様式 C-19

科学研究費補助金研究成果報告書

平成　22 年　6 月　8 日現在

研究種目：基盤研究(B)
研究期間：2006 ～ 2009
課題番号：18300028
研究課題名（和文）　メモリ階層を考慮した高速検索アルゴリズムと
そのハードウェア化の研究
研究課題名（英文）　Hardware approach for high-speed search algorithms
with memory layer optimization

研究代表者　稲葉　真理（INABA MARY）
東京大学・大学院情報理工学系研究科・准教授
研究者番号：60282711

研究成果の概要（和文）：
プロセッサに比してメモリの速度向上が遅い結果,キャッシュミスのペナルティは相対的に増大しつつある．現実の計算においてもメモリバンド幅が計算ボトルネックになることも多く，メモリ階層を意識し，上位階層のメモリの効率的利用を行なうことで，メモリ遅延をできるかぎり隠蔽することが重要である．本研究では，メモリ階層構造を考慮しハードウェアを利用する圧縮簡潔データ構造を利用した高速検索方式の提案および実装実験を行った．また本研究の副産物としてキャッシュミスペナルティーを削減するためのメモリマップを利用したプリフェッチ方式の提案も行なった．

研究成果の概要（英文）：
With the progress of semiconductor process technology, memory access gap, and thus, the pipeline stall time due to cache misses are increasing. It is often observed in the real world computation that the memory band width plays the role of computation bottleneck. Thus, considering memory hierarchy, it is important to hide the memory access latency with efficient utilization of upper layer memory. In this work, we propose and evaluate a hardware solution to accelerate for full text search with compressed succinct data structure. In addition, as a by-product, we propose an efficient prefetch method using memory map data structure.

交付決定額

（金額単位：円）

	直接経費	間接経費	合　計
2006 年度	4,200,000	1,260,000	5,460,000
2007 年度	3,800,000	1,140,000	4,940,000
2008 年度	3,300,000	990,000	4,290,000
2009 年度	3,100,000	930,000	4,030,000
総　計	14,400,000	4,320,000	18,720,000

研究分野：総合領域
科研費の分科・細目：(分科)情報学,(細目)計算機システム・ネットワーク
キーワード：超高速情報処理，ディレクトリ・情報検索，アルゴリズム，
インターネット高速化，コンテンツアーカイブ

様式 C-19

科学研究費補助金研究成果報告書

１．研究開始当初の背景

近年の急速なネットワーク速度とメモリやディスクの記憶容量の増大により，膨大な量のデータから必要な情報を効率良く短時間で抽出する仕組みは，情報科学的な研究基盤のみならず，情報社会における社会基盤としてなくてはならないものとなってきている．本申請の計画が完了する平成 22 年には，10Gbps のネットワークが日常となり，幹線と海外回線を作る科学技術用インターネットバックボーンは 100Gbps 以上となり，また科学技術計算として2ペタフロップスを超えることを予測していた．このような情報基盤上に蓄積された膨大な量のデータを如何に有効に活用できるかは，今後の我が国の科学技術研究ポテンシャルにとり大きな重要性を持っている．本研究がターゲットとする検索問題は情報抽出のための基礎技術であり，内外で精力的に研究が行われてきており，メモリアクセスにかかる時間を一定と仮定するモデルの元では理論的にほぼ下限値であるアルゴリズムが開発されていた。しかしながらインターネット上の Web データや有機化合物反応データベースといった超大規模データからの任意のオブジェクトの高速検索は困難であり，さらなる性能向上が望まれていた．大規模検索システムの本質的な難しさは，記憶装置の階層構造の取り扱いにあると言っても過言ではない．メモリと二次記憶のアクセス速度を区別する計算モデル上でのアルゴリズムは従来から存在するが，現代のシステムにはオンチップキャッシュ・キャッシュ・メモリ・キャッシュディスク・ハードディスク・ネットワークデータといった，それぞれレイテンシーが一桁違うメモリ階層が存在しており、これらを考慮した高速化が必要となってきていた．

