Arm Sophia Internships 2024

1. Duration: 6 months
2. For students in their final year at university or engineering school, in electronics or computer science
3. Internship location is Sophia-Antipolis, France
4. Apply on careers.arm.com/ (keyword "Internship" and location "France")

As we are shifting in a new era of AI, integrating our engineering teams in Sophia Antipolis will offer you the excitements of shaping the Future Wave of Computing that will be fundamental in making the world we live in, more efficient and more sustainable.

In the next pages, the internship subjects for this season are listed per team:

CPU Microarchitecture & Design

To apply for this offer:

- Head to https://careers.arm.com/.
- Search for offers with keyword “Internship” and location “France”
- Click on offer “CPU Microarchitecture and Design Intern”

[2024-D1] Hardware Stack protection mechanism

Return oriented programming (ROP) is a common method used by attackers to execute malicious code and bypass security defences.

The goal of this internship is to implement a hardware mechanism to record and check the return addresses in a superscalar processor.

During this internship, you will:

- Study the existing behaviour of different processors
- Model a solution proposed by architects to automatically protect this Call Stack
- Design and test a hardware implementation of this solution

Desired Skills: Computer Architecture, HDL, Security

[2024-D2] Design of High-Performance and Energy Efficient Data Speculation Memory Subsystem

The increasing disparity of CPU speed and memory response time led to the invention of several prefetching techniques aiming to hide memory latency by exploiting spatial or temporal relations in memory access patterns.

Highly efficient Data Prefetchers require, however, a remarkable amount of storage to avoid misprediction and the consequent cache pollution.

This internship proposal focuses on the investigation of a new approach to increase the efficiency of our latest CPU's memory system design and the implementation of a new Memory Subsystem to integrate in Arm's next generation CPU.

In this internship you will:

- Model in C through Arm's CPU model a new memory subsystem for speculation algorithms.
- Carry investigations to study the different trade-offs between a dedicated storage per each speculation algorithm and a more efficient solution aiming to share resources and reduce the amount of unexploited area.
- Test your final implementation in SystemVerilog in RTL and measure its PPA impact.

Desired Skills: Computer Architecture, C/C++, HDL

[2024-D3] Instruction Fusion techniques

In modern out-of-order CPUs reducing the number of dispatched macro-operation often helps reducing power consumption and increasing IPC. Some instruction patterns allow to fuse multiple instruction in only one macro-operation.

The goal of the internship is to study different patterns and implement the RTL support for them.

In this internship you will:

- Study different instruction pattern to be fused
- Implement the RTL support in different modules of the CPU
- Analyse the performance/efficiency gain

Desired Skills: Computer Architecture, HDL

[2024-D4] Matrix acceleration on processor using state-of-the-art technology corners

Matrix operations are widely used in popular domains such artificial intelligence, machine learning and computer vision, thus are key contributors to next-generation processors on performance, power and area (PPA).

The goal of this internship is to optimize the matrix multiplication accelerator according to state-of-the-art technology corners, to have a high performance, power efficient and robust design.

In this internship you will:

- Study the differences between advanced technology nodes
- Investigate different matrix multiplication algorithms and implementations to have the highest performing based on selected technology corners.

Desired Skills: Multiplier and Adder datapath knowledges, Math, HDL

[2024-D5] Memory Copy improvements on small size copies

Efficiently copying large portions of memory is a complex task for modern CPUs. Optimize some special cases can helps both performance and efficiency.

The goal of this internship is to familiarize with memory copy techniques and optimise some of them.

In this internship you will:

- Study the microarchitecture of various part of the CPU, from decoder to memory sub-system
- Familiarize with memory copy techniques and find ways to optimize them
- Implement in RTL those techniques

Desired Skills: Computer Architecture, HDL

[2024-D6] Memory system critical load predictor

Power efficiency is a key factor in the design and implementation of CPU; as complexity increases, we have to find more specific critical scenarios. Memory system (among other things) implements specific pipelines for load execution; this kind of instructions have to communicate (broadcast) to the core to wake up dependent resources: this specific wake-up operation is power expensive. One effective way for avoiding this scenario is to forecast if a load, along its execution pipeline, will encounter some slowing event (such as MMU miss, cache miss), and then avoid the broadcast.

