秘

評価後回収資料

複写厳禁

資料6-3

「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5月9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．

⇨ 本プロジェクトで開発されたシステムは，広範な応用分野の研究開発基盤として利用されるものであり，また我が国の競争力を高める目的を持つことから，我々のシステム開発方針及び種々の基本条件を考慮したシステム最適化の考え方は妥当と考える．

システム開発の方針

プロジェクトの基本方針

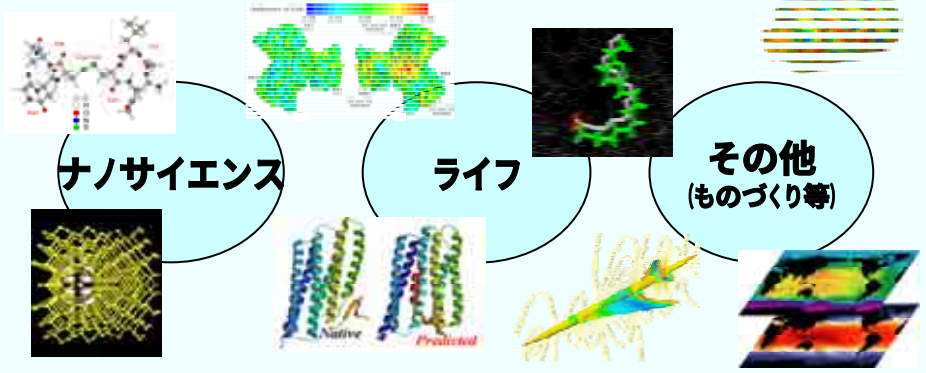
- 計算機シミュレーションにより，科学技術･産業の競争力を維持，高めること．
- スーパーコンピュータの開発力を国内に保持し，継続的な開発を可能とすること．
- 完成時に世界最速と内外から広く認められること．

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．

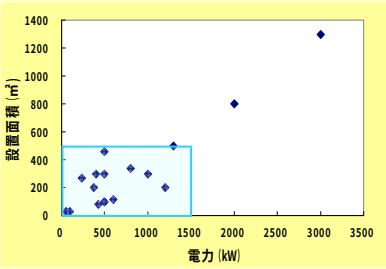
システム最適化の考え方

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

フラグメント分子軌道法
大規模分子動力学法
高速変換
空間探索法
有限要素法
RISM
実空間密度汎関数法
モンテカルロ法
有限差分法
行列ソルバ

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

ナノサイエンス (6本)	ライフ (6本)	環境・防災 (3本)	工学 (4本)	物理・天文 (2本)

【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

ブランク・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を達成する(平成23年6月のスーパーコンピュータサイトTOP500でランキング第1位を奪取)．
- HPC CHALLENGE全28項目中，過半数以上の項目で最高性能を達成する．

- **本システム構成案は，汎用性，アプリケーションからの要求，電力・設置面積等の制約条件，費用対効果，革新性，発展性，展開性等を重視しつつ，Linpackで10PFLOPSを達成するために最適なものと考えている．**
- LINPACK 10PFLOPSについては，以下の試算により達成可能．

＜ユニットA＋ユニットBでLINPACK 10PFLOPS超を達成＞

【試算】

- ユニットA: 11.2PFLOPS（ピーク性能）x 85%（LINPACK効率）=9.52PFLOPS
- ユニットB:　3.1PFLOPS（ピーク性能）x 90%（LINPACK効率）=2.79PFLOPS

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については，資料6-1のとおり．**

システム構成案比較

		単独で Peak１３ＰＦ超		Peak１０ＰＦ超＋Peak３ＰＦ超		
		F	NH	案1	案2	案3
Linpack 性能	10PFLOPS	○	○	○	○	○
	Top1	○+ ×単独では 10P 超まで	○+ ×単独では 10P 超まで	△ ×スケジュール遅延の見込み	△ ×スケジュール遅延の可能性	○
ｱﾌﾟﾘ実効性能 (<3P)		○	○	◎	◎	◎
ｱﾌﾟﾘ実効性能 (3～10P)		○	○	○	○	○
ｱﾌﾟﾘ実効性能 (>10P)		○ ×単独では 10P 超まで	○ ×単独では 10P 超まで	△+	△+	△
汎用性		△+ CPU○ ネットワーク△	△ CPU△ ネットワーク○	○	○	○
アプリ資産の活用		△	△−	○	○	○
消費電力及び設置面積	消費電力	○	○	△ ×ルーターの消費電力	△ ×NIC 等の消費電力	○
	設置面積	○	○	△ ×ルーターの物量	△ ×NIC 等の物量	○
技術力の強化への寄与等	革新性	○	○	◎ ○異機種間の密結合	◎− ○異機種間の密結合(標準規格)	○+
	発展性	△	△	○ ×異機種間密結合の必要性？	○ ×異機種間密結合の必要性？	○
	拡張性	○	○	○	○	◎
	ビジネス展開性	○	○	△ ×F は単独展開困難	◎	◎
費用対効果		△ ×民間資金は半分	△ ×民間資金は半分	× ×密結合は利用が見込まれず	△ ×密結合は利用が見込まれず	○
スケジュール		○	○	× ×遅延の見込み	△ ×遅延の可能性	○
下方展開		○	○	△ ×F 単独の整備が困難	◎	◎
要素技術の波及効果		○	○	◎	◎	◎
多様なユーザーの効率的な利用		○	○	◎	◎	◎
(システムの一体性)		◎	◎	○	○−	△

