Blake Haugen & Nick McNutt
COSC 557
Fall '13

Visualizing Trends in Supercomputing

http://www.bhaugen.com/vis/top500vis.html

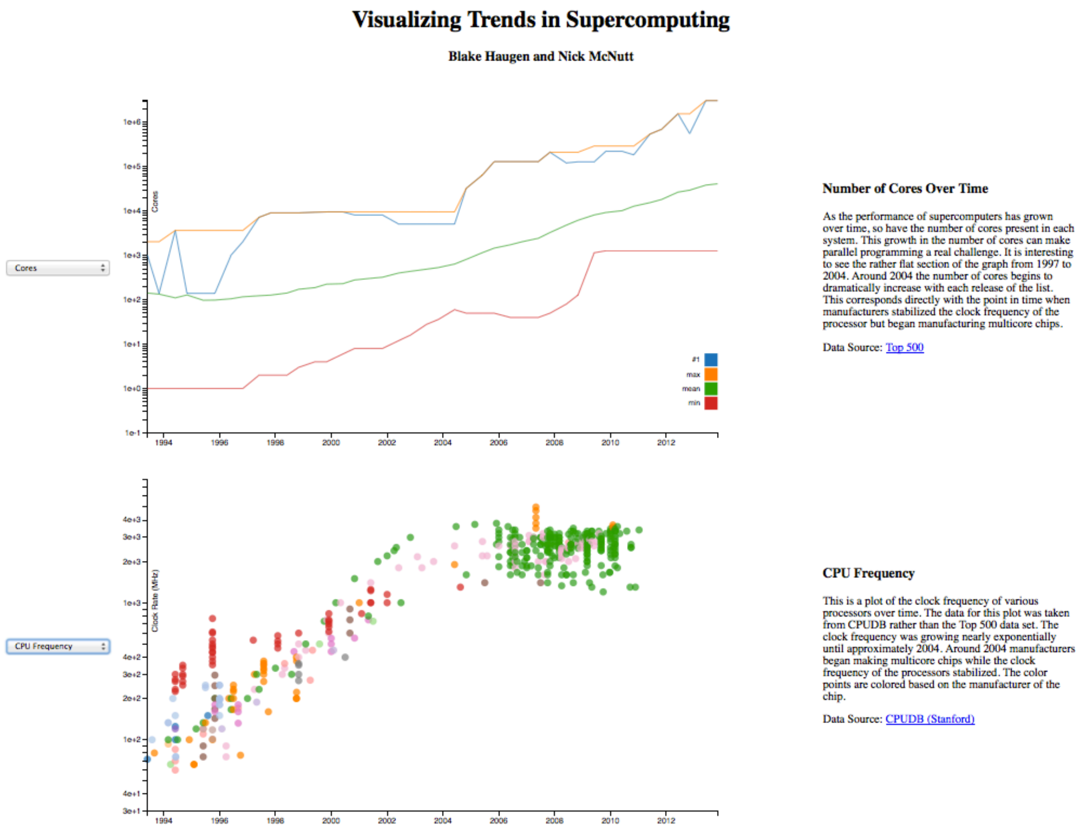
The last twenty years have shown extremely fast growth in the power of supercomputers. The Top 500 list compiled by Hans Meuer, Erich Strohmaier, Jack Dongarra, and Horst Simon has chronicled the rise of supercomputing as well as some many of the trends in the HPC community.

The majority of the data used in this project is taken directly from the Top 500 website. The list has been compiled every June and November since June of 1993. This list describes the performance of the 500 fastest supercomputers in the world and the characteristics of each system. We also used data from the CPUDB project at Stanford. While this is not part of the Top 500 project it shows how processor design and manufacturing has progressed over the years.

The goal of the project was to use the Top 500 data set in order to create a visualization tool that was accessible to novice users while still providing interesting insights for seasoned HPC veterans. The basic framework is a combination of two plots that are stacked on top of each other (see the screenshot below). The user has the ability to select change the plots that are shown in each window. This allows the user to see how two different variables in the data set may be related. Appendix B of this document contains screenshots of each of the individual components. The user can also see a molecular dynamics simulation that can be adjusted to demonstrate the growth in the computational power of supercomputers over the last two decades. Details on the molecular dynamics simulation can be found in Appendix A.

