman, A. Yesil, and Y. Babacan, "DDCC-based meminductor circuit with hard and smooth switching behaviors and its circuit implementation," *Microelectron. J.*, vol. 125, Jul. 2022, Art. no. 105462.

[16] P. B. Petrovic´, "A new electronically controlled floating/grounded meminductor emulator based on single MO-VDTA," *Anal. Integr. Circuits Signal Process.*, vol. 110, no. 1, pp. 185–195, Jan. 2022.

[17] P. K. Sharma, S. Tasneem, and R. K. Ranjan, "A new electronic tunable high-frequency meminductor emulator based on a single VDTA," *IEEE Can. J. Electr. Comput. Eng.*, vol. 46, no. 2, pp. 179–184, Spring 2023.

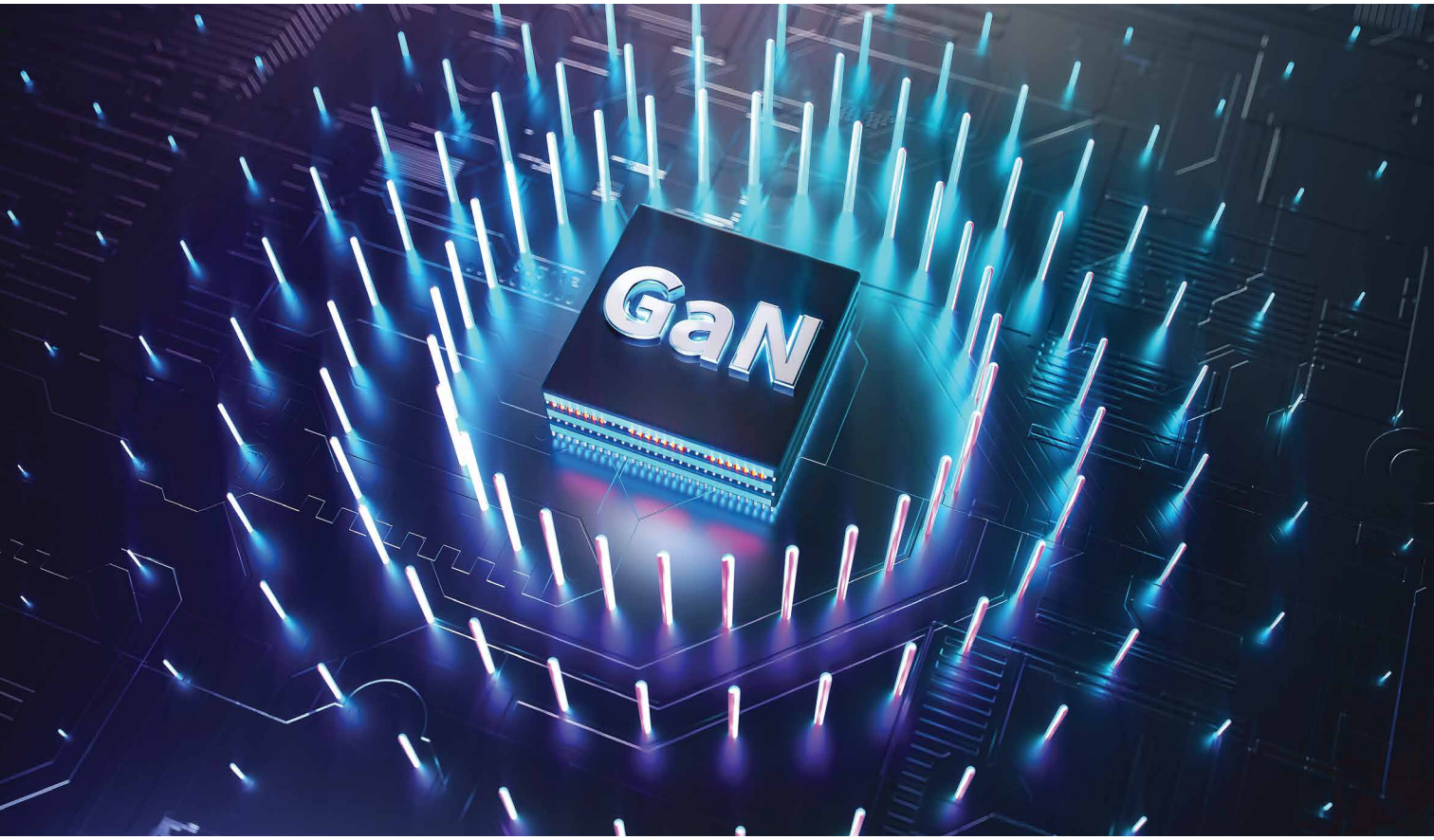
[18] M. O. Korkmaz, Y. Babacan, and A. Yesil, "A new CCII based meminductor emulator circuit and its experimental results," *AEU-Int. J. Electron. Commun.*, vol. 158, Jan. 2023, Art. no. 154450.

[19] B. Aggarwal, S. K. Rai, and A. Sinha, "New memristor-less, resistorless, two-OTA based grounded and floating meminductor emulators and their applications in chaotic oscillators," *Integration*, vol. 88, pp. 173–184, Jan. 2023.

[20] A. Yesil, "A new grounded memristor emulator based on MOSFET-C," *AEU-Int. J. Electron. Commun.*, vol. 91, pp. 143–149, Jul. 2018.

[21] A. Yesil, Y. Babacan, and F. Kacar, "An electronically controllable, fully floating memristor based on active elements: DO-OTA and DVCC," *AEU-Int. J. Electron. Commun.*, vol. 123, Aug. 2020, Art. no. 153315.

[22] F. Z. Wang et al., "Adaptive neuromorphic architecture (ANA)," *Neural Netw.*, vol. 45, pp. 111–116, Sep. 2013.



Chiplet-GAN: Chiplet-Based Accelerator Design for Scalable Generative Adversarial Network Inference

Yuechen Chen, *Member, IEEE*, Ahmed Louri, *Fellow, IEEE*, Fabrizio Lombardi, *Fellow, IEEE*, and Shanshan Liu, *Senior Member, IEEE*

Digital Object Identifier 10.1109/MCAS.2024.3359571
Date of current version: 15 August 2024

Abstract

Generative adversarial networks (GANs) have emerged as a powerful solution for generating synthetic data when the availability of large, labeled training datasets is limited or costly in large-scale machine learning systems. Recent advancements in GAN models have extended their applications across diverse domains, including medicine, robotics, and content synthesis. These advanced GAN models have gained recognition for their excellent accuracy by scaling the model. However, existing accelerators face scalability challenges when dealing with large-scale GAN models. As the size of GAN models increases, the demand for computation and communication resources during inference continues to grow. To address this scalability issue, this article proposes Chiplet-GAN, a chiplet-based accelerator design for GAN inference. Chiplet-GAN enables scalability by adding more chiplets to the system, thereby supporting the scaling of computation capabilities. To handle the increasing communication demand as the system and model scale, a novel interconnection network with adaptive topology and passive/active network links is developed to provide adequate communication support for Chiplet-GAN. Coupled with workload partition and allocation algorithms, Chiplet-GAN reduces execution time and energy consumption for GAN inference workloads as both model and chiplet-system scales. Evaluation results using various GAN models show the effectiveness of Chiplet-GAN. On average, compared to GANAX, SpAtten, and Simba, the Chiplet-GAN reduces execution time and energy consumption by 34% and 21%, respectively. Furthermore, as the system scales for large-scale GAN model inference, Chiplet-GAN achieves reductions in execution time of up to 63% compared to the Simba, a chiplet-based accelerator.

Index Terms—GAN inference, chiplet, scalability, interconnection network, machine learning.

I. Introduction

Generative Adversarial Networks (GANs) have excellent performance in content generation, such as image synthesis and natural language processing (NLP) tasks [1], [2], [3], [4]. With the most recent advancement in training algorithms and models, the latest large-scale GAN models have achieved very good performance [5], [6], [7], [8], [9], [10], [11]. For example, TransGAN [6] has shown excellent accuracy on image synthesis tasks by incorporating transformers into the model. To deploy these GAN models, inference is a quintessential step for the model to generate content with reduced delay [12]. The increasing complexity and size of the large-scale GAN model for capturing more complex patterns and generating higher-quality outputs present challenges in architectural design when it comes to handling the demanding computational and communication workloads associated with model inference.

Existing research has proposed several architectures to accelerate GAN inference [12], [13], [14], [15], [16]. Different from convolutional neural networks, a GAN contains transposed convolution operations (TC) and transformers [17], which demand a significant amount of computation and communication resources. Thus, specialized accelerators are developed to handle these workloads. In GANAX [12], a SIMD-MIMD mixed architecture is developed to handle TC. In SpAtten [13], a specialized unit is designed to accelerate the computation of attention mechanisms in transformers. Existing architectures primarily excel in improving efficiency for specific functionalities, such as TC or the attention mechanism. However, as the parameters and the number of layers increase to expand the widespread usage and accuracy of GAN models, the task of scaling these architectures to meet the growing computational and communication demands becomes an increasingly pressing concern. All existing accelerators [12], [13], [14], [15] are implemented with a global controller and global buffer, which are not scalable to support more computation units for the inference of large-scale GAN models.

Recent research has provided compelling evidence of the advantages of chiplet systems in terms of computation performance and scalability [18], [19], [20], [21]. A chiplet system provides a modular approach to semiconductor design and manufacturing, in which multiple smaller chip components (i.e., chiplets) are integrated into a single system. The chiplets are interconnected via network-on-packet (NoP) with either active or passive links on a common interposer [22], [23], [24]. Within each chiplet, a multi-core system is implemented by using network-on-chip (NoC) to connect cores, caches, and memory interfaces. Packets are used to carry data in NoCs, and routers are implemented to direct packets to their destination. With such a design, the system offers scalability by allowing users to scale their systems by adding chiplets as needed. Existing research and commercial products have proven that these systems can integrate a large number of computation units at a lower cost and higher flexibility compared to conventional single-chip designs [20], [22], [24], [25]. The flexibility and scalability are particularly useful in GAN model inference, in which varying levels of performance and functionality are required.

To the best of the authors' knowledge, there has been no prior research conducted on designing a scalable

Yuechen Chen is with the Department of Computer Science and Information Technologies, Frostburg State University, Frostburg, MD 21532 USA (e-mail: ychen@frostburg.edu).
Ahmed Louri is with the Department of Electrical and Computer Engineering, The George Washington University, Washington, DC 20052 USA (e-mail: louri@gwu.edu).
Fabrizio Lombardi is with the Department of Electrical and Computer Engineering, Northeastern University, Boston, MA 02115 USA (e-mail: lombardi@ece.neu.edu).
Shanshan Liu is with the School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: ssliu@uestc.edu.cn).

GAN inference accelerator by employing a chiplet design approach. Existing research has proposed the use of a chiplet system for deep convolutional neural networks (DCNN), which mainly consists of convolutional layers [18], [19]. Different from DCNN models, TC and transformers require more computation resources with vastly different communication patterns. Compared to existing single-chip accelerator designs [12], [13], [14], [15] for GAN inference that can efficiently process only one type of layer (e.g., matrix multiplication or attention mechanism), a chiplet system excels at efficiently processing multiple layers in parallel. Furthermore, the scalability inherent to a chiplet design approach ensures efficient inference not only for existing large-scale GAN models but also for future models with increasingly demanding computation and communication requirements.

This article presents Chiplet-GAN, a chiplet-based accelerator designed specifically for scalable and efficient GAN inference. Through a detailed analysis of the computation and communication requirements involved in GAN inference, it is observed that operations such as matrix multiplication, reduction operations, up-sampling, and matrix reshape operations consume a significant portion of the inference time and incur intensive communication and computation. Recognizing the distinct communication patterns exhibited by these operations, an adaptive network topology is developed to effectively address the communication demands of the workload. This adaptive interconnection network is implemented in each chiplet, enabling dynamic switching between a concentrated mesh (C-Mesh) and a mesh topology based on the functionality of the layer. For efficient communication between chiplets as Chiplet-GAN scales, NoP is implemented with both passive and active network links on the interposer to efficiently handle both short-distance and long-distance communications. Together with the NoP design and adaptive interconnection network, the efficiency of traffic forwarding is improved during GAN inference, specifically for the aforementioned operations. To complement the adaptive network topology and passive/active network links, the workload partitioning and allocation algorithms are developed. These algorithms partition the workload and strategically distribute the workloads to ensure low communication latency, which further reduces the overall execution time. The key contributions of this article are then summarized as follows.

- This article proposes Chiplet-GAN, a scalable and efficient GAN inference accelerator, by employing a chiplet design approach.
- Chiplet-GAN includes an adaptive interconnection network in chiplet to ensure low communication latency during inference as the GAN model scales.
- Network-on-packet (NoP) is designed with both active and passive network links on the interposer with a workload partitioning and allocation algorithm to ensure low communication latency between chiplets as the system scales.
- Simulation results show that compared to GANAX, SpAtten, and Simba, Chiplet-GAN reduces execution time and energy consumption by 34% and 21%, respectively. Furthermore, as the system scales for large-scale GAN inference, Chiplet-GAN achieves reductions in execution time of up to 63% compared to the existing chiplet-based accelerator (Simba).

II. Motivation and Challenges

A. GAN Inference

The recent large-scale GAN model for inference poses challenges mainly due to the deep model architecture with a large number of parameters and activations. The main challenges of the scaling of the layers that are widely used in GAN models are listed as follows.

- **Fully Connected/Linear/Convolution Layer:** These layers involve matrix multiplications between the input activations and the learned weight matrices. These matrix multiplication operations are computationally expensive, especially when large input activations and weights are used in the model.
- **Attention Mechanism:** The attention mechanism, commonly used in transformer-based GANs, requires pairwise computations between all positions in the input activations. This involves computing attention weights for each position, which entails a significant amount of computation due to many matrix multiplications. Also, the transformer involves matrix reshape operations, such as matrix transpose, and incurring extensive communication traffic between computation units.
- **Up-Sampling Operations:** GAN models utilize up-sampling operations, such as zero insertion or nearest-neighbor up-sampling. Up-sampling operations involve interpolating or expanding input activations to increase the spatial resolution, so requiring extensive data communication between computation units.
- **Activation Functions:** The application of non-linear activation functions, such as *ReLU* (Rectified Linear Unit), and linear activation functions, such as *SoftMax* and normalization in the model, also contributes to the complexity of computation and communication. These activation functions

are applied elementwise to the activations and require additional computation. The normalization and *SoftMax* functions need to survey the inputs before an element-wise operation, that requires a reduction operation prior to an elementwise operation.

As the GAN model keeps scaling to capture more complex patterns and generate higher-quality outputs, efficient inference in such a model requires more communication and computation capabilities for storing, accessing, and processing input activations and parameters. Thus, a scalable system is needed to handle the increasing demand in both computation and communication for GAN inference.

B. Chiplet Design Approach

Chiplet systems offer significant advantages in terms of performance and scalability for architecture design and semiconductor manufacturing [18], [19], [20], [21], [25]. By integrating specialized chip components, known as chiplets, into a unified system, designers can better utilize each chiplet for specific tasks, resulting in enhanced performance. The modular nature of chiplet systems further enables scalability, allowing for nearly effortless addition or removal of chiplets as required. This flexibility allows system upgrades for large-scale GAN model inference workloads.

To support easy system upgrade and scaling, a network-on-packet (NoP) is implemented to connect chiplets. As shown in Fig. 1, the connection between chiplets is accomplished with a passive or active network link via micro-bump (μbump). Recent work [24] has studied both designs by analyzing the positive and negative of the two designs. Active network links enable better scalability compared to passive network links with larger throughput for long-distance communication. However, implementing active network links needs routers being implemented in the interposer, which requires the packet to traverse routers in both chiplet and interposer. On the other hand, passive network links achieve lower network latency by only relying on the router in the chiplet, even though the link has lower bandwidth.

Within each chiplets, processing elements (PEs), caches, and memory interfaces are connected using NoC. All the data is transmitted using packets in both NoC and NoP. Overall, chiplet systems present a promising approach to improve performance, efficiency, and scalability in GAN inference accelerator designs.

As chiplet systems rely heavily on the interconnection network to connect and communicate between different computation units and memories, designing an efficient interconnection network poses a major challenge for efficient and scalable GAN inference. Energy and latency represent two vital aspects when designing interconnect solutions for efficient GAN inference. With the addition of more chiplets to accommodate the expanding GAN model, the NoP inevitably becomes larger; consequently, the traversal of packets through the expanded network incurs higher time and energy costs. According to the existing research [18], [19], [20], [24], [25], network latency and energy consumption significantly impact inference efficiency with longer execution time and high energy consumption. Thus, reducing latency and energy consumption for interconnection networks is crucial for GAN inference on a chiplet system, which often requires real-time responsiveness during deployment. Specifically, the main challenges for designing a chiplet system for GAN inference are listed below.

- **Diverse Communication Pattern:** Inference involving various layers in GAN models results in markedly diverse communication patterns. Given the significant effects of latency and power consumption when a packet is redirected by the routers, there's a pressing need for innovative NoC designs, which should efficiently address the distinct communication patterns observed during GAN inference within a chiplet.
- **Scalable and Efficient NoP:** The communication between chiplets is a major bottleneck for chiplet-based design as the system scales. One major challenge in NoP design is the usage of passive and active network links. Existing research [24] indicates that passive links can handle high data throughput over short distances. However, their scalability is constrained by limited communication range. Conversely, active links facilitate long-distance communication, but their bandwidth is restricted due to the incorporation of routers on the interposer. Considering the benefits and limitations of both active and passive

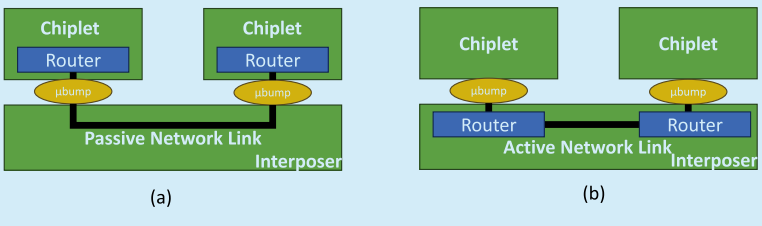


Figure 1. NoP design with a passive or active network link. (a) NoP with passive network link. (b) NoP with active netwrok link.

links, a novel NoP design, grounded in a strategic analysis of GAN inference workloads, is crucial for the efficiency of the chiplet-based accelerator. Additionally, the network-on-package (NoP) should be synergistically designed alongside workload partitioning and allocation algorithms, aiming to harness the strengths of both interposer types.

III. Chiplet-GAN Design

A. Overview

The goal of the proposed accelerator design is to attain both improvements in efficiency and scalability for GAN inference workloads by employing a chiplet design approach. This is mainly accomplished through the following steps.

- A thorough analysis of the computation and communication requirements specific to the GAN model is conducted. This analysis involves examining the architecture of various GAN models used for image synthesis workloads and meticulously assessing the computation and communication demands of each layer as well as the overall model.
- Chiplet-GAN is proposed, featuring an adaptive network topology design for the interconnection network to facilitate the scaling of individual layers.
- Workload partition and allocation algorithms are introduced to effectively utilize the benefits of the adaptive topology design when dealing with large-scale GAN models.

B. Layer Analysis

Table 1 shows the list of large-scale GAN models used in this article. These models contain tens of millions of parameters and multiple layers to generate high-resolution images depending on the configuration. Additionally, newer variations of these models contain more parameters to achieve better image resolution. The basic structure of these models includes both transposed convolution (TC) and transformers, which are the two design trends for the GAN model. The TC-based model includes convolution, up-sampling functions, normalization functions, and activation functions. The transformer-based model contains attention mechanisms, linear layers, up-sampling functions, normalization functions, and activation functions. The functionality of these layers consists of several basic operations, including matrix multiplication, reduction operation, and up-sampling/matrix reshape operations. Communication and computation are needed when computation units execute these basic operations. The analysis of the computation and communication requirements for these basic operations is presented next.

Table 1.
GAN models.

GAN Models	Basic Structure	Number of Parameters	Task
BigGAN-deep [9]	TC	50 million	Image synthesis
ViTGAN-Base [11]	Transformer	38 million	Image synthesis
TransGAN-Base [6]	Transformer	85 million	Image synthesis
StyleGAN2 [5]	TC	24 million	Text-to-Image
CycleGAN [7]	TC	35 million	Image-to-Image Translation

1. Matrix Multiplication

Fully connected, linear, and convolution layers are widely used in both transformer and TC-based GANs. In a transformer, the linear layers are widely used before the attention mechanism. Also, the transformer utilizes matrix multiplication to calculate attention values. In TC-based GANs, the linear layer is used at the end of each processing block to capture the complex relationship between input and output data. Eq. (1) shows the basic operation for a fully connected layer and linear layer, i.e., matrix multiplication. The convolution layer can also be accomplished using matrix multiplication by reshaping the input activation and weight matrices.

$$O_{mp} = A_{mn} * B_{np} \tag{1}$$

Eq. (1) shows two input matrices (i.e., A, B) and one output matrices (i.e., O). m, n denotes the number of rows and columns of matrix A, and n, p denotes the number of rows and columns of matrix B. To perform matrix multiplication, the computation unit needs to load matrices A and B, which requires data communication of $m * n + n * p$ elements. Then, the matrix multiplication requires n $* m * p$ multiplications followed by $(n-1) * m * p$ addition operations. After the computation, the computation unit generates and sends out the matrix O with $m * n$ elements.

2. Reduction Operation

Both *SoftMax* and the normalization layer scale the input matrix into a specific format based on the statistics of the input activations. Since these layers improve the accuracy of the model, they are widely used in both

transformer and TC-based GANs. These layers typically contain two steps. The first step calculates the statistics of the input, which involves a reduction operation. Eqs. (2) and (3) show the function for statistic computation for *SoftMax* and normalization operations.

$$s = \sum_{j=1}^{i} e^{x_j} \tag{2}$$

$$\begin{cases} \mu = \frac{1}{i}\sum_{j=1}^{i} x_j \\ \sigma^2 = \frac{1}{i}\sum_{j=1}^{i} (x_j - \mu)^2 \end{cases} \tag{3}$$

In Eqs. (2) and (3), the input matrix x contains a total of i elements. Both functions need to access all elements in the input matrix to get the statistics. Assume the input matrix has a size of $m * n$. *SoftMax* needs data communication of $m * n$ elements to gather all inputs, so $m * n$ exponent computations and $m * n - 1$ additions to get the variable s. For the normalization layer, $m * n$ elements are gathered to calculate the mean value of the inputs through $m * n - 1$ addition and 1 division operations. Then, the variance is calculated through $m * n$ subtraction and square operations, followed by $m * n - 1$ additions and one division. Both of these layers require a reduction operation, which requires data communication, to gather the input activation and calculate the statistics. After the calculation, the element-wise operations are applied to the input matrix.

3. Up-Sampling/Matrix Reshape Operation

Up-sampling operations are widely used in GANs to increase the size of the output matrix, such as zero insertion and nearest neighbor. These layers require a matrix with $m * n$ elements as the input and upscale it a time, which requires $m * n + a * m * n$ communication. The matrix reshape operations, such as transpose, take a matrix with $m * n$ elements as the input and output a matrix with $n * m$ elements. Both up-sampling and matrix reshape operations typically do not require any computation on the input activation, but they require communications to send data to the correct destination. Table 2 summarizes the layer analysis with the computation and communication ratio ($R_{c\&c}$) as defined in

$$R_{c\&c} = \frac{\text{Number of Computation}}{\text{Size of Input and Output Matrice}} \tag{4}$$

In Eq. (4), the amount of communication is estimated by adding the size of both input and output matrices. The layer requires more computation when the $R_{c\&c}$ value is higher. However, a layer requires more communication with a low $R_{c\&c}$ value. By comparing the $R_{c\&c}$ value for different layers, it can be seen that except for the fully connected/linear layer, in which computation is significantly higher than communication, the remaining layers require significantly more communication compared to the fully connected/linear layer.

C. Static GAN Model Analysis

Fig. 2 shows the breakdown of the communication requirements by statically analyzing the GAN models layer by layer in Table 1. The analysis assumes the inference of the GAN model is executed on sixteen chiplets, which contain multiplication and addition arrays for matrix multiplication and computation units for activation functions and up-sampling operations. The workload is evenly divided and distributed by the size of the activation for each layer. Based on the analysis, the reduction operations and up-sampling/reshape operations occupy more than 72% of the total communication. 36% of the communication is due to the reduction operation to calculate the statistics for the *SoftMax*/normalization layer. Thus, the challenges for the inference of large-scale GAN models on a scalable chiplet system are listed below.

1) **Scaling Reduction Operation**. The wide usage of the normalization layer results in significant communication. Most of these communications are due to the reduction operation for statistics calculation and element-wise operation. Moreover, as the size of activation increases for high-resolution image generation, supporting a large-scale reduction operation on chiplet systems becomes a challenge.
2) **Scaling Matrix Multiplication**. As the GAN models scale, the matrix multiplication workloads increase dramatically. To handle the increasing demand in computation for large matrix multiplication, a scalable chiplet system is needed.
3) **Distributed Up-Sampling/Matrix Reshape**. As the GAN models generate high-resolution data, these operations need to process a large activation matrix during inference, which involves significant communication traffic. Since existing accelerators use a scratch pad memory, both operations require a centralized global controller to reorganize

Table 2.
Summary of layer analysis.

Layer Name	$R_{c\&c}$
Fully Connected/Linear	$(n^2\text{-}n)mp/mn+np+mp$
Layer Normalization	3mn/2mn = 1.5
SoftMax	(3mn-1)/2mn≈1.5
Upscaling/Matrix Reshape	0

the data mapping in the memory, which is not scalable for a chiplet system. For a chiplet system, these operations must be distributed rather than centralized, which incurs a large amount of communication between chiplets for transmitting activations. This communication incurs a high latency, especially when the packet must travel a long distance.

D. Architecture Design

To solve the three challenges for scalable and efficient GAN inference on a chiplet system, the architecture of Chiplet-GAN includes the following features.

- **Adaptive NoC Topology**: To satisfy the communication requirement for the reduction operation, designing a concentrated network allows an efficient gathering of data for the operation. However, such a network has low efficiency for scaling matrix multiplication operations, as matrix multiplication needs frequent data exchange between neighbors. To support both reduction and matrix multiplication operations, Chiplet-GAN utilizes an adaptive network topology design, which dynamically switches between mesh and concentrated mesh (C-Mesh) topologies for efficient on-chip communication. This design ensures the efficiency of the execution of both reduction and matrix multiplication operations.
- **NoP With Both Active and Passive Network Links:** To efficiently handle both long-distance and short-distance communication between chiplets as the system scales, both active and passive network links are implemented in the NoP. Specifically, active network links are implemented for efficient long-distance communication between chiplets and passive network links for low latency communication between two chiplets.
- **Workload Partitioning and Allocation:** To reduce the communication latency for GAN inference, specialized workload partition and allocation algorithms are developed based on static model analysis. The proposed partition and allocation algorithm utilizes the characteristics of both the proposed NoP design and adaptive network topology in NoC for efficient and scalable GAN inference on Chiplet-GAN.

