HAL
open science

La consommation des interconnexions dans les SOC

Antoine Courtay

To cite this version:

Antoine Courtay. La consommation des interconnexions dans les SOC. GDR SOC SIP, Jun 2007, Paris, France. pp.La consommation des interconnexions dans les SOC. hal-00169287

HAL Id: hal-00169287

https://hal.science/hal-00169287

Submitted on 3 Sep 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

La consommation des interconnexions dans les SOC

Antoine Courtay
IRISA - Université de Rennes 1
6, rue de Kerampont
22300 Lannion, France
courtay@irisa.fr
LESTER - Université de Bretagne Sud
rue saint Maudé
56321 Lorient, France
antoine.courtay@univ-ubs.fr

1. INTRODUCTION

Avec l'évolution des paramètres technologiques, la maîtrise de l'évolution du délai et de la consommation est de plus en plus difficile à contrôler. Aujourd'hui les applications portables sont de plus en plus complexes et nécessitent beaucoup de ressources de calculs, ce qui implique un fort volume de données à stocker ou à faire transiter d'une unité à une autre. Les prévisions de l'ITRS [1] montrent une diminution des dimensions des transistors et des fils ce qui se traduit par une évolution du comportement du circuit tout particulièrement au niveau temporel. Ainsi, le délai d'un fil devient largement supérieur à celui d'une porte [2]. Cette augmentation est due à l'évolution des paramètres résistif et capacitifs des interconnexions qui tendent toujours à augmenter. L'augmentation des phénomènes capacitifs se traduit également par une augmentation de la part de la consommation due aux interconnexions qui peut représenter jusqu'à 50% de la consommation totale ainsi que de la surface occupée sur la puce [3].

Il devient donc indispensable de prendre en compte la consommation des interconnexions lors de l'évaluation de la consommation d'une puce. La suite de cet article présentera dans un premier temps les paramètres à prendre en compte lors de la modélisation (délai et consommation) des interconnexions ainsi que les modèles de consommation développés. La troisième partie présentera les résultats expérimentaux obtenus sur l'évaluation des techniques de réduction du $crosstalk$.

2. MODÉLISATION ET ESTIMATION DE LA CONSOMMATION

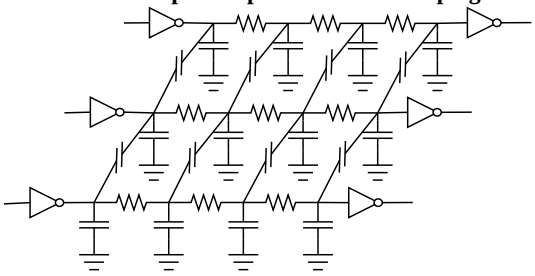
2.1 Modélisation de la consommation des bus

La consommation des interconnexions est liée à deux paramètres principaux qui sont la résistance et la capacité des lignes d' interconnexions.

Les grandeurs élémentaires que l'on trouve dans les $DesignKit$ des constructeurs permettent de calculer la résistance (R) ainsi que la capacité (C) du fil en fonction de ses dimensions. Afin de modéliser le plus précisément le fil, un modèle où R et C sont distribués sera utilisé afin d'avoir une meilleure précision au niveau du délai.

Généralement, les lignes sont regroupées afin d'obtenir un bus de données permettant la transmission d'information entre blocs. Le fait d'utiliser un bus fait apparaître un autre phénomène capacitif qui est le couplage capacitif entre fil aussi appelé $crosstalk$. La capacité de $crosstalk$ dépend quant à elle de la surface en regard entre les fils et varie donc en fonction des dimensions technologiques (épaisseur et espacement). Le bruit dû au $crosstalk$ est relativement localisé ; en général un système soumis au $crosstalk$ est modélisé en négligeant les ordres supérieurs au premier. La figure 1 présente le modèle physique du bus que nous avons utilisé pour mener nos expérimentations.

Figure 1: Modèle complet $\pi 3$ pour 3 fils avec couplage *crosstalk*.



