

RAM DISKを利用した SAS処理時間短縮の試み

Takeda Pharmaceutical Company Limited
Koji Iwamoto
SAS Institute Japan Ltd.
Akihiko Hirata

背景

- PCにSASをインストールし、個別にプログラムを実行
- 例数設計等でのシミュレーションの際、非常に時間がかかる(時には1日以上を要することも)
- 夜間バッチで途中でダウンすることも、しばしばあった
- ユーザーから要望として
 - 高速に処理したい
 - シミュレーション専用PCがほしい

原因

- 個々のPCによる性能のバラツキ
 - Pentium3/1GB ~ Pentium4/2.8G
- PCの限界？
- 個々のプログラミング能力
 - 非効率なプログラム
 - 非効率なロジック
 - 不必要なログやリスティング出力
- とはいえ...何かしなければならない事情があった

プログラムの効率化

<http://support.sas.com/sassamples/quicktips/04jun/performance.html>

CPU Performance Tuning Techniques for Beginners

By Kirk Paul Lafler

As SAS software becomes increasingly popular, guidelines for its efficient use are critical. Areas deserving special consideration include program execution, I/O, disk space and programming time. This tip relates to the area of performance tuning. It highlights techniques related to gaining better CPU efficiencies by improving the time a SAS program takes to decode and execute. The following tips are not meant as an exhaustive review of all known CPU efficiency techniques, but as a sampling of proven methods that can provide some measure of efficiency gain:

- Use KEEP or DROP statements or KEEP= or DROP= data set options to retain desired variables.
- Learn proper placement of KEEP= or DROP= data set options.
- Construct ANDed logic using the most restrictive conditions first.
- Utilize macros for redundant code.
- Use IF-THEN/ELSE statements to process data.
- Use the DATASETS procedure COPY statement to copy data sets with indexes.
- Use the SQL procedure to consolidate the number of steps.
- Turn off the Macro facility when not needed.
- Try to sort data once — avoid unnecessary sorting — and plan its use.
- Collapse large SAS data sets using the SUMMARY procedure.
- Create and use indexes with large data sets.
- Use CLASS statements in procedures when available to combine and group output.
- Use WHERE statements to subset SAS data sets.

解決策

- サーバーの導入
 - マルチCPUによる併行処理 (SASの対応)
- ハイパースレッド (HT) の利用
 - SASも対応している
- PCの性能向上
 - 各種設定のチューニング
 - メモリの増設とRAM Diskの利用
 - 高速ディスクなど足回りを改善
(シリコンディスク、メモリスティック)

今回はRAM DISKの導入を検討した

RAM DISKについて

- RAM DISKとは

- 仮想的にメモリをディスクのように使用して利用する仕組み
 - たとえば、1GBのメインメモリを2分割し、500MBをメモリに割り当てて、残りの500MBをハードディスクの代わりとして利用する。
- メモリのアクセス速度はハードディスクとはくらべものにならないくらい高速

