Saving Energy with "Free" Cooling and the Cray XC30

Brent Draney, Jeff Broughton, Tina Declerck and John Hutchings
National Energy Research Scientific Computing Center (NERSC)
Lawrence Berkeley National Laboratory
Berkeley, CA, USA

***Abstract*—Located in Oakland, CA, NERSC is running its new XC30, Edison, using "free" cooling. Leveraging the benign San Francisco Bay Area environment, we are able to provide a year-round source of water from cooling towers alone (no chillers) to supply the innovative cooling system in the XC30. While this approach provides excellent energy efficiency (PUE ~ 1.1), it is not without its challenges. This paper describes our experience designing and operating such a system, the benefits that we have realized, and the trade-offs relative to conventional approaches.**

Keywords- free cooling, energy efficiency, data center, high performance computing

I. INTRODUCTION

The National Energy Research Scientific Computing Center (NERSC) at the Lawrence Berkeley National Laboratory (LBNL) is the primary scientific production computing facility for The Department of Energy's (DOE's) Office of Science. With more than 4,500 users from universities, national laboratories, and industry, NERSC supports the largest and most diverse research community of any computing facility within the DOE complex. NERSC provides large-scale, state-of-the-art computing, storage, and networking for DOE's unclassified research programs in high energy physics, biological and environmental sciences, basic energy sciences, nuclear physics, fusion energy sciences, mathematics, and computational and computer science.

NERSC recently acquired the first Cray XC30 system, which we have named "Edison." Since its founding in 1974, NERSC has a long history of fielding leading-edge supercomputer systems, including an early Cray-1, the first Cray-2, the Cray T3E-900, and more recently "Hopper," one of the first XE-6 systems. The XC30 is the first all-new Cray design since Red Storm. It incorporates Intel processors and the next-generation Aries interconnect, which uses a dragonfly topology instead of a torus. The XC30 also utilizes a novel water-cooling mechanism that can operate with much warmer water temperatures than earlier supercomputers. This novel cooling method is a focus of this paper.

II. USING THE ENVIRONMENT TO REDUCE ENERGY USE

NERSC is currently located in Oakland, CA at the University of California Oakland Scientific Center (OSF). Occupied in 2001, OSF was a modern data center for its time. It used two 800-ton chillers to provide chilled water for air conditioning and direct liquid cooling, and was able to achieve a PUE (power usage effectiveness[1]) of 1.37. That is, the energy overhead for cooling the data center was 37% of the power consumed by the systems. Through a sustained campaign of energy efficiency optimizations as well as improvements in system design, OSF has reached a PUE of 1.23. Continued improvements in energy efficiency are critical to NERSC's future plans to deliver exascale-class systems. These systems are expected to consume tens of megawatts of power, so even modest improvements in the power required for cooling can translate to megawatts and mega dollars saved per year.

Our new data center, the Computational Research and Theory Facility (CRT), is now being constructed on the LBNL main campus in the hills of Berkeley, CA. Occupancy is planned for early 2015. CRT is being designed to provide the kind of world-class energy efficiency necessary to support exascale systems.

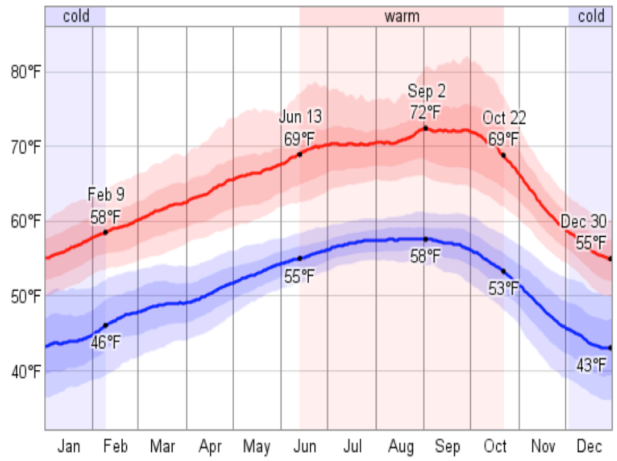

Figure 1. The daily average low (blue) and high (red) temperature for Oakland with percentile bands (inner band from 25th to 75th percentile, outer band from 10th to 90th percentile). Data from weatherspark.com.

