

Increasing Data Center Efficiency with Server Power Measurements

Our analysis of computing efficiency has already delivered benefits, enabling us to add capacity at a data center location that we previously thought was power-constrained.

Executive Overview

Intel IT defined methods for analyzing computing energy efficiency within our design computing environment, using measurements of actual server power consumption and utilization. We used these methods to identify trends and opportunities for improving data center efficiency, and to implement a pilot project that increased data center computing capacity.

Most efforts to measure and increase data center efficiency have been focused at the facilities level. However, we can also obtain significant benefits by employing server-level measurements to analyze and increase computing efficiency. Improvements at this level can enable us to expand data center capacity, reduce capital expenditure, reduce operational power and cooling costs, and scale compute and facility power consumption in proportion to the actual work done.

We undertook several initiatives, including:

- Defining a computing energy-efficiency metric that reflects server performance/watt in design computing production use.
- Using this metric, together with measurements of server power consumption and utilization, to analyze

computing energy efficiency across all servers in an Intel data center containing four server generations.

- Identifying efficiency improvements associated with server refresh.
- Identifying opportunities to reduce power and cooling costs without negatively impacting Intel product design workload throughput.

Our analysis of computing efficiency has already delivered benefits, enabling us to add capacity at a data center location that we previously thought was power-constrained. We plan to use these methods at other data centers, and we are continuing to build on our work with proof of concept (PoC) projects that explore the energy efficiency opportunities we have identified.

Ravi A. Giri
Staff Engineer, Intel IT

Anand Vanchi
Solutions Architect, Intel Data Center Group

Contents

Executive Overview	1
Business Challenge	2
Compute Energy Efficiency	2
Measuring and Increasing Computing Energy Efficiency	3
Computing Energy Efficiency Metric	3
Analyzing Data Center Computing Energy Efficiency	5
Energy Efficiency Improvements and Opportunities	6
Conclusion and Next Steps	8
Additional Opportunities	8
Contributors	8
Acronyms	8

IT@INTEL

IT@Intel is a resource that enables IT professionals, managers, and executives to engage with peers in the Intel IT organization—and with thousands of other industry IT leaders—so you can gain insights into the tools, methods, strategies, and best practices that are proving most successful in addressing today's tough IT challenges. Visit us today at www.intel.com/IT or contact your local Intel representative if you'd like to learn more.

BUSINESS CHALLENGE

Microprocessor design requires enormous computing capacity, and the requirements increase significantly with each processor generation. Because of this, silicon design workloads are the primary driver of compute capacity growth at Intel. This in turn results in rapid growth in data center power and cooling needs.

Today, our design computing environment includes about 65,000 servers, with a very high compute utilization of 85 percent on average.

We refresh these servers based on a four-year cadence; this enables us to take advantage of the substantial increases in performance and energy efficiency that each new server generation delivers. By replacing aging servers on a regularly scheduled cadence, Intel has realized operational cost savings, avoided incremental data center capital spending, and gained capacity. Our studies have shown that we can achieve 10:1 consolidation ratios on average when replacing four-year-old servers based on single-core processors.¹

¹ See the IT@Intel white papers "Staying Committed To Server Refresh Reduces Cost" and "Realizing Data Center Savings with an Accelerated Server Refresh Strategy."

However, newer server technologies such as blade servers are providing significant growth in rack power density and cooling needs. With the shift to blades, the number of servers per rack has increased 1.6x from 40 to 64 over the past five years, while the rack power envelope has increased 3x to 24 kilowatts (kW). As a result:

- Power and cooling costs can be a very large component of overall server total cost of ownership (TCO).
- Power and cooling requirements constrain IT equipment capacity at some data centers. This can result in data center rack and floor space that cannot be utilized.

Compute Energy Efficiency

Rising power and cooling costs have resulted in widespread efforts to analyze and increase data center efficiency within the industry and IT organizations. These efforts have mostly been focused at the facilities level. However, as shown in Figure 1, a holistic approach to energy efficiency is required to gain maximum benefits.

Changes toward the middle and the top of the pyramid—in the server and CPU workloads—can have a dramatic impact in the data center efficiency, defined as the amount of "useful work" done, or performance delivered per watt consumed.

