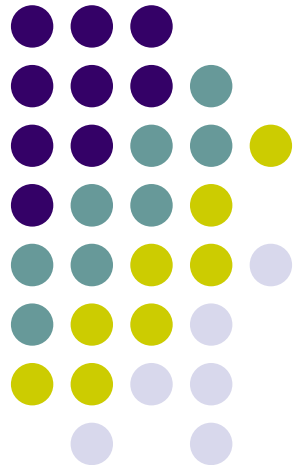
GF(2)上疎行列線形解法の現状と評価

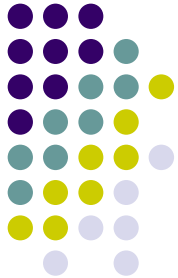
中央大学21世紀COEプログラム
JST CREST
西田 晃



背景

- 情報システムの安全性
  - 公開鍵暗号システムに依存
  - 最新の計算機環境による素因数分解のコストを常に正確に評価する必要

関連研究

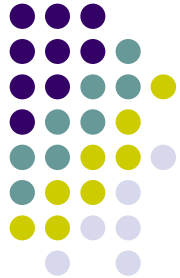


- 公開鍵暗号
  - GF(2) 上の疎行列線形方程式系を効率的に解く必要
    - 連分数法 (CFRAC, Morrison and Brillhart, 1975)
      - 合同式 $x^2 \equiv y^2 \bmod N$ の自明でない解を求める
      - ユークリッドの互除法により，$(x+y, N)$ から合成数 N の因数 p を求める
    - 二次篩法 (Pomerance, 1982)
    - 数体篩法 (Lenstra and Lenstra, 1993)
  - 計算コストの正確な評価

GF(2)の線形解法

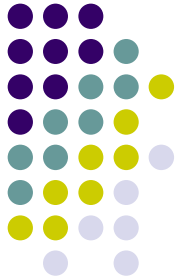
- 大規模行列 A の列に従属性を見つける
  - $Ax=0$, $x \neq 0$ を満たす x を計算
  - 乱数ベクトル x_0 から $b=Ax_0$ を計算し，$Ax=b$ を解く
  - $A(x-x_0)=0$
  - Aの列が線形従属ならば，$x \neq x_0$ である可能性が高い
- 非零ベクトルも自身に対して直交する可能性
  - $y^T y = 0$, $y \neq 0$
- Look-ahead Lanczos のアイデアを活用 (Montgomery's block Lanczos)

素因数分解の記録



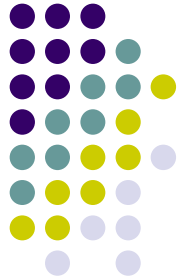
- 数体篩法による素因数分解の記録
  - Digits　Dates　Description　By
  - 200　05/2005　RSA-200　Bonn Univ. et al.
  - 193　11/2005　RSA-640　Bonn Univ. et al.
  - 176　04/2005　cofactor of $11^{281}+1$　Rikkyo Univ. et al.
  - 174　12/2003　RSA-576　Bonn Univ. et al.
  - 164　12/2003　cofactor of $2^{1826}+1$　Rikkyo Univ. et al.
  - 160　04/2003　RSA-160　Bonn Univ. et al.
  - 158　01/2002　co-factor of $2^{953}+1$　Bonn Univ. et al.
  - 155　08/1999　RSA-155　CWI et al.
- 外挿すると… (Brent, 2000)
  - $D^{1/3}=(Y-1928.6)/13.24$ (ムーアの法則から)
  - 2006年で200桁

計算時間



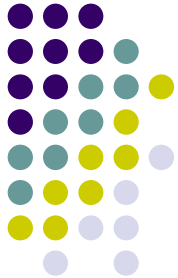
- RSA-155 (1999)
  - 8000MIPS-年
  - 篩部分
    - 8000MIPS-年
  - 行列計算部分
    - Cray C916 で 224CPU時間
      ≒6MIPS-年
- RSA-200 (2005)
  - 70000MIPS-年
  - 行列計算部分
    - 2.2GHz Opteron CPU 1台で55年
      ≒50000MIPS-年
  - 行列計算部分
    - 80台の 2.2GHz Opteron クラスタで約3ヶ月
      ≒20000MIPS-年

観察



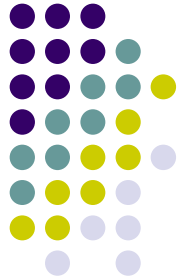
- GF(2) 上のスケーラブルな並列ソルバが必要
- どのように実現するか？

研究概要



- 特定領域研究「情報学」
  - 「最先端の情報通信システムを活用した新しい研究手法」(下條・松岡班)
- InfiniBand ＋ PCI Express の組み合わせで高度な通信性能を備えた PC クラスタ環境を構築
- 今回は主に広帯域環境での処理性能について紹介

反復法による線形系 *Ax=b* の求解



- 一般的な PC クラスタ環境では十分スケールしないことがある
  - NAS Parallel Benchmark CG kernel で評価可能
  - 共役勾配法を使用した疎行列計算ベンチマーク
- 原因
  - ネットワーク性能
    - （PCI バスを含む）帯域幅
    - レイテンシ
  - アルゴリズム
    - 通信が多く，データの待ち時間が長い

