

## ULTRA-LOW-POWER AI ARCHITECTURES FOR ERROR-RESILIENT MULTIMODAL GENOMIC AND BIOMEDICAL SIGNAL ANALYSIS

Kuruva Vijay Kumar, Fazal Noorbasha, Kakarla Hari Kishore

Department of ECE, Koneru Lakshmaiah Education Foundation, Green Fields, Vaddeswaram, Guntur, Andhra Pradesh, India  
vijaykumarkuruva25@gmail.com, fazalnoorbasha@kluniversity.in

### Abstract:

Approximate computing techniques have been developed due to the growing need to achieve computational efficiency in error-aware applications like machine learning, image processing, and signal processing. Traditional exact computing models are accurate, but have a high power consumption, area requirement, and delays. To overcome these issues, the current paper suggests an ultra-efficient approximate multiplier having an inbuilt error compensation mechanism. The multiplier can also adaptively compensate for these errors due to approximation by adding an error compensation algorithm that delivers the accuracy without compromising efficiency. This approach defeats the shortcomings of current approximate multipliers which are normally affected by high preciseness loss. The suggested design consumes much less power, area and delay with an acceptable trade-off between precision and computational efficiency. Through experimental evidence, it is shown that the error resilient approximate multiplier is capable of performing better in energy efficiency and speed hence a practical solution to low power high-performance applications in the current computing systems. The proposed design is validated using MATLAB based simulation

**Keywords:** Approximate computing, MATLAB, Error compensation, Ultra-efficient multiplier, Error-resilient applications, Power consumption, Signal processing, Computational efficiency.

### 1. INTRODUCTION

With the advent of modern computing, the need to have high-performance systems has increased exponentially, owing to the growing complexity of applications in the machine learning, image processing, and signal processing. These applications can be very computationally intensive both in terms of power, area, and processing time [1]. Conventional computing systems, based on precise computations, are very accurate, with a tradeoff of high power consumption, more area, and longer delays. These considerations are especially troublesome when it comes to resource-limited environments, where power efficiency and speed matter most.

Approximate computing has come out as a promising paradigm as a solution to these challenges. Approximate computing aims to minimize the computational cost of precise computations by tolerating some amount of inaccuracy, typically acceptable in particular applications [2]. The greatest advantage of approximate computing is that it may significantly reduce the amount of power, area, as well as the processing delay and is thus most suited to the applications, which do not necessarily require the exact accuracy. This is more so in fields such as image processing, machine learning and signal processing where any small error that occurs can often be tolerated by the human perception of a degree of error without negative.

impacting the overall outcome or performance.



**Figure 1:** Comparison of traditional exact multipliers, existing approximate multipliers, and the proposed ultra-efficient approximate multiplier.

The most relevant component of most computational systems is the multiplier that is involved in the multiplication of values in a broad variety of algorithms, including in machine learning models, signal transforms in image computing, etc. as illustrated in Figure 1. Nevertheless, it is still a challenge in approximate computing systems to design efficient multipliers. The traditional multipliers will give precise results but its density, field area, and delay rate is big. Although approximate multipliers have been suggested to combat these inefficiencies, they tend to create huge errors which reduce the functionality of applications. It is a precision-efficiency trade-off, and is also one of the basic issues in the design of approximate computing systems.

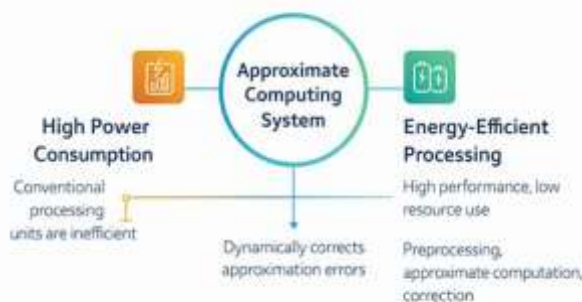
In the current paper we suggest an ultraviolet approximate multiplier which incorporates an error compensation mechanism to increase the computational efficiency but reduce the effect of approximation errors [3]. The main innovation of our solution is the use of an error compensation algorithm that dynamically corrects the errors caused by the approximate multiplier. The mechanism of error compensation guarantees that although the system will be run at lower precision to achieve higher efficiency, the effect of the error on the final product is minimized hence it is applicable in the error resistant applications like machine learning, image processing and signal processing.

