Selecting an Industry-Standard Metric for Data Center Efficiency

By Neil Rasmussen

White Paper #157

APC

Legendary Reliability®

Executive Summary

The development of standards for the measurement and specification of data center efficiency is an essential step in the global effort to reduce the environmental impact of data centers. This paper explains some of the metrics that have been used to describe the efficiency of data center physical infrastructure and suggests which metric is the most effective.

Introduction

In today's environment, it is good public policy and good business to consider the options to control data center energy consumption. A 1MW high-availability data center can consume $20,000,000 worth of electricity over its lifetime. Recent articles suggest that for some customers the cost of electricity is greater than the cost of IT hardware. Many companies are beginning to consider the carbon consumption of their ongoing operations and realizing that data centers are significant contributors to the environmental burden of business and industry. Research by APC-MGE and others is showing that efficiency varies widely across similar data centers, and – even more significant – the actual efficiencies of real installations are well below the practical achievable best-in-class values.

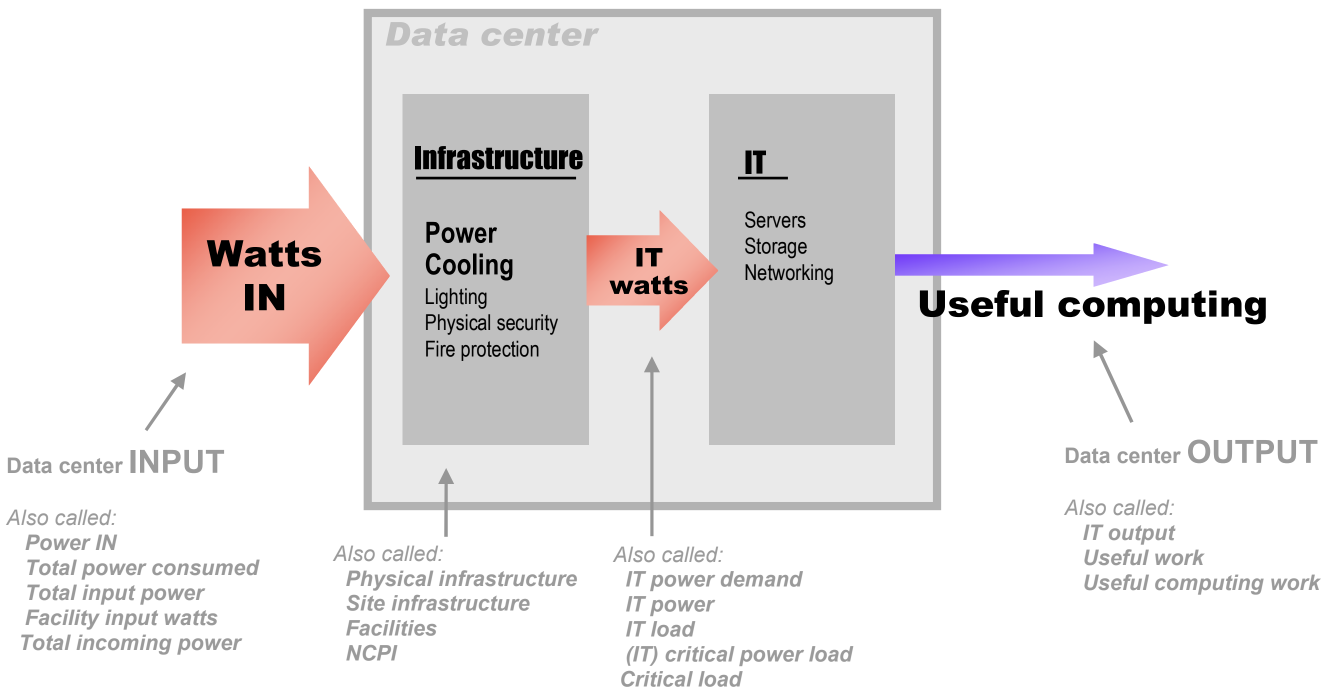
The result is a strong desire among users, manufacturers, and public policy makers to incent and realize improved electrical efficiency in data centers. In this paper, we will focus on the power consumption and efficiency of data center power, cooling, and other miscellaneous infrastructure, which consumes 50% or more of the electrical power in a typical installation. The power consumption and efficiency of the IT equipment *supported* by the infrastructure, while critically important to overall energy use, is not the subject of this paper.

Efficiency metrics are needed to allow users to compare alternative scenarios, to allow specific data centers to be benchmarked against similar facilities, and to measure progress against internal or external targets.

