GREEN500への軌跡と展開

黒川　原佳※、鳥居 淳#

※理化学研究所 情報基盤センター ユニットリーダー
#株式会社ExaScalar CTO

アウトライン

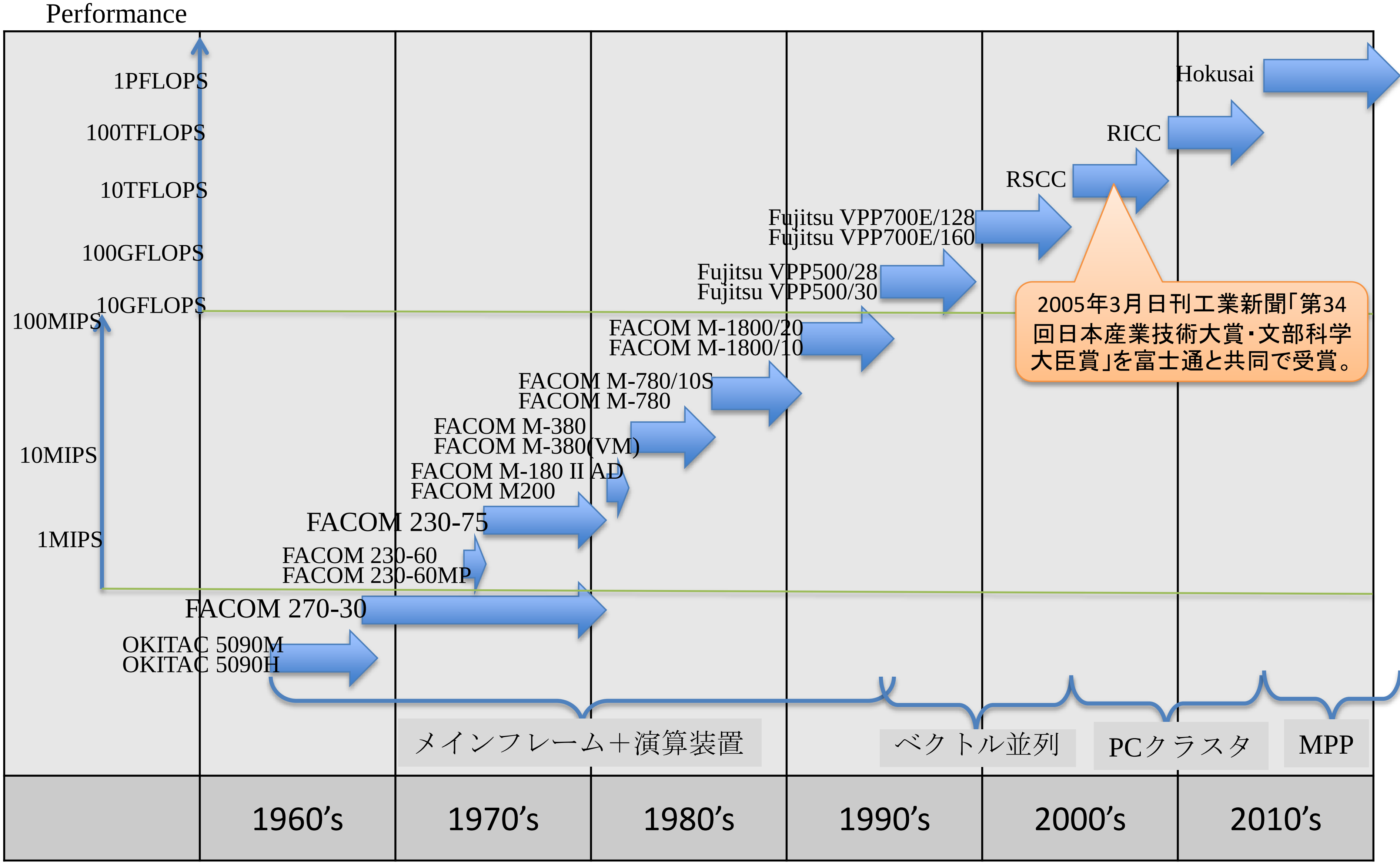
- 発表者や組織の概略
- Shoubuシステムについて
- エネルギー効率の動向
- Shoubuシステム設置とTOP1の顛末
- 今後の展開

私および情報基盤センターの紹介

- 経歴
  - 元々は流体実験やCFD（計算流体力学）の研究者だったが、2002年に理研に来たときコンピュータシステムに興味を持ち、コンピュータ・システム全体の設計・管理・運用などに興味が移行しました。
  - 京の開発プロジェクトも最初期から参加していました。
  - 今は以下の業務に従事

- 理研全体のIT基盤（スーパーコンピュータ、ネットワーク、サーバなど）の整備・運用・サポートおよび情報関連の研究開発を行っている部署。
- 今回は、スーパーコンピュータに特化した内容の紹介をします。
  - 現在運用中のスーパーコンピュータ・システム
    - HOKUSAI GreatWaveシステム
      - Fujitsu製PRIMEHPC FX-100（京の後継機）を中核としたシステム。
  - 共同研究のスーパーコンピュータ・システム
    - 2015年6月期にGREEN500のTOP1を獲得したShoubuシステム
    - 今回のプレゼンのメインテーマ

情報基盤センターでのスーパーコンピュータ導入の歴史



HOKUSAI GreatWaveシステム

- 2段階のシステム稼働を行うシステム
  - 現在は1段階目のシステムが稼働中。
    - 1段階目 : GreatWave (浪裏)
    - 2段階目 : BigWaterfall (瀑布)　2016年頃導入予定

