Energy Efficiency in Data Centers: A New Policy Frontier

By Joe Loper and Sara Parr
Alliance to Save Energy

with support from
Advanced Micro Devices

January 2007

Acknowledgments

Funding for this study came from Advanced Micro Devices (AMD) and corporate supporters of the Alliance to Save Energy. We are grateful for the technical assistance provided by AMD staff, particularly Donna Sadowy, Larry Vertal, Reed Content, Steve Kester, and also, Holly Evans, Strategic Counsel LLC.

Other people who provided helpful comments and insights include Dale Sartor, William Tschudi and Jon Koomey (Lawrence Berkeley National Laboratory), Christian Belady (Hewlett-Packard), Abir Trivedi, Gail Hendrickson, Steve Capanna, Jeff Harris, Lowell Ungar and Kateri Callahan (Alliance to Save Energy), Peter Douglas and Miriam Pye (New York State Energy Research and Development Authority) and Leslie Cordes (Environmental Protection Agency, formerly Alliance to Save Energy).

Of course, the authors are responsible for any errors or omissions.

About the Alliance to Save Energy

The Alliance to Save Energy promotes energy efficiency worldwide to achieve a healthier economy, a cleaner environment and greater energy security. Founded in 1977., the Alliance is co-chaired by Senator Mark Pryor (D-AR) and James Rogers (CEO, Duke Energy), the Alliance promotes energy efficiency policies, conducts research on various energy-related topics, and increases awareness and knowledge about the many ways that energy consumptions can be reduced in the United States and throughout the world.

Data Center Energy and Impacts

Computers are now an integral part of people's lives throughout the globe. Massive computational power and data storage, combined with global networking, has facilitated enormous improvements in living standards, including everything from medical science breakthroughs to home shopping, disaster preparedness to online banking, education of poor people in remote areas to instant access to music and videos.

Powering these information and product exchanges are thousands of data centers, which house about 10 million computer servers in the United States and 20 million worldwide.[1] Operating these devices – running 24 hours a day, 7 days a week – requires significant amounts of electricity. Industry competition and technological advances have helped increase computing power at the chip level while reducing their per-unit cost. Cheaper and cheaper computing power has driven consumers to demand more and more performance and server and processor manufacturers to deliver it.[2]

Until recently, the energy performance of chips and servers has been of less concern to consumers than their computational performance. This balance has shifted over the last couple years, especially as prices of electricity and natural gas (used for electricity generation) have increased. Computers have become much more efficient in terms of computational output per watt, however the power density has increased for the same volume or box size. Increased computing capabilities, combined with a trend toward centralization of corporate computing into data centers, has resulted in the doubling and tripling of data center energy intensities.[3]

Much of the increased power requirements are for ancillary equipment such as power and distribution equipment, and to remove heat from the servers and data centers, rather than the processors and servers themselves. All said, data centers typically consume 15 times more energy per square foot than a typical office building and, in some cases, may be 100 times more energy intensive.[4]

Data centers in the United States now consume an estimated 20 to 30 billion kilowatt hours of electricity annually,[5] roughly equal to the electricity consumption of Utah.[6]

---

[1] Timothy Morgan, "Server Market Begins to Cool in Q4," *The Linux Beacon*, February 28, 2006, http://www.itjungle.com/tlb/tlb022806-story03.html.
[2] Darrell Dunn, "Power Surge," *InformationWeek*, February 27, 2006.
[3] Kevin J. Delaney and Rebecca Smith, "Surge in Internet Use, Energy Costs, Has Big Tech Firms Seeking Power," *Wall Street Journal*, June 13, 2006, p. A1.
[4] Gary Shanshoian, Michele Blazek, Phil Naughton, Robert S. Seese, Evan Mills, and William Tschudi, *High-Tech Means High-Efficiency: The Business Case for Energy Management in High-Tech Industries*, Lawrence Berkeley National Laboratory, November 15, 2005 (retrieved August 16, 2006 from <http://repositories.cdlib.org/lbnl/LBNL-59127/) and Steve Greenberg, Evan Mills, Bill Tschudi, Peter Rumsey and Bruce Myatt, "Best Practices for Data Centers: Lessons from Benchmarking 22 Data Centers," *ACEEE Summer Study on Energy Efficiency in Buildings*, 2006, p. 3-76.
[5] Alliance calculations, based on a value of roughly 300W per each of 10 million servers suggest U.S. data center electricity consumption being around 30 TWh per year, depending on the assumed number of operating hours per year.

Nationwide, data center electric bills are now roughly $2 to $3 billion every year[7] and require roughly 30 power plants.

For some utilities, data centers have become a major portion of their load requirements. For example, Austin Energy, which serves a high-tech region in Texas, estimates that about 8.5 percent (200 MW) of its power is sold to data centers. Large server users such as Yahoo and Google are increasingly mindful of electric costs and building new server "farms" in places like the Pacific Northwest to take advantage of the region's low electricity rates. [8]

The number of installed servers is expected to increase by 40 to 50 percent nationally in the next four years, with sales of new servers numbering about 7 million per year.[9] If the current rates of growth continued and data center efficiencies remained unchanged, data center electric bills and power requirements would double in less than ten years, and data center electricity bills would increase by an additional $200 to $300 million each year.[10]

Fortunately, data center energy use can be reduced significantly. [11] There are many measures to reduce energy use, many with quick paybacks. The challenge is to increase awareness and comfort levels among data center owners and operators about the opportunities that exist and motivate them to make the necessary investments in time and equipment. In the balance of this paper, we discuss these various energy saving opportunities and policies and suggest a range of government policies and programs that, combined with industry knowledge and market reach, could increase deployment of energy-saving technologies and practices in data centers.

---

[6] Energy Information Administration, *State Electricity Profile: 2004*, released June 2006 (http://www.eia.doe.gov/cneaf/electricity/st_profiles/e_profiles_sum.html), reports retail sales of 24 billion kWh in Utah.

[7] Expenditures total about $3.3 billion according to Darrell Dunn, "Power Surge," *InformationWeek*, February 27, 2006. Alliance calculations based on 9 cents/kWh reported in Energy Information Administration, "Average Retail Price of Electricity to Ultimate Customers: Total by End-Use Sector, *Electric Power Monthly*, July 11, 2006 (http://www.eia.doe.gov/cneaf/electricity/epm/table5_3.html). To the extent that large data center operations are being located in regions of the country with low-priced electricity, the actual expenditures could be lower.

