

Report to Congress on Server and Data Center Energy Efficiency
Public Law 109-431

U.S. Environmental Protection Agency
ENERGY STAR Program

August 2, 2007

Table of Contents

Table of Contents.....	i
Executive Summary	4
Background.....	4
Energy Use in Data Centers Through 2011	7
Incentives and Voluntary Programs to Promote Energy Efficiency	11
Recommendations.....	13
Conclusions.....	16
1. Introduction.....	17
1.1. Background.....	17
1.2. Data Center Energy Use.....	17
1.3. Data Center Characteristics.....	18
1.4. Energy Efficiency	23
1.5. Purpose of this Report.....	23
2. Trends in Growth and Energy Use Associated with Servers and Data Centers in the U.S. ..	25
2.1. Overview of Data Center Growth Trends.....	27
2.2. Estimates of U.S. Server and Data Center Energy Use	31
2.3. Energy Use Associated with Federal Government Servers and Data Centers.....	39
3. Potential Energy and Cost Savings through Improved Energy Efficiency.....	41
3.1. Expected Energy Savings from Current Energy Efficiency Trends	41
3.2. Opportunities for Additional Energy-Efficiency Savings.....	50
4. Electric Utility Impacts from Energy Efficiency in Servers and Data Centers	59
4.1. Methodology	60
4.2. Electricity Generation Impacts	62
4.3. Discussion of Transmission and Distribution Impacts	63
5. Potential Impacts of Energy Efficiency on Product Performance, Reliability, Features, and Overall cost.....	67
5.1. Performance Impacts Resulting from IT Energy Efficiency Improvements	68
5.2. Impact of Facilities Energy Efficiency on IT Equipment Performance	71
5.3. Summary	72
6. Distributed Generation and Combined Heat and Power Systems in Data Centers.....	73
6.1. Benefits of Clean DG and CHP for Data Centers.....	73
6.2. DG Applications at Data Centers.....	80
6.3. Issues Affecting Implementation of DG in Data Centers	82
7. Current Energy Efficiency Programs Applicable to Data Centers	84
7.1. Barriers to Energy Efficiency	84
7.2. Current Energy Efficiency Incentives and Voluntary Programs	89
8. Recommendations for Incentives and Voluntary Programs	106
8.1. Policy Recommendations.....	107
8.2. Recommendations for Research and Development Activities.....	117
8.3. Recommendations for Further Analysis	119
References.....	121
Acknowledgments.....	129

Acronyms and Abbreviations

AC	alternating current
AEO	Annual Energy Outlook
AHU	air handling unit
ANSI	American National Standards Institute
ASHRAE	American Society of Heating, Refrigeration, and Air-Conditioning Engineers
Btu	British thermal unit
CAGR	compound annual growth rate
CEMS	continuous emissions monitoring system
CEO	chief executive officer
CFO	chief financial officer
CIO	chief information officer
CHP	combined heat and power
CO	carbon monoxide
CO ₂	carbon dioxide
CPU	central processing unit
CRAC	computer room air conditioner
CRAH	computer room air handler
CW	chilled water
DARPA	Defense Advanced Research Projects Agency
DC	direct current
DCE	data center efficiency
DG	distributed generation
DOE	U.S. Department of Energy
DR	demand response
DSM	demand-side management
EERE	Energy Efficiency and Renewable Energy
EIA	Energy Information Administration
EPA	U.S. Environmental Protection Agency
EPAct 2005	Energy Policy Act of 2005
ESCO	energy service company
ESPC	energy services performance contract

FEMP	Federal Energy Management Program
GPS	global positioning system
GSF	gross square foot
GW	gigawatt
HDD	hard disk drive
HVAC	heating, ventilation, and air conditioning
ICE	Integrated Critical Environment
IECC	International Energy Conservation Code
IT	information technology
kW	kilowatt
kWh	kilowatt-hour
LBNL	Lawrence Berkeley National Laboratory
LEED	Leadership in Energy and Environmental Design
MAID	massive array of idle disks
MCFC	molten carbonate fuel cell
MMTCO ₂	million metric tons of carbon dioxide
MW	megawatt
MWh	megawatt-hour
NAECA	National Appliance Energy Conservation Act
NEMS	National Energy Modeling System
NO _x	nitrogen oxides
NYSERDA	New York State Energy Research and Development Authority
O&M	operation and maintenance
PAFC	phosphoric acid fuel cell
PBF	public benefit fund
PEM	proton exchange membrane
PG&E	Pacific Gas & Electric Company
PUE	power usage effectiveness
PV	photovoltaic
SPEC	standard performance evaluation cooperation
PDU	power distribution unit
PSRR	physical server reduction ratio
PSU	power supply unit

PUE	power usage effectiveness
Quad	quadrillion (10 ¹⁵) Btu
R&D	research and development
RFID	radio frequency identification
SATA	serial advanced technology attachment
SCR	selective catalytic reduction
SGIP	California Self Generation Incentive Program
SI-EER	site infrastructure energy-efficiency ratio
SO ₂	sulfur dioxide
SPEC	Standard Performance Evaluation Corporation
T&D	transmission and distribution
TCO	total cost of ownership
UEC	unit energy consumption
UESC	utilities energy service contract
UPS	uninterruptible power supply
VOIP	voice over internet protocol
VR	voltage regulator

Executive Summary

The United States (U.S.) Environmental Protection Agency (EPA) developed this report in response to the request from Congress stated in Public Law 109-431. This report assesses current trends in energy use and energy costs of data centers and servers in the U.S. and outlines existing and emerging opportunities for improved energy efficiency. It provides particular information on the costs of data centers and servers to the federal government and opportunities for reducing those costs through improved efficiency. It also makes recommendations for pursuing these energy-efficiency opportunities broadly across the country through the use of information and incentive-based programs.

Background

As our economy shifts from paper-based to digital information management, data centers — facilities that primarily contain electronic equipment used for data processing, data storage, and communications networking — have become common and essential to the functioning of business, communications, academic, and governmental systems. Data centers are found in nearly every sector of the economy: financial services, media, high-tech, universities, government institutions, and many others use and operate data centers to aid business processes, information management, and communications functions.

The U.S. data center industry is in the midst of a major growth period stimulated by increasing demand for data processing and storage. This demand is driven by several factors, including but not limited to:

- the increased use of electronic transactions in financial services, such as on-line banking and electronic trading,
- the growing use of internet communication and entertainment,
- the shift to electronic medical records for healthcare,
- the growth in global commerce and services, and
- the adoption of satellite navigation and electronic shipment tracking in transportation.

Other important trends contributing to data center growth in the government sector include:

- use of the internet to publish government information,
- government regulations requiring digital records retention,
- enhanced disaster recovery requirements,
- emergency, health and safety services,
- information security and national security,
- digital provision of government services (e.g., e-filing of taxes and USPS on-line tracking), and
- high performance scientific computing.

During the past five years, increasing demand for computer resources has led to significant growth in the number of data center servers, along with an estimated doubling in the energy used

Public Law 109-431

SECTION 1. STUDY.

Not later than 180 days after the date of enactment of this Act, the Administrator of the Environmental Protection Agency, through the Energy Star program, shall transmit to the Congress the results of a study analyzing the rapid growth and energy consumption of computer data centers by the Federal Government and private enterprise. The study shall include--

- (1) an overview of the growth trends associated with data centers and the utilization of servers in the Federal Government and private sector;
- (2) analysis of the industry migration to the use of energy efficient microchips and servers designed to provide energy efficient computing and reduce the costs associated with constructing, operating, and maintaining large and medium scale data centers;
- (3) analysis of the potential cost savings to the Federal Government, large institutional data center operators, private enterprise, and consumers available through the adoption of energy efficient data centers and servers;
- (4) analysis of the potential cost savings and benefits to the energy supply chain through the adoption of energy efficient data centers and servers, including reduced demand, enhanced capacity, and reduced strain on existing grid infrastructure, and consideration of secondary benefits, including potential impact of related advantages associated with substantial domestic energy savings;
- (5) analysis of the potential impacts of energy efficiency on product performance, including computing functionality, reliability, speed, and features, and overall cost;
- (6) analysis of the potential cost savings and benefits to the energy supply chain through the use of stationary fuel cells for backup power and distributed generation;
- (7) an overview of current government incentives offered for energy efficient products and services and consideration of similar incentives to encourage the adoption of energy efficient data centers and servers;
- (8) recommendations regarding potential incentives and voluntary programs that could be used to advance the adoption of energy efficient data centers and computing; and
- (9) a meaningful opportunity for interested stakeholders, including affected industry stakeholders and energy efficiency advocates, to provide comments, data, and other information on the scope, contents, and conclusions of the study.

SEC. 2. SENSE OF CONGRESS.

It is the sense of Congress that it is in the best interest of the U.S. for purchasers of computer servers to give high priority to energy efficiency as a factor in determining best value and performance for purchases of computer servers.

by these servers and the power and cooling infrastructure that supports them. This increase in energy use has a number of important implications, including:

- increased energy costs for business and government,
- increased emissions, including greenhouse gases, from electricity generation
- increased strain on the existing power grid to meet the increased electricity demand, and
- increased capital costs for expansion of data center capacity and construction of new data centers.

For these reasons, there has been mounting interest in opportunities for energy efficiency in this sector. To its credit, the information technology (IT) industry is actively investigating and developing solutions, such as power-managed servers and adaptive cooling.

The direct energy use of IT and infrastructure equipment is not, however, the only way that data centers affect energy use. The data processing and communication services provided by data centers can also lead to indirect reductions in energy use in the broader economy, which can exceed the incremental data center energy expenditures in some cases.¹ For instance, e-commerce and telecommuting can reduce both freight and passenger transportation energy use. Nonetheless, even though IT equipment may improve energy efficiency in the economy as a whole, pursuit of energy efficiency opportunities in data centers remains important because of the potential for rapid growth in direct energy use in this sector and the resulting impact on both the power grid and U.S. industries.

Role of EPA

EPA has a more than 15-year history of advancing energy efficiency in IT equipment as well as commercial buildings, beginning with the first ENERGY STAR specifications for computers established in 1992 and the Green Lights program established in 1991. Through the ENERGY STAR program, EPA now qualifies a wide array of IT products, including personal computers, imaging equipment, printers, and monitors. EPA has made particular strides in addressing standby energy and power management for these products, demonstrating that it is possible to encourage rapid development and adoption of energy-efficient technologies and practices. The energy savings from efficiency improvements in these products are currently in the billions of dollars per year (US EPA 2006). EPA has also developed an innovative commercial building rating system that helps owners and managers assess the energy performance of their buildings and target efficiency improvements.

In January 2006, EPA convened the first national conference dedicated to examining energy savings opportunities for enterprise servers and data centers. Representatives from the utility, financial services, healthcare, internet, and manufacturing sectors attended the conference (<http://www.energystar.gov/datacenters>). EPA is now working on the first priority identified in that conference, the development of objective measurements of server energy performance, on which future efficiency criteria would be based.

To develop this report, EPA convened a study team led by researchers from the Lawrence Berkeley National Laboratory. The study team offered stakeholders multiple opportunities to give input to and review this report, including:

- conducting preliminary calls with key stakeholders to help plan the study;
- holding a public workshop on February 16, 2007 (attended by approximately 130 people) to solicit input on the topic of energy efficiency in servers and data centers;
- following up on workshop attendees' offers of assistance, to gather and refine information for the study;
- posting on the ENERGY STAR web site an open call for interested parties to submit information, as well as a list of data needs;
- posting on the ENERGY STAR web site a public review draft of this report; and
- incorporating into the final version of this report comments on the public review draft from more than 50 organizations and individuals.

¹ The magnitude of indirect energy reductions attributable to IT equipment is uncertain; one of this report's recommendations is that research should be conducted to better understand this effect.

Energy Use in Data Centers Through 2011

The energy used by the nation's servers and data centers is significant. It is estimated that this sector consumed about 61 billion kilowatt-hours (kWh) in 2006 (1.5 percent of total U.S. electricity consumption) for a total electricity cost of about \$4.5 billion. This estimated level of electricity consumption is more than the electricity consumed by the nation's color televisions and similar to the amount of electricity consumed by approximately 5.8 million average U.S. households (or about five percent of the total U.S. housing stock). Federal servers and data centers alone account for approximately 6 billion kWh (10 percent) of this electricity use, for a total electricity cost of about \$450 million annually.

The energy use of the nation's servers and data centers in 2006 is estimated to be more than double the electricity that was consumed for this purpose in 2000. One type of server, the volume server, was responsible for the majority (68 percent) of the electricity consumed by IT equipment in data centers in 2006. The energy used by this type of server more than doubled from 2000 to 2006, which was the largest increase among different types of servers. The power and cooling infrastructure that supports IT equipment in data centers also uses significant energy, accounting for 50 percent of the total consumption of data centers. Among the different types of data centers, more than one-third (38 percent) of electricity use is attributable to the nation's largest (i.e., enterprise-class) and most rapidly growing data centers.

These energy consumption estimates were derived using a bottom-up estimation method based on the best publicly available data for servers and data centers. The estimation was performed as follows:

- estimated the U.S. installed base of servers, external disk drives, and network ports in data centers each year (based on industry estimates of shipments and stock turnover);
- multiplied by an estimated annual energy consumption per server, disk drive, or network port; and
- multiplied the sum of energy use for servers, storage, and networking equipment by an overhead factor to account for the energy use of power and cooling infrastructure in data centers.

This method was also used to develop five-year projections for future energy use. A five-year time horizon was chosen for the scenarios because this is the period for which equipment shipment forecasts were available, and a period for which change in the rapidly evolving IT sector can be reasonably forecasted. Two baseline scenarios were analyzed to estimate expected energy use in the absence of expanded energy-efficiency efforts. The "current efficiency trends" scenario projected the current energy use trajectory of U.S. servers and data centers based on recently observed efficiency trends for IT equipment and site infrastructure systems. The "historical trends" scenario did not reflect these current energy efficiency trends but simply extrapolated observed 2000 to 2006 energy-use trends into the future. The historical trends scenario projected the energy use of U.S. servers and data centers if no energy-efficiency improvements were made, and therefore indicates the energy savings associated with efficiency trends that are already under way.

Under current efficiency trends, national energy consumption by servers and data centers could nearly double again in another five years (i.e., by 2011) to more than 100 billion kWh (Figure ES-1), representing a \$7.4 billion annual electricity cost. The peak load on the power grid from

these servers and data centers is currently estimated to be approximately 7 gigawatts (GW), equivalent to the output of about 15 baseload power plants. If current trends continue, this demand would rise to 12 GW by 2011, which would require an additional 10 power plants.

These forecasts indicate that unless energy efficiency is improved beyond current trends, the federal government's electricity cost for servers and data centers could be nearly \$740 million annually by 2011, with a peak load of approximately 1.2 GW.

These estimates of data center energy use should be considered approximate because limited data are available on current data center energy use, and there is significant uncertainty about the effects of future technology trends, such as server consolidation and developments in network and storage technologies. However, these estimates and projections illustrate the magnitude of energy use in data centers and the need for effective energy-efficiency strategies. Energy consumption monitoring and reporting may be needed to both improve these estimates and inform future policy initiatives.

Energy-Efficiency Opportunities in Servers and Data Centers

There is significant potential for energy-efficiency improvements in data centers. Although some improvements in energy efficiency are expected if current trends continue, many technologies are either commercially available or will soon be available that could further improve the energy efficiency of microprocessors, servers, storage devices, network equipment, and infrastructure systems. For instance, existing technologies and design strategies have been shown to reduce the energy use of a typical server by 25 percent or more. Even with existing IT equipment, implementing best energy-management practices in existing data centers and consolidating applications from many servers to one server could reduce current data center energy usage by around 20 percent. Energy-efficiency strategies could be implemented in ways that do not compromise data center availability, performance or network security, which are essential for these strategies to be accepted by the market. To develop a better understanding of energy-efficiency opportunities that would accelerate adoption of energy-efficient technologies beyond current trends, three energy-efficiency scenarios were explored:

- The “**improved operation**” scenario includes energy-efficiency improvements beyond current trends that are essentially operational in nature and require little or no capital investment. This scenario represents the “low-hanging fruit” that can be harvested simply by operating the existing capital stock more efficiently.
- The “**best practice**” scenario represents the efficiency gains that can be obtained through the more widespread adoption of the practices and technologies used in the most energy-efficient facilities in operation today.
- The “**state-of-the-art**” scenario identifies the maximum energy-efficiency savings that could be achieved using available technologies. This scenario assumes that U.S. servers and data centers will be operated at maximum possible energy efficiency using only the most efficient technologies and best management practices available today.

Details of the key energy-efficiency assumptions used in this analysis are shown in Table ES-1. These assumptions represent only a subset of the energy-efficiency strategies that could be employed in practice; it is not a comprehensive list of all energy-efficiency opportunities available in U.S. data centers.

Table ES-1. Summary of Assumptions for Analysis of Alternative Efficiency Scenarios

Scenario	Data Center Subsystem	
	IT Equipment	Site Infrastructure (Power and Cooling)
Improved operation	<ul style="list-style-type: none"> Continue current trends for server consolidation Eliminate unused servers (e.g., legacy applications) Adopt “energy-efficient” servers to modest level Enable power management on 100% of applicable servers Assume modest decline in energy use of enterprise storage equipment 	30% improvement in infrastructure energy efficiency from improved airflow management
Best practice	<p>All measures in “Improved operation” scenario, plus:</p> <ul style="list-style-type: none"> Consolidate servers to moderate extent Aggressively adopt “energy-efficient” servers Assume moderate storage consolidation 	<p>Up to 70% improvement in infrastructure energy efficiency from all measures in “Improved operation” scenario, plus:</p> <ul style="list-style-type: none"> improved transformers and uninterruptible power supplies improved efficiency chillers, fans, and pumps free cooling
State-of-the-art	<p>All measures in “Best practice” scenario, plus:</p> <ul style="list-style-type: none"> Aggressively consolidate servers Aggressively consolidate storage Enable power management at data center level of applications, servers, and equipment for networking and storage 	<p>Up to 80% improvement in infrastructure energy efficiency, due to all measures in “Best practice” scenario, plus:</p> <ul style="list-style-type: none"> direct liquid cooling combined heat and power

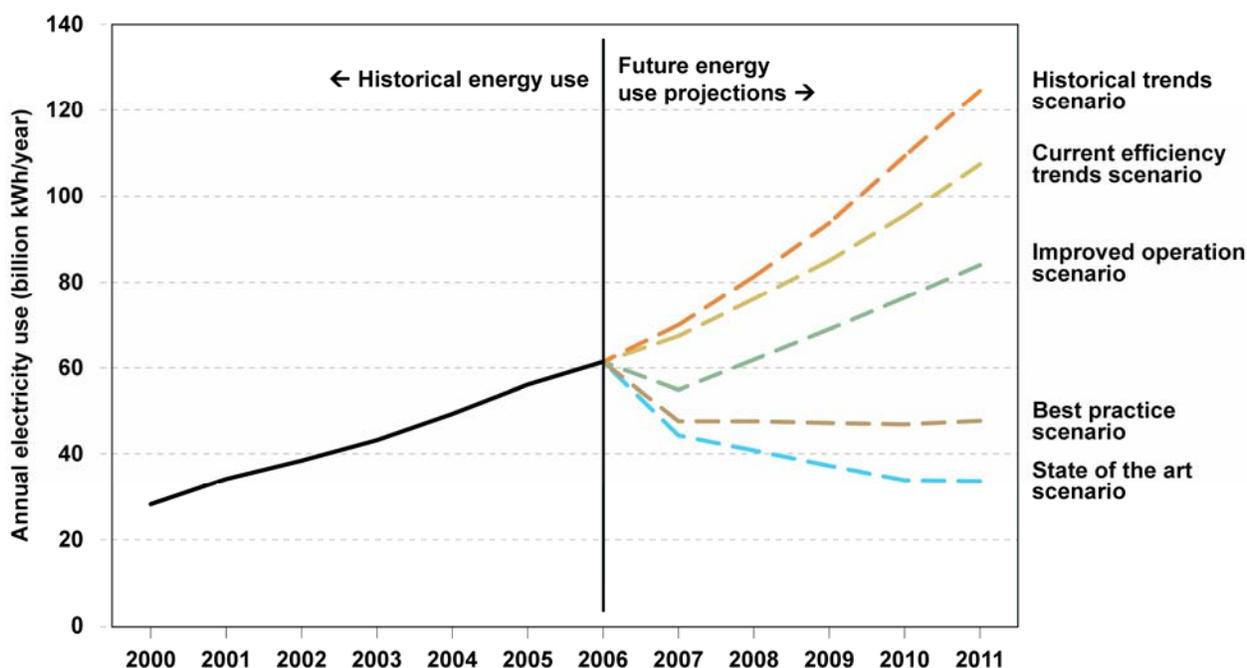
Note: These measures should be considered illustrative of efficiency opportunities in a typical data center. Some measures may only be applicable in new or expansion data centers or may be infeasible for a given data center because of local constraints. Selection of efficiency measures for a particular facility should be based on a site-specific review.

Because the best practice and state-of-the-art scenarios imply significant changes to data centers that may only be feasible to implement during major facility renovations, it was assumed in these scenarios that the site infrastructure measures requiring new capital investments would apply to only 50 percent of the current stock of data centers. For IT equipment, it was assumed that the entire existing stock turns over within the five-year forecast period.

These scenarios, based on the assumptions outlined above, illustrate significant potential for efficient technologies and practices to improve the energy efficiency of servers and data centers by 2011:

- The state-of-the-art scenario could reduce electricity use by up to 55 percent compared to current efficiency trends, representing the maximum technical potential.
- The best practice scenario could reduce electricity use by up to 45 percent compared to current trends, with efficiency gains that could be realized using today’s technologies.
- The improved operational management scenario offers potential electricity savings of more than 20 percent relative to current trends, representing low-cost energy efficiency opportunities.

Figure ES-1. Comparison of Projected Electricity Use, All Scenarios, 2007 to 2011



These scenarios show annual savings in 2011 of approximately 23 to 74 billion kWh compared to current efficiency trends, which reduces the peak load from data centers by the equivalent of up to 15 new power plants and reduces annual electricity costs by \$1.6 billion to \$5.1 billion. The projected savings in electricity use correspond to reductions in nationwide carbon dioxide (CO₂) emissions of 15 to 47 million metric tons (MMT) in 2011. The best practice scenario shows that electricity use in servers and data centers can be reduced below its 2006 level during the next five years rather than almost doubling, which would be the result if current efficiency trends continue.

Based on the assumption that the federal sector accounts for about 10 percent of electricity use and electricity costs attributable to servers and data centers, the annual savings in electricity costs in 2011 to the federal government range from \$160 million (for the improved operation scenario) to \$510 million (for the state-of-the-art scenario).

Table ES-2. Annual Savings in 2011 by Scenario (Compared to Current Efficiency Trends)

Scenario	Electricity consumption savings (billion kWh)	Electricity cost savings (\$billion 2005)	Carbon dioxide emissions avoided (MMTCO₂)
Improved operation	23	1.6	15
Best practice	60	4.1	38
State-of-the-art	74	5.1	47

These efficiency gains appear to be achievable without compromising product or data center performance. Because energy efficiency is a secondary attribute of the equipment used in data centers, changes that would compromise performance will generally not be implemented. In other words, data center designers and managers will first ensure that primary needs – performance and availability – are satisfied and will only then differentiate among products and practices based on energy efficiency. In some situations, improved energy efficiency increases performance and availability. For instance, better cooling distribution in data centers can eliminate hotspots and thereby prevent equipment faults. Finally, it is important to note that the energy-efficiency improvements addressed in this report reduce the costs of excess energy use and excessive power and cooling infrastructure.

The analysis in this report includes consideration of use of fuel cells and other distributed generation (DG) technologies in data centers. DG resources can reduce data center energy costs, particularly when used in combined heat and power (CHP) systems, which use waste heat to provide cooling. CHP systems can produce attractive paybacks and are well suited to the steady power and cooling loads of data centers. Clean DG also has the environmental benefits of reduced criteria pollutants and greenhouse gas emissions. Fuel cell DG systems offer many attractive qualities, such as DC power output, for use in data centers. But fuel cells, as a new-market entrant, have a premium price over more traditional DG systems. So while DG systems based on traditional gas turbine or engine technologies can be considered cost effective without incentives, fuel cells, in many cases, will need financial incentives to be cost effective. Finally, DG systems, particularly fuel cells, do not have a long track record in high power quality, high availability applications such as data centers. Given the high cost of outages for these types of facilities, more demonstration and conclusive information about system availability are needed before most facility designers and operators would likely be willing to adopt DG and CHP technologies.

Incentives and Voluntary Programs to Promote Energy Efficiency

To realize the potential benefits from greater energy efficiency in the nation’s data centers, a number of market barriers need to be addressed. The adoption of energy-efficient technologies and practices is often impeded by barriers such as higher first cost, lack of knowledge, institutional failures, and perverse incentives, and these issues apply equally to data centers. The barriers that prevent data centers from adopting changes that offer very reasonable paybacks are typically not technological but organizational. Three barriers of particular importance in data centers are:

- *Lack of efficiency definitions:* It is difficult to define energy efficiency for a complex system such as a data center or a server. “Energy efficient” is usually defined based on the delivery of the same or better service output with less energy input, but for servers and data centers service output is difficult to measure and varies among applications. Data center operators need standard definitions of productivity in order to purchase energy-efficient equipment, operate it in an optimal way, and design and operate the buildings to house it.
- *Split incentives:* In many data centers, those responsible for purchasing and operating the IT equipment are not the same people that are responsible for the power and cooling infrastructure, who in turn typically pay the utility bills. This leads to a split incentive, in which those who are most able to control the energy use of the IT equipment (and therefore the data center) have little incentive to do so.
- *Risk aversion:* With the increasing importance of digital information, data centers are critical to businesses and government operations. Thus, data center operators are particularly averse to making changes that might increase the risk of down time. Energy efficiency is perceived as a change that, although attractive in principle, is of uncertain value and therefore may not be worth the risk.

These barriers are not unique to data centers but may be more pronounced in this sector. There is a long history of incentive and informational programs to address barriers like these in other sectors – e.g., government agencies, public and private utilities. Although there are few current programs that specifically target data centers, existing energy policies and programs that promote high efficiency buildings and equipment – such as product labeling programs, commercial building technical assistance programs, financial incentives, and government procurement – may be applicable to data centers. These programs include:

- *Product labeling:* Labels identify products that meet certain specifications for performance, including high energy performance, based on standard methods for measuring energy efficiency. These labels can make customers aware of the energy costs associated with their purchasing decisions and encourage consumer acceptance and recognition of high-efficiency products. The performance specifications that underlie the labels form clear purchasing guidelines. This in turn encourages manufacturers to make increasing numbers of efficient products.
- *Commercial building technical assistance:* The growth of data centers is a relatively recent phenomenon, so best practices for design and operation are only recently being developed. Technical assistance programs provide information to facility designers and operators to help them effectively incorporate energy efficiency in the design and operation phases for their facilities. Newer practices in this area include establishment of whole-building energy performance benchmarking. Technical assistance can be provided by government agencies, electric utilities, professional organizations, and industry groups.
- *Financial incentives:* Electric utilities and governments often offer financial incentives to encourage investments in energy-efficiency measures. Financial incentives help buy down

the additional cost of more efficient products when initial product costs are higher than for less-efficient products, help compensate for the increased effort needed to learn about and locate energy-efficient equipment, draw attention to technologies, and legitimize these technologies in the eyes of consumers. The most active utility in the data center sector is Pacific Gas and Electric Company, which offers incentives for server consolidation, among other strategies.

- *Government procurement:* Federal, state, and local governments spend tens of billions of dollars annually on energy-consuming products, which means that there are thousands of opportunities to reduce government energy use through the purchase of energy-efficient products. Government procurement programs help raise awareness of new-to-market energy-efficient products, increase comfort levels as efficient products are put into use, and reduce costs of manufacture through economies of scale. The federal government is required by law to purchase energy-efficient products unless these products are proven to be not cost-effective. The government has developed energy performance specifications for more than 70 types of products.

EPA has begun addressing the energy performance of equipment in data centers by supporting development of energy-performance metrics for servers. In addition, governments and utilities are exploring program mechanisms for promoting improved efficiency.

Recommendations

A mix of programs and incentives is necessary to achieve a significant portion of the potential savings identified in the energy-efficiency scenarios above. Improvements are both possible and necessary at the level of the whole facility (system level) and at the level of individual components. Although it is not possible to optimize data center components without considering the system as a whole, it is also true that efficient components are important for achieving an efficient facility (for instance, efficient servers generate less waste heat which reduces the burden on the cooling system). Nevertheless, the greatest efficiency improvements will likely result from a comprehensive approach, given that there are opportunities for improvement in many areas of the IT equipment and infrastructure systems.

Based on a review of a range of incentives and voluntary programs that have been used in other sectors, and considering the unique aspects of the server and data center market, a number of recommendations can be made to pursue improved energy efficiency in the near term. These recommendations include:

- **Standardized performance measurement for data centers** — Data center operators need standard metrics to assess and report the energy performance of their facilities. The federal government and industry should work together to develop an objective, credible energy performance rating system for data centers, initially addressing the infrastructure portion but extending, when possible, to include a companion metric for the productivity and work output of IT equipment. These metrics should account for differences in data centers in areas such as computing output and availability requirements.

- **Federal leadership** — The federal government can act as a model in encouraging improved data center efficiency. The government should commit to: publicly reporting the energy performance of its data centers once standardized metrics are available, conducting energy efficiency assessments in all its data centers within two to three years, and implementing all cost-effective operational improvements. Additionally, the Architect of the Capitol should implement the server-related recommendations from the Greening of the Capitol report (Beard 2007).
- **Private-sector challenge** — The federal government should issue a challenge to private-sector chief executive officers (CEOs) to conduct DOE Save Energy Now energy-efficiency assessments, implement improvements, and report energy performance of their data centers. These assessments require protocols and tools that should be jointly developed by government and industry.
- **Information on best practices** — Objective, credible information is needed about the performance of new technologies and about best practices as well as the effect of both on data center availability. This information will help raise awareness of energy-efficiency issues in this sector and reduce the perceived risk of energy-efficiency improvements in data centers. The government should partner with industry to develop and publish information on field demonstrations and case studies of best practices. This information should be disseminated as part of a campaign to make data center managers aware of the benefits of energy efficiency in addressing power and cooling constraints in data centers.
- **Standardized performance measurement for data center equipment** — Purchasers of data center equipment, such as servers, storage, network equipment, and uninterruptible power supplies (UPSs), need objective, credible energy performance information if they are to purchase efficient products.
 - The federal government should work with industry to develop objective, credible energy performance metrics for this equipment.
 - Using these metrics, the government should also investigate whether development of ENERGY STAR specifications for these product categories would be an effective strategy to complement the whole-facility approaches outlined above.
 - If and when ENERGY STAR specifications are developed, federal procurement specifications that build on ENERGY STAR should be implemented.
- **Research and development**—The federal government, in collaboration with industry, universities, electric utilities, and other stakeholders, should initiate a comprehensive research and development (R&D) program to develop technologies and practices for data center energy efficiency. Specific research needs are identified in Chapter 8 (R&D recommendations) of this report, covering the following topics: computing software, IT hardware, power conversion, heat removal, controls and management, and cross-cutting activities.
- **Public/private partnership for energy efficiency in data centers**—The federal government should engage stakeholders to formulate a common initiative (including public policies and private-sector actions) to promote energy efficiency in data centers to continue the dialog that

this report initiates. Logical next steps would include defining priorities for the various strategies outlined in this report, developing timelines, defining roles for the various stakeholders, and identifying gaps and issues that require further assessment.

In addition to these near-term actions, several other actions can also play an important role in saving energy used by servers and data centers:

Federal Government:

- Develop a procurement specification for the energy performance of outsourced data centers.
- Work with industry to develop better tools, such as life-cycle risk models and total cost of ownership models that incorporate energy costs, for management of energy in data centers.
- Separately meter all federally owned data centers with significant energy use.
- Charge data center tenants for energy consumption of IT equipment in government-owned data centers.
- Partner with electric utilities, universities, and the data center industry to develop one or more neutral, “real-world” testing and demonstration centers (“National Center for Data-Center Best Practices”) to verify new technologies for reducing energy consumption in data centers.
- Help organize a technology procurement program to bring to market energy-efficient products for data centers.
- Partner with training organizations to develop education and training information and curricula about energy efficiency in data centers.
- Target data centers for efficiency upgrades using energy services performance contracts (ESPCs) and utility energy service contracts (UESCs).
- Provide technical assistance for demonstration projects of energy efficiency in data centers.
- Conduct demonstration and education projects for fuel cells and other clean, efficient DG technologies used for CHP in data centers.
- Develop a procurement specification to improve the efficiency of high-performance computing facilities.

State and Local Governments:

- Consider requiring separate utility meters on large data centers, either through utility regulation or building codes.
- Consider offering financial incentives for clean, efficient technologies used for CHP in high-availability installations (data centers, telecom facilities, etc.).

Electric Utilities:

- Consider offering incentives for energy-efficient data center facilities and equipment, based on the metrics described above.
- Consider partnering with the federal government to develop a neutral, “real-world” testing and demonstration center to verify new technologies for reducing energy consumption in data centers.

- Consider partnering with the federal government to develop a technology procurement program for efficient products.
- Consider offering education and training resources as a component of energy-efficiency programs for data centers.
- Consider offering financial incentives for clean, efficient DG and CHP in data centers.

Data Center Industry:

- Consider partnering with the federal government to develop an objective, credible energy-performance rating system for data centers.
- Consider partnering with the federal government to develop improved tools, such as “energy aware” total cost of ownership models and life-cycle risk models, for management of energy in data centers.
- Consider partnering with the federal government to develop a neutral, “real-world” testing and demonstration center to verify new technologies for reducing energy consumption in data centers.

Conclusions

This report helps define a vision for achieving energy efficiency in U.S. data centers. Although the growing energy use of servers and data centers makes this a challenging goal, there are large opportunities for savings. These savings will not be easy to achieve, given the barriers outlined in this report, but there are many policies available to overcome the barriers. Realizing these efficiency gains will take coordination and collaboration among many stakeholders: the government, the IT industry, data center operators, electric utilities, and others. The outlook for efficiency gains is encouraging, though, because industry is very engaged with these issues and is working with customers who are demanding solutions to the growing energy use in data centers. Federal initiatives should build on these efforts and partner in ways that develop objective, credible information, benchmarks, metrics, and industry standards. Finally, as a significant operator of data centers itself, the federal government can help facilitate change by changing the way it designs and operates its own facilities.

1. Introduction

During recent decades, computers have become more and more integral to the American lifestyle. From the mundane (email), to the miraculous (breakthroughs in medicine), increased computational power, combined with advances in data storage and global networking, have facilitated significant improvements to our quality of life. Despite these benefits, computers and data centers (buildings that house large numbers of networked computer servers) have come under scrutiny because of the increasing amounts of energy they consume (Loper and Parr 2007).

From 2000 to 2006, the energy used by U.S. servers and data centers and the power and cooling infrastructure that supports them has doubled (see Chapter 2). At the same time, these capital-intensive facilities are increasingly essential to the functioning of our digital economy. For these reasons, there has been mounting concern about the rapid growth of data centers energy consumption and interest in the opportunities for energy efficiency in this sector. Reflecting this concern, Congress passed Public Law 109-431, directing the U.S. Environmental Protection Agency (EPA) to study data center energy use, equipment, and opportunities for energy efficiency.

1.1. *Background*

A data center contains primarily electronic equipment used for data processing (servers), data storage (storage equipment), and communications (network equipment).² Collectively, this equipment processes, stores, and transmits digital information and is known as “information technology” (IT) equipment. Data centers also usually contain specialized power conversion and backup equipment to maintain reliable, high-quality power, as well as environmental control equipment to maintain the proper temperature and humidity for the IT equipment.

As our economy and society continue to shift from paper to digital information management, data centers have become ubiquitous -- they are found in nearly every sector of the economy -- and essential to the function of communications, business, academic, and governmental systems. All but the smallest companies have some kind of data center, and larger companies often have many tens, or even hundreds, of data centers. Smaller data centers are commonly located within larger commercial buildings, and larger data centers tend to be buildings constructed specifically to serve their purpose that can be up to several hundred thousand square feet in size.

Universities, municipalities, and government institutions also use and operate many data centers for information management and communication functions.

1.2. *Data Center Energy Use*

Regardless of their use and configuration, most data centers are more energy intensive than other buildings. This is due to the high power requirements of the IT equipment and the power and cooling infrastructure needed to support this equipment. In fact, data centers can be more than 40 times as energy intensive as conventional office buildings (Greenberg et al. 2006), meaning

² This study excludes from the definition of “data center” any facilities that are primarily devoted to communications (e.g., telephone exchanges), including network equipment located in telecom data centers. The definition does, however, follow market research firm IDC’s convention of including in the definition any room that is devoted to data processing servers, i.e., server closets and rooms (Bailey et al. 2006).

that large data centers more closely resemble industrial facilities than commercial buildings with respect to energy use. A brief summary of data center energy use and associated issues highlights the following facts:

- In 2006, electricity consumed by servers in U.S. data centers (including cooling and auxiliary infrastructure) represented about 1.5 percent of national electricity use (see Chapter 2).
- A single data center can house hundreds or thousands of servers, storage devices, and network devices. Continued growth in the number of servers is expected as companies look to expand data center capabilities.
- Power densities are also increasing in data centers (see Figure 2-5). As facility managers try to squeeze more computing power into less space, the energy consumption of a single rack of servers can exceed 20 kW.
- According to AFCOM's Data Center Institute (AFCOM 2006), power failures and limits on power availability will interrupt data center operations at more than 90 percent of all companies over the next five years.
- A survey conducted by Ziff-Davis (Ziff Davis 2005) suggests that more than 70 percent of operators identify IT power and cooling as a primary issue in data center management.
- More and more companies are reporting their greenhouse gas emissions as part of corporate responsibility initiatives, and government agencies are striving to meet energy use goals. Energy-intensive data centers can become a significant portion of an organization's energy use and greenhouse gas emissions, and thus challenge an organization's ability to meet these targets (Raths 2006).

1.3. Data Center Characteristics

Data centers are designed for computers, not people. As a result, data centers typically have no windows and minimal circulation of fresh air. They can be housed in new construction designed for the purpose or in existing buildings that have been retrofitted. Data centers range in size from small rooms (server closets) within a conventional building to large buildings (enterprise-class data centers) dedicated to housing servers, storage devices, and network equipment. Large data centers are becoming increasingly common as smaller data centers consolidate (Carr 2005).

Data center rooms are filled with rows of IT equipment racks that contain servers, storage devices, and network equipment. Data centers include power delivery systems that provide backup power, regulate voltage, and make necessary alternating current/direct current (AC/DC) conversions. Before reaching the IT equipment rack, electricity is first supplied to an uninterruptible power supply (UPS) unit. The UPS acts as a battery backup to prevent the IT equipment from experiencing power disruptions, which could cause serious business disruption or data loss. In the UPS the electricity is converted from AC to DC to charge the batteries. Power from the batteries is then reconverted from DC to AC before leaving the UPS. Power leaving the UPS enters a power distribution unit (PDU), which sends power directly to the IT equipment in the racks. Electricity consumed in this power delivery chain accounts for a substantial portion of overall building load.

Electricity entering servers is converted from AC to low-voltage DC power in the server power supply unit (PSU). The low-voltage DC power is used by the server's internal components of,

such as the central processing unit (CPU), memory, disk drives, chipset, and fans. The DC voltage serving the CPU is adjusted by load specific voltage regulators (VRs) before reaching the CPU. Typical power levels for these various server components are shown in Table 1-1 (Fan et al. 2007). Electricity is also routed to storage devices and network equipment, which facilitate the storage and transmission of data.

