Noname manuscript No.
(will be inserted by the editor)

The Green500 List: Escapades to Exascale

Tom Scogland, Balaji Subramaniam, and Wu-chun Feng
{`tom.scogland,balaji,wfeng`}`@vt.edu`
Department of Computer Science
Virginia Tech
Blacksburg, VA 24060

Abstract Energy efficiency is now a top priority. The first four years of the Green500 have seen the importance of energy efficiency in supercomputing grow from an afterthought to the forefront of innovation as we near a point where systems will be forced to stop drawing more power. Even so, the landscape of efficiency in supercomputing continues to shift, with new trends emerging, and unexpected shifts in previous predictions.

This paper offers an in-depth analysis of the new and shifting trends in the Green500. In addition, the analysis offers early indications of the track we are taking toward exascale, and what an exascale machine in 2018 is likely to look like. Lastly, we discuss the new efforts and collaborations toward designing and establishing better metrics, methodologies and workloads for the measurement and analysis of energy-efficient supercomputing.

1 Introduction

As with all the great races in modern history, the supercomputing race has been myopically focused on a single metric of success. With the space race, the metric was which country could reach the moon first. In supercomputing, the race is more open-ended, but it focuses on maximum achievable performance. This persistent drive toward more speed at any cost has brought about an age of supercomputers that consume enormous quantities of energy, resulting in the need for extensive and costly cooling facilities to operate these supercomputers (Markoff and Hansell, 2006; Atwood and Miner, 2008; Belady, 2007).

In 2007, we created the Green500 (Feng and Cameron, 2007) to bring awareness to this issue and to provide a venue for supercomputers to compete on efficiency as a complement to the TOP500's (Meuer, 2008) focus on speed. Since that time, many milestones have been achieved, most notably, (1) the first petaflop supercomputer and (2) the first supercomputer GFLOP/watt supercomputer. Concurrently, efficiency has entered the consciousness of the supercomputing community at large and is now a primary concern in the design of new supercomputers. In this paper, we

- Explore the emergence of green high-performance computing (HPC).
- Track trends across the past four years of the Green500.
- Discuss the results of the shift toward thinking green.
- Investigate the implications of the above as we move into the future.

Of particular interest are the implications of past and current trends on the feasibility of exascale computing systems in the timeframe discussed in the DARPA IPTO exascale study (Bergman et al, 2008). The study itself indicates that power will be the determining factor in the success or failure of any exascale program. Past trends in the Green500 have led us to believe that at least one of the targets listed in the study may be feasible, but as we look at the progression in total power use and the aggregate progression of the list, some questions are raised as to whether the optimistic figure of 20 megawatts (MW) will be possible by 2018.

The rest of the paper is organized as follows. First, Section 2 discusses the background of and motivation behind

This work was supported in part by NSF grant CCF-0848670, a DoD National Defense Science & Engineering Graduate Fellowship (NDSEG), and Supermicro.

T. R. W. Scogland
2202 Kraft Drive
Blacksburg, VA 24060

B. Subramaniam
2202 Kraft Drive
Blacksburg, VA 24060

W. Feng
2202 Kraft Drive
Blacksburg, VA 24060

the Green500. Section 3 offers a high-level analysis of efficiency, along with power, and a more in-depth analysis of the efficiency characteristics of the different types of machines that achieve high energy efficiency. Projections from current and past lists to exascale are discussed along with their implications in Section 4. Section 5 presents the innovations and collaborations behind the continuing evolution of the Green500 as well as directions for future growth. Finally, Section 6 presents concluding remarks along with future work.

2 Background

Large-scale, high-performance computing (HPC) has reached a turning point. Prior to 2001, the cost of purchasing a 1U server exceeded the *annualized infrastructure and energy (I&E)* cost for that server , as depicted in Figure 1. By 2001, this annualized I&E cost for a server matched the cost of purchasing the server. By 2004, the *annualized infrastructure cost* by itself matched the cost of purchasing the server, and by 2008, the *annualized energy cost* by itself matched the cost of purchasing the server. Ignoring power and energy efficiency to pursue performance at any cost is no longer feasible. Data centers and HPC centers are already feeling the pinch across the industry from Yahoo! (Filo, 2009) to Google (Markoff and Hansell, 2006) to the National Security Agency (Gorman, 2006; croptome.org, 2008).

