



次世代スーパーコンピュータ開発の 現状について

平成22年3月24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

横川 三津夫

革新的ハイパフォーマンス・コンピューティング・インフラ (高機能演算研究基盤) の構築 (平成22年度から)

【概要】

次世代スーパーコンピュータプロジェクトを進化・発展させ、開発側視点から利用者側視点に転換し、多様なユーザーニーズに応える革新的な計算環境を実現する。

- ナンバーワンの世界最先端・最高性能を目指した次世代スーパーコンピュータの開発・整備
- 次世代スパコンと国内のスパコンをネットワークで結び協調的に利用するナンバーワンの「革新的ハイパフォーマンス・コンピューティング・インフラ (HPCI)」の構築

1. 次世代スーパーコンピュータの開発・整備

我が国のハイパフォーマンスコンピューティングの中核となる次世代スパコンを平成24年の完成を目指し開発・整備する。
(平成22年度末一部稼動, 平成24年6月までに10ペタFLOPS級を達成)

2. 革新的ハイパフォーマンス・コンピューティング (HPC) に必要な研究開発

文部科学省資料より

開発スケジュール (平成18年度－平成24年度)

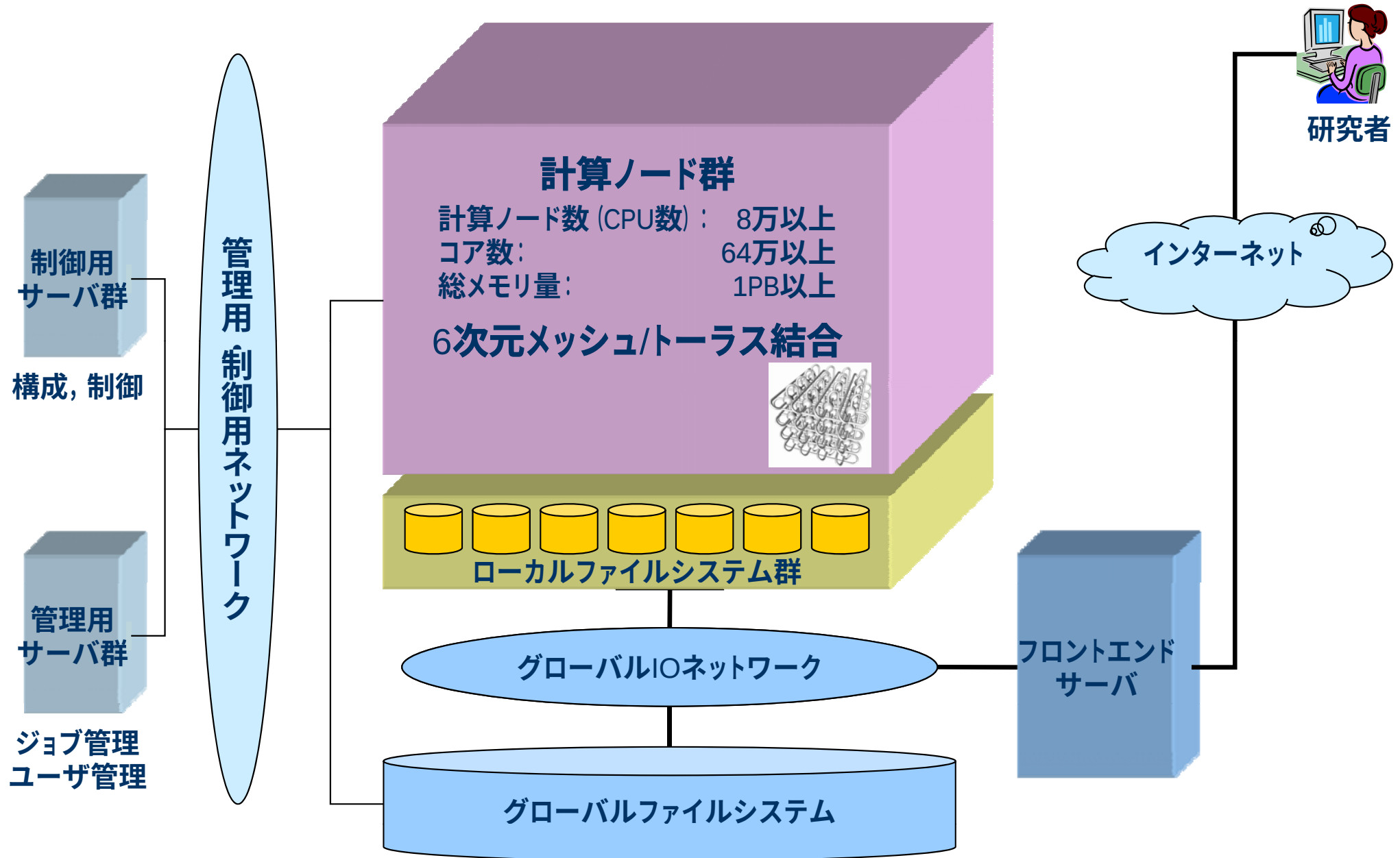
現在

		平成18年度 (2006)	平成19年度 (2007)	平成20年度 (2008)	平成21年度 (2009)	平成22年度 (2010)	平成23年度 (2011)	平成24年度 (2012)	
システム		概念設計		詳細設計		試作・評価・製造			性能 チューニ ング
ソフトウェア プラットフォーム	次世代ナノ統合 シミュレーション	開発・製作・評価					実証	↑ 供用開始	
	次世代生命体統合 シミュレーション	開発・製作・評価					実証		
施設	計算機棟	設計		建設					
	研究棟	設計		建設					



次世代スーパーコンピュータの概要

システム構成概要

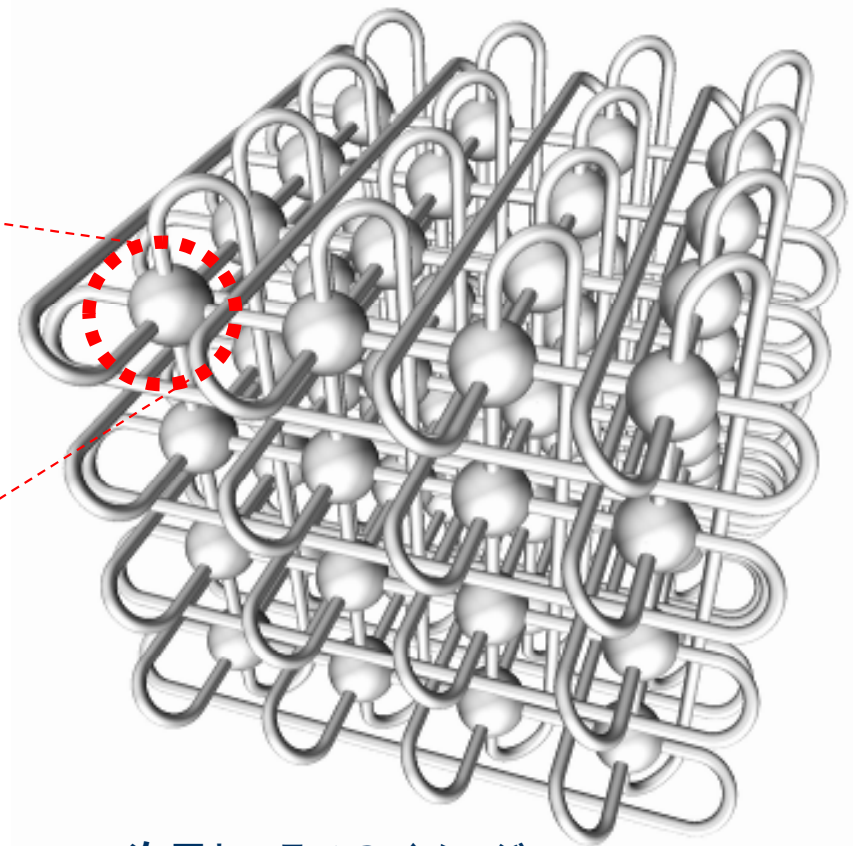
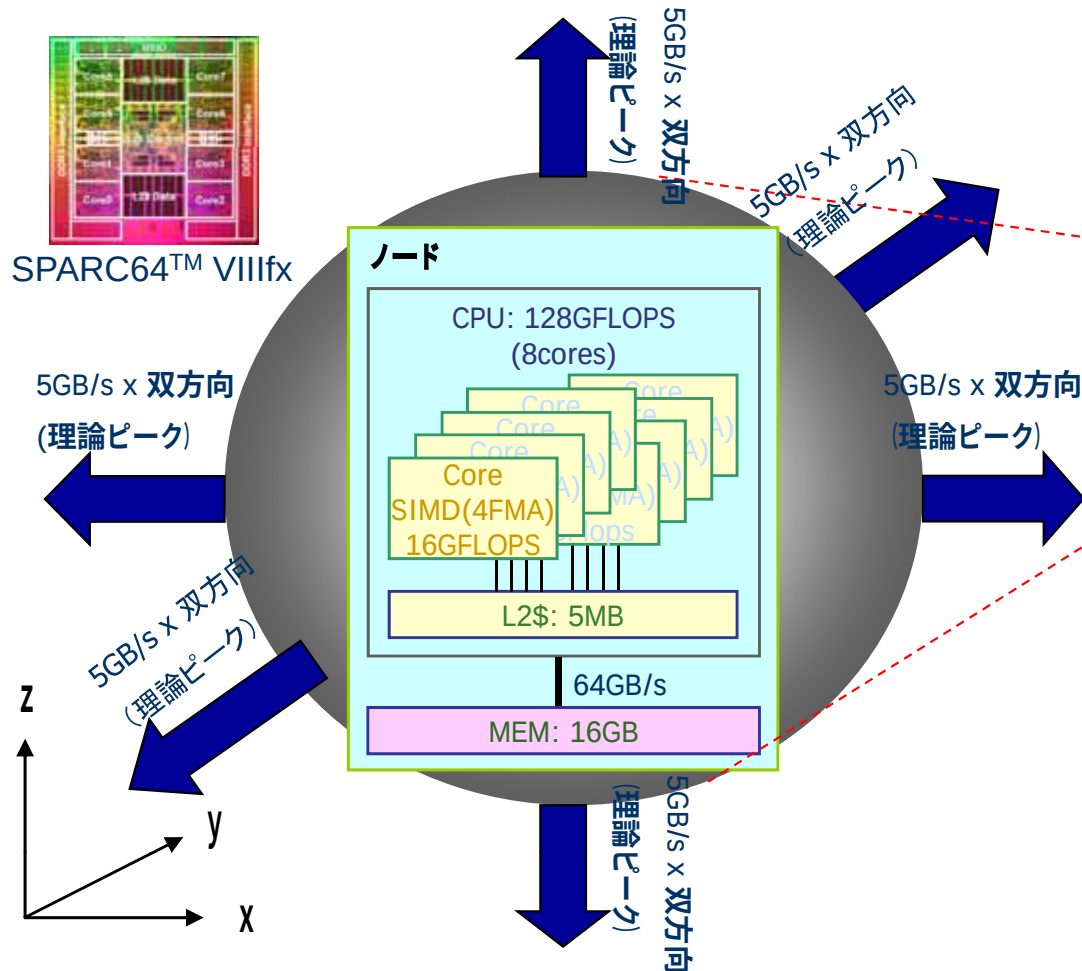