- RAM DISKを選んだ理由

- メモリが安くなった(1GBで2～3万円/ノーブランド)
- 既存のPCにメモリを増設するだけである
- RAM DISKの設定は簡単である

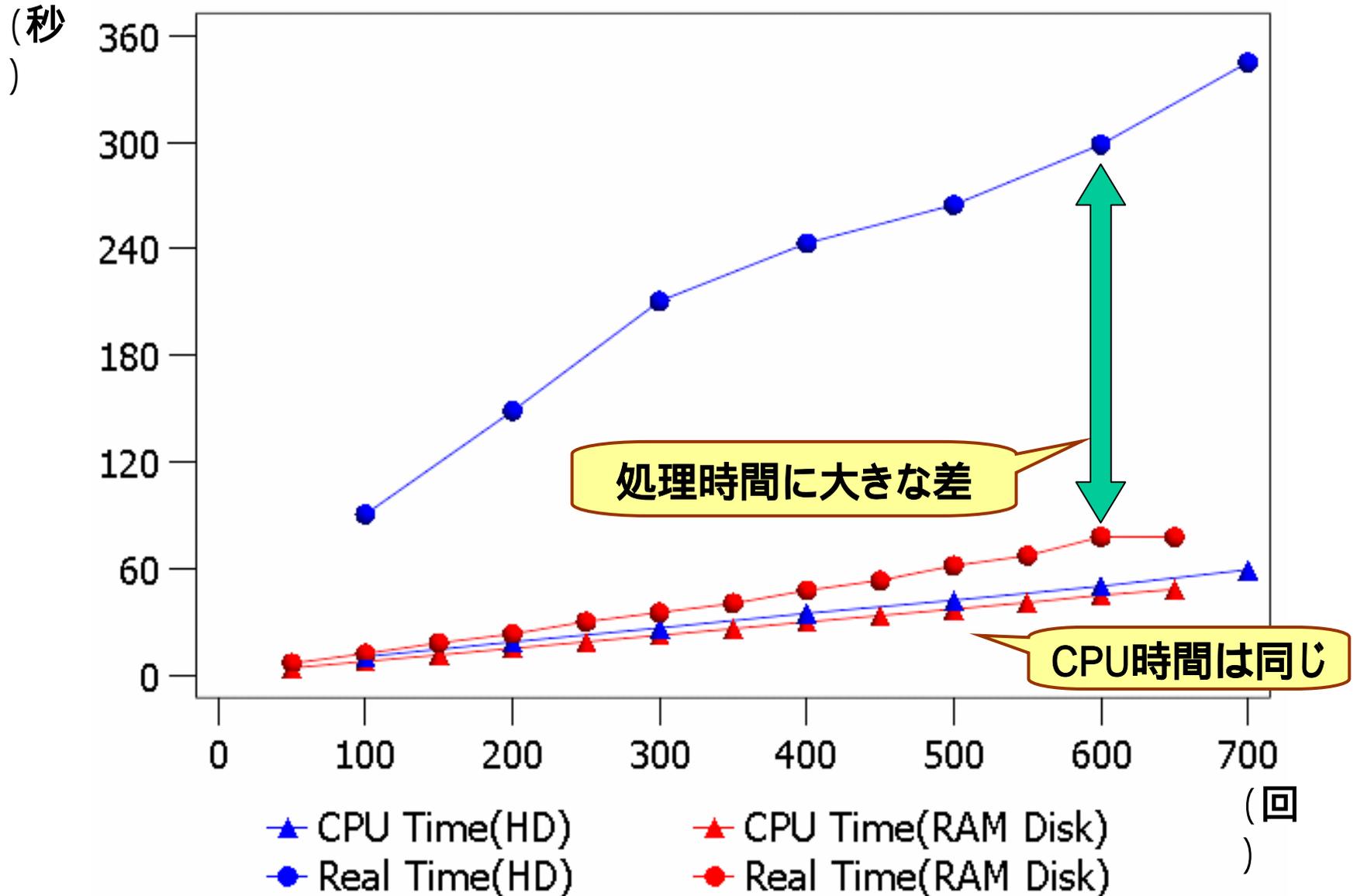
RAM DISKの種類

- WINSOFT RamDisk 2.3 for Windows 2000
- AR RAM Disk V1.20
- ERAM for Windows 2000 V1.12
- Qsoft RAMDisk V9.0
 - 以上はフリーウェアでMS純正のRAM Diskドライバを使用
- RamDisk for Windows 2000
 - Winsoft社が \$32.-USDで販売、独自のドライバを使用

調査結果 - 平田さん(1) -

- RAM DISKとしてQsoft RAMDisk V9.0を使用
- 使用したPCスペック
 - Dell Note PC Latitude D600 (Pentium4 1.6GHz)
 - 2GBメモリ(DDR SDRAM PC2700)
 - Windows2000 Profesional SP4
- 条件
 - SASのWorkファイルをRAM DISK(600MB)上に割り当てた
 - 繰り返し処理をするシミュレーションプログラムを使用
 - 繰り返しは100～700回
 - 使用データサイズは18MB～129MB
 - DATAステップが5割、他はSORT, SUMMARY, RANK, TRANSPOSE, MEANS, TABULATEプロシジャ

調査結果 - 平田さん(2) -



調査結果 - 平田さん(3) -

- CPU時間は、RAM DISKとHard Disk間で差が少ない
 - 若干、RAM DISKの方が速い
- 処理時間は、RAM DISKがHard Diskより約4倍速い
- CPU時間及び処理時間とも処理量とともに単調増加

