Proposed PhD Title: **Adaptive Logic-Based Machine Learning Grids**

Supervision Team:
Tousif Rahman (**Early Career Academic**),
Rishad Shafik (Professor in Microelectronic Systems + Director of Microsystems AI (MAI) Lab, Co-Founder of Literal Labs),
Alex Yakovlev (Professor of Computer System Design, Co-Founder of Literal Labs)

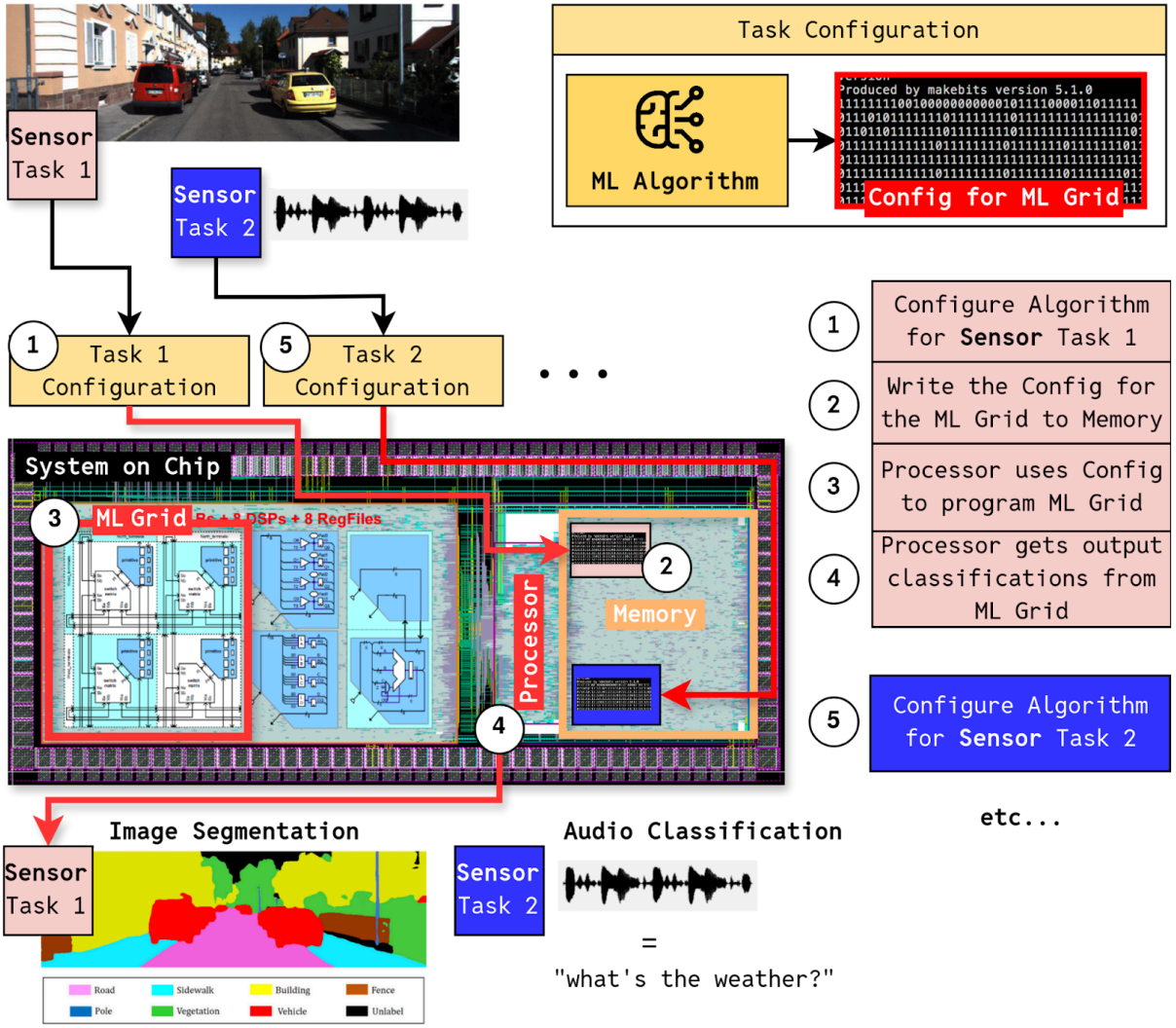

Fig 1: Visual overview of the project. Creating an ultra-small **reconfigurable ML Grid** Integrated Circuit (IC) System-on-Chip for IoT sensor tasks like image segmentation or audio classification. The chip takes design cues from [1] and [3].

