



京速コンピュータ「京」の開発状況について

平成23年2月16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 & 計算科学研究機構

横川 三津夫

「次世代スーパーコンピュータ」改め

- 「京(けい)」は、次世代スーパーコンピュータの愛称
 - 一般からの応募の中から選定(平成22年7月).
- 「京」が、1秒間に1京回の足し算, 掛け算ができることに由来
 - “万=10⁴”, “億=10⁸”, “兆=10¹²”, “京=10¹⁶”
- 「アーチ状の門」を表している※)
 - 「京」の活用による輝かしい未来への入り口を表す文字として相応しい漢字.
- ロゴマークも同時に決定

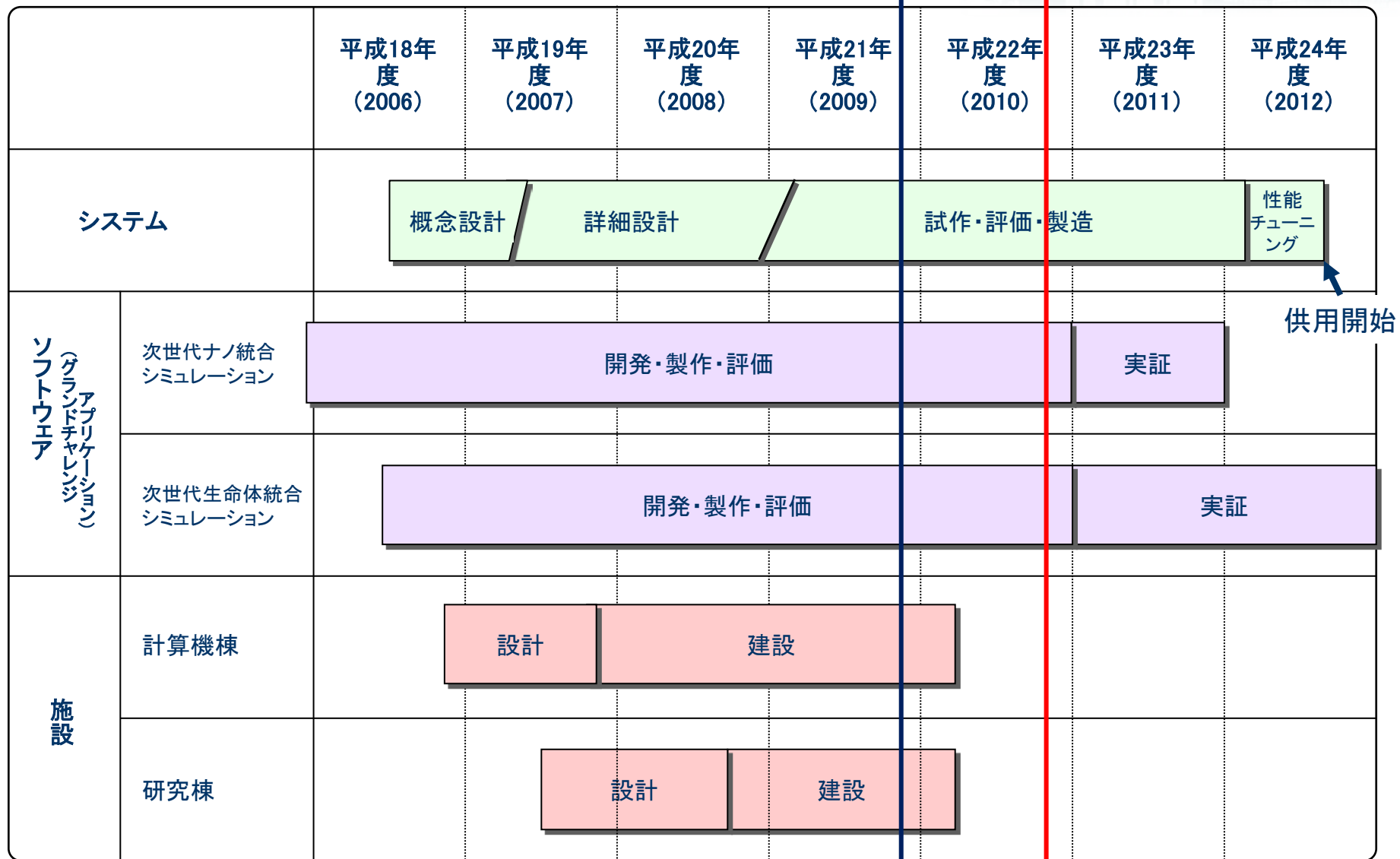
※) 白川静, 字源辞典「字統」, 平凡社 (1984)



唐古・鍵遺跡 <http://inoues.net/ruins/karakokagi.html>

開発スケジュール(平成18年度－平成24年度)