Holistic View of Data Center Efficiency

In a whole-system view, the data center takes in *watts* and performs *useful computing* (**Figure 1**). Within the data center, there are two major subsystems: (1) the data center's ***infrastructure***, which provides support and protection for (2) the ***IT*** equipment, which performs the real "work" of the data center.

***Figure 1** – Basic elements of data center efficiency analysis*



Using the model in the above figure, overall data center efficiency can be represented by the following simple equation:

$$\textbf{Overall data center efficiency} = \frac{\textbf{Useful computing}}{\textbf{Watts IN}}$$

This basic view of data center efficiency, with the two-part internal breakdown shown in **Figure 1**, forms the foundation of the data center efficiency research and analysis currently being done by the US EPA and the EC Directorate-General.

The two components of data center efficiency: Infrastructure efficiency and IT efficiency

Overall data center efficiency can be logically and mathematically separated into *infrastructure efficiency* and *IT efficiency*.

These two components of overall efficiency are referenced in the 2007 EPA Report to Congress, excerpted below.

From the EPA report: [1]

[1] U.S. Environmental Protection Agency ENERGY STAR Program, *Report to Congress on Server and Data Center Energy Efficiency, Public Law 109-431*, August 2007, page 94

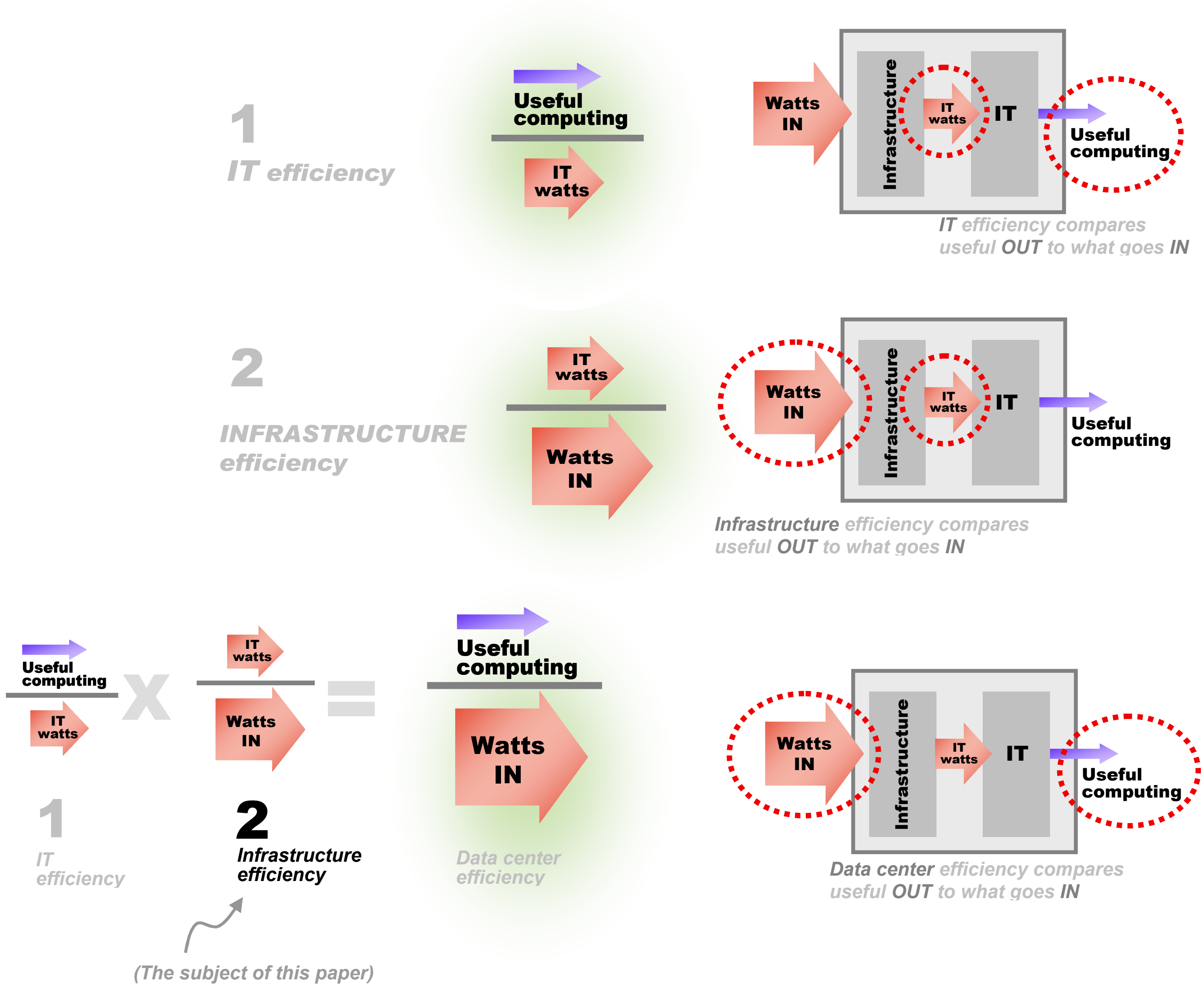
"Brill (2007) defines the aggregate energy productivity of a data center as the product of two terms:

