

Meeting the Demands of Computer Cooling with Superior Efficiency

The Cray ECOphlex™ Liquid Cooled Supercomputer offers Energy Advantages to HPC users

Dorian Gahm
Mark Laatsch

Contents

Introduction.....	3
Cooling History.....	3
Best Practice.....	4
Cooling Technologies of Yesterday	4
Air Cooling.....	4
Liquid Cooling.....	5
Cray High Efficiency (HE) Cabinet with ECOphlex Liquid Cooled Technology	6
Cost of Ownership / Power Usage Effectiveness.....	8
Cray Liquid Cooling Benefits.....	10
Conclusion	13
Acknowledgement.....	13

Introduction

The worldwide energy crunch has seen the issues of energy costs and efficient power usage thrust upon the users of high powered computers in the past several years. Vendors and users alike have been running just to keep up with the demand for compute power at low cost. Modern high performance computing (HPC) demands are calling for computer systems to do more computing in less space, and Moore's law has kept processing power up to this demand. However, this means individual compute racks are currently pushing 40 kW, or more, and there is a need to rapidly adapt cooling on both the facility and rack level.

Efficient energy use is needed to serve the needs of the HPC community. Limited supplies of energy along with rising costs make it important to look for energy savings even down to the fraction of a percentage. With 35% or more of the power in an average modern computer facility going towards cooling, from the chiller, to pumps and fans, it is easy to see how efficient cooling can translate into large power savings. Efficient cooling not only shows up at the bottom line, but in positive environmental impacts. A reduction in cooling power consumption means less coal and gas needs to be wasted, or renewable energy diverted, to moving heat out of the computer room. This allows datacenters to be both good stewards of the environment and be recognized in the communities for their diligence.

Cooling History

HPC users may one day look back with amusement at the designs required to cool the first super computers. The first systems were immensely complex and did the work of a device that could now fit in a pocket. Early systems required massive heat removal and had to be kept cool at any cost. These systems were dense and processing inefficiencies led to a large amount of waste heat.

The original supercomputing designer, Seymour Cray, used to say, with tongue in cheek, that he was little more than a glorified plumber. The world's first supercomputer, the Cray 1, pumped Freon in pipes throughout the computer while peaking at 250 MFlops of compute power. The next iterations of cooling technology were even more complex. The Cray 2 system did away with the piping all together, with inert liquids flowing right over the boards. Extensive pumping was required to remove and replace the boards. This kind of maintenance would not be acceptable in the modern marketplace.

The transition to CMOS microprocessors in the late 1980's had a profound impact on how mechanical engineers looked at cooling computers. The efficiency of CMOS processors dramatically decreased power consumption, and along with several other advances in computing technology, allowed for a re-thinking of computer cooling. Density fell, computers could spread out, and air became more than sufficient to cool the processors. Unfortunately, or fortunately, for cooling engineers, Moore's law continued to hold true. Transistor density doubles every 18-24 months, along with this heat increases at a similar pace.

Compute density has once again increased, driven by the need for network and memory improvements to keep pace with processing capability. The foot print of a large system also strains the limited space available in many compute facilities. The increasing heat loads of these

large, dense systems are pushing system vendors to develop and adopt new cooling techniques. The power densities are also pushing the limits of simple air cooled technologies. Many facilities are seeing the footprint for the Computer Room Air Conditioners (CRACs) expand beyond that of the computer system itself. As air cooling reaches its limits, the necessity for more efficient liquid cooling designs has become apparent.

Best Practice

Typically, all heat generated by a computing system must be removed through the building chilled water in order to maintain a safe computer room environment. Due to chiller design efficiency there is a relation between the increase in water temperature and energy used. System cooling designs are varied by the way in which the heat is transferred to the water, how the water is distributed, and by the chilled water temperature necessary for heat removal. Data centers want to transfer the heat into the chilled water in the most efficient way possible, while minimizing cooling power requirements.

