PCクラスタワークショップ

日立のテクニカルコンピューティングへの取り組み

2010/2/19

株式会社　日立製作所
中央研究所
清水 正明

HITACHI
Inspire the Next

Copyright © Hitachi,Ltd.2008 All rights reserved

目 次

1 日立テクニカルサーバラインナップ

2 日立サーバラインナップ

3 物理乱数発生カードの紹介

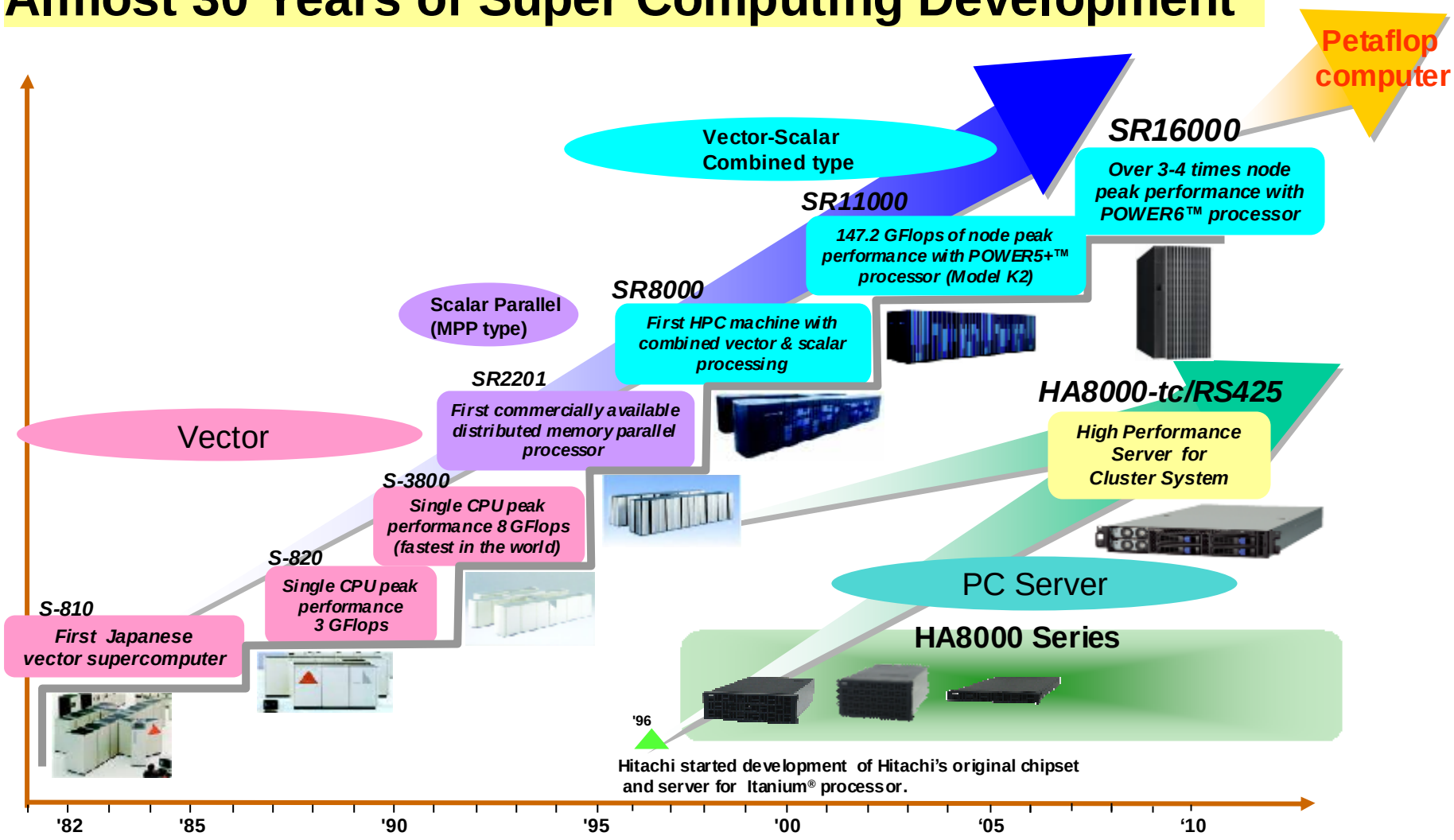
4 GPUコンピューティングへの取り組み

5 TOP500状況

Copyright © Hitachi,Ltd.2008 All rights reserved

1-1 日立テクニカルサーバ：History & Future

Almost 30 Years of Super Computing Development



POWER 6 is a trademark of IBM corporation.
POWER 5+ is a trademark of IBM corporation.