計算ノード群の構成



- 計算ノード数 (CPU数) : 8万以上
 - コア数: 64万以上
- ピーク演算性能: 10PFLOPS以上
- メモリ総容量: 1PB以上 (ノード当り16GB)

- ネットワーク: ユーザービューは3次元トラス
 - 帯域: 3次元の正負各方向にそれぞれ 5GB/s x 2 (双方向) [理論ピーク]
 - ケーブル: 約200,000本, 約1200km

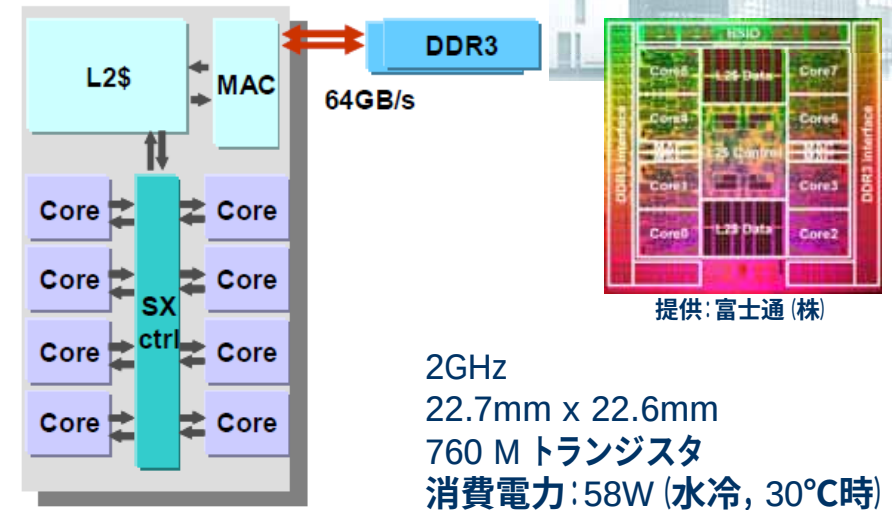


3次元トラスのイメージ

提供: 富士通 (株)

プロセッサ構成

- 8コア構成, 各コア256本の浮動小数点レジスタを備えたスーパースカラ方式
 - SIMD拡張 (積和演算器2個 x 2セット)
 - コア当り16GFLOPS, CPU当り128GFLOPS
- コア共有の2次キャッシュ (5MB, 10way)
 - ハードウェアバリア機構
 - プリフェッチ機構
 - セクタキャッシュ機能 (次ページ)
- データ供給能力
 - レジスタ・L1キャッシュ間: 4B/FLOP
 - L1キャッシュ・L2キャッシュ間: 2B/FLOP
 - L2キャッシュ・主記憶間: 0.5B/FLOP