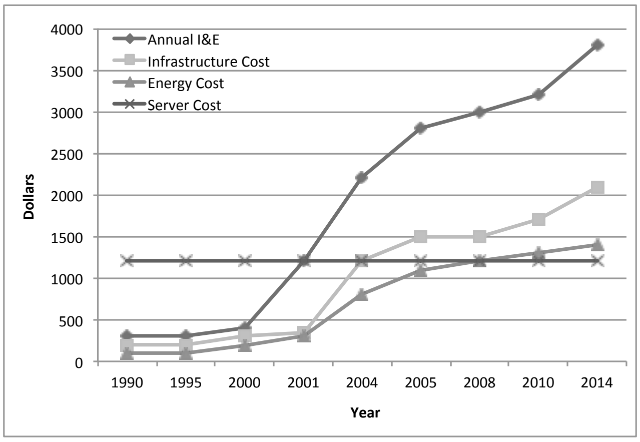

Fig. 1 Annual amortized costs in the data center (Belady, 2007)

The supercomputing community has been particularly guilty of seeking speed at the exclusion of all else. When we released the first Green500 in November of 2007, the computers that resulted from the event known colloquially as *Computenik*, were still in the TOP500, namely the Earth Simulator supercomputer from Japan and ASCI Q from the U.S. The former far outstripped the performance of the latter by *five-fold*. Furthermore, the Earth Simulator had an efficiency of only 5.60 MFLOPS/W while ASCI Q had an even lower score of 3.65 MFLOPS/W.

In 2004, the first U.S. machine to regain the number one position on the TOP500 was an IBM BlueGene/L prototype. It delivered *two orders of magnitude* better energy efficiency than the Earth Simulator and ASCI Q and debuted at 205 MFLOPS/W on the first Green500. Clearly, there was an inkling of the need for efficiency to achieve the greatest possible performance, but even so, it continues to be an uphill battle.

The original goal of the Green500 was to *raise awareness* of the state of energy efficiency in supercomputing and bring the importance of energy efficiency to the community on par with the importance of performance. In order to measure and rank supercomputers in terms of their energy efficiency, the Green500 employs the LINPACK benchmark, as provided by the TOP500 list, combined with power measurements. LINPACK solves a dense linear algebra problem using LU factorization and backward substitution. It is designed and tuned for load balancing and scalability.

While the Green500 remains a ranking of the efficiency of the 500 fastest supercomputers in the world, we launched three additional lists in 2009, based on feedback from the HPC community: the Little Green500, the HPCC Green500 and the Open Green500. As of now, the HPCC Green500 and Open Green500 have been discontinued due in part to lack of participation and interest from the community. The Little Green500 continues to operate and broadens the definition of a supercomputer to help guide purchasing decisions for smaller institutions. To be eligible for the Little Green500, a supercomputer must be as "fast" as the 500th-ranked supercomputer on the TOP500 list 18-months prior to the release of the Little Green500.

3 Efficiency

Across these last four years, we have observed a steady climb in the energy efficiency of the Green500. We have been tracking the average efficiency as compared to Moore's Law and finding that the average does track closely, while the maximum surges ahead, improving at a rate faster than Moore's Law. The extreme weight at the high end of this list, exemplified by IBM's BlueGene/Q machines occupying the top *five* slots of the current Green500, draws the average up to closely track Moore's law while the median lags well behind, as shown in Figure 2. With each release, we see the distance between the high efficiency machines and the average grow wider, to the point where many of the top machines can be considered to be outliers as they are more than 1.5 times the interquartile range above the median.

While on the topic of the high end of the list, we have a first this release — for the first time, the maximum efficiency of the list actually decreased from the June 2011 list to the November 2011 list. The BlueGene/Q supercomputer at #1 grew in size and evidently lost efficiency as a result. Even so, the four production BlueGene/Q computers hold the top four slots with an energy efficiency of approximately 2 GFLOPS/W, as shown in Table 1. The orig-

#	Gf/W	Computer
1	2.026	BlueGene/Q Custom (IBM Rochester)
2	2.026	BlueGene/Q Custom (IBM Watson)
3	1.996	BlueGene/Q Custom2 (IBM Rochester)
4	1.988	BlueGene/Q Custom (DOE/NNSA/LLNL)
5	1.689	Blue Gene/Q Prototype 1 (NNSA/SC)
6	1.378	DEGIMA: ATI Radeon GPU (Nagasaki U.)
7	1.266	Bullx B505, NVIDIA 2090 (BSC-CNS)
8	1.010	Curie: Bullx B505, NVIDIA M2090 (GENCI)
9	0.963	Mole-8.5: NVIDIA 2050 (CAS)
10	0.958	Tsubame 2.0: NVIDIA GPU (TiTech)
11	0.928	HokieSpeed: NVIDIA 2050 (VaTech)
12	0.901	Keeneland: NVIDIA Fermi (GaTech)
13	0.891	PLX: iDataPlex DX360M3, NVIDIA 2070 (SCS)
14	0.891	JUDGE: iDataPlex DX360M3, NVIDIA 2070 (FZJ)
15	0.889	Chama: Xtreme-X GreenBlade (SNL)

Table 1 The greenest 15 of the Green500 by rank

inal BlueGene/Q prototype holds the fifth slot. Below the BlueGene/Q machines, two GPU-accelerated supercomputers separate themselves from the rest of the GPU-accelerated pack: the DEGIMA cluster of AMD/ATI GPUs at #6 and the Bullx B505 machine at the Barcelona Supercomputing Center at #7. The GPU-accelerated supercomputers continue all the way down through #14, namely Curie, Mole-8.5, Tsubame 2.0, HokieSpeed, Keeneland, PLX, and JUDGE, respectively. The trend of GPU-accelerated supercomputers is broken by a surprisingly efficiency commodity CPU machine, with Intel Sandy Bridge - EP processors, from Appro at #15, followed by NVIDIA 2090-based GPU supercomputers down to #30 and then two more of the aforementioned Appro clusters and ending with the K computer at #33, which also happens to be the fastest supercomputer in the world on the TOP500.

