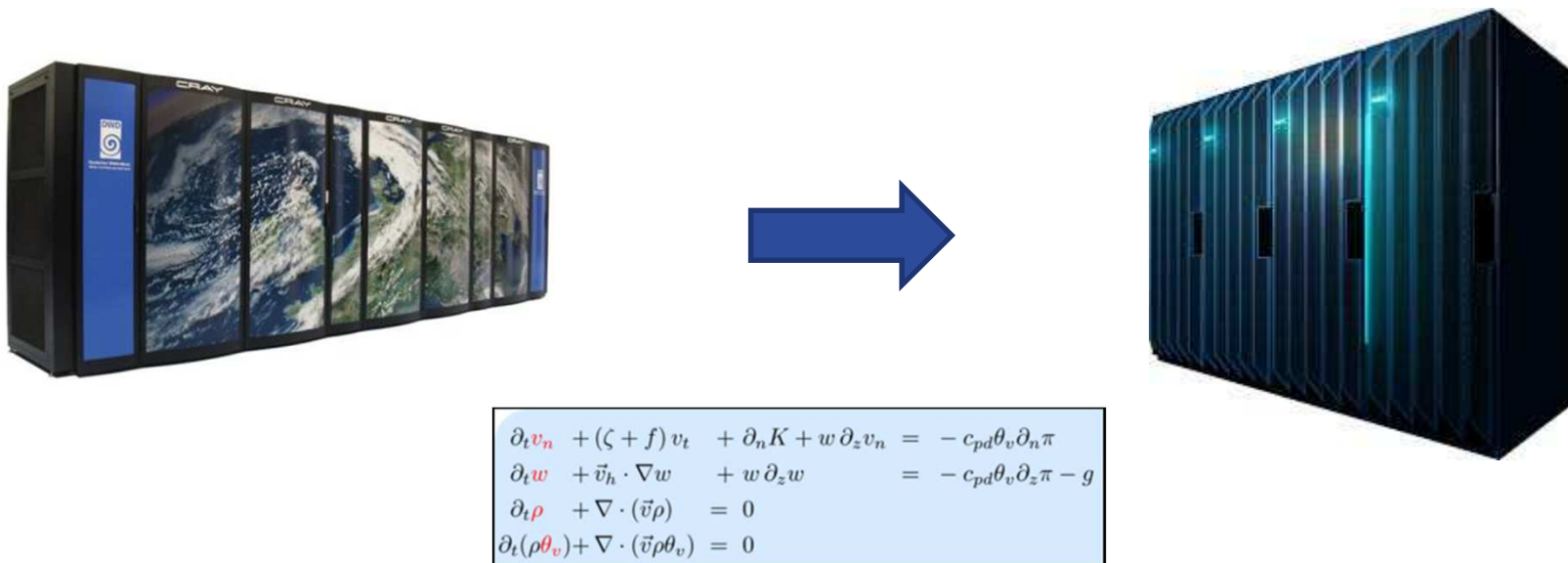


New HPC System at DWD



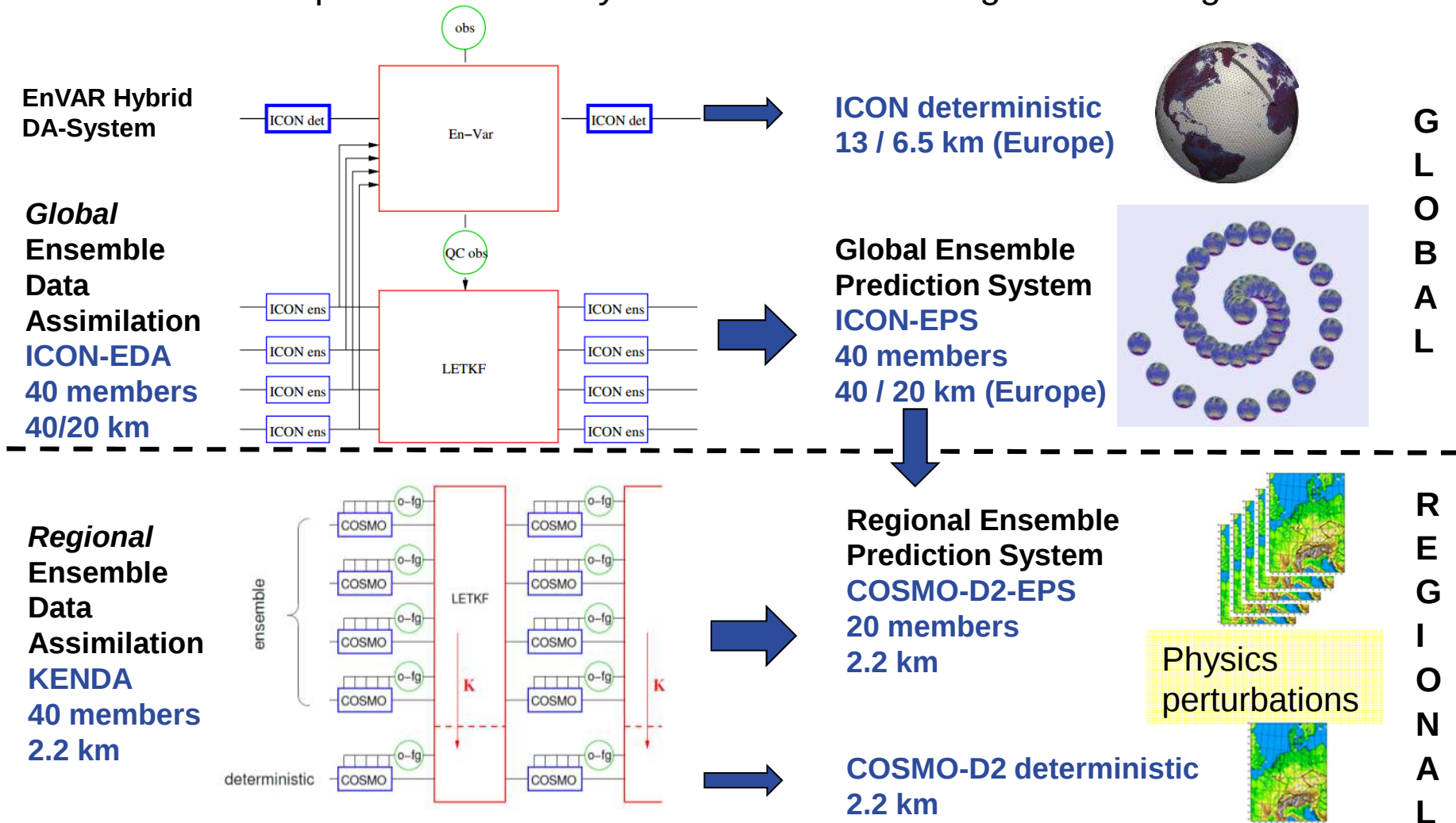
Ulrich Schättler

Department for Numerical Modeling (FE1)

most slides courtesy of Detlev Majewski, Henning Weber (and colleagues)

