



HAL
open science

Ultra Low Power Ambient Artificial Intelligence

Antoine Bonneau, Frédéric Le Mouël, Fabien Mieleville

► **To cite this version:**

Antoine Bonneau, Frédéric Le Mouël, Fabien Mieleville. Ultra Low Power Ambient Artificial Intelligence. Journées non thématiques du GDR RSD, Jan 2023, Villeurbanne, France. hal-03986137

HAL Id: hal-03986137

<https://hal.science/hal-03986137>

Submitted on 13 Feb 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Abstract

The increase in autonomy of ambient intelligence devices has led to the evolution of Internet of Things (IoT) towards the notion of Edge Computing and is driving Wireless Sensor Networks (WSN) (considered as the Low End Devices of IoT) towards total energy autonomy by recovering ambient energy. While the majority of current Machine Learning applications are still centralized in Clouds, we could take advantage of those frugal developments to propose systems that take into account their hardware and energy limitations for collaborative and adaptive distributed learning.

Thesis Goals

- ▶ Deployment of a WSN as a monitoring and adaptive action solution
- ▶ Minimal footprint energy aiming for autonomy (battery-less)
- ▶ Exploiting networking AI to improve and maintain itself
- ▶ Energy efficiency as a cross-cutting topic
- ▶ Focus on Time Series data (reduced feature set and samples)

