

# On the Interpretation of Top500 Data

Dror G. Feitelson

Institute of Computer Science

The Hebrew University of Jerusalem

91904 Jerusalem, Israel

<http://www.cs.huji.ac.il/~feit>

## Abstract

A novel representation of Top500 data is proposed, where the systems are plotted in rank and number-of-processors coordinates. This rendering brings to light various trends, such as complete Japanese domination of the range of high rank achieved with a small number of processors. In addition, an evolutionary trend whereby the rank of a given machine doubles every year is identified. This in turn leads to a characterization of the distribution of computing power across the whole list.

## 1 Introduction

Serious analysis of a field of study requires a compilation of the pertinent raw data. When studying the market for high-performance computing systems, the obvious data is about computer architectures and the products offered by different vendors. Indeed, several detailed surveys have been compiled, listing the available products [5], and identifying trends in the design and implementation of systems [3].

However, the real impact of architectures and products does not depend only on their architectural novelty and performance potential, but also on various economical and social factors that are all-important in determining how well they sell. The Top500 list reflects an understanding of this issue. It lists the top 500 supercomputer installations worldwide, as measured by the Linpack benchmark [2]. Such lists have been published semi-annually (in June and November) since 1993. This study is based on the year-end lists.

Naturally, only a relatively small number of systems are represented in the Top500 list, and this group changes with time. The dominant players at this time are SGI with its Origin 2000 and the Cray line of T3D and T3E MPPs, Sun with its HPC, IBM with its SP2, and the large Japanese supercomputer manufacturers Hitachi, NEC, and Fujitsu. Smaller players include Intel, Compaq, and HP. While many of these systems are architecturally similar (e.g. using top-of-the-line commodity microprocessors, distributed memory, and wormhole routing) and can scale to very large system sizes, their actual usage patterns differ. Our goal is to bring such patterns to light.

## 2 Rendering Top500 Data

The Top500 compilation includes various graphs that interpret the data, and present different classifications of systems and computing power (measured in peak performance). Such classifications include the distribution according to vendor, usage, and continent. However, all these views of the data treat the whole Top500 list as a unit, and do not differentiate between the high end and the low end.

We propose a novel representation of the same data, in which all 500 data points are plotted in rank and number-of-processors coordinates. Rank correlates to performance, so we can see differences between the most powerful systems and other less powerful systems. The number of processors required to achieve a certain rank correlates with architecture and technology. Color coding is used to differentiate between the systems according to some chosen criteria. Patterns seen in such plots enable us to make some new observations that cannot be made based on the traditional views used.

## 3 Usage Patterns

### 3.1 Architectures and Origin

Fig. 1 shows a plot in which systems are divided into three categories: systems manufactured by the three large Japanese companies, Cray Y-MP systems, and MPPs manufactured by American and European companies. It provides a fascinating example of how the uncoordinated activities of thousands of individuals across the globe, whether engineers, managers, or administrators, leads to combined patterns that are simple and easy to characterize.

A basic question in the design of high-performance systems is how to achieve the high performance: use a small number of very powerful vector processors, or a large number of commodity microprocessors, as is done in MPPs. Our plot shows that at the present time neither technology has a clear advantage: MPPs and vector machines compete across the whole list. However, vector machines have more presence in the top half, and parallel machines (most of which are not “massively” parallel) dominate the bottom half. Obviously, MPPs require more processors to achieve the same performance for each rank, so the plot is effectively divided into the MPP area on top and the vector area below it.

The figure also gives striking evidence of complete Japanese control of the lower-left corner — the area of highest performance with the fewest processors. Thus the Japanese companies take the technology pioneered by Cray to the limit, leading to the following simple pattern: MPPs on top, Japanese machines at lower-left, and Cray Y-MP machines forming the boundary between these two domains. This pattern persists in all the Top500 lists to date, but the spread is becoming smaller, and the Y-MP machines are on the verge of disappearing.