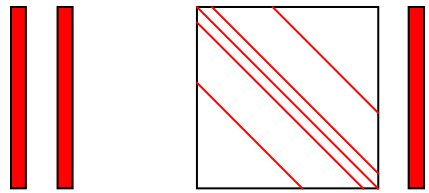
例

- 共役勾配法のアルゴリズム

1. Choose x_0
2. $p_0=r_0=b-Ax_0$

$k=0$

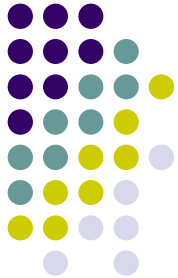
3. $\alpha_k=(r_k,p_k)/(p_k,Ap_k)$
4. $x_{k+1}=x_k+\alpha_k p_k$
5. $r_{k+1}=r_k-\alpha_k Ap_k$
6. $\beta_k=(r_{k+1},r_{k+1})/(r_k,r_k)$
7. $p_{k+1}=r_{k+1}+\beta_k p_k$
8. If not convergent, goto 3.



背景

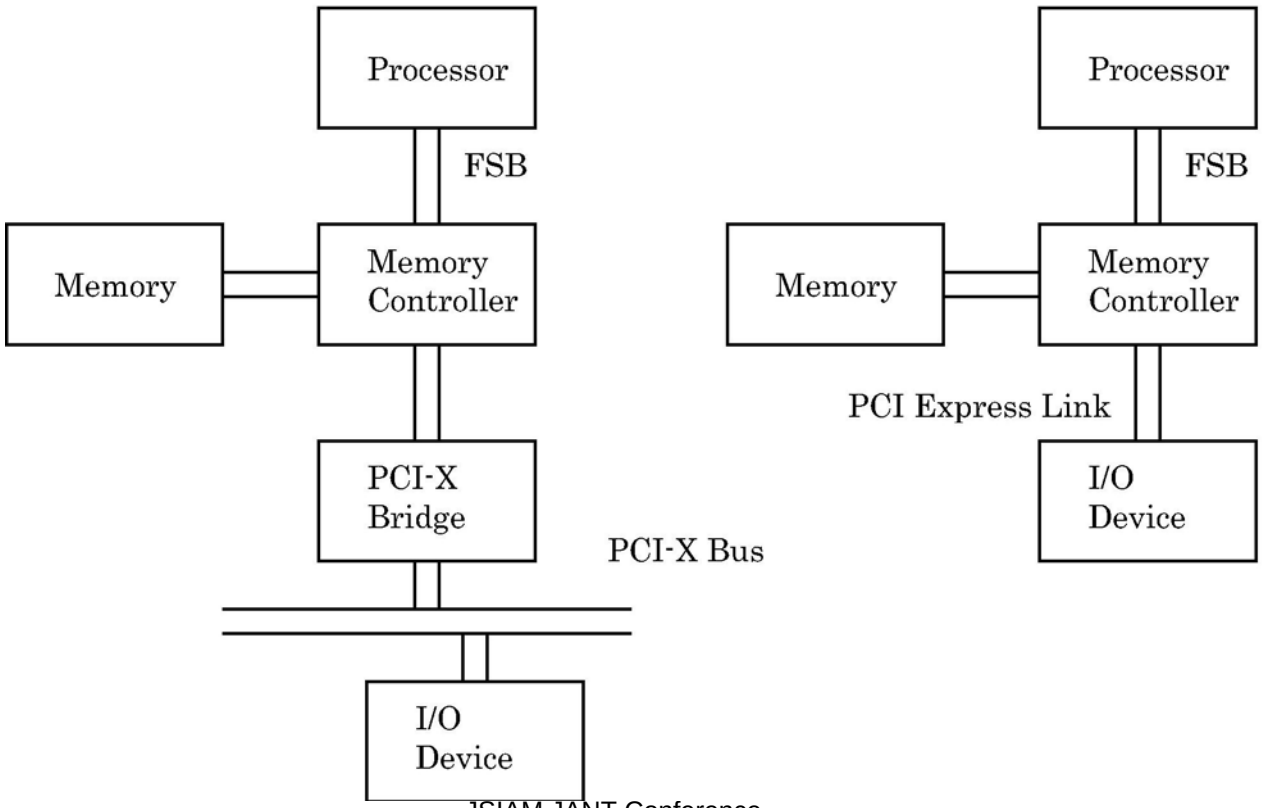
- 既存のクラスタ向けインターコネクト技術
  - Myrinet, Quadrics, GbE, etc.
  - PCI-X バスを利用
  - （アプリケーションによっては）帯域幅に限界
- 最新の技術を活用することで，より広帯域のクラスタを実現することはできないか？

PCI Express

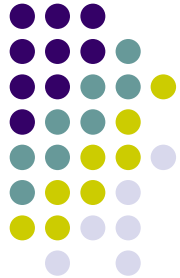


- PCI-X との互換性
- シリアル転送（片方向2.5Gbps／レーン × 最高32），point-to-point 接続
- AGP バスも統合
- ビデオカード用の x16 PCIe スロットを通信用に利用可能
- 2004年から実用化

PCI-X vs PCI Express



InfiniBand



- HCA（Host Channel Adapter）
  - 富士通，Mellanox Technologies, Inc.
  - Mellanox は PCI Express に対応（2004年から）
  - ～片方向2.5Gbps× 4／ポート × 2／アダプタ

