

MICRO-RÉSEAU POUR SYSTÈMES INTÉGRÉS : RÉALISATION D'UN RÉSEAU *SPIN* À 32 PORTS

Adrijean ANDRIAHANTENAINA

Alain GREINER

Département ASIM, laboratoire LIP6, Université Pierre et Marie CURIE
4, place Jussieu, 75252 Paris Cedex 05

{Adrijean.Andriahantenaina, Alain.Greiner}@lip6.fr

RÉSUMÉ : Dans un domaine où la capacité d'intégration double tous les deux ans, et permet aujourd'hui d'intégrer plusieurs dizaines de millions de transistors sur une seule puce, l'enjeu des prochaines années dans le secteur des semi-conducteurs est d'intégrer sur une même puce des systèmes contenant plusieurs dizaines de processeurs ou coprocesseurs hétérogènes. Le goulot d'étranglement au niveau architectural réside dans la gestion des communications entre les processeurs et la mémoire. Dans ce cadre, nous avons développé le réseau d'interconnexion *SPIN* (*Scalable Programmable Integrated Network*), qui est capable de fournir aux concepteurs systèmes une bande passante qui augmente linéairement avec le nombre de processeurs intégrés.

I – INTRODUCTION

L'idée directrice du projet *SPIN* est de remplacer le bus partagé par un réseau d'interconnexion multi-étages utilisant une technique de routage bien connue dans le monde des machines parallèles, ainsi que dans le domaine des télécommunications : la commutation de paquet. Le réseau *SPIN* est un réseau avec une topologie en arbre quaternaire élargi, la technique de routage utilisée est distribuée, adaptative et de type *wormhole*. Le réseau *SPIN* utilise des liens point-à-point bidirectionnels et un contrôle de flux de type crédit. Le réseau *SPIN* repose sur trois macro-cellules VLSI. Les routeurs *RSPIN* se chargent du routage des paquets jusqu'à leurs destinations finales. Les *wrappers* *VCI/SPIN* et *SPIN/VCI* [5] se chargent de l'interface du réseau *SPIN* avec les abonnés en fournissant à ceux-ci une interface respectant le standard *VCI* (*Virtual Component Interface*) à chacun de ses ports pour que les processeurs, mémoires et autres coprocesseurs qui l'utilisent n'aient pas à gérer le protocole *SPIN*. Dans cet article, on se concentre exclusivement sur le réseau de routeurs *RSPIN*.

II – LE RÉSEAU *SPIN*

II-1 – Topologie

La topologie du réseau *SPIN* est un arbre quaternaire élargi qui est caractérisée par le fait que chaque noeud a quatre fils et que les pères sont répliqués quatre fois à chaque niveau de l'arbre [2]. Cette topologie est intrinsèquement redondante, puisque les quatre pères offrent quatre chemins équivalents pour acheminer un message entre deux fils possédant le même père. Dans cette topologie, le plus court chemin est celui passant par un plus proche ancêtre commun. L'arbre quaternaire élargi a les avantages suivants : son diamètre (nombre maximal de liens entre deux abonnés) reste raisonnable ($2 * \log_4 n$, où n est le nombre de couche du réseau). Cette topologie est extensible et utilise un petit nombre de routeurs pour un nombre donné d'abonnés. Elle est non bloquante et a une structure naturellement hiérarchique dont on peut tirer partie dans les systèmes intégrés : Comme dans les

systèmes d'interconnexion utilisant une hiérarchie de bus, une manière très simple d'optimiser le réseau est de mettre dans le même cluster les abonnés appartenant à un même sous-système. Les communications internes à un même sous-système resteront donc locales, ce qui permet de réduire à la fois la latence et la consommation.

II-2 – Format des paquets *SPIN*

Le réseau *SPIN* est un réseau à commutation de paquets. Du point de vue du routeur *RSPIN*, un paquet est composé d'un nombre quelconque de mots de 36 bits. Le réseau considère un paquet comme une information atomique. Il existe un marqueur de début de paquet et un marqueur de fin de paquet. Une fois que le début du paquet est émis sur un lien entre deux routeurs, le lien est occupé et restera occupé jusqu'à l'émission du marqueur de fin de paquet. Un paquet *SPIN* est composé de deux parties distinctes : l'entête et le corps du paquet. L'entête du paquet est un mot de 36 bits qui contient le marqueur de début de paquet, un bit de parité généré et vérifié par les *wrappers* pour le contrôle d'intégrité de bout en bout, ainsi que toutes les informations de routage. Le corps du paquet est composé d'un nombre quelconque (qui peut être nul) de mots de 36 bits. Chaque mot de 36 bits du corps du paquet contient 32 bits de données utiles, un bit de parité généré et vérifié par les *wrappers* pour le contrôle d'intégrité de bout en bout, ainsi que 3 bits qui caractérisent le mot. Le dernier mot contient le marqueur de fin de paquet [4].

