次世代超並列計算機開発
—— 多粒子系向け超並列計算機の開発

牧野淳一郎

東京大学大学院理学系研究科

講演概要

1. 「計算科学」のなかでの計算機ハードウェア開発の位置づけ
2. 「粒子モデル」向け計算機の意義
3. プロジェクト概要
4. 今年度の成果
    - 開発進行状況
    - ゴードン・ベル賞
5. まとめ

「計算科学」のなかでの計算機ハードウェア開発の位置づけ

「計算科学」は計算機ハードウェアを与えられたものとして進めていけるわけではなく、積極的に新しい計算機の構成を提案していく必要がある

例：

- QCD 計算用並列計算機
- Linux PC クラスター

どちらも、計算機メーカーや計算機アーキテクチャ研究者ではなく、現場の研究者が作ってきたもの

自前のハードウェアを作る必要性

計算機メーカー、計算機アーキテクチャ研究者は、計算科学の必要に本当にマッチしたものをなかなか作ってくれない。

- 要求が多様 ＝ それぞれの市場が小さ過ぎる
- 計算機は「汎用」 ＝ 応用毎に別のものを作るのは計算機科学の研究ではない？
- そもそも計算機科学研究者は計算機を作っていない？

近年の汎用計算機の技術トレンド

- ベクトル並列 $\longrightarrow$ **PC** クラスタ。**PC** クラスタのほうが
  - ピーク性能あたりでは安価
  - 性能向上が小刻みに起こるため平均的には高性能
  - 小規模システムから導入可能
- ソフトウェアの欠如
  - 自動並列化、言語レベルでの記述から、**PVM**、**MPI** 等、メッセージパッシングモデルへの後退
- 大規模計算の困難さの増大
  - ハードウェア並列度の増加
  - プロセッサ間通信のバンド幅、レイテンシの相対的な増加

粒子系シミュレーション

系を記述する方程式：

$$\frac{d^2 \boldsymbol{x}_i}{dt^2} = \sum_{j \neq i} \boldsymbol{f}_{ij}$$

- 多くの場合、相互作用は中心力
- 積分時間が非常に長いことが多い
  ：必要メモリ小さい
- 相互作用の計算が計算時間の大半を占める
  ：必要メモリバンド幅小さい

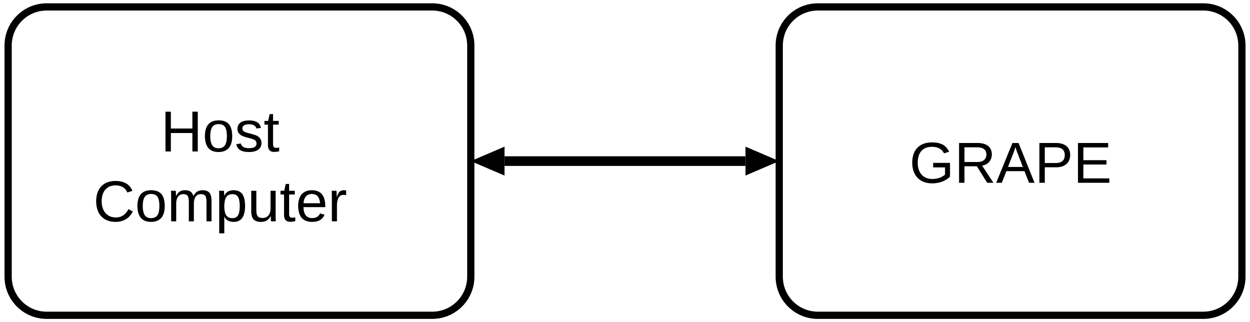
$\begin{cases} \text{「近距離」力：粒子が相互作用するのは}10^2 \sim 10^3\text{個} \\ \text{遠距離力：系のすべての粒子と相互作用} \end{cases}$

粒子 (的) モデル

- 粒子系、長距離相互作用の自然なモデル
  - クーロン力、重力
  - 渦糸法
  - 分子動力学
- 短距離相互作用への適用
  - 圧縮性流体： Smoothed Particle Hydrodynamics 等
  - 構造解析： Element-Free Galerkin method
  - 非圧縮性流体： MPS 等
  - 境界要素法

粒子系向け専用計算機の基本的考え

(GRAPE, GRAvity PipE)



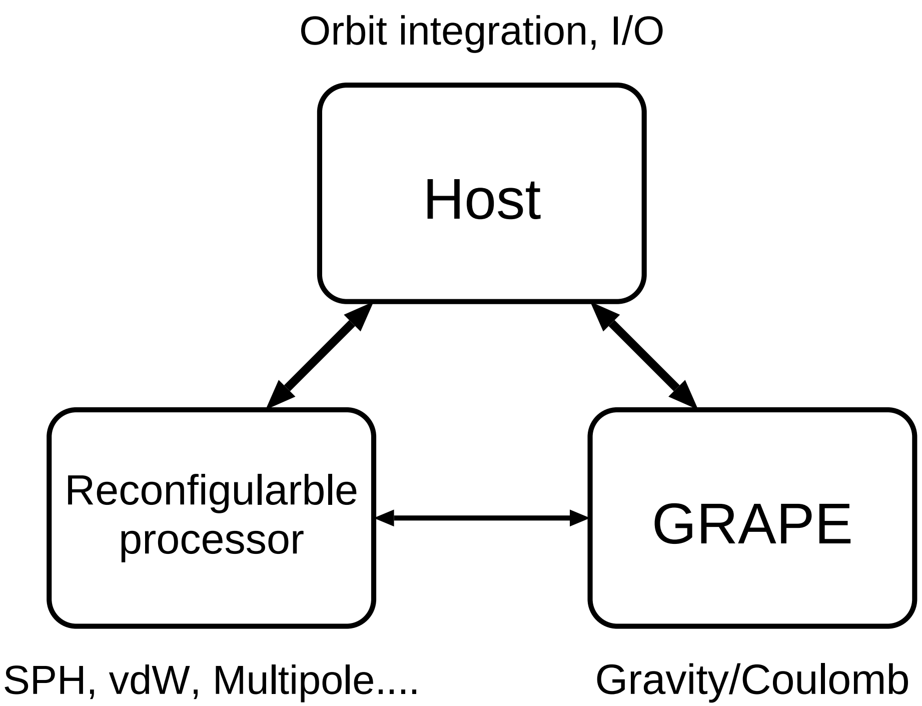
Time integration etc.　　Interaction calculation

