

5

Estudo de Casos

Neste capítulo, serão descritos os sistemas físicos e as suas propriedades calculadas com o objetivo comparar o desempenho computacional de uma GPU com o desempenho obtido com uma CPU.

Na realização deste estudo foi utilizada uma GPU Tesla C1060, a qual possui 240 elementos de processamento (1.3 GHz) e 4 GB de memória RAM com largura de banda de acesso à memória de 102 GB/segundo através de um barramento de 512 bits (NVIDIA, 2010). O computador utilizado possuía quatro destas GPUs e um processador AMD Phenom II (4 núcleos), trabalhando na freqüência de 3 GHz, com 16 GB de memória RAM. Todas as comparações foram realizadas entre uma GPU e um núcleo da CPU. A largura de banda do barramento PCI Express (*Peripheral Component Interconnect*) pode ser considerada baixa. Pois enquanto a largura de banda das transferências internas, entre memória GDDR3 e elementos de processamento, de uma Tesla C1060 é de 102 GB/segundo, a transferência pela PCI Express fica em torno de 2 GB/segundo.

5.1.

Caso 1 – Cálculo da Energia e Estrutura de Bandas de Nanotubos

Os nanotubos de carbono são estruturas cristalinas (cilíndricas) formadas por átomos de carbono (estruturas alotrópicas de carbono). Possuem alta resistência a tensão mecânica, podendo ser usados como aditivos em compostos para melhorar suas características. Dependendo do diâmetro ou da quiralidade do tubo, os nanotubos de carbono são condutores ou semi-condutores, podendo ter aplicações em circuitos micro e nano-eletrônicos.

5.1.1. Estrutura do Nanotubo

O vetor *quiral* define a estrutura atômica dos nanotubos e são usadas denominações especiais para cada caso: os nanotubos (n,n) são denominados *armchair*, enquanto os nanotubos $(n,0)$ são denominados *zigzag*. Os nanotubos (n,m) , com $n \neq m$ e $m \neq 0$, são denominados genericamente quirais, em oposição aos nanotubos *armchair* e *zigzag*, que são aquirais.

A ilustração da Figura 39 (a) mostra o cruzamento das bandas de valência e condução no nível de Fermi, sendo este comportamento característico dos nanotubos metálicos ou nanotubos *armchair* (n,n) . A Figura 39 (b) ilustra a estrutura de banda de um nanotubo $(8,0)$ semicondutor. Os nanotubos (n,m) com $n - m \neq 3i$, onde i é um número inteiro, apresentam um comportamento semicondutor. Já os nanotubos (n,m) com $n - m = 3i$ são classificados como semicondutores com *gap* quase nulo. Para este caso, o nanotubo $(12,0)$ é ilustrado na Figura 39 (c).

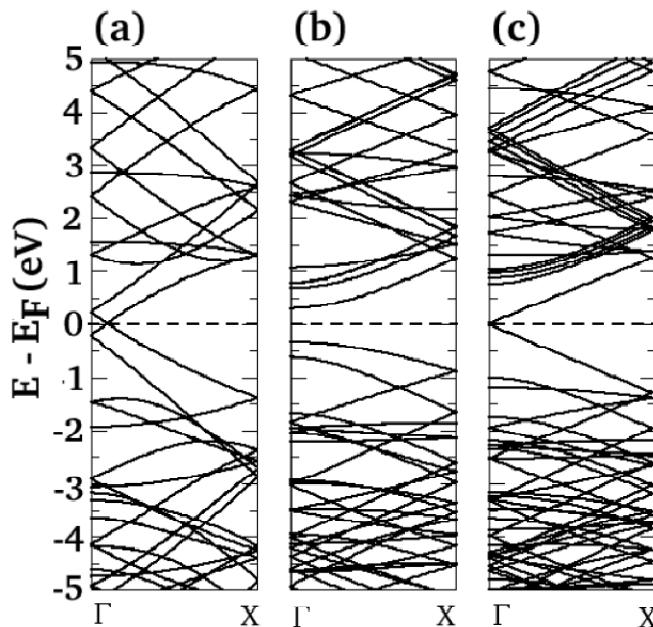


Figura 39 – Estrutura de bandas de nanotubos (a) *armchair* $(5, 5)$, (b) *zigzag* $(8, 0)$ e (c) *zigzag* $(12, 0)$. O nível de Fermi está deslocado para o zero, indicado pela linha tracejada (Silva, 2008).

