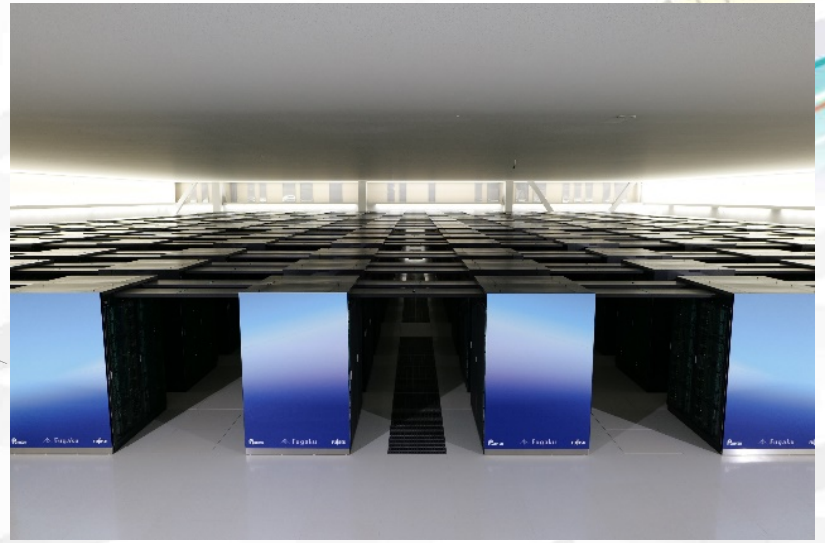
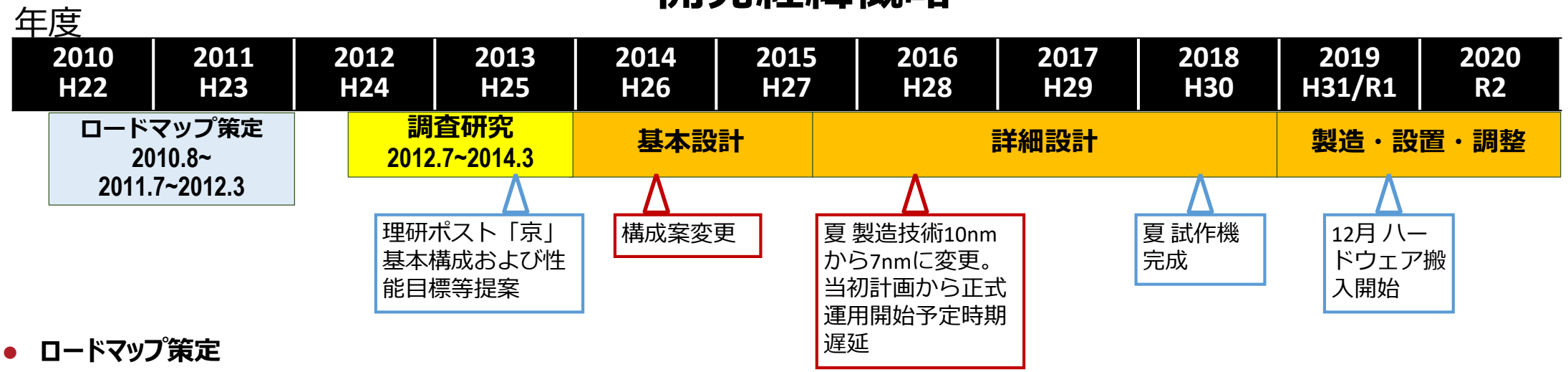
スーパーコンピュータ「富岳」の開発経緯

理化学研究所　計算科学研究センター
フラッグシップ2020プロジェクトリーダ
石川　裕



2021年1月21日（木）13:55-14:45

開発経緯概略



- **ロードマップ策定**
  - 2020年代に実現されるべきアプリケーションとそれを可能とするスパコンの将来像をまとめた以下の白書を策定。当初、草の根的に始まった
    - 計算科学研究ロードマップ白書
    - HPCI技術ロードマップ白書
- **調査研究**
  - 東北大、筑波大、東大の3代表機関が企業と共にそれぞれスパコン基本方式とその性能見積もり等の調査研究を行った
  - 理研はアプリケーション開発者と共に将来の社会的・科学的課題とその解決のためのスパコンで動かすアプリケーションをまとめた
- **基本設計**
  - コンピュータハードウェアおよびソフトウェア、設置条件などの仕様（開発すべき項目等）を決めた
- **詳細設計**
  - 基本設計で確定した仕様を元に、具体的なハードウェア・ソフトウェア設計を行い試作機を制作し評価した
- **製造・設置・調整**
  - 富岳ハードウェアを製造・設置し、ハードウェアおよびソフトウェアの安定化および正式運用に向けた調整作業を行った

