Towards Efficient Supercomputing: Searching for the Right Efficiency Metric

Chung-Hsing Hsu, Jeffery A. Kuehn, and Stephen W. Poole
Computer Science and Mathematics Division
Oak Ridge National Laboratory
Oak Ridge, Tennessee, USA
{hsuc,kuehn,spoole}@ornl.gov

ABSTRACT

Efficiency in supercomputing has traditionally focused on execution time. In early 2000's, the concept of total cost of ownership was re-introduced, with the introduction of efficiency measure to include aspects such as energy and space. Yet the supercomputing community has never agreed upon a metric that can cover these aspects completely and also provide a fair basis for comparison. This paper examines the metrics that have been proposed in the past decade, and proposes a vector-valued metric for efficient supercomputing. Using this metric, the paper presents a study of where the supercomputing industry has been and where it stands today with respect to efficient supercomputing.

Categories and Subject Descriptors

C.4 [**Computer Systems Organization**]: Performance of Systems—*measurement techniques, performance attributes*; D.2.8 [**Software Engineering**]: Metrics—*performance measures*

General Terms

Design, Measurement, Performance, Standardization

Keywords

Energy efficiency, TOP500, Green500, SPECpower

1. INTRODUCTION

Efficiency in supercomputing has traditionally focused on definitions based on execution time and is often conflated with performance. It is commonly measured in terms of a calculation rate such as floating point operations per second (FLOPS) or instructions per second. In fact, this type of metric is conveniently used to define a supercomputer [36]. For example, the TOP500 project [39] ranks computers by how quickly each can solve the LINPACK benchmark [6]; the first 500 are called "supercomputers". The LINPACK benchmark has a fixed number of algorithmic steps to take for a given problem size, thus the quoted MFLOPS metric is a reference to number of such steps per second.

ICPE'12, April 22-25, 2012, Boston, Massachusetts, USA
Copyright 2012 ACM 978-1-4503-1202-8/12/04 ...$10.00.

In early 2000's, the concept of total cost of ownership (TCO) was re-introduced into the supercomputing community, with the introduction of efficiency measure to include aspects such as energy, space, reliability, and availability. All of these had been considered before, but were seemingly of less importance prior to this. There was a concern that we were designing new supercomputers with little consideration for the overall TCO. Many of the current leaders in the current TOP500 list consume multiple megawatts to just run the LINPACK benchmark, costing agencies like U.S. Department of Energy one million U.S. dollars per megawatt per year. As a result, there has been a substantial increase in the interest in pursuing efficient supercomputing.

An immediate question is how to quantify efficiency in supercomputing [16]. One possible metric is the performance-power ratio. For example, the Green500 project [13] re-ranks TOP500 supercomputers by LINPACK performance per watt (or equivalently, algorithmic steps per joule), referred to as FLOPS/W. However, this "miles per gallon" type of metric is criticized as being inappropriate for ranking supercomputers, due to its inability to track machines by size which may or may not reflect the total capability [36].

Today, the supercomputing community is still searching for an appropriate metric that can cover all major aspects of efficiency, while providing a fair basis for comparison [7]. This paper presents our journey in this search, focused on including both time and energy into the metric. We noticed that the struggle is not strictly limited to the supercomputing community. The enterprise server industry, for example, is also searching for a similar metric [8]. As a result, new metrics have been proposed. We observed a trend shared by many new metrics: shifting from a scalar-valued metric to a vector-valued metric, which inspired our work in this paper.

The contribution of this paper is a vector-valued metric for efficient supercomputing. The metric consists of two scalars, one for performance and the other for energy efficiency, in order to reflect the view that energy is as important as performance. Using the metric, the paper presents a study of where the supercomputing industry has been and where it stands today with respect to efficient supercomputing. In fact, the paper is more concerned about the *dimensionality* of the metric space, trying to make a case for vector metric. It is less concerned about the measurement rules for acquiring each scalar value. As we will see later, this decoupling allows us to plug into the real measurement results from various sources to conduct the analysis.

