VI. 高性能計算システム研究部門

1. メンバ

教授　　佐藤 三久，朴 泰祐，児玉 祐悦

准教授　建部 修見，高橋 大介，塙 敏博

助教　　多田野 寛人

2. 概要

本研究グループは，高性能計算システムアーキテクチャ，省電力システムアーキテクチャ，並列数値処理の高速化研究，広域分散環境におけるデータ共有を中心とするグリッド計算技術等の研究を行っている．

3. 研究成果

■ **戦略的国際科学技術協力推進事業（日仏共同研究）「ポストペタスケールコンピューティングのためのフレームワークとプログラミング」を開始（佐藤）**

【省電力／高性能／ディペンダブル並列システムに関する研究】

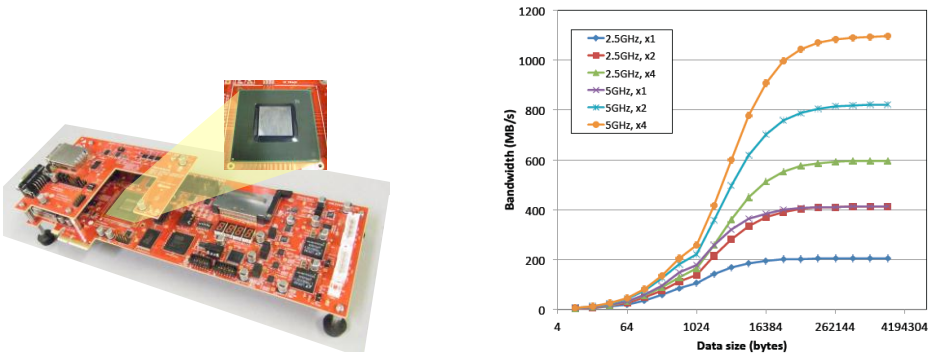
JST-CREST 研究領域「実用化を目指した組込みシステム用ディペンダブル・オペレーティングシステム」における研究課題「省電力高信頼組込み並列プラットフォーム」として，以下の研究を行った．

(1) クラウド技術を用いた高信頼並列ソフトウエアのテスト環境 D-Cloud の開発（佐藤，塙）

クラウドコンピューティングシステムを用いての計算資源の管理をベースとして，テストシステムの構築，フォルトインジェクションによるハードウェア故障エミュレーション，様々なテストケース・シナリオの自動化機能を提供するクラウドコンピューティングシステム D-Cloud のプロトタイプを開発し，デモンストレーションシステムを通じてその有効性を示した．D-Cloud を用いることにより短期間で効率的なプログラムテストが可能となった．これに関連し，SpecC によるハードウェアモデルと仮想マシンを統合したフォルトインジェクションツール FaultVM-SpecC を開発し，D-Cloud 環境への対応を行った．

(2) 高性能・省電力・耐故障並列システムリンクに関する研究（朴，塙）

ルネサステクノロジ社との共同研究により，PCI-Express gen.2 テクノロジに基づく並列システム向け相互結合網リンク PEARL の開発を行った．これを実現する通信スイッチチップのプロトタイプである PEACH チップの実装と実際に PC サーバ等に適用可能な PCI-Express 仕様のテストボードの開発と量産を行った．これにより，過去４年間に渡り研究開発を行ってきた PEACH チップに基づく PEARL ネットワークの基本通信部分が完成した．



完成したテストボード（PEACH チップ搭載）と基本通信性能評価

(3) 不均質な Ethernet トランクによる高性能・耐故障ネットワーク（朴，塙）

Gigabit Ethernet を複数本束ねることにより高性能・耐故障性を持つ汎用仮想ネットワークを構築する技術について，特にノード間接続のトランク本数が不均一な場合におけるトラフィックの偏りを自動的に検出し，常に最適なトラフィックバランスを保つフィードバック機能を持つネットワークシステムを開発した．従来のネットワークは均質な構造を想定していたが，コストパフォーマンスの観点からクライアントノードは少数リンクで，サーバノードは多数リンクで構築するのが自然である．このような状況ではトラフィックに空間的・時間的なばらつきが生じるが，本システムはこれを動的に検出し最適化し，常にバランスの取れた性能を実現する．

【次世代並列処理言語 XcalableMP の研究開発】（佐藤，朴）

E-Science プロジェクト「並列プログラミング言語に関する研究開発 」において，分散メモリ構成を基本とする大規模並列処理システムにおける並列 HPC アプリケーションのため，並列プログラミング言語 XcalableMP（XMP）の開発を行った．言語仕様の基本部分の完成と，PC クラスタを対象としたプロトタイプ実装を行った．XMP を用いた各種アプリケーションカーネルのコードについて性能及びスケーラビリティを評価し，現在の言語仕様の下で典型的な大規模科学技術コードの性能可搬性を持つ記述が可能であることを確認した．C 言語版の XMP コンパイラ 0.5 をリリースし，国際会議 SC10 における HPC Challenge Class2 において Honorable Mention を受賞した．

