

Measuring Server Energy Proportionality

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

ABSTRACT

In performance engineering, metrics are often used to track the progress over time. Concerning the potential bias of using a single metric, performance engineers tend to use multiple metrics for reasoning. However, this approach has its own challenges. In this work we study one of the challenges in the context of analyzing trends in server energy proportionality. We examine a wide range of metrics for measuring energy proportionality, trying to determine which metrics are essential and which are redundant. We do this by comparing the trend curves of the metrics for the published results of the SPECpower_ssj2008 benchmark. While the context is specific, the proposed analysis method is quite general. We hope that this method would help us do performance engineering more effectively.

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

Metrics; Energy proportionality; SPECpower

1. INTRODUCTION

In performance engineering, metrics are often used to track the progress over time. Concerning the potential bias of using a single metric, performance engineers tend to use multiple metrics for reasoning. However, this approach has its own challenges. For example, what if two metrics give the opposite indications? How does one evaluate the added

value of each metric? In this work we study the latter problem in the context of analyzing the progress in server-level energy proportionality.

In an energy-proportional server, the power consumption of the server is proportional to its load. When the server is idle, the server would ideally consume no power. As the load increases, the server consumes gradually more power. In 2007, Barroso and Hölzle made a case why this is important [1]. Specifically, most servers in a data center stay at low utilization levels most of the time in order to accommodate occasional load spikes. Unfortunately, a traditional server has low energy efficiency at low utilization levels. An energy-proportional server, in contrast, has the same peak energy efficiency across all utilization levels. While this new server design does not address the issue of low utilization, it alleviates its consequence. As the annual energy cost for a server surpasses its purchase cost [2], more and more efforts have been made to improve server-level energy proportionality (e.g., [11, 14]).

Seven years have passed, and there have been some studies analyzing trends in server energy proportionality (e.g., [6, 7, 13, 14]). Many of these studies use the SPECpower_ssj2008 benchmark results published by the Standard Performance Evaluation Corporation (SPEC). However, they use different metrics because there is no consensus on how to measure energy proportionality. In this work we examine these metrics as well as several other new metrics, trying to determine which metrics are similar in their evaluation of server energy proportionality. To the best of our knowledge, this has never been done before.

In this work metrics can be similar empirically or analytically. We determine whether two metrics are *empirically* similar by calculating the correlation between two time series, each of which is derived from one metric applied to the common historical data archive. The higher the correlation, the more similar the metrics empirically. We say that two metrics are *analytically* similar if we can establish a mathematical equation between them. Empirical similarity gives us a hint on the possible analytical similarity. A new metric has no added value if it is similar to an existing metric.

Taking the aforementioned approach with the data being all 459 SPECpower_ssj2008 benchmark results published by SPEC until June 2014, we have found that the metric EP [7] can be considered as a good metric for measuring server energy proportionality. We have also established a mathematical relationship between EP and the overall score of the SPECpower_ssj2008 benchmark set forth by SPEC. The relationship allows us to understand how much improve-

(c) 2014 Association for Computing Machinery. ACM acknowledges that this contribution was authored or co-authored by an employee, contractor or affiliate of the United States government. As such, the United States Government retains a nonexclusive, royalty-free right to publish or reproduce this article, or to allow others to do so, for Government purposes only.

ICPE'15, Jan. 31–Feb. 4, 2015, Austin, Texas, USA.

Copyright © 2015 ACM 978-1-4503-3248-4/15/01 ...\$15.00.

<http://dx.doi.org/10.1145/2668930.2688049>.

ment in energy efficiency comes from improvement in energy proportionality. The paper also argues for the importance of another aspect of energy proportionality, linearity. This aspect is often overlooked but it impacts the potential for achieving the greatest energy proportionality at cluster level. Since none of existing linearity metrics provides this kind of insight, the paper calls for the development of a good linearity metric in the context of energy proportionality.

