IJETIR

ISSN (Online) 2583-0554

International Journal of Emerging Technologies and Innovative Research (IJETIR)

Volume 6, Issue 4, April 2026

Impact Factor: 5.731

Machine Learning-Based Power Optimization Techniques for CMOS VLSI Circuits in Internet of Things Applications

Atava G V Karthik Raju[1] and Dr. Deepak Dalal[2]

[1]Research Scholar, Department of Electronics & Communication Engineering

[2]Assistant Professor, Department of Electronics & Communication Engineering

Mind Power University, Bhimtal, Nainital

Abstract: *The proliferation of Internet of Things (IoT) devices has created unprecedented demand for energy-efficient Very Large-Scale Integration (VLSI) circuits capable of delivering high performance within stringent power budgets. Traditional VLSI design methodologies, which rely on deterministic optimization rules, are increasingly inadequate for achieving the required balance between performance and power consumption in resource-constrained IoT applications. This paper presents a comprehensive examination of machine learning (ML)-based power optimization techniques for CMOS VLSI circuits in IoT systems. The review synthesizes recent advances across multiple abstraction levels—from device physics to algorithmic optimization—and evaluates the effectiveness of reinforcement learning, genetic algorithms, Bayesian optimization, and neural network-based modeling in addressing dynamic and static power challenges. Empirical evidence demonstrates that AI-directed optimization frameworks can achieve power savings of up to 43.3%, delay reduction of 29.7%, and energy savings of 52% compared to conventional VLSI systems. Furthermore, the integration of ML-driven adaptive power management, dynamic voltage and frequency scaling (DVFS), and intelligent power gating enables significant extension of device lifetime in energy-constrained IoT nodes. The paper concludes by identifying key challenges and future research directions for ML-driven VLSI power optimization.*

Keywords: Machine Learning, Power Optimization, CMOS VLSI, Internet of Things, Energy Efficiency, Design Space Exploration

1. INTRODUCTION

The convergence of Artificial Intelligence (AI) and Internet of Things (IoT) technologies has led to exponentially increasing computational demands at the network edge. IoT applications spanning healthcare, smart cities, agriculture, and renewable energy management require low-power, high-performance computing architectures that traditional centralized cloud systems cannot adequately provide due to latency, energy efficiency, and real-time processing constraints. This challenge has catalyzed significant research into AI-optimized VLSI architectures that can deliver intelligent, energy-efficient computation at the network edge.

CMOS technology remains the fundamental building block in VLSI system development due to its inherent low energy usage. However, as devices continue to scale according to Moore's law, power density, thermal challenges, and interconnect delays have become critical bottlenecks in hardware design. The total power dissipation in a CMOS circuit consists of dynamic power—consumed during switching activity—and static power, primarily caused by leakage currents. Dynamic power is expressed as:

$$P_{dyn} = C_L \cdot V_{dd}^2 \cdot \alpha \cdot f$$

where CL is load capacitance, Vdd is supply voltage, α is activity factor, and f is clock frequency. The quadratic relationship between supply voltage and dynamic power makes voltage scaling particularly effective for low-power design.

Traditional VLSI design methodologies, which rely on deterministic optimization rules, increasingly fail to achieve the dual objectives of high performance and low power consumption simultaneously. This limitation has prompted the integration of machine learning techniques into the design flow, enabling adaptive, intelligent circuit optimization paradigms. By employing ML algorithms such as reinforcement learning, neural network-based modeling, and genetic algorithms, designers can automatically optimize placement, routing, and synthesis parameters to minimize power consumption without compromising computational accuracy.

This paper provides a comprehensive review of ML-based power optimization techniques for CMOS VLSI circuits in IoT applications. The structure of the paper is as follows: Section 2 examines power consumption fundamentals in CMOS circuits. Section 3 explores machine learning approaches for VLSI design-space exploration. Section 4 analyzes adaptive power management techniques enabled by ML. Section 5 discusses emerging hardware architectures for energy-efficient IoT. Section 6 presents empirical results and performance comparisons. Section 7 concludes with challenges and future directions.