	仕 様
CPU性能	128GFLOPS (16GFLOPSx8コア)
コア数	8個
浮動小数点演算器構成 (コア当り)	積和演算器: 2×2個 (SIMD) (逆数近似命令: SIMD動作) 除算器: 2個 比較器: 2個 ビジュアル演算器: 1個
	浮動小数点レジスタ(64ビット): 256本 グローバルレジスタ(64ビット): 188本
キャッシュ構成	1次命令キャッシュ: 32KB(2way) 1次データキャッシュ: 32KB(2way) 2次キャッシュ: 5MB(10way)コア間共有
メモリバンド幅	64GB/s(0.5B/F)

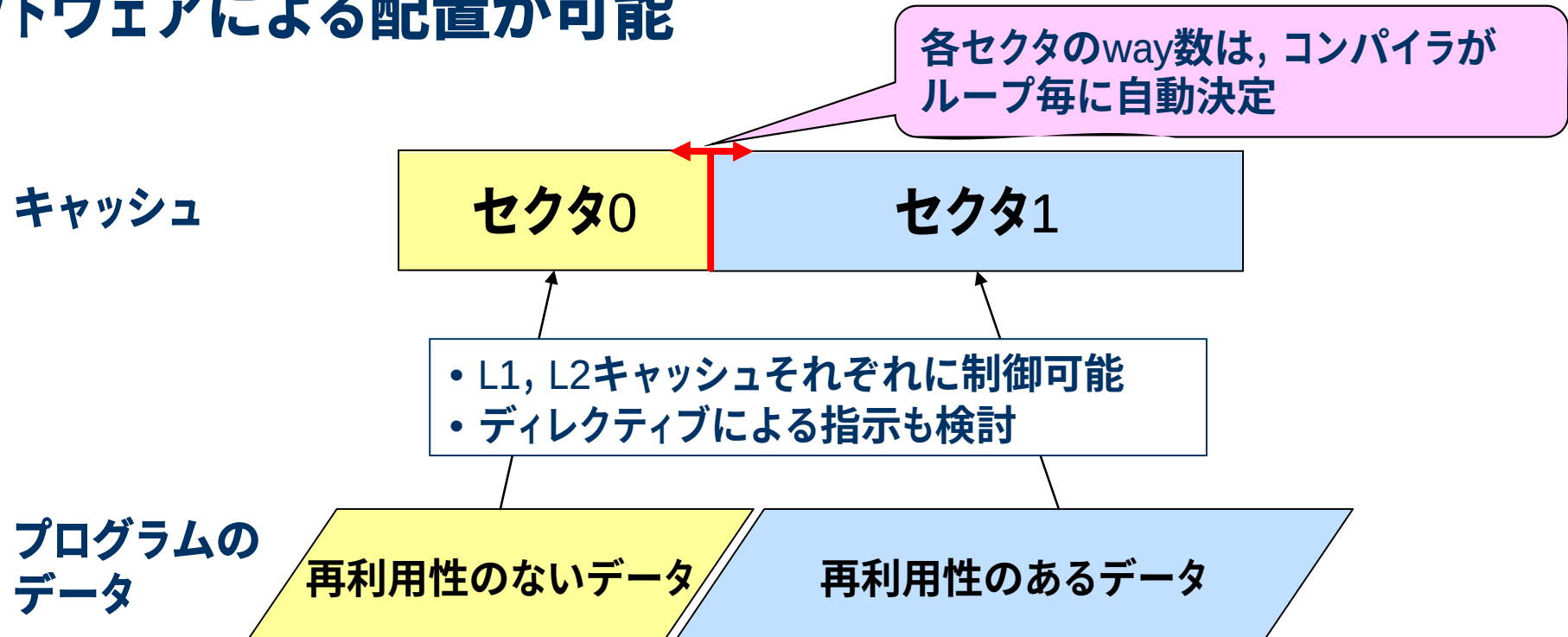
より詳細な情報は, 「SPARC64™ VIII fx Extensions」を参照のこと