2. システム構成案の妥当性

⑴ 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．

⇨ システム構成案の消費電力及び設置面積あたりの演算性能は，完成時点の技術水準や，下方展開性及び米国の動向等を踏まえて妥当と考える．

ブランク・ページ

消費電力及び設置面積あたりの演算性能

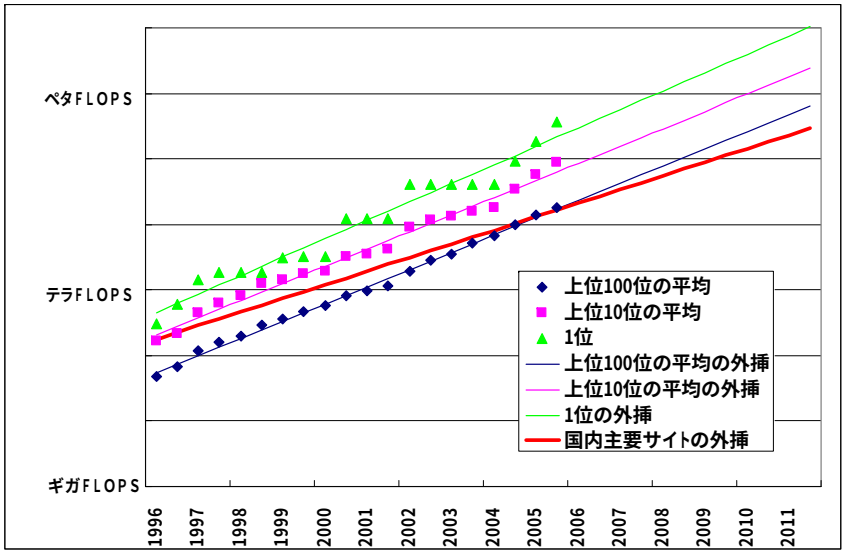
- **概念設計における性能目標**
  - **ピーク性能** 10PFLOPS**以上**
  - **メモリ容量** 2.5PB**以上**
  - **消費電力**　30MW**以下（周辺機器，空調機器を含む）**
  - **設置面積**　3,200㎡**以下（周辺機器を含む）**

- **統合汎用スーパーコンピュータシステムの概要**
  - **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**2,800㎡**以下（本体システム），約**3,800㎡（**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**193㎡/PFLOPS

概念設計の性能目標設定の根拠

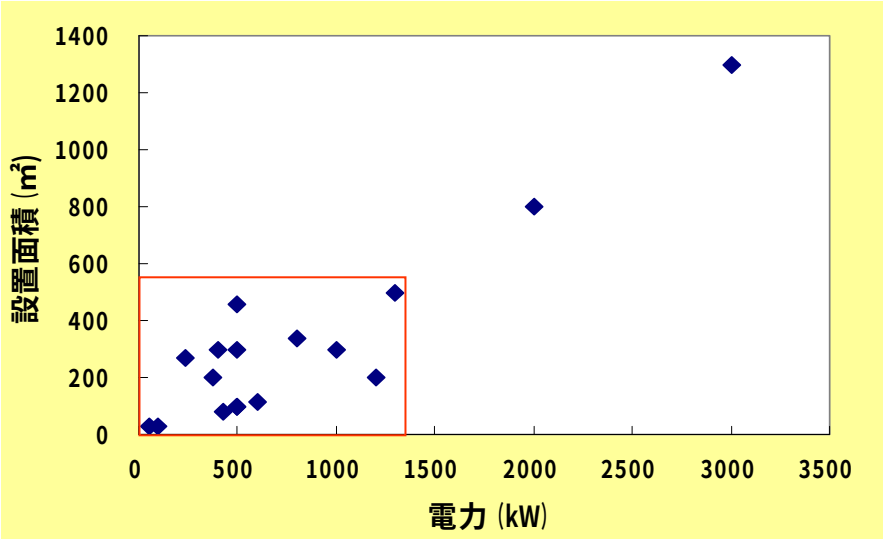
TOP500における国内外の計算機性能
上昇トレンドの比較



- 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)