The rest of the paper is organized as follows. We first present an overview of the metrics proposed in the past decade and analyze their trends in Section 2. We then present a vector-valued metric, focusing on performance and energy efficiency in Section 3. We delay the introduction of the new metric until we provide a historical basis and enough groundwork to enable the the reader to judge the improvements offered. Following that, we discuss in details the use of the metric to study the historical trend of computer systems with respect to efficient supercomputing in Section 3.2. As an illustration, we compare the new ranking produce by the metric with the TOP500 list and the Green500 list. Finally, we conclude the paper in Section 4.

2. RELATED WORK

This section presents an overview of the performance and energy metrics proposed from the supercomputing community to the circuit design community. The emphasis will be on the *type* of metric and on the *trend* of shifting from a scalar-valued metric to a vector-valued metric. To start with, we give a general definition of efficiency: efficiency describes the extent to which effort or resource is *well* spent for the intended task. It is a measurable concept, quantitatively determined by the ratio of output to input.

2.1 Performance Benchmarks

Performance measurement of computer systems has been a focus of much standardization effort. Multiple industry consortia formed by competing vendors participate to improve the quality and ease of comparison for a particular audience [5], e.g., the Transaction Processing Performance Council (TPC) and the Standard Performance Evaluation Corporation (SPEC). Each consortium addresses a different audience or type of application. For example, TPC addresses on-line transaction processing and has two active benchmarks that measure the computer system performance in terms of transactions per second (TPS).

For supercomputer vendors, the performance on the LINPACK benchmark is currently the de facto standard. The benchmark solves a dense linear algebra problem. However, it is criticized as not being representative enough for typical supercomputer workload. As a result, there emerge efforts to add other supercomputing relevant benchmarks. For example, Graph 500 [11] measures the performance of graph search in edge traversals per second. SPEC MPI2007 [25] composes 13 benchmarks from several application domains. HPCC [15] consists of 7 tests for various system features.

Most standard benchmarks measure performance in terms of services per unit of time, although the definition of service is different. SPEC MPI2007 is an exception. It uses the speedup with respect to a reference machine as the metric for each benchmark. The final rating, a scalar value, is given by the geometric mean of these speedups. Of late there has been a trend towards using suites of benchmarks which report multiple performance values.

2.2 Energy Benchmarks

Standard energy benchmarks are also emerging. TPC, for example, augmented all its performance benchmarks with methods to measure and report energy consumption as joules per transaction (W/TPS). SPEC currently releases three benchmarks of this kind. One of them, SPECpower_ssj2008, measures transactions per joule at eleven different load levels [12]. JouleSort [30] represents an academic effort, which measures the energy required to sort a fixed number of randomly permuted records, reporting it as sorted records per joule. Poess et al. has a survey [29] comparing energy benchmarks from major industry consortia.

The proceeding benchmarks are generally service oriented. There are advocates for hardware oriented benchmarks because energy consumption strongly depends on workload [14], system configuration [28] and load level [33]. SWEEP [3] is one such example. It is an academic attempt to evaluate the energy efficiency of a server across the workload space through synthetic workload generation. Molka et al. [23] have a similar effort but for parallel workload generation. SERT [18] developed by SPEC is yet another example, and it targets server-class computer systems.

Most benchmarks measure energy efficiency in terms of services per unit of energy (i.e., the performance-power ratio). In terms of the trend, both performance and energy efficiency are reported, although some benchmarks distinguish between the primary metric and the secondary ones. The drive to hardware oriented benchmark will require multiple energy values to be reported.

For supercomputer vendors, the choices of energy benchmarks are limited. There is the Green500 project [13] that uses the LINPACK benchmark to rank supercomputers by their performance-power ratios starting from November 2007. This power data was also added to the TOP500 list in June 2008. The performance-power ratio is criticized, because it is an intensive metric and thus cannot be used to rank supercomputers by size; however, the ratio is useful for ranking technologies [36].

