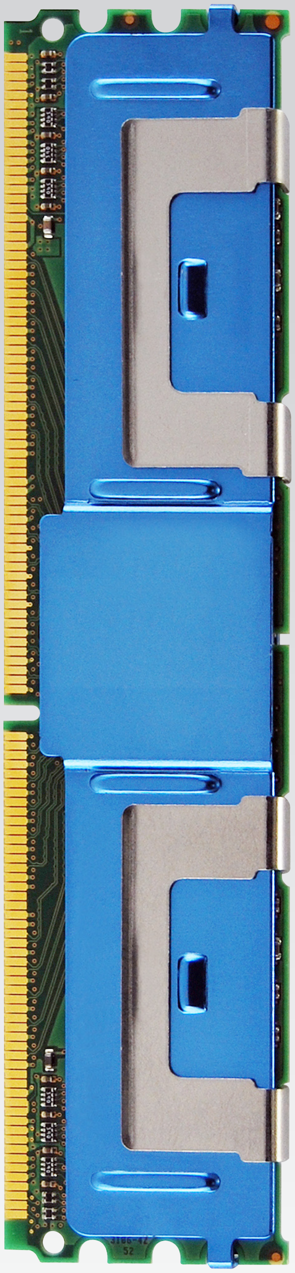


Summary

This white paper addresses the energy consumption of DRAM in computing applications and the opportunities to maximize energy savings by targeting more efficient products for data center servers. Micron estimates module power savings at 24 percent; this has the potential to achieve energy savings of 5.5 billion kilowatt hours (kWh) on a global basis annually. At typical industrial power costs1 ($0.06 per kWh), the savings are more than $300 million per year.

Michael Sporer

Regional Sales Manager, Micron Memory Products Group

The data centers' incremental overhead power consumption due to inefficiencies and cooling is estimated to be equal to the amount that is consumed by servers, storage, and networking

Introduction

The U.S. EPA Energy Star program is conducting a study to assess opportunities for energy efficiency improvements to computer servers and data centers[2]. This is in response to Public Law 109-431, which was passed and signed into law December 20, 20063. This legislation requires an investigation down to the microchip level. As a manufacturer of semiconductor memory products used in server systems, Micron intends to proactively address these opportunities.

Justification

A recent study conducted by Dr. Jonathan Koomey[4] with the Lawrence Berkeley National Laboratory (LBL) and funded by Advanced Micro Devices (AMD), illustrated the significant and growing energy use by data centers. Data centers are rooms, floors or sometimes entire buildings that house computer, storage, and networking equipment. Data centers can serve up Web pages, stream media, enable Internet access, and run simulations of any kind of research. They can also provide computing power for traditional and private uses like banking or other financial transactions.

The computers in data centers, called servers, are similar to PCs in that they have the same basic microchips—the CPU and memory. Unlike PCs, servers in data centers are packed together as densely as possible and use substantial amounts of electricity, the majority of which ends up in the form of heat, which then must be removed from the servers. The power delivery to the systems is provided through uninterruptible power supplies (UPS) that are not 100 percent efficient and also produce copious amounts of heat as well. The heat must be carefully and continuously managed to keep the systems running within their specified operating temperature and humidity ranges. Regardless of the type and efficiency of the cooling system, the heat must be removed from the data center in one way or another. To do so requires additional energy be used to operate the cooling infrastructure.

Introduction

The data centers' incremental overhead power consumption due to inefficiencies and cooling is estimated to be equal to the amount that is consumed by servers, storage, and networking. The user of a single PC, workstation, or laptop doesn't see system heat generation ias a concern, but for data centers, managing the overhead is as important as the servers themselves. If system power is reduced, then the available overhead can handle a greater IT load and perform more useful work in the same power envelope.

