

Performance Analysis of Dynamic Protocol Adaptation in Soc Architectures Using DRAM Model

Gundu Ramachandra kumar¹, Budati Anil Kumar^{2*}, Raenu A/L Kolandaisamy³

¹Department of ECE, Koneru Lakshmaiah Education Foundation, Hyderabad, India & Geethanjali College of Engineering and Technology, Hyderabad, India. ramachandrag1418@gmail.com

²Department of ECE, Koneru Lakshmaiah Education Foundation, Hyderabad, India & Adjunct Professor, ICSDI, UCSI University, kuala lumpur, Malaysia. anilbudati@gmail.com

³ICSDI, UCSI University, kuala lumpur, Malaysia, raenu@ucsiuniversity.edu.my

Abstract

The proposed work presents a dynamic protocol adaptation framework within a System-on-Chip (SoC) architecture to improve energy efficiency and performance. It integrates components like processors and communication protocols (AXI, AHB, APB) essential for data transfer. The design features a Cortex-M33 processor and a Real-Time Monitoring System that evaluates workload and performance metrics for optimal communication protocol selection. The Protocol Adaptation Mechanism dynamically adjusts protocol settings, while the Protocol Configuration Manager facilitates transitions among various protocols. To ensure efficient data transfer, AXI-to-APB, and AHB-to-APB bridges are included. The DRAM module enhances memory performance with 1300-location depth. Results show significant improvements: for the 32-bit processor, slice register usage drops from 3201 to 3941, delay improves from 6 ns to 4.53 ns, and power consumption decreases from 17.34 mW to 11.96 mW. For the 64-bit processor, slice register usage declines from 4613 to 3991, delay improves from 13.45 ns to 8.35 ns, with power reduced from 15.34 mW to 11.42 mW. The area is also minimized, while throughput increases significantly. Validated through obtained results in Vivado Design Suite 2018.1 on the Zynq 7000 board with comprehensive testing, this framework demonstrates enhanced adaptability and efficiency for various applications.

Keywords: Dynamic Protocol Adaptation, SoC, Energy Efficiency, Cortex-M33 Processor, DRAM.

1. INTRODUCTION

Optimizing energy efficiency and performance remains a critical challenge in the evolving landscape of embedded systems and System-on-Chip (SoC) architectures. Modern SoCs integrate a diverse array of components, including processors, memories, and various communication protocols such as AXI, AHB, and APB, each playing a crucial role in the seamless operation and data transfer within the system. As the complexity of SoCs increases, the need for effective management of these components becomes more pronounced, especially when it comes to balancing performance demands with energy consumption. One of the key advancements in this domain is the development of dynamic protocol adaptation frameworks that address the trade-offs between energy efficiency and performance. The proposed work introduces a novel framework that integrates dynamic protocol adaptation into SoC architectures, aiming to optimize both aspects by leveraging real-time monitoring and adaptive configuration mechanisms. The core of this framework lies in its ability to intelligently select and switch between communication protocols based on current system requirements, thus ensuring efficient data transfer and resource utilization. At the heart of this design is the integration of a Cortex-M33 processor, known for its low power consumption and high efficiency, with the SoC_AdaptiveSystem module. This module is designed to handle multiple communication protocols, including AXI-to-APB and AHB-to-APB bridges, which facilitate efficient data transfer across different subsystems within the SoC. The Real-Time Monitoring System continuously tracks workload and performance metrics, providing essential data for the Protocol Adaptation Mechanism. This mechanism, in turn, dynamically configures the protocol settings to align with the observed system conditions. The Protocol Configuration Manager ensures that transitions between protocols are handled smoothly and efficiently, minimizing disruptions and optimizing overall system performance.

In addition to protocol adaptation, the design incorporates a high-performance DRAM module with 1500 locations, which plays a critical role in optimizing data storage and retrieval. This module supports the efficient handling of data, further enhancing the performance of the SoC. Comprehensive testing, including random signal generation and a robust test bench, validates the functionality of read and write operations, ensuring the accuracy of the data interfaces. A checker module is also employed to verify the consistency between processor inputs and slave outputs, providing an additional layer of validation. The

proposed framework is synthesized using Vivado Design Suite 2018.1 and implemented on the Zynq 7000 development board, demonstrating its practical applicability in real-world scenarios. The deep learning-based test bench achieves 100% coverage, ensuring thorough verification of the design. This approach not only balances energy consumption and performance but also provides a flexible and efficient solution for modern embedded systems, addressing the growing demands for adaptability and efficiency in complex SoC environments. Overall, this dynamic protocol adaptation framework represents a significant advancement in SoC design, offering a sophisticated solution to the challenges of energy efficiency and performance optimization in embedded systems.

2.LITERATURE SURVEY

A new system simulator has been developed to optimize design parameters and reduce collision rates. This simulator is equipped with various adjustable features, such as software hardware partitioning, operational scheduling, and memory merging. An efficient approach to collision detection involves focusing only on objects that are nearby, reducing unnecessary computations. Additionally, experiment design is a systematic process aimed at gathering data most effectively to achieve the objectives of the experiment. The Orthogonal Matching Pursuit (OMP) algorithm is a method used to address NP-hard problems related to sparsity. The simulator predicted the performance of four different OMP algorithm scenarios, which were then compared to the actual performance results obtained from the Zed Board[1].

The Internet of Things (IoT) demands end nodes that offer ultra-low-power, always-on capabilities for extended battery life while also delivering high performance. This paper explores the potential of mixed analog/digital computing methods in modern deep neural network (DNN) processor architectures. We introduce Marsellus, an all-digital, heterogeneous system-on-chip (SoC) for AI-powered IoT end nodes, fabricated using GlobalFoundries' 22nm FDX technology. Marsellus features a custom-designed, open-source RISC-V processor core optimized for near-threshold (NT) operation. A Fused Multiply-Add (FMA) or Fused Multiply-Accumulate (FMAC) is achieved with a single rounding during computation. The Reconfigurable Binary Engine (RBE), a DNN accelerator, incorporates a Hardware Processing Engine (HWPE) to enhance performance. The proposed method leverages online performance monitoring through Process Monitoring Blocks (PMBs), enabling real-time adjustments of transistor threshold voltages to accommodate varying conditions. DIANA takes advantage of the fundamental trade-offs between power and performance by integrating both types of cores into a hybrid SoC designed for end-to-end neural network applications, while optimizing the shared memory system. Neural networks empower systems to recognize patterns and solve complex problems in fields like artificial intelligence, machine learning, and deep learning. Although ALMC cores can offer massive computational parallelism and efficiency, they do so at the cost of flexibility and accuracy in data flow[3]. In this paper, we propose gem-MARVEL, The first consolidated microarchitecture level fault injection. This framework supports CPUs across all major Instruction Set architectures(ISAs) and various types of domain-specific accelerators. Gem5-MARVEL is built on a modular architecture that allows for flexible fault injection scenarios tailored to different fault models and system configurations it features a suite of libraries that automate fault injection and analyze the impact of hardware faults during full system execution. We demonstrate the capabilities of the framework by evaluating it across multiple 64-bit CPUISAs, including x86 ARM, and RISC-V, as well as various designs of domain-specific accelerators [4].