2.3 Energy Efficiency Metrics

Besides the performance-power ratio (i.e., services per joule), there are other useful types of energy metrics. One metric is the average power (i.e., joules per second). Another metric is the ratio of the energy consumption on the target machine relative to the reference machine [3], similar to the way the performance is defined in SPEC MPI2007.

A third metric is PUE (Power Usage Effectiveness) [37]. PUE measures how much of the total electricity used by a data center goes to the IT equipment, as opposed to being used on cooling systems and the power infrastructure. PUE was developed by an IT industry group, The Green Grid (TGG), in 2007, and is now widely uses. PUE is a percentage metric, with both input and output measured in the same dimension, energy.

PUE can be viewed as a measure for energy proportionality [1]. Energy proportionality is originally a design principle to ensure the energy consumption is proportional to the executed workload. This paper interprets it as the extent to which energy is *well* spent to the services delivered. In this sense, PUE considers energy to be well spent when it is consumed by the IT equipment, not by other equipments or lost during transmission or conversion.

There exist measures to quantify the energy proportionality of a server. For example, Varsamopoulos and Gupta [40] proposed the IPR metric to measure the power range and the LDR metric to measure the linearity. Ryckbosch et al. [31] proposed the EP metric to measure how closely the actual system approaches the ideal case (i.e., the power consumption is linear to the rate of service). SPECpower_ssj2008 includes some notion of energy proportionality, but does not explicitly quantify it [31].

For supercomputers, a task force from TOP500, Green500, TGG and the EE HPC Working Group [7] have been formed to develop a stronger set of energy efficiency metric(s). PUE is taken as an external constraint and held as an independent variable. The metric of interest is currently workloads per unit of energy where workloads are to leverage well established benchmarks. PUE is considered insufficient because it only compares energy use relative to the support infrastructure. As a result, an energy inefficient system may still have an excellent PUE value if its support infrastructure provides efficient power delivery and cooling. Some argue that an absolute metric, such as the performance-power ratio, should be reported as well [26]. Furthermore, while PUE is meant for tracking datacenter progress over time, it is now mis-used as a comparison tool between different data centers [38].

2.4 Composite Metrics

A composite metric in this case is the end product of trying to combine both performance and energy efficiency into a scalar-valued measure. They generally take the multiplicative form of

$$(\text{performance})^{\alpha} \cdot (\text{energy efficiency})^{\beta}$$

with parameters α and β. It is the choice of the parameter values that make composite metrics seem artificial.

For example, the low-power circuit design community typically uses a single index to guide design tradeoffs and faces a similar challenge to simultaneously optimize performance and power. As a result, researchers have proposed several metrics. Some of the popular metrics are in the form of ED^n [27] where E is the energy, D is the circuit delay, and n is a nonnegative integer; for example, the power-delay product (PDP, n=0), the energy-delay product (EDP, n=1) [10], and the energy-delay-squared product (ED2P, n=2) [20]. The larger the n, the more emphasis on performance.

The ED^n metric is also used for high-end computer systems. We see suggestions on using PDP for workstations, and EDP for servers [4] and supercomputers [32]. Ge et al. [9] proposed to generalize ED2P as $E^{(1-\gamma)}D^{2(1+\gamma)}$, $-1 \leq \gamma \leq 1$, for computer clusters. Bekas and Curioni [2] argued that D should be replaced by an application dependent function of D. Clearly, there is no consensus on the choices for α and β, if we interpret E and D as the reciprocal of energy efficiency and performance, respectively.

What may be agreed upon is the value of the multiplicative form. This can be seen from a figure of merit for mobile devices [19] to the SWaP metric for single servers [22] to a parameterized utility metric for supercomputers [36]. The construction principle is to multiply all desired quantities (such as performance) and divide them by undesired quantities (such as power and size).