以上のことから、RAM DISKの導入を決定した

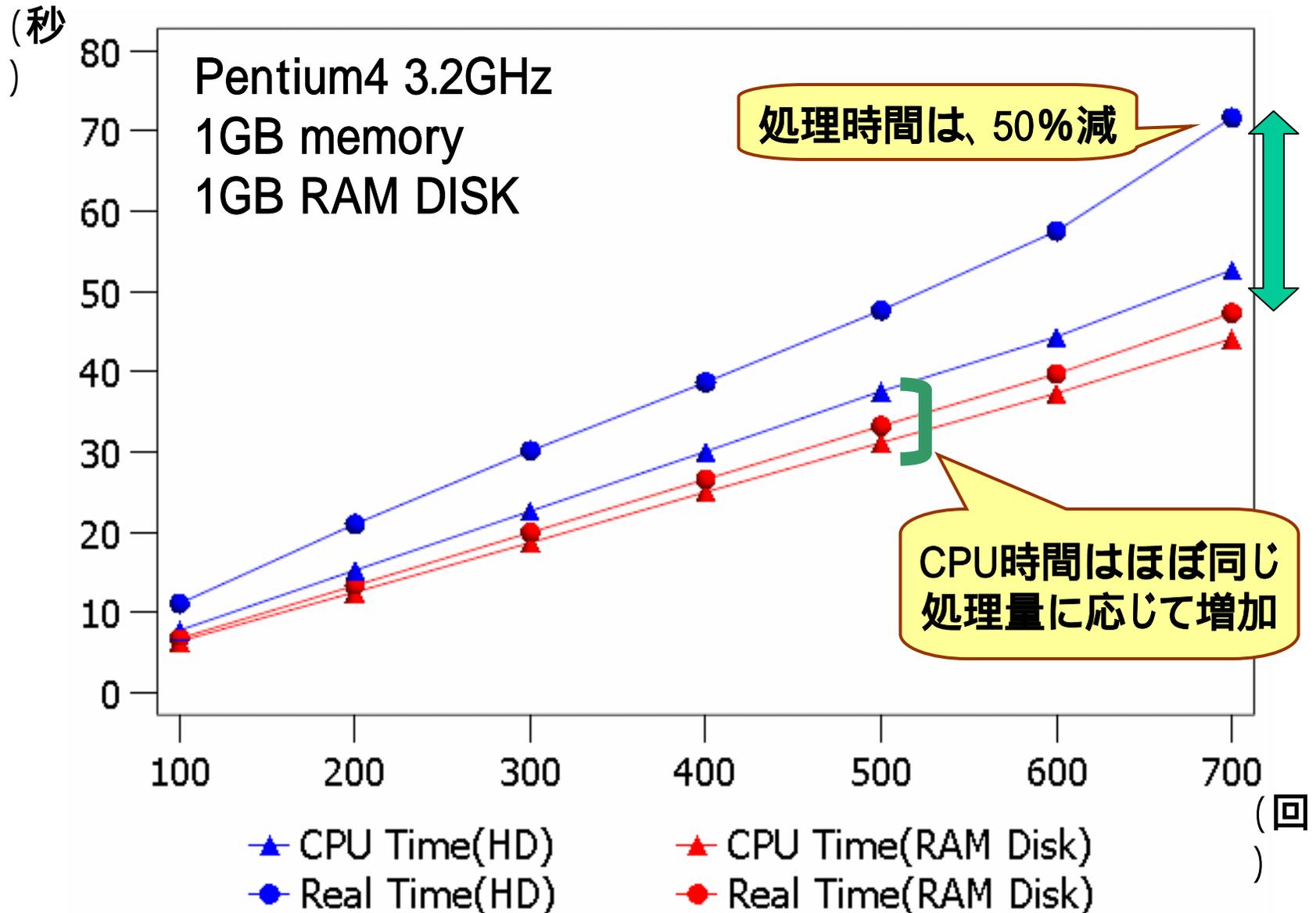
しかし、弊社のIBMマシンでは、Qsoft RAMDisk V9.0では、RAMDISKが128MBまでしか設定できなかった