The Role Memory Plays in the Challenges of Servers and Energy Use

The memory content in servers has been growing at a rapid pace and is expected to continue to do so for a variety of reasons. In general, software with more functionality requires both greater computational ability as well as a larger memory footprint. However, some factors are more applicable to servers than PCs.

The memory content in servers has been growing at a rapid pace and is expected to continue to do so for a variety of reasons

First is the proliferation of multi-core CPUs executing single-threaded applications. Each thread requires its own memory space, therefore doubling the number of CPU cores requires doubling the memory. A recent seminar[5] on server design cited this rule of thumb:

1GB per (1 GHz × # cores)

This equation reinforces the idea that each CPU core mandates an increase in memory space. Another factor driving memory content growth and server power consumption is the adoption of virtualization technologies. A server running a virtualized environment is able to achieve a higher utilization which, in turn, increases the total power consumption of the server. Once again, the importance of energy efficient component selection increases. By analogy, a car will burn very little fuel if it isn't driven. Virtualization, or anything that increases server operation, is like adding a new driver to the mix. Now the car gets driven more and energy efficiency becomes a greater concern.

Traditionally, the CPU has been the component that consumes the most power in the system. Improvements in CPU power consumption now place a greater scrutiny on the other components. Where memory once played a distant second to CPUs in the ranking of system power consumption, now, in some cases, it exceeds the power consumption of the CPU.

The combined savings of these two technologies—1Gb-based, DDR2 reduced chip count modules and 1.5V DDR2 SDRAM devices—is estimated to be 24 percent of the memory DIMM power consumption

Energy Efficient Memory Advantages

Micron's new energy efficient Aspen Memory® product line includes several new products that have a lower power consumption compared to legacy standard products. These technologies are intended for use in both client machines—PCs, laptops, workstations—as well as in servers.

The new products are 1Gb-based, DDR2 reduced chip count (RCC) modules; and 1.5V DDR2 FBDIMMs. The 1.5V DDR2 SDRAM operates at 1.5V instead of 1.8V. The 1Gb-based, DDR2 RCC modules provide the same memory capacity and performance as a DIMM built using legacy (currently 512Mb) devices, but use half as many higher density (1Gb) DRAM devices (see Appendix A for product details). The combined savings of these two technologies is estimated to be 24 percent of the memory DIMM power consumption. For reasons previously cited, extrapolating the savings to the system level is difficult; instead, we have directly measured power savings at the power supply input of the server under test.

Under minimum and maximum loading conditions and using commercially available systems, Micron has measured between 1.5 to 1.8 watts per 2GB DIMM improvement in power consumption when comparing standard legacy products to 1Gb-based, DDR2 reduced chip count DIMMs.

Measurements made in a lab environment using modified commercial hardware with the adaptations required to support 1.5V DDR2 on FBDIMMs show power reductions in the 1.5 to 2.0 watts per 4GB 1.5V DIMM attributed only to the DRAM. Additional power savings could be possible using a low-voltage advanced memory buffer (AMB) chip, which is also on the FBDIMM.

For the purpose of the analyses that follow, it is necessary to convert these savings into a percentage basis. We will assume a conservative 24 percent DIMM-level savings for the 1Gb-based, DDR2 RCC DIMM and 1.5V DDR2 SDRAM.

Data Centers and Energy Use

According to Dr. Koomey's report, data center servers consumed 616 billion kWh worldwide in 2005. The historical growth rate of this figure has been 15 percent annually from the year 2000 to 2005.

