

Extrait du rapport d'activité 2005-2009

Gipsa-lab, Département DIS, Equipe GPIG
Thème : Adéquation Algorithme Architecture

16 décembre 2009

1 Composition

Vincent Fristot (MC INPG)
Dominique Houzet (PR INPG)
Sylvain Huet (MC INPG)
Stéphane Mancini (MC INPG)

2 Réalisations scientifiques

La complexité croissante des systèmes électroniques intégrés, dont les capacités d'intégration et les performances doublent tous les 18 mois (vérifiant ou dépassant la loi de Moore), rend leur conception de plus en plus difficile et nécessite la définition d'une nouvelle génération d'outils informatiques de conception intégrant des architectures génériques. Les propositions de méthodologies de synthèse d'architecture vers SoPC ou de programmation parallèle des Cell et GPU, nécessitent d'intégrer d'une part des travaux théoriques de modélisation des algorithmes et architectures mais aussi des facteurs liés à l'activité même de conception qui nécessite des compromis sur de nombreux critères, souvent antagonistes. Ainsi, nous devons répondre à des questions comme la délimitation de l'activité du concepteur et celle de l'outil, ou encore de déterminer quel est le langage de modélisation/programmation le plus approprié. Comment conserver la sémantique d'un calcul et son niveau d'abstraction sans passer par sa description dans un langage fonctionnel qui est déjà une implémentation (par exemple une FFT peut être codée de différentes façons et il y a une perte d'informations importante lorsque l'on fournit un code C d'une FFT à un outil de CAO car ce code spécifie une implémentation particulière)? Comment pouvons-nous garantir l'équivalence entre les modèles d'entrée et le résultat de la synthèse d'architecture? Comment intégrer dans le modèle les objectifs (ou la fonction de coût à minimiser) de la synthèse d'architecture tout en ayant un code portable et multi-plateforme?

Bien entendu, ces questions sont vastes et il est difficile d'y répondre globalement; ce thème vise à apporter une contribution à leur résolution.

L'Adéquation Algorithme Architecture (AAA) des systèmes intégrés (ASIC, SoC, MPSoC, SoPC, FPGA, GPU, Cell) s'intéresse à faire converger Algorithme et Architecture de façon à produire des systèmes numériques de traitement de l'information (image, signal, etc) efficaces, selon des critères variées (temps réel, consommation, coût/surface, etc), en fonction des contraintes respectives de ces domaines et à l'aide de méthodologies adaptées (automatiques ou "manuelles"). L'Adéquation Algorithme/Architecture s'intéresse donc à trois objets :

- des algorithmes (TDSI) ;
- des architectures de systèmes numériques intégrés (multiprocesseur : GPU, Cell, MPSOC ; hiérarchie mémoire ; réseau sur puce) ;
- des méthodologies (langages, modèles de calcul, modèles de programmation, exploration).

Ces trois points sont traités dans un même mouvement car il est impératif de tenir compte de leurs spécificités et de leurs interactions. Pour gérer la complexité croissante des systèmes, des méthodologies de haut niveau deviennent nécessaires. Cette activité en AAA est essentiellement tournée vers le traitement d'images et concerne l'imagerie médicale (traitement et capteurs) (thèse Arnaud Peizerat 2006), l'IHM et la perception. Les projets en cours, s'intéressent à l'architecture d'opérateurs matériels et de systèmes embarqués mixtes logiciel/matériel ainsi qu'aux méthodologies associées. Autour de la thématique de l'architecture, la définition de mécanismes systèmes (hiérarchie mémoire, contrôle des systèmes complexes, réseaux sur puce, etc) basés sur le traitement de signal ou l'automatique semble une voie prometteuse. La définition de stratégies de gestion de la mémoire devient un axe important compte tenu de l'augmentation considérable de la quantité de mémoire intégrée dans les circuits telle que prévue par des organismes de prospective comme l'ITRS. Le fabuleux potentiel d'intégration des nouvelles technologies, dont l'évolution surpasse aujourd'hui la loi de Moore, est à l'origine de problématiques nouvelles autour des mécanismes de gestion des systèmes intégrés. Alors qu'il est envisageable d'intégrer des centaines de millions de transistors sur un circuit, ce qui représente des milliers de microprocesseurs embarqués, nous pouvons imaginer des mécanismes systèmes élaborés de complexité croissante intégrant par exemple des solutions empruntées au domaine de l'automatique (en collaboration avec le Département d'Automatique de GIPSA-lab). En particulier, l'accès aux données par des opérateurs de plus en plus nombreux et performants donne une nouvelle ouverture à des problématiques telles que la hiérarchie mémoire distribuée ou le partage de ressources de communication dans les réseaux intégrés de type NoC (Network on Chip).