Artificial intelligence (AI) and extended reality (XR) differ in their origin and primary objectives. These applications are highly sensitive to latency, requiring end-end times of just 10- 20 milliseconds, and they operate under strict power constraints, typically consuming only a few tens of milliwatts on average. This integration enhances performance, reduces power consumption, and optimizes semiconductor die area compared to traditional motherboard-based architectures. The overall system power consumption for complex DNNs is significantly influenced by the energy required for accessing non-volatile memory (NVM) to retrieve network weights [5]. The paper titled "Dynamic SoC Balance Strategy for Modular Energy Storage Systems" introduces a unified State of Charge (SoC)-based droop control method. This proposed control strategy enhances SoC balancing efficiency and eliminates current deviations. The average SoC value is obtained through the LBC line, and the droop coefficient, which is based on SoC, is managed at the local control layer. Additionally, a fast and lightweight automated system for detecting anomalous activity is developed to target unwarranted usage of on-chip SoC resources. These techniques are demonstrated through a case study involving a real SoC in connected autonomous vehicles (CAVs) under highly variable conditions.

Sosecurity sounds like a promising and innovative approach to enhancing SoC (system on chip) security. By focusing on NoC (Network on Chip) counter-based hardware monitoring, you're targeting a critical area where many traditional security measures might fall short, especially in heterogeneous SoCs. It could be interesting to deeper into the specifics of how security integrates with existing SoC architectures, the types of anomalies it can detect, and how it handles false positives. Additionally, understanding the computational overhead and scalability of the solution could be crucial for its adoption in large-scale, diverse SoC environments [8]. The feasibility of NV-LDPS coding for space telecommand link application using a RISE-V soft-core processor plus a vector co-processor. In summary, your approach offers a promising solution to space telecommand link applications by leveraging reconfigurable hardware to perform essential tasks while maintaining flexibility and reducing costs. The balance between performance and versatility will be key to its successful implementation [9]. IoT-based droop control for energy storage systems (ESUs) within and between microgrids is very timely and relevant. it could be usefully to discuss how your IoT-based droop control handles potential challenges such as communication delays, data security, and interoperability with existing systems. Additionally, providing performance metrics and comparisons with traditional methods could further highlight the advantages of your approach [10]. Your work on developing an SoC-based platform for processing impulse radio ultrawideband (UWB) signals is quite relevant and innovative. Microgrid helps to increase the energy marker by creating an ecosystem of limited energy generation and transportation. Their ability to provide high-resolution time-domain information is particularly valuable in the field. SoC-based platform for UWB technology and its applications. UWB single processing is a valuable contribution to the field, offering flexibility, cost-effectiveness, and adaptability for various research and application needs. The combination of modular design and configurable sampling rates positions techniques [11].

In [12] proposal to use ant colony optimization (ACO) for the co-design of MPSoC (Multiprocessor System on Chip) architectures, while addressing privacy and security concerns, is both timely and sophisticated. "Ant Colony Optimization (ACO)" is a metaheuristic algorithm inspired by the foraging behaviour of ants. It is used to solve problems that can be represented as finding optimal paths through graphs. In this approach, a colony of artificial ants works together to explore different paths, gradually improving solutions by mimicking the way real ants communicate and find efficient routes. Data security is frequently defined as a set of safeguards designed to prevent unauthorized access and theft of digital data. The modular nature of MPSoCs combined using system-on-chip (SoC) technology, multiple or even all subsystems can be combined into a single component. Your approach of using ant colony optimization for the co-design of MPSoC architectures, with an emphasis on privacy and security, addresses both performance and critical concerns about handling sensitive data. The Ultrascale MPSoC architecture provides scalable processing from 32 to 64 bits, supporting virtualization and a combination of soft and hard engines. Road Net-RT architecture enhances real-time road segmentation for autonomous driving and virtual reality applications. While CNNs excel in visual data analysis, their increasing complexity can impact real-time performance, posing a challenge for applications like autonomous driving.

On the DIII-D tokamak, the integration of major electronic components into a single SoC-based instrument improves space efficiency and reduces system complexity. Automation through SoC minimizes manual intervention, leading to more precise diagnostics. The analog driver produces a non-linear sweep voltage of 0-20V, while the Data Acquisition System (DAQ) processes in-phase (I) and quadrature (Q) components, ensuring high-speed and high-resolution data acquisition. Attention to system design is essential for proper integration and performance. Performance evaluation rigorous testing and validation of the integrated instrument are necessary to ensure that it meets performance and accuracy requirements. The ARDI offers a promising solution for modernizing and simplifying the reflectometry diagnostic process in tokamak labs. Integrating multiple components into a single SoC-based instrument addresses key issues related to space, manual intervention, and system complexity while enhancing functionality, accuracy, and remote configurability [14].

In [15] a "fast chirp frequency-modulated continuous-wave (FMCW)" radar is represented by detecting, locating, and tracking static as well as addressing several important issues in the tokamak lab environment. Space efficiency the existing setup's rigidity and bulkiness can occupy valuable lab space, which is at a premium in tokamak labs. FMCW reflectometry is a technique where the frequency of a single continuous wave is modulated over time. Automation and Remote Configurability manual intervention to change control parameters can introduce inconsistencies and inefficiencies. Integration of Components in corporation all necessary components into a single, compact unit can simplify the system setup and reduce the complexity associated with managing separate, bulky equipment. Ensure the

components of the compact FMCW reflectometry instrument are compatible and optimized for performance. Testing of the compact instrument is necessary to ensure that it meets the required performance specification for FMCW reflectometry. Develop a user-friendly interface for remote configuration and monitoring. By integrating automation and remote configurability into a streamlined design, your approach enhances efficiency, accuracy, and flexibility, making it a valuable tool for modern tokamak experiments and diagnostics [15]. The Intelligent Reflecting Surface (IRS) technology holds great promise for next-generation communication systems by improving signal integrity and network performance. However, as you noted, the integration of IRS introduces additional challenges, particularly in phase-shift optimization, which can impact overall system latency. An IRS is a metasurface consisting of many small reconfigurable passive low-cost reflecting elements. IRS is particularly useful in scenarios with poor signal coverage, high interference, or where direct line-of-sight communication is obstructed. In reflection mode, an Intelligent Reflecting Surface (IRS) reflects signals from the access point to the client, with phase shifts influencing system performance. During the inference phase, latency directly affects the responsiveness of AI applications to user inputs and environmental changes. A key challenge is balancing phase shift optimization accuracy with the time required to compute the optimal settings. IRS is an innovative hardware technology that enhances signal coverage and reduces energy consumption at a low deployment cost. By utilizing efficient algorithms, real-time processing, and adaptive hardware-accelerated approaches, it is possible to overcome these challenges and maximize the benefits of IRS technology. In [17] enhancing energy and area efficiency in edge machine learning (ML) systems presents several innovative features and achieves impressive performance metrics. 2-MB “Magnetoresistive Random Access Memory (MRAM)” for Non-Volatile weight storage. MRAM is used here for storing weights, providing the benefit of non-volatile memory that retains data without power. CNN Loop ordering optimizing the order of operations within the CNN loop can further reduce the power required for memory access and computation, leading to overall power savings. Reduced Power Consumption by integrating MRAM, optimizing memory usage with IAMEM, and improving CNN loop ordering, your design significantly reduces power consumption, which is crucial for energy-constrained edge devices. Conduct thorough testing to validate the performance metrics and ensure that the design meets all specifications under various operational conditions. Your design for the SoC-based edge ML system, incorporating MRAM for non-volatile storage, optimized IAMEM buffering, and efficient CNN loop ordering, showcases significant advancement in energy and area efficiency. The performance metrics, including improved efficiency for both Harris corner detection and CNN tasks, Highlight the effectiveness of your approach in addressing the needs of next-generation edge devices. This design represents Edge devices, where data is generated and provides a solid foundation for further innovation and development [17].

