Enhancement in IoT through Custom Instruction Set Architectures and TinyML: Review

Sandhya P

Department of Computer Science and Engineering National Institute of Technology Calicut Calicut, India sandhya.july1995@gmail.com

Priya Chandran

Department of Computer Science and Engineering
National Institute of Technology Calicut
Calicut, India
priya@nitc.ac.in

Abstract—The Internet of Things or IoT means to the confluence of the network of interconnected devices, the technologies that enable communication among these devices, and between these devices and the cloud. IoT contains embedded CPUs for performing a variety of functions. The processors for enhancing IoT applications form an active area of research in the domain of Computer Architecture. Embedded processors are resource and power-constrained challenging the execution of deep learning and machine learning algorithms. One proposed solution in literature is TinyML (Tiny Machine Learning) to overcome this limitation. TinyML architecture for IoT aims to deliver low latency, and effective bandwidth usage, reinforce data security, enhance privacy, and save costs. TinyML enables IoT devices to perform ML computations to be performed in the edge and avoid frequent access to the cloud which increases the performance. This paper presents a review of processor designs that are designed for TinyML architecture. For addressing other important issues in IoT a review of custom ISAs (Instruction Set Architectures) developed is also presented in this review.

Index Terms—TinyML, IOT, RISC-V, NN, ISA, ISE

I. INTRODUCTION

IoT is defined as a network of actual real items that are equipped with sensors, software, and connectivity. These objects collect and exchange data online, allowing for the automation and optimization of several operations. The term "Internet of Things" was first introduced by Kevin Ashton in the late 1990s, and around the 2000s, scientists and engineers started creating the first IoT prototypes. A new era of connectivity began in 2008 when the number of internet-connected gadgets for the first time surpassed the number of people on the planet. Hence maintenance and security of IoT have become a major challenge.

IoT cannot accommodate massive ML algorithms [1], which are in demand now, due to resource and power constraints. Hence, TinyML, which combines ML and IoT, was proposed by [1] and many other researchers in the literature. IoT devices generally have embedded processors for performing specific tasks. An embedded system (ES), is an electronic-mechanical device intended to perform a certain task. An embedded system (ES) is a device that combines electronic and mechanical components to perform a particular software-based function. An ES is developed to perform particular tasks and can either be based on a microprocessor or a microcontroller, depending on the requirements. [2].

Embedded systems are crucial to Internet of Things (IoT) systems because of their special qualities and capacities, such as low power consumption, reactive computing, and low general and operational expenses. The main distinction between embedded systems and IoT is that, in general, once embedded systems software has been created, it is never modified. On the other hand, updates to the IoT software are made often.

The Internet of Things (IoT) platform uses the network to link various systems or devices. IoT systems have a number of vulnerabilities since it combines technologies including embedded systems, real-time computing, actuation, and wireless sensor networks (WSNs). Because of these weaknesses, attackers can easily access IoT systems [3]. IoTs require infrastructure and processor designs to address issues like physical, network, software, and encryption attacks and other major issues like the efficiency of IoT devices.

It is crucial to protect Internet of Things systems and the whole networks to which these devices are connected against attacks and intrusions. Security can be ensured at the hardware level, software level, and network level and it is crucial to improve the performance of IoT systems along with security. This review intends to analyze the literature on how IoT enhancement can be achieved through TinyML and Custom ISAs.

The review is structured as described, Section 2 explains TinyML and its usage in IoT and the processors designed to use TinyML. Section 3 describes custom ISAs in IoT, NN (Neural Networks), and its various advantages. Section 4 concludes the paper by saying that Custom ISAs and TinyML can bring improvements in IoT security and performance.