The continuous operation of IT equipment and power delivery systems generates a significant amount of heat that must be removed from the data center for the equipment to operate properly. Cooling in data centers is often provided by computer room air conditioning (CRAC) units, where the entire air handling unit (AHU) is situated on the data center floor. The AHU contains fans, filters, and cooling coils and is responsible for conditioning and distributing air throughout the data center. In most cases, air enters the top of the CRAC unit and is conditioned as air passes across coils containing chilled water pumped from a chiller located outside of the data center room. The conditioned air is then supplied to the IT equipment (primarily servers) through a raised floor plenum. Cold air passes through perforated floor tiles, and fans within the servers then pull air through the servers. The warmed air stratifies toward the ceiling and eventually makes its way back to the CRAC unit intake.

Most air circulation in data centers is internal to the data center zone. The majority of data centers are designed so that only a small amount of outside air enters. Some data centers provide no ductwork for outside air to directly enter the data center area. Instead, outside air is only provided by infiltration from adjacent zones, such as office space. Other data centers admit a relatively small percentage of outside air to keep the zone positively pressurized.

Data centers use a significant amount of energy to supply three key components: IT equipment, cooling, and power delivery. These energy needs can be better understood by examining the electric power needed for typical data center equipment in and the energy required to remove heat from the data center.

Table 1-1. Component Peak Power Consumption for a Typical Server

Component	Peak Power (Watts)
CPU	80
Memory	36
Disks	12
Peripheral slots	50
Motherboard	25
Fan	10
PSU losses	38
Total	251

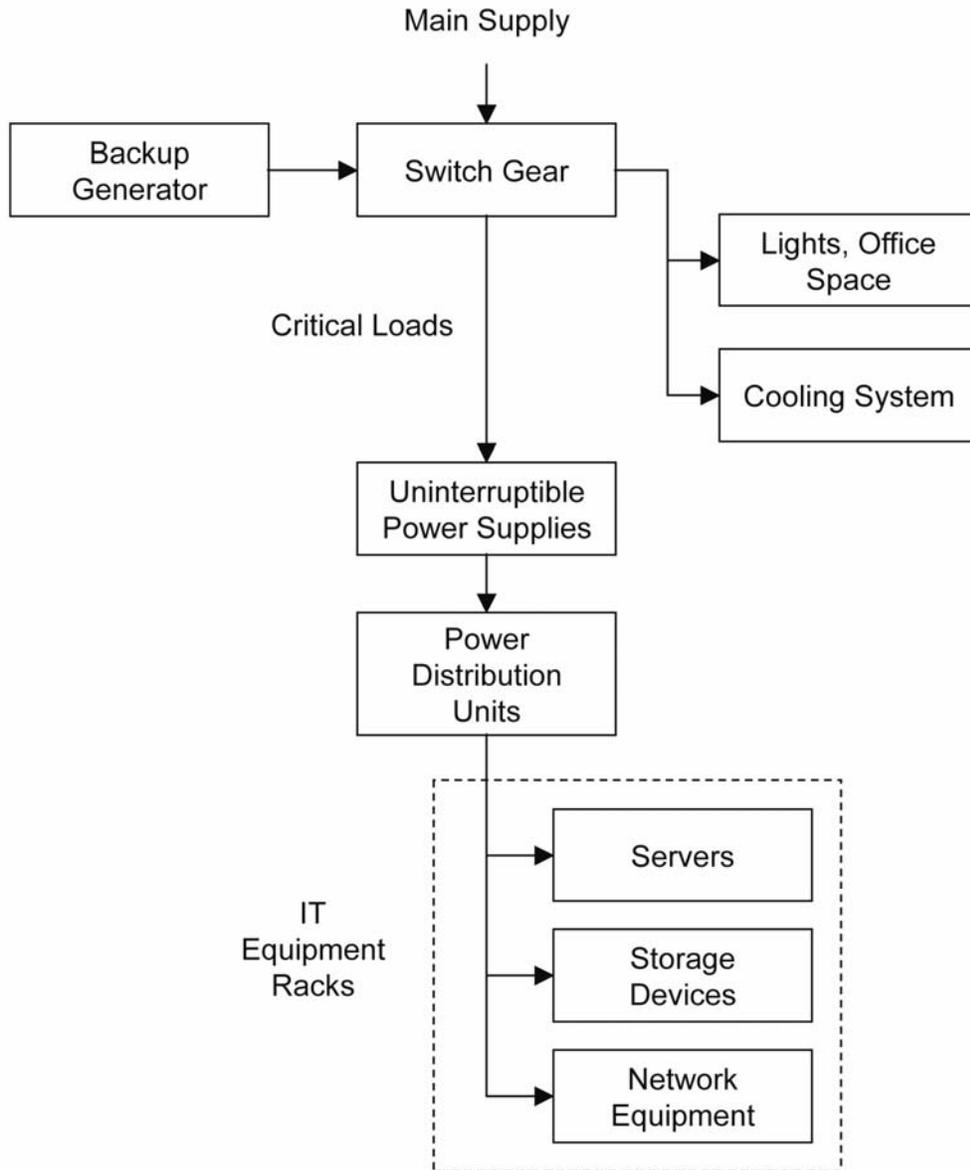
Source: derived from Fan et al. (2007)

1.3.1. Electric Power

Figure 1-1 shows the energy-consuming equipment in a typical data center. This includes equipment that performs primary IT functions as well as equipment that ensures continuous

operation. In many data centers, operation of IT equipment is critical, and UPS equipment is designed to maintain electricity supply even during utility grid disruptions.

Figure 1-1. Typical Electrical Components in a Data Center³



Source: derived from Fan et al. (2007) and Turner et al. (2005)

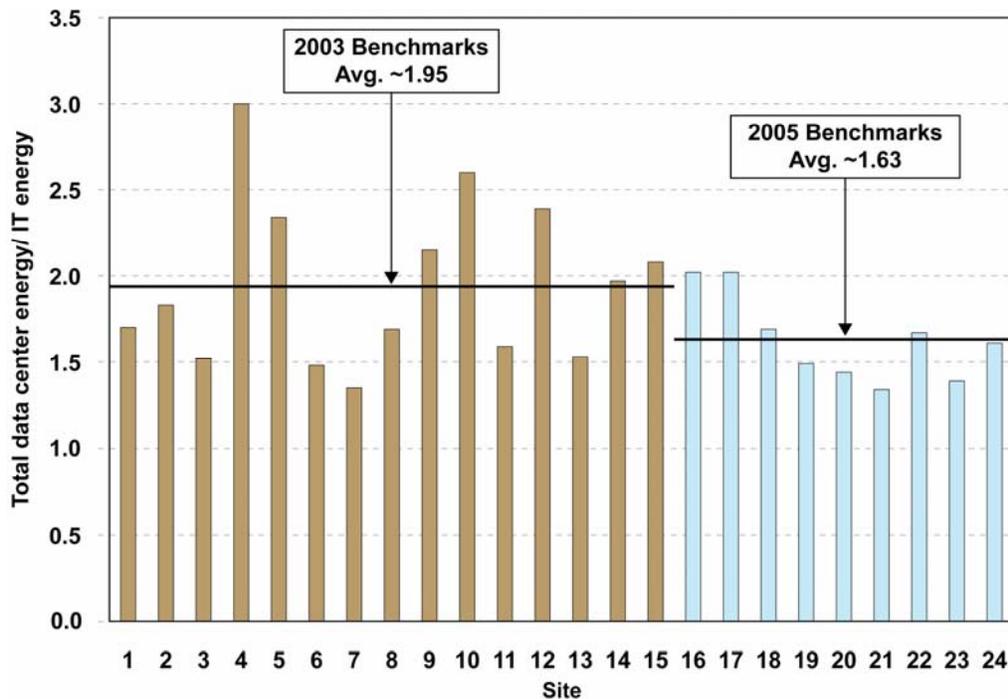
Data center equipment generally exhibit high power intensities with all of the electric power converted to heat. A recent survey of power usage in more than 20 data centers found that a data

³ The electricity used for lights and office spaces is typically only a small fraction of total data center energy use. Thus, the treatment of energy use and energy efficiency opportunities for data centers in this report focuses exclusively on the major users of energy; namely, IT equipment and power and cooling systems.

center's IT equipment alone can use from about 10 to almost 100 Watts per square foot of raised floor area (Greenberg et al. 2006, LBNL 2006).⁴ Moreover, power intensities have been increasing over time, largely because of the increasing heat density of data-processing equipment. If the power and cooling overhead needed to support the IT equipment are factored in, only about half the power entering the data center is used by the IT equipment (see Figure 1-2). The rest is expended for power conversions, backup power, and cooling. Peak power usage for data centers can range from tens of kilowatts (kW) for a small facility, to tens of megawatts (MW) for the largest data centers.

Increasing power density can lead to a situation in which companies are forced to build new data centers not because they are running out of floor space but because they need power and cooling beyond what can be provided in their existing data centers. This situation has driven much of the recent interest in energy-efficiency improvements for data centers. If the power consumed (and resulting heat generated) in data centers can be reduced through energy-efficiency measures, the existing infrastructure can continue to meet cooling and power needs, and costly investments in new data centers can be deferred.

Figure 1-2. Data Center Energy Benchmarking Results for 24 sites
(Total Data Center Energy ÷ IT Equipment Energy)



Source: (Greenberg et al. 2006, LBNL 2006)

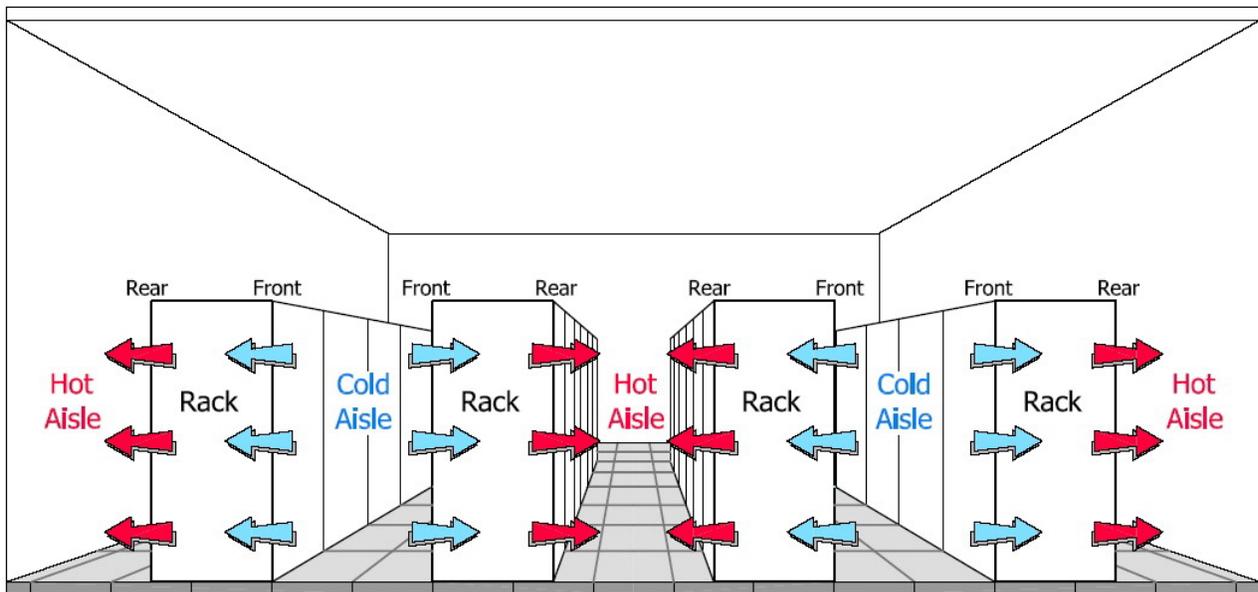
⁴ It is important that power densities be defined consistently, to allow comparison among different data centers. Appendix 1 describes the conventions used in this report.

1.3.2. Heat Removal

The role of data center air-conditioning systems is to keep the components of the electronic equipment within the manufacturers' specified temperature/humidity range. Electronic equipment in a confined space generates a significant amount of heat, and the equipment's reliability is reduced if it is not adequately cooled. High and low relative humidity levels have also been shown to increase failure rates of electronic components in data centers. Recommended upper and lower relative humidity limits are set to limit these problems (ASHRAE 2004b).⁵ To put the cooling load in perspective, consider that a fully populated rack of blade servers requires up to 20-25 kW of power to operate (Hughes 2005). This is equivalent to the peak electricity demand of about 15 typical California homes (Brown and Koomey 2003), but the servers are concentrated in a 2' x 3.5' x 6' space. All of this electric power is converted to heat. As a result, each of these racks of servers is expected to require an additional 20-25 kW of power for the cooling and power conversion equipment that supports it.

There are many IT equipment configurations in a data center room. In some data centers (especially smaller ones), IT equipment can be placed haphazardly throughout the room. One common best-practice configuration is shown in Figure 1-4, in which the racks are in alternating aisles, called a hot aisle/cold aisle layout, with the hot air removed overhead. The IT equipment is mounted in racks that are positioned together in long rows. The racks are placed on a raised floor, which delivers conditioned air. In fact, it is standard terminology to refer to the computer room floor area as the *raised floor area* (even though some computer rooms don't have raised floors).

Figure 1-4. Typical Data Center HVAC Hot Aisle / Cold Aisle Layout



Source: ASHRAE (2004)

⁵ ASHRAE is a professional association for the heating, ventilation, air conditioning, and refrigeration industry.

1.4. Energy Efficiency

The move to increase energy efficiency – to reduce primary energy input for providing services – has found its way into the American mainstream because energy efficiency offers many benefits: reduced energy bills, increased reliability and grid support for the entire electric utility system, avoided capital investment for power plants, reduced dependence on fossil fuels, and avoided emissions of both criteria air pollutants and greenhouse gases. These benefits would accrue from increasing the energy efficiency of data centers as well.

Until recently, data center designers and operators worried primarily about data center reliability, with little or no focus on energy efficiency. Now that data center power density is leading to power and cooling limitations, there is a growing interest in energy efficiency as a potential solution to these problems. As subsequent chapters of this report explain in greater detail, there are many energy efficiency opportunities in data centers, in the IT equipment as well as the supporting power and cooling equipment.

1.5. Purpose of this Report

The report presents findings of an analysis of the rapid growth and energy consumption of U.S. data centers operated by both the federal government and private enterprise. The report fulfills the following analyses called for in Public Law 109-431:

- (1) an overview of the growth trends associated with data centers and the utilization of servers in the federal government and private sector;
- (2) an analysis of industry migration to the use of energy-efficient microchips and servers designed to provide energy-efficient computing and reduce the costs associated with constructing, operating, and maintaining large- and medium-scale data centers;
- (3) an analysis of the potential savings to the federal government, large institutional data center operators, private enterprise, and consumers from the increasing the energy efficiency of data centers and servers;
- (4) an analysis of the potential savings and benefits to the energy supply chain from increasing the energy efficiency of data centers and servers, including reduced energy demand, enhanced capacity, and reduced strain on existing grid infrastructure, and consideration of secondary benefits, including potential impact of related advantages associated with substantial domestic energy savings;
- (5) an analysis of the potential impacts of energy efficiency on product performance, including computing functionality, reliability, speed, and features, and overall cost;
- (6) an analysis of the potential savings and benefits to the energy supply chain from the use of stationary fuel cells for backup power and distributed generation;
- (7) an overview of current government incentives offered for energy-efficient products and services and consideration of similar incentives to encourage the increased energy efficiency in data centers and servers; and

(8) recommendations regarding potential incentives and voluntary programs that could advance the energy efficiency of data centers and computing.

Public Law 109-431 also directs the EPA to provide “a meaningful opportunity for interested stakeholders, including affected industry stakeholders and energy efficiency advocates, to provide comments, data, and other information on the scope, contents, and conclusions of the study.”

To develop this report, EPA convened a study team led by researchers from the Lawrence Berkeley National Laboratory. The study team offered stakeholders multiple opportunities to give input to and review this report, including:

- conducting preliminary calls with key stakeholders to help plan the study;
- holding a public workshop on February 16, 2007 (attended by approximately 130 people) to solicit input on the topic of energy efficiency in servers and data centers;
- following up on workshop attendees’ offers of assistance, to gather and refine information for the study;
- posting on the ENERGY STAR web site an open call for interested parties to submit information, as well as a list of data needs;
- posting on the ENERGY STAR web site a public review draft of this report; and
- incorporating into the final version of this report comments on the public review draft from more than 50 organizations and individuals.

2. Trends in Growth and Energy Use Associated with Servers and Data Centers in the U.S.

This chapter addresses the Public Law 109-431 requirement for:

- An overview of the growth trends associated with data centers and the utilization of servers in the federal government and private sectors.

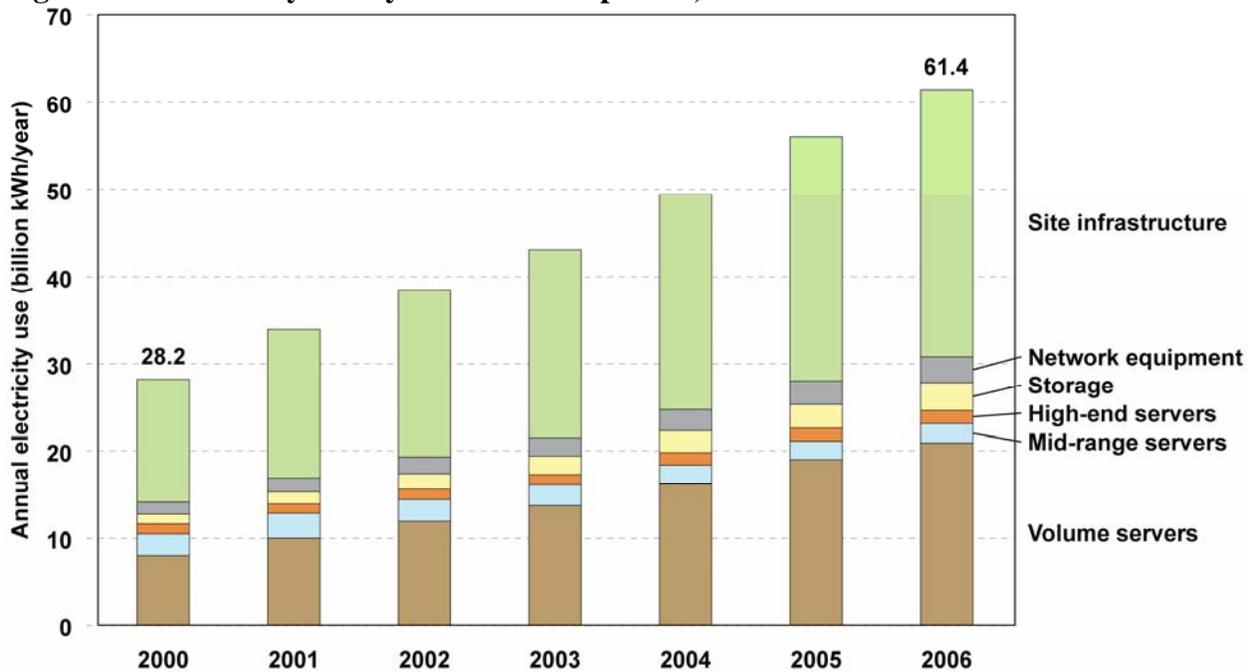
This chapter gives a brief overview of current trends in demand for data processing, exchange, and storage that are driving the rapid growth of data centers and server utilization in the U.S. It also estimates total electricity use of the nation's servers and data centers for the period 2000 to 2006 to illustrate the implications of this growth for the U.S. energy supply system.

Electricity use associated with the nation's servers and data centers grew significantly from 2000 to 2006. As of 2006, the electricity use attributable to the nation's servers and data centers is estimated at about 61 billion kilowatt-hours (kWh), or 1.5 percent of total U.S. electricity consumption (US DOE 2007a). This electricity use has more than doubled since 2000 and amounts to about \$4.5 billion in electricity costs. It is equivalent to the electricity consumed by 5.8 million average U.S. households (which represent 5% of the U.S. housing stock) and is similar to the amount of electricity used by the entire U.S. transportation manufacturing industry (which includes the manufacture of automobiles, aircraft, trucks, and ships) (US Census Bureau 2006, US DOE 2005).

It is estimated that federal servers and data centers accounted for roughly 6.1 billion kWh (10%) of this electricity use in 2006 at a total cost of \$450 million.

One server class—volume servers—was responsible for the majority (68%) of the electricity used by U.S. servers and related IT equipment (i.e., storage devices and network equipment) in 2006, as shown in Figure 2-1. Volume servers also experienced the greatest growth in energy use among all server classes, more than doubling from 2000 to 2006 at a compound annual growth rate (CAGR) of 17%. Infrastructure systems necessary to support the operation of IT equipment (i.e., power delivery and cooling systems) also consumed a significant amount of energy, comprising 50% of total annual electricity use.

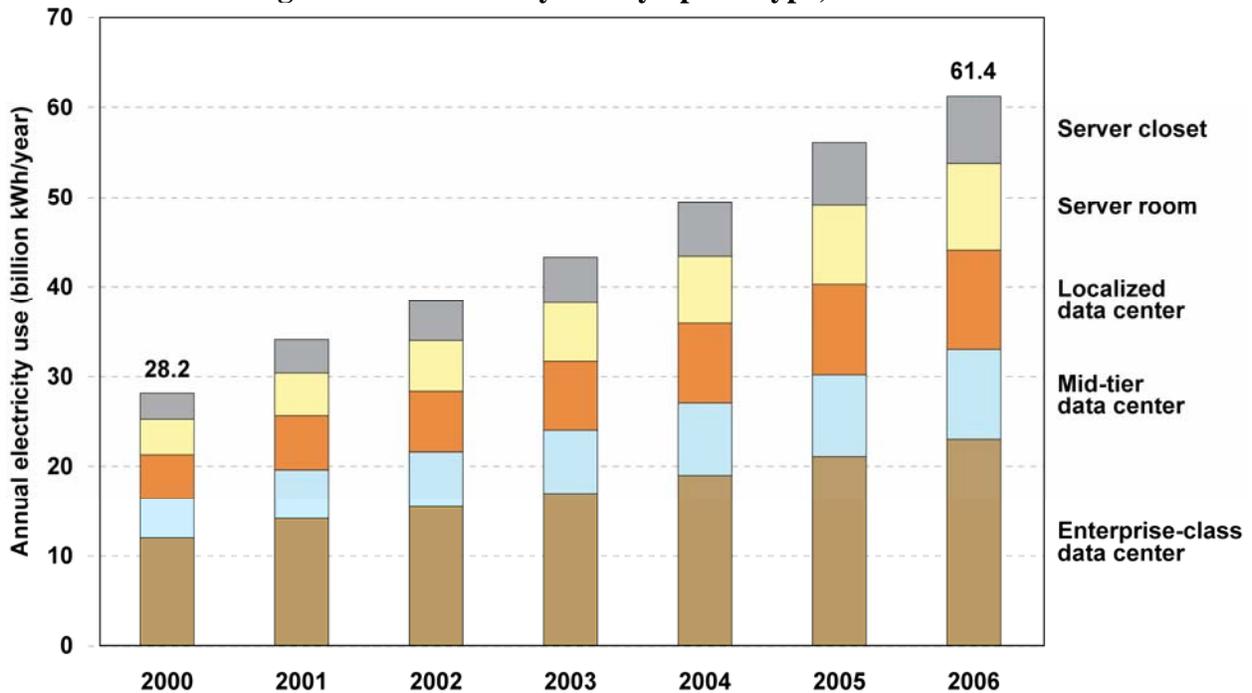
Figure 2-1. Electricity Use by End-Use Component, 2000 to 2006



End use component	2000		2006		2000 – 2006 electricity use CAGR
	Electricity use (billion kWh)	% Total	Electricity use (billion kWh)	% Total	
Site infrastructure	14.1	50%	30.7	50%	14%
Network equipment	1.4	5%	3.0	5%	14%
Storage	1.1	4%	3.2	5%	20%
High-end servers	1.1	4%	1.5	2%	5%
Mid-range servers	2.5	9%	2.2	4%	-2%
Volume servers	8.0	29%	20.9	34%	17%
Total	28.2		61.4		14%

As Figure 2-2 shows, more than one-third (38 percent) of the electricity use is attributable to the nation’s largest (i.e., enterprise-class) data centers.

Figure 2-2. Electricity Use by Space Type, 2000 to 2006



Space type	2000		2006		2000 – 2006 electricity use CAGR
	Electricity use (billion kWh)	% Total	Electricity use (billion kWh)	% Total	
Server closet	3.0	11%	7.5	12%	17%
Server room	3.9	14%	9.7	16%	16%
Localized data center	4.9	17%	11.1	18%	15%
Mid-tier data center	4.4	16%	10.0	16%	15%
Enterprise-class data center	12.0	43%	23.0	38%	12%
Total	28.2		61.4		14%

Section 2.2 gives a brief overview of the key trends driving this rapid growth. Section 2.3 summarizes the methods and data assumptions used to develop the estimates in Figures 2-1 and 2-2. Energy-efficiency opportunities for servers and data centers are explored in detail in the remainder of this report.

2.1. Overview of Data Center Growth Trends

The U.S. data center industry is in the midst of a major growth period. The increasing reliance on digital data in our society is driving a rapid increase in the number and size of data centers. This growth is the result of several factors, including growth in the use of internet media and communications, ongoing digital conversion of business applications, establishment of new regulations that require retention of digital records, and requirements related to disaster recovery.

Internet usage is increasing at approximately 10 percent per year worldwide (comScore Networks 2007), leading to a 13-20 percent CAGR in internet and hosting data centers (Wong 2007). This growth is driven primarily by increased popularity of music downloads, video on demand, on-line gaming, e-commerce, social networking interfaces, and voice over internet

protocol (VOIP) communication. For instance, between 2005 and 2010, VOIP is projected to grow at approximately a 33 percent CAGR (Telegeography 2006).

IT is an increasingly important factor in the growth of businesses, which leads to increased demand for and investment in key business applications and functions. Bailey et al. (2007) cite several key trends that are driving the need for increased data services:

- Healthcare moving to electronic medical records,
- Manufacturing moving to global networked organizations,
- Banking migrating away from paper-based business models (e.g., online banking),
- Financial services moving to digital transactions,
- Insurance database needs growing,
- Retail moving toward real-time inventories and supply chain management, and
- Transportation moving toward global positioning system (GPS) navigation and radio-frequency identification (RFID) tracking.

The importance to business of the electronic applications involved in the trends listed above requires that they be hosted in highly reliable data centers with sufficient capacity to meet peak and growing loads. Many of these applications require not just business processing, which drives computing demand, but also large amounts of data storage, which drives the demand for storage equipment in data centers.

In addition to application needs that are generated by “internal” business processes, companies are increasingly subject to regulations that require the collection and storage of digital information. The most well-known of these regulations is the Sarbanes-Oxley Act, which requires long-term storage of financial information, including electronic records such as email (Apiki 2005). In some industries, it is estimated that the number of records that must be retained is growing at a CAGR of 50 percent or greater (Warmenhoven 2005).

Disaster recovery needs are also strong motivators for increased storage (e.g., duplicate data sets) and redundant data center equipment and facilities (Bailey et al. 2007).

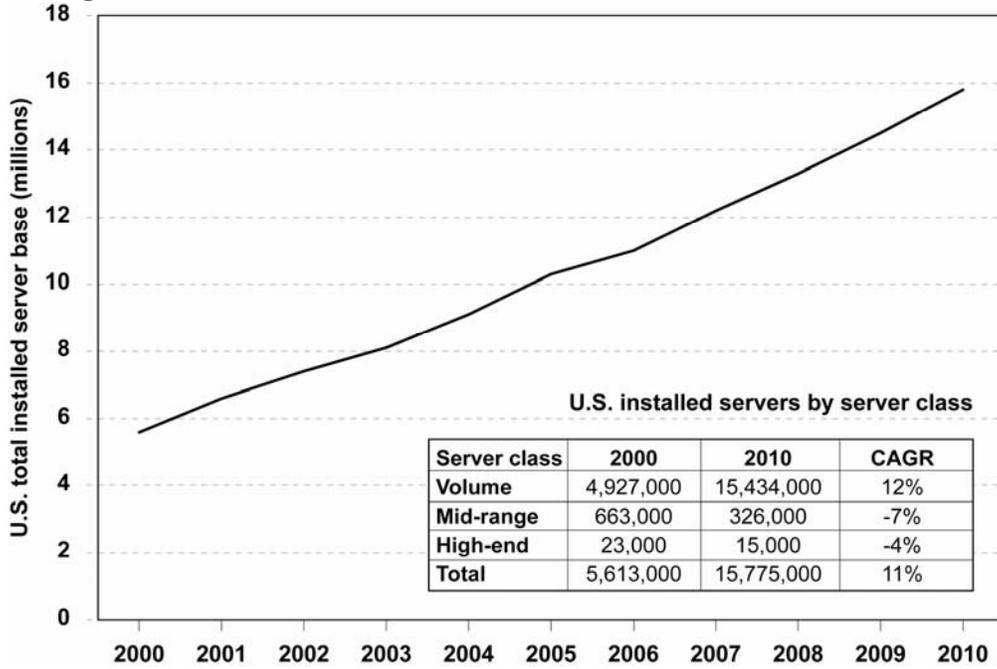
Federal, state, and local governments are also subject to many of the above trends, as well as the following:

- Increased website hosting for public information, online reports, and information;
- Increased movement toward digital services (e.g., e-filing of taxes, on-line tracking of items sent through the U.S. Postal Service);
- Requirements related to homeland security, from e-passports to cyber-security; and
- Scientific computing in the national laboratories and other government research institutions.

The above drivers are causing significant growth in the shipments of servers and related IT hardware. Figure 2-3 shows that the total number of installed servers in the U.S. is expected to grow to around 15.8 million by 2010, which is nearly three times the number of installed servers in 2000 (IDC 2007b). Increasing demand for data storage is also driving rapid growth in storage technologies, including storage area networks, network-attached storage, and external hard disk drive (HDD) arrays.

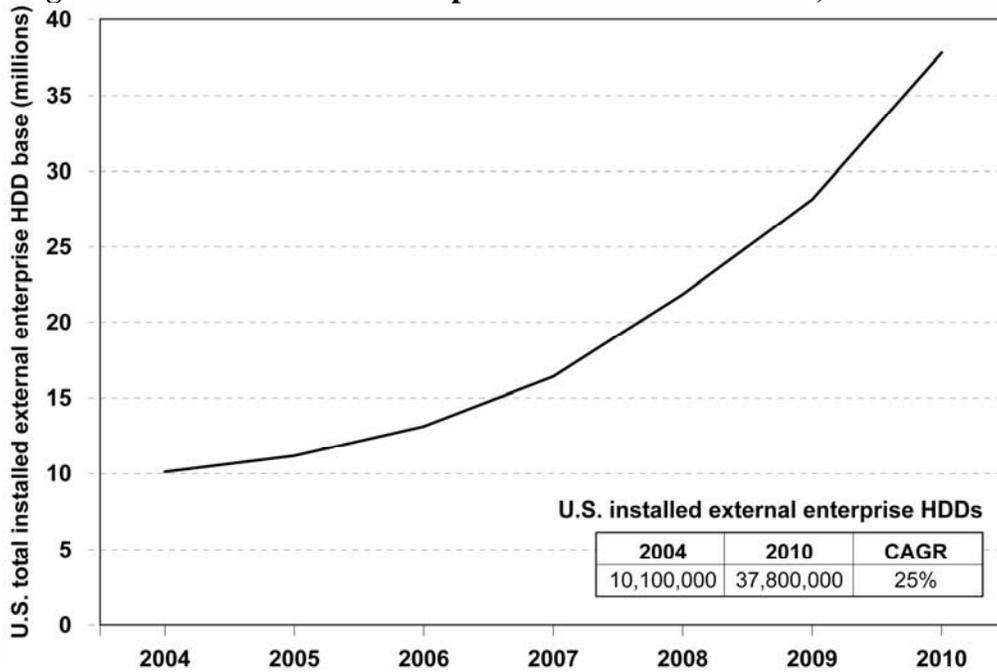
For example, as depicted in Figure 2-4, the number of installed enterprise (i.e., external) HDD storage devices is expected to nearly quadruple from 2004 to 2010 (Osterberg 2007).

Figure 2-3. U.S. Installed Server Base Growth Trends, 2000 to 2010



Source: IDC (2007b)

Figure 2-4. U.S. Installed Enterprise HDD Growth Trends, 2004 to 2010



Source: Osterberg (2007)

In addition the installed base of hardware, the number and floor area of data centers is expected to grow, although at a lower annual growth rate than hardware shipments. Data center floor space is expected to grow at five to ten percent per year (Kumar 2006, Wong 2007), which is lower than the growth rate for IT hardware shipments for several reasons.

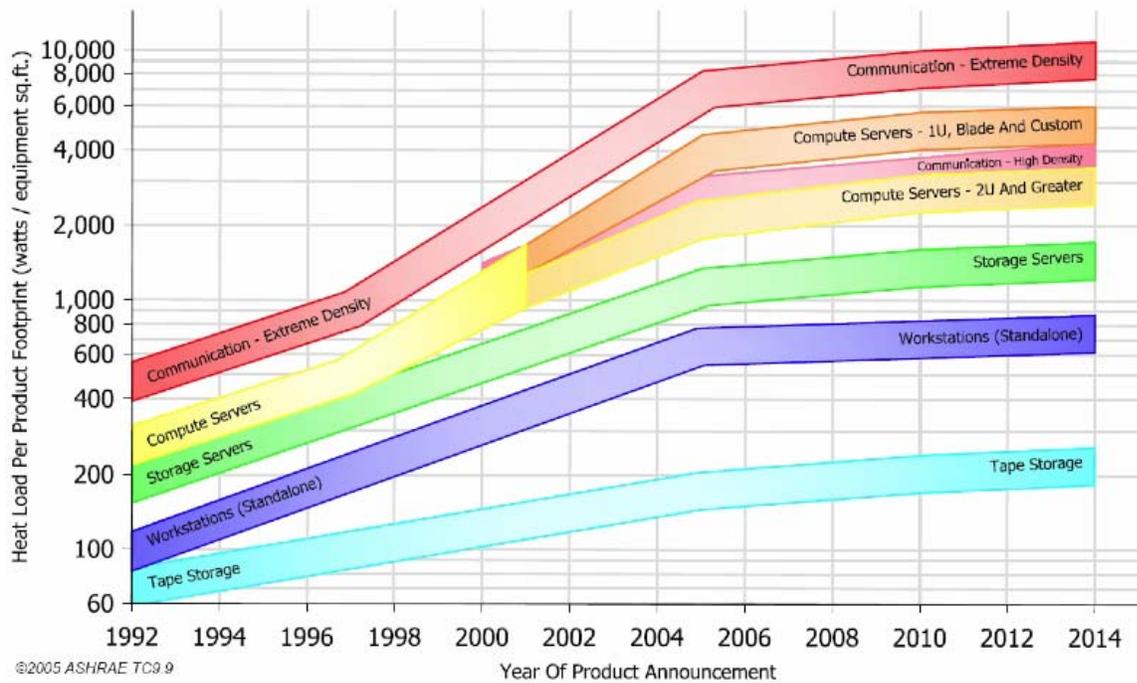
First, there is a general trend toward consolidating many smaller data centers into one or a few large data centers (Antonopoulos 2006, Bailey et al. 2007, Ranganathan and Jouppi 2005). These centralized data centers can support higher equipment densities and more effectively use available floor space than a number of smaller centers can. Second, computing hardware is increasingly moving to space-efficient form factors such as 1U rack servers and blade servers, which allow higher computing density in a given floor space. Blade server shipments are increasing at a 20-30 percent CAGR, which is significantly higher than the shipment rate for the overall server market (IDC 2007d). Finally, the number and floor area of data centers are growing more slowly than computing demand because of a move toward “virtualization,” which allows IT demand to be met through fewer physical servers (IDC 2007a) and storage devices. Although virtualization technology is becoming more common, the pace of its adoption is difficult to predict because it can present management challenges that may limit its applicability in certain data center environments (Dubie 2007).

The above trends toward increasing server density are resulting in higher power densities in many data centers. Figure 2-5 shows recent estimates that suggest that power densities in various data center types have grown remarkably in recent years and will continue to grow for the foreseeable future (Belady 2007). Note that the vertical axis of Figure 2-5 is a logarithmic scale, meaning that power density has been growing exponentially until very recently. The impact of these dramatically higher power densities is that, in many cases, power and cooling capacity are the primary constraints to expansion of computational capacity within a data center. As a result, data center managers must either invest in upgrading the power and cooling infrastructure of existing data centers or build new facilities; either choice requires significant capital investment.

Another impact of higher energy densities is that server hardware is no longer the primary cost component of a data center. Figure 2-6 shows that the purchase price of a new (1U) server has been exceeded by the capital cost of power and cooling infrastructure to support that server and will soon be exceeded by the lifetime energy costs alone for that server (Belady 2007). This represents a significant shift in data center economics that threatens to overwhelm the advances in chip efficiency that have driven the growth of digital information during the past 30 years.

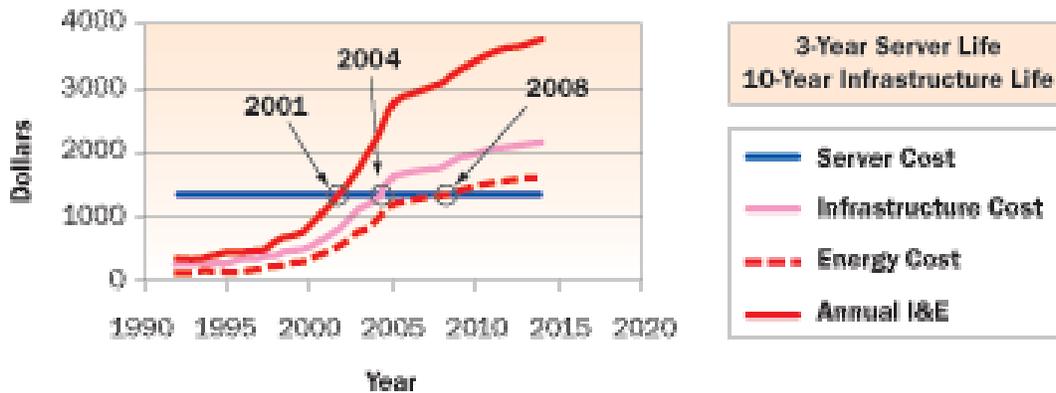
These trends are a significant motivation for the data center industry’s current interest in energy efficiency, which is explored further in the next chapter.

Figure 2-5. ASHRAE TC 9.9 Equipment Power Density Projections



Source: ASHRAE (2005).

Figure 2-6. Annual Amortized Cost for Purchase and Operation of a 1U Server



Source: Belady (2007). Note: I&E is infrastructure + energy cost.

2.2. Estimates of U.S. Server and Data Center Energy Use

As the number of installed servers has grown, the total energy use attributable to servers and data centers has also grown significantly from 2000 to 2006. This growth has had important implications for U.S. electricity use, as shown by the estimates in Figures 2-1 and 2-2.

The estimates in Figures 2-1 and 2-2 were compiled using the following information to provide as complete a picture as possible of server and data center energy use:

- Estimates from 2005 of the energy use of U.S. servers and the energy required for server cooling and auxiliary systems, based on Koomey (2007);
- Information on the growth of the U.S. installed server base, to extend the 2005 energy use estimates to 2006;
- Disaggregation of the U.S. installed server base into five different space types (server closets, server rooms, localized data centers, mid-tier data centers, and enterprise-class data centers); and
- Information on the energy use of storage devices and network equipment not included in Koomey (2007).

The estimates described in this section were based on best available information and data at the time of this study. Nonetheless, there are inherent uncertainties associated with the data and assumptions employed in this study, so the estimates presented in this section should be regarded as preliminary. Recommendations for future work to reduce the uncertainties associated with these estimates are presented in Chapter 8.