The rest of the paper is organized as follows. Section 2 provides information on how trends in server energy proportionality were studied in the past. The method for analyzing the empirical similarity of two metrics is detailed in Section 3. Section 4 shows how the method can be used to reach the conclusion that EP is a good metric. Sections 5 highlights the many ways to measure the linearity aspect of server energy proportionality, and Section 6 discusses why they all fail to provide important insight relating to cluster-level energy management. Section 7 concludes the paper with some future research directions.

2. RELATED WORK

This section describes how trends in server energy proportionality were studied in the past. Two key elements of these studies are the *data* and *methods* used for analysis. We discuss each element in details.

2.1 The SPECpower_ssj2008 Benchmark

To study trends in server energy proportionality, many past studies [6, 7, 13, 14] use the SPECpower_ssj2008 benchmark results published by the Standard Performance Evaluation Corporation (SPEC). This benchmark takes the approach of graduated workload [3], thus allowing energy proportionality to be measured. Specifically, the benchmark measures the performance and power of eleven load levels from zero to 100% of a given server’s full capacity to process business transactions with a server side Java application.

The results published by SPEC are arguably the best data available for analyzing trends in server energy proportionality. Many other power benchmarks [5] have very limited data, making it impossible to do any meaningful trend analysis. For example, GBench implements graduated workload but does not have much data [10]. The Green500 [4] has more data, but its workload does not have any sense of graduated workload, making it impossible to measure energy proportionality. In addition, there are serious concerns about its measurement bias [8]. In contrast, the SPEC-published results may have less measurement bias. In order for a result to be published by SPEC, a set of rigorous run rules must be complied with, including the proper setup for taking the measurements.

SPEC maintains an archive ¹ of all the results it has published since December 2007. The archive currently contains over 450 results from major server vendors such as Acer, Dell, Fujitsu, HP, Huawei, and IBM. As the previous studies, our study also uses these results for trend analysis.

2.2 Measuring Energy Proportionality

Any trend analysis of energy proportionality requires a measure of energy proportionality. Unfortunately there is no consensus on what metric should be used. As a result,

¹http://www.spec.org/power_ssj2008/results/

the previous studies differ mainly in the metric each uses for quantifying energy proportionality.

Historically, server energy proportionality has been measured by the difference in power between the 0% load level and the 100% load level [1]. In this paper we call this metric DR (Dynamic Range) [14]:

$$DR = \frac{P(1) - P(0)}{P(1)} \quad (1)$$

where $P(\ell)$ represents the power consumption (in watts) of a server at load level ℓ , $0 \leq \ell \leq 1$. DR is between 0 and 1, with 1 being fully energy-proportional.

Another commonly used metric is called EP (Energy Proportionality) [7]. Like DR, EP is between 0 and 1, with 1 being fully energy-proportional. In contrast to DR, EP accounts for the power usage in intermediate load levels:

$$EP = 2 - \frac{\int_0^1 P(\ell) d\ell}{\int_0^1 P_E(\ell) d\ell} \quad (2)$$

where

$$P_E(\ell) = P(1) \cdot \ell \quad (3)$$

represents the power-consumption behavior of the fully energy-proportional server, i.e., the EP (also DR) of $P_E(\ell)$ is 1.

Some metrics measure energy disproportionality instead. For example, the metric IPR (Idle-to-peak Power Ratio) [13],

$$IPR = \frac{P(0)}{P(1)}, \quad (4)$$

is also between 0 and 1, but the IPR of $P_E(\ell)$ is 0. In fact, IPR and DR are duals, i.e., $IPR + DR = 1$.

