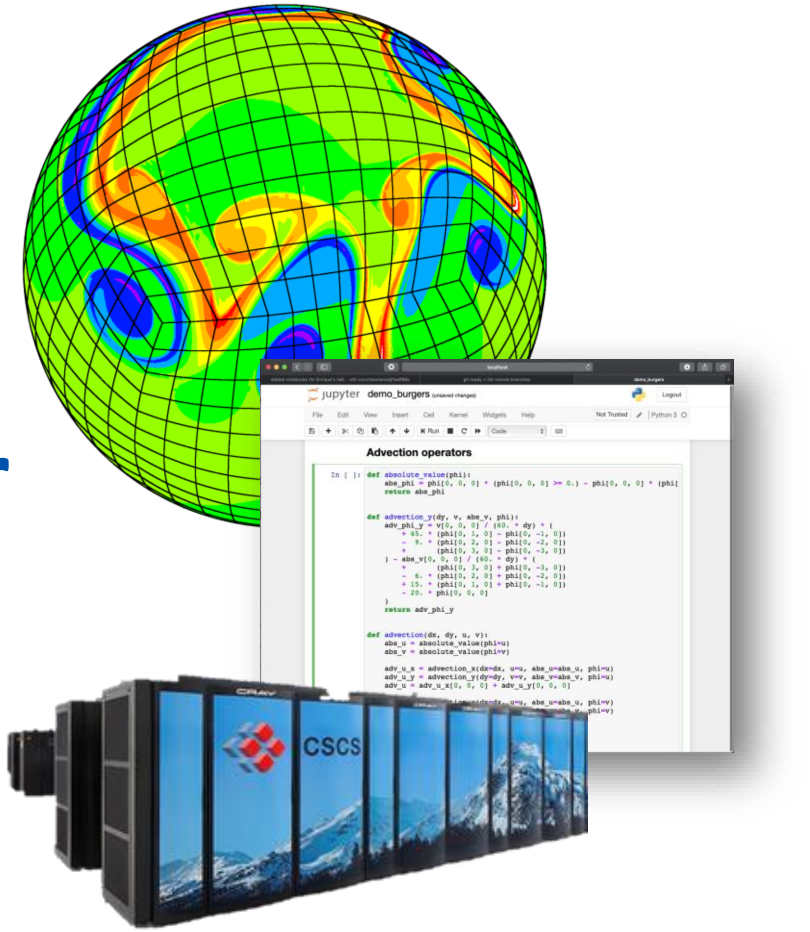
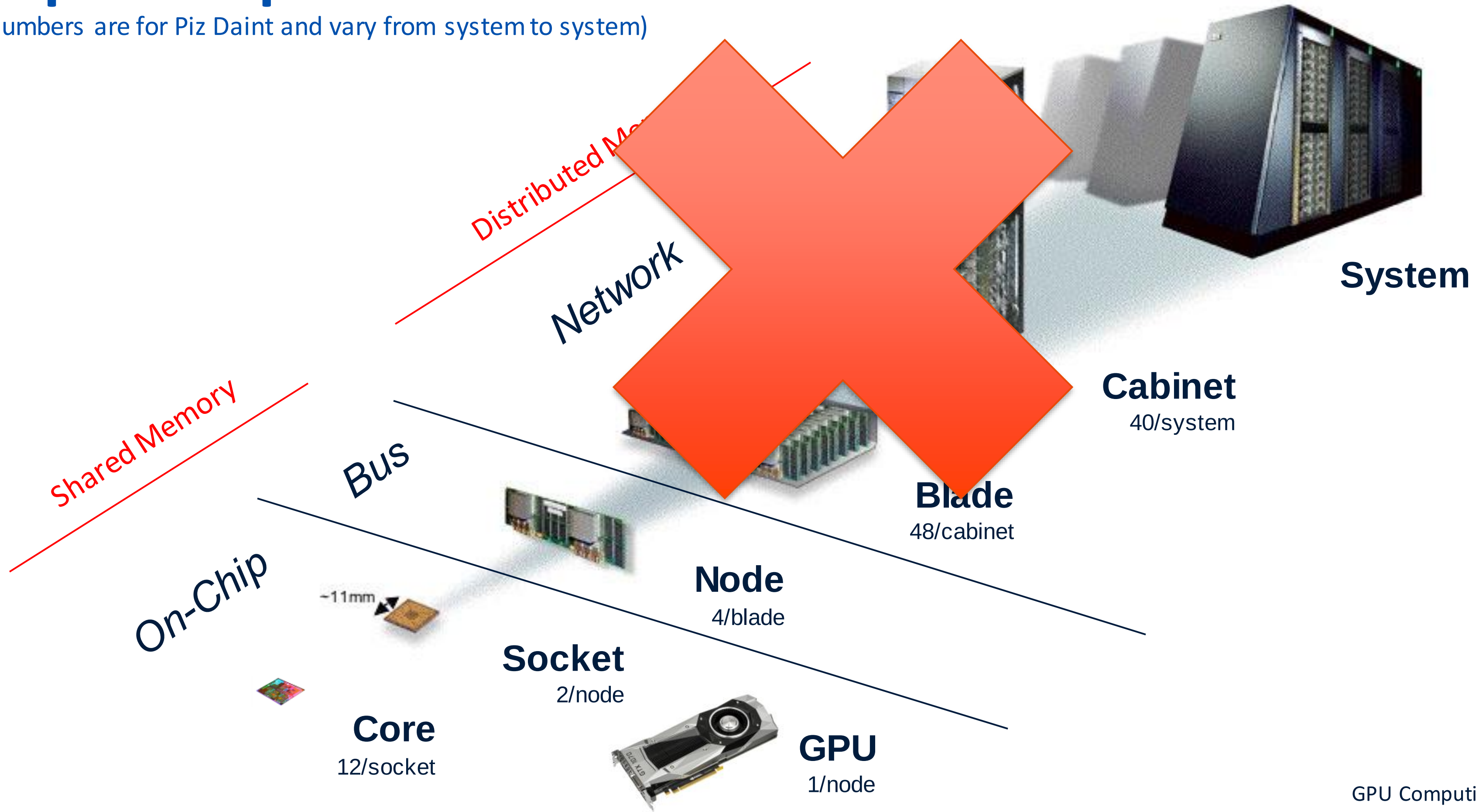
High Performance Computing for Weather and Climate (HPC4WC)