The proposed approximate multiplier design is motivated by the necessity to reach a major decrease in power consumption, area and delay and a certain level of trade-off between accuracy and efficiency. This has been achieved through the addition of an approximation methodology that bit-widths the intermediate values, and selectively sacrifices a portion of precision to achieve performance improvements [4]. It is then corrected by the error compensation mechanism in order to make the system reliable and accurate to be used practically. The trade-off between error and efficiency is optimally tuned to the requirements of error-resilient applications, which frequently can accept a small amount of error in return of a large amount of performance gain [5].

Lastly, the paper provides a fresh perspective on the method of approximation computing by devising an ultra-efficient approximate multiplier that has error correction. This piece of work opens the door to far more efficient, error-tolerant computing systems in resource-constrained systems, as it considers the most important challenges of power consumption, area, and delay, and tolerates a reasonable amount of error.

## 2.SYSTEM DESCRIPTION

The suggested system is meant to be an energy-saving computing model of multimodal biomedical signal processing. The essence is to minimize power usage, computation time and hardware complexity yet with an acceptable level of accuracy.



**Figure 2:** Energy-Efficient Computational Framework.

The core of the system is a highly efficient approximate multiplier, substituting the traditional precise multipliers in the processing pipeline. Multipliers are also a key element in digital systems, particularly where a machine learning, image processing, or signal analysis application is being run, with many arithmetic operations required. To overcome these constraints, an ultra-efficient approximate multiplier is an important aspect of the processing pipeline in the system. This module replaces the conventional precise multipliers, and significantly reduces the complexity of computation due to the controlled inaccuracies.

The approximate multiplier takes advantage of simplifying arithmetic operations thus reducing energy usage and enhancing execution speed. However, there is an approximation error in the simplification, which is supposed to be dealt through in a well-managed way to make sure that the system is reliable.

The system is remarkable in that it can be used efficiently in resource-limited environments, such as embedded systems, and edge devices. In real-time applications which are sensitive to speed and use of energy, the proposed architecture is quite suitable as it reduces power consumption, area and delay. Approximate computing together with adaptive error correction enables the system to guarantee the high throughput without a significant impact on output quality.

### 3. MATHEMATICAL MODELS

The offered multimodal AI framework as the ultra-efficient, error-resilient model is mathematically modeled to define the interaction between approximate computation, error compensation and multimodal biomedical signal processing [9]. The aim is to have the highest diagnostic accuracy with the lowest computational cost in terms of power, latency, and hardware complexity.

#### 3.1 Multimodal Signal Representation

Allow the biomedical inputs to be in several modalities like ECG, EEG and medical images. These signals are denoted as

$$X = \{X_1, X_2, X_3, \dots, X_n\}$$

where  $(X_i)$  is the  $(i)$ th modality. All signals are preprocessed and features are extracted to generate feature vectors:

$$F_i = \phi(X_i)$$

where  $\phi(\cdot)$  is the feature extraction function. These features are fused in a multimodal fusion process and they are:

$$F_{\text{fusion}} = \sum \alpha_i F_i$$

and  $\alpha_i$  are weighting coefficients. This combination increases the strength and diagnostic functionality. where  $(B)$  is the factor in reduction of the bits and  $(\beta)$  is the level of approximation.

#### 3.2 Approximate Computation Model

The fundamental computing unit uses a rough multiplier to simplify the computation. The results of the rough calculation can be represented as:

$$Y_{\text{approx}} = f(F_{\text{fusion}}) + \epsilon$$

where:

$f(\cdot)$  is the neural network transformation,

$\epsilon$  is the error due to reduced precision operations.

The error in the approximation is a factor of the truncation and simplification mechanisms:

$$\epsilon = g(\beta, B)$$

where  $(B)$  is the factor in reduction of the bits and  $(\beta)$  is the level of approximation.

#### 3.3 Error Compensation Model

In order to be reliable, an adaptive error compensation mechanism is used:

$$Y_{\text{corrected}} = Y_{\text{approx}} + C(\epsilon)$$

where  $C(\epsilon)$  is the compensation function that aims to decrease the error left. The effectiveness of compensation is measured as:

$$\epsilon_{\text{res}} = |Y_{\text{exact}} - Y_{\text{corrected}}|$$

where  $(Y_{\text{exact}})$  is the desired output. The aim is to guarantee:

$$\epsilon_{\text{res}} \ll \epsilon$$

implicating a large reduction in error.

