第100回　お試しアカウント付き
並列プログラミング講習会

Reedbush スパコンを用いたGPU ディープラーニング入門

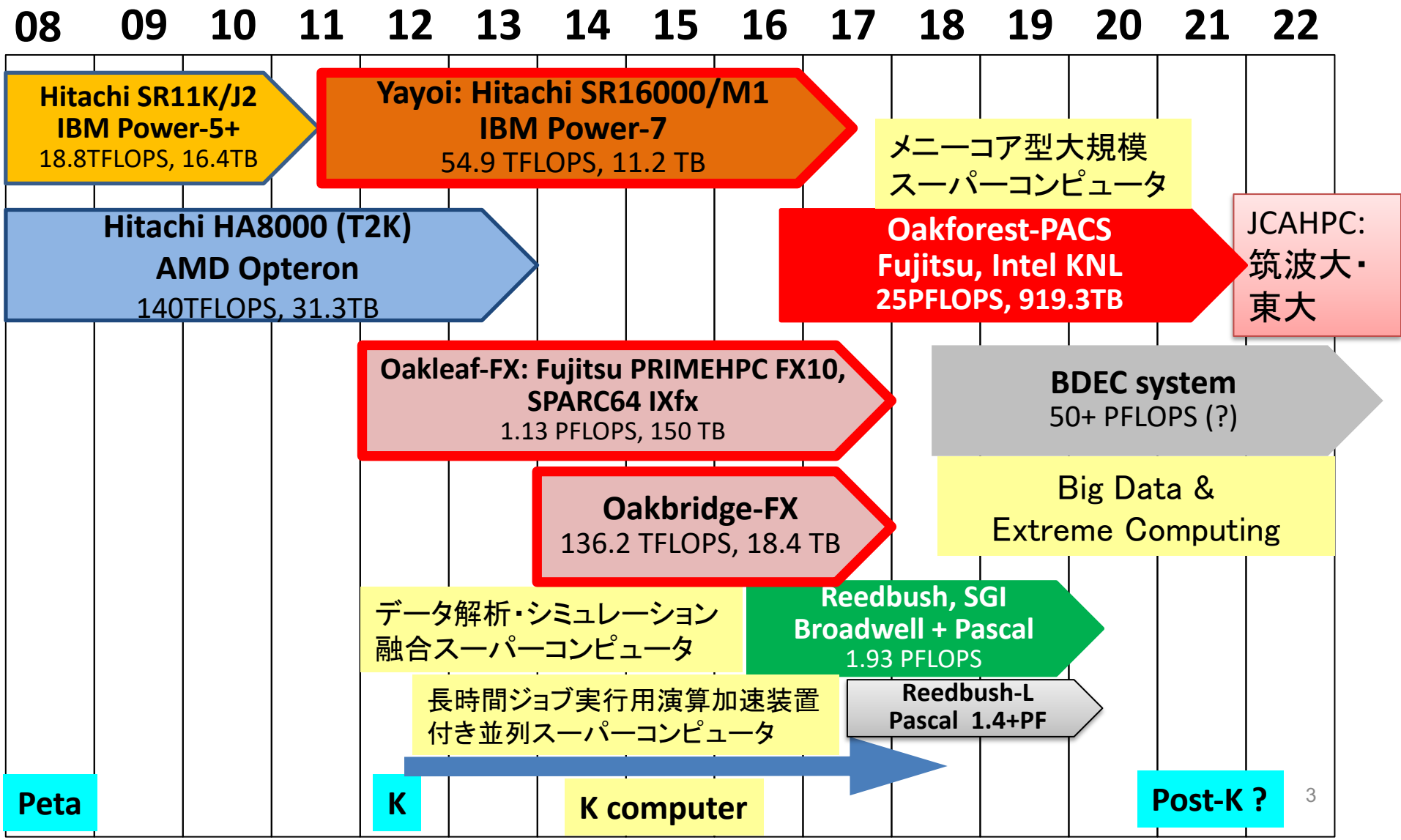
東京大学 情報基盤センター

担当：下川辺 隆史
shimokawabe @ cc.u-tokyo.ac.jp

東大情報基盤センター
スーパーコンピュータの概略

東大センターのスパコン

2基の大型システム，6年サイクル



現在2システム運用中

- Oakleaf-FX (富士通 PRIMEHPC FX10)
  - ✓ 1.135 PF, 京コンピュータ商用版, 2012年4月 ～ 2018年3月

- Oakbridge-FX (富士通 PRIMEHPC FX10)
  - ✓ 136.2 TF, 長時間実行用（168時間）, 2014年4月 ～ 2018年3月
- Reedbush (SGI, Intel BDW + NVIDIA P100 (Pascal))
  - ✓ データ解析・シミュレーション融合スーパーコンピュータ
  - ✓ 3.361 PF, 2016年7月～ 2020年6月
  - ✓ 東大ITC初のGPUシステム (2017年3月より), DDN IME (Burst Buffer)

- Oakforest-PACS (OFP) (富士通、Intel Xeon Phi (KNL))
  - ✓ JCAHPC (筑波大CCS＆東大ITC)
  - ✓ 25 PF, TOP 500で6位 (2016年11月) (日本で1位)
  - ✓ Omni-Path アーキテクチャ, DDN IME (Burst Buffer)

東京大学情報基盤センター スパコン（1/2）