II. TINYML

Tiny machine learning (TinyML), a new idea for the IoT, argues that it integrates deep learning and ML techniques into the IoT device (Fig 1 shows the integration). Embedded Linux enables the integration of Linux into small and resource-constrained devices, providing flexibility, customization, and access to a wide range of software libraries and tools. Edge computing is a computing model that involves performing data processing and analysis in proximity to where the data is generated, instead of accessing the cloud, enabling faster

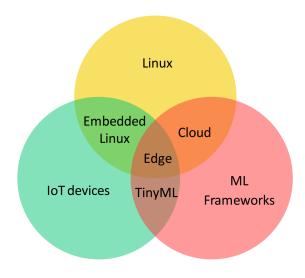


Fig. 1. IoT, ML and TinyML [4]

processing, reduced latency, and real-time decision-making capabilities. The objective of the TinyML framework developed for IoT is to ensure efficient utilization of bandwidth, minimize latency, enhance privacy, ensure data security, and reduce expenses [5]. With TinyML, IoT devices can offer precise machine learning services while operating reliably without the need for continuous cloud service access [6]. In IoT devices or networks, the MicroController Unit (MCU) that communicates with sensors is commonly used to transmit the data obtained to the cloud for further analysis and storage. Since data is created at the network's edge, edge data processing is thought to be more effective than cloud data processing. Frequent cloud access elevates privacy risks and has a major effect on the independence of the edge device. One major obstacle that prevents the integration of machine learning algorithms into MCUs is the challenge to overcome the disparity between the extensive memory requirements of machine learning algorithms and the constrained memory capabilities of microcontroller units [5]. With the increasing quantity of data generated by IoT systems, it is essential to identify and allocate resources to IoT solutions that can alleviate the difficulty of excessive data access and exchange to the cloud. TinyML integration on IoT devices is a remedy to get over this problem [5].

TinyML has many frameworks developed for incorporating ML in IoT. Some examples are TFLM, STM Cube AI, ELL, ARM-NN AI, MicroMLGen, etc [7]. TinyML has various applications for machine learning models, and speech recognition is a commonly used one [1]. TinySpeech is a framework that enables on-device speech recognition with low accuracy by using attention condensers to optimize CPU usage and memory footprint. AttendNets is an on-device image recognition framework that is designed to operate with low precision. Another practical use of TinyML is keyword spotting (KWS), which can be easily trained and deployed for use with different accents and speakers [8]. In the IoT-based arena, voice activity detection (VAD) is growing in popularity. The efficiency of IoT devices can be tremendously increased by incorporating

TinyML. The development of TinyML benchmarks is still in its early stages. However, the TinyMLPerf working group is currently working on creating a set of benchmarking codes specifically designed for deep neural networks that are aware of TinyML [9].

A. Processors designed for TinyML

The Cortex M-55 is the newest member of Arm's well-known Cortex-M series for microcontrollers and is intended to be the most AI-capable Cortex-M core Arm has ever produced [7]. The M55 is the first Arm processor to incorporate the new Helium vector processing technology, which offers performance gains of 15x for machine learning and 5x for DSP (Digital Signal Processing) when compared to earlier Cortex-M versions [10]. The existing Ethos NPU (neural processing unit) family has a micro-version called the Ethos-U55 machine learning accelerator. The Ethos-U55 and Ethos-U65, a microNPU particularly made to speed up ML inference in embedded and IoT devices, are two new machine learning (ML) processors from Arm. Speech recognition, gesture recognition, and biometric applications may all be run on the Cortex-M55 and Ethos-U55 combination [11].

- 1) Hardware Examples for TinyML [12]:
- General purpose cores with low-power and system-on-achip(SoCs) with low-power
 - Arm Cortex-M55, Greenwaves GAP8.
 - Wide (vector) registers is one feature that supports ML workloads.
 - Another peculiarity is the Dot-product operation and matrix multiply operation.
- Micro-Neural Processing Unit with low power
 - Samsung Edge-NPU, Arm Ethos-U55.
 - These processors will provide assistance for lower precision (<= 4b) functions, sparsity/zero-skipping, and Neural network functions through hardware.
- · Analog and Compute-in-Memory
 - Mythic