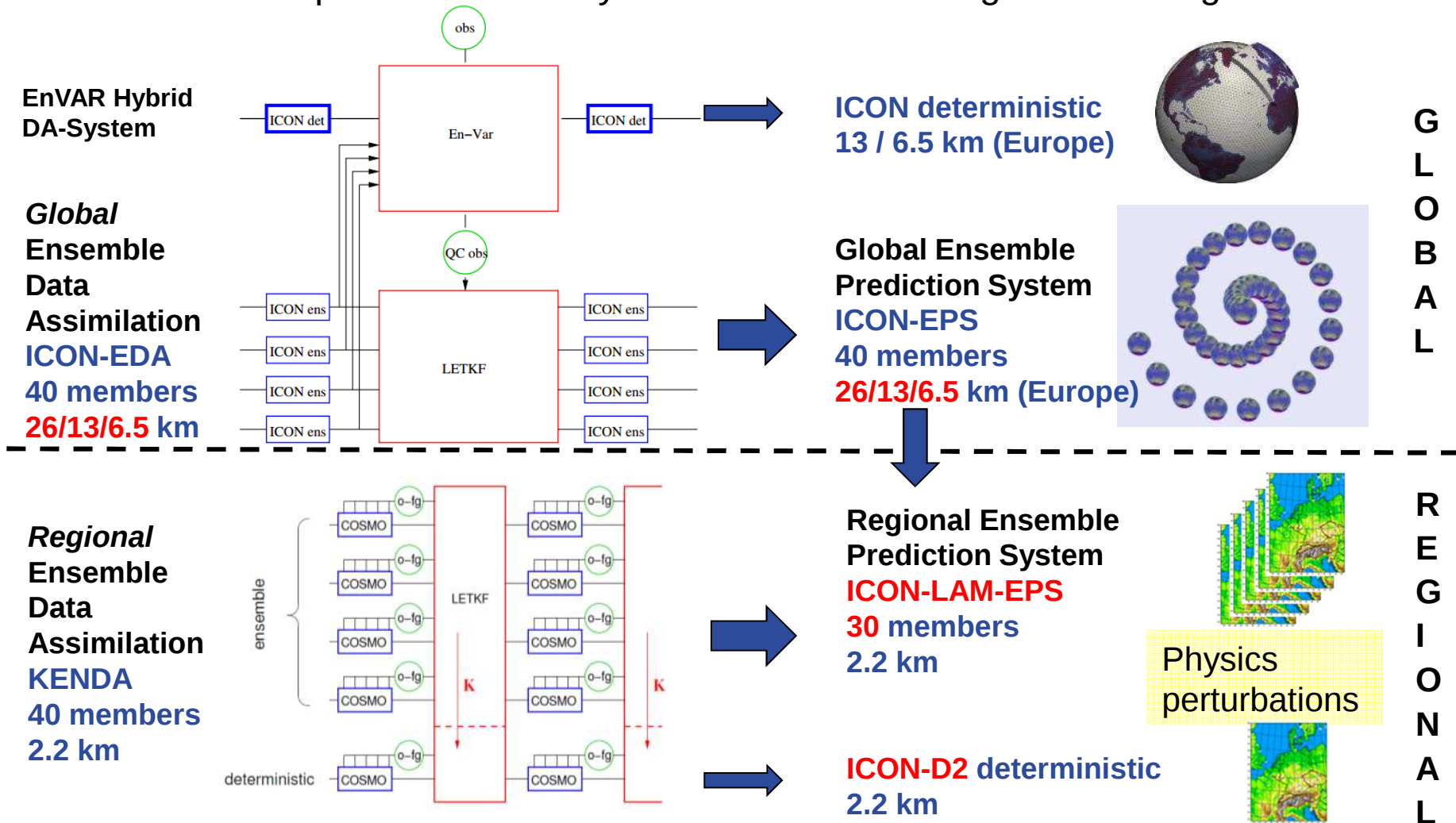
Today's Operational System

Deterministic and probabilistic analyses and forecasts on global and regional scales