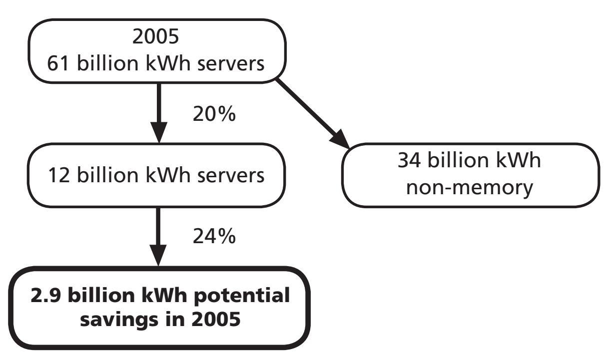
Estimating the power consumption attributed to memory is a difficult challenge. A computer system has multiple memory sockets that can be fully or partially populated with memory modules, and the memory module density can also vary. In addition to these physical variations, the portion of power attributed to memory also depends on the type of workload and memory utilization. Certainly, further study is needed in this area. For the purpose of this paper we are going to assume memory accounts for 20 percent[7] of the total system power budget in a server. The reader can adjust this assumption as needed.

Micron has come up with an alternative method for estimating DRAM power consumption (summarized here; details in Appendix C). This method incorporates analysts' data to estimate the total DRAM production in a given year and the DRAM consumption by market segment. This method also makes assumptions regarding hours of operation as well as system utilization to estimate power consumption and potential savings. (Figure 2 provides the estimate for calendar year 2008.)

The production of DRAM is quantified in terms of 512Mb equivalent units. For example, a single 1Gb DRAM is equivalent to two 512Mb devices. First, we divide the market into three categories: server, client, and other. The client-machine category includes desktops, laptops, and workstations. The other category is a catch-all for non-computing markets and is not considered in this analysis. Given these market segments, it's apparent that the client market consumes four times as many DRAM equivalent units as the server market.

Next, we consider the hours of operation and system utilization on an annual basis. Servers operate 24 hours a day, 7 days a week; client machines operate approximately 8 hours a day, 5 days a week. Server utilization is assumed to be 15 percent; client, 5 percent. By applying the usage model to each DRAM market, we conclude that despite the four-to-one difference in shipments, the DRAM in servers consumes more power than all DRAM in the client machines.

Estimation of Memory Power Consumption and Potential Savings



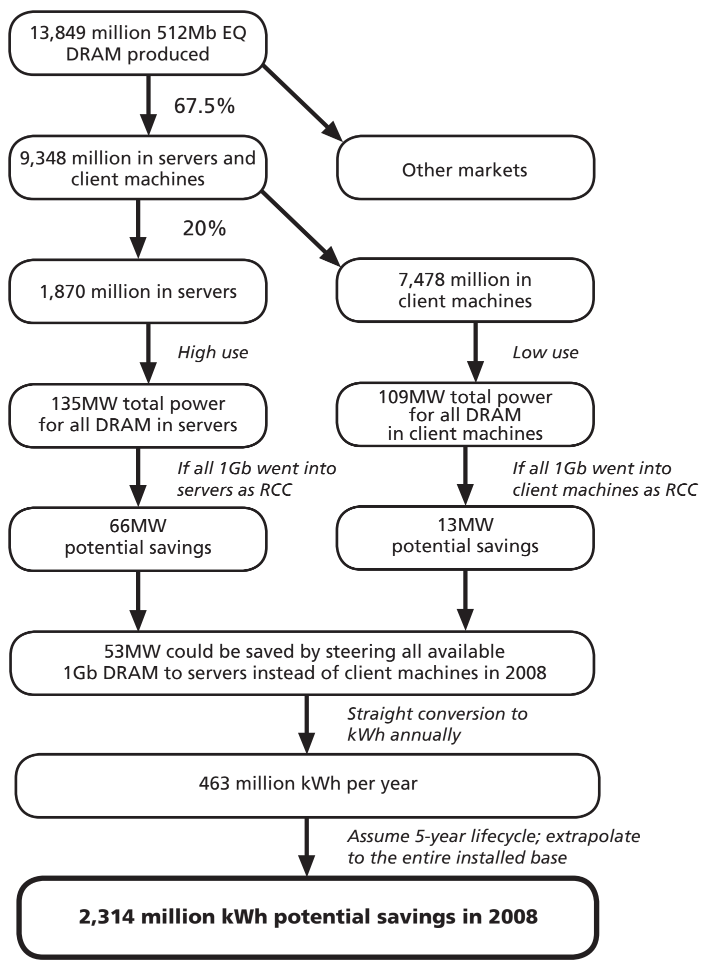
Illustrates the potential savings. For a detailed description, see Appendix B.