Some metrics measure nonlinearity because the geometrical interpretation of $P_E(\ell)$ is a line. For example, the metric LD (Linear Deviation) [14] quantifies the deviation from a linear function $P_L(\ell)$:

$$LD = \frac{\int_0^1 P(\ell) d\ell}{\int_0^1 P_L(\ell) d\ell} - 1 \quad (5)$$

where

$$P_L(\ell) = P(0) + [P(1) - P(0)]\ell \quad (6)$$

represents a particular power-consumption behavior. A server is called superlinearly energy-proportional if $LD > 0$ and sublinearly energy-proportional if $LD < 0$. Geometrically, sublinearity occurs when the curve $P(\ell)$ lies under the line $P_L(\ell)$. The LD of $P_E(\ell)$ is 0, so is that of $P_L(\ell)$.

Another metric, LDR (Linear Deviation Ratio) [13], measures the deviation differently:

$$LDR = \max_{\ell} \left| \frac{P(\ell)}{P_L(\ell)} - 1 \right| \quad (7)$$

where \max is the maximum by absolute-value comparison so as to retain the sign of the maximum value. Like LD, LDR distinguishes between superlinearity and sublinearity. The LDR of $P_E(\ell)$ is 0, so is that of $P_L(\ell)$.

This paper will discuss the similarity of these metrics and whether new metrics are still needed.

3. THE ANALYSIS METHOD

How can one tell that the insights from two metrics are the same? This section will present a method that casts

the problem as computing the similarity of two time series called *trend curves* in this paper. This method is in contrast to visual analysis used in the previous studies.

Define a trend curve as a sequence of n elements sorted according to some temporal order. Each metric has a corresponding trend curve. In this work, the trend curve of a metric contains $n = 459$ elements, and each element holds the metric value for one SPECpower_ssj2008 benchmark result. For example, the EP value for a benchmark result can be approximated via the trapezoidal rule for integration as

$$EP \approx 2 - \sum_{j=1}^{m-1} (\ell_j - \ell_{j-1}) \left[\frac{P_{j-1} + P_j}{P_{m-1}} \right] \quad (8)$$

where the result contains $m = 11$ data points $\{(\tau_j, P_j)\}$, $0 \leq j \leq m-1$ and $\tau_0 < \tau_1 < \dots < \tau_{m-1}$, describing the performance and power behavior of a server. These data points are viewed as the samples of $P(\ell)$ where $P(\ell_j) = P_j$ for $\ell_j = \tau_j/\tau_{m-1}$. The value of a metric can then be calculated based on these samples.

The similarity of two trend curves $\vec{u} = (u_1, u_2, \dots, u_n)$ and $\vec{v} = (v_1, v_2, \dots, v_n)$ can be computed using Pearson's r :

$$r(\vec{u}, \vec{v}) = \frac{\sum_i (u_i - \bar{u})(v_i - \bar{v})}{\sqrt{\sum_i (u_i - \bar{u})^2} \sqrt{\sum_i (v_i - \bar{v})^2}} \quad (9)$$

where

$$\bar{u} = \frac{1}{n} \sum_i u_i \quad \text{and} \quad \bar{v} = \frac{1}{n} \sum_i v_i. \quad (10)$$

Pearson's r measures the strength of the linear relationship between variables \vec{u} and \vec{v} based on the data $\{(u_i, v_i)\}$. It ranges from -1 to 1, with -1 (or 1) being a perfect negative (or positive) linear relationship and 0 no linear relationship. One nice property of this measure is that the two trend curves do not have to be on the same scale, thereby allowing us to study a larger set of metrics.

According to our method, the trend curves of DR and EP have a similarity (or linear correlation) of 0.91. In other words, the two curves are similar but not identical. This justifies the existence of both metrics.

3.1 Verification

Now consider visual check. Figure 1 shows the two trend curves. We see that server energy proportionality has stagnated since 2009 according to DR. In contrast, there was a uplift in 2012 according to EP. In other words, EP captures the uplift that DR does not. On the other hand, DR provides an insight that the *early* improvement in EP can be correlated to the widening dynamic range of the power. Therefore, both metrics are valuable, and EP can be considered better for the current time. This exercise also shows that the proposed method enables us to determine whether two metrics provide different insights or not. The uplift in 2012 has been attributed to the result of using dynamic processor over-clocking (e.g., Intel Turbo Boost) [6].