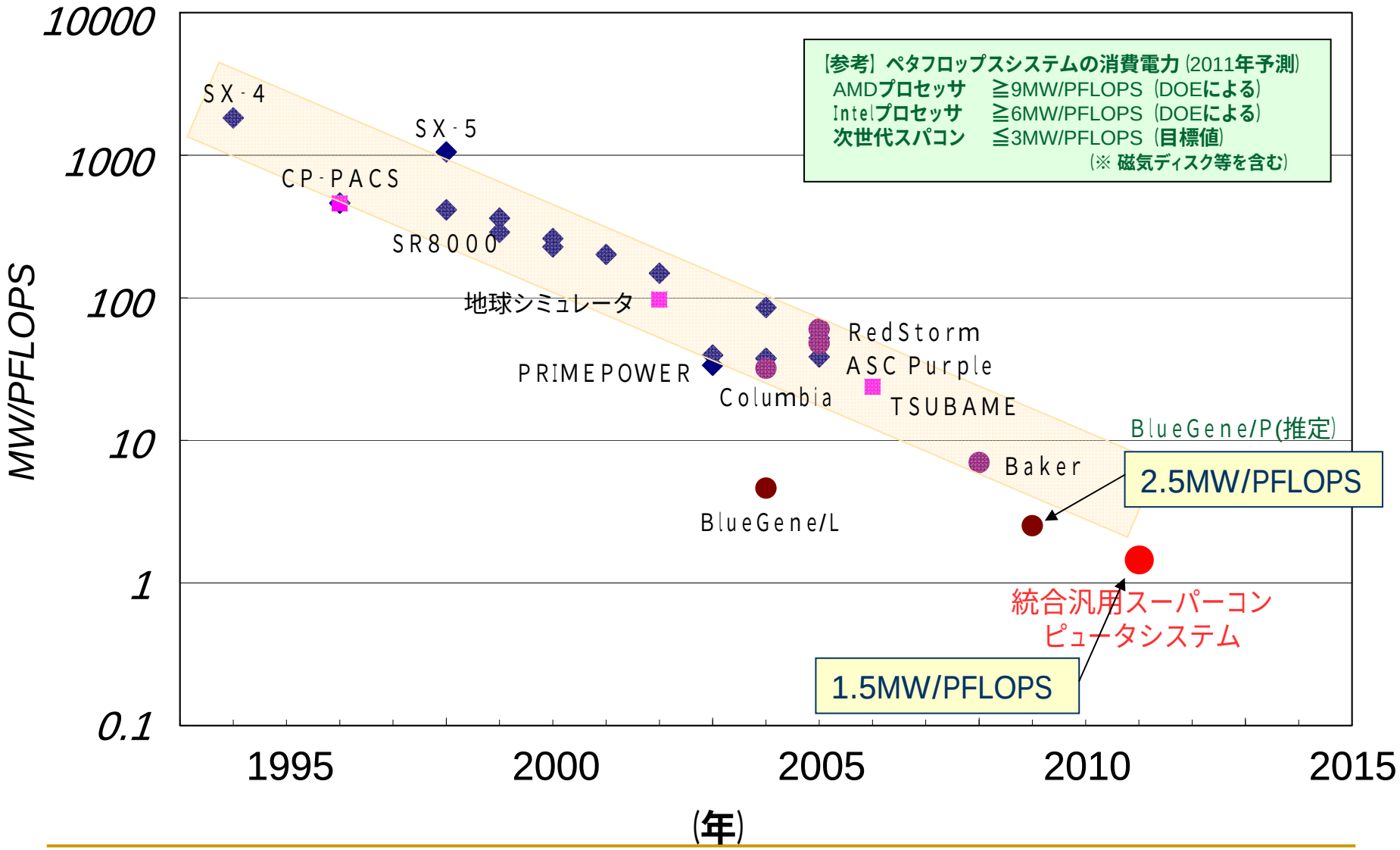
各センターの電力的および設置面積的制約



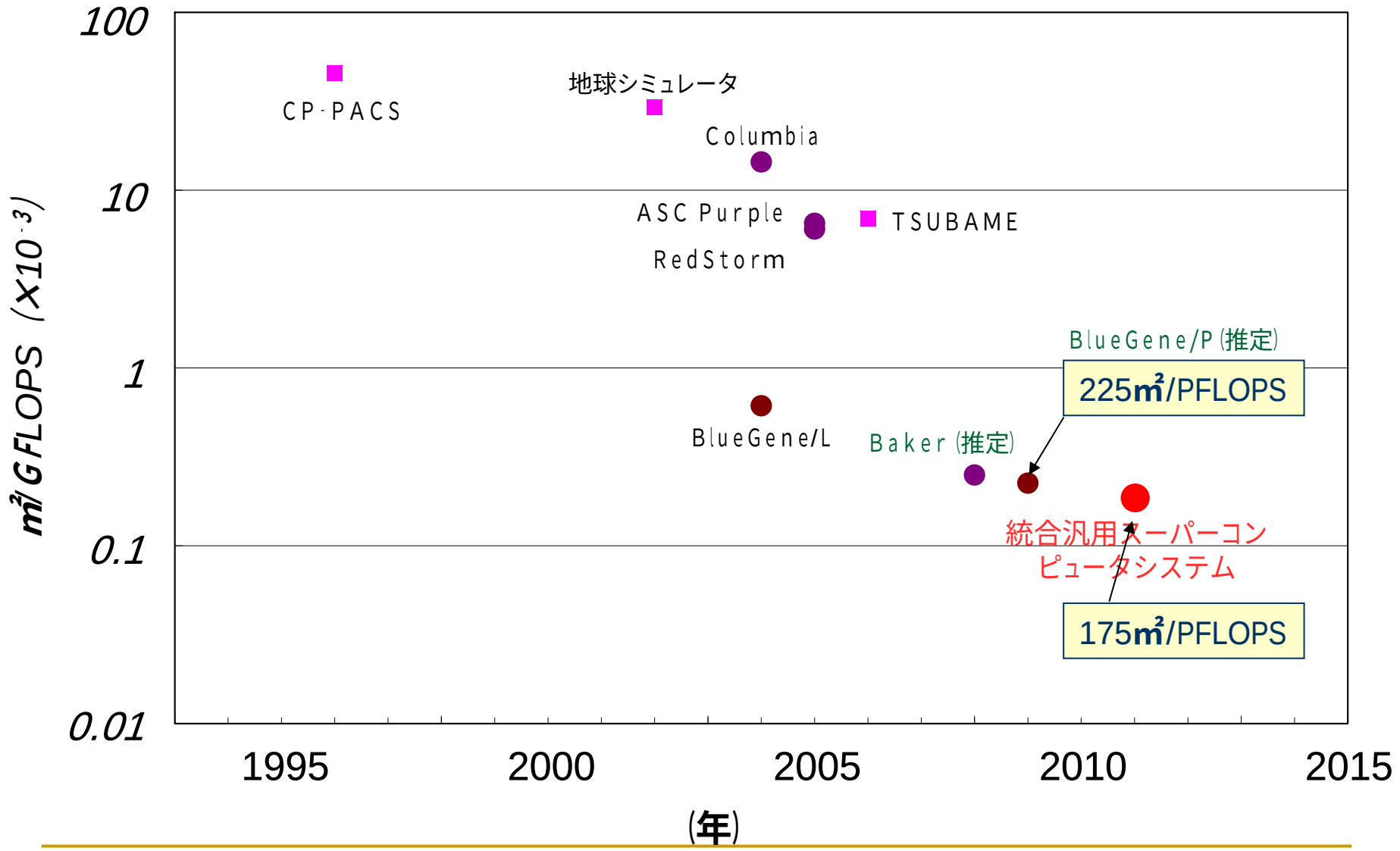
- 設置面積，受電設備許容量には強い制約がある．
- ほとんどのスパコンセンターの
  - •設置面積は約600㎡以下
  - •受電設備容量は1.5MW以下

