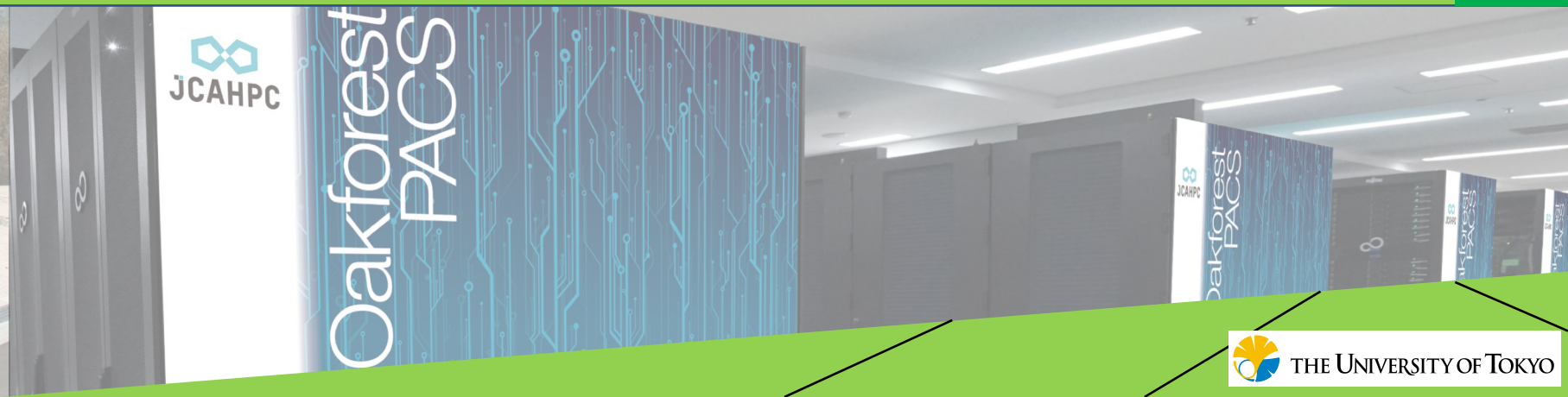
INFORMATION TECHNOLOGY CENTER
THE UNIVERSITY OF TOKYO