Reedbush（SGI Rackable クラスタシステム）

Reedbush-U (2016/7/1 ~)

- 理論性能：508TFlops
- ノード数：420
- ノード構成：Intel Xeon Broadwell x2

Reedbush-H (2017/3/1 ~)

- 理論性能：1418TFlops
- ノード数：120
- ノード構成：Intel Xeon Broadwell x2 + **NVIDIA P100 GPU x2**

Reedbush-L (**2017/10/1 ~**)

- 理論性能：1435TFlops
- ノード数：64
- ノード構成：Intel Xeon Broadwell x2 + **NVIDIA P100 GPU x4**

東京大学情報基盤センター スパコン（2/2）

筑波大学計算科学研究センターと共同運用

Oakforest-PACS （Fujitsu PRIMERGY CX600）

Total Peak performance	: 25 PFLOPS
Total number of nodes	: 8,208
Total memory	: 897.7 TB
Peak performance per node	: 3.046 TFLOPS
Main memory per node	: 96 GB (DDR4)
	+ 16 GB(MCDRAM)
Disk capacity	: 26.2 PB
File Cache system (SSD)	: 960 TB

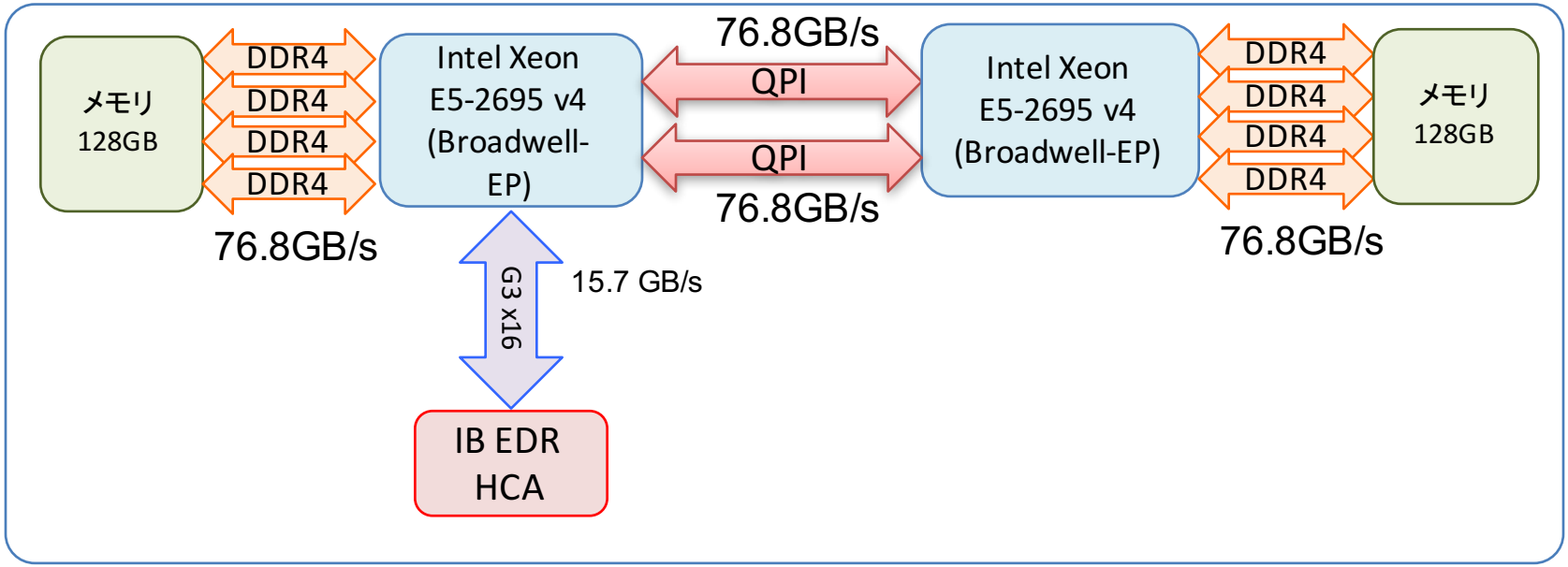
Intel Xeon Phi 7250 1.4 GHz 68 core x1 socket