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

⑵ システム構成案の詳細及び性能

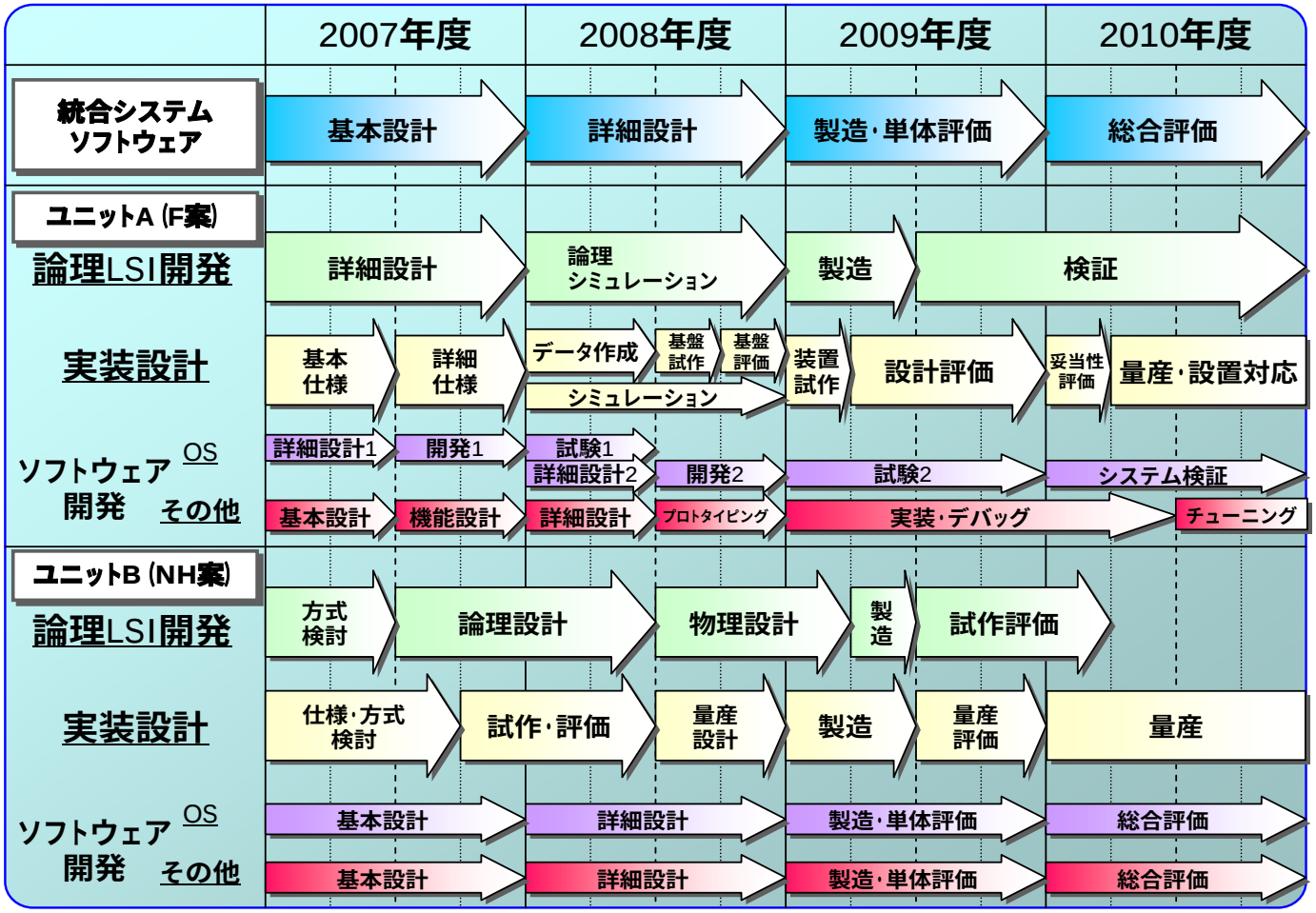
システム構成案を実現するための要素技術は，現在の技術水準及び今後の見通しから判断して，システムの製作時期までに開発可能か．

