東京大学
THE UNIVERSITY OF TOKYO

FX10スーパーコンピュータシステムについて

東京大学情報基盤センター
スーパーコンピューティング部門
http://www.cc.u-tokyo.ac.jp/
問合先：uketsuke@cc.u-tokyo.ac.jp

2012年6月13日版

- 背景
- FX10スーパーコンピュータシステム概要
- スケジュール
- 運用・サービス
  - トークン制
  - 教育利用，若手利用
  - 企業利用
  - トライアルユース
  - 大規模HPCチャレンジ
- 試験運転期間のサービス
- 将来展望
- 質疑

東大センターのスパコン（～2011.09E）

1システム～6年，3年周期でリプレース

HITACHI SR11000 model J2

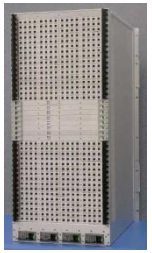
Total Peak performance	: 18.8 TFLOPS
Total number of nodes	: 128
Total memory	: 16384 GB
Peak performance per node	: 147.2 GFLOPS
Main memory per node	: 128 GB
Disk capacity	: 94.2 TB

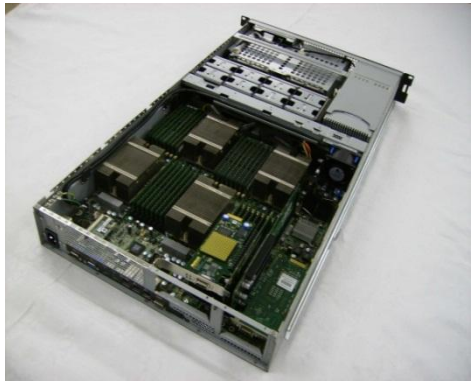
IBM POWER5+ 2.3GHz

T2K（東大）（HA8000クラスタシステム）

Total Peak performance	: 140 TFLOPS
Total number of nodes	: 952
Total memory	: 32000 GB
Peak performance per node	: 147.2 GFLOPS
Main memory per node	: 32 GB, 128 GB
Disk capacity	: 1 PB