For deploying Machine Learning methods on edge devices the recommended hardware is the Arduino Nano 33 BLE Sense [7]. This device boasts a 64MHz, 32-bit, ARM Cortex-M4F microcontroller with 1MB of program memory and 256KB of RAM, which provides sufficient power for running TinyML models. In addition, the device comes equipped with various sensors for measuring color, brightness, proximity, gesture, motion, vibration, orientation, temperature, humidity, and pressure, as well as a Bluetooth low energy (BLE) module and a digital microphone. For most applications, this set of sensors is more than enough. Embedded Learning Library (ELL), an open-source library, has been released by Microsoft, demonstrating the company's support for the TinyML community [7], which enables the creation and implementation of pre-trained machine learning models on limited platforms such as the ARM Cortex-A and Cortex-M-based architectures found on devices like the Arduino, Raspberry Pi, and micro bit.

III. CUSTOM PROCESSORS FOR IOT

This section looks into literature where processors are enhanced with custom instructions or extensions for improving IoT devices in terms of security, efficiency, and incorporation of deep learning and ML algorithms. Table 1 summarizes RISC-V (Reduced Instruction Set Computer-five) Custom ISAs and their advantages in different areas mainly in IoT.

A. Custom processors for IoT Security

IoT security can be added through custom processor ISAs, which offer a special set of features and instructions catered to the needs of IoT applications. Custom processor ISAs can considerably minimize the attack surface of IoT devices, increasing their resistance to cyberattacks by lowering the complexity and size of the instruction set. Moreover, hardware security features like encryption and authentication can be incorporated into bespoke processor ISAs, which can increase the security of IoT devices. Let's look into some of the custom ISAs and ISEs in literature for enhancing security.

For data integrity and information security, Internet of Things (IoT) devices are increasingly using blockchain technology. However, given their limited resources, IoT devices face significant challenges in implementing complex hash algorithms due to their high energy consumption and longer processing time. A method was proposed to integrate blockchain technology into IoT applications by utilizing the RISC-V processor with memristor-based in-memory computing (IMC) [13]. The extensibility of the RISC-V instruction set architecture allowed for the creation of IMC-adapted instructions for the Keccak hash algorithm (ISA) [13]. To integrate a memristor-based IMC with a RISC-V processor in an areaefficient manner, the researchers utilized the Hummingbird E200 open-source core developed for IoT applications. The resulting method achieved significant improvements in performance and energy consumption with only minimal overhead.

ChaCha is a stream cipher designed for high performance and strong security on software platforms [14]. RISC-V is a new technology that includes multiple instructions set expansions (ISEs) to enable a compact RISC-V ISA to handle various applications, including cryptographic acceleration. RISC-V architectures, in order to uphold ChaCha, a minimal ISE is recommended, especially for embedded systems like IoT edge systems that lack a vector engine. The different versions of ISE (Instruction Set Extension) noticeably enhance the performance of the ChaCha block function by reducing the time it takes to execute and the amount of memory required, thus making it a better fit for devices with limited resources.

An additional research study put forth a hardware-based architecture for Dynamic Information Flow Tracking (DIFT) that is designed to be used with cores of RISC-V processor [15]. The researchers implemented and designed this architecture using PULPino, which is an open-source platform that facilitates the creation of RISC-V cores specifically intended for the Internet of Things applications. The DIFT architecture is practical, with no runtime overhead and minimal resource

usage, making it an efficient way to secure the core of RISC-V. The architecture is capable of detecting and preventing various memory corruption attacks, including format string and buffer overflow attacks that are commonly targeted by security threats.

IoT devices now use virtualization technology since security and performance are major concerns. The hypervisor, a new software layer introduced by virtualization, can, however, amplify the hazards already inherent in IoT devices. Architectural Support for Memory Isolation, or ASMI, is a novel architecture [16] that has been proposed to enhance virtualization in IoT devices in terms of security and performance. The performance and security of virtualization technologies can be enhanced by adapting the ASMI architecture to well-known virtualization platforms like MIPS. The greater memory isolation offered by this novel design increases security and lowers the danger of memory-based attacks while providing higher performance than conventional systems.