Ampere's active control platform

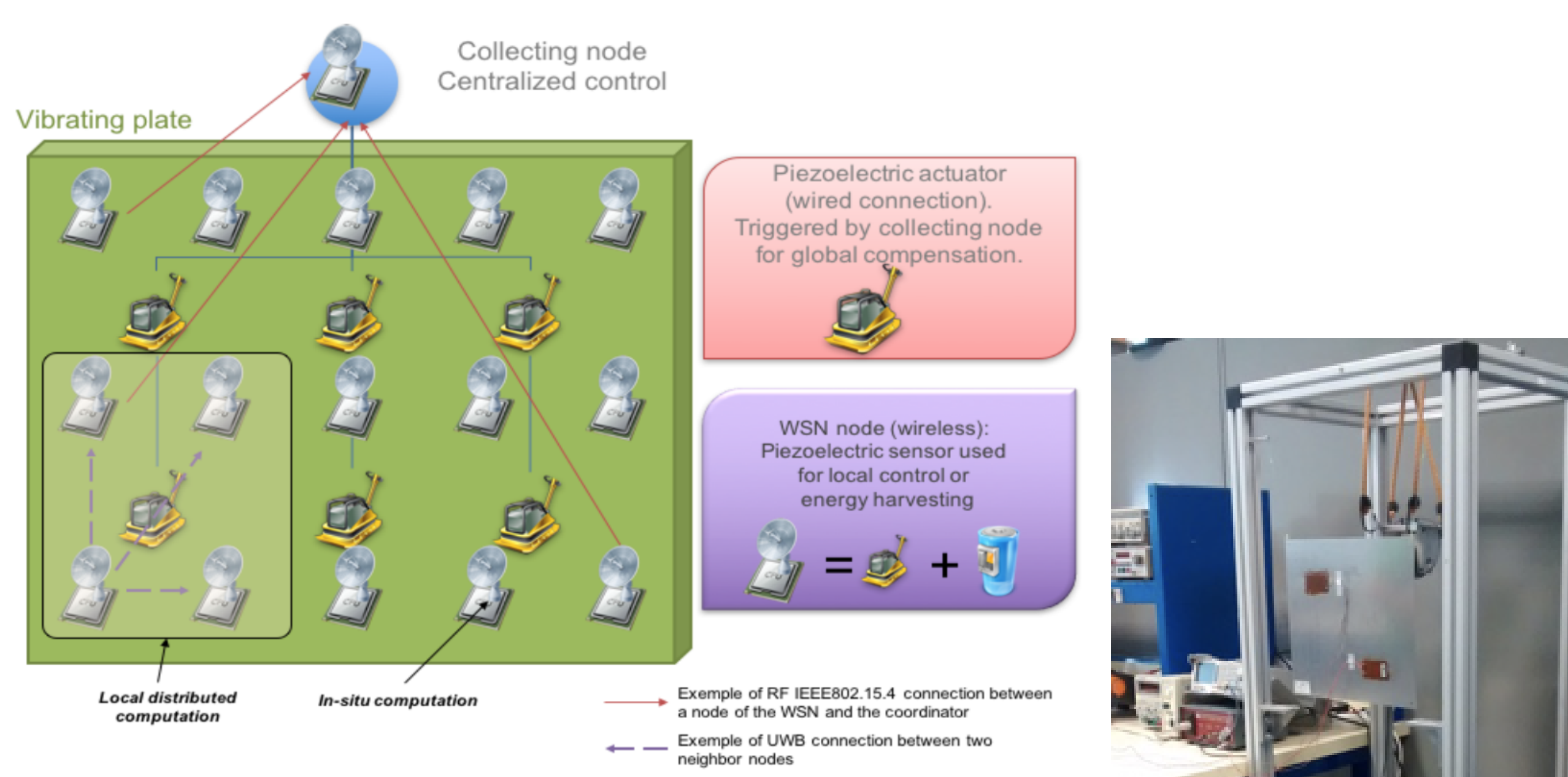


Figure 1: Ampere's vibrating bench

- ▶ Experimental Issues
 - ▶ Low computational power but High Energy Harvesting (EH) potential
 - ▶ Reduce vibration under a threshold locally, collaboratively and globally
 - ▶ Keep WSN up through piezoelectric EH and energy balancing