The cooling design can be made most efficient through the use of best practices and a combination of technologies. A system should be designed to generate the least amount of unnecessary secondary heat. Pumps and fans need to be limited and designed for efficient flow and speed, while still providing adequate reliability. In cutting edge facilities, running complex, time-consuming codes, on state of the art supercomputers, reliable computer operation is imperative. Designs should minimize the danger that the cooling system could damage or disable the supercomputer itself and should also maintain a near constant operating environment in order to achieve the best performance from the machine.

For further information on facilities design best practices, the [ASHRAE](#) organization has extensive datacenter design and operation literature available.

Cooling Technologies of Yesterday

Air Cooling

Air cooling has been the standard practice for over a decade in the HPC market. Air cooling is commonly used in low power density systems and is easy to install and maintain. Air is a very efficient medium for transferring heat from system components, if properly designed, but is usually inefficient at moving heat from the rack to the facility CRAC units in large room environments. Fans can effectively move the hot air out of the racks and into computer rooms, but often create hot spots, which are inefficient and can be problematic for the system if they become severe. Left unchecked these hotspots could overheat the computers to the point of failure very quickly. Alleviating these hotspots often requires detailed analysis, including computation fluid dynamics (CFD) modeling of the datacenter, to properly position more fans to move hot air through the room and to the CRAC units. In large datacenters, as the hot air mixes with the cold air in the other areas of the room, CRAC efficiency plummets with a small temperature differential, requiring over-sizing of CRAC units.

The complexities involved in efficiently air cooling the datacenter can lead to intricate methods of manipulating the heat distribution. In order to avoid hot and cold aisle mixing, walling can be used to separate the hot and cold aisles, or ducting can be used to take the heat to the CRAC units. Raised floors are required in average air cooled configurations, and this can be a helpful place to distribute wiring, piping, and communications infrastructure. CRAC fans create a large amount of pressure gradient under the floor, an effect that can be difficult to predict using CFD modeling. The large pressure gradient also decreases CRAC efficiency.

While air can be inefficient and awkward to deal with at the datacenter level, in the supercomputer itself air is an effective way to remove heat from components, even in dense systems. In air movement designs within the supercomputer itself, fan size and placement is an important design choice. Small fans are common in many systems because they spread the air easily over a large region. However, this increases the quantity of fans necessary. Rows of small fans also require significant upkeep. The downtime can be managed over the small region of fan failure with the other fans ramping up to accommodate a higher load. However, the frequency of failure requires significant maintenance time and costs.

Larger system fans are used in several newer designs, which include the Cray XT4 and XT5 cabinets. As fan size increases, the airflow efficiency also increases, requiring less power to move comparable amounts of air, this allows for larger pressure differentials across more densely populated boards. A single large reliable fan provides a significant advantage in MTBF over numerous small fans. The air cooled Cray XT4 and XT5 cabinets are examples of highly efficient systems operating with a single blower.

Liquid Cooling

Liquid cooling has again become an important player in heat removal in high density applications. Fluid heat removal comes in several forms today, and all fluid systems must address common issues. There are two places where liquid heat exchange is commonly designed to occur. Air can be used to efficiently remove heat from the system components and the heat can then be extracted from the air at the point it enters and/or leaves the cabinet (a door or panel). Secondly, heat can be directly removed by a liquid and air combination at the board or processor level with invasive liquid cooling.

Chilled water heat exchangers are now offered by many HPC vendors. Hot air passes over the exchanger and loses heat to liquid as it exits into the data center. Cold liquid water comes into the coils and is heated by the air and warm liquid water goes back to the chiller. Water is up to 4000 times more efficient than comparable volumes of air at absorbing heat. Yet water has inherent risks associated with being near a computer. Concerns with leakage require careful leak management, and condensation concerns require strict temperature and flow management. This requires the heat to be exchanged to the facility water via a secondary water loop that monitors chilled water flow and temperature to the cabinets. The water also needs to be treated to address the issues of bio-contamination and corrosion.