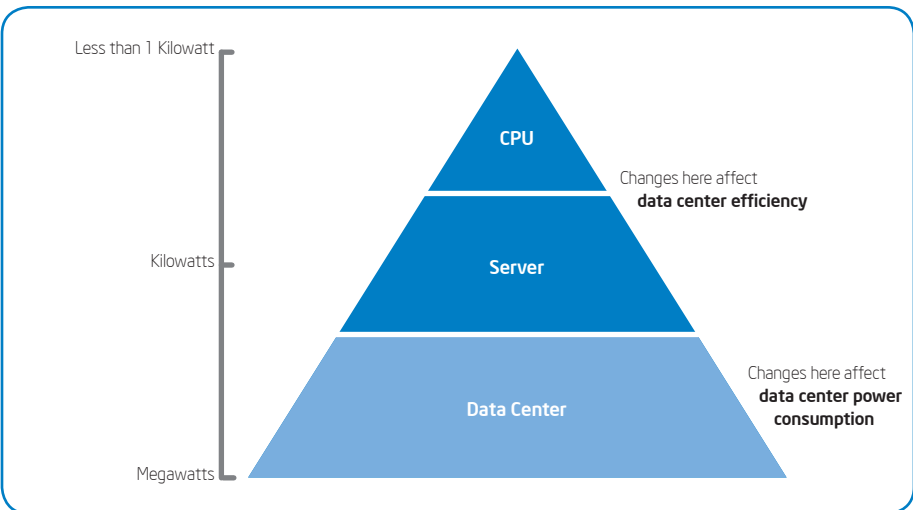


Figure 1. Data center energy pyramid. From "Musings about Data Center Energy Usage," Enrique Castro-Leon, Intel Corporation. <http://communities.intel.com>

These increases in computing energy efficiency, as indicated by performance/watt, can play a crucial role in countering the growth in power and cooling requirements. To achieve improvements in this area, we need to gain a better understanding of server power consumption and how it relates to server function and performance. Applied across many servers, improvements in compute efficiency can significantly reduce overall data center power consumption and Intel's carbon footprint. Potential benefits include increased data center capacity and reduced capital expenditure as well as reduced power and cooling costs with power-aware job scheduling.

INCREASED DATA CENTER CAPACITY AND REDUCED CAPITAL EXPENDITURE

Data center power capacity includes buffers intended to absorb spikes in power use caused by peaks in resource utilization. These buffers are typically based on either nameplate, or nominal, server power consumption or power consumption measured at peak utilization with specific workloads. If we can better understand trends in server power use and map them to server functions and performance, we can reduce the size of these buffers. This substantially increases the effective capacity of existing data centers. As a result, we can avoid or defer capital investment to retrofit data centers and reduce capital expenditure associated with provisioning power for new data centers.

REDUCED POWER AND COOLING COSTS WITH POWER-AWARE JOB SCHEDULING

Ideally, jobs should always run on the most powerful, energy-efficient systems. If we can adjust job-scheduling capabilities to preferentially direct jobs to the most energy-efficient servers available, we can reduce operational power costs and the corresponding cooling needs.

MEASURING AND INCREASING COMPUTING ENERGY EFFICIENCY

We have undertaken several initiatives to analyze and improve computing energy efficiency within our design computing environment, including a long-term study and pilot project conducted at a data center in India.

These efforts included:

- Defining a computing energy-efficiency metric that reflects design computing production use.
- Using this measure—together with other data such as the number of servers, their compute utilization, and power utilization—to analyze computing energy efficiency across all servers in an Intel data center.
- Identifying efficiency improvements due to server refresh.
- Identifying opportunities to reduce power and cooling costs without negatively impacting design workload throughput. We took advantage of one of these opportunities to increase data center capacity by reducing a power buffer.

Computing Energy Efficiency Metric

Our first step toward improving computing energy efficiency was to define a useful measure based on actual Intel design workloads.

