

これまでの議論に対する補足説明

2009年4月22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

詳細設計の値等は現時点の設計目標値であり, 詳細設計終了までに変更される場合がある.

論点と補足説明項目

【論点1】 スカラ部とベクトル部の性能に大きな差はないが、ベクトル部の意義はあるか？

- 設計のコンセプト
- システム構成決定の考え方
- ベクトル部の技術的意義

【論点2】 スカラ部とベクトル部を統合したシステムを構築する意義はあるか？

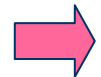
- 統合システムとしての意義

【論点3】 ベクトル部の設計・製造スケジュールの実現は困難ではないか？

論点1: スカラ部とベクトル部の性能に大きな差はないが、ベクトル部の意義はあるか？

■ 概念設計のコンセプト

- LINPACK 10PFLOPS達成可能なシステム
- 高い電力性能比, 高い面積性能比を持つシステム



【結果】F案(スカラ部), NH案(ベクトル部)ともほぼ同等の性能

- スカラ部: SIMD機構付きスカラプロセッサ
 - SIMD機構付加とレジスタ数増加によるHPC向け改良アーキテクチャ
- ベクトル部: キャッシュ付きベクトルプロセッサ
 - 従来のベクトル機(地球シミュレータなど)とは異なるアーキテクチャ

■ システム構成決定の考え方

- 両演算部でのLINPACKで10PFLOPSを達成できること.
- 開発予算の範囲内で, トータルな計算機資源を最大化すること.
- スパコン開発の技術力, 国際競争力向上に貢献すること.



スカラ部とベクトル部からなる複合システムが最適と判断.

- ベクトル部の詳細設計では, メモリ性能の強化を実施(0.5B/FLOP
1B/FLOP)

概念設計の評価結果(1/2)

概念設計仕様

評価結果

(4) メーカーからのシステム提案に対する評価(概要)

➤ 両提案の比較概要

- 概念設計の要求仕様(ピーク性能10PFLOPS以上, メモリ容量2.5PB以上, 消費電力30MW以下, 設置面積3,200㎡以下など)を満足.
- ベンチマーク・テスト(BMT)による性能推定結果, 電力性能比等はほぼ同等.

➤ CPUに対する評価

- F案は, 既存スカラプロセッサと親和性が高く, より幅広い技術展開が可能.
- NH案は, ベクトルプロセッサの課題を解決し, 高い演算性能を容易に達成.

➤ ネットワークに対する評価

- F案の新規性・将来性は評価できるが, 汎用性, 運用性, 実績などに優れたNH案が好ましい.

