

The CNRS logo consists of the letters 'cnrs' in a white, lowercase, sans-serif font, enclosed within a solid red circle. The background of the slide features a complex network of glowing orange and yellow lines and nodes, resembling a particle detector or a cosmic web, set against a dark, starry space background with a nebula-like structure on the right side.

cnrs

IN2P3
Les deux infinis

Institut national de physique nucléaire
et de physique des particules

Sonder les infinis : des particules au cosmos

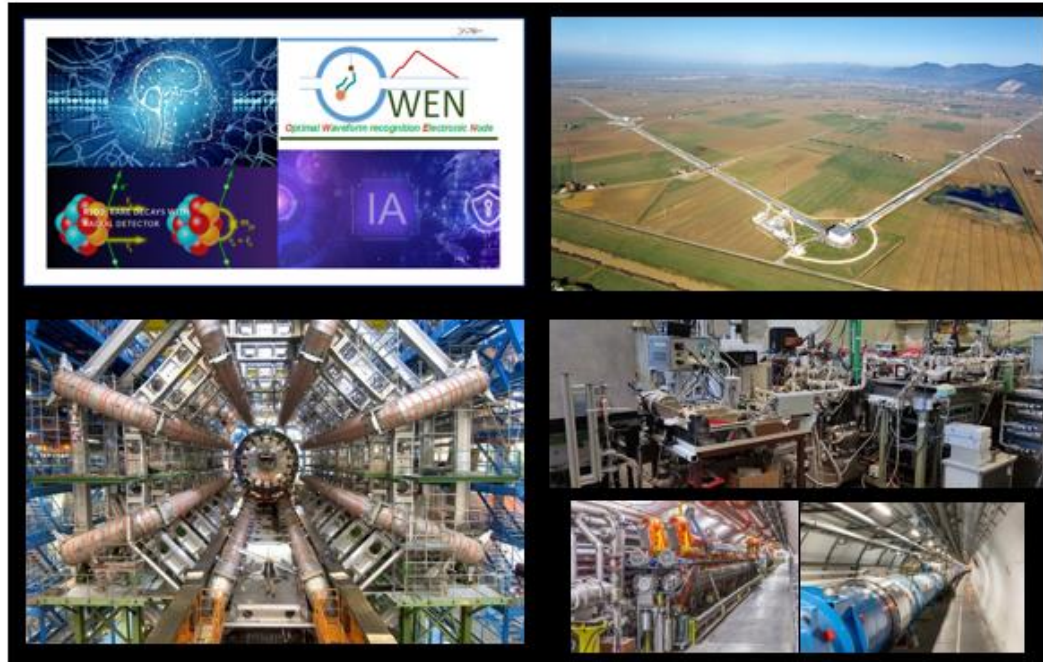
*Entretien Annuel Projet:
Revue de projet*

Projet « THINK »

*M. Frederic Druillole –LP2I Bordeaux
frederic.druillole@lp2ib.in2p3.fr*

Rappel de l'organisation – évolutions éventuelles

L'ambition de THINK est de comprendre et évaluer la possibilité d'implanter des algorithmes de réseaux de neurones relativement en amont pour nos futurs besoins en instrumentation.



LP2I Bordeaux

*F. Druillolle, B. Mansoux,
J. Domange*

LLR Palaiseau

F. Magniette, M. Melennec

CPPM Marseille

G. Aad, R. Bertrand

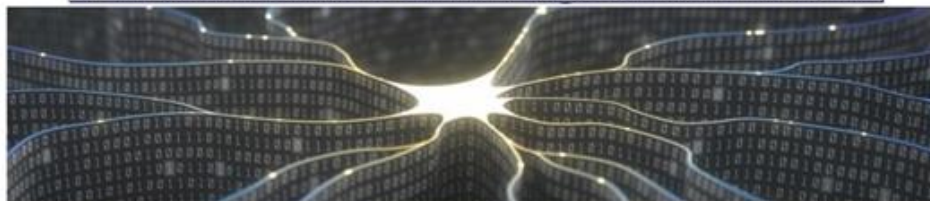
IP2I Lyon

*S. Viret, G. Galbit, Q.
David (2025), X. Chen*

IPHC Strasbourg

*S. Geiger, W. Perrin, M.
Fotzé, J. Wurtz*

Test of Hardware Inferring Neural network



Reseau DAQ – Hardware-Firmware-Software-

13 pers. actives / 23
annoncées

Contexte

- Think Phase 1

2020-2022



Formation des ingénieurs à ces techniques



Faisabilité de portage sur plusieurs types de hardware (succès 3 sur 4)