II-3 – Stratégie de routage

Le réseau *SPIN* possède une stratégie de routage distribuée. Chaque routeur *RSPIN* est indépendant, et prend les décisions d'aiguillage du paquet et/ou de mémorisation sans l'intervention d'une synchronisation centrale. Le routage des paquets est du type *wormhole* pour limiter la latence. Chaque routeur *RSPIN* decode l'entête du paquet qui se présente à son entrée et ouvre le chemin adéquat pour l'envoyer vers une de ses sorties sans attendre de recevoir le paquet en entier. Cela permet d'avoir des *FIFOs* d'entrée de petite taille. Un paquet peut s'étendre sur plusieurs routeurs et peut créer ainsi une contention en rendant indisponibles plusieurs liens si sa longueur est trop importante. Les chemins ouverts par les routeurs *RSPIN* successivement traversés par un paquet dépendent directement d'une part de la topologie du réseau *SPIN*, puisque la décision d'aiguillage d'un paquet doit être cohérente avec la position du routeur *RSPIN* dans le réseau *SPIN*, et d'autre part de l'adresse de destination qui est placée dans l'entête du paquet par le producteur. Ceci signifie que le schéma de communication dans le réseau *SPIN* n'est pas prédéfini, et les canaux de communications sont ouverts automatiquement selon les besoins des abonnés sans aucune reconfiguration globale du système. Le réseau *SPIN* est donc un réseau dynamique.

II-4 – Liens inter-routeur

Les liens du réseau *SPIN* sont des liens bidirectionnels point-à-point. Chaque routeur *RSPIN* possède quatre ports vers des liens montants et quatre ports vers des liens descendants. Chaque lien est full-duplex, il est constitué de deux nappes de 38 bits dont 36 bits qui constituent un mot d'un paquet et 2 bits pour un mécanisme de contrôle de flux bas niveau, par crédit d'émission de mots. Son rôle est de garantir un fonctionnement sans perte : en cas de contention, cette contention se rétro-propage vers l'émetteur primaire. Le contrôle de flux à crédit présente l'avantage de permettre l'introduction de barrières de pipeline au milieu de chaque lien, lorsque le temps de transmission entre deux routeurs devient le facteur limitant la fréquence de fonctionnement du réseau *SPIN*. A une fréquence de fonctionnement de 200 MHz, le réseau *SPIN* offre, sur chacun de ses liens, une bande passante maximale de 12,8 Gbit/s dans chaque direction. Le débit réel est évidemment plus faible que la bande passante théorique. Les simulations effectuées, sur un réseau de 32 processeurs avec des paquets de 20 mots, montrent que la saturation du réseau intervient lorsqu'on dépasse 50% de la charge maximale, ce qui signifie un débit pratique d'environ 3 Gbit/s dans chaque direction pour chaque abonné. La topologie en arbre quaternaire élargi du réseau *SPIN* étant extensible, la bande passante cumulée augmente linéairement avec le nombre d'abonnés [1].

III – LE ROUTEUR *RSPIN*

En détectant un paquet à son entrée, chaque routeur *RSPIN* decode l'entête du paquet et ouvre le chemin adéquat dans son *crossbar* pour aiguiller le paquet vers la sortie qui le rapproche de sa destination finale. Le routeur *RSPIN* possède 8 ports externes bidirectionnels [3]. Il est composé d'une partie opérative contenant huit files d'attente de type *FIFO* (un par port d'entrée), d'un *crossbar* 10 * 10 réalisé par un mélange de multiplexeurs et de portes *tristate*, et d'une partie contrôle qui contient les automates responsables du routage. Dans la partie opérative, les *FIFOs* sont utilisées par le mécanisme de contrôle de flux et ont par ailleurs une fonction de stockage en cas de contention. En effet, un paquet qui se trouve bloqué à l'entrée d'un routeur peut être stocké provisoirement dans la *FIFO* et ainsi libérer le lien et le routeur *RSPIN* en amont. Toutefois, si un paquet est bloqué à l'entrée d'un routeur *RSPIN*, les paquets qui sont derrière lui sont aussi bloqués même si les sorties qu'ils souhaitent emprunter sont libres. C'est le phénomène de *Head of Line Blocking*. Ce problème est résolu par deux queues centrales. Il y a

une queue centrale pour les paquets descendants qui viennent du haut et une autre pour les paquets descendants qui viennent du bas, pour éviter l'interblocage. Notons que la présence de ces queues centrales n'augmente en aucune manière la taille du routeur *RSPIN*, puisque celle-ci est définie par les nappes de fils du *crossbar*.