Uma ilustração da estrutura geométrica da célula unitária do nanotubo $(4,2)$ com 56 átomos de carbono, utilizado para os cálculos de comparação de desempenho entre CPU e GPU realizados neste trabalho, é apresentada na Figura

40 para a vista frontal e a lateral. Uma vez que $4 - 2 \neq 3i$, então um comportamento semicondutor é esperado.

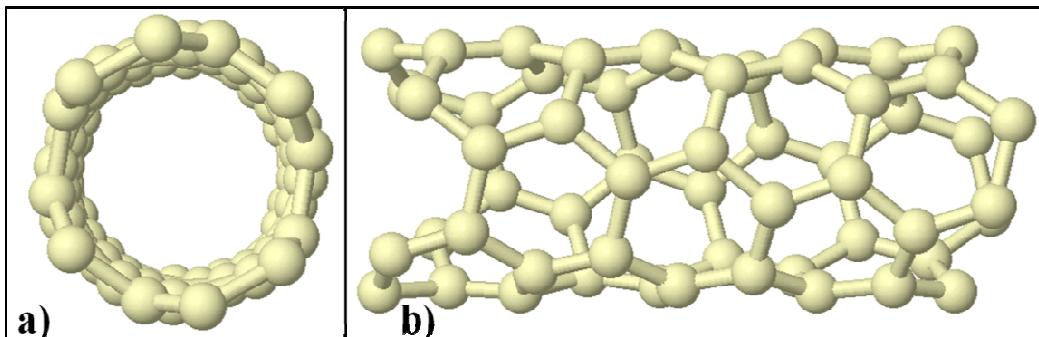


Figura 40 – Nanotubo (4,2). a) Vista Frontal, b) Vista Lateral.

5.1.2.

Resultados e Testes de Desempenho

As propriedades físicas, energia e estrutura de bandas, do nanotubo (4,2) foram calculadas por primeiros princípios fazendo-se o uso da DFT, com o emprego do programa SIESTA. A aproximação para o termo de troca e correlação utilizada foi a LDA, com a parametrização desenvolvida por Ceperley & Alder (1980).

Os cálculos de energia para o nanotubo (4,2) foram realizados duas vezes para cada teste. Uma vez com a versão seqüencial original do SIESTA na CPU e em seguida com a versão que emprega funções alteradas para execução em GPU. Isto foi feito para verificar a acurácia dos resultados produzidos com emprego de funções CUDA e para comparação de desempenho. Conforme o apresentado nas Tabela 10, Tabela 11 e Tabela 12, é possível verificar, que para este caso, o erro da Energia Total encontra-se na sexta casa decimal.

A resolução da Equação de Poisson na GPU é utilizada para calcular o potencial de Hartree. A energia de Hartree e a contribuição para as forças atômicas de deslocamento, devidas ao potencial de Hartree, envolvem somatórios e encontram-se combinadas na mesma porção do código para resolução da Equação de Poisson, sendo também calculadas na GPU. A Tabela 10 mostra os resultados obtidos com a utilização da primeira versão do *kernel* da Equação de Poisson, a qual apresentava um nível de ocupação do multiprocessador de 75 % e foi descrito no Capítulo 4.

Tabela 10 – Resultados de Aceleração com a primeira versão do *kernel* da Equação de Poisson, para o nanotubo (4,2).

	CPU	GPU	Aceleração
Energia Total	-8592.057279 eV	-8592.057281 eV	
Tempo da Equação de Poisson	988.0372 s	34.7373 s	28.4431 X

Um segundo *kernel* foi desenvolvido para resolução da Equação de Poisson na GPU. Conforme o descrito no Capítulo 4, este *kernel* possui uma maior intensidade aritmética e um menor nível de ocupação do multiprocessador, 50 %. Os resultados obtidos para este caso, Tabela 11, são ligeiramente superiores aos do primeiro *kernel*.

Tabela 11 – Resultados de Aceleração com a segunda versão do *kernel* da Equação de Poisson, para o nanotubo (4,2).

