The Green500

And the Importance of Efficient High-Performance Computing

Julian Hölz

Sommerakademie in Leysin, August 2016

Abstract

The Article tries to discuss the value of the Green500 list as an Indicator of Green HPC and Supercomputing. Hence, it displays the history, the motivation leading to the existence of the Green500, and discusses its usefulness. It also tries to define the term Supercomputing and HPC Computing to a level that an uneducated reader would easily understand. Some of the first cutting edge computer systems in GreenIT build during the time from 2002 till 2009 are introduced to the reader and are discussed regarding their energy efficiency and their usefulness in real application. At last the the concepts of power-aware and low-power computing are explained to the reader.

Contents

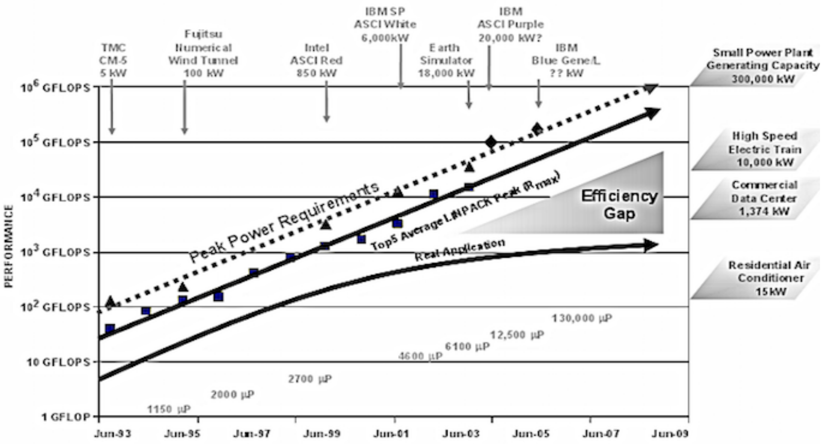

Figure 1: Source: NSF Career Proposal Submission, K. W. Cameron, July 2003

1 Introduction

In an era of environmental awareness and 'Green Thinking' an article similar to this, discussing the topic of energy efficiency in Supercomputers, is very likely to gain some interest. Although only a few years ago such an article would only be read by environmental activists, but would definitely be frown upon by serious HPC developers and architects. Hence, the concept of saving energy in very powerful computer systems is only as old as 10 to 15 years. This doesn't make this principle less important, thought, as this article will illustrate.

Looking back to 1993 we'll find that although computers were not nearly as powerful as comparable competitors on the Top500 list 14 years later (as seen on Fig. 1). More astonishingly is the fact that in 1993 the TMC CM-5, a early supercomputer by the former Thinking Machines Cooperation, has computed at around 12 MFLOPS/watt of energy consumed. The 2003 Japanese Earth Simulator which was much more sophisticated with its approximate 35,86 TFLOPS of average LINPACK performance, however only gets to 5.6 MFLOPS/watt despite technological advances. This is clearly a step into the wrong direction. The trend in supercomputing derived from Figure 1 is pretty clear: Architectures are scaling towards high performance systems while energy efficiency figures are slowly dropping. This will eventually lead to some problems regarding costs, liability and scalability. Evidence of this is that Figure 1 is one of the first graphs of its kind showing the discussed trend and that the author of the graph, Kirk W. Cameron has spent almost 4 months obtaining the data for the graph.

2 The Green500 List

2.1 Beginning

The idea of the Green500 was the brainchild of Kirk W. Cameron and Wu Feng, both working in Virginia at different technical institutes, in 2006. The objective was to create a list of the most power efficient supercomputers akin to the preexisting Top500 list, which solely listed the computers by their performance on the LINPACK benchmarking software (Metrics: FLOPS - Floating Point Operations per Second).