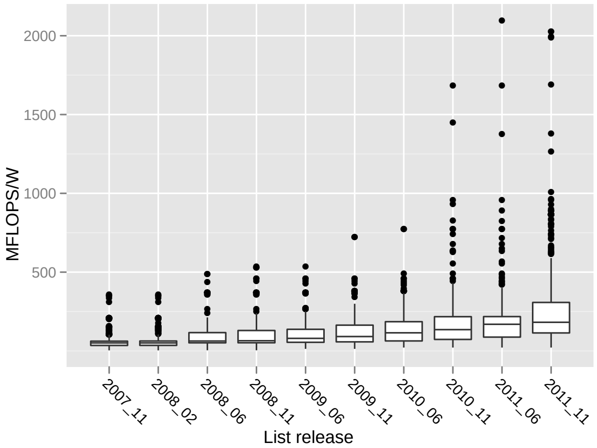

Fig. 2 Efficiency statistics across Green500 releases.

Figure 3 shows the overall power characteristics of the list. The focus of the Green500 is energy efficiency, trying to do the most work for the least energy. That said, the power drawn by machines on the list has not decreased. Far from decreasing, it has not yet stopped increasing, nor has it meaningfully slowed. While the average energy efficiency of the list has increased by four times, the power has more than doubled. In fact, despite its efficiency, the K computer currently at #1 on the TOP500 draws a whopping 12 *megawatts* of power, more than half the optimistic estimate for the power required to run an exaflop supercomputer in the latter part of the decade at 1/100th of the performance. We will discuss trends toward exascale in greater detail in Section 4.

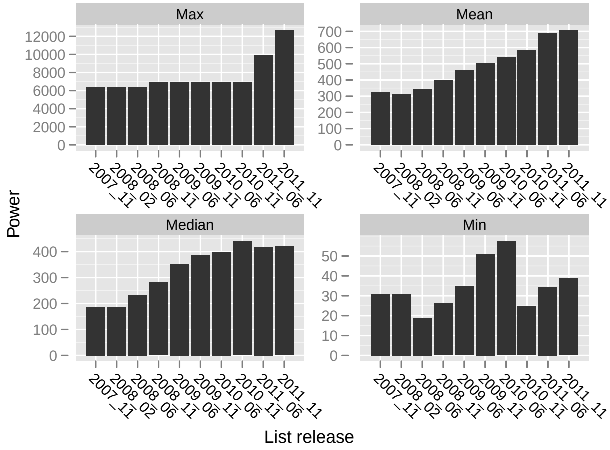

Fig. 3 Power statistics over time

Energy efficiency and power do not represent the whole picture, however. The performance efficiency of machines can be equally important; that is, the percentage of a machine's peak performance that is actually achieved when running LINPACK. Ideally, every machine should have a performance efficiency of 100%, but as you can see in Figure 4, that is far from the case. The worst-case performance efficiency was found in 2010 at below 20% and a large number of machines in the midrange of energy efficiency had performance efficiency below 40% in 2009. However, as each list goes by, more and more machines at the *high* end drop below 50% efficiency.

This trend typifies the split currently occurring in the supercomputing community. Every supercomputer that is at the top of the Green500 this year is based on aggregating large numbers of lower power cores. In the beginning, the first list in November of 2007, multi-processor machines were relatively common, one to four CPUs, each with one to two cores. At that time, the most efficient machines on the list were IBM BlueGene systems with 64 cores or 128 cores for L and P versions, respectively. Since then, the number of cores in a node of the most efficient systems has grown continually. Last year, the greenest 10 supercomputers contained on the order of 2,300 SIMD cores per node while the rest of the list averaged 10-15 cores per node. Overall, the more tightly coupled and smaller core-based machines rise to the top, but what kind of cores and how they are aggregated is important. There are three different approaches being taken toward building energy-efficient supercomputers.