3. A VECTOR-VALUED METRIC

This section presents the proposed vector-valued metric for efficient supercomputing. It starts with the definition of the metric, followed by illustrations on how the metric can be used. The trend of efficient supercomputing is analyzed and a new ranking induced by the metric is compared with the TOP500 list and the Green500 list. After that, the desired properties of the metric are listed and discussed.

3.1 The Metric

We view performance and energy efficiency as two separate dimensions of the efficiency metric. In other words, the metric is represented by a two-dimensional vector. We feel that any composite metric is biased one way or another. Transparency preserves the context of the data and enables end users to assess the relevance of the results to their specific application environments [18].

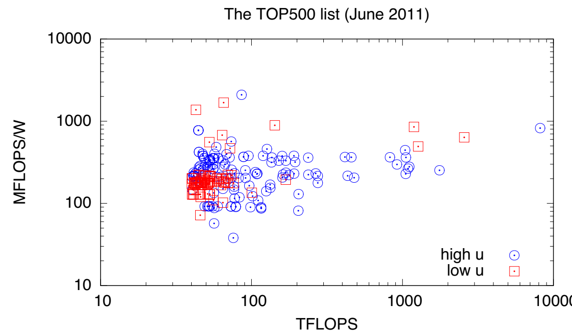

Figure 1: TOP500 in June 2011.

Performance and energy efficiency are defined in the typical way. They are measured in algorithmic steps per second and per joule, respectively. The metric can be visualized as a scatter plot, with the x-axis representing performance and the y-axis representing energy efficiency. The x-axis can also be viewed as the timeline as performance increases over time. A similar plot has been used elsewhere [3].

Note that the vector-valued metric assumes that energy is as important as performance. Furthermore, the two dimensions are separate but not independent. There have been several studies on characterizing the energy-time tradeoffs of a supercomputing application. Finally, although the metric cannot be used to create a total order of the computer systems, it can help generate a *partial* order.

3.2 The Use of the Metric

As an illustration of the value of the new metric, we first take the subset of TOP500 systems which have the power consumption data, and visualize the distribution of their metric values. The data is from the TOP500 list released in June 2011 with 186 unique metric values. Figure 1 shows the distribution of these values. We see that fastest machines have slightly better energy efficiency in general. As a result, these machines still remain at the top ranks in the Green500 list. This may ease the concern that energy efficiency is an intensive quantity, and ranking based on it would favor smaller-scale (and thus slower) systems. On the other hand, these systems do not have the highest energy efficiency. In fact, the energy efficiency of the TOP500 supercomputers are more similar than different, indicating that many of them are built with similar technologies.

The figure also shows that there is no clear winner in two competing designs for an advanced supercomputer. The CPU-based design achieves over 70% of the computational capability whereas the GPU-based design only achieves 50% utilization. GPU-based machines are typically advertised for their potential energy efficiency, but the low utilization of the computational capability leads them to have only comparable efficiency with CPU-based machines. (Note that the low utilization for the less powerful systems is due to the use of slower interconnect.) In the same spirit, there is an ongoing debate in the processor design community as to whether

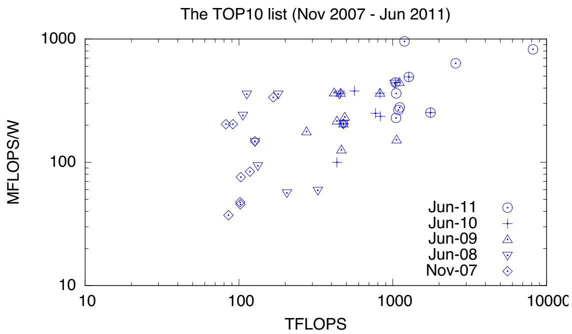

Figure 2: TOP10 over time.