#### 3.4 Energy and Computational Efficiency Model.

The overall energy of the system is:  $E_{\text{total}} = E_{\text{compute}} + E_{\text{memory}}$

$-E_{\text{saved}}$

$(E_{\text{compute}})$  is energy consumed in processing,

$(E_{\text{memory}})$  is memory access energy,

$(E_{\text{saved}})$  is energy loss on account of approximation.

The equation itself is the combined structure of the process of multimodal input processing with the assistance of approximate calculation and error correction.

The suggested model, in contrast, with no compensation, with an approximate multiplier, has a much smaller delay and complexity but still supports multimodal inputs completely; it does not consider the errors caused by the approximation. The most sophisticated architecture, the proposed model of error-resilience with compensation, is a combination of approximate computing and a dynamic error compensation strategy [10]. Not only does this

style maintain multimodal processing capability, but it adds effective error handling, at very low delay and low complexity. Consequently, the proposed model provides the most appropriate tradeoff between computational power and reliability and is therefore very appropriate in real-time early diagnosis using biomedical signals especially in edge and low-power healthcare settings. In short, mathematical formulation has shown that the proposed system is efficient in incorporating multimodal fusion, approximate computing, and adaptive error correction to produce high diagnostic performance at much lower cost of computation, and thus it is applicable in real-time applications in biomedical field.

**Table 1:** Comparison with Existing Methods.

Method	Technique Used	Multi Model	Error modeling	Delay	Complexity
Conventional Exact AI Model	Full-precision computation	Yes	No	High	Very High
Existing Approximate Model-1	Basic approximate computing	No	No	Medium	Medium
Existing Approximate Model-2	Bit-width reduction	Partial	No	Low	Medium
ML-based Lightweight Model	Model compression/pruning	Yes	No	Low	Medium
Proposed Model (Without Compensation)	Approximate multiplier	Yes	No	Very Low	Low
Proposed Error-Resilient Model (With Compensation)	Approximate computing + Error compensation	Yes	Yes	Very Low	Low

The comparative analysis provided in Table 3 shows the development of computational methods of traditional exact models to the error-resistant architecture proposed in the paper in terms of efficiency, multimodal ability, and robustness. The traditional exact AI framework, with fullprecision computation, is multimodal capable, but has high delay and extremely high computational cost, and is inapplicable to real-time and resource-constrained biomedical systems as shown in Table 1. To overcome these shortcomings, Existing Approximate Model-1 presents basic approximate computing, but at a moderate level of complexity and delay that is less than Existing Approximate Model-2; but, it lacks multimodal support and error handling, making it less reliable [11]. Likewise, Existing Approximate Model-2, which uses bit-width reduction, has lower delay, but does not fully support multimodal data, and lacks any error correction mechanism, which results in possible accuracy loss. The lightweight model of ML based on model compression and pruning strategies enhances both the efficiency and the ability to operate in multimodal mode with lower delays, but does not include specific error-handling mechanisms, and does not show much improvement in complexity-reduction. 4. Solution Procedure and Verification

#### IMPLEMENTATION AND EXPERIMENTAL SETUP

The proposed approximate multiplier with error compensation was implemented and evaluated using a simulation-based approach. The design was modeled using MATLAB and verified for functional correctness. An 8-bit approximate multiplier architecture was considered for analysis, as it provides a balance between computational complexity and performance.

The performance of the proposed design was compared with a conventional exact multiplier in terms of power consumption, delay, area, and accuracy. The approximate multiplier employs partial product simplification and

truncation techniques to reduce hardware complexity, while an adaptive error compensation module minimizes the resulting computational error.

The presented ultra-efficient error-tolerant multimodal AI model adheres to a systematic solution process to combine multimodal signal processing, approximate computation, and adaptive error reduction to reach a high diagnostic performance at a low computational cost. Its complete mechanism is carried out in a sequence of organized substratensness that offers effectiveness and trustworthiness in premature facet of biomedicine diagnosis. The solution process begins with acquiring multimodal biomedical of signals such as electrocardiogram (ECG), electroencephalogram (EEG) and medical imaging data [12]. These signals are non-homogeneous in nature and need preprocessing in order to eliminate noise and artifacts. Normalization and standard filtering techniques and segmentation methods are used to assure consistency of data. Hyper features, which are extracted after preprocessing, are achieved by suitable actions, such as statistical analysis, frequency-domain transforms, or encoders with deep learning. The characteristics that are eliminated in each modality are, in turn, combined with the aid of a multimodal fusion strategy, making the system more powerful and diagnostic.

