

# Modélisation et Optimisation de la Consommation Énergétique des Plateformes Reconfigurables

Robin BONAMY      CAIRN-IRISA — LEAT

Journées scientifiques MCSOC, 29-30 juin 2011

## Résumé

Ce rapport présente les travaux de thèse concernant la modélisation et l'optimisation de la consommation de la puissance et de l'énergie des plateformes reconfigurables et en particulier de la reconfiguration dynamique. Ces travaux s'inscrivent dans le cadre du projet ANR *Open-PEOPLE* dont le but principal est de mettre à disposition une plateforme de modélisation, d'estimation, de mesure et d'optimisation de la consommation en puissance et en énergie. La première partie de ce rapport porte sur un rapide état de l'art de la consommation dans les circuits, la partie suivante concerne l'étude de la consommation d'énergie des blocs accélérateurs matériels. La troisième partie concerne le déroulement de la reconfiguration dynamique. La quatrième partie est une étude de la possible réduction de la consommation d'énergie en utilisant la reconfiguration dynamique. Enfin, la conclusion de ces travaux est présentée avec les études en cours.

## Introduction

Un système sur puce (System on Chip, SoC) est un circuit comprenant plusieurs fonctions diverses telles que un ou plusieurs processeur(s), zone(s) reconfigurable(s), processeur(s) de traitement de signal (DSP), divers périphériques, des mémoires ou des parties analogiques. Ces circuits sont de plus en plus utilisés en raison de leur taille réduite et d'une réduction des coûts par rapport à l'utilisation de plusieurs circuits pour réaliser la même fonction. L'utilisation de SoCs est de plus en plus courante dans les applications embarquées où la question de la consommation est très importante. La taille de la batterie dans le cadre d'un système autonome, le dimensionnement des dissipateurs ou le choix des rails d'alimentation sont des paramètres à prendre en compte très tôt dans la conception d'un système.

Pour cette raison, il y a un besoin de pouvoir estimer, rapidement et au plus tôt, la consommation en puissance ou en énergie d'un système hétérogène complexe.

Les travaux portent ici sur la consommation des architectures reconfigurables (de type FPGA).

## 1 État de l'Art

La consommation des circuits actuels (CMOS) a deux principales origines [Julien06], [Garcia99], [TI97] :

- **la consommation dynamique** qui est liée à la tension ( $V_{cc}$ ) d'alimentation du circuit, sa fréquence ( $F$ ) de fonctionnement, son activité ( $\alpha$ ) et la capacité ( $C$ ), liée à la longueur des inter-connexions.

$$P_d = V_{cc}^2 \times F \times \alpha \times C$$

- **la consommation statique** qui est liée à la surface ( $A$ ) et au taux d'occupation et la tension ( $V_{cc}$ ) d'alimentation.

$$P_s = V_{cc} \times A$$

L'estimation de la consommation peut se faire à grain fin. Cette estimation prend en compte l'activité et la longueur des connections, le nombre de bascules utilisées et les opérations effectuées. Le problème de cette estimation est que le design doit être presque terminé pour pouvoir estimer la consommation. Les éventuelles modifications sont donc coûteuses. De plus cette estimation nécessite des temps de calculs très importants. Il existe alors une estimation à gros grain qui prend en compte des paramètres applicatifs, l'utilisation d'un type de mémoire, la surface utilisée et le cadencement. Cette estimation peut être effectuée plus tôt dans la conception du système et elle est moins coûteuse en quantité de calcul donc l'estimation peut être effectuée plus rapidement. L'inconvénient de cette estimation à gros grain est une mauvaise précision. Il existe aussi d'autres types d'estimations en fonction des besoins, par exemple pour modéliser la consommation d'une machine d'états [Benini00] ou pour estimer les pics de consommations [Gupta03] qui peuvent avoir des effets sur les systèmes. Selon le degré de précision souhaité dans la modélisation, il peut être intéressant de prendre en compte les glitches [Ragh96] dans les circuits lors de l'estimation de la consommation.

Les travaux présentés ici portent sur l'opportunité de la réduction de la consommation énergétique en utilisant des blocs accélérateurs matériels avec la reconfiguration dynamique partielle.

## 2 Consommation des Accélérateurs Matériels

Un accélérateur matériel permet de décharger un processeur des charges de calculs lourdes et/ou répétitives. Celui-ci consiste en l'implémentation matérielle d'une tâche. Cet accélérateur permet de améliorer l'efficacité des traitements : le temps d'exécution est plus faible et la consommation d'énergie plus faible [AlteraAN531].