II. POWER CONSUMPTION FUNDAMENTALS IN CMOS VLSI CIRCUITS

2.1 Sources of Power Dissipation

Understanding the fundamental sources of power consumption is essential for effective optimization. Power dissipation in CMOS circuits comprises dynamic power, short-circuit power, and static power:

$$P_{total} = P_{dynamic} + P_{short} + P_{leakage}$$

Dynamic power is consumed during circuit switching and is dominated by the charging and discharging of load capacitance. Short-circuit power occurs when both pull-up and pull-down networks are momentarily ON due to finite input rise and fall times. Static power, consumed when the circuit is idle, is primarily caused by subthreshold and junction leakage currents:

$$P_{static} = I_{leakage} \cdot V_{dd}$$

The growing significance of leakage power in advanced technology nodes has made static power optimization increasingly critical.

2.2 Power Optimization Across Abstraction Levels

VLSI designers can intervene at various abstraction levels to achieve power optimization, with each level presenting distinct trade-offs between computational resources and result accuracy. The hierarchy spans from system-level strategies down to device-level techniques.

At the **algorithmic level**, power consumption depends on program code type, word length, modular interfaces, and technology implementation. Energy-efficient algorithms minimize the total number of operations and switching elements during execution. Recent advances combine heuristic algorithms—such as simulated annealing and genetic algorithms—with machine learning models to facilitate dynamic optimization of design parameters. Experimental studies on benchmark circuits (ISCAS85, ISCAS89, ITC99) have demonstrated that these algorithmic methods can achieve 25–30% power reduction compared to traditional design processes when implemented using industry-standard tools like Cadence Virtuoso and Synopsys Design Compiler.

III. MACHINE LEARNING APPROACHES FOR VLSI DESIGN-SPACE EXPLORATION

3.1 AI-Optimized Design Frameworks

The integration of AI and ML into VLSI design flows represents a paradigm shift from deterministic design methodologies to adaptive, intelligent optimization paradigms. Research demonstrates that AI-optimized VLSI architecture incorporating ML-based design-space exploration, adaptive power management, and energy harvesting schemes can achieve significant performance-per-watt advantages.

Key ML techniques applied to VLSI optimization include:

Reinforcement Learning (RL): Enables adaptive resource allocation and dynamic circuit optimization
Genetic Algorithms: Facilitate intelligent optimization of synthesis and layout parameters
Bayesian Optimization: Supports efficient exploration of power-performance-area trade-offs
Neural Network-Based Modeling: Enables accurate power and performance prediction
A comprehensive survey of AI/ML techniques in VLSI chip design automation categorizes existing research into three broad areas: design result estimation, design optimization and correction, and design construction. ML approaches have demonstrated particular effectiveness in automating complex design tasks, accelerating design cycles, and improving overall design quality.

3.2 Hardware-Software Co-Design

The effectiveness of ML-driven optimization depends significantly on hardware-software co-design principles. Recent work on AI-driven microprocessors for smart devices demonstrates that integrating ML accelerators, neuromorphic cores, and domain-specific AI processing units onto general-purpose microprocessors enables low-latency inference, adaptive learning, and superior energy efficiency. Prototype implementations show up to 10× latency reduction and 5× improvement in energy efficiency while maintaining competitive accuracy.

IV. ADAPTIVE POWER MANAGEMENT TECHNIQUES

4.1 Dynamic Voltage and Frequency Scaling (DVFS)

DVFS remains one of the most effective techniques for reducing dynamic power consumption. The quadratic relationship between supply voltage and dynamic power makes voltage scaling particularly impactful. Modern System-on-Chip (SoC) designs implement DVFS controllers for individual components, enabling fine-grained power optimization.
Innovative approaches combine DVFS with clock scheduling, where effective clock frequencies are achieved by periodically killing clock cycles of a master clock. By dithering between voltage levels and implementing clock scheduling, researchers have achieved results close to ideal DVFS systems. In Network-on-Chip (NoC) applications, implementing per-switch DVFS controllers has demonstrated average power savings of 33% in realistic video application scenarios.

4.2 AI-Enabled Power Gating

Power gating (PG) reduces static power by switching off power supply to unused circuit blocks via pMOS headers or nMOS footers. AI-based power gating extends conventional approaches by employing ML algorithms to predict idle periods and optimize gating decisions. Research demonstrates that partitioning buffers into banks and applying PG only to inactive banks can achieve approximately 40% leakage power savings without performance impact.