2016年12月1日試験運転開始

2017年4月3日正式運用開始

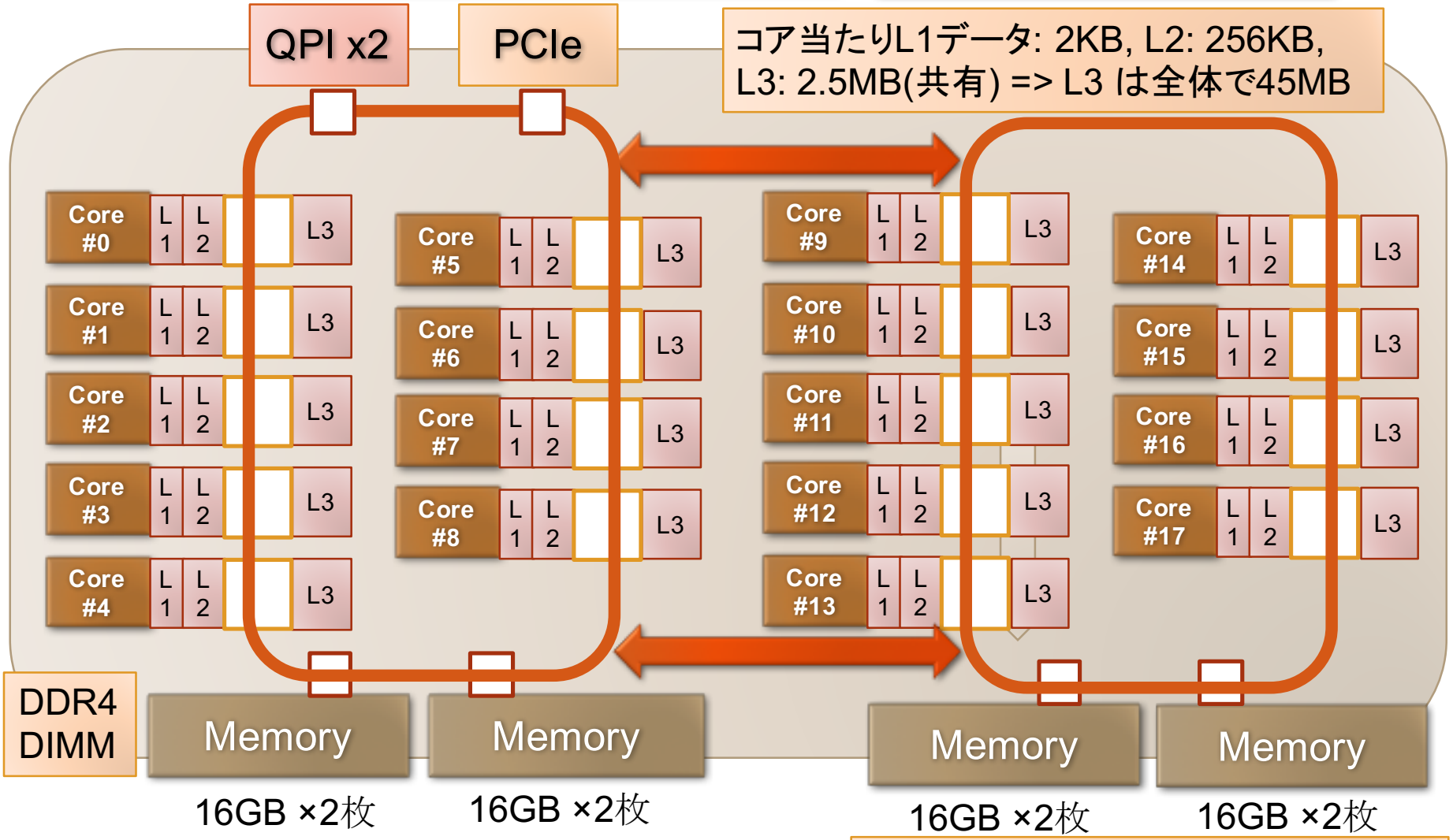
Reedbush-Uノードのブロック図

- メモリのうち、「近い」メモリと「遠い」メモリがある
  => NUMA (Non-Uniform Memory Access)
  (FX10はフラット)