L'implémentation de ces blocs étant matérielle, il est possible de faire varier certains paramètres comme le niveau de parallélisme.

Cette étude porte sur l'étude de l'impact du niveau de parallélisme sur la consommation énergétique d'un accélérateur matériel.

Les mesures sont effectuées sur une plateforme FPGA Virtex 5 de Xilinx, le coeur est alimenté sous 1V, la fréquence de fonctionnement est de 100MHz et à température constante. L'utilisation de la synthèse de haut niveau permet de générer aisément plusieurs implémentations par le biais du déroulage de boucle, par exemple, qui permet de faire varier le niveau de parallélisme.

La multiplication matricielle est choisie en tant que benchmark. Les résultats de la campagne de mesure sont présentés sur la figure 1.

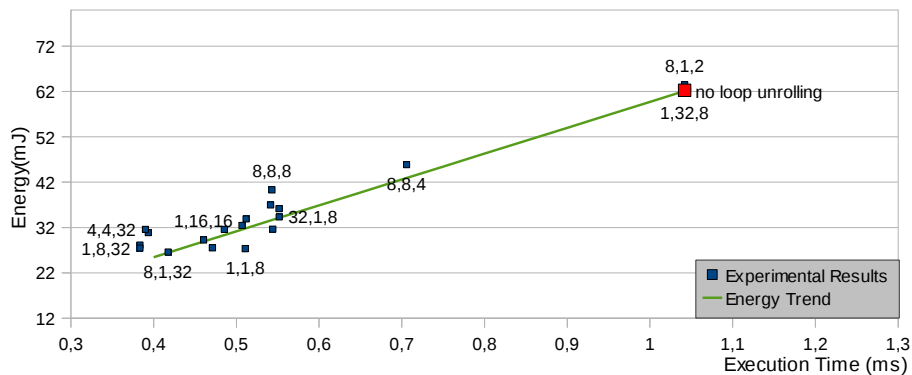


FIGURE 1 – Énergie consommée par la multiplication matricielle en fonction du temps d'exécution. Les nombres séparés par des virgules correspondent au nombre de déroulage de chaque boucle (Loop Unrolling Index, LUI) de la multiplication matricielle (qui en contient trois) ( $LUI1$ ,  $LUI2$ ,  $LUI3$ ). La tendance de cette courbe est de la forme :  $E = A + B \times t$ .

La tendance de la consommation énergétique est proportionnelle au temps d'exécution de la tâche. Plus l'exécution de la tâche est rapide, grâce à l'utilisation du parallélisme, plus la consommation d'énergie nécessaire à l'exécution du calcul diminue.

Une tâche implémentée avec un certain niveau de parallélisme est plus efficace du point de vue de l'énergie qu'une tâche séquentielle, cependant l'énergie correspond à la puissance consommée par la tâche pendant le temps nécessaire à son exécution. Lorsque la tâche n'est plus utilisée, sa consommation *idle* est plus importante pour une tâche occupant plus d'espace. La reconfiguration dynamique a un sens ici pour optimiser l'utilisation de la surface du circuit et pour limiter la consommation lorsque la tâche n'est plus utilisée.

### 3 Consommation de la Reconfiguration Dynamique Partielle

La reconfiguration dynamique partielle permet d'apporter de la flexibilité à une architecture matérielle, elle permet de réduire la surface nécessaire pour une application en réutilisant les zones non utilisées. Grâce à cette flexibilité, il est possible de réduire la consommation en permettant l'exécution de l'application sur un circuit de taille plus faible et donc de réduire sa consommation statique, mais il est aussi possible de réduire la consommation en effectuant des choix d'implémentations d'accélérateurs matériels.

Cependant, la reconfiguration dynamique partielle a un surcoût en temps et en puissance [Savary07], [Becker03] et donc en énergie.

Les mesures de consommations sont effectuées sur la même plateforme que précédemment, en mettant en place le projet proposé par Xilinx pour la reconfiguration dynamique partielle [UG744]. Celui-ci contient un *microblaze*, coeur microprocesseur logiciel de Xilinx, pour la gestion de la reconfiguration et un stockage des données en mémoire externe, CompactFlash.

Les mesures sont présentées sur les figures 2 et 3.

