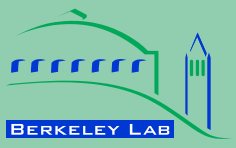
LBNL-1005775




ERNEST ORLANDO LAWRENCE
BERKELEY NATIONAL LABORATORY

# United States Data Center Energy Usage Report

**Arman Shehabi, Sarah Smith, Dale Sartor, Richard Brown, Magnus Herrlin**
*Environmental and Energy Impact Division, Lawrence Berkeley National Laboratory*

**Jonathan Koomey**
*Steyer-Taylor Center for Energy Policy and Finance, Stanford University*

**Eric Masanet**
*McCormick School of Engineering, Northwestern University*

**Nathaniel Horner, Inês Azevedo**
*Climate and Energy Decision Making Center, Carnegie Mellon University*

**William Lintner**
*Federal Energy Management Program, U.S. Department of Energy*

**June 2016**

This work was supported by the Federal Energy Management Program of the U.S. Department of Energy under Lawrence Berkeley National Laboratory Contract No. DE-AC02-05CH1131

## Disclaimer

## Acknowledgments

The Green Grid Association provided comments to draft versions of this report that were contributed by their members representing major data center networking equipment manufacturers, server and storage equipment manufacturers, major software providers, and large data center end users/owners.

Input data for this report was provided by the Green Grid Association, Data Center Dynamics, the IT Industry Council, and the International Data Corporation.

We would also like to acknowledge and thank the expert reviewers from approximately 30 organizations that took the time to answer questions from the authors and provided comments on draft versions of the report.

The research reported in this report was conducted by Lawrence Berkeley National Laboratory with support from the Department of Energy Federal Energy Management Program. Lawrence Berkeley National Laboratory is supported by the Office of Science of the United States Department of Energy and operated under Contract Grant No. DE-AC02-05CH11231.

## Citation

Shehabi, A., Smith, S.J., Horner, N., Azevedo, I., Brown, R., Koomey, J., Masanet, E., Sartor, D., Herrlin, M., Lintner, W. 2016. United States Data Center Energy Usage Report. Lawrence Berkeley National Laboratory, Berkeley, California. LBNL-1005775

# Table of Contents

# Table of Figures

# Table of Tables

# Executive Summary

This report estimates historical data center electricity consumption back to 2000, relying on previous studies and historical shipment data, and forecasts consumption out to 2020 based on new trends and the most recent data available. Figure ES-1 provides an estimate of total U.S. data center electricity use (servers, storage, network equipment, and infrastructure) from 2000-2020. In 2014, data centers in the U.S. consumed an estimated 70 billion kWh, representing about 1.8% of total U.S. electricity consumption. Current study results show data center electricity consumption increased by about 4% from 2010-2014, a large shift from the 24% percent increase estimated from 2005-2010 and the nearly 90% increase estimated from 2000-2005. Energy use is expected to continue slightly increasing in the near future, increasing 4% from 2014-2020, the same rate as the past five years. Based on current trend estimates, U.S. data centers are projected to consume approximately 73 billion kWh in 2020.

Many factors contribute to the overall energy trends found in this report, though the most conspicuous change may be the reduced growth in the number of servers operating in data centers. While shipments of new servers into data centers continue to grow every year, the growth rate has diminished over the past 15 years. From 2000-2005, server shipments increased by 15% each year resulting in a near doubling of servers operating in data centers. From 2005-2010, the annual shipment increase fell to 5%, partially driven by a conspicuous drop in 2009 shipments (most likely from the economic recession), as well as from the emergence of server virtualization across that 5-year period. The annual growth in server shipments further dropped after 2010 to 3% and that growth rate is now expected to continue through 2020. This 3% annual growth rate coincides with the rise in very large “hyperscale” data centers and an increased popularity of moving previously localized data center activity to colocation or cloud facilities. In fact, nearly all server shipment growth since 2010 occurred in servers destined for large hyperscale data centers, where servers are often configured for maximum productivity and operated at high utilization rates, resulting in fewer servers needed in the hyperscale data centers than would be required to provide the same services in traditional, smaller, data centers.

Along with total server count, the power demand for each server has also changed. While server power requirements were observed to be increasing from 2000-2005, power demand appears to have stayed fairly constant since 2005. Additionally, servers are improving in their power scaling abilities, thus reducing power draw during idle periods or when at low utilization. Efficiency improvements in storage, network and infrastructure also influence the electricity estimates in this report. Storage devices are becoming more efficient on a per-drive basis, with the growth in drive storage capacity projected to outpace increases in data storage demand by 2020, ultimately reducing the number of physical drives needed throughout data centers. Recent estimates of network port power consumption are now much lower than estimates from the past decade. Increased awareness in data center infrastructure operations (e.g. cooling) has resulted in improved efficiency across data center types, though the most significant infrastructure impact observed in this report is the recent growth in hyperscale data centers that are often innovatively designed to maximum infrastructure efficiency.

The combination of these efficiency trends has resulted in a relatively steady U.S data center electricity demand over the past 5 years, with little growth expected for the remainder of this decade. It is important to note that this near constant electricity demand across the decade is occurring while simultaneously meeting a drastic increase in demand for data center services; data center electricity use would be significantly higher without these energy efficiency improvements. A counterfactual scenario was created for this study that estimates what data center energy consumption would have been if industry energy-savings efforts were halted in 2010. For this scenario, the follow metrics remain static at 2010 industry-wide levels from 2010-2020:

- Average server utilization
- Server power scaling at low utilization
- Average power draw of hard disk drives
- Average power draw of network ports
- Average infrastructure efficiency (i.e., PUE)

The resulting electricity demand, shown in Figure ES-1, indicates that more than 600 additional billion kWh would have been required across the decade.

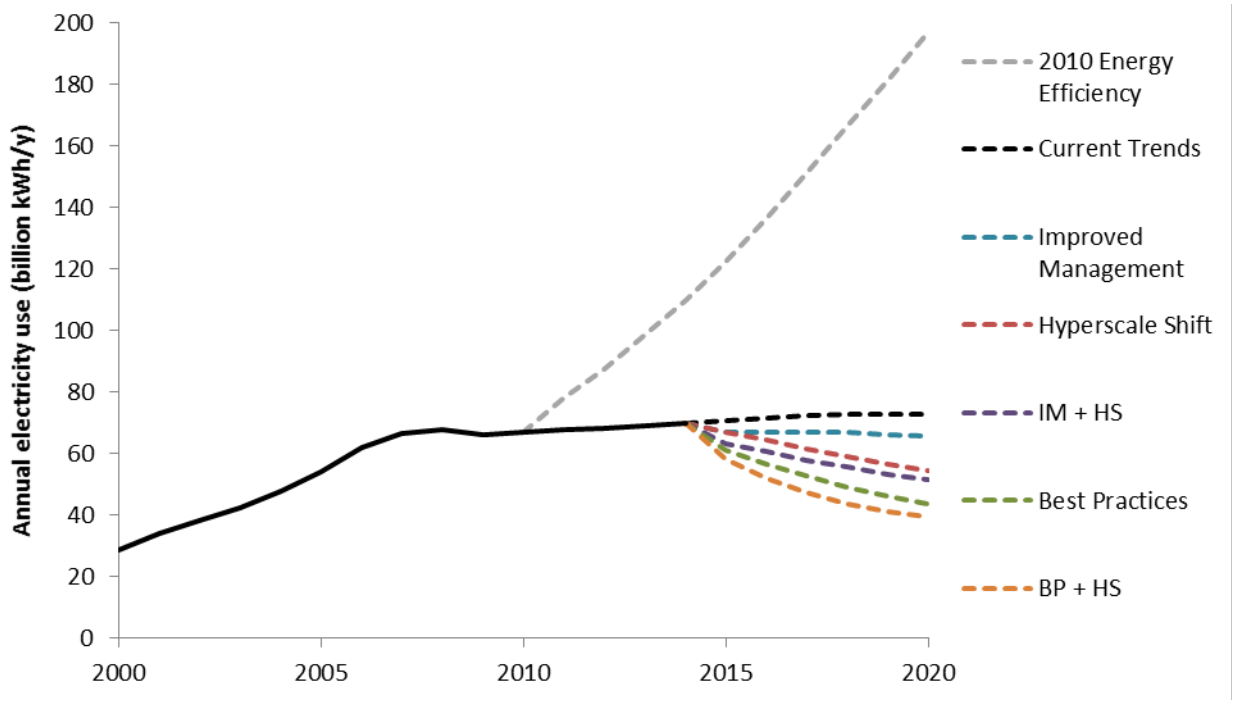



**Figure ES-1 Projected Data Center Total Electricity Use**

*Estimates include energy used for servers, storage, network equipment, and infrastructure in all U.S. data centers. The solid line represents historical estimates from 2000-2014 and the dashed lines represent five projection scenarios through 2020; Current Trends, Improved Management (IM), Best Practices (BP), Hyperscale Shift (HS), and the static 2010 Energy Efficiency.*

Note that this scenario does not halt the technological advancements of the computing industry in terms of performance, and therefore metrics such as computational performance (i.e., computations/second per server), the electrical efficiency of computations (i.e., computations per kWh), storage capacity (i.e., TB per drive), and port speeds (i.e., Gb per port) are all assumed to progress as normal. See Section 2.3.5 in the main body of this report for more details regarding the assumptions in this counterfactual scenario.

Along with the considerable energy efficiency resource already achieved, there are additional energy efficiency strategies and technologies that could significantly reduce data center electricity use below the approximately 73 billion kWh demand projected in 2020. Many of these efficiency strategies are already successfully employed in some data centers while others are emerging technologies that will be commercially available in the near future. Recently observed efficiency trends are incorporated into a "current trends" scenario. The potential impact from a more aggressive adoption of the energy efficiency strategies is explored through additional projections that apply a combination of the three following efficiency scenarios:

- The "improved management" scenario includes energy-efficiency improvements beyond current trends that are either operational or technological changes that require minimal capital investment. This scenario represents a focus on improving the least efficient components of the data center stock by employing practices already commonly used in data centers.

- The "best practices" scenario represents the efficiency gains that can be obtained through the widespread adoption the most efficient technologies and best management practices applicable to each data center type. This scenario focuses on maximizing the efficiency of each type of data center facility.

- The "hyperscale shift" scenario represents an aggressive shift of data center activity from smaller data centers to larger data centers. While the current trend scenario already incorporates some movement towards more server use in large data centers, this scenario assumes the majority of servers in the remaining small data centers are also relocated.

In addition to applying each of these scenarios independently, two additional scenarios demonstrate the combination of a "hyperscale shift" scenario in conjunction with either the "improved operation" or "best practices" scenario. Figure ES-1 shows that these five scenarios yield an annual saving in 2020 up to 33 billion kWh, representing a 45% reduction in electricity demand when compared to current efficiency trends.

# 1 Introduction

Data centers primarily contain electronic equipment used for data processing (servers), data storage (storage equipment), and communications (network equipment). 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 towards increased digital information management, data centers have become ubiquitous – they are found in nearly every sector of the economy – and are essential to the function of communication, business, academic, and governmental systems. All but the smallest companies have some kind of data center needs, and larger companies often have tens, or even hundreds, of data centers. Smaller data centers are commonly located within large commercial buildings, while larger data centers tend to be buildings constructed specifically for their use that can be up to several hundred thousand square feet in size. Universities, municipalities, and government institutions also use and operate data centers for information management and communication functions.

The energy used by the nation’s servers and data centers is significant. In a 2007 Report to Congress,[1] the data center sector was estimated to have 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 (2006 dollars). This estimated level of electricity consumption is similar to the amount of electricity consumed by approximately 5.8 million average U.S. households. The electricity use of the nation’s servers and data centers in 2006 was more than double the electricity that was estimated to have been consumed for this purpose in 2000. The accompanying methodology article[2] to the 2007 Report, published in 2011 with updated methods and inputs, estimated 2008 data center electricity demand to be 69 billion kWh, 1.8% of total U.S. electricity sales. Another study published in 2011[3] revealed that electricity use by U.S. data centers in 2010 constituted about 2% of overall electricity use in that year, and that the rate of growth in energy use slowed significantly in the period 2005-2010, compared to 2000-2005. This trend was likely a result of the 2008-09 economic crisis and the increased adoption of virtualization and other energy efficiency practices in the data center industry.[3] This report provides updated estimates of current data center energy use, updated historical estimates of energy use back to the year 2000, and projections for energy use through 2020.

## 1.1 Report Scope

This report builds on previous modeling efforts and updates key inputs to the data center model used in the 2007 Environmental Protection Agency Report to Congress on Server and Data Center Efficiency, Public law 109-4311.[1] The scope of this report includes updates to the following sections from the 2007 study:

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

- Potential Energy and Cost Savings through Improved Energy Efficiency

Additionally, this report provides the following insights as outlined in the North American Energy Security and Infrastructure Act of 2015:[4]

1. A comparison and gap analysis of the estimates and projections contained in the original report with new data regarding the period from 2008 through 2015;
2. An analysis considering the impact of information technologies, including virtualization and cloud computing, in the public and private sectors;
3. An evaluation of the impact of the combination of cloud platforms, mobile devices, social media, and big data on data center energy usage;
4. An evaluation of water usage in data centers and recommendations for reductions in such water usage; and
5. Updated projections and recommendations for best practices through fiscal year 2020.

## 1.2 Report Organization

This report serves as an update to the 2007 Environmental Protection Agency Report to Congress on Server and Data Center Efficiency, Public law 109-4311[1] (henceforth referred in this study as the 2007 Report). The 2007 Report provides detailed background information such as data center equipment layout and cooling configurations, which are therefore not discussed here. Rather, this report focuses on new trends and data available since the 2007 Report, methodology of the current estimate, and details of the new results and findings. Chapter 2 of this report describes the methodology used for the current energy and water use estimates. Although the current methodology follows the framework presented in the 2007 Report, recent trends such as the proliferation of “unbranded” servers and storage equipment, also referred to as “self-assembled,” “whitebox,” or original design manufacturer (ODM) servers require additional data input and characterization. Unbranded servers are associated with the increasing number of very large “hyperscale” data centers and often bypass large, branded server vendors (e.g. Hewlett-Packard, Dell) by being bought directly from the manufacturing companies that build servers (e.g., Quanta, Wistron, and Foxcom). See Chapter 2 for a more detailed description of unbranded servers. Chapter 3 of the report is devoted to the discussion of servers and data centers owned and operated by the federal government. Chapter 4 describes a number of key trends observed in the data center industry, such as server consolidation, colocation, and the shift to cloud-based platforms. These trends, along with energy efficiency strategies, are used to generate alternative energy use projection scenarios out to the year 2020. While rapid growth in the Internet industry has resulted in significant direct energy use in data centers, it can also indirectly impact certain resources in other sectors and possibly promote other efficiencies in society that were previously not achievable. These observations are discussed in Chapter 5. Finally, Chapter 6 provides suggestions for future research to address challenges that have been identified but not addressed in this report, which could improve future energy use estimates in such a rapidly evolving industry with little publically available data.

# 2 Estimates of U.S. Data Center Energy Use

## 2.1 Methodology Overview

The 2007 Report provided a series of data center energy use projections for 2007-2011 developed for five different efficiency scenarios. Detailed assumptions used for those efficiency measures in each scenario can be found in Brown et al. (2007),[1] while the modeling algorithms developed for these energy use projections are presented in Masanet et al. (2011).[2] Additional studies were conducted by Koomey[3 5 6] to assess the growth in data center electricity use from 2000 to 2010 for the U.S. and the world, using methods similar to the 2007 Report. The 2010 data center electricity use estimated by Koomey (2011) most closely aligns with the "improved operation scenario" projection in the 2007 Report.

The current study differs from previous work in its categorization of IT equipment and types of data centers. The number of data center space types has been expanded from the 2007 Report to account for hyperscale data centers, which are large warehouse-sized data centers that have emerged with the growth in cloud platforms, mobile devices, social media, and big data. Previously considered size-based space types are now disaggregated into two categories: "internal" data centers and "service provider" data centers. Internal data centers represent traditional facilities that support businesses and institutions while service provider data centers account for the more specialized facilities that provide the core services of businesses, such as communication and social media companies. New server classifications are added to distinguish between servers that only contain one processor socket ("1S") and those that contain 2 or more ("2S+"). Additionally, the new server classifications distinguish between servers that are branded and sold by global computer companies (e.g., Dell, Hewlett-Packard, IBM), and servers that bypass these brand vendors in the supply chain and are bought directly from the manufacturing companies that build servers (e.g., Quanta, Wistron, and Foxconn), often by large Internet companies and typically for use in hyperscale data centers, as shown in Figure 1. The latter is defined in this report as "unbranded" servers and the former as "branded" servers. These categories are associated with different usage and data center type placement assumptions, as described in Section 2.2. The current study also includes updates to previous storage power consumption estimates, which are now derived from terabyte shipment data and disaggregated into hard disk drive (HDD) and solid-state drive (SSD) categories. Network consumption is now estimated through a combination of switch port shipment data of various speeds and port power consumption from publically available data and published literature. The equipment types (servers, storage, and networking) modeled in this study are shown in Figure 2.

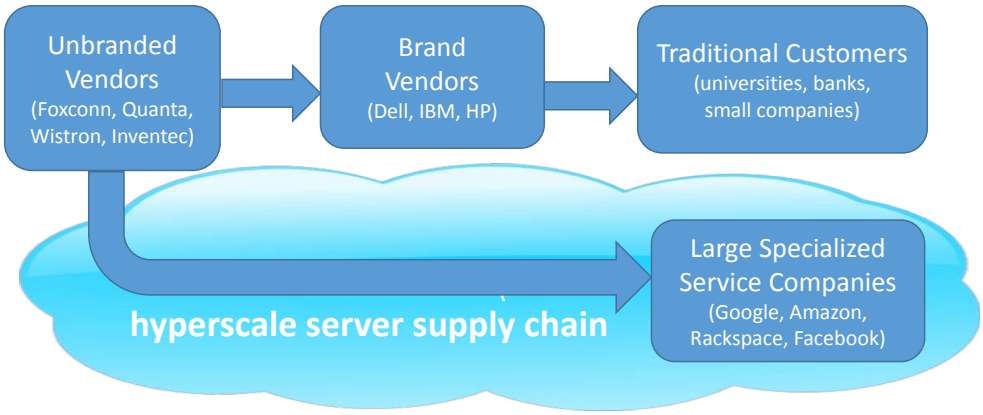



**Figure 1. Volume Server Supply Chain**

**Similar to previous U.S. data center energy estimates,[1] [2] [3] [4] [5] this study uses data provided by the market research firm International Data Corporation (IDC) to derive numbers of data center servers, as well as storage and network equipment, installed in the United States. Power draw assumptions are then applied to the estimated installed base of equipment to determine overall IT equipment energy consumption. IT equipment is disaggregated across various data center space types, each of which has an associated power use effectiveness (PUE) value. The PUE, when multiplied by equipment power consumption, gives an estimate of the total power needed to run the data center, including the data center infrastructure (i.e. cooling, lighting, controls).** Figure 3 shows a schematic of this modeling approach, and Section 2.2 discusses data sources in more detail.

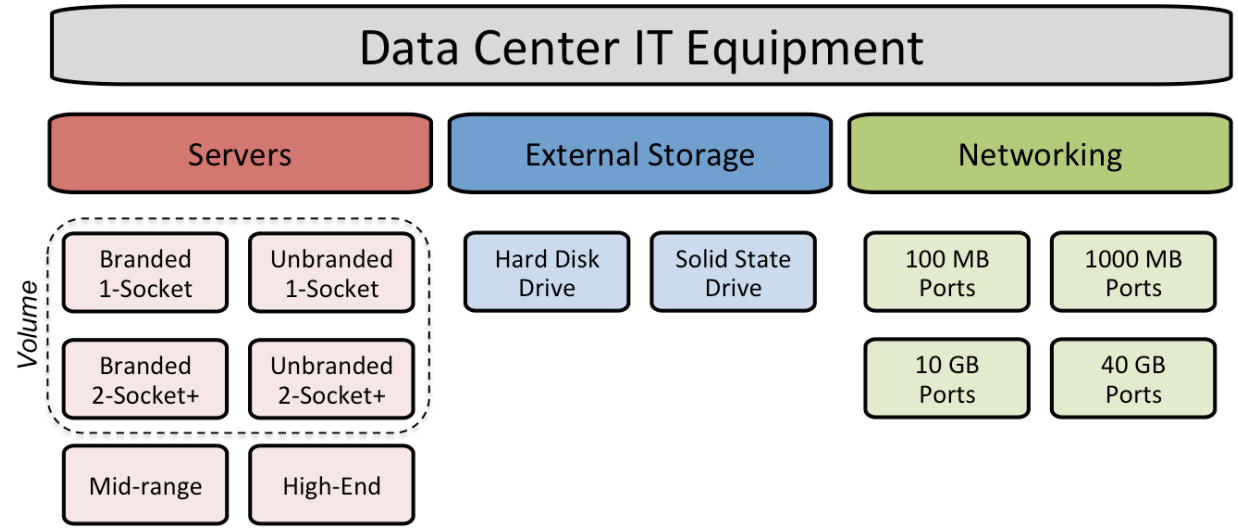



**Figure 2. Equipment Types Modeled in Energy Estimation**

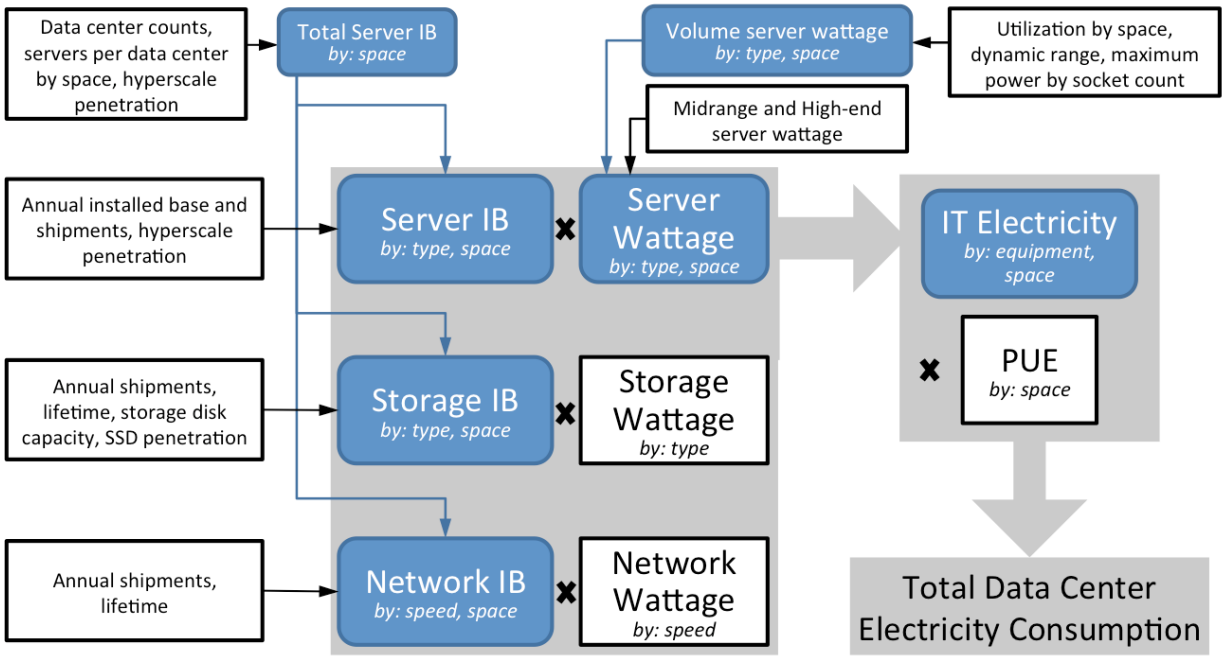



**Figure 3. Schematic of Modeling Approach**

## 2.2 Data

### *2.2.1 IDC Worldwide Trackers[7]*

IDC's Worldwide Quarterly Server, Storage, and Network trackers were used as the basis for many equipment estimates in this study. Data was obtained in 2014 Q4, and therefore all values beyond 2014 are considered forecasts. These trackers include historical and forecasted shipments for each equipment type, as well as installed base estimates for servers.[8] [9] [10] [11] This data is referred throughout the report as "IDC data"

### *2.2.2 SPEC Database[15]*

The SPEC Power benchmark suite, created by the Standard Performance Evaluation Corporation (SPEC), measures power and performance of servers. SPECpower_ssj2008 is an industry-standard benchmark application that has been used since 2007, with users self-submitting results to a database that is reviewed and released to the public quarterly. Data from 2007 to 2015 Q4 was used in this study, and will be referred to as "the SPEC database".