Un des effets du $crosstalk$ est de faire augmenter le délai de propagation sur le bus, il introduit un facteur de délai g comme décrit dans le tableau 1 où r est le ratio de la capacité de $crosstalk$ C_c par rapport à la capacité du fil C_s (capacité du fil par rapport au substrat).

Dans ce tableau, $\uparrow$ représente une transition montante, $\downarrow$ représente une transition descendante et $-$ signifie qu'il n'y a pas de transition sur le fil. Dans le meilleur cas lorsque les trois fils effectuent une transition dans le même sens, le délai sur le fil victime (le fil central) est le délai sans $crosstalk$ (i.e. $g = 1$) ; mais le cycle d'horloge doit être dimensionner en tenant compte exclusivement du délai pire cas (i.e. $g = 1 + 4.r$) afin d'assurer une transmission correcte des données.

Le $crosstalk$ représente également une source de bruit ; en effet, à cause de la capacité introduite entre les fils, une transition sur un

Table 1: Capacité parasite (C_L) et facteur de délai (g) du fil victime en fonction du type de transition.

C_L	Types de transition				g
C_s	$(\uparrow,\uparrow,\uparrow)$	$(\downarrow,\downarrow,\downarrow)$			1
$C_s + C_c$	$(-,\uparrow,\uparrow)$	$(-,\downarrow,\downarrow)$	$(\uparrow,\uparrow,-)$	$(\downarrow,\downarrow,-)$	1+r
$C_s + 2.C_c$	$(-,\uparrow,-)$ $(\uparrow,\uparrow,\downarrow)$	$(-,\downarrow,-)$ $(\uparrow,\downarrow,\downarrow)$	$(\downarrow,\uparrow,\uparrow)$	$(\downarrow,\downarrow,\uparrow)$	1+2.r
$C_s + 3.C_c$	$(-,\uparrow,\downarrow)$	$(-,\downarrow,\uparrow)$	$(\uparrow,\downarrow,-)$	$(\downarrow,\uparrow,-)$	1+3.r
$C_s + 4.C_c$	$(\uparrow,\downarrow,\uparrow)$	$(\downarrow,\uparrow,\downarrow)$			1+4.r

des fils à pour effet de causer un pic de tension sur le fil voisin ce qui peut conduire à des erreurs à la réception si ce pic dépasse la tension de seuil des buffers.

2.2 Estimation de la consommation

Afin d'établir nos modèles de consommation du bus, nous avons besoin de connaître les caractéristiques physiques des interconnexions.

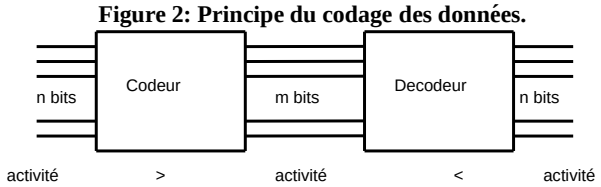
Pour cela nous avons effectué les différentes expérimentations avec un simulateur SPICE (ELDO 5.7) pour plusieurs technologies ($130nm$, $90nm$ et $65nm$) afin d'obtenir des modèles précis au niveau physique en terme de consommation et de vitesse de transmission des données. Au terme des expérimentations, nous avons développé un outil qui fournit à l'utilisateur des résultats en termes de surface, de consommation (instantannée, moyenne et maximale), d'activité des données et de vitesse maximale de transmission. Ces résultats permettent alors à l'utilisateur de dimensionner son bus. Ils vont également permettre d'analyser efficacement les performances des différentes techniques de réduction de la consommation et du délai trouvées dans la littérature.

3. IMPACT DES TECHNIQUES D'OPTIMISATION

On trouve dans la littérature beaucoup de techniques qui visent à réduire l'effet du $crosstalk$ ayant pour but l'accélération de la propagation des informations et également la réduction de la consommation des interconnexions. Ces techniques interviennent à plusieurs niveaux d'abstractions.

