NEC SX Aurora Tsubasa Use for Weather and Climate

Radmila Brožková

Numerical Weather Prediction Department

CHMI

Czech Hydrometeorological Institute

Outline

- About CHMI and the Numerical Weather Prediction ALADIN System;
- Numerical Weather Prediction and Climate simulations – typical compute problems prototypes;
- Return to the vector – selection of the NEC SX-Aurora Tsubasa System;
- Running complex operational works and research, first experiences and comparisons.

About Czech Hydrometeorological Institute (CHMI)

- **Mission** – to provide high level professional data, information and services in the fields of:

  - Meteorology;
  - Climatology;
  - Hydrology and water quality;
  - Air quality.

- **CHMI** is part of the critical infrastructure of the state.

CHMI headquarters at Prague

About the ALADIN System

- The ALADIN System is an ensemble of codes and tools for the **Numerical Weather Prediction** and **Regional Climate Simulations**;

- ALADIN serves for regional very high-resolution forecast;

- The code is mainly written in Fortran, and the source reaches more than 2 million lines. The code shares the main structure, including data flow and parallelization strategy, with the global models ARPEGE (Météo-France) and IFS (ECMWF);

- There are 26 national meteorological services cooperating on the scientific and technical development of the system, forming a new consortium ACCORD.

Weather forecast – a time critical application

- The NWP System has the following major components:

  - Dynamical core;
  - Schemes to parameterize physics of the atmosphere and underlying surface;
  - Assimilation of incoming observations to update the knowledge on the actual atmospheric state.

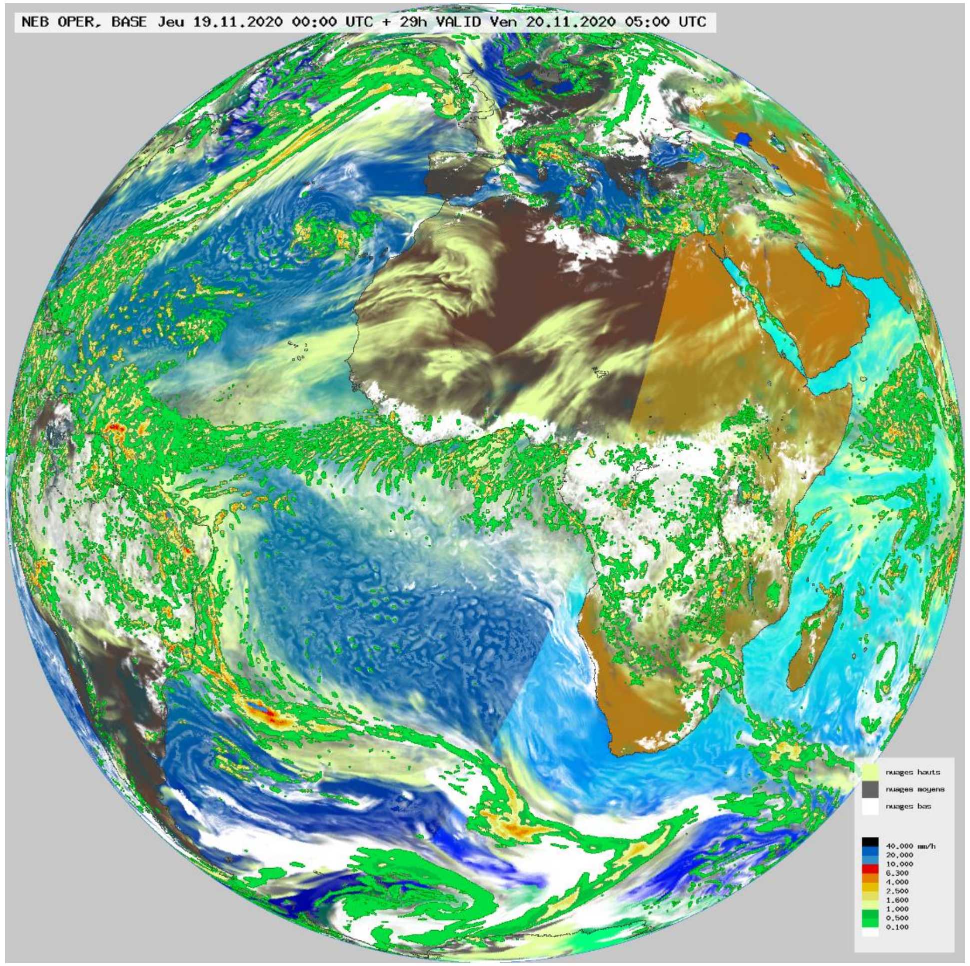
- One forecast production run must be completed within about 1 hour of wall-clock time, employing a complex set of tasks.