### *2.2.3 SERT[12] Database*

The Server Efficiency Rating Tool (SERT) was created by SPEC for the U.S. Environmental Protection Agency's (EPA) ENERGY STAR program. This tool uses a set of synthetic worklets to test discrete system components, providing detailed power consumption data at different load levels. Data from this tool is submitted to the EPA by manufacturers, and is collected and maintained by the Information Technology Industry Council (ITI). Data collected by ITI through March 2016 was used in this report, and will be referred to as "the SERT database".

## 2.3 IT and Infrastructure Equipment Estimates

### *2.3.1 Server Energy Use*

*Server Installed Base*

Estimates for the U.S. installed base of servers are based on IDC's "Worldwide Quarterly Server Tracker – Installed Base",[11] which contains annual data for historical (2006-2014) and forecasted (2014-2018) server shipments and installed base of servers. These numbers are provided for each of the three primary server classes (volume, midrange, and high-end) as well as by processor count (1S or 2S+). The three server classes are an IDC taxonomy based on average sales value, with volume, midrange, and high-end servers representing < $25,000, $25,000-$250,000, and >$250,000, respectively. Volume servers typically contain x86 processors and represent the overwhelming majority of servers in data centers. High-end servers are often large RISC-based systems that operate on Unix and include many high-performance computing supercomputers. Midrange servers can contain aspects of both volume and high-end servers. The installed base estimates were extended from 2018 to 2020 using the compound annual growth rate (CAGR) of the installed base from 2014-2018, the years for which IDC has forecasted installed base growth.

Each of the 1S and 2S+ categories for volume servers is further disaggregated into "branded" and "unbranded" classes using additional data provided by IDC and assumptions about future penetration of hyperscale data centers. First, the percent of all servers located in hyperscale data centers from 2000-2020 is estimated. (See discussion in Section 2.2.4.) Then, a sigmoid curve shape (i.e., "S-shape") is applied to estimate the penetration of unbranded servers within hyperscale data centers over time. Resulting unbranded installed base estimates for 2000-2020 are calibrated to (1) estimates for 2008-2010, which were provided through communications with IDC[13], and (2) IDC historical server shipment data for 2010-2014, which included disaggregation by vendor type.[8] The assumed hyperscale data center server capacity, sigmoid curve, data-based estimates, and resulting installed base are shown in Figure 4.

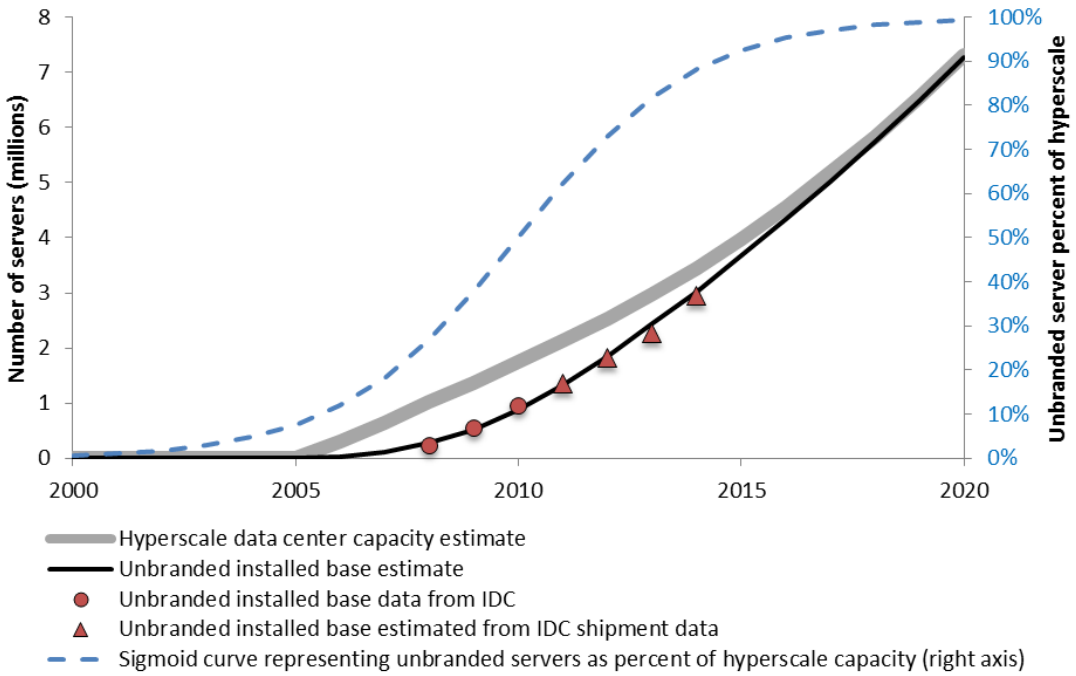



**Figure 4. Unbranded Server Installed Base and Underlying Assumptions**

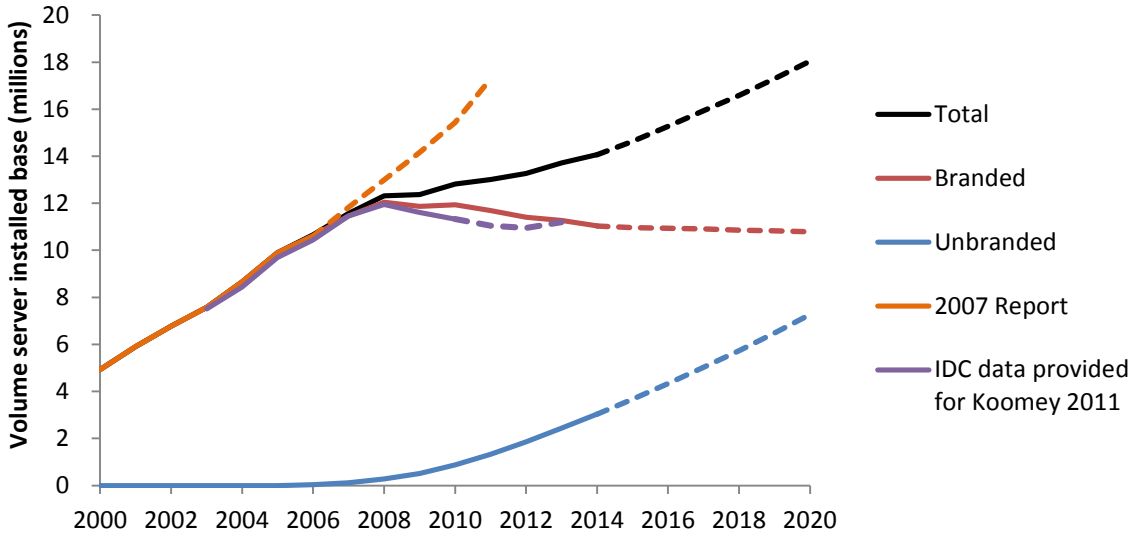



Figure 5 shows the current and projected total server count for 2000-2020 estimated in this study, as well as server count estimates from the 2007 Report and from Koomey (2011). The 2007 Report projected much higher server counts than other estimates, with growth from 2007-2011 following the same trend as 2000-2007, due to these estimates being made prior to the 2008 financial crisis that greatly affected sales. Data from Koomey (2011) is similar to the current study's branded server estimates, with server installed base leveling off then slightly decreasing after 2007. Similarity between these two estimates is expected, as unbranded servers were not being explicitly tracked by IDC at the time of the Koomey (2011) study. Figure

6 shows the volume server installed base estimates from the current study disaggregated by processor count (1S or 2S+) and vendor type (branded or unbranded).

Server installed base and shipment data provided by IDC was also used to determine the approximate lifetime of servers. Observation of data showed that for any given year, the number of servers in the installed base was more than the sum of the previous 4 years' shipments, but less than the previous 5. The exact portion of the 5th year's shipments that were still in the installed base was found using Equation 1. For 2006-2020, this portion averaged 0.4, with very little deviation, indicating an approximate server lifetime of 4.4 years. This value was later used during storage and network installed base calculations based on a report from The Green Grid, which estimates that servers, storage, and network equipment all have a lifetime of 3-5 years.[14]

**Equation 1**

$$f = \frac{IB_y - \sum_{i=y-3}^{y} S_i}{S_{y-4}}$$

Where f = fraction of 5th year shipments in installed base
IBy = installed base in year y
Sy = shipments in year y

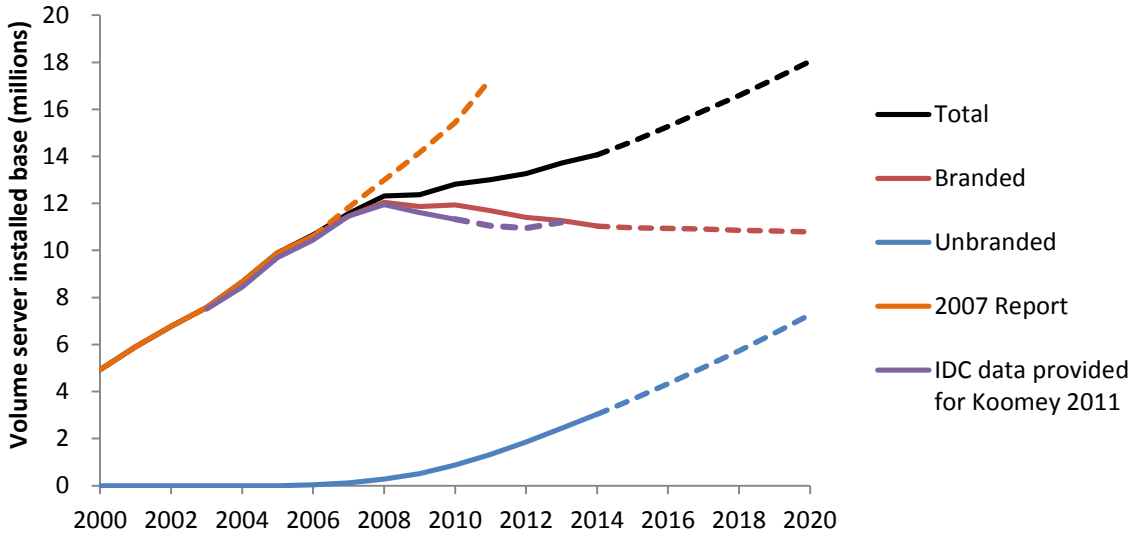



**Figure 5. Total Volume Server Installed Base Estimates from Three Studies**

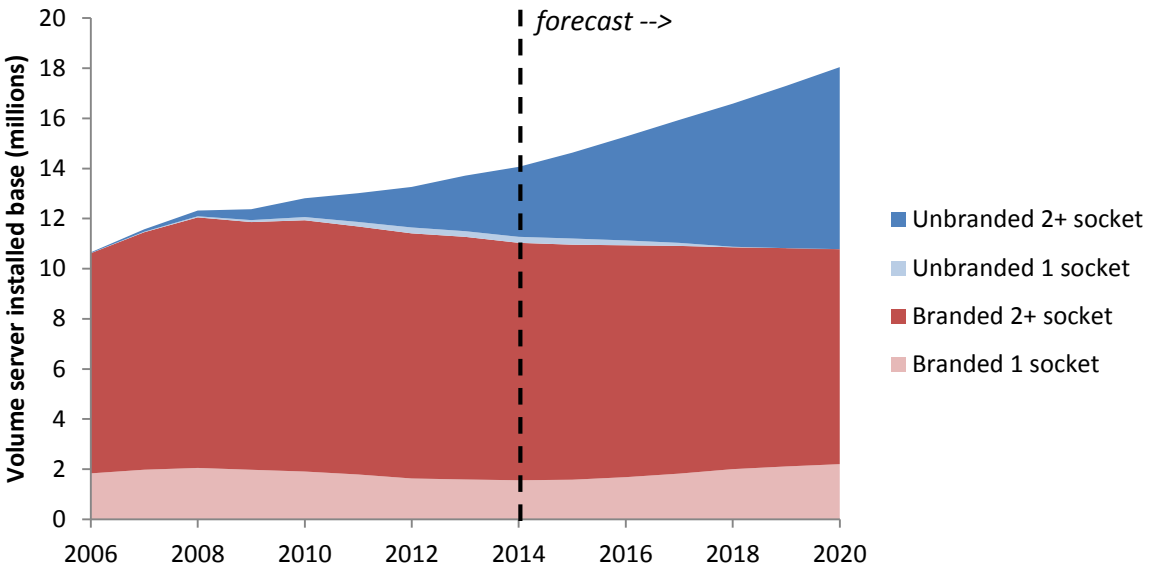



**Figure 6. Volume Server Installed Base 2000-2020**

*Server Power Draw*

Average power use of volume servers was calculated for each year using assumptions for: (1) maximum power consumption (2) average server utilization and (3) average ability for servers to scale power use with utilization. The maximum power represents the wattage of the server when being used at 100% utilization. Average server utilization represents the percent of computing ability used, on average, over the entire year. Average server scaling ability is the extent that servers, across the entire installed base, are able to use less-than-maximum power at non-maximum utilizations.

Maximum power in 2013 for 1S and 2S+ volume servers was estimated from the SERT database to be 118 W and 365 W respectively.[12] These values result in an overall volume server average maximum wattage of ~330 watts, which is the same assumption used in the later years of the 2007 Report. Maximum wattage for 2000-2007 was calculated for 1S and 2S+ servers by assuming that the estimates used in the 2007 Report represented a weighted average of 1S and 2S+ servers and that the ratio between the maximum power of the two types was the same as the 2013 ratio. Maximum power was held constant from 2007-2020 due to two observations: (1) the calculated 2007 values for 1S and 2S+ servers were very close to the SERT (2013) values, and (2) information in the SPEC database[15] shows that maximum server power has remained approximately constant from 2007-2015. While the wattages reported in the SPEC database were not used directly due to the assumed self-selection bias towards high efficiency servers in the database, the general temporal trends are assumed to be representative of all servers. The assumption that the average maximum power remains constant after 2005 is consistent with assumptions made by Heddgren et al.[16] (power draw based on 50% utilization) and the lower bound assumption by Koomey.[3] No difference in maximum power is assumed between branded and unbranded servers. Power consumption for

mid-range and high-end servers is estimated at the overall average level, with utilization and scaling assumptions incorporated. These average wattage values for 2000-2007 were taken from the 2007 Report and extended to 2020 using the 2000-2007 CAGR of approximately 7%, which is consistent with projections through 2020 provided during industry review. Results are shown in Figure 7. These wattages assume underlying growth in both utilization levels and power scaling ability.

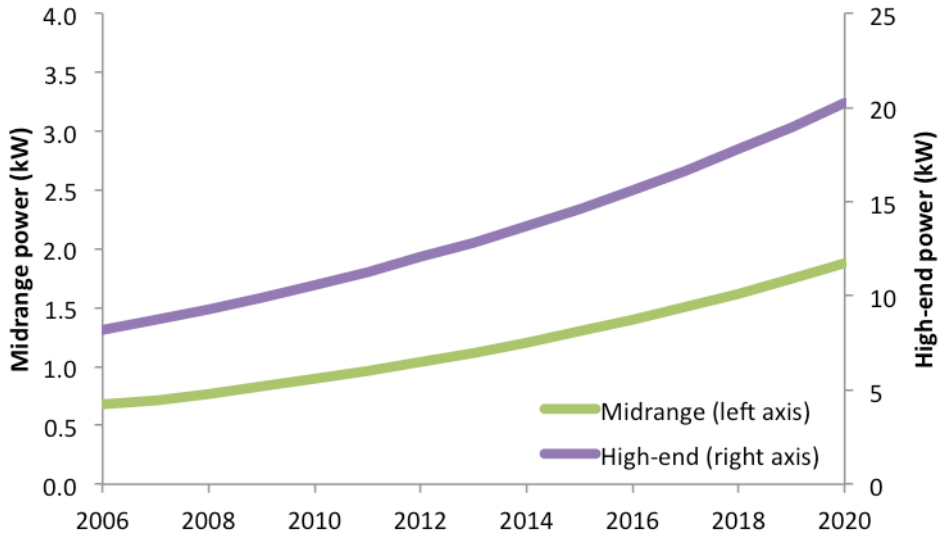



**Figure 7. Average Power Draw Assumptions for Mid-Range and High-End Servers**

Average utilization for volume servers is assumed to vary by space type, as shown in Table 1, and is assumed to be constant from 2000-2010. A steady increase from 2010 to 2020 is assumed, to account the prevalence of virtualization opportunities in data centers. Service provider data centers are assumed to run at higher utilizations than internal data centers, as the servers in service provider data centers are often configured for more specialized and predictable operations. Hyperscale data centers are assumed to run at higher utilizations than other service provider and internal data centers based on server utilization estimates in cloud and non-cloud data centers.[1] [17] [18] The values shown in Table 1 represent the average of active servers, and therefore the inclusion of inactive servers (assumed to be 10% of internal and 5% of service provider and hyperscale data centers) slightly lowers the overall averages. (See discussion in Section 4.2 for more information on inactive servers.)

**Table 1. Average Active Volume Server Utilization Assumptions**

| Space Type | 2000-2010 | 2020 |
|---|---|---|
| Internal | 10% | 15% |
| Service Provider | 20% | 25% |
| Hyperscale | 45% | 50% |

The amount of power consumed at average utilization is dependent on how closely servers come to achieving power-proportionality, where power consumption scales directly with utilization. In other words, perfect power-proportionality would mean that a server would only use 10% of maximum power when run at 10% utilization. One metric used to quantify this behavior is the dynamic range (DR), which is the ratio between the lowest power level (idle power) and the maximum power. Minimal documentation exists on the average DR of the U.S. data center servers stock. The DR varies among different server types and is influenced by hardware properties, power management software, and settings for server-specific operations, all of which continue to change. This study estimates the DR of the installed base by first bounding possible values by a "Maximum DR" trend, which represents the lowest-performing servers (i.e. those that use a large amount of power at idle) and "Minimum DR" trend, which represents the highest-performing servers. The assumed Maximum DR is 0.67 in 2007 (the assumption used in the 2007 Report) and slopes linearly to 0.44 in 2020, as shown in Figure 8. The SERT database and recent publications[15] [18] show servers performing with an average dynamic range of ~0.44, and therefore this study assumes that, at a minimum, *all* volume servers will achieve a DR of 0.44 by 2020. The Minimum DR trend is derived from the SPEC database values from 2007-2015, which are shown in Figure 9, as this database is generally understood to represent well-performing servers due to self-selection bias in the server entries. The SPEC dynamic range trend is modeled as an exponential equation that asymptotically approaches a DR of 0.1, resulting in the Minimum DR trend shown in Figure 8.

The Maximum and Minimum dynamic range trends provide reasonable bounds of volume server DR, but an assumption must be made as to where the average DR lies between these bounds. This study estimates the annual installed base average as a 90/10 mix between the Maximum and Minimum trends for the Current Trends scenario, resulting in the effective average scaling trend shown in Figure 8. The DR is assumed to be constant across vendor types (branded and unbranded) and processor counts (1S and 2S+).