ピーク性能 : 1 PFLOPS（5ラックで実現）

齊藤さん, PEZY Computing, ExaScalerの紹介

- 齊藤さん：元々は放射線科の医師、シリコンバレー他の医療画像処理系ベンチャー数社の創始者

- 株式会社PEZY Computing
  - 創業：2010年1月
  - 事業内容：プロセッサ、PCBボード、システム、ソフトウェアの開発と販売

- 株式会社ExaScaler
  - 創業：2014年4月
  - 事業内容：高効率の液浸冷却装置、及びシステムの開発・製造・販売

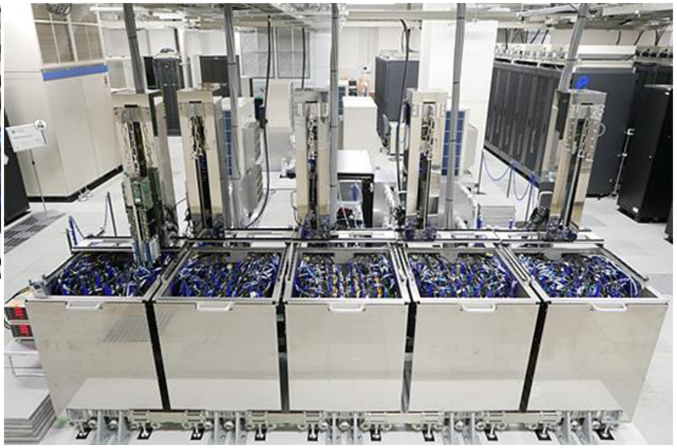
- Shoubuシステムは両者の技術を基盤として構築されています。

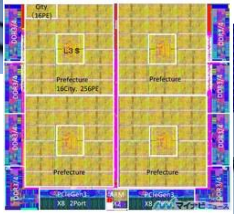
GREEN500 TOP1@Jun/2015 に認定されたシステム
現在最も省エネルギーなスーパーコンピュータ

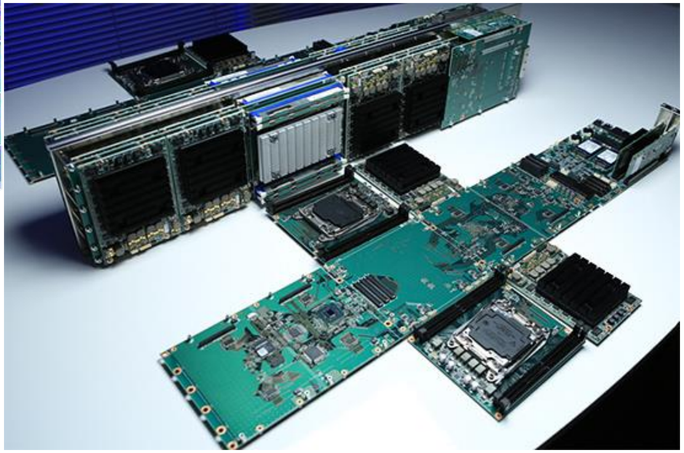
SHOUBUシステムについて

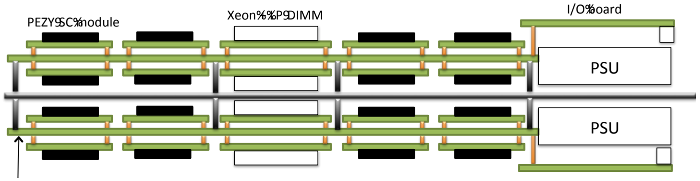
Shoubu（菖蒲）システム

理研-PEZY Computing-ExaScalerの
共同研究のためのシステム









Shoubuシステム構成

- ピーク性能：1.5～2PFLOPS級
  - ZettaScaler※1.4ベース
  - 5液浸槽構成
- ノード数：320ノード
  - 80Brick（ブリック）、16ブリック/1液浸槽
- ノード構成：1CPU＋4PEZY-SCチップ
  - PEZY-SCチップ：1280個

※：ExaScalerあらためZettaScaler

ZettaScalerおよびPEZY-SCの特徴

- ZettaScaler
  - 液浸冷却システム
    - 空気の代わりに液体（フロリナート：フッ素系化合物）を使って冷却するシステム。
  - 液浸（使う液体・方式は色々）は現在世界的に研究開発がホットなもの。
    - ZettaScalerは、高効率冷却、高密度実装、設置場所の制約を低く、システム・ランニングのトータルの低コスト化を狙う。
- PEZY-SC
  - 独自開発低消費電力メニーコアプロセッサ
    - MIMD型で1,024コアを集積した演算専用LSIチップ
  - 倍精度ピークでは1.5TFLOPSを実現可能
    - チップ単体では25GFLOPS/Wに
  - 演算の低消費電力化に多大に貢献

