

# A High Performant Computing data format generator

3rd OBELICS workshop

Thomas Vuillaume, on behalf of Pierre Aubert\* and the LAPP team

## Advantages of a data format generator:

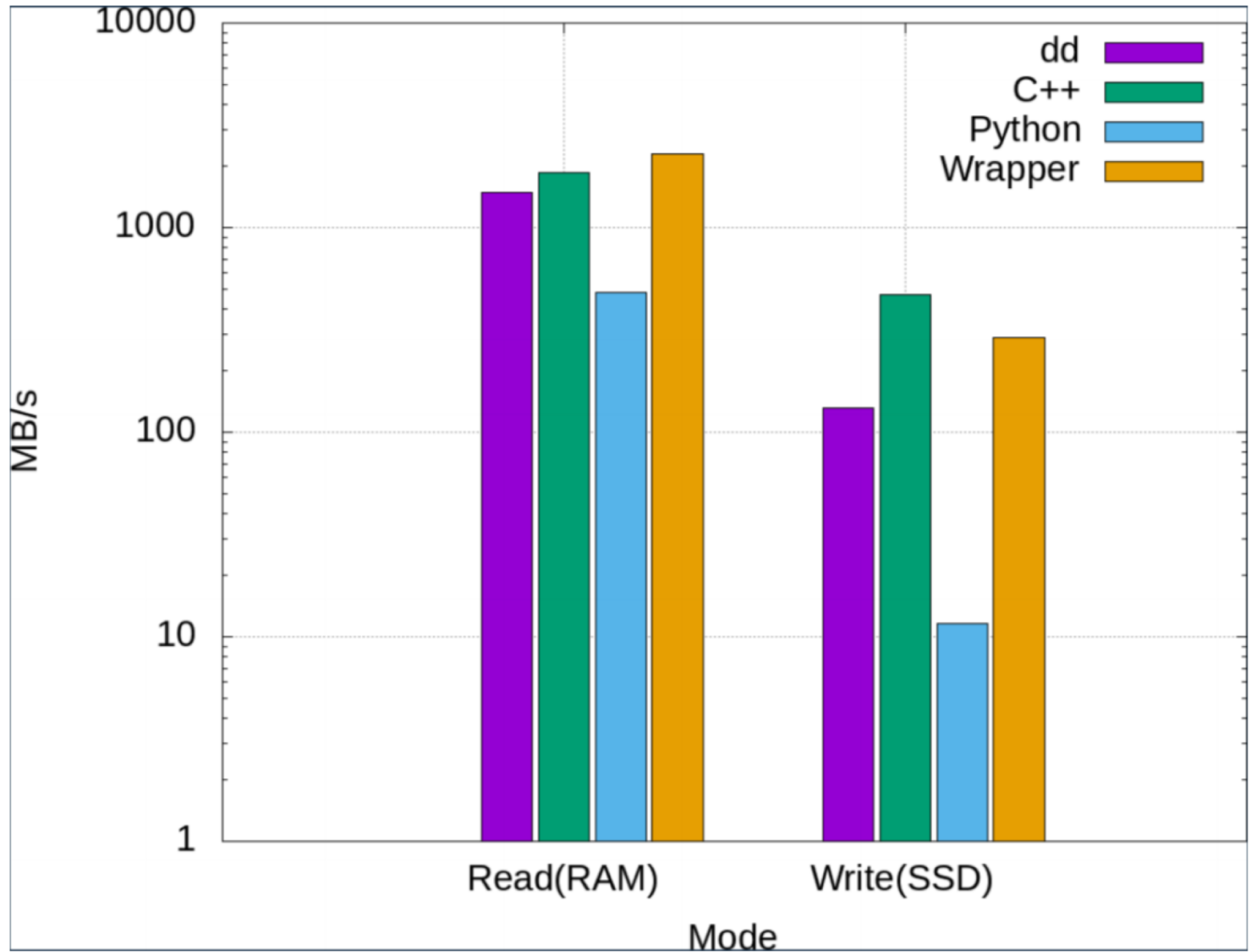
- Save development time
- Handle repetitive task
- Generate library
  - Copy
  - Read/Write Binary Files

Advantages of [this](#) data format generator:

- Save development time
- Handle repetitive task
- Generate library
  - Copy
  - Read/Write Binary Files
  - In C++ and Python
  - Full/Block Compression → see talk in D-INT
- Version compatibility
- HPC compliant

## HPC compliant ?

- Data at reading is guaranteed to be:
  - memory aligned
  - contiguous
- 2D or 3D matrices: 1st element of each row has a memory address modulo the length of the vectoriel register



- Performances increase during data analysis without changing any analysis code  
→ this is because contiguous data allows CPU data prefetching

	Speed (cy/el)	Speed Up
H.E.S.S. root data format	2126	1
Generated data format	53	40

- Generator delivered as open-source library  
[https://gitlab.in2p3.fr/CTA-LAPP/PLIBS\\_8](https://gitlab.in2p3.fr/CTA-LAPP/PLIBS_8)
- Already in use in IN2P3 laboratories
- Format generated for CTA to be benchmarked versus other formats



H2020-Astronomy ESFRI and Research Infrastructure Cluster (Grant Agreement number: 653477).