

秘

回収資料

「評価項目と評価の視点又は基準」 に関する説明資料

平成19年4月27日

理化学研究所
次世代スーパーコンピュータ開発実施本部

1. システム開発方針の適切性

理化学研究所が設定したシステム開発方針(システム最適化の考え方を含む)は、文部科学省におけるプロジェクトの目的及び目標に照らして妥当か。

システム開発の方針

■ プロジェクトの基本方針

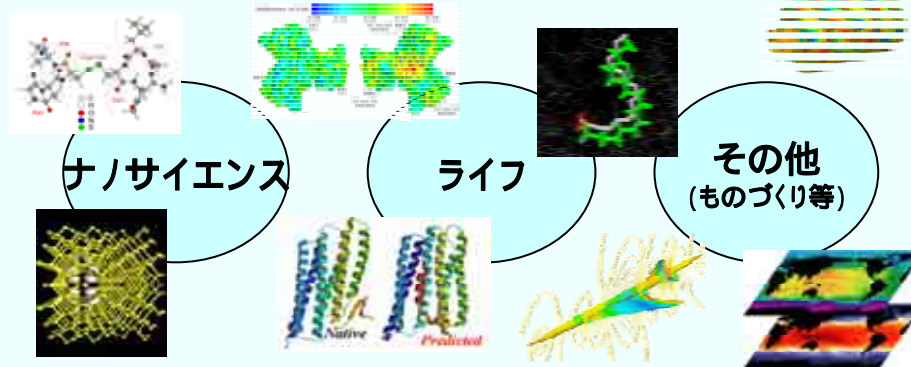
- 計算機シミュレーションにより、科学技術・産業の競争力を維持、高めること。
- スーパーコンピュータの開発力を国内に保持し、継続的な開発を可能とすること。
- 完成時に世界最速と内外から広く認められること。

■ システム開発方針

- 理論性能やLINPACK性能(10PFLOPS以上)を考慮しつつ、実効性能(アプリ性能)を重視したシステム構築を目指す。
- 幅広い活用を促すため、低コストを実現しつつ、利便性の高い汎用機により目標性能を達成することを目指すとともに、アクセラレータの検討も行う。
- 低消費電力CPUなど、新規性の高い技術をベースとした波及効果の高いハードウェア技術の開発を目指す。

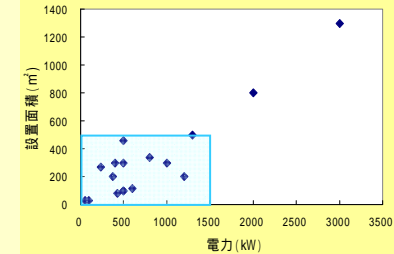
システム最適化の考え方

グランドチャレンジからの要求要件



制約条件

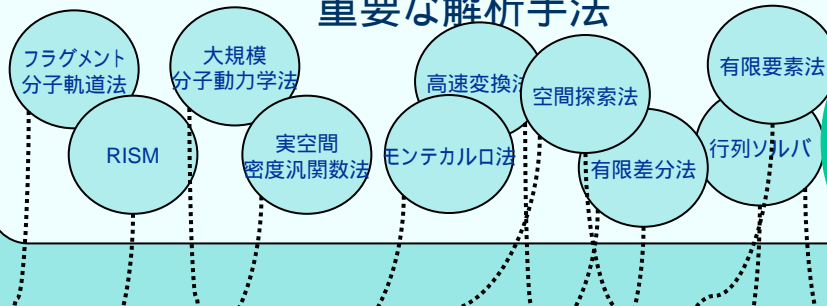
電力, 設置面積



信頼性, 保守性

コスト(開発費, 製造費, 保守費等)

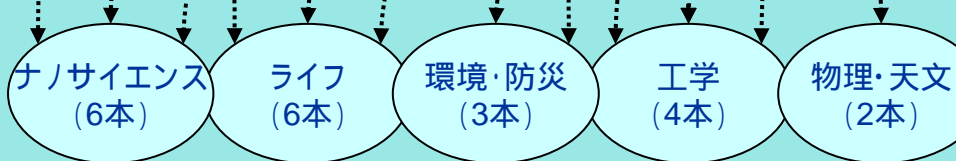
重要な解析手法



最適なシステム構成

世界最速 (完成時)

ターゲットアプリケーションによるシステム検討
- 5分野, 21本のベンチマークテストを抽出



【海外調査】
HPC分野の動向
(開発計画, 予算等)

【国内技術調査】
システム
アーキテクチャ

【運用・利用】
(メモリ容量, ファイル容量, システム運用,
ユーザー管理, 保守条件等)

【要素技術】

半導体製造
プロセス

低消費電力化
SOI
Low-k
High-k

光伝送技術

ソフトウェア
OS, コンパイラ等

産業への波及効果
技術条件, 運用条件

空白・ページ

2. システム構成案の妥当性

(1) システム構成案の詳細及び性能

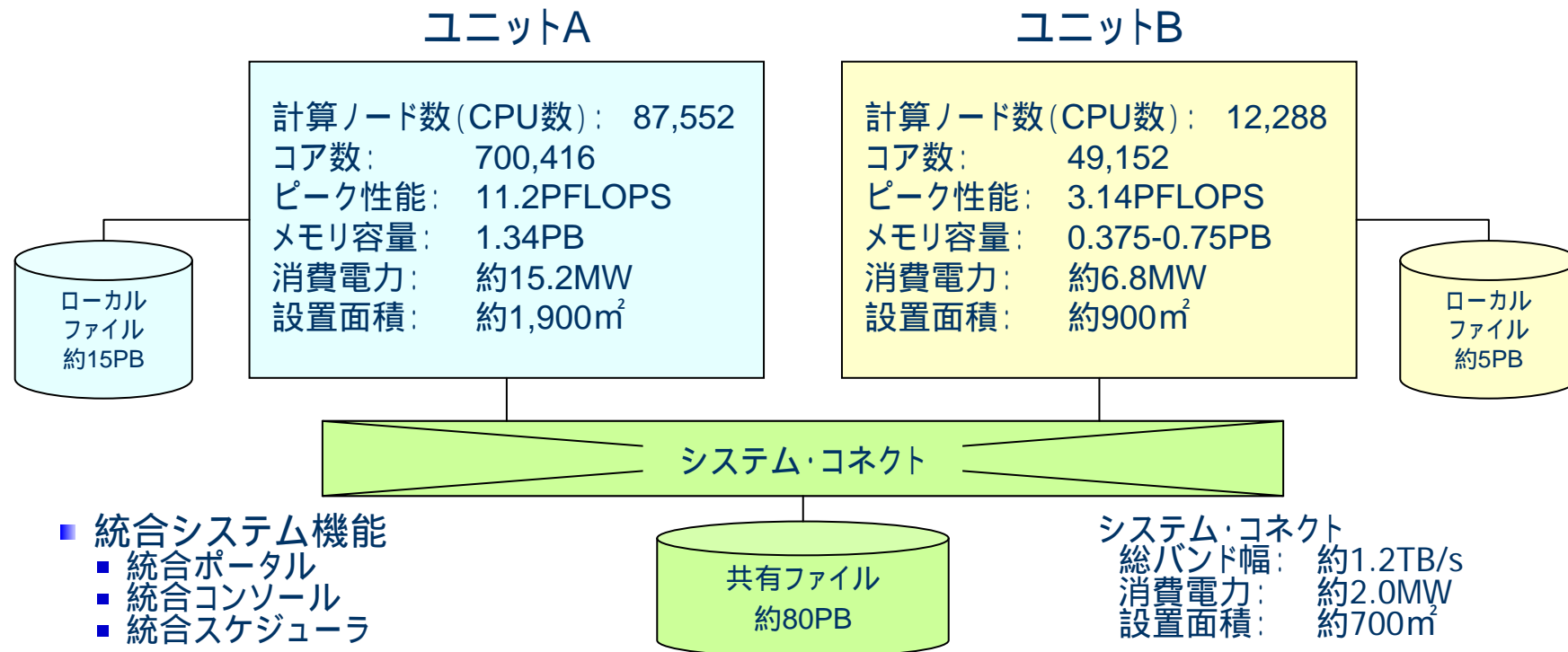
次の性能目標を実現する上で、システム構成案(プロセッサ、メモリ、ネットワーク等の構成)は適切か。

- Linpackで10ペタFLOPSを達成(平成23年6月)。
- HPC CHALLENGE全28項目中、過半数の項目で最高性能を達成。

空白・ページ

統合汎用スーパーコンピュータシステムの構成

- ノード数 (CPU数): 99,840
- コア数: 749,568
- ピーク性能: 14.3PFLOPS
- メモリ容量: 1.7-2.1PB
- 磁気ディスク容量: 約100PB
- 消費電力: 約24MW (空調を除く)
- 設置面積: 約3,800m² (空調を除く)
- 電力性能比: 約1.68MW/PFLOPS
- 面積性能比: 約266m²/PFLOPS

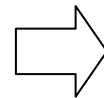


性能目標について

- ユニットA + ユニットBでLINPACK 10PFLOPS超を達成する。

【試算】

- ユニットA: $11.2\text{PFLOPS (ピーク性能)} \times 85\% (\text{LINPACK効率}) = 9.52\text{PFLOPS}$
- ユニットB: $3.1\text{PFLOPS (ピーク性能)} \times 90\% (\text{LINPACK効率}) = 2.79\text{PFLOPS}$



ユニットA+BのLINPACK性能

| | |
|---------|-------------|
| 90%の場合: | 11.08PFLOPS |
| 85%の場合: | 10.46PFLOPS |
| 80%の場合: | 9.85PFLOPS |

- HPC Challengeの性能については検討中。

2. システム構成案の妥当性

(1) システム構成案の詳細及び性能

システム構成案は、消費電力及び設置面積あたりの演算性能において妥当であるか。

空白・ページ

概念設計における性能目標

■ 概念設計の主な要求仕様

- ピーク性能 10PFLOPS以上
- メモリ容量 2.5PB以上
- 消費電力 30MW以下(周辺機器, 空調機器を含む)
- 設置面積 3,200m²以下(周辺機器を含む)

