Improving HPC Cluster Efficiency with 480V Power Supplies

Giri Chukkapalli, Cray R&D Principal Engineer
Maria McLaughlin, Cray Cluster Product Marketing

Cray Inc.

Table of Contents

Introduction

Traditional AC power distribution systems in North America bring 480V AC power into a universal power supply (UPS), where it is converted to DC to charge batteries, and then inverted back to AC. The power is then stepped down to 208V within the Power Distribution Unit (PDU) for delivery to the IT equipment. The power supplies in the IT equipment convert the power back to DC and step it down to lower voltages that are consumed by processors, memory and storage.

Server power supplies that operate directly from 480/277V power distribution circuits can reduce the total cost of ownership (TCO) for a high performance cluster by reducing both infrastructure and operating cost. Most servers used in high performance clusters were originally designed for desktop applications and for operation in office environments. They normally operate from 120/240V single-phase AC power and are cooled by room air conditioning. The reason for using low voltage 120/240V distribution circuits is safety. This equipment is connected and disconnected by untrained personnel creating the potential for electrical shock. When large clusters are constructed using these same commodity servers, the legacy power and cooling designs create inefficiencies. The 120/240V circuits were never designed to deliver large quantities of power. These inefficiencies are acceptable for a single server but the aggregate losses in a large system with thousands of servers are too large to be acceptable. The higher voltages do not present a safety issue for these larger systems since they have better facilities and trained maintenance personnel. These servers need an alternate power supply design that better fits a computer room environment.
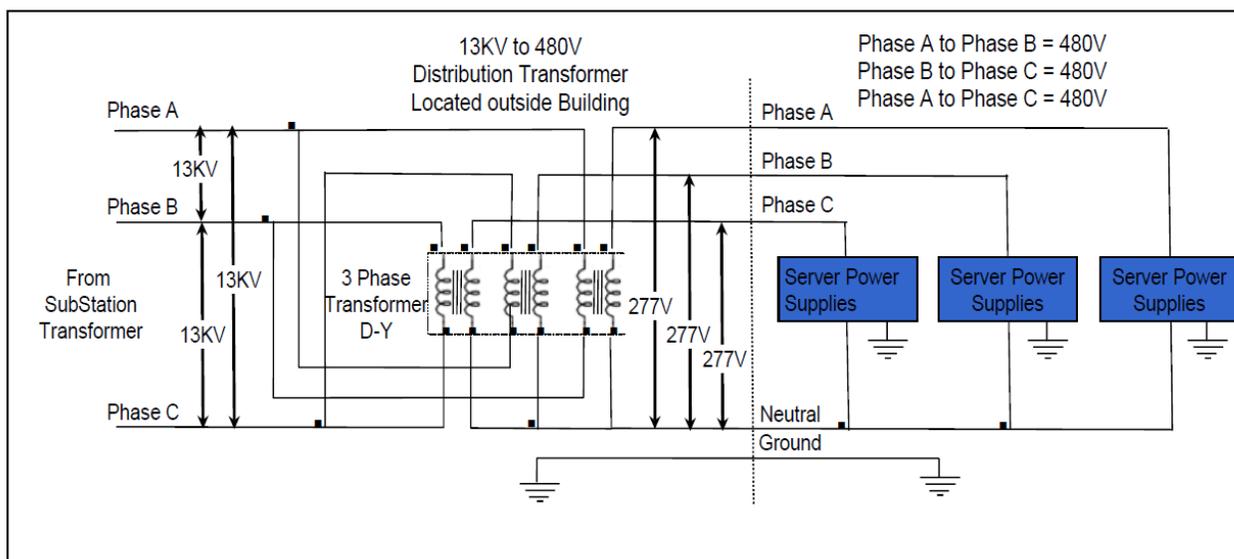

Figure 1: 480V three-phase power distribution directly to server power supplies

Power Distribution

Commercial and industrial users have large power demands. Figure 1 illustrates a typical 480V power distribution system used in commercial facilities. The distribution transformers are generally located outside the buildings. They receive high voltage power directly from a power company substation. The 480V secondary circuits may be capable of delivering up to 2,000 amps, and fuses that protect the transformers are located at the transformers while circuit breakers that protect individual power distribution circuits are located inside the building. The power is distributed from the circuit breakers to equipment inside the building using 480V three-phase circuits. The use of the higher voltage 480V circuits reduces resistive I2R losses by a factor of four to five times compared to using 240V or 208V circuits. This is because 480V circuits require less current to deliver the same amount of power. A load requiring 10KW of power would require 28 amps at 208V, 24 amps at 220V, but only 12 amps at 480V. Most large industrial equipment and mainframe computers are designed to operate directly from 480V three-phase circuits.