Refrigerant has been used for many years in air to liquid heat exchanging across a variety of cooling applications. While the process has several similar features to the water cooled process, where liquid flows into the heat exchanger, there are several key differences. Refrigerant boils at a low temperature due to low latent heat of vaporization. This process absorbs more heat and can do so in a more efficient manner. The two phase process is 10 times more efficient than the single phase water process at a similar flow-rate, and requires a much smaller surface area for heat transfer. In addition, micro-channel heat exchangers used in some refrigeration

processes take advantage of the critical boiling effect and greatly increase heat transfer efficiency over water coils, lowering air pressure drop and decreasing package size. Hot air passes over the heat exchanger and enters the room cool. A vapor/liquid mixture exits the heat exchanger and returns to a condensing unit where it returns to liquid form and is pumped around the system again. Because refrigerants are inert non-conductive liquids, rare cases of leakage pose little threat to the computer boards.

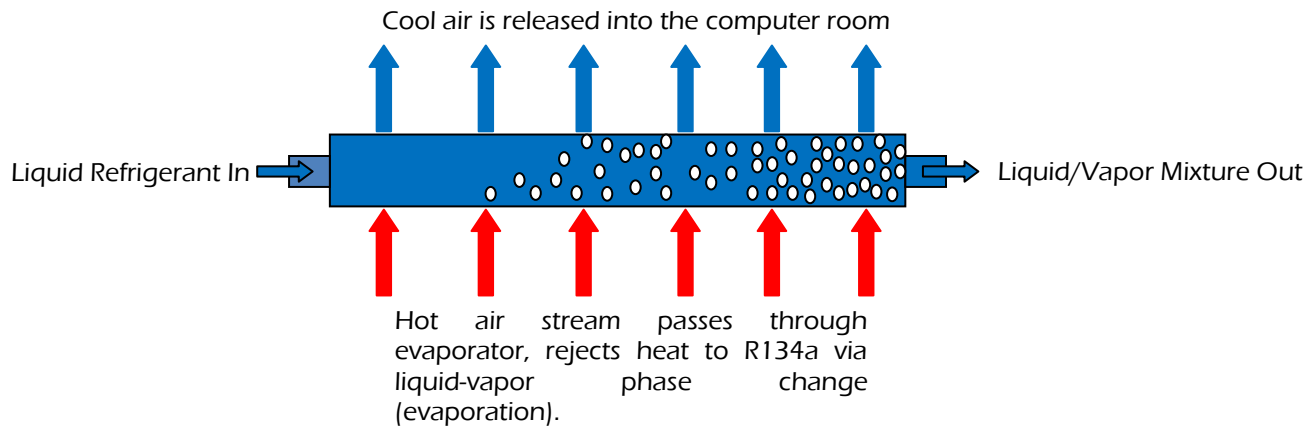


Figure 1: Evaporative Heat Exchanger

Micro-Channel heat sinks pumping the fluids onto the boards are excellent methods for bringing the cold liquid right to the hottest part of the computer, the processor. However, the other computer components such as the memory and I/O must still be cooled via some secondary source such as fans, air conditioning, heat exchangers, or some combination of these. Again, a secondary loop is needed to treat fluid and maintain dew point. The micro-channel cold-plates and piping occupy valuable space on the compute boards, limiting design flexibility and upgrades.

Cray High Efficiency (HE) Cabinet with ECOphlex Liquid Cooled Technology

Cray is now making available a novel, non-invasive approach to heat removal that brings the refrigeration to the cabinet, transferring heat with a patented “flooded coil” cycle. This technology, termed ECOphlex (phase change liquid exchange) uses efficient air flow to remove heat from the base components, and a phase-change refrigerant system to remove heat from the air prior to leaving the cabinet.

Refrigeration has long been used in cooling because it is an inherently efficient process that can remove large amounts of heat. Unlike classic vapor compression cycles which use the expansion of the refrigerant to absorb heat from all of the surroundings, creating massive amounts of condensation; the flooded coil system utilizes the latent heat of vaporization to absorb heat only if there is heat in the air, maintaining a temperature above dew point in the entire coil.