[1] PUE or power usage effectiveness is technically defined as total facility power divided by IT equipment power. The primary non-IT component of total facility power is energy to run cooling equipment but also includes other factors such power distribution losses, lighting, etc. Because of limitations on our instrumentation and estimation models, we are not able to calculate it exactly in all cases. In most cases in this paper, we calculate PUE as only cooling power divided by IT load starting at 480V distribution panels. This approach may understate actual PUE by as much as 5%.

The San Francisco Bay Area, which includes both Oakland and Berkeley, has a climate strongly moderated by the bay. Temperatures remain relatively cool year-round, and on those occasions when temperatures rise above the 70s, humidity stays low. We have cool foggy days, but not hot muggy days. As Mark Twain is purported to have said: "The coldest winter I ever spent was a summer in San Francisco."

The design of CRT leverages our benign environment to provide year-round cooling without mechanical chillers, which consume large amounts of power. On the air side, the building "breaths", taking in outside air to cool systems and expelling the hot exhaust generated by the systems. Hot air can be recirculated to temper and dehumidify the inlet air. (Heat is also harvested to warm office areas.) On the water side, evaporative cooling is used to dissipate heat from systems. Water circulates through cooling towers, where it is cooled by evaporation. This outside, open water loop is connected by a heat exchanger to a closed, inside water loop that provides cool water directly to the systems. The water loop additionally provides cooling for air on hot days.

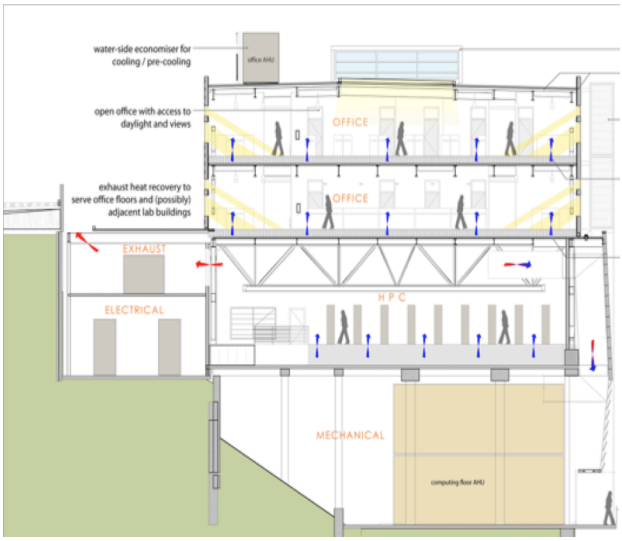

Figure 2: Cross section of NERSC's CRT data center. The bottom mechanical floor houses air handlers, pumps and heat exchangers. The second floor is the data center. The top two floors house offices.

Because it uses the natural environment and not power-hungry mechanical chillers, this approach is called ***"free"*** cooling. Of course, it isn't really free. Power is still used for fans and pumps to circulate air and water. However, the total power needed is less than 1/3 of that required for the best chiller installations. Heat recovery can further offset energy costs.

Many data centers utilize free cooling when the conditions are amenable, and fall back on chillers when temperatures and humidity rise. For example, free cooling may be used in the winter but not in summer, or at night but not during the day. In the Bay Area conditions are favorable all year. In the worst conditions, which occur *only a few hours per year*, CRT can provide 74°F air and 75°F water. For this reason, NERSC has decided to attempt to forgo chillers altogether. As a result, the maximum PUE at CRT is predicted to be less than 1.1 for a likely mix of equipment.

