マッスル・サーバー
(汎用PCクラスタ
＋特定計算向けハードウェア)
の開発
～分子軌道法を例にして～

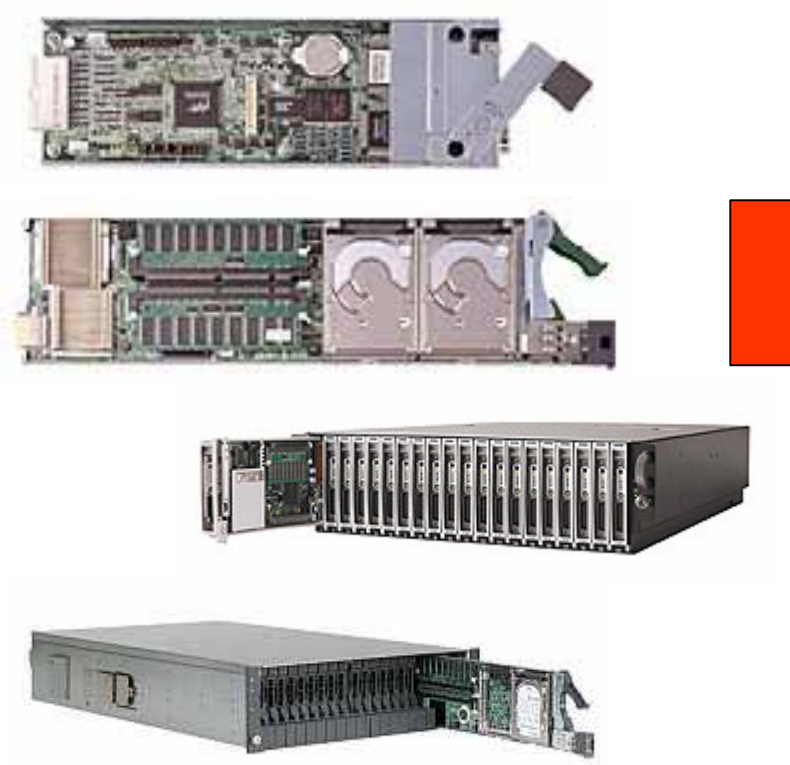
村上和彰

九州大学　情報基盤センター

murakami@cc.kyushu-u.ac.jp

マッスル・サーバー (muscle server) とは？

- 【関連語】ブレード・サーバー (blade server)
  - 高集積型PCサーバー

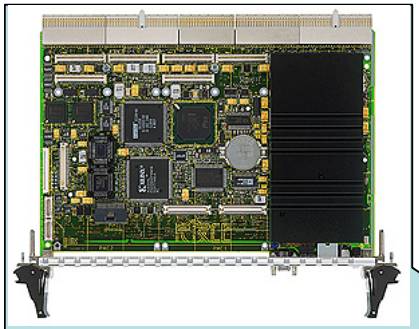


- 【定義】汎用PCクラスタ＋特定計算向けハードウェア

マッスル・サーバー開発例

- EHPC/Eric (Embedded HPC with Eric)
  - 文部科学省・科学技術振興調整費・総合研究「科学技術計算専用ロジック組込み型シミュレータに関する研究」(平成12年度～16年度)
    - 研究代表者：村上和彰 (九州大学)
    - 参加研究機関：九州大学，東京大学，産総研，富士総研，セイコーエプソン，アプリオリ・マイクロシステムズ
  - 構成
    - Compact PCI規格シャーシー
      - Compact PCI規格PC互換ボード
      - Compact PCI規格SH-4マルチプロセッサ・ボード
      - Compact PCI規格二電子積分計算加速ボード (SH-4＋Eric搭載)
    - Eric (二電子積分計算専用プロセッサ)

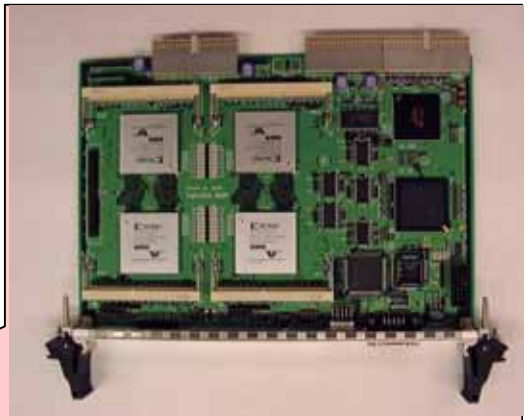
分子軌道法専用マッスル・サーバー
EHPC/Eric



Compact PCI規格
PC互換ボード
(P-II×1)

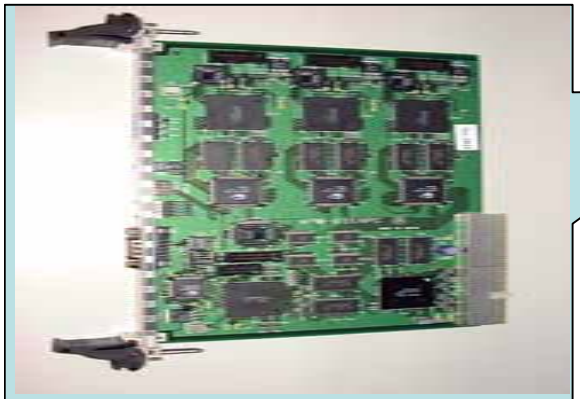
1枚／Compact PCIシャーシー

Compact PCIシャーシー×4



Compact PCI規格
二電子積分計算加速ボード
(SH-4×1 + Eric×4)