The goal of the internship is to design and implement new heuristics for detecting when and how we can avoid a load to wake up core resources.

In the internship you will:

- Find which kind of event can be forecasted
- Model some heuristics for ensuring that the mechanism is not impacting performances
- Implement one chosen method in the RTL for finally measure the power impact

Desired Skills: Computer Architecture, HDL

[2024-D7] Register Cache allocation bandwidth optimization

High performance CPUs often support register caching to reduce power consumption of the Register File's read ports and hide read latency. The allocation bandwidth of the register cache is crucial to increase the probability to have these caches filled with useful data.

The goal of this internship is to study and implement techniques to maximize the allocation bandwidth of the register cache.

In this internship you will:

- Study the issue queue module and register caching mechanisms in a state-of-art uarch
- Implement in RTL different techniques of allocation/eviction of register caches
- Analyse performance and PPA

Desired Skills: Computer Architecture, HDL

CPU Physical Implementation

To apply for this offer:

- Head to https://careers.arm.com/.
- Search for offers with keyword “Internship” and location “France”
- Click on offer “CPU Physical Implementation Intern”

[2024-I1] Exploring Power Improvement features of advanced physical implementation flow

Thanks to CPU architecture and advance technology nodes enhancement, performance is increasing but on the other side power consumption is going higher.

The goal of this internship is to explore, enable new power reduction technics to ensure that we are implementing our architecture in an optimized way according to a given technology node.

You will first start to investigate if our current implementation flow is correctly configured to get the best power consumption for our given CPU architecture. You will then try power reductions technics and EDA features to get additional power benefit exploring various techniques:

- Activity based optimization
- Clock tree optimization and multibit cells
- Layers/vias adjustment
- Threshold voltage/drive size adjustment

The candidate should ideally have a good understanding of implementation flow basic knowledge, power/frequency trade-off in CMOS design and preferably knows TCL or similar language.

Desired Skills: Physical implementation, Performance/Power/Area

[2024-I2] HDL processor power and frequency continuous refinement investigation

Improving power and frequency of microprocessor Soft IP requires continuous refinement of its HDL description and careful monitoring of incremental modifications.

The goal of this internship is to explore and improve some in-house solutions and compare with EDA emerging proposals.

During this internship, you will:

- first familiarize yourself with physical implementation flow and RTL variants comparisons
- look at in-house tool for HDL variant diff capabilities, propose and implement improvement
- evaluate EDA solutions versus in-house tools
- depending on internship progress, you will extend analysis to dynamic activity/power comparison between variants

Proactive and organised, you have basic understanding of HDL, knows TCL or similar language and scripting with Python.

Desired Skills: Physical implementation, Performance/Power/Area

[2024-I3] Advanced signoff optimization exploration

Signoff optimization is an essential part of any modern physical implementation flow. It enables to solve the last timing violations and power reduction using accurate although runtime expensive algorithms

this internship intends to investigate a new way to do signoff optimization, promising a better quality or result for a lowe runtime

You will:

- first start by familiarizing yourself with our physical implementation flow and standard signoff optimization techniques
- then you will put in place a new methodology taking advantage of hierarchical partitioning of a processor
- Compare various timing, power, area and runtime metrics and conclude on interest of this methodology
- if proven beneficial you will integrate this newly developed script in a shared implementation flow environment

Organised and driven, you have some knowledge of power/frequency/area trade-off in CMOS design. You are not afraid of scripting in TCL or similar language and will interact on a regular basis with tool provider support and R&D

Desired Skills: Physical implementation, Performance/Power/Area

[2024-I4] CPU Memory BIST Analysis Investigation

Testability is key to get working silicon, especially on memory. Optimal memory testability is usually obtained thanks to Memory Built In Self Test (MBIST).

This internship aims at evaluating impact of MBIST insertion on a real CPU

- You will first discover a standard CPU implementation flow from synthesis to signoff
- then you will ramp up on MBIST concepts and understand our MBIST insertion flow
- Once MBIST inserted, you will have to extract timing, area, and power both in functional mode and in MBIST mode, including IRdrop analysis
- a comparative report with and without MBIST will have to be produced
- Depending on progress though the internship, a comparison with a different MBIST implementation will be considered.