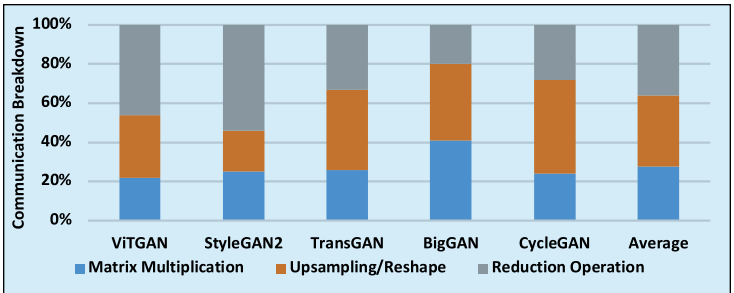


Figure 2. Communication breakdown through static GAN model analysis.

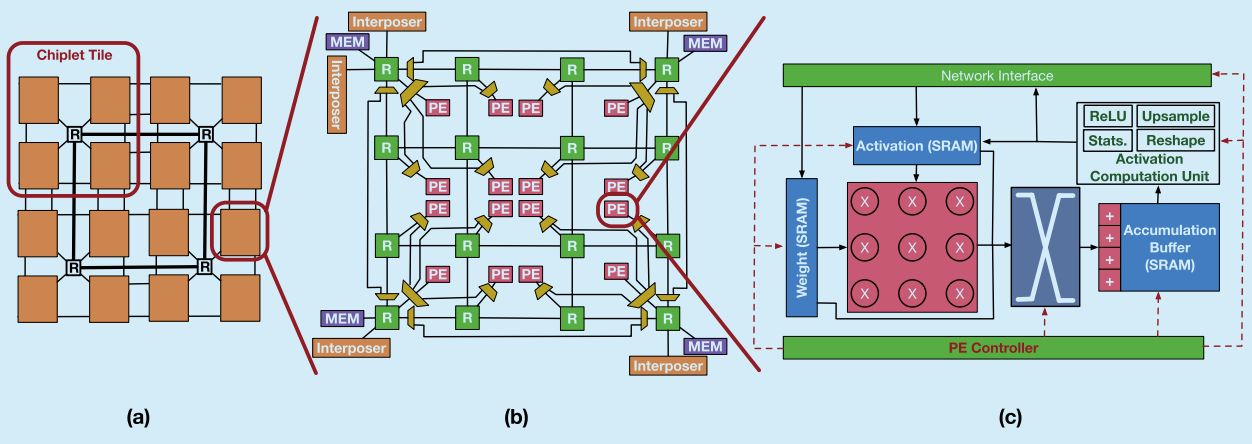


Figure 3. Architecture of Chiplet-GAN. The system package is scalable to include more chiplets tiles. Each chiplet tile consists of four chiplets and is connected to a router (R) on the interposer. The active link is implemented to connect routers for chiplet groups. The passive link is implemented to connect two chiplets. Within each chiplet, an adaptive network is implemented with routers (R) to connect the processing element (PE), the memory interface (MEM), and the active/passive interposer. The passive interposer connects the corner routers of the two adjacent chiplets. Each PE contains a PE controller, SRAMs for storing input and output matrices, a crossbar to forward the product of the multiplier array, and an activation computation unit for processing ReLU, activation statistics (Stats.) calculation, and up-sampling/matrix reshape operations. (a) System package. (b) Chiplet. (c) Processing element (PE).

Fig. 3 shows the architecture of Chiplet-GAN. The system package (Fig. 3(a)) contains multiple chiplets, which is scalable for the inference of large GAN models. As shown in Fig. 3(a), each chiplet group contains four chiplets, which are connected to a router via μbumps. This design handles long-distance communication between chiplets by connecting the corner router on each chiplet and the NoP router in a chiplet group using active network links. These active network links connect chiplets tiles, which not only enables the scaling of the chiplet system with additional tiles but also allows high bandwidth data exchange between tiles. For short-distance communication between adjacent chiplets, passive network links are implemented to enable quick data exchange between the corner routers of the two chiplets.

Fig. 3(b) shows the detailed design of the chiplet. The chiplet includes an adaptive network to efficiently process both reduction operation and matrix multiplication operation by switching between C-mesh and mesh topology. This is achieved by controlling the multiplexers and demultiplexers. By switching to the C-mesh topology, the latency of communication between chiplets is reduced by skipping routers and network links.

Fig. 3(c) illustrates the detailed design of each processing element (PE). Each PE contains an array of multipliers and an accumulative buffer for multiplication and addition operations. The activation computation unit is designed to execute several functions, including ReLU, activation statistics computation, and matrix up-sampling/reshape. Each PE contains a PE controller, which controls the computation process and activation

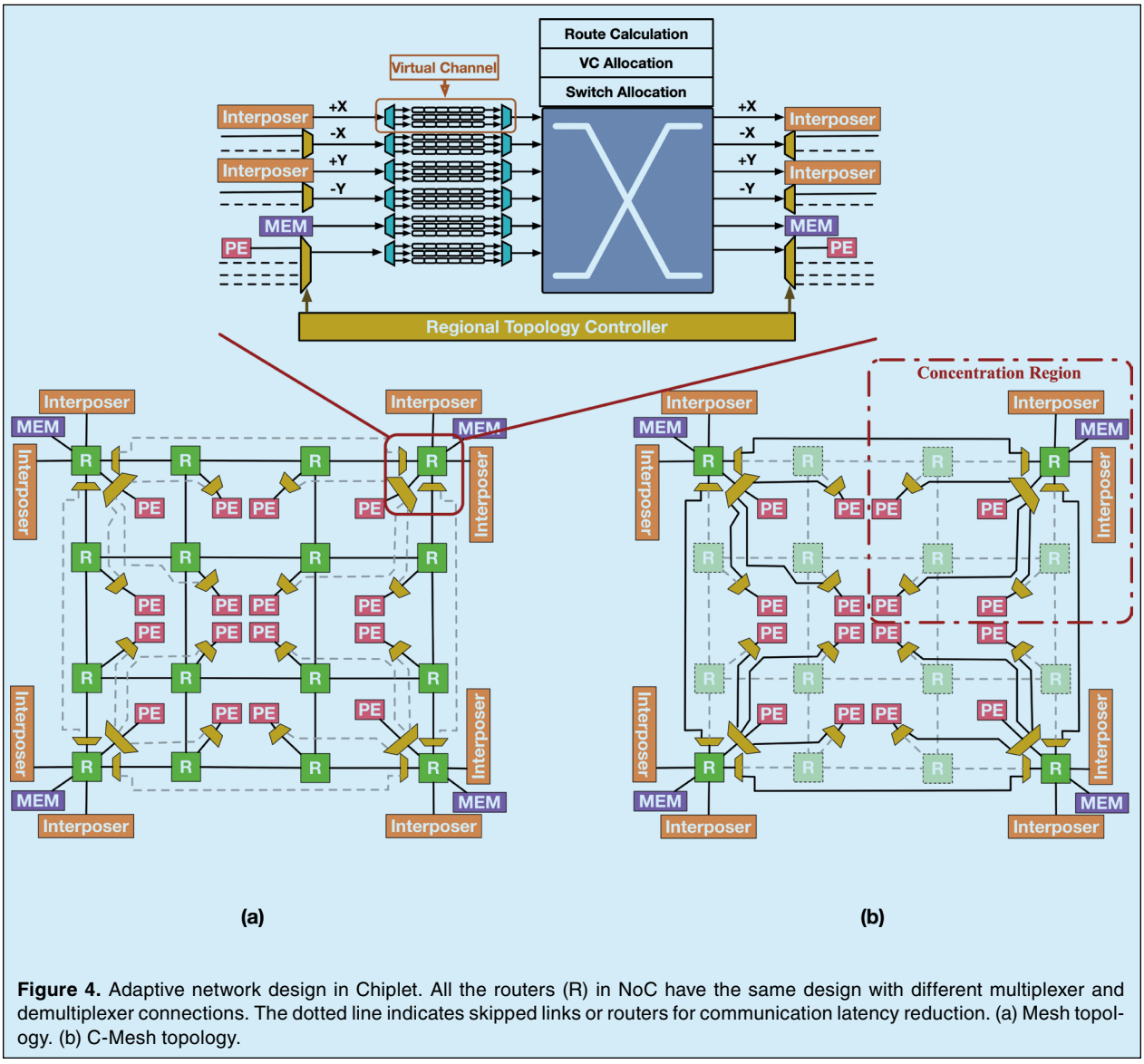


Figure 4. Adaptive network design in Chiplet. All the routers (R) in NoC have the same design with different multiplexer and demultiplexer connections. The dotted line indicates skipped links or routers for communication latency reduction. (a) Mesh topology. (b) C-Mesh topology.

computation unit. The proposed PE utilizes an output-stationary data flow to reduce the frequent exchange of partial sum during the matrix multiplication. With this data flow, the GAN inference workload is partitioned by the size of the output activation of a layer. Section III.E discusses the details of workload partitioning and allocation.

Fig. 4 shows the detailed design of the routers and the two configurations of the topology of NoC in each chiplet. Each chiplet consists of 16 PEs, which can be configured into a 4 * 4 mesh topology (Fig. 4(a)) or into a 2 * 2 C-Mesh topology (Fig. 4(b)). The PEs on each chiplet are partitioned into 4 concentration regions with 4 PEs for each region. The corner router of each chiplet is connected to one memory interface and the interposer to communicate with the neighbor chiplet. The router in Fig. 4 transmits a packet with a five-stage pipeline, which includes route calculation, virtual channel (VC) allocation, switch allocation, switch traversal, and link traversal. The multiplexers and demultiplexers are added between the router and network interface to allow each region to switch to a concentrated topology (i.e., C-Mesh) for efficient data gathering and scattering for PEs in the region. Each region is independently controlled by the regional topology controller, and the topology is changed dynamically during the GAN inference.

To prevent the loss of in-flight packets when switching between the two topologies, the following process is developed in the regional topology controller. After the chiplet completes the inference of one layer, the topology controller prepares to switch topology by monitoring the flits in VC. When the controller observes the last flit in VC is the tail flit, the multiplexers and demultiplexers for that VC are switched immediately. Otherwise, switching of topology is delayed until the tail flit enters the VC. With this switching process, there will be no information in the network link when the controller switches network topology, as the multiplexers and demultiplexers are implemented before the VC. Also, this process ensures the packet is intact when it is traversed through the router.

Compared to the NoP and NoC designs in Simba [18], the Chiplet-GAN facilitates not only efficient communication for reduction and matrix multiplication operations but also low latency communication between chiplets. This is mainly due to the implementation of an active interposer with both active and passive network links and the implementation of the adaptive network to skip excessive links and routers for communication by utilizing the C-Mesh topology. Compared to the passive-only interposer and network link for the NoP in Simba, the proposed NoP design utilizes chiplet tiles and active links for long-distance communication as systems scale and passive links for low latency communication between the adjacent chiplets. Moreover, within each chiplet, the C-Mesh topology also enables quick DRAM access through the NoC as well as easy adapting to mesh topology by only adding several multiplexers and demultiplexers to the conventional NoC routers. Section III.F discusses the configuration process of the adaptive network during the GAN inference.

E. Workload Partitioning and Allocation

Long-distance communication between chiplet impacts the execution time of GAN inference. Although the long-distance communication cost between chiplets can be reduced by utilizing active network links and routers in NoP, the GAN inference workload must be carefully partitioned to fully utilize this feature. Moreover, for layers that require a large amount of data movement (e.g., Matrix Up-Sampling/Reshape) within a short distance, the communication cost between chiplet can be reduced by utilizing the C-Mesh topology in each chiplet and the passive network link between chiplets. Thus, the proposed workload partitioning and workload allocation algorithms are designed to utilize these network features to reduce the communication cost for GAN inference.

1) Workload Partitioning

Since each PE contains an activation computation unit, the workload is first partitioned by the layer functions. Specifically, the GAN model is partitioned by groups, which contain multiple layers with matrix multiplication operations and one layer for activation computation units. Then, the workload for each group is divided and allocated to a set of chiplet tiles depending on the size of the largest activation and the size of the accumulative buffer in each chiplet tile due to the implementation of output-stationary dataflow in each PE.

$$\textit{Chiplet tile } \# = \left\lceil \frac{\text{max activation size in the group}}{\text{accumulative buffer size / chiplet tile}} \right\rceil \tag{5}$$

Eq. (5) illustrates the calculation of the number of chiplet tiles needed for the inference of each layer group, which is the ceiling of the maximum activation size divided by the size of the accumulative buffer in each chiplet tile.
With the proposed workload partitioning method, the inference of each group is allocated to a set number of chiplet tiles. As the last layer of each group is an activation layer, which involves reduction operation or matrix up sampling and reshape, the NoP router at the center of each chiplet tile is utilized for low latency communication, specifically, for a single chiplet tile, the router in

NoP along with the C-Mesh topology in the four chiplet forms a concentrated network for efficient exchange of the activation statistics during reduction operation. For the group that requires more than one chiplet tile, the active network link in NoP provides low-latency communication between the adjacent chiplet tiles. Moreover, after the inference of the activation layer, the NoP is utilized to transmit the activations to the next chiplet tiles for the inference of the next layer. To reduce the communication distance between the layer groups, an effective workload allocation algorithm is essential for the efficiency of Chiplet-GAN.

2) Workload Allocation

With the number of chiplet needed for each group, Algorithm 1 is developed to allocate the workloads on the chiplet system. The greedy algorithm is used in the allocation process to reduce the communication latency between chiplets. According to the workload partition, each layer group is assigned more than one chiplet tile. Thus, the allocation algorithm first calculates the communication cost for all the possible allocations for each layer group, then allocation with the minimal communication cost is selected.

As shown in Algorithm 1, suppose the GAN model M is partitioned into x groups (i.e., $M = [group\_1, group\_2, \ldots, group\_x]$) and the system C has n chiplets tiles (i.e., $C = [chiplet\_tile\_1, chiplet\_tile\_2, \ldots, chiplet\_tile\_n]$). The *Allocated_Group* stores the results of the workload allocation. In the *map_groups_to_chiplets* function, the function selects and allocates the inference workload to the chiplet tiles group by group. The set for the chiplets is selected by the *find_best_chiplet* function, where the minimal communication cost for all the possible allocations is found by evaluating the communication cost for all the possible workload allocations. The *calculate_communication_cost* function calculates the time needed for transmitting data through the network. According to the static workload analysis, the two types of data that occupy the majority of the traffic are input activations and temporary data, which include activation statistics and partial sums. These data are transmitted through either NoP or NoC during the inference. Thus, the communication cost (C) includes the network latency of input (C_i) and the network latency of temporary data (C_{tp} and C_{tc}). Considering the communication for temporary data can go through NoC and NoP, the C_{tp} and C_{tc} represent the communication latency for NoP and NoC, respectively.

Specifically, after the allocation of the group, the input activation is transmitted from another chiplet tile, which relies on the active links in NoP to transmit activation. The network latency of input activation (C_i) is calculated based on the physical delay of the NoP through the active link (T_{rp_late}), the number of NoP routers traversed (H_i), the size of one flit in a packet (S_f), the size of input activations (S_i), and transmission time for one flit in NoP (T_{fp}). In this function, T_{rp_late} is the physical delay caused by the network's physical aspects, which include switch and link delays. Notability, the latency for NoC communication is constant regardless of the allocation of the workload in the function for the network latency calculation of input activation (C_i), as all the activation data have to traverse the NoC in that chiplet to reach the PEs for inference. Thus, it is assumed that the communication latency in NoC is

Algorithm 1
Workload Allocation

```
def calculate_communication_cost(placed_group):
    H_i = # of NoP routers traversed for input activations of a group
    S_i = the size of input activations
    H_tp = # of NoP routers traversed for temporary data
    S_tp = the size of temporary data traversed through the active NoP link
    N_pl = # of traversed passive NoP link for temporary data
    H_tc = # of NoC routers traversed for temporary data
    S_tc = the size of temporary data traversed through NoC
    C_i = T_rp_late*H_i + (S_i/S_f)*T_fp
    C_tp = T_rp_late*H_tp + (S_tp/S_f)*T_fp
    C_tc = T_rc_late*H_tc + (S_tc/S_f)*(T_fc + L_pl*N_pl)
    C = C_i +C_tp +C_tc
return C
def find_best_chiplet(group, available_chiplets):
    min_cost = infinity
    best_chiplet = None
    for chiplet in available_chiplets:
        Place group to chiplet
        for placed_group in available_allocation:
            cost = calculate_communication_
              cost(placed_group)
            if cost < min_cost:
              min_cost = cost
              best_chiplet = chiplet
return best_chiplet
def map_groups_to_chiplets(groups, chiplet_tiles):
    mapped_groups = {}
    available_chiplets = chiplet_tiles
    for group in groups:
    best_chiplet = find_best_chiplet(group, available_chiplets)
    mapped_groups[group] = best_chiplet
    available_chiplets.remove(best_chiplet)
return mapped_groups
GAN Model M = [group_1,group_2, … , group_x]
Chiplet System C = [chiplet_tile_1, chiplet_tile_2, …, chiplet_tile_n]
Allocated_Group = map_group_to_chiplets(M,C)
```

the same, and only the communication cost for NoP is calculated.

Considering the difference in design for the links in NoC and NoP, the T_{rp_late} and T_{rc_late} represent the physical delay for active links in NoP and NoC, respectively, for the calculation of C_{tp} and C_{tc}. The calculation of C_{tp} includes the network latency for the data traversed through the active link in the NoP, where T_{rp_late} is used. In terms of the calculation of C_{tc}, the network latency includes two parts, namely, the latency of traversing NoC and passive NoP links. When a layer group is allocated on more than one chiplet, the communication in the NoC is short-distance communication; thus, passive NoP links are utilized to transmit temporary data. As the change in workload allocation affects the number of passive NoP links traversed during the transmission of temporary data, the network latency of traversing both NoC and passive NoP links has to be counted. In the calculation of the network latency for NoC, T_{fc} represents the transmission time for one flit in NoC. In the calculation of the network latency for traversing passive NoP links, the L_{pl} represents the transmission time for one flit in passive NoP links. Notability, for the NoC communication cost, the algorithm calculates the communication latency under the C-mesh topology configuration, which is lower compared to mesh topology. The following section illustrates the configuration of the adaptive network in chiplet, which maintains low latency communications during GAN inference.

F. Adaptive Network Configuration

The network topology in each chiplet is dynamically changed depending on the function and the allocation of the layers to the accelerator. The regional topology controller for each concentration region changes the network topology to support the partitioned and placed workload for efficient communication between PEs and between chiplet. The controller selects C-Mesh topology for DRAM access, chiplet-to-chiplet communication, and reduction communication. Thus, the communication latency for chiplet-to-chiplet communication is calculated for the C-Mesh topology in each chiplet. For PE-to-PE communication, the mesh topology is selected. The controller monitors the computation process of four PEs in the concentration region. When the PEs in the region finish one layer, the controller selects the next topology based on the communication requirement for the next layer. The configuration process for matrix multiplication, reduction operation, and up-sampling/matrix reshape operations are listed as follows.

1) **Matrix Multiplication.** The NoC is configured into the C-Mesh topology when loading the SRAM in each PE. Then, the network is configured into mesh for efficient exchange of data during the matrix multiplication. For the matrix multiplication workloads, which are allocated to more than two chiplets, the network is configured into a C-Mesh topology for efficient chiplet-to-chiplet communication or DRAM access after the calculation of the local partial sum.
2) **Reduction Operation.** The NoC is configured into C-Mesh topology for reduction operation. For reduction operation on a single chiplet, the parameter is calculated at the four corner PEs. For reduction operation across multiple chiplets, the corner PEs in each chiplet with the most connection to the neighbor in the previous layer are selected for calculating the parameter. Then, the parameters are sent back to PEs for the next operation through the concentrated network.
3) **Up-Sampling/Matrix Reshape.** For the matrix reshape operation, which needs to either store the activation to DRAM or communication between chiplets, the NoC is configured into the C-Mesh topology for efficient communication.

IV. Evaluation

A. Simulation Setup

In this section, the performance of Chiplet-GAN is evaluated by using the SMAUG [26] simulator. The cycle-accurate SMAUG simulation model is modified to support

Table 3.
Simulation setup.

PE Parameter	Value
Multiplier	32 bits FLOAT
Accumulator	32 bits FLOAT
Number of Multiplier	16
Number of Adder	16
Accumulation Buffer	144*32 bit
Activation/Weight SRAM Size	244*32 bit
Chiplet Parameter	Value
Number of PE	4*4
Link width	512 bit
Packet Size	4 flit * 512 bit
Payload/Packet	3 flit * 512 bit
Accelerator Parameters	Value
Number of Chiplets	4*4
DRAM Bandwidth	1024 GB/s
DRAM Size	32 GB DDR4
Frequency	2 GHz

Chiplet-GAN. Table 3 shows the settings for the SMAUG simulator for the Chiplet-GAN. PyTorch is used with the SMAUG simulator to control the entire GAN inference process. The accelerators are synthesized using Synopsys Design Compiler with TSMC 16nm to obtain power dissipation during GAN inference and area consumption of the system.

Chiplet-GAN is compared against GANAX [12], SpAtten [13], and Simba [18] on execution time, communication latency, energy consumption, area, and scalability. GANAX is a GAN inference accelerator for TC acceleration with a single-chip design. SpAtten is an accelerator designed to accelerate the inference of transformers, which also incorporates a single-chip design. Simba is a chiplet-based accelerator that focuses on the DCNN model inference. All accelerators are implemented with the same number of computation units as well as the same size of SRAM, DRAM, and DRAM bandwidth. The inference of the models listed in Table 1 is executed on these accelerators to generate two images during evaluation.

B. Execution Time

Fig. 5 shows the execution time reduction normalized to Chiplet-GAN. The proposed design reduces execution time by 27%, 29%, and 45% on average, compared to GANAX, SpAtten, and Simba, respectively. Compared to Chiplet-GAN, SpAtten enhanced the efficiency of inference attention mechanism, but it takes a longer time to process up-sampling operations. Thus, Chiplet-GAN achieves a 16% execution time reduction on average for transformer-based GANs (i.e., ViTGAN and TransGAN). For TC-based models (i.e., StyleGAN2, BigGAN, and CycleGAN), which require more frequent up-sampling operations, Chiplet-GAN reduces 38% in execution time on average compared to the SpAtten. Compared to GANAX, which is optimized for up-sampling and matrix reshaping, the Chiplet-GAN achieves a 21% execution time reduction on average.

The main reason for execution time reduction is the implementation of the adaptive interconnection network design, active/passive network link in NoP, and workload allocation strategy. These design features not only reduce the communication latency but also increase the utilization of computation units with fewer idle cycles. The reduction in communication latency is shown in Fig. 6, which is measured by the time elapsed between the access of the data and the beginning of the computation. Chiplet-GAN reduces communication latency by 46%, 52%, and 68% on average, compared to GANAX, SpAtten, and Simba, respectively.

Since both Simba and GANAX mainly focus on accelerating convolutional operations, the up-sampling/reshape and reduction operations are not fully accelerated when executing GAN inference. Specifically, TC acceleration, which contains zero insertion optimization, is incorporated in GANAX design. However, a notable drawback in GANAX is the lack of support for reduction operations. This limitation leads to the need for temporary storage of activations in both the DRAM and global buffer, resulting in a considerable increase in data communication

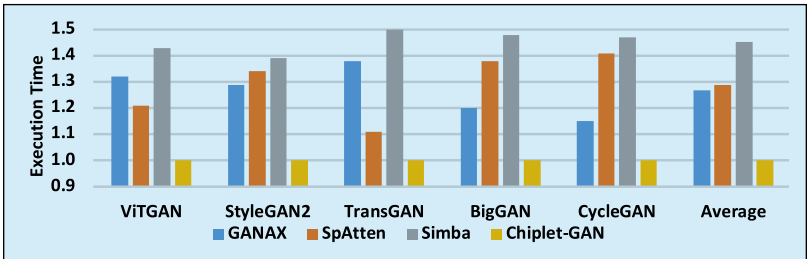


Figure 5. Execution time. Results are normalized to Chiplet-GAN.

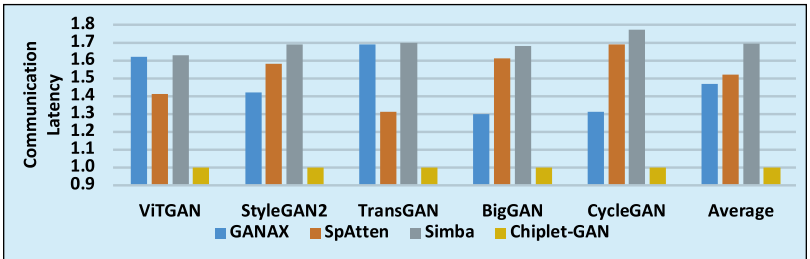


Figure 6. Communication latency. Results are normalized to Chiplet-GAN.

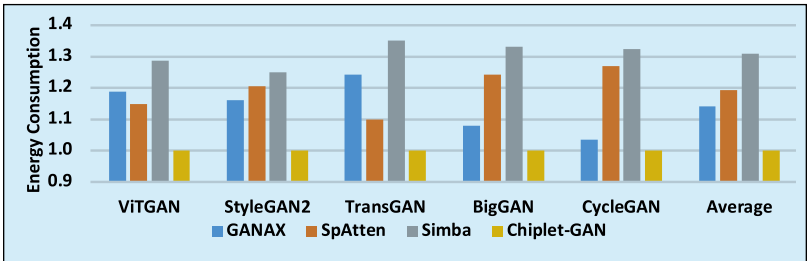


Figure 7. Energy consumption. Results are normalized to Chiplet-GAN.

Chiplet-GAN features an adaptive topology design for the interconnection network and incorporates workload partition and allocation algorithms to efficiently handle large-scale GAN models.

and communication latency for GANAX, especially for the models that require frequent reduction operations (e.g., ViTGAN and TransGAN). As shown in Fig. 6, the GANAX and Simba achieve similar communication latency compared to Chiplet-GAN for ViTGAN and TransGAN.

On the other hand, Simba only focuses on the acceleration of DCNN inference. Thus, the lack of communication latency reduction techniques for both up-sampling/reshape and reduction operations incurs more communication delays and longer execution time. As shown in Fig. 6, Simba incurs 23% and 17% more communication latency compared to GANAX and SpAtten, respectively. Apart from lacking support for the operations during GAN inference, the long-distance communication between chiplets is another issue for Simba. This is mainly caused by the close placement of routers and passive links only NoP design, which further increases the communication latency. As a result, compared to Chiplet-GAN, the Simba incurs significantly longer execution times during GAN inference.

