

Sujet de thèse DGA:
Méthodologie d'exploration “consciente de l'architecture”
pour les systèmes multi-cœurs embarqués reconfigurables et adaptatifs

Ecoles doctorale : MSTIC, Université Paris EstN

Nom / label de l'unité de recherche : IGM- LabInfo UMR CNRS 8049

Localisation : Institut d'électronique et d'informatique Gaspard-Monge, Marne la Vallée

Equipe : Algorithmes, Architectures, Analyse et Synthèse d'Images (A3SI)

Sujet proposé par : T. Grandpierre et E. Dokladalova

Contact : T. Grandpierre, (t.grandpierre@esiee.fr), Groupe ESIEE Paris, Cité Descartes, 93162 Noisy le Grand Cedex, tel/fax: 01 45 92 65 53 / 66 99

Contexte :

Depuis quelques années nous sommes devenus les témoins d'un tournant technologique où les grands constructeurs de processeurs (Intel, AMD, Texas Instruments, IBM) s'engagent dans l'exploitation du parallélisme spatial qui avait été abandonné préalablement au profit d'augmentations des fréquences de fonctionnement. Ainsi, pour proposer plus de puissance de calcul, ils misent maintenant sur la multiplication des cœurs homogènes ou hétérogènes de processeurs au sein d'une même puce [ISPA06]. Il est attendu que cet effort s'intensifie dans les années à venir car certaines applications ont besoin d'un rapport d'amélioration de performances de 100x [MPSOC09b].

Ce tournant technologique est d'autant plus marquant que nous observons le même phénomène dans le domaine de l'embarqué. Ici, deux voies sont explorées : les **multi-cœurs embarqués (MPSoC) type multi-GPP, multicore DSP** (BlackFin/AnalogDevice, MSC8154/Freescale Trimédia, C6474/TexasInstrument) et les architectures émergentes à base de **multi-softcœurs** embarqués sur des FPGA (multi-microblaze[ISFPGA09], multi-nios etc.).

De plus en plus, sous la pression des innovations applicatives, nous nous approchons des systèmes embarqués utilisés selon le paradigme du « *ubiquitous* ou *pervasive computing* » où les ressources de calculs sont interconnectés, communicants, mettant en œuvre une interaction avec l'environnement. Ainsi, les concepteurs font appel à des possibilités technologiques récentes [ISFPGA09] pour **réaliser des systèmes (auto)reconfigurables et/ou (auto)adaptatifs**, dotés d'une capacité d'adaptation de leur structure interne pour fournir le maximum de **performances** et la « **qualité de service** » en fonction des situations particulières.

En conséquence, nous disposons de nombreux systèmes type **MPSoC homogènes et hétérogènes**, dotés de **niveaux de parallélisme hétérogènes** (niveau données, instructions, tâches où même des applications). De plus, ces systèmes offrent souvent un certain degré d'adaptation (de reconfiguration) à différentes échelles pour fournir à tous les instants le maximum de performances ; **ainsi nous avons la puissance de calcul disponible sur chip**.

Le déficit d'aujourd'hui et des années à venir est de résoudre le problème de l'exploitation de cette **puissance de calcul et de « l'intelligence du système » par l'utilisateur** car actuellement, il n'existe aucune **méthodologie générale et globale** qui permettrait d'explorer l'espace d'implantations optimales des applications sur des systèmes embarqués MPSoC adaptatifs ou reconfigurables.

Si de nombreuses approches à l'exploration des implantations des applications ont été proposées dans le passé pour des systèmes embarqués [KLUWER98], [ICCAD01], [CODES02] et [JSA08], aucune

n'apporte de solution satisfaisante dans le cadre des nouvelles architectures MPSoC embarquées reconfigurables dans le contexte présent.