2.2.1. Estimates of U.S. Server Energy Use by Space Type

Koomey (2007) estimates that, in 2005, the electricity use of the nation's servers and associated cooling and auxiliary infrastructure was roughly 45 billion kilowatt-hours (kWh), or 1.2 percent of total U.S. electricity consumption. The Koomey (2007) study methods and results form a solid foundation from which to estimate 2006 energy use attributable to U.S. servers and data centers. Koomey (2007) uses data on the number of volume, mid-range, and high-end servers installed in the U.S. from the market research firm IDC (IDC 2007b), coupled with measured data and estimates of power used by the most popular models in each server class.⁶ Figure 2-7 shows a basic schematic of this modeling approach. Further details on this approach and its key assumptions may be found in Koomey (2007).

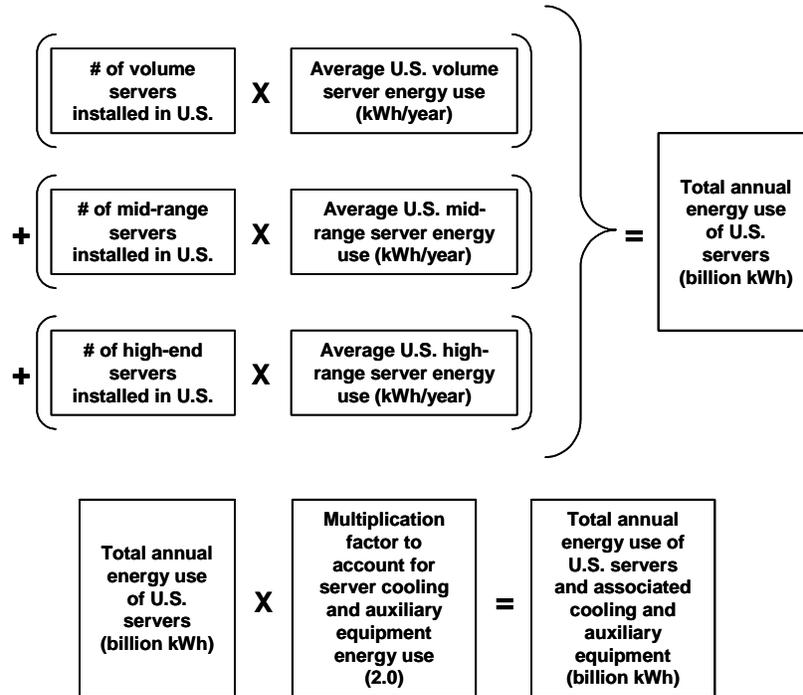
The installed base of the three major server classes in the U.S. was disaggregated into five different space types: server closets, server rooms, localized data centers, mid-tier data centers, and enterprise-class data centers. These types of spaces are defined by IDC (Bailey et al. 2007). Assumptions about the major differences among these five types of spaces are listed in Table 2-1.

This disaggregation was undertaken because a significant fraction of U.S. servers were expected to be located in server closets and server rooms, which can have significantly different IT equipment and infrastructure characteristics than larger data centers (as summarized in Table 2-1). Determining the distribution of different servers and types of spaces allowed for more accurate characterization of the total energy use associated with different server environments. It also allowed better characterization of energy costs because most server closets, server rooms, and localized data centers were expected to be subject to commercial electricity rates whereas

⁶ IDC defines volume servers as those costing less than \$25,000 per unit, mid-range servers as those costing between \$25,000 and \$500,000 per unit, and high-end servers as those costing more than \$500,000 per unit.

larger mid-tier and enterprise-class data centers were expected to be subject to industrial electricity rates. Finally, the disaggregation provided better information for assessing energy efficiency opportunities (see Chapter 3).

Figure 2-7. Schematic of Koomey (2007) Modeling Approach



The installed U.S. server base for the years 2000 to 2006 U.S. was disaggregated by space type by:

- Deriving a distribution of the 2005 installed U.S. server base by server class and space type, based on 2005 data from IDC on U.S. installed servers by CPU type and space type (Bailey et al. 2007, IDC 2007b). Table 2-2 summarizes this distribution.
- Applying the distribution in Table 2-2 to the total installed server base estimates for all years 2000 to 2006.

Table 2-1. Typical IT Equipment and Site Infrastructure System Characteristics, by Space Type

Space type	Typical size	Typical IT equipment characteristics	Typical site infrastructure system characteristics
Server closet	<200 ft ²	1-2 servers No external storage	Typically conditioned through an office HVAC system. To support VOIP and wireless applications, UPS and DC power systems are sometimes included in server closets. Environmental conditions are not as tightly maintained as for other data center types. HVAC energy efficiency associated with server closets is probably similar to the efficiency of office HVAC systems.
Server room	<500 ft ²	A few to dozens of servers No external storage	Typically conditioned through an office HVAC system, with additional cooling capacity, probably in the form of a split system specifically designed to condition the room. The cooling system and UPS equipment are typically of average or low efficiency because there is no economy of scale to make efficient systems more first-cost competitive.
Localized data center	<1,000 ft ²	Dozens to hundreds of servers Moderate external storage	Typically use under-floor or overhead air distribution systems and a few in-room CRAC units. CRAC units in localized data centers are more likely to be air cooled and have constant-speed fans and are thus relatively low efficiency. Operational staff is likely to be minimal, which makes it likely that equipment orientation and airflow management are not optimized. Air temperature and humidity are tightly monitored. However, power and cooling redundancy reduce overall system efficiency.
Mid-tier data center	<5,000 ft ²	Hundreds of servers Extensive external storage	Typically use under-floor air distribution and in-room CRAC units. The larger size of the center relative to those listed above increases the probability that efficient cooling, e.g., a central chilled water plant and central air handling units with variable speed fans, is used. Staff at this size data center may be aware of equipment orientation and airflow management best practices. However, power and cooling redundancy may reduce overall system efficiency.
Enterprise-class data center	5,000+ ft ²	Hundreds to thousands of servers Extensive external storage	The most efficient equipment is expected to be found in these large data centers. Along with efficient cooling, these data centers may have energy management systems. Equipment orientation and airflow management best practices are most likely implemented. However, enterprise-class data centers are designed with maximum redundancy, which can reduce the benefits gained from the operational and technological efficiency measures.

Source: Derived from Bailey et al. (2007)

Table 2-2. Estimated Distribution of U.S. Installed Servers, by Server Class and Space Type, 2005

Server class	% of servers installed in:					Total
	Server closets	Server rooms	Data centers			
			Localized	Mid-tier	Enterprise-class	
Volume	17%	20%	17%	15%	30%	100%
Mid-range	0%	5%	16%	14%	65%	100%
High-end	0%	0%	16%	14%	71%	100%

Source: Derived from Bailey et al. (2007) and IDC (2007b)

The resulting estimates of the number of servers installed in the U.S. by server class and space type are shown in Table 2-3. These estimates should be viewed as rough approximations for several reasons:

- The IDC 2005 data on the number U.S. servers by space type were the only such spatially disaggregated data available and could not be verified using other data sources.
- No such data were available for other years, so the 2005 distribution data in Table 2-2 had to be applied for all years 2000 to 2006.

Table 2-3. Estimated U.S. Installed Server Base (in 1,000s), by Server Class and Space Type, 2000 to 2006

Volume servers in:	2000	2001	2002	2003	2004	2005	2006
Server closets	861	1,032	1,182	1,324	1,513	1,729	1,798
Server rooms	980	1,175	1,346	1,507	1,722	1,968	2,120
Localized data centers	841	1,008	1,155	1,293	1,478	1,689	1,820
Mid-tier data centers	759	911	1,043	1,168	1,335	1,526	1,643
Enterprise-class data centers	1,486	1,781	2,041	2,285	2,611	2,984	3,215
Total volume	4,927	5,907	6,768	7,578	8,658	9,897	10,597
Mid-range servers in:	2000	2001	2002	2003	2004	2005	2006
Server closets	0	0	0	0	0	0	0
Server rooms	33	35	29	27	22	19	18
Localized data centers	105	111	91	84	68	61	58
Mid-tier data centers	94	99	81	75	61	55	52
Enterprise-class data centers	431	456	373	345	281	252	239
Total mid-range	663	701	574	530	432	387	367
High-end servers in:	2000	2001	2002	2003	2004	2005	2006
Server closets	0	0	0	0	0	0	0
Server rooms	0	0	0	0	0	0	0
Localized data centers	4	4	4	3	4	3	3
Mid-tier data centers	3	3	3	3	3	3	3
Enterprise-class data centers	16	16	16	15	16	16	15
Total high-end	23	22	23	21	23	22	21

- The data in Table 2-3 do not include custom servers used by large internet companies (e.g., Google) because no data were found on the number of custom servers in the U.S. installed server base. However, custom servers likely represent only a small fraction of the total number of U.S. servers as of 2006 (Kooimey 2007), so this omission should not

have a significant impact on the energy use estimates in this chapter.⁷

The energy use of U.S. servers by space type for the period 2000 to 2006 was estimated by applying average unit energy consumption (UEC) values for each server class. Table 2-4 lists these values. The average UEC values in Table 2-4 were derived using the data on average power use per server from Koomey (2007) listed in Table 2-5.⁸ The 2006 values in Tables 2-4 and 2-5 were estimated using the 2000 to 2005 CAGRs for each server class. Table 2-6 summarizes the resulting estimates of total U.S. electricity use by server class and space type from 2000 to 2006.

Table 2-4. Estimated Average UEC (kWh/year) per Server, by Server Class, 2000 to 2006

Server class	2000	2001	2002	2003	2004	2005	2006
Volume	1,634	1,691	1,752	1,813	1,871	1,918	1,969
Mid-range	3,724	4,006	4,298	4,590	5,042	5,475	5,917
High-end	48,611	51,088	53,699	56,309	61,251	67,023	71,509

Source: Derived from Koomey (2007)

Table 2-5. Estimated Average Power Use (W) per Server, by Server Class, 2000 to 2006

Server class	2000	2001	2002	2003	2004	2005	2006
Volume	186	193	200	207	213	219	225
Mid-range	424	457	491	524	574	625	675
High-end	5,534	5,832	6,130	6,428	6,973	7,651	8,163

Source: Derived from Koomey (2007)

**Table 2-6. Estimated Total Electricity Use of U.S. Servers
(in billion kWh/year) by Space Type, 2000 to 2006**

Space type	2000	2001	2002	2003	2004	2005	2006
Server closet	1.4	1.7	2.1	2.4	2.8	3.3	3.5
Server room	1.7	2.1	2.5	2.9	3.3	3.9	4.3
Localized data center	1.9	2.3	2.6	2.9	3.3	3.8	4.2
Mid-tier data center	1.7	2.1	2.3	2.6	3.0	3.4	3.7
Enterprise-class data center	4.8	5.6	6.1	6.6	7.3	8.2	8.8
Total	11.6	13.9	15.6	17.4	19.8	22.6	24.5

⁷ Given the rapid growth of custom server installations by companies such as Google, the contribution of custom servers to the total energy use of U.S. data centers may increase significantly in the future. Chapter 8 recommends further research to account for the growth and energy use of custom server installations.

⁸ The UEC values in Table 2-4 were calculated by multiplying the average power use per server data in Table 2-5 by 8,760 hours (for each of the following years: 2001, 2002, 2003, 2005, and 2006) or by 8,784 hours (for each of the leap years 2000 and 2004).

2.2.2. Estimates of Energy Use by Storage Devices and Network Equipment

This report estimates the energy used by storage devices and network equipment, which was outside the scope of the Koomey (2007) study, to present as complete a picture as possible of the total energy used by IT equipment and associated power delivery and cooling systems.

The energy use of enterprise storage devices can vary widely depending on the need for external storage in a particular data center. Estimates of the energy use of enterprise storage devices were developed using data on the installed U.S. base and energy use of external (i.e., enterprise) HDD storage devices (Osterberg 2007). Table 2-7 summarizes these estimates.

Table 2-7. Estimated Energy Use (billion kWh/year) of Enterprise Storage Devices, by Space Type, 2000 to 2006

Space type	2000	2001	2002	2003	2004	2005	2006
Server closet	0	0	0	0	0	0	0
Server room	0	0	0	0	0	0	0
Localized data center	0.28	0.37	0.46	0.56	0.66	0.73	0.86
Mid-tier data center	0.25	0.33	0.42	0.50	0.59	0.66	0.78
Enterprise-class data center	0.57	0.74	0.90	1.07	1.23	1.35	1.58
Total	1.10	1.44	1.79	2.13	2.49	2.74	3.22

Source: Derived from Osterberg (2007)

The estimates in Table 2-7 were made based on the assumptions that:

- The number of external storage devices installed in U.S. server closets and server rooms is negligible as the necessary storage capacity in these space types is typically provided by internal server HDDs
- Total enterprise storage system energy use (HDD storage devices, power supply losses, controllers, etc.) would be twice the energy use of HDD storage devices alone, based on estimates provided by storage industry experts (Osterberg 2007, Winkler 2007).
- Total enterprise storage system energy use could be allocated across localized, mid-tier, and enterprise-class data centers in a proportional manner, based on the installed number of servers in each respective space type, as shown in Table 2-3.

The estimates in Table 2-7 should be viewed as rough approximations because they were based on a single data set (the only data available) for HDD storage devices and did not include the energy use associated with tape storage devices because no data on these devices were available. However, the energy use of U.S. enterprise tape storage devices is expected to be small compared to the energy use of U.S. enterprise HDD storage devices (Winkler 2007), so this omission should not significantly impact the estimates shown in Table 2-7.

Publicly available data on the average energy use of network equipment in data centers and other server installations in the U.S. are scarce. Estimates of the energy use of network equipment in server rooms, localized data centers, mid-tier data centers, and enterprise-class data centers were made using following assumptions, which were based on conversations with industry experts:

- Current network equipment consumes on average eight watts per port.

- The typical server room and data center will have, on average, three installed network ports per installed volume server.

Table 2-8 shows the resulting estimates of network equipment energy use. The estimates for server rooms, localized data centers, mid-tier data centers, and enterprise-class data centers are in close agreement with IT equipment energy use breakdown data for typical data centers (Pouchet 2007).

Table 2-8. Estimated Energy Use (billion kWh/year) of Network Equipment, by Space Type, 2000 to 2006

Space type	2000	2001	2002	2003	2004	2005	2006
Server closet	0.07	0.09	0.11	0.13	0.15	0.17	0.19
Server room	0.22	0.27	0.32	0.37	0.44	0.51	0.57
Localized data center	0.25	0.30	0.34	0.39	0.44	0.50	0.56
Mid-tier data center	0.22	0.27	0.31	0.35	0.40	0.45	0.50
Enterprise-class data center	0.60	0.71	0.77	0.85	0.95	1.06	1.15
Total	1.36	1.64	1.85	2.08	2.38	2.70	2.97

The estimates for the energy used by network equipment in server closets in Table 2-8 were made using expert judgment based on the expectation that fewer ports (one to two per server) would be required in server closets because these spaces are typically designed for small workgroup support (e.g., file and print server applications). This estimate excludes the network equipment energy use attributable to office equipment (e.g., personal computers) that is typically connected to network equipment in server closets.

Although approximate, the values in Tables 2-7 and 2-8 are first-order estimates of the energy used by U.S. enterprise storage devices and network equipment and help to paint a complete picture of the total energy use of IT equipment in server installations in the U.S.

The data in Tables 2-6, 2-7, and 2-8 suggest that, across the total population of U.S. localized, mid-tier, and enterprise-class data centers, servers will on average account for about 75 percent of total IT equipment energy use, storage devices will account for around 15 percent, and network equipment will account for around 10 percent. However, these percentages can vary greatly depending on the configuration, hosted applications, and data storage requirements of individual data centers (AFCOM 2007, Dietrich 2007, Emerson Network Power 2007, Pflueger and Hanson 2007).

2.2.3. Estimates of Energy Use by Site Infrastructure Systems

Koomey (2007) assumed an average value of 2.0 for the ratio of total data center energy use to total IT equipment energy use to account for the energy use of site infrastructure systems.⁹ This

⁹ Total IT equipment energy use is defined in this study as the combined energy use of servers, storage devices, and network equipment.

ratio is often referred to as the “power usage effectiveness” (PUE) ratio.¹⁰ The PUE ratio estimate of 2.0 was based on recent energy use benchmarking studies of 22 data centers performed by Lawrence Berkeley National Laboratory (LBNL) (Greenberg et al. 2006, Tschudi et al. 2003). The majority of data centers in the benchmarking study were located on the west coast; similarly comprehensive benchmark data from other regions of the U.S. were not available. However, climate is expected to significantly affect infrastructure efficiency only in data centers with economizers, which is currently only a small portion of U.S. data centers. A PUE ratio of 2.0 was assumed to be the average value across all U.S. data centers; however, the PUE ratio can vary widely among individual data centers depending on infrastructure equipment configurations and efficiencies, time of year, and local climate (AFCOM 2007, Emerson Network Power 2007). Despite this variation, the LBNL data are the most accurate measured data, so they were used in this study.

A PUE ratio of 2.0 was applied for all space types to estimate the total energy use attributable to U.S. servers and data centers for the period 2000 to 2006. Although attention is currently being paid to the efficiency of infrastructure systems because of rising electricity costs and increasing data center heat densities, this trend is considered to be mostly a recent phenomenon. Therefore, a PUE ratio of 2.0 should be valid for the past seven years because the average equipment types and efficiencies associated with infrastructure systems can be expected to have been reasonably constant over this time period (US EPA 2007).

2.2.4. Estimates of Electricity Costs

Based on the estimates of electricity use attributable to U.S. servers and data centers by space type, shown in Figure 2-2, the 2006 electricity bill for the nation’s servers and data centers was calculated at approximately \$4.5 billion. This estimate assumes that server closets, server rooms, and localized data centers would be subject to the average 2006 U.S. commercial electricity rate (8.8 cents/kWh) and that mid-tier and enterprise-class data centers would be subject to the average 2006 U.S. industrial electricity rate (6.2 cents/kWh) (US DOE 2007a).

2.3. Energy Use Associated with Federal Government Servers and Data Centers

Public Law 109-431 requests information on federal data centers and servers in the U.S. Based on the limited information available, it was estimated that federal servers and data centers account for about 10 percent of the nation’s server and data center electricity use. This estimate corresponds to roughly 6.1 billion kWh of electricity use at a total cost of \$450 million in 2006.

Because no data could be found in the public domain on the number of servers or data centers operated by (or for) the federal government, our estimates were based on interviews with and advice from industry experts. Interviews were conducted with major U.S. manufacturers of servers and microprocessors to gather information on which to base an estimate the percentage of total server sales to the federal government each year. A sufficient number of responses were obtained to estimate that server shipments to the federal government currently represent approximately five to 10 percent of annual U.S. server shipments.

¹⁰ The reciprocal of the PUE ratio is also often used to describe data center efficiency. See Chapter 7 for more discussion on the PUE ratio and other metrics for characterizing data center energy use.

This range does not capture shipments of servers to facilities that are managed by contractors for the federal government. Additionally, no data on federal server shipments by server class (i.e., volume, mid-range, and high-end) were available. It is possible that the federal government accounts for a significant fraction of high-end server electricity use, given that 28 of the 100 largest supercomputers in the world are housed in U.S. government-owned facilities, including the two largest as well as six out of the top ten (Top500 2006).

The estimate that federal server and data center operations account for 10 percent of the electricity use and electricity costs attributable to servers and data centers in the U.S. is at the high end of the range of estimated annual U.S. server shipments provided by industry experts. The high end of the range was used in an attempt to account for servers shipped to government contractors that are not counted as federal shipments. If this 10 percent estimate is applied across all space types (federal data about space types are not available), the federal share of server and data center energy use could be further estimated on a space-by-space basis.

3. Potential Energy and Cost Savings through Improved Energy Efficiency

This chapter addresses the Public Law 109-431 requirements for:

- Analysis of industry shift to the use of energy-efficient microchips and servers designed to provide energy-efficient computing and reduce the costs associated with constructing, operating, and maintaining large- and medium-scale data centers.
- Analysis of the potential savings to the federal government, large institutional data center operators, private enterprise, and consumers from increasing the energy efficiency of data centers and servers.

This chapter presents current industry trends toward improved energy efficiency and projections of the potential savings in electricity use and costs that these trends will likely yield for the federal government and other key sectors through the year 2011. It also provides estimates of additional savings possible beyond those from these current trends for the same time period, for several energy-efficiency improvement scenarios.

3.1. *Expected Energy Savings from Current Energy Efficiency Trends*

A number of energy efficiency trends are currently underway that will likely reduce the energy used by U.S. servers and data centers in the near future. These trends were analyzed to predict likely energy use during the next five years if no additional steps are taken to increase the energy efficiency of data centers and servers. A five-year forecast window (i.e., 2007 to 2011) was chosen because the technologies and applications associated with servers and data centers can change rapidly, so a forecast over a longer period would inherently be less accurate.

The subsections below present:

- A review of current energy-efficiency trends related to servers, IT equipment, and data centers,
- Estimates of future electricity use and costs during the next five years based on these current energy-efficiency trends, and
- Estimates of the expected savings in electricity use and costs during the next five years that can be attributed to current energy-efficiency trends.

3.1.1. Energy-Efficiency Trends

Several key trends toward more efficient microprocessors, servers, storage devices, and site infrastructure systems were identified that could have a significant impact on the future energy use attributable to the nation's servers and data centers. These key trends are described below.

Microprocessors

Advancements in the design and manufacture of integrated circuits have led to dramatic improvements in microprocessor speed and performance during the past two decades. For

example, from 1986 to 2002, microprocessor performance improved at a rate of roughly 50 percent per year (Hennessy and Patterson 2007). Although gains in microprocessor technology have steadily improved server performance, the increasing computing demands placed on servers combined with only modest gains in performance per watt and a rapidly growing installed server base have, over time, led to a significant rise in the aggregate energy use of U.S. servers (see Chapter 2) (Barroso 2005).

Microprocessor technology is continuously advancing, and three key trends in server microprocessor technology hold great promise for reducing server energy use in the near future: (1) the shift to multiple cores, (2) the development of dynamic frequency and voltage scaling capabilities, and (3) the development of virtualization capabilities.

Multiple-core microprocessors contain two or more processing cores on a single die, which run at slower clock speeds and lower voltages than the cores in single-core chips but handle more work in parallel (with proper software support) than a single-core chip. Additionally, because the cores share architectural components such as memory elements and memory management, signaling can be faster and consume less energy than is the case for single-core systems (Greer 2006). This combination of increased performance and reduced energy consumption has significantly improved performance per watt. Newer low-voltage multiple-core processors (primarily targeting the volume server market) are expected to deliver even greater energy savings. Published estimates on processor-level energy savings attributable to state-of-the-art multiple-core designs range from roughly 40 to 60 percent (Intel 2007, Tremblay 2006).

Dynamic frequency and voltage scaling features allow microprocessor frequency or voltage to ramp up or down to better match the computational demands. Thus, when utilization is low, the microprocessor's clock speed can be reduced, which reduces energy consumption and heat dissipation. Frequency and voltage scaling are done automatically and constantly adjust to changes in computational demand, continuously minimizing processor energy consumption.

Used in combination with energy-efficient multiple-core processor designs, frequency and voltage scaling can lead to significant processor-level energy savings. The energy savings from frequency and voltage scaling are a function of average processor utilization. Recent industry data suggest that the system-level energy use of a server with a dual-core microprocessor could be reduced by as much as 20 percent through frequency and voltage scaling (at 20 percent processor utilization), with no measurable impact on performance (AMD 2006).

Many new server microprocessors are also designed to facilitate virtualization. Virtualization allows organizations to replace several dedicated servers that operate at a low average processor utilization level with a single "host" server that provides the same services and operates at a higher average utilization level.

Virtualization may offer significant energy savings for volume servers because these servers typically operate at an average processor utilization level of only five to 15 percent (Dietrich

2007, US EPA 2007).¹¹ The typical U.S. volume server will consume anywhere from 60 to 90 percent of its maximum system power at such low utilization levels (AMD 2006, Bodik et al. 2006, Dietrich 2007). Virtualization will increase the processor utilization level of the host server (both because multiple virtual servers are running and because of a small processor utilization “overhead” associated with virtualization software), thereby increasing energy use. However, this incremental gain in host server energy use is more than offset by the savings from eliminating the significant energy load associated with running multiple servers at low utilization rates. Virtualization software must also coordinate power-management capabilities across virtualized servers.

The treatment of current trends toward multiple core microprocessors, inclusion of dynamic frequency and voltage scaling, and inclusion of virtualization capabilities in this analysis is explained in the discussion of servers below.

Several major microprocessor manufacturers have reported recent research advances in insulation materials that may reduce transistor leakage at the 45-nanometer range; this development promises to significantly increase chip performance and energy efficiency (Markoff 2007). Other R&D advances aimed at improving the efficiency of microprocessor architecture and materials have also been reported recently (US DOE 2007c). Because these technological advances are still in the R&D stage and there is, therefore, a lack of data on energy performance and future market availability, they were not included in this analysis. However, these developments suggest that the advances in microprocessor energy efficiency described in this section are likely to continue for the foreseeable future.

Servers

Major U.S. server manufacturers are clearly moving toward the production and marketing of more energy-efficient servers. A review of server manufacturer websites revealed several key features of servers currently being marketed as “energy efficient”:¹²

- The use of multiple-core microprocessors with power management (i.e., dynamic frequency and voltage scaling) and virtualization capabilities,
- The use of high-efficiency power supplies, and
- The use of internal variable speed fans for on demand cooling.

¹¹ The widespread underutilization of servers is one of the most often-cited reasons for suboptimal energy efficiency in data centers. Maximizing the utilization of existing servers therefore represents one of the most significant opportunities for energy savings in many U.S. data centers and server installations.

¹² The term “energy-efficient” is used in quotation marks here for several reasons: (1) the references in this report to energy-efficient servers are not meant to imply an EPA label, standard, or any other endorsement of server energy efficiency; (2) to refer to the definition of an “energy efficient” server created for the purposes of scenario modeling in this report; and (3) because the energy efficiency of servers is improving on a continuous basis and thus the quotation marks indicate that the definition is only a current “snapshot” of energy efficiency that may change over time.

These characteristics are typical of servers currently being marketed as “energy efficient” for the volume server market. Much recent attention has been paid to improving the energy efficiency of volume servers because of their enormous market share, the large percentage of typical data center energy use (see Figure 2-5) they represent, and the trend toward increasing their power density (e.g., blade servers). For these reasons, recent low-voltage multiple-core microprocessors have been offered almost exclusively for the volume server market. Moreover, mid-range and high-end servers already typically employ high-efficiency power supplies (Kooimey 2007). Thus, the remaining examination of “energy efficient” servers in the chapter is focused on volume servers.

For many volume servers, the microprocessor, cooling fan(s), and power supply losses combined can account for 50 to 80 percent of total server energy use (Dietrich 2007, Eubank et al. 2003, Patterson et al. 2006, Pouchet 2007). Efficiency gains in these three components can therefore save significant energy.

Recent advancements in memory technology may also help to improve the energy efficiency of future servers. According to manufacturer data, next-generation memory modules may reduce server memory energy consumption by up to 24 percent (Sporer 2007). Given that the rising use of multiple-core processors and virtualization software is expected to lead to a significant increase in required server-level memory, energy-efficiency gains in memory technology hold substantial promise for reducing future server energy use.¹³

On average, a current “energy-efficient” volume server (which contains a more efficient microprocessor, power supply, and cooling fan) will consume approximately 25 percent less energy (at all levels of processor utilization) than similar volume server models. This assumption was based on industry test data comparing the power consumption of a volume server currently being marketed as “energy efficient” to the power consumption of a comparable volume server that it would replace (Dell 2007). These test data suggest that 25 percent less energy would be consumed across a range of typical processor utilization levels.¹⁴

The estimate of 25-percent energy savings was designed to characterize all the current trends toward more energy-efficient microprocessors and server designs discussed above. This estimate was applied on a per-server basis for the purpose of modeling the above energy-efficiency trends in this report.

In addition to estimating the typical energy savings associated with “energy-efficient” volume servers, it was also necessary to estimate their adoption rate during the next five years to project potential energy savings. In the absence of shipment or market penetration data for new “energy-efficient” volume server models, it was estimated that “energy-efficient” volume servers

¹³ Because there is a lack of independent data on the energy savings and market potentials of emerging memory technologies, the modeling of efficiency trends in this report does not explicitly address these issues. However, Chapter 8 recommends future work in this area.

¹⁴ Although estimates for the energy savings associated with “energy-efficient” volume servers currently being marketed range from 20 to 50 percent, the test data cited above offer the most transparent and conservative estimate based on the information available.

would represent approximately five percent of all U.S. volume server shipments in 2007 and that this percentage would increase linearly to 15 percent of all U.S. volume server shipments by 2011. This estimate was based on the following observations:

- Currently only a handful of “energy-efficient” volume server models are available on the market
- A representative of a major server manufacturer with a new “energy-efficient” volume server model expects “somewhere in the 10 to 20 percent range of our customers to be interested in this product and to ultimately migrate to [these “energy-efficient” servers]” (Shankland 2006).

In addition, it was assumed that dynamic frequency and voltage scaling (i.e., power management) is only enabled on around 10 percent of applicable volume and mid-range servers, based on an industry expert’s recent estimate of the current utilization rate of microprocessor power-management features (Wigle 2007).

Furthermore, current trends toward server virtualization are expected to slow the growth of the installed U.S. volume server base during the next five years. According to IDC (2007b), worldwide x86 volume server shipments, which were once projected to increase 61 percent by 2010, are expected to grow just 39 percent during that same period because of increasing server virtualization (IDC 2007c).¹⁵

There is evidence that the current private-sector trend toward virtualization is also occurring in the federal sector. For example, the U.S. Postal Service has been able to eliminate the need for 791 of its 895 physical servers through aggressive server virtualization and has reportedly reduced its data center power consumption by 3.5 million kWh per year (Rowe 2007). Similarly although on a smaller scale, the National Operations Coordination Center, run by the federal Aviation Administration, has implemented plans to reduce its 14 servers to six (Zaino 2006).

Storage Devices

Enterprise HDD storage devices are expected to become more efficient during the next five years in part because of a shift to smaller form factor disk drives and increasing use of serial advanced technology attachment (SATA) drives. Osterberg (2007) predicts that the average power use per drive will decrease by approximately seven percent between 2007 to 2010 as a result of these trends. This estimate is used to reflect expected energy-efficiency gains in enterprise storage devices from 2007 to 2011 in this analysis.

Additionally, improved management of storage resources may result in significant data center energy savings. Management strategies include storage virtualization, data de-duplication, storage tiering, and movement of archival data to storage devices that can be powered down when not in use.

¹⁵ The projected growth in the installed base of servers depicted in Figure 2-3 reflects the reduction in future x86 volume server shipments predicted by IDC. Without this reduction in x86 shipments, the total installed base of servers would otherwise grow to nearly 19 million as predicted in “historical trends” analysis described in Appendix 4.

The use of solid-state flash memory devices may be an emerging energy-efficient storage option in data centers (Gonsalves 2007). However, because of a lack of available data on the energy savings and market potentials of flash storage devices in data center applications, this trend was not included in the analysis.

Site Infrastructure Systems

Recent trends toward more efficient infrastructure systems are important because these systems typically account for 50 percent or more of the total energy consumed in data centers and server installations. Recent information indicates that many data centers are pursuing energy-efficiency improvements to infrastructure systems. These improvements include low-cost measures, such as improved airflow management and optimization of temperature and humidity set points, as well as more capital-intensive measures, such as upgrading to more efficient uninterruptible power supply (UPS) systems and water-cooled chillers with variable-speed fans and pumps.

To capture these trends, it was assumed that there would be roughly a one-percent annual improvement in the energy efficiency of infrastructure systems during the period 2007 to 2011. This estimate was derived from industry experts estimating that recent technological improvements and interest in energy efficiency would likely result in a two percent annual improvement in data center energy efficiency (which includes improvements to IT systems and infrastructure systems) during the next five years (US EPA 2007). It was assumed that one-half of the annual improvement would be attributable to infrastructure systems. Therefore, although the energy use of infrastructure systems was assumed to be of equal magnitude to the energy use of IT equipment in 2006, infrastructure systems are expected to represent about 90 percent of the energy use of IT equipment in 2011. It was also assumed that infrastructure efficiency would improve in a linear fashion from 2007 to 2011.

While the trends discussed in the preceding paragraphs can help to maximize the efficiency of individual data center components, even greater energy-efficiency gains may be realized by optimizing the operation and efficiency of the data center as a holistic system (e.g., through real-time, facility-level energy monitoring and management systems).

3.1.2. Estimates of Future Electricity Consumption With and Without Current Energy Efficiency Trends

Estimates of the future electricity use and costs attributable to the nation's servers and data centers for the years 2007 to 2011 were developed using the basic methodology outlined in Chapter 2. Two separate scenarios were analyzed. A "current efficiency trends" scenario incorporated the energy-efficiency trends described in Section 3.1.1, which were based on observed energy-efficiency trends as of early 2007. The purpose of the current efficiency trends scenario was to project the current energy use trajectory of U.S. servers and data centers for the next five years given the efficiency trends already under way. A "historical trends" scenario was also analyzed, which did not reflect current energy-efficiency trends but simply extrapolated observed 2000 to 2006 energy use trends (from Chapter 2) into the future. The purpose of the historical trends scenario was to project the energy use of U.S. servers and data centers that would otherwise occur if no energy-efficiency improvements were made.

Table 3-1 summarizes the key assumptions associated with each scenario.

Table 3-1. Summary of Historical Trends and Current Efficiency Trends Scenario Assumptions

Scenario	IT Equipment	Site Infrastructure Systems
Historical trends	<ul style="list-style-type: none"> • 2000 to 2006 energy use growth trends for IT equipment extrapolated to 2011 	<ul style="list-style-type: none"> • PUE ratio frozen at 2.0 for all space types for years 2007 to 2011
Current efficiency trends	<ul style="list-style-type: none"> • Volume server virtualization leading to a physical server reduction ratio of 1.04 to 1 (for server closets) and 1.08 to 1 (for all other space types) by 2011 • “Energy efficient” servers represent 5% of volume server shipments in 2007 and 15% of shipments in 2011 • Power management is enabled on 10% of applicable servers • Average energy use per enterprise storage device will drop 7% by 2011 	<ul style="list-style-type: none"> • PUE ratio drops to 1.9 by 2011 for all space types because of improved technological and operation performance

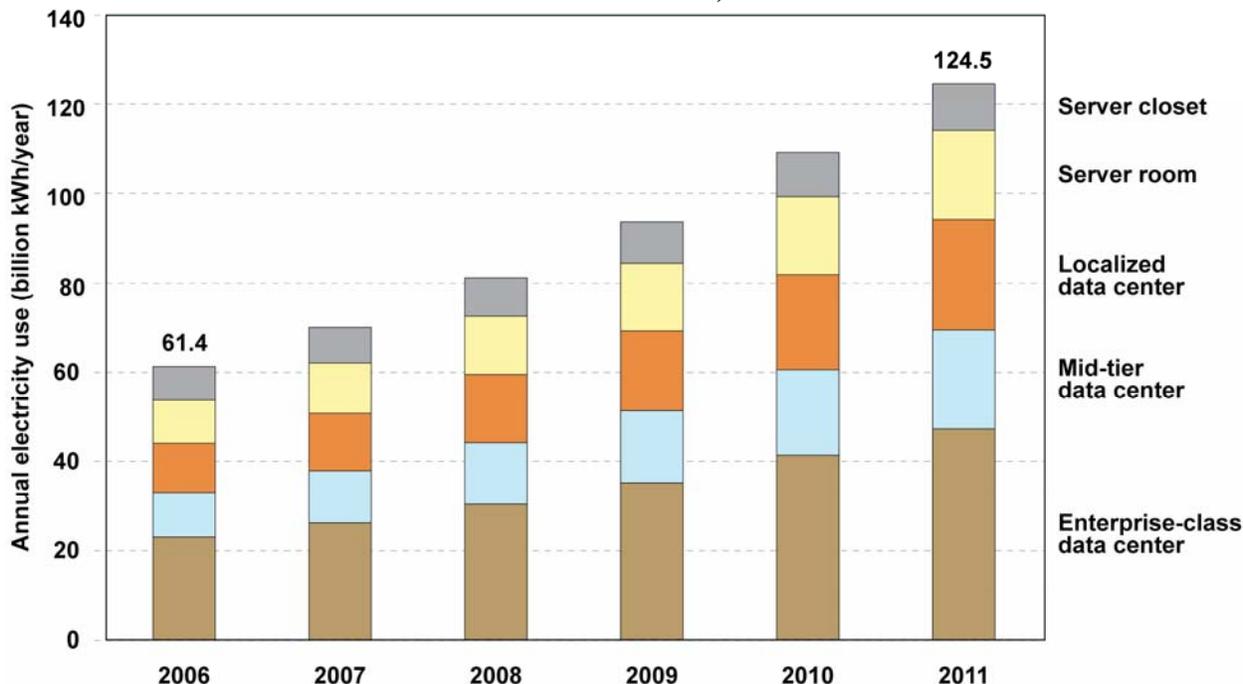
To project future energy use for each scenario, the following values were estimated for each scenario for the period 2007 to 2011, using the general modeling approach outlined in Chapter 2:

- The U.S. installed server base by server class and space type
- Average energy use per server
- Energy use for storage devices and network equipment
- Energy use by infrastructure systems

Appendix 4 describes in detail how these values were estimated.

The results of the scenario analyses indicate that in the absence of any efficiency gains during the next five years, the aggregate electricity use associated with U.S. servers and data centers would more than double from 2006 to 2011, as shown in Figure 3-1. The total 2011 electricity use in the historical trends scenario is estimated at about 125 billion kWh, or about 2.9 percent of total projected nationwide electricity use in 2011 (US DOE 2007a).

Figure 3-1. Projected Electricity Use, by Space Type, Historical Trends Scenario, 2007 to 2011



Given the energy-efficiency trends discussed in this chapter, the aggregate electricity use associated with U.S. servers and data centers is likely to rise more slowly over the next five years than indicated in the historical trends scenario. However, there will still likely be substantial growth, and this sector will consume a significant portion of U.S. electricity. As indicated in Figure 3-2, electricity use is expected to rise to about 107 billion kWh in the current efficiency trends scenario, which is a 75 percent increase over 2006 electricity use and represents about 2.5 percent of total projected nationwide electricity use in 2011 (US DOE 2007a). Figure 3-3 provides a year-by-year comparison of the historical trends and current efficiency trends scenarios, as well as a comparison to the 2000 to 2006 energy use trends discussed in Chapter 2.