- 専用ハードウェアは相互作用計算のみ
- 汎用ホスト計算機が他のすべての面倒を見る

粒子系向け専用計算機の特色

- 専用ハードウェア
  - ハードウェア化した論理 $\longrightarrow$ ソフトウェア不要
  - 相互作用計算専用のパイプライン $\longrightarrow$ 高い性能
  - 並列化、通信についても、専用化により高バンド幅、低レイテンシが実現できる
- 汎用ホスト計算機
  - 高レベル言語が使える (Fortran, C, C++...)
  - 既存のプログラムが多少の手直しで使える
  - ツリーコード、独立時間刻みなども使える

本プロジェクトの概要



1. 重力計算用のサブペタフロップス計算機を実現し、粒子系向け専用計算機の可能性を実証する
2. FPGA を使った粒子系向け多用途計算機を構築する
3. 上の 2 つを組み合わせて、高速で応用範囲の広いシステムを構築する

重力専用計算機

- GRAPE-4 の後継
- 計算速度の向上 (チップ単体で 50–70 倍)
- 全体の性能 $>$ 100Tflops
- 通信速度の向上 ($>$ 2GB/s)
- 周期境界のサポート

粒子系向け多用途計算機の役割

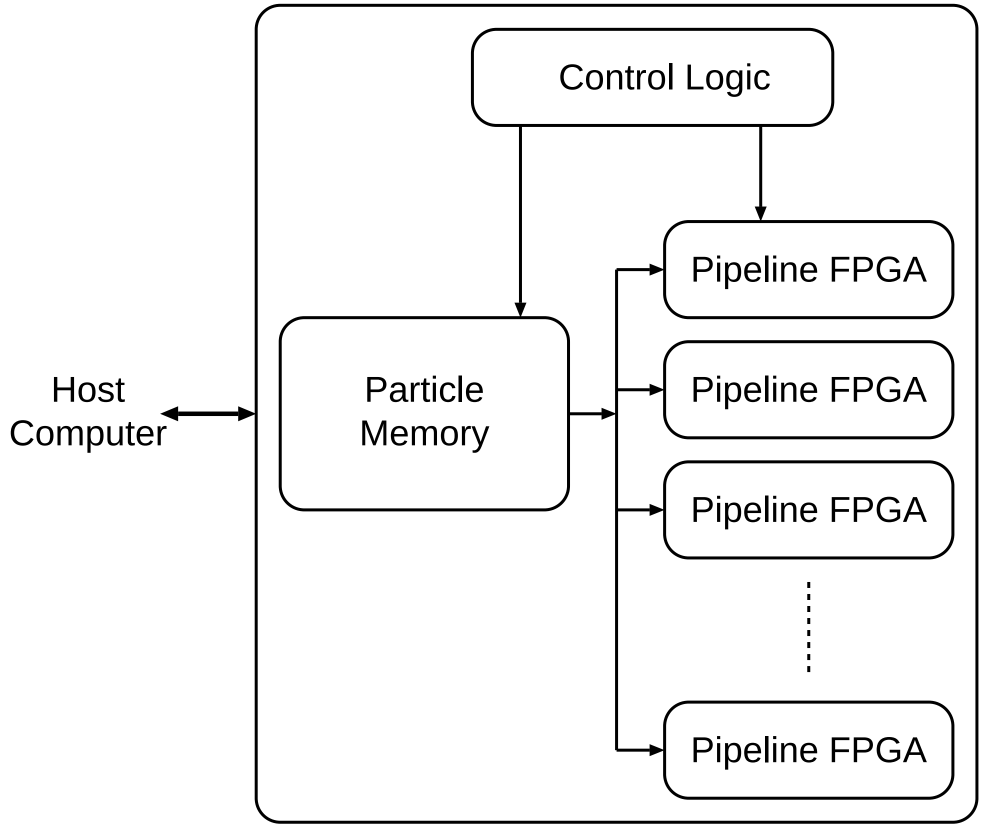
遠距離相互作用に比べて計算量が小さいが、そのかわり多種多様な近距離相互作用を扱う。

例

- **SPH** による自己重力流体シミュレーションでの流体相互作用
- 分子動力学計算での分子間力

多用途計算機の構成

- 基本的発想は GRAPE と同じ
- GRAPE の重力パイプラインの代わりにプログラマブルな FPGA チップをおく
- FPGA の大きさは実効10 万ゲート程度、応用によるがチップあたり1 ～ 10 Gflops 程度が可能。



