Proceedings of the 2nd International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/96/20240968

Mutual advancement: How integrated circuits interact with artificial intelligence

Linye Wang

Beijing Institute of Technology, Beijing, 100081, China

1120220723@bit.edu.cn

Abstract. Integrated circuits and artificial intelligence complement each other. On one hand, the continuous optimization of electron and photon accelerators can reduce the energy consumption of artificial intelligence algorithms and adapt them to edge computing, thereby promoting the development of artificial intelligence. On the other hand, artificial intelligence not only assists in troubleshooting and optimizing of integrated circuits, but also automates the design of integrated circuits, making them more convenient and efficient. The interactive relationship between integrated circuits and artificial intelligence is presented through qualitative analysis and literature review in this paper. The mutual influence of these two technologies creates new breakthroughs and possibilities for the development of science and technology. However, while photonic artificial intelligence accelerators are suitable for high-speed neural networks, their current development is not yet mature due to the high cost, complex manufacturing process, and environmental disturbances. Additionally, artificial intelligence algorithms for integrated circuits design lack sufficient high-quality data for training due to the secrecy of integrated circuit parameters. Thus it requires a long time to train and perfect these algorithms.

Keywords: Integrated Circuits, Artificial Intelligence, Interaction, Deep-Learning.

1. Introduction

Artificial intelligence (AI), including machine learning, deep learning, natural language processing, and computer vision, simulates human cognitive functions, while integrated circuit is crucial in electronic devices, enabling them to become more compact and efficient [1]. They significantly contribute to the progress of science and technology. AI and integrated circuits are closely interconnected, impacting various sectors such as healthcare, data monitoring, and human-computer interaction. And for technological advancement, they introduce novel concepts and create opportunities for significant breakthroughs [2]. The development of integrated circuits and AI circuits mutually complement each other. AI leverages algorithms and data processing to achieve intelligent decision-making and learning, which rely on efficient computing and data processing capabilities [3]. And the high performance and low power consumption of integrated circuits play a vital role in supporting the learning and decision-making processes of AI. On the other hand, the complexity of integrated circuits, with their numerous functional units and electronic components, presents challenges in circuit design [4]. The learning and decision-making capabilities of AI address these challenges, leading to more intelligent and efficient circuit design.

DOI: 10.54254/2755-2721/96/20240968

The paper explores the interplay between integrated circuits and AI, examining the essential role of integrated circuits in AI development and the utilization of AI in integrated circuit design. The close interdependence of these two technologies is underscored, highlighting prospects for deeper integration in future research. Moreover, it seeks to raise awareness about the convergence of these two domains.

2. Interaction Between AI and Integrated Circuit

2.1. AI-Accelerated Chip

The current training of AI models relies on employing algorithms to process and interpret data, enabling the model to learn and adapt to a variety of tasks. However, conventional central processing units (CPUs) are unable to meet the demands for computing resources and time, leading to the emergence of AI chips. As the functional requirements for AI increase, the scale of data grows and algorithms become more intricate. The exceptionally high computing power requirements constrain the advancement of AI. Consequently, novel AI acceleration chips are needed to enhance processing speed and overcome prevailing hardware limitations.

In the existing research, the new chip architecture is mainly aimed at edge computing and power-limited AI application scenarios. The main methods for reducing energy consumption and examples are shown in Figure 1. The challenges in edge AI inference and training are addressed by introducing CHIMERA, a DNN chip optimized for resistive RAM (RRAM) with on-chip weight storage [5]. It achieves peak performance of 0.92-TOPS (trillions (10^{12}) of operations per second) and energy efficiency of 2.2-TOPS/W, overcoming the limitations of current memory technologies at the edge. Some digital AI accelerators featuring low precision, including the UNPU accelerator, the binary neural network (BNN) accelerator presented by Intel, and the 7-nm AI chip presented by IBM, are able to obtain the optimal energy-accuracy trade-off to reduce power consumption [6-8]. Another effective way is to concatenate compute and memory and perform computation in the analog domain, such as resistive and capacitive SRAM in-memory computing (IMC) [9, 10]. This solves the bottleneck that on-chip SRAM's data access energy accounts for more than two-thirds of the total energy due to the requirement of large amounts of memory access and data communication.