La figure 2, concernant la mesure de consommation du coeur du FPGA pendant le début d'une procédure de reconfiguration, montre qu'il y a bien un coût énergétique pour effectuer la reconfiguration. Il y a des pics de puissance correspondant à l'écriture des données dans ICAP (port de reconfiguration interne au FPGA) et qu'il y a des phases d'attente de la compactflash qui est une mémoire assez lente. Dans ce projet, les pics de consommation ( $50\text{ mW}$ ) sont en grande partie dus à l'activité du microblaze.

La figure 3 montre la consommation du coeur du FPGA pendant la configuration d'une tâche existante, dont l'exécution est terminée (puissance idle), vers une nouvelle tâche vide pour réduire la consommation et on voit bien entre le début et la fin de la reconfiguration dynamique que la consommation est réduite de plus de  $20\text{ mW}$ .

### 4 Utilisation de la Reconfiguration Dynamique Partielle pour la Réduction d'Énergie

La section précédente a montré qu'il était possible de réduire la puissance consommée par un périphérique en utilisant la reconfiguration dynamique partielle mais que cette reconfiguration a un coût. Si l'on considère que le FPGA est découpé en plusieurs régions reconfigurables (PRR), on peut alors optimiser l'énergie en utilisant différentes versions d'une tâche comme présenté dans la seconde section de ce rapport.

Quel est alors le moment où utiliser la reconfiguration dynamique partielle est intéressant au niveau de l'énergie ?

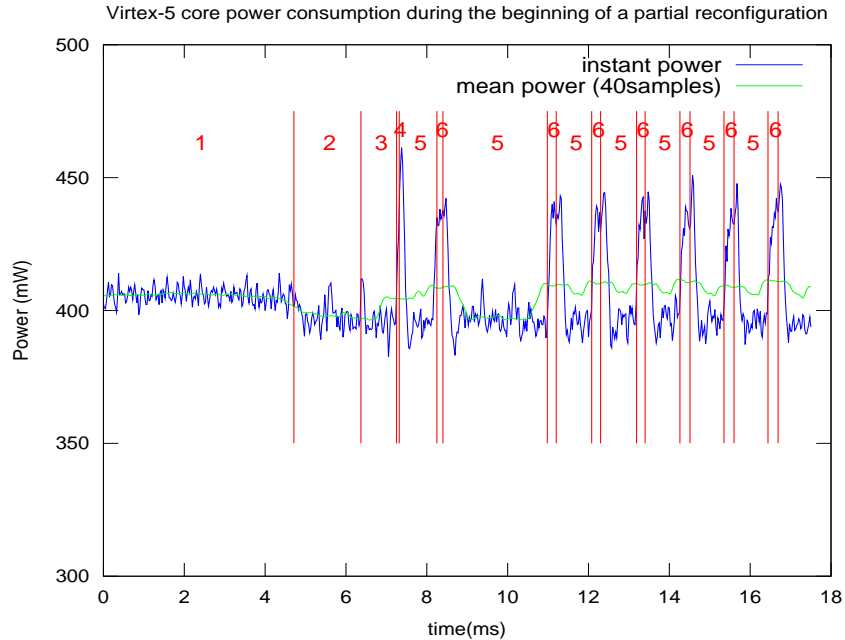


FIGURE 2 – FPGA core power consumption during the beginning of dynamic partial reconfiguration versus time. 1) reconfiguration order arrives 2) opening the bitstream file on the Compact Flash 3) reading the header of the bitstream file 4) analysis of the validity of the header 5) reading the first compact flash sector of the file 6) writing the data into the ICAP (Internal Configuration Access Port) 5 & 6) are repeated until the end of the file.

Il est possible de minimiser simplement l'énergie d'un tel système lorsque que la tâche à exécuter est, par exemple, cadencée pour un certain temps d'utilisation. Dans ce cas, les équations d'énergie correspondent à une somme des énergies des étapes nécessaires :

- configuration de la tâche sur un PRR
- exécution de la tâche
- configuration de la tâche vide
- "exécution" de la tâche vide jusqu'à la fin du créneau de temps

OU

- configuration de la tâche sur un PRR
- exécution de la tâche
- la tâche en *idle* jusqu'à la fin du créneau de temps

la minimisation de l'énergie pour ce créneau de temps passe par le choix de l'équation ayant le résultat le plus petit si l'implémentation est possible.

La figure 4 représente la consommation selon plusieurs choix de configuration possibles.