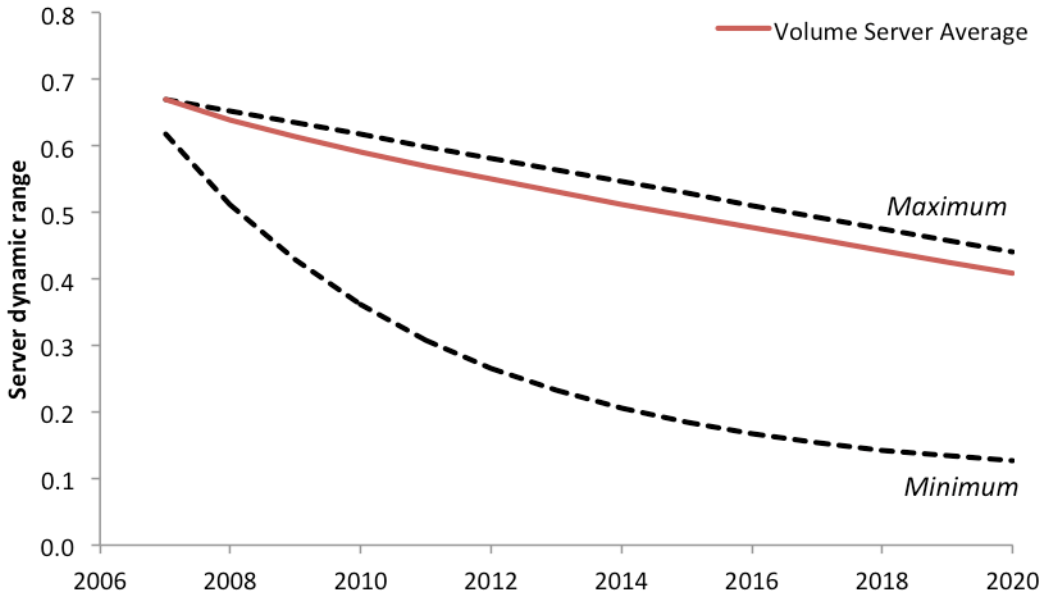


**Figure 8. Assumed Dynamic Range of Volume Servers**

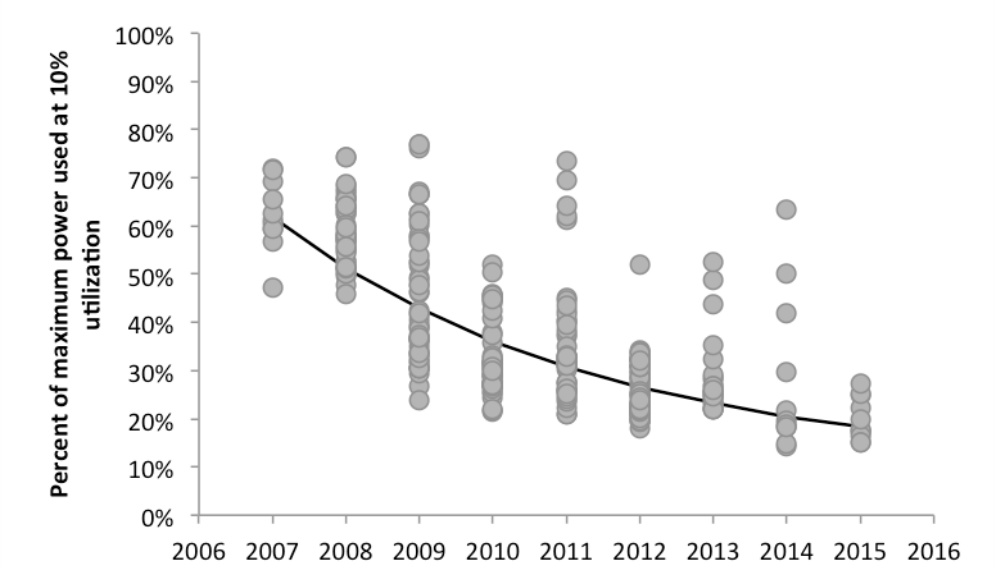



**Figure 9. Dynamic Range of 1- and 2-Socket Servers in SPEC Database**

**Dynamic range and maximum wattage values are used to calculate the slope of the utilization versus power consumption curve for volume servers in each year, as shown in Equation 2. This slope is then used with utilization assumptions to calculate the average wattage of volume servers in each space type in each year. This calculation is shown in Equation 3, and results are shown in Figure 10. These average wattage values, along with installed base estimates, provide the total annual server energy consumption estimates shown in**
Figure 11.

**Equation 2**

$$m = \mathrm{P_{max}} * (1 - DR)$$

Where m = slope of utilization vs. power line
$P_{max}$ = maximum server wattage
DR = dynamic range of server (fraction of max power used at idle)

**Equation 3**

$$P_{avg} = P_{max} - (m * u_{avg})$$

Where $P_{avg}$ = average server wattage
m = slope of utilization vs. power line
$u_{avg}$ = average server utilization

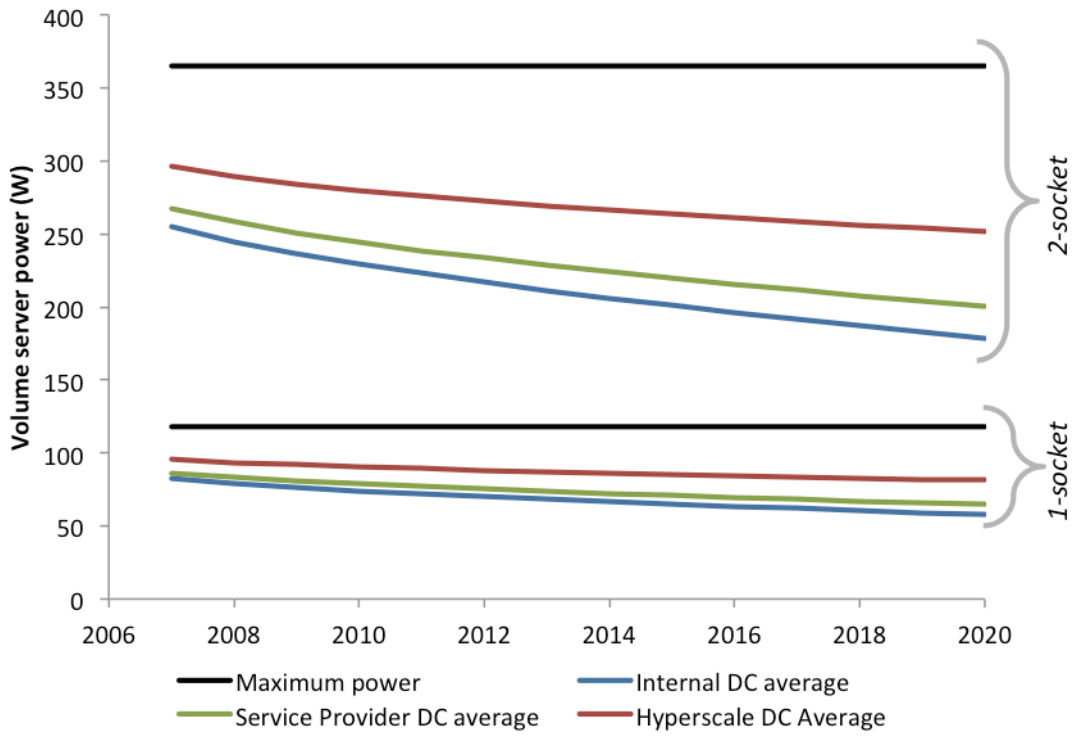



**Figure 10. Maximum and Effective Average Power Estimates for Volume Servers**
*The solid black line represents maximum power of each server size. Servers in hyperscale data centers use more power than those in internal or service provider due to higher utilizations.*

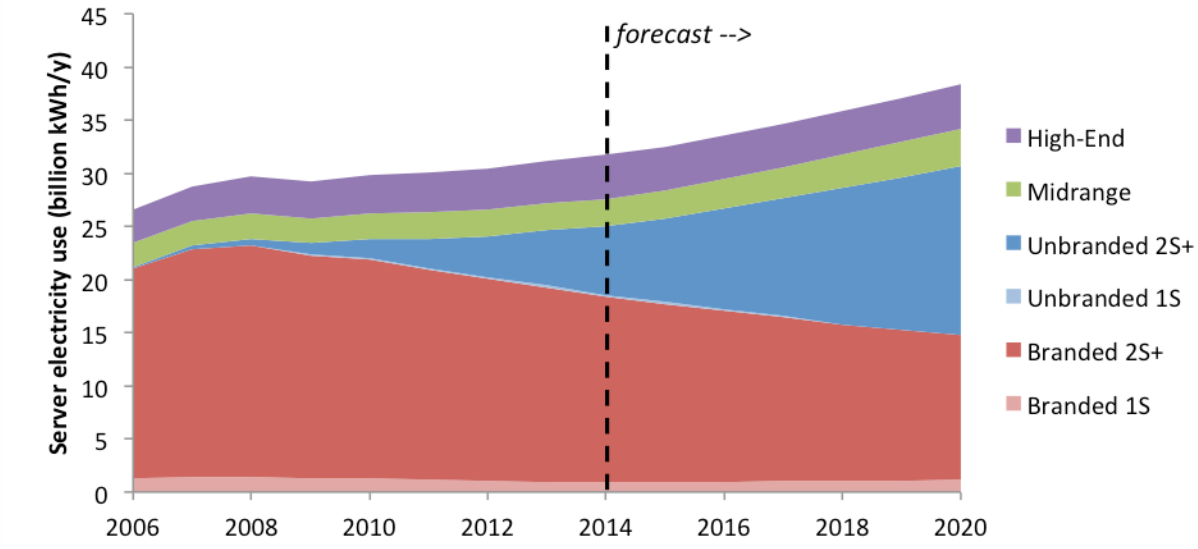



**Figure 11. Total U.S. Annual Direct Server Electricity Consumption by Server Class**

### *2.3.2 Storage Energy Use*

The installed base of data storage equipment was calculated in terabyte (TB) units using IDC historical (2010-2014) and forecast (2015-2019) shipment data. This data includes external storage units from both branded and unbranded vendors as well as internal storage on servers with more than two storage drives. This study estimates the energy consumed by these supplemental storage types collectively, and refers to all types included in the IDC tracking data as either "Data Center Storage" or "External Storage". The provided shipment data is extended back to 2000 and forward to 2020 using the compound annual growth rate (CAGR) from 2010-2019. Then, installed base is calculated assuming an average storage life of 4.4 years, equal to the estimated lifetime of servers (see Section 2.2.1).[14] This installed base is then disaggregated into hard disk drive (HDD) and solid-state drive (SSD) storage categories using a 2015 ASHRAE report that estimated that in SSD accounted for 8% of non-tape storage in 2012, and would grow to 22% by 2017.[20] The final installed base estimate in TB is shown in Figure 12.

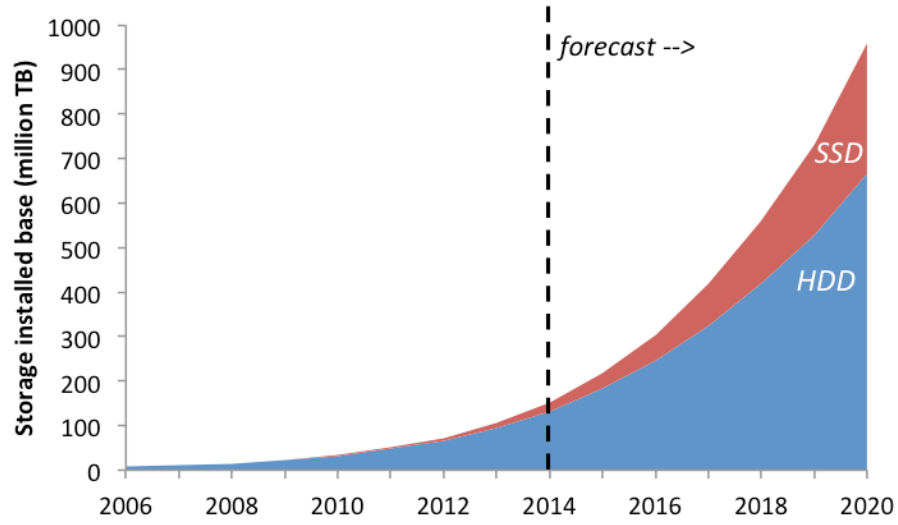



**Figure 12. Total U.S. Data Center Storage Installed Base in Capacity (TB)**

Power consumption of HDDs is relatively fixed on a per-disk level, and is generally not dependent on the capacity of the disk.[3] [19] For SSD units, power consumption is more closely related to capacity (more specifically, to read/write frequency), but is still often reported on a per-disk level. Therefore, it is useful to convert installed base estimates of both HDD and SSD storage from capacity (TB) to number of drive units. To do so, the trend of TB/drive over time for each storage type is first estimated. For 2000-2007, the TB/drive metric for HDD is calculated from (1) this study's installed base estimate in TB, for external branded drives only, and (2) the 2007 Report's disk installed base data, which was provided by storage manufacturers at the time. In 2020, HDDs are assumed to provide an average capacity of 10 TB/disk from drive capacity improvements and capacity optimization methods, based on industry feedback. Data between the 2007 and 2020 estimates are populated using an exponential trend, resulting in the trend shown in Figure 13. For SSD, an average capacity of 5 TB/drive is assumed to be

reached by 2020 based on industry feedback, and the same annual growth rate seen in HDD TB/disk (approximately 27%) is assumed for SSD for all years prior, as shown in Figure 13.

Annual TB/drive estimates are multiplied by the TB installed base estimate (Figure 12) to get the estimated storage installed base in number of drives. This estimate is shown in Figure 14, which shows growth in number of installed drives through 2018, after which point HDD count begins to decrease (SSD count continues to grow). Note that this is the *number* of installed hard disk drives, and does not indicate slowing storage demand. Decreasing drive count is due to HDD capacity per disk (TB/disk, Figure 13) growing at a faster rate than projected HDD capacity demand (TB, Figure 12). SSDs make up 47% of the installed drive base in 2020, which is within the 40-50% range estimated by industry.

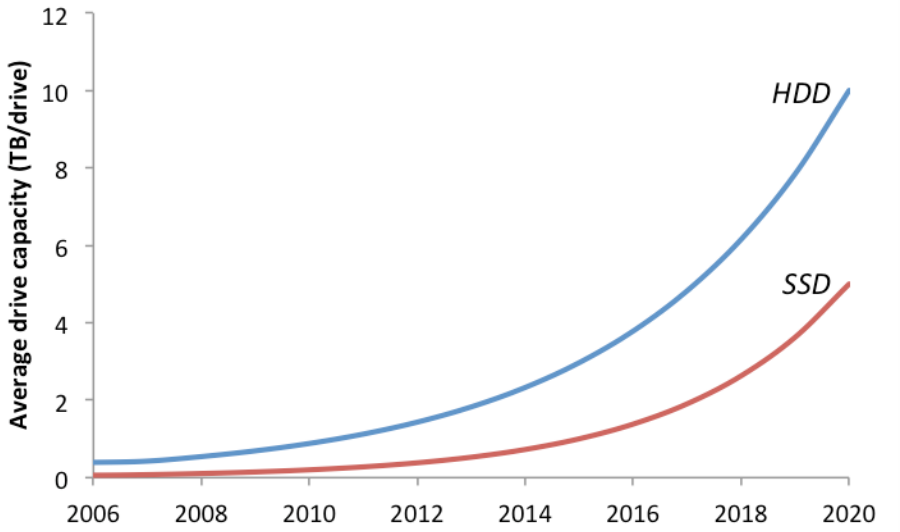



**Figure 13. Estimated Average Capacity of U.S. Data Center Storage Drives**

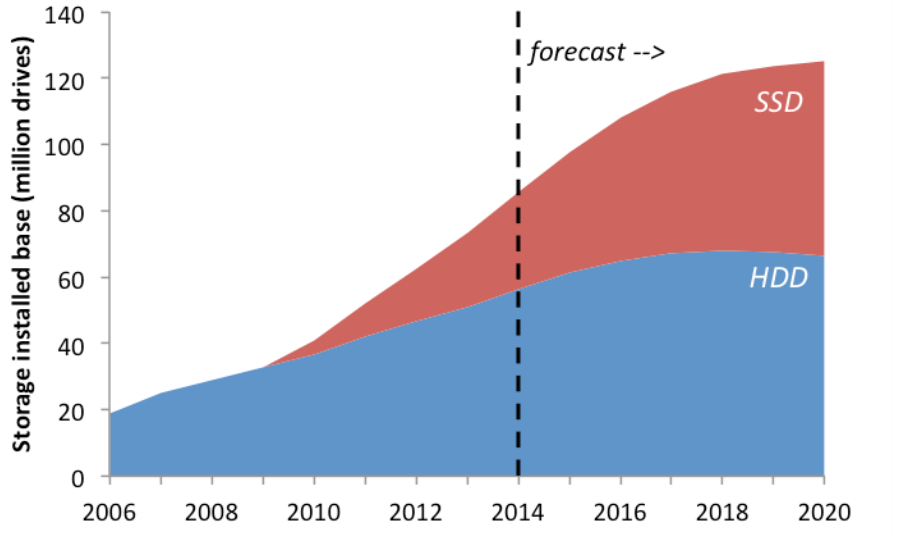



**Figure 14. Total U.S. Data Center Storage Installed Base in Drive Count**

Power consumption for HDDs was assumed to be 14 W/disk from 2000-2006 as estimated by Seagate for the 2007 Report. A 2015 ASHRAE report[20] stated minimum, average, and maximum power usage for various speeds and sizes of HDDs, while a NYSERDA report[21] estimated the breakdown of these various drive types in the installed base. Using these two reports, the average wattage of HDDs is estimated to be 8.6 W/disk in 2015. This represents a 5% annual decrease in disk wattage from 2006-2015, which is assumed to continue through 2020, resulting in 6.5 W/disk as shown in Figure 15. This improvement in disk efficiency represents more efficient disk drive components, lower power use in idle states, and use of capacity optimization methods.

SSD wattage is assumed to be constant at 6 W/drive, as shown in Figure 15, based on both IDC[22] and ASHRAE[20] reports. Constant drive wattage paired with our TB/drive estimates aligns with industry expectations that SSD capacity per watt will increase three to four-fold by 2020.

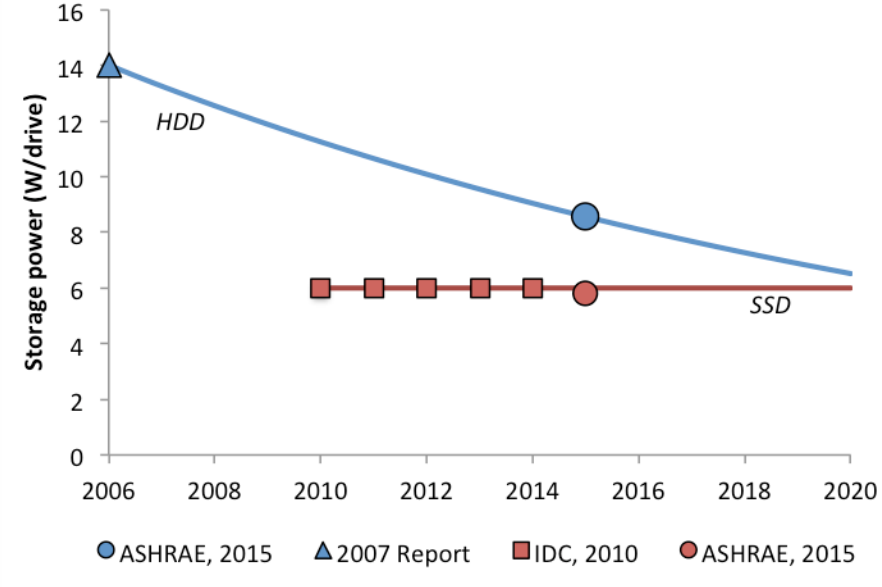



**Figure 15. Average Wattage of Storage Drives in U.S. Data Centers**

Electricity consumption for all data center storage is calculated as the product of the estimated installed base of drives and the assumed power consumption per drive. Additional operational consumption was then added to account for the controller and associated components required to operate external storage systems. This operational energy is assumed to equal 25% of the storage energy, based on industry comment, and only applies to storage that is external to servers. Resulting total storage energy consumption is calculated according to Equation 4 and results are shown in Figure 16.

**Equation 4**

$$E_y = \sum_{t=HDD,SSD} (I_{t,y} * P_{t,y}) * h_y * \left(1 + O * \frac{C_{external}}{C_{total}}\right)$$

Where $E_y$ = Storage electricity consumption in year y
$I_{t,y}$ = Installed base of storage type t in year y
$P_{t,y}$ = Per-unit power consumption of storage type t in year y
$h_y$ = Number of hours in year y
O = Operational energy as a fraction of storage energy
$C_{external}$ = Capacity of the external storage installed base
$C_{total}$ = Capacity of the total storage installed base

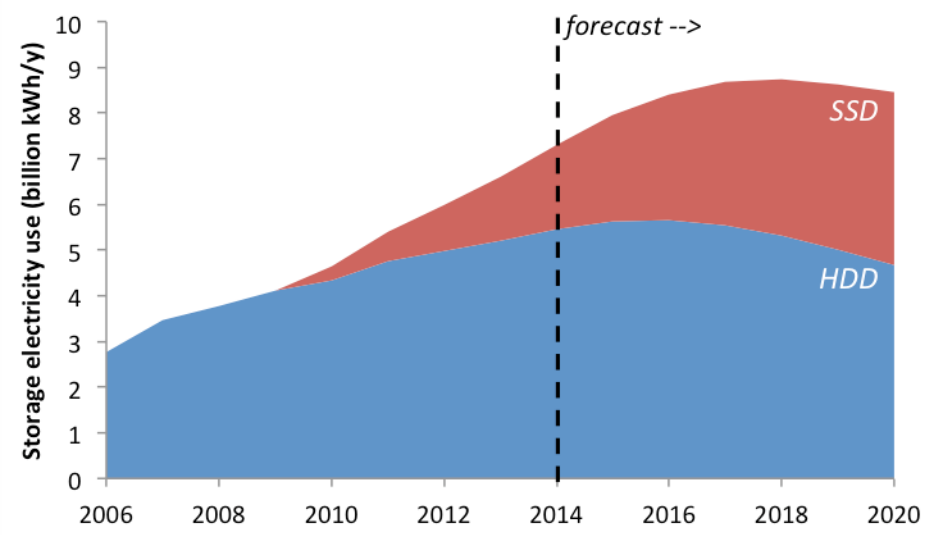



**Figure 16. Total U.S. Data Center Storage Electricity Consumption**

### *2.3.3 Network Energy Use*

Previous reports estimated network energy use either as a percentage of server energy use[1] [16] or as a product of number of servers, ports per server, and watts per port.[2] This study expands upon these previous efforts by using port installed base estimates disaggregated by port speed along with port wattage estimates for each speed, as described below.

Network equipment data from IDC includes historical (2008-2014) and forecasted (2015-2019) shipments of network ports disaggregated into four port speed categories: 100 MB, 1000 MB, 10 GB, and 40 GB. Similar to storage installed base calculations, this shipment data is extrapolated to estimate number of shipments for 2000-2007 and 2020, then used to calculate the installed base of ports by assuming an average network equipment life of 4.4 years, as estimated from server data (see Section 2.2.1).[14] Resulting port installed base estimates are shown in Figure 17.