[8] Kevin J. Delaney and Rebecca Smith, "Surge in Internet Use, Energy Costs Has Big Tech Firms Seeking Power," *Wall Street Journal*, June 13, 2006, p. A1.

[9] See Darrell Dunn, "Power Surge," *InformationWeek*, February 27, 2006 and Timothy Morgan, "Server Market Begins to Cool in Q4," *The Linux Beacon*, February 28, 2006, http://www.itjungle.com/tlb/tlb022806-story03.html.

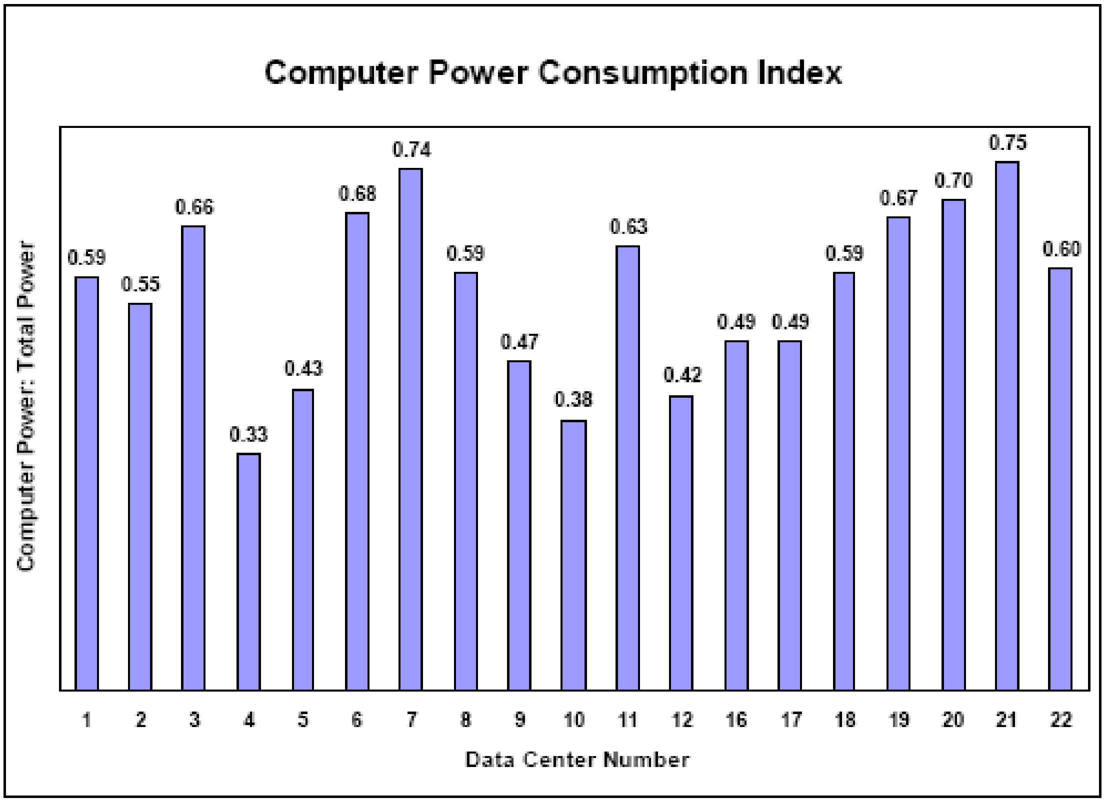
[10] Alliance estimates assuming energy costs increase proportionally to installed servers.

[11.] Christian Belady, "How to Minimize Data Center Energy Bills," *E-Business News*, September 5, 2006, http://www.line56.com/articles/default.asp?ArticleID=7881. See additional discussion in Christopher Malone and Christian Belady, "Metrics to Characterize Data Center and IT Equipment Energy Use", *Proceedings of 2006 Digital Power Forum*, September 2006.

Data Center Energy Use

Recent studies show that in themselves, computing operations -- including processors, memory, and in-box power supplies – comprise from one third to three-fourths of total data center energy us (for example, see Figure 1). The balance of data center energy is for cooling equipment to eliminate the heat that the servers create, to distribute power to the servers, and for lighting (see Figure 2).[12]

Figure 1



Source: Steve Greenberg, Evan Mills, Bill Tschudi, Peter Rumsey and Bruce Myatt, "Best Practices for Data Centers: Lessons from Benchmarking 22 Data Centers," *ACEEE Summer Study on Energy Efficiency in Buildings*, 2006, p. 3-77. Energy use will vary, with the specific operating conditions that are found in the data center under study. Efforts to identify standard metrics to quantify IT equipment power/total facility power, are ongoing.

Inside the server, only about one half of the power actually makes it to the processor (100 to 200 watts), even when operating at the manufacturers' rated capacity.[13] The rest goes to the in-box power supply and memory. Memory makes up a relatively small, but