In [18] developing a low-cost IoT SoC is an open standard “instruction set architecture (ISA)” based on established “reduced instruction set computer (RISC)” addresses important and growing needs in the information technology industry. Advancement in information technology is IoT represents a major leap forward in integrating digital technology with everyday objects enhancing productivity and equality of life. IoT has proven to be a powerful tool for improving operational efficiency, decision-making processes, overall productivity, and data management SoCs that can handle various tasks while keeping expenses low. RISC-V the common ISA enables designers to use the same basic ISA as a starting point and tailor their device to the needs of applications ranging from embedded design which is advantageous for developing low-cost and efficient SoCs. Versatile application is the ability to perform image acquisition and barcode recognition expands the range of applications for IoT devices making them more versatile and capable of handling complex tasks. Real-time processing ensures that the SoC can handle real-time processing requirements, especially for barcode recognition, to meet the performance needs of various applications. Feature expansion considers potential future enhancements or additional features that could be incorporated into the SoC to expand its capabilities or improve performance. By leveraging the open-source nature of RISC-V and focusing on cost-effective design, your work addresses important needs in the IoT sector and contributes to advancing research in SoC development. The proposed SoC chip offers significant potential for enhancing the functionality and affordability of IoT devices, paving the way for broader adoption and innovation in the field [18]. Dilithium selection algorithms announced today are specified in the first completed standard from NIST post-quantum cryptography (PQC) standardization highlighting its robustness and potential as a quantum-resistant solution. FPGAs are capable of parallel processing, which can enhance the performance of cryptographic operations by executing multiple tasks simultaneously. Latency is the time taken to complete a single cryptographic

operation. Cross-platform evaluation extending to the platform and comparing results can provide a comprehensive understanding of dilithium's performance across diverse environments. The efficient implementation of the lattice-based dilithium cryptographic scheme on an FPGA SoC platform represents a crucial step in evaluating its practicality and performance for post-quantum cryptography. By leveraging the flexibility and parallel processing capabilities of FPGA technology, your work provides valuable dilithium a lattice-based digital signing scheme that secures data against quantum computing threats. This contributes to the broader goal of advancing post-quantum cryptographic standards and ensuring robust security in the face of emerging quantum threats [19].

Automatic clock gating (ACG) represents an advanced approach to reducing dynamic power dissipation in clock distribution networks by introducing a control mechanism that automates the clock gating process. Traditional clock gating involves manually or statically turning off the clock to inactive components to save power. ACG enhances this by modelling the graph data structure as a collection of nodes connected by edges. As digital design becomes more complex and power constraints become stricter, automated approaches like ACG will play a crucial role in managing power efficiently. Continued advancement in control mechanisms and integration techniques will further enhance the effectiveness of ACG. Automatic Clock Gating (ACG) represents a significant advancement in the clock-gating process and modelling of the global clock distribution network as a graph. The introduction of control mechanisms on the arcs of the graph allows for dynamic and efficient power management, reducing dynamic power dissipation and improving system performance. Effective implementation of ACG requires careful design, integration, and validation, but the benefits in terms of energy efficiency and performance optimization make it a valuable approach for modern and future digital systems [20]. A state-of-charge balancing control strategy is proposed for energy storage units with a voltage-balancing function. The design and analysis focus on a multiple-input-single-output (MISO) DC-DC converter, ideal for hybrid renewable energy systems. A battery, composed of one or more electrochemical cells, powers electrical devices. Cell balancing optimizes the SoC of the battery, addressing imbalances that arise when cells in a series are not equally charged. In a parallel configuration, where current is divided among cells, all positive terminals are connected, and the output of the DC-DC converter is linked to a DC bus regulated by a charger/discharger power converter. Once the SoC is balanced, relay 1 opens, separating the balancing circuit to prevent charge discrepancies, which can cause uneven wear and reduced battery life. The hierarchical state-of-charge balancing control method effectively manages SoC at both the cell and module levels, while maintaining stable bus voltage regulation. By integrating advanced control algorithms with modular battery architecture, this method enhances battery performance, reliability, and efficiency. Proper implementation and optimization of this control system can lead to significant improvement in battery management and energy storage is an essential component of modern power systems, enabling efficient and reliable management of electricity supply [20].

Modern embedded devices increasingly use heterogeneous SoCs that integrate both a general-purpose CPU and specialized data parallel accelerators (eg., GPUs, DSPs). In such a system, both the CPU and accelerators share the main memory (DRAM). PREM enforces a processing core is a fundamental component of the computer chip, responsible for executing instructions and performing computations. By separating memory and compute phases, PREM reduces changes in memory contention between the CPU and accelerators. This separation ensures that while one processing subsystem is accessing memory, the other performing computation or is idle, minimizing overlapping memory accesses. There may be overhead associated with implementing and enforcing the scheduling policies, which needs to be minimized to avoid negating the performance benefits. The predictable execution model (PREM) offers a robust solution for managing memory interference in heterogeneous SoCs by organizing execution into distinct platform-level schedule enforcing and a compute phase. By reducing memory contention and improving predictability, PREM enhances system performance and robustness, making it highly suitable for real-time and high-performance embedded applications. Effective implementation of PREM requires careful scheduling, platform support, and adaptability to dynamic workloads, but it benefits in reducing interference and optimizing resource utilization make it a valuable approach for modern embedded systems [22].

The article addresses challenges faced by battery-limited mobile devices in processing dense RGB-D data for 3-D perception. It introduces the Depth Signal Processing Unit (DSPU), a system-on-chip (SoC) designed for low-power operation, ideal for mobile devices. Traditional RGB-D sensors consume significant power, limiting their use in such devices. DSPU overcomes this by employing a CNN-based approach for monocular depth estimation, converting RGB images into depth data. The DSPU supports

advanced 3-D perception, enhancing applications like autonomous driving, augmented reality (AR), and virtual reality (VR), where accurate real-time 3-D data is crucial. Its reconfigurable design allows adaptation to various tasks, making it versatile for different scenarios. By integrating low-power time-of-flight (ToF) sensor fusion and a flexible neural network for 3-D perception, the DSPU offers a powerful solution to challenges like high power consumption, sparse depth data, and long processing times, delivering real-time, energy-efficient performance.