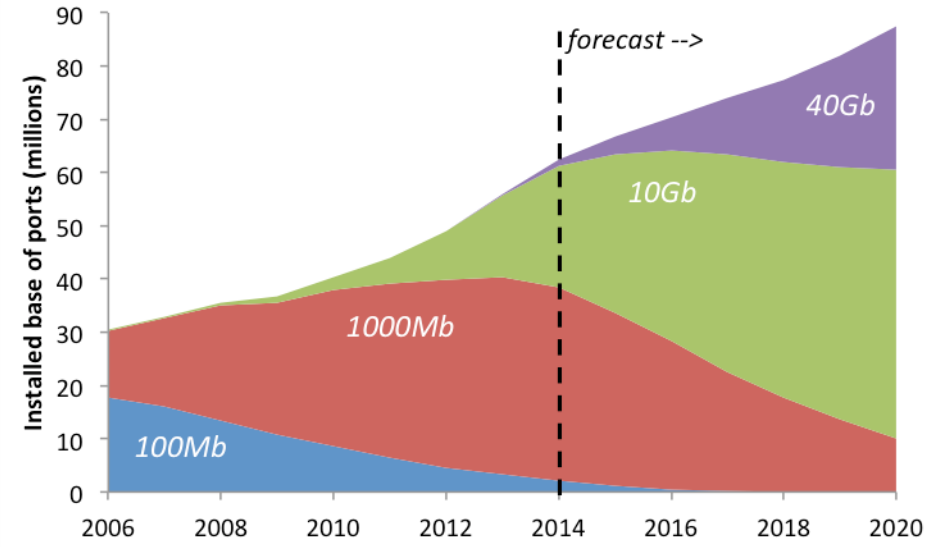



**Figure 17. Total U.S. Data Center Installed Base of Network Ports**

The power draw of network equipment was estimated on a per-port basis using different values for each port speed. Per-port wattage trends were estimated based on: the 2007 Report, which estimated an average of 8 watts (W) across the installed base; a 2012 empirical study on network energy in small to medium sized data centers,[23] which reported estimates of 1.4, 2.3, and 3.6 W for 100 MB, 1000 MB, and 10 GB ports, respectively; a survey conducted as part of this study of 51 technical specification sheets from network manufacturers, which indicated that 40 GB ports use approximately 1.7 times the power of 10 GB ports; and industry comment, which suggested 1 W/port for 1000 MB ports in 2020. These values were used to create the estimated trends shown in Figure 18. The approximate range of W/port values for various port speeds applied in this study is consistent with power estimates in previous network studies.[24] [25]

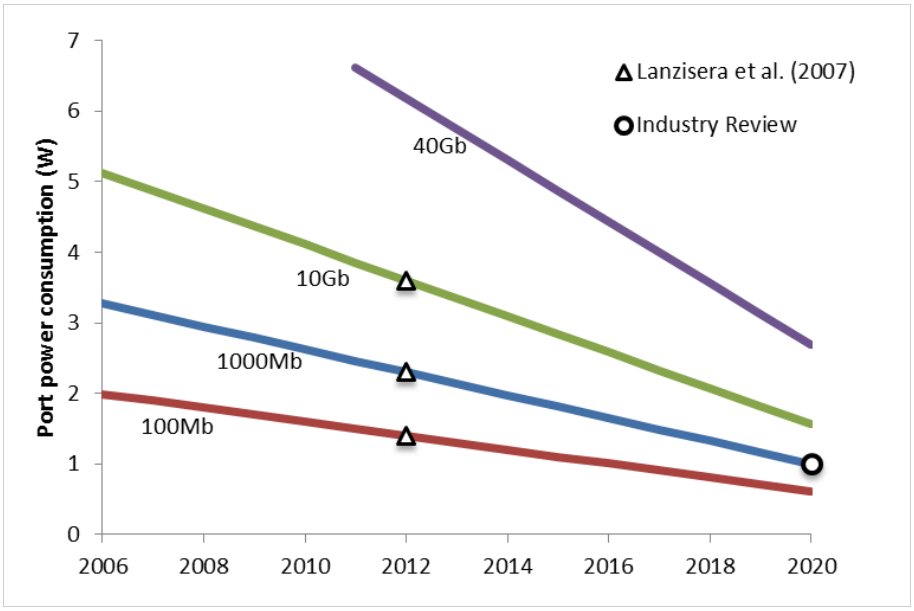



**Figure 18. Assumed Network Power for Four Port Speeds**

Total estimates of network energy consumption were calculated as the product of the number of ports in the installed base and the assumed per-port power draw for each port speed, as shown in Equation 5. Results are shown in Figure 19. While other types of network equipment besides Level 2/Level 3 network ports exist in data centers, detailed shipment and power consumption data are not available for these devices, and this equipment is assumed to contribute minimally to overall data center power consumption.

**Equation 5**

$$E_y = \sum_{s \in S} N_{s,y} * P_{s,y} * h_y$$

Where $E_y$ = Network electricity consumption in year y
S = set of port speeds: 100 MB, 1000 MB, 10 GB, 40 GB
$N_{s,y}$ = Number of installed ports of speed s in year y
$P_{s,y}$ = Power consumption of ports of speed s in year y
$h_y$ = number of hours in year y

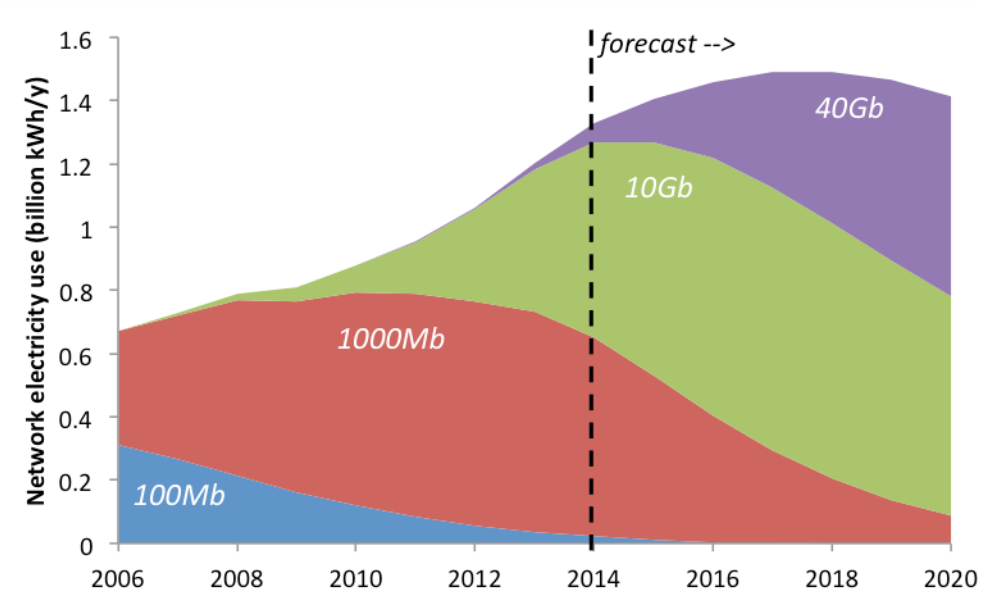



**Figure 19. Total U.S. Data Center Network Equipment Electricity Consumption**

### *2.3.4 Data Center Space Type Classifications*

Installed base estimates of servers, storage, and network equipment were disaggregated into 11 different data center building space types: five internal data center types that include server closets, server rooms, localized data centers, mid-tier data centers, and high-end data centers, as well as six service provider data center types that include server closets, server rooms, localized data centers, mid-tier data centers, high-end data centers, and hyperscale data centers. Internal data centers represent facilities operated by an organization for internal

activities, including email and productivity software, and are typically associated with finance, education, and other institutions not directly involved with providing IT services. Large internal data centers are often referred to as enterprise or corporate data centers. Service provider data centers (also called production data centers, and inclusive of colocation facilities) contain IT equipment used to provide communication services often associated with the core product of a business, such as the services provided by Google, Amazon, Facebook, and other telecommunication and social media companies. Service provider data centers include the same five size categories as internal data centers, as well hyperscale sized data centers, which are sometimes referred to as “mega” or “warehouse scale” data centers and are often associated with cloud data centers. These types of spaces are defined by IDC as shown in Table 2.

Space disaggregation is necessary to estimate energy use, as half of U.S. servers are located in server closets and server rooms[26] that can have significantly different IT equipment and infrastructure characteristics than larger data centers, and because service provider data centers associated with cloud platforms, mobile devices, social media, and big data have grown significantly in recent years.[27] In the context of this study, “infrastructure” consists of the data center equipment that is not used solely for the purpose of performing computations or the storage or transmission of data. This includes cooling systems, lighting, power supplies, security, and so on. Determining the distribution of different servers across space types allows for more accurate characterization of the total energy use associated with different server environments. It also allows for better characterization of electricity costs because most server closets, server rooms, and localized data centers are expected to be subject to commercial electricity rates, whereas larger mid-tier and enterprise-class data centers are expected to be subject to industrial electricity rates. Disaggregation into the eleven space types is a significant expansion on the 2007 study’s five space types. An overview of the methods used to allocate IT equipment to the various data center types are shown in Table 3, with detailed descriptions provided below.

The number of data centers in each space category in the U.S. for 2005 was derived from a 2006 IDC report,[26] while 2012-2018 estimates were taken directly from a 2014 IDC report.[27] The 2005 numbers only included five space categories, and therefore were split into internal and service provider categories (based on the 2012 ratios). No hyperscale data centers were assumed to exist in 2005. These 2005 numbers were then assumed to grow linearly to 2012 values. The number of data centers for 2000-2004 and 2019-2020 were estimated using the 2005-2009 (as reported by Bailey[26]) and 2012-2018 CAGRs, respectively, for each of the eleven room types.

Bailey[26] also reported average number of servers per data center for five space categories in 2005. Servers per data center for each of the five space sizes were assumed to be applicable to both internal and service provider data centers in that category. These numbers were then adjusted for each year so that, when multiplied by the number of data centers, the total number of servers was consistent with this study’s estimates of total server installed base (excluding servers assigned to hyperscale data centers). The resulting total number of servers in each space category are shown in Figure 20.

**Table 2. Typical IT Equipment and Site Infrastructure System Characteristics by Space Type**

| Space type | Typical size | Typical infrastructure system characteristics |
|---|---|---|
| Internal server closet | < 100 $ft^2$ | Often outside of central IT control (often at a remote location) that has little to no dedicated cooling. |
| Internal server room | 100-999 $ft^2$ | Usually under IT control, may have some dedicated power and cooling capabilities. |
| Localized internal datacenter | 500-1,999 $ft^2$ | Has some power and cooling redundancy to ensure constant temperature and humidity settings. |
| Midtier internal datacenter | 2,000-19,999 $ft^2$ | Superior cooling systems that are probably redundant. |
| High-end internal datacenter | > 20,000 $ft^2$ | Has advanced cooling systems and redundant power. |
| Point-of-presence server closet | < 100 $ft^2$ | At local points of presence for OSS and BSS services. Typically leverages POP power and cooling. Space is often a premium. |
| Point-of-presence server room | 100-999 $ft^2$ | Secondary computer point of presence for OSS and BSS services. Typically leverages POP power and cooling. |
| Localized service provider datacenter<br><br>Including subsegment: containerized datacenter | 500-1,999 $ft^2$ | Has some power or cooling redundancy to ensure constant temperature and humidity settings. These are typically facilities set up by VARs to provide managed services for clients. |
| Midtier service provider datacenter<br><br>Including subsegment: prefabricated datacenter | 2,000-19,999 $ft^2$ | Location for small or midsize collocation/hosting provider. Also includes regional facilities for multinational communications service providers. Has superior cooling systems that are probably redundant. |
| High-end service provider datacenter | > 20,000 $ft^2$ | Primary server location for a service provider. May be subdivided into modules for greater flexibility in expansion/refresh. Has advanced cooling systems and redundant power. |
| Hyperscale datacenter | Up to over 400,000 $ft^2$ | Primary server location for large collocation and cloud service providers. Based on modular designs, with individual modules of 50,000 sq ft on average in up to 8 modules. Employs advanced cooling systems and redundant power. |

While IDC shipment estimates include servers destined for hyperscale data centers, the percentage of servers that are located in hyperscale data centers was not available. Therefore, this study uses a simple estimation that the percent of servers housed in hyperscale data centers will grow linearly from 0% in 2005 to 40% in 2020. This assumption is a conservative

estimate of future unbranded server installed base as calculated by shipment forecasts, and aligns with other industry projections of the unbranded server market[28] and expectations that at least 40% of the data in 2020 is expected to be "touched" by the cloud at least once.[29] This estimate also aligns with current IDC estimates of the global average of servers per data center for high-end and hyperscale data centers.[30]

**Table 3. Allocation of Data Center Equipment Across Space Types**

| Step | Equipment | Allocation Method |
|---|---|---|
| 1 | Total Servers | • Set percentage (varies annually) to Hyperscale<br>• Remaining based on estimated data center counts and 2005 servers per data center estimate |
| 2 | Midrange Servers | • 5% Server Rooms<br>30% Localized and Mid-tier Data Centers<br>65% Enterprise Data Centers |
| 3 | High-End Servers | • 30% Localized and Mid-tier Data Centers<br>70% Enterprise Data Centers |
| 4 | Unbranded 1S and 2S+ Volume Servers | • 100% Hyperscale Data Centers |
| 5 | Branded 2S+ Volume Servers | • Fill remaining spots in Hyperscale |
| 6 | Branded 1S and 2S+ Volume Servers | • Fill remaining spots in all other data centers, keeping 1S and 2S+ in equal proportion |
| 7 | Storage | • None in Server Closets or Rooms<br>• Allocated to all other spaces based on total server count |
| 8 | Network Ports | • Total allocated based on total server count, with higher speeds trending towards larger data centers |

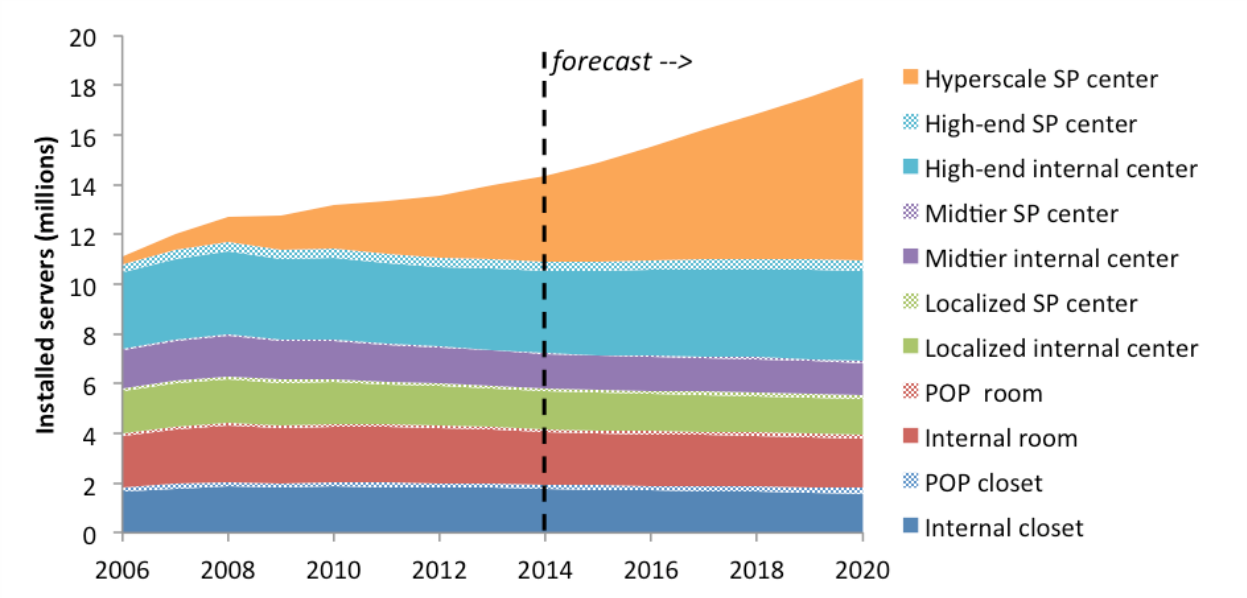



**Figure 20. Total Server Installed Base by Data Center Space Category**

Once total server installed base is allocated to various room types, the prevalence of each of the six server classifications in this study (volume branded 1S and 2S+, volume unbranded 1S and 2S+, mid-range, and high-end) is allocated to the various space categories. Mid-range and high-end servers were allocated based on a set of assumptions used in the 2007 Report: (1) no mid-range or high-end servers are in server closets (2) 5% of mid-range are in server rooms and (3) 65% and 75% of mid-range and high-end servers, respectively, are in high-end data centers. The remainder of mid-range and high-end servers were split evenly between localized and mid-tier data centers. Within a space type, servers were split between internal and service provider space categories based on the total number of data centers in each category. All volume unbranded servers were assumed to be in hyperscale data centers. Lastly, branded volume servers were assumed as the difference between the total number of servers assigned to a given space category, and the number of mid-range, high-end, and unbranded servers assigned in that category.

External storage is assumed to be present in all data center space types except server closets and server rooms, and is allocated based on the number of servers in each space type in the given year. Network ports are also distributed among space types in direct proportion to the number of servers in the given space type. While total number of ports per server is constant across space categories (in a given year), the speed of the ports installed varies between space categories. Space categories were grouped into “fast” (hyperscale and high-end), “medium” (mid-tier and localized), and “slow” (rooms and closets), and the installed base of ports was distributed accordingly.

Infrastructure energy consumption (cooling equipment, uninterrupted power supplies, lighting, etc.) is calculated using the power usage effectiveness (PUE) metric.[31] This metric represents the total energy required by the data center in relation to the energy needed for the IT equipment. A data center with PUE of 1 would use no electricity other than the IT equipment. At

the time of the 2007 Report, the average PUE was estimated to be 2.0, indicating that non-IT energy use was about equal to the IT energy use in a data center. Other studies at that time showed many data centers with PUE values greater than 3.0.[32] [33] [34] More recent studies show that while a wide range of PUE values is still observed, the average PUE has only modestly improved to about 1.8-1.9.[35] [36] The slower rate of efficiency improvement in PUE relative to IT equipment is partially due to the slower turnover rate of a data center's infrastructure relative to the IT equipment. The opportunities to improve data center PUE increase with larger data centers that have the ability to develop better airflow management and employ more efficient cooling equipment or advanced cooling technologies such as liquid cooling. Consequently, smaller data centers are still being measured with PUE values greater than 2.0[37] while large hyperscale cloud data centers are beginning to record PUE value of 1.1 or less.[38] [39] [40] Table 4 presents the characteristic infrastructure equipment and corresponding average PUE values assumed for each data center size for 2014. PUE values for each data center size are based on previously published values, expert elicitation, and energy modeling results for different data center infrastructure configurations.[41] For the current trends scenario, average PUE values for each size data center are assumed to be the same for internal and service provider data centers and, given the very modest improvements observed in published data for average PUE values, to improve by 1% per year through 2020, except for closets. The PUE for closets is held constant at 2.0 throughout the analysis period, given that the lack of dedicated cooling and electrical equipment in this space type limits opportunity for improvemetn. Applying the PUE values in Table 4 to IT energy use in each data center type provides the total energy consumption by data center component (Figure 21) and by space type (Figure 22).

**Table 4. 2014 PUE by Space Type**

| Space Type | IT | Transformer | UPS | Cooling | Lighting | Total PUE |
|---|---|---|---|---|---|---|
| Closet | 1 | 0.05 | - | 0.93 | 0.02 | 2.0 |
| Room | 1 | 0.05 | 0.2 | 1.23 | 0.02 | 2.5 |
| Localized | 1 | 0.05 | 0.2 | 0.73 | 0.02 | 2.0 |
| Midtier | 1 | 0.05 | 0.2 | 0.63 | 0.02 | 1.9 |
| High-end | 1 | 0.03 | 0.1 | 0.55 | 0.02 | 1.7 |
| Hyperscale | 1 | 0.02 | - | 0.16 | 0.02 | 1.2 |

### *2.3.5 Total Data Center Energy Consumption*

As shown in Figure 21, electricity consumption of data centers has been relatively flat in recent years, which is attributable to many factors. The growth rate of server shipments has diminished over the past 15 years. From 2000-2005, server shipments increased by 15% each year resulting in a near doubling of servers operating in data centers. From 2005-2010, the annual shipment increase fell to 5%, most likely driven by the economic recession as well as the emergence of server virtualization during that period. The annual growth in server shipments further dropped after 2010 to 3% and that growth rate is now expected to continue through 2020. This 3% annual growth rate coincides with the rise in hyperscale data centers and an increased popularity of moving previously localized data center activity to colocation or cloud

facilities. In fact, nearly all server shipment growth since 2010 occurred in servers destined for hyperscale data centers, where servers are often configured for maximum productivity and operated at higher utilization rates, resulting in fewer servers needed than would be required to provide the same services in traditional, smaller, data centers.

Along with total server count, the power demand for each server has also changed. While server power requirements were observed to be increasing from 2000-2005, power demand appears to have stayed fairly constant since 2005. Additionally, servers are improving in their power scaling abilities, thus reducing power draw during idle periods or when at low utilization. Efficiency improvements in storage, network and infrastructure also influence the electricity estimates in this report. Storage devices are becoming more efficient on a per-drive basis, with the growth in drive storage capacity projected to outpace increases in data storage demand by 2020, ultimately reducing the number of physical drives needed throughout data centers. Recent estimates of network port power consumption are now much lower than estimates from the past decade. Increased awareness in data center infrastructure operations (e.g. cooling) has resulted in improved efficiency across data center types, though the most significantly in large cloud data centers that are innovatively designed to maximum infrastructure efficiency.

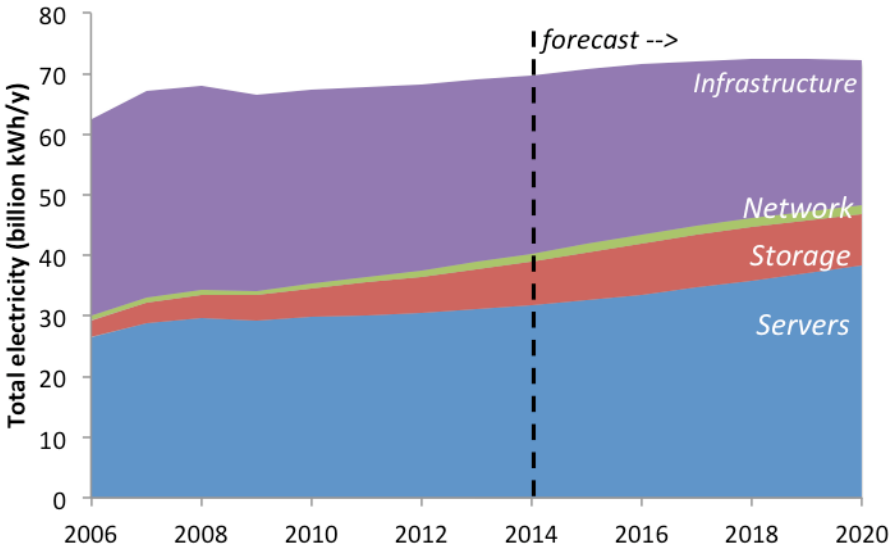



**Figure 21. Total Electricity Consumption by Technology Type**

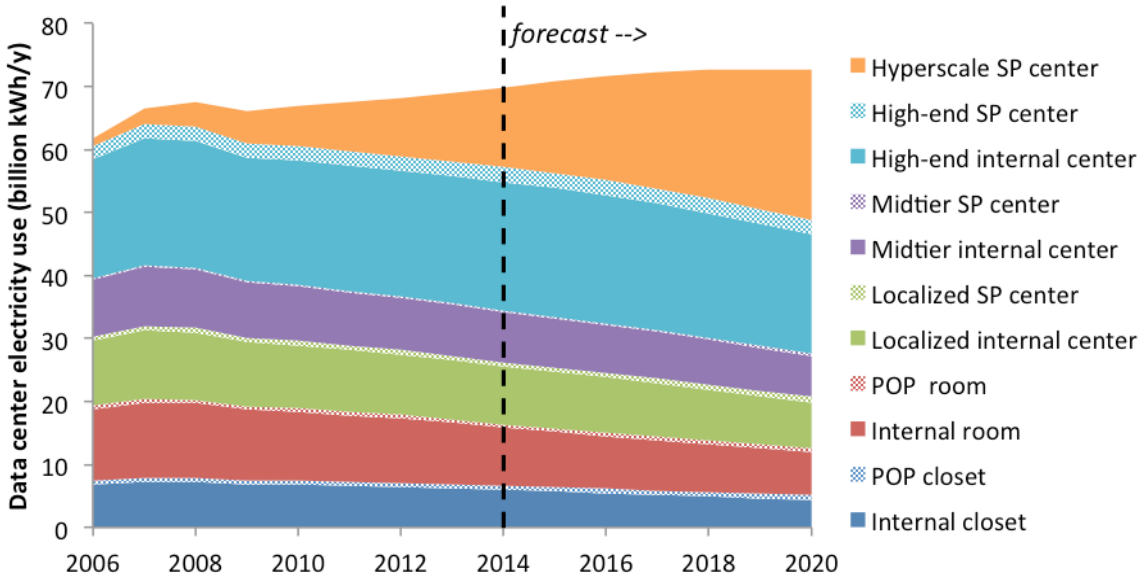



**Figure 22. Total Electricity Consumption by Space Type**

Results from this study for 2000-2014 are shown in Figure 23 alongside results from three previous studies: the 2007 Report; Masanet et al. (2011), which built upon the 2007 Report to estimate consumption in 2008; and Koomey (2011), which estimated 2010 consumption using updated server shipment data after the 2009 financial crisis. While data center operations since 2008 likely benefited from some adoption of energy efficiency measures, such as increased server consolidation and improved IT power management, the distinct reduction in electricity use observed in the later studies immediately after 2008 was likely also due to a lower installed server base associated with the economic recession and the related efficiency improvements driven by pressure to cut energy and equipment costs.

