Pontifícia Universidade Católica
do Rio de Janeiro

Cleomar Pereira da Silva

Programação Genética Maciçamente Paralela em GPUs

Tese de Doutorado

Tese apresentada ao Programa de Pós–graduação em Engenharia Elétrica do Departamento de Engenharia Elétrica da PUC–Rio como requisito parcial para obtenção do título de Doutor em Engenharia Elétrica

Orientador: Prof. Marco Aurélio Cavalcanti Pacheco
Co–Orientador: Prof. Douglas Mota Dias
Co–Orientadora: Profa. Cristiana Barbosa Bentes

Rio de Janeiro
Setembro de 2014

Pontifícia Universidade Católica
do Rio de Janeiro

Cleomar Pereira da Silva

Programação Genética Maciçamente Paralela em GPUs

Tese apresentada ao Programa de Pós–graduação em Engenharia Elétrica do Departamento de Engenharia Elétrica do Centro Técnico Científico da PUC–Rio como requisito parcial para obtenção do título de Doutor em Engenharia Elétrica. Aprovada pela Comissão Examinadora abaixo assinada.

Prof. Marco Aurélio Cavalcanti Pacheco
Orientador
Departamento de Engenharia Elétrica — PUC–Rio

Prof. Douglas Mota Dias
Co–Orientador
Departamento de Engenharia Elétrica — PUC–Rio

Profa. Cristiana Barbosa Bentes
Co–Orientadora
Universidade do Estado do Rio de Janeiro – UERJ

Prof. Ricardo Cordeiro de Farias
Universidade Federal do Rio de Janeiro – UFRJ

Prof. Renato Portugal
Laboratório Nacional de Computação Científica – LNCC

Profa. Noemi Rodriguez
Pontifícia Universidade Católica do Rio de Janeiro – PUC–Rio

Prof. Esteban Walter Gonzalez Clua
Universidade Federal Fluminense – UFF

Prof. José Franco Machado do Amaral
Universidade do Estado do Rio de Janeiro – UERJ

Prof. José Eugênio Leal
Coordenador Setorial do Centro Técnico Científico — PUC–Rio

Rio de Janeiro, 11 de Setembro de 2014

Cleomar Pereira da Silva

Possui graduação em Engenharia Elétrica pela Universidade Federal de Santa Maria (2008) e mestrado em Engenharia Elétrica pela Pontifícia Universidade Católica do Rio de Janeiro (2011).

Ficha Catalográfica

Silva, Cleomar Pereira da

Programação genética maciçamente paralela em GPUs / Cleomar Pereira da Silva ; orientador: Marco Aurélio Cavalcanti Pacheco ; co-orientador: Douglas Mota Dias ; co-orientadora: Cristiana Barbosa Bentes. – 2014.
134 f. : il. (color.) ; 30 cm

Tese (doutorado)–Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Elétrica, 2014.
Inclui bibliografia

1. Engenharia elétrica – Teses. 2. Programação genética. 3. Inspiração quântica. 4. GPUs. 5. Código de máquina. I. Pacheco, Marco Aurélio Cavalcanti. II. Dias, Douglas Mota. III. Bentes, Cristiana Barbosa. IV. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Engenharia Elétrica. V. Título.

CDD: 621.3

À minha namorada Daniela, aos meus pais Antonio Coraci e Maria Rosa e
aos meus irmãos Edimar e Rudinéia.

Agradecimentos

Ao CNPq, à PUC-Rio e ao ICA pelo suporte financeiro prestado ao longo destes anos.

Ao Instituto Federal Catarinense por todo apoio e liberação de atividades de trabalho em benefício da minha formação.

Ao meu orientador Professor Marco Aurélio Cavalcanti Pacheco pelo incentivo e por sempre acreditar na viabilidade do meu trabalho.

Ao meu co-orientador Professor Douglas Mota Dias pelo auxilio prestado nas longas discussões técnicas sobre o funcionamento do modelo.

À minha co-orientadora Professora Cristiana Barbosa Bentes pela enorme contribuição prestada na organização das ideias.

Aos meus pais Antonio Coraci e Maria Rosa e aos meus irmãos Edimar e Rudinéia por terem me proporcionado um ambiente de vivência em família.

E em especial, à minha namorada Daniela, pela pessoa especial e querida que ela é, e por ser a principal razão de seguir em frente.