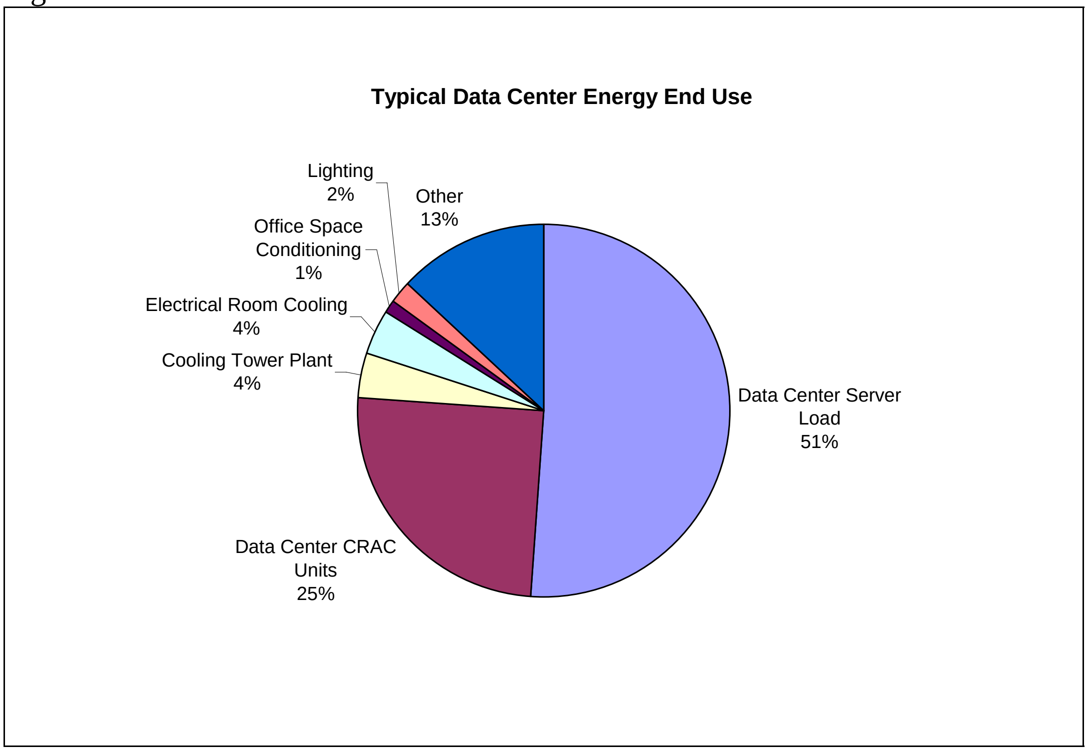
[12] William Tschudi, Tengfang Xu, Dale Sartor, and Jay Stein, "High-Performance Data Centers: A Research Roadmap," Lawrence Berkeley National Laboratory, March 30, 2004. Other sources of data center energy use include, for example, reference 11.

[13] The percentage of power consumed by the processor will vary depending on whether it is operating at full or partial capacity. As discussed below, processors typically operate at far below rated maximum capacity.

growing, share of server energy use.[14] Virtually all of the electricity that actually makes it to the processor ends up as heat that must be removed from the data center.

One study found that power distribution equipment outside the server – including transformers, uninterruptible power supplies, AC/DC inverters, and wiring – can make up 11 percent of data center energy consumption.[15] Lighting can make up about two to three percent of the energy consumed by data centers.[16]

Figure 2



Source: William Tschudi, Tengfang Xu, Dale Sartor, and Jay Stein, "High-Performance Data Centers: A Research Roadmap," Lawrence Berkeley National Laboratory, March 30, 2004. See reference 11 for alternative data (IT load 33%, cooling load 63%, UPS load 4% of data center power usage).

Most of the energy that goes into servers and ancillary equipment eventually ends up as waste heat that then must be managed by cooling systems.[17] Heat from processors, power supplies, power distribution equipment and lighting must be removed from the

[14] Stephen Shankland, "Electric Slide for Tech Industry," C/Net News.com, April 11, 2006. As manufacturers incorporate more efficient power supplies and processors into their servers (and the share of energy used by processors and power supplies decreases), the share of energy consumed by memory will increase (it is expected to consume more than half of a computer's power by 2008).

[15] Richard Sawyer, *Calculating Total Power Requirements for Data Centers*, White Paper #3, 2004 American Power Conversion, Inc., p.4.

[16] Richard Sawyer, *Calculating Total Power Requirements for Data Centers*, White Paper #3, 2004 American Power Conversion, Inc., p.4.

[17] David Moss, "Guidelines for Assessing Power and Cooling requirements in the Data Center," *Dell Power Solutions*, August 2005, pp.62-65; Neil Rasmussen, Electrical Efficiency Modeling of Data Centers, White Paper 13, American Power Conversion, 2005.

data center to avoid damaging servers and other equipment. Cooling systems – including air conditioners, fans, dehumidifiers, and pumps – typically represents about one-fourth to one-half or more of the electricity consumed by data centers. Indeed, cooling power requirements can (and often do) exceed the power used for the IT equipment itself.[18] Due to the high energy demand for cooling systems in data centers, cooling responsibility is beginning to shift from the data center operator to server vendors, some of whom are reportedly becoming "air-cooling specialists."[19]

Energy Efficiency Opportunities

While not all of the energy losses can be avoided, energy efficiency opportunities exist throughout most data centers. To reduce impacts and costs of energy use, data center operators have four basic choices: (1) they can reduce computational power of their data centers through efficient application management, (2) they can improve the efficiency of their servers; (3) they can improve the efficiency of their power supplies and distribution, or (4) they can improve the efficiency of their cooling systems. Some of these efficiencies may be more readily achieved by outsourcing the management of certain services and equipment.

In data centers, the capacities of various systems are much greater than necessary in order to compensate for inefficiencies in other areas of the data center, accommodate overly conservative system redundancy, and overestimation of future needs. "Right-sizing" – whether applications or cooling equipment, lighting or power supplies –offers probably the greatest number of efficiency opportunity in a data center. To the extent that systems are too big and thus operated at less than optimal loads, the actual efficiency will be less than the manufacturer's rated efficiency. Conversely, until maximum efficiencies have been realized throughout the system, it is difficult to optimally size servers, power supplies and cooling equipment.

Management of Applications

As the cost of computing power has decreased, the number of applications for computing power has increased. Software that can operate more efficiently presents one interesting option for reducing the energy consumed in data center operations. In addition, if data center processing requirements can be reduced through better management of applications, then data centers will need fewer servers and processors to achieve the same processing functionality. If data centers need fewer processors to achieve the same functions, they will require less power and, therefore, fewer power supplies and less distribution equipment, and also less cooling capacity.

Often, these applications are part of the operating system. When software and operating systems are not performing work functions, servers may operate in an idle state that is

---

[18] Richard Sawyer, *Calculating Total Power Requirements for Data Centers*, White Paper #3, 2004 American Power Conversion, Inc., p.4.
[19] Darrell Dunn, "Power Surge," *InformationWeek*, February 27, 2006.

still consuming power. This is somewhat analogous to leaving the light on when you leave the room. Additional applications, largely unbeknownst to the users, accumulate over time as "cookies" and are intentionally or inadvertently loaded on machines as programs are installed and uninstalled. Removing these applications from machines every few months could reduce processing requirements and energy use.