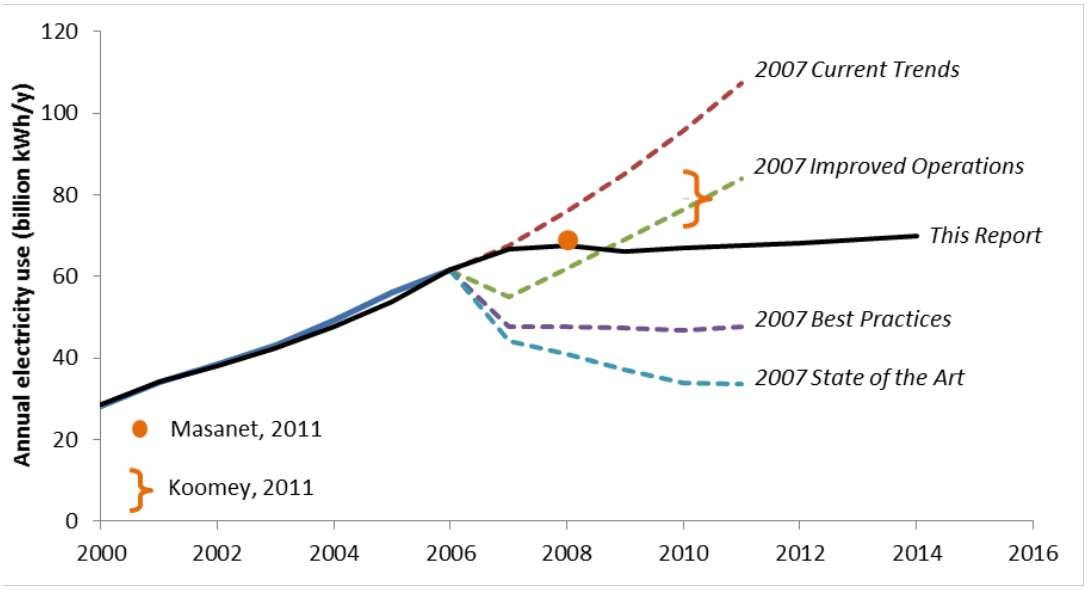



**Figure 23. Historical Data Center Total Electricity Use**

Growth in data center energy consumption has slowed drastically since the previous decade. However, demand for computations and the amount of productivity performed by data centers continues to rise at substantial rates.[42] Technology advancements have made IT equipment more efficient by being able to perform more work on a given device, while other design and management efforts have made the industry more energy efficient. A counterfactual scenario was created for this study that estimates what data center energy consumption would have been if industry energy-savings efforts were halted in 2010. The metrics frozen in 2010 include:

- Industry-wide average server utilization remains at 14%
- Dynamic range of server remains at 0.59 (i.e., servers use 59% of their maximum power at idle) (see Figure 8)
- Wattage per HDD remains at about 11.3 watts (see Figure 15)
- Wattage of network ports remains at 1.6, 2.6, 4.1, and 7.1 for 100Mb, 1000Mb, 10Gb, and 40 Gb, respectively (see Figure 18)
- Industry-wide weighted average PUE remains at 1.89

This scenario does not halt the technological advancements of the computing industry in terms computational performance (i.e., computations/second per servers) and the electrical efficiency of computations (i.e., computations per kWh). Computational performance and the electrical efficiency of computations are assumed to continue to improve in parallel and at similar rates[43] (thereby cancelling each other out), since the main trend driving both advancements are smaller transistors.[42] Additionally, wattage per HDD and per network port are held at 2010 levels but storage capacity (i.e., TB per drive) and a shift towards fasters ports are assumed to progress as normal.

The results of this scenario are shown in

Figure 24, with 2014 energy use that is 60% higher than currently estimated, and projected 2020 use 170% higher than the Current Trends scenario. Energy savings of the industry are therefore estimated to be 100 billion kWh from 2010-2014, and an additional 520 billion kWh from 2015-2020. The overwhelming majority of these savings come from the servers and infrastructure. Server savings are driven by the increase in industry-wide average utilization that results from consolidation efforts and the growth data centers that operate at higher utilization levels (i.e., hyperscale and other service provider data centers). Infrastructure savings result from the reduced amount of IT equipment that require cooling and electrical services as well as the decrease in industry-wide average PUE, brought down by the growth in data centers with very low PUE values (i.e., hyperscale data centers).

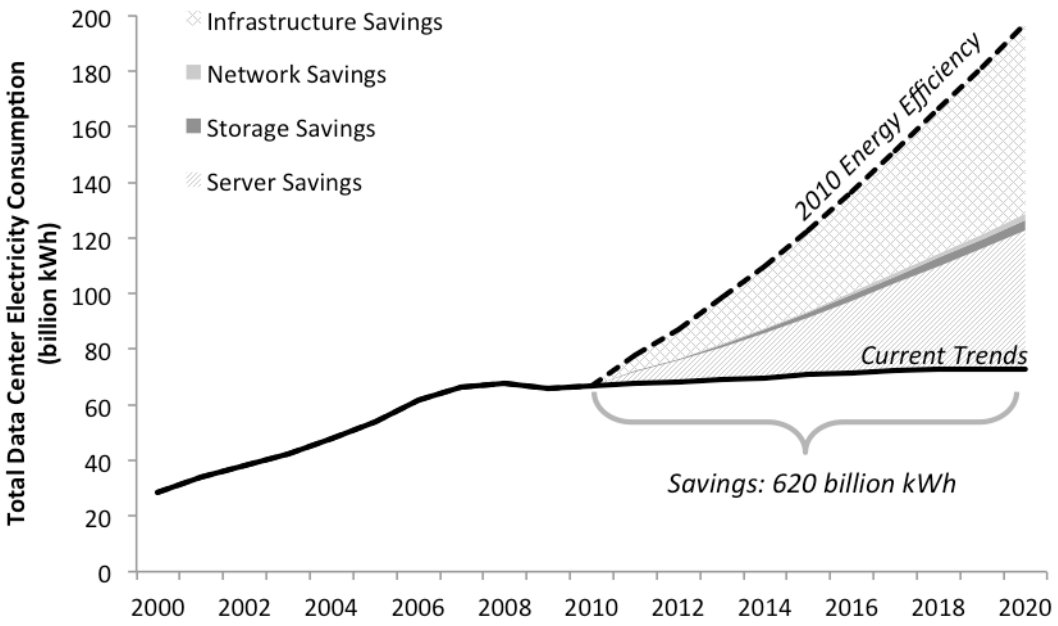



**Figure 24. Data Center Electricity Consumption in Current Trends and 2010 Energy Efficiency Scenarios**

*The 2010 Energy Efficiency scenario assumes that data center energy-related design and operational efforts do not continue past 2010, which indicates that current trend energy efficiency practices will have saved 620 billion kWh of electricity over the period 2010-2020.*

### *2.3.6 Data Center Water Consumption*

Along with electricity demand, data centers require significant water consumption during operation. Water is consumed at two key points during data center operation.[44] First, water is required during the generation of electricity from primary energy that is eventually transmitted for use at the data center site. A national average of 7.6 liters (2.0 gallons) of water are consumed for each kWh when weighting the water losses at both thermoelectric and hydroelectric plants in the U.S.[45] Second, in medium to larger size data centers that employ cooling tower based chillers to improve energy efficiency, water is consumed at the data center site itself. Cooling towers use water evaporation to reject heat from the data center causing losses approximately equal to the latent heat of vaporization for water, along with some additional losses for drift and blowdown. In larger data centers this on site water consumption can be significant, with data centers that have 15 MW of IT capacity consuming between 80-130 million gallons annually.[46] [47] In this study, on-site water consumption is estimated at 1.8 liters (0.46 gallons) per kWh of total data center site energy use for all data centers except for closet and room data centers, which are assumed to use direct expansion (air-cooled chillers). With these assumptions, approximately 626 billion liters of water was estimated to be consumed in 2014 for data centers, with that number reaching 660 billion liters in 2020. Data center water consumption is shown in Figure 25 and Figure 26. Note that water consumption associated with the generation of electricity varies significantly by primary energy type and power plant

efficiency, such that the actual water consumption attributable to any individual data center will depend on its specific location and electricity provider.

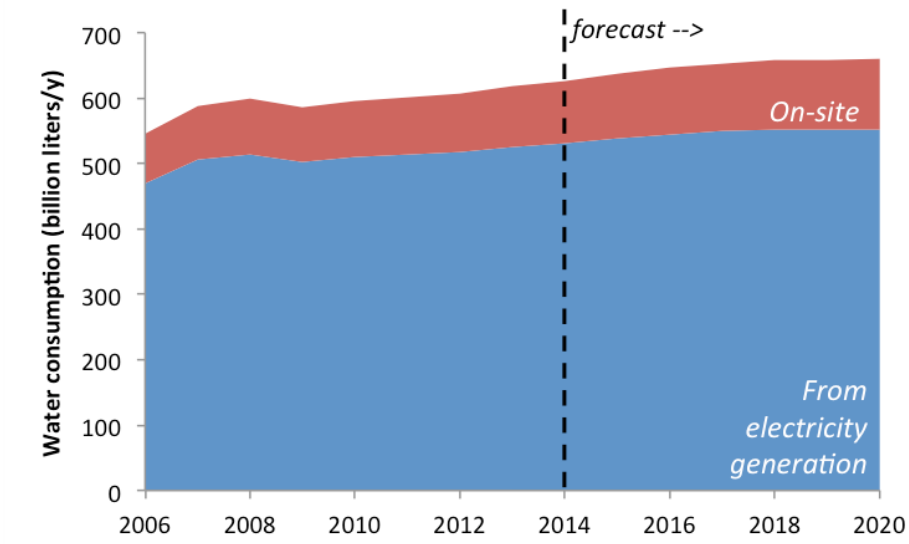



**Figure 25. Direct vs. Indirect U.S. Data Center Water Consumption**

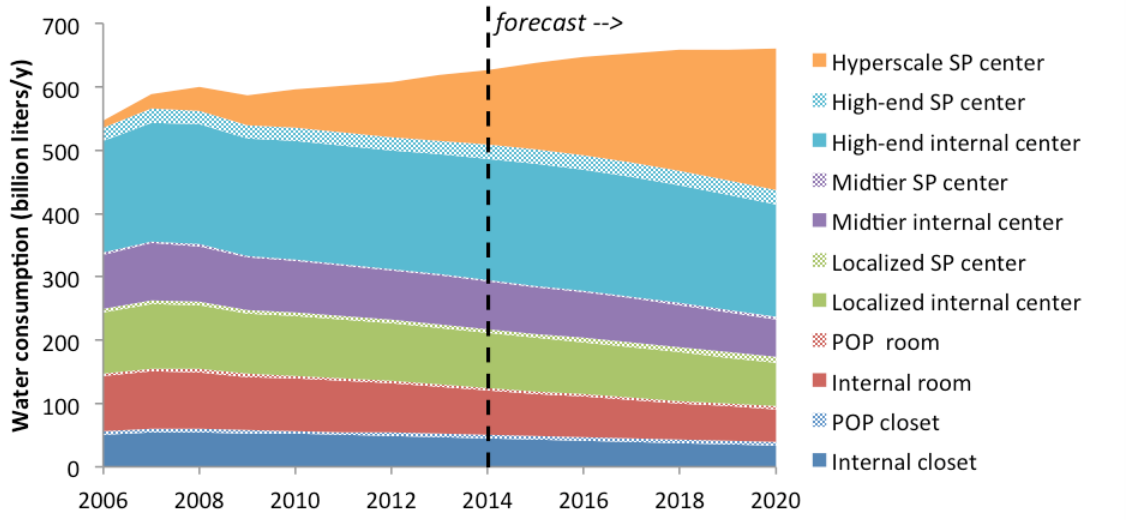



**Figure 26. Total U.S. Data Center Water Consumption by Space Type**

# 3 Energy Use Associated with Federal Government Servers and Data Centers

The Federal Data Center Consolidation Initiative (FDCCI), established in 2010 to reverse the historic growth of Federal data centers, has been conducting surveys on federally operated data centers to verify existing records and expand inventory data for these facilities (e.g. address, operational status and cost, square footage, server count, etc.).[48] For this study, server count

estimates were provided by the Office of Management and Budget (OMB) for 2012-2014, corresponding to approximately 1-2% of nationwide server installed base estimates, as shown in Table 5.

**Table 5. Servers in Federal Data centers Tracked by OMB**

| | 2012 | 2013 | 2014 |
|---|---|---|---|
| Total Server Count | 238,709 | 295,316 | 118,851 |
| Percent of National Installed Base | 1.8% | 2.1% | 0.8% |

Estimates provided in Table 5 do not capture shipments of servers to facilities that were managed by contractors for the federal government and adequate information to estimate the percent of federal server use that occurs in these facilities is not available. Therefore, these numbers provide only a portion and not a complete representation of the impact of the federal government associated with data center operation.

This estimate from OMB is a sharp decrease from the 2007 Report where federal servers and data centers were estimated to be about 10% of the U.S. total. At that time, no data could be found in the public domain on the number of servers or data centers operated by (or for) the federal government and therefore the server estimate was based on interviews conducted with major U.S. manufacturers. These interviews led to the conclusion that server shipments to the federal government represented about 5-10% of annual U.S. server shipments and the higher end of this range was used in an attempt to account for servers shipped to government contractors that were not counted as federal shipments.

FDCCI efforts have resulted in the closing of approximately 1.7 million square feet of data center space since 2010.[49] According to current estimates of server density in various space types this correlates to the removal of approximately 60,000 servers from the federal inventory or about half of the 2014 estimate. However, even without these consolidation efforts the 2014 server count would correspond to only 1.2% of the national installed base. This large discrepancy compared to the 2007 Report estimate implies OMB's server accounting may not be capturing a significant portion of federal servers in the U.S. data center stock.

No data on federal server shipments by server class (i.e., volume, mid-range, and high-end) are available. It is possible that the federal government accounts for a significant fraction of high-end server electricity use, given that about one third of the 100 largest supercomputers in the world are housed in U.S. government-owned facilities, including four out of the top ten in the world.[50] This larger representation of high-end servers would potentially put federal energy consumption at a higher percentage of the national installed base than the percentage represented in Table 5.

# 4 Expected Energy Savings Opportunities

## 4.1 Energy Efficiency Trends

A number of energy efficiency trends have been shaping server operations over the past decade, including virtualization and consolidation, which are expected to continue to generate energy savings and reduce server footprint. Physical parts of the server such as the microprocessor, cooling fan, and power supply are also improving in energy efficiency (or being eliminated, such as the fan in servers with liquid cooling), thus further reducing server power consumption per unit of computing output. More efficient storage devices such as SSDs are coming down in cost and their increasing prevalence will continue to drive efficiency. Furthermore, the growing trend of cloud computing has resulted in significantly larger data centers that are more efficient, both in terms of server utilization and infrastructure PUE, compared to traditional enterprise data centers. Along with the current trend previously described in this report, three additional scenarios and two combinations of scenarios were modeled to estimate near-term energy efficiency opportunities in U.S. data centers.

## 4.2 Improved Management Scenario

The Improved Management (IM) scenario consists of two components: improved PUE and removal of inactive servers. Improved PUE represents an effort for smaller data centers to reduce their infrastructure (e.g. cooling, lighting) energy demand through improved airflow and thermal management, such hot/cold isle isolation and reducing set point temperatures. PUE values trend linearly after 2014 to the improved numbers shown in Table 6. Inactive servers (also referred to as comatose or “zombie” servers) represent obsolete or unused servers that consume electricity but provide no useful information services. Previous studies have estimated that inactive servers represent 10-30% of servers in U.S. data centers.[51] [52] [53] [54] Inactive server removal represents an impact of raised awareness on part of data center operators as to what equipment is being utilized in the data center. In order to model this impact, servers are first assigned to each space type as described in Section 2.2.4, then the number of volume servers in internal data centers is decreased by 10%, and in service provider data centers by 5%, for each year after 2014. The percent decrease in service provider data centers is assumed to be smaller because these data centers tend to have a lower rate of inactive servers due to better management practices that avoid the institutional problems of dispersed responsibility between IT and facility departments which often plagues internal data centers.[55]

The IM scenario results in total energy savings in 2020 relative to the Current Trends scenario of 10% which is composed of: 2.5% from server energy savings due to the removal of inactive servers; no savings in storage or network energy; and 7.5% from savings in infrastructure energy due to the reduced server energy along with improved PUE.

**Table 6. PUE and Redundancy Values for Efficiency Scenarios**

| Space Type | 2014 PUE | 2020 PUE | | | Redundancy |
|---|---|---|---|---|---|
| | | Current Trends | Improved Management | Best Practices | |
| Closet | 2.0 | 2.00 | 2.00 | 2.00 | N+0.5N |
| Room | 2.5 | 2.35 | 1.70 | 1.50 | N+1 |
| Localized | 2.0 | 1.88 | 1.70 | 1.50 | N+1 |
| Midtier | 1.9 | 1.79 | 1.70 | 1.40 | N+0.2N |
| High-end | 1.7 | 1.60 | 1.51 | 1.30 | N+0.5N |
| Hyperscale | 1.2 | 1.13 | 1.13 | 1.10 | N |

## 4.3 Best Practices Scenario

The Best Practices (BP) scenario builds upon the improvements in the IM scenario. These best practices include further improved PUE values (Table 6) from using more efficient infrastructure components and employing economizer or liquid cooling when applicable, as well as improved dynamic range (power scaling ability) of servers, server and network consolidation efforts, and reduced storage disk and network port power consumption.

In all scenarios, average dynamic range is calculated by determining what mix of servers follow a minimum versus a maximum dynamic range trend (Figure 8). For the CT scenario, the average dynamic range of volume servers is assumed to be represented by 90/10 mix of the maximum and minimum dynamic range trends. The BP scenario has a more aggressive penetration of servers that follow the maximum dynamic range trend and assumes that this ratio reaches 50/50 by 2020. This results in the dynamic range trend shown in Figure 27, where volume servers reach a dynamic range of 0.28 in 2020. While these number are possible, it should be noted that getting below 20-25% of maximum power at idle can require powering down specific functionality that may increase idle-to-active wake-up latency and result in delayed response times. Attempts to improve the idle power savings versus latency include establishing different levels of inactive modes with varying degrees of power savings and response times.[56] [57] As discussed in Section 2.2.1, these numbers are used to calculate the slope of the utilization versus power line (Equation 2) and therefore the power consumption at average utilization levels (Equation 3).

Server consolidation entails replacing multiple servers running at low utilization with a single server running at a higher utilization. Eighty percent of the volume server installed base is assumed to be consolidated onto servers that operate at the utilization levels presented in Table 7. These parameters are chosen based on the assumption that servers in internal and service provider data centers could be consolidated through methods such as virtualization and containerization to the current high end of utilization observed in servers in small and medium size data centers[17] [18] and servers in hyperscale data centers could be consolidated to 75% utilization through improved CPU bandwidth provision techniques, which have been shown to achieve average data center server utilizations as high as 90%.[58] There is a small additional benefit to consolidation by reduction in redundant servers required, which is calculated

according to the data center size-specific redundancy frameworks presented in Masanet (2013),[59] shown in Table 6. The formulas in the "Redundancy" column represent the total number of servers needed for a data center containing N functional servers. For example, redundancy of "N+1" means that there is one redundant server present in each data center, while redundancy of "N+0.1N" means that there is one redundant server for every 10 functional servers. For data centers where the number of redundant servers scales with server count (i.e. closets, mid-tier, and high-end enterprise), consolidation of servers reduces the number of redundant servers required.