Resumo

Silva, Cleomar Pereira; Pacheco, Marco Aurélio Cavalcanti; Dias, Douglas Mota; Bentes, Cristiana Barbosa. **Programação Genética Maciçamente Paralela em GPUs**. Rio de Janeiro, 2014. 134p. Tese de Doutorado — Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

A Programação Genética permite que computadores resolvam problemas automaticamente, sem que eles tenham sido programados para tal. Utilizando a inspiração no princípio da seleção natural de Darwin, uma população de programas, ou indivíduos, é mantida, modificada baseada em variação genética, e avaliada de acordo com uma função de aptidão (*fitness*). A programação genética tem sido usada com sucesso por uma série de aplicações como projeto automático, reconhecimento de padrões, controle robótico, mineração de dados e análise de imagens. Porém, a avaliação da gigantesca quantidade de indivíduos gerados requer excessiva quantidade de computação, levando a um tempo de execução inviável para problemas grandes. Este trabalho explora o alto poder computacional de unidades de processamento gráfico, ou GPUs, para acelerar a programação genética e permitir a geração automática de programas para grandes problemas. Propomos duas novas metodologias para se explorar a GPU em programação genética: compilação em linguagem intermediária e a criação de indivíduos em código de máquina. Estas metodologias apresentam vantagens em relação às metodologias tradicionais usadas na literatura. A utilização de linguagem intermediária reduz etapas de compilação e trabalha com instruções que estão bem documentadas. A criação de indivíduos em código de máquina não possui nenhuma etapa de compilação, mas requer engenharia reversa das instruções que não estão documentadas neste nível. Nossas metodologias são baseadas em programação genética linear e inspiradas em computação quântica. O uso de computação quântica permite uma convergência rápida, capacidade de busca global e inclusão da história passada dos indivíduos. As metodologias propostas foram comparadas com as metodologias existentes e apresentaram ganhos consideráveis de desempenho. Foi observado um desempenho máximo de até 2,74 trilhões de GPops (operações de programação genética por segundo) para o *benchmark* Multiplexador de 20 bits e foi possível estender a programação genética para problemas que apresentam bases de dados de até 7 milhões de amostras.

Palavras–chave

Programação genética; inspiração quântica; GPUs; código de máquina.

Abstract

Silva, Cleomar Pereira; Pacheco, Marco Aurélio Cavalcanti (Advisor); Dias, Douglas Mota; Bentes, Cristiana Barbosa. **Massively Parallel Genetic Programming on GPUs**. Rio de Janeiro, 2014. 134p. PhD Thesis — Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

Genetic Programming enables computers to solve problems automatically, without being programmed to it. Using the inspiration in the Darwin's Principle of natural selection, a population of programs or individuals is maintained, modified based on genetic variation, and evaluated according to a fitness function. Genetic programming has been successfully applied to many different applications such as automatic design, pattern recognition, robotic control, data mining and image analysis. However, the evaluation of the huge amount of individuals requires excessive computational demands, leading to extremely long computational times for large size problems. This work exploits the high computational power of graphics processing units, or GPUs, to accelerate genetic programming and to enable the automatic generation of programs for large problems. We propose two new methodologies to exploit the power of the GPU in genetic programming: intermediate language compilation and individuals creation in machine language. These methodologies have advantages over traditional methods used in the literature. The use of an intermediate language reduces the compilation steps, and works with instructions that are well-documented. The individuals creation in machine language has no compilation step, but requires reverse engineering of the instructions that are not documented at this level. Our methodologies are based on linear genetic programming and are inspired by quantum computing. The use of quantum computing allows rapid convergence, global search capability and inclusion of individuals' past history. The proposed methodologies were compared against existing methodologies and they showed considerable performance gains. It was observed a maximum performance of 2,74 trillion GPops (genetic programming operations per second) for the 20-bit Multiplexer benchmark, and it was possible to extend genetic programming for problems that have databases with up to 7 million samples.

Keywords

Genetic programming; quantum inspired; graphics processing units; machine code.

Sumário

Lista de figuras

Lista de tabelas

Eu tentei 99 vezes e falhei, mas na centésima tentativa eu consegui, nunca desista de seus objetivos mesmo que esses pareçam impossíveis, a próxima tentativa pode ser a vitoriosa.

Albert Einstein