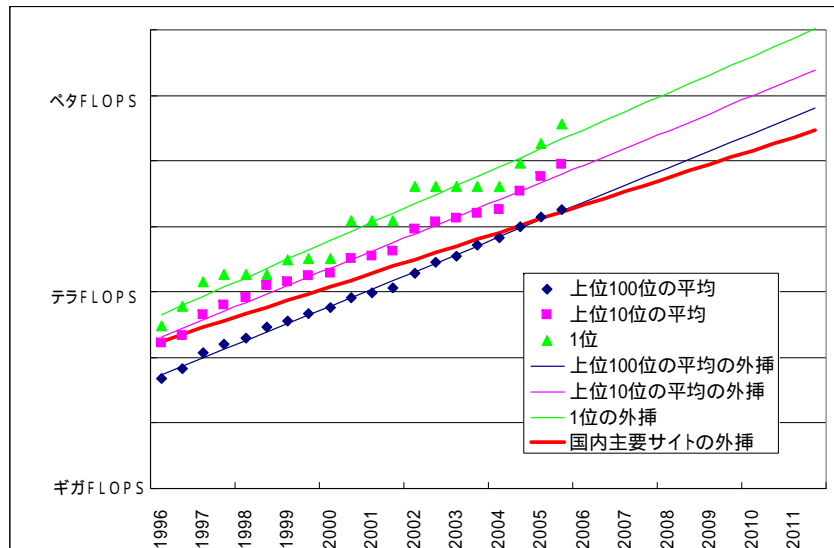
■ 統合汎用スーパーコンピュータシステム

- ピーク性能 14.5PFLOPS
- 消費電力 約21.4MW(本体システム), 約24MW(周辺機器を含む)
- 設置面積 約2,800m²以下(本体システム), 約3,800m²(周辺機器を含む)

- 本体システムのピーク性能あたりの消費電力: 約1.5MW/PFLOPS
- 本体システムのピーク性能あたりの面積: 約193m²/PFLOPS

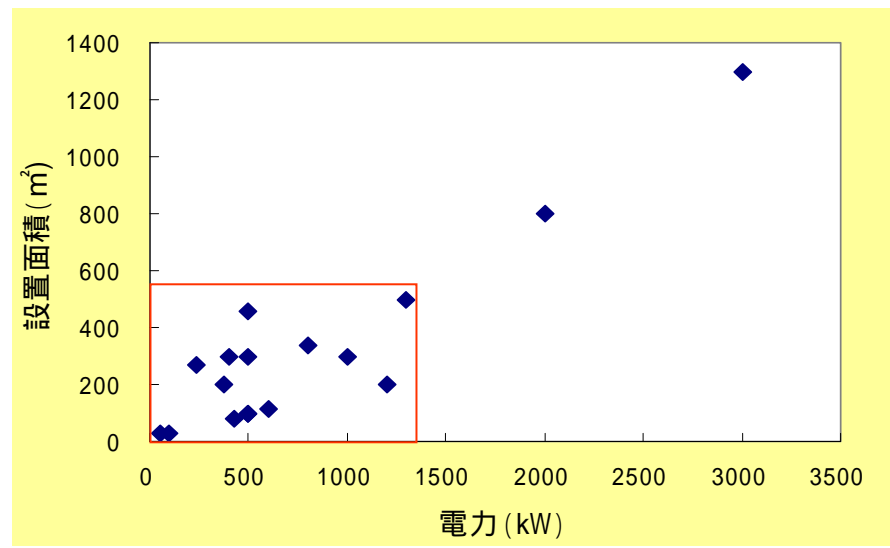
スーパーコンピュータセンター調査結果

TOP500における国内外の計算機性能
上昇トレンドの比較



- 国内計算機センターのスーパーコンピュータ性能は長期低落傾向にある
- 国内の計算機センターは年率約1.6倍の性能向上
- 世界的には年率約1.8倍で性能が上昇
(TOP500リストによる)