Meteorological task typical compute problems

- Numerically efficient integration of Navier – Stokes type of equations requiring solvers, in ALADIN we profit from the spectral formulation;

- Run for more physical realism and precision; the most compute demanding is the gaseous radiative transfer calculation;

- The initial value problem solves a variational task of the dimension of 10^8 .



ARPEGE forecast: synthesis image of clouds and precipitations

ALADIN System HPC platforms portability (1/2)

- Starting from the ARPEGE/IFS, we have been following the evolution of HPC technologies over years;
- The original development was on vector machines: Cray-2, Cray C90, Fujitsu VPP series, NEC-SX series at ECMWF, Météo-France, CHMI …;
- Scalar platforms came progressively in from 2000’, e.g. IBM Power series, then x86.

CHMI vector supercomputers used from 1998 to 2017

ALADIN System HPC platforms portability (2/2)

- In recent years, the x86 architecture became dominant. The code keeps its high vectorization capacity. The classical hybrid MPI-OpenMP parallelization is available.

- No profit yet from GPU; a major rewriting would be necessary, a far too big work investment with uncertain outcomes.

CHMI x86 LX supercomputer used from 2017

HPC platform performance requirements

- To meet the major application needs, **ALADIN benchmarks** are key selection criterions;

- The benchmarks reflect the **future use of the platform**, like foreseen resolution and scientific content of the model;

- The ALADIN reference on x86 Broadwell was 24h forecast, needing 90 nodes (180 sockets) to be completed within 800 s. A performance gain factor w.r.t. the reference is constructed (timeliness of a single copy on a given number of nodes, number of simultaneous copies to run on time);

- These **applicative benchmarks** can be completed by technical tests, such as IO to verify the file-system performance, even if the model benchmarks reflect it indirectly as well.

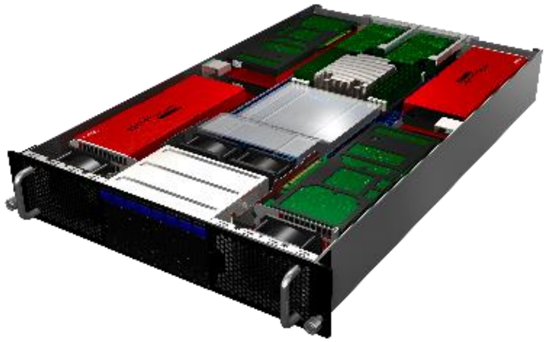
The SX-Aurora selection

- In the CHMI latest procurement, the proposed solution by NEC based on the SX-Aurora Tsubasa technology was ***the most performant by a wide margin***.