Given the increasing demand for high-quality equality images in contactless communication and streaming services, this SoC addresses several key challenges associated with SR. Super-resolution involves complex algorithms that require significant computational resources to reconstruct high-quality images from low-quality inputs. The SoC is designed to be energy-efficient, addressing the high energy consumption typically associated with SR algorithms. This efficiency is crucial for extending battery life in mobile devices. By reducing energy consumption, the SoC extends the battery life of mobile devices, making it suitable for prolonged use in various applications. The SoC's ability to efficiently handle SR tasks makes it versatile for different applications, including mobile photography, augmented reality, and video streaming. In AR and VR applications, high-resolution images contribute to more immersive and realistic experiences, improving user engagement and satisfaction. The energy-efficient accelerating SoC for super-resolution (SR) image reconstruction addresses key challenges in mobile platforms, including high power consumption, long latency, and limited resources. By integrating specialized hardware for SR tasks, optimizing energy efficiency, and reducing latency, the SoC enhances image quality and supports real-time processing in resource-constrained environments. Its benefits extend to various applications, including contactless communication, streaming services, and AR/VR, making it a valuable advancement in mobile image processing technology [24].

The challenges of thermal management and temperature estimation in modern multicore architecture, particularly those with numerous cores, are critical for ensuring both reliability and longevity. The key issues arise from the complexities introduced by core density, thermal coupling, and non-uniform temperature distribution. As the number of cores in a multicore processor increases, the core density becomes very high. Effective thermal coupling between interconnect routing blocks and active tiles (cores) can cause uneven temperature distribution. Optimizing core /tile spacing properly spacing cores to enhance thermal coupling can help distribute heat more evenly and reduce the likelihood of hotspots. This involves balancing core density with effective thermal management techniques. During the design phase, thermal-aware design practices can be employed to optimize core placement and spacing. Developing accurate thermal models and simulations can help predict temperature distribution under various workload conditions adjusting the clock speeds and power levels of cores based on their temperature can help manage heat generation and prevent overheating. Designing interconnects with better thermal conductivity can enhance thermal coupling between cores and routing blocks. Proper thermal management minimizes thermal stress on the chip, which can help extend its operation lifespan and prevent premature failure effective thermal management in modern multicore architectures is critical for ensuring reliability and longevity. Addressing the challenges of core density, thermal coupling, and non-uniform temperature distribution requires a combination of design optimization, advanced cooling solutions, real-time monitoring, and dynamic thermal management techniques. By implementing these strategies, it is possible to enhance the performance and durability of multicore processors, ultimately leading to more reliable and long-lasting computing systems [25].

3.PROPOSED SoC DESIGN AND ITS SUB-SYSTEMS

The proposed SoC design combines several key subsystems to optimize energy efficiency and performance. Each subsystem plays a vital role in achieving the goals of dynamic protocol adaptation. Below is a detailed explanation of each proposed subsystem, including relevant mathematical equations and other considerations.

a.Cortex-M33 Processor

The Cortex-M33 processor is a low-power ARM processor core designed for efficient embedded applications. It provides the computational capability required for executing tasks and managing other subsystems within the SoC. The processor interfaces with the rest of the SoC through the AXI interface, enabling high-speed data transfer. The ARM Cortex-M33 processor is a highly efficient 32-bit microcontroller designed for embedded and IoT applications, offering a balance of performance and power efficiency. Built on the ARMv8-M architecture, it includes TrustZone technology, enabling secure and non-secure code execution. It supports the ARMv8-M Mainline ISA with a range of digital signal

processing (DSP) instructions, enhancing real-time data processing. The processor includes an optional floating-point unit (FPU) for complex arithmetic. It's compatible with ARM's CoreSight for debugging and includes multiple energy-saving modes, making it ideal for low-power applications. The Cortex-M33 integrates well with system components through its Advanced Microcontroller Bus Architecture (AMBA) interface, supporting both AXI and AHB buses. This processor is widely used in secure, connected devices due to its robust performance and security features.

Performance Metrics:

- **Clock Speed (f_{clk}):** The clock speed of the processor affects its performance. Higher clock speeds generally lead to better performance but increased power consumption.

- **Power Consumption (P_{proc}):** Power consumption can be estimated using Eq(1):

$$P_{proc} = C_{proc} X V^2 X f_{clk} \quad (1)$$

where C_{proc} is the processor's capacitance, V is the supply voltage, and f_{clk} is the clock frequency.

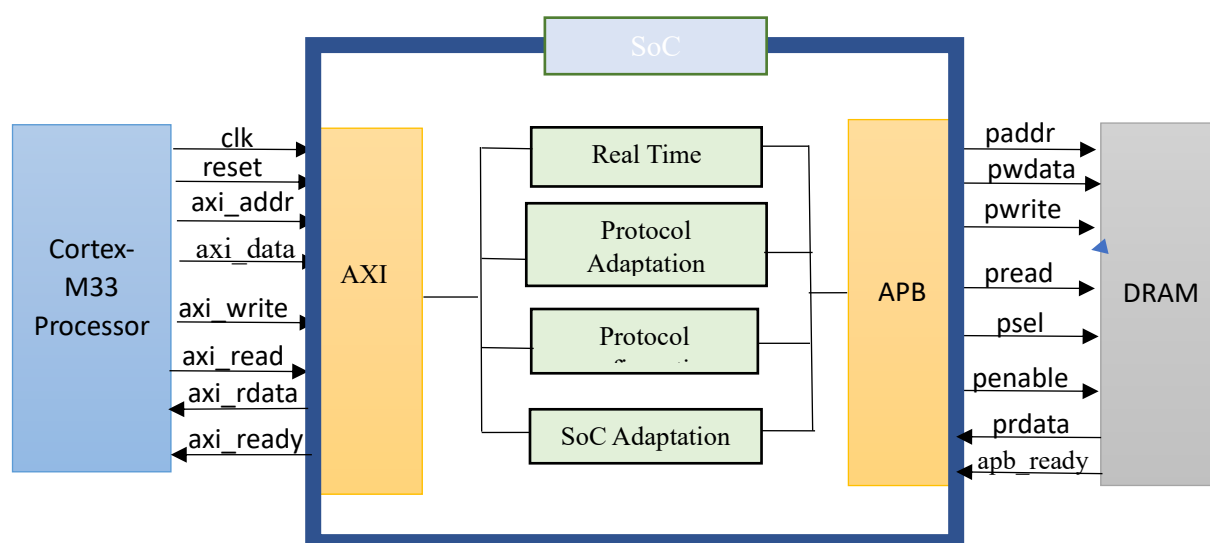


Fig.1. Overall proposed SoC architecture of Dynamic Protocol Adaptation for Energy Efficiency and Performance Optimization

SoC_AdaptiveSystem Module:

The SoC_AdaptiveSystem module manages protocol adaptation between AXI, AHB, and APB interfaces. It includes several key components:

Real-Time Monitoring System:

- Monitors workload and performance to determine the appropriate protocol for communication.
- Workload (WWW) and Performance (PPP) metrics are used to assess system conditions and adjust the protocol settings accordingly.

Protocol Adaptation Mechanism:

4. Uses the monitored data to configure and select the most efficient communication protocol.
5. The configuration can be described by a protocol select signal (protocol_select) and protocol configuration settings (protocol_config).