菖蒲(Shoubu)の名前の由来

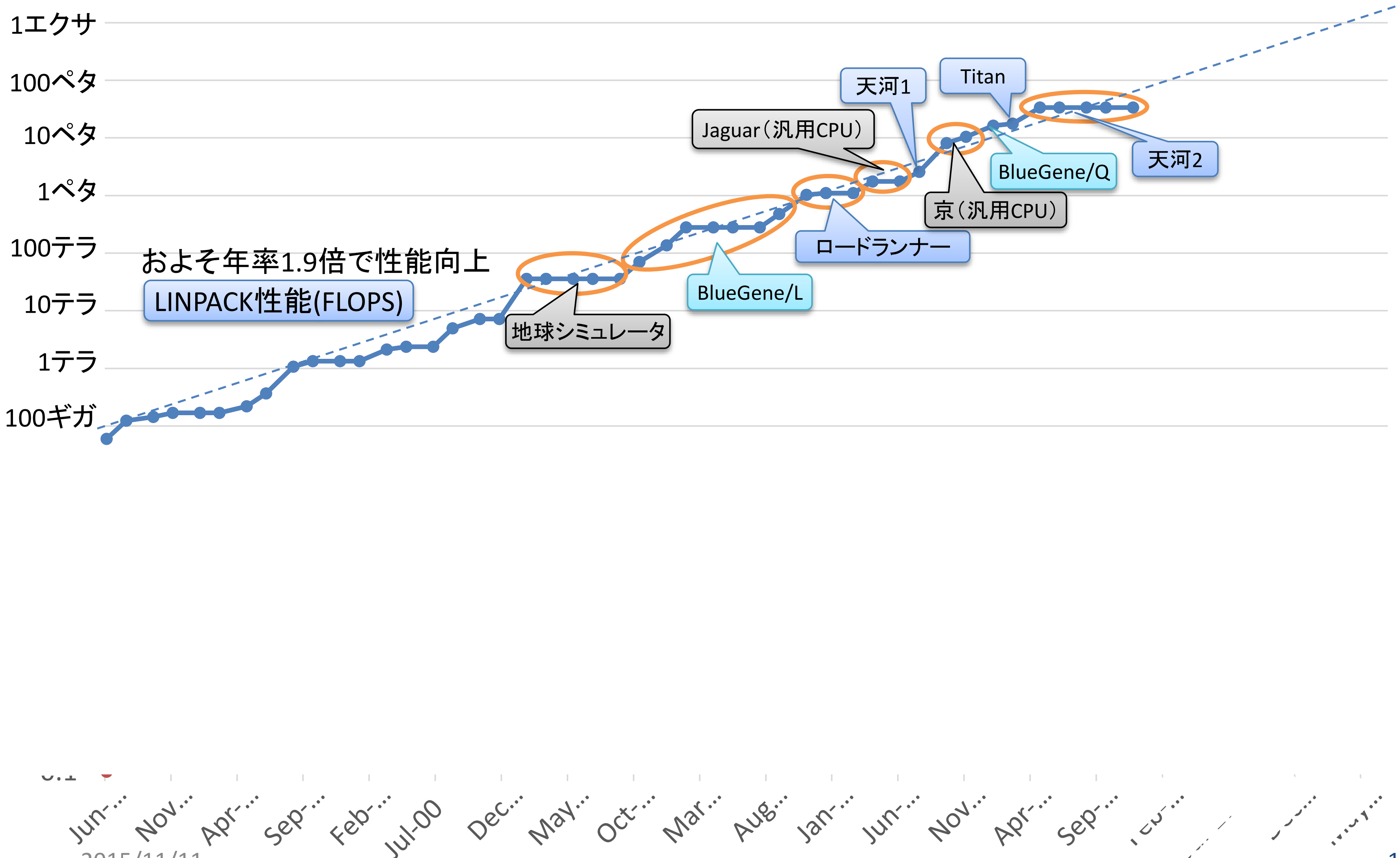
- 古来より縁起の良い植物。
- 水辺の植物。
- 5-6月に花を咲かせる。
- 葉が扁平で積層になっている。
- 「勝負」にも通じる。

Name Acorus calamus Family Acoraceae Original book source: Prof. Dr. Otto Wilhelm Thomé Flora von Deutschland, Österreich und der Schweiz 1885, Gera, Germany