The first Challenge was to design the power measurement run rules by which the systems are

Green500 Rank	MFLOPS/W	Site*	Computer*	Total Power (kW)
1	**6,673.84**	Advanced Center for Computing and Communication, RIKEN	Shoubu - ZettaScaler-1.6, Xeon E5-2618Lv3 8C 2.3GHz, Infiniband FDR, PEZY-SCnp	149.99
2	**6,195.22**	Computational Astrophysics Laboratory, RIKEN	Satsuki - ZettaScaler-1.6, Xeon E5-2618Lv3 8C 2.3GHz, Infiniband FDR, PEZY-SCnp	46.89
3	**6,051.30**	National Supercomputing Center in Wuxi	Sunway TaihuLight - Sunway MPP, Sunway SW26010 260C 1.45GHz, Sunway	15,371.00

Figure 2: The first three places of the Green500 list of June 2016. Source:Green500.org, 2016.

measured for the first list. To make this as easy as possible Cameron suggests that it would be enough to measure only a single node and from that extrapolate the power consumed by the whole system while running LINPACK. The rules were especially designed to encourage the easy measurement of the systems and the submission to the list. This created as much relatively accurate data as possible. The energy consumption of systems that weren't submitted were then estimated to complete the list.

The first list was then completed and released in November 2007. Ironically, just prior to the launch, the power problem in huge data centers has gained the interest of the public and has become subject to open discussion. By then, many agreed that HPC systems definitely need to become more efficient. [1]

2.2 Requirements and Power Measurement

The idea of the the Green500 list itself is not very complex. It uses the preexisting list of the Top500 Supercomputers and rearranges the list's order placing the most efficient machine at the top. It harnesses the before mentioned metrics FLOPS/watt to do so.

Right after the release of the list some sceptics also had one or two things to say about the list:

1. The list was criticised for only ranking the best 500 performing supercomputers. Critics said that there are even more efficient computers that should definitely be listed.

2. The power measurement rule established at the beginning at the list might encourage more system owners to measure their systems, though, by measuring on only one single node the measurements could be very inaccurate or even worse: subject to gaming.

To fight the critics the power measurement rules have been altered by now. Since the list has gained much popularity after it was released, system figures were much more easily collected than in the beginning, thus there was no longer a need to encourage the system administrators to submit energy efficiency figures by making measurements as easy as possible. Therefore, new rules and standards have been introduced.

The measurements are now being ranked into 3 different levels of quality which give a great overview of how much effort has been put into measuring and how much data is being submitted. Especially important is what the levels say about how the energy has been measured.

1. Level: The power consumption only needs to be measured on a single node of the system

2. Level: The power consumption needs to be measured on a good fraction of nodes

3. Level: The power consumption of the complete system needs to be measured.

Green500 Rank	MFLOPS/W	Site*	Computer*	Total Power (kW)
1	5,271.81	GSI Helmholtz Center	ASUS ESC4000 FDR/G2S, Intel Xeon E5-2690v2 10C 3GHz, Infiniband FDR, AMD FirePro S9150	57.15
2	4,945.63	High Energy Accelerator Research Organization /KEK	ExaScaler 32U256SC Cluster, Intel Xeon E5-2660v2 10C 2.2GHz, Infiniband FDR, PEZY-SC	37.83
3	4,447.58	GSIC Center, Tokyo Institute of Technology	LX 1U-4GPU/104Re-1G Cluster, Intel Xeon E5-2620v2 6C 2.100GHz, Infiniband FDR, NVIDIA K20x	35.39

Figure 3: The first three places of the little Green500 list of November 2014. Source:Green500.org, 2016.

For measurements done for the Green500 list the measurements must be at least of quality level two or three.

It is also pretty clear that the power can not be measured at any given time. The workload on each system must approximately be about the same for each measurement. For the Green500 the measurements must therefore be done while running the LINPACK benchmarking software. The standard for the Green500, though, goes even further by stating that the power must be measured in constant intervals during the test, since the problem with LINPACK is that the test is done in multiple states. First the Software uses a random number generator to generate a high density matrix, then it performs linear operations on the matrix to test the calculation power of the system. During the different states the computer will most likely consume different levels of energy. The Green500 then uses the average of the energy consumption values that have been measured during the test.

More informations on the power measurement methodologies can easily be found on the Green500 website [2], [7].