Curiosity, Proactivity, Organisation, Basic understanding of HDL will be needed for this internship

Desired Skills: Design for test, Performance/Power/Area

CPU Verification

To apply for this offer:

- Head to https://careers.arm.com/.
- Search for offers with keyword “Internship” and location “France”
- Click on “CPU Verification Intern”

[2024-V1] Auto-adjust test selection via coverage feedback loop

Functional coverage is one of the most important metrics used during the verification process of a design.

It helps the verification team to be confident that all functionalities of the design under test are being stressed.

On the other hand, fine-tuning the testbench to have the best coverage values can be frustrating given the complexity of a modern verification environment and the huge number of parameters.

The goal of this internship is to study how we can improve this process by creating a loop which auto-adjusts the testbench based on the status of the coverage database.

In this internship you will:

1. Get familiar with the team’s testbench and simulation environment
2. Implement an API to read coverage databases
3. Feedback the coverage result to the testbench for test selection
4. Make the testbench adjust the test selection based on the feedback

Desired Skills: Python, C++, SystemVerilog

[2024-V2] Optimization of a data structure consumed by a Machine Learning algorithm

Verification is an important part of CPU design, and a Machine Learning flow has been implemented to increase stimulus quality of tests launched.

The goal of the internship is to improve the data structure used to train the model. To speed up the process of candidate generation from the flow and allow for a quicker and more efficient model creation.

In this internship you will:

1. Ramp up on a complex testbench
2. Analyse the existing data structure
3. Interact with the Machine Learning team to find area of improvements
4. Suggest solutions to improve these areas
5. Implement modifications chosen with the internship tutor.

Desired Skills: Python, C++

[2024-V3] Study and improve stress stimuli of test generator

RIS tools, which are verification software generating embedded smart stressful stressing self-checking tests, are an essential step of a CPU verification process. They are usually adopted before the CPU is physically implemented, but they can still find hardware bugs on manufactured silicon.

GenASM-MP is a RIS tool capable of generating and running real assembly code on numerous testing platforms including simulation, FPGA and even silicon.

The scenarios and coverage of the generated tests is growing every year and reaching a complexity difficult to perceive and analyse for a human, making it difficult to understand and detect the possible missing areas, performance issues.

The goal of this internship is to create a visualization tool that will help the user understand what a test is really doing, before expanding the possibilities of the tool by incorporating new scenarios and feature to cover the detected areas of improvement.

In this internship you will:

1. Study the Arm RIS tool and its usage for CPU verification
2. Discuss with the internship tutor about means to analyse the tests execution performance and improve debug-ability
3. Implement the selected solution
4. Analyse the results
5. Implement new features that covers hole discovered with the visualization solution

Desired Skills: Python, C++

[2024-V4] Prototype a hardware-execution monitoring framework

The goal of the internship is to create a hardware monitoring capture solution. This tool will be used to sample key events happening during RTL testing (in simulation, emulation, FPGA). The tool will be written in Python. The users will be able to specify the signals to sample as well as optional processing of the signals. The tool will help the teams monitoring the RTL execution flow.

In this internship you will:

1. Study the existing monitoring tools used in verification
2. Implement a Python framework that enables teams to sample key RTL signals
3. Test the framework in several RTL testing environments (simulation, emulation…)
4. Document the usage of the tool

Desired Skills: Python, C++

CPU Microarchitecture Modelling & Performance Analysis

To apply for this offer:

- Head to https://careers.arm.com/.
- Search for offers with keyword “Internship” and location “France”
- Click on “CPU Performance Analysis and Modelling Intern”

[2024-P1] – CPU performance analysis – statistical clustering

Performance analysis of modern processors is becoming increasingly challenging due to their complexity and the quantity of metrics to review.

The goal of this internship is to explore new methodologies to identify the performance bottlenecks in the CPU, based on statistical clustering methods applied to our profiling data.

During the internship, you will be exposed to the entire performance analysis workflow: running benchmarks on microarchitectural CPU models, understanding key metrics in the microarchitecture and improving the way to analyze them.

