Barcelona Supercomputing Center
Centro Nacional de Supercomputación

EXCELENCIA SEVERO OCHOA

Improving the throughput of Earth System Models using an asynchronous parallel I/O server

Xavier Yepes-Arbós

17/04/2018

8th JLESC Workshop, Barcelona

Index

- Introduction
- Motivation
- I/O overview for Earth system models
- Case study: IFS-XIOS
- Evaluation
- Conclusions

Introduction

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Introduction

- Scientific applications have benefited of the exponential growth of supercomputing power
- This allows to use more complex computational models to find more accurate solutions
- However, this implies to generate a huge amount of data to meticulously represent accurate solutions

Introduction

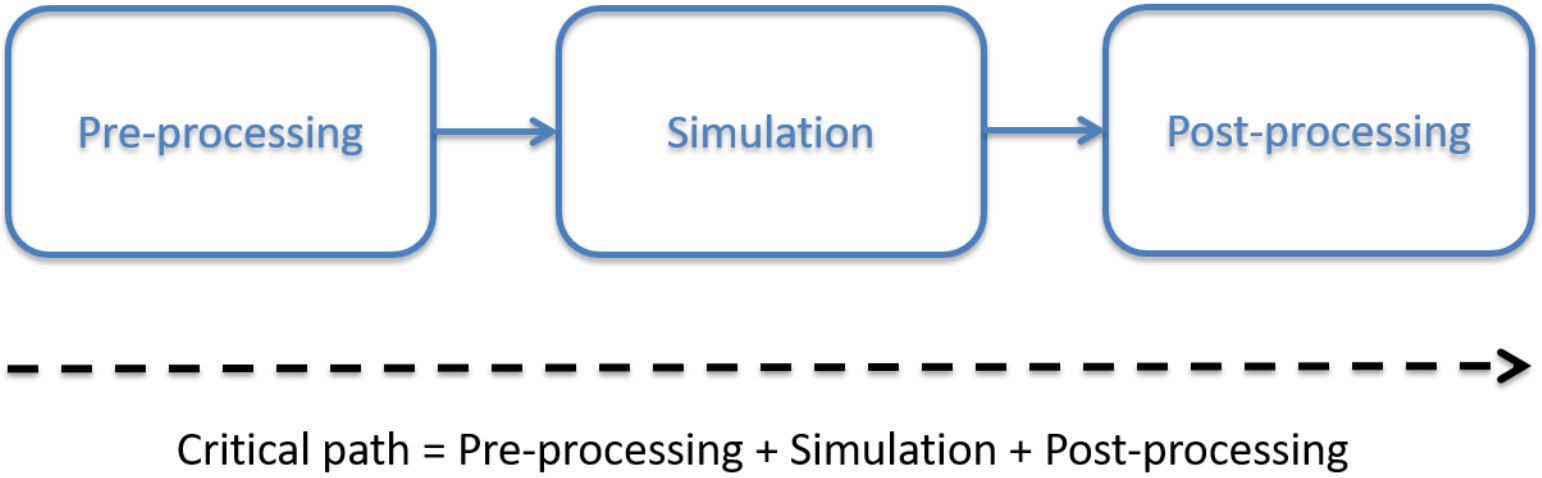
- Earth system models are a particular case that solve complex models and generate a lot of data
- More accurate forecasts, predictions and projections through higher grid resolutions
- This implies to use efficient HPC techniques, such as optimized MPI communications
- However, the I/O part has almost been forgotten, since it was not significant enough in the past

Introduction

- Some Earth system models output data using inefficient sequential I/O schemes
- This type of scheme requires a serial process:
  - Gather all data in the master process of the model
  - Then, the master process sequentially writes all data
- This is not scalable for higher grid resolutions, and even less, for future exascale machines

Introduction

- In addition, Earth system models run experiments that have other tasks in their workflow
- Post-processing task can perform data format conversion, compression, diagnostics, etc.