Given the relative inexpensive cost of processing power, software developers may not have considered these issues when writing software code. As concerns about data center energy use become more prominent, some software developers may be able to differentiate their products based on the energy demands that result from their code. More stable software could reduce the processing requirements of the primary server but also reduce the need for redundant processing capability, thereby, decreasing energy use.[20]

A significant opportunity for energy savings is to increase the utilization of each server. In most data centers, systems are larger than necessary to compensate for inefficiencies in other areas of the data center, as well as to provide an overly conservative system that allows for future growth. To the extent that servers and other equipment are operated at less than full capacity, the actual efficiency will be less than the manufacturer's rated efficiency. Correct sizing, therefore, is an important criterion for data center planning - from applications to cooling equipment.

Although "underutilization" of servers is sometimes intentional to improve reliability, "asset sprawl," in which applications are running on oversized servers reportedly pervades the nation's data centers.[21] Computer utilization rates of 10 to 15 percent are not uncommon.[22] A study by Hewlett Packard Lab of six corporate data centers found that most of their 1,000 servers were using only 10 to 25 percent of their capacity.[23] There are many examples in the trade press describing the reductions in server and related energy requirements resulting from better distribution of applications on right-sized servers could cut the installed base of servers in half.[24]

---

[20] For example, the 2003 version of Microsoft Exchange was reportedly much more stable and could handle larger number of users than previous versions.

[21] Andrew Hiller, "A Quantitative and Analytical Approach to Server Consolidation," *CiRBA White Paper*, January 2006, p.4.

[22] Also Dale Sartor, Personal correspondence, August 9, 2006.

[23] A. Andrzejak, M. Arlitt and J. Rolia, "Bounding the resource Savings of utility Computing Models," working paper HPL-2002-339, Hewlett-Packard Laboratories, Palo Alto, California, November 27, 2002. as cited in Nicholas G. Carr, "The End of Corporate Computing," MIT Sloan Management Review, Volume 46, Number 3, Spring 2005, pp. 67-73.

[24] See Timothy Morgan, "Server Market Begins to Cool in Q4," *The Linux Beacon*, February 28, 2006, http://www.itjungle.com/tlb/tlb022806-story03.html; Paul Allen, "CIO Challenge: Energy Efficiency," *Wall Street & Technology*, August 21, 2006, http://www.wallstreetandtech.com/showArticle.jhtml?articleID=192202377; and Joe Spurr, "Sun Fires running VMware cool hot data center," *Data Center News*, April 19, 2006, http://searchdatacenter.techtarget.com/originalContent/0,289142,sid80_gci1182957,00.html.

High levels of reliability require redundancy, so system operators often want a back-up server in case the primary server fails or when maintenance is required. However, consolidation of applications onto a single machine can often be done without sacrificing performance or reliability. [25] A number of companies[26] offer applications optimization services or software which tell clients how well servers are being used, identify periods and levels of maximum application use, suggest opportunities to merge servers, and predict future capacity needs.[27] Asset optimization – through "virtualization," consolidation of applications, and other methods -- can reduce software licensing costs, maintenance costs, and hardware replacement costs, in addition to energy costs.[28]

Another opportunity being explored is dynamically shifting load with a computer center to reduce hot spots, or shifting load from one data center to another (perhaps half way around the world). The capability to do this is driven by redundancy and reliability; however the energy and demand response opportunities are significant.[29]

Server Processors

Over the last several years, manufacturers of servers, processors, and other components have been working hard to increase the efficiency of their products. Most of the major server and processor companies are now touting improved performance per watt – performance improvements of 35% to 150% or more over previous generations.[30]

The three major areas of processor efficiency improvements include: "multi-core" processors, low voltage processors, and smaller chips made with advanced materials. Multi-core processors, in which a single processor has two or more processors that can run simultaneously when needed, use demand-based switching that allows only one processor to operate and draw power when processing loads are light. Chip manufacturers have also introduced low voltage chipsets which, for many applications and uses, offer sufficient performance and greater performance per watt. While leakage current increases with smaller geometries, both, Intel and AMD are developing smaller and smaller devices and utilizing new materials such as strained silicon, in order to enable lower leakage current and heat losses.[31]

---

[25] Andrew Hiller, "A Quantitative and Analytical Approach to Server Consolidation," *CiRBA White Paper*, January 2006, p.4.

[26] Companies include IBM, Microsoft, Sun Microsystems, Dell, CiRBA, VMWare, TeamQuest, Unisys, Information Systems Manager, Inc. (PerfMan),

[27] Correspondence from John Kean, Account Executive, ISM, Inc. August 17, 2006. PerfMan advertising "case history" of PacifiCorp use of PerfMan software.

[28] Christopher Lindquist, "Those Incredible Shrinking Servers, *CIO*, June 15, 2006. Bell Mobility was reportedly able to consolidate its server pool by 25 percent and save $1.5 million in hardware replacement expenses over two years. According to Joe Spurr, "Bell Mobility to save $200,000 per year on server consolidation," *Data Center News*, May 10, 2006, due to this consolidation, Bell Mobility expects savings of $200,000 in energy and software costs

[29] Dale Sartor, LBL, personal correspondence, November 17, 2006.

[30] This is a rough range of performance improvements based on scanning manufacturer literature and reviews.

[31] See http://www.intel.com/technology/silicon/si10031.htm , and, http://www.amd.com/us-en/Corporate/VirtualPressRoom/0,,51_104_543_13744~114609,00.html. For excellent summaries of

Importantly, processor efficiency improvements can significantly reduce the energy required in other areas of the data center, including power supplies, distribution systems and cooling systems.

Power Supply and Distribution

Existing servers in data centers may have in-box power supply units that are 60 to 70 percent efficient and have significant room for improved efficiency.[32] Many servers being shipped today, however, contain power supplies that are more efficient, even at lower loads.[33] A recent evaluation of a small sample of data center grade server power supplies with single 12V outputs found peak efficiencies in the range of 85-87%.[34] These power supplies are more efficient than the multiple-output, desktop-derived server power supplies evaluated in earlier studies; In addition to reducing the power losses in the power supplies themselves, efficiency improvements would reduce losses in distribution and cooling systems

Importantly, the rated efficiencies of power supplies are often much higher than actual efficiencies because power supplies seldom operate at a load under which the efficiency was originally calculated.[35] As noted previously, IT equipment is often operated at less than full capacity and may sit in standby mode for much of the time. Low-load losses in power supplies and other equipment are therefore a very important part of the efficiency equation, yet it is difficult to anticipate and include in manufacturers' ratings. New power management technologies for power supplies are intended to address the issue of decreased efficiencies at lower loads.