Moderate Upgrade

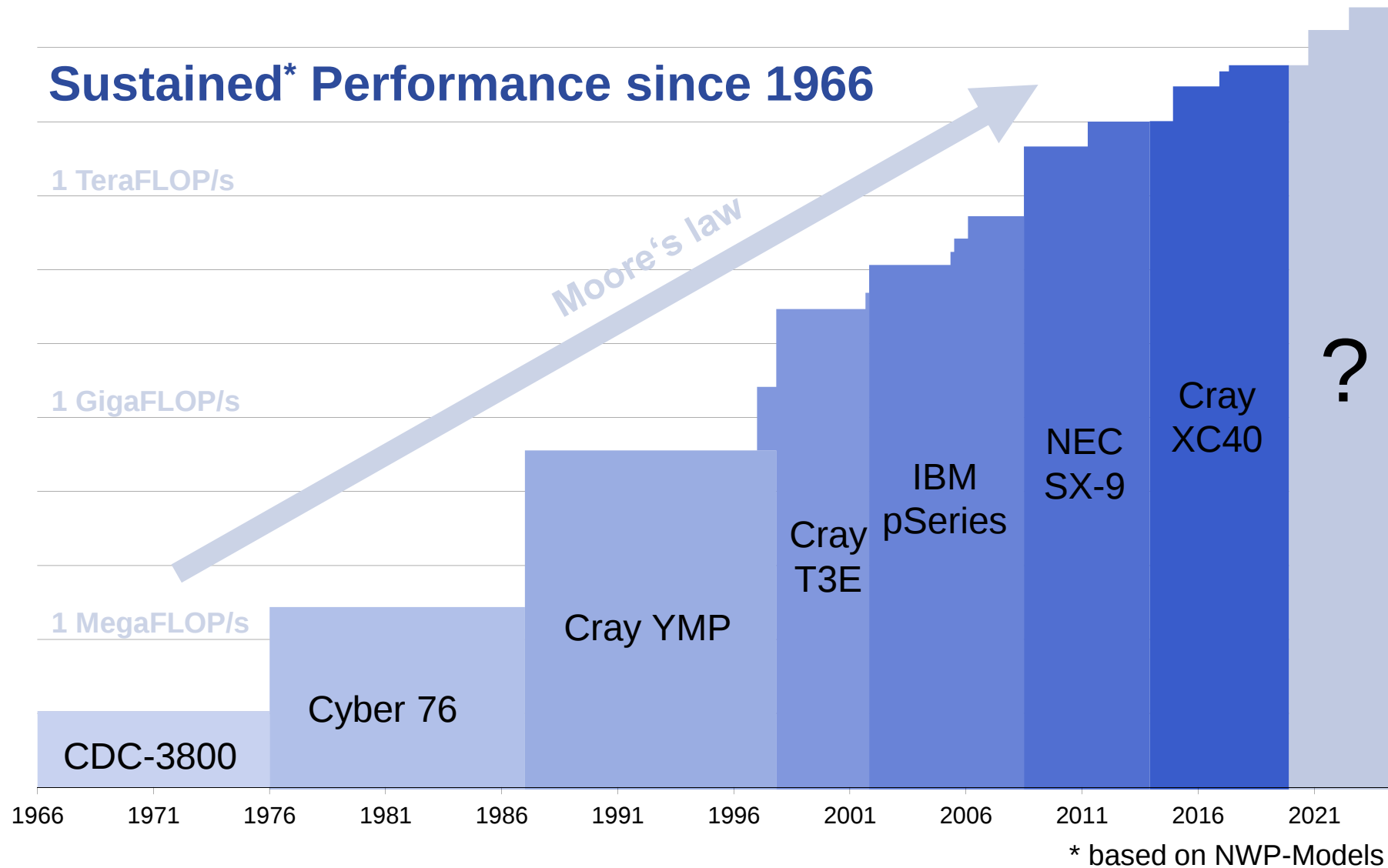
Deterministic and probabilistic analyses and forecasts on global and regional scales