Using an individual cooling cabinet, the Cray ECOphlex system can do the cooling of nearly three CRAC units in less than a fourth of the space. In a three by three footprint (less than half of an industrial CRAC unit), a Refrigerant Pumping Unit (RPU) can remove 240 kW of heat. Such large heat removal capacity can easily remove the heat generated by multiple modern compute cabinets that consume about 35kW each. This form of heat removal also allows for the progression of up to 60 kW of heat removal per rack, spread over four racks, and the capacity for future growth through RPU development is enormous. The directly mounted vaporization heat exchangers take heat away as it exits the top of the rack. The heat therefore does not enter the computer room and eliminates the need for CRAC units. This allows for a “room neutral” system, i.e. it has no effect on the room environment in many datacenters, depending on ambient room and chilled water temperatures.

In addition, compared to water cooling designs, the refrigerant cooling cycle keeps the water, and it's inherent damage potential, further away from the internal supercomputer components. Dew point is monitored to ensure no condensation occurs at any point on the refrigerant loops. In air to water HEU designs, some condensation is nearly unavoidable (which is why they require a drip tray) and preventing condensation on chilled water pipes requires significant insulation. Cray has also implemented a new peak top plenum design. This triangular plenum allows air to slow down and become more uniform before passing through the evaporators. This spreads the heat over the evaporator making the evaporator more efficient. A variation of this peak design allows the system a lower profile while spreading the evaporators further to the front and rear of the unit.

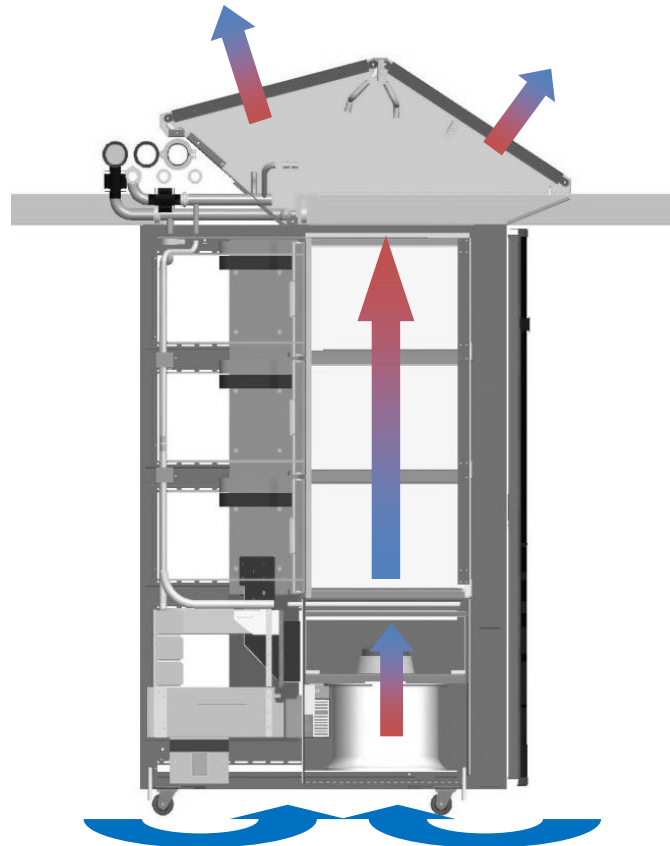


Figure 2: Vertically aligned ECOphlex cabinet cooling diagram

To complement the ECOphlex vertically cooled evaporative system, the HE cabinet design uses a single, large, high-efficiency fan built with an industrial motor and ceramic bearings for better reliability. This custom designed turbo fan is very efficient, converting 77% of electrical power to wind power as it moves a volume of air comparable to many smaller, less efficient, “off-the-shelf” fans. The smaller fans take up a larger foot print and operate at about 40% efficiency. The fan design allows it to push the air against the pressure of three consecutive blade enclosures and remove heat from all compute components with maximum efficiency. Cray’s patented air cooling technology allows blades to be in series, while all key components are maintained at a constant junction temperature. This design also allows for a 300% increase in the quality of heat in the cooling air, which allows for an increase in heat exchanger efficiency of 300% compared to front-to-back or side-to-side designs. The vertical air flow direction also minimizes footprint and eliminates any horizontal “hot aisle” temperature gradient in the room. The Cray XT5 system today, using the HE cabinet, is capable of achieving well over 11TFlops of performance per rack, or approximately 1TFlop per square foot for a five rack system including the RDP. The HE cabinet cooling can sustain much higher densities and will be expandable and upgradeable to future processor, memory and network technologies.