The most commonly used data energy efficiency metrics today are power usage effectiveness (PUE) and its reciprocal, data center infrastructure efficiency (DCiE). These metrics are defined by the Green Grid consortium, of which Intel is a member. PUE is defined as *total facility power/IT equipment power*. While PUE is very valuable for establishing facilities efficiency, it measures only the proportion of total power that goes to IT equipment, not the useful work that is done with that power.

EDA MIPS: A MEASURE OF DESIGN COMPUTING PERFORMANCE

We needed an energy-efficiency measure that reflected actual design computing server use within our environment. We decided to base our approach on measurements of performance/watt with design workloads. This is analogous to the Green Grid's recently proposed (but not finalized) data center energy productivity (DCeP) metric, which is defined as *useful work/energy consumed*.

To be realistic, our metric needed to be based on workloads representative of our actual design workloads. It needed to encompass both capacity and utilization, measure the performance of the entire server platform, and be applicable across different platforms. Because of these requirements, simplistic measures such as the number of cores or industry-standard CPU-specific metrics such as SPECint* were not adequate.

As a basis, we selected electronic design automation meaningful indicator of performance (EDA MIPS), an internal measure we had previously defined. It was originally created to compare the throughput of different Intel® server platforms for the purpose of calculating server refresh ratios—the number of new servers required to replace a given number of older systems to provide the same level of throughput.

Intel IT measures EDA MIPS by running a cross-section of real Intel electronic design automation (EDA) workloads on each platform. Figure 2 shows EDA MIPS ratings for several generations of Intel server platforms. Intel IT also measures server peak power consumption during the same tests that are used to determine the EDA MIPS ratings. This measurement provides the server's expected peak power consumption; Intel IT uses this to help determine data center capacity.

We defined computing energy efficiency as *utilized EDA MIPS/consumed watts*. To calculate this, we need to measure both system utilization and actual server power consumption.

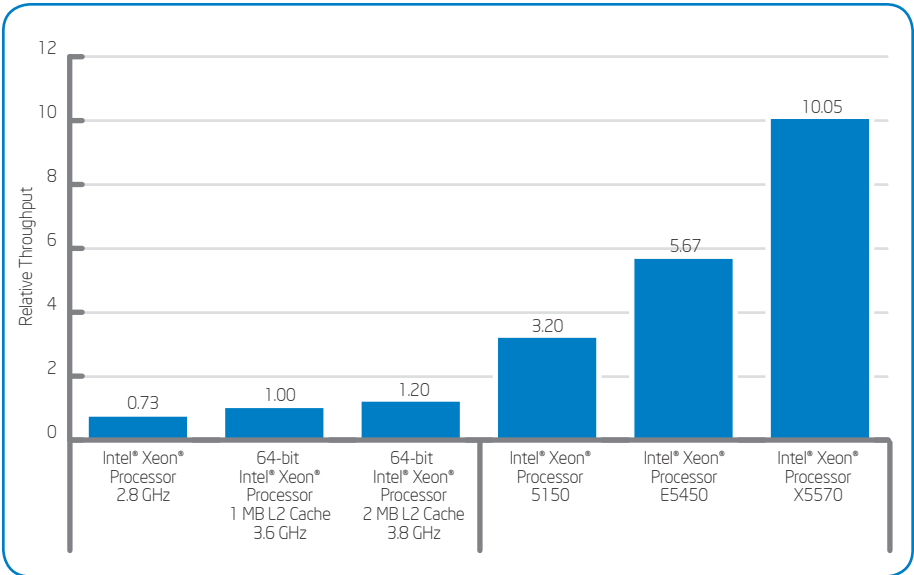


Figure 2. Relative performance of different Intel® server platforms measured in electronic design automation meaningful indicator of performance (EDA MIPS).