2.3 The Little Green500 List

Since we now know how gaming with the test results can be easily prevented by introducing stricter measuring rules, we can focus on how the Green500 organisation has solved the criticism number one. Whilst some embraced now having a list of low-performance systems and touted the Top500 supercomputers, others said that the Green500, having a focus on energy efficiency instead of performance, should not be limited to a few 500 of the best systems world wide. There are many more efficient systems that have not made their way on the Top500 list. The solution to that was simple. In 2009, the little Green500 list was released. The little List has a different ruleset, which qualifies for participation:

1. Similar to the normal list, the little list also only ranks 500 computers.

2. To qualify for the list a computer must be at least as fast as the slowest supercomputer on the Top500 list 18 months ago.

This slightly less strict ruleset takes care of also representing computers, which are less powerful and thus less expensive than the ones listed on the Green500. This circumstance makes the little list a better representative of efficient computing for the broad mass than the Green500. Even though the little list has a great ideology that supports it, the list lacks regular updates. In August 2016 the latest list published was the list of November 2014 whilst the Green500 had its last update in June 2016.

Figure 4: The Green Destiny Computer as a single rack solution to HPC. Source: http://i1-news.softpedia-static.com/images/news2/Green-Destiny-2.jpg.

3 Supercomputers in Their Most Different Forms

The term "Supercomputer" reminds most of us on large facilities holding kilometres of racks filled with servers and hard drives, on data centers used by Facebook and Co., and on very fast, almost instantaneous computing. In some cases what we imagine a supercomputer to be is close to the reality, whereas in other cases the term "Supercomputer" can be quite misleading. It might be better to just call it High-Performance Computing. This less biased term is less known to many people and less filled with prejudice about form, size and capability of the systems. The reason this needs to be mentioned is because, in this Chapter, a few 'Firsts' regarding high-performance computing will be introduced, which don't fit into the standard picture of supercomputing.

3.1 The Green Destiny

Since also in HPC systems money is an important factor to consider, it is best to fit the system to the type of calculation work which needs to be done on the computer. One step of taking HPC or supercomputing to a small-sized level, for less power hungry calculations, was the Green Destiny (Figure 4), which was released in 2002. The Green Destiny is a 240-CPU Linux-based cluster that fits on about a half of a square meter. Short: a one rack solution to high-performance computing. It wasn't only the size which staggered computer architects, it also achieved low-power performance by using only as much energy as 2 to 3 hairdryers and unprecedented reliability by not overheating the processors. Despite its competitive performance then, many still felt that Green Destiny sacrifices to much performance to achieve the above. Still, Green Destiny has done an amazing job being one of the first of its kind, especially during a time in which energy efficient computing was merely frown upon. The figures in Table 1 simply underline this statement. The bold numbers mark a superlative and give a comparison between comparable supercomputers of the same time. For these figures the computer is also sometimes called the 'Green Density' [3].

Interestingly enough is the fact that even though the 11.6 MFLOPS/watt in a supercomputer where cutting edge at this time it wouldn't be enough to be ranked on the Green500 any more. Luckily, the computer architectures of HPC-systems have improved so much that even the least efficient computer on the Green500 is better in efficiency than the Green Destiny. This is good evidence that Green Awareness has already influenced the thinking of most system architects.

Machine	Avalon (1996)	ASCI Red (1996)	ASCI White (2000)	ASCI Q (2002)	Green Destiny+ (2002)
Performance (GFLOPS)	18	600	2,000	**8,000**	58
Area (m^2)	11.15	148.64	921.6	1950.96	**0.56**
Power (kW)	18	1,200	2,000	3,000	**5**
Compute Density (MFLOPS/m^2)	1,614	4,037	2,713	4,101	**103,571**
Power Efficiency (MFLOPS/watt)	1.0	0.5	1.2	2.7	**11.6**

Table 1: Performance and Efficiency Numbers for Clusters and Supercomputers. Source: The Importance of Being Low Power in High Performance Computing, Wu-Chun Feng, 2005.

Figure 5: The Orion DS-96.

3.2 High-Performance Workstations

Rather than just sticking with the idea of evolving the concept of the pre-existing Green Destiny as done by the MegaScale Computing Project, which used low-power architectures similar to the Green Destiny computer, a second more modest architectural approach to low power computing has emerged. This is a more practical and commercial evolution of the Green Destiny, trying to pack the performance and the energy efficiency in a box that would fit under your desk. The HPC workstation was born.