シンポジウム2009 現在

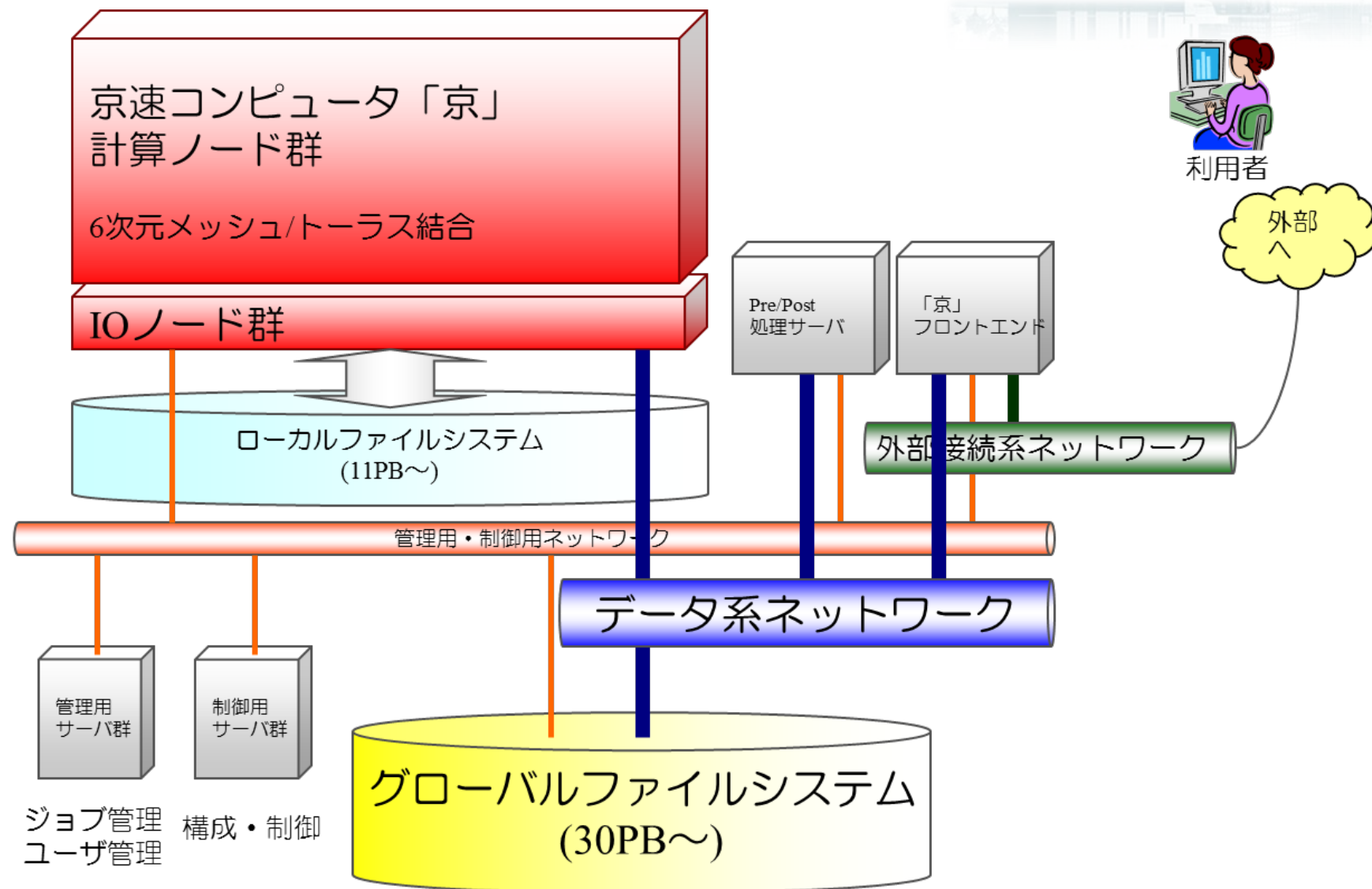


システム全体概要



- ターゲット性能
 - LINPACK 10PFLOPS
- 多くのアプリケーションに高い実効性能を提供
 - ハイパフォーマンスコンピューティング向けCPUを採用
 - マルチコアCPU(8コア), 演算アクセラレータ付き(SIMD機構), FPレジスタ数増強
 - ペタフロップス級のアプリケーション実効性能を目指す
- 高性能・低消費電力CPU
 - 2.2GFLOPS/W, ワット当たり世界トップレベル
 - 45nm CMOS プロセス
 - 小規模システムとして829MFLOPS/W(GREEN500 4位, TOP500 170位)
- 高信頼性システム
 - 「壊れにくい」「壊れても全てが止まらない」「壊れた部分はすぐ直す」
- 利用者にとって使いやすいシステム
 - 世界標準のソフトウェア環境を提供
 - デファクトスタンダードの開発環境を提供