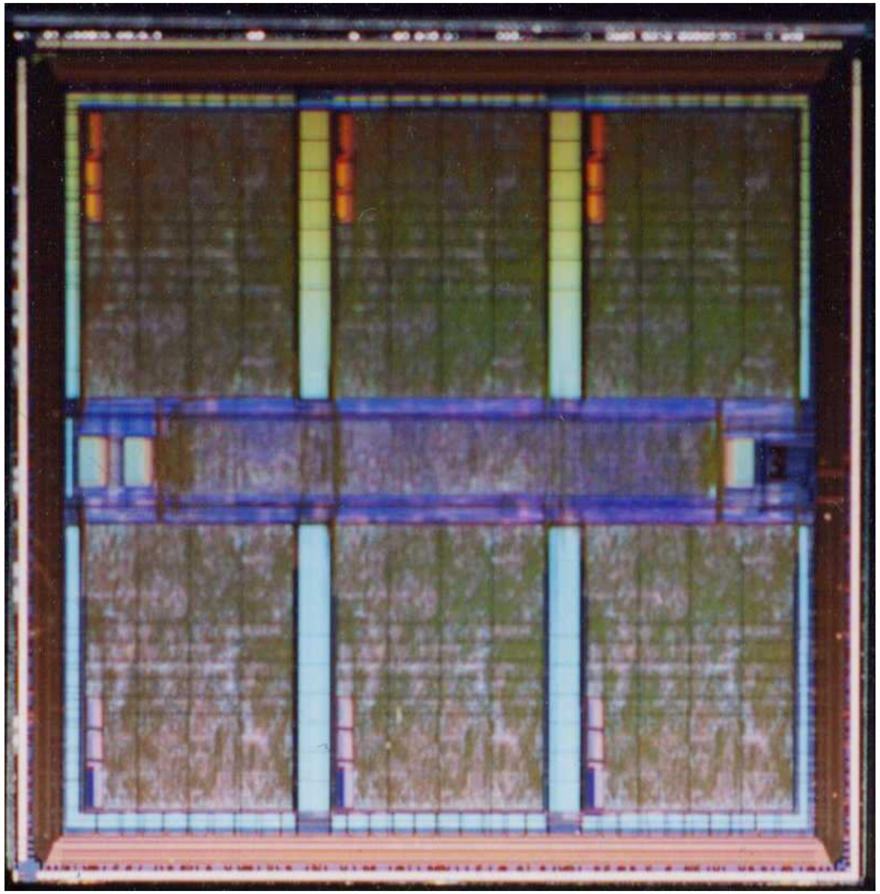
今年度の成果

今年度 (2000 年 1 月まで) の主な成果

- 重力計算システムの開発進行
  - プロセッサチップ完成、プロトタイプシステムによる評価終了
  - ホストインターフェースボード (第 1 版) 完成、評価終了
  - プロセッサボード試作
  - 匡体試作
- GRAPE-5 システムでの応用計算で1999 年ゴードン・ベル賞獲得
- (疑似粒子多重極法の改良、評価 — 今日は省略)