C. Energy Consumption

Fig. 7 shows the evaluation results for the energy consumption of Chiplet-GAN, GANAX, SpAtten, and Simba. All results are normalized to Chiplet-GAN. The energy is the product of execution time and power dissipation. Power dissipation includes two parts: static power and dynamic power.

Chiplet-GAN reduces energy consumption by 20% on average compared to the existing accelerators. Specifically, compared to GANAX, SpAtten, and Simba, Chiplet-GAN reduces energy consumption by 14%, 19%, and 30% on average, respectively. Compared to the reduction in execution time, the energy reduction is less; this is mostly due to the extra power consumed by the adaptive interconnection network in Chiplet-GAN and the active links implemented in the interposer. The GANAX and SpAtten have simplified networks with the predetermined data flow; however, fixed data flow in existing designs results in an unscalable architecture, which significantly impacts execution time as the system scales.

D. Area Evaluation

Table 4 summarizes the area required for the accelerators with the configuration given in Table 3. The results are from the synthesis report of the Synopsys Design Compiler.

Overall, Chiplet-GAN occupies less area compared to SpAtten despite the additional area needed for the connection between chiplets and interposer. This is mainly due to the implementation of a crossbar for on-chip communication, which incurs a large on-chip area in SpAtten. Compared to SpAtten, the proposed

Table 4. Area.

	GANAX	SpAtten	Simba	Chiplet-GAN
Area (mm2)	5.98	7.48	6.03	6.34

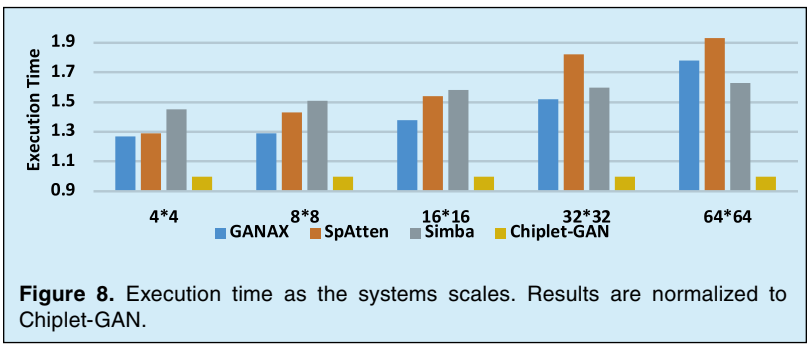


Figure 8. Execution time as the systems scales. Results are normalized to Chiplet-GAN.

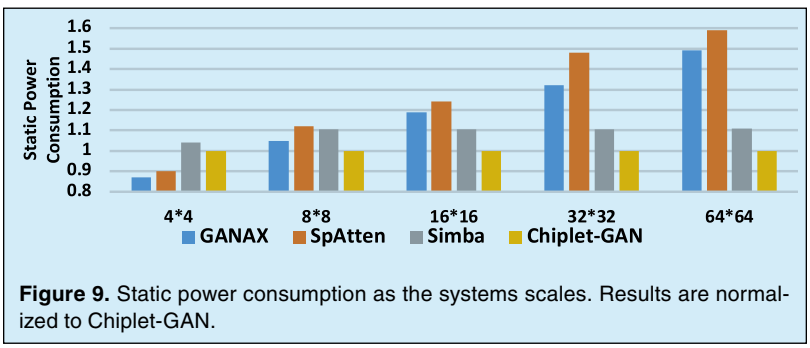


Figure 9. Static power consumption as the systems scales. Results are normalized to Chiplet-GAN.

accelerator reduces the area by 15%. Compared to Simba and GANAX, Chiplet-GAN requires 5% and 6% more area, respectively. This is mainly due to the additional hardware needed to support adaptive interconnection networks and active/passive network links. However, when considering the significant reduction in execution time and energy consumption, the increase in area is rather marginal for implementing the proposed accelerator.

E. Scalability

To demonstrate the scalability of the system, the GAN models are scaled by increasing the size of both activation and weights as Chiplet-GAN scales to 8 * 8, 16 * 16, 32 * 32, and 64 * 64 chiplets. For example, a layer in a model with a weights size of 4 * 4 * 3 and activations size of 128 * 128 * 3 is executed on the system with 4 * 4 chiplets. The weights and activations are scaled to sizes of 8 * 8 * 3 and 256 * 256 * 3, respectively, as the system scales from 4 * 4 to 8 * 8. The DRAM size and DRAM bandwidth are scaled accordingly by doubling the size and bandwidth to handle the significant increase in memory port and memory traffic. The SpAtten, GANAX, and Simba are also scaled accordingly with the same amount of computation units and on-chip SRAM.

Figs. 8 and 9 show the comparison of average execution time for all the GAN inference workloads and static power consumption under these configurations. Compared to a chiplet-based design (Simba), the Chiplet-GAN achieves up to a 63% reduction in execution time as the system scales to 64 * 64 chiplets. Compared to GANAX and SpAtten, the Chiplet-GAN reduces execution time by up to 93% as the system scales. Simba achieves better scalability compared to single-chip designs (i.e., GANAX and SpAtten) due to the chiplet design. However, compared to Chiplet-GAN, Simba still requires more execution time. This is mainly due to the high latency communication between chiplets during inference. In terms of static power consumption, both chiplet systems show an advantage in scalability as the number of components and power consumption linearly increases. Due to the simplicity of GANAX and SpAtten, they consume less power when the system is small (i.e., 4 * 4). However, the power consumption for GANAX and GANPU increases dramatically as the complexity of the global controller must be increased significantly to handle the increase in the computation units. Specifically, compared to Chiplet-GAN, the chiplet-based design (i.e., Simba) consumes up to 10% more power, whereas GANAX and SpAtten consume up to 49% and 59% more static power, respectively, as the system scales.

V. Conclusion

In this article, we analyzed the communication and computation requirements for large-scale GAN model inference and identified three major challenges. Addressing these challenges, we proposed Chiplet-GAN, a chiplet-based accelerator for scaling reduction operation, scaling matrix multiplication, and distributed up-sampling/matrix reshape operations. To the best of our knowledge, this is the first work that proposes a chiplet-based design approach for GANs. The proposed design introduces a novel interconnection fabric with adaptive topology, active/passive network links in NoP, and a workload partition and allocation algorithm. The novel interconnection fabric enables low communication latency during GAN inference. The workload partition and allocation algorithms further reduce communication latency with a greedy algorithm. We conducted extensive simulation studies to demonstrate the effectiveness of Chiplet-GAN. Our detailed evaluation shows that Chiplet-GAN reduces the execution time by 34% and the energy consumption by 21% on average compared to GANAX, SpAtten, and Simba. As the accelerator scales to enhance computation capability for large-scale GAN inference, the proposed design reduces the execution time by up to 63% compared to the existing chiplet-based accelerator (Simba).

Acknowledgment

This work was supported by NSF under Grant CCF-1953961, Grant CCF-1812467, Grant CCF-1812495, Grant CCF-1953980, Grant CCF-1702980, and Grant CCF-1901165.

Yuechen Chen (Member, IEEE), received the Ph.D. degree in computer engineering from George Washington University, Washington, DC, in 2024. He is currently an Assistant Professor with the Department of Computer Science and Information Technologies, Frostburg State University, Maryland, USA. His research interests include parallel computing systems, network-on-chip (NoC), machine learning algorithms, applications, and accelerators.

Ahmed Louri (Fellow, IEEE) received the Ph.D. degree in computer engineering from the University of Southern California, Los Angeles, CA, in 1988. From 1988 to 2015, he was a Professor of electrical and computer engineering at the University of Arizona. From 2010 to 2013, he

served as a Program Director of the National Science Foundation's (NSF) Directorate for Computer and Information Science and Engineering. He is currently the David and the Marilyn Karlgaard Endowed Chair Professor of electrical and computer engineering with George Washington University, Washington, DC, USA, which he joined in August 2015. His research interests include interconnection networks and network on chips for multicores, and the use of machine learning techniques for energy-efficient, reliable, high-performance and secure many-core architectures and accelerators. He was the Editor-in-Chief of IEEE TRANSACTIONS ON COMPUTERS from 2019 to 2023. In 2024, he is serving on the Computer Society Publication Board Executive Committee. He is also the Chair of the Transactions Operations Committee.

Fabrizio Lombardi (Fellow, IEEE), received the B.Sc. degree (Hons.) in electronic engineering from the University of Essex, U.K., in 1977, the master's degree in microwaves and modern optics and the Diploma degree in microwave engineering from the Microwave Research Unit, University College London, in 1978, and the Ph.D. degree from the University of London in 1982. He is currently the International Test Conference (ITC) Endowed Chair Professorship with Northeastern University, Boston, USA. His research interests include bio-inspired and nano manufacturing/computing, VLSI design, testing, and fault/defect tolerance of digital systems. He was the President of the IEEE Nanotechnology Council in 2022–2023.

Shanshan Liu (Senior Member, IEEE) received the Ph.D. degree in microelectronics and solid-state electronics from the Harbin Institute of Technology, Harbin, China, in 2018. She was a Post-Doctoral Researcher at Northeastern University, Boston, USA, from 2018 to 2021, and an Assistant Professor with New Mexico State University, Las Cruces, from 2021 to 2023. She is currently a Professor with the University of Electronic Science and Technology of China, Chengdu, China. Her research interests include fault tolerance design in high performance computer systems, VLSI design, dependable machine learning, stochastic computing, and error correction codes.

References

[1] T. Karras et al., "Training generative adversarial networks with limited data," in *Advances in Neural Information Processing Systems*. Red Hook, NY, USA: Curran Associates, 2020, pp. 12104–12114.

[2] M. Chen et al., "Generative pretraining from pixels," in *Proc. 37th Int. Conf. Mach. Learn.*, Nov. 2020, pp. 1691–1703.

[3] P. Esser, R. Rombach, and B. Ommer, "Taming transformers for high-resolution image synthesis," presented at the *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 12868–12878.

[4] H. Abdelaziz et al., "Rethinking floating point overheads for mixed precision DNN accelerators," *Proc. Mach. Learn. Syst.*, vol. 3, pp. 223–239, Mar. 2021.

[5] E. Richardson et al., "Encoding in style: A styleGAN encoder for image-to-image translation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 2287–2296.

[6] Y. Jiang, S. Chang, and Z. Wang, "TransGAN: Two pure transformers can make one strong GAN, and that can scale up," in *Advances in Neural Information Processing Systems*. Red Hook, NY, USA: Curran Associates, 2021, pp. 14745–14758.

[7] J.-Y. Zhu et al., "Unpaired image-to-image translation using cycle-consistent adversarial networks," 2020, *arXiv:1703.10593*.

[8] J. Park and Y. Kim, "Styleformer: Transformer based generative adversarial networks with style vector," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 8973–8982.

[9] A. Brock, J. Donahue, and K. Simonyan, "Large scale GAN training for high fidelity natural image synthesis," 2018, *arXiv:1809.11096*.

[10] H. Zhang et al., "Self-attention generative adversarial networks," in *Proc. 36th Int. Conf. Mach. Learn.*, May 2019, pp. 7354–7363.

[11] K. Lee et al., "ViTGAN: Training GANs with vision transformers," 2021, *arXiv:2107.04589*.

[12] A. Yazdanbakhsh et al., "GANAX: A unified MIMD-SIMD acceleration for generative adversarial networks," in *Proc. ACM/IEEE 45th Annu. Int. Symp. Comput. Archit. (ISCA)*, Jun. 2018, pp. 650–661.

[13] H. Wang, Z. Zhang, and S. Han, "SpAtten: Efficient sparse attention architecture with cascade token and head pruning," in *Proc. IEEE Int. Symp. High-Perform. Comput. Archit. (HPCA)*, Feb. 2021, pp. 97–110.

[14] S. Kang et al., "GANPU: An energy-efficient multi-DNN training processor for GANs with speculative dual-sparsity exploitation," *IEEE J. Solid-State Circuits*, vol. 56, no. 9, pp. 2845–2857, Sep. 2021.

[15] T. J. Ham et al., "A^3: Accelerating attention mechanisms in neural networks with approximation," in *Proc. IEEE Int. Symp. High Perform. Comput. Archit. (HPCA)*, Feb. 2020, pp. 328–341.

[16] W. Luo et al., "Rethinking motivation of deep neural architectures," *IEEE Circuits Syst. Mag.*, vol. 20, no. 4, pp. 65–76, 4th Quart., 2020, doi: 10.1109/MCAS.2020.3027222.

[17] A. Vaswani et al., "Attention is all you need," 2017, *arXiv:1706.03762*.

[18] Y. S. Shao et al., "Simba: Scaling deep-learning inference with multi-chip-module-based architecture," in *Proc. 52nd Annu. IEEE/ACM Int. Symp. Microarchitecture (MICRO)*. New York, NY, USA: Assoc. Comput. Machinery, Oct. 2019, pp. 14–27.

[19] R. Hwang et al., "Centaur: A chiplet-based, hybrid sparse-dense accelerator for personalized recommendations," in *Proc. ACM/IEEE 47th Annu. Int. Symp. Comput. Archit. (ISCA)*, May 2020, pp. 968–981.

[20] M.-S. Lin et al., "A 7-nm 4-GHz arm^{1}-core-based CoWoS^{1} chiplet design for high-performance computing," *IEEE J. Solid-State Circuits*, vol. 55, no. 4, pp. 956–966, Apr. 2020.

[21] J.-F. Zhang and Z. Zhang, "Machine learning hardware design for efficiency, flexibility, and scalability," *IEEE Circuits Syst. Mag.*, vol. 23, no. 3, pp. 35–53, 2023.

[22] N. E. Jerger et al., "NoC architectures for silicon interposer systems: Why pay for more wires when you can get them (from your interposer) for free?" in *Proc. 47th Annu. IEEE/ACM Int. Symp. Microarchitecture*, Dec. 2014, pp. 458–470.

[23] A. Kannan, N. E. Jerger, and G. H. Loh, "Enabling interposer-based disintegration of multi-core processors," in *Proc. 48th Annu. IEEE/ACM Int. Symp. Microarchitecture (MICRO)*, Dec. 2015, pp. 546–558.

[24] D. Stow et al., "Cost-effective design of scalable high-performance systems using active and passive interposers," in *Proc. IEEE/ACM Int. Conf. Comput.-Aided Design (ICCAD)*, Nov. 2017, pp. 728–735.

[25] G. Shan et al., "Architecture of computing system based on chiplet," *Micromachines*, vol. 13, no. 2, p. 205, Jan. 2022.

[26] S. L. Xi et al., "SMAUG: End-to-end full-stack simulation infrastructure for deep learning workloads," *ACM Trans. Archit. Code Optim.*, vol. 17, no. 4, pp. 39:1–39:26, Nov. 2020.

Innovations Corner

Michael Friebe, Jie Chen, and Fakhrul Zaman Rokhani

Sparking Technological Innovation Through CASS Educational Entrepreneurship Initiative

Abstract

Recognizing the increasing relevance of entrepreneurship skill-sets to engineers, CASS launched an initiative to educate on entrepreneurship with technology solutions aligned with CAS Society's visions. In this special issue, this article introduces the motivation, course setup, and the results of an intensive six-week hybrid course stimulating entrepreneurial thinking. The course was based on a novel iterative and agile innovation framework designed to lead up to the BIOCASS 2023 conference in Toronto, inviting the best proposals for an in-person presentation. The learners confirmed a changed mindset towards innovation generation.

I. Introduction

Health innovations often originate from technical and scientific departments but are mainly connected to the clinical needs raised by the current clinical stakeholders. This is not bad but typically leads to incremental innovations. Moreover, it will not solve the major challenges we face concerning healthcare delivery, e.g., unequal access, demographic changes/increased longevity and healthcare costs, and the lack of focus on prevention and home care.

A more forward-looking view considering the technological advancements could address these challenges even though the business models supporting these developments do not yet exist [1], [2].

There is hope that entrepreneurial activities focusing on the problem space can do so. We believe that general entrepreneurial offerings must be complemented (or replaced) by dedicated programs focusing on health-related challenges (e.g., SDG 3: Health + Wellbeing for Everyone).

Michael Friebe is with the FOM University of Applied Sciences, 45127 Essen, Germany, and also with the AGH University of Krakow, 30-059 Krakow, Poland (e-mail: friebe@agh.edu.pl).
Jie Chen is with the University of Alberta, Edmonton, AB T6G 2R3, Canada, and also with the Academy for Engineering and Technology, Fudan University, Shanghai 200433, China.
Fakhrul Zaman Rokhani is with Universiti Putra Malaysia, Serdang 43400, Malaysia.

Digital Object Identifier 10.1109/MCAS.2024.3358279
Date of current version: 15 August 2024

II. Course Setup and Teaching Methods

A 6-week course consisting of 30 videos (each between 7 and 15 minutes) offered on the CASS Microlearning platform [3], with an equal number of assignments and complementing literature resources, was developed. The regular online mentoring session provides interaction and opportunities for learners seeking further clarity from instructors and other learners. At the end of each week, learners' understanding is tested through a summative assessment consisting of 25 multiple-choice questions.

Every week had a different main theme: 1) *EXPLORATION + FUTURE PROJECTION*: Problem statement. 2) *DESIRABILITY*: Value propositions and key stakeholder needs. 3) *VIABILITY*: Business Models embedded in a Massive Transformative Purpose. 4) *FEASIBILITY*: Minimal Viable Prototypes. 5) *VALIDATION + ITERATION,* and finally, 6) Creating a *PRESENTATION*.

The innovation and entrepreneurial mindset framework and methodology were based primarily on the Purpose Launchpad [4] and the adapted Health version (see Fig. 1) [2], [5]. The attendees were introduced to Financial planning, Value Proposition + Business Model Canvas, Team + Empathy Canvas, SWOT analysis, Elevator Pitch, the "How Might We So That" concept, Exponential and MoonshotVisionMissionValues Canvas, One-Page Summary + creation of a Pitch Deck [2].

III. Results and Conclusion

The course started with over 30 learners and quickly lost more than half. This was expected based on the weekly workload and different expectations towards the course. The remaining learners turned in excellent results with very positive feedback and convincing improvements throughout the course (see Fig. 2). Seven finalists participated in the competition, and six of them are featured in this special issue. Detailed reports of the journey of these finalists in discovering CAS-based technology innovations for various healthcare needs are explained within this special issue. We hope that many more teaching institutions will use such a course in the future.

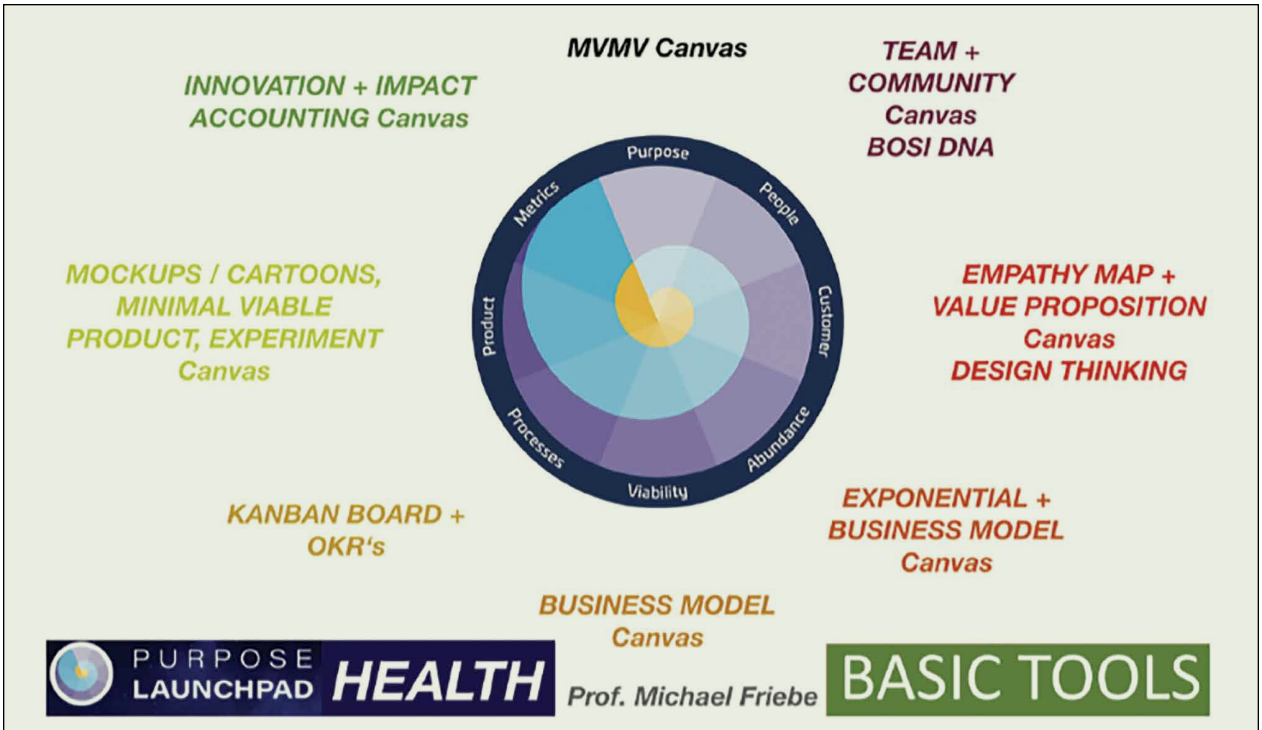


Figure 1. PURPOSE LAUNCHPAD HEALTH framework and the recommended tools for each of the eight segments starting with PURPOSE.

Figure 2. The projects—from a deep-dive Problem Exploration to a conceptual idea for Validation with a (middle) typical Purpose Launchpad Progress.

Seven finalists participated in the competition, and six of them are featured in this special issue. Detailed reports of the journey of these finalists in discovering CAS-based technology innovations for various healthcare needs are explained within this special issue.

References

[1] M. Friebe, "Healthcare in need of innovation: Exponential technology and biomedical entrepreneurship as solution providers," *SPIE Medi. Imag.*, vol. 11315, p. 113150T, Mar. 2020, doi: 10.1117/12.2556776.

[2] M. Friebe, Ed. *Novel Innovation Design for the Future of Health*. Cham, Switzerland: Springer Nature, doi: 10.1007/978-3-031-08191-0 .

[3] *IEEE CASS Microlearning Platform*. Accessed: Oct. 1, 2023 [Online]. Available: https://ml.ieee-cas.org

[4] F. Palao. *Purpose Launchpad Guide and Assessment Information*. Accessed: Oct. 1, 2023. [Online]. Available: www.purposelaunchpad.com

[5] M. H. Friebe et al., "The PLH—Purpose launchpad health—Meta-methodology to explore problems and evaluate solutions for biomedical engineering impact creation," in *Proc. EMBC*, 2022, pp. 3299–3302, doi: 10.1109/EMBC48229.2022.9871218.

Michael Friebe received the Ph.D. degree in medical physics in Germany. He spent five years in San Francisco, as a Research and Design Engineer with an MRI and ultrasound device manufacturer. Since 2022, he has been a Professor of biomedical engineering innovation with the AGH University of Science and Technology, Krakow, Poland. He is currently the Director of the Center of Innovation, Business Development and Entrrepreneurship, FOM University, Essen, Germany; a Research Fellow with the Technical University of Munich, Munich; an Adjunct Professor with the Queensland University of Technology, Brisbane; and an Honorary Professor of image guided therapies with Otto von Guericke University, Magdeburg, Germany. He is a listed inventor of more than 80 patents and has authored over 300 articles, three books, and was part of over 35 medical technology start-ups. From 2016 to 2018, he was a Distinguished Lecturer of the IEEE EMBC Teaching Innovation Generation and the MedTec Entrepreneurship.

Jie Chen received the Ph.D. degree in electrical and computer engineering from the University of Maryland, College Park, MD, USA. He is currently an Inaugural Dean of biomedical engineering, Fudan University, China. He was a Professor with the Department of Electrical and Computer Engineering and an Adjunct Professor with the Biomedical Engineering Department, University of Alberta, Edmonton, AB, Canada. He has coauthored three books and more than 230 journal articles and conference proceeding papers (Citations over 8400, H-index 44). He is the Board of Governors for the IEEE Circuits and Systems Society. He was the Committee Chair of the Technical Program of Bio Nanotechnology and BioMEMS (BNM) and IEEE EMBS.

Fakhrul Zaman Rokhani received the Ph.D. degree from the University of Minnesota, USA. He was with the Intel Penang Design Center and Huawei Technologies as a Visiting Professor, a Visiting Scholar with the ASIC and Systems State Key Laboratory, Fudan University, and a Visiting Professor at several other universities. He is currently an Associate Professor with the Universiti Putra Malaysia and the Head of the Smart System and System-on-Chip (S3oC) Research Group. His current research interests include low-power/energy-efficient system-on-chip (SoC) design and automation, the IoT system integration, and sensors for food quality applications. In 2023, he co-founded a startup company focusing on low power IPs. He serves as the Vice President of Education and Communications for CAS Society and the Steering Committee of the IEEE Future Direction—Global Semiconductors at IEEE. On the publication front, he has been contributing as an Associate Editor of IEEE Transactions on Circuits and Systems—I: Regular Papers, an Associate Editor of the CAS Society newsletter, the Guest Editor of IEEE Transactions on Circuits and Systems—I: Regular Papers and IEEE Transactions on Circuits and Systems—II: Express Briefs, and the technical program/publication chair/track chairs/embedded workshop chair of several IEEE CASS conferences.

Yuqi Wu, Wenshan Huang, Xuanjie Ye, and Jie Chen, *Fellow, IEEE*

AīCare: An Affordable, Reliable, and Intelligent Senior Care Ecosystem

Abstract

Global population aging has become an issue nowadays, which will eventually lead to a huge senior care market in the future. In this article, a senior care ecosystem, named AīCare, was proposed to revolutionize the future senior care industry. The AīCare senior care ecosystem combined exponential technologies, such as IoT, cloud computing, and AI, to deliver affordable, reliable, and intelligent senior care services and products, which aims to amplify the efficacy of current systems and contributes significantly towards the evolution of senior care, ultimately augmenting the quality-of-care provision.

I. Introduction

The global population is aging rapidly due to factors like increased longevity and lower fertility rates. This trend affects families, social welfare, support services, and health systems, with significant implications for older individuals' well-being. Addressing these challenges is crucial for the welfare of the growing elderly population. United Nations statistics predict around 1.6 billion people aged 65 and above globally, emphasizing the importance of senior care in both research and application [1].