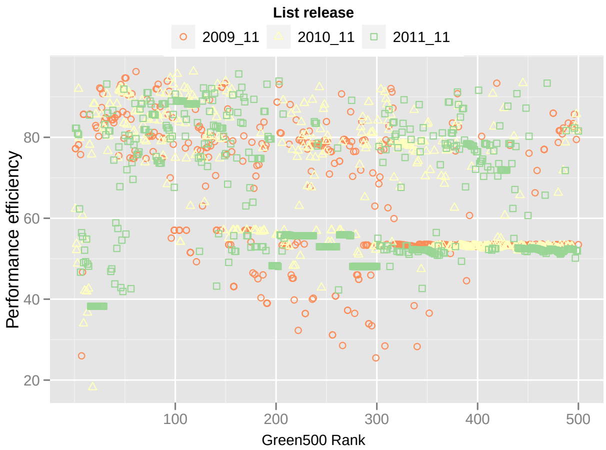

Fig. 4 Performance efficiency of systems over time

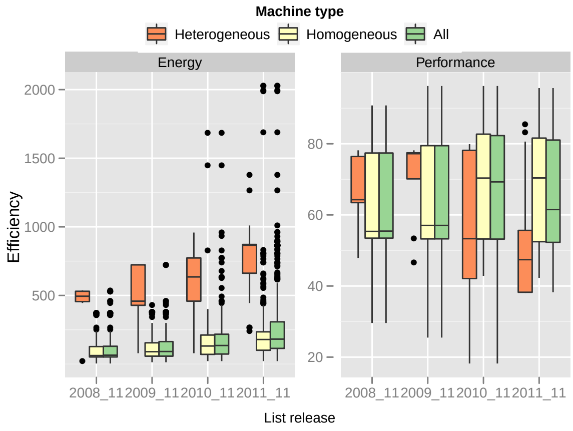

Fig. 5 Performance efficiency of heterogeneous and homogeneous systems over time

3.1 Heterogeneous Accelerator-Based Clusters

In the current list, this could just as well be labeled GPUs, as the only other heterogeneous accelerator supercomputers left are CellBE machines, which are no longer real contenders for the top of the list. At present, the accelerator supercomputer is a commodity cluster enhanced with high-bandwidth accelerators. High bandwidth means custom memories, e.g., GDDR or XDR, combined with a large number of simpler cores that can execute large amounts of simple calculations at a high rate. These accelerators do not handle latency-bound tasks very well, however, and as such tend to achieve very poor performance efficiency (35-60%) and account for nearly all the low performance-efficiency machines at the top of the list, as depicted in Figure 4.

Their increasing popularity, coupled with extreme energy efficiency and performance inefficiency, is enough to significantly sway the overall efficiency of the list as a whole. For example, the left half of Figure 5 shows that the energy efficiency of the last four November list releases exhibits a clump of much higher energy-efficient machines using accelerators, along with a number of outliers, which will be discussed later. In contrast, the right half of Figure 5 points to the median performance efficiency being dragged down almost 10% by the accelerator-based machines for the latest edition of the list. Previous lists did not show such a drop because of the high performance efficiency of the CellBE-accelerated machines, which were more common in those past lists.

3.2 High-Density Custom Supercomputers

This group of supercomputers has held the eyes of the efficient computing world ever since its debut in November 2004 with IBM BlueGene/L. For the inaugural Green500 in November 2007, IBM BlueGene/P topped the list. Since then, IBM's BlueGene line continues to make great strides in efficiency, culminating in the top five machines on our current list. That said, IBM is not the only player in this game, the #1 machine on the TOP500, the K computer, can be classified similarly. These machines achieve both very high energy efficiencies and very high performance efficiencies.

Interconnect	2007	2008	2009	2010	2011
Custom/Proprietary	26	12	18	16	5
InfiniBand	4	18	12	13	25
Gig-E	0	0	0	1	0
Green 30 % custom	87%	40%	60%	53%	17%
Overall % custom	13%	12%	9%	10%	12%

Table 2 Interconnect statistics for the greenest 30 machines

While accelerator-based systems frequently depend on the efficiency of the accelerators to enhance otherwise commodity systems, custom systems are designed from the ground up for efficiency. One of the key differences is the design of the interconnects in these systems. Some accelerator systems take advantage of custom interconnects, but they are significantly more common in custom designed machines, allowing for much more efficient communication and higher performance efficiency. Systems with Infiniband interconnects tend to top out around 85% efficiency, the benefits have been underscored by the K Computer, which attained a performance efficiency of 96% in June even at the scale of the fastest computer on the TOP500. Table 2 summarizes the statistics for the 30 greenest supercomputers over time. The table shows a traditional split between custom interconnects and Infiniband. The major shift this year is that the custom interconnects have become far less common at the top of the list, while becoming more common again in the list as a whole. This correlates highly with the emergence of accelerator-based systems. In fact, the only reason that there are custom interconnects in the top 30 at all this year is the set of five IBM BlueGene/Q systems in the top five slots.

3.3 Commodity Clusters

Commodity clusters continue to comprise the majority of the Green500. Advances in processor technology and the industry's embracing of lower-power CPUs and cloud computing have given birth to a series of increasingly efficient

List Portion	2007	2008	2009	2010	2011
Overall	76.85	61.52	53.71	49.302	41.26
Greenest 20	120	76.75	71.25	46.45	35.40
Top 20	93.75	69.25	66.5	56.05	43.40
Greenest 20 w/ accel	120	76.75	71.25	48.68	40.83

Table 3 Average minimum feature size in nanometers

general-purpose processors. A number of clusters using Intel's Sandy Bridge and AMD's Opteron 12-core processors can now surpass the efficiency of the K computer and even the NVIDIA C2090 GPUs for efficiency. One such system at #45 is the first acknowledged submission of a cloud resource, a piece of Amazon's self-made Sandy Bridge-based EC2 cloud infrastructure.