More efficient uninterruptible power supply (UPS) systems are also available. UPS systems provide back-up power (using batteries or flywheels) for servers in the event of a power outage. Typically, some power is always running through these systems so that

---

processor efficiency opportunities, see Ben Williams, "Maximizing Data Center Efficiencies through Microprocessor Innovation," presented to *Conference on Enterprise Servers and Data Centers*, January 31, 2006, http://www.energystar.gov/ia/products/downloads/BWilliams_Keynote.pdf and Arshad Mansoor, EPRI PEAC Corporation, and Brian Griffith, Intel Corporation, "Enabling High-Efficient Power Supplies for Servers," *Intel Technology Symposium*, 2004.

32 Chris Calwell (Ecos Consulting) in Matt Stansberry, "Vendor Tussle Over Measuring Server Efficiency," *SearchDataCenter.com*, February 7, 2006.

33 Jack Pouchet and Dave Douglas, "The Intelligently-Adaptive Data Center," *Proceedings of 2006 Digital Power Forum*, September 2006. For additional examples: http://h2000.ww2.hp.com/bc/docs/support/SupportManual/c00816246/c00816246.pdf?jumpid=reg_R1002_USEN ; also, http://www.ibm.com/news/nl/nl/2006/11/nl_nl_news_20061116.html; http://www.dell.com/content/topics/global.aspx/corp/environment/en/energy?c=us&l=en&s=corp.

34 Brian Fortenbery, Baskar Vairamohan, Peter May-Ostendorp, Power Supply Efficiency in High Density Data Center Servers, *Proceedings of 2006 Digital Power Forum*, September 2006.

35 Richard Sawyer, *Calculating Total Power Requirements for Data Centers*, White Paper #3, 2004 American Power Conversion, Inc., p. 6.

they can be started up instantly. A typical UPS system is about 90 percent efficient when fully loaded, with 95-percent and higher efficiencies available.[36]

Cooling

After reducing cooling requirements by improving the efficiency of the applications, the servers and the power supply systems, data center operators can take additional measures to improve cooling system efficiencies, including: right-sizing,"

As with servers, "right-sizing" cooling equipment is probably the most cost-effective opportunity for efficiency in data centers. The cooling equipment will run more efficiently, saving energy costs, and the up-front costs for the purchase of the equipment will decrease. There is a tendency to overcool data centers. Many data centers maintain a temperature of 68 degrees and 50 percent humidity to ensure moisture is not allowed to damage delicate equipment.[37] According to LBNL, these temperature and humidity levels are too conservative and the fluctuation of the temperature and humidity is probably more important than maintaining absolute levels.[38] Equipment manufacturers often do not provide environmental specifications with their products and data center operators often do not call and ask.

In addition, the use of "smart" or "adaptive" cooling solutions that dynamically modify data center cooling to match existing heat loads holds great promise as a power savings technology.

As the power and heat densities of data centers increases, it is important to ensure that hot and cool air are going to the right areas within the centers. Cool supply air to servers should not be allowed to mix with hot air being blown away from servers. Data centers should be wary of mixing servers that have different air flow (for example, the vents for exhaust air for one unit placed in front of another server) since this results in the mixing of hot and cold air, which reduces the efficiency of the cooling system and creates hot spots in the server rooms, which can damage equipment. The concept of cool aisle/hot aisle is now widely accepted practice, at least in large data centers. The efficiency of this system can be improved by making sure that all cable and other openings that would allow cold air in hot aisles and hot air in cold aisles are filled.[39]

Most data centers are still cooled with air. Liquid cooling can be far more efficient for data centers with high power densities that need to shed large concentrations of heat.

---

[36] Steve Greenberg, Evan Mills, Bill Tschudi, Peter Rumsey and Bruce Myatt, "Best Practices for Data Centers: Lessons from Benchmarking 22 Data Centers," *ACEEE Summer Study on Energy Efficiency in Buildings*, 2006, p. 3-83
[37] Don Beaty, PE, "Using ASHRAE specs for data center metrics," *Data Center Management Advisory Newsletter*, July 26, 2006. Beaty is the ASHRAE Technical Committee (TC 9.9) Chair.
[38] Lawrence Berkeley National Laboratory, *Data Center Energy Benchmarking Case Study*, Prepared by Rumsey Engineers, February 2003, p. 28.
[39] Lawrence Berkeley National Laboratory, Data Center Energy Benchmarking Case Study, Prepared by Rumsey Engineers, February 2003, p. 28. Also, Bob Sullivan, "Understanding the Impacts and Savings Potential for More Efficient Enterprise Servers and Data Centers," Conference on Enterprise Servers and Data Centers, January 2006. www.energystar.gov/serverconference.

Liquid cooling eliminates the mixing of cool air with hot air. In large data centers, air cooling systems often require about one watt for every watt of data center systems peak load that is being supported. By contrast, chilled water systems will require only about 70 percent of the system wattage.[40]

Outsourcing Data Center Operations

Rather than a company operating its own data center, some companies outsource their data center operations, either to acquire temporary capacity for special projects or, to avoid investment in hardware and human resources necessary to operate a reliable and efficient data center.[41] Some industry observer think such information technology "utilities" could be the wave of the future: "After pouring millions of dollars into in-house data centers, companies may soon find that it's time to start shutting them down. IT is shifting from being an asset companies own to a service they purchase."[42]

Increased outsourcing could have significant implications for reducing data center energy use if it results in higher rates of server utilization and brings greater operational efficiencies to overall data center operations.

Barriers to Improved Efficiency

Although energy saving opportunities in data centers are widespread and can reap significant cost savings, cost and other economic barriers often impede their adoption.

Despite the rising cost of electricity and the relatively large share of data center costs represented by energy expenses, data center energy costs remain a relatively small portion of overall costs for most data center clients, especially those with small and medium sized data centers.[43] Compounding the problem is that most data center managers never see the energy bill for their facilities, and their performance is not based on energy costs.[44] While the company or institution using the data center may benefit overall, the data center will probably see little reward. Separately metering data center operations would facilitate accountability and possibly provide an economic incentive directly to data center operators for improved energy performance.

