BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC BÁCH KHOA HÀ NỘI

HỒ VĂN PHI

TỐI ƯU HÓA VÀ ĐÁNH GIÁ HIỆU NĂNG CỦA TỔ CHỨC CACHE TRONG HỆ THỐNG VI XỬ LÝ THẾ HỆ SAU

LUẬN ÁN TIẾN SĨ KỸ THUẬT VIỄN THÔNG

Hà Nội - 2014

BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC BÁCH KHOA HÀ NỘI

HỒ VĂN PHI

TỐI ƯU HÓA VÀ ĐÁNH GIÁ HIỆU NĂNG CỦA TỔ CHỨC CACHE TRONG HỆ THỐNG VI XỬ LÝ THẾ HỆ SAU

Chuyên ngành: Kỹ thuật Viễn thông
Mã số: 62520208

LUẬN ÁN TIẾN SĨ KỸ THUẬT VIỄN THÔNG

NGƯỜI HƯỚNG DẪN KHOA HỌC:

1. TS. HỒ KHÁNH LÂM

2. TS. NGUYỄN VIẾT NGUYÊN

Hà Nội - 2014

LỜI CAM ĐOAN

Tôi xin cam đoan đây là công trình của riêng tôi. Tất cả các ấn phẩm được công bố chung với các cán bộ hướng dẫn khoa học và các đồng nghiệp đã được sự đồng ý của các tác giả trước khi đưa vào Luận án. Các kết quả trong Luận án là trung thực, chưa từng được công bố trong bất kỳ công trình nào.

Tác giả Luận án

Hồ Văn Phi

LỜI CẢM ƠN

Đầu tiên, tôi xin bày tỏ lời cảm ơn chân thành và sự kính trọng đến các Thầy giáo TS. Hồ Khánh Lâm và TS. Nguyễn Viết Nguyên, các Thầy đã nhận tôi làm nghiên cứu sinh và hướng dẫn tôi rất nhiệt tình trong suốt thời gian học tập, nghiên cứu và thực hiện bản Luận án này. Các Thầy đã tận tình chỉ bảo và giúp đỡ tôi cả về lĩnh vực khoa học cũng như trong cuộc sống. Tôi vô cùng biết ơn sự kiên trì của các Thầy, các Thầy đã dành nhiều thời gian để đọc cẩn thận và góp nhiều ý kiến quý báu cho bản thảo của Luận án. Những kiến thức mà tôi nhận được từ các Thầy không chỉ là bản Luận án mà trên hết là cách nhìn nhận, đánh giá cũng như phương thức giải quyết vấn đề một cách toàn diện và khoa học.

Tôi xin trân trọng cảm ơn Lãnh đạo Trường Đại Học Bách khoa Hà Nội, Viện Sau Đại học, Viện Điện tử - Viễn thông và Bộ môn Điện tử & Kỹ thuật máy tính đã tạo điều kiện thuận lợi cho tôi được học tập và làm nghiên cứu sinh, luôn quan tâm động viên tôi trong suốt quá trình học tập và nghiên cứu.

Tôi xin chân thành cảm ơn sự giúp đỡ tận tình của các GS, PGS, TS, các Thầy, Cô giáo trong Bộ môn Điện Tử & Kỹ thuật máy tính, Viện Điện tử - Viễn thông, các Nhà khoa học trong và ngoài Trường Đại học Bách khoa Hà Nội.

Tôi xin trân trọng cảm ơn Lãnh đạo Trường Đại học Quy Nhơn và Khoa Kỹ thuật & Công nghệ - Trường Đại học Quy Nhơn, cũng như bạn bè đồng nghiệp đã ủng hộ và tạo mọi điều kiện thuận lợi giúp đỡ tôi trong suốt thời gian học tập, nghiên cứu và hoàn thành Luận án.

Cuối cùng, tôi muốn dành lời cảm ơn đến những người thân yêu nhất của tôi. Bản Luận án này là món quà quý giá tôi xin được kính tặng cho cha mẹ, vợ và các con thân yêu của tôi.

