

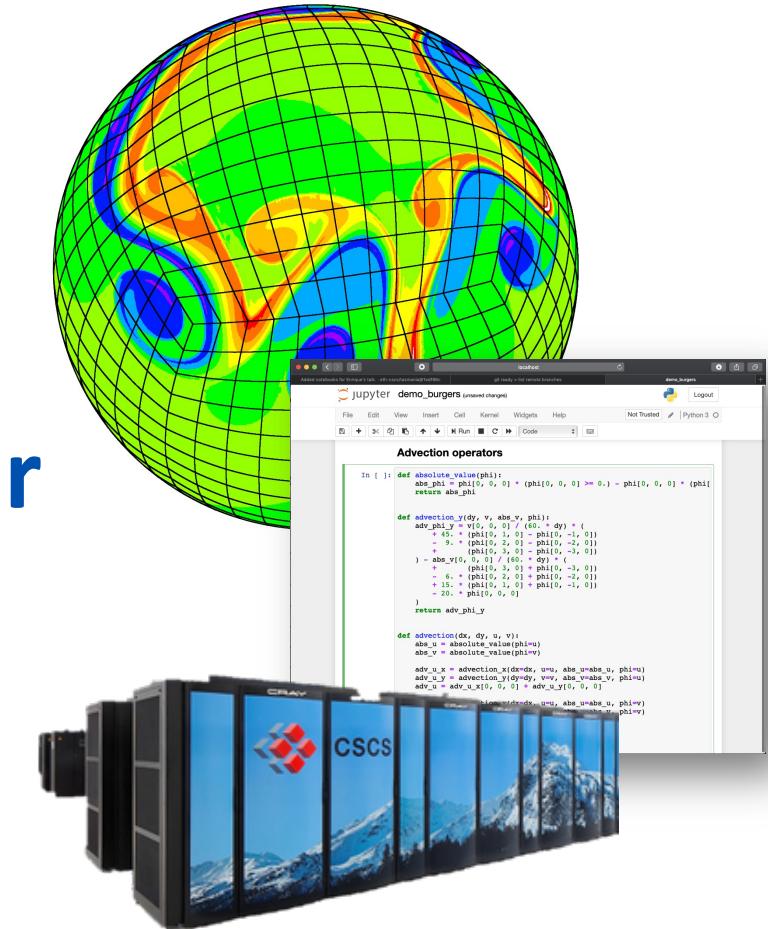
# High Performance Computing for Weather and Climate (HPC4WC)

Content: Graphics Processing Units

Lecturers: Oliver Fuhrer

Block course 701-1270-00L

Summer 2021



# Learning goals

- Understand why specialized hardware such as GPUs is become the new “normal”
- Learn how to program a GPU using a high-level programming language
- Understand potential and difficulties of GPU-computing

# Power, power, power!



Scalability tests with IFS on Piz Daint for simulations with 1.45km grid spacing (Düben et al., 2020)

