IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 1, June 2023

Low Power Implementation of Mitchells Approximate Logarithimic Multiplication for Convolutional Neural Networks

Kanuparthi Venkata Siva Prasad Reddy, Pamuluru Ganesh, Pala Mohan Sai, Gorla Prateesh
Department of Electronics and Communication Engineering
Prathyusha Engineering College, Thiruvallur, Tamil Nadu, India

Abstract: *Approximate computing (AC) is an emerging paradigm that leverages the inherent error tolerance of many applications—such as image recognition, multi- media processing, and machine learning (ML)—to allow some accuracy to be traded off to save energy consumption. AC techniques can be applied at both the circuit and/or architecture levels, possibly in coordination with software- level techniques. Multiplication is one of the most resource- and power-hungry operations in many error-tolerant computing applications, such as image processing, neural networks (NN), and digital signal processing (DSP). In this research project, we focus on the design and implementation of hardware-efficient approximate computing circuits, aiming to simplify the multiplication operation and/or to reduce the number of required multiplications. Two 4×4 approximate multiplier designs are proposed in which approximation is employed in the partial product reduction tree, the most expensive part of the design of a multiplier. The two proposed designs are then used to construct larger approximate multipliers. Multiplication is the computational bottleneck in NNs. For the first time, we attempt to find the critical features in an approximate multiplier that make it superior to others for use in a NN. Inspired by the insight that adding small amounts of noise can improve the performance of NNs, we replaced the ex- act multipliers in two representative NNs with 600 approximate multipliers and then experimentally measured the effect on classification accuracy. Interestingly, some approximate multipliers improved the performance of NNs. Insight into which features of an approximate multiplier make it superior to others in the NN applications was gained by training a statistical predictor that anticipates how well a given approximate multiplier is likely to work in a NN application. In the logarithmic number system (LNS) the multiplication operation is converted into simple shift and addition operations. We have proposed a novel exact leading-one detector (LOD) to speed up the calculation of the base-2 logarithm of the input operands to a logarithmic multiplier. In addition, since the logarithmic multipliers that use LODs always underestimate the actual multiplication product, a nearest-one detector (NOD) is proposed for a logarithmic multiplier that has a double-sided error distribution. Finally, we investigate the design of multiply-accumulate (MAC) units. An approximate logarithmic MAC (LMAC) unit is proposed for the first time. Furthermore, a soft-dropping low-power (SDLP) architecture is specifically designed for convolutional neural networks (CNNs) that, unlike the existing accelerators that simplify the multiplication/addition operations, reduces the number of required multiplications. The SDLP takes advantage of the spatial dependence between the input image pixels and skips some of the multiplications during the convolution operation and, thereby, reduces the energy consumption of the CNN inference calculation*

Keywords: Approximate computing

REFERENCES

[1]. Venkataramani, S.; Chakradhar, S.T.; Roy, K.; Raghunathan, A. Approximate computing and the quest for computing efficiency. In Proceedings of the 52nd Design Automation Conference, San Francisco, CA, USA, 8–12 June 2015. [Google Scholar]

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 1, June 2023

[2]. Breuer, M.A. Multi-media applications and imprecise computation. In Proceedings of the 8th Euromicro Conference on Digital System Design, Porto, Portugal, 30 August–3 September 2005. [Google Scholar]

[3]. Zhang, H.; Putic, M.; Lach, J. Low power GPGPU computation with imprecise hardware. In Proceedings of the 51st Design Automation Conference, San Francisco, CA, USA, 1–5 June 2014. [Google Scholar]

[4]. Shoushtari, M.; Rahmani, A.M.; Dutt, N. Quality-configurable memory hierarchy through approximation. In Proceedings of the 14th International Conference on Compilers, Architecture, and Synthesis for Embedded Systems, Taipei, Taiwan, 9–14 October 2011. [Google Scholar]

[5]. Sarwar, S.S.; Srinivasan, G.; Han, B.; Wijesinghe, P.; Jaiswal, A.; Panda, P.; Raghunathan, A.; Roy, K. Energy efficient neural computing: A study of cross-layer approximations. IEEE J. Emerg. Sel. Top. Circuits Syst. 2018, 8, 796–809. [Google Scholar] [CrossRef]

