WGNE Computer table

WGNE meeting “in the cloud”, 3 Nov. 2020

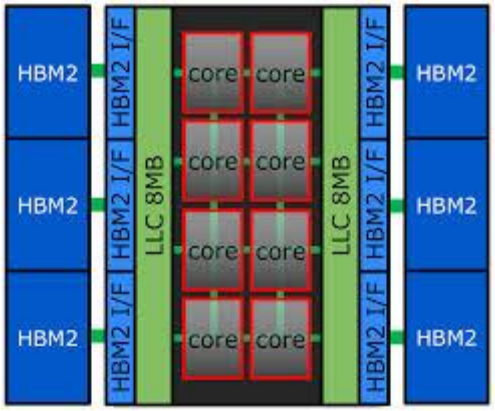
Günther Zängl

Overview

- DWD's experience with the migration to NEC Aurora
- News from other centers
- Discussion items

DWD's migration experience

Overview of our system – a soft flavour of „heterogeneous“



- A computing node consists of
  - a vector host: 24-core AMD Rome
    (2.8 GHz; 512 GB memory)
  - 8 vector engines: SX-Aurora 1 TSUBASA Typ 10AE
  - Every vector engine has
    - 8 cores (1.584 GHz; 304.1 GF/s (DP); 608.3 GF/s (SP) per core)
    - 48 GB HBM2 3D-stacked memory (6 GB/core; 1.35 TB/s)
    - and is direct liquid cooled
- Number of nodes / engines

Phase	Operations	Experiments
0	178 / 1424	232 / 1856
1	224 / 1792	292 / 2336
2	224+101/3408	292+132 / 4448

Why did we decide for such an ‚exotic' architecture?

- NEC made the best offer based on our procurement benchmarks (with quite some lead over the competitors)
- Energy efficiency (taking ICON with current OpenACC port as a reference) is comparable to GPUs
- However, the amount of work needed for porting the ICON code to the NEC was at least two orders of magnitude less than for the GPU port

DWD's migration experience

- Installation of the machine was delayed by 2 months … because the vector host CPUs from AMD arrived late
- In addition, the migration phase took 2 months longer than calculated because of compiler bugs, incomplete batch system functionalities and hardware instabilities (officially, the covid pandemic was blamed, but I don't think fixing these issues would have progressed faster otherwise)
- Yet, we are not the only ones facing this problem (see below)

- Porting the data assimilation code was hampered by numerous compiler bugs which were difficult to isolate
- (But again, we should not grumble too much – the latest Cray compiler version is unable to translate the ICON code correctly)
- Still not satisfying: hardware stability (repeated failures of filesystem access on individual nodes, Infiniband issues slowing down communication speed) and stability of batch system

News from other centers

- Planned resolution upgrades for global system have been put into operation planned at Météo France and NRL
- NCEP: ‘NGGPS’ based on FV3 dycore is now also used for the ensemble prediction system
- ECMWF: start of operation of new HPC in Bologna is delayed even more than for our computer (current estimate Nov 2021)
- Very uncertain situation for strategic planning in Brazil

- Next major change at DWD: Replacement of COSMO-D2(-EPS) by ICON-D2(-EPS) for convection-permitting regional forecasts in February 2021

- For more details please see updated WGNE table at meteoinfo.ru

Possible discussion items

- **The trend is going towards seamless prediction, but ‚doing a little bit of everything' is not a useful strategy, neither in terms of development resources nor in terms of computing resources. How do you prioritize tasks and applications?**
- **How do you decide about the resource partitioning between deterministic and ensemble runs?**