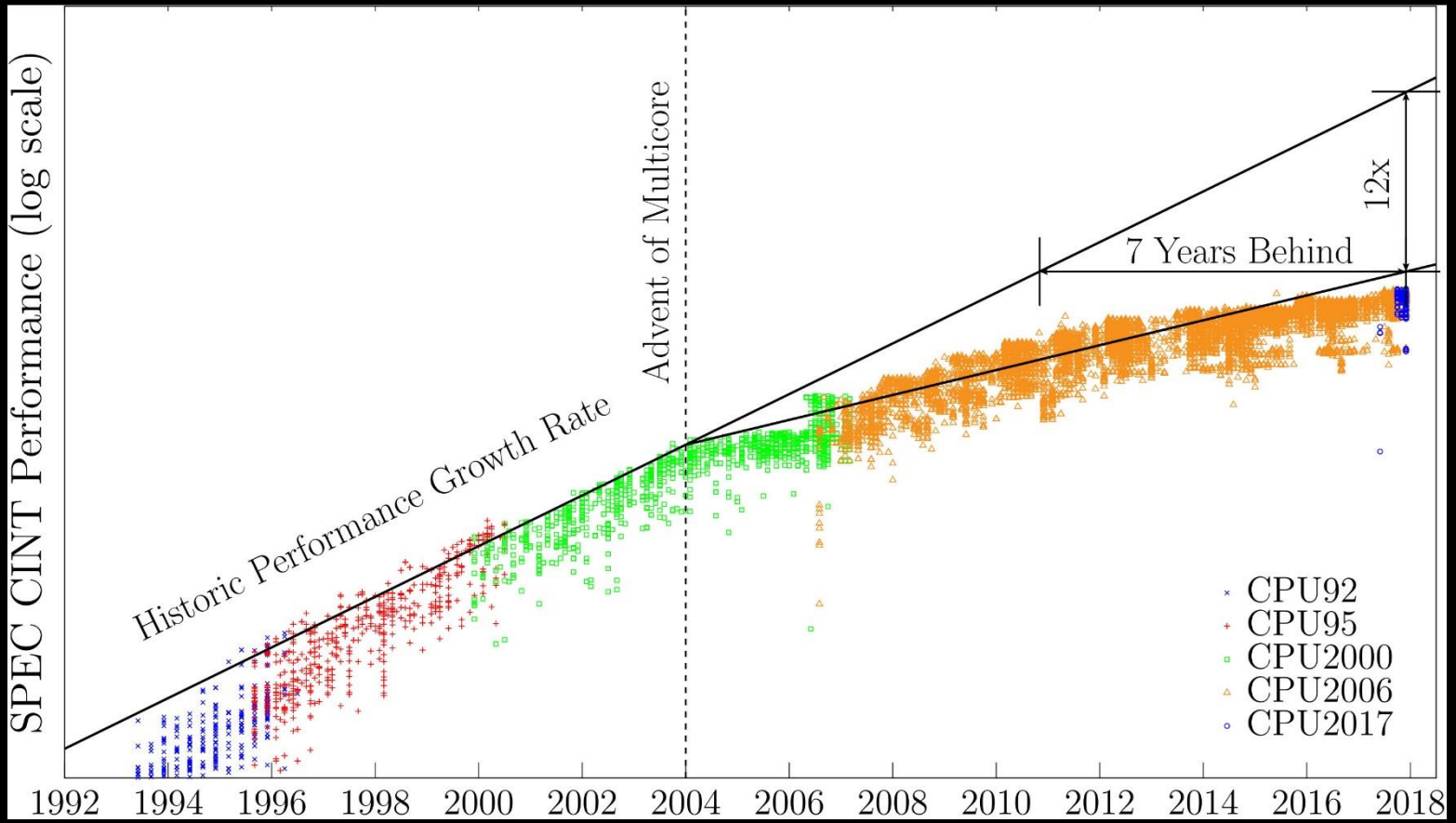
It's a Multicore World

John Urbanic
Parallel Computing Scientist
Pittsburgh Supercomputing Center

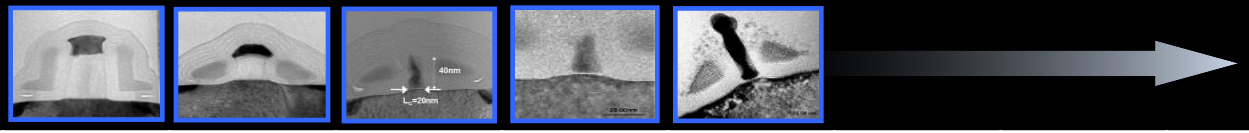
Moore's Law abandoned serial programming around 2004



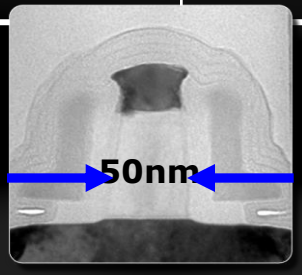
Courtesy Liberty Computer Architecture Research Group

But Moore's Law is only beginning to stumble now.