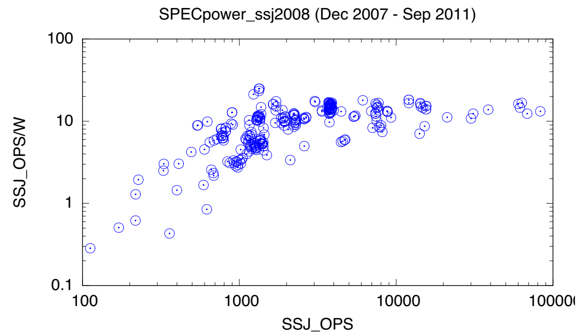

Figure 4: SPECpower_ssj2008 @ 100% over time.

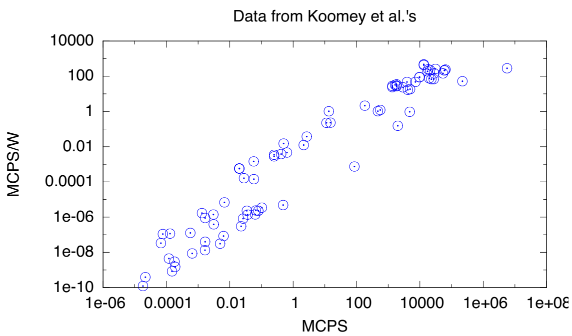

Figure 3: Koomey et al.'s data [17].

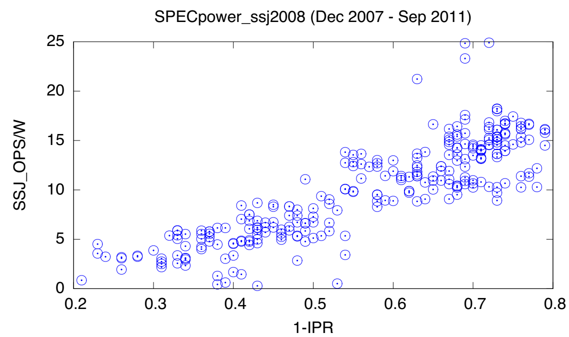

Figure 5: The correlation between energy efficiency and energy proportionality.

slower but more energy efficient "wimpy" processors, aggregated in large numbers, beat "brawny" processors [21, 24].

Figure 2 shows the metric values for the TOP10 supercomputers over different time periods from November 2007 until June 2011. We can see that both performance and energy efficiency are improved over the years. A similar observation can be made for the 10 most energy efficient supercomputers. The multi-dimensional improvement is most likely due to the combined effects of smaller transistor sizes, custom interconnects and more processing elements [34]. The figure also seems to suggest that performance and energy efficiency are improved at similar rates. This observation matches well with Koomey's Law [17] described below.

Koomey and his colleagues recently published a study showing that both performance and energy efficiency tracks very well with Moore's Law [17]. Specifically, they found that the energy efficiency of computation has doubled every 1.57 years from 1946 onward. This rate of improvement is slightly slower than that for personal computers (PCs), which saw efficiency double every 1.52 years from 1975 to 2009. Performance for PCs is doubled every 1.5 years during that time period. For comparison, we plot Koomey et al.'s data [17] as shown in Figure 3. The figure shows a linear correlation, meaning that both performance and energy efficiency are improved at steady rates. This correlation is quite different from what we see in Figure 1.

In order to make more sense out of the results above, we plot the metric values of individual servers benchmarked through SPECpower_ssj2008 at the 100% load-level. The data is from the public records [35] released between December 2007 and September 2011 with 270 unique metric values. Figure 4 shows the result. This figures is similar to Figure 1 in that systems with higher performance have similar energy efficiency. The figure also shows that the improvement over energy efficiency becomes much slower in recent years.