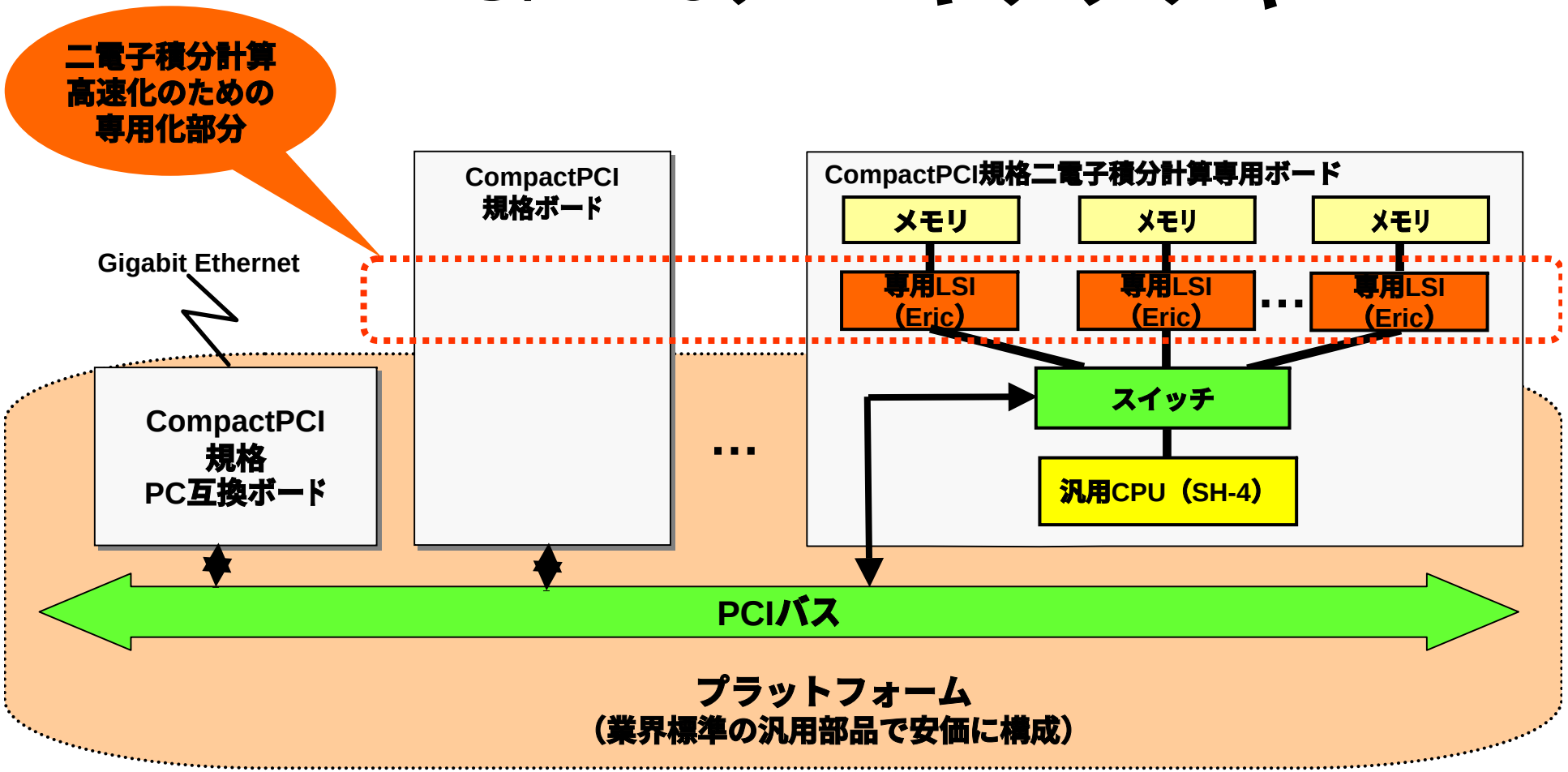
最大7枚／Compact PCIシャーシー



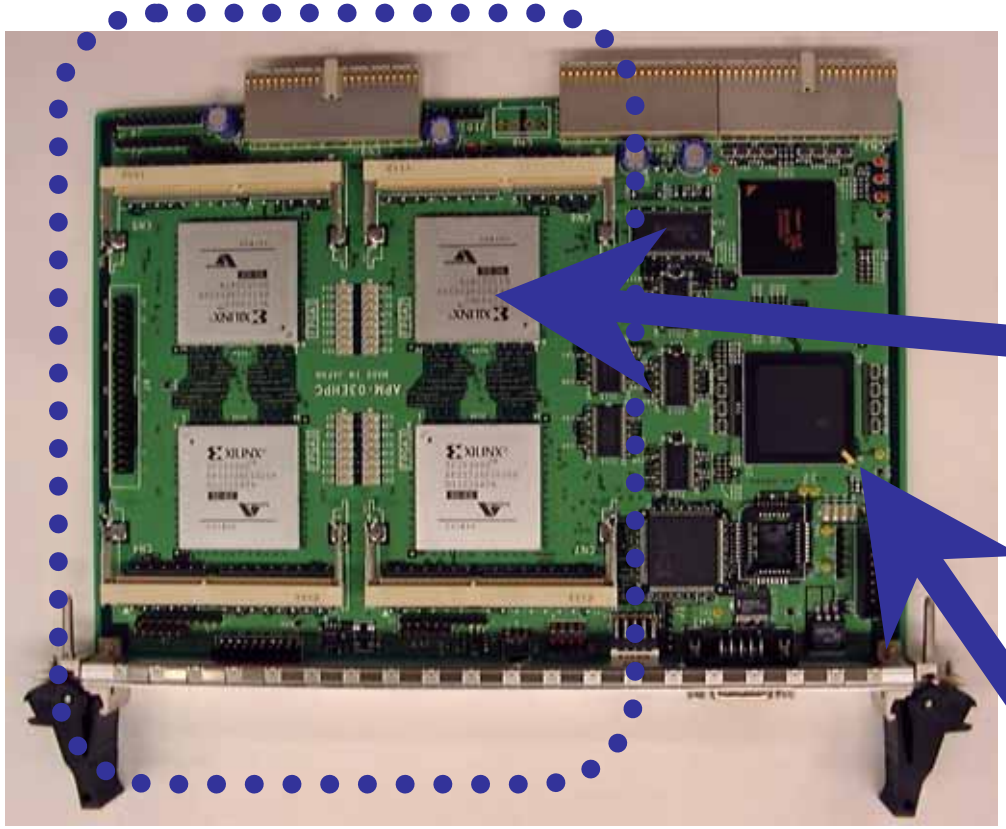
Compact PCI規格SH-4
マルチプロセッサ・ボード
(SH-4×4)

(最大7枚／Compact PCIシャーシー)

EHPC/Ericアーキテクチャ



Compact PCI規格
二電子積分計算加速ボード



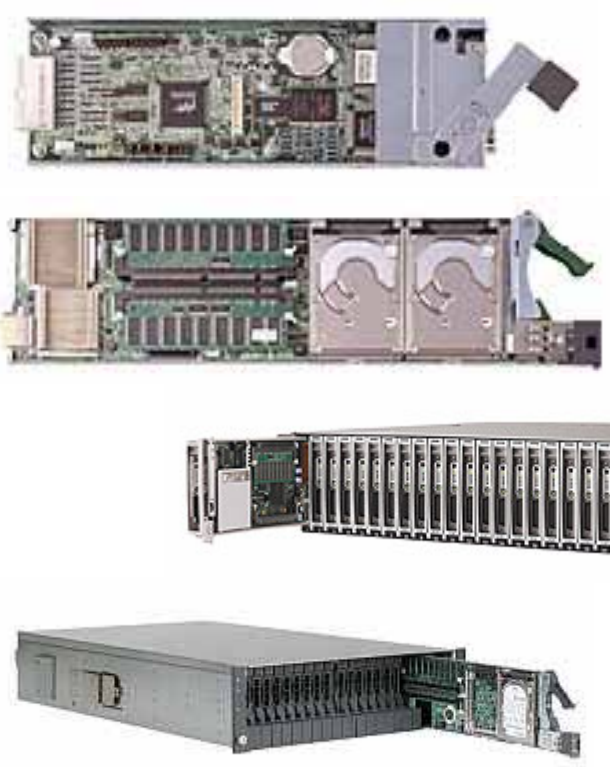
- Compact PCI規格に準拠したプリント基板
- Eric (二電子積分計算専用プロセッサLSI) ×4
- SDRAM
  - 各Eric当り1GB
- 汎用MPU (SH4) ×1
- PCIバスI/F, Ethrenet, 等

マッスル・サーバー vs. 他のHPCアーキテクチャ

	ベクトル プロセッサ	MPP	SMPクラスタ	PCクラスタ (ブレード・サーバー)	マッスル・サーバー
プロセッサ	専用ベクトルプロセッサ	汎用高性能マイクロプロセッサ	汎用高性能マイクロプロセッサ	汎用高性能マイクロプロセッサ	汎用／組込みマイクロプロセッサ＋特定計算向けハードウェア
メモリシステム	専用	専用	専用	汎用	汎用
ネットワーク	専用	専用	• ローカル：専用 • グローバル：専用または汎用	汎用 (Ethernet, etc.)	汎用 (Compact PCI, Ethernet, etc.)
高速化技術	• ベクトル処理 • 並列処理 • 高速通信	• 並列処理 • 高速通信	• スレッド並列処理 • 並列処理 • 高速通信	• 高クロック周波数 • 並列処理	• 特定計算向けハードウェア • 並列処理