Figure 3-2. Projected Electricity Use, by Space Type, Current Efficiency Trends Scenario, 2007 to 2011

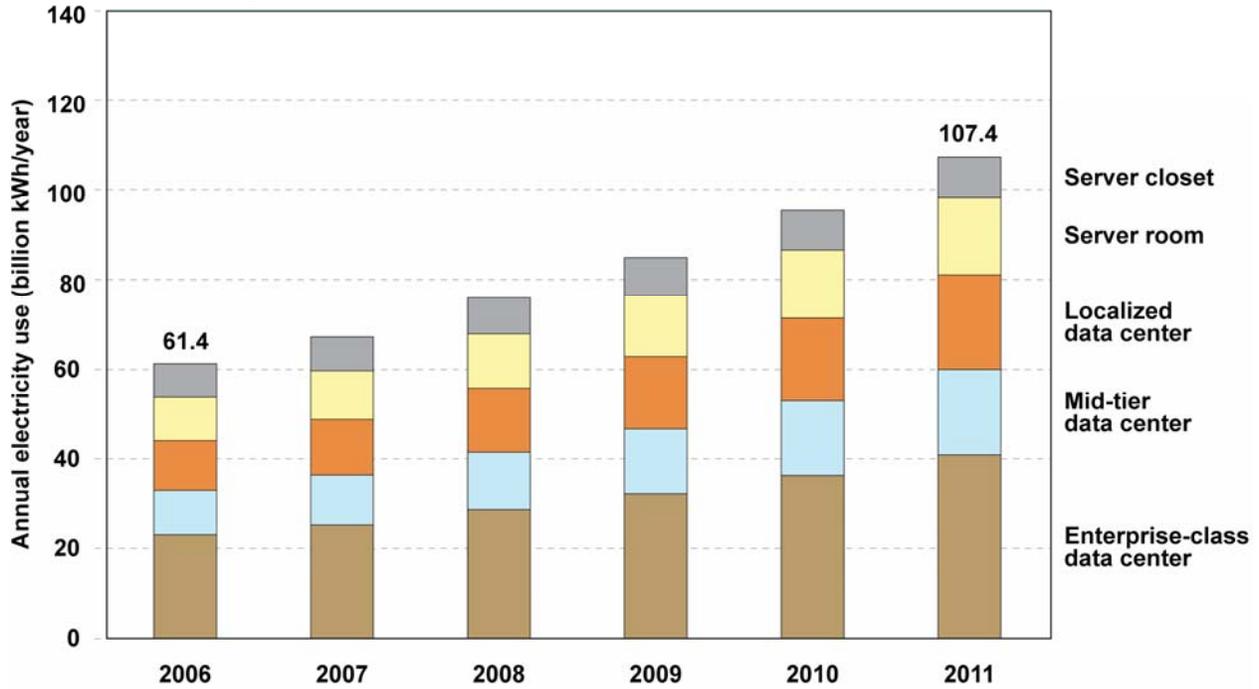
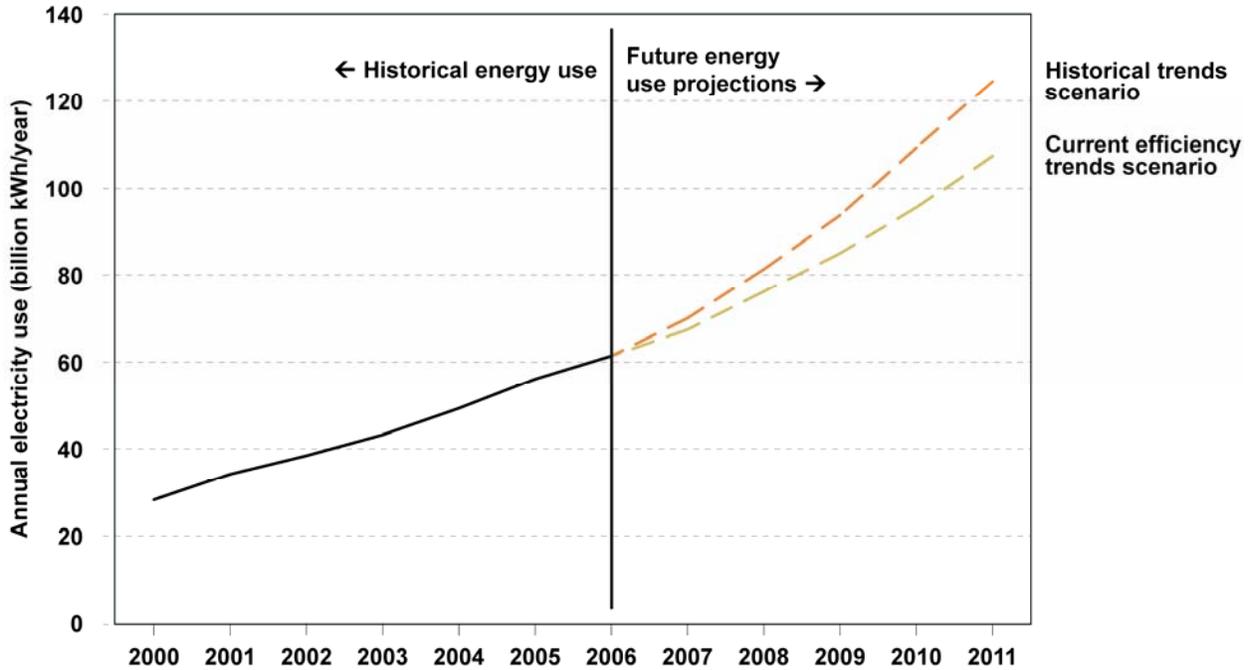


Figure 3-3. Comparison of Projected Electricity Use, Historical Trends and Current Efficiency Trends Scenarios, 2007 to 2011



3.1.3. Estimates of Expected Savings in Electricity Use and Costs from Current Energy-Efficiency Trends

The shift toward greater energy efficiency in data centers and servers, as reflected in the current efficiency trends scenario, is expected to yield significant savings in electricity use during the next five years, as shown in Table 3-2. By 2011, about 17 billion kWh will be saved annually or about 14 percent of total energy use. In addition, the annual growth rate in electricity use declines to 12 percent during the next five years, compared to 16 percent in the historical trends scenario.

Table 3-2. Projected Annual and Cumulative Electricity Use (in billion kWh), Historical Trends and Current Efficiency Trends Scenarios

Scenario	2007	2008	2009	2010	2011	2007-2011 Total	2007-2011 CAGR	% of historical trends scenario
Historical trends	70.0	81.3	93.7	109.3	124.5	478.8	16%	100%
Current efficiency trends	67.4	76.1	85.0	95.5	107.4	431.4	12%	90%

Significant savings in electricity costs are also expected, as shown in Table 3-3. Electricity cost savings are estimated at about \$1.2 billion per year in 2011, representing about 14 percent annual savings, and a 10 percent savings over the five-year period.¹⁶

Table 3-3. Projected annual and Cumulative Electricity Costs (in billions of 2005 dollars), Historical Trends and Current Efficiency Trends Scenarios

Scenario	2007	2008	2009	2010	2011	2007-2011 Total	% of historical trends scenario
Historical trends	5.2	6.0	6.8	7.8	8.6	34.5	100%
Current efficiency trends	5.0	5.7	6.2	6.8	7.4	31.1	90%

Assuming that the federal sector accounts for about 10 percent of the electricity use and electricity costs attributable to servers and data centers in the U.S. (see Chapter 2), current energy-efficiency trends may lead to cumulative energy savings of about 4.7 billion kWh through 2011 with a total cumulative cost savings of about \$340 million.

3.2. Opportunities for Additional Energy-Efficiency Savings

There are opportunities for greater savings in electricity use and electricity costs over the next five years through the adoption of energy-efficient technologies and practices. The savings would be in addition to those predicted in the current efficiency trends scenario. To understand

¹⁶ These costs were estimated using projections for average U.S. commercial and industrial electricity rates for the next five years (U.S. DOE 2007) and by applying the electricity rates by space type as outlined in Chapter 2.

these opportunities, three additional energy-efficiency scenarios were developed and analyzed. This section presents:

- The three energy-efficiency scenarios and their key assumptions and underlying information,
- Estimates of future electricity use and costs for each scenario, and
- A discussion of the expected savings in electricity use and costs associated with these energy efficiency scenarios.

3.2.1. Additional Energy-Efficiency Savings Scenarios

The three additional energy-efficiency savings scenarios considered in this report are:

- *Improved operation.* The “improved operation” scenario includes energy-efficiency improvements over and above current trends that are essentially operational in nature and require little or no capital investment. This scenario characterizes the efficiency gains that can be realized through optimizing the existing capital stock using such strategies as adjusting temperature and humidity set-points and enabling IT equipment with energy management features. This scenario represents the “low hanging fruit” that remains after current trends are accounted for.
- *Best practice.* The “best practice” scenario represents the efficiency gains that can be obtained by more widespread adoption of the practices and technologies found in the most energy-efficient facilities in operation today. These facilities employ proven technologies and management practices and represent the “best in class” of today’s data centers. This scenario represents technologies and practices that are cost-effective today though perhaps not in widespread use.
- *State of the art.* The “state-of-the-art” scenario represents how far energy-efficiency savings could be taken using available technologies. It assumes that all U.S. servers and data centers will be operated at maximum energy efficiency using only the most efficient technologies and best management practices available today. Thus it describes the maximum technical potential for energy savings during the next five years (significant market intervention would be required to attain these savings).

Table 3-5 lists the assumptions used to project the energy use and costs attributable to U.S. servers and data centers for the period 2007 to 2011 for each of the three scenarios above. These assumptions represent typical energy-efficiency opportunities that could be employed in each scenario and are indicative of the increasingly aggressive energy-efficiency activities that would be encompassed by these types of scenarios. The scenarios are meant to characterize the total technical potential for energy-efficiency improvement associated with the activities that are included in each scenario. As such, the scenarios are meant to provide credible estimates of future energy savings potential but are not predictions of future behavior. The scenarios and assumptions were developed from a review of a wide range of energy-efficiency opportunities, as shown in Table 3-6. The modeling framework employed in these scenario projections (discussed in detail in Appendix 4) ensured that “double counting” of energy savings across the assumed efficiency measures was strictly avoided so that the most realistic technical potentials could be presented.

The list in Table 3-6 reflects the main categories of energy-efficiency options that may be available and that could be pursued through one or more of the scenarios outlined above. Although every effort was made to capture the full range of potential efficiency opportunities available to U.S. data centers, the list of opportunities in Table 3-6 should not be interpreted as inclusive of all available opportunities. Furthermore, the inclusion of a particular technology in this list does not imply endorsement of it, and exclusion from this list does not imply that a technology or policy would be an ineffective means of improving data center efficiency.

Table 3-5. Summary of Alternative Efficiency Scenario Assumptions

Scenario	IT Equipment	Site infrastructure systems
Improved operation	<ul style="list-style-type: none"> • Volume server virtualization leading to a physical server reduction ratio of 1.04 to 1 (for server closets) and 1.08 to 1 (for all other space types) by 2011 • 5% of servers eliminated through virtualization efforts are not replaced (e.g., legacy applications) • “Energy efficient” servers represent 5% of volume server shipments in 2007 and 15% of shipments in 2011 • Power management enabled on 100% of applicable servers • Average energy use per enterprise storage drive declining 7% by 2011 	<ul style="list-style-type: none"> • PUE ratio declining to 1.7 by 2011 for all space types assuming: • 95% efficient transformers • 80% efficient UPS • Air cooled direct exchange system chiller • Constant speed fans • Humidification control • Redundant air handling units
Best practice	<ul style="list-style-type: none"> • Moderate volume server virtualization leading to a physical server reduction ratio of 1.33 to 1 (for server closets) and 2 to 1 (for all other space types) by 2011 • 5% of servers eliminated through virtualization efforts are not replaced (e.g., legacy applications) • “Energy efficient” servers represent 100% of volume server shipments 2007 to 2011 • Power management enabled on 100% of applicable servers • Average energy use per enterprise storage drive declining 7% by 2011 • Moderate reduction in applicable storage devices (1.5 to 1) by 2011 	<ul style="list-style-type: none"> • PUE ratio declining to 1.7 by 2011 for server closets and server rooms (using previous assumptions) • PUE ratio declining to 1.5 by 2011 for data centers assuming: • 98% efficient transformers • 90% efficient UPS • Variable-speed drive chiller with economizer cooling or water-side free cooling (in moderate or mild climate region) • Variable-speed fans and pumps • Redundant air-handling units
State-of-the-art	<ul style="list-style-type: none"> • Aggressive volume server virtualization leading to a physical server reduction ratio of 1.66 to 1 (for server closets) and 5 to 1 (for all other space types) by 2011 • 5% of servers eliminated through virtualization efforts are not replaced (e.g., legacy applications) • “Energy efficient” servers represent 100% of volume server shipments 2007 to 2011 • Power management enabled on 100% of applicable servers • Average energy use per enterprise storage drive declining 7% by 2011 • Aggressive reduction of applicable storage devices (~2.4 to 1) by 2011 	<ul style="list-style-type: none"> • PUE ratio declining to 1.7 by 2011 for server closets and server rooms (using previous assumptions) • PUE ratio declining to 1.5 by 2011 for localized and mid-tier data centers (using previous assumptions) • PUE ratio declining to 1.4 by 2011 for enterprise-class data centers assuming • 98% efficient transformers • 95% efficient UPS • Liquid cooling to the racks • Cooling tower (in moderate or mild climate region) • Variable-speed fans and pumps • CHP

**Table 3-6. Potential Energy-Efficiency Improvement Opportunities*
for Servers and Data Centers**

Computing software
Design software to avoid excess code and inefficiencies (treat CPU cycles as a finite resource)
Provide developer tools to help improve efficiency of software
Enable shifting of computational load among systems for maximizing energy efficiency
Upgrade applications no longer supported on latest technology and/or operating systems, allowing removal of legacy servers
Implement virtualization to allow consolidation of server and storage hardware
IT hardware (computing, storage and network)
<i>Operational Improvements</i>
Turn off (ideally remove) dead, obsolete, or excess equipment
Turn off or power-manage equipment that won't be used for extended periods of time (e.g., development systems not in active use, systems for future expected increases in activity, etc.)
Enable power-management features on existing equipment (e.g., frequency/voltage scaling)
Maximize utilization of storage capacity through shared data storage, data compression, and data de-duplication
<i>Design Improvements</i>
Accept high-efficiency power supplies over full operating range (including DC-DC conversions) or directly accept moderate DC voltage
Digitally control power supplies to better match output to load
Use high-efficiency variable speed fans (within IT equipment)
Reduce energy use at lower utilizations (whether the resource is processing capacity, memory, communications, or etc.). Applies to individual systems and to clusters.
Improve microprocessors to lower leakage current, increase system integration, etc.
Use storage virtualization and massive array of idle disks (MAID) technologies to allow storage power management
Use centralized servers (large systems) to improve sharing of computer resources
Improve hardware support for virtualization
Use built-in power monitoring
Electrical Systems
Use high-efficiency power distribution (i.e., higher-voltage AC or moderate-voltage DC (50-600 VDC))
Use premium-efficiency motors in fans and pumps
Use high-efficiency UPS units over full range of load
Use rotary-based UPS units
Right size power distribution and conversion to optimize efficiency
Use on-site generation with grid as back-up
Heat Removal
Improve airflow management (i.e., use hot/cold aisle configuration or penetration sealing)
Adjust environmental conditions (temperature and humidity set points) to allow wider range while still meeting manufacturer specifications
Optimize data center airflow configuration using visualization tools (computational fluid dynamics modeling or infrared tomography)
Use high-efficiency variable-speed air-handler fans and chilled water pumps
Use variable-speed chillers
Use variable-speed, primary-only chilled water pumping
Use high-efficiency chiller and chilled water supply motors
Use high-efficiency CRAC units
Use air-side economizers (outdoor air) when outdoor conditions permit (preferably in mild climate locations)

Use water-side economizers (cooling tower) when outdoor conditions permit (preferably in mild climate locations)
Commission infrastructure systems to ensure set points are at proper values, sensors are in calibration, airflow is within design tolerances, etc.
Rebalance air-handler system after significant IT reconfiguration
Size systems and configure redundancy to maximize efficiency (e.g., use redundant air-handler capacity in normal operation)
Increase chilled water supply and return temperature difference to reduce chilled water flow
Optimize chilled water plant (cooling tower)
Reuse waste heat for space heating
Use direct liquid cooling (water or other dielectric liquid) with currently available technology (i.e., in-rack or in-row cooling) and emerging technology (i.e., in-chassis or chip-level)
Controls and Management
Use system management hardware/software that enables powering down (to sleep and/or off) parts of server clusters during times of low utilization
Dispatch non-time-sensitive computational operations to reduce peak computing load and allow reduction of total server count
Provide for standard reporting of power use, platform temperature, and processor utilizations to assist operators in understanding and managing energy use in their data centers
Use shared computing models, such as grid computing
Optimize cooling controls to dynamically match the cooling supply to the IT heat load
Dynamically optimize the assignment of work across the data center to ensure maximum efficiency
Monitor power in real time
Distributed Generation
Use combined heat and power
Use renewable energy (e.g., photovoltaic panels)
Use fuel cells

*Each opportunity is not necessarily applicable to all data centers

3.2.2. Estimates of Future Electricity Use, Costs, and Savings for Each Scenario

To project future energy use for all three scenarios, the values below were estimated for each scenario for the period 2007 to 2011, using the general modeling approach outlined in Chapter 2:

- The U.S. installed server base by server class and space type
- Average energy use per server
- Energy use for storage devices and network equipment
- Energy use by infrastructure systems

Appendix 4 describes how these values were estimated.

Figure 3-4 compares total projected annual electricity use for the period 2007 to 2011 for all scenarios considered in this study. For the three additional energy-efficiency scenarios that address efficiency opportunities beyond current efficiency trends, the projected energy savings beyond current trends can be clearly seen.

Figure 3-4. Comparison of Projected Electricity Use, All Scenarios, 2007 to 2011

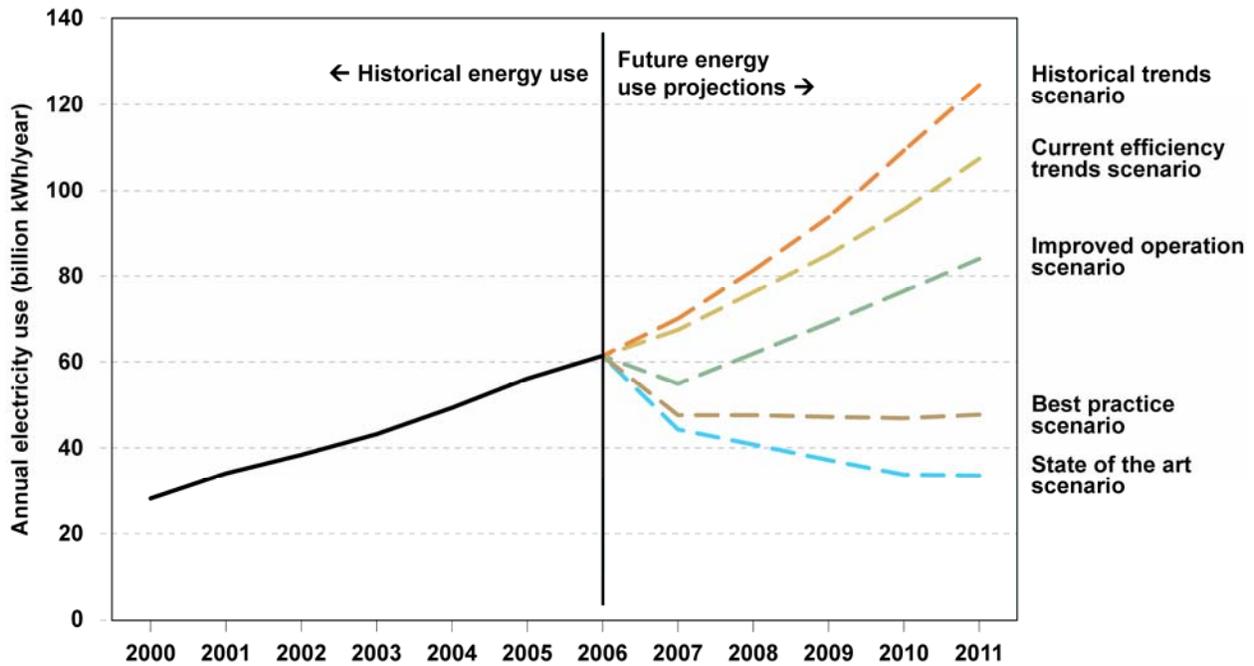


Table 3-7 summarizes the projected aggregate electricity use estimated for each scenario, both annually and cumulatively from 2007 to 2011.¹⁷

Table 3-7. Projected Electricity Use (in billion kWh/year), All Scenarios, 2007 to 2011

Scenario	2007	2008	2009	2010	2011	2007-2011 Total	% of historical trends scenario	% of current efficiency trends scenario
Historical trends	70.0	81.3	93.7	109.3	124.5	478.8	100%	111%
Current efficiency trends	67.4	76.1	85.0	95.5	107.4	431.4	90%	100%
Improved operation	54.9	61.9	69.0	76.4	84.0	346.3	72%	80%
Best practice	47.6	47.6	47.2	46.9	47.7	236.9	49%	55%
State-of-the-art	44.3	40.8	37.2	33.8	33.6	189.7	40%	44%

The results in Table 3-7 demonstrate the significant potential for energy savings from aggressive pursuit of energy-efficiency improvement opportunities beyond the current trends. In 2011, the state-of-the-art scenario would reduce electricity use by approximately 70 percent compared to current efficiency trends. The projected energy savings associated with the state-of-the-art scenario represent the maximum technical potential. The energy savings associated with the

¹⁷ The column titled “% of historical trends scenario” compares the cumulative 2007-2011 total in each scenario to that of the historical trends scenario. Likewise, the column titled “% of current efficiency trends scenario” compares the cumulative 2007-2011 total in each scenario to that of the current efficiency trends scenario.

improved operation scenario (22 percent less than would result from current efficiency trends in 2011) and best practice scenario (56 percent less than would result from current efficiency trends in 2011) are representative of efficiency gains that might more realistically be realized through opportunities that are proven and cost effective today.

As shown in Table 3-8, even with improved operation, the electricity use in 2011 associated with the nation’s servers and data centers is expected to grow, as a percentage of total U.S. electricity use, compared to the energy use in 2006 (when servers and data centers accounted for 1.5 percent of U.S. electricity use) (US DOE 2007a). With more aggressive best-practice and state-of-the-art improvements, however, it will be possible to reduce this share of total U.S. electricity use significantly.

Table 3-8. Projected Electricity Use as a Percentage of Total U.S. Electricity Use, 2011

Scenario	Projected electricity use of U.S. servers and data centers as % of total U.S. electricity use, 2011
Historical trends	2.9%
Current efficiency trends	2.5%
Improved operation	2.0%
Best practice	1.1%
State-of-the-art	0.8%

Table 3-9 summarizes the projected annual electricity costs for each scenario.¹⁸ For the three efficiency scenarios that incorporate improvements beyond current trends, the cumulative savings in electricity costs from 2007 to 2011 (as compared to the costs in current efficiency trends scenario) range from roughly \$6.2 billion (for the improved operation scenario) to roughly \$17.4 billion (for the state-of-the-art scenario).

Table 3-9. Projected Electricity Costs (in billions of 2005 dollars), All Scenarios, 2007 to 2011

Scenario	2007	2008	2009	2010	2011	2007-2011 Total	% of historical trends scenario	% of current efficiency trends scenario
Historical trends	5.2	6.0	6.8	7.8	8.6	34.5	100%	111%
Current efficiency trends	5.0	5.7	6.2	6.8	7.4	31.1	90%	100%
Improved operation	4.1	4.6	5.0	5.4	5.8	24.9	72%	80%
Best practice	3.5	3.5	3.4	3.3	3.3	17.1	50%	55%
State-of-the-art	3.3	3.0	2.7	2.4	2.3	13.7	40%	44%

¹⁸ These cost projections were calculated in the same manner as the cost projections in Table 3-3.

As discussed in Chapter 2, it was estimated that the federal sector accounts for about 10 percent of the electricity use and electricity costs attributable to servers and data centers in the U.S. Based on this assumption, the cumulative savings in electricity costs to the federal government that could be realized through the pursuit of efficiency beyond current trends over the next five years range from \$620 million (for the improved operation scenario) to \$1.7 billion (for the state-of-the-art scenario).

Table 3-10 summarizes the projected annual emissions of carbon dioxide (CO₂) associated with the electricity use of U.S. servers and data centers in each scenario. The estimates in Table 3-10 were derived by applying an average CO₂ emissions factor for U.S. electricity generation from the U.S. DOE's *Annual Energy Outlook* (U.S. DOE 2007a) to the scenario projections for electricity use summarized in Table 3-7. Current efficiency trends are expected to reduce nationwide CO₂ emissions by nearly 11 million metric tons (MMT CO₂) by 2011 compared to the historical trends scenario. However, efficiency improvements beyond current efficiency trends could further reduce nationwide CO₂ emissions by an additional 15 to 47 MMT CO₂ in 2011.

Table 3-10. Projected CO₂ Emissions Associated with the Electricity Use of U.S. Servers and Data Centers (in MMT CO₂/year), All Scenarios, 2007 to 2011

Scenario	2007	2008	2009	2010	2011	2007-2011 Total	% of historical trends scenario	% of current efficiency trends scenario
Historical trends	44.4	51.2	59.2	69.2	78.7	302.8	100%	111%
Current efficiency trends	42.8	47.9	53.6	60.5	67.9	272.8	90%	100%
Improved operation	34.8	39.0	43.5	48.4	53.1	219.0	72%	80%
Best practice	30.2	30.0	29.8	29.7	30.1	149.8	49%	55%
State-of-the-art	28.1	25.7	23.5	21.4	21.2	119.9	40%	44%

4. Electric Utility Impacts from Energy Efficiency in Servers and Data Centers

This chapter addresses the Public Law 109-431 requirement for:

- Analysis of the potential cost savings and benefits to the energy supply chain through increasing the energy efficiency of data centers and servers, including reducing demand, enhancing capacity, and reducing strain on existing grid infrastructure, and consideration of secondary benefits, including potential impact of related advantages associated with substantial domestic energy savings.

Data centers operate almost entirely on electricity, so their increasing demand on the power grid has caused concern among utility planners, particularly in regions that have large concentrations of data centers. Nationwide, the energy use estimate for data centers from Chapter 2 translates into a peak load of more than 7 GW in 2007 (equivalent to the output of about 15 baseload power plants), growing to about 12 GW if current growth trends continue. Energy-efficiency improvements in the state-of-the-art scenario have the potential to actually reduce the peak load attributable to data centers, limiting the peak load contribution to approximately 5 GW through 2011. This load reduction could result in the avoidance of about 1 GW of new generating capacity that would otherwise need to be built by 2011. Looking at the regional distribution of these load impacts on the electricity transmission grid, it is possible to draw the conclusion that most of the metropolitan areas that have high concentrations of data centers are also the areas with the most strained power grids. Thus, the peak-load reductions achievable through energy-efficiency improvements could play a significant role in relieving capacity constraints in these grids.

The issue of data center impacts on the power grid first came to the public's attention during the late-1990s internet boom, when many large (10 MW or greater peak load) data centers were proposed, and utility planners were faced with a large spike in demand growth. Many of these facilities were never built, but the result of these proposals was that the idea that internet data centers were consuming vast amounts of power penetrated the public consciousness, most notably as a result of a *Forbes* magazine article by Huber and Mills (1999), which claimed that electricity use of all information technology equipment (including the internet infrastructure) would equal 50 percent of U.S. electricity consumption within 10 years. These claims were subsequently shown to be many times higher than the growth that could be reasonably expected (Kawamoto et al. 2002, Koomey et al. 2002, Koomey et al. 2004, Koomey et al. 1999), but the concern about rapidly increasing energy use of data centers persists and is gaining visibility because of the accelerating pace of data center construction. For instance, the utility that serves Northern California, Pacific Gas & Electric Company (PG&E), estimates that data centers are currently responsible for 500 MW of peak electrical load in their service territory (Thibodeau 2007) (equivalent to the generating capacity of a large baseload power plant), out of a systemwide peak load of about 20 GW. Likewise, Austin Energy, the utility serving Austin, Texas, estimates that its demand from stand-alone data centers is currently about 2.5 percent of the system load in their service territory (Johnson 2007). Given these demand levels, there is understandable interest in the utility industry to better understand this load and the potential to reduce it.

To assess the impact of energy-efficient data centers on the nation’s electricity infrastructure, this study includes both a quantitative estimate of reduced peak load and avoided generating capacity resulting from electricity savings, as well as a qualitative assessment of the regional impacts of data centers on the U.S. regional transmission and distribution grids. This report does not separately quantify the cost savings in the power generation and transmission system resulting from reduced electrical demand from data centers. It is assumed that these cost savings would be reflected in retail electricity tariffs and thus captured in the energy bill savings estimates presented in Chapter 3. This report also does not consider impacts on the natural gas system from reduced power consumption in data centers.

4.1. Methodology

To estimate the impact of projected energy savings on generating capacity, the savings estimates from the previous chapter are translated into peak power demand impacts using the National Energy Modeling System (NEMS). NEMS is a large, multi-sectoral, partial-equilibrium model of the U.S. energy sector that has been developed over several years by the Department of Energy (DOE) Energy Information Administration (EIA), primarily for the purpose of preparing the *Annual Energy Outlook* (AEO) (US DOE 2007a). NEMS produces a widely recognized baseline forecast for the U.S. through 2030 and is in the public domain. This analysis was conducted using the NEMS-BRS model,¹⁹ which is the AEO 2006 version of NEMS with minor modifications, used for DOE’s appliance standards rulemaking analysis (US DOE 2002a). More details on the analysis method can be found in Hamachi-LaCommare et al. (2004).

The scenario simulation runs are executed by reducing electricity consumption in the NEMS-BRS Commercial Demand Module by an amount equivalent to the total electricity savings for each of the scenarios from the previous chapter. For time-of-use modeling purposes, actual load data from representative data centers were not available, so the savings are modeled using a representative end use from the commercial building sector, which has a relatively flat load shape, similar to that of data centers. The load shape data are from the default AEO 2006 reference case. NEMS uses a different hourly load shape for each of two day-types (weekday and weekend) for each month of the year, in order to allocate the annual consumption values estimated in Chapters 2 and 3 to hourly demand values for each hour of the year. Figure 4-1 illustrates the representative end-use weekday load shape for the two months with the highest and lowest loads. This load shape has a load factor of 0.85. Loads are higher during the summer months because of reduced cooling efficiency, which results from higher outdoor temperatures.

To account for differences between the U.S. regional electricity grids, NEMS models electricity consumption at the level of the U.S. census divisions. The national savings estimates for each scenario were allocated among the U.S. census divisions based on the estimated fraction of national data center energy use in each region (Table 4-1), using information about the location of data centers discussed in Section 4.3 (Discussion of Transmission and Distribution Impacts).

Although data center load shapes are currently very flat, in the advanced efficiency scenarios analyzed here, the load shapes may become less flat because of efficiency improvements. The

¹⁹ “BRS” stands for the Building Research & Standards Office of DOE.

ability to use outside air for cooling (and the efficiency of doing so) and the tracking of IT equipment power to computing load should both affect the shape of the load curve. In both cases, nighttime load could drop relative to daytime. On the other hand, server virtualization and macro-scheduling of computational workloads can smooth out the load distribution for the IT equipment, thus flattening the facility’s load. The net impact of these effects is unknown, and these effects were not considered in this analysis.

Figure 4-1. Weekday Hourly Load Shapes for Representative Commercial-Building End Use

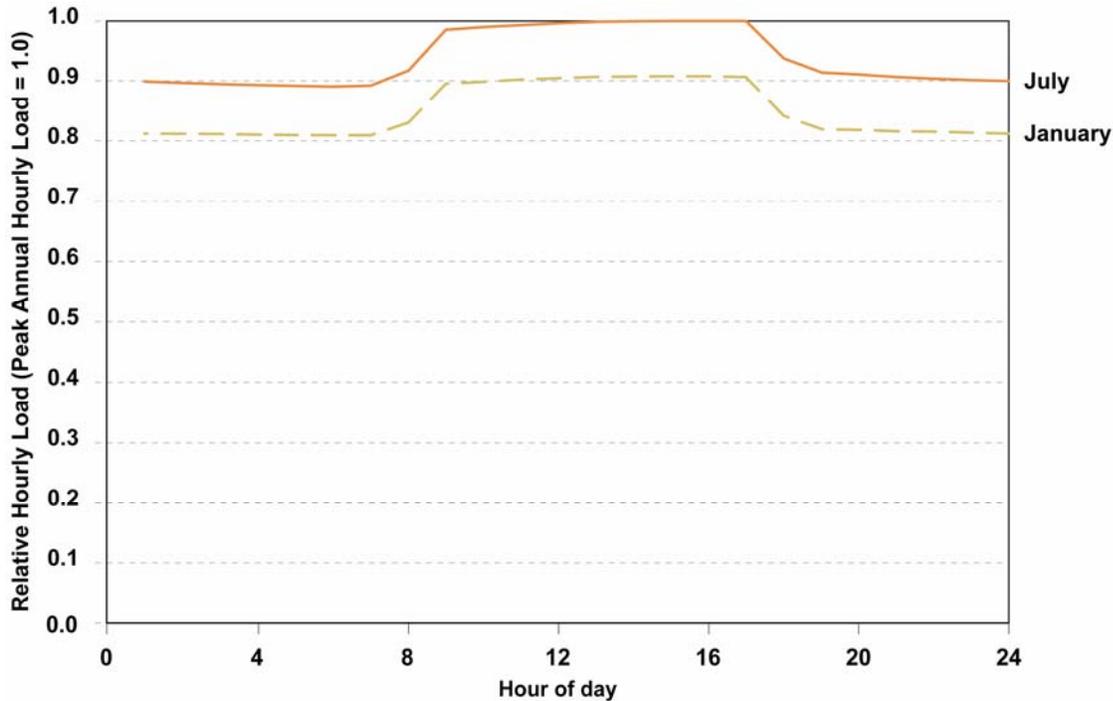


Table 4-1. Regional Share of National Data Center Energy Use

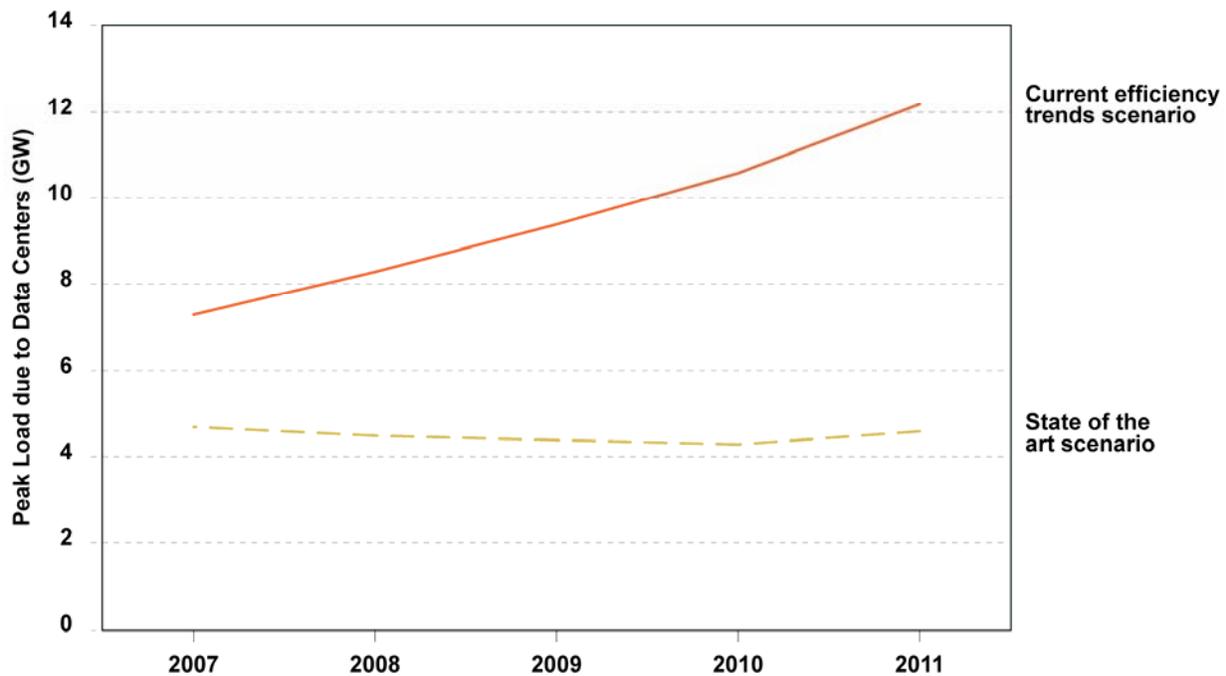
Census Division	%
New England	5
Middle Atlantic	20
South Atlantic	15
East North Central	10
East South Central	5
West North Central	5
West South Central	15
Mountain	5
Pacific	20
	100

Source: LBNL estimate

4.2. Electricity Generation Impacts

As a bounding case, the peak load analysis is presented here for the state-of-the-art scenario, based on the NEMS analysis method described above. Figure 4-2 shows that under the current efficiency trends scenario, data centers (including servers, storage, and network equipment, along with their supporting infrastructure) have a peak load of approximately 7.3 GW in 2007, rising to more than 12 GW in 2011. The significant efficiency improvements in the state-of-the-art scenario actually reduce the peak load from data centers, maintaining a nearly constant load of about 4.5 GW.

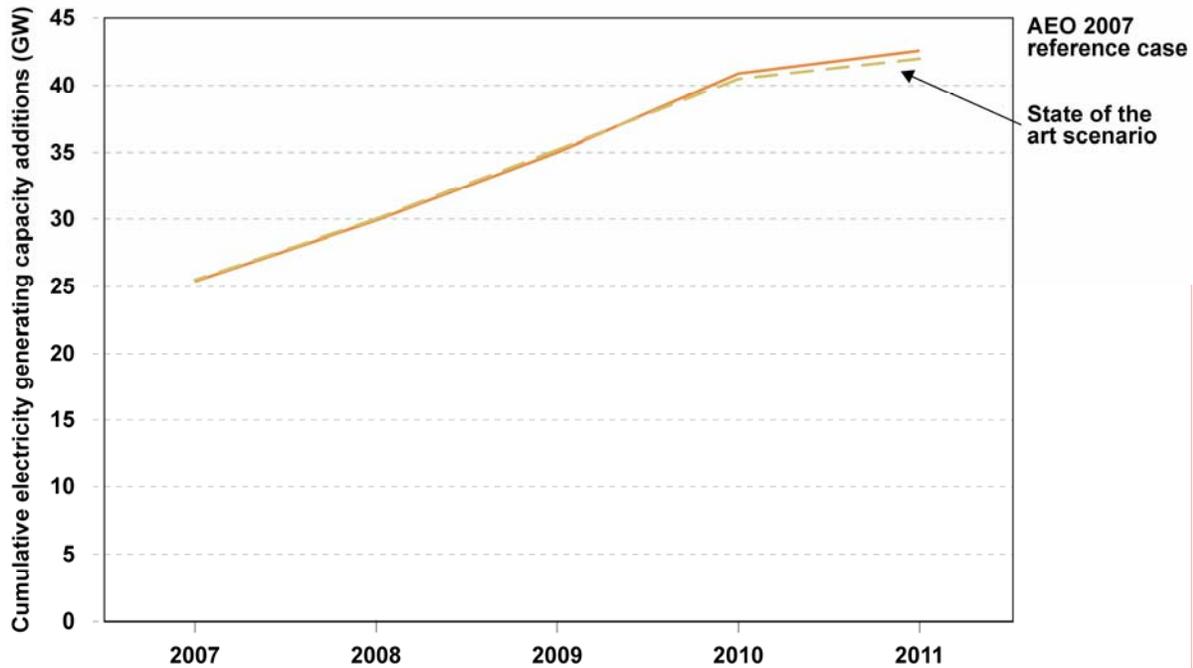
Figure 4-2. Peak Electricity Load Due to U.S. Data Centers



These peak load reductions translate into a modest reduction in generating capacity additions by 2011. Figure 4-3 compares the expected cumulative capacity additions in the AEO 2007 reference case (equivalent to the current trends scenario) to the capacity additions expected under the state-of-the-art scenario. As a result of the energy-efficiency improvements, about 600 MW of new power plant capacity is avoided.²⁰ Although not shown in Figure 4-3, the NEMS model was also used to forecast capacity additions through 2020, assuming that data center efficiency continues to grow at the rate of the state-of-the-art scenario. After 2011, the impact of efficiency improvements on peak load and thus on generating capacity is significantly larger. This lag in the impact of peak-load reductions is a result of the significant lead time in power plant construction. Most of the power plants that will come on line by 2011 are already in the planning stages, so their construction is difficult to avoid, but plants to be commissioned after 2011 can still be avoided through reductions in peak power demand.

²⁰ Note that Figure 4-3 shows total capacity additions for the U.S., not just capacity needed to satisfy growth in demand from data centers. Capacity is measured at the power plant busbar, i.e., it accounts for transmission and distribution losses.

Figure 4-3. U.S. Electric Power Sector Generating Capacity Additions (cumulative, post-2005)



Source: AEO reference case from U.S. DOE (2007a).

4.3. Discussion of Transmission and Distribution Impacts

The benefits of load reductions to the electricity transmission and distribution (T&D) grids are very time and location dependent. To conduct a rigorous quantitative analysis of these benefits is beyond the scope of this study. However, it is possible to qualitatively discuss how reductions in peak load can benefit T&D grids.