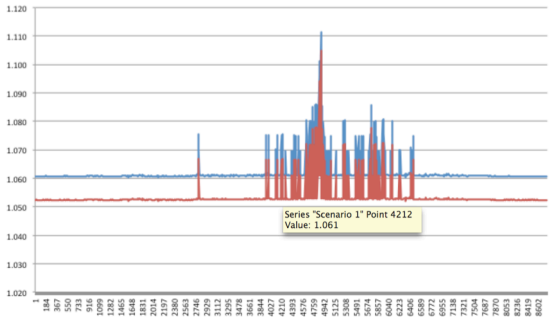

Figure 3. Predicted PUE in CRT with two different mixes of air and water cooled systems.

The Edison system has been installed at OSF, and will move to CRT when the building is ready to be occupied. For this reason, a primary requirement was that it be able to operate within the 75°F water envelop of CRT. The new cooling design of the XC30 is able to meet this requirement. As we will describe below, we also had the opportunity to deploy Edison with free cooling in OSF, allowing us to gain experience with the technique before moving into CRT.

III. Understanding your system's needs

Before installing any major computing system, there must first understand the environmental requirements needed for continuous operation. Air cooled systems have three basic requirements that need to be met: power, air flow, and air temperature. These requirements are normally aggregated per rack and are measured in Kilowatts (kW), cubic meters per hour (CMH) or cubic feet per minute (CFM), and Centigrade (C) or Fahrenheit (F) depending on your location and equipment manufacturer.

Liquid cooled systems such as a Cray-2 or Cray T3E have the same three basic environmental needs but with a different fluid medium, power, liquid flow, and liquid temperature. A fourth requirement of pressure differential (ΔP) measured in pounds per square inch (PSI) or Pascals (Pa) is added to insure correct liquid flow.

Both air and liquid cooling methods are thermodynamically the same but since water has a specific heat four times that of air and a volumetric heat capacity 4,000 times greater, it becomes possible to support power densities with a liquid cooled system that are infeasible with air cooling. The downside to liquid cooling is that the mechanical support systems are more expensive and special fluids (e.g. Fluorinert, pure water) may also be required. Also liquid cooling makes servicing equipment more complex. These costs and complexities led to the demise of liquid cooling in the 90's in favor of air-cooling.

As the power densities increased in the late 2000's liquid cooling entered a renaissance when air cooling

became a limiting factor for both systems and data centers. Immersion cooling is back on the table with systems like (TACC), liquid cooled heat sinks are now used in PERCS systems, and there are a vast array of near liquid cooling solutions using rear door radiators and intercooler coils that create and air/liquid hybrid system.

The Cray XC30 is one such hybrid system with some unique characteristics. Hybrid systems greatly simplify serviceability and reduce the costs compared to liquid systems, but increase the number of environmental requirements a facility must provide combining both the air and liquid sets (i.e. power, air flow, air temperature, liquid flow, liquid temperature, and pressure differential).

Water quality is a key factor in systems that use water for cooling. At one extreme, systems such as the IBM BlueGene/Q, may require ultra-pure, reverse osmosis, (RO) or de-ionized (DI) water. Use of DI water requires class of plastic pipes. Other systems may accomodate "tap" water with only limited treatment to control mineral content, corrosion and biological contamination. Conventional steel, copper or brass piping and other components can be used.

IV. Cray XC30 Environmental Requirements

The Cray XC30 system has an innovative cooling mechanism. Instead of a more typical front to back airflow this system uses a side-to-side airflow where the exhaust air from one rack becomes the intake air of the next. Each cabinet contains a water intercooler (radiator) on the outlet side. This transfers heat from the air blowing through the cabinets to the water loop and cools air delivered to the next cabinet in the row. Fan cabinets are placed at the intake and outlet ends of the row as well as interspersed between pairs of two compute racks.

