



A novel design of high throughput power efficient multiply accumulate unit

M Jayasanthi¹ · R Kalaivani² · Kilvani Palanisamy Sampooram³ · Vijay Nath⁴ 

Received: 9 April 2025 / Revised: 9 April 2025 / Accepted: 25 July 2026 / Published online: 20 August 2026
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2026

Abstract

The demand for high-performance hardware accelerators is growing, which are widely used nowadays for implementation of Deep Learning methods like that of Convolutional Neural Networks (ConvNet / CNN) on reconfigurable devices. The Multiply-Accumulate (MAC) units are the warhorses behind the remarkable speed and efficiency of convolution operations. Double MAC approach combines two units of MAC operations into a single Digital Signal Processing module of a reconfigurable Field Programmable Gate Arrays (FPGA), thereby increasing the throughput of Convolutional Neural Network accelerators. By dynamically adjusting the precision of MAC operations to 4-bit, 8-bit, and 16-bit fixed and floating-point representations, MAC units can be optimized for efficient resource utilization and power consumption. In this paper, a Double multiply-accumulate unit, dynamically changing precision with 8-bit and 16-bit, is proposed. By leveraging the inherent parallelism and pipelining capabilities of Field Programmable Gate Arrays (FPGA), Double multiply-accumulate approach optimizes the utilization of hardware resources while minimizing the power consumption. The flexible precision scaling in the proposed work proves advantageous in convolution layers, where greater precision is required for final classification and smaller precision in initial stage. Simulation, Synthesis and FPGA Implementation of the proposed Double MAC unit with dynamic precision scaling have been done and it is found successful. The proposed Double MAC approach has showed twofold improvement in throughput and 15% improvement in power consumption. These techniques are highly useful for fast data processing in wireless communication.

Keywords Field Programmable Gate Arrays · Convolutional Neural Networks · Digital Signal Processing · Multiply-and-Accumulate · Single-Instruction Multiple-Data · Fixed point and floating-point unit

1 Introduction

Since the deep learning techniques become more prominent, there is a growing requirement for accelerators and their hardware implementation. Convolutional neural networks are subsystems in the deep neural networks that have several advantages over other types of networks when it comes to hardware acceleration. These advantages include high computational complexity, high structural regularity, good recognition performance and broad applicability. FPGAs are among the most widely used systems due to their incredibly high levels of flexibility and parallelism. Consequently, a great deal of effort has gone into creating better CNN accelerators on FPGAs.

Generally, Deep Neural Network (DNN) accelerators operate on fixed precision arithmetic and typically 16-bit or 32-bit floating-point representations. While this ensures accuracy, it does not exploit the potential benefits of lower

✉ Vijay Nath
vijaynath@bitmesra.ac.in

M Jayasanthi
jayasanthiranjith@gmail.com

R Kalaivani
kalaitss72@gmail.com

Kilvani Palanisamy Sampooram
drkpsampooram1970@gmail.com

¹ PSG Institute of Technology and Applied Research, Coimbatore, Tamilnadu, India

² Erode Sengunthar Engineering College Erode, Erode, Tamilnadu, India

³ Bannari Amman Institute of Technology, Erode, Tamilnadu, India

⁴ Birla Institute of Technology, Mesra, Ranchi, Jharkhand, India

Table 3 Comparison of run time of convolution operations with fixed and dynamic precision

Time in ns	8-bit floating point	16-bit floating point	Single MAC unit with dynamic precision	Floating point with dynamic precision & Double MAC (Proposed)
MAC operation(2)	0.26	0.50 [31]	0.38	0.19
Convolution operation	7.15	13.8	10.45	5.23

4.4 Performance analysis with convolution operation

Table 3 compares the run time of convolution operations using proposed MAC architecture and conventional ones, comparing 8-bit floating point operations, 16-bit floating point operations, single MAC with dynamic precision and double MAC with Dynamic Precision units. For a filter of size 3×3 with a single channel input, an output feature map of size 5×5 , one output channel, the run time for the convolution operations are shown in Table 3 for each case.