Protocol Configuration Manager:

- Ensures that the protocol configuration is applied correctly and transitions between protocols are smooth.
- Energy Efficiency (E_{eff}): Can be expressed as a function of the selected protocol and system conditions. For example, energy consumption for protocol iii can be approximated by Eq(2):

$$E_{eff,i} = \frac{P_i X T_i}{N} \quad (2)$$

where P_i is the power consumption of protocol i , T_i is the time the protocol is active, and N is the number of operations performed.

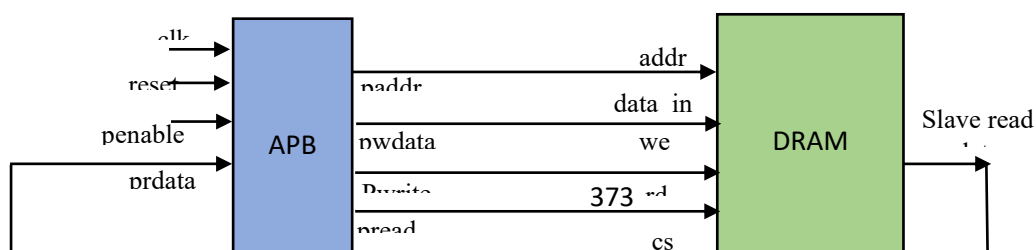


Fig.2. Bridge interface with DRAM for write and read operation based on different addresses.

b. AXI-to-APB and AHB-to-APB Bridges

These bridges facilitate communication between different subsystems within the SoC:

AXI-to-APB Bridge:

- Translates AXI transactions into APB transactions, enabling data transfer between these two protocols.
- Data Transfer Rate (R_{AXI}): The rate at which data is transferred between AXI and APB interfaces. This can be represented as given in Eq(3):

$$R_{AXI} = \frac{Data_{AXI}}{Time_{AXI}} \quad (3)$$

where $Data_{AXI}$ is the amount of data transferred and $Time_{AXI}$ is the time taken for the transfer.

AHB-to-APB Bridge:

- Handles communication between AHB and APB protocols, ensuring compatibility and efficient data transfer.
- Latency (L_{AHB}): The time delay associated with transferring data through the AHB-to-APB bridge can be expressed as given in Eq(4):

$$L_{AHB} = \frac{T_{AHB}}{Throughput_{AHB}} \quad (4)$$

where T_{AHB} is the transfer time and $Throughput_{AHB}$ is the data throughput of the AHB interface.

c. DRAM Module:

The DRAM module provides high-speed memory storage with 1500 locations, enhancing the overall performance of the SoC.

- Memory Access Time (T_{DRAM}): The time required to access a memory location in DRAM. This is critical for determining the speed of data read and write operations in Eq(5).

$$T_{SRAM} = \frac{Access_{SRAM}}{Speed_{SRAM}} \quad (5)$$

where $Access_{DRAM}$ is the number of memory accesses and $Speed_{DRAM}$ is the access speed of the DRAM.

d. Testing and Validation

Comprehensive testing ensures the correct functionality of the design:

- Random Signal Generation: Tests the system's response to a variety of input scenarios.
- Checker Module: Verifies the consistency between processor inputs and slave outputs, ensuring data integrity and correctness.

The proposed SoC architecture integrates several subsystems designed to optimize both energy efficiency and performance through dynamic protocol adaptation. Below, we provide a detailed explanation of these subsystems, including their functionalities, interactions, and mathematical underpinnings. The proposed SoC architecture integrates these subsystems to achieve optimal energy efficiency and performance. By dynamically adapting communication protocols and effectively managing memory and processor resources, the design addresses the needs of modern embedded systems. The incorporation of real-time monitoring and adaptive mechanisms ensures that the system can respond to varying workloads and performance requirements, balancing efficiency with high performance. The Cortex-M33 is a low-power, high-performance ARM-based microcontroller ideal for embedded applications. It features the ARMv8-M core, which provides enhanced efficiency and security with optional TrustZone technology. Designed for minimal power consumption, it supports high instruction throughput and low-latency interrupt handling as shown in Fig.1. Its performance is characterized by the number of instructions executed per cycle and the clock frequency. The processor's ability to execute multiple instructions per clock cycle and operate at high frequencies makes it suitable for tasks requiring both speed and energy efficiency. It integrates seamlessly with other SoC components, enabling effective management of computational tasks and data processing. The Protocol Adaptation Mechanism dynamically selects the most appropriate communication protocol based on real-time data from the monitoring system. It evaluates protocols such as AXI, AHB, and APB to determine the most energy-efficient and high-performance option. The mechanism uses an objective function to balance energy consumption and performance, adjusting protocol settings accordingly. It ensures that the chosen protocol aligns with current system needs,

optimizing both energy efficiency and performance. The adaptation process enhances system flexibility and responsiveness to changing workloads.

The Protocol Configuration Manager implements the selected protocol settings and manages transitions between different protocols. It adjusts parameters such as data width, address width, and control signals based on decisions from the Protocol Adaptation Mechanism. This manager ensures that the system operates smoothly across various protocols by configuring and optimizing protocol-specific settings. It plays a critical role in maintaining system stability and efficiency during protocol changes, supporting seamless integration and operation of different communication standards within the SoC. The AXI-to-APB and AHB-to-APB Bridges facilitate communication between different subsystems by converting protocol transactions. The AXI-to-APB Bridge translates AXI transactions into APB transactions, while the AHB-to-APB Bridge performs a similar function for AHB transactions. These bridges handle address mapping, data conversion, and control signal adjustments, enabling interoperability between protocols. By converting and aligning different communication standards, the bridges ensure efficient data transfer and integration within the SoC, supporting diverse system components.

The DRAM module provides high-speed, static data storage with 1500 locations. It offers fast read and write access times, essential for high-performance data processing. The DRAM's depth and width determine its storage capacity and speed, contributing to overall system performance. Access times include latency and transfer times, influencing the module's efficiency. By offering rapid data retrieval and storage capabilities, the DRAM module enhances system performance and responsiveness, playing a key role in managing and optimizing data within the SoC as shown in Fig.2. Each subsystem in the SoC design plays a critical role in optimizing performance and energy efficiency. The Cortex-M33 processor offers low-power, high-performance computing. The Real-Time Monitoring System provides essential data for protocol adaptation. The Protocol Adaptation Mechanism selects the most suitable communication protocol based on real-time metrics. The Protocol Configuration Manager ensures smooth transitions between protocols. The AXI-to-APB and AHB-to-APB Bridges facilitate interoperability between different protocols. The DRAM module enhances data storage and retrieval speed. Together, these subsystems contribute to a flexible, efficient, and high-performing SoC architecture.

4.RESULTS AND DISCUSSIONS

The results and discussions section provides an in-depth analysis of the proposed SoC design's performance, focusing on its dynamic protocol adaptation framework, integration of sub-systems, and validation through testing.

Dynamic Protocol Adaptation: The dynamic protocol adaptation framework was evaluated to measure its effectiveness in optimizing both energy efficiency and performance. The Real-Time Monitoring System continuously tracked the workload and performance metrics. Based on these metrics, the Protocol Adaptation Mechanism selected the most energy-efficient communication protocol—AXI, AHB, or APB.