Cost of Ownership / Power Usage Effectiveness

One of several metrics for computer room cooling efficiency is in the Total Cost of Ownership (TCO) for the cooling system. This metric includes all the annual costs associated with use in addition to the upfront costs of acquiring a cooling system. Along with the initial purchase, these costs include maintenance, power consumption, and floor space. All these metrics can change in value depending on location, yet it is clear that reduced TCO comes from effective management, low maintenance and low power in a minimized floor space, at a low price.

The Power Usage Effectiveness (PUE) of a system is another driving factor in determining the most effective cooling process. PUE is defined as the ratio of total system power to IT power. Total power includes all aspects related to the cooling and operation of a computer. This would include the entire data center cooling infrastructure, such as CRAC units and chillers. A computer operator wants this ratio to be low, as this would mean more of the limited power resources are being used for compute performance rather than to compensate for cooling inefficiencies to remove heat.

Ideally a system PUE would be equal to 1, if a system operated in a cool natural environment that was always well lit and supplied with fresh filtered air. Normally this ratio ranges from 1.5, for a very efficient system in an efficient datacenter, upwards to 2.5 for inefficient cluster systems or for inefficient datacenter designs. A ratio of 1.5 would indicate 1/3 of total power is going to cooling and power loss, and in a normal system most of this goes into cooling the water. About 20% of the total power goes to chilling the water. The rest of the cooling power usually goes to the secondary cooling system (fans and pumps) to move air and liquids around the computer room.

This chart shows a typical power usage profile for a standard 1.8 PUE facility. A typical HPC system operates with a high efficiency 90% power conversion rate, losing 9% of the total system power to voltage conversion.