Content: Graphics Processing Units
Lecturer: Tobias Wicky
Block course 701-1270-00L
Summer 2022



Supercomputer Architecture

(Numbers are for Piz Daint and vary from system to system)

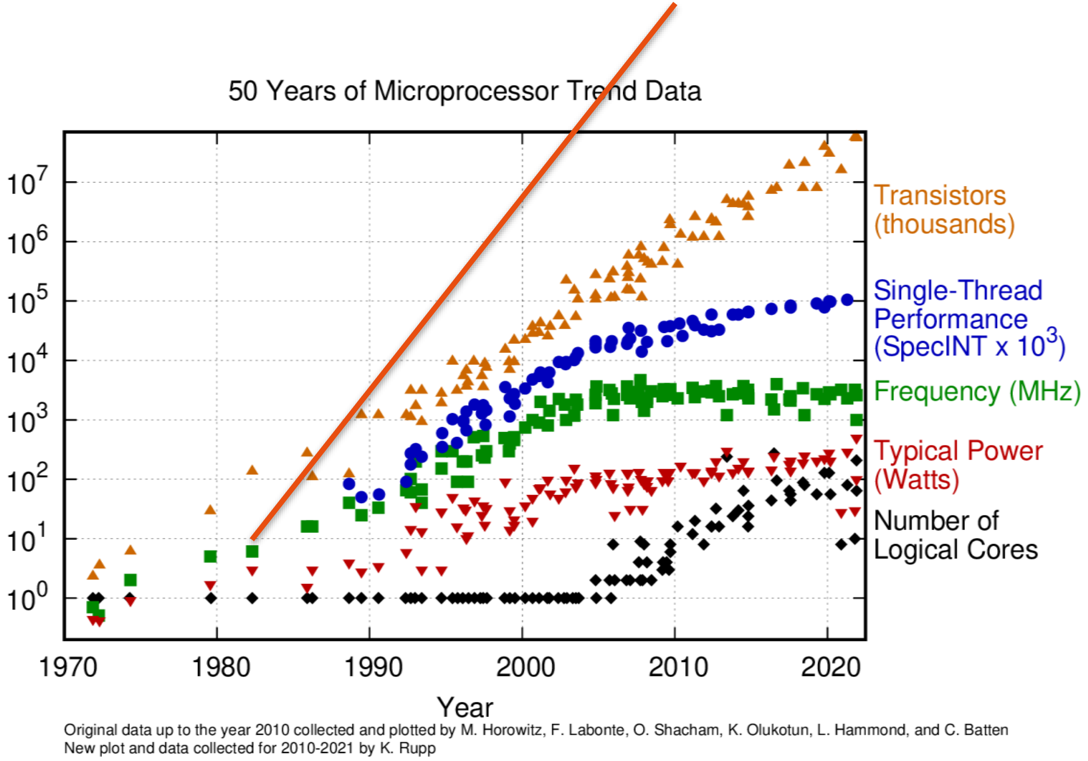


Learning goals

- 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

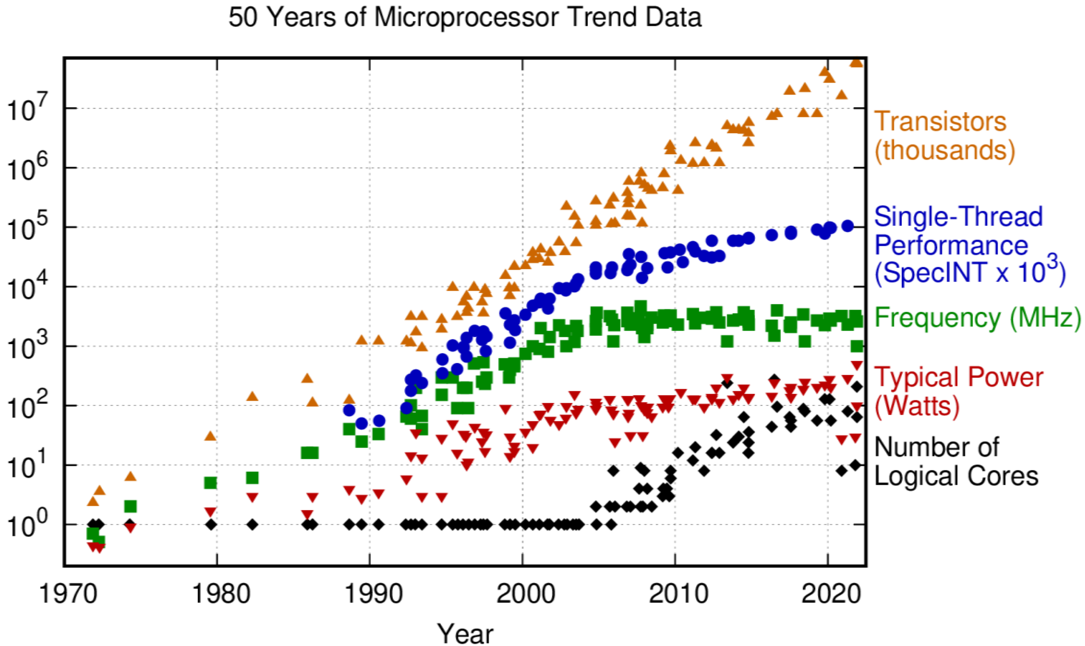
Moore's Law

- "The number of transistors in a dense integrated circuit will double every two years"