Intel process technology capabilities

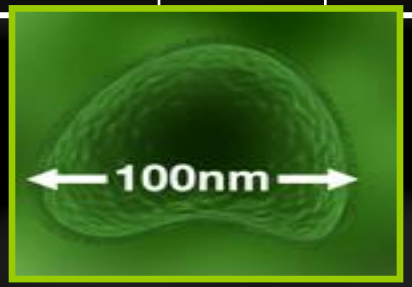


High Volume Manufacturing	2004	2006	2008	2010	2012	2014	2018	2021
Feature Size	90nm	65nm	45nm	32nm	22nm	14nm	10nm	7nm
Integration Capacity (Billions of Transistors)	2	4	8	16	32	64	128	256



Transistor for 90nm Process

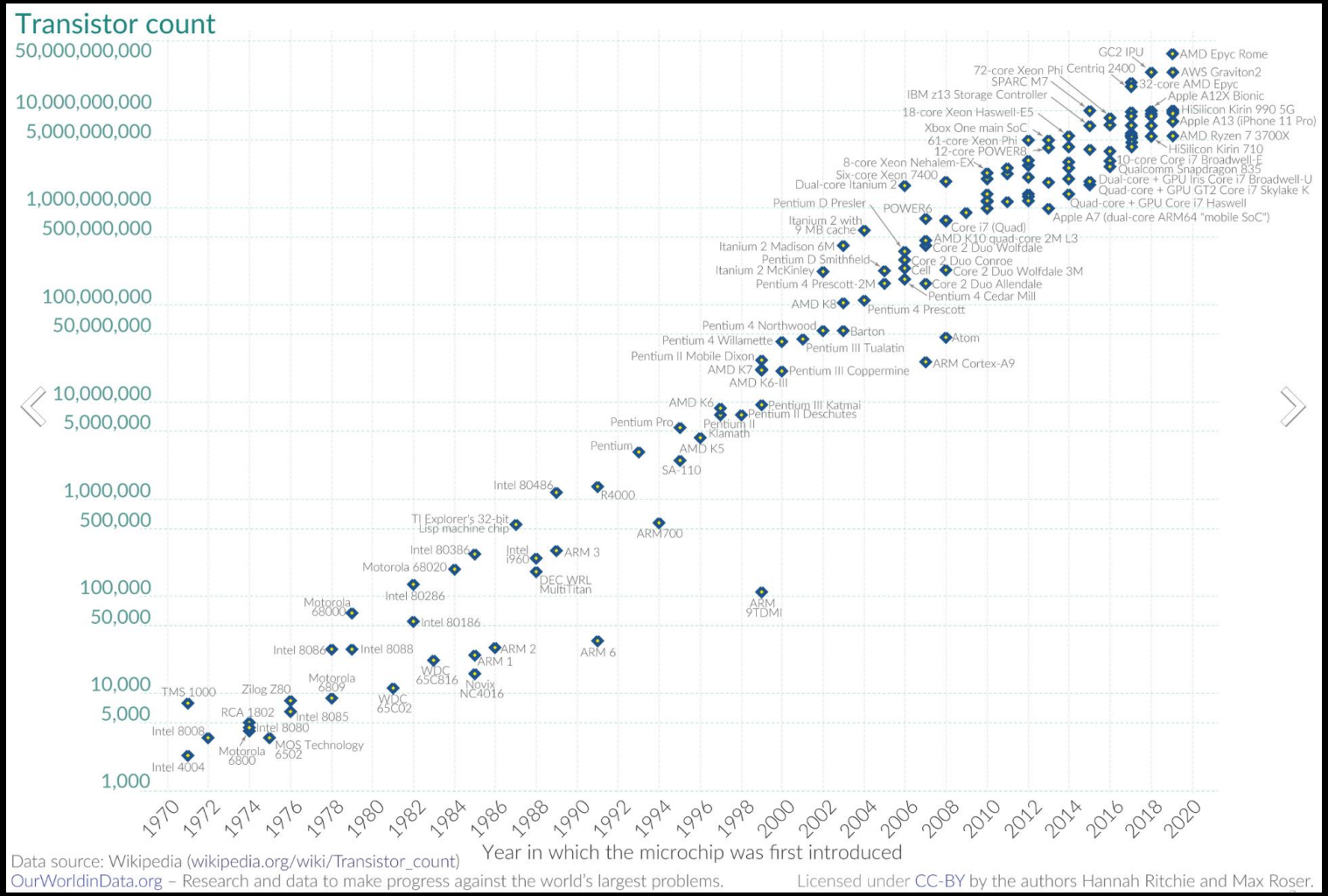
Source: Intel



Influenza Virus

Source: CDC

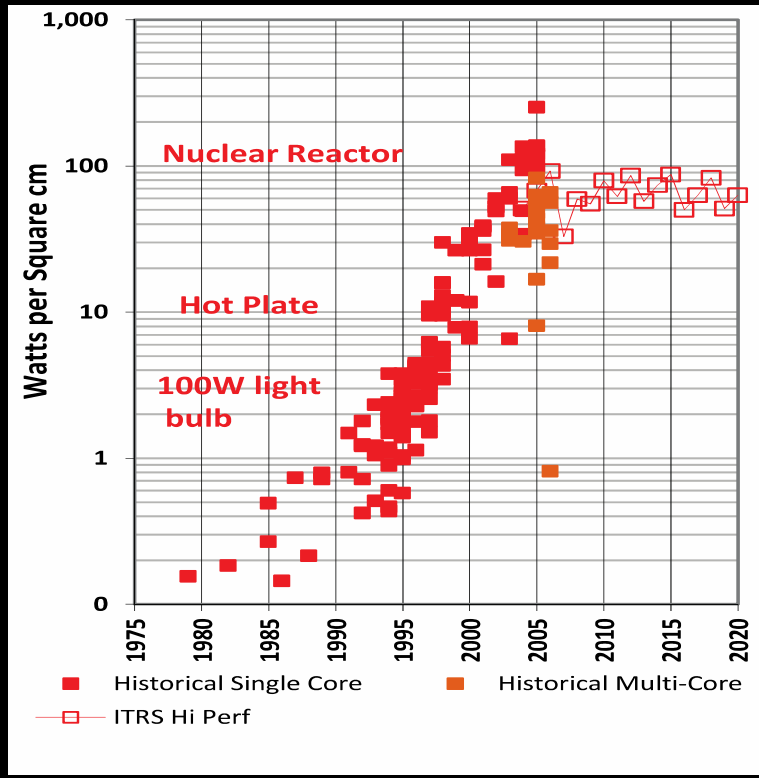
And at end of day we keep using getting more transistors.