### 3.2 MPP Vendors

Fig. 2 makes the distinction between major MPP vendors (vector machines are not distinguished). This plot shows the following:

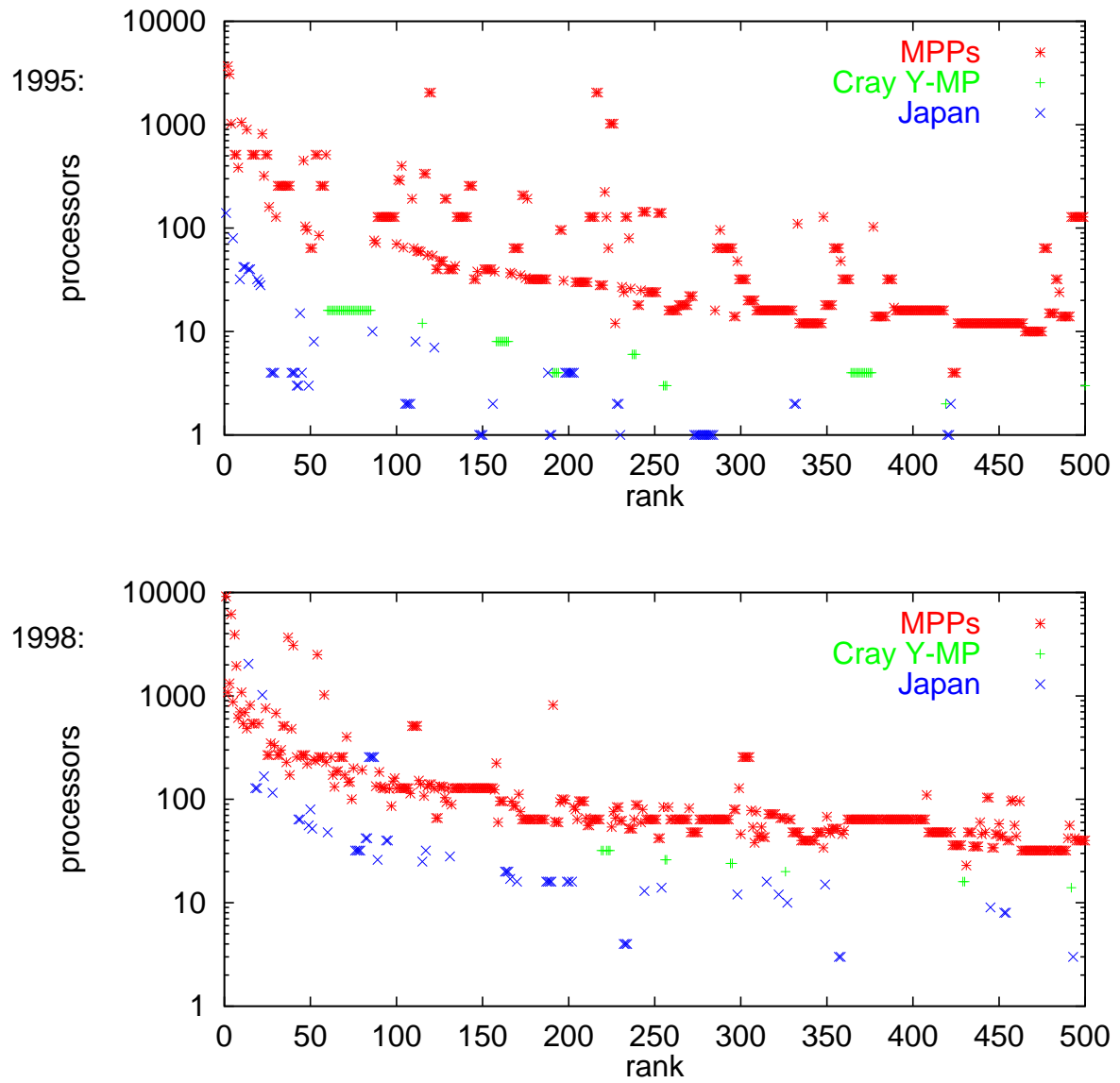


Figure 1: *The Cray Y-MPs divide the rank/processors space between Japanese machines and American MPPs.*