⇨ プロジェクトの全体スケジュールに合わせて，各ユニットを構成する要素技術の開発スケジュールが計画されており，システム製作開始時期までに開発可能と判断している．

※ 各要素技術の開発スケジュールは別紙．

要素技術開発の可能性について

- 各ユニットを構成する要素技術の開発スケジュールは，以下のとおりである．製作開始時期までに要素技術が開発されると判断している．



2. システム構成案の妥当性

⑴ 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．

⇨ 本システム構成案は，次頁のとおり，革新性，発展性，拡張性及び展開性に優れたものである．

　また，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技術オプションを発展させ，将来の国際競争力の一層の向上を図るものである．

- **革新性**：統合システムにより，複雑系シミュレーションなど計算科学の質的向上につながる計算環境を提供することができる．また，メモリ量やアプリケーション実効性能を確保しつつ，性能当たりの電力，設置面積を極めて低く抑えている．
  さらに，ユニットA及びBのそれぞれのＣＰＵ及びネットワークは次のような革新性を有している．CPUは最新のプロセス技術で製造（45nm）．

  ユニットA　CPU：スカラ型にSIMD演算加速機構を搭載しつつ，多機能，高信頼性を実現
  　　NW：超高並列に対応できる新規性，拡張性の高いネットワーク
  ユニットB　CPU：従来のベクトル型の課題を解決する新規のアーキテクチャ
  　　NW：汎用性，信頼性が高く，かつ超省電力の光インタコネクション・ネットワーク

- **発展性**：シミュレーション技術の向上により，統合システムのニーズは高まっていくことが予想される．一方，両ＣＰＵ技術を発展させることにより，次々世代以降のCPU開発において，両者の技術を融合させることも視野に入れられる．

- **拡張性**：両ユニットとも10PFLOPSを大きく上回ることが可能なスケーラブルなシステム構成となっている．両ユニットそれぞれが拡張可能なため，単一アーキテクチャに比しても拡張性大．

- **展開性**：ユーザーのニーズに応じ，Ｆ及びＮＨ両者のシステムのそれぞれの展開及び複合システムとしての展開の両方が可能．

2. システム構成案の妥当性

⑴ システム構成案の詳細及び性能

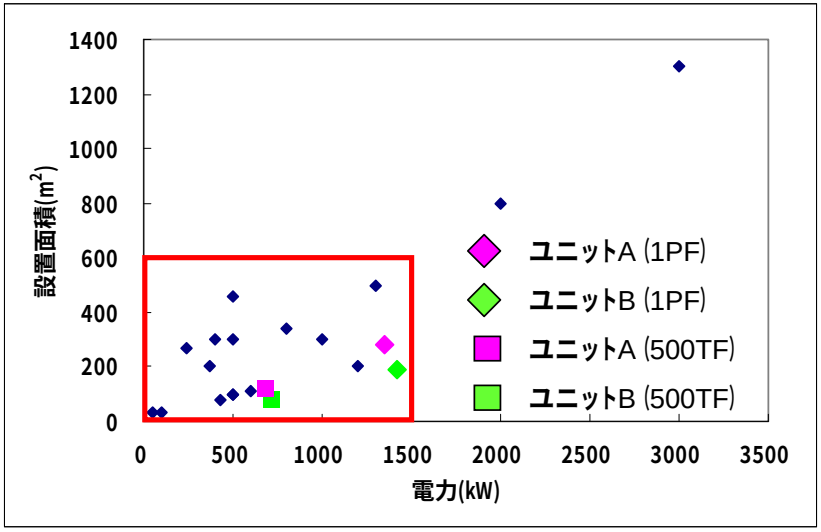
システム構成案は，それを基に大学や研究機関向けの計算機システムを構築することを可能とするものか．また，それを実施する場合に，消費電力，設置面積及び将来の拡張性の面で，適当なものとなるか．

⇨ システム構成案は，下方展開時の制約条件を満足しており，上記評価項目を満足している．また，筐体単位の消費電力や設置面積等から判断して，将来の拡張性も十分である．

　また，各計算センターのアプリケーションに対するニーズは多様であり，本システム構成が導入される可能性が高い．

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600m^{2}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