Step-down transformers are necessary to convert the 480V three-phase circuits to 208V three-phase and 120/240V single-phase circuits. These circuits are used to power office equipment, desktop computers and coffeemakers. Resistive losses are minimized by locating the 480V to 208V step-down transformers near the loads. Using these lower voltage circuits to power a high performance computer results in unnecessary cost compared to using the 480V circuits directly. It also results in decreased efficiency when 208V circuits are used to operate 120/240V power supplies because they are operated at the low end of their operating range. [1]

The HPC Power Issue

A large high performance computing (HPC) cluster may have thousands of individual servers. Sixty-four or more servers can be located in a single equipment rack requiring up to 30KW. A single equipment rack can require more power than is required by a small office. A typical HPC cluster with 1,024 compute nodes may require 16 or more equipment racks and need more than 500KW of electrical power. Figure 2 illustrates how 208V three-phase power is distributed in current computer rooms. These applications typically use multiple step-down transformers with ratings between 15KW and 150KW to reduce the voltage from 408V three-phase circuits to 208V three-phase circuits. The only reason for the step-down transformers is to accommodate the 120/240V power supplies used in the servers. In fact, current power supplies would be more efficient if the transformers output was 415/240V. Figure 3 illustrates the reduced efficiency of operating 120/240V power supplies on 208V or on 120V. The efficiency quoted for 120/240V power supplies is always measured when running on 240V. They are never as efficient when running on 208V or 120V.

A 1,024-node HPC cluster consisting of 16 equipment racks with 64 compute nodes in each rack would typically require up to approximately 28KW per equipment rack. The servers in each equipment rack would be powered by between 24 and 32 separate power supplies, each with a 1,600- to 1,800-watt capability. The power would typically be distributed to each rack by two 208V 50-amp three-phase PDUs. In many clusters built with commodity servers, up to four distribution circuits per rack are required because of the limited size of commercial rack PDUs. Our example would require approximately 32 separate 208V three-phase power distribution circuits with circuit breakers rated at 50 amps. The power would typically be supplied by approximately eight 75KVA step-down transformers with each transformer supporting four 208V three-phase outputs.
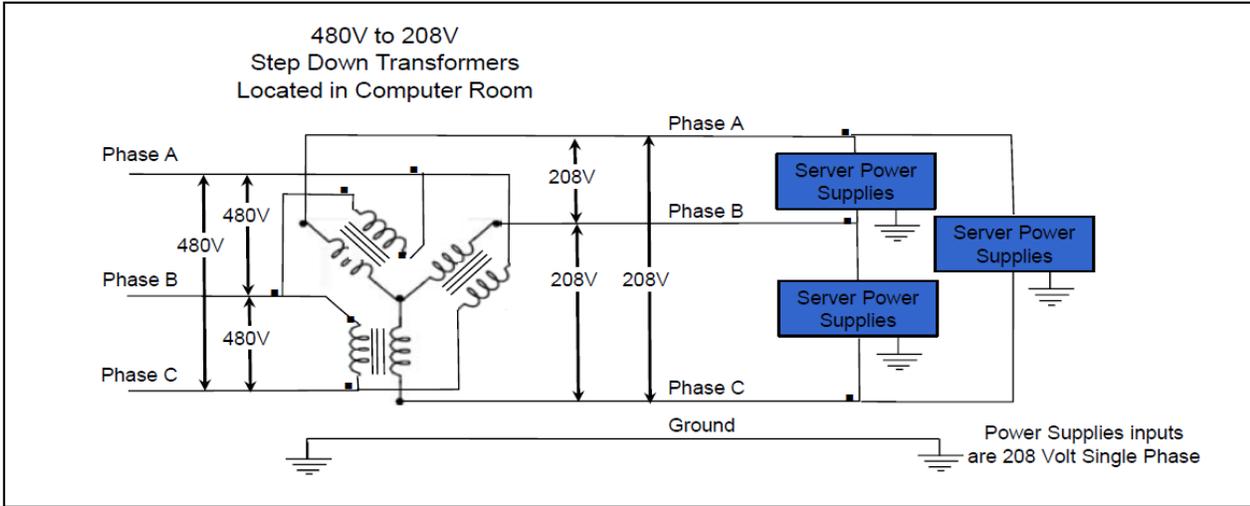
Figure 2: Power from 480V three-phase to 208V three-phase step-down transformer