While the performance and salary increase of most data center operators will likely never be based on the center's energy costs, disruption of operations resulting from attempts to

---

[40] Richard Sawyer, *Calculating Total Power Requirements for Data Centers*, White Paper #3, 2004 American Power Conversion, Inc., p.6.
[41] For example, see promotional information on Sun Grid Compute Utility at http://www.sun.com/service/sungrid/overview.jsp.
[42] Nicholas G. Carr, "The End of Corporate Computing," *MIT Sloan Management Review*, Volume 46, Number 3, spring 2005, pp. 67-73.
[43] Kevin J. Delaney and Rebecca Smith, "Surge in Internet Use, Energy Costs Has Big Tech Firms Seeking Power," *Wall Street Journal*, June 13, 2006, p. A1. According to Yahoo CFO, while electricity represents 20-50% of running a data center, it is a minor part of Yahoo's overall costs.
[44] Darrell Dunn, "Power Surge," *Information Week*, February 27, 2006.

institute new and untested software, hardware, or cooling innovations could very well threaten their jobs. For example, a leak in a liquid cooling system could be disastrous. Some software manufacturers advise operators not to run on virtual servers because they have not tested the virtualization software with their software and do not want to be held liable for any resulting instability.[45] It is not surprising that data center operators might not immediately adopt every new technology or idea. Successful adoption of new technologies and processes, especially by peers, is the most compelling information for a data center operator.

Data center operators need confidence in manufacturers' claims that operations will not be disrupted, they also need sufficient knowledge to convince budget decision makers that the measure is worth paying for. The rapid growth of data centers and disconnects between architects, financial managers, and data center operators often leads to poor data center design and operations. The challenge is in weighing the energy and other benefits against the potential risk to data center operations, and then communicating those benefits and risks to financial decision makers. Indeed, making a financial case for an efficiency improvement may be a greater hurdle to data center operators than implementation of the efficiency measures itself. Facilitating conversations between these groups of people and providing guidance that can be used in those communications is a critical need.

Perhaps the greatest barrier to energy efficiency improvements has been the rapid increase in new computer applications combined with the rapidly falling cost of processing power, especially if procurement criteria do not include energy efficiency. It is difficult to continually optimize data centers in such a rapidly changing environment and even more difficult to predict future requirements.[46] Because HVAC systems have long-lived components, while IT equipment has a relatively short life, HVAC systems often end up being mismatched with IT equipment and its cooling requirements.[47] Consequently, data center operators tend to err on the side of too much capacity rather than too little, resulting in less efficient operation.

As discussed above, outsourcing of data center operations can help concentrate expertise and consolidate applications to allow more efficient use of servers and other equipment. However, pricing for these services is sometimes based on the number of servers that are used. Consequently sellers of data center services may have little incentive to encourage efficient use of those servers through the consolidation and virtualization of applications and other efficiency improvements. Outsourcing has no energy efficiency improvement unless it results in a higher utilization rate for the servers being used or the data center service provider has made investments in efficiency that take advantage of economies of scale.

---

[45] Abir Trivedi, Alliance to Save Energy, personal Communication, August 17, 2006.
[46] Roger Schmidt and Don Beaty, "ASHRAE Committee Formed to Establish Thermal Guidelines for Datacom Facilities," *Electronics Cooling*, February 2005.
[47] Schmidt and Beaty.

Programs and Policies for Spurring Greater Energy Efficiency

To date, there has been little public sector role in promoting and facilitating improved data center energy efficiency and overcoming the barriers discussed above. As data centers comprise a growing portion of electricity use, governments and electric utilities are increasingly recognizing the need to take actions to reduce data center energy consumption. Because of the societal benefits achieved by data centers, it is important that these actions do not discourage innovation or impede server performance. Data center energy performance can be enhanced through any number of measures, including ones discussed below: energy performance metrics, design and operation guidance, government procurement, and financial incentives.

Energy Performance Metrics

Defining "energy performance" is a prerequisite for most energy efficiency policies and programs, and programs targeting data centers are no different. The way that performance is measured is critical, as it can determine if a server or other equipment will meet a consumer's needs or will be eligible for utility rebates or required as part of procurement requirements.

Developing performance metrics for even the simplest types of equipment can prove difficult and controversial.[48] For example, it has taken several years to develop widely accepted testing methods and energy efficiency standards for power supplies. The 80plus® performance specification, whose development was funded by electric utilities, requires computer and server power supplies to operate at 80-percent efficiency at 20, 50, and 100 percent of rated capacity.[49] More than 20 power supply manufacturers have submitted products for 80plus® certification and some computer manufacturers are now selling computers with certified power supplies.[50] Due in part to the efforts of the 80plus® program, the ENERGY STAR program is considering developing ENERGY STAR labels for power supplies in the future.[51]

---

[48] PowerPulse.net, "EPA and DOE's Energy Star Seek Comments on Proposed Energy Management Protocol," *PowerPulse.net*, August 22, 2006. Retrieved August 24, 2006 from http://www.powerpulse.net/news/story_print.php?storyID=15822 Also, Greg Papadopoulos, "Impacts and Importance of Energy Efficiency, Industry Viewpoint", Conference on Enterprise Servers and Data Centers, January 2006 at www.energystar.gov/serverconference; Christopher Malone and Christian Belady, "Metrics to Characterize Data Center & IT Equipment Energy Use, *Proceedings from 2006 Digital Power Forum*, September 2006; Magnus Herrlin, "Prerequisites for Successful Energy and Thermal Management in Data Centers: Metrics," *Proceedings of 2006 Digital Power Forum*, September 2006.

[49] As with power supplies, servers are often operated at far less than its rated capacity. As noted earlier, manufacturers' labels have traditionally been based on operation at maximum capacity, which can be misleading because power supplies, like servers and other data center equipment, often operate at partial loads.

[50] See 80-Plus web site at http://www.80plus.org/cons/cons.htm.

[51] Andrew Fanaro, Environmental Protection Agency, personal correspondence. Note that ENERGY STAR labels exist for external power supplies, just not internal power supplies, which are sold as OEM equipment in computers and servers.