Further examination indicates that the energy efficiency of servers improves at a steady pace but slower than the performance improvement. A major driving force for the improvement of energy efficiency is the improvement of energy proportionality. Figure 5 shows the correlation between energy efficiency and energy proportionality using the SPECpower_ssj2008 data. We use 1-IPR as the measure for energy efficiency where IPR is the ratio of the idle power to the peak power. The figure shows that a more energy proportional server tends to have a higher energy efficiency.

Unfortunately, we cannot conduct similar analysis to the TOP500 supercomputers since the list does not report the idle power of each supercomputer. We conjecture that supercomputers have relatively lower (but more similar) energy proportionality, and therefore the variation of their energy efficiency is not as significant.

Finally, we want to comment on how the proposed metric can help create a ranking among supercomputers. Although the metric is vector-valued and thus cannot create a total order, it can generate a partial order. We define the partial order in the typical, mathematical manner. Specifically, system A is better than system B if both of its performance and energy efficiency are higher; otherwise, they are incomparable, meaning each system has different advantages. The partial order among all systems creates a directed acyclic graph

Table 1: The 10 most efficient supercomputers.

Rank	η_T	η_E	$\eta_T \cdot \eta_E$	(η_T, η_E)
1	M1	M109	M1	{M1,M5,M109}
2	M2	M165	M2	
3	M3	M430	M5	
4	M4	M5	M4	{M2,M22,M54,M165}
5	M5	M54	M10	
6	M6	M1	M3	
7	M7	M406	M8	
8	M8	M407	M6	{M3,M4,M430}
9	M9	M408	M12	
10	M10	M22	M7	

which can be converted into a layered graph. The layers in the graph enable us to create a ranking, one rank for each layer. Table 1 shows this new ranking with respect to other rankings. The notation Mn means the n^{th} supercomputer in the TOP500 list. Notations η_T and η_E represent performance and energy efficiency respectively. Notation $\{A, B\}$ means systems A and B are incomparable.

We see multiple supercomputers at the same rank, meaning that none of them dominates the other in terms of efficiency. For example, there are three systems M1, M5 and M109 at rank 1. Both M1 and M109 aggregate many low-power processors whereas M5 uses energy efficient accelerators. M1 delivers the highest performance; M5 and M109 are more energy efficient. M5 is GPU-based, consisting of 73,278 cores. In contrast, M109 is Cell-based with 8,192 cores. In other words, although M109 is more energy efficient than M5, M5 is larger in size and provides higher performance.

One novel aspect of the new ranking is that it identifies new "middle" classes. Consider Figure 1, a scatter plot of the TOP500 list with performance as the x-axis and energy efficiency as the y-axis. System M1 is the rightmost point in the plot whereas system M109 is the topmost point. System M5 represents a new class: machine that computes faster than M109 (the most energy-efficient supercomputer) and consumes less energy than M1 (the best-performance supercomputer). Ideally, the multi-dimensional space also provides a natural mechanism to capture some sense of distance, for example, the Euclidean distance which can then be used to cluster systems.

3.3 Further Discussion

In the following we list and briefly discuss a set of desired properties for a good efficiency metric.

1. *Higher is better*: This property looks for a "normalized" metric such that it represents efficiency, not inefficiency. There may or may not exist an upper bound for the metric value.

2. *Capture energy proportionality*: A consensus in the community is to define energy efficiency as the *useful* IT work per joule where how to measure the usefulness is not yet agreed upon.

3. *Not utilization based*: There is an expectation that the usefulness is not closely dependent on how the computer system is utilized. The 100% system utilization does not necessarily imply that the progress of the intended task is also at the full speed.

4. *Not biased*: The hope is to induce a fair comparison. However, this property is rather subjective. For example, some suggest not to favor large-scale machines [16] whereas others want this favoritism [36].

5. *Insightful*: A good metric not only identifies the best system but also finds the distance between two systems so as to help driving design decisions between two drastically different design directions.