The 208V three-phase circuits from the step-down transformers are four-wire circuits, which have conductors for each of the three phases plus a safety ground. The server power supplies are connected between two of the phases. This feature provides them with 208V single-phase power. Since the 208V circuits are three-phase circuits, it is important that the server power supplies be distributed equally across the three phases to balance the current in the phases. The PDUs located in the racks would be 208V three-phase units delivering an aggregate of approximately 40 amps to 12 to 16 outlets. Under the North American electrical codes, the branch-level circuit breakers are 80 percent rated so that a 50A circuit can deliver only 40A to a load.
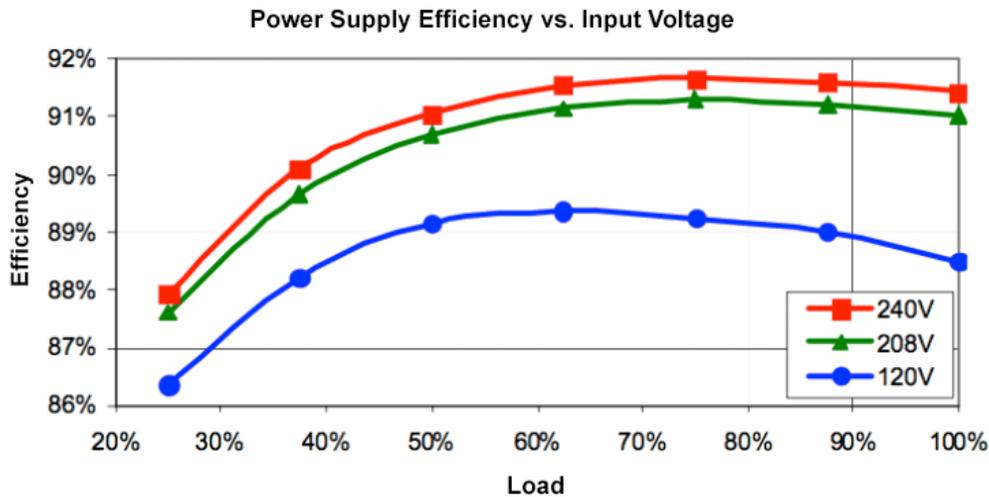

Figure 3: Power supply efficiency versus input voltage

The 480V to 208V transformers used in these applications are designed and rated on their ability to deal with non-linear loads, such as those found in the switching power supplies used for servers. Transformers are rated by the K-factor system, which ranges from K-1 units, which are designed to work only with resistive loads, to K-40 units, which are best suited for use with non-linear loads. Non-linear loads cause harmonics in the transformers that result in increased losses in both the wiring and the transformer cores. In datacenters the commonly available transformer types are K-1, K-13 and K-20. K-1 transformers, which are not designed for non-linear loads, are the least expensive and are normally used for resistive heating and lighting loads. K-13 transformers are somewhat better but more expensive, while K-20 or K40 transformers would be the preferred choice for datacenters using switching power supplies.

Most low-cost transformers used for these applications have aluminum windings and no better than K-13-grade core material. This is driven by installation cost considerations. Users have not been so concerned about the design of these transformers in the past and electrical contractors tend to choose the least expensive units. Using more expensive, higher efficiency transformers could cut transformer losses in half. Transformers that use copper windings instead of aluminum have lower I2R losses. Transformers that use higher quality core material have lower magnetic circuit losses. These transformers are also more expensive, and may offset any savings in electrical power, resulting in approximately the same overall cost. Contractors will typically install K-1 or K-13 transformers if possible because they are more concerned about the installation cost than the operating cost. The best solution, however, is to reduce the number of transformers required.

In a typical 1,024-node cluster, requiring 500KW, the power loss in the 480V to 208V step-down transformer is approximately 3 percent. The losses are due to a combination of resistive and magnetic core losses. A 500KW system would typically have transformer losses of approximately 15KW. This power loss is converted to heat inside the transformers. If the transformer is located inside the computer room, the heat must then be removed by the computer room air conditioning equipment. The losses in the transformers and in the wiring both add to the total load on the 480V three-phase circuits and to the air conditioning load. [2]

The typical life of a high performance cluster is three years. Approximately $47,000 is wasted over the life of a 500KW computer based on a power cost of 12 cents per kilowatt hour. If step-down transformers are located inside the computer room, the increased air conditioning load could cost an additional $20,000 to remove the heat generated by the transformers. For a 500KW system the installed cost of the step-down transformers would add a minimum of $40,000. This adds up to a total life cycle cost of $107,000. This is wasted money. It is wasted as heat and the purchase of equipment that would not be required if the server power supplies operated directly from the 480V three-phase circuits.