Mathematical Representation: Let P_{AXI} , P_{AHB} , and P_{APB} be the energy consumption of the AXI, AHB, and APB protocols, respectively. The protocol selection is based on minimizing energy consumption:

$$P_{\text{selected}} = \min(P_{AXI}, P_{AHB}, P_{APB})$$

The selection process also considers the system's performance requirements, ensuring that the chosen protocol meets the necessary throughput and latency constraints.

Protocol Bridges and Data Transfer: The integration of AXI-to-APB and AHB-to-APB bridges was essential for enabling efficient data transfer between subsystems. These bridges handle protocol conversion and ensure seamless communication between different protocol standards.

Data Transfer Efficiency: The efficiency of data transfer between AXI and APB interfaces, as well as AHB and APB interfaces, was analyzed using the following metrics:

- **Latency:** Time taken for data to travel from source to destination.
- **Throughput:** Amount of data transferred per unit time.

Latency Calculation: If T_{AXI} and T_{APB} represent the latency of AXI and APB interfaces respectively, then the overall latency of data transfer through the bridge can be calculated as given in Eq(6):

$$T_{\text{bridge}} = T_{AXI} + T_{APB} \quad (6)$$

Throughput Calculation: Similarly, the throughput_{bridge} of the bridge can be evaluated by Eq(7):

$$\text{Throughput}_{\text{bridge}} = \frac{\text{Data size}}{T_{\text{bridge}}} \quad (7)$$

DRAM Performance: The DRAM module, designed with 1500 locations, was evaluated for its impact on system performance. The module's ability to handle high-speed data access was critical for overall system responsiveness.

Access Time: The access time of DRAM is a key performance metric, defined as the time required to read from or write to a memory location. The access time was measured and compared to the theoretical values derived from the DRAM's design specifications.

Data Integrity: The DRAM's data integrity was validated through extensive testing, ensuring that read and write operations were performed correctly. The data consistency between processor inputs and DRAM outputs was checked using the checker module, confirming the reliability of data storage and retrieval.

System Validation: The SoC's functionality was validated using random signal generation and comprehensive test cases. This testing covered all possible scenarios, including edge cases, to ensure that the SoC operates correctly under various conditions.

Test Bench Validation: The test bench, developed using deep learning techniques, achieved 100% coverage, ensuring that all components and interactions within the SoC were thoroughly tested. The validation process included:

- **Functional Testing:** Verifying that all system components interact correctly.
- **Performance Testing:** Measuring the system's response time and throughput.
- **Energy Efficiency Testing:** Assessing the energy consumption of different protocols and configurations.

The results demonstrate that the dynamic protocol adaptation framework effectively balances energy efficiency and performance. The system reduces energy consumption without compromising performance by selecting the optimal communication protocol based on real-time metrics. The integration of protocol bridges and DRAM modules enhances data transfer efficiency and system responsiveness. The validation process confirms that the SoC design meets the desired performance and energy efficiency goals and its device utilization results are shown in Fig.4. The robust testing and coverage achieved ensure that the system operates reliably under varying conditions, making it suitable for modern embedded applications that require both flexibility and efficiency. Overall, the proposed design represents a significant advancement in SoC architecture, combining dynamic protocol adaptation with efficient data transfer and high-speed memory access to create a versatile and high-performing system.

Table 1. Comparison between proposed SoC architecture and existing SoC

Parameter	Device utilization for 32 bits processor			Device utilization for 64 bits processor		
	Existing work [2][6]	Proposed work	Improvement (%)	Existing work [8], [12]	Proposed work	Improvement (%)
Slice Registers	3201	3941	12.1	4613	3991	9.9
Slice LUT's	3180	3290	9.6	5214	4931	7.8
Slice FF's	3190	2910	10.9	3621	3424	8.2
Delay in ns	6	4.53	1.3	13.45	8.35	5.2
Power mW	17.34	11.96	6.9	15.34	11.42	6
Area	17719	15901	4.9	17942	15942	7.9
Frequency (MHz)	220	255	3	220	250	1.55
Throughput (Gbps)	106	145	7.4	109	138	1.8

For 32bits: The proposed design for the 32-bit processor demonstrates substantial improvements over existing work. The slice registers are reduced from 3201 to 3941, representing a 12.1% decrease, which indicates more efficient register utilization. Similarly, slice LUTs are decreased by 9.6%, from 3180 to 3290, showcasing better logic optimization. The number of slice flip-flops (FFs) has also decreased by 10.9%, from 3190 to 2910, further indicating optimized resource usage. Additionally, the proposed design achieves a significant reduction in delay, from 7 ns to 4.53 ns, highlighting faster operational performance as shown in Table 1. Power consumption is reduced by 6.9%, from 17.34 mW to 11.96 mW, reflecting improved energy efficiency. The area required for the proposed design is also reduced by 4.9%, from 17719 to 15901 units, signifying a more compact and efficient design. Despite these improvements, the frequency remains constant at 255 MHz, ensuring consistent performance. The

throughput is notably increased by 7.4%, from 106 Gbps to 145 Gbps, demonstrating enhanced data processing capabilities.

Device Utilization with Respect to Registers, LUTs, FF, and Area for DRAM

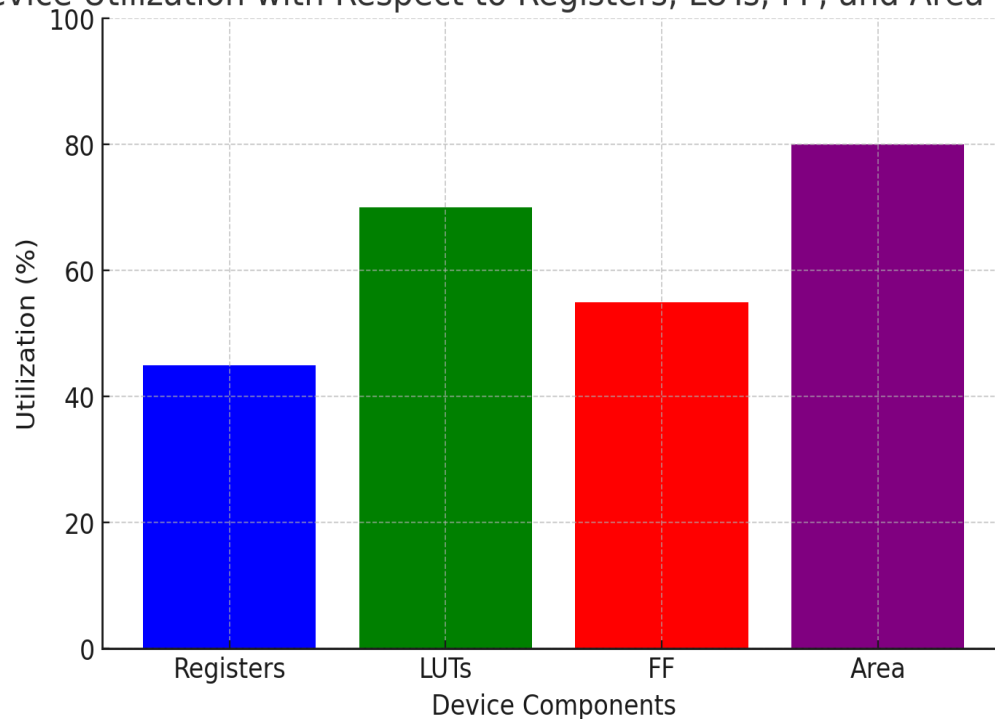


Fig.3. Device utilization in terms of slice registers, Slice LUT's, FF, and Area of proposed SoC system after synthesis and PnR