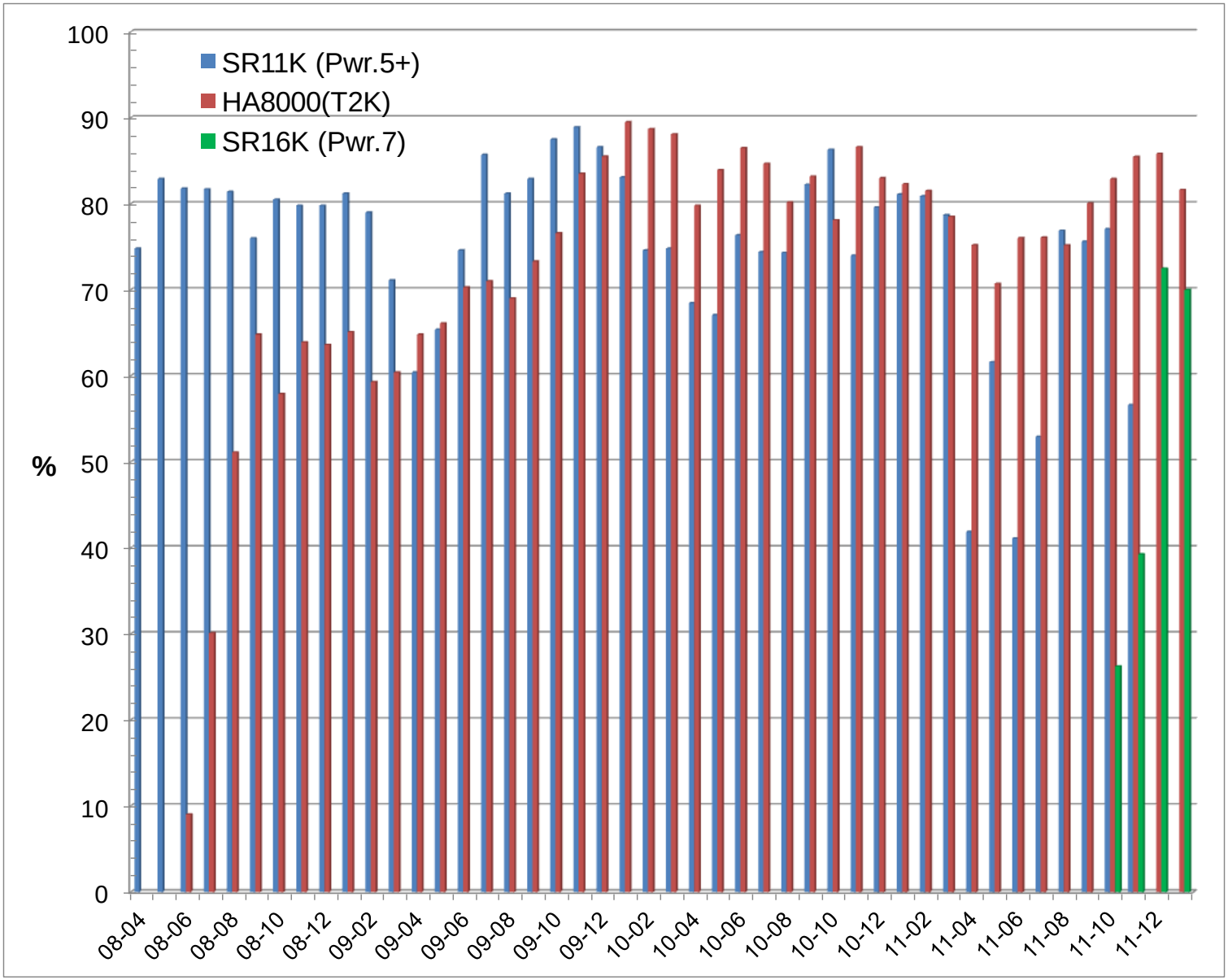
AMD Quad Core Opteron 2.3GHz



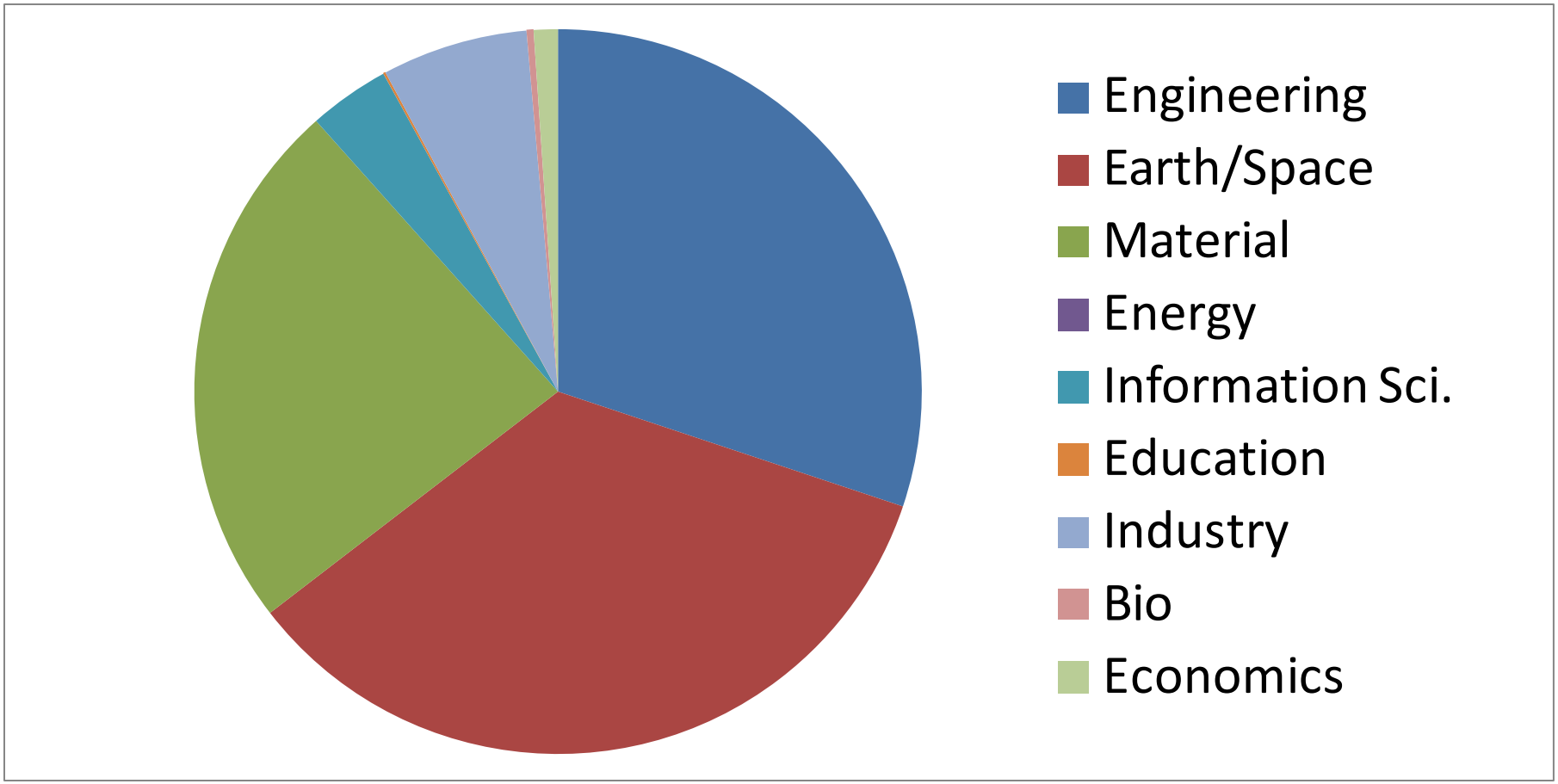


東大センターのスパコン（~2012.01E）

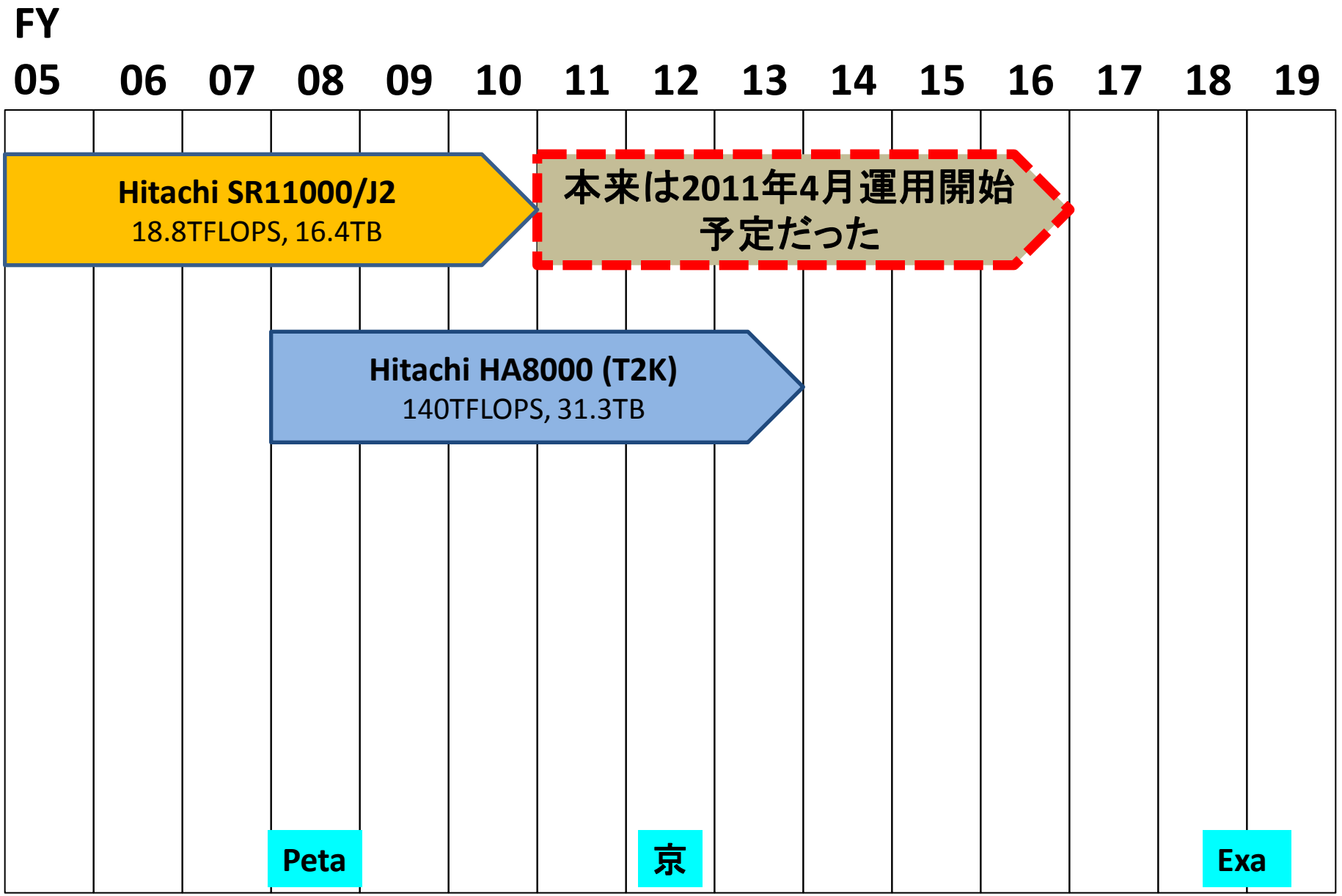
利用者：SR11K-約490名，SR16K-約360名，HA8000-約1,100名



利用ノード時間積による利用分野
T2K：FY.2011（2012.1月末時点）
専用キュー＋教育＋企業



東大情報基盤センターのスパコン



調達の経緯(1/3):2つのシステム

- 2009年4月頃から次期システムに関する検討を開始
- 2システムの導入
  - PFLOPS級のMPPを入れたい，でも「いわゆるSRユーザー」も無視できない
  - 複合型超並列スーパーコンピューターシステム(当初案)
    - 大規模SMP並列スーパーコンピューター(150 TFLOPS級)(SMP)
    - 大規模超並列スーパーコンピューター(850 TFLOPS級)(MPP)
      - コプロセッサ，アクセラレータ無し
  - 最終的に調達を分離
    - SMP(50TFLOPS):SR後継機
      - 継続性重視
    - MPP(1PFLOPS，総メモリバンド幅400TB/sec以上)
      - 計算性能～消費電力のバランス，コンパクト性
      - ファイルシステム性能
      - オープンソースライブラリ・アプリケーション

調達の経緯(2/3):柏移転

- 柏キャンパスへの移転
  - 敷地，電力容量がこのまま(浅野地区)では不足
  - 2009年秋に3部局合同棟(第2総合研究棟，既取得地)への入居決定
    - この時点で2011年3月竣工予定
- その他様々な要因により，2010年10月時点で以下のように決定
  - SMP(稼働開始:2011年10月)，MPP(同:2012年1月)共に柏の新棟へ設置
  - Hitachi SR11000/J2のレンタルを2011年10月まで延長

調達の経緯(3/3):東日本大震災

- 東日本大震災
  - 既存システムの運用に影響
  - 「第2総合研究棟」竣工遅延
    - 電源・空調工事遅延
  - MPP
    - 入札中止(2011年4月末)
    - 仕様・調達スケジュール変更
      - 電力事情の考慮
        - » 性能～消費電力から，限られた消費電力(空調込み2MW)で最大の性能を得られるシステムへ(性能に加点)
      - ピークカットを考慮し，柔軟な運用が可能となるような要求を付加
        - » 90分以内に設定変更可能
    - 2012年4月運用開始へ
  - SMP
    - Hitachi SR16000/M1に決定済(開札済)
    - 電源・空調工事が間に合わない:柏での運用を断念

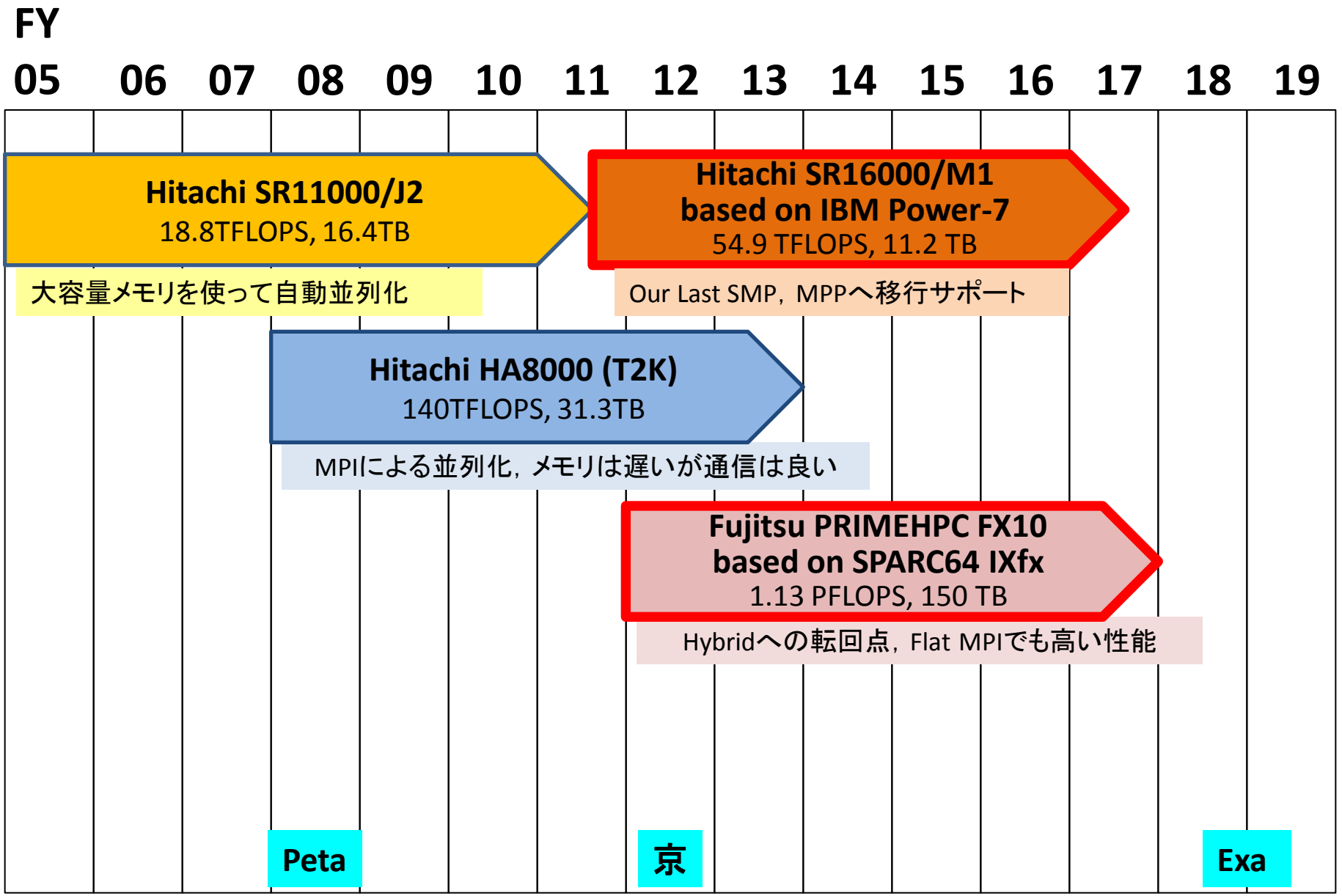
MPP仕様の変遷

		2011年2月25日	2011年8月26日
全 体	**ピーク性能（TFLOPS）**	**1,000以上**	**650以上**
	メモリ容量（TB）	150以上	100以上
	総メモリバンド幅（TB/sec.）	400以上	225以上
	ディスク容量（TB）	1,500以上	2,000以上
	最大消費電力（Linpack） SR11K+T2Kで2MW弱 （空調等込み）	**2.6 MVA以下**	**1.4 MW以下**
稼動開始時期		2012年1月17日	2012年4月 2日
入札説明会		2011年2月25日	2011年8月26日
応札締切日		2011年4月26日	2011年10月6日
開札日		2011年5月31日	2011年11月4日

新システム

- **SMP：Hitachi SR16000/M1**
  - SR16000システム（SMP）（Yayoi）
  - ピーク性能 54.9 TFLOPS
  - 56計算ノード
    - IBM POWER 7, 32 cores/node, 200 GB/node
  - 2011年10月3日より試行運用，11月25日より本運用開始
  - 大容量メモリノードを有するタイプのシステム（SMPと呼んでいる）の導入はこれで最後（データサーバー等除く）
    - 利用者は6年以内に並列化を進め，MPP等へ移行する
      - センターも講習会，個別相談などできる限りのサポートをする
- **MPP: Fujitsu PRIMEHPC FX10**
  - FX10スーパーコンピュータシステム（Oakleaf-FX）
  - ピーク性能　1.13 PFLOPS
  - 4,800計算ノード
    - SPARC64 IXfx，16 cores/node, 32GB/node