The configuration in a system using 277V power supplies is different from the 208V system. The power supplies connect from the phases to neutral instead of between the phases. This provides 277V single-phase power to the power supplies. Only a small increase in the input voltage rating of the power supplies is required to accommodate the change from 240V to 277V operation. In fact, when operating from 208V most server power supplies operate at the low end of the 240V input range and are not as efficient as when operating on 120 or 240 volts. Even though the phase-to-phase voltage is 480V, the voltage at the receptacles for the server power supplies is only 277V.

A 208V system has higher I2R wiring losses than the 480V system. This is because the lower voltage power supplies require more current to deliver the same power. The current required for the 208V power supplies is 277V/208V=1.33 times higher than for the 277V power supplies. Each of the 208V three-phase circuits carries 1.33 x 1.73 = 2.3 times the current because of the current sharing in a three-

phase circuit with the power supplies connected phase to phase instead of phase to neutral. Since the losses are proportional to the current squared times the circuit resistance, the losses for a 208V system is up to 5 times greater than for a 480V system. The loss will actually be a bit less since the 208V system will need more lower-resistance wiring, which also adds to the installation cost.

To determine the savings, one first needs to compute the losses for the 208V system. Then, subtract the losses for an equivalent 480V system. This example still uses the same 1,024-node 500KW system used in the previous example. Thirty-two 208V three-phase 50 amp PDUs would be required for each equipment rack in the system. Each 208V three-phase circuit would have a maximum rating of approximately 18KW. The wiring on each phase would be required to handle a total of approximately 40 amps per phase. This would require No. 4 gauge wiring that has a resistance of 0.25 ohms per 1,000 feet. Assuming the step-down transformers are located within 50 feet of the power supplies, the wiring resistance would be 0.0125 ohms per phase. The resistive losses in the 32 circuits would be 1.9 KW.

Loss = Phases x Current2 x Resistance x Number of Circuits
Loss = 3 x 40 x 40 x 0.0125 x 32 = 1.9KW or approximately 0.4 percent

When using 480V power supplies, illustrated in Figure 4, the system still requires 32 power distribution circuits. The number of circuits is determined by the number of servers being supported by the circuit. These circuits are each required to supply a total of approximately 19 amps. If No. 8 wires were used for these connections, the same length of wire would have 0.03 ohms of resistance, resulting in a loss of 1KW or approximately 0.2 percent.

Loss = 3 x 19 x 19 x 0.03 x 32 = 1KW

Since the 480V wiring may need to be routed further to supply the 480V to 208V step-down transformers, the 480V wiring is necessary in both 480V and 208V systems. The additional losses in a 208V system are approximately 1,900 watts. The savings over three years at 12 cents per kilowatt hour are approximately $6,000.
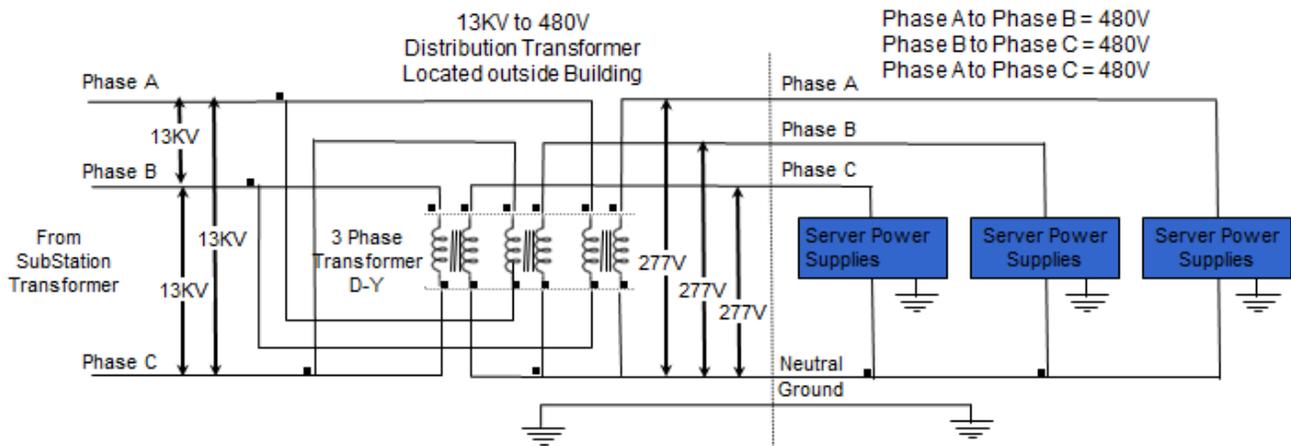