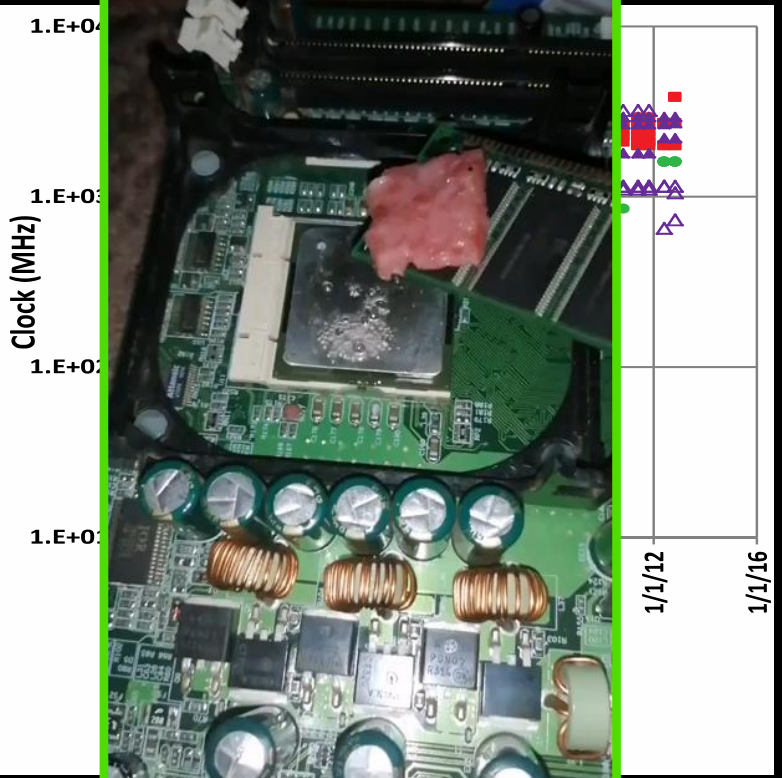


Data source: Wikipedia (wikipedia.org/wiki/Transistor_count)

OurWorldinData.org – Research and data to make progress against the world's largest problems.

That Power and Clock Inflection Point in 2004… didn't get better.

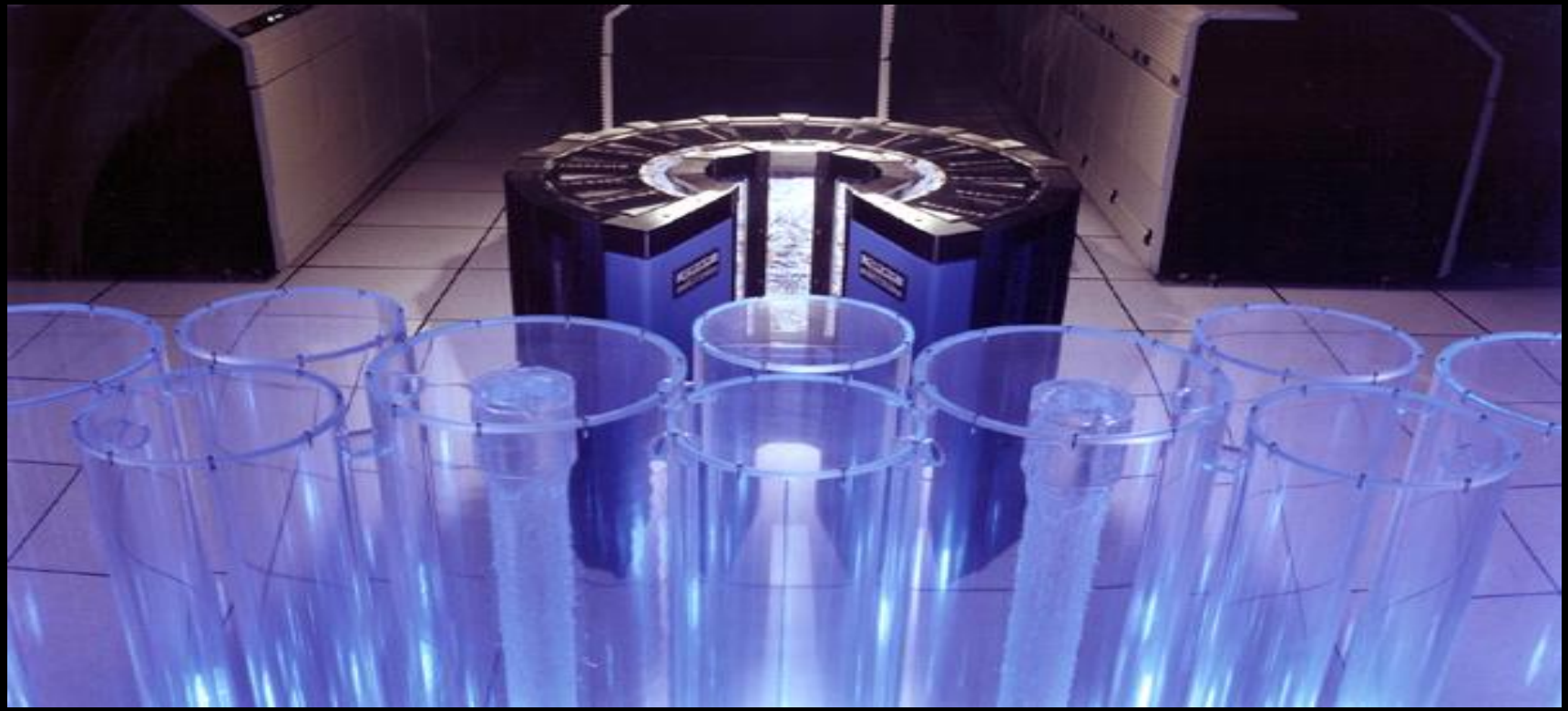




Fun fact: At 100+ Watts and <1V, currents are beginning to exceed 100A at the point of load.