東大情報基盤センターのスパコン

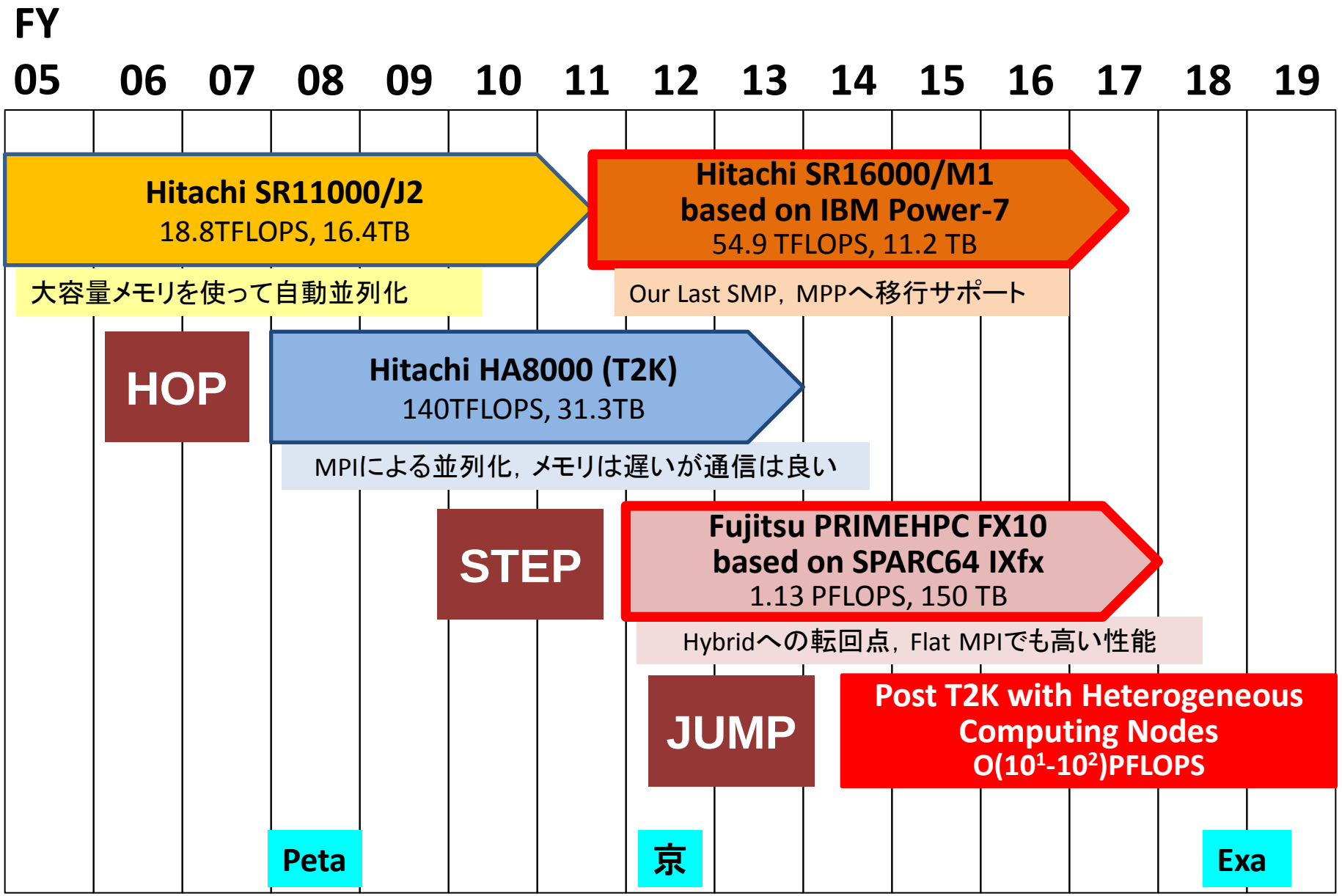


- 背景
- **FX10スーパーコンピュータシステム概要**
- スケジュール
- 運用・サービス
  - トークン制
  - 教育利用，若手利用
  - 企業利用
  - トライアルユース
  - 大規模HPCチャレンジ
- 試験運転期間のサービス
- 将来展望
- 質疑

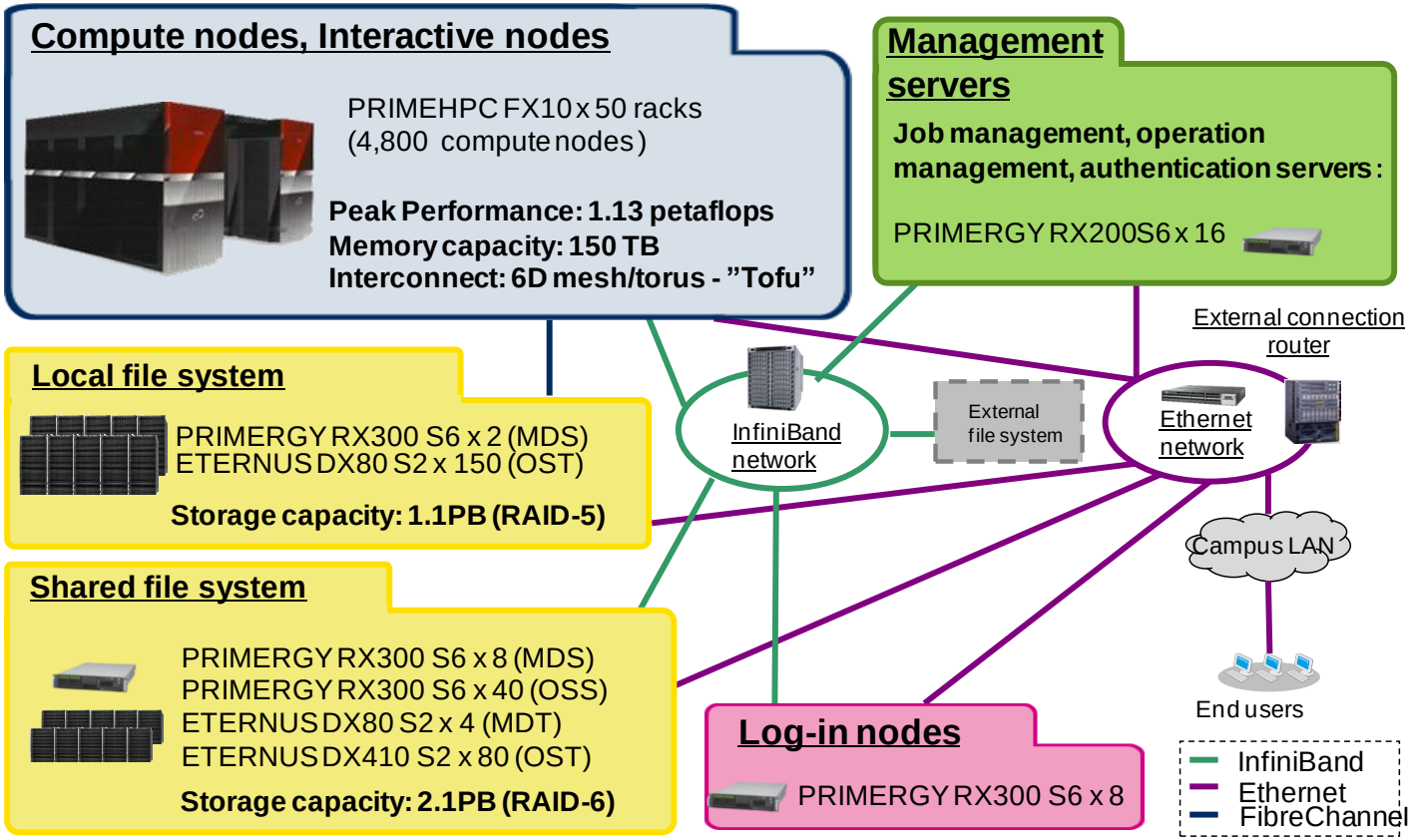
新MPPの位置づけ:三段跳びの「Step」

- Hop
  - HA8000（T2K）, Homogeneous Compute Nodes
  - O(10^{-1}) PFLOPS
  - Flat MPI
- Step
  - PRIMEHPC FX10, Homogeneous
  - O(10^{0}) PFLOPS
  - MPI + OpenMP，但しFlat MPIも充分速くなければ使えない
- Jump
  - Post T2K, Heterogeneous
    - 省電力，メモリバンド幅：Heterogeneousな計算ノード
  - O(10^{1}-10^{2}) PFLOPS
  - MPI + X（OpenMP, CUDA, OpenCL … OpenACC）
- その先にExaがあるはず

東大情報基盤センターのスパコン



FX10 Supercomputer System

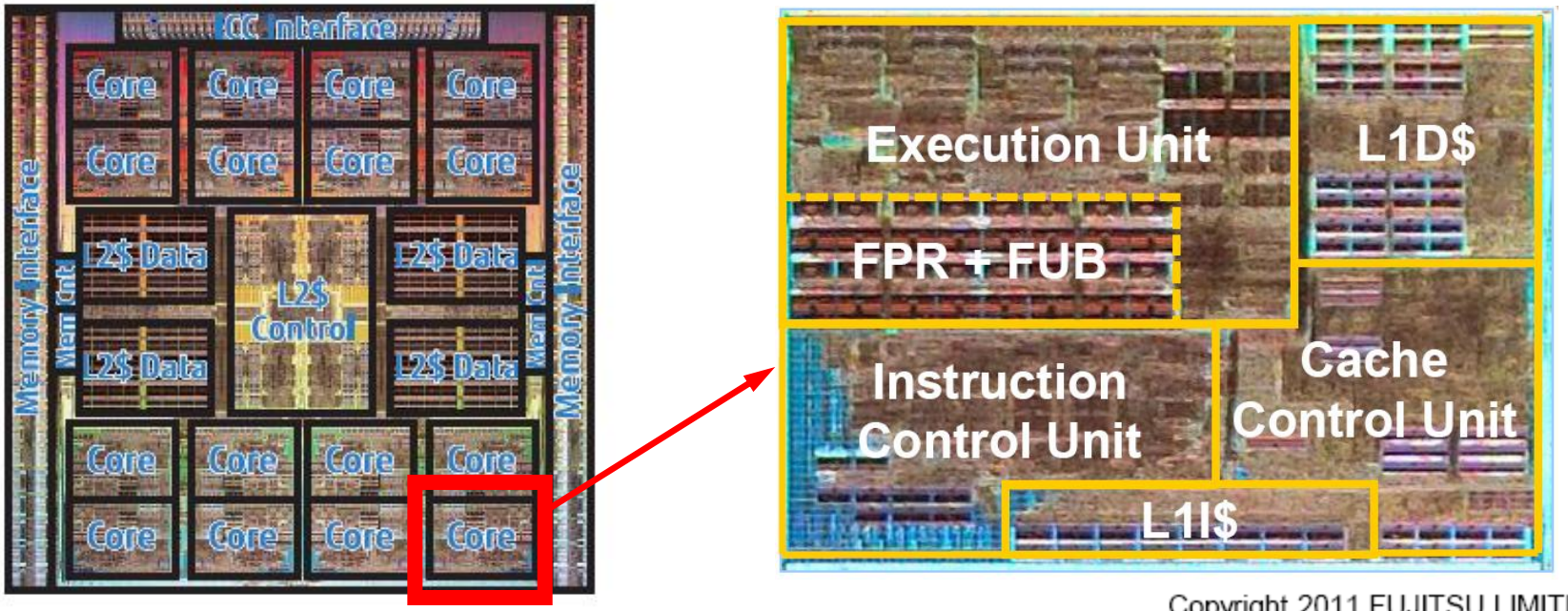


- Aggregate memory bandwidth: 398 TB/sec.
- Local file system for staging with 1.1 PB of capacity and 131 GB/sec of aggregate I/O performance (for staging)
- Shared file system for storing data with 2.1 PB and 136 GB/sec.
- External file system: 3.6 PB

FX10（Oakleaf-FX）の概要

- ピーク性能1.13PFLOPS
- 周辺装置込み最大消費電力<1.40MW（Linpack最大時）
  - 空調を含めても2.00MW未満
- 6次元メッシュ／トーラスネットワーク
  - Tofuインターコネクト
  - リンク当りバンド幅：5GB/sec×2，Bi-Sectionバンド幅：6 TB/sec
- 高性能ファイルシステム
  - FEFS（Fujitsu Exabyte File System）（Lustreベース）
- 通常運転～省電力運転の柔軟な切り替え
- 「京」との互換性
- 多様なオープンソースライブラリ・アプリケーション
- Flat-MPI，Hybrid共に高い計算性能