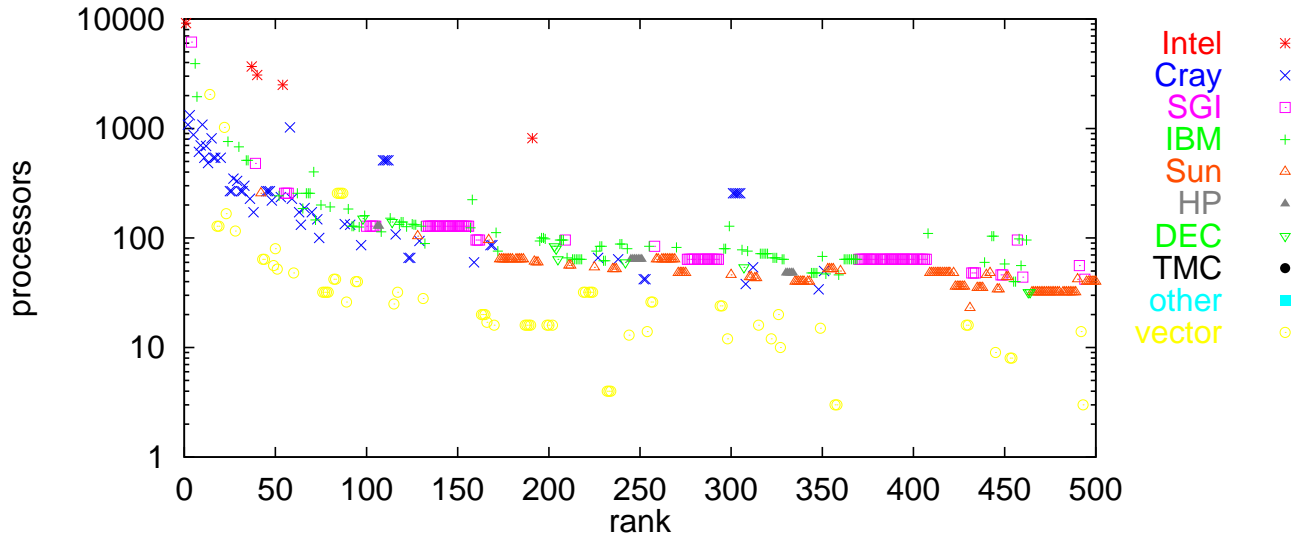


Figure 2: *Machines from major MPP vendors create distinct patterns.*

- The Intel Paragon uses weak processors — it consistently needs more processors to achieve the same performance as its competitors. Therefore the Paragon data points form a line across the top of the plot. The node performance of all other systems is remarkably similar to each other.
- The Cray T3D and T3E machines have significant presence in the top quarter of the list, and the SGI Origin 2000 have significant presence in the rest. Together, SGI/Cray cover the spectrum. However, it seems that the Origin has a scalability problem: they have a number of discrete models that sell well, but nothing in between. The actual performance differences between these models are significant (note that the rank coordinate correlates with performance, but this is certainly not a linear correspondence).
- The IBM SP2 scales well across the whole spectrum, mainly the top two thirds of the list. Models with any number of nodes can (and have) been installed, leading to a continuous line across the plot. Similarly, Sun HPC systems span the bottom two thirds of the list.

### 3.3 Usage

The Top500 data makes a distinction among five types of installations: research, academia, industry, classified/government, and vendor (for client benchmarking and similar activities). This distinction is used in Fig. 3. The plot shows that research is literally ahead of industry: Research and academic installations dominate the top of the list, while industry and government dominates the lower parts. The fractions of the list change with time: in 1994 industry and government use were scattered mainly in the bottom half of the list, in 1996 they dominated the bottom two thirds, and in 1998 they dominated all but the top 20%.

Another interesting observation is that the overall spread of the entries is being reduced with time: in 1994 entries were scattered all over the processors/rank space, but by 1998 they are