In fact, we can establish a mathematical relationship between EP and DR as follows:

$$EP \approx 2 - (2 - DR)(LD + 1) \quad (11)$$

where LD is a nonlinearity measure. This equation indicates two ways to increase EP: to increase DR or to decrease LD. Since the trend in DR has stagnated since 2009, much of the *later* improvement in EP comes from the decrease in

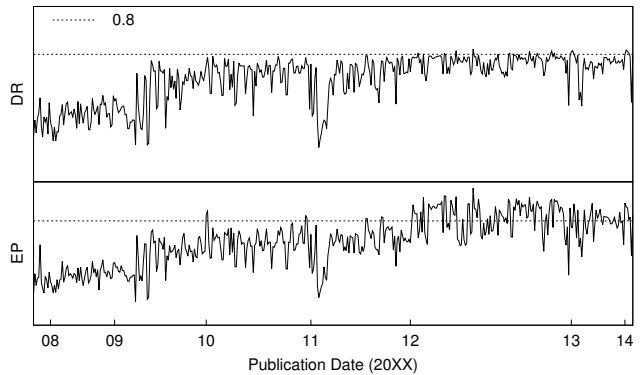


Figure 1: The trend curves of DR and EP.

LD. Wong and Annavaram noted that DR and LD can affect EP [14], but here we establish the math form for that observation. This exercise also shows how empirical similarity can give us a hint on the possible analytical similarity.

4. METRICS BASED ON DEVIATION

In this section and next, we examine a series of metrics for measuring server energy disproportionality. The metrics discussed in this section all measure the deviation from the behavior described by $P_E(\ell) = P(1) \cdot \ell$.

Given a set of m samples $\{(\ell_j, P_j)\}$, the following are a few possible metrics for measuring the deviation:

$$L_\infty = \max_j |\varepsilon_j|, \quad (12)$$

$$L_1 = \sum_j |\varepsilon_j|, \quad (13)$$

and

$$L_2 = \sqrt{\sum_j |\varepsilon_j|^2}, \quad (14)$$

where

$$\varepsilon_j = \frac{P_j - P_E(\ell_j)}{P(1)}. \quad (15)$$

These are common deviation metrics, but we are not aware of any previous work that uses one of these metrics to evaluate energy disproportionality.

Based on our method, these metrics are very similar to each other. The curves of L_∞ and L_1 have the least similarity which is 0.98. In fact, these metrics can be viewed as the dual of EP. Mathematically, we can establish the following relationship between EP and L_1 :

$$EP \approx 1 - \frac{2L_1}{m-1}. \quad (16)$$

In summary, EP is a good energy-proportionality metric because metrics L_∞ , L_1 , and L_2 do not add any new insight.

4.1 The Impact to Energy Efficiency

Ryckbosch et al. [7] compared EP with the overall score (SCR) of the SPECpower_ssj2008 benchmark. While SCR is a metric for measuring server energy *efficiency*, its calculation involves the power numbers at multiple load levels

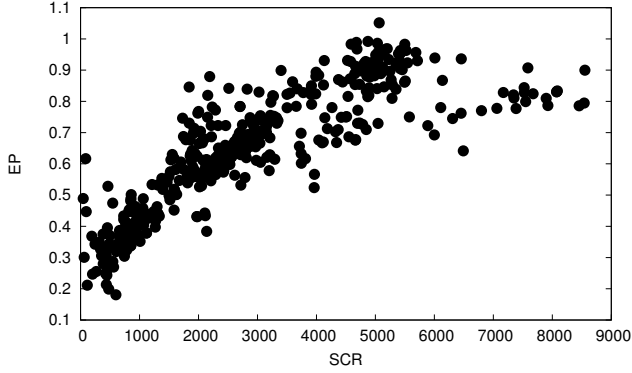


Figure 2: The imperfect correlation between metrics SCR and EP.

and thus includes some notion of energy proportionality:

$$\text{SCR} = \frac{\sum_j \tau_j}{\sum_j P_j}. \quad (17)$$

Through visual analysis, Ryckbosch et al. noted that SCR correlates well with EP, but not perfect. A server with a high EP does not necessarily have a high SCR, and vice versa. Figure 2 shows this imperfect correlation. Using our method, the correlation can be quantified as 0.83.