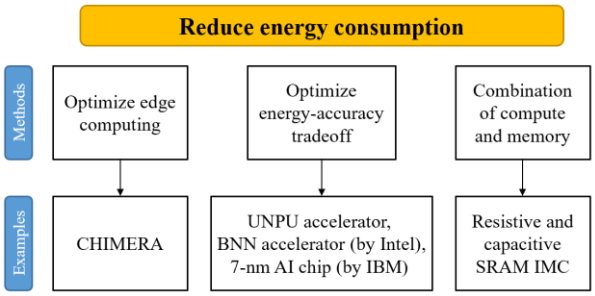

Figure 1. Methods for reducing energy consumption

In addition to traditional electronic integrated circuits, photonic acceleration chips are now more widely used. A fully optical neural network (ONN) is proposed, which leverages high-detection-rate and high-sensitivity photon detectors to enable rapid and energy-efficient neural networks [11]. Also, a universal optical vector convolutional accelerator was proposed that operates at more than ten-TOPS [12]. Such an accelerator generates convolutions of images with 250,000 pixels, which is large enough for facial image recognition and can be trained for more complex networks. In addition, a scalable chiplet-based DNN accelerator with photonic interconnects (ASCEND) is also presented, which introduces a novel photonic network and tailored dataflow with 71% and 67% reduction in execution time and energy consumption, respectively [13-15]. These photonic circuits are highly suitable for the implementation of advanced neural networks due to their exceptional interconnectivity and linear computational abilities.

DOI: 10.54254/2755-2721/96/20240968

Overall, the advancement of AI is primarily constrained by computing resources and speed. Continuous improvements in AI acceleration chips are crucial for reducing energy consumption and enabling AI to deploy intricate algorithms for training and learning from extensive datasets. This progress has facilitated the development of AI, paving the way for overcoming the impediments attributable to limited computing power.

2.2. AI in Integrated Circuit

While the design, troubleshooting, and parameter selection of integrated circuits are hampered by numerous challenges, the advanced learning and computing capabilities of AI are currently a crucial breakthrough in addressing these complexities, promising to enhance design efficiency and lower costs.

AI has been employed to support circuit design, specifically in parameter optimization and troubleshooting. Integrated circuits consist of various parameters, such as process and device parameters, making it impossible to optimize all parameters simultaneously. Therefore, the optimization requirements for integrated circuits differ based on their specific needs. Several optimization targets, including maximizing electrical-thermal performance, optimizing TSV array and decoupling capacitor design, and improving the performance of chaogate circuit and thermal placement, are outlined in Table 1.

One or more AI algorithms used by researchers have a significant impact on parameter optimization. For instance, utilizing machine learning with Bayesian optimization algorithm enhanced temperature gradient, CPU time, and skew by 4.4%, 31.1% and 6.9%, respectively. Also, EO has an 18.2% increase by using the optimal method with reinforcement learning, while coverage of state space increases for 10^{12}~10^{13} order with the similar method. There is a 47% and 12% improvement in □□ and □□ with the Genetic algorithm and an 87% decrease in runtime with the method merging Green's function and Bayesian optimization as well. All these examples indicate that AI enhances parameter optimization efficiency.

Table 1. Literature summary of parameter optimization.

Optimization objective	Optimization parameters	Optimization methods	Results
Maximize the electrical – thermal performance [16]	Multi-parameters	Machine learning combined with Bayesian Optimization	4.4%, 31.1%, and 6.9% improvement in temperature gradient, CPU time, and skew
Optimizing through silicon via (TSV) array design in high-bandwidth memory (HBM) [17]	Eye opening(EO)	Policy-based Reinforcement-Learning (RL)	18.2% increase in EO
Optimize parameters of the chaogate circuit [18]	Lyapunov exponent (λ) and entropy (H)	Bayesian Optimization Genetic Algorithm	47%, 12% improvement in λ_m and H_m
Propose optimal decoupling capacitor design method [19]	The coverage of the state space	Deep RL based optimal decap design method	Approximately 10^{12}~ 10^{13} order increase in coverage
Improve thermal placement optimization of multiple chips on silicon interposer [20]	Achieve lowest maximum temperature in shortest runtime (t)	Rapid optimization method that merges Green's function and Bayesian optimization	T_{max} difference not exceeding 0.1 degree and 87% decrease in runtime