The above results demonstrate that the total power consumption and throughput has been improved using the proposed method at the cost of increase in LUTs, Flops and IOs. The above table describes the total runtime for all the 55 convolution operations. It can be observed that there is a twofold improvement in throughput of the MAC unit and convolution operations.

5 Conclusion

A method of increasing the computational rate and reducing the power consumption in the MAC unit is proposed in this paper. Convolutional Neural Networks pack multiple MAC operations into one DSP block of FPGAs. The proposed approach can accomplish a good sense of balance amid power consumption and throughput with reusability in implementation. The approach has been validated through Verilog simulation and Implementation using EDA tools. Since machine learning or deep learning algorithms like Convolutional Neural Networks (CNNs) are highly demanding technologies nowadays, the proposed approach has a wider scope in its area. Many existing applications use machine learning that either consume more energy or has high processing speed. In the proposed Double MAC approach, the power consumption as well as throughput have been improved which is considerably efficient than the existing Single MAC and Double MAC approaches. This approach can be incorporated to applications that require algorithms, which shall process data faster with variable precision and should consume lesser energy. This method is highly useful for wireless communication systems for fast data processing.

Author contributions M. Jayasanthi prepared the draft article and did the required simulation. Dr. R. Kalaivani and Dr KP Sampooram cross check the article and updated the results and verified. Dr. Vijay Nath read the complete article did the required correction and submitted the article to journal and have taken responsibility of further updates article of journal reviews and corrections.

Funding There is no funding support.

Data availability No datasets were generated or analysed during the current study.

Declarations

Competing interests The authors declare no competing interests.

References

- Chen, Z. (2015). et.al. 'Optimizing FPGA-based accelerator design for deep convolutional neural networks' In *Proceedings of the 2015 ACM/SIGDA international symposium on field-programmable gate arrays*, pp. 161–170. <https://doi.org/10.1145/2684746.2689060>
- Jiantao, Q. (2016). et.al. 'Going deeper with embedded FPGA platform for convolutional neural network', In *Proceedings of the 2016 ACM/SIGDA international symposium on field-programmable gate arrays*, pp. 26–35. <https://doi.org/10.1145/2847263.2847265>
- Atul, R. (2017). et.al. 'Design space exploration of FPGA accelerators for convolutional neural networks', In *Design, Automation & Test in Europe Conference & Exhibition (DATE)*, pp. 1147–1152. <https://doi.org/10.23919/DATE.2017.7927162>
- Ramana, K., et al. (2023). Modified CNN accelerator using double MAC design. *AIP Conference Proceedings*, 2492(1). <https://doi.org/10.1063/5.0118186>
- Alex, K. et.al (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6), 84–90. <https://doi.org/10.1145/3065386>
- Yunji, C. (2014). et.al. 'Dadiannao: A machine-learning supercomputer' In *2014 47th Annual IEEE/ACM International Symposium on Microarchitecture*, pp. 609–622. <https://doi.org/10.1109/MICRO.2014.58>
- Yufei, M. et.al (2018). Optimizing the convolution operation to accelerate deep neural networks on FPGA. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 26(7), 354–1367. <https://doi.org/10.1109/TVLSI.2018.2815603>
- Julian, F. et.al (2019). AddNet: Deep neural networks using FPGA-optimized multipliers. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 28(1), 115–128. <https://doi.org/10.1109/TVLSI.2019.2939429>
- Fan, S. et.al (2017). 'A high-performance accelerator for large-scale convolutional neural networks', In *2017 IEEE International Symposium on Parallel and Distributed Processing with Applications*