Figure 4: 480V distributed directly to the server power supplies

This makes the total savings from lower resistive losses and eliminating the step-down transformers more than $110,000. Additional savings for the 480V system would also result from the use of less expensive wiring and power distribution units. The savings would be even higher if the wiring runs were longer than the estimated 50 feet.

Power Distribution Examples

Figure 5 illustrates a one-line diagram for a 208V power distribution system with 480V to 208V step-down transformers that provide power to the PDUs and servers. Figure 6 illustrates the same diagram for a system with 277V power supplies that use the 480V circuits directly. The visible difference is the requirement for step-down transformers in the 208V system. The less obvious difference is the increased resistive loss in the 208V system.

Another system that should be mentioned is one with an uninterruptible power supply (UPS), as illustrated in Figure 7. Many UPS systems have 480V inputs and supply 240V to the load. These systems require transformers inside the UPS system to transform 480V to 240V. This is particularly true for bypass mode, where the UPS is bypassed during maintenance. The transformers in a UPS may be of higher quality but still experience the same losses. The best UPS solution is a 480V UPS with 480/277V outputs and power supplies that use the 480/277V outputs directly. These 480/277V power supplies are an integral part of any high-efficiency power distribution system for a datacenter. [3]
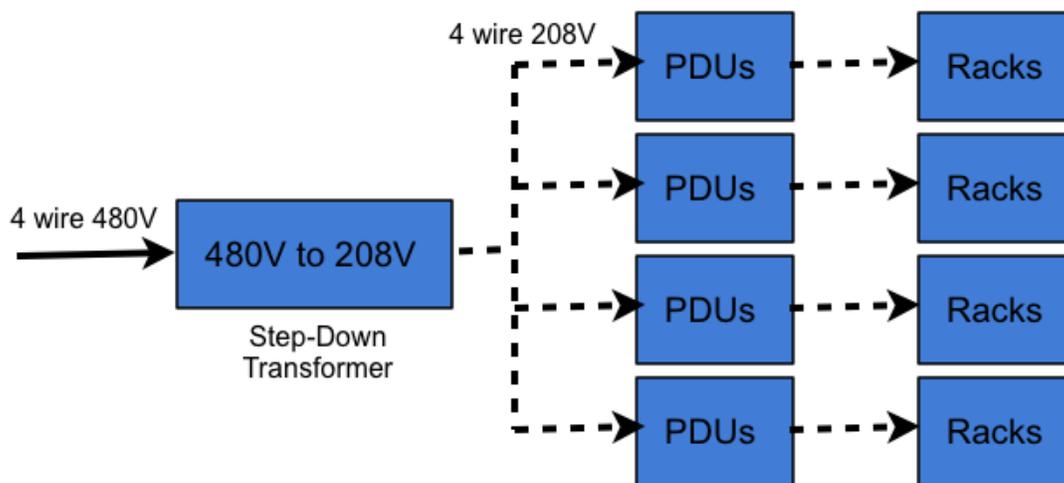

Figure 5: One-line diagram for 208V power distribution

Total cost of ownership is improved by:

1) Elimination of the 480V to 208V three-phase step-down transformers

2) Lower resistive losses because of lower current in the power distribution circuits

3) Lower cost power distribution units

4) Lower copper wiring cost

5) Higher power supply efficiency by operating the power supplies at the appropriate input voltage; power supplies designed for 240V operation are less efficient when operated at 208V.

Negative impacts of using 480V power supplies:

1) Miscellaneous equipment such as data routers, switches and storage still require 120/240V inputs. This will require separate 120/240V PDUs. This has an upside because the power supplies in this equipment will be more efficient operating from 120/120 than on 208V circuits.