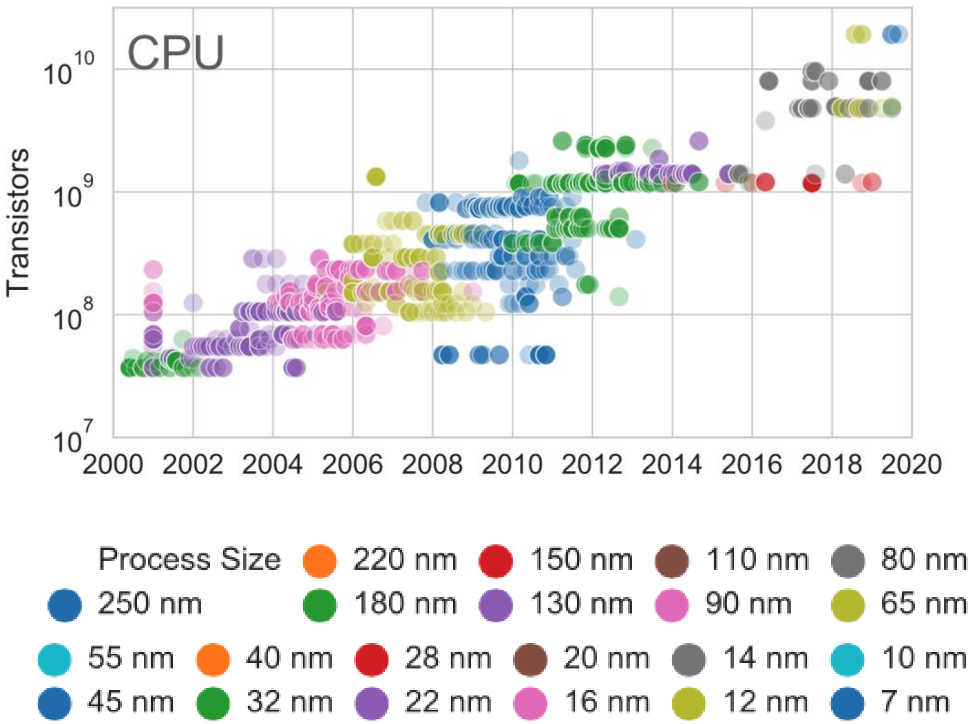
Dennard Scaling

- "If the transistor density doubles, power consumption (with twice the number of transistors) stays the same."

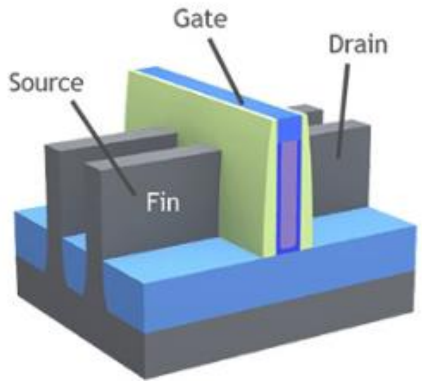


Original data up to the year 2010 collected and plotted by M. Horowitz, F. Labonte, O. Shacham, K. Olukotun, L. Hammond, and C. Batten
New plot and data collected for 2010-2021 by K. Rupp