各センターの電力的および設置面積的制約

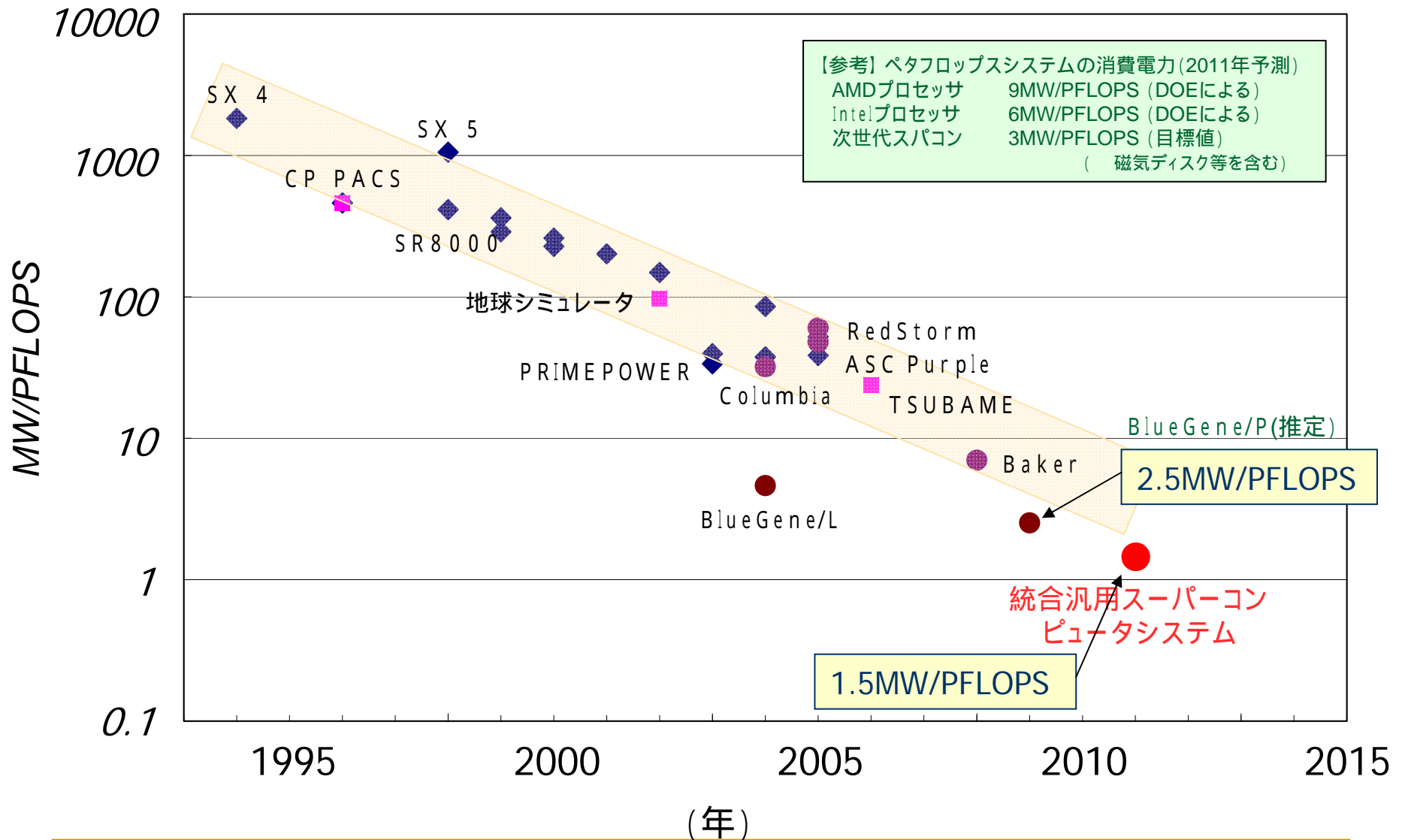


- 設置面積, 受電設備許容量には強い制約がある.
- ほとんどのスパコンセンターの
 - 設置面積は約600m²以下
 - 受電設備容量は1.5MW以下

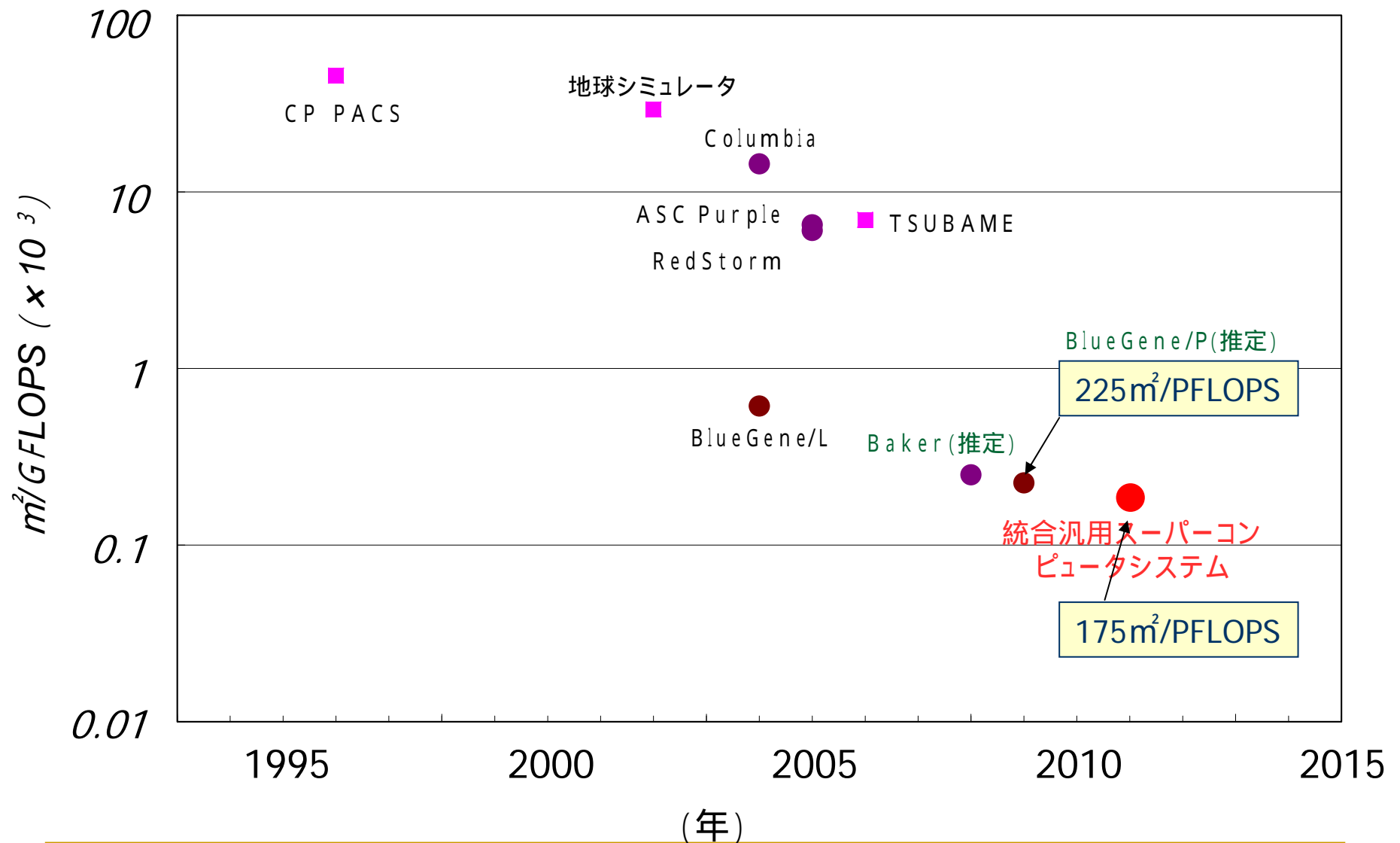


これらの調査結果と、次世代スーパーコンピュータセンターとしての設備制約条件から、前記目標を設定。

本体システムのピーク性能あたり消費電力



本体システムのピーク性能あたりの面積



2. システム構成案の妥当性

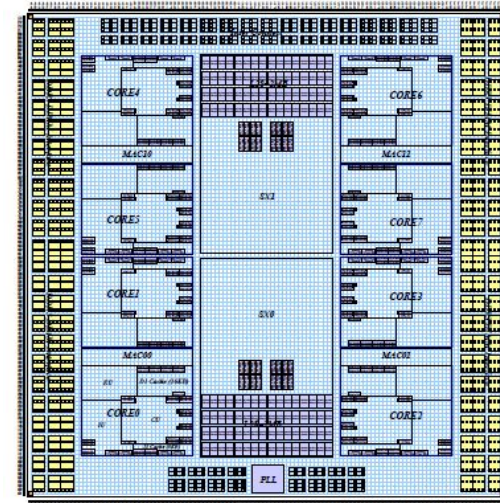
(2) システム構成案の詳細及び性能

システム構成案を実現するための要素技術は、現在の技術水準及び今後の見通しから判断して、システムの製作時期までに開発可能か。

空白・ページ

【ユニットA】論理LSIの開発要素技術 (1/2)

論理LSIフロアプラン



- **高性能技術**
 - 富士通45nm半導体プロセス技術
 - マルチコアLSI技術(8コア)
 - HPC向け機能拡張コア
 - コア当たり4つの浮動小数点積和演算器
 - SIMD機構
 - 共有2次キャッシュ(6MB)
 - 高速バリア機構
- 高いピーク性能(128GFLOPS)に加え, 実環境でも高い性能を発揮

- **低消費電力技術**
 - アクティブ電力, リーク電力削減技術
 - 低温動作
- 消費電力/性能を当社従来比10分の1
- 小型 ~ 大型/ブレードサーバへ適用可能

| 低消費電力技術 | 主な狙い (: 該当) | |
|------------------|--------------|---------|
| | アクティブ電力削減 | リーク電力削減 |
| クロック制御 | | - |
| RAMの電力削減 | | - |
| 省電力ラッチの採用 | | |
| Vth最適化 | - | |
| トランジスタサイズ最適化 | | |
| キャパシタセルのゲートリーク削減 | - | |
| チップごとのVdd, Vbs制御 | | |

【ユニットA】論理LSIの開発要素技術 (2/2)

高信頼性回路技術

- 高信頼技術
 - メインフレーム用高信頼性回路技術
 - ハードによるエラー検出 & 自己回復
 - 低温動作
 - 故障率の低減
- ➔ ペタスケール (およそ10万個のプロセサ相当)のシステム運用を実現
- 論理LSI開発スケジュール

| | | 種別 | エラー検出方法 |
|----------|-------------|---|----------------------|
| 内蔵RAM | L1命令\$ | データ部 | パリティビット |
| | | タグ部 | パリティビット |
| | L1データ\$ | データ部 | SEC・DED ECC |
| | | タグ部 | パリティ |
| | L2\$ | データ部 | SEC・DED ECC |
| | | タグ部 | SEC・DED ECC |
| | 命令・データ mTLB | | パリティビット |
| ブランチヒストリ | タグ部, データ部 | パリティビット | |
| 実行部 | レジスタ | GPR, FPR, GUB, FUB, PC, PSTATE, 演算入出力ラッチ等 | パリティビット |
| | 演算器 | ALU, SHIFT, 除算器, グラフィック演算器, 比較器 | パリティ予測回路 |
| | | 乗算器, FMA | 剰余チェック回路 パリティ予測回路 |

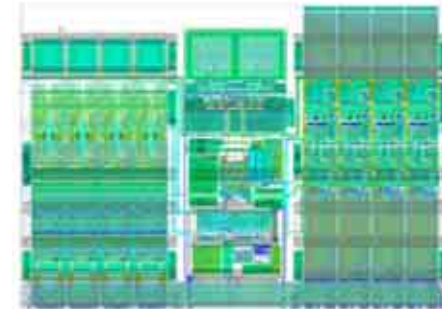


【ユニットA】インターコネクットの開発要素技術 (1/2)

■ シリアルI/Oマクロ技術

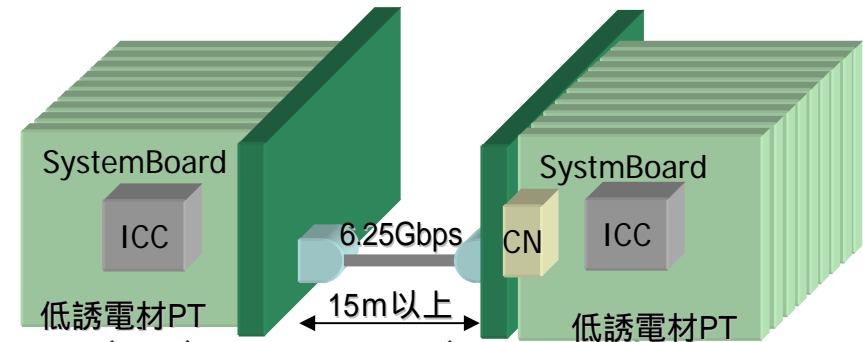
- 6.25Gbps高速シリアル伝送
 - プリエンファシス, イコライゼーション技術の確立
 - 小面積, 低消費電力を実現
- 主流となりつつある, 高速シリアル伝送技術を発展
汎用シリアルバス方式の高速化に寄与

3.125Gbpsシリアルマクロ



■ 高速信号伝送技術

- 6.25Gbpsデータ伝送技術の開発
 - 低損失コネクタの開発
 - 低誘電材料を使用したPT板開発
 - ラック間15m以上のデータ伝送を実現
- 高速伝送が必要とされるIDCでの汎用サーバ, ブレードサーバへの応用



【ユニットA】インターコネクットの開発要素技術 (2/2)

■ ToFuインターコネクット

- 高帯域, 低遅延伝送プロトコル
- 耐故障性にすぐれたシステム運用機構
- 集合通信処理を高速化するMPI処理オフロード機構
- 大規模メモリシステムを実現する分散メモリ機構
- 論理ピーク性能100PetaFlopsを超えるスケーラビリティ

- ペタスケールコンピューティングに最適なインターコネクットの実現
- HPCだけでなく汎用サーバへの転用も可能

■ インターコネクット開発スケジュール

| 2007年度 | 2008年度 | 2009年度 | 2010年度 |
|--------|--------|--------|--------|
| 仕様検討 | 詳細設計 | 製造 | システム検証 |



【ユニットA】部品単位水冷技術 (1/2)

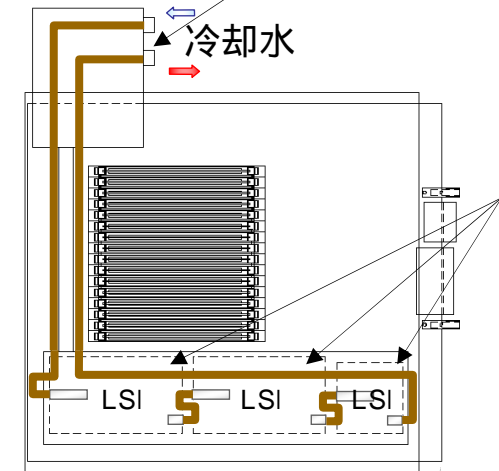
■ 高密度水冷機構

- 小型コールドプレート, 小型冷却水カプラを開発し, 従来のは水冷機構にはない高密度実装を実現
- 空冷用ダクト空間削減による実装高密度化

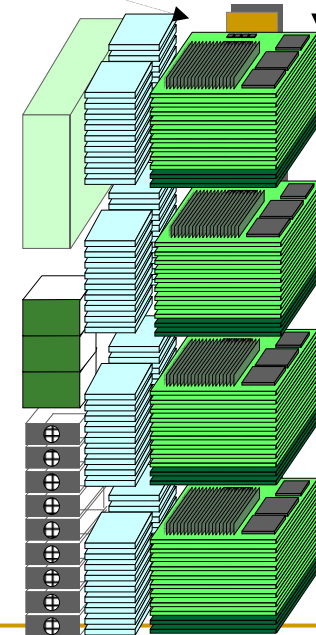
■ ボード活性交換との両立

- 故障したシステムボード(SB)の交換を, 他SBの稼働を継続したまま行える冷水制御, 漏水防止機構の開発

【システムボード(SB)】 冷却水カプラ



【計算機筐体】



【ユニットA】部品単位水冷技術 (2/2)