Cette thématique AAA est déterminante pour le maintien et la pérennisation d'une activité qui s'intéresse à l'architecture pour le traitement d'image et de signal au sein du laboratoire GIPSA-lab. En effet, une activité sur les méthodologies, au delà de leur intérêt propre, est indispensable pour rester compétitif dans un secteur de recherche et industriel aux évolutions rapides. Le renforcement du thème Adéquation Algorithme Architecture est à mettre en perspective, sur la période, avec le recrutement de deux enseignants-chercheurs (un professeur en 2006 et un maître de conférences en 2007) ainsi que le retour à temps plein d'un maître de conférences en 2008.

Ces travaux de recherche s'appuient sur la plate-forme SoPC. Par ailleurs, une plate-forme multi-GPU est en construction, elle intégrera plusieurs serveurs de calcul haute performance quadri-GPU en réseau pour l'étude et le prototypage d'applications parallèles.

Les membres du thème AAA ont organisé à Grenoble en 2007 la conférence DASIP ; cette conférence a été créée par la communauté AAA du GDR ISIS. Ils ont aussi mis en place début 2007 un premier stage de formation continue au niveau national sur les processeurs graphiques (GPU).

Il est aussi à noter l'investissement fait depuis plusieurs années d'un permanent pour le développement et la diffusion d'un logiciel de Conception Assistée par Ordinateur pour le routage des cartes électroniques, appelé Kicad

2.1 Hiérarchie mémoire et localité

Cette action traite de la génération conjointe de hiérarchie mémoire et d'expression de localité spatiale et temporelle. La problématique est de formaliser les méthodes utilisées empiriquement dans les projets en cours (projet BQR ArchiTEP - thèse Nicolas Gac [8], projet RNTS TELMA) et d'automatiser l'optimisation de hiérarchies mémoires conjointement à la transformation des applications afin d'exprimer leur localité spatiale et temporelle [6, 5]. Des applications comme les transformations d'images simples (perspective, log-polaire, zoom local, etc), ou complexes (ondelettes, vision et interprétation de scènes, etc), la 3D (parcours de grille, application de texture) ou encore la reconstruction tomographique (intégrales curvilinéaires sur des structures 2D ou 4D) posent des problèmes d'accès à la base de données car elles reposent sur le calcul complexe d'indices et d'organisation de boucles imbriquées. L'approche courante consiste à parcourir le résultat par blocs ou tuiles pour générer de la localité spatiale et temporelle dans les données initiales. D'autres méthodes de réorganisation des traitements se sont avérées efficaces et leur formalisation pourrait être incluse dans des méthodologies plus générales. En traitement d'images, il est tout à fait possible d'envisager des hiérarchies mémoire adaptées qui fournissent des services de haut niveau tels qu'un adressage par coordonnées des images, avec des mécanismes de cache ou de compression/décompression à la volée. Les travaux sur le cache 2D adaptatif et prédictif ont montré l'intérêt d'une hiérarchie mémoire adaptée au traitement d'images [3]. Ce cache exploite la localité temporelle et spatiale 2D pour prédire les prochains accès à une image et pré-charger les données correspondantes. L'architecture proposée est modulaire, avec les possibilités d'adapter le mécanisme de prédiction à l'application de traitement et de construire des hiérarchies complexes pour fournir du parallélisme d'accès à l'image. Cette démarche s'inscrit en liaison avec le projet "Rétina" sur la modélisation des systèmes de vision embarquée [thèse BDI CNRS STMicroelectronics Rosilde Corvino - 2009] qui s'intéresse aux méthodologies de synthèse HLS [4, 1] de type C vers RTL et plus particulièrement à la possibilité d'utiliser et d'adapter des logiciels tels que PICO, ImpulseC, CatapultC pour des applications de traitement d'images.