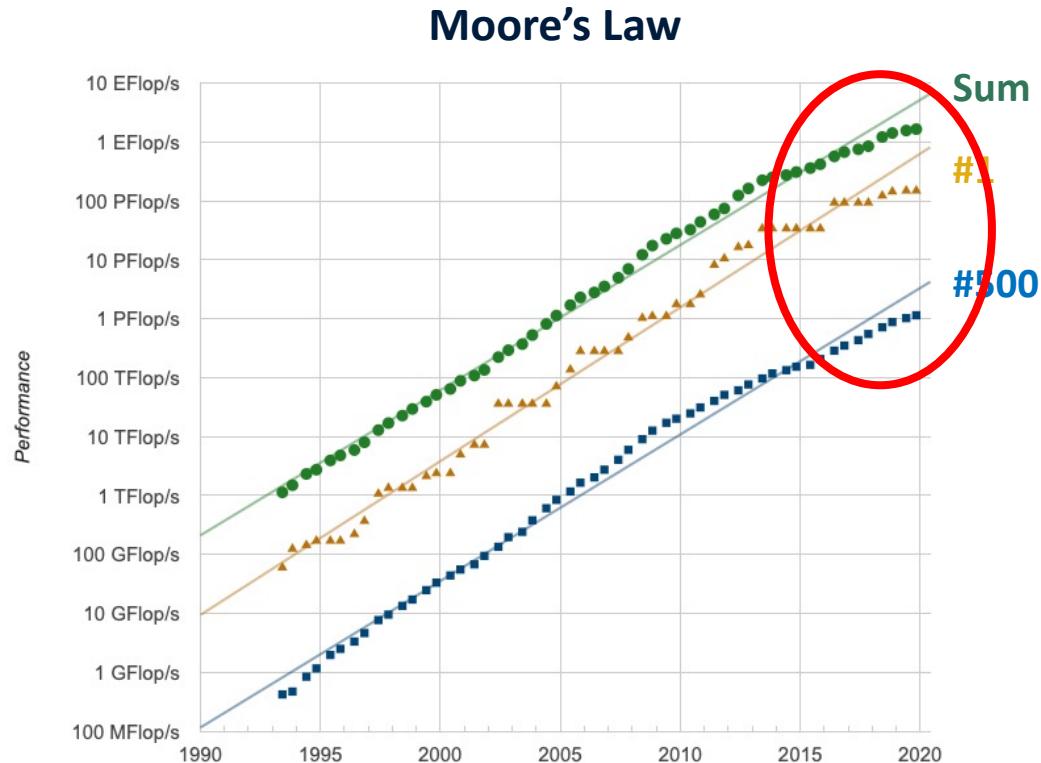
Dycore option	#tasks and threads	Energy consumption per year	Throughput
Hydrostatic	4880 tasks; 12 threads per task	85.21 MWh/SY	0.190 SYPD
Non-hydrostatic	9776 tasks; 6 threads per task	191.74 MWh/SY	0.088 SYPD
Non-hydrostatic	4880 tasks; 12 threads per task	195.30 MWh/SY	0.085 SYPD

$$191.74 \text{ MWh/SY} * 0.088 \text{ SYPD} = 16873 \text{ kWh/Day}$$

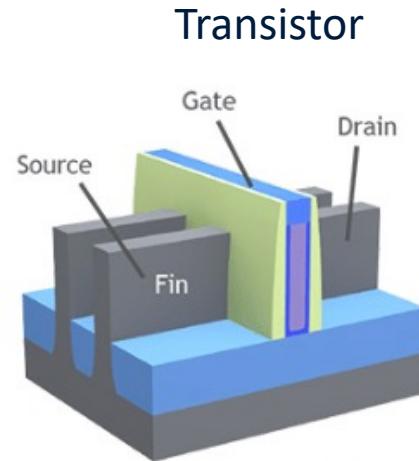
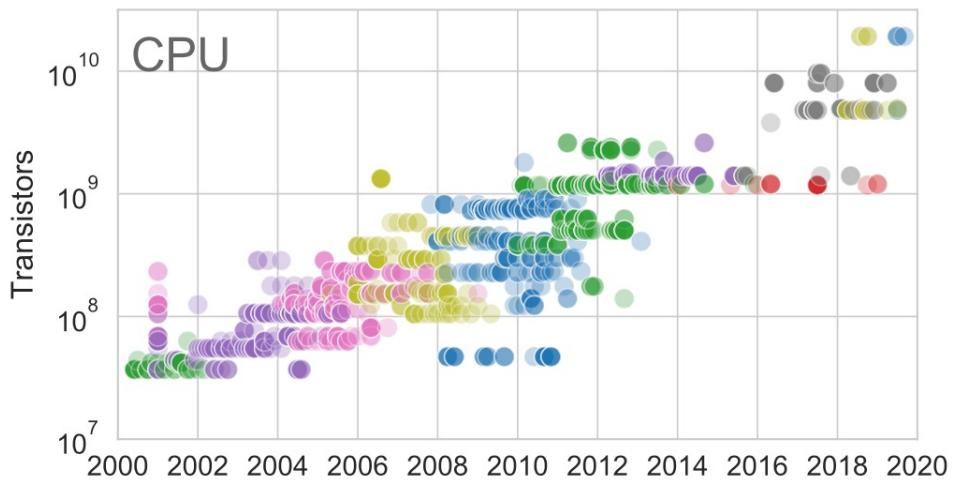
Average electricity consumption of U.S. residential utility:  $\approx 29 \text{ kWh/day}$