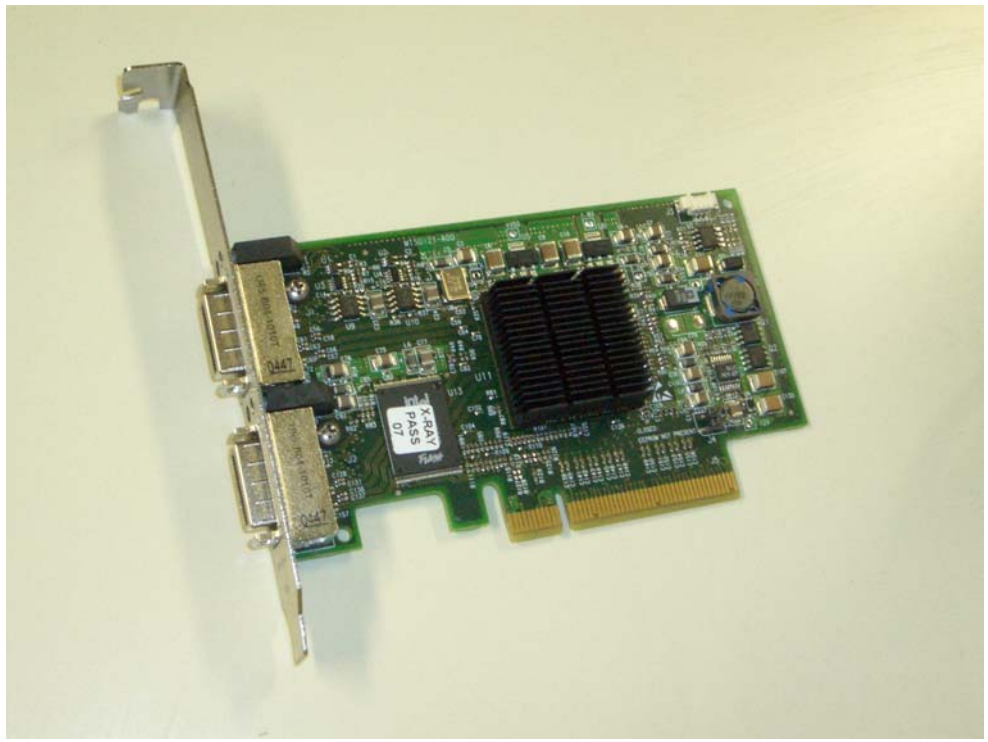
  （8B/10B データ符号化）→ 2GB/s
- Kernel 2.6.11 から Linux 標準カーネルにドライバ（OpenIB.org 版）を統合
- PCI Express 対応 Myrinet 10G も2006年から利用可能（今回は評価せず）

PCI Express 用 IB HCA

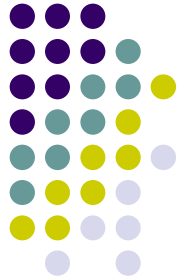
- Mellanox MHEL-CF128-T (128MB メモリ搭載）
- 約15万円

PCI Express 用 IB HCA

- 同 MHEA28-XT（メモリ非搭載版）
  - PCI Express の帯域を利用
- 約7万円

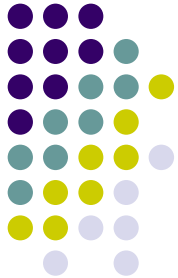


クラスタの構築



- 平成15年度
  - PCI Express, InfiniBand ＋ Opteron で計画
- 平成16年度
  - PCI Express 対応 InfiniBand HCA の登場
    - 16年秋より出荷開始（Mellanox）
    - HCA，24ポートスイッチを導入（ポート単価約3万円）
  - AMD の PCI Express への対応
    - NVIDIA nForce4 chipset
    - 年末から搭載マザーボードが出荷開始
      - Athlon64 … nForce4 SLI etc.（Asus, 16年12月）
      - Opteron … nForce Professional 2200（Rioworks, 17年3月）

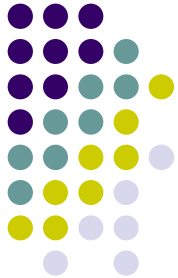
評価環境



- HCAは Mellanox InfiniHost チップ搭載の
  - MHEL-CF128-T（128MBメモリ搭載版）
  - MHEA28-XT（メモリ非搭載版）

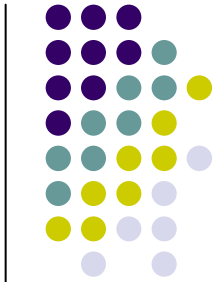
  を使用
- 通信ライブラリ
  - InfiniBand に対応した MPI ライブラリ
    - MVAPICH（Ohio State Univ.)
      - ストライピング機能を実装
        - 複数ポートにデータを分配可能
    - LAM MPI
      - Mellanox HCA も使用可
    - MPICH/Score
      - Cisco（旧 Topspin, InfiniBand ベンダ）の機器に対応

クラスタ構成



- ノードは Opteron ＋ nForce で構成
  - チップセット
    - nForce Professional 2200
  - まず 2-way 構成で
    - CPU: AMD Opteron 246 2.0GHz，1MB L2
    - M/B: Rioworks HDAM Express
      … 2005年3月に出荷
    - メモリ：512MB PC3200 DDR x 4
  - 8ノード16CPUで構築