Broadwell-EPの構成

1ソケットのみを図示

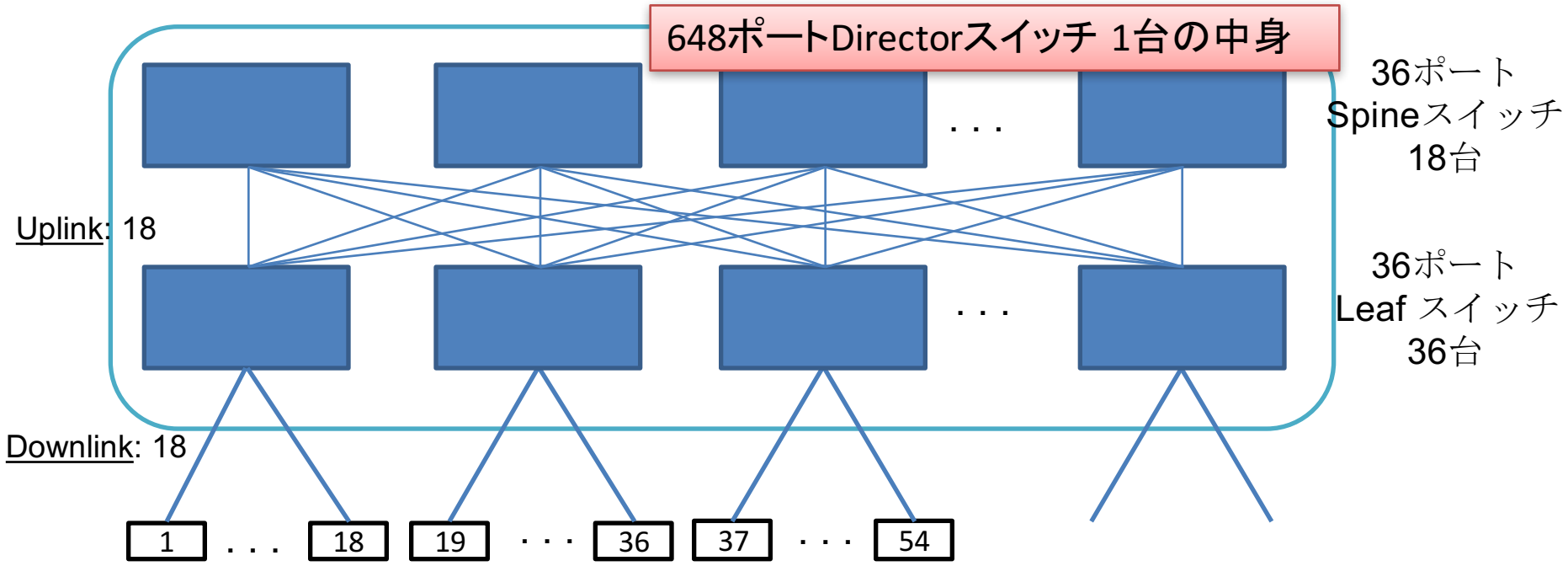


ソケット当たりメモリ量：16GB×8＝128GB