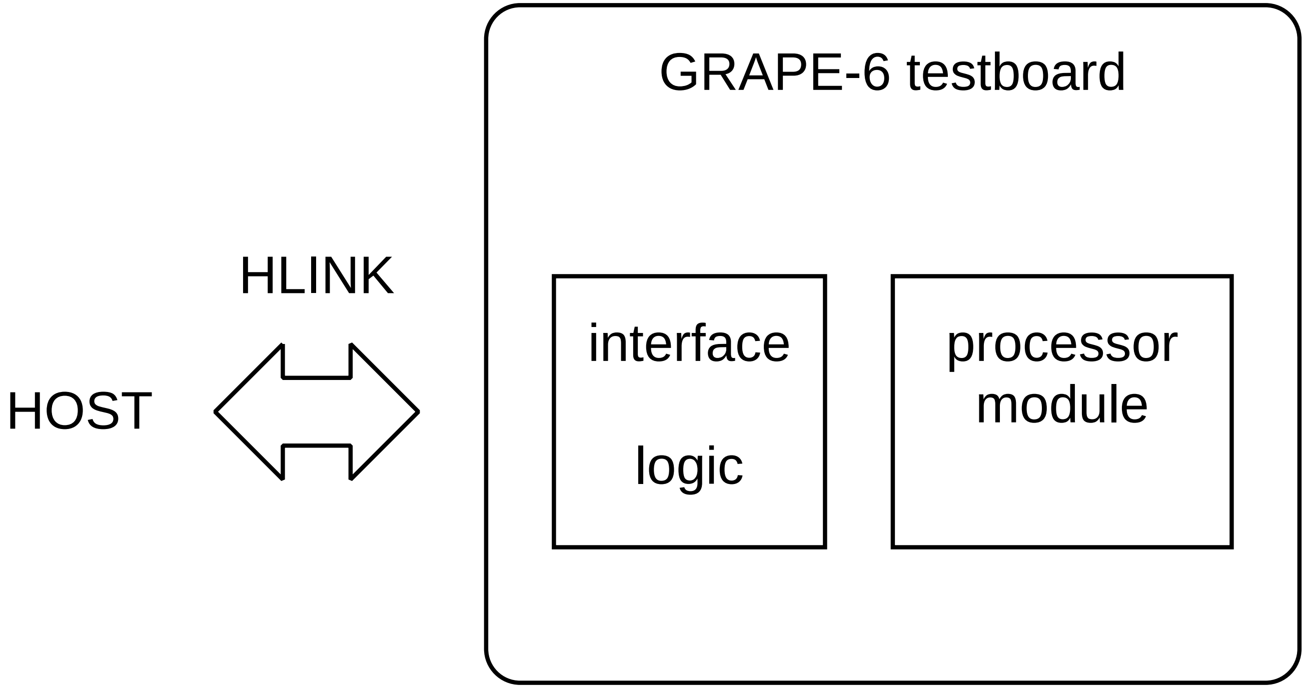
重力計算システムの開発進行状況

重力計算チップ



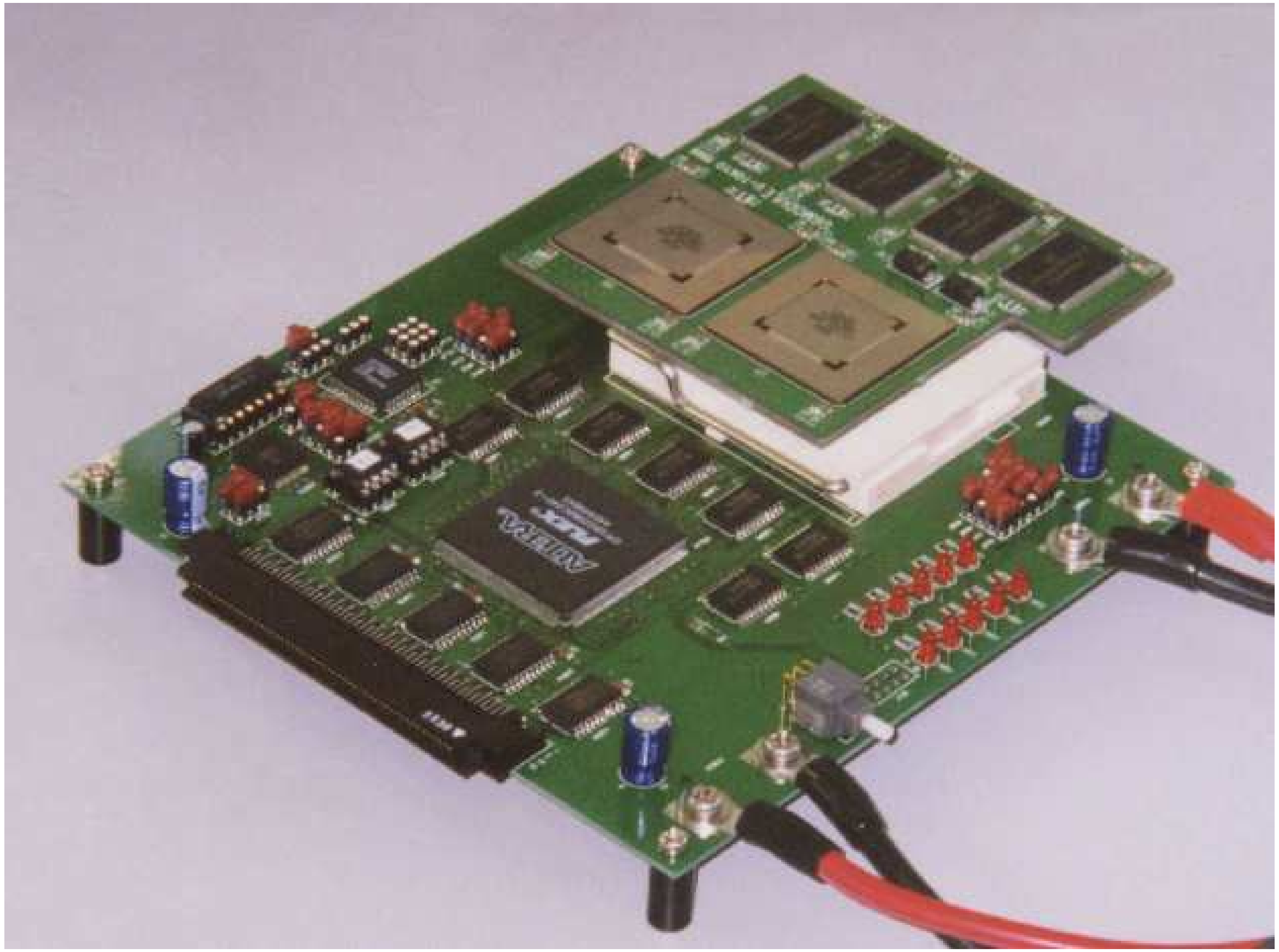
- 0.25 μm ルール (東芝 TC-240, 1.8M gates)
- 2.5 V 電源, 100 MHz 動作
- 1 チップに 6 パイプライン、クロック当たり360 演算
- 36 Gflops （相当）
- 2 チップ＋メモリでモジュールを構成

