



CSC\_5RO06\_TA/PROJET/2024/v.1

## Cours CSC\_5RO06\_TA

Accélérateurs matériels pour l'IA et la robotique

### Projet 1

Hardware-Software codesign

(Hw-Sw codesign)

Omar Hammami

[omar.hammami@polytechnique.edu](mailto:omar.hammami@polytechnique.edu)

Hervé LeProvost

[herve.le-provost@cea.fr](mailto:herve.le-provost@cea.fr)



hammam@ensta.fr

## Description

Nous souhaitons dans ce projet évaluer les performances en temps d'exécution et en ressources d'une fonction de multiplication de matrices largement utilisée en IA embarquée et traitement d'images/vision en robotique dans 3 configurations possibles : 1. sur processeur généraliste disponible sur PC (Intel, AMD) 2. sur processeur embarqué ARM9 3. en accélérateur matériel sur circuit reconfigurable FPGA XC7Z020 disponible sur carte [zedboard](#).

L'objectif étant de sélectionner la configuration présentant le meilleur compromis performance-cout pour le contexte applicatif.

Nous utiliserons pour ce projet une carte zedboard dans la configuration ci-dessous ou équivalente

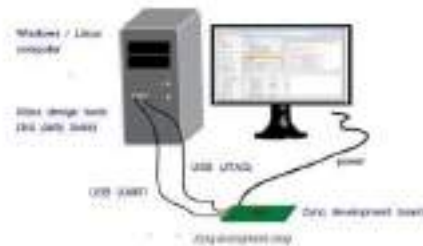


Fig1 - environnement de conception et d'évaluation de performances

La fonction F1 sera donc évaluée selon les méthodes usuelles sur PC et dans les 2 configurations suivantes :

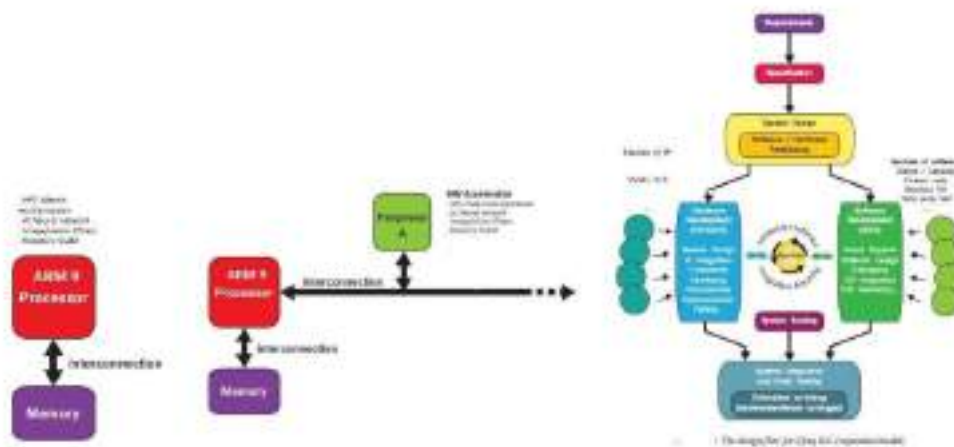


Fig2 – (a) fonction F1 en logiciel embarqué sur processeur ARM9 (b) fonction F1 comme accélérateur matériel sur circuit FPGA Xilinx Zynq (c) flot hw/sw codesign

La fonction choisie sera codée en C/C++ et vous définirez un jeu de tests (ensemble de données de tailles variables) pour la validation en considérant les contraintes mémoires de la zedboard (512 MB DDR3 –4 GB SDcard).



CSC\_FIDI\_IA/PROJET/2024/v.1

9/17/24, 11:43



CSC\_PRCM\_TA/PROJET/2024/v.1

barhami@ensta.fr

### Q3. Estimation de performances et de ressources par accélérateur matériel sur circuit FPGA

Vous concevrez un accélérateur de l'application précédente depuis C/C++ en utilisant l'environnement Vivado 2019 HLS et générerez une synthèse de ce circuit.