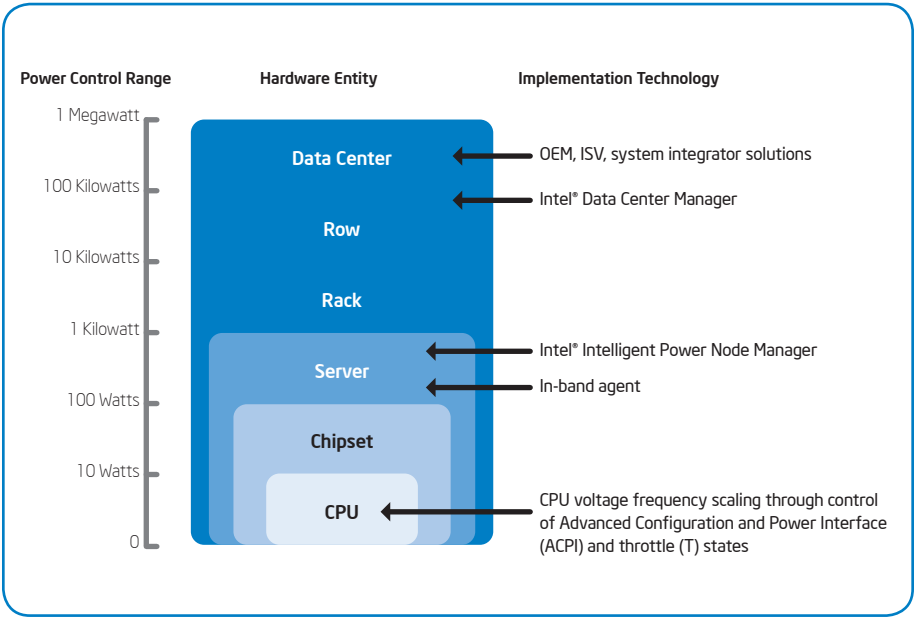


Figure 3. Technologies for measuring energy consumption within the data center. From “A Reference Architecture for Cloud Storage Power Management,” Enrique Castro-Leon, Intel Corporation. <http://communities.intel.com>

MEASURING POWER CONSUMPTION

To analyze and optimize computing energy efficiency, we need to measure server power consumption. However, at many organizations, data center power consumption is primarily measured at the facilities level; there is a lack of more detailed information about power consumption at the server, rack, or row level. This is primarily because there has been limited need to track this information, since the bill for energy consumption and the bill for server purchases are generally handled by different departments. Also, the cost of deploying additional instrumentation for measuring energy consumption can be significant.

There are a growing number of tools and technologies that enable us to gather more granular information about energy use, as shown in Figure 3. For example, we have already used row-level instrumentation to help us increase efficiency at one data center.²

METHODS FOR MEASURING SERVER POWER CONSUMPTION

Several options exist to measure server power consumption, including service processor-based power sensors, estimating power use based on utilization and server function, and technologies such as Intel® Intelligent Power Node Manager and Intel® Data Center Manager.

Service Processor-based Power Sensors

Our primary method of measuring server power consumption was to gather information from server-based power sensors. Most recent generations of server platforms have service processors that include power-monitoring sensors, accessible through a standardized interface.

² See the IT@Intel white paper “Increasing Data Center Efficiency Through Metering and Monitoring Power Usage.”

The Intelligent Platform Management Interface (IPMI) is one of the most common of these. IPMI operates independently of the OS and also allows out-of-band access. A Sensor Data Records (SDR) repository provides the properties of the individual sensors on the system board, including temperature, fan-speed, and power consumption.

For blade servers, the enclosure or chassis itself also has OEM-provided sensors that report power and temperature, among other readings. Some OEMs also provide proprietary methods of accessing such sensor data out-of-band. Since a mix of all these types—IPMI proprietary and blade-enclosure-based—exist in our data center, we developed a “wrapper” application that could use each different method to access service processor data but present a common interface to the user, thereby simplifying the task of data collection.

Estimating Power Use Based on Utilization and Server Function

Many data centers still include older systems lacking service processors or sensors. For these systems, two options exist:

- **Assume that the systems constantly use power at their peak power rating.** For these older machines, this provides a stop-gap approach that can be used until they reach their end of life. However, this method does not tell us how much energy the systems actually consumed.
- **Estimate power usage based on server utilization and function.** System utilization is correlated with power consumption; by taking a series of measurements, we can establish the relationship between system utilization and power consumption for a specific server model. This enables us to measure system utilization in our production environment and then use this measurement to estimate power

consumption. Once we have established the correlation factor for each server model, we can use this method to estimate server power consumption across the data center.

