Abstract:
Analyzing the history of supercomputers: how the industry arrived to where it is in its current form and what technological advancements enabled the movement. Explore the applications of Digital Signal Processor in high performance computing and whether or not tests have already taken place. Examine whether Digital Signal Processors have a future with supercomputers.
# Table of Contents

Table of Contents ..................................................................................................................................................... ii  
List of Figures ............................................................................................................................................................ iii  
Introduction .................................................................................................................................................................. 1  
  What is a supercomputer? ........................................................................................................................................ 1  
  What are supercomputers used for? .......................................................................................................................... 1  
History of Supercomputing ......................................................................................................................................... 2  
  The Past (1964 – 1985)........................................................................................................................................... 2  
  The Present (2012 – Now) ...................................................................................................................................... 3  
DSP’s in Supercomputing ............................................................................................................................................. 5  
  How do they fit it? ..................................................................................................................................................... 5  
  MUSIC: Multi Signal Processor System with Intelligent Communication ............................................................. 6  
Conclusion .................................................................................................................................................................... 7  
References ..................................................................................................................................................................... 8
List of Figures

Figure 1: CDC 6600 ......................................................... 2
Figure 2: Cray-2 ............................................................ 3
Figure 3: Cray Titan ...................................................... 4
Figure 4: Advantech’s Quad DSP half-size PCI-Express co-processor ........................................ 5
Figure 5: Comparative price performance of MUSIC project to common supercomputers ............ 6
Figure 6: Architecture of the MUSIC DSP supercomputer ....................................................... 6
Figure 7: Performance benchmark of MUSIC with competition .................................................. 7
Introduction

What is a supercomputer?

A supercomputer is defined as a machine that has high processing speed and capacity. In other words, these computers are able to perform very complex calculations in a short amount of time and perform these calculations on extremely large data sets. The power of these machines are measured using a benchmark known as ‘LINPACK’, which measures the ‘Floating Operations per Second’ (FLOPS) when solving a dense nxn matrix.

Supercomputers come in two main forms, cluster and distributed. Computer clusters consist of a large amount of processing units within a short distance (usually in the same room). These processing units can be connected using a variety of network topologies, each with specific use cases, to accomplish a task together. Distributed computing is when there are a large amount of processing units spread over long distances. These processing units are often given packets of data to process which are then sent back to a central computer when completed. These processing units can be spread out all over the global and be a small as a regular desktop computer. [1]

Generally, supercomputers are built to perform a specified task and are not meant to run many different programs like consumer computers. Supercomputers employ very specialized operating systems to accomplish these tasks and be able incorporate these complex network topologies.

What are supercomputers used for?

There are many fields and industries that require a large amount of processing power to complete their jobs. This includes but is not limited to scientific research, weather modelling and cryptography.

Scientific research greatly benefits from supercomputers to be able to run simulations for matters including [2]:

- The Evolution of the Cosmos
- Biological system behavior
- Molecular Dynamics

The advent of supercomputers allows researchers to continuously push the envelope of what’s possible to be studied. Each year, supercomputers become faster; allowing more complex simulations to be run.

The roots of supercomputers go back to when electromechanical machines were used to break the enigma code during World War 2. Today, supercomputers are still used in the field of cryptography to break codes, specifically at the National Security Agency (NSA).
History of Supercomputing

The Past (1964 – 1985)

The first supercomputer was born out of the Control Data Corporation (CDC), designed by Seymour Cray. This first machine was known as the CDC 6600 as was released in 1964. As mentioned previously, the roots of supercomputers date back to the Second World War and even further past that, though the CDC 6600 is considered to be the first supercomputer, as the term is known today.

The CDC 6600 was custom built for the European Organization for Nuclear Research (CERN) to analyze bubble-chamber tracks for the experiments that were being performed. [3] The machine was later produced and sent to researcher centers around the world.

Figure 1: CDC 6600

The CDC 6600 performed at approximately 1 MFLOPS, which at the time was the fastest computer in the world until it was succeeded by the CDC 7600 which ran at 10 MFLOPS. Released in 1969 the successor cost $5 million to buy and remained the fastest computer until around 1975. The CDC 7600 utilized a reduced instruction set with 64 15-bit machine codes.

In the late 1960’s, CDC began work on the 8600 which would end up being essentially 4 faster 7600’s utilizing a form of vector processing. The four processor would access shared memory and be able to process the same instruction with different data on each processor. This is an early form of what is now known as ‘Single Instruction Multiple Data’ (SIMD). The computer operated ten times faster than its predecessor (100FLOPS), however, due to financial problem the project was scratched. Seymour Cray left the company in 1972 to form Cray Research.