## Conclusion

Ce rapport présente que la variation du niveau de parallélisme de l'implémentation d'une tâche accélérateur matériel influe sur la consommation d'énergie, mais qu'une tâche parallélisée occupe plus de surface. La consommation idle de cette tâche est donc plus importante et la surface peut être mal utilisée. Pour y remédier, il est possible d'utiliser la reconfiguration dynamique. Or la reconfiguration dynamique est d'autant plus coûteuse en temps et en énergie que la tâche est large. Il y a un compromis en fonction de l'exécution de la tâche (et de son contexte) et son implémentation.

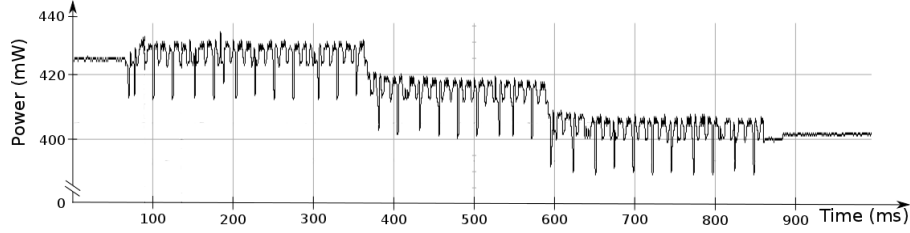


FIGURE 3 – Power consumption of a Virtex-5 core during the reconfiguration of a PRR from a Task to a *low power*, empty task versus time.

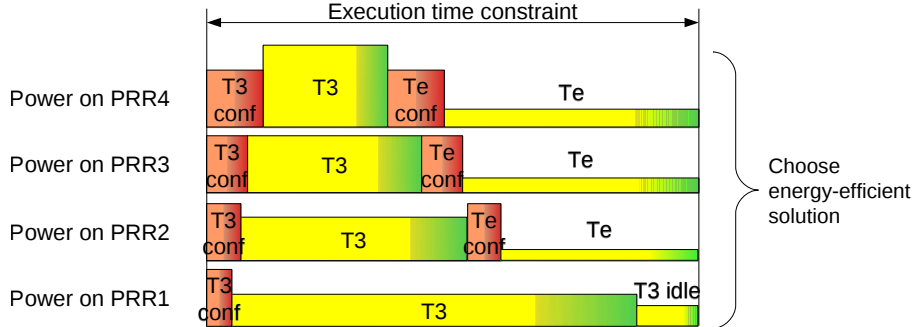


FIGURE 4 – Exemple de choix de configurations possibles (sur 4 PRR) et leur consommation respective pour l'exécution d'une tâche (T3) et éventuellement d'une tâche vide, faible consommation (Te).

La continuité des travaux passe par l'établissement des modèles de consommation de la reconfiguration dynamique (en fonction des contrôleurs de reconfiguration utilisés) et l'écriture des équations complètes de minimisation de l'énergie. Ces équations seront intégrées dans un manager ou un ordonnanceur de la reconfiguration dynamique pour permettre d'effectuer, en ligne, des choix de configuration et de minimiser l'énergie du système.

## Références

- [Julien06] Julien N., Caractérisation et modélisation de la consommation sur FPGA, ECOFAC 2006
- [Garcia99] Garcia A. and al., Power modelling in Field Programmable Gate Arrays (FPGA), In International Workshop on Field Programmable Logic and Applications 1999
- [TI97] Texas Instruments, CMOS Power Consumption and Cpd Calculation, 1997
- [AlteraAN531] Altera Corporation, AN 531 : Reducing Power with Hardware Accelerators, 2008
- [Savary07] Savary Y., Thèse : Etude du Potentiel des Architectures Reconfigurables pour Maîtriser la Consommation dans les Applications Embarquées, 2007
- [Becker03] Becker J. and al., Power estimation and power measurement of Xilinx Virtex FPGAs : trade-offs and limitations, in Integrated Circuits and Systems Design, 2003
- [Benini00] Benini L. and al., Regression-based RTL power models for controllers. In Proceedings of the 10th Great Lakes symposium on VLSI, 2000
- [Gupta03] Gupta S. and al., Energy and peak-current per-cycle estimation at RTL. Very Large Scale Integration (VLSI) Systems, IEEE Transactions on, 2003
- [Ragh96] Raghunathan A. and al., Register-transfer level estimation techniques for switching activity and power consumption. In Proceedings of the 1996 IEEE/ACM international conference on Computer-aided design, 1996
- [UG744] Xilinx Inc. UG744, PlanAhead Software Tutorial : *Partial Reconfiguration of a Processor Peripheral*