NGHIÊN CỨU GiẢI PHÁP CÔNG NGHỆ TÍNH TOÁN HIỆU NĂNG CAO VỚI BỘ XỬ LÝ ĐỒ HỌA GPU VÀ ỨNG DỤNG.

Nguyễn Đức Minh

Trường Đại học Công nghệ - Đại học Quốc gia Hà Nội

Luận văn Thạc sĩ chuyên ngành: Kỹ thuật phần mềm;
Mã số: 60 480 10 3
Người hướng dẫn: TS. Lê Quang Minh
Năm bảo vệ: 2016

Abstract: Trình bày cơ sở lý thuyết về tính toán hiệu năng cao và tính toán song song,trên cơ sở lý thuyết về tính toán song song và kiến thức về GPU và CUDA đưa ra các bài toán ứng dụng trên GPU để làm rõ hiệu nằng tính toán về mặt thời gian so với tính toán đơn thuần trên CPU.
Keywords: Tính toán song song, GPU, CUDA

MỞ ĐẦU

Sự bùng nổ của Internet, sự bùng nổ của xu thế mọi thiết bị đều kết nối (Internet of thing - IOT), sự bùng nổ về nhu cầu thưởng các sản phẩm âm thanh đồ họa độ phân giải cao và chất lượng cao, sự bùng nổ của các dịch vụ lưu trữ đám mây, dịch vụ trực tuyến, đã khiến cho khôi lượng dữ liệu mà vi xử lý (CPU) phải tính toán ngày càng lớn và thực sự đã vượt quá nhanh so với sự phát triển tốc độ của CPU. Không những thế con người mặc dù muốn có nhiều thông tin hơn, thông tin phải tốt hơn lại còn muốn tốc độ xử lý phải nhanh hơn, điều này càng làm cho nhu cầu tính toán trong lĩnh vực khoa học, công nghệ đã và đang trở thành một thách thức lớn. Từ đó các giải pháp nhằm tăng tốc độ tính toán đã được ra đời.

Tư năm 2001 đến 2003 tốc độ của Pentium 4 đã tăng gấp đôi từ 1.5GHz lên đến 3GHz. Tuy nhiên hiệu năng của CPU không tăng tương xứng như mức gia tăng xung nhịp của CPU và việc tăng xung nhịp cũng chỉ đạt tới giới hạn công nghệ. Cụ thể tính đến 2005 xung nhịp của Pentium 4 mới chỉ tăng lên được 3.8GHz. Việc tăng xung nhịp của CPU dẫn đến việc tăng nhiệt độ làm việc của CPU. Các công nghệ làm mát có thể không đáp ứng được do bề mặt tiếp xúc của CPU ngày càng nhỏ. Trước tình hình này các nhà nghiên cứu vi xử lý đã chuyển sang hướng phát triển công nghệ đa lõi nhằm song song hóa các quá trình tính toán để tăng tốc độ và tiết kiệm điện năng. Một trong các công nghệ đa lõi xử lý song song ra đời là bộ xử lý đồ họa GPU (Graphic Processing Unit). GPU ban đầu ra đời chỉ phục vụ cho mục đích xử lý đồ họa, và ngành công nghiệp Game. Tuy nhiên ngày nay với công nghệ CUDA được phát triển bởi hãng NVIDIA từ năm 2007 đã cho phép thực hiện các tính toán song song với các phép tính phức tạp như dấu chấm động. Với hiệu xuất cả ngàn lệnh trong một thời điểm. Chính vì vậy một xu hướng nghiên cứu mới đã ra đời đó là phát triển các thuật toán song song thực hiện trên GPU. Với CUDA các lập trình viên có thể nhanh tróng phát triển các ứng dụng tính toán song song cho rất nhiều ứng dụng khác nhau như: Điện toán, sắp xếp, tìm kiếm, xử lý tín hiệu số, ảnh,…

Việc nghiêp cứu áp dụng CUDA để tăng tốc độ tính toán cho các bài toán mà cần phải xử lý một khối dữ liệu đầu vào khổng lồ hoặc các bài toán yêu cầu tính thời gian thực đã thực sự trở thành một vấn đề cấp thiết trong thực tế. Xuất phát từ như

cầu này mà tôi đã chọn đề tài **:** NGHIÊN CỨU GiẢI PHÁP CÔNG NGHỆ TÍNH TOÁN HIỆU NĂNG CAO VỚI BỘ XỬ LÝ ĐỒ HỌA GPU VÀ ỨNG DỤNG.

Luận văn gồm 3 chương chính:

Chương 1: Tổng quan về tính toán song song và GPU, chương này giới thiệu những kiến thức tổng quan về tính toán song song, từ đó tìm hiểu những kiến thức cơ bản về bộ xử lý đồ họa GPU và cách thức ứng dụng tính toán trên đó.

Chương 2: Tính toán song song trên GPU trong CUDA,. Chương này cung cấp các kiến thức về môi trường lập trình, ngôn ngữ lập trình, cách thiết lập chương trình và các chỉ dẫn hiệu năng khi cài đặt ứng dụng tính toán trên GPU.

Chương 3:Tăng tốc độ tính toán một số bài toán sử dụng GPU. Trên cơ cở các kiến thức được trình bày ở các chương trên, tác giả luận văn đã tiến hành cài đặt và thử nghiệm mô phỏng bài toán trên CPU và GPU. Từ đó có những so sánh, nhận xét về năng lực tính toán vượt trội của GPU so với CPU truyền thống. Đồng thời cũng mở ra các hướng cải tiến hiệu năng mới cho bài toán chạy trên GPU.