SPARC64™ IXfx

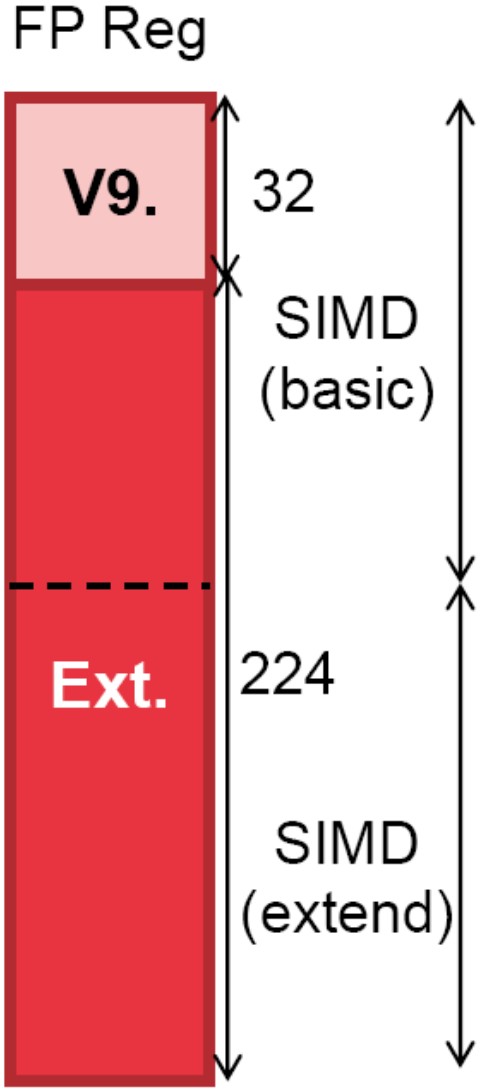


CPU	SPARC64™ IXfx 1.848 GHz	SPARC64™ VIIIfx 2.000 GHz
Number of Cores/Node	16	8
Size of L2 Cache/Node	12 MB	6 MB
Peak Performance/Node	236.5 GFLOPS	128.0 GFLOPS
Memory/Node	32 GB	16 GB
Memory Bandwidth/Node	85 GB/sec (DDR3-1333)	64 GB/sec (DDR3-1000)

HPC-ACE：HPC向け命令セット拡張

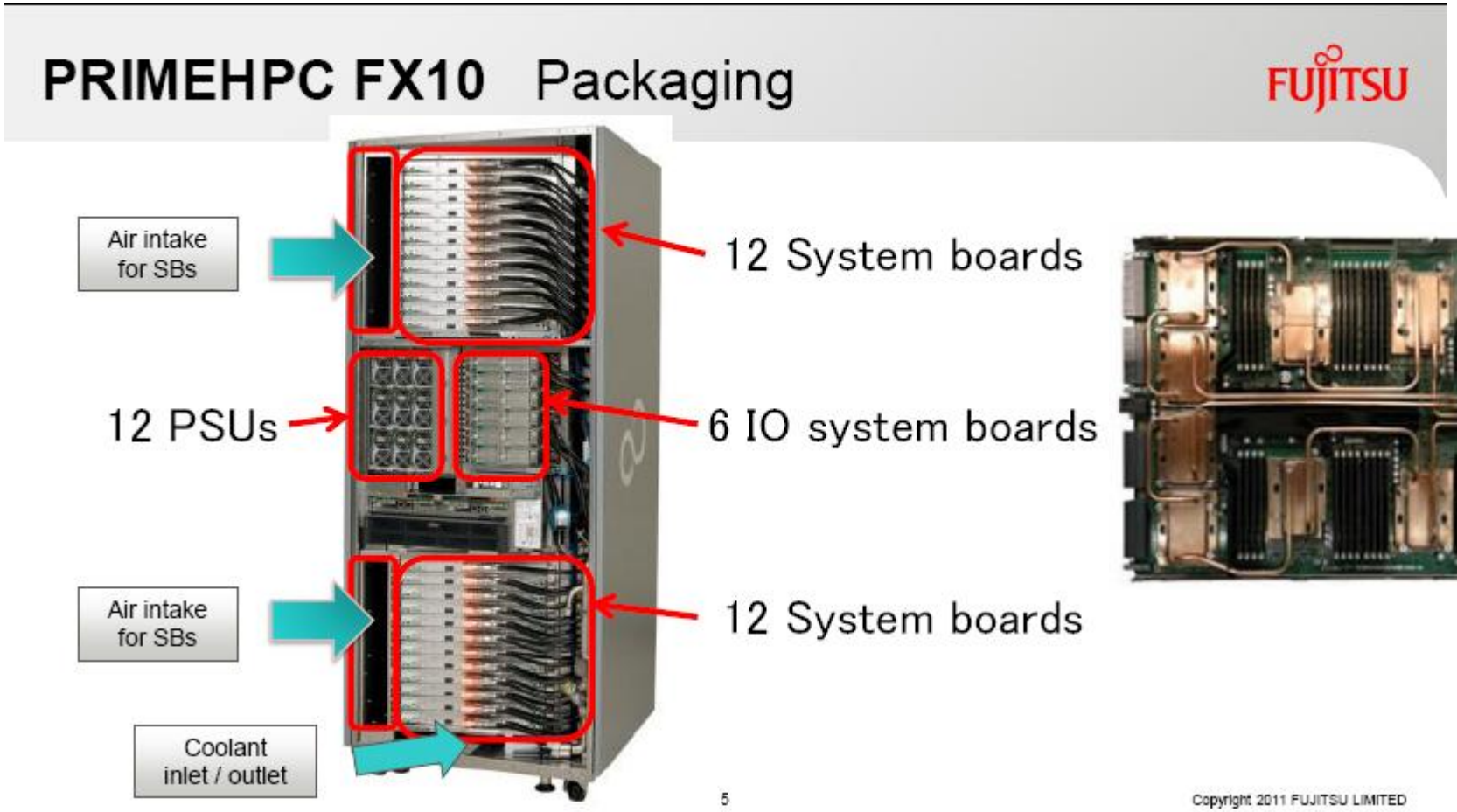
High Performance Computing – Arithmetic Computational Extensions

- SPARC-V9命令セットアーキテクチャに対するHPC向け拡張命令セット
  - 高性能＋省電力
- レジスタ数拡張
  - 浮動小数点演算レジスタ32→256
- ソフトウェア制御可能キャッシュ
  - セクタキャッシュ
  - 再利用頻度の高いデータをキャッシュメモリに保持
- 高速化・最適化
  - 条件付実行命令（if文を含むループ）
  - 三角関数，除算，平方根



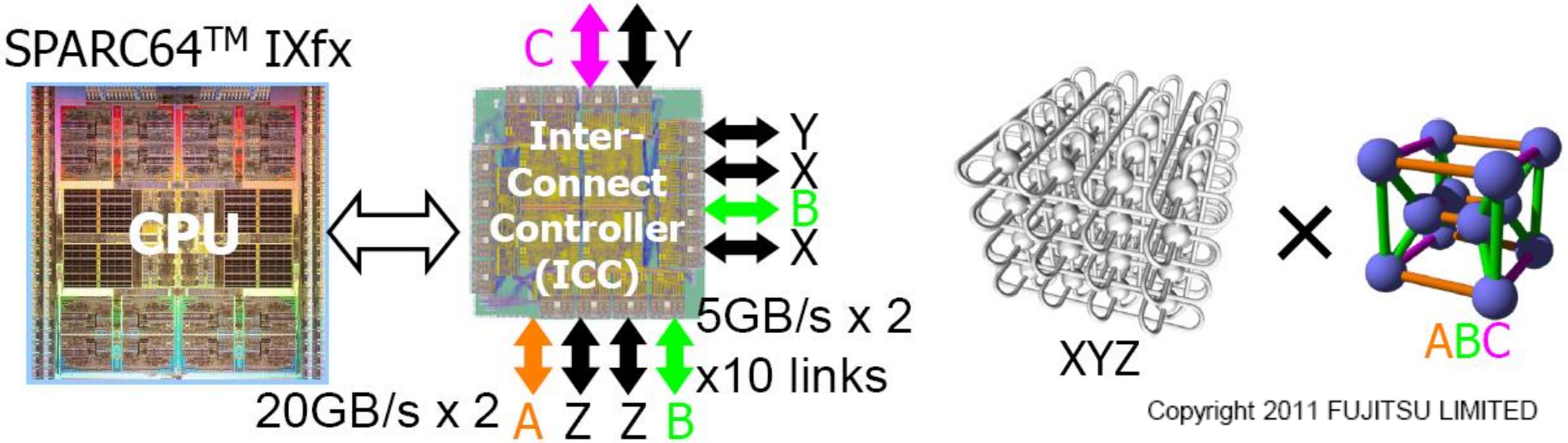
ラック構成

- システムボード：4ノード
- 1ラック：24システムボード，96ノード
- 50ラック，4,800ノード，76,800コア



Tofuインターコネクト

- ノードグループ
  - 12ノード
  - A軸・C軸：システムボード内4ノード結合，B軸：3ボード結合
- 6D：（X,Y,Z,A,B,C）
  - ABC 3D Mesh：ノードグループの12ノードを結合：　2×2×3
  - XYZ 3D Mesh：ABC 3D Meshグループを結合：　10×5×8
- ネットワークトポロジーを指定したJob Submission可能
  - 実行されたXYZは知ることができる



ファイルシステム

- 3システム（ローカル・共有・外部ファイルシステム）
  - HA8000 クラスタシステム，SMPとはファイル共有していない
  - バックアップは各自で実施してください
- ローカルファイルシステム（ステージング用）
  - 1.1PB，131GB/sec.
- 共有ファイルシステム
  - 2.1PB，136GB/sec.
  - ログイン・インタラクティブ・計算ノードからアクセス可能
  - ホームディレクトリ
    - グループ：全体，個人（50GB）
- 外部ファイルシステム
  - 3.6PB
  - ワーク領域．
  - 月末処理日に，最終更新日から1年以上経過したファイルを削除．
  - ただしファイルシステムを圧迫するようであれば，予定に関係なく削除．

ソフトウェア構成

項目	計算・インタラクティヴノード	ログインノード
OS	専用OS（XTCOS）	Red Hat Enterprise Linux
コンパイラ	富士通社製コンパイラ Fortran 77/90 コンパイラ C／C++ コンパイラ GNU コンパイラ GCC，g95 （FX向け拡張無し）	富士通社製コンパイラ（クロス環境） Fortran 77/90 コンパイラ C／C++ コンパイラ GNU コンパイラ（クロス環境） GCC，g95 （FX向け拡張無し）
ライブラリ	富士通社製ライブラリ SSL II (Scientific Subroutine Library II)，C-SSL II，SSL II/MPI その他ライブラリ BLAS，LAPACK，ScaLAPACK，FFTW，SuperLU，PETSc，METIS， SuperLU_DIST，Parallel NetCDF	
アプリケーション	OpenFOAM，ABINIT-MP，PHASE， FrontFlow/blue FrontISTR，REVOCAP	
ファイルシステム	FEFS	
フリーソフトウェア	bash, tcsh, zsh, emacs, autoconf, automake, bzip2, cvs, gawk, gmake, gzip, make, less, sed, tar, vim など	