Exploration des problématiques solvables par DNN



Exploration des performances



Exploration des outils

2023



Restitution d'informations à la communauté

THINK Phase 2



FASTER V3.0 avec CPU+GPU (LPC Caen)
Synchro WR Idrogen (IJCLab)

BILAN - Objectifs

3 personnes de Think phase 1

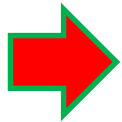


Objectifs différents

5 personnes en formation

Les cours de CosmoStat, AIM (Joana Frontera-Pons)

Les cours FIDLE



Recherche des modèles de DNN pour répondre au besoin d'expériences dans nos instruments.

Déploiement d'exemples en openSource sur Git

Mise en commun des connaissances et de l'expérience de chacun

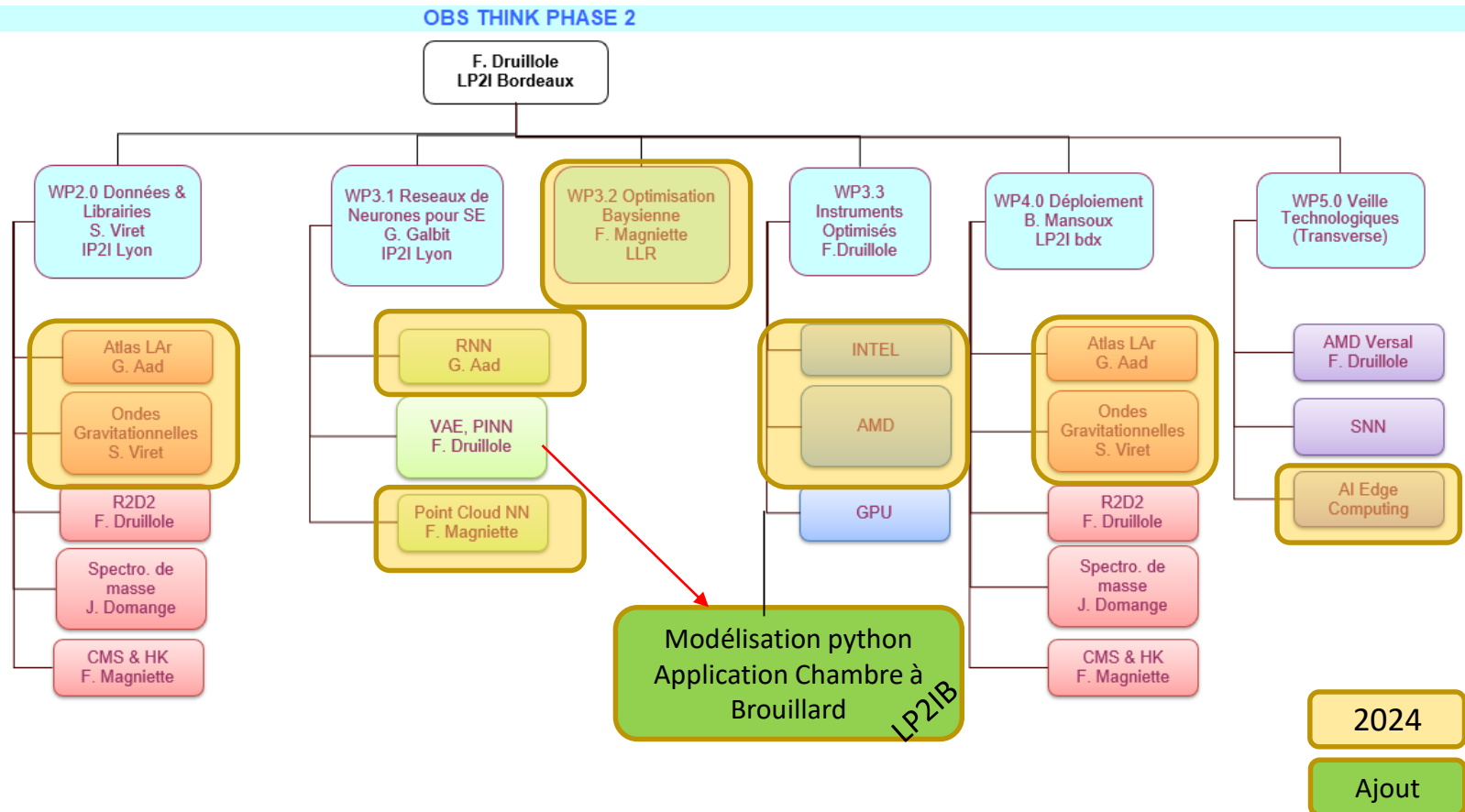
2024: Organisation & Mise en œuvre des outils