- 半導体の稼働温度を下げて故障率を大幅に低減
 - 空冷装置と比較してCPUの固定故障率を約0.006倍へ低減(理論計算値)
- 半導体の稼働温度を下げてリーク電流を低減
- 水冷による高効率な冷却の実現
 - 効率良く熱を部屋外へ移動できるため、データセンターで生じている廃熱問題を解決
 - 冷却の高効率化は、計算機本体だけでなく空調設備の省電力化、静音化、省スペース化も実現
- 小型～大型/ブレードサーバへの展開
 - 省電力化、静音化、高密度実装へ寄与
- 部品単位水冷開発スケジュール

| 2007年度 | 2008年度 | 2009年度 | 2010年度 |
|---------|--------|--------|--------|
| 方式/部品設計 | 試作/評価 | 製造 | 実機検証 |

【ユニットA】 SIMD化コンパイラの開発要素技術 (1/2)

- SIMD機構の活用:
コンパイラの命令スケジューリング機能を活用
 - 並列化オーバーヘッドのない細粒度の並列実行
 - Basic, Extendの両ユニットで2並列実行
 - 2演算/1命令
 - Basic, Extendを独立使用し, 条件実行時も両ユニットを並行動作
 - SIMD機能の2演算を独立に使用し, 条件付演算もSIMD機構を活用して並列実行

```
DO I=1,N
  IF (条件(I)) then
    A(I)=B(I)+C(I)
  ELSE
    X(I)=Y(I)*Z(I)
  ENDIF
ENDDO
```



```
DO I=1,N,2
  IF (条件(I)) then
    IF (条件(I+1)) then
      A(I)=B(I)+C(I)  A(I+1)=B(I+1)+C(I+1)
      SIMD演算
    ELSE
      A(I)=B(I)+C(I)  X(I+1)=Y(I+1)*Z(I+1)
      BASIC側で計算  EXTEND側で計算
    ENDIF
  ELSE
    IF (条件(I+1)) then
      X(I)=Y(I)*Z(I)  A(I+1)=B(I+1)+C(I+1)
      BASIC側で計算  EXTEND側で計算
    ELSE
      X(I)=Y(I)*Z(I)  X(I+1)=Y(I+1)*Z(I+1)
      SIMD演算
    ENDIF
  ENDIF
ENDDO
```

- プリフェッチによるメモリアクセス高速化
 - L2, L1の両キャッシュ向けに 目的に応じてプリフェッチ命令を配置

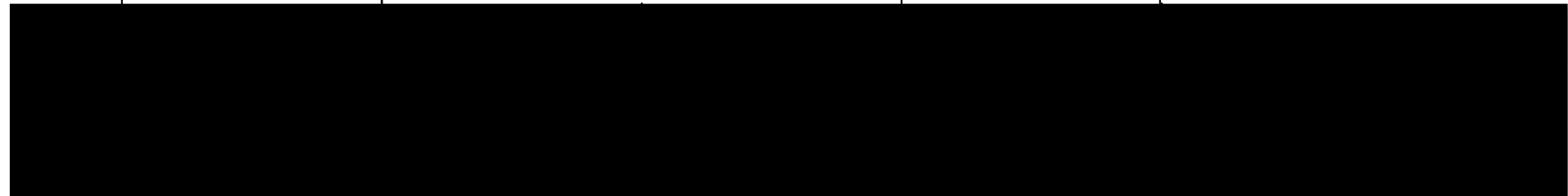
【ユニットA】SIMD化コンパイラの開発要素技術 (2/2)

- 自動並列化機能:
Venusの8コアを1つの高性能コアのように活用
 - 最内ループの並列化
 - コア間共有キャッシュ, コア間高速同期機構の活用
粒度の小さい最内ループも自動並列化
 - ベクトルより広範囲の適用可能性を追求
 - マルチスレッドの特徴
 - 制御オーバーヘッドを縮小し, 回転数の少ないループでも高速化が可能
回転数(ベクトル長)の確保が必須なベクトルに対する優位性

■ SIMD化コンパイラ開発スケジュール



| 2007年度 | 2008年度 | 2009年度 | 2010年度 |
|--------|---------------|--------|-----------------|
| 詳細設計 | プログラミング/ 機能検証 | | 実機検証 /チューニング |

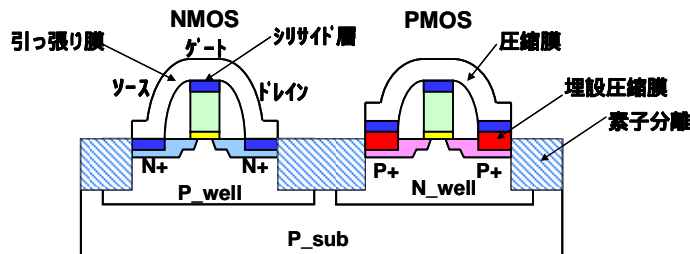


【ユニットB】論理LSI技術(1/2)

(1) 要素技術の内容

- ・最先端半導体プロセス

トランジスタ構造

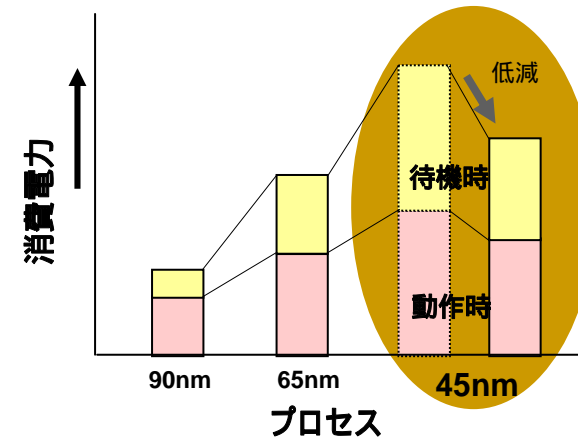


論理LSI概要

| | |
|----------|--------------------|
| プロセス | 45nm CMOS, 13層Cu配線 |
| 消費電力 | 最大140W |
| 総トランジスタ数 | 最大7億 |
| 実装方式 | フリップチップ, ペアチップ実装 |
| クロックサイクル | 2GHz |

- ・LSI回路技術
 - ✓高速・高集積LSIを実現する回路技術
 - ✓動作時/待機時両方の消費電力低減技術
- ・LSI設計技術
 - ✓高い設計品質の確保を目指した検証技術
 - ✓高速・高集積を実現するレイアウト設計技術

消費電力の低減



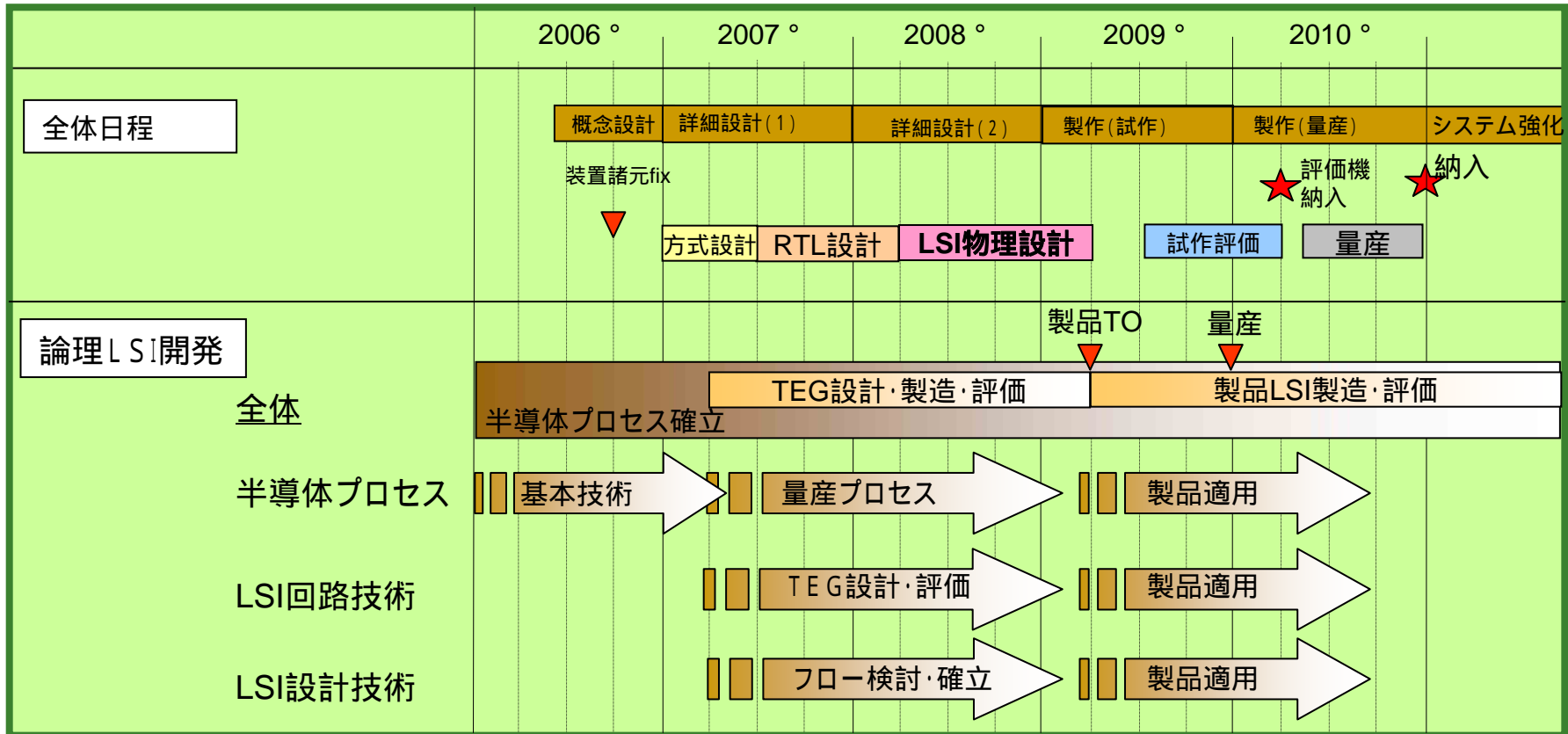
(2) 革新性, 発展性, 技術力強化への寄与

- ・最先端・高性能45nmCMOSプロセスの採用 (歪みシリコン, Low etcの先進技術適用)
- ・高速・高集積回路技術確立 (高速内部セル/SRAM, 高速インタフェース, 高速クロック分配etc)
- ・高速性と両立する低消費電力回路技術の確立 (マルチVth, クロックゲーティングetc)
- ・上記確立した技術を, 他の製品領域 (高性能サーバー, デジタル家電etc) に展開

