

# パネル討論

## 次世代先端的計算基盤のあり方を考える

理化学研究所 計算科学研究センター  
 プロセッサ研究チーム チームリーダー  
 佐野 健太郎

2020/12/22



### 理研 計算科学研究センター

#### 高性能計算 (HPC)分野を牽引する研究センター

- ✓ スーパーコンピュータ富岳の開発と運用
- ✓ スーパーコンピュータを用いた研究のための最先端インフラの整備と普及
- ✓ HPCに関する先端的研究

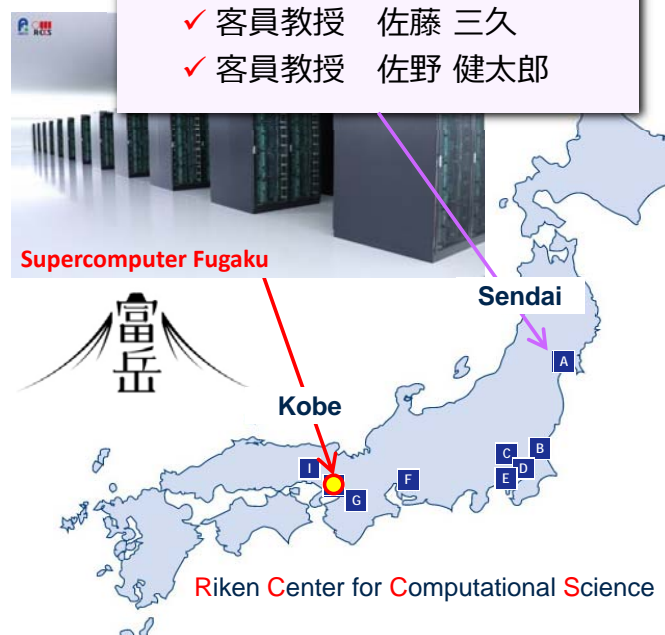
#### プロセッサ研究チーム

- ✓ 将来の高性能計算機アーキテクチャの探求
- ✓ 現在のHPCシステムの高度利用の探求
  - **PI Dr. Kentaro Sano**
  - Dr. Tomohiro Ueno, Dr. Takaaki Miyajima
  - Dr. Jens Huthmann, Dr. Atsushi Koshiba
  - Dr. Boma A. Adhi

**研究員募集中!**

#### 東北大学 理研連携講座

- ✓ 大学院情報科学研究科 先進的計算システム論講座
- ✓ 客員教授 佐藤 三久
- ✓ 客員教授 佐野 健太郎



# NGACIホワイトペーパーでの役割

## デバイス・回路・アーキテクチャWG とりまとめ

- ✓ 特に「4.2 専用システム混載型および新たなる可能性」を担当  
(デジャブ：前回のWPでも、「アクセラレータ」部分を担当したような・・・)

## 今後のHPC開発に重要と思われる部分

- ✓ 準専用 または 汎用アクセラレータ  
アーキテクチャとそのプログラミング方式
- ✓ チップ内拡張型 (SoC または MCP)  
とメモリサブシステム
- ✓ アクセラレータ間のデータ移動機構

