NERO:
A Near High-Bandwidth Memory Stencil Accelerator for Weather Prediction Modeling

Gagandeep Singh, Dionysios Diamantopoulos, Christoph Hagleitner,
Juan Gómez-Luna, Sander Stuijk, Onur Mutlu, and Henk Corporaal

30th FPL, Sweden
31th August 2020

TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY
ETH Zürich
SAFARI
IBM Research | Zurich

Stencil Computation in Weather Modeling

COSMO (Consortium for Small-Scale Modeling)

- Around **80 complex stencils**

- Horizontal diffusion
  Vertical advection

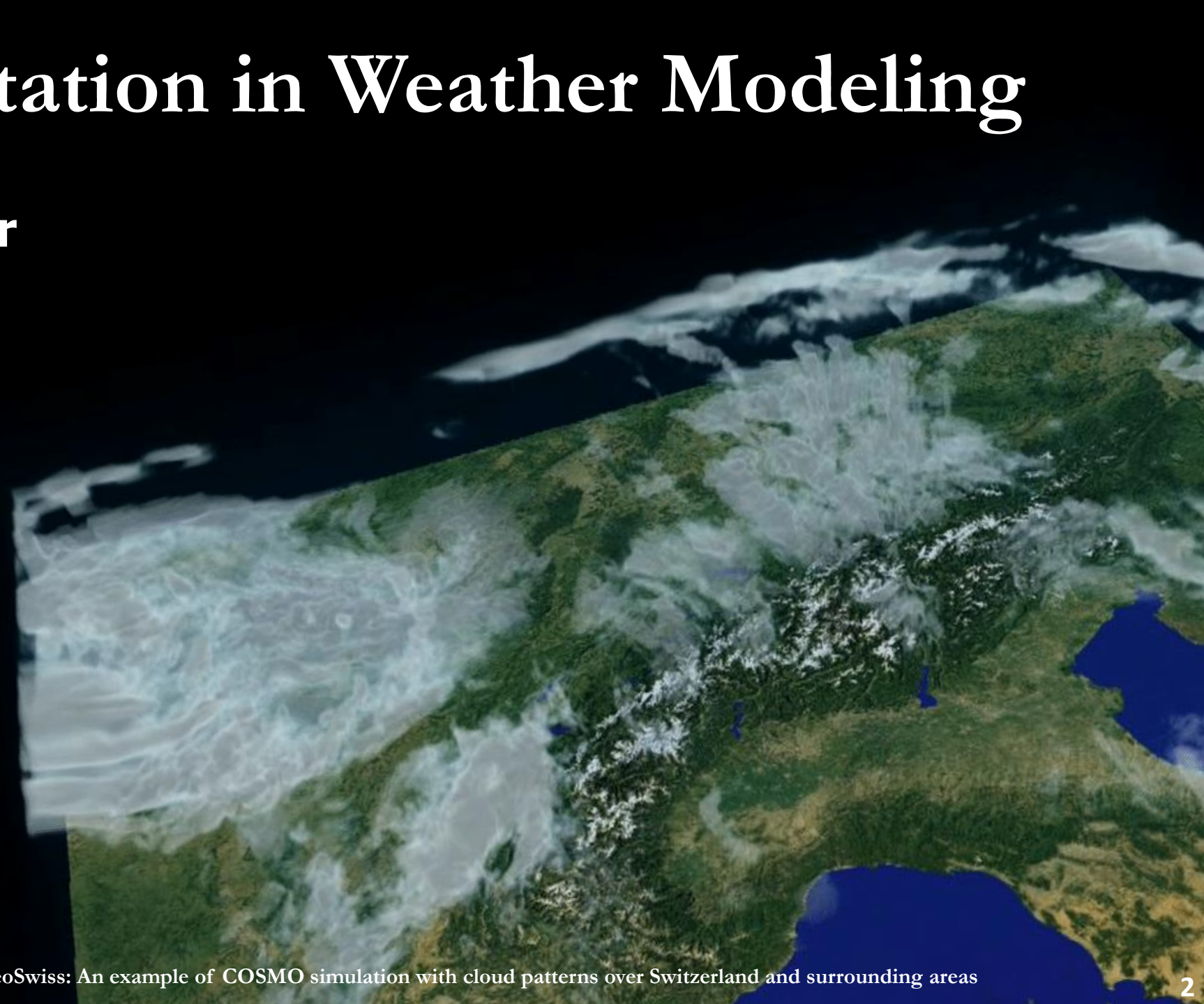
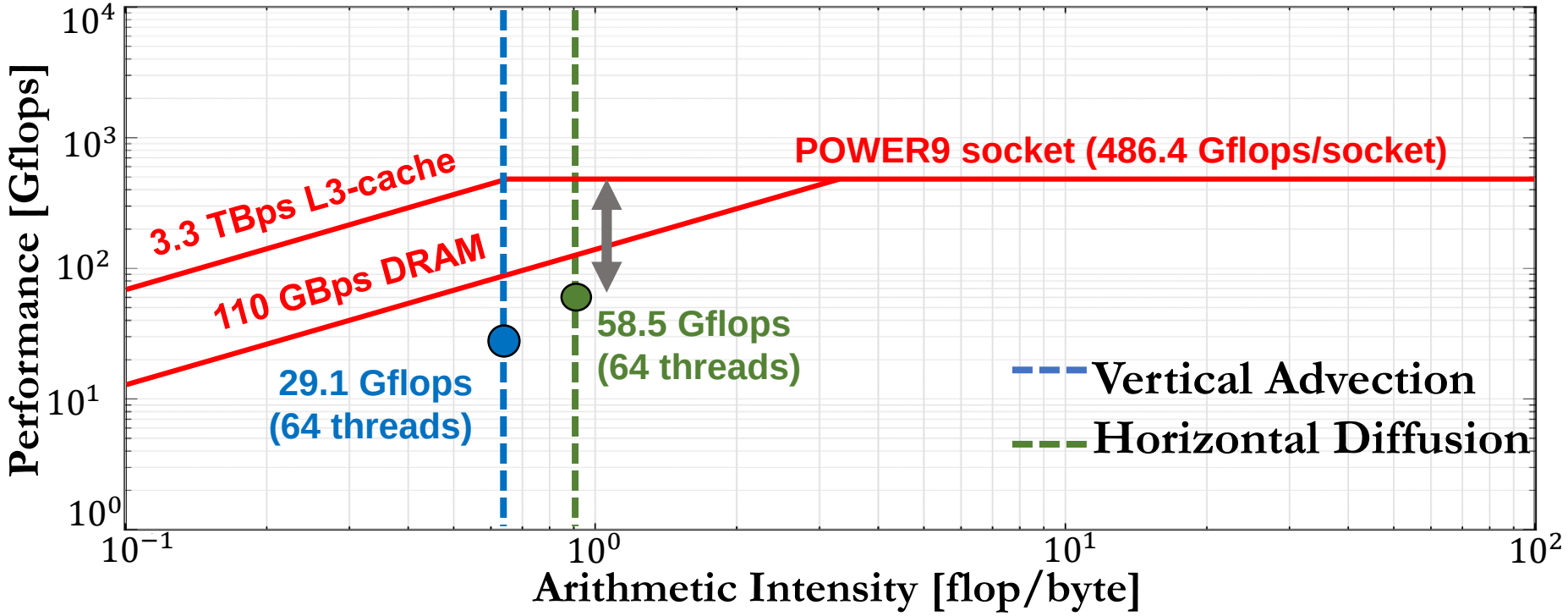