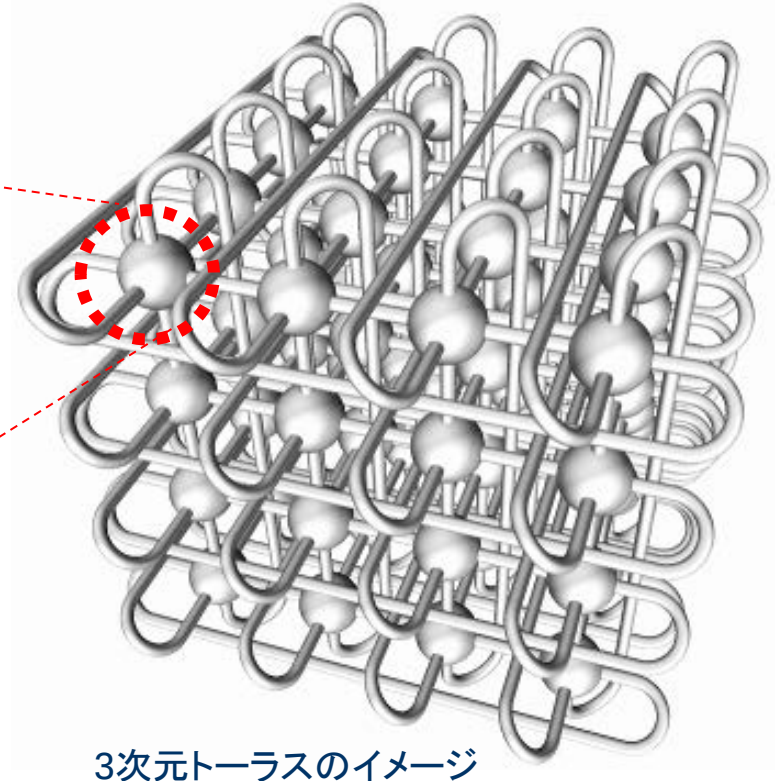
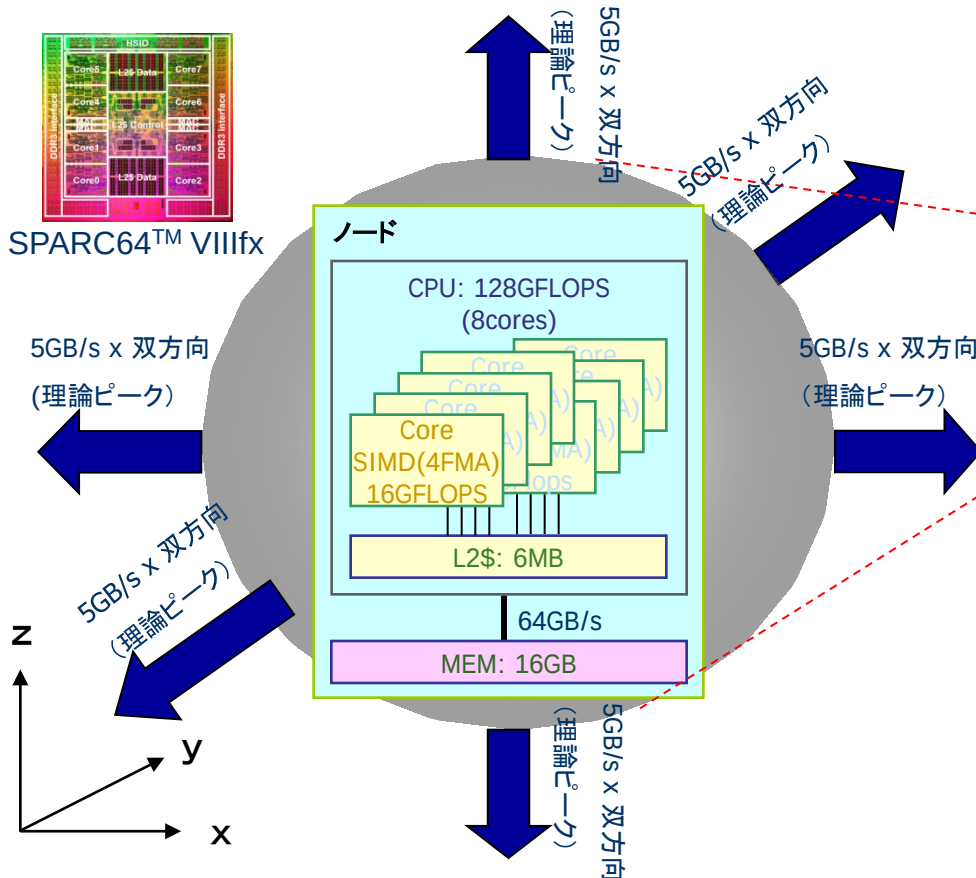
システム構成概要



システム全体の構成

- 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)

- 3次元メッシュトーラスネットワーク: Tofu
 - ユーザ・ビューは3次元トーラス
 - 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1000km