The screenshot below demonstrates just one of the many insights that can be gained by viewing two plots in the same screen. The number of cores plot (on the top) is being compared with the CPU frequency data set (on the bottom). This shows that the number of cores in the computers on the Top 500 list begins to grow rather quickly around 2004. At the same time we see the frequency of processors begin to stabilize. This coincides directly with the shift to multicore processor design.

Blake Haugen & Nick McNutt
COSC 557
Fall '13



The visualizations are completely implemented in Javascript using the D3 library. All of the data wrangling was done in advance using Python and the Pandas library. The original data set had to be reconciled because the fields in each of the individual lists had changed over the years. This also allowed us to reduce the size of the data set (originally ~5.9MB) to allow for faster loading and interaction.

In conclusion, we believe that our visualization allows the user to explore the Top 500 data set in an intuitive way. We hope that it will be useful for novice users and HPC experts alike. The visualization is available at http://www.bhaugen.com/vis/top500vis.html.

Blake Haugen & Nick McNutt
COSC 557
Fall '13

Appendix A

In order to better understand the evolution of supercomputing capability throughout the years, a representative scientific research task that requires immense computational resources was selected. Molecular dynamics simulations are used to understand atomic-scale phenomena. We chose to simulate a small (100 atom) Lennard-Jones liquid. This liquid represents a real material whose atoms interact with van der Waals forces but without chemical bonds. The simulation is physically accurate and produces the same kinetic and potential energy values as LAMMPS, a software program designed by Sandia National Laboratories that is designed to run on massively parallel machines.

The unique nature of our program is that all of the simulation code is written in Javascript. By its nature, Javascript was not designed to be a high-performance language; however, the code is only about 40% slower than an equivalent optimized C++ implementation. This indicates that Chrome's V8 engine and other similar browser engines work sufficiently well for allowing computationally-intensive tasks to be performed in the web browser. This kind of interactive, widely-accessible visualization option was not available until recently.

While the simulation calculations are computed in Javascript, the visualization code is a combination of Javascript and WegGL. WebGL allows the graphics-processing load to be shifted to the GPU, freeing the CPU for the task of simulation. The visualization code is also custom-written and allows an interaction with molecular systems similar to that of the desktop-based software VMD. To our knowledge, this code is the first that combines simulation and visualization into one HTML5 package that is usable by anyone with a relatively modern computer.

The simulation consists of a periodic box of 100 atoms, and each atom's color corresponds to its speed. As the average kinetic energy of a system is proportional to temperature, regions of bright colors indicate high temperature.

To the left of the simulation frame is a slider that allows a relative scaling of performance to match that of the world's best supercomputers throughout the past two decades. It is notable that moving the slider any earlier than about year 2006 brings the motion of the atoms to a standstill; this visualization represents the exponentially improving rate of computational performance in a way that numbers alone cannot.

To the right of the simulation frame are two radio buttons that allow the relative performance of the simulation to be scaled to the top supercomputer of a given year or to the 500th most powerful computer of that year. It is notable that there is a vast difference in aggregate floating-point operations per second (FLOPS) between the two computing systems.

