

Proceeding Series of the Brazilian Society of Computational and Applied Mathematics

Melhorando o Desempenho da Computação de Estênceis em GPUs

Ricardo K. Lorenzoni¹

DCEEng, Universidade Regional do Noroeste do Estado do Rio Grande do Sul (UNIJUI), Ijuí, RS

Matheus S. Serpa²

Instituto de Informática, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS

Edson Luiz Padoin³

DCEEng, Universidade Regional do Noroeste do Estado do Rio Grande do Sul (UNIJUI), Ijuí, RS

Philippe O. A. Navaux⁴

Instituto de Informática, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS

1 Introdução

A computação de estênceis está presente em uma grande gama de aplicações científicas, e é responsável por boa parte do tempo de computação destas aplicações. Devido a isto, a utilização de Computação de Alto Desempenho nestas aplicações científicas ganha importância, especialmente a computação em arquiteturas paralelas almejando reduzir o tempo de execução de problemas científicos [1].

Uma das características da computação de estênceis é que para calcular o valor de um ponto, depende-se do seu valor e de seus vizinhos no passo temporal anterior [3]. Além disso, este tipo de cálculo possui baixa intensidade computacional, sendo assim caracterizado como *memory-bound*. Considerando isto, um dos aspectos mais importantes a se considerar no desenvolvimento de uma aplicação de computação de estênceis é a organização da hierarquia de memória [2]. Nesse sentido, este trabalho tem como objetivo melhorar o desempenho uma aplicação de computação de estênceis em GPU através da otimização dos acessos a memória do dispositivo.

2 Metodologia

A aplicação selecionada realiza a computação de um estêncil em formato de estrela com tamanhos de braço de 1, 2, 3, 4 e 5 pontos em cada direção a partir do ponto central. Para avaliação, foram executados testes em uma matriz tridimensional com 32 a 2048 pontos

¹ricardo.lorenzoni@unijui.edu.br

²msserpa@inf.ufrgs.br

³padoin@unijui.edu.br

⁴navaux@inf.ufrgs.br

na direção X variando em passos de 32 e 256 pontos nas direções Y e Z . Os testes foram executados em uma placa gráfica (GPU) NVIDIA Tesla K80. Os resultados são média de 30 execuções com intervalo de confiança de 95%.

3 Resultados e Conclusões

A Figura 1 apresenta os resultados da implementação com e sem otimização de acesso a memória cache somente leitura. Com o seu uso busca-se localidade dos dados e reuso de registradores de memória.

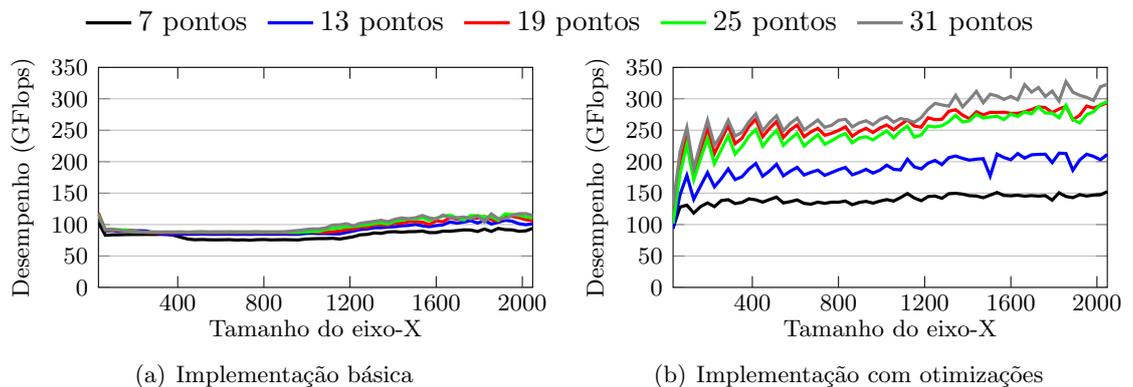


Figura 1: Resultados das implementações de computação de *estênceis*

Os resultados demonstram que o uso correto da memória da GPU possibilita ganhos desempenho. Também percebe-se que, quanto maior o tamanho do braço do estêncil, maior o desempenho da aplicação. O uso correto da hierarquia de memória da GPU é fundamental para acelerar este tipo de computação obtendo ganhos desempenho de até 2,82 vezes para um estêncil de 31 pontos.

Referências

- [1] C. A. Navarro, N. Hitschfeld-Kahler and L. Mateu. A Survey on Parallel Computing and its Applications in Data-Parallel Problems Using GPU Architectures, *Communications in Computational Physics*, 2014.
- [2] M. S. Serpa, E. H. M. Cruz, F. B. Moreira, M. Diener, P. O. A. Navaux. Impacto do Subsistema de Memória em Arquiteturas CPU e GPU, *XVII Simpósio em Sistemas Computacionais de Alto Desempenho*, 2016.
- [3] T. C. Nasciutti and J. Panetta. Impacto da arquitetura de memória de gpgpus na velocidade de computação de estênceis. *XVII Simpósio de Sistemas Computacionais*, 2016.