1) The ratio of IT output from the data center (like web pages served or calculations completed over some period) divided by the average watts of power demanded by the computing, storage, and networking equipment within the data center [termed the information technology (or IT) critical power load].

2) The ratio of IT critical power load to the total incoming power consumed by the datacenter.

Multiplying these terms together yields a ratio of total IT output (e.g., web pages served) by total input power to the data center facility."

Figure 2 *– EPA breakdown of data center efficiency*
(Refer to ***Figure 1****)*



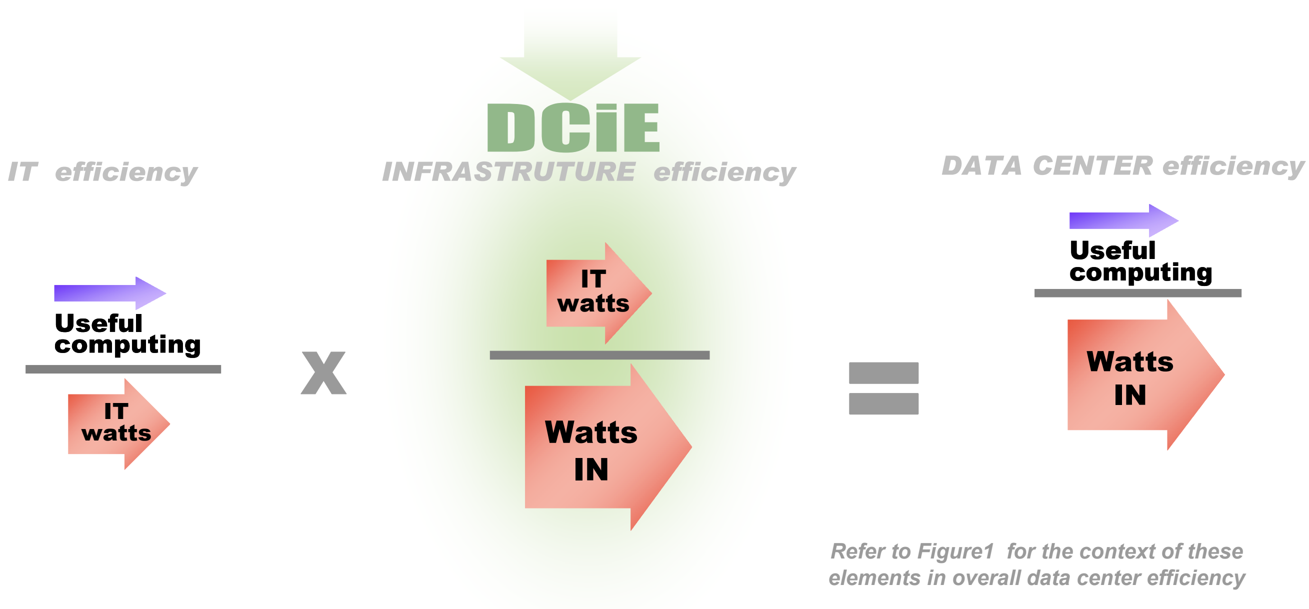
EPA criteria for an infrastructure efficiency metric

The EPA report further clarifies requirements for the metric to be used to specify the ***infrastructure*** efficiency component (called ***site infrastructure*** in the EPA report – efficiency component #2 above):

> **"A meaningful measure of site infrastructure efficiency needs to meet the following criteria: 1) clearly convey its meaning, 2) correctly define an efficiency metric in which the output metric is in the numerator and the energy input in the denominator, and 3) apply only to the site infrastructure portion of the data center."**

To review, the EPA report defines data center efficiency as the combined (multiplied) efficiency of the data center's two major components – *infrastructure* and *IT* (**Figure 3**). The second term in the multiplication is *infrastructure* efficiency. This has begun to appear in the industry literature as **data center infrastructure efficiency**, or **DCiE** – a metric that aligns precisely with the EPA definition quoted above. The defense of DCiE as the correct metric for infrastructure efficiency is the subject of this paper.

