Mapping a Coupled Earth-System Simulator onto the Modular Supercomputing Architecture

Sam Hatfield[1], Olivier Marsden[1], Kristian Mogensen[1], Ioan Hadade[1], Mathieu Stoffel[2]

1: European Centre for Medium-Range Weather Forecasts, 2: Eviden; samuel.hatfield@ecmwf.int

The Modular Supercomputing Architecture, or *MSA*

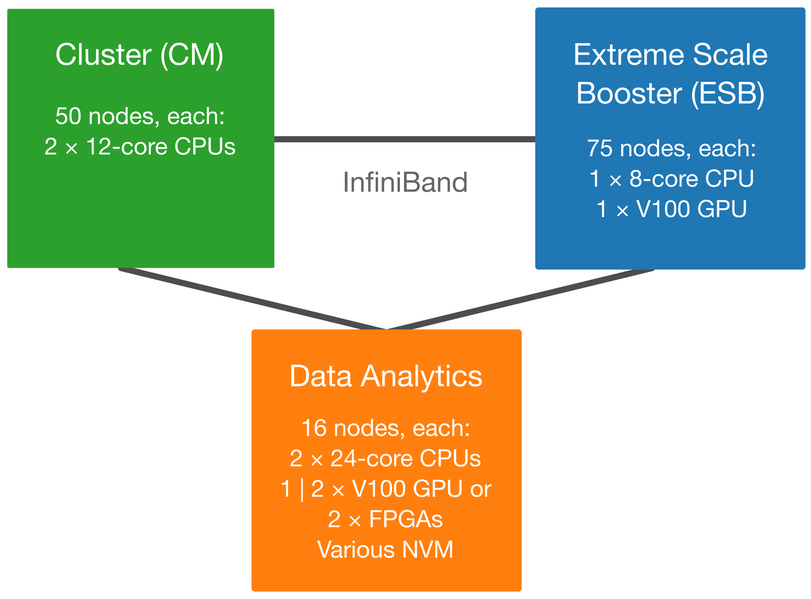
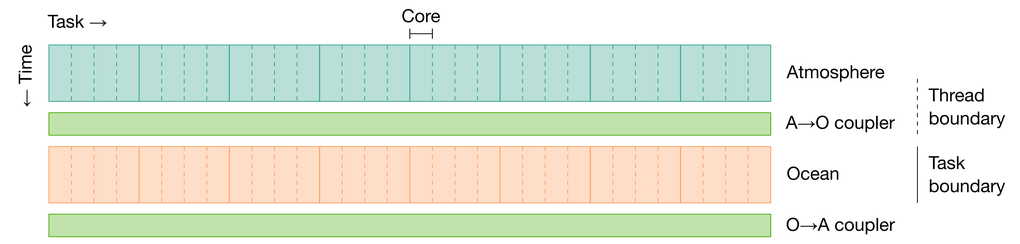


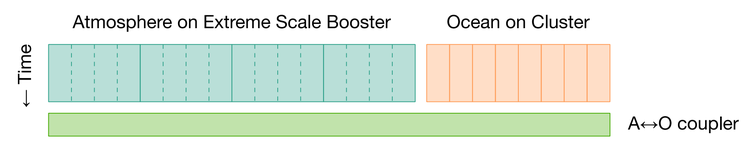
Figure 1: The DEEP Modular Supercomputing Architecture prototype at the Jülich Supercomputing Centre.

- The **Modular Supercomputing Architecture (MSA)** is a style of heterogeneous computing platform composed of multiple "modules" (i.e. partitions) with a **common interconnect**
- One module might provide **general purpose CPU-based compute ("Cluster")** while another may expose **GPUs for targeted acceleration ("Booster")**
- Applications can be allocated nodes **across multiple modules**
- The DEEP project series has developed a **prototype MSA** based at the Jülich Supercomputing Centre (Figure 1)
- Here we discuss efforts under the DEEP-SEA project to exploit the MSA with the ECMWF coupled Earth-system model: the IFS

Mapping the ECMWF coupled model, the IFS, to the MSA



(a) Old: tightly coupled



(b) New: loosely coupled

Figure 2: Two approaches for coupled atmosphere-ocean simulations with the IFS.

- The IFS is composed of **multiple components** for representing the Earth system, notably the **atmosphere** and **ocean**
- We will run these two components on **separate modules, chosen to match their computational characteristics**
- The atmospheric part of IFS can now run partially on GPUs (namely, the spectral transform kernel) — the atmosphere should therefore be pinned to the Extreme Scale Booster module (ESB)
- The ocean part (the NEMO model) does not support GPUs, however, so this should be pinned to the Cluster module (CM)
- To allow this it was necessary to move from a **tightly coupled** approach, where the atmosphere and ocean share MPI tasks, to a **loosely coupled approach**, where the two execute on separate MPI tasks with their own communicators (Figure 2)
- Now, the atmosphere and ocean compute can take place **concurrently on separate modules**

