

1. PENDAHULUAN

1.1. LATAR BELAKANG MASALAH

Pada saat ini, teknologi mengalami kemajuan yang sangat pesat, khususnya pada perkembangan komputer. Adapun salah satu teknologi yang mendukung cepatnya perhitungan dalam komputasi paralel dan juga dalam permasalahan grafik pada komputer. Dalam hal permasalahan tersebut pemrograman yang berbasis GPU (*Graphic Processing Unit*) dapat membantu mengatasinya. GPU adalah sebuah prosesor khusus untuk bagian grafis 3D dari *microprocessor*. GPU modern sangat efisien dalam memanipulasi komputer grafis dan struktur paralel. Pada PC, GPU biasanya terdapat di *video card* atau *motherboard*, lebih dari komputer desktop dan notebook mempunyai GPU yang terintegrasi, yang biasanya jauh berbeda daripada yang ada di *video card* [2].

GPU dipopulerkan oleh NVIDIA. NVIDIA Corporation adalah sebuah perusahaan produsen prosesor grafis, kartu grafis, dan media alat-alat komunikasi untuk komputer, dan konsol permainan Sony Playstation3. Produk paling terkenal dari NVIDIA adalah seri GeForce yang digunakan untuk komputer [5]. NVIDIA juga mengembangkan sebuah teknologi yang bernama CUDA (*Compute Unified Device Architecture*). Arsitektur perangkat lunak ini menggunakan bahasa pemrograman C dan C++. CUDA (*Compute Unified Device Architecture*) adalah sebuah arsitektur perangkat keras dan perangkat lunak untuk mengelola komputasi secara paralel pada perangkat keras GPU. Setiap perangkat GPU berkemampuan CUDA dapat bertindak sebagai sebuah perangkat komputasi data paralel secara massal dengan jumlah memori yang besar [14].

CUDA merupakan pemrograman paralel yang terdiri dari tiga *core* atau inti, yaitu *thread*, *shared memory*, dan sinkronisasi. Pemanfaatan *threads* pada GPU inilah yang dapat mempercepat dalam perhitungan dengan melakukan pembagian tugas perhitungan tersebut dan

menyebarkan ke *thread-thread* lain untuk dilakukan komputasi pada masing-masing *thread* tersebut. *Shared memory* berfungsi untuk mempermudah penggunaan *resource* oleh *thread* yang melakukan komputasi hingga tidak perlu keluar *block* untuk mencari *resource* utama. Dalam hal tersebut, akan mengurangi waktu komputasi dan mempercepat proses pengiriman dan penggunaan *resource*. Sinkronisasi antar *thread* juga mengimplementasi proses komputasi, karena pada perhitungan yang dilakukan oleh *threads* akan dilakukan penggabungan dalam satu perhitungan [1].

Pada CUDA akan ada pendeklarasian fungsi *kernel*. *Kernel* adalah suatu perangkat yang menjadi bagian utama dalam melakukan pengerjaan untuk sebuah proses. Dengan menggunakan CUDA, maka pada tugas akhir ini dilakukan pengujian pada suatu *kernel* dan melakukan pembagian proses kedalam *block* dan *thread* yang ada di dalam GPU.

Maka dari itu, dari segi pemanfaatan kemampuan dari CUDA yang sudah dijelaskan, diharapkan ada performansi waktu dalam perhitungan komputasi paralel dengan memanfaatkan kemampuan GPU. Dalam tugas akhir ini akan dilakukan analisis perkalian matriks secara paralel pada GPU secara eksplisit, yaitu dengan pembagian proses kedalam *block*, *thread*, dan kombinasi antara *block* dan *thread*. Adapun tujuan dari hal tersebut adalah untuk mempercepat melakukan perhitungan pada perkalian matriks dalam jumlah *block* dan *thread* yang akan ditentukan. Pada dasarnya, jumlah *block* dan *thread* tergantung pada kemampuan suatu GPU. Semakin tinggi spesifikasi suatu GPU, maka semakin banyak pula jumlah *block* dan *thread* pada suatu GPU tersebut. Kasus perkalian matriks dapat digunakan untuk berbagai penerapan dalam proses citra digital, penyelesaian program linier dan bidang-bidang lainnya.

1.2. PERUMUSAN MASALAH

Berdasarkan latar belakang sebelumnya, maka terdapat beberapa rumusan permasalahan yang dapat diangkat dalam tugas akhir ini antara lain :

1. Bagaimana melakukan perkalian matriks dengan menggunakan GPU CUDA?
2. Bagaimana menjalankan fungsi *kernel* GPU CUDA secara eksplisit?
3. Bagaimana performansi waktu implementasi perkalian matriks dengan menggunakan fungsi jumlah *block* dan *thread* yang berbeda-beda?