Overview: This project will develop an adaptable Machine Learning (ML) hardware architecture to solve Artificial Intelligence (AI) classification tasks using Internet of Things (IoT) sensor data. In contrast to high power GPUs, this will be a small system-on-chip designed to operate on the edge (i.e. close to the sensor). The project will explore whether emerging logic-based ML algorithms can be translated into *smaller*, *faster*, more *energy efficient* and *cost-effective* hardware compared to the current state-of-the-art. The design will involve an ML grid which will adapt to

tasks in real time such that grid elements can be powered on only upon use (see Fig.1). The project will involve aligning the in-house algorithm-to-hardware development of the Micro-Systems Research Group at Newcastle University ([3],[5],[7]) with next-generation Field Programmable Gate Array (FPGA) hardware technologies ([1],[6]).

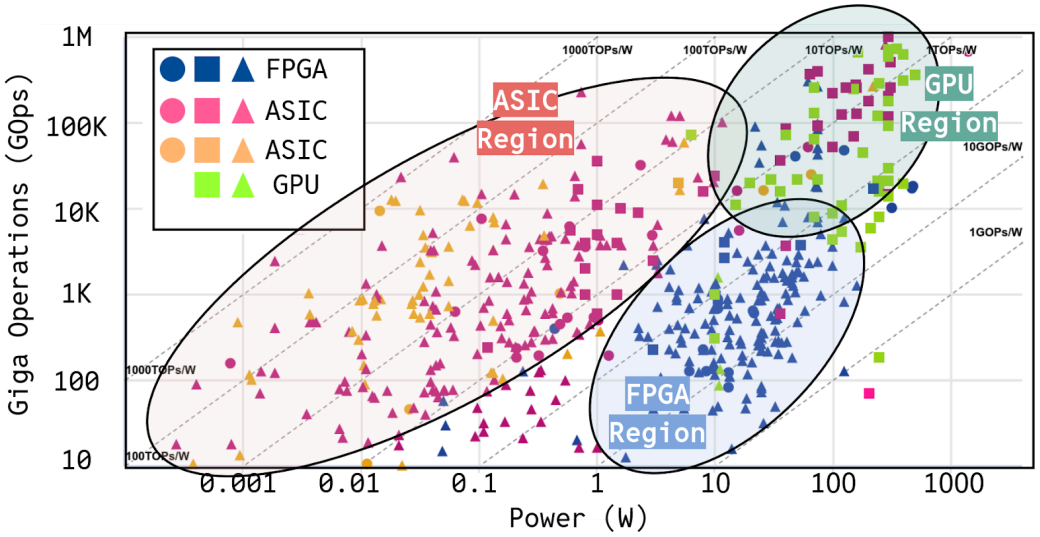

Fig 2: All major ML accelerators (2018-Now revised from [2]).

Creating an Impact in the ML Hardware Accelerator Landscape: Machine Learning (ML) accelerators are specialised hardware platforms designed to offer better energy efficiency and sustainability to AI tasks. The core design principle is *maximising the number of operations* that can be computed within the *smallest power budget*. This is shown in Fig.2 (GOps/W). Application Specific Integrated Circuits (ASICs) offer the best GOps/W. FPGAs are naturally grid-like, they offer similar GOps but use more power. However, they are much cheaper and require less development time. This project will develop custom power efficient FPGA based ML grids that operate in the ASIC region.

Methodology, Novelty and Knowledge Exchange: Off-the-shelf FPGAs have large, complex grid elements, however, open-source frameworks like Open-FPGA [6] and FABULOUS [1] offer smaller and customizable grids. The project aim is to efficiently map logic-based ML algorithms like Tsetlin Machines to smaller, custom lower power ML grids (Fig. 3). The project presents an opportunity for greater collaboration with the FABULOUS [1] and Open-FPGA developers at University of Manchester, Heidelberg and Utah as well as those who regularly attend the Open-FPGA summer school [4].