Blake Haugen & Nick McNutt
COSC 557
Fall '13

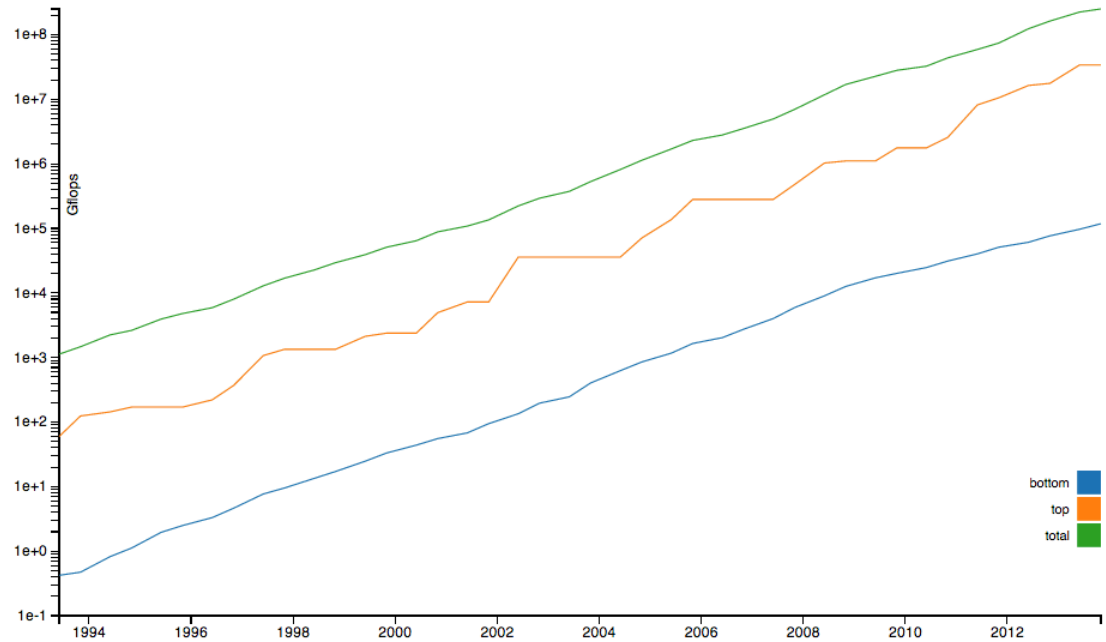
The simulation makes use of the libraries jQuery, jQuery UI, CoffeeScript, and ThreeJS. All of these libraries provide auxiliary assistance in coding, and none of them contain any of the fundamental simulation or visualization capabilities.

Blake Haugen & Nick McNutt
COSC 557
Fall '13

Appendix B

The following screen captures from each of the individual components that make up the visualization. The user has the ability to select two of the components to view side by side.