Courtesy Horst Simon, LBNL

Not a new problem, just a new scale…



Cray-2 with cooling tower in foreground, circa 1985

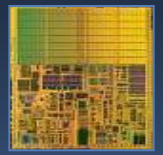
And how to get more performance from more transistors with the same power.

A 15% Reduction In Voltage Yields

RULE OF THUMB

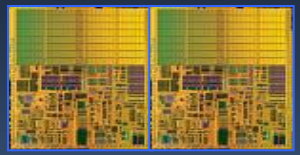
Frequency Reduction	Power Reduction	Performance Reduction
15%	45%	10%

SINGLE CORE



Area = 1
Voltage = 1
Freq = 1
Power = 1
Perf = 1

DUAL CORE

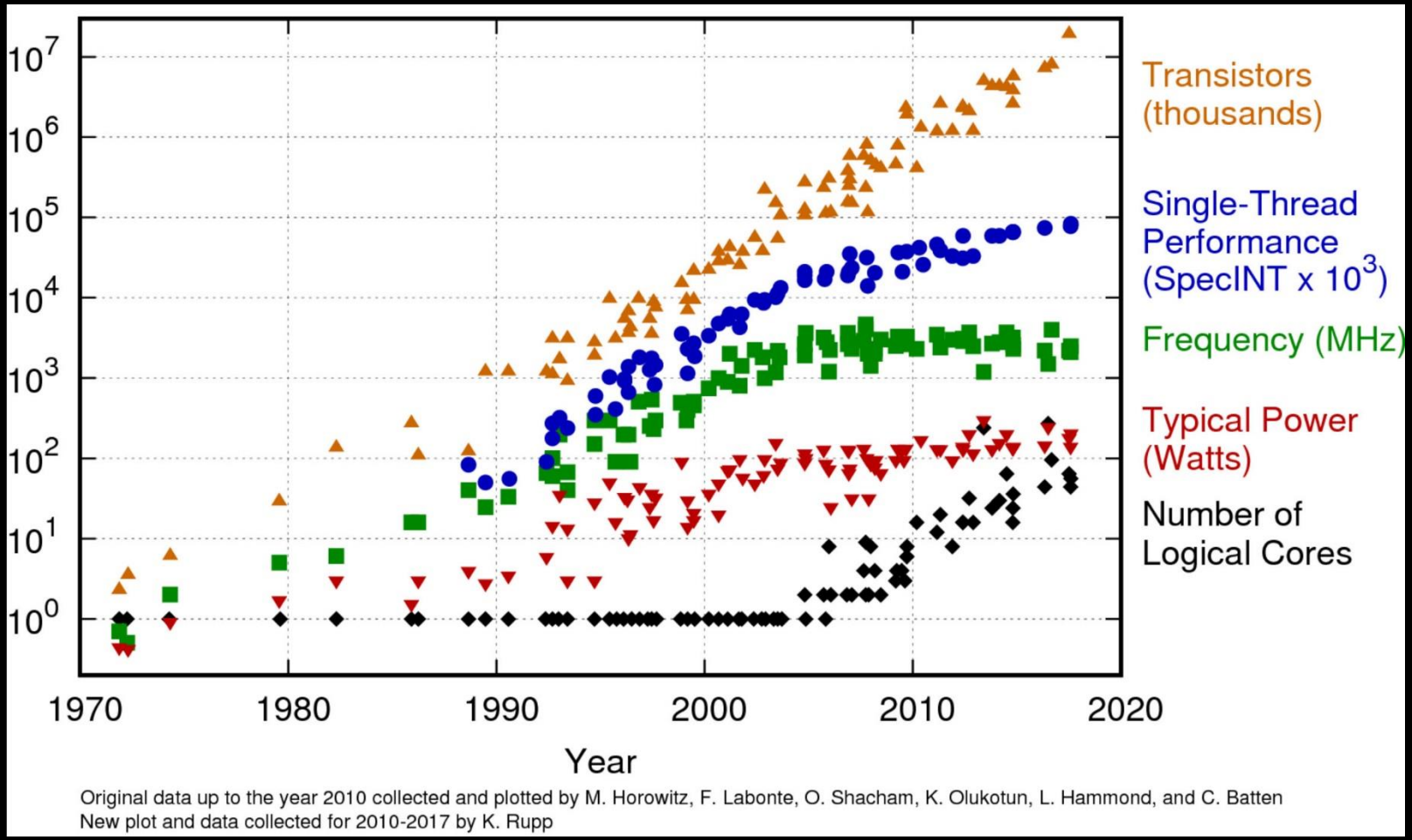


Area = 2
Voltage = 0.85
Freq = 0.85
Power = 1
Perf = ~1.8

Single Socket Parallelism

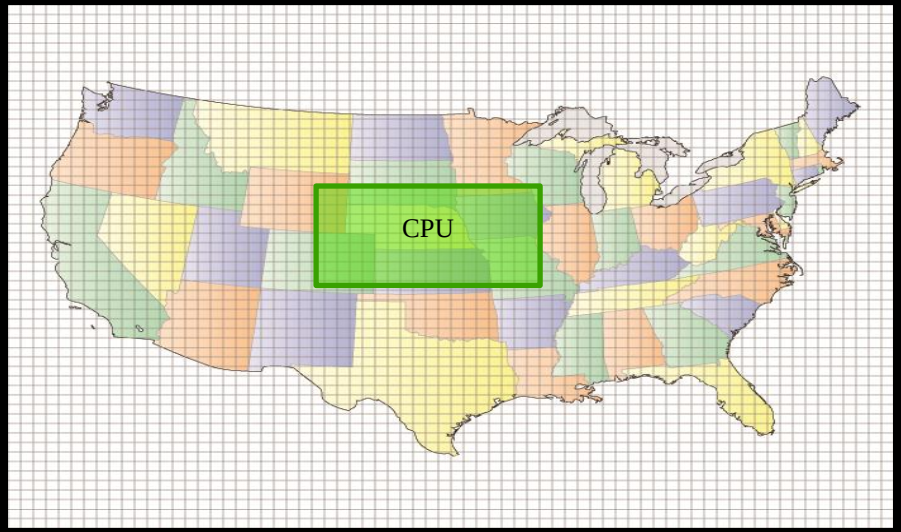
Processor	Year	Vector	Bits	SP FLOPs / core / cycle	Cores	FLOPs/cycle
Pentium III	1999	SSE	128	3	1	3