Hà Nội, tháng 06 năm 2014

Tác giả Luận án

Hồ Văn Phi

MỤC LỤC

DANH MỤC CÁC KÝ HIỆU VÀ CHỮ VIẾT TẮT

DANH MỤC CÁC BẢNG

DANH MỤC CÁC HÌNH, ẢNH VÀ ĐỒ THỊ

MỞ ĐẦU..... 1

1. Tính cấp thiết của luận án 1

2. Mục đích nghiên cứu của luận án 4

3. Đối tượng và phạm vi nghiên cứu của luận án 4

4. Phương pháp nghiên cứu của luận án 4

5. Ý nghĩa khoa học và thực tiễn của luận án 5

6. Cấu trúc của luận án 5

Chương 1. TỔNG QUAN VỀ KIẾN TRÚC CHIP ĐA XỬ LÝ, ĐA LUỒNG 6

1.1. Giới thiệu 6

1.2. Kiến trúc của chip đa xử lý, đa luồng 6

1.2.1. Kiến trúc chung của chip đa xử lý, đa luồng 6

1.2.2. Kiến trúc chip đa xử lý, đa luồng đồng thời 11

1.2.3. Mạng liên kết trên chip 12

1.2.4. Phân cấp hệ thống nhớ 16

1.3. Kết luận chương 1 18

Chương 2. NGHIÊN CỨU TỔ CHỨC CACHE, CHÍNH SÁCH THAY THẾ CACHE TRONG KIẾN TRÚC CHIP ĐA XỬ LÝ, ĐA LUỒNG 19

2.1. Tổ chức cache trong kiến trúc chip đa xử lý, đa luồng 19

2.1.1. Cache và các nguyên tắc làm việc của cache 19

2.1.1.1. Vị trí tạm thời 20

2.1.1.2. Vị trí không gian 20

2.1.1.3. Vị trí tuần tự 20
2.1.2. Các thành phần của cache 20
2.1.3. Các tổ chức cache 21
2.1.3.1. Cache liên kết đầy đủ 22
2.1.3.2. Cache sắp xếp trực tiếp 24
2.1.3.3. Cache liên kết tập hợp 26
2.2. Các đặc tính hiệu năng của cache 28
2.2.1. Các tỷ số trúng cache và trượt cache 29
2.2.1.1. Trúng cache 29
2.2.1.2. Trượt cache 29
2.2.1.3. Tỷ số trúng cache, trượt cache và trượt penalty 29
2.2.1.4. Bus bộ nhớ, kích thước từ nhớ, kích thước khối và trượt penalty 31
2.2.1.5. Trượt cache cục bộ và toàn cục 31
2.2.1.7. Ảnh hưởng của tổ chức cache đến trượt penalty 33
2.2.1.8. Kích thước khối cache và tỷ số trượt 34
2.2.1.9. Các loại trượt cache 35
2.2.1.10. Tổ chức cache ảnh hưởng đến tốc độ của CPU 36
2.2.2. Các giải pháp tăng hiệu năng của cache 38
2.3. Các chính sách thay thế dòng cache 38
2.3.1. Chính sách thay thế cache LRU 39
2.3.2. Chính sách thay thế cache LFU 39
2.3.3. Chính sách thay thế cache FIFO 39
2.3.4. Chính sách thay thế cache Random 39
2.3.5. Chính sách thay thế cache NRU 40
2.3.6. Chính sách thay thế cache SRRIP 40
2.3.7. Chính sách thay thế cache DRRIP 42
2.4. Ghi và đọc cache 42

2.4.1. Ghi cache.....42
2.4.1.1. Ghi thông qua.....43
2.4.1.2. Ghi trở lại.....44
2.4.2. Đọc cache.....46
2.4.2.1. Đọc bên cạnh.....46
2.5.2.2. Đọc thông suốt.....47
2.5. Cache chia sẻ thông minh 48
2.5.1. Tổ chức phân cấp cache trong các chip đa xử lý 48
2.5.2. Cache chia sẻ thông minh.....49
2.6. Tính nhất quán cache trong các chip đa xử lý, đa luồng 50
2.6.1. Thế nào là nhất quán cache.....50
2.6.2. Các giao thức nhất quán cache.....52
2.7. Kết luận chương 2 52
Chương 3. PHÂN TÍCH ĐÁNH GIÁ HIỆU NĂNG CỦA TỔ CHỨC CACHE TRONG KIẾN TRÚC CHIP ĐA XỬ LÝ, ĐA LUỒNG..... 53
3.1. Cơ sở lý thuyết để phân tích đánh giá hiệu năng của tổ chức cache..... 53
3.1.1. Kiến trúc chip đa xử lý, đa luồng là mạng xếp hàng đóng đa lớp có dạng tích các xác suất (MCPFQN).....53
3.1.1.1. Khái quát mạng xếp hàng đóng.....53
3.1.1.2. Khái quát mạng xếp hàng đóng có dạng tích các xác suất 56
3.1.1.3. Kiến trúc chip đa xử lý, đa luồng là mạng xếp hàng đóng đa lớp có dạng tích các xác suất (MCPFQN).....58
3.1.2. Thuật toán phân tích giá trị trung bình (MVA) đánh giá hiệu năng cho các mạng xếp hàng đóng có dạng tích các xác suất.....59
3.1.2.1. Mạng xếp hàng đóng đơn lớp có dạng tích các xác suất 59
3.1.2.2. Mạng xếp hàng đóng đa lớp có dạng tích các xác suất 61
3.2. Mô hình tổ chức cache trong kiến trúc chip đa xử lý, đa luồng..... 62