Moreover, due to the intricate nature of integrated circuits, troubleshooting can be highly labor-intensive and costly. However, the incorporation of AI algorithms simplifies this task. An effective

DOI: 10.54254/2755-2721/96/20240968

thermal image processing technique to identify defective ICs on PCBs is proposed by Pittu et al., which achieves a 97 percent accuracy in detecting faulty and non-faulty images with minimal computational time while alerts about circuit fault conditions to users [14]. Yang, Meng and Wang present a fault diagnosis approach for analog circuits using a compact 1D-CNN model, which enhances the classification accuracy and ensures resistance to interference, thereby making it well-suited for real-time fault detection and monitoring. These suggest that automatic fault removal in ICs is highly possible in the future.

In addition to assisted design, there is an increasing focus on using AI to automate integrated circuit design. The intricate and time-consuming nature of integrated circuit design necessitates significant investments in manpower, materials, and financial resources. Any design errors can lead to considerable losses. Therefore, automating integrated circuit design systems, such as automated RL framework (AutoCkt) and supervised pre-training approach, can reduce human errors, eliminate repetitive tasks, reduce costs, and handle complex designs and rapid iterations [21-22].

3. Discussion

Nowadays, both AI and integrated circuits are advancing rapidly. While electronic AI-accelerated chips are more mature, the application of photonics still faces numerous challenges. Optical technology has proven to offer significant advantages in computational speed, data parallelization, and power consumption, making it well-suited for high-performance implementation of neural networks. However, the high cost and complexity of manufacturing photonic integrated circuits pose a major obstacle. Furthermore, environmental light interference and the absence of optical random access memory also restrict the development of photonic accelerated chips [23]. But high-speed and low-power photoelectric conversion remain increasingly crucial during this transitional phase. Regarding integrated circuit design, incorporating AI brings substantial benefits, yet obtaining well-trained AI algorithmic models poses a primary challenge. The training process requires a massive amount of data, which is often highly confidential in IC design, resulting in insufficient quantity and quality for AI training. Additionally, the high accuracy required in IC design places significant computational and processing demands on AI. Therefore, refining AI algorithms may necessitate extended time and ongoing updates to achieve superior IC designs.

Despite the numerous obstacles in the development and integration of AI and integrated circuits, the potential for both fields is undeniable. Both technologies are widely applied in areas such as autonomous vehicles, medical imaging diagnosis, and smart homes. The ongoing advancement of photonic computing also offers opportunities to overcome the constraints of electronic integrated circuit development. It is foreseeable that as technology continues to progress, the collaborative evolution of AI and integrated circuits will lead to breakthroughs and innovation in numerous other fields.

4. Conclusion

The paper explores the dynamic interaction between integrated circuits and AI. Through the ongoing optimization of electron and photon accelerators, it is possible to reduce the energy consumption of AI algorithms, enabling their adaptation to edge computing and fostering AI development. AI contributes to troubleshooting, optimization, and automated design of integrated circuits, but has challenges including high costs, complex manufacturing processes, and inadequate high-quality data for training AI algorithms. Despite these difficulties, both AI and integrated circuits exhibit promising potential for future scientific and technological progress.

Although this article primarily explores the potential for AI engagement with integrated circuits in scientific research, it lacks specific elaboration on details such as optical computing acceleration methods (e.g., matrix transformation) and the categorization of AI learning algorithms along with their respective training techniques. Therefore, the author looks forward to gaining further insights into specific approaches for utilizing AI in integrated circuit design and advocating for the concurrent advancement of both fields through the application of AI in circuit design.