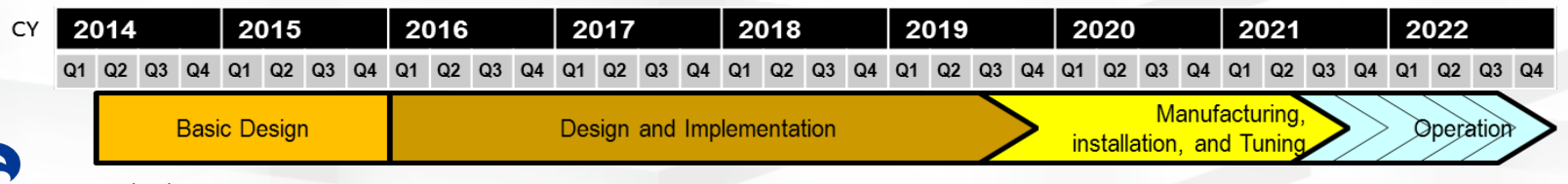
ポスト京（富岳）の開発

- **設計方針**
  1. 課題解決型
     - ポスト京運用後に成果が期待されるアプリケーション群が必要とする性能要求に応える
  2. 協調設計
     - アプリケーション開発者と計算機システム開発者の協調によりアプリケーションおよびシステムを協調設計(co-design)していく
  3. 使い勝手の向上
     - より多くの利用者が容易に使えるようにする
  4. Total Cost of Ownership
     - 省エネ、製造・運用保守経費削減
  5. 拡張性＆社会が欲するニーズに即応
     - ビッグデータ、人工知能

CSTI（総合科学技術・イノベーション会議）において示した目標性能 (2014年10月)

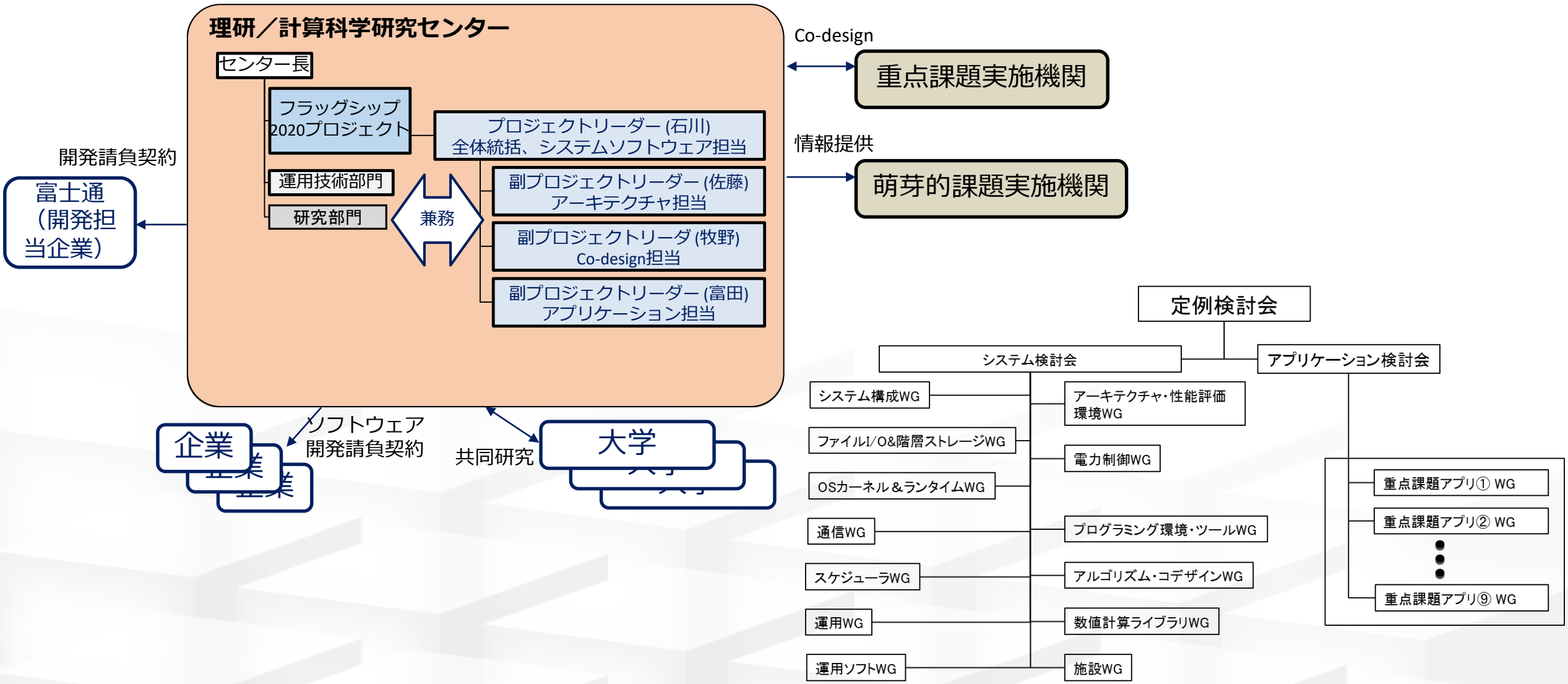
アプリ名	目標性能
GENESIS	100 倍
GENOMON	15 倍
GAMERA	15 倍
NICAM+LETKF	75 倍
NTChem	40 倍
ADVENTURE	15 倍
RSDFT	35 倍
FFB	20 倍
LQCD	50 倍