重力計算チップ評価システム



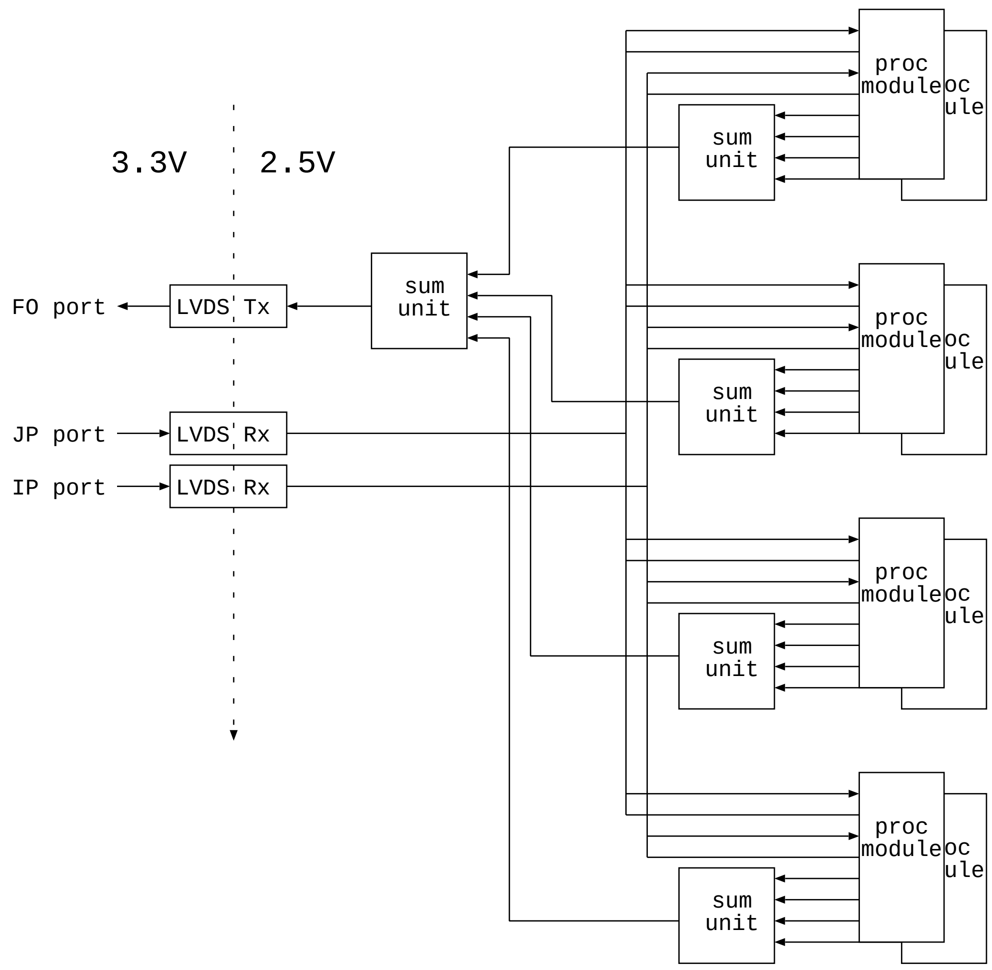
- 2 チップによるチップ評価システム
- PCI インターフェース

重力計算チップ評価システム（続き）

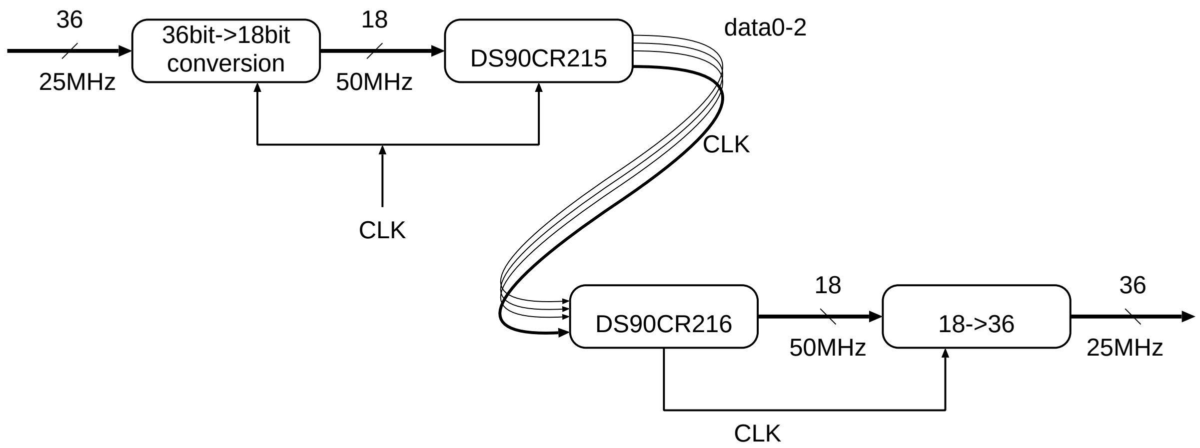


重力計算システムプロセッサボード

- プロセッサモジュール 8 個搭載 (チップ 16 個)
- ピーク性能 576 Gflops
- ホストまたは通信・制御ボードへの接続: LVDS による高速セミシリアル方式 (100MB/s ×3)
- ボード上で計算結果をリダクション