Pentium IV	2001	SSE2	128	4	1	4
Core	2006	SSE3	128	8	2	16
Nehalem	2008	SSE4	128	8	10	80
Sandybridge	2011	AVX	256	16	12	192
Haswell	2013	AVX2	256	32	18	576
KNC	2012	AVX512	512	32	64	2048
KNL	2016	AVX512	512	64	72	4608
Skylake	2017	AVX512	512	96	28	2688

Putting It All Together

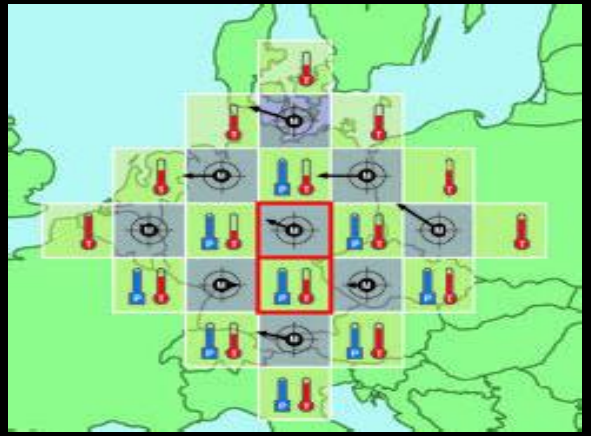


Original data up to the year 2010 collected and plotted by M. Horowitz, F. Labonte, O. Shacham, K. Olukotun, L. Hammond, and C. Batten
New plot and data collected for 2010-2017 by K. Rupp

Prototypical Application: Serial Weather Model

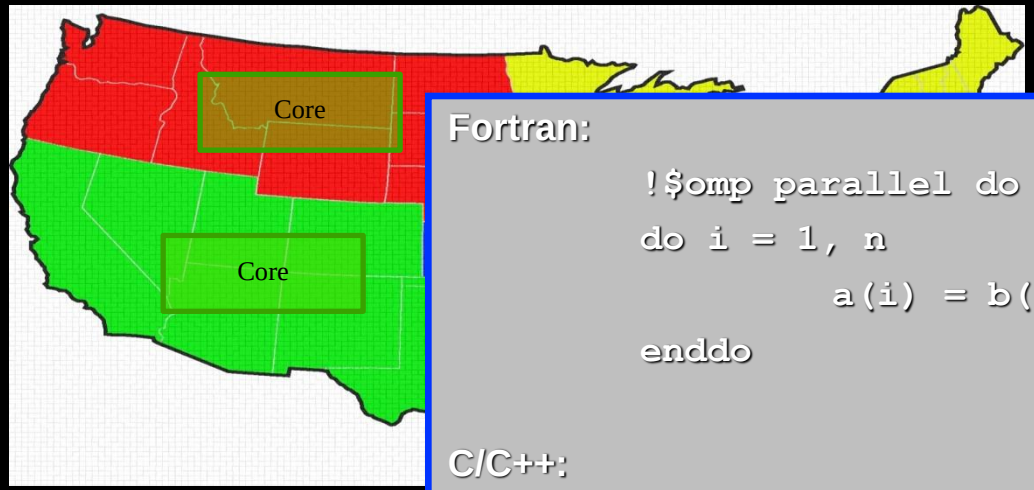


First Parallel Weather Modeling Algorithm: Richardson in 1917



Courtesy John Burkhardt, Virginia Tech

Weather Model: Shared Memory (OpenMP)