- 1 visio par mois (8 / 10 mois)*
- 1 stage au LP2IB financé par Univ Bdx*

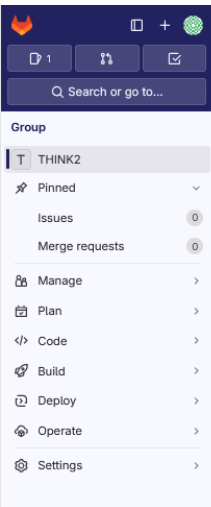
Organisation du (Master) Projet (Bilan)

4 Thèmes applicatifs

Thèmes	Responsable	Labos
<i>Les particules de hautes énergies</i>	George Aad	CPPM
<i>La physique des neutrinos</i>	Frédéric Druillole	LP2I Bordeaux
<i>Les ondes gravitationnelles</i>	Sébastien Viret	IP2I Lyon
<i>L'instrument optimisé</i>	Jocelyn Domange	LP2I Bordeaux



DIFFUSION, FORMATION: Git en place



THINK2

Search or go to...

Group

- THINK2
- Pinned
- Issues
- Merge requests
- Manage
- Plan
- Code
- Build
- Deploy
- Operate
- Settings

THINK2

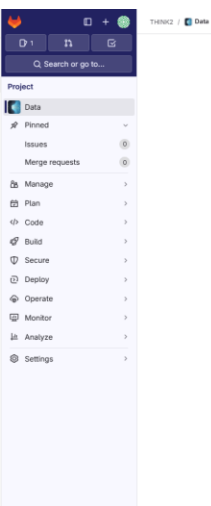
THINK2

New subgroup New project

<https://think.in2p3.fr>

Subgroups and projects Shared projects Inactive

Search (3 character minimum)		Name	
Data	0		4 months ago
Models	0		3 weeks ago

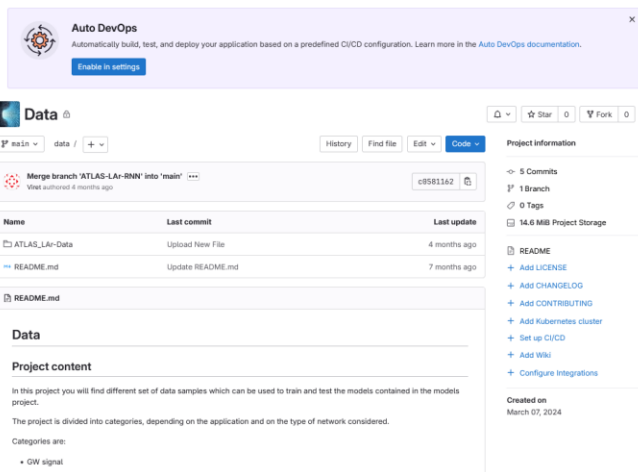


THINK2 / Data

Search or go to...