The major reason that the energy efficiency of commodity systems continues to increase is the ever-shrinking feature size on newer processors, allowing lower power draw for the same work. In other words, if a smaller feature size means a more efficient processor, then it stands to reason that the most energy-efficient machines should have (on average) processors with the smallest feature size. To test this, we created Table 3, which shows the minimum feature size (on average) for CPUs of machines on the Green500 over time. On average, the greenest machines had larger feature size than the rest of the list before last year, but now it seems to be coming out as expected. Even when accounting for the accelerator in a system as the primary processing element, rather than the CPU, the greenest 20 have (on average) a smaller feature size than machines lower down on the list. It is an interesting shift, which correlates with the rise of accelerator-based machines and efficient machines made of mass-market components. The custom machines that kept the feature size large in the past are becoming less common, and even the custom machines, such as the BlueGene/Q, are getting processors with a more comparable feature size as well.

4 Projections

The next major target for the supercomputing industry is the exascale system. Planning for that milestone, DARPA's Exascale Computing Study (Bergman et al, 2008) analyzes the different paths and hurdles between the current state of the art and the goal of an exaflop system in 2018-2020. A major focus of the study is power, with predictions that an exascale system may require as much as 100 MW to run, but with the goal being a 20-MW exaflop supercomputer. Given that we have seen the #1 supercomputer on the TOP500 list consuming over 12 MW at a mere fraction of the performance, the 100 MW mark even seems a lofty goal. As the time draws nearer, we have begun to look at the data collected as part of the Green500 to see how well technologies are progressing toward these goals.

One telling measure of our progress is a simple projection of the amount of power it would take to run an exaflop machine made entirely out of components available today. To make that prediction we choose a machine, in this case either the top machine from the Green500 or TOP500 lists, and naïvely expand it to exascale by assuming linear scaling of both performance and power. The result of doing this for every top machine since the first Green500 in 2007 is plotted in Figure 6 along with a trend-line and 30% confidence interval for each series. First, note that the Y-axis of the figure, representing power consumption, is measured in *gigawatts* (GW). Efficiency has come a long way over the past four years. The most powerful machine in November 2007 gives us a prediction of almost 5 GW for exascale while the most energy-efficient machine in November 2007 extrapolates out to a marginally more palatable figure at slightly under 3 GW. Now, however, we have reached a point where a system might be built at as low as 500 MW, assuming the greenest machine as a basis. Still that's a far cry from the goal of 20 MW in 2018, and that's where the trend comes in. We look forward to the coming years in eager anticipation as the current trend is so drastically dropping in power consumption that a bound must be fast approaching. While that is a strong statement, if the current trend continued, our most efficient supercomputers would be *generating* electricity within the next two years.

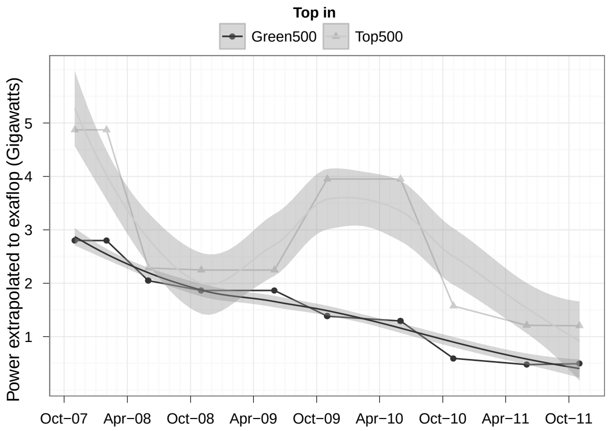

Fig. 6 Power projections for exascale systems

Another way to measure progress is to take a more holistic view of the list and include all the machines that do not meet the extreme conditions of Figure 6. When we plot every machine as a point on a graph of power vs. efficiency, as shown in Figure 7, certain trends become clear. First, for each subsequent list, the bottom moves outward. That is, we see power and efficiency increasing in roughly equal measure along the log-scale plot. Second, the plot shows a clear trend that has not shifted in four years. To show where that trend is taking us, two machines have been added to the plot for 2018, one at the 100-MW predicted position and one at the 20-MW DARPA target. The data shows us on track to hit the 100-MW number dead on,but that it will take a major outlier to achieve the 20-MW goal.