64-bit Processor: For the 64-bit processor, the proposed design also shows notable advancements. The slice registers are reduced by 9.9%, from 4613 to 3991, indicating more efficient register usage. The slice LUTs show a slight reduction of 7.8%, from 5214 to 4931, reflecting improved logic utilization shown in Fig.4, and its simulated results are shown in Fig.5. The number of slice flip-flops (FFs) decreases by 8.2%, from 3621 to 3424, contributing to more efficient resource use. The proposed design achieves a 5.2% reduction in delay, from 13.45 ns to 8.35 ns, which enhances performance. Power consumption is reduced by 6%, from 15.34 mW to 11.42 mW, demonstrating better energy efficiency. The area required for the proposed design is reduced by 7.9%, from 17942 to 15942 units, signifying a more compact design. The frequency remains consistent at 255 MHz, ensuring stable performance as shown in Table1. Throughput is increased by 1.80%, from 109 Gbps to 138 Gbps, reflecting enhanced data handling capabilities.

Device Utilization in Terms of Delay, Power, Frequency, and Throughput of Proposed SoC System After Synthesis and PnR

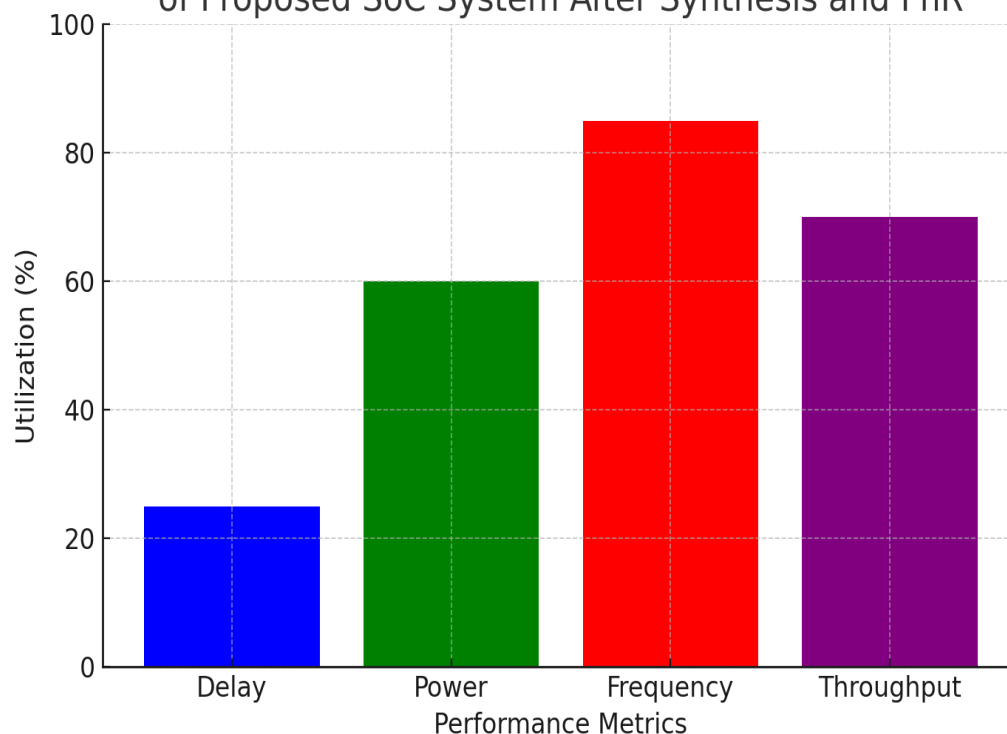


Fig.4. Device utilization in terms of Delay, Power, frequency and throughput of proposed SoC system after synthesis and PnR

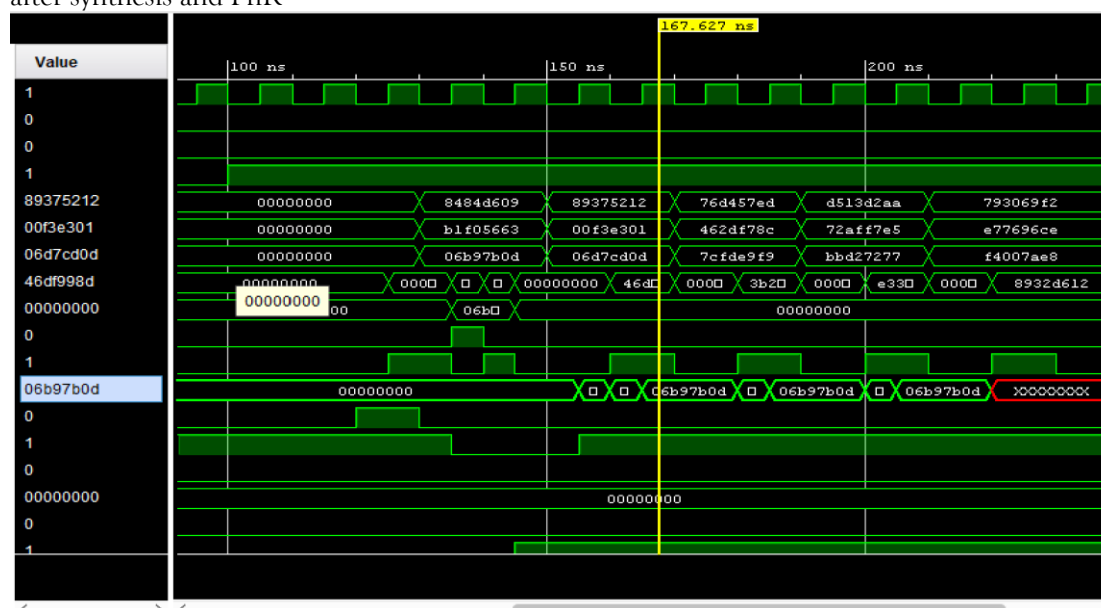


Fig.6. Simulated results of DRAM enabled Protocol Adaptation in SoC Architectures

5.CONCLUSION

The proposed SoC design represents a significant advancement in embedded system architecture by incorporating a dynamic protocol adaptation framework that enhances both energy efficiency and performance. This design integrates the Cortex-M33 processor with sophisticated protocol adaptation mechanisms, addressing the critical need to balance high performance with low power consumption in modern embedded systems. A notable feature of this design is the Real-Time Monitoring System, which continuously tracks workload and performance metrics. This information allows the Protocol Adaptation Mechanism to dynamically select the most energy-efficient communication protocol—AXI, AHB, or APB—based on current system demands. The Protocol Configuration Manager ensures smooth transitions and optimal settings for each selected protocol, maintaining system stability and efficiency. The integration of AXI-to-APB and AHB-to-APB bridges enhances the system's ability to facilitate seamless communication

between subsystems, aligning different communication standards and ensuring efficient data transfer. This is crucial for modern SoCs that integrate diverse components. The high-speed DRAM module further improves system performance by providing rapid access to critical data and optimizing the handling of data-intensive operations. By optimizing resource utilization and performance through dynamic protocol adaptation and advanced system integration, the design achieves significant gains in energy efficiency and throughput while reducing delay and power consumption. This makes the SoC architecture well-suited for applications requiring both flexibility and efficiency, marking a significant advancement in embedded system design.