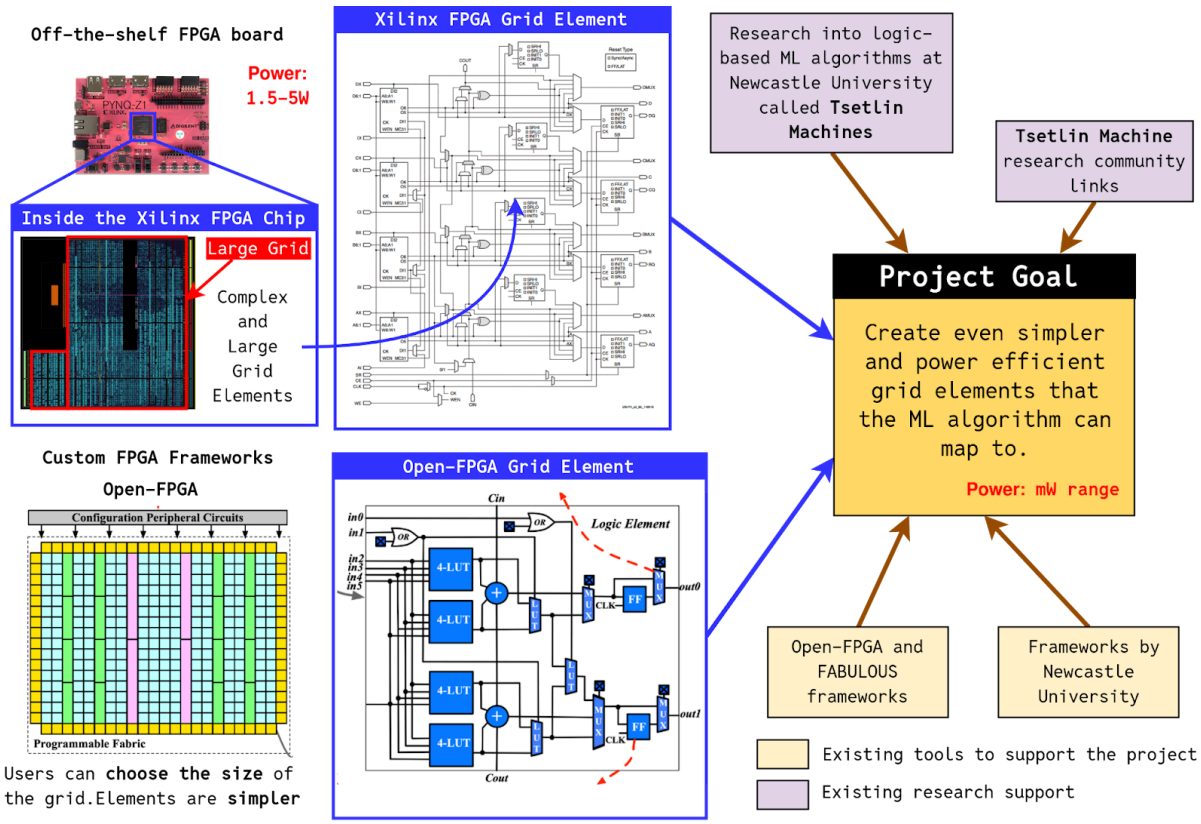

Fig.3: Visualising the project goal.

Research Questions:

Main Question: Can ML applications (like the sensor tasks above) be mapped to reprogrammable custom ML grids using custom FPGA technology and offer similar GOps/W as ASIC platforms?

1) Can sparse logic based algorithms like Tsetlin Machines be translated to open source frameworks like Open-FPGA to develop reprogrammable custom ML grids? What are the overheads? How does it compare against standard FPGAs and similar algorithms?
2) How scalable are designs like this? Can they be automated to produce larger ML grids for more complex problems but retain power efficiency?

The timeline has been designed with adherence to implementation requirements from similar projects (Fig. 4). Each year the Open-FPGA summer school serves as a checkpoint for evaluating progress.