In fact, we can do more. Through some mathematical manipulation, we can establish the following relationship:

$$\text{SCR} \approx \left(\frac{\tau_{10}}{P_{10}} \right) \left(\frac{1.1}{2.1 - \text{EP}} \right). \quad (18)$$

This equation indicates two ways to increase SCR: to improve the performance-power ratio at the 100% load level or to improve EP. Since the improvement in EP has stagnated since 2012, much of the recent improvement in SCR has come from the increasing performance-power ratio at the 100% load level.

5. METRICS BASED ON LINEARITY

In this section we examine metrics for measuring nonlinearity. We start by comparing metrics LD and LDR.

Based on our method, the similarity of the two metrics is only 0.85. So both metrics are valuable. We are unable to establish a mathematical equation connecting them. However, we have found that their dissimilarity is largely due to the different *design* bias. In the following we will discuss where they are similar and dissimilar.

On the similarity side, both metrics indicate the same trend. As shown in Figure 3, the metric values were mostly positive in 2007–2011. Beginning 2012, more negative values have appeared. This timing matches what we have discussed earlier about the EP’s uplift in 2012. Clearly, these curves provide additional evidence on why the uplift occurs. They also supplement Equation (11) by showing that LD not only decreases but also falls below zero. In other words, modern servers show more sublinear energy proportionality.

On the dissimilarity side, the two metrics treat the deviation at the same load level unequally. LDR weighs more on deviations at low load levels. In contrast, LD treats deviations at all load levels equally. Specifically, both metrics

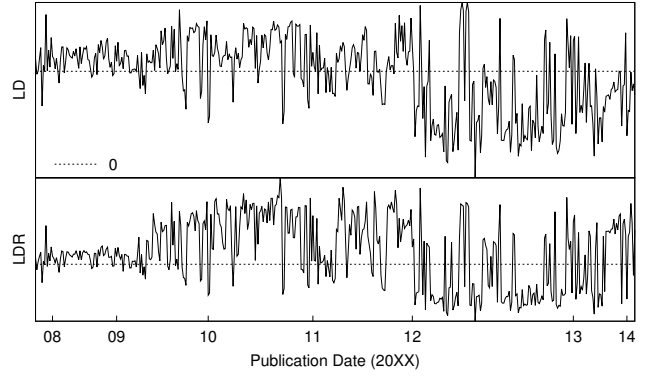


Figure 3: The trend curves of LD and LDR.

can be rewritten as

$$\text{LD} \approx \frac{2}{(m-1)(\text{IPR}+1)} \sum_j \frac{P_j - P_L(\ell_j)}{P(1)}. \quad (19)$$

and

$$\text{LDR} \approx \max_j \frac{1}{\text{IPR} + (1 - \text{IPR})\ell_j} \cdot \frac{P_j - P_L(\ell_j)}{P(1)}. \quad (20)$$

While both metrics calculate the deviation from the same linear function $P_L(\ell)$ as

$$\varepsilon_j = \frac{P_j - P_L(\ell_j)}{P(1)}, \quad (21)$$

LDR weighs more on deviations at low load levels. Due to their different weight assignments, the trend curves of LD and LDR become somewhat dissimilar.

In terms of insight, Varsamopoulos et al. [13] noted two trends in LDR: one positive trend and one negative trend. However, the positive trend in LD is not as obvious in Figure 3. In other words, LDR shows the continuing difficulty in achieving sublinear energy proportionality at low load levels. LD shows that either sublinearity occurs at more load levels or sublinearity becomes higher or both.