- and 2017 IEEE International Conference on Ubiquitous Computing and Communications (ISPA/IUCC), pp. 622–629. <https://doi.org/10.1109/ISPA/IUCC.2017.00099>
10. Abdelilah, H., et al. (2014). Implementing Convolutional Neural Networks on FPGA: A Survey and Research. *ITM Web of Conferences*, 52(1). <https://doi.org/10.1051/itmconf/20235202004>
 11. Shah, N., et al. (2018). Runtime programmable and memory bandwidth optimized FPGA-based coprocessor for deep convolutional neural network. *IEEE Transactions on Neural Networks and Learning Systems*, 29(1), 5922–5934. <http://eprints.iisc.ac.in/id/eprint/61218>
 12. Marco, B. (2017). et.al. ‘A convolutional neural network fully implemented on FPGA for embedded platforms’, In *2017 New Generation of CAS (NGCAS)*, pp. 49–52. <https://doi.org/10.1109/NGCAS.2017.16>
 13. Chaoyang, Z., et al. (2020). An efficient hardware accelerator for structured sparse convolutional neural networks on FPGAs. *IEEE Transactions on Very Large-Scale Integration (VLSI) Systems*, 28(9), 1953–1965. <https://doi.org/10.1109/TVLSI.2020.3002779>
 14. Nguyen, D. (2017). Doubling the performance of convolutional neural networks on modern FPGAs. *Design Automation & Test in Europe Conference & Exhibition (DATE)*, 890–893. <https://doi.org/10.23919/DATE.2017.7927113>
 15. Chi, Z. (2017). et.al. ‘Frequency domain acceleration of convolutional neural networks on CPU-FPGA shared memory system’, *Proceedings of the 2017 ACM/SIGDA International Symposium on Field-Programmable Gate Arrays*. <https://doi.org/10.1145/302007.8.3021727>
 16. Liu, L. (2020). ‘Duet: Boosting deep neural network efficiency on dual-module architecture’, In *2020 53rd Annual IEEE/ACM International Symposium on Microarchitecture (MICRO)*, pp. 738–750. <https://doi.org/10.1109/MICRO50266.2020.00066>
 17. Ma, Y. (2017). An automatic RTL compiler for high-throughput FPGA implementation of diverse deep convolutional neural networks. *2017 27th International Conference on Field Programmable Logic and Applications (FPL)*. <https://doi.org/10.23919/FPL.2017.8056824>
 18. Yizhi, W., et al. (2018). FPAP: A folded architecture for energy-quality scalable convolutional neural networks. *IEEE Transactions on Circuits and Systems I*, 66(1), 288–301. <https://doi.org/10.1109/TCSI.2018.2856624>
 19. Huynh, V., et al. (2021). ‘FPGA-based acceleration for convolutional neural networks on PYNQ-Z2’. *International Journal of Computing and Digital System*. <https://doi.org/10.12785/ijcds/110136>
 20. Mário, V., et al. (2022). Efficient design of low bit width convolutional neural networks on FPGA with optimized dot product units. *ACM Transactions on Reconfigurable Technology and Systems*, 16(1), 1–36. <https://doi.org/10.1145/3546182>
 21. Seyedramin, R., et al. (2023). *Rethinking FPGA Architectures for Deep Neural Network applications*. PhD diss.
 22. Stylianos, I. V., et al. (2018). Tool flows for mapping convolutional neural networks on FPGAs: A survey and future directions. *ACM Computing Surveys (CSUR)*, 51(3), 1–39. <https://doi.org/10.1145/3186332>
 23. Ilja, I. (2019). et.al. ‘Implementation Analysis of Convolutional Neural Networks on FPGAs’, *Diss. Master Thesis, University of Technology Eindhoven*.
 24. Heshan, Z. (2019). et.al. ‘A block-floating-point arithmetic-based FPGA accelerator for convolutional neural networks’, In *2019 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 1–5.
 25. Yangqing, J. (2014). ‘Caffe: Convolutional architecture for fast feature embedding’, In *Proceedings of the 22nd ACM international conference on Multimedia*, pp. 675–678. <https://doi.org/10.48550/arXiv.1408.5093>
 26. Clément, F. (2009). et.al. ‘CNP: An FPGA-based processor for convolutional networks’, In *2009 International Conference on Field Programmable Logic and Applications*, pp. 32–37. <https://doi.org/10.1109/FPL.2009.5272559>
 27. Jayasanthi, M. (2023). Implementation of Power Efficient Multiply Accumulate Unit for DSP Applications. *2023 International Conference on Intelligent Technologies for Sustainable Electric and Communications Systems (iTech SECOM)*, 546–551. <https://doi.org/10.1109/iTechSECOM59882.2023.10435014>
 28. Atul, R. (2016). et.al. ‘Efficient FPGA acceleration of convolutional neural networks using logical-3D compute array’, *2016 Design, Automation & Test in Europe Conference & Exhibition (DATE)*.
 29. Krishnan, V. (2019). et.al. ‘A comparative study on the performance of FPGA implementations of high-speed single-precision binary floating-point multipliers’, In *2019 International Conference on Smart Systems and Inventive Technology (ICSSIT)* (pp. 1041–1045).
 30. Arish, S. (2015). et.al. ‘Run-time reconfigurable multi-precision floating point multiplier design for high speed, low-power applications’, In *2015 2nd International Conference on Signal Processing and Integrated Networks (SPIN)* (pp. 902–907).
 31. Zhang, H., et al. (2019). New flexible multiple-precision multiply-accumulate unit for deep neural network training and inference. *IEEE Transactions on Computers*, 69(1), 26–38.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.