２．研究の目的

本研究課題「メモリ階層を考慮した高速検索アルゴリズムとそのハードウェア化の研究」は，大容量ストレージやネットワーク上に蓄積された膨大なデータを科学技術研究の現場で活用するための礎となる実用的高速検索システムを構築することを目標としている．本提案の特徴は，実用性のために各記憶階層を意識した(1)圧縮情報の直接検索による高速化および(2)ハードウェア・並列分散化による高速化を行うところに特徴がある．

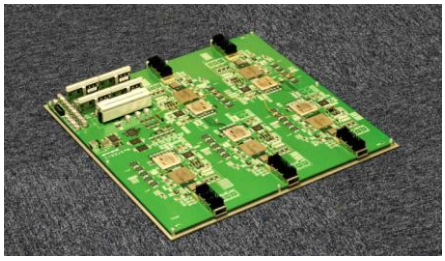
メモリ階層を考慮した圧縮情報の直接検索による高速化：一般に検索システムは前処理として与えられたデータを高速に検索するためのデータ構造（以下索引部と記述する）を作成し，検索時には索引部およびデータ部それぞれに対しデータアクセスが発生する．メモリ階層の上位階層ほど記憶容量が少なくなるため検索時にアクセスする索引部およびデータ部の容量あたりの情報量が大きければ大きいほど，(a)より上位階層に保持できる情報量が増え，(b)単位時間あたりに下位階層から上位階層に転送できる情報量が増えるため，検索をより高速に行うことができる．我々はこれを「見かけのメモリ増加」と呼ぶ．索引部あるいはデータ部を圧縮し圧縮したデータを直接検索するためのオーバーヘッドが圧縮による見かけのメモリ増加の効果よりもはるかに小さければ，記憶領域の節約と高速検索を同時に達成することができる．分担者定兼はメモリサイズを意識した圧縮接尾辞配列(Compressed Suffix Array, CSA)を提案しその有効性を理論的にそして実験的に明らかにしてきた．しかしながらこの理論的解析はメモリのアクセス速度が一定という仮定の下でのみ成り立つため，現実の計算機システムでの超巨大データ処理における性能は保証されない．本研究ではメモリ階層を意識し，CSA を発展させたデータ構造の研究を行う．

メモリ階層を考慮したハードウェア・並列分散化による高速化： 稲葉・菅原らは FPGA を利用しホストマシンの計算を加速するためのプログラマブルボードを開発している．一方，検索時のメモリアクセスを考えるとキャッシュライン単位で検索を高速に行なうことは重要な鍵となっている．本研究では，まず，このプログラマブルボードを利用し，キャッシュライン内の一括高速検索を実現する．検索を行なう回路はキャッシュラインに依存しないため，(a)ラインごとの検索がハードウェア化により高速に行え，(b)キャッシュライン内索引部が不要になることで索引部サイズも軽減され，(c)索引部データへのアクセス効率の改善が期待できる．また，このボード搭載方式はホスト CPU に負荷をかけないという特長を持つため，１台のホストに複数のボードを搭載し超高速検索エンジンを作成，ホストのメイン CPU

は，検索結果に対する処理，たとえば統計処理やデータマイニングなど，より複雑な処理を行うことを目的とする応用システムを開発する．

３．研究の方法

本研究は，上記目的の達成のため，(1) 階層構造を利用したメモリアクセスレイテンシー隠蔽のためのジェネラルな方式の提案と，(2) プログラマブルハードウェアである FPGA(Field Programmable Gate Array)を利用した検索機構の並列化，高速化という２通りのアプローチを採り課題を遂行した．
具体的には，(1)メモリの階層構造を利用する事でメモリアクセスレイテンシーの隠蔽を行なうために，(1.1) ループ最適化のため，整数計画法を利用したメモリ階層間通信の最適化を行なうためのコンパイラのスケジューリング方式の提案，およびその拡張として，従来研究では独立に行なわれていたパイプラインを意識したブロックスケジューリングと階層間メモリ通信を同時に最適化するためスケジューリング方式の提案および実装というソフトウェア方式，および，(1.2)近い将来に利用するであろうデータをあらかじめ予想しておき，そのデータを上位階層メモリに配置するプリフェッチ方式，および，近い未来に利用されることはないと予測されるデータはキャッシュに載せないようなキャッシュリプレースメントアルゴリズムの提案と，シミュレーションを行なうハードウェア方式の研究を行なった．