Performance Plot

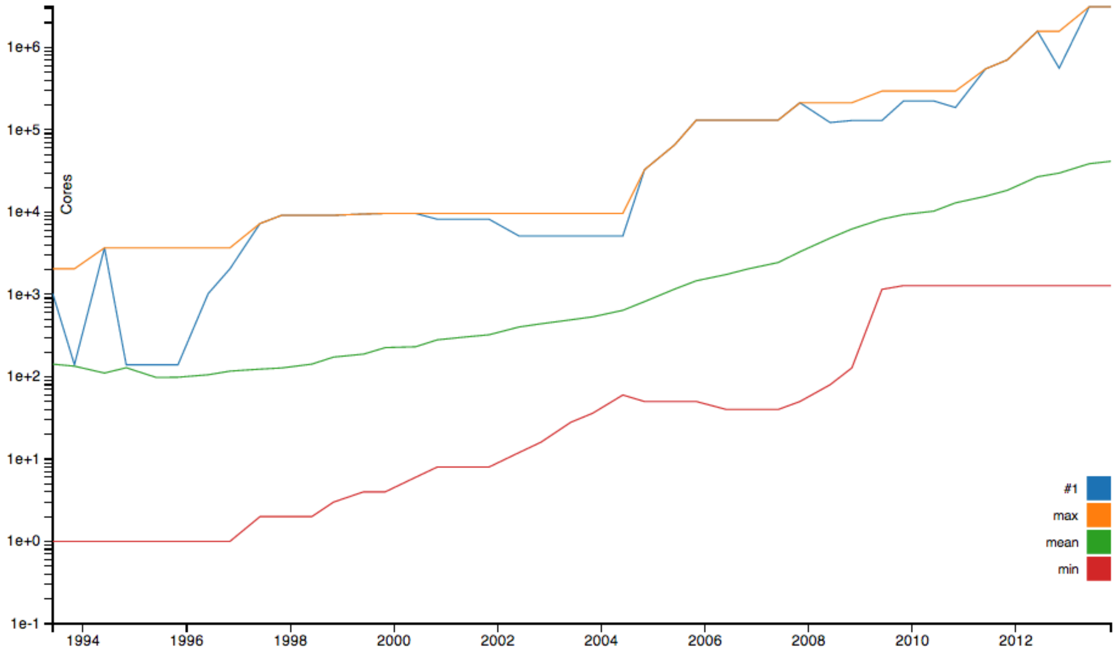


The performance plot shows that the speed of supercomputers has been increasing exponentially for the last 20 years. The plot shows the performance of the #1 system on the Top 500 list at each point in time. It also shows the performance of the #500 machine and the sum of the performance of all machines. The performance is measured using the Linpack benchmark. This plot leaves many people wondering how long the exponential growth in performance will continue.

Data Source: [Top 500]

Blake Haugen & Nick McNutt
COSC 557
Fall '13

Number of Cores

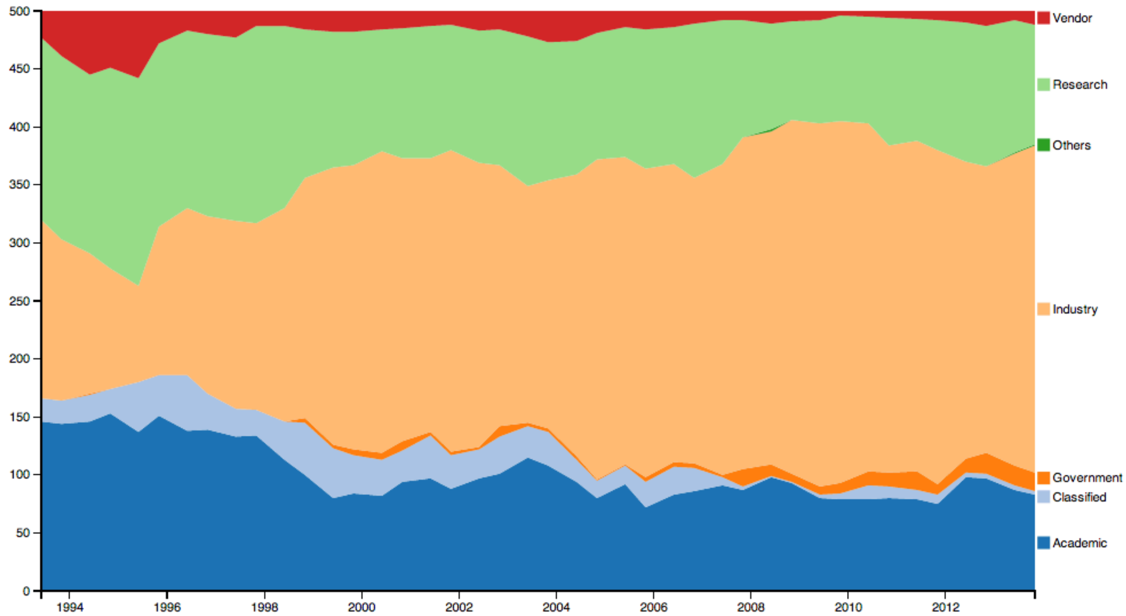


As the performance of supercomputers has grown over time, so have the number of cores present in each system. This growth in the number of cores can make parallel programming a real challenge. It is interesting to see the rather flat section of the graph from 1997 to 2004. Around 2004 the number of cores begins to dramatically increase with each release of the list. This corresponds directly with the point in time when manufacturers stabilized the clock frequency of the processor but began manufacturing multicore chips.

Data Source: [Top 500]