Copyright © Hitachi,Ltd.2008 All rights reserved

1-2 日立テクニカルサーバ　ラインアップ

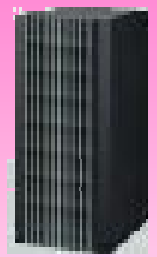
PCクラスタからスカラSMPまでラインアップ拡充

●最新POWER6™プロセッサ32way搭載SR16000
●Opteron™ 4ソケットプロセッサ
●今後の最新cpuにも対応予定

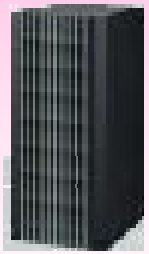
Opteron™
2.3/1.9GHz 16way
HA8000-tc/RS425

PCクラスタ

POWER6™
5.0GHz 64way
空冷ﾓﾃﾞﾙ
SR16000
VL1ﾓﾃﾞﾙ



POWER6™
4.7GHz 32way
水冷ﾓﾃﾞﾙ
SR16000
L2ﾓﾃﾞﾙ



POWER6™
3.5GHz 32way
空冷モデル
SR16000
L1ﾓﾃﾞﾙ

スカラSMPクラスタ

Copyright © Hitachi,Ltd.2008 All rights reserved

2

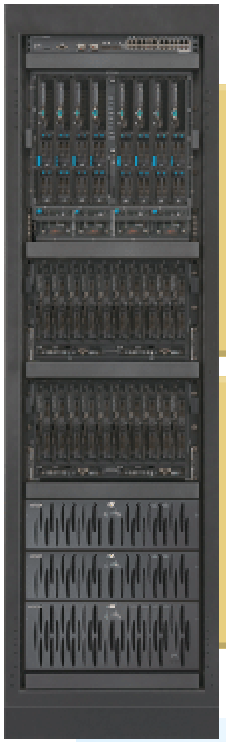
日立サーバラインナップ

・ブレードサーバ
・ラックマウントサーバ

Copyright © Hitachi,Ltd.2008 All rights reserved

2-2 BladeSymphonyラインアップ

規模や用途を広範囲にカバー。
単なる統合ではなく、全体最適化したシステムを提供します。



ターゲット	ブレードサーバ部	ストレージ部	ネットワーク部
Web3階層統合システム, 大規模データベース等	ハイエンドモデル BS2000	Hitachi Storage Solutions	CommuniMax ロードバランサ AX2500
PCサーバ統合 データセンター 中小規模の部門サーバ	小型高集積モデル BS320	BR1600 SMS100 BR20	

サーバ仮想化　日立サーバ仮想化機構 *Virtage*　VMware® ESX　Windows Server® 2008 Hyper-V™

かんたん導入モデル
BladeSymphony SP
導入後すぐに使える
ブレードサーバ

システム管理ソフトウェア　BladeSymphony Manage Suite

ミドルウェア　JP1　HiRDB　Cosminexus　主要ISV ソフトウェア

プラットフォームソリューション／サポートサービス

Copyright © Hitachi,Ltd.2008 All rights reserved

2-3 ハイエンドモデルBS2000の特徴

ミッションクリティカル・システムにフォーカ BS2000

■ 仮想統合を実現する高性能・高信頼ブレードサーバ

8ブレード/10U

- 仮想化による集約、高速処理に対応した性能・拡張性
  - インテル®Xeon®プロセッサー搭載、メモリ最大144GB
  - I/O性能：162Gbps(内部バス・単方向)
  - I/O拡張性・ シャーシ：内蔵スイッチ6台
    サーバブレード：拡張カードx2, PCI-Expressカードx2
    I/Oスロット拡張装置：PCI-Expressカードx16

- 日立サーバ仮想化機構 *Virtage* 標準搭載(*1)
- メインフレームの高信頼・高可用化技術を継承
  - Linux環境高信頼化, 稼動中の保守性, 保守作業容易性

- 世界最高基準の高効率電源
  - CSCI Gold基準適合, 80 PLUS® GOLD認証取得(*2)
- 長期サポート(ロングライフサポートサービス：7年)

Windows Server 2008　redhat.　vmware

*1：Essentialモデル
*2：電源負荷50%時の変換効率92%を実現

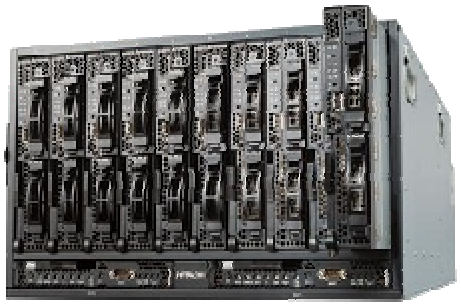
Copyright © Hitachi,Ltd.2008 All rights reserved

