

Distributed Predictive Maintenance Architecture for Edge Sensors Networks: An Optimal Regression Based Machine Learning Model

Jacob Greasley^{1,2}, Oluyomi Simpson², Iosif Mporas²*

¹MBDA

²University of Hertfordshire

*corresponding author: j.greasley2@herts.ac.uk

In predictive maintenance (PdM), implementing machine learning models (ML) on edge sensor hardware is particularly challenging. This is due to power constraints which significantly reduce computational performance in conventional embedded processors such as central processing units (CPUs) and microcontroller units (MCUs). However, Field Programmable Gate Arrays (FPGAs) have been identified as an ideal processing unit to overcome this, providing hardware acceleration of models on the edge. With low-precision data, FPGAs have been shown to outperform conventional processing units both in terms of giga-operations-per-second (GOPS) and power consumption. This research seeks to establish an effective methodology for implementing high-level ML regression models on FPGAs within edge sensors.

Keywords: FPGAs; machine learning; edge sensors; hardware acceleration; regression models.

Introduction

Advancements in modern machine learning (ML) techniques have been incorporated with predictive maintenance (PdM) to great effect, allowing data analysts and maintenance engineers to uncover hidden anomalies and interpret their meanings in ways never before possible [1]. However, machine learning alone cannot account for the rise of PdM. Its great potential is enabled by advancements in a whole host of other interdependent technologies. Today, high-level computational tools are more powerful than ever before, allowing for greater precision and accuracy in simulation and analysis of engineering systems through physics-based modelling. Telecommunication technologies are also more powerful, facilitating faster, higher-capacity interconnectivity between systems, end-users, and databases - the concept referred to as the internet-of-things (IoT) [2].

Under the IoT umbrella, so-called edge sensors used for condition monitoring (CM) are also ongoing vast technological advancements, whereby embedded computing enables a smart gateway between the physical world and the digital infrastructure of PdM. Now, in distributed architectures machine learning models can be pushed to the edge to overcome common challenges such as latency and bandwidth requirements associated with networking, cloud computing and the wider IoT [3]. As a result, edge sensors themselves need to evolve, to improve symbiosis with machine learning technology.

Implementing ML models on sensor hardware is challenging, as their complexity and the inherent low power constraints of edge devices, greatly reduce computational performance in common embedded processors such as CPUs and MCUs. Graphics processing units (GPUs) demonstrate greater performance in the training and validation of ML models than CPUs and MCUs, while also comfortably handling floating point datatypes, but their high-power consumption and physical sizes make them unsuitable for low-power embedded application. Hence, they are not well suited to typical remote sensors and IoT nodes. On the other hand, research has shown that for low precision data, FPGAs outperform GPUs both in terms of giga-operations-per-second (GOPS) and power consumption, with a better performance per watt in the execution of mathematically complex operations. They also have better form factor and heat dissipation.

For speed, ASICs are the best option for hardware acceleration, however they are the most costly and time consuming to develop [5] [6]. In comparison, FPGAs are considerably cheaper and quicker to develop, and still

provide superior speed to CPUs, MCUs and GPUs due to their parallelism. This parallelism especially thrives in performing computationally repetitive tasks. Furthermore, many modern FPGAs contain hardened functionality to increase execution speed of complex functions. Therefore, to accelerate ML models on FPGAs, we must identify parts of the code that are good candidates for utilizing parallelization and hardened functionalities, then we must optimize their conversion to register transfer level (RTL) abstraction [4] [7].

Armed with an understanding of the target platform, this research seeks to identify, train and adapt a suitable ML regression algorithm for deployment on FPGAs in edge sensors. Planned activities include investigating the capabilities of ML libraries and platforms in generating light weight models for edge deployment and the application of high-level synthesis (HLS) tools.

Experimental/Simulation

The synthesized schematic shown below was generated from initial attempts to implement signal processing algorithms on the target Artix 7 FPGA through the use of HLS tools.