Project

- Data
- Pinned
- Issues
- Merge requests
- Manage
- Plan
- Code
- Build
- Secure
- Deploy
- Operate
- Monitor
- Analyze
- Settings



Auto DevOps

Enable in settings

Data

History Find file Edit Code

Merge branch 'ATLAS-LAr-RNN' into 'main'

Name	Last commit	Last update
ATLAS_LAr-Data	Upload New File	4 months ago
README.md	Update README.md	7 months ago

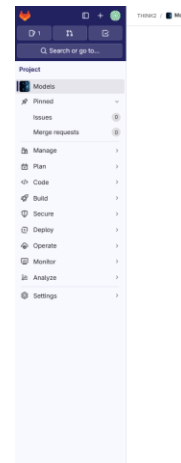
Project information

- 5 Commits
- 1 Branch
- 0 Tags
- 14.6 MB Project Storage

README

- Add LICENSE
- Add CHANGELOG
- Add CONTRIBUTING
- Add Kubernetes cluster
- Set up CI/CD
- Add Wiki
- Configure integrations

Created on March 07, 2024

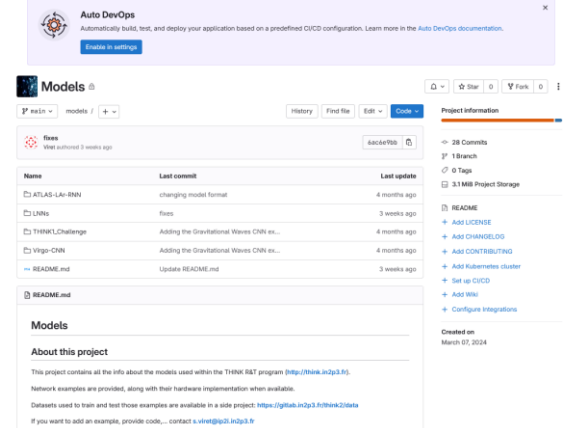


THINK2 / Models

Search or go to...

Project

- Models
- Pinned
- Issues
- Merge requests
- Manage
- Plan
- Code
- Build
- Secure
- Deploy
- Operate
- Monitor
- Analyze
- Settings



Auto DevOps

Enable in settings

Models

History Find file Edit Code

Name	Last commit	Last update
ATLAS-LAr-RNN	changing model format	4 months ago
LNNS	files	3 weeks ago
THINK_Challenge	Adding the Gravitational Waves CNR es...	4 months ago
Virgo-CNR	Adding the Gravitational Waves CNR es...	4 months ago
README.md	Update README.md	3 weeks ago

Project information

- 28 Commits
- 1 Branch
- 0 Tags
- 3.1 MB Project Storage

README

- Add LICENSE
- Add CHANGELOG
- Add CONTRIBUTING
- Add Kubernetes cluster
- Set up CI/CD
- Add Wiki
- Configure integrations

Created on March 07, 2024

Les Frameworks d'inférence des modèle d'IA

AI Edge computing

- **Ax8i (max78000)**
- **ST
Microelectronics**
- CubelIDE + Cube-AI
- **VITIS AI (AMD)**
- **OpenVino (Intel)**
- **N2D2 –
DeepGreen
Eclipse Aidge**
- **Edge Impulse
(société)**

Spatial Accelerators

- **VIVADO HLS**
- HLS4ML
- FiNN (Xilinx R&D)
- ChaiDNN
- **HADDOC2**
- **DNNWeaver**
- ...

→ Framework servant à réaliser l'accélération des calculs des modèles IA pour l'inférence.
Utilise des modèles éprouvés
Difficile de créer des inférences sur des composants à faibles ressources (multi-voies, proche des capteurs)



Critères de comparaison des matériels

Critères	CPU	CPU+AI-CoPro	GPU	FPGA	ASIC
Adaptabilité (à une variété de modèles)	Elevé	Elevé	Moyen	Faible	Aucune
Puissance de calcul	Faible	Moyen-Elevé	Elevé	Elevé	Moyen
Latence	Moyen (ms)	Moyen (ms)	Faible (μ s)	Faible (10 ns)	Très faible
Flux d'entrée	Faible	Moyen	Elevé	Elevé	Elevé
Parrallelisme	Faible	Moyen	Elevé	Elevé	Elevé
Efficacité de la consommation électrique	Moyen	Moyen	Moyen	Moyen	Elevé
Mise en Oeuvre	Facile	Moyen	Moyen	Complexe	Très Complexe
Emconbrement/Densité des NN	Moyen	Faible	Faible	Elevé	Elevé