【ユニットB】論理LSI技術(2/2)

論理LSI開発日程

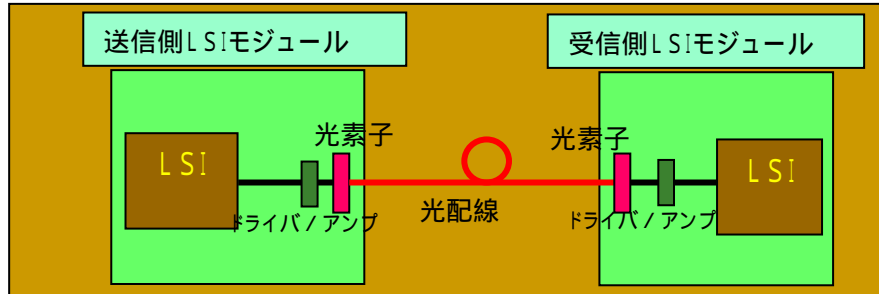
| | | |
|--------------|---|------------|
| TEG設計開始 | ; | 2007年度/1Q末 |
| 製品LSI テープアウト | ; | 2009年度/1Q末 |
| LSI製造量産開始 | ; | 2009年度/4Q末 |



【ユニットB】光インターコネクト技術(1/2)

(1) 技術開発内容

LSI間の信号伝送を従来の電気伝送に変わって**光で伝送**, 部分**水冷技術**



具体的な研究項目

- ・高速光素子に関する開発
発光素子, 受光素子
- ・高密度実装に関する開発
高速光電気実装技術, 高密度光配線技術, 冷却モジュール技術, 他

(2) 革新性, 発展性, 技術力強化への寄与

開発目標

- ・高速化: 20Gbps超SerDes技術 **ITRS予測の2倍**
電気伝送の限界5~10Gbpsを打破
- ・小型化: 1000信号/LSI **従来の1/200**
- ・消費電力: **従来比1/100**
- ・高信頼度, 低コスト化



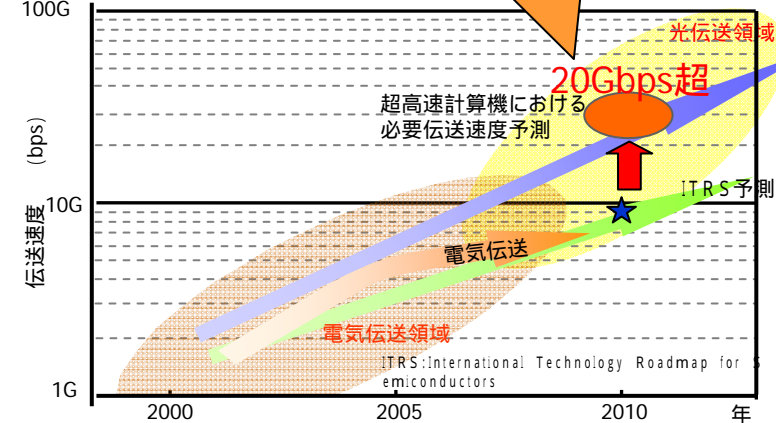
次世代スパコン実現に寄与

さらに

- ・ハイエンドサーバ, ファイルサーバネットワーク機器
 - ・PC, 携帯端末, 情報家電
 - ・医療機器, 車 ……
- への活用が期待される

光技術よりITRS予測の2倍の高速化を目標

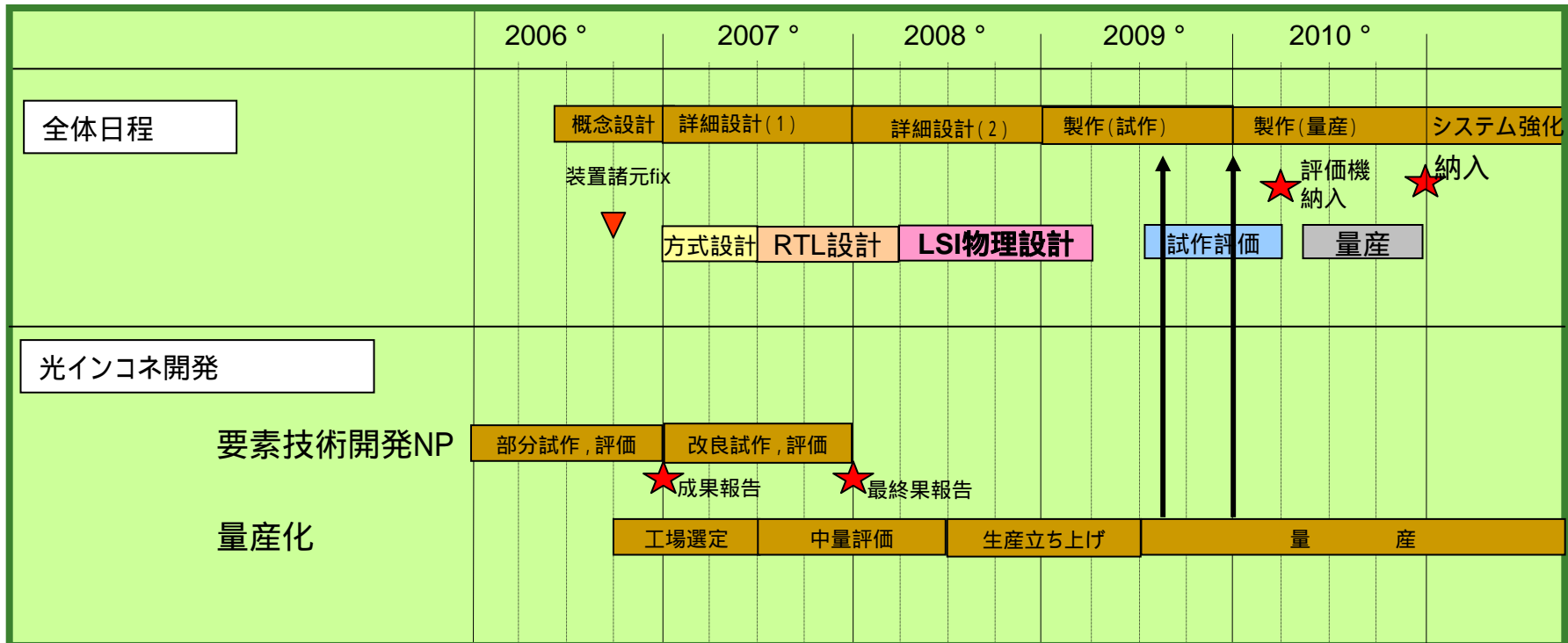
インターコネクト技術のトレンド



【ユニットB】光インターコネクタ(2/2)

光インターコネクタ開発日程

| | |
|---------------|--------------|
| 要素技術開発完了 | : 2007年度/4Q末 |
| インターフェイス仕様FIX | : 2007年度/4Q末 |
| 量産工場選定 | : 2007年度/2Q末 |
| 中量評価完了 | : 2008年度/2Q末 |
| 生産立ち上げ | : 2009年度/2Q末 |
| 量産開始 | : 2009年度/3Q~ |



【ユニットB】キャッシュ対応ベクトル化コンパイラ(1/2)

(1) 技術開発内容

データをキャッシュに登録するかしないかをソフトウェア(コンパイラ)から制御する技術で、再利用性の高いデータを選んでキャッシュに登録することにより、キャッシュ容量の有効利用を図る。

具体的な技術項目

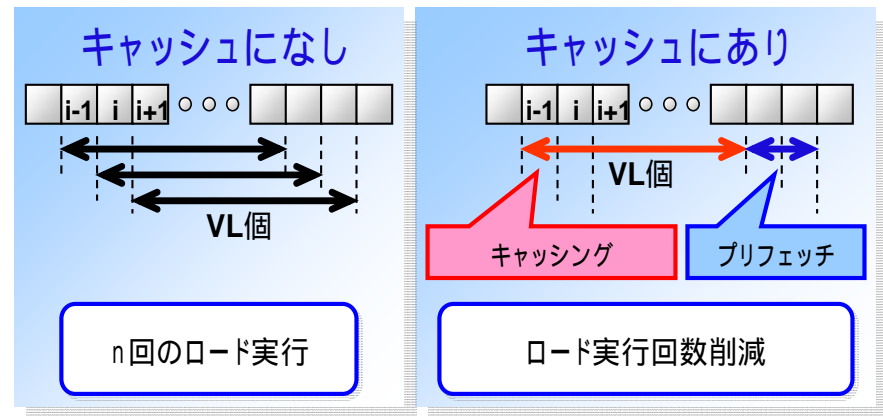
- ・自動的に適用配列を選択する技術
- ・自動的に非適用配列を選択する技術
- ・指示行で適用配列を指定する技術
- ・指示行で非適用配列を指定する技術

```
Program
DO i = 1, n
...
= ...+B(i-1)+...
= ...+B(i)+...
= ...+B(i+1)...
...
END DO
```

