Modeling HPC applications for in situ Analytics

Valentin Honoré, Brice Goglin, Guillaume Aupy, Bruno Raffin

Modeling HPC applications for *in situ* Analytics

Valentin Honoré[1] - Brice Goglin[1], Guillaume Aupy[1], Bruno Raffin[2]

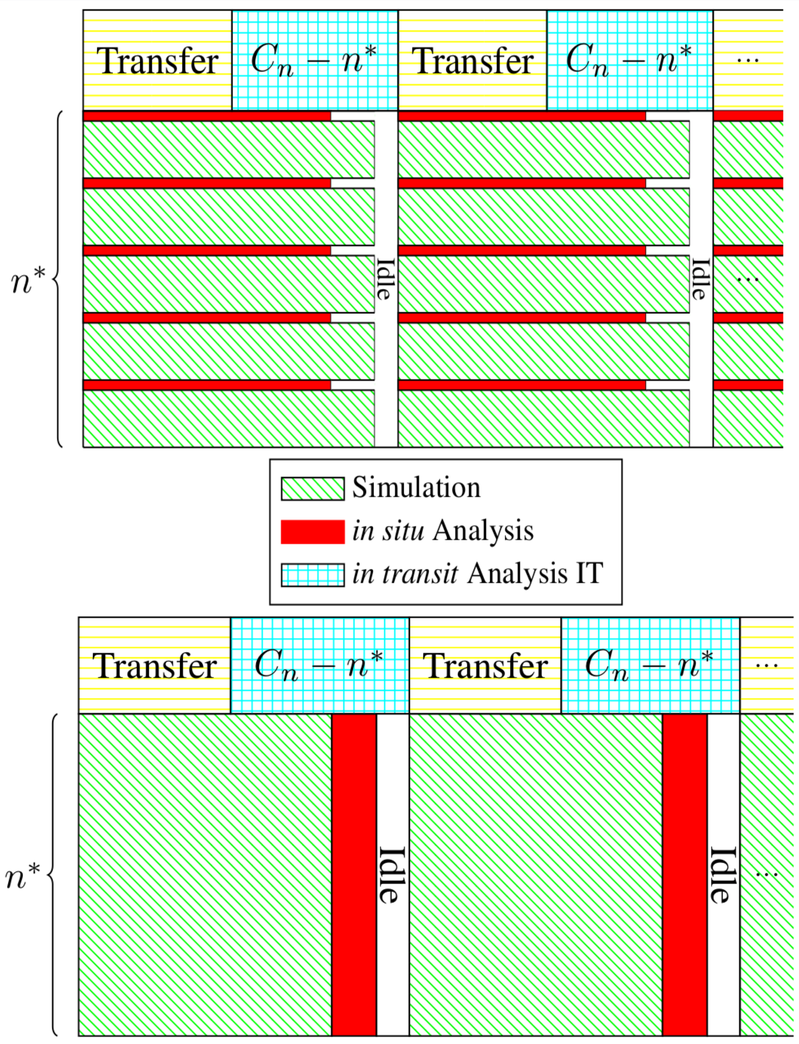
TADaaM & SATANAS - [1]Inria & Univ. Bordeaux, LaBRI [2]Inria & Univ. Grenoble, LIG

LaBRI

High Performance Computing

- **Application**
  - Simulation: compute-intensive
  - Analysis: post-processing of data
- **Data management**, **fault tolerance** & **large-scale scheduling**
- *out-of-machine* vs *in-machine* paradigm

Application Models & Platform Features



- Asynchronous (overlap) vs Synchronous (no overlap)

Objective

- **Optimize resource allocation** between the analysis and the simulation
- **Provide efficient algorithms** to perform both
  - resource partitioning
  - analysis allocations

Contributions

- **General model for HPC applications** (simulation, analysis, target architecture · · ·)
- **New algorithms** for
  - resource partitioning problem (theoretical analysis of the model)
  - scheduling problem (optimal non-polynomial + greedy)
- **Evaluation on synthetic applications:** extract key elements for *in situ* functions

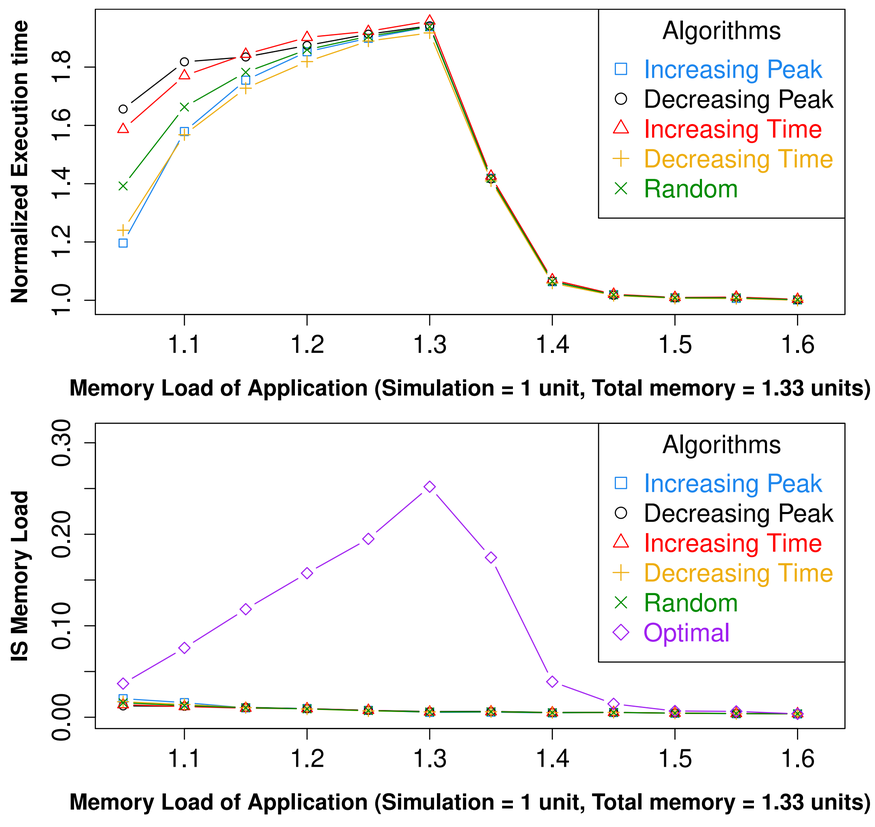
Resource Partitioning Problem

- **Application time:** maximum between simulation phase (T_S) and analysis (T_A^{IS} and T_A^{IT})
- *Given a scheduling* $(\mathcal{A}^{IS}, \mathcal{A}^{IT})$*, find* $n^* \leq C_n$*, and* $c^* \leq c$ *that minimize*

$$\max\left(T_S(n^*, c^*), T_A^{IS}(\mathcal{A}^{IS}, n^*, c^*), T_A^{IT}(\mathcal{A}^{IT}, n^*)\right)$$

- **How to determine the scheduling** $(\mathcal{A}^{IS}, \mathcal{A}^{IT})$**?**

Simulation Results: Asynchronous Scenario



Simulation Results: Importance of *in situ* Resource Usage