The End of General Purpose Computing

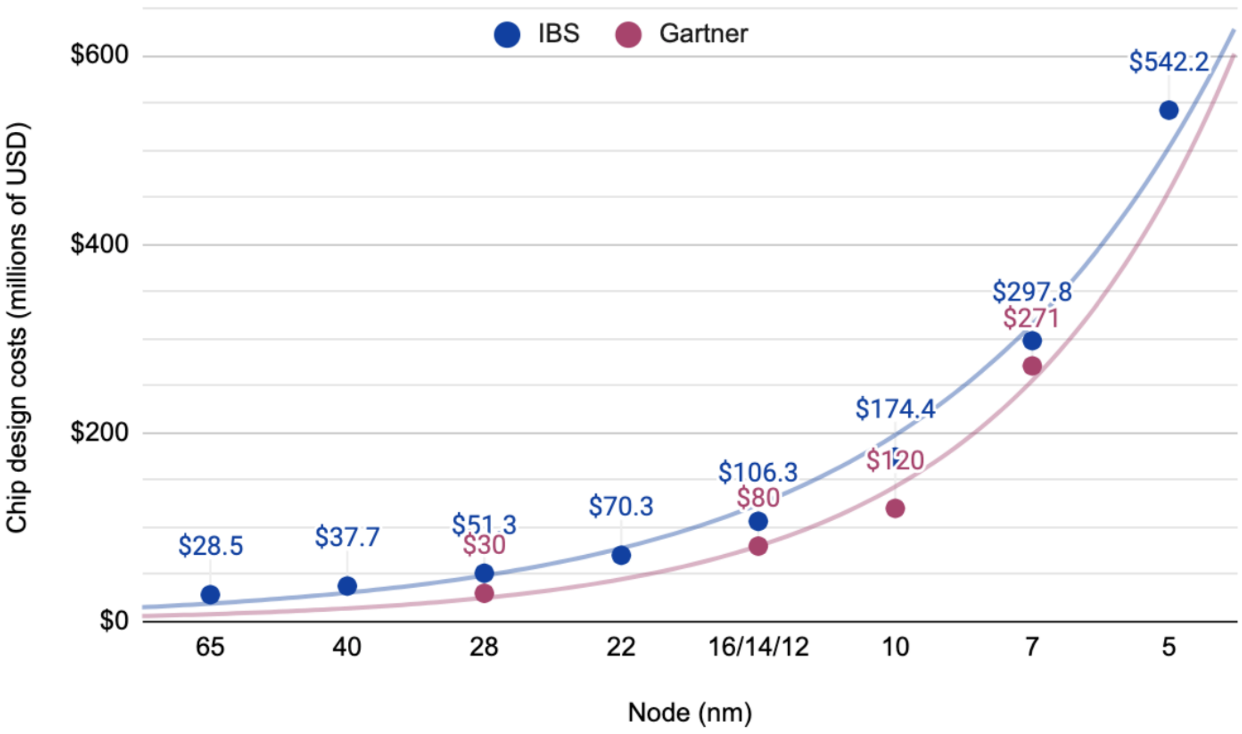


Transistor



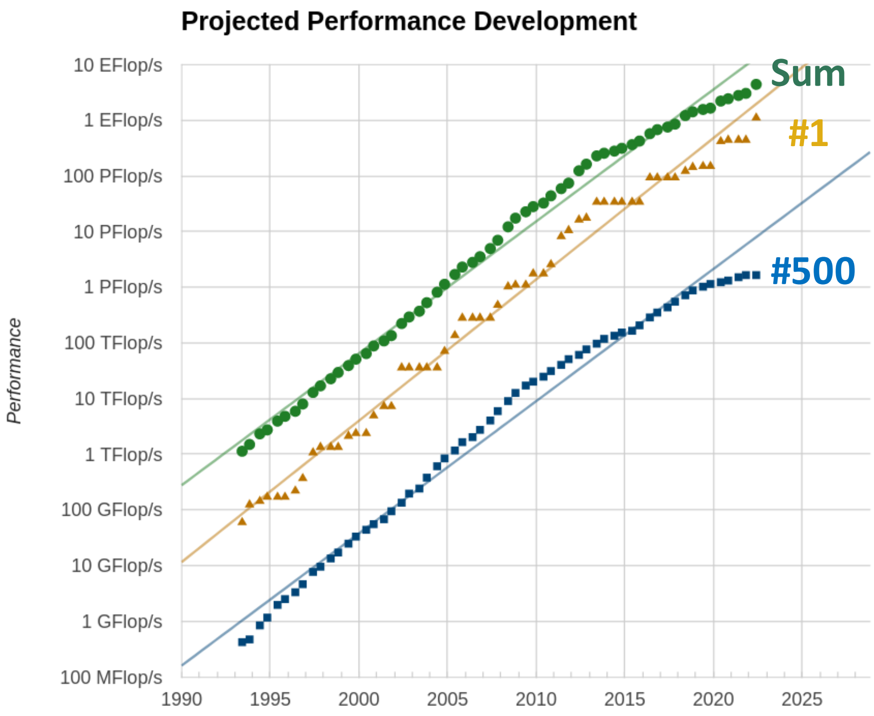
Distance between Si-atoms is 0.5 nm!

Chip Design Costs



Source: top500.org

How does performance of our machines behave

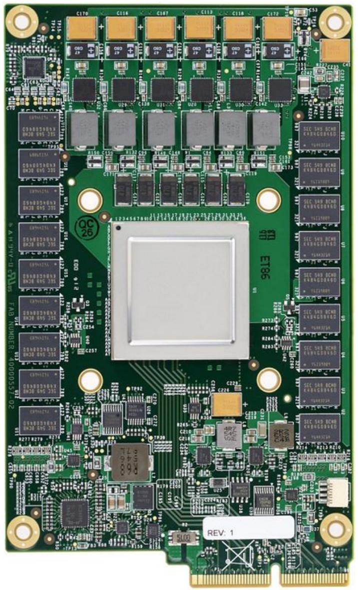


Source: top500.org

So why are we still ok?