### 準専用アクセラレータ

- ・マトリックス/テンソルプロセッサ
- ・グラフィックプロセッサ
- ・データフロープロセッサ等

### 汎用アクセラレータ

- ・FPGAやCGRA等

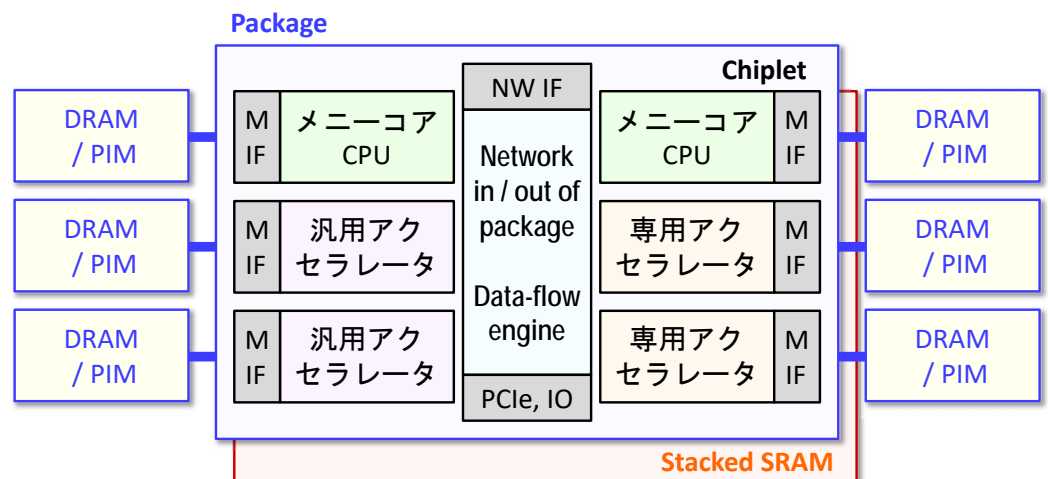
### 複数チップレットによる

- ・メニーコアCPU
- ・汎用アクセラレータ
- ・ネットワークインパッケージ (NiP)
- ・外部ネットワークインターフェース

## 例えば

### ● CPUとアクセラレータ間のデータ移動・共有方法が重要

- ✓ オーバーヘッドを低減、隠蔽
- ✓ 制御・同期・コヒーレンスを局所化、スケーラブルに



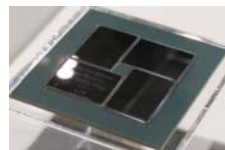
# 1) 10年後にあるべきスパコンの姿として、汎用指向 vs 専用指向のどちらを目指した開発をすべきか？

## 専用指向を目指すべき

- ✓ 汎用のみでは目標性能に到達しない
- ✓ 高い性能を必要とする個々のアプリドメインや計算カーネルに対し、**専用機構により汎用機の性能を補完**すべき
- ✓ アクセラレータといっても、**汎用、準専用**（特定のドメイン向け）、**完全専用**まで様々
- ✓ **トレードオフ**：
  - 性能 (得られる性能・電力性能)
  - 可搬性 (カバーされるドメインの広さ)
  - 生産性 (プログラム容易性)
  - 費用対便益 (かかるコストに見合う利用が見込めるか)



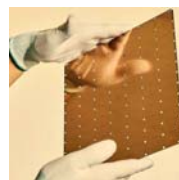
Graphcore



PFN MN-Core



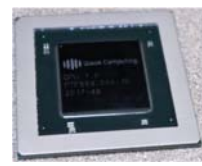
Groq



Cerebras



SambaNova



Wave computing



XILINX  
VERSAL  
(ACAP)

# 2) Society5.0関連のアプリの、従来のHPCアプリとの違いは何か。システム面や運用面での課題は何か？

## Society5.0関連アプリ（例えばAIやリアルタイムCPSアプリ）

- ✓ 高スループットに加え、**遅延低減が重要**な場合も（特にCPS）。リアルタイム性も？
- ✓ 数値計算としても、**必要精度が異なる**（例 DLのBFloat16, 8, Binary, Ternary。適応精度）
- ✓ 主要演算が**数値計算以外の演算**の場合も（論理演算、Look-up、マッチング、比較）
- ✓ より複雑・アプリ毎にカスタマイズされた**メモリ参照パターン**
- ✓ 求められる**並列性が異なる？**（Embarrassingly parallel, Batch parallelで十分だったり）

## システム面や運用面での課題は何か？ > 特にシステム面について

- ✓ 異なるメモリ参照パターンに適合してデータ再利用を可能とする  
コンフィギャラブル キャッシュメモリ・オンチップバッファ
- ✓ 数値計算以外の演算サポート
- ✓ 低遅延、リアルタイム性への要求に対するサポート

### 3) 今後最も重点的に取り組むべき研究開発要素は何か？

#### 専用・準専用計算機アーキテクチャ

- ✓ 特に、汎用CPUと密に結合して動作する（準）専用計算機構
- ✓ 汎用CPUとの結合方式・メモリ空間共有方式（物理的にも、論理的にも）
- ✓ 専用計算機構のメモリサブシステム（例：コンフィギャラブルバッファ）
- ✓ 専用機間のデータ移動方式（専用網など）
- ✓ 専用化すべきアプリやドメインの探求



#### 事例：

- ✓ FPGAクラスタ試作システム（既存マシンや富岳から利用、FPGA間通信網）
- ✓ CGRA設計空間探索（ラインバッファなどを構築できるリコンフィギャラブルバッファタイル）

#### そのコンパイラ、プログラミング方式

- ✓ 新しく書かせるコードで楽に高性能を実現できれば良いのではないかと（未来志向）
- ✓ 既存のコード資産利用については、議論の余地あり  
（とりあえず動作することが重要で、性能最適化を行ってほしい）

### 4) 今後のスパコン開発にあたってHPC関連ベンダーには技術面や運用面などで何を期待しているか？

#### 既存のメニーコアプロセッサ以外の新しいアーキテクチャへの投資

- ✓ 大学や研究所などとの積極的な共同研究等
- ✓ ハードウェアのみならず、コンパイラも