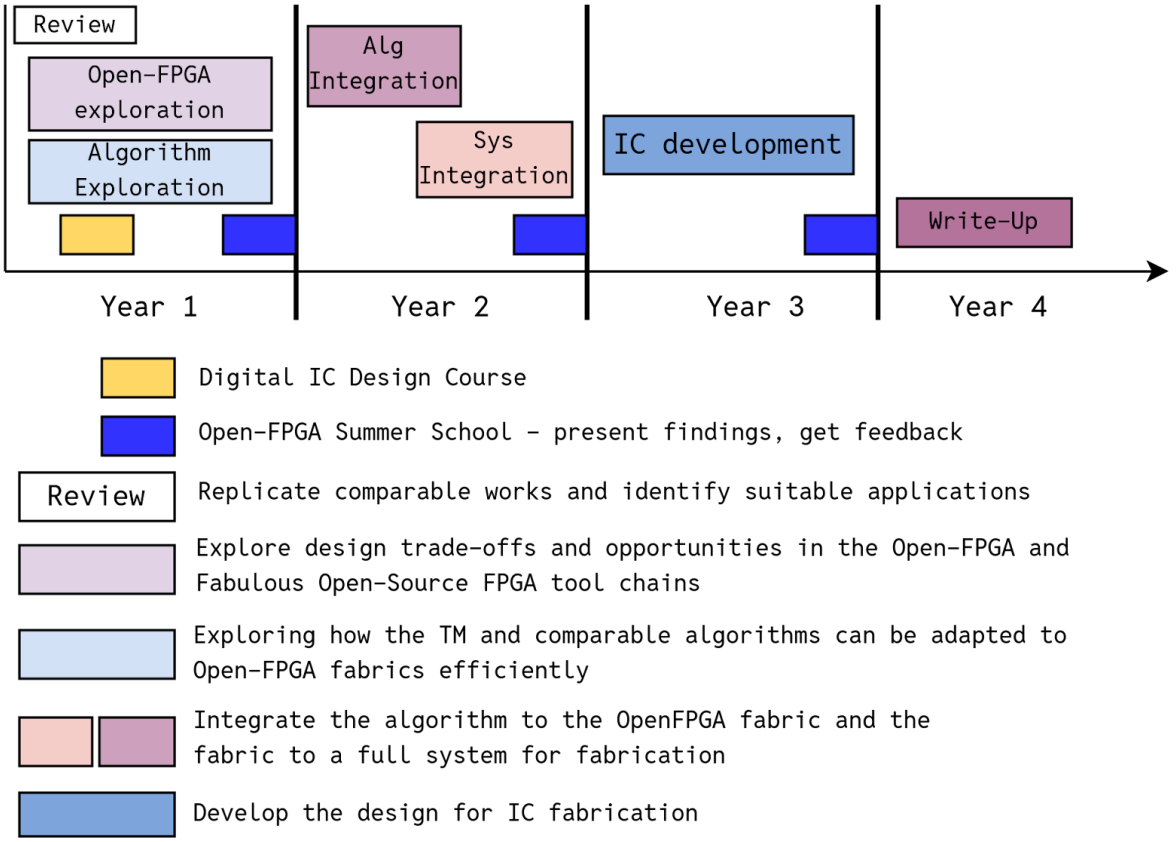

Fig.4: Proposed Project Timeline

References

1. D. Koch, N. Dao, B. Healy, J. Yu, and A. Attwood. “Fabulous: an embedded fpga framework,” In The 2021 ACM/SIGDA International Symposium on Field-Programmable Gate Arrays, FPGA '21, 45–56. New York, NY, USA, 2021. Association for Computing Machinery. URL: https://doi.org/10.1145/3431920.3439302
2. NICSEF Project Accelerator Comparison. Available at: https://nicsefc.ee.tsinghua.edu.cn/project.html (Accessed: 25 November 2024).
3. T. Rahman, G. Mao, S. Maheshwari, R. Shafik, and A. Yakovlev, “MATADOR:Automated System-on-Chip Tsetlin Machine Design Generation for Edge Applications,” in 2024 Design, Automation and Test in Europe Conference and Exhibition (DATE), 2024, pp. 1–6. DOI: 10.23919/DATE58400.2024.10546779
4. IGNITE FPGA Summer School (2024), Available at: https://fpga-ignite.github.io/ (Accessed: 25 November 2024).
5. S. Maheshwari, T. Rahman, R. Shafik, A. Yakovlev, A. Rafiev, L. Jiao, and O.-C. Granmo, “REDRESS: Generating Compressed Models for Edge Inference Using Tsetlin Machines,” IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 1–16, 2023, DOI: 10.1109/TPAMI.2023.3268415
6. X. Tang, E. Giacomin, B. Chauviere, A. Alacchi and P. -E. Gaillardon, "OpenFPGA: An Open-Source Framework for Agile Prototyping Customizable FPGAs," in IEEE Micro, vol. 40, no. 4, pp. 41-48, 1 July-Aug. 2020, DOI: 10.1109/MM.2020.2995854.
7. O. Ghazal *et al.*, "IMBUE: In-Memory Boolean-to-CUrrent Inference ArchitecturE for Tsetlin Machines," 2023 IEEE/ACM International Symposium on Low Power Electronics and Design (ISLPED), Vienna, Austria, 2023, pp. 1-6, doi: 10.1109/ISLPED58423.2023.10244315.