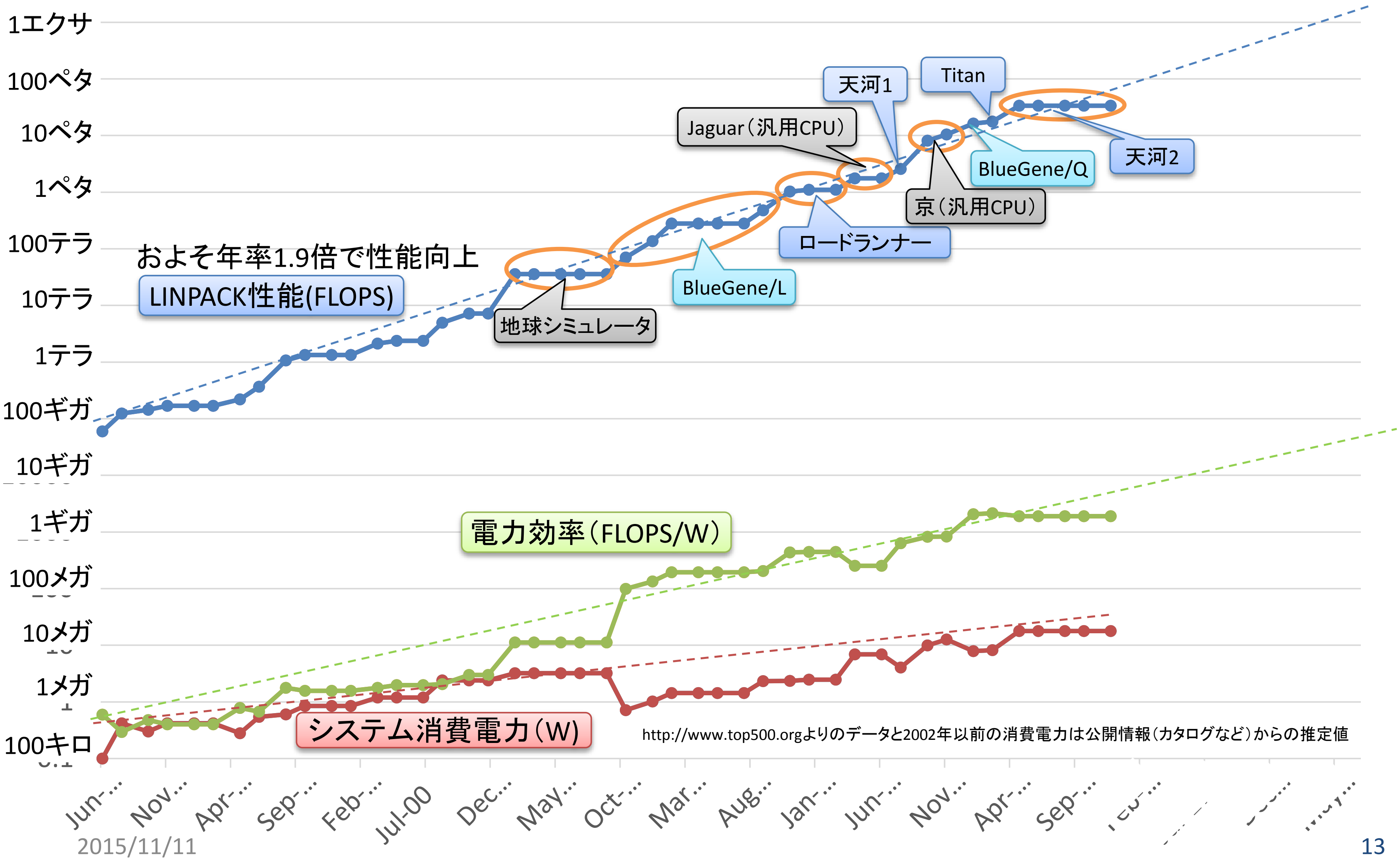
演算あたりのエネルギー効率の過去からの推移とこれから。
MFLOPS/W（1ワットあたりの演算回数）を評価する。