そこで、RamDisk for Windows 2000を導入した

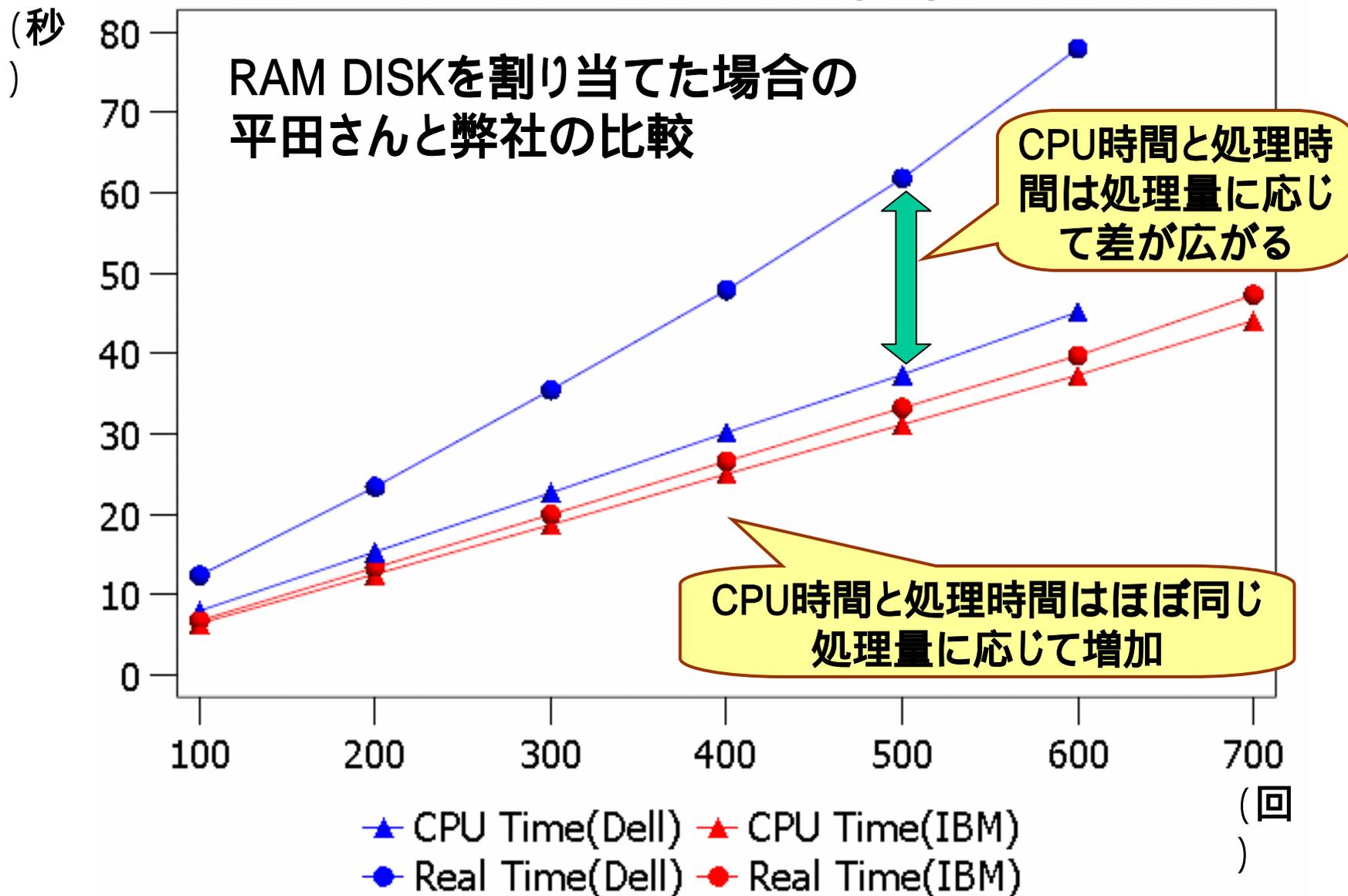
調査結果 - 岩本(1) -

- RAM DISKとしてRamDisk for Windows 2000使用
- 使用したPCスペック
 - IBM Desktop PC A50
 - Pentium4 2.6GHz, 256MB/512MBメモリ(SDRAM PC2700)
 - Pentium4 2.8GHz, 256MB/512MBメモリ(SDRAM PC2700)
 - Pentium4 3.2GHz, 2GBメモリ(SDRAM PC2700)
 - Windows2000 Profesional SP3
- 条件
 - 平田さんと同じ、ただしRAM DISKに1GBを割り当てた