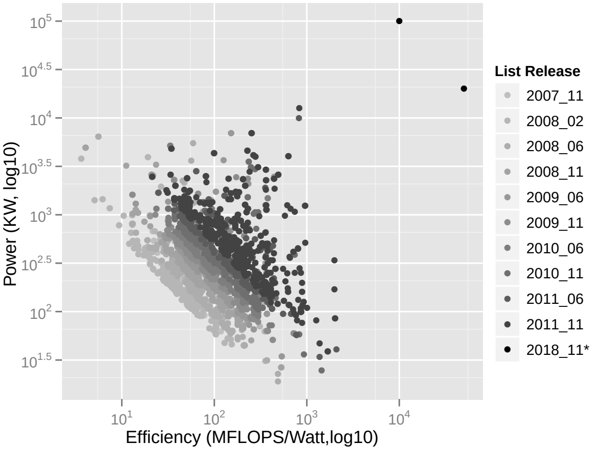

Fig. 7 Power vs. energy efficiency across lists

5 Future Directions

The Green500 was officially launched in 2007 to bring the issue of energy efficiency to the consciousness of the supercomputing community and to give a venue to those who wanted to innovate and improve this area. Now, as then, we strive to further the cause of energy efficiency in supercomputing. Since the first list, we have introduced three exploratory lists, established official run rules apart from those of the TOP500 list, and expanded our analysis extensively. In this section, we discuss some of the future directions envisioned for the list, along with how they have started to take shape over the last year. These fall into three primary categories: workloads, metrics, and methodologies.

5.1 Workloads

The workload required for a submission to the Green500 is the LINPACK benchmark, as run for the TOP500 list. This decision has been an important one, as the LINPACK benchmark focuses on one particular subset of the components in a supercomputer and always assumes the user wants to extract the best possible performance from that machine, which may not be the configuration which produces the most energy-efficient run.

Several solutions have been proposed to solve the issues with LINPACK, the most common being to use a different benchmark altogether such as HPCC or the Graph500 benchmark. In response, we created the HPCC Green500, and to date there has been one submission to that list. The result, while not what we would like, was not unexpected. Optimizing even a single benchmark for a supercomputer and holding the machine from production long enough to get results is trouble enough for most, asking them to run multiple benchmarks makes submissions highly unlikely. To ease the process of submission and to capture the efficiency of the different subsystems of a supercomputer, the Energy Efficient High Performance Computing Working Group, TOP500, The Green Grid, and Green500 collaboratively proposed the idea of using a subset of benchmarks with the intention of stressing different components of the system at SC11. For example, LINPACK and RandomAccess benchmarks can be used to stress compute and memory subsystems of a supercomputer, respectively.

	2007	2008	2009	2010	2011
Average Rank	76	123	162	106	129
Lowest Rank	176	496	445	404	328
Highest Rank	1	1	2	1	5

Table 4 Statistics on the TOP500 ranks of the 30 greenest supercomputers over time

As an alternative, we have been investigating the possibility of a load-varying LINPACK (Subramaniam and Feng, 2010a). Why would one want to vary the load since any variance from the maximum will lower performance? While true, systems do not always achieve their highest efficiency at highest load. Being able to vary the behavior of LINPACK in this way, we can use a benchmark that the industry is highly familiar with to produce a more useful energy-efficient result.

5.2 Metrics

For now, our chosen metric is MFLOPS/Watt, or millions of floating-point operations per second per watt. Much like the LINPACK workload, the MFLOPS/Watt metric has been under debate from the beginning. For example, MFLOPS/Watt appears to favor smaller machines over larger ones. As discussed in (Hsu et al, 2005; Feng and Lin, 2010) the performance of benchmarks, LINPACK included, scales at most linearly with the addition of new nodes to a cluster, while the power increases at least linearly. In other words, the larger a machine is, the less performance gain is achieved for a given increase in power. Given that truth, smaller supercomputers should be more energy efficient than their larger counterparts. However, Table 4 does not support that conclusion. Since 2007, only one November list has had the fastest, and largest, supercomputer outside of the 30 most efficient machines and that is this November. It is worth noting however that the K computer only falls out of that range by two at #32. Large, powerful machines have a tendency to fare well on the Green500. We believe this is a result of more effort being put into ensuring they are efficient as a result of their size.

As mentioned earlier, the Green500 and collaborative entities from the TOP500, the Green Grid, and EEHPCWG have been investigating the use of multiple benchmarks in order to capture the "true" energy efficiency of the system. However, using more than a single benchmark results in a mixture of benchmark outputs and it leads to the following question: "What metric should be used to capture all the benchmark results in a single number in order to rank the system?"

Prior to the above collaborative efforts, the Green500 team was involved with creating and investigating metrics like the Green Index (TGI) (Feng, 2010) to address this issue. The key idea behind TGI is to measure the energy efficiency of an HPC system with respect to a reference system. This approach is similar to the approach adapted by *Standard Performance Evaluation Corporation (SPEC)* (SPEC, 2012) for comparing system performance as shown in Equation (1).

$$\text{SPEC rating} = \frac{\text{Performance of Reference System}}{\text{Performance of System Under Test}} \tag{1}$$

The TGI of a system can be calculated by using the following algorithm:

1. Calculate the energy efficiency (EE), i.e., performance-to-power ratio, while executing different benchmark tests from a benchmark suite on the supercomputer:

$$\text{EE}_i = \frac{\text{Performance}_i}{\text{Power Consumed}_i} \tag{2}$$

where each i represents a different benchmark test.