4.3.1. Data Center Locations

The impact of data centers on T&D grids depends heavily on the locations of existing and future data centers. Comprehensive data about the locations of data centers are not readily available for several reasons: 1) organizations are concerned about the physical security of these critical infrastructure facilities, 2) private data centers are seen as a strategic asset that companies want to keep as confidential as possible for competitive reasons, 3) many data centers are part of larger commercial buildings and campuses and therefore not identified uniquely as data centers (and they usually are not separately metered), and finally 4) data centers have only recently been seen by government agencies as important infrastructure and an indicator of economic activity, so there are not yet any comprehensive, public statistics compiled on data center location, size, value, etc.

Based on the current structure of the data center market, it was assumed that the majority of existing data centers (in terms of electricity demand) are in major metropolitan areas. Furthermore, because high-speed communication networks have been developed in large part to serve the networking needs of data centers, the major nodes in today's high-speed wide-area

networks are a good indicator of the locations of today's large data centers. These assumptions were used to develop the list of metropolitan areas with major data center concentrations shown in Table 4-2, based on maps published by network backbone providers (such as Sprint, AT&T, and XO Networks) and the trade press (Data Center Knowledge 2007).

Although data centers currently tend to be in more developed metropolitan locations, it appears, based on reports in the press, that large new data centers are increasingly being located in non-metropolitan areas (Dunn 2006b, Mehta 2006). There are many factors motivating this shift, primarily having to do with lower land, construction, labor, and power costs, as well as improved reliability through redundancy and dispersed locations. In assessing T&D impacts, it is assumed that these non-metro data centers have relatively less impact on the power grid because of their location in relatively uncongested areas. Nevertheless, these non-urban data centers still tend to be built in clusters that could very well contribute to new areas of T&D grid congestion in the future. This trend warrants observation by power planners and utility regulators to anticipate and avoid future grid congestion.

4.3.2. Areas of Grid Congestion

The power grid can be thought of like our highway system: in most places and at most times traffic flows smoothly, but certain critical locations have insufficient capacity to meet the demand at peak periods. As with travel demand, the term *congestion* is used to describe areas of the power grid that have insufficient capacity. At a national level, the greatest concern about the power grid today is with the transmission system, which is used for long-distance transport of high-voltage electricity from generating resources to load centers (and, increasingly, between load centers). DOE recently completed a national study to identify regions of the transmission grid that are congested (US DOE 2006b). As Table 4-2 indicates, many of the metropolitan areas that are identified here as having many data centers are also areas with transmission congestion identified by DOE. The study identified several cities from the list below that are *Critical Congestion Areas* (The Atlantic coastal area from metropolitan New York southward through Northern Virginia, and Southern California) or *Congestion Areas of Concern* (New England, the Phoenix-Tucson area, the Seattle-Portland area, and the San Francisco Bay area), as shown in Figure 4-4. Other cities in Table 4-2 were identified as having transmission constraints but not rising to the level of concern that the previous areas warrant. It is important to note that two of the cities with the highest concentration of data centers — New York City and San Francisco — are geographically isolated areas with relatively limited electricity transmission resources. Anecdotal evidence suggests that power-grid vulnerability in these two cities is a factor that data center developers consider in deciding whether to develop new data center assets there.

Given these findings, there is reason for concern about the adequacy of the transmission system to serve today's data centers in certain areas. Insofar as energy efficiency can help reduce peak demand from data centers, it is therefore also a strategy to help reduce stress on the transmission grid in its most vulnerable areas. More definitive conclusions are not possible given the limited data currently available.

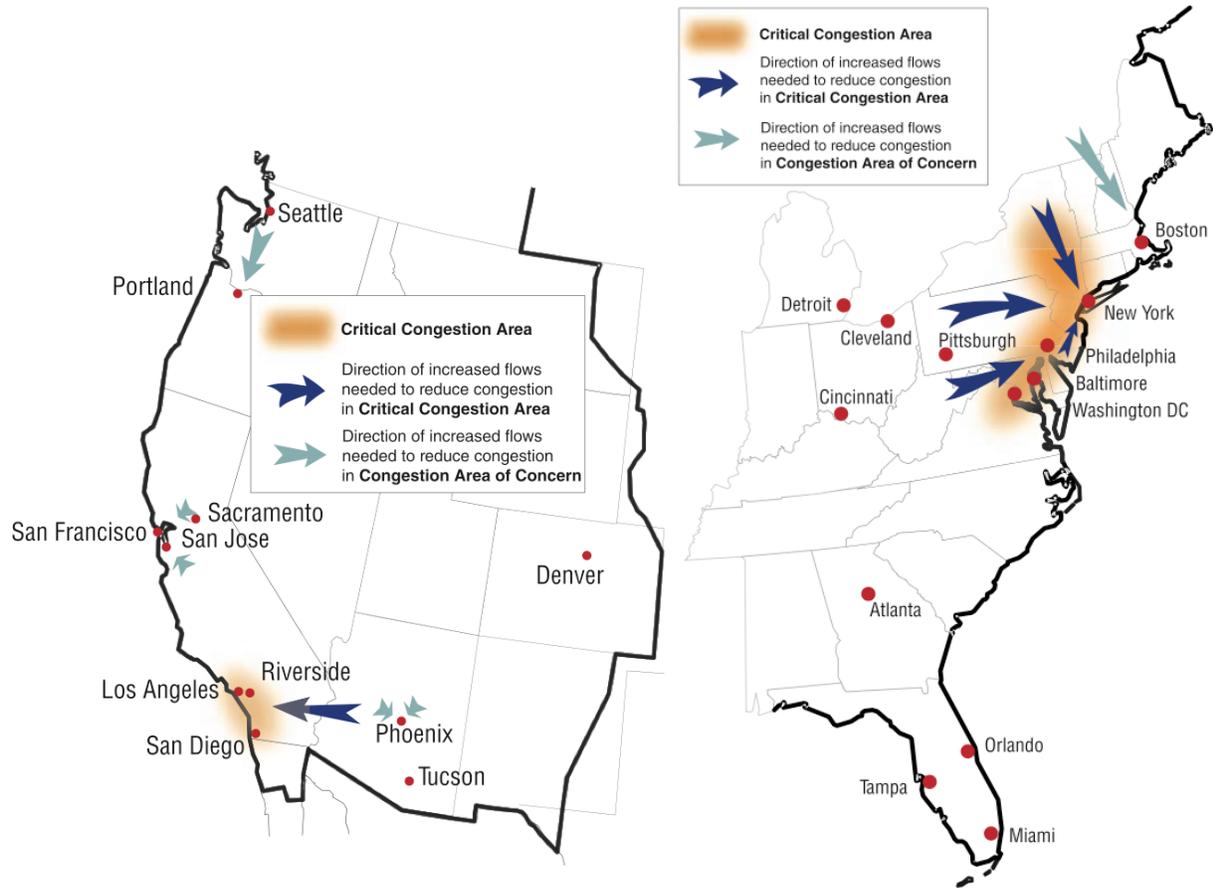
Table 4-2. Transmission Congestion Status in U.S. Metropolitan Areas with Largest Concentration of Existing Data Centers

U.S. Metropolitan Area	Transmission Congestion Severity		
	Critical	Area of Concern	Other
New York City / Northern New Jersey	√		
San Francisco Bay Area CA		√	
Chicago IL			
Dallas TX			
Washington DC area	√		
Austin TX			
Los Angeles CA	√		
Atlanta GA			√
Miami FL			√
Seattle WA			√
St. Louis MO			
Denver CO			√
Boston MA		√	
Houston TX			
San Antonio TX			
Phoenix AZ		√	
Kansas City MO			√
Sacramento CA			

Sources: LBNL analysis and U.S. DOE (US DOE 2006b).

As mentioned above, T&D congestion is not a constant phenomenon but varies in time depending on electricity system demand. Congestion tends to be worst at periods of peak power system demand. Solutions to grid congestion include both supply-side options (e.g., building more transmission lines) and demand-side options (i.e., reducing end-user loads at peak periods through energy efficiency and demand response). Two demand-side solutions have great potential to reduce data centers’ impacts on the T&D grid: distributed generation (DG) and demand response (DR). DG is discussed at length in Chapter 6 and will not be covered here. DR refers to mechanisms to reduce the demand from electricity customers in response to supply conditions, for example, either at times of critical peak system load or in response to high wholesale prices for power. Although it is often assumed that data centers are not good candidates for load management because of the critical function they perform, high-reliability data centers are in fact designed to continue operating when the power grid is unavailable — using on-site power generation and storage, which suggests that they can also reduce the energy drawn from the grid at times of peak load. There is anecdotal evidence that data centers have participated in utility DR programs and demonstrated the ability to make meaningful reductions in demand, particularly facilities that use thermal energy storage for cooling backup. On the other hand, DR using backup generators in data centers may have limited potential due to equipment not being rated for long run hours and limited availability of air emission permits. Nevertheless, DR strategies for data centers should be explored further as a way to reduce impact on the power grid.

Figure 4-4. Critical Areas and Areas of Concern for Transmission Congestion



Source: (US DOE 2006b)

5. Potential Impacts of Energy Efficiency on Product Performance, Reliability, Features, and Overall cost

This chapter addresses the Public Law 109-431 requirement for:

- Analysis of the potential impacts of energy efficiency on product performance, including computing functionality, reliability, speed, and features, and overall cost.

Although interest in energy efficiency has been growing as a solution to power and cooling problems in data centers, one of the main impediments to its adoption has been a perception that some energy-efficiency strategies may degrade data center performance or reliability. Although poorly implemented energy-efficiency measures have the potential to impair performance, measures that are implemented well may not have negative consequences and in some cases can actually improve performance. Performance, including reliability, is a key attribute of high-quality data center equipment, and manufacturers will not compromise it to achieve energy-efficiency improvements. This chapter reviews the potential effects of energy-efficiency measures on IT product performance.

Because energy-efficient data center products and operation could have positive or negative consequences on data center performance, it is important to delineate the direction and relative magnitude of these possible effects. Among the aspects of product performance called out in the text of the law, the one that stands out is “reliability.”

For data center operators, systems must be more reliable than many of the individual components are, so redundant equipment and capacity is provided. The metric of most concern to operators is then the “availability” of the system to meet the data processing needs it is advertised to meet.

In general, server manufacturers and customers will not accept degraded product performance in the interest of increasing energy efficiency. This is not to say that energy efficiency has no effect on performance, just that designers and users take the effect into account to ensure that products and systems meet expectations. Although concern naturally arises regarding energy-efficiency effects that could reduce reliability, there are examples showing that more efficient systems can increase reliability.

In assessing how the energy efficiency of IT equipment could affect performance, it is essential to identify the method by which efficiency is accomplished because different methods have different impacts. The following general categories of improvements are commonly discussed:

- Systems that consume less power at peak performance,
- Systems that consume less power at lower levels of utilization,
- Reducing the performance or capability deployed (e.g., by consolidating and virtualizing or relaxing latency limits).

The IT industry is reducing energy use and/or increasing performance in each successive generation of products. This discussion considers changes that go beyond this “business as usual” change.

For data center operation, some feel that reliability could be reduced by relaxing environmental conditions – temperature and relative humidity – and that higher contamination levels could result from increased use of outside air. ASHRAE (2004b) has set recommended and allowable limits to which server manufacturers have agreed. Because most data center operators adhere to more stringent limits for their facilities that exceed the ASHRAE guidelines (lower temperatures and narrower humidity limits), simply changing operations to match the ASHRAE guidelines will not hurt reliability and will save considerable energy. Operating experience as well as studies currently under way (e.g., Herrlin 2006) have shown that greater use of outside air does not significantly increase contamination, and any concerns can be mitigated by improved filtration.

Although interest in data center energy consumption is partly motivated by environmental concerns, the main concern is cost. There is the direct cost of purchasing power, the cost of excessive power and cooling infrastructure, the cost of building new facilities if existing ones run out of power/cooling capacity, the risks to reliability from hot spots, and the personnel costs of managing excessive IT infrastructure. The energy-efficiency improvements addressed in this report in general reduce costs, often substantially, relative to the alternatives. The magnitude of savings for different measures depends heavily on a data center's existing characteristics: its IT equipment, infrastructure equipment, and the operation of each.

5.1. Performance Impacts Resulting from IT Energy Efficiency Improvements

The subsections below address the performance, cost, and reliability impacts of energy-efficiency improvements to data center computing functionality.

5.1.1. Computing Functionality and Features

More efficient servers will not be less functional, nor will they have fewer needed features. Manufacturers may offer models, configurations, or options that lack specific features as a way to reduce power consumption, but those will only be purchased by customers who don't need the omitted features.

5.1.2. Performance

In systems for which performance is paramount (e.g., high-performance computing), performance will remain the preeminent goal even when attention is given to energy efficiency. Some of the technologies for accomplishing energy efficiency introduce higher latencies (delays) than is currently typical, either on an ongoing basis or when making the transition between performance states. These latencies are often small, infrequent, or irrelevant. Server operators will only adopt energy-saving features to the degree that such latencies are acceptable. A likely trend is that systems will enable users to specify needed levels of responsiveness and for the system to automatically monitor responsiveness and adjust its behavior and performance to stay within the specified parameters. In contrast, current systems mostly operate at maximum performance even when this greatly exceeds actual requirements.

5.1.3. Overall Cost

Some methods to reduce power consumption do not increase product cost. Other methods do increase cost but will typically only be implemented when the savings are considerably larger and offset the increased cost. Savings can result from reduced hardware purchases, reduced maintenance costs, reduced IT energy costs, reduced power and infrastructure operating costs, and avoided or deferred facility upgrades and expansions.

More efficient hardware may cost modestly more than the alternatives because of extra, sophisticated, higher-quality components or more rigorous testing. Virtualization software (and hardware) adds cost, and the complexity of power management could increase time to market for products and so indirectly increase costs and delay the onset of the savings. Other solutions (e.g., distribution of high-voltage DC current) could lower hardware costs, save large amounts of energy, and improve reliability.

An aspect of integrated circuit production that is not generally apparent is a distribution in efficiency among seemingly identical components. For example, one product off the production line might consume 10 percent less than average, and the next product could consume 10 percent more. Variation among production lines of “identical” products is often even larger. This variation could become an explicit part of the market with more efficient products costing more (similar to the current phenomenon of processors that run at faster-than-normal speeds carrying a price premium).

Some operational changes may increase costs by requiring more staff time to implement and monitor than the traditional methods of greatly over-provisioning capacity.

Some energy-efficiency measures actually reduce the cost of building a new data center. Examples are supplying DC power (avoiding some power conversion and cooling equipment), sizing for the right load (and providing ways to efficiently add on later), improving system design and layout, and using free cooling (to reduce or eliminate the need for chillers).

As noted above, all of these potential cost increases are expected to be much less than the value of lifetime energy savings and avoided site infrastructure costs from more efficient products, often with a very short payback period, so the more efficient systems have a lower total cost of ownership (TCO). In general, however, it is not possible to draw general conclusions about the additional cost of energy-efficient products because there is significant variation among product types, the efficiency strategies applied vary among products, and the rapid pace of change in technology makes it difficult to collect cost data.

5.1.4. Reliability

Manufacturers have reliability standards in their equipment designs, and it is safe to assume that these will be maintained at current levels for new equipment. However, the following strategies that can be applied to improve energy efficiency may also affect reliability:

Energy efficiency strategies that may increase reliability

- Lower interior temperatures,
- Lower data center temperatures overall, and
- Reduced cost to provide for redundancy.

Energy efficiency strategies that may reduce reliability

- Increased system complexity,
- More dynamism in system states and activity, and
- Higher average utilization.

One example of the relationship between reliability and product design is the spinning down of disk storage drives during long periods of non-access to save energy. If drives were not designed for routine spin-up/spin-down, this could result in a reduction in reliability. However, many drives are designed with this usage in mind so that frequent spin-down will not reduce (and may even increase) overall reliability. Today, only lower-speed drives are available that can spin down, not “server-class” drives, so this strategy can only be used in situations where data access speeds are not critical.

Many of the quantitative data about reliability and its relation to performance are anecdotal and/or proprietary. An exception is the study by Pinheiro et al., (2007) on failure trends in disk drives. They noted that the literature on disk drive failure patterns is sparse and found that “temperature and activity levels were much less correlated with drive failures than previously reported.” Moreover, Guha and Ouderkirk (2006) found that actual drive failure rates in MAID systems were reported to be significantly lower than the rates experienced in traditional always-spinning drives. Similarly, Bodik et al. (2006) argue that efficiency improvements will actually increase reliability because the reliability gain from not running equipment continuously can more than offset the reliability reduction from on/off cycles. This is an indication that some feared reliability risks of efficiency measures may be illusory.

Server power supplies interact with reliability in several ways. For example, removing the AC/DC conversion and directly feeding DC could improve reliability and improve efficiency (fewer potential points of failure), or it could reduce reliability (if there are fewer redundant components used in the system).

An increasing number of operators utilize redundant servers and data centers so that eliminating dual power supplies could improve efficiency and decrease local — but not overall — reliability, assuming the DC system was designed to provide the same level of redundancy. Virtualization technology can facilitate easy movement of applications among servers to increase reliability.

More efficient cooling inside the chassis of IT equipment can support both efficiency and reliability (especially in blade servers). This can be accomplished by use of efficient fans, variable-speed drives, aqueous or dielectric liquid cooling, and heat pipes. In addition, some of these solutions could be coupled with efficient building systems to effectively move heat out of the building, which might reduce building infrastructure cost as well.

Ultimately, however, equipment failure modes need to be analyzed as part of understanding the effect of energy efficiency on reliability. In practice, this analysis is very complex as there are many possible failure modes, and this analysis is specific to the particular product being examined. For this reason, it is not possible to undertake a complete analysis here; instead, this report treats qualitatively some of the factors that affect reliability.

5.2. **Impact of Facilities Energy Efficiency on IT Equipment Performance**

Energy efficiency-motivated changes to facilities design and operation, particularly power infrastructure and cooling systems, can affect the reliability of IT equipment. In fact, the tight environmental condition limits common in data centers were instituted originally out of concern about reliability. Sophisticated analysis is required to determine how adverse environmental conditions can lead to failures because there are many possible failure modes, only a subset of which are shown in Table 5-1. For this reason, it is only possible here to discuss general aspects of the factors that affect reliability.

Generally, achieving reliability through redundancy in HVAC and electrical systems decreases efficiency, but in facilities with “excessive” redundancy (i.e., more than is required to meet the desired reliability requirements), efficiency may actually be improved by implementing a more appropriate redundancy scheme. Different redundancy strategies vary in energy efficiency and in level of backup provided. Another energy-efficiency opportunity is evaluating actual mean time to failure for data center components, and only providing redundancy on the equipment that fails most often and on the equipment most critical to the online reliability requirements of users or installation.

New data centers offer an opportunity to achieve energy efficiency through system optimization, potentially resulting in large capital cost savings for avoided power and cooling infrastructure. Menuet and Turner (2006) note that with higher power densities of IT equipment, there is a greater reliability risk from power outages. Often, back-up power for IT equipment is maintained by a UPS and on-site generator, but power for the cooling equipment is not backed up in the same way. Under these conditions, the temperature that a piece of higher-density IT equipment experiences will rise quickly during a grid power outage, so there is little time before critical equipment temperatures are reached. Energy-efficient IT equipment, by contrast, produces less heat (per unit of computing output) and will exhibit a slower temperature rise. Efficiency from virtualization—which concentrates computing—could, however, exacerbate this effect.

Table 5-1. Example Equipment Failure Modes and Their Causes

Environmental Condition	Failure Mode
Low Relative Humidity	Electrostatic Discharge
High Relative Humidity	Hygroscopic Dust Failures Corrosion
Airborne Particles	Hygroscopic Dust Failures Corrosion
Low Temperature	Timing Hygroscopic Dust Failures
High Temperature	Diffusion Corrosion

Temperature Cycling	Thermo-Mechanical Fatigue
Temperature Shock	Multiple
Volatile Organic Compounds	Contact Erosion

Source: Herrlin 2006

A final observation is that it is common for IT managers and data center operators to simply build the most reliable data center system they can afford, without carefully analyzing the reliability requirements of the IT service being offered. Data center managers are inherently reluctant to take risks; perceived reliability issues are perhaps more of a problem than actual reliability issues. Lack of reliability data perpetuates a resistance to change, and organizational structures are another key impediment to progress. In this context, lack of reliability data is a significant barrier to adoption of energy-efficient technologies. This problem is discussed in Chapter 7.

5.3. Summary

As data center equipment and operation move toward greater efficiency, ensuring that high performance and reliability standards are cost effectively maintained will be challenging but readily achievable. Changes that would compromise reliability will generally not be implemented. A key research/policy need is to identify areas of uncertainty about performance and reliability and provide information that manufacturers, purchasers, and operators can use to confidently make decisions.

6. Distributed Generation and Combined Heat and Power Systems in Data Centers

This chapter addresses the Public Law 109-431 requirement for an:

- Analysis of the potential cost savings and benefits to the energy supply chain through the use of fuel cells for backup power and distributed generation.

In addition to analyzing the use of fuel cells for backup power and distributed generation (DG), this chapter also analyzes comparable DG technologies, such as reciprocating engines and gas turbines, for use in data centers. These DG resources can save significant energy in data centers, particularly when integrated with other energy-efficiency measures. DG resources used in CHP systems that use waste heat to provide cooling can provide the greatest benefits and lowest paybacks throughout the supply chain and are ideally suited to the steady power and cooling loads of data centers. For instance, in a CHP configuration, fuel cell systems can have paybacks of about 10 years when state and utility incentives are included whereas gas turbine and microturbine systems can have paybacks of less than five years. Clean DG – whether fuel cells or other clean energy prime movers – can also have environmental benefits both regionally, in the reduction of criteria pollutants that otherwise would be emitted by grid-generated electricity, and globally, in the reduction of greenhouse gas emissions. In addition to these financial and environmental benefits, DG systems can also significantly improve the reliability of data center power supplies.

The DG technologies discussed in this chapter include fuel cells (typically 10 kW to 2 MW in size), microturbines (30 kW to 250 kW), gas turbines (500 kW to 20MW), and reciprocating engines (100 kW to 3 MW). Solar photovoltaic power (PV) could be a supplementary source of power as well, but because of the intermittent and “as-available” nature of PV, it would not be considered a reliable, primary source of power for a data center. The focus of this section is on DG that can be dispatched on demand to support both the energy and reliability needs of data centers. It is also important to note that the optimal use of DG/CHP in data centers requires that all cost-effective efficiency improvements in server and building operations be implemented before consideration of a DG/CHP investment.

6.1. *Benefits of Clean DG and CHP for Data Centers*

Use of DG and CHP in data centers result in cost savings to the facility operator in the form of:

- Reduced energy-related costs and enhanced economic competitiveness – from reduced fuel and electricity purchases, resulting in lower operating costs;
- Increased reliability and decreased risk from outages – as a result of reliable on-site power supply; and
- Increased ability to meet facility expansion timelines – by avoiding the need for utility infrastructure upgrades.

In addition to these cost savings, DG/CHP provide the following benefits to the energy supply chain:

- Increased economic development value – through energy cost savings, which can make businesses more economically competitive in a global market, maintaining local employment and economic health;
- Increased energy efficiency – from generating electricity and useable thermal energy from a single fuel source;
- Reduced greenhouse gas emissions – from reduced fuel use, which results in lower levels of fossil fuel combustion and reduced emissions of CO₂;
- Reduced emissions of criteria air pollutants – from reduced fuel use and cleaner technology, which result in lower emissions of carbon monoxide (CO), nitrogen oxides (NO_x), and sulfur dioxide (SO₂); and
- Increased reliability and resource adequacy for the grid – through customer-sited generation, which can provide support and stability to the distribution grid and reduce or defer the need for regional power plant and transmission construction.

6.1.1. Reduced Energy Costs

Because of their very high electricity consumption, data centers have high power costs. Installation of CHP systems with absorption cooling can often reduce energy costs by producing power more cheaply on site than can be purchased from the utility supplier. In addition, waste heat from the power generation can drive absorption chillers that displace electric air conditioning loads.

Table 6-1 shows simple annual savings and paybacks for four actual CHP systems installed in California and the northeast. Capital costs for the first three DG systems were based on average capital costs for completed installations under the California Self Generation Incentive Program (SGIP) plus an assumed capital increase for absorption cooling of \$1,200/ton based on the cooling capacity required for each system. Capital costs for the gas turbine were based on an estimated price for a complex installation with a double-effect absorption chiller (\$350/ton), selective catalytic reduction (SCR) and CO control with a continuous emissions monitoring system (CEMS), and average U.S. construction costs.

Given the performance of these systems and a favorable differential between the price of fuel and the price of electricity (called the spark spread) that can be found in the Northeast or in California, annual energy cost savings range between \$239/kW and \$697/kW. For example, one data center in New York recently installed a 1.4-MW fuel cell CHP system that saves \$680,000 per year on the center's energy costs.

The high cost of fuel cell systems makes payback periods longer and more difficult to justify than for other efficiency improvements. Very generous incentive programs are available in some areas, such as the California SGIP fuel cell incentive of \$2,500/kW, which can be combined with the federal tax credit for fuel cells of \$1,000/kW. Microturbines are eligible for smaller incentives under both of those programs.

Table 6-1. Energy Cost Savings Comparison for CHP in Data Centers

CHP System	Molten Carbonate Fuel Cell/Chiller	Phosphoric Acid Fuel Cell/Chiller	Microturbine/Chiller Package	Gas Turbine/Chiller
Capacity, kW	1,000	200	200	3,364
Heat Rate, Btu/kWh	8,060	9,480	14,300	13,930
Capital Cost, \$/kW	\$7,238	\$7,805	\$4,088	\$2,312
Total Capital Cost	\$7,238,000	\$1,560,900	\$817,600	\$7,778,200
O&M Cost*, \$/kWh	\$0.032	\$0.029	\$0.022	\$0.022
Annual O&M Cost	\$266,304	\$48,268	\$36,617	\$615,895
Annual Gas Cost**	\$503,065	\$118,339	\$178,507	\$2,924,767
Annual Avoided Electricity** Savings	(\$1,103,497)	(\$220,756)	(\$354,668)	(\$5,153,526)
California SGIP	(\$2,500,000)	(\$500,000)	(\$160,000)	(\$800,000)
Federal Fuel Cell Tax Credit	(\$1,000,000)	(\$200,000)	(\$40,000)	
Net Capital Cost	\$3,738,000	\$860,900	\$617,600	\$6,978,200
Net Unit Capital Cost, \$/kW	\$3,738	\$4,305	\$3,088	\$2,074
Payback without Incentives, years	21.7	28.8	5.9	4.8
Payback with Incentives, years	11.2	15.9	4.4	4.3

*O&M costs are based on having an annual service contract in place, as is typical; no facility staff would be required to provide system maintenance or service.

**Based on gas price of \$7.50/MMBtu and electricity price of \$0.13/kWh reflective of electricity price in California and the northeast.

6.1.2. Increased Reliability

Data centers require both high-quality and extremely reliable power. Of all customer types, data centers, telecommunication facilities, and other mission-critical computer systems have the highest costs associated with power outages or lapses in power quality, ranging from \$41,000 per minute for cellular communications to as high as \$30 million dollars per minute for data center operations during peak periods (Bryson et al. 2001, Digital Realty Trust 2007).

Data centers almost always have UPS systems to condition power and eliminate momentary outages, sags, surges, and other deviations from a clean, in-phase sinusoidal power signal. Data centers often have more than one utility feed and associated seamless switching equipment to increase the reliability of service. Battery backup is generally used to provide a short-term outage ride-through of a few minutes to an hour. Longer-term outages are typically handled with standby diesel generators.

On-site power generation, whether it is an engine, fuel cell, microturbine, or other prime mover, supports the need for reliable power by protecting against long-term outages beyond what the UPS and battery systems can provide. DG/CHP systems that operate continuously provide additional reliability compared to emergency backup generators that must be started up during a utility outage. Backup generators typically take 10 to 30 seconds to pick up load in the case of an outage. The increased reliability provided by continuously operating DG/CHP allows for a reduction in the amount of battery backup that is typically designed into premium secure power systems because, for example:

- A DG system that is already running during a grid power disturbance does not need to be started. Therefore, the likelihood of a non-start is eliminated.
- A continuously operating DG system serves the function of a second power feed, significantly reducing non-available power times and reducing the need for battery backup.
- A continuously operating DG system can provide voltage support for the facility, reducing the likelihood of voltage sags and other power-quality disturbances from the grid.

The availability factor²¹ of DG/CHP systems is an important component in determining overall system reliability. The most recent comprehensive review of DG/CHP availability was conducted for Oak Ridge National Laboratory in 2003 (Energy and Environmental Analysis 2004). Of the systems studied, reciprocating engine availability factors averaged 96 to 98 percent. Gas turbines had availability factors from 93 to 97 percent. Fuel cells, which are a new-market entrant, have demonstrated the potential to achieve comparable availability factors as the fuel cell market matures and the servicing infrastructure becomes more readily available.

Reliability of power supply to the facility can be increased by using parallel utility feeds and on-site DG/CHP generation. A typical onsite DG/CHP system, as described above, could have an availability factor of 97 percent; a utility service feed availability is roughly 99.7 percent. The availability of power with onsite DG/CHP plus one independent utility feed would therefore be 99.97 percent, resulting in 43 minutes of expected outages per year. With two independent utility feeds plus CHP, the availabilities would be at a level of 99.99998 percent with only 7 seconds of expected outages per year. These short-term power supply outages can be handled with on-site energy storage, typically batteries.

6.1.3. Facility Expansion

Developing on-site power sources gives data center operators increased flexibility in both the expansion and design of new facilities. Upgrading older, smaller data centers with new equipment can result in a large increase in power demand to the facility that the utility might not be able to meet in the near term. Incorporating continuous prime power DG/CHP options can facilitate expansion and facility development on a more rapid schedule than can sometimes be possible by relying solely on the existing utility grid. Minimizing external power demand also reduces additional utility infrastructure requirements and associated costs that might be required for new or expanded facilities.

6.1.4. Economic Development

When operating costs are reduced and economic value is added a facility, such as through reduced outage costs, the facility becomes more economically competitive. This economic development value can be important to facilities that are competing globally. Reduced operating costs translate into greater profitability and more competitive pricing of services.

²¹ The availability factor is the proportion of hours per year that a unit “could run” (based on planned and unplanned maintenance), divided by the total hours in the year.

Increasing the economic competitiveness of a facility's operation benefits not only the facility operator but also the employees and the surrounding economy. Money generated by the facility in terms of wages and materials purchased creates a multiplier effect for the local economy.

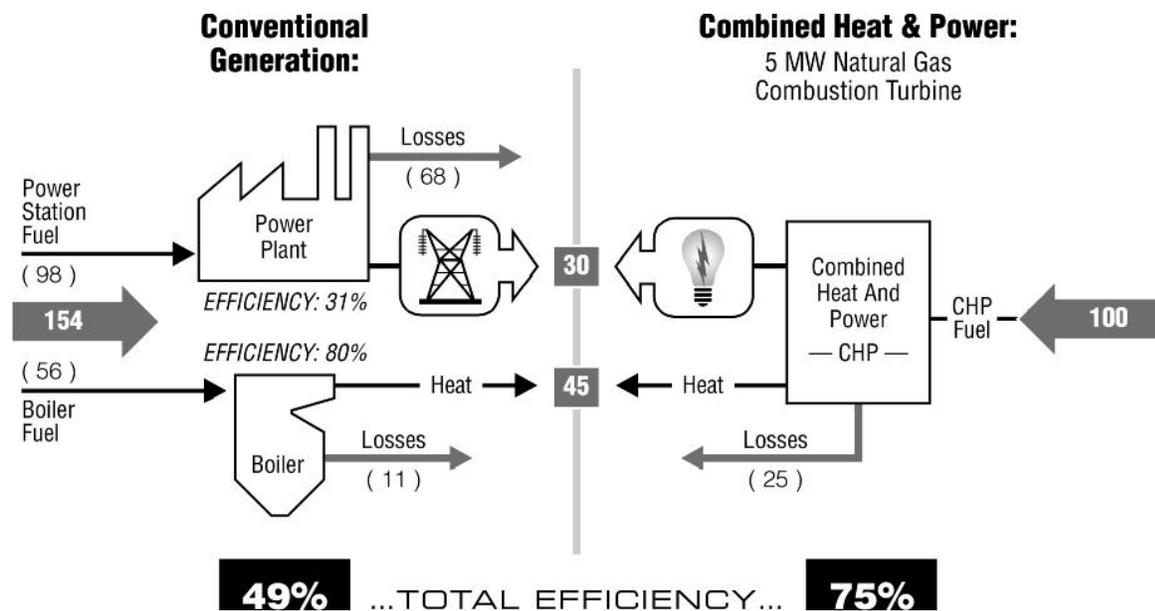
6.1.5. Increased Energy Efficiency

Energy efficiency reduces dependence on fossil fuels, decreases emissions of greenhouse gases and other air pollutants, reduces pressure on energy prices, and increases economic viability and sustainability.

The average efficiency of fossil-fuel power plants in the U.S. is 33 percent and has remained virtually unchanged for four decades. This means that two-thirds of the energy in the fuel is lost—vented as heat—at most U.S. power plants. About 5 to 10 percent of the electricity generated by central plant power stations is also lost before it reaches an end user as a result of resistance in the transmission and distribution (T&D) system.

The alternative to these line losses is to generate electricity at or near customer load centers and to capture and use the heat energy resulting from electricity generation. By using waste heat recovery technology, CHP systems typically achieve total system efficiencies of 60 to 80 compared to only 49 percent when electricity and thermal energy are produced separately (see Figure 6-2). In data centers where the thermal load is almost entirely cooling rather than heating, CHP can still provide an overall efficiency advantage. The waste heat from the generator is used in absorption chillers to produce cooling, which displaces electricity-powered electric chillers rather than displacing direct fuel purchases for heating. In addition to the electricity produced by the generator, the total electricity displaced by a combined cooling and power system can be up to 35 percent of the on-site generator capacity. Therefore, the total electricity provided and displaced by a combined cooling and power system can be up to 135 percent of the on-site generator capacity.

Figure 6-2. Combined Heat and Power Efficiency Advantage Compared to Central Power and Onsite Boiler Fuel



6.1.6. Emissions Benefits

Data centers have very low site emissions, but their high energy intensity results in a large emissions signature from a source perspective (i.e., the utility's emissions to produce electricity for the data center). CHP systems reduce emissions of criteria air pollutants—CO, NO_x, and SO₂—through increased efficiency and the use of cleaner technologies. Emissions reductions can be particularly significant when state-of-the-art CHP equipment replaces outdated and inefficient existing equipment (generally used for backup power) at the site. In fact, fuel cells have such low emissions that they can emit fewer criteria pollutants in a year as the primary power source for a data center than a diesel generator will emit in 24 hours operating as a backup generator. CHP also reduces emissions that contribute to global warming; increased fuel efficiency allows facilities to achieve the same levels of output with lower levels of total fossil fuel combustion, reducing CO₂ emissions.

Table 6-4 compares annual emissions from four CHP options with central station generation:

1. 1,000-kW molten carbonate fuel cell with a double-effect absorption chiller,
2. 200-kW phosphoric acid fuel cell with a single-effect absorption chiller utilizing roughly half of the available thermal energy from the system,
3. 200-kW microturbine system with a double-effect absorption chiller, and
4. 3,400-kW gas turbine system with a double-effect absorption chiller and emissions after treatment (SCR and CO reduction) with a continuous emissions monitoring system.

The CHP emissions factors are shown on a per-unit-of-power-output basis and incorporate a credit for the displaced emissions from avoided air conditioning load met by the absorption chillers. The CHP emissions are compared to the average emissions from fossil-based power generation in the U.S.²² Each of the four systems shows reductions in NO_x, SO₂, and CO₂ compared with the average for central fossil-based power generation.

Table 6-5 compares two backup power options—a proton exchange membrane (PEM) fuel cell and a standard diesel engine.²³ The table shows significant emissions reductions from switching to a PEM fuel cell from a diesel engine even if only in backup power mode.

What is most striking in comparing Table 6-4 and Table 6-5 is that the backup diesel engine has higher emissions of criteria pollutants from operating just 24 hours during the year than the fuel cell CHP systems show for operating continuously throughout an entire year.

²² Combustion-based power generation uses fossil fuels, biomass, and waste fuels. Nuclear, hydro, solar, and wind power are not included because these sources would typically not be displaced by DG.

²³ Diesel technology emissions can be reduced through the use of low-sulfur diesel fuel and incorporation of engine modifications and exhaust after-treatment. These technologies are being developed for the transportation market, but they are not currently being used for standby engines.

Table 6-4. Emission Benefits of Combined Heat and Power

	Combined Heat and Power				Central Power
	Molten Carbonate Fuel Cell/Chiller	Phosphoric Acid Fuel Cell/Chiller	Microturbine/Chiller Package	Gas Turbine/Chiller	U.S. Fossil Power Average*
Capacity, kW	1,000	200	200	3,364	
Heat Rate, Btu/kWh	8,060	9,480	14,300	13,930	
CHP Emissions Factors (for generated electricity minus avoided air conditioning emissions)					
NO _x , lb/MWh	0.0058	0.0273	0.2583	0.0751	N/A
SO ₂ , lb/MWh	0.0003	0.0003	0.0003	0.0376	N/A
CO ₂ , lb/MWh	848	998	937	1056	N/A
Annual Emissions (based on 8,760 hours/year)					
NO _x , lb/MW-year	51	239	2,262	658	21,725
SO ₂ , lb/MW-year	2.36	2.77	2.60	329	44,501
CO ₂ , tons/MW-year	3,716	4,370	4,103	4,626	6,899

*Based on 2000 eGRID emissions data.

Table 6-5. Emission Benefits of Fuel Cell Backup Power

	PEM Fuel Cell	Diesel Generator
Capacity, kW	150	600
Heat Rate, Btu/kWh	9,750	10,000
Emission Factors (generated electricity)		
NO _x , lb/MWh	0.100	20.282
SO ₂ , lb/MWh	0.006	2.900
CO ₂ , lb/MWh	1,170	1,650
Annual Emissions (based on 24 hours/year)		
NO _x , lb/MW-year	2	487
SO ₂ , lb/MW-year	0.14	69.60
CO ₂ , tons/MW-year	14	20

6.1.7. Reliability and Resource Adequacy

DG/CHP can increase reliability and resource adequacy for the power grid as a whole, not just for the facility operator. Rapid growth in local power demand can create localized shortages of power and impact power quality and grid reliability, as discussed in Chapter 4. Considerable concern emerged about the concentration of data centers in the Silicon Valley area of California during the power and fuel crises of 2001. The problems led to local concerns regarding pollution from over-use of diesel generators, and to migration of new and some existing data centers out of the region in search of more reliable power supplies. Based on the ability of utilities to expand supply, transmission, and substation capacity, it could take several years before such shortages are eliminated. DG/CHP can reduce or delay infrastructure investments, making the grid more reliable for all customers.

6.2. DG Applications at Data Centers

DG can be applied in a variety of configurations to meet a hierarchy of facility needs, including standby/backup power, continuous prime power, and CHP (adapted from APC 2003). DG has been successfully employed in data centers using both fuel cells and other types of prime movers, such as reciprocating engines, gas turbines, and microturbines. Appendix 3 lists recent fuel cell installations in data centers throughout the country. It also includes three case studies of successful CHP system applications in data centers in New York and California.

6.2.1. Standby/Backup

The traditional emergency or backup generator is a form of DG. In standby mode, the electric utility is the primary source of power, and on-site power generation is used only as backup during a scheduled shutdown or failure of the utility feed(s). A UPS is used to bridge the time delay while the standby system starts. This mode of operation is used in more than 99 percent of network rooms and data centers that have local power generators (APC 2003).