(2) 革新性, 発展性

キャッシュ対応技術により、メモリバンド幅維持と短レイテンシ化を実現。

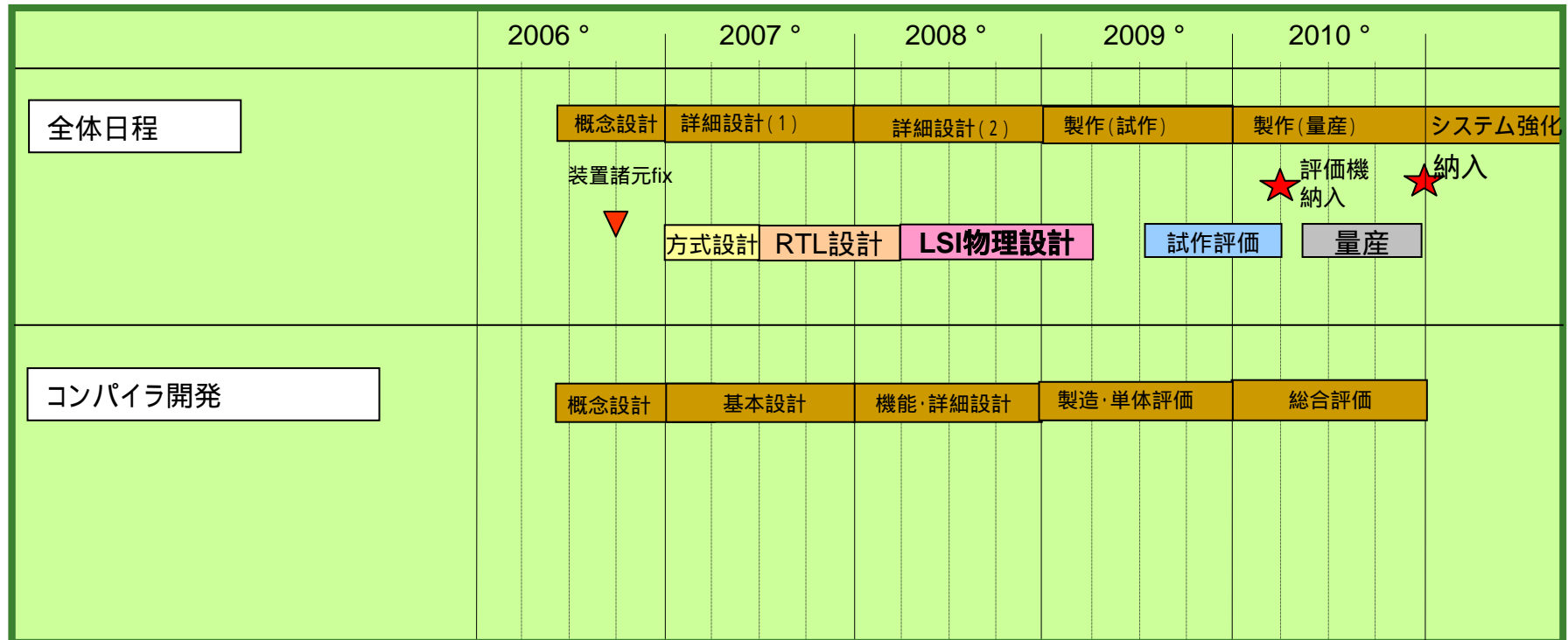
- ・登録データの選択によるキャッシュの有効利用
 - メモリレイテンシ隠蔽
 - キャッシュ格納済み配列のロードバンド幅高
 - バンク競合率低減
 - 短ベクトル長時の実行性能向上
- ・SWプリフェッチ機能によるメモリレイテンシ隠蔽



【ユニットB】キャッシュ対応ベクトル化コンパイラ(2/2)

コンパイラ開発日程

基本設計完了 : 2007年度/4Q末
 機能・詳細設計完了 : 2008年度/4Q末
 製造・単体評価完了 : 2009年度/4Q末
 総合評価完了 : 2010年度/4Q末



2. システム構成案の妥当性

(1) システム構成案の詳細及び性能

システム構成案は、革新性、発展性、拡張性及び展開性を有するものであるか。また、我が国が継続的にスーパーコンピュータを開発していくための技術力の強化に寄与するものであるか。

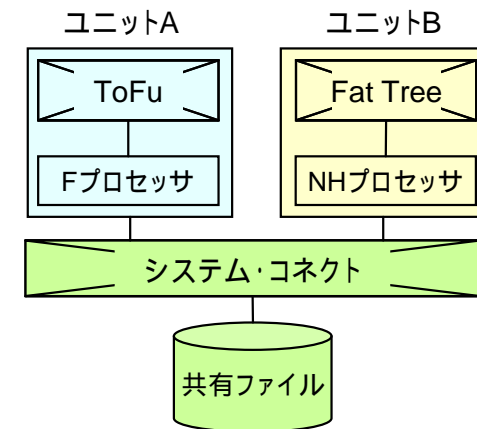
本システムの基本構成と特長(1/3)

F及びNH両者の提案は、それぞれに優れた特長を持ち、一方のみを選択することは、スパコン技術の将来の可能性とリスクを考慮すると適切とは言えず、共同開発の条件を満たし、かつ技術的な実現見通しがついたことから、両者の技術を開発して、一つのシステムを構成することが最善と判断。



- 演算加速機構を付加すること等により高性能化したスカラプロセッサ及び拡張性の高い新規のネットワーク構成によるスカラユニット(理論性能約10PFLOPS超)と、画期的な構成により高い演算性能を達成するベクトルプロセッサ及び汎用性・運用性の高い省電力ネットワーク構成によるベクトルユニット(理論性能約3PFLOPS超)の両者をシステムコネクで結合した統合汎用システムを開発する。
- これにより、少なくとも目標性能であるLinpackの実行性能10PFLOPSを達成する。

統合汎用スーパーコンピュータシステム



本システムの基本構成と特長(2/3)

- 将来の我が国のスパコン開発の技術力, 国際競争力等の向上に一層貢献する.

世界的主流となっているスカラプロセッサに演算加速機構を付加したプロセッサと, 我が国が強みを持つベクトルプロセッサの改良型となる新しい汎用プロセッサを同時に開発することにより, 次世代以降のプロセッサの技術オプションを発展させ, 将来に向け国際競争力の一層の向上を図る(= 次々世代の開発に繋げるために必要) .

世界的に大多数のスパコンシステムが採用しているスカラプロセッサにSIMD型演算加速機構を付加すること等によりさらに高性能化を図り, 将来に向け国際競争力を高めることは極めて重要.

制御構造が単純なベクトルプロセッサは, スカラプロセッサに比べ演算性能を向上させることがより容易であり, 将来に向け更に高度化を図り技術を発展させることにより, 我が国の重要な基幹技術の一つとなり得る.

FとNH両者の競争関係が維持され, 国際的により強い技術を開発できる可能性が高まるとともに, 次々世代以降のCPU開発においては, 両者の技術を融合させることも視野に入れることが可能.

本システムの基本構成と特長 (3/3)

- ユニットAのスカラプロセッサは、SIMD型演算加速機構などを付加した高性能プロセッサであり、将来的にも高い国際競争力を持っている。
- ユニットAのToFuネットワークは、完全結合と三次元トラスを組合わせた新規のネットワークであり、容易に大規模構成が可能な拡張性の高いネットワークである。
- ユニットBのプロセッサは、我が国が強みを持つベクトルプロセッサを改良したプロセッサである。制御構造が単純で、演算性能を向上させることが容易であり、さらなる高度化を図り、技術を発展させて我が国の重要な基幹技術へと発展させることができる。
- ユニットBのFat-Treeは、光インターコネクションによる汎用性、運用性に優れた省電力ネットワークである。
- 本システムは性格の異なったシステムを、システム・コネクで結合し、統合システム機能により、多様なアプリケーションへの対応が可能かつ計算資源を有効に活用することができる統合汎用スーパーコンピュータシステムである。
- スカラとベクトルの両プロセッサの技術を発展・保持できることで、次々世代以降のプロセッサ開発において、両技術による新しい汎用プロセッサの開発を視野に入れることが可能になる。

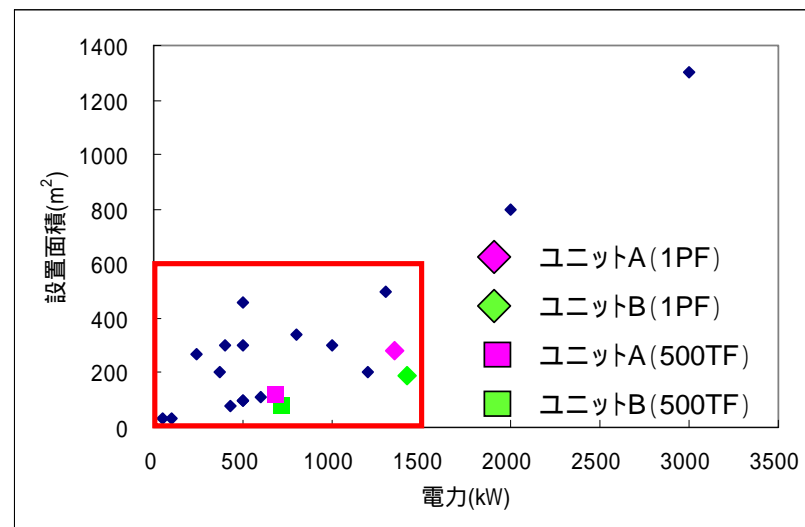
2. システム構成案の妥当性

(1) システム構成案の詳細及び性能

システム構成案は、それを基に大学や研究機関向けの計算機システムを構築することを可能とするものか。また、それを実施する場合に、消費電力、設置面積及び将来の拡張性の面で、適当なものとなるか。

下方展開について

- 両案とも,スーパーコンピュータセンター調査で明らかになったほとんどの計算センターの制約条件である設置面積600m²以下,消費電力1.5MW以下を達成.