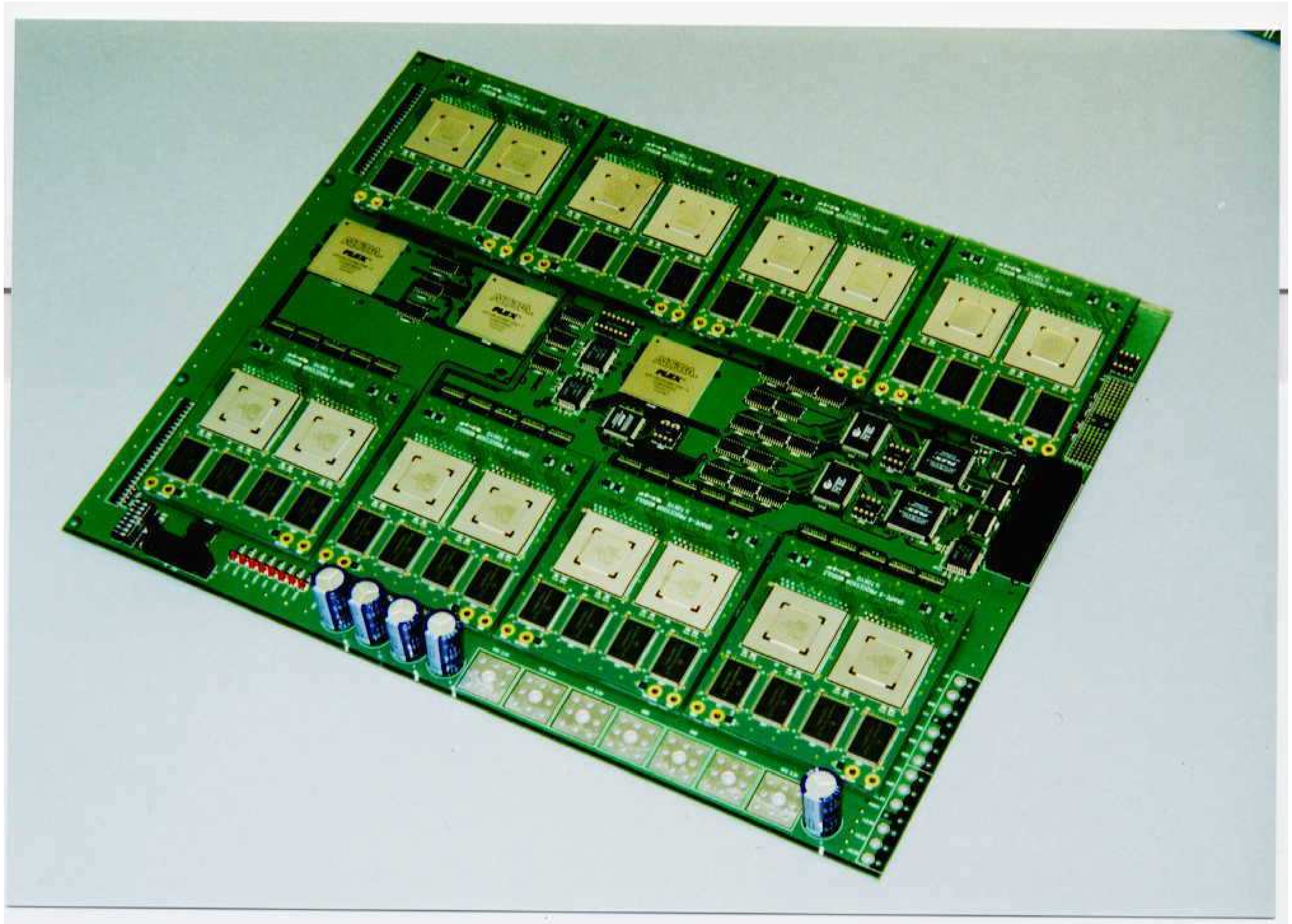


LVDS によるボード間通信

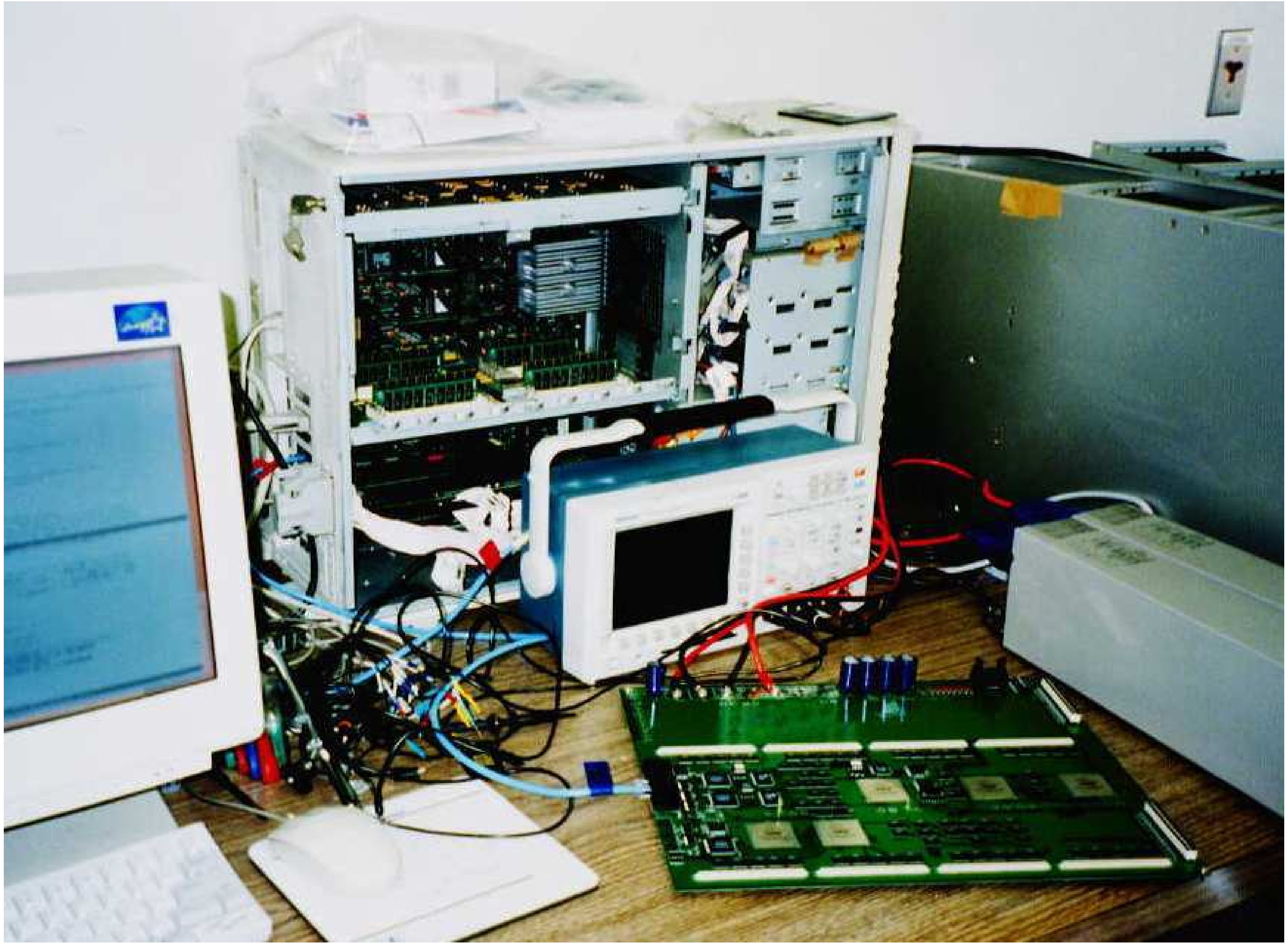


- 信号レベルはLVDS (Low Voltage Differencial Signal, EIA-644)。
- 液晶パネル等に使われる LVDS-link LSI を採用。
- 信号 3 ペア + クロックを通常のカテゴリー 5 ツイストペアケーブルで伝送。コネクタも通常のモジュラジャック。
- 伝送実験で高い信頼性を確認 (エラーレート 10^{-12} 以下)