Savoir-Faire



Le Monde de Think Phase 2

1 Mparam

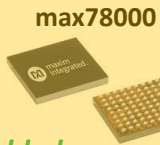
New 2024

Edge AI

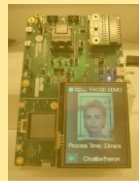
Embedded system



ESP32



max78000



stm32



FPGA/ SoC



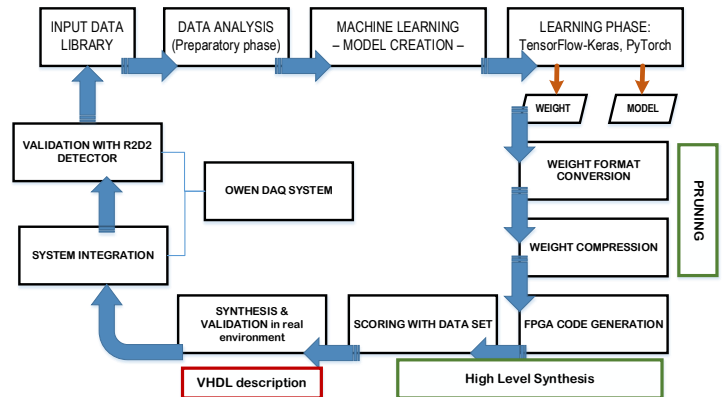
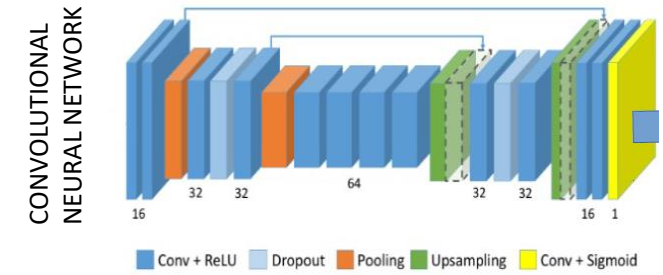
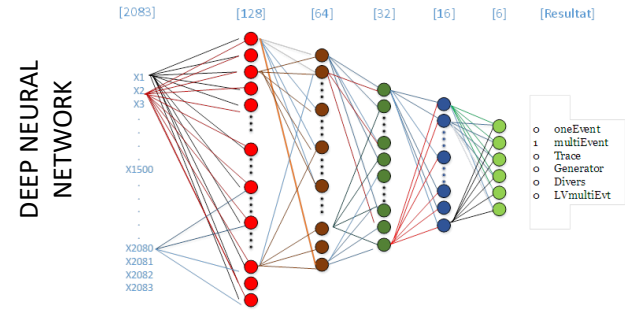
Puce neuromorphique

1 Gparam

GPU



DGX A100



Assemblée Générale le 7 et 8 Octobre 2024 au LP2I Bordeaux:

→ Travaux Pratiques sur Zynq SoC et HLS4ML

Diffusion des modèles sur Lar et OG

Formation et diffusion des modèles

Bilan des Matériels pour l'IA embarqué (Think phase 1)

Optimisation matériels sur Stratix et Agilex

Connaissances validées

Etienne Marie Fortin expertise IA sur Intel

→ *OK pour venir au CPPM*

Optimisation Bayésienne des Réseaux de Neurones

Stratégies d'entraînement

Les convolutions de Graphe et les modèles pour la physique des particules

Les Liquid Neural Network et les modèles pour la physique

Mise en commun des savoir-faire

Appels à projets

ANR 2025: OSCARI (soumis) → R&T ADC ?

ERC 2024: Novelty@LHC (soumis)

ANR 2026: FASTLIQUID

ACTION DE COMMUNICATION & RESEAU

ARTIFACT: ARTifical Intelligence For Accelerators, user Communities and associated Technologies

HORIZON-INFRA-2024-TECH-01-01



Echec

A. Ghribi

Organisation CNRS/Univ/EPIC d'un réseau

M4CAST

(Multiphysics Modelling, Machine learning and Model-based Control in Accelerator Sciences and Technologies) is a French network dedicated to applications of artificial intelligence to accelerator sciences and technologies.