2-4 小型高集積モデルBS320の特徴

より軽く、より小さく　高密度実装を追求　BS320

■ 幅広い用途に対応する高集積・省電力ブレードサーバ

- 高さ6U(約27cm)に最大10ブレード搭載可能
- 最大重量約98kg/シャーシの軽量設計
- 用途に応じた多彩なサーバブレードをラインアップ
- 日立サーバ仮想化機構 *Virtage*に対応(*1)
- 高効率電源 (CSCI Silver基準適合, 80 PLUS® SILVER認証取得(*2))
- 長期サポート (ロングライフサポートサービス：7年)



標準サーバブレード　SAN専用サーバブレード　HDD拡張サーバブレード　PCI拡張サーバブレード　ロードバランサブレード

Windows Server 2008　redhat.　vmware

*1：PCI拡張サーバブレード*Virtage*モデルで提供
*2：電源負荷50%時の変換効率89%以上を実現

Copyright © Hitachi,Ltd.2008 All rights reserved

2-5 HA8000製品ラインアップ

(4Processor)

HA8000/RS440 Xeon(X7460/E7450/E7450 /E7220)

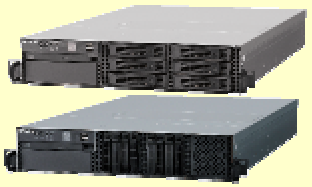
2コア 4コア 6コア
SAS 2.5型HDD
日立製RAID
LANポート

xJモデル

(2Processor)

HA8000/RS220
HA8000-es/RS220

Xeon(X5570/E5540/E5520/E5502)
Xeon(L5520)*2



2コア 4コア
SAS 3.5/2.5型HDD
SATA 3.5型HDD
日立製RAID
オンボードSVP
消費電力表示
WSSモデル

HA8000/TS20

Xeon(X5570/E5540/E5520/E5502)

2コア 4コア
SAS 3.5/2.5型HDD
SATA 3.5型HDD
日立製RAID
静音(全モデル)
オンボードSVP
消費電力表示

HA8000/RS210
HA8000-es/RS210

Xeon(X5570/E5540/E5520/E5502)
Xeon(L5520)*2

2コア 4コア
SAS 3.5/2.5型HDD
SATA 3.5型HDD
日立製RAID
オンボードSVP
消費電力表示

xJモデル

(1Processor)

HA8000/TS10 Xeon(X3430/X3460 /X3470

2コア 4コア
SAS 3.5型HDD
SATA 3.5型HDD
日立製RAID
静音
WSSモデル

HA8000/RS110 Xeon(X3430/X3460 /X3470)