# Where do the Watt's go?

- CPUs are very complex and built for general purpose
- For weather and climate modeling, we often don't use all of the capabilities of a general purpose computer chips
- Historically, chips were custom-built for scientific computing
- Moore's law and the economics of developing and manufacturing chips (\$30 - \$80 M) changed that
- Era of "General Purpose Computing"

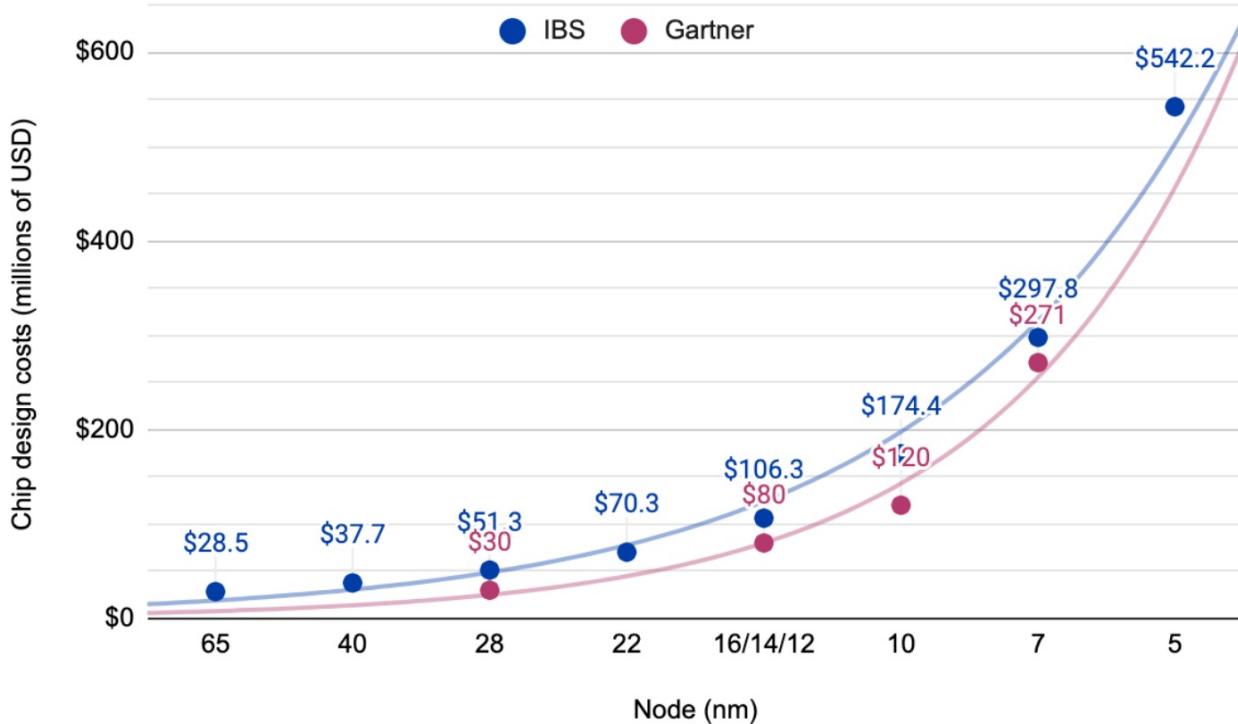


# The End of General Purpose Computing

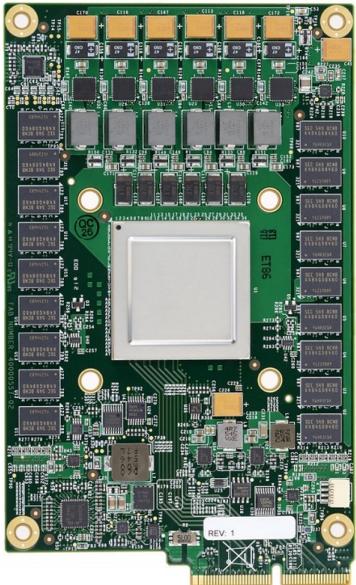


Distance between Si-atoms is 0.5 nm!

# Chip Design Costs



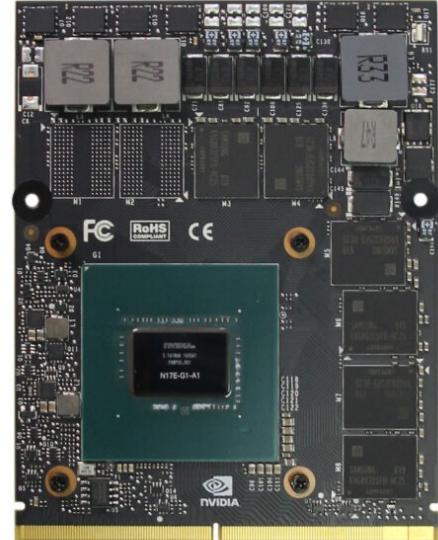
# Specialized Chips are on the Rise!



**Google's TPU**  
(e.g. machine learning)