Intel® Intelligent Power Node Manager and Intel® Data Center Manager

Intel Intelligent Power Node Manager and Intel Data Center Manager (Intel® DCM) are promising new technologies that we plan to implement in the near future. Intel Intelligent Power Node Manager provides power monitoring and policy-based power management for an individual server, while Intel DCM scales these functions to racks and groups of servers. It implements group-level policies that aggregate data across the entire rack or data center to track metrics and historical data. These policies and data can also be integrated into other management consoles through the Intel DCM software development kit (SDK).

Analyzing Data Center Computing Energy Efficiency

We applied our methods for measuring server performance and energy use to analyze computing energy efficiency trends across a data center containing four generations of Intel®-based servers.

COMPUTING CAPACITY AND POWER CONSUMPTION BASELINE

To provide a baseline for measuring energy efficiency improvements, we first needed to calculate the data center's total computing capacity and expected peak power draw. To do this, we summed the EDA MIPS ratings and the expected peak power consumption for all servers across the entire data center. This involved accurately mapping server hostnames to server models, then mapping the server models to their EDA MIPS ratings and peak power consumption.

This provided a baseline expected performance/watt—based on the EDA MIPS ratings and expected peak power consumption of all servers—defined as *available EDA MIPS/expected consumption in kW*.

COMPUTING ENERGY EFFICIENCY

We define computing energy efficiency as *utilized EDA MIPS/consumed kW*. To calculate this, we:

- Captured system utilization data for all servers using in-house tools as well as open-source applications. We then multiplied the total data center EDA MIPS capacity by the percentage utilization to calculate utilized EDA MIPS.
- Measured power consumption using the appropriate method for each server:
 - For systems with service processor-based sensors, we gathered power consumption information using an IPMI-based tool.
 - For models lacking these sensors, we correlated server utilization and function to power consumption.
 - We augmented this data with power consumption data gathered from rack-level metered power distribution units and row-level external energy meters.

Using these measurements of utilized EDA MIPS and power consumption, we calculated the actual performance/watt achieved within the data center.

ANALYZING TRENDS

Our goal was to plot and analyze this data over time, in order to capture trends and identify opportunities to improve efficiency. We were able to do this because we had accurate data documenting the number of servers, when they were purchased, when they were retired, and their locations within the data center.

Figure 4 shows these statistics gathered over a 15-month period at the data center. This high-level view helped identify trends as well as opportunities for improving energy efficiency.

During this period, we periodically replaced older servers as part of our refresh strategy. Figure 4 clearly shows the effects of this strategy as well as the impact of significantly improved server utilization and initial efforts to improve data center energy efficiency.³ We found:

- A large decrease in the number of servers resulted in a corresponding but smaller drop in compute capacity. This is because we were removing the older, less-powerful servers from the environment.
- Small increases in the number of servers resulted in a relatively large increase in compute capacity. This is because we were adding new servers that are much more powerful than previous generations.
- Over time, the actual computing energy efficiency increased and exceeded our expected energy efficiency. The gap between the actual *utilized EDA MIPS/consumed kW* and the *expected EDA MIPS/kW* in Figure 4 reflects this sustained

improvement. This was due to significantly improved server utilization combined with lower-than-expected power consumption. The smaller variations in actual energy efficiency over the period shown are due to variations in server utilization.

- The data center compute capacity—available EDA MIPS—increased significantly over time; however, the number of servers required to provide this capacity actually decreased.
- The performance/watt—*utilized EDA MIPS/consumed kW*—increased substantially during 2009 as the proportion of newer, more energy-efficient server increased.

Energy Efficiency Improvements and Opportunities

In general, there are two approaches to improving energy efficiency as measured by performance/watt:

- Increase the numerator (performance). Possible steps include:
 - Increasing available EDA MIPS by making it possible to add more servers.
 - Increasing server utilization so that the utilized EDA MIPS become closer to the available EDA MIPS.

- Decrease the denominator (power consumption). Possible steps include:
 - Reducing the power consumed by the servers, even at high levels of utilization.
 - Reducing the power needed to cool the servers at high levels of utilization.

The data helped us identify opportunities in both of these areas for improving energy efficiency in 2009 and beyond. We have already taken advantage of one of these opportunities in a pilot project to increase data center capacity.