次世代スパコンのための SW/HW最適化

有間　英志

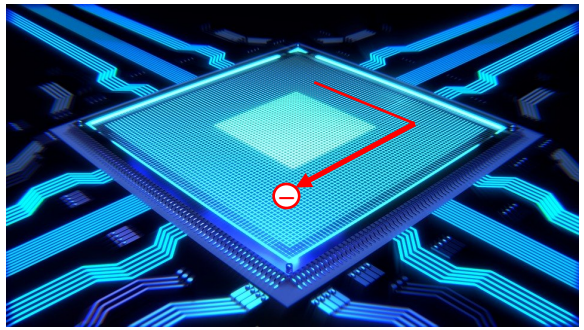


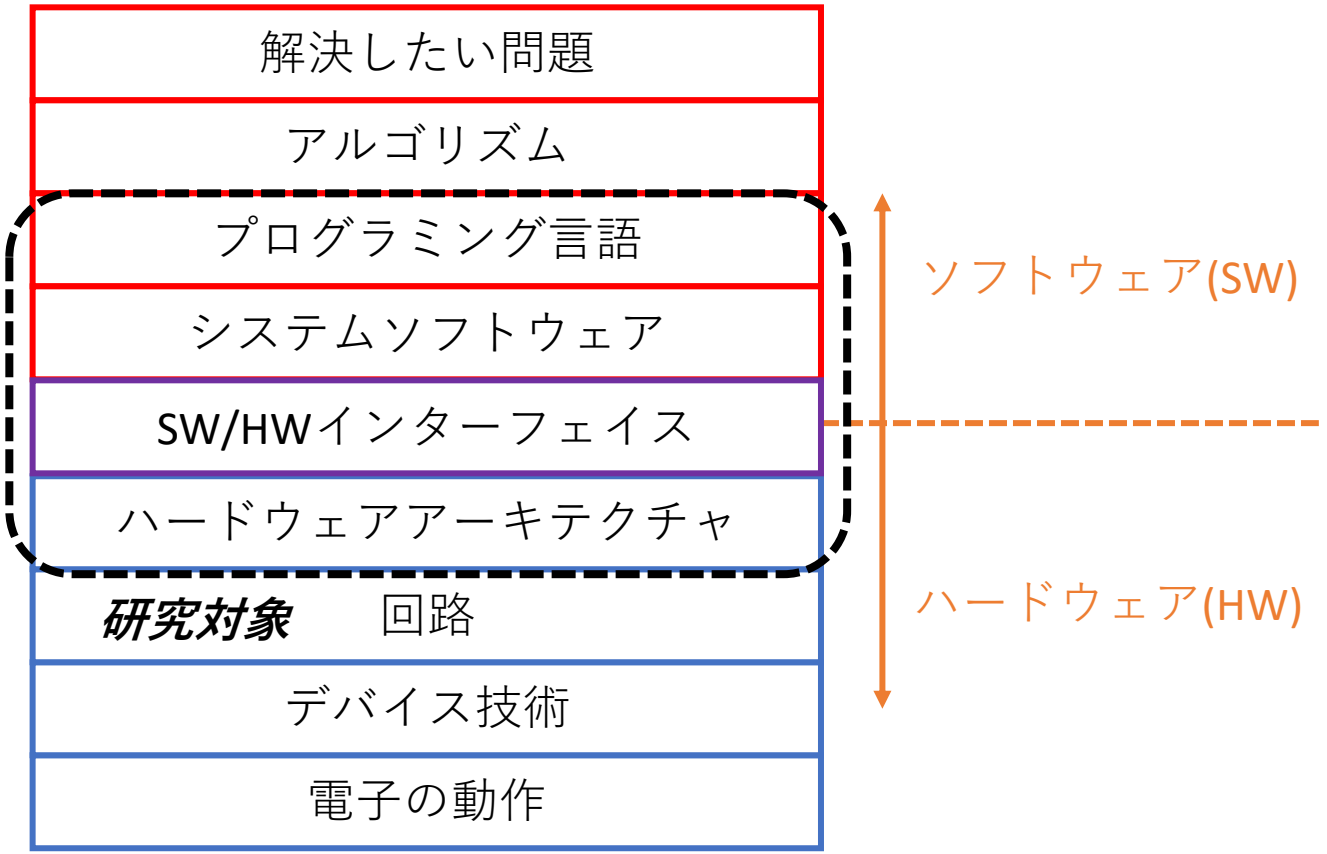
THE UNIVERSITY OF TOKYO

研究対象

問題をより高速、省電力、(高信頼)に解決するための、基盤ソフトウェア、SW/HWインターフェイス、ハードウェアアーキテクチャ、並びにその協調に関する研究が対象

$$\begin{aligned} \max \quad & \textstyle\sum_{i \in I} v_i x_i \\ \text{s.t.} \quad & \textstyle\sum_{i \in I} w_i x_i \leq W \\ & x_i \in \mathbb{N} \qquad (\forall i \in I) \end{aligned}$$





より具体的には

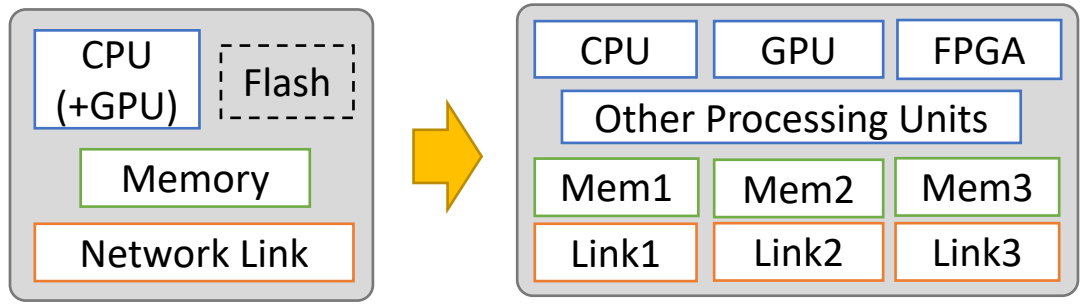
現在～将来に至るスパコンを対象とし、その高性能化・省電力化をソフトウェア・ハードウェアの両観点から、特に**電力制御・データ管理**に着目して行う