**FPGA**  
(e.g. bitcoin mining)



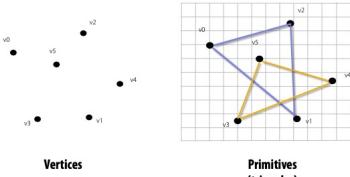
**GPU**  
(e.g. gaming)



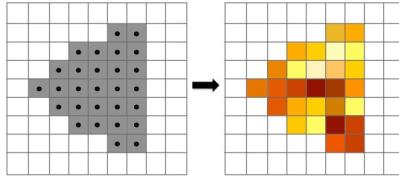
GPU's are for gaming, right?

# GPU Computing

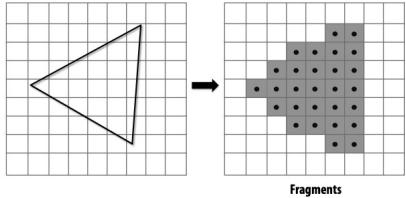
## Primitive computation



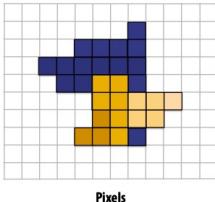
## Fragment processing



## Rasterization



## Pixel operations



**Parallel operations!**

As video games became more complex, the hardware and software environment evolved to be more flexible.

With *OpenGL*, *OpenCL*, *CUDA*, ... programming languages started to appear that made general purpose computing on GPUs possible.