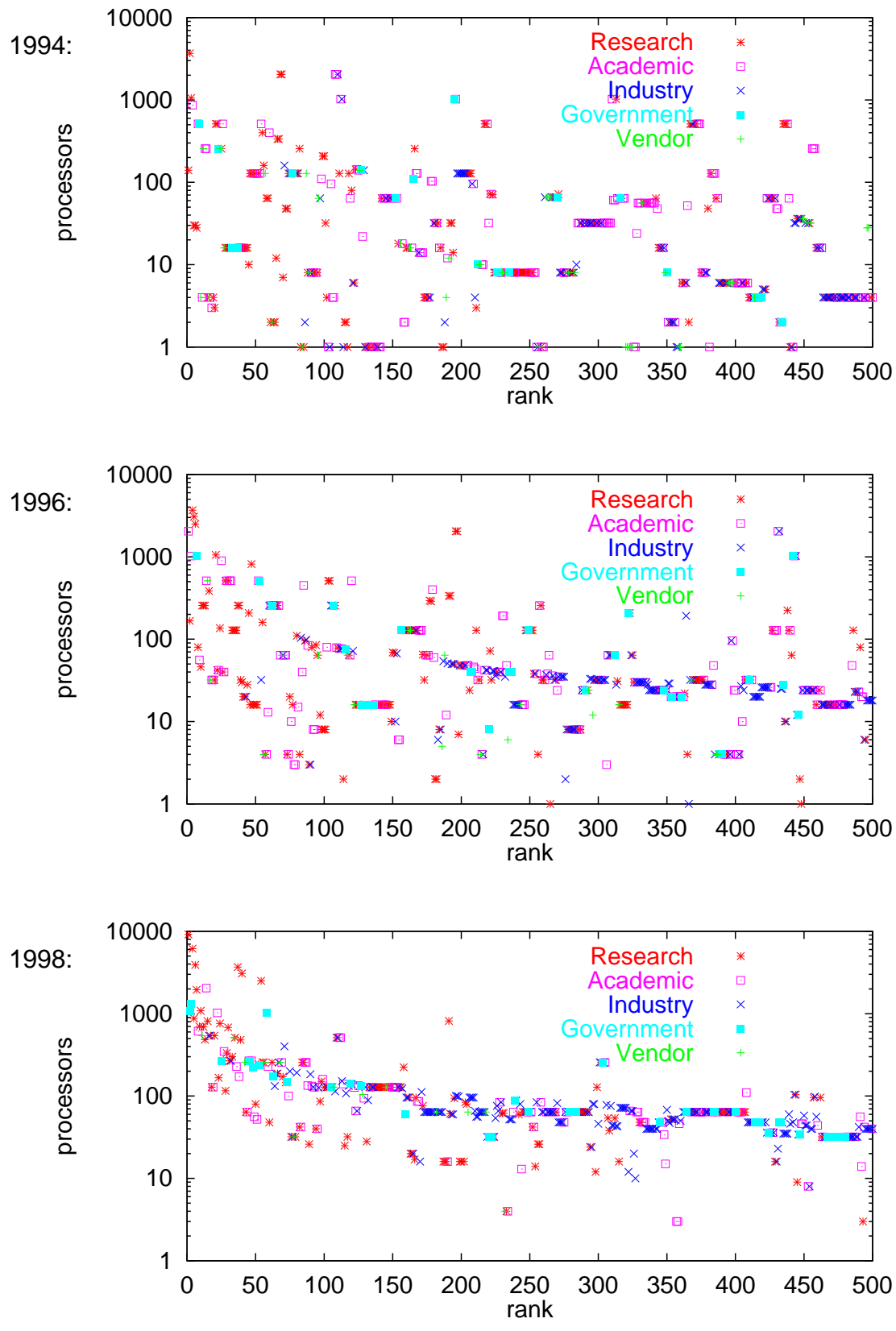


Figure 3: *Research is ahead of industry and more widely dispersed.*

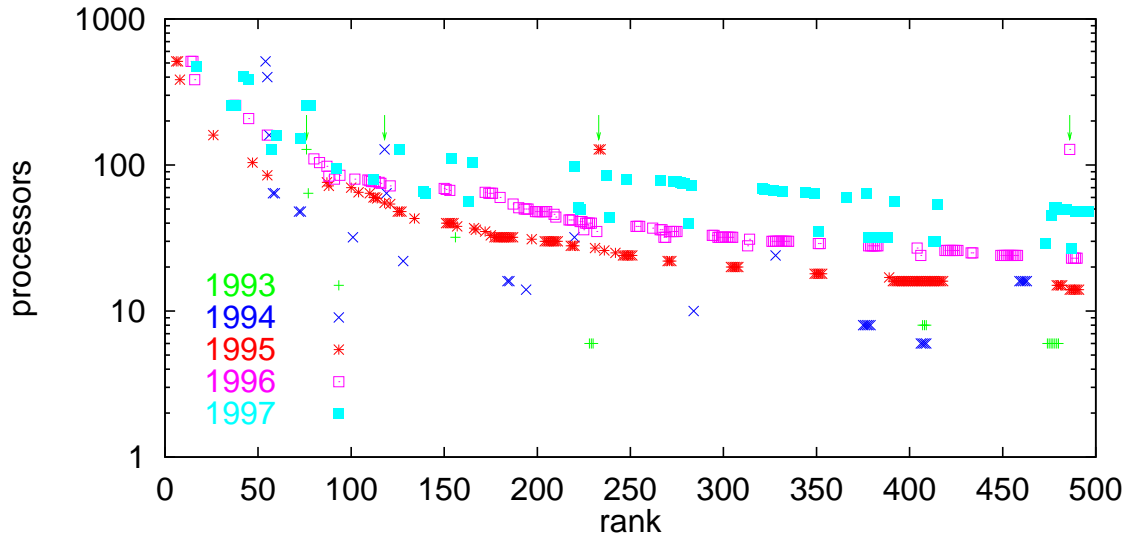


Figure 4: *The placing of IBM SP machines in different lists.*

concentrated in a relatively narrow swath. Within this general spread, research and academic sites are much more scattered, implying the use of many different models and configurations. Industry and government entries tends to clump in one continuous line built of clusters representing favored models and configurations, be they SGI Origin 2000s, IBM SP2s, or Sun HPCs.

## 4 Evolutionary Trends

All the plots so far have included entries from a single Top500 list. It is also instructional to compare related entries in several lists. For example, Fig. 4 shows the placing of IBM SP machines over the years. It shows a scattering of initial machines in 1993 and 1994, the consolidation into a solid line in 1995 and 1996, and the split into two lines when the strongest Power2 processors were introduced in 1997.

### 4.1 Rate of Change

An interesting feature that stands out in the IBM plot is the set of entries for the 128-node SP1 installed at Argonne National Lab (marked by small arrows). It seems that the distances between these points double each time, leading to the conclusion that the rank of a given machine doubles each year. This has been verified for a number of other representative machines, and holds except for a handful of entries at the very top of the list (Fig. 5). It is also possible to see the effects of the relative slump of 1993–1995, when the top of the list stayed stable, followed by the aggressive installment of top-notch machines in 1996–1998.

The fact that the rank doubles every year implies that new procurements are distributed in the same manner as existing installations. Thus each year the bottom half of the list gets bumped off, and new systems are inserted uniformly throughout the list, somewhat like a perfect shuffle of old