2コア 4コア
SATA 3.5型HDD
WSSモデル

*1:HA8000のみ *2:HA8000-esのみ

Copyright © Hitachi,Ltd.2008 All rights reserved

3

物理乱数発生カードの紹介

Copyright © Hitachi,Ltd.2008 All rights reserved

3

物理乱数発生カードの用途：

・モンテカルロシミュレーション等、シミュレーション
　において高品質な乱数が求められる。

利点：

・ＣＰＵに負荷を与えず乱数を生成可能

Copyright © Hitachi,Ltd.2008 All rights reserved

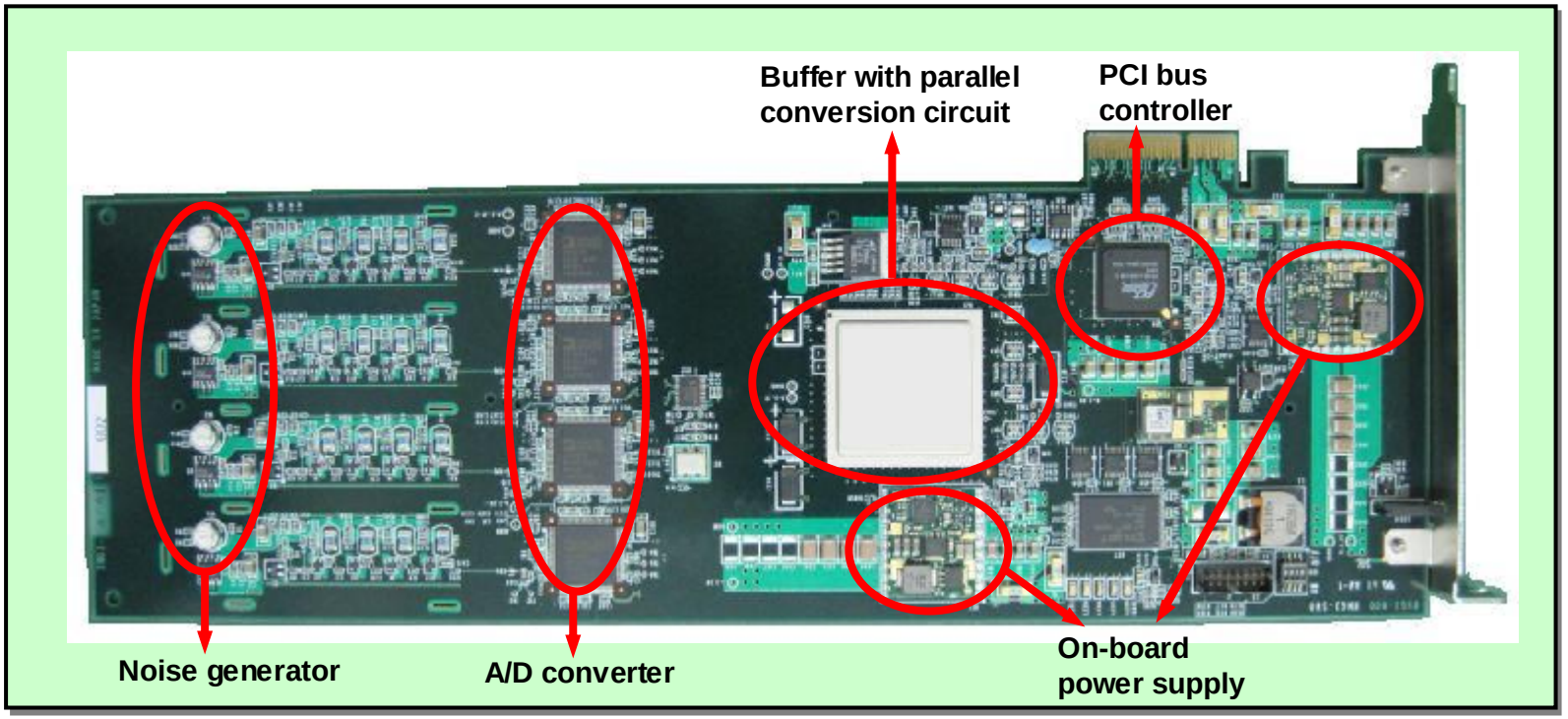
3-1 物理乱数発生カード

- 日立SRシリーズ用に物理乱数カードを独自開発
- 第三世代ではノードに最大4枚を搭載可能とし生成性能を向上

	第一世代 (1999年1月出荷)	第二世代 (2004年12月出荷)	第三世代 (2010年1月出荷予定)
搭載サーバ	SR8000	SR11000	SR16000
理論乱数生成速度	25MB/s	200MB/s	200MB/s (単体), 400MB/s (ノード)
乱数源, 乱数生成方式	白色ノイズ, A/D変換	白色ノイズ, A/D変換 + 並列	白色ノイズ, A/D変換 + 並列
物理インターフェース	PCI 2.1準拠	PCI 2.2準拠 (64bit, 66MHz)	PCI Express x4
ノード搭載枚数	1セット(2枚)	1枚	1～4枚