基本設計：約2年間
詳細設計：約3年間
製造・設置・調整：約2年



	Target Application	
	Program	**Brief description**
①	**GENESIS**	**MD for proteins**
②	**Genomon**	**Genome processing (Genome alignment)**
③	**GAMERA**	**Earthquake simulator (FEM in unstructured & structured grid)**
④	**NICAM+LETK**	**Weather prediction system using Big data (structured grid stencil & ensemble Kalman filter)**
⑤	**NTChem**	**molecular electronic (structure calculation)**
⑥	**FFB**	**Large Eddy Simulation (unstructured grid)**
⑦	**RSDFT**	**an ab-initio program (density functional theory)**
⑧	**Adventure**	**Computational Mechanics System for Large Scale Analysis and Design (unstructured grid)**
⑨	**CCS-QCD**	**Lattice QCD simulation (structured grid Monte Carlo)**

フラッグシップ2020開発体制



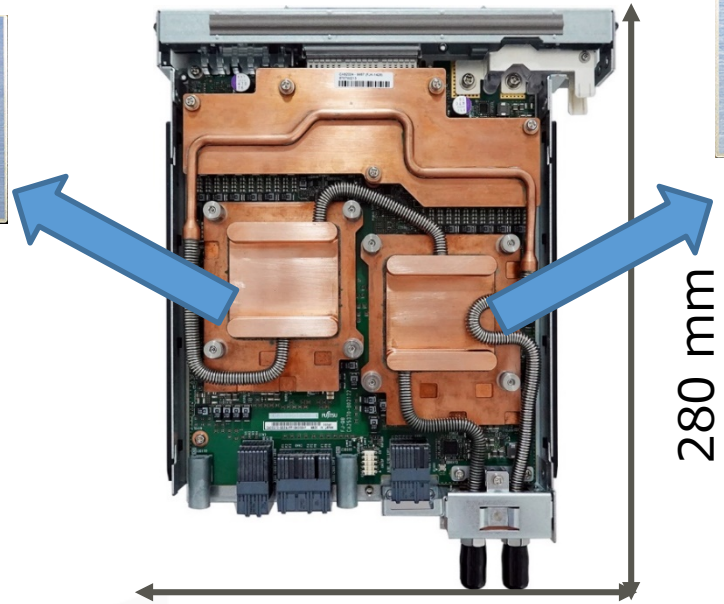
スーパーコンピュータ「富岳」

CMU (CPU Memory Unit)
2ノード x 8

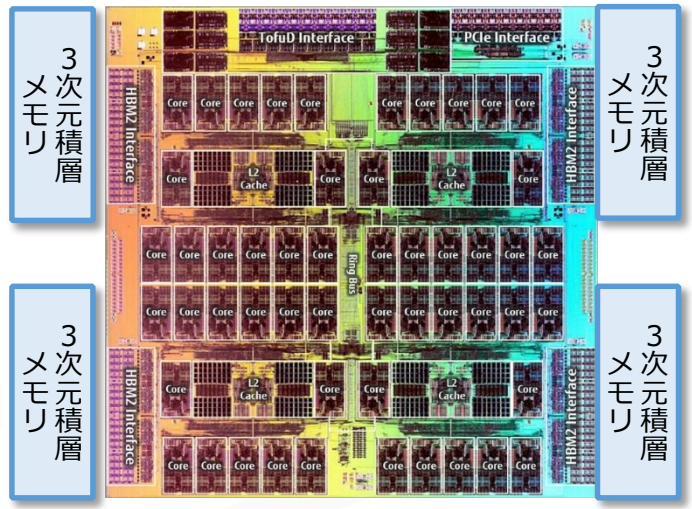
BoB (Bunch of Blade)
16ノード x 3

Shelf
48ノード x 8

Rack
384ノード
写真は2ラック分
85cm x 140 cm x 2,000 cm



A4判紙（210x297mm)位の大きさ



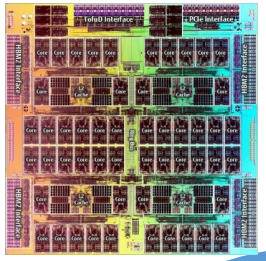
- **storage system**
  - 1st Layer (1.6 TB/16CN)
    - ~ Cache for global file system
    - ~ Temporary file systems
      - Local file system for compute node
      - Shared file system for a job
  - 2nd Layer
    - Fujitsu FEFS: Lustre-based global file system, about 150 PB

総演算性能	**通常モード（CPU動作クロック周波数2GHz）**	**倍精度理論最高値（64bit）488 PFLOPS 単精度理論最高値（32bit）977 PFLOPS 半精度（AI学習）理論最高値（16bit）1.95 EFLOPS 整数（AI推論）理論最高値（8bit）3.90 EOPS**
	ブーストモード（CPU動作クロック周波数2.2GHz）	倍精度理論最高値（64bit）537 PFLOPS 単精度理論最高値（32bit）1.07 PFLOPS 半精度（AI学習）理論最高値（16bit）2.15 EFLOPS 整数（AI推論）理論最高値（8bit）4.30 EOPS
総メモリ容量		4.85 PiB
総メモリバンド幅		163 PB/s
総ノード数		158,976ノード
総ラック数		432ラック(394ノード x 396, 192ノード x 36)