4.3 Near-Memory and In-Memory Computing

The growing memory wall challenge—where data movement dominates energy consumption—has motivated research into near-memory and in-memory computing architectures. AI-optimized VLSI architectures increasingly incorporate these approaches to minimize data movement energy.
Memristor-based Bayesian machines represent a notable advance in this direction. The logarithmic memristor-based Bayesian machine, fabricated in a hybrid CMOS/hafnium-oxide memristor process, demonstrates superior accuracy and energy efficiency compared to stochastic computing approaches for edge AI applications. This design converts probability multiplications into additions using near-memory integer adders, simplifying the computational model while effectively handling low-probability events.

V. EMERGING HARDWARE ARCHITECTURES FOR ENERGY-EFFICIENT IOT

5.1 Neuromorphic and Bio-Inspired Systems

Neuromorphic computing systems—inspired by biological neural processing—have demonstrated exceptional energy efficiency for AI workloads. Processors such as IBM's TrueNorth and Intel's Loihi achieve high power efficiency through event-driven computation, local memory access, and parallel processing. Similarly, Eyeriss, a spatial dataflow

architecture for convolutional neural networks, delivers up to 10× higher energy efficiency compared to conventional digital signal processors.

5.2 Hardware Accelerators for Edge Inference

The deployment of deep learning models on edge devices faces challenges of computational resources, power consumption, and memory bandwidth limitations. Hardware accelerators—including FPGAs, ASICs, and Neuromorphic Processing Units (NPUs)—enable efficient implementation of deep learning algorithms on edge platforms. Model compression techniques such as quantization, pruning, and knowledge distillation further reduce computational costs while maintaining accuracy.

A programmable and scalable bit-sliced VLSI architecture for decision tree-based machine learning inference demonstrates significant efficiency advantages over multilayer perceptrons, achieving 97–98% area and power savings. A 3-level decision tree implementation in 0.5µm CMOS technology occupies 90 mm² with 10.3 mW power consumption at 12.8 MHz.

5.3 ULTRA-LOW-POWER WAKE-UP CIRCUITS

Always-on smart sensors require ultra-low-power circuits for continuous monitoring. A recently demonstrated all-digital CMOS always-on smart sensor accelerator employing Hyperdimensional Computing (HDC) achieves extremely low power consumption of 39.21–92.59 µW in typical applications, consuming only 120.01 nJ/inference. The novel systolic similarity search unit and energy-efficient near-memory fused item memory + multiply-accumulate encoder enable this remarkable efficiency.

VI. EMPIRICAL RESULTS AND PERFORMANCE COMPARISONS

6.1 Quantitative Performance Gains

Simulation results using Cadence and Synopsys design tools demonstrate significant performance improvements from AI-optimized VLSI architectures:

Metric	Improvement
Power Saving	43.3%
Delay Reduction	29.7%
Energy Saving	52.0%

These results validate the effectiveness of ML-based optimization compared to conventional VLSI systems.

6.2 Neural Learning Unit Performance

A 16-bit floating-point Neural Learning Unit (NLU) fabricated in GlobalFoundries 22nm FDSOI technology demonstrates the feasibility of on-device learning for edge IoT applications. Key performance metrics include:

Metric	Value
Silicon Area	0.015 mm²
Power Consumption	0.379 mW
Training Speedup vs RISC-V-FPU	24.38×

IJETIR
ISSN (Online) 2583-0554
International Journal of Emerging Technologies and Innovative Research (IJETIR)
Volume 6, Issue 4, April 2026
Impact Factor: 5.731

Metric	Value
Energy Reduction vs RISC-V-FPU	37.37×
Energy Efficiency vs RISC-V-Vector	4.2× higher

The NLU achieves 89.34% training accuracy on keyword spotting tasks using only 40% of the training dataset, demonstrating practical feasibility for real-world edge learning applications.

6.3 Comparative Analysis

Research on AI/ML-driven VLSI system design shows that RL-based approaches, neural network-based circuit operation, and real-time fault detection capabilities enable significant improvements in power efficiency and system performance. These AI-enhanced designs are more adaptive, future-ready, and energy-efficient compared to traditional approaches, with reduced development time.