<http://img.jp.fujitsu.com/downloads/jp/jhpc/sparc64viii-fx-extensions.pdf>

セクタキャッシュとは？



- 再利用性のあるデータを選択的にキャッシュに残す仕組み
- ソフトウェアによる配置が可能



- 再利用性のあるデータを含むプログラム例

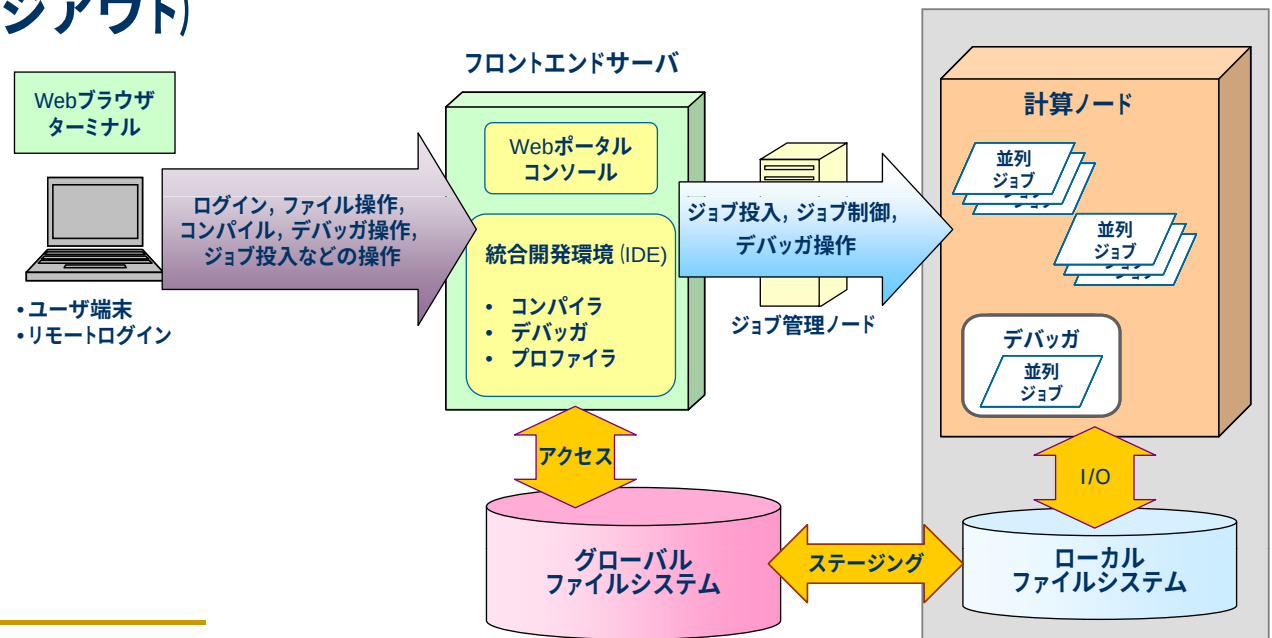
```
do j=1,n
  do i=1,n
    a(i) = a(i) + b(i,j)
  enddo
enddo
```