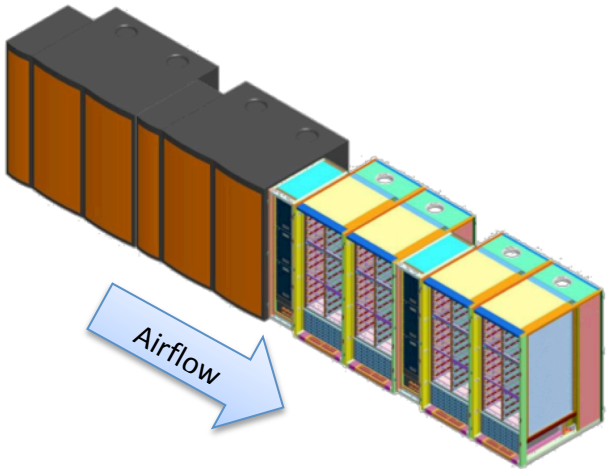

Figure 4: Diagram of a Cray XC30. Air flow is through the cabinets from upper left to lower right. There is a blower at the start of the row and after every pair of cabinets. At the exit of each cabinet is a vertical radiator (intercooler, light blue) that cools hot air exiting the cabinet and entering the next.

Flowing the air from side-to-side has unique advantages. First, additional cabinets in a row do not require more air from the computer room than was supplied to the first rack. Less overall air requirements means fewer building air handlers and greater overall center efficiency.

Additionally, the surface area of the side of the rack is greater than the front and the width of the rack is less than the depth. Both of these contribute to less fan energy required to move the same volume of air through a compute rack. With the air moving more slowly across an intercooler a greater amount of heat is extracted per volume. This leads to a closer temperature approach (or simply, approach) between the water used to cool the system and the air that is being cooled, and a greater change in temperature (ΔT) of both the air (cooling) and the water (warming) through the system.

Approach is the difference between the water temperature and the exiting air temperature. The slower the air moves through the coil (more time to transfer the heat) and the more efficient the coil is the smaller the approach will be. Approach can get to a few degrees on a very efficient system but can never get to zero. Depending on the direction of the water through the intercooler, the approach can be the difference between the exiting water and the exiting air (parallel flow) or the entering water and the exiting air (counter flow). Since the entering water is always colder than the exiting water it will be more efficient to have the water flow in the opposite direction of the air. Counter-flow is required for efficient free cooling.

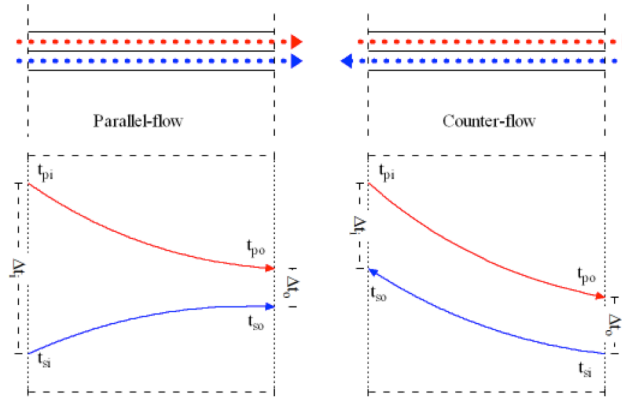

Figure 5: Entry and exit temperatures of parallel and counter-flow designs for water intercoolers. "Approach" is the difference between the water (blue) temperature and exiting air (red) temperature.

V. Understanding Your Environment

Our Cray XC30 was targeted to occupy the floor space previously occupied by a retired Cray XT4 (Franklin). This section of the floor was cooled by a set of chillers and cooling towers in the mechanical yard adjacent to the computer floor. While these chillers and cooling towers had sufficient cooling capacity to meet the cooling needs of the new system Edison, the chillers were designed to accommodate a system requiring much lower temperatures. To support Edison, it would have been necessary for the

cooling plant to run 10 degrees above the maximum design temperature of the chillers (65°F based on the Cray control algorithms). Alternatively, we could have modified our single (primary) loop system to convert it to a primary/secondary configuration to work around the limitation that the chiller could only reduce the water temperature by 10°F in a single pass. Since, it was obvious that some modification to the cooling plant was needed to support the Cray XC30, the challenge was finding the right design.

Based on the environmental capabilities of the Cray XC30 and the unique operating environment of Oakland and Berkeley, NERSC commissioned a study to determine if chiller free cooling was both practical and economical for Edison. The pertinent specifications from Cray, as well as the specifications for the existing cooling plant were given to an engineering firm. The firm analyzed the equipment and compared it with the Typical Meteorological Year 3 (TMY3) data set for Oakland, California.