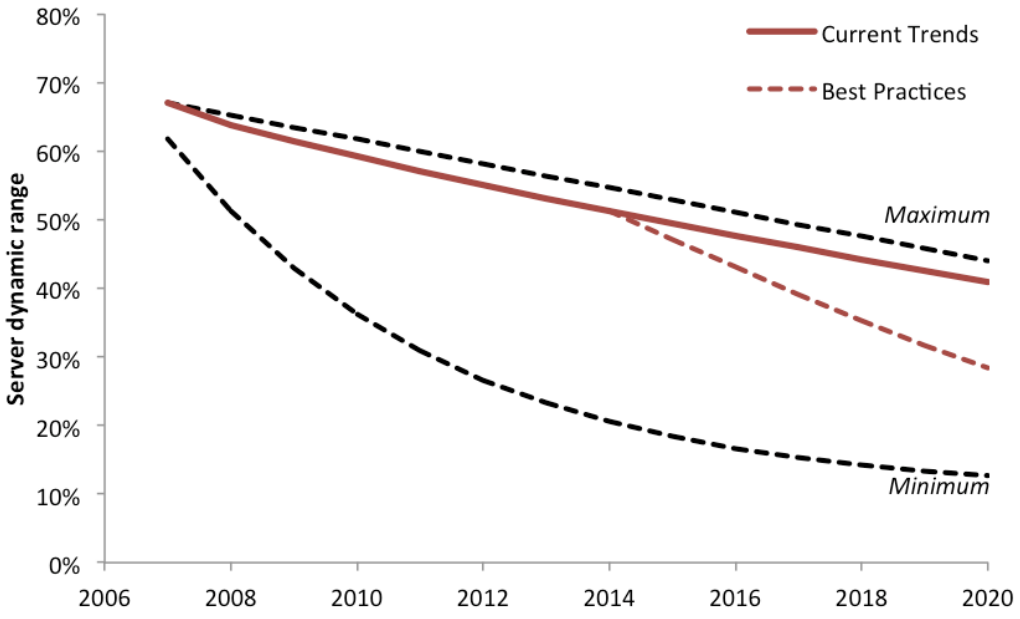



**Figure 27. Average Volume Server Dynamic Range for Current Trends and Best Practices Scenarios**

**Table 7. Best Practices Scenario Consolidation Parameters**

| Space Type | Percent of installed base consolidated by 2020 | Utilization | | Consolidation utilization overhead |
|---|---|---|---|---|
| | | Pre-consolidation | Post-consolidation | |
| Internal | 80% | 10-15% | 45% | 5% |
| Service Provider | 80% | 20-25% | 55% | 5% |
| Hyperscale | 80% | 45-50% | 75% | 5% |

Equation 6 describes the number of servers that are consolidated each year from 2015-2020. Equation 7 describes the number of pre-consolidation servers that can be replaced by each post-consolidated server, based on pre- and post- consolidation utilization, utilization "overhead", and redundancy considerations. Utilization overhead is estimated as 5% per post-consolidation server and accounts for the applications that must be run on the server to balance multiple workloads. In essence, this means that consolidation of two servers running at 10% utilization would result in one server running at 25% utilization. Because the scenario's post-

consolidation utilizations are defined, Equation 7 is used to calculate the number of pre-consolidation servers that are replaced by each post-consolidation server. The number of servers consolidated each year (Equation 6) is then divided by this metric to determine how many servers (out of those that are consolidated) remain in the installed base, as shown in Equation 8. Finally, the new post-consolidation installed base is calculated as shown in Equation 9. Consolidation impacts on server installed base are shown in Figure 28.

**Equation 6**

$$S_{con,y} = C_{y_f} * \frac{y - y_i}{y_f - y_i} * IB_y$$

Where $S_{con,y}$ = number of servers to be consolidated in year y
$C_{yf}$ = Percent of installed base consolidated in final year
$y_i$ = year consolidation begins
$y_f$ = year consolidation ends
$IB_y$ = baseline installed base in year y

**Equation 7**

$$N_{con} = \frac{u_{post} - u_o}{u_{pre}} * \frac{1}{r_{pre}}$$

Where $N_{con}$ = number of servers that can be consolidated into one
$u_{post}$ = post-consolidation server utilization
$u_o$ = consolidation utilization overhead
$u_{pre}$ = pre-consolidation server utilization
r = fraction of pre-consolidation servers that are non-redundant

**Equation 8**

$$\boldsymbol{S_{res,y} = \frac{S_{con,y}}{N_{con}}}$$

Where $S_{res,y}$ = number of consolidated servers resulting in year y
$S_{con,y}$ = number of servers consolidated in year y
$N_{con}$ = number of servers that can be consolidated into one

**Equation 9**

$$\boldsymbol{IB_{y,post} = IB_y - S_{con,y} + S_{res,y}}$$

Where $IB_{y,post}$ = installed base in year y post-consolidation
$IB_y$ = baseline installed base in year y
$S_{con,y}$ = number of servers to be consolidated in year y
$S_{res,y}$ = number of consolidated servers resulting in year y

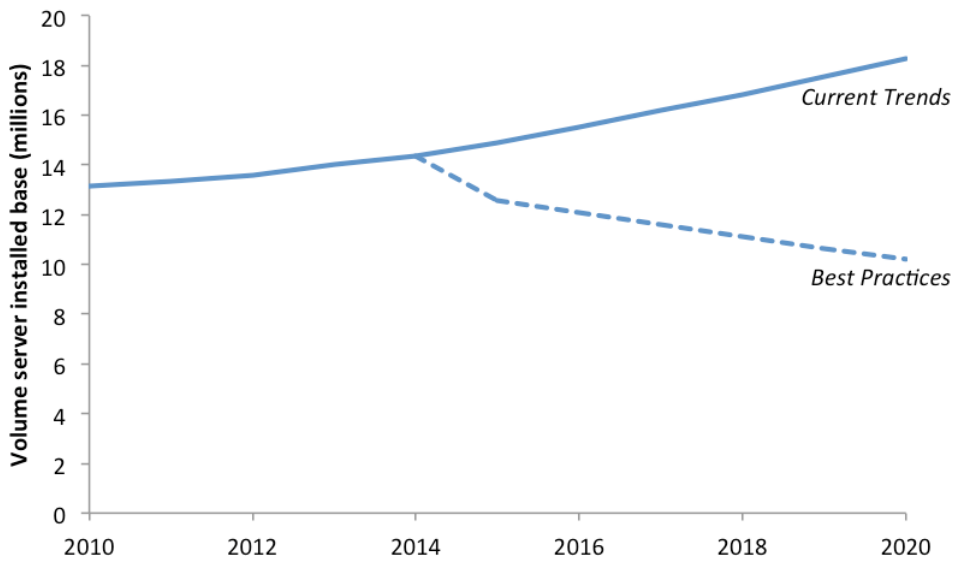



**Figure 28. Volume Server Installed Base for Current Trends and Best Practices Scenarios**

Network port consolidation is similar to server consolidation. Eighty percent of the installed base of 10 GB ports are assumed to be consolidated into 40 GB ports, which can transmit 4 times the data at only 1.7 times the power. In addition, the BP scenario assumes that ports installed in high-end enterprise and hyperscale data centers will become 25%[60] more efficient by 2020. This reduction in port wattage is based on network improvements in network topology, dynamic link rate adaptation, and link and switch sleep modes.[25] BP scenario impacts on port installed base are shown in
Figure 29.

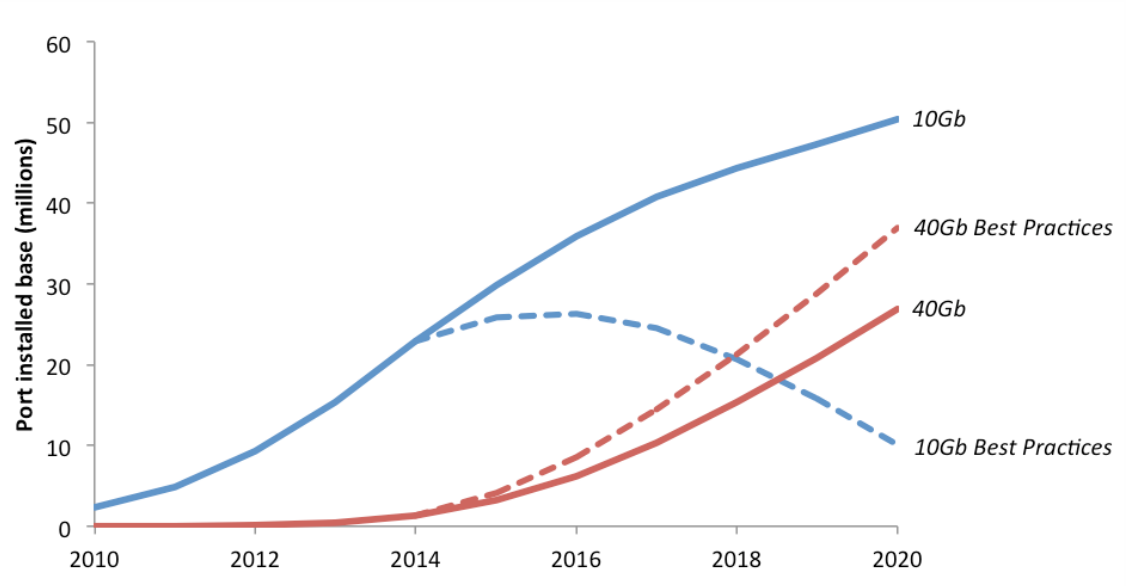



**Figure 29. Network Installed Base for 10 GB and 40 GB Ports in Current Trends and Best Practices Scenarios**

The BP scenario also includes a provision to improve the efficiency of storage disks by 25% in 2020. This improvement increases linearly, and is based upon a scenario where future storage disks have reduced energy usage in their idle state. The resulting disk power consumption is shown in Figure 30.

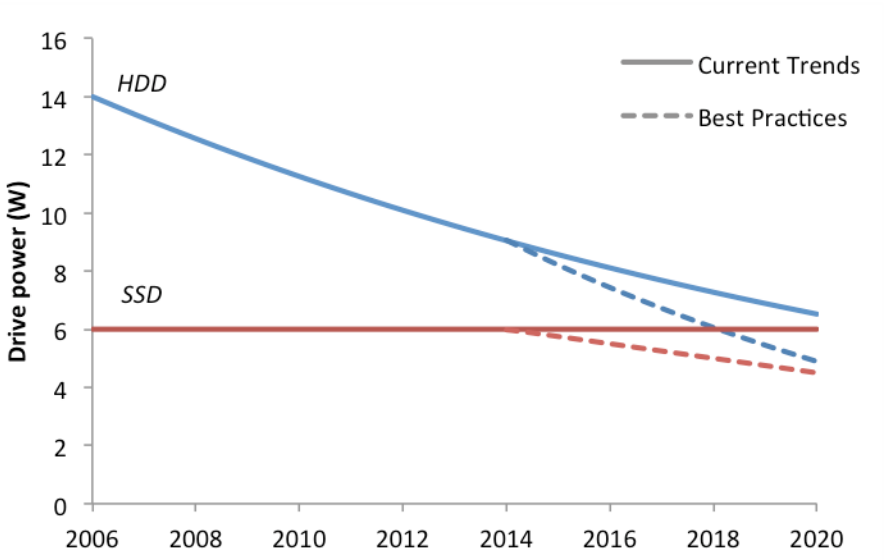



**Figure 30. Storage Disk Power Consumption for Current Trends and Best Practices Scenarios**

The BP scenario results in total energy savings in 2020 relative to the Current Trends scenario of 40% which is composed of: 15% from server energy savings due to the removal of inactive servers, the improvement in server scaling, and consolidation of servers; 3% from savings in storage energy due to improved disk efficiency, 1% from savings in network energy due to port consolidation; and 22% from savings in infrastructure energy due to the reduced IT energy along with improved PUE.

## 4.4 Hyperscale Shift Scenario

The Hyperscale Shift (HS) scenario involves the consolidation of 80% of the servers in non-hyperscale data centers into hyperscale data centers, excluding servers in Service Provider Rooms and Closets, which are defined to include local point-of-presence facilities that would still be needed to support management functions for hyperscale data centers.[27] The 80% shift is in addition to the shift already accounted for in the Current Trends scenario. Hyperscale data centers operate servers at higher utilizations in infrastructure-efficient (low-PUE) spaces, and include cloud-based platforms that remove the need for dedicated redundancy servers, so that consolidating IT services from many small disparate data centers into hyperscale data center can yield significant energy savings. HS scenario impacts on volume server installed base are shown in Figure 31. As shown, the installed base in non-hyperscale data centers decreases to approximately one quarter of the CT scenario value by 2020. The impact on the installed base in hyperscale data centers is smaller since, on average, one server in a hyperscale data center can replace 3.75 servers in non-hyperscale data centers. This is because servers in hyperscale

data centers are assumed to run at roughly 3 times the utilization of non-hyperscale data centers and have no redundancy requirements.[61]

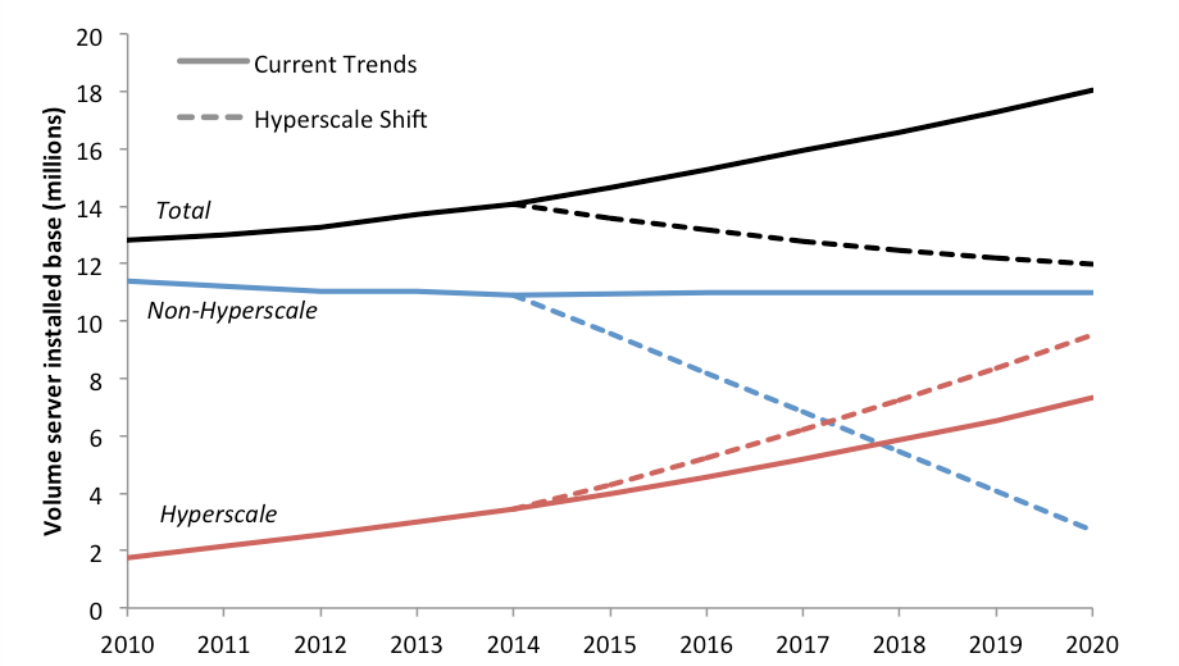



**Figure 31. Volume Server Installed Base in Current Trends and Hyperscale Shift Scenarios**

The HS scenario results in total energy savings in 2020 relative to the Current Trends scenario of 25% which is composed of: 10% from server energy savings, due to the consolidation; no savings in storage or network energy; and 15% from savings in infrastructure energy due to the reduced server energy along with the relocation of servers into data centers with lower PUE values.

## 4.5 Scenario Results

Server energy use is affected in all scenarios. Relative to the CT scenario, IM, BP, and HS measures reduce server energy consumption by 4%, 18%, and 28% respectively. When HS measures are applied to the IM and BP scenarios, server energy consumption is reduced an additional 16% and 3% of the CT value, respectively. These results are shown in Figure 32.

Storage energy use is only affected in the BP scenario where disks are assumed to be 25% more efficient in 2020 than in the CT scenario. Therefore, storage energy is 25% less in the BP and BP+HS scenarios relative to CT. Network energy is also only affected in the BP scenario due to both port consolidation and port efficiency improvements. The result is a 38% reduction in network energy consumption in BP and 40% in BP+HS scenarios relative to CT. This is shown in Figure 33.

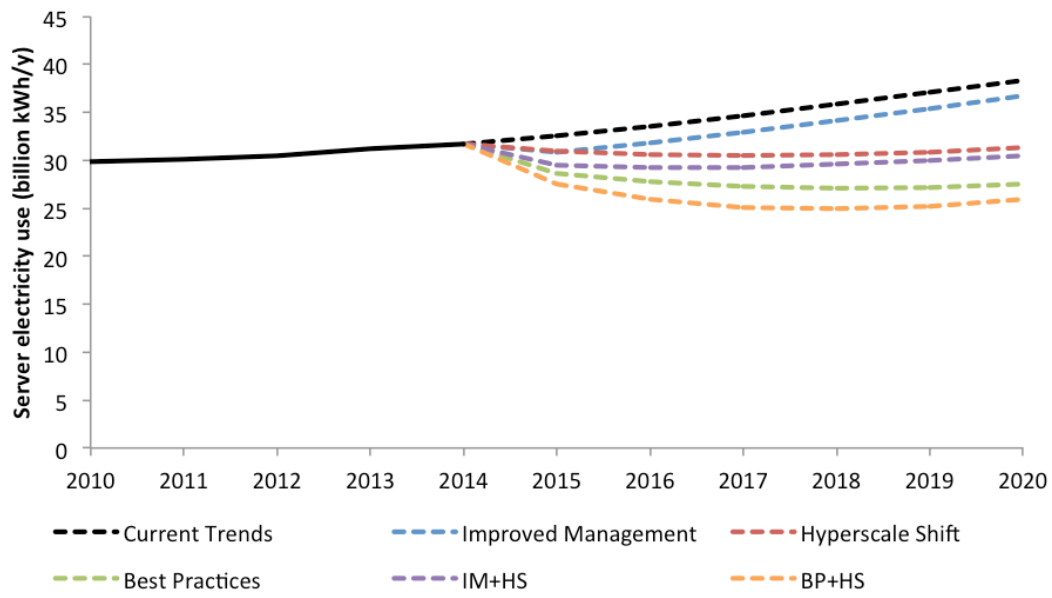



**Figure 32. Server Electricity Use for All Scenarios**

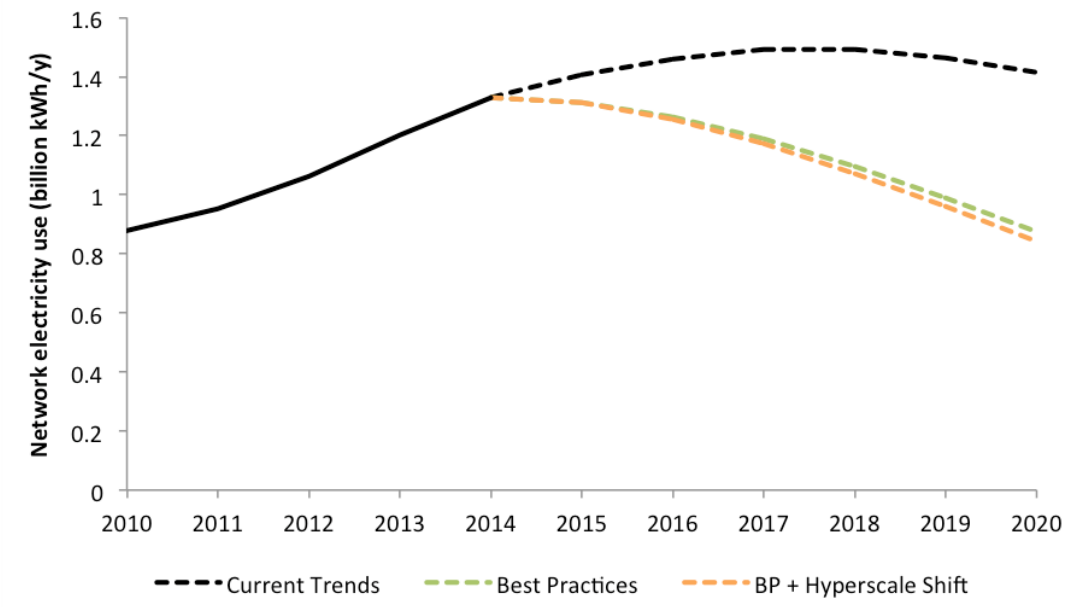



**Figure 33. Network Electricity Use for Current Trends and Best Practices Scenarios.**

*Improved Operation, Hyperscale Shift, and IO+HS scenarios are identical to Current Trends. BP+HS network electricity is slightly lower than BP alone due to the higher prevalence of network ports in hyperscale data centers where the 25% wattage reduction is assumed to occur.*

Infrastructure energy is affected in all scenarios, as shown in Figure 34. Infrastructure energy is the product of the space type PUE and the space IT energy consumption, so reductions in both of these areas compound for very large energy savings relative to the CT scenario. The HS

scenario has infrastructure energy savings of 46% relative to the current trends scenario. This is due both to the reduction in server energy as well as the shift of this energy into hyperscale data centers, which have a lower PUE than other space types. The IM scenario saves 22% of infrastructure energy relative to CT due to both the reduction in IT energy and improvement in PUE across smaller space types. IM and HS in conjunction save 54% of infrastructure energy. Lastly, the BP and BP+HS scenarios have infrastructure energy savings of 64% and 75% respectively due to the large decrease in server energy, additional decrease in storage and network energy, and strong improvement in PUE across all space types.

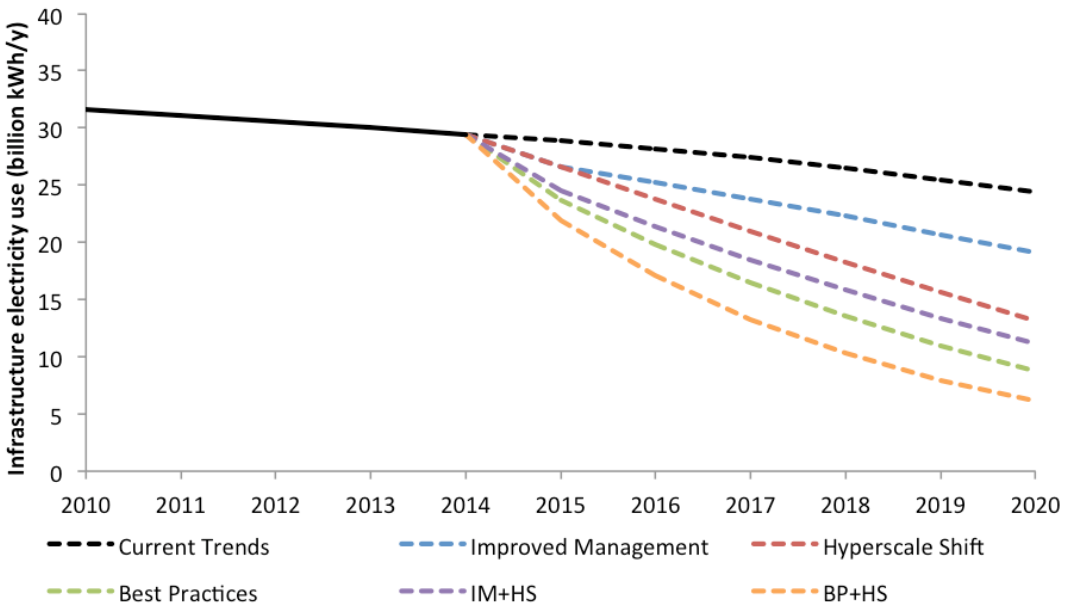



**Figure 34. Infrastructure Electricity Use for All Scenarios**

Total data center energy consumption for all scenarios is shown in
Figure 35. Total electricity savings in 2020 for the IM, BP, and HS scenarios are 10%, 40%, and 25% respectively, relative to the CT scenario. When HS measures are applied to the IM and BP scenarios an additional 20% and 6% of CT total energy is saved. Water savings are very similar with scenario 2020 savings as follows: IM, 9%; BP: 39%; HS: 24%; IM+HS, 28% (18% more than IM alone); and BP+HS, 45% (5% more than BP alone).