システム利用環境

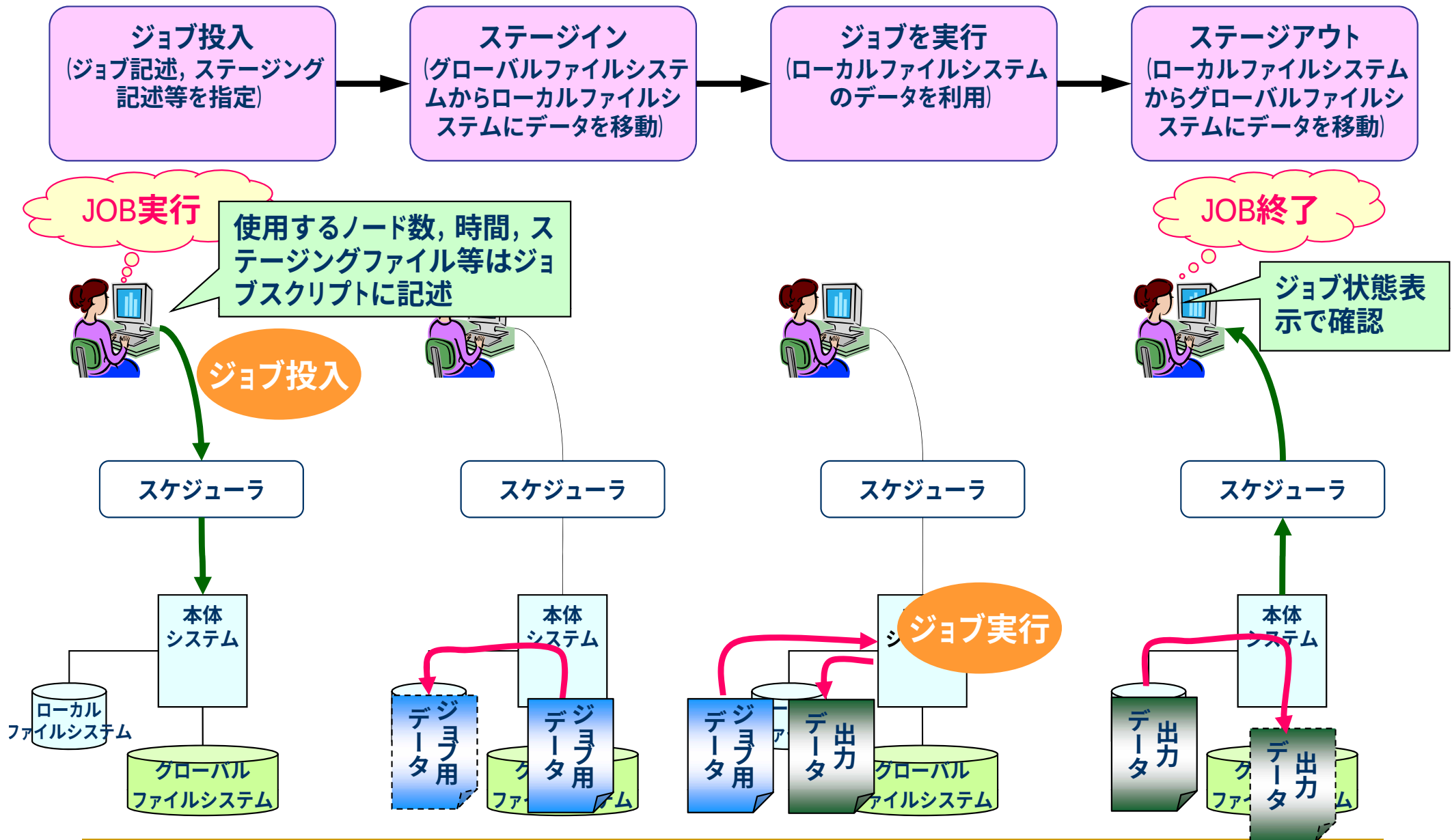


- OS: Linuxをベースとしたオペレーティングシステム
 - POSIX規格に準ずるコマンド群を提供
- 大規模分散ファイルシステム (2階層のファイルシステム)
 - ファイルステージング機能
 - ジョブ実行前にグローバルファイルシステムからローカルファイルシステムへファイルを転送 (ステージイン)
 - ジョブの出力ファイルをローカルファイルシステムからグローバルファイルシステムへ転送 (ステージアウト)

- バッチジョブを主体としたジョブ実行環境
 - デバッグ用に会話型環境を用意 (予定)



バッチジョブ実行時の処理の流れ



プログラム言語, コンパイラ



- Fortran, C, C++, XPFortran
- GNU C/C++ **拡張仕様**
- **4倍長精度演算をサポート: IEEE754R及びdouble-double形式**

- SPARC64™ VIIIfxの機能を有効活用するコンパイラ機能
 - SIMD機構の活用
 - 自動ベクトル化を応用したSIMD命令の自動生成
 - IF文を含むループのSIMD化 (マスク付きSIMD化)
 - 大容量レジスタ (倍精度浮動小数点 256本) の有効活用
- セクタキャッシュの利用
 - セクタキャッシュを考慮したプリフェッチ命令の自動生成
 - セクタキャッシュをユーザが意識して利用するためのディレクティブ
- 自動並列化
 - マルチスレッド化, パイプライン並列化機能