DOI: 10.54254/2755-2721/96/20240968

References

[1] Pavlidis, V. F., Savidis, I., & Friedman, E. G. (2017). Three-dimensional integrated circuit design. Newnes.

[2] Lambrechts, J. W., Sinha, S., Sengupta, K., Bimana, A., Kadam, S., Bhandari, S., Preez, J. D., Shao,Z., Huang, X., Liu, Z., Karahan, E. A., Blundo, T., Allam, M., Ghozzy, S., Zhou, J., Fang, W., & Valliarampath, J. (2024). Intelligent Integrated Circuits and Systems for 5G/6G Telecommunications. IEEE Access.

[3] Nolan, A. (2024). Artificial intelligence in science: Challenges, opportunities and the future of research.

[4] Hu, Y., Li, J., Klemme, F., Nam, G. J., Ma, T., Amrouch, H., & Xiong, J. (2024). SyncTREE: Fast Timing Analysis for Integrated Circuit Design through a Physics-informed Tree-based Graph Neural Network. Advances in Neural Information Processing Systems, 36.

[5] Prabhu, K., Gural, A., Khan, Z. F., Radway, R. M., Giordano, M., Koul, K., Doshi, R., Kustin, J. W., Liu, T., Lopes, G. B., Turbiner, V., Khwa, W., Chih, Y., Chang, M., Lallement, G., Murmann, B., Mitra, S. & Raina, P. (2022). CHIMERA: A 0.92-TOPS, 2.2-TOPS/W edge AI accelerator with 2-MByte on-chip foundry resistive RAM for efficient training and inference. IEEE Journal of Solid-State Circuits, 57(4), 1013-1026..

[6] Lee, J., Kim, C., Kang, S., Shin, D., Kim, S., & Yoo, H. J. (2018). UNPU: An energy-efficient deep neural network accelerator with fully variable weight bit precision. IEEE Journal of Solid-State Circuits, 54(1), 173-185.

[7] Knag, P. C., Chen, G. K., Sumbul, H. E., Kumar, R., Hsu, S. K., Agarwal, A., Kar, M., Kim, S., Anders, M. A., Kaul, H., & Krishnamurthy, R. K. (2020). A 617-TOPS/W all-digital binary neural network accelerator in 10-nm FinFET CMOS. IEEE journal of solid-state circuits, 56(4), 1082-1092.

[8] Agrawal, A., Lee, S. K., Silberman, J., Ziegler, M., Kang, M., Venkataramani, S., Cao, N., Fleischer, B., Guillorn, M.l., Cohen, M., Mueller, S., Oh, J., Lutz, M., Jung, J., Koswatta, S., Zhou, C., Zalani, V., Bonanno, J., Casatuta, R., Chen, C., Choi, J., Haynie, H., Herbert, A., Jain, R., Kar, M., Kim, K., Li, Y., Ren, Z., Rider, S., Schaal, M., Schelm, K., Scheuermann, M., Sun, X., Tran, H., Wang, N., Wang, W., Zhang, X., Shah, V., Curran, B., Srinivasan, V., Lu, P., Shukla, S., Chang, L., & Gopalakrishnan, K. (2021, February). 9.1 A 7nm 4-core AI chip with 25.6 TFLOPS hybrid FP8 training, 102.4 TOPS INT4 inference and workload-aware throttling. In 2021 IEEE International Solid-State Circuits Conference (ISSCC) (Vol. 64, pp. 144-146). IEEE.

[9] Shanbhag, N. R., & Roy, S. K. (2022, April). Comprehending in-memory computing trends via proper benchmarking. In 2022 IEEE Custom Integrated Circuits Conference (CICC) (pp. 01-07). IEEE.

[10] Mingtao Zhan, Jiaxin Liu, Jian-Wei Su, Yen-Lin Chung, Yue, J., Feng, X., He, Y., Huang, Y., Wang, Y., Yuan, Z., Zhan, M., Liu, J., Su, J., Chung, Y., Wu, P., Hung, L., Chang, M., Sun, N., Li, X., Yang, H., & Liu, Y. (2021, February). 15.2 A 2.75-to-75.9 TOPS/W computing-in-memory NN processor supporting set-associate block-wise zero skipping and ping-pong CIM with simultaneous computation and weight updating. In 2021 IEEE International Solid-State Circuits Conference (ISSCC) (Vol. 64, pp. 238-240). IEEE.