また(2) プログラマブルハードウェア FPGA を利用した圧縮簡潔データ構造の並列化，高速化については，(2.1) 圧縮簡潔データ構造からの情報抽出のためのハードウェアアルゴリズムと，FPGA への実装の研究を行ない，(2.2) FPGA 5 個搭載しそれぞれを完全グラフによって結合する FPGA ボードの作成を行なった．

４．研究成果

(1.1) メモリ間階層構造を利用した最適化の研究に関しては，まず，平成 18 年度から 19 年度にかけて，明示的なコピー操作を行うことができるスクラッチパッドメモリを階層的にもつシステムをターゲットとした最適化手法を提案，「メモリ階層構造を意識した最適化コンパイラ，MCAMP」を実装し，ベンチマークテストを行ってその有効性を検証した．具体的には，コピーキャンディディトをグラフのノードとして可能なコピーをエッジとした階層的コピーキャンディデーとグラフを作成，各エッジにコピーの際のコストを割り付け，整数計画法を用いて費用最小となる最適化パス抽出することで最適化を行った[文献 20]．この成果を拡張するため，平成 21 年度は，プロセッサおよびメモリ階層の定式化および最適化の研究に引き続き取り組み，メモリ階層に加え，パイプラインを意識し，プログラムコードをもとにして行うデータ転送コスト静的最適化の研究に特に力を入れ，その成果をネットワーク層にも応用した．また基礎的な離散最適化の研究を行なう[文献 6]．

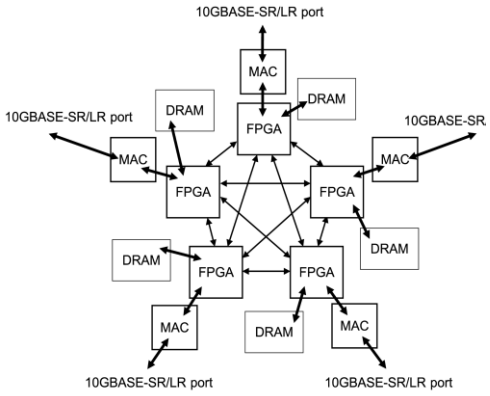
さらに，大容量ストレージやネットワーク上に蓄積された膨大なデータを活用するため，遠隔地分散巨大ストレージ環境を構築し，分散環境で大容量データを取り扱うシステムを構成した．具体的には，ソリッドステートドライブ (SSD) 8 台構成の RAID システムを複数構成し，Linux の sendfile 機能を用いて，データ送出を行い，性能の評価を行った．sendfile に対応する受け取り側の recvfile にあたるものは，Linux では提供されていないため，recvfile に相当する受け取り機能の実装を行い単体および複合性能評価を行い，日仏間および日米間の実際のネットワーク回線を用いて実際に大規模なファイルデータ転送実験を行った[文献 2,7]．特に 2009 年 11 月に行った日米間の転送実験においては，ユーザインターフェースとして，ウェブブラウザを利用することでユーザビリティの向上につとめた．具体的には，一般に広く使われている Firefox ブラウザのプラグインを開発，チューニングした Apache と組み合わせることにより，一般ユーザがネットワークを意識することなく利用可能なアプリケーション UsadaFox を，コモディティ PC の上に実現した．米国ポートランドで行われた SC09 においてこの UsadaFox によるデータ転送方式が評価され，

Supercomputing 2009, SC09 Bandwidth Challenge で, Impact Award を受賞した[文献 5].

(1.2) 近い将来に使われるデータおよび使われることがないデータを予測するために, メモリアクセスマップデータ構造を提案した. このメモリアクセスマップデータ構造を利用することにより, まず, 平成 20 年度から 21 年度にかけて, ハードウェアコストをさして使わずに, 効率よく近未来に使われるデータを予測しあらかじめデータを上位メモリ階層に配置するプリフェチ方式を提案した[文献 10]. またこの方式を, The Journal of Instruction-Level Parallelism が主催する Data Prefetching Championship に出場, First JILP Data Prefetching Championship (DPC-1)で champion となった. 平成 22 年度は, このデータ構造を, 近い未来に再利用されないデータの予測に利用, 再利用されないデータをキャッシュに残さないことで, キャッシュ等, 上位階層メモリの効率的な利用をはかるためのキャッシュリプレースメントアルゴリズムを提案した. [文献 1].