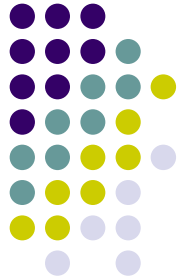
クラスタ構成



- InfiniBand
  - スイッチ:
    - 24ポート InfiniBand スイッチ MTS2400（Mellanox）
  - HCA
    - ノード 0－3 MHEL-CF128-T（128MBメモリ搭載版）
    - ノード 4－7 MHEA28-XT（メモリ非搭載版）
- GbE
  - Dell PowerConnect 2724
  - NIC: 32bit, 66MHz 対応 RTL8169 チップ搭載カード
- OS
  - SuSE Linux 9.1 Professional

IBD クラスタ

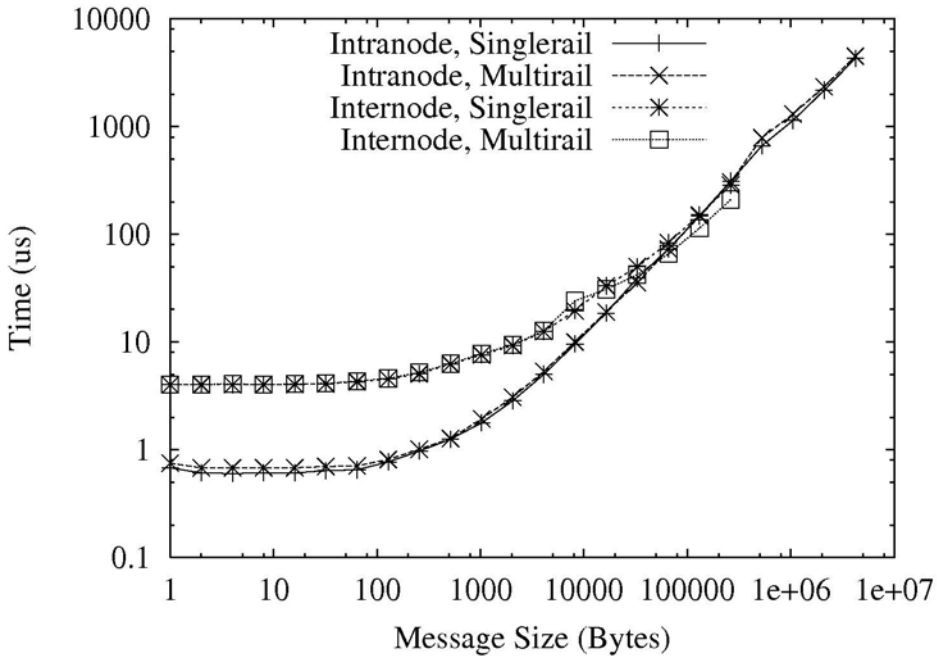
比較対象



- SGI Altix 3700
  - Intel Itanium2 Processor 1.3GHz，3MB L3 ×32
  - 主記憶 PC3200 DDR 32GB
  - このうち隣接した16CPUを使用
  - ノード
    - 2CPUを搭載
    - ノード内は 6.4GB/s, ノード間は3.2GB/s の SGI NUMAflex ネットワークで接続（fat tree）