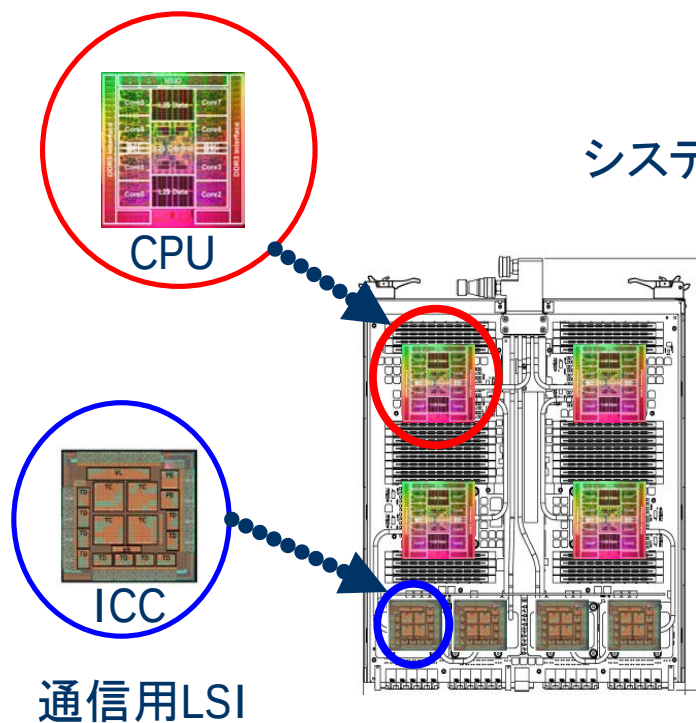


提供: 富士通(株)

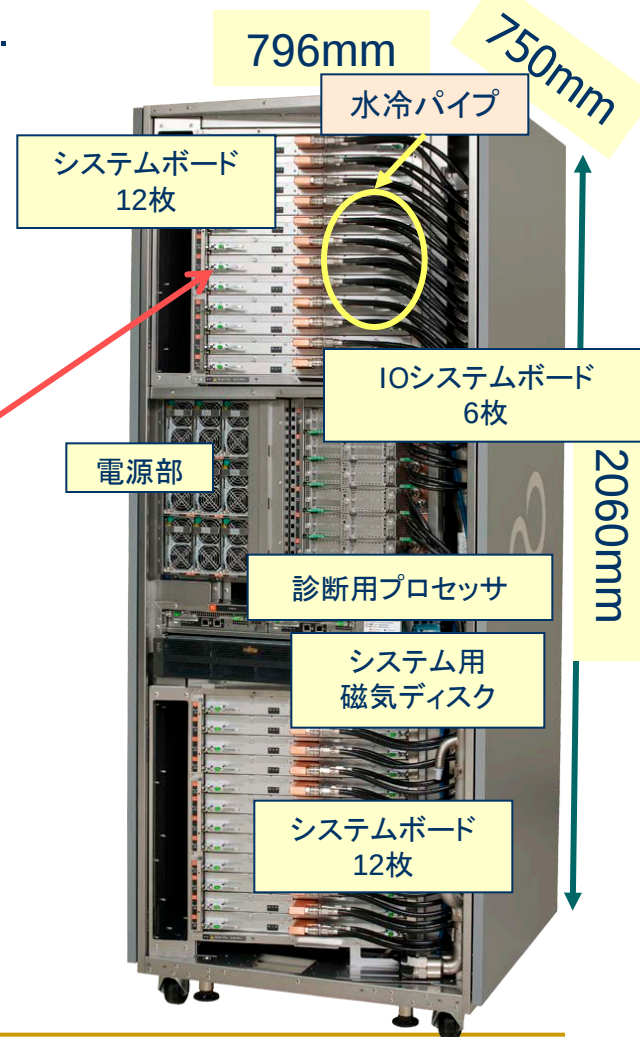
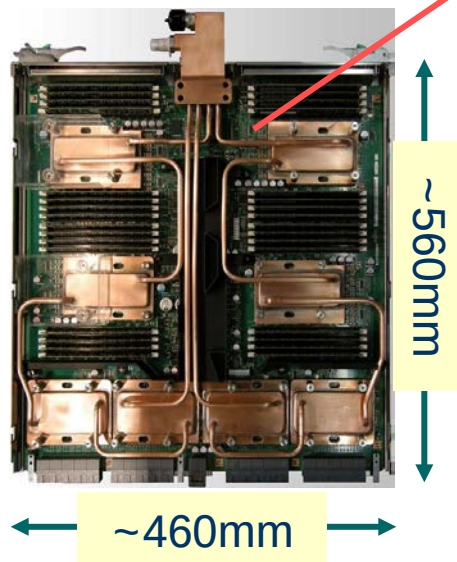
筐体(ラック)の内部構造