(2.1) 一方, ハードウェア作成グループはメモリ階層を意識し, プログラマブルハードウェアを利用した高速化の研究を行った. (2.1) 圧縮簡潔データ構造からの情報抽出に関しては, FPGA は, ビットマニュピレーションが高速に行なえることを利用し, 圧縮簡潔データ構造からの直接情報抽出の FPGA ファームウェアへの実装を行い, 検索専用アクセラレータの開発を行なった. 平成 18 から 19 年度にかけて, アルゴリズムグループが行った圧縮データ構造の研究および理論的解析に基づき, ハードウェアグループがプロトタイプシステム上で FPGA 実装を行なった. 圧縮データ構造を使った検索システムについては FPGA 上で, ビットマニピュレーションを高速に行うことで探索が高速化されている. 開発は当初の予定より早く完了し, 実在のゲノムデータを中心とした実証実験を行った. 実際の DNA の塩基配列のデータを用いた実験で, 6 倍程度の高速化が達成された. 当初目標にしていた ソフトウェアに比べて 10 倍以上という値に比して, 想定していたレベルの性能がでていないことがわかり, 性能低下要因解析のためサイクルレベルシミュレーションおよび様々なチューニングを行った. その結果, ホストメモリとアクセラレータメモリのバンド幅ボトルネックが性能に大きな影響を与えることがわかり, 大規模化した際に要求スペックで正常稼働させるための問題点が顕在化した. ホストメモリとアクセラレータメモリのバンド幅ボトルネックが性能に大きな影響を与えることがわかり, 大規模化した際に要求スペックで正常稼働させるための問題点が顕在化した[文献 8]. この問題点を解決するため, アルゴリズム面の検討とソフトウェア開発と全体 FPGA システム仕様再設計を行った. このような経緯で 19 年度予算の一部を繰越し, 平成 20 年度にはメモリバンド幅ボトルネックを隠蔽するため大容量低速メモリをメモリディスクに組み込む機構を含んだシステムを設計した. 平成 20 年 10 月をもってこのシステムを利用した解析を完了させた.

(2.2)限られたクロックサイクルで高速処理を行うため 5 個の FPGA をボードに搭載, データ検索において, メモリバンド幅がボトルネックになりがちであるため, 各 FPGA 間を Rocket-IO で高速に通信させ, 外部との通信は, 10Gbps



の高速 Ethernet で接続するボードを作成した. このボードは, FPGA のファームウェアを流し込むことで, 計算を行なわせることができるため, 専用アクセラレータボードとして利用できる. このバンド幅ボトルネックとなる計算を中心とする, 並列化方式の研究や通信実験を行なった[文献 9,11,12].

また, 提案者稲葉は, 平成 16 年度から平成 20 年度にかけて「分散共有型研究データ利用基盤の整備」GRAPE-DR システムの開発にも携わっていたが, このデータ利用基盤は, 独自開発した ASIC と FPGA を搭載した加速ボードが 2 枚ずつ, 500 台のホストサーバに搭載されたクラスタシステムで超並列高速計算機構とネットワーク分散ストレージからなっている[文献 21]. 現在も, この超並列

システムを使いこなすことを目標として，より使いやすいスクリプト言語による超並列計算サポートの枠組を，計算の最適化手法と並行して検討，検証を行っている．

研究成果の総括

半導体プロセスの発達により，メモリの階層構造は，より深化していく傾向にある．また近年のマルチコアの発展により，キャッシュがどのコアに利用されるか等，新たな問題が出現しつつある現在，メモリ階層を意識した，メモリアクセスレイテンシーの隠蔽技術の研究に取り組むことは，計算機科学の発展に大きな意味をもっている．