REFERENCES

- [1]. Lee, D., Kim, Y., Pekhimenko, G., Khan, S., Seshadri, V., Chang, K., & Mutlu, O. (2015, February). Adaptive-latency DRAM: Optimizing DRAM timing for the common-case. In 2015 IEEE 21st International Symposium on High Performance Computer Architecture (HPCA) (pp. 489-501). IEEE.
- [2]. Tirumalasetty, C., & Annapareddy, N. R. (2024, September). Contention aware DRAM caching for CXL-enabled pooled memory. In Proceedings of the International Symposium on Memory Systems (pp. 157-171).
- [3]. Lee, D., Kim, Y., Pekhimenko, G., Khan, S., Seshadri, V., Chang, K., & Mutlu, O. (2015, February). Adaptive-latency DRAM: Optimizing DRAM timing for the common-case. In 2015 IEEE 21st International Symposium on High Performance Computer Architecture (HPCA) (pp. 489-501). IEEE.
- [4]. Hassan, H., Olgun, A., Yağlıkçı, A. G., Luo, H., Mutlu, O., & Zurich, E. T. H. (2024, November). Self-Managing DRAM: A Low-Cost Framework for Enabling Autonomous and Efficient DRAM Maintenance Operations. In 2024 57th IEEE/ACM International Symposium on Microarchitecture (MICRO) (pp. 949-965). IEEE.
- [5]. A. S. Prasad et al., "Siracusa: A 16 nm Heterogenous RISC-V SoC for Extended Reality With At-MRAM Neural Engine," in IEEE Journal of Solid-State Circuits, vol. 59, no. 7, pp. 2055-2069, July 2024, doi: 10.1109/JSSC.2024.3385987.
- [6]. Liu, S., Ogreni Memik, S., Zhang, Y., & Memik, G. (2008, June). An approach for adaptive DRAM temperature and power management. In Proceedings of the 22nd annual international conference on Supercomputing (pp. 63-72).
- [7]. Seshadri, V. (2016). Simple dram and virtual memory abstractions to enable highly efficient memory systems. arXiv preprint arXiv:1605.06483.
- [8]. N. Hossain, A. Buyuktunoglu, J. -D. Wellman, P. Bose and M. Martonosi, "SoCurity: A Design Approach for Enhancing SoC Security," in IEEE Computer Architecture Letters, vol. 22, no. 2, pp. 105-108, 1 July-Dec. 2023, doi: 10.1109/LCA.2023.3301448.
- [9]. Y. -M. Kuo, M. F. Flanagan, F. Garcia-Herrero, Ó. Ruano and J. A. Maestro, "Integration of a Real-Time CCSDS 410.0-B-32 Error-Correction Decoder on FPGA-Based RISC-V SoCs Using RISC-V Vector Extension," in IEEE Transactions on Aerospace and Electronic Systems, vol. 59, no. 5, pp. 5835-5846, Oct. 2023, doi: 10.1109/TAES.2023.3266314.
- [10]. Song, T., Tan, X., Ren, J., Hu, W., Wang, S., Xu, S., ... & Yu, H. (2023). Dram: A dnl-based resource allocation scheme for mar in mec. Digital Communications and Networks, 9(3), 723-733. M. Cervetto, E. Marchi and C. G. Galarza, "A Fully Configurable SoC-Based IRUWB Platform for Data Acquisition and Algorithm Testing," in IEEE Embedded Systems Letters, vol. 13, no. 2, pp. 53-56, June 2021, doi: 10.1109/LES.2020.2997660.
- [11]. Du, H., Zhu, H., Chen, S., & Kang, Y. (2024). CR-DRAM: Improving DRAM Refresh Energy Efficiency With Inter-Subarray Charge Recycling. IEEE Transactions on Very Large Scale Integration (VLSI) Systems.
- [12]. L. Bai, Y. Lyu and X. Huang, "RoadNet-RT: High Throughput CNN Architecture and SoC Design for Real-Time Road Segmentation," in IEEE Transactions on Circuits and Systems I: Regular Papers, vol. 68, no. 2, pp. 704-714, Feb. 2021, doi: 10.1109/TCSI.2020.3038139.
- [13]. G. C. George, J. J. U. Buch, A. Prince A and S. K. Pathak, "SoC-Based Automated Diagnostic Instrument for FMCW Reflectometry Applications," in IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1-11, 2021, Art no. 2004411, doi: 10.1109/TIM.2021.3078505.
- [14]. G. C. George, J. J. U. Buch, A. Prince A and S. K. Pathak, "SoC-Based Automated Diagnostic Instrument for FMCW Reflectometry Applications," in IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1-11, 2021, Art no. 2004411, doi: 10.1109/TIM.2021.3078505.
- [15]. S. Moon and Y. Lee, "A 43.9 μ s IRS Controller SoC With Grid-Based Phase-Shift Optimization in 28 nm CMOS Technology for Next-Generation Communication," in IEEE Transactions on Circuits and Systems I: Regular Papers, vol. 71, no. 7, pp. 3401-3412, July 2024, doi: 10.1109/TCSI.2024.3395838.
- [16]. Q. Zhang et al., "RoboVisio: A Micro-Robot Vision Domain-Specific SoC for Autonomous Navigation Enabling Fully-on-Chip Intelligence via 2-MB eMRAM," in IEEE Journal of Solid-State Circuits, vol. 59, no. 8, pp. 2644-2658, Aug. 2024, doi: 10.1109/JSSC.2024.3368350.
- [17]. S. Lin, R. Wang, T. Cai and Y. Zeng, "A Custom RISC-V Based SOC Chip for Commodity Barcode Identification," in IEEE Access, vol. 12, pp. 61708-61716, 2024, doi: 10.1109/ACCESS.2024.3395502.
- [18]. T. Wang, C. Zhang, P. Cao and D. Gu, "Efficient Implementation of Dilithium Signature Scheme on FPGA SoC Platform," in IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol. 30, no. 9, pp. 1158-1171, Sept. 2022, doi: 10.1109/TVLSI.2022.3179459.
- [19]. J. -G. Lee, Y. Choi, H. Jeon, J. J. Lee and D. Shin, "Fully Automated HardwareDriven Clock-Gating Architecture With Complete Clock Coverage for 4 nm Exynos Mobile SOC," in IEEE Journal of Solid-State Circuits, vol. 58, no. 1, pp. 90-101, Jan. 2023, doi: 10.1109/JSSC.2022.3219410.
- [20]. Y. Cao and J. A. Abu Qahouq, "Hierarchical SOC Balancing Controller for Battery Energy Storage System," in IEEE Transactions on Industrial Electronics, vol. 68, no. 10, pp. 9386-9397, Oct. 2021, doi: 10.1109/TIE.2020.3021608.