2. Obtain the relative energy efficiency (REE) for a specific benchmark by dividing the above results with the corresponding result from a reference system:

$$\text{REE}_i = \frac{\text{EE}_i}{\text{EE}_{Ref_i}} \tag{3}$$

where each i represents a different benchmark test.

3. For each benchmark, assign a TGI component (or weighting factor W) such that the sum of all weighting factor is equal to one.
4. Use the weighting factors and sum across product of all weighting factors and corresponding REEs to arrive at the overall TGI for the system.

$$TGI = \sum_i W_i * \text{REE}_i \tag{4}$$

The Green500 team will further investigate the use of TGI and its incorporation into the Green500 in the future.

5.3 Methodologies

As more and more petaflop systems enter the Green500, one of the major concerns that needs to be addressed is how to measure the power consumed by such large-scale systems. In particular, the questions that begs to be asked in our quest to standardize the methodology (Subramaniam and Feng, 2010b) for power measurement in order to improve our run rules are as follows:

1. When should the power be measured? (For a certain period of time or for the entire benchmark execution?)
2. How should the power be measured? (Extrapolate from a single node or measure the power consumed by the entire system?)

To standardize methodologies for power measurement, it is critical to understand the computational characteristics of the benchmark in use (which is LINPACK in our case pending a shift to a new benchmark). LINPACK is a linear algebraic package which solves a dense system of linear equations. It runs in four stages: (1) random matrix generation, (2) LU factorization of the matrix, (3) backward substitution to solve, and (4) correctness checking. The second and third steps are used for calculating the LINPACK score (in GFLOPS) and require $O(N^3)$ and $O(N^2)$, respectively. Note that as the application progresses, the effective matrix size reduces and there is a corresponding drop in FLOPS as depicted in Figure 8, making the portion of the run that a measurement is made in highly important.

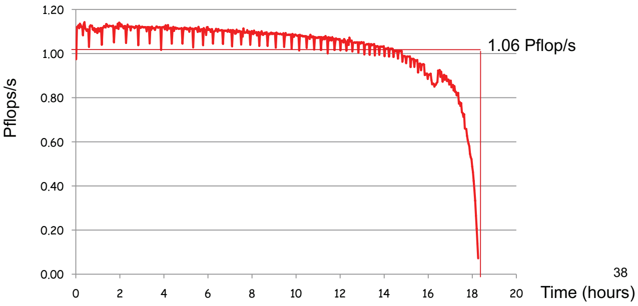

Fig. 8 Instantaneous FLOPS Rating Running LINPACK on Jaguar Supercomputer (Dongarra, 2010)

We expect the power profile of the LINPACK run to have trends related to its computational characteristics as time progresses. We analyze the profile of the newly installed HokieSpeed cluster at Virginia Tech to track the power consumption during a run. Figure 9 shows the power profile of HokieSpeed for a CUDA-LINPACK run extrapolated from one full rack (neither optimized to achieve best performance nor best energy efficiency).

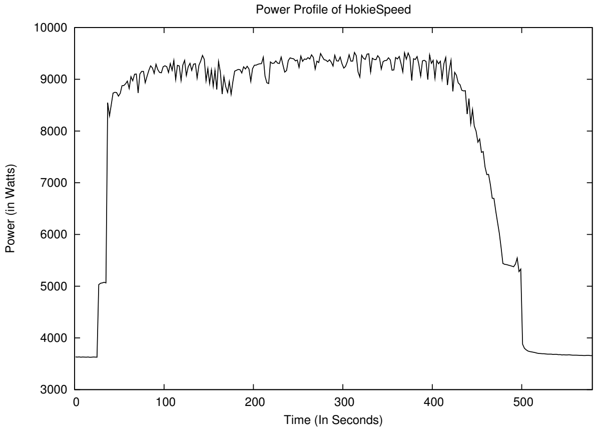

Fig. 9 Instantaneous Power Profile of HokieSpeed Supercomputer.

The trend in power profile clearly demarcates step 2 (factorization phase) and step 3 (solve phase) of the LINPACK benchmark from the other steps. The power consumed by the system ramps up as soon as the step 2 starts and gradually decays at the end of step 3. This indicates the region where the power consumption of the system would be at its highest and helps us narrow down the phases in which the power should be measured. The instantaneous minimum and maximum power consumed during these two steps are 8711 and 9518 watts respectively, which is a variation of approximately 8.4% (calculated as (max-min)/max power consumed), indicating that the accuracy of the power measurement increases as we measure larger percentage of the run. Such insights into the power profile of LINPACK led us to release a more rigorous update to the run rules for the Green500 in summer of 2010.