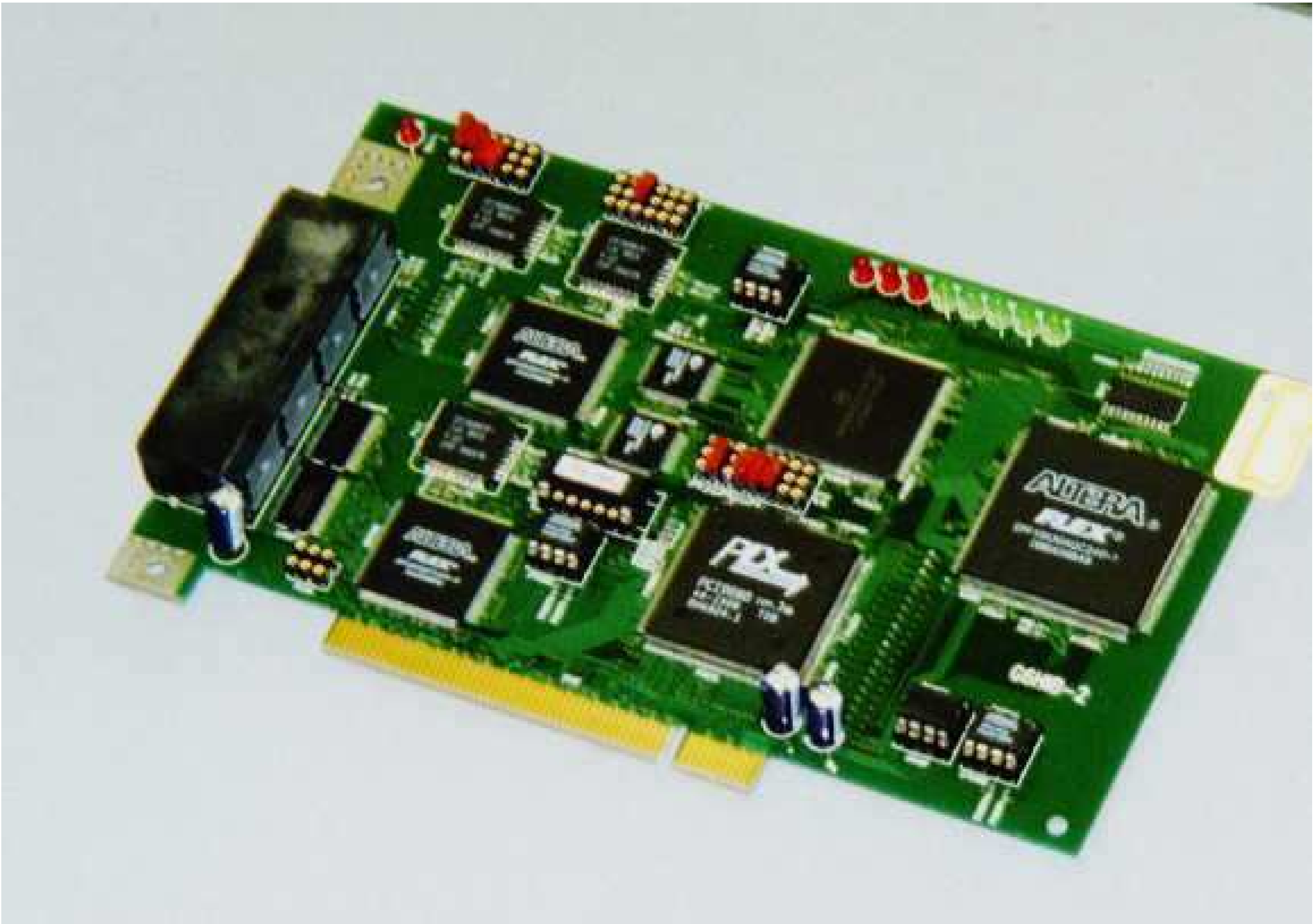
プロセッサボード



調整中のプロセッサボード



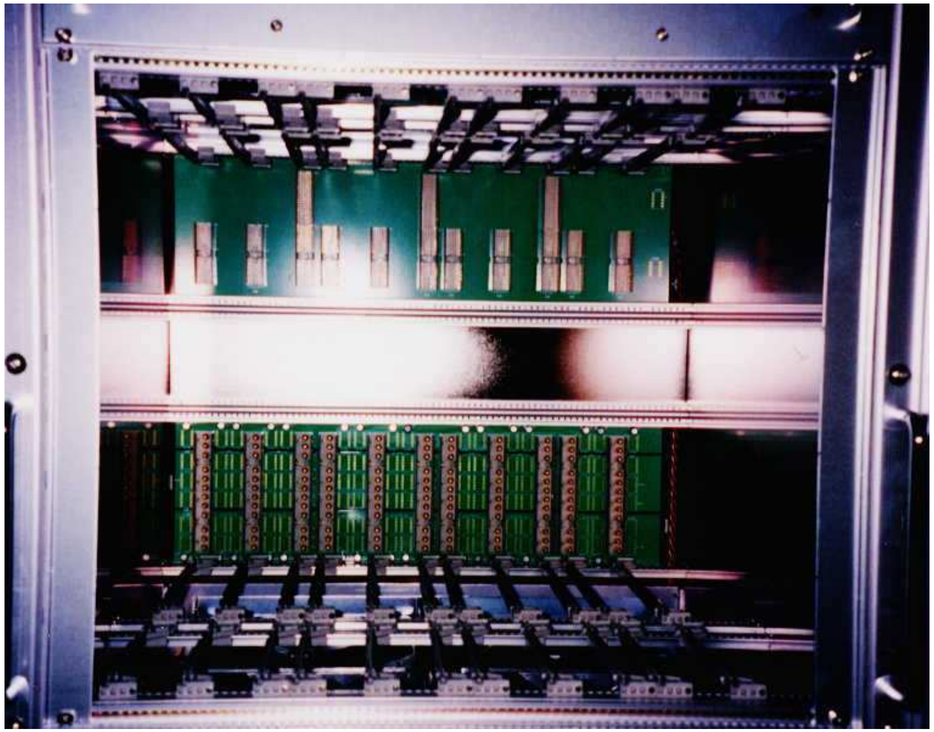
インターフェースボード



インターフェースボード概要

- **LVDS** リンク (出力 **2** 本 / 入力 **1** 本)
- **PCI 32 bit/33MHz** 動作
- バースト転送速度 **133MB/s**
- **DMA**、 バースト **PIO** の両方をサポート
- ユーザープロセスからの直接 **DMA** 起動による低レイテンシ通信を実現
- **64bit** に拡張する予定