[6]. Sampson, A.; Deitl, W.; Fortuna, E.; Gnanapragasam, D.; Ceze, L.; Grossman, D. EnerJ: Approximate data types for safe and general low-power computation. In Proceedings of the 32nd ACM SIGPLAN Conference on Programming Language Design and Implementation, San Jose, CA, USA, 4–8 June 2011. [Google Scholar]

[7]. Sampson, A.; Nelson, J.; Strauss, K.; Ceze, L. Approximate storage in solid-state memories. In Proceedings of the 46th Annual IEEE/ACM International Symposium on Microarchitecture, Davis, CA, USA, 7–11 December 2013. [Google Scholar]

[8]. Nair, R. Big data needs approximate computing: Technical Perspective. Commun. ACM 2015, 58, 104. [Google Scholar] [CrossRef]

[9]. Panda, P.; Sengupta, A.; Sarwar, S.S.; Srinivasan, G.; Venkataramani, S.; Raghunathan, A.; Roy, K. Cross-layer approximations for neuromorphic computing: From devices to circuits and systems. In Proceedings of the 53rd Annual Design Automation Conference, Austin, TX, USA, 5–9 June 2016. [Google Scholar]

[10]. Jiang, H.; Santiago, F.J.H.; Mo, H.; Liu, L.; Han, J. Approximate arithmetic circuits: A survey, characterization, and recent applications. Proc. IEEE 2020, 108, 2108–2135. [Google Scholar] [CrossRef]

[11]. Scarabottolo, I.; Ansaloni, G.; Constantinides, G.A.; Pozzi, L.; Reda, S. Approximate logic synthesis: A survey. Proc. IEEE 2020, 108, 2195–2213. [Google Scholar] [CrossRef]

[12]. Jiang, H.; Liu, C.; Liu, L.; Lombardi, F.; Han, J. A review, classification, and comparative evaluation of approximate arithmetic circuits. ACM J. Emerg. Technol. Comput. Syst. 2017, 13, 1–37. [Google Scholar] [CrossRef][Green Version]

[13]. Garside, J.D. A CMOS VLSI implementation of an asynchronous ALU. In Proceedings of the IFIP Working Conference on Asynchronous Design Methodologies, Manchester, UK, 31 March–2 April 1993. [Google Scholar]

[14]. Wanhammar, L. DSP Integrated Circuits, 1st ed.; Academic Press: Cambridge, MA, USA, 1999; ISBN 9780127345307. [Google Scholar]

[15]. Raha, A.; Jayakumar, H.; Raghunathan, V. Input-based dynamic reconfiguration of approximate arithmetic circuits for video encoding. IEEE Trans. VLSI Syst. 2016, 24, 846–857. [Google Scholar] [CrossRef]

[16]. Ercegovac, M.D.; Lang, T. Digital Arithmetic; Morgan Kaufmann: Burlington, MA, USA, 2003; ISBN 978-1558607989. [Google Scholar]

[17]. Jiang, H.; Liu, C.; Maheshwari, N.; Lombardi, F.; Han, J. A comparative evaluation of approximate multipliers. In Proceedings of the IEEE/ACM International Symposium on Nanoscale Architectures, Beijing, China, 18–20 July 2016. [Google Scholar]

[18]. Vai, M.M. VLSI Design; CRC Press: Boca Raton, FL, USA, 2000; ISBN 978-0849318764. [Google Scholar]

[19]. Mahdiani, H.R.; Ahmadi, A.; Fakhraie, S.M.; Lucas, C. Bio-inspired computational blocks for efficient VLSI implementation of soft-computing applications. IEEE Trans. Circuits Syst. I Regul. Pap. 2010, 57, 850–862. [Google Scholar] [CrossRef]

[20]. Balasubramanian, P.; Maskell, D.L. Hardware efficient approximate adder design. In Proceedings of the IEEE Region 10 Conference, Jeju, Korea, 28–31 October 2018. [Google Scholar]