**GPUs are great for some workloads / algorithms, but not so great for others!**

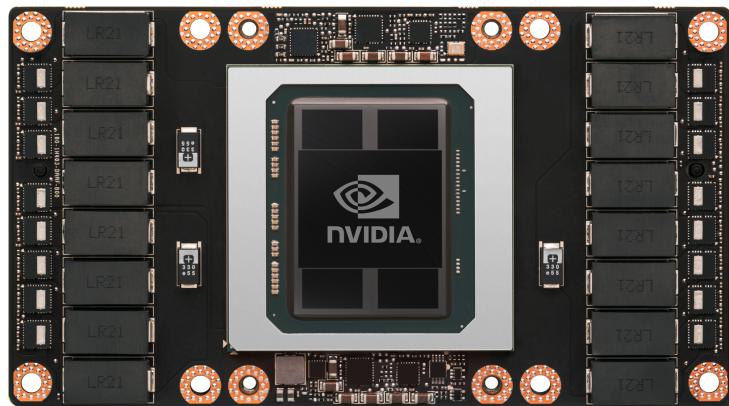
# Performance / Watt

Intel Xeon E5-2690 v3 + DRAM



~ 200 W    0.5 TFLOP/s    70 GB/s

NVIDIA Tesla P100

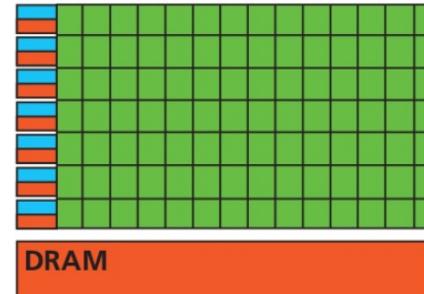
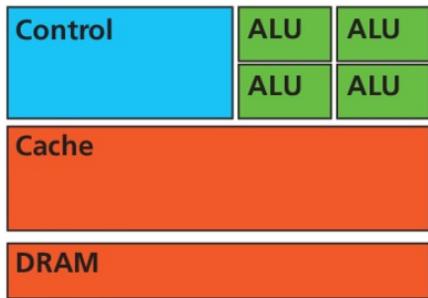


~ 300 W    5.3 TFLOP/s    720 GB/s

# CPU vs. GPU

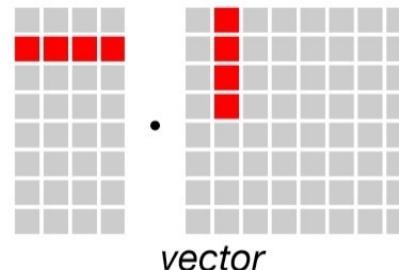
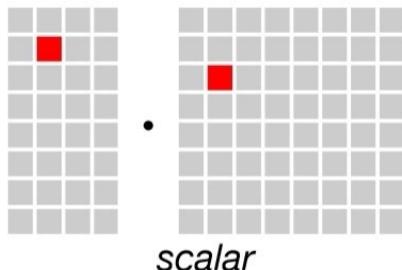
## Architecture

Few complex cores  
Few fast threads  
Complex ILP  
Little die surface for FLOPs