***Figure 3 – DCiE** meets EPA criteria for an **infrastructure** efficiency metric*



DCiE meets the requirements of the EPA report, plus has the convenient benefit of having the same units in the numerator and denominator (watts), so the resulting DCiE ratio is dimensionless and can be expressed as an ordinary percent from zero to 100%.

Data center infrastructure efficiency

$$\text{DCiE} = \frac{\text{Watts delivered to IT load}}{\text{Watts IN to data center}}$$

Ratio expressed as percent %

Alternative Metrics For "Infrastructure Efficiency"

There are a number of different metrics currently in use to represent the EPA concept of "infrastructure efficiency," which can be confusing to both the data center community and the public. These different metrics are summarized in **Table 1**.

***Table 1** – Various metrics for **infrastructure efficiency** currently found in the literature*

Metric	Meaning	Value Range	Status
DCiE **Data Center Infrastructure Efficiency**	**Ratio of IT load power to facility input power**	**0-100%**	Leading term compliant with EPA and EC specifications
DCE **Data Center Efficiency**	**Same as DCiE**	**0-100%**	Being replaced by DCiE because it was not clear whether it included the IT part of the data center
PUE **Power Utilization Effectiveness**	**Ratio of facility input power to IT load power (1/DCiE)**	**Infinity - 1**	Non-compliant with EPA specification; non-intuitive because efficiency goes **up** when value goes **down**
SI-EER **Site Infrastructure Energy Efficiency Ratio**	**Same as PUE**	**Infinity - 1**	An alternative term for PUE that is not widely used
SI-POM **Site Infrastructure Power Overhead Multiplier**	**Same as PUE**	**Infinity - 1**	A recent term that has been proposed to replace PUE because the language is more intuitive – i.e., **lower overhead** is **better**

Table 1 shows that there are really only two metrics, **DCiE and PUE** – the others are simply different terminology representing one of these two metrics.

Various public reports, such as the EPA Report to Congress, describe the confusion over metrics and the need to develop a consensus around a single metric, and a single term to describe it. We believe that under close examination, the DCiE emerges as the best and most effective metric for describing data center infrastructure efficiency. While the PUE metric conveys the same information mathematically, it is more difficult to understand and communicate effectively.

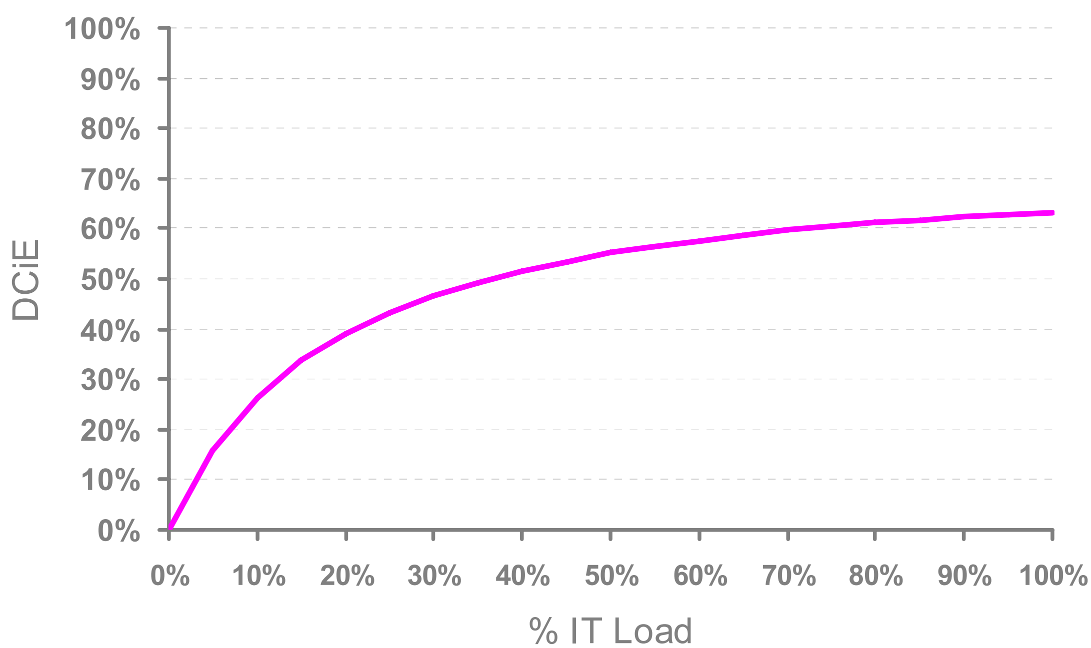
Comparing the DCiE and PUE Metrics

There are a variety of ways to compare and evaluate these two metrics in order to discover good reasons to select one over the other.