II. Promblem Statement

Current senior care systems face challenges, including a shortage of trained providers, especially in low-income areas, and rising costs due to high labor expenses and limited government funding. These systems often prioritize physical health, neglecting the mental and emotional needs of seniors. The elderly people are at risk of loneliness and related mental health issues due to factors like declining health and social isolation. To address these issues, our venture plans to introduce a new senior care model using technologies like cloud computing, AI, and IoT to offer affordable, reliable, and holistic care for the elderly, ensuring their overall well-being.

Yuqi Wu and Xuanjie Ye are with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6G 2R3, Canada. Jie Chen is with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6G 2R3, Canada, and also with the Academy for Engineering & Technology, Fudan University, Shanghai 200433, China (e-mail: jc65@ualberta.ca).
Wenshan Huang is with the Department of Psychology, University of Alberta, Edmonton, AB T6G 2R3, Canada.

Digital Object Identifier 10.1109/MCAS.2024.3358277
Date of current version: 15 August 2024

III. Solution Proposal

The AīCare ecosystem offers a platform integrating senior care IoT devices, such as smart remotes, wheelchairs, and watches, with AI-driven software. As illustrated in Fig. 1, these elements are interconnected, with health data stored securely in the cloud using Fully Homomorphic Encryption, ensuring data privacy during transmission. This encryption allows for computations on encrypted data directly. The system employs AI for data analysis, health monitoring, and user-specific controls, continuously refining its AI model for personalized care. With advancements in large language models (LLMs), the virtual assistant within the system becomes more adept and tailored, improving its support capabilities as it learns from user data.

IV. Innovation Framework

The Purpose Launchpad (PLP), as our primary innovative framework, serves as our primary navigational tool for this project [2]. We developed the project around a pre-defined Massive Transformative Purpose and utilized Business Model Canvas (Fig. 2) to summarize our business plan.

V. Learning

We carried out multiple interviews with prospective users to gather opinions on our project concepts. From these discussions, we gleaned several crucial understandings: 1) The ecosystem should be accessible and cater to those with disabilities. 2) The product's

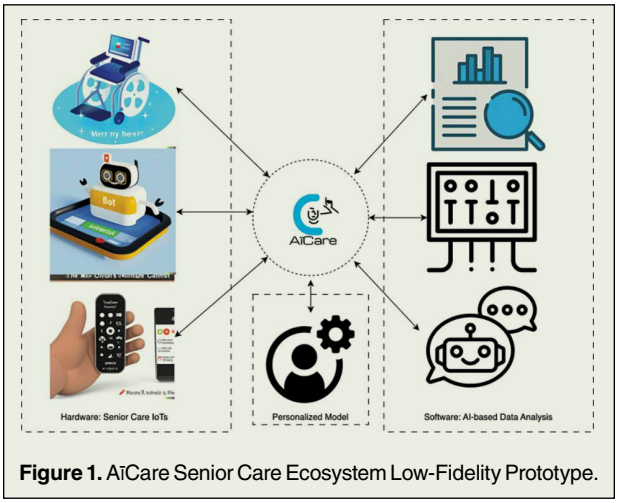


Figure 1. AīCare Senior Care Ecosystem Low-Fidelity Prototype.

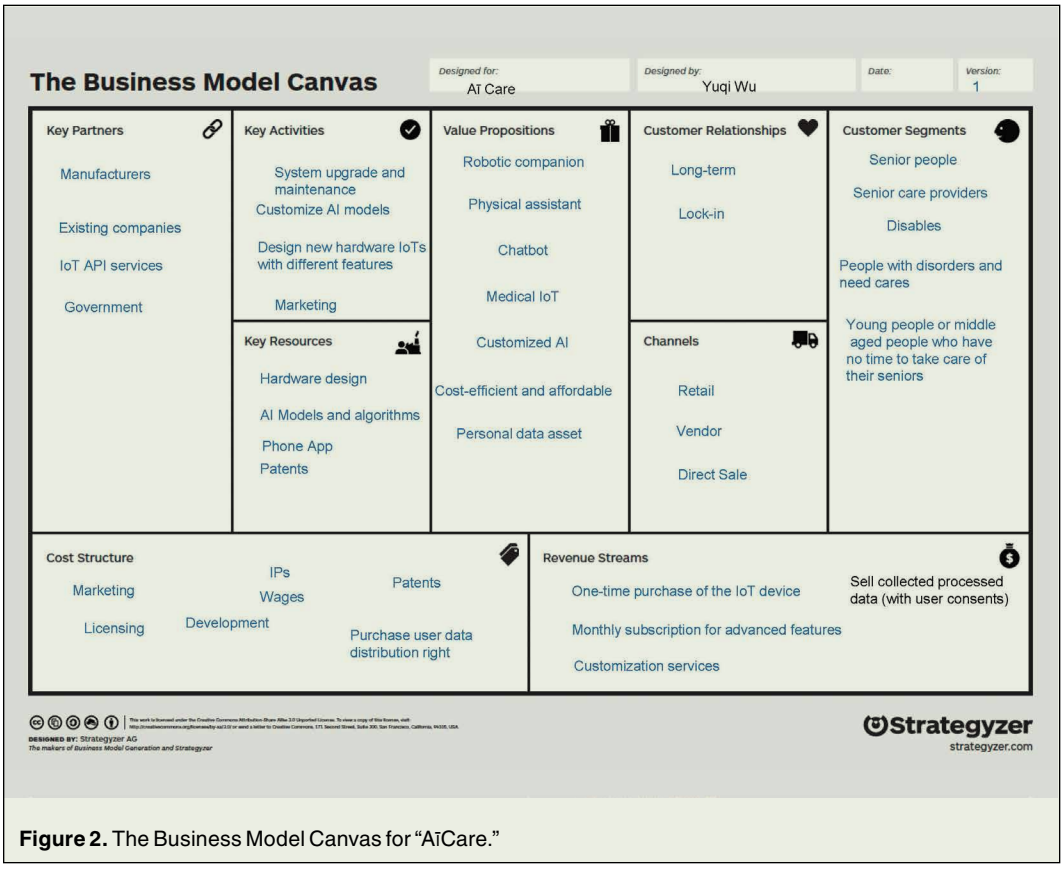


Figure 2. The Business Model Canvas for "AĪCare."

sensitivity should be top-notch. 3) The system must safeguard against any breaches of privacy or data loss.

VI. Conclusion

Utilizing the PLP as our design framework, we've gained deeper insights into the market, our potential customers, and their needs. We firmly believe that our AĪCare senior care ecosystem has the potential to emerge as a leading platform for upcoming senior care technologies.

References

[1] *World: Population (Age 65+)*, United Nations, Department of Economic and Social Affairs, Population Division, New York, NY, USA, 2022.
[2] F. Palao. (2023). *Purpose Launchpad Guide V2.0*. Purpose Alliance. Accessed: Jul. 30, 2023. [Online]. Available: https://www.purpose-launchpad.com/methodology

Yuqi Wu received the B.Sc. degree in electrical and computer engineering from the University of Alberta, Edmonton, AB, Canada, where he is currently pursuing the Ph.D. degree with the Electrical Computer Engineering Department. His work focuses specifically on artificial intelligent methods in affective computing and natural language processing.

Wenshan Huang received the B.Sc. degree in psychology from the University of Alberta, Edmonton, AB, Canada. She is currently pursuing the M.Sc. degree in project management with Northeastern University.

Xuanjie Ye received the Ph.D. degree in electrical and computer engineering from the University of Alberta, Edmonton, AB, Canada.

Jie Chen (Fellow, IEEE) received the Ph.D. degree in electrical and computer engineering from the University of Maryland, College Park, MD, USA. He is currently a Professor with the Department of Electrical and Computer Engineering and an Adjunct Professor with the Biomedical Engineering Department, University of Alberta, Edmonton, AB, Canada. He has coauthored three books, more than 230 journal articles and conference proceeding papers (Citations: 8234).

Innovations Corner

Valery Chebet, *Member, IEEE*, Allan Kipkirui Koech, *Member, IEEE*, and Claude Omosa, *Member, IEEE*

CardioCare for Faster Medical Emergency Response

Abstract

CardioCare helps people with health setbacks like strokes by providing real-time heart monitoring through a wearable device and mobile app. An algorithm analyses data to determine emergencies. The wearable device sounds an alarm and sends notifications for ambulance dispatch or ideal travel routes to medical facilities. It also keeps a QR-coded medical history for proactive care and faster medical response.

Valery Chebet is with the Department of Electrical Engineering, Kenyatta University, P.O. Box 43844-00100, Nairobi, Kenya (e-mail: valerychebet@ieee.org).
Allan Kipkirui Koech is with PSET Developer Enablement, Autodesk, P.O. Box 66488-00800, Nairobi, Kenya, and also with the Department of Computer Science and Digital Technologies, University of East London, E16 2RD London, U.K. (e-mail: allankoech@ieee.org).
Claude Omosa is with the Department of Computer Science, Kenyatta University, P.O. Box 43844-00100, Nairobi, Kenya (e-mail: claudeomosa@gmail.com).

Digital Object Identifier 10.1109/MCAS.2024.3358786
Date of current version: 15 August 2024

I. Problem Analysis and Statement

Emmergencies, ranging from health crises to accidents, can strike anytime and often affect children more than diseases. Breathing issues, shock, and subsequent cardiac arrests can all result from injuries. Stroke is a critical problem, with one in every five people dying within a month and repeated incidents affecting 18% of people within three months and 10% yearly [1]. Stroke survivors and carers endure stress, planning challenges, and communication gaps, emphasizing the need for a solution. Issues develop in maternal healthcare due to delays in accessing hospitals caused by unpreparedness, failure to choose delivery facilities despite antenatal treatment, and transportation obstacles. The lack of medical records containing crucial health information impedes accurate diagnosis and prompt treatment, particularly for new patients [2], [3].

II. CardioCare

CardioCare combines a wearable device with a mobile app for real-time heart rate and vital sign monitoring (Fig. 1).

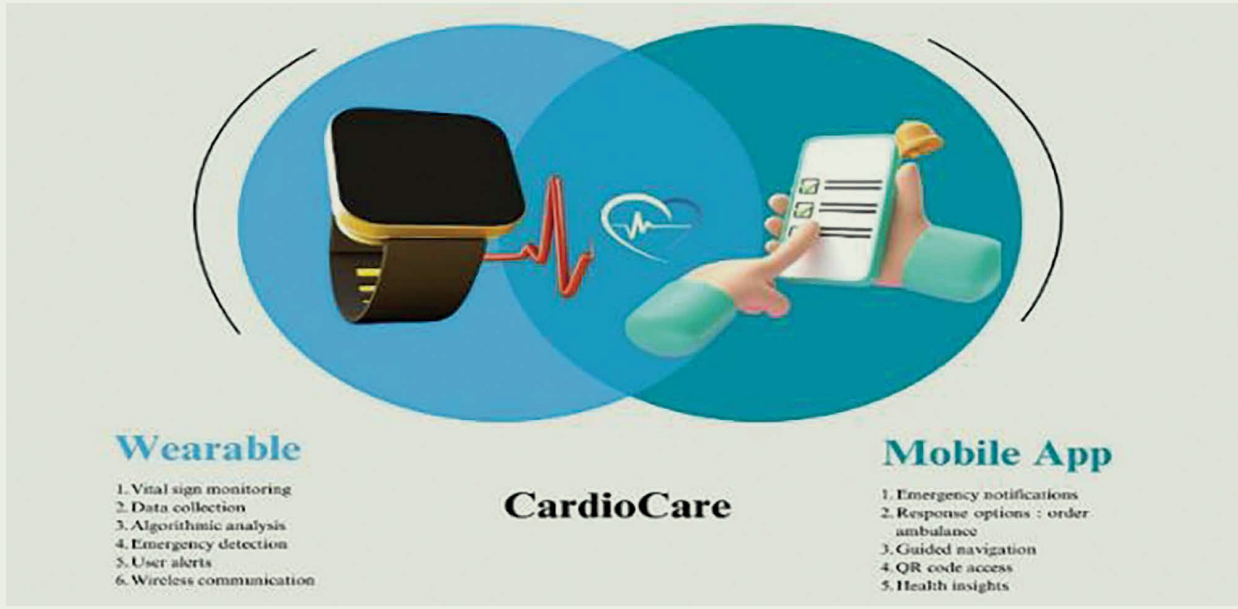


Figure 1. CardioCare features.

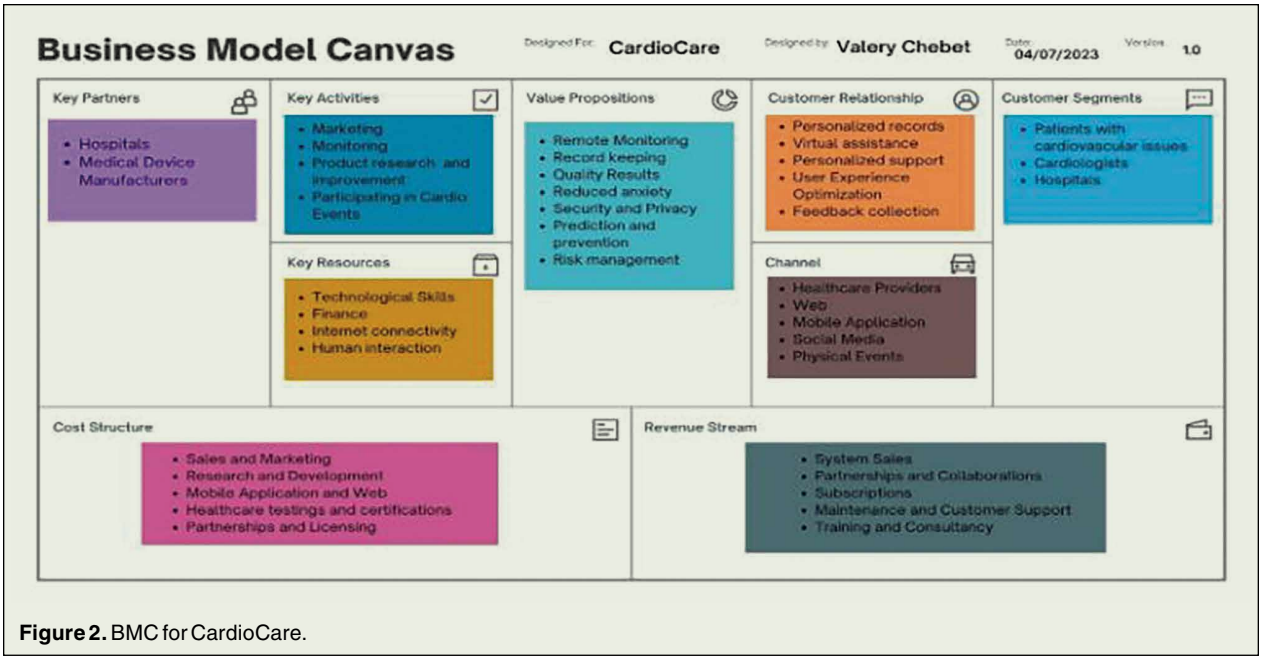


Figure 2. BMC for CardioCare.

This system detects emergencies and suggests appropriate actions by utilising advanced algorithms for vital sign analysis via Bluetooth-enabled wireless technology. It provides secure digital recordkeeping and instant notification of pre-designated emergency contacts. The wearable device uses smartwatch technology to measure critical indicators such as heart rate, breathing rate, temperature, and blood pressure, triggering alarms for unusual patterns and granting access to the user's detailed medical history via a QR code. With user consent, the app analyses data to offer personalized insights, effectively merging technology and healthcare for faster responses and better healthcare management.

III. Innovation Framework

The system processes real-time patient information via a LAN server connected to self-triaging stands. The local database stores all patient data and syncs it online. Transceivers in the hospital connect to the server via Ethernet cables, and integration with existing HMSs is possible through local servers or cloud APIs.

IV. Learnings

We gained insights while using the framework and interviewing the primary market to improve our innovation. Integrating healthcare into existing setups was challenging due to its broad reach (Fig. 2). Patients had unique requirements differing from hospitals. Our solution could be marketed as a lifestyle item, providing a fresh viewpoint on its potential.

V. Conclusion

The PLP approach has helped us better understand value propositions and market dynamics [4]. Our innovative approach attempts to improve healthcare access and effectiveness. With strategic finance, we're focusing on testing and perfecting CardioCare. The system will be optimized in the future, as will collaboration with medical specialists and adherence to global standards.

References

[1] M. W. Moses et al., "Performance assessment of the county healthcare systems in Kenya: a mixed-methods analysis," BMJ Global Health, vol. 6, no. 6, p. e004707, Jun. 2021, doi: https://doi.org/10.1136/bmjgh-2020-004707.
[2] M. Lampi, J. Junker, J. Tabu, P. Berggren, C.-O. Jonson, and A. Wladis, "Potential benefits of triage for the trauma patient in a Kenyan emergency department," BMC Emergency Medicine, vol. 18, Nov. 2018, doi: https://doi.org/10.1186/s12873-018-0200-7.
[3] M. Wakaba et al., "The public sector nursing workforce in Kenya: a county-level analysis," Human Resources for Health, vol. 12, no. 1, Jan. 2014, doi: https://doi.org/10.1186/1478-4491-12-6
[4] "Purpose Launchpad Methodology: Introduction." Novel Innovation Design for the Future of Health: Entrepreneurial Concepts for Patient Empowerment and Health Democratization, edited by Michael Friebe, Cham, Springer International Publishing, 2022, pp. 189–205, doi. org/10.1007/9783031081910_19

Allan Kipkirui Koech, *Member*, *IEEE*, and Valery Chebet, *Member*, *IEEE*

LinderCare: Your Triage Solution to Better Health

Abstract

LinderCare is a triaging medical device that combines a wristband for continuous patient vital monitoring and geolocation within the hospital for real-time patient triaging. This solution is designed to speed up the triaging process, easing the workload on medical personnel and enabling them to respond to emergency cases quickly. LinderCare can enhance service delivery by streamlining the triage process, benefiting patients and healthcare providers. By implementing this innovative system, we expect significant improvements in healthcare services, ultimately leading to better well-being for patients and healthcare providers in Kenya.

Allan Kipkirui Koech is with PSET Developer Enablement, Autodesk, P.O. Box 66488-00800, Nairobi, Kenya, and also with the Department of Computer Science and Digital Technologies, University of East London, University Way, E16 2RD London, U.K. (e-mail: allankoech@ieee.org).
Valery Chebet is with the Department of Electrical Engineering, Kenyatta University, P.O. Box 43844-00100, Nairobi, Kenya (e-mail: valerychebet@ieee.org).

Digital Object Identifier 10.1109/MCAS.2024.3360374
Date of current version: 15 August 2024

I. Problem Analysis

Kenya's healthcare system is under strain from rising populations and income disparities, resulting in overcrowded public facilities and extended wait times [1]. Triage, essential for prioritizing care, is often slow and flawed [2], [3]. While some hospitals have implemented queue systems, these do not always consider the urgency of medical conditions, leading to potential complications for patients.

II. LinderCare

The system includes self-triaging kiosks where users enter information and receive a linked wristband (Figure 1). This wristband features geo-tagging to locate users within the facility and has a screen displaying navigational instructions.

The system processes real-time patient information via a LAN server connected to self-triaging stands. The local database stores all patient data and syncs it online. Transceivers in the hospital connect to the server via Ethernet cables, and integration with existing HMSs is possible through local servers or cloud APIs.

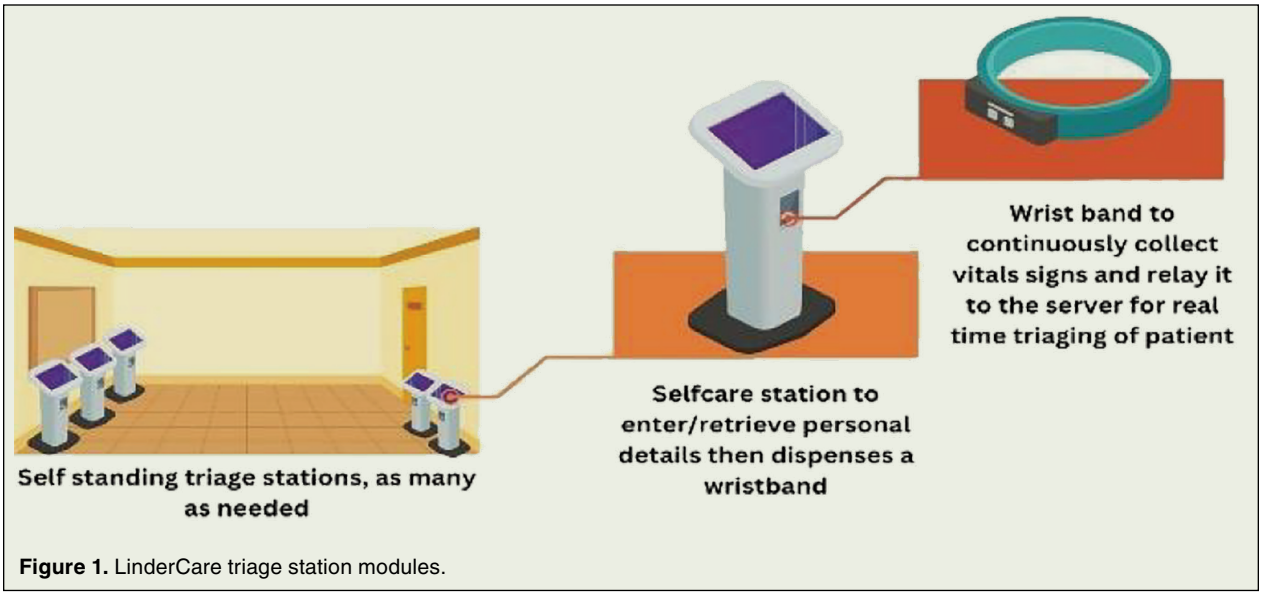


Figure 1. LinderCare triage station modules.

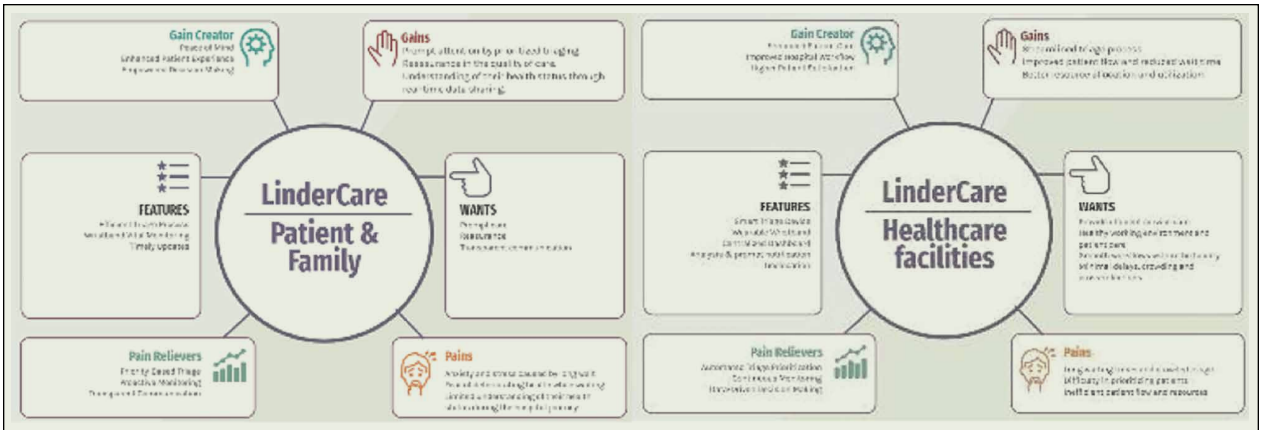


Figure 2. VPCs for patients and healthcare providers.

III. Innovation Framework

LinderCare uses the continuous innovation framework to improve technology and client alignment. The wristband automates patient prioritization in hospitals, monitors vital signs, and provides real-time data, enhancing workflow and reducing wait times (Figure 2). This approach leads to better patient care and increased satisfaction for healthcare providers.

III. Key Learnings

LinderCare was developed based on Purpose Launchpad radars (PLP) insights, interviews, and SWOT analysis [4]. The smart triage device and bracelet cut wait times, while continuous vital sign monitoring improves emergency medical aid. The centralized dashboard improves patient care efficiency. A key learning was the critical need for continuous triaging of patients, unlike the traditional one-off triaging on arrival.

IV. Conclusion

Proper assessment and prioritization are crucial for high-quality patient care. Implementing these practices streamlines hospital workflows and improves healthcare delivery. Prioritizing these changes benefits all parties involved.

References

[1] M. W. Moses et al., "Performance assessment of the county healthcare systems in Kenya: A mixed-methods analysis," *BMJ Global Health*, vol. 6, no. 6, Jun. 2021, Art. no. e004707, doi: 10.1136/bmjgh-2020-004707.

[2] M. Lampi et al., "Potential benefits of triage for the trauma patient in a Kenyan emergency department," *BMC Emergency Med.*, vol. 18, no. 1, p. 49, Nov. 2018, doi: 10.1186/s12873-018-0200-7.

[3] M. Wakaba et al., "The public sector nursing workforce in Kenya: A county-level analysis," *Hum. Resour. Health*, vol. 12, no. 1, p. 6, Jan. 2014, doi: 10.1186/1478-4491-12-6.

[4] M. Friebe, "Purpose Launchpad methodology: Introduction," in *Novel Innovation Design for the Future of Health: Entrepreneurial Concepts for Patient Empowerment and Health Democratization*. Cham, Switzerland: Springer, 2022, pp. 189–205, doi: 10.1007/9783031081910_19.

Tianmin Kong, *Member, IEEE*, and Ava Hedayatipour, *Member, IEEE*

Oximeter for All: An Innovative Look in Inclusive Physiological Monitoring

Abstract

The "Oximeter for all—OXIfA" is an innovative wearable device designed to address common challenges in PPG sensors. It offers accurate blood oxygen measurements, regardless of skin pigmentation and environmental factors, thanks to a dynamic feedback system. Additionally, it integrates various algorithms for providing comprehensive health insights.

Index Terms—PPG, heartrate, analog design, CMOS, innovation.

The authors are with the Department of Electrical Engineering, California State University, Long Beach, CA 90840 USA (e-mail: ava.hedayatipour@csulb.edu).

Digital Object Identifier 10.1109/MCAS.2024.3363608
Date of current version: 15 August 2024

The wearable electronics market has witnessed significant growth due to rising demand for health monitoring and bio- metric data access from companies like Apple, Garmin, and Fitbit. Projections suggest a compound annual growth rate of 13.89%, reaching a market cap of $392.4 billion by 2030. This growth has spurred technological innovations, with a focus on both software and hardware enhancements [1].