EPA has also been working with industry and other stakeholders to develop measurement protocols for servers and, in November 2006, EPA distributed a final server energy efficiency measurement protocol drafted by Lawrence Berkeley Laboratory with other stakeholders.[52] Note that power supply performance is relatively straightforward—i.e., unit of energy output per unit of energy input – compared to measuring server energy performance. Applications processing is the desired output of a server, which is more difficult to measure than energy output.[53] At least one challenge is in the range of applications; Different computers (and chips) are optimum for different applications. This makes simple across-the-board comparisons difficult.

The server protocol presents a method for adding energy measurements to existing performance metrics (e.g., web pages served per minute). The result is a measure that will demonstrate the wattage consumed by a server at different loads. This will allow consumers to compare the energy consumption of different manufacturers' servers based on performance metrics that consumers and manufacturers are already familiar with and are using.

The protocol avoids, at least for now, debates about which performance metrics are better. Meanwhile, the Standard Performance Evaluation Corporation (SPEC), a non-profit organization that develops benchmarks for computers, expects to complete its first energy performance protocol—for small to medium-sized servers—in the beginning of 2007. SPEC will use its current application benchmarks as the basis for generating loads that are "typical of day-to-day server use."[54]

The importance of these efforts cannot be overstated. Standard measures of performance will increase the ability of data operators to compare products based on energy performance, making them more likely to consider energy performance in their purchasing decisions. An energy performance measurement standard, based on established and familiar server performance measures, will help increase users' confidence that the overall performance of the energy-efficient server will not be lower than the standard server. Last, but not least, widely accepted energy performance metrics and criteria for servers – and for entire datacenters - will provide the groundwork for other policies that will promote the development and purchase of energy efficient servers –including, for example, government procurement policies and financial incentives, such as tax rebates and credits.

---

[52] The protocol is available form EPA's web site at www.energystar.gov/datacenters and was written by LBL in collaboration with representatives from AMD, Dell, Hewlett Packard, IBM, Intel, Sun Microsystems, Stanford University, Rumsey Engineers, the California Data Center Design Group, and the Uptime Institute.

[53] If measured on the same basis as power supplies, servers would have almost zero-efficiencies because almost all of the energy that goes into the server is exhausted in the form of waste heat. Although this waste heat is undesirable, it does not mean that the server has failed at successfully running a database or web browsers.

[54] Standard Performance Evaluation Corporation, "SPEC to develop energy metrics for servers: New committee explores intersection of power use & performance," May 18, 2006. http://www.spec.org/specpower/pressrelease.html

Design and Operation Guidance

The growth of data centers is a relatively recent phenomenon, thus design and operational best practices have only recently been developed. Lawrence Berkeley National Laboratory (LBL) has been a pioneer in the field, providing technical resources to help designers and operators improve data center energy performance. The LBL website contains a wealth of information, including benchmarking tools, case studies, and technical reports covering everything from power supply efficiency measurements to surveys of data center operations.[55] To date, LBL has focused on power supplies, distribution and cooling and offers few resources pertaining to applications management and servers.

The American Society of Heating, Refrigeration, and Air-Conditioning Engineers (ASHRAE) develops standards for buildings and building systems, versions of which have been incorporated into most state and/or municipal governments' commercial codes. ASHRAE Technical Committee 9.9 Mission Critical Facilities, Technology Spaces, and Electronic Equipment was established to "provide better communications between electronic/computer equipment manufacturers and facility operations personnel to ensure proper and fault tolerant operation of equipment within data processing and communications facilities."[56]

ASHRAE Technical Committee 9.9 has developed guidance to address at least three aspects of data center design and operations that were discussed above:

1) Thermal Guidelines for Data Processing Environments—provide temperature and humidity guidance for data centers (as well as for other types of computer environments). [57]
2) Air flow protocols for servers—provide a common airflow scheme so exhaust from one manufacturer's equipment is not being ingested by another manufacturer's air inlet—i.e., the front face of all equipment is facing the cold aisle.[58]
3) Power Trend Charts—Address the mismatch of HVAC and IT equipment due to different lifetimes and help data center operators/designers predict future IT requirements. [59]

Other organizations that provide resources for data center operators and designers that may be useful in improving data center efficiency are summarized below.[60]

---

[55] LBL's data center energy efficiency web site is at http://hightech.lbl.gov.
[56] Roger Schmidt and Don Beaty, "ASHRAE Committee Formed to Establish Thermal Guidelines for Datacom Facilities," *Electronics Cooling*, February 2005.
[57] ASHRAE, 2004, Special Publication, "Thermal Guidelines for Data Processing Environments." American Society of Heating, Refrigerating and Air Conditioning Engineers, Inc. Atlanta , Georgia.
[58] ASHRAE 2004, Special Publication, "Thermal Guidelines for Data Processing Environments." American Society of Heating, Refrigerating and Air Conditioning Engineers, Inc. Atlanta , Georgia .
[59] ASHRAE, 2005, Special Publication, "Datacom Equipment Power Trends and Cooling Applications," American Society of Heating, Refrigerating and Air Conditioning Engineers, Atlanta, Ga..
[60] A number of additional organizations exist in universities or as part of labs associated with industry companies – e.g., IBM's research lab in Austin, Texas.

- *GreenGrid*: An association of information technology professionals seeking to lower the overall consumption of data centers globally. The GreenGrid allows members to share best practices in data center power management. (www.thegreengrid.org)
- *Efficient Power Supplies*: A website created by EPRI Solutions Inc. and Ecos Consulting to encourage a global discussion of energy-efficient power supplies. (www.efficientpowersupplies.org)
- *ITherm*: "An International conference for scientific and engineering exploration of thermal, thermomechanical, and emerging technology issues associated with electronic devices, packages, and systems." (www.itherm.com)
- *Consortium for Energy Efficient Thermal Management*: A collaboration of Georgia Institute of Technology and the University of Maryland to conduct "research on thermal and energy management of electronics and telecommunications infrastructure." (http://www.me.gatech.edu/CEETHERM/)
- *7x24*: An association facilitating the exchange of information for "those who design, build, use, and maintain mission-critical enterprise information infrastructures…7x24 Exchange's goal is to improve the end-to-end reliability by promoting dialogue among these groups." ((http://www.7x24exchange.org/index.html)
- *Uptime Institute*: Facilities exchange of information for improving reliability in data centers and information technology organizations. The *Institute* sponsors meetings, tours, benchmarking, best practices, research, seminars, and training. (http://www.uptimeinstitute.org)
- *AFCOM*: AFCOM started as an association of "a handful of data center managers looking for support and professional education." AFCOM membership now includes "more than 3,000 data centers" worldwide. AFCOM provides information to data center managers through annual conferences, published magazines, research and hotline services, industry alliances, and more. (http://www.afcom.com).