Motivation

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Motivation

- In particular, we experience this problem with EC-Earth, a coupled climate model
- EC-Earth has started to run ultra-high resolution experiments under the H2020 PRIMAVERA project
- However, it suffers a considerable slowdown, where the I/O represents about 30% of the total execution time

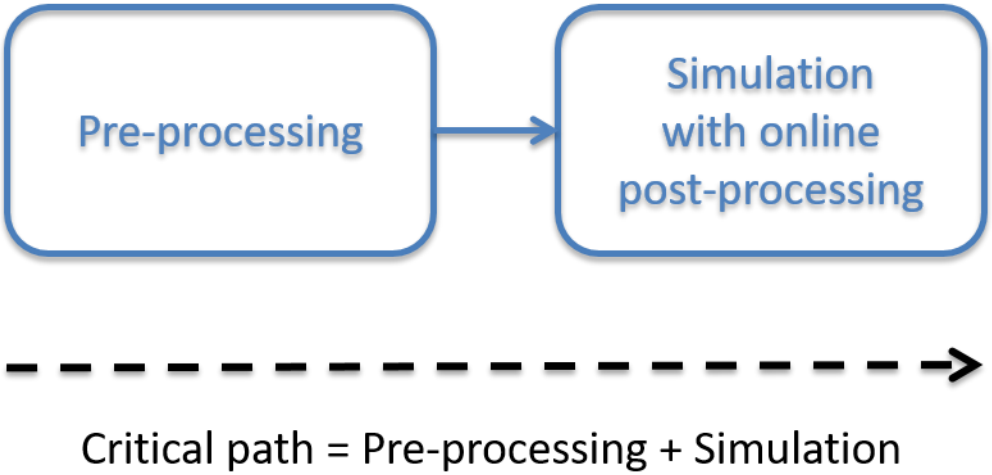
Motivation

- In order to address the I/O issue, we have to select a suitable tool that fulfills a series of needs:
  1. It must be a parallel, efficient and scalable I/O tool
  2. Data must be written using netCDF format and must follow the CMIP standard
  3. It must perform online post-processing along with the simulation, such as interpolations or data compression
- There is a tool designed to that end: XIOS
- XIOS is an I/O server

Motivation

Use a tool such as XIOS, has a twofold effect:

- Improve the computational performance and efficiency of a model, and thus, reduce the execution time
- Reduce the critical path of its workflow by avoiding the post-processing task

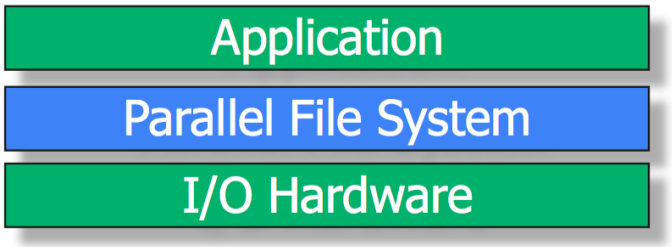


I/O overview for Earth system models

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Sequential I/O

- In parallel applications there are two approaches:
  - Each process outputs its own local data
  - Data is gathered by the master process, which sequentially writes the whole global data
- It is typically done using the POSIX I/O API: open, write, close, etc.

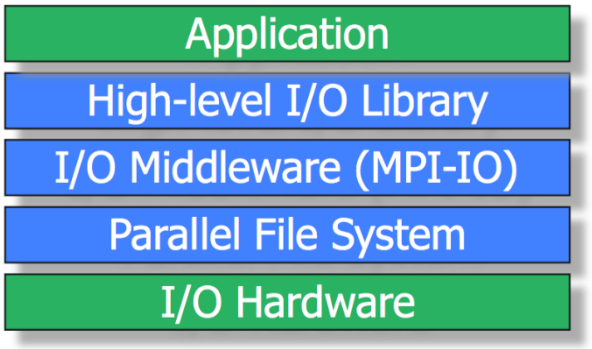


Parallel I/O

- In order to make the I/O scalable, it is typically used parallel I/O
- Parallel I/O is the ability to perform multiple input/output operations at the same time, such as:
  - Simultaneously write several files
  - Concurrently write into different regions of the same file from different processes
- Different writing modes: multiple file, one file and intermediate approach

