

Computação Astronômica com GPUs

Resumo

A presente proposta, assinada por uma expressiva fração dos astrônomos brasileiros que desenvolvem ou utilizam-se de simulações numéricas de alto desempenho, visa a aquisição de uma servidora GPU com processamento de pico de ~4 Tflops, e a realização de um workshop de computação em GPUs no âmbito do INCT-A. Essa servidora estará disponível a todo pesquisador brasileiro interessado em desenvolver novas aplicações, ou testar suas aplicações existentes, nesta arquitetura. A tecnologia de *GPU computing* permite atingir velocidades de processamento muito elevadas a um custo muito baixo, mas atualmente sua utilização, fundamentada em uma estrutura de *software* própria, freqüentemente requer um importante investimento em desenvolvimento por parte do pesquisador. O objetivo desta proposta é prover à comunidade um ambiente eficiente para desenvolvimento e testes de aplicações numéricas em GPU. O pequeno investimento solicitado justifica-se plenamente tendo-se em vista a crescente demanda da comunidade por processamento, pois fornecerá os subsídios necessários para fundamentar tecnicamente qualquer futuro investimento nesse sentido.

Introdução

“(...) l’Astronomie nous apprenait à ne pas nous effrayer des grands nombres.” H. Poincaré¹

No decorrer da história, a Astronomia teve várias vezes o importante papel de quebrar paradigmas da humanidade, ao ponto de Poincaré ter declarado que é ela a responsável por nos transformar em uma alma capaz de compreender a natureza. Para isso a Astronomia sempre desenvolveu, ou utilizou, as mais modernas técnicas

¹“(...) a Astronomia nos ensinava a não temer os grandes números (...)”, H. Poincaré, in *La Valeur de la Science*

numéricas e analíticas disponíveis em cada época, literalmente desbravando os limites do cálculo.

Não por acaso, atualmente a pesquisa em Astronomia tem se defrontado com problemas teóricos cada vez mais complexos e com um crescente volume de dados, que demandam recursos computacionais cada vez maiores e algoritmos cada vez mais inteligentes. Logo, o investimento em processamento de alto desempenho, tanto através da capacitação para o uso de novas tecnologias, quanto através da aquisição de novos equipamentos, é essencial para o desenvolvimento e para a competitividade da Astronomia brasileira durante a próxima década.

Muitos dos objetivos estratégicos do INCT-A dependem, direta ou indiretamente, da expansão da capacidade de processamento instalada no Brasil. A implantação de uma rede de observatórios virtuais, por exemplo, passa pela criação de uma rede de processamento em grade (*grid processing*) eficiente, assim como a redução e/ou análise do grande volume de dados gerados pelos projetos da Astronomia moderna (GEMINI, SOAR, SDSS, e futuramente os ELTs, o LSST e o Gaia).

Devido a impossibilidade de se manter o ritmo de aumento de velocidade de cálculos seqüenciais, atualmente a computação paralela vem tomando cada vez mais espaço, como pode ser observado a partir da crescente popularização de processadores complexos com dois ou quatro núcleos. Entretanto, existe uma grande classe de problemas que pode ser tratada por meio de um grande número de núcleos mais simples, com a função de repetir muitas vezes uma mesma operação. Denominado SIMD/SPMD (*Single Instruction/Process Multiple Data*), essa classe de problemas aparece em diversas áreas da Astronomia, e uma recente tecnologia denominada *GPU computing* tem atraído a atenção da comunidade científica por apresentar um excepcional custo benefício em sua resolução.

Uma GPU (*graphic processing unit*) é um processador especializado, que até poucos anos atrás era usado exclusivamente em aplicações gráficas. GPUs modernas têm uma estrutura altamente paralelizada, composta atualmente por várias centenas de núcleos de processamento. Quando devidamente integradas com uma CPU, as

unidades de processamento com GPUs são capazes de atingir desempenho muito superiores aos de CPUs isoladas para cálculos numéricos. Por exemplo, uma servidora GPU Tesla M1060, da nVidia, composta de dois processadores Xeon 5500 e duas GPUs com 240 núcleos, é capaz de atingir 2 Tflops (2×10^{12} operações de ponto flutuante por segundo), sendo que um processador Xeon 5500 isolado atinge apenas 40 Gflops. Quando comparamos o custo por Gflop de uma CPU (tipicamente US\$ 50/Gflop) com um sistema com GPUs (aproximadamente US\$ 6/Gflop) compreende-se o crescente interesse recente da comunidade científica em *GPU computing*. Um exemplo do ganho proporcionado por esse tipo de processamento pode ser encontrado a seguir.