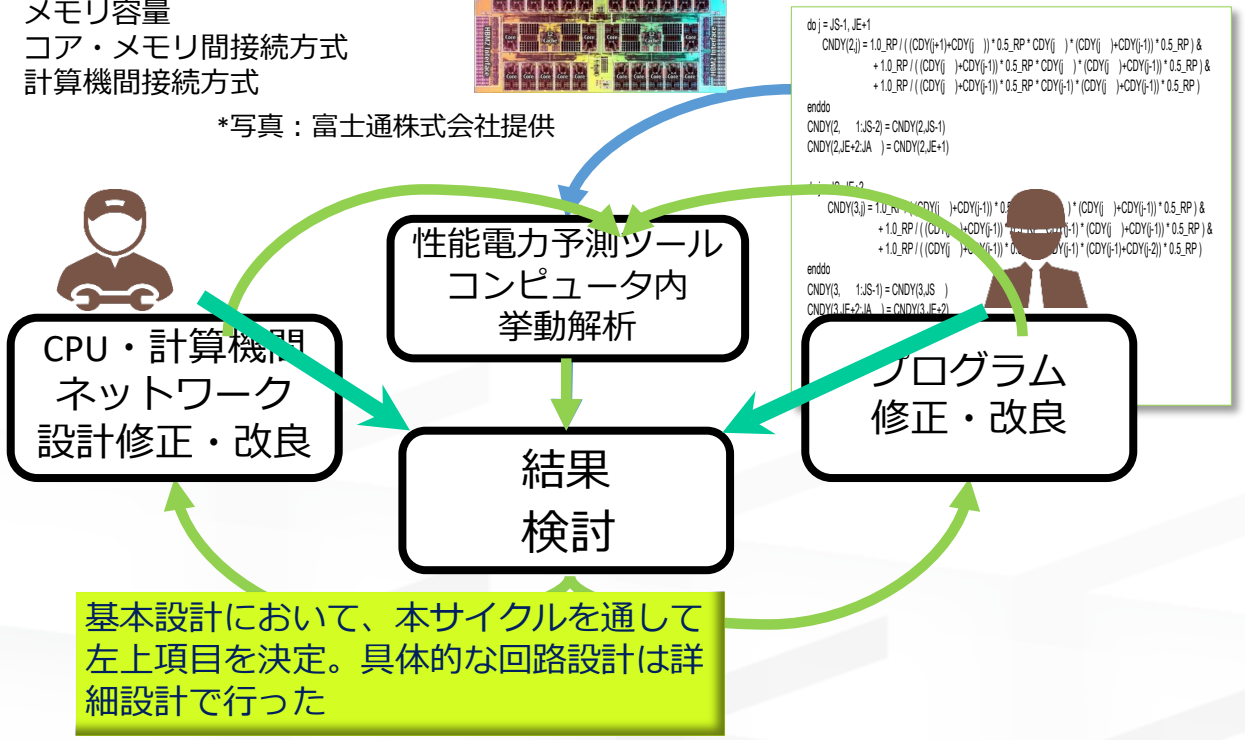
コデザイン

計算機の性能を左右する項目

- ✓ コア数
- ✓ 演算回路数
- ✓ キャッシュ（高速メモリ）容量
- ✓ データ転送性能
- ✓ メモリ容量
- ✓ コア・メモリ間接続方式
- ✓ 計算機間接続方式



*写真：富士通株式会社提供



性能電力予測ツール

- プロジェクト開始時富士通の最新商用スパコンFX100を用いてカーネルコード実行時の性能情報に基づき開発CPUの性能を予測する
- カーネルコードとはアプリケーションコード中実行時間の長いコード断片を切り出したもの

*The Icons made by Freepik from www.flaticon.com

	Target Application	
	Program	**Brief description**
①	GENESIS	MD for proteins
②	Genomon	Genome processing (Genome alignment)
③	GAMERA	Earthquake simulator (FEM in unstructured & structured grid)
④	NICAM+LETKF	Weather prediction system using Big data (structured grid stencil & ensemble Kalman filter)
⑤	NTChem	molecular electronic (structure calculation)
⑥	FFB	Large Eddy Simulation (unstructured grid)
⑦	RSDFT	an ab-initio program (density functional theory)