マッスル･サーバー向きの応用は？

- 汎用PC
  - 各種雑多な処理
- 特定計算向けハードウェア
  - “Compute Intensive”な処理を担当



マッスル･サーバーの性能
～アムダールの法則～

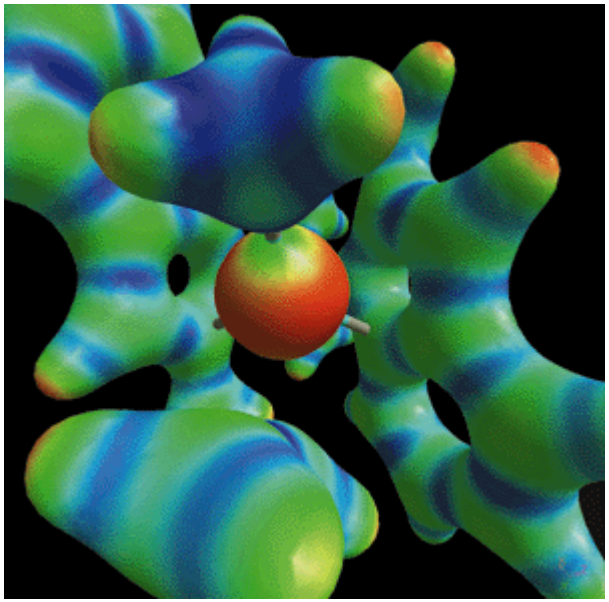
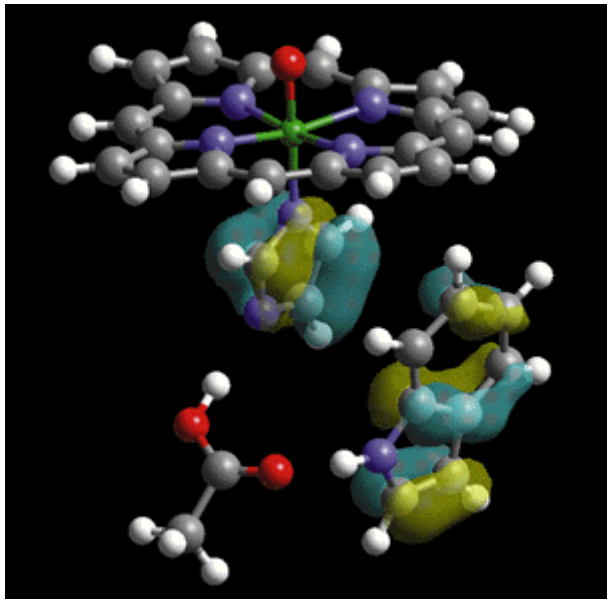
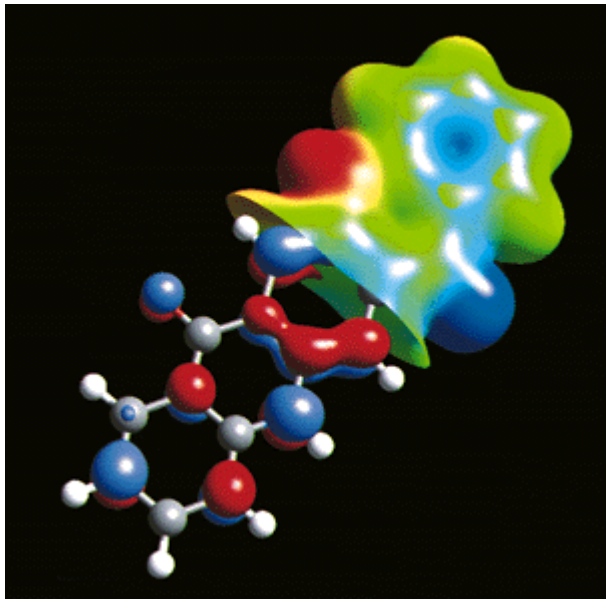
- p：**並列化可能な処理の割合** $(0 \leq p \leq 1)$
- o：**特殊演算向けハードウェアにオフロード可能な処理の割合** $(0 \leq o \leq 1)$
- N：**マシン並列度**
- H：**特定計算向けハードウェアによる性能向上率**
- S：**マッスル･サーバーの単体PCに対する性能向上率**

$$S = \frac{1}{(1-p) + \frac{p}{N}\left\{(1-o) + \frac{o}{H}\right\}}$$

マッスル･サーバーの分子軌道法への応用

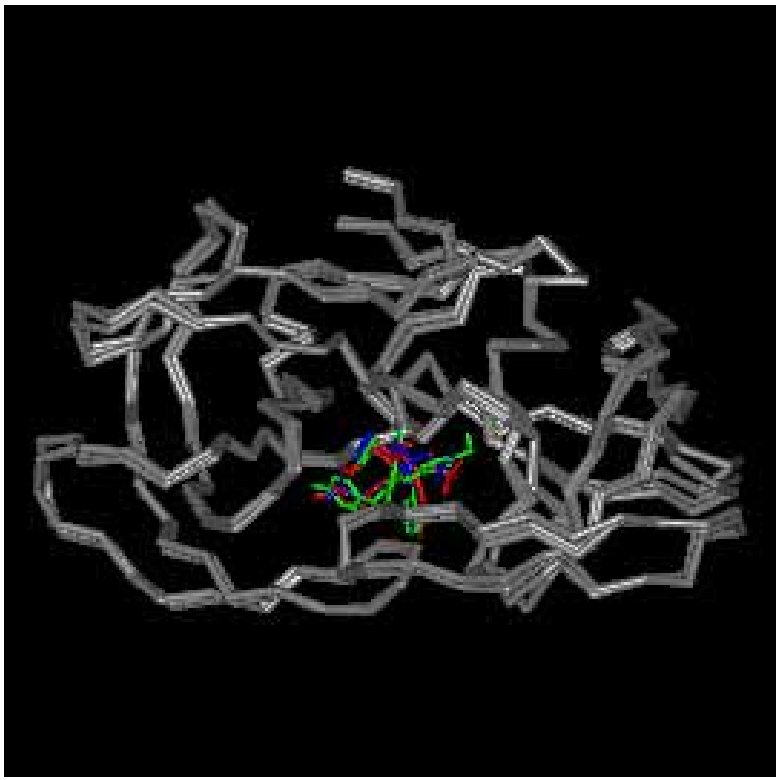
分子軌道法

- 電子が分子（を構成する原子核）の周りでどのような分布状況にあって，どのようなエネルギーを持っているかを計算により求める
  - 創薬
  - 材料開発
  - etc.