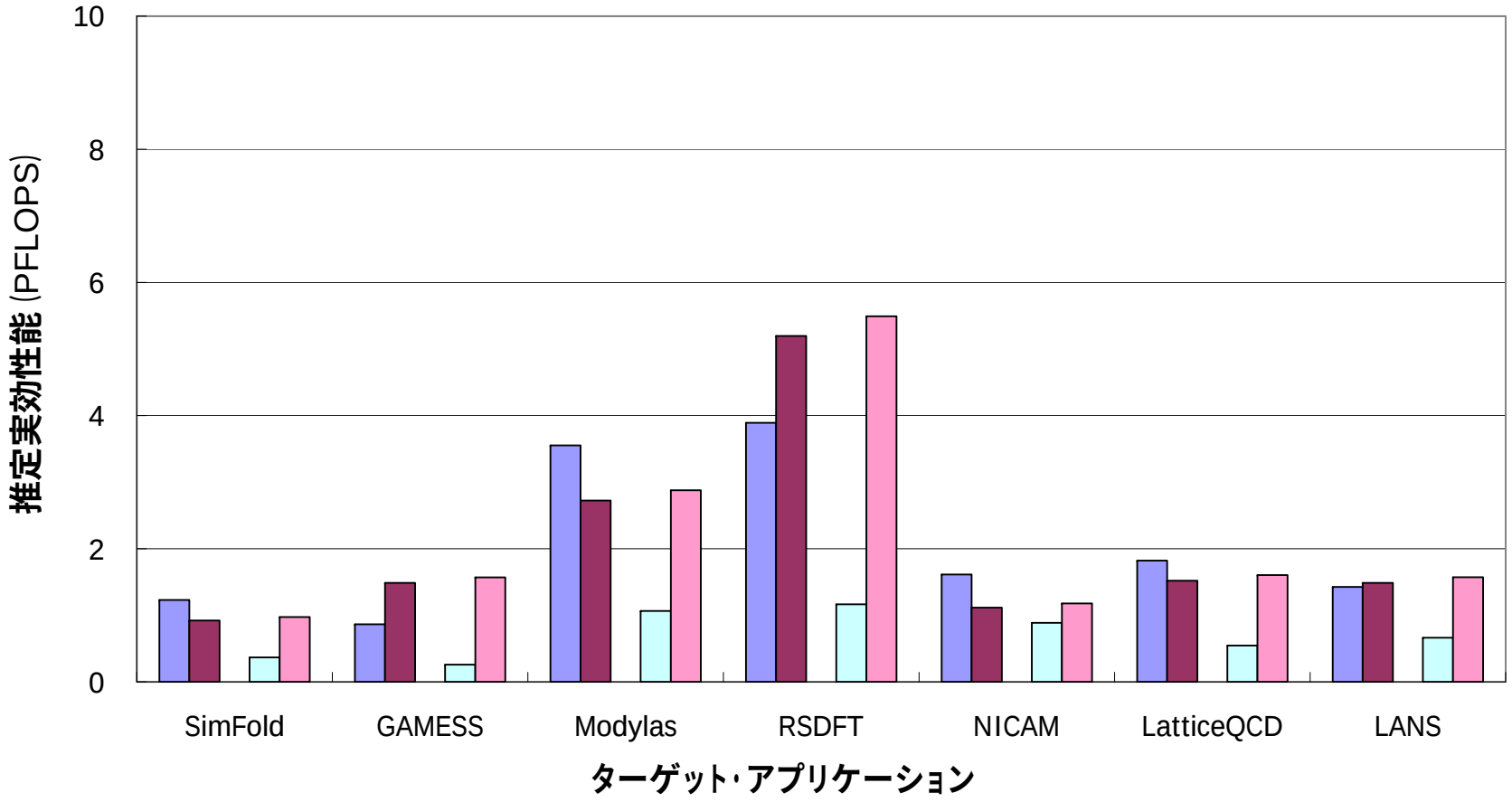
2. システム構成案の妥当性

⑵ 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．

⇨ 主要なターゲットアプリケーション7本に対して，ペタフロップス級の実効性能が見込まれ，十分であると考えている．

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

⑵ 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.

⇨ 評価したターゲットアプリケーションには，代表的なアルゴリズムが含まれており，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考えている.