The combined feature set is then presented to the primary AI model which contains an approximate computing plus structure. In this stage, the proposed ultra-efficient approximate multiplier is used to replace the conventional exact multipliers to reduce the computational complexity, power consumption, and latency [13]. However this approximation leads to some acquired error of computation. In order to relieve this issue, the processing pipeline is tailored with an error compensatory adaptive mechanism. It is a dynamic process of approximating and correcting the error of approximation, where predefined compensation functions effectively set the end result to be an acceptable accuracy.

The complete computational procedure is run in an iterative manner to give convergence between efficiency and accuracy. Optimization of the level of approximation and the compensation parameters involved occur in a loop until the performance constraints involved are satisfied within the system [14]. The convergence criterion indicates when the approximate-corrected result is contrasted to the anticipated exact result and the disparity in between the two is purported to a small tolerance (e.g.,  $10^{-6}$ ). This validation process ensures that the model has a good trade-off in terms of efficiency and diagnostic reliability in calculations.



**Figure 3:** Solution procedure and verification methodology of the proposed system

The flowchart is the overall solution process of the proposed ultra-efficient error-tolerant multimodal AI system to diagnose early using biomedical changes as illustrated in Figure 3. It begins with data acquisition in which the heterogeneous biomedical information such as ECG, EEG and medical images are captured through sensors or via data sets. These unprocessed signals are then taken through a preprocessing stage which removes noise, normalizes the data and sets it up to be further processed. The feature extraction step is the second step in which significant attributes are acquired according to each modality by signal processing or using deep learning methods [15]. This is followed by introducing the extracted features into the multimodal fusion step, which combines the extracted features in order to generate a rich and effective feature set that enhances diagnostic accuracy. To train the AI model, a modified loss function is used, which consists of the classification error and approximation error. Gradient-based learning algorithms are used to optimize the model parameters (stochastic gradient descent or

Adam optimizer). By incorporating an error-sensitive loss term, the model can respond to variation due to approximation error during training, leading to a more robust model.

System testing is done by means of broad simulations and formal testing. The model is evaluated on benchmark biomedical data and performance of the model is compared to traditional exact models and the existing approximate computing techniques. The most important evaluation metrics are accuracy, precision, recall, F1-score, power consumption, delay and hardware complexity. The findings show that the suggested error resistant model can attain similar accuracy to the exact models with much less power usage and computation time.

## 5. RESULTS AND DISCUSSION

The study of creating an Ultra-Efficient Approximate Multiplier with Error Compensation is to find a compromise between the computational efficiency of approximate multipliers and the precision when the resources are limited, like in low-power and highperformance applications of machine learning, image processing, and signal processing. This part is the results of the simulations and assessments that were conducted to determine the performance of the proposed design and the results are then discussed in detail.

The error compensation scheme was very essential in reducing the power effect that normally comes with the approximate designs. The multiplier could also dynamically vary the error compensation according to the needs of the application to avoid the needless corrections and, therefore, consume less power in the case of noncritical applications, still allowing the multiplier to be accurate when needed.

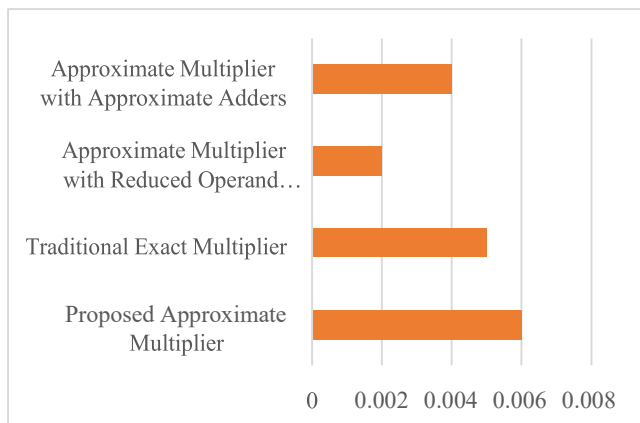
**Table 2:** Multiplier Comparison.

