Scaling IFS to 1 km and beyond

21st edition of the ECMWF workshop on high-performance computing in meteorology

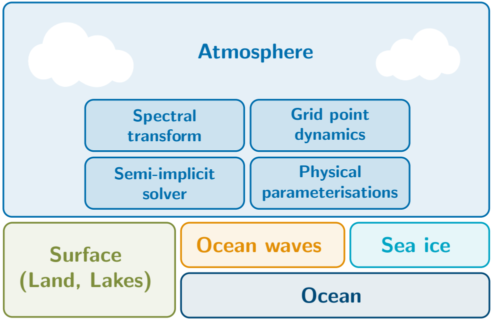
Ioan Hadade

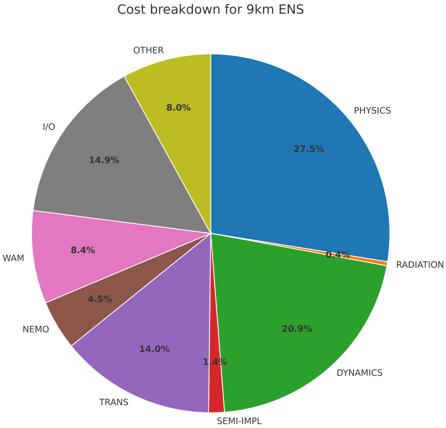
ioan.hadade@ecmwf.int

European Centre for Medium-Range Weather Forecasts

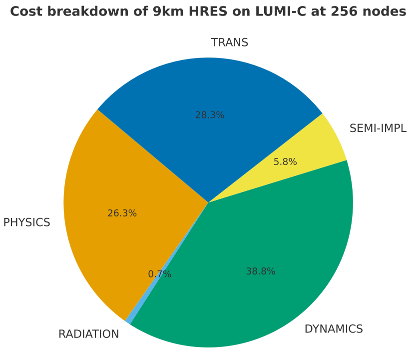
ECMWF

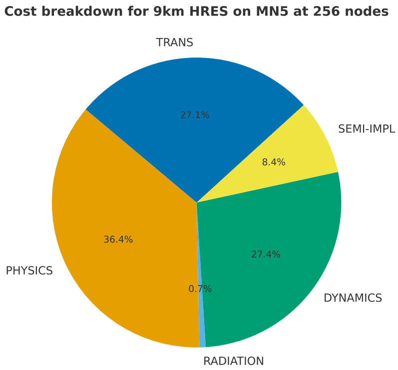
Integrated Forecasting System (IFS)