Blake Haugen & Nick McNutt
COSC 557
Fall '13

Segments of the Economy

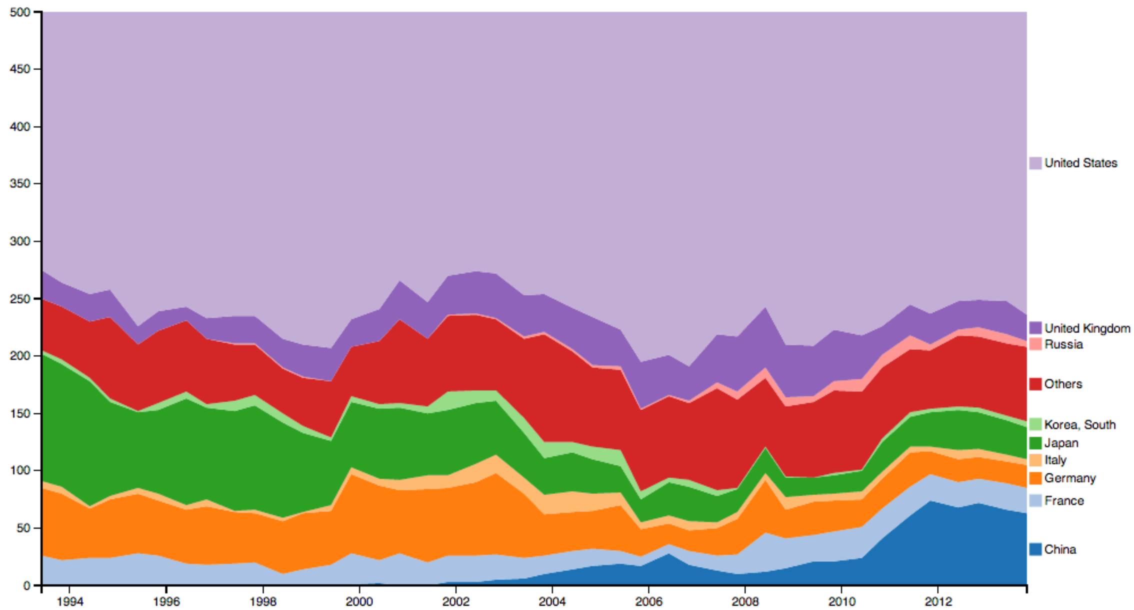


This plot shows the segments of the economy that are using supercomputers (and have submitted Linpack results). It is often assumed that universities and government-run research facilities have the vast majority of supercomputers. Over the last 20 years, however, industry machines have become a large portion of the list.

Data Source: [Top 500]

Blake Haugen & Nick McNutt
COSC 557
Fall '13

Countries

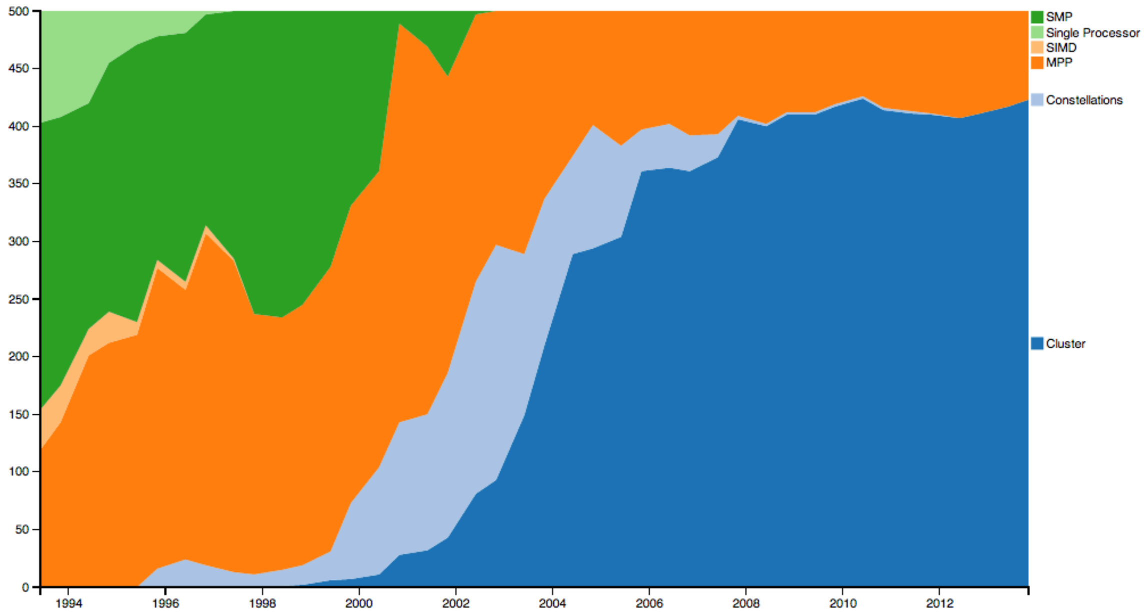


There are a number of countries that have supercomputers but the U.S. still has the largest number of machines in the Top 500 list. One of the most interesting parts of this plot is the growth in the number of Chinese machines on the list. It is even more interesting if you take some time to examine the lists and notice that many of the Chinese machines are quite highly ranked, including the #1 machine as of November 2013.