Copyright © Hitachi,Ltd.2008 All rights reserved

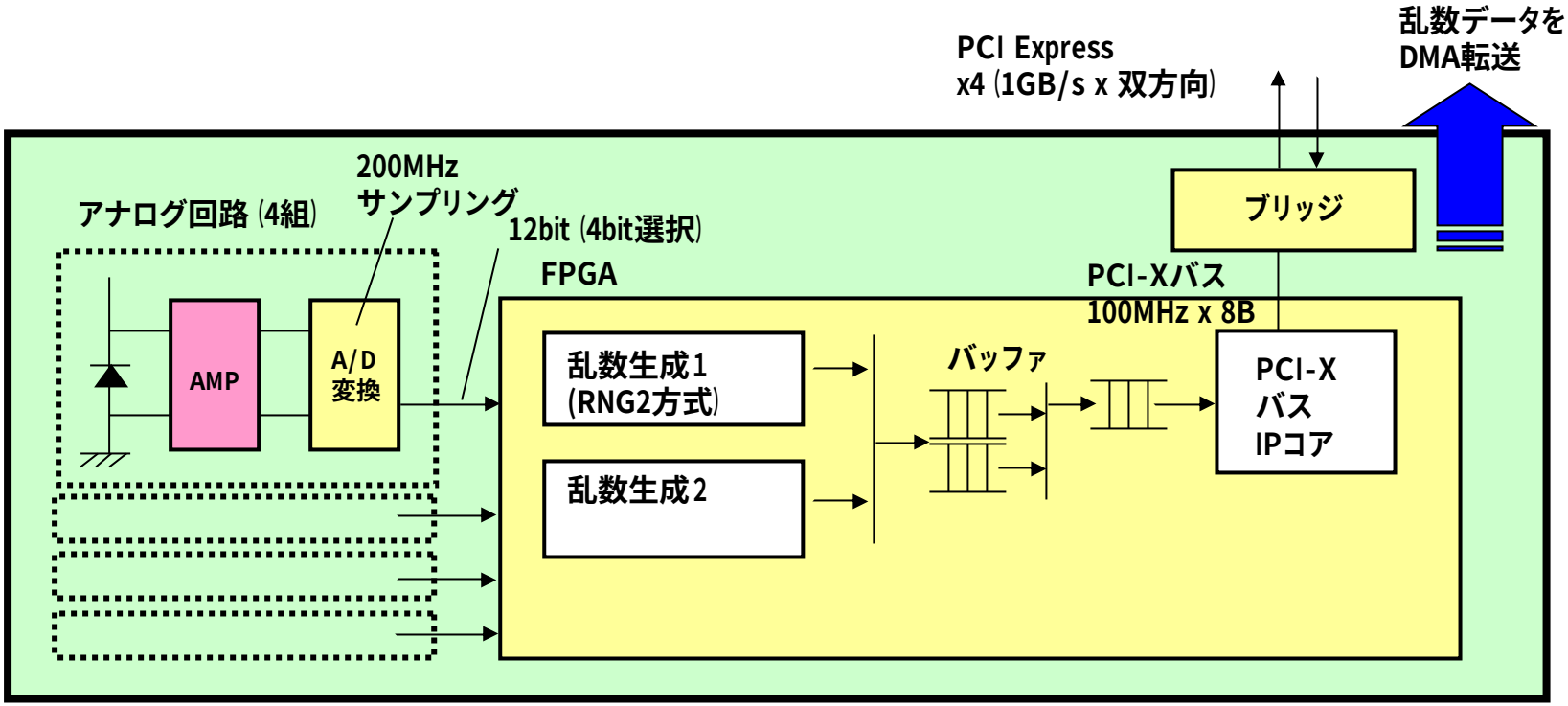
3-2 新 (第三世代) 物理乱数カードの概観



- 乱数カード・ホストI/F：PCI-Exp 4x
- 物理乱数発生源：　白色ノイズ (熱雑音)
- 理論乱数生成性能：200MB/s/カード，400MB/s/ノード