GeoFEM Benchmark: ICCG for FEM Performance of a Node: Flat MPI

	SR11K/J2	SR16K/M1	T2K	FX10	京
Core #/Node	16	32	16	16	8
Peak Performance (GFLOPS)	**147.2**	980.5	**147.2**	236.5	128.0
STREAM Triad (GB/s)	**101.0**	264.2	**20.0**	64.7	43.3
B/F	0.686	0.269	0.136	0.274	0.338
GeoFEM (GFLOPS)	**19.0**	72.7	**4.69**	16.0	11.0
% to Peak	12.9	7.41	3.18	6.77	8.59
LLC/core (MB)	**18.0**	4.00	**2.00**	0.75	0.75

疎行列ソルバー：Memory-Bound

ノード当りピーク性能は同じなのにこの差！

GeoFEM Benchmark: ICCG for FEM Performance of a Node: Flat MPI

	SR11K/J2	SR16K/M1	T2K	FX10	京
Core #/Node	16	32	16	16	8
Peak Performance (GFLOPS)	**147.2**	**980.5**	147.2	236.5	128.0
STREAM Triad (GB/s)	101.0	264.2	20.0	64.7	43.3
B/F	**0.686**	**0.269**	0.136	0.274	0.338
GeoFEM (GFLOPS)	**19.0**	**72.7**	4.69	16.0	11.0
% to Peak	**12.9**	**7.41**	3.18	6.77	8.59
LLC/core (MB)	18.0	4.00	2.00	0.75	0.75

B/F悪い割にSR16K/M1は健闘

Linpackは約880 GFLOPS/node（89.8%）

GeoFEM Benchmark: ICCG for FEM Performance of a Node: Flat MPI

	SR11K/J2	SR16K/M1	T2K	FX10	京
Core #/Node	16	32	16	16	8
Peak Performance (GFLOPS)	147.2	980.5	**147.2**	**236.5**	**128.0**
STREAM Triad (GB/s)	101.0	264.2	20.0	64.7	43.3
B/F	0.686	0.269	**0.136**	**0.274**	**0.338**
GeoFEM (GFLOPS)	19.0	72.7	**4.69**	**16.0**	**11.0**
% to Peak	12.9	7.41	**3.18**	**6.77**	8.59
LLC/core (MB)	18.0	4.00	2.00	0.75	0.75

空調込み消費電力あたり性能
（年間推定電気料金より推算）

T2K:　1.00 unit
FX10:　7.65 unit

GeoFEM Benchmark: ICCG for FEM Performance of a Node: Flat MPI

	SR11K/J2	SR16K/M1	T2K	FX10	京
Core #/Node	16	32	16	16	8
Peak Performance (GFLOPS)	147.2	980.5	147.2	236.5	128.0
STREAM Triad (GB/s)	101.0	264.2	20.0	64.7	43.3
B/F	0.686	**0.269**	0.136	**0.274**	0.338
GeoFEM (GFLOPS)	19.0	72.7	4.69	16.0	11.0
% to Peak	12.9	**7.41**	3.18	**6.77**	8.59
LLC/core (MB)	18.0	**4.00**	2.00	**0.75**	0.75

B/FではFX10がやや上回るが
キャッシュサイズも影響か?

- 背景
- FX10スーパーコンピュータシステム概要
- **スケジュール**
- 運用・サービス
  - トークン制
  - 企業利用
  - トライアルユース
  - 教育利用，若手利用
  - 大規模HPCチャレンジ
- 試験運転期間のサービス
- 将来展望
- 質疑

スケジュール

- 試験運転
  - 4月2日（月）10:00～6月29日（金）09:00
  - 負担金無料
  - 試験運転期間中は，システムの設定変更等のため，予告なく運用の停止，運用仕様の変更を行う場合がありますので，予めご了承ください。
- 正式運用
  - 7月2日（月）09:00開始
- 新規利用申込み
  - 受付中
  - 申込み状況によっては、新規利用申込みを打ち切ることがあります。
  - 詳しくは、HP(http://www.cc.u-tokyo.ac.jp/)でご確認ください。

情報・問い合わせ

- 全般
  - http://www.cc.u-tokyo.ac.jp/system/fx10/
- 試験運転開始のお知らせ
  - http://www.cc.u-tokyo.ac.jp/system/fx10/fx10_test.html
- 利用コース
  - http://www.cc.u-tokyo.ac.jp/system/fx10/fx10_course.html
- ジョブクラス
  - http://www.cc.u-tokyo.ac.jp/system/fx10/fx10_job.html
- 利用者支援（利用申込・利用負担金、刊行物、成果登録）
  - http://www.cc.u-tokyo.ac.jp/support/
- FAQ
  - http://www.cc.u-tokyo.ac.jp/support/faq/fx10_faq.html
- 東京大学情報基盤センター研究支援係
  - 電話（平日09-12，13-17） 03-5841-2717
  - uketsuke(at)cc.u-tokyo.ac.jp

- 背景
- FX10スーパーコンピュータシステム概要
- スケジュール
- **運用・サービス**
  - **トークン制**
  - 教育利用，若手利用
  - 企業利用
  - トライアルユース
  - 大規模HPCチャレンジ
- 試験運転期間のサービス
- 将来展望
- 質疑

利用コース

- FX10スーパーコンピュータシステムの利用コース
  - パーソナルコース：研究者個人単位（大学・公共機関）
  - グループコース：研究グループ単位（大学・公共機関，企業）
  - HA8000の「パーソナルコース」「専用キュー」に対応
    - 「ノード固定」は無し：全系運転を可能とするため
  - よりフレキシブルな利用ができるようになっている
- 利用するコース (パーソナル・グループコース)，利用申込したノード数に応じて，計算ノードの利用可能時間である「トークン」を割当てます。
- 割り当てられたトークン内であれば (一部のコース、サービスを除き) 利用できるノード数制限はなく、最大利用可能ノード数まで，バッチジョブの実行を可能。
  - HA8000では申込みノードまでしか実行できなかった，また同時実行ジョブの合計ノード数が申込みノード数以下でなければならなかった。FX10ではそのような制限は無くなる。

トークン（token）

- トークンは，バッチジョブ（インタラクティブノードを利用したバッチジョブを除く）実行ごとに消費。
  - ノード時間積「経過時間 ×ノード数 × 消費係数」により消費
  - 消費係数は、申込ノード数までは 1.00
    - 申込ノード数を超えた部分については，2.00
- トークンを使い果たすとジョブ実行不可となる
  - 1日以下の単位でモニター，トークン不足の場合もジョブはsubmitできない
  - 計算資源に余裕がある場合にのみ，トークンを追加することが可能（発行トークン量：ノード時間数×1.25目安）
- トークンは，利用を許可された有効期間内に全量が利用できることを保証するものではありません。
- 利用を許可された期間のみを有効期間としているため，次年度への繰り越しや返金等は不可。

トークン制の利点・欠点

- 基本的にはHA8000の「専用キュー」の考え方と類似しているが，よりフレキシブルな運用が可能，需要があるときに集中的に計算資源を利用できる
  - 同時実行数，同時投入数などの制限はあります
- 使いすぎに注意

- できるだけ長期間（1年単位，本年度は9ヶ月）の契約がお得，短い期間でのトークン追加は割高です。

パーソナルコース

- パーソナルコースは研究者が個人単位でお使いいただくためのコースで，FX10 スーパーコンピュータシステムで利用できる最大ノード数により以下の 2 コースを用意しています。

コース	利用負担金 (年額，税込) 大学・公共機関等	利用可能 ノード数	トークン	ディスク量 /home
パーソナル コース1	120,000円/年	最大 24ノード	25,920 ノード時間 消費係数 : 12 ノードまでは 1.00 12 ノード超過は 2.00	200GB
パーソナル コース2	250,000円/年	最大 96ノード	51,840 ノード時間 消費係数 : 24 ノードまでは 1.00 24 ノード超過は 2.00	

グループコース

- グループコースは，研究グループなどで利用されるためのコースで，12 ノード単位で利用申込が行えます
  - 提供できる資源量に限りがあるため、利用申込単位 (ノード数) によってはご希望に添えない場合があります。
- 標準で割り当てられるディスク容量は12 ノードあたり4TB
  - それ以外にもグループに所属する利用者ごとに 50 GB のディスク容量が割り当てられます。
- グループに登録できる利用者数は無制限
  - 割り当てられたトークンをグループに登録された利用者で共有
- グループ管理者（利用申込み時に設定）にはグループ内の利用者ごとに割り当てるトークン量を変更できる仕組みを導入。

グループコース負担金（年）
12ノード当りの金額です

利用負担金 (年額，税込) （12ノード当り）		利用可能 ノード数	トークン	ディスク量 /group /home
大学・公共機関等	企業			
500,000円/年	1,400,000円/年	最大 1,440ノード	103,680 ノード時間 消費係数 : 申込ノードまでは 1.00 申込ノード超過は 2.00	グループ: 12ノード当り 4TB グループ各 ユーザー: 50GB／人

グループコースのトークンについて

- 申込み期間は原則として1年（本年度は9ヶ月）
  - 一ヶ月単位で申込みは可能
  - 年度はまたげない
    - 本年度から「年度の途中でやめる」ことが可能となった
- 「12ノード・1年」で申し込むとは?
  - 360日×24時間×12ノード＝103,680ノード時間
  - 次年度に余ったトークンを持ち越すことはできない
- 「12ノード・1ヶ月」で申し込むとは
  - 30日×24時間×12ノード＝8,640ノード時間
  - 「延長」はできない（資源が空いていたら新規申込可）
    - 余ったトークンを持ち越すことはできない
  - 360÷12よりはだいぶ割高