Image Source: NVIDIA/MeteoSwiss: An example of COSMO simulation with cloud patterns over Switzerland and surrounding areas

Motivation and Goal

- **Memory bound** with **limited performance** and **high energy consumption** on **IBM POWER9 CPU**



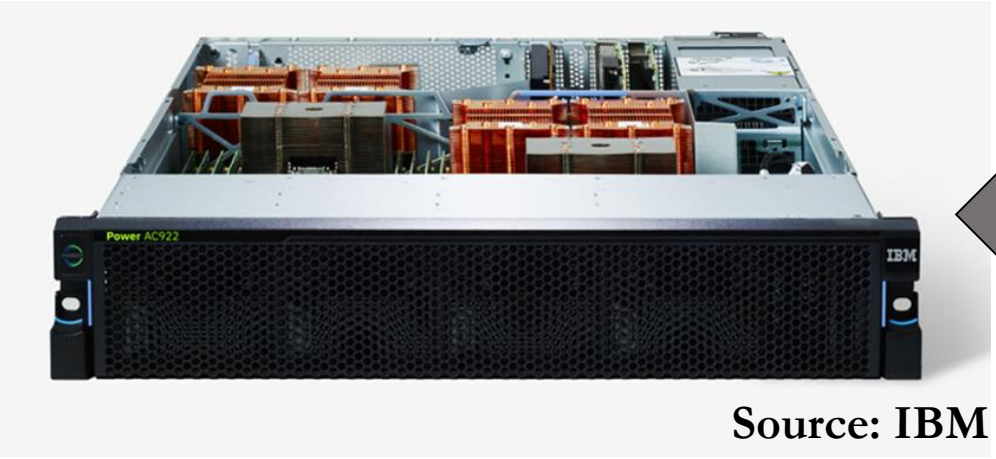
Goal:

- **Mitigate** the **performance bottleneck** of weather prediction kernels in an **energy-efficient way**
- Evaluate the use of **near-memory acceleration** using a **FPGA+HBM** connected through **IBM CAPI2** (Coherent Accelerator Processor Interface)

NERO: A Near High-Bandwidth Memory Stencil Accelerator for Weather Prediction Modeling

- First **near-HBM FPGA-based** accelerator for representative kernels from a **real-world weather prediction application**
- Data-centric caching with **precision-optimized tiling** for a heterogeneous memory hierarchy
- In-depth **scalability analysis** for both DDR4 and HBM-based FPGA