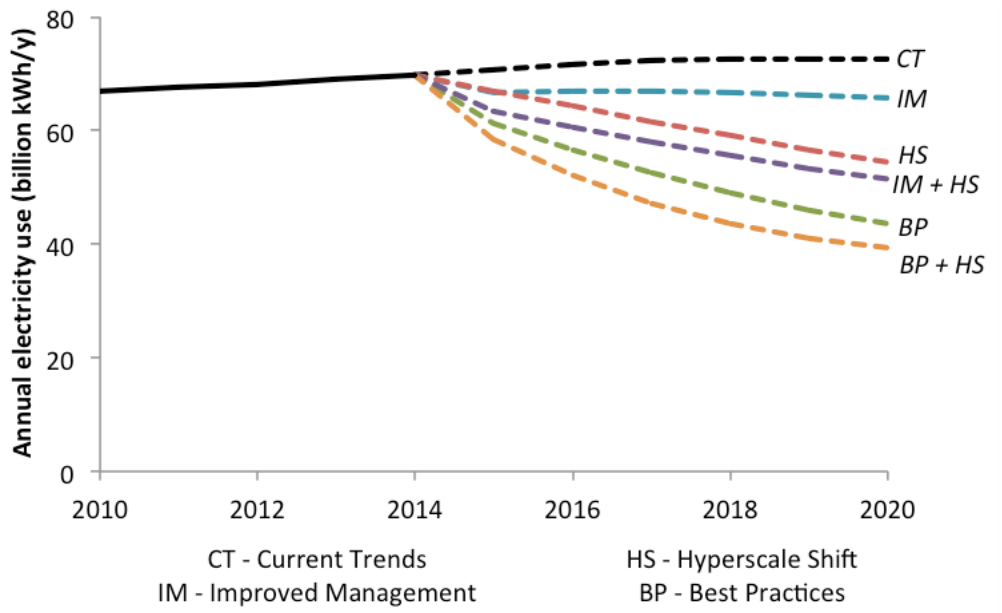



**Figure 35. Total Electricity Consumption for All Scenarios**

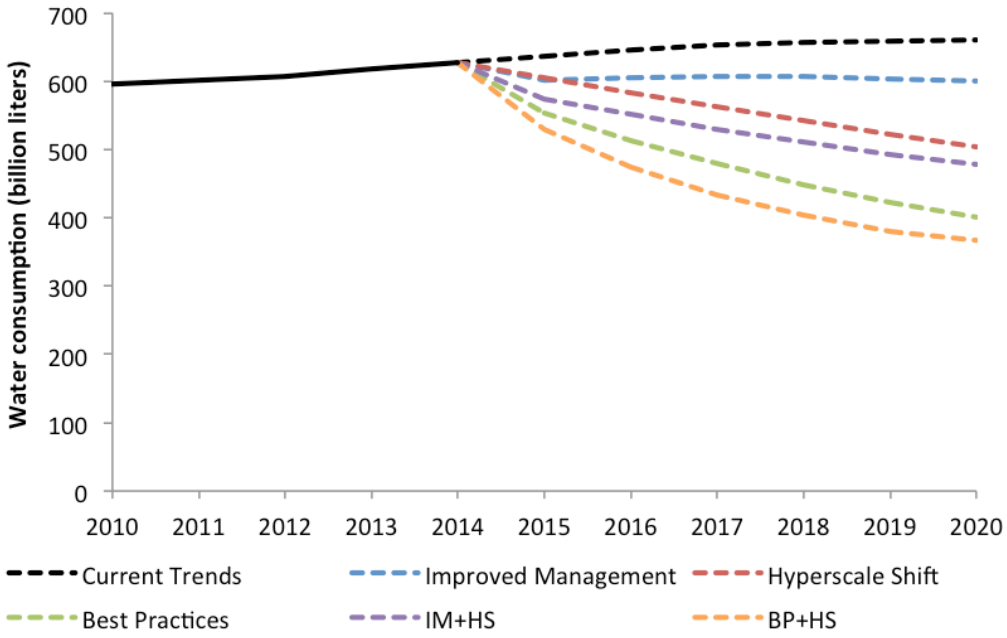



**Figure 36. Water Consumption for All Scenarios**

# 5 Indirect Energy Impacts

Data centers constitute a foundational component of the information and communication technology (ICT) that provides the services—such as streaming media, email, Internet content, and electronic recordkeeping—now ubiquitous throughout much of the world.[a] Today, our computers, mobile devices, sensors, and networks are more likely than ever to utilize centralized information stored on a data center in the cloud. While the primary focus of this report is to estimate the electricity directly consumed in data centers, it is important to take a broader perspective and note that the services provided by data centers can profoundly affect how—and how much—energy[b] is used elsewhere by society. For example, ICT can enable existing products and services to become more efficient or create "e-substitutes" for physical products. Other, higher-order effects occur when the introduction of ICT causes a change in consumption or production elsewhere in the economy.

These implications for broader energy use are known as *indirect energy impacts*, and they might either reduce or increase overall energy use (Figure 37). From an energy savings standpoint, if the indirect energy impacts offset the direct consumption by data centers and other ICT equipment, then the synergy between ICT and societal energy savings is positive; if, instead, the introduction of ICT causes an amplifying effect in which overall energy consumption increases, the synergy is negative. Characterizing this "net" impact of ICT deployment on societal energy consumption has been of great interest, as evidenced by the emergence of an ICT for Sustainability research community,[62] two special issues in the *Journal of Industrial Ecology*[63] [64] an OECD effort to link statistical indicators between the ICT and environment research fields,[65] work in "green computing" from the computer science field,[66] and a variety of other reports.[67] [68] This report chapter provides brief summary of the literature review and interpretation found in Horner et al.[69] on the indirect impacts of ICT.

There is, in fact, no consensus on the magnitude or even the sign of ICT's indirect energy impact. Generally in the positive synergy camp are Romm et al.[70] and a series of reports published by the American Council for an Energy Efficient Economy,[71] [72] who anticipate ICT-enabled energy efficiency gains across broad sectors of the economy, and work commissioned by the industry-sponsored Global e-Sustainability Initiative,[73] which estimates a greenhouse gas (GHG) abatement potential of 20% by 2030 due to ICT deployment. More cautionary is Rattle,[74] who posits that higher-order effects are likely to swamp these sorts of energy savings projections. Berkhout and Hertin[75] argue for moving "beyond the dichotomy between pessimism and optimism" to recognize that the relationship between ICT and energy impacts is "complex, interdependent, deeply uncertain and scale-dependent." Other literature reviews point to an

[a] The term information technology (IT) has been used in this report to refer to the servers, network, and storage equipment in the data center. In this chapter, *ICT* is used as a broader term encompassing data center IT components, other network infrastructure, and a wide variety of end-use devices.

[b] Resource usage and GHG emissions are, of course, concomitants of ICT energy consumption; while this report focuses on energy, much of the literature includes these associated impacts.

ambiguous net impact or acknowledge that this complexity and uncertainty confound attempts to verify a general belief that the net energy savings of ICT should be positive. [76] [77] [78] [79]

## 5.1 Energy Impact Taxonomy

Direct energy consumption, which in addition to the operational energy use estimated in this report also includes the energy required to manufacture and dispose of ICT equipment, is likely the simplest and ultimately the least important ICT energy effect[80] although it is by no means small. One particular issue is that embodied energy can dominate operational consumption for some types of ICT equipment, such as mobile devices.[78] However, the *indirect* energy effects are likely to be of much greater magnitude,[78] owing to the breadth of the various mechanisms by which ICT services alter energy use. Table 8 breaks out individual effects, organizes them into a taxonomy of increasing scope, and maps them to other terms used in the literature, while Figure 37 shows this taxonomy graphically.

Working from narrow to broad scope, ICT adoption first leads to *efficiency* in and *substitution* for conventional products and services. Efficiency occurs when, for example, smart building technology reduces air conditioning energy consumption by tailoring climate-control to the real-time needs of building occupants. An example of substitution is the replacement of air travel with teleconferencing. There is no guarantee, however, that the substituted ICT service will be less energy intense than the conventional service it replaces, and even evaluation of simple cases is not always straightforward.

Any energy reduction achieved through efficiency or substitution can be plagued by *rebound effects*, in which expected gains are offset by induced additional consumption. Azevedo[81] and Gillingham et al.[82] provide comprehensive introductions to rebound effect types, and Borenstein[83] contains a more technical analysis. Rebound is typically broken into direct rebound, indirect rebound, and economy-wide effects. *Direct rebound* effects are own-price-elasticity effects: as prices fall (due to improvements in efficiency or productivity), substitution and income effects increase consumption. For an ICT example, if an e-book is less costly than a conventional book, then consumers might purchase more books. Alternatively, these savings could be spent on other goods and services, which are *indirect rebound* effects. Indirect rebound effects result from cross-price elasticity of demand for other products and services due to increased real consumer income.

**Table 8. Taxonomy of ICT Energy Effects from Horner et al.[69]**

*Scope of effect increases from top to bottom. The third column provides and example of each effect type related to the deployment of Global Positioning System (GPS) Technology.*

| Taxonomy summarized in this report | | | Alternate taxonomies | | | |
|---|---|---|---|---|---|---|
| **Effect** | **Scope** | **GPS System Example** | **Hilty[84]** | **Berkhout & Hertin[75]** | **Williams[91]** | **Rattle[74]** |
| Embodied energy | Direct | Energy to produce a GPS system | 1st-order | Direct effects | ICT infrastructure and devices | |
| Operational energy | | Energy to operate a GPS system | | | | |
| Disposal energy | | Energy to dispose of a GPS system at end-of-life | | | | |
| Efficiency | Indirect: Single-service | More efficient traffic flow due to GPS-enhanced routing | 2nd-order | Indirect effects | Applications | Optimization |
| Substitution | | Replacement of paper maps | | | | Substitution |
| Direct rebound | | More travel due to lower cost of traffic congestion | 3rd-order | Structural & behavioral effects | Effects on economic growth and consumption patterns | Induction |
| Indirect rebound | Indirect: Comple-mentary services | Energy consumed during time saved by more efficient travel | | | | Supplement-ation |
| Economy-wide rebound (Structural change) | Indirect: Economy-wide | GPS enables autonomous vehicles, causes growth of intelligent transportation system manufacturing | | | | Creation |
| Systemic Transformation | Indirect: Society-wide | Autonomous vehicles alter patterns in where people choose to live and work | | | Systemic effects on technology convergence & society | |

*Economy-wide* effects occur when the ICT introduction causes macroeconomic adjustments across economic sectors. That is, the ICT industry can promote or inhibit growth in other sectors of the economy, inducing structural changes that have energy use implications of their own. For example, e-commerce is having broad effects on the logistics industry,[85] including growth in urban freight vehicle sales and changing patterns in distribution center floor space,[86] increased trucking and adoption of new pricing strategies by freight carriers[87] and use of more specialized packaging and a broader range of box sizes.[88]

Finally, *transformational effects* refer to the altering of human preferences and economic and social institutions caused in part by the development of ICT.[89] [90] Historical examples include the advent of the telephone and automobile, which heavily altered where and how people lived and

worked. We might conceive of a similar transformation (one of many possible ICT-enhanced futures) in which the fundamental constraints on where people live and work continue to loosen: e-commerce and home delivery make proximity to traditional retail outlets less important, seamless telework results in less commuting, and driverless vehicles allow for more productive use of the commuting time.

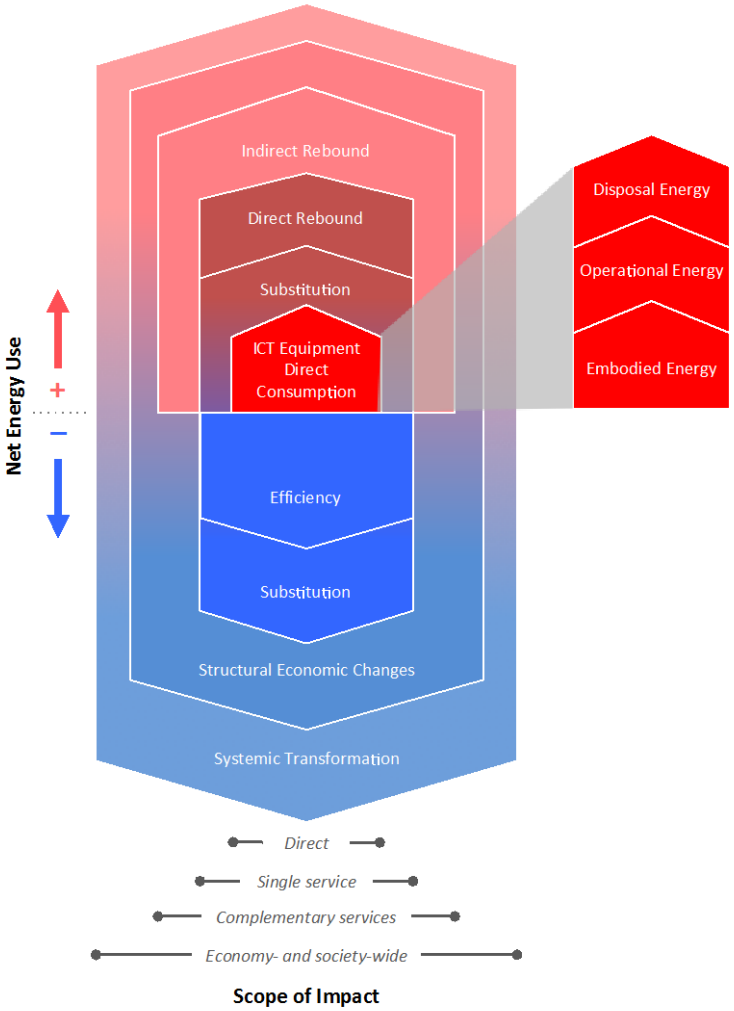



**Figure 37. Taxonomy of Energy Effects from Adoption of ICT, from Horner et al.[69]**

*Red effects increase energy use, blue effects decrease energy use, and shading intensity decreases as effect scope increases.*

As noted by Börjesson Rivera et al.,[76] the existing literature uses several different sets of terms for this hierarchy of effects, collected in the right half of Table 8. ICT energy effects are frequently grouped into first-order impacts due to direct consumption, second-order effects resulting from process changes, such as efficiency, and third-order effects due behavioral and economic changes.[75,84] Williams[91] adds a fourth level, essentially breaking third-order effects into rebound effects and broader systemic change.

## 5.2 Energy Impact Estimation

The literature on characterizing these indirect impacts can generally be divided into two broad types. One body of literature typically assesses the impact of a single service, such as e-commerce, telecommuting, or smart buildings. Such analyses often use lifecycle assessment (LCA),[92] [93] modeling & simulation,[94] or case studies[95] [96] and generally address only direct consumption, efficiency, substitution, and occasionally direct rebound. A separate body of literature focuses on the higher-order indirect energy impacts brought about by ICT-induced changes to complementary services, the broader economy, and societal systems. These studies typically rely on econometric analysis of macroeconomic indicators,[97] scenario analysis,[98] [99] or anecdotal evidence.[100]

Importantly, neither type of study uniformly finds a positive or negative effect: results from both are highly dependent on modeling choices and assumptions used in scoping the study. For example, the energy savings reported in LCA studies of e-commerce for book retail range from negative 500% (i.e., a 5x increase in energy use in the e-commerce case) to nearly 50% (a reduction in energy use by half).[69] This variation results from differences in the system boundary and from sensitivity to assumptions, including population density, freight mode, product return rate, proportion of multipurpose trips, and packaging type. Case studies often show energy savings in specific deployment scenarios and provide valuable lessons on how to deploy ICT in such a way that energy savings are attained; however, scalability and higher-order effects remain uncertain in such work. Macroeconomic studies also show highly variable results.[67]

Thus, while both conceptual discussion and analytical modeling of ICT energy and environmental impacts have been occurring for at least two decades, the jury is still out on the net effects of ICT adoption for several reasons. First, the complexity and variability of ICT deployment schemes makes it difficult to isolate a standard implementation to analyze and to compare study results. Second, the lack of empirical data on how human users interact with ICT systems hinders the ability to assess actual, instead of potential, energy effects. Third, the difficulties in disentangling the causes of interconnected effects lead to a tendency to fall back on theory—and on modeling exercises that conform to these theories. Finally, as the impact scope increases up the effect taxonomy (Table 8), the potential effect's magnitude and uncertainty increase dramatically.

The current state of understanding can be summarized with three related statements: the technical potential of ICT net energy savings is likely positive; the sign and magnitude of realized net energy savings are highly sensitive to the parameters that characterize the ICT deployment and are not guaranteed; and, finally, the actual net energy effect is unclear and difficult to assess, especially when higher-order impacts are considered.

## 5.3 Pathway Forward

Just as implementation of best practices in data center design and operation has the potential to drastically reduce direct energy consumption (Figure 35), optimizing the manner in which we integrate ICT into our lives can have a large impact on overall energy consumption. During the first decade of the century, data center energy consumption grew rapidly (Figure 23). IT

managers focused on service provisioning, with the power bill being a much lesser concern. However, the industry is now converging on a paradigm of virtualization and centralization that has both business and energy co-benefits, and such reductions in direct energy use should continue to be pursued.

The broader evolution of ICT is perhaps on a similar path: new systems and services are being developed rapidly without much consideration of energy impacts, and as a result it seems likely that ICT services are often deployed in a way that does not achieve their full potential to achieve energy savings. However, it also seems likely that more optimal deployment plans—those that create energy savings while maintaining the value of the service—exist, and more focus on characterizing these "system optima" is warranted. [101]

The danger in waiting to identify these deployment plans is that society-wide systems, structures, and habits that become entrenched can be much more difficult to alter. A server has a typical lifetime less than five years; the economic and social infrastructures built out through new ICT services can last much longer. Thus, the important role of analysis in this area is to identify the important drivers of ICT indirect energy impacts and gather data on actual, rather than potential, energy savings, so that these results can inform both public policy and private decision making on the implementation and use of ICT. For this reason, the field would benefit from more focus on empirical case studies and on understanding the behavioral aspects of how various stakeholders use ICT services in practice. Additional work on characterizing uncertainty in energy effect estimates would also benefit discussions in this area.

# 6 Future Work

Estimating U.S. data center energy use requires developing inputs and assumptions for an industry with rapidly evolving technologies and limited publically available energy use data, which ultimately limits the potential scope of analysis. Through the challenges of developing data center energy use estimates for this report, additional areas of research were identified that could improve future growth estimates in data center energy consumption and the potential impacts for specific efficiency efforts. Below are key areas identified that warrant future research.

## 6.1 Server Utilization and Power Proportionality

Due to the limited data on utilization rates for servers in U.S. data centers, this study generalizes server utilization using a single average (per space type), with no information about the actual distribution of utilization over the average year. This generalization prevents distinguishing between a server that is run at 40% utilization constantly and one that is run at 80% utilization half of the time and idled the other half of the time. Additionally, while this study assumes a linear relationship between utilization and power consumption (i.e., the "scaling curve") of servers, the actual relationship is generally nonlinear[18] and therefore there is a loss of accuracy in modeling server energy use at the average utilization level. Without the ability to model nonlinear scaling curves, it then becomes difficult to understand the impacts of certain

efficiency measures, such as the targeted lowering of idle state power consumption (e.g. Emerson Network Power's proposed 10 Minus standard[102]) or powering down servers during idle times. As consolidation efforts like virtualization are increasing server utilization levels across the industry, it is even more important to understand how future energy savings opportunities associated with increasing dynamic ranges and low-power idle states are going to be affected.

## 6.2 Workload Variation

In addition to understanding server utilization over time, further efforts should be made to understand the distribution of various types of server workloads and their associated hardware requirements. This will help to quantify opportunities for energy savings associated with optimizing hardware for specific workloads as opposed to using a "one-size-fits-all" approach prevalent across large data centers today. This could include using single-socket designs and workload-optimized processors (RISC, FPGA, GPUs, etc.). Understanding workload variations among different servers within data centers would also assist in identifying strategies to further increase overall data center utilization loads and increase cooling efficiency by creating the opportunity to provide server-specific cooling demand.

## 6.3 Barriers to Hyperscale Shift

While there has been significant growth in hyperscale data centers, this report shows that a significant portion of servers are still expected to reside in small room or closet data centers. Better understanding the different barriers, including technical, legal, and security barriers, that are preventing movement to colocation or to the cloud can help drive solutions that increase the shift to large data centers and tailor energy efficiency strategies for the small data centers that remain.

## 6.4 Beyond PUE

There is a need in the data center industry for performance metrics that better capture the efficiency of a given data center. The limitations of PUE, the most commonly discussed metric of efficiency, are generally understood,[103] but a key issue it that PUE only measures the efficiency of the building infrastructure supporting a given data center and indicates nothing about the efficiency of the IT equipment itself. Metrics that capture the functionality of the data center (e.g. amount of computations it performs) and relate that to energy use can help industry better understand where progress has been made and where there are opportunities to reduce energy use. Initial efforts to accomplish this include The Green Grid's Data Center Productivity (DCP) and Data Center energy Productivity (DCeP),[104] the Uptime Institute and McKinsey's Corporate Average Data center Efficiency (CADE),[105] and JouleX's Performance per Watt (PPW),[106] but PUE is still the dominant metric broadly observed in the data center industry.

## 6.5 Beyond 2020

The significant energy efficiency improvements in the design and operation of data centers over the past decade have allowed U.S. data center energy use to remain nearly constant while

simultaneously meeting a drastic increase in demand for data center services. However, the data available at the time of this study limited the scope of future projection to 2020. The key efficiency strategies identified in this report, improved PUE, increased server utilization rates, and better power proportionality all have theoretical and practical limits and the current rate of improvement indicates that these limits may be reached in the not too distant future. The potential for data center services, especially from a global prospective, are still in a fairly nascent stage and future demand could continue to increase after our current strategies to improve energy efficiency have been maximized. Understanding if, and when, this transition may occur, and the ways in which data centers can minimize their costs and environmental impacts under such a scenario, is an important direction for future research. This report highlights the success of the data center industry to stabilize electricity demand, but further investigation and technological breakthroughs in energy efficiency across the ICT equipment spectrum will be needed to insure that success is not simply a plateau before an increase in electricity demand resumes at a rate proportional to future growth of data center services.