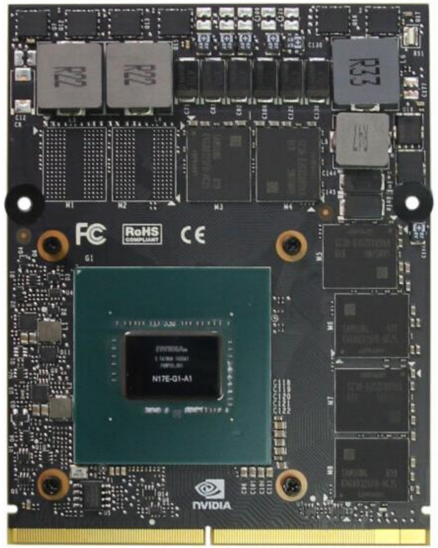
Rank	System	Cores	Rmax (PFlop/s)	Rpeak (PFlop/s)	Power (kW)	
1	**Frontier** - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, **HPE** DOE/SC/Oak Ridge National Laboratory United States	8,730,112	1,102.00	1,685.65	21,100	AMD GPU
2	**Supercomputer Fugaku** - Supercomputer Fugaku, A64FX 48C 2.2GHz, Tofu interconnect D, **Fujitsu** RIKEN Center for Computational Science Japan	7,630,848	442.01	537.21	29,899	
3	**LUMI** - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, **HPE** EuroHPC/CSC Finland	1,110,144	151.90	214.35	2,942	AMD GPU
4	**Summit** - IBM Power System AC922, IBM POWER9 22C 3.07GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband, **IBM** DOE/SC/Oak Ridge National Laboratory United States	2,414,592	148.60	200.79	10,096	NVIDIA GPU
5	**Sierra** - IBM Power System AC922, IBM POWER9 22C 3.1GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband, **IBM / NVIDIA / Mellanox** DOE/NNSA/LLNL United States	1,572,480	94.64	125.71	7,438	NVIDIA GPU
6	**Sunway TaihuLight** - Sunway MPP, Sunway SW26010 260C 1.45GHz, Sunway, **NRCPC** National Supercomputing Center in Wuxi China	10,649,600	93.01	125.44	15,371	
7	**Perlmutter** - HPE Cray EX235n, AMD EPYC 7763 64C 2.45GHz, NVIDIA A100 SXM4 40 GB, Slingshot-10, **HPE** DOE/SC/LBNL/NERSC United States	761,856	70.87	93.75	2,589	NVIDIA GPU
8	**Selene** - NVIDIA DGX A100, AMD EPYC 7742 64C 2.25GHz, NVIDIA A100, Mellanox HDR Infiniband, **Nvidia** NVIDIA Corporation United States	555,520	63.46	79.22	2,646	NVIDIA GPU
9	**Tianhe-2A** - TH-IVB-FEP Cluster, Intel Xeon E5-2692v2 12C 2.2GHz, TH Express-2, Matrix-2000, **NUDT** National Super Computer Center in Guangzhou China	4,981,760	61.44	100.68	18,482	Xeon Phi
10	**Adastra** - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, **HPE** Grand Equipement National de Calcul Intensif - Centre Informatique National de l'Enseignement Suprieur (GENCI-CINES) France	319,072	46.10	61.61	921	AMD GPU

Specialized Chips are on the Rise!



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

FPGA
(e.g. bitcoin mining)



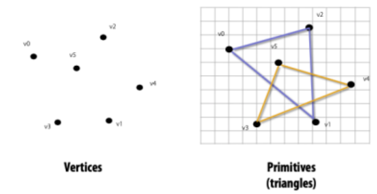
GPU
(e.g. gaming)

Who has experience with programming GPUs?

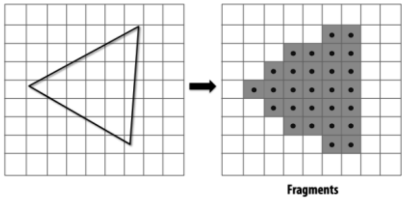
GPU’s are for gaming, right?

GPU Computing

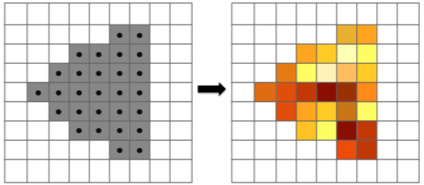
Primitive computation



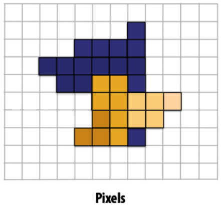
Rasterization



Fragment processing



Pixel operations



Parallel operations!