Four meteorologists in t

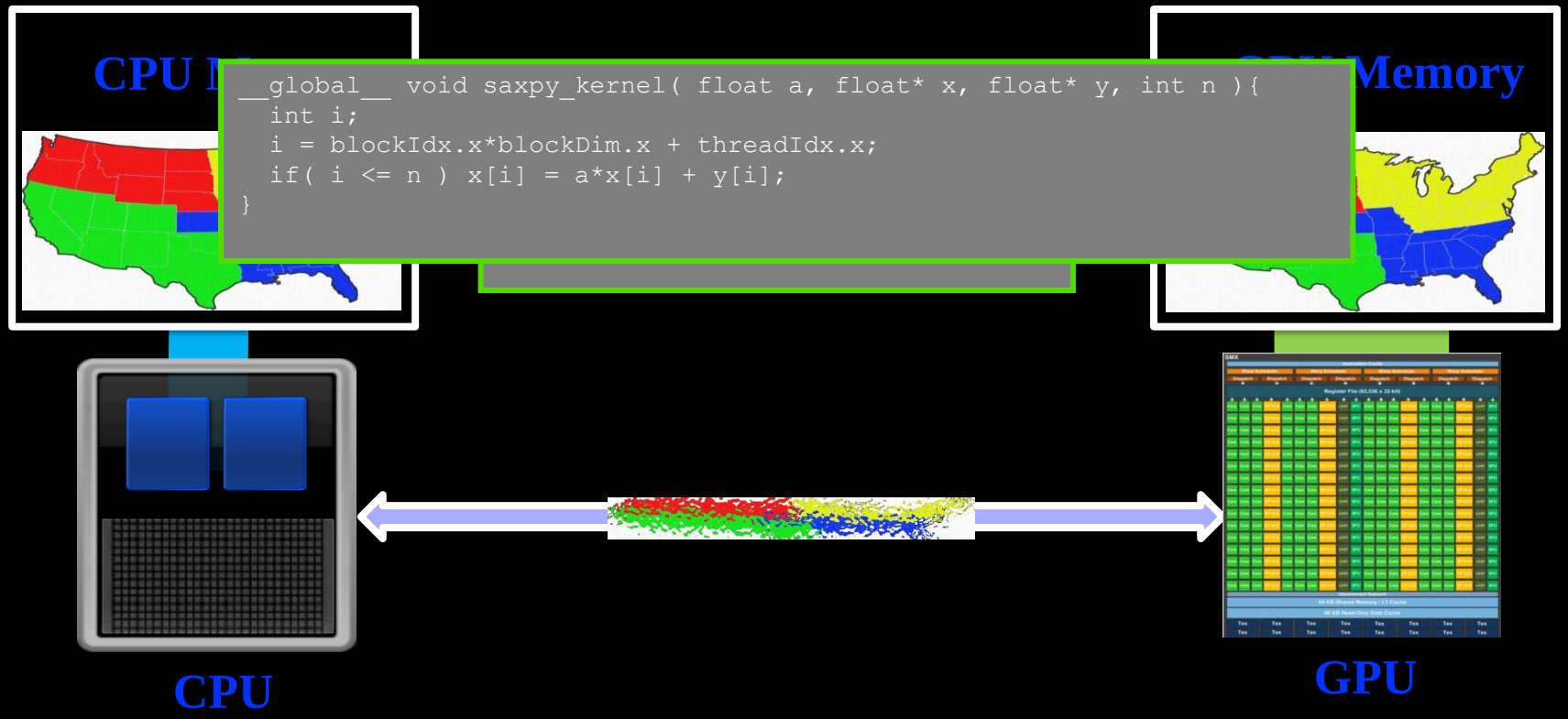
Fortran:

```
!$omp parallel do
do i = 1, n
        a(i) = b(i) + c(i)
enddo
```

C/C++:

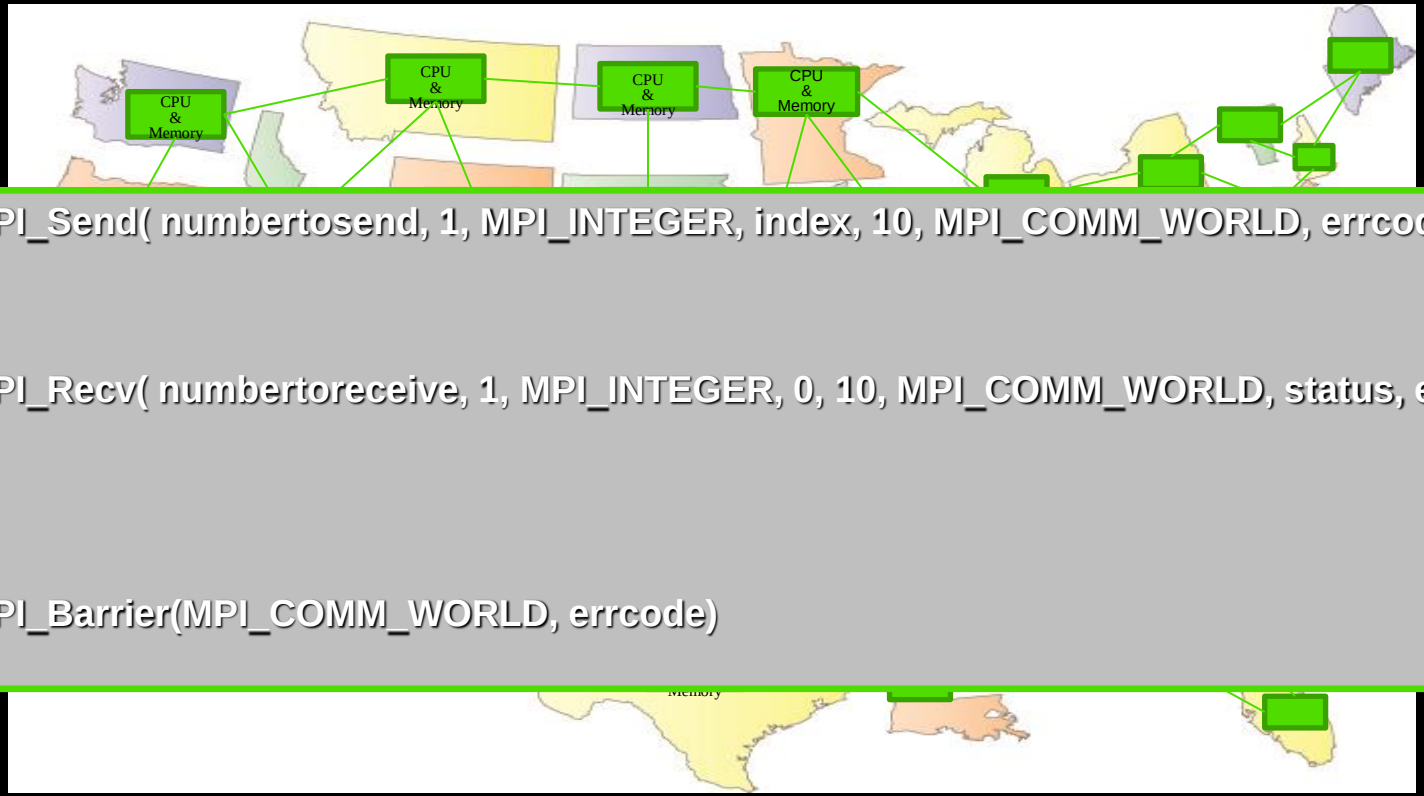
```
#pragma omp parallel for
for(i=1; i<=n; i++)
        a[i] = b[i] + c[i];
```