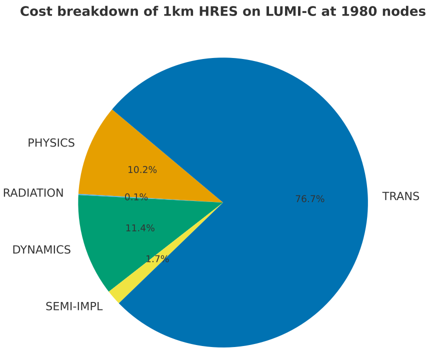


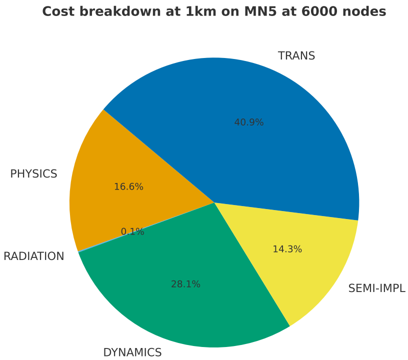
Breakdown of cost for 9 km HRES





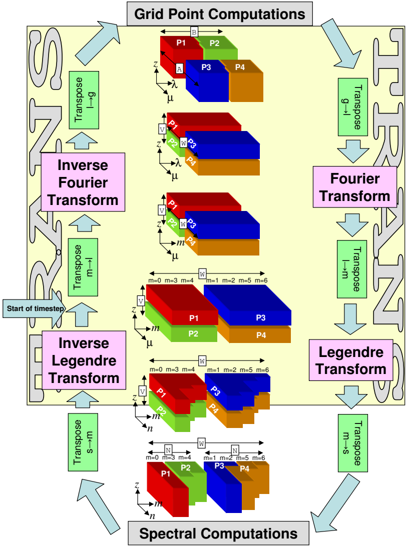
Breakdown of cost for 1 km HRES



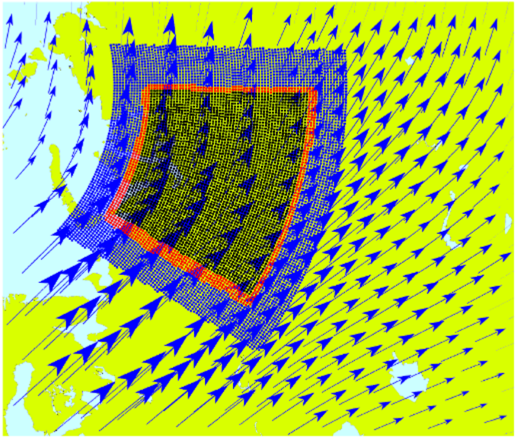


Spectral transforms and semi-implicit

- one large direct and inverse transform per timestep
- main scalability bottleneck at high resolution and high node counts
- 1D FFTs in Fourier space and GEMMs in spectral space
- Point-to-point communication and AlltoAllv communications



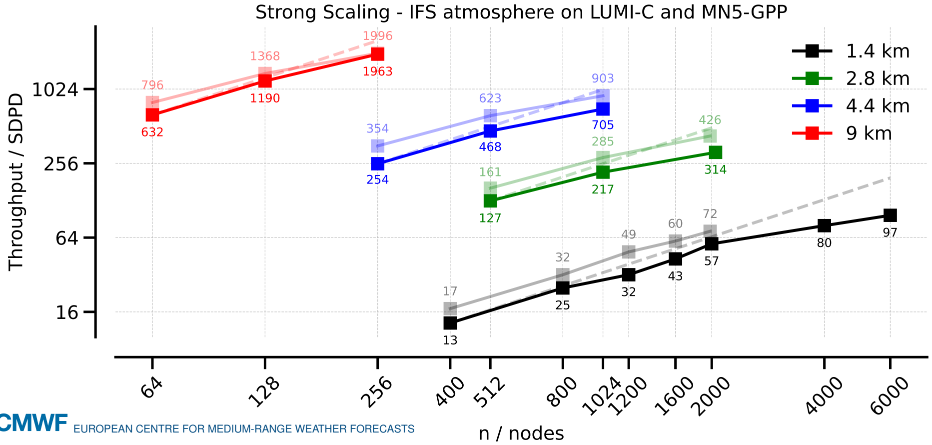
Semi-Lagrangian advection



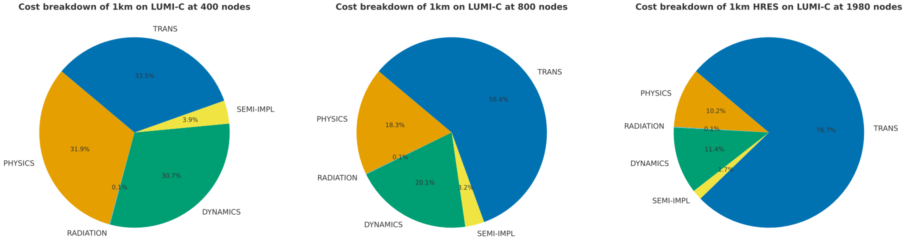
- Main source of communications in the dynamics
- Non-blocking point-to-point
- Halo-width (blue) assumes wind speeds much larger than observed in atmosphere
- Two-phase communication:
  - Entire blue halo filled for departure point computation (SLCOMM)
  - Once departure point is known, only the red halo is filled with the points around departure point (SLCOMM2A)

IFS atmosphere strong scaling from 9 km to 1 km

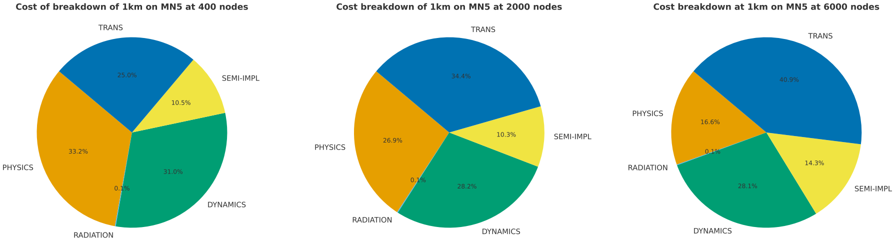
- Very good strong scaling up to 2.8 km
- Scalability drops at 1 km once the 2000 node count boundary is crossed