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

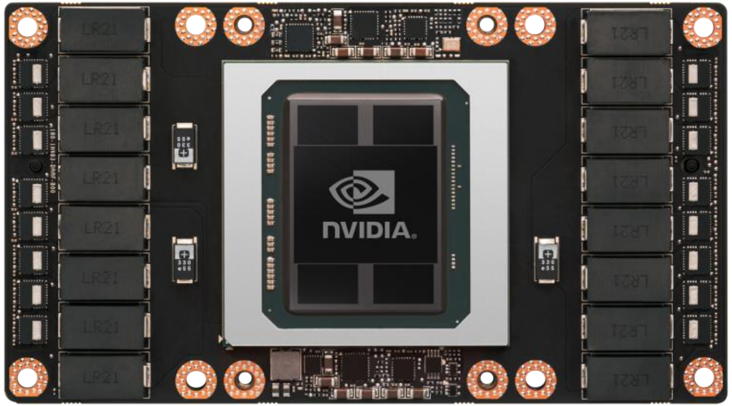
With *OpenGL*, *OpenCL*, *CUDA*, … programming languages started to appear that made general purpose computing on GPUs possible.

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

Performance / Watt

Intel Xeon E5-2690 v3 + DRAM

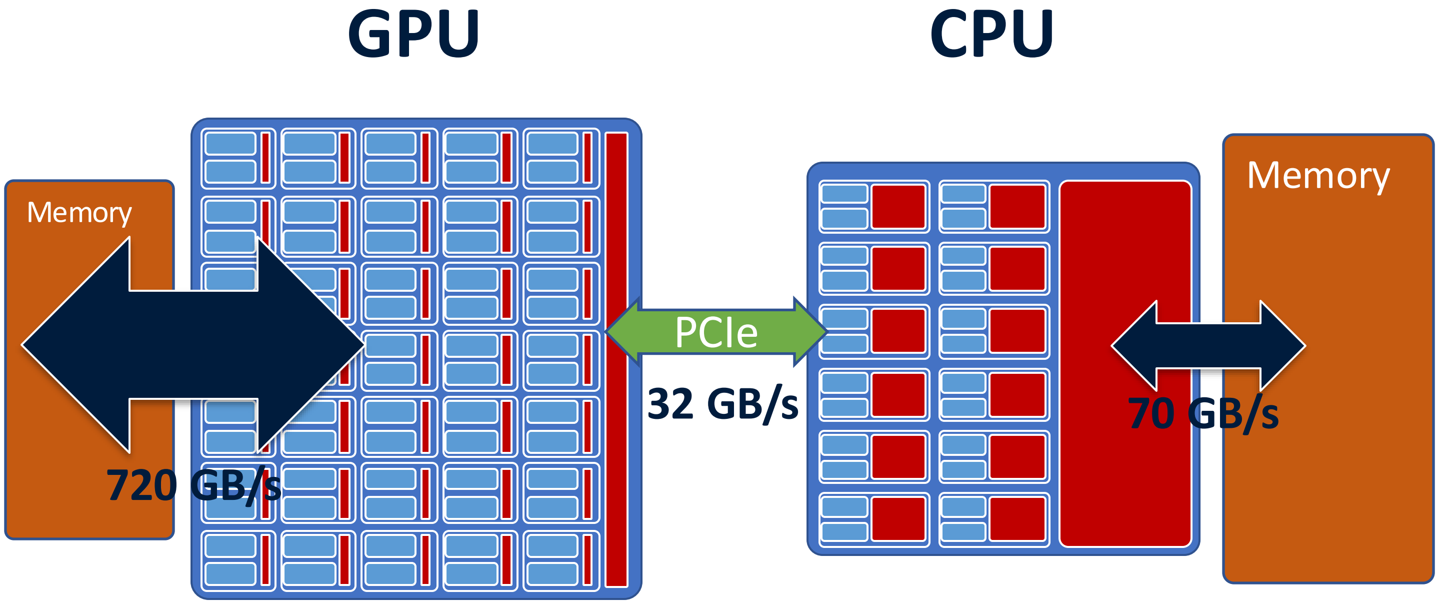
NVIDIA Tesla P100



~ 200 W   0.5 TFLOP/s   70 GB/s

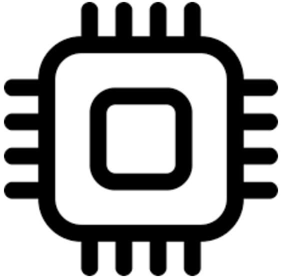
~ 300 W   5.3 TFLOP/s   720 GB/s

Node Architecture

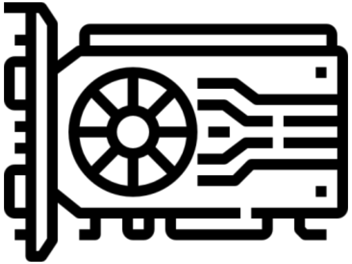


Crucial to minimize memory transfers between CPU and GPU!