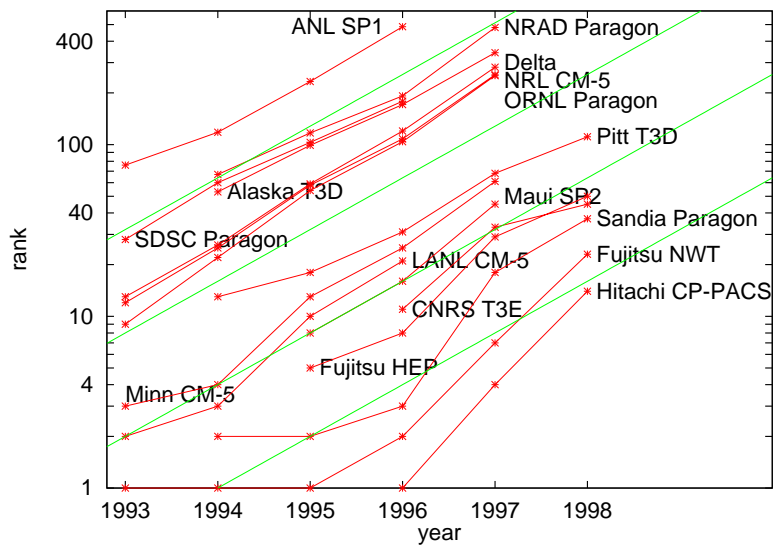


Figure 5: *The rank of a machine doubles every year.*

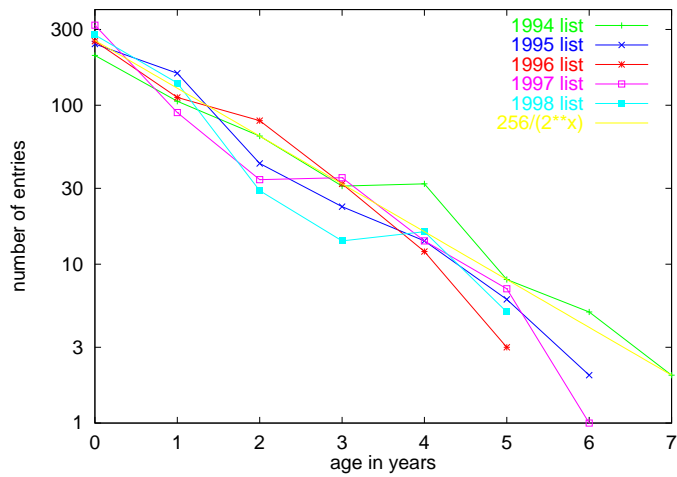


Figure 6: *The number of machines of a given age drops exponentially.*

<i>year</i>	<i>rank</i>	<i>Machine</i>	processors
1995	491	IBM SP2	14
1996	500	SGI Challenge	18
1997	500	Sun Ultra	22
1998	489	Sun HPC	32

Table 1: *The lowest degree of parallelism in microprocessor-based machines doubles every three years.*

and new. Simple counting indeed shows that at any year approximately half of the list is composed of new machines, and the number of older surviving machines drops off exponentially (Fig. 6). This high rate of exchange testifies to the continuous and aggressive progress in the use of high performance parallel machines.

## 4.2 Rate of Progress

It is tempting to measure the rate of progress by the power of the strongest machine each year. However, the rank 500 machine may be more representative of the state of the high-performance computing arena as a whole. To a good approximation, this number doubles every year.

Given the common assumption of doubling microprocessor performance every 18 months [4], doubling system performance every 12 months implies that more and more processors need to be used. The difference between doubling every year and doubling every 18 months leads to the prediction that the number of processors used should double every three years. This agrees with data from the last four years (Table 1), disregarding the 1993 and 1994 lists, when vector machines dominated microprocessor-based machines. In addition, it should be noted that up to 1995 the number of uniprocessor vector machines that made the list has dropped steadily, and since 1996 there are no more uniprocessors in the lists.

Remarkably, such doubling of performance every year is very close to the stated ASCI goal of tripling performance every 18 months [1]. (to be precise, tripling every 18 months is equivalent to multiplication by a factor of 2.08 every year, and the minimal rmax values are closely modeled by multiplication by a factor of 2.05 every year.) Thus it might seem that the ASCI program is actually not very aggressive. However, as we'll see below, the top end of the list (which is the part targeted by ASCI) is not keeping up with this rate of progress.

## 5 The Distribution of Power

The previous section established two observations:

1. The rank of a given machine doubles every year.
2. The rmax value of the rank 500 machine doubles every year.

The combination of these two observations allows us to define the distribution of rmax values across the whole list.