4. CONCLUSIONS

This paper has examined the metrics to quantify efficient supercomputing in terms of performance and energy efficiency. Some metrics are driven by industry consortia while others are borrowed from the low-power circuit design community. Although there is not yet a consensus in the supercomputing community on what the right efficiency metric is, there is a trend of shifting from a scalar-valued metric to a vector-valued metric. We follow the same trend and propose a vector-valued metric for efficient supercomputing. Using the metric, the paper presented a study of historical data and current state of the art with respect to efficient supercomputing.

5. ACKNOWLEDGMENTS

This work was partially supported by the Extreme Scale Systems Center at Oak Ridge National Laboratory. The submitted manuscript has been authored by a contractor of the U.S. Government under Contract No. DE-AC05-00OR22725. Accordingly, the U.S. Government retains a non-exclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for U.S. Government purposes. We wish to thank the anonymous reviewers for their valuable comments.

6. REFERENCES

[1] L. Barroso and U. Hölzle. The case for energy-proportional computing. *IEEE Computer*, 40(12):33–37, Dec. 2007.

[2] C. Bekas and A. Curioni. A new energy aware performance metric. In *International Conference on Energy-Aware High Performance Computing*, Sept. 2010.

[3] K. D. Bois, T. Schaeps, S. Polfliet, F. Ryckbosch, and L. Eeckhout. SWEEP: Evaluating computer system energy efficiency using synthetic workloads. In *International Conference on High Performance Embedded Architectures and Compilers*, Jan. 2011.

[4] D. Brooks, P. Bose, S. Schuster, H. Jacobson, P. Kudva, A.Buyuktosunoglu, J.-D. Wellman, V. Zyuban, M. Gupta, and P. Cook. Power-aware microarchitecture: Design and modeling challenges for the next generation microprocessors. *IEEE Micro*, 20(6):26–44, November/December 2000.

[5] T. Conte and W. Hwu. Advances in benchmarking techniques: New standards and quantitative metrics. *Advances in Computers*, 41:231–253, 1995.

[6] J. Dongarra, P. Luszczek, and A. Petitet. The LINPACK benchmark: Past, present, and future. *Concurrency and Computation: Practice and Experience*, 15(9):803–820, Aug. 2003.

[7] Energy Efficient High Performance Computing Working Group. http://eehpcwg.lbl.gov.

[8] A. Fanara, E. Haines, and A. Howard. The state of energy and performance benchmarking for enterprise servers. In *TPC Technology Conference on Performance Evaluation and Benchmarking*, Aug. 2009.

[9] R. Ge, X. Feng, and K. Cameron. Improvement of power-performance efficiency for high-end computing. In *Workshop on High-Performance, Power-Aware Computing*, Apr. 2005.

[10] R. Gonzalez and M. Horowitz. Energy dissipation in general purpose microprocessors. *IEEE Journal of Solid-State Circuits*, 31(9):1277–1284, Sept. 1996.

[11] The Graph 500 list. http://www.graph500.org.

[12] L. Gray, A. Kumar, and H. Li. Workload characterization of the SPECpower_ssj2008 benchmark. In *SPEC International Performance Evaluation Workshop*, June 2008.

[13] The Green500 list: Environmentally responsible supercomputing. http://www.green500.org.

[14] D. Hackenberg, R. Schöne, D. Molka, M. Müller, and A. Knüpfer. Quantifying power consumption variations of HPC systems using SPEC MPI benchmarks. In *International Conference on Energy-Aware High Performance Computing*, Sept. 2010.

[15] HPC Challenge benchmark. http://icl.cs.utk.edu/hpcc/.

[16] C. Hsu, W. Feng, and J. Archuleta. Towards efficient supercomputing: A quest for the right metric. In *Workshop on High-Performance, Power-Aware Computing*, Apr. 2005.

[17] J. Koomey, S. Berard, M. Sanchez, and H. Wong. Implications of historical trends in the electrical efficiency of computing. *IEEE Annals of the History of Computing*, 33(3):46–54, July-September 2011.