Data Source: [Top 500]()

Blake Haugen & Nick McNutt
COSC 557
Fall '13

Architecture

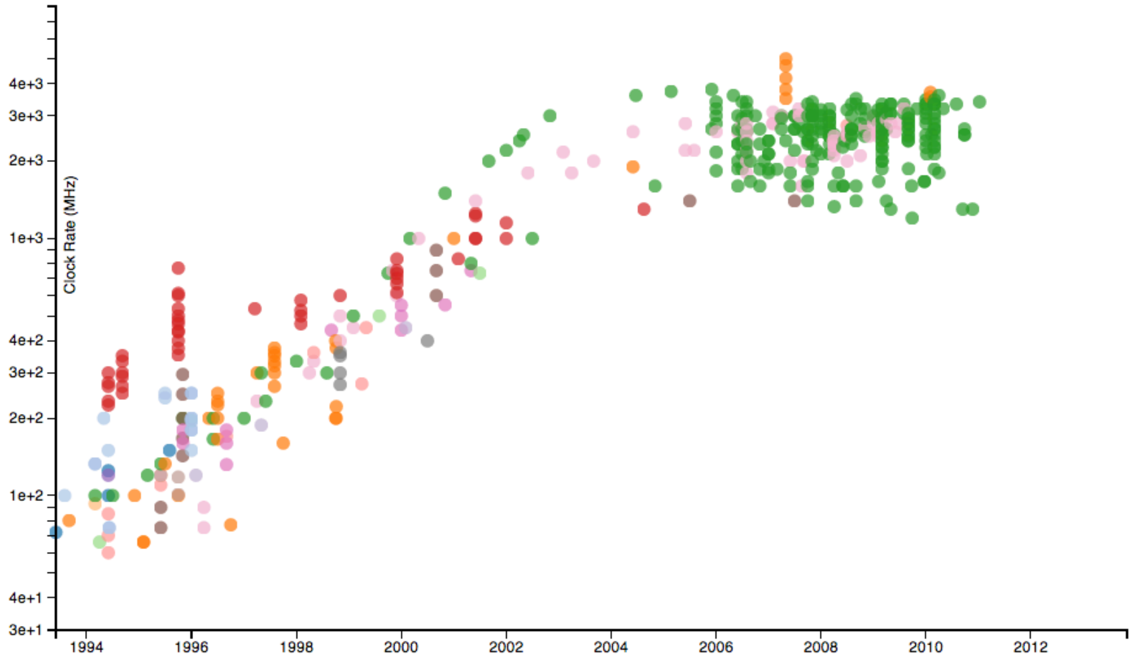


The way supercomputers are built has also changed over the last 20 years. In the early years the list was dominated by SIMD (Single Instruction Multiple Data), SMP (Symmetric Multiprocessor), and MPP (Massively Parallel Processor) supercomputers. The last 20 years of supercomputing have marked the rise and fall of constellation supercomputers as well as the dominance of cluster supercomputers today.