Nous proposons de nous inspirer de l'approche utilisée dans le domaine des compilateurs conscients (« aware ») des caractéristiques particulière du système telles que l'organisation de mémoire, la consommation énergétique, les communications, etc. [MPSOC09a], [MAAC06], [DATE06]. Il a été prouvé que l'intégration de cette « connaissance » du système aux méthodologies d'explorations permet d'obtenir des résultats intéressant [DATE98], [JSC07]. Le problème est qu'en générale, une seule caractéristique est intégrée à la fois ou un seul système est ciblé. Alors que pour une méthodologie globale, il faut savoir intégrer toutes les caractéristiques cruciales pour les performances : différents niveaux de parallélisme, communications, organisation de mémoire, niveau de reconfiguration/adaptation et ceci à toutes les échelles.

Objectifs scientifiques de la thèse :

L'enjeu est donc de proposer une méthodologie globale d'implantation optimisée des applications sur des architectures émergentes MPSoC : embarquées, multi-cœurs, reconfigurables ou adaptatives. Cette méthodologie devrait avoir conscience des caractéristiques clés de l'architecture visée et permettre une rapide convergence vers une solution optimisée respectant les contraintes temps réel données.

Verrous technologiques et scientifiques :

Pour implanter efficacement une application, en générale comportant plusieurs algorithmes de complexité variée, sur un système permettant de mettre en œuvre différents niveaux de parallélisme, de nombreuses décisions doivent être prises :

1. partitionner l'application en respectant les contraintes et les caractéristiques du système de calcul utilisé (homogène, hétérogène, organisation des mémoires, ...),
2. communiquer, échanger des données entre les différentes parties, en fonction des moyens de communication disponibles (point à point, passage de message, mémoire partagée ...),
3. définir l'ordre d'exécution,
4. synchroniser les processus,
5. générer, compiler, télécharger le programme de chaque processeur.

Il est évident que le nombre des possibilités à explorer pour exploiter efficacement le parallélisme disponible de ces architectures est quasi-infini. Il est souvent effectué en utilisant le schéma classique de développement : compilation, exécution, profiling de l'exécution; avec de nombreuses itérations successives dont la convergence vers un optimum (respect des contraintes temps réel, « embarquabilité ») est presque impossible à garantir.

Pour proposer une nouvelle méthodologie, permettant d'intégrer des caractéristiques des systèmes, il faudra étudier quatre axes

- 1) Faire état de l'art des différentes architectures multi-softcore actuelles et une projection des systèmes à venir en vue de définir un modèle générique de ces architectures qui permettrait par la suite de systématiser et d'accélérer la tâche d'implantation.
- 2) Définir des métriques utilisées pour l'aide à la prise de décision lors de la distribution de l'application et d'indicateurs d'exploitation du parallélisme présent dans l'architecture.
- 3) Etudier l'implantation de plusieurs applications sur des architectures représentatives choisies dans l'état de l'art (par exemple l'architecture [ISFPGA09]).
- 4) Proposer une méthodologie, voire un outil, permettant l'exploitation optimisée de ces nouvelles

architectures. Ce travail sera validé par l'implantation d'algorithmes de vision développés au sein de notre équipe.

Ces travaux se baseront sur l'expertise approfondie de la méthodologie AAA et des outils SynDEx [MEMO03] et SynDEx-IC [ESD07] que nous développons conjointement avec l'INRIA. Cette méthodologie couvre actuellement le prototypage rapide d'applications temps réel embarqués sur architecture multi-composants programmables (SynDEx) [TG00] et reconfigurables (SynDEx-IC) [LK04]. Elle est basée sur la modélisation des architectures et des algorithmes à l'aide de graphe. L'implantation de l'algorithme sur l'architecture est modélisée par des transformations formelles de graphes. Le choix d'une implantation optimisée respectant les contraintes temps réel est réalisé à l'aide d'heuristiques implantées dans ces outils qui couvrent aussi la prédiction de performances ("simulation" avant exécution), et la génération automatique de code distribué/Circuit [ECSJS04].

Ces travaux s'appuieront sur l'expertise de l'équipe dans le domaine des architectures adaptatives orientées traitement d'images qui ont permis de mettre en évidence l'importance des architecture adaptatives/reconfigurable [MEDEADAC06, ISCAS10] d'un côté et les performances des portages d'application sur des tels architectures multi-cœurs embarquées [JASP05] et d'une compétence dans le domaine du parallélisme [SPIE97, ACIVS08].