Probably a better way to gain the above kind of insight is to look into the *range* of the deviations at various load levels. Using Equation (21) as the deviation measure, Figure 4 shows the maximum and minimum of these deviations. We see that both extremes tend to have the same sign before 2013. They are either all positive or all negative. Starting 2013, the maximum is positive and the minimum is negative. In other words, curve $P(\ell)$ and line $P_L(\ell)$ intersect.

This insight has led us to study the impact of the shape of $P(\ell)$. The next section will discuss how it plays an important role in achieving great *cluster-level* energy proportionality.

5.1 Other Possible Metrics

The rest of this section shows that there are many other ways to measure linearity or nonlinearity. Table 1 lists a sample of them in three different types. The first two types are related to curve fitting. For the first type, we see that many functional forms can be used to describe the curve, and fitness or function parameter can be used to measure linearity. For the second type, linearity is measured in terms of complexity. A line is expected to be modeled by a simple function with a good fitness (or numerical accuracy). The more complex the function, the more nonlinear the curve.

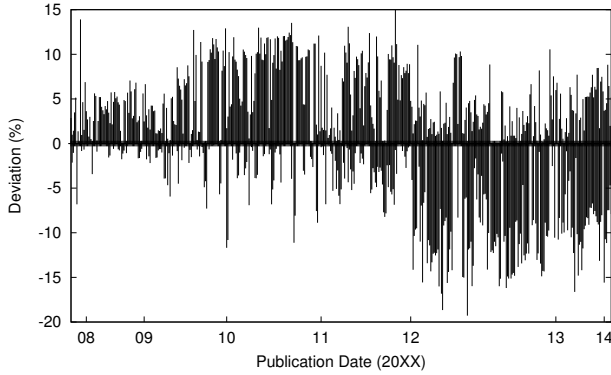


Figure 4: The range of the linearity deviations based on Equation (21).

The third type comes from image processing. In image processing applications, a line is often needed to be recognized. There are also many ways to recognize the line pattern. AO is considered to be one of the most effective methods for this task [9]. AO assumes the curve as a piecewise-linear function which passes through all data points $\{(\ell_j, P_j)\}$. Since there are 11 points, there are 10 line segments in this curve. AO uses the average of all the normal unit vectors of these line segments to measure linearity.

According to our method, the trend curves of these metrics are different from each other, except that the curves of RMSE and MAPE have a similarity of 0.99. We have tried other correlation measures such as Spearman’s ρ and Kendall’s τ , but none helped. Smoothing the curves before computing their similarity helped but not much. We were only able to derive a high similarity of POW and COE using the median trend-lines of their curves.

More importantly, we cannot associate any of these metrics with implications on how to improve energy proportionality. As we will discuss in the next section, a metric that characterizes the shape of $P(\ell)$ may provide such insight.

6. IS LINEARITY IMPORTANT?

This section tries to justify the importance of a good shape of $P(\ell)$ that the existing metrics may have overlooked. We start with the shape of $P_E(\ell) = P(1) \cdot \ell$.

Consider a cluster of homogeneous servers. Curve $P_E(\ell)$ has two properties that affect cluster-level energy management. The property of $P(0) = 0$ affects server provisioning. Servers can be added at any time without power penalty. The property of linearity affects load distribution. Load can be distributed in any way without power penalty. In the case of two servers, this means there are many ways to serve the input load level of 0.8 with the same amount of aggregate power usage, whether it be one server at the 0.8 load level, two servers at the 0.4 load level each, or two servers at the 0.2 and 0.6 load levels respectively.

In reality, the first property is not met by any real server. In other words, any server that is on will introduce some power penalty. To achieve better energy proportionality, cluster management can use the Packing scheme which provisions a minimum number of servers to handle the input load level. In the case of two servers, provisioning one server at the 0.8 load level consumes the least power.