On peut citer par exemple au niveau technologique la possibilité de jouer sur les dimensions des fils (espacement, épaisseur) afin de diminuer les capacités parasites. La technique du blindage est également couramment utilisée ; elle consiste à insérer un fil relié à la masse (par exemple) entre deux fils adjacents de manière à supprimer les pires cas de transition du tableau 1. C'est au niveau architectural que l'on rencontre le plus grand nombre des techniques [4, 5, 6, 7]. Elles consistent toutes en un codage des données tel que le montre la figure 2. Le principe du codage est de transmettre l'information sur m bits avec ($m \geq n$) tel que l'activité des données codées soit inférieure à celle des données non codées. Le codage des données ayant pour but de supprimer les pires cas du tableau 1 (i.e. $g = 1 + 4.r$ et/ou $g = 1 + 3.r$).

Figure 2: Principe du codage des données.



Afin de confirmer la pertinence de ces méthodes nous nous sommes attachés dans un premier temps à vérifier leur performances en termes de variation de l'activité, des capacités parasites, de la surface et de la consommation énergétique sur le bus. Ces résultats sont expliqués dans [8] où il est fait une synthèse des performances de ces méthodes. Dans un second temps nous avons regardé si les transitions pire cas en temporel sont bien les mêmes qu'en consommation. Les expérimentations effectuées montrent que cela n'est pas toujours le cas. De plus, nous avons remarqué que pour différents flots de données (image, musique, parole) la répartition des différentes classes de transitions montre que les techniques n'éliminent pas les transitions qui ont le plus fort pourcentage d'apparition.

Les expérimentations ont également montré que lors d'un changement de technologie ($130nm \rightarrow 90nm$ par exemple) le délai de propagation pire cas est multiplié par trois pour une bufferisation simple de l'interconnexion. D'où la nécessité d'utiliser les méthodes d'insertion de buffer [9] afin de respecter les cadences de fonctionnement que l'on se fixe sur le bus. Malheureusement l'insertion de buffer est pénalisante puisque l'on insère des portes supplémentaires qui vont contribuer à alourdir le budget consommation.

4. BIBLIOGRAPHIE

[1] ITRS, "Technical report," *International Technology Roadmap for Semiconductors*, 2004.

[2] R. Ho, K. Mai, and M. Horowitz, "The future of wires," *Proceedings . IEEE*, vol. 89, no. 4, pp. 490–504, April 2001.

[3] N. Magen, A. Kolodny, U. Weiser, and N. Shamir, "Interconnect-power dissipation in a microprocessor," in *the Proceedings of the 2004 international workshop on System level interconnect prediction (SLIP)*, Paris, France, 2004, pp. 7–13.

[4] L. Benini, G. D. Micheli, E. Macii, D. Sciuto, and C. Silvano, "Asymptotic zero-transition activity encoding for address busses in low-power microprocessor-based systems," in *the Proceedings of the 7th Great Lakes Symposium on VLSI (GLS)*, Urbana, USA, 1997, p. 77.

[5] J. Philippe, S. Pillement, and O. Sentieys, "Area efficient temporal coding schemes reducing crosstalk effects," in *the Proceedings of the International Symposium on Quality Electronic Design (ISQED)*, San Jose, USA, 2006, pp. 334–339.

[6] M. Stan and W. P. Burleson, "Bus-invert coding for low-power i/o," in *IEEE Trans. on Very Large Scale Integration Systems*, vol. Vol 3, no. No.1, 1995, pp. pp.49–58.

[7] C. Su, C. Y. Tsu, and A. M. Despain, "Saving power in the control path of embedded processors," *Design & Test of Computers, IEEE*, vol. 11, pp. 24–31, 1994.

[8] A. Courtay, O. Sentieys, and N. Julien, "Interconnexions et consommation : où en sommes nous ?" in *the Proceedings of the 4th MajecSTIC*, Lorient, France, 2006.

[9] A. Nalamalpu and W. Burleson, "A practical approach to dsm repeater insertion : Satisfying delay constraints while minimizing area and power," in *the Proceedings of the 14th Annual IEEE Internationnal ASIC/SOC Conference*, Arlington, USA, 2001, pp. 152–156.