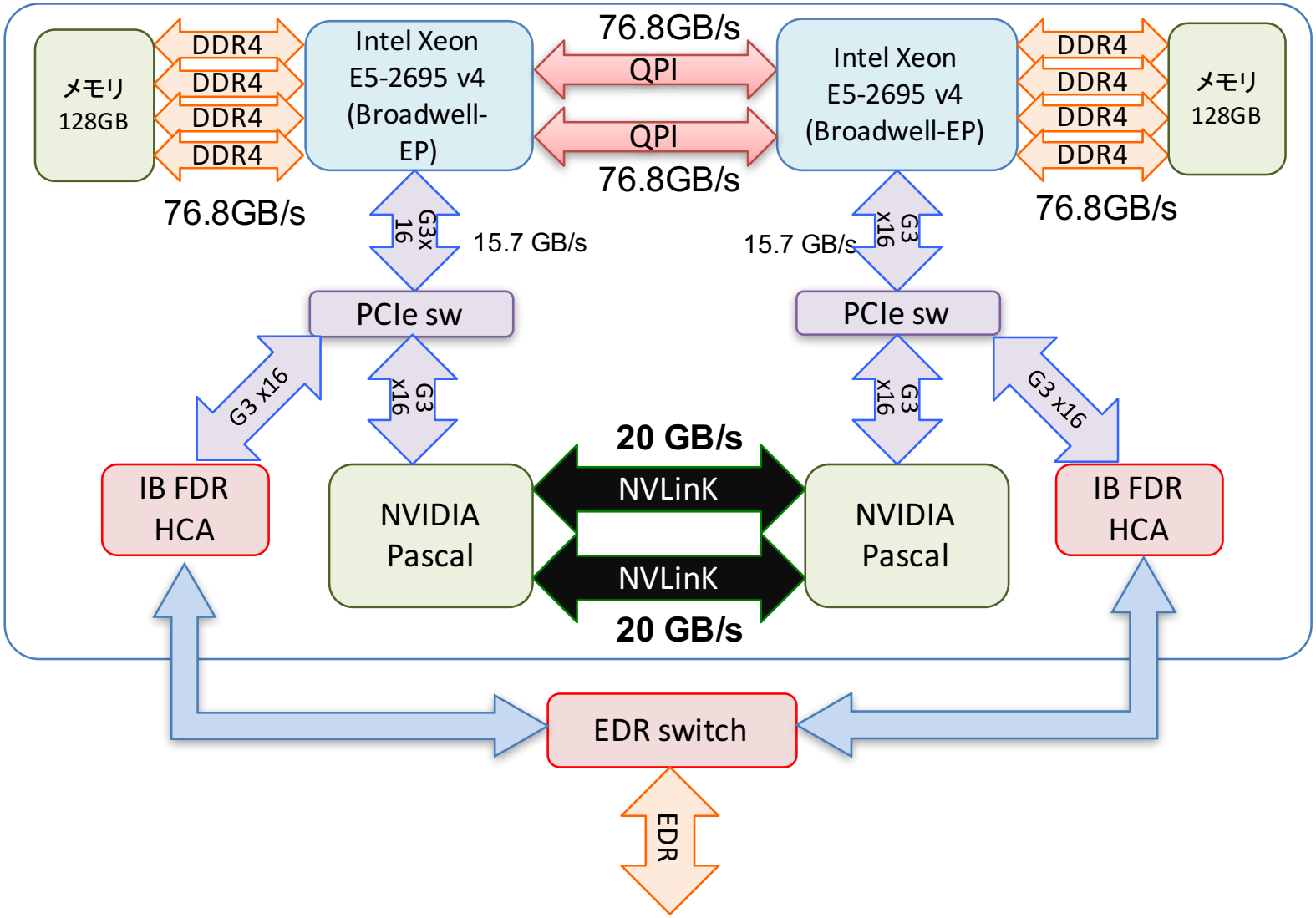
76.8 GB/秒
=(8Byte×2400MHz×4 channel)