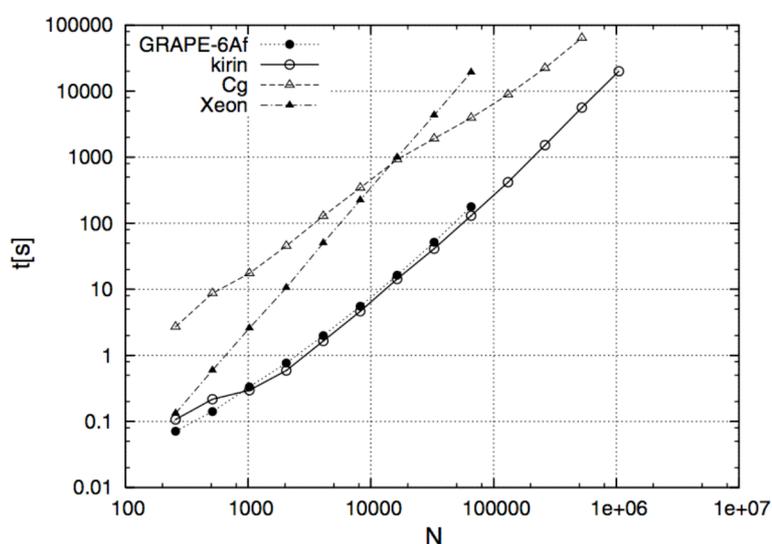


Figura 1. Comparação de performance para um código de N-corpos entre um hardware especializado (GRAPE-6Af), uma GPU programada em linguagem CUDA (kirin), uma GPU programada em Cg (Cg) e uma CPU Intel Xeon. (Belleman et al. 2008, *New Astron.*, 13, 103)

Nos últimos anos tem-se observado alguns exemplos da utilização de GPUs tanto em Astronomia observacional (eg. Murchison Widefield Array) quanto em teórica (eg. Simulações de N-corpos de Belleman, Bédorf e Portegies Zwart). O interesse no uso desta nova tecnologia pode ser verificado pelo aumento nos últimos dois anos do número de artigos científicos voltados especificamente para os aspectos computacionais do uso das GPUs em problemas como simulações numéricas complexas, de transferência radiativa e, mais recentemente, de correlação em interferômetros. Dado seu ótimo custo-benefício, a tecnologia GPU apresenta-se

como uma boa opção para futuros investimentos em computação de alto desempenho, desde que o *overhead* de programação não seja demasiado grande para o problema em questão, pois a utilização de GPUs em toda sua potencialidade implica muitas vezes em alterações no código fonte, e ocasionalmente até mesmo no algoritmo, utilizado.

Projetos

“On aurait pu raisonner de même du temps de Ptolémée; alors aussi, on croyait tout savoir, et on avait encore presque tout à apprendre.” H. Poincaré²

A proposta aqui apresentada visa principalmente a capacitação e alavancagem no uso de GPUs na Astronomia brasileira. Isso significa que a infra-estrutura montada atenderá a qualquer demanda de qualquer pesquisador do INCT-A que necessite testar e desenvolver seu código em um tal ambiente. Para se ter uma idéia da real demanda existente hoje na comunidade e do potencial impacto em pesquisa advindo do uso da computação em GPUs, descreve-se, no anexo C, alguns dos projetos dos signatários desta proposta que farão uso da infra-estrutura computacional proposta.