(参考)筐体単位の消費電力および設置面積

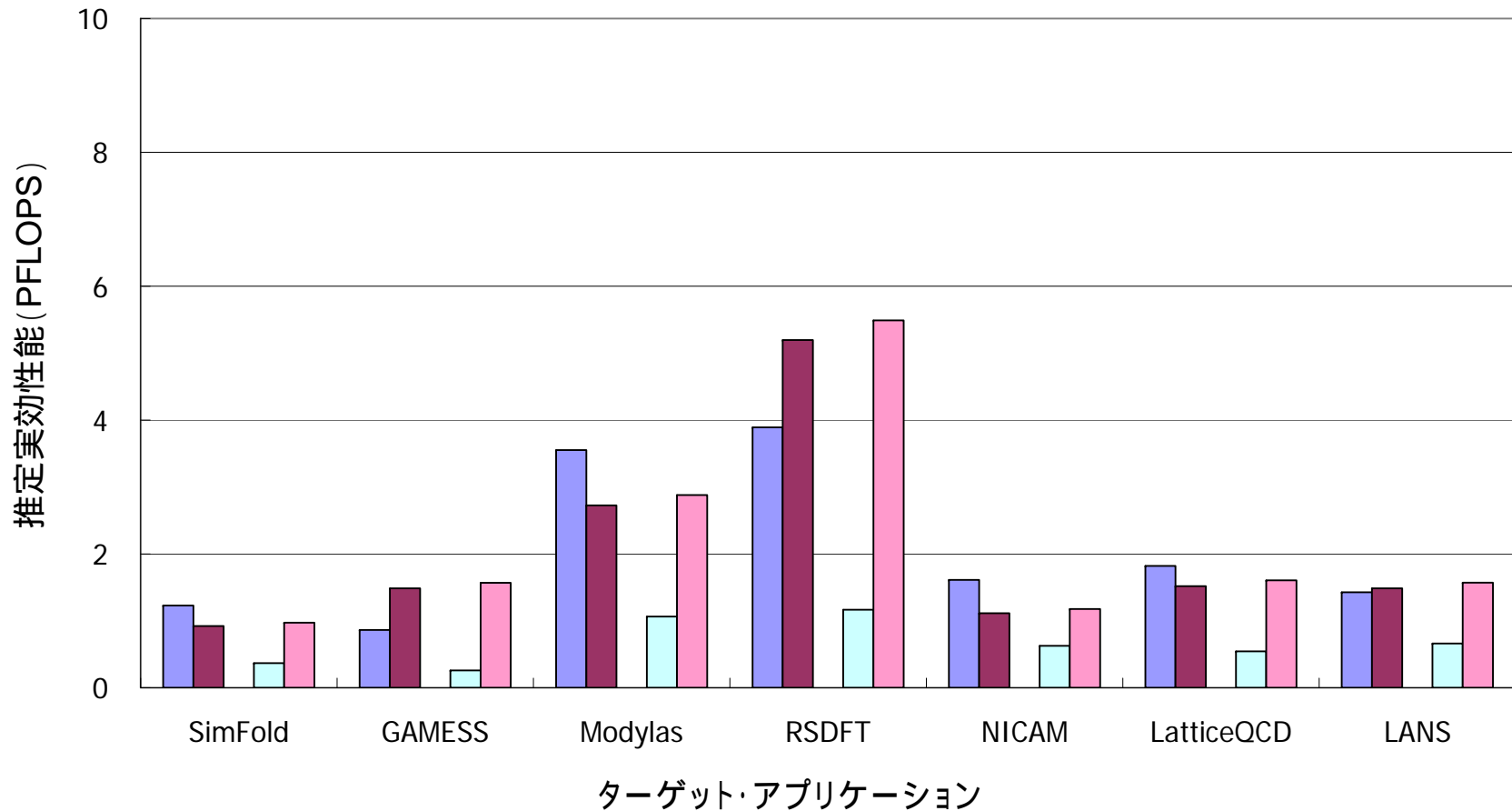
| | 演算性能(TF) | 消費電力(KW) | 設置面積(m ²) |
|-------|----------|----------|-----------------------|
| ユニットA | 約18 | 約24 | 1.2 |
| ユニットB | 約16 | 約21 | 2.0 |

2. システム構成案の妥当性

(2) システムの機能

ターゲットアプリケーションについての実効性能は、十分であると評価されるか。

ターゲット・アプリケーションによる性能推定



- NH案: ピーク性能10.48PFLOPS時の推定性能
- F案: ピーク性能10.61PFLOPS時の推定性能
- 統合システム(ユニットA): ピーク性能11.2PFLOPS時の換算値(ピーク性能比)
- 統合システム(ユニットB): ピーク性能3.14PFLOPS時の換算値(ピーク性能比と約1B/FLOP性能向上を考慮)

2. システム構成案の妥当性

(2) システムの機能

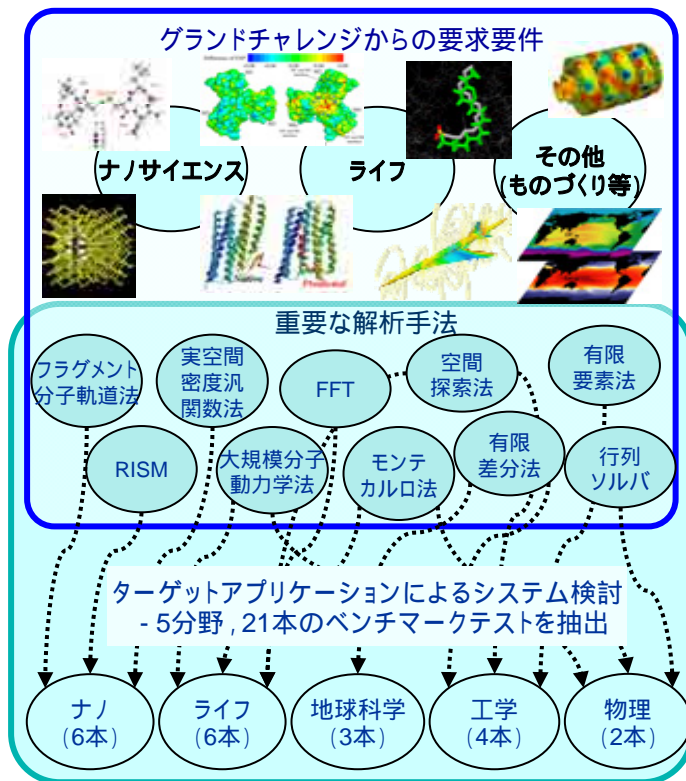
その他の広範な分野におけるアプリケーションについても十分な実効性能を出すことが可能か。

ターゲット・アプリケーションの選定

- 次世代スーパーコンピュータのアーキテクチャ検討に資するため、2010年頃に重要となるアプリケーション・ソフトウェアを検討。
- 次世代スーパーコンピュータ開発戦略委員会の下にアプリケーション検討部会を設置し、ターゲット・アプリケーションを選定。
 - 平成18年1月から平成19年3月までに計7回の会議を開催。
 - 5分野からターゲット・アプリケーション21本を選定。
 - その他の討議事項
 - 概念設計における運用・管理システム検討のための運用指針(案)の検討
 - システム構成案について
 - COE形成について

評価したターゲット・アプリケーション

- 代表的なアルゴリズムを含む主要なアプリケーションを評価した。その他の広範なアプリケーションにおいても、高い実効性能が得られると推測。



| ベンチマーク | 分野 | アプリケーション |
|------------|--------|-----------------|
| SimFold | ライフ/ナノ | ライフタンパク質立体構造の予測 |
| GAMESS/FMO | | 分子軌道法計算 |
| Modyas | | 高並列汎用分子動力学計算 |
| RSDFT | | 実空間第一原理分子動力学計算 |
| NICAM | 地球科学 | 全球雲解像大気大循環モデル |
| LatticeQCD | 物理 | 格子QCDシミュレーション |
| LANS | 工学 | 圧縮性流体計算 |