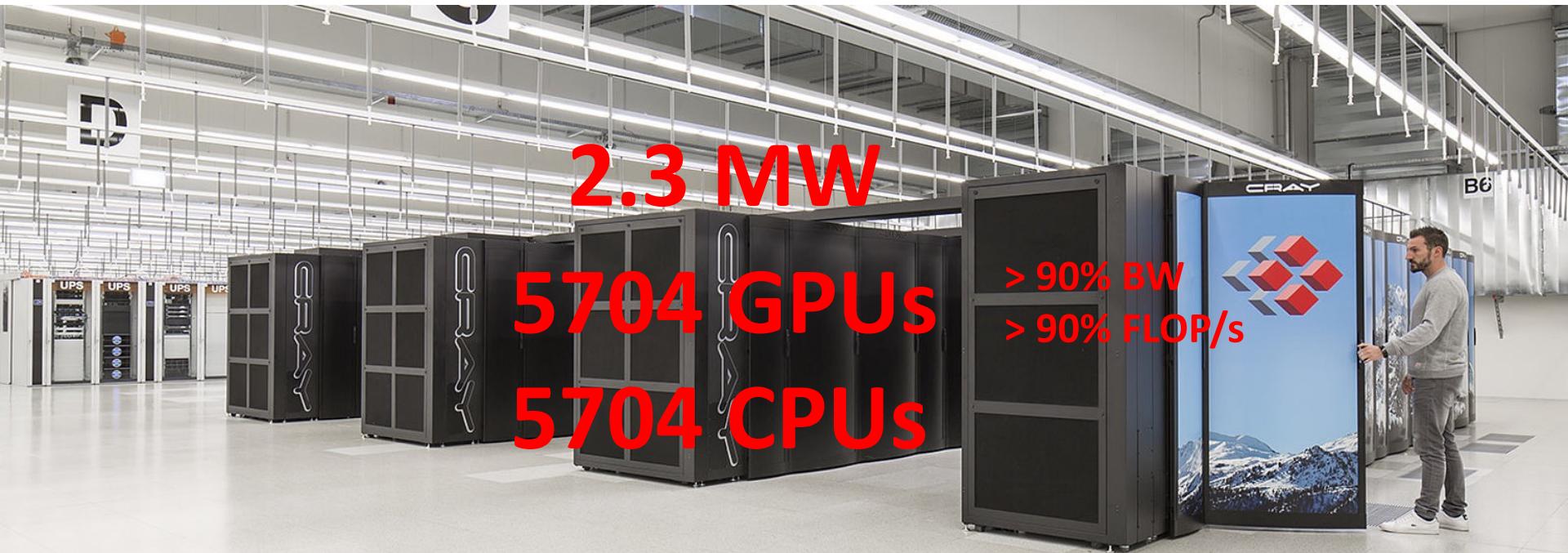


Many simple cores  
Many concurrent threads  
Maximize FLOP-throughput

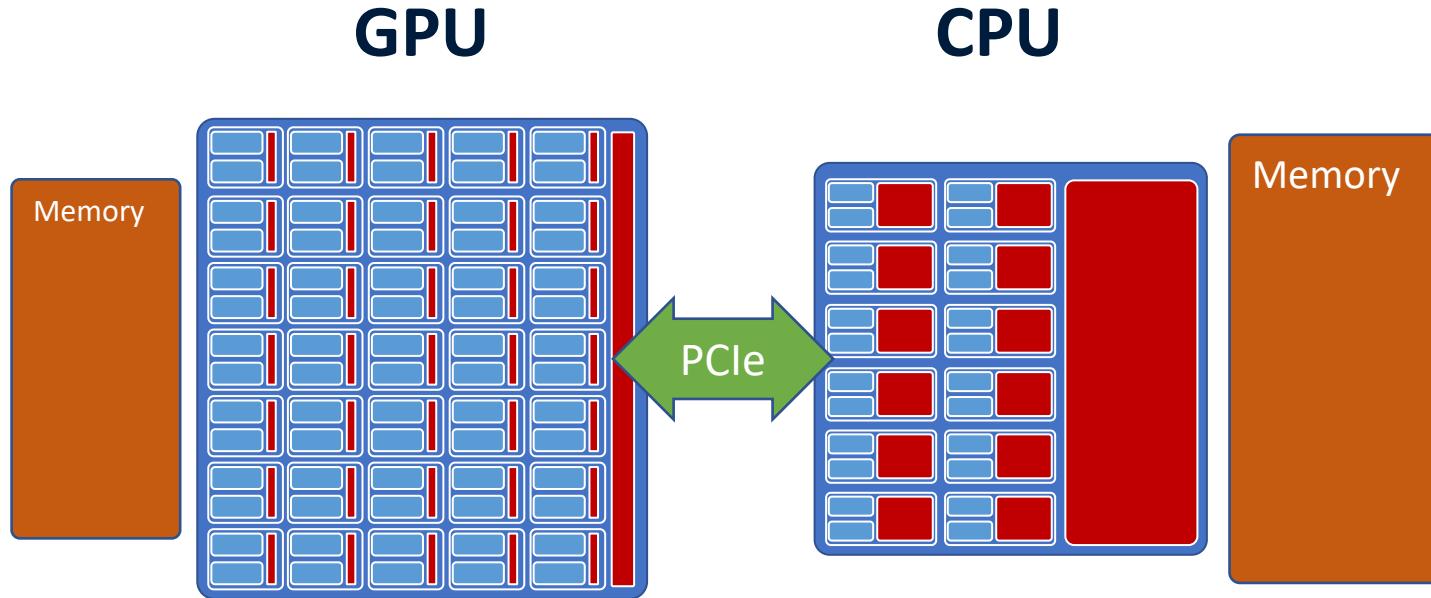
## Compute primitive



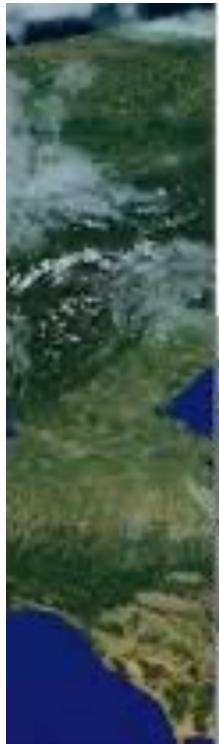
# Hybrid Supercomputer



# Node Architecture



# Weather and Climate on GPUs

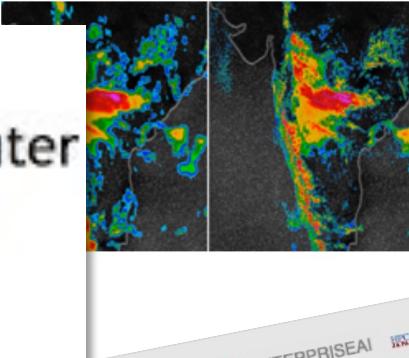


## MeteoSwiss New Weather Supercomputer

World's First GPU-Accelerated Weather Forecasting System



2x  
48  
192 Tesla  
> 90% of FLOPs  
Operational



HPC | TOP STORIES

Jan 10, 2019

New GPU-accelerated Weather Forecasting System Dramatically Improves Accuracy

By Nefi Alarcon

Tags: News, OpenACC, Tesla

Discuss

TABOR NETWORK: DATANAMI ENTERPRISEAI HPCWIRE JAPAN ADVANCED SCALE FORUM HPC ON WALL STREET

**HPC** wire

Since 1987 - Covering the Fastest Computers  