Weather Model: Accelerator (OpenACC)



1 meteorologists coordinating 1000 math savants using tin cans and a string.

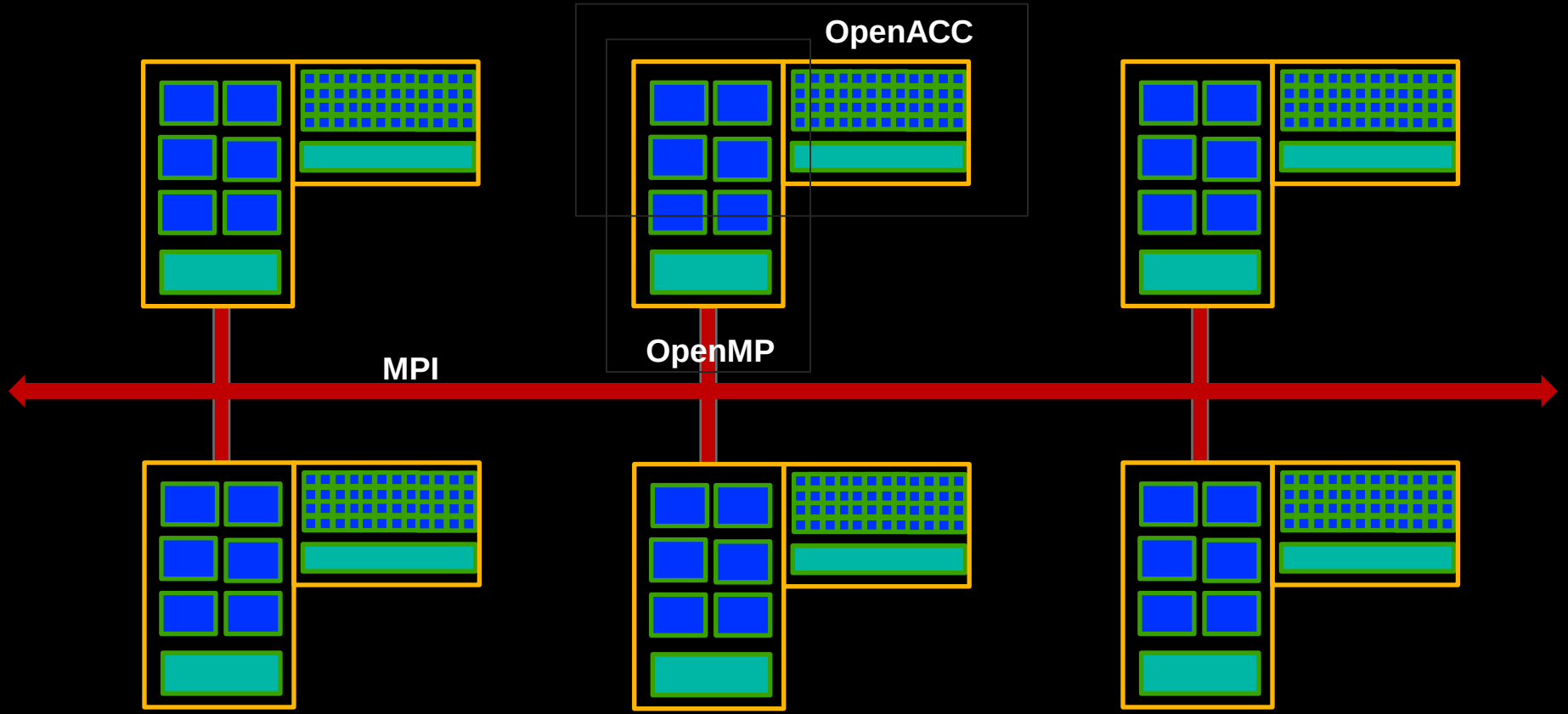
Weather Model: Distributed Memory (MPI)




```
call MPI_Send( numbertosend, 1, MPI_INTEGER, index, 10, MPI_COMM_WORLD, errcode)
.
.
call MPI_Recv( numbertoreceive, 1, MPI_INTEGER, 0, 10, MPI_COMM_WORLD, status, errcode)
.
.
.
call MPI_Barrier(MPI_COMM_WORLD, errcode)
.
```

50 meteorologists using a telegraph.

The pieces fit like this…



Many Levels and Types of Parallelism

- Vector (SIMD)
- Instruction Level (ILP)
  - Instruction pipelining
  - Superscaler (multiple instruction units)
  - Out-of-order
  - Register renaming
  - Speculative execution
  - Branch prediction

Compiler
(not your problem)

OpenMP

- Multi-Core (Threads)
- SMP/Multi-socket

OpenACC

- Accelerators: GPU & MIC

MPI

- Clusters
- MPPs

Also Important

- ASIC/FPGA/DSP
- RAID/IO

Cores, Nodes, Processors, PEs?