Data Source: Top 500

Blake Haugen & Nick McNutt
COSC 557
Fall '13

CPU Frequency

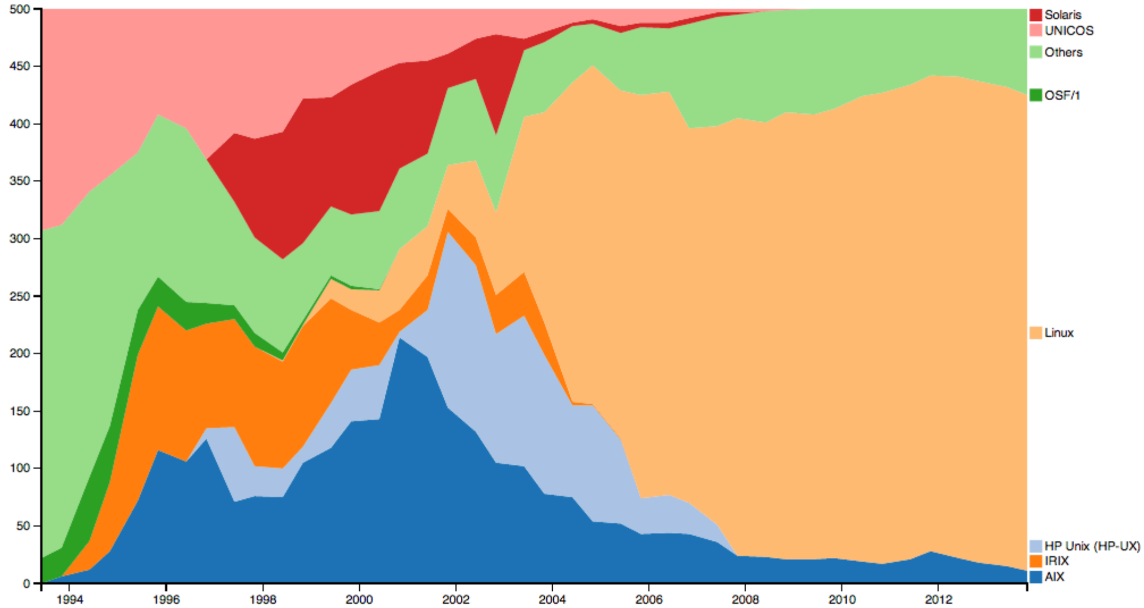


This is a plot of the clock frequency of various processors over time. The data for this plot was taken from CPUDB rather than the Top 500 data set. The clock frequency was growing nearly exponentially until approximately 2004. Around 2004 manufacturers began making multicore chips while the clock frequency of the processors stabilized. The color points are colored based on the manufacturer of the chip.

Data Source: [CPUDB (Stanford)]