After this update to the run rules, the aforementioned collaborative effort began to standardize requirements for energy measurements of supercomputers. The first concrete results of this collaboration debuted at SC11 in the form of the EEHPCWG power measurement specification, which describes three levels of power measurement quality for evaluating supercomputers. In broad strokes, the levels are as follows.

1. One averaged power measurement over at least 10% of the run or one minute whichever is larger and measuring at least 1/64 of the machine or 1 kW worth, whichever is larger.
2. A series of equally spaced averaged measurements that begin before the run and continue after it is over, with at least 100 of these being during the run, the subset required is 1/8 of the machine or 10 kW.
3. A series of total energy readings, measured with a continual energy measurement device, includes the entire machine.

Within this system, the derived numbers reported on the Green500 may be considered to be level 0, as they do not represent actual measurements of hardware. Some results showing the difference between the levels as measured on a large DOE cluster were presented at SC. A particular benefit of higher levels is that it is much harder to game the system by using a small subset of machines which happen to have higher than average efficiency, a common issue with submissions today. Moving forward we hope that this and other methodological enhancements will make it easier to determine the quality of measurements not just on the Green500, but other resources as well.

6 Conclusion

Now that the Green500 has seen its fourth year come to an end, we have presented a comprehensive analysis of trends from this year back to the founding of the list. We have shown that while energy efficiency has become a prime concern, power usage has failed to stop, or even materially slow even as efficiency skyrockets. On the path to exascale, we have shown that some of the goals set by the DARPA exascale study may be feasible, but the optimistic figure is just that, along with some of the aspects of successful machines which may make it possible to reach even the most optimistic goal.

GPUs, which have been rising for years, really rose to the forefront of supercomputing this past November 2011, where nearly all the top 30 slots on the Green500 were GPU-based, a feat which has never before been matched by a new technology. We have seen several of our long-standing trends shift, as the unconventional highly commodity nature of GPU machines does not fit the previous model of a successful green supercomputer.

We also presented our work towards new metrics, methodologies and workloads for the measurement and analysis of green supercomputing. All of these areas are ongoing, and represent areas of future work, as we work continually to improve the list.

Acknowledgements

The authors wish to thank Mark Gardner, Heshan Lin, and Kirk Cameron for their support, and the Green500 community for their continued support and feedback, in particular participants at the Green500 Birds-of-a-Feather sessions over the last four years.

References

Atwood D, Miner JG (2008) Reducing Data Center Cost with an Air Economizer. White Paper: Intel Corporation

Belady C (2007) In the Data Center, Power and Cooling Cost More Than the IT Equipment It Supports. Electronics Cooling Magazine 13(1)

Bergman K, Borkar S, Campbell D, Carlson W, Dally W, Denneau M, Franzon P, Harrod W, Hill K, Hiller J, Karp S, Keckler S, Klein D, Lucas R, Richards M, Scarpelli A, Scott S, Snavely A, Sterling T, Williams RS, Yelick K, Kogge P (2008) Exascale Computing Study: Technology Challenges in Acheiving Exascale Systems

croptome.org (2008) NSA Electrical Power Upgrade. http://cryptome.org/nsa010208.htm

Dongarra J (2010) LINPACK Benchmark with Time Limits on Multicore and GPU Based Accelerators

Feng W (2010) Personal Communication with Sicortex

Feng W, Cameron K (2007) The Green500 List: Encouraging Sustainable Computing. IEEE Computer

Feng W, Lin H (2010) The Green500 List: Year two. In: Parallel & Distributed Processing, Workshops and Phd Forum (IPDPSW), 2010 IEEE International Symposium on, IEEE, pp 1–8

Filo D (2009) Serving up greener data centers. http://ycorpblog.com/2009/06/30/serving-up-greener-data-centers/

Gorman S (2006) NSA Risking Electrical Overload. In: The Baltimore Sun

Hsu C, Feng W, Archuleta J (2005) Towards Efficient Supercomputing: A Quest for the Right Metric. In: IPDPS '05: Proceedings of the 19th IEEE International Parallel and Distributed Processing Symposium - Workshop 11, Washington, DC, USA

Markoff J, Hansell S (2006) Hiding in Plain Sight, Google Seeks More Power. In: The New York Times

Meuer H (2008) The TOP500 Project: Looking Back over 15 Years of Supercomputing Experience. www.top500.org

SPEC (2012) The Standard Performance Evaluation Corporation (SPEC). http://www.spec.org/

Subramaniam B, Feng W (2010a) Load-varying linpack: A benchmark for evaluating energy efficiency in high-end computing

Subramaniam B, Feng W (2010b) Understanding power measurement implications in the green500 list. In: Green Computing and Communications (GreenCom), 2010 IEEE/ACM Int'l Conference on & Int'l Conference on Cyber, Physical and Social Computing (CPSCom)