3.2.1. Khái quát.....62

3.2.2. Mô hình tổ chức cache trong kiến trúc chip đa xử lý, đa luồng.....63

3.3. Phân tích đánh giá hiệu năng của tổ chức cache trong kiến trúc chip đa xử lý, đa luồng 64

3.3.1. Mô hình thực hiện phân tích hiệu năng của kiến trúc chip đa xử lý, đa luồng.....64

3.3.1.1. Mô hình tổng quát.....64

3.3.1.2. Mô hình rút gọn 66

3.3.2. Kết quả mô phỏng và đánh giá hiệu năng cho kiến trúc CMP đa luồng.....72

3.3.2.1. Kết quả mô phỏng cho các kiến trúc CMP đa luồng.....72

3.3.2.2. Đánh giá hiệu năng các chip đa xử lý, đa luồng 83

3.4. Kết luận chương 3 84

Chương 4. GIẢI PHÁP TỐI ƯU HÓA HIỆU NĂNG CỦA TỔ CHỨC CACHE TRONG KIẾN TRÚC CHIP ĐA XỬ LÝ, ĐA LUỒNG 85

4.1. Kiến trúc cụm lõi cho chip đa xử lý, đa luồng 85

4.1.1. Khái quát.....85

4.1.2. Mô hình MCPFQN cho kiến trúc cụm lõi.....86

4.1.2.1. Mô hình MCPFQN tổng quát của kiến trúc cụm lõi 86

4.1.2.2. Mô hình MCPFQN rút gọn của kiến trúc cụm lõi 88

4.1.3. Kết quả mô phỏng và đánh giá hiệu năng cho kiến trúc cụm lõi.....90

4.1.3.1. Kết quả mô phỏng cho kiến trúc cụm lõi 90

4.1.3.2. Đánh giá hiệu năng cho kiến trúc cụm lõi.....93

4.2. Lựa chọn cấu hình mạng liên kết trên chip 94

4.2.1. Khái quát.....94

4.2.2. Đề xuất công thức tính trễ truyền thông trung bình cho mạng liên kết trên chip 94

4.2.3. Lựa chọn cấu hình mạng liên kết trên chip 99

4.2.3.1. Kết quả mô phỏng 99

4.2.3.2. Đánh giá kết quả 104

4.3. Kết luận chương 4 104

KẾT LUẬN 106

TÀI LIỆU THAM KHẢO 107

DANH MỤC CÁC CÔNG TRÌNH ĐÃ CÔNG BỐ CỦA LUẬN ÁN 115

PHỤ LỤC 116

Phụ lục 1. Kịch bản mô phỏng đánh giá hiệu năng của tổ chức cache trong kiến trúc CMP đa luồng có 2 cấp cache và 3 cấp cache 116

Phụ lục 2. Kịch bản mô phỏng đánh giá hiệu năng của tổ chức cache trong kiến trúc cụm lõi 3 cấp cache 130

DANH MỤC CÁC KÝ HIỆU VÀ CHỮ VIẾT TẮT

Ký hiệu	Tên tiếng Anh	Tên tiếng Việt
A		
ACM	Association for Computing Machinery	Hiệp hội kỹ thuật tính toán Hoa Kỳ
ALU	Arithmetic Logic Unit	Đơn vị số học và logic
AMAT	Average Memory Access Time	Thời gian trung bình truy nhập bộ nhớ
AMC	Asymmetric Multi-core Chip	Chip đa lõi bất đối xứng
ANSI	American National Standards Institute	Viện tiêu chuẩn quốc gia Hoa Kỳ
ASCII	American Standards Code for Information Interchange	Chuẩn mã trao đổi thông tin Hoa Kỳ
ASIC	Application-Specific Integrate Circuit	Mạch tích hợp ứng dụng chuyên biệt
B		
BCE	Base Core Equivalents	Tương đương lõi cơ sở
BIP	Bimodal Insertion Policy	Chính sách chèn hai phương thức
BRRIP	Bimodal Re-Reference Interval Prediction (Bimodal RRIP)	Hai phương thức dự báo khoảng tham chiếu lại
C		
CAD	Computer Aided Design	Thiết kế bằng máy tính
ccNUMA	cache-coherent Non-Uniform Memory Access	Truy cập bộ nhớ không đồng nhất tương quan cache
CD	Compact Disk	Đĩa compact
CISC	Complex Instruction Set Computer	Máy tính có tập lệnh phức