Blake Haugen & Nick McNutt
COSC 557
Fall '13

Operating System

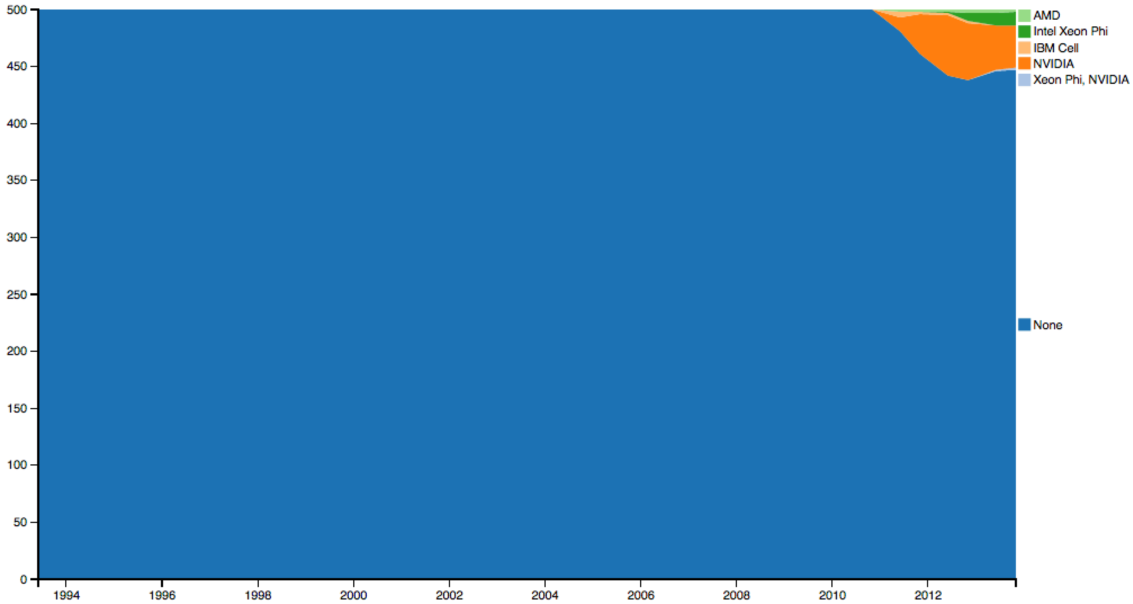


This plot shows how operating systems have developed over time in the HPC community. The "Others" category is a collection of other operating systems that have a relatively small market share. The real message in this plot is that Linux has taken over the supercomputing community.

Data Source: [Top 500]

Blake Haugen & Nick McNutt
COSC 557
Fall '13

Accelerators/Co-Processors

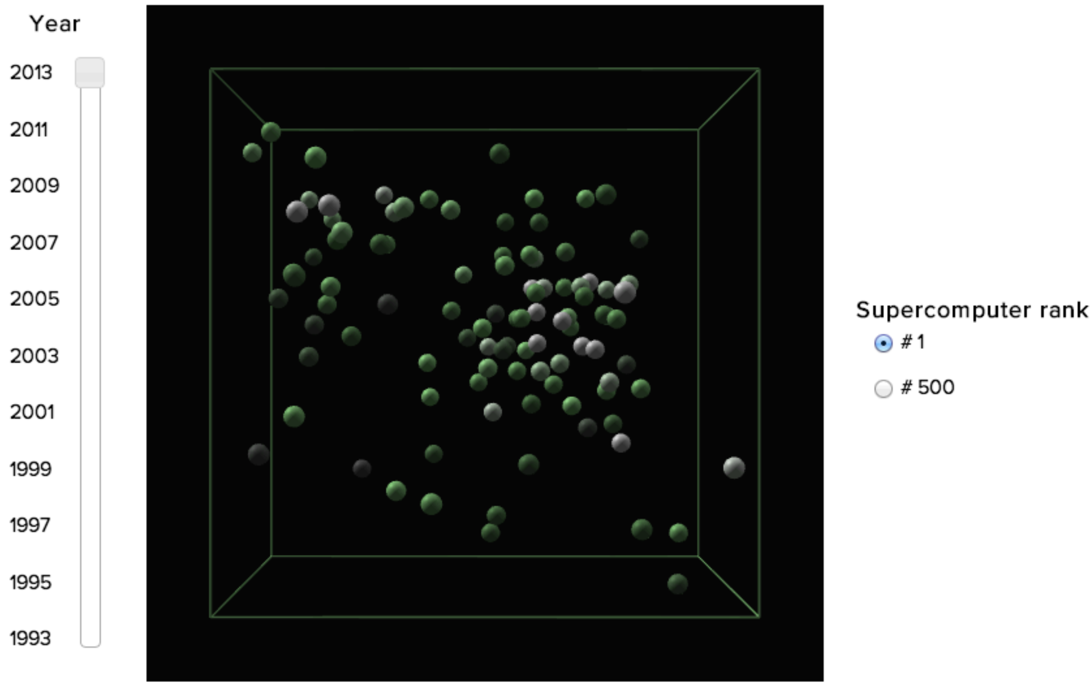


In recent years supercomputers have started to use accelerators or co-processors in order to improve performance. There is a lot of hype surrounding GPU computing and accelerators. For this reason many people believe that the Top 500 list has a large number of machines that employ accelerators. This graphs shows that this is simply not true. In fact, only a small number of systems use accelerators.

Data Source: Top 500

Blake Haugen & Nick McNutt
COSC 557
Fall '13

Molecular Dynamics Simulation



This is a physically accurate Lennard-Jones liquid simulation using the velocity-verlet timestep integration algorithm. It is a to-scale representation of a real simulation that is performed on supercomputing systems. This molecular dynamics simulation illustrates the evolution of supercomputing performance in floating point operations per second (FLOPS) over the last two decades. This clearly demonstrates the power of exponential growth.

The brightness of each atom corresponds to the atom's speed.