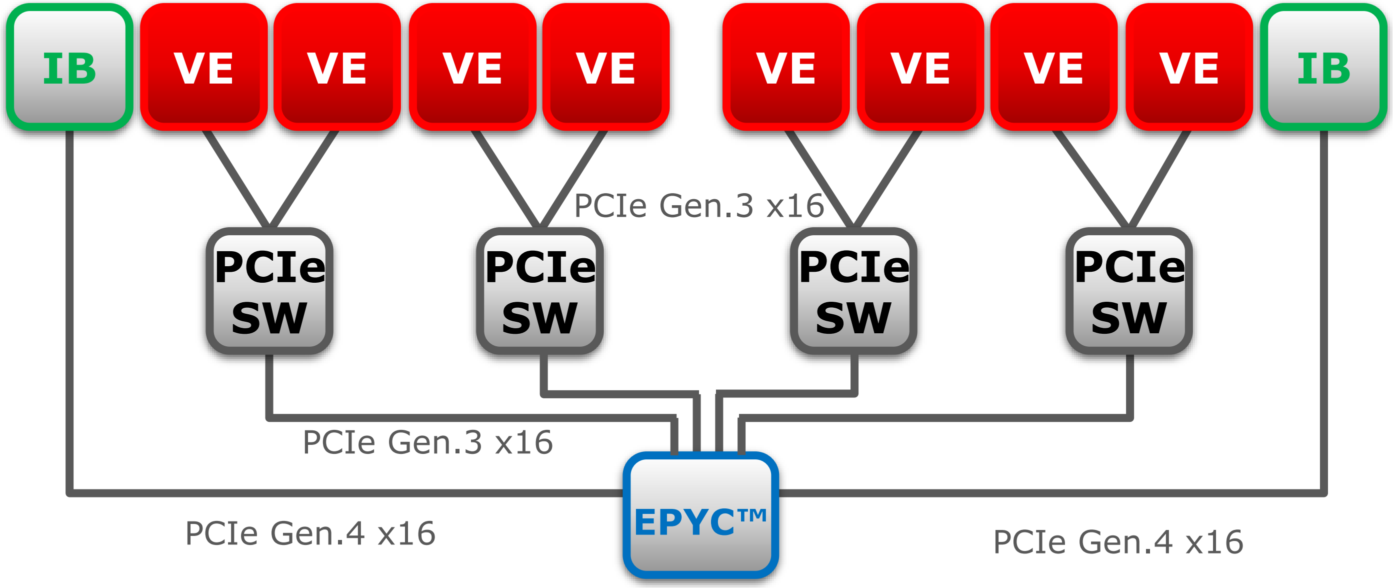
- The compute performance was the decisive factor, while the system had to meet the ***infrastructure limits***.

CHMI SX-Aurora Tsubasa supercomputer accepted in December 2020.

The configuration at CHMI: Aurora node B401-8



VE Type 20B



- AMD EPYC Procesor, 24 cores, Linux OS, 512 GB DDR4 RAM;
- 8 Vector Engines type 20B, each with 8 cores and 48 GB HBM2 memory;
- 2 x 200 Gbit HDR network;
- Direct liquid cooling;
- 16 Aurora nodes (Vector Hosts) per rack, 48 nodes in total.