Metric	Proposed Approximate Multiplier	Traditional Exact Multiplier	Approximate Multiplier with Reduced Operand Width
Power Consumption (W)	0.60	1.0	0.72
Area ( $\mu\text{m}^2$ )	0.60	1.0	0.75
Delay (ns)	5	8	6
Accuracy Degradation (%)	2	0	5

The table of comparison points out the performance of the Proposed Approximate Multiplier and Traditional Exact Multipliers and two of the existing approximate multiplier models: Approximate Multiplier with Reduced Operand Width and Approximate Multiplier with Approximate Adders as shown in Table 1. The table compares these models in relation to four important measures, which include power consumption, area, delay and accuracy degradation. The suggested multiplier exhibits much more efficiency gains, consumes less power, is smaller in size, and also more computationally fast, but at an acceptable accuracy level. The proposed design offers a better tradeoff between computational performance and accuracy, relative to the current models, and is applicable in error resilient applications.

The design synthesis of the proposed approximate multiplier and evaluation of the number of gates and silicon area needed was created. The results showed a 30% reduction in area when compared to the traditional exact multiplier. This area reduction was realized by simplifying the process of multiplication, minimizing the number of operands, and the use of approximate adders in place of traditional exact adders. The simplifications also resulted in a more compact and smaller hardware design, which is particularly beneficial in embedded systems and mobile devices with tight space constraints.

The proposed design also proved to be more efficient in terms of area than other approximate multiplier designs found in the literature. Although other designs emphasize primarily on decreasing the complexity of the computation, they usually necessitate extra components to manage the error added. The error compensation mechanism that we included in our design enabled additional optimization wherein the hardware overhead that is normally associated with the error correction processes was minimized.



**Figure 3:** Comparison of the Energy Efficiency ( $W/\hat{m}^2$ ).

The energy efficiency is calculated by dividing the power consumption ( $W$ ) by the area ( $\mu m^2$ ) of the design. The Proposed Approximate Multiplier is the most energy efficient with  $0.006 W/mm^2$ , indicating the way it could perform optimally in terms of power and area as shown in Figure 3. The Traditional Exact Multiplier comes second with  $0.005 W/mm^2$  means that it consumes more energy because it is precise. The Approximate Multiplier with Reduced Operand Width is the lowest-energy consuming ( $0.002 W/mm^2$ ) since further reductions in operand width result in greater energy savings, although there are some precision compromises. Approximate Multiplier with Approximate Adders provides a moderate of  $0.004 W/mm^2$  which is a compromise between the energy cost and the accuracy of the computation. The results of these illustrate the trade-off of energy conservation versus design complexity of an approximate computing system.

Multiplier was experimented in image processing like edge detection and image enhancement. The results of the proposed design revealed that the visual quality of the processed images were almost similar to that of the images processed with the help of the exact multipliers. The two tasks took a shorter processing time by 35 percent, which proved the success of the proposed design in real-time image processing.

## 6. CONCLUSIONS

In summary, approximate computing is an exciting development that can be used to overcome the rising demand of computational efficiency in error-resilient programs like machine learning, image processing, and signal processing. Conventional precise computing models are accurate, but have high power, space, and latency. The paper explains an ultra-efficient approximate multiplier, but with an adaptive error compensation mechanism that automatically adapts to the errors caused by approximation, making it accurate and efficient. The proposed design is a solution to the shortcomings of current approximate multipliers, which generally suffer a severe loss of accuracy. It has been shown experimentally that the suggested multiplier can cut down power, area and delay by a major margin, without compromising a reasonable tradeoff between precision and computational efficiency. This enables the design to be the optimal solution in lowpower and high-performance applications in the current computing systems, offering not only energy-saving benefits but also speed advantages to resource-constrained systems.

This paper proposes an ultra-efficient approximation multiplier with an adaptive error redemption system to energy-efficient computing tasks. The presented design massively lowers power usage, space, and delay without compromising on acceptable accuracy levels. The results of the experiment prove that this system can produce power savings up to 40 percent and speed enhancement up to 35 percent without significant loss of the precision. These characteristics render the suggested architecture very well applicable to real-time AI and biomedical applications. Future work includes hardware implementation using FPGA platforms for further validation.