Concurrent atmosphere-ocean simulations on the MSA

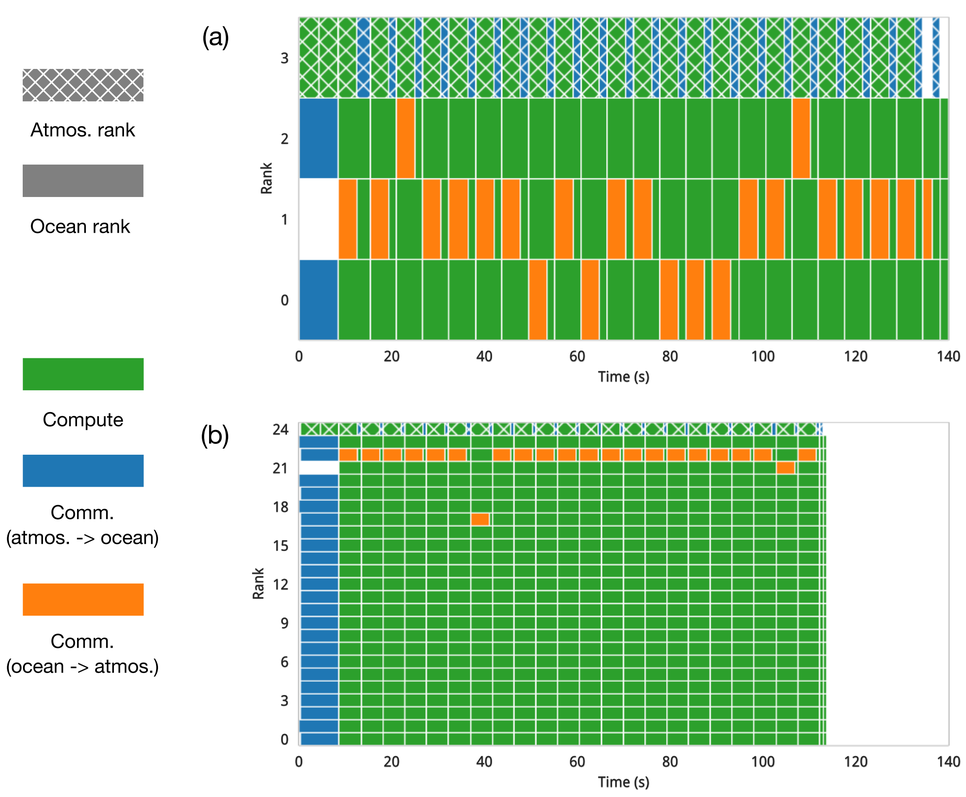


Figure 3: Traces of heterogeneous IFS runs on the DEEP testsystem.

- Figure 3 shows traces of two heterogeneous (CPU-only) coupled IFS runs on the DEEP testsystem with 1 CM node (atmosphere) and 1 ESB module node (ocean)
- In (a) we use the same **MPI+OpenMP shape** for the atmosphere and the ocean
- In (b) we take advantage of the flexibility of the new approach by **running the ocean in pure MPI mode**
- The NEMO ocean model is known to perform better in pure MPI mode so this delivers a **20% acceleration** in overall performance

Energy optimisation on the DEEP testsystem

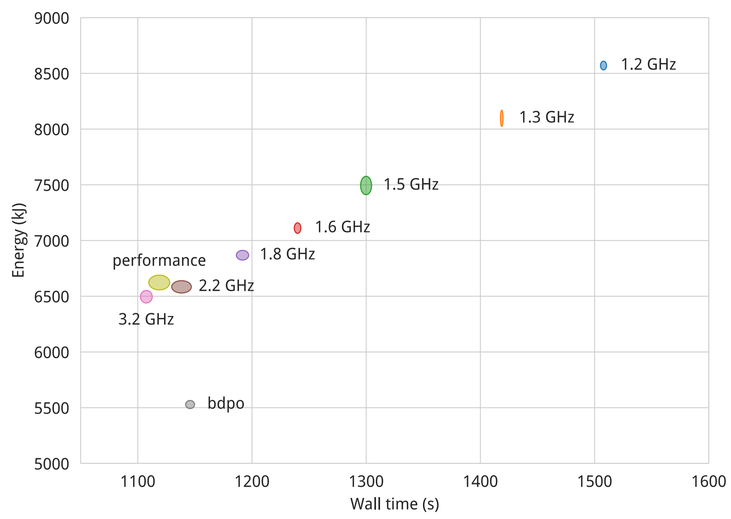


Figure 4: An energy consumption and wall time benchmark of ecTrans with several configurations. "performance" is the reference, relying on the CPU's performance governor.

- We have also investigated tools for **measuring** and **tuning** the **energy consumption** of applications on the DEEP testsystem
- We focused on the **Bull Dynamic Power Optimiser (BDPO)**
- This identifies periods of **low CPU activity** while an application is running and **dynamically scales down the CPU frequency** in order to **reduce power consumption**
- Any application that isn't **purely compute-bound** could benefit
- Figure 4 shows the **energy consumption** and **wall time** of a standard benchmark of the IFS spectral transform module, **ecTrans**, performed on 16 nodes of DEEP
- Each datapoint is averaged over 10 repetitions
- By enabling BDPO as a Slurm submission option we can **reduce the energy consumption of ecTrans by 16%** with only **3% increase in wall time**
- This compares favourably with the naive approach of uniformly decreasing CPU frequency which **increases both wall time and energy**

This project has received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 955606. The JU receives support from the European Union's Horizon 2020 research and innovation programme and Germany, France, Spain, Greece, Belgium, Sweden, United Kingdom, Switzerland.