Vous explorerez différentes directives d'optimisations pour la synthèse HLS et noterez les temps de latence, les débits et les ressources matérielles utilisées du circuit XC7Z020 (logic cells (**% de 85K**), LUTs (**% de 53200**), Block RAM (**% de 4,9 Mb**), DSP slices (**% de 220**)).

Vous tracerez des courbes d'estimation de performance (temps)/ressources du circuit seul pour l'ensemble des configurations sélectionnées (algorithme, code, options de compilations). Confirmez que la courbe a la forme d'une courbe de Pareto au sens de l'optimisation mathématique multiobjective.

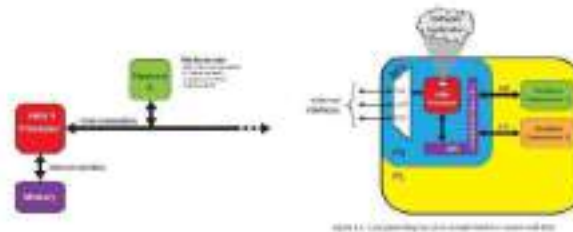
### Q4. Mesures de performances par accélérateur matériel sur circuit FPGA

Vous implémenterez en utilisant Vivado 2019 votre accélérateur sur circuit Zynq et le connecterez au processeur ARM9 en utilisant le réseau AXI.

Vous explorerez différentes directives d'optimisations pour la synthèse HLS et noterez les temps de latence, les débits et les ressources matérielles utilisées du circuit XC7Z020 (logic cells (**% de 85K**), LUTs (**% de 53200**), Block RAM (**% de 4,9 Mb**), DSP slices (**% de 220**)).

Vous tracerez des courbes de performance **par exécution** (temps)/ressources du circuit complet pour l'ensemble des configurations sélectionnées (algorithme, code, directives de synthèse et d'optimisations).

Confirmez que la courbe a la forme d'une courbe de Pareto au sens de l'optimisation mathématique multiobjective.



#### Références

- [Vivado Design Suite User Guide High-Level Synthesis - UG902 \(v2019.1\) July 12, 2019](#)
- [Vivado Design Suite Tutorial High-Level Synthesis](#)
- [Vivado HLS Optimization Methodology Guide UG1270 \(v2017.4\) December 20, 2017](#)

### Q5. Optionnel : Automatisation (TCL) et parallélisation (+5 pts)

Le travail précédent peut être automatisé et permettre ainsi une exploration plus large de l'espace de conception (**DSE - Design Space Exploration**). Cette automatisation permet d'améliorer la productivité de conception de circuits complexes (SOC – System on Chip) de manière considérable.

La base est d'utiliser des scripts TCL pour la synthèse et l'implémentation. Le logiciel Vivado 2019 est installé sur l'ensemble du réseau de l'ENSTA et vous pouvez lancer en parallèle N copies de Vivado permettant d'explorer un nombre très important de configurations de SOC.

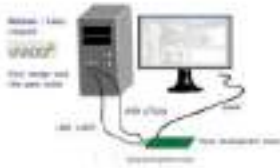




ENSTA

15 rue de la

hammam@ensta.fr



#### Reference

- Vivado Design Suite User Guide Using Tcl Scripting - UG094 (v2019.1) May 22, 2019
- [Vivado Design Suite Tcl Command Reference Guide UG035 \(v2019.1\) May 22, 2019](#)
- <https://www.xilinx.com/video/hardware/using-the-vivado-hls-ocl-interface.html>

## Q6. Validation Projet et Rapport

### Q 6.1 – présentation

Vous soumettrez une présentation par groupe de 4 élèves qui répond aux questions précédentes pour le **14/10/2024** (23H59) dans le répertoire eCampus du cours.

Une validation sur carte sera effectuée le **15/10/2024** de 10H00 à 12H00 après votre soutenance.

### Q 6.2 – rapport + fichiers projet

Vous soumettrez un rapport par groupe de 4 élèves qui répond aux questions précédentes et fait la synthèse des résultats obtenus en utilisant le template suivant :

[https://www.ieee.org/publications\\_standards/publications/authors/author\\_templates.html](https://www.ieee.org/publications_standards/publications/authors/author_templates.html)

pour le **14/10/2024** (23H59) dans le répertoire eCampus dédié à votre groupe.