システム利用のイメージ

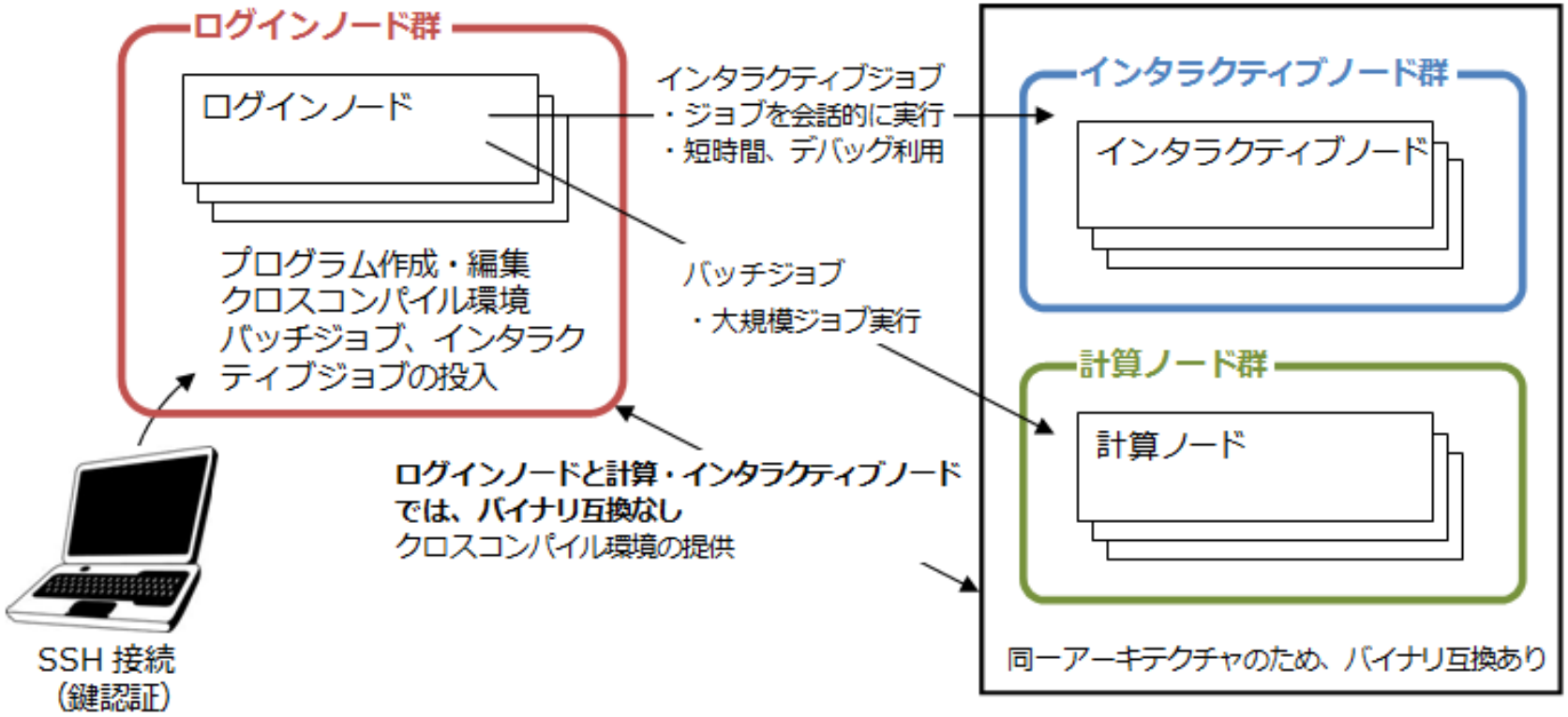


図２．システムのご利用イメージ

ジョブ実行

- インタラクティヴジョブ
- バッチジョブ

- 1-ユーザー，1-IDとする
  - 複数のプロジェクトに属している場合（パーソナルコースを含む）も統一のIDを使用。
  - バッチジョブ実行時に「財布（プロジェクトコード）」を指定する必要がある

ログイン

- プログラムの編集，コンパイルやバッチジョブの投入などに利用するための環境として，ログインノードを 6 台用意しています。
  - oakleaf-fx-n.cc.u-tokyo.ac.jp
    - n= 1～6
    - oakleaf-fx.cc.u-tokyo.ac.jp でいずれかに振り分けられる
  - ログインノードは計算ノードとは異なるアーキテクチャのため，計算ノード (インタラクティブノード) 上で実行可能なバイナリを生成するためのクロスコンパイル環境等を用意しています。
  - ログインノードは本センターでサービスを行っている他のスーパーコンピューターシステムと同様に，公開鍵認証方式による接続となります。

インタラクティヴジョブ

- 小規模，短時間のテスト用ジョブは，インタラクティヴノード(50ノード，計算ノードと同じアーキテクチャ)によって実行
  - ログインノードからインタラクティヴに実行
    - 実際は上記50ノードを使ったバッチジョブが動く
    - インタラクティヴノードが全て使われていると「待ち」状態になる
    - ジョブ投入時に、インタラクティヴジョブであることをオプションで指定(--interact)
  - HA8000のdebugキューに近い
  - インタラクティヴ実行によってトークンは消費されない(トークンが無くても実行可能)

代表キュー名	キュー名	最大ノード数	実行制限時間（経過時間）	ノード当りメモリー容量(GB)
interactive	interactive_n1 interactive_n8	1 8	2 h 10 min	28 28

バッチジョブ

- 長時間複数ノード利用ジョブは，バッチジョブとして実行
  - HA8000 クラスタシステム，SMP等で利用しているジョブ実行スクリプトとは互換性は無い
- Queueは全ユーザーで共通（「教育」ユーザー除く）
- 以下は本運用（2012年7月以降）の設定
  - 試験運転期間中は短め（後述）

代表キュー名	キュー名	最大ノード数	実行制限時間（経過時間）	ノード当りメモリー容量 (GB)
debug	debug	1-240	30 min	28
short	short	1-72	6 h	28
regular	small medium large x-large	12-216 217-372 373-480 481-1,440	48 h 48 h 48 h 24 h	28 28 28 28

バッチジョブ実行方法

- 利用者は、「トークン量」がある限り，好きなキューにジョブを投入することができる
  - 最大利用ノード数
    - パーソナルコース1： 24ノード
    - パーソナルコース2： 96ノード
    - グループコース： 1,440ノード
  - 申込ノード数を超過したジョブについては、超過部分について消費係数が高め（=2.00）に設定されます。
- 「regular」を指定してジョブ投入すると，ノード数によって small，medium，large に振り分けられる

バッチジョブ実行の詳細

- 最大受付本数制限（グループ全体での数）
  - 同時実行可能数の4倍
- 最大同時実行本数制限（パーソナルコース）
  - コース1：2本
  - コース2：4本
- 最大同時実行本数制限（グループコース）
  - 96ノードまでのグループ：
    グループあたり4本
  - 以下24ノード増えるごとに1本増加
  - 例：192ノード：8本、180ノード：7本、
    144ノード：6本、156ノード：6本

グループ管理者機能

- グループコース申込者のうち、それぞれのグループコースのユーザを管理する管理者（グループ管理者）を登録できる
- グループ管理者は、登録ユーザについて、コマンドおよびポータルを通して、以下の制限をかけることができる（ポータルは準備中）
  - トークン量の上限値
  - 最大利用ノード数
- 各ユーザは、コマンドを通して、自分に設定されている制限情報を閲覧できる
  - 利用可能トークン量、消費したトークン量
  - 実行可能キュー名とその状態
  - 同時実行可能数、同時受付可能数

- 背景
- FX10スーパーコンピュータシステム概要
- スケジュール
- **運用・サービス**
  - トークン制
  - **教育利用，若手利用**
  - **企業利用**
  - **トライアルユース**
  - **大規模HPCチャレンジ**
- 試験運転期間のサービス
- 将来展望
- 質疑

教育利用

- http://www.cc.u-tokyo.ac.jp/service/education/
- 大学院や学部の授業・演習にスーパーコンピューター資源を提供
  - 大学，高等専門学校教員が担当する大学院，大学学部，高等専門学校における講義・演習（集中講義・講習会を含む）を対象とします
  - 申込を随時受け付けています。東大以外の大学からも申し込めます（東大以外の大学・大学院・高専の講義で利用できます）
- 2011年度まではHA8000で実施，今後はFX10へ移行
  - 2012年7月以降，募集開始予定
  - 講義時間中：12ノード優先利用，学期中12ノードまで利用可
  - 無料
  - 成果報告
    - 広報誌への記事執筆，報告会（センター主催ワークショップ等）

若手利用

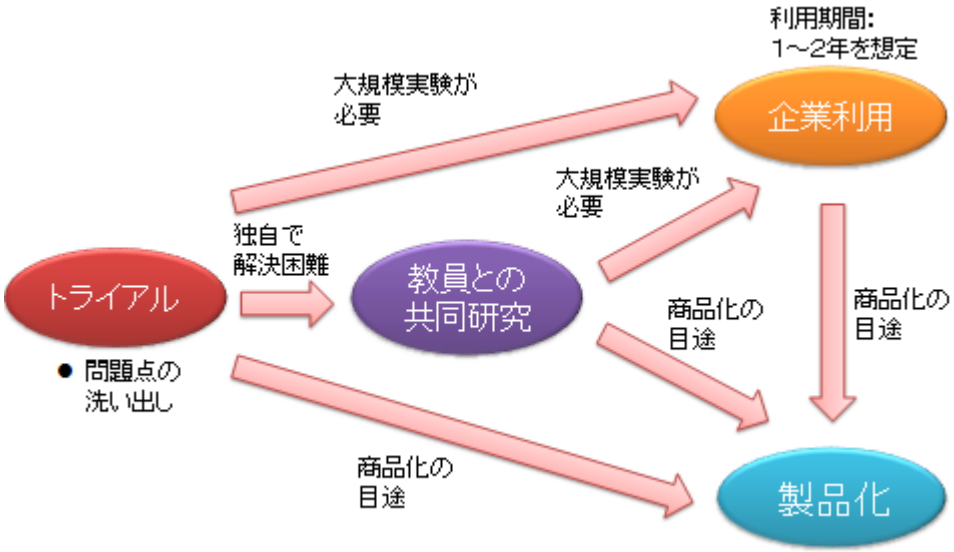
- http://www.cc.u-tokyo.ac.jp/service/wakate/
- 概ね45歳以下の若手研究者（学生を含む）を対象とした利用者向け推薦制度による課題を公募している。
  - スーパーコンピューティング部門の教員により審査の上，採択された課題の計算機利用負担金をセンターが負担
  - 年2回公募，年間4件程度の優れた研究提案を採択する予定
    - 期間は1回半年，継続申請・再審査により最大1年間の無料利用可能
- 成果報告
  - 広報誌への記事執筆，報告会（センター主催ワークショップ等）
- 本制度に採択された課題は終了後，得られた成果をもとに，科研費，学際大規模情報基盤共同利用共同研究拠点（8センター）公募型研究への進展が期待される
- 2011年度までHA8000利用，今後はFX10へ移行
  - 2012年7月以降募集開始
  - 内容について若干の変更を検討中（共同研究ベースのものを検討中）

企業利用

- http://www.cc.u-tokyo.ac.jp/service/company/
- 社会貢献の一環として，企業に対して有償でFX10の一部（全資源の10%以内）を提供
  - 2011年度まではHA8000で実施していたが，2012年度よりFX10へ移行
  - 単なる企業の計算需要の肩代りを行うのではなく，大規模高性能並列計算普及のための支援を主眼とする。
  - 「成果公開利用」のみ。
- 詳細は「東京大学情報基盤センタースーパーコンピューター企業利用審査要項」を参照
  - http://www.cc.u-tokyo.ac.jp/service/company/shinsayoko-company.pdf

企業利用(分類)

- 企業利用
  - 通常有償利用
  - トライアルユース(グループ，パーソナル)(後述)
  - 共同研究型
    - 「共同研究型」は，共同研究協約をセンターと結んで実施，負担金は大学・公共機関並



- 通常有償利用，トライアルユース(グループ)は審査あり
  - 年4回募集
  - 次回は第3期募集
  - 8月中旬締切、9月上旬審査、10月1日利用開始、(いずれも予定)

企業利用（通常有償利用：割引制度）

- 以下の条件を満たす場合，利用開始年度の最初の3ヶ月分については大学・公共機関並負担金を適用する
  - トライアルユース（後述）を利用していない
  - 年度末まで利用契約している
  - 割引きは利用開始年度のみ有効
- 通常有償利用において、採択企業名、採択課題名は公開される
- 利用終了後、利用報告書が公開される。

