

1 Introdução

As Unidades de Processamento Gráfico (GPUs), ou Placas Gráficas, foram originalmente desenvolvidas com o propósito de renderização gráfica. Contudo, nos últimos anos, o desempenho na realização de operações de ponto flutuante das GPUs superou em muito o desempenho das CPUs (Unidades Centrais de Processamento). As GPUs possuem um elevado número de elementos de processamento e uma grande largura de banda de memória. Constituem assim, uma poderosa ferramenta, tanto para o processamento gráfico, quanto para a execução de trechos de código de programas de propósito geral.

Recentemente, muitas aplicações científicas foram portadas para a plataforma de hardware das GPUs. Assim, podemos citar o estudo de Anderson, Goddard & Schröder (2007) para execução do Método de Monte Carlo Quântico nesta plataforma. As simulações envolvendo Elementos Finitos também se apresentam adequadas (Göddeke, Strzodka e Turek, 2007). Stone, Phillips, *et al.* (2007) relataram a aceleração de cálculos de mecânica molecular e Meel, Arnold, *et al.* (2008) descrevem o emprego de GPUs para simulações de dinâmica molecular clássica. Outra área que se beneficia da alta capacidade de cálculo das GPUs é a Simulação de Nanodispositivos, tais como dispositivos nano-fotônicos (Kelmelis, Durbano, *et al.*, 2006; Price, Humphrey e Kelmelis, 2007).

No campo dos cálculos quânticos da estrutura eletrônica, Yasuda (2008b) propôs um algoritmo para o cálculo da integral da repulsão de dois elétrons em GPUs. Mais recentemente, Ufimtsev & Martínez (2008b) revisaram este algoritmo e propuseram novas estratégias, obtendo resultados promissores. O campo autoconsistente do algoritmo de Hartree-Fock, utilizado no GAMESS (*General Atomic and Molecular Electronic Structure System*), foi acelerado por GPUs (Ufimtsev e Martínez, 2008a). Métodos de maior acurácia na energia total calculada, como os que empregam a teoria da perturbação de segunda ordem, foram parcialmente modificados para execução em GPUs por Vogt, Olivares-Amaya, *et al.* (2008).

A Teoria do Funcional da Densidade (DFT) está entre os mais populares e versáteis métodos disponíveis para estudos de química computacional e física do estado sólido. É usada para estudar a estrutura eletrônica (principalmente o estado fundamental) de sistemas de muitos corpos, em particular átomos, moléculas e sólidos. O seu formalismo foi estabelecido a partir dos dois teoremas de Hohenberg & Kohn (1964). Eles demonstraram que em princípio a densidade eletrônica contém toda a informação que pode ser obtida da função de onda de muitos elétrons. Conforme será detalhado no Capítulo 2, o cálculo da energia de um sistema físico por DFT geralmente encontra-se subdividido em cinco termos principais.

Em trabalhos relacionados à DFT, Yasuda (2008a) obteve bons resultados com o termo de troca e correlação da energia total em GPU. Os termos da energia relativos ao pseudopotencial do algoritmo BigDFT, o qual usa conjuntos de base com funções *wavelets*, foi estudado por Genovese, Ospici, *et al.* (2009), obtendo-se acelerações de até 20 vezes em algumas operações. Genovese, Ospici, *et al.* (2009) calcularam a solução da Equação de Poisson na CPU e comentaram que a solução da mesma na GPU teria eliminado a necessidade de duas transferências de memória para a variável densidade. A transferência de memória entre CPU e GPU, através do barramento PCI Express, constitui um dos potenciais gargalos que deve ser evitado para obter bom desempenho.

O SIESTA (*Spanish Initiative for Electronic Simulations with Thousands of Atoms*) é um método autoconsistente da Teoria do Funcional da Densidade que usa o pseudopotencial para representação do caroço iônico, orbitais atômicos numéricos e funções de base localizada (Soler, Artacho, *et al.*, 2002; Ordejón, Artacho e Soler, 1996). Este algoritmo representa um esforço para o desenvolvimento de um método *ab initio* DFT autoconsistente de ordem N , $O(N)$, para CPU. A determinação do hamiltoniano autoconsistente em $O(N)$ operações é difícil de ser obtida usando-se ondas planas como funções de base. Isto levou à escolha de um conjunto de bases atômicas localizadas para o SIESTA. Contudo, algumas funções ainda necessitam um número superior de operações, tais como a Transformada de Fourier empregada na resolução da Equação de Poisson, a qual é de ordem $N \log(N)$.

1.1. Motivação

Este trabalho envolve o estudo de duas áreas distintas do conhecimento, a Teoria do Funcional da Densidade (DFT) e a Programação de Propósito Geral em Unidades de Processamento Gráfico (GPGPU). O poder computacional crescente das GPUs pode ser empregado para auxiliar na realização dos cálculos envolvidos na DFT. Assim, o estudo destas duas áreas distintas visa prestar uma contribuição no sentido de apontar as alterações da DFT para execução acelerada por GPU. A modificação de partes do código e a realização de simulações de propriedades físicas empregando o SIESTA são utilizadas para comprovar os ganhos de desempenho que podem ser alcançados pela colaboração entre estas duas áreas.