The SKIVA [17] processor is a specialized processor designed to aid in the creation of defenses against implementation threats. It includes custom instruction-set extensions that enable security-sensitive and recurring operations, such as fault detection, redundant logic calculation, secret-share generation, and bit slicing. This processor provides a flexible and efficient way for researchers and developers to experiment with different security measures and explore new ways to enhance the security of hardware systems. The custom instruction-set extensions in the SKIVA processor offer an effective means of implementing security measures in hardware systems.

There are numerous examples of custom processor ISAs and ISEs that can be utilized to enhance the security of IoT devices. Custom instruction-set extensions can provide specialized operations and improved performance for security-sensitive tasks. Overall, these specialized processor designs offer a promising approach to ensuring the security of IoT devices.

B. Custom processors for IoT Neural Networks

Custom processors are becoming increasingly popular for IoT neural networks, as they can be tailored to meet the unique requirements of neural network operations and accelerate performance and security. Let's look into some of the custom ISAs and ISEs in the literature for IoT Neural Networks.

IoT technologies commonly employ Convolutional Neural Networks (CNNs) for various applications. A reconfigurable coprocessor [18] based on the RISC-V ISA has been created to improve the processing efficiency of IoT CPUs. The developed coprocessor is connected with RISC-V and it becomes a separate entity that operates in parallel with the main CPU core and provides a flexible and scalable way to extend the functionality of CPU cores, allowing for optimized performance for specific applications or workloads. By comparing the running cycles of the convolution, pooling, ReLU, and matrix addition algorithms on the coprocessor's unique instruction set to the conventional RISC-V instruction set, the coprocessor's performance is assessed. The convolution algorithm accelerates

much more with the custom coprocessor instruction set than with the normal instruction set, according to the results.

Similarly, based on RISC-V architecture, a new CNN processor is created [19]. The processor is more versatile and can benefit from the parallelism of CNN in addition to having advantages from specially created instructions. Convolution operation instructions, vector store instructions, vector load instructions, and vector addition instructions are all made to hasten the convolution process execution in CNN. It can run customized instructions in addition to completing common instructions.

A specialized programmable processor has been created and put into action, which has a distinct instruction set architecture specifically designed to efficiently implement Artificial Neural Networks (ANNs) for embedded applications that have limitations in terms of space and power [20]. This processor is capable of handling ANNs of moderate size. The ANN processor's design permits the usage of an unlimited amount of layers and artificial neurons, and it can accommodate both feed-forward and dynamic recurrent networks, along with the flexibility to create ANNs with link architectures tailored to individual needs.

Cambricon [21] is an innovative Instruction Set Architecture (ISA) that has been proposed for machine learning applications. It provides the flexibility to handle various ML techniques, including both traditional ML techniques and NN approaches. When compared to x86 and MIPS on 10 different NNs techniques, Cambricon demonstrated significantly higher code density. While the current NN and ML accelerators, like DaDianNao and PuDianNao, can only handle a restricted range of ML methods (3 and 8, respectively), the Cambricon prototype accelerator can support all 16 ML approaches. Although the prototype accelerator's performance and energy efficiency have only been tested with partial benchmarks, it achieves similar results to the state-of-the-art accelerators, without any significant overheads, in contrast to the cutting-edge accelerator.

C. Custom processors for increasing efficiency of IoT

In order to handle the enormous amounts of data created by IoT devices, there is an increasing demand for processors that are more efficient and specialized. For IoT applications, custom processors can be created and optimized, resulting in higher energy efficiency, lower power requirements, and quicker processing rates. This section elaborates more on the custom processors for enhancing the efficiency of IoT applications.

Instruction-set extension based on RISC-V ISA that has focused on accelerating complex arithmetic used in physical-layer protocols of IoT communication schemes as well as ultra-low power (ULP) software-defined wireless IoT transceivers has been described in the study by Amor et al. [22].