Reedbush-Uの通信網

- フルバイセクションバンド幅を持つFat Tree網
  - ✓ どのように計算ノードを選んでも互いに無衝突で通信が可能
- Mellanox InfiniBand EDR 4x CS7500: 648ポート
  - ✓ 内部は36ポートスイッチ (SB7800)を (36+18)台組み合わせたものと等価



Reedbush-Hノードのブロック図



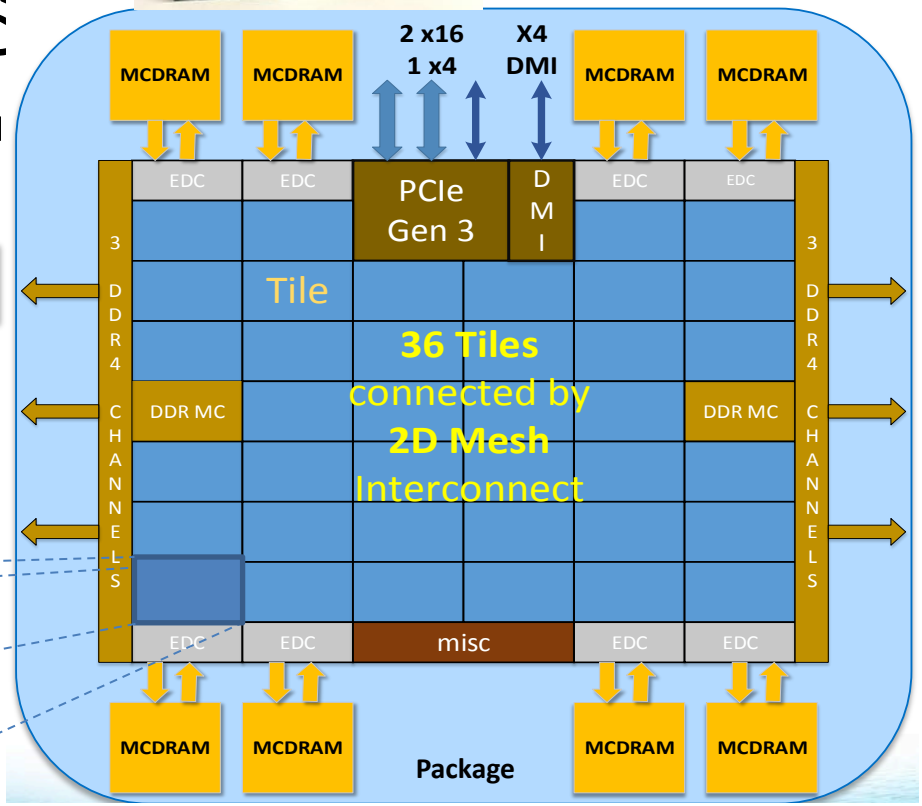
Oakforest-PACS 計算ノード

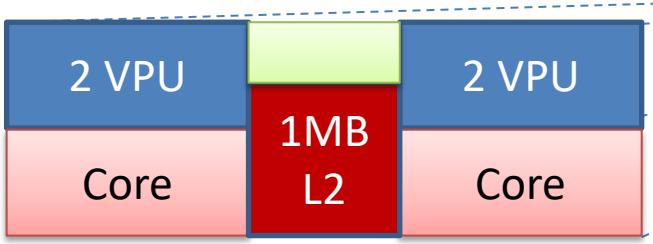
- Intel Xeon Phi (Knights Landin
  - 1ノード1ソケット
- MCDRAM: オンパッケー
  の高バンド幅メモリ16G
  + DDR4メモリ