1.2. Objetivos

O objetivo geral desta pesquisa é investigar o apoio das GPUs na aceleração dos cálculos envolvidos na Teoria do Funcional da Densidade.

Entre os objetivos específicos encontra-se a identificação dos trechos do código SIESTA que seriam adequados ao paralelismo de dados oferecido pelas placas gráficas. A alteração de algumas destas partes identificadas como sendo adequadas para GPUs. Além da realização de cálculos de propriedades físicas em estruturas, tais como nanotubos de carbono e fulerenos, com a finalidade de comprovar o ganho de velocidade da execução em GPUs para as partes alteradas.

1.3. Descrição do Trabalho

Descrição das principais etapas seguidas no desenvolvimento deste trabalho:

1. Estudo da Teoria do Funcional da Densidade:

Para isso, uma revisão bibliográfica foi realizada visando o entendimento das equações fundamentais da Teoria do Funcional da Densidade

2. Estudo do Modelo de Programação de Propósito Geral em Unidades de Processamento Gráfico:

Revisão do conhecimento bibliográfico disponível na literatura envolvendo a programação CUDA (*Compute Unified Device Architecture*) para GPUs.

3. Seleção de um Algoritmo DFT:

Foi realizada uma busca por um algoritmo eficiente, de código fonte aberto, para servir de ponto de partida para estudar a alteração de trechos de código para GPU. Diversos pacotes de programas, usados em cálculos de química quântica, foram compilados, optando-se finalmente por estudar mais detalhadamente o código fonte do SIESTA.

4. Identificação e Alteração de Partes do SIESTA:

O estudo das aproximações DFT e da forma particular usada no SIESTA para representar os termos da energia total possibilitou a identificação de trechos de código com potencial de ganhos para execução em GPU. Algumas destas partes identificadas foram reescritas usando a linguagem CUDA, tais como o cálculo do potencial de Hartree e da energia de Hartree, usando a resolução da Equação de Poisson na GPU. A contribuição para as forças de deslocamento atômicas, devidas ao potencial de Hartree, também são computadas em GPU.

5. Comprovação do Desempenho:

Verificação da correção dos resultados obtidos com a versão das funções executadas em GPU, quando comparada às funções sequenciais originais. Comparação dos tempos das simulações para avaliar a aceleração obtida para estas funções alteradas para executar em GPUs.

6. Resultados:

Propriedades físicas de nanotubos e de fulerenos foram calculadas na CPU e na GPU para comparação do desempenho. A solução da Equação de Poisson na GPU para calcular o potencial de Hartree e a energia de Hartree pode ser acelerada cerca de 30 vezes. A reordenação de dados em GPU apresenta importância para reduzir o volume de transferências de memória através do barramento PCI Express. O cálculo do dipolo elétrico pode ser acelerado em até duas ordens de grandeza.

7. Sugestões de Trabalhos Futuros:

A continuação do trabalho, com a alteração do código fonte para execução em GPUs, dos demais termos da energia total do programa SIESTA. Sugestão do emprego de técnicas inteligentes para redução do esforço necessário para obtenção de código fonte compilável para execução em GPU.

1.4. Estrutura da Dissertação

No Capítulo introdutório são listadas aplicações gerais de GPUs no meio científico, e aplicações voltadas, especificamente, para os métodos empregados na química quântica.

No Capítulo 2 é feita uma revisão bibliográfica sobre as principais aproximações empregadas na Teoria do Funcional da Densidade, visando tornar possível a obtenção de uma solução aproximada para a equação de Schrödinger em tempo hábil. Ao final deste Capítulo é feita uma ligação da teoria DFT com a forma particular empregada no SIESTA para fazer a representação dos termos da energia total no hamiltoniano.

O Capítulo 3 é dedicado a fazer uma revisão do conhecimento disponível na literatura sobre a Computação de Propósito Geral em Unidades de Processamento Gráfico. Este conhecimento será empregado para realizar as alterações de partes do código SIESTA para execução acelerada por GPU.

As partes do código SIESTA selecionadas para alteração são apresentadas no Capítulo 4. São explicadas as modificações realizadas fazendo-se uma ligação com os termos da energia do SIESTA explicados no final do Capítulo 3.

O Capítulo 5 descreve os cálculos de propriedades físicas, tais como a energia e a estrutura de bandas de nanotubos e a otimização da geometria estrutural de fulerenos. Estas simulações foram realizadas com a finalidade de comprovar a eficiência das GPUs em acelerar trechos de códigos empregados na DFT.

Por fim, as conclusões e os trabalhos futuros, identificados durante a realização desta pesquisa, são apresentados no Capítulo 6.