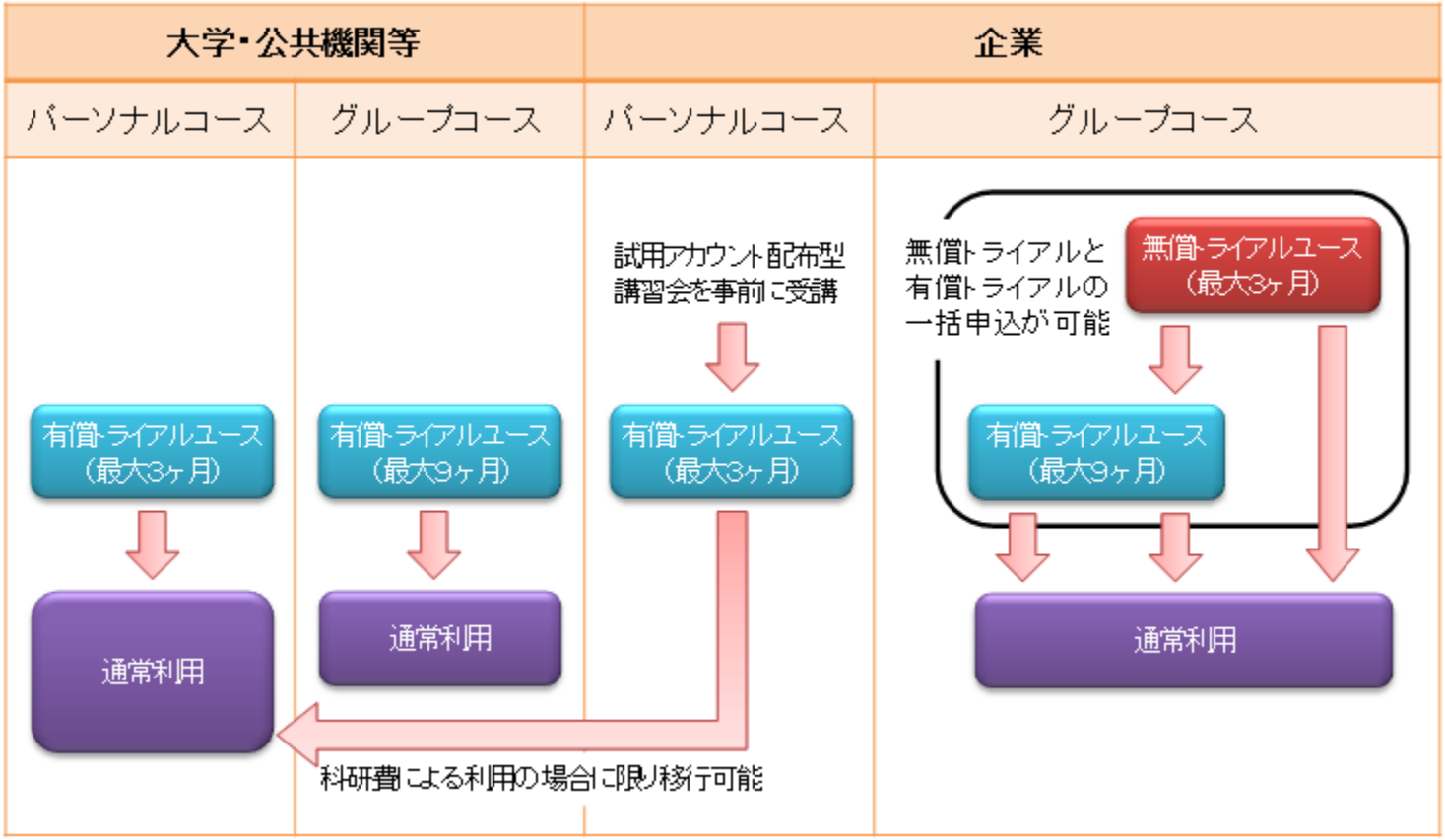
トライアルユース（全般）

- http://www.cc.u-tokyo.ac.jp/service/trial/fx10.html
- 連続した一定期間（1ヶ月～12ヶ月），通常の負担金とは異なる体系の料金で試験的利用（トライアルユース）可能
  - 大学・公共機関等，企業に対してそれぞれパーソナル，グループコースがある（全部で4種類）
- グループコース（企業）「以外」は，HA8000に準ずる。
  - 上限はパーソナル：3ヶ月，グループ（大学・公共機関等）：9ヶ月
  - パーソナルコース（企業）利用者は試用アカウント配布型講習会受講者に限定，HA8000向け講習会の受講者も（HA8000のトライアルユースを利用していない場合は），FX10を利用できる
  - コースに応じたトークンを割り当てる（グループは12ノード×月相当），トークンの追加はできない

トライアルユース（グループコース（企業））

- 無償トライアルユース（3ヶ月），有償トライアルユース（通算最大9ヶ月（1ヶ月単位），大学・公共機関等料金を適用）を設ける
  - 12ノード×月相当のトークンを割り当てる
    - 追加はできない
  - トライアルユース開始に当たっては利用資格審査が必要となる
  - トライアルユースから通常有償利用に移行する場合は改めて利用資格審査が必要となる。
- トライアルユース期間は年度にまたがってもよいが，同一年度内継続の場合と比較して利用負担金は高くなる。
- トライアルユースでトークンが余った場合，通常有償利用に持ち越すことはできない
- 採択企業名、課題名は公開される

トライアルユースの仕組み



大規模HPCチャレンジ

- http://www.cc.u-tokyo.ac.jp/service/4800hpc/
- 月1回1日（24時間），4,800ノード（全計算ノード）を1グループで占有して実行できる，公募制，無料。
  - 実施中は一般ユーザーはログインノード，インタラクティヴジョブのみ利用できる。
- FX10ユーザー以外も応募可能である。
- 成果公開を義務づける
  - センター広報誌への寄稿
  - センター主催各種催しでの発表，各種外部発表への情報提供
  - 速報結果の査読付国際会議への投稿等による迅速，国際的な成果公開が望ましい。
- 企業からの申し込みも受け付ける（成果公開を義務づけ）
- 自作プログラム，オープンソースプログラム利用に限定
- 次期公募締切：2012年6月18日 (月) 10:00
- 今後の実施予定については、HPを参照ください。

大規模HPCチャレンジ（予定）

実施時期	募集締切	審査	採択通知
2012年07月26日 (木) 9:00 ～ 27日 (金) 9:00 2012年08月30日 (木) 9:00 ～ 31日 (金) 9:00 2012年09月27日 (木) 9:00 ～ 28日 (金) 9:00 2012年10月25日 (木) 9:00 ～ 26日 (金) 9:00 2012年11月29日 (木) 9:00 ～ 30日 (金) 9:00	2012年 6月18日（月） 10:00【締切】	2012年 7月上旬	2012年 7月中旬
2012年12月27日 (木) 9:00 ～ 28日 (金) 9:00 2013年01月24日 (木) 9:00 ～ 25日 (金) 9:00 2013年02月21日 (木) 9:00 ～ 22日 (金) 9:00 2013年03月29日 (木) 9:00 ～ 30日 (金) 9:00	2012年 11月19日（月） 10:00【締切】	2012年 12月上旬	2012年 12月中旬

- 背景
- FX10スーパーコンピュータシステム概要
- スケジュール
- 運用・サービス
  - トークン制
  - 教育利用，若手利用
  - 企業利用
  - トライアルユース
  - 大規模HPCチャレンジ
- **試験運転期間のサービス**
- 将来展望
- 質疑

試験運転期間中のサービス・お願い

- 基本的に7月以降利用するコース，ノード数を申し込んでいただくようお願いします。トークンの途中追加は可能。
- 試験運転期間中の残余トークンは本運用に持ち越せない
- 試験運転期間中は「トライアルユース」は設けない
  - 4月～6月は企業利用の「無償トライアルユース」期間に含めない
- 試験運転期間中の企業利用
  - 7月から「通常有償利用」を予定する場合は，その利用ノード数，「トライアルユース」の場合は12ノードに対応したトークンがそれぞれ与えられる。
- ジョブの実行時間は本運用時より短くする

試験運転期間中のバッチジョブ実行制限時間の設定

代表キュー名	キュー名	最大ノード数	実行制限時間（経過時間）	ノード当りメモリー容量 (GB)
debug	debug	1-240	30 min	28
short	short	1-72	2 h	28
regular	small medium large x-large	12-216 217-372 373-480 481-1,440	12 h 12 h 12 h 6 h	28 28 28 28

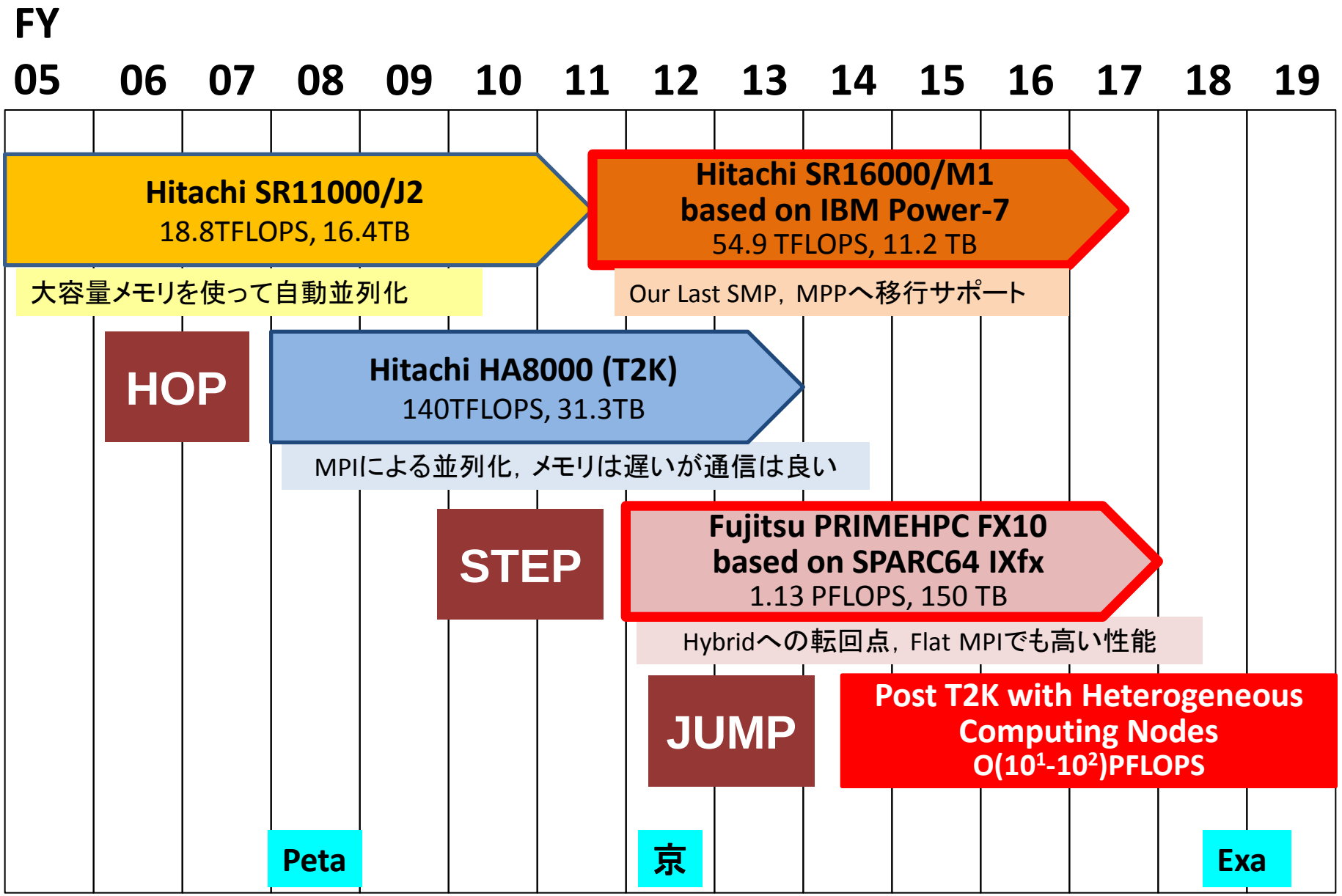
- 背景
- FX10スーパーコンピュータシステム概要
- スケジュール
- 運用・サービス
  - トークン制
  - 教育利用，若手利用
  - 企業利用
  - トライアルユース
  - 大規模HPCチャレンジ
- 試験運転期間のサービス
- **将来展望**
- 質疑