【仮想マシン環境における省電力化技術の研究とデータセンター電力ベンチマーク】（佐藤）

NEDO グリーン IT「エネルギー利用最適化データセンタ基盤技術の研究開発／データセンタのモデル設計と総合評価」において，データセンター電力ベンチマークの検討を行うとともに，現在，データセンターの運用で注目されている仮想マシン環境において，省電力化を行うための研究を進めた．

【大規模広域分散ファイルシステム及びグリッド／クラウド技術に関する研究】

(1) 大規模環境における広域ファイルシステムの評価（建部）

オープンソースで研究開発を進めている Gfarm ファイルシステムの大規模広域環境における性能評価を行った．全国主要大学機関を結ぶネットワーク上での実証実験を行い，その有効性を示した．また，Gfarm ファイルシステム上で自動ファイル複製政策，更新型複製間一貫性制御機能を実現した．これにより，広域分散環境における一層の性能向上と信頼性の向上が実現される．また，ファイルデータのアクセス及び維持に関する信頼性向上を図り，システムの実用性を高めた．

(2) JLDG におけるグループ内広域ファイル共有の実現（建部, 佐藤）

素粒子物理研究グループとの共同研究により，国内の素粒子物理学研究者における大規模データ共有基盤 JLDG の構築, 運用を行っている．前年度より開始された KEK，金沢大，京大，阪大，広島大の各拠点との共同運用の下，ファイル共有の実運用を開始した．また，今後より大規模な広域分散環境における運用を見据え，特に HPCI におけるファイル共用のプロトタイプとしての検討を行った．

【高性能並列数値計算に関する研究】

(1) 並列高速フーリエ変換（FFT）の自動チューニング手法に関する研究（高橋）

科学技術計算で広く用いられている並列 FFT の性能を改善するために，自動チューニング手法の研究を行った．並列三次元 FFT において，二次元分割により通信時間を削減すると共に，演算と通信をオーバーラップさせることで，従来の実装に比べてさらに性能を改善した．また，理化学研究所計算科学研究機構で運用開始予定の「京」コンピュータでの実用に向けた FFT アルゴリズムのチューニングを開始した. 他方, FFT 性能に重要な影響を与える並列全対全通信に着目し，二段階処理による全対全通信の開発と自動チューニングを導入し，MPI コミュニケータプロセス数の自動調整による高速化を達成した．

(3) Block Krylov 部分空間反復法に関する研究（多田野）

複数本の右辺ベクトルをもつ連立一次方程式を高速・高精度で解くための Block Krylov 部分空間反復法の研究を行った．これまで開発した Block BiCGGR 法は高精度近似解が生成できる一方で，右辺ベクトル数が多い場合は残差が発散することがあった．同法よりも高い収束性を持ち，かつ高精度近似解を生成する新たな方法を開発し，数値実験を通して解法の頑健性を示した．日本応用数理学会 2010 年度年会においてこの解法を発表し, 同学会第 7 回若手優秀講演賞を受賞した．