MAKING MORE COMPUTING CAPACITY AVAILABLE BY REDUCING POWER BUFFERS

Data centers typically are able to use about 80 percent of the power capacity provided to the facility. The remainder acts as a buffer to absorb spikes in IT equipment power usage. We undertook a pilot project to determine whether, by analyzing actual power consumption, we could reduce the size of the buffer and increase data center capacity as a result.

We have traditionally based data center capacity on measurements of peak server power consumption taken during the performance benchmark tests used to establish the server's EDA MIPS rating.

³ See the IT@Intel white paper "Increasing Data Center Efficiency Through Metering and Monitoring Power Usage."

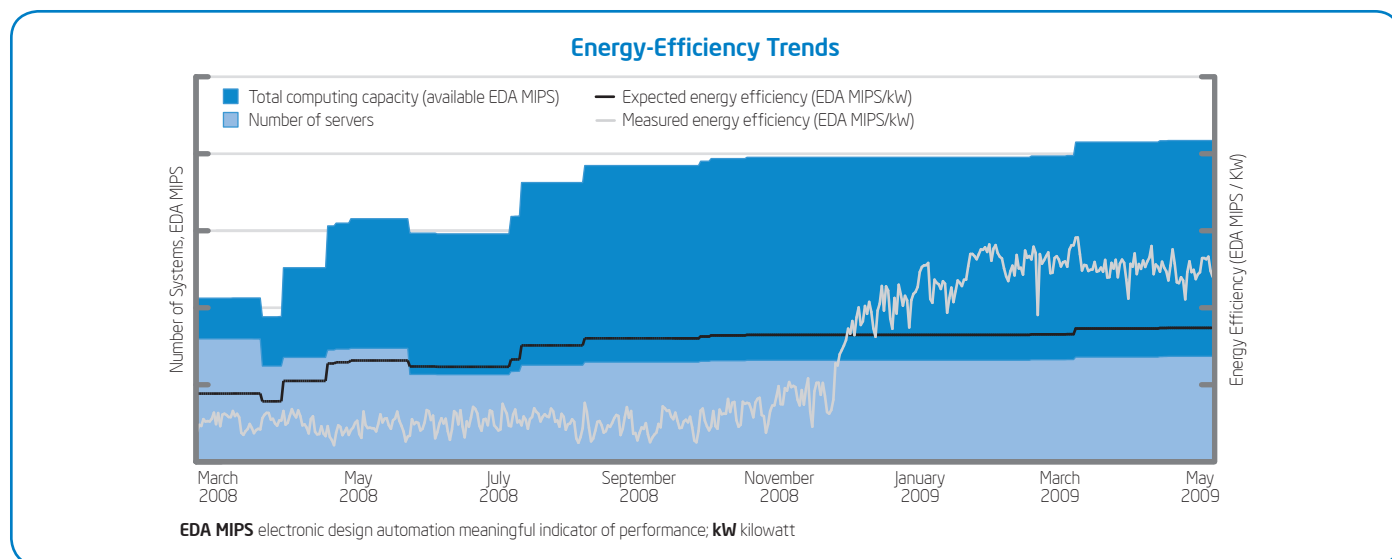


Figure 4. Compute energy-efficiency trends at an Intel data center. From March 2008 to May 2009, the number of servers has decreased by 14.2%, while compute capacity (EDA MIPS) has increased by 96.4% and energy efficiency (EDA MIPS/kw) has increased by 309%.

As shown in Figure 4, actual energy efficiency in 2009 increased to above the expected levels. This was because server utilization improved significantly—the utilized EDA MIPS approached the available EDA MIPS—while actual power draw was lower than the expected peak power consumption.

To understand why power consumption was lower than expected, we needed to correlate trends in actual server power consumption with specific functions and workloads as well as different levels of utilization. By performing this analysis, we identified specific workload types that resulted in significantly lower-than-average processor utilization. This allowed us to make informed decisions to increase the compute capacity that can be landed in a given power footprint.

