

---

# LightAMR standardized data structure and associated lossless compression algorithms: RAMSES datasets use cases.

Loic Strafella\*<sup>1</sup>

<sup>1</sup>Institut de Recherches sur les lois Fondamentales de l'Univers – Commissariat à l'énergie atomique et aux énergies alternatives : DRF/IRFU/DEDIP/LILAS – France

## Abstract

### Résumé:

Les sorties du code RAMSES sont aujourd'hui uniquement de type protection/reprise et sont principalement basées sur un format binaire dédié. Chaque snapshot d'une simulation peut occuper plusieurs téraoctets de données répartis sur plusieurs dizaines de milliers de fichiers pénalisant la charge du réseau et le système de fichiers. De plus, bien que le format de protection/reprise permette de relancer une simulation en cas d'arrêt impromptu, il nécessite une conversion coûteuse vers un format de "dépouillement", plus léger, pour la visualisation et l'analyse des données à l'aide d'outils spécifiques. Les objectifs de ces travaux sont d'intégrer dans RAMSES, un format standard pour la description de la grille AMR et ses données physiques associées tout en réduisant significativement le volume total de données. Nous présenterons le nouveau format LightAMR ainsi que les algorithmes de réduction de redondance et de compression sans perte associés à ce nouveau format.

### Abstract:

As of today, RAMSES code outputs are checkpoint/restart outputs only, written as binary files in a specific format. Every simulation snapshot can be very large up to terabytes, split on dozens of thousands of files, thus hindering the network and filesystem load resulting in large I/O times and in a poor scalability of RAMSES code in the exascale context. Moreover, even if the checkpoint/restart file format contains every information required by the code in case the simulation needs to be restarted upon failure, it needs to be post-processed to a lightweight format before any visualization or data analysis by dedicated tools. The main objectives of this work are to integrate in RAMSES a new standardized AMR format for the description of the AMR mesh as well as the physical quantities in order to significantly reduce the I/O volume while keeping the hierarchy description of the mesh. The new lightAMR format will be presented as well as the associated tree pruning algorithm that reduce the redundancy in RAMSES datasets and the lossless compression algorithm for mesh description and floating point data.

---

\*Speaker