One of the front runners of such systems where the Orion DT-12 and the DS-96 (Figure 5). The two systems are arranged in clusters of 12 reps. 96 nodes. This architecture utilised certain trends which could be observed at the time of 2005:

1. The increasing popularity of cluster based HPC-systems.
2. The slow increase in workstation power compared to normal PCs (as seen in Figure 6
3. The high availability of open source cluster software

The rapid increase in supercomputers' performance and the stagnating performance gains in workstation power has left a wide gap (also Figure 6) for a cluster-based workstation such as the DT-12 or the DS-96.

Workstations do also have multiple advantages over traditional supercomputers [6]:

- Advantages:
  1. No or little need for external cooling since workstations aren't usually as densely placed as traditional rack-mount servers

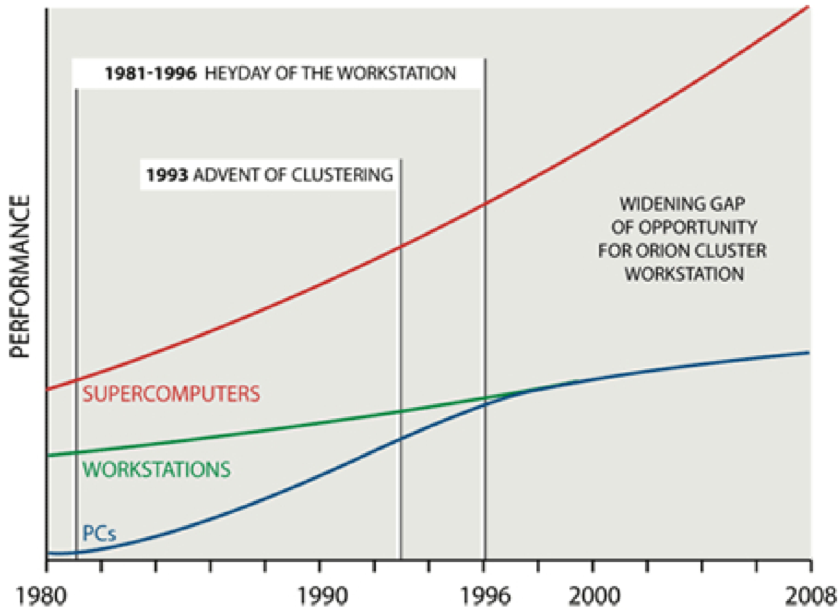

Figure 6: The Widening Performance Gap Between PCs and Supercomputers. Source: The Importance of Being Low Power in High Performance Computing, Wu-Chun Feng, 2005.

2. Workstation are much quieter than rack-mount servers
3. Normal work can be done on workstations as well, thus no need to purchase additional client computers

- Disadvantages:
  1. Not the ideal solution for very powerful HPC solutions that need many nodes to be arranged in a cluster
  2. Resources can not be allocated as efficiently as if a central HPC Unit manages the computations
  3. Some users might not even need workstation power

4 GreenIT - Just a Trend Towards a Cleaner Planet?

The Green500 organisation clearly states that environmental Protection is an important objective, which the Green500 can help to achieve. They say on their webpage:

> "In order to raise awareness to other performance metrics of interest [. . .], the Green500 offers lists to encourage stakeholders ensure that supercomputers are only simulating climate change and not creating [it]." (Green500.org, 2016)

Though GreenIT comes with many more advantages over traditional high-performance computing. Some that are even more tangible and relevant to architects that care less about the environment as the Green500 list does.

4.1 Total Cost of Ownership

The Cost of Ownership is something any architect would consider before getting into building the system. It can be split in two major factors. (1) the cost of acquisition which is obviously a one time cost including procurement, negotiation and purchase and (2) the Cost of Operation which can include the electrical bill, the cooling, rent or depreciation cost for the space in which the computer is deployed and the cost of administration, maintenance, or downtime [4].