In 1977, the Cray Research released their first supercomputer: Cray-1. Performing at 80 MFLOPS and introducing the first usage of integrated circuits for Seymour Cray, along with instruction parallelism and other enhancements; the Cray-1 was a spectacular entry into the super computer market for Cray Research.
The successor to the Cray-1, released in 1982, had a similar architecture but increased the memory bandwidth and introduced shared-memory parallel vector processing. This SIMD-style processing allowed the Cray X-MP to achieve approximately 400MFLOPS of power. Later on in 1985, the Cray-2 was introduced with a new architecture, more vector processors and a max performance of 1.9 GFLOPS.

![Figure 2: Cray-2](image)

What can be seen in the early days of supercomputing is that there was a lot of experimentation going on. These designers were always trying to find ways to increase performance. The advent of vector processing was a huge boon for the designers and foreshadowed what was to come in the future. What is interesting is that the history of Seymour Cray's computers followed a revolutionary-evolutionary cycle. Introduce a new model with a radical new design, then in the next model just improve on it. This can be paralleled to how Intel currently releases their microprocessors, on a ‘tick-tock’ cycle.

The Present (2012 – Now)

Fast-forwarding to the present, supercomputers have become a far cry from their historical predecessors both in terms of power and scale. Whereas before these computers were relatively small discrete machines, today's supercomputers have dedicated buildings and warehouses with their own industrial cooling system.

The processing units in many current supercomputers are known as heterogeneous systems; utilizing both traditional Central Processing Units (CPUs) as well as Graphics Processing Units (GPUs). The Cray Titan was built in 2012 for the Oak Ridge National Laboratory consisting of 18,688 16-core AMD CPUs and nVidia Tesla GPUs, producing 17.59 PFLOPS of performance at the cost of 8.2 MW of power. The Titan is made of fairly off-the-shelf parts where the CPU’s are consumer available, and the Tesla GPUs are very similar to the aptly named nVidia Titan GPU.
What makes the GPUs special is that they are made of many Reduced Instruction Set (RISC) processors which allows for highly parallelized tasks to be run. Having both Complex Instruction Set (CISC) and RISC processor gives the benefit of both worlds, which is why heterogeneous systems are very common. China has built the current fastest supercomputer, known as: ‘TIANHE-2’ [4]. This is a heterogeneous system like the Cray Titan, however, it uses a completely Intel based solution. Utilizing off-the-shelf Intel Xeon server processors along with the ‘Xeon Phi’ coprocessor (GPU equivalent), the TAINHE-2 has a peak performance of 33.86 PFLOPS using 17.6 MW of power.

What started with the introduction of vector processing between two processing cores in the early ‘80s has grown to thousands upon thousands of vector processing units now in 2013. It will continue to grow on into the future with more General Purpose GPUs (GPGPU) being made specifically to be general enough for many different computational scenarios. The Xeon Phi coprocessors in the Chinese supercomputer is evolving to a consumer product from Intel to be known as ‘Knight’s Landing’ to directly compete with nVidia’s Tesla [5].

Supercomputers come with their fair share of problems. Currently we want to build the largest computer we possibly can and one of the obvious limiting factors is money. These computers are expensive and will continue to be expensive, though given an unlimited budget one should be able to build and infinitely large computer. One of the other limiting factors is heat, where these processors running at full speed will generate enough heat to require dedicated cooling systems. Along with that comes the power requirements because it is important to note how much the processors take to power as well powering the cooling system. The upfront cost of these supercomputers is only the start; their real expensive comes through there usage. This is where power efficiency comes into play, where engineers attempt to get the most value per FLOPS. If we look at the Cray Titan and TAINHE-2, they achieve 2.14 and 1.92 GFLOP/Watt, respectively. The most efficient supercomputer is the TSUBAME-KFC which achieves 4.5 GFLOPS/Watt utilizing an oil-cooled heterogeneous architecture [6].
DSP’s in Supercomputing

How do they fit it?

Digital Signal Processors (DSP) by their very nature are efficient. Commonly built for embedded low power devices and being highly parallelized they use very little power while retaining fast floating point operations. Texas Instruments added floating point hardware to their chips in 2010 and were able to achieve performance of 12.8 GFLOPS/Watt. This is much better than the supercomputer efficiency, though that is a single 8-core chip. [7] The obvious question is whether or not we can connect many of these DSPs together into a supercomputer architecture to utilize their power while keeping the efficiency.