M Jayasanthi received PhD from Anna University and ME in VLSI Design in Government college of Technology, Coimbatore, Tamilnadu, India. She is currently working as Professor at PSG Institute of Technology and Applied Research, Tamilnadu. She is an active member of ISTE, IE, IAENG, and IEEE. Her current research interests include ASIC Design, Low power IC design and mixed signal IC design and Machine Learning.



Kalaivani R is a Professor in the Department of Electronics and Communication Engineering (ECE), Erode Sengunthar Engineering College, Tamil Nadu, India. She has over 24 years of teaching and research experience and completed her Ph.D. in Engineering in 2015. Her research interests include VLSI, Wireless Communication, Signal Processing, IoT, and emerging communication technologies. She has published research papers in Scopus-indexed journals and conferences and is actively involved in research, funded projects, innovation, and academic research coordination.



Kilvani Palanisamy Sampooranam received her BE degree in electronics and communication engineering from VLB Janaki Ammal College of Engineering & Technology, Coimbatore, India in 1990 and ME degree in VLSI Design from Government College of Technology, Coimbatore, India in 2005. She has completed her Ph.D from Anna University, Chennai, India in November 2013. Her areas of interest include Wireless Communication, Networking and Embedded Systems.



Vijay Nath received his BSc degree in physics from DDU University Gorakhpur, India in 1998 and PG Diploma in computer networking from MMM University of Technology Gorakhpur, India in 1999 and MSc degree in electronic from DDU University Gorakhpur, India in 2001 and PhD degree in electronics from Dr. Ram Manohar Lohiya Avadh University Ayodhya, India in 2008. Presently, he is

Professor Incharge of VLSI Design Lab, Department of Electronics and Communication Engineering, Birla Institute of Technology, Mesra, Ranchi, Jharkhand, India. He is accelerator of India Semiconductor Mission at State and National Level. He have more than 250 Scopus & SCI/SCIE publications in top world class journals and proceedings. He is Brand Ambassador of IIC, AICTE New Delhi. His areas of interest are low power very large-scale integration design, ultra-low power complementary metal oxide semiconductor temperature sensors for missiles applications, micro-electronics engineering, mixed signal design, application specific integrated circuit design, embedded system design, cardiac pacemaker, internet of things, artificial intelligence & machine learning, and computational intelligence.