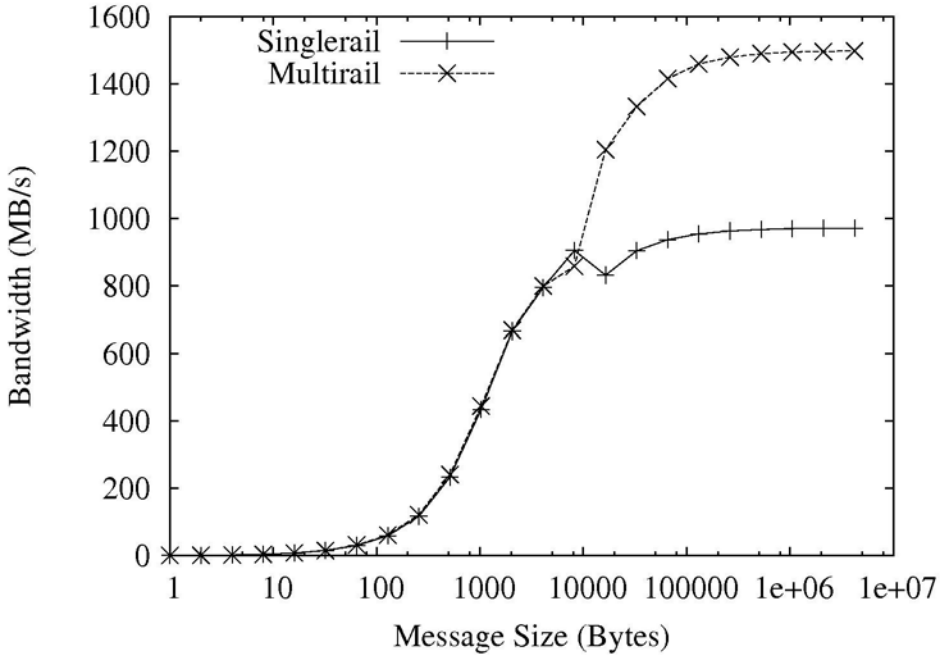
通信性能

- InfiniHost HCA（MHEL-CF128-T）のMPIレイテンシ



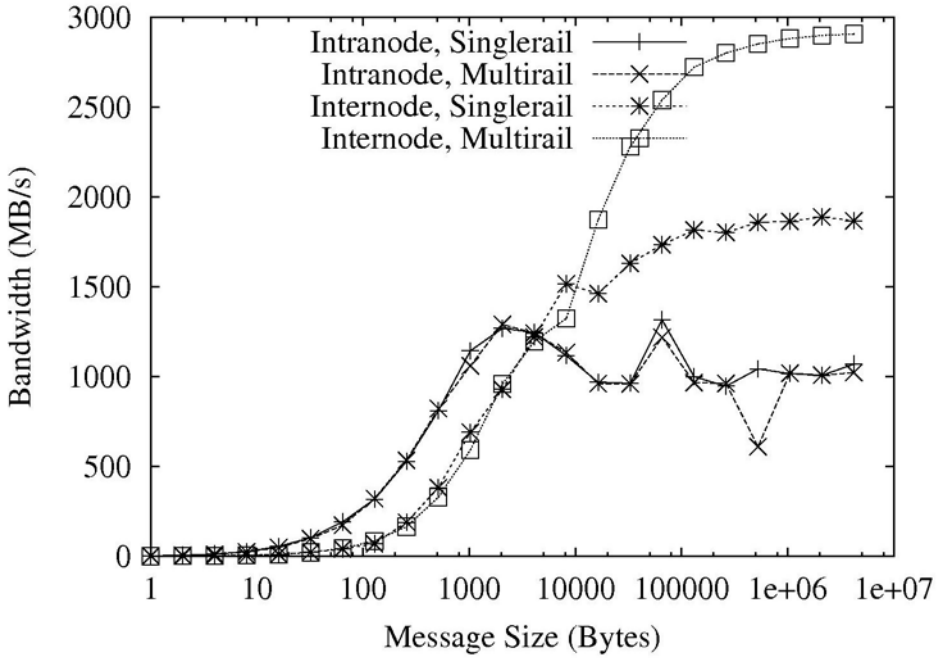
通信性能

- InfiniHost HCA（MHEL-CF128-T）の片方向帯域幅



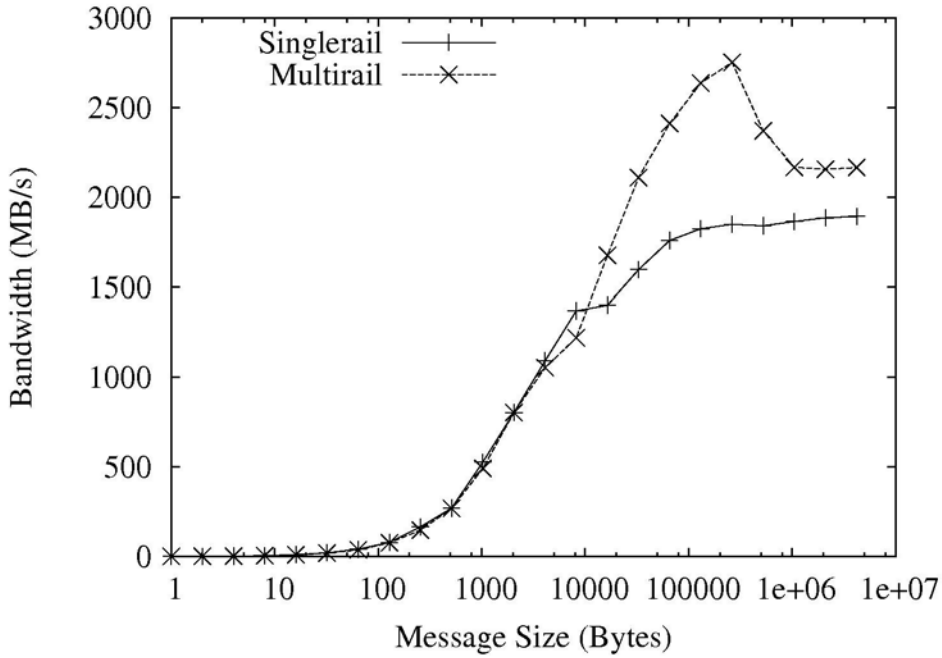
通信性能

- InfiniHost HCA（MHEL-CF128-T）の双方向帯域幅



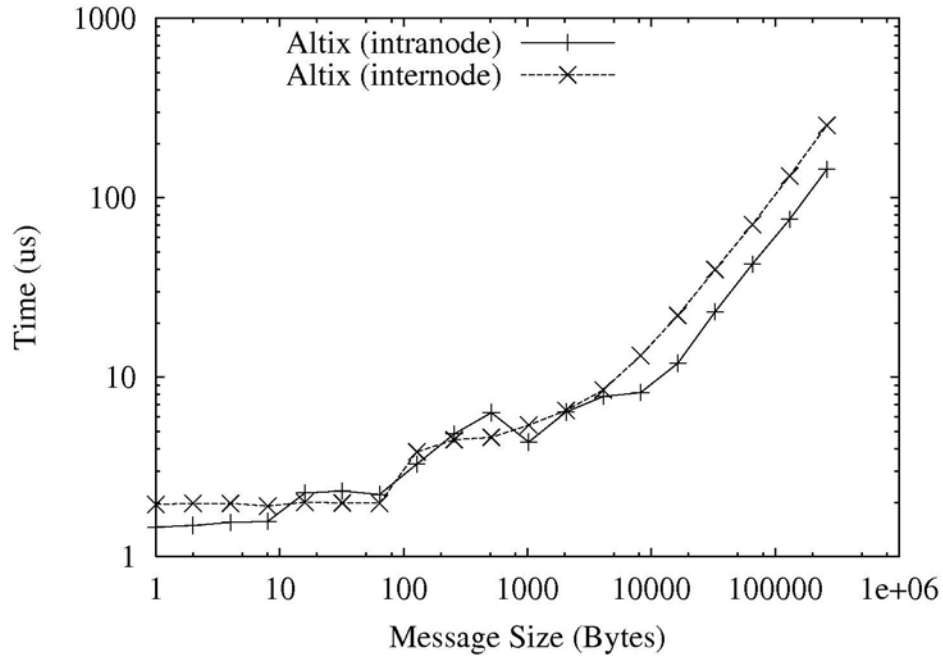
通信性能

- InfiniHost HCA（MHEA-28-XT）の双方向帯域幅



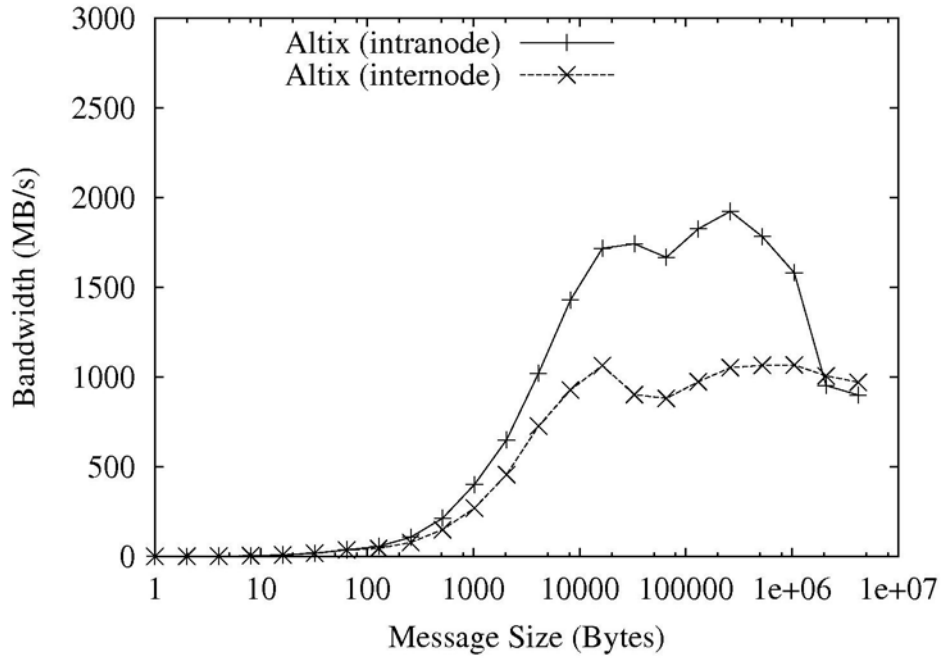
通信性能

- SGI Altix 3700 上での MPI レイテンシ