Desired skills: Python, Analytic skills

[2024-P2] – CPU microarchitecture– criticality-based cache replacement policies

Cache memories play a critical role in modern CPU performance. Cache replacement policies aim at optimizing the use and efficiency of this scarse resource.

The goal of this internship is to investigate the potential benefits of criticality-based policies, a specific family of cache replacement algorithms.

Your internship will start with a biblio study of existing algorithms, before implementing and tuning a chosen solution in an existing CPU microarchitectural model, and assess its potential benefits on a wide range of benchmarks.

Desired skills: CPU microarchitecture, C programming, Analytic skills

[2024-P3] – CPU microarchitecture exploration assisted by genetic algorithms

The number of interdependent parameters in a CPU design is exploding, making it impossible to exhaustively explore all possible options.

The goal of this internship is to adapt a dedicated framework, based on genetic algorithms, to help and explore the design space more efficiently while adjusting the many configurable parameters.

Once familiar with our internal CPU microarchitectural model and its surrounding infrastructure, you will study and improve the existing genetic algorithm framework to optimize a given microarchitecture across multiple dimensions (Performance, Power & Area).

You will then use the tool and demonstrate its efficiency by tuning specific microarchitectural areas like data prefetchers.

Desired skills: Python, Analytic skills, CPU microarchitecture

Productivity Engineering (PE) Hardware

[2024-E1] Containerise EDA compilation task to run on Cloud

Functional verification is the biggest consumer of cluster in the company. Past years we have been able enable the simulation and coverage merge stages to run massively parallel in the Cloud. We have used Cloud native technologies like Docker or Object Stores. We have solved lot of problematics to make it happened.

The goal of this internship is to enable the primary stage of this flow to run in a similar way. This step actually has lot of implication with the user working environment which makes it more exciting! Such application is really valuable since we will be able to dispatch the task more easily on any platform, region ate low cost with high efficiency.

In this internship you will:

1. Get familiar with the team the BlkVal flow, the EDA tools and the Cloud approach
2. Work on our current demo to get a POC (Proof of Concept) for one EDA tool
3. Port the same to the 2 other EDA tools as well
4. Select a lead partner and make sure the solution is viable

Desired Skills: EDA tools, software programming, scripting, Python, Cloud

[2024-E2] Constraint Programming

A configuration space is defined with parameters (boolean, enum, integer) that have constraints.

A constraint may involve multiple parameters, and a parameter may be involved in multiple constraints.

In order to implement/verify/document the configuration space, it is required to be able to solve all constraints and produce sets of parameters. A configuration space may be hierarchical and may involve multiple hierarchy of instances, each instance comes with its own configuration space.

The goal of the internship is to use various algorithm to split the complete problem resulting of all the parameter constraints into independent problems, to be able to explore a reduced set of parameters.

In the internship, you will:

- research the most efficient problem to groups related parameters based upon their constraints.
- research the most efficient algorithm/library for hierarchical constraint satisfaction and optimization.
- develop a solver resolving valid parameter values for each set (n-wise, exhaustive).
- implement coverage metrics to quantify the quality of the value set.

Desired Skills: Data-Structure, Python, Object Oriented Programming, Test Driven Development

Central Technology (CT) CPU

To apply for this offer:

- Head to https://careers.arm.com/.
- Search for offers with keyword “Internship” and location “France”
- In the list select the offer specified in the description below

[2024-T1] Device Performance Analysis

Offer name: “Device Performance Analysis Intern”

CPU development and validation rely on simulation and FPGA. However, CPU performance study on silicon also heavily contribute to the continuous improvement of the simulation accuracy and target performance coverage. For this reason, a CPU performance and its sensitivity are regularly screened through benchmarks and more standard workloads running on multiple real devices or evaluation boards. In order to produce reliable data and draw meaningful conclusions, this assessment shall be as deterministic as possible, by notably ensuring the device starts and returns to a state (CPU operating point control, memory status, power modes control, thermal conditions, OS framework settings, tracing capabilities ...) which yields the same performance level. This internship will consist of defining/scoping this device state and developing means to force this state and control it for all performance use cases considered today.