While most people would expect the cost of acquisition to be much higher for expensive processors, main boards, and memory, while the cost of operation would be very low, since this is the case on most electrical devices used at home (otherwise everyones electrical bill would simply go through the roof), the reality looks much different. It has been found that some 1U servers' annual cost of operation exceeded the cost of acquisition after 2001. In the year of 2008, the cost of acquisition already matched the mere cost of energy consumed [5].

Since the power put into the systems converts to almost 100% into heat, cooling is an important thing to consider. In the infamous Lawrence Liverpool National Laboratory located in the Silicon Valley which has been the host of many supercomputers before. There, for every watt consumed by the computers an additional 0.7 watts was needed to cool the system down again. This leaves the LLNL with an annual cooling bill of $6M and a total of $14.6M including maintenance et cetera [3], [8].

This above trends can be very costly to operators of large scale systems. Low power-architectures or Power-aware architectures could end this trend and make supercomputing more scalable and affordable. The principle itself is very simple: a system that isn't able to produce as much heat doesn't need much cooling.

According to the same principle the Green Destiny Computer was engineered. The computer is now capable of operating in an 30 degree celsius dusty warehouse at around 2500m above sea level without any additional cooling apart from the internal cooling. The computer is simply not capable of producing so much heat that the processors could overheat.

4.2 Reliability

The next thing that needs to be considered is reliability or why a processor needs cooling at all. Well, obviously we want to prevent the processors from overheating and breaking. But we want to know why an operation temperature of 50 degrees is better than one of 80 degrees which would still be in the range where a processor would not break. The Answer is pretty simple (and for those who thoughtfully read the title of this paragraph, obvious). With increasing heat the computers tend to become less reliable. There also is a rule of thumb explaining this behaviour: For every 10 degree increase of temperature the mean time between two errors halves.

This can also be explained from the microelectronic perspective using the Arrhenius equation applied to microelectronics. It states that the time a condensator needs to be sufficiently charged to flip the switch in a CMOS-transistor increases with rising temperature. If a condensator can not be sufficiently charged within a clock cycle, which is usually an constant time in most older processors, the switch comes to a different result as expected, e.g. an error.

Wu-Chun Feng stated in his paper of 2005 [3] that while running a LINPACK benchmark on a dense beowulf cluster in 2002:

> "The cluster produced an answer outside the residual (i.e., a silent error) after only ten minutes of execution."

We could easily solve this problem by putting a ton of fans on top of each processor and next to every main board, though, this would not change the fact that the energy bills of most systems are already to high. Not producing as much heat achieved by utilising energy efficient hardware would obviously be the better way to go.

5 Power-Awareness or Low-Power

That using energy efficient hardware as suggested before is not the perfect solution has been said by many critics before. Low-Power architectures simply sacrifice to much of the performance, which is needed to be considered a 'real' HPC system. Generally, there are two different concepts in HPC One called capability computing, the other capacity computing. Each concept has a different approach to save power.

5.1 Capability Computing

Capability Computing is usually the term with which the ability to solve a single huge problem in the shortest time possible is described. A computer designed for this must be ready to be challenged with large scale simulations and tasks, which usually run for quite a while. This is the "Ferrari" of HPC-systems. These are the systems the critics of low-power architectures compare them to.

If a computer needs to be designed just for a single purpose that is to be as fast as possible a low-power approach, which sacrifices lots of performance is probably not ideal. Though, rather than using special low-power equipment, these computers, even when being designed for power-awareness, use high-performance, high-power CPUs that support a mechanism called *dynamic voltage and frequency scaling*. The basic idea, which is using high-performance gear, allows the user to demand as much calculation power as needed whilst being able to scale down the power usage and the performance by cutting down the voltage and the clock speed of the processor. You can read in another paper of the GreenIOT series how the concept works in detail.

The results using this technique, though, are astonishing: On some systems the could reduce the average CPU energy consumption by 20% whilst only losing 3% of performance (2005). In more resent test these numbers have been far surpassed due to advances in the software that controls the DVFS. This would be the power-aware approach of saving energy in computing.

