



GPU-body: Simulações de n-corpos em GPU

Lucas Brahm, César Augusto Missio Marcon (orientador)

Faculdade de Informática, PUCRS

Resumo

Evoluções na tecnologia de fabricação de Circuitos Integrados (CIs) e o crescimento do número de transistores por unidade de área nos CIs permitem implementar diversos componentes em um único CI. Quando esses componentes atuam como um sistema, este CI é denominado de SoC (*System-on-Chip*). Buscando atender requisitos de escalabilidade, paralelismo de comunicação, vazão e consumo de energia eficiente, foi estabelecida uma nova arquitetura de comunicação denominada NoC (*Network-on-chip*) [1]. Em uma NoC, dezenas de elementos de processamento, em inglês *Processing Elements* (PEs), podem ser integrados perfazendo um sistema multiprocessado intrachip, do inglês *Multiprocessor SoC* (MPSoC) [2].

NoCs são usualmente implementadas em duas dimensões (2D), onde os PEs ficam dispostos em um mesmo plano. Porém, tal disposição apresenta limitações referentes à latência e ao consumo de energia; limitações estas que se tornam ainda mais impactantes conforme aumentam a quantidade de PEs e a distância entre estes. Desta forma, a evolução natural de NoCs foi utilizar planos sobrepostos, adicionando mais uma dimensão ao processo de fabricação de CIs, resultando em NoCs tridimensionais (NoCs 3D) [3]. A terceira dimensão permite diminuir a distância de comunicação e o número de saltos necessários para o tráfego de mensagens, reduzindo a latência e aumentando a vazão na rede, e consequentemente reduzindo o tempo de execução da aplicação.

O objetivo deste trabalho é implementar uma NoC 3D, onde os roteadores têm latência mínima de roteamento, i.e. um ciclo de relógio [4]. Esta implementação objetiva minimizar a latência média das mensagens, aumentar a vazão da rede, minimizar o consumo de energia, além de proporcionar roteadores de menor área.

A validação do projeto de uma NoC 3D requer aplicações com alto grau de paralelismo, cuja funcionalidade pode ser distribuída entre os processadores do MPSoC. Assim, utiliza-se aqui o problema de n-corpos como principal aplicação paralela para validação da plataforma. N-corpos é uma técnica em que deve ser calculado a interação entre todas as n partículas (i.e. corpos), o que normalmente implica utilizar um algoritmo de complexidade $O(n^2)$ [5][6]. Esta técnica permite obter alto grau de paralelismo durante as iterações, o que torna adequado para arquiteturas como MPSoCs e GPUs. Sendo que a aplicação construída sobre uma GPU pode ser posteriormente mapeada para um MPSoC.

Palavras-chave

GPU; NoC 3D; Aplicações paralelas

Referências Bibliográficas

- [1] C. Marcon, M. Kreutz, A. Susin, N. Calazans. **Models for embedded application mapping onto NoCs: timing analysis.** *IEEE International Symposium on Rapid System Prototyping (RSP)*, pp. 17-23, 2005.
- [2] E. Carvalho, C. Marcon, N. Calazans, F. Moraes. **Evaluation of static and dynamic task mapping algorithms in NoC-based MPSoCs.** *International Symposium on System-on-Chip (SOC)*, pp. 87-90, 2009.
- [3] Y. Ghidini, T. Webber, E. Moreno, I. Quadros, R. Fagundes, C. Marcon. **Topological impact on latency and throughput: 2D versus 3D NoC comparison.** *Symposium on Integrated Circuits and Systems Design (SBCCI)*, pp. 1-6, 2012.
- [4] Y. Ghidini, M. Moreira, L. Brahm, T. Webber, N. Calazans, C. Marcon. **Lasio 3D NoC vertical links serialization: Evaluation of latency and buffer occupancy.** *Symposium on Integrated Circuits and Systems Design (SBCCI)*, pp. 1-6, 2013.
- [5] Y. Miki, D. Takahashi, M. Mori. **Highly scalable implementation of an image-body code on a GPU cluster.** *Computer Physics Communications*, v. 184, Issue 9, pp. 2159-2168, Sep. 2013.
- [6] J. Bédorf, E. Gaburov, S. Zwart. **A sparse octree gravitational N-body code that runs entirely on the GPU processor.** *Journal of Computational Physics*, v. 231, Issue 7, pp. 2825-2839, Apr. 2012.