2.2 Hiérarchie mémoire distribuée

Cette étude porte premièrement sur les méthodologies de synthèse d'architecture système intégrant une hiérarchie mémoire distribuée, réseau de communication, multiprocesseur, sur technologie multi-FPGA/ASIC avec exploration et aide au paramétrage des mécanismes architecturaux impliqués. Cela passe par la modélisation à haut niveau (avec des outils comme Cofluent et des langages comme SystemC ou Haskell [BQR MEMNOC et thèse Tomasz Toczek 2010] avec simulation (en liaison avec le projet de plate-forme nationale SoCLiB) et exploration automatique par algorithmes génétiques (codirection de thèse de Zoukhun Wang avec l'ENSTA). L'intégration des mécanismes architecturaux génériques (NoC, caches, mémoires, synchronisation, reconfiguration) au sein même du flot de conception et de manière unifiée doit permettre d'atteindre l'efficacité recherchée, car chaque élément du système a un impact sur l'ensemble [2, 7]. Cette étude vise également à aborder de nouvelles cibles technologiques (GPU, Cell) comme étape préliminaire de prototypage/validation d'algorithmes parallèles, ou parfois même comme alternative aux circuits intégrés. Cela concerne l'adaptation à ces cibles des méthodologies d'adéquation algorithmes architecture étudiées au sein de l'équipe dans le contexte du TDSI, avec prototypage d'applications étudiées au sein d'autres équipes du laboratoire, voir d'autres laboratoires (financement BQR). Ces travaux ont été initiés fin 2006 avec l'arrivée de ces composants dans le commerce.

Références

- [1] Rosilde Corvino, Stéphane Mancini, Roberto Guizzetti, and Pascal Urard. Optimisation of a memory subsystem for nonlinear references in a high level synthesis flow. In *DCIS'08 proceedings*, page 1, grenoble, France, 2008. hal-00348538.
- [2] Dominique Houzet and Salim Ouadjaout. Generation of embedded Hardware/Software from SystemC. *EURASIP Journal on Embedded Systems*, page ID18526, 2006. hal-00127973.
- [3] Zahir Larabi, Yves Mathieu, and Stéphane Mancini. Efficient data access management for fpga-based image processing socs. In IEEE Computer Society, editor, *Efficient data access management for FPGA-Based image processing SoCs*, pages 159–165, Paris, France, 2009. hal-00368532.
- [4] Bertrand Le Gal, Emmanuel Casseau, and Sylvain Huet. Dynamic memory access management for high-performance dsp applications using high-level synthesis. *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, 16(11) :1454–1464, 2008. hal-00361870.
- [5] Stéphane Mancini, Michel Desvignes, Nicolas Gac, and Dominique Houzet. High speed 3d tomography on cpu, gpu, and fpga. *EURASIP Journal on Embedded Systems*, 2008(Article ID 930250) :<http://www.hindawi.com/GetArticle.aspx?doi=10.1155/2008/930250>, 2008. hal-00367321.
- [6] Stéphane Mancini, Nicolas Gac, Michel Desvignes, Olivier Bourrion, and Olivier Rossetto. Application d'un cache 2d prédictif à l'accélération de la rétroprojection tep 2d. *Traitement du Signal*, 23 :Numéro 5–6, 2006. hal-00122552.
- [7] Fabienne Nouvel, Florent Berthelot, and Dominique Houzet. A flexible system level design methodology targeting run-time reconfigurable fpgas. *EURASIP Journal on Embedded Systems*, 2008 :793919, 2008. hal-00320192.
- [8] Thèse tel-00330365.