⑧	Adventure	Computational Mechanics System for Large Scale Analysis and Design (unstructured grid)
⑨	CCS-QCD	Lattice QCD simulation (structured grid Monte Carlo)

	アプリ側改良	システムへの要請
件数	105件	68件
例	SIMD演算利用率向上	TofuDインターコネクト機能強化

- ✓ ファイルI/Oミドルウェア(LLIO)機能およびストレージ性能はアプリのファイルI/Oパターンに基づいて設計した

8~10年先の技術を予測した開発の困難さ

- **最先端半導体微細加工技術を用いた、消費電力に優れ、アプリケーションレベルで高い実効性能を有するCPUの開発**
  - **困難１：**FS2020プロジェクトは、製造・設置完了8年前(2012年)の技術予測に基づいた開発
    - 2世代先の微細加工技術の性能予測に基づく目標性能設定の危険(想定は10nm技術だった)
    - 2015~2016年全世界の半導体製造メーカーの技術発展が鈍化した
    - 目標設定見直し vs.開発遅延
      - ~ 目標再設定：量産体制に入っている16nm技術を採用
      - ~ 遅延：次の7nm技術を採用
        - 開発遅延を選択
        - 遅延期間を用いてAI(深層学習)実行性能を高められるCPU機能(半精度浮動小数点演算）を追加
        - メモリ技術も最先端技術に変更
  - **困難2:**ターゲット微細加工技術決定後の半導体性能（集積度、動作周波数、動作電圧）予測
    - チップ面積を固定した場合に左右されるパラメータ例