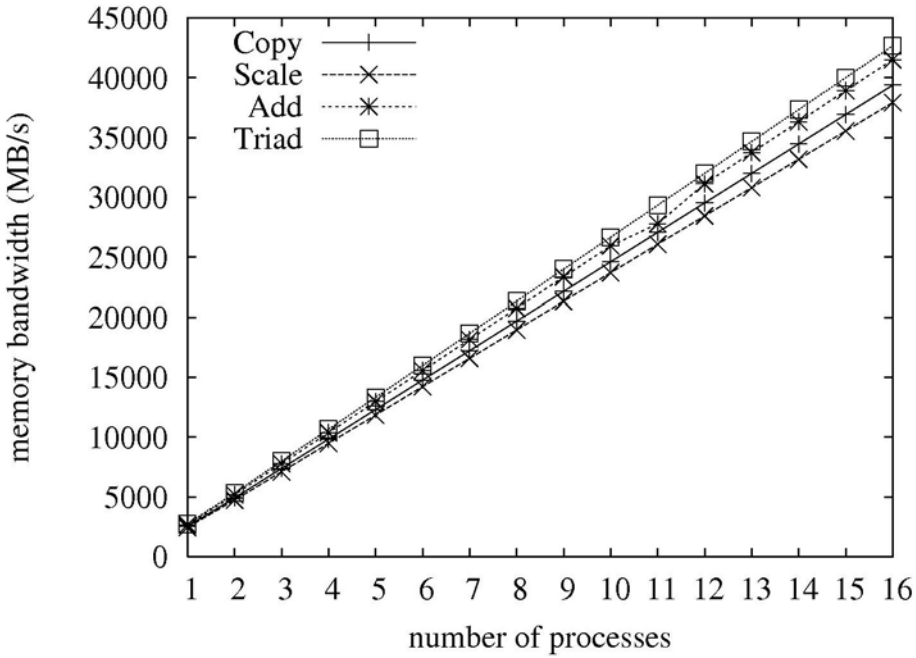
通信性能

- SGI Altix 3700 上での双方向帯域幅



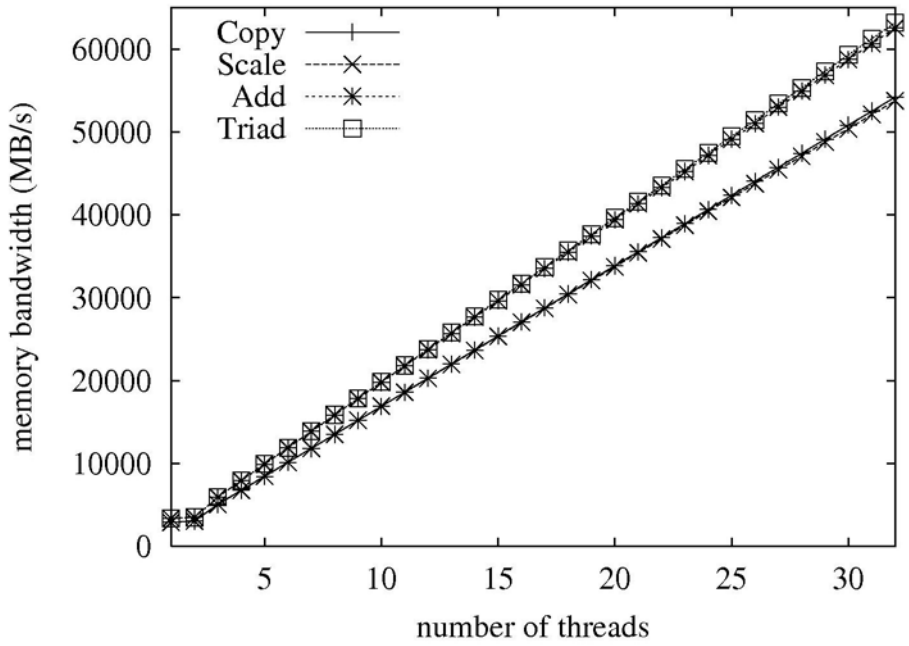
メモリ帯域幅

- ノード当たり2MPIプロセスで実行した場合のSTREAM benchmark 性能

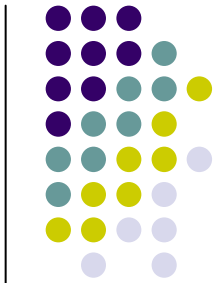


メモリ帯域幅

- SGI Altix 3700 上でのSTREAM benchmark 性能

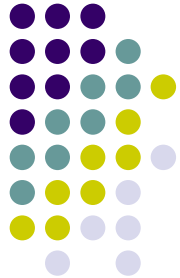


NAS Parallel Benchmark CG

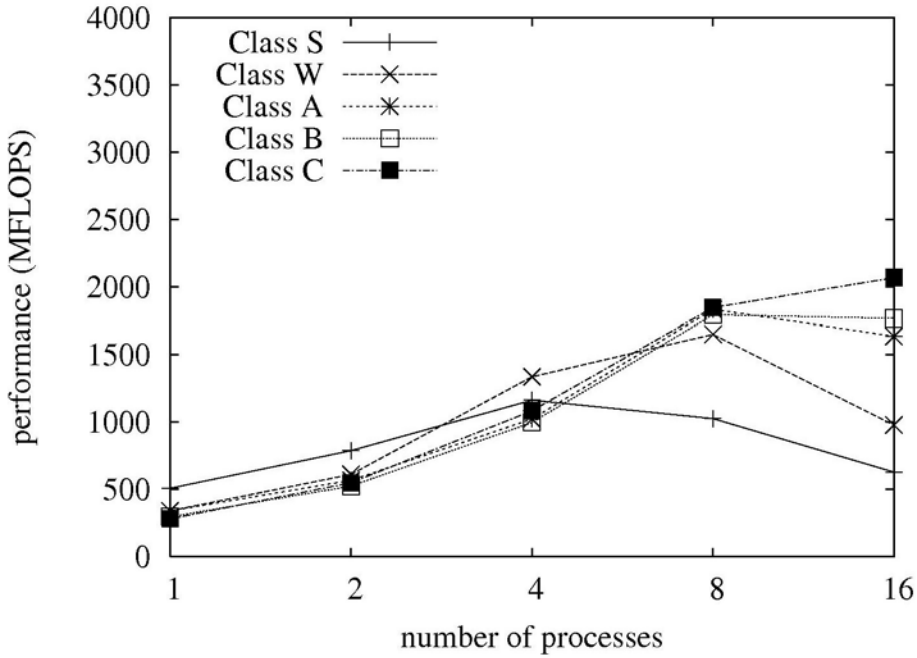


- 対称正定値行列の最小固有値を逆反復法と共役勾配法で計算