[21]. Balasubramanian, P.; Maskell, D.L. Hardware optimized and error reduced approximate adder. Electronics 2019, 8, 1212. [Google Scholar] [CrossRef][Green Version]

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 1, June 2023

[22]. Balasubramanian, P.; Nayar, R.; Maskell, D.L.; Mastorakis, N.E. An approximate adder with a near-normal error distribution: Design, error analysis and practical application. IEEE Access 2021, 9, 4518–4530. [Google Scholar] [CrossRef]

[23]. Balasubramanian, P.; Nayar, R.; Maskell, D.L. An approximate adder with reduced error and optimized design metrics. Accepted for publication. In Proceedings of the 17th IEEE Asia Pacific Conference on Circuits and Systems, Penang, Malaysia, 22–26 November 2021. [Google Scholar]

[24]. Balasubramanian, P.; Nayar, R.; Maskell, D.L. Approximate array multipliers. Electronics 2021, 10, 630. [Google Scholar] [CrossRef]

[25]. Balasubramanian, P.; Nayar, R.; Min, O.; Maskell, D.L. Image blending using approximate multiplication. In Proceedings of the IEEE 32nd International Conference on Microelectronics, Nis, Serbia, 12–14 September 2021. [Google Scholar]

[26]. Approximator. Available online: https://github.com/OkkarMin/approximator-tool (accessed on 7 November 2021).

[27]. Approximator Tool Documentation. Available online: https://tool-documentation.vercel.app (accessed on 7 November 2021).

[28]. Zhu, N.; Goh, W.L.; Zhang, W.; Yeo, K.S.; Kong, Z.H. Design of low-power high-speed truncation-error-tolerant adder and its application in digital signal processing. IEEE Trans. VLSI Syst. 2010, 18, 1225–1229. [Google Scholar]

[29]. Albicocco, P.; Cardarilli, G.C.; Nannarelli, A.; Petricca, M.; Re, M. Imprecise arithmetic for low power image processing. In Proceedings of the 46th Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, USA, 4–7 November 2012. [Google Scholar]

[30]. Gupta, V.; Mohapatra, D.; Raghunathan, A.; Roy, K. Low-power digital signal processing using approximate adders. IEEE Trans. Comput. Aided Des. Integr. Circuits Syst. 2013, 32, 124–137. [Google Scholar] [CrossRef]

[31]. Dalloo, A.; Najafi, A.; Garcia-Ortiz, A. Systematic design of an approximate adder: The optimized lower part constant-OR adder. IEEE Trans. VLSI Syst. 2018, 26, 1595–1599. [Google Scholar] [CrossRef]

[32]. Seo, H.; Yang, Y.S.; Kim, Y. Design and analysis of an approximate adder with hybrid error reduction. Electronics 2020, 9, 471. [Google Scholar] [CrossRef][Green Version]

[33]. Bovik, A. Handbook of Image and Video Processing, 2nd ed.; Academic Press: Orlando, FL, USA, 2005; ISBN 978-0080533612. [Google Scholar]

[34]. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. IEEE Trans. Image Processing 2004, 13, 600–612. [Google Scholar] [CrossRef] [PubMed][Green Version]

[35]. Chan, W.-T.J.; Kahng, A.B.; Kang, S.; Kumar, R.; Sartori, J. Statistical analysis and modeling for error composition in approximate computation circuits. In Proceedings of the 31st IEEE International Conference on Computer Design, Asheville, NC, USA, 6–9 October 2013. [Google Scholar]

[36]. Balasubramanian, P.; Maskell, D.L. Factorized carry lookahead adders. In Proceedings of the IEEE 14th International Symposium on Signals, Circuits and Systems, Iasi, Romania, 11–12 July 2019. [Google Scholar]

[37]. Synopsys SAED_EDK32/28_CORE Databook. Revision 1.0.0, January 2012. Available online: https://www.synopsys.com/community/university-program/teaching-resources.html (accessed on 27 September 2021).

[38]. Yamamoto, T.; Taniguchi, I.; Tomiyama, H.; Yamashita, S.; Hara-Azumi, Y. A systematic methodology for design and analysis of approximate array multipliers.