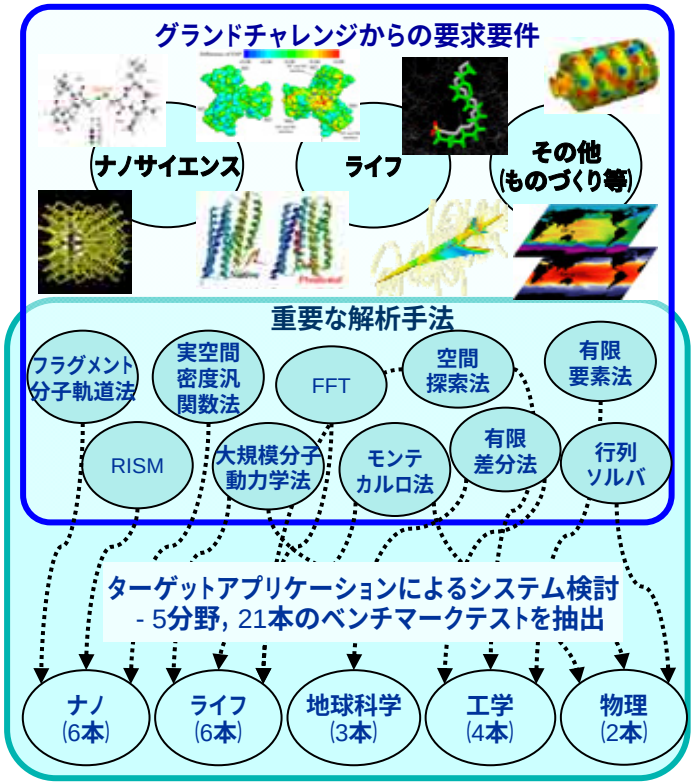
ターゲット・アプリケーションの選定

- 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．

ブランク・ページ

2. システム構成案の妥当性

(2) システムの機能

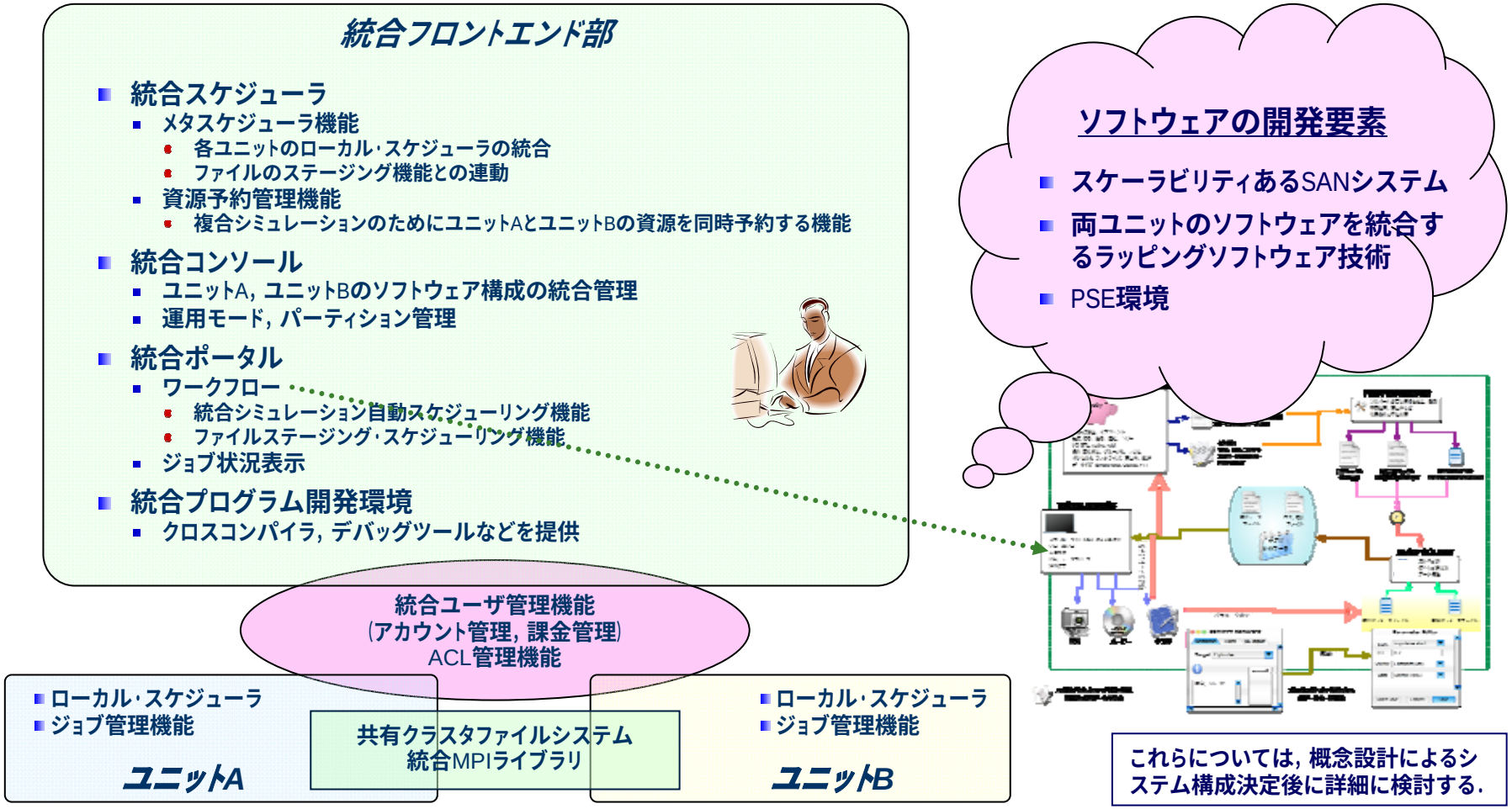
システムソフトウェア (OS, ライブラリ, コンパイラ等) はシステムの性能を十分引き出すものであるか.

⇨ システムソフトウェアは, 本システムの性能を十分引き出すものと考えている.

なお, 機能の詳細については, システム構成決定後の詳細設計において, 詳しく検討し確定することとしている.