Displaying efficiency data in graphical form

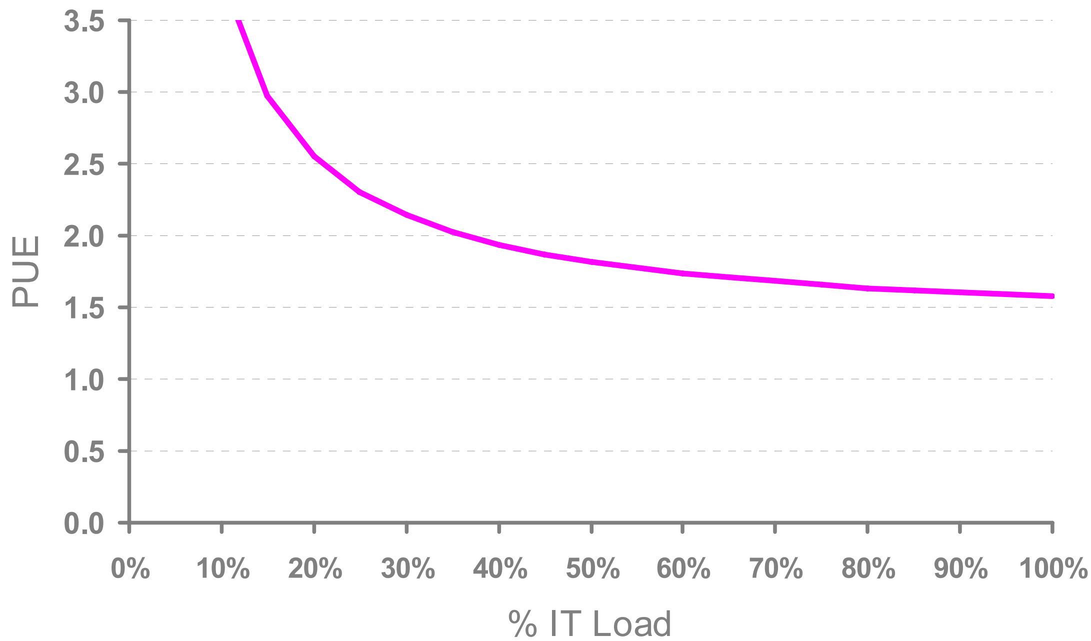
Every data center has a characteristic efficiency curve as a function of IT load, which is an essential description of the capabilities of the data center. In the future, this curve will be provided as part of a data center design, as it is today for various components such as UPS. A typical data center efficiency curve is shown in **Figure 4** using the DCiE metric.

Figure 4 *– Typical infrastructure efficiency curve, using DCiE as the metric*



The same data center can have its efficiency represented by the PUE metric, as shown in **Figure 5**.

***Figure 5** – Typical infrastructure efficiency curve, using PUE as the metric*



When these curves are compared, it is immediately apparent that part of the data on the PUE curve, below 12% IT load, is truncated. This is because PUE always approaches infinity as the IT load approaches zero. If more of the curve is shown for light load by expanding the vertical scale, then the PUE scale becomes compressed and it is difficult to read PUE values of the curve at higher IT loads. By contrast, the DCiE curve is always constrained to the range 0-100% so all plots are standardized and no data is missing.

This same problem affects instrumentation displays. If a display has a range limit then PUE data may be truncated. By contrast, a 0-100% display always can show any DCiE value.

From this comparison, it is clear that the PUE is less effective for conveying graphical efficiency data for data center designs.

Public communication

Virtually every written publication about data center power consumption uses the term “efficiency.” Efficiency has a plain meaning to most people, which is that *increased* efficiency is desirable. If PUE is used as an “efficiency” metric then the idea that the number must go **down** to increase efficiency is contrary to expectation. In discussions of PUE it is common to hear comments such as “I want to increase my PUE,” not understanding or remembering that they actually want to *decrease* PUE.