FX10に関する様々な情報

- 随時，HP，広報誌（スーパーコンピューティングニュース）で最新情報を提供
  - チューニング関連
- MPI基礎に関する講習会を、7月2日、3日に柏で開催。
- 講習会の開催予定は、以下のHPでご確認ください。

http://www.cc.u-tokyo.ac.jp/support/kosyu/schedule_kosyu.html

東大情報基盤センターのスパコン



ポストT2Kシステムへの展望

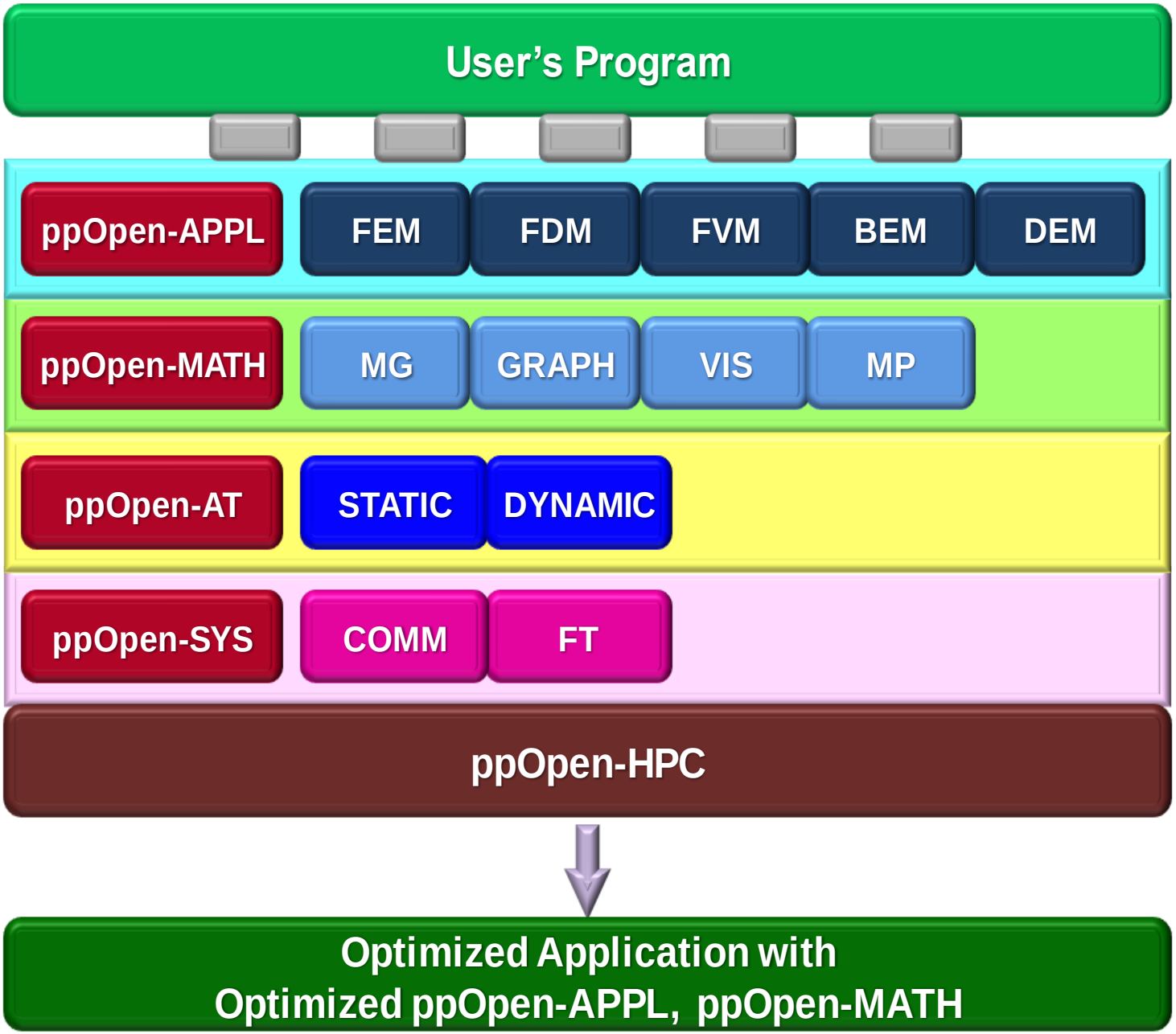
- 限られた電力消費量において，高い計算性能を有し，メモリ性能とのバランスのとれたシステムを目指すためには，（当面は）ヘテロジニアスな計算ノードを有するシステムが有力
- マルチコアCPU+GPGPU，マルチコアCPU+メニーコア（e.g. Intel MIC）
  - TSUBAME 2.0（東工大）
  - 筑波大，京大の新システム
- プログラミングの困難さ
  - MPI+OpenMPですら結構大変
    - 陽解法はそれでもまだ簡単
  - CUDA，OpenCL，OpenACC

ppOpen-HPC（1/2）

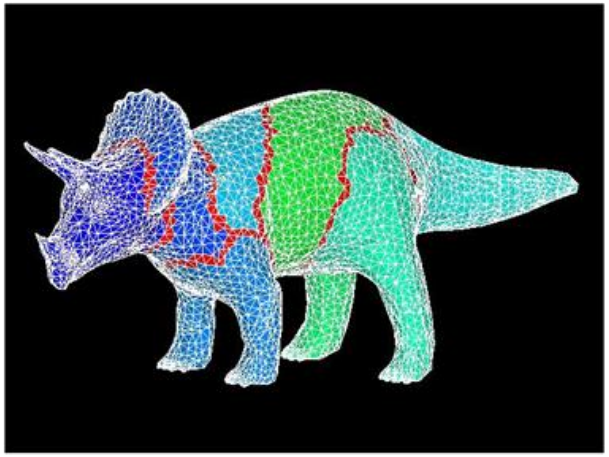
- 東京大学情報基盤センターでは，ヘテロジニアスなアーキテクチャによる計算ノードを有するポストペタスケールシステムの処理能力を充分に引き出す科学技術アプリケーションの効率的な開発，安定な実行に資する「自動チューニング機構を有するアプリケーション開発・実行環境：ppOpen-HPC」を開発している。
  - 科学技術振興機構戦略的創造研究推進事業（CREST）研究領域「ポストペタスケール高性能計算に資するシステムソフトウェア技術の創出」（2011～2015年度）
  - 東大（大気海洋研究所，地震研究所，人工物工学研究センター），京大学術情報メディアセンター，海洋研究開発機構

ppOpen-HPC（2/2）

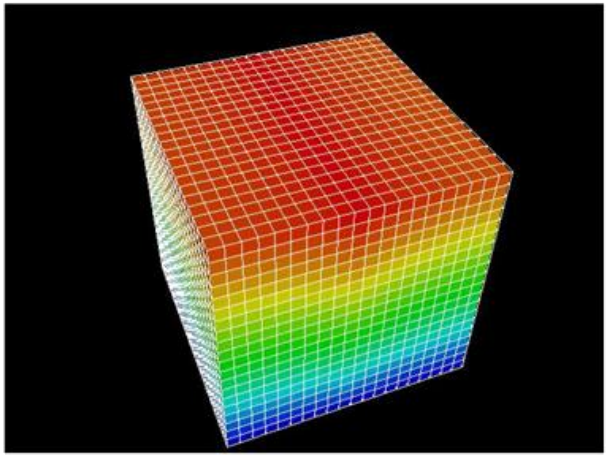
- 対象離散化手法を有限要素法，差分法，有限体積法，境界要素法，個別要素法に限定し，各手法の特性に基づきハードウェアに依存しない共通インタフェースを有するアプリケーション開発用ライブラリ群，ノード間通信ライブラリ，耐故障機能を含む実行環境を提供する。
- 自動チューニング技術の導入により，様々な環境下における最適化ライブラリ，耐故障機能を持つ最適化アプリケーションの自動生成を目指す。
- 2014年度に東京大学情報基盤センターに導入予定のポストT2Kシステムをターゲットとし，同システム上で実アプリケーションによって検証，改良し，一般に公開する。
  - 本年9月にT2K，FX10向けのプロトタイプを公開予定
  - ポストT2Kシステムへのスムースな移行を目指す


User's Program
ppOpen-APPL
FEM
FDM
FVM
BEM
DEM
ppOpen-MATH
MG
GRAPH
VIS
MP
ppOpen-AT
STATIC
DYNAMIC
ppOpen-SYS
COMM
FT
ppOpen-HPC
Optimized Application with
Optimized ppOpen-APPL, ppOpen-MATH

ppOpen-HPC covers …

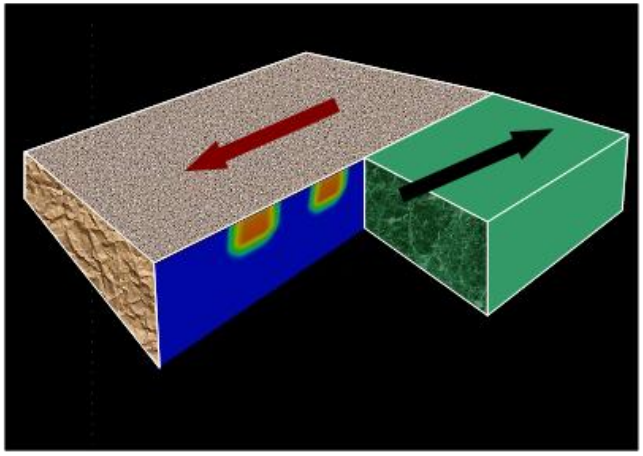


FEM
Finite Element Method

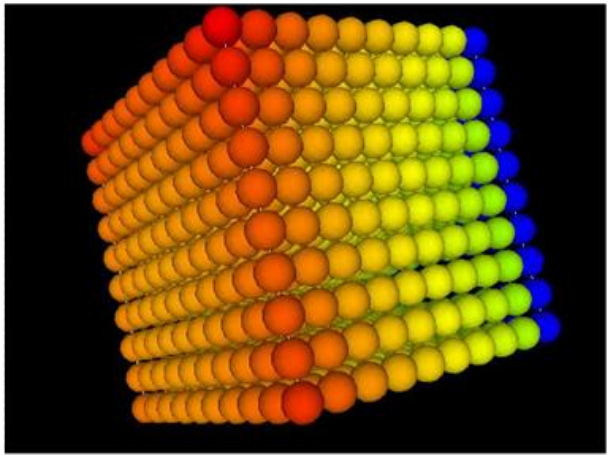


FDM
Finite Difference Method

FVM
Finite Volume Method



BEM
Boundary Element Method



DEM
Discrete Element Method