I. Problem Statement

PPG sensors on wearables face signal noise from motion artifacts, skin pigment variations, and environmental factors. Studies on Apple and Garmin devices revealed inaccuracies linked to tattoos and skin tones. Results showed inaccuracies in SaO_2 measurements, more pronounced in darker skin tones, particularly at SaO_2 levels

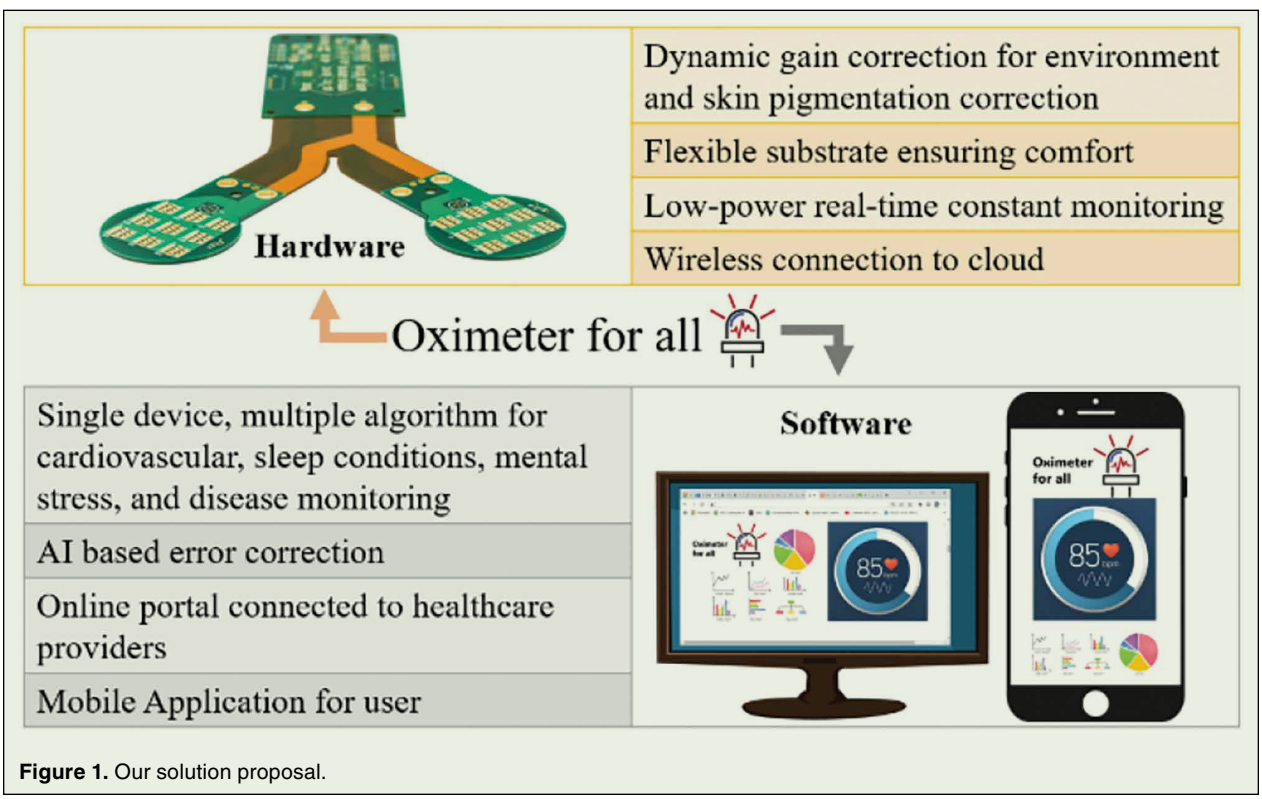


Figure 1. Our solution proposal.

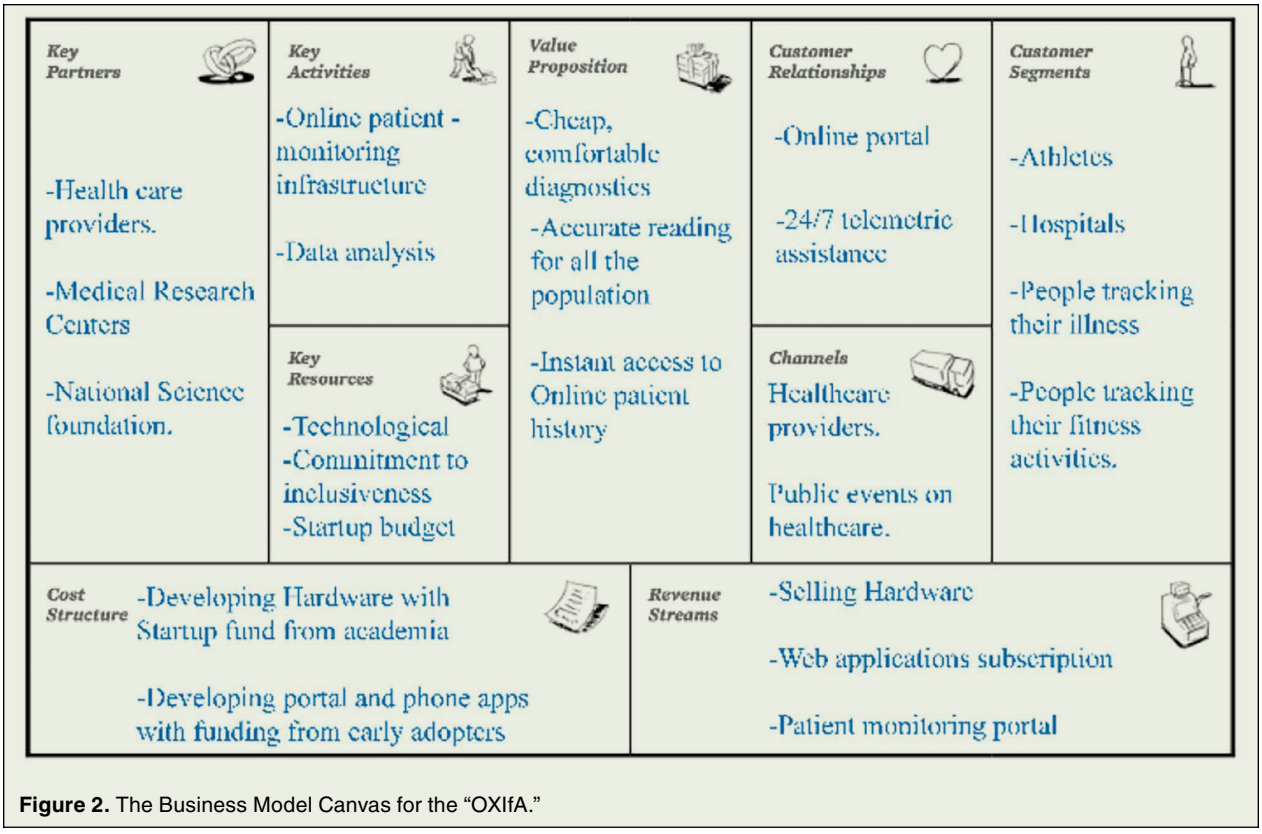


Figure 2. The Business Model Canvas for the "OXIfA."

of 60%–70%. Dark skin had an average overestimate of 3.56 +/- 2.45%, while light skin had .37 +/- 3.20%. Correcting this error is crucial for building trust.

II. Solution Proposal

The OXIfA is an affordable, flexible wearable device with a Photoplethysmography (PPG) sensor. It uses dynamic feed- back to provide accurate blood oxygen measurements despite variations in skin pigmentation, light, and temperature and includes multiple algorithms for cardiovascular monitoring, sleep analysis, stress assessment, and disease diagnostics. The system consumes minimal power and compensates for temperature variations. Developed in 45 nm CMOS technology, it leverages AI, machine learning, and cloud computing for personalized insights (Fig. 1).

III. Innovation Framework

Our innovation approach from problem to solution involves using a framework and step-by-step guidance (Purpose Launchpad—PLP) to foster innovation and early translation from exploration to implementation [2]. We developed a Massive Transformative Purpose (MTP), a vision and mission to provide equitable health monitoring through affordable, accessible, and comfortable tech-powered solutions. Value propositions were formulated through "How Might We" (HMW) and "So That" (ST) questions, addressing skin tone variations, environmental noise, comfort, precision, and usability. We also developed a business model (Fig. 2) that involves key partners, activities, resources, and channels promoting the OXIfA.

IV. Learnings

Interviews with potential users are essential for product development. They offer valuable insights into user needs, preferences, and pain points. Through these interviews, we discovered key insights: 1) Battery life is a critical feature. 2) Leveraging AI can enhance marketing. 3) Users prefer a smartphone app over a web portal. Engaging with users, designers, and developers helps identify and address specific issues, tailoring the product to meet real-world demands leading to a higher adoption rate.

V. Conclusion

The PLP process helped us to define the needed value propositions and to understand the market dynamics. We are now more confident that our OXIfA could offer a promising solution in the monitoring wearable market.

References

[1] (Aug. 2022). *Wearable Technology Market*. [Online]. Available: https://www.precedenceresearch.com/wearable-technology-market
[2] M. Friebe, *Novel Innovation Design for the Future of Health: Entrepreneurial Concepts for Patient Empowerment and Health Democratization*. Cham, Switzerland: Springer, 2022.

Innovations Corner

Ravi Durbha and Valencia Koomson

ChromaSense-Empowering Health, Empowering You

Abstract

ChromaSense, an advanced wearable oximetry sensor, provides superior accuracy compared to traditional vital sign monitors by mitigating noise sources due to user variations (e.g., skin tone, skin thickness, body mass index (BMI), perfusion index), environmental perturbations (e.g., motion artifact, sensor placement), and environmental factors. With advanced sensors and intelligent algorithms, it not only offers precise oxygen saturation readings but also measures important vital signs, such as heart rate, respiration rate, heart rate variability, blood pressure and arterial stiffness based on signal quality of the of recorded photoplethysmogram (PPG) signals.

I. Introduction

Wearable technology has transformed healthcare monitoring with lightweight, continuous devices, particularly aiding chronic conditions like asthma, COPD. These devices benefit individuals, healthcare pros, and researchers, propelling health studies and practices forward. Leading companies like Apple, Fitbit, and Samsung dominate the health wearable market, with forecasts projecting pulse oximeters at $14 billion and blood pressure devices at $25.3 billion by 2028 [1]. Despite challenges, ongoing tech improvements and regulations promise to enhance well-being through these devices.

The authors are with the Department of Electrical and Computer Engineering, Tufts University, Medford, MA 02155 USA (e-mail: ravi.durbha@tufts.edu; valencia.koomson@tufts.edu).

Digital Object Identifier 10.1109/MCAS.2024.3360025
Date of current version: 15 August 2024

II. Problem Statement

Pulse oximeters are vital in healthcare settings for respiratory assessment but can suffer accuracy issues due to various noise sources like ambient light, motion artifacts, and electromagnetic interference. Study by Sjoding and Michael [2] highlights biases in readings based on skin pigmentation resulting in delays in medical treatment. Limited diversity in calibration data exacerbates this problem, potentially impacting healthcare disparities, diagnoses, and treatment accuracy.

III. Solution Proposal

ChromaSense (Figure 1), patented in March 2023, is a comprehensive, noninvasive health monitor, ensuring accuracy across skin tones using adaptive calibration

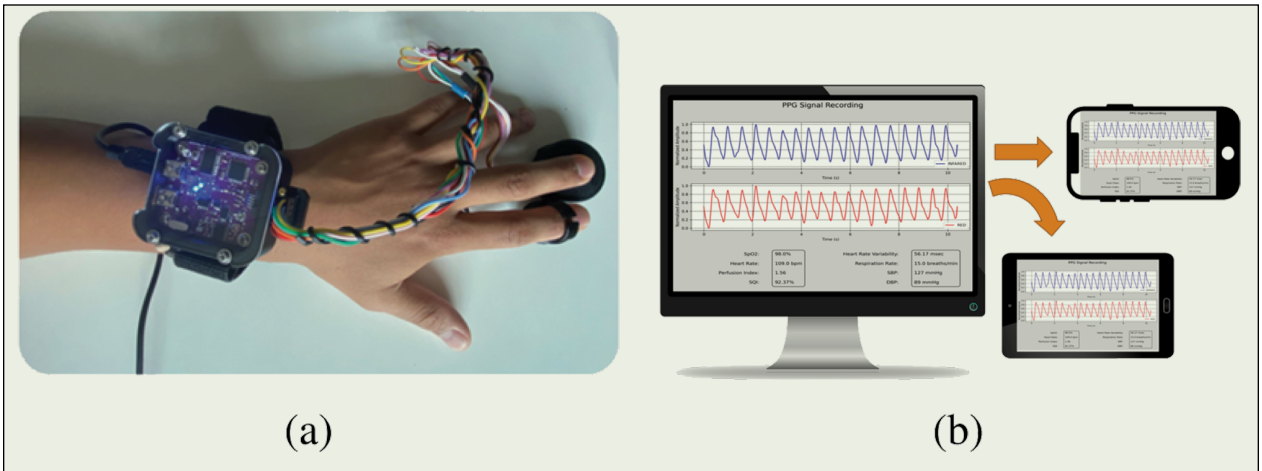


Figure 1. ChromaSense-MVP. (a) ChromaSense—Device. (b) ChromaSense—GUI.

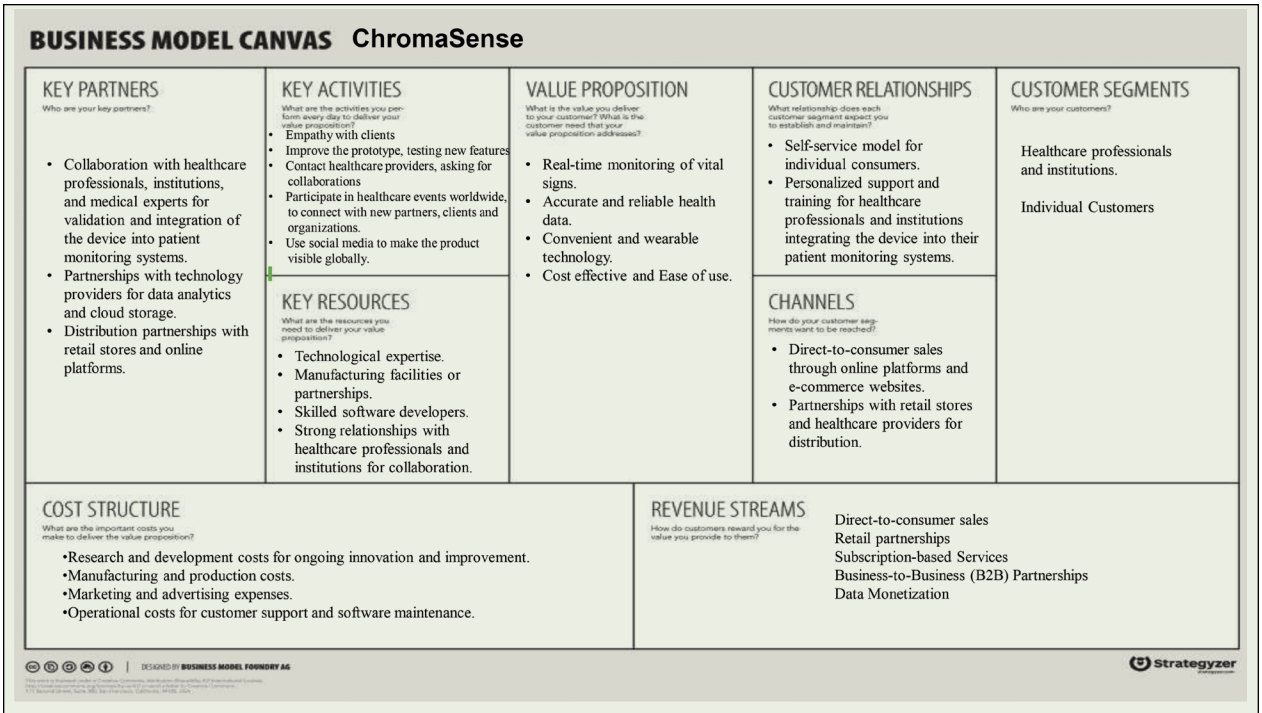


Figure 2. Business Canvas Model.

technique implemented using feedback circuit and advanced algorithms. Initial funding from NSF, NIH, Tufts University, U.S. Army NSRDEC, and Open Oximetry challenge drives hardware and algorithm enhancements. This low-cost sensor architecture provides multiple vital signs to monitor diverse health conditions and research in cardiovascular and respiratory diseases.

IV. Innovation Framework

Integrating customer insights through interviews into the Purpose Launchpad Radar [3] enhanced our Value Proposition Canvas and shaped our business model, as shown in Figure 2 By addressing "How might we" and "So that" questions, we discovered key features for the minimum viable product: accessibility (home care/clinical setting), cost effectiveness, and reliability.

V. Conclusion

ChromaSense is a groundbreaking innovation ensuring accurate health monitoring for diverse populations, offering unique value across healthcare sectors and promoting health equity. By addressing biases and reducing costs, it fosters equal access to quality care, benefiting individuals, professionals, and society at large through improved health outcomes and accessibility.

References

[1] Insight Partners. (2022). *North America mHealth Market.* [Online]. Available: https://www-emis-com.ezproxy.library.tufts.edu/php/search/docpdf?docid =760632768

[2] M. W. Sjoding et al., "Racial bias in pulse oximetry measurement," *New England J. Med.,* vol. 383, no. 25, pp. 2477–2478, 2020.

[3] Francisco Palao. (2022). *Purpose Lauchpad.* [Online]. Available: https://www.purposelaunchpad.com

Zillah Nangobi, *Member, IEEE*, Mutegeki Rodgers, *Member, IEEE*, and Clare Nadunga, *Member, IEEE*

Your Health Application

Abstract

Your Health Application, is a democratized health application built to avail the general public in Uganda with updated and authentic health care information that will ease the delivery of health care services, reduce costs and make healthcare easily accessible to everyone. The information will be used by anyone to know more about illnesses, their causes, signs and symptoms and how to prevent them from reoccurring, the cost of the various treatments, doctors' availability and which hospitals to go to for treatment.

Index Terms—Health application, health democratization.

I. Introduction

LACK of health care equity in Uganda has become an increasing concern as many countries have entered a period of simultaneous high infectious and non-communicable disease burdens, which requires a robust primary care network to diagnose, treat, and monitor patients due to a long continuum of care, drugs and equipment shortages, physical location of health centers, lack of transportation to obtain care, and lack of confidence in the quality of services [1].

II. Problem Analysis

Health care in our Country is sometimes mishandled and patients end up paying the heavy price with the consequences that result from it as a result of having few big hospitals in the country. Some patients end up dying due to improper care among others and health care is sometimes costly and some people end up not receiving the health care they are required to have.

III. Innovation Framework

To help tackle the above problem by evolving initial ideas into viable healthcare innovations and scalable businesses, the Purpose Launchpad framework dedicated to Health Innovations is introduced. The framework is comprised of three phases (Exploration/Discovery,

This article is among the eight submissions for the First Entrepreneurship Initiative by IEEE CASS.

Digital Object Identifier 10.1109/MCAS.2024.3360960
Date of current version: 15 August 2024

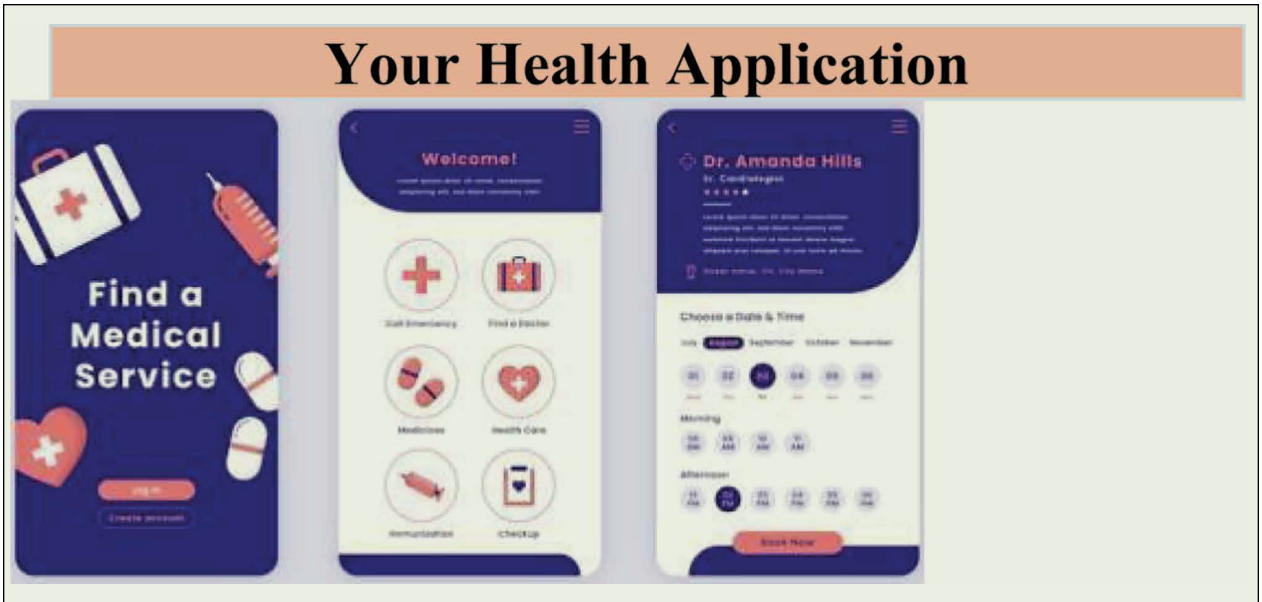


Figure 1. Your health care application.

Evaluation/Validation, and Implementation/Impact Generation) under which are (Purpose, People, Customer, Abundance, Viability, Process, Product, Metrics) which are used in the Innovation process and to ensure a structured learning process [2].

IV. Solution Proposal

Your Health application will provide the following services to its users to ensure health equality in Uganda. It will provide them with information on illnesses with their causes, signs and symptoms, preventive measures and which hospitals handles them best, specialists' schedule in hospitals, first aid that can be given to illnesses, standard cost for the various treatments at the hospitals and emergency services that the patient can easily access (see Figure 1).

V. Learnings

Some sicknesses show different signs and symptoms, it is therefore not easy to exactly tell what one is suffering from, Information on natural home remedies that can reduce the impact of some illnesses should be added to the application/website.

VI. Conclusion

It is important for health care services to be accessible to everyone in order to increase health longevity. Exponential technologies can ease the delivery of health services, reducing costs and making healthcare easily accessible by everyone.

Acknowledgment

We are grateful to IEEE CASS Society, Prof. Dr. Michael Friebe and Prof. Jie Chen and Fakhrul Zaman for providing this training that has been so impactful to us.

References

[1] N. Dowhaniuk, "Exploring country-wide equitable government health care facility access in Uganda," *Int. J. Equity Health*, vol. 20, Jan. 2021, Art. no. 38, doi: 10.1186/s12939-020-01371-5.

[2] O. Morbach and M. Friebe, "Purpose launchpad methodology: Introduction," in *Novel Innovation Design for the Future of Health*, M. Friebe, Ed. Cham, Switzerland: Springer, 2022, pp. 89–205, doi: 10.1007/978-3-03108191-0_19.

CASS Conference Highlights

Rajiv Joshi, *Fellow, IEEE*, Matthew Ziegler, *Senior Member, IEEE*, Jin-Ping Han, and Kaoutar El Maghraoui

6th IBM IEEE CAS/EDS AI Compute Symposium (AICS'23)

The 6th IBM IEEE CAS/EDS AI Compute Symposium was held hybrid at the T. J. Watson Research Center on 28 November 2023. The event was extremely successful and well attended by over 2000 folks from all over the world (in-person and virtual). The symposium featured 8 distinguished speakers (7 from industry and 1 from academia), over 30 student in-person posters, best poster awards, and a panel discussion. The registration list spanned citizens of 53 countries. The theme of the symposium, "From Chips to Chiplets," turned out to be an opportune and important topic for the current semiconductor industry direction. The symposium served as an educational as well as a brainstorming session for industry/academia/students across the world. The symposium covered a range of topics from emerging device technology, innovative circuits, chip and chiplet architecture, advanced packaging technologies, such as 2D to 3D packaging elements, and how these topics drive the rapid growth of AI and generative AI. Dr. Rajiv Joshi, General Chair and IEEE Life Fellow opened the symposium with welcoming remarks along with the goals and accomplishments of this symposium under the auspices of CAS and IBM.

Huiming Bu, VP of Global Semiconductors R&D and Albany Operations, IBM gave the first keynote talk—**"Path to 1 Trillion Transistors in the Era of AI"** (Fig. 1, left). He discussed how artificial intelligence is transforming our world and the demand for computing power is increasing at an unprecedented pace. Then he showed how semiconductor technology innovations in materials, transistors, interconnects, chip architectures, and advanced packaging are being pursued to meet this demand. He emphasized technology development in the semiconductor ecosystem is the key ingredient to making this happen, He shared the roadmap, challenges, and enablers of IBM research's Artificial Intelligence Unit (AIU) Chiplet and advanced packaging technologies are vital to enable the next generation of AIU.

Rob Aitken, Distinguished Architect, Synopsys gave a vivid talk on **"Impedance Matching AI, EDA, Chips, and Chiplets"** (Fig. 1, right). Approximately 35 high-school student participants from two local high schools applauded loudly and commented that they understood most of it and conveyed the following messages. As Moore's law has slowed and Dennard scaling has all but vanished, the technical world is exploring a wide variety of approaches to improve the performance, power, and cost of digital systems. Four key methods are explored in the talk: expanded use of AI, especially generative AI, improvements in design automation, novel chip architectures, and the use of multi-die systems, informally "chiplets." Each of these approaches has already demonstrated the ability to recapture some amount of classical scaling, yet while they are not orthogonal to one another, neither are they always aligned. The talk focused on some of the challenges of the approaches and the resulting opportunities for them to work together, which can be thought of metaphorically as a kind of impedance matching designed to preserve and amplify the improvements available from each source.

Tulika Mitra, Vice-Provost (Academic Affairs) and Provost's Chair Professor of Computer Science at the National University of Singapore (NUS), talked about **"Next-Generation Accelerator Design: Specialize or Generalize?"** (Fig. 2, left). In the AI era, where parallelism is essential for modern workloads, several domain-specific AI hardware accelerators rooted in dataflow computing models have emerged to cater to the demanding performance and energy-efficiency needs of such workloads. Yet, this specialization shift challenges the programmability of generalized solutions cherished by software developers. The central question arises: Can the advantages of both paradigms be harmonized? Similar to how versatile foundation models in generative AI can be fine-tuned to a wide range of specialized tasks, can we find a way to blend the adaptability of general-purpose processors with the strength of specialized accelerators? Further, software-defined hardware accelerators, an innovative, domain-agnostic embodiment

The authors are with the IBM T. J. Watson Research Center, Yorktown Heights, NY 10598 USA.