- MPI 版 Class S, W, A-C で評価

NAS Parallel Benchmark CG

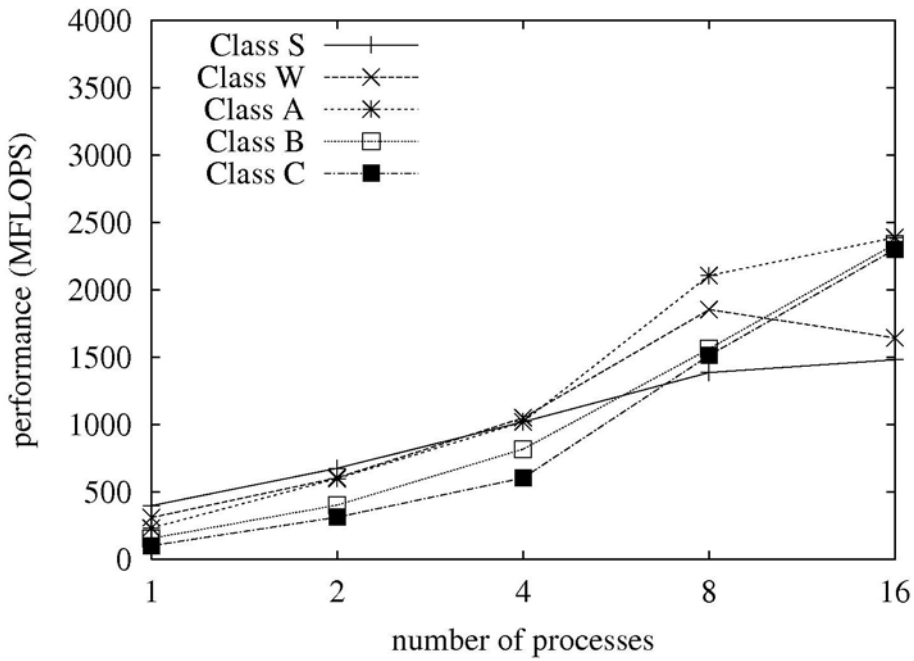


- SGI Altix 3700 上での MPI 版 CG の演算性能



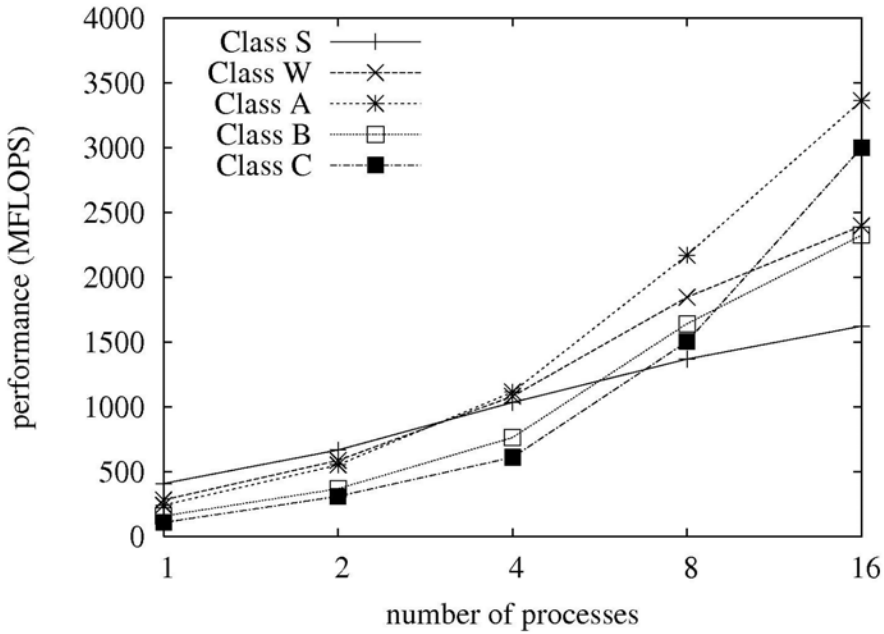
NAS Parallel Benchmark CG

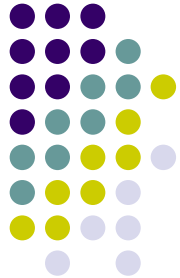
- クラスタ上で InfiniHost HCA 1ポートのみを使用した場合の演算性能



NAS Parallel Benchmark CG

- クラスタ上で InfiniHost HCA 2ポートを使用した場合の演算性能

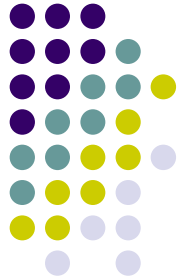




観察

- CGの性能は通信帯域幅によって決まる
  - 参考データ(7月8日現在)
    - Cray XT3　　Opteron 1シリーズを3Dトーラスで結合
      MPI 片方向帯域幅 1160MB/s
      MPI 双方向帯域幅 2080MB/s
      MPI レイテンシ 6.4us
      (通信性能に改善の余地)
    - Myri-10G　　PCI Express 対応 Myrinet カード
      MPI 片方向帯域幅 1204MB/s
      MPI 双方向帯域幅 2397MB/s
      MPI レイテンシ 2.4us

大規模素因数分解に向けて



- PCI Express 対応の通信カードを用いることにより，専用並列計算機と同性能の広帯域かつ低価格なクラスタ環境が構築可能
- ネットワーク性能の疎行列数値処理に与える影響を評価
  - 疎行列計算ではネットワーク帯域幅が律速