CITI's Cortex Lab

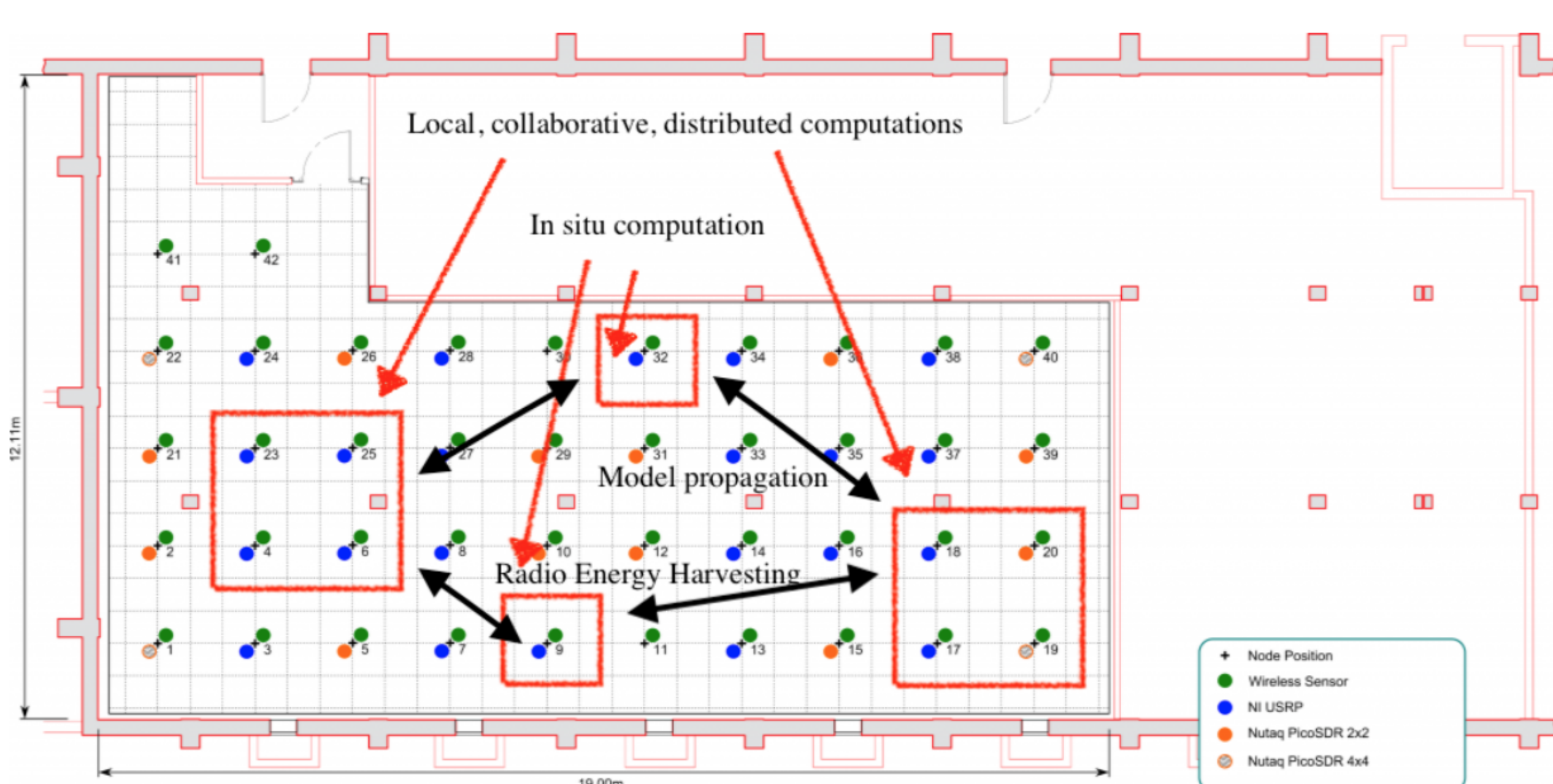
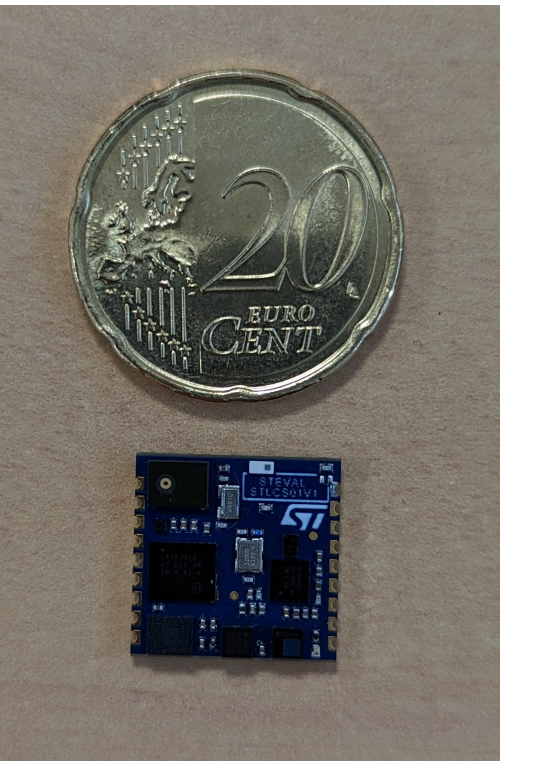


Figure 2: Nodes and activation example in Cortex Lab platform

- ▶ Experimental Issues
 - ▶ High computational power but Low EH potential
 - ▶ Modify antennas behaviour (reallocation, beamforming, power off, ...)
 - ▶ Maximize covering, bandwidth, signal homogeneity, ...
 - ▶ Minimize interferences

On-Device Learning : Resources constrained learning

- ▶ Memory constraints → Compressed Models [1]
 - ▶ Pruning (adaptative, distributed, ...)
 - ▶ Quantization (aware training, partial, ...)
 - ▶ Huffman Encoding.
- ▶ Computation constraints → Preprocessed Models [2]
 - ▶ Use pre-trained models on boards.
 - ▶ Sparsely fine tune model depending on the model.