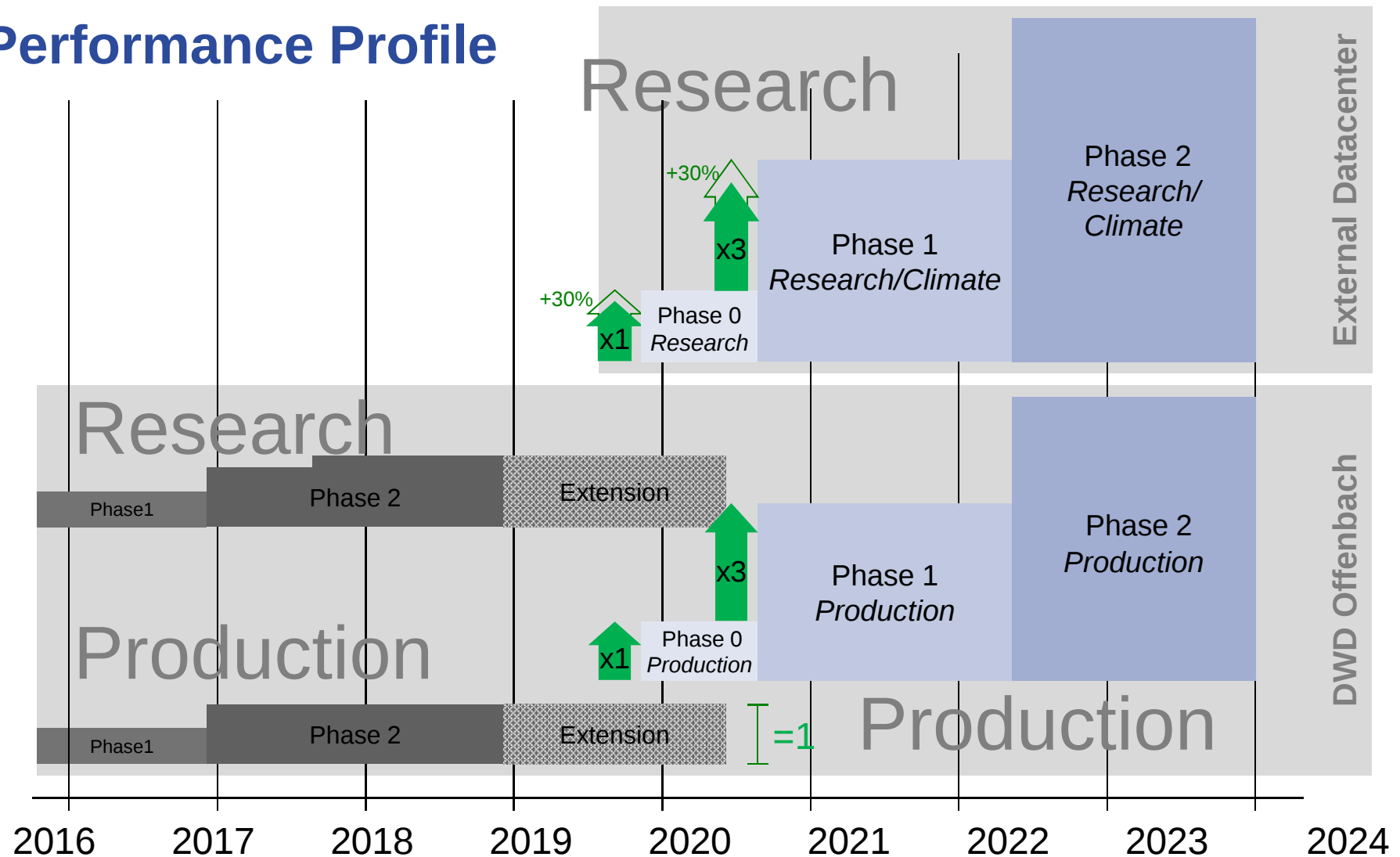
Plans for the Operational System

- Moderate upgrade to existing components of the NWP System
- New component SINFONY:
 - Seamless Integrated Forecasting System for Nowcasting and very short range forecasts
 - Combination of both approaches to an ensemble-based seamless forecast product for forecast ranges from 0 to 6 (12) h with a rapid update rate (≤ 1 h).
- Emergency response system with ICON-ART (26/13/6.5 km) on demand.
- Especially SINFONY will need a major upgrade of the HPC system.
 - Estimated growth of compute performance: at least a factor of 3 – 4.5
 - Estimated growth of data production (including parallel suites):
 - 18 to 190 Tbyte / day !