Next, we consider the total available 1Gb DRAM which could be used to build the 1Gb-based, DDR2 RCC DIMMs (see Appendix A for RCC details). For 2008, all 1Gb DRAM production is estimated to be enough to provide for 98 percent of the total demand for servers or 24 percent of the total demand for client machines. Finally, we look at the potential power savings for all the 1Gb-based, DDR2 RCC DIMMs if they were installed into either client machines or servers.

When we analyze this power-savings comparison, we see that although the DIMM power requirements and potential savings are identical in either application, the cumulative energy savings is substantial for the server market due the longer hours of operation and higher utilization factors of server platforms. In this example, 462 million kWh energy savings would be achieved for energy efficient DRAM devices sold this year. This approach would conserve 2.3 billion kWh over a five-year product lifecycle.

Savings in servers are much greater because, unlike client machines, servers are always running

Estimation of power conusmption based on annual DRAM manufacturing and market usage

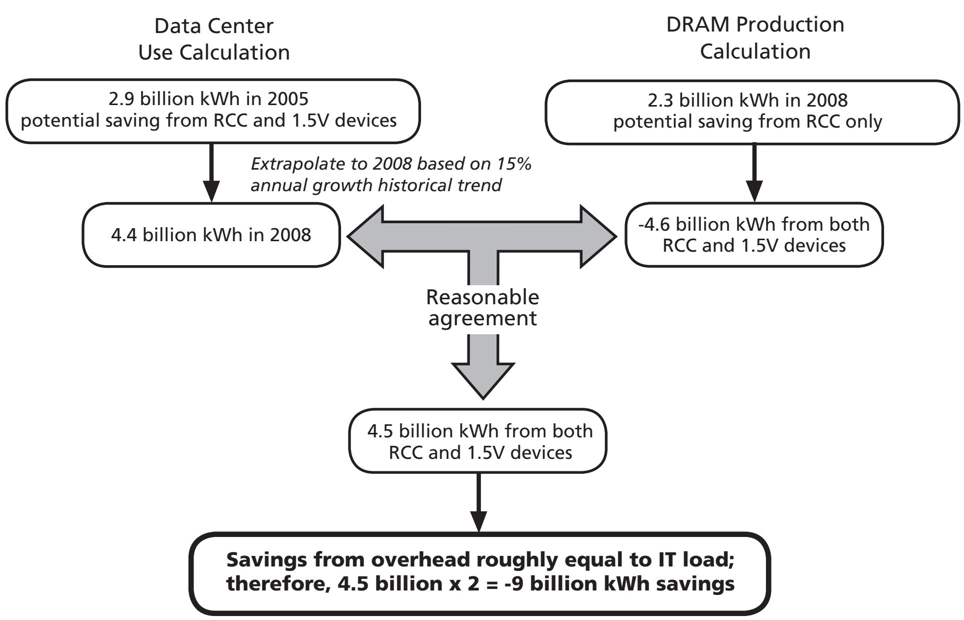


Comparing the Two Methods

Let's attempt to correlate the two estimation methods. First we need to look at the differences so we can compensate accordingly. The first method is based on data from 2005 and assumes both 1Gb-based, DDR2 RCC DIMMs and 1.5V devices are placed into the installed base. The second method is based on data for 2008 and assumes only 1Gb-based, DDR2 RCC DIMMs are used (see figure below).