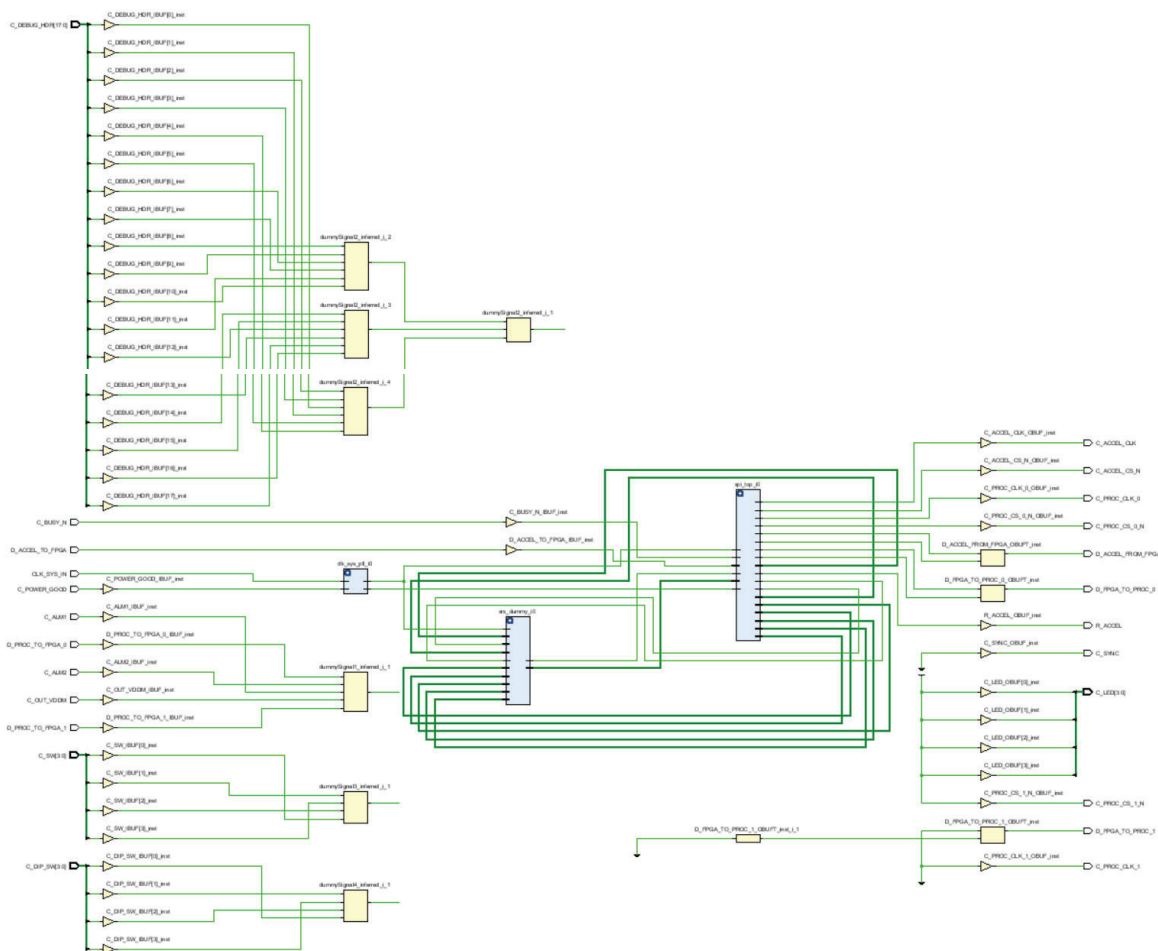


Figure 1. Schematic of synthesized algorithm to be implemented on edge sensor.

In figure 2. a power estimation of the design has been generated. Here the on-chip power consumption has been elaborated.