世のスパコン

Oakbridge-CX

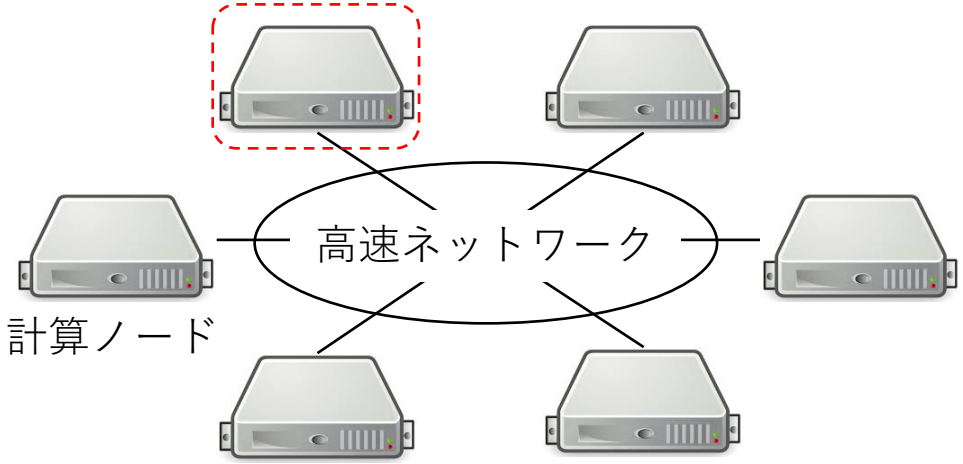
Oakforest-PACS

Fugaku†



現在の計算ノード構成

将来的な計算ノード構成



†https://www.riken.jp/en/news_pubs/news/2020/20200623_1/

Free server icon: https://www.freeiconspng.com/img/2306

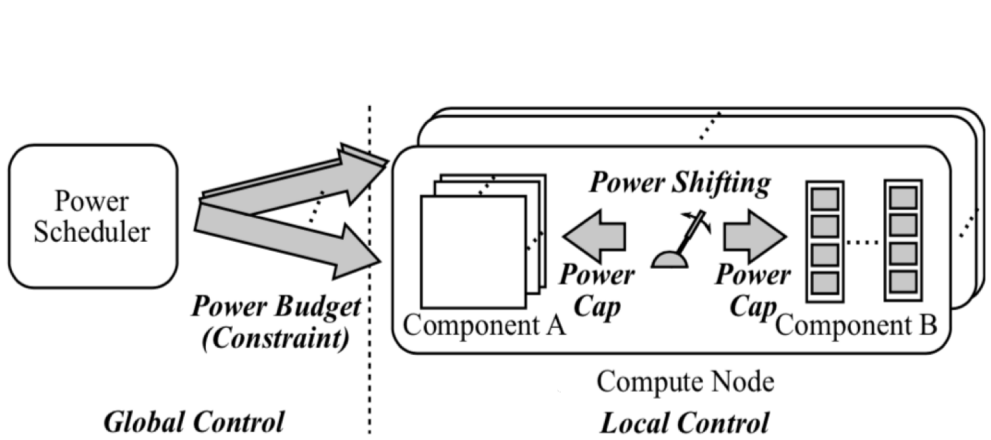
電力制御に関する研究事例1: Footprint-Aware Power Capping [E. Arima+ISC'20]

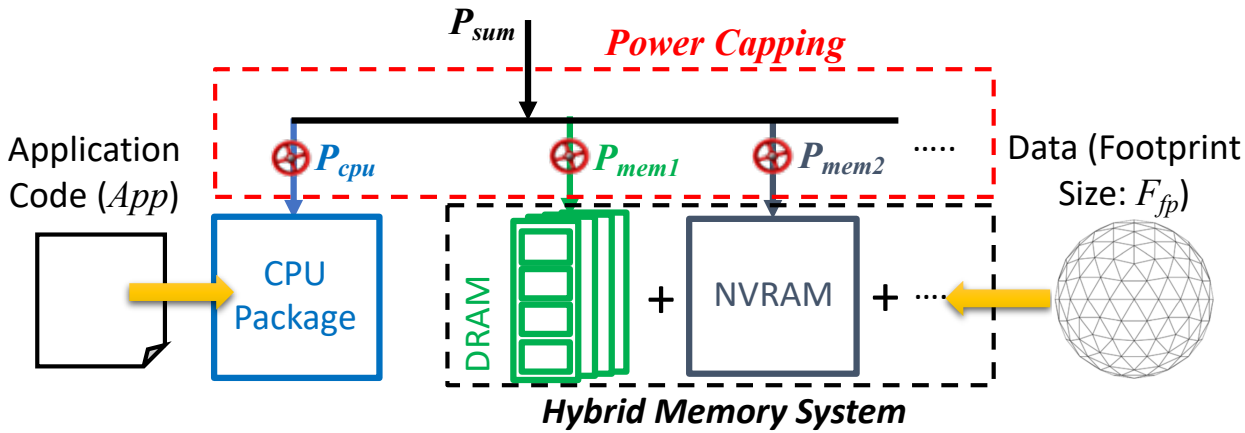
前提: 階層的な電力制御

- 電力スケジューラ(司令塔)が各計算ノードに適切に電力予算を割り当て、各計算ノードはそれに従って、各コンポーネント(CPU、メモリ等)への電力割り当てを最適化する

着眼点: 各計算ノードにおける最適な電力割り当ては、アプリケーションの扱う問題サイズ(データフットプリントサイズ)に依存

- 特に異種メモリを混載したシステム(例: Oakforest-PACS)上では顕著





電力制御に関する研究事例1: Footprint-Aware Power Capping [E. Arima+ISC'20]