in the World and the People Who Run Them

- Home
- Technologies
- Sectors
- COVID-19
- AI/ML/DL



Swiss Weather Forecasting Achieves 1.1km Resolution on 'Piz Kesch'

April 1, 2016

# GPUs to Power the Exascale Era



Home > CPUs

## El Capitan Supercomputer Detailed: AMD CPUs & GPUs To Drive 2 Exaflops of Compute

by Ryan Smith on March 4, 2020 1:00 PM EST

55

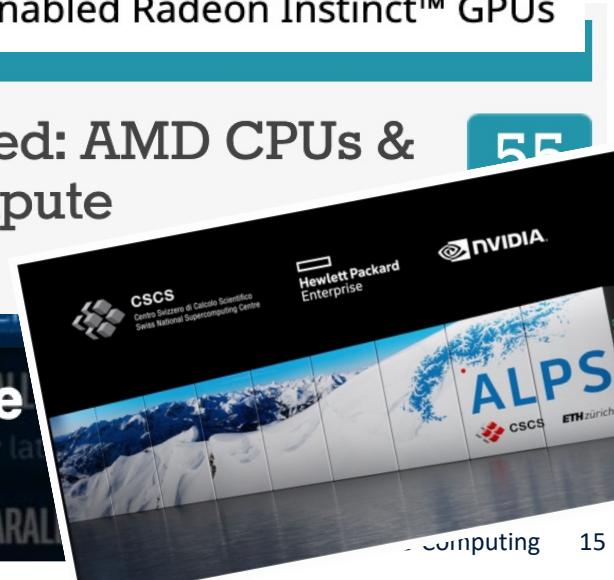
## Intel Xe-HP GPU Deployed for Aurora Exascale Development

By Tiffany Trader

ONEAPI

## U.S. Department of Energy and Cray Deliver Record-Setting Frontier Supercomputer at ORNL

- Custom AMD EPYC™ CPUs optimized for HPC and AI
- Purpose-built HBM-enabled Radeon Instinct™ GPUs



# Lab Exercises

## 01-GPU-programming-cupy.ipynb

- Introduction to GPU programming using a high-level programming language