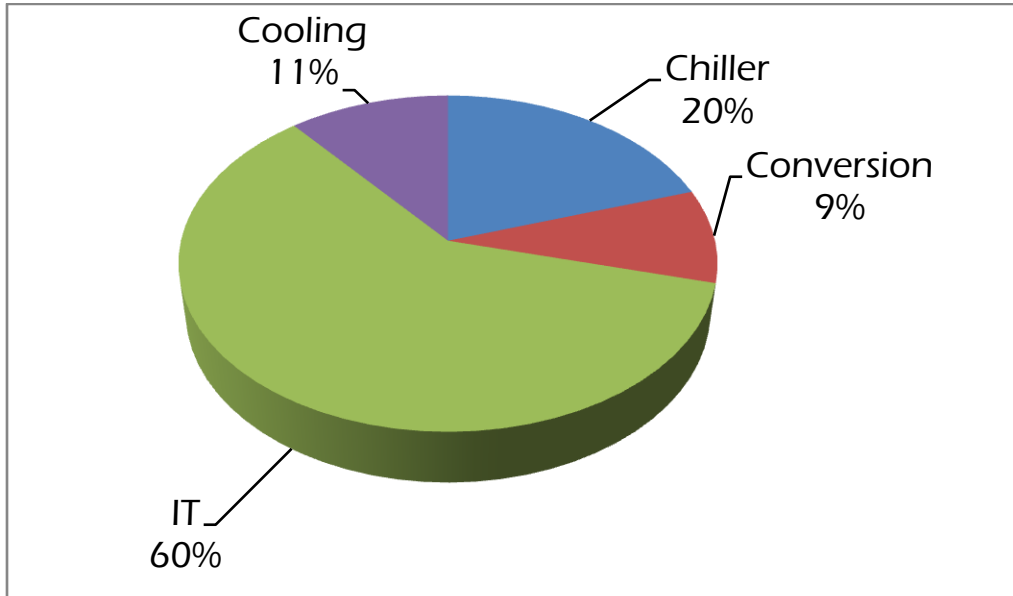


Figure 3: Standard HPC Facility Power Distribution

As the largest component of the cooling cycle, the building water chiller plant is the ideal place to first tackle cooling costs. The main power consumer in a chiller plant is the compressor, acting as a large inefficient pump unit. Due to the design of a water chiller, on a variable speed device each one degree increase in water temperature can decrease power consumption 2-4%. In order to take advantage of higher water temperatures with standard cooling designs, however, would require an increase in room temperature comparable to that of the chilled water, a significant increase in design footprint, or an increase in cooling capacity for inefficient CRAC units. The average chilled water plant includes centrifugal chillers, pumps, and cooling tower fans. They run at an efficiency of about 1.25 kW per ton of cooling. This number could vary based on the age of the central plant and energy conservation strategies. A small change in chiller plant efficiency can have a large impact on system PUE.

Without major changes to the chiller plant requirements, the next place to look to power efficiency is in the supplemental cooling technology.

CRAC units have been the standard for computer room cooling for many years, and their cooling capabilities and costs can be summarized here. A standard water cooled CRAC unit uses about 7.5 kW of fan power to remove 120 kW of sensible heat, i.e. 1 Watt power: 16 Watts cooling. These can operate as low as 50% efficiency and up to 75% efficiency in room, translating to at best 90 kW of real cooling or 11 to 1 cooling to power ratio. Installation of a CRAC unit requires facility piping, insulation, fittings, valves, controls, and precise engineering. Maintenance for CRAC units includes upkeep for fans, fan motors, air filters, and chilled water connections. The building automation system can be used to monitor and control the CRAC from a remote location. Each CRAC unit also occupies a significant area in a computer room, typically about 3 by 8 feet, along with an additional 3 foot plenum to develop flow, or 48 square feet in total.

For any liquid cooled system, such as the liquid cooled Cray ECOphlex, a secondary fluid loop must be maintained in the refrigerant pumping unit (RPU). The refrigerant pumping unit of the ECOphlex system exchanges heat between the refrigerant and the building water. Current designs can remove 240 kW of power using a RPU with a three foot by three foot footprint. Using only a 2.5 kW pump to perform this cooling relates to a nearly 100 to 1 cooling to power ratio, approximately 650% more cooling efficiency than a CRAC unit. Because the system operates at a large air temperature differential, it also performs the cooling at nearly full efficiency. At a system installation cost comparable to CRAC units, the secondary pumping units require significantly less maintenance due to a simple liquid pumping design.

Fan usage will also figure into a facility PUE. There are two factors that determine how efficiently fans can move air; size and rotation speed. Larger fans move more air at lower rotation speeds, moving relatively more air with less power consumption. Smaller, "off the shelf" fans also face several maintenance issues, as they are generally not built as reliably. Even comparing to a large fan with similar reliability, many more small fans will be required for a similar workload, and the system will generally need more frequent maintenance.

Cray Liquid Cooling Benefits

In an examination of the tangible benefits of the various available systems, three basic systems will be compared. These systems will be equivalent in terms of compute performance and total system wattage. The systems will only vary in terms of the cooling used. We will look at a large 50 cabinet system where each cabinet draws 35kW for a total of 1750 kW, and 550 teraflops (550 trillion floating operations per second) of compute performance. All systems will be evaluated on a standard chiller plant operating at 1.25kW per cooling ton. This chiller specification may cause some PUE calculations to be higher than estimations based on high efficiency chiller plants, and are most useful as a comparison rather than an absolute value. All HPC systems will be compared on a high efficiency 90% power conversion basis.

System 1: An efficient white-box, HPC system will use 32 small fans to air cool the cabinets horizontally.

System 2: The Cray XT5 system with the HE cabinet design with air cooling only.

System 3: The Cray XT5 system with the HE cabinet with ECOphlex cooling technology.

First we must compare the need for CRAC units to cool the various systems, using assumptions for high efficiency CRAC units described previously. Both air cooled systems would require twenty CRAC units. Using Cray ECOphlex technology up to two CRAC units may be necessary for overall room humidity control or excess heat generated by other devices (storage, etc).