Responsibilities:

- You will have to understand the important parameters that influence the performance of a given system, study their influence and understand how to control them.
- Then, you will be in charge of defining and developing a SW tool to control most of these key operating systems and devise parameters that can influence the performance simulation results (this should be applicable to various systems).
- You will work with real Android devices, learn how to remotely run applications or your own benchmark on them, how to profile these SW applications using profiling tools like Perfetto...

Desired Skills: Python skills, General knowledge of phone sw/hw architecture, CPU microarchitecture

[2024-T2] CPU Microarchitecture Characterization

Offer name: "CPU Microarchitecture Characterization Intern"

Micro-benchmarks are a common tool to understand the underlying micro-architecture of a physical device. This encompasses getting intelligence on CPU pipeline, memory hierarchy, virtual memory behavior. The goal of this internship is to develop infrastructure and tests to characterize Android devices. Emphasis will be placed on understanding how real device constraints (e.g. OS behavior) impact the visibility of micro-architectural properties.

Responsibilities:

- You will learn the details of CPU micro-architecture key structures and the underlying algorithms used. For example: branch history or prefetchers or caches...
- You will then define and develop, together with the team, an infrastructure and some new micro-benchmarks to be able to understand the detailed structure of given micro-architecture aspects.
- You will then test these on real devices and ensure the portability of the micro-benchmarks that you have developed.

Desired Skills and Experience: Assembly, Python, C/C++, Basic understanding of cpu pipeline, Basic OS understanding (Virtual Memory)

Solution Engineering (SE) Design Enablement (DE)

[2024-DE1, DE2] Implementation workload optimization

Building the latest generations of processors and systems on the most advanced technology nodes requires always increasing compute resources.

The goal of this internship is to benchmark, analyse and optimize workload associated to processors and systems implementation to define best models and improve overall computing cost while maintaining the required turnaround time.

During this internship, you will:

1. Analyse the workload and computing required to build latest generation processors and systems
2. Benchmark different solutions and trade off (multi host multi core, cloud spot instance and on demand, Arm architecture and x86, EDA tools flexibility etc…) on real design cases
3. Document the model considering turnaround time and cost
4. Define guidelines and deployment recommendation for the best practices

Desired skills:

- Great curiosity, energy and willingness to learn
- Strong interpersonal and teamwork skills
- A keen curiosity about CPU, GPU, and digital systems.
- Exposure or knowledge on synthesis, place and route tools and techniques.

SE Memory (MEM)

[2024-MEM1] Low power design solutions

The continuous demand for higher clock frequencies combined with large capacitive load of a clock tree has led to large power dissipation within the clock network. In some applications, the clock network can draw up to 40% of the total power.

During the internship you will:

1. Start with a state-of-the-art review for deep understanding of the following:
   a. Clock Tree Synthesis (CTS)
   b. Main characteristics and constraints of clock signals
2. Design and simulate at schematic level a first low-power clock driver cell
3. Simulate a clock path using the designed low-power clock driver
4. Evaluate the benefits of the designed low-power clock driver on power consumption.

Desired Skills: Circuit theories, Mixed-signal design, EDA tools, Scripting

[2024-MEM2] Strengthening of memory IPs design for automotive applications

The aim of this internship is to explore, establish, new ways to further strengthen the reliability and robustness of our memory IPs. The expected outcome of the internship is to provide the memory team with tools and methodologies to assess the reliability and robustness of our IPs early in the development process.

During the internship you will:

1. Learn how state-of-the-art memory IPs operate, how they are architected and tested, and ultimately laid out (mask design)
2. Participate to the detection of potential design weaknesses coming from excessive transistor variability and design improvements suggestions
3. Help to improve the detection of insufficient test coverage of memory IP internal components and suggest micro-architecture update to alleviate potential issues
4. Collaborate to the reporting of quality metrics of our memory mask designs and participate to the elaboration of layout guidelines to meet high-quality criteria

Desired Skills: Circuit theories, Mixed-signal design, EDA tools, Scripting

SE Front End Design (FED)

[2024-FED] Cortex X IPC optimisation: Hardware accelerator for external memory access bandwidth and throughput improvement

In the context of the latest generation of test chip integrating latest ARM core generation on the latest technology node (3nm), to explode the maximum IPC, significant memory access improvements are required. Explore technique for efficient memory mapped IO operations to access external devices (DDR, SPI Flash, HyperRAM) while maximizing data throughput. Develop hardware algorithms and processes for parallelizing external memory access in multi core and distributed computing environment.