ソケット当たりメモリ量：16GB×6＝96GB

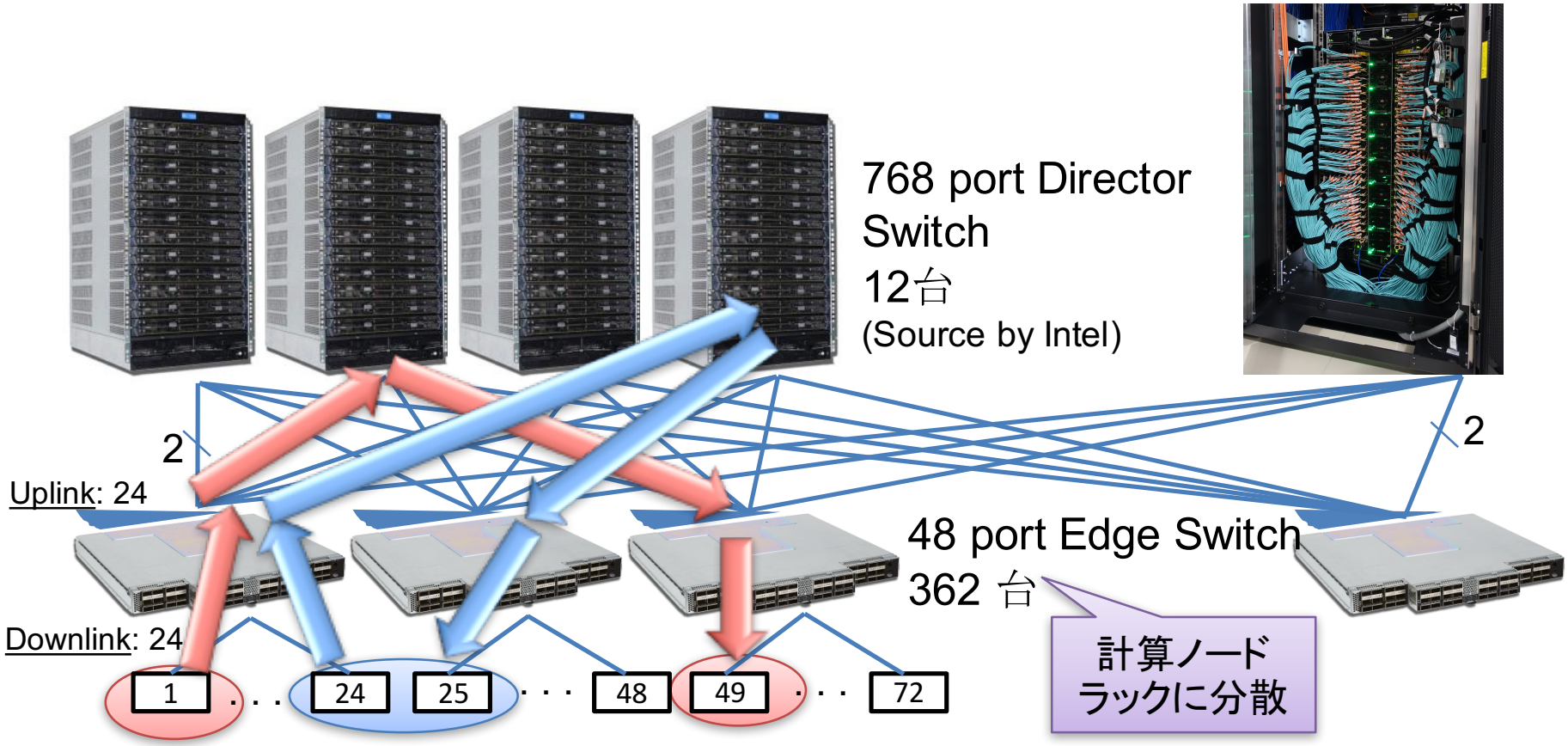
MCDRAM: 490GB/秒以上（実測）
DDR4: 115.2 GB/秒
=(8Byte×2400MHz×6 channel)