It is very likely that people will continue to discuss data center efficiency and the need to increase it. The efficiency metric should reflect this intuitive "higher is better" concept, as DCiE does, and not reflect the opposite, as PUE does. Based on our interviews with data center users and members of the press, it is fair to say that DCiE is a more intuitive metric than PUE.

Instrumentation

Computing either DCiE or PUE requires mathematical calculations from instrumentation readings. Neither value can be read directly from basic instruments. There is no reason to prefer one over the other to simplify instrumentation. For purposes of display, however, PUE has the disadvantage of going to infinity, while DCiE can always be displayed on a 0-100% scale.

DCiE vs. PUE in analysis

Efficiency analysis requires breaking down the contributions to inefficiency and assigning them to different subsystems within the data center. For this type of calculation, PUE has an interesting feature: the **PUE value minus one** is equal to the sum of the "per unit" losses of all the power, cooling, and lighting devices. The "per-unit" calculation is a common methodology among power system engineers, which accounts for the early emergence of PUE as an efficiency metric.

DCiE has the disadvantage that one additional calculation – 1/ X, or taking its inverse – is needed to convert to additive "per-unit" values. However, in the future this calculation will be performed automatically by data center efficiency monitoring instrumentation. Therefore, this minor computational distinction should be a non-issue for data center users and public policy makers.

Comparing benchmarks

As efficiency becomes a pervasive – even dominating – concern in the industry, it will become common to compare efficiency benchmarks. Consider the following comparative descriptions of two data centers:

> Data center **A** has a DCiE of 50% and data center **B** has a DCiE of 60%
>
> Data enter **A** has a PUE of 2.0 and data center **B** has a PUE of 1.6

Carefully consider these two sentences and see which one most clearly communicates which has better efficiency, and by how much. Consider which one is most likely to be misinterpreted.

Summary comparison of DCiE and PUE

Table 2 summarizes the differences between DCiE and PUE as the metric for data center infrastructure efficiency.

***Table 2** – Comparison of DCiE and PUE metrics for data center infrastructure efficiency*

	DCiE	PUE
Worst value	0%	Infinity
Best value	100%	1
Direction of improvement	Higher is better	Lower is better
Relation to other metric	1 / PUE	1 / DCiE
Meets metric requirement established by EPA report?	Yes	No

Use of Efficiency Metrics

It is very important to understand that any efficiency metric is incomplete without data regarding the averaging period and the operating conditions of the data center. The efficiency performance of a data center varies with load and other conditions, in much the same way as the fuel economy of a car varies. To better understand how to measure efficiency, how to benchmark it, and how to interpret it, refer to APC White Paper #154, *Electrical Efficiency Measurement For Data Centers*.

Conclusion

Many metrics have been discussed in the industry to describe the efficiency of data center infrastructure. All the metrics in the literature convey the same basic information, but using a number of different acronyms and two possible mathematical representations – one the inverse of the other. With the discussion of data center efficiency now becoming widespread and intense, it is time for a thoughtful consideration of the alternatives in order to lead to industry standardization on a single metric.

This paper argues for ***data center infrastructure efficiency*** (DCiE) as the best metric of the ones currently appearing in efficiency discussions. Its name and concept are intuitive, it can be easily displayed on a standard scale, and it relates to our everyday understanding of what an efficiency metric should represent – higher is better, and 100% is perfect. It is consistent with the requirements for an effective metric as established in the 2007 US EPA Report to Congress.

The most commonly used alternative, PUE, has the mathematically obscure (and graphically difficult) scale of one to infinity, and the non-intuitive property that it is a measure of efficiency that *decreases* as efficiency *increases*. A measure of effectiveness that goes *down* when effectiveness goes *up* can easily lead to unnecessary confusion and misstated comparisons. Furthermore, the PUE metric does not meet the requirements set out in the EPA guidelines to Congress regarding data center efficiency metrics.

About the Author

Neil Rasmussen is the Chief Technical Officer of APC-MGE. He establishes the technology direction for the world's largest R&D budget devoted to power, cooling, and rack infrastructure for critical networks. Neil is currently leading the effort at APC-MGE to develop high-efficiency, modular, scalable data center infrastructure solutions and is the principal architect of the APC-MGE InfraStruXure system.

Prior to founding APC in 1981, Neil received his Bachelors and Masters degrees from MIT in electrical engineering where he did his thesis on the analysis of a 200MW power supply for a tokamak fusion reactor. From 1979 to 1981, he worked at MIT Lincoln Laboratories on flywheel energy storage systems and solar electric power systems.