エネルギー効率の動向

TOP500リストのNo.1の性能変遷



TOP500リストのNo.1の数値変遷

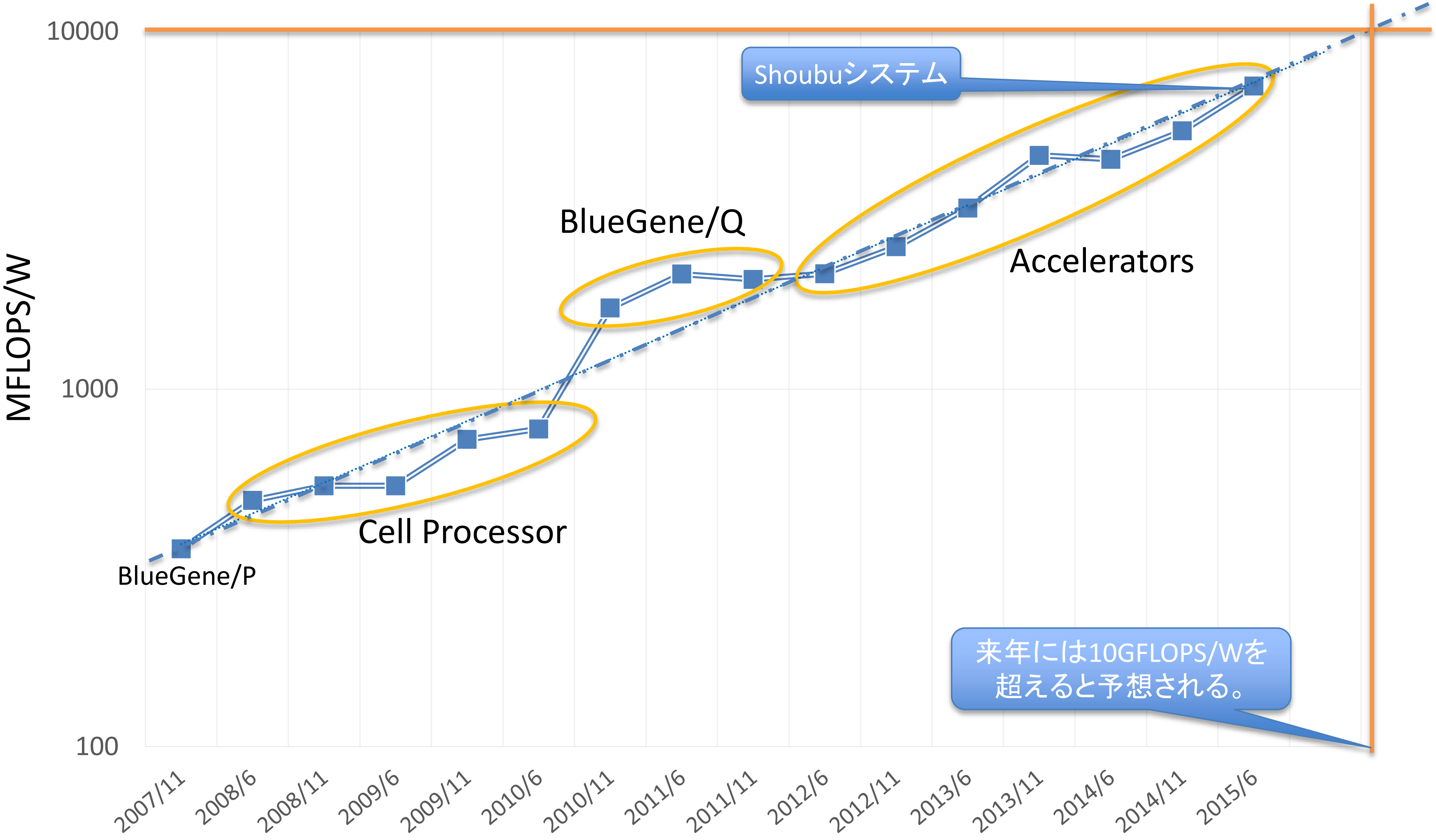


GREEN500リスト（2015年6月期）

Green500 Rank	MFLOPS/W	Site*	Computer*	Total Power (kW)	
1	7,031.58	RIKEN	Shoubu - ExaScaler-1.4 80Brick, Xeon E5-2618Lv3 8C 2.3GHz, Infiniband FDR, PEZY-SC	50.32	日本
2	6,842.31	High Energy Accelerator Research Organization /KEK	Suiren Blue - ExaScaler-1.4 16Brick, Xeon E5-2618Lv3 8C 2.3GHz, Infiniband, PEZY-SC	28.25	日本
3	6,217.04	High Energy Accelerator Research Organization /KEK	Suiren - ExaScaler 32U256SC Cluster, Intel Xeon E5-2660v2 10C 2.2GHz, Infiniband FDR, PEZY-SC	32.59	日本
4	5,271.81	GSI Helmholtz Center	ASUS ESC4000 FDR/G2S, Intel Xeon E5-2690v2 10C 3GHz, Infiniband FDR, AMD FirePro S9150	57.15	ドイツ
5	4,257.88	GSIC Center, Tokyo Institute of Technology	TSUBAME-KFC - LX 1U-4GPU/104Re-1G Cluster, Intel Xeon E5-2620v2 6C 2.100GHz, Infiniband FDR, NVIDIA K20x	39.83	日本
6	4,112.11	Stanford Research Computing Center	XStream - Cray CS-Storm, Intel Xeon E5-2680v2 10C 2.8GHz, Infiniband FDR, Nvidia K80	190.00	米国
7	3,962.73	Cray Inc.	Storm1 - Cray CS-Storm, Intel Xeon E5-2660v2 10C 2.2GHz, Infiniband FDR, Nvidia K40m	44.54	米国
8	3,631.70	Cambridge University	Wilkes - Dell T620 Cluster, Intel Xeon E5-2630v2 6C 2.600GHz, Infiniband FDR, NVIDIA K20	52.62	英国
9	3,614.71	TU Dresden, ZIH	Taurus GPUs - Bull bullx R400, Xeon E5-2680v3 12C 2.5GHz, Infiniband FDR, Nvidia K80	58.01	ドイツ
10	3,543.32	Financial Institution	iDataPlex DX360M4, Intel Xeon E5-2680v2 10C 2.800GHz, Infiniband, NVIDIA K20x	54.60	米国

GREEN500 Web site http://www.green500.org より

GREEN500 TOP1のMFLOPS/Wのトレンド



GREEN500 Websiteより http://www.green500.org

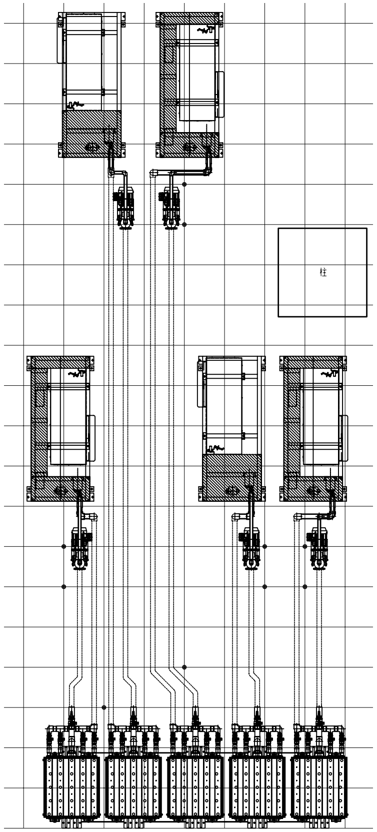
SHOUBUシステム設置とTOP1の顛末

はじめに

- 4月9日：電子情報通信学会集積回路研究会アクセラレーション技術討論発表会の場で姫野に話があった。
- システム設置も含めて一緒にやっていけるところを探している。というお話だったそうです。
  - できるかぎり、今回（2015/7）のGREEN500を狙いたい。とのこと。
  - 今年はISCの開催が1ヶ月遅れたため可能性はあるかな。。
- 4月下旬頃、PEZY+ExaScalerの具体的な話をしようという会議が設定されていた。
- 当日行ってみると、PEZY関係者、ExaScaler関係者、戎崎主任やAICSの牧野さんも座っている。
- 今のところ、場所と電気はあるけど、、冷却機への冷媒の管路が無い。
  - まあ、冷却能力として閉じるなら、冷却機を室内においたらいけるか。。
  - ただし、全開運転の場合、冷却能力としてギリギリの所のため、他のシステムを少し止めないと無理。
  - まあ、意気込みは買います。やってみますか。ということで始まる。
- 共同研究契約を2015年4月付けで行って、その範囲内で実施。
- 6月の時点では、冷却機は室内に設置していた。
  - 現在は冷却機は室外に設置しています。