La partie contrôle du routeur *RSPIN* est distribuée : Il existe ainsi un automate de contrôle pour chaque port d'entrée, et un automate de contrôle pour chaque port de sortie. La partie contrôle d'entrée (*icb*) decode l'entête du paquet pour en extraire l'adresse de destination. En fonction de cette adresse et de la position du routeur *RSPIN* dans le réseau *SPIN*, le bloc *icb* envoie une requête vers le port de sortie qui rapproche le plus le paquet de sa destination finale. Pour choisir la sortie adéquate, le bloc *icb* prend en compte l'état des différents ports de sortie, qui peuvent être désactivés, ou alloués à un autre port d'entrée. Le routage des paquets descendants dans le réseau *SPIN* est déterministe. Si la sortie voulue est occupée, le paquet a la possibilité de faire un tour dans la queue centrale afin de ne pas encombrer le réseau. En revanche, les quatre liens montants du routeur *RSPIN* ont la particularité d'être complètement équivalents, le routage des paquets montants est donc un routage adaptatif, ce qui permet de limiter les contentions.

Le bloc de contrôle de sortie (*ocb*) est conçu comme une machine d'état fini qui gère l'état d'allocation des sorties aux entrées, c'est à dire la configuration dynamique du *crossbar*. Une priorité tournante garantit l'absence de famine.

Pour minimiser le temps de cycle, la décision d'allocation d'un port de sortie à un port d'entrée est pipelinée en deux cycles. Avec cette stratégie d'allocation pipelinée en deux cycles, le temps de traversée moyen d'un routeur *RSPIN* est de 2,5 cycles.

IV – MÉCANISME DE TEST

Le test est un problème crucial : Un réseau *SPIN* à 32 ports contient 16 routeurs *RSPIN* interconnectés par 80 nappes de 76 fils chacune (soit plus de 6000 fils). Le problème du test se pose donc à un double niveau : Il faut tester individuellement chaque routeur *RSPIN*, dans toutes les configurations possibles du *crossbar*, et il faut également tester l'ensemble des nappes de fils entre les routeurs *RSPIN*. Le routeur *RSPIN* se prête mal aux méthodes de test habituellement utilisées dans le domaine de la micro-électronique. Le taux de couverture du test aléatoire appliqué au routeur *RSPIN* ne dépasse pas les 30%. Nous avons donc introduit différents mécanismes matériels dans le routeur *RSPIN* pour faciliter le test [3]. Le routeur *RSPIN* contient un mécanisme d'autotest (*BIST*), fonctionnant à la fréquence nominale, et permettant de tester de façon exhaustive tous les chemins de données à travers le *crossbar* et les *FIFOs*, ainsi que toute la logique de contrôle contenue dans les automates. Dans un réseau *SPIN*, il est possible de tester en parallèle tous les routeurs *RSPIN* du réseau. Le routeur *RSPIN* contient également un mode pour le test des nappes de fils d'interconnexion entre les routeurs *RSPIN*. Dans ce mode, la fonction de routage des routeurs *RSPIN* (c'est à dire la décision d'allouer un port de sortie à un port d'entrée) est modifiée de façon à avoir un routage totalement déterministe. Il est alors possible de construire un chemin eulérien, tel qu'un paquet injecté sur un port particulier du réseau *SPIN* traverse successivement tous les routeurs *RSPIN* du réseau *SPIN*, en passant une fois à travers chaque lien, avant de revenir à son point de départ.

V – RÉALISATION MATÉRIELLE

Le *placement/routage* du routeur *RSPIN* a été réalisé en deux étapes. Le *placement/routage* du chemin de données a été optimisé et décrit de façon explicite au moyen d'un langage procédural. Le *placement/routage* de la partie contrôle a été effectué ensuite avec un outil automatique, auquel on a donné certaines directives de localité, dans la place restée disponible.

Les dimensions du routeur *RSPIN* sont déterminées en hauteur par la hauteur des *FIFOs*, et en largeur par les nappes de fils des ports d'entrée/sortie. La macro-cellule *RSPIN* a une surface totale de $0,24 \text{ mm}^2$ pour un procédé de fabrication 0.13 micron de *ST Microelectronics* (fig. 1).

Le réseau *SPIN* à 32 ports est composé de 16 routeurs (fig. 2), soit deux rangées de huit routeurs *RSPIN*. Il occupe une surface totale de 3.9 mm^2 (fig. 3) dont environ 30% est occupée par les *FIFOs* (en noir sur la fig. 3). Toutes les nappes de fils d'interconnexion entre les routeurs peuvent être dessinées au dessus des routeurs eux-même en métal 4,5 et 6.