	CPU	GPU	Aceleração
Energia Total	-8592.057279 eV	-8592.057282 eV	
Tempo da Equação de Poisson	988.0372 s	33.796 s	29.2353 X

Ao portar novas funções de cálculo do dipolo elétrico e de reordenação de dados para execução em GPU, foi reduzida a necessidade de transferência de memória através do barramento PCI Express. A matriz de densidade, necessária para a resolução da Equação de Poisson, já se encontrava na memória da GPU devido à execução das outras funções. Isso contribui para aumentar a aceleração da Equação de Poisson, conforme o observado na Tabela 12.

O Cálculo do Dipolo Elétrico é uma pequena porção do código, quando comparada a Equação de Poisson, envolvendo apenas três somatórios. Não há a necessidade de operações mais dispendiosas computacionalmente, tais como a Transformada de Fourier, e não são realizadas transferências de memória. Esta função é empregada sobre a matriz de densidade, a qual é transferida para a GPU durante a Reordenação de Dados. Isso resultou num elevado valor de aceleração, de duas ordens de grandeza, Tabela 12.

A Reordenação de Dados é responsável pelas transferências de memória relacionadas com a matriz de densidade. Apesar de a transferência ser realizada empregando transferência assíncrona e execução concorrentes, a baixa intensidade aritmética desta função resulta num valor menor de aceleração de 2.9 quando comparado às demais funções, Tabela 12. Esta função foi portada para GPU por atuar sobre os mesmos conjuntos de dados que a Equação de Poisson. Portar todas as funções relacionadas com a matriz de densidades para execução em GPU

poderá eliminar a necessidade de sua transferência através do barramento PCI Express. Seria necessário transferir somente os parâmetros de inicialização e o resultado das integrais.

Tabela 12 – Resultados de Aceleração Reduzindo-se a Transferência de Dados através do Barramento PCI Express, para o nanotubo (4,2).

	CPU	GPU	Aceleração
Energia Total	-8592.057279 eV	-8592.057282 eV	
Tempo de Cálculo do Dipolo Elétrico	51.6835 s	0.2093 s	246.9350 X
Tempo de Reordenação de Dados	105.5779 s	39.0563 s	2.7032 X
Tempo da Equação de Poisson	988.0372 s	26.6871 s	37.0230 X

O tempo total de execução do SIESTA, para o nanotubo (4,2), com a Equação de Poisson, o Cálculo do Dipolo Elétrico e a Reordenação de Dados em GPU, foi de 353.2323 segundos. A versão seqüencial original do SIESTA consumiu 1390.1599 segundos. Assim, a aceleração global do código levando em consideração as chamadas de função C a partir do Fortran, para este caso, foi de 3.9355.

A estrutura de bandas do nanotubo (4,2), obtida nos cálculos realizados com o SIESTA, é apresentada na Figura 41. Pode se verificar que está coerente com o fato do nanotubo (4,2) ser um semicondutor ($4 - 2 \neq 3i$), pois não ocorre o cruzamento das bandas de valência e condução no nível de Fermi, de forma semelhante ao discutido anteriormente para a Figura 39 (b).

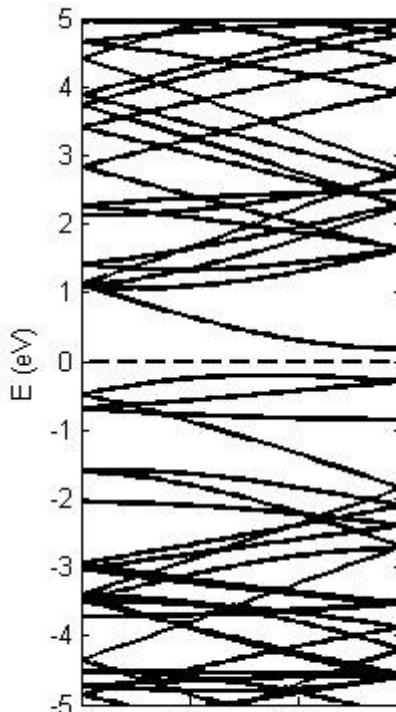


Figura 41 – Estrutura de bandas para o nanotubo (4,2). O nível de Fermi está deslocado para o zero, indicado pela linha tracejada.