To extrapolate from 2005 to 2008, we will assume a 15 percent annual growth rate consistent with the previous five years. As a first approximation we will assume that the savings from 1.5V DDR2 devices and 1Gb-based, DDR2 RCC modules are equal. As shown below, both methods demonstrate 4.5 billion kWh annual potential savings for DRAM in servers. The aggregate data center energy savings would be doubled when considering the incremental overhead and cooling energy costs. Alternately, instead of reducing power consumption, these savings could be used to support more IT equipment within the existing infrastructure, resulting in better asset utilization and deferring the need for new data center construction.

Comparing the Two Methods for Estimating Energy Savings



The greatest opportunity for power savings is in data center applications due to servers' high utilization

Conclusion

This paper brings together three important findings for memory with respect to energy consumption in computing applications. First, we highlight the growing memory content per server due to the increasing deployment of multi-core CPUs. We also discuss the relative importance of memory as CPUs and other sub-assemblies are being optimized for lower energy consumption. Second, we demonstrate two methods for estimating the energy consumption and potential savings of DRAM in both the general computing market and data centers. This also reinforces the idea that the greatest opportunity for power savings is in data center applications due to servers' high utilization. Finally, we show how a significant reductions in power consumption can be achieved by adopting high-density 1Gb-based, DDR2 reduced chip count modules with 1.5V DDR2 SDRAM devices.

Micron's Aspen Memory® product line features modules that are optimized for low power consumption

Appendix A: Product Availability

Micron has introduced a product line which features products that are optimized for low power consumption and have superior performance compared to standard products.

1.5V DDR2 Devices

The 1.5V DDR2, DIMMs, and motherboards that can use this technology are currently under development. Please contact Micron for the latest status.

1Gb-based, DDR2 Reduced Chip Count DIMMs

Currently, 1Gb-based, DDR2 reduced chip count modules are available for a wide range of computer applications. These 1Gb-based, DDR2 RCC modules provide the same memory capacity and performance as a DIMM built using legacy (currently 512Mb) devices while using half as many higher density (1Gb) DRAM chips.

For existing systems that can address 1Gb DRAM technology, the 1Gb-based, DDR2 RCC modules should easily work. Systems that use registered DIMMs or fully buffered DIMMs (FBDIMMs) and can support a 4GB density should be able to use 2GB reduced chip count DIMMs, which are built using the same 1Gb DRAM technology.

Some systems require DIMMs to be installed in matched pairs. For these systems, pairing a reduced chip count DIMM with a standard DIMM could reduce system performance or possibly cause the system to stop functioning. Oftentimes, a memory upgrade or a firmware or BIOS update will solve the problem. Refer to your system manufacturer for compatibility questions.

For some of the more common system questions, check the Micron® motherboard compatibility page from www.micron.com. For additional compatibility questions, refer to your system manufacturer.

TABLE 1: Quick Reference for Reduced Chip Count DIMMs

DIMM Density	Number of Chips on a DIMM	
	Standard	Reduced Chip Count
512MB	8 or 9	4 or 5
1GB	16 or 18	8 or 9
2GB	32 or 36	16 or 18

Note: Even numbers are for standard DIMMs; odd numbers are for ECC (error correction code) DIMMs.

Appendix B:
Derivation from LBL/AMD[8] White Paper

Calculating energy use and potential energy savings has not yet become a standard practice for data centers. Because of this, many calculations for determining actual energy use can be inaccurate. This situation is further complicated by the fact that power equipment efficiency is commonly calculated as the difference between power out and power in.

Power consumed by memory in servers varies significantly depending on many factors. The two primary factors are the memory capacity of the server relative to the power consumed by the rest of the system and the second factor is the actual amount of memory installed. We assume 20 percent of the power is consumed by memory.

$$\textit{61 billion} \times 0.2 = \textit{12 billion kWh}$$

By implementing 1.5V chips in reduced chip count server modules, data centers could reduce system memory power consumption by approximately 24 percent, which would be a reduction of 2.9 billion kWh.

$$\textit{12 billion} \times 0.24 = \textit{2.9 billion kWh}$$

Assuming a power cost of $0.06 per kWh a 24 percent drop in power consumption translates into an average annual savings of $0.174 billion ($174 million)[9].