# References


[1] Brown, R., Masanet, E., Nordman, B., Tschudi, W., Shehabi, A., Stanley, J., Koomey, J., Sartor, D., Chan, P., Loper, J., Capana, S., Hedman, B., Duff, R., Haines, E., Sass, D., and A. Fanara. (2007). *Report to Congress on Server and Data Center Energy Efficiency: Public Law 109-431.* Lawrence Berkeley National Laboratory, Berkeley, California. LBNL-363E.

[2] Masanet, E, Brown, R E, Shehabi, A, Koomey, J G, and B Nordman (2011). "*Estimating the Energy Use and Efficiency Potential of U.S. Data Centers. Proceedings of the IEEE*, Volume 99, Number 8.

[3] Koomey, J.G. (2011). *Growth in Data Center Electricity Use 2005 to 2010*. Analytics Press, Oakland, California. http://www.analyticspress.com/datacenters.html

[4] Upton, F (2015). *North American Energy Security and Infrastructure Act of 2015. H.R. 8, 114th Congress.* https://www.congress.gov/bill/114th-congress/house-bill/8

[5] Koomey, Jonathan. 2008. Worldwide electricity used in data centers. *Environmental Research Letters*. vol. 3, no. 034008. September 23. [http://stacks.iop.org/1748-9326/3/034008]

[6] Koomey, Jonathan G. 2007. *Estimating Total Power Consumption by Servers in the U.S. and the World.* February 15. http://www.mediafire.com/file/exywo1hf6ionskw/AMDserverpowerusecomplete-final.pdf.

[7] IDC Worldwide Quarterly Server Tracker: http://www.idc.com/tracker/showproductinfo.jsp?prod_id=7

[8] International Data Corporation (IDC). 2015. *IDC's Worldwide Quarterly Server Shipment Tracker, 2010-2018*, Framingham, MA, March.

[9] International Data Corporation (IDC). 2015. *IDC's Worldwide Quarterly Disk Storage Systems Tracker, 2010-2019*, Framingham, MA, March.

[10] International Data Corporation (IDC). 2015. *IDC's Worldwide Quarterly Data Center Networks, 2008-2019*, Framingham, MA, March.

[11] International Data Corporation (IDC). 2014*. IDC's Worldwide Quarterly Server Tracker – Installed Base, 2006-2018.* Framingham, MA: IDC. December.

[12] Dietrich, J. *2014 ITIC Analysis of SERT Worklet Results*. Information Technology Industry Council. Available at: https://www.energystar.gov/sites/default/files/specs/ITI%20Analysis%20of%20SERT%20Data.pdf

[13] International Data Corporation (IDC). 2015. Personal communication with Lidice Fernandez, Program Vice President, WorldWide Tracker Research, September 9.

[14] The Green Grid, Data Centre Life Cycle Assessment Guidelines. *White Paper #45, v2, 2012*. Available at: http://www.thegreengrid.org/~/media/WhitePapers/WP45v2DataCentreLifeCycleAssessmentGuidelines.pdf

[15] SPEC (2015). *SPECpower_ssj2008 Results.* Downloaded September 10, 2015 from https://www.spec.org/power_ssj2008/results/

[16] Van Heddeghem, W., Lambert, S., Lannoo, B., Colle, D., Pickavet, M., & Demeester, P. (2014). *Trends in worldwide ICT electricity consumption from 2007 to 2012. Computer Communications*, *50*, 64-76.

[17] NRDC and WSP (2012). *The Carbon Emissions of Server Computing for Small- to Medium-Sized Organizations: A Performance Study of On-Premise vs.The Cloud.* WSP Environment & Energy, LLC and Natural Resources Defense Council. October 2012.

[18] Barroso, L. A., Clidaras, J., & Hölzle, U. (2013). *The datacenter as a computer: An introduction to the design of warehouse-scale machines.* Synthesis lectures on computer architecture, Morgan Claypool Publishers

[19] Hylick, A., Ripduman, S., Rice, A., Jones, B. 2008. *An Analysis of Hard Drive Energy Consumption.* University of Cambridge, Computer Laboratory, October. Available at: https://www.cl.cam.ac.uk/~acr31/pubs/hylick-harddrive2.pdf

[20] ASHRAE (2015). *Data Center Storage Equipment – Thermal Guidelines, Issues, and Best Practices*. Technical Committee 9.9.

[21] Cadmus (2015). *New York State Data Center Market Characterization*. New York State Energy Research and Development Authority (NYSERDA) Report Number 15-06. October 2015.

[22] Reinsel, David (2010). *A Plateau in Sight for the Rising Costs to Power and Cool the World's External Storage?* IDC Opinion, IDC#225016. September 2010.

[23] Lanzisera, S., Nordman, B., & Brown, R. E. (2012). *Data network equipment energy use and savings potential in buildings.* Energy Efficiency, 5(2), 149-162.

[24] Reviriego, P., Maestro, J. A., & Larrabeiti, D. (2010). *Burst transmission for energy-efficient ethernet.* Internet Computing, IEEE, 14(4), 50-57.

[25] Dudkowski, D., Hasselmeyer, P*. Energy-Efficient Networking in Modern Data Centers*. In Konstantinos Samdanis, Peter Rost, Andreas Maeder, Michaela Meo, Christos Verikoukis: Green Communications: Principles, Concepts and Practice, John Wiley & Sons, 2015. ISBN 978-1-118-75926-4

[26] 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.

[27] Villars, Richard L (2014). *"U.S. Datacenter Census and Construction 2014-2018 Forecast: Realigning Workloads, Managing Obsolescence, and Leveraging Hyperscale*". IDC #252712.

[28] Venture Outsource, 2015. *ODM Quanta focus on white boxes bypasses Dell, HP, Cisco traditional model for OEMs.* VentureOutsource.com: https://www.ventureoutsource.com/contract-manufacturing/focus-odm-quanta-it-shift-cloud-infrastructure-leaving-dell-hp-traditional

[29] IDC (2014). *The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things*. April 2014, sponsored by EMC. http://idcdocserv.com/1678.

[30] International Data Corporation (IDC). 2015. Personal communication with Rick Villars, Program Vice President, Data Center and Cloud Research, November 12.

[31] *The Green Grid datacenter power efficiency metrics: PUE and DCiE.* Available at http://www.thegreengrid.org/sitecore/content/Global/Content/white-papers/The-Green-Grid-Data-Center-Power-Efficiency-Metrics-PUE-and-DCiE.aspx.

[32] 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.

[33] 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.

[34] C. Malone and C. Belady. *Metrics to characterize datacenter & IT equipment energy use.* In Proceedings of the Digital Power Forum, Richardson, TX, September 2006.

[35] Sullivan, A. 2010. *Energy Star for Data Centers*. Green Grid Forum. February 4, 2010

[36] Uptime Institute, *Important to recognize the dramatic improvement in data center efficiency*.http://blog.uptimeinstitute.com/2012/09/important-to-recognize-the-dramatic-improvement-in-data-center-efficiency/

[37] Cheung, I. H., Greenberg, S., Mahdavi, R., Brown, R., & Tschudi, W. (2014, August). *Energy Efficiency in Small Server Rooms: Field Surveys and Findings.* Proceedings the 2014 ACEEE Summer Study on Energy Efficiency in Buildings. LBNL- 6952E

[38] Google 2015. "*Our energy-saving data centers*." Accessed December 2, 2015. http://www.google.com/about/datacenters/efficiency/internal/index.html#measuring-efficiency

[39] Gelber, R. 2012. "*Facebook showcases green datacenter." HPCwire*. Accessed December 2, 2015. http://www.hpcwire.com/hpcwire/2012-04-26/facebook_showcases_green_datacenter.html

[40] Masanet, E., Shehabi, A., Ramakrishnan, L., Liang, J., Ma, X., Walker, B., & Mantha, P. (2013). *The Energy Efficiency Potential of Cloud-Based Software: A US Case Study*. Lawrence Berkeley National Laboratory, Berkeley, California.

[41] Shehabi, A., Masanet, E., Price, H., Traber, K., Horvath, A., and W.W. Nazaroff. 2011. "*Data Center Design and Location: Consequences for Electricity Use and Greenhouse-Gas Emissions.*" *Building and Environment*, Volume 46, Issue 5.

[42] Koomey, J.G., Berard, S., Sanchez, M. and Wong, H., 2011. *Implications of historical trends in the electrical efficiency of computing*. *Annals of the History of Computing, IEEE*, *33*(3), pp.46-54.

[43] Koomey, Jonathan, and Samuel Naffziger. 2015. "*Efficiency's brief reprieve: Moore's Law slowdown hits performance more than energy efficiency*." In IEEE Spectrum. April. pp. [http://spectrum.ieee.org/computing/hardware/moores-law-might-be-slowing-down-but-not-energy-efficiency]

[44] The Green Grid. 2011. *Water Use Effectiveness: A Green Grid Data Center Sustainability Metric. White Paper #35.*

[45] P. Torcellini, N. Long, and R. Judkoff. *Consumptive Water Use for U.S. Power Production*. 2003, NREL/TP-550- 33905, http://www.nrel.gov/docs/fy04osti/33905.pdf

[46] Miller, R. 2009. *Data Centers Move to Cut Water Waste. Data Center Knowledge*. Available at: http://www.datacenterknowledge.com/archives/2009/04/09/data-centers-move-to-cut-water-waste

[47] Fitzgerald, D. 2015. *Data Centers and Hidden Water Use.* The Wall Street Journal. Available at: http://www.wsj.com/articles/SB10007111583511843695404581067903126039290

[48] Federal Data Center Consolidation Initiative (FDCCI), 2014. *GSA FDCCI Inventory Data Fields.* Available at: https://cio.gov/wp-content/uploads/downloads/2012/10/FDCCI-Inventory-Data-Fields-Release-2.pdf. Accessed: March 25, 2014.

[49] Boyd, A. 2015. *Experts: Data center consolidation goals not aggressive enough*. Federal Times.com. Available at: http://www.federaltimes.com/story/government/it/data-center/2015/02/17/data-center-consolidation-goals-not-aggressive/23556995/

[50] Top500. 2016. *Top500 List – November 2015.* http://www.top500.org/lists/2015/11/

[51] Kaplan, JM., Forrest, W., Kindler, N. 2008. *Revolutionizing Data Center Efficiency McKinsey and Company.* http://www.mckinsey.com/clientservice/bto/pointofview/pdf/revolutionizing_data_center_efficiency.pdf

[52] *The Uptime Institute estimate*; https://uptimeinstitute.com/research-publications/asset/comatose-server-savings-calculator

[53] Koomey, J., Taylor J. 2015. *New data supports finding that 30 percent of servers are 'Comatose' , indicating that nearly a third of capital in enterprise data centers is wasted.* http://anthesisgroup.com/wp-content/uploads/2015/06/Case-Study_DataSupports30PercentComatoseEstimate-FINAL_06032015.pdf

[54] McMillian, R. 2015. *Zombie Servers: They're Here and Doing Nothing but Burning Energy.* The Wall Street Journal. September 13, 2015. http://www.wsj.com/articles/zombie-servers-theyre-here-and-doing-nothing-but-burning-energy-1442197727

[55] Delforge, P., & Whitney, J. 2014. *Issue Paper: Data Center Efficiency Assessment Scaling up Energy Efficiency Across the Data Center Industry: Evaluating Key Drivers and Barriers*. *Natural Resources Defense Council (NRDC).*

[56] Ourghanlian, B. 2010. *Improving Energy Efficiency: And End User Perspective*. The Green Grid. The Green Grid Technical Form 2010. Available at: http://www.thegreengrid.org/~/media/EMEATechForums2010/Improving%20Energy%20Efficiency%20-%20An%20End%20User%20Perspective_Paris.pdf?lang=en

[57] United EIF, Inc. 2015*. Advanced Configuration and Power Interface Specification*. Available at: http://www.uefi.org/sites/default/files/resources/ACPI_6.0.pdf

[58] Lo, D., Cheng, L., Govindaraju, R., Ranganathan, P., & Kozyrakis, C. (2015, June). *Heracles: improving resource efficiency at scale.* In Proceedings of the 42nd Annual International Symposium on Computer Architecture (pp. 450-462). ACM.

[59] Masanet, E., Shehabi, A., Ramakrishnan, L., Liang, J., Ma, X., Walker, B., Hendrix, V., and P. Mantha (2013). *The Energy Efficiency Potential of Cloud-Based Software: A U.S. Case Study*. Lawrence Berkeley National Laboratory, Berkeley, California.

[60] Heller, B., Seetharaman, S., Mahadevan, P., Yiakoumis, Y., Sharma, P., Banerjee, S., & McKeown, N. (2010, April*). ElasticTree: Saving Energy in Data Center Networks*. In NSDI (Vol. 10, pp. 249-264).

[61] Masanet, E., A. Shehabi, L. Ramakrishnan, J. Liang, X. Ma, B. Walker, V. Hendrix, and P. Mantha (2013). *The Energy Efficiency Potential of Cloud-Based Software: A U.S. Case Study.* Lawrence Berkeley National Laboratory, Berkeley, California.

[62] Hilty, L.M., Aebischer, B. (Eds.), 2015. *ICT Innovations for Sustainability, Advances in Intelligent Systems and Computing.* Springer International Publishing, Cham.

[63] Masanet, E.R., Matthews, H.S. (Eds.), 2010. *Environmental Applications of Information and Communication Technology* [special issue]. J. Ind. Ecol. 14, 685–862.

[64] Rejeski, D. (Ed.), 2002. *E-Commerce, the Internet, and the Environment* [special issue]. J. Ind. Ecol. 6, 1–161.

[65] Roberts, S., 2009. *Measuring the Relationship between ICT and the Environment*. Organization for Economic Co-operation and Development. http://www.oecd.org/sti/43539507.pdf

[66] Auweter, D. Kranzlmüller, A. Tahamtan, A. M. Tjoa, D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, and G. Weikum, Eds., *ICT as Key Technology against Global Warming*, ser. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, vol. 7453, ISBN: 978-3- 642-32605-9. [Online]. Available: http://link.springer.com/10.1007/978-3-642-32606-6

[67] Erdmann, L., Hilty, L., Goodman, J., Arnfalk, P., 2004. *The Future Impact of ICTs on Environmental Sustainability (Technical Report No. EUR 21384 EN).* Institute for Prospective Technological Studies, Seville, Spain.

[68] Jørgensen, M.S., Andersen, M.M., Hansen, A., Wenzel, H., Thoning, T., Pedersen, U.J., Falch, M., Rasmussen, B., Olsen, S.I., Willum, O., 2006. *Green Technology Foresight about environmentally friendly products and materials.* Environ. Prot. Agency Den. http://www2.mst.dk/udgiv/publications/2006/87-7052-216-2/pdf/87-7052-217-0.pdf.

[69] Horner, N., Shehabi, A., Azevedo, I. *Known unknowns: Indirect energy effects of information and communication technology.* (In Review)

[70] Romm, J., Rosenfeld, A., Herrmann, S., 1999. *The internet economy and global warming: A scenario of the impact of e-commerce on energy and the environment*. Cent. Energy Clim. Solut. Dec. Httpwww Cool-Co. Orgenergycecs Cfm.

[71] Laitner, J.A., Ehrhardt-Martinez, K., 2008. *Information and Communication Technologies: The Power of Productivity (E081).* American Council for an Energy-Efficient Economy, Washington, D.C.

[72] Elliott N, Molina, M., and Trombley, D. 2012, *A defining framework for intelligent efficiency* (Washington, DC: American Council for an Energy-Efficient Economy) Online: http://aceee.org/sites/default/files/publications/researchreports/e125.pdf

[73] Accenture. 2015. *SMARTer2030: ICT Solutions for 21st Century Challenges* Online: http://smarter2030.gesi.org/downloads/Full_report2.pdf

[74] Rattle, R., 2010. *Computing Our Way to Paradise?: The Role of Internet and Communication Technologies in Sustainable Consumption and Globalization.* Rowman & Littlefield.

[75] Berkhout, F., Hertin, J., 2004. *De-materialising and re-materialising: digital technologies and the environment.* Futures 36, 903–920. doi:10.1016/j.futures.2004.01.003

[76] Börjesson Rivera, M., Håkansson, C., Svenfelt, Å., Finnveden, G., 2014*. Including second order effects in environmental assessments of ICT.* Environ. Model. Softw. 56, 105–115. doi:10.1016/j.envsoft.2014.02.005

[77] Erdmann, L., Hilty, L.M., 2010. *Scenario Analysis: Exploring the Macroeconomic Impacts of Information and Communication Technologies on Greenhouse Gas Emissions*. J. Ind. Ecol. 14, 826–843. doi:10.1111/j.1530-9290.2010.00277.x

[78] Koomey, J.G., Matthews, H.S., Williams, E., 2013. *Smart Everything: Will Intelligent Systems Reduce Resource Use?* Annu. Rev. Environ. Resour. 38, 311–343. doi:10.1146/annurev-environ-021512-110549

[79] Yi, L., Thomas, H.R., 2007. *A review of research on the environmental impact of e-business and ICT.* Environ. Int. 33, 841–849. doi:10.1016/j.envint.2007.03.015

[80] Allenby, B., Unger, D., 2001. *Information technology impacts on the US energy demand profile*. RAND Corp.

[81] Azevedo, I.M.L., 2014. *Consumer End-Use Energy Efficiency and Rebound Effects*. Annu. Rev. Environ. Resour. 39, 393–418. doi:10.1146/annurev-environ-021913-153558

[82] Gillingham, K., Rapson, D., Wagner, G., 2015. *The rebound effect and energy efficiency policy.*

[83] Borenstein, S., 2013. *A microeconomic framework for evaluating energy efficiency rebound and some implications.* National Bureau of Economic Research.

[84] Hilty, L.M., Arnfalk, P., Erdmann, L., Goodman, J., Lehmann, M., Wäger, P.A., 2006. *The relevance of information and communication technologies for environmental sustainability – A prospective simulation study.* Environ. Model. Softw. 21, 1618–1629. doi:10.1016/j.envsoft.2006.05.007

[85] Hesse, M., 2002. *Shipping news: the implications of electronic commerce for logistics and freight transport.* Resour. Conserv. Recycl. 36, 211–240.

[86] Harrington, D., 2015. *From first mile to last mile: Global industrial & logistics trends*. Colliers International.

[87] Shorr Packaging Corp, 2015. *The Amazon Effect: Impacts on Shipping and Retail* Available at: http://www.shorr.com/packaging-news/2015-06/amazon-effect-impacts-shipping-and-retail (accessed 11.30.15).

[88] Mohan, A.M., 2014. *E-commerce packaging pitfalls & opportunities.* Packag. World.

[89] Greening, L.A., Greene, D.L., Difiglio, C., 2000. *Energy efficiency and consumption -- the rebound effect -- a survey.* Energy Policy 28, 389–401.

[90] Plepys, A., 2002. *The grey side of ICT*. Environ. Impact Assess. Rev. 22, 509–523.

[91] Williams, E., 2011*. Environmental effects of information and communications technologies*. Nature 479, 354–358. doi:10.1038/nature10682

[92] Weber, C. *et al. Life cycle comparison of traditional retail and e-commerce logistics for electronic products: A case study of buy.com*. *Green Des. Inst. Carnegie Mellon Univ.* (2008). at http://www.ce.cmu.edu/~greendesign/research/Buy_com_report_final_030209.pdf

[93] Shehabi, A., Walker, B. & Masanet, E. *The energy and greenhouse-gas implications of internet video streaming in the United States. Environ. Res. Lett.* **9,** 054007 (2014).

[94] Kitou, E. & Horvath, A. *External air pollution costs of telework. Int. J. Life Cycle Assess.* **13,** 155–165 (2008).

[95] Henderson, P. & Waitner, M. *Real-time energy management: A case study of three large commercial buildings in Washington, D.C.* (Natural Resources Defense Council, 2013). at http://www.nrdc.org/business/casestudies/files/tower-companies-case-study.pdf

[96] Seidel, S. and Ye, J. 2012. *Leading by example: using information and communication technologies to achieve Federal sustainability goals* (Center for Climate and Energy Solutions) Online: http://www.c2es.org/docUploads/federal-sustainability-ict.pdf

[97] Laitner, J. A. & Ehrhardt-Martinez, K. *Information and Communication Technologies: The Power of Productivity*. (American Council for an Energy-Efficient Economy, 2008). at http://aceee.org/sites/default/files/publications/researchreports/E081.pdf

[98] Baer, W. S., Hassell, S. & Vollaard, B. A. *Electricity requirements for a digital society*. (RAND Corporation, 2002).

[99] Hilty, L. M. *et al. The relevance of information and communication technologies for environmental sustainability – A prospective simulation study*. *Environ. Model. Softw.* **21,** 1618–1629 (2006).

[100] Romm, J., Rosenfeld, A. & Herrmann, S. *The internet economy and global warming: A scenario of the impact of e-commerce on energy and the environment. Cent. Energy Clim. Solut. Dec. Httpwww Cool-Co. Orgenergycecs Cfm* (1999). at http://www.fraw.org.uk/files/tech/romm_1999.pdf

[101] Laitner, J.A., McDonnell, M.T., and Keller, R.M. 2015. *ICT-Enabled Intelligent Efficiency: Shifting from device-specific approaches to system optima.* Online: http://cda.iea-4e.org/document/11/ict-enabled-intelligent-efficiency-shifting-from-device-specific-approaches-to-system-optima

[102] Pouchet, J. 2016. *Time for a Server Idle Performance Standart – Introducing 10 Minus*. Emersion Network Power Blog. Available at: http://blog.emersonnetworkpower.com/efficiency/time-for-a-server-idle-performance-standard/

[103] Horner, N., Azevedo, I. 2016. *Power usage effectiveness in data centers: overloaded and underachieving*. The Electricity Journal, Volume 29, Issue 4, May 2016, Pages 61-69

[104] Azevedo, D., Rawson, A. 2008. *Measuring Data Center Productivity. Metrics and Measurements Work Group.* The Green Grid. Available at: http://www.thegreengrid.org/~/media/TechForumPresentations2008/Measuring_Data_Center_Productivity.pdf?lang=en

[105] Kaplan, J., Forrest, W., Kindler, N. 2008. *Revolutionizing Data Center Energy Efficiency*. McKinsey & Company. Available at: http://www.sallan.org/pdf-docs/McKinsey_Data_Center_Efficiency.pdf

[106] Cappiello, C., Chen, D., Ferreria, A.M., Henis, R., Jiang, T., Kat, I., Kipp, A., Liu, J., Sotnikov, D., Vitali, M. 2011. *Usage Centric Green Performance Indicators.* Available at: http://www.sigmetrics.org/sigmetrics2011/greenmetrics/green11_Kat_email.pdf