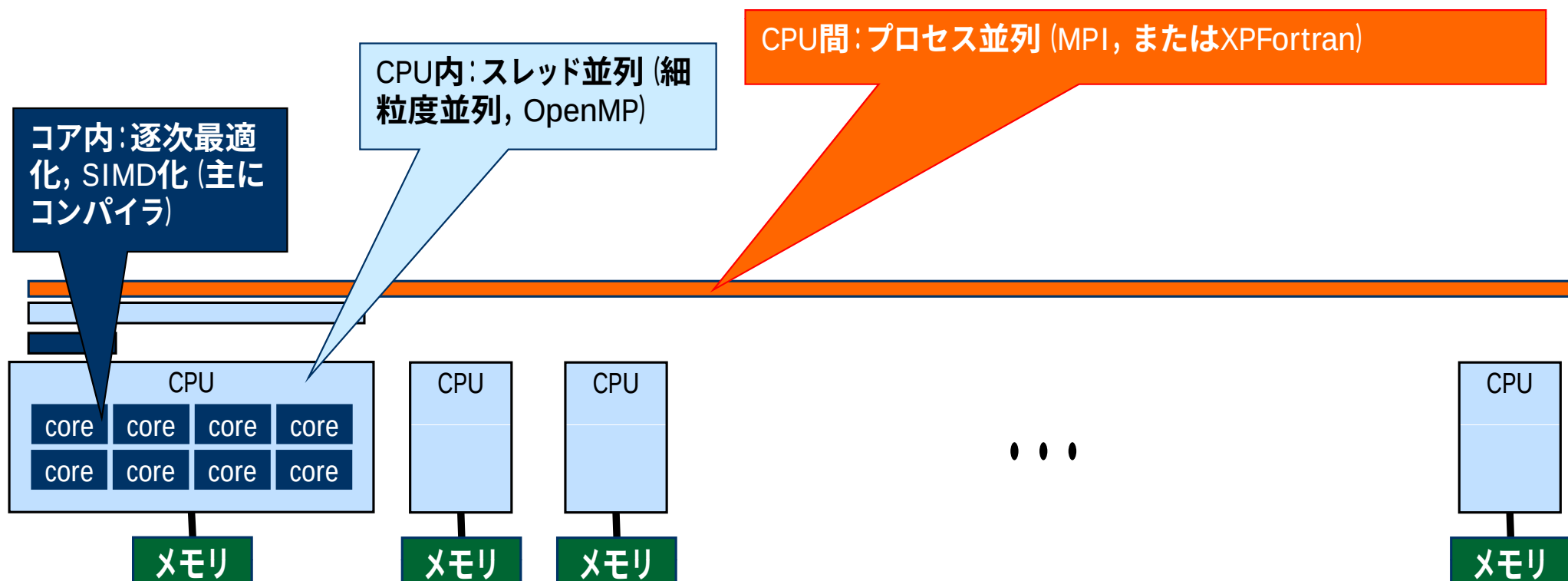
ライブラリ及びプログラム開発支援環境

- MPIライブラリ (MPI-2.1に対応)
 - 低レイテンシ・高スループットの実現
 - トポロジ構成を意識した集団通信関数を提供
 - Bcast /Allgather /Alltoall /Allreduce
 - インターコネクトのハードウェアバリア機構を用いたハードバリア/リダクション演算への活用
- 数値計算/科学技術計算ライブラリ
 - システムにチューニングされたBLAS, LAPACK, SSL II (富士通製科学技術計算用ライブラリ), FFTWを提供
- 開発支援ソフトウェア
 - デバッガ: DWARF2対応
 - 性能解析ツール: デバッグツール, プロファイラ, MPIトレーサ等の連携