分子軌道法高速化の必要性

- **創薬，材料開発のためには大規模分子の電子状態計算を数分のオーダーで実行する必要がある**



HIV-1 protease

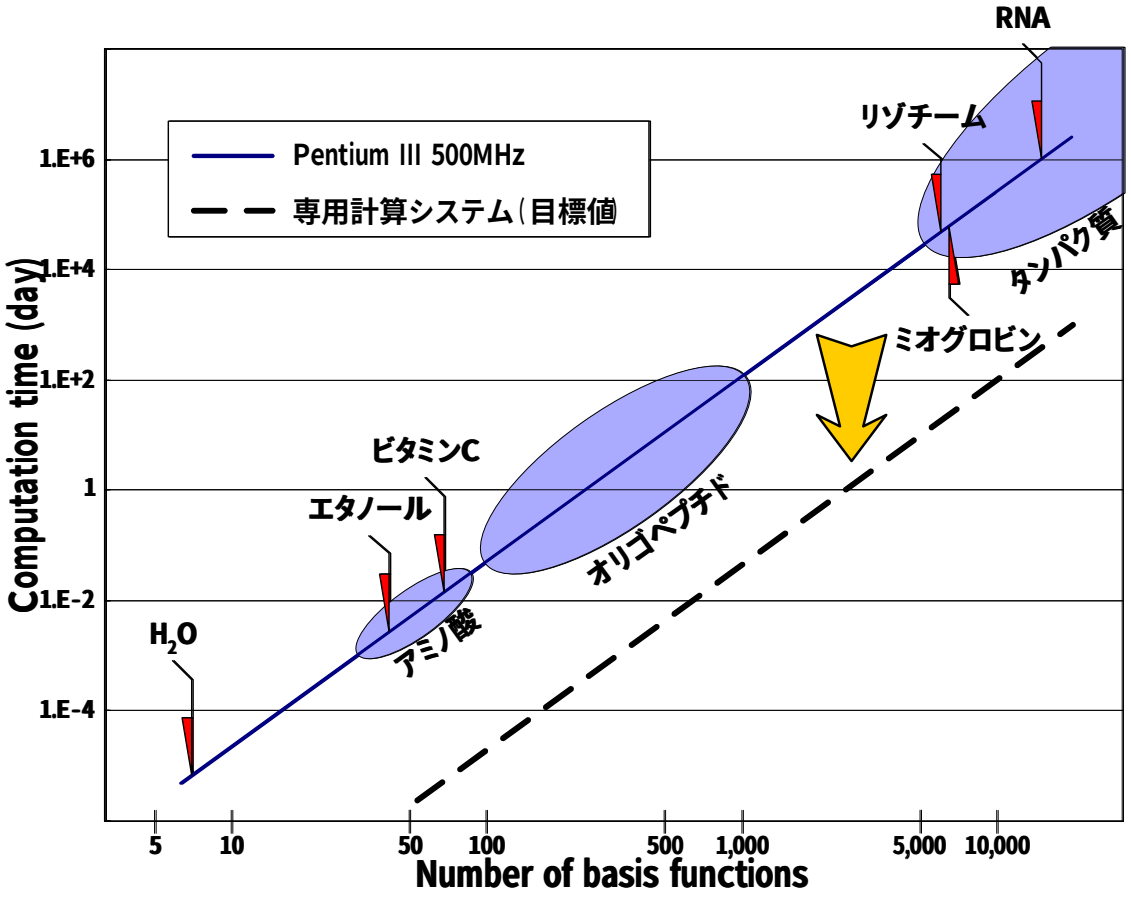
<u>現在</u>

PC**クラスタ** (Xeon (2.8GHz)×80，**主記憶** 512GB/processor) **で約2時間** (**基底関数** STO-3G)

<u>将来</u>

1分以下で計算できるようにしたい！

非経験的分子軌道計算の計算時間

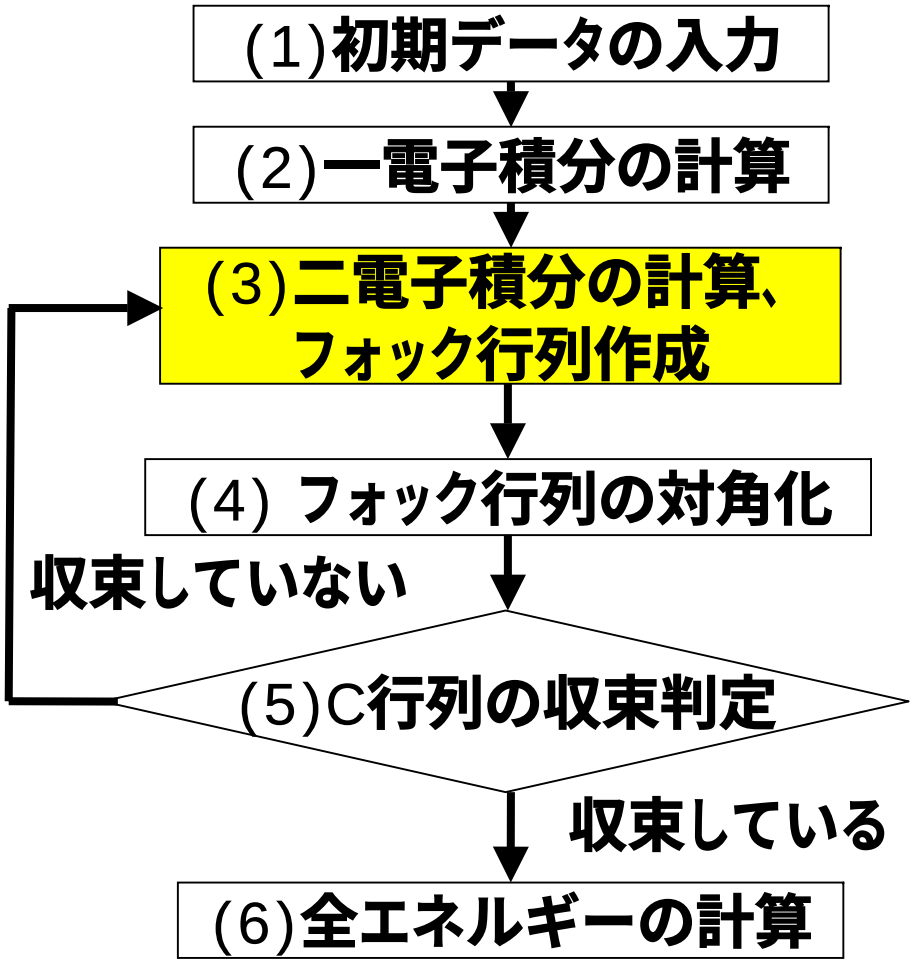


- **膨大な計算時間**
  - **低分子量タンパク**
    10^4日　27年
  - RNA
    10^6日　2740年

現実と理想のギャップ

- **ユーザー**
  「1週間～10日は待てる！」
- **5000基底を10日で！**

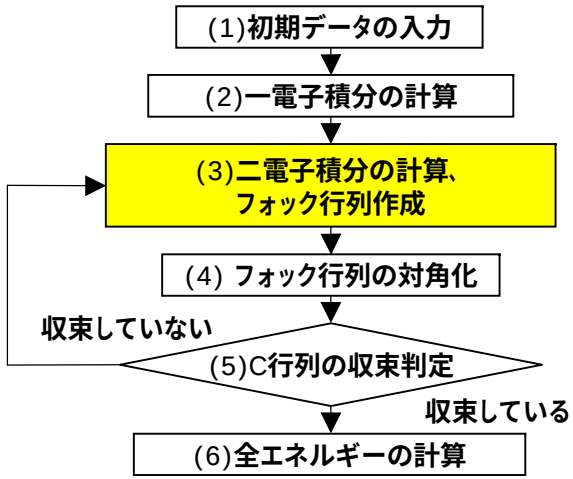
非経験的分子軌道法の計算手順



$$F_{IJ} = T_{IJ} + V_{IJ} + \sum_K \sum_L P_{KL} \left\{ (IJ, KL) - \frac{1}{2}(IK, JL) \right\}$$

$$\sum_{I=1}^{N} F_{IJ} C_{aI} = \sum_{I=1}^{N} S_{IJ} C_{aI} \varepsilon_a$$

非経験的分子軌道計算の計算時間



➔ 全計算時間の98%は二電子積分計算とフォック行列生成に費やされる！

➔ 二電子積分計算専用LSIを開発！

分子 (ペプチド分子)		G	GA	GAQ	GAQM	GAQMY
原子数		10	20	37	58	75
基底関数の数		55	110	207	316	427
計算時間（秒）	初期データの入力	0.1	0.6	4.4	18.9	57.3
	一電子積分の計算	0.1	0.3	1.5	5.0	10.1
	二電子積分の計算 フォック行列作成	**22.9 (96.6%)**	**269.4 (98.7%)**	**1871.0 (98.6%)**	**8482.1 (98.8%)**	**23284.3 (98.6%)**
	フォック行列の対角化	0.2	1.7	11.0	60.9	211.7
	全エネルギーの計算	0.1	0.3	2.2	9.15	27.5
	Total	23.7	272.9	1892.7	8584.9	23614.5