The standard generation technology for this mode is a diesel generator. Emergency diesel generators are relatively inexpensive, can pick up load rapidly on start-up, and are reliable if properly maintained. Lack of maintenance and testing, however, can reduce system reliability and lead to the engines failing to start when needed. In addition, diesel generators for emergency use emit significant pollution and can produce visible smoke, noise, and odor that can lead to local complaints when a facility tests or is forced to use their generators.

Some data centers are considering a switch to cleaner technologies for backup power, including gas- or propane-fired engines, microturbines, or fuel cells. These technologies take longer to start and pick up load than a diesel engine, so they might require additional energy storage in the form of batteries or flywheels. Fuel-cell-based backup power systems are just emerging as commercial products. Fuel cells for backup power are being developed by a number of companies with remote telecommunications facilities as an early target market. In 2005, a fuel-cell product was introduced to the market that provides backup power to data centers using bottled hydrogen as a fuel. Another fuel cell product on the market fits within a single 42U rack and provides 30 kW of backup power. Fuel cell technology is currently aimed at data center applications where conventional backup power is impractical and where backup needs are longer than can be provided by conventional diesel generator-based UPS systems.

6.2.2. Continuous Prime Power

In continuous prime power mode, DG is the primary source of power, and utility-supplied power is used primarily as a backup during a scheduled shutdown or failure of the on-site generator. Use of the utility feed in this way is not free, but most utilities have established standard tariffs for providing standby power to a customer that has its own generator. In some states, these rates are high enough to discourage customer-sited generation; in other states standby tariffs are more favorable to DG/CHP. UPS and battery backup are used to maintain system power during switchover from on-site to utility power and to maintain power quality. Using DG in continuous prime power mode without capture and use of waste heat can be cost effective in areas with high overall electricity costs and/or high demand charges. (One example of a prime power installation is a 200-kW fuel cell system installed at the Hamilton Sundstrand data center in Windsor Locks, Connecticut.)

One efficiency measure that works particularly well for continuous prime power integrated with DG systems is the use of direct current (DC) power as the underlying power backbone (Robertson 2003). In a DC-based system, a dual DC ladder bus aggregates all of the outputs of the distributed generators and provides multiple paths for the power supply to reach the loads. The system can use any combination of generators—turbines, natural gas reciprocating engines, diesel engines, microturbines, fuel cells, or even photovoltaics—operating at any voltage or frequency. Distribution feeders from the utility can be a source of generation, too. A major advantage of this architecture—demonstrated through research and development and limited commercial installation—is the elimination of all paralleling and switching required by traditional alternating current (AC)-based DG and backup power systems.

6.2.3. Combined Heat and Power (CHP)

CHP mode can be configured in the same ways as prime power mode but uses available heat from the on-site generator to meet facility heating and cooling loads. As previously stated, data centers require so much power—all of it converted to heat as it is used within the facility—that the most useful CHP thermal configuration provides thermally activated air conditioning.

Heat recovered from the on-site power generation unit in the form of steam or hot water can be utilized in an absorption chiller. High-temperature hot water near or above the boiling point (under pressure) can power a single-effect absorption chiller. A single-effect absorption chiller needs about 17,000 Btu of high-temperature hot water or low-pressure steam to produce a ton of cooling (12,000 Btu). Double-effect chillers can produce a ton of cooling with 10,000 Btu of steam. Absorption chillers would be used in a data center with computer room air handlers, or, less commonly, the chilled water could be fed to water cooled racks. Absorption chillers require larger cooling-tower capacity than standard electric chillers and more pumping power.

Gas turbines, reciprocating engines, high-temperature fuel cells like solid oxide and molten carbonate, and microturbines are capable of producing steam that can be used in either a single-effect or a double-effect chiller. The commercial phosphoric acid fuel cell has a high-temperature heat recovery option that can produce 250 °F hot water for powering a single-effect absorption chiller. PEM fuel cells do not produce temperatures high enough to use in an absorption cycle.

Although this is not a widespread practice at this time, there are a number of commercial CHP installations in dedicated data centers or in office buildings, banks and communications facilities where data processing is a major activity within the building, as shown in Table 6-6. As shown, a variety of technologies have been used successfully, including fuel cells, reciprocating engines, gas turbines, and microturbines.

Table 6-6. CHP Installations in Data Center and Communications Facilities

Facility Name	City	State	Prime Mover	Capacity (kW)	Operating Year
Telecommunications Facility	Burlingame	CA	Microturbine	120	2003
Chevron Accounting Center	Concord	CA	Recip. Engine	3,000	1988
Guaranty Savings Building	Fresno	CA	Fuel Cell	600	2004
Citibank West FSB Building	La Jolla	CA	Microturbine	60	2005
QUALCOMM, Inc.	San Diego	CA	Gas Turbine	11,450	1983/2006
WesCorp Federal Credit Union	San Dimas	CA	Microturbine	120	2003
ChevronTexaco Corporate Data Center	San Ramon	CA	Fuel Cell	200	2002
Network Appliance Data Center	Sunnyvale	CA	Recip. Engine	825	2004
Flint Energies Service Center Facility	Warner Robins	GA	Fuel Cell	5	2002
Zoot Enterprises	Bozeman	MT	Recip. Engine	500	2003
First National Bank of Omaha	Omaha	NE	Fuel Cell	800	1999
AT&T	Basking Ridge	NJ	Recip. Engine	2,400	1995
Continental Insurance Data Center	Neptune	NJ	Recip. Engine	450	1995
Verizon Communications	Garden City	NY	Fuel Cell	1,400	2005

Source: Energy and Environmental Analysis CHP Database (2006)

Appendix 3 includes three case studies of CHP systems being used successfully at data centers in New York and California.

6.3. Issues Affecting Implementation of DG in Data Centers

Fuel cells and other forms of DG power and thermally activated cooling can provide the reliability and power-quality that data centers require while also providing environmental benefits and energy cost savings. However, widespread adoption of such systems in data center applications has been impacted by a number of issues:

- Power outage costs are so high that many facility operators are reluctant to deviate from the standard design of UPS, battery storage, and standby diesel generators. The failure modes of these systems are well known, and design to ensure reliability has become standard practice.
- Fuel cells in particular but also other DG systems do not have much of a track record in these high-power-quality, high-reliability applications; therefore, not all failure modes are completely known. (Specifically, failures from the interaction of the fuel cell or microturbine system power electronics with the UPS and switching systems are relatively untested.) Demonstration systems are in place throughout the country, which are providing important opportunities to prove reliability and improve operational practices.
- Fuel-cell DG systems have high costs. Higher production levels combined with engineering and materials advances are needed to bring costs down. In addition, not enough fuel cells have run for an extended duration to provide statistically significant results to show when stack replacement is necessary, which is a major cost.

- Other types of DG systems are more cost competitive, but these systems could also benefit from reductions in packaging, site engineering, and installation costs.
- Part of the value of CHP is the integration of thermally activated cooling from central chillers, typically used only in large facilities. Some work is needed to demonstrate the use of chillers in smaller facilities.
- Fuel-cell-based backup power systems are emerging as commercial products. These small fuel-cell systems tend to compete more with battery systems than with diesel generators.
- Battery systems represent a costly component of standard systems in terms of dollars and floor space. Integration of DG systems with other forms of energy storage, such as flywheels, could be attractive for facility operators to consider as a means to reduce required battery capacity.

For these reasons, continued R&D support and incentives are needed for fuel cells and other forms of DG to become more common in data centers. These needs are discussed further in Chapter 8.

7. Current Energy Efficiency Programs Applicable to Data Centers

This chapter addresses the Public Law 109-431 requirement for:

- An overview of current government incentives offered for energy-efficient products and services and consideration of similar incentives to encourage the adoption of energy efficient measures in data centers and servers.

For the past 30 years, government agencies and public utilities have promoted energy-efficient products and practices through incentive programs. These programs have been developed to address the “efficiency gap”: the tendency for consumers and firms to forego apparently beneficial and cost-effective investments in energy-efficient products and practices. There are many reasons for this behavior, which are collectively known as “barriers to energy efficiency.” This chapter first explores how these barriers manifest in the data center industry and then summarizes current government and utility programs and initiatives that have been developed to address these barriers.

7.1. *Barriers to Energy Efficiency*

Although there are many opportunities to save energy in data centers, which can produce significant benefits such as energy cost savings and improved reliability, many factors also impede the adoption of these efficiency measures, including first cost, lack of knowledge, institutional failures, and perverse incentives. Most of these barriers are common to both private- and public-sector data centers (Loper and Parr 2007).

7.1.1. *Defining Efficiency*

It is difficult to define energy efficiency for a complex system such as a data center. An energy efficient product or facility is usually defined as delivering the same or better service output with less energy input. Ideally, efficiency is represented by an energy-based performance metric in which the output is measured in terms of an energy service that end users value, in the same way that household cooling equipment is rated based on its thermal cooling output. However, a data center’s output is a complex mix of computational processing; data storage; network communication; and server availability, reliability and responsiveness. The performance metric that is used affects the various competitive interests of the suppliers of data center equipment; any definition of energy efficiency will benefit some suppliers and disadvantage others.

The difficulty in defining efficiency is especially true for a key piece of equipment in the data center – the server. Measuring a server’s output is difficult for several reasons: 1) for a given server, the output is dependent on the hardware and software configuration; 2) the primary application of the server (e.g., web applications, email, database, etc.) leads to different processing and memory requirements, which affect energy performance; and 3) an individual server can provide multiple services simultaneously. Nonetheless, significant progress has been made during the past year toward creating energy performance metrics for servers. This includes an agreement that was reached among the various stakeholders on standard assumptions about server capacity utilization for web servers (Kooimey et al. 2006). However, this agreement is

applicable to one type of equipment that serves just one type of function in the data center, so much more work needs to be done.

Additional challenges emerge in attempts to derive measurements of overall data center energy performance that include power distribution, cooling, and lighting. Different data centers will have different performance characteristics depending on their type of application, level of criticality, location, and size. LBNL has proposed energy-performance metrics based on the ratio of total data center energy consumption to the energy consumed by the servers. The higher this ratio is, the more energy is being transformed into waste heat or is being used to supply power to the servers themselves. Installing more efficient cooling, lighting, power supplies, and even servers would produce a lower ratio.²⁴ A measure along these lines may offer more short-term promise than a server metric.

Although it is a difficult process, defining “energy performance” is a prerequisite for implementation of many government and efficiency policies and programs. Energy performance measures are necessary to establish eligibility for utility rebates or tax incentives, for example, or to create procurement requirements.

Energy-efficiency metrics can also be used as a benchmark for individual facilities to gauge their performance against other similar facilities. Data center operators need to understand how much energy their equipment is using, how it fares against similar equipment, the trade-off between energy performance and other types of performance, and how these factors can be improved. These goals can only be achieved by defining a standard metric. In the absence of a metric, surveys show that while data center operators/owners identify IT power and cooling as the primary issue facing their data centers, most say that neither IT power consumption nor cooling concerns affect their server purchases (Ziff Davis 2005).

7.1.2. First Cost

Many data center efficiency improvements have a high initial cost, which are difficult for facility managers to justify, even if the projected payback period is only a matter of months. Money spent upfront on redesigning a data center seems more tangible than the money not spent on energy costs in the future. This view can also dissuade vendors and consultants from recommending cost-effective data center improvements. It is less risky for a vendor to offer a less-expensive, inefficient package of recommendations than to try to convince a data center operator that the high price of the efficient design will eventually be repaid through energy savings. Interviews with equipment vendors indicate that the efficient designs are a tougher sell and could result in the loss of a potential customer.

Despite the rising cost of electricity and the relatively large share of data center costs attributable to energy expenses, data center energy costs remain a relatively small portion of overall facility costs for even the largest data center operations. According to Yahoo’s chief financial officer (CFO), for example, while electricity represents 20-50 percent of the cost of running a data center, it is still a minor part of overall costs (Delaney and Smith 2006). If this is true for a company like Yahoo, whose entire business depends on servers, then data center energy costs

²⁴ A more efficient server would generate less waste heat that must be removed by the data center cooling system.

must garner even less attention in a more typical entity, for which data centers make up a smaller portion of overall operations.

7.1.3. Split Incentives

Attention to data center energy use is often motivated by electricity supply, cooling, and building space constraints than by electricity costs.²⁵ This is not surprising because most data center operators do not know how much their energy costs are (these costs are typically not reported separately on the energy bill for the larger facility). This is the barrier of split incentives.

Split incentives occur when the entity responsible for paying energy bills is different than the entity responsible for capital investment decisions. There are two common forms of split incentives that are relevant to data centers: when data centers operate in leased buildings, and when the data center operator is unaware of or indifferent to the energy use of the equipment.

Many data centers are housed in buildings that are not owned by the equipment operator. If the data center owner is not the sole tenant of a given building, the utility bills are frequently prorated based on square footage occupied rather than by actual electricity consumption of each tenant and thus may not represent the actual electricity consumption of each tenant's equipment. The high energy use per square foot of data centers is allocated among all tenants; thus, the data center's energy use is in effect subsidized by other tenants and building uses. Tenants of leased buildings generally have little incentive to make long-term investments in building infrastructure.

The second type of split incentive is especially problematic in companies or organizations in which the company that controls the facility is just hosting the IT equipment, or in which IT equipment and facilities are managed by separate organizations within the same company. In either case, while the IT side is responsible for purchasing and managing the IT equipment, the facility side is responsible for providing the power and cooling infrastructure and paying the energy bills. Under this arrangement, most IT managers never see the energy bill for their equipment, and their job performance is not evaluated based on energy costs (Dunn 2006a). While improved energy efficiency may benefit the organization overall, the data center manager will see little reward. Communication between the two areas is frequently limited, with little or no common management or interaction below the "C-level" (CFO/CIO/CEO) of the company.

7.1.4. Risk Aversion

The two biggest concerns of most data center operators are security breaches and interrupted operations. Measures to avert these risks frequently inhibit adoption of energy-saving measures.

Data center owners are often view their centers as competitive assets. Data center specifics, including size, makeup, and even location are often considered confidential. This constrains the willingness of organizations to make improvements to their data centers; the more people (including contractors) that work on a data center, the more likely it is that its closely guarded

²⁵ Several vendors (including representatives from APC, VMware, Siemon Company, and Emerson Network Power) at FOSE 2007 government technology conference repeated this claim when asked to explain the current focus on energy efficiency. Almost everybody mentioned that the principal reason for efficiency improvements in the federal sector that they were familiar with stemmed from a lack of electricity capacity.

specifics will be leaked to a competitor or, in the case of the government (especially the military), to an enemy.

As insurance against the outage of critical data center operations, most data centers have some level of system redundancy: if one server goes down, another server is positioned to take over the first server's functions. The costs associated with additional inefficient energy loads as a result of redundancy are perceived as small compared to the disaster of even a short outage, and, as noted elsewhere, the costs are probably invisible to the data center operator.

Eliminating redundancy will be impossible and probably undesirable. The key is to ensure that the redundancy is achieved in the most efficient way possible and that operators do not create more redundancy than is necessary, running servers whose sole function is to take over in case of failure in the principal servers, or cooling below even the overly conservative range suggested by manufacturers.

Although data center operators' job performance is rarely (if ever) based on the center's energy costs, interrupted operations resulting from attempts to institute new and untested software, hardware, or cooling innovations can threaten their jobs. It is thus not surprising that operators may not immediately adopt every new technology or idea. For example, a water leak in a liquid cooling system could be disastrous.²⁶ A data center operator is unlikely to be willing to pursue an innovation if there is any uncertainty in timely response by well-trained personnel to maintenance problems that might affect system reliability.

Equipment suppliers may also contribute to excessive system redundancy. Some application software vendors reportedly advise operators not to run on virtual servers because the applications have not been tested with virtualization software and thus have an unknown effect on reliability.²⁷ And manufacturer-specified nameplate rating for server power is often much greater than actual use because it is typically a safety limit. For instance, a facilities manager at a health care data center explained data center operators may have to specify 12 amps of power capacity even though the server really only draws 3 amps.²⁸

Successful adoption of new technologies and processes, especially by peers, is probably the best way to convince a data center operator to implement a new technology, especially given perceptions that energy-efficient equipment is less reliable than standard equipment (see Appendix 2, working group 1 notes). And, currently, these operators cannot independently test equipment without being subject to a vendor's sales pitch (see Appendix 2, working group 4 notes). Greater knowledge of the actual extent of the risk of interrupted performance resulting from implementation of a given improvement could make data center operators less cautious. There are technical solutions as well; for example, installation of energy monitoring equipment

²⁶ Not all liquid cooling systems are water based.

²⁷ This is not to say that this advice is correct. Virtualization software is being used by all of the Fortune 100 companies, as well as the federal government, and is supported by several major information technology companies.

²⁸ ASHRAE TC 9.9 addressed this issue by requiring server manufacturers to provide actual current draw along with nameplate ratings. This should consequently not be as significant a problem in the future. Nonetheless, without proper training and readily available information, this may be a difficult distinction for data center operators to make.

would enable data center managers to run closer to established power and cooling limits without fear of affecting data center availability.

7.1.5. Learning Curve

Data center operators are trained to manage servers and the IT equipment they support. Their primary concern is to ensure that the equipment runs, not that the equipment is energy efficient. Even if data center operators are aware of the benefits of energy efficiency, they may not have the know how to monitor and evaluate the energy performance of the data center's cooling system, lights, or power supplies. Acquiring that information can be time consuming and costly. Likewise, architects and others involved in the front-end design of a data center may not know enough about data center operations to be able to question IT requests.

When considering energy-efficiency measures, data center operators need sufficient background knowledge to be able to vet claims by vendors and then communicate the benefits, costs, and risks to financial decision makers. Creating a financial case for an efficiency improvement may be a greater hurdle to data center operators than implementation of the efficiency measure itself. Facilitating conversations among these groups of people and providing guidance that can be used in those communications is a critical need.

Data centers are also frequently treated as if they were any other aspect of commercial office space, rather than the energy-intensive operations they are. This can result in inappropriate application of designs and equipment that, in other circumstances, would be considered efficient, such as a typical office-building cooling tower. Careful and knowledgeable application of equipment is vital to achieving adequate energy efficiency (Brill 2007).

7.1.6. Quickly Changing Technology

One of the greatest barriers to improved energy efficiency in data centers is the rapid increase in new computer applications that has occurred as the cost of processing power has fallen and the rapid pace at which IT equipment becomes obsolete and is replaced. Because the components of HVAC systems last a long time (10 to 25 years) and IT equipment has a relatively short life (two to five years), HVAC systems often end up being mismatched with the cooling requirements of IT equipment (Schmidt and Beaty 2005). UPS systems also have a relatively long life compared to IT equipment, which can block the adoption of newer, more efficient technologies. It is difficult to continually optimize data centers in such a rapidly changing environment and even more difficult to predict future requirements (Schmidt and Beaty 2005).

Similarly, in the time it takes to construct a building, IT will have changed dramatically. Even the most conscientious data center designer may end up with a brand new, state-of-the-art facility designed to outdated criteria. The rapid pace of growth in data centers encourages data centers to be built with more space, cooling, and power supply capacity than is actually needed, resulting in less-efficient operation as equipment operates at partial capacity.

This has led some observers to recommend “tuning-up,” or recalibrating, equipment in data centers rather than building a more efficient substitute or replacing existing equipment with brand new, more efficient models. The criticality of data centers' function frequently inhibits facility managers from making major overhauls simply to realize energy-efficiency

improvements. Tuning up existing equipment can be done with relatively little risk. By some estimates, continuous commissioning in the average data center can result in 20 percent efficiency improvements with little capital investment (Brill 2007).

7.1.7. Lack of Energy Monitoring

A major barrier to improved energy efficiency is the difficulty of collecting data on the energy consumption of individual components of data centers, and the lack of data collection on many data centers overall. Most data centers have not been outfitted with advanced electric meters or other monitoring systems that accurately measure and report data centers' energy consumption and efficiency.

Better energy data collection would not only help to quantify the energy load of data center operations, thus highlighting the importance of energy-efficiency improvements and facilitating right-sizing of equipment to match the energy load, it would also allow data center operators to monitor and evaluate the energy savings resulting from specific energy-efficiency improvements. Seeing concrete savings may also spur data center operators to implement further efficiency measures.

7.2. *Current Energy Efficiency Incentives and Voluntary Programs*

Although specific aspects of the barriers discussed above apply uniquely to data centers, in general the barriers that impede the adoption of efficiency measures in data centers are similar to those in other sectors. Thus, existing energy policies targeting buildings and equipment – such as labeling programs, financial incentives, and government procurement – will likely also be applicable to data centers.

EPA's ENERGY STAR program is beginning the process of addressing energy performance of equipment in data centers by supporting development of energy performance measures for servers. ENERGY STAR also provides labels for some equipment, such as air conditioners, that might be used in data centers but are typically used in other space types.

Other equipment used in data centers, such as UPS systems, or computer room air conditioners, could benefit from ENERGY STAR labeling. LBNL developed a draft labeling criteria for UPS systems patterned after a proposed European standard. An ASHRAE standard exists that could be used as a basis for a CRAC labeling program.

DOE's Save Energy Now program can provide tools, training, best practice information and energy assessments directly to data centers or to companies owning multiple data centers for them to deploy.

Beyond federal programs, energy performance for data centers as a whole is just starting to be addressed by state governments and utilities. The New York State Energy Research and Development Authority (NYSERDA) is examining the feasibility of such a program now. One utility, Pacific Gas & Electric (PG&E), currently offers incentives for data center efficiency improvements, and another, Austin Energy, is in the process of launching a similar program. Utility programs may be further encouraged by energy-efficiency portfolio standards, which

have been instituted in several states. In addition, EPA and many utilities run programs promoting CHP, which is applicable to data centers. All of these activities are described in more detail below.

7.2.1. Financial Incentives

The financial incentives offered by governments to spur energy-efficiency investment come in a variety of forms. They are most commonly tax incentives – e.g., income tax credits or deductions, accelerated depreciation, sales tax exemptions – and special project financing. Financial incentives can be useful for gaining the attention of financial managers and other supervisors in an organization that might not otherwise be concerned about a facility's energy consumption.

7.2.1.1. Federal Tax Incentives

The Energy Policy Act of 2005 (EPAct 2005, Public Law 109-58) created a number of new tax incentives for energy-efficient buildings, equipment, and vehicles. Prior to EPAct 2005, federal tax incentives for energy efficiency were limited to a provision that allowed individuals to exclude utility incentives for efficiency measures from taxable income and a \$2,000 deduction for the purchase of clean vehicles (including hybrid-electrics).

Of the EPAct 2005 tax incentives, only the commercial building provision (EPAct 2005, section 1331) appears applicable to data centers. It provides a tax deduction of \$1.80 per improved square foot to owners of new or existing buildings who construct or reconstruct their buildings to reduce the building's total heating, cooling, ventilation, water heating and interior lighting energy cost by 50 percent or more compared to the ASHRAE Standard 90.1-2001 reference building.

Partial deductions of \$0.60 per square foot are available for owners installing improvements to the building envelope, lighting, or HVAC of a building that reduces total heating, cooling, ventilation, water heating, and interior lighting energy by 16.7 percent compared to the ASHRAE reference building (The Tax Incentives Assistance Project 2007). An interim lighting goal exists which allows building owners to receive a prorated deduction of \$0.30 to \$0.60 per square foot for reductions in the building's lighting power density of 25 percent to 50 percent beyond the ASHRAE reference building (NEMA 2006).

7.2.1.2. State Tax Incentives

At least four states (Maryland, Nevada, New York, and Oregon) offer financial incentives to commercial buildings that make energy-efficiency improvements. New York and Maryland offer relatively similar green building tax credits for building owners and tenants whose buildings or tenant space meet certain green standards. Maryland's credit is based on the U.S. Green Building Council's Leadership in Energy and Environmental Design (LEED) standards and is worth up to eight percent of the total cost of the building (MEA 2006). New York's credit can be worth up to \$2 million per applicant and references both LEED standards and New York State building codes (NYDEC 2007).

Nevada also offers an incentive for green buildings, in the form of a partial property tax exemption of up to 50 percent for up to 10 years for buildings achieving LEED-silver ratings (State of Nevada 2007). Oregon provides business energy tax credits for investments in energy-efficient equipment and sustainable buildings, equal to 35 percent of the incremental cost of the efficiency investment (Oregon DOE 2007). As all of these incentives reference LEED, and LEED does not currently have any data-center-specific requirements (although it is beginning to consider some) (Fuertes 2007), it is unclear whether improvements in data center efficiency would be eligible for these credits and exemptions.

7.2.1.3. State Loan Programs

At least 19 states offer no- or low-interest loans (usually around three to five percent per year) for energy-efficiency improvements. As of 2002, state loan programs had commitments in place for more than 24,000 energy-efficiency projects worth \$620 million (ASE 2002). Most of these loan funds are available only for improvements in public buildings or entities although some states also make them accessible to non-public entities, especially non-profit organizations. Eligibility is usually not very restrictive; data center projects may be eligible. Typical projects include HVAC system improvements, lighting improvements, motor controls, building envelope modifications, and commissioning.²⁹

7.2.2. Labeling and Benchmarking

Energy labeling programs have long been a key part of federal and state government efforts to encourage greater voluntary energy efficiency. Labels make customers more aware of the energy costs associated with their purchasing decisions and encourage consumer acceptance and recognition of high-efficiency products, which in turn encourages manufacturers to make more of them. Labeling programs complement regulatory policies such as standards and codes, which prohibit sales of the least-efficient product models. They can also work well with consumer awareness programs that “pull” the market forward by encouraging purchase and manufacture of high-efficiency products (Loper et al. 2005).

The ENERGY STAR product labeling program (www.energystar.gov/productdevelopment), administered by the EPA and DOE, offers energy-efficiency performance criteria for more than 50 residential and commercial product categories. The ENERGY STAR mark represents the most energy-efficient models found in the marketplace and is now recognized in the U.S. and worldwide as the symbol for energy efficiency. In the U.S., more than 60 percent of American households recognize the ENERGY STAR mark. Awareness is even greater – more than 70 percent – in areas where energy-efficiency program sponsors are actively promoting ENERGY STAR. On a more global level, EPA has ENERGY STAR partnerships with Australia, Canada, the European Union, New Zealand, and Taiwan. These partnerships are intended to unify voluntary energy-efficiency labeling programs in major global markets and make it easier for partners to participate by providing a single set of energy-efficiency qualifications instead of a patchwork of varying country-specific requirements.

The ENERGY STAR label, for some product categories, is also the basis for determining eligibility for state tax incentives in at least three states – Connecticut, Minnesota, and Maryland

²⁹ See Appendix 5 for more details about these energy-efficiency financing funds.

– and in Washington D.C. In addition, a number of states and cities follow federal procurement practices and require purchase of ENERGY STAR products wherever practical (ASE 2005, Harris et al. 2004). Many utilities, including in California, also use ENERGY STAR criteria to determine eligibility for rebates, particularly for appliances and heating and cooling equipment (ASE 2005, Brown et al. 2002, CEE 2007a, DSIRE 2007).

An important consideration for the program is that the credibility of the ENERGY STAR label depends on product eligibility requirements that are meaningful. If all models in a product category are labeled as ENERGY STAR, either because of “eligibility creep” as markets change over time or to attempts to accommodate too many industry definitions of energy efficiency, the label will not usefully differentiate products for buyers. In contrast, if the eligibility requirements exclude some products that could reasonably be considered efficient by another definition, the label’s usefulness is compromised. These will be the major challenges associated with the eventual development of standards for servers and data centers as a whole.

In addition to labeling products, the ENERGY STAR program offers benchmarks and labels for whole building energy use. The national energy performance rating system assesses commercial buildings and expresses their energy performance on a scale of 1 to 100. This rating is based on measured energy consumption and is normalized to adjust for specific operating characteristics and geographical location. By assessing the entire building, the benchmark is able to provide a holistic evaluation of performance that accounts for the interactions among various system components. Buildings that earn a 75 or higher are in the top quartile of their peer group and are eligible to apply for the ENERGY STAR label. To date, more than 3,200 buildings have earned the ENERGY STAR rating and more than 26,000 have used EPA’s on-line tool to benchmark energy performance on the 1-to-100 scale.

The normalization performed in the EPA whole-building assessment is based on the type of commercial building. At this time, space types that can be rated using the EPA tool represent more than 50 percent of the commercial building stock. Available spaces are: offices, K-12 schools, hotels, supermarkets, hospitals, warehouses, dormitories, and medical offices. Because of data limitations, there is no current benchmark for data centers.

7.2.2.1. Labeling Servers and Data Centers

Development of labels for data centers and data center equipment is still in its nascent stages. The challenge is that, as discussed earlier, developing performance metrics for computer equipment can prove difficult and controversial (PowerPulse.net 2006). For example, ENERGY STAR took more than two years to revise the computer specification for single-user desktop and notebook computers). In the past, the ENERGY STAR computer specification only addressed low-power modes, i.e., when the computer is not being used for its primary purpose. But the new specification included active-power modes, which caused industry stakeholders to become concerned about capping the energy needed to support certain emerging technologies. Another significant challenge, which is similar to that facing the data-center market, is the lack of existing industry standards for measuring energy efficiency. Developing test procedures for server energy performance should be similar in complexity to the process for the ENERGY STAR computer specification.

Despite the challenges, EPA has been working with industry and other stakeholders to develop measurement protocols for servers. An industry consensus process led to an initial energy measurement protocol for server energy efficiency (Koomey et al. 2006). This protocol describes 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 power consumed by a server at different computing loads. A more detailed process led by the Standard Performance Evaluation Corporation (SPEC 2007) should yield results by the end of 2007. SPEC will use its current application benchmarks as the basis for generating loads that are “typical of day-to-day server use.” But this will not cover the multitude of other applications. Additional benchmarks will be needed to cover the full range of data center applications.

In December 2006, EPA announced that it will explore development of an ENERGY STAR specification for servers. This process will build on the measurement protocols developed by SPEC and others.

Besides servers, it may be possible to label the performance of UPS systems and other infrastructure products. Performance assessment of UPS systems is relatively straightforward and could be useful in optimizing selection of efficient systems. Labeling in conjunction with industry application criteria could lead to significant efficiency gains.

Besides servers and data center equipment, it may be possible to develop whole-facility labels for data centers. The EPA is convening a workgroup to identify the most appropriate metric to benchmark entire data centers.³⁰ Also, industry and government (DOE, EPA and the National Institute for Standards and Technology – NIST) are currently exploring a way in which a manufacturing plant could obtain an ANSI (American National Standards Institute) certification for energy management and energy intensity improvement. This certification process under development may also prove useful in recognizing data centers that have adopted good practice in energy management and are on a path of continual improvement in energy efficiency.

The development of meaningful energy-performance metrics for data centers and their equipment is crucial to advancing energy-efficiency efforts in this sector. Standard measures of performance increase the ability of data operators to compare products based on energy performance, making operators more likely to consider energy performance in their purchasing decisions. An energy performance measurement standard, based on established and familiar server performance measures, also helps increase users’ confidence that overall performance will not suffer with energy-efficient servers. Lastly, widely accepted energy performance metrics and criteria for servers – and for entire data centers – lay the groundwork for other policies that promote the development and purchase of energy-efficient servers, including, for example, government procurement policies and financial incentives, such as tax rebates and credits.

7.2.2.2. Recent industry focus on metrics

The data center industry has recently focused great attention on the issue of metrics, with mixed success. The server performance metrics mentioned above have been one area of relative success.

³⁰ Updates on EPA’s research can be found at <http://www.energystar.gov/datacenters>.

The issue of benchmarks for data center infrastructure has been the focus of some industry discussions and agreement on concepts, but there is continued disagreement about terminology. It is important that these disagreements be resolved so the industry can move forward with actually measuring parameters of interest. What is most critical is that terms like “efficiency” be used in ways that are technically accurate, intuitive (e.g., for efficiency, an increasing number between 0 and 1 indicates higher efficiency and a decreasing number indicates lower efficiency), and most evocative of the term’s true meaning.

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 data center.

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.

There seems to be general agreement that the first measure above is needed, and that the SPEC process will eventually yield something that can be used to measure that term. There is less agreement, however, about the language used to define the second term in the equation even though there is general agreement on the concepts. 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.

There are currently several terms being proposed by industry to define the energy efficiency of the infrastructure portion of the data center. Brill (2007) calls the ratio of total data center load to IT load the “Site Infrastructure Energy Efficiency Ratio” or SI-EER. The Green Grid (2007) calls that same ratio the “Power Usage Effectiveness” (PUE) and its inverse “Data Center Efficiency” (DCE). Rasmussen (2006) also uses the term DCE. Before significant progress can be made in benchmarking data centers, basic terminology that meets the above criteria needs to be agreed upon by the various stakeholders, and additional details related to measurement protocols for the site infrastructure efficiency need to be worked out.

7.2.3. Electric Utility Demand-Side Management Programs

As part of billing and supply functions, distribution utilities have unequalled access to virtually every household and business in their service territories. This allows utilities to be effective conduits for communicating information and providing financial incentives to encourage investments in energy-efficiency measures. In exchange for monopoly franchise over a certain service territory, utilities have an “obligation to serve” the customers in that area and to ensure that electric and gas supplies are provided in a cost-effective manner. It is usually more cost-effective for utilities to “free up” electricity through demand-side management (DSM) activities rather than to increase generating capacity (Loper et al. 2005).

Electric utilities spent \$2.6 billion in 2006 for energy-efficiency and demand-response programs (CEE 2007b). Although DSM programs may be more cost effective than supply-side investments, utilities are not always eager to implement DSM programs, which reduce utility revenues and profits under most states' ratemaking formulas. Consequently, regulators have generally had to direct utilities to institute DSM programs.

Six states – California, Hawaii, Illinois, Nevada, Pennsylvania, and Texas – have either instituted energy-efficiency portfolio standards, requiring utilities to reduce their customers' demand by a certain percentage a year, or allow energy-efficiency measures to count toward renewable portfolio standards, which mandate that a certain amount of electricity generated by utilities in the state must come from renewable energy sources.³¹ For utilities that must reduce their customers' demand, a project to assist data centers in reducing energy load could help these utilities meet their goals.

As of 2001, approximately 20 states and the District of Columbia had public benefits funds (PBFs) dedicated to energy-efficiency improvements (DSIRE 2007, Kushler et al. 2004, US EPA 2005). Financed by a surcharge on customers' electric bills, PBFs frequently support commercial energy-efficiency or research and development projects that could target data centers.

Electric utility incentives help buy down the additional cost of more efficient data center equipment and are perhaps even more effective than government incentives because they are available immediately rather than at tax time. Besides the reduced cost of the upgrades, utility incentives help compensate for the increased effort needed to learn about and locate energy-efficient equipment, draw attention to technologies, and legitimize the technologies in the eyes of consumers, who see that the government and/or electric utilities are essentially endorsing these technologies.

As noted earlier, few utilities are offering incentives specifically for data center equipment although utilities and state efficiency funds that are not actively marketing energy-efficiency data center improvements may be amenable to including data-center efficiency improvements in their commercial special projects funds. Another possible future source of rebates is the 80 Plus® program, in which participating utilities agree to provide a \$10 rebate to manufacturers of personal computers and pedestal servers that install power supplies certified to be at least 80 percent efficient.³² In less than two years, the 80 Plus® program has generated more than \$5 million in incentives although that number includes utility and administrative fees. A utility-funded initiative to create equivalent 80 Plus® standards for data center computing equipment power supplies is currently in development. Preliminary results are expected in the second quarter of 2007.

³¹ See Appendix 5 for details of and links to all of the relevant portfolio standards.

³² Utility-related participants listed on 80 Plus®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 Company, Sacramento Municipal Utility District, Southern California Edison, Western Massachusetts Electric, and XCel Energy. <http://www.80plus.org/util/util.htm>.

In August 2006, PG&E became the first utility to offer rebates to its non-residential customers that implement virtualization/server consolidation projects (Sun Microsystems 2006). The value of the rebate is determined by the energy savings that are directly attributable to the removal of computing equipment (i.e., not to indirect savings from a reduced cooling load). The utility customer receives eight cents in up-front incentive payment per kWh of annual savings, for a total of up to half the original cost of the virtualization software (PG&E 2006).

In addition, Avista Utilities, serving the Pacific Northwest, began offering a prescriptive rebate program to commercial and government customers operating data centers that deploy energy-efficient liquid-cooled racks in November 2006. The rebate is \$100 per server, or up to \$3,600 per rack, and is available for both new construction and retrofits (SprayCool 2006).

Other electric utilities are considering similar rebates and incentives, including Austin Energy, which is planning to introduce a data center program in the second half of 2007. According to Austin Energy staff, the package will include rebates on a variety of efficiency measures, including server virtualization and efficient cooling practices. The utilities that are developing efficiency programs for data centers have formed a committee to exchange ideas and coordinate their programs, and the Consortium for Energy Efficiency is beginning an initiative to address energy use in data centers.

NYSERDA, with funding from electric utilities, supports the development and demonstration of emerging technologies, including power-reduction technologies for high-density electric load applications, such as data centers (NYSERDA 2007). NYSERDA is reportedly examining potential incentives for data center procurement and energy management but does not expect to finalize any incentives until the second half of 2007 (Bull 2007).

7.2.4. Technology Procurement

The incentive and labeling policies described above are generally used to promote and increase the market share of commercially available products. In some cases, however, there is little variation in energy performance among available products, or a dramatically more efficient technology has been developed but is not being incorporated into commercially available products for one reason or another. In this situation, traditional labeling and incentive programs are not effective because they can really only promote products and technologies that are commercially available.

To solve this problem, a strategy called technology procurement (a.k.a, “the golden carrot”) has been developed. Technology procurement is a method for pulling new, highly efficient and affordable products into the marketplace through competitive procurement backed by large-volume buyers. This process benefits buyers because they get products that better meet their needs, and it can benefit product vendors because it helps them reduce the risks involved in technology development and helps them better understand customer needs. The technology procurement process involves writing technical specifications for the desired attributes of new products, issuing competitive solicitations, selecting winning products, and promoting the winning products. Often, financial incentives such as utility rebates are used to promote the selected products. This model was successfully applied to refrigerators in the early 1990s, and

more recently has been used to bring new lighting and air conditioning technologies into the market (US DOE 2002b).

7.2.5. Energy Performance Standards and Building Energy Codes

Energy performance standards are a regulatory approach to requiring energy efficient products, generally by banning the sale of products and equipment that do not meet certain minimum efficiency thresholds. Likewise, building energy codes require that new or renovated buildings meet minimum efficiency standards in order to receive building or occupancy permits.

7.2.5.1. Energy Performance Standards

The federal energy performance standards program began in 1987 following the proliferation of sometimes differing state standards, including those in California, Connecticut, Florida, Massachusetts, and New York, which prompted appliance manufacturers to seek uniform national standards through federal law. In response to calls for federal leadership and the establishment of national standards, Congress passed the Energy Policy and Conservation Act of 1975 (EPCA), which established a federal energy conservation program for major household appliances and appliance efficiency targets (US DOE 2006a).

The National Appliance Energy Conservation Act (NAECA) and the Energy Policy Act of 1992 provided the framework for establishing minimum energy-efficiency standards for more than two dozen types of appliances and equipment.³³ NAECA was enacted with support from product manufacturers concerned about the costs of trying to comply with varying state standards. NAECA prohibits state standards for products that are subject to federal standards.³⁴

NAECA and, subsequently, the Energy Policy Act of 1992 (EPAct), prescribed the level of efficiency to be required by the standards for many products. For others, DOE was instructed to develop standards through a rulemaking process, allowing opportunities for advice and comment from various stakeholders. In its rulemakings, DOE is required to provide extensive analysis of the impacts of proposed standards on consumers, the economy, product manufacturers, and the environment. DOE serves essentially as a single gatekeeper for federal standards. It is tasked with creating a process that is transparent and credible and that provides an opportunity for all stakeholders to be heard. In fact, for all intents and purposes, DOE is responsible for negotiating standards with those stakeholders (Loper et al. 2005).