L'ensemble des fichiers de votre projet devront être soumis pour le **14/10/2024** (23H59) dans un drive que vous communiquerez directement aux enseignants.

Toute soumission après cette date entraîne une pénalité de 3pts par jour de retard et toute soumission avant cette date entraîne un bonus de 3 pts par semaine complète d'avance pour un projet ayant une valeur de 14/20 au moins. Pour rappel les modalités d'évaluation sont comme suit : 70% projet 30% examen. La note du projet étant décomposée en : 40 % groupe et 60% individuel.



ENSTA

18750

hammami@ensta.fr

CSC\_PRCM\_TA/PROJET/2024/v.1

## Références

### Applications Xilinx

- « [A Zynq Accelerator for Floating Point Matrix Multiplication Designed with Vivado HLS](#) » - XAPP1170 (v2.0) January 21, 2016 - Author: Daniele Bagni, A. Di Fresco, J. Noguera, F. M. Vallina
- [Zynq All Programmable SoC Sobel Filter Implementation Using the Vivado HLS Tool](#) Author: Fernando Martinez Vallina, Christian Kohn, and Pallav Joshi - XAPP890 (v1.0) September 25, 2012

### Articles de recherche (exemples)

1. HW/SW Co-Design of the HOG algorithm on a Xilinx Zynq SoC. Jens Romkowski ; Andrew Boutros ; Diana Göhrings, Journal of Parallel and Distributed Computing Volume 109, November 2017, Pages 50- 62
2. CNN-Grinder: From Algorithmic to High-Level Synthesis descriptions of CNNs for Low-end-low-cost FPGA SoCs, Panagiotis G. Mousoullos ; Loukas P. Petrou , Microprocessors and Microsystems Volume 73, March 2020, 102990
3. A fast and scalable architecture to run convolutional neural networks in low density FPGAs, Mário P. Véstias ; Horácio C. Neto<sup>1</sup> , Microprocessors and Microsystems Volume 77, September 2020, 103136
4. Design of hand skeleton extraction accelerator for a real-time hand gesture recognition, Seonyoung Lee ; Haengwon Son ; Yunjong Kim ; Kyoungwon Min 2019 International SoC Design Conference (ISODC)
5. The AXIOM Project: IoT on Heterogeneous Embedded Platforms, IEEE Design & Test, 2019.
6. MulMapper: Towards an Automated FPGA-Based CNN Processor Generator Based on a Dynamic Design Space Exploration, Mudun Haillesellasse ; Syed Rafay Hasan ; Osmane Ait Mohamed 2019 IEEE International Symposium on Circuits and Systems (ISCAS)
7. Automatic Energy-Minimized HW/SW Partitioning for FPGA-Accelerated MPSoCs, Gereon Fültr ; Seyit Halil Hamurcu ; Diego Pala ; Thomas Grass ; Rainer Leupers ; Gerd Ascheid ; Juan Fernando Euse IEEE Embedded Systems Letters 2019
8. [Acceleration and implementation of convolutional neural networks based on FPGA](#), Digital Signal Processing Volume 141, September 2023,
9. MVSym: Efficient symbiotic exploitation of HLS-kernel multi-versioning for collaborative CPU-FPGA cloud systems, Integration Volume 93, November 2023
10. [SoC-based real-time SVM classification with integrated training using HLS and PYNQ](#), Microprocessors and Microsystems Volume 101, September 2023
11. Efficient Hardware Implementation of Artificial Neural Networks on FPGA, Kasem Khalil; Tamador Mohalad; Mahmood Darwich; Ashok Kumar; Magdy Bayoumi, 2024 IEEE 6th International Conference on AI Circuits and Systems (AICAS)
12. [iMAC: Image-to-Column and General Matrix Multiplication-Based Dense CNN Accelerator, Accelerators for Convolutional Neural Networks](#), 2024
13. Real-Time Multi-Learning Deep [Neural Network](#) on an MPSoC-FPGA for Intelligent Vehicles: Harnessing Hardware Acceleration With Pipeline, Güner Tatar; Salih Bayar; İhsan Çiçek, IEEE Transactions on Intelligent Vehicles, 2024.

Organisation v.1 2020/09/15