- The most unambiguous way to refer to the smallest useful computing device is as a Processing Element, or PE.
- This is usually the same as a single core.
- “Processors” usually have more than one core – as per the previous list.
- “Nodes” is commonly used to refer to an actual physical unit, most commonly a circuit board or blade with a network connection. These often have multiple processors.

I will try to use the term PE consistently here, but I may slip up myself. Get used to it as you will quite often hear all of the above terms used interchangeably where they shouldn’t be.

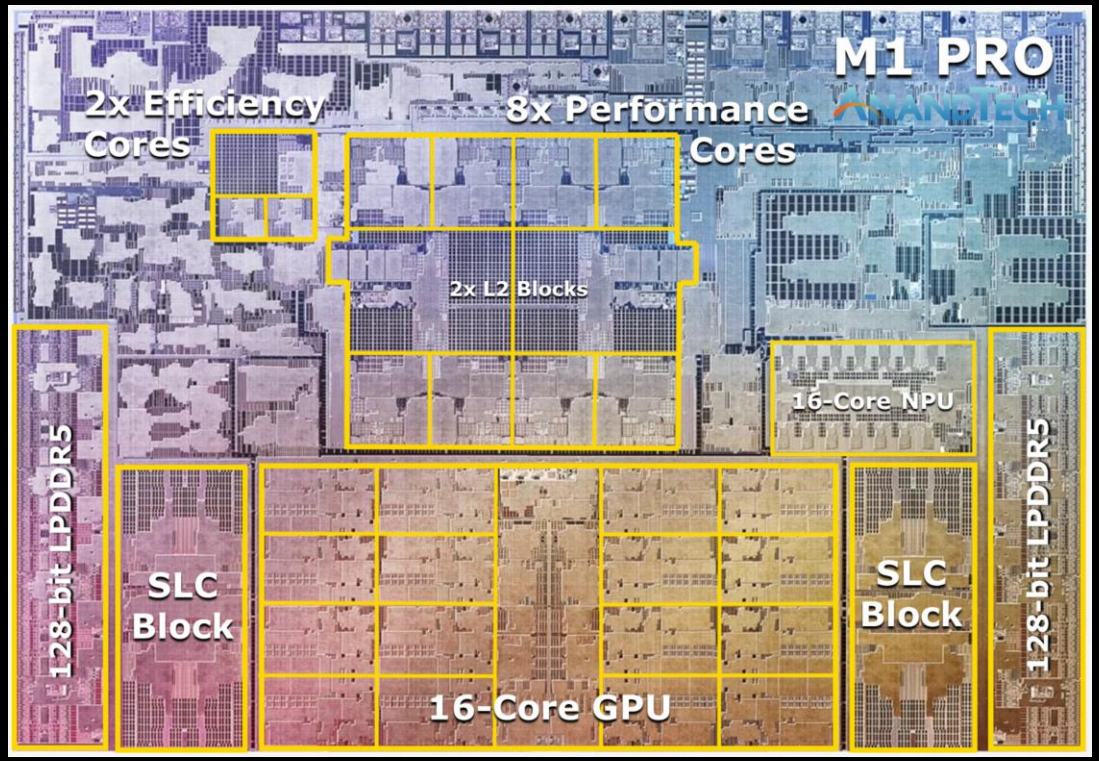
Top 10 Systems as of November 2024

#	Computer	Site	Manufacturer	CPU Interconnect [*Accelerator*]	Cores	Rmax (Pflops)	Rpeak (Pflops)	Power (MW)
1	**El Capitan**	Lawrence Livermore National Laboratory **United States**	HPE	AMD EPYC 24C 1.8GHz Slingshot-11 AMD Instinct MI300A	**11,039,616**	**1742**	**2746**	**30**
2	**Frontier**	Oak Ridge National Laboratory **United States**	HPE	AMD EPYC 64C 2GHz Slingshot-11 AMD Instinct MI250X	**9,066,176**	**1353**	**2055**	**25**
3	**Aurora**	Argonne National Laboratory **United States**	HPE	Intel Xeon Max 9470 52C 2.4GHz Slingshot-11 Intel Data Center GPU Max	**9,264,128**	**1012**	**1980**	**39**
4	**Eagle**	Microsoft **United States**	Microsoft	Intel Xeon 8480C 48C 2GHz Infiniband NDR NVIDIA H100	**1,123,200**	**561**	**846**	
5	**HPC6**	Eni S.p.A. **Italy**	HPE	AMD EPYC 64C 2GHz Slingshot-11 AMD Instinct MI250X	**3,143,520**	**477**	**606**	**8**
6	**Fugaku**	RIKEN Center for Computational Science **Japan**	Fujitsu	ARM 8.2A+ 48C 2.2GHz Torus Fusion Interconnect	**7,630,072**	**442**	**537**	**29**
7	**Alps**	Swiss National Supercomputing Center **Switzerland**	HPE	NVIDIA Grace 72C 3.1GHz Slingshot-11 *NVIDIA GH200*	**2,121,600**	**434**	**574**	**7**
8	**LUMI**	EuroHPC **Finland**	HPE	AMD EPYC 64C 2GHz Slingshot-11 AMD Instinct MI250X	**2,752,704**	**379**	**531**	**7**
9	**Leonardo**	EuroHPC **Italy**						**304**	**7**
10		Lawrenc							

500 ThinkSystem SR590, Xeon Gold 5218 16C 2.3GHz, 10G Ethernet, Lenovo Service Provider T China 108,800 2.31 4.00

The word is *Heterogeneous*

And it's not just supercomputers. It's on your desk, and in your phone.



AnandTech

How much of this can you program?

In Conclusion…