プログラミングモデル

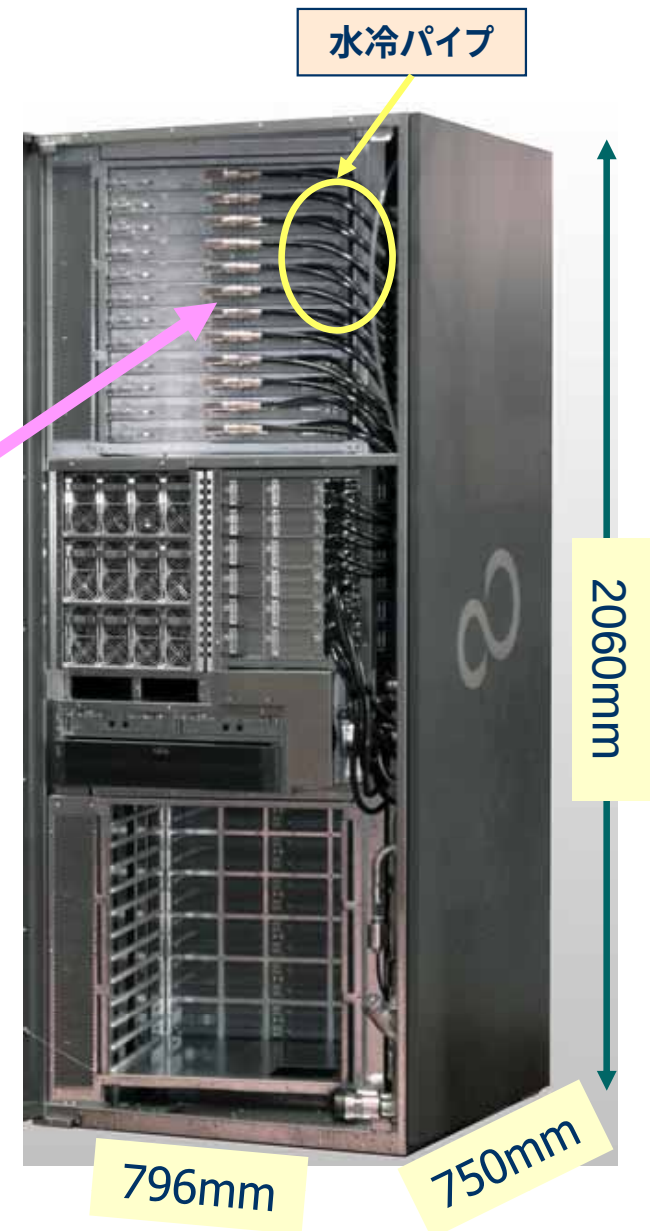
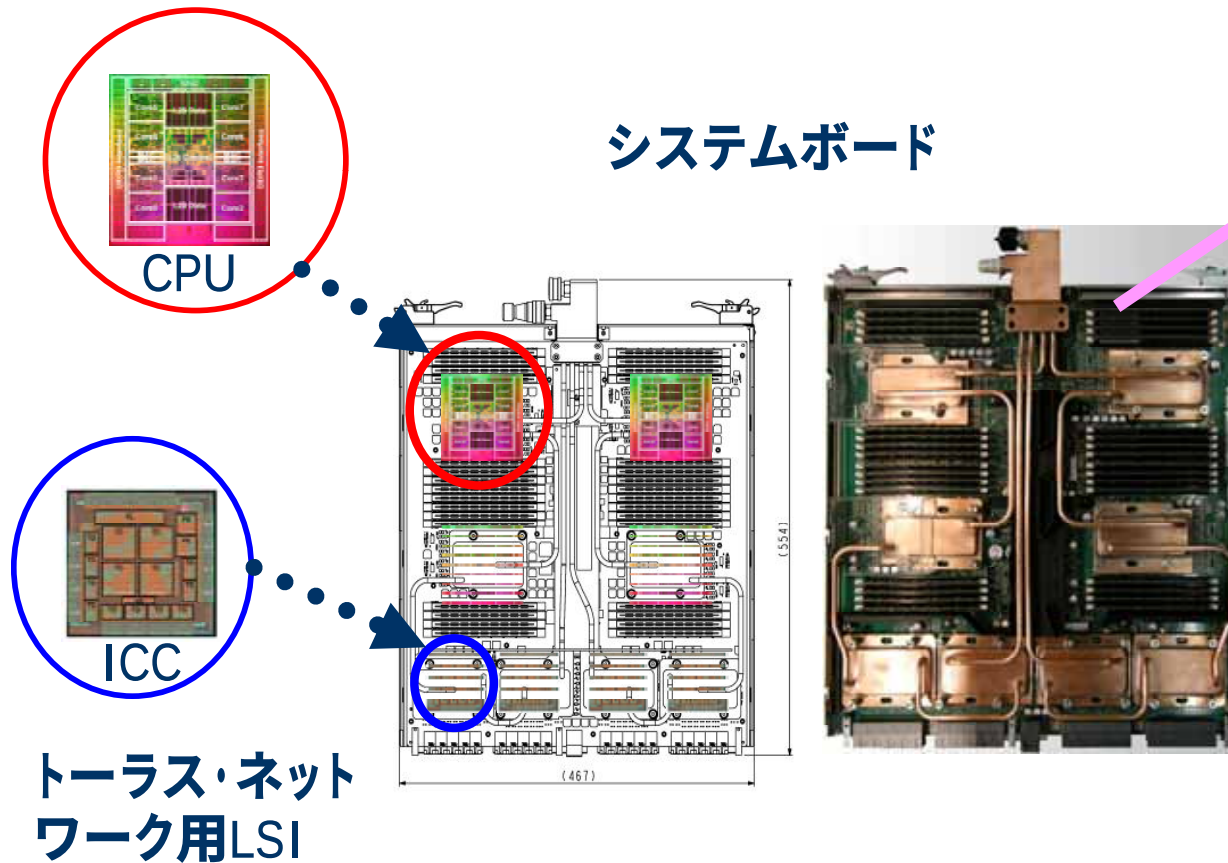


- スレッド並列＋プロセス並列のハイブリッド型を推奨
 - コア内：コンパイラによる逐次最適化, SIMD化
 - CPU内：スレッド並列 (自動並列化, OpenMP)
 - CPU間：プロセス並列 (MPI, XPFortran)
- フラット型も可能



システム開発の状況

- 試作機が完成. ハードウェア及びソフトウェアの試験を実施中.





次世代スーパーコンピュータ施設について

次世代スーパーコンピュータ施設



450km (280miles)
west from Tokyo

兵庫県神戸市中央区港島南町7丁目 (ポートアイランド第2期内)
ポートアイランド南駅より徒歩約1分, JR新神戸駅から25分



建屋完成イメージ



【計算法棟】

- 延床面積 約10,500㎡
- 建築面積 約 4,300㎡
- 構造 鉄骨造り地上3階地下1階（免震構造）

【研究棟】

- 延床面積 約9,000㎡
- 建築面積 約1,800㎡
- 構造 鉄骨造り地上6階地下1階（免震構造）



施設の建設風景 (平成22年2月2日)



研究棟

計算機棟

居室	計算機室	
居室	計算機筐体	
居室	空調機械室	
居室	空調機	
居室	居室	計算機室
居室		グローバルファイルシステム
空調機械室等	空調機械室	
	空調機	



施設内部

次世代スパコン設置フロア (計算機棟3階)



冷凍機 (熱源機械棟)



太陽光発電パネル (屋上)



フリーアクセス架台



研究棟