2.9 billion × 0.06 = $174 million > ~$150 million

Including the savings in overhead power raises this to 5.8 billion kWh and more than ~$300 million, respectively.

Appendix C:
DRAM Energy Consumption Based on Manufacturing and Market

Another method for determining total energy consumption focuses on cumulative DRAM production and the applications into which DRAM is placed. According to market analysts although 13.5 percent of total DRAM gets placed into servers; the majority—54 percent—goes into workstations, PCs, and laptops (client machines). (The remainder goes into market segments not covered in this paper.)

A typical client machine is operated approximately eight hours a day, five days a week. Utilization of clients is typically cited at 5 percent. Many government regulatory agencies have instituted energy efficiency requirements, with EnergyStar in the United States as one such example. Given these factors, the total energy consumed by DRAM in client machines is reasonably low, even when the power switch is on throughout the day. Compare that to a typical server in a data center that is powered on twenty-four hours a day, seven days a week. Utilization of servers is typically cited at about 15 percent. Table 2 shows the difference in total DRAM power consumption between client and server machines.

TABLE 2: Use Percentages of Client Machines and Servers

	Clients	Servers	Units
Percent of DRAM market	54	13.5	%
Annual power-on hours	24	100	%
Utilization	5	15	%
DIMM Power			
Utilized	6	6	Watts per DIMM
Idle	2	2	Watts per DIMM
Sleep	0	0	Watts per DIMM
Total Power:	0.29	0.35	
Percent of total	45	55	%

Note: Client machines limited to desktop, laptop, and workstations.

In Table 2, total power equals the sum of :

Percent of DRAM market × [Power-on hours × [%Utilization × DIMM Power (Utilized) + (1-%Utilization) × DIMM Power (Idle)]]

Table 2 shows that, although client machines have four times more DRAM than servers, the total DRAM power consumed by servers is nearly equal, if not slightly higher, than power consumed by client machines. Initially, the implications might not be apparent. Of course, putting more energy efficient DRAM in either application will save power. The problem, however, is that advanced, energy efficient DRAM technology is not widely available. Given the limited availability, the question is what is the best use of what little is available?

To determine what is available, we need to examine the total worldwide production of advanced DRAM products. Market analyst data in Table 3 shows the distribution of forecasted DRAM production and use for 2007 and 2008.

TABLE 3: Projected Distribution of DRAM Production

	Units	2007	2008	Notes
TOTAL production	512Mb EQ	9,203	13,849	1
1Gb as % total	%	4%	13%	1
Client	%	54%	54%	2
Server	%	13.5%	13.5%	2
DRAM in Client: Power	MW	72	109	3
DRAM in Server: Power	MW	90	135	3
1Gb supply/demand client	%	8%	24%	4
1Gb supply/demand server	%	31%	98%	4
Max Client savings	MW	2.80	13.28	4, 5
Max Server Savings	MW	13.93	66.11	4, 5
Annual delta from putting 1Gb in servers	MW	11.13	52.83	4
Annual delta Power	million kWh	97.5	462.8	4
5 year lifecycle savings	million kWh	488	2,314	4

Notes:

1. Average of Gartner, iSupply, and IDC.
2. iSupply 4Q06 report.
3. 6W/DIMM typical for system in use; 2W for system idle; 0W for client in E-star or off.

   www.sun.com/servers/coolthreads/t2000/calc/

   www.sun.com/servers/x64/x2200/calc/
4. Calculated.
5. Maximum savings assumes all 1Gb DRAM goes into this market segment.

The key will be the availability of 1Gb DRAM within the 2007-2008 time frame. A 1Gb DRAM built on advanced process technology will have power consumption on par with a 512Mb device built on older processes. The 1Gb DRAM enables a 2GB DIMM to be built using 18 chips rather than the 36 chips required with 512Mb DRAM.

