



HAL
open science

Modeling HPC applications for in situ Analytics

Valentin Honoré, Brice Goglin, Guillaume Aupy, Bruno Raffin

► **To cite this version:**

Valentin Honoré, Brice Goglin, Guillaume Aupy, Bruno Raffin. Modeling HPC applications for in situ Analytics. IPDPS 2019 - 33rd IEEE International Parallel and Distributed Processing Symposium, May 2019, Rio de Janeiro, Brazil. hal-02567370

HAL Id: hal-02567370

<https://inria.hal.science/hal-02567370>

Submitted on 7 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Modeling HPC applications for *in situ* Analytics

Valentin Honoré¹ - Brice Goglin¹, Guillaume Aupy¹, Bruno Raffin²

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

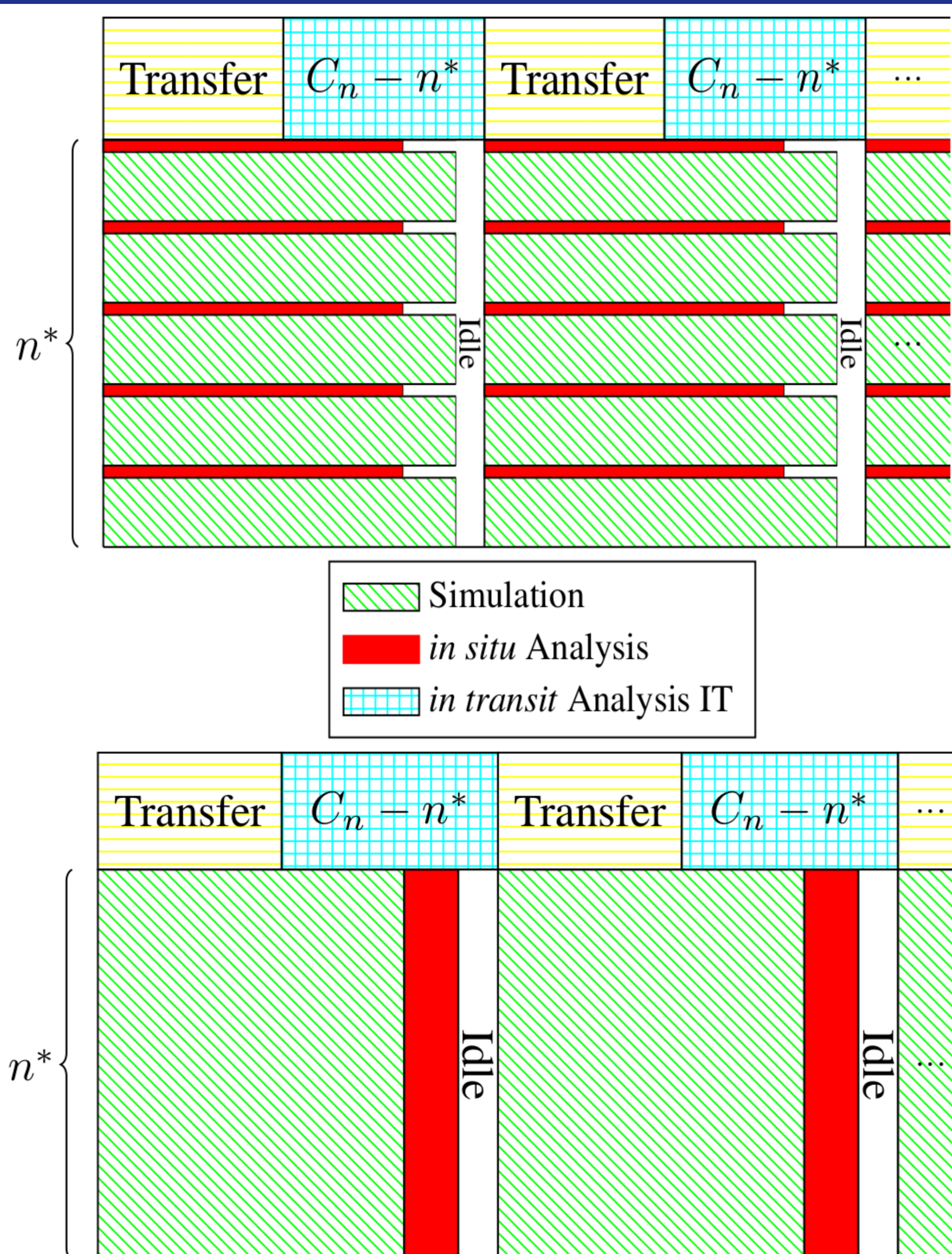