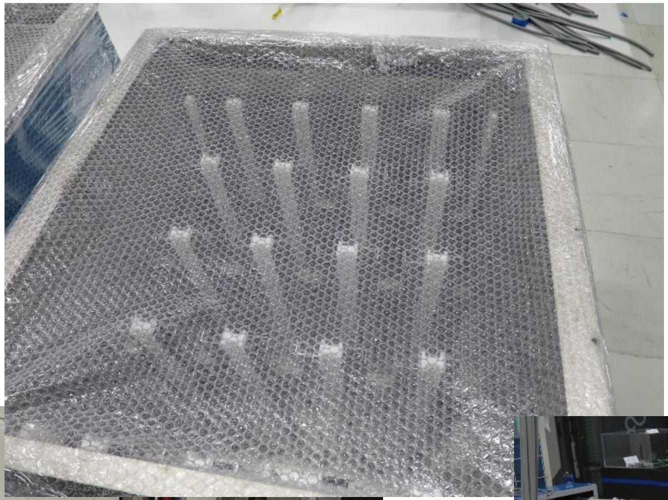
設置・計測のタイムライン(1/7)

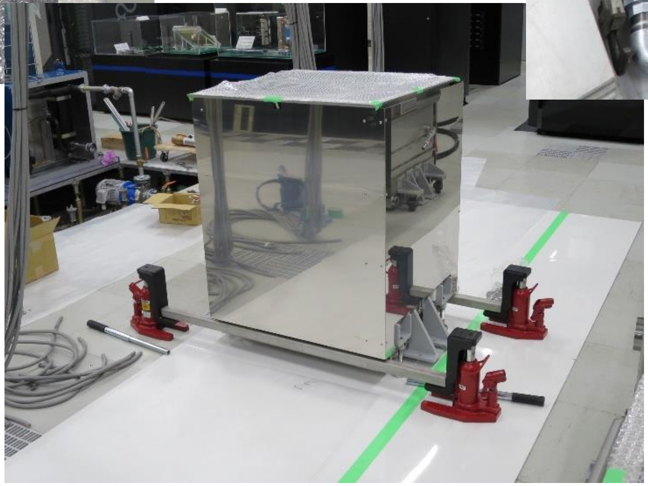
- 5月中旬(-50日):工事準備開始(墨だしなど)
- 5月下旬(-40日):冷却機を室内に設置(4台分)
- 6月1週目(-30～25日):電気工事とポンプ搬入



設置・計測のタイムライン(2/7)