[18] K.-D. Lange and M. Tricker. The design and development of the server efficiency rating tool (SERT). In *International Conference on Performance Engineering*, Mar. 2011.

[19] T. Makimoto, K. Eguchi, and M. Yoneyama. The cooler the better: New directions in the nomadic age. *IEEE Computer*, 34(4):38–42, Apr. 2001.

[20] A. Martin. Towards an energy complexity of computation. *Information Processing Letters*, 77:181–187, Feb. 2001.

[21] D. Meisner and T. Wenisch. Does low-power design imply energy efficiency for data centers. In *International Symposium on Low Power Electronics and Design*, Aug. 2011.

[22] S. Microsystems. The SWaP (Space, Watts and Performance) metric, 2007.

[23] D. Molka, D. Hackenberg, R. Schöne, T. Minartz, and W. Nagel. Flexible workload generation for HPC cluster efficiency benchmarking. In *International Conference on Energy-Aware High Performance Computing*, Sept. 2011.

[24] T. Mudge and U. Hölzle. Challenges and opportunities for extremely energy-efficient processors. *IEEE Micro*, 30(4):20–24, July/August 2010.

[25] M. Müller, M. van Waveren, R. Lieberman, B. Whitney, H. Saito, K. Kumaran, J. Baron, W. Brantley, C. Parrott, T. Elken, H. Feng, and C. Ponder. SPEC MPI2007:an application benchmark suite for parallel systems using MPI. *Concurrency and Computation: Practice and Experience*, 22(2):191–205, Feb. 2010.

[26] V. Pel. Energy efficiency aspects in Cray supercomputers. In *International Conference on Energy-Aware High Performance Computing*, Sept. 2010.

[27] P. Pénzes and A. Martin. Energy-delay efficiency of VLSI computations. In *ACM Great Lakes Symposium on VLSI*, Apr. 2002.

[28] M. Poess, R. Nambiar, and K. Vaid. Optimizing benchmark configurations for energy efficiency. In *International Conference on Performance Engineering*, Mar. 2011.

[29] M. Poess, R. Nambiar, K. Vaid, J. J.M. Stephens, K. Huppler, and E. Haines. Energy benchmarks: A detailed analysis. In *International Conference on Energy-Efficient Computing and Networking*, Apr. 2010.

[30] S. Rivoire, M. Shah, P. Ranganathan, and C. Kozyrakis. JouleSort: A balanced energy-efficiency benchmark. In *International Conference on Management of Data*, June 2007.

[31] F. Ryckbosch, S. Polfliet, and L. Eeckhout. Trend in server energy-proportionality. *IEEE Computer*, 44(9):69–72, Sept. 2011.

[32] V. Sarkar. ExaScale software study: Software challenges in extreme scale systems. DARPA IPTO Report, Sept. 2009.

[33] D. Schall, V. Hoefner, and M. Kern. Towards an enhanced benchmark advocating energy-efficient systems. In *TPC Technology Conference on Performance Evaluation & Benchmarking*, Aug. 2011.

[34] T. Scogland, B. Subramaniam, and W. Feng. Emerging trends on the evolving Green500: Year three. In *Workshop on High-Performance, Power-Aware Computing*, May 2011.

[35] SPECpower_ssj2008 Results. http://www.spec.org/power_ssj2008/results/.

[36] E. Strohmaier. Generalized utility metrics for supercomputers. June 2009.

[37] The Green Grid. The Green Grid power efficiency metrics: PUE and DCiE, 2007.

[38] The Green Grid. Recommendations for measuring and reporting overall data center efficiency (version 2), May 2011.

[39] TOP500 supercomputing sites. http://www.top500.org.

[40] G. Varsamopoulos and S. Gupta. Energy proportionality and the future: Metrics and directions. In *International Workshop on Green Computing*, Sept. 2010.