High Performance Computing



- ▶ **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

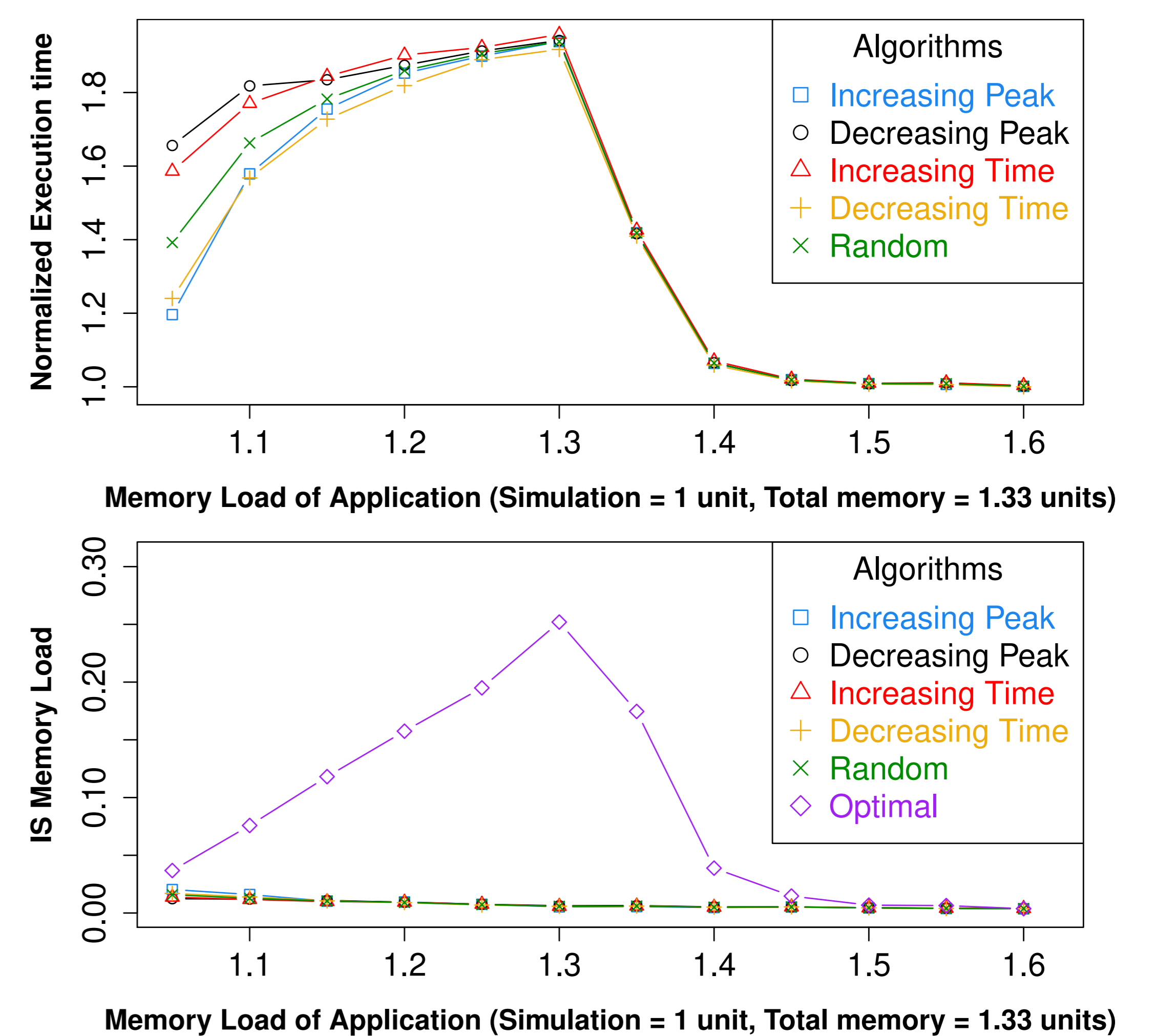
- ▶ **General model for HPC applications** (simulation, analysis, target architecture ...)
- ▶ **New algorithms** for
 - ▷ resource partitioning problem (theoretical analysis of the model)
 - ▷ scheduling problem (optimal non-polynomial + greedy)
- ▶ **Evaluation on synthetic applications:** extract 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

$$\max(T_S(n^*, c^*), T_A^{IS}(\mathcal{A}^{IS}, n^*, c^*), T_A^{IT}(\mathcal{A}^{IT}, n^*))$$
- ▶ **How to determine the scheduling** $(\mathcal{A}^{IS}, \mathcal{A}^{IT})$?

Simulation Results: Asynchronous Scenario



Simulation Results: Importance of *in situ* Resource Usage