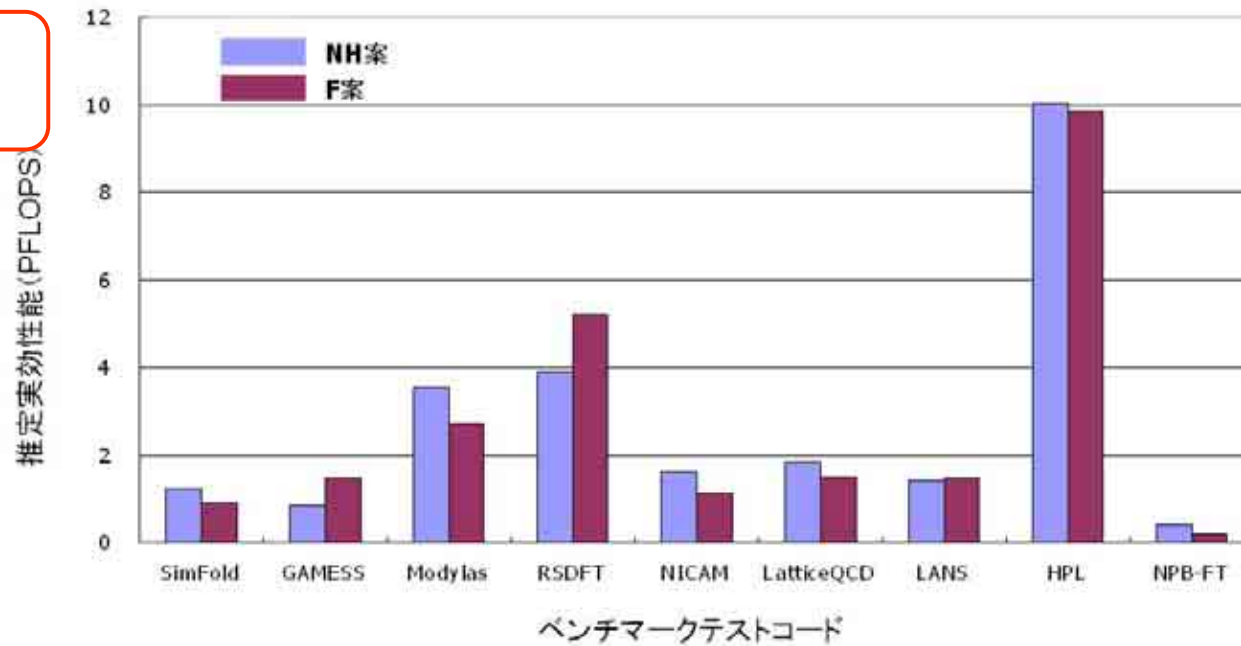
2007/5/28

次世代スーパーコンピュータ概念設計評価作業部会(第7回)資料

27

概念設計の評価結果(2/2)

ベンチマーク・テストによる性能予測(詳細9本)



NH案のメモリ性能
0.5B/FLOPで評価

- ターゲット・アプリケーションから7本のベンチマーク・テスト, 及びHPL, NPB-FTについて, 実効性能を推定.
- いずれのベンチマーク・テストもほぼ同等の性能.

詳細設計において, ベクトル部の
メモリ性能1B/FLOPに強化

システム構成案検討の考え方

(5)システム構成案検討の考え方(2006年12月末)

■ 選択肢

- 2者のいずれかを選択(2者択一)。

2者の提案は明らかな優劣をつけ難いものであるが、対外的に分かりやすい答えを出すためにも総合判断としていずれか一方を選択する

- 2者の案をベースに共同開発。ただし、以下を満たすことが条件。

- 共同開発のシステム構成の方が単独開発のシステムより、性能が上がること。
- 共同開発により、将来の我が国のスパコン開発の技術力、国際競争力、ビジネス展開力等の向上に一層貢献すること。
- 開発予算の範囲内で、共同開発システムが構築できること。

- 両演算部によりLINPACK実行
- システム全体のジョブスループット

- 汎用機による目標性能達成の見込みが確認できたため、アクセラレータの採用は考慮しない。

システム構成案の決定(1/2)

<ユニットA, Bの性能(1)>

10PF

★ 単一アーキテクチャで10PFLOPS級のアプリケーションを実行し、世界最高のアプリケーション実行性能を達成するとともに、グランドチャレンジの成果を挙げる。
(アプリ例:①ナノデバイスの高精度シミュレーション(例1)、②巨大生体分子の全電子計算など)

多くの利用者がいるPCクラスタや並列サーバによる開発・実行環境からの連続性、プログラムの移植性を重視し、スカラプロセッサでこれを達成する

➤FのCPUで理論性能:10PFLOPS超

3PF

★ 地球シミュレータなどベクトルプロセッサ向けのプログラム資産の有効利用等において、ペタフロップス級のアプリケーション実行性能を確保するとともに、世界最高性能のベクトル計算環境を提供する。

(アプリ例:①雲解像モデルによる気候変動予測(例2)、②乱流遷移現象を取り入れた航空・宇宙機の設計など)