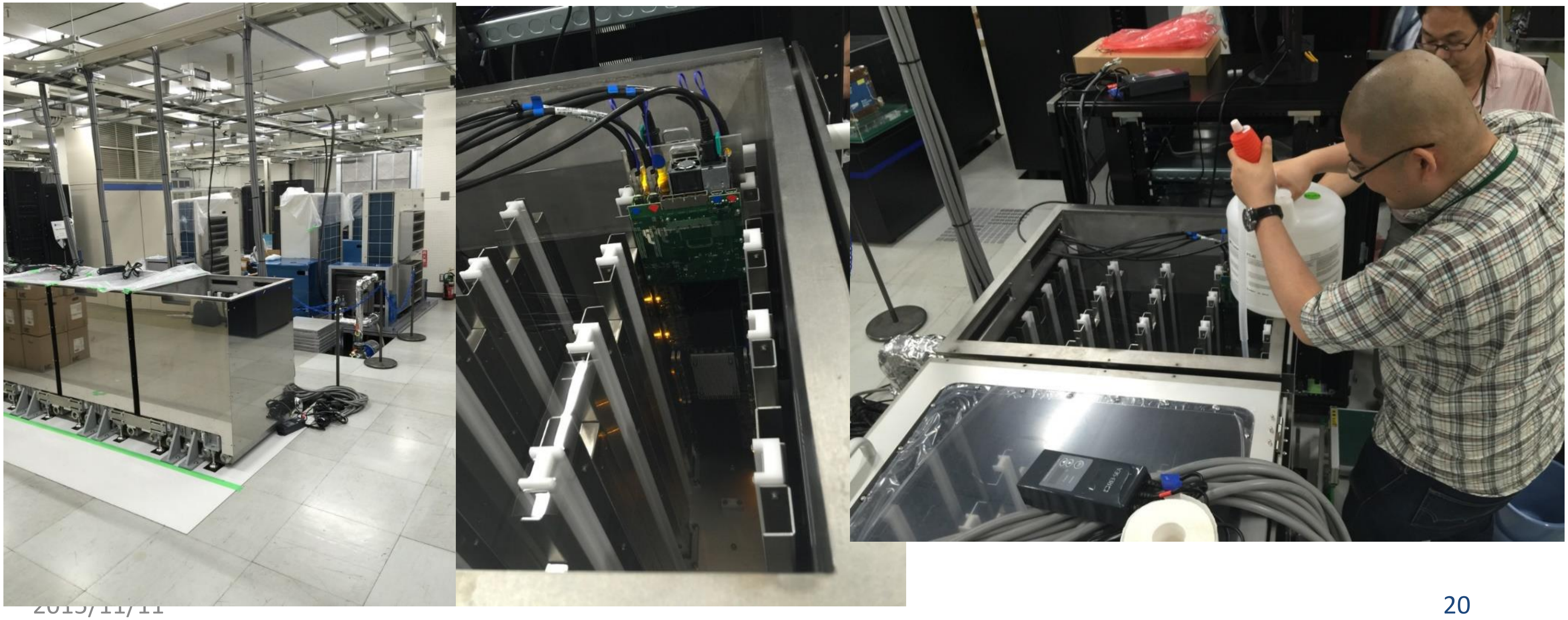
- 6月7日(-23日): 水槽5台搬入
  - 順次配管工事





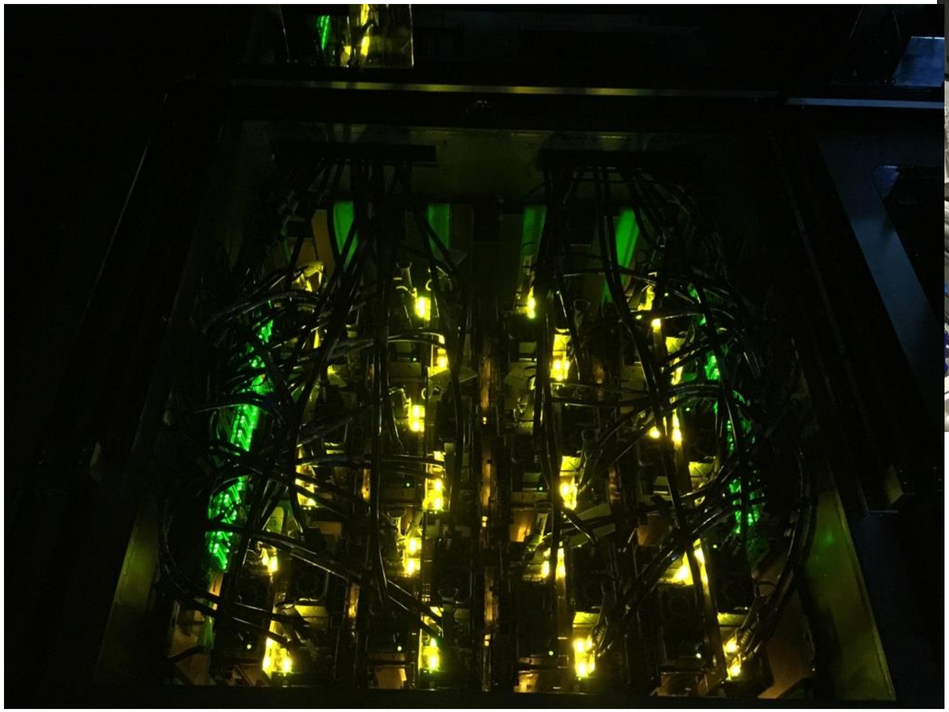
設置・計測のタイムライン(3/7)

- 6月11日（-19日）: 1水槽分の配管が終わり
  - フロリナート注入
- 6月15日（-15日）: 1Brick動き始めた。
  - Brickは理研内で組み立てられてました。。



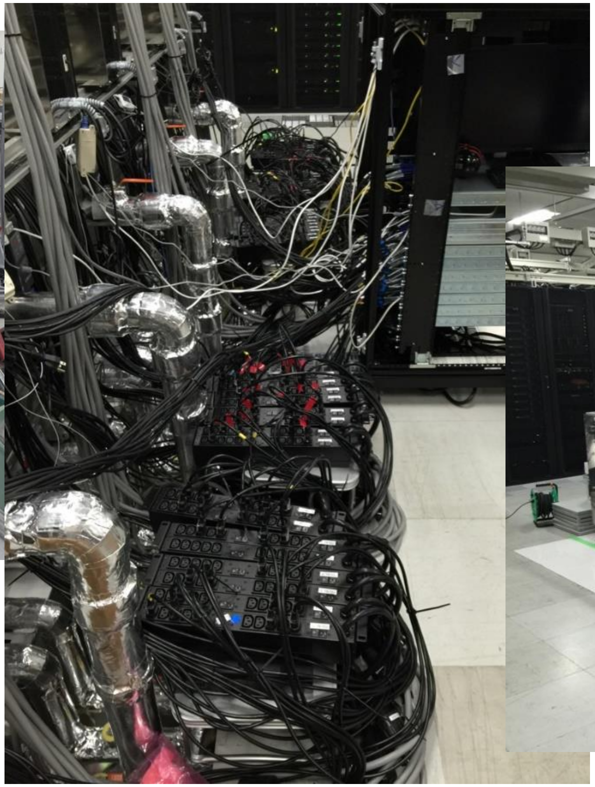
設置・計測のタイムライン(4/7)

- 6月18日（-12日）: 1水槽分動き始めて、HPL完走。
  - 113TFLOPS



設置・計測のタイムライン(5/7)