4-31G Basis, GAMESS, PentiumIII500MHz, 512MB

二電子積分計算の特徴

```
for(I = 0;  I <  Nshell; I++)
  for( J = 0; J < I;  J++)
    for (K = 0;  K < I; K ++)
      for( L = 0; L < I; L++)
        for(i = 0;  i <  Ni; i++)
          for(j = 0; j < Nj;  j++)
            for (k = 0;  k < Nk; k++)
              for(l = 0; l < Nl; l++)
                <si sj sk sl> の計算
              forend
            forend
          forend
        forend
        <aI aJ aK aL>の計算(漸化計算部分)
      forend
    forend
  forend
forend
```

初期積分計算部分

- 小原法を基に「新小原法」を開発
- 初期積分計算部分
  - ４重ループ構造
  - 並列性が低い
  - 複雑な倍精度浮動小数点演算を含む
    - 除算
    - 開平逆数演算
    - 指数関数演算
- 漸化計算部分
  - 並列性が高い
  - 多数の積和演算からなる

Eric：二電子積分計算専用LSI
－設計方針－

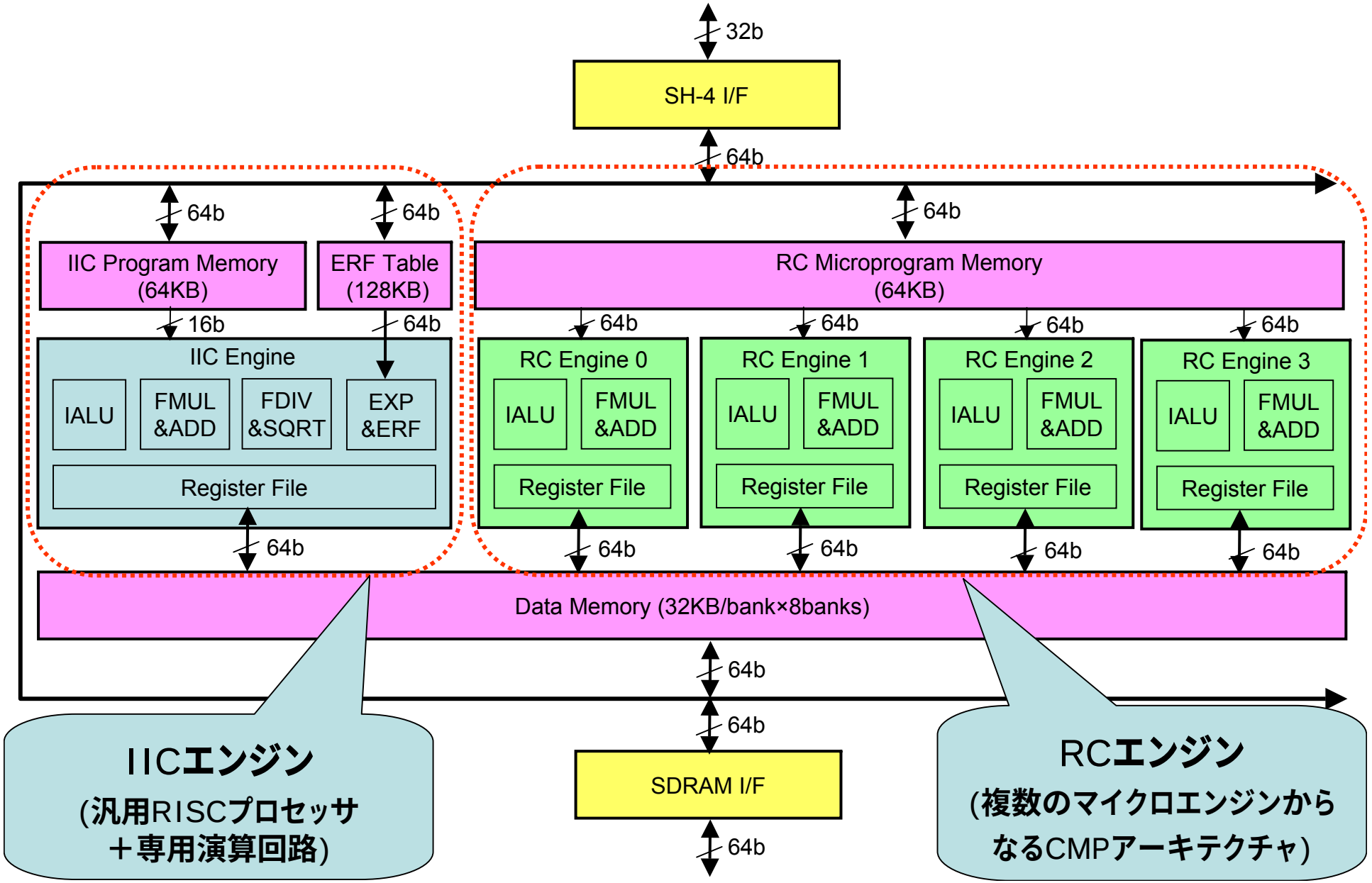
	内在する演算	並列度
初期積分計算	• 浮動小数点加減乗除算 • 開平逆数演算 • 指数関数演算 • 誤差関数計算	低
漸化計算	• 積和演算	高