Conseil & Expertise sur 2 projets INSU:

Advanced-Plasma
[LOA/Photons] Univ. Lille
GeoSciences: Toulouse-inp &
Univ. Rennes

Séminaire à l'IRAP Toulouse le 14 Octobre



ENSEIRB-MATMECA
Option SE:
IA embarqué (12h)

ANF : Aquisition de données intelligentes

- Modélisation avec python (génération de données)
- Initiation à l'IA embarqué sur Zynq

Bilan Financier & ETP

Types Budget	2024	2025
Equipements	5000€	5000€
Fonctionnements	2500€	2500€
Missions	20000€	20000€ 10k€
Dépenses effectives		
Demandé	27500€	12500€
Obtenu	10000€	
Dépensé	7000€	
À terminaison	0€	
<i>Achat RP</i>	<i>1000 €</i>	
<i>Stage LP2IB</i>	<i>2700€ (3 mois)</i>	

Investissement
Matériel

2 AG par an

1 AG par an

Investissement
Matériel → 3k€
Fin d'exercice

Bilan Financier & ETP

Laboratoires	RH	ETP 2024	Tâches	ETP 2025	Tâches
LP2IB	Jocelyn Domange	0,05	Formation & appli Chambre à Brouillard	0,1	finalisation appli chambre, test pour appli débruitage sur spectres PIAGARA
	Bruno Mansoux	0,1	Formation	0,1	Déploiement Modèle LP2IB
	Frédéric Druillole	0,3	Management, Test sur CPU+CoPro, Essai sur FPGA Z020 et ARTY avec HLS4ML. Travail sur la modélisation de chaine analogique pour la génération de données dans le cadre de R2D2	0,6	Management, modèle deconvolution, débruitage et GenAI pour Actar et R2D2, Chambre à brouillard, publication et communication externe.
IP2I	Sébastien Viret	0,2	Mise en œuvre Github, Veille techno LNN, gestion R&D OG	0,3	
	Geoffrey Galbit	0,2	Formation/veille techno - Début du développement du démonstrateur de portage de réseau machine learning vers rtl (library hdl, scripts de transformation, simulation et synthèse avec flow Quartus Pro).	0,5	finaliser le démonstrateur complet (jusqu'à l'implémentation sur cible Stratix 10) et enrichissement de celui-ci avec la prise en charge de différentes architectures et types de NN.
	Quentin David	0		0,1	Formation & suivi projet

Bilan Financier & ETP

Laboratoires	RH	ETP 2024	Tâches	ETP 2025	Tâches
IPHC	Sébastien Geiger	0,2	Formation, acculturation	0,5	Apprentissage et création de modèles NN
	Willy Perrin	0,2		0,2	FORMATION, test de descente IA sur FPGA
	Marcellin Fotze	0,2		0,2	FORMATION, test de descente IA sur FPGA
	Jacques Wurtz	0,05	Gestion et suivi de Think	0,05	Gestion et suivi de Think
LLR	Frédéric Magniette	0,9	Travaux convolutions de graphe Adaptation pour implémentation hls	0,9	Gcn jusqu'en 2026 Implémentation xilinx prévu 2025/26 Etp stables pour l'année prochaine 1.8 ETP Une publication en 2025 dans les actes de la conférence Chep24
	Mathieu Melenec	0,9		0,9	
CPPM	Georges Aad	0,05	Supervision de Raphael	0,05	Supervision de Raphael
	Raphael Bertrand	0,2	Optimisation des réseaux pour l'application LAr ATLAS notamment en remplant des couches RNN par des couches denses. Fournir (sur le git THINK) un réseaux + données pour exercer différentes implémentations sur FPGAs.	0,2	Optimisation des réseaux pour l'application LAr ATLAS.

ETP 2024: 3,55



Prévu: 2,75

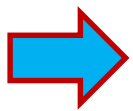
ETP 2025: 4,7



Prévu: 2,85

Points spécifiques à discuter

- THINK Phase 2 2025:
 - ➔ Connaissance maîtrisée:
 - ➔ Publication CGN
 - ➔ Publication LaR
 - ➔ Publication Hardware for Physics instrumentation (THINK phase 1)
- Financement de stages non fléchés:
 - ➔ exemple: D.U finance à hauteur de 1500 € (40% d'un stage ingénieur)
- R&T et EL2I: Management



Recherche de modèles efficients sur FPGA pour nos instruments
Transposable sur ASIC