Target : WSN Nodes / IoT low-end devices

MCU Arch	Clock speed	On-Board Storage	RAM	Current
16 or 32 bit	≤ 50 MHz	≤ 512 KB	≤ 64 KB	≤ 10 mA

Intermittent Learning : Duty Cycling

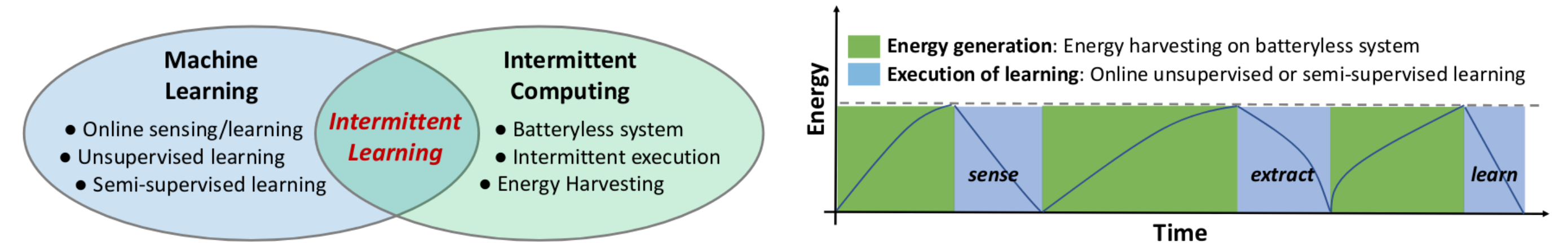


Figure 3: Intermittent Learner duty [3]

- ▶ Extreme power scarcity → Energy Harvesting
 - ▶ Save computations through gradient / energy-aware checkpointing
 - ▶ Decompose monolithic ML with task-based approaches [3]
- ▶ Changes in energy availability → Energy-aware reconfiguration [4]
 - ▶ Power on / off different peripherals and change sensors sampling rates
 - ▶ Adapt compute complexity through early exiting
 - ▶ Change focus from learning to inference