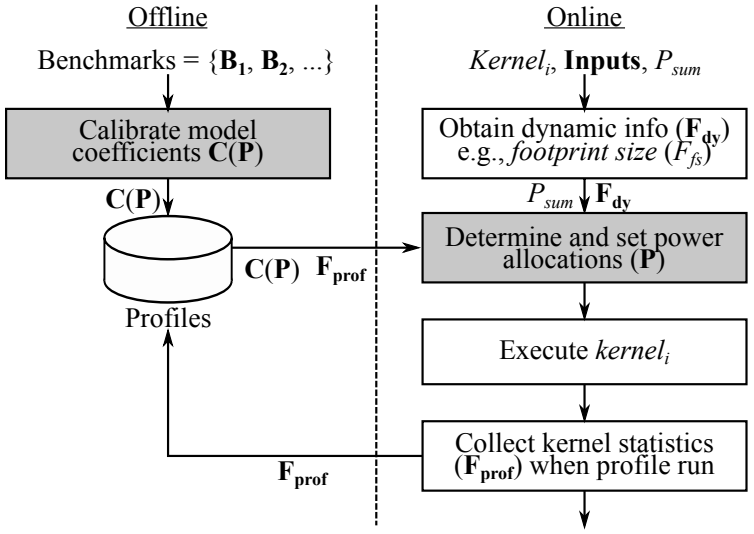
提案の概要:

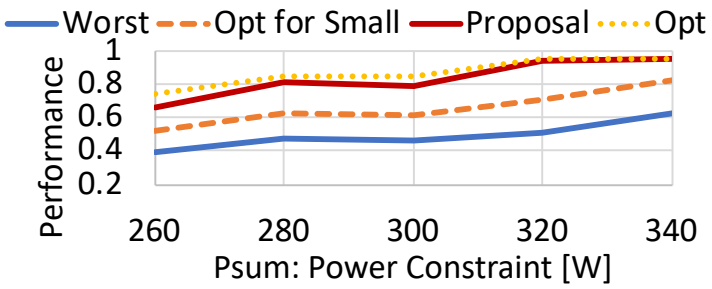
- 当該電力割り当てを最適化問題として定式化し、これを解く為の性能モデリングを行なった
- 上記定式化に基づき、電力割り当てを最適化するソフトウェアフレームワークを提案(効率的モデル係数調整手法及び電力割り当て最適化アルゴリズムから成る)

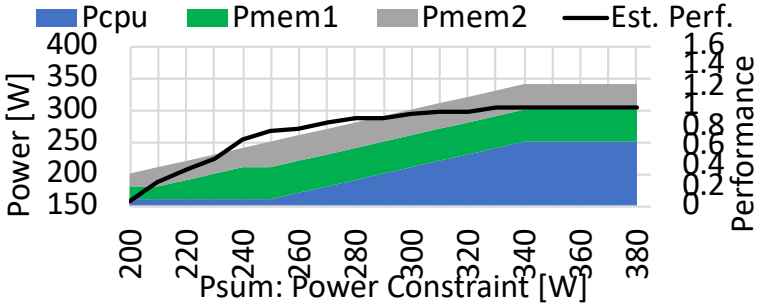
評価結果: ほぼ最適な電力割り当てが可能であることを確認

Given *Kernel*, **Inputs**, P_{sum} ($\Rightarrow$ **F**, P_{sum})
Max *Obj*(**P**, **F**)
s.t. $\Sigma P_x \leq P_{sum}$
$P_x \in S_{Px}$ (x = cpu, mem1, …)

Kernel: target kernel region **Inputs**: inputs for the app = (arg1, arg2, …) **F**: app feature parameters (e.g., F/B rate)
Obj(**P**, **F**): objective function
P_{sum}: given total power budget [W] **P**: power allocations = (P_{cpu}, P_{mem1}, …) S_{Px}: Set of power caps for component *x* (e.g., ={20, 30, 40})

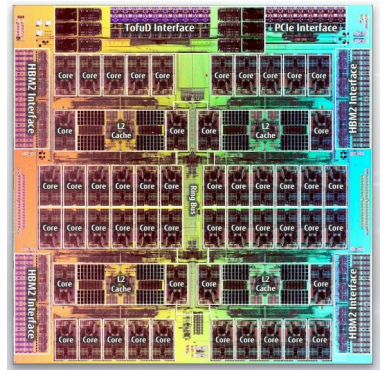






電力制御に関する研究事例2: スパコン「富岳」を利用した電力評価

- Evaluation of Power Controls on Supercomputer Fugaku [Y. Kodama+EEHPC@CLUSTER'20]
  - 富岳に搭載されているA64FXプロセッサの電力制御機能を2万ノード以上を用いて評価
    - Eco mode: 浮動小数点演算器の利用を制限し、電力を抑える
    - Boost mode: 動作周波数及び電圧を増加させ、電力と引き換えに性能を向上
    - Core retention: 利用しないコアを低電力モードに
  - 運用上役立つ幾つかの知見をレポート
- 次世代システムの為の性能・電力評価環境整備
  - 富岳の為に利用したツールを拡張し、スパコン向けプロセッサの未来を予想