A company by the name of Advantech has created a board, shown in figure 4, with 4 of the aforementioned TI chips pushing out 192 GFLOPS of double precision performance. [8] They believe they will be able to double the performance on a larger board while only consuming 110 watts. The biggest obstacle for a company like TI entering the high performance computing industry is not to do with hardware though. nVidia already has an easy to use and scalable platform for utilizing their hardware in such manner, with Compute Unified Device Architecture (CUDA). TI would need to create a similar software platform to allow developers easy access to the hardware.

This is not to say that such an endeavor in using DSPs for in a supercomputer platform hasn’t already been attempted. A high performance computer company by the name of Mercury Systems created such a design in the mid ‘90s. In 1994 they unveiled a heterogeneous architecture consisting of PowerPC RISC processors with Analog Device’s SHARC DSPs. In 1996 they were able to achieve record performance at the time with their SHARC-based system, having 5GFLOPS of performance on one module. [9] As well, out of the Supercomputing ’92 conference, Swiss researchers unveiled their M.U.S.I.C. project.
MUSIC: Multi Signal Processor System with Intelligent Communication

At the supercomputing ‘92 conference in Minneapolis, researchers from the Swiss Federal Institute of Technology created a supercomputer comprised entirely of Motorola 96002 DSP processors. This computer was comparable and even faster than some at the time supercomputers on the market. The computer they built was faster, lower cost and used less power than their traditional counterparts. [10]

![Comparative price performance of MUSIC project to common supercomputers](image)

*Figure 5: Comparative price performance of MUSIC project to common supercomputers*

The architecture of the MUSIC project was designed to fit in one 19” server rack, consisting of 20 boards. Each module contained 3 processing elements (PE), in this case being the Motorola 96002 for the DSP. Communication happens over the video interface in a 5MHz pipelined ring buffer. In figure 6, all of these elements can be seen together. Each board has its own manager connected to the host interface of the DSP.

![Architecture of the MUSIC DSP supercomputer](image)

*Figure 6: Architecture of the MUSIC DSP supercomputer*
In software, the communication for the computer is handled by only 3 functions. First is ‘Init_comm’, which initializes communication and transmits parameters such as: array dimensions, pointers, etc. Then ‘Send_data’, which informs the PE that there is new data read to be sent. ‘Synchronize’ is the last function whose purpose is to wait for all data to be ready for computation. The computation is done via Single Process Multiple Data (SPMD), so each processing element executes the same code but on unique data.

The MUSIC project, under execution of an algorithm known as MD-Atom, was faster than the Cray Y-MP and NEC SX-3 (shown in figure 7) which were some of the fastest supercomputers at the time. Achieving performance such as this in one server rack was quite a feat. Though the real interesting aspect is the power efficiency where the computer consumes only around 800 watts while producing 3.8GFLOPS of performance, which was very low power for the time.

<table>
<thead>
<tr>
<th></th>
<th>Model 1 (125/1000)</th>
<th>Model 2 (1000/100)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sun IPX</td>
<td>118 seconds</td>
<td>643 seconds</td>
</tr>
<tr>
<td>CRAY Y-MP</td>
<td>3.7 seconds</td>
<td>12.2 seconds</td>
</tr>
<tr>
<td>NEC SX-3</td>
<td>1.4 seconds</td>
<td>4.4 seconds</td>
</tr>
<tr>
<td>MUSIC-10 (30 PE)</td>
<td>1.3 seconds</td>
<td>3.8 seconds</td>
</tr>
<tr>
<td>MUSIC-20 (60 PE)</td>
<td>0.91 seconds</td>
<td>2.02 seconds</td>
</tr>
</tbody>
</table>

Figure 7: Performance benchmark of MUSIC with competition

Conclusion

Through the history of supercomputing, we see that these computers evolved into highly paralleled vector processing units. Adding more process cores, introducing SIMD processing and increasing the speed of each core; It makes sense that DSP could be a part of this. Since such a huge expense of modern day supercomputers is just in powering them and trying to keep them cool, efficiency should be one of the biggest goals. For every penny you save on running costs means more can be put in the initial costs in making a bigger computer. Given that DSPs are already highly paralleled, but also extremely power efficient they should be able to compete in the high performance computing industry. Texas Instruments has begun testing these sorts of high performance applications.

Two years after Texas Instruments initially start looking into high performance computing, we have not seen very much movement forward. Whether or not this is because of limitations of software where there are already established platforms such as CUDA. It may be hard for DSPs to enter the current race among the heterogenous architectures of CPUs and GPGPUs, especially when some computers are already achieving 4+ GFLOPS/Watt performance. Despite already shown to be a viable alternative, it may be too late in the game for DSP manufacturers.
References