匡体

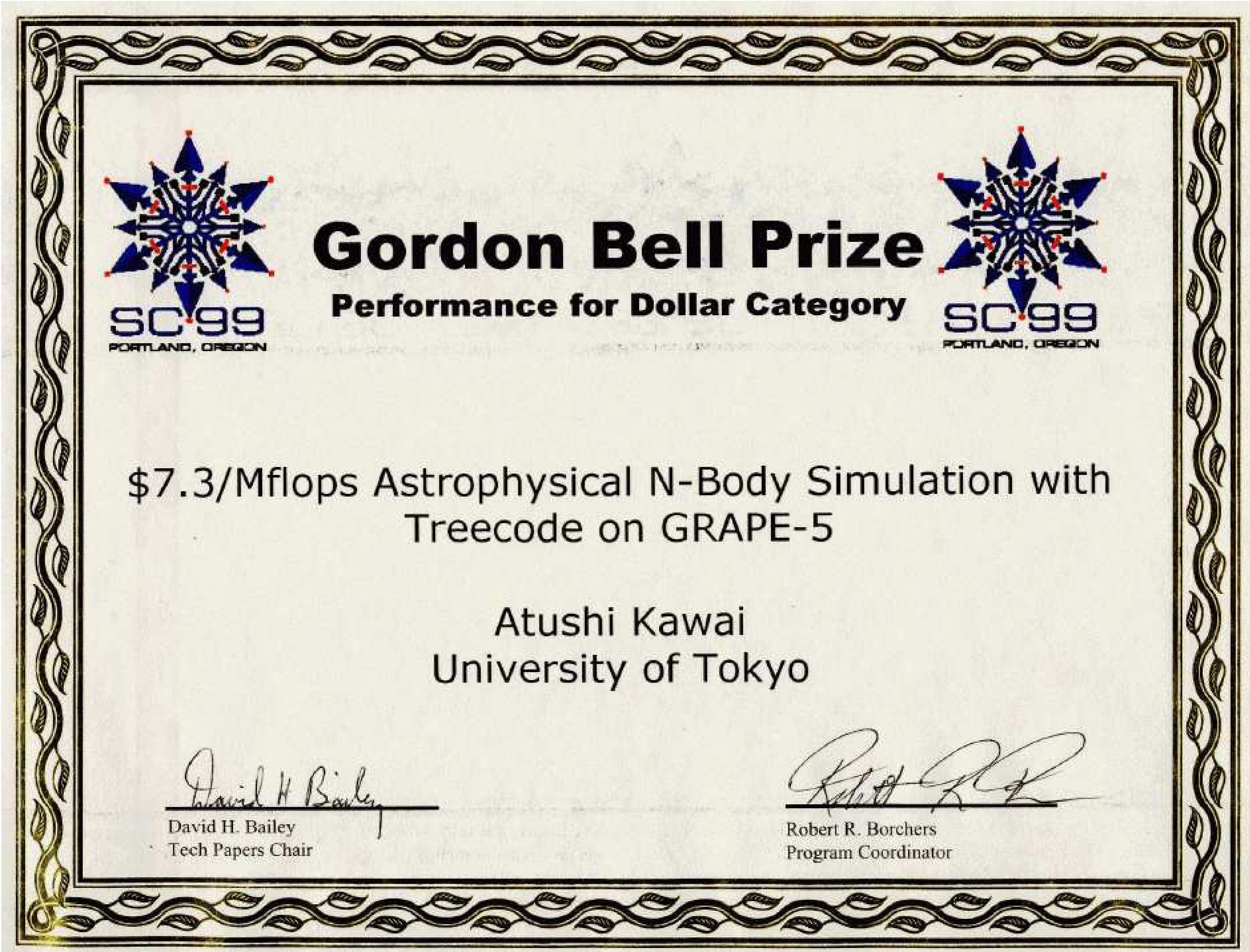


開発状況のまとめ

- プロセッサチップ、モジュールは完成、評価終了、量産中
- プロセッサボードは完成、調整・評価中
- インターフェースボードは完成、評価終了

ほぼ予定通り順調に開発が進んだ。

GRAPE-5 ゴードンベル賞



SC'99
PORTLAND, OREGON

Gordon Bell Prize

Performance for Dollar Category

SC'99
PORTLAND, OREGON

$7.3/Mflops Astrophysical N-Body Simulation with
Treecode on GRAPE-5

Atushi Kawai
University of Tokyo

David H. Bailey
Tech Papers Chair

Robert R. Borchers
Program Coordinator

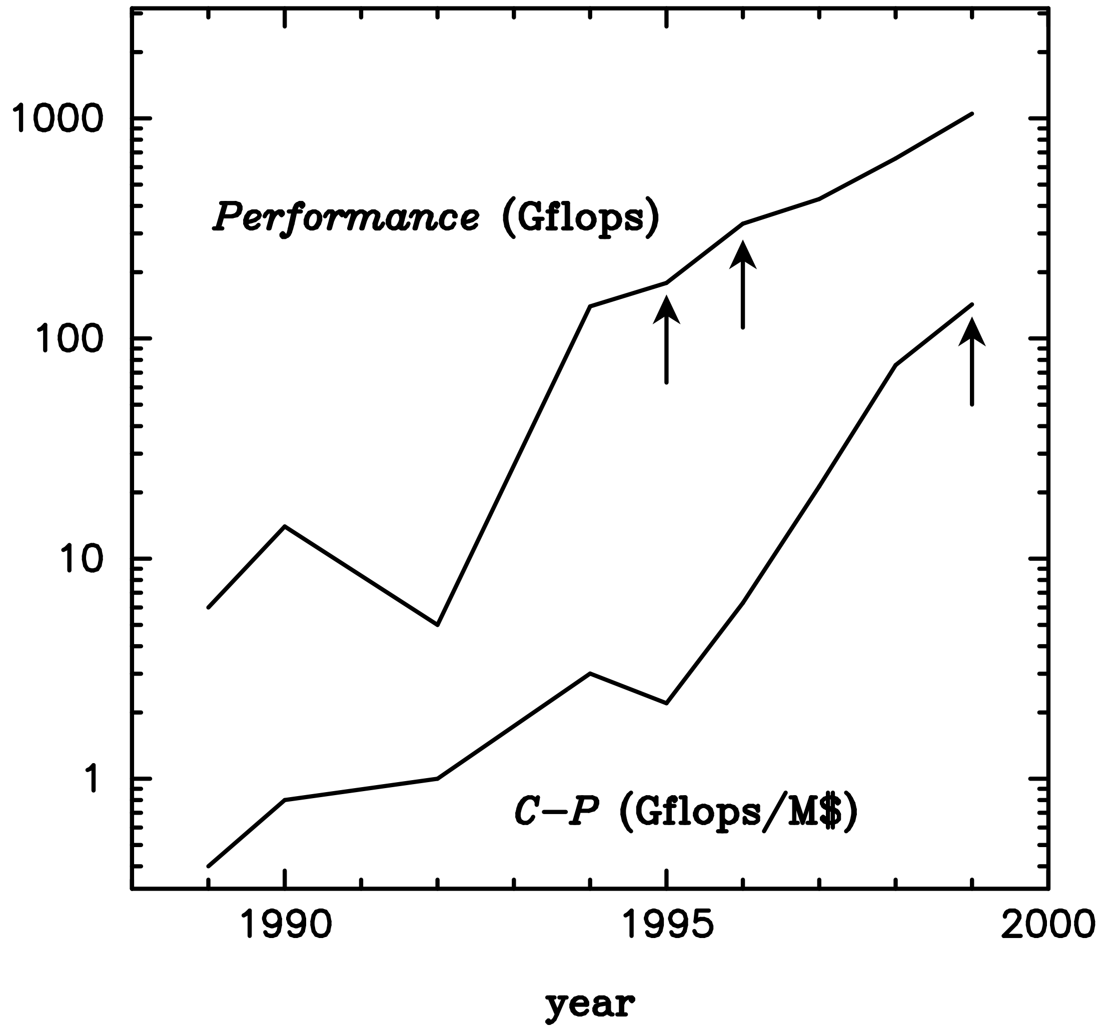
ゴードンベル賞とは？

- 元 **DEC** のゴードン・ベルが **1987** 年に創設。
- 実用的な科学技術計算の並列化に対して与えられる。
- 現在は
  - 実効性能
  - 価格性能比
  - 特別部門（特に並列化困難な応用での達成）

  の **3** 部門。
- 毎年、応募のあったもののなかから選ばれる。

過去の受賞者の達成

- 過去の 10 年の計算機の性能向上を表す良い指標
- 1995, 1996 年は GRAPE-4
- 1997 年の価格性能比は PC クラスタ
- 1998 年の価格性能比は QCD 専用機



今回の我々の計算

- GRAPE-5 × 2 ボード (16 チップ、ピーク 80Gflops) +Compaq XP1000
- Barnes-Hut ツリー法による計算を実行。単純な $O(N^2)$ 直接計算ではない
- 前年の同じ Barnes-Hut ツリー法での応募 (70 台の Alpha クラスタ、一台 $1,700 で購入) に比べて3 倍近い価格性能比の改善、同等の実効性能を実現

ツリー法のような高速アルゴリズムとの組み合わせでも汎用計算機、そのなかでも価格性能比の高いとされる PC クラスタに比べて専用計算機が優れていることを実証した。

GRAPE-5 と Alpha クラスタ

$\simeq$

まとめ

- 本プロジェクトでは、これまでの GRAPE の成果を拡張し、多様な粒子系に適用可能な超高速並列計算機を開発する
- 今年度はほぼ計画通りの成果を達成した。
- 具体的には、重力計算チップが完成し、それを搭載するプロセッサボード、匡体なども現在調整ないし評価中である。
- GRAPE-5 のゴードン・ベル賞価格性能比部門での受賞で、ツリー法のような高速アルゴリズムとの組み合わせでも専用計算機が有効であることをデモンストレートできた。