[11] Shen, Y., Harris, N. C., Skirlo, S., Prabhu, M., Baehr-Jones, T., Hochberg, M., Sun, X., Zhao, S., Larochelle, H., Englund, D., & Soljačić, M. (2017). Deep learning with coherent nanophotonic circuits. Nature photonics, 11(7), 441-446.

[12] Xu, X., Tan, M., Corcoran, B., Wu, J., Boes, A., Nguyen, T. G., Chu, S. T., Little, B. E., Hicks, D. G., Morandotti, R., Mitchell, A., & Moss, D. J. (2021). 11 TOPS photonic convolutional accelerator for optical neural networks. Nature, 589(7840), 44-51.

[13] Li, Y., Wang, K., Zheng, H., Louri, A., & Karanth, A. (2022). Ascend: A scalable and energy-efficient deep neural network accelerator with photonic interconnects. IEEE Transactions on Circuits and Systems I: Regular Papers, 69(7), 2730-2741.

DOI: 10.54254/2755-2721/96/20240968

[14] Pittu, P. S. K. R., Sankaran, V., Alagu Mariappan, P., Pramod, G., Nikita, & Sharma, Y. (2023). Design of Artificial Intelligence-Based Novel Device for Fault Diagnosis of Integrated Circuits. Engineering Proceedings, 58(1), 77.

[15] Yang, H., Meng, C., & Wang, C. (2020). Data-driven feature extraction for analog circuit fault diagnosis using 1-D convolutional neural network. IEEE Access, 8, 18305-18315.

[16] Park, S. J., Bae, B., Kim, J., & Swaminathan, M. (2017). Application of machine learning for optimization of 3-D integrated circuits and systems. IEEE transactions on very large scale integration (VLSI) systems, 25(6), 1856-1865.

[17] Kim, K., Park, H., Kim, S., Kim, Y., Son, K., Lho, D., Son, K., Shin, T., Sim, B., Park, J., Park, S., & Kim, J. (2024). Policy-Based Reinforcement Learning for Through Silicon Via Array Design in High-Bandwidth Memory Considering Signal Integrity. IEEE Transactions on Electromagnetic Compatibility.

[18] Acharya, R. Y., Charlot, N. F., Alam, M. M., Ganji, F., Gauthier, D., & Forte, D. (2021, April). Chaogate parameter optimization using bayesian optimization and genetic algorithm. In 2021 22nd International Symposium on Quality Electronic Design (ISQED) (pp. 426-431). IEEE.

[19] Park, H., Park, J., Kim, S., Cho, K., Lho, D., Jeong, S., Park, S., Park, G., Sim, B., Kim, S., Kim, Y., & Kim, J. (2020). Deep reinforcement learning-based optimal decoupling capacitor design method for silicon interposer-based 2.5-D/3-D ICs. IEEE Transactions on Components, Packaging and Manufacturing Technology, 10(3), 467-478.

[20] Xiao, C., Zheng, W., Tian, Q., Rao, X., & Zhang, H. (2024). An efficient Green's function-based Bayesian optimization method for the thermal optimization of multi-chips on a silicon interposer. International Communications in Heat and Mass Transfer, 153, 107379.

[21] Settaluri, K., Liu, Z., Khurana, R., Mirhaj, A., Jain, R., & Nikolic, B. (2021). Automated design of analog circuits using reinforcement learning. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 41(9), 2794-2807.

[22] Hakhamaneshi, K., Nassar, M., Phielipp, M., Abbeel, P., & Stojanovic, V. (2022). Pretraining graph neural networks for few-shot analog circuit modeling and design. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems.

[23] Wu, J., Lin, X., Guo, Y., Liu, J., Fang, L., Jiao, S., & Dai, Q. (2022). Analog optical computing for artificial intelligence. Engineering, 10, 133-145.