4. 研究業績

(1)研究論文

1. D. Takahashi: Parallel implementation of multiple-precision arithmetic and 2,576,980,370,000 decimal digits of π calculation, Parallel Computing, Vol.36, No.8, pp.439-448, 2010.
2. Y. Sato, D. Takahashi and R. Grimbergen: A Shogi Program Based on Monte-Carlo Tree Search, ICGA Journal, Vol.33, No.2, pp.80-92, 2010.
3. 佐藤佳州，高橋大介：探索結果を利用した実現確立探索，情報処理学会論文誌， Vol.51, No.11, pp.2021-2030, 2010.
4. J. Iwata, D. Takahashi, A. Oshiyama, T. Boku, K. Shiraishi, S. Okada and K. Yabana: A massively-parallel electronic-structure calculations based on real-space density functional theory, Journal of Computational Physics, Vol. 229, No. 6, pp. 2339-2363, 2010.
5. I. Yamazaki, H. Tadano, T. Sakurai, and K. Teranishi. A Convergence Improvement of the BSAIC Preconditioner by Deflation. JSIAM Letters, Vol. 3, pp. 5-8, 2011.
6. Y. Futamura, H. Tadano, and T. Sakurai. Parallel Stochastic Estimation Method of Eigenvalue Distribution. JSIAM Letters, Vol. 2, pp. 127-130, 2010.
7. H. Ohno, Y. Kuramashi, T. Sakurai, and H. Tadano. A Quadrature-Based Eigensolver with a Krylov Subspace Method for Shifted Linear Systems for Hermitian Eigenproblems in Lattice QCD. JSIAM Letters, Vol. 2, pp. 115-118, 2010.
8. J. Asakura, T. Sakurai, H. Tadano, T. Ikegami, and K. Kimura. A Numerical Method for Polynomial Eigenvalue Problems Using Contour Integral. Japan J. Indust. Appl. Math., Vol. 27, Iss. 1, pp. 73-90, 2010.
9. I. Yamazaki, M. Okada, H. Tadano, T. Sakurai, and K. Teranishi. A Block Sparse Approximate Inverse with Cutoff Preconditioner for Semi-Sparse Linear Systems Derived from Molecular Orbital Calculations. JSIAM Letters, Vol. 2, pp. 41-44, 2010.
10. H. Umeda, Y. Inadomi, Y. Watanabe, T. Yagi, T. Ishimoto, T. Ikegami, H. Tadano, T. Sakurai, and U. Nagashima. Parallel Fock Matrix Construction with Distributed Shared Memory Model for the FMO-MO Method. J. Comput. Chem., Vol. 31, Iss. 13, pp. 2381-2388, 2010.
11. H. Tadano, Y. Kuramashi, and T. Sakurai. Application of Preconditioned Block BiCGGR to the Wilson-Dirac Equation with Multiple Right-Hand Sides in Lattice QCD. Comput. Phys. Comm., Vol. 181, pp. 883-886, 2010.
12. 米元大我，塙敏博，三浦信一，朴泰祐，佐藤三久，"トラフィック量に適応する非対称マルチリンクEthernetトランキング"，情報処理学会論文誌コンピューティングシステム, Vol.3, No.1, pp.25-37, 2010.
13. 鈴木克典，建部修見，”PCクラスタ間ファイル複製スケジューリング”， 論文誌コンピューティングシステム(ACS)， 情報処理学会， Vol.3，No.3， pp.113-125, 2010.

14. O. Tatebe, K. Hiraga, N. Soda, "Gfarm Grid File System", New Generation Computing, Ohmsha, Ltd. and Springer, Vol.28, No.3, pp.257-275, 2010.
15. V.-T. Tran, G. Antoniu, B. Nicolae, L. Bougé and O. Tatebe, "Towards a Grid File System Based on a Large-Scale BLOB Management Service", Grids, P2P and Services Computing, Springer, pp.7-19, 2010.

(2)学会発表

(A)招待講演

1. M. Sato: Trends in Post-petascale computing -- from Japanese NGS project "The K computer" to Exascale computing, The 13th IEEE International Conference on Computational Science and Engineering (CSE-2010), Hong-Kong, 2010.

(B)その他の学会発表

1. D. Mukunoki and D. Takahashi: Implementation and Evaluation of Quadruple Precision BLAS Functions on GPU, Proc. Workshop on State of the Art in Scientific and Parallel Computing (PARA2010), LNCS, Springer-Verlag, 2010.
2. 椋木大地，高橋大介：GPU による 4 倍・8 倍精度 BLAS の実装と評価，2010 年ハイパフォーマンスコンピューティングと計算科学シンポジウム HPCS2011 論文週，pp.148-156, 2010.
3. T. Sakurai, H. Tadano, T. Ikegami. A Hierarchical Parallel Eigenvalue Solver: Parallelism on Top of Multicore Linear Solvers. 6th International Workshop on Parallel Matrix Algorithms and Applications (PMAA'10), Switzerland, Jun. 2010.
4. Y. Futamura, H. Tadano, T. Sakurai. Parallel Stochastic Estimation Method for Matrix Eigenvalue Distribution. 6th International Workshop on Parallel Matrix Algorithms and Applications (PMAA'10), Switzerland, Jun. 2010.
5. S. Otani, H. Kondo, I. Nonomura, A. Ikeya, M. Uemura, Y. Hayakawa, T. Oshita, S. Kaneko, K. Asahina, K. Arimoto, S. Miura, T. Hanawa, T. Boku, M. Sato, "An 80Gb/s Dependable Communication SoC with PCI Express I/F and 8 CPUs", Proc. of ISSCC2011, San Francisco, CD-ROM, 2011.
6. T. Hanawa, T. Boku, S. Miura, M. Sato, K. Arimoto, "PEARL: Power-aware, Dependable, and High-Performance Communication Link Using PCI Express", Proc. of IEEE/ACM International Conference on Green Computing and Communitations (GreenCom2010), pp. 284-291, Hangzhou, 2010.
7. T. Hanawa, T. Boku, S. Miura, M. Sato, and K. Arimoto, "Power-aware, Dependable, and High-Performance Communication Link Using PCI Express: PEARL," Proc. of IEEE International Conference on Cluster Computing (Cluster2010), poster, 4 pages, Creta Island, Sep. 2010.
8. M. Nakao, J. Lee, T. Boku, M. Sato, "XcalableMP Implementation and Performance of NAS Parallel Benchmarks", Proc. of PGAS10, New York, 2010.
9. J. Lee, M. Nakao, M. Sato. SC10 HPC Challenge Submission for XcalableMP (selected as a finalist of

HPCC Class2), SC10, New Orleans, Louisiana, USA, Nov., 2010.