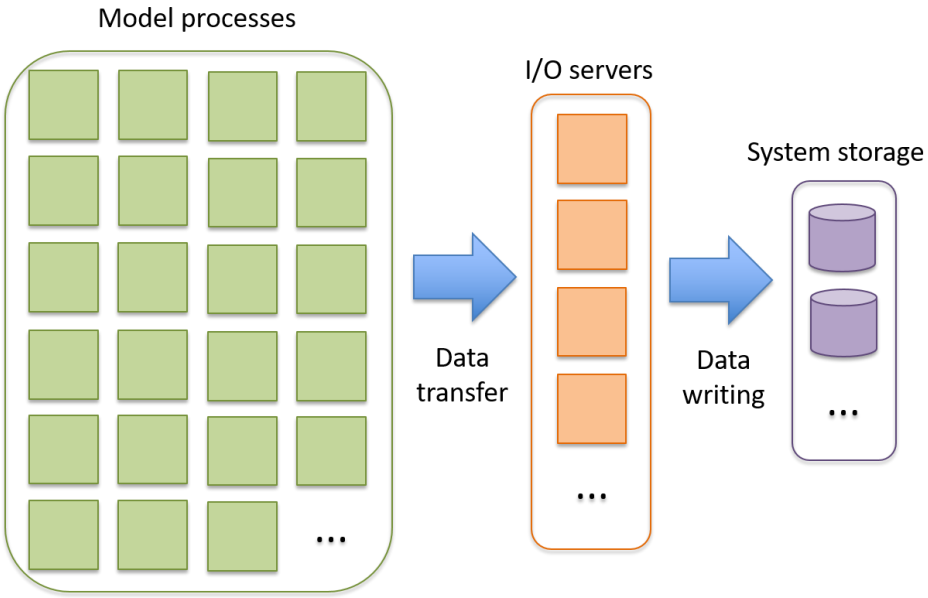
Parallel I/O

- POSIX cannot offer the possibility to implement parallel I/O
- Built upon MPI-IO for the I/O Middleware layer
- Built upon netCDF or HDF5 for the High-level layer



I/O servers

- I/O servers are exclusively dedicated resources to perform input/output
- Model processes do not deal with the I/O, so they can continue with the simulation
- Disk latency is hidden



Case study: IFS-XIOS

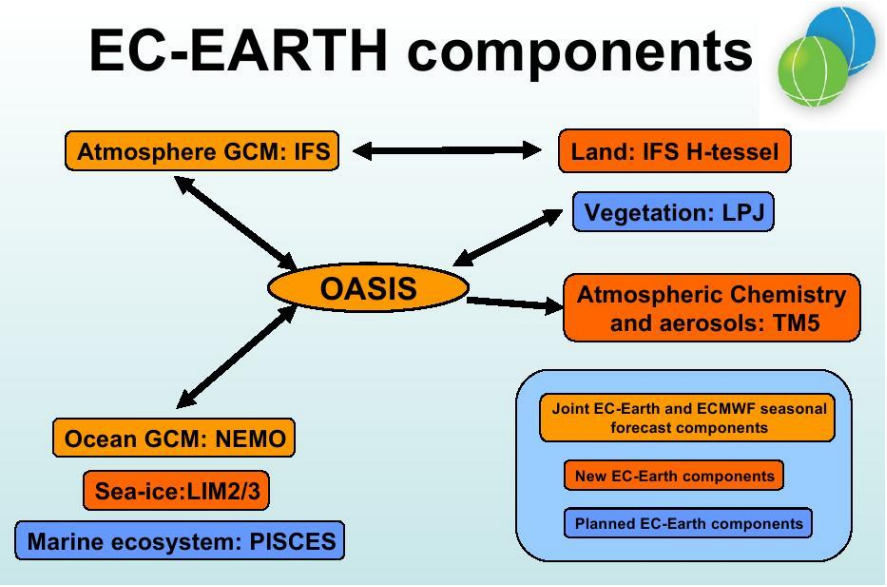
Barcelona Supercomputing Center
Centro Nacional de Supercomputación

IFS

- The Integrated Forecast System (IFS) is a global data assimilation and forecasting system developed by ECMWF
- It uses an inefficient sequential I/O scheme
- It writes using the GRIB format (weather forecast)
- When IFS is used for climate modeling, post-processing is needed to:
  - Convert GRIB to netCDF files
  - Transform data to be CMIP-compliant
  - Compute diagnostics

EC-Earth

- EC-Earth is a global coupled climate model, which integrates a number of component models in order to simulate the Earth system
- The two main components are **IFS as the atmospheric model** and NEMO as the ocean model