Digital Object Identifier 10.1109/MCAS.2024.3395580
Date of current version: 15 August 2024

Fig. 1. Dr. Huiming Bu (Left) delivering a keynote talk **"Path to 1 Trillion Transistors in the Era of AI"** and Dr. Robert Aitken (right) giving a presentation related **"Impedance Matching AI, EDA, Chips, and Chiplets."**

Fig. 2. Prof. Tulika Mitra (Left) presenting **"Next-Generation Accelerator Design: Specialize or Generalize?"** and Dr. Dae-Woo Kim (right) describing **"3D package solution, the new boundary of Si and package technology."**

of the dataflow computing model are presented. This approach envisages a synergistic hardware-software co-design strategy, allowing the same silicon chip to be morphed and instantiated to support various dataflows via software. Consequently, the system offers the efficiency of specialized accelerators while preserving the flexibility needed to accommodate diverse applications through generalization. The challenges posed by this novel paradigm were described along with a spotlight on the substantial opportunities it presents in the realm of accelerator design.

Andre Tost, Distinguished Engineer, IBM Watsonx Client Engineering, presented recent developments in the talk **"Generative AI in the Enterprise World with Watsonx"** (Fig. 3, left). The advent of generative AI technology has revolutionized the enterprise world, transforming business processes and user experiences alike. One key aspect of this shift is the integration of Large Language Models within IT landscapes, enabling increased automation and natural language interfaces. This talk delves into real-world use cases that leverage generative AI, showcasing concrete tasks assigned to Large Language Models across diverse industries and regions. These ideas address unique challenges that arise with the implementation of generative AI. Then strategies for addressing privacy, compliance, and trustworthiness

concerns, as well as approaches for managing the increased resource consumption associated with this technology are explored. Then examples of generative AI's impact on various industries, such as finance, manufacturing, and retail are highlighted. The usage of Large Language Models to streamline processes, enhance customer interactions, and unlock new business opportunities are showcased. Furthermore, the ethical considerations that accompany the use of generative AI, by examining the importance of transparency in AI decision-making, and the need for compliant and representative training data are stressed.

Dae-Woo Kim, Corporate VP, AVP, Samsung Electronics talked about **"3D package solution, the new boundary of Si and package technology"** (Fig. 2, right). Despite the high cost of the silicon fabrication process, chip sizes would like to increase beyond the reticle size limit by adding more and more functional blocks for high performance computing. In particular, with the continuous demand for high performance and high capacity in memory products, the amount of data created, processed, stored, and transferred is increasing tremendously. To overcome these challenges, advanced packages based on RDL (Re-Distribution Layer), flip chip bonding, and TSV (Through Silicon Via) have been actively used for heterogeneous integration in electronic packages for the past decade. Heterogeneous integration using advanced packaging technology (2.5D and 3D) and chiplets have been attracting a lot of attention as these approaches enable higher bandwidth with low power consumption at a reduced cost. Advanced packaging has been developed for optimum chip-to-chip interconnections, so more and more silicon fabrication processes are being adopted in package technology. The 2.5D silicon interposer architecture has been widely used for horizontal interconnection between logic to logic and logic to high bandwidth memory integration. The 3D stacking architecture is for vertical interconnections enabling small form factor, increasing signal speed, and reducing power consumption and power dissipation. To provide a bonding pitch solution less than 10 um for extreme I/O density, power and signal integrity, and thermal property, HCB (Hybrid Cu Bonding) must be considered for next generation bonding solutions. In this talk, recent advanced package technology and the roadmap for the Samsung AVP business unit was shared for memory, mobile, and HPC products.

Samuel Naffziger, SVP and Corporate Fellow, Advanced Micro Devices, talked about **"Technology and Architecture Requirements for Energy Efficient AI"** (Fig. 3, right). While the explosion in the capabilities of generative AI and the associated benefits has received a tremendous amount of attention, both in the technology world and in popular press, so has the unprecedented amount of compute and energy required to train and serve these extremely complex models. Many of these concerns are well founded given the hundreds of Mega-Watt-hours required to train large language models (LLMs). To frame the challenge, the trends in energy use for generative AI and where the power is consumed for modern training systems need to be understood. In this context, the talk proposes adopting a holistic approach to energy efficiency to address issues related to how to avoid overwhelming the world's power grid and energy generation capabilities. This approach involves the reduction of energy per computation through the use of lower precision math formats and associated algorithms, the adoption of advanced packaging and 3D architectures to reduce data movement power, and advanced optical interconnects. The most powerful lever

Fig. 3. Mr. Andre Toast (Left) presenting a talk related to **"Generative AI in the Enterprise World"** and Samuel Naffziger (right) describing **"Technology and Architecture Requirements for Energy Efficient AI."**

Fig. 4. Dr. Arif Khan (Left) talked about **"The More-than-Moore Juggernaut for Differentiation and Disaggregation for Processors and SoCs in the Generative AI World"** and Dr. Debendra Das Sharma (right) presented **"Universal Chiplet Interconnect Express™ (UCIe™): An Open Interconnect Standard for Innovation with Chiplets."**

ties these improvements together with algorithmic and hardware-software synergy for efficiently mapping AI problems onto the optimized system. If these approaches are adopted, then the history and recent developments in our industry point to an ability to deliver the benefits of LLMs while reining in energy use.

Arif Khan, Sr. Group Director, Cadence, presented a great talk **"The More-than-Moore Juggernaut for Differentiation and Disaggregation for Processors and SoCs in the Generative AI World"** (Fig. 4, left). This past year, ChatGPT was quite the phenomenon as generative AI hit the peak of the hype cycle and made AI part of our everyday lexicon. This presentation discussed the key market trends driving AI and the demand for newer processor/SoC, chip-to-chip, and module architectures that address the needs of this space. The unsustainable pace of AI die-size growth has come up against the reticle limit. The rise in cost per transistor (CPT) is outweighing scaling benefits from advances in generational process technology. The need for high numerical aperture EUV (High NA-EUV) at nodes beyond 3 nm reduces the reticle size by half. These diminishing silicon economies of scale have pushed foundries, EDA companies, and the manufacturing ecosystem to enable chiplet designs. New developments in packaging technology (through-silicon vias and stacking, interposers, bridging, bump-pitch scaling) and standardization of die-to-die interfaces are providing technology gains to offset the challenges of die sizes and CPT.

Standards bodies and IP implementers have risen to the challenge of providing solutions to interface bottlenecks. This talk also highlighted important standards in memory, such as the latest LPDDR and HBM versions, along with key interface standards such as 112G/224G, peripheral component interconnect express (PCIe), and Compute Express Link (CXL), and chiplet and die-to-die interfaces such as Universal Chiplet Interconnect Express **(**UCIe) that are critical to these new architectures for AI products. The juggernaut for differentiation and disaggregation in the *more-than-Moore* era continues to build momentum. These trends and recent technology advances, while exploring future directions—including questions such as "Can AI be used to design the next 3D-IC AI processors?" are explored through this talk.

Debendra Das Sharma, Intel Senior Fellow, co-GM Memory and I/O Technologies, Intel, presented **"Universal Chiplet Interconnect Express™ (UCIe™): An Open Interconnect Standard for Innovation with Chiplets"** (Fig. 4, right). High-performance workloads demand on-package integration of heterogeneous processing units, on-package memory, and communication infrastructure such as co-packaged optics to meet the demands of the computing landscape in the generative AI era. On-package interconnects are a critical component to deliver power-efficient performance with the right feature set in this evolving landscape.

Universal Chiplet Interconnect Express (UCIe), is an open industry standard with a fully specified stack that comprehends plug-and-play interoperability of chiplets on a package; like the seamless interoperability on board with well-established and successful off-package interconnect standards such as PCI Express® and Compute Express Link (CXL)®. The usages and key metrics associated with different technology choices in UCIe and how this open standard could potentially evolve to incorporate more compelling usage models in the future were described.

During the symposium, a poster session as well as a panel discussion was conducted (Fig. 5). The details of the program are listed in the following link. https://www.zurich.ibm.com/thinklab/AIcomputesymposium.html

Fig. 5. Dr. Rajiv Joshi (Left) gave **opening/closing remarks** and Dr. Arvind Kumar (right) moderated the **panel discussion**.

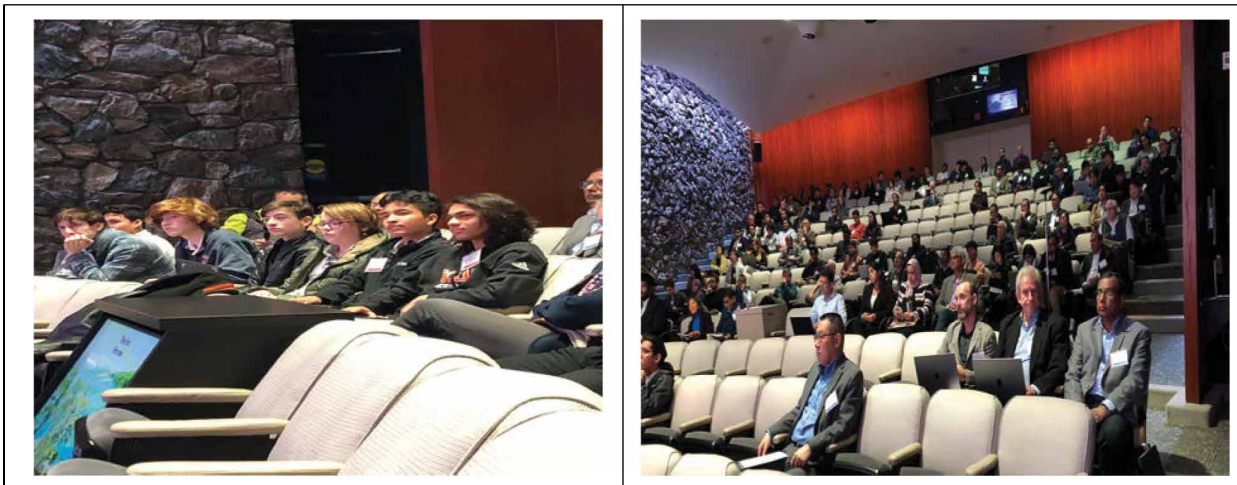


Fig. 6. Yorktown High School Students (left) and the audience (right) engrossed in talks.

This year, the AI Compute Symposium has expanded its participant base to include local Yorktown Heights high school students for the first time (Fig. 6). This initiative marks a significant step towards engaging and nurturing young talent in AI. It provides a unique platform for these students to immerse themselves in the latest trends in AI hardware and software. During the symposium, these bright young minds had the chance to contribute actively, particularly in the poster sessions. Their involvement went beyond mere participation—in a remarkable achievement, one high school student co-authored and presented a poster showcasing their research and insights alongside seasoned professionals.

Overall this symposium included industry veterans and exposed the young and the mature audience to future directions in technology and motivated all to pursue abundant opportunities in the field of chips, chiplets, and systems.

General Chairs
Rajiv Joshi, Matthew Ziegler, and Jin-Ping Han

Technical Program Committee
Anna Topol, Kaoutar El Maghraoui, Krishnan Kailas, Xin Zhang, Arvind Kumar, Linda Rudin, Cheng Chi, Atom Watanabe, and John Rozen

CASS Conference Highlights

Ayan Datta, *IEEE CAS Bengaluru Chapter*

IEEE WINTECHCON 2023: Women in Technology Conference

Celebrating Women in Circuits and Systems: Highlights From IEEE WINTECHCON 2023
21 September 2023|Leela Palace, Bengaluru
Organized by: IEEE Bangalore Section, IEEE CAS Bangalore Chapter, WIE AG Bangalore Section
Theme: Emerging Technologies from Silicon to Software for a Sustainable Future
Website: https://wintechcon.com/

Empowering Innovation: Reflecting on IEEE WINTECHCON 2023

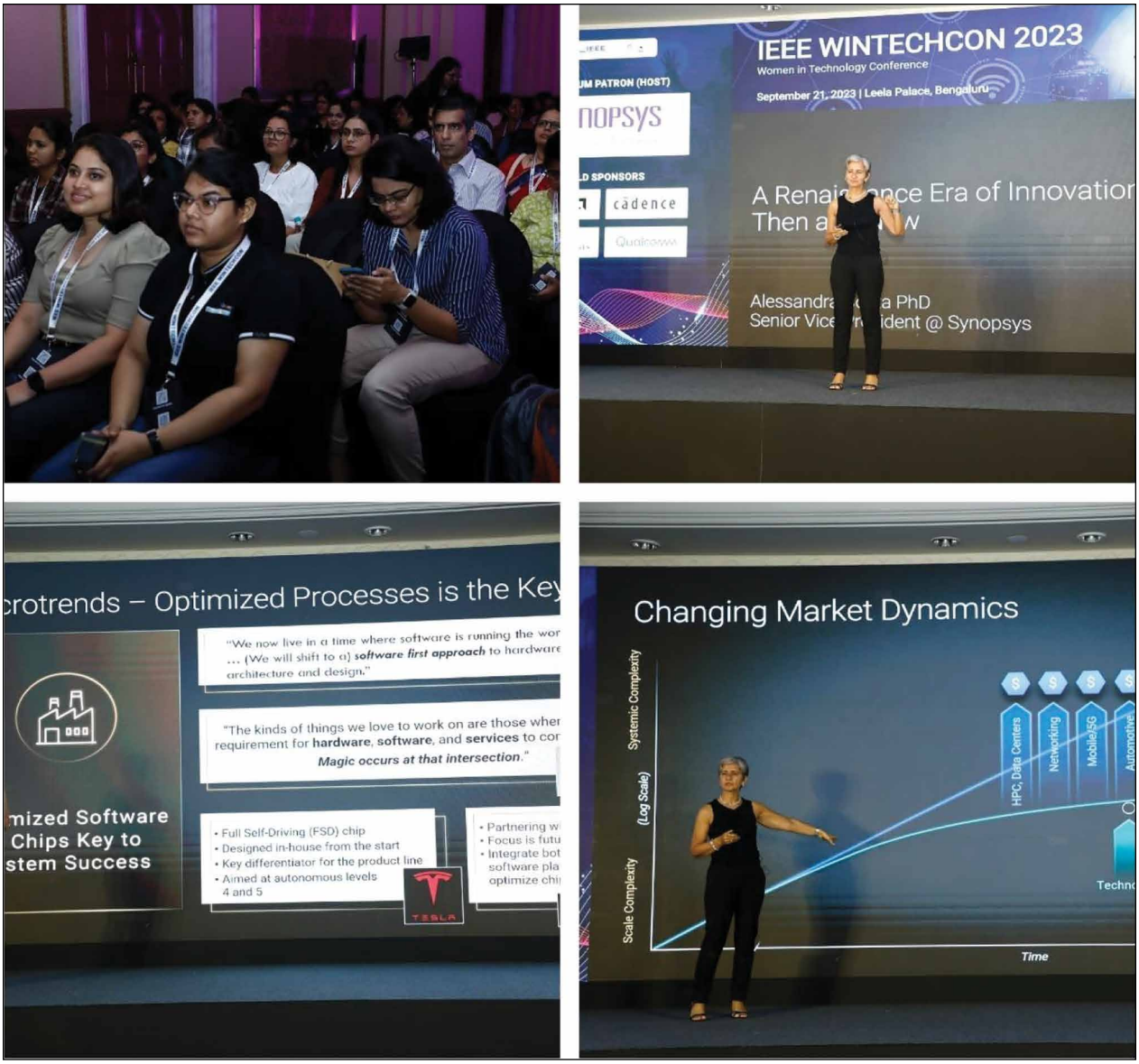
In the bustling metropolis of Bengaluru, amidst the vibrant tapestry of technological innovation, the Leela Palace stood as a beacon of inspiration during the IEEE WINTECHCON 2023. This annual Women in Technology Conference, a collaborative effort spearheaded by senior women technologists alongside IEEE Bangalore Section, IEEE CAS Bangalore Chapter, and IEEE WIE AG Bangalore Section, served as a nexus of brilliance, creativity, and empowerment.

Themed "Emerging Technologies from Silicon to Software for a Sustainable Future," WINTECHCON 2023 emerged as a platform to celebrate the remarkable achievements and breakthroughs of women in technology. From semiconductor design to artificial intelligence, the conference traversed the spectrum of innovation, illuminating the path forward in the ever-evolving landscape of technology.

The inaugural ceremony, graced by esteemed dignitaries including Dr. Aloknath De (IEEE Bangalore Section Chair) and Mrs. Alessandra Costa (SVP Synopsys Inc.), set the stage for the conference's proceedings. Dr. De, in his address, lauded the pivotal role of women technologists in India's semiconductor journey, emphasizing the conference's focal point on Silicon-to-Software innovations. Mrs. Costa, in her keynote speech, invoked the spirit of Renaissance, urging young women to become catalysts for transformative change in the tech realm.

Ayan Datta, Chair of IEEE CAS Bangalore Chapter, aptly captured the essence of the conference, stating,

Digital Object Identifier 10.1109/MCAS.2024.3402189
Date of current version: 15 August 2024



"Breaking Barriers and Building Tomorrow: IEEE WinTechCon 2023 Shines a Spotlight on Women in Technology, where visionaries and innovators converge to celebrate the limitless potential of female technologists! Its a platform where 600+ women from 15+ Semicon companies come together to celebrate this Technutsav. A conference which enables Narishakti for India Semiconductor Mission."

One of the defining features of WINTECHCON 2023 was its comprehensive program, comprising four distinct tracks: VLSI & EDA, Automotive, AI/ML & IoT, and Connectivity & Cloud. These tracks served as conduits for exploration, providing attendees with insights into the latest advancements and trends shaping their respective domains. The technical program of WINTECHCON 2023 was marked by excellence and innovation. With over 650 paper submissions, a meticulous review process involving a pool of 400+ reviewers ensured the selection of 23 papers for presentation. From project demonstrations to poster sessions, the conference served as a canvas for showcasing disruptive innovations driving technological progress.

Keynote speeches by industry luminaries such as Mrs. Alessandra Costa and Dr. Sugandhi Gopal provided profound insights into the transformative power of technology. Mrs. Costa's call for inclusive environments and Dr. Gopal's exploration of medical advancements fueled by technology underscored the profound impact of women in driving innovation across sectors.

Beyond presentations and speeches, WINTECHCON 2023 was committed to nurturing the next generation of technologists. Sixteen students from government colleges in Bangalore were sponsored to attend the conference, offering them a firsthand glimpse into the dynamic world of technology. This initiative underscored the conference's dedication to fostering diversity and inclusion in the tech industry. The conference also featured an array of engaging sessions, including panel discussions and tutorials. The panel discussion on "Learning and Growing your career in the Generative AI era" brought together experts from leading companies to explore the implications of AI/ML in various domains. Discussions ranged from ethical considerations to societal impacts, providing attendees with valuable insights into the evolving landscape of technology.

As the conference drew to a close, the overwhelming response from various sectors of the industry served as a testament to the growing recognition of the importance of diversity, equity, and inclusion in the tech ecosystem. Nithya Raghavan, General Chair of the conference, expressed her satisfaction at the event's success, noting the unprecedented turnout and the palpable enthusiasm among attendees.

In essence, IEEE WINTECHCON 2023 was not merely a conference; it was a celebration of women's contributions

to technology, a testament to their ingenuity, resilience, and unwavering commitment to shaping a sustainable future. As we reflect on the insights and innovations shared at the conference, we are reminded of the transformative power of diversity in driving technological progress and propelling humanity towards a brighter tomorrow.

CASS Conference Highlights

Mariana Siniscalchi and Florence Podevin

Women in Circuits and Systems (WiCAS) at LASCAS 2024

Held on Wednesday, 28 February 2024, from 17:30 to 20:30, in Punta del Este, Uruguay, the Women in Circuits and Systems (WiCAS) event was a collaborative effort organized by Mariana Siniscalchi (Universidad de la República, Uruguay) and Florence Podevin (Université Grenoble Alpes, France). The event was integrated into the International Symposium of the IEEE Circuits and Systems Society (LASCAS 2024), emphasizing the theme "Women in Circuits and Systems: Rhythm of Progress, Focusing on Inclusion." The event aimed to celebrate and empower women in engineering while advocating for diversity and equal opportunities. It comprised three segments:

Segment 1: Speak Up (60 minutes)

The event commenced with an insightful keynote/tutorial session featuring Carolina Mora-López from Imec, Belgium, an esteemed expert in circuits and technologies for implantable biomedical devices. Carolina's presentation, depicted in Fig. 1 as she addresses the audience, not only showcased her technical expertise but also served as an inspiration for aspiring women engineers. Her remarkable career journey exemplified the impact of empowered women in the field, setting the tone for the event's focus on empowerment and inclusivity. Her talk was particularly appreciated, drawing an audience of about 70 attendees to the event.

Digital Object Identifier 10.1109/MCAS.2024.3393890
Date of current version: 15 August 2024

Segment 2: Fostering Inclusive Views (30 minutes)

Following the keynote, participants engaged in an interactive workshop aimed at addressing discrimination against women in the engineering workplace. Two typical workplace scenarios in engineering were enacted to illustrate common microaggressions. In the first scene, shown in Fig. 2, a male professor and his two Ph.D. students, Julia (she) and Rodrigo (he), were portrayed. The professor assigned technical tasks to Rodrigo and administrative tasks to Julia, reflecting a long-standing distribution of roles. The second scene, shown in Fig. 3, depicted a female client consulting with a team of partners developing a product, Rosina (she) and Andrés (he). The client assumed Andrés, the male, to be the team leader, disregarding Rosina's contributions. After the scenes were enacted, attendees discussed potential actions they would have taken if they were involved in the situations presented, facilitated by Florence Podevin as the moderator as seen in Fig. 4. They explored strategies to support the targeted woman from the perspective of various characters. Audience participation enriched the discussion, providing diverse viewpoints and fostering learning opportunities for all. The inclusive

Figure 1. Carolina Mora-López giving the keynote speech.

Figure 2. Scene portraying the Professor and the two students.

Figure 3. Scene portraying the client and the two partners.

nature of the event, with a diverse group of participants, greatly facilitated meaningful exchanges. An open-minded discussion emerged, particularly focusing on the nature of solutions that should encompass both vertical, top-down approaches, commonly seen in WiCAS event organization worldwide, and horizontal approaches, akin to the stance individuals adopt in their daily lives.

Segment 3: Embrace the Rhythm of Diversity and Taste and Socialize (90 minutes)

Participants engaged in an inclusive dance workshop, delving into traditional Uruguayan dances, celebrating diversity and fostering a sense of empowerment and self-expression. Led by seasoned instructors, the workshop showcased Uruguayan culture and underscored

Figure 4. Co-chair Florence Podevin moderating the interactive workshop.

Figure 5. Musicians at the dancing workshop.

Figure 6. Dancing workshop and cocktail.

the shared experiences of individuals in engineering and dance.

The limited presence of women at the conference quickly became apparent, prompting reflection among the audience. As a result, men partnered up for the dance, which may have posed a challenge for some, further stimulating thought and discussion. Eventually, participants adapted to the circumstances and fully embraced the activity, seizing the opportunity to learn something new. Figures 5 and 6 visually depict this segment of the event.

The event smoothly transitioned into a relaxed cocktail reception, facilitating networking and reflective discussions, fostering connections, and camaraderie among attendees.

The WiCAS co-chairs express their gratitude to the LASCAS 2023 organizing committee and deeply thank the IEEE Circuits and Systems Society for the financial support. Special thanks to Mariana del Castillo, Rocío Cabral, Andrés Seré, Julia Azziz, Varinia Cabrera, Leonardo Steinfeld, Rodrigo García, and Rosina D'Eboli.

CASS Conference Highlights

Tuba Ayhan and Okan Zafer Batur

DEICAS-YPCAS Event at ICECS 2023

On 5 December 2023, a joint Diversity, Equity, and Inclusion in Circuits and Systems (DEICAS) and Young Professionals in Circuits and Systems (YPCAS) event took place in the 30th IEEE International Conference on Electronics, Circuits and Systems (ICECS), Istanbul, Turkey, which was supported by IEEE DEICAS and IEEE YPCAS committees. In alignment with the principles outlined in the IEEE Diversity Statement, which asserts "*IEEE is committed to advancing diversity in the technical profession, and to promoting an inclusive and equitable culture*," this event focused on fostering an equitable culture within the circuits and systems field. Moreover, this event also supports the career development for IEEE Circuits and Systems Society (CASS) members, particularly those at early career stages. This event provided a platform for professionals from diverse backgrounds, ages and occupations in CAS to engage in discussions regarding the essential conditions and opportunities for creating an inclusive and equitable environment.

A panel on "Innovation and Start-ups in CAS," organized by Tuba Ayhan (MEF University) and Okan Zafer Batur (Bilgi University) on 5 December 2023, catered specifically to individuals in the early steps of their CAS career or those who would like to implement their idea by themselves, or just exploring the options. The panel featured four distinguished speakers: **Özlem Özbay**, an

Figure 1. Panelists of the joint DEICAS-YPCAS event that took place during the ICECS 2023, from left to right, Özlem Özbay, Anıl Akseki, Akın Şibay, Melike Atay Karabalkan and moderator Tuba Ayhan.

Digital Object Identifier 10.1109/MCAS.2024.3368968
Date of current version: 15 August 2024

Across diverse fields, varying definitions emerge for what qualifies as an innovative process or product, i.e., cost reduction and the inclusion of security parameters.

experienced Electronics Engineer affiliated with companies of various scales; **Melike Atay Karabalkan**, a seasoned researcher, designer, and trainer with a decade of experience in CAS projects; **Anıl Akseki** and **Akın Şibay**, enthusiastic co-founders of a young interdisciplinary research company. The panelists are seen in Figure 1.

After a brief introduction to DEICAS, YPCAS, and an overview of the panelists, the session unfolded in four rounds. The subsequent sections provide key highlights of the discussion.

The first round centered on innovation. Across diverse fields, varying definitions emerge for what qualifies as an innovative process or product, i.e., cost reduction and the inclusion of security parameters. However, panelists agree that innovations and businesses developed by start-ups are the results of sustained effort. This sentiment, articulated by M. Atay Karabalkan, underscored the importance of patience and not giving up. Ö. Özbay explained the evolution of innovation in research and design through years. She emphasized the monumental shifts in communication and design tools, providing us with opportunities for visualization, simulation, and creation. Recently, the ascendancy of digitalization and data-driven technologies steer numerous sectors. Reflecting on her role as a digital circuit designer, M. Atay Karabalkan remarked, "*we see that the boundaries of the digital design world are being pushed*."