A64FX†

†https://www.nextplatform.com/2019/11/13/a64fx-arm-chip-gets-a-big-push-from-cray/

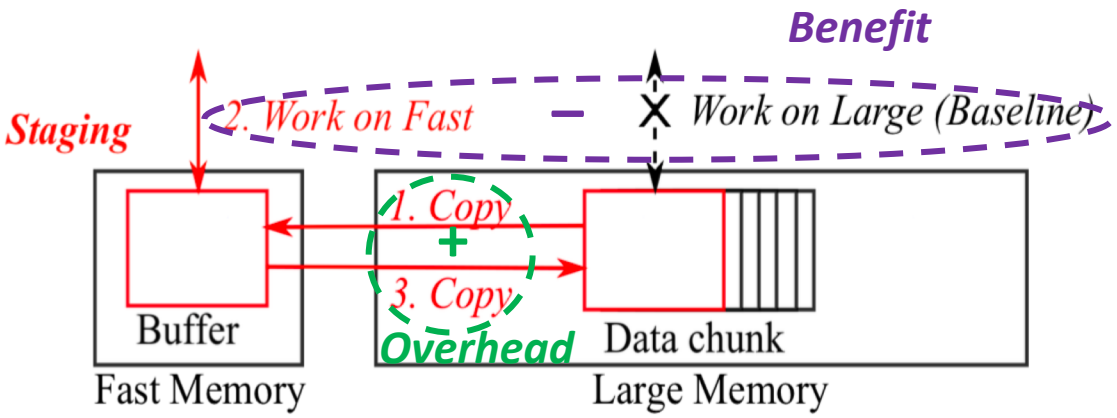
データ管理に関する研究事例1: Pattern-Aware Staging [E. Arima+ISC'20]

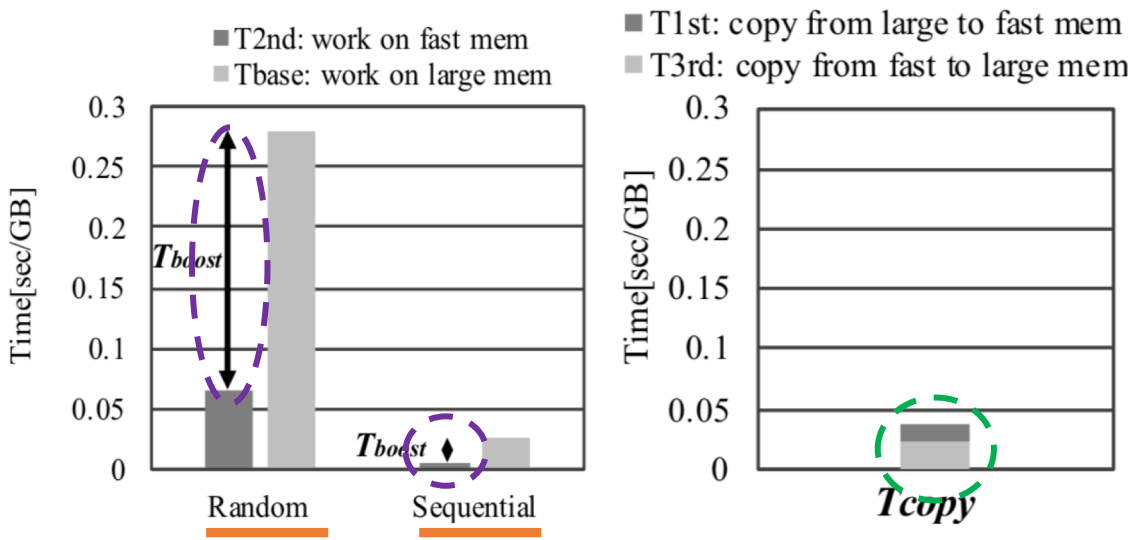
問題: 1計算ノード内に「小容量で高速なメモリ」と「大容量で低速なメモリ」が存在する場合に、どの様にして高速性と大容量を両立すべきか？