CPU vs. GPU



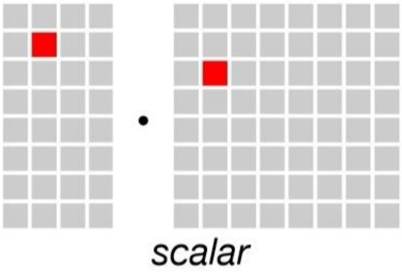
Architecture



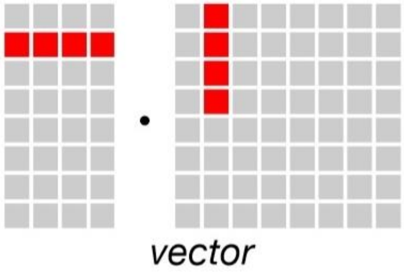
Latency

Optimization

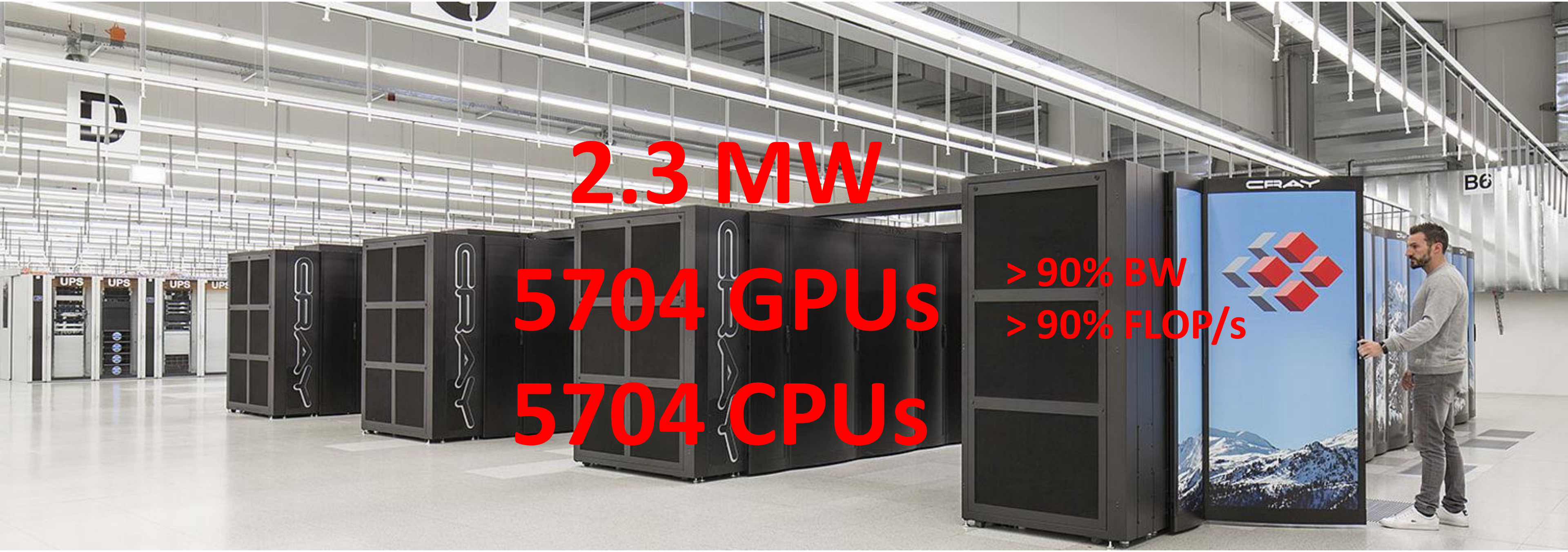
Bandwidth



Compute primitive



Hybrid Supercomputer



Source: www.eia.gov

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 MWh / SY * 0.088 SY / day = 16874 kWh / day

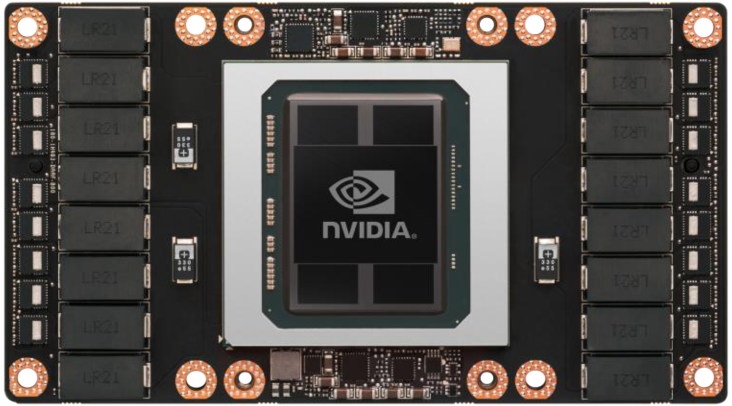
Average electricity consumption for one household ~ 39 kWh / day