Clearly, a wealth of resources is available to data center operators trying to improve the energy efficiency of their operations. The challenge is to let data center operators know these resources are available and motivate them to take advantage of these resources.

Government Procurement

Federal, state and local governments spend tens of billions of dollars annually on energy-consuming products, thus offering thousands of opportunities to reduce government energy use through the purchase of energy efficient products. Furthermore, government procurement programs are used to help raise awareness of new-to-market energy efficient products, increase comfort levels with their use, and reduce costs of manufacture through economies of scale.

The federal government is required by law to purchase energy efficient products unless they are proven to not be cost effective. To help federal employees comply with these requirements, the Federal Energy Management Program at the Department of Energy (DOE-FEMP) along with EPA's ENERGY STAR program have developed energy performance specifications for approximately 70 types of products, including lighting equipment, heating and air conditioning, office equipment and more.[61] Typically, as a starting point, eligible products are in the top-25 percent of efficiencies of their product class.

An 80plus® type specification could be added to the federal product specifications to award the use of the most efficient server power supplies. In fact, some large institutional consumers are reportedly already specifying 80plus® requirements in their procurement policies. The 80plus® program web site provides downloadable procurement specifications to make it easy.[62]

Once energy performance measurements are developed for servers, DOE-FEMP or EPA could develop procurement requirements for them as well. Once minimum energy performance requirements are established for servers, federal agencies would be required to purchase only servers meeting those requirements (unless they provide written justification for not doing so).

Energy performance measurements could also facilitate financing of data center efficiency improvements through energy services performance contracts (ESPCs) and utility energy service contracts (UESCs). These alternative financing vehicles allow federal agencies to pay for the up-front costs from the stream of energy cost savings; Energy Services Companies and local utilities make the up-front investments.

Finally, establishing ENERGY STAR or DOE-FEMP eligibility requirements for servers could increase purchases of energy efficient servers by local, state and foreign governments, some of whom use ENERGY STAR and DOE-FEMP eligibility requirements in their own procurement practices.

Financial Incentives

Financial incentives can help buy down the additional cost of more efficient data center equipment, compensate for increased "hassle factor," draw attention to technologies, and legitimize the technologies in the eyes of the consumer, who sees that government and/or electric utilities are essentially endorsing these technologies. An important element of the 80plus® program is that participating utilities agree to provide a $10 rebate to manufacturers of servers that install power supplies that are certified to be at least 80 percent efficient. [63] These types of rebates are similar to rebates that have been offered by governments and utilities over the last couple of decades to a wide range of energy efficient products.

---

[61] http://www1.eere.energy.gov/femp/pdfs/eep_productfactsheet.pdf
[62] http://www.80plus.org/cons/cons.htm
[63] Utility-related participants listed on 80plus®Program web site are Efficiency Vermont, Midwest Energy Efficiency Alliance, National Grid, New York State Energy Research and Development Authority, Northwest Energy Efficiency Alliance, NSTAR, Pacific Gas & Electric, Sacramento Municipal Utility District, Southern California Edison, Western Massachusetts Electric, and XCel Energy. http://www.80plus.org/util/util.htm.

Financial incentives may also be catching on for servers. In August of 2006, Pacific Gas and Electric Company (PG&E) began offering $700 to $1,000 rebates for efficient servers (up to about 25 percent of the cost of servers). This is reportedly the first utility rebate program ever offered for servers.[64]

The challenge currently with this type of program is the lack of standardized server metrics. Once widely accepted energy performance measures are in place, such as those proposed by EPA and LBL (as discussed above), these kinds of programs could be offered more widely by utilities, as well as state and local, and foreign governments.[65]

Policy Recommendations

To date, the role of governments in data center efficiency has been limited. The environmental and other social benefits associated with the reduction of electricity consumption and demand warrants an expanded government role, specifically in the following areas:

Metering data center energy use – Governments should encourage submetering of data centers to help isolate energy efficiency opportunities among various loads and over time.

Energy performance measurement – Government should support efforts to develop server and power supply energy performance metrics.

Energy performance standards – Governments should impose minimum energy performance standards for power supplies at the 80-Plus level or better. Once performance measures have been established for server power supplies, governments should consider adoption of minimum energy performance standards.

Building codes – Governments should ensure that data center best practices are included in commercial building codes. The federal government should work with ASHRAE to ensure inclusion of data center systems that ensure a minimum of energy waste in commercial building standards including an analysis of the sizing of the cooling systems for server areas.

Financial incentives – Where energy performance measures are obtainable, systems are comparable, and budgets are available, governments should establish tax and/or utility incentives for servers, power supplies and other data center equipment and even best practices, such as virtualization and consolidation of applications.

---

[64] Sun Microsystems, Inc, press release, August 15, 2006, http://www.sun.com/smi/Press/sunflash/2006-08/sunflash.20060815.2.xml. According to Sun's press release, its "Cool thread" servers are currently the only ones that qualify for the rebate.

[65] Matt Stansberry, "PG&E to offer energy rebates for Sun servers," *Data Center News*, August 15, 2006, http://searchdatacenter.techtarget.com/originalContent/0,289142,sid80_gci1210722,00.html

Research – Governments should support research in at least the following three areas: 1) server and data center energy performance measures; 2) potential savings and costs from more efficient coding; and 3) potential savings and costs from better applications management

Pilot/Government Program Implementation – The federal government should establish a program and/or pilot projects that mimic real-word data center challenges. In this way, the government can test new energy-saving technologies and ensure compatibility with server software, minimizing the risk to server managers and making energy saving equipment more attractive.

Awareness – Governments should work with industry and others to raise awareness of existing information resources about data centers and promote awareness that significant opportunities exist. Government should establish best practices guidelines for distribution and assistance to companies and data center managers.