Copyright © Hitachi,Ltd.2008 All rights reserved

3-3 新（第三世代）物理乱数カードの論理構成図

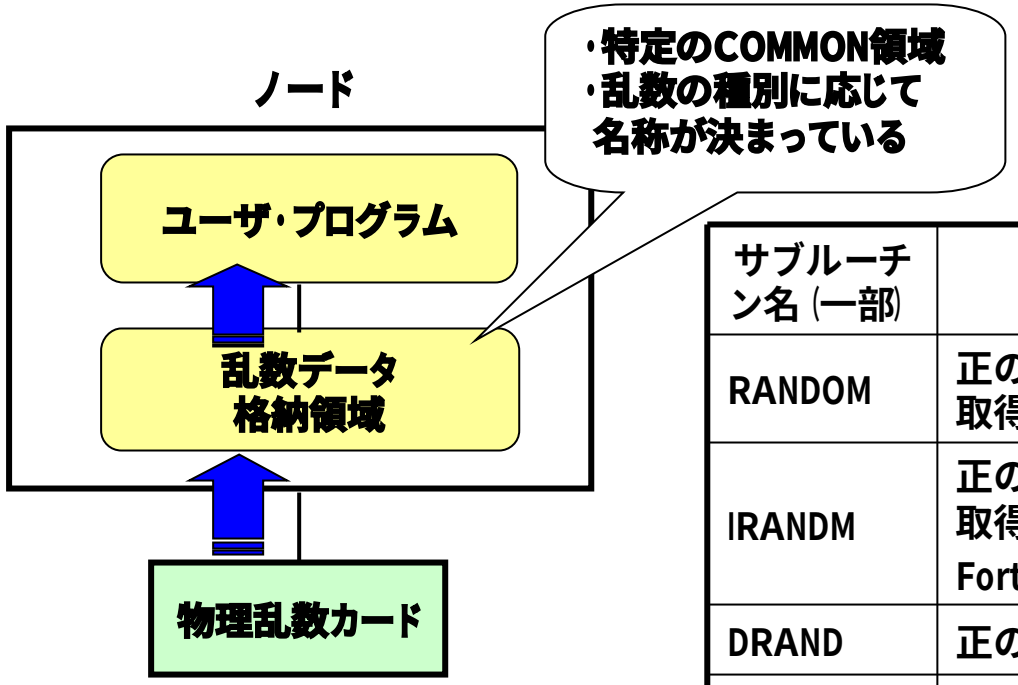


- ダイオード電圧の熱雑音をA/Dコンバータでサンプリング
- 独立した4系統の生成源を統合し乱数生成

Copyright © Hitachi,Ltd.2008 All rights reserved

3-4 物理乱数カードの使い方（プログラムI/F）

- Fortranのユーザ・インタフェース（サブルーチン、関数）を提供
  - ユーザは乱数データの個数と型を指定するだけ
  - 利用可能データ型は単精度整数、単精度実数、倍精度実数



サブルーチン名（一部）	機能
RANDOM	正の整数乱数（整数型）を指定した領域に**1個**取得
IRANDM	正の整数乱数（整数型）を指定した領域に**N個**取得 FortranのCOMMON領域名：**IRANAR**
DRAND	正の実数乱数（倍精度型）を**戻り値**として取得
DURANDM	正の実数乱数（倍精度型）を指定した領域に**N個**取得　　* Nは最大 3,000,000個 FortranのCOMMON領域名：**DURANAR**

Copyright © Hitachi,Ltd.2008 All rights reserved

4

GPUコンピューティングへの取り組み

Copyright © Hitachi,Ltd.2008 All rights reserved

4-1 最新GPUの動向

HPC向け機能の拡充

- 倍精度性能の飛躍的向上
- メモリエラー訂正機能(ECC)のサポート
- GPU機種に依存しないヘテロジニアスプログラミング環境OpenCLのサポート

	NVIDIA Tesla20 S2070	AMD/ATI Radeon HD5970
性能	550～625GFLOPS×4(倍精度)	2.32TFLOPS×2(単精度) 464GFLOPS×2(倍精度)
メモリ	GDDR5, 6GB×4, ECCサポート	GDDR5, 1GB×2, 256GB/s
電力	900W(typ)	Idle 51W, max 294W
プログラミング環境	C for CUDA, CUDA Fortran(PGI) 自動GPUコンパイラ(PGI) OpenCL DirectX 11 統合開発環境Nexus	Stream SDK(Brook+, CAL) OpenCL DirectX 11 開発環境KernelAnalyzer
その他	IEEE 754-2008準拠	

Copyright © Hitachi,Ltd.2008 All rights reserved
17

4-2 GPGPUのメリット・デメリット

メリット・デメリット		内容
メリット	高演算性能	単精度1~3TFLOPS，倍精度数百GFLOPS
	高メモリバンド幅	150~250GB/s
	低価格	HPC向けGPUで~30万円，グラフィクス向けGPUで~数万円
	信頼性向上の動き	メモリエラー訂正機能サポート(NVIDIA)
	プログラミング環境 汎用化の動き	OpenCL（GPU,Cell，マルチコアCPUの共通言語仕様） PGIコンパイラ（GPUコード自動生成機能）
デメリット	プログラムの互換性	プログラム書き換え必須， 性能をひきだすための独特チューニング要
	メモリ容量	GPUあたり1～数GBで大規模科学技術計算には不足
	ホストCPUとのデータ 転送による性能低下	プログラムの全データをGPUに保持して処理できれば 10~100倍加速， プログラムの一部データをGPUに転送して処理すると 加速率は10%～3倍　(*)

(*)東京工業大学　青木尊之教授講演より
http://www.kgt.co.jp/avs_conso/event/vc15/summary/data/2-4p.pdfほか