統合システムとしての機能

- 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

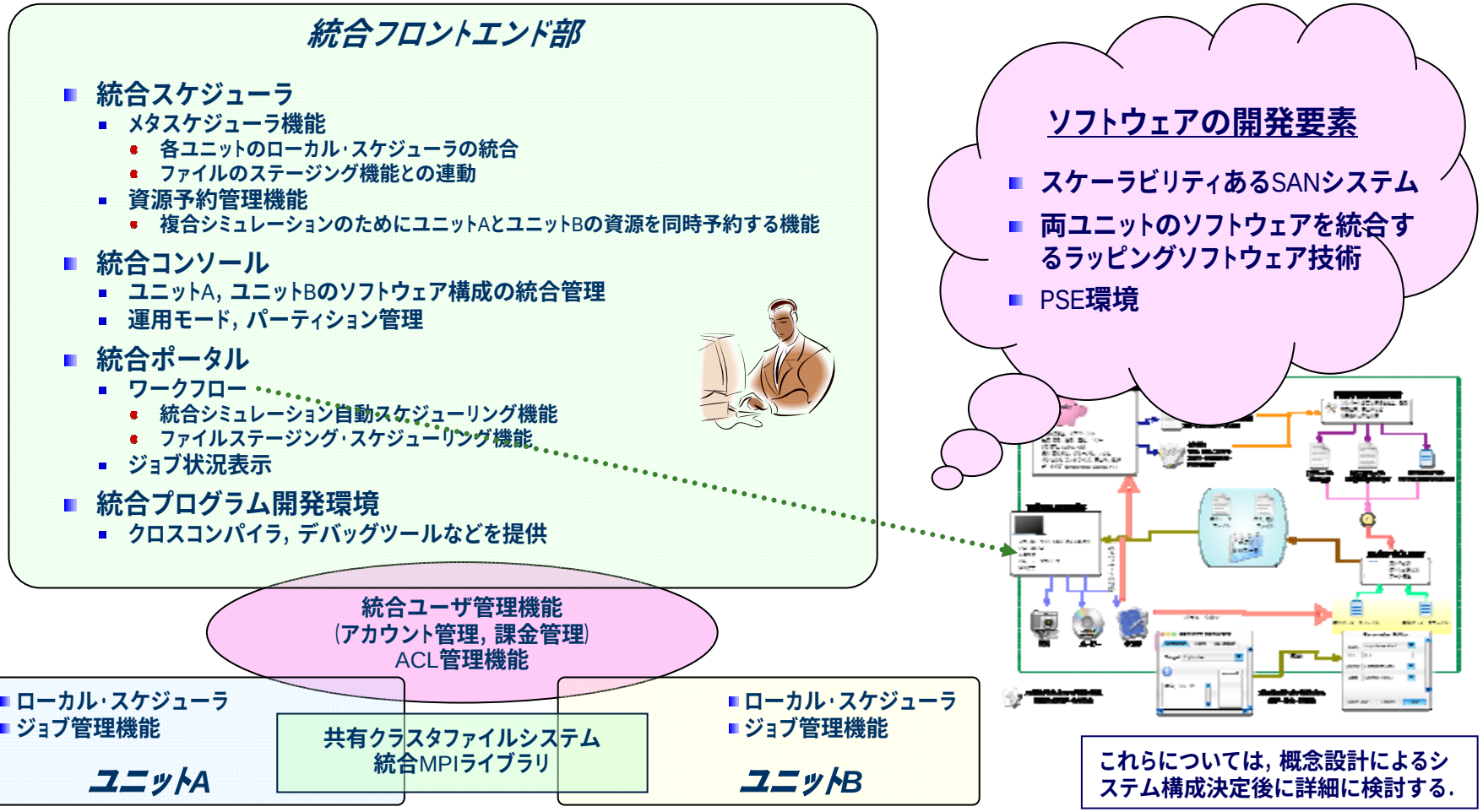
システムソフトウェア (OS, ライブラリ, コンパイラ等) は幅広い利用者が利用することが可能なものか.

⇨ システムソフトウェアは幅広い利用者の利用に対し, 十分な機能を持つと考えている.

なお, 機能の詳細については, システム構成決定後の詳細設計において, システム運用方針と照らし合わせて検討することとしている.

統合システムとしての機能

- システム・コネクトは，コモディティ技術（InfiniBand，または10GbE）で構成する．
  - システム全体の性能及びコストを最適化する観点で，両ユニットを結合する案として最適と判断．



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

⑶ システムの運用

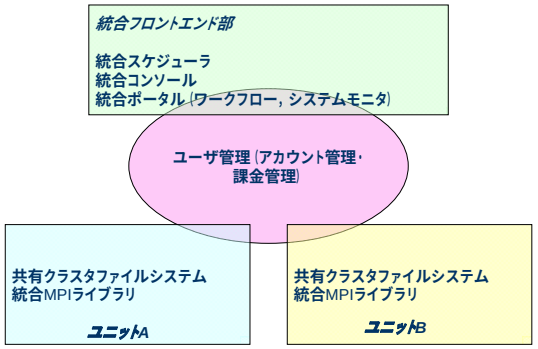
計算機資源の効率的な配分等により，多数の利用者がシステムを多様な用途に利用することが可能か．

⇨ 計算機資源を配分する機能などユーザが利用する視点での機能を備えており，利用可能性は十分あると考えている．システム構成決定後の詳細設計において，さらに詳細な機能について検討する．

なお，運用方針については今後検討されることになっている．

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次元トーラスを2シャシ単位で複数の小規模3次元トーラスに分割し，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

⑶ 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．

⇨ ハードウェアの基本的RAS機能は十分と考えている．運用方針に関連して，ソフトウェアを含むシステム全体の保守，運用については，詳細設計で検討することとしている．

ブランク・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ー貫性を確保
  - 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