To date, DOE has established federal minimum energy-performance standards, or has begun conducting rulemakings, for at least 27 primarily electric residential and commercial product categories (US DOE 2007b). None of the standards is explicitly focused on equipment found in data centers, although a few standards, like those for commercial lighting, are relevant. Standards for servers, power supplies, or data centers would not be possible until widely accepted energy-performance metrics, as described previously, are developed. The standard-

³³ More than a decade earlier, EPCA established a federal energy conservation program for major household appliances by calling for appliance efficiency targets, but little progress was made until the 1980s (US DOE 2006a).

³⁴ States may receive a waiver if they persuade the Secretary of Energy that the state standard is “needed to meet unusual or compelling state or local energy or water interests” 42 U.S.C. 6291.

setting process is lengthy; it usually takes several years from the time that DOE announces its intent to develop a standard until the standard becomes effective.

Besides the federal minimum efficiency standards, California and several other states also have minimum efficiency standards for a range of residential and commercial products. California, in particular, has been very active during the past few years in setting standards for products that have traditionally not been regulated for their energy use, such as ceiling fans and consumer electronics. The California standards do not currently cover servers or other equipment used in data centers although California does regulate the efficiency of external power supplies and has considered regulating computer power supplies.

Power supplies and UPS systems are probably the types of data center equipment most amenable to standards. Servers and other types of IT equipment, by contrast, generally evolve so quickly that the lengthy federal standard-setting process would have difficulty keeping abreast of the type of products on the market and their energy performance. Equipment performance standards are also not the most appropriate way to address products whose energy performance is strongly determined by their in-field installation and configuration. These situations are best addressed by building energy codes and other types of whole-building policies.

7.2.5.2. Building Energy Codes

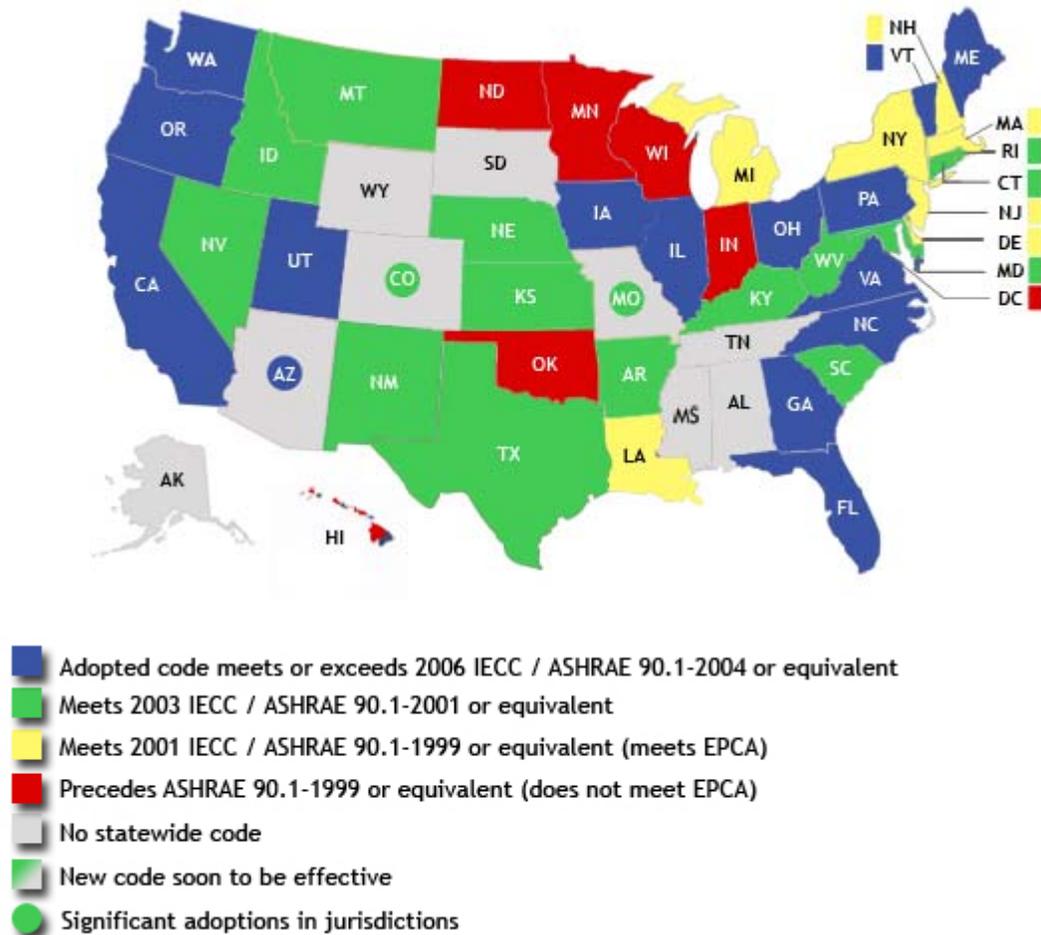
Building energy codes were first adopted by California in 1974. To help states trying to adopt their own codes, national model energy codes are developed and updated every few years. Under federal law, states are free to adopt the updated model energy code, adapt the model code to the state's particular needs, adopt an older version of the model code, adopt a code that is completely different from the national model code, or institute no energy code at all.

Currently, at least 41 states have adopted some form of energy building code. In several states that have not adopted codes, major cities have adopted them. For example, Phoenix, Chicago, and Denver each have adopted a version of the International Energy Conservation Code (IECC) even though their states have either no or a less-stringent building code (Loper et al. 2005). Figure 7-1 shows the current status of building codes for commercial buildings in each state.

Model energy codes are used as the basis for state codes in roughly four out of five states. From 1991-2005, national model energy codes delivered cumulative energy cost savings exceeding \$7 billion and saved enough energy to meet the needs of 4.7 million homes for one year (Mueller 2007).³⁵ Nadel and Prindle (2004) estimate that the U.S. saved 0.5 quad in the year 2000 alone thanks to commercial and residential building energy codes.

³⁵ These numbers may underestimate the total savings from buildings codes because they do not include savings from states that have adopted energy-efficiency codes that were significantly different from the model energy codes, nor states that have adopted the IECC 2003, which has not been endorsed by DOE. For example, California does not rely on IECC or ASHRAE for its format, scope, or stringency requirements for commercial or residential codes. Florida has its own residential code, but its commercial code relies heavily on the national model. Seattle takes the ASHRAE code and commits to improving it by 20 percent.

Figure 7-1. Status of Commercial-building Energy Codes, by State



Source: (BCAP 2007).

Building codes are administered by 50 states and thousands of local authorities. Commercial building code recommendations are developed and offered to state and local governments both by the IECC and ASHRAE. The ASHRAE 90.1 standard – once adopted as a mandatory state or municipal building energy code – is applicable to all new commercial buildings and all major remodeling or renovation of existing commercial buildings (Loper et al. 2005). The current ASHRAE standards (90.1-2004), however, exclude energy that is primarily used for commercial, industrial, or manufacturing processes. Thus ASHRAE 90.1 excludes data center heating and cooling (which mainly cools equipment, not people) and IT equipment (ASHRAE 2004a, Section 2.3). However, ASHRAE is developing guidance to address aspects of data center design, as detailed later in this chapter, and plans to evaluate whether data centers could be included within standard 90.1.

7.2.6. Technical Assistance and Awareness Building

The growth of data centers is a relatively recent phenomenon, so best practices for design and operation have only recently been developed.

LBNL has had an active research program for several years on energy use in data centers and provides technical resources to help designers and operators improve data center energy performance (<http://hightech.lbl.gov/datacenters>). LBNL's information resources focus on energy efficiency in power supplies, power distribution, and cooling. The DOE Save Energy Now program is building upon the LBNL information resources to create an energy assessment protocol, tools and training (see www.eere.energy.gov/datacenters/).

Besides the ASHRAE building codes described previously, ASHRAE Technical Committee 9.9 (covering 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" (Schmidt and Beaty 2005).

ASHRAE Technical Committee 9.9 has developed guidance to address at least three aspects of data center design and operations that have been discussed in this report:

- 1) Thermal Guidelines for Data Processing Environments – temperature and humidity guidance for data centers (as well as for other types of computer environments) (ASHRAE 2004b).
- 2) Airflow protocols for servers – provide a common airflow scheme so exhaust from one manufacturer's equipment is not drawn into another manufacturer's air inlet; i.e., the front face of all equipment is facing the cold aisle (ASHRAE 2004b).
- 3) Power Trend Charts – Address the mismatch of HVAC and IT equipment resulting from the shorter lifetime of IT equipment, and help data center operators/designers predict future IT requirements (ASHRAE 2005).

ASHRAE TC9.9 is publishing two upcoming books on data center management highlighting energy efficiency and best practices: "*Improving Energy Efficiency of Datacom Equipment Centers*" and "*High Density Data Centers – Case Studies and Best Practices.*"

In addition, ASHRAE is preparing guidance to address energy efficiency and total cost of ownership.

Other organizations that provide resources for data center operators and designers that may be useful in improving data center efficiency are listed below.³⁶

- *7x24 Exchange*: 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>).
- *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>).

³⁶ A number of additional organizations exist in universities or as part of labs associated with companies in the IT industry – e.g., IBM's research lab in Austin, Texas.

- *Consortium for Energy Efficiency (CEE)*: A nonprofit organization promoting energy-efficient products and services, CEE's board recently endorsed an exploration of efficient data center operations. Their first conference call with relevant stakeholders and advocates occurred in late March 2007 (http://www.cee1.org/resrc/news/07-02nl/04_data.html).
- *Consortium for Energy Efficient Thermal Management (CEETHERM)*: 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," CEETHERM operates a testbed data center to test new power and cooling technologies and practices (<http://www.me.gatech.edu/CEETHERM/>).
- *Critical Facilities Round Table*: A non-profit organization dedicated to the success of owners and operators of data centers and other critical facilities, the round table focuses on the sharing of information and the publication of "cutting edge" technical solutions for the data center community, with the intention of proactively influencing the management, institutions, and regulators of the critical facilities industry (www.cfroundtable.org).
- *Data Center Dynamics*: Hosts a conference and expo series for professionals involved in the design, construction and operation of 24/7 mission critical IT facilities in the worlds top business cities. Also publishes ZeroDownTime magazine. (www.datacenterdynamics.com).
- *Efficient Power Supplies*: This is an initiative created by EPRI Solutions Inc. and Ecos Consulting to encourage a global discussion of energy-efficient power supplies (www.efficientpowersupplies.org).
- *The Green Grid*: An association of information technology professionals seeking to increase the energy efficiency of data centers globally through an industry-based response to efficiency issues, the Green Grid allows members to share best practices in data center power management. Companies involved in the Green Grid include end users, all the major server manufacturers, processor manufacturers, suppliers of power and cooling equipment, and major providers of server software (www.thegreengrid.org).
- *ITherm*: This organization describes itself as "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).
- *Pacific Northwest National Laboratories (PNNL)*: Like CEETHERM, PNNL is conducting research into energy-efficient liquid-cooling technologies for use in data centers, called the "Energy Smart Data Center" project. This DOE-funded research is publicly available. (<http://hpc.pnl.gov/projects/spraycool>)
- *Uptime Institute*: The institute has 85 large (avg. 60,000 ft²) data center members in the U.S. and 10 in Europe. It sponsors benchmarking; abnormal incident and failure collection; the development of industry best practices; and presents symposiums, design charrettes, seminars, certifications, and training for its member companies. The Uptime Institute has recently focused on the creation of Integrated Critical Environment (ICE) teams, which are cross-functional data center management teams that integrate expertise from both the facility and IT parts of the organization. The institute's website has a free library of white papers addressing energy efficiency, cooling, reliability, and best practices to improve data center uptime efficiency (<http://www.uptimeinstitute.org>).

7.2.7. Federal Energy Management Policies and Programs

Over the past two decades, the federal government has instituted several goals, requirements, and initiatives to improve the energy efficiency of its buildings. Federal agencies are required by

presidential executive order and law to reduce energy per square foot in their facilities and to purchase only ENERGY STAR equipment where available (OFEE 2007, EAct 2005). Appropriated funds have generally not been sufficient to pay for these mandates, so alternative financing mechanisms have been authorized by Congress, which allow agencies to enter into performance based contracts for energy services with Energy Services Companies (ESCOs) and local utilities. Other initiatives to assist agencies include audits, demonstration projects, and technical assistance.

7.2.7.1. Energy Reduction Requirements

Federal agencies are required by presidential executive order to reduce building energy use per gross square foot (GSF) by three percent annually, including energy-intensive operations such as data centers (OFEE 2007).³⁷ The requirements were only recently increased from the two percent annual reductions contained in EAct 2005 (enacted August 2005), which was an increase from annual targets of roughly 1.5 percent in prior years.

Most agencies have complied with the energy reduction targets and are expected to do so in the future. As noted elsewhere, it appears that with the new three percent annual reduction targets, agencies are feeling pressure to identify new efficiency opportunities, which may put a spotlight on data centers.

7.2.7.2. Procurement Requirements

Federal, state, and local governments spend tens of billions of dollars annually on energy-consuming products, 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 (FEMP 2006b). Typically, as a starting point, eligible products are in the top 25 percent of efficiencies for their product class.

Some of the existing ENERGY STAR and Recommended Product categories could be applicable to data centers, for example, air conditioning systems, although most types of equipment found in data centers are not covered by the procurement recommendations.

In addition, the Electronic Product Environmental Assessment Tool (EPEAT) provides a registration label for products that meet certain environmental and energy attributes. Initially,

³⁷ In past years, some agencies have excluded their data center operations from building energy reduction targets, citing them as energy intensive operations (FEMP 2006a). This loophole has been closed in EAct 2005, which effectively eliminated the distinction between buildings and energy-intensive operations for the purpose of determining the applicability of mandated reductions.

EPEAT addresses desk-top computers; however, with time, the program is expected to cover servers and other data center-type products.

The federal government is also a major customer in the market for high-performance computing (i.e., supercomputers). These are large computers with correspondingly large energy consumption, used for numerically intensive applications such as high-energy physics. When purchasing new supercomputers, the government typically focuses on computing performance and does not consider energy efficiency in its purchasing criteria.

7.2.7.3. Building Performance Standards

EPAct 2005 required federal commercial buildings to meet the ASHRAE 90.1-2004 energy code (updated from a previous requirement to meet ASHRAE 90.1-1989). This same provision of EPAct 2005 also requires new federal commercial buildings to be designed for 30 percent energy savings compared to the ASHRAE 90.1-2004 reference building, if cost efficient (EPAct 2005, Section 109).

As discussed above, however, ASHRAE does not include provisions for data center heating and cooling use or for the IT equipment itself. If ASHRAE were to incorporate data center requirements into its commercial energy code, as it is considering, when federal building standards are next updated they would also likely incorporate data center requirements, which could have a market transformation effect that would lead to greater efficiency. ASHRAE currently has four books covering various aspects of data center design, and a fifth book in preparation that addresses energy efficiency and total cost of ownership.

7.2.7.4. Alternative Financing

Appropriations for energy-efficiency improvements have never been sufficient to exploit more than a small fraction of the energy-saving opportunities in federal facilities. Alternative financing sources – namely ESCOs and utilities – have been relied on heavily, providing funding for nearly half of federal efficiency investments made to date. The alternative financing is provided as part of an Energy Savings Performance Contract (ESPC), in which ESCOs invest in energy-saving projects in federal facilities, and agencies, in return, pay for the efficiency upgrade through the guaranteed savings in their energy bill. Once the investment is fully paid, the facility or agency enjoys the full benefit of the lower utility costs. Utilities can enter into similar agreements with government entities, called a Utilities Energy Service Contract (UESC).

Several ESCOs already have ESPCs in place with government agencies, but it appears the ESPCs generally do not include data center efficiency improvements even though there are no legal barriers to inclusion. This exclusion of an area with much potential for energy savings is a result of several factors.

Data center operations are vital to the success of a given facility; if they go down, most of the facility's functions simply cannot be performed. Consequently, many agencies treat data center operations as sensitive and are not willing to grant ESCOs full access (Wagner 2007). This becomes even more important the more sensitive the operation of a facility; the Department of Defense is more likely to be worried about security than other agencies, for instance. This is

significant because, in fiscal year 2005, 15 of the 20 government ESPCs were awarded by the Department of Defense (FEMP 2006a).

Even if there are facilities that are willing to have an ESCO work on their data centers, many facility managers may not realize the enormous potential energy savings in their centers and will therefore not ask ESCOs to make improvements. ESCOs will usually only suggest energy-saving measures that are typically included in ESPCs, focusing on the standard efficiency improvements along with those areas that the facility operators may specifically request. Some ESCOs would likely need to train their employees and familiarize them with data-center-specific efficiency improvements so that they would know the types of measures that would be appropriate to propose and implement.

Furthermore, ESPCs are attractive in large part because they remove financial barriers to energy-efficiency improvements. But data center operators usually have enough money to carry out any improvements they see fit because continuous operation of data centers is so vital. This may also lead data center operators to be relatively unconcerned about energy consumption because performance is so much more important than efficiency, and they know that their bills will be paid more or less without question (Kueck 2007).

On July 12, 2007, James L. Connaughton, Chairman of the Council on Environmental Quality (CEQ) signed a memo to Heads of Agencies prompting a more aggressive use of ESPCs in the Federal community. Because data centers are part of the Federal facility portfolio already covered under Executive Order 13423 and the ESPC memo, an increase in ESPC use for data centers is likely.

7.2.7.5. Building Metering

EPAct 2005 (section 103), with subsequent guidance from FEMP, requires agencies to install electric meters where cost effective on “all buildings” by 2012. In response, agencies were required to submit metering plans to FEMP in August of 2006. Progress in meeting the plans is being overseen by the Office of Management and Budget and FEMP (FEMP 2006c). Although there are no specific requirements to submeter within a building – e.g., to submeter data center operations – some agencies may recognize the benefit of obtaining data on energy use in their data centers and include data centers in their metering plans. Based on discussion with federal energy managers, it appears that separately metered data centers are rare in the federal sector.

Advanced metering would provide data center managers with the information they need to save energy and money as part of an effective operations and management (O&M) practice. Beyond simply measuring electricity consumption, metering also facilitates bill allocation and energy management and helps data center managers identify energy savings opportunities, verify savings, and participate in utility demand reduction programs.

Many UPSs installed in data centers contain built-in meters already. If data centers began using these meters to monitor the energy consumption of their IT equipment, this could be a quick, inexpensive solution to obtaining better energy usage data, at least for equipment protected by the UPSs. It would not monitor HVAC consumption, however, or losses in other parts of the power infrastructure.

7.2.7.6. Demonstration Projects and Technical Assistance

DOE provides numerous forms of technical assistance to help raise awareness about efficiency opportunities, increase familiarity with various types of measures, and demonstrate peer acceptance of energy-saving technologies. DOE-EERE (Office of Energy Efficiency and Renewable Energy) is currently working with LBNL in conducting pilot assessments in data centers. This work includes development of an assessment protocol that can be used by qualified "assessors" in a national data center assessment campaign. This campaign will be included under the ongoing "Save Energy Now" energy assessments for industry, which comprehensively examine the energy systems in buildings and facilities. The DOE program plans to develop and disseminate best practice information, conduct demonstrations of energy efficiency solutions, and provide for research of new solutions to improve data center efficiency.

FEMP's monthly magazine, *FEMP Focus*, provides information on various demonstration projects, mostly for renewable energy, and documents and publicizes energy-efficiency projects undertaken by agencies. FEMP also provides technical assistance to facilities considering or negotiating ESPCs and agencies trying to comply with various energy-efficiency requirements such as efficient purchasing and building metering requirements.

8. Recommendations for Incentives and Voluntary Programs

This chapter addresses the following Public Law 109-431 requirement:

- Recommendations regarding potential incentives and voluntary programs that could be used to advance the adoption of energy efficient data centers and computing.

This report's previous chapters have documented that data centers consume a significant amount of electricity, resulting in large energy costs, and that the rapid growth of data centers can contribute to power grid reliability problems. This report has also identified ways to significantly reduce this consumption. Incentives, voluntary programs, and research can play an important role in overcoming market barriers to energy efficiency and thus to the achievement of this savings potential. These efforts may be of interest to utilities, state governments, and others interested in the policies that can reduce electric utility costs, improve power grid reliability, and avoid pollutant emissions.

Given the growing contribution of data centers to national energy consumption, it is perhaps surprising that so few policies and programs have emerged. This may be a result of several factors:

- 1) Energy costs were traditionally a relatively small part of overall data center costs.
- 2) There is a lack of energy performance metrics to distinguish energy-efficient products and facilities.
- 3) Rapidly changing technology makes it hard for government programs and policies to stay relevant.
- 4) Existing high-level industry interest, as demonstrated for example by the recent spate of launches of new, more efficient server products, suggests there may only be a small role for government in this market transformation.
- 5) There is no easy solution for how government policies or programs can overcome some of the key barriers, such as split incentives.
- 6) These issues are not unique to data centers so may be adequately addressed through generic energy-efficiency policies and programs.
- 7) There is a lack of knowledge or there is inexperience with data center energy-efficiency opportunities because of rapid growth of the market and reliance on outdated methods and technologies.
- 8) There is a lack of energy-efficiency research focused on data-center-specific needs.

This chapter outlines a set of incentives and voluntary programs intended to address the most significant market barriers that hinder energy efficiency in data centers. It recommends actions that the federal government can take as well as those that can be taken by other organizations: state and local governments, public utilities, and the data center industry (including IT equipment and software manufacturers). It concludes with recommendations for research and development (R&D) activities that can further enhance efficiency opportunities in this sector.

These recommendations are drawn from several sources. The first is a workshop on data-center energy use convened by EPA on February 16, 2007, which engaged stakeholders including

equipment vendors, component manufacturers, industry trade groups, and data center operators (see Appendix 2 for summary notes from the workshop). Input was also received from many stakeholders following the workshop. The second source is interviews with individual data center operators conducted by LBNL and ASE during March 2007.³⁸ Another source is a recent report by the Alliance to Save Energy on policy options to promote energy efficiency in data centers (Loper and Parr 2007). Finally, the R&D recommendations are based on a roadmap prepared by LBNL (Tschudi et al. 2003) and subsequent work in this market including assisting in developing R&D plans for DOE.

Several principles guide these recommendations:

- Efficiency measures need to address all parts of the data center. Both a holistic approach to the data center as a system, as well as attention to specific data center components, will be needed to achieve dramatic improvements in energy efficiency.
- Top-down changes in organizations that own and operate data centers will be required to achieve efficiency in these facilities. Executive-level managers (CEOs, CIOs, CFOs), need to be part of the solution.
- Companies are not now measuring or minimizing total cost of ownership (TCO), but a focus on TCO is the single most important step an organization can take to address the cost problems in the data center.
- Objective, credible information is needed by data center operators to accurately evaluate the performance of their facilities and the products they purchase.
- Prescriptive approaches that require specific technologies will probably not be effective because of the rapid pace of change and innovation in this sector.
- Financial incentives can be effective in some situations. Although the government and utility funds that could be used for such incentives are relatively small in comparison to the capital routinely invested in data centers, they can leverage large amounts of capital. Financial incentives can also play an important role in raising awareness of efficiency options.
- The federal government has an opportunity to lead by example, so all of the recommendations below include a role for the data centers used by the federal government.
- Action to address data center energy use should be part of a comprehensive energy efficiency effort for commercial buildings and industry.
- The purpose of these policy recommendations is to articulate desired outcomes and a process for achieving those outcomes, not to advocate particular technical solutions for improving energy efficiency. The R&D recommendations identify emerging practices and technologies that warrant further investigation.

8.1. Policy Recommendations³⁹

Goal: *Executive-level managers use information about the energy efficiency of their data centers as a standard management metric.*

³⁸ Layne Maly of AFCOM assisted with recruiting survey respondents.

³⁹ Note: each of the policy goals is independent of the other goals; the list is not cumulative.

Executives have traditionally not been aware of the energy use of their data centers. This has allowed “siloes” organizations to develop, in which responsibility for meeting data center energy needs is the responsibility of facilities managers while the driver of this energy use – IT equipment – is the responsibility of IT managers who typically report to a different executive-level manager. Improved information about data center energy performance allows organizations to manage these assets in a more integrated way and break down split-incentives and other barriers within the organization.

Strategy 1: Standardized performance measurement for data centers — Develop an objective, credible energy-performance rating system for data centers, to allow tracking of facility performance over time and comparison among facilities. This rating system may initially address only the infrastructure portion of the data center although it should be extended (or a companion metric developed), when possible, to include the productivity and work output of IT equipment. These metrics should account for factors such as computing output and availability requirements that vary among data centers and also allow for dynamic reporting of energy performance in near-real time. Comprehensive data about the energy use and characteristics of a wide range of data centers is essential to developing a rating system, and compiling these data should be the highest priority activity in order to meet this goal.

Policies:

Federal government: Partner with industry to develop an objective, credible energy-performance rating system for data centers, which accounts for the large diversity in data center characteristics, such as reliability and types of computing performed. Compile a database of whole-building energy use data for use in developing this rating system. Develop a voluntary performance rating system for data centers (e.g., ENERGY STAR), based on these industry-accepted metrics. Also, work with industry to develop a method to certify a data center facility’s energy management practices and energy intensity improvement. This voluntary certification process, currently under development, may prove useful in recognizing a data center’s adoption of best energy management practices and their continual improvement in energy efficiency.

Data center industry: Consider partnering with the federal government to develop an objective, credible energy-performance rating system for data centers, including assisting with compiling whole-building energy use data for use in developing a metric.

Electric utilities: Consider using the rating system developed as a basis for data center energy-efficiency incentive programs in their service territories. Coordinate efficiency programs with those of other utilities.

Strategy 2: Private-sector challenge — Challenge executive-level managers to commit to conduct energy-efficiency assessments, implement improvements, and report energy performance of their existing data centers. These assessments require protocols and tools that should be jointly developed by government and industry.

Policies:

Federal government: Issue a “CEO Challenge,” which would challenge private-sector CEOs to conducting energy-efficiency assessments, implement any improvements that reduce TCO, and report energy performance of their data centers annually. Provide public recognition for best performers, both in absolute terms and in relative improvement. DOE’s Save Energy Now program could work with industry to develop

assessment protocols and best practices information (including no-cost and low-cost measures) to support private-sector data centers in meeting their commitments.

Organizations that operate data centers: Consider publicly committing to conduct energy-efficiency assessments, implement any improvements that reduce TCO, and report energy performance of their data centers.

Strategy 3: Federal leadership — The federal government can act as a model in encouraging improved efficiency of data centers, by reporting energy performance, conducting assessments, and implementing efficiency improvements in its own data centers.

Policies:

Federal government: Commit to publicly reporting the energy performance of its data centers, once standardized metrics are available. The federal government should commit to conducting energy-efficiency assessments in all its data centers within two to three years and implement all cost-effective operational improvements. To the extent that reporting on data center performance is a logical outcome of existing requirements under the EPA Act 2005, Executive Order 13423 (High Performance Buildings Goal and Energy Efficiency Goal), as well as OMB Scorecard on Energy Management, government agencies should ensure that data centers are included within existing reporting efforts. Additionally, the Architect of the Capitol should implement the server-related recommendations from the Greening of the Capitol report (Beard 2007). For outsourced data centers (not owned by the government) and services, develop a procurement specification that requires ENERGY STAR-rated data centers and adoption of best practices, or compliance with government/industry voluntary standards for energy management. For high-performance computing facilities, develop procurement specifications that require an evaluation of energy use as well as computational ability – e.g. Flops/Watt.

Goal: *Improve the management of data centers so that capital investments and operational decisions incorporate energy-performance considerations.*

In many organizations, it is standard practice to base IT equipment and software purchasing decisions on TCO, which includes the lifetime maintenance and support costs. These TCO calculations for IT equipment rarely include energy as a factor, but they should include both the cost of electricity to run and cool the hardware as well as the cooling, power conditioning, and other *capital* equipment required to support the IT hardware in the data center. Data centers should also be managed in a way that avoids split incentives, draws on expertise from all parts of the organization, and uses standard metrics and tools to allow continual monitoring and improvement in energy management. Data center tenants, who control the IT equipment that drives energy consumption, usually do not pay the power costs caused by that equipment, and thus have little incentive to reduce its energy use. If a data center is part of a larger building or campus, it often is not separately metered, which makes it difficult to accurately charge tenants for the energy they use (and difficult to benchmark its energy performance). Metering energy use (including chilled water, with separate metering of IT loads) and charging end users for their energy consumption eliminate the split incentive between those who purchase the equipment and those who pay the electricity costs. This also helps align incentives for data center capital cost because the vast majority of the costs in building a data center are directly related to the amount

of power use.

Strategy 1: TCO methods — Develop TCO methods, data, and tools that account for energy consumption and capital investments required for IT hardware, software, and data center infrastructure systems to provide required level of computing output and operational risk mitigation.

Policies:

Federal government: Support the development and sharing of “energy aware” TCO methods, data and tools, through Save Energy Now and FEMP and federally funded IT research through the National Science Foundation, DARPA (Defense Advanced Research Projects Agency), etc.

Strategy 2: Energy management methods — Develop tools and standards for continual improvement in data center energy performance.

Policies:

Federal government: Work with industry to develop tools to assist in energy management and continual improvement in energy performance of data centers, leading eventually to an industry standard (e.g., ISO standard) for management of data center energy. Standards setting could be coordinated with activities to create ANSI certification for energy management and energy intensity improvement in the manufacturing sector.

Strategy 3: Cross-functional teams — Develop and disseminate best practices for energy management policies that integrate across functional groups of a datacenter.

Policies:

Federal government: Through the federal Chief Information Officer working group, encourage the use of cross-functional management teams and select one or two agencies to use as case studies for testing and publicizing this approach. Develop technical guidance that stresses the importance of cross-functional teams and provides suggestions on how to implement them. Add data-center-specific guidance to the ENERGY STAR Guidelines for Energy Management.

Electric utilities and data center industry: Consider using and promoting the ENERGY STAR Guidelines for Energy Management, which include components on strategic energy management and cross-functional teams.

Strategy 4: Data center metering — Meter data center electricity use. Metered energy use data are essential for user billing and for benchmarking data center energy performance. Charge data center tenants for energy use.

Policies:

Federal government: Ensure federal government-owned data centers with significant energy use are included in agency electricity metering plans, with separate metering of data centers at the facility level and key systems level (e.g., chilled water, IT equipment). In government-owned data centers, use a charge-back system based on the energy consumption of IT equipment to charge data center tenants. Develop tools and standards for chargeback systems, and document best practices. For outsourced data centers, require that hosting contracts charge for peak power demand and energy use instead of per square foot as is almost universally done now.

State public utility commissions and electric utilities: Consider requiring separate utility revenue meters on large data centers when electric service is established even if a data center is part of a larger building.

State and local building codes: Consider requiring separate utility meters on large data centers.

Goal: *Make it easier for purchasers to easily identify and specify data center equipment with the best energy performance.*

Currently, it is difficult to identify energy-efficient data center equipment (especially IT equipment) because there is neither a standard test method to determine efficiency, nor a consistent way for vendors to communicate this information to purchasers. Clear efficiency metrics help facility, IT, and senior management personnel communicate about data center energy consumption and make optimal purchasing decisions. These metrics would complement, but are distinct from, the whole-facility energy performance metrics described earlier.

Strategy 1: Standardized performance measurement for data center products —

Develop objective, credible energy performance metrics for data center equipment, including servers, storage, network equipment, UPSs, CRAC units, chilled water plants, and other specialty equipment such as liquid-cooled cabinets. Metrics need to realistically represent energy consumption under typical usage conditions (such as significant time at low load), cover a broad range of product types and applications (using a workable classification system for IT products), and be updated periodically. The metrics should cover collections of servers as well (e.g., rack, cluster, and blade servers). Including realistic peak-power measurements as an output of these metrics may help avoid oversizing of data center infrastructure.

Policies:

Federal government: Work with industry to develop objective, credible energy-performance metrics for this equipment. Harmonize metrics globally where possible.

Data center industry: Consider initiating an effort in partnership with the federal government to develop an objective, credible energy-performance rating system for data center equipment.

Standards bodies: Consider developing consensus energy metrics for these equipment types. Server energy use should be evaluated using industry-accepted metrics such as the SPEC Power and Performance and benchmark metric when it becomes available. Other appropriate standards bodies and coordinating organizations may include IEEE and the Green Grid Association.

Strategy 2: ENERGY STAR specifications for data center products — Label products that have the best energy performance on the market, to make it easy for purchasers to identify and buy these products. In addition to energy performance, labels should also include prescriptive requirements for energy-information reporting by IT equipment, such as standard communication protocols to query over the network the power consumption of an IT product, and standard reporting formats for obtaining and aggregating data relevant to power such as system utilization rates. Consider going beyond ENERGY STAR to support coverage of data center products under EPEAT.

Policies:

Federal government: Using the metrics defined above, investigate whether the development of ENERGY STAR specifications for these product categories would be an effective strategy to complement the whole-facility approaches outlined earlier. For volume servers, consider an initial specification that uses an 80 Plus® type specification for power supplies.

Electric utilities: Consider using the ENERGY STAR product labeling specifications as a basis for data center efficiency programs in their service territories.

Strategy 3: Federal procurement specifications for data center products — Modify procurement policies and specifications to require energy-efficient equipment.

Policies:

Federal government: If and when ENERGY STAR specifications are developed, federal procurement specifications that build on ENERGY STAR should be implemented, starting with volume servers, storage equipment, and network equipment. Develop purchasing tools to help federal purchasers estimate energy and data center infrastructure savings from efficient data center equipment.

Goal: *Data center operators are willing to adopt new energy-efficiency technologies and practices because the reliability/performance/cost risks can be reduced to acceptable levels.*

Performance, availability, and maintainability will always be the primary concerns for data center operators. These attributes are what make IT valuable to organizations. Energy efficiency is currently seen as a change that risks degrading these primary attributes. Much of this perceived risk is a result of lack of information on the part of data center operators, both about the true cost of inefficiency and about the benefits of energy-efficient technologies. The first problem can be addressed through the energy benchmarking described above, and the second can be tackled through better information about the “non-energy” (e.g., availability) impacts of new energy-efficiency technologies and practices. Lastly, information about technology demonstrations and data center best practices needs to be publicly available.

Strategy 1: Demonstrations and case studies — Provide objective, credible information about the effect of energy-efficient technologies and practices on data center availability and performance.

Policies:

Federal government: Provide technical assistance for design and implementation of demonstrations of new or underutilized technologies and publicize results. Develop case studies of efficient data centers. Partner with electric utilities, universities, industry associations, and the data center industry to develop a neutral, “real-world” testing and demonstration center(s) (“National Center(s) for Data-Center Best Practices”) to verify technologies for reducing energy consumption in data centers. Emphasis should be given to technologies that reduce risk to data center operators through modular approaches to scaling, rapid deployment, etc.

Electric utilities, universities, data center industry: Consider partnering with the federal government to develop a neutral, “real-world” testing and demonstration center to verify new energy-efficient products and technologies. The CEETHERM test data center mentioned in Chapter 7 (or other university-operated test centers) could serve as a model

for this type of facility. Publish testimonials and case studies to document the benefits of energy-efficiency technologies in data centers.

Strategy 2: Technology procurement — Provide incentives to bring new technologies and products that are both energy efficient and reduce performance and reliability risks to market.

Policies:

Federal government: Partner with electric utilities and energy-efficiency organizations to organize a technology procurement program to bring products into the market that are energy efficient while meeting performance and reliability requirements that match or exceed those of currently available products.

Electric utilities: Consider partnering with the federal government to develop a technology procurement program for efficient products, and promote the winning products through financial incentives, marketing information, etc.

Strategy 3: Life-cycle risk model — Develop an integrated, full life-cycle risk model to assess the performance and availability risks from new energy-efficiency technologies. Certain technologies, such as dynamic provisioning or DC power, involve a change in the way that data centers are designed and operated, so their operators need to be able to objectively weigh the costs, risks, and benefits of these new technologies. The model needs to address risk classification, quantification, and mitigation.

Policies:

Federal government: Fund development of a life-cycle performance and reliability risk model for data centers. Require life-cycle risk analysis for new data centers and major reconfigurations. Develop the risk models based on real data collected through continuous monitoring programs.

Goal: *Data center operators and managers are aware of the problems caused by inefficient data center design and operation, and know about strategies to address these problems.*

Currently, energy use in data centers is mostly viewed as a problem of insufficient power or cooling capacity, which can only be solved by increasing the power and cooling infrastructure. This view ignores the fact that greater energy efficiency is usually the most cost-effective solution to solving these power and cooling constraints (when considered from the perspective of the TCO for the whole company). In particular, IT managers often do not view their decisions as part of the problem of, or a potential solution to, energy use in data centers. In smaller companies and data centers, these problems stem from a lack of awareness about the true costs of energy use and opportunities to reduce energy use. When data center operators have better information about energy use and organizations are modified to eliminate perverse incentives, the operators will be able to make decisions that reduce life-cycle energy costs and total cost of ownership. A wealth of resources is already available to data center operators trying to improve the efficiency of their operations. The challenge is to make these resources user friendly, credible, and readily accessible, and to motivate operators to take advantage of them.

Strategy 1: Information on best practices — Objective, credible information about the performance of new technologies and best practices, as well as their effect on data center availability, is needed to help raise awareness and reduce the perceived risk of energy

efficiency in data centers. Develop education and training information and curricula about energy efficiency in data centers, emphasizing best practices and cross-functional approaches that span both IT and facilities. This information should build on the best-practices information already available and include other recommendations from this report, such as whole-facility energy performance benchmarking.

Policies:

Federal government: Partner with industry to develop and publish information on field demonstrations and case studies of best practices. These techniques should be identified through assessments of existing data centers which will lead to the identification of best practices. DOE's Save Energy Now program is planning to develop and disseminate best practices and tools through collaboration with industry associations such as Green Grid. Conduct demonstrations using standard metrics to quantify the actual efficiency value of technologies, practices, and designs. Partner with organizations that conduct training for commercial building operators and data center operators to develop education and training resources on energy-efficiency measures and how to implement them without affecting performance and reliability. Require the Federal CIO Council to make these resources available to federal data center operators.

Electric utilities: Offer education and training resources as a component of energy efficiency programs for data centers.

Strategy 2: Awareness campaign — Organize a campaign to make data center operators and managers aware of the benefits of energy efficiency in addressing power and cooling constraints in data centers. Encourage data center operators to take initial steps in monitoring energy performance and conducting energy-efficiency assessments.

Policies:

Federal government: Partner with electric utilities, the data center industry, and energy-efficiency advocacy groups to develop an energy awareness campaign targeted at data center operators both in the federal government and in the private sector. Highlight management techniques that maximize energy-efficiency potential without adversely affecting performance and availability.

Electric utilities, the data center industry and energy-efficiency advocacy groups: Consider partnering with the federal government to develop an energy awareness campaign targeted at data center operators both in the federal government and in the private sector. Highlight management techniques that maximize energy-efficiency potential without adversely affecting performance and availability.

Goal: *Data center efficiency improvements have minimal cost premium over standard practice.*

There are many types of cost associated with data center efficiency improvements, including hardware costs, engineering and design costs, risk, operating costs, and staff training and labor. These costs are a significant barrier to adoption of energy efficiency because operators do not perceive that the value of energy efficiency outweighs the costs. Although other policies described here are intended to address the problem of “perceived value,” in the short run it is often most effective to use financial incentives to directly reduce the incremental cost of efficient products and services. As these energy-efficient products are manufactured in greater volume, their cost tends to decline along with the need for financial incentives. Even though financial

incentives such as tax deductions and utility rebates are most clearly linked to reducing the implementation cost of efficiency improvements, it is important to keep in mind that most policies or programs – including labeling, benchmarking, technical assistance, training, pilot or demonstration projects, research and development – will eventually reduce the overall cost of efficiency improvements, and thus increase the likelihood that these improvements will be undertaken. In other words, direct financial incentives are not the only strategy to reduce the cost of efficiency improvements, but they are a useful first step.