The second and third rounds focused on the "Start-up Journey" from both a knowledge and product perspective, respectively. We compared the methods to address design challenges that extend beyond the team's expertise. Ö. Özbay initiated the discussion by presenting the critical factors to consider before making decisions on hiring, training, subcontracting, and outsourcing. These factors encompassed budget constraints, time considerations, relevance to the project, scope and depth of knowledge required,

Figure 2. More than 40 people attended the panel, some of them are in the group photo.

and accessibility of knowledge sources. Drawing from their early experiences in the startup realm, Akseki and Şibay candidly shared insights, cautioning the audience about seeking external support before tasks become overwhelming. M. Atay Karabalkan said they never outsource in their company, but for other companies they provide training, offer design support, or a combination of both. This underscores that the knowledge acquisition method is project-specific, while maintaining know-how within a company depends on cultivating a supportive environment. Then the audience was briefly informed about finding financial support for design and productization. The panelists concluded that the funding agencies and practices vary based on the target sector and economic region. Nevertheless, stability of the team and low team circulation in long term projects are the key factors for taking cutting-edge research and designs from the laboratory to the market.

Finally, the panelists shared advice for young processionals. M. Atay Karabalkan emphasized the significance of combining academic expertise with hands-on experience, underscoring its positive impact on startup ventures.

Drawing attention to the difficult paths on their start-up journey, Akseki and Şibay suggested to work on different projects concurrently to widen knowledge in different fields. However, they cautioned against the expectation of commercializing everything simultaneously. Özbay shared an inspiring perspective, stating "*I have never seen anyone who worked diligently and did their utmost with good intention not succeed*." Her advise included focusing on strengths, working projects that ignite excitement, keeping up with the new developments, specializing, seeking a job where state-of-the-art technology is used, and maintaining a balance between work and private life.

The panel was well attended (see the group photo in Figure 2). The ICECS 2023 conference DEICAS and YPCAS co-chairs thank all attendees and panelists, as well as the IEEE YPCAS co-chairs Chi-Seng Lam and Yang Jiang, and IEEE DEICAS co-chairs Giulia Di Capua and Yoko Uwate, for their support during the planning of this event. They also thank the IEEE Circuits and Systems Society for the financial support received for organizing the event.

This report was prepared by Tuba Ayhan (MEF University) and Okan Zafer Batur (Bilgi University).

CASS Conference Highlights

Ricardo Reis *(CASS BoG R9 Representative)* and Antonio Augusto *(UFPB)*

IEEE CASS Tour Paraiba 2023

I. The IEEE CASS Tour Paraíba 2023

The IEEE CASS Tour Paraíba 2023 was held at the Federal University of Paraíba (UFPB), at João Pessoa, in 18 September. The Tour included a set of technical talks as well as a special talk about the IEEE Circuits and Systems Society, presenting CASS activities and opportunities. It was organized a panel about "Education on Computing and Microelectronics," with the participation of professors **Antonio Augusto** (UFPB), **Antonio Cavalcanti** (UFPB), **Raimundo Freire** (UFCG), **Romulo Câmara** (UFPB), and **Ricardo Reis** (UFRGS). Figure 1 shows the full program of the event. The panel was an opportunity for the colleagues of the Electrical Engineering and Computer Science from the UFPB and UFCG (Federal University of Campina Grande) to discuss the challenges in education to cope with trends in circuits and systems (Figure 2).

The audience was composed by students and professors from computer science and electrical engineering of UFPB, and professors and students from UFCG that came by bus from Campina Grande, the location of UFCG. It is a piece of evidence that continuity of actions should take place to keep motivating students and professors. CASS has a key role in this process.

The CASS Tour Paraíba is part of a set of CASS Tours being organized in R9. It is possible to know about some past tours by reading the reports published in past editions of the IEEE Circuits and Systems Magazine [1], [2], [3], [4], [5], [6], [7].

Digital Object Identifier 10.1109/MCAS.2024.3358278
Date of current version: 15 August 2024

Figure 1. Flyer of the IEEE CASS Tour at João Pessoa.

Figure 2. Photo taken during the Panel on Education.

The next edition of 37th Symposium on Integrated Circuits and Systems Design, SBCCI 2024, will be hosted by the team of UFPB, 26–30 August 2024. The CFP deadline is 1 April 2024. SBCCI is co-sponsored by IEEE CASS. More information at www.sbcci.org.br.

II. Conclusion

It was a great experience to organize the IEEE CASS Tour Paraíba 2023. The participation of professionals and students was great, and they are motivated to do a petition to launch a CASS Student Branch Chapter in the region. We expect an increase in CASS membership in the region.

Acknowledgment

To all ones that helped in the organization of the CASS Tour Paraiba 2023.

References

[1] R. Reis, V. Grimblatt, and M. Delgado-Restituto, "IEEE CASS tour Argentina 2022," *IEEE Circuits Syst. Mag.*, vol. 23, no. 1, pp. 87–89, 1st Quart., 2022, doi: 10.1109/MCAS.2023.3244430.

[2] F. Olivera and R. Reis, "IEEE CASS Rio de Janeiro tour 2022," *IEEE Circuits Syst. Mag.*, vol. 23, no. 2, pp. 64–65, 2nd Quart., 2023, doi: 10.1109/MCAS.2023.3267939.

[3] R. Reis and R. Murphy, "IEEE CASS tour Mexico 2022," *IEEE Circuits Syst. Mag.*, vol. 23, no. 2, pp. 66–70, 2nd Quart., 2023, doi: 10.1109/MCAS.2023.3267987.

[4] R. Reis and R. Taco, "IEEE CASS tour Ecuador 2023," *IEEE Circuits Syst. Mag.*, vol. 23, no. 3, pp. 72–75, 3rd Quart., 2023, doi: 10.1109/MCAS.2023.3302394.

[5] R. Murphy and R. Reis, "CASS tour Mexico 2023," *IEEE Circuits Syst. Mag.*, vol. 23, no. 3, pp. 76–80, 3rd Quart., 2023, doi: 10.1109/MCAS.2023.3302395.

[6] R. Reis and M. Delgado-Restituto, "IEEE CASS tour Puerto Rico 2023," *IEEE Circuits Syst. Mag.*, vol. 23, no. 3, pp. 80–83, 3rd Quart., 2023, doi: 10.1109/MCAS.2023.3302396.

[7] R. Reis and M. Delgado-Restituto, "IEEE CASS tour Panama 2023 and the IEEE R9 meeting," *IEEE Circuits Syst. Mag.*, vol. 23, no. 3, pp. 83–86, 3rd Quart., 2023, doi: 10.1109/MCAS.2023.3302397.

CASS Conference Highlights

Ricardo Reis *(CASS BoG Member)*, Manuel Delgado-Restituto *(CASS President)*, Victor Grimblatt *(CASS BoG Member)*, Walter Calienes-Bartra *(CASS Peru Chapter Chair)*, and Carlos Silva *(PUCPeru)*

IEEE CASS Tour Peru 2023

I. The IEEE CASS Tour Peru 2023

The IEEE CASS Tour Peru 2023 was held from 4 to 8 September, with a set of talks in four locations: Trujillo, Arequipa, and two locations in Lima. The set of talks included technical ones, a special talk about the IEEE Circuits and Systems Society and opportunities, and why to become a CASS member. The talk related to CASS and opportunities present CASS activities such as conferences, workshops, symposiums, seasonal schools, publications, student design contests, the distinguished lecturer program, pre-doctoral grants, student travel grants, and educational talks promoted by CASS. The list of calls organized by CASS was also presented to support the organization of CASS activities.

The main goal of the CASS Tours being organized in R9 is to motivate professionals and students to be active members of CASS, showing them a whole set of opportunities and support the society provides.

The CASS Tour in Trujillo, was organized at the Universidad Privada Antenor Orrego. Trujillo is situated in the north of Peru and has a large set of impressive constructions made in adobe from 1000 to 500BC. Figure 1 presents a photo with the audience in Trujillo.

Figure 2 shows Prof. Walter Calienes-Bartra during his talk in Trujillo. The program of the Tour in Trujillo is presented in Figure 4.

On 6 September, the CASS Tour went to the Universidad Católica Santa María at Arequipa. Figure 3 shows Prof. Manuel Delgado-Restituto talking about CASS opportunities,

Digital Object Identifier 10.1109/MCAS.2024.3358271
Date of current version: 15 August 2024

Figure 1. Photo with part of the participants of the CASS Tour at the Universidad Privada Antenor Orrego in Trujillo.

Figure 2. Prof. Walter Calienes-Bartra during his talk about Radiation Effects on Integrated Circuits.

Figure 3. Prof. Manuel Delgado-Restituto giving his talk in the CASS Tour at Arequipa.

Figure 4. Program of the IEEE CASS Tour in Trujillo.

Figure 5. Program of IEEE CASS Tour in Arequipa.

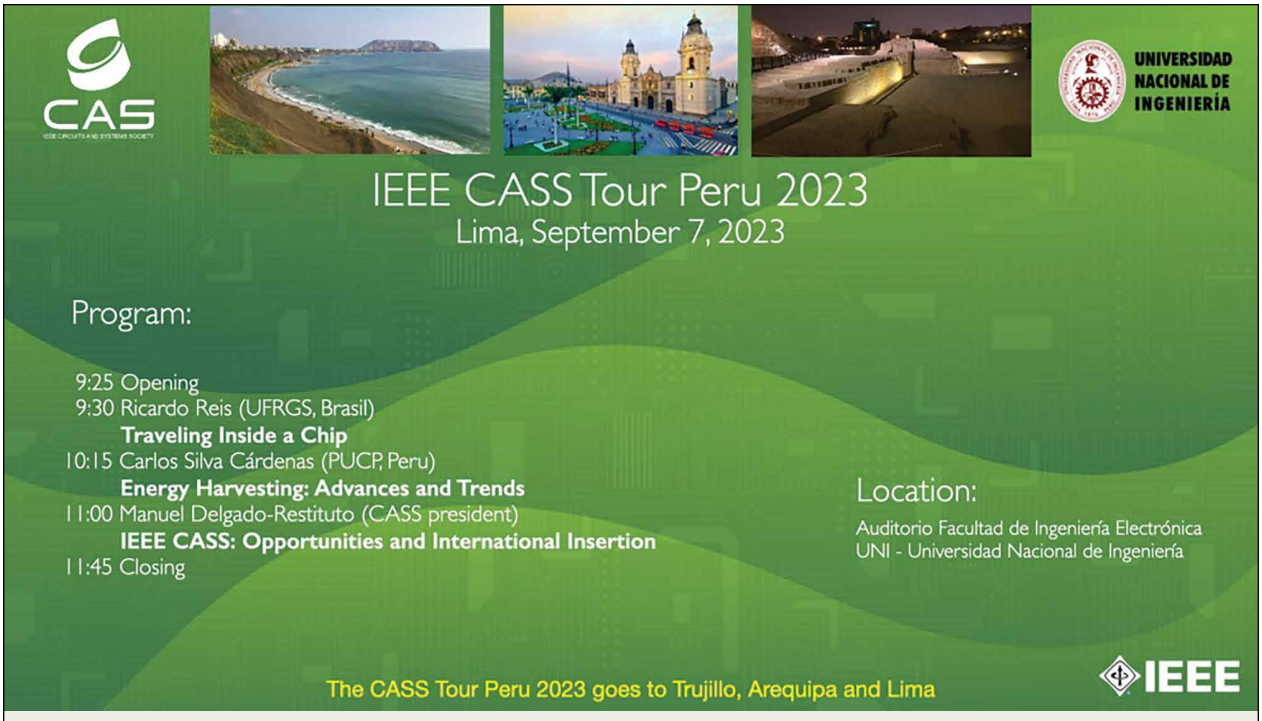


Figure 6. Program of IEEE CASS Tour in Lima, at UNI.

Figure 7. Program of IEEE CASS Tour in Lima, at PUCP.

Figure 8. Some of the participants of the CASS Tour at Universidad Católica Santa María, in Arequipa.

Figure 9. Some of the participants of the CASS Tour at UNI, in Lima.

including the new UNIC-CASS (Universalization of IC Design in CASS) program that gives access to silicon. The Full Program in Arequipa is presented in Figure 5, and Figure 8 presents a group photo done in Arequipa.

On 7 September, the CASS Tour occurred in Lima at UNI (Universidad Nacional de Ingeniería). Figures 6 and 7 present the program at the two locations in Lima. Figure 9 shows some of the participants at UNI, and Figure 10 shows Prof. Carlos Silva

Cardenas doing his talk about Energy Harvesting.

On 8 September, the CASS Tour was organized at PUCP (Pontificia Universidad Católica del Perú) in Lima. Figure 11 shows Dr. Victor Grimblatt presenting his talk about Digital Integrated Circuits Design Flow.

II. Conclusion

It was a pleasant experience to organize the IEEE CASS Tour in Peru, going to 4 locations in 3 cities. The participation of professionals and students was excellent and helped to increase CASS visibility to IEEE members who are not yet CASS members. We expect an increase in CASS membership in all visited locations and the start of new CASS Student Branch Chapters.

Other CASS Tours in Latin American Countries were organized in 2022 and 2023, and it is possible to find some reports about them in the IEEE CASS Magazine [1], [2], [3], [4], [5], [6], [7]. Other CASS Tours are being planned as, for instance, an IEEE CASS Tour in Chile, another in Colombia, and Fortaleza, Brazil.

Figure 10. Prof. Carlos Silva Cardenas is doing his talk at UNI.

Figure 11. Dr. Victor Grimblatt is doing his talk at PUCP.

Acknowledgment

We thank the local organizers, responsible for the logistic and organization of the CASS Tour Peru: Prof. Prof Filiberto Azabache in Trujillo, Prof Mario Urrutia in Arequipa, and Yahaida Apaza at UNI in Lima.

References

[1] R. Reis, V. Grimblatt, and M. Delgado-Restituto, "IEEE CASS tour Argentina 2022," *IEEE Circuits Syst. Mag.*, vol. 23, no. 1, pp. 87–89, 1st Quart., 2022, doi: 10.1109/MCAS.2023.3244430.

[2] F. Olivera and R. Reis, "IEEE CASS Rio de Janeiro tour 2022," *IEEE Circuits Syst. Mag.*, vol. 23, no. 2, pp. 64–65, 2nd quart., 2023, doi: 10.1109/MCAS.2023.3267939.

[3] R. Reis and R. Murphy, "IEEE CASS tour Mexico 2022," *IEEE Circuits Syst. Mag.*, vol. 23, no. 2, pp. 66–70, 2nd quart., 2023, doi: 10.1109/MCAS.2023.3267987.

[4] R. Reis and R. Taco, "IEEE CASS tour Ecuador 2023," *IEEE Circuits Syst. Mag.*, vol. 23, no. 3, pp. 72–75, 3rd quart., 2023, doi: 10.1109/MCAS.2023.3302394.

[5] R. Murphy and R. Reis, "CASS tour Mexico 2023," *IEEE Circuits Syst. Mag.*, vol. 23, no. 3, pp. 76–80, 3rd quart., 2023, doi: 10.1109/MCAS.2023.3302395.

[6] R. Reis and M. Delgado-Restituto, "IEEE CASS tour Puerto Rico 2023," *IEEE Circuits Syst. Mag.*, vol. 23, no. 3, pp. 80–83, 3rd quart., 2023, doi: 10.1109/MCAS.2023.3302396.

[7] R. Reis and M. Delgado-Restituto, "IEEE CASS tour Panama 2023 and the IEEE R9 meeting," *IEEE Circuits Syst. Mag.*, vol. 23, no. 3, pp. 83–86, 3rd quart., 2023, doi: 10.1109/MCAS.2023.3302397.

CASS Conference Highlights

Ricardo Reis *(CASS BoG R9 Representative)* and Calebe Conceição *(IFSul)*

IEEE CASS Tour Sergipe 2023

I. The IEEE CASS Tour Sergipe 2023

The IEEE CASS Tour Sergipe 2023 was held at the Federal University of Sergipe (UFS), at Aracaju, in 2 October. The UFS was launched in 1968. The Tour included a set of technical talks as well as a special talk about the IEEE Circuits and Systems Society, presenting CASS activities and opportunities. It was organized a panel about "Education on Computing and Microelectronics," with the participation of professors Edward Moreno (UFS Computer Science), Elyson Carvalho (UFS—Electrical and Electronics Engineering), Calebe Conceição (IFSul), and Ricardo Reis (UFRGS). Figure 1 shows the full program of the event. The panel was an opportunity for the colleagues of the Electrical Engineering and Computer Science from the UFS to discuss the challenges in education to cope with trends in computing and microelectronics.

The audience was composed by students and professors from Computer Science and Electrical Engineering of UFS (Federal University of Sergipe). Figure 2 was taken during the presentation of prof. Calebe Conceição. Figure 3 shows the speakers with the students of the IEEE Student Branch ExCom. It is a piece of evidence that continuity of actions should take place to keep motivating students and professor. CASS has a key role in this process.

The CASS Tour Sergipe is part of a set of CASS Tours being organized in R9. It is possible to know about some past tours by reading the reports published in past editions of the IEEE Circuits and Systems Magazine [1], [2], [3], [4], [5], [6], [7].

Digital Object Identifier 10.1109/MCAS.2024.3358282
Date of current version: 15 August 2024

Figure 1. Flyer of the IEEE CASS Tour at Aracaju.

Figure 2. Photo taken during the talk done by Prof. Calebe Conceição.

Figure 3. Photo with speakers and students of the IEEE Student Branch.

II. Conclusion

It was a great experience to organize the IEEE CASS Tour Sergipe 2023. The participation of professionals and students was great, and they have stated that it will be done a petition to launch a CASS Student Branch Chapter in the region. It is expected an increase in CASS membership in the region.

Acknowledgment

To all ones that helped in the organization of the CASS Tour Sergipe 2023, specially to Alysson Felipe Virgulino

da Silva, Chair of the IEEE Student Branch of the Federal University of Sergipe.

References

[1] R. Reis, V. Grimblatt, and M. Delgado-Restituto, "IEEE CASS tour Argentina 2022," *IEEE Circuits Syst. Mag.*, vol. 23, no. 1, pp. 87–89, 1st Quart., 2022, doi: 10.1109/MCAS.2023.3244430.

[2] F. Olivera and R. Reis, "IEEE CASS Rio de Janeiro tour 2022," *IEEE Circuits Syst. Mag.*, vol. 23, no. 2, pp. 64–65, 2nd Quart., 2023, doi: 10.1109/MCAS.2023.3267939.

[3] R. Reis and R. Murphy, "IEEE CASS tour Mexico 2022," *IEEE Circuits Syst. Mag.*, vol. 23, no. 2, pp. 66–70, 2023, doi: 10.1109/MCAS.2023.3267987.

[4] R. Reis and R. Taco, "IEEE CASS tour Ecuador 2023," *IEEE Circuits Syst. Mag.*, vol. 23, no. 3, pp. 72–75, 2023, doi: 10.1109/MCAS.2023.3302394.

[5] R. Murphy and R. Reis, "CASS tour Mexico 2023," *IEEE Circuits Syst. Mag.*, vol. 23, no. 3, pp. 76–80, 2023, doi: 10.1109/MCAS.2023.3302395.

[6] R. Reis and M. Delgado-Restituto, "IEEE CASS tour Puerto Rico 2023," *IEEE Circuits Syst. Mag.*, vol. 23, no. 3, pp. 80–83, 2023, doi: 10.1109/MCAS.2023.3302396.

[7] R. Reis and M. Delgado-Restituto, "IEEE CASS tour Panama 2023 and the IEEE R9 meeting," *IEEE Circuits Syst. Mag.*, vol. 23, no. 3, pp. 83–86, 3rd Quart., 2023, doi: 10.1109/MCAS.2023.3302397.

CASS Conference Highlights

Ricardo Reis *(CASS BoG R9 Representative)* and Ricardo Jacobi *(CASS Center-North Chapter Chair)*

IEEE CASS Tour Brasília 2023

I. The IEEE CASS Tour Brasilia 2023

The IEEE CASS Tour Brasilia 2023 was held at the Federal University of Brasilia (UNB), in 15 September. The Federal University of Brasilia is a young university, starting activities in 1962, just 2 years after the construction of Brasilia, a planned city, constructed to be the Capital of Brazil.

The Tour included a set of technical talks as well as a special talk about the IEEE Circuits and Systems Society, presenting CASS activities and opportunities.

It was organized a panel about "Education in Computing and Microelectronics," with the participation of **Ricardo Jacobi** (UNB), **Carlos Llanos** (UNB), **Sandro Haddad** (UNB), **Alexandre Romariz** (UNB), **Ney Calazans** (UFRGS), **Hamilton da Silva** (MCTI—Ministry of Science, technology and Innovation). and **Ricardo Reis** (UFRGS). Figure 1 shows the full program of the event. The panel was an opportunity for the colleagues of the Electrical Engineering and Computer Science from the UNB to discuss the challenges in education to cope with trends in circuits and systems. Figure 2 shows the audience during the presentation of Prof. Ney Calazans.

The number of participants was 69, with the participation of students and professors from 2 campus of the UNB. The CASS Tour was well received by the participants. It is a piece of evidence that continuity of actions should take place to keep motivating students and professors, and CASS has a key role in this process.

The CASS Tour Brasilia is part of a set of CASS Tours being organized in R9. It is possible to know about some past tours by reading the reports published in past

Digital Object Identifier 10.1109/MCAS.2024.3358288
Date of current version: 15 August 2024

Figure 1. Flyer of the IEEE CASS Tour at Brasilia.

Figure 2. The audience during the talk done by Prof. Ney Calazans.

editions of the IEEE Circuits and Systems Magazine [1], [2], [3], [4], [5], [6], [7].

II. Conclusion

It was a great experience to organize the IEEE CASS Tour Brasilia 2023. The participation of professionals and students was great, and they have stated that it will be done an effort to increase the engagement of students in the CASS Student Branch Chapter in the region. It is also expect an increase in CASS membership in the region.

Acknowledgment

To all ones that helped in the organization of the CASS Tour Brasilia 2023.

References

[1] R. Reis, V. Grimblatt, and M. Delgado-Restituto, "IEEE CASS tour Argentina 2022," *IEEE Circuits Syst. Mag.*, vol. 23, no. 1, pp. 87–89, 1st Quart., 2022, doi: 10.1109/MCAS.2023.3244430.

[2] F. Olivera and R. Reis, "IEEE CASS Rio de Janeiro tour 2022," *IEEE Circuits Syst. Mag.*, vol. 23, no. 2, pp. 64–65, 2nd Quart., 2023, doi: 10.1109/MCAS.2023.3267939.

[3] R. Reis and R. Murphy, "IEEE CASS tour Mexico 2022," *IEEE Circuits Syst. Mag.*, vol. 23, no. 2, pp. 66–70, 2023, doi: 10.1109/MCAS.2023.3267987.

[4] R. Reis and R. Taco, "IEEE CASS tour Ecuador 2023," *IEEE Circuits Syst. Mag.*, vol. 23, no. 3, pp. 72–75, 2023, doi: 10.1109/MCAS.2023.3302394.

[5] R. Murphy and R. Reis, "CASS tour Mexico 2023," *IEEE Circuits Syst. Mag.*, vol. 23, no. 3, pp. 76–80, 2023, doi: 10.1109/MCAS.2023.3302395.

[6] R. Reis and M. Delgado-Restituto, "IEEE CASS tour Puerto Rico 2023," *IEEE Circuits Syst. Mag.*, vol. 23, no. 3, pp. 80–83, 2023, doi: 10.1109/MCAS.2023.3302396.

[7] R. Reis and M. Delgado-Restituto, "IEEE CASS tour Panama 2023 and the IEEE R9 meeting," *IEEE Circuits Syst. Mag.*, vol. 23, no. 3, pp. 83–86, 3rd Quart., 2023, doi: 10.1109/MCAS.2023.3302397.

CASS Conference Highlights

Ricardo Reis *(CASS BoG Member)*, Manuel Delgado-Restituto *(CASS President)*, and Victor Grimblatt *(CASS BoG Member)*

IEEE CASS Tour Chile 2023

I. The IEEE CASS Tour Chile 2023

The IEEE CASS Tour Chile 2023 was held from 23 to 27 October, with a set of talks in three locations: Valparaíso, Punta Arenas and Santiago. The set of talks included technical ones, a special talk about the IEEE Circuits and Systems Society and opportunities, and why to become a CASS member. The talk related to CASS and opportunities present CASS activities such as conferences, workshops, symposiums, seasonal schools, publications, student design contests, the distinguished lecturer program, pre-doctoral grants, student travel grants, and educational talks promoted by CASS. The list of calls organized by CASS was also presented to support the organization of CASS activities.

Digital Object Identifier 10.1109/MCAS.2024.3358272
Date of current version: 15 August 2024

The main goal of the CASS Tours being organized in R9 is to motivate professionals and students to be active members of CASS, showing them a whole set of opportunities and support the society provides.

The CASS Tour in Valparaiso, was organized at the Universidad Técnica Federico Santa Maria. The University starts its activities in 1926 (https://usm.cl/en/home/). Valparaíso is situated is situated about 100 km west of Santiago. The full program of the CASS Tour in Valparaíso is presented in Figure 1. A photo with some participants of the CASS Tour in Valparaiso is presented in Figure 2.

On 25 October, the CASS Tour went to the Universidad Magallanes at Punta Arenas, the southern university in Chile. Punta Arenas is located about 3400 km south Santiago, by road, and about 3 hours by airplane. The University was founded in 1961 (more information at

Figure 1. Program of the IEEE CASS Tour in Valparaíso.

Figure 2. Group photo with some of the participants of the CASS Tour at Valparaíso.

Figure 3. Program of IEEE CASS Tour in Punta Arenas.

Figure 4. Group photo with some participants of the CASS Tour at Punta Arenas.

Figure 5. Program of IEEE CASS Tour in Santiago, at Synopsys.