Copyright © Hitachi,Ltd.2008 All rights reserved

4-3 日立の研究開発へのGPGPU適用可能性

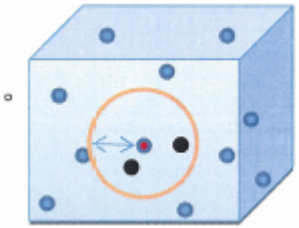
計算科学を用いた日立の研究開発分野

- 材料物性・ナノシミュレーション
- 機械（熱流体，構造，振動）
- 電磁場
- ライフサイエンス

Copyright © Hitachi,Ltd.2008 All rights reserved

4-4 GPGPU性能評価

⑴ 分子動力学

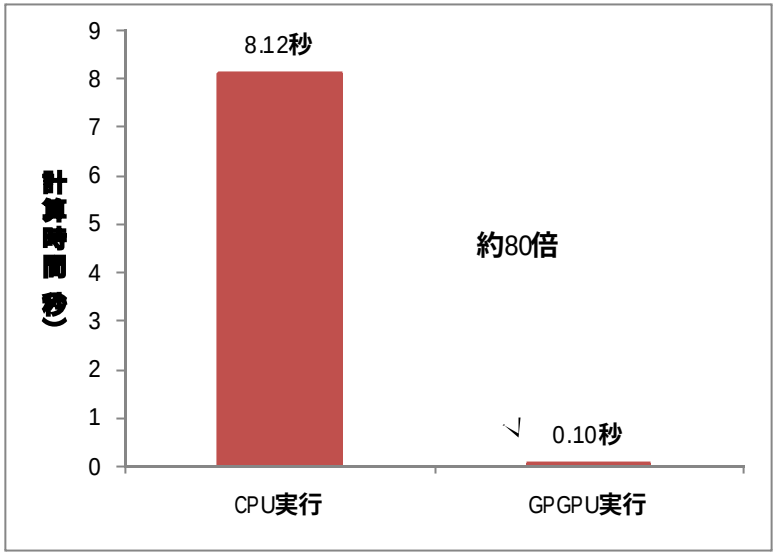


評価対象：隣接粒子リスト作成処理
　全粒子間のすべての組み合わせに対して距離を計算し，
　カットオフ半径以下のものについてポテンシャル計算の対象とする
　27,343粒子，領域サイズ120x120x70Å，カットオフ半径5.2Å