5.2 Capacity Computing

Capacity Computing, opposed to capability computing, is describing a type of system that are typically used to solve a multitude of smaller problems simultaneously. High-performance is therefore not the main objective of these computers. Since the difference between capability and capacity computing is that capacity computing system is never expected to solve a large problem in the shortest time possible, the hardware can easily be fitted to the systems needs. The main criticism on low-power systems such as Green Destiny, MegaProto, and the Orion DS-96 is that they can not reach such a potential. Though, for capacity computing a high computing potential is not needed.

Making these computers as efficiently as possible is then a case of purchasing the right hardware in the first place. High-performance CPUs are not the way to go in those systems, instead they use low-power CPUs with lower clock speed to reduce the energy consumption in the system.

6 Resume

The Green Thinking in the IT branch is here to stay. The rapid improvements in power efficiency evidenced by the first few ranks of the Green500 list is enough proof for that. Though, the Green500 list is not to be seen as the holy grail of efficient computing. As with any ranking system it has faults that either can not be repaired or are infeasible to conquer. This includes the measurement of power, which is not made easier with the growing size of some systems, the restriction to only some of the best supercomputers and some problems with gaming the list, which are, due to some rules, probably still possible. It is also to keep in mind that every computer is designed for a single purpose in which it is supposed to be best or the most efficient at. If it wasn't designed for linear algebra solving it might do worse at the LINPACK benchmark than a computer thats only purpose is to do well on this test.

Though, with all this criticism in mind the main idea of the List was to propel the concept of energy efficiency in large scale computers (also some that are not included in the Top500 list, evidenced by the little list). And wether it was the list or just common sense of computer architects that for once were looking at their electrical bill, the downtime of the systems and the cooling, which was needed to encouraged the rethinking of power-awareness in supercomputing doesn't really matter here. As long as the list is looked at from a sceptical point of view one will easily be able to understand the intention of the list.

As a result of the GreenIT movement we will most likely expect to see already today and in the future that systems especially in capability computing, will be more and more based on power-aware solutions to bringing the power-consumption down and also increase reliability and availability of the systems. Low-power architectures will probably only be reserved to capacity systems, which also plays a central role in High-Performance Computing nowadays. We will therefore hope to see many more projects such as the Green Destiny and the Orion DS-96.

References

[1] Cameron, Kirk W., *HPC Power Efficiency and the Green500*, HPCwire.com, 20.11.2013, http://www.hpcwire.co/2013/11/20/hpc-power-efficiency-green500/, 08.08.2016.

[2] *Energy Efficient High Performance Computing Power Measurement Methodology*, Green500.org, https://www.top500.org/static/media/uploads/methodology-2.0rc1.pdf, 15.08.2016.

[3] Feng, Wu-Chun, *The Importance of Being Low Power in High Performance Computing*, CTWatch Quaterly, Volume 1, Number 3, August 2005, http://www.ctwatch.org/quarterly/articles/2005/08/the-importance-of-low-power-in-high-performance-computing/, 08.08.2016.

[4] Feng, Wu-Chung, *Making a Case for Efficient Supercomputing*, ACM Queue, Volume 1, Issue 7, 05.12.2003, http://queue.acm.org/detail.cfm?id=957772, 10.08.2016.

[5] Scogland, Tom, Balaji Subramaniam, and Wu-chun Feng, *The Green500 list: escapades to exascale*, Springer-Verlag, 23.05.2012, http://people.cs.vt.edu/~balaji/docs/papers/green500_isc12_balaji.pdf, 13.08.2016.

[6] Smith, Jan, *Workstations, Servers and Clusters - What is HPC in 2014?*, Microway, 13.02.2014, https://www.microway.com/hpc-tech-tips/workstations-servers-and-clusters-what-is-hpc-in-2014/, 15.08.2016.

[7] Subramaniam, Balaji, and Wu-chun Feng, *Understanding Power Measurement Implications in the Green500 List*, IEEE Computer Society, 2010, https://www.top500.org/files/green500/balaji-power-meas-green500-greencom10.pdf, 13.08.2016.

[8] Woods, Andy, *Cooling the Data Center*, ACM Queue, Volume 8, Issue 3, 08.03.2010, https://queue.acm.org/detail.cfm?id=1737963, 08.08.2016.