DANH MỤC THUẬT NGỮ

	Tiếng Anh	**Tiếng Việt**
1	API	Application Program Interface: một API định nghĩa một giao diện chuẩn để triệu gọi một tập các chức năng.
2	coproccessor	bộ đồng xử lý
3	gpgpu	tính toán thông dụng trên GPU
4	GPU	Bộ xử lý đồ họa
5	kernel	hạt nhân
6	MIMD	Multiple Instruction Multiple Data: đa lệnh đa dữ liệu
7	primary surface	Bề mặt chính, khái niệm dùng trong kết cấu
8	proccessor	Bộ xử lý
9	Rasterization	Sự quét mành trên màn hình
10	SIMD	Single Instruction Multiple Data: đơn lệnh đa dữ liệu
11	stream	Dòng
12	streaming processor	Bộ xử lý dòng
13	texture	Kết cấu: cấu trúc của đối tượng, nó được xem như mô hình thu nhỏ của đối tượng.
14	texture fetches	Hàm đọc kết cấu
15	texture reference	Tham chiếu kết cấu
16	warp	Mỗi khối được tách thành các nhóm SIMD của các luồng.

DANH MỤC HÌNH VẼ

Content

TÓM TẮT LUẬN VĂN THẠC SỸ

Luận văn gồm 3 chương chính:

Chương 1: Tổng quan về tính toán song song và GPU, chương này giới thiệu những kiến thức tổng quan về tính toán song song, từ đó tìm hiểu những kiến thức cơ bản về bộ xử lý đồ họa GPU và cách thức ứng dụng tính toán trên đó.

Bao gồm:

- Lịch sử ra đời lý do mục đích của tính toán song song.
- Các mô hình tính toán song song.
- Các mô hình lập trình song song.
- Nguyên lý thiết kế giải thuật song song.
- Nhận thức về những bài toán chương trình có thể song song hóa được.
- Tổng quan về GPU.

Chương 2: Tính toán song song trên GPU trong CUDA,. Chương này cung cấp các kiến thức về môi trường lập trình, ngôn ngữ lập trình, cách thiết lập chương trình và các chỉ dẫn hiệu năng khi cài đặt ứng dụng tính toán trên GPU.

Bao gồm:

- Tổng quan về CUDA (lịch sử ra đời, cấu tạo,thành phần….).
- Môi trường lập trình và cơ chế hoạt động của 1 chương trình trong CUDA
- Cách thức lập trình ứng dụng với CUDA và các ví dụ

Chương 3: Tăng tốc độ tính toán một số bài toán sử dụng GPU. Trên cơ cở các kiến thức được trình bày ở các chương trên, tác giả luận văn đã tiến hành cài đặt và thử nghiệm mô phỏng bài toán trên CPU và GPU. Từ đó có những so sánh, nhận xét về năng lực tính toán vượt trội của GPU so với CPU truyền thống. Đồng thời cũng mở ra các hướng cải tiến hiệu năng mới cho bài toán chạy trên GPU.

Bao gồm :

- Giới thiệu các bài toán

- Chương trình demo
- Kết quả chạy thử nghiệm
- So sánh kết quả về mặt thời gian xử lý .

KẾT LUẬN

Luận văn đã nghiên cứu tổng quan về tính toán song song, đó là điều kiện cần để phát triển ứng dụng GPU cho mục đích thông dụng. Tác giả luận văn cũng đã tìm hiểu về cơ chế hoạt động của GPU, các kiến trúc bên trong nó, mô hình lập trình trên GPU. Trong chương 2, luận văn đã tìm hiểu công cụ lập trình GPU phổ biến nhất hiện nay là CUDA. Tác giả luận văn cũng trình bày chi tiết các mô hình lập trình, thiết lập phần cứng trên card đồ họa của Nvidia, giao diện lập trình cũng như các chỉ dẫn hiệu năng khi chạy ứng dụng trên card đồ họa.

Từ các hiệu biết trên, tác giả đã thực hiện thử nghiệm năng lực tính toán của GPU so sánh với CPU để kiểm chứng những điều mà lý thuyết đã nói. Các kết quả thử nghiệm được trình bày chi tiết trong chương 3 của luận văn.

Với các kết quả đạt được, tác giả mong muốn có các nghiên cứu thêm về cải tiến hiệu năng bài toán mô phỏng tiếp tục nghiên cứu phát triển cài đặt các thuật toán, các phương pháp xử lý tín hiếu số, ảnh áp dụng mạng Nơron trên nền tảng GPU Mong rằng các kết quả nghiên cứu trong tương lai của luận văn sẽ đạt được điều đó.

TÀI LIỆU THAM KHẢO

Tài liệu tiếng việt

[1] Trương Văn Hiệu (2011), "*Nghiên cứu các giải thuật song song trên hệ thống xử lý đồ họa GPU đa lõi*", luận văn thạc sĩ, trường Đại học Đà Nẵng.

[2] Nguyễn Việt Đức – Nguyễn Nam Giang (2012), "*Xây dựng thuật toán song song tìm đường đi ngắn nhất với CUDA*", luận văn thạc sỹ, trường Đại học Công nghệ Hồ Chí Minh.

[3] Nguyễn Thị Thùy Linh (2009), "*Tính toán hiệu năng cao với bộ xử lý đồ họa GPU và ứng dụng*", luận văn thạc sĩ, trường Đại học Công nghệ Hà Nội.

Tài liệu tiếng anh

[4] Jason Sanders, Edward Kandrot, "*CUDA by example*", an introduction to General- Purpose GPU programming.

[5] Maciej Matyka, "*GPGPU programming on example of CUDA*", Institute of Theoretical Physics University of Wroclaw.

[6] NVIDIA, "*High performance computing with CUDA*", Users Group Conference San Diego, CA June 15, 2009.