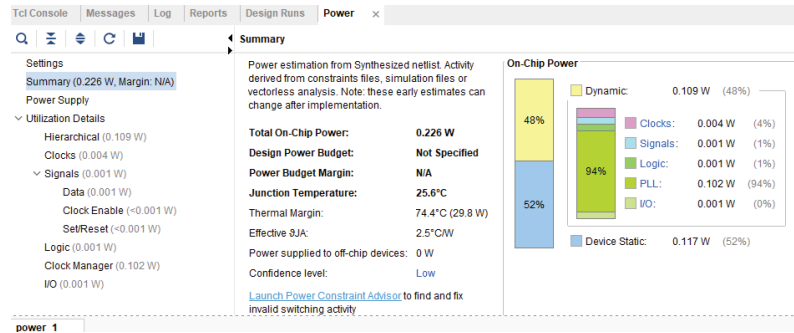


Figure 2. Estimated power consumption of synthesized algorithm.

Results and discussion

The power estimates in Vivado indicate a power consumption of 48% for the DSP algorithm used. The size and complexity of the model now serves as a benchmark for the lightweight ML algorithms that are to be explored in this research. The netlist properties show that the most used primitives are flip flops and look up tables (LUTs). However, with ML model synthesis, this is expected to be very different with a higher usage of multiply accumulate units (MACs) and LUTs.

Conclusion

Considering the resource utilization of the DSP algorithm, it's safe to assume that greater optimization will be needed to successfully implement a ML model on the same device, as ML models are typically far more complex. It is possible that after further investigation this research work may conclude that the Artix 7 is simply not suitable to provide sufficiently high-resolution sensor data for effective predictive analytics. However, with the wide array of ML model development tools and platforms, there is certainly potential for an innovative solution.

References

- [1] P. Ivanovic, A. Ganbarov and P. Neumann, "System Architecture for Real-time Condition Monitoring and Anomaly Detection on Ships," in *2023 22nd International Symposium on Parallel and Distributed Computing (ISPDC)*, Bucharest, 2023.
- [2] H. H. Dewey and D. R. DeVries, "Case study in utilizing the Internet of Things as a PHM architecture for aerospace applications," in *2018 IEEE Aerospace Conference*, Big Sky, MT, USA, 2018.
- [3] P. Nunes, J. Santos and E. Rocha, "Challenges in predictive maintenance – A review," in *CIRP Journal of Manufacturing Science and Technology*, 2023.
- [4] Rijad Sarić et al., "Design of ASIC and FPGA system with Supervised Machine Learning Algorithms for Solar Particle Event Hourly Prediction," *IFAC-PapersOnLine*, pp. 230-235, 20 July 2022.
- [5] A. Boutros and V. Betz, "FPGA Architecture: Principles and Progression," *IEEE Circuits and Systems Magazine*, vol. 21, no. 2, pp. 4-29, 2021.
- [6] A. M. Abdulla and M. M. Al Khalidy, "Technical methods for accelerating CNNs for IoT using FPGAs and machine learning," in *3rd Smart Cities Symposium (SCS 2020)*, Online Conference, 2020.
- [7] M. Langhammer, "FPGA Processing Block for Machine Learning or Digital Signal Processing Operations". USA Patent US-20210326111-A1, 2021.
- [8] S. Ahmad, K. Styp-Rekowski, S. Nedelkoski and O. Kao, "Autoencoder-based Condition Monitoring and Anomaly Detection Method for Rotating Machines," in *2020 IEEE International Conference on Big Data (Big Data)*, Atlanta, GA, USA, 2020.
- [9] Pradip Kumar Sharma et al., "A Software Defined Fog Node Based Distributed Blockchain Cloud Architecture for IoT," *IEEE Access*, vol. 6, pp. 115 - 124, 2017.
- [10] Nasir Abbas et al., "Mobile Edge Computing: A Survey," *IEEE Internet of Things Journal*, vol. 5, no. 1, pp. 450 - 465, 2018.