➤NHのCPUで理論性能:3PFLOPS超

★ 上記の値をそれぞれのユニットが最低限実現すべき理論性能として、プロジェクト予算総額の範囲内のシステム構成を検討。

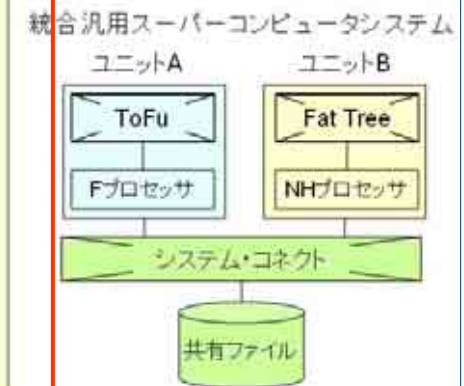
システム構成案の決定(2/2)

(9) 結論

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、経費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。



複合システムの決定

システム構成の意義(技術の維持・発展)

本システム構成案を採用する理由 I. (共同開発の条件b.)

- 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).

- ① 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.
- ② 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.
- ③ 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, 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.

ベクトル部の技術的意義については後述

システム構成の意義(アプリケーション利用と運用)

本システム構成案を採用する理由Ⅱ。(共同開発の条件a.)

- 効率的なシミュレーションの実行及びシステムの運用が可能。

両者のCPUを用いた統合汎用システムにより、ソフトウェア資産のより有効な利用や共用施設として効率的なユーザー対応ができるだけでなく、多くのアプリケーションで有効な複合シミュレーションのための最適なシステム環境を構築できる。

- ① スカラプロセッサ向けに開発された多くのアプリケーションと地球シミュレータに代表されるベクトルプロセッサ向けに最適化されたアプリケーションの両方を容易かつ発展的に活用でき、ソフトウェア資産のより多様な有効活用が可能となる。
- ② 多様なアプリケーションをそのアプリケーションに適したプロセッサで実行することにより、単独プロセッサによるシステムよりも計算資源を有効活用できる。
- ③ 計算科学の多くのアプリケーションで見られる複合シミュレーションにおいて、最適な統合システム環境(スカラ+ベクトル)の構築が可能となる。

計算科学技術推進ワーキンググループ
第2次中間報告,平成17年8月24日

システム構成の意義（投資効果）

本システム構成案を採用する理由Ⅲ.（共同開発の条件b.）

- 本プロジェクトの波及効果を最大化できる。
 - ① 大学等の計算センターの多様なニーズを踏まえ、必要とされるスパコンの規模やアプリケーションに対応したより柔軟な下方展開が期待される。また、メーカーの競争関係も維持される。
 - ② CPU技術及びネットワーク技術の様々な要素技術が、家電や電子機器、ビジネスサーバー等へ幅広く展開されることが期待される。
 - ③ 人材を含めた両者のリソースをより多く投入できる共同開発により、費用対効果の高いシステム開発が可能。

開発予算の範囲内でシステム全体の性能を最大化

概念設計評価作業部会(平成19年6月)の評価結果

理化学研究所のシステム開発方針及び最適化の考え方は、文部科学省のプロジェクトの目的・目標に照らして妥当である。また、そのシステム開発方針及び最適化の考え方に基づき、理化学研究所が作成したシステム構成案は、我が国の最先端・高性能汎用スーパーコンピュータのシステムを構築する上で、適切なものであり引き続き研究開発を進めるべきであると評価する。

本プロジェクトの大きな目標の一つに、「世界最先端・最高性能」を獲得することが挙げられている。現時点で計画されている他のプロジェクトや技術開発動向等を勘案して設定されたLinpack10ペタFLOPSの達成を目指した研究開発は妥当であると考えながら、世界的な開発競争の中で不確定要素を含むものであることから、これに柔軟に対応できる取組みが必要であると考える。

これまで述べてきたとおり、本システム構成案は複合システムであるという大きな特徴を有し、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、我が国が強みをも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ものである。

この複合システムは、

- 多様なアプリケーションの効率的な実行
- 高い拡張性、下方展開性
- 技術オプションの確保による技術力の強化、国際競争力の向上

等の観点から非常に有効であるが、同時に本システムを一体的に運用するためのトータルシステムソフトウェアの開発について、より一層の検討、取組みが必要である。

「次世代スーパーコンピュータ概念設計評価報告書」科学技術・学術審議会 研究計画・評価分科会
情報科学技術委員会 次世代スーパーコンピュータ概念設計評価作業部会(平成19年6月)より抜粋