集積度	搭載コア数	コア内演算回路数	キャッシュ容量

    - 消費電力を固定した場合に左右されるパラメータ例

チップ面積ｘ集積度	動作周波数＆電圧	計算ノード数

    - ターゲットアプリケーション性能予測を通してこれらパラメータを決定

昔と現在のロードマップを比較しても予測の難しさが分かる

Logic/Foundry Process Roadmaps (for Volume Production)

	2013	2014	2015	2016	2017	2018	2019
Intel		14nm finFET		14nm+	14nm++	10nm	10nm+
GlobalFoundries	28nm		14nm finFET		22nm FDSOI	7nm 12nm	12nm FDSOI
Samsung	28nm	20nm	14nm finFET	28nm FDSOI	10nm	8nm	7nm EUV / 18nm FDSOI
SMIC	28nm						14nm finFET
TSMC		20nm	16nm+ finFET		10nm	7nm 12nm	7nm+ EUV
UMC	28nm				14nm finFET		

出典：https://www.icinsights.com/news/bulletins/Revenue-Per-Wafer-Rising-As-Demand-Grows-For-Sub7nm-IC-Processes/

現在の情報

Logic/Foundry Process Roadmaps (for Volume Prod

	2015	2016	2017	2018	2019	2020
Intel		14nm+	10nm (limited) 14nm++		10nm	10nm
Samsung	28nm FDSOI	10nm		8nm	7nm EUV 6nm EUV	18nm FDSOI 5nm
TSMC	16nm+ finFET	10nm	7nm 12nm		7nm+ EUV	5nm
GlobalFoundries	14nm finFET		22nm FDSOI	12nm finFET	12nm FDSOI	
SMIC	28nm				14nm finFET	
UMC			14nm finFET		22nm planar	

Note: What defines a process "generation" and the start of "volume" production varies fro company, and may be influenced by marketing embelishments, so these points of transition seen as very general guidelines.

Sources: Companies, conference reports, IC Insights

終わりに

- **10年後の要素技術進歩の予測に基づいた目標設定の難しさを経験した**
  - 計画見直し、開発遅延、価値を高めるための機能（AI学習実行高速化）追加
- **当初計画では2020年度からの正式運用だったが1~2年遅れる可能性があった。2020年春には新型コロナ禍対策等の利用者に限定的ではあるが提供できた**

2020年6月、11月ベンチマーク結果

		ノード数	周波数(GHz)	測定値	6月からの性能向上	ピーク性能	効率	使用ノード数割合	第2位性能	2位との性能差（倍率）
2020年11月	Top500	158,976	2.2	442.01 PF	6.4%	537.21 PF	82.3%	100%	148.60 PF	3.0
	HPCG	158,976	2.2	16.00 PF	19.8%	537.21 PF	3.0%	100%	2.92 PF	5.5
	HPL-AI	158,976	2.2	2.00 EF	40.8%	2.14 EF	93.2%	100%	0.55 EF	3.6
	Graph500	158,976	2.2	102.95 TTeps	45.0%			100%	23.75 Tteps	4.3
2020年6月	Top500	152,064	2.2	415.53 PF		513.85 PF	80.9%	96%	148.60 PF	2.8
	HPCG	138,240	2.2	13.36 PF		467.14 PF	2.9%	87%	2.92 PF	4.6
	HPL-AI	126,720	2.0	1.42 EF		1.55 EF	91.3%	80%	0.55 EF	2.5
	Graph500	92,160	2.2	70.98 TTeps				58%	23.75 TTeps	3.0

性能向上の要因はノード数が増えただけでなく前回に比べてシステムソフトウェアの調整が進んだ結果（OSノイズ削減、ユーザ利用可能メモリ容量増大）に加えて
- HPLは通信時間の削減や1ノード上の問題サイズを大きく出来た結果。
- HPCGは計算の高性能化を図っている。
- HPL-AIはさらなる計算・通信の高性能化を図っている。
- Graph500は、１ノードあたりの問題サイズを大きくしている。