Breakdown of cost for IFS 1 km strong scaling on LUMI-C

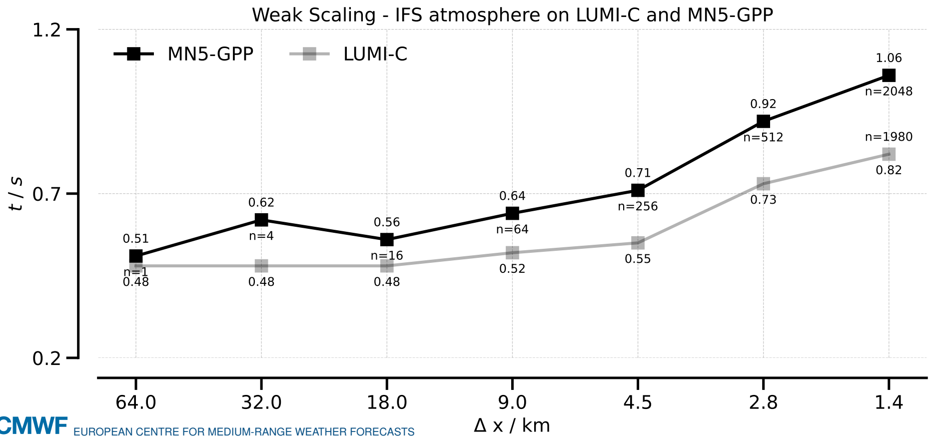


Breakdown of cost for IFS 1 km strong scaling on MN5

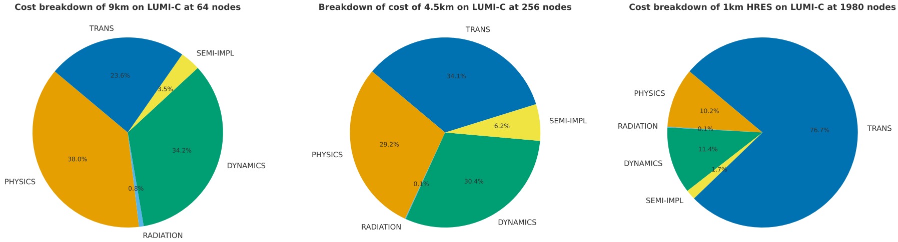


IFS atmosphere weak scaling

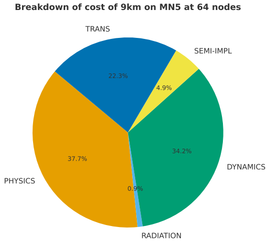
- Very good weak scaling up to 4.5 km (90% on LUMI-C and 72% on MN5)
- Drops down when going to 1 km (60% on LUMI-C and 49% on MN5)