L'intérêt pour la communauté scientifique et industrielle sera de disposer d'un état de l'art exhaustif des différentes architectures émergentes, et surtout de pouvoir enfin disposer d'une méthodologie transversale de l'exploration consciente de l'architecture reconfigurable ou adaptable et d'aide à l'implantation pour le calcul haute performance avec en particulier l'implantation parallèle des algorithmes innovant du laboratoire (squelettisation 3D, ligne de partage des eaux topologique).

Profil du candidat :

Le candidat devra posséder une bonne connaissance en architecture des processeurs et programmation parallèle. Il devra posséder les bases théoriques correspondantes à ces disciplines ainsi que les compétences informatiques et techniques permettant la réalisation d'un prototype. Des compétences en traitement des images sont souhaitables.

Peuvent candidater :

- Ressortissants de l'union européenne et de la Suisse
- Âgés de moins de 27 ans au 1er octobre de l'année considérée,
- En cours de préparation d'un Master pendant l'année de dépôt de la candidature ou titulaires d'un Master ou d'une équivalence leur permettant de s'inscrire en thèse.

Bibliographie :

[ISPA06] J-L. Gaudiot, "The Walls of Computer Design, (abstract)," Lecture Notes in Computer Science, Volume 4330/2006, pp.1, Proceedings of 2006 International Symposium on Parallel and Distributed Processing and Applications (ISPA 2006), Sorrento, Italy, December 4-7, 2006.

[ISFPGA09] Li, X. and Hammami, O. 2009. Small scale multiprocessor soft IP (SSM IP): single FPGA chip area and performance evaluation. In Proceeding of the ACM/SIGDA international Symposium on Field Programmable Gate Arrays (Monterey, California, USA, February 22 - 24, 2009). FPGA '09. ACM, New York, NY, 278-278. DOI=
<http://doi.acm.org/10.1145/1508128.1508180>

[MPSOC09a] Wen-mei Hwu, Many-core Parallel Computing - Can compilers and tools do the heavy lifting?, 9th International Forum on Embedded MPSoC and Multicore, 2-7 August 2009, Savannah, Georgia, USA

[MPSOC09b] Ahmed Jerraya, SoC integration beyond Hardware and Software, 9th International Forum on Embedded MPSoC and Multicore, 2-7 August 2009, Savannah, Georgia, USA

[MAAC06] Lars Wehmeyer, Peter Marwedel, Fast, Efficient and Predictable Memory Accesses: Optimization Algorithms for Memory Architecture Aware Compilation, Fast, Efficient and Predictable Memory Accesses: Optimization Algorithms for Memory Architecture Aware Compilation, Springer-Verlag New York, Inc., August 2006.

[DATE06] Martino Ruggiero, Alessio Guerri, Davide Bertozzi, Francesco Poletti, Michela Milano, Communication-aware allocation and scheduling framework for stream-oriented multi-processor systems-on-chip., DATE 2006: 3-8

[DATE98] Rabaey, J. and Wan, M. 1998. An energy-conscious exploration methodology for reconfigurable DSPs. In Proceedings of the Conference on Design, Automation and Test in Europe (Le Palais des Congr s de Paris, France, February 23 - 26, 1998). Design, Automation, and Test in Europe. IEEE Computer Society, Washington, DC, 341-342.

[KLUWER98] Catthoor, F., Greef, E. d., and Suytack, S. 1998 Custom Memory Management Methodology: Exploration of Memory Organisation for Embedded Multimedia System Design. Kluwer Academic Publishers.

[ICCAD01] Givargis, T., Vahid, F., and Henkel, J. 2001. System-level exploration for pareto-optimal configurations in parameterized systems-on-a-chip. In Proceedings of the 2001 IEEE/ACM international Conference on Computer-Aided Design (San Jose, California, November 04 - 08, 2001). International Conference on Computer Aided Design. IEEE Press, Piscataway, NJ, 25-30.