CRAC units will also have a major impact on system footprint when cooling high powered machines. At about one CRAC for every 2.5 air cooled 35kW compute cabinets, the datacenter footprint for CRAC units will be larger than the compute footprint when the restricted air flow area around the CRAC is accounted for. Each CRAC unit accounts for approximately 48 square feet compared to 35 square feet for the corresponding compute power and access aisles.

Refrigerant pumping units represent a significant cost savings opportunity. 10 RPUs are needed for a liquid cooled ECOphlex system of this size generating the same 1.75 MW of heat.

RPU units represent a large decrease in system footprint when compared to CRAC units. At about one RPU for every five 35kW machines, and with approximately the same area as a single cabinet, the cooling footprint will be only 20% of the full system footprint. This represents a very significant savings in data center floor space. In effect, the data center space occupied by cooling infrastructure decreases by over 80% when compared to the air cooled system.

When a variance in blower design is considered, a noticeable difference in power consumption can also be seen. A single custom fan design uses about 20% less fan power than the equivalent cooling from 32 smaller “off the shelf” fans. This results in a net savings of 35 kW over a 50 cabinet machine or \$40,000 savings over an average 8740 hour work year at 10 cents per kWh. The additional annual savings from reduced CRAC usage and efficient RPUs comes to \$125,000 for a total savings of \$165,000 compared to a standard air cooled HPC system. This is a savings of 20% for the cooling power, or 7% of total system energy against a modern HPC air cooled design.

To further illustrate the importance of computer cooling efficiency, a final comparison is made to a typical cluster system assuming the same performance and power efficiency, with the only variable being cooling efficiency. For this example, the cluster is assumed to have a PUE of 2.0, which is fairly typical. When compared to a 550 teraflop cluster, the savings would be nearly \$800K annually. This does not even take into account the power inefficiencies and much larger foot print associated with “off the shelf clusters” which can take up three times as much space and operate at 20% lower power efficiency.

System	Energy Cost Savings per Year	Percentage Savings	PUE
Standard 2.0 PUE Cluster	base	base	2.00
System 1; 32 fan	\$616,619	43%	1.66
System 2; 1 fan	\$656,771	45%	1.64
System 3; 1 fan LC Cray	\$781,815	54%	1.55

Table 1:

Savings on cooling energy between several HPC cooling schemes, compared to a standard cluster

Because the air cooled systems rely so heavily on the CRAC units, CRAC efficiencies have a large effect on system performance. Comparing the air and liquid cooled Cray systems as CRAC efficiency varies shows that the difference in PUE between the systems increases exponentially as CRAC efficiency decreases. The ECOphlex system maintains an almost constant PUE at low levels of CRAC performance, while the air cooled system’s power consumption skyrockets from the compounded issues of running more CRAC units and requiring more chilled water.