During this internship, you will:

1. Study state of the art in term of scatter-gather technics and bandwidth improvement through data compression and decompression methods
2. Propose a set of hardware algorithm to be implemented in HDL
3. Provide Models for those set of algorithms
4. Code in RTL and implant those algorithms in current existing design Test Chip
5. Conduct comparative study with simulation RTL, bandwidth improvement measurement, pro and cons for each solution

Desired Skills: Computer Architecture, HDL language

SE Logic (LO)

[2024-LO] Developing a MIMCAP Compiler

Traditional strategies are no more sufficient to address the IR drop problem SoC Designer is facing. MIMCAPs (Metal Insulator Metal capacitances) is one of the potential solution arm is proposing. This is an efficient decoupling capacitor with no area penalty. The goal of this internship is to study the structure and define the best MIMCAP architecture for a targeted application. Then, you will specify and develop a demonstrator of the suggested automation.

During this internship, you will:

1. Evolved as part of a dynamic and diverse team and will directly impact the performance of arm products
2. Study MIMCAP architectures and proposed the best solution for the targeted application in an advanced technology
3. Develop a 'demonstrator' to automate the MIMCAP generation

Desired Skills: Honesty, Curious, Motivated to grow technically

SE Analog/Mixed Signal (AMS)

[2024-AMS1] Adaptive Clocking

Adaptive Clocking (ADCK) is one of the key Analog & Mixed-Signal solution allowing to increase CPU performance by handling fast supply voltage variations and dynamically adapting the clock period. Arm silicon proven IP in one of the most advanced technology nodes is the basis for this work.

The goal of the internship is to simulate and investigate a key evolution of the circuit to allow very small grain frequency modulation, keeping an ultra-low response latency.

During this internship, you will:

1. Evolve in a highly motivated and experienced team in Analog and MS circuits
2. Study the behaviour and merits of existing frequency shifters
3. Simulate and tune the novel frequency shifters created in the team
4. Validate the circuit to place on the testchip

Desired Skills: Analog and Mixed-Signal micro-electronics, Analog / spice simulations, Essential digital circuits

[2024-AMS2] Operational Transconductance Amplifier

Developing efficient Operational Transconductance Amplifiers (OTA) is essential in many analog and mixed-signal circuits such as ADCs. The OTA is one of the key building blocks of Arm IPs involved in advanced chip solutions. Arm OTA circuit in one of the most advanced technology nodes is the basis for this work.

The goal of the internship is to simulate and investigate a key evolution of the circuit to improve characteristics especially of the temperature sensor that uses the ADC.

During this internship, you will:

1. Evolve in a highly motivated and experienced team in Analog and MS circuits
2. Study the behaviour and merits of existing OTA structures
3. Simulate, adapt and tune the evolution of the OTA with target characteristics
4. Validate the circuit to place it on the testchip

Desired Skills: Analog and Mixed-Signal micro-electronic circuits, Analog / spice simulations

SE Silicon Enablement Implementation (IM)

[2024-IM] Implementation Physical Design

Meeting Area, Frequency, and Power requirements with a reasonable runtime is one of most meaningful activities we encounter when implementing physical IPs, CPU and GPU cores.

As part of the Silicon Enablement Physical Design implementation team, you will be given ownership of an investigation topic directly related to physical implementation performance push. You will gain the opportunity to learn how implement high performance processors.

We will allocate a dedicated expert engineer to follow the internship development.

During this internship, you will:

1. Analyse timing correlation between the main physical implementation steps (from standard cells placement to detail routing)
2. Correlate critical timing paths between static timing analysis and spice simulation
3. Improve physical implementation setup and recipes to improve timing convergence
4. Quantify the impact of improved timing convergence on max achievable frequency

Desired skills:

- Exposure or knowledge on synthesis, place and route tools and techniques
- Curiosity about CPU, GPU and digital systems
- Excellent attention to detail and creativity