- 例えば、Oakforest-PACSで導入されており、将来的には主流となる

着眼点: 「小容量で高速なメモリ」によって性能を享受できるかどうかは、メモリアクセスパターンに強く依存

- 不規則なアクセスか？疎なアクセスか？



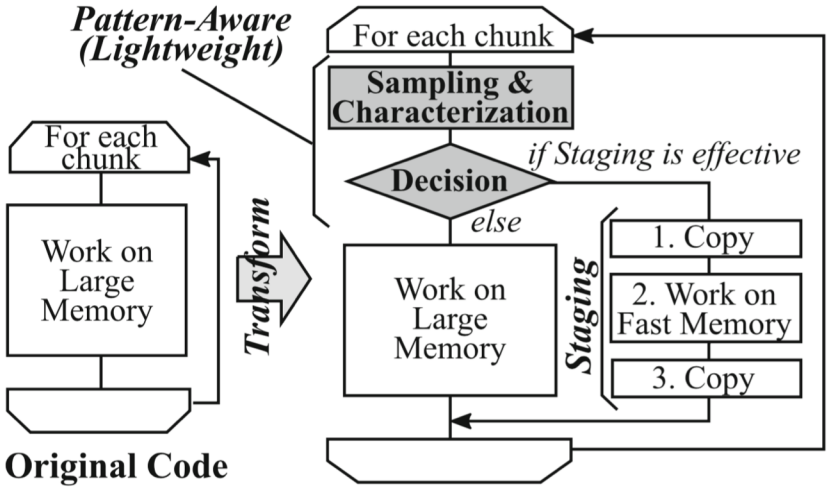


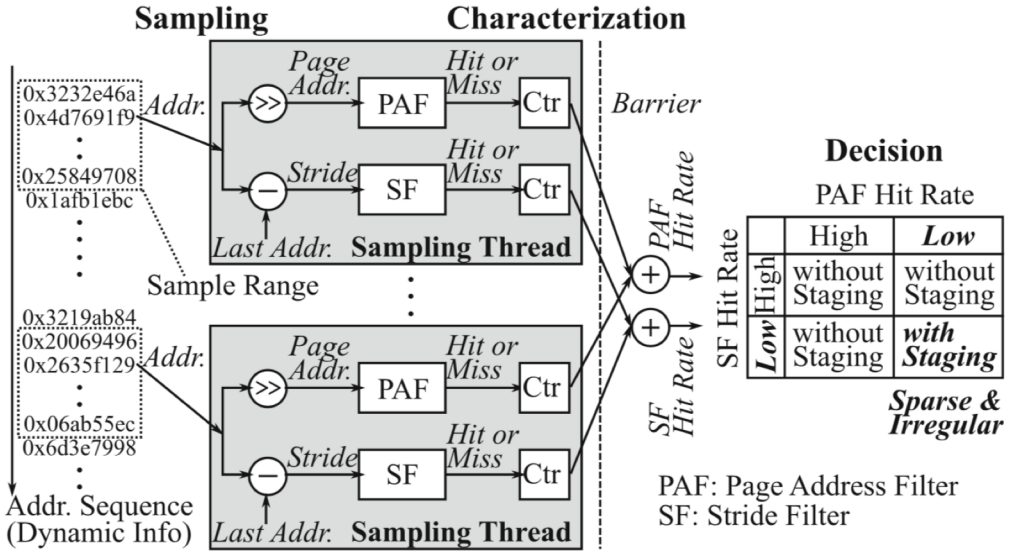
データ管理に関する研究事例1: Pattern-Aware Staging [E. Arima+ISC'20]

提案の概要: 元のコードを変換し、以下の機能を追加(コンパイラによる自動化可)

- **Sampling:** Helper Threading [e.g., M. Kamruzzaman+ ASPLOS'11, J. Lee+ TPDS'09]を応用した、メモリアクセスシーケンス取得手法
- **Characterization:** Bloom Filterと呼ばれる確率的データ構造を用いたアクセスパターン識別
- **Decision:** データを動かすべきかどうかを上記識別結果を元に判断

評価結果: 様々なHPCカーネルで評価し、平均1.9倍、最大で3倍の性能向上を達成



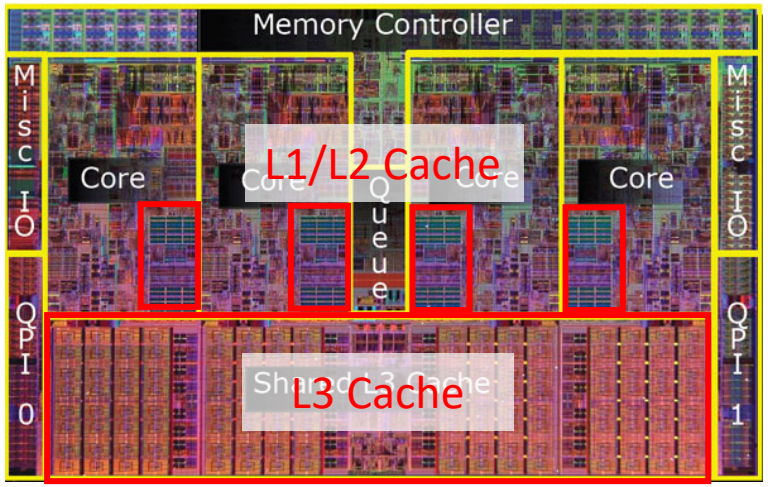


データ管理に関する研究事例2: 先進的ハードウェアキャッシュ制御 [E. Arima@DSD'20]