In our pilot project, our goal was to analyze power consumption and determine whether we could land additional servers in a data center row that we had previously considered to be power-constrained. To do this, we:

- Analyzed trends in peak energy consumption and server utilization over a one-year period, as shown in Figure 5, and mapped

this information to specific server models and functions.

- Mapped the lower-than-expected power consumption to two specific types of workloads that were not processor-intensive. These workloads ran on dedicated systems, identified as Type 1 and Type 2 in Figure 5.
- Reduced these buffers by lowering the expected power usage thresholds of servers used for specific functions.
- Quantified the additional capacity gained as a result of lowering these thresholds and established the feasibility of landing new servers based on other constraints such as physical space and network ports.
- Added automated monitoring controls to help ensure that, if the thresholds are breached or server functions change, the server processors are “locked” (barred from accepting new workloads) and throttled down (reducing power consumption by changing the processor performance [P]-states). The data center and server operational teams are also automatically notified.

Figure 5 shows the results of analyzing actual utilization data for the two types of servers and mapping this against the expected utilization. This analysis showed that actual peak utilization (1 in the figure) during a one-year period was only about 60 percent—lower than the average 80 percent utilization in the design computing environment.

This meant that there was a utilization buffer of about 40 percent (2). Since we wanted to retain a buffer of only 15 to 20 percent, this represented an opportunity to leverage the remaining 20 percent buffer to add computing capacity. By mapping this utilization buffer to actual power consumption using the estimation technique previously described, we determined that this opportunity represented 10 percent of the available power for that data center row.

Based on the analysis, we reset our expected peak consumption (3), allowing us to retain a buffer of 20 percent while providing additional usable capacity (4); we used this capacity to land additional servers with power consumption equivalent to about 10 percent of the row's total power capacity.

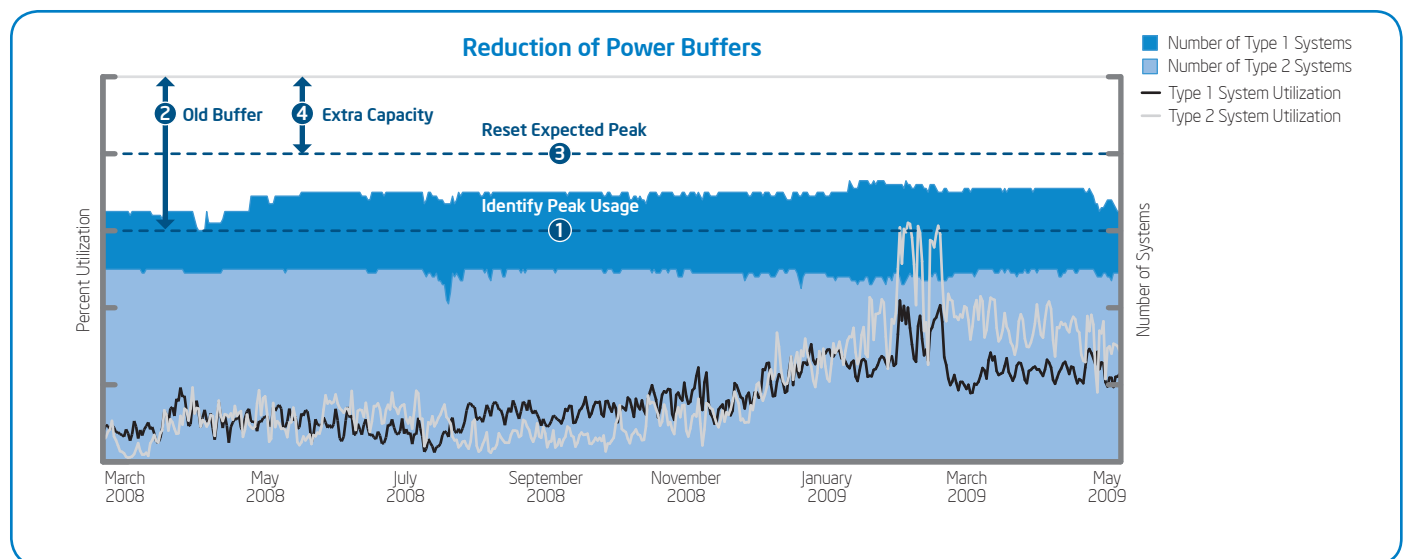


Figure 5. Utilization over time for one data center row.