Table 3 projects that in 2007, 1Gb DRAM shipments will only be 4 percent of total production, but will increase in 2008 to 13 percent. Since the server market is roughly one-fourth the size of the client market, it is possible to achieve a much higher market penetration in the server market. In 2008 the available 1Gb DRAM will be large enough to service 98 percent of the projected demand for servers.

The client machines' low power-on hours and low utilization shown in Table 2, combined with the market size estimates in Table 3, indicate a baseline power consumption 109MW in 2008. However, since the available 1Gb DRAM could only serve 24 percent of the total client machine market, the potential savings would be 13.28MW.

By comparison, the available 1Gb DRAM could serve 98 percent of the total server market. Applying the same mathematical computation as before produces a power savings of 66.1MW—a difference of 52MW. Significant power savings is achieved simply by channeling a scarce DRAM resource into a market segment where utilization is the highest.

These results represent a reduction of 0.463B kWh for equipment installed during 2008. If we assume this represents only one-fifth of data center servers, and if we extrapolate that to the entire installed base, we find the estimated savings would be 2.3B kWh. This estimate only assumes reduced chip count technology; adding in 1.5V DDR2 FBDIMMs would add another 2.3B KWh, which brings the total annual IT load savings to 4.6B kWh.

Finally, when the overhead power consumption is added into the mix, an equal amount of overhead energy can be saved with regard to reduced cooling, supply inefficiencies, etc. In other words, another 4.6B kWh could be saved, bringing the total potential savings to more than 9B kWh annually.

Currently, the effective service life of servers is five years or longer due to the depreciation schedule imposed by Internal Revenue Service. Dr. Koomey's report indicates a server life of three to five years. Micron's investigation revealed on a limited dataset that if a server was no longer capable of meeting customer requirements inside the 5 year schedule, it would be repurposed for a less intensive workload. Thus, most servers can be expected to run for a minimum of five years, with many running much longer. In contrast, client machines are typically expensed or depreciated on a shorter service life schedule.

Footnotes:

1. Energy Information Administration: Official energy statistics from the U.S. Government: www.eia.doe.gov/cneaf/electricity/epm/table5_3.html
2. Enterprise Server and Data Center Energy Efficiency Initiatives: www.energystar.gov/index.cfm?c=products.pr_servers_datacenters
3. http://clerk.house.gov/evs/2006/roll369.xml
4. http://enterprise.amd.com/Downloads/svrpwrusecompletefinal.pdf
5. University of Washington Television Webcast: www.uwtv.org/programs/displayevent.aspx?rID=2879 : Urs Holzle presenting.
6. http://enterprise.amd.com/Downloads/svrpwrusecompletefinal.pdf
7. Ibid.
8. Memory power consumption in a fully loaded server is estimated in the range of 25% to 66%. This analysis assumes 20% on the basis that we do not have a clear understanding of memory content per system; www.energystar.gov/index.cfm?c=products.pr_esads_conf_media The presenter is Gregg Papadopoulos, CTO of Sun Microsystems.
9. http://enterprise.amd.com/Downloads/svrpwrusecompletefinal.pdf. Note (7) of this document states that total electricity consumption (including cooling and auxiliary equipment) is twice that of the direct server power consumption, based on typical industry practice

About Micron

Micron Technology, Inc., one of the world's most efficient and innovative semiconductor companies, manufactures and markets a full line of DRAM components and modules, NAND Flash memory, CMOS image sensors, and other semiconductors. Our broad product line includes both legacy and leading-edge solutions, offered in multiple generations, densities, configurations, and packages to meet the diverse needs of our customers. With operations in 18 countries, customers can count on us to deliver the expert design, manufacturing, sales, and technical support—and ultimately, the high-performance, advanced semiconductor solutions—that go into successful product designs.