Ações Propostas e Itens Solicitados

Propomos a compra, no âmbito do INCT, de uma servidora Tesla da nVidia com desempenho de aproximadamente 4 Tflops. A existência no país de uma máquina científica dotada de GPUs permitirá que os astrônomos brasileiros avaliem diretamente o tamanho do *overhead* de programação para suas aplicações e ganhem experiência prática neste tipo de paralelismo. Complementando esta iniciativa, propomos a realização de um workshop (aos moldes do AstroGPU realizado no IAS, Princeton), voltado aos aspectos práticos da programação em GPU. Acreditamos que esse workshop gerará um momento inicial entre os pesquisadores brasileiros para utilização dessa promissora tecnologia.

² “Podia-se pensar o mesmo no tempo de Ptolomeu; então também, acreditava-se tudo saber mas tinha-se quase tudo para aprender.”, H. Poincaré, in *La Valeur de la Science*

Finalmente, propomos também o estabelecimento de nós locais de capacitação em GPUs, através da compra de placas de vídeo comerciais (tais como a GTX 280 da nVidia, a um custo de aproximadamente US\$ 500,00). Estas placas cumpririam duas importantes funções: 1) por estarem instaladas na instituição do pesquisador, o desenvolvimento inicial dos aplicativos poderá ser feito localmente, com o benefício de se evitar os gargalos de rede existentes no Brasil; 2) o desenvolvimento local possibilitará um uso mais eficiente da servidora nacional, pois ela seria utilizada apenas nas fases finais de desenvolvimento e, o que é mais importante, para o processamento dos modelos.

Os itens solicitados são:

- 1) Duas servidoras Tesla M1060, da nVidia, a serem instaladas no Laboratório de Computação Científica Avançada (LCCA) do CCE/USP. **Custo aproximado: US\$ 22.000,00** (orçamento e especificações detalhadas no anexo A)
- 2) 5 licenças do Fortran Accelerator da Portland. Este é o único compilador Fortran atualmente compatível com GPUs. **Custo: US\$ 1.869,00** (fonte: <http://www.pgggroup.com>)
- 3) Verba para organização de um workshop sobre GPUs em Astronomia, a ser realizado no primeiro semestre de 2010 na Universidade Cruzeiro do Sul, em São Paulo. **Custo aproximado: R\$ 14.000,00** (Anexo B).
- 4) Aquisição de seis placas GTX 280 que serão distribuídas às instituições dos signatários desta proposta para capacitação. **Custo aproximado: US\$ 3.000,00.**

Signatários da proposta

1. Alberto G. O. Krone-Martins (IAG/USP)
2. Alex Cavaliéri Carciofi (IAG/USP)
3. Alexandre Zobot (UFSC)
4. Antônio Mário Magalhães (IAG/USP)

5. Antonio Kanaan (UFSC)
6. Basílio Santiago (UFRGS)
7. Cláudia Villega Rodrigues (INPE)
8. Charles José Bonatto (UFRGS)
9. Diego Falceta Gonçalves (Unicsul)
10. Eduardo Balbinot (UFRGS)
11. Elisabete Maria de Gouveia Dal Pino (IAG/USP)
12. Fabrício Ferrari (Unipampa)
13. Gastão B. de Lima Neto (IAG/USP)
14. Guilherme Gonçalves Ferrari (UFRGS)
15. Grzegorz Kowal (IAG/USP)
16. Horácio Dottori (UFRGS)
17. Joaquim E. R. Costa (INPE)
18. Laerte Sodr  Junior (IAG/USP)
19. Natalia Vale Asari (UFSC)
20. Marcos Tadeu dos Santos (IAG/USP)
21. Marcos Tadeu dos Santos (IAG/USP)
22. Miriani G. Pastoriza (UFRGS)
23. Paula R. T. Coelho (Unicsul)
24. Roberto Cid Fernandes Jr. (UFSC)
25. Rodrigo S. Nemmen (UFRGS)
26. Rog rio Riffel (UFRGS)
27. Sylvio Ferraz-Mello (IAG/USP)
28. Tatiana A. Michtchenko (IAG/USP)
29. Xavier Haubois (IAG/USP)
30. William Schoenell (UFSC)
31. Zulema Abraham (IAG/USP)