HotChips27
KNLスライドより





Oakforest-PACS: Intel Omni-Path Architecture による フルバイセクションバンド幅Fat-tree網



コストはかかるがフルバイセクションバンド幅を維持
- システム全系使用時にも高い並列性能を実現
- 柔軟な運用：ジョブに対する計算ノード割り当ての自由度が高い

東大情報基盤センターOakforest-PACSスーパーコンピュータシステムの料金表（2018年4月1日）

- パーソナルコース
  - 個人研究者（大学・公共機関）のみ
  - 1口：100,000円=17,280トークン（2ノード年相当）
- グループコース
  - 1口：400,000円 (企業 480,000円) =69,120トークン（8ノード年相当）

- 基準、最大ノード数
  - 基準（口数×8）ノード、最大2,048ノードまで

- トークン消費について
  - 基準ノードまでは、トークン消費係数が1.0
  - 基準ノードを超えると、超えた分は、消費係数が2.0になる
  - 大学等のユーザはReedbushとの相互トークン移行も可能

東大情報基盤センターReedbushスーパーコンピュータシステムの料金表（2018年4月1日）

- パーソナルコース
  - 個人研究者（大学・公共機関）のみ
  - 1口：150,000円=17,280トークン（Reedbush-U 2ノード年相当）
- グループコース
  - 1口：300,000円 (企業 360,000円) =34,560トークン（ Reedbush-U 4ノード年相当）
- 基準、最大ノード数
  - Reedbush-U: 基準（口数×4）ノード、最大128ノードまで
  - Reedbush-H: 基準（口数×1）ノード、最大32ノードまで
  - Reedbush-L: 基準（口数×1）ノード、最大16ノードまで
- トークン消費について
  - 基準ノードまでは、トークン消費係数がRB-U: 1.0、RB-H: 2.5、RB-L: 4.0
  - 基準ノードを超えると、超えた分は、それぞれの消費係数が2倍となる
  - 大学等のユーザはOakforest-PACSとの相互トークン移行も可能
  - ノード占有利用（ノード固定）制度あり

JPY (=Watt)/GFLOPS Rate

Smaller is better (efficient)

System	JPY/GFLOPS
Oakleaf/Oakbridge-FX (Fujitsu) (Fujitsu PRIMEHPC FX10)	125
Reedbush-U (SGI) (Intel BDW)	62.0
Reedbush-H (SGI) (Intel BDW+NVIDIA P100)	17.1
Oakforest-PACS (Fujitsu) (Intel Xeon Phi/Knights Landing)	16.5

トライアルユース制度について

安価に当センターのReedbush-U/H/L, Oakforest-PACSシステムが使える「無償/有償トライアルユース」制度があります。

- アカデミック利用（有償のみ）
  - パーソナルコース、グループコース（1ヶ月～3ヶ月）
- 企業利用
  - パーソナルコース（１ヶ月～３ヶ月、有償）
    - 最大ノード数：　RB-U: 16、RB-H: 4、 RB-L: 4、OFP: 16
    - 講習会いずれかの受講が必須、審査無
  - グループコース
    - 無償トライアルユース：（１ヶ月～３ヶ月）
      - 最大ノード数：　RB-U: 128、RB-H: 32、 RB-L: 16、OFP: 2048
    - 有償トライアルユース：（１ヶ月～最大通算９ヶ月）
      - 最大ノード数：無償と同等
    - スーパーコンピュータ利用資格者審査委員会の審査が必要（年2回実施）
  - 双方のコースともに、簡易な利用報告書の提出が必要

スーパーコンピュータシステムの詳細

- 以下のページをご参照ください
  - 利用申請方法
  - 運営体系
  - 料金体系
  - 利用の手引

  などがご覧になれます。

https://www.cc.u-tokyo.ac.jp/guide/