Strategy 1: Financial incentives — Provide financial incentives to encourage adoption of energy-efficient IT equipment and data centers. These incentives are most cost effective if targeted at “upstream” market actors, such as data center design firms or equipment manufacturers. Given that ESPCs are a key priority of the current Administration, they are expected to play an important role going forward.

Policies:

Federal government: Expand the DOE-EERE Save Energy Now assessments to include data center operations. Provide technical assistance for demonstration projects of energy efficiency in data centers, to offset engineering and design costs that would normally be incurred by the facility operator. For federal data centers, target ESPCs and UESCs to data centers while addressing concerns about security and data center reliability. Ensure that ESPC and UESC contractors are educated about energy-efficiency opportunities in federal (both owned and outsourced) data centers.

Electric utilities: Consider providing financial incentives for data center buildings and equipment that meet the ENERGY STAR performance specifications described above. Include in these incentives the cost for operators to conduct energy-efficiency audits, feasibility studies, implement cross-functional teams, etc.

Goal: *Combined heat and power (CHP), based on fuel cells and other technologies, is considered a viable alternative for improving the energy efficiency of data centers.*

Although data center operators are familiar with distributed generation (DG) as a backup power source, the idea of DG as a primary source of power in a CHP arrangement is still considered novel in data centers. Because of concerns about power reliability, data center operators are reluctant to significantly change their power systems. In addition, fuel-cell systems currently have a significant cost premium over standard systems. Increasing the acceptance of DG and CHP in data centers will require incentives and research to reduce the first cost of these technologies, as well as demonstration projects to document the benefits and reduce the perceived risk from these technologies.

Strategy 1: DG/CHP demonstrations — Conduct demonstration and education projects to improve information about fuel cells and other DG technologies used for CHP in data centers.

Policies:

Federal government: The EPA CHP Partnership Program should provide information on available DG/CHP systems, cost, and performance, as well as information for system developers and packagers about the market opportunity within the data center market.

EPA should also define data center configurations, power, and cooling requirements that can be used to simplify the design process, and support selected data center partners in doing preliminary feasibility studies of DG/CHP installations.

Strategy 2: DG/CHP financial incentives — Provide financial incentives to reduce the capital cost of fuel cells and other DG technologies used for CHP in data centers.

Policies:

State governments: Consider offering financial incentives for fuel cells and other technologies used for CHP in high-reliability installations (data centers, telecom facilities, etc.).

Electric utilities: Consider designing and marketing financial incentives for DG and CHP to target the data center sector.

Goal: *Develop an ongoing initiative to bring about lasting improvements in data center energy efficiency.*

Strategy 1: Public/private partnership for energy efficiency in data centers— formulate a common initiative (including public policies and private-sector actions) to promote energy efficiency in data centers. Logical next steps would include defining priorities for the various strategies outlined in this report, developing timelines, defining roles for the various stakeholders, and identifying gaps and issues that require further assessment.

Policies:

Federal government: continue the dialog begun with this report by engaging stakeholders to formulate a common initiative (including public policies and private-sector actions) to promote energy efficiency in data centers.

Data center industry, electric utilities, and other stakeholders: consider partnering with the federal government on a common initiative to promote energy efficiency in data centers.

8.2. Recommendations for Research and Development Activities

Although the policy recommendations outlined above offer immediate opportunities to improve the energy efficiency of data centers, there are also significant longer-term opportunities to improve data center energy efficiency through the development of new technologies and practices. These opportunities span the entire breadth of the data center: computing software; computing, storage and network hardware; the power conversion chain; heat removal; and controls. In all these areas, however, research is needed to better understand the effect of energy-efficient technologies on data center performance, availability, and cost.

The list below identifies R&D strategies that have the significant potential to affect energy use in data centers. It is not a comprehensive list but includes those items identified in the course of researching this report. A final recommendation to prepare a comprehensive R&D plan is included in this list.

Data center R&D may be pursued by federal or state governments or by industry. In particular, the government can play a valuable role by undertaking high-risk early-stage research; conducting research that is objective, vendor-neutral, and publicly available; and by ensuring that the public interest (i.e., energy efficiency) is incorporated into research goals.

8.2.1. Computing Software

- Improve available data, metrics, and test standards to measure computational (or other “output”) energy efficiency of IT equipment (servers, storage, and network equipment); also, facilitate public availability of test applications and workloads that realistically represent the complexity and variability of computational loads.
- Improve virtualization software for all aspects of IT equipment (computation, storage, and networking) to reduce virtualization overhead, allow easier configuration and management, and dynamically match hardware resources to the load.
- Develop, validate, and demonstrate the effectiveness of virtualization software-based system power management.
- Improve software development tools and techniques to allow software to more efficiently use chip-level multiprocessing (improved parallelization).
- Develop better metrics, software development processes, and tools to improve the efficiency of software.
- Improve application resiliency to delayed response times from power-managed storage (or other IT equipment).

8.2.2. IT Hardware (Computing, Storage and Network)

- Improve energy performance of hardware-based virtualization technologies (reduce virtualization overhead).
- Improve server platform power management capabilities.
- Develop lower power states for use at lower utilization levels.
- Improve power management for storage systems, to allow many disks to remain off most of the time with little impact on performance; investigate impact of storage latency.

- Investigate the reliability impacts of thermal cycles in a power-managed server and ways to retard thermal fatigue and thermal failures in low-cost packages.
- Investigate application of solid-state (non-mechanical) storage technologies to data centers.

8.2.3. Power Conversion Chain

- Improve the efficiency of power conversion and backup systems across all load levels and under various redundancy practices.
- Develop efficient modular and scalable power systems.
- Develop components, architectures, and standards for DC-powered data centers.
- For CHP/DG systems, continue technology development in the areas of cost reduction, system control, absorption and adsorption chiller systems, and data-center-specific cooling configurations.
- For CHP/DG installations at data centers, quantify the user and social benefits of increased reliability, reduction in environmental emissions, and operating cost reductions.

8.2.4. Heat Removal

- Develop scientific understanding of the impact of environmental conditions (temperature, humidity, particulates and other pollutants) on IT equipment operation and reliability, to expand operating ranges without decreasing reliability.
- Explore tradeoffs between “hardening” IT equipment for more extreme environments vs. energy costs associated with maintaining tight environmental conditions.
- Develop systems and guidelines for reliable use of outside air economizers (air treatment requirements, necessary filter technology, controls and monitoring systems)
- Test, demonstrate, and evaluate emerging pre-engineered products for liquid-based energy-efficient rack, row, and in-chassis cooling.

8.2.5. Controls and Management

- Improve server operating systems to increase power management enabling rates in servers.
- Develop improved computing control strategies (such as statistical machine learning or control theory) to allow better power management of IT equipment at the system, cluster, and data center level.
- Improve hardware and control algorithms for close-coupled control of IT and HVAC systems.
- Develop active power management strategies for high-performance computing systems, e.g., taking advantage of workload imbalances to reduce the power of lightly loaded system components.
- Develop standard communication protocols to allow continuous energy monitoring and interoperability among IT equipment and data center infrastructure products.
- Develop best practices for improving energy efficiency through storage optimization and server virtualization.

8.2.6. Cross-Cutting Activities

- Develop improved public data sets of data center energy use, performance, and physical characteristics, to allow improved whole-building benchmarking models. Particular emphasis is needed to include the performance of IT equipment in these data sets to allow development of end-to-end data center performance metrics (e.g., Total Data Center Energy Use ÷ Data Processing Output). One avenue to collect these data would be the EIA's Commercial Building Energy Consumption Survey. Also need better information about the energy-use characteristics of non-data center locations for IT equipment (e.g., server closets and rooms).
- Understand and mitigate the impact of energy-efficiency strategies on performance, availability, and cost of the data center (IT, power, and cooling);
 - Identify ways in which energy efficiency affects IT equipment reliability (e.g., disk drive power management or processor thermal cycling).
 - Develop technologies to improve component reliability while improving energy efficiency.
 - Produce information for manufacturers, purchasers, and operators to use to confidently make decisions about the level of redundancy actually required to meet desired levels of reliability.
- Analyze and quantify the net impact of data centers and IT equipment generally on overall U.S. energy use (e.g., trade off between e-commerce and transportation energy).
- Undertake a more comprehensive effort to assess R&D opportunities in data centers, including basic research (e.g., semiconductor materials) and opportunities in software and IT equipment that have the potential for large efficiency gains.

8.3. Recommendations for Further Analysis

The estimates described in this report were based on best available information and data at the time of this study. However, there are inherent uncertainties associated with the data, assumptions, and modeling techniques employed in this study, which limit the accuracy of the resulting estimates. It is therefore recommended that future research be considered in the following areas to improve the study methodology:

- Alternative methods for estimating the energy use of IT equipment and site infrastructure systems beyond the Koomey (2007) approach
- The establishment of uncertainty ranges on the data assumptions and final results to provide indications of data confidence
- Improved estimates on the breakdown of installed U.S. servers by space type over time, including servers in co-located facilities
- Estimates of the numbers of installed custom-built servers, their average energy use, and their projected growth trends
- Improved estimates for storage device energy use and efficiency trends, which could include tape storage devices
- Improved estimates for network equipment energy use and efficiency trends
- Estimates of the energy-efficiency trends of memory technologies
- Improved understanding of where data centers are located and what electricity rates are charged

- Improved estimates for number of servers and data centers operated by and for the federal government
- Improved estimates of typical load shapes by space type
- Inclusion of peak demand charges for calculation of electricity costs

References

- AFCOM. 2006. Five Bold Predictions For The Data Center Industry That Will Change Your Future [Keynote Slides]. AFCOM Data Center Institute. March.
http://www.afcom.com/files/PDF/DCI_Keynote_Final.pdf (accessed April 6, 2007).
- AFCOM. 2007. *Personal Communication: Survey Results & Analysis for Energy Efficiency in the Data Center*. April 9.
- AMD. 2006. *Power and Cooling in the Data Center*. Advanced Micro Devices. 34246C.
http://enterprise.amd.com/Downloads/34146A_PC_WP_en.pdf.
- Antonopoulos, Andreas. 2006. The Four Main Data Center Trends. *NetworkWorld magazine*, May 16.
<http://www.networkworld.com/newsletters/datacenter/2006/0515datacenter1.html>.
- APC. 2003. *Alternative Power Generation Technologies for Data Centers and Network Rooms*. American Power Conversion. #64, Revision 1 (also published in Data Center Journal).
- Apiki, Steve. 2005. Sarbanes-Oxley: Driving the Storage Compliance Boom. February 25.
<http://www.enterprisestorageforum.com/continuity/features/article.php/3485651>.
- ASE. 2002. *Funds for Energy Efficiency Projects*. Washington, DC: Alliance to Save Energy. April 23. http://www.ase.org/uploaded_files/intl_eefunds_april23.pdf.
- ASE. 2005. State Energy Efficiency Index: Tax Incentives [web page]. Alliance to Save Energy.
<http://www.ase.org/content/article/detail/2607> (accessed March, 2007).
- ASHRAE. 2004a. *ASHRAE Standard: Energy Standard for Buildings Except Low-Rise Residential Buildings, I-P Edition*. Atlanta, GA: American Society of Heating, Refrigerating and Air Conditioning Engineers. Standard 90.1.
- ASHRAE. 2004b. *Thermal Guidelines for Data Processing Environments*. Atlanta, GA: American Society of Heating, Refrigerating and Air Conditioning Engineers.
- ASHRAE. 2005. *Datacom Equipment Power Trends and Cooling Applications*. Atlanta, GA: American Society of Heating, Refrigerating and Air Conditioning Engineers.
- Bailey, M., M. Eastwood, T Grieser, L. Borovick, V. Turner, and R.C. Gray. 2007. *Special Study: Data Center of the Future*. New York, NY: IDC. IDC #06C4799. April.
- Barroso, Luiz André. 2005. The Price of Performance: An Economic Case for Chip Multiprocessing. *ACM Queue*, September.
<http://labs.google.com/papers/priceofperformance.html>.
- BCAP. 2007. Building Codes Assistance Project. March. http://www.bcap-energy.org/map_page.php.
- Beard, Daniel P. 2007. *Green the Capitol Initiative: Final Report*. Washington, DC: U.S. House of Representatives, Chief Administrative Officer. June 21.
<http://cao.house.gov/greencapitol/green-the-capitol-final-report.pdf>.
- Belady, Christian L. 2007. In the Data Center, Power and Cooling Costs More Than the IT Equipment it Supports. *Electronics Cooling*. vol. 13, no. 1. <http://electronics-cooling.com/articles/2007/feb/a3/>.
- Bodik, P., M. Armbrust, K. Canini, A. Fox, M. Jordan, and D. Patterson. 2006. *A Case for Adaptive Datacenters to Conserve Energy and Improve Reliability*. Berkeley, CA: University of California, Berkeley Reliable Adaptive Distributed (RAD) Systems Laboratory. <http://radlab.cs.berkeley.edu/>.

- Brill, Kenneth G. 2007. *Data Center Energy Efficiency and Productivity*. Santa Fe, NM: The Uptime Institute.
[www.uptimeinstitute.org/symp_pdf/\(TUI3004C\)DataCenterEnergyEfficiency.pdf](http://www.uptimeinstitute.org/symp_pdf/(TUI3004C)DataCenterEnergyEfficiency.pdf).
- Brown, Elizabeth, Patrick Quinlan, Harvey Sachs, and Daniel Williams. 2002. *Tax Credits for Energy Efficiency and Green Buildings: Opportunities for State Action*. Washington, DC: American Council for an Energy-Efficient Economy. E021. March.
<http://aceee.org/getfile.cfm?publicationid=23>.
- Brown, Richard E., and Jonathan G. Koomey. 2003. Electricity Use in California: Past Trends and Present Usage Patterns. *Energy Policy*. vol. 31, no. 9, pp. 849-864 (also LBNL-47992). July.
- Bryson, T., K. Darrow, K. Davidson, and B. Major. 2001. *Advanced Microturbine System (AMTS) Market Study, prepared for DOE and Capstone Turbine Corporation*. Onsite Energy Corporation. April.
- Bull, Pierre. 2007. *Personal Communication*. NYSERDA. March 26.
- Carr, Nicholas G. 2005. The End of Corporate Computing. *MIT Sloan Management Review*. vol. 46, no. 3, pp. 67-73. spring.
- CEE. 2007a. Consortium for Energy Efficiency. <http://www.cee1.org>.
- CEE. 2007b. Estimated 2006 Total Energy-Efficiency Budgets for Electric and Gas Programs [web page]. Consortium for Energy Efficiency. <http://www.cee1.org/ee-pe/budgets-main.php3> (2007).
- CNET. 2006. Verizon heeds call of fuel cells. CNET News.com. August 7.
http://news.com.com/Verizon+heeds+call+of+fuel+cells/2100-1033_3-6102552.html.
- comScore Networks. 2007. comScore Networks Releases Top Web Properties Worldwide for December; Reviews Biggest Gainers for 2006. comScore Networks. January 31.
<http://www.comscore.com/press/release.asp?press=1193> (accessed April 10, 2007).
- Data Center Knowledge. 2007. Regional Markets [Weblog]. March 24.
<http://www.datacenterknowledge.com/> (accessed March 30, 2007).
- Delaney, Kevin J., and Rebecca Smith. 2006. Surge in Internet Use, Energy Costs Has Big Tech Firms Seeking Power. *Wall Street Journal*, June 13, A1.
- Dell, Inc. 2007. Datacenter Capacity Planner. <http://www.dell.com/calc>.
- Dietrich, Jay. 2007. *Personal Communication*. IBM. March 19.
- Digital Realty Trust. 2007. *Comments submitted regarding April 23, 2007 Public Review Draft—Report to Congress on Server and Data Center Energy Efficiency Public Law 109-431*.
- DSIRE. 2007. Database of State Incentives for Renewables & Efficiency.
<http://www.dsireusa.org/>.
- Dubie, Denise. 2007. Virtualization ROI Hard to Quantify. *NetworkWorld*, 27 March.
<http://www.networkworld.com/news/2007/032707-ca-virtualization-roi.html>.
- Dunn, Darrell. 2006a. Power Surge. *Information Week*, February 27.
- Dunn, Darrell. 2006b. The Best And Worst Cities For Data Centers. *InformationWeek*, October 23. <http://www.informationweek.com/story/showArticle.jhtml?articleID=193401105>.
- Emerson Network Power. 2007. *Personal Communication*: Spring 2007 Data Center Users' Group Survey Results. May 9.
- Energy and Environmental Analysis. 2004. Distributed Generation Operational Reliability and Availability Database, Oak Ridge National Laboratory. January. http://www.eea-inc.com/dgchp_reports/FinalReportORNLDGREL.pdf.

- Energy and Environmental Analysis. 2006. Combined Heat and Power Database, maintained for U.S. Department of Energy and Oak Ridge National Laboratory. <http://www.eea-inc.com/chpdata/index.html>.
- Engle, David. 2005. Power Sufficiency with 'Chilling' Efficiency. *Distributed Energy Journal*, July/August.
- Eubank, H., J. Swisher, C. Burns, Jen Seal, and B. Emerson. 2003. *Design Recommendations for High-Performance Data Centers: Report of the Integrated Design Charrette*. Snowmass, CO: Rocky Mountain Institute. February 2-5.
- Fan, Xiaobo, Wolf-Dietrich Weber, and Luiz André Barroso. 2007. Power Provisioning for a Warehouse-sized Computer. Proceedings of the 34th International Symposium on Computer Architecture in San Diego, CA. Association for Computing Machinery, ISCA '07, http://labs.google.com/papers/power_provisioning.pdf.
- FEMP. 2006a. *Annual Report to Congress on Federal Government Energy Management and Conservation Programs Fiscal Year 2005*. Washington, DC: US Department of Energy, Federal Energy Management Program. September 26. <http://www1.eere.energy.gov/femp/pdfs/annrep05.pdf>.
- FEMP. 2006b. *ENERGY STAR-qualified and FEMP-designated Products* Washington, DC: US Department of Energy, Federal Energy Management Program. April 25. http://www1.eere.energy.gov/femp/pdfs/eep_productfactsheet.pdf.
- FEMP. 2006c. *Guidance for Electric Metering in Federal Buildings*. Washington, DC: US Department of Energy, Federal Energy Management Program. DOE/EE-0312. February 3. http://www1.eere.energy.gov/femp/pdfs/adv_metering.pdf.
- Fuertes, Gwen. 2007. *Personal Communication*. Washington, DC: US Green Building Council. March 23.
- Gonsalves, A. 2007. Micron Predicts Flash Drives In The Data Center. *Information Week*, May 23.
- Greenberg, Steve, Evan Mills, Bill Tschudi, Peter Rumsey, and Bruce Myatt. 2006. Best Practices for Data Centers: Lessons Learned from Benchmarking 22 Data Centers. Proceedings of the ACEEE Summer Study on Energy Efficiency in Buildings in Asilomar, CA. ACEEE, August. Vol 3, pp 76-87. <http://eetd.lbl.gov/emills/PUBS/PDF/ACEEE-datacenters.pdf>.
- Greer, D. 2006. Chip Makers Turn to Multicore Processors. *IEEE Computer*, May.
- Guha, Alope, and Shawn Ouderkirk. 2006. *Disk Failure Rates and Implications of Enhanced MAID Storage Systems*. Copan Systems. April.
- Hamachi-LaCommare, Kristina , Etan Gurrman, Chris Marnay, Peter Chan, and Katie Coughlin. 2004. *A New Approach for Modeling the Peak Utility Impacts from a Proposed Commercial Unitary Air-Conditioning Standard*. Berkeley, CA: Lawrence Berkeley National Laboratory. LBNL-55269. October. <http://eetd.lbl.gov/ea/emp/reports/55269.pdf>.
- Harris, Jeffrey, Matt Brown, John Deakin, Steve Jurovics, Afroz Khan, Ed Wisniewski, James Mapp, Barbara Smith, Melissa Podeszwa, and Alison Thomas. 2004. Energy-Efficient Purchasing by State and Local Government: Triggering a Landslide Down the Slippery Slope to Market Transformation. Proceedings of the 2004 ACEEE Summer Study on Energy Efficiency in Buildings in Asilomar, CA. ACEEE, <http://www.dc.lbl.gov/LBNLDC/publications/Energy%20Efficient%20Purchasing%20by%20State%20and%20Local%20Government.pdf>.

- Hennessy, J., and D. Patterson. 2007. *Computer Architecture: A Quantitative Approach, 4th edition*. San Francisco: Morgan Kaufman. (accessed
- Herrlin, Magnus. 2006. *Relationship Between Equipment Reliability and Energy Efficiency*. Sunnyvale, CA: ENERGY STAR Conference on Enterprise Servers and Data Centers: Opportunities for Energy Savings. February 1.
- Huber, Peter, and Mark P. Mills. 1999. Dig more coal—the PCs are coming. *Forbes*, May 31. pp. 70-72. <http://forbes.com/forbes/99/0531/6311070a.htm>.
- Hughes, Ron. 2005. Data Centers of the Future [web page]. May 24. http://www.datacenterjournal.com/News/Article.asp?article_id=319 (accessed April 2, 2007).
- IDC. 2007a. *Demand for Increased IT Efficiency Drives Worldwide Server Virtualization Adoption, IDC Says*. Framingham, MA: IDC. 22 January. <http://www.idc.com/getdoc.jsp?containerId=prUS20528507>.
- IDC. 2007b. *IDC's Worldwide Installed Base Forecast, 2007-2010*. Framingham, MA: IDC. March.
- IDC. 2007c. *Virtualization and Multicore Innovations Disrupting the Worldwide Server Market, According to IDC*. Framingham, MA: IDC. March 20. <http://www.idc.com/getdoc.jsp?containerId=prUS20609907>.
- IDC. 2007d. *Worldwide Server Market Gains Momentum in Fourth Quarter as Market Revenues Exceed \$52 Billion in 2006, According to IDC*. Framingham, MA: IDC. 26 February. <http://www.idc.com/getdoc.jsp?containerId=prUS20577707>.
- Intel. 2007. Intel Xeon Processors [web page]. Intel Corp. http://www.intel.com/business/xeon/?cid=cim:ggl|xeon_us_quadxeon|k78C6|s (accessed March 30, 2007).
- Johnson, Anne. 2007. *Personal Communication: Growth trends in data centers power use*. Austin, TX: Austin Energy. email, March 9.
- Kawamoto, Kaoru, Jonathan Koomey, Bruce Nordman, Richard E. Brown, Maryann Piette, Michael Ting, and Alan Meier. 2002. Electricity Used by Office Equipment and Network Equipment in the U.S. *Energy--The International Journal*. vol. 27, no. 3, pp. 255-269 (also LBNL-45917). March.
- Koomey, Jonathan, Christian Belady, Henry Wong, Rob Snevely, Bruce Nordman, Ed Hunter, Klaus-Dieter Lange, Roger Tiple, Greg Darnell, Matthew Accapadi, Peter Rumsey, Brent Kelley, Bill Tschudi, David Moss, Richard Greco, and Kenneth Brill. 2006. *Server Energy Measurement Protocol*. Oakland, CA: Analytics Press. November 3. <http://energystar.gov/ia/products/downloads/Finalserverenergyprotocol-v1.pdf>.
- Koomey, Jonathan, Chris Calwell, Skip Laitner, Jane Thornton, Richard E. Brown, Joe Eto, Carrie Webber, and Cathy Cullicott. 2002. "Sorry, wrong number: The use and misuse of numerical facts in analysis and media reporting of energy issues." In *Annual Review of Energy and the Environment 2002*. Edited by R. H. Socolow, D. Anderson and J. Harte. Palo Alto, CA: Annual Reviews, Inc. (also LBNL-50499). 119-158 pp.
- Koomey, Jonathan, Huimin Chong, Woonsien Loh, Bruce Nordman, and Michele Blazek. 2004. Network electricity use associated with wireless personal digital assistants. *The ASCE Journal of Infrastructure Systems*. vol. 10, no. 3, pp. 131-137 (also LBNL-54105). September.
- Koomey, Jonathan G. 2007. *Estimating Total Power Consumption by Servers in the U.S. and the World*. February 15. <http://enterprise.amd.com/Downloads/svrpwrusecompletefinal.pdf>.

- Koomey, Jonathan, Kaoru Kawamoto, Bruce Nordman, Mary Ann Piette, and Richard E. Brown. 1999. *Initial comments on 'The Internet Begins with Coal'*. Berkeley, CA: Lawrence Berkeley National Laboratory. LBNL-44698. December 9.
<http://enduse.lbl.gov/projects/infotech.html>.
- Kueck, Jim. 2007. *Personal Communication*. Honeywell. March 8.
- Kumar, Rakesh. 2006. *A Message From Data Center Managers to CIOs: Floor Space, Power and Cooling Will Limit our Growth*. Gartner RAS. G00142393.
<http://www.sun.com/servers/coolthreads/overview/docs/Sun1617.pdf>.
- Kushler, Martin, Dan York, and Patti Witte. 2004. Five Years In: An Examination of the First Half-Decade of Public Benefits Energy Efficiency Policies [web page]. American Council for an Energy-Efficient Economy. April.
http://www.aceee.org/briefs/apr04_01.htm.
- LBNL. 2006. High-Performance Buildings for High-Tech Industries, Data Centers. Lawrence Berkeley National Laboratory. <http://hightech.lbl.gov/datacenters.html> (accessed April 6, 2007).
- Loper, Joe, and Sara Parr. 2007. *Energy Efficiency in Data Centers: A New Policy Frontier*. Washington, DC: Alliance to Save Energy. January.
http://www.ase.org/files/3581_file_data_center_energy.pdf.
- Loper, Joe, Lowell Ungar, David Weitz, and Harry Misuriello. 2005. *Building on Success: Policies to Reduce Energy Waste in Buildings*. Washington, DC: Alliance to Save Energy. July. <http://www.ase.org/images/lib/buildings/Building%20on%20Success.pdf>.
- Markoff, John. 2007. Intel Plans Faster Chips That Also Save Power. *New York Times*, March 29. <http://www.nytimes.com/2007/03/29/technology/29chip.html>.
- MEA. 2006. Green Building Tax Credit. Maryland Energy Administration.
<http://www.energy.state.md.us/programs/commercial/greenbuilding/index.html> (accessed March, 2007).
- Mehta, Stephanie N. 2006. Behold the server farm! Glorious temple of the information age! *Fortune*, August 7.
http://money.cnn.com/magazines/fortune/fortune_archive/2006/08/07/8382587/index.htm
- Menuet, Rob, and W. Pitt Turner. 2006. *Continuous Cooling is Required for Continuous Availability*. Santa Fe, NM: The Uptime Institute.
http://www.uptimeinstitute.org/Onsite_Downloads/pdf/4-TUI-%20CONTINUOUS%20COOLING_Sec.pdf.
- Mueller, Karen. 2007. *Personal Communication*: Benefits from DOE's Building Energy Code Program. Pacific Northwest National Lab. email,
- Nadel, Steven, and Bill Prindle. 2004. *Supplementary Information on Energy Efficiency for the National Commission on Energy Policy*. Washington, DC: American Council for an Energy Efficient Economy. July.
<http://www.energycommission.org/files/finalReport/III.2.c%20-%20Supplemental%20Info%20on%20EE.pdf>.
- NEMA. 2006. The Commercial Buildings Deduction (Interim Lighting Rules) [web page]. National Electric Manufacturers Association.
http://www.lightingtaxdeduction.org/tax5_2.html (accessed March, 2007).

- NYDEC. 2007. New York State Green Building Initiative [web page]. New York State Department of Environmental Conservation.
<http://www.dec.state.ny.us/website/ppu/grnblgd/index.html> (accessed March, 2007).
- NYSERDA. 2007. Next Generation Emerging Technologies for End-Use Efficiency: Program Opportunity Notice 1105. New York State Energy Research and Development Administration. <http://www.nyserdera.org/funding/funding.asp?i=2>.
- OFEE. 2007. *Strengthening Federal Environmental, Energy, and Transportation Management*. Washington, DC: Office of the Federal Environmental Executive. Executive Order 13423. January 24. http://www.ofee.gov/eo/EO_13423.pdf.
- Oregon DOE. 2007. Business Energy Tax Credits. Oregon Department of Energy. <http://egov.oregon.gov/ENERGY/CONS/BUS/BETC.shtml> (accessed March, 2007).
- Osterberg, Ken. 2007. *Personal Communication*: Data Center Hard Disk Drive Shipments and Power Levels. Seagate Technology. March 13.
- Patterson, M. K., A. Pratt, and P. Kumar. 2006. *From UPS to Silicon: an End-to-End Evaluation of Data Center Efficiency*. Santa Clara, CA: Enterprise Servers and Data Centers: Opportunities for Energy Savings Conference. January 31.
http://www.energystar.gov/ia/products/downloads/MPatterson_APratt_Case_Study.pdf.
- Pflueger, J., and S. Hanson. 2007. *Data Center Efficiency in the Scalable Enterprise*. Dell Power Solutions. February. <http://www.dell.com/downloads/global/power/ps1q07-20070210-CoverStory.pdf>.
- PG&E. 2006. High Tech Energy Efficiency Incentives. Pacific Gas & Electric Co. http://www.pge.com/biz/rebates/hightech/htee_incentives.html.
- Pinheiro, Eduardo, Wolf-Dietrich Weber, and Luiz André Barroso. 2007. Failure Trends in a Large Disk Drive Population. Proceedings of the 5th USENIX Conference on File and Storage Technologies (FAST'07) in San Jose, CA. February.
http://labs.google.com/papers/disk_failures.html.
- Pouchet, Jack. 2007. *Personal Communication*. Emerson Network Power/Liebert Corporation. March 9.
- PowerPulse.net. 2006. EPA and DOE's Energy Star Seek Comments on Proposed Energy Management Protocol. PowerPulse.net. August 22.
<http://www.powerpulse.net/story.php?storyID=15822> (accessed August 24, 2006).
- Ranganathan, Parthasarathy, and Norman Jouppi. 2005. Enterprise IT Trends and Implications for Architecture Research. Proceedings of the 11th Int'l Symposium on High-Performance Computer Architecture (HPCA-11 2005) in San Francisco. February 12-16.
http://www.hpcacnf.org/hpca11/papers/24_x_ranganathan-enterpriseittrends.pdf.
- Rasmussen, Neil. 2006. *Electrical Efficiency Modeling of Data Centers*. APC. White paper #113.
<http://www.apc.com>.
- Raths, David. 2006. Energy Hogs on the Server Farm. *Public CIO*, December 19.
<http://www.public-cio.com/story.print.php?id=102970>.
- Robertson, Chris. 2003. *Rx for Health Care Power Failures*. Distributed Energy, Primen. November.
- Rowe, Joe. 2007. *Personal Communication*. VMware.
- Schmidt, Roger, and Don Beaty. 2005. ASHRAE Committee Formed to Establish Thermal Guidelines for Datacom Facilities. *Electronics Cooling*, February.

- Shankland, S. 2006. Dell sells premium energy-efficient servers. CNET News.com. December 4. http://news.com.com/Dell+sells+premium+energy-efficient+servers/2100-1010_3-6140669.html.
- SPEC. 2007. SPEC Power and Performance Committee. Standard Performance Evaluation Corporation. <http://www.spec.org/specpower/>.
- Sporer, M. 2007. *The Power Demands of Data Centers Require Memory Innovations*. Boise, Idaho: Micron Technology. http://download.micron.com/pdf/whitepapers/low_power.pdf.
- SprayCool, Inc. 2006. SprayCool's SprayCool Technology Qualifies for Avista Utilities Rebate Program to Help Address Data Center Cooling, Power Issues [press release]. November 28. <http://www.spraycool.com/news/PR/061128.asp>.
- State of Nevada. 2007. Nevada Revised Statutes, Chapter 361.0775 [web page]. <http://www.leg.state.nv.us/NRS/NRS-361.html> (accessed March, 2007).
- Sun Microsystems, Inc. 2006. Sun Teams with PG&E On First Ever Rebate For Servers, Up To \$1000 In Savings [press release]. August 15. <http://www.sun.com/smi/Press/sunflash/2006-08/sunflash.20060815.2.xml>.
- Telegeography. 2006. US VoIP Revenue & Subscribers Post Triple Digit Gain in 2005. March 7. <http://www.telegeography.com/press/releases/2006-03-07.php> (accessed April 5, 2007).
- The Green Grid. 2007. *Green Grid Metrics: Describing Data Center Power Efficiency*. The Green Grid, Technical Committee. February 20. http://www.thegreengrid.org/downloads/Green_Grid_Metrics_WP.pdf.
- The Tax Incentives Assistance Project. 2007. Business Tax Incentives [web page]. http://www.energytaxincentives.org/business/commercial_buildings.php (accessed March, 2007).
- Thibodeau, Patrick. 2007. Utility looks to cut energy use in data centers: Pacific Gas & Electric is pushing efforts to save on energy costs [web page]. Computerworld. March 28. <http://www.computerworld.com/action/article.do?command=viewArticleBasic&articleId=9014674> (accessed March 28, 2007).
- Top500. 2006. Top500 List – November 2006. <http://www.top500.org/list/2006/11/100>.
- Tremblay, M. 2006. New Approaches to Efficient Computing. *Electronic Products*, December.
- Tschudi, William, Tengfang Xu, Dale Sartor, and Jay Stein. 2003. *High Performance Data Centers: A Research Roadmap*. Berkeley, CA: Lawrence Berkeley National Laboratory. LBNL-53483. http://hightech.lbl.gov/documents/DataCenters_Roadmap_Final.pdf.
- Turner, W. Pitt, John H. Seader, and Kenneth G. Brill. 2005. *Industry Standard Tier Classifications Define Site Infrastructure Performance*. Santa Fe, NM: Uptime Institute.
- US Census Bureau. 2006. *Annual Survey of Manufactures: Statistics for Industry Groups and Industries*. Washington, DC: U.S. Census Bureau. M05(AS)-1. November.
- US DOE. 2002a. *Technical Support Document: Energy Efficiency Standards for Consumer Products: Residential Central Air Conditioners and Heat Pumps*. Washington, DC: U.S. Department of Energy, Assistant Secretary, Energy Efficiency and Renewable Energy, Office of Building Research and Standards. May. http://www.eere.energy.gov/buildings/appliance_standards/residential/ac_central_1000_r.html.
- US DOE. 2002b. Technology Procurement. Energy Efficiency and Renewable Energy, Building Technologies Program, Emerging Technologies. October.

- <http://www.eere.energy.gov/buildings/emergingtech/printable/page1.html> (accessed April 4, 2007).
- US DOE. 2005. Home Energy Use and Costs: Residential Energy Consumption Survey (RECS) [web page]. EIA, Energy Information Administration, U.S. Department of Energy, Washington, DC. August 22. <http://www.eia.doe.gov/emeu/recs/contents.html> (accessed October 31, 2005).
- US DOE. 2006a. History of Federal Appliance Standards [web page]. U.S. Department of Energy, Building Technologies Program. August 10. http://www.eere.energy.gov/buildings/appliance_standards/history.html (accessed March, 2007).
- US DOE. 2006b. *National Electric Transmission Congestion Study*. Washington, DC: U.S. Department of Energy, Office of Electricity Delivery & Energy Reliability. August. http://www.oe.energy.gov/epa_sec1221.htm.
- US DOE. 2007a. *Annual Energy Outlook 2007, with Projections to 2030*. Washington, DC: Energy Information Administration, U.S. Department of Energy. DOE/EIA-0383(2007). February. [http://www.eia.doe.gov/oiaf/aeo/pdf/0383\(2007\).pdf](http://www.eia.doe.gov/oiaf/aeo/pdf/0383(2007).pdf).
- US DOE. 2007b. Appliances & Commercial Equipment Standards. US Department of Energy, Energy Efficiency and Renewable Energy. February. http://www.eere.energy.gov/buildings/appliance_standards/.
- US DOE. 2007c. Intel and Raytheon Developing More Efficient Computer Processors [EERE Network News]. U.S. Department of Energy, Energy Efficiency and Renewable Energy. April 4. <http://www.eere.energy.gov/news/enn.cfm>.
- US EPA. 2005. *Summary of State Clean Energy-Environment Actions*. Washington, DC: Environmental Protection Agency. November. <http://www.epa.gov/cleanenergy/pdf/summary-matrix.pdf>.
- US EPA. 2006. *ENERGY STAR Overview of 2006 Achievements*. Washington, DC: U.S. Environmental Protection Agency, Climate Protection Partnerships Division. March 1. http://www.epa.gov/appdstar/pdf/2006_Achievements_Overview.pdf.
- US EPA. 2007. *Working Group Notes from the EPA Technical Workshop on Energy Efficient Servers and Datacenters*. Santa Clara, CA: U.S. Environmental Protection Agency. February 16.
- Wagner, Mark. 2007. *Personal Communication*. Johnson Controls, Inc. March 8.
- Warmenhoven, Dan. 2005. Three Years Later, A Look At Sarbanes-Oxley. *Forbes*, July 27. http://www.forbes.com/home/columnists/2005/07/27/warmenhoven-opinion-regulatoincz_dw_0726compliance.html.
- Wigle, Lorie. 2007. *Personal Communication*. Intel. January 30.
- Winkler, Katherin. 2007. *Personal Communication*. EMC. March 8.
- Wong, Henry. 2007. *Personal Communication: EPA Datacenter Study IT Equipment Feedback Summary*. Intel Digital Enterprise Group. March.
- Zaino, J. 2006. *Deft Consolidation: Agencies turn to server virtualization to reverse the risk and liability of costly computer sprawl*. fcw.com. December 4. <http://www.fcw.com/article96933-12-04-06-Print>.
- Ziff Davis. 2005. *Power Consumption and Cooling in the Data Center: A Survey*. New York, NY: Ziff Davis. December. http://internet.ziffdavis.com/amd_assets/AMD_PowerConsumptionCooling.pdf.

Acknowledgments

Andrew Fanara, ENERGY STAR Product Development team lead, Climate Protection Partnerships Division, U.S. EPA, oversaw the development of this report. Other EPA staff that contributed to this report include Jill Abelson, Ann Bailey, Kim Crossman, Robin Shudak, Alexandra Sullivan, Maria Vargas, and Mike Zatz.

Technical support was provided by staff from *Lawrence Berkeley National Laboratory*, including Rich Brown, Peter Chan, Joe Eto, Sondra Jarvis, Jonathan Koomey, Eric Masanet, Bruce Nordman, Dale Sartor, Arman Shehabi, John Stanley, and Bill Tschudi.

Other organizations providing technical and logistical support in the preparation of this report include:

The Alliance to Save Energy: Steve Capana and Joe Loper;
ICF Incorporated: Rebecca Duff, Evan Haines, and Bruce Hedman;
ERG Incorporated: Danielle Sass.

The following Federal agencies provided inputs to and comments on this report:

U.S. Department of Energy Industrial Technologies Program
U.S. Department of Energy Federal Energy Management Program
U.S. Department of Commerce

We would like to acknowledge and thank the following organizations for providing input data to this study and commenting on draft versions of the report:

3M
Active Power
Advanced Micro Devices
AFCOM
American Power Conversion
American Society of Heating, Refrigerating and Air-Conditioning Engineers
Austin Energy
Bloom Energy
CGI Federal
Clustered Systems
Consortium for Energy Efficiency
Copan Systems
Datastructure Management
Degree Controls
Dell
Digital Realty Trust
The Green Grid Association
E Source
Eaton Electrical
Ecos Consulting
EMC
Emerson Network Power

EYP Mission Critical Facilities
FileTek
Google
Hewlett-Packard
Honeywell
IDC
Information Technology Industry Council
Intel
International Business Machines
Lehman Brothers
LV Power
Michaud Cooley Erickson
Micron Technology
Modius
Morrison Hershfield
Natural Resources Defense Council
New York State Energy Research and Development Authority
ON Semiconductor
Pacific Gas & Electric Company
Pacific Northwest National Laboratory
RPM Electric
Sabey DataCenters
Seagate Technology
Silicon Graphics
Southern California Edison
SprayCool
Standard Performance Evaluation Corporation
Sun Microsystems
Supermicro Computer
Sutter Health
Syska Hennessy Group
Uptime Institute
U.S. Green Building Council
Verizon
VMWare