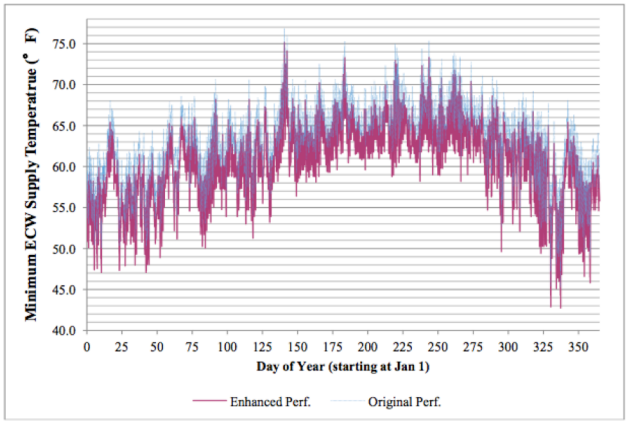

Figure 6: Maximum water temperature delivered to the Cray XC30. Blue signifies the temperature with a single cooling tower; red, with two towers operational.

Since cooling is achieved by evaporation of water in cooling towers, the outside wet-bulb temperature is the primary factor driving the delivered temperature. Wet bulb temperature is the temperature when air is cooled to saturation (i.e. 100% relative humidity). When air is dry, the cooling effect can be significant. For this reason, cool water can be generated even when the outside temperature is very hot.

Mechanical efficiency also affects the delivered temperature. With a 12°F Δ T on the cooling towers achieved by running two cooling tower cells in parallel and assuming a 2° approach on a heat exchanger the engineering firm calculated the expected number of hours a year that a chiller free cooling system would exceed the design point of 75°F. We were delighted to find out that we should expect to exceed 75°F about 1 hour a year and that 98.5% of the time we should be at or below 70° F. This analysis made the idea of a chiller free cooling system, also known as a Water Side Economizer (WSE), quite feasible.

TABLE I. HOURS PER YEAR OF HIGH WATER TEMPERATURES.

Supplied water temp	<70F	70-71F	71-72F	72-73F	73-74F	74-75F	>75F	
One Cell	8258	261	101	64	45	19	12	Hours per year
Two Cell	8636	63	32	19	7	2	1	

(In practice, we have learned that using two heat exchangers in parallel—in addition to the two cooling towers—lowers pressure loss (drop) for both open and closed loops pumps and results in approximately a 1 degree approach. This improves efficiency and may lessen the number of hours that the closed loop is in the upper temperature ranges.)

VI. THE ECONOMICS OF CHILLER FREE COOLING

Once convinced that removing chillers from the equation was technically feasible we had to decide if it was economically feasible— there was a reasonable payback period within the life of the system. The same engineering firm was asked to perform a power study on the components in both options. A bid estimator assembled projected install costs and Berkeley Lab power recharge rates gave us initial operating costs. Power savings were calculated at 80% or 216 kW per year, and at $0.10/kWh that adds up to $189,200 in savings per year!

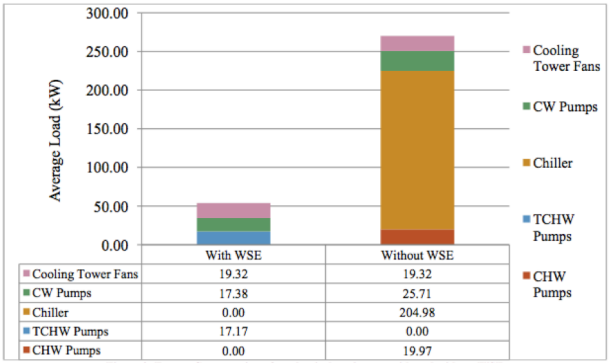

Figure 7: Comparative energy usage with a water-side economizer (free cooling) and without.