- 24枚のシステムボード, IO用システムボード, システム用磁気ディスク装置, 電源部などが一つの箱の中に組込まれている。

- 性能: 12.3TFLOPS/ラック
- サイズ: 796mm x 750mm x 2060mm



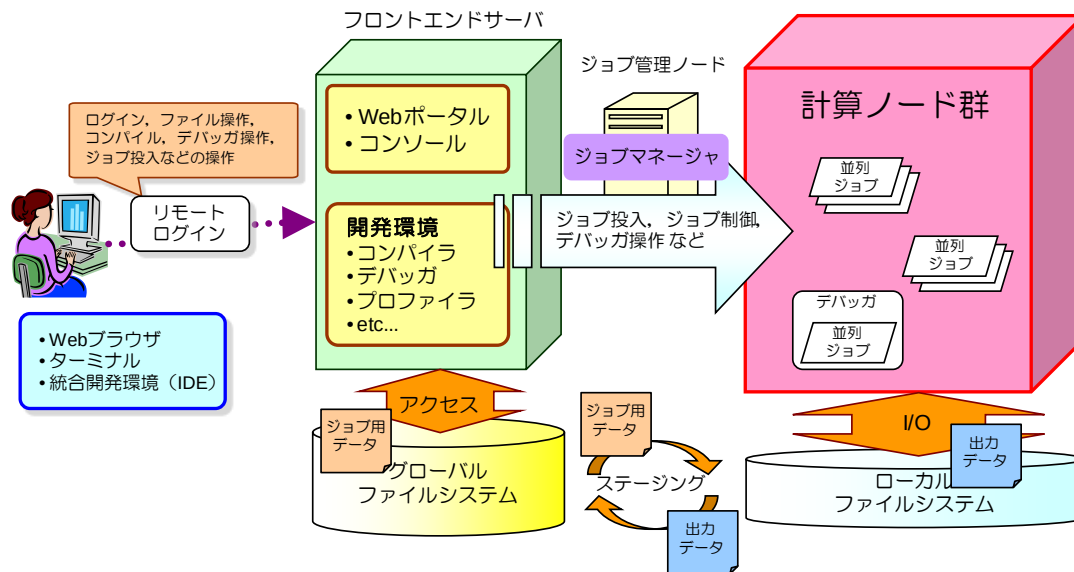
システムボード



システム利用環境概要

- フロントエンドサーバ
 - Linux(RHEL)上で開発環境を提供
 - ターミナルとWebブラウザ(およびWebサービス)のUIを提供
- ステージングによる2階層のファイルシステム
 - グローバルファイルシステム(GFS)とローカルファイルシステム(LFS)
 - ジョブ実行前にGFSからLFSへ実行・入力ファイルを転送(ステージイン)
 - ジョブ終了後にLFSからGFSへ出力ファイル転送(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)



ジョブ実行シナリオ



- グローバルファイルシステム上にファイルを用意
 - 外部から転送の場合もあり
 - 必要に応じてPre/Postサーバなどでデータを処理
- フロントエンド(IDE, ポータル, CUI)からジョブスクリプトを投入
- フロントエンドからジョブを操作
 - 状態監視, キャンセルなど
- 実行結果を取得
 - 必要に応じてPre/Postサーバなどでデータを処理
- 必要に応じて外部へ転送

システムソフトウェア環境



■ ノードOS

■ Linuxベース

- 計算ノードではOSによる演算処理への外乱を極力排除
 - ④ 超並列環境での低ノイズ環境(アプリ実行時間のブレが少ない環境)を目指す
- パフォーマンスカウンタI/FとしてPAPIも提供
- IOノード
 - ④ PCI-ex を有し, Tofu, GbE, FiberChannel, InfiBandの各インターフェイスを搭載

■ ファイルシステム

■ Lustreをベースとした拡張ファイルシステムを採用

- GFS, LFS共にLustreベースのファイルシステム
 - ④ LFSはラック単位にローカリティによる配置最適化やジョブ毎に帯域を(ある程度)保証する機能
 - ④ ステージング帯域のQoS機能

ジョブ実行環境



■ ジョブ実行環境

■ ジョブタイプ

- 会話ジョブ, バッチジョブ, ステップジョブ, バルクジョブなど

■ 様々なMPIジョブの実行が可能

- ハイブリッドMPI, フラットMPIどのジョブが実行可能
- MPIジョブの形態としてSPMD, MPMDなどが実行可能
 - ② 実行に際して様々な制約がある

■ リソース指定

- ノード数, 仮想トポロジ, ステージングのファイル容量, 経過時間, ノード割付ポリシーなどが指定可能

■ ノード割付

■ 全てのジョブに論理的1次元~3次元トラス空間を提供

- ジョブスクリプトに1~3次元形状を指定
 - ② Rankマッピングは様々な方法が可能
 - Rank番号をXYZで自動マッピングやhostfileによるユーザ指定など
- IO占有を指定
 - ② IO性能を確保したいジョブのための指定

プログラム開発環境(1/2)



■ 提供言語

■ Fortran, XPFortran, C/C++

- C/C++: ISO/IEC9899:1999(C99)準拠, ISO/IEC14882:2003(C++2003), (GCCのC/C++拡張を一部対応)
- Fortran: ISO/IEC1539-1:1997(Fortran95)準拠およびFortran2003の一部機能に対応
- OpenMP 3.0に準拠
- CPU拡張を踏まえた最適化コンパイラを提供
(SIMD, 256 FPレジスタ, セクタキャッシュ, コア間バリアなど)

■ 4倍精度演算のサポート

■ Fortran, C/C++

- ソフトウェア処理としてbinary128データ形式をコンパイラで提供
- ハードウェア処理としてdouble-doubleデータ形式の演算ライブラリを提供

■ MPIライブラリ

- MPI2.1規格(OpenMPI 1.4.1をベース)に対応
- Tofuネットワーク(直接網)に最適化されたライブラリ
 - 低レイテンシかつ広帯域

プログラム開発環境(2/2)