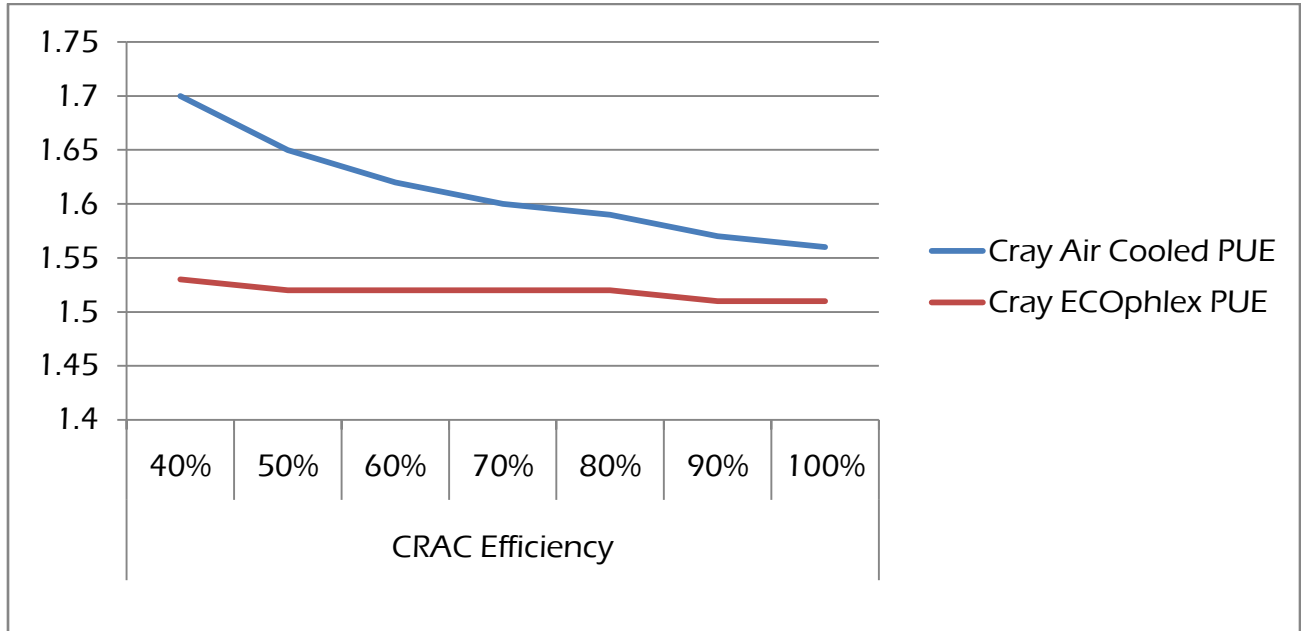


Figure 4: Air Cooled vs. ECOphlex Systems; PUE as function of CRAC Efficiency

Fan design variations also show a large benefit for system maintenance needs of the ECOphlex design. With a MTBF of 80,000 hours for a custom blower in the Cray rack, and a MTBF of approximately 20,000 hours for an “off-the-shelf” fan, the fact that the Cray design only uses one fan (or 2% as many fans as a standard system) means the blower system on a standard HPC assembly will need to be serviced 200 times as often as a Cray ECOphlex rack.

With the Cray liquid cooled ECOphlex technology, an additional flexibility is the ability to run the inlet water temperature to the RPU at higher water temperatures (e.g. 60 degrees Fahrenheit), depending on the data center location and cooling strategy. This would be dependent upon such factors as datacenter ambient temperature. The CRAC units typically require 45 degree or cooler inlet water temperatures. So this is yet another place in the cooling system that can result in very substantial energy savings for the cabinets incorporating Cray’s ECOphlex technology. For the 50-cabinet system in the comparison example, the \$130K annual savings could easily grow to \$500K by taking advantage of warmer inlet water temperatures, which in turn can drive chiller plant savings. When combined with data center design best practices, this design could eliminate the need for a chiller entirely. Without the need for a chiller, power usage would plummet, and datacenters could possibly achieve the highly sought after goal of free cooling.

Conclusion

Increasing energy concerns have forced computer users and vendors alike to search for efficiency wherever it can be found. An array of technologies are available to design more efficient cooling systems. In lieu of complete facilities design one must find the most efficient way to get heat to the building water, while still protecting costly high performance computers and ensuring a good operating environment. Cray combines an innovative phase change process with efficient air movement to remove heat. The result is excellent system density and energy efficiency, which greatly reduce the total cost of ownership and allow for upgrades well into the future with little to no change in power and cooling infrastructure.

Acknowledgement

This material is based upon work supported by the Defense Advanced Research Projects Agency under its Agreement No. HR0011-07-9-0001. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Defense Advanced Research Projects Agency.

© 2009 Cray Inc. All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior permission of the copyright owners.

Cray is a registered trademark, and the Cray logo and Cray XT are trademarks of Cray Inc. Other product and service names mentioned herein are the trademarks of their respective owners.

Corporate Headquarters
901 Fifth Ave, Suite 1000
Seattle, WA 98164
Phone: 206.701.2000
Fax: 206.701.2500