The AI-PiM (Artificial Intelligence-Processing-in-Memory) architecture is presented in [23] as a method of tightly integrating PiM accelerators into the RISC-V processor, allowing

Area Processor Details Advantages Blockchain technology in IoT [13] Security in IoT RISC-V ISE for ChaCha (Stream Cipher) Ultra-low power IoT transceivers [22] AI-PiM [23] RISC-V and custom ISE Accelerated non-AI workloads Reconfigurable RISC Custom ISE Accelerate
technology in IoT [13] Security in IoT [14] ChaCha(stream Cipher) Ultra-low power IoT transceivers [22] AI-PIM [23] RISC-V and custom ISE RISC-V and custom of the speedup on a sine dege processor for AI anon-AI workloads
IoT [13] Security in IoT ChaCha(stream Cipher) Ultra-low power IoT transceivers [22] AI-PiM [23] RISC-V and custom ISE RISC-V and custom ISE RISC-V and custom offers speedup on a sin edge processor for AI a non-AI workloads
Security in IoT
[14] ChaCha(stream Cipher) ChaCha Ultra-low power IoT transceivers [22] AI-PiM [23] RISC-V and custom ISE Offers speedup on a sin edge processor for AI a non-AI workloads
Cipher) Ultra-low power IoT transceivers [22] AI-PiM [23] RISC-V and custom ISE RISC-V and custom edge processor for AI a non-AI workloads
Ultra-low power IoT transceivers [22] AI-PiM [23] RISC-V ISE Accelerated comp arithmetic operations Offers speedup on a sin edge processor for AI a non-AI workloads
IoT transceivers [22] AI-PiM [23] RISC-V and custom ISE Offers speedup on a sin edge processor for AI a non-AI workloads
[22] AI-PiM [23] RISC-V and custom ISE Offers speedup on a sin edge processor for AI a non-AI workloads
AI-PiM [23] RISC-V and custom ISE Offers speedup on a sin edge processor for AI a non-AI workloads
ISE edge processor for AI a non-AI workloads
non-AI workloads
Reconfigurable RISC Custom ISE Accelerate
CNN-accelerated performance by a fac
Co-processor for of 6.27
IoT CPU [18]
Information Crypto-extension Good Flexibility, high
Security and with RISC-V ergy efficiency
Energy efficiency
in IoT [24]
[25] RISC-V processor Instruction PRESE
and PRINCE, go
encryption performan
low cost, high efficience
Security in IoT DIFT architecture for Security and efficiency [26] RISC-V
CNN [19] RISC-V custom ISA Accelerates convolut
process in CNN
Approximate RISC-V ISE Enhancing ML appli
computing in IoT tions
[27]
AES hardware Custom RISC-V ISA Lowered power consum
accelerator in tion, encryption time a
IOT [28] code size
NLP [29] RISC-V based known Improved code dens
as RISC-VTF and performan
as Rise-VII and performan
ANN in IoT [20] RISC-V based cus- An infinite number of l.
tom ISA ers and artificial neuro
can be supported
ISA for ML [21] Cambricon This prototype accelera
is able to support 16 M
techniques
IoT virtualization MIPS with ASMI performance and secur
22 The state of th
and security [16] improved
and security [16] improved IoT [30] RISC-V scalar micro- Encourages RISC-V
and security [16] improved IoT [30] RISC-V scalar micro- out-of-order proces- IoT improved Encourages RISC-V
and security [16] improved IoT [30] RISC-V scalar micro- out-of-order proces- sor:AnnikaCore Interval
and security [16] improved IoT [30] RISC-V scalar micro- out-of-order proces- IoT improved Encourages RISC-V

them to function as efficient execution units. This architecture makes use of a co-design approach for hardware, ISA, and software. AI-PiM offers a software framework that can seamlessly integrate PiM accelerators into the software stack. Additionally, AI-PiM also incorporates PiM functional units into the RISC-V processor hardware and extends the RISC-V ISA with custom instruction extensions. By doing so, AI-PiM enables the processing of both AI and non-AI workloads on the same processor without any issues, resulting in a significant performance boost. Specifically, when running the MLPerf Tiny benchmark, The AI-PiM provides a significant boost in speed compared to the RV64IMC RISC-V processor and Arm

Cortex-A72 processor, with an average speedup of 2.74x and 2.45x, respectively.