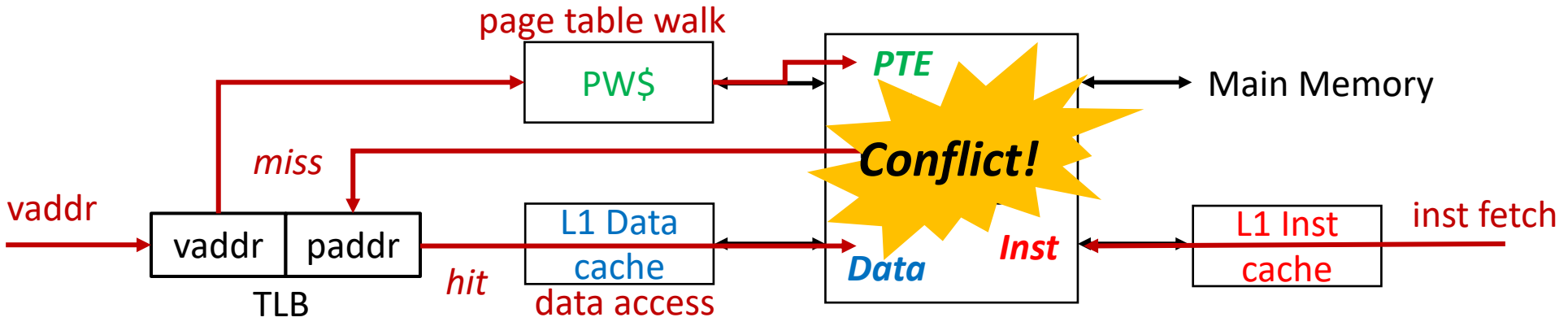
対象: CPU内のハードウェアキャッシュ(高頻度アクセスデータの一時保存領域)

着眼点: 異種データ間で起きるキャッシュの競合

- 命令コード、データ、PTE(アドレス変換表)
- 昨今のアプリケーションにて顕著: 命令コードサイズの肥大、扱うデータサイズの増大、アクセスパターンの複雑化



Intel Nehalem Processor†



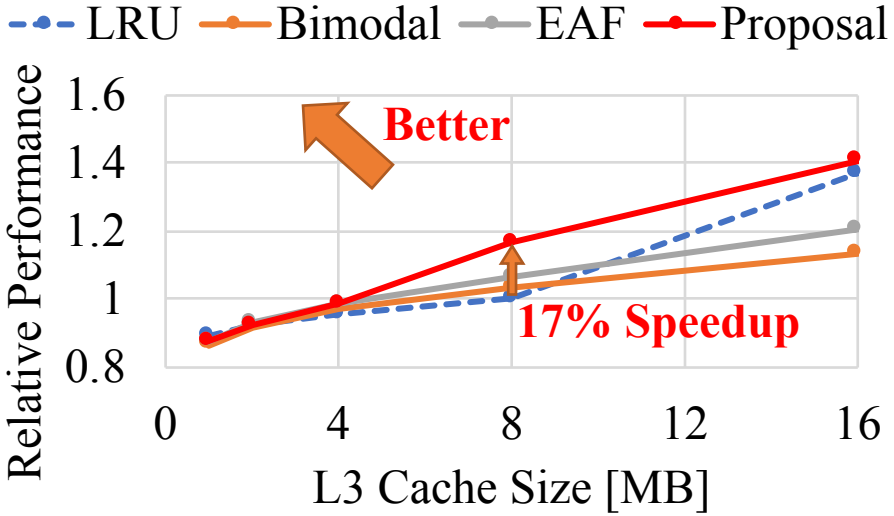
†https://pcper.com/2008/08/inside-the-nehalem-intels-new-core-i7-microarchitecture/