実行環境：
- Core i7(2.67GHz)+GeForce GTX285
- CUDA　2.2

実行結果
- CPU実行：8.12秒，0.8GFLOPS
- GPU実行：0.10秒，64.9GFLOPS

約80倍高速化



Copyright © Hitachi,Ltd.2008 All rights reserved

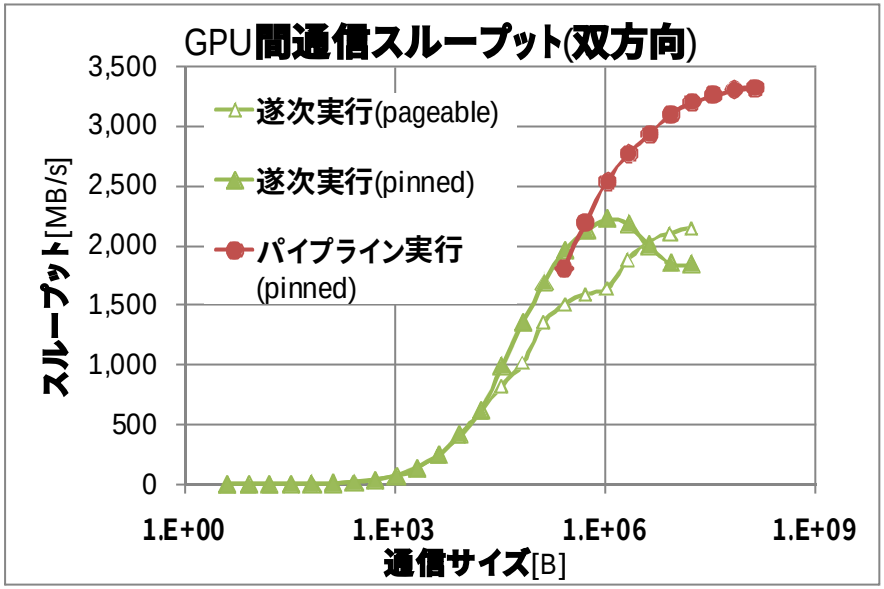
4-5 GPGPU性能評価

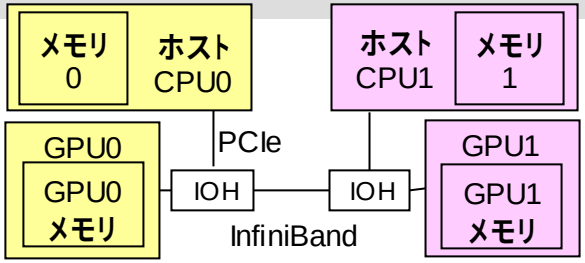
(2) 並列GPU

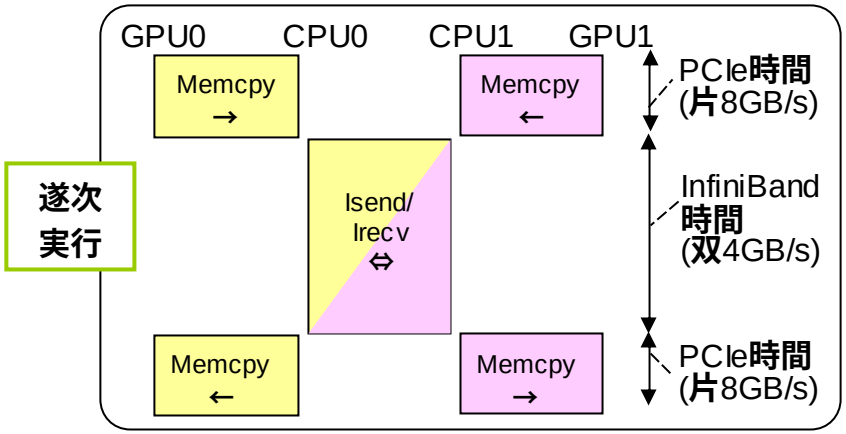
GPU間通信スループットの評価
　複数CPUノード+複数GPUでMPI並列
　データ交換スループットを評価

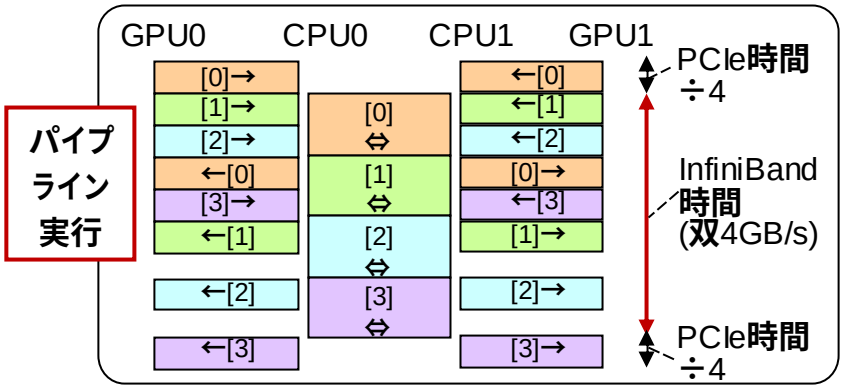
実行環境
- (Core i7+GTX280)+(Core i7+GTX295)+ InfiniBand(x4 DDR)
- CUDA 2.2, MVAPITCH2

実行結果
- パイプライン実行により3.3GB/sに向上









Copyright © Hitachi,Ltd.2008 All rights reserved

4-6 今後の取り組み

ノウハウの蓄積：
- 各種のHPCアプリケーションのGPGPU適合性を検証
- CUDA，PGIコンパイラ(指示行ベース)，OpenCLを用いた最適化技術を蓄積
- 並列GPU評価の推進

製品展開，ソリューション展開：
- PCクラスタ製品の一展開として推進
  （ご相談ください）

Copyright © Hitachi,Ltd.2008 All rights reserved

5

TOP500状況

Copyright © Hitachi,Ltd.2008 All rights reserved

23

5-1 TOP500状況 http://www.top500.org/

東京大学情報基盤センター殿システム

Linpack測定値；82.98Tflops

理論ピーク；113Tflops = (512+256) x 147.2

効率；73.4%

(2009/11発表)

Linpack測定値；101.74Tflops

理論ピーク；139Tflops = (944) x 147.2

効率；73.1%

Copyright © Hitachi,Ltd.2008 All rights reserved

2009/11 発表の性能

HITACHI
Inspire the Next

2008/11 発表の性能 → 2009/11

- 4SMP BLAS (Goto r25) →(Goto r26)
  N=10,000 32.07Gflops 87.1%
  33.15Gflops 90.1%
- LINPACK (512+256 : 113Tflops) → (944 : 139Tflops)
  N=1,433,600 P=32,Q= 96 4SMP-3072MPIs
  N=1,740,800 P=32,Q=118 4SMP-3776MPIs
  Block size NB=160 , Bcast = 1ring

  82.98 Tflops ($6^{hr}34^{min}30^{sec}$) 73.4%
  101.74 Tflops ($9^{hr}36^{min}07^{sec}$) 73.1%

Copyright © Hitachi,Ltd.2008 All rights reserved

HITACHI

Inspire the Next

Copyright © Hitachi,Ltd.2008 All rights reserved