ベクトル部の技術的意義【技術的問題と解決方法】

- 3つの壁(Wall)とそれらへの対応
 - ILP(Instruction level Parallelism) Wall
マルチコア化, ベクトル化により, 単位時間あたりに処理できる命令数を増大させる(チップあたりの性能向上)
 - Memory Wall
キャッシュ容量増, キャッシュ有効利用(ソフトウェア制御)により, 主記憶へのアクセスを削減し, 全体の処理時間を短縮させる.
 - Power Wall
マルチコア化, ベクトル化により電力あたりの性能を向上させる.

ベクトル部の技術的意義【ベクトル部とx86の将来】

- コア数の増大(両者共通)
- コア性能の向上
 - レジスタ増, 演算器増
 - ベクトル部は, 演算器とレジスタは独立に増加することが可能
命令セット不変 コンパイラ不変
 - x86は演算器とレジスタは一体的に増加
命令セット変更・追加が必要
 - ベクトル部は将来の性能向上に対し, アーキテクチャ上の柔軟性あり.
 - x86系では, レジスタ大容量化等に伴うISAの変更・追加が必須.
- 省電力化: 電力あたりの演算性能向上(演算密度向上)
 - 長ベクトル化により, 命令発行数削減可能. これにより, 命令制御の簡素化と命令制御動作削減により消費電力の削減が可能.

ベクトル部の技術的意義【結論】

- x86系も長ベクトル化の方向にある。
 - Intel Sandy Bridge (32nm) で採用されるAVXは、レジスタ長がSSE4の128bitから256bitに拡張。
 - Intel Haswell (22nm) は、ベクタプロセッサを搭載の計画あり。
 - ただし、x86系のSIMD強化は、ISA上、レジスタ数増加は限界 (AVXでは最大1024bitまで)。
 - 以下の参考文献では、x86/Power系CPUにベクトル部CPUとほぼ同様の機能搭載の提案あり (VL制御によるベクトルレジスタ、リストベクトル/ストライド・アクセスなど)。

【参考文献】 J. Gebis and D. Patterson, "Embracing and Extending 20th-Century Instruction Set Architectures," Computer, Vol.40, No.4, pp.68-75 (2007).

- **ベクトル部CPUは、将来方向を先取り**したものであり、**国家基幹技術**として、**維持、発展**させるべき技術の一つである。

論点2: スカラ部とベクトル部を統合したシステムを構築する意義はあるか?

■ 概念設計時の統合システムとしての意義

- 性能目標を達成するために、**両演算部による統合LINPACK**により、LINPACK性能 10PFLOPSを達成すること。
- ユーザに対し、On-the-fly複合シミュレーションなどの**連携アプリケーションの実行環境**を提供すること。



ユーザの利便性を確保するために、ユーザが統合システムを一体的に利用できる統合システム機能(トータルシステムソフトウェア)を開発することとした。

- 統合システム(トータルシステムソフトウェア)の機能概要については、設計・製造計画評価作業部会及び中間評価作業部会において説明済み。

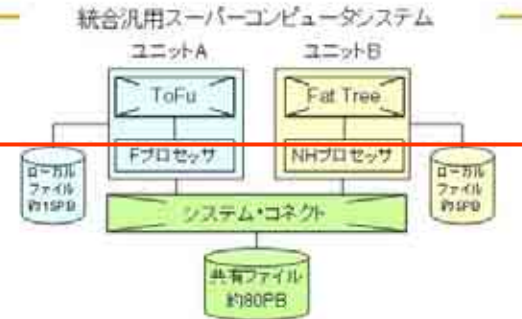
統合LINPACKについて【概念設計評価時】

システム構成案

- 性能目標
 - ユニットA+ユニットBでLINPACK 10PFLOPS超を達成する。
 - ユニットAの理論性能: 10PFLOPS超
 - ユニットBの理論性能: 3PFLOPS超