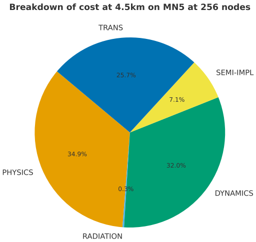


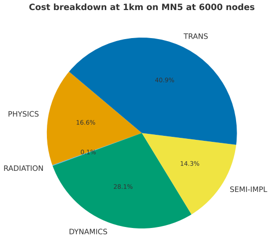
IFS atmosphere weak scaling on LUMI-C



IFS atmosphere weak scaling on MN5

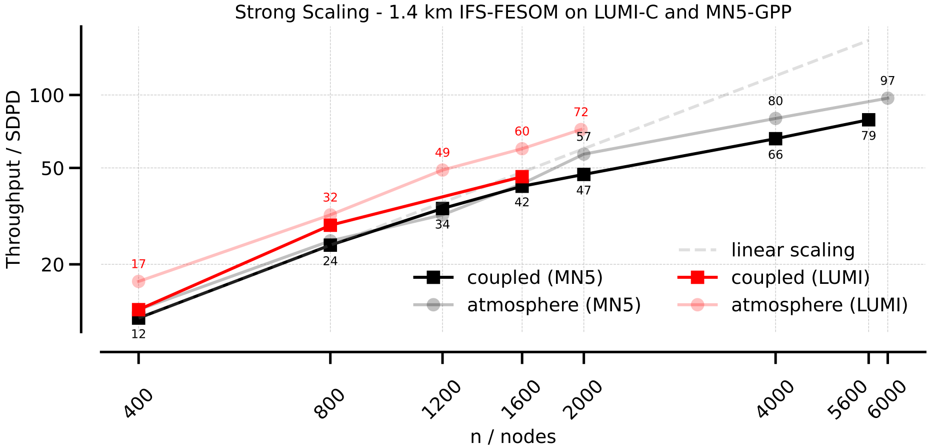






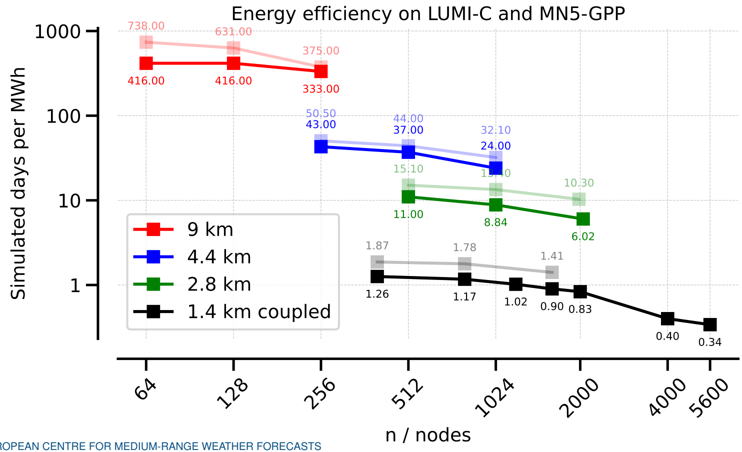
IFS coupled strong scaling at 1 km

- 1 km IFS coupled with 1 km FESOM2 ocean



IFS energy efficiency

- Unsurprisingly, best energy efficiency is obtained at the lower end of the scaling curve



Can spectral methods on the sphere live for ever?

By **Clive Temperton**

European Centre for Medium-Range Weather Forecasts

3. Problems at high resolution

In the early days of massively parallel processing, it was widely held that spectral methods would run inefficiently on MPP architectures because of their need for global communication. This proved to be much less of a problem than had been feared; relatively simple communica-

tion strategies have proved adequate to ensure efficient scaling on up to 1000 processors (*Barros et al.*, 1995).

A more serious problem is that for a horizontal spectral truncation at total wavenumber N the number of gridpoint computations per timestep scales as $O(N^2)$, the work involved in Fourier transforms scales as $O(N^2 \log N)$, while the computation of Legendre transforms scales as $O(N^3)$. Thus, it is argued that *eventually* the Legendre transforms will dominate the computation and each doubling of the horizontal resolution will result in an eightfold increase in the computational work per timestep (compared with a fourfold increase for a gridpoint model).

To put this problem in perspective, Table 1 shows the cost of the transforms as a percentage of the work per timestep at the current operational resolution of the ECMWF model, and as measured at two experimental higher resolutions. It is clear that although the relative cost of the Legendre transforms is growing as expected, it will remain bearable beyond T1000. Nevertheless, for horizontal resolutions which might be contemplated in the more distant future, it is evident that the current approach cannot be sustained for ever.

TABLE 1. COST OF TRANSFORMS AS PERCENTAGE OF MODEL TIMESTEP

	Fourier	**Legendre**
T319	2%	9%
T511	2%	12%
T799	2%	15%

Beyond 1 km and next steps

- First IFS 700m simulation performed on Leonardo Booster by Nils Wedi (30s per timestep on 1024 GPU nodes)
- We will explore this further on JUPITER in both hydrostatic and non-hydrostatic configurations
- This is of interest from both technical and scientific perspective
- At the same time, we will try to reach 1 SYPD at 1 km
- To achieve this, we clearly need to focus more on improving the communication schemes in the IFS
- Work in this area is already happening:
  - Overlap computation and communication (with Rich Graham and his team at NVIDIA)
  - Single-sided communications in semi-Lagrangian advection and semi-implicit solver