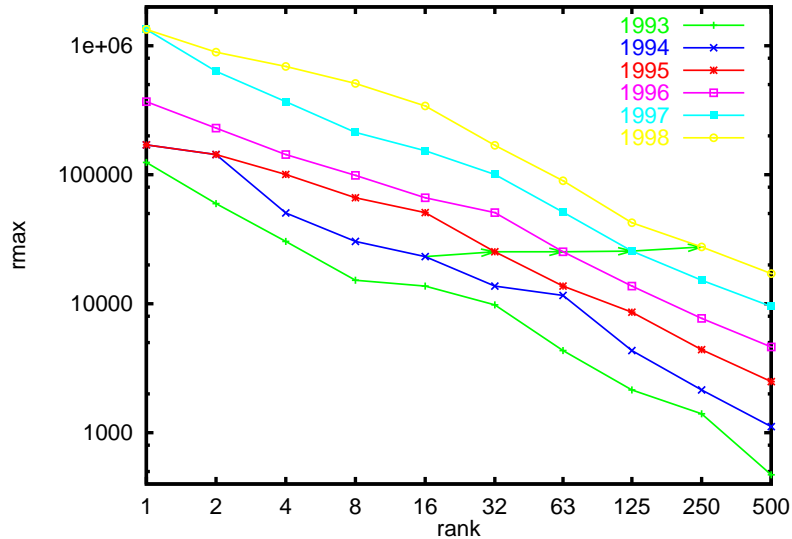


Figure 7: *Within each list, the power of machines drops proportionally to the rank.*

Consider the rank 250 machine in a given list. One can expect it to be close to the rank 500 machine the next year. The rank 500 machine next year is expected to have double the power of the rank 500 machine this year. Thus the rank 250 machine should have approximately double the power of the rank 500 machine. Similar arguments indicate that the rank 125 machine should have double the power of the rank 250 machine, and four times the power of the rank 500 machine, and so on. In short, the power should drop in proportion to the rank, leading to a Zipf distribution.

Plotting the rmax values against rank on a log-log scale indeed leads to rather straight lines, and in most cases equal performance is seen at doubling ranks across the years (arrows). But a

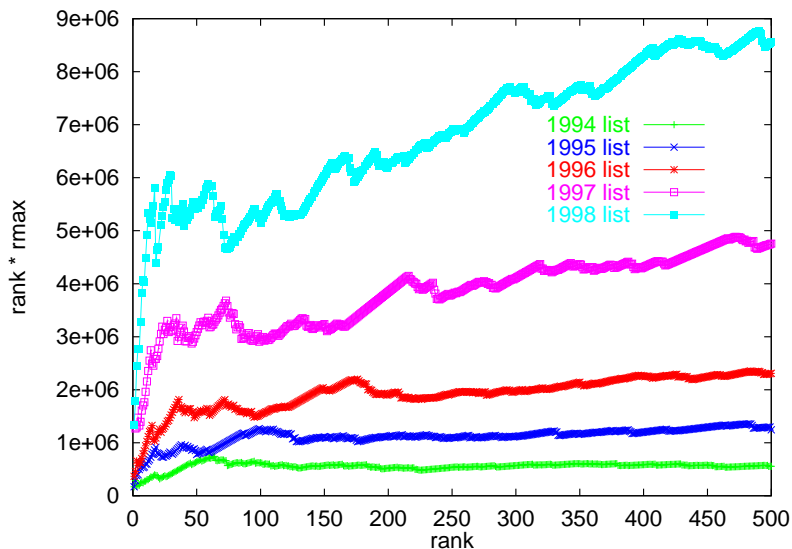


Figure 8: *The quality of the Zipf-distribution model is deteriorating.*

better verification is to plot the product of the rank and rmax values as a function of the rank (Fig. 8). Compliance with a Zipf distribution should then lead to a straight horizontal line. The graphs show that in 1994 and 1995 the Zipf distribution was an excellent model, but since then, it has deteriorated. Specifically, the low end of the list has more power than predicted by this model. In addition, the very top of the lists never fits the model: these data points are always too low. It remains to be seen whether the ASCI program will boost the performance of the top machines to the levels anticipated according to their lesser relatives.

## References

- [1] D. Clark, “*Breaking the teraflops barrier*”. *Computer* **30(2)**, pp. 12–14, Feb 1997.
- [2] J. J. Dongarra, H. W. Meuer, and E. Strohmaier, “*Top500 supercomputer sites*”. <http://www.netlib.org/benchmark/top500.html>. (updated every 6 months).
- [3] G. Lerman and L. Rudolph, *Parallel Evolution of Parallel Processors*. Plenum Press, 1993.
- [4] R. R. Schaller, “*Moore’s Law: past, present, and future*”. *IEEE Spectrum* **34(6)**, pp. 52–59, Jun 1997.
- [5] A. Trew and G. Wilson (eds.), *Past, Present, Parallel: A Survey of Available Parallel Computing Systems*. Springer-Verlag, 1991.