PhD Thesis: Study of the environmental impact of edge IoT platforms running machine learning processes

Marie-Anne LACROIX, Robin GERZAGUET and Pascal SCALART
Univ Rennes, IRISA, GRANIT Team
www-granit.irisa.fr

1 Thesis Context

For several years, the Internet of Objects (IoT) sphere has taken up the issue of artificial intelligence (AI). In order to embed this technology on communicating objects, which are subject to severe constraints in terms of memory, computing power and energy capacity, numerous studies have been conducted to reduce the number of parameters or operations for these algorithms [LLM23]. These works reduce the computation complexity of AI and make it possible to run such an algorithm on hardware target. Furthermore, awareness of the ecological emergency raises the question of AI energy consumption, whatever its use [BCJL24].

The work conducted to reduce computational complexity directly concerns the use phase of the devices, which mainly impacts on climate change. However, electronics is known for having a strong environmental impact on other elements such as water toxicity or mineral depletion, to name but a few. This impact is mainly concentrated on the resource extraction, manufacturing and end-of-life phases [ADE23]. The environmental impact of electronic systems is also linked to the potential for reuse of all or part of the hardware after their original use (the so-called end-of-life phase in life-cycle analysis). So it is important to look at the non-use phases, which together make up the life cycle of electronic devices.

During this thesis, we aim to study the environmental impact of the various hardware targets enabling AI to be used in the IoT. These include a wide variety of possibilities, depending on application constraints: from the simple microcontroller to the specific NPU (neural processing unit) processor, via GPUs and FPGAs [FHL+24, SBA23]. To the best of our knowledge, no data or models currently exist to help us understand how to arbitrate between these different execution targets, taking into account different environmental impacts.

The environmental impacts of electronic devices and communicating objects are addressed by the CMA ESOS (https://esos.insa-rennes.fr/en/) and CMA RIS3 (https://ris3.insa-rennes.fr/) projects respectively, who are co-financing this thesis.

2 Thesis Objectives

The thesis focuses on the environmental impact of AI-enabled embedded devices. Several runtime platforms will be looking at, and we wil consider two IoT use cases : an industrial real-time sensor network for sound event classification (in association with the ANR Edge AI "Light-swift") and a sensor network dedicated to environmental monitoring on ultra-low consumption platform (in association with ANR Thématiques Spécifiques en IA - Circuits "OWL"). The aim of the thesis is to assess the environmental impact of the entire life cycle of IoT devices by studying the influence of the computing platform and the communication stack.

- **General or specialized architecture?**
  Given the particular constraints of AI and its increasingly widespread use, specialized architectures such as NPUs are integrated into some embedded chips. Nevertheless, these chips generally execute a limited instruction set, which restricts their use to some applications and may prevent their use if we wish to upgrade the system with new layers or neural network architectures. In this thesis, we will develop a methodology for assessing the environmental impact of different available architectures (microcontroller, GPU, FPGA, NPU) based on models of these architectures and models of AI algorithm execution. One of the challenges and a profound originality with the state of the art will be to integrate into the methodology metrics/factors that are interrelated and others that are very different, for different phases of the life cycle (extraction/manufacturing, use, end of life) [CABJ24]. This will enable us to compare the environmental impact of different architectures, specialized or more generalist, for different use cases, and in particular the two mentioned above.

- **Local or distant computing : when communicate?**
  Wireless communication is generally considered the most energy-intensive part of IoT [FLMFA21]. However, with embedded AI and depending on the use case (AI algorithm, frequency of AI calculations, intermittent or non-intermittent communication, amount of data to be transmitted, ...), the shares of consumption are highly variable. The aim is to understand the extent to which, in the use phase of the IoT system, the energy consumed by 1) the communication part of the IoT object (transmitter-receiver), 2) wireless communication itself, can have an impact on the system's environmental impact, considering the whole life cycle. The two use cases considered will be used to support answers to this open question.

- **Standard battery or recheargeable battery with energy recovery?**
  We can envisage two diametrically different types of power supply for communicating objects: a standard battery dimensioned for the entire lifetime of the system, immovable but well dimensioned, or a rechargeable battery coupled to an energy recovery system (in this thesis, we will consider conventional energy recovery, via a photovoltaic collector [Aoudia2018]). On the basis of these two types of systems [Hasan2023] and for the two use cases considered, we will try to identify the most relevant criteria to consider when carrying out an environmental impact analysis, taking into account the entire life cycle.

3 Provisional schedule

Part 1 : State of the art

- environmental impacts of execution targets (microcontroller, GPU, FPGA, NPU) in the extraction/manufacturing phases and relative impact models ;
- energy consumption models of execution targets ;
- approaches for analyzing the environmental impact of electronic circuits and systems.

Part 2 : Proposal of impact models and a methodology for analyzing the environmental impact of AI-enabled embedded systems.

Part 3 : Experimental part

- development of a tool to analyze the environmental impact of embedded systems over their entire life cycle ;
- comparison of energy consumption models with measurements on different targets; benchmarking; model refinement; (2 use cases)
- use of this tool for analyzing environmental impacts of standard or rechargeable battery-powered systems (2 use cases).

4 Skills

Holder of a master's degree or an engineering degree, you have skills in embedded systems and artifical intelligence. You have an interest in research and ecological issues.

5 Information and Contacts

The thesis is carried out in Lannion, in the GRANIT team at the IRISA laboratory.

Marie-Anne LACROIX	marie-anne.lacroix@irisa.fr
Robin GERZAGUET	robin.gerzaguet@irisa.fr
Pascal SCALART	pascal.scalart@irisa.fr

References

[ADE23] ADEME. Evaluation de l'impact environnemental du numérique en france et analyse prospective, 2023.

[BCJL24] Adrien Berthelot, Eddy Caron, Mathilde Jay, and Laurent Lefèvre. Estimating the environmental impact of generative-ai services using an lca-based methodology. *Procedia CIRP*, 122:707–712, 2024.

[CABJ24] Fan Chen, Shahzeen Attari, Gayle Buck, and Lei Jiang. Iotco2: Assessing the end-to-end carbon footprint of internet-of-things-enabled deep learning. *arXiv preprint arXiv:2403.10984*, 2024.

[FHL+24] William Fabre, Karim Haroun, Vincent Lorrain, Maria Lepecq, and Gilles Sicard. From near-sensor to in-sensor: a state-of-the-art review of embedded ai vision systems. *Sensors*, 24(16):5446, 2024.

[FLMFA21] Hamid Reza Farahzadi, Mostafa Langarizadeh, Mohammad Mirhosseini, and Seyed Ali Fatemi Aghda. An improved cluster formation process in wireless sensor network to decrease energy consumption. *Wireless Networks*, 27:1077–1087, 2021.

[LLM23] Zhuo Li, Hengyi Li, and Lin Meng. Model compression for deep neural networks: A survey. *Computers*, 12(3):60, 2023.

[SBA23] Matthew Sibanda, Ernest Bhero, and John Agee. Ai edge processing-a review of distributed embedded systems. In *2023 31st Southern African Universities Power Engineering Conference (SAUPEC)*, pages 1–6. IEEE, 2023.