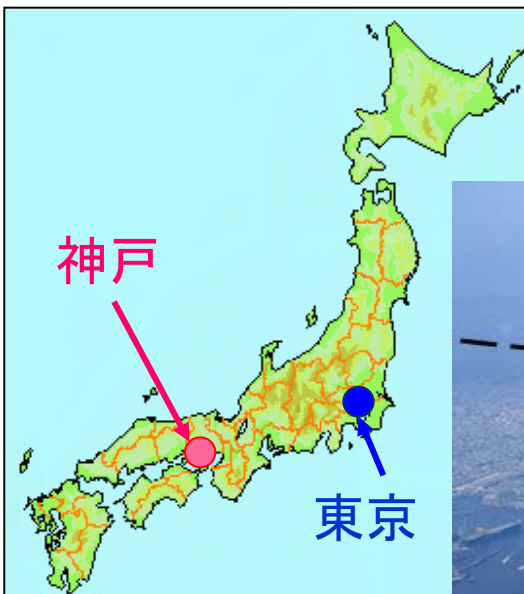


- 数値・科学技術計算ライブラリ
 - BLAS, LAPACK, FFTW(1.2,1.3)およびSSL-II(富士通製科学技術計算ライブラリ)
- IDE環境
 - ユーザ端末で実行するリッチクライアントを提供
- デバッグ
 - 会話型デバッガ(ソースコードデバッガ)
 - プログラム埋め込みデバッグ(ユーザ関数埋め込み)
 - 実行時デバッグ(実行時のデバッグ出力)
- パフォーマンス・チューニング
 - プロファイラ
 - サンプルング情報取得の手段を提供
 - ② ルーチン/プログラム区間の割合/経過時間/ハードウェアモニタ情報/通信時間/プロセス間・Thread間バランスなどのプロファイル情報を採取
 - トレーサー
 - 並列アプリケーションの時系列情報の取得手段を提供
 - ② OTF形式をサポート