CONCLUSION AND NEXT STEPS

The use of performance/watt to measure energy efficiency aligns well with the IT goals of improving utilization and reducing power consumption without impacting throughput of workloads.

Our analysis of compute energy efficiency has highlighted opportunities to achieve both of these goals. It has already delivered benefits, enabling us to add data center capacity by reducing power buffers. We are planning to use this method in other data centers during the coming year.

Additional Opportunities

We are continuing to build on our work with PoCs that explore some of the energy efficiency opportunities we have identified. We also intend to extend our approach to other computing areas within Intel.

POWER-AWARE JOB SCHEDULING

Energy-efficiency opportunities that we have identified include power-aware job scheduling.

When determining where a job should run, the most power-efficient and higher-performance systems in the data center should always be preferred. However, current batch job scheduling algorithms and configurations are tuned only to optimize performance; energy efficiency has been ignored.

We plan to study methods for implementing power-aware job scheduling, including

automated actions such as resubmitting jobs to more energy-efficient systems (as indicated by their performance/watt metric), as well as reconfiguring less energy-efficient systems for lower-priority work or moving them to low-power mode by changing the processor P and T states when idle.

Such opportunities will be significantly easier to leverage using Intel Intelligent Power Node Manager and Intel DCM with new Intel server platforms. These technologies will allow us to take more granular and accurate measurements as well as provide the capability to set and manage policies to govern power consumption.

EXTENDING SCOPE TO OTHER COMPUTING AREAS

To date, we have focused on applying our energy-efficiency metric—performance/watt or useful work/watt—to the Intel IT silicon design environment, since design computing has the largest data center footprint within Intel and accounts for about 70 percent of the installed servers.

However, the metric is generic enough to be applicable beyond the design environment. Our intention is to extend the scope of the energy efficiency measurements, targeting opportunities for improvement in the Intel IT office and enterprise computing domains. We also intend to extend this approach to storage; this will require a different metric to represent performance or useful work, with different normalization requirements.

CONTRIBUTORS

Ravindranath Madras

ACRONYMS

ACPI	Advanced Configuration and Power Interface
DCeP	data center energy productivity
DCiE	data center infrastructure efficiency
EDA	electronic design automation
EDA MIPS	electronic design automation meaningful indicator of performance
Intel® DCM	Intel® Data Center Manager
IPMI	Intelligent Platform Management Interface
kW	kilowatt
P	performance
PoC	proof of concept
PUE	power usage effectiveness
SDK	software development kit
SDR	Sensor Data Records
T	throttle
TCO	total cost of ownership

For more straight talk on current topics from Intel's IT leaders, visit www.intel.com/it.

Performance tests and ratings are measured using specific computer systems and/or components and reflect the approximate performance of Intel products as measured by those tests. Any difference in system hardware or software design or configuration may affect actual performance. Buyers should consult other sources of information to evaluate the performance of systems or components they are considering purchasing. For more information on performance tests and on the performance of Intel products, reference www.intel.com/performance/resources/benchmark_limitations.htm or call (U.S.) 1-800-628-8686 or 1-916-356-3104.

This paper is for informational purposes only. THIS DOCUMENT IS PROVIDED "AS IS" WITH NO WARRANTIES WHATSOEVER, INCLUDING ANY WARRANTY OF MERCHANTABILITY, NONINFRINGEMENT, FITNESS FOR ANY PARTICULAR PURPOSE, OR ANY WARRANTY OTHERWISE ARISING OUT OF ANY PROPOSAL, SPECIFICATION OR SAMPLE. Intel


disclaims all liability, including liability for infringement of any proprietary rights, relating to use of information in this specification. No license, express or implied, by estoppel or otherwise, to any intellectual property rights is granted herein.

Intel, the Intel logo, and Xeon are trademarks of Intel Corporation in the U.S. and other countries.

* Other names and brands may be claimed as the property of others.

Copyright © 2010 Intel Corporation. All rights reserved.

Printed in USA
0110/KAR/KC/PDF

 Please Recycle
322021-001US