Federated Learning : Bypass computational weaknesses

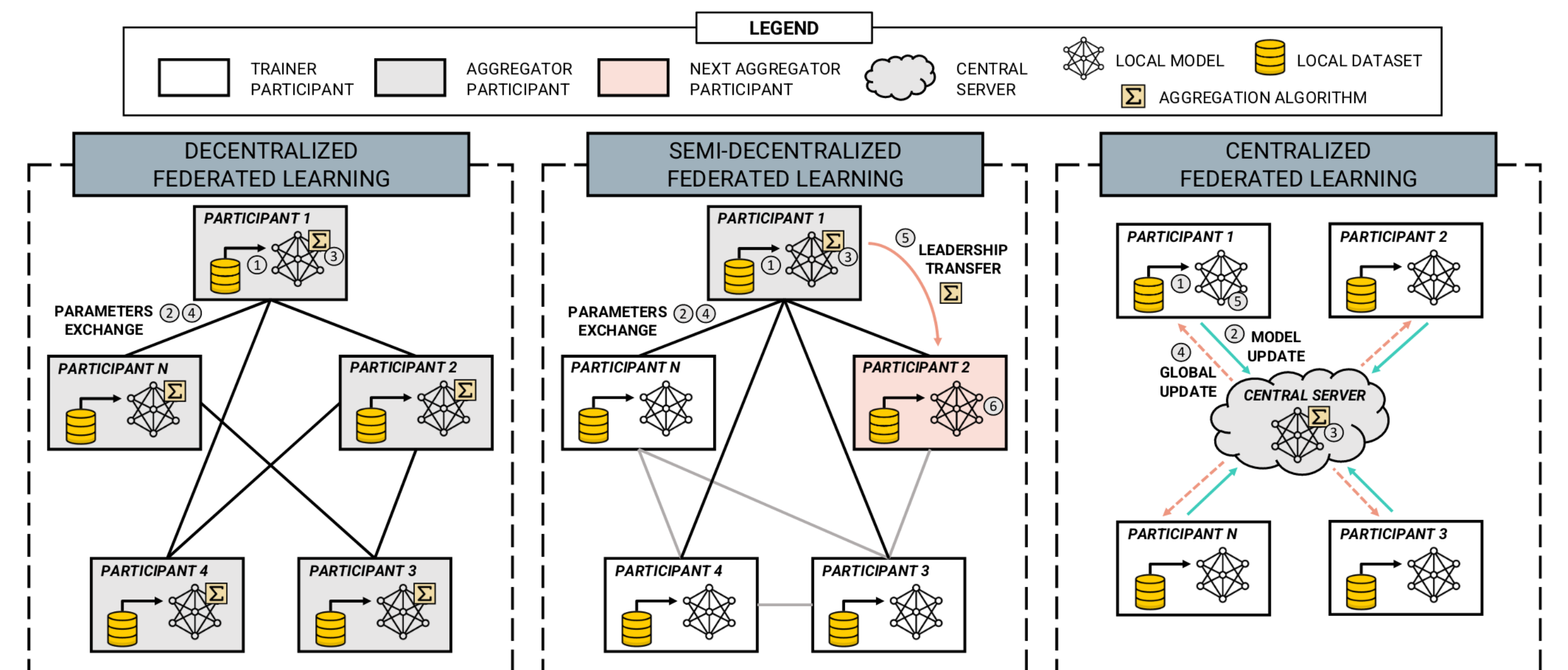


Figure 4: DFL architectures [5]

- ▶ Learning in a spatially correlated group → Adaptive Neighbor Matching
 - ▶ Select neighbors based on similar performances (gossip and PENS)
 - ▶ Establish selection that can significantly improve convergence speed and training loss.
- ▶ Communication is energy-consuming → Optimize transmission
 - ▶ Reduce shared model size with compressed sensing [6]
 - ▶ Choosing communication-efficient aggregators (e.g Dynamic Average Consensus-Based)

Perspectives

- ▶ Build a simulator on top of ns3 to evaluate communication energy consumption of Federated Learning for different topologies.
- ▶ Varying microcontrollers to determine how low you can go : Raspberry Pi, STM32 F302R8, STEVAL-STLK01, MSP430
- ▶ Plugging different types of Energy Harvesting modules (Solar, RF, piezoelectric) and adapting intermittent schedulers to federated learning.
- ▶ Comparing energy consumption to traditional AI approaches (ARIMA)

References

- [1] C. Profentzas, M. Almgren, and O. Landsiedel, "MiniLearn: On-Device Learning for Low-Power IoT Devices," presented at the International Conference on Embedded Wireless Systems and Networks, Sep. 2022. [Online]. Available: <https://research.chalmers.se/en/publication/531933>.
- [2] J. Lin, L. Zhu, W.-M. Chen, W.-C. Wang, C. Gan, and S. Han, "On-Device Training Under 256KB Memory," Jul. 2022. doi: 10.48550/arXiv.2206.15472.
- [3] S. Lee, B. Islam, Y. Luo, and S. Nirjon, "Intermittent Learning: On-Device Machine Learning on Intermittently Powered System," *Proceedings ACM Interactive Mobile Wearable Ubiquitous Technologies*, Dec. 2019. doi: 10.1145/3369837.
- [4] A. Bakar, A. G. Ross, K. S. Yildirim, and J. Hester, "REHASH: A Flexible, Developer Focused, Heuristic Adaptation Platform for Intermittently Powered Computing," *Proceedings ACM Interactive Mobile Wearable Ubiquitous Technologies*, Sep. 2021. doi: 10.1145/3478077.
- [5] E. T. M. Beltrán, M. Q. Pérez, P. M. S. Sánchez, et al., "Decentralized Federated Learning: Fundamentals, State-of-the-art, Frameworks, Trends, and Challenges," Nov. 2022. doi: 10.48550/arXiv.2211.08413.
- [6] C. Li, G. Li, and P. K. Varshney, "Communication-Efficient Federated Learning Based on Compressed Sensing," *IEEE Internet Things J.*, Oct. 2021. doi: 10.1109/JIOT.2021.3073112.