2) The 480V systems require better trained personnel and more concern with the higher voltages because of the consequences of shorts or arcing.
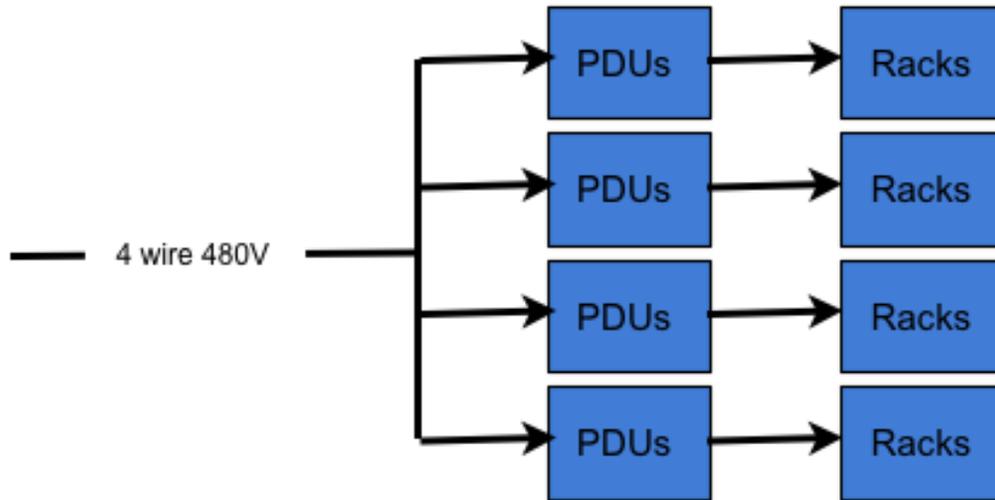

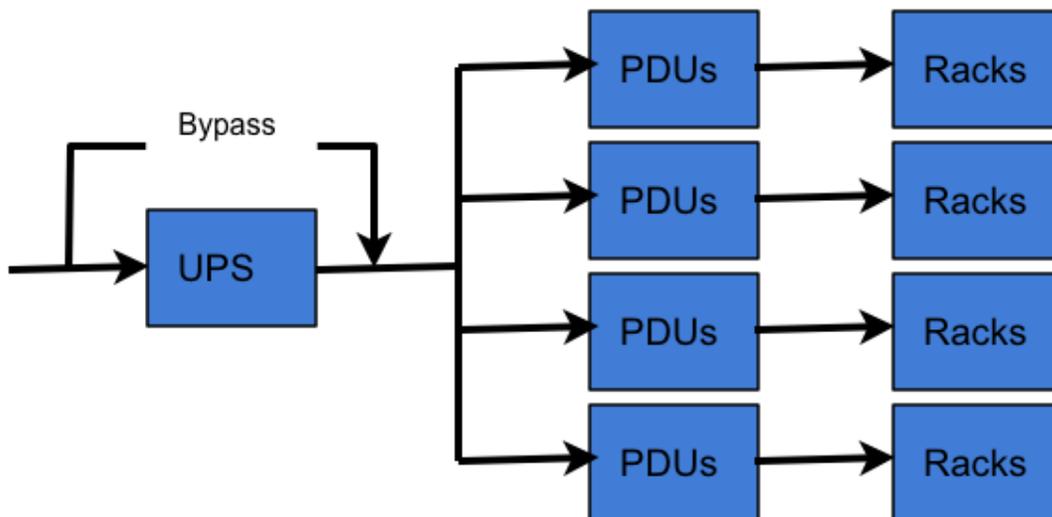**Figure 6:** One-line diagram for 480V power distribution

Figure 7: System with uninterruptible power supply

Cray introduces 480V power supplies to be used in the Cray CS300™ cluster supercomputer series. This has been driven by requirements from customers such as U.S. national labs and the U.S. Department of Defense for more efficient power supplies and power distribution systems.

Summary

Many of the servers used in computers on the HPC TOP500® list use 208V power supplies and very inefficient air cooling solutions. Some of these systems have tens of thousands of servers. This represents millions of dollars in wasted electric power and added equipment infrastructure costs. In today's energy-sensitive environment, these practices need to change. The switch to 480/277V power supplies for the next generation of high performance clusters is a small change with large rewards.

References

[1] Qualitative Analysis of Power Distribution Configurations for Data Centers – The Green Grid. Retrieved 2007
http://www.thegreengrid.org/~/media/WhitePapers/TGG_Qualitative_Analysis.pdf?lang=en

[2] Data Center Transformer Types and the K Factor – Search Datacenter
http://searchdatacenter.techtarget.com/tip/Data-center-transformer-types-and-the-K-Factor

[3] The Role of Isolation Transformers in Data Center UPS Systems – APC by Schneider Electric
https://www.insight.com/content/dam/insight/en_US/pdfs/apc/apc-role-of-isolation-transformaers-in-data-center-ups-systems.pdf

Acknowledgements