For a cluster of s servers, the following formula calculates the cluster-level energy proportionality \widehat{EP} if managed by the Packing scheme:

$$\widehat{EP} = 1 - \frac{1 - EP}{s}. \quad (22)$$

Assume each server is very energy disproportional, say $EP = 0.1$. A cluster of 100 such servers with the Packing scheme can easily achieve the cluster-level energy proportionality of 0.9, and this applies to all shapes of $P(\ell)$.

The above formula is only correct in the ideal scenario in which servers can be provisioned instantaneously. According to a simulation-based study done by Wong and Annavaram [15], \widehat{EP} saturates around 0.8 in more realistic scenarios. They suggested to forego dynamic server provisioning and distribute the input load evenly among all servers. Using this Uniform scheme, \widehat{EP} is equivalent to EP . Since some of modern servers have their EP greater than 0.8 (see Figure 1), the Uniform scheme effectively improves cluster-level energy proportionality.

The Uniform scheme is sometimes optimal. In the case of two servers, the scheme will provision two servers at the 0.4 load level each. If the server behavior can be described as $P(\ell) = 0.2 + 0.8 \cdot \ell^2$, then the total power usage based on this scheme is the lowest. In fact, the Uniform scheme remains optimal if $P(\ell)$ depicts a *convex* curve. Once the curve is not convex, the optimality will not be guaranteed. This is where the shape of $P(\ell)$ can impact how to achieve the best cluster-level energy proportionality.

As we continue to create solutions, such as [12, 15], to further improve server energy proportionality, we may need to consider its impact on the complexity of achieving near-optimal cluster-level energy management. If linearity is desired, we should increase its weight when creating these new solutions. If convexity is desired, we need metrics that can capture this. Metric LD falls short because $LD < 0$ does not imply that $P(\ell)$ is convex. If curve $P(\ell)$ and line $P_L(\ell)$ intersect, $P(\ell)$ is definitely not convex.

In short, we are lacking metrics that can provide insight on the good shapes of $P(\ell)$ for enabling simple yet effective cluster-level energy management.

7. CONCLUSIONS AND FUTURE WORK

This paper starts with the question of how to evaluate the added value of each metric when multiple metrics are used during performance engineering. This is important because a high degree of redundancy in results hinders the efficiency in gaining insight. This paper presents a general method to address the question. The method is demonstrated in the context of analyzing trends in server energy proportionality. It enables the identification of a good energy-proportionality metric from which more focused studies, such as mathematical relationships, can be conducted. While the identified metric is considered better than other existing metrics, this paper argues that the metric fails to capture an important aspect of server energy proportionality, linearity. The paper outlines why this aspect is also equally important and concludes that none of the existing metrics has captured it in a satisfactory manner.

Obviously, the lack of a good linearity metric is one of several directions for the future work. In addition, the argument provided in the paper is rather analytical. Simulation-based studies can be conducted to further strengthen the

Table 1: The many ways to measure linearity. SEG uses piecewise linear functions of different line segments. COD and AO measure linearity, and all the other metrics measure nonlinearity.

Name	Definition	Notes	Range	
Type 1: Fitting by the least squares method				
COD	$[r(\vec{\ell}, \vec{P})]^2$	$P(\ell) = c_0 + c_1 \cdot \ell$	[0, 1]	
RMSE	$\sqrt{\frac{1}{m} \sum_j \left[\frac{P(\ell_j) - P_j}{P_{m-1}} \right]^2}$		[0, 1]	
MAPE	$\frac{1}{m} \sum_j \frac{ P(\ell_j) - P_j }{P_j}$		$[0, \infty]$	
POW	α		$P(\ell) = c_0 + c_1 \cdot \ell^\alpha$	$[0, \infty]$
COE	c_2		$P(\ell) = c_0 + c_1 \cdot \ell + c_2 \cdot \ell^2$	$[0, \infty]$
Type 2: Finding the smallest k such that RMSE is below 0.01				
DEG	k	$P(\ell) = \sum_{i=0}^k c_i * \ell^i$	$[0, m - 1]$	
SEG	k	$P(\ell) = c_0 + c_1 * x + \sum_{i=2}^k c_i * x - \beta_i $	$[1, m]$	
Type 3: Recognizing pattern in planar point sets				
AO	$\frac{1}{m-1} \sqrt{(\sum_j \frac{-s_j}{\sqrt{s_j^2+1}})^2 + (\sum_j \frac{1}{\sqrt{s_j^2+1}})^2}$	$s_j = \frac{P_{j+1} - P_j}{\ell_{j+1} - \ell_j}, 0 \leq j < m$	$[2/\pi, 1]$	