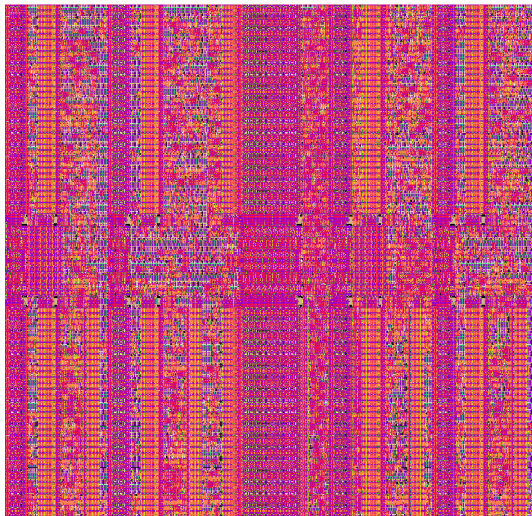


FIG. 1 – Layout du routeur *RSPIN*

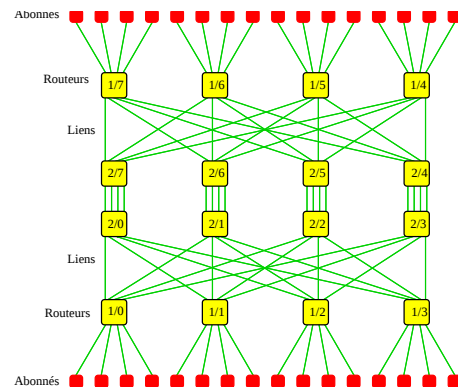


FIG. 2 – Topologie du réseau *SPIN* à 32 ports

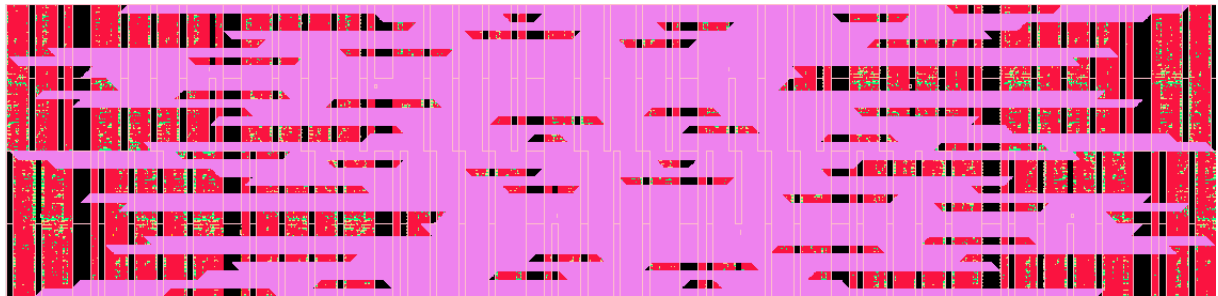


FIG. 3 – Layout du réseau *SPIN* à 32 ports

VI – CONCLUSION

Ce travail a permis de démontrer qu'il était possible d'implanter effectivement un réseau *SPIN* à 32 ports dans une surface de silicium inférieure à 4 mm^2 , pour un procédé de fabrication CMOS 0.13 micron . Le dessin des masques utilise la technologie de "dessin symbolique" de la chaîne *Alliance*, ce qui permet d'exploiter la macro-cellule *SPIN32* pour n'importe quel procédé de fabrication comportant au moins 6 niveaux de métallisation. La bande passante d'un tel réseau permet de fournir à chaque abonné une bande passante de 3 Gbit/s dans les deux sens, dans une hypothèse de fonctionnement à 200 MHz.

RÉFÉRENCES

- [1] *A generic architecture for on-chip packet-switched interconnections*, P. Guerrier, A. Greiner, Proceedings of Design Automation and Test in Europe, 2000.
- [2] *Un réseau d'interconnexion pour systèmes intégrés*, P. Guerrier, Thèse, Université Pierre et Marie Curie, 2000.
- [3] *Réseau SPIN à 32 ports*, A. Andriahantenaina, A. Greiner, Rapport technique, Université Pierre et Marie Curie, 2001.
- [4] *S.P.I.N.*, A. Andriahantenaina, A. Greiner, Rapport technique, Université Pierre et Marie Curie, 2001.
- [5] *Les wrapper VCI/SPIN*, H. Charlery, A. Greiner, Rapport technique, Université Pierre et Marie Curie, 2001.