専用演算器を用いて高速化

多数の積和演算器を搭載し並列性を活用して高速化

専用LSIを2種類のエンジンに分割:
- 初期積分計算 (IIC) エンジン
- 漸化計算 (RC) エンジン

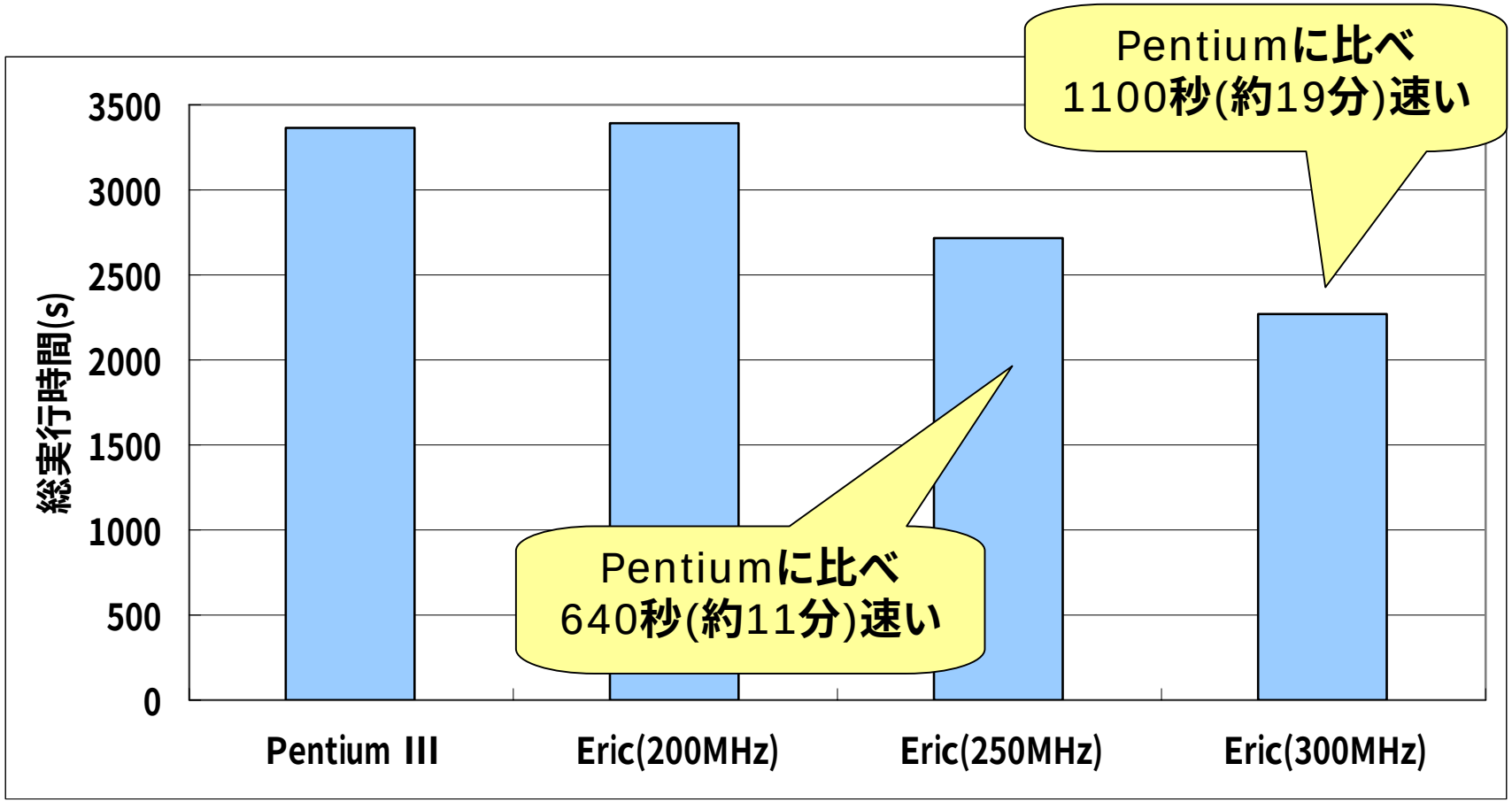
Eric:二電子積分計算専用LSI



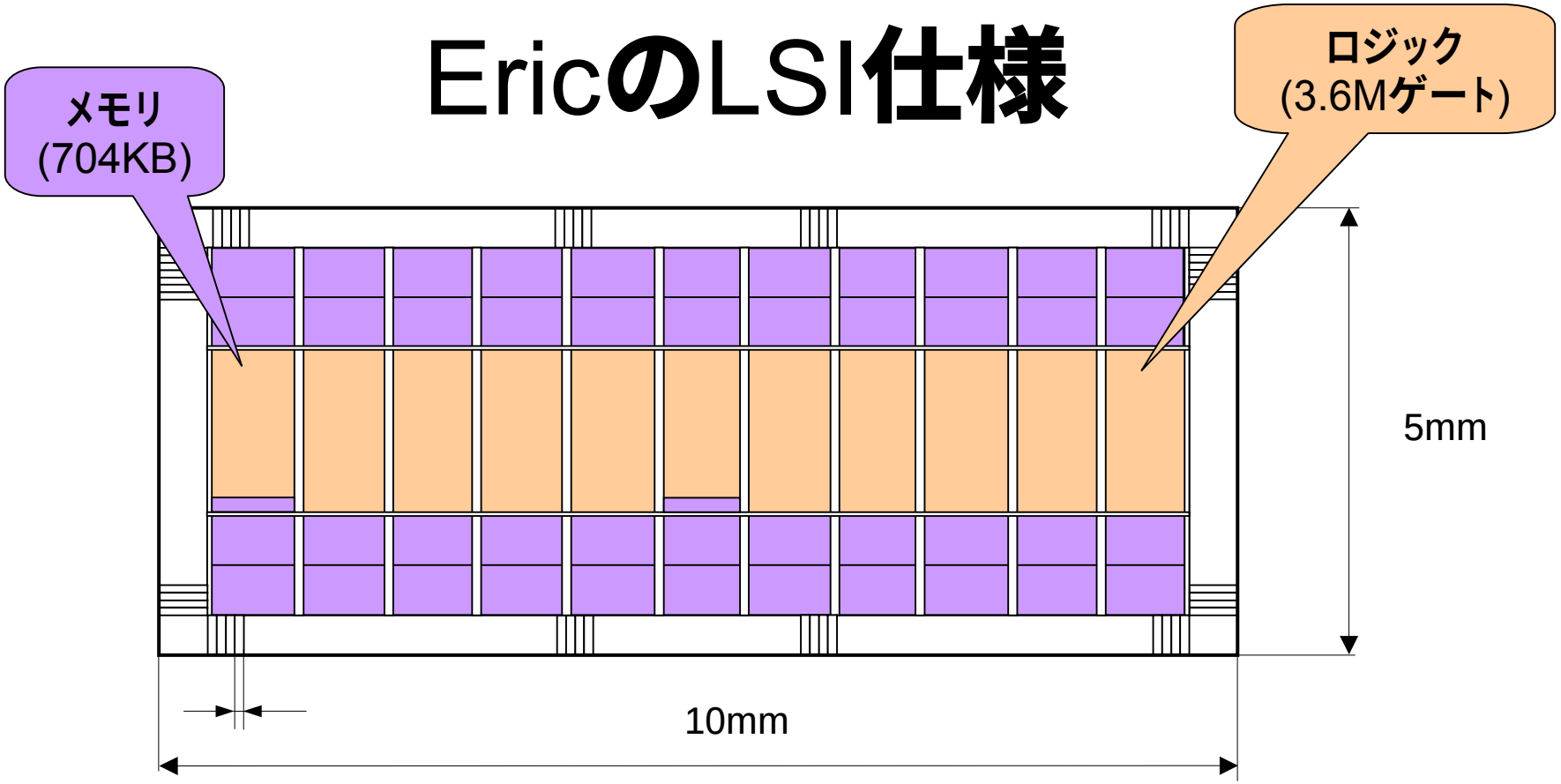
二電子積分計算専用LSI Eric vs. Intel P4

	二電子積分計算専用LSI Eric	Intel P4
マイクロアーキテクチャ	• 非均質CMP (チップ･マルチプロセッサ) – 初期積分計算エンジン×1 – 漸化計算エンジン×4	• シングルプロセッサ • スーパースカラ･プロセッサ
クロック周波数	200MHz (最低達成目標)	3.2GHz (2003年10月時点最速モデル)
実装する倍精度浮動小数点演算器，および，同時実行可能な演算器数	• 積和演算×（1＋4） • 除算／開平逆数×1 • 指数関数／誤差関数×1	• 加算／乗算／除算／開平×1
倍精度浮動小数点演算性能 (ピーク値)	10演算／クロックサイクル	1演算／クロックサイクル