【試算】

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

- 統合システム
 - 統合システム機能:
 - 統合スケジューラ, 統合ポータルなど
 - 統合共有クラスタファイルシステムなど
 - システム・コネクタ: 総バンド幅 約1.2TB/秒
- メモリ容量, 及びファイル容量の目標仕様
 - ユニットAのメモリ容量は, 理論性能当たり1/8 (B/FLOPS)
 - ユニットBのメモリ容量は, 理論性能当たり1/4 - 1/8 (B/FLOPS)
 - ファイルシステムの容量は, 全体で約100PBとする。
 - ユニットA, 及びBのローカルファイルシステムの容量は, メモリ容量の10倍とする。
 - 残りを共有ファイルシステムの容量とする。
- ユニットAとBの機能及び性能上の特徴を活かし, 詳細設計においてさらに最適化を行う。(例: ユニットBのメモリバンド幅の強化)
- 詳細設計終了後, 製造に入る前に最終構成を決定する。

統合LINPACKに対する考え方

- 詳細設計において、LINPACK性能、整備スケジュールを詳細に検討。
 - 平成24年6月までにスカラ部でLINPACK 10PFLOPS達成可能である。
 - 詳細設計が進捗した結果、統合システムでのLINPACK計測は、両演算部の完成後、平成24年4月以降の計測と成らざるを得ない。計測のための期間は約2ヶ月¹⁾。

1) 運用時のシステム構成を全体で動作する動作モードにし、計測後に再度運用時のシステム構成に戻すまでの期間

- 統合システムによるLINPACK計測を行うよりも、**早い段階でアプリケーションの成果を得ることが重要**。
 - 統合LINPACKを計測しなければ、早い時期にユーザへの共用が開始できると考えている。

 スカラ部でLINPACK 10PFLOPSを達成する。

連携アプリケーションの実行環境について

- プロジェクト開始当初は、文科省の計算科学技術推進ワーキンググループの報告のとおり、**複合型計算機での連携計算が有効であるアプリケーション(14本)**が例示されていた。
(「計算科学技術推進ワーキンググループ 第2次中間報告」,平成17年8月24日)
- 概念設計評価時には、これらのアプリケーションを対象として複合システムを有効に利用できると考えていた。
- 詳細設計時には、**統合システムとしての機能設計に重点を置きつつ**、具体的な連成アプリケーションの選定を試みたが、適切なアプリケーションが現れなかった。
- 設計・製造計画評価作業部会において、「統合システムを実現するシステムコネクタに関しては、**実際のアプリケーションにおける統合MPI機能やファイルシステムの実効性能を評価して最適な構成を決めるべきである**」との指摘を受け、**連携アプリケーション(RISM-OpenFMO, MSSG-放射モデル)**を選定して評価を開始したところである。
- ただし、各演算部に適したアプリケーションがあることから、ファイル渡しによる連携アプリケーションの需要はあり、**統合システムの機能は必要**である。また、戦略委員会(第6回)では原子力/防災の分野において連成アプリケーションの要望が示されている。

利用・運用を一体化するシステム機能【概念設計時】

統合システム機能

システム・コネクは、コモディティ技術(InfiniBand, または10GbE)で構成し、統合的な利用・運用を可能とするためのソフトウェアを開発。

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機能との連動
 - 資源予約管理機能
 - 複合シミュレーションのためにユニットAとユニットBの資源を同時予約する機能
- 統合コンソール
 - ユニットA, ユニットBのソフトウェア構成の統合管理
 - 運用モード, パーティション管理
- 統合ポータル
 - ワークフロー
 - 統合シミュレーション自動スケジューリング機能
 - ファイルステージング・スケジューリング機能
 - ジョブ状況表示
- 統合プログラム開発環境
 - クロスコンパイラ, デバッガツールなどを提供