本研究では，圧縮簡潔データ構造による検索システムをターゲットとし，メモリ階層構造を利用したアクセスレイテンシーの隠蔽の研究を行なった．

離散最適化アルゴリズムを応用し，整数計画法を用いた最適化により，階層間メモリ通信最適化のために定式化を行い，スケジューリング方式を提案，SIMD コンピュータを対象とする一般的な枠組みを提示し，パイプライン化を考慮した拡張を行い，実際にコンパイラへの実装・実験を行った．

また，大容量データの効率的な利用のための枠組みを提案し，それぞれ大きな成果をあげた．さらにこれと同時に，当初の計画にはなかった，プリフェッチアルゴリズムおよびキャッシュリプレースメントアルゴリズムの提案を行い，研究が想定していた以上の発展を見せた．

FPGA を利用し，高速ビットマニュピレーションを多用する圧縮簡潔データ構造を利用した検索については，実際の DNA の塩基配列のデータを用いた実験で，6 倍程度の高速化が達成された．これは，当初の目標値 10 倍という性能より低かったため，解析を行ったところ，ホストメモリとアクセラレータメモリのバンド幅ボトルネックが性能に大きな影響を与えることがわかった．この結果をもとに FPGA を５台搭載して，高速に FPGA が通信を行なうことのできるボードの開発を行なった．

５．主な発表論文等

〔雑誌論文〕（計２件）

[1] Y. Sugawara, M. Inaba, K. Hiraki, “Flow Balancing Hardware for Parallel TCP Streams on Long Fat pipe Network”, International Journal of Software Engineering and Its Applications(FGCN, selected paper), vol2 no 2, pp. 1-12

[2] N. Tanida, K. Hiraki, M. Inaba, “Efficient disk-to-disk copy through long-distance high-speed networks with background traffic”, Fusion Engineering and Design. An International Journal for Fusion Energy and Technology devoted to Experiments, Theory, Methods and Design, To appear (2010)

〔招待解説論文〕（計１件）

[3] 平木敬,稲葉真理,菅原豊,吉野剛史,玉造潤史,加藤朗：”Internet2 Land Speed Record 長距離 TCP 通信高速化への挑戦”, 情報処理学会論文誌招待解説論文, vol.49, no.2, pp.179-186, 2008

〔学会発表〕（計 19 件）

[4] Y. Ishii, M. Inaba, K. Hiraki, “Map-Based Adaptive Insertion Policy”, CRC-1 Workshop of the ACM IEEE International Symposium on Computer Architecture, to appear, (2010)

[5] Y. Iguchi, N. TANIDA, K. Koizumi, M. Inaba, K. Hiraki, “ USADAFOX: Ultra-High-Speed file-Acquisition-system over Distance with Apache and fireFOX”, TERENA Networking Conference 2010 (TNC2010)

[6] T. Sonobe, M. Inaba, “SAT CNF Encoding with Multi-modeling”, The 3rd Annual Meeting of the Asian Association for Algorithms and Computation (AAAC2010), pp.45, 2010

[7] N. Tanida, K. Hiraki, M. Inaba, “ Highly efficient data transmission facility through very long distance high-speed networks ”, Seventh IAEA Technical Meeting on Control, Data Acquisition, and Remote Participation for Fusion Research, http://www-fusion-magnetique.cea.fr/tmiaea2009/website/data/articles/000025.pdf, 2009

[8] N. Tanida, M. Inaba, K. Hiraki, and T.Yoshino, “ Hardware Accelerator for Full-Text Search (HAFTS) with Succinct Data Structure”, International Conference on ReConFigurable Computing and FPGAs (RECONFIG09), pp.155-160, 2009

[9] K. Koizumi, M. Inaba, K. Hiraki, Y. Ishii, T. Miyoshi, K. Yoshizoe, “Triple Line-based Playout for Go － An Accelerator of Monte Carlo Go ”, International Conference on