Performance / Watt

Intel Xeon E5-2690 v3 + DRAM

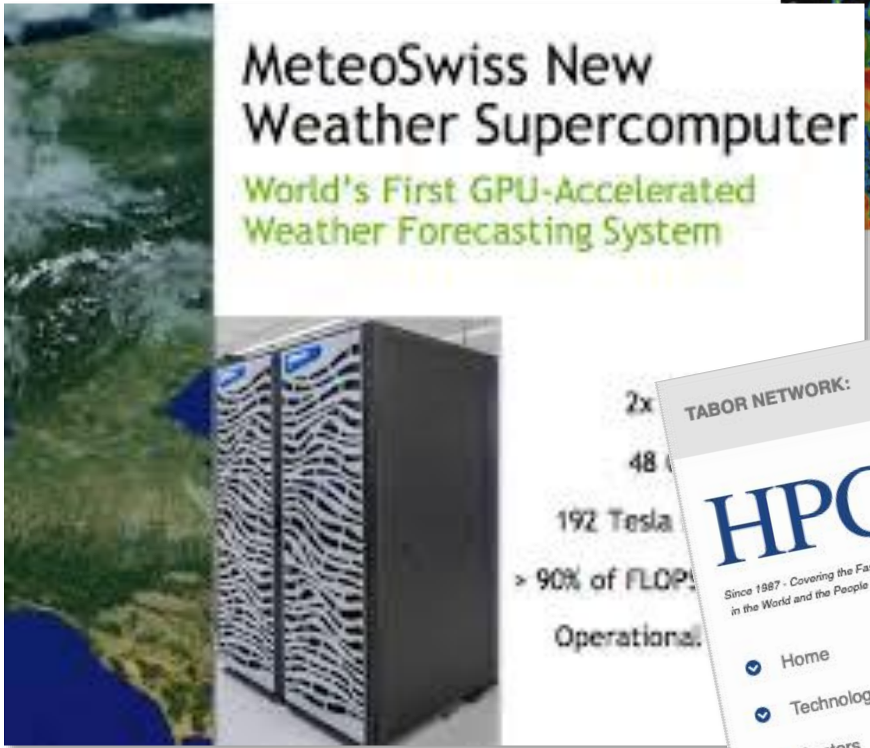
~ 200 W 0.5 TFLOP/s 70 GB/s

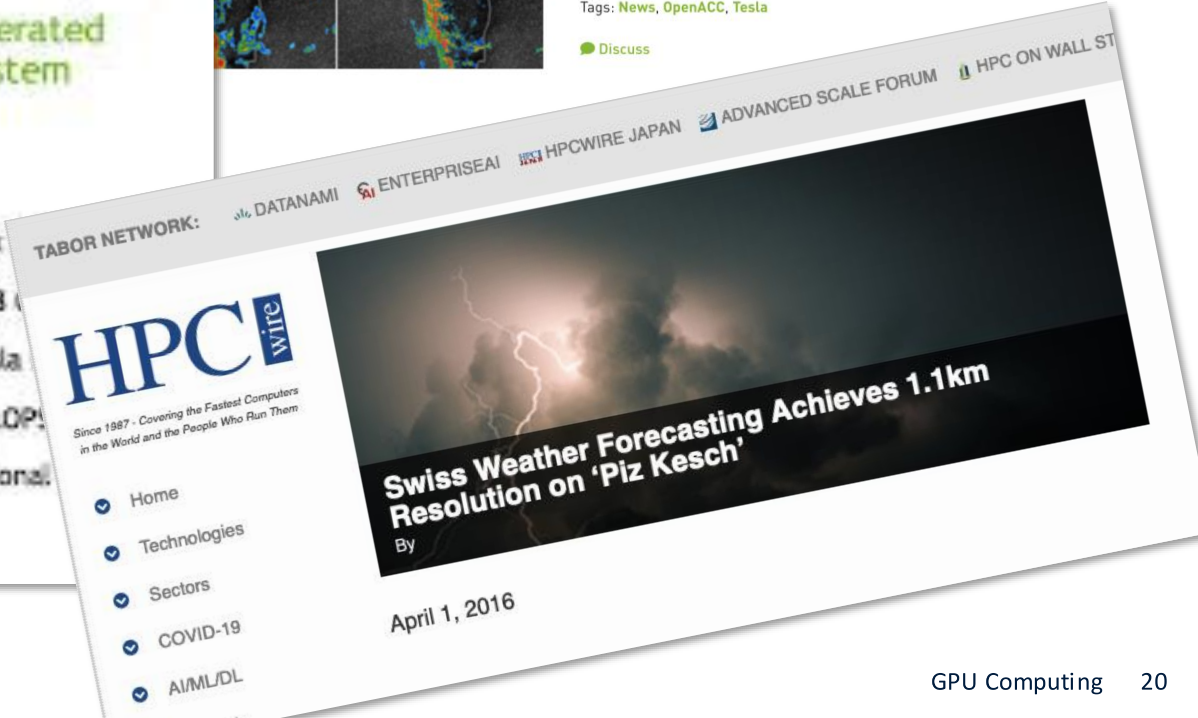
NVIDIA Tesla P100



~ 300 W 5.3 TFLOP/s 720 GB/s

Weather and Climate on GPUs





Questions?

Lab Exercises

01-GPU-programming-cupy.ipynb

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

Note. You will be asked to write a GPU version of stencil2d.py. To be able to use Matplotlib to plot the results, please issue this command from a terminal:

```
export PYTHONPATH=/users/classXXX/HPC4WC_venv/lib/python3.8/site-packages:$PYTHONPATH
```

Remarks

- When running a GPU notebook, you may experience this error:

  cupy_backends.cuda.api.runtime.CUDARuntimeError: cudaErrorDevicesUnavailable: all CUDA-capable devices are busy or unavailable

  Don’t worry, it’s not your fault! Just restart the kernel and the error should disappear. If it persists, reach out to us.

- To let multiple tasks access the same GPU: `export CRAY_CUDA_MPS=1`

Let’s go!