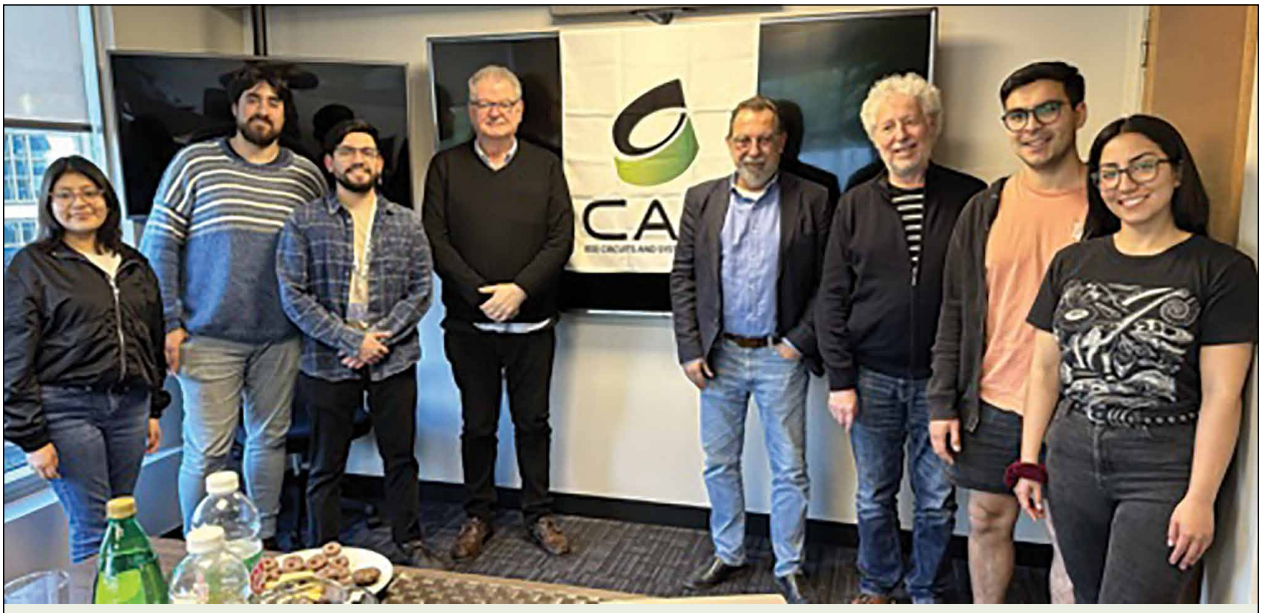


Figure 6. Meeting with the IEEE Student Chapter at the Synopys Offices in Santiago de Chile.

http://www.umag.cl/en/). Between the Departments of the Engineering Faculty, there is an Electrical Engineering Department and a Computer Engineering one. The Full Program in Punta Arenas is presented in Figure 3. In Figure 4, it is possible to see a group photo with some participants of the CASS Tour at Punta Arenas.

On 26 October, the CASS Tour occurred in the offices of Synopsys Chile (Program in shown in Figure 5). Figure 6 shows the participants of the meeting with the IEEE Student Chapter of the University of Santiago, where the CASS President presented the opportunities in being a CASS member. It is expected to have a new CASS Student Branch Chapter at the University of Santiago.

II. Conclusion

It was a pleasant experience to organize the IEEE CASS Tour in Chile, going to 3 locations in 3 cities. The participation of professionals and students was excellent and helped to increase CASS visibility to IEEE members who are not yet CASS members. We expect an increase in CASS membership in all visited locations and the start of new CASS Student Branch Chapters.

Other CASS Tours in Latin American Countries were organized in 2022 and 2023, and it is possible to find some reports about them in the IEEE CASS Magazine [1], [2], [3], [4], [5], [6], [7]. Other CASS Tours are being planned as, for instance, an IEEE CASS Tour in Colombia, Fortaleza (Brazil) and Belo Horizonte (Brazil) where is starting a new IEEE CASS Chapter Minas Gerais.

Acknowledgment

We thank the local organizers, responsible for the logistic and organization of the CASS Tour Chile: Prof. Ioannis Vourkas in Valparaíso and Profs. Iván Andrade y Bedrich Magas in Punta Arenas.

References

[1] R. Reis, V. Grimblatt, and M. Delgado-Restituto, "IEEE CASS tour Argentina 2022," IEEE Circuits Syst. Mag., vol. 23, no. 1, pp. 87–89, 1st Quart., 2022, doi: 10.1109/MCAS.2023.3244430.

[2] F. Olivera and R. Reis, "IEEE CASS Rio de Janeiro tour 2022," IEEE Circuits Syst. Mag., vol. 23, no. 2, pp. 64–65, 2nd Quart., 2023, doi: 10.1109/MCAS.2023.3267939.

[3] R. Reis and R. Murphy, "IEEE CASS tour Mexico 2022," IEEE Circuits Syst. Mag., vol. 23, no. 2, pp. 66–70, 2nd Quart., 2023, doi: 10.1109/MCAS.2023.3267987.

[4] R. Reis and R. Taco, "IEEE CASS tour Ecuador 2023," IEEE Circuits Syst. Mag., vol. 23, no. 3, pp. 72–75, 3rd Quart., 2023, doi: 10.1109/MCAS.2023.3302394.

[5] R. Murphy and R. Reis, "CASS tour Mexico 2023," IEEE Circuits Syst. Mag., vol. 23, no. 3, pp. 76–80, 3rd Quart., 2023, doi: 10.1109/MCAS.2023.3302395.

[6] R. Reis and M. Delgado-Restituto, "IEEE CASS tour Puerto Rico 2023," IEEE Circuits Syst. Mag., vol. 23, no. 3, pp. 80–83, 3rd Quart., 2023, doi: 10.1109/MCAS.2023.3302396.

[7] R. Reis and M. Delgado-Restituto, "IEEE CASS tour Panama 2023 and the IEEE R9 meeting," IEEE Circuits Syst. Mag., vol. 23, no. 3, pp. 83–86, 3rd Quart., 2023, doi: 10.1109/MCAS.2023.3302397.

CASS Conference Highlights

Ricardo Reis *(CASS BoG R9 Representative)* and Jarbas Silveira (*UFC*)

IEEE CASS Tour Ceará 2023

I. The IEEE CASS Tour Ceará 2023

The IEEE CASS Tour Ceará 2023 was held at an auditorium of the Federal University of Ceará (UFC) on 1 December. The Federal University of Ceará is a public university in Brazil that started activities in 1955. The SBCCI 2017 (30th Symposium on Integrated Circuits and Systems Design), which is co-sponsored by IEEE CASS, was organized in Fortaleza, State of Ceará.

The Tour included a set of technical talks and a special talk about the IEEE Circuits and Systems Society, presenting CASS activities and opportunities. Figure 1 shows the entire program, including talks from academia and industry. It was organized a panel about "Education in Computing and Microelectronics," with the participation of professors Heleno de Souza Castro (UFC), Lírida Naviner de Barros (Telecom Paris, France), João César Mota (UFC), Jarbas Silveira (UFC), and Ricardo Reis (UFRGS). It was discussed important aspects of microelectronics education in Brazil. The central discussion was about creating a critical number of students in microelectronics without having foundries in Brazil.

There was a nice number of participants, including students, professors, and professionals from the industry. The participants received the event well, and Figure 2 shows a group photo with some of the participants.

The program included a set of interesting technical lectures. In the first lecture, Prof. Jarbas Silveira talked about the Nascerr Project, which is a 2U CubeSat project whose mission will be the investigation of

Digital Object Identifier 10.1109/MCAS.2024.3376649
Date of current version: 15 August 2024

Figure 1. Flyer of the IEEE CASS Tour in Fortaleza.

Figure 2. A Group Photo with the audience of the CASS Tour at UFC.

dynamic error correction codes algorithmics to protect external static RAMs against errors by radiation. The program also included the real-time virtual presentation of the CASS Talk with Prof. Xinmiao Zhang (Ohio State University, USA) about Error-Correcting Codes and Cryptography.

The CASS Tour Ceará is part of a set of CASS Tours organized in R9. It is possible to know about some past CASS Tours organized in R9 during 2022 and 2023 by reading the reports published in the past IEEE Circuits and Systems Magazine editions.

II. Conclusion

It was a pleasant experience to organize the IEEE CASS Tour Ceará 2023. The participation of professionals and students was great, and they have stated that it will be done an effort to launch CASS Student Branch Chapter in the university. An increase in CASS membership in the region is also expected.

Acknowledgment

To all ones that helped in the organization of the CASS Tour Ceará 2023.

CASS Conference Highlights

Ricardo Reis *(CASS BoG R9 Representative)*, Victor Grimblatt *(CASS BoG)*, and Faruk Fonthal *(CASS Colombia Chapter Chair)*

IEEE CASS Tour Colombia 2023

I. The IEEE CASS Tour Colombia 2023

The IEEE CASS Tour Colombia 2023 was held from 21 to 23 November, in two locations: Cartagena and Barranquilla. The CASS Tour Colombia organized in 2022, was held in Bogotá and Barranquilla (location of ANDESCON 2022) [1].

A main goal of the set of CASS Tours being organized in R9 is to motivate professionals and students to be active members of CASS, showing them a whole set of opportunities and support provided by the society.

The Tour in Cartagena was done at Universidad Tecnológica de Bolivar, and was composed by two technical talks as well as a special talk about the IEEE Circuits and Systems Society, presenting CASS activities such as conferences, workshops, symposiums, seasonal schools, publications, student design contests, the distinguished lecturer program, pre-doctoral grants, student travel grants, and talks promoted by chapters. The Full Program of the CASS Tour in Cartagena is presented in Figure 1.

It was used the opportunity to visit the Cartagena Convention Center, that is a huge and very nice convention center that can host events bigger than ISCAS. The convention center is well located, just in front of the wonderful old city.

Digital Object Identifier 10.1109/MCAS.2024.3376639
Date of current version: 15 August 2024

II. The CASS Tour at IEEE C3 2023

From 22 to 25 November, it was organized the IEEE Colombian Caribbean Conference (C3) in Barranquilla, at the Campus of the Universidad del Norte. C3 2023 was

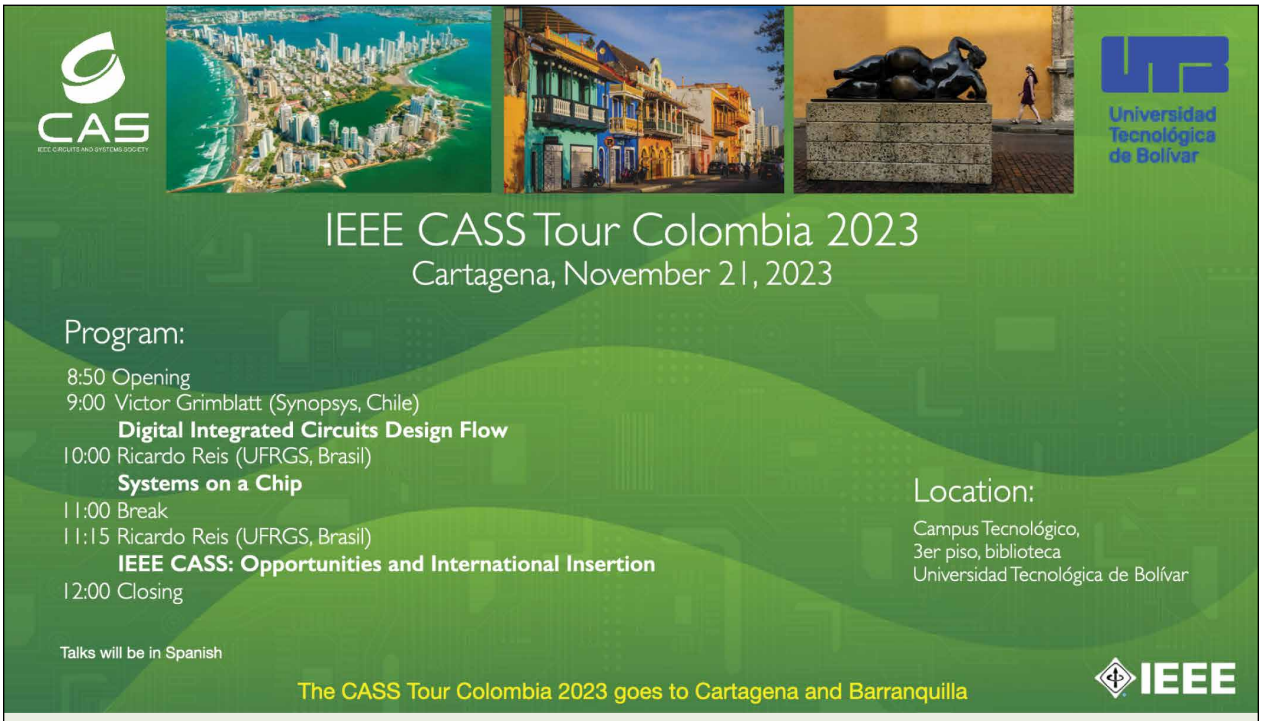


Figure 1. Program of the IEEE CASS Tour in Cartagena.

Figure 2. Photo taken during the CASS tutorial at IEEE C3.

Figure 3. Photo of IEEE C3 audience.

organized by the IEEE Colombian Caribbean Section and supported by the IEEE Andean Council, and Universidad del Norte. More details, as the C3 Full Program are available at the conference webpage https://attend.ieee.org/c3/.

IEEE CASS had a nice activity set embedded in C3 with 3 presentations. On Wednesday, 22 November, Ricardo Reis has done a tutorial with the title "Trends on Micro and Nanoelectronics" (Figure 2). A plenary Keynote talk with the title "How to feed a growing population while conserving the planet's resources—IoT to the Rescue," by Victor Grimblatt was done in 23 November and the talk about CASS and Opportunities was done in 23 November. Both talks were plenary sessions (Figure 3 shows a photo with the audience). The flyer with the full CASS Program at C3 is shown in Figure 4.

IEEE C3 will be organized each two years in a different city of the region covered by the IEEE Section Colombia Caribe. It is also a great opportunity to meet IEEE and CASS members from several countries from R9. IEEE C3 2025 will be organized in Santa Marta.

During the Tour Colombia, the relationship with the leaders of the Colombia Caribbean section of the IEEE and people from different universities in the region allowed us to motivate the creation of the IEEE CASS professional chapter for the IEEE Colombia Caribbean section. This enabled Colombia to have two professional chapters of the IEEE CAS Society.

Figure 4. Program of IEEE CASS Tour in Barranquilla, embedded in the IEEE Colombia Caribbean Conference.

Other CASS Tours in Latin American Countries were organized in 2022 and 2023, and it is possible find some reports about them in past IEEE CASS Magazine editions as the report related to the CASS Tour Colombia 2022 [1]. Other R9 CASS Tours are being planned for 2024.

III. Conclusion

It was a nice experience to organize the IEEE CASS Tour Colombia 2023, going to 2 locations in 2 cities in one week. The participation of professionals and students was great, and the embedded CASS talks at IEEE C3 helped to increase CASS visibility to IEEE members that are not yet CASS members. We expect an increase in CASS membership in all visited locations. As a consequence of the Tours in the Colombian Caribbean Section, it was formed a new CASS chapter, the IEEE CASS Colombian Caribbean Chapter.

Acknowledgment

To the local organizers of the CASS Tour Colombia 2024, with a special thanks to the organizers of the IEEE C3 headed by César Viloria Núñez, for the opportunity of CASS presented a tutorial and two plenary sessions.

References

[1] R. Reis and F. Faruk, "IEEE CASS Tour Colombia 2022," *IEEE Circuits Syst. Mag.*, vol. 23, no. 2, p. 71, 2nd Quart., 2023, doi: 10.1109/MCAS.2023.3277489.

CASS Conference Highlights

Ricardo Reis *(CASS BoG R9 Representative)* and Janier Arias (*CASS Minas Gerais Chapter Chair*)

IEEE CASS Tour Minas Gerais 2023

I. The IEEE CASS Tour Minas Gerais 2023

The IEEE CASS Tour Minas Gerais 2023 took place at the auditorium of the Engineering School of the Federal University of Minas Gerais (UFMG) on 4 December. The Federal University of Minas Gerais is one of the main public universities in Brazil, starting activities in 1927.

The Tour featured a series of technical talks and a special presentation on the IEEE Circuits and Systems Society, highlighting CASS activities and opportunities. Figure 1 displays the entire program, encompassing talks from academia and industry. The event garnered a significant number of participants, including students, professors, and industry professionals. Belo Horizonte is home to Cadence Brazil. This CASS Tour was unique as it marked the inaugural activity of the NEW CASS Chapter Minas Gerais. A celebration of the new Chapter concluded the event. Now, all IEEE Sections in Brazil boast an IEEE CASS Chapter. Participants warmly received the event, and Figure 2 captures a group photo with the attendees.

The CASS Tour Minas Gerais is part of a series of CASS Tours organized in Region 9 (R9). Information about past CASS Tours held in R9 during 2022 and 2023 can be found by reading reports published in previous editions of the IEEE Circuits and Systems Magazine.

Digital Object Identifier 10.1109/MCAS.2024.3376652
Date of current version: 15 August 2024

Figure 1. Flyer of the IEEE CASS Tour at Belo Horizonte.

Figure 2. A Group Photo with the audience of the CASS Tour at UFMG.

II. Conclusion

Organizing the IEEE CASS Tour Minas Gerais 2023 was a rewarding experience, especially as it served as the inaugural activity of the new Chapter. The enthusiastic participation of professionals and students was noteworthy, with many expressing their intent to make efforts to launch CASS Student Branch Chapters in the region. An increase in CASS membership in the area is also anticipated.

Acknowledgment

To all ones that helped in the organization of the CASS Tour Minas Gerais 2023.

CASS Conference Highlights

Fabián Olivera, *Member, IEEE*, and Ricardo Reis, *Life Senior Member, IEEE*

IEEE CASS Workshop Rio 2023 Current Trends in IC Design

The IEEE Circuits and Systems Society Workshop Rio (CASSW Rio), entitled Current Trends in IC Design, was organized as a co-event of the SBCCI (Symposium on Circuits and Systems Design, Chip in Rio [1], which is the most important microelectronic conference organized in Brazil since 1983 (in Rio, it was celebrated 40 years of the event) and co-sponsored by the CASS Society since 2004. The 36th edition of SBCCI took place in the city of Rio de Janeiro, at both, the Rio Art Museum and Museum of Tomorrow (see Figure 1), with the particular incentive that it was the first in-person edition after a long COVID-19 pandemic virtual time. The IEEE CASSW Rio, organized by the CASS Rio de Janeiro Chapter, was intended for academic and industrial sectors, allowing networking for both students and professionals.

The workshop was moderated and opened by an introductory talk by Professor Fabián Olivera (Federal Center for Technological Education of Rio de Janeiro), Chair of the IEEE CASS Rio de Janeiro Chapter. In the opening, Professor Fabián presented some details of the first call for the Universalization of IC Design from CASS (UNIC-CASS) [2], which had been closed on June 22.

The three invited speakers were distinguished professionals with significant work in the field of circuits and systems. Then, the IEEE CASSW Rio was composed of the following invited talks:

- **Fernando Silveira** (University of the Republic, Uruguay), from 14:20 PM to 15:00 PM, "Ratio Based Analog/RF Design: A Generalization of gm/ID and Inversion Coefficient Methods."
- **Ricardo Reis** (Federal University of Rio Grande do Sul, Brazil), from 15:05 PM to 15:45 PM, "Physical Design: New Solutions Inspired in the Past."
- **Victor Grimblatt** (R&D Group Director and General Manager, Synopsys, Chile), from 15:50 PM to 16:30 PM, "The Tangled Tree of Technology."
- **Ricardo Reis**, from 16:35 PM to 17:20 PM, "CASS News and Opportunities."

Digital Object Identifier 10.1109/MCAS.2024.3358286
Date of current version: 15 August 2024

The event was publicized on the IEEE CASS Rio Chapter's channels, such as the website [3], e-mail lists, and Instagram, using the publicity flyer in Figure 2. Also, a CASS banner and folders were placed on the SBCCI poster session, as shown in Figure 3. Since it was a co-event with Chip in Rio, their participants had free access to the workshop. However, the public of the workshop was not limited to only the conference participants, and 15 IEEE CASS Rio de Janeiro members participated in the workshop, with previews and free registration either by chapter website [3] or by IEEE vTool. Figure 4 shows the IEEE CASSW Rio speakers, in

Figure 1. SBCCI—Chip in Rio event venue: a) Rio Art Museum; b) Museum of Tomorrow.

Figure 2. IEEE CASSW Rio publicity flyer.

Figure 3. Publicity banner and folders of CASS at the poster session location of Chip in Rio conference.

Figure 4. IEEE CASSW Rio speakers: Fabián Olivera, Fernando Silveira, Ricardo Reis and Victor Grimblatt.

Figure 5. Workshop room at the first lecture given by Fernando. *Silveira*.

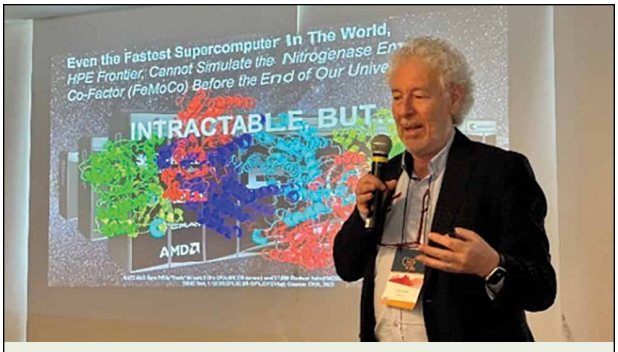


Figure 6. Lecture "The Tangled Tree of Technology" by Victor Grimblatt.

which from left to right are Fabián Olivera, Fernando Silveira, Ricardo Reis, and Victor Grimblatt.

In all lectures, there was a space for questions and answers. The IEEE CASSW Rio, Current Trends in IC Design, featured the participation of 13 students and 15 professionals, as can be appreciated in part from Figure 5 during the lecture of Professor Fernando Silveira. Figure 6 shows Victor Grimblatt at his lecture, "The Tangled Tree of Technology."

The next SBCCI (Chip on the Cliffs) will be at João Pessoa, State of Paraíba, from 26 to 30 August 2024, and it should be organized another CASS Workshop.

References

[1] Souza, P., "Call for papers: SBCCI 2023," *IEEE Circuits Syst. Mag.*, vol. 22, no. 4, pp. 79–79, 4th Quart., 2022, doi: 10.1109/MCAS.2022.3214853.
[2] ieee-cas.org. *Universalization of IC Design From CASS (UNIC-CASS)*. Accessed: Dec. 2, 2023. [Online]. Available: https://ieee-cas.org/universalization-ic-design-cass-unic-cass
[3] r9.ieee.org/rdj-cas. *IEEE CASS Workshop Rio: Current Trends in IC Design*. Accessed: Dec. 2, 2023. [Online]. Available: https://r9.ieee.org/rdj-cas/event/current-trends-in-ic-design/

CASS Conference Highlights

Ricardo Reis *(CASS BoG Member)*

IEEE CASS at SBRM 2023

I. The IEEE R9 SBRM 2023

IEEE R9 Student Branches Regional Meeting (SBRM) was organized from 29 to 31 October at Cartagena, Colômbia. The SBRM is a meeting point for students from the different regional IEEE student branches to diversify their knowledge in soft and hard skills. This goal has been achieved by creating the perfect space for attendees to learn about different topics and meet and interact with students and professionals from all over the region. This allows a constant exchange of ideas, primarily leading to developing activities for local or regional benefit.

Between the sponsors, there were 6 IEEE societies: Photonics, PES, Computer, ComSoc, Oceanic Engineering, and CASS. IEEE CASS was also a sponsor of the gala dinner, and being a sponsor, CASS had space for a presentation of the Society for all participants during the dinner.

The agenda of the meeting can be seen at https://r9.ieee.org/sbrm/agenda/, as well as a set of photos. Between the activities included in the program, on the first day, a session was composed of a set of "Inspirational talks" by speakers from the six sponsoring societies. In the program, the order of the talks was not announced, as the order was defined on the fly at the beginning of the session using an electronic roulette wheel. The first Society selected by the electronic wheel was CASS. In the Inspirational Talk, we focused on showing the students the importance of knowing how to design a chip, as more and more computing and electronic systems end in a chip (Figure 1). After the six talks, there was a space for a set of questions and answers the audience could address to any of the speakers.

In the afternoon of 29 October, there was a fair where the sponsoring societies had a booth. The CASS booth was well visited by the students, showing interest in opening new CASS Student Branch Chapters. Figure 2 shows a photo of the CASS booth with some students who visited it.

On 30 October, it was presented a set of talks by the sponsoring societies. In the talk related to IEEE CASS, it was presented the opportunities provided by CASS as well CASS activities such as conferences, workshops,

Digital Object Identifier 10.1109/MCAS.2024.3376644
Date of current version: 15 August 2024

Figure 1. Ricardo Reis during the "Inspirational Talk" presentation.

Figure 2. Photo with some visitors of the CASS Booth during the Societies Fair.

symposiums, seasonal schools, publications, student design contests, the distinguished lecturer program, pre-doctoral grants, student travel grants, and educational talks promoted by CASS. The list of calls organized by CASS to support the organization of CASS activities was also presented.

After the presentations of the societies, there was a panel with the speakers representing the sponsoring societies, where the panel members first had to answer some questions prepared by the organizers, followed by a space for questions from the audience.

On the evening of 30 October, the Gala Dinner, sponsored by CASS, was organized. In the center of each table, there was a small flag with the CASS logo. At the beginning of the dinner, as CASS was the sponsor, we had the opportunity to do a short presentation of CASS, including a video about CASS. At the end of the dinner, there was a band playing local music as well as some dancers showing some local dances.

The total attendance of the SBRM exceeded 200 people, including speakers, logistics personnel, and participants in general. Out of this number, 170 were student and professional attendees from 40 student branches, representing 13 countries belonging to IEEE Region 9 (Argentina, Brazil, Chile, Colombia, Costa Rica, Ecuador, El Salvador, Mexico, Panama, Peru, Puerto Rico, Trinidad and Tobago, Uruguay). Figure 3 shows a group photo taken on the evening of the second day.

It is important to observe that there was an important space in the program for presentations and activities related to the IEEE sponsoring societies, allowing IEEE Student Branches representatives and other students to know more about the opportunities provided by each Society. A form was provided to evaluate the event by the attendees. One question was: "What do you think of the motivational spaces by the societies on the first day?" and 72.2% of the answers gave the maximum positive evaluation. Another question was: "What do you think of the motivational spaces by the societies on the first day?" with 77.8% of maximum positive evaluation. So, the students liked the participation of the sponsoring societies a lot.

II. Conclusion

It was a great experience to participate in the R9 Student Branch Regional Meeting in Cartagena, Colombia. A CASS Tour Colombia 2023 (20–23 November) was also organized at Cartagena and Barranquilla. The CASS participation in the SBRM, as well as the organization of the CASS Tours in Colombia in 2022 [1] and 2023, were important in the process of launching a new CASS Chapter in the IEEE Section Colombia Caribe. IEEE has formally approved the new CASS Chapter Colombia Caribe.

The SBRM is an excellent opportunity to meet Student Branch Chapter Chairs and to promote CASS to students of R9. The location of the next SBRM will be decided soon between El Salvador and Chile. It will be great to have CASS as a sponsor of the next editions.

Figure 3. Group photo with participants of the R9 SBRM.

Acknowledgment

We thank the local organizers responsible for the logistics and organization of the R9 SBRM.

References

[1] R. Reis and F. Faruk, "IEEE CASS Tour Colombia 2022," *IEEE Circuits Syst. Mag.*, vol. 23, no. 2, p. 71, 2nd Quart., 2023, doi: 10.1109/MCAS.2023.3277489.