統合ユーザ管理機能
(アカウント管理, 課金管理)
ACL管理機能

- ローカル・スケジューラ
- ジョブ管理機能

ユニットA

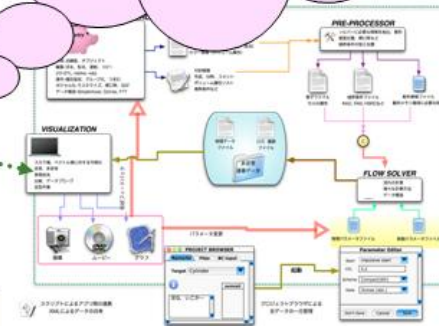
共有クラスタファイルシステム
統合MPIライブラリ

- ローカル・スケジューラ
- ジョブ管理機能

ユニットB

ソフトウェアの開発要素

- スケーラビリティあるSANシステム
- 両ユニットのソフトウェアを統合するラッピングソフトウェア技術
- PSE環境

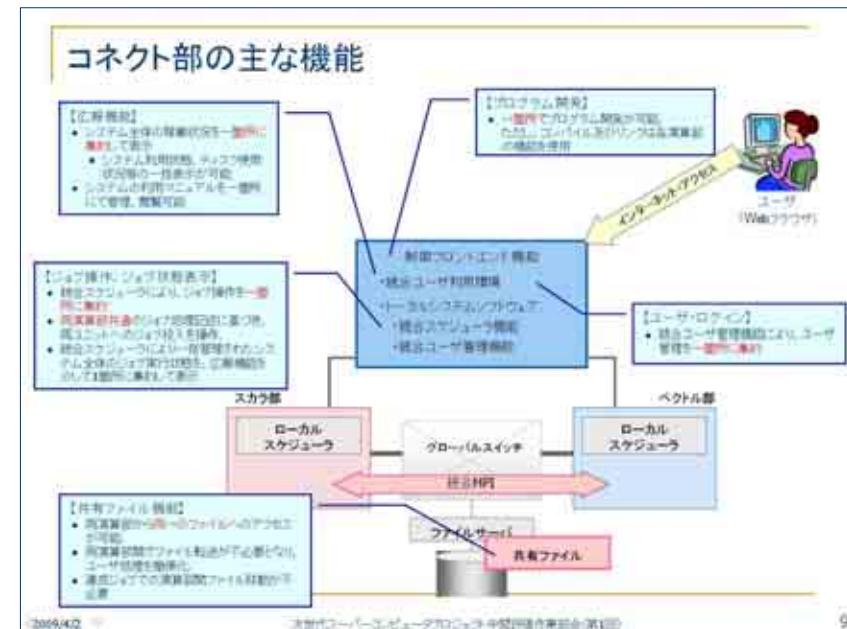


これらの機能は、概念設計によるシステム構成決定後に詳細に検討する。

【設計結果】一体的なユーザ利用環境を提供

■ 概念設計評価後, ユーザが統合システムを一体的に利用できる環境を設計

- 複合システムをユーザに隠蔽する機能を充実
 - ワンストップサービス
 - 高いユーザビリティ
- 統合ユーザ利用環境
 - 1箇所でシステムの利用が可能
- 統合スケジューラ
 - ジョブスクリプトを統一
 - ジョブを一元的に管理
 - 連携・連成ジョブの制御
- 共有ファイルシステム
 - ファイルを一元的に管理
 - 両演算部間でファイル転送が不必要



- 機能に関しては, 設計・製造計画評価作業部会において「一体的な利用環境を提供すること, 及び複合システムとしての性能を引き出す機能を備えることについては概ね妥当である。」との評価を得ている.

論点3: ベクトル部の設計・製造スケジュールの実現は困難ではないか？

- ベクトル部については、概念設計時のスケジュールと比較して、製造開始(Tape Out)が2ヶ月遅延。
- 量産時期は予定通りであるが、65nm半導体プロセスによる設計、製造の経験(SXシリーズ)を踏まえ、製造スケジュールは実現可能であると判断している。
- LSI製造については、45nm半導体プロセスによるメモリ混載の実績があり、実現性は高いと考える。
- ベクトル部のLSI量産期間とLSI生産能力を考慮すると、搬入・据付調整期間は6ヶ月あり、平成22年度末1PFLOPSのシステム稼動は可能と考える。