[CODES02] Palesi, M. and Givargis, T. 2002. Multi-objective design space exploration using genetic algorithms. In Proceedings of the Tenth international Symposium on Hardware/Software Codesign (Estes Park, Colorado, May 06 - 08, 2002). CODES '02. ACM, New York, NY, 67-72. DOI= <http://doi.acm.org/10.1145/774789.774804>

[JSA08] Papadopoulos, L., Baloukas, C., and Soudris, D. 2008. Exploration methodology of dynamic data structures in multimedia and network applications for embedded platforms. J. Syst. Archit. 54, 11 (Nov. 2008), 1030-1038. DOI= <http://dx.doi.org/10.1016/j.sysarc.2008.04.009>

[JSC07] Dimitroulakos, G., Galanis, M. D., and Goutis, C. E. 2007. Design space exploration of an optimized compiler approach for a generic reconfigurable array architecture. J. Supercomput. 40, 2 (May. 2007), 127-157. DOI= <http://dx.doi.org/10.1007/s11227-006-0016-1>

Publications du laboratoire

[MEMO03] "From Algorithm and Architecture Specifications to Automatic Generation of Distributed Real-Time Executives: a Seamless Flow of Graphs Transformations", T; Grandpierre, Y. Sorel, in MEMOCODE2003, Formal Methods and Models for Codesign Conference, Mont Saint-Michel, France, June 2003,

[TG00] GRANDPIERRE Thierry, Thèse de doctorat Paris Sud Orsay, Spécialité électronique "Modélisation d'architectures parallèles hétérogènes pour la génération automatique d'exécutifs distribués temps réel optimisés", Jury : Alain Mérigot, Michel Auguin, Ahmed Amine Jerraya, Anne-Marie Déplanche, soutenue le 30 novembre 2000, Université Paris XI.

[LK03] KAOUANE Linda, Thèse de doctorat "Formalisation et optimisation d'applications s'exécutant sur architectures reconfigurables", H. Mehrez : Rapporteur, e. Sanchez : Rapporteur, JM. Delosme : Rapporteur, M. Crochemore : Examineur, T. Grandpierre : Examineur, M. Akil : Directeur de thèse, Y. Sorel : Co-directeur de thèse, le 17 Décembre 2004 Université de Marne la Vallée, France.

[ESD07] "Implementing Real-Time Algorithms using the AAA Prototyping Methodology", P.Niang, T. Grandpierre, M. Akil, Embedded System Design: Topics, Techniques and Trends, ISBN 978-0-387-72257-3, 2007

[ECSJS04] "A methodology to implement real-time applications on reconfigurable circuits", L.Kaouane, M.Akil, T. Grandpierre, Y.Sorel, numéro spécial : Engineering of Configurable Systems of the Journal of Supercomputing, Kluwer Academic Publisher, vol. 30. 2004. pp. 283-301

[SPIE97] Laurent Perroton and Zouina Aktouf. Parallelization of an Efficient 3D Thinning Algorithm. SPIE Proceedings (Parallel and Distributed Methods for Image Processing). vol. 3166. 1997. pp. 128–134.

[ACIVS08] Petr Matas, Eva Dokládalo, Mohamed Akil, Thierry Grandpierre, Laurent Najman, M. Poupas, and V. Georgiev. Parallel Algorithm for Concurrent Computation of Connected Component Tree. Advanced Concepts for Intelligent Vision Systems (ACIVS'08). vol. 5259/2008. Lecture Notes in Computer Science. October 2008. pp. 230–241. Springer-Verlag.

[JASP05] Eva (Dokládalo) Dejnokova and Petr Dokládalo. Embedded real-time architecture for level-set-based active contours. EURASIP Journal on Advances in Signal Processing. vol. 17. 2005. pp. 1–16.

[MEDEADAC06] Eva Dokládalo, R. Schmit, S. Pajaniradja, and S. Amadori. Carvision: SOC architecture for dynamic vision systems from image capture to high level image processing. MEDEADAC. 2006. Note: electronic version (4 pp.)

[ISCAS10] Nicolas Ngan, E. Dokládalo, M. Akil, F. Contou-Carrère, Dynamically adaptable architecture for real-time video processing, IEEE International Symposium on Circuits and Systems (ISCAS) 2010, to appear