ReConFigurable Computing and FPGAs (RECONFIG09), pp.161-166, 2009
[10] Y. Ishii, M. Inaba, K. Hiraki, “Access Map Pattern Matching for Data Cache Prefetch”, ACM International Conference on Super-computing (ICS'09), pp.499-500 (poster), 2009
[11] M. Inaba, K. Koizumi, T. Yoshino, Y. Sugawara, J. Tamatsukuri, H. Tezuka, K. Hiraki, 2The Effect of the Buffer of the Path-Bottleneck Switch of Long Fat-pipe Network”, International Workshop on Protocols for Future, Large-Scale & Diverse Network Transports (PFLDNeT09), pp.7-12, 2009
[12] K. Koizumi, T. Yoshino, Y. Sugawara, M. Inaba, K. Hiraki, “A Reconfigurable Hardware Mechanism for Harmonizing Parallel TCP Streams of 10 Gigabit Ethernet”, International Workshop on Protocols for Future, Large-Scale & Diverse Network Transports (PFLDNeT09), pp.41-46, 2009
[13] Y. Ishii, M. Inaba, and K. Hiraki, “Access Map Pattern Matching Prefetch: Optimization Friendly Method”, The Journal of Instruction-Level Parallelism Data Prefetching Championship The 1st JILP Data Prefetching Championship (DPC-1), http://www.jilp.org/dpc/online/papers/03ishii.pdf, 2009
[14] Y.Ito, Y. Sugawara, M. Inaba, K. Hiraki, “CVC: The C to RTL compiler for callback based verification model”. International Conference on Field Programmable Logic and Applications (FPL2009), CD-ROM (short paper), 2008
[15] Y. Sugawara, T. Yoshino, H. Tezuka, M. Inaba, K. Hiraki, “Effect of Parallel TCP Stream Equalizer on Real Long Fat-pipe Network”, The 7th IEEE International Symposium on Network Computing and Applications (IEEE NCA08) (short paper), pp.279-282, 2008
[16] Y. Sugawara, T. Yoshino, H. Tezuka, M. Inaba, K. Hiraki, “Effect of Packet Shuffler on Parallel TCP Stream Network”, IEEE The Seventh International Conference on Networking (ICN 2008) pp. 180-185, 2008
[17] Y. Sugawara, T. Yoshino, M. Inaba, and K. Hiraki, “Fine Tune for parallel TCP Streams on Long Fat-pipe Network using Hardware Engine”, 6th International Workshop on Protocols for FAST Long-Distance Networks (PFLDnet2008), CD-ROM, 2008
[18] Y. Sugawara, M. Inaba, K. Hiraki, “Flow Balancing Hardware for Parallel TCP Streams on Long Fat pipe Network”, International Journal of Software Engineering and Its Applications(FGCN, selected paper), Vol.2, No.2, pp.1-12, 2008
[19] T. Yoshino, Y. Sugawara, K. Inagami, J. Tamatsukuri, M. Inaba and K. Hiraki, “Performance Optimization of TCP/IP over 10 Gigabit Ethernet by Precise Instrumentation”, SuperComputing2008(SC08), USB-stick, Austin, Texas, U.S.A.
[20] H. Hayashizaki, Y. Sugawara, M. Inaba and K. Hiraki, “MCAMP: Communication Optimization on Massively Parallel Machines with Hierarchical Scratch-pad Memory”, Parallel Architectures and Compilation Techniques (PACT) 2008, pp.102-111, Toronto, Canada
[21] J. Makino, K. Hiraki, and M. Inaba, GRAPE-DR: 2-Pflops massively- parallel computer with 512-core, 512-Gflops processor chip for scientific computing”, SuperComputing2007(SC07), USB-stick, Reno USA
[22] M. Inaba, K. Hiraki, “Network Processing Hardware”, Springer LNCS4311, Proc. Second ASIAN INTERNET ENGINEERING CONFERENCE, AINTEC 2006, pp.103-112

〔図書〕(計 0件)

〔産業財産権〕
○出願状況 (計 0件)
○取得状況 (計 0 件)

〔その他〕
なし

６．研究組織
(1)研究代表者
稲葉 真理・東京大学大学院情報理工学系研究科・准教授
研究者番号：60282711
(2)研究分担者
なし
(3)連携研究者
今井 浩・東京大学院情報理工学系研究科・教授
研究者番号：80183010
定兼 邦彦・国立情報学研究所情報学プリンシプル研究系・准教授
研究者番号：20323090