Information security is ever-increasing regarding linked devices is a problem that is unavoidably occurring and is garnering greater attention. Yet, it is challenging to deploy cryptographic techniques on limited power-efficient IoT systems which are battery-powered. Wang et al. proposed an energy-efficient crypto-coprocessor [24] that enables cryptography methodologies including AES, ECC, and SHA. The design of this technology incorporates cryptographic primitives and utilizes 128-bit or 256-bit data pathways through a unified pipelined architecture. By combining the RISC-V core with a crypto-extension, the resulting architecture achieves both high energy efficiency and flexibility.

ReonV, a customized version of LEON3, is a RISC V processor designed specifically for IoT systems and equipped with robust, efficient security features from the outset. The ReonV processor core's standard instruction set architecture includes the suggested cryptographic instructions (PRESENT and PRINCE) [31]. They have shown that the suggested approaches not only have excellent encryption performance, great efficiency, and low cost but are also sufficiently secure to survive the majority of threats.

The concept of approximate computing can effectively reduce power consumption by allowing for imprecise computations to be performed [27]. This approach takes into account the acceptable level of inaccuracy in the computation and aims to optimize energy efficiency while achieving the desired outcome. It is anticipated that the upcoming iteration of IoT devices designed for learning is expected to have approximation computing features in their processors. Based on these characteristics, an approximation of an IoT processor is developed that takes advantage based on the RISC-V ISA and is aimed at ML techniques like classification and clustering [27]. Implementing approximation operations in a processor can result in up to 23% reduction in power consumption for ASIC, while still achieving a minimum accuracy of 90% on both the trained models and test data set [27].

Another effort for limiting power usage is the introduction of an AES hardware accelerator in the literature with the aim of reducing the encryption time, and code size, besides lowering the power consumption of an IOT circuit. In order to implement the AES accelerator on an IBEX RISC-V core [28], RISC-V ISA is enhanced. The hardware implementation of AES and two software AES algorithms, TinyAES and OpenSSL AES, have been contrasted to show the advantages of adding a new instruction into the RISC-V ISA. The energy consumption of the AES accelerator is 662 and 44.9 times lower than that of TinyAES and OpenSSL AES, respectively [28]. The effectiveness of this approach highlights its value and promotes its adoption in IoT applications. The demonstrated success of this design emphasizes the benefits of using approximate computing in the context of IoT and encourages further exploration and implementation of similar techniques in other related areas.

Natural Language Processing (NLP) professionals fre-

quently employ Transformers, a deep learning model. Computational power is becoming more and more important in the field of NLP, especially when using the Transformer deep learning model. Yet, despite their abundant hardware resources, conventional general-purpose processors like CPUs and GPUs have shown to be ineffective for this task. In order to overcome these difficulties and better meet the particular needs of NLP and other deep learning applications, new hardware architectures, and specialized processing units have been created [29]. Custom ISAs prove to be effective for Natural Language Processing (NLP) as well. For the Transformer model, a hardware-friendly ISA has been designed based on RISC-V. Other than basic instruction, the activation instruction, softmax instruction, matrix load/store instruction, and other user-defined instructions are defined for the intensive and general computing portion of the model in accordance with the rules of RISC-V ISA. RISC-VTF offers improved code density and performance efficiency compared to the conventional common ISA such as x86, arm, or MIPs [29].