ペプチド分子GAQMY分子での漸化計算にかかる総実行時間

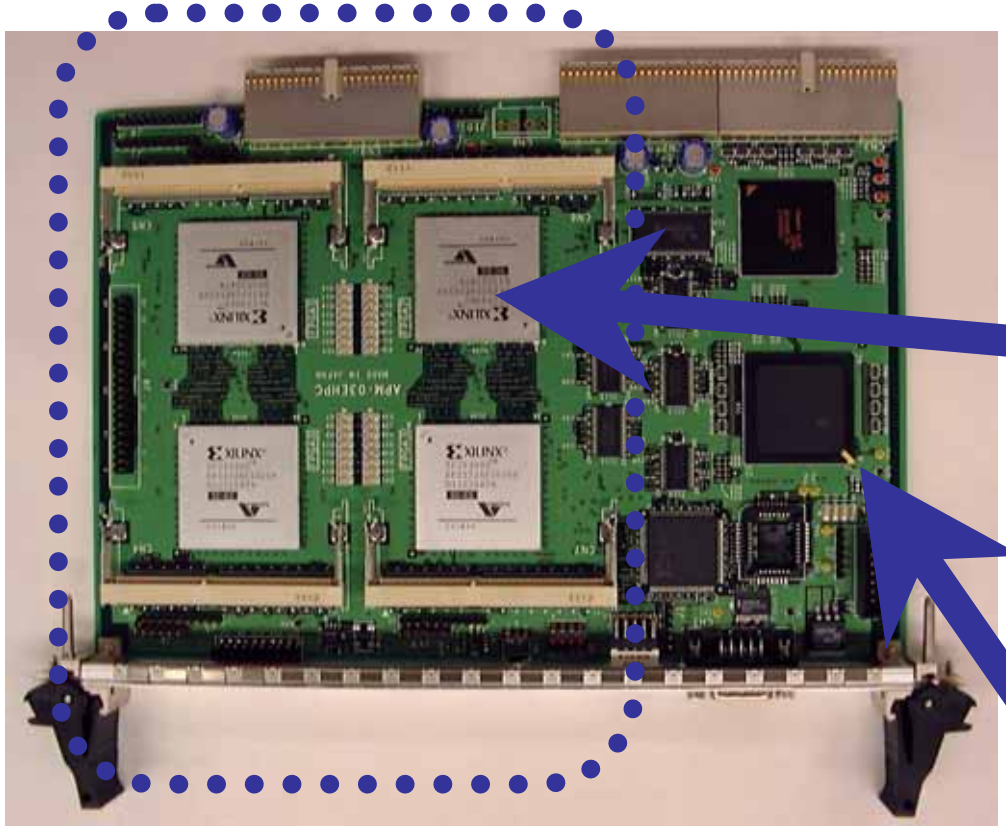


EricのLSI仕様



- **製造プロセス技術**：TSMC 0.13μm**シャトルサービス**
- **論理規模**：3.6M**ゲート**
- **メモリ・サイズ**：704KB
- **面積**：5mm×10mm
- **動作クロック周波数**:200MHz
- **消費電力**：10W
- **ピーク性能**：2GFlop/s(**倍精度**)

Compact PCI規格
二電子積分計算加速ボード



- Compact PCI規格に準拠したプリント基板
- Eric (二電子積分計算専用プロセッサLSI) ×4
- SDRAM
  - 各Eric当り1GB
- 汎用MPU (SH4) ×1
- PCIバスI/F, Ethrenet, 等

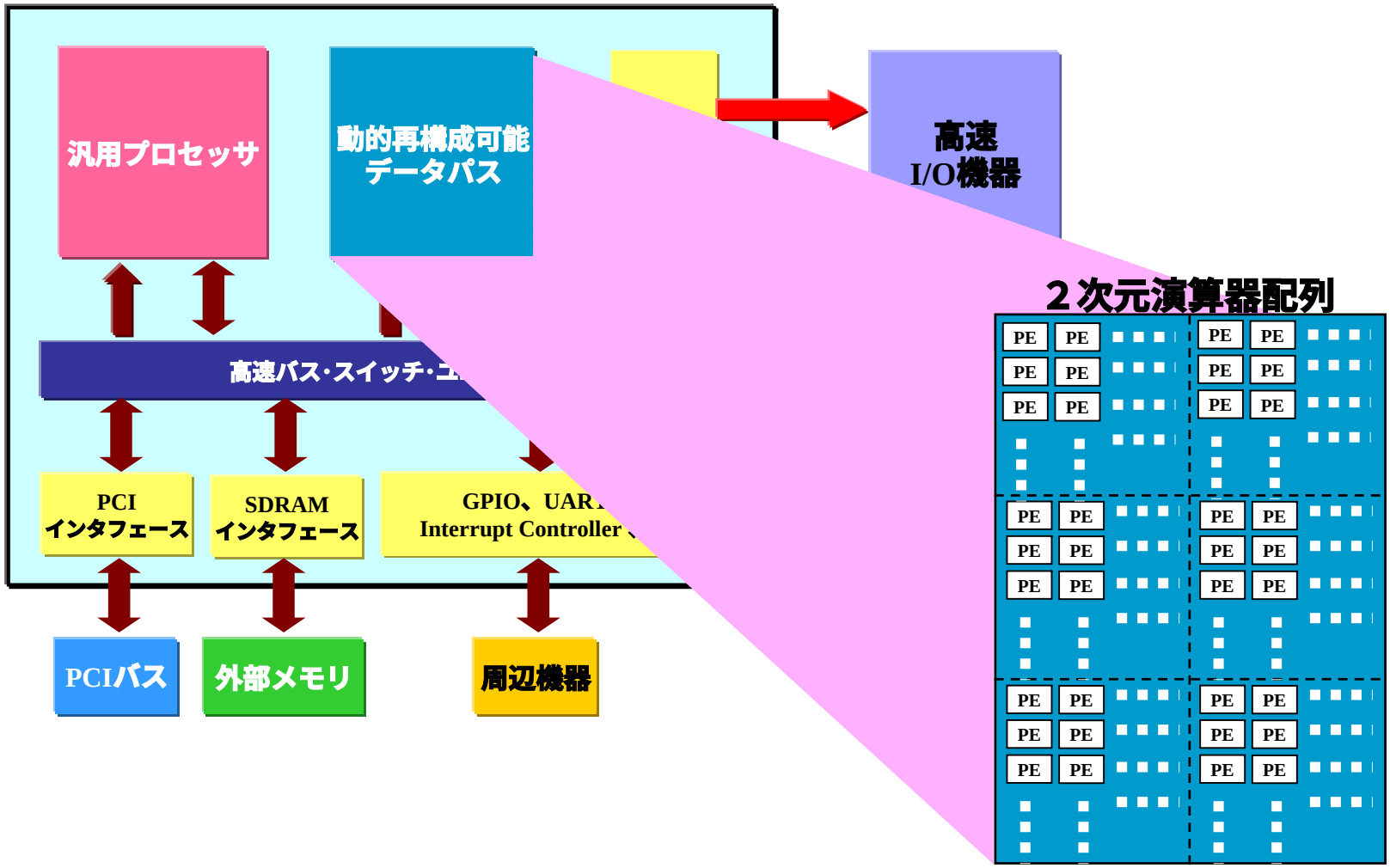
分子軌道法専用マッスル・サーバー
EHPC/Eric

- 1ノード
  - 二電子積分計算加速ボード×7枚
    =Eric×28個
  - PCボード
  - ハードディスク，等
- システム全体
  - 任意数のノードをEthernet接続
  - 4ノード構成の場合
    - Eric112個による並列計算