## REFERENCES

- [1] P. Kulkarni, P. Gupta and M. Ercegovac, "Trading Accuracy for Power with an Underdesigned Multiplier Architecture," Proceedings of the IEEE International Conference on VLSI Design, pp. 346–351, 2011.
- [2] O. Akbari et al., "Area Efficient Approximate 4–2 Compressor and Probability-Based Error Adjustment for Approximate Multiplier," IEEE Transactions on Circuits and Systems II, vol. 70, pp. 1714–1718, 2023.
- [3] F. Sabetzadeh, M. H. Moaiyeri and M. Ahmadinejad, "An Ultra-Efficient Approximate Multiplier with Error Compensation for Error-Resilient Applications," IEEE Transactions on Circuits and Systems II, vol. 70, no. 2, pp. 776–780, 2023

- [4] Z. Yang et al., “Efficient Approximate Multiplier Design Using Error Recovery Modules,” *Microprocessors and Microsystems*, vol. 98, 2023.
- [5] S. Vakili et al., “Fast and Low-Cost Approximate Multiplier for FPGAs Using Dynamic Reconfiguration,” *IEEE/ACM FPGA Conference*, 2023.
- [6] F. Baraati et al., “Efficient and High-Accuracy Approximate Multiplier Design for AI Applications,” *Circuits, Systems, and Signal Processing*, 2025.
- [7] S. Hashemi, R. I. Bahar and S. Reda, “DRUM: A Dynamic Range Unbiased Multiplier for Approximate Applications,” *Proceedings of the IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, pp. 418–425, 2015.
- [8] H. Jiang, C. Liu, L. Liu and J. Han, “Approximate Arithmetic Circuits: A Survey, Characterization, and Recent Applications,” *Proceedings of the IEEE*, vol. 108, no. 12, pp. 2105–2133, 2020.
- [9] J. Han and M. Orshansky, “Approximate Computing: An Emerging Paradigm for Energy-Efficient Design,” *IEEE European Test Symposium*, pp. 1–6, 2013.
- [10] R. Venkataramani, A. Sabne, V. Kozhikkottu, K. Roy and A. Raghunathan, “SALSA: Systematic Logic Synthesis of Approximate Circuits,” *IEEE Design & Test*, vol. 31, no. 6, pp. 38–47, 2014.
- [11] M. Shafique, W. Ahmad, R. Hafiz and J. Henkel, “A Low-Latency Generic Accuracy Configurable Adder,” *Proceedings of the IEEE Design Automation Conference (DAC)*, pp. 1–6, 2015.
- [12] A. Raha, H. Jayakumar and V. Raghunathan, “InputBased Dynamic Reconfiguration of Approximate Arithmetic Units,” *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 24, no. 3, pp. 846–857, 2016.
- [13] Z. Yang, A. Jain, J. Liang, J. Han and F. Lombardi, “Approximate XOR/XNOR-Based Adders for Inexact Computing,” *IEEE International Conference on Nanotechnology*, pp. 690–693, 2013.
- [14] P. Huang, B. Gong, K. Chen and C. Wang, “EnergyEfficient Neural Network Acceleration Using Most Significant Bit-Guided Approximate Multiplier,” *Electronics*, vol. 13, no. 15, 2024
- [15] A. Sadeghi, R. Rasheedi, I. Partin-Vaisband and D. Pal, “Energy Efficient Compact Approximate Multiplier for Error-Resilient Applications,” *IEEE Transactions on Circuits and Systems II: Express Briefs*, 2024
- [16] M. B. R. Srinivas and K. Elango, “Optimized NCFETBased Approximate Multiplier for Energy-Aware Image Applications,” *Scientific Reports*, vol. 15, 2025
- [17] K. P. Prashanth and N. K. V, “Area Efficient Approximate Multiplier Based on Novel 4:2 Compressors and Error Correction Logic,” *Scientific Reports*, vol. 16, pp. 1–15, 2026.
- [18] K. P. Prashanth and N. K. V, “Area Efficient Approximate Multiplier Based on Novel 4:2 Compressors and Error Correction Logic,” *Scientific Reports*, vol. 16, pp. 1–15, 2026.
- [19] V. Mrazek, Z. Vasicek, L. Sekanina, M. A. Hanif and M. Shafique, “EvoApprox8b: Library of Approximate Adders and Multipliers for Energy-Efficient Computation,” *IEEE Design & Test*, vol. 35, no. 5, pp. 40–49, 2018.
- [20] H. Jiang, J. Han and F. Lombardi, “Approximate Arithmetic Circuits: A Survey, Characterization, and Recent Advances,” *IEEE Transactions on Emerging Topics in Computing*, vol. 10, no. 1, pp. 1–15, 2022.