Sustained* Performance since 1966

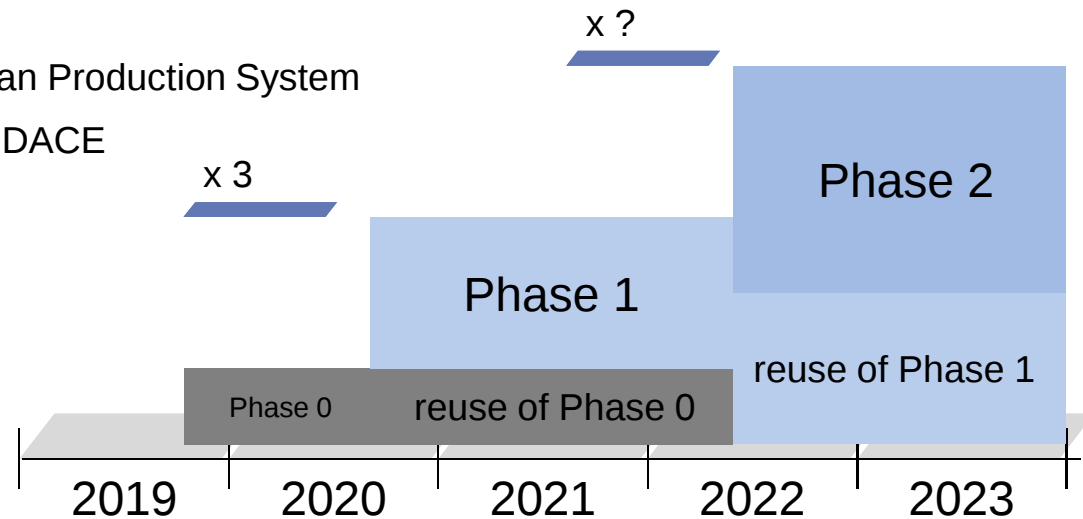


Performance Profile



Performance Profile

- Two systems: Operation / Research
- Two sites: Datacenter in OF (DWD) / External Datacenter (HPC-Vendor)
- Three Phases:
 - Phase 0: for migration only – same performance as current operational system required!
 - Phase 1: upgrading Phase0-System to at least x3 performance without introducing a new technology
 - Phase 2: upgrading or replacing Phase1-system. In case of upgrading no more than two homogeneous partitions allowed.
- Research System always 30% larger than Production System
- Benchmarks based on ICON, COSMO, DACE



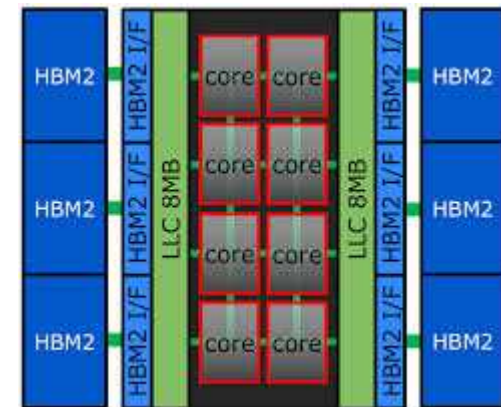
Offers

- There were offers based on
 - Intel Xeon Cooper Lake / Sapphire
 - AMD Rome / Genoa
 - Vector processors
- with a performance factor for Phase 2 ranging from 4.45 to 6.0
- And the Winner is NEC:
 - Phase 1: 3x performance growth
 - Phase 2: another 2x performance growth
 - all-in-all: Factor of 6.0!



Technical Details

- A computing node consists of
 - a vector host: 24-core AMD Rome
(2.8 GHz; 2.048 GByte 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))
 - 48 GB HBM2 3D-stacked memory (6 GB/core; 1.35 TB/s)
 - and is direct liquid cooled
- Peak performance of the vector node:
 - double: $8 \text{ engines} \times 8 \text{ cores} \times 304.1 \text{ GF/s} = 19462.4 \text{ GF/s}$
 - single: $8 \text{ engines} \times 8 \text{ cores} \times 608.3 \text{ GF/s} = 38931.2 \text{ GF/s}$



Number of Nodes per Phase

→ Where is the Performance Growth?

→ Phase 0 would need less nodes

1.3x ✓

6x ?
3x ?

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

→ Why do we have so many nodes in Phase 0? Because of a special trick for DACE Benchmark: must only use 6% of the full machine. And needs 8 – 10 nodes (NEC is still working on it!)

Performance Growth Measured by ICON-EPS

➔ Number of members / vector engines used (always 11 per member in Phase 0/1)

Phase	Operational members / engines	Experimental members / engines
0	49 / 539	64 / 704
1	161 / 1771	210 / 2310
2	161 / 1771 133 / 798	210 / 2310 173 / 1038

- ➔ Additional computing nodes in Phase 2: Aurora 3 with higher performance (need only 6 vector engines for one ICON ensemble member)
- ➔ Why do we have so many nodes in Phase 0? Because of a special trick for DACE Benchmark: must only use 6% of the full machine. And needs 8 – 10 nodes (NEC is still working on it!)

Some More Technical Details

- Infiniband HDR
- Peak Performance (in TFlops/s; Phase 2 only approximated)

Phase	Double		Single	
	Operations	Research		
0	3464.3	4515.3	6929.8	9032.0
1	4359.6	5683.0	8720.6	11367.9
2	8332.0	10874.0	16665.4	21749.9

- typical power usage in Phase 2: 777 kW (operations); 981 kW (research);
- This is about 70 % compared to Intel and about 80 % compared to AMD, which only provided a performance factor of about 4.5!



Thank you
very much
for your
attention