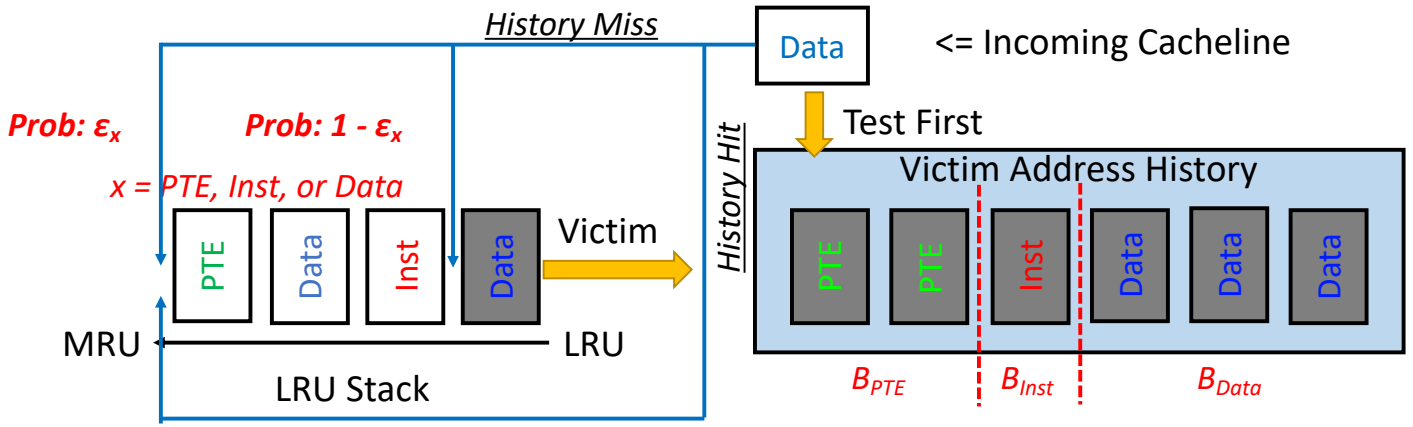
データ管理に関する研究事例2: 先進的ハードウェアキャッシュ制御 [E. Arima@DSD'20]

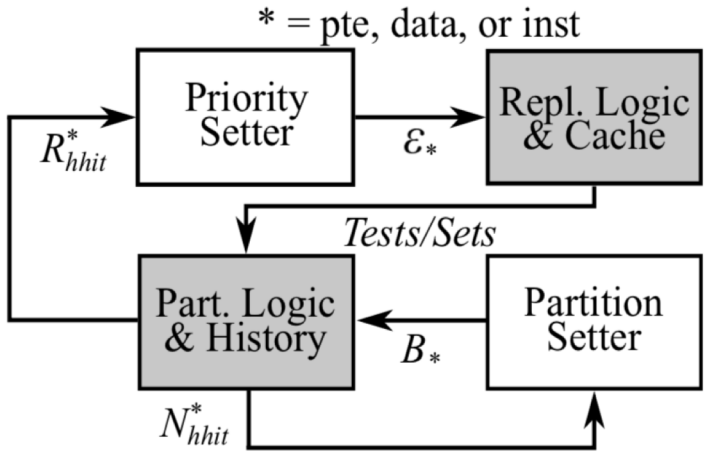
アプローチ: (1) 各種類毎に異なる配置優先度を設定(左図ε_x); (2) 再利用性推定を行う為の履歴機構を各種類毎に分割して管理(左図B_x); (3) ε_x及びB_xを自動で最適化(右下図)

- これらのHWコストはキャッシュ全体のサイズと比較して無視できる程小さい

評価結果: 容量辺りの性能を大きく向上(右上図)







まとめと参考文献

まとめ:

- スパコンの高性能化・省電力化・(高信頼化)の為には、SW/HWの両面を考慮した最適化が不可欠であり、その余地は数多く残されている
- 今回は、電力制御・データ管理の観点から、幾つかの研究事例を紹介した

参考文献:

- Eishi Arima, Toshihiro Hanawa, Carsten Trinitis, Martin Schulz "Footprint-Aware Power Capping for Hybrid Memory Based Systems" In *Proceedings of ISC High Performance*, pp.347--369, Jun. (2020) Youtube
- Yuetsu Kodama, Tetsuya Odajima, Eishi Arima, and Mitsuhisa Sato "Evaluation of Power Controls on Supercomputer Fugaku" In *Proceedings of CLUSTER (EEHPC volume)*, pp.xx--xx, Sep. (2020) Youtube
- Eishi Arima, Martin Schulz "Pattern-Aware Staging for Hybrid Memory Systems" In *Proceedings of ISC High Performance*, pp.474--495, Jun. (2020) Youtube
- Eishi Arima "Classification-Based Unified Cache Replacement via Partitioned Victim Address History" In *Proceedings of DSD*, pp.101--108, Aug. (2020)

その他の研究、国際会議等活動: https://www.cspp.cc.u-tokyo.ac.jp/arima/index-e.html