The bid estimates plus LBNL burden and management fees for converting to chiller free cooling came to $665,141. That would give us a simple payback on the investment of 3.5 years. This was within the life of the system but not within the expected length of time that the system would remain at OSF. Two other factors helped tip the scale in the direction of chiller free cooling. First, the current plant would have to be modified to include an additional set of pumps to create a primary/secondary loop at a cost of about $200k. The second and most fortuitous was a Pacific Gas & Electric (PG&E) energy efficiency project rebate program with an initial estimate of about $125k. With the difference in costs and the potential for the rebate, the payback period would be cut in half and the chiller free cooling would pay for itself while Edison was installed at

OSF. When we received the written rebate estimate based on the design it totaled an incredible $416K. That would reduce the payback to about three months! The rebate will not be official rebate until PG&E completes verification.

VII. Lessons Learned, Observations and Things to Consider

Conventional HVAC coils are designed to work with a 5-10 psi ΔP with chiller temperatures closer to 55°F. As the water temperature increases, the amount of water needed goes up substantially to keep the air at set point. This leads to substantially higher-pressure requirements as the flow resistance goes up with the cube of the velocity. As the water inlet approaches the coil outlet temperature the flow requirement grows asymptotically. A mechanical engineer should calculate the pressure drop for your entire system to insure that the maximum flow rate can be met with your pumps.

The Cray XC30 comes with special food grade stainless steel cooling connectors, the same type that you would see used in the Midwest dairy industry or in Napa Valley wineries. The connectors are custom welded to threaded end fittings that attach to the cooling pipes. These stainless end fittings are hard to seal because the threads are softer and can deform under torque. Plumbers must be careful and use pipe dope with Teflon tape to get a leak free seal. Apply the pipe dope carefully and make sure to properly flush the cooling system before use.

The Cray XC30 can operate in one of two modes: room neutral and fixed set point. In room neutral mode, the system attempts to maintain an exhaust temperature from the last rack in the row equal to the input temperature of the first rack. With a fixed set point, the system attempts to maintain a fixed exhaust temperature. The XC30 is set by default to room neutral. We discovered this by accident, as we went through experiments to raise the water temperature to the upper limits of our operating range. In OSF, most of the machine room is cooled with chillers and can maintain a lower ambient temperature than the free cooled side. As a result, the XC30 struggled to produce room neutral exhaust when the supplied water temperatures were high, and as described above, the required flow rates became unusually high. Cray service personnel modified the system for a fixed set point slightly above room temperature to address the issue.

The Cray XC30 control systems have some safety margin and will automatically adjust fan speeds and water flow upwards if they detect a hot spot. There is not currently an automatic mechanism to return the controls back to their original set point. Moreover the control parameters are lost across a reboot and must be manually maintained. This includes whether or not the system is in a room neutral mode or a fixed set point mode.

The temperature in the OSF computing center tends to stratify as with most data centers. This is evident in the nodes in the upper chassis running about 2°F warmer than those in the bottom chassis. This is not currently an issue but it can be addressed by sealing the air intake to the floor similar to the method used with a BlueGene system. Other methods would be to add a precooler to the Cray but we have found a small number of standard air handling units sufficient to control humidity and reduce the exit air temperature by an additional 3 degrees to prevent condensation.

A future area of investigation is interaction between the Cray XC30 and the building control system (BCS). Currently, an XC30 rack modulates its exhaust temperature by adjusting the water flow through the intercooler. If the water temperature is higher, more flow is required to maintain a desired exhaust temperature. The BCS sees this as a demand for higher ΔP, and must increase pump speed to compensate. However, there is a delay of several minutes in responding because of the length of pipe between the system and the pumps. Temperature and pressure in the cooling system and XC30 may oscillate as a result. Direct feedback from the XC30 sensors to the BCS could mitigate this effect.

Additionally, the BCS should adapt to the workload and environmental conditions. The energy consumed by microprocessors (with a constant workload) increases with ambient temperature due to leakage current within the chip. To minimize power consumed by the processors, it is desirable to keep the system as cool as possible. This is in conflict with the desire to minimize energy used by the cooling system by operating at higher temperatures. Ideally, we will identify optimal water and XC30 exhaust air temperatures for certain processor power levels (workloads) and maintain those weather permitting. Both the BCS and XC30 environmental control system need to participate.