1.3. BATASAN MASALAH

Batasan masalah yang digunakan dalam penelitian tugas akhir ini adalah sebagai berikut :

1. Kasus yang dilakukan dengan menggunakan perkalian matriks persegi ($n \times n$).
2. Bilangan pada matriks merupakan bilangan *floating point* dan *random*.
3. Perangkat komputer yang digunakan pada penelitian tugas akhir ini adalah sebagai berikut:
 - GPU : *NVIDIA GeForce GTX 670*
 - CPU : *Intel core i5 2500 Quad Core*

1.4. TUJUAN PENELITIAN

Tujuan dari tugas akhir ini adalah sebagai berikut :

1. Melakukan perkalian matriks dengan menggunakan GPU CUDA,
2. Menjalankan fungsi *kernel* dalam kasus perkalian matriks pada GPU CUDA secara eksplisit,
3. Mengetahui performansi implementasi waktu pada perkalian matriks dengan menggunakan jumlah *block* dan *thread* yang berbeda-beda.

1.5. METODOLOGI PENELITIAN

Metodologi penelitian yang digunakan dalam penyelesaian tugas akhir ini adalah sebagai berikut :

1. Studi Literatur

Melakukan pencarian referensi teori yang relevan atau sesuai dengan permasalahan yang ditemukan. Referensi tersebut berisikan tentang :

- ❖ GPU (*Graphic Processing Unit*)
- ❖ CUDA (*Compute Unified Device Architecture*)
- ❖ *Kernel*
- ❖ *Block dan Thread*
- ❖ ANOVA *ONE & TWO WAY*

Referensi ini dicari dari buku, kumpulan artikel laporan penelitian, jurnal, dan situs-situs di internet.

2. Perancangan Algoritma

Merancang algoritma pada kasus perkalian matriks, dan diikut sertakan untuk pemanggilan fungsi *kernel* untuk proses pembagian pada *block* dan *thread*.

3. Implementasi Perancangan Algoritma

Mengimplementasi rancangan algoritma kasus perkalian matriks tersebut pada GPU NVIDIA CUDA.

4. Analisis Hasil Implementasi Sistem

Menganalisis hasil implementasi dari kasus yang sudah dilakukan pada perancangan algoritma yang berupa performansi waktu dalam proses running.

5. Pengambilan kesimpulan dan penyusunan laporan

Membuat kesimpulan dari hasil perancangan dan analisis yang telah dilakukan, yang kemudian disusun dalam bentuk laporan.

1.6. SISTEMATIKA PENULISAN

Sistematika penulisan laporan tugas akhir ini disusun sesuai dengan rencana sebagai berikut:

BAB 1 PENDAHULUAN

Pada bab pertama ini akan dijelaskan latar belakang masalah, rumusan masalah, tujuan penelitian, metodologi penelitian, dan sistematika penulisan.

BAB 2 DASAR TEORI

Pada bab kedua ini akan dijelaskan uraian dasar teori dari pengumpulan teori yang berkaitan tentang GPU, CUDA, matriks, *kernel*, *block* dan *thread*, dan teori lain yang mendukung dengan penyelesaian tugas akhir.

BAB 3 PERANCANGAN SISTEM

Pada bab ketiga ini akan dijelaskan mengenai proses analisis perancangan dan implementasi mengenai parameter dan performansi yang akan dibangun.

BAB 4 ANALISIS HASIL PENGUJIAN

Pada bab keempat ini akan dijelaskan hasil dari analisis perancangan dan implementasi pada metode yang digunakan dalam sistem yang telah dibangun pada bab sebelumnya.

BAB 5 PENUTUP

Pada bab terakhir ini akan dijelaskan kesimpulan dari hasil uji analisa yang telah dilakukan, beserta saran-saran untuk pengembangan tugas akhir selanjutnya.

1.7. RENCANA KEGIATAN

Rencana kegiatan yang akan dilakukan dalam penyelesaian tugas akhir ini dijadwalkan dengan representasi tabel berikut :

No.	Kegiatan	Bulan ke-1	Bulan ke-2	Bulan ke-3	Bulan ke-4	Bulan ke-5	Bulan ke-6
1	Studi literatur	■	■	■	■	■	■
2	Pengumpulan data			■	■	■	■
3	Perancangan Algoritma	■	■	■			
4	Implementasi Algoritma			■	■	■	■
5	Analisis hasil implementasi			■	■	■	■
6	Pembuatan laporan	■	■	■	■	■	■

Tabel 1.1. Tabel Rencana Kegiatan