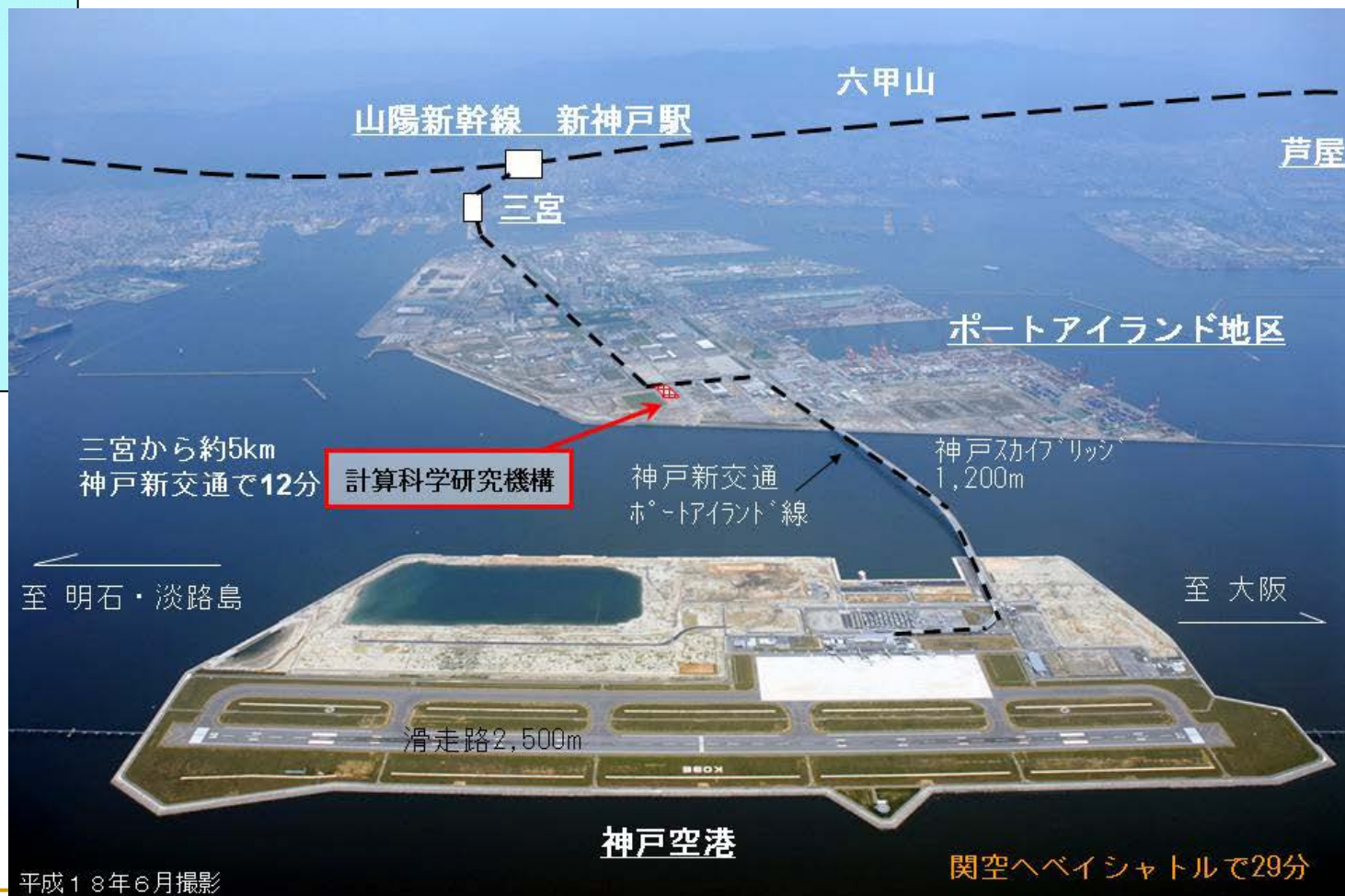
計算科学研究機構 施設の概要

計算科学研究機構



東京から西に450km

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

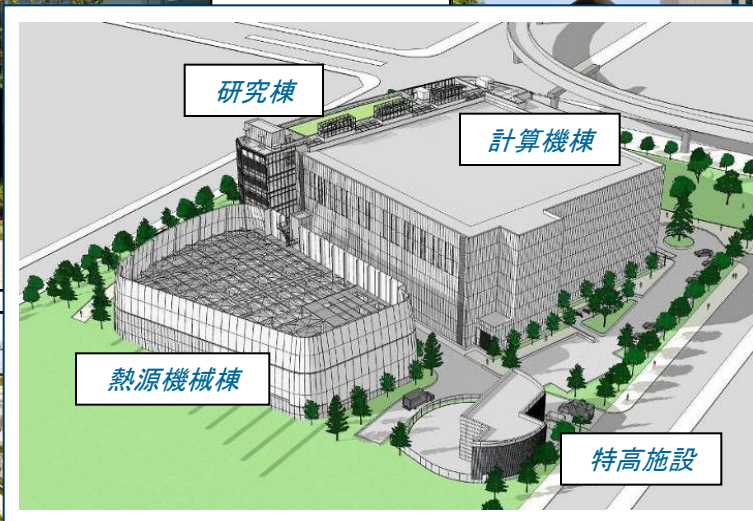


平成18年6月撮影

平成22年5月31日完成



建物の配置



施設の概要

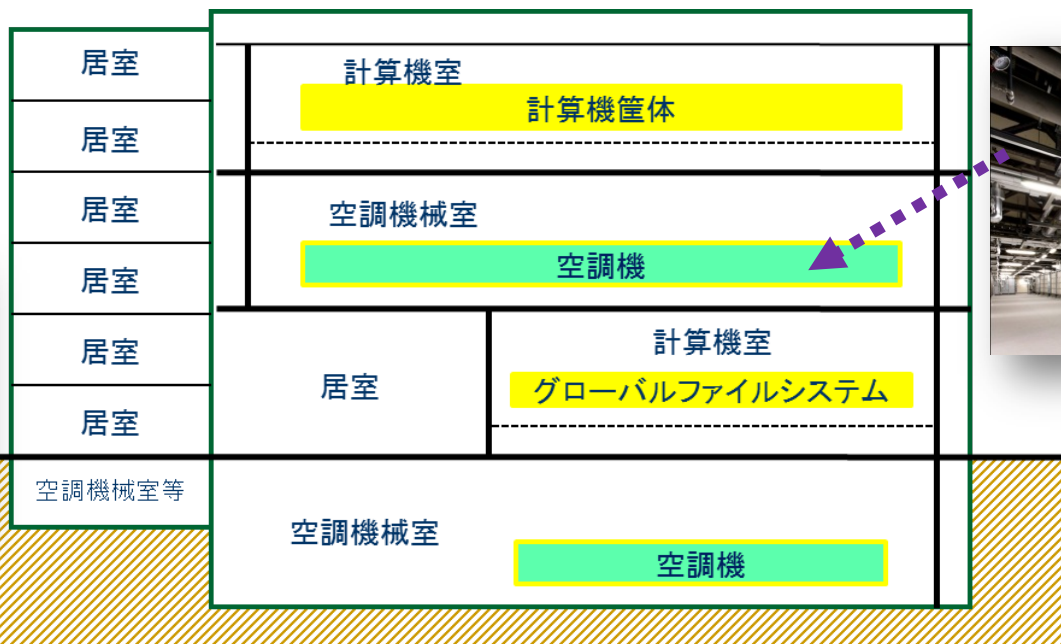


研究棟

- 地上6階, 地下1階(鉄骨造り)
- 建築面積 $\sim 1,800\text{m}^2$, 延床面積 $\sim 9,000\text{m}^2$

計算機棟

- 地上3階, 地下1階(鉄骨造り)
- 建築面積 $\sim 4,300\text{m}^2$, 延床面積 $\sim 10,500\text{m}^2$



熱源機械棟(面積 1900m^2)