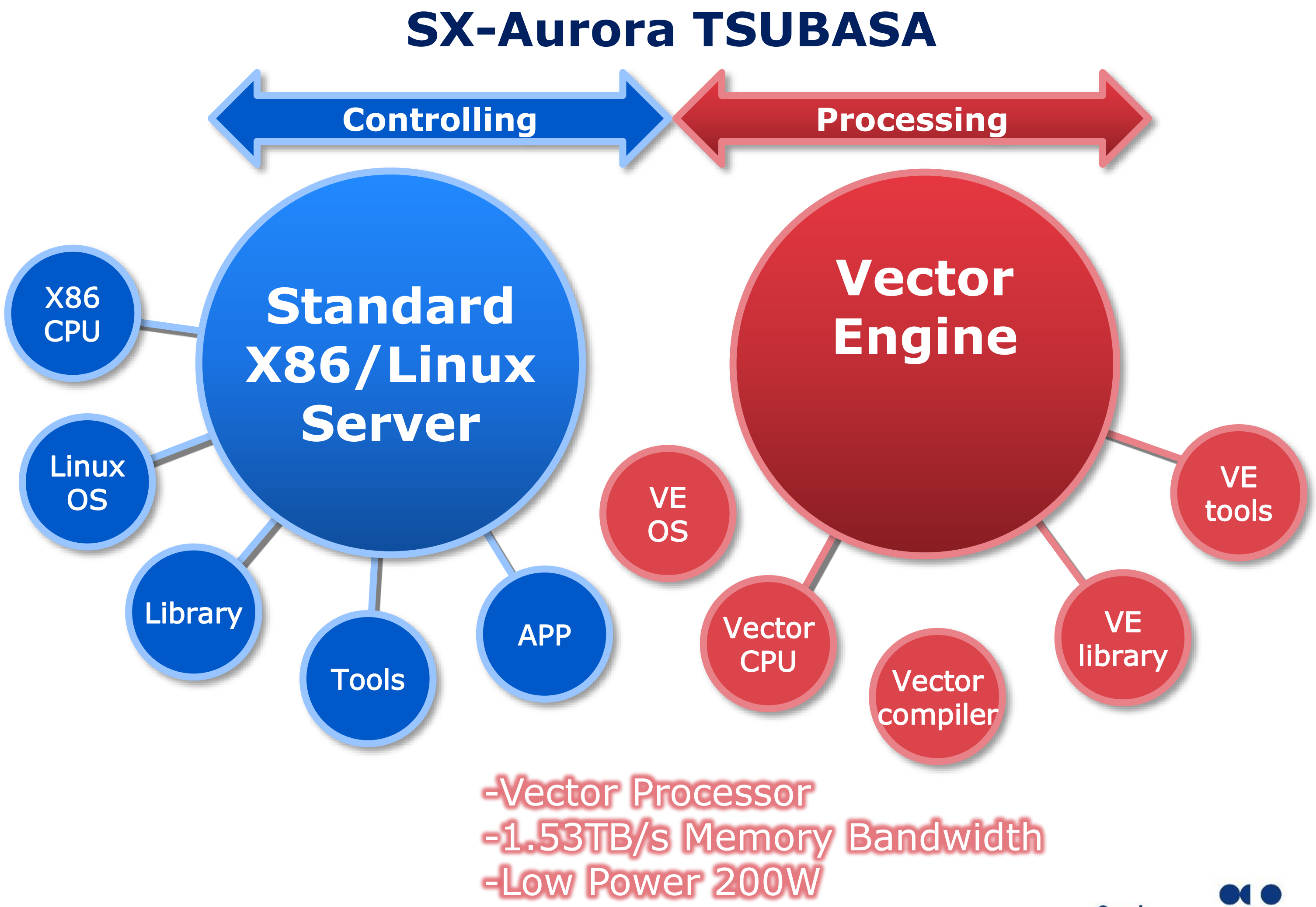
Comfort and optimal use of resources

The major difference with respect to the previous vector platforms is **the hybrid architecture**;

No need to port small applications or those unsuitable for the vector treatment to Vector Engines (VE);

Such tasks can run on Vector Hosts (VH); many standard SW packages run on x86.

We optimize the operational works suite using the VH/VE hybrid features.



Porting ALADIN configurations to Aurora VE

- Highly practical cross-compiler kit is installed on the so-called service nodes (front-ends) to compile the ALADIN code;
- Some deep technical libraries are compiled directly on a VE if needed;
- The forecast configuration - ready to use thanks to benchmarks;
- Other main configurations, like data assimilation, have been ported smoothly and are running.

Code Optimizations (1/2)

- The compiler makes the vectorization automatically most of the time;
- In contrast to a possible profit from GPUs, no major code re-factorization is necessary;
- The ALADIN code has high values of vector operation ratio;
- Optimization are localized to few routines, code changes remain small.

Code Optimizations (2/2)

Essential: code changes do ***not compromise the code portability***:

- Compiler directives;
- No performance loss on the x86 architecture.

To find a code for optimizations, a tracing profiler is available by NEC

Example from data assimilation variational job

V.OP RATIO	AVER. V.LEN	VECTOR TIME	L1CACHE MISS	CPU PORT CONF	VLD LLC HIT E.	PROC.NAME	
92.98	87.3	50.264	0.002	1.254	98.57	jgvcor_$1	before optimization
99.43	251.9	16.690	0.004	5.910	96.11	jgvcor_$1	optimized: only two loops reorganized

The example of the radiative transfer

The most expensive computation in the forecast;

Intermittent computations: precise expensive gaseous transmissions are not calculated at every time step.

Going to full precise computations:	**x86**	model cost x **2.54**
	Aurora	model cost x **1.78**
		=> *higher efficiency*

Forecast performance / power consumption ratio

We compare power needed to get the ALADIN forecast executed at ***the same wall-clock time***:

• Broadwell 12 cores:	**24.75**	kW	measured
• AMD Roma 64 cores:	12.2	kW	estimate
• Aurora B401-8:	**6**	kW	measured

Conclusion

- The vector technology remains suitable for our meteorological problem, and likely to similar scientific computations;
- NEC SX-Aurora Tsubasa hybrid solution offers a flexibility of using the standard x86 and vector engines where appropriate;
- Our first experiences show a rather straightforward use of the system.

Thank you for your attention

Radmila Brožková

🖂*radmila.brozkova@chmi.cz*

Czech Hydrometeorological Institute

Question:
Can you say more about the decision to return to a vector supercomputer?