分子軌道法専用マッスル・サーバー
EHPC/Ericの性能

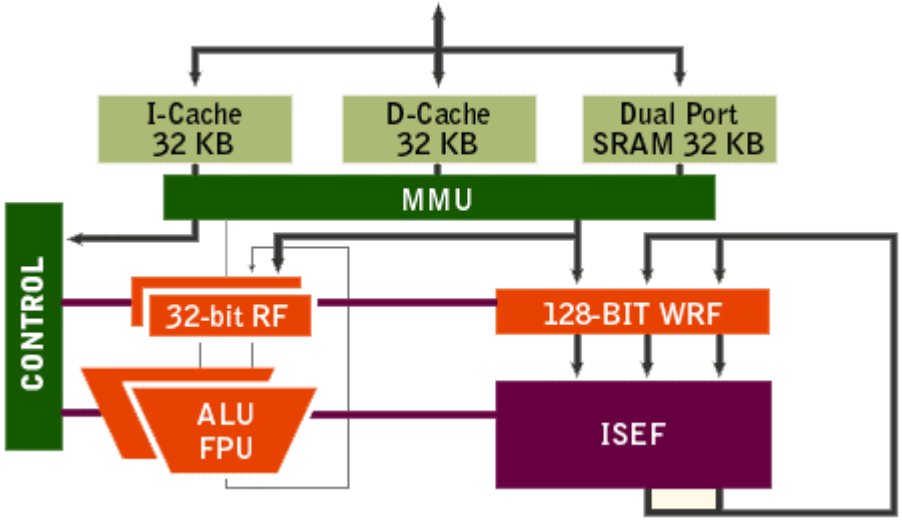
	 EHPC/Eric	 SR8000/64
プロセッサ数	112	512
ピーク性能 (GFlop/s)	224	512
MO計算時間(時)	**(見積もり値) 2**	**4**
体積 (W×D×H:mm)	1200×650×1010	4720×3274×1785
消費電力 (KW)	**(見積もり値)** 2	212
価格 (M$)	**(見積もり値)** 0.1	15

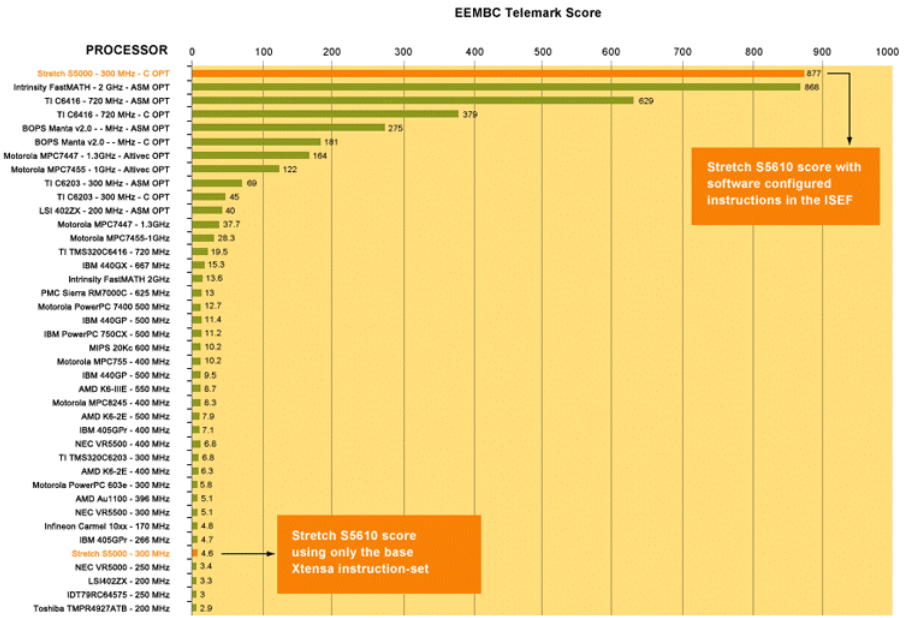
新しいマッスル・サーバー・アーキテクチャの可能性
～動的再構成可能プロセッサのHPC応用～



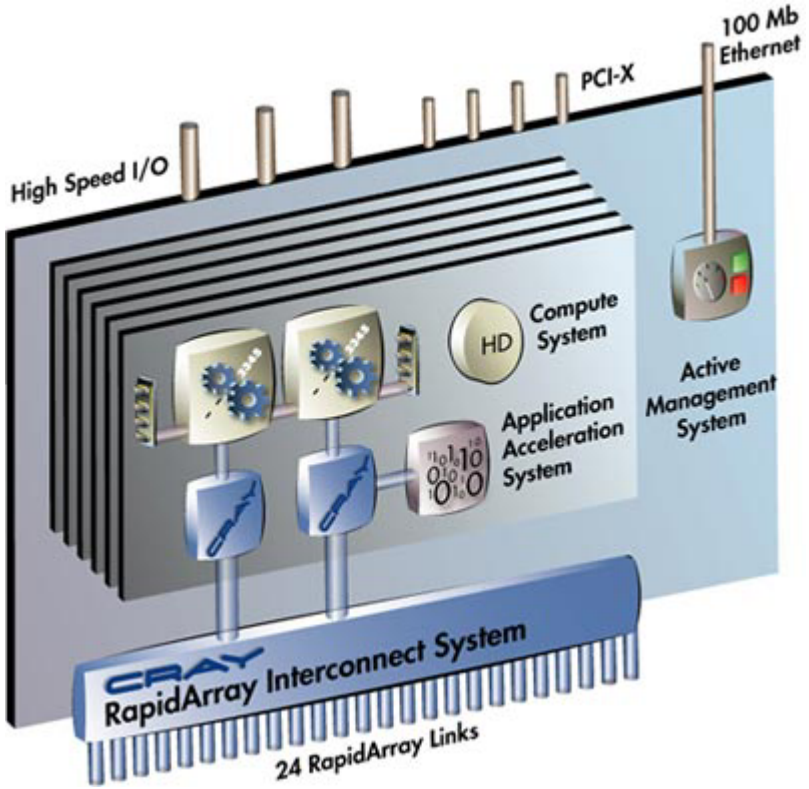
Stretch

- 300 MHz, 32-bit Xtensa-based processor
- 16- and 24-bit instructions
- FPU
- MMU with TLB
- Stretch Instruction Set Extension Fabric
  - Aligned load and store
    - 8, 16, 32, 64, and 128 bit
  - Unaligned load and store
    - Up to 16 bytes variable byte streaming I/O
    - Up to 32 bits variable bit streaming I/O
- User-defined extensions to the core ISA
  - Defined in C/C++
  - Fully pipelined and interlocked
- Low power consumption
- Support for standard operating systems





CRAY XD1



	Chassis	Each Rack
Compute Processors	12	144
Performance	53 GFlop/s	633 GFlop/s
Aggregate Switching Capacity	96 GB/s	1152 GB/s
Interprocessor Latency	1.6 us	1.8 us
Aggregate Memory Bandwidth	77 GB/s	922 GB/s
Maximum Memory	96 GB	1152 GB
Maximum Disk Storage	296 GB	

マッスル・サーバー
(汎用PCクラスタ
＋特定計算向けハードウェア)
の開発
～分子軌道法を例にして～

村上和彰

九州大学　情報基盤センター

murakami@cc.kyushu-u.ac.jp