吸収式冷凍機
x 4

ターボ型冷凍機
x 3

CGS
(5MW)
x 2

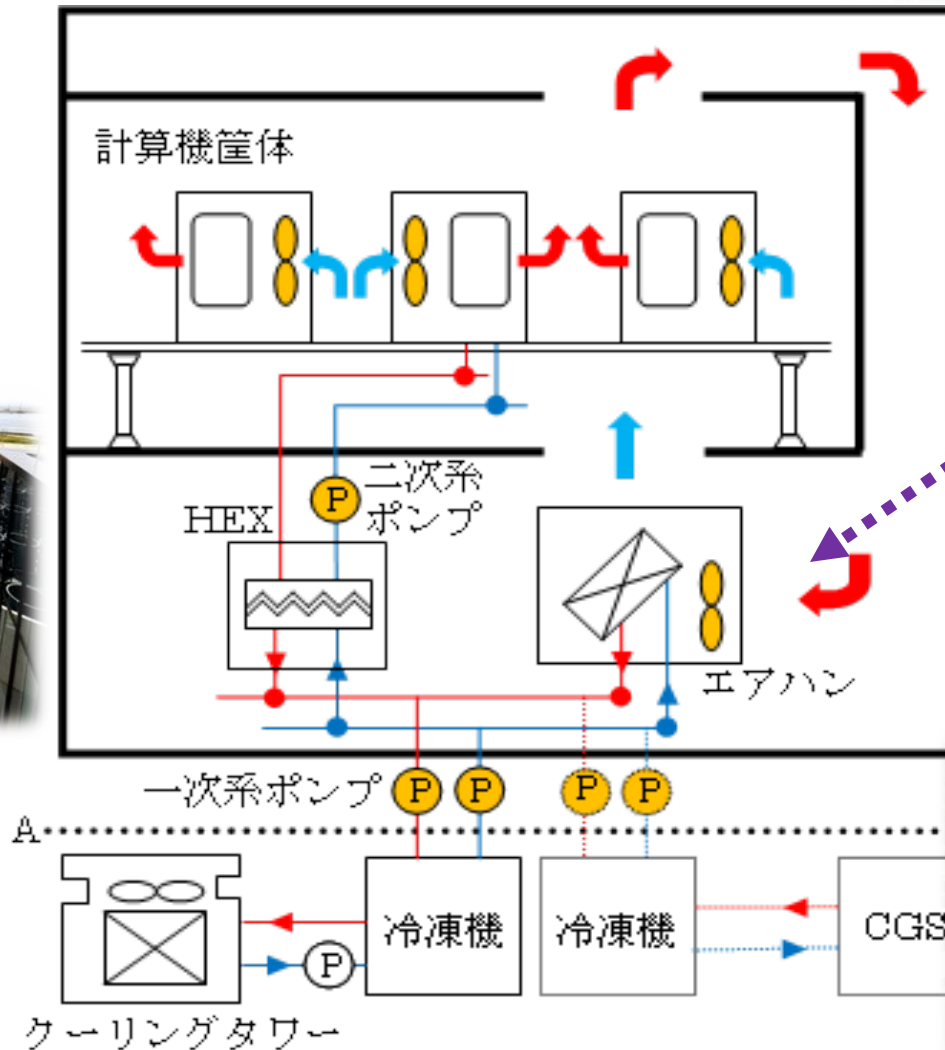
特別高圧変電施設(面積 200m^2)

30MW
77,000V(受電)
→ 6,600V

冷却システム



クーリングタワー



空調機



冷凍機

施設の特徴(1)

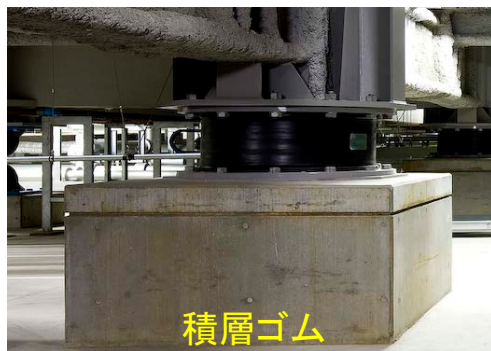


■ 基礎

- 地盤支持力の強化, 液状化対策等を目的として, 埋立部分(地下20m)の地盤改良を実施
地盤改良部を支持層とする2mの直接基礎

■ 耐震性

- 研究棟及び計算機棟は免震構造. 積層ゴムによる免震装置を49か所設置. 振動を素早く抑える鉛ダンパーや鋼性U型ダンパーを設置
- 耐震グレードはSグレード(震度6強レベルの大地震が起きても主要な機能を確保)



■ 耐久性(海浜地区の塩害対策)

- 重要な構造体には防錆対策などを実施
- 計算機棟外壁をアルミパネル, 研究棟外壁をガラス基調として建設

施設の特徴(2)

■ 計算機室構造

- 設置レイアウトの自由度の確保, 均一配置, スパコンの相互接続時の通信ケーブル長の短縮化に対応するため, 計算機室は無柱.
 - 200,000本のケーブル
- 計算機設置フロアの床耐荷重は平均約1トン/m²
 - 「京」筐体当り最大1,500kg



ご清聴ありがとうございました。

A photo in the early evening