RISC-V scalar micro-out-of-order processor named Annika-Core [30] with a 3-stage pipeline is suggested in order to encourage the use of RISC-V processors in the IoT space. The results of the simulation and verification of AnnikaCore demonstrate that the planned processor complies with the RISC-V architecture specification and is functionally correct.

IV. CONCLUSION

IoT security is still a topic that needs more exploration. The literature review focuses on custom ISAs based on RISC-V, which have demonstrated the potential to uplift efficiency and IoT device security. Several unique instructions set architectures have been proposed to enhance the security of IoT systems. Custom modifications to the RISC-V architecture have been shown to have a significant impact on the IoT domain. TinyML, combined with custom ISAs and CPUs, has the potential to further advance IoT capabilities. TinyML-enabled processors with custom instructions allow for the efficient processing of data on edge devices. As a future scope, continuous exploration and investigation of the advantages offered by TinyML and custom ISAs, IoT devices can continue to outperform in the present as well as in the coming era.

ACKNOWLEDGMENT

This work is supported and sponsored by the Science and Engineering Research Board (SERB), Government of India through the Core Research Grant (CRG) Scheme.

REFERENCES

- [1] P. P. Ray, "A review on tinyml: State-of-the-art and prospects," pp. 1595–1623, 4 2022.
- [2] S. Babar, A. Stango, N. Prasad, J. Sen, and R. Prasad, "Proposed embedded security framework for internet of things (iot)," 2011.
- [3] M. Kasim, "Security of the embedded and iot systems: Threats, attacks, and countermeasures," 2021. [Online]. Available: www.ijcrt.org
- [4] H. Doyu, R. Morabito, and M. Brachmann, "A TinyMLaaS Ecosystem for Machine Learning in IoT: Overview and Research Challenges," 2021 International Symposium on VLSI Design, Automation and Test, VLSI-DAT 2021 - Proceedings, no. April, 2021.

