

Soutenance de thèse

Morgan GAUTHIER soutiendra sa thèse de doctorat, préparée au sein de l'équipe d'accueil doctoral ISAE-ONERA DECISIO et intitulée «*Déploiement haut niveau d'applications automobiles sur les architectures E/E futures*»

Le 3 octobre 2024 à 9h00, salle des thèses, ISAE-SUPAERO

devant le jury composé de

M. Rob VINGERHOEDS	ISAE-SUPAERO	Directeur de thèse
M. Emmanuel CAILLAUD	CNAM	Rapporteur
M. Frédéric MALLET	Université Cote d'Azur	Rapporteur
Mme Claude BARON	INSA Toulouse	
M. Christophe MERLO	ESTIA	
M. Jean-Baptiste CHAUDRON	ISAE-SUPAERO	Invité
M. Phippe CUENOT	Continental Automotive	Invité

Résumé : Pendant de longues années, les besoins croissants en puissance de calcul ont été satisfaits en suivant deux voies principales : l'intégration de mécanismes micro-architecturaux de plus en plus complexes permettant d'augmenter le nombre d'instructions traitées par cycle d'horloge (l'IPC) tels que les caches, les mécanismes de prédiction de branchement, d'exécution spéculative, etc. et, surtout, l'augmentation continue de la fréquence d'horloge des processeurs. Cependant, les contraintes de dissipation thermique – le fameux power wall – rendent désormais cette dernière voie impraticable. L'amélioration des performances est aujourd'hui essentiellement obtenue en multipliant le nombre d'unités de traitement, que ce soit sous la forme de cœurs (architectures multi-cœurs ou many-cœurs), de GPU, d'accélérateurs spécialisés implémenté en silicium ou sur FPGA. Le domaine automobile n'a échappé ni à un accroissement très important des besoins de traitements, ni à l'évolution technologique. Aujourd'hui, l'intégration de fonctions complexes telles que l'assistance à la conduite (ADAS) ou la conduite autonome nécessite une refonte majeure de l'architecture électronique et électrique des véhicules afin de satisfaire conjointement les exigences de temps de réponse, de poids, et de puissance électrique consommée (SWaP), de sûreté de fonctionnement et de coût de développement. Les « Systems on Chip » (SoC) récents sont d'une très grande complexité. Ils combinent des unités de traitement multiples, hétérogènes, et de niveau de déterminisme temporel différents (plusieurs cœurs d'architectures différentes, des unités SIMD, des accélérateurs ML, etc.). Aussi, exploiter de façon optimale ou simplement de manière efficace des plateformes présentant ce niveau de diversité est une tâche extrêmement complexe. Elle requiert de combiner des compétences multiples, très spécifiques et généralement dispersées (architecture logicielle, architecture FPGA, développement parallèle, DSP, IA...), de prendre en compte de contraintes très diverses (ségrégation/co-localisation, temps de calcul, déterminisme, énergétique...), de maîtriser des technologies et outils très variés. Elle implique de réaliser des choix architecturaux justifiés par des méthodes et des moyens techniques. À ce jour ces méthodes et moyens couvrant l'ensemble de ces problèmes n'existent pas ou ne traitent que partiellement le problème. Cette thèse a pour objectif de donner les moyens techniques et méthodologiques permettant à un ingénieur de rationaliser ses choix architecturaux, notamment du point de vue du déploiement d'une application logicielle sur une structure d'exécution matérielle comportant plusieurs cœurs et ou accélérateurs. Bien que formalisé dans un cadre général, le problème sera résolu dans le cadre spécifique des applications et des technologies du domaine automobile. Dans ce but, les travaux porteront conjointement sur (i) la définition claire et rigoureuse du problème d'optimisation architecturale, (ii) la proposition des méthodes et techniques (outillées ou non) permettant de résoudre ce problème. L'originalité de la

thèse réside dans son caractère « holistique » puisqu'il s'agit de prendre en compte l'ensemble des dimensions et contraintes du problème (HW, SW, temps de réponses, consommation, etc.), tout en assurant que la solution proposée restera applicable sur des cas concrets et réels. Il est important de noter que l'on ne cherche pas à automatiser complètement la recherche d'un optimum global sur un problème simplifié, mais à automatiser partiellement la recherche d'une solution (potentiellement sous-optimale) mais sur un problème réel.

Mots clés : Automobile, Système, Algorithme

Summary: For many years, the growing needs in computing power have been met by following two main paths: the integration of increasingly complex micro-architectural mechanisms making it possible to increase the number of instructions processed per clock cycle (IPC) such as caches, branch prediction mechanisms, speculative execution, etc. and, most importantly, the continuous increase in the clock frequency of processors. However, the constraints of heat dissipation - the famous power wall - now make this last route impractical. The improvement in performance is today mainly obtained by multiplying the number of processing units, whether in the form of cores (multi-core or many-core architectures), GPUs, specialized accelerators implemented in silicon or on FPGA. The automotive industry has not escaped a very significant increase in the need for treatment, or technological change. Today, the integration of complex functions such as driver assistance (ADAS) or autonomous driving requires a major overhaul of the electronic and electrical architecture of vehicles in order to satisfy jointly the requirements of response time, weight, and consumed electrical power (SWaP), operational reliability and development cost. the recent 'Systems on Chip' (SoC) are very complex. They combine multiple, heterogeneous processing units with different levels of temporal determinism (several cores of different architectures, SIMD units, ML accelerators, etc.). Also, optimally or simply efficiently using platforms with this level of diversity is an extremely complex task. It requires combining multiple, very specific and generally dispersed skills (software architecture, FPGA architecture, parallel development, DSP, IA, etc.), taking into account very diverse constraints (segregation / co-location, computing time, determinism, energy ...), to master a wide variety of technologies and tools. It implies making architectural choices justified by methods and technical means. To date these methods and means covering all of these problems do not exist or only partially address the problem. This thesis aims to provide the technical and methodological means allowing an engineer to rationalize his architectural choices, in particular from the point of view of the deployment of a software application on a hardware execution structure comprising several cores and / or accelerators. Although formalized in a general framework, the problem will be solved within the specific framework of automotive applications and technologies. To this end, the work will focus jointly on (i) the clear and rigorous definition of the architectural optimization problem, (ii) the proposal of methods and techniques (equipped or not) to solve this problem. The originality of the thesis lies in its "holistic" character since it involves taking into account all the dimensions and constraints of the problem (HW, SW, response times, consumption, etc.), while ensuring that the proposed solution will remain applicable to concrete and real cases. It is important to note that we are not trying to completely automate the search for a global optimum on a simplified problem, but to partially automate the search for a solution (potentially sub-optimal) but on a real problem.

Keywords: Automotive, System, Algorithm