5.2.

Caso 2 – Otimização da Geometria Estrutural de Fulereno

O fulereno é uma estrutura formada por átomos de carbono organizados nos vértices de um icosaedro truncado que tem a forma de uma bola de futebol (com pentágonos e hexágonos).

5.2.1.

Estrutura do Fulereno

A estrutura do fulereno é esférica, formada por hexágonos interligados por pentágonos, sendo estes últimos responsáveis pela sua curvatura e, consequentemente, por sua forma tridimensional (Kroto, Allaf e Balm, 1991). O representante mais conhecido da família dos fulerenos é o C₆₀ (com 60 carbonos), com um diâmetro de aproximadamente 1 nm. Este fulereno é apresentado na Figura 42 e é nosso segundo estudo de caso.

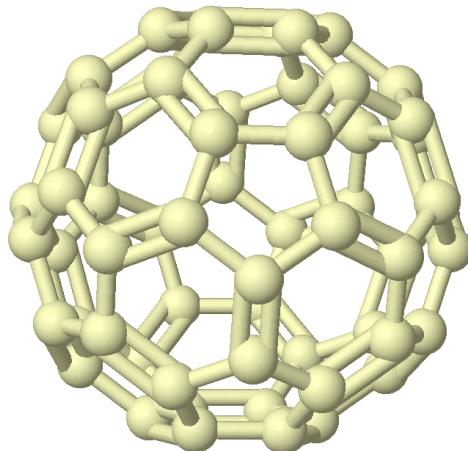


Figura 42 – Fulereno C60.

5.2.2.

Resultados e Testes de Desempenho

Uma estrutura muito próxima da estrutura final foi utilizada como ponto de partida para otimização da geometria estrutural do fulereno C60. A energia total da estrutura é calculada e pequenas variações das coordenadas atômicas são realizadas, de acordo com os resultados obtidos para as forças de deslocamento atômicas. Foi utilizada a DFT do programa SIESTA, com a aproximação GGA para o termo de troca e correlação da energia. Os cálculos são repetidos duas vezes, em CPU e em CPU-GPU, para comparação de desempenho. Com relação à acurácia, para este caso, o Erro da Energia Total encontra-se na quarta casa decimal, conforme pode ser observado nas Tabela 13 e Tabela 14. A otimização da geometria estrutural requer a realização de mais de um cálculo de energia. Isto justifica o aumento do erro, neste caso, pois o erro do cálculo de energia é acumulado entre os passos do processo iterativo de otimização da geometria.

Juntamente com o código para o cálculo do potencial de Hartree pela resolução da Equação de Poisson na GPU, encontra-se, de forma combinada, o código para o cálculo da energia de Hartree e a contribuição para as forças de deslocamento atômicas, devidas ao potencial de Hartree. A aceleração obtida pela resolução da Equação de Poisson na GPU é apresentada na Tabela 13. Neste caso foi utilizada somente a segunda versão do *kernel* da Equação de Poisson, conforme o descrito no Capítulo 4.

Tabela 13 – Resultados de Aceleração com a segunda versão do *kernel* da Equação de Poisson, para o fulereno C60.

	CPU	GPU	Aceleração
Energia Total	-7853.604264 eV	-7853.604698 eV	
Tempo da Equação de Poisson	187.1953 s	10.7989 s	17.3346 X

Portar um maior volume de operações, realizadas sobre a mesma matriz de densidades, para a GPU contribui para melhorar o desempenho. Comparando a Tabela 13 com a Tabela 14, podemos observar o aumento de desempenho da Equação de Poisson na GPU, pois a matriz de densidade havia sido usada por outras funções GPU, não necessitando a transferência através do barramento PCI Express.

Tabela 14 – Resultados de Aceleração Reduzindo-se a Transferência de Dados através do Barramento PCI Express, para o fulereno C60.

	CPU	GPU	Aceleração
Energia Total	-7853.604264 eV	-7853.604698 eV	
Tempo de Cálculo do Dipolo Elétrico	14.4850 s	0.0774 s	187.1447 X
Tempo da Equação de Poisson	187.1953 s	8.9205 s	20.9848 X