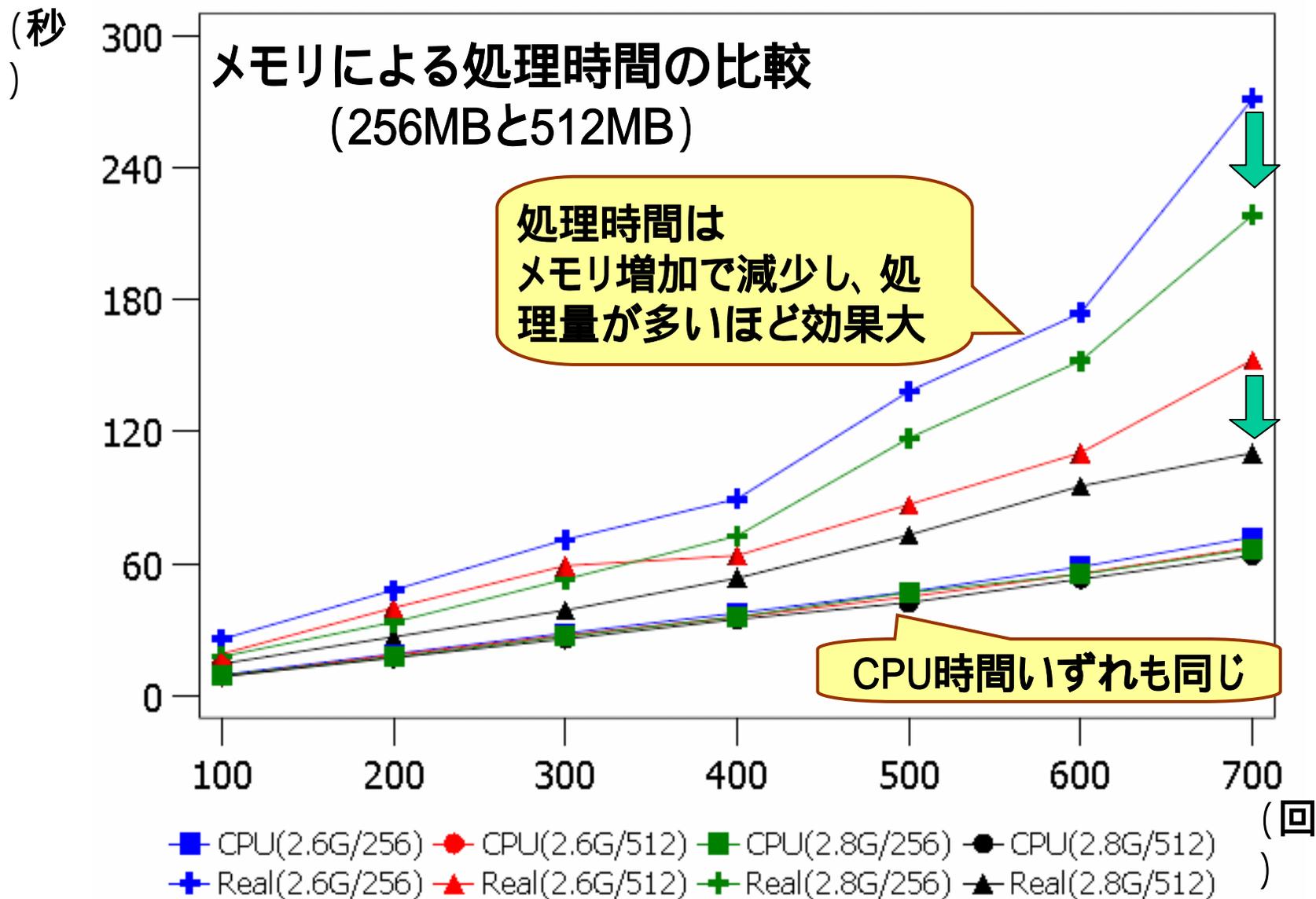
調査結果 - 岩本(2) -



調査結果 - 岩本(3) -



調査結果 - 岩本(4) -



まとめ

- RAM DISKの導入により、処理時間を短縮可能
- メインメモリも多い方が速い(500MB以上)
- CPUのクロック数は2.6GHz以上では差が少ない
 - 3.2GHz, 1.GBで処理時間:53秒
 - 2.8GHz, 512MBで処理時間:64秒
 - 2.6GHz, 512MBで処理時間:67秒
- ただし、CPUのメモリには限界(最大4GB?)があるので、大きなデータに対するものには不向きと思われる