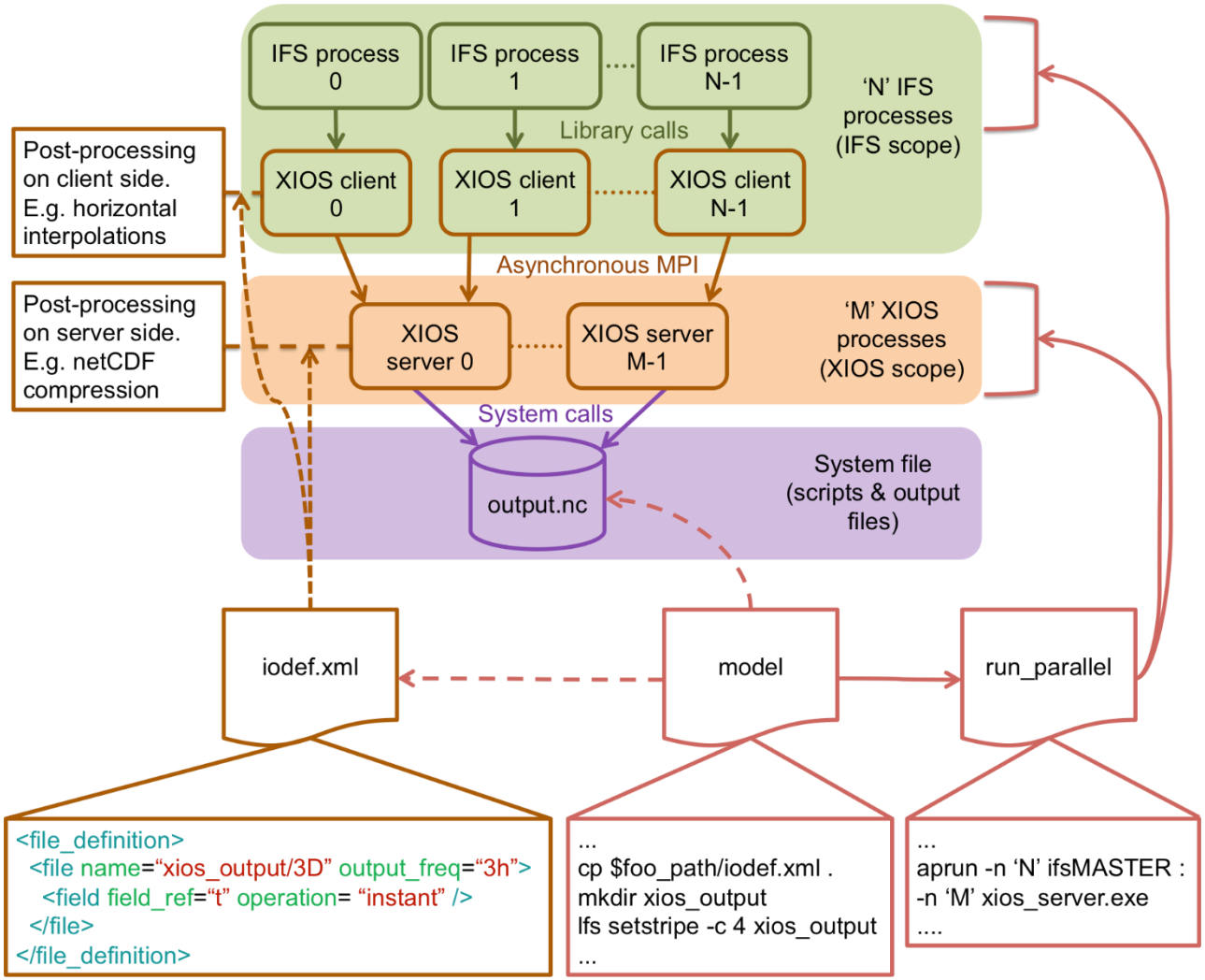
XIOS

- The XML Input/Output Server (XIOS) is an asynchronous MPI parallel I/O server
- It writes using the netCDF format
- Written data is CMIP-compliant
- It is able to post-process data online to generate diagnostics