VIII. Water Connections Versus Seismic Isolation

The Bay Area is well-known to be a seismically active area. To protect both people and equipment from damage in an earthquake, we provide seismic isolation for all our systems. In Oakland at the OSF, we use ISO-Base from WorkSafe Technologies. With this approach, system cabinets are placed on top of platforms that ride on ball bearings. When an earthquake occurs, the cabinets stay relatively still due to inertia, while the building and floor move underneath it. The ball bearings sit in a dish that limits travel to 8 inches and provides a self-centering action.

Figure 8. ISO-Base ball and cone seismic isolation platform.

Power, network and water connections are a challenge for systems placed on ISO-Base. Care must be taken to ensure that there is adequate slack and travel for the connections at maximum displacement. Otherwise, the

cable may bind the movement of the system or be cut by a guillotine effect. Water connections are especially tricky. Of course, flexible hoses must be used. Since hoses are much thicker that even power cords, their width had to be considered when calculating slack and displacement.

Unfortunately, when Cray designed the XC30, they did not consider how the systems would interact with seismic isolation. The bottom of the system is flat and is designed to distribute weight over a raised floor; power and water connections are made through small openings in the cabinet base. There is no room for these connections to wiggle. In addition, the width of two system cabinets together and one blower cabinets is 3.75 tiles -- meaning that the point of connection varies from tile to tile across a raised floor. With a stable floor, there is no interference with raised floor stringers. But with eight inches of travel, there can be interference.

NERSC and Cray worked together to design a solution to this problem. All connection points are in the rear of the cabinet, behind the backplane and blade chassis and substantially behind the center of gravity. As a result, it was possible to cantilever the rear 14 inches of the cabinet off the back of the ISO-Base to make room for power and water connections. A double-high top plate was also designed to provide adequate clearance. Wide cutouts were made in the raised floor to give play for the water hoses, but the location of the cutouts varied down the row. In some cases, hoses ran through a hole cut underneath the neighboring cabinet.

At the CRT, we have designed a seismically isolated floor in which the whole floor will move when an earthquake occurs. Cabinets will sit directly on the floor and will not move relative to the floor as on an ISO-Base. This eliminates potential problems with cabinet placement and cutouts, and under floor space is adequate to cover connection displacement.

IX. Results and Projections

Our operational data is limited at this point. The four-cabinet Edison subset ("Phase 1") has only been installed since December. Power and cooling needs are much less than the final system, and we only have experience in the cooler part of the year. Nonetheless, we can say that the observed performance of the system has been quite good.

By replacing our earlier Cray XT4 ("Franklin"), Edison has had a material impact on the operation of the center. Franklin required chilled water at 41°F, and required us to run the NERSC facility quite cold. Franklin's removal has enabled us to incrementally increase the chilled water temperature to 48.5°F. (Additional increases are not expected to improve energy usage as the chillers are optimized for lower temperatures.) Room air temperatures have also been increased (from approximately 55°F to over 70°F) because our air-cooled midrange systems can operate at ASHRAE temperatures.

Edison promises substantially improved energy efficiency over that of our current systems. Ignoring cooling for the moment, it will use 30% less energy than Hopper and provide approximately 70% more computational throughput for an improvement in performance per watt of nearly 140%. The added savings from free cooling only compound these benefits.

We have been pleasantly surprised with how efficient the system is running actual workloads. Cray originally told us to expect that the system would use 84KW per cabinet at peak (i.e. Linpack), and "something in the 70s" day-to-day. We have observed the Phase 1 system to run at an average power of 55KW per cabinet with a peak of 62KW for the whole month of March. During an earlier Linpack run, the system managed to reach only 71KW per cabinet. We believe that the achieved power levels are due to use of low-power processor bins, aggressive power management in the Intel processors and the effectiveness of the cooling technology.

