

Nouvelles topologies pour les réseaux sur puce

Yann Kieffer¹

G-SCOP ; Institut Polytechnique de Grenoble ; 46 avenue Félix Viallet, F-38031 Grenoble Cedex, France
yann.kieffer@g-scop.grenoble-inp.fr

Mots-Clés : *réseaux sur puce, topologies, diamètre, contraintes physiques*

1 Contexte

L'intégration toujours grandissante de composants au sein d'un unique circuit intégré a permis le développement de systèmes sur puce (System on Chip : SoC) qui contiennent tous les composants fonctionnels d'un système entier : circuits dédiés à certains types de calculs, mémoires, pilotes de périphériques, etc. Ainsi, ce qui était obtenu auparavant par l'assemblage de plusieurs puces sur une carte, est aujourd'hui réalisé sur une unique puce.

Mais avec la miniaturisation des circuits viennent de nouveaux défis. L'un d'entre eux concerne les interconnexions. En effet, la diminution de la période d'horloge et de la finesse de gravure empêchent aujourd'hui de réaliser des circuits totalement synchrones ; on parle alors de circuits localement synchrones et globalement asynchrones (globally asynchronous, locally synchronous : GALS).

Le circuit est alors organisé en îlots de synchronisme, et ces îlots sont connectés par un réseau sur puce (Network on Chip : NoC) [1]. Ces réseaux sont conçus à l'image des réseaux informatiques classiques, avec néanmoins certaines spécificités, qui découlent de l'environnement de ces réseaux.

2 Le cas des SoCs homogènes

Le cas particulier de SoCs qui nous intéresse ici est celui d'un système organisé géométriquement en un pavage en grille de tuiles rectangulaires de mêmes dimensions (tile-based SoC). Les réseaux utilisés peuvent alors être choisis réguliers ; c'est le choix naturel lorsque les tuiles ont la même fonctionnalité, comme dans le cas des multi-processeurs sur puce (CMP : chip multiprocessor).

Les topologies usuellement employées sont celles provenant de la culture des réseaux : anneau, grille, tore, hypercube. Mais les contraintes de performance et les spécificités propres des réseaux sur puce - comme typiquement la limite sur le nombre de sorties d'un routeur - nous poussent à réexaminer de plus près la question des topologies pour cet usage particulier.

3 Pour une remise en cause des usages en matière de topologies

Dans le monde industriel, la seule topologie effectivement employée à ce jour pour les réseaux sur puce homogènes est la grille. Les raisons généralement invoquées pour justifier ce choix sont la simplicité de l'architecture du réseau et la simplicité du routage.

D'autres architectures peuvent être envisagées. En particulier, la réduction du diamètre semble être importante pour réduire à la fois la latence, le débit et la congestion du réseau. Pour ces qualités, le réflexe usuel est d'envisager les réseaux en hypercube.

Nous nous proposons de mathématiser ce problème, afin de pouvoir à la fois donner des pistes de topologies intéressantes, mais aussi des preuves d'impossibilité.

4 Modélisation et analyse

Une topologie d'interconnexion est représentée par un graphe orienté, les topologies symétriques donnant alors lieu à des graphes symétriques, c'est-à-dire non orientés. Les performances du réseau, en termes de débit et de latence, sont liées aux longueurs des plus courts chemins entre les sommets du graphe. C'est pourquoi le diamètre est un indicateur très regardé ; mais nous nous intéresserons également à la distance moyenne entre les sommets du réseau.

Les contraintes physiques de réalisation des routeurs imposent une limite sur le degré maximum sortant des sommets, qu'il faut donc s'efforcer de garder bas. Enfin, il doit être possible de plonger le graphe dans la grille, de sorte que les images de ses arcs soient les moins longues possibles, au sens de la distance de Manhattan.

Après l'exposé de ce modèle, nous présenterons les caractéristiques des architectures classiquement envisagées pour les réseaux. Nous donnerons ensuite des exemples de topologies prometteuses : une topologie particulièrement adaptée aux réseaux à 20 sommets, et un schéma général pour construire des topologies pour la grille à partir de topologies linéaires. Nous démontrerons également quelques résultats d'impossibilité.

Enfin, nous proposerons une explication probable du fait que l'hypercube ait pu, sur un plan culturel, entraver la recherche de topologies alternatives.

Références

- [1] Giovanni De Micheli and Luca Benini, editors. *Networks on Chips : Technology and Tools*, The Morgan Kaufmann Series in Systems on Silicon. Morgan Kaufmann, July 2006.
- [2] Mohammad Hosseinabady, Mohammad Reza Kakoei, Jimson Mathew, and Dhiraj K. Pradhan. De bruijn graph as a low latency scalable architecture for energy efficient massive nocs. *Design, Automation and Test in Europe Conference and Exhibition*, 0 :1370–1373, 2008.
- [3] Lih-Hsing Hsu and Cheng-Kuan Lin. *Graph Theory and Interconnection Networks*. CRC, September 2008.
- [4] Vijay Lakamraju, Israel Koren, and C. M. Krishna. Filtering random networks to synthesize interconnection networks with multiple objectives. *IEEE Trans. Parallel Distributed Systems*, 13, 2002.