argument. If linearity is indeed very critical, new server designs will need to be judged based on this aspect as well. In terms of trend analysis, one idea is to approximate $\{(\ell_j, P_j)\}$ with a curve, measure the energy proportionality of the curve, and see if it makes any difference.

8. ACKNOWLEDGMENTS

This work was supported by the Extreme Scale Systems Center at Oak Ridge National Laboratory and funded by the United States Department of Defense. This manuscript has been authored by UT-Battelle, LLC, under Contract No. DE-AC0500OR22725 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for the United States Government purposes.

9. REFERENCES

- [1] L. Barroso and U. Hözl. The case for energy-proportional computing. *IEEE Computer*, 40(12):33–37, Dec. 2007.
- [2] C. Belady. In the data center, power and cooling costs more than the IT equipment it supports. *Electronics Cooling Magazine*, Feb. 2007.
- [3] L. Gray, A. Kumar, and H. Li. Workload characterization of the SPECpower_ssj2008 benchmark. In *SPEC International Performance Evaluation Workshop*, June 2008.
- [4] The Green500 list: Environmentally responsible supercomputing. <http://www.green500.org>.
- [5] C. Hsu, J. Kuehn, and S. Poole. Towards efficient supercomputing: Searching for the right efficiency metric. In *International Conference on Performance Engineering*, Apr. 2012.
- [6] C. Hsu and S. Poole. Revisiting server energy proportionality. In *International Workshop on Power-Aware Systems and Architectures*, Oct. 2013.
- [7] F. Ryckbosch, S. Polfiet, and L. Eeckhout. Trends in server energy proportionality. *IEEE Computer*, 44(9):69–72, Sept. 2011.
- [8] T. Scogland, C. Steffen, T. Wilde, F. Parent, S. Coghlan, N. Bates, W. Feng, and E. Strohmaier. A power-measurement methodology for large-scale, high-performance computing. In *International Conference on Performance Engineering*, Mar. 2014.
- [9] M. Stojmenović, A. Nayak, and J. Zunic. Measuring linearity of planar point sets. *Pattern Recognition*, 41(8):2503–2511, Aug. 2008.
- [10] B. Subramaniam and W. Feng. GBench: Benchmarking methodology for evaluating the energy efficiency of supercomputers. In *International Supercomputing Conference*, June 2012.
- [11] B. Subramaniam and W. Feng. Towards energy-proportional computing for enterprise-class server workloads. In *International Conference on Performance Engineering*, Apr. 2013.
- [12] B. Subramaniam and W. Feng. Enabling efficient power provisioning for enterprise applications. In *International Symposium on Cluster, Cloud and Grid Computing*, May 2014.
- [13] G. Varsamopoulos, Z. Abbasi, and S. Gupta. Trends and effects of energy proportionality on server provisioning in data centers. In *International Conference on High Performance Computing*, Dec. 2010.
- [14] D. Wong and M. Annavaram. Scaling the energy proportionality wall with KnightShift. *IEEE Micro*, 33(3):28–37, May/June 2013.
- [15] D. Wong and M. Annavaram. Implications of high energy proportional servers on cluster-wide energy proportionality. In *International Symposium on High-Performance Computer Architecture*, Feb. 2014.