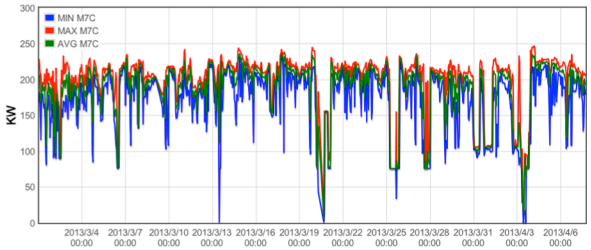

Figure 9: Minimum (blue), maximum (red) and average (green) power readings on an hourly basis during March, 2013 for Edison Phase 1.

Free cooling provides a substantial improvement in PUE over our current systems. With only Hopper and midrange systems operational, the PUE at OSF prior to the installation of Edison averaged approximately 1.23. As shown in the figure below, the PUE varies due to many factors, but primarily system load and non-linear behavior in performance of the cooling plant. PUE is defined in such a way that increased system power consumption without a proportional increase in cooling plant power will cause the PUE to *decrease*, and vice versa. Thus, it is actually harder to improve PUE with systems that are more efficient in terms of performance/watt.

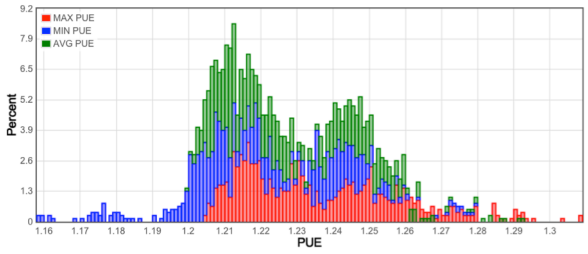

Figure 10. Distribution of minimum (blue), maximum (red) and average (green) hourly PUE readings at the OSF for a one month period in Oct/Nov, 2012. This represents Hopper and midrange systems with cooling from chillers.

Table II shows the power consumed and PUE for Phase 1 (measured) and for the full Phase 2 system (estimated). For comparison, the measurements for Hopper are also

shown. Note that in the XE6, XT4 and most conventional servers, the fans are powered from the same power feeds as the processor and memory, and are treated as part of the IT load for purposes of computing PUE (because they cannot be separated out). In the XC30, the blowers run off a separate power circuit from the processors, and we can account for its usage separately. We show two values for PUE: one with blowers included in the system (IT) load (for comparison with other systems) and one with the blowers included in the mechanical load (which more correctly represents "true" PUE).

TABLE II. EDISON POWER AND PUE

	Phase 1 (measured)	**Phase 2 (projected)**	**Hopper (measured)**
System Cabinets	220 KW	1,540 KW	2,215 KW
Login Nodes and Storage	66 KW	69 KW	125 KW
Blowers	15 KW	90 KW	Incl. in system
Cooling Plant	38 KW	80 KW	538 KW
PUE (Blowers in IT load)	1.116	1.047	1.23
PUE (Blowers in mech. load)	1.175	1.106	N/A

Notice that the power to cool the system does not increase substantially from Phase 1 to Phase 2. That is because most of the energy is expended moving water through pipes. Currently, water connections for all 28 cabinets are open, and the input and output connections for missing cabinets are "jumpered" together. This was done initially to ensure that debris had been flushed from the system and then to exercise the pumps prior to Phase 2 installation.

X. CONCLUSION

Even at this early stage, the use of free cooling with the very efficient Edison system has resulted in positive improvements in energy utilization at OSF. Air temperatures within the machine room have been increased; computation power efficiency has been improved; power required for cooling has been reduced; and costs have been lowered. Significant further improvements will be seen when the full system is installed. We estimate the center-wide PUE (representing a mix of Edison, Hopper, and less energy efficient midrange systems, networks and storage) will drop below 1.18.

Our experience with Edison Phase 1 has been valuable in understanding how to operate in a free-cooling environment, before we move into the CRT building. As a result, we have increased pump capacity in CRT and have added requirements for flow rate, differential pressure, and water quality to RFPs for future systems.

Much can still be learned. The coming summer months will reveal the highest tower water temperatures that we will experience, and the larger, full-scale system will tax the cooling plant to the fullest extent.