

Master 2 Research Internship and Ph.D. Proposal

Laboratory: Centre de Nanosciences et de Nanotechnologies (C2N)
Université Paris-Saclay/CNRS.

Address: 10 Boulevard Thomas Gobert, 91120 Palaiseau, France

Advisor: Damien QUERLIOZ

Website: <https://sites.google.com/site/damienquerlioze/>

E-mail: damien.querlioze@universite-paris-saclay.fr

université
PARIS-SACLAY



Exploiting the Physics of Memory for Energy-Efficient Generative AI

You will investigate how electronic memory devices can use their physical behavior to perform computations needed by generative AI. Depending on your background, the project will center on device measurements and physical modeling, or on circuit design and simulation. It is particularly suited to students in physics, nanoelectronics or electrical engineering who want to apply their expertise to modern AI.

Generative AI places considerable demands on electricity and computing hardware. Developing more energy-efficient ways to run these systems is an important technological and environmental challenge. Part of the energy cost comes from storing and repeatedly accessing information as a model processes text and generates responses. In widely used language models called transformers, an “attention” mechanism draws on representations of previously processed text, stored in a working memory called the key-value cache, or KV cache. This memory grows with the length of a conversation, increasing storage requirements and data movement. Recent approaches known as state-space models offer another possibility: maintaining a compact “running summary” of earlier inputs, which is updated as new information arrives [1]. This creates an opportunity to connect AI with device physics: could this summary be stored and updated directly as electrical charge or ferroelectric polarization?

Our goal is to develop physical building blocks for this kind of AI memory, with the ambition of making generative AI easier to run on compact, low-power hardware. You will investigate how electrical signals and simple circuits can control how much stored information is retained and how new information is added. By performing these updates directly within memory devices, we aim to reduce the computation and data movement needed to process a sequence.

The project will explore technologies such as hafnium-oxide ferroelectric devices and gain-cell memories using indium gallium zinc oxide (IGZO), an oxide semiconductor. Gain cells store information as electrical charge, while the very low leakage of IGZO transistors helps preserve it. These technologies offer complementary possibilities for retaining, updating and reading information. Our recent demonstration of a combined ferroelectric-memristor memory illustrates how different devices can support different roles within a learning system [2].

The project builds on our research published in Nature Electronics, Nature Communications and Nature Reviews Electrical Engineering. Our results include a memristor-based Bayesian machine [3], experimental neural networks that use memory-device physics to estimate uncertainty [4], and learning algorithms that balance the retention of knowledge with the ability to learn and forget [5]. Together with our work on ferroelectric memories [2] and our perspective on Bayesian electronics [6], these contributions provide foundations for connecting physical memory mechanisms with modern AI.

Funding: The Master’s internship and Ph.D. thesis are fully funded through European funds.

During the Master’s internship, you will develop a first proof of concept showing how a device or small circuit can perform an elementary update of an AI model’s working memory. Depending on your background, you will investigate this through device measurements, physical modeling, or circuit simulation. You will then use the results in a simple neural-network simulation to assess how the physical behavior affects its ability to remember and process information. The necessary deep-learning concepts will be introduced during the project.

During the PhD, you will extend these ideas towards memory arrays and more complete computational blocks inspired by state-space models and recurrent attention. A central challenge will be to control information retention and updates despite noise and differences between devices. You will assess the resulting trade-offs between energy consumption, memory capacity and model performance. One promising direction is

to combine memories with different physical properties to support information that needs to be updated or retained over different timescales.

For a student with an integrated-circuit design background, the PhD can include designing a prototype chip, sending it for fabrication at our partner CEA-Leti and testing it experimentally. **For a student with a stronger physics background, the emphasis can be on experimental devices**, their dynamics and the computational mechanisms they enable. Both routes will involve collaboration with researchers developing learning algorithms.

You will join the interdisciplinary INTEGRANO group at C2N, Université Paris-Saclay/CNRS. This group associates research on nanodevices, bioinspired computing and artificial intelligence, with researchers and students of very diverse backgrounds, making it an exciting environment that fosters interdisciplinary thinking. This work will involve collaboration with Elisa Vianello's group at CEA-Leti. Previous experience with deep learning is welcome, but its foundations can be acquired during the project. You will develop expertise connecting device physics, electronic systems and AI, with a concrete objective: reducing the resources required for generative computation. For more technical details on the project, contact us!

Selected references

- [1] A. Gu and T. Dao, "Mamba: Linear-Time Sequence Modeling with Selective State Spaces", arXiv:2312.00752.
- [2] M. Martemucci, ..., D. Querlioz, E. Vianello, "A ferroelectric–memristor memory for both training and inference", **Nature Electronics** 8, 921–933 (2025).
- [3] K.-E. Harabi, ..., D. Querlioz, "A memristor-based Bayesian machine", **Nature Electronics** 6, 52–63 (2023).
- [4] D. Bonnet, ..., D. Querlioz, E. Vianello, "Bringing uncertainty quantification to the extreme-edge with memristor-based Bayesian neural networks", **Nature Communications** 14, 7530 (2023).
- [5] D. Bonnet, ..., D. Querlioz, "Bayesian continual learning and forgetting in neural networks", **Nature Communications** 16, 9614 (2025).
- [6] D. Querlioz and E. Vianello, "Bayesian electronics for trustworthy artificial intelligence", **Nature Reviews Electrical Engineering** 2, 846–855 (2025).