- [5] D. L. Dutta and S. Bharali, "Tinyml meets iot: A comprehensive survey," Internet of Things, vol. 16, p. 100461, 12 2021.
- [6] N. N. Alajlan and D. M. Ibrahim, "Tinyml: Enabling of inference deep learning models on ultra-low-power iot edge devices for ai applications," *Micromachines*, vol. 13, 6 2022.
- [7] A. Osman, U. Abid, L. Gemma, M. Perotto, and D. Brunelli, "Tinyml platforms benchmarking."
- [8] V. J. Reddi, "'tinymlperf: Deep learning benchmarks for embedded devices"," 2021.
- [9] C. Banbury, V. J. Reddi, P. Torelli, J. Holleman, N. Jeffries, C. Kiraly, P. Montino, D. Kanter, S. Ahmed, D. Pau, U. Thakker, A. Torrini, P. Warden, J. Cordaro, G. Di Guglielmo, J. Duarte, S. Gibellini, V. Parekh, H. Tran, N. Tran, N. Wenxu, and X. Xuesong, "MLPerf Tiny Benchmark," 2021. [Online]. Available: http://arxiv.org/abs/2106.07597
- [10] "Arm Cortex-M55 Processor Most AI-Capable Cortex-M Processor."
- [11] "Arm Ethos-U55 microNPU Embedded ML Inference for Cortex-M systems."
- [12] "TinyML," https://cms.tinyml.org/wp-content /uploads/asia2020/tinyMLAsia2020d1p2-Mattina.pdf, 2020.
- [13] X. Xue, C. Wang, W. Liu, H. Lv, M. Wang, and X. Zeng, "A risc-v processor with area-efficient memristor-based in-memory computing for hash algorithm in blockchain applications," *Micromachines*, vol. 10, 8 2019.
- [14] B. Marshall, D. Page, and T. H. Pham, "A lightweight ise for chacha on risc-v," vol. 2021-text. Institute of Electrical and Electronics Engineers Inc., 7 2021, pp. 25–32.
- [15] C. Palmiero, G. Di Guglielmo, L. Lavagno, and L. P. Carloni, "Design and Implementation of a Dynamic Information Flow Tracking Architecture to Secure a RISC-V Core for IoT Applications," 2018 IEEE High Performance Extreme Computing Conference, HPEC 2018, nov 2018.
- [16] R. Jithin and P. Chandran, "Secure and dynamic memory management architecture for virtualization technologies in iot devices," *Future Inter*net, 2018.
- [17] P. Kiaei, D. Mercadier, P. E. Dagand, K. Heydemann, and P. Schaumont, "Custom Instruction Support for Modular Defense Against Side-Channel and Fault Attacks," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics*), vol. 12244 LNCS, pp. 221–253, 2021.
- [18] N. Wu, T. Jiang, L. Zhang, F. Zhou, and F. Ge, "A reconfigurable convolutional neural network-accelerated coprocessor based on risc-v instruction set," pp. 1–19, 6 2020.
- [19] Z. Li, W. Hu, and S. Chen, "Design and implementation of cnn custom processor based on risc-v architecture." Institute of Electrical and Electronics Engineers Inc., 8 2019, pp. 1945–1950.
- [20] D. Valencia, S. F. Fard, and A. Alimohammad, "An artificial neural network processor with a custom instruction set architecture for embedded applications." [Online]. Available: https://github.com/dlvalencia/ANN-Processor.
- [21] Y. Chen, H. Lan, Z. Du, S. Liu, J. Tao, D. Han, T. Luo, Q. Guo, L. Li, Y. Xie, and T. Chen, "An instruction set architecture for machine learning," ACM Transactions on Computer Systems, vol. 36, 8 2019.
- [22] H. B. Amor, C. Bernier, and Z. Prikryl, "A risc-v isa extension for ultra-low power iot wireless signal processing," *IEEE Transactions on Computers*, vol. 71, pp. 766–778, 4 2022.
- [23] V. Verma and M. R. Stan, "Ai-pim—extending the risc-v processor with processing-in-memory functional units for ai inference at the edge of iot," *Frontiers in Electronics*, vol. 3, 8 2022.
- [24] W. Wang, J. Han, X. Cheng, and X. Zeng, "An energy-efficient cryptoextension design for risc-v," *Microelectronics Journal*, vol. 115, 9 2021.
- [25] W. E. H. Youssef, A. Abdelli, F. Dridi, R. Brahim, and M. Machhout, "An efficient lightweight cryptographic instructions set extension for iot device security," 2022.
- [26] I. of Electrical and E. Engineers, 2018 IEEE High Performance Extreme Computing Conference (HPEC): the Westin-Hotel Waltham-Boston, 70 Third Avenue, Waltham, Massachusetts USA, 25-27 September 2018.
- [27] İbrahim Taştan, M. Karaca, and A. Yurdakul, "Approximate cpu design for iot end-devices with learning capabilities," 1 2020.
- [28] A. Zgheib, O. Potin, J. B. Rigaud, and J. M. Dutertre, "Extending a RISC-V core with an AES hardware accelerator to meet IOT constraints," SMACD / PRIME 2021 International Conference on Synthesis, Modeling, Analysis and Simulation Methods and Applications to Circuit Design and 16th Conference on PhD Research in Microelectronics and Electronics, no. July, pp. 244–247, 2021.

- [29] Q. Jiao, W. Hu, F. Liu, and Y. Dong, "Risc-vtf: Risc-v based extended instruction set for transformer." Institute of Electrical and Electronics Engineers Inc., 2021, pp. 1565–1570.
- [30] Y. Zhang, Z. Guo, J. Li, F. Cai, and J. Zhou, "Annikacore: Risc-v architecture processor design and implementation for iot," vol. 2021-October. IEEE Computer Society, 2021, pp. 200–203.
- [31] W. El Hadj Youssef, A. Abdelli, F. Dridi, R. Brahim, and M. Machhout, "An Efficient Lightweight Cryptographic Instructions Set Extension for IoT Device Security," Security and Communication Networks, vol. 2022, 2022.