- 6月20日（-10日）: 5水槽分の配管完了。
  - 各水槽毎に調整を開始。



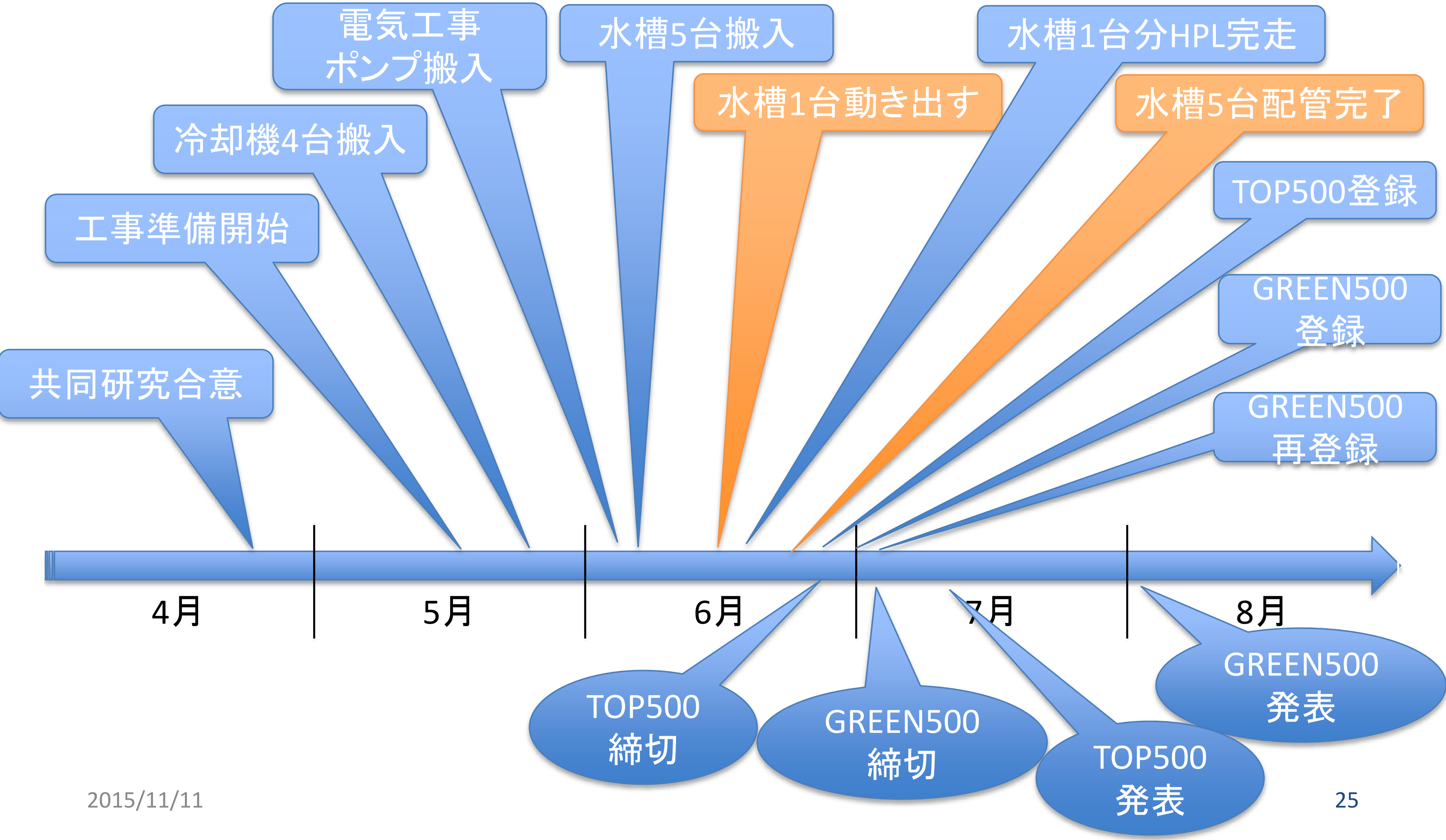
設置・計測のタイムライン(6/7)

- 6月27日（-3日） TOP500締切
  - 4水槽分320ノード（Rmax:413TFLOPS）で登録
- -月--日（-日） GREEN500締切
  - この時点では6.834GFLOPS/W(84ノード）でした。。
    - でも、登録Webが締め切られないので、、、

設置・計測のタイムライン(7/7)

- -月--日(-日) GREEN500再計測
- 20% ルール上ながら、7GFLOPS/Wを記録！！
- 本来は7月12日からのISC2015中に発表されるはずだったが、、、
  - ISC2015では次期からのGREEN500のレギュレーション変更が発表
- 8月2日 GREEN500発表
  - 1～3位をZettaScalerが独占

タイムラインをまとめると



今後の展開

- 次回2015年11月のリストでは計測レギュレーションも変更されるが、同等以上の性能結果を出したいと考えている。
- PEZY-SCや液浸システムは、まだ開発が始まって間もない。
  - これらはまだまだ最初の一歩です。
  - 今後、研究開発を進めてスーパーコンピューティングについて特徴的な方向性が見いだせる可能性を秘めている。
- PEZY-SCチップ（アクセラレータ）
  - 応用アプリケーションへどのように展開するかが鍵。
    - 今までに無い方式のアクセラレータであるため、アプリケーションの向き不向きもまだよく分かっていない。
  - GREEN500で1位などで成果は得られているが、実アプリケーションを効率的に動作させるための技術開発やノウハウの蓄積が重要となる。
- 液浸技術
  - 実際に動かしていく中で、問題点や検討課題のあぶり出しを行い、研究開発を進めていく。
  - 液浸をベースとして高効率な冷却を行い、データセンター全体に対する冷却電力をいかに下げていけるかが重要となる。
- 今後も、色々な形の展開が見られると思います。こうご期待！！