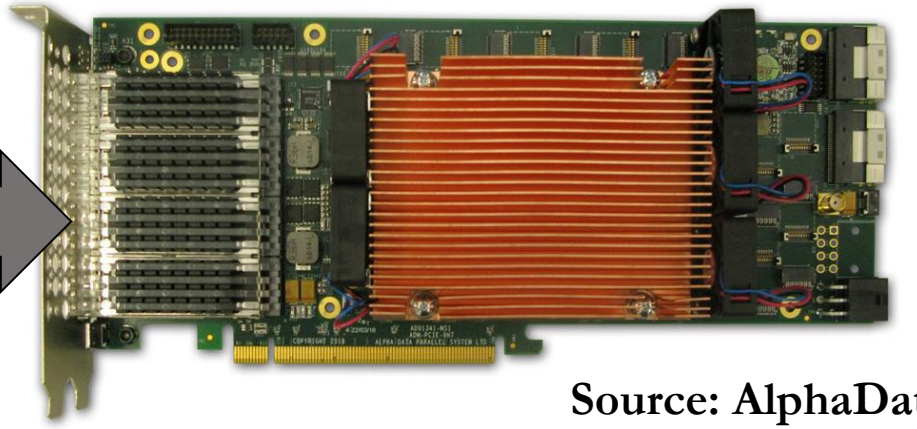
Heterogeneous System: CPU+FPGA



Source: IBM

POWER9 AC922

CAPI2



Source: AlphaData

HBM-based AD9H7 board

We evaluate two POWER9+FPGA systems:

1. HBM-based AD9H7 board

Xilinx Virtex Ultrascale+™ XCVU37P-2

Heterogeneous System: CPU+FPGA

Source: IBM

CAPI2

Source: AlphaData

POWER9 AC922

DDR4-based AD9V3 board

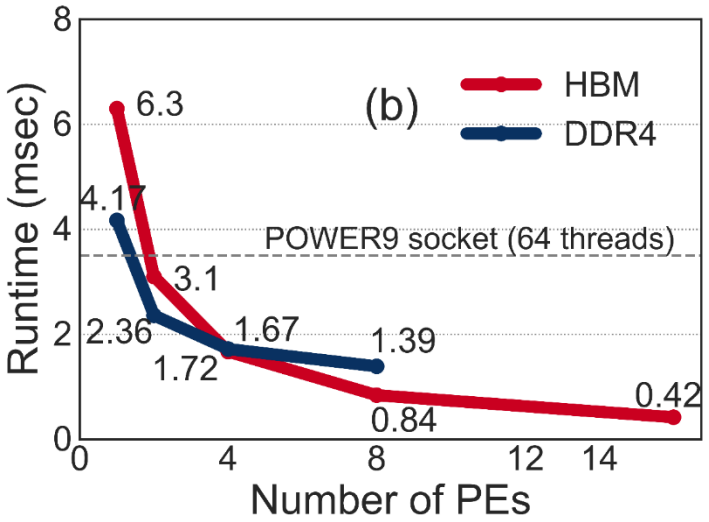
We evaluate two POWER9+FPGA systems:

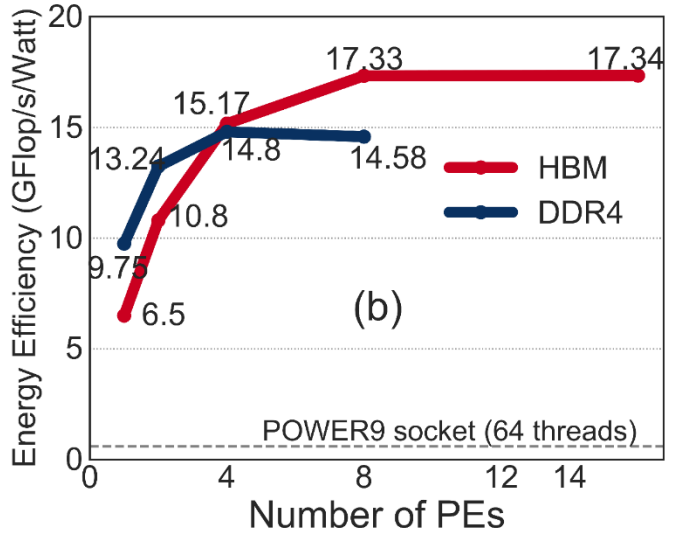
1. HBM-based AD9H7 board
Xilinx Virtex Ultrascale+™ XCVU37P-2

2. DDR4-based AD9V3 board
Xilinx Virtex Ultrascale+™ XCVU3P-2

Results

- NERO **outperforms** a **16-core IBM POWER9** system by 4.2x and 8.3x when running two compound stencils
- NERO **reduces energy** consumption by 22x and 29x
- NERO **provides energy efficiency** of 1.5 GFLOPS/Watt and 17.3 GFLOPS/Watt





Hardware acceleration on an FPGA+HBM is a promising solution for weather modeling

NERO:
A Near High-Bandwidth Memory Stencil Accelerator for Weather Prediction Modeling

Gagandeep Singh, Dionysios Diamantopoulos, Christoph Hagleitner,
Juan Gómez-Luna, Sander Stuijk, Onur Mutlu, and Henk Corporaal

30th FPL, Sweden
31th August 2020

TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY
ETH Zürich
SAFARI
IBM Research | Zurich