10. H. Kimura, T. Imada and M. Sato: Runtime Energy Adaptation with Low-Impact Instrumented Code in a Power-scalable Cluster System 10th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGRID2010), pp. 378-387, 2010
11. T. Hanawa, T. Banzai, H. Koizumi, R. Kanbayashi, T. Imada, and M. Sato, "Large-Scale Software Testing Environment Using Cloud Computing Technology for Dependable Parallel and Distributed Systems," the 2nd International Workshop on Software Testing in the Cloud (STITC2010), co-located with the 3rd IEEE International Conference on Software Testing, Verification, and Validation (ICST 2010), pp. 428-433, Apr. 2010.
12. T. Banzai, H. Koizumi, R. Kanbayashi, T. Imada, T. Hanawa, and M. Sato, "D-Cloud: Design of a Software Testing Environment for Reliable Distributed Systems Using Cloud Computing Technology", the 2nd International Symposium on Cloud Computing (Cloud 2010) in conjunction with the 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing (CCGrid 2010), pp. 631-636, May 2010.
13. T. Hanawa, H. Koizumi, T. Banzai, M. Sato, and S. Miura, "Customizing Virtual Machine with Fault Injector by Integrating with SpecC Device Model for a software testing environment D-Cloud," the 16th IEEE Pacific Rim International Symposium on Dependable Computing (PRDC'10), pp. 47-54, Dec. 2010.
14. T. Hanawa, T. Boku, S. Miura, M. Sato and K. Arimoto, "PEARL: Power-aware, Dependable, and High-Performance Communication Link Using PCI Express", IEEE/ACM International Conference on Green Computing and Communitations (GreenCom2010), pp. 284-291, Dec. 2010.
15. S. Otani, H. Kondo, I. Nonomura, A. Ikeya, M. Uemura, K. Asahina, K. Arimoto, S. Miura, T. Hanawa, T. Boku, M. Sato, "An 80Gb/s Dependable multicore Communication SoC with PCI Express I/F and Intelligent Interrupt Controller," IEEE Symposium on Low-Power and High-Speed Chips (COOL Chips XIV), 3 pages (regular), CD-ROM, 2011.
16. T. Amagasa, N. Ishii, T. Yoshie, O. Tatebe, M. Sato, H. Kitagawa: A Faceted-Navigation System for QCDml Ensemble XML Data. 3PGCIC 2010: pp.132-139, 2010.
17. T. Hanawa, T. Boku, S. Miura, M. Sato and K. Arimoto, "Power-aware, Dependable, and High-Performance Communication Link Using PCI Express: PEARL," IEEE International Conference on Cluster Computing (Cluster2010), poster, 4 pages, Sep. 2010.
18. Adnan and M. Sato: Flexible Fine Grain Threads Management By StackThreads/Mp Library for OpenMP Task Implementation, International Workshop on OpenMP2010 (IWOMP2010, poster session), 2010.
19. 鈴木克典，建部修見，「クラスタ間並列複製作成のためのファイル分割を許さないスケジューリング」，ハイパフォーマンスコンピューティングと計算科学シンポジウム HPCS2011 論文集，情報処理学会，pp.10-19，2011.
20. M. Tanaka, O. Tatebe, "Pwrake: A parallel and distributed flexible workflow management tool for wide-area data intensive computing", Proc. of ACM International Symposium on High Performance Distributed

Computing (HPDC), pp.356-359, 2010.

21. 田中昌宏，建部修見, “グラフ分割による広域分散並列ワークフローの効率的な実行”，先進的計算基盤システムシンポジウム SACSIS 2010 論文集， 情報処理学会，電子情報通信学会， pp.63-70, 2010.
22. 鈴木克典，建部修見， “PC クラスタ間ファイル複製スケジューリング”，先進的計算基盤システムシンポジウム SACSIS 2010 論文集，情報処理学会，電子情報通信学会，pp.71-78，2010.

5. 連携・国際活動・社会貢献、その他

1. 戦略的国際科学技術協力推進事業（日仏共同研究）「ポストペタスケールコンピューティングのためのフレームワークとプログラミング」を開始（佐藤）
2. 国際会議 International Conference on Supercomputing 2010 (ICS2010), Tsukuba, Jun. 2010 を開催, Conference General Chair（朴）
3. 国際会議 International Workshop on OpenMP 2010 (IWOMP2010), Tsukuba, Jun. 2010 を開催, Workshop General Chair（佐藤）