- アプリケーションの特性に適合したユニットを選択し、高い実効性能を得ることが出来るシステムである。

空白・ページ

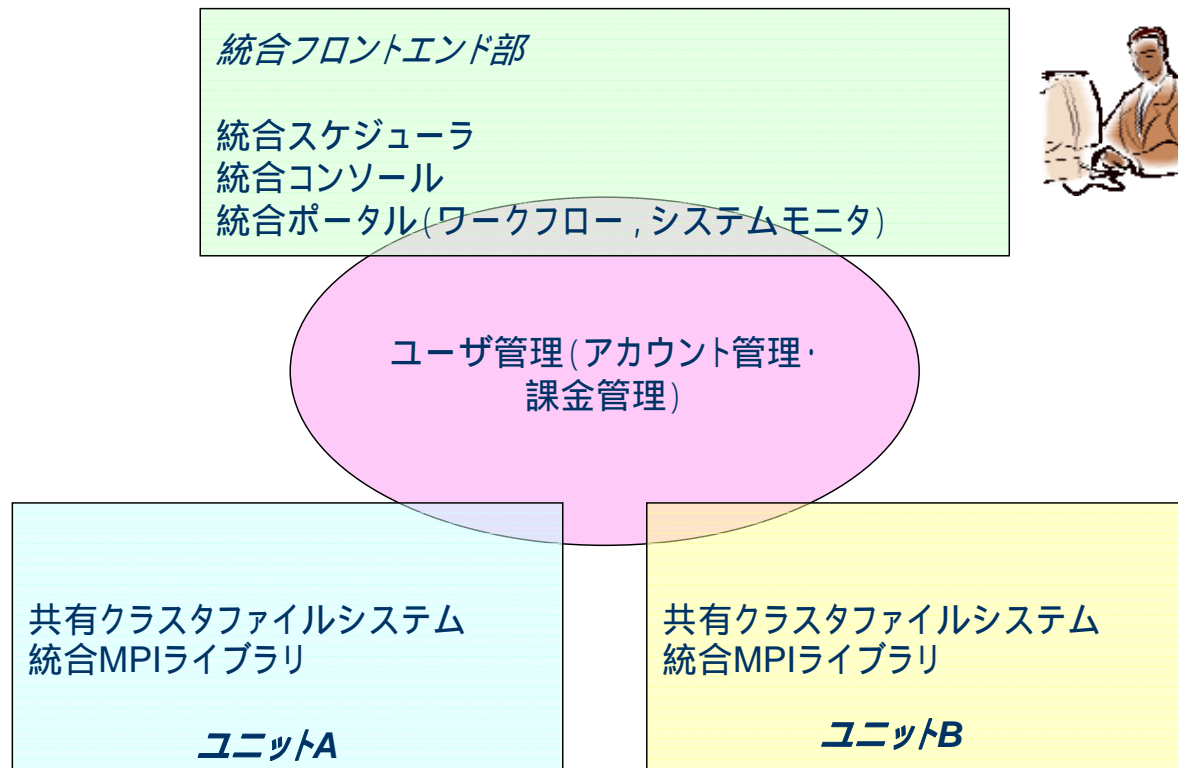
2. システム構成案の妥当性

(2) システムの機能

システムソフトウェア(OS, ライブラリ, コンパイラ等)はシステムの性能を十分引き出すものであるか.

統合システム・ソフトウェアの機能

- 統合汎用スーパーコンピュータシステムの効率良い運用のために、以下の統合システム・ソフトウェア機能の開発を予定.



システムソフトウェアの機能【ユニットA】

■ コンパイラ

■ プロセッサ内のSIMD機構の有効活用

- 命令スケジューリング機能を応用し、オーバーヘッドのない細粒度の並列実行を実現
- SIMD拡張機能の効率的な並列実行を実現

■ 自動並列化

- 1CPU内のマルチコアを自動並列化により高速単一コアのように利用
- 並列化に伴うオーバーヘッドを最小化しプログラム実行を高速化

■ ライブラリ

■ MPIライブラリの最適化

- 並列度の大規模化への対応, 及び, ネットワークトポロジに適した通信性能の実現
- 計算ノード間の集合演算支援機能による通信性能の最適化

■ 科学技術計算用ライブラリの最適化

- 新規開発プロセッサに適合した最適化, マルチコア効率利用・分散メモリ型の双方の最適化

システムソフトウェアの機能【ユニットB】

■ コンパイラ

■ プロセッサ内のベクトル演算機構とRDB機能の有効活用

- 自動ベクトル化機能
- RDB機能を用いたメモリアクセスオーバーヘッドの削減

■ 自動並列化

- 1CPU内のマルチコアを自動並列化により高速単一コアのように利用
- 並列化に伴うオーバーヘッドを最小化しプログラム実行を高速化

■ ライブラリ

■ MPIライブラリの最適化

- 並列度の大規模化への対応, 及び, ネットワークトポロジに適した通信性能の実現
- 計算ノード間の集合演算支援機能による通信性能の最適化

■ 科学技術計算用ライブラリの最適化

- 新規開発プロセッサに適合した最適化, マルチコア効率利用・分散メモリ型の双方の最適化

2. システム構成案の妥当性

(2) システムの機能

システムソフトウェア(OS, ライブラリ, コンパイラ等)は幅広い利用者が利用することが可能なものか.

空白・ページ

システムソフトウェアの汎用性【ユニットA】

- OS
 - POSIX規格に準ずるUNIX系オープンOSの採用
- コンパイラ
 - Fortran95規格, C99規格, JIS規格C++等の標準規格に準拠
 - OpenMPのAPIをサポート
- ライブラリ
 - MPIライブラリの標準規格準拠
 - 並列言語HPF, CAF, XPF(富士通の現有並列言語)等のサポート
 - BLAS, PBLAS, ScaLAPACK等, 広く利用されている標準的な科学技術計算用ライブラリのサポート

システムソフトウェアの汎用性【ユニットB】

- OS
 - I/Oノードに汎用性高いLinuxを適用
 - 計算ノードにPOSIX規格に準ずるUNIX系OSを採用
- コンパイラ
 - Fortran95規格, HPF2.0, C99規格, JIS規格C++等の標準規格に準拠
 - OpenMP API 2.5に準拠
- ライブラリ
 - MPI-1, MPI-2に準拠したMPIライブラリ
 - BLAS, PBLAS, ScaLAPACK等, 広く利用されている標準的な科学技術計算用ライブラリのサポート

2. システム構成案の妥当性

(3) システムの運用

計算機資源の効率的な配分等により、多数の利用者がシステムを多様な用途に利用することが可能か。

統合システムとしての運用ソフトウェア機能

■ 統合フロントエンド部

■ 統合スケジューラ

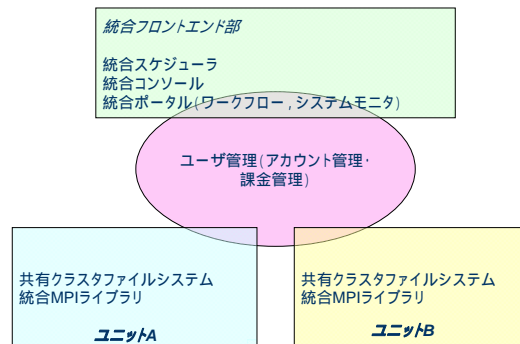
- メタスケジューラ機能
 - ☞ 各ユニットのローカルスケジューラの統合
 - ☞ ファイルのステージング連動
- 資源予約機能
 - ☞ ユニットAとユニットBの資源を同時予約し連携ジョブを実行

■ 統合コンソール

- ソフトウェア構成管理
- パーティション管理
- 運用モード設定管理
 - ☞ チェックポイント取得 & マイグレーション

■ 統合ポータル

- ワークフロー
 - ☞ ユニット間連携計算自動スケジューリング
 - ☞ ファイルのステージング支援
- システムモニタ
 - ☞ ジョブ状況表示



■ 統合フロントエンド部(続き)

■ 統合プログラム開発環境

- クロスコンパイラ
- デバッグツール
- チューニングツール

■ 共通機能

- ユーザ管理
 - アカウント管理
 - 課金管理
- ACL機能

■ 各ユニット

- ローカル・スケジューラ
- 共有クラスタファイルシステム
- 統合MPIライブラリ
 - 共通API仕様
 - ユニット間高速通信インターフェース

システム運用：資源の効率的な配分【ユニットA】

■ ジョブマッピング

- 通常は18CPUからなるシャシ単位で計算ノードの割り当てが可能
- パラメタスイープ型のアプリケーションに対しては1CPU単位の割り当てが可能

■ パーティショニング

- 仮想3次元トラスを2シャシ単位で複数の小規模3次元トラスに分割し、パーティションを形成
- パーティション毎に計算資源を管理、ジョブスケジューリングを実施

■ 運用ソフトウェア

- 計算資源を一括して管理し、計算ノード、ファイルシステムと連携した資源管理、及び、ジョブ実行管理を実現

システム運用：資源の効率的な配分【ユニットB】

- ジョブマッピング
 - 32CPUからなるNノード単位で計算ノードの割り当てが可能
 - パラメタスイープ型のアプリケーションに対しては1CPU単位の割り当てが可能
- パーティショニング
 - Nノード単位でパーティションを形成
 - パーティション毎に計算資源を管理，ジョブスケジューリングを実施
- 運用ソフトウェア
 - 計算資源を一括して管理し，計算ノード，ファイルシステムと連携した資源管理，及び，ジョブ実行管理を実現

2. システム構成案の妥当性

(3) システムの運用

システムの部分的な故障時等に、全体の運用に影響を及ぼさない仕組みは構築されているか。また、迅速な修理等は可能か。

空白・ページ

システム運用:RAS機能【ユニットA】

- CPU
 - キャッシュ部でのECC機能,内蔵RAM全体での徹底したパリティチェックと自動修正機能によりデータ一貫性を確保
 - 演算部ではパリティチェックに加え剰余チェックによるデータ保護,さらに命令リトライ機能により実行結果を保証
 - これら高信頼設計を徹底することでメインフレーム計算機レベルの信頼性を達成
- 計算ノード間ネットワーク
 - 障害リンク,及び障害ノードの検出と回避ルートへの自動切り替え機能
 - 障害発生時にも仮想的な3次元トーラスのユーザビューを維持
- ストレージ・ファイルシステム
 - ディスク,及び計算ノードからのパスの二重化によるフェイルオーバ
- 運用ソフトウェア
 - 計算ノード,ファイルシステム,フロントエンド,及びシステム制御サーバの的確な連携とシステム全体の信頼性の確保

システム運用:RAS機能【ユニットB】

- CPU/メモリ
 - ハードウェア診断回路
 - ECCチェック, パリティチェック, 2重化チェック, MOD-Nチェック, Out-of-Nチェック, 制御回路のシーケンス/タイミング/タイムアウトチェック, Built In Self Test回路
 - 診断プログラム
 - 自動診断プログラム実行によるパトロールチェック機能
 - 障害時の縮退運転
 - 汎用機プロセッサ並みの故障診断機能および故障検出率を達成
- 計算ノード間ネットワーク
 - ノード間通信のECCによるエラー訂正
 - スイッチ障害時のプレーン切り離しによる縮退運転
- ストレージ・ファイルシステム
 - ディスク, 及び計算ノードからのパスの二重化によるフェイルオーバ
 - RAID6の採用
- 運用ソフトウェア
 - 計算ノード, ファイルシステム, フロントエンドの的確な連携とシステム全体の信頼性の確保