VII. CHALLENGES AND FUTURE DIRECTIONS

7.1 Current Challenges

Despite significant advances, several challenges remain in integrating ML into VLSI design flows:

Data Availability and Quality: ML models require large, high-quality datasets for training, which may not be available for specific design tasks.

Model Generalization: ML models trained on specific design scenarios may not generalize well to novel applications or technology nodes.

Integration into Existing EDA Workflows: Incorporating ML techniques into established Electronic Design Automation (EDA) tools presents technical and practical barriers.

Power-Accuracy Trade-offs: Approximate computing and aggressive power optimization may compromise computational accuracy, requiring careful balancing for safety-critical applications.

7.2 Future Research Directions

Continual On-Device Learning: Developing systems capable of adapting to new data without cloud retraining remains a key research priority.

Standard Design Tool Chains: Establishing standardized ML-based design tool chains and model portability frameworks would facilitate broader adoption.

Post-CMOS Technologies: Exploring quantum computing, neuromorphic systems, and emerging memory technologies such as memristors and STT-MTJ could revolutionize low-power VLSI design.

3D Integrated Circuits: 3D IC technology offers potential for significant power reduction through reduced interconnect length and improved integration density.

Sustainability and Green Electronics: The integration of energy harvesting circuits and AI-managed power distribution enables self-powered IoT nodes and zero-power standby operation, supporting clean-energy-powered IoT systems.

VIII. CONCLUSION

Machine learning-based power optimization techniques have emerged as a transformative approach for addressing the stringent energy requirements of CMOS VLSI circuits in IoT applications. By integrating AI-driven design-space exploration, adaptive power management, and intelligent resource allocation, these techniques enable significant performance-per-watt improvements that traditional deterministic design methodologies cannot achieve.

Empirical evidence demonstrates that AI-optimized VLSI architectures can achieve power savings of 43.3% and energy savings of 52% compared to conventional approaches. Specialized neural learning units fabricated in advanced CMOS technologies show remarkable efficiency gains, with 24× training speedup and 37× energy reduction.

As IoT applications continue to proliferate across healthcare, smart cities, autonomous systems, and industrial automation, the demand for energy-efficient computational platforms will only intensify. Continued research into ML-driven optimization, emerging memory technologies, neuromorphic computing, and sustainable design practices will be essential for realizing the next generation of intelligent, energy-efficient edge computing systems.

REFERENCES

[1] AI-Optimized VLSI Architecture for Energy-Efficient and Sustainable IoT Systems. *IEEE Xplore*, 2026.
[2] NLU: An Adaptive, Small-Footprint, Low-Power Neural Learning Unit for Edge and IoT Applications. *IEEE Xplore*, 2025.
[3] A comprehensive review of low-power design techniques in VLSI systems: From device physics to algorithmic optimization. *Springer*, 2026.
[4] Karaosmanoğlu, M. F. Design of Energy-Efficient CMOS Circuits for AI Applications. *Human Computer Interaction*, 2024.
[5] Innovative VLSI System Design and Embedded Architectures Empowered by AI and Machine Learning Advancements. *IEEE Xplore*, 2025.
[6] Constructing AI-Driven Microprocessors for Smart Devices: Integrating On-Chip Intelligence to Boost Cutting-Edge Decision-Making Performance. *IEEE ICCCA*, 2025.
[7] Sabarad, J. M. AI/ML for VLSI chip design automation: a comprehensive survey. *Dimensions for DTIC*, 2025.
[8] The logarithmic memristor-based Bayesian machine. *Communications Engineering (Springer)*, 2025.
[9] Somesula, R. S. Programmable and Scalable Bit-Sliced VLSI Architecture for Decision Tree Based Machine Learning Edge Inference. *USF Tampa Graduate Theses and Dissertations*, 2024.
[10] A Extensive Review of Recent Technologies and Trends in Low-Power VLSI Design. *R Discovery*, 2024.
[11] A 40 µW 8-bit Accelerator Wake-Up Circuit for Always-on Smart IoT Sensor Monitoring. *IEEE Xplore*, 2026.
[12] Implementing Deep Learning Models Efficiently in Edge Devices and Internet of Things Applications via VLSI-Driven Embedded Systems. *IEEE Xplore*, 2026.