IFS-XIOS integration

- Integrate IFS with XIOS
- Analyze and optimize the integration
- The goal is to improve the IFS performance to get benefit of the XIOS features in EC-Earth:
  - Reduce the IFS execution time
  - Avoid the post-processing task

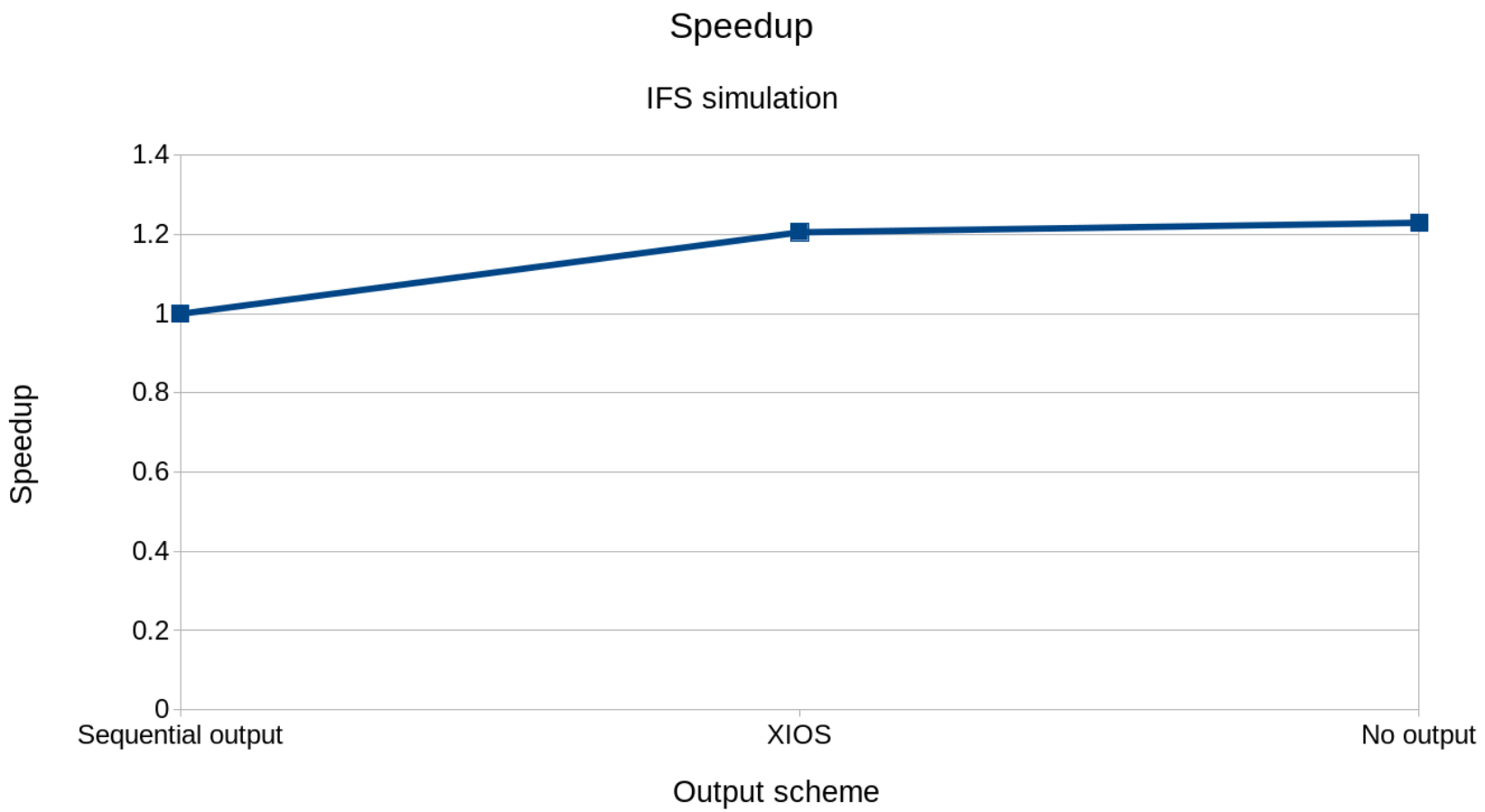
IFS-XIOS integration



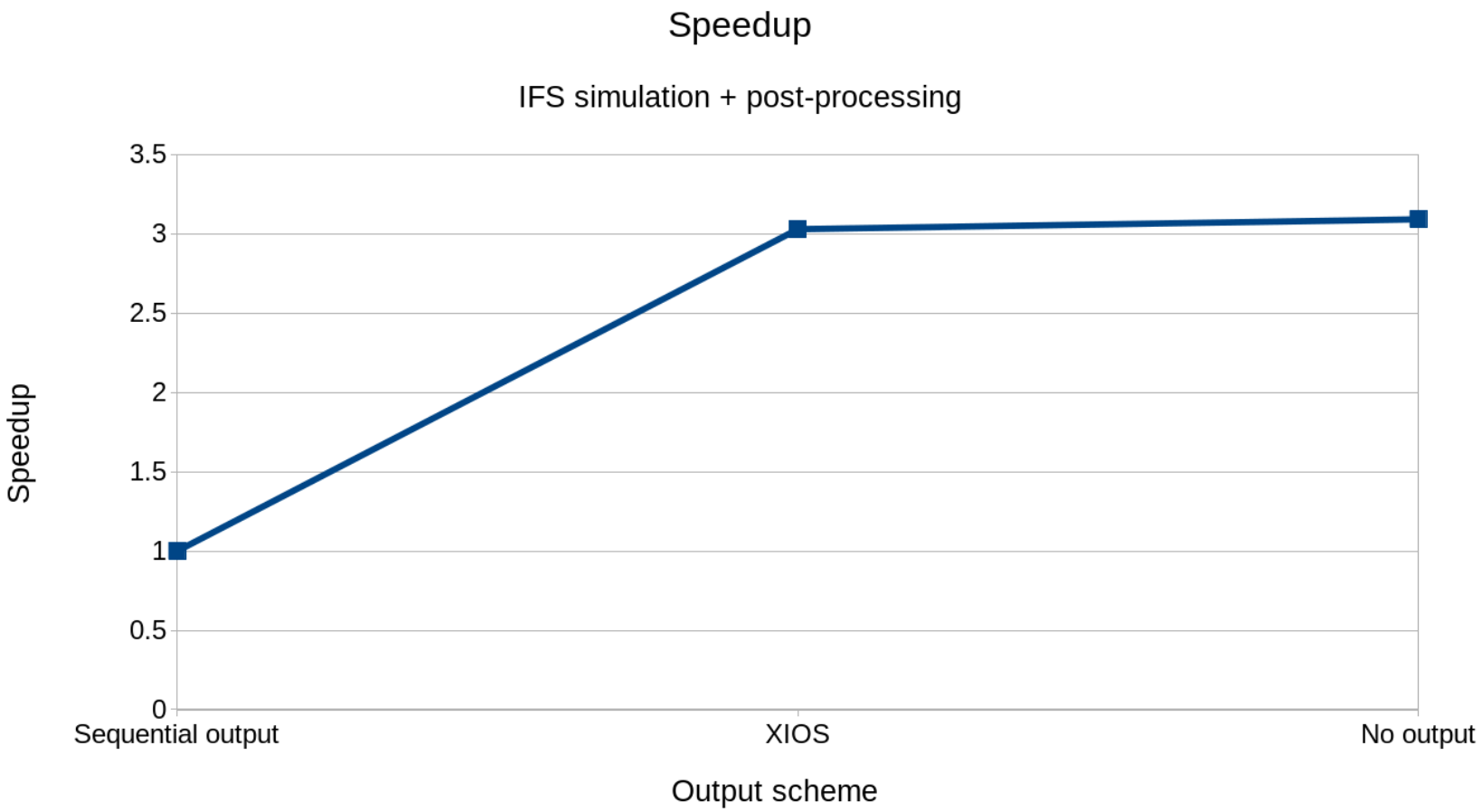
Evaluation

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

IFS simulation



IFS simulation + post-processing



Discussion

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Discussion

- Is there any other tool similar to XIOS? If so, is it more efficient?
- Do you think there are other parallel I/O techniques that are more suitable than XIOS?
- Are I/O servers a good approach to address the I/O bottleneck for future exascale machines?

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

EXCELENCIA SEVERO OCHOA

Thank you

xavier.yepes@bsc.es